The emergence of a “twin” transition scientific knowledge base in European regions

Giacomo Damioli[1] Stefano Bianchini[2], and Claudia Ghisetti[3]

[1] Faculty of Business Studies and Economics, University of Bremen, Bremen, Germany; CAPP – Centre for the Analysis of Public Policies, University of Modena and Reggio Emilia, Modena, Italy giac.damioli@gmail.com

[2] BETA, Université de Strasbourg, Strasbourg, France; s.bianchini@unistra.fr

[3] Department of Sociology and Social Research, University of Milan-Bicocca, Milan, Italy; claudia.ghisetti@unimib.it

Abstract

This study aims to uncover spatial patterns and cognitive combinations of digital and environmental scientific knowledge bases, for the generation of new knowledge in the “twin” transition domain. Its recent and rapid diffusion in European regions has not followed clearly defined spatial patterns and has been a dynamic process of actors reconfiguration. Yet, regions with a strong green and digital science base have a higher propensity to produce more, better quality and more visible twin knowledge. Among the prevalent “twin” knowledge fields emerge AI- and IoT-based applications for energy storage, distribution and consumption, environmental monitoring and modeling, and urban planning.

JEL codes: O30, Q53, Q55, R11

Keywords: twin transition, scientific knowledge, digital eco-system, environmental innovation

Acknowledgment

Claudia Ghisetti acknowledges financial support under Research Projects of National Interest (PRIN) by the Italian Minister of University and Research (MUR) under the project titled “Digital and green transitions in Italian businesses: An assessment of the main antecedents” n. 2022CS5X8Y

1. Introduction

The ongoing digital transformation – and recent advances in artificial intelligence (AI) particularly – has brought some optimism about how to meet the grand challenges of the new millennium. While the notion of "AI for social or public good" is not free from controversies on who will benefit from future digital developments (Taylor 2016, Moore 2019), hopes are diffusing that advanced digital technologies may provide solutions to the tremendous challenges humanity is facing (Cowls 2020). Among these is the concern of how to best mitigate human-induced climate change and limit environmental degradation. Two centuries of sustained economic development has caused severe damages to the environment, making mankind one of the new geological forces (Rockstrom et al. 2009, Lewis and Maslin 2015). Thus, economic growth should be increasingly accompanied by the incessant introduction of new environmental technologies to improve (material and energy) efficiency (Ayres and van den Bergh 2005).

Digital technologies could be a force for improvements in this direction. Cockburn et al. (2018) have suggested that AI – and "deep learning" in particular – acts as a "general method of invention", combining properties of general-purpose technologies (Bresnahan and Trajtenberg 1995) and new methods of invention (Griliches 1957). Recent studies support this intuition by showing that deep learning makes it possible to tackle new problems in highly unconventional ways and speed up the process of scientific discovery, affecting a multiplicity of scientific fields (Besiroglu et al. 2022; Bianchini et al. 2022, Thu et al. 2022).

Digital technologies are indeed increasingly being applied to tackle environmental issues via prominent applications (Goralski and Tan, 2020). Some studies elicited expert opinions to discuss the potential beneficial and adverse environmental impacts of certain digital technologies (Vinuesa et al. 2020, Guenat et al. 2022). Others showed that firms' and regions' strength in digital technologies contribute to environmental innovation (Cicerone et al. 2022, Montresor and Vezzani 2022), though the overall net environment impact is ambiguous (Bianchini et al. 2023).

While such literature provides interesting insights on the environmental impacts of digital technologies, comprehensive evidence on the combinatory potential of green and digital technologies is lacking. This paper fills this gap by focusing on the combination of digital and green knowledge in science. Its main objective is to uncover the spatial patterns and knowledge base used to successfully generate new knowledge in an emerging field, which we label – following the recent European Union (EU) policy discourse – the "*twin transition*". In general terms, the twin transition involves the development of new digital solutions to help society, as a whole, becoming more sustainable.

To that purpose, we leveraged scientific publications with authors affiliated in Europe jointly dealing with digital and green technologies. The focuses on science and Europe are both relevant in our context. On the one side, scientific knowledge has been shown to be one of the most important drivers of technological change (Fleming and Sorensen 2004, Murray 2002, Rosenberg and Nelson 1994). On the other side, the EU devoted increasing resources to policies aimed to promote the twin transition, through a wide range of coordinated flagship initiatives, such as the European Green Deal, the European Industrial Strategy, the Digital Strategy and the NextGenerationEU recovery package.

Drawing on data from Web of Science (WoS) and Altmetric.com, we created three measures of regional scientific strength in the digital, green, and twin domains: i) the *quantity* in research outputs; ii) its scientific *quality* and iii) its *online visibility*. Each measure allowed us to capture a different facet of impact: scientific quantity, proxied through the number of publications, is a raw indicator of scientific capacity; scientific quality, proxied through the number of citations received, mirrors the importance of a contribution within the scientific community; online visibility, approximates the dissemination of scientific knowledge outside the "ivory tower", such as in the policy community and, more generally, in the general public.

We retrieved digital and green publications over the period 2000-2020 through a keyword approach, which we framed to disentangle several digital and green subdomains, and then assigned them on the basis of authors' affiliations to 287 NUTS 2 regions in 30 European countries. Twin publications were flagged as those contributions that lie at the intersection of the digital and green ones.

The results of the analysis, which combines various statistical methods including econometric and natural language processing (NLP) techniques, show that twin scientific knowledge creation diffused in the last twenty years to an increasing number of European regions. Its diffusion did not follow clearly defined spatial patterns and was accompanied by a reshuffling of regions' rankings across decades. Regions with a strong green and digital science base have a higher propensity to produce more, better quality and more visible twin knowledge. Twin domains are emerging on the study of AI- and IoT-powered applications for energy storage, distribution and consumption, environmental monitoring and modeling, and urban planning.

The paper is organized as follows. Section 2 outlines the conceptual background. Section 3 describes data, variables and methods used. Section 4 presents the results. Section 5 concludes with some implications for policy.

2. Conceptual framework

Scientific knowledge is an important driver of technological innovation (Rosenberg and Nelson 1994, Murray 2002, Fleming and Sorensen 2004,). Scientific discoveries guide research and development (R&D) efforts, for instance reducing trial-and-errors and, therefore, the time needed for the development and deployment of new technologies (Dasgupta and David 1994) and indicating opportunities for the re-combination of existing knowledge (Cassiman et al. 2004). A large body of empirical research confirms a strong nexus between major scientific advances and new technologies in various domains, such as ICT (Mazzucato 2014), semiconductors (Dibiaggio et al. 2014), biotechnology (Magerman et al. 2015), and wind turbines (Lacerda 2019).

In what follows, we outline the role that scientific knowledge plays in driving the twin transition and its underlining digital and green domains (Subsection 2.1); the evidence available to date on the twin transition (Subsection 2.2); and iii) the role that economic geography may have for digital and green knowledge creation and recombination (2.3).

2.1 Scientific knowledge for the twin transition

Scientific knowledge is central for the development of complex technologies (Sorenson et al. 2006, Arthur 2009, Nightingale 2009), including environmental and digital technologies, whose recombination is at the heart of the twin transition.

Environmental technologies, on the one hand, are complex technologies that tend to rely on multifaceted and interdisciplinary knowledge domains, often close to the scientific frontier (Barbieri et al. 2020a). Certain environmental technologies, such as those concerning renewable energy sources, have an analytical knowledge content that makes them rely more closely on the local scientific knowledge from universities and research labs, especially for exploratory search into other technological domains to benefit from cross fertilization of ideas (Ocampo et al. 2021, Moreno and Ocampo 2022).

Digital technologies, on the other hand, include a range of heterogeneous technologies having increasingly complex interactions (Schwab 2017, De Propris and Storai 2019, OECD 2019). Within digital technologies, AI (and machine learning in particular) is expected to play a central role for the recombination of existing technologies and the generation of new inventions. According to many scholars, AI combines the properties of "general-purpose technologies" (Goldfarb et al. 2023) – wide scope for continuous improvement and wide complementarity with other (existing and new) technologies (Bresnahan and Trajtenberg 1995) – with the ability to generate new innovations as a "method of invention" (Cockburn et al. 2018). Consistently with this idea, recent studies show that AI intervenes at different stages of the scientific pipeline, helping scientists formulate the research question, collect and manage data, and analyze and interpret the results (Raghu and Schmidt 2020). The impact of AI on scientific discovery occurs in virtually all scientific fields, and can be significant, though often uncertain (Bianchini et al. 2022, Thu et al. 2022).

Digital technologies other than AI can also be extremely valuable for the twin transition. Some IoT devices and applications – e.g., wearable, sensors, connected cars, smart cities – permit the generation and collection of increasing volumes of (heterogenous) data through complex interactions of a range of different *milieus* (private and domestic, industrial, service-related). Thanks to computational infrastructures and advanced capabilities to leverage big data analytics, these data are often applied in scientific activity (Chen et al. 2015, Mourtzis et al. 2016).

In summary, digital technologies – especially AI, big data, and IoT – seem particularly suited to generate new discoveries interacting with other scientific domains. We therefore expect them to exert an important role in driving the emergence of the twin scientific domain that is combinatorial in nature.

2.2 Existing evidence on the twin transition

Empirical research on the twin transition is still in its infancy, though a growing number of recent studies contributed to frame this work. Such an emerging literature broadly confirms the idea that new digital technologies can facilitate the green transformation, thus supporting the twin transition rationale. Yet, a number of environmental challenges deriving from the use of digital technologies are also stressed (Creutzig et al. 2022).

A first literature strand encompasses a series of studies focusing on how digital technologies can contribute to meet sustainable development goals (SDGs). For instance, Goralski and Tan (2020) discuss a few case studies of AI-based applications, including AI-powered smart water management systems enabling to reduce water overuse and degradation. Based on experts' opinions, Vinuesa et al. (2020) and Guenat et al. (2022) point to both positive – e.g. more efficient energy distribution through smart grids or better monitoring of remote ecosystems – and negative – e.g. increasing pollution due to high energy demand and disposal of devices at the end of their life-cycle – contributions to SDGs achievement, with a balance moderately favoring the positive ones.

Another related strand of literature has studied how digital technologies may have contributed to environmental innovations (De Marchi and Di Maria 2020, Ardito et al. 2021). Firms' investments in AI have been found to support Italian firms' ability to adopt environmental innovations (Montresor and Vezzani 2022), though with substantial differences across firms' location in urban vs rural areas (Cattani et al. 2023). Based on patent data, it has been observed that regional strength in AI technologies promotes further regional specialization in green ones (Cicerone et al. 2022), though at the expenses of an adverse environmental impact in terms of growing greenhouse gas emissions (Bianchini et al. 2023).

While exploiting expert opinions, case studies, companies' surveys, and patents provide insights on the impacts of digital technologies on various environmental outcomes, no study to date has dealt with the knowledge recombination process leading to the creations of twin technologies. This study aims to fill this gap.

2.3 Spatial and cognitive patterns of the twin transition

Regional and evolutionary economists have long studied the emergence of new technologies. Half a century of scholarly research firmly supports the idea that new technology results from the combination of existing components (Usher 1954, Arthur 2009).

Cognitive and spatial embeddedness facilitate knowledge flows among the various actors involved in the process of innovation and spur technological innovation (Freeman 1987, Jaffe et al. 1993, Capello 2002, Storper 2018). Because the learning process is cumulative and knowledge spillovers often bounded in space, new scientific domains and technologies are more likely to emerge in places with a greater endowment of pre-existing knowledge bases (Boschma 2017). Such spatial dependence can have long-lasting consequences on the concentration and polarization of knowledge and capabilities (Dosi 1982, Bellandi et al. 2018). In this respect, many studies have shown – typically through the analysis of patent data – that both digital (World Bank 2017, Ciffolilli and Muscio 2018, Martinelli et al. 2021) and green (Corradini 2019, Truffer and Coenen 2012, Montresor and Quatraro 2020) technologies are highly spatially embedded, their development being highly dependent on the local presence of related and enabling technologies.

