Fuzzy Sets and Systems 533 (2026) 109819

Contents lists available at ScienceDirect

Fuzzy Sets and Systems

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

Measuring population decline through a composite fuzzy index: Evidence from Italian municipalities

Federico Bacchi [a,*], Laura Neri [b]

[a] Department of Statistical Sciences "Paolo Fortunati", University of Bologna, Via delle Belle Arti 41, Bologna, 40126, Italy
[b] Department of Economics and Statistics, University of Siena, Piazza San Francesco 7–8, Siena, 53100, Italy

ARTICLE INFO

Keywords:
Temporal persistence
Time-weighted aggregation
Longitudinal analysis
Census data

ABSTRACT

The implications of population decline have been widely examined in the literature, with particular attention to rural, mountain, and peripheral areas. Most research in the field relies on approaches that measure depopulation by imposing fixed thresholds or by classifying geographical areas as "declining" versus "non-declining". Such methods suffer from several shortcomings, including the arbitrariness of cut-off values, the treatment of depopulation as a dichotomous phenomenon, and the neglect of the timing and persistence of decline. To address these limitations, this study proposes the Composite Fuzzy Demographic Index (CFDI), defined as the convex combination of two components: one based on the log-differences of the demographic variable P between two subsequent time points, and the other on the persistence of the decreasing state of P. The computation of these components involves a time-weight vector that increases with time proximity and accounts for the variability of P across different time points. The proposed methodology is applied to census data on the resident population of Italian municipalities between 1951 and 2021. Results show that the CFDI produces spatial patterns consistent with established evidence on depopulation, confirming its empirical validity. At the same time, the proposed index uncovers novel insights that traditional threshold-based measures fail to capture.

1. Introduction

Italy's demographic landscape is characterized by a persistent decline in population, with the resident population standing at 58.9 million people as of June, 2025, marking a decrease of 24,789 inhabitants compared to the beginning of the year (equivalent to a decline of 0.4 per thousand). The total fertility rate is alarmingly low at 1.18 children per woman, which is one of the lowest in the world [1]. Comparing the period January – September 2024 and 2025, we observe a marginal worsening of the natural balance. Although the change is small relative to its level (−7,377 units), it indicates a high degree of persistence in the negative natural dynamic [2].

Migration patterns are evolving, with immigration from abroad decreasing slightly (−0.8% compared to January 2024), while emigration has fallen significantly to 68,000 (−39.8% compared to January 2024). In contrast, internal migration within Italy has increased by 2.4% to 719,000, indicating greater mobility among the resident population [3].

The population is also ageing rapidly, with one in five Italians over 65 years old. The urban population is around 70.4%, with major cities like Rome, Milan, and Naples experiencing significant growth [4,5], while Italy's inner areas are experiencing significant depopulation.

* Corresponding author.
E-mail address: federico.bacchi3@unibo.it (F. Bacchi).

https://doi.org/10.1016/j.fss.2026.109819
Received 6 November 2025; Received in revised form 5 February 2026; Accepted 9 February 2026
Available online 12 February 2026

These demographic trends underscore the challenges facing Italy, including low fertility rates, an ageing population, and shifting migration patterns.

The Italian case is not unique. As of January 1, 2025, Europe recorded an overall population decline of 0.1% compared to the beginning of 2024, alongside a share of citizens aged 65 and over equal to 20.9% [6]. Although the total fertility rate is less critical than Italy's, it remains low at 1.41 children per woman. At the same time, the proportion of the population living in urban areas is even higher than in Italy, reaching 73.0%.

Even though Europe was the only continent to experience a population decline between 2024 and 2025, population growth rates in the remaining continents have exhibited a downward trend in recent years. Europe is, in fact, the continent with both the lowest total fertility rate and the highest share of population aged 65 and over, with the former showing a declining trend and the latter an increasing one over time. By contrast, Oceania is the only continent with a lower share of urban population than Europe, amounting to 70.8%.

Traditional approaches to measuring depopulation typically rely on fixed thresholds or classify municipalities in binary terms - declining versus non-declining [7]. Such methods are conceptually rigid and display limited sensitivity to temporal and spatial variation. In particular, they fail to capture the gradual, uncertain, and multidimensional nature of demographic shrinkage. As a result, they are poorly suited for early detection, comparative analyses, and the design of targeted policy interventions.

To address these limitations, this paper introduces a fuzzy approach to the measurement of depopulation. We propose a fuzzy index that translates long-term population dynamics into degrees of membership to the concept of a "depopulated municipality". Unlike conventional crisp indices, the index accounts not only for the magnitude of population decline but also for its persistence and timing, thereby providing a richer and more flexible representation of demographic trajectories.

Drawing on complete census data for all Italian municipalities from 1951 to 2021, the paper addresses the following research questions:

1. What spatial and temporal patterns of population decline have characterized Italian municipalities over the past seven decades?
2. How do conventional threshold-based measures and fuzzy-based representations differ in capturing the spatial and temporal dynamics of depopulation?
3. What theoretical and interpretive advantages does a fuzzy index provide for understanding the multifaceted nature of depopulation?

The remainder of the paper is structured as follows. Section 2 reviews the relevant literature on depopulation measurement and fuzzy approaches to demographic analysis. Section 3 outlines the methodological framework and develops the fuzzy index. Section 4 presents the case of study and the data analysed. Section 5 describes the computation of the index and a series of robustness checks. Section 6 reports the empirical findings, comparing traditional and fuzzy-based measures of depopulation across Italian municipalities. Section 7 concludes with theoretical and policy implications, as well as suggestions for future research.

2. Methodological approaches to depopulation measurement

Research on depopulation has long emphasized its implications for regional development, service provision, and social cohesion. A substantial body of work has documented the economic and social consequences of population decline, particularly in rural, mountain, and peripheral areas [7–9]. These studies provide valuable contextual insights, highlighting the challenges faced by shrinking territories. Less attention has been paid to the methodological issue of how depopulation is defined and measured. Nevertheless, the existing literature can be broadly organized into two main methodological approaches: dichotomous and fuzzy.

2.1. Dichotomous approach

The dichotomous approach represents the most traditional and widely adopted strategy for identifying depopulation processes. Within this framework, empirical analyses of depopulation rely on crisp definitions of decline, typically operationalized through fixed thresholds (e.g., a minimum percentage of population loss over a given period) or binary classifications of municipalities as "declining" versus "non-declining" [7,10].

These approaches, while intuitive and straightforward, are limited by the dichotomous treatment of depopulation, which relies on arbitrary thresholds to classify areas as depopulating. This approach obscures the degree, intensity, and temporal trajectory of population decline, overlooking the gradual processes through which decline unfolds. National and supranational analysis frameworks often employ fixed cut-offs (e.g., an average annual change of −1% over a set period) that are context-dependent and ignore the continuity of decline over time [11–13]. Population decline is both an outcome and a process [14] , and analyses of decline trajectories reveal substantial heterogeneity in how places experience and sustain depopulation. Consequently, binary classifications risk overlooking transitional and borderline cases, which are critical for a nuanced assessment of territorial vulnerability [15].

As a result, the dichotomous approach is poorly suited for comparative analyses across territories, for the early detection of emerging shrinkage processes, and for guiding targeted policy interventions [16,17].

2.2. Fuzzy approaches to population decline

Demographic shrinkage is inherently a gradual, dynamic, and context-dependent phenomenon, which cannot be fully captured through rigid or dichotomous classifications. A methodological framework capable of representing degrees, intensity, and temporal dynamics is therefore required to adequately reflect the patterns of population decline across territories.

Within this perspective, fuzzy set theory provides a natural and flexible approach. By allowing units to be characterized by degrees of membership to a given concept, fuzzy methods capture situations that are neither fully "in" nor fully "out" of a category, and are thus particularly suited to phenomena marked by continuity, uncertainty, and transition. Applied to depopulation, fuzzy indices enable areas to be described in terms of degrees of decline, accounting for both the intensity and the temporal trajectory of population change, while explicitly accommodating transitional and borderline cases. Depopulation is therefore conceptualized as a continuum rather than a binary state, and population decline as a gradual, membership-based phenomenon.

Fuzzy approaches have been extensively applied in the measurement of social phenomena characterized by gradation and uncertainty, most notably in poverty and social exclusion analysis, where they allow for the representation of intermediate and overlapping states of disadvantage [18–21], following the original formulation of fuzzy sets introduced by Zadeh [22]. Despite their demonstrated effectiveness in these domains, applications of fuzzy methods to demographic decline and depopulation remain limited and comparatively underexplored, leaving substantial room for methodological development.

Some recent contributions have begun to explore fuzzy methodologies in the analysis of demographic phenomena, for instance, applying fuzzy transform methods to spatio-temporal demographic balance data to improve forecast accuracy and interpretability [23], and employing fuzzy clustering techniques for the analysis of longitudinal population sequences without imposing crisp groupings [24].

Addressing this gap, the paper advances a fuzzy-set approach to depopulation measurement, providing a flexible framework capable of considering the persistence and representing intensity, and borderline cases in population decline analysis.

3. A composite fuzzy demographic index

The main goal of the present work was to define a synthetic index that represents the propensity of each statistical unit to be subject to depopulation. This could be mathematically translated into the definition of a membership function in the fuzzy set of Italian municipalities subject to depopulation.

Consider a population of $N \in \mathbb{N}$ geographical units observed in $(T+1)$ fixed time points. Let $t = 0, \dots, T$, with $t \in \mathbb{N}$, denote the time points (where $t = 0$ is the furthest and $t = T$ the closest to the present), $i = 1, \dots, N$ the geographical units, and $\boldsymbol{p}_i = (p_{i,0}, \dots, p_{i,T})$ the vector of demographic information P for the unit i and assume $p_{i,t} > 0$, for all i and for all t.

The most intuitive way to infer the propensity to be subject to depopulation is to compare, for each i, the demographic information at each time point $t = 1, \dots, T$ with its previous $(t-1)$. More specifically, this comparison might be done in many possible ways. However, the interest of the present work was on two main aspects: (i) the magnitude of the relative changes of P for each unit i in each time interval $(t-1; t)$; (ii) the persistence of negative changes of P for each unit i.

Referring to the former, this comparison might be conceived as a simple chained index, i.e. $p_{i,t}/p_{i,t-1}$. However, it seems more appropriate to consider the log-difference of the two elements:

$$l_{i,t} = \log(p_{i,t}) - \log(p_{i,t-1}) = \log\left(\frac{p_{i,t}}{p_{i,t-1}}\right), \quad i = 1, \dots, N, \quad t = 1, \dots, T.$$

The rationale of this choice could be found in the desirable properties characterizing log-differences [25]. So, a Fuzzy Depopulation Index (FDI) might be defined as a function:

$$FDI_i(\boldsymbol{p}_i) = 1 - \text{norm}\left\{\sum_{t=1}^{T} \delta_t l_{i,t}\right\} = 1 - \text{norm}\{I_T(\boldsymbol{p}_i)\} \in [0; 1], \tag{1}$$

where $\delta_t \in (0; 1)$ is the inter-temporal weight associated to time $t = \{1, \dots, T\}$ subject to the condition $\sum_{t=1}^{T} \delta_t = 1$, and $\text{norm}(X)$ is the classical normalization function, for a generic indicator X, in the unit interval defined as:

$$\text{norm}(X) = \frac{X - \min(X)}{\max(X) - \min(X)}.$$

In the case of FDI, the function that is normalized is $I_T(\boldsymbol{p}_i) = \sum_{t=1}^{T} \delta_t l_{i,t}$ introduced in (1). It is worth noting that FDI is defined in (1) as a one's complement. This choice was due to an interpretation point of view: log-difference is higher when the demographic information P shows an increase, but the purpose of the work is to obtain a synthetic index representing the opposite information.

On the other hand, a possible way to account for the persistence of the decreasing state of demographic variable P is to define a sequence of discrete variables for each $i = 1, \dots, N$ and for each $t = 1, \dots, T$ as follows:

$$d_{i,t}(\tau) = \begin{cases} +1, & l_{i,t} > \tau; \\ 0, & -\tau \le l_{i,t} \le \tau; \\ -1, & l_{i,t} < -\tau; \end{cases}$$

where $\tau > 0$ is a threshold value useful to distinguish between a steady state and a dynamic evolution of P. Then, a fuzzy persistence index of depopulation might be defined as:

$$PERS_i(\boldsymbol{p}_i) = 1 - \text{norm}\left\{\sum_{t=1}^{T} \delta_t d_{i,t}(\tau)\right\} = 1 - \text{norm}\{J_T(\boldsymbol{p}_i, \tau)\} \in [0; 1], \tag{2}$$

where $J_T(\boldsymbol{p}_i, \tau) = \sum_{t=1}^{T} \delta_t d_{i,t}(\tau)$ represents the actual function of the vector of demographic information for statistical unit i that is normalized.

The two components might be combined via a weighted mean to obtain a Composite Fuzzy Demographic Index (CFDI):

$$CFDI_i(\lambda, \boldsymbol{p}_i) = \lambda FDI_i(\boldsymbol{p}_i) + (1 - \lambda) PERS_i(\boldsymbol{p}_i), \tag{3}$$

with $\lambda \in [0; 1]$ representing a constant weight for all statistical units. Being defined as a convex combination of two indices varying in the unit interval, it is evident that also $CFDI$ is bounded by the same values, i.e. $CFDI_i(\lambda, \boldsymbol{p}_i) \in [0; 1]$.

3.1. Properties

From an analytical point of view, it is important to verify that the proposed composite index satisfies a set of desirable properties. More specifically, a major interest was on the following:

1. Scale-invariance;
2. Increase-decrease Monotonicity;
3. Proximity Monotonicity.
4. Distributional Stationarity;
5. Robustness to Local Noise.

The proofs are provided in Appendix A.

The first property, scale-invariance, is related to the fact that the proposed index does not change in the presence of different scales of measurement.

