Unveiling spatial patterns of population in Italy

Davide Fiaschi [*] Angela Parenti[†] Cristiano Ricci[‡]

February 28, 2026

Abstract

We study the evolution of local population density in Italy from 1985 to 2020, moving beyond traditional administrative classifications by adopting a fine spatial grid of approximately 82,500 cells (locations). We propose a novel empirical approach to characterise the dynamics of spatial distribution and derive five location types from projected long-run demographic trajectories, shedding new light on urbanisation, depopulation, and spatial clustering. We illustrate policy relevance through two applications: evaluating the Italian National Strategy for Inner Areas and constructing a forward-looking taxonomy of NUTS-3 regions that complements standard urban–rural classifications.

For space reasons, figures are not included. The full paper is available at: `https://www.dropbox.com/scl/fi/yo2hmgcbwobp597nlomkb/main_feb_2026.pdf?rlkey=jn249kea01tkub72pp28s9`
`dl=0`

JEL Classification Numbers: C14, O18, R12, R23

Keywords: random vector field, spatial dependence, spatial distribution dynamics, densification, depopulation, urban–rural dichotomy

[*]Davide Fiaschi, University of Pisa, Via Ridolfi, 10, 56124 Pisa, Italy (e-mail: davide.fiaschi@unipi.it).

[†]Angela Parenti (corresponding author), University of Pisa, Via Ridolfi, 10, 56124 Pisa, Italy (e-mail: angela.parenti@unipi.it).

[‡]Cristiano Ricci, University of Pisa, Via Ridolfi, 10, 56124 Pisa, Italy (e-mail: cristiano.ricci@unipi.it).

Contents

1 Introduction **4**

2 Conceptual framework **7**

2.1 Administrative borders 8

2.2 Spatial spillovers 8

2.3 Classifying locations 9

3 The Italian population at grid-cell level **10**

3.1 Data sources and sample selection 10

3.2 A first view on population density 11

4 The evolution of the population density **13**

4.1 The estimation of spatial dynamics 13

4.2 The estimate of RVF for Italian population density 15

5 Taxonomy of locations **16**

5.1 Types of locations within municipalities 18

6 Applications **19**

6.1 The National Strategy for Inner Areas (SNAI) 19

6.2 Classification of the NUTS 3 regions 21

7 Concluding remarks **22**

A Technical details on the estimation of RVF and forecast **29**

A.1 The adaptive kernel estimator with full covariance matrix 29

A.2 The choice of optimal estimation parameters h and α 30

A.3 The forecast via RVF in continuos time 31

B The procedure for the taxonomy of locations **32**

C Population and population density of Italian municipalities in Figure 4 **33**

1 Introduction

Population density has been the subject of numerous studies due to its association with various factors such as economic development, labour productivity, and overall individual welfare (see, e.g., Milanovic, 2018, Ahlfeldt and Pietrostefani, 2019, and Cattaneo et al., 2023). The dynamics observed result from a competition between the opposing forces of agglomeration and dispersion. On the one hand, factors such as wages, local amenities, and knowledge spillovers drive people toward densely populated areas, namely cities, at the expense of rural areas. On the other hand, factors such as house prices and congestion contribute to urban sprawl (Krugman, 1994, Glaeser, 2010, and Overman et al., 2010).

Despite apparent similarities in population density dynamics across European regions, there is considerable spatial heterogeneity, influenced not only by country-specific effects resulting from national policies, but also by different geographical characteristics and historical patterns (Krugman, 1996 and Pike et al., 2016). Italy represents a particularly relevant case in this context: its historical unification and administrative evolution have shaped both the number and spatial distribution of urban areas, affecting their density and long-term demographic trends. Dealing with this heterogeneity requires an approach that focuses on the *spatial distribution dynamics* of population density, a crucial piece of information for improving the design of European regional policies (McCann, 2013, Brezzi and Veneri, 2015 and Fiaschetti et al., 2021). This paper aims to provide such an approach, offering new insights into Italy's demographic local evolution and its implications for place-based policies.

A growing body of work addresses the spatial heterogeneity by proposing taxonomies of geographic units at multiple levels of aggregation (from regular grid cells to municipalities and larger regions) to distinguish urban cores, peripheries, and rural areas. These classifications rely on different methodological families. Some adopt fixed thresholds based on density and/or population size (e.g., OECD and Eurostat typologies; Brezzi et al., 2011, EUROSTAT, 2024), which are transparent and easy to communicate but may embed arbitrariness and limited portability across contexts (Duranton, 2021). Other contributions implement data-driven clustering approaches, including density-based algorithms such as DBSCAN (Arribas-Bel et al., 2021) and message-passing methods such as *affinity propagation* (AP) (Fiaschetti et al., 2021). These approaches can capture richer spatial structure than pure thresholds, but they typically depend on tuning parameters (e.g., density cut-offs or preference parameters) and often operate on static snapshots or synthetic summaries.

This paper contributes to this literature by introducing a tool to estimate and classify the *spatial distribution dynamics* of population density at a fine spatial scale, explicitly modelling local outcomes jointly with those of neighbouring areas. The key innovation is that locations are grouped not only by their current density level, but by their *projected long-run demographic trajectory* implied by the estimated vector field in Moran space. This dynamic perspective is particularly valuable for place-based policy design, because it separates areas that may look similar in cross-section but face systematically different long-run challenges (McCann, 2013, Brezzi and Veneri, 2015, Fiaschetti et al., 2021). In this sense, our approach complements existing urban-rural typologies and trajectory-based perspectives that emphasise heterogeneity in spatial development paths (Verburg et al., 2010, Van Eupen et al., 2012).

Relative to threshold-based frameworks (OECD/Eurostat), our method reduces arbitrariness by avoiding a priori cut-offs and by recovering endogenous partitions from convergence patterns in Moran space. Relative to clustering methods that rely on density thresholds (e.g., DBSCAN) or on preference parameters (e.g., AP), our approach leverages long-run convergence implied by the estimated RVF and selects the partition using multiple, conceptually distinct criteria (PDM, KL, Silhouette, WANG, and PAM), thereby strengthening robustness and transparency (see Appendix B for details). Moreover, because the classification is based on *basins of attraction*, it yields centroids that are directly tied to the support of observed data, providing a meaningful summary of each group's dynamic regime. At the same time, we acknowledge potential limitations: the approach is computationally intensive, requires long and spatially detailed time series, and its interpretation rests on the stability of the estimated dynamics over the projection horizon. For these reasons, we complement the baseline with robustness analyses and alternative specifications.

We apply our methodology to estimate population density dynamics across approximately 82,500 cells (hereafter referred to as locations) covering nearly the entire Italian territory over the period 1985-2020. Spatial dependence between locations is assumed to occur within Italian municipalities, given the crucial role played by administrative boundaries in residential choices (Nechyba and Strauss, 1998; Schirmer et al., 2014). We find evidence that population growth is concentrated in already dense locations and has become more spatially clustered over time. Coherently, two main patterns are identifiable - densification and depopulation - alongside strong local heterogeneity. From the clustering procedure five significant basins of attraction emerge and, consequently, five types of locations: (1) Urban Periphery Increasing Density (increasing population density starting from medium level and belonging to a high-populated municipality), (2) Urban Core Increasing Density (increasing population density starting from an already very high population density and belonging to a high-populated municipality), (3) Sparse Neighbourhood Decreasing Density (decreasing population density and belonging to a medium/low-populated municipality), (4) Remote (decreasing population density and belonging to a very low-populated municipality), and (5) Sparse Pocket (low populated and belonging to a low/medium populated municipality). The majority of the population (about 70%) is expected to belong to the *Urban Core Increasing Density* locations, which, however, cover only 16% of the territory. In contrast, the *Sparse Neighbourhood Decreasing Density* locations account for 44% of the total territory but only 19% of the total population. The *Urban Periphery Increasing Density* locations account for a similar share of population and territory (11% vs. 17%). Finally, *Remote* and *Sparse Pocket* locations account for negligible shares of the population (0.017% and 0.003%, respectively). However, while Remote locations cover nearly 23% of the territory, Sparse Pocket locations represent only 0.5%. A high level of heterogeneity is therefore present within municipalities, although some typical patterns can be found. In particular, most municipalities belonging to already urbanised areas tend to experience densification, which is mostly driven by the increase in the density of their belt (exceptions appear in the presence of natural parks or mountains). On the other hand, in more rural areas, municipalities, including a medium-sized city, tend to increase the density of this city at the expense of depopulation of surrounding locations, while those without a major city experience a depopulation at all. Importantly, the main qualitative patterns are robust to alternative modelling choices, including a different spatial weights matrix (10-nearest neighbours rather than within-municipality dependence) and a shorter estimation window (2000–2020 instead of 1985–2020).