At the same time, new technologies and innovation can bring profound structural changes, reshuffling the configuration of actors, industry and market structure (Malerba and Orsenigo 1996). Thus, the turbulence created by technological change may reshape the spatial distribution of knowledge development. The advent of (advanced) digital technologies, often referred to as the “fourth industrial revolution”, is causing deep transformations of economies and industries as well as society at an unprecedented scale (Allen 2017, Mitchell and Brynjolfsson 2017, Schwab 2017). It has partly reshuffled the spatial landscape of European regions with the emergence of new technological “leaders” who were/are not as advanced in ICT (Balland and Boshma 2021, Capello and Lenzi 2021). As for green knowledge, existing studies suggest that green technologies are more complex and novel than non-green ones as they tend to recombine with a wide range of knowledge sources, technological components and knowledge inputs that may come from cognitively distant domains (Barbieri et al. 2020a). Accordingly, their uptake is positively correlated with a diversified knowledge base across unrelated technological fields (Barbieri et al. 2020b). This makes the advent of advanced digital technologies a potential opportunity for further evolution and diffusion of green technologies.

The lack of empirical evidence, coupled with the high emphasis on the twin transition in the ongoing policy discourse, makes particularly interesting to study how scientific knowledge on the twin transition emerged across European regions and how much it builds on prior art, disentangling the mechanism of “combinatorial evolution” shaping its creation, which constitute the research questions to which this study aims to answer.

3. Methodological approach

3.1 Data sources

We construct an original dataset of ‘green’ and ‘digital’ scientific publications in EU27, UK, Norway and Switzerland. In particular, we retrieved publication data from the Web of Science (Wos) Core

Collection by identifying a list of selected keywords in the title, abstract, or keywords.[1] Our query filters all documents that contain at least one author affiliation in the selected countries for the period 2000-2020.

For digital publications, we used an enriched version of the keyword list recently proposed in Bianchini et al. (2023), shown in Appendix (Figure A1), which is largely inspired by the OECD digital taxonomy (2019, p.18), recent contributions on AI mapping (Van Roy et al. 2020), and a set of components of the Industry 4.0 paradigm (Martinelli et al. 2021). The digital ecosystem in our study includes the following non-exclusive (i.e., a publication can fall in more than one) dimensions: *additive manufacturing*, *AI*, *blockchain*, *big data*, *computing infrastructures*, *IoT* and *robotics*.

As for green publications, we developed an original comprehensive list of keywords[2], shown in Appendix (Figure A2), based on words' occurrence in the description of the Cooperation Patent Classification (CPC) and International Patent Classification (IPC) patent classes in the OECD Env-Tech classification of environmental-related technologies (Haščič and Migotto, 2015*)*. Accordingly, green knowledge space comprises the following non-exclusive dimensions: *biodiversity*; *GHG capture*; *mitigation* in *buildings*, *energy*, *production*, *transport* and *wastewater*; *environmental management*; and *adaptation* in *water*.

3.2 Measures

The set of documents that include at least one digital *and* at least one green keywords in their title, abstract or keywords lists constitutes what we label "twin" publications. We retrieved a total of 377,665 digital, 218,981 green and 7,886 twin publications, which we assigned to 287 NUTS2 regions by geo-locating authors' affiliations.[3] We endorsed a "full counting" approach assigning a publication to each reach region with at least one author affiliation to reflect the idea that knowledge is not fractionable.[4]

We use the number of publications, which enters the empirical analysis under the label "*quantity*", to measure the volume of scientific output of domains and regions. Panel a) of Figure 1 shows the evolution across time of scientific quantity in digital, green and twin domains in Europe. Twin publications were virtually non-existent in Europe in the early 2000s, but their number gradually rose across the past two decades, with an accelerated growth in recent years, driven by the boom of digital publications. Beside increasing in number, twin publications diffused to a rising number of regions, from about 10% (of the 287 NUTS 2 regions) in the early 2000s up to 85% in 2020 (panel b) in Figure 1). Also, the fractions of regions with at least one digital or green publication rose in the period, but more mildly, increasing from about 80% in the early 2000s to more than 95% in recent years.

---

[1] Some readers could be concerned by the reliance of our empirical analysis on data from the Web of Science, which could underestimate the extent of publishing activities in some increasingly popular venues, particularly in computer science disciplines and AI/ML, such as open access journals (often not indexed by WoS) and open repositories like arXiv. To assess the magnitude of this potential discrepancy, we compared the number of records we obtain by applying our AI-related keywords in the open catalog OpenAlex and found that, one once purged from pre-prints not published in peer-reviewed journals, our WoS sample covers about 95.5% of the records found in OpenAlex.

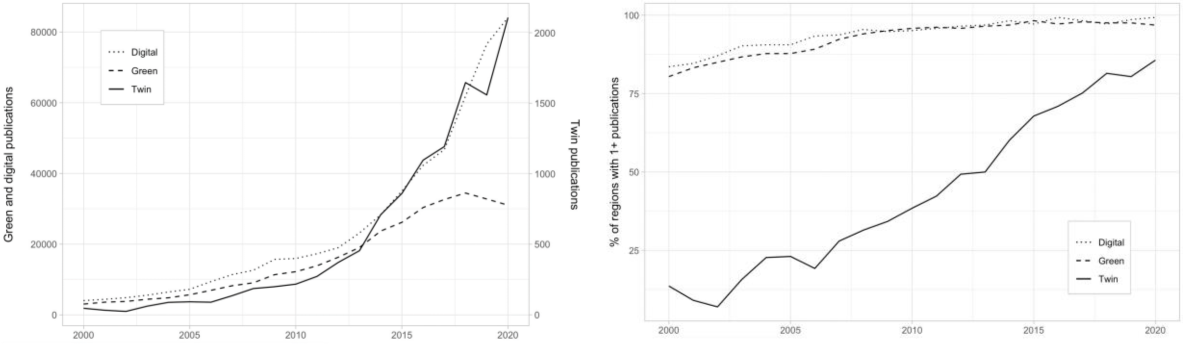
[2] David Popp (2016) was the first providing a rich list of keywords to flag publications in environmental domains, but his study was restricted to the specific domain of renewable energy, while we aimed to capture more ample environmental dimensions.

[3] A low number of author affiliations could not be geo-located due to lack of information and/or misspellings of addresses. We lost 2.1% of green publications and 1.5% of digital publications.

[4] To test the robustness of the findings to this assumption, we performed the full analysis also using a "fractional counting" scheme assigning to each region its share of authors affiliations. Our findings are confirmed using this alternative assignation scheme – see Table A4 in the Appendix.

Table 1 shows the composition of publications across the digital and green subdomains. AI and IoT are the digital subdomains that account for the largest shares of both twin and digital publications. Yet, their importance changes considerably in twin and digital publications. IoT is considerably more prevalent in twin (31%) than in digital (13%) publications. By contrast, AI is more diffused in digital (49%) than in twin (34%) publications. Alike for AI, the share of publications on robotics is larger in digital (17%) than twin (11%) publications. Differences in the shares of green subdomains computed over twin and green publications are more nuanced. A slightly larger prevalence in twin than green publications is observed in the buildings (18% of twin vs 10% of green publications) and energy (38% vs 33%) subdomains. By contrast, the GHG capture (2% vs 9%) and water (6% and 11%) subdomains are more diffused in green than twin publications. The remaining digital and green subdomains account for similar shares of publications in the twin, digital and green domains: environmental management accounts for 16% of twin and green publications, transport for 11% of twin and green publications, computing infrastructures for 10% of twin and digital publications, and the other subdomains for smaller shares.

Figure 1. Development and diffusion of digital, green and twin publications in European regions



a) Number of publications

b) Percentage of regions with 1+ author affiliation

Notes: percentages are computed over a total of 287 NUTS 2 regions.

Table 1. Digital and green subdomains in digital, green and twin publications

Digital subdomain	Digital publications		Twin publications		Difference in normalised shares
	Share of publications	Normalised share	Share of publications	Normalised share	
Additive manufacturing	0.06	0.05	0.05	0.05	-0.01
AI	0.55	0.49	0.39	0.34	-0.15
Big data	0.06	0.06	0.09	0.08	0.02
Blockchain	0.01	0.01	0.02	0.01	0.00
Computing infrastructures	0.11	0.10	0.11	0.10	0.00
IoT	0.14	0.13	0.36	0.31	0.19
Robotics	0.19	0.17	0.13	0.11	-0.06
Total	1.12	1.00	1.14	1.00	0.00

Green subdomain	Green publications		Twin publications		Difference in normalised shares
	Share of publications	Normalised share	Share of publications	Normalised share	
Biodiversity	0.02	0.01	0.01	0.01	-0.01
GHG capture	0.14	0.09	0.04	0.02	-0.06
Buildings	0.16	0.10	0.28	0.18	0.09
Energy	0.53	0.33	0.59	0.38	0.06
Production	0.04	0.03	0.05	0.03	0.00
Transport	0.17	0.11	0.17	0.11	0.00

Waste-water	0.12	0.08	0.07	0.04	-0.03
Environmental management	0.25	0.16	0.24	0.16	0.00
Water	0.18	0.11	0.09	0.06	-0.05
Total	1.63	1.00	1.53	1.00	0.00

Notes: the total share of publications is larger than one as publications may belong to multiple subdomains.

The second measure we constructed is "*quality*", measured as the average yearly number of citations received by publications. The third measure is "*online visibility*". Using the digital object identifier (DOI) of each document, we retrieved from Altmetrics.com various indicators of publication's online activity, which we use to measure their visibility beyond the scientific community. In particular, we used the aggregate "attention score" that combines mentions of publications in mainstream media (such as online newspapers), online blogs and forums, social networks (such as Twitter and Facebook), Wikipedia, policy documents, and discussions in research blogs.

Table 2 provides some descriptive statistics. Although twin publications are much lower in number – there was 1 twin publication every 46 digital publications and every 29 green ones – they received more attention in the scientific community, with an average of 7.5 citations per publication as compared to 4.9 and 4.2 citations received by digital and green publications. By contrast, green publications received the largest degree of online visibility. The growing importance of the twin knowledge domain is reflected in its higher growth rate between the decades 2001-2010 and 2011-2020. Hence, twin publications account for an increasing share of digital and green publications, rising from about 1% in the early 2000s to about 3% of digital and 6% of green publications at the end of the 2010s.[5]

Table 2. Quantity, scientific impact and online visibility of twin, digital and green publications

	Digital	**Green**	**Twin**
Average yearly number of publications, per region (quantity)	88.28	55.37	1.90
Average number of scientific citations, per publication (quality)	4.90	4.16	7.49
Average score of online mentions, per publication (online visibility)	1.33	2.61	1.64
citations in blogs	1.55	2.41	1.50
citations in Twitter	1.11	1.63	1.00
citations in mainstream media (newspapers, magazines …)	0.06	0.15	0.09
citations in Facebook	0.02	0.03	0.02
citations in Wikipedia	0.01	0.02	0.03
Average number of authors, per publication	4.47	4.77	4.49
Average number of European NUTS 2 regions with 1+ affiliations, per publication	1.41	1.52	1.45

3.3 Methods

We analyzed the three measures of scientific knowledge combining various statistical methods including econometric (described in Subsection 3.3.1) and NLP (Subsection 3.3.2) techniques.

3.3.1 Econometric analysis

The econometric analysis aims to assess the role that the green and digital knowledge bases play for the local emergence of the twin knowledge. In particular, it allows us to correlate the quantity, quality and

---

[5] Figure A3 in the Appendix shows the yearly shares of twin versus digital and green publications.

online visibility of the twin domain with the corresponding measures of green and digital ones.

