

Caldarola, Bernardo; Mazzilli, Dario; Napolitano, Lorenzo; Patelli, Aurelio; Sbardella, Angelica

Working Paper

Economic complexity and the sustainability transition: A review of data, methods, and literature

JRC Working Papers on Corporate R&D and Innovation (CoRDI), No. 01/2024

Provided in Cooperation with:

Joint Research Centre (JRC), European Commission

Suggested Citation: Caldarola, Bernardo; Mazzilli, Dario; Napolitano, Lorenzo; Patelli, Aurelio; Sbardella, Angelica (2024) : Economic complexity and the sustainability transition: A review of data, methods, and literature, JRC Working Papers on Corporate R&D and Innovation (CoRDI), No. 01/2024, European Commission, Joint Research Centre (JRC), Seville

This Version is available at:

<https://hdl.handle.net/10419/301914>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.



<https://creativecommons.org/licenses/by/4.0/>

Economic complexity and the sustainability transition: A review of data, methods, and literature

*JRC Working Papers on Corporate
R&D and Innovation (CoRDI)*
No 01/2024

Caldarola, B., Mazzilli, D., Napolitano, L., Patelli, A., Sbardella, A.



This publication is a working paper by the Joint Research Centre (JRC), the European Commission's science and knowledge service. It aims to provide evidence-based scientific support to the European policymaking process.

The contents of this publication do not necessarily reflect the position or opinion of the European Commission. Neither the European Commission nor any person acting on behalf of the Commission is responsible for the use that might be made of this publication. For information on the methodology and quality underlying the data used in this publication for which the source is neither Eurostat nor other Commission services, users should contact the referenced source. The designations employed and the presentation of material on the maps do not imply the expression of any opinion whatsoever on the part of the European Union concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries.

The *JRC Working Papers on Corporate R&D and Innovation (CoRDI)* address economic and policy questions related to industrial research and innovation and their contribution to the European fair and sustainable competitiveness and growth. They are produced under editorial supervision by the Industrial Innovation & Dynamics (IID) team of Unit B6 Industrial Strategy, Skills and Technology Transfer in the JRC Directorate B Fair and Sustainable Economy of the European Commission.

The CoRDI Working Papers are accessible electronically from the following repositories: RePec (<http://repec.org/>), the JRC Industrial Research and Innovation webpage (<http://iri.jrc.ec.europa.eu/home>), and the JRC Publication Repository (<https://publications.jrc.ec.europa.eu/repository/>).

Editorial Board: The *JRC Working Papers on Corporate R&D and Innovation* are published under the editorial supervision of Pietro Moncada-Paternò-Castello and Alexander Tübke in collaboration with Fernando Hervás, Koen Jonkers, Francesco Rentocchini at the European Commission – Joint Research Centre, and in cooperation with Sara Amoroso (German Institute for Economic Research, DEU), Michele Cincera (Solvay Brussels School of Economics and Management, Université Libre de Bruxelles, BEL), Alex Coad (Waseda University, Tokyo, JAP), Enrico Santarelli (University of Bologna, ITA), Daniel Vertesy (International Telecommunication Union, CHE – and UNU-MERIT, NLD), Antonio Vezzani (Roma Tre University, ITA); Marco Vivarelli (Università Cattolica del Sacro Cuore, Milan, ITA) and Zoltan Csefalvay (Mathias Corvinus Collegium, HUN). Any comments and submissions can be sent by email to: JRC-B6-secretariat@ec.europa.eu.

Contact information:

Pietro Moncada-Paternò-Castello

Address: European Commission – Joint Research Centre. Edificio Expo. c/ Inca Garcilaso, 3. E-41092 Seville (Spain)

E-mail: pietro.moncada-paterno-castello@ec.europa.eu

More information, including JRC.B6 activities and publications, is available at: <https://iri.jrc.ec.europa.eu/home/>.

EU Science Hub

<https://joint-research-centre.ec.europa.eu>

JRC137954

Seville: European Commission, 2024

© European Union, 2024



The reuse policy of the European Commission documents is implemented by the Commission Decision 2011/833/EU of 12 December 2011 on the reuse of Commission documents (OJ L 330, 14.12.2011, p. 39). Unless otherwise noted, the reuse of this document is authorised under the Creative Commons Attribution 4.0 International (CC BY 4.0) licence (<https://creativecommons.org/licenses/by/4.0/>). This means that reuse is allowed provided appropriate credit is given and any changes are indicated. For any use or reproduction of photos or other material that is not owned by the European Union permission must be sought directly from the copyright holders.

How to cite this working paper: European Commission, Joint Research Centre, Caldarola, B., Mazzilli, D., Napolitano, L., Patelli, A. and Sbardella, A., *Economic complexity and the sustainability transition: A review of data, methods, and literature*, European Commission, Seville, 2024, JRC137954.

Contents

Abstract.....	2
Acknowledgements.....	3
1 Introduction	4
2 Data and methods	7
3 Review of the literature	21
4 Conclusions.....	42
References	47
Appendix.....	59

Abstract

Economic Complexity (EC) methods have gained increasing popularity across fields and disciplines. In particular, the EC toolbox has proved particularly promising in the study of complex and interrelated phenomena, such as the transition towards a greener economy. Using the EC approach, scholars have been investigating the relationship between EC and sustainability, proposing to identify the distinguishing characteristics of green products and to assess the readiness of productive and technological structures for the sustainability transition. This article proposes to review and summarize the data, methods, and empirical literature that are relevant to the study of the sustainability transition from an EC perspective.

We review three distinct but connected blocks of literature on EC and environmental sustainability. First, we survey the evidence linking measures of EC to indicators related to environmental sustainability. Second, we review articles that strive to assess the green competitiveness of productive systems. Third, we examine evidence on green technological development and its connection to non-green knowledge bases. Finally, we summarize the findings for each block and identify avenues for further research in this recent and growing body of empirical literature.

Acknowledgements

DM, AP, and AS acknowledge the financial support under the National Recovery and Resilience Plan (NRRP), Mission 4, Component 2, Investment 1.1, Call for tender No. 104 published on 2.2.2022 by the Italian Ministry of University and Research (MUR), funded by the European Union – NextGenerationEU– Project Title "WECARE" – CUP 20223W2JKJ by the Italian Ministry of Ministry of University and Research (MUR).

DM, AP, and AS acknowledge the financial support under the National Recovery and Resilience Plan (NRRP), Mission 4, Component 2, Investment 1.1, Call for tender No. 1409 published on 14.9.2022 by the Italian Ministry of University and Research (MUR), funded by the European Union – Next Generation EU – Project Title "Triple T – Tackling a just Twin Transition: a complexity approach to the geography of capabilities, labour markets and inequalities" – CUP F53D23010800001 by the Italian Ministry of Ministry of University and Research (MUR).

The views expressed in this working paper do not necessarily reflect the official position of the European Commission.