Property 1 (Scale-invariance). *Let $\lambda = \lambda^*$ be a fixed value set a priori, $\min_t(p_t) = \min_i(p_{i,t})$, and $\max_t(p_t) = \max_i(p_{i,t})$, be fixed for each $t = 0, \dots, T$. For all $T \in \mathbb{N}$, for all possible levels $\boldsymbol{p}_i^* = (p_{i,0}^*, \dots, p_{i,T}^*) \in \mathbb{R}^{(T+1)}$ of the demographic variable P and for all $b \in \mathbb{R}$:*

$$CFDI_i(\lambda^*, b\boldsymbol{p}_i^*) = CFDI_i(\lambda^*, \boldsymbol{p}_i^*).$$

The second property is related to the fluctuations of the demographic variable P between two time points: a positive variation of P produces a decrease in the propensity of being a member of the fuzzy set D, whereas a negative variation produces an increase in the fuzzy membership function. To simplify the notation, the number of time points is supposed to be fixed $(T + 1) = 2$.

Property 2 (Increase-decrease Monotonicity). *Let $\lambda = \lambda^*$ be a fixed value set a priori, $\min_t(p_t) = \min_i(p_{i,t})$, and $\max_t(p_t) = \max_i(p_{i,t})$, be fixed for $t = 0, 1$, $\Delta p > 0$ denote the absolute variation of P between two time points, and p_i^* the level of P for statistical unit i at time $t = 0$. For all $p_i^*, \Delta p > 0$ $(\Delta p < p_i^*)$:*

$$CFDI_i(p_i^*, p_i^* + \Delta p) < CFDI_i(p_i^*, p_i^*) < CFDI_i(p_i^*, p_i^* - \Delta p).$$

The third property, like the previous one, is related to the fluctuations of the demographic variable P. Generally speaking, it would be desirable for the proposed composite index to take into account the fact that variations that occur further in the past should have a lower impact on the propensity to be subject to depopulation. This might be achieved by requiring the time weights δ_t to increase with time proximity, i.e. $\delta_1 < \delta_2 < \dots < \delta_T$.

Property 3 (Proximity Monotonicity). *Let $\lambda = \lambda^*$ be a fixed value set a priori, $\min_t(p_t) = \min_i(p_{i,t})$, and $\max_t(p_t) = \max_i(p_{i,t})$, be fixed for each $t = 0, \dots, T$, $\Delta p > 0$ denote the absolute variation of P between two time points, and $p_{i,t}^*$ a fixed level of P for statistical unit i at time t. In addition, it is assumed that for each i, the sequence $\boldsymbol{p}_i^*$ is constant in all time intervals, except the one where the variation Δp occurs. If the vector of time weights satisfies the condition of being increasing with time proximity $(\delta_1 < \delta_2 < \dots < \delta_T)$, then, for all $T \in \mathbb{N}$, for all $p_i^*, \Delta p > 0$ $(\Delta p < p_i^*)$:*

$$\begin{aligned} &CFDI_i(p_{i,0}^*, \dots, p_{i,T-1}^*, p_{i,T-1}^* + \Delta p) < \\ &CFDI_i(p_{i,0}^*, \dots, p_{i,T-2}^*, p_{i,T-2}^* + \Delta p, p_{i,T-2}^* + \Delta p) < \\ &\qquad \dots < \\ &CFDI_i(p_{i,0}^*, p_{i,0}^* + \Delta p, \dots, p_{i,0}^* + \Delta p). \end{aligned}$$

The fourth property concerns its statistical stability with respect to the length of the observed time window $(T + 1)$. In particular, the interest is in verifying whether the distribution of $CFDI_i(\lambda, \boldsymbol{p}_i)$ remains invariant (in a statistical sense) when the number of time periods T increases. This type of behavior can be interpreted as a form of distributional stationarity with respect to T.

Property 4 (Distributional Stationarity). *Let $CFDI_i(\lambda, \boldsymbol{p}_i)$ be the Composite Fuzzy Demographic Index defined in Eq. (3), where the demographic information $p_{i,t}$ follows a stochastic process of the form:*

$$p_{i,t} = p_{i,t-1} \exp\left(\varepsilon_{i,t}\right), \quad \varepsilon_{i,t} \sim \mathcal{D}(0, \sigma_\varepsilon^2),$$

with $\mathcal{D}$ being a distribution with zero mean and finite variance. Then, the distribution of $CFDI_i(\lambda, \boldsymbol{p}_i)$ does not change systematically as the time horizon T increases.

Table 1
Main summary statistics of the municipality resident population at each census time (31 December).

Year	Min.	Q_1	Median	Q_3	Max.	Mean
1951	74	1,387	2,648	5,234	1,632,402	6,013
1961	90	1,294	2,500	5,038	2,163,555	6,406
1971	51	1,152	2,348	5,034	2,750,370	6,850
1981	32	1,105	2,370	5,372	2,805,109	7,156
1991	31	1,077	2,414	5,627	2,733,908	7,184
2001	33	1,074	2,439	5,853	2,546,804	7,212
2011	30	1,068	2,531	6,336	2,617,175	7,520
2021	31	987	2,404	6,234	2,749,031	7,469

The last property is related to the occurrence of small random shocks or anomalous fluctuations that should not undermine the evaluation of the propensity for depopulation. This property is particularly useful in applied contexts where demographic data may suffer from revisions, rounding, or sporadic measurement inaccuracies (e.g., due to census interpolation or administrative gaps). The robustness of $CFDI$ ensures that such irregularities do not bias the resulting spatial comparison.

Property 5 (Robustness to Local Noise)**.** *Let $CFDI_i(\lambda, \boldsymbol{p}_i)$ be the Composite Fuzzy Demographic Index defined in Eq. (3). Suppose that the time series of the demographic variable P for each unit i is affected by a localized random perturbation $\eta_{i,t}$ in one or more time points t, such that:*

$$p_{i,t} = p_{i,t-1} \exp\left(\eta_{i,t}\right), \quad \eta_{i,t} \sim \mathcal{D}(0, \sigma_\eta^2),$$

where $\eta_{i,t}$ is a zero-mean noise term independent of the underlying trend. Then, under mild regularity conditions on the weights δ_t and the normalization operator, the variation of $CFDI_i$ induced by such perturbations is bounded and asymptotically negligible for sufficiently large T.

4. Data

Census data about population have been downloaded from Italian National Statistical Institute (ISTAT) website. The target variable P to evaluate the phenomenon of depopulation is the resident population at municipality level; such information was historically collected by the Population Census. The first one was in 1951 and, until 2011, it was conducted every ten years. Since 2018, ISTAT has started the so called "permanent census" which allows to collect data yearly. Thus, at present, census data on the resident population at municipality level are available for the years 1951, 1961, 1971, 1981, 1991, 2001, 2011, and for every year starting from 2018 on. However, among the information provided by the permanent census, only data referring to year 2021 were considered in order to work on homogeneous ten-year time intervals.

The statistical units of analysis were $N = 7{,}903$ Italian municipalities existing at a fixed time (01/01/2021), whose borders were kept stable throughout the analysis time horizon. This condition was necessary to make accurate comparisons. Geographical data on municipalities (shapefiles) are provided by ISTAT [26]. The main summary statistics of the distribution of the resident population at municipality level, for each time point, are presented in Table 1.

Empirical and policy research shows that Italian "Aree Interne" are often characterized by high levels of demographic decline and marginalization, linked to structural isolation and limited access to essential services. This has been formally addressed in national strategies such as the SNAI, which aim to enhance service accessibility and promote population sustainability. In addition, natural hazards, such as seismic risk, interact with demographic vulnerabilities, further exacerbating the fragility of rural communities. Research also highlights the role of the digital divide as an additional factor influencing local population retention and economic opportunities.

To explore these relationships and to support the interpretation and validation of the proposed Composite Fuzzy Demographic Index, several contextual variables were considered, such as information on the demographic composition and geographical characteristics of the municipality, municipal classification according to the classification of the Agenzia per la Coesione Territoriale, quality and types of Internet connections, and seismic risk indicators. Although these variables are not used in the construction of the CFDI itself, they provide crucial information to characterize municipalities and contextualize the dynamics of the observed population. These contextual variables were collected from different data sources.

First, the Agency for Territorial Cohesion (Italian: Agenzia per la Coesione Territoriale) provided useful information about the $N = 7,903$ Italian municipalities linked to the classification in "Aree Interne" and "Centri", such as the sub-categories of the two groups and the average journey time from the closest "Centro" [27].

Following the field literature [28], information on the demographic composition of each municipality at 31/12/2021 was downloaded from ISTAT website [29]. Moreover, from the same source, information on some geographical characteristics of the municipality was gathered, namely on the altitude, on the altimetric zone, and on the degree of urbanization.

An original data source was identified in the Authority for Communications Guarantees (AGCOM; Italian: Autorità per le Garanzie nelle Comunicazioni) related to the quality of Internet connections have been downloaded to evaluate the goodness of the infrastructures. These data [30]) were collected in December 2018 and referred to the 7,998 existing municipalities on 31st December 2016.

Table 2
Values of sequences of variables computed for few Italian municipalities.

Code		1961	1971	1981	1991	2001	2011	2021
001001	$l_{i,t}$	−0.049	0.128	−0.085	0.017	−0.019	0.027	−0.032
	$d_{i,t}$	0	1	−1	0	0	0	0
001002	$l_{i,t}$	−0.031	0.385	0.705	0.068	0.089	0.072	−0.043
	$d_{i,t}$	0	1	1	1	1	1	0
…	…	…	…	…	…	…	…	…
111106	$l_{i,t}$	0.143	0.149	0.101	0.031	−0.032	−0.030	−0.038
	$d_{i,t}$	1	1	1	0	0	0	0
111107	$l_{i,t}$	0.212	0.079	0.190	0.059	0.105	0.212	0.052
	$d_{i,t}$	1	1	1	1	1	1	1

Since the research interest was on the number of different types of Internet connection addresses, it has been easy to obtain the values of the same variables for the 7,903 municipalities existing on 1st January 2021. Indeed, the former municipalities were merged into a contiguous one or united in a new one. So, it has been only necessary to make a bunch of additions.

Finally, the analysis was enriched by considering a feature specific to Italy: the seismic risk. It was possible to obtain the information about this variable at municipality level thanks to the database of Civil Protection Department [31] (Italian: Dipartimento della Protezione Civile).

5. Results

The Composite Fuzzy Demographic Index ($CFDI$) described in Section 3 was computed on the data presented in Table 1. The values of the fixed parameters λ and τ were chosen a priori and, then, compared with some alternative values in order to check the robustness of the final index (Section 5.2), whereas the time weights δ_t's were chosen after having carefully evaluated a set of alternatives.

The computation procedure of $CFDI_i(\lambda, \boldsymbol{p}_i)$ is described in the next subsection.

5.1. Computation of CFDI

The first step of the procedure was the computation of the sequence of variables $l_{i,t}$ and $d_{i,t}$ defined in Section 3, for all $i = 1, \dots, 7,903$ and for all $t = 1961, 1971, 1981, 1991, 2001, 2011, 2021$. For the computation of the latter, τ was set to 0.05. Table 2 presents the values computed for the top two and for the bottom two statistical units of the data set.

Then, the two components of the composite index, i.e. FDI and $PERS$, could be computed according to Eqs. (1) and (2). With the aim of simplifying the notation, let $\boldsymbol{t}^* = \{1961, 1971, 1981, 1991, 2001, 2011, 2021\}$ denote the vector of time points. First, it is necessary to define the time weights, subject to the condition $\sum_{t \in \boldsymbol{t}^*} \delta_t = 1$. A total of six different vectors of weights were compared.

First, the condition of equal weights was explored, i.e. $\boldsymbol{\delta}_{eq} = \{0.143; 0.143; 0.143; 0.143; 0.143; 0.143; 0.143\}$. However, as explained in the previous sections, the final index should account for two important features: (i) the time proximity, and (ii) the variability observed in the same time interval, i.e. column-variability. For this reason, a set of alternative vectors of weights were defined.

With the aim of accounting for time proximity, a vector of linearly increasing weights was obtained by solving the following equation:

$$\delta + 2\delta + 3\delta + 4\delta + 5\delta + 6\delta + 7\delta = 1,$$

whose solution is $\delta = 1/28$, leading to the vector:

$$\begin{aligned}\boldsymbol{\delta}_{lin} &= \{\delta; 2\delta; 3\delta; 4\delta; 5\delta; 6\delta; 7\delta\} = \\ &= \{0.036; 0.071; 0.107; 0.143; 0.179; 0.214; 0.250\}.\end{aligned}$$

A more penalizing alternative for further time intervals was explored by defining a vector of exponentially increasing weights, i.e. solving the equation:

$$\delta^7 + \delta^6 + \delta^5 + \delta^4 + \delta^3 + \delta^2 + \delta = 1,$$

whose solution is $\delta \approx 0.502$, leading to the vector:

$$\begin{aligned}\boldsymbol{\delta}_{exp} &= \left\{\delta^7; \delta^6; \delta^5; \delta^4; \delta^3; \delta^2; \delta\right\} = \\ &= \{0.008; 0.016; 0.032; 0.064; 0.127; 0.252; 0.502\}.\end{aligned}$$

On the other hand, in order to account for the column-variability, a vector of weights proportional to the standard deviation of $l_{i,t}$ for each $t \in \boldsymbol{t}^*$ was defined:

$$\boldsymbol{\delta}_{col} = \{0.176; 0.217; 0.156; 0.125; 0.115; 0.124; 0.087\}.$$

Table 3
Spearman correlation coefficients between the six $CFDI$'s obtained by using different weighting schemes.

