Socio-Economic Planning Sciences 94 (2024) 101917

Contents lists available at ScienceDirect

Socio-Economic Planning Sciences

journal homepage: www.elsevier.com/locate/seps

The functional distance-based approach: An application on long-term Metropolitan Development

Francesca Fortuna [a,*], Alessia Naccarato [a], Luca Salvati [b]

[a] Department of Economics, Roma Tre University, Via Silvio D'Amico 77, 00145 Rome, Italy
[b] Department MEMOTEF, Sapienza University, Via Del Castro Lurenziano 9, 0061 Rome, Italy

ARTICLE INFO

Keywords:
Metropolitan development
FDA
Distance correlation

ABSTRACT

Metropolitan development is a complex phenomenon depending on the joint action of multiple socioeconomic and territorial factors, which are generally quantified through indicators interacting across spatial and temporal scales. A comprehensive understanding of metropolitan development requires a detailed examination of the dependence structure within the development trajectories of both urban and suburban areas. Multidimensionality, non-linearity, and evolutionary components at the base of metropolitan development should be jointly considered when investigating the socioeconomic mechanisms underlying the long-term expansion of cities and their surrounding regions. In this multidimensional perspective, the Functional Data Analysis approach enables the simultaneous investigation of the level and evolution over time of multiple socioeconomic processes through summary functional measures, such as the modified hypograph index and the weighted integrated first derivative. Starting from these tools, a functional distance-based approach is proposed to study the relationship between urban and suburban areas from a twofold perspective: focusing on the entire set of metropolitan development indicators, and considering each dimension individually. This approach contributes to overpass the limitations of standard econometric approaches offering a multivariate measure of change in local systems experiencing complex development trajectories. The proposal is illustrated via a simulation study and a real data set.

1. Introduction

Metropolitan Development (MD) is a multidimensional phenomenon that reflects the joint action of multiple factors, called MD indicators (MDIs), referring to territorial and socioeconomic processes. The level and the evolutionary behavior of such factors characterize MD, which in turns is the result of a complex interplay of demographic, social, economic, institutional, ecological and organizational variables interacting across spatial and temporal scales. Despite the different theoretical interpretations of city growth in several fields, urbanization and suburbanization are considered the most important stages of MD [1,2]. For a long time, a comprehensive description of such processes was performed within the framework of the Spatial Cycle Theory (SCT) [1,3]. The SCT assumes metropolitan regions as composed of a core surrounded by a suburban ring, which interact through hierarchical (core–periphery) interdependencies [4]. The analysis of core–periphery relations was basically performed using descriptive statistical tools [5,6] to quantify the intrinsic rates of change of key MDIs, such as population density, per-capita income, unemployment rates, among others. However, the study of single MDIs gives a partial evidence of city dynamics, and highlights the importance of multidimensional approaches to urban complexity [7–9]. Multi-centered structures constituting the majority of metropolitan regions in advanced economies, and especially in Europe, result in a continuous interaction between central cities and peripheries, reflecting cooperation and complementarity [7]. Assuming the bi-directional linkage between urban and suburban development processes [10], econometric models were frequently implemented in studying the decline of inner cities and peri-urban expansion [11–14]. However, these models show critical issues mainly concerning the linearity of the relationships [4]. In this context, analytical tools reflecting non-linearity and the intrinsic multidimensionality of MD processes seem to be particularly appropriate [15]. In addition, MDIs vary over time, thus their temporal variability should be considered in the MD evaluation.

Based on these premises, this paper aims to investigate the uni-directional or bi-directional relationship between urban and suburban areas, while simultaneously considering the non-linearity, the multi-dimensionality, and the evolutionary nature of MD. To this end, the Functional Data Analysis (FDA) approach [16,17] is considered. Thus, the values of each MDI at different times represent discrete observations

* Corresponding author.
E-mail addresses: francesca.fortuna@uniroma3.it (F. Fortuna), alessia.naccarato@uniroma3.it (A. Naccarato), luca.salvati@uniroma1.it (L. Salvati).

https://doi.org/10.1016/j.seps.2024.101917
Received 26 March 2024; Received in revised form 29 April 2024; Accepted 3 May 2024
Available online 8 May 2024
0038-0121/© 2024 The Author(s). Published by Elsevier Ltd. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

of an underlying smooth function [18,19]. The main advantage of the FDA approach lies in the use of functional tools, which reflect distinctive features of the curves, such as their level and evolution [20–26]. These measures can be used as coordinates of each MDI in a Cartesian plane. The two functional measures considered in this paper are the Modified Hypograph Index (MHI) [27] and the Weighted Integrated First Derivative (DW) [18,19]. MHI and DW are scalar measures, which summarize the behavior of the functions over time. Specifically, MHI gives information on the indicator's level, while DW quantifies its evolutionary behavior, by distinguishing how long the function shows a certain trend.

The use of hypograph together with other functional tools was already proposed in the literature, especially to study the shape and the magnitude of curves and provide an ordering of the functions [28–30]. In particular, Pulido et al. [31] consider the combination of hypograph and epigraph applied to both the original curves and their derivatives for clustering purposes. The authors demonstrate that the joint use of these indexes largely characterizes aspects of the curves in the sample. Our approach differs from that of Pulido et al. [31] in that we consider MHI together with DW to capture both the level and the time period in which a curve displays a specific trend.

Since each MDI is represented as a point on the $MHI - DW$ plot, it is possible to compute both a global distance between MD of different areas and pairwise distances between MDIs of different areas. Specifically, a functional distance-based approach is suggested to evaluate the synergic relationship between urban and suburban areas. The proposed method consists of two distance measures. The first one is the distance correlation coefficient [32], which establishes a global and multidimensional measure of independence in long-term MD pattern between urban and suburban regions. The second one is based on pairwise distances between MDIs of different areas and reflects differences due to specific MD dimensions. With reference to the second distance measure, a further proposal is to identify the role of the level and evolutionary components in defining differences between MD processes. Thus, the analysis of MD is conducted from two distinct perspectives. Firstly, by comparing urban and suburban areas based on the entire set of MDIs; secondly, by comparing the two areas according to each MDI.

The proposed approach introduces significant advantages over classical methods, offering a more nuanced analysis of MD, that considers its multidimensional, non-linear, and evolving nature. First, the FDA approach, by its nature, captures non-linearity in relationships. Second, this approach incorporates temporal variability of MDIs into the evaluation of MD by representing each MDI as discrete observations of an underlying smooth function. Finally, FDA utilizes functional tools that reflect distinctive features of curves, such as their level and evolution. These functional tools allow to implement a functional distance-based approach, to investigate MD of urban and suburban areas from both a global and dimension-specific perspective. Specifically, the distance correlation coefficient is a global and multidimensional measure of relationship between urban and suburban areas. Indeed, unlike traditional methods, which focus on individual indicators at a specific time span, the distance correlation coefficient considers the joint action of several MDIs taking into account their dynamic over time.

The proposal is applied to a real data-set, concerning two interacting districts of Athens' metropolitan region (Greece): central city and the surrounding suburbs for which a multivariate set of MDIs observed over a long time frame is considered.

