Working Paper

Modelling Urban Networks

Sustainable Progress

Project CP2018_6.1.2

December 2018

Title page

Modelling Urban Networks Sustainable Progress

Joan Marull[a], *Mercè Farré*[b], *Rafael Boix*[c]

[a] Barcelona Institute of Regional and Metropolitan Studies, Autonomous University of Barcelona, E-08193 Bellaterra, Spain.

[b] Department of Mathematics, Autonomous University of Barcelona, E-08193 Bellaterra, Spain

[c] Department of Economic Structure, University of Valencia, Avda. dels Tarongers, E-46022 Valencia, Spain

Index

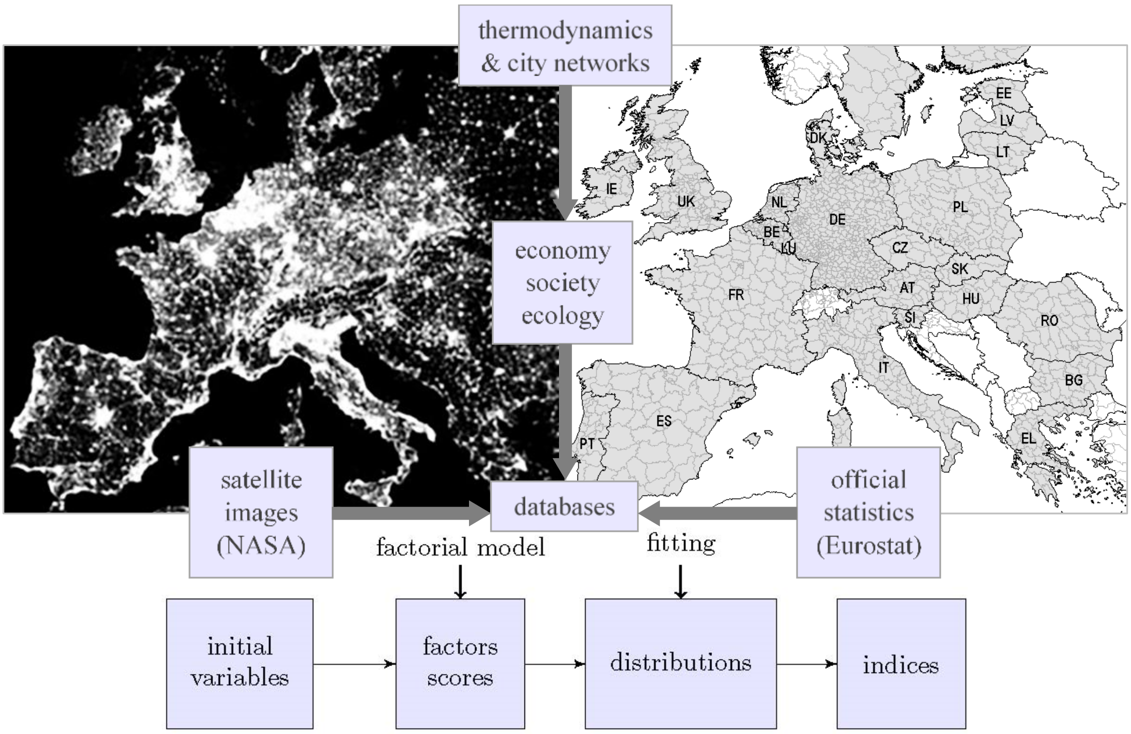
Abstract

In this paper, we analyse the relations between thermodynamics and city networks: an increase in the complexity and the organized information in such urban systems leads to less demand for resources and less social entropy, which overall makes them more efficient and stable. The goal of this study is to propose a method to measuring city networks sustainable progress based on statistical models, derived from Eurostat databases and NASA satellite images, and capable of analyzing different conceptual scenarios of urban development in Europe. The obtained probability-based indices enable us to evaluate the dynamics of city networks in terms of three components of sustainable progress – economic activity, social cohesion and urban ecology – and help us understand the properties that a regional and megaregional economy must possess in order to optimize the inclusive development of the urban system. The results have implications for pro-active policies, and for urban and territorial planning at supra-local level.

Keywords

Beyond GDP, inclusive development, city networks, factor analysis, probability-based indices, Europe

Graphical abstract



Highlights

- Sustainable urban system progress from a thermodynamic point of view
- Development of a reproducible urban system sustainable progress model
- Probably-based indices evaluate the sustainable progress of urban networks
- Impact on regional and megaregional planning and land-use policies

1. Introduction

Urban growth has been dominated by the acceleration of the processes of urbanisation and the development of city networks (Camagni, 1993), and has given rise to complex, large-scale trans-metropolitan structures (Lang and Nelson, 2009). The level of organisation at megaregional scale could lead to an acceleration in global change (Grazi et al., 2008) due to a concentration of a large amount of the planet's productivity and innovation, which is associated with higher per capita income and greater creativity (Ross, 2009); however, it could also increase metabolic inefficiency and social inequality. Conversely, the traditional approach focused on GDP and per capita income (Kuznets, 1934) is not a good indicator of urban progress and contributes to social instability and environmental deterioration (Wilkinson and Pickett, 2009).

Two important transformations have occurred and become consolidated during the first years of the 21st century: i) a change in scale in urban systems, which are now structured in megaregions consisting of dense networks of interconnected cities (Ross, 2009); ii) an awareness of the limitations of the traditional notions of growth and sustainable development (Costanza et al., 2009) when confronting the complexity of human organizations within these new spatial units. The sheer size of these new megaregional urban systems, as well as the internal complexity of connections between cities (Marull et al., 2015), raises doubts as to whether or not the multiple interrelated dimensions of cities at megaregional scale are sustainable (Wheeler, 2009); likewise, questions exist as to whether or not traditional notions are still valid when evaluating cities' sustainability or it would be better to move towards complex system analysis (Baynes, 2009).

Theoretically, it would seem that a change in the scale of urban systems should lead to a greater consumption of resources, lower environmental quality and more social entropy.

Nevertheless, the experience of complex systems suggests that such systems are subject to the laws of thermodynamics (Pulselli et al., 2006) and, in fact, the more complex they are, the more efficient they become and the fewer resources they consume. Nevertheless, one question that has received less attention in the literature (Campbell, 2009) is: May megaregions become more efficient in resource consumption and human well-being?

In this work, we explore the hypothesis that megaregions are an organisational response by cities structures to be more efficient and to reduce their consumption of resources, which will also help improve levels of internal social cohesion and economic performance. To study this hypothesis, we have developed new indices based on a factorial model that enable us to measure sustainable progress in city networks based on four conceptual scenarios: i) *economic development* (dominant economic model); ii) *social sustainability* (mainly based on social equality); iii) *environmental sustainability* (mainly based on resource consumption); iv) *inclusive development* (taking into account a balance between economic, social and ecological factors).

The concepts and tools used habitually to measure development in cities are only of limited use for understanding the sustainable progress of city networks, above all when they join to create megaregions. The combination of official data (Eurostat) and images from satellites (NASA) enabled us to identify the megaregions that exist in Europe and estimate the appropriate indices during the period of analysis (between 1995 and 2010).

This study introduces three relevant innovations into current models analysing change in European regions (Brandsma et al., 2015; Capello et al., 2017; Varga, 2017): i) it delimits and uses a unit of analysis that reflects the socioeconomic and ecological reality in a more realistic way than other studies and models; ii), it introduces the megaregion as a supra-regional unit of which the region is a part, unlike current models that use the country as a basis; and iii) it takes into account an ecological perspective that current models usually ignore since they concentrate on economic variables.

The objective of this paper is to propose a method based on a statistical and probabilistic model, derived from Eurostat databases and NASA satellite images, capable of analyzing different conceptual scenarios of urban networks sustainable progress in Europe.

The paper is structured as follows. After the introduction, Section 2 associates the idea of complex urban systems with city networks in megaregions and introduces the notion of sustainable progress for the study of urban networks. Section 3 elaborates a factorial model of sustainable urban networks progress that leads to the definition of indices to quantify different scenarios based on economic development, social sustainability, environmental sustainability, and inclusive development. Section 4 presents the results for the European regions and megaregions. Section 5 concludes.

2. Urban complexity and sustainable progress

2.1. Urban complexity

2.1.1. The metabolism of the cities

In certain ways, cities resemble living organisms: they continuously exchange, process and store energy, matter and information with their surroundings, and their growth is limited by their energy needs and its availability. Like living creatures, we can talk about a city's metabolism (Pulselli et al., 2006) for as a city grows – just like a cell – its needs tend to increase quicker than the available resources, which puts a limit on its growth. To overcome these limitations in a thermodynamic sense, it is essential to increase energy efficiency, the amount of organized information and transportation possibilities.

The strategy that consists of increasing the complexity of cities without increasing the dissipation of energy is an alternative to the conventional urban development model that bases its competitiveness on increasing resource consumption (Wilson, 2009). Innovation has always been an important element in the growth of large cities: for example, the development of railway networks historically increased transport efficiency, while the use of steam machinery improved productive capacity. Albeit more complex than the thermodynamic systems most commonly analysed (not much more complex, however, that a cell), the visions of urban and territorial planners can be enriched by taking into account a thermodynamic perspective of an urban system (Filchakova et al., 2007).

2.1.2. City networks

Cities are not isolated entities but are rather interconnected and form part of a larger network. City networks have been defined as a series of interactive relationships between similar or complementary centres that foment the development of inter-dependent economies (cooperation and innovation) that through specialization (division of work and function) complement each other (Camagni, 1993). In regional politics, there is habitually a trade-off between economic growth and social-environmental quality (Batabyal and Nijkamp, 2009). Large urban conurbations should be the product of economic models based more on knowledge than on resource consumption (energy adds up, whereas information multiplies). This is the main challenge that sustainable progress in city networks has to overcome.

In these networks, cities benefit from the economic advantages that derive from both their urban setting and the efficiency of their complex of relationships (organized information). The importance of the change in scale is fundamental for achieving positive results in terms of economic efficiency and, probably, social cohesion and environmental quality. Thus, the object of analysis must go beyond the city and their respective metropolitan areas. A particularly relevant case occurs when interconnected networks of metropolitan areas join together to form units – megaregions – that operate on an even greater scale.

2.1.3. The formation of megaregions

Megaregions are global economic units that are the product of the expansion, coalescence and direct or indirect networking between centres of innovation, production and consumption (Florida et al., 2008). They represent a novel functional unit of analysis that has emerged as metropolitan areas expand and, not only grow and become denser, but also spread and join other metropolis (Ross, 2009). A megaregion can thus be defined as a polycentric supra-metropolitan network of cities.

Although the development of megaregions is based on the theories of the economies of agglomeration (Florida et al., 2008; Trullén et al, 2013) in which, traditionally, most attention is focused on economic growth, the scale of the inherent metabolic processes can lead to serious modifications in the surroundings and, in turn, can have an impact at global scale (Grazi et al., 2008). The equivalent of the economies of agglomerations can be achieved not only through concentrated and diverse economic and social structures but through the relationships that develop in city macropolitan networks, which should be referred to as 'spatially mobile' or 'network' economies (Trullén et al., 2013).

2.2. Sustainable progress

2.2.1. GDP and its limitations as a measure of well-being

At this point the question arises as to which is the best way of measuring the inherent advantages and disadvantages of this new organizational scale for cities. The traditional way of measurement since the first third of the twentieth century always involved calculating the city's growth, that is, its GDP or per capita income. This index takes into account essentially commercial transactions but ignores environmental impacts and social inequalities (Kubiszewski et al., 2013).

Yet, GDP and its growth do not explain how income is distributed in terms of production to the people who live in a particular city, or how it is related to other factors such as general well-being including health and the environment (Van den Bergh, et al., 2009). If we start from a premise that society should aim to increase people's quality of life in an fair and environmentally sustainable fashion, it is clear that a country's GDP is not a sufficiently accurate measure of human well-being, as a number of reports have shown including those published by the Commission on the Measurement of Economic Performance and Social Progress (Stiglitz et al., 2009), the Center for the Study of the Longer-Range Future (Costanza et al., 2009) and the European Commission's *Beyond GDP* initiative (http://ec.europa.eu/environment/beyond_gdp/index_en.html).

2.2.2. From GDP to sustainable urban development

One of the most popular alternatives to the GDP is the notion of "sustainable development", which applied to cities can be defined as a process of synergetic integration and co-evolution among the great subsystems making up a city (economic, social, physical and environmental), which guarantees the local population a non-decreasing level of well-being in the long term, without compromising the possibilities of development of surrounding areas and contributing by this towards reducing the harmful effects of development on the biosphere (Camagni, 2017).

A number of studies have proposed more accurate ways of measuring the notion of "sustainable development", based usually on suggestions made by panels of experts. Some of the best known such proposals include: United Nations Human Development Index and Social Progress Index, OECD Better Life Initiative, or the Inclusive Development Index promoted by the World Economic Forum. Currently, we can also found several approaches and frameworks for measuring sustainable urban development using indicators and multicriteria indexes, for example: Lynch et al. (2011), Abu Bakar and Chen (2013), Liang et al. (2016) or European Commission (2018). As Zegras et al. (2004) conclude, frameworks and indicators must face the relevant scale of analysis, information availability, and reflect political reality.

2.2.3. From sustainable development to sustainable progress in city networks

The notion of “sustainable development” is limited by the size and the resources available in the environment of the urban system in question (Borowy, 2014). It takes into account the internal structure and processes of an urban system but not how it interrelates and even integrates with other such systems. Popa et al. (2014) propose the notion of “sustainable progress” to take into account not only environments consisting of the functional structure of a system but also the assemblies of systems – that is, systems composed of systems. Following this latter approach, we can use the notion of sustainable progress in networks of cities, defined by how limits of sustainability in urban systems can be overcome via mobility and migration towards more favourable environments.

Sustainable progress in city networks is defined by the internally and externally cyclical increase in economic competitiveness, urban complexity, functional integration, metabolic efficiency, and overall social well-being. Megaregions are incorporated as large-scale macropolitan networks of cities. These city networks are complex adaptive systems, with multiple variables and dimensions whose dynamic relationships must be taken into account when attempting to understand change and to redress certain tendencies and move towards more sustainable urban and territorial planning.