We illustrate two applications that connect our dynamic cell-level classification to widely used administrative and statistical frameworks. First, we aggregate the results at the municipal level to inform the discussion on inner areas and to assess how local dynamics map into policy-relevant units, such as those targeted by Italy's National Strategy for Inner Areas (Barca et al., 2014). Second, we build a forward-looking taxonomy for Italian NUTS 3 regions and compare it with Eurostat's urban-rural typology, showing how a trajectory-based classification can complement cross-sectional categories by distinguishing "currently urban" from "urbanising" regions and by identifying transitional territories.

The remainder of the paper is organised as follows. Section 2 presents the conceptual framework. Section 3 describes the data and provides a first characterisation of spatial dynamics in Italian population density. Section 4 introduces the methodology. Section 5 develops the location taxonomy based on long-run spatial dynamics. Section 6.1 aggregates the location taxonomy to municipalities to discuss policy implications for Italy's National Strategy for Inner Areas, while Section 6.2 extends the approach to NUTS 3 regions and benchmarks our forward-looking classification against Eurostat's urban–rural typology. Finally, Section 7 concludes.

2 Conceptual framework

In this section, we address four key, interconnected issues related to the analysis of local population dynamics and place-based policies. First, we discuss why *administrative borders* pose significant limitations when studying the spatial distribution of population dynamics. Second, we highlight the importance of considering *spatial spillovers* from neighbouring locations, as these are critical to understanding population dynamics. In the context of Italy, we argue that the municipality serves as the most appropriate unit for capturing spatial spillovers in the dynamics of population density and for identifying heterogeneity within municipal administrative boundaries. Third, we explore how the results of our cell-level analysis can be *conveniently aggregated* to align with the relevant units of observation (e.g., municipalities, NUTS-2 regions, etc.) used in the design of various place-based policies. Finally, we also critically examine the existing literature on *classifying locations* into categories such as urban, rural, and suburban, particularly when detailed population density data is available at the cell level.

2.1 Administrative borders

The spatial distribution of population density exhibits significant variation, with distinct patterns emerging depending on the scale and level of observation. Demographic changes at finer geographical scales reveal a complexity that transcends large-scale studies and cannot be adequately captured by the static urban-rural dichotomy (Wilson, 1995; Krugman, 1996). For instance, within rural areas, some regions experience counter-urbanisation, while others face depopulation (Scott et al., 2007). Similarly, in urban areas, population growth can manifest as either concentrated development or urban sprawl (Fioretti et al., 2020). The lack of affordable housing often exacerbates urban sprawl, raising concerns about commuting patterns and the efficiency and sustainability of transport systems (Bhatta, 2010). Recognising the nuanced differences between broad area classifications and their specific local dynamics can offer valuable insights for the design, implementation, and monitoring of policies (Pateman, 2011). In this context, the analysis of local population density dynamics has been the focus of numerous studies, with objectives ranging from understanding urban agglomeration dynamics to identifying typical spatial patterns (Moreno-Monroy et al., 2021; Fioretti et al., 2020).

2.2 Spatial spillovers

The definition of the spatial spillover extent depends on the phenomenon under scrutiny. In this respect, our methodology is highly flexible. The perspective adopted here shifts the focus to a finer spatial scale than typical administrative borders (e.g., municipalities). However, even though population dynamics are examined at this granular level, we cannot overlook the role of spatial dependence, particularly spatial spillovers. The extent of these spillovers is an empirical question; we argue that, for Italian population dynamics, these neighbour relations should be defined at the municipal level. This is motivated by the fact that administrative boundaries play a crucial role in residential decisions, as they delineate the provision of key public services - such as schools, healthcare, and social infrastructure - as well as local fiscal policies (Nechyba and Strauss, 1998; Schirmer et al., 2014). In other words, studying local dynamics within municipalities while considering neighbour relations at the municipal scale allows for capturing how spatial dependencies emerge in a framework where decisions are influenced not only by geographical proximity but also by institutional and service-related boundaries. For instance, population shifts within a city may be shaped by school district zoning, healthcare accessibility, or other administrative regulations that do not necessarily align with purely geographical or functional spatial structures (Bayoh et al., 2006). Municipal boundaries remain fundamental in shaping daily life and residential choices. From a broader perspective, we argue that to properly account for the role of institutional settings in population distribution, it is essential to adopt a dual approach: a fine-grained analysis of local variations combined with an acknowledgement of the overarching influence of municipal boundaries.

2.3 Classifying locations

As noted, "the discretization of space into categories such as rural or urban involves a loss of information" (Duranton, 2021, p. 2). Mitigating this loss requires moving beyond the rural-urban dichotomy and implementing a richer, more flexible classification of locations. However, this may involve establishing specific rules, such as thresholds, or employing sophisticated statistical approaches (Duranton, 2021), which introduces a degree of arbitrariness into the analysis. Our methodology enhances existing approaches by focusing on the joint dynamics of population changes in a location and its neighbouring locations, rather than relying solely on a static snapshot of current population density. This approach has the significant advantage of being based solely on population density data and the definition of neighbouring units - information that can be easily gathered over extended periods, which is particularly important given the long timescale of demographic processes. Moreover, the clustering algorithm employed to determine the number of location types achieves an optimal balance between information loss and the level of synthetic representation necessary for effective policy design and evaluation. The long-run horizon of our analysis also allows for the study of the evolution of locations, i.e., how the system has evolved toward its present configuration. This is a crucial feature for designing future policies. From a policy-oriented perspective, the growing emphasis on the role of EU Cohesion Policy has led to a reshaping of subregional spatial policies and fostered the understanding of the functionality of territories (Mendez et al., 2021). This trend reflects evolving perspectives on the purpose of regional policy within the narratives and mechanisms of place-based approaches in EU Cohesion Policy. Scholars such as Barca (2009), Barca et al. (2012), Dąbrowski (2014), and Mendez (2013) have highlighted this shift, which began with the "Urban Agenda for the EU" and culminated in the "Sustainable Urban Development and Rural Development strategies" of the European Union (Fioretti et al., 2020).

3 The Italian population at grid-cell level

In this section, we first discuss the data sources for the analysis of the dynamics of Italian population density at granular grid cells. We then highlight its most important spatial patterns in the period 1985-2020.

3.1 Data sources and sample selection

To account for the high heterogeneity of population density across the Italian territory, we employ the *Global Human Settlement Layer Population Grid* (GHS-POP). This dataset provides residential population estimates at five-year intervals between 1975 and 2020 for a total number of cells of 47,752,226, with grid cells averaging 0.0064 km^2 in size, covering the entire Italian territory.[1]

To address computational constraints while maintaining a sufficiently small unit of observation, we aggregate the initial grid cells into larger cells using an aggregation factor of 22. This results in a total of 101,500 cells, each with an average size of 3.1214 km^2. For comparison, the average size of Italian municipalities is approximately 38 km^2. Although some municipalities are smaller than the aggregated cells, the latter account for only about 2.5% of the total. We exclude from the analysis cells where no population is present at the initial and final years, reducing the size of the

[1]For further details, see Pesaresi et al. (2024) and `https://human-settlement.emergency.copernicus.eu/ghs_pop2023.php`.

sample to 82,536.

A key data source on the Italian population is the Census, conducted every ten years since 1861. Since 1982, ISTAT has also provided annual intercensal population estimates by updating the most recent census figures with data on births, deaths, and both internal and international migrations reported by municipalities. From October 2018, ISTAT has been conducting an annual sample survey to collect key demographic characteristics at the municipal level. As a result, *yearly* population estimates are available for 1982–2021. To ensure the accuracy of the GHS-POP dataset, we validate its estimated data using official administrative population figures at the municipal level. In the analysis, we therefore restrict our attention to the period 1985–2020, as 1985 marks the first year when intercensal estimates became sufficiently established to allow reliable integration with census data for validation at this geographic scale. As discussed in Section 2.3, our analysis of local population dynamics considers neighbour relations at the municipal scale as a proxy for spatial spillovers. Since between 1985 and 2020 municipal boundaries underwent several modifications due to territorial reconfigurations and municipal mergers, to ensure comparability over time, we have harmonised municipalities to their 2020 definitions.

3.2 A first view on population density

Figure 1a displays the spatial distribution of population density across the 82,536 Italian cells (locations) in 2020. The population density distribution is strongly influenced by geographical features: low-density areas are prevalent in mountainous regions (such as the Alps and Apennines), while high-density areas are concentrated in the Po Valley and along the coasts. Spatial clusters of high-density locations highlight the presence of urban agglomerations (e.g., regional capitals like Florence and Bologna) and major metropolitan areas (e.g., Rome, Naples, Milan, and Turin).