1. Introduction

The notion of Economic Complexity (EC) has been widely used to encompass a set of methods that characterize the productive and technological composition of economies (countries, regions, cities) relying on complex systems science (Hidalgo and Hausmann, 2009; Tacchella et al., 2012). EC methods have proved to be particularly effective in predicting future patterns of economic growth using information on the export basket of countries (Tacchella et al., 2018). The key intuition behind the EC approach is that economic development and growth are the result of the specialization and diversification patterns of economies, which emerge from underlying hidden interactions between elements in the society (Balland et al., 2022; Pugliese et al., 2019a).

More specifically, EC focuses on the role played by the accumulation of (unobserved) productive and technological capabilities in driving economic diversification and growth (Hausmann et al., 2007, 2011; Pugliese et al., 2017; Sbardella et al., 2018b). By preserving information on *what* economies produce, rather than merely *how much*, the EC literature is able to “describe and compare economies in a manner that eschews aggregation” (Hausmann and Hidalgo, 2011). The EC approach yields a complementary perspective on several domains of economically relevant human activity (e.g., trade, technical innovation, scientific research) with respect to conventional (i.e., aggregate) indicators on productive inputs or performance. This emphasis on the content of the activity baskets of countries or regions resonates with other approaches in economics that understand economic growth through the lens of sectoral allocation of productive factors, such as the structuralist literature (Prebisch, 1950; Lewis, 1954; Hirschman, 1958; Lin et al., 2011), and that identify in the role of capabilities the main factor driving innovation, embodied in the evolutionary economics literature (Cimoli and Dosi, 1995; Dosi and Nelson, 1994; Nelson and Winter, 1982; Teece and Pisano, 1994).

The quality and diversity of an economy’s productive and technological portfolio have broader implications than simply economic growth. For instance, the composition of economic and technological specialization has strong implications for the environment, as the footprint of different products and technologies can differ substantially. By the same token, the accumulation of technological capabilities may

put countries on trajectories that can mitigate or exacerbate climate change. In this respect, EC methods can prove particularly useful to understand and guide the sustainability transition.

The transition towards a more sustainable and low-carbon socioeconomic system is a top policy priority, with the EU Green Deal aiming at climate neutrality by 2050,¹ the Inflation Reduction Act in the United States (Bistline et al., 2023) and the ambitious renewable energy targets of China's 14th Five-Year Plan For Renewable Energy With Quantitative Targets For 2025 (Li et al., 2019; Yang and Wang, 2023). However, decarbonizing the economy by phasing out high-emission and energy-intensive industries will require radical transformations and profound structural change at the core of socioeconomic systems. Moreover, this will have to account for the heterogeneous capacity of geographical areas and industries to achieve climate neutrality along with the possible long-lasting effects on income, spatial, and environmental inequalities.

To inform policy on how to address such a complex transformation, wherein geographical, structural and institutional elements interact, in the last few years an approach has emerged in the literature (see among others Barbieri et al., 2020, 2022; Montresor and Quatraro, 2020; Santoalha and Boschma, 2021) that draws from sustainability studies, evolutionary economic geography (Boschma and Frenken, 2006, 2018) and EC. Embracing a complexity perspective may in fact be more effective than traditional approaches in accounting for the interconnected nature of this process of change (Common and Stagl, 2005) providing policy-relevant, data-driven and granular evidence to embrace socioeconomic complexity at different geographical scales.

While environmental goods and technologies have been mainly studied as homogeneous, aggregate quantities in the extant literature, they actually display high degrees of heterogeneity in terms of functions, underlying know-how, life cycle stages, and links to pre-existing specialisation patterns. Therefore, the capacity of EC methodologies to keep away from aggregation and to provide feasibility diagnostics at the level of single products or technologies may prove extremely relevant in analyzing the potential directions of green development for each country or region,

Coupling the geographical distribution of productive and technological

¹ <https://eur-lex.europa.eu/legal-content/EN/TXT/?qid=1596443911913&uri=CELEX%3A52019DC0640#document2>

capabilities with environmental and socioeconomic variables, various scholars have applied the EC toolbox to try to answer three main broad questions: (i) What is the relationship between complexity and sustainability? (ii) What are the properties of green products and technologies, and are these inherently different from non-green ones? (iii) How can we assess the readiness of national/regional knowledge bases and productive structures for the green transition?

Contributions addressing the first question investigate the empirical relationship between countries' environmental outputs, such as Greenhouse Gases (GHG) emissions, and the complexity of their productive structures, assessed by focusing on the export dimension through various EC metrics. These studies are far from unanimous, as they vary in their samples or estimating methods. This yields contrasting results on the effect of EC on the environment, ranging from linear positive to non-linear relationships *à la* Kuznets curve (1955).

The second and third questions have roots in the evolutionary economic geography literature and are more varied in terms of techniques and scales of observation. They mainly examine the geography of green competitiveness of national/regional knowledge bases or productive structures by analyzing specialization profiles in green products or technologies, different dimensions of relatedness between green and non-green technological or productive activities, and their relationship with a series of socioeconomic indicators – e.g., inequality or digital literacy. As is common in the EC literature, these works especially stress the potential complementarity between existing capabilities and future specialization in low-carbon technologies or products.

The EC approach emphasises the role played by productive and technological capabilities in promoting advancement across domains of human activity. From the point of view of green goods and innovative activities, this implies that as countries and regions strive to move towards sustainable practices, EC offers a framework to analyze the composition and diversification of their economies, and to identify the dynamic set of capabilities that are most conducive to green productive or technological specialization, allowing countries to progress along the sustainability ladder. As we show in this review, EC provides a valuable lens through which policymakers and researchers can understand and navigate the intricacies of the sustainability transition.

This literature review aims to summarize the empirical evidence produced so far to answer the three aforementioned questions from an EC perspective. It should

be noted that other recent contributions have tackled the link connecting EC and socioeconomic sustainability from several angles. Notably, the recent review by Montiel-Hernández et al. (2024) focuses on the ways in which EC has proven to be useful to relevant socio-economic topics such as economic growth, innovation, employment, education levels, poverty reduction, life expectancy, environmental impacts, public policies. The paper identifies five areas of interest for EC, namely investigating the causal pathways linking EC to sustainability dimensions; developing new indicators of EC; exploring how EC can be used to promote sustainable development policies as well as international research cooperation on this topic; and integrating EC into tertiary education programs. At the same time, several contributions have ventured into a detailed analysis of how EC indicators correlate with empirical measures of socioeconomic sustainability – some of which are reviewed by Hidalgo (2021), and more extensively in Section 3.1 of this paper. What makes this review complementary to the aforementioned contributions is that it focuses explicitly on the methodologies and data that have been used to link EC to environmental sustainability only, in an attempt to answer the macro-questions listed above. Our effort is particularly timely given the rapid expansion of this literature, and it provides scholars with a clear roadmap on how to ensure comparability between new empirical evidence and the existing body of knowledge.

