Journal of Open Innovation: Technology, Market, and Complexity

MDPI

Article

Spatio-Temporal Dynamics of European Innovation—An Exploratory Approach via Multivariate Functional Data Cluster Analysis

Imke Rhoden [1,*], Daniel Weller [2] and Ann-Katrin Voit [3]

[1] Institute of Energy and Climate Research—Systems Analysis and Technology Evaluation, Forschungszentrum Jülich, 52425 Jülich, Germany
[2] Institute for International Law of Peace and Armed Conflict, Center for Research Methods, Ruhr-Universität Bochum, 44801 Bochum, Germany; daniel.weller@rub.de
[3] FOM Hochschule für Oekonomie und Management, 45127 Essen, Germany; ann-katrin.voit@rub.de
* Correspondence: i.rhoden@fz-juelich.de

Abstract: We apply a functional data approach for mixture model-based multivariate innovation clustering to identify different regional innovation portfolios in Europe, considering patterns of specialization among innovation types. We combine patent registration data and other innovation and economic data across 225 regions, 13 years, and eight patent classes. The approach allows us to form several regional clusters according to their specific innovation types and captures spatio-temporal dynamics too subtle for most other clustering methods. Consistent with the literature on innovation systems, our analysis supports the value of regionalized clusters that can benefit from flexible policy support to strengthen regions as well as innovation in a systematic context, adding technology specificity as a new criterion to consider. The regional innovation cluster solutions for IPC classes for 'fixed constructions' and 'mechanical engineering' are highly comparable but relatively less comparable for 'chemistry and metallurgy'. The clusters for innovations in 'physics' and 'chemistry and metallurgy' are similar; innovations in 'electricity' and 'physics' show similar temporal dynamics. For all other innovation types, the regional clustering is different. By taking regional profiles, strengths, and developments into account, options for improved efficiency of location-based regional innovation policy to promote tailored and efficient innovation-promoting programs can be derived.

Keywords: functional data analysis (FDA); innovation concentration; spatio-temporal cluster modeling; multivariate cluster analysis; European innovation; cluster algorithm

Citation: Rhoden, I.; Weller, D.; Voit, A.-K. Spatio-Temporal Dynamics of European Innovation—An Exploratory Approach via Multivariate Functional Data Cluster Analysis. *J. Open Innov. Technol. Mark. Complex.* **2022**, *8*, 6. https://doi.org/10.3390/joitmc8010006

Received: 15 November 2021
Accepted: 22 December 2021
Published: 5 January 2022

1. Introduction

Innovation is a key driver of Europe's sustainable economic success. The topics of innovation, geography, clustering and their interdependencies can be investigated by a variety of approaches. If innovation is considered in the context of geography and economic growth, there is no single theoretical framework, as there are too many interlinkages between these topics to find a universal approach [1]. Thus, there are multiple schools of thought regarding the temporal and spatial evolution of innovative activity [2]. In this paper, we focus on approaches related to investigating innovation clustering such as that of Fornahl and Brenner [3], who find that heterogeneous types of innovation cluster differently, which points to the relevance of considering innovation as a differentiated subject. Knowledge spillovers are another link between innovation and geography to consider, as knowledge (tacit or understanding) is often only transferred locally or regionally. Innovation is thus prone to spillover, as research shows (e.g., McCann and Simonen [4], Costantini et al. [5], and Aldieri et al. [6]). This is confirmed by Bottazzi and Peri [7], who correlate data on research and development (R&D) and patents, finding that R&D spending can increase innovation output, but limited to a local scale. Giannitsis and Kager [8] analyze links between technology and specialization as these can determine market positions and competitive success. Thus, it is vital to know how static and dynamic conditions interact and how they contribute to the emergence of innovation. They note that it is important for policy to adapt effectively and in a timely fashion to changing circumstances, as technical specification can drive industry, and thus competitive advantages. Here, policy can promote progress through innovation.

There are various national and supranational approaches for delineating innovation profiles and analyzing the European innovation landscape on several levels, the most renowned being the European Innovation Scoreboard (EIS) and its regional equivalent (RIS). Ranking regions according to their innovation strength is important for identifying and analyzing the characteristics of innovation leaders, so conditions in lagging regions can be improved, e.g., via regional or innovation policy [9,10].

Capello and Lenzi [11] search for patterns of knowledge, attitudes, and innovation behaviors in innovative European regions utilizing a cluster analysis. They cluster the degree of knowledge and innovation that the selected regions produce, taking into account the different stages of the innovation process. Above all, the results indicate that policy measures at a regional level are useful and necessary, as innovation trajectories diverge due to regional characteristics. Innovation, the authors propose, is much more complex than just the divide between agglomerated and peripheral regions. Moreover, they suggest policies that are closely oriented towards the respective clusters and their specific innovation patterns, leading to a "smart" Europe.

Spielkamp and Vopel [12] explicitly combine innovation systems and cluster theory to find innovation clusters in Germany. They assume the existence of agents in technological environment networks to create, use, and diffuse technology. By linking this view with other innovation variables they found a system of innovation and of firms emerging that leads to certain patterns. Furthermore, they emphasize that due to the extremely high complexity of innovation systems, a multitude of approaches are possible. Several variables are used in their clustering approach, the most important of which are innovation, knowledge, information, and industry characteristics.

Common among these approaches is the fact that innovation should not be considered without a spatial component, nor a temporal component. Turkina and Van Assche [13] examine innovation performance in clusters and find that linkages along the horizontal and vertical supply chain are key to increasing knowledge intensity and thus innovation. Pełka [14] analyzes innovation clusters using symbolic density-based ensemble clustering, taking into account innovation policy. They investigate European countries and use the Regional Innovation Scoreboard as well as multiple innovation and other indicators. They calculate clusters with standard methodologies (e.g., k-means) and investigate the heterogeneity of the clusters. The result is a ranking of innovation leadership.

Ionela-Andreea and Marian [15] use data from the European Patent Office and calculate the Malmquist index for total factor productivity in knowledge performance [16]. They also identify differences and similarities in the development of innovation capacities between the resulting clusters. Zabala-Iturriagagoitia et al. [17] investigate the increasing territorial disparities in Europe using production theory and also apply the Malmquist index. They note that advances in innovation are not necessarily synonymous with technological progress and that there is no innovation convergence via which lagging regions can catch up with leading regions.

Pelau and Chinie [18] conduct a cluster analysis of European regions focusing on innovation and sustainable development, linking innovation and sustainability for an improved economic growth process. They use a static multivariate analysis to characterize regional clusters and find three major innovation-sustainability clusters ranked by degree of achievement. They also relate their approach to the literature on innovation systems, emphasizing the importance of the regional context of innovation. Kim and Bae [19] apply clustering as a step in forecasting potentially promising technology. Based on the information contained in classified patents, which can indicate the technologies involved in development, they find technology-specific clusters. Their aim is then to derive potential trends for developing technologies based on the clusters.

As innovation is a highly complex matter, it is crucial to focus on regional innovation profiles and align policy programs with regional characteristics, as there are several and significant difficulties in targeting an increase in innovation activity and possibly resulting economic growth. Furthermore, there is a need to look past national levels and investigate regional strengths and weaknesses for a more efficient adaptation of policy mixes, as policies are often not able to address regional needs [20]. Insights into specific branches of innovation via patent analysis, supported by the inclusion of further knowledge indicators and regional characteristics, can provide levers for improving policies, thus harnessing not only innovation potentials but also regional potentials of a European cohesion policy [21]. Thorough investigations and a precise understanding of the different types of innovation, their place of inception, and their evolution over time are crucial for aiding Europe's path to a sustainable economic future.

We apply a mixture model-based clustering analysis for multivariate functional data proposed by Schmutz et al. [22] to explore the spatio-temporal dynamics of European regional innovation activities and uncover groups of regions with homogeneous innovation profiles. To achieve this, the analysis is based on the functional data analysis paradigm (FDA), which allows us to analyze latent functional forms, inherent dynamics, and other features in time series of multiple innovation indicators too subtle to be captured by classical time series or clustering approaches. As innovation is a heterogenous phenomenon, we use several time series for main patent classes as proxies for innovation activity as well as other closely related indicators to generate individual innovation profiles for Europe's regions.

To our knowledge, a functional data analysis for multivariate innovation clusters taking into account different innovation types, multiple measures and temporal and spatial dimensions, which allows for a very detailed investigation of innovation profiles, has not been conducted so far. We will illustrate the procedure in the following chapters. First, a theoretical overview of relevant innovation literature concerning general principles and related approaches for identifying innovation clusters and the role of policy is given. The second section describes the time series data used for the statistical analysis, the general principles of the functional data paradigm and the mixture model-based multivariate functional clustering algorithm applied. A presentation of the clustering results is given in the third section before the paper concludes with a discussion in the fourth section.

1.1. Innovation Theory

In economics, the spatial dimension has played an increasingly important role since the beginning of the 1990s, as the publication of Krugman's 'Geography and Trade' [23] broadened the economic view for a better understanding of the global economy through its spatial dimension.

While some countries may experience lower growth or investment rates, other countries may suffer from higher unemployment rates. Krugman linked this globally and explained that global competition leads to more challenges that need to be considered. To have a sufficient number of qualified jobs for their population, countries have to consider their advantages or disadvantages due to location and innovation. In modern economies, more companies are forced to export their products and are therefore subject to a higher level of international competitiveness. To manage this successfully, companies are likely to settle into industry clusters and may consider relocating to gain from location advantages.

Krugman also shows the crucial role of innovation in economies by saying '(t)he more you know, the more you can learn' [24]. He pointed out that countries must find regional strengths and weaknesses to ensure their success, which depends to a large extent on the development of innovation-promoting structures.

Accordingly, spatial factors and innovation activities are closely connected as, e.g., seen in the Silicon Valley area in California, USA [25]. This leads to the assumption that regional factors influence the operational innovation process, as these processes have a geographic origin and are thus spatially differentiated. Spatial proximity promotes innovation-relevant interactions and, as a result, regional innovation and technology policy support measures are effective [26].

Krugman also considers transport costs affecting regional growth rates [26] as producers and consumers make spatial decisions based on prices and revenues to optimize their profits. Consequently, producers will try to increase profits by minimizing the costs of transportation, which can lead to a relocation of businesses. This logic also applies to the end-users, as they will relocate their demand to the regions with the lowest transport costs. As companies are more likely to produce most efficiently where their required production factors are sufficiently available, they have a latent incentive to locate in their customer's geographical vicinity. This will result in low transport costs for both parties, the supply side, and the demand side. However, it has to be considered that transport costs have been significantly reduced by transportation technology and have consequently become less important in many industries in recent years [27].