We estimated the following model:

$$TWIN_{i;t,t+2} = \alpha + \beta_1 GREEN_{i;t,t-2} + \beta_2 DIGIT_{i;t,t-2} + \gamma Controls_{i;t-2} + \tau Time_t + \sigma_i + \varepsilon_{i;t} \qquad (1)$$

where *i* indexes 270 NUTS 2 regions in the EU27, Norway and the UK, and *t* indexes years from 2000 to 2020.

The dependent variable (*TWIN*) is a measure of twin scientific knowledge across the three different dimensions of quantity, quality, and online visibility, on which we estimate separate models.

The explanatory variables of main interest – *GREEN* and *DIGIT* – capture publications in the green and digital scientific domains along the three dimensions of levels, quality, and visibility (after having subtracted the twin ones to avoid double counting).[6]

Controls include regional economic performance (gross domestic product, *GDP*), population density (population per squared kilometre, *DENSITY*), technological development (patent applications stock constructed applying a yearly depreciation rate of 15%, *PAT*), economic structure (share of production units operating in energy intensive sectors, *DIRTY*), and R&D intensity (share of R&D expenditures over GDP, *RD_GDP*). To better grasp the role of green and digital regional technological development, we replaced an alternative specification the overall regional patent stock (PAT) with green (based on environmental patent applications flagged using the OECD Env-Tech classification, *GREENTECH)* and digital (based on digital patent applications flagged using the identification strategy adopted by Bianchini et al. 2023, *DIGITECH)* ones. All models include individual regional fixed effects (σ), as suggested by the Hausman test, and year dummies (τ). The definition and descriptive statistics of all variables are reported in the Appendix (Table A1).

We smoothened the volatile distribution of publications and patent applications by applying a 3-year moving average. To allow a sufficient time lag for past scientific and technological knowledge bases to influence twin scientific knowledge creation, we constructed the moving averages of the dependent variables using forward years (*t*, *t+1* and *t+2*), and the moving averages of the explanatory variables using lagged ones (*t*, *t-1* and *t-2*). We transformed all variables (excluding shares) using an inverse hyperbolic sine transformation, which is well fitted to handle distributions with extreme values (including zeros) and right-skewed (Burbidge et al., 1988), and allows interpreting parameters estimates as elasticities (Aihounton and Henningsen, 2021). We did not use models accounting for spatial autocorrelation because standard Lagrange multiplier tests of the form proposed by Anselin et al. (1996) indicate the absence of spatial autocorrelation issues in the estimated models.[7]

3.3.2 NLP techniques

In a second step, we investigate which green and digital subdomains most likely generate twin scientific

---

[6] When the dependent variable is *TWIN* quantity (or quality or online visibility), the focal covariate enters also are *GREEN* and *DIGIT* quantity (or quality or online visibility).

[7] We report LM test for spatial errors and spatial lags in Table A4 in the Appendix. the spatial lag models with a queen weighting matrix W built using the 5th-nearest neighbouring scheme. As robust LM tests indicate that spatial lag models should be preferred to spatial error ones in the models on online visibility, we report results of spatial autoregressive models in Table A5 of the Appendix. In line with the very low spatial autoregressive terms, results are very close to those of models that do not control for spatial autocorrelation. We are very grateful to two referees for their guidance in the implementation and interpretation of these tests.

knowledge. To this end, we applied topic modelling – specifically Latent Dirichlet Allocation (LDA, Bei et al. 2003)[8] – to the abstracts of twin publications. Our approach adheres to established practices in the literature, including the removal of stop-words and the lemmatization and stemming of the corpus (Jelodar et al. 2019). LDA is one of the most popular NLP approaches for topic discovery and semantic mining from unordered documents. It is an unsupervised generative probabilistic model of a corpus, which treats any document as a random mix over latent topics, where a topic is characterized by a distribution over words. Using LDA, the terms in the cumulated corpora (twin abstracts in our case) generate a vocabulary that is then applied to discover hidden topics. Once topics are learned from the data, the probability of a document belonging to a specific topic is determined by the prevalence of terms from the vocabulary associated to that topic. In our context, we employ common metrics leading to the selection of the optimal number of six topics: coherence (having a score equal to 0.39), assessing the interpretability of topics, and perplexity (score equal to -7.80), measuring the model accuracy on a held-out set of documents.

4. Results

This section reports and discusses the results of the analysis. We first focus on the spatial patterns of twin scientific knowledge (Subsection 4.1), and then how green and digital scientific knowledge combine in its creation (Subsection 4.1).

4.1 Spatial patterns of twin scientific knowledge

Figure 2 maps the spatial distribution of twin scientific knowledge across NUTS 2 regions of our sample, in absolute counts and scaled by population, for quantity, quality and online visibility.[9] The maps show some well-defined spatial patterns with strong production of twin scientific knowledge in regions in Finland, Italy, Portugal, and Spain, and weaker twin scientific production in many Eastern countries' regions. Yet, most neighbour regions belong to different quartiles of twin, resulting blurred overall spatial patterns. This visual impression is confirmed by the results of the I Moran's tests of spatial correlation, which we report in Table A2 in the Appendix and indicate a moderate degree of spatial correlation of twin scientific knowledge (as well as of digital and green ones), with I Moran's values comprised between 0.03 for online visibility and 0.15 for quantity scaled by population. These spatial patterns differ quite substantially to those typically observed for patent activities, which tend to be more spatially clustered (as discussed, e.g., in Maurseth and Verspagen 2002; Moreno et al. 2005), indicating that scientific knowledge is less spatially embedded than technological one.

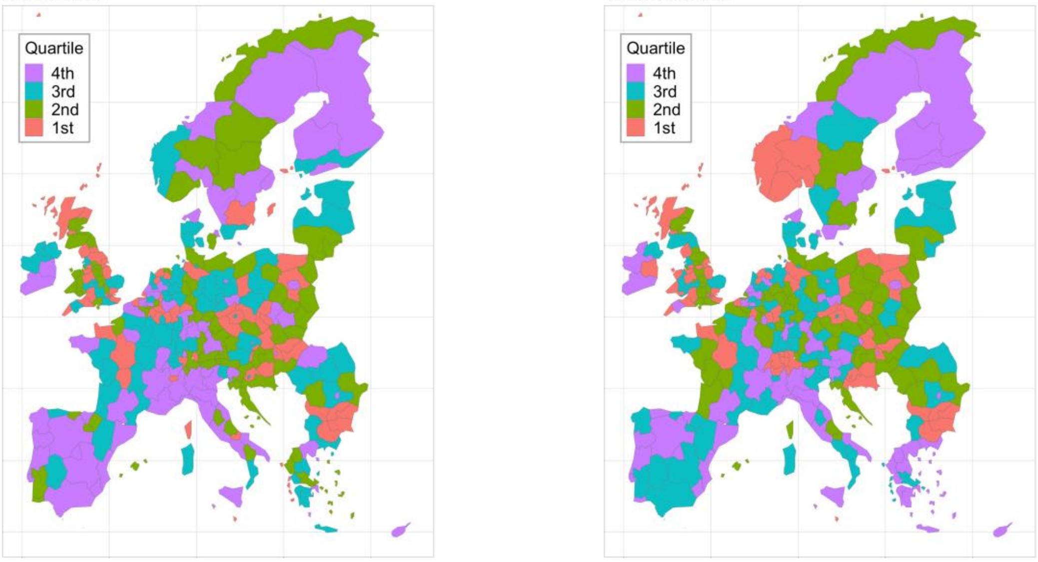
---

[8] We also experimented with BERTopic (Grootendorst 2022); however, the outcomes were unsatisfactory as the method exclusively identified two topics, namely green and digital.

[9] Figures A4 and A5 in the Appendix show analogous maps for digital ad green scientific knowledge.

Figure 2. Spatial distribution of twin publications, 2000-2020

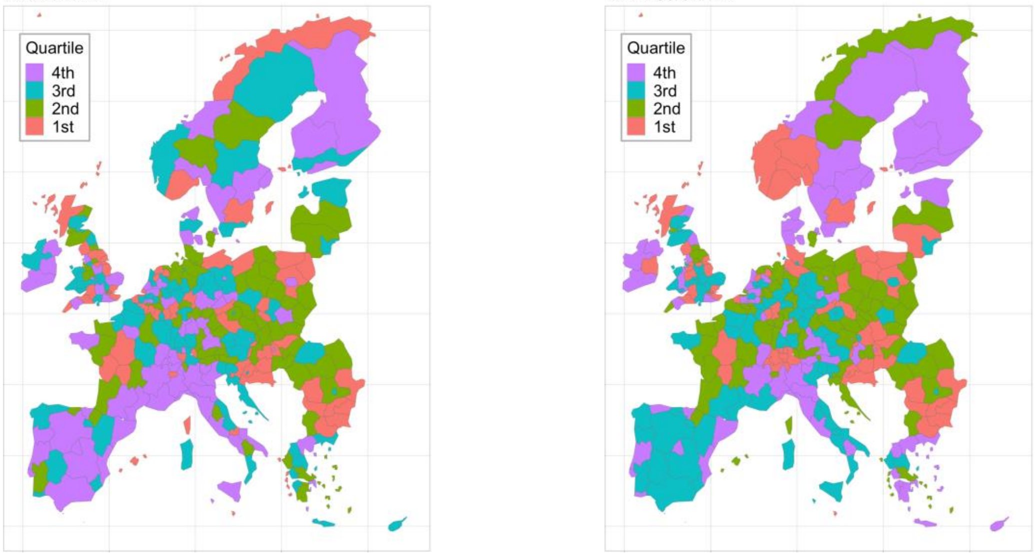
Quantity



Absolute number

Scaled by population

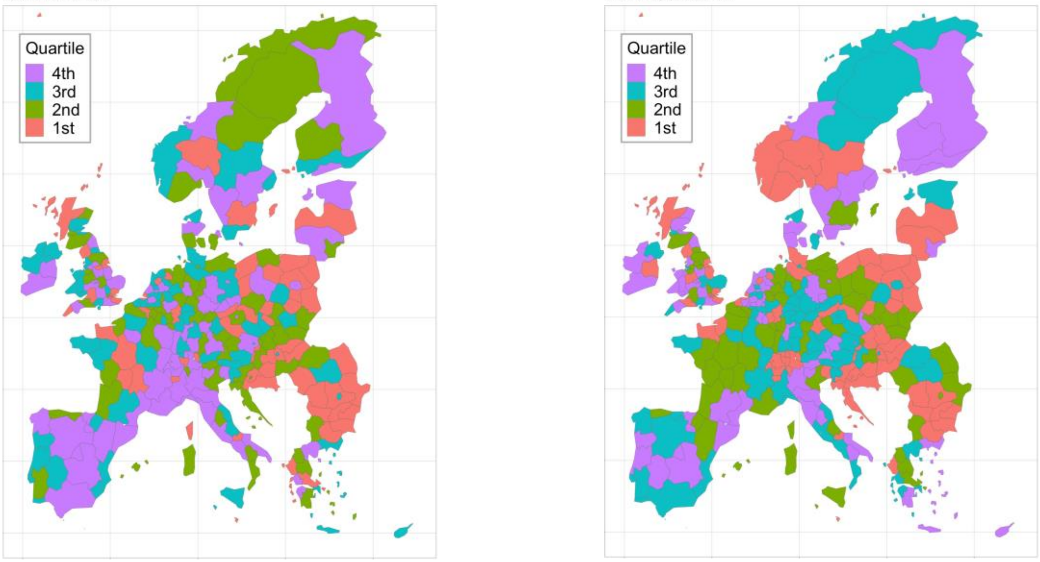
Quality



Absolute number

Scaled by population

Online visibility



Absolute number

Scaled by population

Table 3. Regional dynamics in scientific performance between 2001-2010 and 2011-2020

