Accepted Article

Cluster mapping in Spain: Exploring the correlation between industrial agglomeration and regional performance

Rudy Fernández-Escobedo, Begoña Eguía-Peña**, Leire Aldaz-Odriozola****

Recieved: 16 August 2022
Accepted: 22 January 2024

Abstract:

This paper presents a quantitative cluster mapping methodology for traded industries, adapted for the Spanish case; also explores the correlation between the existence of clusters and regional performance. The study is made at NUTS-2 level, and a total of forty-seven out of eighty-eight 2-digits codes for CNAE-2009 are analyzed; ICT Index and Industry 4.0 Index are also designed and computed. A six-step methodology is applied departing from cross-industry linkages and implementing clustering algorithms; one set of clusters is elected and mapped over territory. The correlation analysis shows that a high number of clusters based on absolute employment data is positively correlated with variables associated with competitiveness, education, ICT adoption, and Industry 4.0, while no significant correlation is found for GDP per capita nor earning per worker.

Keywords: Cluster analysis; agglomeration economics; Industry 4.0; classification methods; industrial agglomeration.
JEL Classification: O18; R12; O52.

Mapeo de clústers en España: Explorando la correlación entre la aglomeración industrial y el desempeño regional

Resumen:

Esta investigación presenta una metodología cuantitativa para el mapeo de clústers en España; también explora la correlación entre la existencia de dichos clústers y el desempeño regional. El estudio se realiza a nivel NUTS-2, y se analizan un total de cuarenta y siete códigos CNAE-2009; asimismo se diseñan y calculan el índice TIC y el índice Industria 4.0. Se aplica una metodología de seis pasos basada en vínculos interindustriales y en un análisis clúster; se selecciona un grupo de clústers y se mapea sobre el territorio. El análisis de correlación muestra que un alto número de clústers basado en datos absolutos de empleo está positivamente relacionado con variables asociadas a la competitividad, educación, adopción de las TIC e Industria 4.0, mientras que no se encuentra correlación significativa con el PIB per cápita ni con las ganancias por trabajador.

Palabras clave: Análisis clúster; economías de la aglomeración; Industria 4.0; métodos de clasificación; aglomeración industrial.
Clasificación JEL: O18; R12; O52.

* Departamento de Políticas Públicas e Historia Económica, Universidad del País Vasco/Euskal Herriko Unibertsitatea (UPV/EHU). Bilbao. España. rfernandez@rfaconsulting.com.mx
** Departamento de Políticas Públicas e Historia Económica, Universidad del País Vasco/Euskal Herriko Unibertsitatea (UPV/EHU). Bilbao. España. bego.eguia@ehu.eus
*** Departamento de Políticas Públicas e Historia Económica, Universidad del País Vasco/Euskal Herriko Unibertsitatea (UPV/EHU. Donostia-San Sebastián. España. leire.aldaz@ehu.eus
Corresponding Author: rfernandez@rfaconsulting.com.mx

1. Introduction

The relevance of industrial agglomeration is undeniable in a highly competitive and complex economy, in which productivity and innovation are key elements looking constantly for fertile ground to flourish (Yelkikalan et al., 2012). Additionally, urbanization and localization have proven to be an essential condition for economic development in the long term (Jofre-Monseny et al., 2014).

There are multiple models of industrial agglomeration. However, the industrial district (Becattini, 1990) and the industrial cluster (Porter, 1990) have been particularly popular for the last three decades, while the former has reached high levels of institutionalization in Europe and US (Ortega-Colomer et al., 2016).

The efforts for empirically identifying such agglomerations over territory have led to the development of mapping tools, as an effort to help policy makers, industrials, and practitioners to understand and capitalize the industrial agglomeration phenomenon. The largest institutional efforts in this matter are the Cluster Mapping Project directed by the Institute of Competitiveness (US), and the European Cluster Collaboration Platform sponsored by the European Observatory for Clusters and Industrial Change (Europe). There are also national efforts for mapping industrial districts departing from manufacturing industries (Lorenzini & Lombardi, 2018).

However, while the Cluster Mapping Project departs from Cluster Category Definitions (CCD)[1] derived from an empirical methodology designed to identify cross-industry linkages across the US economy, the European Observatory for Clusters departs from the homologation of US cluster definitions for the European context (Ketels & Protsiv, 2021; Szanyi et al., 2010), assuming industrial and environmental heterogeneity between EU countries and US (Brodzicki, 2010). Moreover, the mapping of industrial districts relies on Local Labor Markets (LLM) as territorial units (Boix & Trullén, 2010), which are not harmonized for all European countries.

This represents a relevant gap in the literature for Europe, since a comprehensive cluster mapping initiative should develop a quantitative methodology based on common data, methodology, and literature, capable of being implemented in a comprehensive way across any particular economy to identify specific CCD for the geographic region being analyzed (Ketels, 2017).

Is it possible to complement the existing efforts of cluster mapping at a national level through the implementation of a comprehensive and quantitative methodology using domestic raw data? This paper pretends to tackle that research question testing the methodology of Delgado et al. (2016) over Spain, not only because such country has been object of multiple institutional efforts to implement industrial agglomeration policies (Ortega-Colomer et al., 2016), but also because there are previous exercises of industrial agglomeration mapping that suggest sufficient data for the analysis (Boix & Galletto, 2009). Furthermore, Spain brings the opportunity to test the methodology in a country with different geographical and industrial structure when compared to other advanced economies like US and Germany (ICEX España Exportación e Inversiones, n.d.)

Since this is the first time such methodology is fully applied using domestic raw data outside US, the paper aims to: (I) present a robust cluster analysis methodology for the Spanish context to create domestic CCD and a cluster map; (II) discuss the methodological implications of the research and its differences with other exercises of cluster identification; and (III) explore the correlation between the existence of clusters and multiple economic variables. Besides, two indexes are built to summarize the regional adoption of ICT (ICT Index) and the regional adoption of technologies associated to Industry 4.0 (Industry 4.0 Index); this is the first time such regional analysis is made for Spain, helping to fill another gap in literature.

The remainder of the paper is structured as follows. The first section presents a theoretical background for the industrial cluster concept, common methodologies for cluster mapping, and externalities of this phenomenon. The second section of the article presents the quantitative methodology

[1] In the context of cluster mapping initiatives, a Cluster Category Definition is a brief description of a group of industries that share different linkages related to employment, know-how, and value-chain, among others.

implemented for the cluster mapping exercise. The third and fourth parts present results and discusses them in the frame of previous research, respectively. Finally, main conclusions and limitations are presented, together with implications for cluster scholars and policymakers.

2. Theoretical background

Academics make broad efforts to consolidate empirical and theoretical literature about industrial agglomeration, its causes, identification, and effects. However, for the last thirty years, the concept of industrial cluster has reached a high level of popularity and institucionalization around the world, becoming a central element for industrial policy and creating a common language for regional development that could not be matched by other related concepts (Babkin et al., 2017; Hermans, 2021; Ortega-Colomer et al., 2016; Skokan & Zotyková, 2014).

In the next paragraphs, the paper presents literature framed by previous research about the industrial cluster, its externalities, and mapping methodologies.

2.1. The industrial cluster concept

Although the seminal work of Marshall (1920) laid the foundations of the cluster concept, it did not reached relevance among researchers and policy makers until the 90's, influenced mainly by the research of Becattini (1990), Krugman (1991), and Porter (1990).

Since then, this idea has been evolving from the basic viewpoints of networking and competitiveness to most complex and multidisciplinary approaches like knowledge management and the triple helix of innovation (Caloffi et al., 2018). Moreover, the cluster has adopted ideas or even competed with other models of industrial agglomeration; such is the case of the industrial district concept, from which the industrial cluster adopted its socio-economic and geographical notion (Sforzi, 2015).

In its current form, the industrial cluster concept rests on geographical, economic, competitive, and sociologic foundations (Jofre-Monseny et al., 2014) (Figure 1). Furthermore, literature reveals that the historical foundations also play a relevant role in the cluster genesis and evolution when studied under the path-dependence model (Elola et al., 2012; Zhu & Pickles, 2016).

Therefore, industrial clusters can be defined as groups "of companies and institutions geographically concentrated, whose relationships have as main characteristics the collaboration and exchange of resources, which implies a high cognitive proximity among actors" (Tavares et al., 2021, p 193).

Finally, although there are multiple definitions for the cluster, all of them fit the idea of a geographic space where economics of agglomeration manifest themselves among related organizations (Delgado et al., 2016).