In terms of spatial clustering, agglomeration effects are of crucial relevance as several advantages as well as disadvantages can arise from localized concentrations of companies. To reach optimal levels of competitiveness within clusters, both internal and external agglomeration effects have to be differentiated and considered. Internal agglomeration effects are also known as economies of scale and provide advantages reducing fixed costs by the production of larger quantities of goods, while external agglomeration effects refer to the proximity of companies in the same value chain. As clearly shown by [28], companies with similar activities can profit from sharing access to skilled labor-by-labor pooling, sharing inputs from common suppliers, and benefiting from knowledge spillovers. Thus, companies can maximize their profits by shared use of workers, infrastructure, services and information. If a company has access to all its key resources in its vicinity, it can experience a competitive advantage [29]. Potential disadvantages of agglomeration effects can arise in form of higher environmental pollution [30], increasing property prices, higher competition in the local area, or overstrained infrastructure.

Further effects of agglomeration are localization and urbanization [31], where localization effects can be described as advantages arising from a company's proximity to other companies of its industry. Those advantages can be, e.g., an industry-relevant job market in the area, R&D facilities, and therefore patenting activity or the emergence of a specialized supply industry [32]. The urbanization effects develop over time as different industries lead to more infrastructure as well as urbanization of affected areas, and therefore to an increase in economic activity in general [31].

However, the analysis is limited due to other factors, e.g., psychological factors such as security, social factors such as welfare, tax, and subsidies, education and overall trends [33,34]. Even if governments try to encourage specific regions to grow or develop, there will be no guarantee of sustainable success in innovation, but firms are considerably more likely to benefit than in an environment without supporting structures [35]. The economic importance of innovation in general and in the regional economies of Europe in particular is highlighted due to the benefits of innovation to a system, whether through patents as a measure of analysis, or through other innovation indicators. To be globally competitive in the long run, a high level of innovation activity within countries or regions is required, which policy can support by promoting approaches to assessing technological progress. Funding R&D, promoting open innovation, and the successful use of intellectual property rights are important steps [35]. This paper shows the importance of assessing innovation in a differentiated way and links data to innovation agglomeration and the systems theory behind it, which complements the views of Krugman [23] and Cooke [36]. Furthermore, it underlines that regional clusters are not only important for facilitating knowledge and innovation transfer [37], but that clusters differ spatially in terms of technology.

1.2. Europe and Innovation Support

As some regions in Europe are highly successful and others are lagging behind, theories of innovation can provide means to identify and understand these disparities in terms of employment, infrastructure, availability of services, and economic success in general. Often disparities arise due to the unequal distribution of natural resources, decisions of the public sector, or other locational reasons. Within metropolitan areas, disparities can be seen between central and peripheral areas with peripheral areas increasing faster in value than the centers. Therefore, migration and relocations from centers to suburban areas can be seen, which can potentially lead to the devaluation of certain areas and neighborhoods, thus creating social inequalities [38].

In the European Union, regional development is a key factor for equal living conditions in and between the member states. Therefore, the EU targets economic development via spatial planning, state planning, and regional planning. Particularly, the European Regional Development Fund (ERDF) and the Trans-European Networks (TEN) are means to detect national and regional disparities and to support the realization of equal living conditions. The newest European funding program and a key focus of the European Commission is the European Green Deal, which aims to fight environmental degradation and climate change, while simultaneously searching for new and sustainable growth strategies in order to be competitive in the future. The European Green Deal focuses on investing in environmentally friendly technologies, supporting innovation in industries, and introducing cleaner, cheaper, and healthier forms of private and public transport. In addition, decarbonizing energy production, ensuring higher energy efficiency of buildings, and improving global environmental standards via international cooperation are core values [39].

To achieve these goals, the EU must ensure a high level of labor skills as well as high levels of investment in R&D. Given this focus, it is vital to gain a precise understanding of the spatio-temporal dynamics within the European innovation system. The potential of innovation for mitigating climate change through new, efficient technologies that promote sustainable growth can help to avoid lock-in and turn innovation into 'green innovation' [40]. This applies not only to identifying the types and drivers of Member States' innovation strengths but also to investigating locational differences in innovation. More knowledge about the structure of innovation and its place of inception can be used to understand innovation emergence as well as its inherent geographical nature, provide insights into the success of policy programs and help structure future policy programs for a sustainable innovation climate in the EU.

Structures for innovation clusters, networks, and interrelations can benefit from open innovation, which can contribute to sustainable knowledge growth. The role of public institutions in supporting knowledge generation is particularly important in regional systems [36], where collaboration and networking facilitate innovation by increasing social capital [41–43]. Policy is needed where local actors are lacking in expertise concerning open innovation practices [44]. This is particularly important for small firms, as they rely heavily on spatial proximity to generate innovation, as shown by Leckel et al. [35]. Open innovation as a paradigm [45] can promote the creation of innovation and strengthen inter- and intra-regional knowledge exchange by emphasizing different regional characteristics [46].

2. Materials and Methods

In practice, several indicators can be used to approximate innovation, but we choose mainly to use patent data to indicate the type of innovation, with the patent classification scheme allowing a distinction between different types of inventions. The classification is based on the type of innovation group to which a patent belongs and must be indicated when filing an application. The patent classes we use are the eight major classes (A: 'human necessities'; B: 'performing operations and transporting'; C: 'chemistry and metallurgy'; D: 'textiles and paper'; E: 'fixed constructions'; F: 'mechanical engineering, lighting, heating, weapons, and blasting'; G: 'physics'; H: 'electricity' [47]). As suggested by Griliches [48] and noted by several other researchers, the inclusion of patents as an indicator of innovation is justified by the intentions pursued by filing a patent, i.e., an intended commercial use. Other innovation indicators are R&D personnel and researchers as well as internal R&D expenditures as a percentage of the gross domestic product. In addition, we use a human capital indicator approximated by human resources in science and technology. These variables are suitable to support patents as an indicator of innovation, as they are directly related to the emergence of innovation and can lead to patents or other forms of innovation.

In the model, variants of these variables are used. First, we compute Innovation Gini indicators according to Rhoden [49] for each IPC class, which provide a measure of the degree of innovation variation in regions. Then, we calculate the labor density and relate it to the regional GDP, the share of R&D labor, the human capital density, and the R&D expenditure per R&D labor. These measures are used to indicate labor productivity, human capital accumulation per worker, and R&D expenditure productivity. This step results in a set of five covariates that are included in the clustering process of the Innovation Ginis (see Table 1). In this way, multivariate spatio-temporal innovation dynamics of European regions can be aggregated into eight sets of clusters showing the similarities and differences of regional structures for each of the eight patent classes.

Table 1. Variable Declaration, source: Own calculations [50–52].

Variable	Description	Eurostat Datasets Used
Innovation Gini	Innovation Gini for the relevant IPC class of patents (patent applications to the EPO by priority year); Normalization Factor: Economically active population in 1000	PAT_EP_RIPC LFST_R_LFP2ACT
Labor Density	Economically active population per square kilometer in 1000	LFST_R_LFP2ACT DEMO_R_D3AREA
Share of R&D Labor	R&D personnel and researchers directly engaged in R&D per economically active population in 1000	RD_P_PERSREG LFST_R_LFP2ACT
GDP per Labor	Gross Domestic Product at current market prices in Billion Euro per economically active population in 1000	NAMA_10R_3GDP LFST_R_LFP2ACT
Human Capital Density	Human resources in science and technology (Persons with education in science and technology) per economically active population in 1000	HRST_ST_RCAT LFST_R_LFP2ACT
R&D Investment per R&D Labor	Internal R&D investment in Billion Euro per R&D personnel and researchers directly engaged in R&D in 1000	RD_E_GERDREG RD_P_PERSREG

As our analysis focuses on European regions, our dataset consists mainly of data from Eurostat [50] from 2000 to 2012, with supplements from other statistical offices and organizations (i.e., [51,52]) used for filling missing values in the main datasets after checking for plausibility. However, there are still large numbers of remaining missing values, which we choose to impute via natural spline interpolation using the annual cross-sections of our datasets as knots [53,54]. This imputation is applied when less than 30% of values for a region are missing and the pattern of missingness can be reasonably handled by spline interpolation, i.e., when there are enough values next to the missing values. Although this may seem like an arbitrary choice, sensitivity analyses have shown that this procedure strikes a more robust balance between the highest number of regions to cluster and the least amount of imputation bias compared to other approaches (e.g., Honaker and King [55]).

Spatially, we focus on the European regions at NUTS-2 level, which necessitates the creation of a custom reference, as several revisions of the NUTS classification were made over the periods covered by our data. This reference is based on NUTS 2016, which corresponds to most of our data but adopts NUTS 2010 regions where later revisions differ from the regions in our dataset. We also create a custom shapefile to correctly represent the statistical geographical level, which we then apply throughout our calculations. In total, we use 225 distinct regions in our mixture model-based multivariate functional cluster analysis. All computations are realized in the software R [56] using the packages *fda* [57] and *funHDDC* [22].

2.1. Functional Data Paradigm

Although the concept of functional data dates back to Grenander [58] and Rao [59], the actual term functional data for objects that can naturally be viewed as smooth curves rather than a set of discrete observations was coined by Ramsay [60], Ramsay and Dalzell [61] and Rice and Silverman [62]. In statistical terms, functional data are random variables usually observed at multiple discrete points on an infinite dimensional or functional continuum such as time, space, or other variables describing continua [63]. Accordingly, a set of functional variables for multiple observations is called the functional dataset. In line with Kokoszka and Reimherr [64], we refer to functional data as

$$X_n(t_{n,p}) \in \mathbb{R}^p;\ t_{n,p} \in [T_{min}, T_{max}];\ n = 1, \ldots, N;\ p = 1, \ldots, P.$$

In this notation, functional data are given by a set of N independent curves X_n observed in discrete sets of values $\{t_{n,p}, y_{n,p}\}$ along an interval $[T_{min}, T_{max}]$ over potentially infinite dimensions P. Functional data analysis can thus be performed not only with random curves but also p-dimensional random surfaces. In most fields of research, however, the focus is still on the analysis of curves, which is why the term curve data [62,65,66] is often used for the analysis of the special case of a one-dimensional continuum. A comprehensive review of the history of functional data analysis, its methods, and applications in different fields of research is given by Wang et al. [67].

In general, functional data are considered as independent and identically distributed samples from L_2-continuous stochastic processes whose mean and covariance estimators are given by $\hat{\mu}(t_p) = \frac{1}{n}\sum_{i=1}^{n} x_i(t_{n,p})$ and $\hat{v}(t_p) = \frac{1}{n-1}\sum_{i=1}^{n}(x_i(s_{n,p}) - \hat{\mu}(s_{n,p}))(x_i(t_{n,p}) - \hat{\mu}(t_{n,p}))$. As Deville [68] has shown, both estimators converge to $\mu(t_p)$ and $\nu(s_p, t_p)$ in L_2-norm, which is consistent with the assumption of a latent functional form in the form of smooth curves rather than mere sequences of observations as a basic principle of functional data analysis [69].