The paper is organized as follows: Section 2 deals with the MDIs representation in the FDA context and introduces the two functional measures MHI and DW. Section 3 illustrates the two distance measures within the analytical framework. In Section 4 a simulation study is performed. In Section 5 long-term development of the Athens' metropolitan region is analyzed using 14 MDIs for both central city and suburbs on a year base between 1966 and 2008. Discussion and concluding remarks are provided in Section 6.

2. Metropolitan development in a functional framework

Let I_{il} be the observed value of the ith MDI, $i = 1, 2, \dots, n$, at the time point t_l, $l = 1, \dots, L$. These time measurements represent discrete and noisy observations of an underlying smooth function, $I_i(t)$, which belongs to a temporal domain T, with $t \in T$. Let us assume that the smooth functions belong to the Hilbert space of square integrable functions with the usual inner product [33].

Starting from the discrete temporal sequence of the ith indicator, the smooth function $I_i(t)$ is reconstructed via a linear expansion in terms of B known basis functions, $\{\psi_1(t), \psi_2(t), \dots, \psi_B(t)\}$, as follows [16]:

$$I_i(t) = \sum_{b=1}^{B} c_{ib}\psi_b(t), \tag{1}$$

where c_{ib} is the ith basis coefficient, $\psi_b(t)$ represents the bth basis function, and B is the total number of basis functions. Basis expansion methods represent the potentially infinite dimensional world of functions within the finite dimensional framework of vectors of basis coefficients. After selecting an appropriate basis system, the shape of the curves is defined by the basis coefficients. Thus, the problem of finding the smooth function becomes a problem of finding the coefficients of the basis expansion.

The expansion in Eq. (1) also provides expressions of the derivatives in terms of the derivatives of the basis functions; so that the first and second order derivatives are the following:

$$I_i'(t) = \sum_{b=1}^{B} c_{ib}\psi_b'(t); \quad I_i''(t) = \sum_{b=1}^{B} c_{ib}\psi_b''(t). \tag{2}$$

To control the amount of smoothness, the basis coefficients are obtained by minimizing, for each i, the following penalized residual sum of squares:

$$PENSSE_i = \sum_{l=1}^{L}\Big(I_{il} - \sum_{b=1}^{B} c_{ib}\psi_b(t)\Big)^2 + \lambda \int [I''(t)]^2\, dt, \tag{3}$$

where λ is a smoothing parameter, and $\int [I''(t)]^2\, dt$ is the total curvature of the function, given by the integrated squared second derivative. The total curvature is a penalty term, which measures the roughness of the function. Indeed, highly variable functions can be expected to yield high values of the total curvature because their second derivatives are larger over at least some of the range of interest [16]. The smoothing parameter λ reflects the trade-off between data fit and the variability of the function. Low values of λ result in model over-fitting, whereas high values of λ result in over-smoothing. Generalized cross-validation is typically used to choose the smoothing parameter λ [16,34]. Reinsch [35] and De Boor [36] demonstrated that the function that minimizes the penalized residual sum of squares in Eq. (3) is a cubic spline with knots at the data points t_l. Compared to the classical least squares criterion, the roughness penalty approach produces better results in the estimation of derivatives, ensuring that the latter are smoothed. For further details on the FDA approach refer to Ramsay and Silverman [16].

To provide a complete picture of MD, both the magnitude and evolutionary behavior of the MDIs should be taken into account in the evaluation process. To this end, two functional measures are considered: MHI and DW, which reflect the magnitude and the evolutionary behavior of each indicator, respectively.

The MHI of a given curve $I(t)$, with respect to a sample of functions, $I_i(t)$, $i = 1, 2, \dots, n$, is defined as follows [27]:

$$MHI\Big(I(t)\Big) = \sum_{i=1}^{n} \frac{\tau(\{G(I) \subseteq hyp(I)\})}{n\tau(T)}; \tag{4}$$

where $G(I) = \{(t, I(t)), t \in T\}$ is the graph of the functional indicator $I(t)$; $hyp(I) = \{(t, y) \in T \times \mathbb{R} : y \le I(t)\}$ is the hypograph of $I(t)$, that is the area under its graph, and $\tau(\cdot)$ stands for the Lebesgue measure on $\mathbb{R}$. MHI represents the proportion of T in which the curves of the sample are below $I(t)$. Thus, a high value of MHI indicates that many functions are contained in the hypograph of a given curve; whereas a MHI value equal to 0.5 indicates that the curve is located in a central position.

The main disadvantage of MHI is that it does not account for how variable a function is in the time span so that different shapes of the curves can yield equal values of MHI [18]. DW provides a solution to this limitation by reflecting the evolutionary behavior of the function and distinguishing how long it shows a certain trend. Considering $M+1$ sub-intervals of the domain, according to a set of points, $m = 1, 2, \ldots, M$, which are defined by the changes in the first derivative sign, DW is obtained as a linear combination of the integrated first derivatives in each sub-interval, with coefficients equal to the proportion of the width of each sub-interval [18]:

$$DW_i = \sum_{j=1}^{M+1} D_{ij} w_{ij}, \tag{5}$$

where D_{ij} is the integrated first derivatives of the ith function in the jth sub-interval of the domain, and the weights w_{ij} are computed as follows:

$$w_{ij} = \frac{(t_m - t_{m-1})}{(t_L - t_0)}, \tag{6}$$

where t_0 and t_L are the start and end points of the domain of the curves.

To jointly consider the magnitude and the evolution of a functional indicator, Fortuna et al. [18] proposed a visualization tool called $MHI - DW$ plot, that is a Cartesian plane defined by $MHI \in [0, 1]$ (the x-axis) and $DW \in (-\infty, +\infty)$ (the y-axis). The $MHI - DW$ plot is divided into four quadrants, which result from the intersection of the axes at the coordinate point $MHI = 0.5$ and $DW = 0$. The value $MHI = 0.5$ discriminates among functions that are below or above the median magnitude, and the value $DW = 0$ distinguishes between functions that experience better ($DW > 0$) or worse ($DW < 0$) performances.

3. Functional distance-based measures

The FDA approach enables to represent each MDI as a point on the $MHI - DW$ plot and, consequently, define functional distance-based measures to explore both the overall relationship among neighboring areas and the dimension-specific pairwise dissimilarities among indicators, as well as the contribution of each dimension (i.e. MHI and DW) to these dissimilarities. Referring to the first aspect, the distance correlation coefficient [32] is considered. It is a measure of dependence between two random vectors of arbitrary and not necessarily equal dimension, which permits to assess the dependence of a MD process within different composing dimensions (in our case, MDIs). The distance correlation coefficient is equal to zero if and only if the random vectors are independent, and, unlike the Pearson's correlation coefficient, it measures both linear and non-linear association between two random variables [32,37].