Scientific digital production in 2001-2010	Scientific digital production in 2011-2020					
	Zero or first quintile	Second quintile	Third quintile	Fourth quintile	Fifth quintile	Total
Zero or first quintile	**51**	7	0	0	0	58
Second quintile	7	**40**	10	0	0	57
Third quintile	0	10	**43**	5	0	58
Fourth quintile	0	0	5	**46**	7	58
Fifth quintile	0	0	0	6	**50**	56
Total	58	57	58	57	57	287

Scientific green production in 2001-2010	Scientific green production in 2011-2020					
	Zero or first quintile	Second quintile	Third quintile	Fourth quintile	Fifth quintile	Total
Zero or first quintile	**51**	8	0	0	0	59
Second quintile	7	**41**	8	0	0	56
Third quintile	0	8	**43**	9	0	60
Fourth quintile	0	0	7	**40**	8	55
Fifth quintile	0	0	0	8	**49**	57
Total	58	57	58	57	57	287

Scientific twin production in 2001-2010	Scientific twin production in 2011-2020					
	Zero or first quintile	Second quintile	Third quintile	Fourth quintile	Fifth quintile	Total
Zero or first quintile	**48**	30	15	3	0	96
Second quintile	15	**12**	17	8	2	54
Third quintile	0	8	**13**	8	5	34
Fourth quintile	0	3	13	**27**	13	56
Fifth quintile	0	0	3	8	**36**	47
Total	63	53	61	54	56	287

The emergence of the twin research domain and its spatial diffusion came along with a reshuffling of regional rankings along the distribution of twin knowledge (quantity).[10] Table 3 shows transition matrixes of regions across quintiles of the distribution of the number of publications over the 2001-2010 and 2011-2020 decades. Twin publications show the larger degree of reshuffling of regional rankings. In particular, slightly more than half of European NUTS 2 regions (52.3%) changed the quintile they belong to in the distribution of the number of twin publications across decades. Most of changing regions belonged to the first three quartiles, indicating a stronger persistence of regions with larger amount of twin publications. Yet, more than one third (35.7%) of the regions in the top quintile of the distribution of publications in 2011-2020 came from lower quantiles in the previous period. This implies that a large number of regions became scientific leaders in twin publications in less than a decade. By contrast, only about 20% of regions changed their quintile in the distribution of digital and green publications, and only a relatively small proportion (12.3% for digital; 14% for green) of regions managed to enter the top quintile in 2011-2020. All in all, we can conclude that the spatial distributions of green and digital publications are very persistent across regions, while twin publications are much more dynamic and volatile.

---

[10] See also Table A3 in the Appendix listing NUTS2 regions in the first decile of the distribution of twin, digital and green quantity.

4.2 Cognitive recombination of green and digital scientific knowledge into twin one

This section explores the role that digital and green knowledge bases play in the creation of twin scientific knowledge. We first compare the relative contribution of the overall digital and green domains (Subsection 4.2.1), and next focus on the recombination of green and digital subdomains characterizing twin knowledge creation (Subsection 4.2.2).

4.2.1 The role of green and digital scientific knowledge bases

Table 4 shows the results of regional fixed-effects models on twin scientific knowledge. Parameters' estimates can be interpreted as elasticities, with the cautionary remark that no claim of causality should be made. Green and digital scientific knowledge bases are positively and significantly associated with twin knowledge. Green scientific production is more strongly associated than digital one with the quantity of twin scientific production. The associations of green and digital scientific quality with twin one are broadly comparable in their magnitude, though the association of digital quality is at the edge of significance at the 10% level. The picture reverses when looking at online visibility of scientific production, with the online visibility of digital publications being more important than the one of green publications for the online visibility of twin ones.[11]

As for control variables, population density is strongly associated with all considered dimensions of twin scientific knowledge. R&D activities are positively associated with twin quantity and, to a lower extent, quality. General technological development is not associated with twin knowledge creation. By contrast, digital and green technological developments are positively associated with twin knowledge, especially digital technologies that are significant for all the dimensions of quantity, quality and visibility, while green ones only for quantity. Finally, the economic performance and the structural composition of the regional economy are not associated with twin scientific knowledge production.

4.2.2 Combinations of green and digital subdomains

Table 5 shows the most prevalent 20 combinations of digital and green subdomains. We first observe the high fragmentation of combinatory classes, with the largest of them – IoT-energy – for just 8.2% of twin publications and the largest 20 combinations for 62.2%. The most prevalent classes combine IoT and AI digital subdomains with the energy and environmental management green ones. Other subdomains include, for what concerns digital subdomains, big data, computing infrastructure and robotics, and for what concerns green ones, buildings and transport, often coupled with energy.

We used NLP methods to uncover the topics covered by the twin scientific literature. Table 6 shows the results of LDA topic modelling implemented on the corpus of twin publications' abstracts. Results provide a concrete idea of the main research topics covered in the combinatory classes shown in Table 5.

---

[11] The magnitude of the parameter's estimate of digital in the twin online visibility models reduces when using the "fractional counting" scheme, suggesting a similar role of digital and green domains (Table A4 in the Appendix reports results of the econometric model using the "fractional counting" scheme).

Table 4. Regional fixed-effects models on twin scientific knowledge

	Quantity		Quality		Online visibility	
	(1)	(2)	(3)	(4)	(5)	(6)
GREEN	0.1974***	0.1821***	0.1245***	0.1141***	0.1306***	0.1269***
	(0.0474)	(0.0425)	(0.0473)	(0.0427)	(0.0229)	(0.0235)
DIGIT	0.0991**	0.0993**	0.0852*	0.0769	0.2698***	0.2574***
	(0.0473)	(0.0434)	(0.0494)	(0.0467)	(0.0296)	(0.0303)
GDP	-0.0195	-0.2205	-0.2138	-0.5265	-0.0457	-0.0827
	(0.2476)	(0.2257)	(0.3752)	(0.3577)	(0.1880)	(0.1752)
DENSITY	1.7106**	1.0567*	3.2584***	2.3768**	2.2738***	2.1158***
	(0.6759)	(0.5986)	(1.2232)	(1.1456)	(0.6346)	(0.6360)
RD_GDP	0.1106***	0.0891**	0.1275*	0.0965	0.0694	0.0649
	(0.0419)	(0.0384)	(0.0731)	(0.0693)	(0.0463)	(0.0469)
DIRTY	0.0260	0.0180	0.0240	0.0131	-0.0153	-0.0173
	(0.0182)	(0.0174)	(0.0349)	(0.0344)	(0.0191)	(0.0188)
PAT	0.0186		0.0080		0.0001	
	(0.0602)		(0.1163)		(0.0538)	
GREENTECH		0.0820**		0.1198		-0.0537
		(0.0399)		(0.0818)		(0.0402)
DIGITECH		0.2080***		0.2907***		0.0902**
		(0.0366)		(0.0668)		(0.0407)
Constant	-10.4541**	-6.1364*	-18.2643**	-12.3856*	-12.9829***	-11.8400***
	(4.1481)	(3.6832)	(7.4031)	(6.9566)	(3.8254)	(3.8428)
N	4590	4590	4590	4590	4590	4590
Regions	270	270	270	270	270	270
Adj. R^2	0.642	0.662	0.639	0.648	0.614	0.616
Log-likelihood	-2557.1	-2425.4	-5715.6	-5650.9	-3757.8	-3745.2

Notes: robust standard errors are shown in parentheses. All models include year dummies. Variables other than the shares of energy-intensive industries (DIRTY) and R&D expenditures over GDP (RD_GDP) are transformed using the inverse hyperbolic sine transformation. *p<0.10, **p<0.05 and ***p<0.01.

Table 5. Combinations of the most frequent digital-green subdomains in twin publications

Digital-Green combination	**% of twin publications**	**Cumulative %**
IoT - Energy	8.2	8.2
IoT - Buildings/Energy	7.9	16.0
AI - Energy	7.0	23.0
AI - Environmental Management	5.4	28.4
IoT - Environmental Management	4.4	32.8
AI - Buildings/Energy	3.5	36.3
AI - Energy/Transport	2.7	39.1
Digital Infrastructure - Buildings/Energy	2.5	41.6
AI - Waste-Water	2.5	44.0
AI - Adaptation: Water	2.5	46.5
Robotics - Transport	2.3	48.8
IoT - Energy/Transport	2.3	51.1
Robotics - Environmental Management	1.9	52.9
AI - Transport	1.9	54.8
Big data - Buildings/Energy	1.5	56.3
Robotics - Energy	1.4	57.7
Robotics - Buildings/Energy	1.2	58.9
Digital Infrastructure - Energy	1.2	60.0
AI - Buildings	1.1	61.2
Big data/ Digital Infrastructure - Buildings/Energy	1.1	62.2

Notes: the table shows the 20 largest combinations of digital-green subdomains in twin publications.

The first column of Table 6 shows a label for each of the six topics. Such labels were provided to us by ChatGPT as an output to the following query: “We applied LDA to a set of documents and obtained 6 topics. Here are the most recurrent terms for each topic [list of the 50 most frequent bigrams and trigrams of each topic]. Can you assign a label to each topic?”. After verifying the appropriateness of labels, and slightly modifying one of them[12], we came to the following six descriptions: i) energy management (distribution, efficiency, use); (ii) water systems management and modeling; (iii) wireless sensor networks for environmental monitoring; (iv) environmental sustainability in urban planning; (v) air quality and pollution modeling; and (vi) energy storage and harvesting technologies.

The topic “energy management (production, consumption, distribution)” is the most prevalent, being covered in about 30% of twin publications. Papers in this category mainly deal with smart grids, power systems, and energy management for IoT applications in smart homes. Slightly less frequent are the topics “water systems management and modeling” (22%) and “wireless sensor networks for environmental monitoring” (20%). In the former category, we identified studies primarily focusing on machine learning predicting water quality and wastewater modeling, while the latter category includes contributions mostly addressing energy conservation in wireless sensor networks and IoT applications for environmental monitoring. It is worth noting that these two categories are substantially less important when weighted by citations and online mentions. By contrast, the category “energy storage and harvesting technologies”, which accounts for just about 7% of the twin publications, exceeds 20% of total citations and more than 25% of total online mentions. Similar, though less pronounced, is the pattern for the category “air quality and pollution modeling”.

Table 6. Combinatory classes of twin scientific knowledge

Class	Representative terms	Digital and green subdomains	Examples of twin papers	Share		
				Quantity	Quality	Visibility
Energy management (distribution, efficiency, use)	smart grid, electric vehicle, energy management, internet of things (IoT), energy saving, smart home, neural network, power system, power grid	Digital: AI, IoT Green: Buildings, Energy, Transport	A blockchain-based smart grid: towards sustainable local energy markets – https://doi.org/10.1007/s00450-017-0360-9 5G network-based Internet of Things for demand response in smart grid: A survey on application potential – https://doi.org/10.1016/j.apenergy.2019.113972 A novel energy management approach for smart homes using Bluetooth low energy – 10.1109/JSAC.2015.2481203	0.284	0.245	0.322
Water systems management and modeling	neural network, machine learning, wastewater treatment, treatment plants, water	Digital: AI Green: Wastewater, Water	Monitoring and detecting faults in wastewater treatment plants using deep learning – https://doi.org/10.1007/s10661-020-8064-1	0.223	0.106	0.078

---

[12] ChatGPT label of topic i) was “Smart Grid and Energy Management”. We opted for a more general re-labeling as “energy management (distribution, efficiency, use)”.