As crucial as smoothness may be for the analysis of functional data, it may not be obvious in raw datasets, as observations are often contaminated or distorted by random noise, measurement errors, or other types of bias [69]. These effects can be viewed as fluctuations in the smooth curves that we include by extending our earlier notion of functional data:

$$S_n(t_{n,p}) = X_n(t_{n,p}) + \epsilon_{n,p},$$

where $S_n(t_{n,p})$ is the realized and observable functional form and $\epsilon_{n,p}$ the representation of noise, disturbance, or error. We would like to refer to Ferraty and Vieu [63], Ramsay and Silverman [69], and Kokoszka and Reimherr [64] for a complete overview of the theoretical foundations of functional data analysis.

As our imputed data are still in their raw form, we use basis expansion to reconstruct their functional forms, which is necessary for any kind of functional data analysis [70]. Ideally, this basis function is similar in shape and form to the observed functions, as the curves can then be easily approximated by a linear combination of the chosen basis function [64]. As there is no clear rule for choosing the most efficient shape and number of basis functions for multivariate functional clustering [71], we follow the suggestions of Schmutz et al. [72] and choose a set of B-spline functions whose size corresponds to the number of years for every variable, while applying a small roughness parameter to reduce potential biases due to our earlier spline imputation.

2.2. Multivariate Functional Clustering

Cluster analyses are used to find homogeneous groups of observations in datasets without prior knowledge of latent group relationships, which can be achieved with a wide variety of algorithms that have been proposed for clustering of functional data. However, due to the potentially infinite-dimensional nature of functional data, several issues arise that are of lesser importance for classical cluster analyses, such as the reduction of functional dimensionality, which need to be solved. To address these issues, several methodological approaches for clustering functional data have recently been published, ranging from the simple transfers of classical algorithms to the functional domain, to complex model-based clustering after applying statistical filtering (see Jacques and Preda [71] for a review).

However, most of these approaches focus on clustering univariate functional data (see e.g., Abraham et al. [73], Serban and Wasserman [74], Coffey et al. [75], Peng and Müller [76], Li and Chiou [77], Chiou and Li [78], James and Sugar [79], Bouveyron and Jacques [80], Bouveyron et al. [81], Jacques and Preda [82], Bongiorno and Goia [83]), while there are still only a few concepts dedicated to multivariate functional clustering. Among these concepts, model-based approaches have received more attention in recent years, as they have proven to be suitable for complex statistical structures and relationships (see e.g., Schmutz, Jacques, Bouveyron, Cheze and Martin [72], Bouveyron and Jacques [80], Ieva and Paganoni [84], Kayano et al. [85], Jacques and Preda [86], Traore et al. [87]).

In our cluster analysis, we follow the mixture model-based approach proposed by Schmutz et al. [72] to cluster multivariate functional data of regional innovation activities in order to investigate spatio-temporal similarities and differences in the European innovation system. This approach builds on previous work by Bouveyron and Jacques [80] and Jacques and Preda [86] by circumventing the curse of dimensionality [88] with a multivariate functional principal component analysis (MFPCA) and considers the analytical scores to be random variables with cluster-specific probability distributions. By reprojecting the previously infinite- onto a finite-dimensional problem, the cluster-specific probability distributions can then be approximated via expectation maximization (EM) [89], which makes this approach highly flexible as additional assumptions can easily be imposed on the model.

2.3. Multivariate Functional Principal Component Analysis

The use of principal component analysis for functional data as a means for dimensionality reduction was already proposed by Ramsay and Silverman [69]. Multivariate functional data require more adaptive approaches, as shown by Jacques and Preda [86] and Schmutz et al. [72]. Specifically, MFPCA aims to find the eigenvalues and eigenfunctions to solve the decomposition equation of the covariance operator

$$v f_j = \lambda_l f_j$$

where λ_j is a finite group of j positive eigenvalues, principal scores, and f_j is a group of corresponding multivariate eigenfunctions, principal factors. Following Schmutz et al. [72], we assume that the latter are part of a linear space spanned by a matrix ϕ:

$$f_j(t) = \phi(t) b_j'$$

Consequently, we can reformulate the eigenproblem using the covariance estimator

$$\hat{v}(s,t) = \frac{1}{n-1} \phi(s) C' C \phi'(t)$$

which leads to

$$\frac{1}{n-1} \phi(s) C' C W b_l' = \lambda_l \phi(t) b_l'$$

where $W = \int_0^T \phi'(t)\phi(t)$ is a $R \times R$-Matrix containing the inner product of our basis functions. The principal component analysis is then reduced to an eigenvalue decomposition of the matrix

$$\frac{1}{\sqrt{n-1}} C W^{\frac{1}{2}}$$

allowing each multivariate curve $S_n(t_{n,p})$ to be identified by its scores $\delta_i = (\delta_{ij})$ into the basis of multivariate eigenfunctions (f_j) for $j \geq 1$ (see Jacques and Preda [86] and Schmutz et al. (2020) for proofs).

2.4. Mixture Model-Based Clustering of Multivariate Functional Data

Model-based clustering assumes that population data are a mixture of groups so that the elements of this mixture can be modeled by their conditional probability distribution. Therefore, the latent finite mixture model for the approach by Schmutz et al. [72] can be formulated as

$$g(s) = \sum_{k=1}^{K} \pi_k f_k(s_n)$$

where $g(s)$ is the probability density function of s, the mixture proportion of the k-th cluster is given by π_k with $\sum_{k=1}^{k} \pi_k = 1$ and $f_k(s_n)$ being the conditional density function. However, a feature of functional random variables is the lack of general notion of probability density functions [90], which necessitates the use of a parametric approximation:

$$g(s) = \sum_{k=1}^{K} \pi_k f_k(s_n;\ \theta_k)$$

with θ_k being the parameter vector of the k-th mixture element. Given this approximation, the likelihood of the mixture model proposed by Schmutz et al. [72] is then given by

$$l(\theta; s; z) = \sum_{n=1}^{N} \sum_{k=1}^{K} z_{kn} \log(\pi_k f(s_n; \theta_k))$$

where z_{kn} is a latent group variable equal to 1 if multivariate curves belong to cluster k or 0 otherwise. Finally, we can obtain a fully parameterized form of the likelihood by including the Gaussian density function $f(s_n; \theta_k))$ (see Schmutz et al. [72] for proofs):

$$l(\theta; s; z) = -\frac{1}{2} \sum_{k=1}^{K} n_k \left[-2\log(\pi_k) + \sum_{j=1}^{d_k} \log\left(a_{kj}\right) + \sum_{j=d_k+1}^{R} \log(b_k) + \sum_{j=1}^{d_k} \frac{q_{kj}^t W^{1/2} C_k W^{1/2} q_{kj}}{a_{kj}} + \sum_{j=d_k+1}^{R} \frac{q_{kj}^t W^{1/2} C_k W^{1/2} q_{kj}}{b_k} \right] + \frac{nR}{2} \log(2\pi)$$

where a_{kj} and b_k are a direct result of the MFPCA, since it is assumed that the scores of the n_k curves of the k-th cluster $\delta_n{}^k$ follow a Gaussian distribution with mean function $\mu_k \in \mathbb{R}$ and a covariance matrix Δ_k. The latter is crucial for both parameters as they are diagonal matrix elements:

$$\Delta_k = \left(\begin{array}{cc} \begin{bmatrix} a_{k1} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & a_{kd_k} \end{bmatrix} & 0 \\ 0 & \begin{bmatrix} b_k & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & b_k \end{bmatrix} \end{array} \right) \begin{array}{l} \left.\vphantom{\begin{matrix}a\\a\\a\end{matrix}}\right\} d_k \\ \left.\vphantom{\begin{matrix}a\\a\\a\end{matrix}}\right\} R - d_k \end{array}$$

Due to this covariance matrix, the variance of the first d_k principal components can be modeled much more accurately, while the other components can be retained and modeled via the parameter b_k, which provides a model with much higher degrees of clustering flexibility [72].

2.5. Model Inference via Expectation Maximization

An expectation maximization algorithm [89] is used to estimate the parameters of the complete likelihood given in the previous section, as this type of algorithm is reliable and reproducible in maximizing the likelihood of model-based clustering approaches. The algorithm uses two stages to estimate the model parameters and constantly alternates from one to the other until an optimal solution is found (Schmutz et al., 2020).

In the expectation step, the conditional expectation of the log-likelihood is calculated using the most recent parameter estimates. Then, the maximization step updates these estimates by maximizing the expected log-likelihood conditionally. The process is stopped when the difference of two successive estimations is smaller than 10^{-6} or a limit of 200 iterations has been reached. However, the algorithm must first be initialized by either providing initial values or using random values. We choose to initialize the clustering analysis by applying a k-means algorithm by Hartigan and Wong [91] with four partitions to the discretized values of our functional dataset to obtain initial values for the functional partitions. Although Schmutz et al. [72] suggest using multiple initialization strategies to prevent convergence to a local maximum, we found this approach to result in nearly identical cluster solutions as random initialization.

To obtain optimal cluster solutions for each IPC class, a series of models covering all parameter constraints provided by Schmutz et al. (2020) is estimated for a range of 2 to 10 clusters. We retain the solutions with the lowest Bayesian information criterion (BIC) [92] as our final cluster solution. The BIC is defined by

$$BIC = l(\theta; s; z) - \frac{m}{2} * \log(n)$$

where $l(\theta; s; z)$ is the maximum log-likelihood value, the number of model parameters is given with m and n is the number of individuals, which allows the log-likelihood to be penalized by model complexity. This procedure is in line with proposals by Schmutz et al. (2020).

3. Results

In the following sections, the results of our mixture model-based cluster analysis are presented for the eight IPC classes for the innovation indicators. As these cluster solutions are the results of multivariate functional dynamics, differentiation of the clusters is based on a simultaneous evaluation of all modeling variables, i.e., Innovation Gini coefficients, labor density, the share of R&D labor, GDP per labor, human capital density, and R&D investment per R&D labor (see Table 1). This ensures that subtle spatio-temporal regional dynamics in the modeled indicators are captured and regional disparities can be shown more clearly. To optimize the cluster solution, a range of models with various parameter constraints is used for up to 10 clusters, with the lowest BIC indicating the best cluster solution for a given set of variables. Accordingly, the number of clusters varies across the eight patent classes, but the size of the clusters is not limited, i.e., the numbers of regions per cluster only depend on regional similarities in innovation dynamics.