Figure 1.
Foundations of the industrial cluster concept

Geographic closeness		
Economics	Competitiveness	Sociology/History
• Input-output linkages. • Labor market pooling. • Knowledge spillovers.	• Cost of economic factors. • Demand conditions. • Strategy and rivalry in the industry. • Support organizations. • Economic policies.	• Path dependence. • Social context and networking. • Legal framework. • Luck. • Mutual confidence and reciprocity.

Source: Authors' elaboration based on reviewed literature.

2.2. The externalities of industrial clusters

The conceptual heterogeneity of clusters, added to the difficulty to establish their geographic delimitations and fully identify valuable networks and participants, makes it difficult for researchers to generalize empirical findings about the impact of clusters on economic development. Skokan and Zotyková (2014) raise the next question as one of the most important for the study of clusters: how to measure the benefits of clusters on economy?

The most influential studies about the positive impact of clusters on economy are focused on innovation (Delgado, Porter, et al., 2014; Tavares et al., 2021; Ybarra & Domenech-Sanchez, 2012), showing that the access of cluster members to specialized inputs, skilled labor, market intelligence, and supportive infrastructure, has a positive effect on such variable.

Likewise, there are empirical evidence about positive externalities related to the improvement of competitiveness, productivity, salaries, unemployment, and GDP. Slaper et al. (2018) found that regions with high prevalence of industrial clusters outperformed regions with low prevalence of them in variables like GDP per capita, wage level and total income per worker. Similarly, Babkin et al. (2018) observed a positive and significant relation between the existence of industrial agglomeration phenomena and competitiveness.

Empirical studies also show that clusters, as innovation networks, enhance collaboration among government, industry, and research centers, creating more stable and less uncertain business environments in which digital transformation and Industry 4.0 have better probabilities to evolve and improve the innovation capabilities (Babkin et al., 2018; Fernandez-Escobedo et al., 2023; Götz & Jankowska, 2017; Grashof et al., 2021; Jasinska & Jasinski, 2019; Vlaisavljevic et al., 2020). Furthermore, research made on different models of industrial agglomeration has reached similar results (Hervás-Oliver, 2021).

However, the conclusions about cluster externalities are far for being definitive. Research shows that the life-cycle stage of clusters moderates the externalities of such agglomeration phenomenon (Elola et al., 2017; Skokan & Zotyková, 2014). Additionally, studies have shown that clusters can fall into technological lock-in, affecting the competitiveness of regions and industries (Elola et al., 2012; Zhu & Pickles, 2016). The best-known negative externality is what some authors call congestion costs, which implies the cost increase of key resources for cluster members, provoking diminishing returns and hurting entrepreneurship, competitiveness, and firm performance (Delgado, Porter, et al., 2014; Grashof & Fornahl, 2021; Slaper et al., 2018).

To conclude, it is important to mention that despite the challenges found by researchers to assess the effects of industrial clusters on economy and their actors, the findings about the positive effect on innovation and productivity tend to be more consistent in clusters that involve high-tech and traded industries, compared with low-tech and local industries (Bathelt & Li, 2014; Grashof & Fornahl, 2021; Slaper & Ortuzar, 2015; Tavares et al., 2021).

2.3. Methodologies for identification of industrial clusters

Researchers have developed multiple tools and approaches to build methodologies for clusters identification. Between the top-down methods and the bottom-up methods, the former fit better the needs of cluster mapping initiatives (Hermans, 2021; Ketels, 2017) as those methods have a quantitative approach based on statistical modeling, and are broadly applicable with nationwide/multi-industry scope.

The top-down methods depend on the definition of specific territories (spatial units for study); once studies define such units, the methodologies aim to analyze data in search of geographical concentration of industries and cross-industry linkages.

The main tools for identifying industrial agglomeration are the indexes and location quotients (LQ). Ellison et al. (2010) proposed an index of industry concentration which have suffered from multiple revisions and adaptations for cluster mapping projects; similarly, the Gini coefficient is another index adapted to measure industrial agglomeration (Burki & Khan, 2011). The LQ is also a popular measure to explore agglomeration; this one revolves around the employment specialization of regions when compared

with others (Slaper et al., 2018). The central limitation of those tools is that they only can be used on well specified industries or group of industries, which make them useless to find cross-industry linkages.

In the case of cross-industry linkages identification, there are tools and methodologies that depart from Marshallian micro-foundations of agglomeration; among them, it is worth mentioning the next ones.

The quantitative input-output analysis (QIOA) was developed to capture linkages related to flow of goods and services, departing from the study of Input-Output matrices (Oosterhaven et al., 2001; Titze et al., 2011). Similarly, the cross-industry patent citations measures and technology-flow matrices were developed to identify agglomeration patterns for knowledge linkages among industries (Ellison et al., 2010; Scherer, 1984). These tools are commonly limited for the availability of the data and the disaggregation level of it.

Most robust methodologies include the locational correlation (LC) analysis and the Sforzi-ISTAT methodology. The first one is capable of combining multiple approaches and capturing cross-industry linkages related to co-location, labor market pooling, input-output relations, and knowledge-flow, and it is the base of contemporaneous cluster mapping efforts (Diodato et al., 2018). However, it is limited for the quality/quantity of the data and is not capable of finding agglomeration patterns by itself. The second one is based on industrial district's literature and departs of the identification of LLM and the definition of the groups of economic activities, which should be made previous to the analysis (Boix & Galletto, 2009). Nevertheless, while the methodology can find agglomeration patterns, it is limited by the need of a harmonized LLM structure for different countries and the ex ante aggregation of industries, which reduces its flexibility and its capacity to find complex cross-industry linkages.

Finally, state-of-the-art methodologies combine multiple of these methods with algorithms of cluster analysis based on Ward's linkage, finding agglomeration patterns and cross-industry linkages at the same time, providing the needed data to create appropriate CCD for specific territories (Delgado et al., 2016). Unfortunately, such methodologies tend to use administrative divisions as spatial units for study, missing the rationale of community that shapes the concept of LLM, which is at the heart of the industrial district mapping (Canello & Pavone, 2016).

Although the presented tools and methods have the mentioned limitations, researchers recognize their valuable potential for cluster mapping, particularly when they are combined, and their results are used for comparison purposes.

3. Methodology

This empirical research has an exploratory, descriptive, non-experimental, and cross-sectional design with a quantitative approach, using the statistical technique known as cluster analysis. The research also uses the Pearson correlation coefficient to explore correlation between pairwise industries, and among CCD and multiple macroeconomic variables.

The presented methodology is focused on traded industries (Delgado, Bryden, et al., 2014) and based on the work of Delgado et al. (2016) which describes the current algorithm used by the Cluster Mapping Project to establish CCD in US.

The analysis is based on the statistical classification of economic activities for Spain (known as CNAE-2009) at 2-digits level and uses autonomous communities as spatial units to analyze data (NUTS-2), excluding Ceuta and Melilla. These decisions are made for two reasons: first, to avoid as much as possible data suppression from the Spanish Statistical Office; and second, to avoid finding artificially high LC across many industries if small regions with low industrial representations are used (Porter, 2003).

The method follows multiple steps: to build the datasets which are arranged as similarity matrices (step one); to build and assess the groups of clusters (steps second to fourth); and to choose the highest quality group of clusters and project it over Spanish territory (steps five and six).

3.1. Data and sources

This research uses multiple open databases from the Spanish Statistical Office, the Spanish Patent and Trademark Office, and the European Commission[2].

A total of 47 out of 88 2-digits codes for CNAE-2009 are analyzed[3]. The first group of data used for cluster analysis includes:

- Statistical structure for business – commerce, industry, and services (year 2019, CNAE-2009 2-digits, NUTS-2).
- Annual national accounting – input-output matrix (year 2016 – rev. 2019).
- Annual national accounting – origin-destination matrices (years 2010 to 2018).
- The labor force survey (year 2011, CNO-2011 2-digits, CNAE-2009 2-digits).

The second group of data is used to explore correlation between CCD and macroeconomic variables, and includes:

- For economic development:
  - The regional accounting for the real GDP per capita (year 2019).
- For population and employment:
  - The labor force survey for regional active population and for regional unemployment rate (average for all four quarters of 2019).
  - The wage structure survey for total income per worker (year 2019).
  - The educational attainment survey for adults with professional education or more (year 2016).
- For innovation:
  - The regional patent application per million inhabitants as innovative activity (average 2018-2019).
- For competitiveness:
  - The Regional Competitiveness Index (RCI) for sub-index "basic" (year 2019).
- For ICT and Industry 4.0:
  - The regional survey on the use of Information and Communication Technologies (ICT).
  - eCommerce in enterprises with more than ten employees (years 2017, 2019, 2020 and 2021, depending on the specific item since different data is collected each year).