The remainder of this paper is organized as follows. We first review the data on products and technologies used to apply EC methods to the study of the sustainability transition, highlighting drawbacks and advantages entailed by different types of data (Section 2.1). Second, we offer a methodological contribution by attempting to unify the methods used to estimate and analyze measures of EC and relatedness (Section 2.2). Third, we canvas the literature that links complexity measures to the concept of environmental sustainability, in order to summarize the debate on the role of complexity in explaining environmental issues and in identifying viable avenues into the sustainability transition (Section 3). Finally, we attempt to identify limitations of the existing literature and methods, proposing further research avenues in the field (Section 4).

2. Data and Methods

The data employed in EC-based analysis of the green economy are drawn from two main sources: patents and trade flows. This section presents in detail the most

commonly used datasets and classifications of green activities, firstly focusing on patent data and secondly on trade data. This data is used in EC following a common empirical framework: a bipartite matrix is constructed by cross-tabulating products (technologies) and geographical areas (countries, regions, cities...). The bipartite matrix is then filtered, projected, ordered, and processed in different ways, depending on the specific purposes of the analysis (see Section 2.2).

2.1. Data

2.1.1. Patent data The main data sources for green or non-green patenting activity at the national or local level are the European Patent Office’s (EPO) PATSTAT and the OECD’s REGPAT. PATSTAT is a comprehensive database covering patent applications filed in more than 70 national and international patent offices, including the most important ones – e.g. the EU, US, and Japan patent offices. REGPAT is instead published annually by the OECD and covers the subset of PATSTAT patent applications filed at the EPO since its inception in 1978. While PATSTAT covers a wider set of patent documents and a longer time-span, REGPAT covers EPO’s applications, that are high-quality and extremely reliable – we refer to the paper’s Appendix 4 for a more detailed comparison between these two data sources.

The technological fields associated to the patents recorded in PATSTAT and REGPAT follow two classifications: the International Patent Classification (IPC) and the Cooperative Patent Classification (CPC). Both classifications follow a similar hierarchical structure, spanning from codes with a very detailed description (e.g., G02B 1/02: optical elements [...] made of crystals, e.g. rock-salt) to codes aggregating many detailed technologies under a broader common technological area (e.g., G02: optics; G: physics). For example, at 4-digits, the classifications have around 600 unique codes, while at 8-digits there are approximately 7000 codes in the IPC and 10000 in the CPC. Despite the strong similarities between the two classifications, only the CPC has a section dedicated to Climate Change Mitigation and Adaptation Technologies (CCMTs), the ‘Y02/Y04S’ tagging scheme, nested under section Y. Table 1 reports the 1-digit codes and the titles of the sections comprising the IPC and CPC classifications.

The contributions to eco-innovation presented in this review rely mainly on the Y02/Y04S tagging scheme of the CPC and the OECD ENV-TECH classification, based on a mixture of IPC and CPC codes, which cover a wide range of a wide range

Table 1: Sections of the IPC and CPC classifications.

Code	Title	IPC	CPC
A	Human necessities	Y	Y
B	Performing operations; Transforming	Y	Y
C	Chemistry; Metallurgy	Y	Y
D	Textiles; Paper	Y	Y
E	Fixed constructions	Y	Y
F	Mechanical engineering and weapons	Y	Y
G	Physics	Y	Y
H	Electricity	Y	Y
Y	New technologies, including CCMT[†] & Smart Grids	N	Y

[†]Climate Change Mitigation and Adaptation Technology.

technologies related to sustainability objectives. These include energy efficiency in buildings, energy generation from renewable sources, sustainable mobility, and smart grids.

In response to the increasing attention and concerns about climate change mitigation and renewable energy generation, there has been a large increase in the number and scope of patent applications in environment-related domains in the recent past (European Patent Office, 2013). However, searching for environment-related patent documents was not straightforward at the beginning because a dedicated classification system for sustainable technologies was not available. In fact, before 2011, no specific branch of the IPC or of other technology classification covered environment-related inventions. A permanent solution to this issue became available in 2013 following the publication of the Cooperative Patent Classification, which was jointly developed by the the EPO and the United States Patent and Trademark Office (USPTO), in an attempt to harmonize their patent classification procedures. Since then, the CPC has become increasingly popular as a classification standard, complementing or substituting the IPC in a growing number of patent offices worldwide.

As illustrated in Table 1, two types of codes can be found in the CPC classification: codes starting with the letters A to H, which are similar to IPC codes and represent the traditional classification of technologies; and codes starting with Y, which are used to tag cross-sectional technologies already indexed somewhere else

Table 2: CPC Y02/Y04S (2018) tagging scheme (European Patent Office, 2018).

Class or Subclass	Title and description
Y02	TECHNOLOGIES OR APPLICATIONS FOR MITIGATION OR ADAPTATION AGAINST CLIMATE CHANGE
Y02A	Technologies for adaptation to climate change
Y02B	Climate change mitigation technologies related to buildings, e.g. housing, house appliances or related end-user applications, including the residential sector
Y02C	Capture, storage, sequestration or disposal of greenhouse gases
Y02D	Climate change mitigation technologies in information and communication technologies, i.e. information and communication technologies aiming at the reduction of their own energy use
Y02E	Reduction of Greenhouse Gases (GHG) emissions, related to energy generation, transmission or distribution, including renewable energy, efficient combustion, biofuels, efficient transmission and distribution, energy storage, and hydrogen technology
Y02P	Climate change mitigation technologies in the production or processing of goods
Y02T	Climate change mitigation technologies related to transportation, e.g. hybrid vehicles
Y02W	Climate change mitigation technologies related to wastewater treatment or waste management
Y04	INFORMATION OR COMMUNICATION TECHNOLOGIES HAVING AN IMPACT ON OTHER TECHNOLOGY AREAS
Y04S	Systems integrating technologies related to power network operation, communication or information technologies for improving the electrical power generation, transmission, distribution, management or usage, i.e. smart grid technologies including hybrid vehicles interoperability.

in the classification.² In addition to the Y02 class, the new subclass Y04S dedicated to smart grids was integrated into the CPC section Y. As shown in Table 2, the Y02 class consists of more than 1000 tags related to sustainable technologies organized in 9 sub-classes.

The enduring popularity of the IPC classification and its predominant coverage of patents filed before the publication of the CPC prompted efforts to maximize the informative content on eco-innovation present both in the IPC and CPC classifications over a longer time period. In this context, in 2015 the OECD (Haščič and Migotto, 2015) developed ENV-TECH, an expert-based catalogue of environment-related technologies based on the IPC classification, last updated

² <https://www.uspto.gov/web/patents/classification/cpc/html/cpc-Y.html>

in 2016 (OECD, 2016), which can be used to tag green patent documents in PATSTAT or other patent datasets. ENV-TECH identifies 94 environment-related technology areas that group 4- to 16-digit IPC and CPC codes, building on the CPC Y02 class whenever possible. The catalogue relies on a keyword search strategy identifying patent documents that correspond to each 'target environmental technology field' (Haščič and Migotto, 2015, p.19), covering CCMTs but also environmental management and water-related adaptation technologies (class 1 and 2). When it is not possible to identify single IPC/CPC classes that define alone the technological field of interest, it employs a combination of different IPC/CPC patent classes, also at different aggregation levels.