3.1. IPC Class A

The clustering process results in ten distinct clusters of spatio-temporal innovation dynamics for the patent group ‘human necessities’ (IPC A, see Figure 1, 1st row, left panel). While clusters 1, 2, and 6 are relatively small and limited to a few regions spread over Central European countries, clusters 4, 5, and 10 consist mostly of neighboring regions in, with few exceptions, large parts of Eastern Europe and the Baltic Area (cluster 4), and Portugal and Spain (cluster 5). In contrast, cluster 10 is significantly less spatially concentrated, containing most parts of France, but also regions in Italy, Austria, Germany, or Finland. Another large cluster is found in Scandinavian regions and Iceland, with regions in the United Kingdom, Germany and Italy also assigned to this group of innovation concentration. East Germany has similar innovation dynamics to regions in northern Spain, northern Ireland, and southern Italy (cluster 8). These regions are often characterized as structurally weak, which seems to be reflected in innovation concentration potentials. Most regions in Central Europe, mainly Germany, Luxembourg, Belgium, and the Netherlands, are highly diverse with neighboring regions not being part of the same innovation cluster.

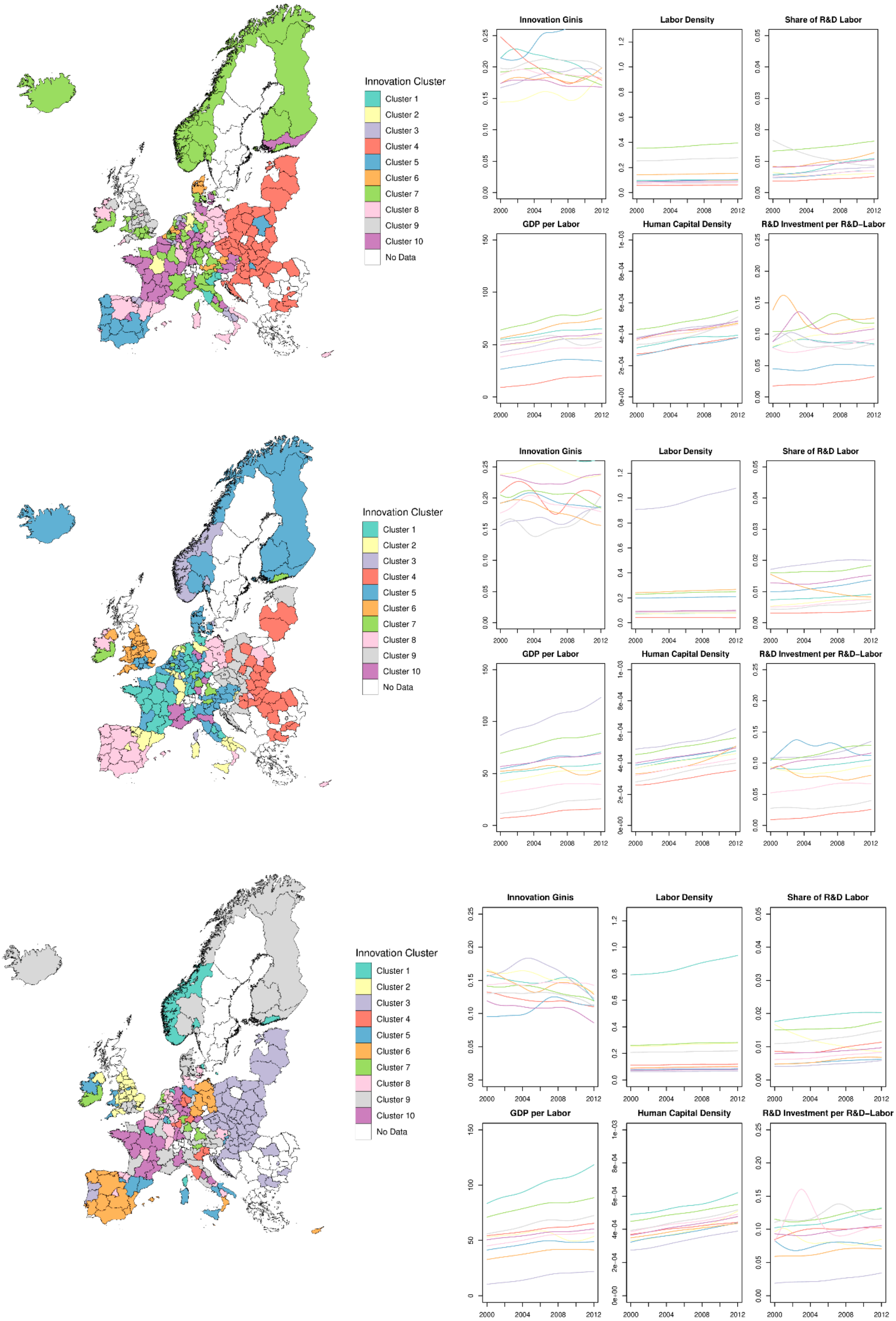

Figure 1. Results of the mixture model-based multivariate functional clustering algorithm. Rows: IPC classes A, B, and C. (A: 'human necessities'; B: 'performing operations and transporting'; C: 'chemistry and metallurgy'). Columns: **left**: Spatial cluster mapping, **right**: Temporal cluster dynamics. Source: Own calculations.

Regarding the mean curve of the Innovation Gini (see Figure 1, 1st row, right panel), it is noticeable that cluster 4 drops significantly compared to the other clusters until 2008, before a slight increase to a stable trajectory sets in. In comparison, the mean curve of cluster 5 rises very sharply from 2004 onwards, reaching a higher level of concentration than any of the other clusters. Most other cluster mean curves either show a stale trajectory or increase slightly until 2008 before declining. Despite these temporal dynamics, there are noticeable crossings of most cluster mean curves, with cluster 5 being an exception, i.e., the clusters only develop in a relatively narrow range overall in terms of innovation concentration, but evolve highly variably in this range. The mean curves of the variables for labor density, the share of R&D labor, GDP per labor, and human capital density from 2000 to 2012 show a relatively even and almost linear increase. In terms of human capital density, the clusters do not differ much. In principle, these results hold for the mean curves of all indicators, but the decreasing function of cluster 10 deviates from the other stable or slightly increasing cluster trajectories of the share of R&D labor. The mean curves for R&D investment per R&D labor are very differentiated. While the curves for clusters 6 and 10 initially increase, then decrease until 2005, clusters 4 and 5 show a stable level, which is, however, well below all other clusters.

3.2. IPC Class B

Ten clusters are found for the patent group 'performing operations and transporting' (IPC B). Again, Central Europe is quite fragmented in terms of cluster memberships, as the regions in this part of Europe are assigned to clusters 1, 2, 5, 7, and 10 (see Figure 1, 2nd row, left panel). Other regions in cluster 5 are located in England, large parts of Norway, Iceland, and Finland, but also in Austria, northern Italy, and France, while cluster 8 is mainly located in Spain, Portugal, and East Germany, which corresponds to the clustering previously observed in these regions for IPC class A. Eastern Europe is largely composed of regions in clusters 4 and 8, with slightly more variation than for class A. Ireland and Northern Ireland are divided into three different clusters (6, 7, and 8), while Norway is divided into two large clusters (3 and 5). Most Regions belonging to cluster 2 are located in southern Italy and the northeastern parts of Spain and France, as well as parts of the Netherlands and Germany.

Regarding the temporal dynamics of the cluster for this IPC class, there are significant differences in the mean curves for innovation concentration (see Figure 1, 2nd row, right panel). For example, clusters 4 and 9, both located mainly in Eastern Europe, differ strongly. While the mean curves for most clusters decrease over time, cluster 3 seems to be an exception, as regions in this cluster seem to slightly increase their degree of concentration. In general, cluster 10 is the most stable in terms of innovation concentration. However, the mean curves of most clusters have slightly decreased since 2005, which corresponds to a decrease in innovation concentration. The other covariates, with exception of R&D investment per R&D labor, mostly show stable or slight, almost linear, increases over time. The mean curves of all clusters are very close to each other and similar in terms of human capital density, which again is consistent with the results for IPC class A. The mean curves for GDP per labor show wide variation in terms of the level, with only cluster 6 and its exclusive focus on the UK showing a slight decline from 2007 onwards, while the regions of cluster 3, which are mostly located in Norway, show the highest overall mean values. In terms of the share of R&D labor, cluster 6 again diverges from the other clusters, showing a steady decline over time, while the other clusters remain largely stable. In terms of labor density, the mean curve for cluster 3 is significantly higher than all other curves, which are stable over time. The mean curves of clusters 5 and 6 show opposite trends in R&D investment per R&D labor, with one cluster increasing while the other decreases and vice versa. With the exception of the last four years, cluster 5 is mostly higher than the other clusters, which show a slight and steady increase over time.

3.3. IPC Class C

For the patent class 'chemistry and metallurgy', the spatio-temporal clustering process again resulted in ten clusters (see Figure 1, 3rd row, left panel). Essentially, the clustering appears to be similar to the results of the previous patent class, with a few exceptions. For example, clusters 6, 9, and 10 are mostly identical, with two Portuguese regions now belonging to the cluster mainly located in Eastern Europe. The latter is no longer divided, as all Eastern European regions have innovation profiles that make them part of the same homogeneous cluster. Compared to the results of the previous patent class, a few regions in cluster 1 are showing changes in the clustering pattern. While Central Europe is again fragmented compared to patent class B, and this also applies to a higher degree to Spain and Portugal, the homogeneous structure of Eastern Europe represents a clear contrast to the rest of Europe.

The mean curves for the Innovation Ginis are quite similar in their temporal dynamics, with several curves intersecting each other, but most remaining within a narrow, slightly declining corridor (see Figure 1, 3rd row, right panel). While cluster 3, with its focus on Eastern Europe, has the highest level of mean curves but declines sharply from 2008 onwards, clusters 5 and 6 seem to develop comparably from 2008 onwards with slight time lags, whereas previously they had complementary trajectories. As in the previous cluster results, the labor-related covariates show a slight, but constant increase in the mean curves. In terms of human capital density, the mean curves are again close to each other and also increase linearly. There is an increase in GDP per labor for all clusters, with a slight dip in 2008 and the Eastern European regions of cluster 3 showing the lowest mean curve values. The share of R&D labor is more or less stagnant for all clusters, with cluster 5 again showing the highest mean curve values. In terms of R&D investment in R&D labor, the spatially-spread cluster 8 shows a strong increase in 2003, followed by a similarly long decline until 2008. Regarding labor density, all clusters show linear trajectories at very low levels, except for cluster 5, which is spread over half a dozen regions across Europe and shows significantly higher and slightly increasing mean curve values. Across all covariates, mean curves of cluster 3 are lower than all others, with the exception of Innovation Gini curves.