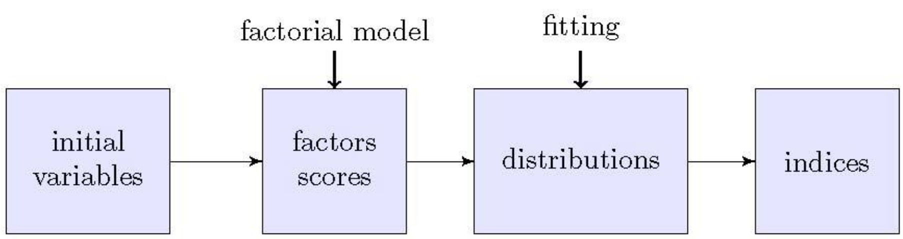
3. Modelling urban networks sustainable progress

We propose a new method to measure sustainable progress of urban networks, based on a statistical model, derived from Eurostat databases and NASA satellite images, and capable of analyzing different conceptual scenarios of urban development in Europe. The proposed Urban Networks Sustainable Progress Model integrates current knowledge of how social, economic and ecological factors contribute collectively to establishing and measuring sustainable progress in close relation to territorial variables.

The measurement of sustainable progress in city networks has two main inherent problems: i) how to define the scope of these urban systems at a megaregional level; ii) how to choose the factors, which have to be adaptable to spatial and temporal changes, and establish their relative weights in the measurement of sustainable progress. Economic, social and ecological factors cannot be measured directly but rather have to be calculated through a series of specific variables. For this reason, we employed a factorial analysis as a way to detect and measure indirectly the hidden dimensions of sustainable progress in city networks. Law-based indices are derived from the factors' distributions on the territorial units.

The factorial model and the indices developed from it will have to: i) aid understanding of the relationship between social, economic and ecological factors within a framework of the territory in which they exist; ii) provide indices for the main conceptual scenarios regarding sustainable progress in city networks at regional and megaregional scales, that will be useful to compare different territorial units in several years. The methodological approach starts in the variables measured on the territorial units and ends developing a set of compound indices that try to capture complex scenarios (Scheme I).

Scheme I. Conceptual framework of the Urban Networks Sustainable Progress Model



Firstly, we define the territorial units and describe the data sources (Section 3.1). Secondly, we describe how the factorial analysis is performed (Section 3.2). Finally, we introduce the sustainable progress indices (Section 3.3).

3.1. Urban networks and sustainable progress factors

3.1.1. Delimiting urban networks at megaregional scale

From a systemic perspective, the process that leads to the formation of megaregions begins with the growth of chain interaction between nearby cities in dense areas. The interaction expands first in short distances and then widens in distance while the interrelation chains multiply exponentially. Growing at the same time, the networks of different urban areas become connected at some point, so that this chain mechanism widens its scale again. The procedures to delimit the megaregions try to reproduce, in one way or another, this process.

To delimit megaregions a number of different methodologies exist that are based mainly on census data and a group of structural criteria including transport networks, demographic growth and land use (Lang and Dhavale, 2005; Dewar and Epstein, 2007). In this study, we use NTL satellite data (Marull et al., 2013), an improvement over official statistics since this type of data allows us to delimit urban systems and estimate indicators for non-administrative units of analysis. Thus, the main database used for defining megaregions consists of images from the satellite DMSP-OLS supplied by the NASA. These images are in GeoTiff format with a spatial resolution of approximately 1 km^2 per pixel (30'). Each satellite pixel sensor assigns a specific value to the light intensity known as DN (Digital Number), which has a radiometric resolution of 6 bits and a range of values between 0 and 63.

The delimitation of megaregions (Florida et al., 2008) was performed using the following criteria: i) a megaregion consists of a continuous illuminated zone harbouring more than one large city or metropolitan region, and emitting over 100,000 million dollars of LRP (Light-based Regional Product); ii) a megaregion is characterised by the physical contiguity of its human settlements, which implies a minimum DN threshold (DN=10) and a minimum distance between its illuminated areas (3 km). This accumulative procedure applied to a series of annual data sets (from 1995 to 2010) enables us to identify and measure the evolution of 12 megaregions that exist in Europe (Figure 1). The names of these regions are abbreviated as follows: AMB (*Amsterdam-Brussels-Antwerp*); BAL (*Barcelona-Lyon*); BER (*Berlin*); FRG (*Frankfurt-Stuttgart*); GLB (*Glasgow-Edinburgh*); LIS (*Lisbon*); LON (*London*); MAD (*Madrid*); PAR (*Paris*); PRA (*Prague*); RMT (*Rome-Milan-Turin*); VIB (*Vienna-Budapest*); and NMR (No MegaRegion, i.e., does not belong to any megaregion).

The delimiting of the networks of the cities that form part of the megaregions is only a good approximation since due to a series of technical problems (Small et al., 2005) it is not possible to define an exact relationship between illuminated areas detected by the satellites and built-up areas. Nevertheless, the use of a single criteria and a common database is a guarantee that the defined entities are comparable (Nel·lo et al, 2017).

3.1.2. Data sources and management

The most common framework for sustainable development lies in the economic, social and environmental dimensions. The thermodynamic approach also takes into account knowledge and organization. The latent factors involved in the urban network sustainability have to be deduced from a set of evaluable variables, whether official indicators or others deduced from them. We work at both, regional (NUTS 3) and megaregional scales, thus dealing with two raw data sets. The procedures have been implemented in the R-environment (R Core Team, 2013). Here we show how the data files are defined and managed to obtain the urban networks sustainable progress factors**.**

The region-year data file: 20704 cases, corresponding to the NUTS 3 in four different years (1995, 2000, 2005 and 2010), and six variables. We used a database consisting of three traditional variables obtained from Eurostat (economic growth –GDP, employment –GRE and knowledge –PAT), together with three other variables estimated using satellite data (energy consumption –PEC, urban growth –URG and urban density –URB). We use a technique based on satellite images, described in Marull et al., 2013 and founded on a strong direct linear relation between light intensities and the several indicators of urban development, that enables us to project the values of these variables - available at the country level- onto the related regional units with sufficient accuracy. In this way, the NUTS 3-year dataset is built.

The detailed description of the six variables are: GPDpc (Gross Domestic Product in PPA; thousands of euros at parity of acquisitive power per inhabitant); GREpc (Gross Rate Employment; number of workers per 1,000 inhabitants); PATth (Patent Applications to the European Patent Office; number of patents applied for per 1,000 inhabitants); URDpsk (Urban Density; number of inhabitants per km^2 of illuminated surface area NTL); URGpor (Urban Surface; percentage of illuminated surface area NTL per NUTS 3); PECpc (Primary Energy Consumption; millions of tonnes of petrol equivalent per inhabitant). Once the regions dataset is given, then the megaregions data file is constructed similarly using population or satellite data as projection weightings, depending on the variable (Appendix A).

It is advisable to mention the existence of units of analysis where one or more variables have missing values. There are several approaches for imputing multivariate incomplete data; here we used the multivariate imputation by chained equations (MICE) (Rubin, 1987). In our study, the differences between imputed and complete correlation matrices are lower than 0.08 and most are below 0.02. With this controlled allocation, a group of cases have been recovered (from 13237 complete cases initially, to 17363 complete cases after imputing). The cases containing imputed values are not going to be used in determining the factorial model. Computations involving missing values affect only the scatter plots in Figures 4, 5 and 6.

3.2. Factorial model

The factorial analysis (Thomson, 1951) is a method for investigating whether a number of observed variables of interest can be expressed as a linear combination of a smaller number of unobservable latent factors, allowing a residual or unexplained term. According to conceptual information, based on the fact that sustainable progress involves social, economic and ecological aspects, we propose fitting a structural equations model (Jöreskog, 1969) consisting in three correlated factors. The model is defined imposing that each factor affects only a specific subset of variables, despite the fact that all the factors can be either directly or indirectly correlated with each one of the variables to one degree or another. In structural equations models (Westland, 2015), the assumptions made on the factors have certain implications that allow validating whether the data fits the hypothesized measurement model.

3.2.1. Development of the factorial model

The factorial model is based on the variables correlation matrix. Thus, its equation is expressed in terms of *Z*, the vector of the scaled variables: zGPDpc, zGREpc, zPATth, zURDpsk, zURGpor and zPECpc. The model is developed in the following stages:

a) *Model equation and assumptions*: The equation relating the initial indicators and the latent factors can be expressed as:

$$Z = QF + U \qquad (1)$$

where Z is the column vector containing the six (scaled) variables, $F=(F_1, F_2, F_3)$, is a three latent factors column vector, Q is the factors loadings matrix and U the residual factors column vector. We assume that: i) latent factors have unit variance; ii) there exist correlation between each pair of latent factors (Σ_F is not diagonal, allowing an oblique rotation); iii) specific factors are pairwise incorrelated and incorrelated with the latent factors. Finally, we fix to zero several factor loadings, that is, each factor loads only on a specific subset of variables:

$$\begin{aligned}
\text{zPATth} &= q_{11} F_1 + 0\, F_2 + 0\, F_3 + U_1\\
\text{zGDPpc} &= q_{21} F_1 + 0\, F_2 + q_{23} F_3 + U_2\\
\text{zPECpc} &= 0\, F_1 + q_{32} F_2 + 0\, F_3 + U_3\\
\text{zGREpc} &= 0\, F_1 + 0\, F_2 + q_{43} F_3 + U_4\\
\text{zURDpsk} &= 0\, F_1 + q_{52} F_2 + 0\, F_3 + U_5\\
\text{zURGpor} &= q_{61} F_1 + 0\, F_2 + 0\, F_3 + U_6
\end{aligned}$$

This imposes a specific simple structure in the matrix Q in (1). These hypotheses are common in confirmatory factorial analysis, where typically some of the factor loadings are fixed to be zero. Furthermore, these assumptions are consistent with the idea that sustainability is based on three interrelated components (economic, social and ecological) and that all the factors influence directly or indirectly the whole set of variables (the direct correlations coming from the loadings and the indirect ones due to the within-factors correlations).

b) *Model implications*: The assumptions made on the model imply the decomposition

$$R = Q\Sigma_F Q^t + \Psi \qquad (2)$$

where $R = (r_{ij})$ is the symmetric and positive definite correlation data matrix, Q is the loadings matrix and Ψ the residual covariance matrix:

$$Q = \begin{pmatrix} q_{11} & 0 & 0 \\ q_{21} & 0 & q_{23} \\ 0 & q_{32} & 0 \\ 0 & 0 & q_{43} \\ 0 & q_{52} & 0 \\ q_{61} & 0 & 0 \end{pmatrix} \quad \Sigma_F = \begin{pmatrix} 1 & \alpha & \beta \\ \alpha & 1 & \gamma \\ \beta & \gamma & 1 \end{pmatrix} \quad \psi = \begin{pmatrix} \psi_{11} & 0 & 0 & 0 & 0 & 0 \\ 0 & \psi_{22} & 0 & 0 & 0 & 0 \\ 0 & 0 & \psi_{33} & 0 & 0 & 0 \\ 0 & 0 & 0 & \psi_{44} & 0 & 0 \\ 0 & 0 & 0 & 0 & \psi_{55} & 0 \\ 0 & 0 & 0 & 0 & 0 & \psi_{66} \end{pmatrix}$$

The proof of (2) is given in Appendix B.

3.2.2. Parameters estimation and factorial scores

Once the factorial model has been defined, the next step is to adjust it to the data. The function *cfa()* in *lavaan* R-package (Rosseel I., 2012) was used. The free parameters are estimated (Section 4) and the model has been shown to reproduce faithfully the original correlation matrix (Appendix C). Several numeric and graphical procedures have been implemented enabling us to perform an exhaustive analysis based on:

a) *The pattern matrix*: The analysis of the loadings estimates will permit us to understand the latent factors meaning. Due to the conceptual constraints on the loadings, the factors can already be labelled at this time as *economic activity* (F_1), *urban ecology* (F_2) and *social cohesion* (F_3). The numerical estimates of Q (Section 4) shall confirm this labelling.

b) *The factors correlation matrix*: The within latent factors correlations illustrate the trend and the intensity of the linear relationship between them.

c) *The structure matrix:* Contains the correlations between the initial variables and the latent factors, and it turns out to be the pattern matrix times the factors correlation matrix (Appendix B):

$$R_{XF} = Q\Sigma_F \qquad (3)$$

d) *The factorial scores:* Are the predictions of underlying factors punctuations in each analysis unit (region or megaregion – year). The factorial scores have been calculated using the Thompson regression method (Thomson, 1951),

$$F = ZR^{-1}Q\Sigma_F \quad (4)$$

Formula (4) is proved in Appendix B, under the models constraints. It gives us the equations for the prediction in the given units and for new observations too (Section 5). Expressing the normalized variables Z in terms of the raw original ones, $Z = f(X,\mu,\sigma)$, the scores F depend directly on X (see (6)).

3.3. Probability-based indices

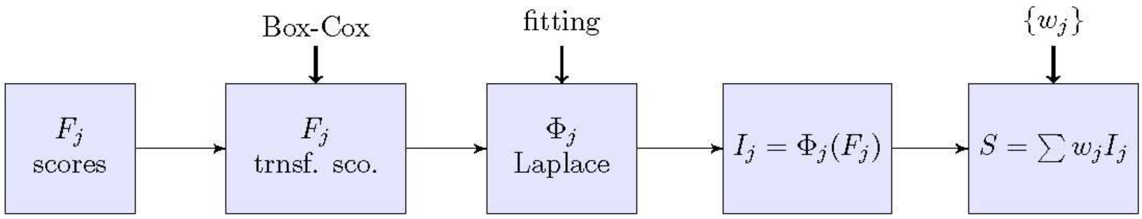
We propose the term "probability-based index" to refer to any statistic that is a function of several variables (indicators) and satisfies: i) it takes values in a bounded interval; ii) it can be used to rank a set of observations; iii) it depends on a parametric family of distributions adjusted to the data sample. Following this criterion, we develop three probability-based indices, denoted I_1, I_2 and I_3, and called "simple-indices". These indices are derived from the empirical distributions of the factorial scores F_1, F_2 and F_3, and therefore they should be able to capture the various components of sustainability in urban networks (social, economic or urban). We call "compound-index" any linear convex combination of simple-indices.