3.2. Step one: Building the similarity matrices

Similarity matrices M_{ij} provide information about the relatedness between pairs of industries i and j. To build a unidimensional matrix, it is required to transform one or more indicators into a single

[2] Data were retrieved from https://www.ine.es for economy, labor market and ICT; http://consultas2.oepm.es/ipstat/faces/IpsBusqueda.xhtml for industrial property, and https://ec.europa.eu/regional_policy/es/information/maps/regional_competitiveness/ for competitiveness.

[3] Information for 21 codes was not available by the Spanish Statistical Office; another 31 codes were grouped into 11 provisional codes to homologize the CNAE-2009 with the industrial classification of the input-output matrix. Due to statistical confidentiality, there is incomplete information for specific industries in particular regions; this data was disregarded.

similarity measure; multidimensional matrices are built combining similarity measures from unidimensional matrices.

The indicators and measures used in this research are chosen to capture as many cross-industry linkages as possible (e.g., knowledge, skills, supply, or demand links). Table 1 shows the specifications of each matrix built.

Table 1.
Similarity matrices used to generate sets of CCDs

Similarity matrix	Indicators used	Measure computed
Unidimensional matrices		
Co-location pattern for employment (LC_Emp)	Employment size of industry *i* and *j* in region *r*	Locational correlation of employment [-1, 1]
Co-location pattern for establishments (LC_Est)	Establishments of industry *i* and *j* in region *r*	Locational Correlation of establishments [-1, 1]
Geographic concentration of employment (COI)	Employment size of industry *i* and *j* in region *r*	Co-agglomeration Index
Input-Output Links (IO)	Inputs of industry *i* coming from *j*, and outputs of industry *i* going to *j*	Average share of inputs of industry i coming from j, outputs of industry i going to j, and vice versa [0, 1]
Labor Occupation Links (Occ)	Employment size of industry *i* and *j* related to occupation *k*	Occupational correlation [-1, 1]
Multidimensional matrices		
Co-location pattern (LC)	Locational correlation of employment, and locational correlation of establishments	Average of LC_Emp and LC_Est
Co-location pattern and Geographic concentration of employment (LC_COI)	Locational correlation of employment, locational correlation of establishments, and Co-agglomeration Index	Average of (standardized) LC_Emp, LC_Est, and COI
Geographic concentration of employment, Input-Output Links, and Labor Occupation Links (COI_IO_Occ)	Co-agglomeration Index, average share of input-output links, and occupational correlation	Average of (standardized) COI, IO, and Occ
All unidimensional measures (ALL)	All unidimensional measures	Average of (standardized) LC_Emp, LC_Est, COI, IO, and Occ

Source: Authors' elaboration.

3.3. Step two: Identifying traded industries and adjusting similarity matrices

While local industries serve local markets, traded industries are those that produce goods and services that are either exported or sold across regions. Since this research is focused on traded industries (both natural-resource-based and not), it is necessary to identify them and remove local ones from the similarity matrices.

A multi-criterion methodology is applied to assess the 47 CNAE-2009 2-digit industries and find traded industries. For this multi-criterion methodology, the set of traded industries includes all those industries classified as traded by both the gross value-added ratio methodology and the locational Gini Coefficient methodology:

- The export to gross value-added ratio (Mano & Castillo, 2015), based on a single cutoff set by literature, using the average of 2010-to-2018 ratios to reduce overrepresentation of external shocks.
- The locational Gini Coefficient (Carlino & Kerr, 2015), based on a single cutoff set by authors of this research.

3.4. Step three: Setting parameters and running clustering functions

In this step, the following parameters (β) are used: clustering functions are run over raw data as each similarity matrix is built with a common internal scale; starting values for clustering functions are chosen at random; and multiple number of clusters (*numc*) are set, going from seven to 13[4], when functions are run.

Two clustering functions (*F*) for continuous data are used in this research (Delgado et al., 2016; Everitt et al., 2011; Grimmer & King, 2011): the hierarchical function of Ward's minimum variance (squared Euclidean distance) (*H*), and the centroid based function (kmean) (*K*).

Before running clustering functions over the similarity matrices, the algorithm is tested and validated following the method of Delgado et al. (2016), using an artificial similarity matrix based on the first digit of the CNAE-2009 2-digits code for the traded industries.

Let *C* be a single group of clusters given *F* and β, then:

$$C = F(M_{ij}, \beta) \tag{1}$$

The clustering algorithm is run over all nine similarity matrices, using all possible combinations of parameters.

3.5. Step four: Assessing quality of *C*s through Validation Scores

Validation Scores (*VS*) are computed for each *C*, following Delgado et al. (2016) methodology; *VS* are the average of two partial validations scores*: VS-Cluster* and *VS-Industry*. All five unidimensional matrices *Mij* are used to build the validation scores, since the capture of different industry interdependencies is assumed for each of them; a single similarity measure between *i* and *j* represents a relatedness measure.

On the one side, *VS-Cluster* measures whether individual clusters (*c*) in *C* are meaningfully different from each other, and it is made up of two averaged sub-scores. These sub-scores depart from the Within Cluster Relatedness for *c* (WCR_c) measure (as the average relatedness between pairs of industries within a *c*), and the Between Cluster Relatedness for *c* (BCR_c) measure (as the average relatedness between industries in *c* and those in another cluster). *VS-Cluster*'s sub-scores are expressed as follows:

$$VS - Cluster\ Average_C^M = \frac{\sum_c I[WCR_c(M_{ij}) > AvgBCR_c(M_{ij})]}{N_c} * 100 \tag{2}$$

$$VS - Cluster\ Percentile95_C^M = \frac{\sum_c I[WCR_c(M_{ij}) > Pctile95BCR_c(M_{ij})]}{N_c} * 100 \tag{3}$$

where N_c is the number of clusters in *C,* and *I* is an indicator function equal to 1 for a given cluster *c* which met the condition expressed inside brackets.

On the other hand, *VS-Industry* measures whether individual industries (*i*) in *C* are more related to the industries within its own *c* than to industries outside its cluster, and it is also made up of two averaged

[4] As the analysis is based on CNAE-2009 2-digits codes with 27 traded industries, working with numbers of clusters greater than 13 would have increased the chances for the appearance of multiple one-industry groups; the minimum number of clusters is set following to Delgado et al. (2016) who set the minimum number of clusters as the half of the maximum number chosen.

sub-scores. These sub-scores depart from the Within Cluster Relatedness for *i* in *c* (WCR_{ic}) measure (as the average relatedness between *i* and other industries within a *c*), and the Between Cluster Relatedness for *i* in *c* (BCR_{ic}) measure (as the average relatedness between *i* and those in another cluster). *VS-Cluster*'s sub-scores are expressed as follows:

$$VS - Industry\ Average_C^M = \frac{\sum_i I[WCR_{ic}(M_{ij}) > AvgBCR_i(M_{ij})]}{N_i} * 100 \quad (4)$$

$$VS - Industry\ Percentile95_C^M = \frac{\sum_i I[WCR_{ic}(M_{ij}) > Pctile95BCR_i(M_{ij})]}{N_i} * 100 \quad (5)$$

where N_i is the number of industries in *C*.

3.6. Step five: Choosing the *C*s with higher quality and setting CCD

The *C* with the highest position in the *VS* rank (let us call it *C**) is elected to create CCD at the regional level (NUTS-2). CCD are defined arbitrarily for each *c*, looking at the industries which configure each cluster and aiming to suggest names easy to assimilate for researchers, policy makers, and development practitioners.

3.7. Step six: Finding the territorial presence of clusters over Spain

Since each *C* is configured by a set of *c*s, this step is about finding the presence of each *c* over the analyzed regions (spatial units of study).

The US's Cluster Mapping Project recognizes three types of clusters presence over territory based on employment share and location quotients (Delgado et al., 2016; Ketels, 2017): clusters by top employment specialization (TESp), clusters by top employment share (TESh), and clusters by top employment specialization & share (TESS). The results of the analysis of territorial presence are presented in the Results section.

3.8. The correlation analysis

Finally, after exploring the territorial presence of *c*, correlation analysis is made among cluster presence and multiple variables.