The total Italian population increased from approximately 56.6 million in 1985 to 59 million in 2020. However, as highlighted in Figure 1b, this growth is geographically uneven, with population density tending to rise in already high-density areas and decline in low-density ones. A closer examination of the growth rates reveals that the density in major cities (e.g., Florence, Milan, Rome, and Naples) is growing less than in their surrounding locations, while the opposite is true for other urban areas (e.g., Cagliari, Catania, and Palermo). This suggests the possible coexistence of different kinds of *densification*. Additionally, *urban growth* areas appear along the borders of the Po Valley, while *depopulation* is primarily observed in mountainous regions.

This variegated pattern of population density dynamics is further confirmed by Figure 1c, which illustrates the nonlinear relationship between population density in 1985 and the annual growth rate at the location level from 1985 to 2020. The figure highlights that locations with a population density below the threshold of 45.4 (residents per km^2) experienced population decline over time, while those above this threshold saw population growth. Among the latter, locations with a density of around 1,347 exhibited the highest population growth rates.

[Figure 1 about here.]

The descriptive statistics on population density reported in Table 1 confirm our findings. First, while the mean population density increased slightly from 215.9 to 227.5 - reflecting the overall rise in total population - the median population density remained relatively stable, decreasing marginally from 21.1 to 20.6. The large and growing gap between the mean and median indicates that the distribution of population density is highly skewed. A few densely populated locations pull the mean upward, while the majority of locations remain sparsely populated. This suggests that population growth has been concentrated in already dense areas rather than being distributed more evenly across all locations. Additionally, the coefficient of variation (CV) decreased slightly from 3.8 to 3.6, indicating a modest reduction in the overall dispersion of population density.

[Table 1 about here.]

Assuming that the extent of spatial spillovers is primarily determined by membership to the same municipality, as discussed in Section 2.2, Table 1 highlights that, while the total variance and its decomposition between and within municipalities remain relatively stable, the Moran's I index - a measure of spatial autocorrelation - increased significantly from 0.32 to 0.36.[2] This rise suggests that population density has become more spatially clustered over time, with high-density locations increasingly concentrated near other high-density areas and low-density locations near other low-

[2]The 95% confidence intervals of Moran's I in 1985 and 2020 are $(0.316, 0.324)$ and $(0.356, 0.364)$, respectively.

density areas, where "areas" should be meant as municipalities. Overall, we can conclude that while the average population density has increased slightly, its initially highly skewed distribution has become even more skewed. This is further reinforced by the notable increase in spatial agglomeration from 1985 to 2020.

The observed spatial heterogeneity in population density and its evolution over time suggest the need for a methodological approach capable of capturing these dynamics systematically. In particular, understanding how population density evolves at a local level requires an estimation of spatial distribution dynamics that accounts for local interactions and spatial dependence. The following section introduces a methodology based on the estimation of a Random Vector Field (RVF) to systematically analyse these dynamics and provide insights into long-term spatial equilibria.

4 The evolution of the population density

We now proceed to estimate the spatial distribution dynamics of the Italian population density. In Section 4.1, we illustrate our econometric methodology for estimating spatial dynamics, specifically estimating a Random Vector Field (RVF) on the Moran space, while in Section 4.2, we present the outcomes of our RVF estimation for Italy.

4.1 The estimation of spatial dynamics

The dynamics of population density can be estimated using the stochastic kernel proposed by Quah (1996, 1997). This approach allows for tracking intra-distribution dynamics of units, generalising Markov transition matrices to a continuous state space. However, when analysing the spatial evolution of population density, it becomes essential to account for spatial dependence, i.e., spatial spillovers. This introduces a challenge known as the "curse of dimensionality", as highlighted by Silverman (1986).[3] Here, we adopt an alternative methodology based on the estimation of an RVF on a Moran space. This approach estimates the joint dynamics of the population density of a

[3]One attempt to address this issue is the conditioned stochastic kernel proposed by Quah (1997), where the variable of interest is expressed in relative terms with respect to the level of neighbouring units. While this approach incorporates the spatial component, it cannot capture the joint increase or decrease of both municipal density and that of its neighbours when the relative level remains unchanged. Another approach is that of Gerolimetto and

location and the municipality to which it belongs, modelling these dynamics as a vector field.

Let $\mathbf{y}_{t_0} := (y_{1,t_0}, \cdots, y_{N,t_0})$ and $\mathbf{y}_{t_0+\tau} := (y_{1,t_0+\tau}, \cdots, y_{N,t_0+\tau})$, be the observed level of (log) population density for cell j at year t_0 and $t_0 + \tau$ for $j = 1, \ldots, N$, respectively, and N the number of observations in the sample. Moreover, let $\mathbf{W}$ the $N \times N$ the row-standardised *spatial weight matrix* defining for each location its membership to a municipality; hence, the element (i, j) will be equal to zero in the case location i and j do not belong to the same municipality and equal to 1 over the number of locations in the municipality otherwise, with the convention that the element (i, i) is equal to zero. Define $\mathbf{z}_{j,t_0} := (y_{j,t_0}, (\mathbf{Wy})_{j,t_0})$ and $\mathbf{z}_{j,t_0+\tau} := (y_{j,t_0+\tau}, (\mathbf{Wy})_{j,t_0+\tau})$ the observation j in Moran space in year t_0 and $t_0 + \tau$, respectively. Therefore, define the observed movements as $\boldsymbol{\delta}_{j,\tau} := \mathbf{z}_{j,t_0+\tau} - \mathbf{z}_{j,t_0}$. For each point $\mathbf{z}$ in the Moran space, define the *expected* movement as the weighted average of all the observed ones around the point, as:

$$\widehat{\mathbf{RVF}}_{\tau}(\mathbf{z}) := \sum_{j=1}^{N} \omega_{j,t_0}(\mathbf{z}) \boldsymbol{\delta}_{j,\tau} \tag{1}$$

where:

$$\omega_{j,t_0}(\mathbf{z}) := \frac{\frac{1}{Nh^2} K\left(\frac{\mathbf{z} - \mathbf{z}_{j,t_0}}{h}\right)}{\hat{f}_{Y_{t_0}, WY_{t_0}}(\mathbf{z})},$$

$$\hat{f}_{Y_{t_0}, WY_{t_0}}(\mathbf{z}) := \sum_{i=1}^{N} \frac{1}{Nh^2} K\left(\frac{\mathbf{z} - \mathbf{z}_{i,t_0}}{h}\right),$$

where $K(\cdot)$ is a *kernel function* in $\mathbb{R}^2$ and h is a smoothing parameter (*bandwidth*) (Silverman, 1986). Heuristically, Eq. (1) represents the expected movement at point $\mathbf{z}$, obtained by summing all observed movements $\boldsymbol{\delta}_{j,\tau}$ from t_0 to $t_0 + \tau$, with weights $\omega_{j,t_0}(\mathbf{z})$ that depend on the distance of each observed movement from $\mathbf{z}$ at time t_0. The latter corresponds to the relative contribution of observation j to the joint density estimation of observations $\mathbf{z}_{i,t_0}$, evaluated at $\mathbf{z}$.

For estimation, we consider a more sophisticated version of Eq. (1), which accounts for the covariance between the two components of $\mathbf{z}_{j,t_0}$ and incorporates a bandwidth that varies with local

Magrini (2022), where the authors consider spatial dependence arising from omitted variables that are spatially correlated. Although this method accounts for spatial effects in the estimation of the stochastic kernel, thereby avoiding bias from omitted autocorrelation, it does not track the dynamics of neighbouring units.

observation density (Silverman, 1986; Fiaschi et al., 2018).[4] More technical details are provided in Appendix A. The main advantage of using RVF is that it mitigates the curse of dimensionality, as it relies on two-dimensional density estimation rather than a four-dimensional one, as required in the stochastic kernel approach.[5] Specifically, the number of dimensions is reduced from four to two by conditioning observed movements on the initial observation, which encapsulates all relevant dynamic information. The estimated RVF can also be easily visualised by drawing an arrow for each evaluation point in the Moran space, as illustrated in Figure 2.

4.2 The estimate of RVF for Italian population density

Figure 2 displays the estimation of the RVF for our sample of 82,536 locations. The red arrows represent the estimated directions and the intensity of the movement, with their lengths divided by a factor of 5 for graphical clarity. The grey area indicates where RVF is calculated, as determined by the distribution of observations in the Moran space. Yellow and green curves represent the contour lines of the distribution of observations in the years 1985 and 2020 respectively. Additionally, we include the estimated non-parametric Moran curve using the Nadaraya-Watson estimator, along with its 95% confidence bands for both years (blue and purple lines for 1985 and 2020, respectively).[6]

[Figure 2 about here.]