	$\boldsymbol{\delta}_{eq}$	$\boldsymbol{\delta}_{lin}$	$\boldsymbol{\delta}_{\exp}$	$\boldsymbol{\delta}_{col}$	$\boldsymbol{\delta}_{lincol}$	$\boldsymbol{\delta}_{expcol}$
$\boldsymbol{\delta}_{eq}$	1					
$\boldsymbol{\delta}_{lin}$	0.963	1				
$\boldsymbol{\delta}_{\exp}$	0.887	0.969	1			
$\boldsymbol{\delta}_{col}$	0.992	0.924	0.831	1		
$\boldsymbol{\delta}_{lincol}$	0.985	0.994	0.941	0.958	1	
$\boldsymbol{\delta}_{expcol}$	0.906	0.981	0.997	0.854	0.957	1

Lastly, the previously defined vectors of weights were integrated to obtain two additional vectors accounting simultaneously for time proximity and column variability:

$$\delta_{lincol,t} = \frac{\delta_{lin,t} * \delta_{col,t}}{\sum_{t \in t^*} \delta_{lin,t} * \delta_{col,t}};$$

$$\delta_{expcol,t} = \frac{\delta_{exp,t} * \delta_{col,t}}{\sum_{t \in t^*} \delta_{exp,t} * \delta_{col,t}};$$

for all $t \in t^*$. These definitions led to the following vectors:

$$\boldsymbol{\delta}_{lincol} = \{0.050; 0.124; 0.133; 0.143; 0.164; 0.212; 0.173\};$$

$$\boldsymbol{\delta}_{expcol} = \{0.013; 0.032; 0.046; 0.074; 0.136; 0.292; 0.407\}.$$

Then, the computational procedure for FDI and $PERS$ was performed for all the presented weighting structures, with the help of `FuzzyPovertyR` R package [32]. Finally, the two components were combined according to (3), setting $\lambda = 0.5$. As a result, six different distributions of the final index were obtained. These distributions were compared using Spearman correlation coefficients presented in Table 3. All pairs of $CFDI$'s were characterized by a very high Spearman correlation coefficient: the lowest (0.831) is shown by the comparison between the exponential weighting scheme and the one accounting for column-variability.

Although the comparison revealed the robustness of the procedure with respect to the use of different weighting schemes, it was necessary to choose one in order to proceed in the statistical analysis. As stated at the beginning of the current Section, there were two main requirements that the elements of the chosen vector should respect: being increasing with time proximity, and accounting for column variability. The former was achieved by three vectors: the vector of linearly increasing weights ($\boldsymbol{\delta}_{lin}$), the vector exponentially increasing weights ($\boldsymbol{\delta}_{\exp}$), and the vector which combined the exponentially increasing weights with the ones accounting for column variability ($\boldsymbol{\delta}_{expcol}$). On the other hand, column variability was accounted only by the three vectors where it was explicitly considered, i.e. $\boldsymbol{\delta}_{col}$, $\boldsymbol{\delta}_{lincol}$, and $\boldsymbol{\delta}_{expcol}$. Being the latter the only vector of weights satisfying both requirements, it was the final choice.

The geographical distribution of $CFDI$ obtained by using $\boldsymbol{\delta}_{expcol}$ is presented in Fig. 1. At first, it seemed that the presence of Italian municipalities subject to depopulation, characterized by $CFDI$ close to 1.0, was quite high in the Appennine Mountains and in insular regions. The same consideration does not seem to hold for the municipalities located in the Alps. On the other hand, a large proportion of the municipalities characterized by low propensity to depopulation, i.e. $CFDI$ close to 0, seemed to be located in the Po Valley, in some areas of Tuscany, in Lazio (especially close to Rome), and in Apulia. These considerations will be further explored in the analysis presented in Section 6. An extract of the final ranking of Italian municipalities based on $CFDI$, is provided in Table 4. Notice that the number of ranked statistical units is $7,902$, because the municipality of Vajont (PN) was separated from Maniago (PN) in 1971 and it was not possible to retrieve prior data. The whole ranking is available upon request.

5.2. Robustness checks

The procedure of definition of $CFDI$ involved some arbitrary choices. As explained in detail in the previous Section, the weighting procedure itself was a possible source of arbitrariness and, as a consequence, of non-robustness of the index. However, it was shown that the use of different vectors of time weights, defined following different criteria, did not alarmingly affect the distribution of the index.

Nevertheless, at least two additional choices deserve particular attention: (i) the choice of the threshold value τ in the computation of the sequence of discrete variables $d_{i,t}$; (ii) the choice of the constant weight λ in the convex combination (3).

Starting from the former, the robustness $CFDI$ was investigated by considering three alternative values of τ, namely $\boldsymbol{\tau}^+ = \{0.01; 0.025; 0.10\}$. These different values affected only the index accounting for the persistence of the decreasing state. So, it was necessary to compute $PERS_i(\boldsymbol{p}_i, \tau^+)$ for all $i = 1, \dots, 7,903$ and for all τ^+'s. Then, the convex combination was calculated with the original constant weight $\lambda = 0.5$.

The density plots of the three different indices plus the original one are presented in Fig. 2. It is clear that, as τ increases, the densities become more concentrated around the mean value. This is not surprising, since higher values of τ bring as a natural consequence higher frequencies of 0's in the distributions of the sequence of discrete variables $d_{i,t}$.

Despite this expected element of difference, there is extremely high agreement between the four rankings, as testified by the Spearman correlation coefficients provided in Table 5: even between the two most distant threshold values, i.e. 0.01 and 0.10, the coefficient is very high (0.929).

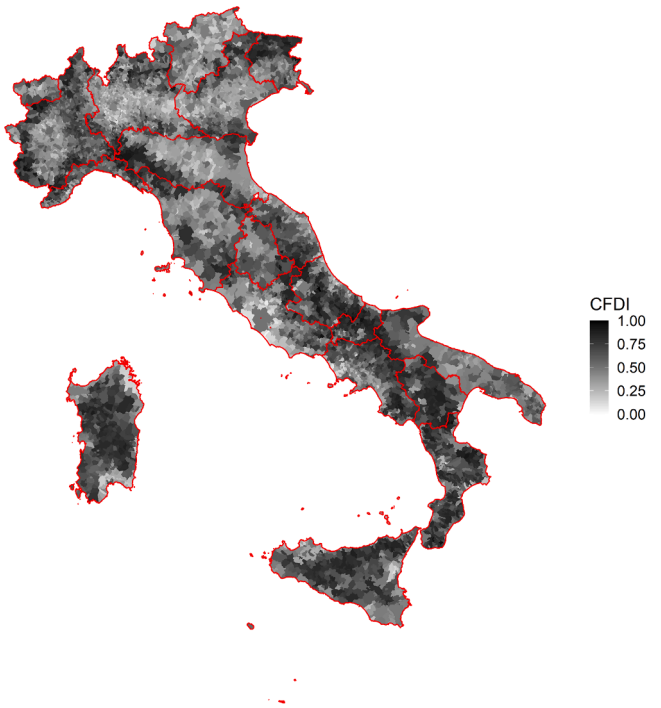

Fig. 1. Cartograph of $CFDI$ obtained by using the exponential column weights δ_{expcol}.

Table 4
Extract of the final ranking of Italian municipalities.

Rank	Code	Name	CFDI	Prov.	Region
1	066104	Villa S. Lucia degli Abruzzi	1.000	AQ	Abruzzo
2	069080	San Giovanni Lipioni	0.968	CH	Abruzzo
3	030034	Drenchia	0.966	UD	Friuli-Venezia G.
4	080090	Staiti	0.965	RC	Calabria
5	070063	San Biase	0.963	CB	Molise
6	069088	Schiavi di Abruzzo	0.958	CH	Abruzzo
7	001165	Noasca	0.955	TO	Piemonte
8	004053	Castelmagno	0.953	CN	Piemonte
9	057036	Marcetelli	0.952	RI	Lazio
10	078032	Castroregio	0.952	CS	Calabria
…	…	…	…	…	…
7,893	061027	Castel Volturno	0.082	CE	Campania
7,894	061053	Orta di Atella	0.080	CE	Campania
7,895	087012	Camporotondo Etneo	0.076	CT	Sicilia
7,896	108055	Roncello	0.070	MB	Lombardia
7,897	056038	Monterosi	0.066	VT	Lazio
7,898	018060	Cura Carpignano	0.059	PV	Lombardia
7,899	058036	Fiano Romano	0.055	RM	Lazio
7,900	018129	Roncaro	0.054	PV	Lombardia
7,901	018043	Ceranova	0.053	PV	Lombardia
7,902	058117	Ardea	0.036	RM	Lazio

Table 5
Spearman correlation coefficients between the four $CFDI$'s obtained by using different threshold values τ.

		τ			
		0.01	0.025	0.05	0.10
τ	**0.01**	1			
	0.025	0.980	1		
	0.05	0.953	0.967	1	
	0.10	0.929	0.936	0.954	1

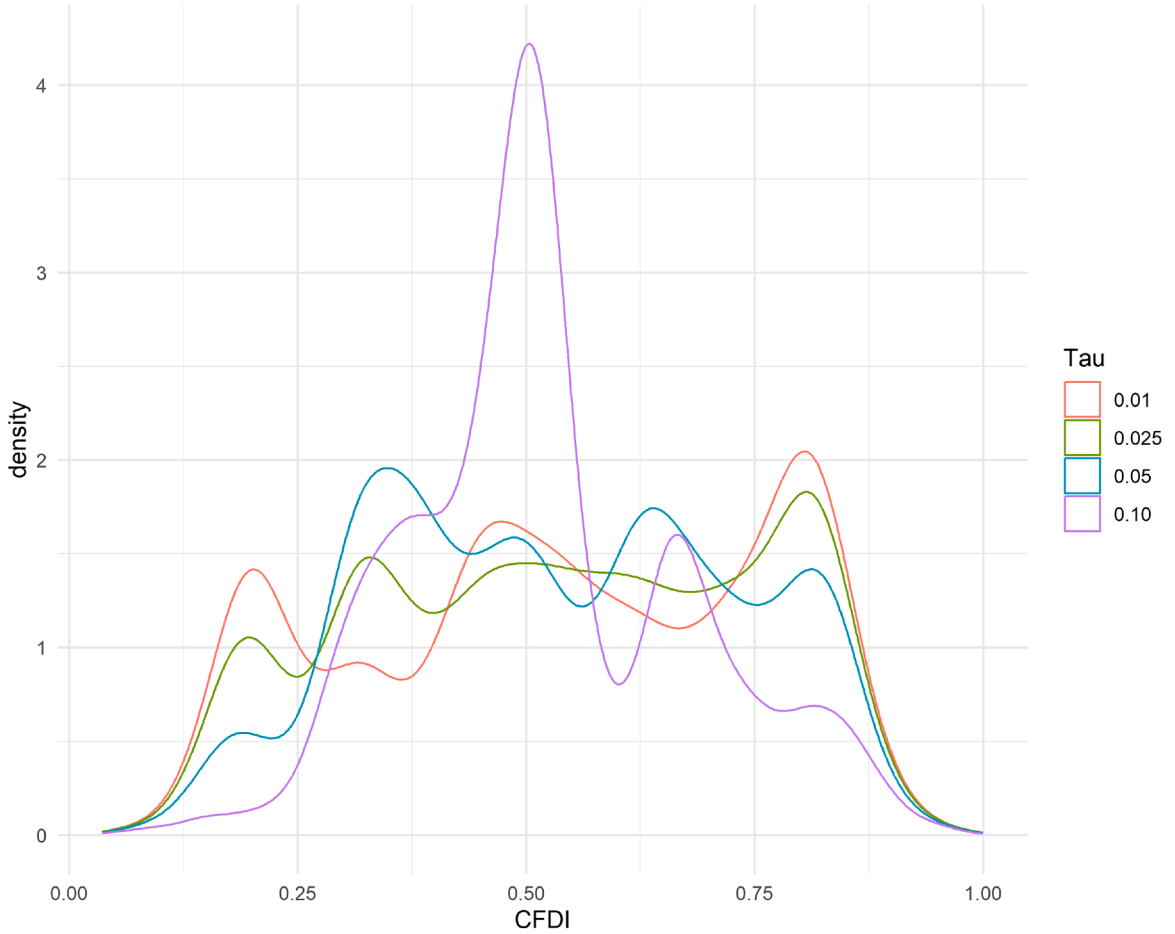

Fig. 2. Density plots of $CFDI$, each computed with a different threshold τ.

The second element of arbitrariness was the choice of the constant weight in the convex combination of FDI and $PERS$. The robustness of $CFDI$ with respect to this possible source of bias was evaluated by considering the following different values of λ:

$$\lambda = \{0.0; 0.1; 0.2; 0.3; 0.4; 0.5; 0.6; 0.7; 0.8; 0.9; 1.0\},$$

where $\lambda = 0.0$ is the limit case where $CFDI \equiv PERS$, $\lambda = 1.0$ is the opposite situation where $CFDI \equiv FDI$, and $\lambda = 0.5$ is the original $CFDI$ computed in Section 5.1. In Fig. 3, the histograms of the 11 different indices are presented.

The presence of different shapes of the density plots, ranging from multimodal to unimodal distributions, reflects the behaviour of the underlying components. On the one hand, $PERS$ (the limiting case where $\lambda = 0$) relies on the weighted sum of a sequence of discrete variables $d_{i,t}(\tau = 0.05)$; this sum tends to take, with relatively high frequency, a wide range of values in the interval $-T, \dots, 0, \dots, T$. On the other hand, FDI ($\lambda = 1.0$) is based on the weighted sum of a sequence of continuous variables $l_{i,t}$; these variables are expected to be symmetrically distributed around a mean value, due to the properties of log-differences. Despite these differences, the agreement between the resulting rankings is extremely high. The lowest Spearman correlation coefficient is observed between $PERS$ and FDI, yet it remains very high (0.922).