The presence of each *c* over regions is arranged as a discrete dichotomous variable (1-0, the cluster is present or not). Also, the total count of *c* (by TESp, TESh, and TESp) in each territory is considered.

Multiple variables are selected to run the Pearson's correlation analysis against the presence of clusters. Variables election is based on the work of Delgado et al. (2014) and Slaper et al. (2018); the calculation and introduction of ICT and Industry 4.0 indexes is a novelty introduced in this research. To build those Indexes, multiple measures are considered following literature about ICT and Industry 4.0 impact on business (Almeida et al., 2020; Atik & Ünlü, 2019; Maresova et al., 2018).

On the one hand, the ICT Index groups ten different measures related to the use of computers, Internet connection, webpage, social networks, ERP, CRM, electronic communications, eGovernment, eSignature, and cybersecurity. On the other hand, the Industry 4.0 Index groups six different measures related to the use of: industrial robots, big data, cloud computing, 3D printing, Internet of things, and artificial intelligence. The grouping methodology for both indexes is based on the World Economic Forum (WEF) (Atik & Ünlü, 2019).

It is also relevant to point out that for competitiveness the RCI basic sub-index is chosen due to the full RCI is configured also by another two sub-indexes (efficiency and innovation) which are highly correlated with other variables chosen for this research, such as population, educational attainment, innovation activity and ICT adoption.

Correlations are presented in the Results section.

4. Results

Descriptive statistics are obtained for each similarity matrix (Table 2). The correlation among all the similarity matrices seems to be significant at 1% level, except for Occ with LC_Emp, LC_Est, and LC.

Table 2.
Descriptive statistics for similarity matrices; 47 industries (CNAE-2009 2-digits codes) and N=2,167

Similarity Matrices M_{ij}	Mean	Median	Std. Dev.	Min.	Max
LC_Emp	0.672	0.767	0.300	-0.540	1.000
LC_Est	0.743	0.822	0.265	-0.914	0.998
COI	-0.001	0.000	0.021	-0.089	0.136
IO	0.014	0.007	0.023	0.000	0.343
Occ	0.130	0.050	0.220	-0.128	0.973
LC	0.708	0.782	0.267	-0.658	0.999
LC_COI	-0.029	0.150	0.780	-3.962	2.613
COI_IO_Occ	-0.074	-0.169	0.530	-1.403	2.847
ALL	-0.053	0.018	0.548	-2.186	1.876

Note: An observation is any pair of industries (ij, $i \neq j$). All unidimensional matrices are based on 2019 data except for IO and Occ, which is based on 2016 and 2011 data, respectively.
Source: Authors´ elaboration.

Using the similarity matrices, a single set of traded industries is configured (Table 3), meeting two key attributes: the exclusion of industries that conceptually are classified as local (e.g., real state, retail, local transportation, and sewerage), and the improving of the correlation between similarity matrices of traded industries when compared with correlation between similarity matrices for all industries. 27 out of 47 industries are categorized as traded.

The cluster algorithm is applied over the nine similarity matrices of traded industries, and 126 *Cs* are obtained (the number is equal to all combinations among F, β and M_{ij}). The quality of individual *Cs* is assessed through the *VS* (Table 4).

Table 3.
List of 27 out of 47 CNAE-2009 2-digit codes classified as traded industries

Code	Description
IN05	Groups: Mining of coal and lignite; Extraction of crude petroleum and natural gas; Mining of metal ores; Other mining and quarrying; Mining support service activities
IN10	Groups: Manufacture of food products; Manufacture of tobacco products
IN13	Groups: Manufacture of textiles; Manufacture of wearing apparel; manufacture of leather and related products
IN16	Manufacture of wood and of products of wood and cork, except furniture; mfg. of articles of straw and plaiting materials
IN17	Manufacture of paper and paper products
IN19	Manufacture of coke and refined petroleum products
IN20	Manufacture of chemicals and chemical products
IN21	Manufacture of basic pharmaceutical products and pharmaceutical preparations
IN22	Manufacture of rubber and plastic products

TABLE 3. CONT.
List of 27 out of 47 CNAE-2009 2-digit codes classified as traded industries

Code	Description
IN23	Manufacture of other non-metallic mineral products
IN24	Manufacture of basic metals
IN25	Manufacture of fabricated metal products, except machinery and equipment
IN26	Manufacture of computer, electronic and optical products
IN27	Manufacture of electrical equipment
IN28	Manufacture of machinery and equipment n.e.c.
IN29	Manufacture of motor vehicles, trailers, and semi-trailers
IN30	Manufacture of other transport equipment
IN31	Groups: Manufacture of furniture; Other manufacturing
IN50	Water transport
IN51	Air transport
IN58	Publishing activities
IN59	Groups: Motion picture, video and television programme production, sound recording and music publishing activities; Programming and broadcasting activities
IN61	Telecommunications
IN62	Groups: Computer programming, consultancy, and related activities; Information service activities
IN72	Scientific research and development
IN73	Advertising and market research
IN79	Travel agency, tour operator and other reservation service and related activities
	Traded Industries to Total Industries ratio **57.4%**

Source: Authors´ elaboration.

TABLE 4.
Validation scores (*VS*), partial validation scores (*VS-Cluster* and *VS-Industry*) and sub-scores (VS-Cluster Avg, VS-Cluster Pctile95, VS-Industry Avg, VS-Industry) for the ten highest-ranked groups of clusters (*Cs*)

Rank (*VS*)	*VS*	Method	Similarity Matrix M_{ij}	Numc	*C* code	Rank (VS-Cluster)	VS-Cluster	VS-Cluster Avg	VS-Cluster Pctile95	Rank (VS-Industry)	VS-Industry	VS-Industry Avg	VS-Industry Pctile95
1	72.9	H	ALL	7	H-ALL-7	4	72.9	80.0	65.7	1	73.0	90.4	55.6
2	72.6	K	COI_IO_Occ	8	K-COI_IO_Occ-8	1	76.3	85.0	67.5	3	68.9	88.9	48.9
3	71.6	H	COI_IO_Occ	7	H-COI_IO_Occ-7	2	74.3	88.6	60.0	3	68.9	88.9	48.9
4	70.3	K	COI_IO_Occ	7	K-COI_IO_Occ-7	3	72.9	74.3	71.4	5	67.8	83.7	51.9
5	66.6	K	Occ	7	K-Occ-7	4	72.9	94.3	51.4	17	60.4	91.1	29.6
6	64.8	K	COI_IO_Occ	9	K-COI_IO_Occ-9	7	63.3	75.6	51.1	7	66.3	87.4	45.2
7	63.3	K	ALL	9	K-ALL-9	9	58.9	64.4	53.3	4	67.8	84.4	51.1
8	62.4	H	COI	9	H-COI-9	8	60.0	71.1	48.9	10	64.8	85.2	44.4
9	62.3	H	Occ	7	H-Occ-7	5	67.1	88.6	45.7	23	57.4	87.4	27.4
10	62.1	K	ALL	7	K-ALL-7	13	57.1	65.7	48.6	6	67.0	86.7	47.4

Notes: Rank shows the relative position of *C* compared with the others when considering the relevant score. For *VS-Cluster* and *VS-Industry* some scores are equal, so the ranks are too. H and K represent the clustering function used (hierarchical and kmean, respectively).
Source: Authors´ elaboration.

Table 5.
Cluster Category Definitions (CCD) and list of industries (by code) configuring each cluster c for *C

c number	CCD	Industry codes
01	Extraction, mining, and agro-industrial cluster	IN05
		IN10
		IN16
		IN23
		IN31
02	Packaging, covers and lining – manufacturing cluster	IN13
		IN17
		IN20
		IN22
03	Fuel and multipurpose vehicles – manufacturing cluster	IN19
		IN30
04	Biotechnological cluster	IN21
		IN26
		IN72
05	Electromechanical and automotive cluster	IN24
		IN25
		IN27
		IN28
		IN29
06	Water-travel cluster	IN50
07	Tourism, ICT, and creativity – services cluster	IN51
		IN58
		IN59
		IN61
		IN62
		IN73
		IN79

Source: Authors' elaboration.

The presence of clusters in regions is presented in Table 6, distinguishing among clusters presence by top employment specialization (TESp), by top employment share (TESh), and by top employment specialization & share (TESS). As shown, Catalonia stands out reaching the maximum number of clusters by TESh. Contrastingly, the number of clusters by TESp is more evenly distributed among regions. Besides, the number of clusters by TESS is smaller since it combines both previous criteria.