3.3.1. Steps in the indices' construction

As we consider indices based on the cumulative distribution function of the factorial scores, a pre-processing of the data is required to achieve a reliable distributional fitting. We performed a Box-Cox monotone power-type transformation (Box and Cox, 1964) of the factors' scores to stabilise the variance and get a more symmetric shape. This technique is applied to the factors F_1, F_2 and F_3, and the transformed scores are then adjusted to a Laplace distribution. The parameters of the Box-Cox transforms and the Laplace distributions on each factor are given in Appendix D. To simplify notation, old and transformed factors are denoted equally.

Simple-indices consist of applying the Laplace cumulative distribution function to the transformed scores. Once the simple-indices have been obtained, the aim is to integrate them into new compound-indices that exhibit sufficient flexibility to respond to the evaluation of differentiated scenarios of sustainable development. To do so, the simple-indices will be pooled using penalty weights (that allow us, for example, to prioritise social factors over economic factors). A representation of the procedure is given in Scheme II.

Scheme II. Methodological approach of the Urban Network Sustainable Progress Model



We start with a factor (for example, F_1) conveniently normalized by Box-Cox. Next, we subject it to a new transformation applying the function Φ_1, which is precisely the cumulative Laplace distribution function adjusted to the F_1 scores. Thus, the simple-index I_1 is defined by: $I_1 = \Phi_1(F_1)$. Therefore, if a region has a score f_{1i} in some specific year, the value $\Phi_1(f_{1i})$ is the accumulative proportion of cases whose punctuations in F_1 are less than or equal to f_{1i} according to the law Φ_1.

Clearly, simple-indices defined in this way lie in (0,1), are probability based, and may be used to rank the sample cases. Each index I_j aggregates multiple indicators, provided that the factor F_j itself combines linearly the initial variables. These indices should be interpreted as follows: If the value I_j=m corresponds to a given analysis unit, then this region-year achieves the m-th quantile (m×100% percentile) in the correspondent factor distribution. The larger the value of I_j, the better is the position of the region-year related to this factor.

As mentioned above, we consider a compound-index to be any convex lineal combination (i.e. a weighted mean) of certain collection of simple-indices. Indeed, given the three simple-indices, we define:

$$S = w_1 I_1 + w_2 I_2 + w_3 I_3; \quad w_1 + w_2 + w_3 = 1 \qquad (5)$$

Obviously, any compound-index lies in (0,1) and the election of the weights determines the meaning of *S*. For instance, take $w_1 = s,\ w_2 = 1 - w,\ w_3 = 0$ and consider $S = I_{12}^{w} = (1 - w)I_1 + wI_2$, depending only on *w*, where *w* < 0.5. The weight *w* can be understood as a *penalty* on I_1 depending on the value in I_2: a positive penalty if I_2 is larger than I_1, and negative conversely.

Figure 2 shows how Φ_1 applies on the scores to obtain I_1, together with the effect of the weight *w* in I_{12}^{w}, for *w* = 0.1 to *w* = 0.5 (increasing by intervals of 0.1). The values are shown taking into account two behaviours: i) red dots correspond to negatively penalised cases: $I_{12}^{w} < I_1$, if $F_2 < F_1$; ii) blue dots correspond to correspond positively penalised cases: $I_{12}^{w} > I_1$, if $F_2 > F_1$. The differences between a low (*w* = 0.1) and a high weight (*w* = 0.5) are clear.

According to the labels of the factors, we call: I_1 (*economic activity*), I_2 (*urban ecology*) and I_3 (*social cohesion*). By means of appropriate weights, these simple-indices are then combined to define compound-indices that capture the simultaneous evolution in more than one factor.

3.3.2. *Compound-indices to define conceptual scenarios*

Taking into account different criteria of sustainable progress in city networks, four conceptual scenarios shall be evaluated in terms of four compound-indices (S_j), defined as certain linear convex combinations of the three interrelated simple-indices (I_1, I_2, I_3) with the following criteria: i) S_1 – *economic development* (I_1, negatively penalized by very low values of I_2 and I_3); ii) S_2 – *social sustainability* (I_3, penalising negatively low values of I_2, and omitting I_1); iii) S_3 – *environmental sustainability* (I_2, negatively penalized by low values of I_3, and omitting I_1); iv) and S_4 – *inclusive development* ($S_{4.1}$, with the same weight for I_1, I_2 and I_3; $S_{4.2}$, where I_1, I_2 and I_3 play a symmetric role but penalising an imbalance in their values). The explicit formulas and other details are given in the sequel. A number of different penalties are proposed in the different scenarios (Table 1). It is important to remark that in all scenarios the weights can be modified as the user wishes.

The index S_1 (*economic development*) tries to measure a trend according to neoclassical economic theory, in which economic growth should be as great as possible and will be poorly affected by any other factor:

$$S_1 = w_1 I_1 + w_2 I_2 + w_3 I_3;\ w_1 + w_2 + w_3 = 1;\ w_1 > w_2 \ \ and \ \ w_1 > w_3$$

To visualise the regions that meet this standard, an approximation to the calculation S_1 is performed taking into account the weights w_1 = 0.8, w_2 = 0.1 and w_3 = 0.1 in order to maximise economic factors (without completely ignoring other factors).

The index S_2 (*social sustainability*), on the other hand, prioritises the factor of social equality. In this case, it is assumed that the economic factor plays no explicit role (although GDPpc loads on F_3 and therefore affects I_3). Thus, S_2 takes into account I_3 and I_2. In this paper, this index is constructed using the weights w_3 = 0.8 and w_2 = 0.2.

$$S_2 = w_2 I_2 + w_3 I_3;\ w_2 + w_3 = 1;\ w_3 > w_2$$

The index S_3 (*environmental sustainability*) focuses on the urban ecology factor. In this case, the economic factor likewise plays no explicit role. Thus, S_3 also takes into account the indicators I_2 and I_3 but is constructed with the reciprocal weights $w_2 = 0.8$ and $w_3 = 0.2$.

$$S_3 = w_2 I_2 + w_3 I_3;\ w_2 + w_3 = 1;\ w_2 > w_3$$

Finally, the index S_4 (*inclusive development*) measures the regions that have a high but balanced value for all three indicators I_1, I_2 and I_3. Here we consider two possible approximations: $S_{4.1}$ (with the same weight for all factors) and $S_{4.2}$ (with the factors playing the same role but penalizing unbalanced values for the factors).

$S_{4.1}$ is an inclusive index because all the indicators I_1, I_2 and I_3 have the same weight:

$$S_{4.1} = \frac{1}{3} I_1 + \frac{1}{3} I_2 + \frac{1}{3} I_3$$

$S_{4.2}$ is also a balanced inclusive index, which penalises imbalances between I_1, I_2 and I_3:

$$S_{4.2} = \left(\frac{1}{3} + 2\beta\right) \mathrm{Min}\{I_1, I_2, I_3\} + \frac{1}{3} \mathrm{Med}\{I_1, I_2, I_3\} + \left(\frac{1}{3} - 2\beta\right) \mathrm{Max}\{I_1, I_2, I_3\}$$

with $\beta \leq \frac{1}{6}$. Specifically, in this study case, we take $\beta = \frac{1}{12}$, and so:

$$S_{4.2} = \frac{1}{2} \mathrm{Min}\{I_1, I_2, I_3\} + \frac{1}{3} \mathrm{Med}\{I_1, \mathrm{I}_2, I_3\} + \frac{1}{6} \mathrm{Max}\{I_1, I_2, I_3\}$$

4. Results and discussion

4.1. Factorial model: estimates, labels and scores

The factorial model has been shown to reproduce faithfully the correlation matrix between the initial variables. Indeed, off the diagonal, the maximum absolute difference between the original and the reproduced correlation matrices is 0.1 (Appendix C). The estimates of the parameters are given in different tables and pictures:

a) *The pattern matrix*: The loadings matrix estimates are in the table on top in Figure 3. An analysis of the loading estimates (signs and absolute values) confirms the factors' labels: *economic activity* (F_1), *urban ecology* (F_2) and *social cohesion* (F_3). Indeed, the pattern matrix (Figure 3, upper table) reveals that in F_1 the variables zPATth, zGDPpc and zURGpor are direct measurements with positive weights (0.50, 0.27 and 0.60, respectively) and reflect economic activity based on innovation and urban development. In F_2, the difference in the signs of the weights (-0.58 and 0.80) of the variables zPECpc and zURDpsk shows that when cities are arranged in a dense, well-connected urban network (e.g. in a polycentric structure), they are more efficient regarding the energy they require to maintain their complexities, a thermodynamic concept present in urban ecology. Finally, in F_3, the variables zGREpc and zGDPpc, with loadings 0.92 and 0.68, are the observable measurements of the latent factor referred to as social cohesion.

b) *The factors correlations matrix*: The estimates of the within latent factors correlations matrix – table on the middle part in Figure 3– show a moderate positive correlation between each pair of them. This can be interpreted in the sense that economic activity, urban ecology and social cohesion have an increasing trend, but not really strong linear relation; in fact, the scores represented in Figures 4 to 6, show a curvilinear relation, as it will be mentioned in Section 4.2.

c) *The structure matrix:* Looking at the structure matrix estimates (Figure 3, lower table), of special interest in F_3 is the negative correlation with the energy consumption (PECpc) and the positive correlation with patent applications (PATth); although this latter correlation is still low in intensity, it is possible to trace a route towards inclusive development.

A graph-based representation provides a good insight on the aforementioned estimates. In the two graphs in Figure 3, the unidirectional arrows represent the regression coefficients that express the initial variables in terms of the latent factors; the bidirectional arrows, on the other hand, show the correlations. The breadth of the lines is proportional to the coefficients absolute value (in red, negative).

The factorial scores are given using the formula (4) and unscaling the data (i.e. expressing the scaled variables in terms of the original ones). In this way, we obtain the explicit equations expressing the scores as linear functions of the original variables:

$$F_1 = 0.00129PAT_{th} + 4\times 10^{-5}GDP_{pc} - 0.01981PEC_{pc} - 0.00163GRE_{pc} + 6\times 10^{-5}URD_{psk} + 0.01481URG_{por} - 2.17892\,;$$

$$F_2 = 7.5\times 10^{-4}PAT_{th} + 0GDP_{pc} - 0.12097PEC_{pc} + 0.00984GRE_{pc} + 5.6\times 10^{-4}URD_{psk} - 0.00206URG_{por} - 0.2246;$$

$$F_3 = -3.4\times 10^{-4}PAT_{th} + 3\times 10^{-5}GDP_{pc} + 0.00402PEC_{pc} + 0.06453GRE_{pc} + 3\times 10^{-5}URD_{psk} + 0.00207URG_{por} - 3.73266. \quad (6)$$

4.2. Factorial scores at regional and megaregional level

In order to evaluate the factorial model used in this study, firstly we describe the behaviour of the factors F_1 (*economic activity*), F_2 (*urban ecology*) and F_3 (*social cohesion*) at regional level (NUTS 3) taking into account the megaregion it is a part of – if, indeed, it belongs to any – during the period of analysis (1995, 2000, 2005, 2010). In Figures 4 to 6 the scores are represented in scatter plots, each point corresponding to a region for a given year. In general, the NUTS 3 that do not belong to any megaregion (NMR) behave worse both in terms of economic activity and urban ecology (F_1 and F_2) if we compare them with the majority of the NUTS 3 that do belong to a megaregion (Figure 4).

In the relationship between economy and ecology (F_1, F_2) (Figure 4) three main trends stand out in the NUTS 3 that were part of a megaregion during the period of analysis (see Section 4.4): *Frankfurt-Stuttgart* (FRG): increased economic activity (F_1); *London* (LON): greater ecological efficiency (F_2); and *Paris* (PAR): high values for both these factors. As it shows the quadratic shape, at low levels for the economic factor, an increase in this factor is associated with negative increases in the ecological factor, whereas at high levels the association is positive.

A quite lineal and increasing tendency can be observed in the relationship between economic activity and social cohesion (F_1, F_3) (Figure 5), with lower values in both factors in the NUTS 3 that do not belong to any megaregion (NMR) and higher values in the NUTS 3 that belong to the most developed megaregions in Europe (e.g. FRG, AMB and PAR). The convexity of the curve indicates that economic activity is associated with a trend that is increasingly somewhat greater than the social cohesion.

Even more interesting is the relationship between urban ecology and social cohesion (F_2, F_3) (Figure 6), where there is a positive association between these two factors that is not observable in the NUTS 3 that do not belong to a megaregion (NMR) but is obvious in those that form part of the most advanced European megaregions such as FRG, AMB, LON and PAR. The concave aspect of the curve is due to a single extreme value without which the relationship would be almost lineal in the positive quadrant. This relationship between ecological efficiency and social equality (F_2, F_3) (Figure 6) is especially remarkable as it indicates that it is possible to have high levels of employment and low levels of energy consumption in the most complex city networks (polycentric urban structures; Marull et al., 2015).

4.3. Indices values at regional and megaregional level

Once the behaviour of the factors has been analysed, for each European region (NUTS 3) the value obtained for the indices of sustainable progress for the city networks is given, according to the scenarios under consideration: S_1 (*economic development*; Figure A1), S_2 (*social sustainability*; Figure A2), S_3 (*environmental sustainability*; Figure A3) and, finally, $S_{4.2}$ (*inclusive development*; Figure A4). The European regions for which there are not sufficient data (in white on the maps) were not included in the analyses.

In general, the maps provide a consistent representation of economic development (Figure A1), social sustainability (Figure A2) and environmental sustainability (Figure A3) of the city networks at NUTS 3 level between 1995 and 2010. The first two indices (S_1 and S_2) improve over time, although S_2 decreases in the latter period, probably due to the financial crisis. This tendency is not so obvious in S_3. European regions seem to be advancing more quickly in economic than in socio-environmental fields.