Class	Representative terms	Digital and green subdomains	Examples of twin papers	Share		
				Quantity	Quality	Visibility
	quality, water distribution		River water modelling prediction using multi-linear regression, artificial neural network, and adaptive neuro-fuzzy inference system techniques – https://doi.org/10.1016/j.procs.2017.11.212 Neural networks modelling of pesticides removal by activated carbon for water treatment - https://doi.org/10.2166/ws.2004.0087			
Wireless sensor networks for environmental monitoring	sensor network, wireless sensor, energy efficiency, energy saving, power consumption, environmental monitoring	Digital: IoT Green: Energy, Environmental management	Energy conservation in wireless sensor networks: a survey – https://doi.org/10.1016/j.adhoc.2008.06.003 Design of a WSN platform for long-term environmental monitoring for IoT applications – 10.1109/JETCAS.2013.2243032 Analysis of three IoT-based wireless sensors for environmental monitoring – 10.1109/TIM.2017.2771979	0.200	0.161	0.165
Environmental sustainability in urban planning	big data, internet of things (IoT), climate change, smart city, renewable energy, waste management, waste collection, sustainable development	Digital: Big data, IoT Green: Energy, Environmental management	Energy sustainability in smart cities: artificial intelligence, smart monitoring, and optimization of energy consumption – https://doi.org/10.3390/en11112869 Big data: the key to energy efficiency in smart buildings – https://doi.org/10.1007/s00500-015-1679-4 How can we tackle energy efficiency in IoT based smart buildings? – https://doi.org/10.3390/s140609582	0.125	0.104	0.064
Air quality and pollution modeling	air pollution, air quality, air conditioning, machine learning, time series, air pollutant, particulate matter,	Digital: AI Green: Environmental management	Air quality index and air pollutant concentration prediction based on machine learning algorithms – https://doi.org/10.3390/app9194069 An application of machine learning methods to PM10 level medium-term prediction –	0.093	0.152	0.114

Class	Representative terms	Digital and green subdomains	Examples of twin papers	Share		
				Quantity	Quality	Visibility
	forecasting model		https://doi.org/10.1007/978-3-540-74829-8_32 Autonomous vehicles opportunities for cities air quality – https://doi.org/10.1016/j.scitotenv.2020.136546			
Energy storage and harvesting technologies	energy storage, energy consumption, energy harvesting, power management, storage device, wearable electronics	Digital: IoT Green: Energy	A review on lithium-ion battery ageing mechanisms and estimations for automotive applications – https://doi.org/10.1016/j.jpowsour.2013.05.040 Lab-on-skin: a review of flexible and stretchable electronics for wearable health monitoring – https://doi.org/10.1021/acsnano.7b04898 Thermoelectric energy harvesting of human body heat for wearable sensors – 10.1109/JSEN.2013.2252526	0.074	0.232	0.257

Notes: Classes (or topics) are identified with the Latent Dirichlet Allocation (LDA) topic modelling method.

5. Conclusions

The promotion of digital technologies applied to environmental issues is a key goal of many governments worldwide. In the EU, it lies at the core of the current policy agenda. This paper studies the spatial and combinatory patterns characterizing the emergence in recent years (2000-2020) in Europe of scientific knowledge in the so called "twin" domain, i.e., scientific research on both green and digital technologies.

This knowledge field is emerging at a fast pace, driven by the surge of scientific knowledge on advanced digital technologies. Twin scientific knowledge creation diffused in the last twenty years to an increasing number of European regions. The diffusion of twin knowledge creation did not follow clearly defined spatial patterns and was accompanied by a reshuffling of regions' rankings: some regions climbed the ladder, others fell behind, in a dynamic process of actor reconfiguration. The presence of geographically dispersed and changing actors certainly raises appealing opportunities, but also poses coordination challenges for policies that – such as the new European Research Area for Research and Innovation that the EU launched in 2020 – aim at supporting science and research cooperation in twin domains.

Twin scientific knowledge is larger, of better quality, and more visible in regions well-endowed with strong green and digital science bases. While the green knowledge base plays a more important role than the digital one for the quantity of twin scientific production, its quality and online visibility depend in a more balanced way from both digital and green science bases. Twin scientific knowledge tends to focus on combinations of the AI and IoT digital domains with the energy and environmental management green ones. Twin domains are emerging on the study of AI- and IoT-powered applications for energy storage, distribution and consumption, environmental monitoring and modeling, and urban planning. An established scientific knowledge base for defined twin domains is valuable for policies that – such as the Horizon Europe EU funding programme for research and innovation and the NextGenerationEU recovery package – aim at guiding research and development activities towards technological advancements contributing to tackle global challenges as environmental degradation and climate change.

This work represents a first step in an attempt to understand the extent to which the combination of digital and green knowledge can promote sustainability and contribute to solving environmental problems. As such, it could be extended in several ways, including the study of the links between science and technology, for instance by looking at the impact on technological change of scientific knowledge recently created in the twin domain. Another direction for future research could be the study of the determinants of online visibility of scientific knowledge. Our analysis suggests that online visibility is, as it could be considered desirable for the social good, related more to the quality than to the quantity of scientific production. In that respect, a dedicated analysis could provide more nuanced understanding of public opinion’s drivers.

.

References

Aihounton, G. B., & Henningsen, A. (2021). Units of measurement and the inverse hyperbolic sine transformation. The Econometrics Journal, 24(2), 334-351.

Allen, R. C. (2017). Lessons from history for the future of work. Nature, 550(7676), 321–324.

Anselin, L., Bera, A. K., Florax, R. and Yoon, M. J. (1996). Simple diagnostic tests for spatial dependence. Regional Science and Urban Economics, 26 (1): 77–104.

Ardito, L., S. Raby, V. Albino, and B. Bertoldi. 2021. "The Duality of Digital and Environmental Orientations in the Context of SMEs: Implications for Innovation Performance." Journal of Business Research 123:44–56.

Arthur, W. B. (2009). The nature of technology: What it is and how it evolves. Simon and Schuster.

Ayres, R.U., van den Bergh, J.C.J.M., 2005. A theory of economic growth with material/ energy resources and dematerialization: interaction of three growth mechanisms. Ecol. Econ. 55 (1), 96–118.

Balland, P. A. and Boschma, R. (2021) Mapping the potentials of regions in Europe to contribute to new knowledge production in Industry 4.0 technologies, Regional Studies, 55:10-11, 1652-1666.

Barbieri, N., Marzucchi, A., & Rizzo, U. (2020a). Knowledge sources and impacts on subsequent inventions: Do green technologies differ from non-green ones? Research Policy, 49(2), 103901.

Barbieri, N., Perruchas, F., Consoli, D., (2020b). Specialization, diversification and environmental technology life-cycle. Economic Geography. Economic Geography, 96(2), 161-186.Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. Journal of machine Learning research, 3, 993-1022.

Bellandi, M., Santini, E., & Vecciolini, C. (2018). Learning, unlearning and forgetting processes in industrial districts. Cambridge Journal of Economics, 42(6), 1671-1685.

Besiroglu, T., Emery-Xu, N., & Thompson, N. (2022). Economic impacts of AI-augmented R&D. *arXiv preprint arXiv:2212.08198*.

Bianchini, S., Müller, M., & Pelletier, P. (2022). Artificial intelligence in science: An emerging general method of invention. Research Policy, 51(10), 104604.

Bianchini, S., Damioli, G., & Ghisetti, C. (2023). The environmental effects of the "twin" green and digital transition in European regions. Environmental and Resource Economics, 84(4), 877-918.

Boschma, R. (2017). Relatedness as driver of regional diversification: A research agenda. Regional Studies, 51(3), 351-364.

Bresnahan, T. F., & Trajtenberg, M. (1995). General purpose technologies 'Engines of growth'?. Journal of econometrics, 65(1), 83-108.

Burbidge, J. B., L. Magee and A. L. Robb (1988). Alternative transformations to handle extreme values of the dependent variable. Journal of the American Statistical Association 83, 123–7.

Capello, R. (2002). Entrepreneurship and spatial externalities: Theory and measurement. The Annals of Regional Science, 36(3), 387-402.

Capello, R., & Lenzi, C. (2021). 4.0 Technologies and the rise of new islands of innovation in European regions. Regional Studies, 55(10-11), 1724-1737.

Cassiman, B. R. Veugelers, P. Zuniga, In search of performance effects of (in)direct industry science links, Ind. Corps Change 17 (2008) 611–646.

Cattani, L., Montresor, S., & Vezzani, A. (2023). Firms' eco-innovation and Industry 4.0 technologies in urban and rural areas. Regional Studies, 1-13.

Chen, F., Deng, P., Wan, J., Zhang, D., Vasilakos, A. V., & Rong, X. (2015). Data mining for the internet of things: literature review and challenges. International Journal of Distributed Sensor Networks, 11(8), 431047.

Cicerone, G., Faggian, A., Montresor, S. & Rentocchini, F. (2022): Regional artificial intelligence and the geography of environmental technologies: does local AI knowledge help regional green-tech specialization?, Regional Studies.

Ciffolilli, A., & Muscio, A. (2018). Industry 4.0: national and regional comparative advantages in key enabling technologies. European Planning Studies, 26(12), 2323-2343.

Cockburn, I. M., Henderson, R., & Stern, S. (2018). The impact of artificial intelligence on innovation: An exploratory analysis. In The economics of artificial intelligence: An agenda (pp. 115-146). University of Chicago Press.

Corradini, C. (2019). Location determinants of green technological entry: Evidence from European regions. Small Business Economics, 52(4), 845–858.

Corradini, C., Santini, E., & Vecciolini, C. (2021). The geography of Industry 4.0 technologies across European regions. Regional Studies, 55(10-11), 1667-1680.Cowls, J. (2021). 'AI for Social Good': Whose Good and Who's Good? Introduction to the Special Issue on Artificial Intelligence for Social Good. Philosophy & Technology, 34(Suppl 1), 1-5.

Creutzig, F., Acemoglu, D., Bai, X., Edwards, P. N., Hintz, M. J., Kaack, L. H., ... & Zumwald, M. (2022). Digitalization and the Anthropocene. Annual review of environment and resources, 47, 479-509.

Dasgupta,P. P. David, Special Issue in Honor of Nathan Rosenberg. Toward a new economics of science, Res. Policy 23 (1994) 487–521.

De Marchi, V., & Di Maria, E. (2020). Achieving circular economy via the adoption of Industry 4.0 technologies: A Knowledge management perspective. Knowledge Management and Industry 4.0: New Paradigms for Value Creation, 163-178

De Propris, L., & Storai, D. (2019). Servitizing industrial regions. Regional Studies, 53(3), 388–397.

Dibiaggio, L. M. Nasiriyar, L. Nesta, Substitutability and complementarity of technological knowledge and the inventive performance of semiconductor companies, Res. Policy 43 (2014) 1582–1593.

Dosi, G. (1982). Technological paradigms and technological trajectories: a suggested interpretation of the determinants and directions of technical change. Research Policy, 11(3), 147-162.

Fleming, L. O. Sorenson, Science as a map in technological search, Strateg. Manag. J. 25 (2004) 909–928.

Freeman, C. (1987). Technology, policy, and economic performance: Lessons from Japan. Pinter.

Goldfarb, A., Taska, B., & Teodoridis, F. (2023). Could machine learning be a general purpose technology? a comparison of emerging technologies using data from online job postings. Research Policy, 52(1), 104653.

Goralski, M. A., & Tan, T. K. (2020). Artificial intelligence and sustainable development. The International Journal of Management Education, 18(1), 100330.

Grootendorst, M. (2022). BERTopic: Neural topic modeling with a class-based TF-IDF procedure. arXiv preprint arXiv:2203.05794.

Guenat, S., Purnell, P., Davies, Z. G., Nawrath, M., Stringer, L. C., Babu, G. R., ... & Dallimer, M. (2022). Meeting sustainable development goals via robotics and autonomous systems. Nature communications, 13(1), 3559.

Haščič, I., & Migotto, M. (2015). Measuring environmental innovation using patent data. OECD Environment Working Papers No. 89, OECD, Paris.

Jaffe, A. B., Trajtenberg, M., & Henderson, R. (1993). Geographic localization of knowledge spillovers as evidenced by patent citations. Quarterly Journal of Economics, 108(3): 577–598.