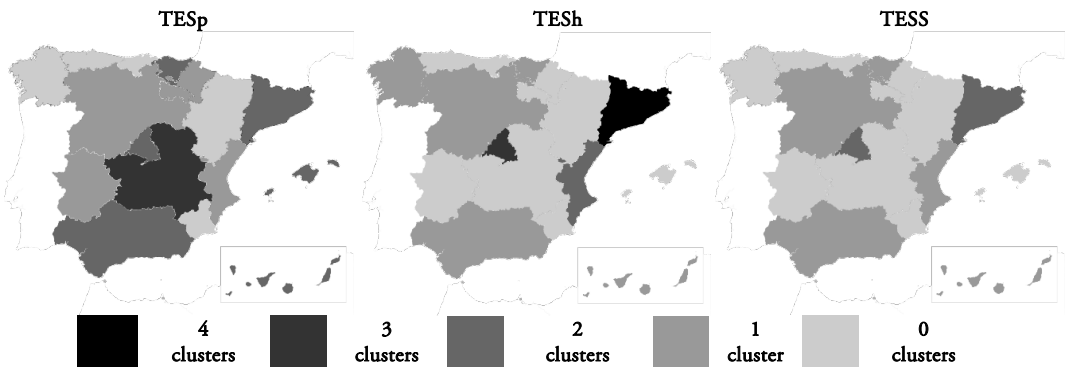
TABLE 6.
Clusters presence by autonomous community (C* set)

	01	02	03	04	05	06	07	TESp	TESh	TESS
Andalusia			***				*	2	1	1
Aragon								0	0	0
Asturias, Principality of								0	0	0
Balearic Islands						*	*	2	0	0
Basque Country			*		***			2	1	1
Canary Islands						***	*	2	1	1
Cantabria								0	0	0
Castile and León	***							1	1	1
Castilla – La Mancha	*	*	*					3	0	0
Catalonia		***		***	**		**	2	4	2
Extremadura	*							1	0	0
Galicia	**							0	1	0
Madrid, Community of			**	***			***	2	3	2
Murcia, Region of								0	0	0
Navarre, Ch. Community of					*			1	0	0
Rioja, La		*						1	0	0
Valencian Community		***				**		1	2	1
Total								**20**	**14**	**9**

Note: The table distinguish clusters presence by top employment specialization (TESp) (*), by top employment share (TESh) (**), and by top employment specialization & share (TESS) (***).
Source: Authors´ elaboration.

Multiple maps can be drawn departing from the result of this research. For example, **Figure 2** shows the intensity of clusters presence by TESp, TESh, and TESS over regions; it draws attention that regions with high population concentration show a high presence of industrial clusters.

FIGURE 2.
Intensity of clusters presence by TESp, TESh, and TESS over autonomous communities (based on *C)**



Source: Authors´ elaboration.

Finally, descriptive statistics are obtained for variables classified as economic development, population and employment, innovation, competitiveness, ICT, and Industry 4.0 (Table 7). ICT Index shows positive and significant correlation with nine out of ten measures grouped (the correlation with social networks is positive but not statistically significant). Industry 4.0 Index shows positive and

significant correlation with five out of six measures grouped (the correlation with use of industrial robots is positive but not statistically significant). Table 8 shows regional performance for both ICT Index and Industry 4.0 Index.

To conclude, full correlation matrix is computed (Table 9).

Table 7.
Descriptive Statistics for autonomous communities' variables (N=17)

Categories	Variables	Mean	Median	Std. Dev.	Min.	Max
ECON.	GDP per capita (euros)	24808.773	23197.379	4930.420	18275.749	34805.061
	Earning per worker (euros)	23642.193	22877.130	2627.158	19940.680	29476.210
	Natural resources dependency	0.040	0.043	0.018	0.012	0.074
POP. &	Population (miles)	2760.900	2038.700	2558.826	314.400	8448.200
EMP.	% Population with a grade or more	0.143	0.136	0.032	0.103	0.231
	Unemployment rate	0.133	0.118	0.042	0.082	0.215
INNOV.	Patent application to million inhab. ratio	29.147	28.500	15.672	7.000	66.000
COMP.	RCI basic sub-index	-0.070	-0.078	0.138	-0.213	0.302
ICT	ICT Index	0.548	0.576	0.164	0.245	0.829
IND. 4.0	Industry 4.0 index	0.466	0.435	0.170	0.212	0.808

Source: Authors´ elaboration.

Table 8.
Regional ICT Index and Industry 4.0 Index

Region	ICT Index	Industry 4.0 Index
Andalusia	0.605	0.373
Aragon	0.648	0.515
Asturias, Principality of	0.526	0.456
Balearic Islands	0.497	0.243
Basque Country	0.588	0.579
Canary Islands	0.332	0.212*
Cantabria	0.245*	0.615
Castile and León	0.454	0.435
Castilla – La Mancha	0.365	0.392
Catalonia	0.829**	0.723
Extremadura	0.321	0.392
Galicia	0.604	0.434
Madrid, Community of	0.800	0.808**
Murcia, Region of	0.572	0.349
Navarre, Ch. Community of	0.576	0.630
Rioja, La	0.631	0.226
Valencian Community	0.715	0.539

Note: ** Highest score. *Lowest score.
Source: Authors´ elaboration.

TABLE 9.
Correlation between prevalence of clusters (C*) and selected variables (N=17).

	TESp	TESh	TESS	01	02	03	04	05	06	07	GDP per capita (euros)	Earning per worker (euros)	Natural resources dependency	Population (miles)	Share of population with a Grade or more	Unemployment rate	Patent application to million inhab. ratio	RCI basic sub-index	ICT Index	Industry 4.0 index
TESp	1.000																			
TESh	0.362	1.000																		
TESS	.496*	.925**	1.000																	
01	0.044	-0.156	-0.223	1.000																
02	0.345	0.326	0.176	0.019	1.000															
03	.645**	0.206	0.375	0.019	0.019	1.000														
04	0.326	.850**	.772**	-0.203	0.228	0.228	1.000													
05	0.246	0.339	0.313	-0.257	0.107	0.107	0.310	1.000												
06	0.246	0.071	0.091	-0.257	0.107	-0.257	-0.169	-0.214	1.000											
07	.576*	.548*	.622**	-0.358	-0.054	0.251	.566*	0.040	0.378	1.000										
GDP per capita	0.160	0.416	0.416	-0.405	-0.011	0.208	.568*	.603*	-0.174	0.169	1.000									
Earning per worker	0.158	0.343	0.375	-0.461	-0.096	0.364	0.466	.727**	-0.284	0.092	.894**	1.000								
Natural resources dependency	-0.254	-0.451	-.571*	.709**	0.025	-0.189	-0.475	-.522*	-0.174	-0.368	-.66**	-.68**	1.000							
Population	0.437	.820**	.817**	-0.159	0.219	0.464	.645**	0.134	0.008	.643**	0.124	0.151	-0.268	1.000						
% Population with a grade or more	0.239	.579*	.623**	-0.294	-0.002	0.459	.635**	0.406	-0.162	0.186	.830**	.833**	-.64**	0.403	1.000					
Unemployment rate	0.244	-0.105	-0.005	0.261	-0.070	0.127	-0.224	-0.436	0.240	0.264	-.79**	-.68**	0.348	0.216	-.557*	1.000				

 ISSN: 1695-7253 e-ISSN: 2340-2717

TABLE 9. CONT.
Correlation between prevalence of clusters (C*) and selected variables (N=17).

	TESp	TESh	TESS	01	02	03	04	05	06	07	GDP per capita (euros)	Earning per worker (euros)	Natural resources dependency	Population (miles)	Share of population with a Grade or more	Unemployment rate	Patent application to million inhab. ratio	RCI basic sub-index	ICT Index	Industry 4.0 index
Patent application to million inhab. ratio	-0.354	0.079	0.045	-0.398	-0.074	0.008	0.159	0.351	-0.334	-0.282	.593*	.577*	-.497*	0.020	.532*	-.62**	1.000			
RCI basic sub-index	0.301	.556*	.544*	-0.199	0.269	0.456	.763**	0.303	-0.307	0.157	.673**	.615**	-.565*	0.359	.764**	-0.467	0.441	1.000		
ICT Index	0.057	.662**	.543*	-0.389	0.304	0.146	.612**	0.339	-0.095	0.264	.592*	.505*	-.488*	.600*	.642**	-0.406	.555*	0.476	1.000	
Industry 4.0 index	-0.017	.593*	.518*	-0.178	0.014	0.243	.665**	.502*	-0.379	0.023	.630**	.691**	-0.456	0.411	.755**	-.493*	.589*	.763**	0.470	1.000

Note: *Coefficients are significant at 5% level. **Coefficients are significant at 1% level.
Source: Authors´ elaboration.