Moran space is divided into four regions based on the sample averages of the population density of locations and of municipalities to which they belong, reported by dashed lines in Figure 2. Most arrows in the HH quadrant (top-right in Moran space) point toward the upper right, indicating an increase both in the location density and in the density of its corresponding municipality. This pattern suggests a pattern of urban *densification* (Bhatta, 2010). Consequently, the non-parametric Moran in the HH quadrant exhibits a slight rotation toward the bisector. In the HL quadrant (bottom-right in the Moran space), arrows predominantly point to the left, though the direction

[4] Specifically, our estimation uses an *Epanechnikov kernel* for $K(\cdot)$, with $h = 0.21$ and $\alpha = 0.0067$ (the parameter adapting bandwidth to local observation density); see Appendix A.

[5] Estimating the stochastic kernel while explicitly maintaining spatial dependence would require a joint non-parametric density estimation of $\left(\mathbf{y}_{t_0}, \mathbf{W}\mathbf{y}_{t_0}, \mathbf{y}_{t_0+\tau}, \mathbf{W}\mathbf{y}_{t_0+\tau}\right)$.

[6] Results are robust to alternative specifications of the spatial weights matrix W (10-nearest neighbours instead of within-municipality contiguity) and to restricting the estimation window to 2000–2020 (instead of 1985–2020).

of movement varies depending on the population density of the municipalities. In particular, two opposing dynamics are observed. Locations belonging to municipalities with relatively high population density have experienced a marked decline, while their municipalities have not followed the same trend. In contrast, locations belonging to municipalities with relatively low population density have maintained their population density despite an increase in the population density of their municipalities. The overall pattern suggests that high-density locations within low-density municipalities are losing population.

In the LL quadrant (bottom-left in the Moran space), most arrows point downward and to the left, signalling a decrease in density at both the location and municipal levels. This pattern is indicative of *depopulation*. As a result, the non-parametric Moran curve in this quadrant also exhibits a slight rotation toward the bisector. Finally, in the LH quadrant (top-left in Moran space), different patterns emerge. A divergent dynamic appears near the vertical dashed line, with arrows pointing in two opposite directions: the arrows closest to the line point toward the upper-right, indicating a trend similar to the one prevalent in the HH quadrant of densification; on the contrary, the arrows slightly farther away point to the (downward-)left, suggesting a drop in location population density while maintaining the same density in their municipality. Finally, the dynamics in the far left of the quadrant are strongly oriented toward an increase in location density.

The estimated RVF highlights distinct spatial dynamics across different areas of the Moran space, revealing patterns of densification and depopulation that vary depending on both local and municipal-level conditions. However, while these findings provide a dynamic representation of population movements, a more structured classification is needed to identify regions with similar long-term demographic trajectories. The next section develops a taxonomy of locations based on their projected evolution, allowing for a clearer categorisation of spatial patterns and their policy implications.

5 Taxonomy of locations

In this section, we introduce an algorithm designed to classify locations based on the forecasted evolution of their population density and that of their municipalities. The expected evolution is computed using the estimated RVF, with a projection horizon of 10,000 years.

A comprehensive description of the classification algorithm is provided in Appendix B. Briefly, we simulate forward the dynamics implied by the estimated RVF from a dense grid of evaluation points in the Moran space to recover the long-run equilibrium associated with each point. We then cluster these equilibria using multiple, conceptually distinct selection criteria - Pham et al. (2005) (PDM), Krzanowski and Lai (1988) (KL), Rousseeuw (1987) (Silhouette), Wang (2010) (WANG), and Kaufman and Rousseeuw (1990) (PAM) - to identify *regions of attraction*. Using several criteria provides a robustness check for the selected partition; the detailed outcomes are reported in Appendix B. In the baseline specification, this procedure yields five regions (Figure 3a), with basins defined by the set of points converging to the same region; results remain very similar under a four-region alternative (see Figure 8 in Appendix B).

[Figure 3 about here.]

Figure 3a and Table 2 characterise five types of locations based on their membership in different basins of attraction. In Figure 3a, locations of Type 1 (light brown region) and Type 2 (dark brown region) are expected - starting from a currently high population density - to further increase in density and converge toward the bisector, i.e., a region where locations have approximately the same density as the municipality they belong to. In particular, locations of Type 1 have an actual population density lower than their surrounding neighbourhood (see Table 2), which suggests labelling them Urban Periphery Increasing Density (UPID) locations. In contrast, locations of Type 2 have an actual population density higher than their neighbourhood (see Table 2), suggesting the label Urban Core Increasing Density (UCID) locations. Locations of Type 3 (aquamarine region) belong to municipalities with very low population density (on average, 65.5 inhabitants per square kilometre; see Table 2). Regardless of their initial density, their population is expected to decline over time, which suggests labelling them Sparse Neighbourhood Decreasing Density (SNDD) locations. Locations of Type 4 (dark green region) currently have an extremely low population density (on average, 0.14; see Table 2) with no substantial expected changes in the future. This suggests labelling them Remote (R) locations. Finally, locations of Type 5 (gold region) also have an extremely low actual population density (on average, 0.99; see Table 2) but are surrounded by high-density neighbourhoods. Since no substantial demographic dynamics are expected, these locations can be labelled Sparse Pocket (SP) locations.

[Table 2 about here.]

Figure 3b highlights the complex distribution of these five types of locations. Each type, except for Type 5, represents a significant portion of Italian locations and population, as shown in Table 2. In particular, SNDD locations cover 44.1% of the total territory but account for only 19.2% of the population and are distributed across the entire country. UCID locations comprise 69.6% of the population but only 15.5% of the territory, with a strong concentration along the coastline and in the Po Valley. R locations host just 0.017% of the total population but span 23.1% of the territory, primarily in mountainous and inland areas.

5.1 Types of locations within municipalities

The extreme granularity of the grid used in the analysis can be appreciated in Figure 4, which examines nine representative Italian municipalities whose territories are partitioned according to our location classification (Section C in the appendix contains some descriptive statistics of these municipalities). Rome, the capital of Italy, contains four types of locations; however, as expected, UCID locations dominate in the centre, while UPID locations prevail around it (Figure 4a). Milan, Italy's third-largest municipality by area, exhibits an even clearer pattern of densification across all its locations within the municipality (Figure 4b). This trend is also evident in several other urban municipalities, such as Modena in the Po Valley (Figure 4c).

Some municipalities with significant mountainous areas, such as Genoa and L'Aquila, show a mix of UCID and UPID locations near R and SP locations (Figures 4d and 4e). Conversely, Grosseto and Sassari exemplify municipalities with a central cluster of UPID locations surrounded mainly by SNDD locations, i.e. an increasing density in central locations strongly contrasts with depopulation in peripheral areas (Figures 4f and 4g). Finally, Brindisi and Pomarance (a municipality in central Tuscany) are predominantly composed of SNDD locations due to their low average population density at the municipal level. However, while (almost) all locations in Pomarance have low density, Brindisi contains some highly populated locations that are expected to experience depopulation in the future.

We conclude by noting that the classification proposed in this section provides a structured framework for understanding the spatial dynamics of population density at a highly disaggregated level. This taxonomy is particularly valuable for informing place-based policies, as it allows for precise identification of areas experiencing different types of demographic transitions. Two illustrative examples are presented in Section 6.

[Figure 4 about here.]

6 Applications

While the classification of locations offers valuable insights at a granular level, policy interventions are typically implemented at broader administrative scales, such as municipalities or functional regions. Our methodology can be readily extended to inform policy design and evaluation across multiple territorial levels by aggregating the identified location types to higher spatial units, thereby enabling comparisons with policy-relevant territorial frameworks. Accordingly, Section 6.1 assesses the extent to which our taxonomy can support the design and validation of the National Strategy for Inner Areas (SNAI), which aims to counteract socio-economic decline in disadvantaged Italian territories. Section 6.2 then benchmarks our classification against Eurostat's urban–rural typology for NUTS 3 regions, a key reference for EU cohesion and rural development policies in identifying spatial disparities in demographics, economic conditions, access to services, and infrastructure.

6.1 The National Strategy for Inner Areas (SNAI)

Since 2012, Italy has developed the National Strategy for Inner Areas (SNAI) to leverage the potential of municipalities that, despite being distant from essential service hubs, possess significant environmental and cultural resources and a long history of human settlement (Barca et al., 2014). Many of these areas face depopulation, declining public and private services, and degradation of cultural and landscape heritage. The ultimate goal of the strategy is to reverse and improve demographic trends (Barca et al., 2014, p. 8).