After having verified the robustness of $CFDI$ with respect to two arbitrary choices, it was of primary interest to investigate how much the final index was affected by the length of the time span. More specifically, since $CFDI$ was conceived also to take into account time proximity, this check was undertaken by removing the furthest time points. The procedure started by removing the intervals: (i) 1951–1961; (ii) 1951–1971; (iii) 1951–1981. In doing so, even keeping fixed the scheme which integrated exponentially increasing weights and the ones accounting for column variability, it has been necessary to recompute the time weights, leading to the following vectors:

$$\boldsymbol{\delta}_{1961-2021} = \{0.033; 0.048; 0.076; 0.138; 0.295; 0.410\}$$
$$\boldsymbol{\delta}_{1971-2021} = \{0.051; 0.080; 0.144; 0.305; 0.420\}$$
$$\boldsymbol{\delta}_{1981-2021} = \{0.087; 0.155; 0.322; 0.435\}.$$

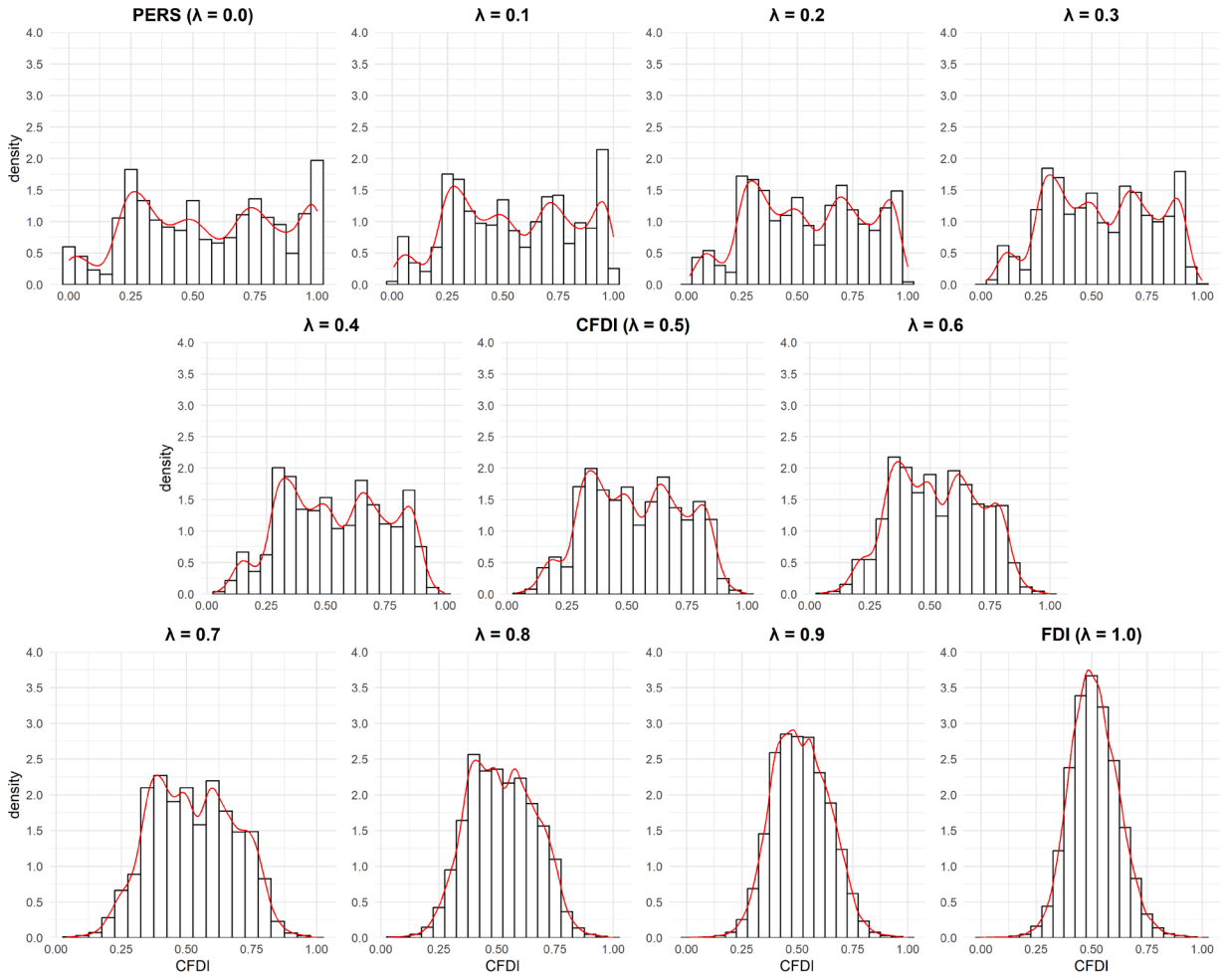

Fig. 3. Histograms of $CFDI$ in correspondence of different values of the weighting constant λ.

Then, in each time horizon, $CFDI$ was computed in the original settings, i.e. with $\tau = 0.05$ and $\lambda = 0.5$. Fig. 4 shows that the density plots of the indices obtained overlap almost perfectly, independently of the time span.

6. Empirical findings

The previous Section led to the computation of $CFDI$ for $7,902$ Italian municipalities, existing on 01/01/2021. Furthermore, the implementation of proper robustness checks revealed that $CFDI$ is not alarmingly affected by the choice of the vector of time weights $\boldsymbol{\delta}_{t^*}$, of the threshold τ differentiating between steady and changing state, and of the weighting constant λ in (3).

After having checked the robustness of $CFDI$, the research interest was shifted to characteristics that could be associated with different propensities to be subject to depopulation. So, an extended set of covariates was considered.

First, particular attention was paid to administrative covariates, such as belonging to higher levels of spatial aggregation of the statistical units. Referring to the European Union's (EU) Nomenclature of Territorial Units for Statistics (NUTS), the interest was on the first (NUTS 1) and on the second (NUTS 2) level of the hierarchy. For Italy, EU recognizes the existence of five major socio-economic regions – NUTS 1: North-West, North-East, Center, South, and Insular – and twenty basic regions – NUTS 2: the twenty Italian Regions.

In Fig. 5 the box and whiskers plot of $CFDI$, conditioned to the NUTS 2 region is presented. The box-plots referring to each NUTS 2 are filled with different colours, on the basis of the NUTS 1 of reference. First, it should be noted that the conditioned distributions of $CFDI$ of municipalities belonging to Southern, with the exception of Puglia, and Insular regions were characterized by larger median values than those belonging to the remaining NUTS 1. Furthermore, municipalities located in regions like Molise and Basilicata, showed a narrow interquartile range, suggesting that depopulation is a widespread problem in these areas. Molisan municipalities also recorded the highest median value of $CFDI$ (0.787), whereas the ones belonging to a North-Eastern region, i.e.

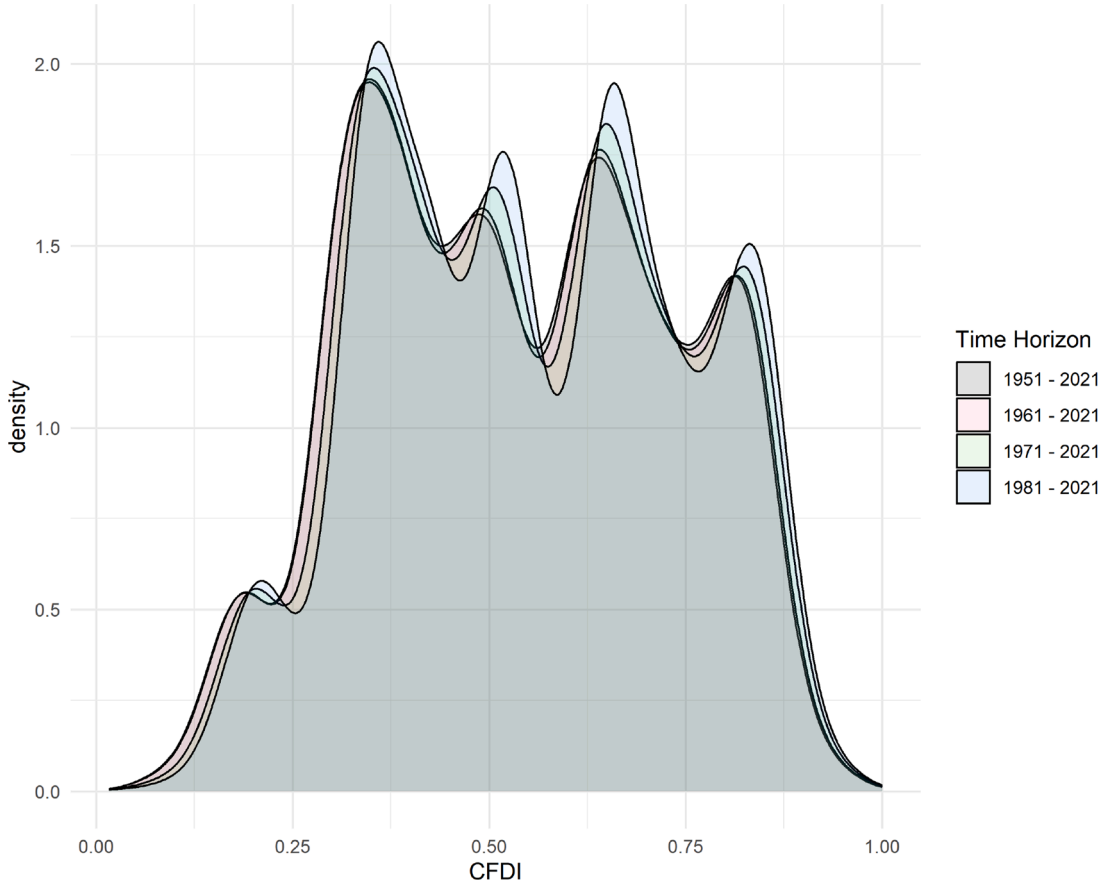

Fig. 4. Density plots of $CFDI$ in correspondence of different time horizons.

Trentino-Alto Adige, recorded the lowest (0.367). The main summary statistics of $CFDI$, conditioned to the different categories of all the categorical covariates, are presented in Table B.7.

As reported in Section 2, the difference between rural and urban areas has been extensively explored in the context of the study of intra-national population dynamics. For this reason, it seemed reasonable to investigate whether the proposed fuzzy index provided consistent results with previous studies. Data downloaded from ISTAT website provide granular information on the degree of urbanization of each municipality, coded according to EU's classification in three mutually exclusive classes: in order from the highest to the lowest level, urban areas, towns and suburbs, and rural areas.

As shown in the conditioned box-plot in the top-left panel of Fig. 6, $CFDI$ seemed to be consistent with previous findings. Indeed, the median level of $CFDI$ within the rural municipalities is significantly higher (0.636) than within the more urbanized statistical units, i.e. 0.396 for urban and 0.377 for towns and suburbs.

Slightly different but, in some sense, related to the urban-rural dichotomy is the classification defined by the Agency for Territorial Cohesion. Based on the level of accessibility to three basic services – hospitals, secondary schools, and railway stations – Italian municipalities are divided into two macro-categories: Aree Interne and Centri. Municipalities lacking at least one of these services belong to the former, while all others belong to the latter.

Furthermore, a more detailed classification is available, which distinguishes six levels according to the distance from the nearest Centro: Poles (A), Intermunicipal Poles (B), and Belts (C) are sub-categories of "Centri", while Intermediate (D), Peripheral (E), and Ultra-peripheral (F) correspond to "Aree Interne".

The conditioned box-plot in the top-right panel of Fig. 6 shows that the propensity to be subject to depopulation tends to be lower for the sub-categories of "Centri", with the lowest median value recorder by Italian municipalities labeled as Intermunicipal Poles (0.423). On the other hand, among the "Aree Interne", $CFDI$ is characterized by higher values as the distance from the closest "Centro" increases, with the highest median value recorded by municipalities labeled as Ultra-Peripheral (0.734). It is worth noting that the values of $CFDI$ for Poles and Intermunicipal Poles were characterized by lower interquartile ranges than the other subcategories. This is, at least in part, due to the fact that Poles and Intermunicipal Poles represent, respectively, only 2.3% and 0.7% of the total number of statistical units.

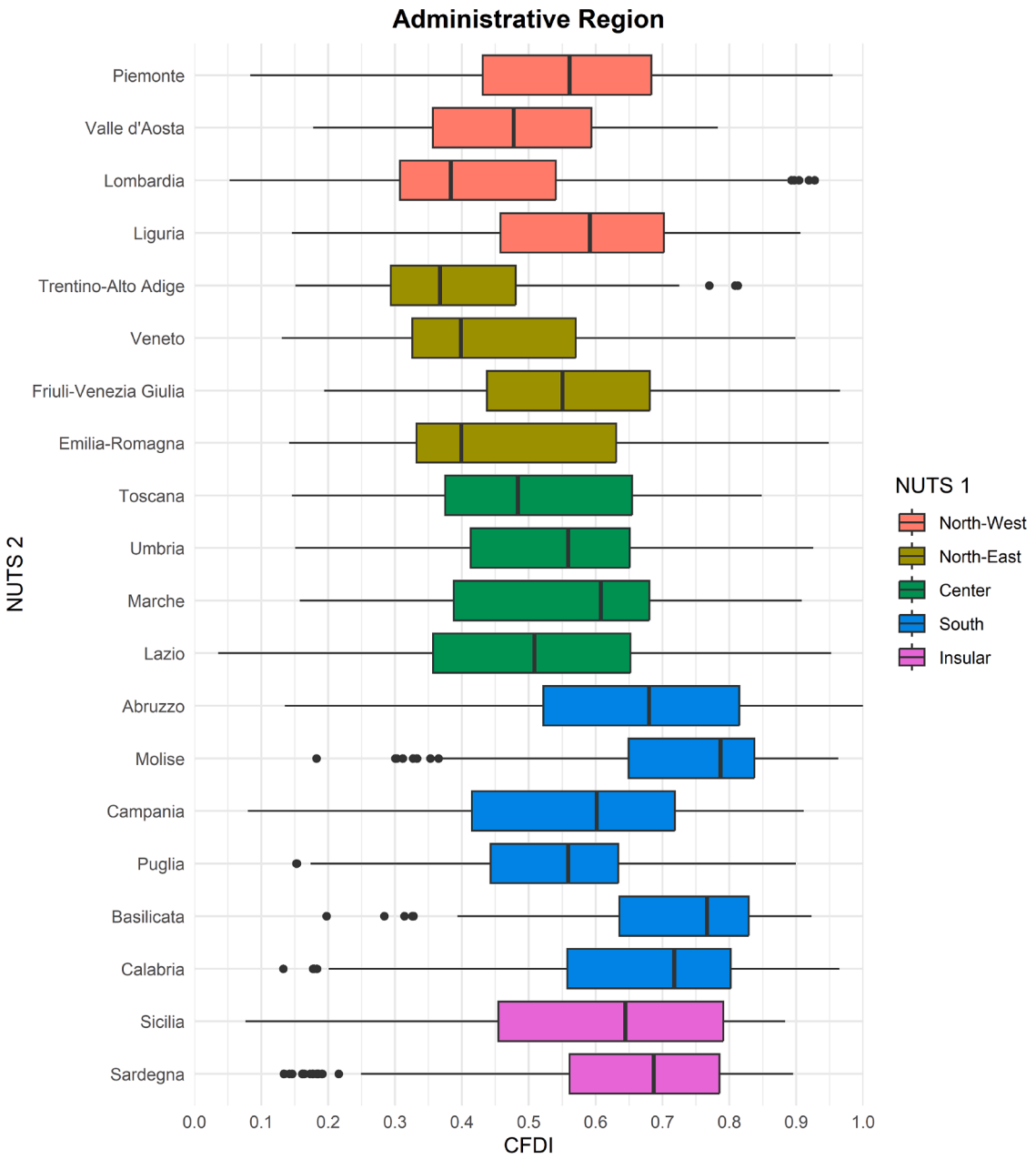

Fig. 5. Box-and-whiskers plots of $CFDI$ conditioned to the administrative regions (NUTS 2).