Jelodar, H., Wang, Y., Yuan, C., Feng, X., Jiang, X., Li, Y., & Zhao, L. (2019). Latent Dirichlet allocation (LDA) and topic modeling: models, applications, a survey. Multimedia Tools and Applications, 78, 15169-15211.

Lacerda, J. S. (2019). Linking scientific knowledge and technological change: lessons from wind turbine evolution and innovation Energy research & Social Science, 55: 92-105.

Lewis, S. L., & Maslin, M. A. (2015). Defining the anthropocene. Nature, 519(7542), 171-180.

Magerman,T. B.V. Looy, K. Debackere, Does involvement in patenting jeopardize one's academic footprint? An analysis of patent-paper pairs in biotechnology, Res. Policy 44 (2015) 1702–1713.

Malerba, F., & Orsenigo, L. (1996). Schumpeterian patterns of innovation are technology-specific. Research policy, 25(3), 451-478.

Martinelli, A., Mina, A., & Moggi, M. (2021). The enabling technologies of industry 4.0: examining the seeds of the fourth industrial revolution. Industrial and Corporate Change, 30(1), 161-188.

Maurseth, P. B., & Verspagen, B. (2002). Knowledge spillovers in Europe: a patent citations analysis. Scandinavian Journal of Economics, 104(4), 531-545.

Mazzucato, M. The entrepreneurial state: debunking public vs. private sector myths, Anthem Frontiers of Global Political Economy, revised ed., Anthem Press, London, 2014.

Mitchell, T., & Brynjolfsson, E. (2017). Track how technology is transforming work. Nature, 544(7650), 290-292.

Montresor, S., and Quatraro, F. (2020). Green technologies and smart specialisation: A European patent-based analysis of the intertwining of technological relatedness and key enabling technologies. Regional Studies, 54(10), 1354–1365.

Montresor, S., and A. Vezzani. 2023. “Digital Technologies and Eco-Innovation. Evidence of the Twin Transition from Italian Firms.” Industry and Innovation 1–35.

Moore, J. (2019). AI for not bad. Frontiers in Big Data, 2, 32.

Moreno, R., Paci, R., & Usai, S. (2005). Spatial spillovers and innovation activity in European regions. Environment and planning A, 37(10), 1793-1812.

Moreno, R., & Ocampo-Corrales, D. (2022). The ability of European regions to diversify in renewable energies: The role of technological relatedness. Research Policy, 51(5), 104508.

Mourtzis, D., Vlachou, E., & Milas, N. J. P. C. (2016). Industrial big data as a result of IoT adoption in manufacturing. Procedia cirp, 55, 290-295.

Murray, F. (2002) Innovation as co-evolution of scientific and technological networks: exploring tissue engineering, Res. Policy 31, 1389–1403.

Nightingale, P. The product–process–organisation relationship in complex development projects, Res. Policy 29 (2000) 913–930.

Noailly, J., and Shestalova, V. (2017). Knowledge spillovers from renewable energy technologies: Lessons from patent citations. Environmental Innovation and Societal Transitions, 22, 1-14.

Ocampo, D., Moreno, R., Suriñach, J., 2021. Knowledge flows and technologies in renewable energies at the regional level in Europe. Reg. Stud. 55 (3), 521–532.

OECD (2019). Going Digital: Shaping Policies, Improving Lives. OECD Publishing, Paris.

Popp, D. (2016). Economic analysis of scientific publications and implications for energy research and development. Nature Energy, 1(4), 1-8.

Rennings, K. (2000). Redefining innovation—eco-innovation research and the contribution from ecological economics. Ecological economics, 32(2), 319-332.

Rockström, J., Steffen, W., Noone, K., Persson, Å., Chapin, F. S., Lambin, E. F., ... & Foley, J. A. (2009). A safe operating space for humanity. Nature, 461(7263), 472-475.

Rosenberg, N. and R.R. Nelson, American universities and technical advance in industry, Res. Policy 23 (1994) 323–348.

Schwab, K. (2017). The Fourth Industrial Revolution. Crown Business.

O. Sorenson, J.W. Rivkin, L. Fleming, Complexity, networks and knowledge flow, Res. Policy 35 (2006) 994–1017.

Storper, M. (2018). Regional innovation transitions. In J. Glückler, R. Suddaby, & R. Lenz (Eds.), Knowledge and institutions: Knowledge and space (Vol. 13, pp. 197–225). Springer.

Taylor, L. (2016). The ethics of big data as a public good: which public? Whose good?. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 374(2083), 20160126.

Thu, M. K., Beppu, S., Yarime, M., & Shibayama, S. (2022). Role of machine and organizational structure in science. PloS one, 17(8), e0272280.

Truffer, B., & Coenen, L. (2012). Environmental innovation and sustainability transitions in regional studies. Regional studies, 46(1), 1-21.

Usher, A. P. (1954). A history of mechanical inventions. Courier Corporation.

Vinuesa, R., Azizpour, H., Leite, I., Balaam, M., Dignum, V., Domisch, S., ... & Fuso Nerini, F. (2020). The role of artificial intelligence in achieving the Sustainable Development Goals. Nature Communications, 11(1), 1-10.

V Van Roy, D Vertesy, G Damioli (2019). AI and robotics innovation. Handbook of labor, human resources and population economics, 1-35.

World Bank (2017). Trouble in the making? The future of manufacturing-led development. International Bank for Reconstruction and Development/The World Bank.

Appendix. Additional statistics

Figure A1. Digital technologies keywords by subdomain

Additive manufacturing	Artificial Intelligence	Big data	Blockchain	Computing infrastructure	Internet of Things	Robotics
3d print	artificial intellig	apache spark	altcoin	hyper connectivity	connected device	autonomous car
3d prototyping	artificial realit	apache cassandra	bitcoin	**5G networks**	connected home	autonomous underwater vehicle
additive fabrication	augmented realit	big data	blockchain	5th generation mobile	cyber-physical system	autonomous vehicle
additive layer manufacturing	automated reasoning	data center	cryptocurrenc	fifth generation mobile	human-machine interface	auv
additive manufacturing	backpropagation	data centre	private blockcha	**Cloud computing**	industrial internet of things	chatbot
direct digital manufacturing	computer-mediated realit	hadoop	public blockchai	cloud application	intelligent factor	drone
layered manufacturing	computer vision	large-scale data		cloud architecture	internet of everything	humanoid robot
rapid prototyping	data mining	mapreduce		cloud broker	internet of things	manipulator
	data science	massive data		cloud client	iot	mobile manipulator
	deep learning			cloud computing	machine-to-enterprise	mobile robot
	expert system			cloud infrastructure	machine-to-human	robot
	face detection			cloud migration	machine-to-machine	robotic
	feature extraction			cloud optimizer	pervasive sensing	self-driving car
	generative advesartial network			cloud portfolio	sensor network	self-driving vehicle
	gesture recognition			cloud provider	smart device	uav
	image classification			cloud server	smart factor	ugv
	image recognition			cloud service	smart home	uncrewed vehicle
	image segmentation			cloud sourcing	smart sensor	unmanned aerial vehicle
	information retrieval			cloud storage	wearable	unmanned air vehicle
	intelligent machine			clould platform	wireless body area network	unmanned ground vehicle
	kernel machine			community cloud	wireless sensor network	unmanned spacecraft
	knowledge representation			dynamic cloud		unmanned underwater vehicles
	machine intelligence			federated cloud		unmanned vehicle
	machine learning			hybrid cloud		unmanned aircraft system
	machine translation			infrastructure as a service		
	meta-learning			inter-cloud computin		
	mixed realit			multi-cloud		
	multilayer perceptron			on-demand computing		
	natural language processing			platform as a service		
	neural net			private cloud		
	object detection			public cloud		
	object identification			software as a service		
	object recognition			**Computing power**		
	pattern recognition			cognitive comput		
	pose estimation			cyberinfrastructure		
	reinforcement learning			data-intensive comput		
	semantic search			hardware accelerator		
	semi-supervised learning			high performance comput		
	sentiment analysis			neuromorphic comput		
	speech recognition			optical comput		
	statistical learning			photonic comput		
	supervised learning			quantum comput		
	text classification			real-time comput		
	transfer learning			supercomput		
	transformer network					
	unsupervised learning					
	virtual realit					
	voice recognition					

Figure A2. Green technologies keywords by subdomain

Biodiversity	GHG capture	Mitigation: buildings	Mitigation: energy	Mitigation: production	Mitigation: transport	Mitigation: waste-water	Environmental management	Adaptation: water
biodiv	absorption	air condit	accumulator	afforestation	altern fuel	altern irrig	**Air pollution abatement**	desalin
ecosyst health	adsorption	bioethan	altern fuel	altern irrig	biodiesel	bio pack	air pollut	purif water
ecosyst serv	bio separ	cogenerat	batter	biofeedstock	bioethan	bio process	emission abat	sanitation
	carbon capt	efficien cook	biodiesel	bio plastic	biofuel	bio reac	emission mitigat	sterili water
	carbon stor	efficien cool	bioethan	bio reac	capacitor	bio treat	emission trad	water collect
	ccs	efficien heat	biofuel	eco design	eco design	disassembl	greenhouse gas	water conserv
	chem separ	efficien light	biogas	efficien input	emission mitigat	landfil	purif air	water distrib
	co2 capt	energ efficien	biomas	efficien output	efficien propuls	purif water	**Environmental management**	water stor
	co2 stor	energ light	efficien input	emission mitigat	efficien engin	remanufact	circular econ	water treat
	greenhouse gas capt	energ reduc	efficien output	environm control	electr motor	sanitation	clim change	
	greenhouse gas stor	energ sav	efficien power	material minimi	electr switc	sterili water	environm control	
	methan capt	energ us	emission mitigat	material process	electr vehic	waste collect	environm manag	
		insulat	energ alternat	material recover	electromobil	waste dismantl	pollut abat	
		led light	energ conserv	minimiz component	engin manag	waste process	**Environmental monitoring**	
		natural heat	energ efficien	minimiz material	filter vehic	waste separ	environm monitor	
		pv cell	energ harvest	modular design	flywheel	waste stor	**Soil remediation**	
			energ light	organic fertil	fuel alternat	waste transf	soil remed	
			energ optim	pesticid alternativ	fuel efficien	waste transport	**Waste management**	
			energ recover	process efficien	fuel pump	waste treat	mater reus	
			energ reduc	produc from waste	fuel sustain	wastewater treat	recycl	
			energ sav	reduc emission	hybrid vehic		remanufact	
			energ stor	reforestation	mech stor		reus	
			energ us	remanufact	natural gas vehic		waste management	
			fluid stor		regenerative brak			
			fuel cell		vehic charg			
			fuel efficien		vehic design			
			geotherm					
			hybrid cell					
			hydro energ					
			hydro power					
			hydroelectric					
			hydrogen					
			marin energ					
			mech stor					
			ocean energ					
			photovolt					
			pump stor					
			ren energ					
			smart grid					
			solar cell					
			solar concentrat					
			solar energ					
			solar heat					
			solar pond					
			superconduct elem					
			therm energ					
			therm stor					
			tidal					
			wind energ					
			wind power					
			wind turbin					