5. Discussion

This study applies for first time this methodology to the Spanish context, using raw data of the country to build specific Spanish CCD at the NUTS-2 level. Such approach separates this effort from others previously made, since they depart from CCD built for US. Moreover, the analysis is sharp enough to show the relevance of industries for specific regions, and reinforces previous findings about regional cluster presence in Spain made through case-studies (Elola et al., 2012; Jofre-Monseny et al., 2014; Molina-Morales et al., 2017; Ortega-Colomer et al., 2016; Vlaisavljevic et al., 2020). Additionally, this cluster mapping exercise groups industries using empirical measures rather than a conceptual aggregation of sectors without a robust theoretical justification, as the industrial district mapping has done before (Boix & Trullén, 2010; Canello & Pavone, 2016).

The study proves the feasibility of the application of an end-to-end methodology to map clusters in Europe, placing serious questions about why the current cluster mapping efforts assume that locational patterns found on US are representative for those found in the EU, and tend to homologate American CCD for Europe (Ketels & Protsiv, 2021). That *representativeness assumption* could not be reasonable for less-large, less-diversified, less-dynamic, and less-industrialized economies (Brodzicki, 2010). Furthermore, Delgado et al. (2016) states that current and past barriers to trade across Europe shaped different patterns of agglomeration when compared with US, and that American CCD aim to be a benchmark for other economies.

This research supports the idea that such representativeness assumption is questionable at least for the Spanish case, due to the next three reasons.

First, the spatial units of study for the American case are the *Economic Areas (EA)*, which represent regional relevant markets delimited for economic purposes. In contrast, in the EU the cluster mapping is made over administrative divisions (generally NUTS-2), which are defined by each member country following local criteria (in the case of Spain, historical and socio-political antecedents shaped the administrative divisions). This is relevant because the nature of the spatial units has an impact over the capacity of the similarity matrices to identify cross-industry linkages, and while for US the LC_Est/IO have the best performance as unidimensional matrices and COI/Occ have the worst ones, for Spain the LC_Est/IO have the worst performance and COI/Occ the best ones. Additionally, for the US case the similarity matrix with the best performance is a multidimensional one (LC_IO_Occ), and the authors never mix the LC and COI as they assume that such indicators capture similar linkages among industries. For the Spanish case that assumption is overlooked, and results show that the similarity matrix with the best performance is one constructed with the COI: the COI_IO_Occ.

Second, while this paper departs from traded industries as the study of Delgado et al. (2016) does, the three-criteria methodology to identify traded industries of the latter study is not capable to effectively discriminate by itself between local and traded industries for the Spanish case. Instead, this study applies a different multi-criterion methodology based on export to gross value-added ratio and the locational Gini Coefficient; for the last criterion, the cutoff is set at 0.01, as multiple cutoffs are tested in incremental ranges of 0.01 looking for the set of traded industries with the maximum overlap compared with the set defined by the three-criteria methodology of Delgado et al. (2014) (the geometric mean is used to measure the industry overlap in each direction).

Third, the North America's industrial classification is not harmonized with the EU's one. Therefore, the adaptation of the American CCD for Europe depends on the reinterpretation of the American industrial codes for the European case, which is not always a straightforward task (Brodzicki, 2010). Additionally, since the cluster algorithm relies on the data of individual industries, the differences on the interpretation of *what is* each industry will have a direct impact on the assessed cross-industry linkages and thus in the identified clusters.

The presented arguments support the idea that a robust and reliable cluster mapping effort must depart for locally-measured relatedness among industries. Otherwise, the adaptation of foreign CCD could disregard local cross-industry linkages and overestimate other less relevant ones. Moreover, this research also demonstrates that depending on the economy being analyzed, the methodology could require the

modification or complementation of procedures, criteria, and indicators with the purpose of improving the results and meeting the conceptual requirements.

This is a call for European researchers, policy makers, and economic development practitioners to take with reservation the data about local agglomeration when it is derived from the adaptation of foreign measures for cross-industry linkages. Failing to do so could lead to deficient industrial policy design, inadequate cluster performance assessment and misinterpretation of cluster's externalities. In addition, initiatives like the European Cluster Collaboration Platform and the European Clusters Excellence program present maps that show and assess presence of *cluster organizations* and not empirical evidence of the presence of industrial clusters, which could lead to the misinterpretation of the existence of industrial clusters as a real agglomeration phenomenon and not as a policy tool.

In a different train of thought, the correlation analysis between clusters' presence and different variables also shows insightful results discussed in the next paragraphs.

The correlations presented have different responses when the clusters' presence is assessed by absolute measures than when it is assessed with relative measures. In other words, the clusters' presence measured by TESh (which departs from absolute measures of employment share for each CCD) presents more statistically significant correlations with other variables than the clusters' presence measured by TESp (which departs from relative measures of employment and establishments based in LQ). Such finding suggests that, at this level of data aggregation, the absolute employment concentration on specific industries could be more useful when exploring the effects of industrial clusters over economy.

This analysis also supports previous findings related to the correlation of clusters' presence and variables like population education level, natural resource dependency, and competitiveness, showing different levels of statistical significance depending on the measure of presence but being consistent in the sign of the coefficients (Babkin et al., 2017; Delgado, Porter, et al., 2014; Slaper et al., 2018). However, at this level of data aggregation, no significant correlation is found between clusters' presence and GDP per capita, earning per worker, innovation, and unemployment variables, which are commonly linked by researchers and policy makers with the industrial agglomeration. These findings reinforce the idea that the clusters' relations with other phenomena are complex and not so evident at meso and macro levels (Grashof & Fornahl, 2021).

Mention apart deserves the correlation between the clusters' presence and the ICT/ Industry 4.0 Indexes: the sign of the correlation is positive in all the cases and statistically significant for absolute measures of presence. These results support previous findings made at micro-level that suggest that industrial clusters improve the rates of ICT and Industry 4.0 adoption.

Moreover, the research provides to researchers and policy makers with insightful data about the overall level of technological adoption in Spanish regions. This approach overcomes limitations of previous research made in Spain and Europe about Industry 4.0 and industrial clusters, since they rely on case studies, specific regions, or specific technologies (Götz & Jankowska, 2017; Grashof et al., 2021; Hervas-Oliver et al., 2019).

The correlation analysis makes it possible to assess the correlation of individual CCD with the elected variables of economic performance. In this matter, two CCD (the 04 and 05) outperform the correlations showed by the other CCD, even showing statistically significant correlations with variables like GDP per capita and earning per worker. Noteworthy, those two CCD involve engineering and manufacturing related to biochemicals, electronics, machinery, and computing, suggesting that positive externalities could find stronger linkages with those industries, as Tavares et al. (2021) suggest.

The findings provide to practitioners and researchers interested in industrial clusters with useful information to focus their efforts on identifying native competitive networks naturally present over their territory, aiming to develop their industrial clusters in a more effective way. Furthermore, for the Spanish case, policy makers could depart from this paper to assess not only their efforts into developing particular clusters over their regions, but also to put the spotlight on overlooked cross-industry linkages and to develop and improve their territorial presence, aiming to boost their returns and reach new clients and suppliers.

Although economic development and technology adoption are complex phenomena to assess, the results of this research not only provide to researchers, government, and industry leaders a solid basis for industrial policy and competitive strategy, but also a solid methodology to explore the existence of industrial clusters in different contexts. Additionally, the final insights invite researchers to explore the impact of industrial clusters using novel approaches, like the Structural Equation Modeling, capable to identify complex relations among multiple variables that could operate as mediators between the industrial cluster presence and the economic development.

6. Conclusions

This research applies, for the first time, a full quantitative methodology of cluster mapping for the Spanish context, adapted from state-of-the-art literature, based on statistical modeling and broadly applicable, with a multi-regional/multi-industry scope. The results find the presence over territory of different industrial clusters based in native cross-industry linkages naturally present over territory, departing from the industrial classification CNAE-2009 2-digits level, and the use of autonomous communities as spatial units to analyze data (NUTS-2), excluding Ceuta and Melilla. Additionally, the study explores the correlations between clusters' presence and a group of relevant variables for the economic development understanding.