Italy is characterised by a significant exposure to seismic risk, which varies substantially across the national territory. Previous studies have shown that seismic events may exacerbate population decline in affected areas [33,34]. However, the present work did not aim to evaluate the impact of specific seismic events; rather, it investigated whether the presence of different levels of seismic risk might be associated with depopulation.

Seismic risk varies across municipalities. The Italian Civil Protection Department classifies the national territory into four seismic zones, ranging from Zone 1, corresponding to the highest level of risk, to Zone 4, representing the lowest. Each municipality is assigned to one of these zones. The analysis therefore focuses on whether different levels of seismic risk are associated with different propensities to depopulation.

The conditioned box plot in the bottom-left panel of Fig. 6 seems to suggest a potential statistical association between the CFDI and seismic risk. Median CFDI values decrease as the risk declines – from Zone 1 (0.700), characterized by a high probability of strong earthquakes, to Zone 3 (0.466), where the probability of strong earthquakes is low. The median value in Zone 4, where the probability of strong earthquakes is extremely low, is slightly higher (0.509) than in Zone 3, but still lower than in Zone 2 (0.579), where

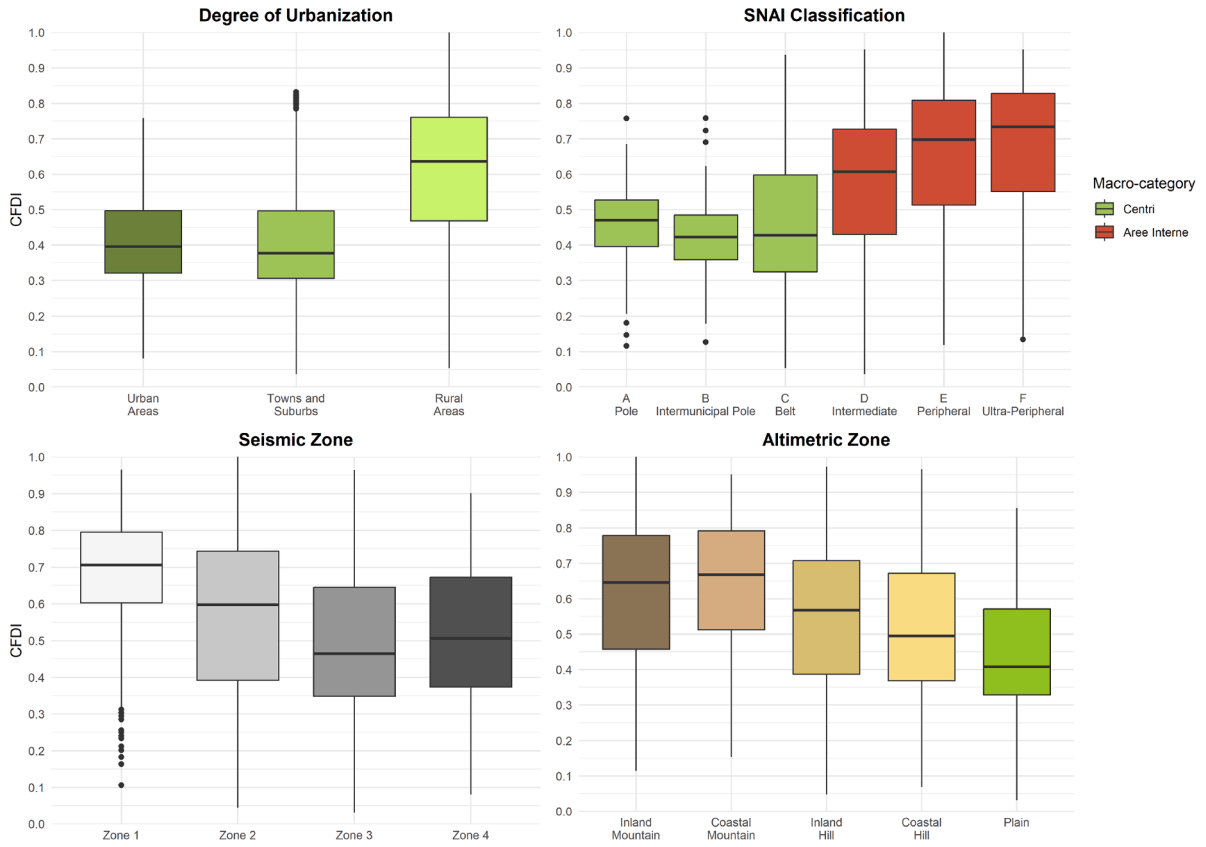

Fig. 6. Box-and-whiskers plots of $CFDI$ conditioned to four covariates: degree of urbanization, SNAI classification, seismic zone, and altimetric zone.

strong earthquakes are possible. Municipalities with the highest and lowest CFDI values are both located in Zone 2 characterized by medium-to-high seismicity, where strong earthquakes are possible but not as frequent or intense as in Zone 1.

Geographic features might play a role in the evolution of population dynamics. For example, in a country that is quite diverse as Italy, it could be interesting to study the role played by altitude and distance from the sea. In this regard, Italian municipalities are divided by ISTAT into five different altimetric zones [35]: (i) inland mountain (Italian: montagna interna), characterized by high altitude (over 600 – 700 meters) and not close to the sea; (ii) coastal mountain (Italian: montagna litoranea), high altitude and proximity to the sea; (iii) inland hill (Italian: montagna interna), mean altitude (approximately 300 – 600 meters) and not close to the sea ; (iv) coastal hill (Italian: collina interna), mean altitude and proximity to the sea; (v) plain (Italian: pianura), low altitude (below 300 meters).

Focusing on the conditioned box-plot in bottom-right panel of Fig. 6, it is worth noting that the municipality with highest $CFDI$ was in the inland mountain, while the one with lowest $CFDI$ was in plain. Furthermore, a decreasing pattern of $CFDI$ as the altitude decreased seemed to be present, with the proximity to the sea playing a key role in the hills, where the median value of $CFDI$ in coastal hill (0.496) was significantly smaller than in inland hill (0.553), but not in the mountains, where median value of $CFDI$ in coastal mountain (0.658) was even slightly higher than in inland mountain (0.639).

An additional feature of geographic areas that might be relevant for population attractiveness is the quality of infrastructures [16]. A modern example of infrastructure is the level of access to broadband services. Previous studies have shown that broadband availability is associated with lower levels of depopulation in rural areas [36,37], supporting the claim that digitalization may contribute to reversing negative demographic trends in disadvantaged areas [38].

In view of the foregoing, the potential role of the quality of households' Internet connections in population dynamics can be analysed. For each Italian municipality, data collected by AGCOM at the end of 2018 provide information on the number of households served by four different types of connections - listed from the worst to the best: Asymmetric Digital Subscriber Line (ADSL), Very-high-bit-rate Digital Subscriber Line (VDSL), Enhanced VDSL (EVDSL), and Fiber To The Home (FTTH) - as well as on connection speeds, classified into five ranges (0 – 2 Mbps, 2 – 30 Mbps, 30 – 100 Mbps, 100 – 500 Mbps, 500 – 1,000 Mbps).

Table 6 presents the Pearson correlation coefficients between, on the left-hand side, $CFDI$ and the share of households served by different types of connection, and, on the right-hand side, between $CFDI$ and the share of households grouped according to the speed ranges. The first two columns, related to the type, shows only negative coefficients. These were increasing in absolute value

Table 6
Pearson correlation coefficients between $CFDI$ and share of households served by four different types of connection, $CFDI$ and share of households grouped according to the speed ranges, and $CFDI$ and share of population per age class.

Type	CFDI	Speed	CFDI	Age class	CFDI
ADSL	−0.172	*No network*	0.282	*0 – 17*	−0.727
VDSL	−0.293	*0 – 2*	0.038	*18 – 64*	−0.563
EVDSL	−0.298	*2 – 30*	0.112	*65 or more*	0.650
FTTH	−0.072	*30 – 100*	−0.261		
		100 – 500	−0.227		
		500 – 1,000	−0.082		

as the quality of the connection increased, with the exception of the highest quality level (FTTH). This might be due to the fact that, in 2018, the distribution of this type of connection was still in the embryonic stage, as suggested by the extremely low value (0.74%) of the average share of households served by FTTH at the municipal level. The following two columns, referring to the speed of the connection, provide even more interesting insights: higher shares of households with no wireline network or connections with low speed (up to 30 Mbps) are positively associated with $CFDI$, whereas higher shares of households with connections characterized by high speed are negatively associated with $CFDI$. Furthermore, the negative coefficient of the highest range is small in absolute value and the reason might be similar to the one described for the share of FTTH connections.

Lastly, since the shrinking population of a local area is clearly dependent on natural growth, it seemed reasonable to analyse the age structure of the Italian municipalities at the end of the time horizon. For this reason, attention was focused on three age classes: (i) minors, i.e., the population under the age of 18; (ii) adult population, aged in the range $[18; 64]$; (iii) population in retirement age, with age greater than 64.

As shown in the last two columns of Table 6, the behaviour of $CFDI$ in relation to the share of inhabitants per age class was consistent with expectations: the coefficients referring to the share of minors and to the share of adults in working age were negative and characterized by quite high absolute values; on the other hand, the coefficient related to the share of people in retirement age is quite high and positive. These results could be seen as an additional signal that the proposed $CFDI$ was able to capture the phenomenon of the local depopulation.

6.1. Spatial data analysis

After having analysed a series of possible characteristics of the statistical units that could have been related to the propensity to depopulation, the research interest was shifted to the spatial distribution of the proposed index. The data analysed in the present study is, as a matter of fact, a spatial data set, as it directly refers to well-defined geographic areas. More specifically, it is a clear example of areal data [39].

When dealing with spatial data, it is interesting to understand whether space plays an active role in the distribution of the variable of interest. With the aim of verifying this analytically, the concept of spatial autocorrelation needs to be introduced. Spatial autocorrelation is the correlation of a variable with itself due to the spatial location of the observations. It is possible to measure it at a global scale, via Global Indices of Spatial Autocorrelation, or at a local scale, via Local Indices of Spatial Autocorrelation (LISA [40]). In the present work, for the variable of interest $CFDI$, both levels of analysis were considered and are described in the following.

Before presenting the results, it is necessary to introduce two additional basic concepts: the neighbourhood and the weight matrices. The neighbourhood matrix of a set of N statistical units ($\boldsymbol{N}$) is a $N \times N$ matrix which codifies the neighbourhood structure of the geographic units. There are several ways to determine whether two distinct regions are neighbours. In this context, a total of five different neighbouring structures were considered (Appendix C) which all led to almost identical findings. So, for the sake of brevity, only one is presented. The following results were obtained from a matrix $\boldsymbol{N}$ based on Delaunay's Triangulations was considered: for a given unit i, the municipalities connected by Delaunay's Triangulation were considered neighbours. Then, the link between two regions, say i and j, was transformed into an element w_{ij} of the weight matrix $\boldsymbol{W}$ as

$$w_{ij} = \begin{cases} \frac{d_{ij}^{-2}}{\sum_r d_{ij}^{-2}}, & \text{if } j \text{ is in the neighbourhood of } i, \\ 0, & \text{otherwise,} \end{cases} \tag{4}$$

denoting with d_{ij} the great-circle distance between the centroids of the regions i and j. The denominator is aimed at obtaining a row-standardization of the weights, i.e. $\sum_j w_{ij} = 1$, for all $i = 1, \dots, N$.