Barca et al. (2014) classify municipalities into "Urban poles of attraction" and "Intermunicipal poles of attraction", which provide a full range of secondary education, emergency care hospitals, and relevant railway stations, and into "Outlying", "Intermediate", "Peripheral", and "Ultra-peripheral" areas, defined by their distance (20, 40, 75+ minutes) from the nearest hub. Intermediate, Peripheral, and Ultra-peripheral areas are collectively labelled as *Inner Areas*, many of which have experienced negative demographic trends since the late 1970s. SNAI targets selected project areas, grouping neighbouring municipalities, based on over 100 socio-economic criteria including demographic trends, land use, education, transportation, and healthcare. From 2014–2020, 72 project areas were selected, covering 1,060 municipalities and roughly 2 million inhabitants.[7]

The methodology proposed in Section 5 can support the identification and validation of project areas by estimating "intervention-free" demographic trajectories (Barca et al., 2014, p. 16). For each municipality, we compute the composition of the five location types and construct a dissimilarity matrix using the Gower (1971) coefficient. Applying PAM clustering (Kaufman and Rousseeuw, 2009) yields four clusters: "Urban", "Peripheral urban", "Peripheral", and "Rural" municipalities. Comparison with Barca et al. (2014) shows that our approach reproduces the classification of inner areas, with expected discrepancies in urban poles due to non-demographic criteria.

Consistent with the EU cohesion principle of *concentration* (whose only exceptions appear related to the regional boundaries) and the presence of budget constraints, only a fraction of Inner Areas are designated as project areas (Figures 5a vs. 5c). Figure 5d illustrates that most project areas are composed of rural and peripheral municipalities, with few urban or peripheral urban municipalities included, largely due to the requirement of geographic contiguity, except in some regions such as southern Apulia.

[Figure 5 about here.]

[7]See `https://politichecoesione.governo.it/âĂę` for a full list.

6.2 Classification of the NUTS 3 regions

Eurostat publishes regional statistics across a wide range of domains, and these data are extensively used to inform EU regional policy. To capture differences in settlement patterns between rural and urban areas, Eurostat adopts an urban–rural typology for NUTS 3 regions based on 1 km^2 population-grid data. In a first step, grid cells are classified using neighbouring-cell criteria: urban clusters are defined as contiguous cells with a population density of at least 300 inhabitants per km^2 and a minimum cluster population of 5,000; all remaining cells are classified as rural (i.e., outside urban clusters). In a second step, these classifications are aggregated within NUTS 3 boundaries to assign each region to one of three categories: *predominantly urban* (at least 80% of the population lives in urban clusters), *intermediate* (more than 50% but less than 80% lives in urban clusters), and *predominantly rural* (at least 50% lives in rural grid cells) (see Figure 6a).

As a possible application of our methodology, we apply the same steps described in the previous section to classify Italian NUTS 3 regions. The optimal number of clusters via PAM (Kaufman and Rousseeuw, 2009) is three, and the resulting taxonomy is shown in Figure 6b. While Figure 6a highlights that Eurostat's urban-rural typology provides a cross-sectional picture of settlement structure based on the current share of population living in urban clusters, our taxonomy is explicitly forward-looking, grouping NUTS 3 regions by their projected long-run population-density trajectory (depopulation vs. moderate vs. high densification, see Figure 6b). Accordingly, the two maps largely coincide for major metropolitan areas where high density is coupled with sustained densification (e.g., Milan, Varese and Venice). At the same time, our approach reclassifies some regions that are "urban" in a purely structural sense but do not exhibit persistent densification, such as Turin, Cagliari, and Rome, towards the medium-density (moderate densification) group. This is particularly evident in the Rome-Latina case: settlement dynamics in the metropolitan area of Rome have been widely described as characterised by urban diffusion rather than sustained densification, whereas Latina displays clearer densification trends (Di Feliciantonio and Salvati, 2015).

Conversely, it identifies as highly densifying a set of polycentric lowland provinces along the Po Valley corridor (including Bologna, Modena, Ravenna and areas around Venice), consistent with evidence of long-run diffuse urbanisation in the Po Valley (Romano and Zullo, 2016). A similar logic applies in Campania: Naples remains within the highly densifying group, but our taxonomy also flags inland provinces such as Benevento and Avellino (classified as less urbanised in Eurostat's framework), reflecting documented processes of demographic change and urban sprawl in these medium-sized cities (Bencardino, 2015). Finally, the relative scarcity of highly densifying regions in Apulia and Sicily is consistent with the weaker demographic momentum projected for the Mezzogiorno (Sarno, 2021).

[Figure 6 about here.]

[Table 3 about here.]

7 Concluding remarks

In this paper, we present a new methodology that aims to improve our understanding of population dynamics by considering the joint dynamics of individual locations and their neighbours. Our methodology breaks away from the traditional urban/rural dichotomy, avoiding rigid classifications or thresholds and overcoming the typical static picture of spatial population density. Using a fine grid of Italian locations and the municipalities they belong to, we reveal nuanced and complex patterns of urbanisation, densification and depopulation well within the municipal level. We provide a more comprehensive insight into overall population dynamics by taking a prospective view, where each location is classified according to its expected characteristics in the long run, i.e. its population density and that of its neighbours.

Although our approach is primarily descriptive, it has the advantage of being based solely on the time series of population density and the definition of neighbouring areas. From a policy point of view, even if the proposed methodology is not able to identify the underlying drivers of population density, it results in a powerful tool to improve the formulation of place-based policies by providing a more accurate classification based on the forecasted demographic dynamics in the long term. In this respect, when applied to Italian municipalities in order to validate the identification of project areas made in the *National Strategy for Inner Areas*, we find that the actual funding strategy is, on the whole, well designed. However, two concerns arise: i) insufficient funding for municipalities expected to become rural or suburban areas in the near future; and ii) a border effect that generates an excessive spatial fragmentation of funding in the presence of positive spatial spillovers, possibly due to the proposal of funding areas being made by NUTS-2 regions.[8] We also show that the same logic can be extended beyond the municipal scale. In Section 6.2, we build a forward-looking taxonomy for Italian NUTS 3 regions and benchmark it against Eurostat's urban-rural typology, highlighting how a trajectory-based classification can complement cross-sectional approaches by distinguishing "currently urban" regions from those undergoing sustained densification.

We conclude by emphasizing that our methodology can be applied to estimate the evolution of spatial patterns in a wide range of socio-economic phenomena characterized by spatial dependence, such as personal income, poverty, and social exclusion, and to assess the design of place-based policies aimed at reducing spatial inequality.

Declaration of 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.

[8]For details, see `https://politichecoesione.governo.it/it/strategie-tematiche-e-territoriali/strategie-territoriali/strategia-nazionale-aree-interne-snai/le-aree-interne-2014-2020/`.

References

Ahlfeldt, G. M. and Pietrostefani, E. (2019). The economic effects of density: A synthesis. *Journal of Urban Economics*, 111:93–107.

Arribas-Bel, D., Garcia-López, M.-À., and Viladecans-Marsal, E. (2021). Building (s and) cities: Delineating urban areas with a machine learning algorithm. *Journal of Urban Economics*, 125:103217.

Barca, F. (2009). *Agenda for a reformed cohesion policy*. European Communities Brussels.

Barca, F., Casavola, P., and Lucatelli, S. (2014). A strategy for inner areas in Italy: Definition, objectives, tools and governance. Materiali UVAL-Documenti: Vol. 31: UVAL.

Barca, F., McCann, P., and Rodríguez-Pose, A. (2012). The case for regional development intervention: place-based versus place-neutral approaches. *Journal of Regional Science*, 52(1):134–152.

Bayoh, I., Irwin, E. G., and Haab, T. (2006). Determinants of residential location choice: How important are local public goods in attracting homeowners to central city locations? *Journal of Regional Science*, 46(1):97–120.

Bencardino, M. (2015). Demographic changes and urban sprawl in two middle-sized cities of campania region (italy): A measurement in the cities of benevento and avellino. In *International Conference on Computational Science and Its Applications*, pages 3–18. Springer.

Bhatta, B. (2010). *Analysis of urban growth and sprawl from remote sensing data*. Springer Science & Business Media.

Bowman, A. W. and Azzalini, A. (1997). *Applied smoothing techniques for data analysis: the kernel approach with S-Plus illustrations*, volume 18. OUP Oxford.

Brezzi, M., Dijkstra, L., and Ruiz, V. (2011). OECD extended regional typology. *OECD Regional Development Working Papers, No. 2011/06, OECD Publishing, Paris.*

Brezzi, M. and Veneri, P. (2015). Assessing polycentric urban systems in the oecd: Country, regional and metropolitan perspectives. *European planning studies*, 23(6):1128–1145.

Cattaneo, M., Morlotti, C., Malighetti, P., and Redondi, R. (2023). Airports and population density: where benefits outweigh costs. *Regional Studies*, 57(3):576–589.

Dąbrowski, M. (2014). Eu cohesion policy, horizontal partnership and the patterns of sub-national governance: Insights from central and eastern europe. *European Urban and Regional Studies*, 21(4):364–383.

Di Feliciantonio, C. and Salvati, L. (2015). "southern" alternatives of urban diffusion: Investigating settlement characteristics and socio-economic patterns in three m editerranean regions. *Tijdschrift voor economische en sociale geografie*, 106(4):453–470.