Finally, in addition to the CPC and ENV-TECH classifications, the World Intellectual Property Organization's IPC Committee of Experts proposed the *IPC Green Inventory* (IPC-GI).³ IPC-GI streamlines patent searches for Environmentally Sound Technologies, as identified by the United Nations Framework Convention on Climate Change, with codes spread across various IPC technical fields and the purpose of consolidating them into a single, easily accessible repository. However, given its limited coverage, the IPC calls for caution in its use for research.

2.1.2. Trade data Since *The environmental goods and services industry manual for data collection and analysis* was published by the OECD (1999), a wide array of lists and taxonomies of green products has been proposed. Here we briefly discuss the lists that are most commonly used in EC analyses, highlighting the main critical issues in classifying environmental goods. Readers should keep in mind that the EC literature has focused, especially in earlier works, on establishing the level of complexity and growth potential of an economy by extracting information from its trade specialization profile. In order to do so, a harmonized global classification of products with a uniform, global interpretation has been crucial to develop the framework. The Harmonized System (HS) classification satisfies these requirements and has been the most widely used in the field. HS is a standardized numerical method of classifying traded products used by customs authorities to identify products; it is maintained by the World Customs Organization (WCO) and it is updated every five years. HS comprises more than 5,000 commodity subheadings,

³ <https://www.wipo.int/classifications/ipc/green-inventory/home>, accessed February 2024.

which are identified by a 6-digit code and arranged according to a nested structure going up to 96 2-digit Chapters and 21 1-digit Sections.

The first issue to take into account when constructing a comprehensive classification of green products is in the very nature of the HS classification system, which was not conceptualized to accommodate a green/non-green dichotomy or information on the goods' efficiency. The introduction in the HS of several 6-digit subheadings which group together new environmental goods was announced (Steenblik, 2020) in 2020; the WCO is slowly addressing this issue by updating the HS system with more codes for green products. However, the updated classification is not yet available. Therefore, a clear-cut identification of environmental goods within existing product classifications may at times prove to be a difficult task. For instance, up until recently, it was impossible to distinguish between combustion engines and electric vehicles. Other classifications or surveys exist, although they focus on specific environmental aspects or on final use, and do not present the standardization and granularity properties required by EC techniques. A second crucial requirement for a comprehensive and effective green goods classification is a unanimous definition of what GREEN PRODUCTS are. This clearly depends on specific needs: for instance, different requirements need to be satisfied to fulfil regulatory and research goals, or to define (tax) incentives towards more sustainable practices and energy sources.

While stressing a series of open questions, partially still unanswered, a 1999 OECD report (OECD, 1999) defines the environmental goods and services industry as the set of activities aimed at measuring, preventing and mitigating environmental damages related to eco-systems, water, air, soil, and noise pollution, or waste management. This includes cleaner technologies, products and services that reduce environmental risk and minimize pollution and resource use. In the report, a list of 121 environmental goods satisfying this definition was proposed. More recently, according to the International Monetary Fund (IMF)⁴ environmental goods are connected to environmental protection or goods that have been adapted to be more environmentally friendly. To achieve an encompassing classification for policy implications, evaluation, and tariff regularization, in 2009 the WTO published a

⁴ The definition of environmental goods of the IMF Climate Change Indicator Dashboard, 'Trade in Environmental Goods' data-set can be found at: https://climatedata.imf.org/datasets/8636ce866c8a404b8d9baeaffa2c6cb3_0/explore, where international trade flows in environmental goods – defined according to the OECD 1999's list (OECD, 1999) as updated by the IMF – are reported.

broader list of green products (WTO, 2009) comprising 480 products. While the list is comprehensive, many inaccuracies and biases have been pointed out by a EUROSTAT report published in the same year (EUROSTAT, 2009). The WTO also published two shorter *friend* and *core* lists composed respectively of 154 and 26 products. Similar efforts were made by the OECD for the 2010 Toronto G20 summit which proposed an updated list of 150 Plurilateral Environmental Goods and Services (PEGS). Moreover, in 2012 the Asian-Pacific Economic Cooperation (APEC) (APEC, 2012) put forward a list of 54 environmental goods subject to reduced tariffs. Finally, by combining and revising the WTO friend, APEC, and PEGS lists, in 2014 the OECD published the most recent list of green products, the Combined List of Environmental Goods (CLEG) (Sauvage, 2014) comprising 248 products relevant to tackling climate change.

In addition to the fact that any list of environmental products cannot be considered final as the HS classification needs constant updates and revision, regardless of the specific definition, two main shortcomings are shared by all green product classifications:

- Final use: it is not possible to ascertain the actual final use of a large number of products. Many commodities labelled as green (e.g., filters, pumps, and pipes) may be used also for non-environmental purposes. While statistics can be computed based on surveys, their reliability on a global scale is yet unclear.
- Greener products: depending on the goal of the study/application, one could consider products that are less raw material-intensive, more energy efficient, easier to dispose of, or have longer life spans. However, these properties crucially rely on the comparison with other products belonging to the same category, comparisons that are, even in principle, very difficult to carry out.

Therefore, a clear trade-off between accuracy and comprehensiveness arises, and, as we will see below, the selected contributions using green products have proposed different identification strategies to, albeit not fully, overcome these issues.

2.2. Methods

The basic intuition of the EC approach is that specific activities, such as exported products, industries, patents, or scientific research are important because they constitute different learning opportunities and development possibilities. EC

explicitly builds on the heterogeneity and interactions between different economic actors, assuming that knowledge grows not by accumulating ‘more’ of the same, but by adding new and different elements to existing capacities. In this framework, economic performance is seen as the result of the accumulation of different types of non-tradeable inputs and capabilities, that are not empirically observable but can be inferred by studying the structure of the specialization profiles of countries. Therefore, EC can be seen as an indirect measure of the capability endowment of a country.

The concept of capabilities was initially devised at the firm level in the evolutionary economics literature to denote the dynamic knowledge necessary for firms to innovate and introduce new products and services (Penrose, 1959; Teece et al., 1997, 1994). In the EC literature, productive capabilities encompass location-specific attributes that extend to intangible elements crucial for fostering and leveraging productive efficiency. These include institutional frameworks, educational systems, policies, and infrastructures necessary for countries to specialize in more complex activities (Hausmann and Rodrik, 2003; Hausmann et al., 2005). Information on capabilities is derived from a binary network linking geographical areas to the activities in which they exhibit a comparative advantage.

In the following, we present the main metrics and tools used in the EC literature analyzing different dimensions of sustainability.