Given two random variables $\boldsymbol{X}$ and $\boldsymbol{Y}$, taking values in $\mathbb{R}^p$ and $\mathbb{R}^q$, respectively and with finite second moments, the distance covariance function, $V(\boldsymbol{X},\boldsymbol{Y})$, is defined as follows:

$$V(\boldsymbol{X},\boldsymbol{Y}) = \left(\int_{\mathbb{R}^{p+q}} | \phi_{\boldsymbol{X},\boldsymbol{Y}}(t,s) - \phi_{\boldsymbol{X}}(t)\phi_{\boldsymbol{Y}}(s) |^2 \, w(t,s)\, dt\, ds \right)^{-\frac{1}{2}}, \tag{7}$$

where $\phi_{\boldsymbol{X},\boldsymbol{Y}}(t,s)$ is the joint characteristic function of $\boldsymbol{X}$ and $\boldsymbol{Y}$; $\phi_{\boldsymbol{X}}(t)$ and $\phi_{\boldsymbol{Y}}(s)$ is the product of the marginal characteristic functions; p and q are the dimensions of $\boldsymbol{X}$ and $\boldsymbol{Y}$, respectively, $(t,s) \in \mathbb{R}^{p+q}$, and $w(t,s)$ is a weight function, defined as follows:

$$w(t,s) = (c_p c_q \mid t \mid^{1+p} \mid s \mid^{1+q})^{-1}, \quad c_d = \frac{\pi^{\frac{1+d}{2}}}{\Gamma\left(\frac{1+d}{2}\right)}; \quad d = p, q, \tag{8}$$

with $| \cdot |$ the complex norm when its argument is complex, otherwise the Euclidean norm; and $\Gamma(\cdot)$ the Gamma function [37]. The distance correlation between $\boldsymbol{X}$ and $\boldsymbol{Y}$, called $R(\boldsymbol{X},\boldsymbol{Y})$, can be defined as the standardized version of Eq. (7):

$$R(\boldsymbol{X},\boldsymbol{Y}) = \left(\frac{V^2(\boldsymbol{X},\boldsymbol{Y})}{\sqrt{V^2(\boldsymbol{X},\boldsymbol{X})V^2(\boldsymbol{Y},\boldsymbol{Y})}} \right)^{-\frac{1}{2}}, \quad \in (0,1). \tag{9}$$

The empirical distance covariance and correlation measures are functions of the double centered distance matrices of the samples. Specifically, let $(\boldsymbol{X}_i, \boldsymbol{Y}_i) \in \mathbb{R}^{p+q}$, $i = 1, \ldots, n$, be a random sample from the joint distribution of $\boldsymbol{X}$ and $\boldsymbol{Y}$. Let us compute the $n \times n$ Euclidean pairwise distance matrices for the $\boldsymbol{X}$ and the $\boldsymbol{Y}$ samples, with elements $a_{ik} = (\|\boldsymbol{X}_i - \boldsymbol{X}_k\|)$ and $b_{ik} = (\|\boldsymbol{Y}_i - \boldsymbol{Y}_k\|)$, respectively. Then, these matrices are double centered so that their row and column means are equal to zero:

$$A_{ik} = a_{ik} - \bar{a}_{i.} - \bar{a}_{.k} + \bar{a}_{..}, \quad B_{ik} = b_{ik} - \bar{b}_{i.} - \bar{b}_{.k} + \bar{b}_{..}, \quad i,k = 1, \ldots, n, \tag{10}$$

where:

$$\bar{a}_{i.} = \frac{1}{n}\sum_{k=1}^{n} a_{ik}, \;\; \bar{a}_{.k} = \frac{1}{n}\sum_{i=1}^{n} a_{ik}, \;\; \bar{a}_{..} = \frac{1}{n^2}\sum_{i,k=1}^{n} a_{ik}. \tag{11}$$

Similarly, the quantities $\bar{b}_{i.}$, $\bar{b}_{.k}$ and $\bar{b}_{..}$ are computed. The sample distance covariance, $\hat{V}(\boldsymbol{X},\boldsymbol{Y})$, can be defined by the following statistic:

$$\hat{V}(\boldsymbol{X},\boldsymbol{Y}) = \left(\frac{1}{n^2} \sum_{i,k=1}^{n} A_{ik} B_{ik} \right)^{-\frac{1}{2}}, \tag{12}$$

and the sample distance correlation, $\hat{R}(\boldsymbol{X},\boldsymbol{Y})$, is obtained as the standardized version of Eq. (12):

$$\hat{R}(\boldsymbol{X},\boldsymbol{Y}) = \frac{\hat{V}(\boldsymbol{X},\boldsymbol{Y})}{\hat{V}(\boldsymbol{X},\boldsymbol{X})\hat{V}(\boldsymbol{Y},\boldsymbol{Y})}, \tag{13}$$

where $\hat{V}(\boldsymbol{X},\boldsymbol{X}) = \sqrt{\frac{1}{n^2}\sum_{i,k=1}^{n} A_{ik}^2}$.

The main advantage of applying the distance correlation coefficient to the points defined by MHI and DW is that it is possible to assess the multidimensional dependence between MD of different areas by taking into account both their level and their evolutionary behavior over the entire time domain. In other words, we can evaluate the long-term global dependence between neighboring areas. Specifically, $\boldsymbol{X}$ and $\boldsymbol{Y}$ are two random vectors whose generic elements are $x_i = (MHI_{x_i}, DW_{x_i})$ and $y_i = (MHI_{y_i}, DW_{y_i})$ for the center and periphery, respectively.

A consistent test of independence is based on the sample distance covariance: large values of $n\hat{V}^2(\boldsymbol{X},\boldsymbol{Y})$ support the alternative hypothesis that $\boldsymbol{X}$ and $\boldsymbol{Y}$ are dependent [32,38]. Following Székely et al. [32], it can be shown that, under the null hypothesis of independence between the random vectors $\boldsymbol{X}$ and $\boldsymbol{Y}$, $n\hat{V}^2(\boldsymbol{X},\boldsymbol{Y})$ converges in distribution to a quadratic form $Q = \sum_{k=1}^{\infty} \lambda_k Z_k^2$, where Z_k are independent standard normal random variables, and λ_k are eigenvalues that depend on the joint distribution of $(\boldsymbol{X},\boldsymbol{Y})$. On the other hand, if $\boldsymbol{X}$ and $\boldsymbol{Y}$ are not independent, $n\hat{V}^2(\boldsymbol{X},\boldsymbol{Y}) \rightarrow \infty$ as $n \rightarrow \infty$ in probability; hence a test that rejects independence for large $n\hat{V}^2(\boldsymbol{X},\boldsymbol{Y})$ is consistent against dependent alternatives. Székely et al. [32] proposed a Monte Carlo permutation test, which represents the standard method for testing independence based on the distance covariance: one first computes the distance correlation (involving the re-centering of Euclidean distance matrices) between two random vectors, and then compares this value to the distance correlations of many shuffles of the data.