Duranton, G. (2021). Classifying locations and delineating space: An introduction. *Journal of Urban Economics*, 125:103353.

EUROSTAT (2024). Methodological manual on territorial typologies — 2024 online edition, `https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Territorial_typologies_manual`.

Fiaschetti, M., Graziano, M., and Heumann, B. W. (2021). A data-based approach to identifying regional typologies and exemplars across the urban–rural gradient in europe using affinity propagation. *Regional Studies*, 55(12):1939–1954.

Fiaschi, D., Gianmoena, L., and Parenti, A. (2018). Spatial club dynamics in European regions. *Regional Science and Urban Economics*, 72:115–130.

Fioretti, C., Pertoldi, M., Busti, M., and Van Heerden, S. (2020). Handbook of sustainable urban development strategies. Technical report, Joint Research Centre (Seville site).

Gerolimetto, M. and Magrini, S. (2022). Distribution dynamics: a spatial perspective. *Spatial Economic Analysis*, pages 1–25.

Glaeser, E. L. (2010). *Agglomeration economics*. University of Chicago Press.

Gower, J. C. (1971). A general coefficient of similarity and some of its properties. *Biometrics*, pages 857–871.

Kaufman, L. and Rousseeuw, P. J. (1990). *Finding Groups in Data: An Introduction to Cluster Analysis*. John Wiley & Sons, New York.

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

Krugman, P. (1994). Complex landscapes in economic geography. *The American Economic Review*, 84(2):412–416.

Krugman, P. (1996). *The self organizing economy*. John Wiley & Sons.

Krzanowski, W. J. and Lai, Y. (1988). A criterion for determining the number of groups in a data set using sum-of-squares clustering. *Biometrics*, pages 23–34.

McCann, P. (2013). *Modern urban and regional economics*. Oxford University Press, USA.

Mendez, C. (2013). The post-2013 reform of eu cohesion policy and the place-based narrative. *Journal of European Public Policy*, 20(5):639–659.

Mendez, C., Van Der Zwet, A., and Borkowska-Waszak, S. (2021). Rescaling urban development policy in the eu: the impact of integrated place-based approaches in cohesion policy. *Regional Studies*, 55(6):1154–1165.

Milanovic, B. (2018). Towards an explanation of inequality in premodern societies: the role of colonies, urbanization, and high population density. *The Economic History Review*, 71(4):1029–1047.

Moreno-Monroy, A. I., Schiavina, M., and Veneri, P. (2021). Metropolitan areas in the world. delineation and population trends. *Journal of Urban Economics*, 125:103242.

Nechyba, T. J. and Strauss, R. P. (1998). Community choice and local public services: A discrete choice approach. *Regional Science and Urban Economics*, 28(1):51–73.

Overman, H. G., Rice, P., and Venables, A. J. (2010). Economic linkages across space. *Regional Studies*, 44(1):17–33.

Pateman, T. (2011). Rural and urban areas: comparing lives using rural/urban classifications. *Regional trends*, 43:11–86.

Pesaresi, M., Schiavina, M., Politis, P., Freire, S., Krasnodębska, K., Uhl, J. H., Carioli, A., Corbane, C., Dijkstra, L., Florio, P., et al. (2024). Advances on the global human settlement layer by joint assessment of earth observation and population survey data. *International Journal of Digital Earth*, 17(1):2390454.

Pham, D. T., Dimov, S. S., and Nguyen, C. D. (2005). Selection of k in k-means clustering. *Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science*, 219(1):103–119.

Pike, A., Rodríguez-Pose, A., and Tomaney, J. (2016). *Local and regional development*. Routledge.

Quah, D. T. (1996). Twin peaks: growth and convergence in models of distribution dynamics. *The economic journal*, 106(437):1045–1055.

Quah, D. T. (1997). Empirics for growth and distribution: stratification, polarization, and convergence clubs. *Journal of economic growth*, 2:27–59.

Romano, B. and Zullo, F. (2016). Half a century of urbanization in southern european lowlands: a study on the po valley (northern italy). *Urban Research & Practice*, 9(2):109–130.

Rousseeuw, P. J. (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, 20:53–65.

Sarno, E. (2021). The silent italian demographic crisis and the issue of south italy. *REVIEW OF HISTORICAL GEOGRAPHY AND TOPONOMASTICS*, 16:77–86.

Schirmer, P. M., Van Eggermond, M. A., and Axhausen, K. W. (2014). The role of location in residential location choice models: a review of literature. *Journal of Transport and Land Use*, 7(2):3–21.

Scott, A., Gilbert, A., and Gelan, A. (2007). *The urban-rural divide: Myth or reality?* Citeseer.

Silverman, B. W. (1986). *Density estimation for statistics and data analysis*, volume 26. CRC press.

Van Eupen, M., Metzger, M., Pérez-Soba, M., Verburg, P., Van Doorn, A., and Bunce, R. (2012). A rural typology for strategic european policies. *Land use policy*, 29(3):473–482.

Verburg, P. H., van Berkel, D. B., van Doorn, A. M., van Eupen, M., and van den Heiligenberg, H. A. (2010). Trajectories of land use change in europe: a model-based exploration of rural futures. *Landscape ecology*, 25:217–232.

Wang, J. (2010). Consistent selection of the number of clusters via crossvalidation. *Biometrika*, 97(4):893–904.

Wilson, P. A. (1995). Embracing locality in local economic development. *Urban Studies*, 32(4-5):645–658.

A Technical details on the estimation of RVF and forecast

In this appendix, we detail the technicalities behind estimating an RVF and its use to compute the forecast of spatial distribution.

A.1 The adaptive kernel estimator with full covariance matrix

The RVF reported in Figure 2 is based on weights computed from the joint density of (Y_{t_0}, WY_{t_0}), estimated using an *adaptive kernel method* and keeping into account the covariance of observations, as suggested by Silverman (1986). In particular:

1. run the *pilot density estimate*, with bandwidth h in evaluation point $\mathbf{z} \in \mathbb{R}^2$, defined as:

$$\tilde{f}_{Y_{t_0},WY_{t_0}}(\mathbf{z}) := \frac{(\det S)^{-1/2}}{Nh^2} \sum_{j=1}^{N} k\left(\frac{(\mathbf{z}-\mathbf{z}_{j,t_0})^T S^{-1}(\mathbf{z}-\mathbf{z}_{j,t_0})}{h^2}\right),$$

   where $k(\mathbf{z}^T\mathbf{z}) = K(\mathbf{z})$, K is a kernel function over $\mathbb{R}^2$ and S is the sample covariance matrix of observations.

2. Calculate the *local bandwidth factors* λ_j as

$$\lambda_j := \left(\frac{\tilde{f}_{Y_{t_0},WY_{t_0}}(\mathbf{z}_{j,t_0})}{g}\right)^{-\alpha},$$

   where g is the geometric mean of $\tilde{f}_{Y_{t_0},WY_{t_0}}(\mathbf{z}_{j,t_0})$, i.e.

$$g := \exp\left(\frac{1}{N}\sum_{j=1}^{N} \tilde{f}_{Y_{t_0},WY_{t_0}}(\mathbf{z}_{j,t_0})\right).$$

   The parameter $\alpha \in (0,1)$ is the sensitivity parameter to local density, i.e. the strength of adaptation of the estimate to local density of observations ($\alpha = 0$ no adaptation) (Silverman, 1986).

3. Finally, run the *adaptive kernel density estimator*:

$$\hat{f}^{a}_{Y_{t_0},WY_{t_0}}(\mathbf{z}) := \frac{(\det S)^{-1/2}}{N} \sum_{j=1}^{N} \frac{1}{h^2\lambda_j^2} k\left(\frac{(\mathbf{z}-\mathbf{z}_{j,t_0})^T S^{-1}(\mathbf{z}-\mathbf{z}_{j,t_0})}{h^2\lambda_j^2}\right).$$

In our case, the adaptive approach is really effective since observations are spread very far apart from the median, with regions displaying a high number of observations and others with only few observations. Therefore, calculate the *RVF adaptive estimator* $\widehat{\mathbf{RVF}}_{\tau}^{adp}$ as

$$\widehat{\mathbf{RVF}}_{\tau}^{adp}(\mathbf{z}) := \sum_{j=1}^{N} \omega_{j,t_0}^{adp}(\mathbf{z})\boldsymbol{\delta}_{j,\tau}, \tag{2}$$

where:

$$\omega_{j,t_0}^{adp}(\mathbf{z}) := \frac{\frac{(\det S)^{-1/2}}{Nh^2\lambda_j^2} k\left(\frac{(\mathbf{z}-\mathbf{z}_{j,t_0})^T S^{-1}(\mathbf{z}-\mathbf{z}_{j,t_0})}{h^2\lambda_j^2}\right)}{\hat{f}^{adp}_{Y_{t_0},WY_{t_0}}(\mathbf{z})}.$$