The findings contribute to literature from four different perspectives.

First, from a methodological perspective the study demonstrates that even when the foundations of the methodology applied remain the same, there are procedures, criteria, and indicators that researchers must modify or complement with the purpose of improving the results of its application in particular economies.

Second, the conceptual perspective makes a call to researchers and policy makers to question the *representativeness assumption* made over the American cross-industry linkages, and to promote the local CCD creation for individual countries or even for Europe, departing from the quantitative assessment of local cross-industry linkages. The use of homologated-and-foreign CCD for the European case could underestimate relevant linkages or overestimate irrelevant ones, misleading conclusions about clusters' presence, performance, and externalities.

Third, the externalities perspective shows that the clusters' presence measured with absolute employment data correlates better with variables related to education, technology adoption and competitiveness, in contrast to the clusters' presence measured with relative employment and establishments data. Besides, the clusters' presence does not have a statistically significant correlation with expected variables like GDP per capita, earning per worker and innovation, but it maintains the expected correlation sign. These final insights invite researchers to explore the impact of industrial clusters using different approaches to find more complex relations among variables.

Fourth, from the practical perspective this paper offers, right out-of-the-box, useful information to take the regional and industrial assessment further. Researchers, policy makers, and practitioners can find the list of industries classified as traded, the groups of industries that shape each CDD, the clusters' location, and even two indexes of technological adoption for all autonomous communities (ICT and Industry 4.0 indexes). The index construction presented in this paper is the first one to group into a single indicator the technology adoption of different regions using harmonized data for all of them, being the first exercise of its kind for Spanish regions.

Nonetheless, the study is limited by the aggregation level of the data, not to mention that complete data for some industries is unavailable or hidden due to statistical confidentiality. Thus, although there are challenges related to more complete and disaggregate data availability, further analysis is recommended at NUTS-3 and CNAE-2009 3-digits to generate more detailed CCD and provide useful information at even more local level. Additionally, this could make it possible deep exploration of relations among variables, using inferential statistics as Ordinary Least Squares regression and Structural Equation Modeling. Furthermore, this research's methodology could be improved including indicators related to

technological similarity, community linkages, and natural advantages, which would be helpful to find novel cross-industry linkages departing from other approaches like the industrial district mapping.

Finally, this research shows a contemporaneous outlook to industrial structure in Spain and expects to be useful not only as a benchmark for future research, but also for policy makers and industry leaders currently working on industrial policy and competitive strategy.

References

Almeida, F., Duarte, J., & Monteiro, J. (2020). The challenges and opportunities in the digitalization of companies in a post-COVID-19 world. *IEEE Engineering Management Review*, *48*(3), 97–103. https://doi.org/10.1109/EMR.2020.3013206

Atik, H., & Ünlü, F. (2019). The measurement of Industry 4.0 performance through Industry 4.0 Index: An empirical investigation for Turkey and European countries. *Procedia Computer Science*, *158*, 852–860. https://doi.org/10.1016/j.procs.2019.09.123

Babkin, A., Plotnikov, V., & Vertakova, Y. (2018). The analysis of industrial cooperation models in the context of forming digital economy. *SHS Web of Conferences*, *44*, 12. https://doi.org/10.1051/shsconf/20184400012

Babkin, A., Vertakova, Y., & Plotnikov, V. (2017). Study and assessment of clusters activity effect on regional economy. *SHS Web of Conferences*, *35*, 63. https://doi.org/10.1051/shsconf/20173501063

Bathelt, H., & Li, P. F. (2014). Global cluster networks-foreign direct investment flows from Canada to China. *Journal of Economic Geography*, *14*(1), 45–71. https://doi.org/10.1093/jeg/lbt005

Becattini, G. (1990). The Marshallian industrial district as a socio-economic concept. In F. Pike, G. Becattini, & W. Sengenberger (Eds.), *Industrial Districts and Inter-Firm Cooperation in Italy* (pp. 37–51). ILO.

Boix, R., & Galletto, V. (2009). Innovation and industrial districts: A first approach to the measurement and determinants of the I-district effect. *Regional Studies*, *43*(9), 1117–1133. https://doi.org/10.1080/00343400801932342

Boix, R., & Trullén, J. (2010). Industrial districts, innovation and I-district effect: Territory or industrial specialization? *European Planning Studies*, *18*(10), 1707–1729. https://doi.org/10.1080/09654313.2010.504351

Brodzicki, T. (2010). Critical review of cluster mapping studies in Poland. *Analizy i Opracowania KEIE UG*, *1*(3), 3–20.

Burki, A. A., & Khan, M. A. (2011). Agglomeration economies and their effects on technical inefficiency of manufacturing firms: Evidence from Pakistan. In *International Growth Centre* (Issue March). https://www.theigc.org/project/agglomeration-economies-and-their-effects-on-productivity-and-efficiency-of-manufacturing-firms-evidence-from-pakistan/

Caloffi, A., Lazzeretti, L., & Sedita, S. R. (2018). The story of cluster as a cross-boundary concept: From local development to management studies. In F. Belussi & J. L. Hervas-Oliver (Eds.), *Agglomeration and Firm Performance. Advances in Spatial Science* (pp. 123–137). Springer. https://doi.org/10.1007/978-3-319-90575-4_8

Canello, J., & Pavone, P. (2016). Mapping the Multifaceted Patterns of Industrial Districts: A New Empirical Procedure with Application to Italian Data. *Regional Studies*, *50*(8), 1374–1387. https://doi.org/10.1080/00343404.2015.1011611

Carlino, G., & Kerr, W. R. (2015). Agglomeration and innovation. In G. Duranton, J. V Henderson, & W. Strange (Eds.), *Handbook of Regional and Urban Economics* (1st ed., Vol. 5, pp. 349–404). Elsevier B.V. https://doi.org/10.1016/B978-0-444-59517-1.00006-4

Delgado, M., Bryden, R., & Zyontz, S. (2014). *Categorization of traded and local industries in the US economy*. 1–9. https://clustermapping.us/sites/default/files/files/page/Categorization of Traded and Local Industries in the US Economy.pdf

Delgado, M., Porter, M. E., & Stern, S. (2014). Clusters, convergence, and economic performance. *Research Policy*, *43*(10), 1785–1799. https://doi.org/10.1016/j.respol.2014.05.007

Delgado, M., Porter, M. E., & Stern, S. (2016). Defining clusters of related industries. *Journal of Economic Geography*, *16*(1), 1–38. https://doi.org/10.1093/jeg/lbv017

Diodato, D., Neffke, F., & O'Clery, N. (2018). Why do industries coagglomerate? How Marshallian externalities differ by industry and have evolved over time. *Journal of Urban Economics*, *106*, 1–26. https://doi.org/10.1016/j.jue.2018.05.002

Ellison, G., Glaeser, E. L., & Kerr, W. R. (2010). What causes industry agglomeration? Evidence from coagglomeration patterns. *American Economic Review*, *100*(3), 1195–1213. https://doi.org/10.1257/aer.100.3.1195

Elola, A., Valdaliso, J., Lopez, S. M., & Aranguren, M. J. (2012). Cluster life cycles, path dependency and regional economic development: Insights from a meta-study on Basque clusters. *European Planning Studies*, *20*(2), 257–279. https://doi.org/10.1080/09654313.2012.650902

Elola, A., Valdaliso, J. M., Franco, S., & López, S. M. (2017). Public policies and cluster life cycles: insights from the Basque Country experience. *European Planning Studies*, *25*(3), 539–556. https://doi.org/10.1080/09654313.2016.1248375

Everitt, B. S., Landau, S., Leese, M., & Stahl, D. (2011). *Cluster analysis*. Wiley.