Once an overall measure of independence is defined, the interest is to investigate dissimilarities between areas for each indicator using the generic r distance defined as follows:

$$d^r(x_i, y_i) = \left(\gamma_i^r + \eta_i^r \right)^{\frac{1}{r}}, \tag{14}$$

where $\gamma_i = | MHI_{x_i} - MHI_{y_i} |$ and $\eta_i = | DW_{x_i} - DW_{y_i} |$ and the parameter r defines a specific distance measure. The distance in Eq. (14) reflects how different an indicator is in the two areas in terms of MHI and DW: the greater this distance, the greater the dissimilarity. Since Eq. (14) measures the distance between different areas considering both the level (MHI) and the evolutionary behavior (DW) of each indicator, it is important to evaluate the contribution of each dimension to this distance emphasizing what primarily drives a different behavior of the same indicator in different areas. For example, an indicator might have the same level across two areas, whenever exhibits different evolutionary patterns, indicating improvement, deterioration, or stability in a specific area. To this end, the contribution of each dimension to the distance in Eq. (14) is computed as follows:

$$\alpha_\gamma^r = \frac{\gamma^r}{\gamma^r + \eta^r}; \quad \alpha_\eta^r = \frac{\eta^r}{\gamma^r + \eta^r}; \tag{15}$$

with $\alpha_\gamma^r + \alpha_\eta^r = 1$. For the pairwise comparisons between MDIs of different areas, we consider $r = 1$ in Eq. (14), that is the Manhattan distance, so that the contribution of each dimension is expressed in the same unit as MHI and DW.

4. Simulation study

A simulation study is performed to investigate the ability of the proposed functional-distance measures in evaluating the long-term global dependence between MD of center and periphery. Let us consider a bivariate functional process, $(X(t), Y(t)) : t \in T$, whose components, $x_1(t), \ldots, x_{n_1}(t)$, and $y_1(t), \ldots, y_{n_2}(t)$ are obtained as realizations of a Gaussian process, $X(t)$ and $Y(t) \in L^2(T)$, of the following form:

$$X(t) = \mu_x(t) + e(t),$$
$$Y(t) = \mu_y(t) + e(t), \tag{16}$$

where $t \in T$, $\mu_x(t)$ and $\mu_y(t)$ are the mean of $X(t)$ and $Y(t)$, respectively, and $e(t)$ is a centered Gaussian process with exponential covariance function, which is common for both components:

$$C(s,t) = \alpha \exp(-\beta|s-t|), \tag{17}$$

where the parameters $\alpha, \beta \geq 0$.

Two different scenarios are considered: in the first one, S_1, the components of the bivariate functional process are independent, while in the second one, S_2, they are strongly dependent. In both scenarios, the same number of functions are considered for each component, hence $n_1 = n_2 = 100$. In S_1, the two components presents the same functional mean, which is computed as follows:

$$\mu_x(t) = \mu_y(t) = \sin(2\pi t). \tag{18}$$

In S_2, $\mu_x(t)$ is computed as in Eq. (18), while $\mu_y(t)$ is obtained as a nonlinear transformation of the first component as follows:

$$\mu_y(t) = \exp(\mu_x(t)). \tag{19}$$

For each scenario, the curves are observed in the interval $[0,1]$ with a grid of $L = 90$ equally spaced points and the procedure is replicated $S = 1000$ times. The measure of dependence between the two components is provided by the average Kendall's tau correlation coefficient for bivariate functional data [39], computed across the S replications:

$$\bar{\tau} = \frac{1}{S}\sum_{s=1}^{S} \hat{\tau}_r, \tag{20}$$

with:

$$\hat{\tau}_r = \binom{n}{2}^{-1} \sum_{i<j} \left[2I(x_i \prec x_j \; and \; y_i \prec y_j) + 2I(x_j \prec x_i \; and \; y_j \prec y_i)\right] - 1, \tag{21}$$

and $n = n_1 + n_2$. Eq. (20) represents a benchmark to evaluate the performance of the distance correlation coefficient.

Fig. 1 shows the simulated components of the bivariate functional process for a single simulation, $s = 1$, in S_1 (top panel) and S_2 (bottom panel). The performance of the proposed method is ascertained using the average distance correlation coefficient across the replications:

$$\bar{R}(X,Y) = \frac{1}{S}\sum_{s=1}^{S} \hat{R}_s(X,Y), \tag{22}$$

and assessing the proportion of correct decisions, PCD:

$$PCD = \frac{\sum_{s=1}^{S} I_s}{S}, \tag{23}$$

where $I_s = 1$ if the sth distance correlation test provides a correct decision, that is to accept the null hypothesis of independence for S_1 and reject the independence hypothesis for S_2, using a significance level of $\alpha = 0.01$, and $I_s = 0$ otherwise.

Table 1
Average Kendall's tau correlation coefficient, average distance correlation coefficient with standard deviations in round brackets, and proportion of correct decision for the distance correlation test for S_1 and S_2.

Scenario	$\bar{\tau}$	$\bar{R}(X,Y)$	PCD
S_1	0.00 (0.07)	0.17 (0.04)	99%
S_2	0.90 (0.01)	0.72 (0.04)	100%

Table 1 reports the simulation results for S_1 and S_2. The results show that the functional distance-based approach is able to identify dependence or independence between the components of the functional bivariate process. Indeed, the values of $\bar{R}(X,Y)$ are close to 0 for S_1, and equal to 0.72 for S_2, indicating independence and dependence between the components, respectively. Furthermore, in both scenarios, a correct decision is reached, as showed by the percentage of correct responses equal to 99% in S_1 and 100% in S_2.

5. Application

The proposed approach is adopted to analyze long-term MD characteristics of Athens region (Greece). To this aim, 14 MDIs representative of multiple dimensions of MD in both central city and suburbs are considered.

5.1. Study area

Attica is the densest (administrative) region in Greece, which covers nearly 3800 km^2 of mainland and insular districts [40]. The regional territory is administered by more than 110 mainland municipalities (both urban and rural) and nearly 10 municipalities governing some (small and medium) islands in the Saronikos gulf (Aegean Sea). Taken together, the region mostly consists of mountains bordering the flat area that hosts the Greater Athens' agglomeration (administered by 58 municipalities), and corresponding with the 'central city'. Being located outside the Greater Athens' area, the coastal plains of Messoghia (including Marathona) and Thriasio host the largest proportion of rural population in Attica [41,42].

5.2. Metropolitan development indicators

A total of 14 MDIs covering the time period between 1966 and 2008 was made available on a year base from official statistics, distinguishing two sub-areas: the Greater Athens' area ('center') and the rest of Attica region ('periphery'). MDIs were calculated from elementary data routinely collected by the National Statistical Service of Greece and reorganized in a geo-database, disseminating a vast set of information quantifying economic performances and social dynamics of Greek regions and prefectures. The selected MDIs represent different