3.4. IPC Class D

In terms of the patent class for innovations in 'textiles and paper', a set of nine distinct clusters was found in the clustering process, possibly due to missing data for some regions included in previous clustering results (see Figure 2, 1st row, left panel). The cluster with the highest number of regions is cluster 3, which includes regions in Finland, most of France, and parts of Italy, Austria, Germany, Belgium, Luxembourg, and the Netherlands. Except for Southern Germany, Belgium, Luxembourg, and the Netherlands, neighboring regions are part of the same cluster. The United Kingdom is divided into four different clusters, with cluster 9 occurring only in England. Most parts of Eastern Germany, Northern Ireland, and parts of Spain are members of cluster 8. As the innovation profiles for this patent class seem to be more homogeneous than in previous results, most regions belong to a few larger clusters, while the remaining regions are divided into highly distinct clusters 1, 4, and 6.

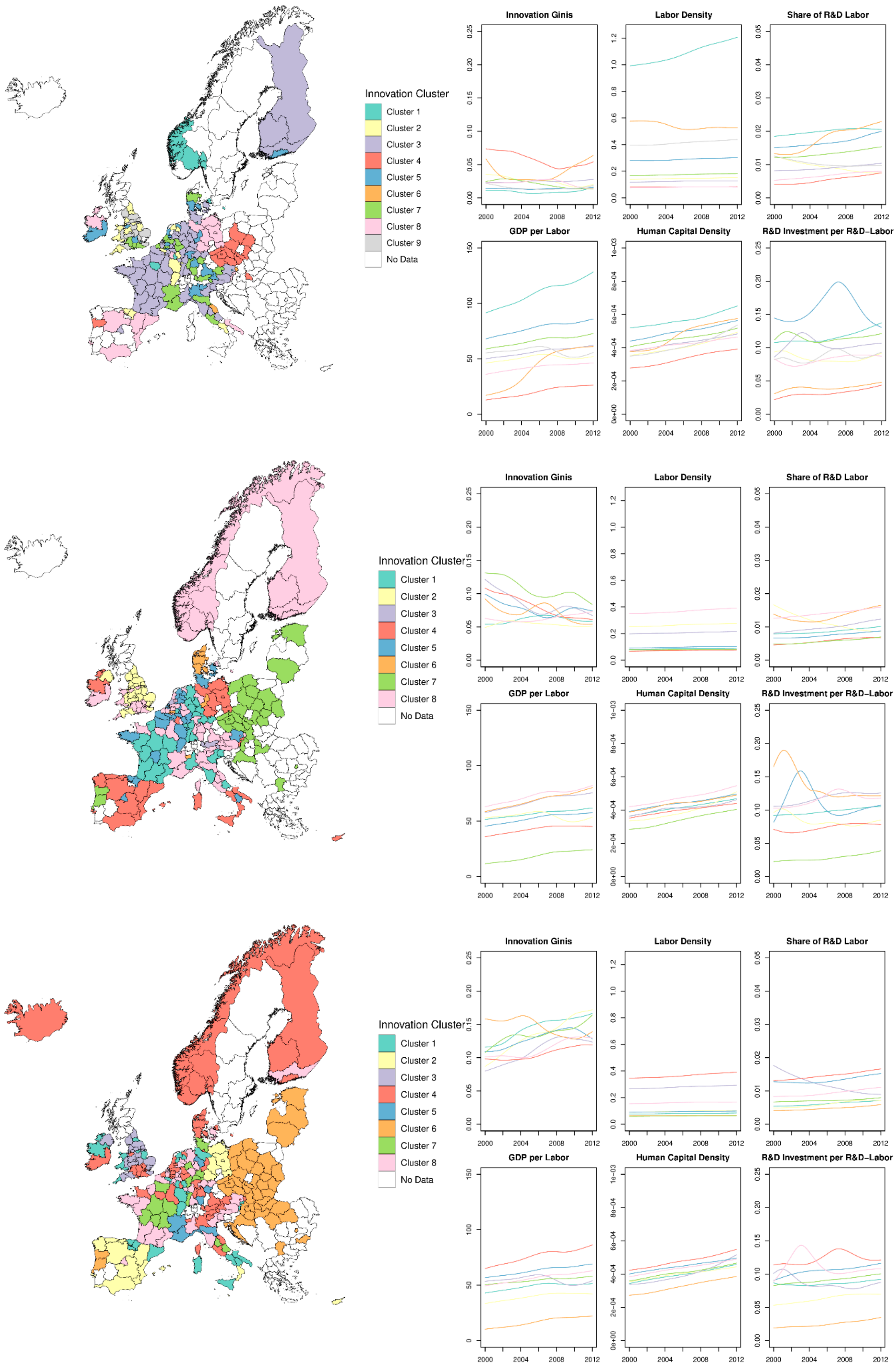

Figure 2. Results of the mixture model-based multivariate functional clustering algorithm. Rows: IPC classes D, E, and F (D: 'textiles and paper'; E: 'fixed constructions'; F: 'mechanical engineering, lighting, heating, weapons, and blasting'). Columns: **left**: Spatial cluster mapping, **right**: Temporal cluster dynamics. Source: Own calculations.

Compared to the previous results, the mean curves for the Innovation Ginis are at a very low level (see Figure 2, 1st row, right panel). Here, the regions of Eastern Europe and Portugal of cluster 4 show the highest values but decline slightly after 2008. This contrasts with cluster 6, which consists of only two regions and follows a U-shaped trajectory so that the curve only rises steadily after 2008 and shows the highest values of all clusters. Regarding the other variables, the mean curves show mostly linear trajectories, with a few exceptions such as cluster 1, which mostly consists of Norwegian regions and shows the highest mean values with increasing trends. This is particularly noticeable for labor density and GDP per labor. Compared to the other clusters, cluster 6 varies the most over time, with its trajectory changing towards 2004 and even increasing non-linearly for both covariates and human capital density. As with the previous results, the most variation across all clusters is found for R&D investment per R&D labor. Here, cluster 5, which is scattered across Europe, shows a sharp increase to the highest mean curve value in 2008, before declining similarly. This is mirrored at a lower level in clusters 3, 7, and 9, with the first two reaching their maximum around 2002.

3.5. IPC Class E

The clustering for innovations in 'fixed constructions' are quite similar to the clustering for the IPC class C, although only eight clusters are found (see Figure 2, 2nd row, left panel). Essentially, clusters 2, 4, and 7 are evidence for this similarity. The fragmentation of Central Europe has shifted slightly to the west, as western German regions are members of the same cluster. The Scandinavian cluster is also found in Central European regions and is scattered across northern Italy, parts of the United Kingdom, and Ireland. Another similarity to the innovation profiles of IPC classes C and E can be seen due to cluster 2, which is exclusively found in the UK.

In comparison with the results of IPC class C, the mean curves for the Innovation Ginis for fixed constructions are on a much lower level, with the Eastern European regions of cluster 7 showing the highest mean curve values (see Figure 2, 2nd row, right panel). Furthermore, clusters 5 and 6 show complementary trajectories and while the mean curves are steadily decreasing, at the same time the dispersion of all mean curves is decreasing over time. For labor density, the share of R&D labor, GDP per labor, and human capital density the temporal dynamics of the mean curves are again comparable to the results of IPC class C, with the exception being that the dispersion across the mean curves is much smaller and no cluster has significantly higher mean values than all other clusters. In terms of R&D investment per R&D labor, clusters 5 and 6 show high maxima in the period from 2000 to 2004 and then converge to the overall corridor of cluster mean curves.

3.6. IPC Class F

For the patent class for 'mechanical engineering, lighting, heating, weapons, and blasting' innovations, eight clusters are found, again showing noticeable similarities to the cluster results for IPC class C (see Figure 2, 3rd row, left panel). Especially the Eastern European regions (cluster 6), Scandinavia and parts of Central Europe (cluster 4), Spain, Portugal, and East Germany (clusters 2 and 6) as well as France (clusters 6 and 7) are the reason for the similarities in the spatial cluster pattern. Nevertheless, some deviations from previous clustering results can be found in western Germany, northern Italy, Austria, and parts of France. Compared to Western Europe, innovation profiles in the Northern and Eastern European regions seem to be more homogenous.

The cluster mean curves for the Innovations Ginis show some temporal variation and an overall increasing trend, with cluster 6 showing the highest level until 2004 before decreasing thereafter (see Figure 2, 3rd row, right panel). The strongest increase is shown by the mostly non-adjacent Central Europe regions of cluster 5 and cluster 2 (East Germany and Spain), while the regions of cluster 4 stagnate at a stable level. Concerning the other covariates, similar temporal trajectories as for IPC class E are shown for the mean curves.

Due to the curve maxima not standing out from the curves as in previous results, the cluster mean curves show smoother trajectories overall.

3.7. IPC Class G

As with most previous results, ten innovation clusters are found for the patent class for 'physics' that resemble the clustering pattern of IPC class C, while sharing a few similarities with IPC classes E and F (see Figure 3, 1st row, left panel), except for Denmark, which is now an independent cluster with a single region in northern Germany (cluster 7) and no longer part of the Scandinavian cluster (cluster 2). In addition, some smaller regions in the Netherlands are assigned differently compared to other IPC classes.