2.2.1. Complexity measures EC analyses often start from the observation of empirical bipartite networks connecting geographical areas – be they countries, regions, or cities – to different types of economically relevant activities – such as patenting (Balland and Rigby, 2017; Pugliese et al., 2019a; Sbardella et al., 2018a), manufacturing of products (Hidalgo and Hausmann, 2009; Tacchella et al., 2012), or scientific research (Patelli et al., 2023) – in which they specialize. These bipartite networks are typically assessed by evaluating the comparative advantage of the geographical area in the selected activity, using Balassa’s Revealed Comparative Advantage (RCA) index (Balassa, 1965). RCA measures the share of an activity in a geographical area with respect to a reference distribution, often the activity’s global share, and is thus interpreted as a proxy for above-average competitiveness in that activity. The index (and its refinements – see Bruno et al., 2023) remains the workhorse for binarizing the adjacency matrices of bipartite networks in the EC literature. In formula, the RCA of geographic area g in activity a of volume W , can

be written as follows:

$$[RCA]_{ga} = \frac{W_{ga}(\sum_{g'a'} W_{g'a'})}{(\sum_{a'} W_{ga'})(\sum_{g'} W_{g'a})} = \frac{W_{ga}}{\sum_{a'} W_{ga'}} \bigg/ \frac{\sum_{g'} W_{g'a}}{\sum_{g'a'} W_{g'a'}} \quad (1)$$

RCA assigns a real and positive value to each combination of geographical areas and activities. $[RCA]_{ga} = 1$ if the weight W of a in the basket of activities of geographical area g is the same as the global weight of activity a relative to all activities. The resulting distribution of RCA values is skewed with large tails (Patelli et al., 2022). High-variability continuous values can be impractical in most applications, especially when the only information required is on the presence/absence of specialization of geographical area g into activity a . Therefore, the RCA matrix is usually binarized by setting a threshold value of 1.⁵ A matrix entry equal to 1 is inserted in the binarized matrix cell when $[RCA]_{ga} \geq 1$ – i.e., g has a comparative advantage in a –, while an entry equal to zero is inserted otherwise:

$$M_{ga} = \begin{cases} 1 & \text{if } [RCA]_{ga} \geq 1 \\ 0 & \text{otherwise.} \end{cases} \quad (2)$$

The binary RCA network displays strong empirical evidence of diversification and nestedness patterns in all the activities studied in EC: many geographical areas are competitive in large set of activities, in contrast to Ricardian specialisation patterns. However, there is more to the binary M matrices than just the diversification of the geographical areas represented by their rows. M matrices are in fact usually nested across domains of activity, irrespective of whether the columns of M represent products, technologies, or scientific fields (Patelli et al., 2022) and scales of analysis (Laudati et al., 2023; Pugliese et al., 2019b; Sbardella et al., 2017). The concept of nestedness originates from ecology (Atmar and Patters, 1993) and refers to the observation that less diversified geographical areas (species) typically have comparative advantages in the activities (ecological niches) pursued by more diversified geographical areas. This pattern can be made explicit when the presence/absence M matrix is shown with the proper ordering of rows and columns, resulting in the emergence of a hierarchical model represented by triangular-like matrices (Cristelli et al., 2015).

Based on the information contained in the binary matrix M , it is possible to define synthetic EC indicators capturing relevant information on the network

⁵ Even though the intuitive unit threshold is standard in the EC literature, different values may be appropriate depending on the specific application.

structure, and thus on the geographical areas' specialization profiles. The rationale of EC algorithms is that the complexity of the economies under analysis and the complexity of the activities in which they are specialized can be determined recursively by taking advantage of the information provided by the composition of their activity portfolio. The first indicator of Economic Complexity was proposed by Hidalgo and Hausmann (2009), the 'Method of Reflections' (MR) aimed at capturing the diversification of a country's export profile. At the n -th iteration, for a geographical area g and activity a , MR proposes an iterative procedure to capture the diversification of geographical area g , k_g^n , and of activity a , k_a^n :

$$\begin{aligned} k_g^{(n)} &= \frac{1}{k_g^0} \sum_a M_{ga} k_a^{(n-1)} \\ k_a^{(n)} &= \frac{1}{k_a^0} \sum_g M_{ga} k_g^{(n-1)} \end{aligned} \tag{3}$$

with initial conditions $k_g^0 = \sum_a M_{ga}$ and $k_a^0 = \sum_g M_{ga}$, the diversification of g and the ubiquity of a . The MR rationale is that the complexity of a geographic area, interpreted as its generalized diversification, is driven by the average complexity of the activities it specializes in. By the same token, the average complexity of each activity is driven by the average complexity of the areas displaying a comparative advantage in it. Since the iterative model in Equation 3 has a trivial solution where each component is 1 (Caldarelli et al., 2012; Kemp-Benedict, 2014), the same authors later proposed a more practical definition of the indices based on the eigenvector associated with the second largest eigenvalue of a matrix derived from the MR (Mealy et al., 2019): the Economic Complexity Index (ECI) of countries, and the Product Complexity Index (PCI), in the case of products (Hausmann et al., 2011, 2014). Both indices converge to those evaluated using MR only after considering the standardized version (removing the mean and dividing by the standard deviation), although there is a certain carelessness in the economic complexity literature that refers to the ECI/PCI only as the second-largest eigenvalue. Their generalization to the case of other activities, such as patents, is straightforward: the same mathematical procedure is applied to the binary matrix based on the data on the activity of interest.

However, the ECI-PCI approach presents a non-negligible conceptual drawback due to the fact that ECI and PCI are defined by averages (Pietronero et al., 2017) and not by sums, hence smoothing out the information on the cumulative nature of capabilities encapsulated in the geographical area-activity networks, a pillar of

the economic reasoning behind the EC approach (Dosi and Nelson, 1994; Teece and Pisano, 1994). This means that, for example, if we consider two fictitious countries, with the first specialized in all the activities present in the network, and the second holding a comparative advantage only in one activity of average complexity, they will paradoxically display the same ECI value. Interestingly, Mealy et al. (2018) define a Green Complexity Index (GCI) simply as the sum of the complexity of green products (while product complexity is still calculated on the entire set of products using ECI), with the final sum accounting for country diversification in green products, partially overcoming this crucial aspect of ECI.