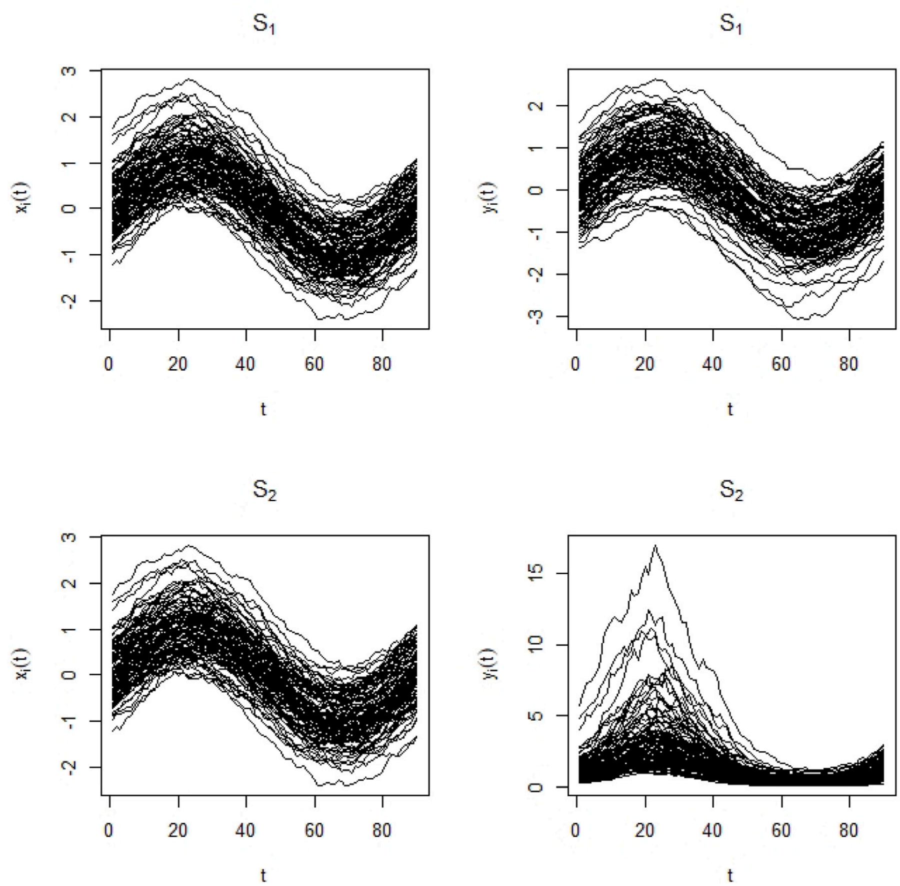

Fig. 1. Simulated components of the bivariate functional process in S_1 and S_2, for $s = 1$.

Table 2
Description of the MDIs of the Athens' metropolitan region.

MDI	Description	MD issue
Den	demographic density (residents per km^2)	PJM
Nat	natural population balance (absolute rate of births to deaths)	PJM
Par	rate of participation (%) in the job market	PJM
New	the ratio of new dwellings per 100 inhabitants	LS
Roo	average number of rooms per new dwelling	LS
Cro	share of cropland (%) in total landscape	LS
Irr	share of irrigated land (%) in total landscape	LS
Ele	per cent share of household electricity consumption in total use	EC
Exp	absolute ratio of expenditures to revenues of municipalities	EC
Car	number of cars per 100 inhabitants	EC
Acc	number of road accidents per 1000 inhabitants	MF
Hos	density (per km^2) of hospital beds	MF
Doc	density (per km^2) of doctors	MF
Dru	density (per km^2) of drugstores	MF

aspects of Athens' long-term MD [8,41,43], referring to four distinctive dimensions (MD issues): (i) PJM: population and job market (3 MDIs); (ii) LS: land-use and settlements (4 MDIs); (iii) EC: economic dynamics (3 MDIs); and (iiii) MF: metropolitan functions (4 MDIs). Table 2 provides a detailed description of each indicator. The MDIs encompassed two sub-periods with different socioeconomic contexts and territorial characteristics [44]: (i) a growth wave dominated by compact urbanization with huge population increase (from mid-1960s to mid-1980s) and (ii) a (less intensive) development wave reflecting spatially dispersed (low-density) settlement expansion with stable or slightly increasing population (from mid-1980s to mid-2000s). Fig. 2 shows the time series of each MDI for the center (red lines in the figure) and periphery (blu lines in the figure) of the Athens' metropolitan region. There are no particular differences between the trends of the MDIs in the two areas.

5.3. MDIs in the functional framework

For the functional representation of MDIs, cubic B-spline basis are considered in Eq. (1) with knots at every data point t_l, $l = 1, \ldots, L$ and $L = 43$ discrete points, following De Boor [36]. Then, the basis coefficients are estimated by adding to the least square criterion a roughness penalty, which involves the curvature of the function as in Eq. (3) and choosing $\lambda = 10^{0.25}$ by generalized cross-validation. The roughness penalized least square smoothing is essential since the first derivative is involved in the analysis. Fig. 3 shows the functional indicators distinguishing the center and periphery of the Athens' metropolitan region with red and blue lines, respectively. To provide information regarding both magnitude and evolutionary behavior of each indicator, MHI and DW are computed as in Eqs. (4) and (5), respectively. Table 3 shows the values of these two functional tools for each MDI, together with the corresponding quadrant in the $MHI - DW$ plot. It is evident that the main difference between the two areas concerns evolutionary behavior rather than magnitude.

Fig. 4 shows the MDIs on the $MHI - DW$ plot. In the first quadrant (top right-hand panel), we find MDIs with high magnitude values ($MHI > 0.5$), and which have improved their performance over time ($DW > 0$). It is worth noting that when DW is close to zero and MHI is close to one, indicators start from very high magnitude levels and maintain them throughout the period considered. In the first quadrant, we find 12 MDIs (6 pertaining to the central area and 6 to the peripheral area). In this quadrant, the MDIs common to both areas are those relating to Den and Exp. In both cases, the evolutionary

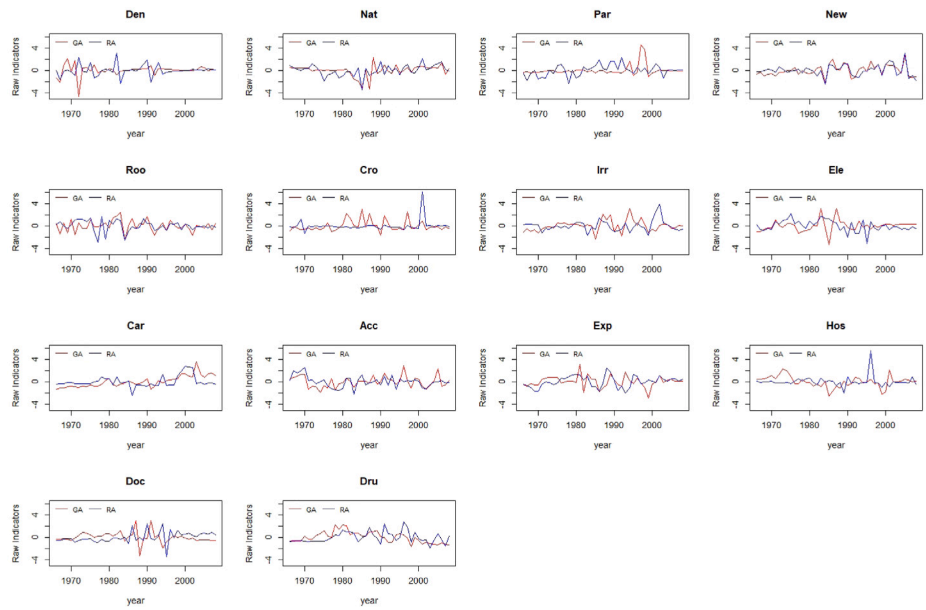

Fig. 2. MDIs for the center (red lines, GA) and periphery (blue lines, RA) of the Athens' metropolitan region.