It is even more interesting to note the behaviour of the inclusive development index (Figure A4) in its final form $S_{4.2}$ ($S_{4.1}$ are $S_{4.2}$ are graphically very similar). This index reveals that the most complex regions have the greatest and most balanced sustainable progress (in social, ecological and environmental terms) and show higher resilience of the city networks (for example, in light of the disturbances generated by the recent financial crisis).

The application of the simple-indices I_1 (*economic activity*), I_2 (*urban ecology*) and I_3 (*social cohesion*) at megaregional scale (Table 2) during the period of analysis (1995-2010) gives the highest values for economic activity for *Paris* (PAR; from I_1 = 0.76 to I_1 = 0.89), *Frankfurt-Stuttgart* (FRG; from I_1 = 0.24 to I_1 = 0.85) and *Amsterdam-Brussels-Antwerp* (AMB; from I_1 = 0.54 to I_1 = 0.76). The results also give high levels of ecological efficiency for *Berlin* (BER; I_2 = 0.89), *Madrid* (MAD; I_2 = 0.84) and *Paris* (PAR; I_2 = 0.82), largely due to their urban density; and greater social cohesion to *Frankfurt-*

Stuttgart (I_3 = 0.89) and *Paris* (I_3 = 0.87). The NUTS 3 that do not belong to any megaregion (NMR) have the lowest values for these three indices (I_1 = 0.05; I_2 = 0.34; I_3 = 0.02).

The compound-indices S_1 (*economic development*), S_2 (*social sustainability*) and S_3 (*environmental sustainability*) at megaregional scale (Table 2) show an overall increase in their values for the period 1995-2010. In 2010, the results show the greatest economic development and social cohesion in the megaregions of *Paris* (S_1 = 0.88; S_2 = 0.86) and *Frankfurt-Stuttgart* (S_1 = 0.85; S_2 = 0.87); and greatest environmental sustainability in *Berlin* (S_3 = 0.87), *Madrid* (S_3 = 0.84) and *Paris* (S_3 = 0.83).

The NUTS 3 regions that do not belong to a megaregion (NMR) have the lowest values for economic development and social sustainability (S_1 = 0.17; S_2 = 0.38), although the *Lisbon* megaregion has the lowest value for environmental sustainability (S_3 = 0.49). The scenario calculation (Table 2) includes the penalty factors (as described in Table 1) and so the compound-indices (S_j) constitute a more precise measurement than that obtained with the simple-indices (I_j).

The S_4 compound-indices (*inclusive development*; in the versions $S_{4.1}$ with equivalent weights for the factors, and $S_{4.2}$ with equivalent weights and balanced values for the factors) calculated at megaregional level (Table 2) give the highest values for *Paris* ($S_{4.1}$ = 0.86; $S_{4.2}$ = 0.85) and *Frankfurt-Stuttgart* ($S_{4.1}$ = 0.84; $S_{4.2}$ = 0.82), and the lowest values for the NUTS 3 regions that do not belong to a megaregion – NMR ($S_{4.1}$ = 0.33; $S_{4.2}$ = 0.26). There is a general increase in the values for these scenarios (1995-2010).

Although results for $S_{4.1}$ and $S_{4.2}$ are similar, it is worth highlighting the fact that the lowest values were obtained for $S_{4.2}$ since in this index the imbalance between factors is penalised (Table 1). Thus, we believe that $S_{4.2}$ is a better approximation to sustainable progress in a city network given the inclusive development principle, which consists of the tendency of urban systems to move towards social equality, ecological efficiency and economic competitiveness.

Figure 7 shows simple-indices (I_j) and compound-indices (S_j) for each megaregion and year. This makes it easy to compare the behaviour and dynamics of city networks during the period of analysis under different conceptual scenarios of sustainable progress; this enables us to detect which indicators follow the best or worst patterns and the balance between the main factors of sustainability in urban systems. This method, based on a factorial model, provides information on the properties of regions and megaregions to citizens, technicians and politicians, which in turn helps them take democratic decisions regarding public policy.

5. Conclusions

This paper is the first attempt to model sustainable progress in complex urban systems from the point of view of thermodynamics. In thermodynamics, as a system becomes more complex its energetic dependence is lessened and the amount of organised information increases. From this approach, regional and megaregional city networks are an organizational response of complex urban systems to changes in economic, ecological and social patterns. As such, they do not have to be unsustainable (Wheeler, 2009), but a response to ensure the long-term viability of these systems (Marull et al., 2015). In particular, we have explored the hypothesis that regions belonging to megaregions have a better performance in economic, ecological and social dimensions, improving sustainable urban progress. The hypothesis is tested using a flexible empirical modeling inferred from data that permits the evaluation of different conceptual scenarios with dominance in each one of economic, social, ecological, and inclusive urban networks development.

We have developed a reproducible Urban Networks Sustainable Progress Model that allows developing a family of simple and compound indices from a set of initial variables on urban regions and mega-regions, derived from official indicators using satellite images to overcome the difficulty of having data at these geographical scales, with the aim of measuring different conceptual scenarios of sustainability. The procedure has two principal elements: a factorial model and a type of distribution-based indices. The factorial model reproduces faithfully the correlations between the initial variables and reduces the dimensionality from six indicators into three latent factors, which may be identified as economic activity, urban ecology and social cohesion. The latent factors can then be reduced to one or more indices trying to reflect conceptual scenarios. Each one of the indices acts on the factorial scores and ranks the analysis units in the sample.

Moreover, new sample data can also be ranked, if all the model variables are given, following the Scheme II steps: i) use equations (6) and obtain the scores; ii) apply specific Box-Cox transforms to have the normalized scores; iii) for each factor, apply the correspondent Laplace cumulative distribution function on its normalized scores and compute the simple-index values; iv) choose convenient weights to define a compound-index trying to reflect a fixed conceptual scenario; and v) analyze the position of the ‘new unit’ in relation to the others. Furthermore, in other experimental settings a similar procedure could be "imitated" in direction of Scheme I, as long as a (likely different) factorial model and some parametric distribution –Laplace or another– can be satisfactorily adjusted to the data.

The results obtained with this modelling show that the most complete and integrated measurement of sustainable progress is given by a compound-index responding to a scenario based on inclusive development with a balance between the factors of economic activity, urban ecology and social cohesion (Figure 7). However, there is a possibility that economic growth, social development and environmental quality become disassociated. Some urban megaregional networks have improved their social cohesion and environmental quality without experiencing above-average economic growth. Based on those results, the main conclusion of this work is that, in fact, urban megaregional systems respond to increasing complexity and, via better access to more information, adapt their structural relationships to become more efficient and stable, and move to more sustainable forms of organisation. The main implication is that it is vital to redirect urban and territorial policies towards greater sustainable progress not only at urban or metropolitan levels but also at regional and megaregional levels. This thermodynamic-territorial strategy reduces the impact of city networks in terms of entropy and increases the organised information available in the urban system.

According to the indices of sustainable progress developed in this study, the experience of the best-positioned European megaregions shows that this is possible. From here, we can derive the fact that a change in the economic model that places greater importance on the economies of agglomerations based on polycentric urban structures (in which knowledge becomes a strategic productive element) will in the future become the motor of change in urban progress. In these sense, our results can be interpreted as a contribution to the dialogue between two lines of investigation, ecological economics and urban ecology, at a new spatial scale that can be used to explore the overall sustainability of urban systems: the consolidated and emergent megaregions. In addition, the results could be related to the current European strategy that is promoting "intelligent, sustainable and inclusive growth", in which the role that new policies at city network scale aimed at achieving these objectives could have a highly relevant role to play.

The study has some limitations, the most relevant are: i) the focus on economic, social and ecological factors, without taking into account other dimensions such as the cultural one of other composite dimensions of well-being; ii) the availability of data at sub-national levels, which limits the number of indicators available and the territorial detail and precision of some of them.

Future research should deepen knowledge of how to formulate urban networks sustainable progress scenarios, and how to use ever-larger data sets to interrelate different observation scales. Finally, it would be interesting to model the temporal variation in satellite NTL intensities. The possibility of making useful sustainable progress predictions related to urban networks scenarios could have an important impact on regional planning and land-use policy at global scale.

6. References

Abu Bakar, A.H., Cheen, K.S. 2013. A Framework for Assessing the Sustainable Urban Development. Procedia - Social and Behavioral Sciences 85, 484-492. https://doi.org/10.1016/j.sbspro.2013.08.377

Batabyal, A., Nijkamp, P. 2009. Sustainable development and regional growth. In: Capello, R., Nijkamp, P. (eds.), Handbook of Regional Growth and Development Theories. Edward Elgar, Cheltenham, UK/Northampton, MA, 282-301.

Baynes, T.M. 2009. Complexity in Urban Development and Management. Historical Overview and Opportunities. Journal of Industrial Ecology 13 (2), 214-227. https://doi.org/10.1111/j.1530-9290.2009.00123.x

Borowy, I. 2014. Defining Sustainable Development: the World Commission on Environment and Development. Brundtland Commission. Routledge.

Box G.E.P., Cox D.R. 1964. An analysis of transformations. Journal of the Royal Statistical Society Series B 26(2), 211-252. https://www.jstor.org/stable/2984418

Brandsma, A., Kancs, d'A., Monfort, P., Rillaers, A. 2015. RHOMOLO: A dynamic spatial general equilibrium model for assessing the impact of cohesion policy. Papers in Regional Science 94 (1), 197-222. https://doi.org/10.1111/pirs.12162

Camagni, R. 1993. From City Hierarchy to City Network: Reflections about an Emerging Paradigm. In Lackshmanan, T.R., Nijkamp, P. Structure and development in the space economy: essays in honour of Martin Beckmann. Berlin. Springer-Verlag.

Campbell, S. 2009. Mega-regions and sustainability. In: Ross, C. (Ed.), Megaregions: Planning for Global Competitiveness. Island Press, 127-139.

Capello, R., Caragliu, A., Fratesi, U. 2017. Modeling Regional Growth between Competitiveness and Austerity Measures: The MASST3 Model. International Regional Science Review, 40 (1), 38-74. https://doi.org/10.1177/0160017614543850

Costanza, R., Hart, M., Posner, S., Talberth, J. 2009. Beyond GDP: The Need for New Measures of Progress. Boston University, 2009. https://www.bu.edu/pardee/files/documents/PP-004-GDP.pdf

Dewar, M., Epstein, D. 2007. Planning for megaregions in the United States. Journal of Planning Literature 22, 118-124. https://doi.org/10.1177/0885412207306615

European Commission. 2018. Indicators for Sustainable Cities. Science for Environment Policy In-Depth Report. https://doi.org/10.2779/121865

Filchakova, N., Robinson, D., Scaterzzini, J-L. 2007. Quo vadis thermodynamics and the city: a critical review of applications of thermodynamics methods to urban systems. International Journal of Ecodynamics 2 (4), 222-223. https://doi.org/10.2495/978-1-84564-654-7/20

Florida, R., Gulden, T., Mellander, C. 2008. The Rise of the Mega Region. Economy and Society 1, 459-476. https://doi.org/10.1093/cjres/rsn018

Grazi, F., van den Bergh, J.C., van Ommeren, J.N. 2008. An empirical analysis of urban form, transport, and global warming. Energy Journal 29 (4), 97-122. https://www.jstor.org/stable/41323183

Jöreskog, K.G. 1969. A general approach to confirmatory maximum likelihood factor analysis. Psychometrika, 34, 183–202. https://doi.org/10.1007/BF02289343

Kubiszewski, I. Costanza, R., Franco, C., Lawn, P., Talberth, J., Jackson, T., Aylmer, C. 2013. Beyond GDP: Measuring and achieving global genuine progress. Ecological Economics 93, 57-68. https://doi.org/10.1016/j.ecolecon.2013.04.019

Kuznets, S. 1934. National Income, 1929–1932. 73rd US Congress, 2d session, Senate document no. 124, 5-7.

Lang, R.E., Dhavale, D. 2005. Beyond Megalopolis: Exploring America's New Megapolitan Geography. Metropolitan Institute at Virginia Tech, Census Report Series.

Lang, R.E., Nelson, A.C. 2009. Defining megapolitan regions. In: Ross, C.B., Contant,C. (Eds.), Megaregions in America. Island Press, Washington, DC.

Liang, X., Zhang, W., Chen, L., Deng, F. 2016. Sustainable Urban Development Capacity Measure - A Case Study in Jiangsu Province, China. Sustainability, 8(3), 1-13. https://doi.org/10.3390/su8030270

Lynch, A.J., Andreason, S., Eisenman, T., Robinson, J., Steif, K., Birch, E.L. 2011. Sustainable Urban Development Indicators for the United States. PENN IUR White Paper Series on Sustainable Urban Development. U.S. Department of Housing and Urban Development. https://penniur.upenn.edu/uploads/media/sustainable-urban-development-indicators-for-the-united-states.pdf

Marull, J., Galletto, V., Domene, E., Trullén, J. 2013. Emerging megaregions: a new spatial scale to explore urban sustainability. Land Use Policy 34, 353-366. https://doi.org/10.1016/j.landusepol.2013.04.008

Marull, J., Font, C., Boix, R. 2015. Modelling urban networks at mega-regional scale: Are increasingly complex urban systems sustainable? Land Use Policy 43, 15-27. https://doi.org/10.1016/j.landusepol.2014.10.014

Marull, J., Boix, R. (eds.). 2016. Megaregions i desenvolupament urbà sostenible. Factors estratègics per a l'àrea metropolitana de Barcelona en el context europeu. Papers 58.

Nel·lo, O., López, J., Martín, J., Checa, J. 2017. Energy and urban form. The growth of European cities on the basis of night-time brightness. Land Use Policy 61, 103-112.