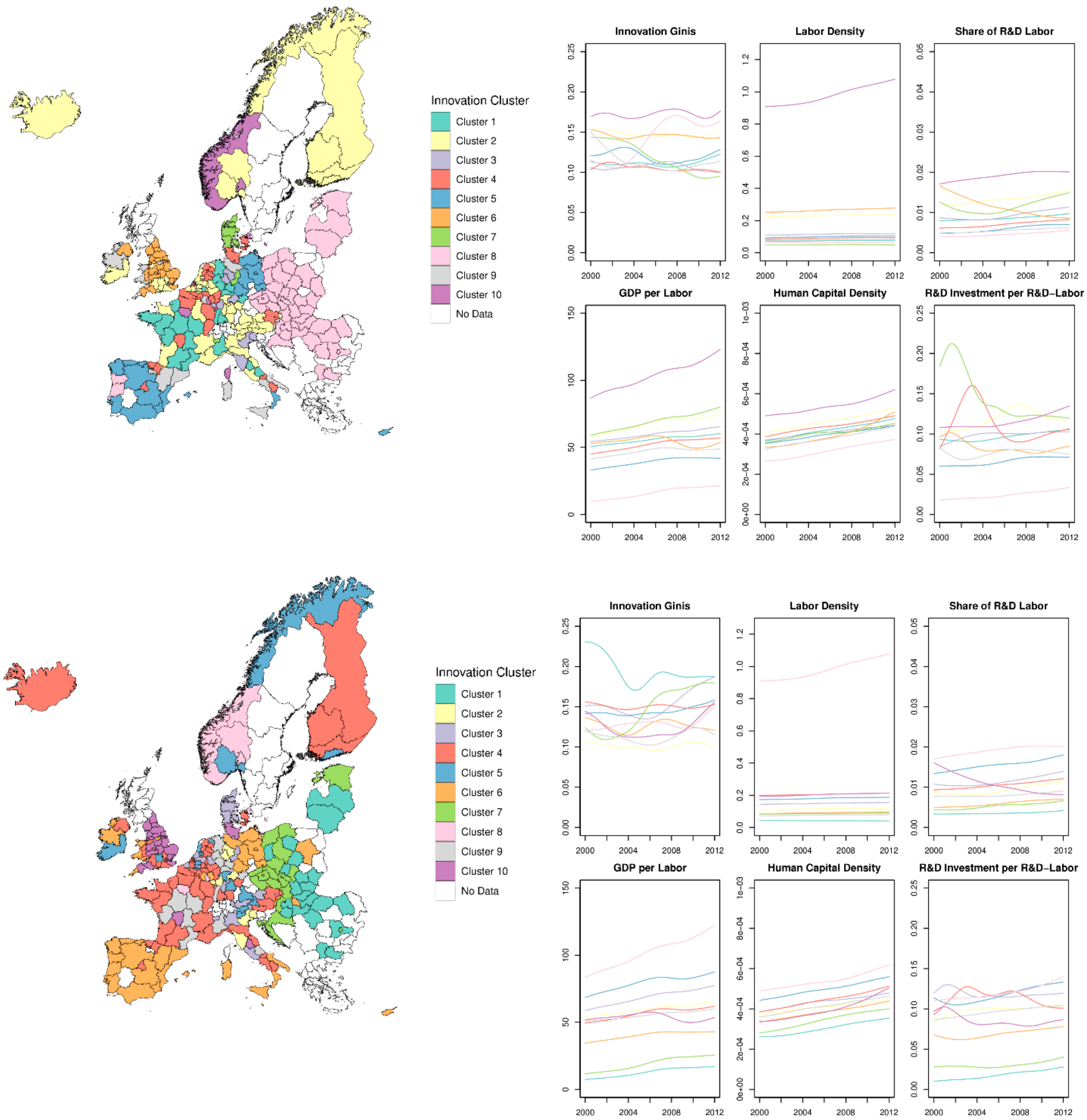

Figure 3. Results of the mixture model-based multivariate functional clustering algorithm. Rows: IPC classes G and H (G: 'physics'; H: 'electricity'). Columns: **left**: Spatial cluster mapping, **right**: Temporal cluster dynamics. Source: Own calculations.

In the cluster mean curves of the Innovation Ginis, both the overall level and the curve maxima are very pronounced in comparison to IPC class C (see Figure 3, 1st row, right panel). The highest mean curve values are shown for cluster 10, which is located mostly in Norway, while cluster 8 increases sharply in 2004 before matching the temporal dynamic of cluster 10 from 2008 onwards. In contrast, most cluster mean curves remain stable for this variable, with cluster 7 being an exception that decreases over time. As far as the other variables are concerned, cluster 10 has the highest level of all mean curves for almost all of these variables. For labor density, the share of R&D labor, GDP per labor, and human capital density the difference between cluster 10 and the other clusters in mean values is very clear. Only in terms of R&D investment per R&D labor is cluster 10 surpassed by the maxima of clusters 6 and 7 until 2005, but as these curves decline again, the mean curve of cluster 10 reaches the highest level again in 2012.

3.8. IPC Class H

The last clustering found a set of ten clusters for electrical innovations (IPC H, see Figure 3, 2nd row, left panel). The East German regions are divided into three larger clusters (clusters 1, 6, and 7), with cluster 6 again consisting of regions in East Germany, Spain, and Portugal that were assigned to the same cluster for other IPC classes. In addition, several regions in Ireland, England and southern Italy are members of this cluster. Scandinavia is also divided into three clusters, with members of cluster 4 found in Finland, France, Austria, and other regions throughout Central Europe. Another cluster is found in the southern regions of Norway as well as the central region of Paris. The rest of Scandinavia is clustered together with southern Ireland, southern Germany, and the central London region (cluster 5). Most regions in England form their cluster, with only a few exceptions in Denmark, Italy, and the southwest of France (cluster 10).

Consistent with all previous results, there is a high degree of variation in the mean curves of innovation concentration, with cluster 1 showing the highest overall mean values, but steadily decreasing over time, with a noticeable minimum in 2005 (see Figure 3, 2nd row, right panel). While most other clusters stagnate at a stable level, cluster 7 increases sharply after a minimum in 2001. In comparison, cluster 10 decreases until 2004, and stagnates until 2009 before finally increasing. In terms of the other covariates, there are large similarities to IPC class G. Regarding labor density, cluster 8 shows the highest mean values and increase, with a clear gap to the other clusters, which remain constant over time and show similar values, with minor variations. For all other variables, cluster 10 differs the most from the other clusters, as the mean curve for the share of R&D labor decreases while all others increase constantly. In addition, there is a dip in GDP per labor in 2007 and human capital seems to be gradually increasing for cluster 10. While the mean curves for most clusters are quite similar to the curves for IPC class G, high maxima for R&D investment per R&D labor are missing.

4. Discussion

Overall, innovation clusters in Europe differ by IPC class, although some regions are more similar than others and some IPC classes more interconnected in terms of innovation concentration. The Innovation Gini is mostly similar in the main regions in Eastern Europe, Spain, Portugal, and East Germany, resulting in these regions being in the same spatio-temporal cluster groups. Regarding the Innovation Gini and the different covariates used in the mixture model-based multivariate functional clustering, it is noticeable that some covariates seem to have opposite functional effects. This is the case when considering regions with the highest values of innovation concentration over time, which is usually accompanied by the lowest values in the covariates. This holds for all IPC classes except for class G ('physics'). The clustering results for the classes E, C, and F are similar, with the pair E/F being more similar than the pair E/C. In addition, cluster solutions for the classes G and C as well as G and H show similar temporal dynamics.

If one relates the clustering results to analyses of innovation-promoting policies from the same period, the clustering clearly shows the various efforts in innovation policy and general economic trends such as the economic crisis of 2008. The crisis is reflected in the functional curves and affects almost all IPC cluster solutions, with some being more affected than others. As Izsák et al. [20] state in their final report for the European Commission, funding focused on innovation development slowed down during the period of our analysis, especially after the economic crisis. Nevertheless, funding shifted towards more collaborative projects which is one reason that our analysis showed the emergence of clusters not only in neighboring regions but also at the supra-regional level. Furthermore, the funding priorities have not shifted in their scientific and technological cores, so the FDA cluster model should be able to capture relevant effects to a large degree.

Considering patents as one of the innovation measures is consistent with the concept of closed innovation, which is solved by adding further innovation indicators in the models. These types of indicators mainly include ex-ante innovation indicators, where the success of innovations is implicit, but considered sufficient for analyzing and forecasting effects on the economy [93–95]. Nevertheless, open innovation concepts can enrich innovation analyses by resolving the opposition to patents. Licensing concepts are not always antithetical to intellectual property rights in the sense of patents, as flexible concepts can increase the value of technologies and facilitate innovation dispersion [96–98]. Our analysis also includes multiple indicators and highlights the importance of regional innovation structures and policy support where innovation is lacking and thus can support the findings of, e.g., Leckel et al. [35]. The literature on innovation systems also proves to be a relevant point that our results can support [36], as Pelau and Chinie [18] show. Moreover, as McPhillips [37] notes, innovation clusters differ in terms of the openness of innovation, and policy can support where barriers exist. Our analyses can provide information about the characteristics of regional clusters and serve as leverage for improved policy targeting. Open innovation can be conducive to further innovation gains, which is likely to benefit from information about specializations in the regional clusters, in line with innovation system literature [36,99].

In general, our analysis would benefit from longer time series of data that could provide further insights into national and regional innovation dynamics. The periods of funding programs often span several years or even decades, and their impact might not be fully captured by the analysis conducted in this paper. Similarly, it is possible that the effects of regional innovation policies have not been significant enough to have lasting effects related to innovation concentration [20].

The concept of the European and Regional Innovation Scoreboards takes into account innovation developments over time and divides nations by regions, but policies derived from the European legislation are relatively inconsistent. The innovation index generally shows little variation between countries, with most countries occupying the same or similar categories of innovation leadership. This is also true across regions, with exceptions due to highly specialized regions (e.g., Malta as a moderate innovator, is among the strongest innovators in digitalization) [9]. As Pełka [14] states, innovation is heterogeneous. Our model supports this result and we further suggest combining multiple measures of innovation indicators, which is in line with Spielkamp and Vopel [12].

Izsák et al. [20] conclude that innovation policy should be location-based and tailored to different conditions to take national characteristics into account. This idea is supported by the results of our analysis, as regional characteristics and differing conditions in the technological mix foster the emergence of heterogeneous innovation portfolios and thus suggest higher policy efficiency if properly taken into account. Moreover, place-based policy strategies should leverage the interconnectedness of industries, authorities, firms, and other actors by mobilizing knowledge internal and external to the region to facilitate innovation [100]. This is also noted by Capello and Lenzi [11] and supported by our model, which describes specialization in clusters corresponding to IPC classes. Policy concepts should be flexible and not apply a fixed scheme to all regions, as they differ with regard to a multitude of characteristics, as our analysis shows [100,101].

5. Conclusions

Knowledge of regional innovation dynamics, leading to different Innovation Ginis that result in clustering regions differently across all of Europe depending on the type of innovation activity, is crucial when designing policies for supporting innovation in Europe. The emergence of innovation clustering in different regions is an important factor for innovation systems, drawing on principles of agglomeration economics. This paper extends the knowledge of clusters by highlighting the technological specialization of regional clusters, which underlines the relevance of innovation systems and also provides options for policy concepts related to open innovation and regional development.

In this paper, a mixture model-based multivariate functional clustering algorithm has been adapted to analyze the spatio-temporal dynamics of European regional innovation activities at the NUTS-2 level from 2000 to 2012. As this analysis is based on the paradigm of functional data, it allows the analysis of latent features and dynamics in multivariate time series of innovation activities that would have been too subtle to be captured by classical time series or clustering approaches, making it possible to cluster 225 European regions according to their temporal innovation profiles. This was achieved through the integrative combination of multivariate functional principal component analysis, mixture modeling, and expectation maximization, overcoming the limitations of classical methods in analyzing high-dimensional data. However, it should be noted that this multivariate clustering approach requires the reconstruction of functional forms from raw time series, which necessitates careful consideration of robust spline interpolations to impute missing data and the selection of appropriate basis expansions. Therefore, functional clustering, especially with regard to multivariate data, is more laborious in the pre-processing stages compared to other methods, as any interpolation or expansion approach may inappropriately influence or change the functional form, which must be avoided through multiple sensitivity and simulation analysis runs. Nevertheless, by using these multivariate innovation time series, multi-characteristic innovation activity is taken into account, reflecting the political efforts of European policy programs. Our measurements for identifying the clusters are innovation- and economy-related variables including innovation concentration indicators that consider different IPC classes of patents. Thus, regions are profiled according to their innovation portfolios.