While ECI is based on the idea that countries are diversified, the observation of nested patterns uncovers complex and interdependent systems that cannot be easily described through linear models and averages. A more suitable metric for the characterization of nested patterns in the M matrix is the Economic Fitness and Complexity (EFC) algorithm (Tacchella et al., 2012): the Fitness of a geographical area g is defined as the sum of the complexity of all the activities in which the area is specialized (i.e., if $RCA_{ga} \geq 1$). Therefore, Fitness increases as the activity basket of g become larger: a geographical area with a more advanced set of capabilities will have a more diversified portfolio of activities, spanning from the most to the least complex ones, and will therefore have a higher Fitness score.⁶ The extensive nature of Fitness is complemented by the definition of the Complexity of activity a , which is driven mainly by the lowest Fitness areas holding a comparative advantage in a , since complex activities are rare and appear almost exclusively in the portfolio of high-Fitness areas. Consequently, an area with low Fitness has a smaller endowment of capabilities and thus operates exclusively in less complex domains. Operationally, the EFC iterative algorithm is defined as the fixed point of the following coupled

⁶ It has been shown that Fitness and Complexity can be interpreted as *potential* functions, related to the bipartite network M_{ga} , which defines a forbidden region of the matrix given by the requirement of efficiency in the allocation of resources of countries and products (Mazzilli et al., 2024).

equations, that are suitably normalized at each iterative step:⁷

$$\left\{ \begin{array}{ll} \tilde{F}_g^{(n)} = \sum_a M_{ga} Q_a^{(n-1)} & F_g^{(n)} = \frac{\tilde{F}_g^{(n)}}{\langle \tilde{F}_g^{(n)} \rangle_g} \\ \tilde{Q}_a^{(n)} = \frac{1}{\sum_g M_{ga} \frac{1}{F_g^{(n)}}} & Q_a^{(n)} = \frac{\tilde{Q}_a^{(n)}}{\langle \tilde{Q}_a^{(n)} \rangle_a} \end{array} \right. \quad (4)$$

It is noteworthy to mention that both Fitness and ECI are relative measures, based on the structure of the geographical area-activity matrix M at a given time of observation of the economic system under analysis. Therefore, their magnitudes cannot be compared longitudinally across time as, at each given time point, the underlying bipartite networks change. This issue is not often discussed in the literature, but it should be a pressing concern when these metrics are included in longitudinal regression settings (many such examples are discussed in Section 3.1). It is however possible to overcome this issue by using ECI/Fitness rankings, albeit losing information on the EC index's magnitude, or more soundly by setting a reference scale (Mazzilli et al., 2024), i.e. by assigning to a reference country a fixed value, or by adding a dummy country that is always specialized in all activities, therefore imposing an invariant of scale across time periods.

Another frequent issue that appears in the EC literature is the need to analyze systems across different geographical scales. Often, at finer geographical scales (i) local data-set do not have global coverage; (ii) data quality is lower since volumes are smaller and errors become more relevant. While issue (ii) is straightforward, issue (i) is often overlooked, although it may pose problems in evaluating a reference distribution for computing the RCA (e.g., it is very different to assess the complexity of products when looking at global trade flows or only at European countries or regions). To overcome these issues, a measure of *exogenous fitness* has been proposed (Operti et al., 2018; Sbardella et al., 2018a). To compute the Fitness of subnational areas, this approach computes the Complexity of economic activities at the country level, and subsequently evaluates the Fitness of subnational areas by plugging the country-level complexity in the computation of the Fitness equation (first term of Equation 4). In such a way, (i) the signal-to-noise ratio in the data is higher; (ii) it is possible to provide a more realistic assessment of complexity

⁷ The normalization step is unnecessary for the evaluation of the fixed point (Mazzilli et al., 2024), however, it is helpful for the numerical procedures and the stability of the code.

evaluating competition dynamics at the global level. In principle, the same rationale may be implemented in the ECI definition (Equation 3). In particular, Green Technology Fitness, introduced by Sbardella et al. (2018a) and used in other works reviewed in the present paper, is a measure of Fitness computed on climate change mitigation technologies (see Section 2.1), which, when computed at the regional level, employs the exogenous fitness rationale.

2.2.2. Relatedness Another key set of techniques from the EC toolbox aims at measuring relatedness⁸, a measure of the pairwise similarity between activities – be they products (Hidalgo, 2009; Zaccaria et al., 2014)⁹, technologies (Breschi et al., 2003; Napolitano et al., 2018), or both technologies and products (Pugliese et al., 2019a; de Cunzio et al., 2022). The measure of relatedness between two activities is based on the observation of their empirical co-occurrences in the specialisation profiles of different geographical areas, and is connected to the probability that having a comparative advantage in the first activity will also lead to a comparative advantage in the second. This is done relying on the assumption that countries that can successfully specialise in an activity have developed a set of capabilities that will enable them to diversify into related activities. Intuitively, activities that share similar inputs will be situated close to each other in the network, and proximity in the network should indicate a relatively high probability of jumping from an activity a to a neighboring one, a' .

Relatedness is usually represented as a network of connected activities, where links between activities are based on their degree of similarity. There are several methods to evaluate relatedness, resulting in different and heterogeneous networks that cannot always be easily compared. However, most of these networks present a core-periphery structure, with hubs and leaves associated with similar types of activities or arranged into meaningful communities. The first approach to relatedness in the EC literature is the Product Space, proposed by Hidalgo et al. (2007), initially based on the country-product international trade network, and later employed also for different set of activities. The similarity between activity a and activity

⁸ These analytical tools can be traced back to Jaffe (1986) and to the measure of corporate coherence introduced by Teece and Pisano (1994).

⁹ See also Tacchella et al. (2023) for a multi-product, non-linear approach based on machine learning.

a' is evaluated using *proximity* (Hidalgo, 2009), a measure of pairwise normalized geographical co-occurrence, i.e. summing over the geographical dimension g :

$$\phi_{aa'} = \min \left\{ \frac{\sum_g M_{ga} M_{g,a'}}{\sum_c M_{ga}}, \frac{\sum_c M_{ga} M_{ga'}}{\sum_c M_{ga'}} \right\}. \quad (5)$$

Originally, proximity was interpreted as the conditional probability of being competitive in activity a given competitiveness in activity a' (see p. 3 of Hidalgo et al., 2007 SI).¹⁰

The EC literature has later proposed a more statistically sound approach to relatedness that allows to evaluate time-lagged connections between pairs of activities, also between two layers of different types of activities such as products and technologies, or technologies and scientific fields (de Cunzio et al., 2022; Pugliese et al., 2019a; Sbardella et al., 2022; Zaccaria et al., 2014), allowing to model time-lagged spillovers across different dimensions of capability. Initially focused on international trade, Zaccaria et al. (2014) put forward a refined methodology to assess relatedness, the Product Progression Network. This approach has three main advantages: i) it is dynamic, as it takes explicitly into account time by comparing the pairwise activity co-occurrences within the same geographical area at different points in time, instead of contemporaneous co-occurrences; ii) it filters each link of the resulting network using a null model so that only the statistically significant links are stored: a crucial advancement, because the presence of a link may be simply due to the ubiquity of a product or to the diversification of a country (Cimini et al., 2019; Saracco et al., 2015, 2017); iii) instead of focusing on pairwise correlation between two activities of the same type (e.g. two products), it has been employed to study also the co-occurrence among different types of activities, such as links between products and technologies (de Cunzio et al., 2022; Pugliese et al., 2019a).