Popa, H.L., Pater, L.R., Cristea, S.L. 2014. A systemic description of sustainable progress. Procedia – Social and Behavioral Sciences 124, 322-330. https://doi.org/10.1016/j.sbspro.2014.02.492

Pulselli, R.M., Ciampalini, F., Galli, A., Pulselli, F.M. 2006. Non-equilibrium thermodynamics and the city: a new approach to urban studies. Annali di Chimica 96, 543-552. https://onlinelibrary.wiley.com/doi/pdf/10.1002/adic.200690056

R Core Team. 2013. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org/

Ross, C.L. 2009. Megaregions: Planning for Global Competitiveness. Island Press, Washington.

Rosseel, I. 2012. Lavaan: An R Package for Structural Equation Modeling. Journal of Statistical Software 48(2), 1-36. http://www.jstatsoft.org/v48/i02/

Rubin, D.B 1987. Multiple imputation for nonresponse in surveys. Wiley series in probability and mathematical statistics. John Wiley & Sons.

Small, C., Pozzi, F., Elvidge, C.D. 2005. Spatial analysis of global urban extent from DMSP-OLS night lights. Remote Sensing of Environment 96 (3/4), 277-291. https://doi.org/10.1016/j.rse.2005.02.002

Stiglitz, J.E., Sen, A., Fitoussi, J.-P. 2009. Report by the Commission on the Measurement of Economic Performance and Social Progress Vol. 12. https://ec.europa.eu/eurostat/documents/118025/118123/Fitoussi+Commission+report

Thomson, G.H. 1951. The Factorial Analysis of Human Ability. London University Press.

Trullén, J., Boix, R., Galletto, V. 2013. An insight on the unit of analysis in urban research. In: Kresl, P.K., Sobrino, J. (eds.). Handbook of Research Methods and Applications in Urban Economies. Edward Elgar, Northampton.

Wheeler, S. 2009. Regions, megaregions, and sustainability. Regional Studies 43, 863-876. https://doi.org/10.1080/00343400701861344

Van den Bergh, J. C. J. M. J. 2009. The GDP paradox. Journal of Economic Psychology 30(2), 117–135. https://doi.org/10.1016/j.joep.2008.12.001

Varga, A. 2017. Place-based, Spatially Blind, or Both? Challenges in Estimating the Impacts of Modern Development Policies: The Case of the GMR Policy Impact Modeling Approach. International Regional Science Review 40 (1), 12-37. https://doi.org/10.1177/0160017615571587

Westland, J.C. 2015. Structural Equation Modeling: From Paths to Networks. Springer, New York.

Wilkinson, R.G. Pickett, K., 2009. The Spirit Level: Why Greater Equality Makes Societies Stronger. Bloomsbury.

Wilson, A. 2009. The thermodynamics of the city. Evolution and complexity science in urban modeling. In Reggiani, A., Nijkamp, P. (eds.). Spatial Networks and Complexity. Springer, Berlin.

Zegras, C., Poduje, I., Foutz, W., Ben-Joseph, E., Figueroa, O. 2004. Indicators for sustainable urban development. In Keiner, M., Zegras, C., Schmied W.A., Salmerón, D. (eds.), From Understanding to Action, 157-189. Kluwer Academic Publishers.

Figure 1 Dynamics of the European megaregions and NTL satellite data (1995-2010)

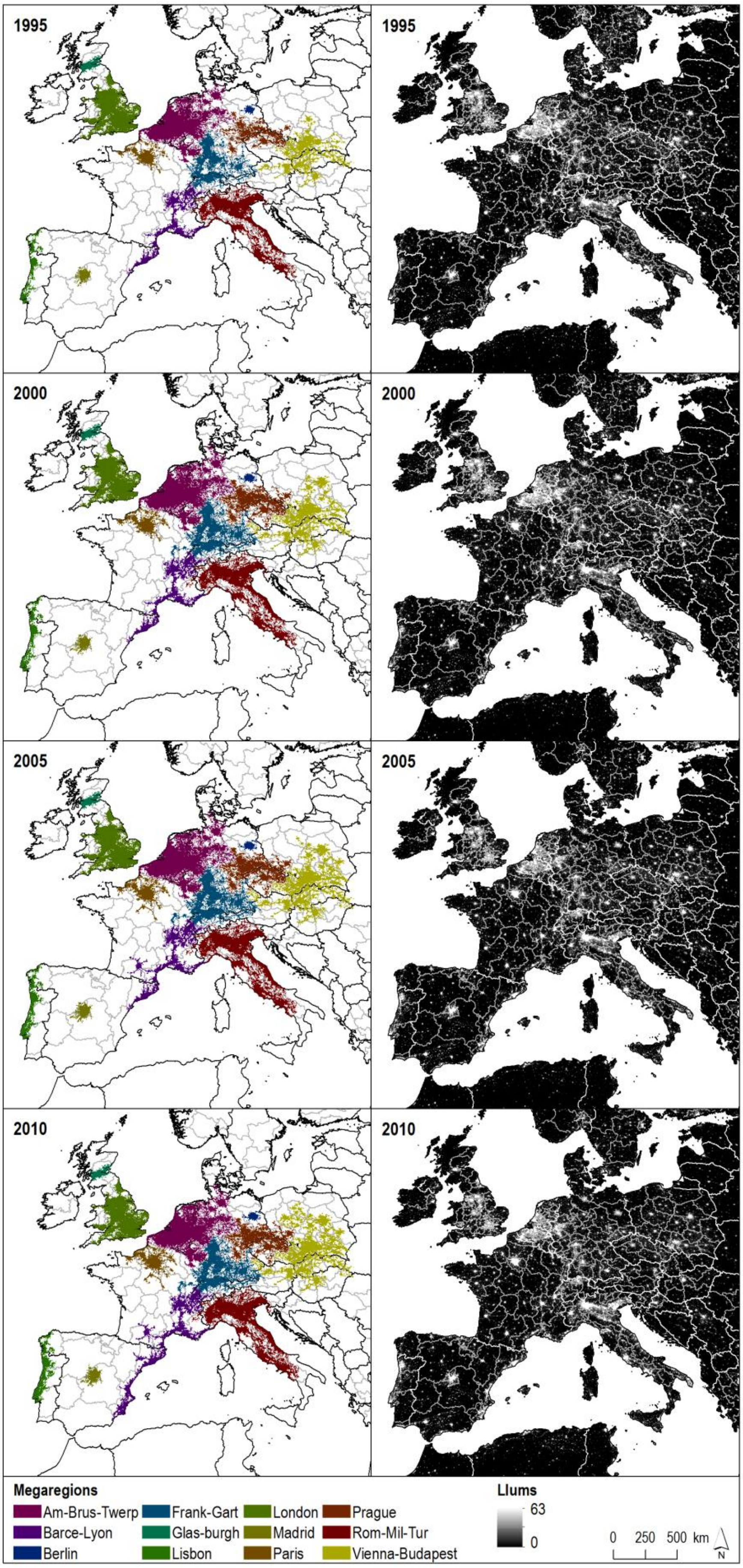


Figure 2 Effects of different penalty values (w between 0 and 0.5 on I_2) to modify I_1 and get the final expression of the index I_{12}^w. The cases in which the index I_{12}^w increases (blue) or decreases (red) with respect to I_1 (grey) are shown

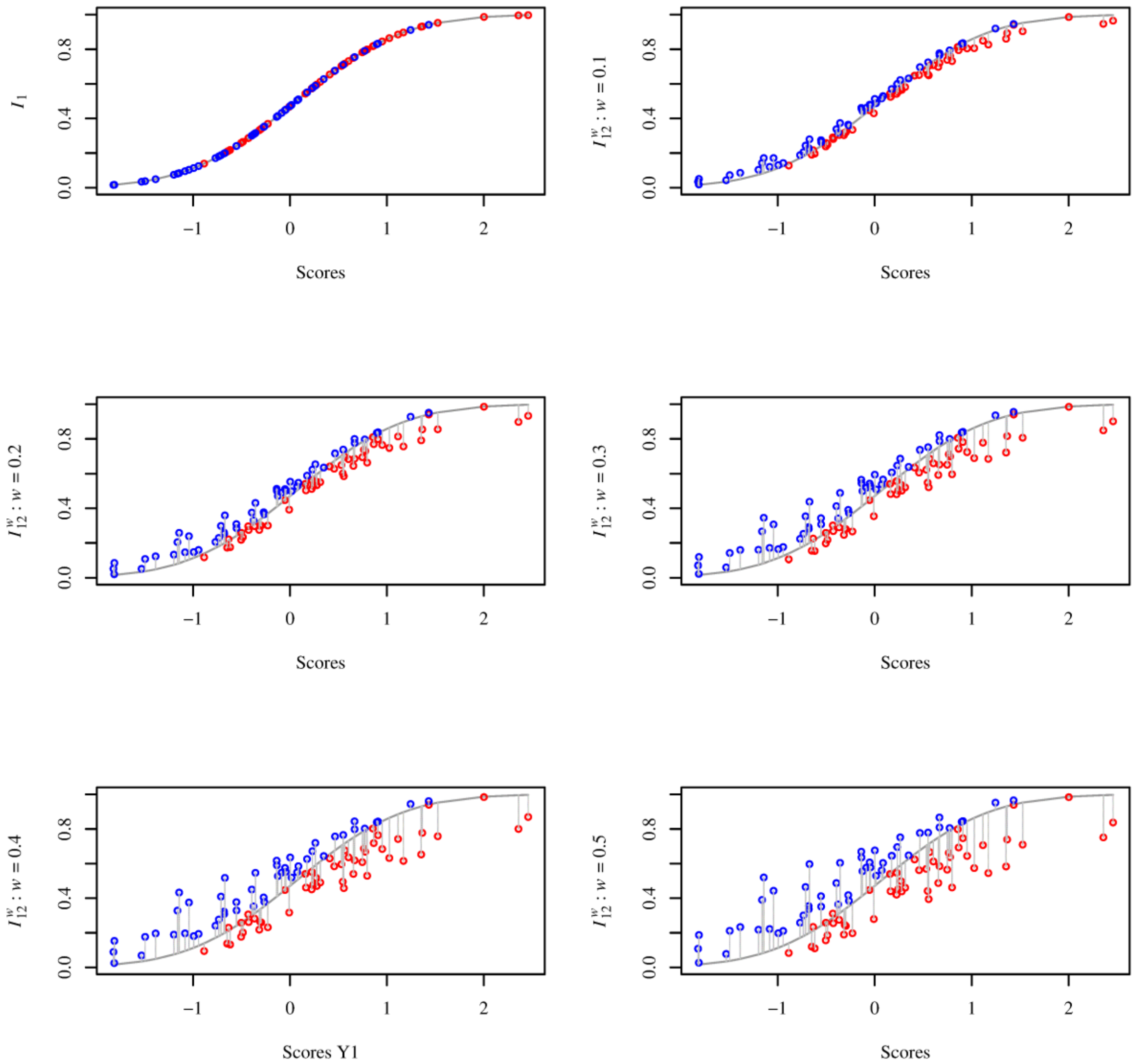
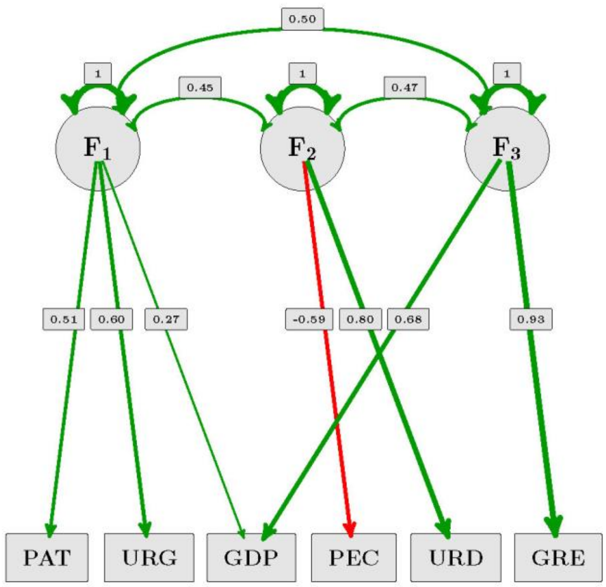
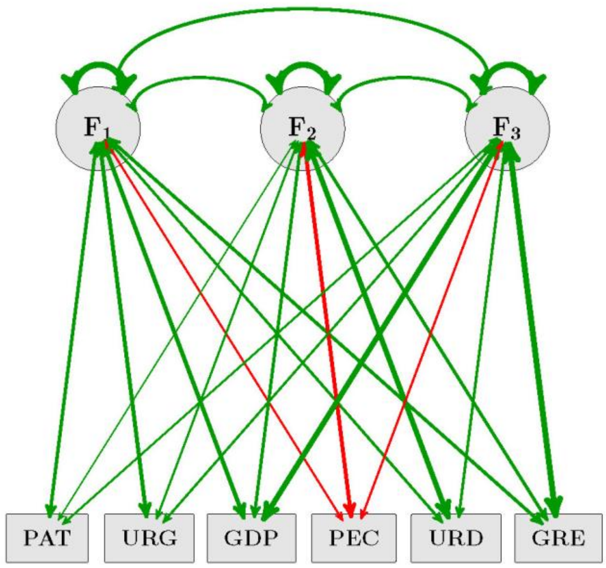


Figure 3 Factorial confirmatory analysis: pattern matrix and factors correlation matrix (above), and structure matrix (below). Estimated values and graphical representation



Variable	*F₁*	*F₂*	*F₃*
zPATth	0,50	0,00	0,00
zGDPpc	0,27	0,00	0,68
zPECpc	0,00	-0,58	0,00
zGREpc	0,00	0,00	0,92
zURDpsk	0,00	0,80	0,00
zURGpor	0,60	0,00	0,00