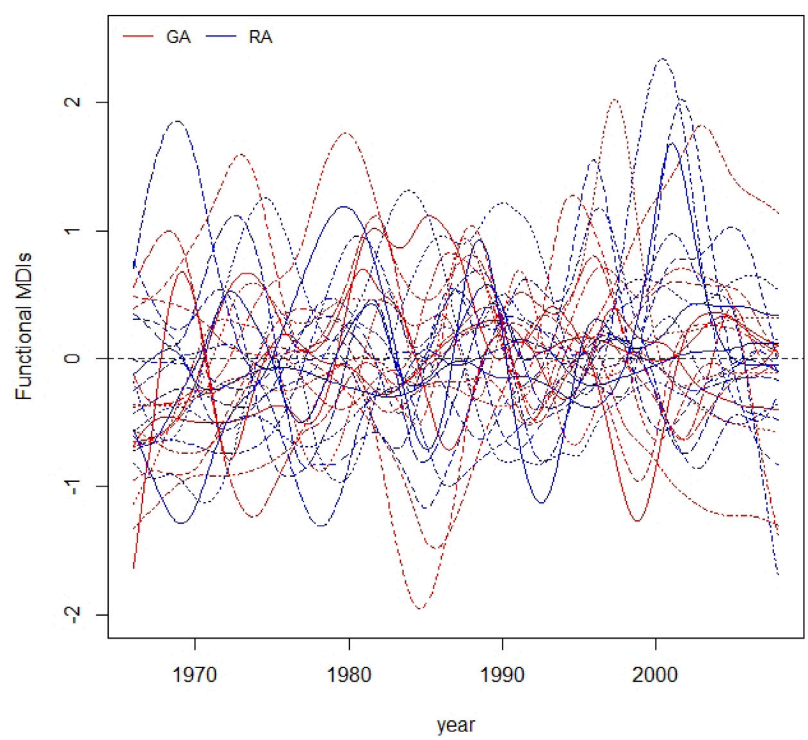

Fig. 3. Functional MDIs for the center (red lines, GA) and periphery (blue lines, RA) of the Athens' metropolitan region.

growth is higher in the periphery, whereas the magnitude is higher in the center in the case of Den and identical in the case of Exp. The second quadrant (bottom right-hand panel) is defined by MDIs that have a magnitude above the median value ($MHI > 0.5$) but whose performance has worsened in the time interval considered ($DW < 0$). In this quadrant, we find 10 MDIs: 5 for the center and 5 for the periphery. The MDIs common to both areas are Nat and Irr, with very similar magnitude values and a decreasing trend that is greater in the center in the case of Nat and greater in the periphery in the case of Irr. The third quadrant (bottom left-hand panel) defines the worst situation because it is defined by MDIs with $MHI < 0.5$ and $DW < 0$. This is the case of Cro for the center and Car for the periphery. Finally, the fourth quadrant (top left-hand panel) includes MDIs with a magnitude below the median ($MHI < 0.5$) but with a predominant improvement of their performance ($DW > 0$). In this quadrant, we find 4 MDIs: 2 for the central area (Par and Car) and 2 for the peripheral area (Cro and Dru).

To assess a global independence (or dependence) among MDIs of the center and periphery, the distance correlation coefficient in Eq. (13) is considered. The value of the sample distance correlation is equal to 0.44, whose statistical significance is assessed using the distance covariance test [32] with 1000 bootstrap replications to obtain the

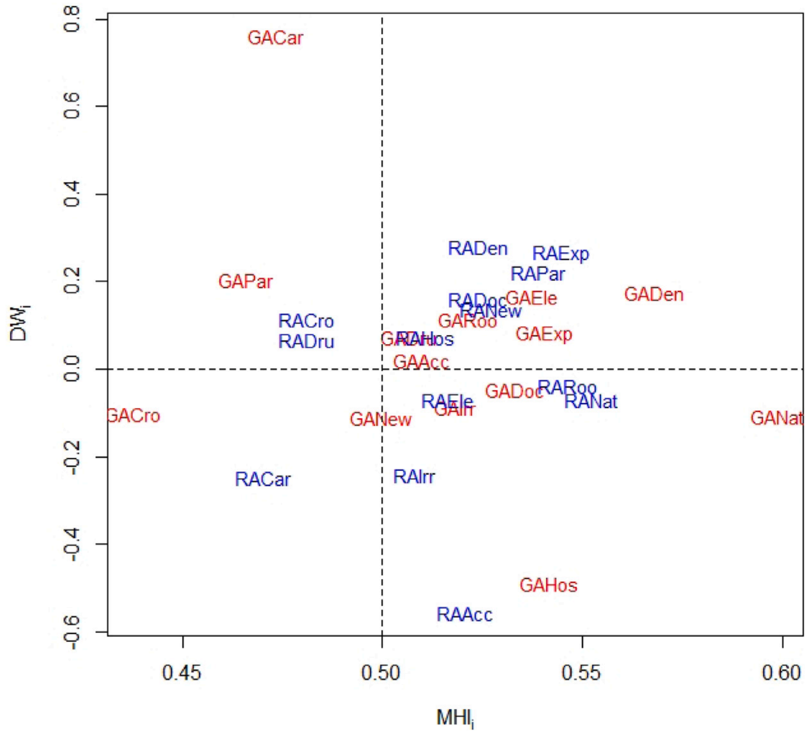

Fig. 4. $MHI - DW$ plot for the MDIs of the center (in red, GA) and periphery (in blue, RA) of Athens'area.

Table 3
MHI and DW values with the corresponding quadrant in the $MHI - DW$ plot for MDIs of Athens' area.

MDI	MHI	DW	Quad.	MDI	MHI	DW	Quad.
GADen	0.57	0.18	1	RADen	0.52	0.28	1
GANat	0.60	−0.11	2	RANat	0.55	−0.07	2
GAPar	0.47	0.21	4	RAPar	0.54	0.22	1
GANew	0.50	−0.11	2	RANew	0.53	0.14	1
GARoo	0.52	0.11	1	RARoo	0.55	−0.04	2
GACro	0.44	−0.10	3	RACro	0.48	0.11	4
GAIrr	0.52	−0.09	2	RAIrr	0.51	−0.24	2
GAEle	0.54	0.17	1	RAEle	0.52	−0.07	2
GACar	0.47	0.76	4	RACar	0.47	−0.25	3
GAAcc	0.51	0.02	1	RAAcc	0.52	−0.56	2
GAExp	0.54	0.08	1	RAExp	0.54	0.26	1
GAHos	0.54	−0.49	2	RAHos	0.51	0.07	1
GADoc	0.53	−0.05	2	RADoc	0.52	0.16	1
GADru	0.51	0.07	1	RADru	0.48	0.07	4

sampling distribution of the test statistic under the null hypothesis of independence. The permutation test accept the null hypothesis yielding a p-value equal to 0.67.

Dissimilarities between areas for each MDI are investigated using the distance measure in Eq. (14) with $r = 1$. The resulting pairwise distances are reported in Table 4 together with the contribution of the magnitude and the evolutionary component (α_γ and α_η, respectively) to these dissimilarities. The largest differences between the two areas refer to the MDIs Car, Acc and Hos, especially with reference to the evolutionary component. Very low values of the distances are recorded for the MDIs Dru, Par and Nat where, the weight of the magnitude component appears to be stronger.