Figure A3 . The share of twin relative to digital and green publications

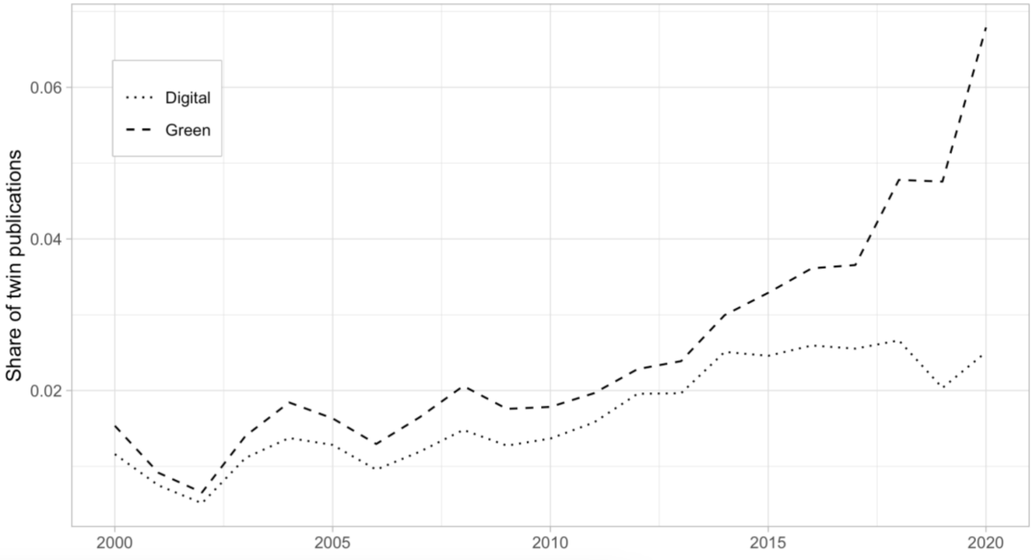
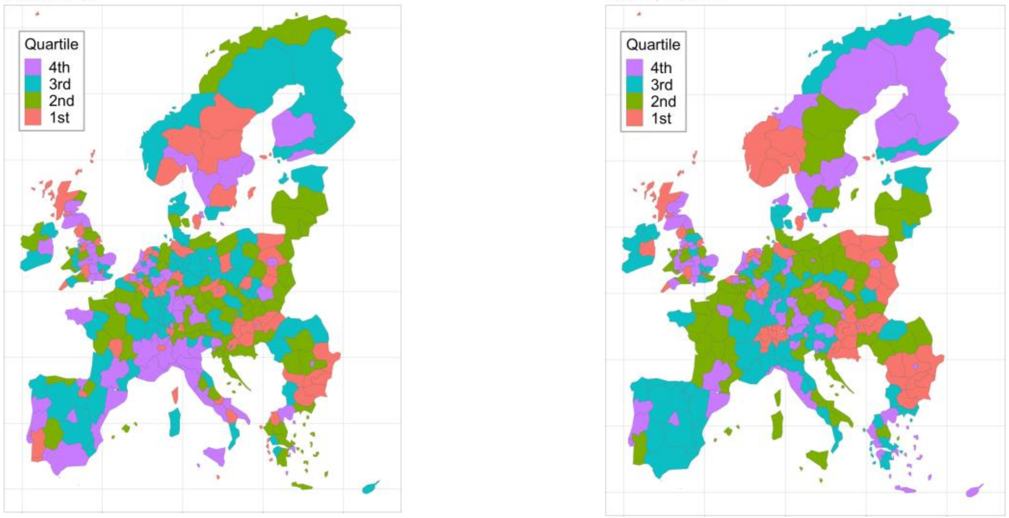
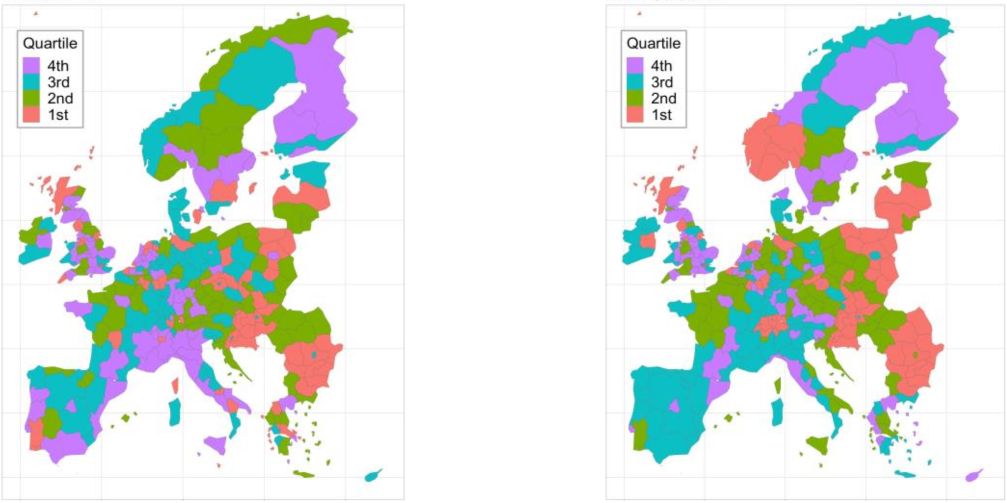


Figure A4. Spatial distribution of digital publications, 2000-2020

Quantity



Quality



Online visibility

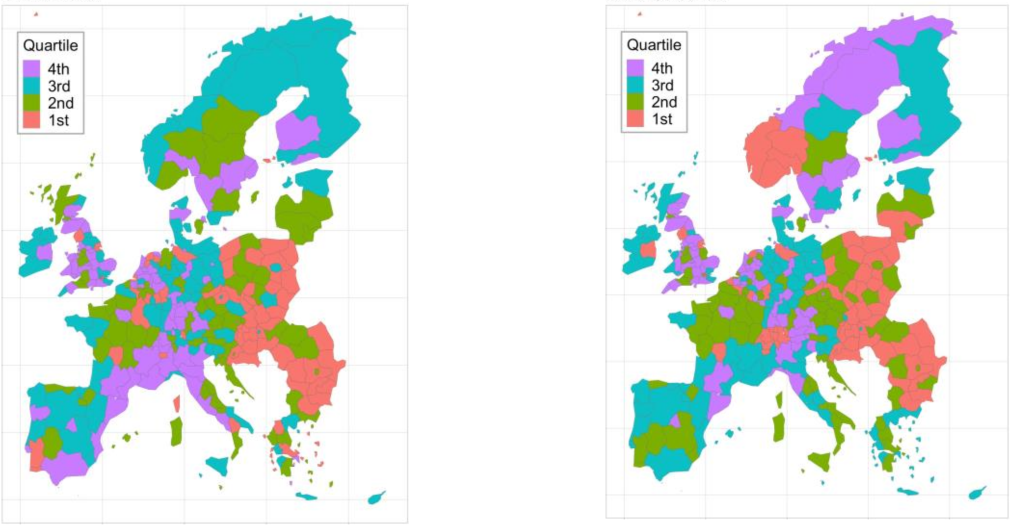
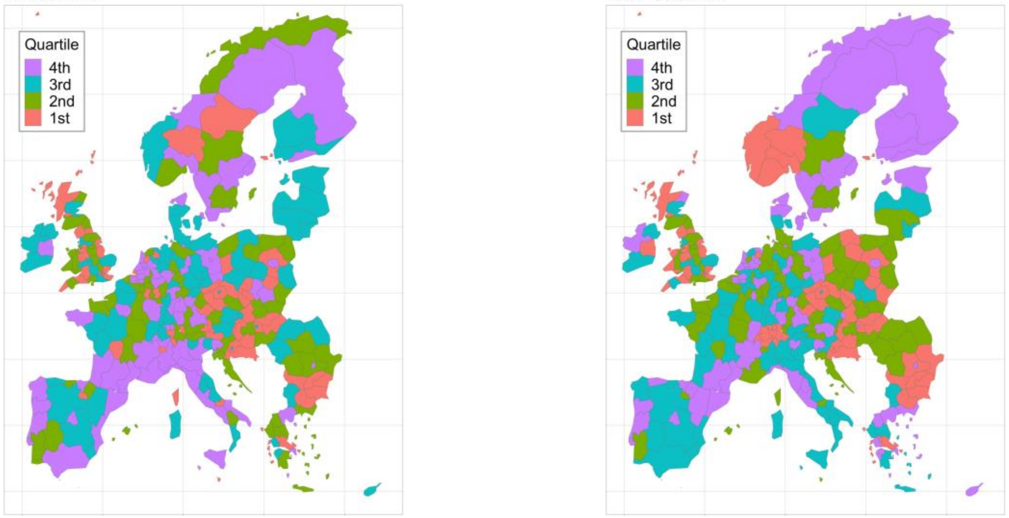
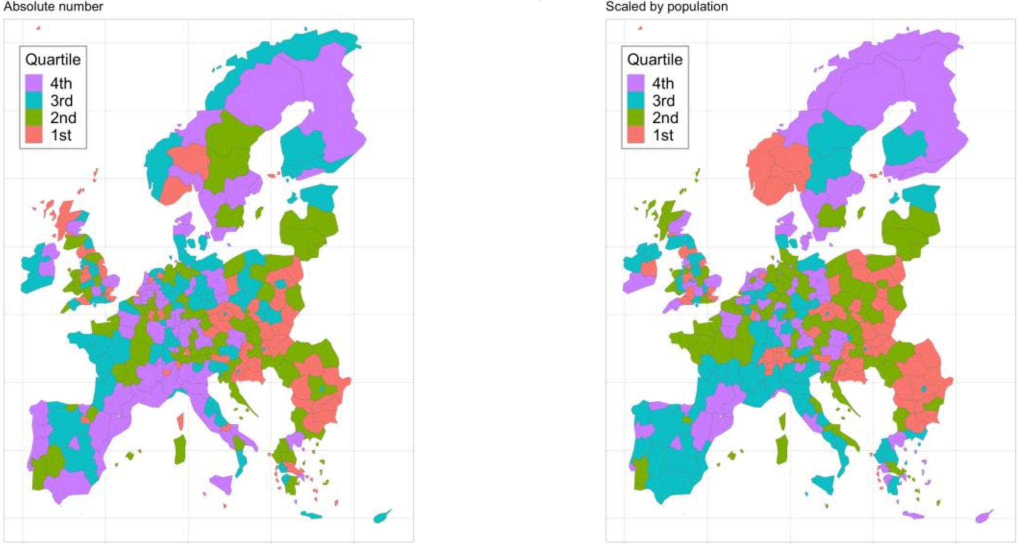