After having defined the neighbourhood structure, it is possible to detect the presence of some kind of spatial structure. A common method to proceed is to focus on the possible presence of autocorrelation on the $CFDI$ by computing the global Moran's I as

$$I_W = \frac{n}{\sum_i \sum_j w_{ij}} \frac{\sum_i \sum_j w_{ij}(CFDI_i - \overline{CFDI})(CFDI_j - \overline{CFDI})}{\sum_i (CFDI_i - \overline{CFDI})^2}, \quad i \neq j.$$

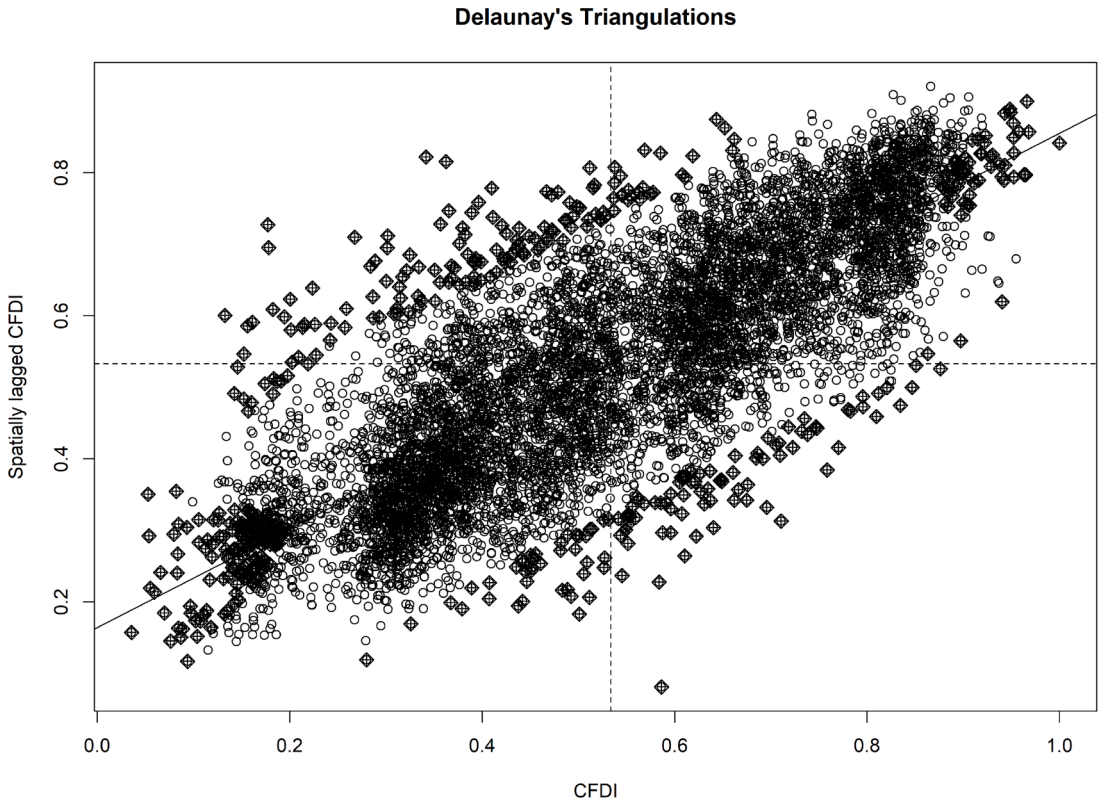

Fig. 7. Moran's diagram of $CFDI$, obtained via R [41] package `spdep` [42].

A value of the Moran's I greater than zero implies the presence of positive autocorrelation. However, it should not be interpreted exactly as the Pearson correlation coefficient because the range of I_W depends on the weight matrix and it does not always coincide with $[-1, 1]$.

In addition to quantifying spatial autocorrelation, Moran's I might be used as a test statistic for the hypothesis of the absence of autocorrelation, since

$$E(I_W) = -\frac{1}{R-1},$$

$$\frac{I_W - E(I_W)}{\sqrt{Var(I_W)}} \sim \mathrm{N}(0,1).$$

The hypothesis system to be tested is the following

$$\begin{cases} H_0 : \text{ Absence of spatial autocorrelation;} \\ H_1 : \text{ Presence of spatial autocorrelation.} \end{cases}$$

In this context, the inference was conducted under the randomisation hypothesis: the estimated statistic computed from the data was compared with the distribution obtained by randomly re-ordering the data. As shown in the second row of Table C.8, the estimated Moran's Index was high, positive (0.691) and significant. This finding suggested that the distribution of $CFDI$ was characterized by a strong phenomenon of positive autocorrelation.

In addition, another useful tool to detect the presence of spatial autocorrelation was considered: Moran's diagram (Fig. 7), i.e. a scatter plot with the values of $CFDI$ in the x-axis and the average values of $CFDI$ for the neighbouring observations in the y-axis. The dashed lines represent the mean value of $CFDI$ (0.534) and divide the diagram into four quadrants. A high frequency of observations in the upper right (i.e. high-high) and in the bottom left (i.e. low-low) quadrants is a sign of the presence of positive autocorrelation, whereas a high frequency of observations in the upper left (i.e. low-high) and in the bottom right (i.e. high-low) quadrants is a sign of the presence of negative autocorrelation. In the analysed data set, the 83.1% of the observations belong to the first and the third quadrant (penultimate column of Table C.8).

These findings suggested that local depopulation is a phenomenon characterized by positive spatial autocorrelation. In other words, Italian municipalities with high propensity to depopulation tend to be surrounded by other municipalities with high propensity to depopulation, and viceversa. This statement is based on the computation of a global measure of spatial autocorrelation, implicitly

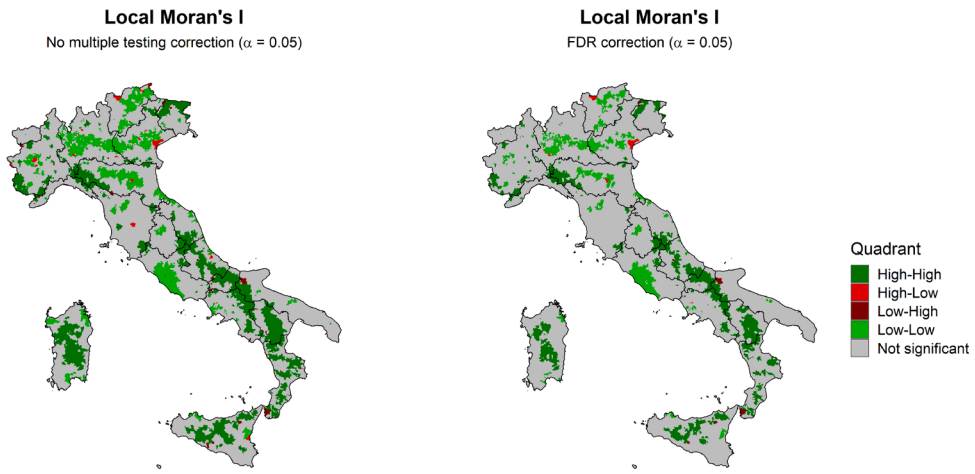

Fig. 8. Cartogram of the significant Local Moran's I (`ggplot2` [45]).

assuming that the spatial phenomenon is the same throughout the space (spatial stationary process). However, this is not always a realistic assumption. For this reason, LISA were considered in the present work. These indicators measure the intensity and significance of local autocorrelation between the value of a variable in a region and the value of the same variable in the neighbouring areas.

One of the most used LISA is the Local Moran's I, defined as

$$I_i = \frac{(CFDI_i - \overline{CFDI})}{\sum_{k=1}^{N}(CFDI_k - \overline{CFDI})^2/(N-1)} \sum_j w_{ij}(CFDI_j - \overline{CFDI}),$$

for every region i. When global autocorrelation is detected, as in the present case, the Local Moran's I helps identifying regions that have a particular impact on the global process or, on the other side, which stand out from it. As the analogous global index, also the Local Moran's I might be used as a test statistic for the hypothesis of absence of local autocorrelation in each of the N regions. Adopting the hypothesis of conditional randomization ([43]), i.e. I_i is compared with its mean value when all the observations except i are permuted, the results in Fig. 8 were obtained. The share of significant statistics ($\alpha = 0.05$) falls from 31.5%, when not adopting any adjustment method of the p-value, to 18.8%, when adopting the False Discovery Rate (FDR) correction in the original formulation proposed by [44].

Analysing the right panel of Fig. 8 along with the second row of Table C.9, it is possible to obtain further useful information on the spatial phenomenon underlying $CFDI$. Despite the presence of positive autocorrelation, there is a not negligible minority of units which show some significant differences from the global context. Most of these have a particular impact on the global phenomenon, being located in the first and in the third quadrant of the Moran's diagram.

The municipalities belonging to the former, High-High (HH) quadrant, are those with high propensity to depopulation and surrounded by other units with similar levels of propensity. These are mostly located in the Southern and in the Insular macro-regions, especially in the Appennines: the highest regional frequency (93) is shown by Abruzzo, followed by Calabria (91) and Sardegna (90); on the other hand, regions like Valle d'Aosta and Trentino-Alto Adige have no units in this quadrant.

The municipalities belonging to the Low-Low (LL) quadrant are characterized by the opposite situation, that is low propensity to depopulation and proximity to other municipalities with low propensity to depopulation. It is quite evident from the map that these municipalities are mostly located in Po Valley, in Trentino-Alto Adige and in the proximity of the capital. The hishest regional frequency is shown by Lombardia (344), followed with long distance by Veneto (92); on the other hand, four regions – Basilicata, Calabria, Friuli-Venezia Giulia, and Molise – have no units in this quadrant.

The municipalities that stand out from the global process are just a small minority. Almost all of them are not provincial capitals, but there are three exceptions that are worth mentioning. In the Low-High (LH) quadrant there is Reggio Calabria: this suggests that this city is not shrinking ($CFDI = 0.491$) while its neighbours are. A possible explanation is that Reggio Calabria has attracted people from the surrounding municipalities. On the other hand, in the High-Low (HL) quadrant, there are Bologna and Venice: this suggests that these two cities, even if not alarmingly subject to depopulation – Bologna showed a value of $CFDI$ equal to 0.549, very close to the average, and Venice equal to 0.588 – are surrounded by municipalities that are growing.

7. Conclusion and discussion

This paper proposed a new fuzzy index (CFDI), designed to capture the phenomenon of population decline without imposing thresholds for classifying geographical areas as "declining" or "non-declining". As a fuzzy index, the CFDI allows for a more flexible representation of the degree of belonging to the set of depopulating areas.

The empirical application to study the propensities of the Italian municipality to be subject to depopulation shows that the fuzzy framework can highlight intermediate situations often overlooked by crisp or purely statistical measures, providing a richer understanding of the phenomenon under study.

More in detail, the conditional distribution of the index to some specific variables, such as administrative regions, degree of urbanization, SNAI classification, seismic zone, and altimetric zone, reveal results consistent with the expectations. Moreover, the analysis on spatial distribution of the proposed index revealed the prevalence of a significant positive autocorrelation, meaning that Italian municipalities with high (low) propensity to depopulation tend to be surrounded by other municipalities with high (low) propensity to depopulation.

The spatial distribution of municipalities reveals heterogeneous depopulation dynamics, highlighting the relevance of context-dependent interpretations. Municipalities characterized by a high propensity for depopulation and surrounded by other municipalities with similar propensity to depopulation (HH quadrant in Fig. 8), predominantly located in Southern and Insular Italy, especially in mountainous areas such as the Apennines, experience the most persistent population decline. In these cases, inter-municipal coordination and shared service provision may represent relevant policy implications derived from the observed spatial patterns.

Reggio Calabria (LH quadrant in Fig. 8) constitutes a distinctive case, as it exhibits population growth despite being embedded in a region characterized by high depopulation propensity. This pattern is likely associated with internal migration from neighbouring municipalities and suggests the importance of explicitly accounting for spatial interaction mechanisms when interpreting local depopulation processes.

A different spatial configuration emerges for Bologna and Venice (HL quadrant in Fig. 8), where demographic stability at the municipal level coexists with population growth in surrounding areas. These two cities are not experiencing significant depopulation, but they are surrounded by municipalities that are growing. This suggests that their influence remains positive in terms of attractiveness, although many people may choose to live in nearby areas, for instance because of high housing and rental costs. This is a typical phenomenon in tourist destinations and university cities. In such cases, house owners could be encouraged to rent dwellings for long-term residential purposes rather than for short-term tourist use, in order to promote population return to the city. This consideration is probably more realistic for Bologna, whereas Venice represents a very specific case, where residential relocation to surrounding areas may also be driven by accessibility constraints associated with living in the Venetian Lagoon.

From a methodological standpoint, the CFDI explicitly incorporates two temporal dimensions: timing and persistence of population decline, combined through a time-weight vector that assigns greater importance to more recent observations and captures variability in the population variable P across municipalities in each time point. However, in its current formulation, the CFDI cannot be regarded as a multidimensional index. A natural extension would consist in embedding the CFDI within a multidimensional fuzzy framework, explicitly distinguishing between depopulation due to natural causes (negative natural balance) and depopulation due to migratory causes (negative migratory balance). Such an extension would enable the representation of multiple, interacting dimensions of depopulation within a unified fuzzy setting.

Socio-economic variables, which were deliberately excluded from the present analysis as they were not considered informative at this stage, could be incorporated in future developments to further enrich the multidimensional structure of the index. Accordingly, future research may focus on: (i) extending the CFDI to a multidimensional fuzzy system integrating demographic and socio-economic information, including census or survey data; and (ii) evaluating its robustness across different spatial scales and national contexts.

CRediT authorship contribution statement

Federico Bacchi: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Formal analysis, Data curation, Conceptualization; **Laura Neri:** Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Methodology, Formal analysis, Data curation, Conceptualization.

Data availability

Research Link Provided

Declaration of interests

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.