The mathematical formulation of this approach relies on the so-called *assist matrix* B , which estimates the probability of having a comparative advantage in activity a in geographical area g in year y_2 , conditional on having a comparative advantage in activity a' in the same area in a previous year $y_1 = y_2 - \Delta t, \Delta t \geq 0$

¹⁰ Please notice that even though each value of $\phi_{pp'}$ is real, positive, and bounded to 1, proximity is not a conditional probability. In fact, a conditional probability $P(A|B)$ requires that $\sum_A P(A|B) = 1$, while $\sum_{p'} \phi_{pp'}$ depends on the dimension of the network (the linear sizes of M_{cp}) and is typically larger than 1.

(typically a lag of $\Delta t = 5$ years is considered) and is defined as follows:

$$B_{aa'}^{\alpha\beta}(y_1, y_2) = \sum_g \frac{M_{ga}^{\alpha}(y_1)}{u_a^{\alpha}(y_1)} \frac{M_{ga'}^{\beta}(y_2)}{d_g^{\beta}(y_2)}, \quad u_a^{\alpha} = \sum_g M_{ga}^{\alpha}, \quad d_g^{\alpha} = \sum_a M_{ga}^{\alpha} \quad (6)$$

where α and β indicate that the matrices may contain information on different activity types (e.g., α being patents, and β products). In this case $B_{aa'}$ is defined as a conditional probability, since $\sum_{a'} B_{aa'} = 1$, $0 \leq B_{aa'} \leq 1$.

However, the assist matrix, like any other relatedness measure, is based on nested networks, where diversified actors are active in most of the activities, and as a consequence, a co-occurrence may not be informative *per se*. Hence, a null-model should be employed to discount the fact that co-occurrences can be random, and are more likely for more ubiquitous activities or more diversified geographical areas (Cimini et al., 2019). The most common null model used in the literature to assess the statistical significance of the conditional probabilities $B_{aa'}(y_1, y_2)$, is the Bipartite Configuration Model (BiCM) (Saracco et al., 2015, 2017), a maximum-entropy approach for the randomization of bipartite networks – we refer to (Cimini et al., 2019) for a more detailed discussion on null models for complex networks.

3. Review of the literature

In this section, we propose a selection of papers analyzing different dimensions of the sustainability transition using EC techniques. This exercise is focused not only on highlighting the main findings but especially on illustrating the variety of methods employed, which lead often to inconclusive results due to their low comparability. For this reason, we conduct a purposive and ad hoc literature review, with the aim of illustrating the empirical applications of the methods discussed in the previous section. Systematically reviewing all the evidence on the use of EC metrics and methods to study the sustainability transition falls beyond the scope of this article; instead, we strive to offer a representative account of the empirical EC applications in the field. To filter from the corpus of academic literature potential candidates for inclusion in this review, we first defined a list of relevant keywords (see Table C1 for details) and used them both individually and in appropriate combinations to query specialized scholarly databases and search engines. We performed this extensive search for the first time in November 2022 to define a shortlist of papers and then pruned the results based on a preliminary reading, in which we assessed contributions

based on their relevance, novelty and representativeness of the areas we set out to cover. We repeated the searching and pruning again in early 2024 to account for newly available relevant publications.

The rationale for the subdivision of the literature in three blocks follows rather naturally from the shared characteristics of the empirical applications of EC tools to study environmental sustainability. We divide the corpus of empirical literature based on the unit of analysis of the empirical study, namely i) geography (national or sub-national), ii) products and iii) technologies. A graphical representation of this representative corpus is summarized by Fig. 1, which identifies a number of features shared across the empirical contributions, namely: unit of observation (group), data, methods, definition of sustainability, and sample adopted. The following subsections reflect the three groups of contributions indicated by the colour coding of Fig. 1.

The first subsection includes contributions that have explored the association between complexity metrics at the geographical level (mainly countries and regions) and variables related to environmental sustainability, such as aggregate CO₂ and GHG emissions (papers highlighted in yellow in Figure 1). As can be inferred from Fig. 1, large share of this literature uses the Economic Complexity Index (computed using Eq. 3) as explanatory variable, with a few exceptions that use instead the Economic Fitness and Complexity (Eq. 4). The association between EC and environmental sustainability is analyzed in regression settings, which sometimes resort to panel dynamic techniques. Occasionally, causal claims are made by the authors. The sample of countries considered is rather heterogeneous, which hampers the rationalization of the findings.

The second subsection (blue rows in Figure 1) focuses on the identification of green products, aiming to measure the green potential for green diversification based on export patterns. As shown in 1, the main tool used in this branch of studies is the product space. Their goal is to analyse the current status of countries' productive system from the point of view of green product. Most of these works do not try to asses relationship with target variables such as CO₂ emission or other measures of sustainability. Mainly following Mealy and Teytelboym (2022), they use distances and proximity in the products space to construct a composite measure of 'greenness' of productive systems. Relying on export data, the geographical coverage is usually extended to most of the world, with few focuses on specific countries or regions.

The third subsection (highlighted in green in Figure 1) is constituted by empirical articles that use patent data to assess the readiness of regions for the

Author	Year	#	Data				Method									Definition of Sustainability								Sample												
			Exported products	Output	Patents	Publications	Occupations	Product space	Relatedness	Null model	Multilayer network	Method of reflections	Fitness	Kernel regression	Time series regression	Panel regression	Dynamic panel regression	GMM	Causal inference	CO2 emissions	GHG emissions	Pollutant concentration	Fossil fuel consumption	Socio-economic	Composite measure	Climate change mitigation and adaptation	Green to non-green technological relatedness	Green tech to exported goods relatedness	Cross country (all)	Cross country (high income)	Cross-country (emerging)	One country (high income)	One country (low-middle income)	Regions (high income)	Regions (low-middle income)	
Can and Gozgor	2017	A1	x							x		x	x					x													x					
Neagu	2019	A2	x							x					x			x										x								
Dong et al.	2020	A3		x				x		x				x							x															
Swart and Brinkman	2020	A4	x							x				x									x												x	
Ylanci and Pata	2020	A5	x							x		x	x											x								x			x	
Boleti et al.	2021	A6	x							x	x							x		x	x			x												
Dogan et al.	2021	A7	x							x				x	x			x		x	x			x				x		x						
Dordmond et al.	2021	A8	x				x	x		x					x	x			x					x					x						x	
Martins et al.	2021	A9	x							x					x			x		x									x							
Pata	2021	A10	x							x		x	x					x						x							x					
Romero and Gramkow	2021	A11	x					x						x		x			x					x				x								
Zheng et al.	2021	A12	x							x								x											x							
Akadiri et al.	2022	A13	x							x								x						x									x			
Rafique et al.	2022	A14	x							x						x	x							x						x						
You et al.	2022	A15	x							x	x			x					x										x							
Esmaeili et al.	2023	A16	x							x									x												x					
Stojkoski et al	2023	A17	x		x	x				x	x								x										x							
Fankhauser et al.	2013	B1	x																					x					x							
Hamwey et al.	2013	B2	x					x																x									x			
Fraccascia et al.	2018	B3	x					x	x					x	x									x					x							
Pérez-Hernández et al.	2021	B4	x					x						x										x												x
Mealy and Teytelboym	2022	B5	x					x						x	x				x										x							
Muller et al.	2023	B6	x					x	x															x					x							x
Sbardella et al.	2018	C1			x							x	x												x		x	x								
Montresor et al.	2019	C2			x				x						x									x		x	x							x		
Barbieri et al.	2020	C3			x										x										x							x				
Bergamini and Zachmann	2020	C4			x				x						x									x		x	x								x	
Perruchas et al.	2020	C5			x				x						x										x		x			x						
Santoalha and Boschma	2021	C6			x				x						x									x		x									x	
Santoalha et al.	2021	C7			x				x						x									x		x	x								x	
de Cunzio et al.	2022	C8	x		x					x	x														x		x	x	x							
Barbieri et al.	2022	C9			x				x	x				x										x		x	x								x	
Napolitano et al.	2022	C10			x									x	x									x		x				x						
Grashof and Basilico	2023	C11			x				x					x										x		x	x									x