Fernandez-Escobedo, R., Eguía-Peña, B., & Aldaz-Odriozola, L. (2023). Economic agglomeration in the age of Industry 4.0: developing a Digital Industrial Cluster as a new policy tool for the digital world. *Competitiveness Review*. https://doi.org/10.1108/CR-07-2022-0095

Götz, M., & Jankowska, B. (2017). Clusters and Industry 4.0 - Do they fit together? *European Planning Studies*, *25*(9), 1633–1653. https://doi.org/10.1080/09654313.2017.1327037

Grashof, N., & Fornahl, D. (2021). "To be or not to be" located in a cluster?—A descriptive meta-analysis of the firm-specific cluster effect. In *Annals of Regional Science* (Vol. 67, Issue 3). Springer Berlin Heidelberg. https://doi.org/10.1007/s00168-021-01057-y

Grashof, N., Kopka, A., Wessendorf, C., & Fornahl, D. (2021). Industry 4.0 and clusters: Complementaries or substitutes in firm's knowledge creation? *Competitiveness Review*, *31*(1), 83–105. https://doi.org/10.1108/CR-12-2019-0162

Grimmer, J., & King, G. (2011). General purpose computer-assisted clustering and conceptualization. *Proceedings of the National Academy of Sciences*, *108*(7), 2643–2650. https://doi.org/10.1073/pnas.1018067108

Hermans, F. (2021). The contribution of statistical network models to the study of clusters and their evolution. *Papers in Regional Science*, *100*(2), 379–403. https://doi.org/10.1111/pirs.12579

Hervás-Oliver, J. L. (2021). Industry 4.0 in industrial districts: regional innovation policy for the Toy Valley district in Spain. *Regional Studies*, *55*(10–11), 1775–1786. https://doi.org/10.1080/00343404.2021.1939861

Hervas-Oliver, J. L., Estelles-Miguel, S., Mallol-Gasch, G., & Boix-Palomero, J. (2019). A place-based policy for promoting Industry 4.0: the case of the Castellon ceramic tile district. *European Planning Studies*, *27*(9), 1838–1856. https://doi.org/10.1080/09654313.2019.1642855

ICEX España Exportación e Inversiones. (n.d.). *Why invest in Spain? - Economy*. Ministerio de Industria, Comercio y Turismo. Retrieved January 9, 2020, from https://www.investinspain.org/en/why-spain/economy

Jasinska, K., & Jasinski, B. (2019). Clusters under industry 4.0 conditions - Case study: the concept of Industry 4.0 cluster in Poland. *Transformations in Business & Economics*, *18*(2B), 802–823. http://www.transformations.knf.vu.lt/47b/article/clus

Jofre-Monseny, J., Marin-Lopez, R., & Viladecans-Marsal, E. (2014). The determinants of localization and urbanization economies: evidence from the location of new firms in Spain. *Journal of Regional Science*, *54*(2), 313–337. https://doi.org/10.1111/jors.12076

Ketels, C. (2017). Cluster mapping as a tool for development. In *Harvard Business School* (Issue June). https://www.hbs.edu/faculty/Pages/item.aspx?num=53385

Ketels, C., & Protsiv, S. (2021). Cluster presence and economic performance: a new look based on European data. *Regional Studies*, *55*(2), 208–220. https://doi.org/10.1080/00343404.2020.1792435

Krugman, P. (1991). Increasing returns and economic geography. *Journal of Political Economy*, *99*(3), 483–499. https://doi.org/10.1086/261763

Lorenzini, F., & Lombardi, S. (2018). L'identificazione dei distretti industriali: una rassegna metodologica [Mapping industrial districts: A methodological review]. *Scienze Regionali, Italian Journal of Regional Science*, *17*(2), 225–260. https://doi.org/10.14650/90223

Mano, R., & Castillo, M. (2015). The level of productivity in traded and non-traded sectors for a large panel of countries. *IMF Working Papers*, *15*(48), 1. https://doi.org/10.5089/9781484392140.001

Maresova, P., Soukal, I., Svobodova, L., Hedvicakova, M., Javanmardi, E., Selamat, A., & Krejcar, O. (2018). Consequences of Industry 4.0 in business and economics. *Economies*, *6*(3), 14. https://doi.org/10.3390/economies6030046

Marshall, A. (1920). *Principles of economics* (8th ed., p. 69). Macmillan (original work published 1890).

Molina-Morales, F. X., Martinez-Chafer, L., & Valiente-Bordanova, D. (2017). Disruptive technological innovations as new opportunities for mature industrial clusters. The case of digital printing innovation in the Spanish ceramic tile cluster. *Investigaciones Regionales-Journal of Regional Research*, *39*, 39–57. https://investigacionesregionales.org/es/article/disruptive-technological-innovations-as-new-opportunities-for-mature-industrial-clusters-the-case-of-digital-printing-innovation-in-the-spanish-ceramic-tile-cluster/

Oosterhaven, J., Eding, G. J., & Stelder, D. (2001). Clusters, linkages and interregional spillovers: Methodology and policy implications for the two Dutch mainports and the rural North. *Regional Studies*, *35*(9), 809–822. https://doi.org/10.1080/00343400120090239

Ortega-Colomer, F. J., Molina-Morales, F. X., & Fernández-de-Lucio, I. (2016). Discussing the concepts of cluster and industrial district. *Journal of Technology Management & Innovation*, *11*(2), 139–147. https://doi.org/10.4067/S0718-27242016000200014

Porter, M. E. (1990). The competitive advantage of nations. *Harvard Business Review*, *62*(2), 73–93. https://www.hbs.edu/faculty/Pages/item.aspx?num=6105

Porter, M. E. (2003). The economic performance of regions. *Regional Studies*, *37*(6–7), 549–578. https://doi.org/10.1080/0034340032000108688

Scherer, F. M. (1984). Using linked patent and R&D data to measure interindustry technology flows. In Z. Griliches (Ed.), *R&D, Patents, and Productivity* (pp. 417–461). University of Chicago Press.

Sforzi, F. (2015). Rethinking the industrial district: 35 years later. *Investigaciones Regionales*, *32*(2015), 11–29. http://www.aecr.org/images/ImatgesArticles/2015/11/2_Sforzi.pdf?_ga=2.131452296.680042952.1678926565-1405984931.1678409478

Skokan, K., & Zotyková, L. (2014). Evaluation of business cluster performance during its lifecycle. *Acta Universitatis Agriculturae et Silviculturae Mendelianae Brunensis*, *62*(6), 1395–1405. https://doi.org/10.11118/actaun201462061395

Slaper, T., Harmon, K., & Rubin, B. (2018). Industry clusters and regional economic performance: A study across US Metropolitan Statistical Areas. *Economic Development Quarterly*, *32*(1), 44–59. https://doi.org/10.1177/0891242417752248

Slaper, T., & Ortuzar, G. (2015). Industry clusters and economic development. *Indiana Business Review*, *90*(1), 7–9. https://www.ibrc.indiana.edu/ibr/2015/spring/pdfs/article2.pdf

Szanyi, M., Iwasaki, I., Csizmadia, P., Illésy, M., & Makó, C. (2010). Cluster development in Hungary: searching for a 'critical mass' of business via cluster mapping. In B. Dallago & C. Guglielmetti (Eds.), *Local economies and global competitiveness* (pp. 113–133).

Tavares, M. S. D., Gohr, C. F., Morioka, S., & da Cunha, T. R. (2021). Systematic literature review on innovation capabilities in clusters. *Innovation and Management Review*, *18*(2), 192–220. https://doi.org/10.1108/INMR-12-2019-0153

Titze, M., Brachert, M., & Kubis, A. (2011). The identification of regional industrial clusters using qualitative input-output analysis (QIOA). *Regional Studies*, *45*(1), 89–102. https://doi.org/10.1080/00343400903234688

Vlaisavljevic, V., Cabello-Medina, C., & Van Looy, B. (2020). The role of policies and the contribution of cluster agency in the development of biotech open innovation ecosystem. *Technological Forecasting and Social Change*, *155*, 119987. https://doi.org/10.1016/j.techfore.2020.119987

Ybarra, J. A., & Domenech-Sanchez, R. (2012). Innovative business groups: territory-based industrial policy in Spain. *European Urban and Regional Studies*, *19*(2), 212–218. https://doi.org/10.1177/0969776411428558 WE - Social Science Citation Index (SSCI)

Yelkikalan, N., Soylemezoglu, E., Kiray, A., Sonmez, R., Ezilmez, B., & Altun, M. (2012). Clustering approach as a regional development tool. *8th International Strategic Management Conference*, *58*, 503–513. https://doi.org/10.1016/j.sbspro.2012.09.1027

Zhu, S., & Pickles, J. (2016). Institutional embeddedness and regional adaptability and rigidity in a Chinese apparel cluster. *Geografiska Annaler, Series B: Human Geography*, *98*(2), 127–143. https://doi.org/10.1111/geob.12095

ORCID

Rudy Fernández-Escobedo https://www.orcid.org/0000-0001-5121-1877

Begoña Eguía-Peña https://www.orcid.org/0000-0003-1852-1624

Leire Aldaz-Odriozola https://www.orcid.org/0000-0002-7960-5410