Appendix A. Properties - Proofs

Property 1 (Scale-invariance). *Let $\lambda = \lambda^*$ be a fixed value set a priori, $\min_t(p_t) = \min_i(p_{i,t})$, and $\max_t(p_t) = \max_i(p_{i,t})$, be fixed for each $t = 0, \dots, T$. For all $T \in \mathbb{N}$, for all possible levels $\boldsymbol{p}_i^* = (p_{i,0}^*, \dots, p_{i,T}^*) \in \mathbb{R}^{(T+1)}$ of the demographic variable P and for all $b \in \mathbb{R}$:*

$$CFDI_i(\lambda^*, b\boldsymbol{p}_i^*) = CFDI_i(\lambda^*, \boldsymbol{p}_i^*).$$

Proof. Referring to Eqs. 1 and 2, the two components of $CFDI$ were both based on the normalization of a function dependent on the realizations of P – respectively, I_T and J_T. Starting from the former, we have:

$$I_T(\boldsymbol{p}_i^*) = \sum_{t=1}^{T} \delta_t \log\left(\frac{p_{i,t}^*}{p_{i,t-1}^*}\right),$$

where $p_{i,t}^*$ is the generic element of the vector $\boldsymbol{p}_{i,t}^*$. Multiplying it by a scaling constant $b \in \mathbb{R}$, we obtain:

$$\begin{aligned} I_T(b\boldsymbol{p}_i^*) &= \sum_{t=1}^{T} \delta_t \log\left(\frac{bp_{i,t}^*}{bp_{i,t-1}^*}\right) = \\ &= \sum_{t=1}^{T} \delta_t \log\left(\frac{p_{i,t}^*}{p_{i,t-1}^*}\right) = I_T(\boldsymbol{p}_i^*). \end{aligned}$$

On the other hand, we have:

$$J_T(\boldsymbol{p}_i^*, \tau) = \sum_{t=1}^{T} \delta_t d_{i,t}(\tau),$$

where $d_{i,t}(\tau)$ is defined in function of the same log-difference $\log\left(p_{i,t}^*/p_{i,t-1}^*\right) \equiv \log\left(bp_{i,t}^*/bp_{i,t-1}^*\right)$. □

Property 2 (Increase-decrease Monotonicity). *Let $\lambda = \lambda^*$ be a fixed value set a priori, $\min_t(p_t) = \min_i(p_{i,t})$, and $\max_t(p_t) = \max_i(p_{i,t})$, be fixed for $t = 0, 1$, $\Delta p > 0$ denote the absolute variation of P between two time points, and p_i^* the level of P for statistical unit i at time $t = 0$. For all $p_i^*, \Delta p > 0$ ($\Delta p < p_i^*$):*

$$CFDI_i(p_i^*, p_i^* + \Delta p) < CFDI_i(p_i^*, p_i^*) < CFDI_i(p_i^*, p_i^* - \Delta p).$$

Proof. In the presence of an absolute change Δp between the two time points, we have:

$$\begin{cases} \log\left(\frac{p_i^*+\Delta p}{p_i^*}\right) > 0 & \Rightarrow I_1(p_i^*, p_i^* + \Delta p) \propto \delta_1 > 0; \\ \log\left(\frac{p_i^*}{p_i^*}\right) = 0 & \Rightarrow I_1(p_i^*, p_i^*) = 0; \\ \log\left(\frac{p_i^*-\Delta p}{p_i^*}\right) < 0 & \Rightarrow I_1(p_i^*, p_i^* - \Delta p) \propto -\delta_1 < 0. \end{cases}$$

As a consequence, recalling the construction of FDI as a one's complement (1) and that $\min(p_t) < p_i^* - \Delta p < p_i^* < p_i^* + \Delta p < \max(p_t)$, we obtain:

$$FDI_i(p_i^*, p_i^* - \Delta p) > FDI_i(p_i^*, p_i^*) > FDI_i(p_i^*, p_i^* + \Delta p). \tag{A.1}$$

On the other hand, for the second component of $CFDI$, it is necessary to take into account the value of τ:

$$\begin{cases} \log\left(\frac{p_i^*+\Delta p}{p_i^*}\right) > \tau & \Rightarrow I_1(p_i^*, p_i^* + \Delta p) \propto \delta_1 > 0; \\ 0 < \log\left(\frac{p_i^*+\Delta p}{p_i^*}\right) \leq \tau & \Rightarrow I_1(p_i^*, p_i^* + \Delta p) = 0; \\ \log\left(\frac{p_i^*}{p_i^*}\right) = 0 & \Rightarrow I_1(p_i^*, p_i^*) = 0; \\ -\tau \leq \log\left(\frac{p_i^*-\Delta p}{p_i^*}\right) < 0 & \Rightarrow I_1(p_i^*, p_i^* - \Delta p) = 0. \\ \log\left(\frac{p_i^*-\Delta p}{p_i^*}\right) < -\tau & \Rightarrow I_1(p_i^*, p_i^* - \Delta p) \propto -\delta_1 < 0. \end{cases}$$

As a consequence, recalling the construction of $PERS$ as a one's complement (2), we obtain:

$$PERS_i(p_i^*, p_i^* - \Delta p) \geq PERS_i(p_i^*, p_i^*) \geq PERS_i(p_i^*, p_i^* - \Delta p). \tag{A.2}$$

Aggregating the results obtained in (A.1) and (A.2), we finally get:

$$CFDI_i(p_i^*, p_i^* - \Delta p) > CFDI_i(p_i^*, p_i^*) > CFDI_i(p_i^*, p_i^* + \Delta p).$$

The extension to a time horizon with $(T + 1) > 2$ is straightforward. □

Property 3 (Proximity Monotonicity). *Let $\lambda = \lambda^*$ be a fixed value set a priori, $\min_t(p_t) = \min_i(p_{i,t})$, and $\max_t(p_t) = \max_i(p_{i,t})$, be fixed for each $t = 0, \ldots, T$, $\Delta p > 0$ denote the absolute variation of P between two time points, and $p_{i,t}^*$ a fixed level of P for statistical unit i at time t. In addition, it is assumed that for each i, the sequence $\boldsymbol{p}_i^*$ is constant in all time intervals, except the one where the variation Δp occurs.*

If the vector of time weights satisfies the condition of being increasing with time proximity ($\delta_1 < \delta_2 < \cdots < \delta_T$), then, for all $T \in \mathbb{N}$, for all $p_i^*, \Delta p > 0$ ($\Delta p < p_i^*$):

$$\begin{aligned} &CFDI_i(p_{i,0}^*, \dots, p_{i,T-1}^*, p_{i,T-1}^* + \Delta p) < \\ &CFDI_i(p_{i,0}^*, \dots, p_{i,T-2}^*, p_{i,T-2}^* + \Delta p, p_{i,T-2}^* + \Delta p) < \\ &\qquad \dots < \\ &CFDI_i(p_{i,0}^*, p_{i,0}^* + \Delta p, \dots, p_{i,0}^* + \Delta p). \end{aligned}$$

Proof. Starting from the first component of $CFDI$, recalling that $p_{i,0}^* = p_{i,1}^* = \cdots = p_{i,T}^*$ by assumption, we have:

$$\begin{cases} I_T(p_{i,0}^*, \dots, p_{i,T-1}^*, p_{i,T-1}^* + \Delta p) & = 0 + \cdots + 0 + \delta_T l_\Delta; \\ I_T(p_{i,0}^*, \dots, p_{i,T-2}^*, p_{i,T-2}^* + \Delta p, p_{i,T-2}^* + \Delta p) & = 0 + \cdots + 0 + \delta_{T-1} l_\Delta + 0; \\ \quad \dots \quad \dots \quad \dots \quad \dots & \quad \dots \quad \dots \quad \dots \quad \dots \\ I_T(p_{i,0}^*, p_{i,0}^* + \Delta p, \dots, p_{i,0}^* + \Delta p) & = \delta_1 l_\Delta + 0 + \cdots + 0; \end{cases}$$

where $l_\Delta = \log\left((p_{i,t}^* + \Delta p)/p_{i,t}^*\right)$ is the log difference computed in the time point t where a positive variation Δp is observed. Being l_Δ a fixed positive quantity, the comparison between the previous equations depends only on the time weights. Then:

$$\begin{aligned} &FDI_i(p_{i,0}^*, \dots, p_{i,T-1}^*, p_{i,T-1}^* + \Delta p) < \\ &FDI_i(p_{i,0}^*, \dots, p_{i,T-2}^*, p_{i,T-2}^* + \Delta p, p_{i,T-2}^* + \Delta p) < \\ &\qquad \dots < \\ &FDI_i(p_{i,0}^*, p_{i,0}^* + \Delta p, \dots, p_{i,0}^* + \Delta p). \end{aligned} \tag{A.3}$$

Shifting the attention to the second component, when $l_\Delta > \tau$, we have:

$$\begin{cases} J_T(p_{i,0}^*, \dots, p_{i,T-1}^*, p_{i,T-1}^* + \Delta p) & = \delta_T; \\ J_T(p_{i,0}^*, \dots, p_{i,T-2}^*, p_{i,T-2}^* + \Delta p, p_{i,T-2}^* + \Delta p) & = \delta_{T-1}; \\ \quad \dots \quad \dots \quad \dots \quad \dots & \quad \dots \\ J_T(p_{i,0}^*, p_{i,0}^* + \Delta p, \dots, p_{i,0}^* + \Delta p) & = \delta_1; \end{cases}$$

whereas, when $l_\Delta \leq \tau$, all the previous equations collapse to 0. Thus:

$$\begin{aligned} &PERS_i(p_{i,0}^*, \dots, p_{i,T-1}^*, p_{i,T-1}^* + \Delta p) \leq \\ &PERS_i(p_{i,0}^*, \dots, p_{i,T-2}^*, p_{i,T-2}^* + \Delta p, p_{i,T-2}^* + \Delta p) \leq \\ &\qquad \dots \leq \\ &PERS_i(p_{i,0}^*, p_{i,0}^* + \Delta p, \dots, p_{i,0}^* + \Delta p). \end{aligned} \tag{A.4}$$

Combining (A.3) and (A.4), we obtain:

$$\begin{aligned} &CFDI_i(p_{i,0}^*, \dots, p_{i,T-1}^*, p_{i,T-1}^* + \Delta p) < \\ &CFDI_i(p_{i,0}^*, \dots, p_{i,T-2}^*, p_{i,T-2}^* + \Delta p, p_{i,T-2}^* + \Delta p) < \\ &\qquad \dots < \\ &CFDI_i(p_{i,0}^*, p_{i,0}^* + \Delta p, \dots, p_{i,0}^* + \Delta p). \end{aligned}$$

□

Property 4 (Distributional Stationarity). *Let $CFDI_i(\lambda, \boldsymbol{p}_i)$ be the Composite Fuzzy Demographic Index defined in Eq. (3), where the demographic information $p_{i,t}$ follows a stochastic process of the form:*

$$p_{i,t} = p_{i,t-1} \exp\left(\varepsilon_{i,t}\right), \quad \varepsilon_{i,t} \sim \mathcal{D}(0, \sigma_\varepsilon^2),$$

with $\mathcal{D}$ being a distribution with zero mean and finite variance. Then, the distribution of $CFDI_i(\lambda, \boldsymbol{p}_i)$ does not change systematically as the time horizon T increases.

Proof. Under the given assumptions, the log-differences $l_{i,t} = \log(p_{i,t}) - \log(p_{i,t-1}) = \varepsilon_{i,t}$ are i.i.d. over t, with

$$\mathrm{E}[l_{i,t}] = 0, \quad \mathrm{Var}(l_{i,t}) = \sigma_\varepsilon^2.$$

The index component $I_T(\boldsymbol{p}_i) = \sum_{t=1}^{T} \delta_t l_{i,t}$ is thus a weighted sum of i.i.d. variables. If the weights δ_t are uniform, i.e., $\delta_t = 1/T$, the variance of I_T is σ_ε^2/T and vanishes as $T \to \infty$. Even for generic bounded weights, as in our general case with $\delta_t \in (0; 1)$, $\sum_t \delta_t = 1$, the Central Limit Theorem ensures convergence in distribution of I_T to a Gaussian random variable. The normalization operator $\mathrm{norm}(\cdot)$ rescales I_T to the $[0; 1]$ interval across units i, and this rescaling preserves distributional convergence. The same applies to the persistence component $J_T(\boldsymbol{p}_i, \tau)$, which is a weighted sum of discretized $l_{i,t}$ values based on a fixed threshold τ. Since $d_{i,t}(\tau)$ are

categorical i.i.d. variables, J_T converges in distribution under the same conditions. As $CFDI_i$ is a convex combination of FDI_i and $PERS_i$, each defined as the complement of a normalized sum, it inherits the distributional convergence of its components. Therefore, the distribution of $CFDI_i$ remains stable for increasing T, and the index is said to be distributionally stationary with respect to the time horizon. □

Property 5 (Robustness to Local Noise). *Let $CFDI_i(\lambda, \boldsymbol{pi})$ be the Composite Fuzzy Demographic Index defined in Eq. (3). Suppose that the time series of the demographic variable P for each unit i is affected by a localized random perturbation $\eta_{i,t}$ in one or more time points t, such that:*

$$p_{i,t} = p_{i,t-1} \exp\left(\eta_{i,t}\right), \quad \eta_{i,t} \sim \mathcal{D}(0, \sigma_\eta^2),$$

where $\eta_{i,t}$ is a zero-mean noise term independent of the underlying trend. Then, under mild regularity conditions on the weights δ_t and the normalization operator, the variation of $CFDI_i$ induced by such perturbations is bounded and asymptotically negligible for sufficiently large T.

Proof. The perturbed log-difference becomes:

$$l^*_{i,t} = \log\left(\frac{p^*_{i,t} \exp(\eta_{i,t})}{p_{i,t-1}}\right) = \log\left(\frac{p_{i,t}}{p_{i,t-1}}\right) + \log\left(\exp(\eta_{i,t})\right) = l_{i,t} + \eta_{i,t}.$$

That is, the log-difference absorbs the perturbation as an additive noise. The corresponding perturbed index component is:

$$I^*_T = \sum_{t=1}^{T} \delta_t (l_{i,t} + \eta_{i,t}) = \sum_{t=1}^{T} \delta_t l_{i,t} + \sum_{t=1}^{T} \delta_t \eta_{i,t} = I_T + \sum_{t=1}^{T} \delta_t \eta_{i,t}.$$

Since $\eta_{i,t}$ are independent with zero mean and bounded variance, and $\delta_t \in (0; 1)$ with $\sum_t \delta_t = 1$, the additional term $\sum_{t=1}^{T} \delta_t \eta_{i,t}$ has expected value zero and bounded variance (specifically, $\leq \sigma_\eta^2$). The same reasoning applies to the persistence component J_T if noise does not systematically alter the sign of $l_{i,t}$ across the threshold τ. Therefore, for moderate noise levels and balanced weights, the value of $CFDI_i$ is only slightly perturbed, and this perturbation does not significantly affect the ranking or interpretation of the index. This behavior supports the claim that CFDI is robust to local noise or measurement errors affecting the demographic time series. □

Appendix B. Contextual Covariates

The main statistics of $CFDI$, conditioned to the different categories of the covariates used in Section 6, are summarized in the Table below.

Table B.7
Main summary statistics of $CFDI$, conditioned to the categories of the contextual covariates presented in Section 6.