The analysis is performed in the R environment [45], using the 'fda' pachage for the functional approximation of MDIs. MHI and DW are performed using the MHI function of the R package 'roahd' [46] and the DW function available in the supplementary material of Fortuna et al. [18], respectively. Finally, the permutation test is implemented in the R package 'energy' [47].

Table 4
Pairwise distances ($d^r(x_i, y_i)$) among MDIs of the center and periphery of the Athens' metropolitan region, together with magnitude (α_γ) and evolutionary (α_η) components.

MDI	$d^r(x_i, y_i)$	α_γ	α_η
Den	0.15	0.33	0.67
Nat	0.09	0.56	0.44
Par	0.08	0.88	0.12
New	0.28	0.11	0.89
Roo	0.18	0.17	0.83
Cro	0.25	0.16	0.84
Irr	0.16	0.06	0.94
Ele	0.26	0.08	0.92
Car	1.01	0.00	1.00
Acc	0.59	0.02	0.98
Exp	0.18	0.00	1.00
Hos	0.59	0.05	0.95
Doc	0.22	0.05	0.95
Dru	0.03	1.00	0.00

6. Discussion and conclusions

The paper employs FDA to explore the dynamics of MD, offering a novel analysis that accounts for the multidimensional, non-linear, and evolutionary characteristics of the phenomenon. This approach enhances our understanding of complex relationships within metropolitan regions, offering insights for a more general analysis of long-term MD in advanced economies. Specifically, the contribution provides interesting functional tools to identify multidimensional socioeconomic dynamics [44,48] and reveals the bi-directional relationship of MD paths in central and peripheral areas.

The FDA approach has proven to be an effective tool for understanding long-term relationships in MD dynamics, as highlighted through the simulation study. The inclusion of temporal variability and the use of synthetic functional measures to evaluate MDIs distinguish this approach. Indeed, the representation of each MDI as a smooth function over time provides a more realistic perspective on the long-term development of cities and surrounding regions. Moreover, functional tools such as MHI and DW, reflect the level and intrinsic dynamics of MD. These two synthetic measures, represent the ground to build a functional distance-based system, able to quantify the relationship between concentric areas of the same territory, from both a global (with the distance correlation coefficient) and dimension-specific (with pairwise distances) perspective.

Our findings allow a rethinking of MD toward complexity. The study challenges the 'traditional' sequence of urbanization-suburbanization waves, introducing counter-intuitive, reverse dynamics that reflect the indirect effects of suburbanization on the recovery of central cities post-economic decline and population shrinkage [49]. The proposed approach proved effective in disentangling the peculiar nexus between territorial dynamics of socioeconomic processes and spatial planning in metropolitan regions, highlighting the latent interplay between central and peripheral locations [5,9,50]. The results outlined how changes over time in all dimensions of MD exerted a latent contribution toward suburbanization [51]. This concurs with previously emphasized literature, confirming that the distinct temporal patterns observed between the central city and suburbs strongly imply spatial heterogeneity in economic development [50,52], suggesting a more dynamic socioeconomic context in the suburbs, particularly since the mid-1980s [53].

Furthermore, the distance correlation coefficient, computed in this study using the same MDIs for both areas, can also be derived in the less common scenario where the number of indicators is different. This extension broadens the applicability of the proposed method to instances involving missing data in specific MD dimensions.

In conclusion, the comprehensive insights derived from the FDA approach have the potential to significantly contribute to the understanding of urban studies and regional planning, providing policymakers with a novel perspective on MD dynamics.

CRediT authorship contribution statement

Francesca Fortuna: Writing – original draft, Software, Methodology, Formal analysis, Conceptualization. **Alessia Naccarato:** Writing – original draft, Methodology, Formal analysis, Conceptualization. **Luca Salvati:** Writing – original draft, Formal analysis, Data curation, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The authors do not have permission to share data.

References

[1] Klaassen L, Molle W, Paelinck J. Dynamics of urban development. New York: Sage; 1981.

[2] Duvernoy I, Zambon I, Sateriano A, Salvati L. Pictures from the other side of the fringe: Urban growth and peri-urban agriculture in a post-industrial city (Toulouse, France). J Rural Stud 2018;57:25–35.

[3] Van den Berg L, Drewett R, Klaassen L, Rossi A, Vijverberg C. A study of growth and decline. Oxford: Pergamon; 1982.

[4] Hall P, Pain K. The polycentric metropolis: Learning from mega-city regions in Europe. Earthscan; 2006.

[5] Gordon P, Cox W. Cities in western europe and the United States: do policy differences matter? Ann Reg Sci 2021;48:565–94.

[6] Delmelle E, Nilsson I, Adu P. Poverty suburbanization, job accessibility, and employment outcomes. Soc Inclusion 2021;9:166–77.

[7] Silverman R. Rethinking shrinking cities: Peripheral dual cities have arrived. J Urban Aff 2020;42(3):294–311.

[8] Salvati L, Serra P. Estimating rapidity of change in complex urban systems: A multidimensional, local-scale approach. Ann Reg Sci 2016;48:132–56.

[9] Pili S, Grigoriadis E, Carlucci M, Clemente M, Salvati L. Towards sustainable growth? A multi-criteria assessment of (changing) urban forms. Ecol Indic 2017;76:71–80.

[10] Perrin C, Nougarèdes B, Sini L, Branduini P, Salvati L. Governance changes in peri-urban farmland protection following decentralisation: A comparison between montpellier (France) and Rome (Italy). Land Use Policy 2018;70:535–46.

[11] Black D, Henderson V. A theory of urban growth. J Political Econ 1999;107:252–84.

[12] Anas A, Pines D. Anti-sprawl policies in a system of congested cities. Reg Sci Urban Econ 2008;38(5):408–23.

[13] Storper M, Scott A. Rethinking human capital, creativity and urban growth. J Econ Geogr 2009;9:147–67.

[14] Li X, Gong P. Urban growth models: progress and perspective. Sci Bull 2016;61:1637–50.

[15] Coccossis H, Economou D, Petrakos G. The ESDP relevance to a distant partner: Greece. Eur Plan Stud 2005;13:253–64.

[16] Ramsay J, Silverman B. Functional data analysis. New York: Springer-Verlag; 2005.

[17] Ferraty F, Vieu P. Nonparametric functional data analysis. New York: Springer-Verlag; 2006.

[18] Fortuna F, Naccarato A, Terzi S. Correction to: Evaluating countries' performances by means of rank trajectories: functional measures of magnitude and evolution. Comput Statist 2022. http://dx.doi.org/10.1007/s00180-022-01313-5.

[19] Fortuna F, Naccarato A, Terzi S. Country rankings according to well-being evolution: composite indicators from a functional data analysis perspective. Ann Oper Res 2022. http://dx.doi.org/10.1007/s10479-022-05072-w.

[20] Fortuna F, Maturo F, Di Battista T. Clustering functional data streams: Unsupervised classification of soccer top players based on Google trends. Qual Reliab Eng Int 2018;34:1448–60.