Figure A5. Spatial distribution of green publications, 2000-2020

Quantity



Quality



Online visibility

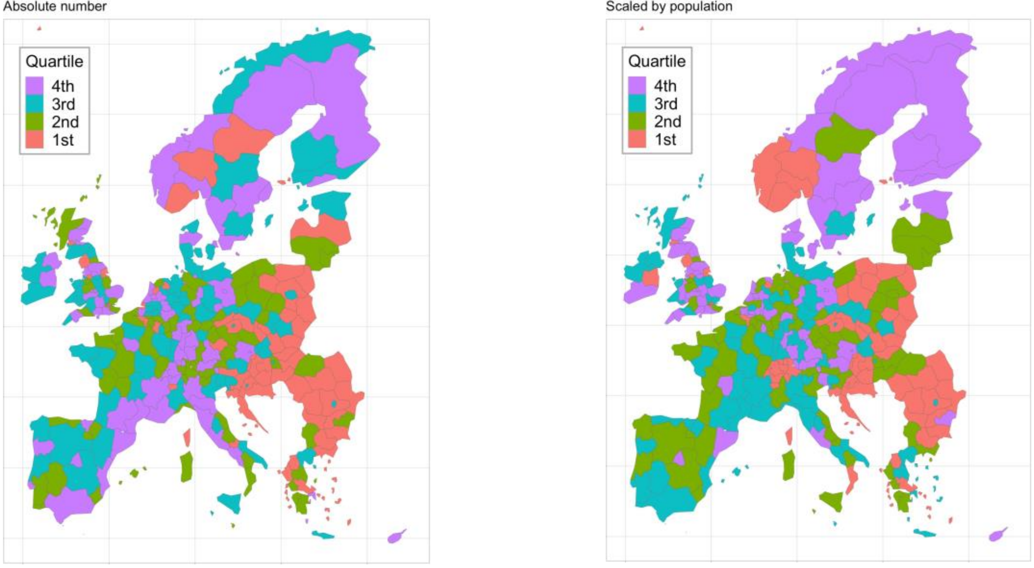


Table A1. Variables description and summary statistics

	Description	N	Mean	SD	Min	Max
TWIN QUANTITY	Publication in “twin”domains	4590	1.57	3.62	0	51
TWIN QUALITY	Average citations of publications in “twin” domains to account for quality	4590	7.66	25.1	0	387.3
TWIN VISIBILITY	Publication score in “twin”domains to account for visibility	4590	0.84	6.28	0	189.2
GREEN QUANTITY	Publications in “green” domains”	4590	51.9	82.4	0	778
GREEN QUALITY	Average citations of publications in “green” domains to account for quality	4590	182	310.9	0	3078.8
GREEN VISIBILITY	Visibility of “green” domains	4590	83	312.9	0	4722.6
DIGIT QUANTITY	Publications in “digit” domains	4590	73.3	123.2	0	2140
DIGIT QUALITY	Average citations of publications in “digit” domains to account for quality	4590	274	651.2	0	17732.7
DIGIT VISIBILITY	Visibility of “digit” domains	4590	35.6	158.8	0	4686.8
GDP	Gross Domestic Product	4590	46.3	51.9	0.8	654.2
DENSITY	Density: Population over Area of the region	4590	408.1	1088.6	3.1	11124.7
RD_GDP	Share of regional R&D expenditures over GDP	4590	1.5	1.2	0.06	12.2
DIRTY	Share or dirty sector in the regional structural economic composition, constructed as the share of energy intensive units belonging to the manufacturing sectors over the total units. Dirty units are those in sectors: Food (C10) Steel and all the basic metals (C24); Chemicals fertilizers and all the manufacture of chemicals (C20); Cement, Lime, Ceramics, Glass and non-metallic mineral products (C23) and paper (C17).	4590	1.7	1.2	0	10.6
PAT	Stock patent applications (15% yearly depreciation rate).	4590	1106.4	3071.7	0	42136.7
GREENTECH	Stock patent applications in environmental technologies (15% yearly depreciation rate), using OECD ENVTECH	4590	110.4	353.2	0	5404.7
DIGITECH	Stock patent applications in digital technologies (15% yearly depreciation rate), following Bianchini et al. (2023)	4590	6.5	22.9	0	432.3

Table A2. MORAN's I tests for spatial correlation in scientific knowledge of European regions, 2000-2020

	DIGIT		**GREEN**		**TWIN**	
	Absolute number	**Scaled by population**	**Absolute number**	**Scaled by population**	**Absolute number**	**Scaled by population**
Quantity	-0.028	-0.027	0.082***	0.233***	0.116***	0.151***
	(0.034)	(0.030)	(0.034)	(0.034)	(0.034)	(0.034)
Quality	0.018	0.024	0.097***	0.227***	0.105***	0.094***
	(0.034)	(0.028)	(0.034)	(0.035)	(0.035)	(0.035)
Online visibility	0.052**	0.050**	0.085***	0.139***	0.030	0.119***
	(0.032)	(0.024)	(0.034)	(0.035)	(0.034)	(0.034)

Notes: standard errors in parentheses; * $p < 0.10$, ** $p < 0.05$, and *** $p < 0.01$.

Table A3. NUTS2 ranking in digital, green and twin publications (2000-2020)

Ranking	**Twin publications**		**Digital publications**		**Green publications**	
	NUTS 2 region	**Number**	**NUTS 2 region**	**Number**	**NUTS 2 region**	**Number**
1	Cataluña	311	Inner London — West	16703	Ile-de-France	8625
2	Lombardia	290	Ile-de-France	15215	Cataluña	7445
3	Ile-de-France	258	Comunidad de Madrid	10634	Lombardia	6386
4	Comunidad de Madrid	229	Lombardia	9677	Comunidad de Madrid	6002
5	Lazio	220	Cataluña	9368	Lazio	5570
6	Andalucía	202	Oberbayern	8913	Hovedstaden	5402
7	Zürich	199	Lazio	7650	Zuid-Holland	5217
8	Emilia-Romagna	185	Andalucía	7066	Rhône-Alpes	5131
9	Rhône-Alpes	181	Rhône-Alpes	6617	Köln	5072
10	Norte	170	Toscana	6608	Andalucía	5071
11	Αττική	166	Zürich	6318	Zürich	5059
12	Piemonte	164	Zuid-Holland	5946	Oberbayern	4341
13	Toscana	160	Berkshire, Buckinghamshire and Oxfordshire	5795	Stockholm	4245
14	Oberbayern	159	Köln	5791	Berlin	4122
15	Área Metropolitana de Lisboa	158	Région lémanique	5687	Karlsruhe	3925
16	Helsinki-Uusimaa	156	Comunitat Valenciana	5563	Helsinki-Uusimaa	3774
17	Campania	150	Berlin	5517	Campania	3615
18	Comunitat Valenciana	143	Karlsruhe	5455	Norte	3463
19	Zuid-Holland	134	Lancashire	5372	Área Metropolitana de Lisboa	3388
20	Sicilia	132	Área Metropolitana de Lisboa	5184	Emilia-Romagna	3383
21	București-Ilfov	129	Αττική	5158	Östra Mellansverige	3363
22	Centro (PT)	126	East Anglia	5098	Gelderland	3330
23	Karlsruhe	123	Norte	4910	Toscana	3305
24	Puglia	122	Emilia-Romagna	4873	Utrecht	3273
25	Köln	120	Eastern Scotland	4865	Αττική	3241
26	Stockholm	113	Campania	4720	Région lémanique	3205
27	Région lémanique	110	Stockholm	4688	Wien	3059
28	Wien	110	Warszawski stołeczny	4634	Oslo og Viken	3055

Notes: the table shows the rankings of the first decile of the distribution of regions in the number of twin, digital and green publications.

Table A4. Regional fixed-effects models on twin scientific knowledge using fractional counting

	(1) QUANTITY	(2) QUANTITY	(3) QUALITY	(4) QUALITY	(5) VISIBILITY	(6) VISIBILITY
GREEN_frac	0.2234***	0.2007***	0.1780***	0.1528***	0.1450***	0.1435***
	(0.0308)	(0.0286)	(0.0343)	(0.0314)	(0.0216)	(0.0219)
DIGIT_frac	0.0842***	0.0808***	0.1243***	0.1074***	0.1303***	0.1268***
	(0.0301)	(0.0282)	(0.0374)	(0.0351)	(0.0212)	(0.0216)
DIRTY	0.0168	0.0120	0.0223	0.0138	-0.0105	-0.0109
	(0.0110)	(0.0105)	(0.0244)	(0.0240)	(0.0087)	(0.0085)
GDP	-0.0985	-0.2068	-0.2805	-0.5522**	0.0305	0.0584
	(0.1712)	(0.1555)	(0.2922)	(0.2670)	(0.0939)	(0.0846)
DENSITY	1.2927***	0.8920**	2.8606***	2.1971***	0.6325**	0.5878*
	(0.4212)	(0.3796)	(0.8239)	(0.7614)	(0.3199)	(0.3290)
RD_GDP	0.0491**	0.0367	0.0781	0.0524	0.0106	0.0115
	(0.0247)	(0.0234)	(0.0522)	(0.0500)	(0.0214)	(0.0215)
PAT	0.0016		-0.0489		0.0208	
	(0.0389)		(0.0817)		(0.0240)	
GREENTECH		0.0348		0.0445		-0.0192
		(0.0230)		(0.0587)		(0.0175)
DIGITECH		0.1300***		0.2481***		0.0156
		(0.0233)		(0.0502)		(0.0204)
Constant	-7.4737***	-4.8783**	-15.5378***	-11.0970**	-3.8610**	-3.5722*
	(2.5757)	(2.3190)	(5.0724)	(4.7006)	(1.8955)	(1.9538)
N	4590	4590	4590	4590	4590	4590
Regions	270	270	270	270	270	270
Adj. R^2	0.531	0.555	0.585	0.601	0.491	0.491
Log-Likelihood	-426.9	-305.9	-3999.8	-3911.5	-889.9	-888.7

Notes: robust standard errors are shown in parentheses. All models include year dummies. Variables other than the shares of energy-intensive industries (DIRTY) and R&D expenditures over GDP (RD_GDP) are transformed using the inverse hyperbolic sine transformation. *p<0.10, **p<0.05 and ***p<0.01.

Table A5. Lagrange multiplier (LM) tests for spatial dependence

	Quantity		Quality		Online visibility	
	(1)	(2)	(3)	(4)	(5)	(6)
LM for spatial error	27.22***	26.51***	25.47***	25.09***	10.08***	11.12***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
LM for spatial lag	25.77***	26.84***	24.96***	23.89***	15.83***	15.55***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Robust LM for spatial error	1.47	0.41	0.52	1.21	1.45	0.44
	(0.22)	(0.52)	(0.47)	(0.27)	(0.29)	(0.51)
Robust LM for spatial lag	0.02	0.73	0.00	0.01	7.22**	4.87**
	(0.89)	(0.39)	(0.95)	(0.91)	(0.01)	(0.03)

Notes: p-values in parentheses. LM tests are based on the . Regional fixed-effects models reported in Table 4. * p < 0.10, ** p < 0.05, and *** p < 0.01.

Table A6. Spatial autoregressive regional fixed-effects models on twin scientific knowledge

	Quantity		Quality		Online visibility	
	(1)	(2)	(3)	(4)	(5)	(6)
GREEN	0.1925***	0.1776***	0.1250***	0.0253**	0.1289***	0.1253***
	(0.0203)	(0.0197)	(0.0227)	(0.0112)	(0.0126)	(0.0126)
DIGIT	0.0913***	0.0919***	0.0758***	0.0344***	0.2669***	0.2546***
	(0.0208)	(0.0202)	(0.0242)	(0.0120)	(0.0164)	(0.0165)
GDP	0.0193	-0.1873**	-0.1107	0.0282	-0.0139	-0.0441
	(0.0890)	(0.0836)	(0.1745)	(0.0824)	(0.1136)	(0.1087)
DENSITY	1.6590***	1.0280***	3.0465***	0.8493***	2.1498***	1.9900***
	(0.2185)	(0.2129)	(0.4357)	(0.2152)	(0.2952)	(0.2932)
RD_GDP	0.1110***	0.0897***	0.1301***	0.0949***	0.0731***	0.0689**
	(0.0211)	(0.0206)	(0.0420)	(0.0208)	(0.0276)	(0.0275)
DIRTY	0.0241***	0.0165**	0.0248	0.0114	-0.0125	-0.0145
	(0.0082)	(0.0080)	(0.0162)	(0.0080)	(0.0107)	(0.0107)
PAT	0.0115		0.0027		0.0056	
	(0.0259)		(0.0513)		(0.0335)	
GREENTECH		0.0775***		0.0869***		-0.0538**
		(0.0192)		(0.0195)		(0.0256)
DIGITECH		0.2055***		0.2080***		0.0891***
		(0.0136)		(0.0138)		(0.0188)
Spatial rho	0.1374***	0.1240***	0.1340***	0.1398***	0.0687***	0.0668***
	(0.0222)	(0.0219)	(0.0230)	(0.0219)	(0.0219)	(0.0218)
N	4590	4590	4590	4590	4590	4590
Regions	270	270	270	270	270	270
R^2	0.106	0.196	0.087	0.179	0.100	0.108
Log-likelihood	-2538.3	-2409.7	-5699	-2472.5	-3752.9	-3740.6

Notes: robust standard errors are shown in parentheses. The matrix W is built using the 5th-nearest neighbouring scheme through the Stata command *spwmatrix* and estimated through the command *xsmle.* All models include year dummies. Variables other than the shares of energy-intensive industries (DIRTY) and R&D expenditures over GDP (RD_GDP) are transformed using the inverse hyperbolic sine transformation. *p<0.10, **p<0.05 and ***p<0.01.