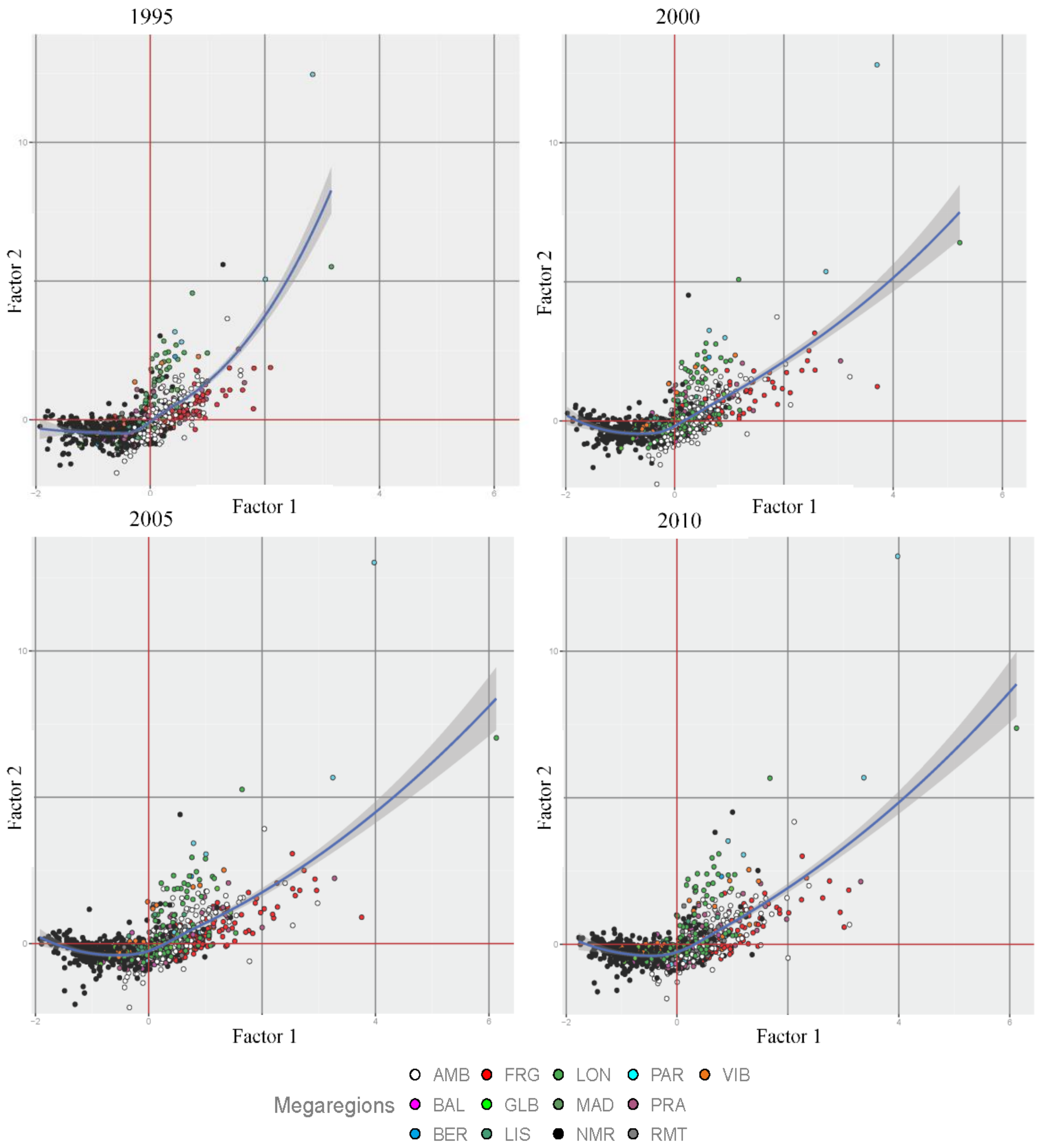
Factor	F_1	F_2	F_3
F_1	1	0,45	0,50
F_2	0,45	1	0,47
F_3	0,50	-0,47	1



Variable	F_1	F_2	F_3
zPATth	0,50	0,23	0,25
zGDPpc	0,62	0,44	0,82
zPECpc	-0,26	-0,58	-0,28
zGREpc	0,47	0,44	0,92
zURDpsk	0,36	0,80	0,38
zURGpor	0,60	0,27	0,30

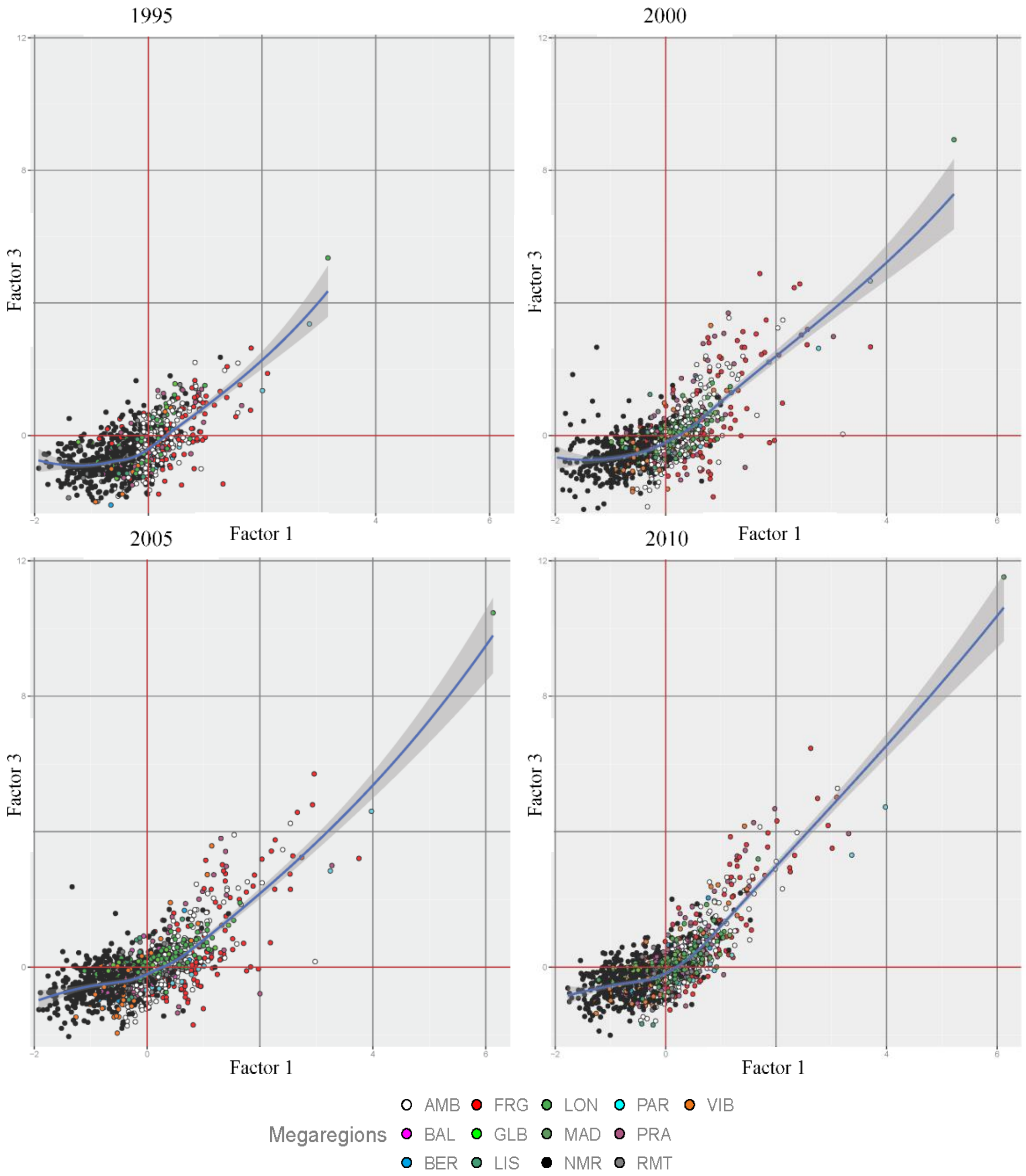
Factors: F_1 (Economic Activity), F_2 (Urban Ecology), F_3 (Social Cohesion). Variables: GPDpc (Gross Domestic Product); GREpc (Gross Rate Employment); PATth (Patent Applications); URDpsk (Urban Density); URGpor (Urban Surface); PECpc (Primary Energy Consumption).

Figure 4 Factor 2 (*economic activity*) *vs* Factor 1 (*urban ecology*) for the NUTS 3 that are part of a megaregion; 1995-2010



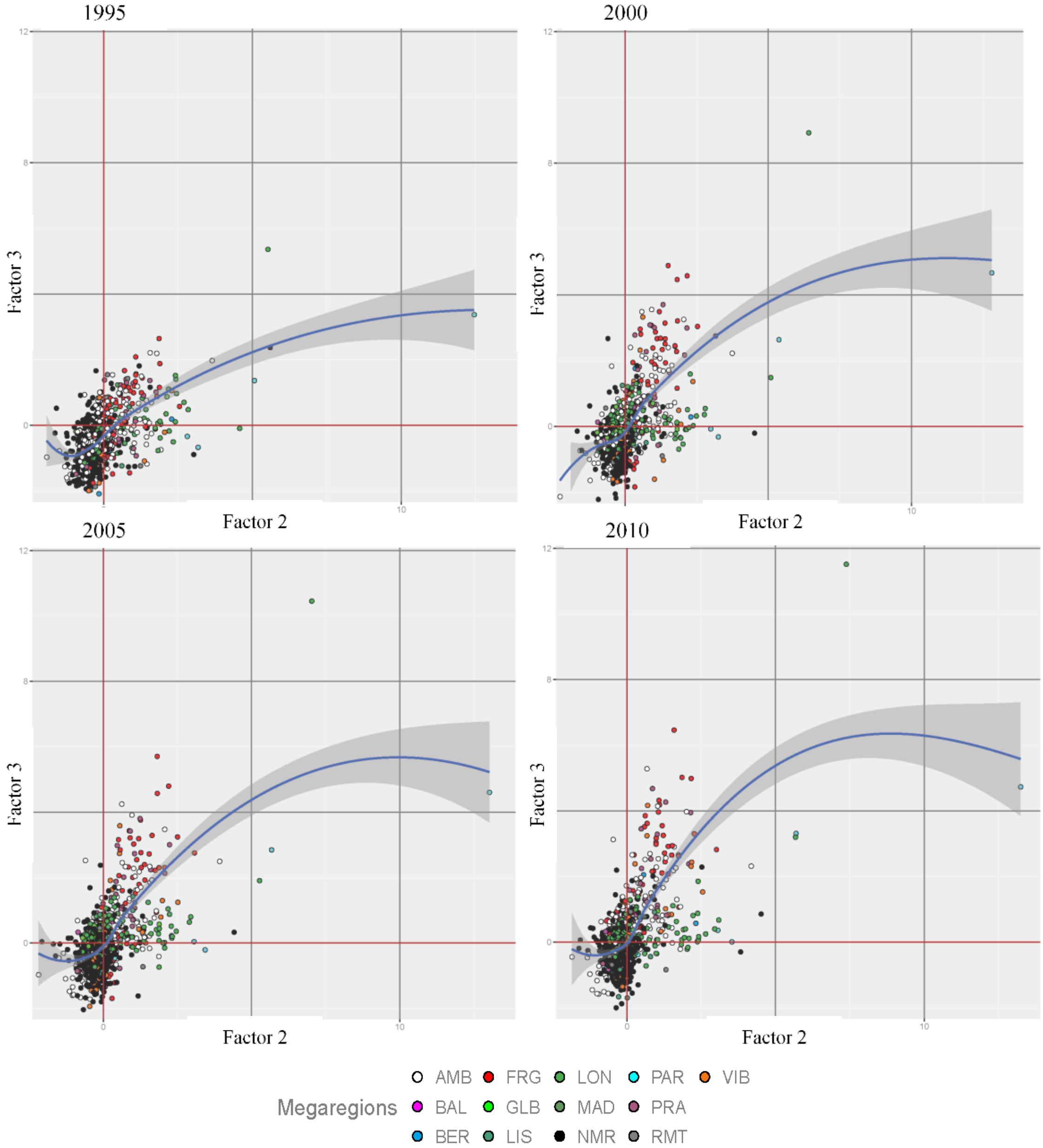
Megaregions: AMB (*Amsterdam-Brussels-Antwerp*); BAL (*Barcelona-Lyon*); BER (*Berlin*); FRG (*Frankfurt-Stuttgart*); GLB (*Glasgow-Edinburgh*); LIS (*Lisbon*); LON (*London*); MAD (*Madrid*); PAR (*Paris*); PRA (*Prague*); RMT (*Rome-Milan-Turin*); VIB (*Vienna-Budapest*); NMR (belongs to no megaregion).