[21] Di Battista T, Fortuna F, Maturo F. Environmental monitoring through functional biodiversity tools. Ecol Indic 2016;60:237–47.

[22] Maturo F, Fortuna F, Di Battista T. Outliers detection in assessment tests' quality evaluation through the blended use of functional data analysis and item response theory. Ann Oper Res 2022. http://dx.doi.org/10.1007/s10479-022-05099-z.

[23] Di Battista T, Fortuna F. Functional confidence bands for lichen biodiversity profiles: A case study in Tuscany region (central Italy). Stat Anal Data Min ASA Data Sci J 2017;10(1):21–8.

[24] Belli E, Vantini S. Measure inducing classification and regression trees for functional data. Stat Anal Data Min ASA Data Sci J 2022;15(5):553–69.

[25] Banerjee B, Laha A, Lakra A. Data-driven dimension reduction in functional principal component analysis identifying the change-point in functional data. Stat Anal Data Min ASA Data Sci J 2020;13(6):529–36.

[26] Tucker J, Lewis J, Srivastava A. Elastic functional principal component regression. Stat Anal Data Min ASA Data Sci J 2019;12(2):101–15.

[27] Lopez-Pintado S, Romo J. A half-region depth for functional data. Comput Statist Data Anal 2011;55:1679–95.

[28] Arribas-Gil A, Romo J. Shape outlier detection and visualization for functional data: the outliergram. Biostatistics 2014;15:603–19.

[29] Martín-Barragan B, Lillo R, Romo J. Functional boxplots based on epigraphs and hypographs. J Appl Stat 2015;43:1088–103.

[30] Franco-Pereira A, Lillo R. Rank tests for functional data based on the epigraph, the hypograph and associated graphical representations. Adv Data Anal Classif 2020;14:651–76.

[31] Pulido B, Franco-Pereira A, Lillo R. A fast epigraph and hypograph-based approach for clustering functional data. Stat Comput 2023;33. http://dx.doi.org/10.1007/s11222-023-10213-7.

[32] Székely G, Rizzo M, Bakirov N. Measuring and testing dependence by correlation of distances. Ann Stat 2007;35:2769–94.

[33] Ramsay J, Silverman B. Applied functional data analysis: Methods and case studies. New York: Springer; 2002.

[34] Craven P, Wahba G. Smoothing noisy data with spline functions: Estimating the correct degree of smoothing by the method of generalized crossvalidation. Numer Math 1979;31:377–403.

[35] Reinsch C. Smoothing by spline functions. Numer Math 1967;10:177–83.

[36] De Boor C. A practical guide to splines. New York: Springer; 2001, revised edition.

[37] Székely G, Rizzo M. On the uniqueness of distance covariance. Stat Probab Lett 2012;82:2278–82.

[38] Székely G, Rizzo M. Brownian distance covariance. Ann Appl Stat 2009;3:1236–65.

[39] Valencia D, Lillo R, Romo J. A Kendall correlation coefficient between functional data. Adv Data Anal Classif 2019;13:1083–103.

[40] European Environment Agency. Mapping guide for a European urban Atlas. Copenhagen: EEA; 2010.

[41] Morelli V, Rontos K, Salvati L. Between suburbanisation and re-urbanisation: revisiting the urban life cycle in a Mediterranean compact city. Urban Res Pract 2014;7:74–88.

[42] Egidi G, Cividino S, Quaranta G, Alhuseen A, Salvati L. Land mismatches, urban growth and spatial planning: A contribution to metropolitan sustainability. Environ Impact Assess Rev 2020;84:106439.

[43] Rontos K, Grigoriadis E, Sateriano A, Syrmali M, Vavouras I, Salvati L. Lost in protest, found in segregation: Divided cities in the light of the 2015 "$\chi\iota$" referendum in Greece. City Culture Soc 2016;7:139–48.
[44] Zambon I, Salvati L. Metropolitan growth, urban cycles and housing in a Mediterranean country, 1910s–2010s. Cities 2019;95. 102412–.
[45] R Core Team. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2020, URL https://www.R-project.org/.
[46] Ieva F, Paganoni A, Romo J, Tarabelloni N. roahd package: Robust analysis of high dimensional data. The R J 2019;11(2):291–307. http://dx.doi.org/10.32614/RJ-2019-032.
[47] Rizzo M, Székely G. Energy: E-statistics: Multivariate inference via the energy of data. 2016, R package version 1.7-0..
[48] Makarem N. Social networks and regional economic development: The Los Angeles and Bay Area metropolitan regions, 1980–2010. Environ Plan C Gov Policy 2016;34:91–112.
[49] Salvati L, Zitti M, Sateriano A. Changes in city vertical profile as an indicator of sprawl: Evidence from a Mediterranean urban region. Habitat Int 2013;38:119–25.
[50] Shen J, Wu F. Paving the way to growth: Transit-oriented development as a financing instrument for Shanghai's post-suburbanization. Urban Geogr 2020;41(7):1010–32.
[51] Haase A, Kabisch S, Steinführer A, Bouzarovski S, Hall R, Ogden P. Spaces of reurbanisation: exploring the demographic dimension of inner-city residential change in a European setting. Popul Space Place 2010;16:443–63.
[52] Bailey M. Retail suburbanization, modernization, and growth in sydney during Australia's Postwar boom. J Urban Hist 2021. http://dx.doi.org/10.1177/00961442211023600.
[53] Salvati L, Sabbi A. Identifying urban diffusion in compact cities through a comparative multivariate procedure. Ann Reg Sci 2014;53:557–75.

Francesca Fortuna is associate Professor of Statistics at the Economic department of "Roma Tre University' of Rome. She received the Ph.D. degree in "Economics and Statistics" from "G. d'Annunzio" University of Chieti-Pescara in 2014. From 2015 to 2020, she was a post-doctoral fellow at the Department of Philosophical, Pedagogical and Quantitative Economic Sciences at "G. d'Annunzio" University of Chieti-Pescara. Her research activities are focused on functional data analysis, ecological diversity, item response theory models and sampling designs.

Alessia Naccarato (Ph.D.) is associate professor of Statistics at the Roma Tre University, Department of Economics. Her main research interests are related to econometric models, time series analysis, spatial analysis, composite indicators. On such themes she published on major international journals like Ecological Indicators, Computational Statistics, Annals of Operation Research, Journal of Official Statistics, Technological Forecasting and Social Change, Spatial Statistics.

Luca Salvati (Ph.D.) is assistant Professor of Economic Statistics at the MOMOTEF Department of the University of Rome 'La Sapienza'. Since 2001, he has been involved in research activities at various institutes of the Council for Agricultural Research and Economics Analysis (CREA), Italian National Institute of Statistics (ISTAT), the National Research Council (CNR) and the Higher Institute for Environmental Protection and Research (ISPRA). He specializes in institutional statistics, territorial statistics, and indicator theory. He has conducted studies on topics including economic statistics, urban economics, regional demography, and sustainable development, also utilizing spatial analysis, geographic information systems for decision support, and more recently developing a research line on the monetary valuation of natural resources.