The resulting innovative activity across the European clusters differs, although some regions in Eastern Europe and on the Iberian Peninsula are reliably constant across innovation types. Accounting for the differences in innovation, clustering for IPC classes E ('fixed constructions') and F 'mechanical engineering, lighting, heating, weapons, and blasting') is almost identical, whereas similarities in regional clustering of classes E and C ('chemistry and metallurgy') are relatively more distinct but still comparable. Clusters of classes G ('physics') and C are correspondent while classes H ('electricity') and G exhibit comparable dynamics over time. This supports a place-based regional innovation policy approach that is not only able to account for differing regional potentials in innovation but also diverging specialization in innovation types.

For future research, it is crucial to consider and address the limited data availability in terms of time series length and missing values. The accuracy of the results could be further increased if these limitations were removed by better, more complete, and more recent data, or if more variables were added to further optimize our existing models. Another possibility would be to analyze regions below the NUTS-2 level if high-quality data from official sources are available.

Author Contributions: Conceptualization, I.R. and D.W.; methodology, D.W.; software, D.W.; validation, I.R. and D.W.; formal analysis, D.W.; investigation, I.R. and D.W.; resources, I.R., D.W. and A.-K.V.; data curation, I.R. and D.W.; writing—original draft preparation, I.R., D.W. and A.-K.V.; writing—review and editing, I.R., D.W. and A.-K.V.; visualization, I.R. and D.W.; supervision, I.R.; project administration, I.R. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: The authors would like to thank Wilhelm Kuckshinrichs, Stefan Vögele, Ashraf Pramanik and three anonymous reviewers for support and valuable comments.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Acs, Z.J.; Varga, A. Geography, endogenous growth, and innovation. *Int. Reg. Sci. Rev.* **2002**, *25*, 132–148. [CrossRef]
2. Rhoden, I. *Space and Time in German Innovative Activity*; Springer: Berlin/Heidelberg, Germany, 2019.
3. Fornahl, D.; Brenner, T. Geographic concentration of innovative activities in Germany. *Struct. Change Econ. Dyn.* **2009**, *20*, 163–182. [CrossRef]
4. McCann, P.; Simonen, J. Innovation, knowledge spillovers and local labour markets. *Pap. Reg. Sci.* **2005**, *84*, 465–485. [CrossRef]
5. Costantini, V.; Mazzanti, M.; Montini, A. Environmental performance, innovation and spillovers. Evidence from a regional NAMEA. *Ecol. Econ.* **2013**, *89*, 101–114. [CrossRef]
6. Aldieri, L.; Carlucci, F.; Vinci, C.P.; Yigitcanlar, T. Environmental innovation, knowledge spillovers and policy implications: A systematic review of the economic effects literature. *J. Clean. Prod.* **2019**, *239*, 118051. [CrossRef]
7. Bottazzi, L.; Peri, G. Innovation and spillovers in regions: Evidence from European patent data. *Eur. Econ. Rev.* **2003**, *47*, 687–710. [CrossRef]
8. Giannitsis, T.; Kager, M. *Technology and Specialization: Dilemmas, Options and Risks. Knowledge for Growth: Prospects for Science, Technology and Innovation*; Publications Office of the European Union: Luxembourg, 2010; Available online: https://data.europa.eu/doi/10.2777/47564 (accessed on 3 August 2021).
9. European Commission. European Innovation Scoreboard Integrated European Innovation Scoreboard Tool. Available online: https://interactivetool.eu/EIS/EIS_2.html (accessed on 3 August 2021).
10. European Commission. Internal Market, Industry, Entrepreneurship and SMEs-European Innovation Scoreboard. Available online: https://ec.europa.eu/growth/industry/policy/innovation/scoreboards_en (accessed on 3 August 2021).
11. Capello, R.; Lenzi, C. Territorial patterns of innovation: A taxonomy of innovative regions in Europe. *Ann. Reg. Sci.* **2013**, *51*, 119–154. [CrossRef]
12. Spielkamp, A.; Vopel, K. Mapping Innovative Clusters in National Innovation Systems. 1998. Available online: https://www.zew.de/en/publications/mapping-innovative-clusters-in-national-innovation-systems-1 (accessed on 3 August 2021).
13. Turkina, E.; Van Assche, A. Global connectedness and local innovation in industrial clusters. *J. Int. Bus. Stud.* **2018**, *49*, 706–728. [CrossRef]
14. Pełka, M. Analysis of innovations in the European Union via ensemble symbolic density clustering. *Ekonometria* **2018**, *22*, 84–98. [CrossRef]
15. Ionela-Andreea, P.; Marian, N. Cluster Analysis of Regional Research and Development Disparities in Europe. *Stud. Bus. Econ.* **2020**, *15*, 303–312. [CrossRef]
16. Caves, D.W.; Christensen, L.R.; Diewert, W.E. The Economic Theory of Index Numbers and the Measurement of Input, Output, and Productivity. *Econometrica* **1982**, *50*, 1393–1414. [CrossRef]
17. Zabala-Iturriagagoitia, J.M.; Aparicio, J.; Ortiz, L.; Carayannis, E.G.; Grigoroudis, E. The productivity of national innovation systems in Europe: Catching up or falling behind? *Technovation* **2021**, *102*, 102215. [CrossRef]
18. Pelau, C.; Chinie, A.C. Cluster Analysis for the Determination of Innovative and Sustainable Oriented Regions in Europe. *Studia Univ. Vasile Goldiș Arad Ser. Științe Econ.* **2018**, *28*, 36–47. [CrossRef]
19. Kim, G.; Bae, J. A novel approach to forecast promising technology through patent analysis. *Technol. Forecast. Soc. Change* **2017**, *117*, 228–237. [CrossRef]
20. Izsák, K.; Markianidou, P.; Radošević, S. *Lessons from a Decade of Innovation Policy-What can Be Learnt from the INNO Policy TrendChart and The Innovation Union Scoreboard*; European Commission: Brussels, Belgium, 2013.
21. European Commission. EU Cohesion Policy. Available online: https://ec.europa.eu/regional_policy/en/2021_2027/ (accessed on 3 August 2021).
22. Schmutz, A.; Jacques, J.; Bouveyron, C. funHDDC: Univariate and Multivariate Model-Based Clustering in Group-Specific Functional Subspaces. 2021. Available online: https://rdrr.io/cran/funHDDC/ (accessed on 3 August 2021).
23. Krugman, P.R. *Geography and Trade*; MIT Press: Cambridge, MA, USA, 1991.
24. Krugman, P. A Model of Innovation, Technology Transfer, and the World Distribution of Income. *J. Political Econ.* **1979**, *87*, 253–266. [CrossRef]
25. Sturgeon, T.J. What really goes on in Silicon Valley? Spatial clustering and dispersal in modular production networks. *J. Econ. Geogr.* **2003**, *3*, 199–225. [CrossRef]
26. Koschatzky, K. *Räumliche Aspekte im Innovationsprozess: Ein Beitrag Zur Neuen Wirtschaftsgeographie Aus Sicht der Regionalen Innovationsforschung*; LIT Verlag Münster: Münster, Germany, 2001.

27. Fernau, A.K. *Standortwahl Als Komponente der Wettbewerbsfähigkeit*; Deutscher Universitätsverlag: Wiesbaden, Germany, 1997.
28. Marshall, A. *Principles of Economics an Introductory Volume Macmillan and Co.*; Macmillan Publishers: London, UK, 1920.
29. Sraffa, P.; Dobb, M.H. *The Works and Correspondence of David Ricardo*; Cambridge University Press: Cambridge, UK, 1951; Volume 1.
30. Breuste, J.; Keidel, T. Urbane und suburbane Räume als Kulturlandschaften-planerische Gestaltungsaufgaben. *Inf. Zur Raumentwickl.* **2008**, *5*, 279–288.
31. Blume, L. Erfolgsfaktoren kommunaler Wirtschaftspolitik in Ostdeutschland. *Raumforsch. Und Raumordn.* **2004**, *62*, 3–17. [CrossRef]
32. Alcácer, J.; Chung, W. Location strategies for agglomeration economies. *Strateg. Manag. J.* **2014**, *35*, 1749–1761. [CrossRef]
33. Méndez-Suárez, M. Headquarters location decisions under conflicts at home: Evidence from a configurational analysis. *J. Small Bus. Strategy* **2021**, *31*, 1–9.
34. Ellram, L.M.; Tate, W.L.; Petersen, K.J. Offshoring and reshoring: An update on the manufacturing location decision. *J. Supply Chain. Manag.* **2013**, *49*, 14–22. [CrossRef]
35. Leckel, A.; Veilleux, S.; Dana, L.P. Local Open Innovation: A means for public policy to increase collaboration for innovation in SMEs. *Technol. Forecast. Soc. Change* **2020**, *153*, 119891. [CrossRef]
36. Cooke, P. Regional innovation systems: Competitive regulation in the new Europe. *Geoforum* **1992**, *23*, 365–382. [CrossRef]
37. McPhillips, M. Trouble in Paradise? Barriers to Open Innovation in Regional Clusters in the Era of the 4th Industrial Revolution. *J. Open Innov. Technol. Mark. Complex.* **2020**, *6*, 84. [CrossRef]
38. Pflüger, M. Regionale Disparitäten und Regionalpolitik: Treiber der Veränderung, Handlungsbedarf und Handlungsoptionen. *Wirtschaftsdienst* **2019**, *99*, 10–18. [CrossRef]
39. European Commission. The EU Green Deal. Available online: https://ec.europa.eu/info/strategy/priorities-2019-2024/european-green-deal_en (accessed on 2 August 2021).
40. Aghion, P.; Hemous, D.; Veugelers, R. No Green Growth without Innovation. 2009. Available online: https://www.bruegel.org/wp-content/uploads/imported/publications/pb_climatervpa_231109_01.pdf (accessed on 13 November 2021).
41. Partanen, J.; Möller, K.; Westerlund, M.; Rajala, R.; Rajala, A. Social capital in the growth of science-and-technology-based SMEs. *Ind. Mark. Manag.* **2008**, *37*, 513–522. [CrossRef]
42. Inkpen, A.C.; Tsang, E.W. Social capital, networks, and knowledge transfer. *Acad. Manag. Rev.* **2005**, *30*, 146–165. [CrossRef]
43. Cooke, P.; Wills, D. Small firms, social capital and the enhancement of business performance through innovation programmes. *Small Bus. Econ.* **1999**, *13*, 219–234. [CrossRef]
44. Pustovrh, A.; Rangus, K.; Drnovšek, M. The role of open innovation in developing an entrepreneurial support ecosystem. *Technol. Forecast. Soc. Change* **2020**, *152*, 119892. [CrossRef]
45. Chesbrough, H.; Bogers, M. Explicating open innovation: Clarifying an emerging paradigm for understanding innovation. In *New Frontiers in Open Innovation*; Oxford University Press: Oxford, UK, 2014; pp. 3–28.
46. Shearmur, R.; Doloreux, D. How open innovation processes vary between urban and remote environments: Slow innovators, market-sourced information and frequency of interaction. *Entrep. Reg. Dev.* **2016**, *28*, 337–357. [CrossRef]
47. WIPO. WIPO IP Portal. Available online: https://www.wipo.int/classifications/ipc/ipcpub/?notion=scheme&version=20210101&symbol=none&menulang=en&lang=en&viewmode=f&fipcpc=no&showdeleted=yes&indexes=no&headings=yes&notes=yes&direction=o2n&initial=A&cwid=none&tree=no&searchmode=smart (accessed on 2 August 2021).
48. Griliches, Z. *Patent Statistics as Economic Indicators: A Survey Part II*; National Bureau of Economic Research (NBER): Cambridge, MA, USA, 1990.
49. Rhoden, I. Innovating in Krugman's Footsteps–Where and How Innovation Differs in Europe: Static Innovation Indicators for Identifying Regional Policy Leverages. 2020. Available online: https://www.econstor.eu/handle/10419/218875 (accessed on 31 March 2021).
50. Eurostat. Eurostat Database. Available online: https://ec.europa.eu/eurostat/data/database (accessed on 31 March 2021).
51. OECD. Database. Available online: https://data.oecd.org/ (accessed on 31 March 2021).
52. Office for National Statistics. Database. Available online: https://www.ons.gov.uk/ (accessed on 31 March 2021).
53. Simonoff, J.S. *Smoothing Methods in Statistics*; Springer Science & Business Media: Berlin/Heidelberg, Germany, 2012.
54. Eubank, R.L. *Nonparametric Regression and Spline Smoothing*; CRC Press: Boca Raton, FL, USA, 1999.
55. Honaker, J.; King, G. What to do about missing values in time-series cross-section data. *Am. J. Political Sci.* **2010**, *54*, 561–581. [CrossRef]
56. R Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2021.
57. Ramsay, J.O.; Graves, S.; Hooker, G. fda: Functional Data Analysis. 2021. Available online: https://cran.r-project.org/web/packages/fda/fda.pdf (accessed on 17 November 2021).
58. Grenander, U. Stochastic processes and statistical inference. *Ark. Mat.* **1950**, *1*, 195–277. [CrossRef]
59. Rao, C.R. Some statistical methods for comparison of growth curves. *Biometrics* **1958**, *14*, 1–17. [CrossRef]
60. Ramsay, J.O. When the data are functions. *Psychometrika* **1982**, *47*, 379–396. [CrossRef]
61. Ramsay, J.O.; Dalzell, C. Some tools for functional data analysis. *J. R. Stat. Soc. Ser. B (Methodol.)* **1991**, *53*, 539–561. [CrossRef]
62. Rice, J.A.; Silverman, B.W. Estimating the mean and covariance structure nonparametrically when the data are curves. *J. R. Stat. Soc. Ser. B (Methodol.)* **1991**, *53*, 233–243. [CrossRef]