A.2 The choice of optimal estimation parameters h and α

The choice of parameters α and h of Eq. (2) is made by minimizing the mean square error (MSE) of the estimate for a set of candidate couples. In particular, for each possible couple of (α, h) estimate the adaptive RVF, $\widehat{\mathbf{RVF}}_{\tau}^{adp}(\mathbf{z})$; then, use the estimated RVF, given the initial observations, to predict their values in the final year, i.e.

$$\hat{\mathbf{z}}_{j,t_0+\tau} := \Phi^{\tau}_{\widehat{\mathbf{RVF}}_{\tau}^{adp}}(\mathbf{z}_{j,t_0}), \tag{3}$$

where we refer to Section A.3 for the details of function $\Phi^{\tau}_{\widehat{\mathbf{RVF}}_{\tau}^{adp}}$ which predicts the observations in the last year given the ones in the first year. Finally, the goodness of the estimation is measured by the mean of the square distance between the predicted observations $\hat{\mathbf{z}}_{j,t_0+\tau}$ and the actual observation at the final year $\mathbf{z}_{j,t_0+\tau}$, i.e.:

$$MSE(\alpha, h) := \frac{1}{N} \sum_{j=1}^{N} ||\hat{\mathbf{z}}_{j,t_0+\tau} - \mathbf{z}_{j,t_0+\tau}||^2.$$

In practice, the optimal couple of (α, h), i.e. the one minimizing $MSE(\alpha, h)$, is searched over a finite grid, which in our case consists of all their possible combinations in the sets $h = \{0.04, 0.06, 0.08, 0.11, 0.15, 0.21, 0.29, 0.41, 0.57, 0.80\}$ and $\alpha = \{0, 0.0005, 0.0012, 0.0028, 0.0067, 0.0158, 0.0375, 0$ for a total of 100 estimated RVFs. The choice of the grid follows the methodology suggested by Bowman and Azzalini (1997). The optimal combination results $(\alpha^\star, h^\star) = (0.0067, 0.21)$.

A.3 The forecast via RVF in continuos time

Differently from Fiaschi et al. (2018), forecast is made in continuous time by function $\Phi^\tau_{\widehat{\mathbf{RVF}}_\tau^{adp}}$ in Eq. (3). Given the vector field $F : \mathbb{R}^2 \to \mathbb{R}^2$ and an observation j at the initial year $\mathbf{z}_{j,t_0} \in \mathbb{R}^2$, first calculate the trajectory that observation j, $\mathbf{z}_{j,t_0}$, will have following the *vector field* F. In particular, introduce the differential equation for the variable $p^{z_{j,t_0}}$ as follows:

$$\begin{cases} \dot{p}^{z_{j,t_0}} = F(p^{z_{j,t_0}}); \text{and} \\ p^{z_{j,t_0}}(t_0) = z_{j,t_0}. \end{cases} \tag{4}$$

Roughly speaking the curve $p^{z_{j,t_0}} : [t_0, +\infty) \to \mathbb{R}^2$ represents the trajectory of observation j, which starts at z_{j,t_0} at time t_0 and follows the direction prescribed by F; $\Phi^\tau_F(z_{j,t_0})$ is therefore defined as:

$$\Phi^\tau_F(z_{j,t_0}) := p^{z^{j,t_0}}(t_0 + \tau) \in \mathbb{R}^2. \tag{5}$$

In other words, $\Phi^\tau_F(z_{j,t_0})$ is the value at time $t_0 + \tau$ of an observation j whose value at time t_0 was z_{j,t_0}, and has followed the arrows prescribed by the vector field F. In our model, $\hat{\mathbf{z}}_{j,t_0+\tau}$ of Eq. (3) represents the position in Moran space that observation j will reach following the estimated RVF for τ periods.

In practice, to numerically solve Eq. (4) we have to compute F in every point of the trajectory of $p^{z_{j,t_0}}$. However, in our case the vector field to predict the dynamic $\widehat{\mathbf{RVF}}_\tau^{adp}$ can only be estimated on a finite set of evaluation points. Therefore, we rely on an interpolation procedure: given a finite regular grid of evaluation points where $\widehat{\mathbf{RVF}}_\tau^{adp}$ has been estimated, and for each point $\mathbf{z}$ for which we have to evaluate $\widehat{\mathbf{RVF}}_\tau^{adp}(\mathbf{z})$ (assuming $\mathbf{z}$ itself is not one of the evaluation points), first consider the four evaluation points defining the corners of the square where $\mathbf{z}$ belongs to. Then, the value of $\widehat{\mathbf{RVF}}_{\tau}^{adp}(\mathbf{z})$ is approximated by linearly interpolating the values of the estimated RVF in the four corners. This produces an approximated estimated RVF in every point $\mathbf{z} \in \mathbb{R}^2$, thus allowing to numerically solve Eq. (4).

B The procedure for the taxonomy of locations

The location-classification procedure is grounded in the dynamics of the Moran space. We begin by constructing a fine grid of 2,500 (50×50) evenly spaced evaluation points spanning the full range of observed values in the Moran space. We then retain a subset of 1,542 points, namely those for which a direction of motion can be estimated from the observed dynamics and, therefore, for which the RVF is defined. The procedure consists of the following steps:

1. For each evaluation point, using the estimated RVF, we simulate the corresponding trajectory forward for 10,000 years to identify the long-run equilibrium associated with that initial condition.

2. The resulting set of long-run equilibria is partitioned into an *optimal* number of clusters ($K = 5$) based on five selection criteria: (i) Pham et al. (2005) (PDM), implemented with k-means; (ii) Krzanowski and Lai (1988) (KL), implemented with k-medians; (iii) Rousseeuw (1987) (Silhouette), implemented with k-medians; (iv) Wang (2010) (WANG), implemented with k-medians; and (v) Kaufman and Rousseeuw (1990), implemented with the k-medoids (PAM) algorithm. These clusters define *attraction regions* in the Moran space.

3. Each evaluation point is then assigned to the attraction region of the equilibrium to which its simulated trajectory converges. This assignment defines the corresponding *basin of attraction* for each attraction region.

We compute the centroids of each attraction region and of its associated basin of attraction. Figure 3a reports these centroids and connects each pair (attractor vs. basin) with an arrow.

Figure 7 reports the detailed outcomes of the five selection criteria. Specifically, Pham et al. (2005), using the standard threshold of 0.85, indicates five clusters; Krzanowski and Lai (1988), excluding the two-cluster solution, suggests six clusters; Rousseeuw (1987) suggests five clusters; Wang (2010) suggests four clusters; and PAM (Kaufman and Rousseeuw, 1990) points to five or six clusters. Overall, the choice of five clusters appears to be the most strongly supported. Figure 8 shows that adopting four clusters instead of five yields only minor differences for the dynamics at locations corresponding to observed points.

[Figure 7 about here.]

[Figure 8 about here.]

C Population and population density of Italian municipalities in Figure 4

Table 4 reports the population density and the total population in 1985 and 2020 for the 9 Italian municipalities of Figure 4.

[Table 4 about here.]

List of Figures

1 The maps of population density (residents per km^2) in 2020, the local annual growth rate over the period 2020-1985, and the latter versus the population density in 1985 for the sample of 82,536 Italian locations. 35
2 The estimated Random Vector Field (RVF) in the Moran space, based on Eq. (1), is shown for the period 1985–2020. The red arrows represent the estimated directions and the intensity of the movement (the length of each arrow has been divided by a factor of 5 for clarity). The grey area indicates where RVF is calculated. In 1985 and 2020, level curves of the densities are coloured in yellow and green, respectively. The blue and purple lines depict the estimated non-parametric Moran curve, along with their 95% confidence bands for 1985 and 2020. Dashed lines represent the sample averages of the population density of locations and of municipalities to which they belong. 36
3 The estimated RVF and the five types of locations identified by using the classification algorithm described in Appendix B: Type 1 (Urban Periphery Increasing Density locations, light brown); Type 2 (Urban Core Increasing Density locations, dark brown); Type 3 (Sparse Neighbourhood Decreasing Density locations, aquamarine); Type 4 (Remote locations, dark green); and Type 5 (Sparse Pocket locations, gold). 37
4 Types of location within the Italian municipalities: Rome, Milan, Modena, Genoa, L'Aquila, Grosseto, Sassari, Brindisi and Pomarance (thick black lines represent municipal borders). Location taxonomy: Type 1 (Urban Periphery Increasing Density locations, light brown); Type 2 (Urban Core Increasing Density locations, dark brown); Type 3 (Sparse Neighbourhood Decreasing Density locations, aquamarine); Type 4 (Remote locations, dark green); and Type 5 (Sparse Pocket locations, gold) 38
5 Maps of the National Strategy for Inner Areas (SNAI) and our proposed taxonomy of municipalities. 39
6 The maps of Eurostat's urban–rural typology classification (2024) and our proposed taxonomy of NUTS 3 regions. 40
7 The outcome of the five algorithms for the choice of the optimal number of clusters. 41
8 The comparison between four and five clusters in the RVF. 42

Figure 1: The maps of population density (residents per km^2) in 2020, the local annual growth rate over the period 2020-1985, and the latter versus the population density in 1985 for the sample of 82,536 Italian locations.

map_popDensAggrCells_2020.pdf

(a) 2020 (black points represent regional capitals).

mapDiffLog_AggrCells_1985_2020.pdf

(b) The annual local growth rate of population (black points represent regional capitals).

growthPath.pdf

(c) The estimated nonlinear relationship between the annual population growth rate and population density in 1985 is made by the Nadaraya-Watson estimator, with 95% confidence bands indicated by dotted lines.