		Summary Statistics				
Variable	Freq.	Min.	Q1	Med.	Q3	Max.
Degree of Urbanization						
Urban Areas	3.2%	0.080	0.321	0.396	0.497	0.758
Towns and Suburbs	33.0%	0.036	0.306	0.377	0.497	0.832
Rural Areas	63.8%	0.053	0.469	0.636	0.761	1.000
NUTS 1						
North-West	37.9%	0.053	0.345	0.481	0.639	0.955
North-East	17.6%	0.131	0.330	0.412	0.596	0.966
Center	12.3%	0.036	0.376	0.525	0.660	0.952
South	22.6%	0.080	0.482	0.650	0.794	1.000
Insular	9.7%	0.076	0.483	0.662	0.789	0.896
SNAI						
Pole	2.3%	0.115	0.396	0.471	0.528	0.757
Intermunicipal Pole	0.7%	0.127	0.359	0.423	0.485	0.758
Belt	48.4%	0.053	0.325	0.428	0.598	0.937
Intermediate	24.4%	0.036	0.430	0.608	0.727	0.953
Peripheral	19.3%	0.118	0.513	0.698	0.809	1.000
Ultra-Peripheral	4.8%	0.134	0.551	0.734	0.828	0.952
Seismic Risk						
Zone 1	9.3%	0.133	0.590	0.700	0.798	0.965
Zone 2	29.9%	0.036	0.385	0.579	0.742	1.000
Zone 3	38.2%	0.053	0.339	0.466	0.634	0.955
Zone 4	22.6%	0.093	0.365	0.509	0.663	0.902
Altimetric Zone						
Inland Mountain	30.0%	0.126	0.458	0.639	0.782	1.000
Coastal Mountain	1.5%	0.178	0.515	0.658	0.796	0.943
Inland Hill	32.0%	0.055	0.383	0.553	0.702	0.968
Coastal Hill	9.9%	0.076	0.362	0.496	0.660	0.965
Plain	26.6%	0.036	0.315	0.404	0.560	0.856

Appendix C. Contiguity and Neighborhood Matrices

In addition to the structure presented in Section 6.1, four alternative neighbouring structures were considered to check the robustness of the results. First, the Queen-contiguity matrix was considered: for each location i, the units sharing at least one point in the space with it were considered neighbours. Then, the link between two neighbours i and j was transformed into a constant element $w_{ij} = 1/\#W_i$ of the weight matrix $\boldsymbol{W}$, with $\#W_i$ denoting the number of neighbours of the unit i. Secondly, the neighbourhood matrix based on sphere of influences (SOI): for a given unit i, the municipalities connected by the sphere of influences graph were considered neighbours. Lastly, the k-Nearest-Neighbour (kNN) method was considered: for a given unit i, the $k = \{5; 6\}$ nearest municipalities were considered neighbours. The w_{ij}'s for the neighbours obtained with Delaunay's Traingulations and with the kNN method were computed following the same procedure described in (4).

Table C.8
Global analysis of spatial autocorrelation with different Neighbourhood $\boldsymbol{N}$ and Weight matrices $\boldsymbol{W}$ [42].

	Moran's I				
W	Range	Statistic	P-value	Q1 - Q3	Q2 - Q4
Queen	[−1.000; 1.116]	0.644	< 0.001	82.1%	17.9%
Delaunay	[−0.994; 1.081]	0.691	< 0.001	83.1%	16.9%
SOI	[−1.084; 1.124]	0.698	< 0.001	82.8%	17.2%
5NN	[−0.606; 1.090]	0.675	< 0.001	82.5%	17.5%
6NN	[−0.551; 1.090]	0.663	< 0.001	82.0%	18.0%

Then, the same neighbourhood structures were used to perform the local analysis of spatial autocorrelation.

Table C.9
Local analysis of spatial autocorrelation with different Neighbourhood $\boldsymbol{N}$ and Weight matrices $\boldsymbol{W}$ [42], considering FDR multiple testing correction ($\alpha = 0.05$).

	Local Moran's I - Quadrant				
W	HH	LH	LL	HL	Not Significant
Queen	11.3%	0.3%	11.0%	0.2%	77.0%
Delaunay	9.7%	0.1%	8.9%	0.1%	81.2%
SOI	7.0%	0.1%	6.5%	0.1%	86.3%
5NN	10.6%	0.3%	10.2%	0.2%	78.7%
6NN	13.5%	0.3%	14.0%	0.3%	71.9%

References

[1] I.N.d. Statistica, Fertility indicators of the resident population, 2026, (https://demo.istat.it/app/?i=FE1&a=2024&l=en). Accessed: 2026-01-31.
[2] I.N.d. Statistica, Monthly Demographic Balance, 2026, (https://demo.istat.it/app/?i=D7B&a=2025&l=en). Accessed: 2026-01-31.
[3] I.N.d. Statistica, Bilancio demografico mensile gennaio-giugno 2025, 2025, (https://www.istat.it/notizia/bilancio-demografico-mensile-gennaio-giugno-2025/). Accessed: 2026-01-31.
[4] M. Battaglini, M. Simone, Indicatori Demografici | Anno 2024, Technical Report, Istituto Nazionale di Statistica, 2025. https://www.istat.it/wp-content/uploads/2025/03/Indicatori_demografici_2024.pdf.
[5] I.N.d. Statistica, Demo - Statistiche demografiche, 2025, (https://demo.istat.it/?l=en). Accessed: 2025-10-30.
[6] U.N.P. Division, World Population Prospects 2024, 2026, (https://population.un.org/wpp/). Accessed: 2026-01-15.
[7] T. Haartsen, V. Venhorst, Planning for decline: anticipating on population decline in the netherlands, Tijdschrift Voor Economische en Sociale Geografie 101 (2) (2010) 218–227. https://doi.org/10.1111/j.1467-9663.2010.00597.x
[8] F. Collantes, V. Pinilla, Peaceful surrender: the depopulation of rural Spain in the twentieth century, Cambridge Scholars Publishing, 2011. https://doi.org/10.1002/psp.1745
[9] R.A. Bijker, T. Haartsen, D. Strijker, Migration to less-popular rural areas in the netherlands: exploring the motivations, J. Rural Stud. 28 (4) (2012) 490–498. https://doi.org/10.1016/j.jrurstud.2012.07.003
[10] F. Barca, P. McCann, A. Rodríguez-Pose, The case for regional development intervention: place-based versus place-neutral approaches, J. Reg. Sci. 52 (1) (2012) 134–152. https://doi.org/10.1111/j.1467-9787.2011.00756.x
[11] I. Turok, V. Mykhnenko, The trajectories of European cities, 1960–2005, Cities 24 (3) (2007) 165–182. https://doi.org/10.1016/j.cities.2007.01.007
[12] OECD, Classifying shrinking regions: shrinking smartly and sustainably, 2025, https://doi.org/10.1787/0185fe09-en
[13] OECD, Shrinking Smartly and Sustainably: Strategies for Action, 2021, https://www.oecd.org/en/publications/rural-well-being_d25cef80-en.HTML.
[14] R.S. Franklin, I come to bury (population) growth, not to praise it, Spat. Econ. Anal. 15 (4) (2020) 359–373. https://doi.org/10.1080/17421772.2020.1802056
[15] N. Newsham, F. Rowe, Understanding trajectories of population decline across rural and urban europe: a sequence analysis, Popul. Space Place 29 (3) (2022) e2630. https://doi.org/10.1002/psp.2630
[16] E.O. N. f.T. Development, Shrinking rural regions in Europe – Towards smart and innovative approaches to regional development challenges in depopulating rural regions, 2017, https://archive.espon.eu/sites/default/files/attachments/ESPON%20Policy%20Brief%20on%20Shrinking%20Rural%20Regions.pdf.
[17] OECD, Rural Well-being: Geography of Opportunities. OECD Rural Studies, 2021, https://www.oecd.org/en/publications/rural-well-being_d25cef80-en.HTML.
[18] A. Cerioli, S. Zani, A fuzzy approach to the measurement of poverty, in: C. Dagum, M. Zenga (Eds.), Income and Wealth Distribution, Inequality and Poverty: Proceedings of the Second International Conference on Income Distribution by Size: Generation, Distribution, Measurement and Applications, Held at the University of Pavia, Italy, September 28–30, 1989, Springer, 1990, pp. 272–284. https://doi.org/10.1007/978-3-642-84250-4_18
[19] B. Cheli, A. Lemmi, A "totally" fuzzy and relative approach to the multidimensional analysis of poverty, Econ. Notes 24 (1995) 115–134.
[20] G. Betti, V. Verma, Fuzzy measures of the incidence of relative poverty and deprivation: a multi-dimensional perspective, Statist. Methods Appl. 17 (2) (2008) 225–250. https://doi.org/10.1007/s10260-007-0062-8
[21] G. Betti, A. Lemmi, Poverty and social exclusion: New methods of analysis, Routledge, 2013. https://doi.org/10.4324/9780203085172
[22] L.A. Zadeh, Fuzzy sets, Infor. Contr. 8 (3) (1965) 338–353. https://doi.org/10.1016/S0019-9958(65)90241-X
[23] B. Cardone, F. Di Martino, A novel spatiotemporal prediction method based on fuzzy transform: application to demographic balance data, Inf. Sci. 634 (2023) 677–695. https://doi.org/10.1016/j.ins.2023.03.117
[24] R. Piccarreta, E. Struffolino, Tools for analysing fuzzy clusters of sequences data, Demogr. Res. 51 (2024) 553–576. https://doi.org/10.4054/DemRes.2024.51.16
[25] L. Törnqvist, P. Vartia, Y.O. Vartia, How should relative changes be measured?, Am. Stat. 39 (1) (1985) 43–46. https://doi.org/10.1080/00031305.1985.10479385
[26] I.N.d. Statistica, Confini delle unità amministrative, 2025, (https://www.istat.it/notizia/confini-delle-unita-amministrative-a-fini-statistici-al-1-gennaio-2018-2/). Accessed: 2025-07-23.
[27] A.p. l.C. Territoriale, Mappa AI 2020 – Elenco e Classificazioni, 2025, (https://view.officeapps.live.com/op/view.aspx?src=https%3A%2F%2Fpolitichecoesione.governo.it%2Fmedia%2F2825%2F20220214-mappa-ai-2020-elenco-e-classificazione-comuni.xlsx&wdOrigin=BROWSELINK). Accessed: 2025-07-23.
[28] A. Goujon, C. Jacobs-Crisioni, F. Natale, C. Lavalle, The Demographic Landscape of EU Territories: Challenges and Opportunities in Diversely Ageing Regions, Publications Office of the European Union, 2021. https://doi.org/10.2760/658945
[29] I.N.d. Statistica, IstatData, 2022, (https://esploradati.istat.it/databrowser/#/en). Accessed: 2022-07-15.
[30] A.p. l. G.n. Comunicazioni, Boradband Map, 2022, (https://maps.agcom.it/arcgis/sharing/rest/content/items/1e8be5d1e1c743cfb384f74faad55625/data). Accessed: 2022-07-15.
[31] D.d.P. Civile, Seismic Classification, 2022, (https://rischi.protezionecivile.it/en/seismic/activities/emergency-planning-and-damage-scenarios/seismic-classification). Accessed: 2022-07-15.
[32] F. Crescenzi, L. Mori, G. Betti, F. Gagliardi, A. D'Agostino, L. Neri, An r tool for computing and evaluating fuzzy poverty indices: the package fuzzypovertyr, J. Appl. Stat. (2025) 1–14. https://doi.org/10.1080/02664763.2025.2481461
[33] M.A. Thiri, Social vulnerability and environmental migration: the case of miyagi prefecture after the great east japan earthquake, Int. J. Disast. Risk Reduct. 25 (2017) 212–226. https://doi.org/10.1016/j.ijdrr.2017.08.002
[34] D. Dottori, The effect of the earthquake in central italy on the depopulation of the affected territories, Reg. Sci. Urban Econ. 105 (2024) 103985. https://doi.org/10.1016/j.regsciurbeco.2024.103985
[35] Istituto Centrale di Statistica, Circoscrizioni statistiche metodi e norme, serie C n. 1, 1958.
[36] O. Lehtonen, Population grid-based assessment of the impact of broadband expansion on population development in rural areas, Telecomm. Pol. 44 (10) (2020) 1–12. https://doi.org/10.1016/j.telpol.2020.102028
[37] F. Merino, M.A. Prats, C.-J. Prieto-Sánchez, The access to broadband services as a strategy to retain population in the depopulated countryside in spain, Cities 144 (2024) 104647. https://doi.org/10.1016/j.cities.2023.104647

[38] A. Garashchuk, F. Isla-Castillo, P. Podadera-Rivera, The empirical evidence of digital trends in more disadvantaged European union regions in terms of income and population density, J. Reg. Sci. 65 (1) (2025) 75–111. https://doi.org/10.1111/jors.12729
[39] N. Cressie, Statistics for Spatial Data, John Wiley & Sons, 2015. https://doi.org/10.1002/9781119115151
[40] L. Anselin, Local indicators of spatial association - LISA, Geogr. Anal. 27 (2) (1995) 93–115. https://doi.org/10.1111/j.1538-4632.1995.tb00338.x
[41] R.C. Team, R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria, 2023. https://www.R-project.org/.
[42] R. Bivand, M. Altman, L. Anselin, R. Assunção, O. Berke, A. Bernat, G. Blanchet, Package 'spdep', Spat. Depend.: Weigh. Schemes, Statist., R Pack. Vers. (2017) 1. https://homepage.divms.uiowa.edu/ dzimmer/spatialstats/spdep.pdf.
[43] R.R. Sokal, N.L. Oden, B.A. Thomson, Local spatial autocorrelation in a biological model, Geogr. Anal. 30 (4) (1998) 331–354. https://doi.org/10.1111/j.1538-4632.1998.tb00406.x
[44] Y. Benjamini, Y. Hochberg, Controlling the false discovery rate: a practical and powerful approach to multiple testing, J. Royal Statist. Soc.: Series B (Methodological) 57 (1) (1995) 289–300. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x
[45] H. Wickham, ggplot2: Elegant Graphics for Data Analysis, Springer-Verlag New York, 2016. https://ggplot2.tidyverse.org.