Figure 5 Factor 3 (*economic activity*) *vs* Factor 1 (*social cohesion*) for NUTS 3 that belong to a megaregion; 1995-2010



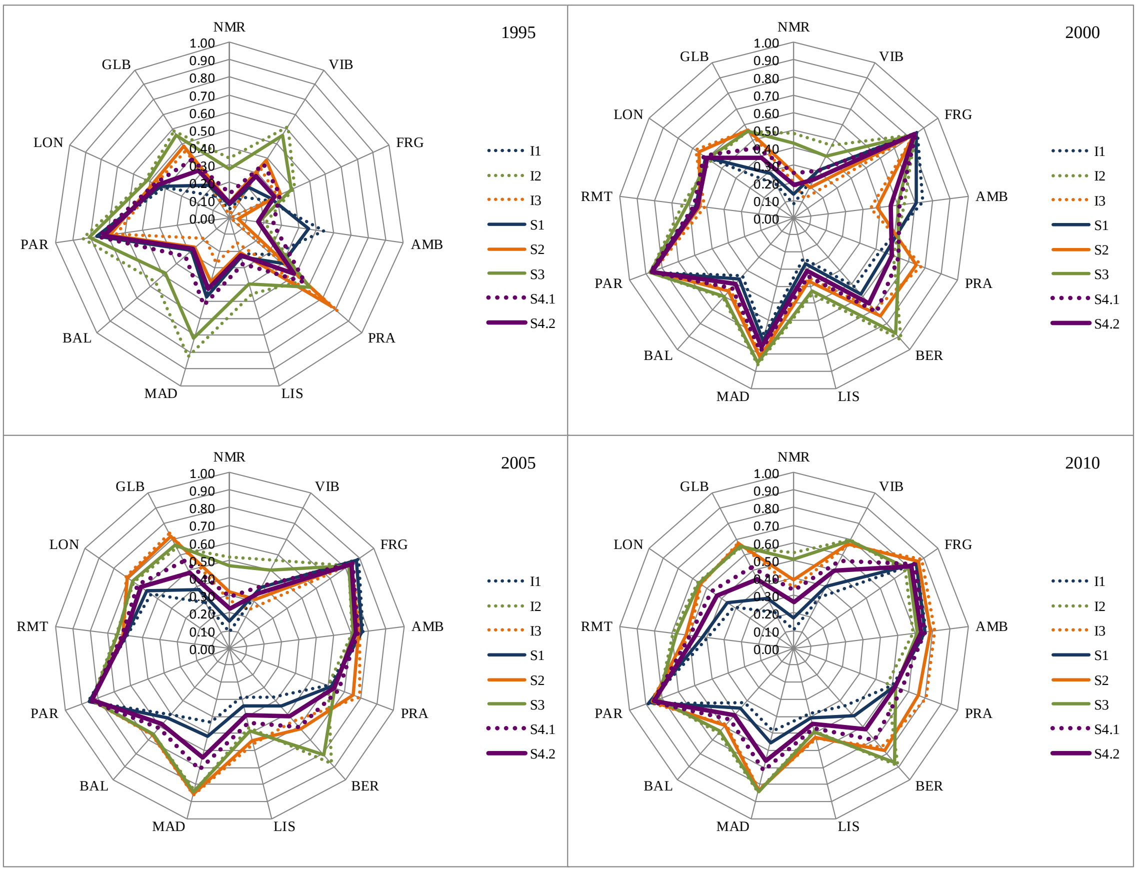
Megaregions: AMB (*Amsterdam-Brussels-Antwerp*); BAL (*Barcelona-Lyon*); BER (*Berlin*); FRG (*Frankfurt-Stuttgart*); GLB (*Glasgow-Edinburgh*); LIS (*Lisbon*); LON (*London*); MAD (*Madrid*); PAR (*Paris*); PRA (*Prague*); RMT (*Rome-Milan-Turin*); VIB (*Vienna-Budapest*); NMR (belongs to no megaregion).

Figure 6 Factor 3 (*urban ecology*) vs. Factor 2 (*social cohesion*) for NUTS 3 that belong to a megaregion; 1995-2010



Megaregions: AMB (*Amsterdam-Brussels-Antwerp*); BAL (*Barcelona-Lyon*); BER (*Berlin*); FRG (*Frankfurt-Stuttgart*); GLB (*Glasgow-Edinburgh*); LIS (*Lisbon*); LON (*London*); MAD (*Madrid*); PAR (*Paris*); PRA (*Prague*); RMT (*Rome-Milan-Turin*); VIB (*Vienna-Budapest*); NMR (belongs to no megaregion).

Figure 7 Comparison of the simple-indices (I_1 –Economic activity, I_2 –Urban ecology, I_3 –Social cohesion) and the compound-indices (S_1 –Economic development, S_2 –Social sustainability, S_3 –Environmental sustainability, S_4 –Inclusive development: $S_{4.1}$ –each factor same weight, $S_{4.2}$ –each factor same role and balanced) at megaregional level; 1995-2010



Note 1: AMB (*Amsterdam-Brussels-Antwerp*); BAL (*Barcelona-Lyon*); BER (*Berlin*); FRG (*Frankfurt-Stuttgart*); GLB (*Glasgow-Edinburgh*); LIS (*Lisbon*); LON (*London*); MAD (*Madrid*); PAR (*Paris*); PRA (*Prague*); RMT (*Rome-Milan-Turin*); VIB (*Vienna-Budapest*); NMR (does not belong to a megaregion). There are NUTS 3 without values for some years and variables (mainly 1995), and so were not taken into account.

Note 2: Values in italics (mainly in bold) are aggregated data with many missing values (not imputed), so they are little representative (mainly correspond to 1995 and come from lost data of GRE and PAT). LIS has little data every year (see Table 2).

Table 1 Compound-indices (S_j) in conceptual scenarios of sustainable progress in city networks taking into account simple-indices (I_j) and their penalty factors (w_j)

S_1 Economic development		
I_1	Economic activity	$w_1 = 0.8$
I_2	Urban ecology	$w_2 = 0.1$
I_3	Social cohesion	$w_3 = 0.1$
S_2 Social sustainability		
I_1	Economic activity	$w_1 = 0.0$
I_2	Urban ecology	$w_2 = 0.2$
I_3	Social cohesion	$w_3 = 0.8$
S_3 Environmental sustainability		
I_1	Economic activity	$w_1 = 0.0$
I_2	Urban ecology	$w_2 = 0.8$
I_3	Social cohesion	$w_3 = 0.2$
$S_{4.1}$ Inclusive development		
I_1	Economic activity	$w_1 = 1/3$
I_2	Urban ecology	$w_2 = 1/3$
I_3	Social cohesion	$w_3 = 1/3$
$S_{4.2}$ Inclusive development (balanced)		
Min (I_1,I_2,I_3)		$w_1 = 1/2$
Med (I_1,I_2,I_3)		$w_2 = 1/3$
Max (I_1,I_2,I_3)		$w_3 = 1/6$

Table 2 Values of simple-indices (I_1 –Economic activity, I_2 –Urban ecology, I_3 –Social cohesion) and compound-indices (S_1 –Economic development, S_2 –Social sustainability, S_3 –Environmental sustainability, S_4 –Inclusive development: $S_{4.1}$ –each factor same weight, $S_{4.2}$ –each factor same role and balanced) at megaregional level; 1995-2010

Year	Megaregion	I_1	I_2	I_3	S_1	S_2	S_3	$S_{4.1}$	$S_{4.2}$
1995	NMR	*0.05*	*0.34*	*0.02*	*0.07*	*0.08*	*0.28*	*0.14*	*0.08*
	VIB	*0.14*	*0.62*	*0.33*	*0.20*	*0.39*	*0.56*	*0.36*	*0.28*
	FRG	*0.24*	*0.41*	*0.30*	*0.26*	*0.32*	*0.38*	*0.31*	*0.29*
	AMB	*0.54*	*0.21*	***0.01***	*0.45*	*0.05*	***0.17***	***0.25***	***0.16***
	PRA	*0.33*	*0.55*	*0.82*	*0.40*	*0.76*	*0.61*	*0.57*	*0.48*
	LIS	***0.22***	***0.45***	***0.15***	***0.23***	***0.21***	***0.39***	***0.27***	***0.22***
	MAD	0.45	0.83	0.27	0.47	0.38	0.72	0.52	0.43
	BAL	0.27	0.56	0.18	0.29	0.26	0.48	0.34	0.27
	PAR	0.76	0.84	0.65	0.76	0.69	0.81	0.75	0.72
	LON	0.41	0.52	0.46	0.43	0.47	0.51	0.47	0.45
	GLB	0.14	0.59	0.46	0.22	0.48	0.56	0.40	0.32
2000	NMR	0.08	0.48	0.21	0.13	0.26	0.42	0.25	0.19
	VIB	*0.31*	*0.47*	*0.13*	*0.31*	*0.20*	*0.40*	*0.30*	*0.24*
	FRG	0.85	0.84	0.84	0.85	0.84	0.84	0.84	0.84
	AMB	0.74	0.64	0.44	0.70	0.48	0.60	0.61	0.56
	PRA	0.53	0.61	0.78	0.56	0.75	0.64	0.64	0.60
	BER	0.53	0.92	0.70	0.58	0.74	0.88	0.72	0.65
	LIS	***0.24***	***0.45***	***0.35***	***0.27***	***0.37***	***0.43***	***0.35***	***0.31***
	MAD	0.69	0.86	0.80	0.72	0.81	0.85	0.78	0.76
	BAL	0.44	0.61	0.54	0.47	0.56	0.59	0.53	0.50
	PAR	0.88	0.88	0.85	0.87	0.86	0.87	0.87	0.86
	RMT	0.56	0.64	0.51	0.56	0.54	0.61	0.57	0.55
	LON	0.60	0.57	0.68	0.61	0.66	0.59	0.62	0.60
	GLB	0.22	0.56	0.57	0.29	0.56	0.56	0.45	0.39
2005	NMR	0.09	0.52	0.27	0.15	0.32	0.47	0.29	0.22
	VIB	*0.39*	*0.57*	*0.25*	*0.39*	*0.31*	*0.50*	*0.40*	*0.35*
	FRG	0.89	0.82	0.85	0.88	0.84	0.83	0.86	0.84
	AMB	0.77	0.70	0.75	0.76	0.74	0.71	0.74	0.73
	PRA	0.60	0.60	0.79	0.62	0.76	0.64	0.67	0.63
	BER	0.38	0.88	0.55	0.44	0.62	0.81	0.60	0.52
	LIS	***0.29***	***0.47***	***0.56***	***0.34***	***0.54***	***0.49***	***0.44***	***0.40***
	MAD	0.43	0.84	0.86	0.52	0.86	0.84	0.71	0.64
	BAL	0.50	0.65	0.66	0.53	0.66	0.66	0.61	0.58
	PAR	0.86	0.84	0.83	0.85	0.84	0.84	0.84	0.84
	RMT	0.59	0.65	0.61	0.60	0.61	0.64	0.61	0.60
	LON	0.54	0.66	0.72	0.57	0.71	0.67	0.64	0.61
	GLB	0.29	0.65	0.74	0.37	0.72	0.66	0.56	0.49
2010	NMR	0.10	0.54	0.35	0.17	0.39	0.51	0.33	0.26
	VIB	0.32	0.70	0.66	0.40	0.67	0.69	0.56	0.50
	FRG	0.85	0.77	0.89	0.85	0.87	0.80	0.84	0.82
	AMB	0.76	0.68	0.81	0.76	0.79	0.71	0.75	0.73
	PRA	0.60	0.57	0.81	0.62	0.76	0.62	0.66	0.62
	BER	0.44	0.89	0.75	0.52	0.78	0.87	0.70	0.62
	LIS	***0.39***	***0.48***	***0.53***	***0.41***	***0.52***	***0.49***	***0.47***	***0.44***
	MAD	0.49	0.84	0.83	0.56	0.83	0.84	0.72	0.66
	BAL	0.42	0.65	0.58	0.46	0.59	0.63	0.55	0.51
	PAR	0.89	0.82	0.87	0.88	0.86	0.83	0.86	0.85
	RMT	0.49	0.69	0.60	0.52	0.62	0.67	0.59	0.56
	LON	0.41	0.66	0.64	0.46	0.64	0.65	0.57	0.53
	GLB	0.23	0.65	0.68	0.32	0.67	0.65	0.52	0.44

Note 1: AMB (*Amsterdam-Brussels-Antwerp*); BAL (*Barcelona-Lyon*); BER (*Berlin*); FRG (*Frankfurt-Stuttgart*); GLB (*Glasgow-Edinburgh*); LIS (*Lisbon*); LON (*London*); MAD (*Madrid*); PAR (*Paris*); PRA (*Prague*); RMT (*Rome-Milan-Turin*); VIB (*Vienna-Budapest*); NMR (does not belong to a megaregion).

Note 2: Values in italics (mainly in bold) are aggregated data with many missing values (not imputed), so they are little representative (mainly correspond to 1995 and come from lost data of GRE and PAT). LIS has little data every year.

Appendices

Appendix A

For the initial variables (y = PATth, GDPpc, PECpc and GREpc), their values in the megaregions-year units are computed as weighted averages of the correspondent NUTS 3 values. Indeed, for each megaregion i and year k, the value of the variable y_{ik} in this unit is a weighted mean of the values y_{jk} of the same variable in the NUTS 3j-year k belonging to the region i, being the weights w_{jk} the corresponding proportion of population:

$$y_{ik} = \sum_{\forall j \in i} y_{jk} \text{ x } w_{jk}, \text{ where } w_{jk} = POPpr_{jk} = \frac{POP_{jk}}{\sum_{\forall m \in i} POP_{mk}}$$

The variables URDpsk and URGpor in a region i in year k are computed as quotients of aggregated values, based on estimates derived from the relation between lighted areas (*LArea*) and total areas (*TArea*):

$$URGpor_{ik} = \frac{LArea_{ik}}{TArea_{ik}} \text{x } 10^2 = \frac{\sum_{\forall j \in i} LArea_{jk}}{\sum_{\forall j \in i} TArea_{jk}} \text{x} 10^2$$

$$URDpsk_{ik} = \frac{POP_{ik}}{TArea_{ik}} \text{x } 10^6 = \frac{\sum_{\forall j \in i} POP_{jk}}{\sum_{\forall j \in i} TArea_{jk}} \text{x} 10^6$$

Appendix B

Proof of formula (2): In the first line, we use: the fact that covariance and correlation coincide for scaled variables; the expression of the covariance for centered vectors; and equation (1). In the second line, we use the assumptions made on the model (U and F are incorrelated and the components of U are incorrelated too), which imply $E[FU^t] = E[UF^t] = 0$ and $E[UU^t] = 0$, and that $E[FF^t] = \sum_F$:

$$R = \text{cov(Z)} = E[ZZ^t] = E[(QF + U)(QF + U)^t]$$

$$= QE[FF^t]\, Q^t + E[UF^t]\, Q^t + QE\,[FU^t] + E\,[UU^t] = Q \textstyle\sum_F Q$$

Proof of formula (3): Using that Z and F are both scaled vectors and analogous arguments that above, we have:

$$R_{XF} = R_{XF} = \mathrm{cov}(Z,F) = E[ZF^t] = E[(QF + U)F^t] = QE[FF^t] + E[UF^t] = Q\Sigma_F$$

Proof of formula (4): Using the Thompson method to compute scores and the formula (2):

$$F = Z(\mathrm{cov}\ (Z))^{-1}\ \mathrm{cov}\ (Z,F) = ZR^{-1}Q\Sigma_F$$

Proof of equations (5): These equations express the scores in terms of the unscaled variables (say *X*: PATth, GDPpc, PCEpc, GREpc, URDpsk, URGpor), instead of formula 4 where scores are expressed in terms of the scaled ones (*Z*: zPATth, zGDPpc, zPCEpc, zGREpc, zURDpsk, zURGpor). Then, to obtain these expressions it suffices to invert the scaling procedure.