Source: our elaborations on Global Human Settlement Layer Population Grid (GHS-POP) data.

Figure 2: The estimated Random Vector Field (RVF) in the Moran space, based on Eq. (1), is shown for the period 1985–2020. The red arrows represent the estimated directions and the intensity of the movement (the length of each arrow has been divided by a factor of 5 for clarity). The grey area indicates where RVF is calculated. In 1985 and 2020, level curves of the densities are coloured in yellow and green, respectively. The blue and purple lines depict the estimated non-parametric Moran curve, along with their 95% confidence bands for 1985 and 2020. Dashed lines represent the sample averages of the population density of locations and of municipalities to which they belong.

RVF_1985-2020_GHSL_wMun_Revised.pdf

Source: our elaborations on Global Human Settlement Layer Population Grid (GHS-POP) data.

Figure 3: The estimated RVF and the five types of locations identified by using the classification algorithm described in Appendix B: Type 1 (Urban Periphery Increasing Density locations, light brown); Type 2 (Urban Core Increasing Density locations, dark brown); Type 3 (Sparse Neighbourhood Decreasing Density locations, aquamarine); Type 4 (Remote locations, dark green); and Type 5 (Sparse Pocket locations, gold).

fiveClustersOverEvalPointsKmedians.pdf

(a) The five basins of attraction, their centroids (black squares), the centroids of the regions of attraction (black rhombs), with the estimated RVF (red arrows).

map_CellsInClusters_noBordi.pdf

(b) The map of Italian locations coloured on the basis of their type.

Source: our elaborations on Global Human Settlement Layer Population Grid (GHS-POP) data.

Figure 4: Types of location within the Italian municipalities: Rome, Milan, Modena, Genoa, L'Aquila, Grosseto, Sassari, Brindisi and Pomarance (thick black lines represent municipal borders). Location taxonomy: Type 1 (Urban Periphery Increasing Density locations, light brown); Type 2 (Urban Core Increasing Density locations, dark brown); Type 3 (Sparse Neighbourhood Decreasing Density locations, aquamarine); Type 4 (Remote locations, dark green); and Type 5 (Sparse Pocket locations, gold)

roma.pdf

(a) Rome

milan.pdf

(b) Milan

modena.pdf

(c) Modena

genova.pdf

(d) Genoa

Laquila.pdf

(e) L'Aquila

grosseto.pdf

(f) Grosseto

sassari.pdf

(g) Sassari

brindisi.pdf

(h) Brindisi

pomarance.pdf

(i) Pomarance (centre of Tuscany)

Source: our elaborations on Global Human Settlement Layer Population Grid (GHS-POP) data.

Figure 5: Maps of the National Strategy for Inner Areas (SNAI) and our proposed taxonomy of municipalities.

`barca_map_Edited.pdf`

(a) Map of municipalities' taxonomy based on Barca et al. (2014) (Figure III.2 in (Barca et al., 2014, p. 26)).

`map_Mediodis.pdf`

(b) Map of municipalities' taxonomy based on the locations' classification of Section 5.

`map_areeStrat_financed.pdf`

(c) Map of municipalities financed in SNAI 2014-2020.

`map_SA_RA_financedCRICCISEIFICO.pdf`

(d) Map of urban/urban peripheral municipalities and rural/peripheral municipalities according to our taxonomy, partitioned between financed and not

Figure 6: The maps of Eurostat's urban–rural typology classification (2024) and our proposed taxonomy of NUTS 3 regions.

NUTS3_classification.png

(a) Map of NUTS 3 regions based on EUROSTAT (2024).

map_MediodisNUTS3.pdf

(b) Map of NUTS 3 regions taxonomy based on the locations' classification of Section 5.

Source: our elaborations on ISTAT data (Italian National Institute of Statistics).

optimalClusterCellsPhamDN05.pdf

(a) Pham et al. (2005)

optimalClusterKrzanowskiLaiKmedians.pdf

(b) Krzanowski and Lai (1988)

optimalClusterSilhouetteKmedians.pdf

(c) Rousseeuw (1987)

optimalClusterWangKmedians.pdf

(d) Wang (2010)

optimalClusterSilhouettePAM.pdf

(e) Kaufman and Rousseeuw (1990)

Figure 7: The outcome of the five algorithms for the choice of the optimal number of clusters.

fiveClustersOverEvalPointsKmedians.pdf

(a) Five clusters in the RVF

fourClustersOverEvalPointsKmedians.pdf

(b) Four clusters in the RVF

Figure 8: The comparison between four and five clusters in the RVF.

List of Tables

1 The mean and median population density, the coefficient of variation (CV), the variance decomposition between and within municipalities, and the Moran's I index (standard error in parenthesis), where municipal membership is the criterion for assessing the extent of spatial spillovers, for the sample of 82,536 Italian locations. . 44
2 The number of locations, total area (in square kilometres), population, mean location density, and mean municipal density in 2020 for the five types of locations identified using the classification algorithm described in Appendix B. 45
3 Characteristics of NUTS 3 regions clusters 46
4 The total population and population density of the Italian municipalities of Figure 4 in 1985 and 2020. 47

Table 1: The mean and median population density, the coefficient of variation (CV), the variance decomposition between and within municipalities, and the Moran's I index (standard error in parenthesis), where municipal membership is the criterion for assessing the extent of spatial spillovers, for the sample of 82,536 Italian locations.

	Mean	Median	CV	Variance			Moran's I
				Total	Within (%)	Between (%)	
1985	215.9	21.1	3.8	675671.8	63.5	36.5	0.32^{***} (0.00188)
2020	227.5	20.6	3.6	678486.8	63.8	36.2	0.36^{***} (0.00192)

Source: our elaborations on Global Human Settlement Layer Population Grid (GHS-POP) data.

Table 2: The number of locations, total area (in square kilometres), population, mean location density, and mean municipal density in 2020 for the five types of locations identified using the classification algorithm described in Appendix B.

Type of location	Number of locations (%)	Area (%)	Population (%)	Mean locations pop. density	Mean municipal pop. density
Type 1 *Urban Periphery* *Increasing Density* (UPID)	17.1	16.8	11.1	126.3	469.2
Type 2 *Urban Core* *Increasing Density* (UCID)	15.7	15.5	69.6	850.2	402.3
Type 3 *Sparse Neighbourhood* *Decreasing Density* (SNDD)	43.7	44.1	19.2	82.3	65.5
Type 4 *Remote* (R)	23.1	23.1	0.017	0.14	72.3
Type 5 *Sparse Pocket* (SP)	0.5	0.5	0.003	0.99	517.1

Source: our elaborations on Global Human Settlement Layer Population Grid (GHS-POP) data.

Cluster	Mean pop 1985	Mean pop 2020	Mean density 1985	Mean density 2020
Low-density with projected depopulation	411.2	398.3	130.8	126.9
Medium-density with projected moderate densification	688.0	728.5	226.8	240.5
Urban with projected high densification	1075.0	1185.1	360.3	397.6

Table 3: Characteristics of NUTS 3 regions clusters

Table 4: The total population and population density of the Italian municipalities of Figure 4 in 1985 and 2020.

Municipality	Population in 1985	Population in 2020	Density in 1985	Density in 2020
Rome	2,507,384	2,743,434	1,988.38	2,175.57
Milan	1,352,562	1,246,732	7,847.46	7,233.44
Modena	175,017	189,941	937.51	1,017.45
Genoa	644,334	507,808	2,742.95	2,161.75
L'Aquila	70,735	60,124	291.16	247.48
Grosseto	66,099	74,626	152.33	171.98
Sassari	123,012	127,322	232.27	240.41
Brindisi	90,589	88,236	315.15	306.97
Pomarance	6,428	5,373	29.87	24.96

Source: our elaborations on Global Human Settlement Layer Population Grid (GHS-POP) data.