63. Ferraty, F.; Vieu, P. *Nonparametric Functional Data Analysis: Theory and Practice*; Springer Science & Business Media: Berlin/Heidelberg, Germany, 2006.
64. Kokoszka, P.; Reimherr, M. *Introduction to Functional Data Analysis*; Chapman and Hall: London, UK; CRC: Boca Raton, FL, USA, 2017.
65. Gasser, T.; Muller, H.-G.; Kohler, W.; Molinari, L.; Prader, A. Nonparametric regression analysis of growth curves. *Ann. Stat.* **1984**, *12*, 210–229. [CrossRef]
66. Gasser, T.; Kneip, A. Searching for structure in curve samples. *J. Am. Stat. Assoc.* **1995**, *90*, 1179–1188.
67. Wang, J.-L.; Chiou, J.-M.; Müller, H.-G. Review of functional data analysis. *arXiv* **2015**, arXiv:1507.05135.
68. Deville, J.-C. Méthodes statistiques et numériques de l'analyse harmonique. *Annales L'inseé* **1974**, *15*, 3–101. [CrossRef]
69. Ramsay, J.; Silverman, B.W. *Functional Data Analysis*; Springer Series in Statistics; Springer: Berlin/Heidelberg, Germany, 1997.
70. Aguilera, A.M.; Escabias, M.; Preda, C.; Saporta, G. Using basis expansions for estimating functional PLS regression: Applications with chemometric data. *Chemom. Intell. Lab. Syst.* **2010**, *104*, 289–305. [CrossRef]
71. Jacques, J.; Preda, C. Functional data clustering: A survey. *Adv. Data Anal. Classif.* **2014**, *8*, 231–255. [CrossRef]
72. Schmutz, A.; Jacques, J.; Bouveyron, C.; Cheze, L.; Martin, P. Clustering multivariate functional data in group-specific functional subspaces. *Comput. Stat.* **2020**, *35*, 1–31. [CrossRef]
73. Abraham, C.; Cornillon, P.-A.; Matzner-Løber, E.; Molinari, N. Unsupervised curve clustering using B-splines. *Scand. J. Stat.* **2003**, *30*, 581–595. [CrossRef]
74. Serban, N.; Wasserman, L. CATS: Clustering after transformation and smoothing. *J. Am. Stat. Assoc.* **2005**, *100*, 990–999. [CrossRef]
75. Coffey, N.; Hinde, J.; Holian, E. Clustering longitudinal profiles using P-splines and mixed effects models applied to time-course gene expression data. *Comput. Stat. Data Anal.* **2014**, *71*, 14–29. [CrossRef]
76. Peng, J.; Müller, H.-G. Distance-based clustering of sparsely observed stochastic processes, with applications to online auctions. *Ann. Appl. Stat.* **2008**, *2*, 1056–1077. [CrossRef]
77. Li, P.-L.; Chiou, J.-M. Identifying cluster number for subspace projected functional data clustering. *Comput. Stat. Data Anal.* **2011**, *55*, 2090–2103. [CrossRef]
78. Chiou, J.M.; Li, P.L. Functional clustering and identifying substructures of longitudinal data. *J. R. Stat. Soc. Ser. B (Stat. Methodol.)* **2007**, *69*, 679–699. [CrossRef]
79. James, G.M.; Sugar, C.A. Clustering for sparsely sampled functional data. *J. Am. Stat. Assoc.* **2003**, *98*, 397–408. [CrossRef]
80. Bouveyron, C.; Jacques, J. Model-based clustering of time series in group-specific functional subspaces. *Adv. Data Anal. Classif.* **2011**, *5*, 281–300. [CrossRef]
81. Bouveyron, C.; Côme, E.; Jacques, J. The discriminative functional mixture model for a comparative analysis of bike sharing systems. *Ann. Appl. Stat.* **2015**, *9*, 1726–1760. [CrossRef]
82. Jacques, J.; Preda, C. Funclust: A curves clustering method using functional random variables density approximation. *Neurocomputing* **2013**, *112*, 164–171. [CrossRef]
83. Bongiorno, E.G.; Goia, A. Classification methods for Hilbert data based on surrogate density. *Comput. Stat. Data Anal.* **2016**, *99*, 204–222. [CrossRef]
84. Ieva, F.; Paganoni, A.M. Risk prediction for myocardial infarction via generalized functional regression models. *Stat. Methods Med. Res.* **2016**, *25*, 1648–1660. [CrossRef] [PubMed]
85. Kayano, M.; Dozono, K.; Konishi, S. Functional cluster analysis via orthonormalized Gaussian basis expansions and its application. *J. Classif.* **2010**, *27*, 211–230. [CrossRef]
86. Jacques, J.; Preda, C. Model-based clustering for multivariate functional data. *Comput. Stat. Data Anal.* **2014**, *71*, 92–106. [CrossRef]
87. Traore, O.; Cristini, P.; Favretto-Cristini, N.; Pantera, L.; Vieu, P.; Viguier-Pla, S. Clustering acoustic emission signals by mixing two stages dimension reduction and nonparametric approaches. *Comput. Stat.* **2019**, *34*, 631–652. [CrossRef]
88. Bellman, R.E. *Dynamic Programming*; Princeton University Press: Princeton, NJ, USA, 1957.
89. Dempster, A.P.; Laird, N.M.; Rubin, D.B. Maximum likelihood from incomplete data via the EM algorithm. *J. R. Stat. Soc. Ser. B (Methodol.)* **1977**, *39*, 1–22.
90. Delaigle, A.; Hall, P. Defining probability density for a distribution of random functions. *Ann. Stat.* **2010**, *38*, 1171–1193. [CrossRef]
91. Hartigan, J.A.; Wong, M.A. Algorithm AS 136: A k-means clustering algorithm. *J. R. Stat. Society. Ser. C (Appl. Stat.)* **1979**, *28*, 100–108. [CrossRef]
92. Schwarz, G. Estimating the dimension of a model. *Ann. Stat.* **1978**, *6*, 461–464. [CrossRef]
93. Arts, S.; Appio, F.P.; Van Looy, B. Inventions shaping technological trajectories: Do existing patent indicators provide a comprehensive picture? *Scientometrics* **2013**, *97*, 397–419. [CrossRef]
94. Arts, S.; Veugelers, R. Technology familiarity, recombinant novelty, and breakthrough invention. *Ind. Corp. Change* **2015**, *24*, 1215–1246. [CrossRef]
95. Verhoeven, D.; Bakker, J.; Veugelers, R. Measuring technological novelty with patent-based indicators. *Res. Policy* **2016**, *45*, 707–723. [CrossRef]
96. Gassmann, O. Opening up the innovation process: Towards an agenda. *R D Manag.* **2006**, *36*, 223–228. [CrossRef]
97. Noh, H.; Lee, S. What constitutes a promising technology in the era of open innovation? An investigation of patent potential from multiple perspectives. *Technol. Forecast. Soc. Change* **2020**, *157*, 120046. [CrossRef]

98. Chesbrough, H.W. *Open Innovation: The New Imperative for Creating and Profiting from Technology*; Harvard Business Press: Boston, MA, USA, 2003.
99. Chesbrough, H.; Vanhaverbeke, W.; West, J. *Open Innovation: Researching a New Paradigm*; Oxford University Press on Demand: Oxford, UK, 2006.
100. Schmidt, S.; Müller, F.C.; Ibert, O.; Brinks, V. Open region: Creating and exploiting opportunities for innovation at the regional scale. *Eur. Urban Reg. Stud.* **2018**, *25*, 187–205. [CrossRef]
101. Tödtling, F.; Trippl, M. One size fits all? Towards a differentiated regional innovation policy approach. *Res. Policy* **2005**, *34*, 1203–1219. [CrossRef]