Appendix C

Function cfa() of Lavaan library is used, estimating parameters by means of the unweighted least squares method (ULS). The optimization procedure converged after 40 iterations, and the out-of-diagonal differences between the original and the reproduced correlation matrices are shown here:

Residuals out of the diagonal

	PATth	GDPpc	PECpc	GREpc	URDpsk	URGpor
PATth	****					
GDPpc	0.07	****				
PECpc	0.08	0.04	****			
GREpc	0.04	0.00	0.02	****		
URDpsk	0.10	0.03	0.00	0.02	****	
URGpor	0.00	0.05	0.09	0.03	0.07	****

Appendix D

Box-Cox transform: It is well known that the Box-Cox $\{\varphi_\lambda, \lambda\}$ family of functions (Box and Cox, 1964) may be helpful to reduce the skewness in data and, often reducing the effects of outliers. The function powerTransform() in R-library *car* (see Fox and Weisberg, 2011) provides tools to estimate the optimal value of the parameter λ. For that, it is necessary to avoid negative scores. In our setting, a (+10) translation assuring positive scores is applied to each factor F_j, followed by a Box-Cox transformation with an optimal value λ_j, giving rise to the transformed scores tF_j, for $j = 1, 2, 3$:

$$tF_j = \frac{(F_j + 10)^{\lambda_j} - 1}{\lambda_j}, \ \lambda_1 \approx 0.964, \ \lambda_2 \approx -4.505, \ \lambda_3 \approx -3.933$$

The Laplace' density and cumulative distribution functions are, respectively:

$$f(x) = \frac{1}{2\beta} \exp\left(-\frac{|x-m|}{\beta}\right); \ \Phi(x) = \frac{1}{2} + \frac{1}{2}\operatorname{sgn}(x-m) - \exp\left(-\frac{|x-m|}{\beta}\right)$$

Given a sample $y_1 \dots y_n$, the maximum likelihood estimators (MLE) of the parameters m and β are the sample median (50th-percentile) and the mean absolute deviations from the sample, respectively:

$$\hat{m} = C_{50} \qquad \hat{\beta} = \frac{1}{n}\sum_{i=1}^{n} |y_i - \hat{m}|$$

Using these estimates for each transformed factor tFj, the corresponding Laplace distribution function Φ_j is compared to the empirical cumulative distribution function (ecdf) of the same transformed scores (Figure AD). In the text, tFj are denoted Fj, for the sake of notational simplicity.

Figure AD. Comparison of the empirical (in black) and Laplace cumulative distribution functions (in green) for the three factors scores after transformation (F_1 –Economic activity, F_2 –Urban ecology, F_3 –Social cohesion)

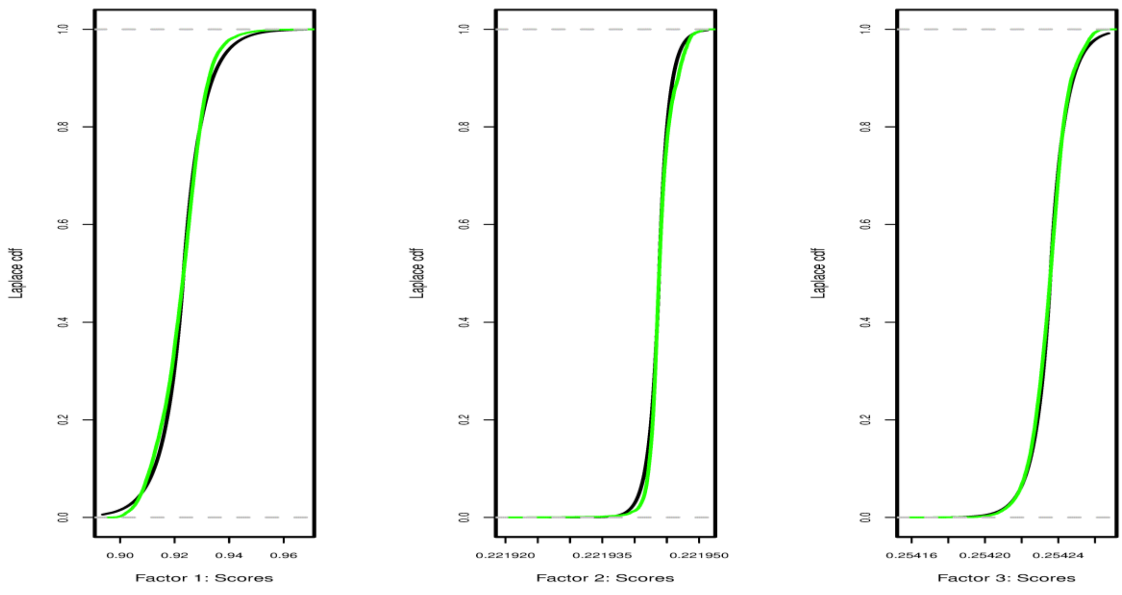


Fig. A1 Index S_1 (*economic development*) at regional level (NUTS 3); 1995-2010

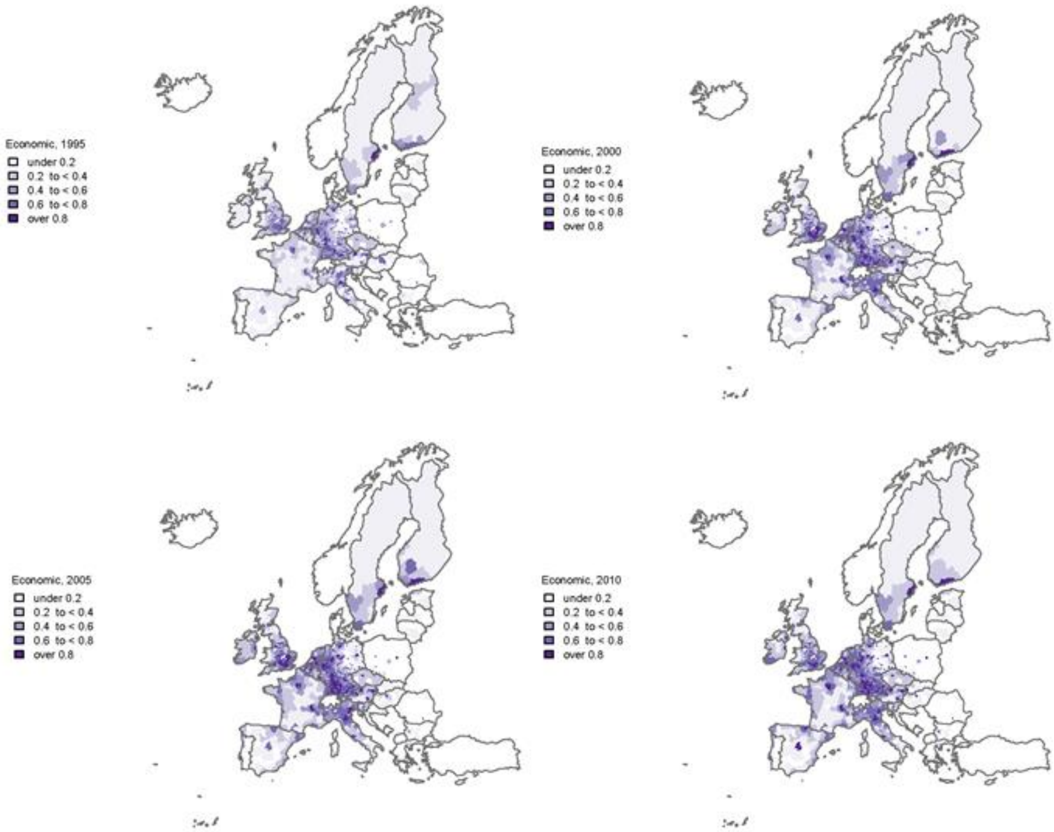


Fig. A2 Index S_2 (*social sustainability*) at regional level (NUTS 3); 1995-2010

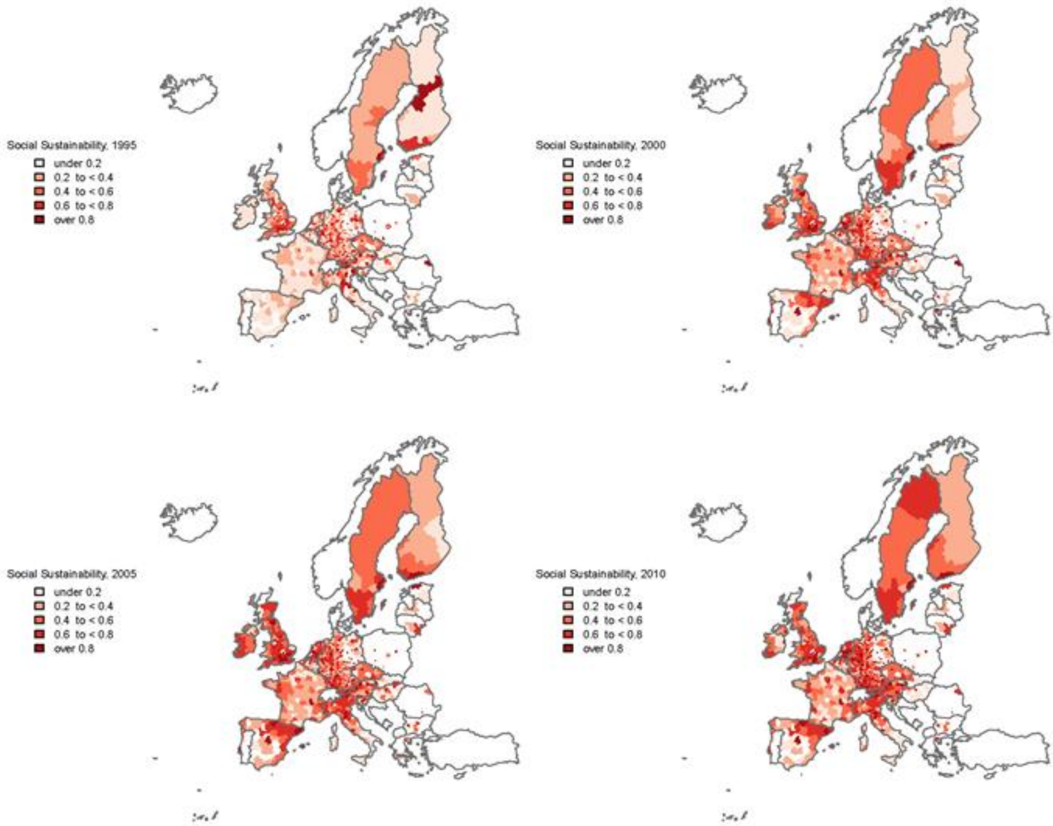


Fig. A3 Index S_3 (*environmental sustainability*) at regional level (NUTS 3); 1995-2010

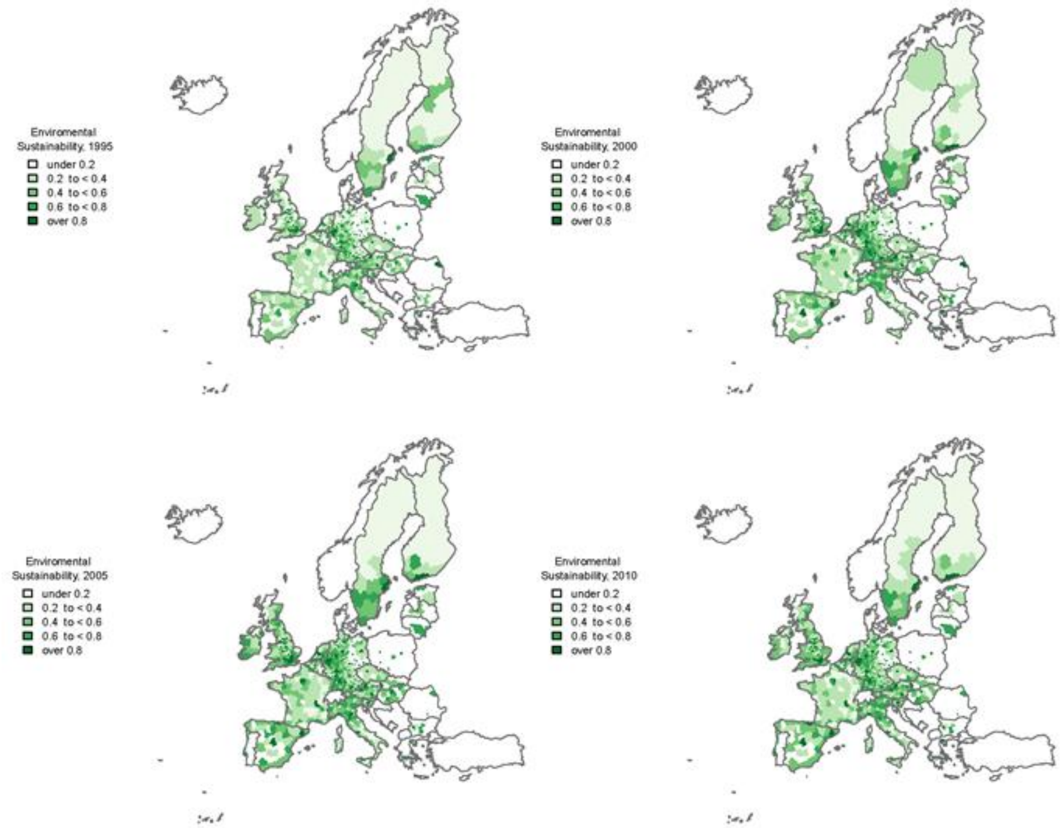


Fig. A4 Index $S_{4.2}$ (*inclusive development*) at regional level (NUTS 3); 1995-2010