Figure 1: Summary of the reviewed articles, classified by comparable dimensions. The papers are sorted and colour-coded by group. Different colours indicate the unit of analysis of the articles, which sets the structure of the subsections that follow. Within the same group, papers are sorted by year. Papers belonging to the first group are highlighted in yellow (category A, country/regional level); second group in blue (category B, products); third group in green (category C, technologies). Each paper is associated to an alphanumeric code, in the first column, to aid the cross-reference with Tables 3, 4 and 5. 'Data' indicates the data sources employed; 'Method' refers to the specific methodologies adopted to analyze the data, beyond the EC toolbox; 'Definition of Sustainability' refers to the main definition of environmental sustainability adopted by each paper; 'Sample' indicates the specific dataset used. The coding of each paper within each dimension is not exclusive, i.e. each paper may draw upon more than one data source, method, sample, or definition of sustainability.

green transition based on their existing capabilities examining the geographical patterns of green innovative capacity. As can be appreciated from Fig. 1, this literature often explores the link and potential complementarity between green and non-green know-how, or productive capacity, mainly through proximity-based measures of relatedness between green and non-green technologies (Eq. 5), with a few exceptions relying on an asset matrix-based definition of relatedness (Eq. 6), and exploring geographical patterns through the definition of Green and non-green technology based measures of Fitness (Eq. 4). These dimensions of relatedness or green technological complexity are used by different contributions as explanatory variables for a series of socioeconomic dimensions, such as inequality, digital literacy, or support for environmental policies.

3.1. Economic Complexity and the environment

EC indices have been widely used by academics and policymakers to predict economic growth (Hidalgo and Hausmann, 2009; Tacchella et al., 2012; Cristelli et al., 2013; Tacchella et al., 2018). However, the pursuit of economic growth has been put under scrutiny by academics and society alike, due to its association to environmental impacts and climate change (Raworth, 2017; IPCC, 2022). Since the measurement of EC relies heavily on the nature of the products exported by countries or regions, which in turn impact the environment with different intensity (as they embody different levels of GHG emissions, have different energy requirements, and produce more or less polluting byproducts), there has been growing interest in understanding the relationship between the complexity of countries' productive structures and environmental impacts.

An ever-growing body of literature has investigated this relationship, looking especially at the export dimension of complexity and at country-level measures of environmental impacts – as highlighted by the "Data" column in Figure 1. Table 3 summarises the articles on the relationship between EC and the environment reviewed here, focusing on their dependent and independent variables, geographical

Table 3: Literature on the relationship between EC and environmental variables

Article	Dep. var.	Indep. var.	Geo	Rel. EC → Env.
A1 Can and Gozgor (2017)	CO ₂ emissions	Exogenous ECI, Energy consumption	France	Linear: ↑ ECI, ↓ CO ₂ ; GDP-based EKC (controlling for ECI)
A2 Neagu (2019)	CO ₂ emissions	ECI, Energy intensity	25 EU countries	Non-linear: ECI-based EKC
A3 Dong et al. (2020)	Coal consumption	Industrial Complexity Index (ICI) and Relatedness density (ECI-based); Energy-saving policies	Chinese provinces	Linear: ↑ ICI, ↓ Coal cons.; energy-saving policies associated to pollution haven hypothesis
A4 Swart and Brinkmann (2020)	Deforestation, Forest fires, Solid waste generation, Air pollution	ECI	Brazilian metropolitan regions	Linear: ↑ ECI, ↑ Fires ↓ Waste; GDP-based EKC
A5 Yilanci and Pata (2020)	Ecological footprint (composite index)	ECI, Energy consumption, Economic growth	China	Linear: ↑ ECI, ↑ Footprint; ECI-based EKC rejected
A6 Boleti et al. (2021)	Environmental Performance Index (EPI)	ECI/Fitness	88 low and high income countries	Linear: ↑ ECI/Fitness, ↑ EPII; GDP-based EKC
A7 Dogan et al. (2021)	CO ₂ emissions	ECI, Renewable energy consumption, Non-renewable energy consumption	OECD Countries	Linear: ↑ ECI, ↓ CO ₂
A8 Dordmond et al. (2021)	Green Jobs Index (GJI)	ECI	Brasilian states	Linear: ↑ ECI, ↑ GJI
A9 Martins et al. (2021)	CO ₂ emissions	ECI, Economic growth, Energy consumption, Globalisation	7 top ECI countries	Linear: ↑ EC, ↑ CO ₂ ; GDP-based EKC
A10 Pata (2021)	Ecological footprint (composite index)	ECI, Economic growth, Energy consumption, Globalisation	US	Non-linear: ECI-based EKC
A11 Romero and Gramkow (2021)	GHG intensity, GHG per capita, Product Emission Intensity Index	ECI	67 countries	Linear: ↑ ECI, ↓ GHG ↓ GHGpc ↓ PEII; centre of the product space has lower emissions
A12 Zheng et al. (2021)	CO ₂ emissions	ECI, Economic growth, Renewable energy consumption	16 leading exporters	Linear and two-way: ↑ ECI, ↓ CO ₂ ; GDP-based EKC
A13 Akadiri et al. (2022)	Ecological footprint (composite index)	ECI, Renewable and non-renewable energy use, Economic growth	China	Linear and two-way: ↑ ECI, ↑ Footprint ↑ Energy cons.

Table 3 continued from previous page

Article	Dep. var.	Indep. var.	Geo	Rel. EC → Env.
A14 Rafique et al. (2022)	Ecological footprint (composite index)	ECI, Human capital, Renewable energy, Urbanization, Economic growth, Export quality, Trade	Top 10 ECI countries	Linear: ↑ ECI, ↑ Footprint
A15 You et al. (2022)	CO ₂ emissions	ECI/Fitness, Economic growth	95 low, middle and high income countries	Linear: ↑ ECI/Fitness, ↓ CO ₂ (high-income); ↑ ECI/Fitness, ↑ CO ₂ (low- and middle-income)
A16 Esmaeili et al. (2023)	CO ₂ emissions	ECI, renewable energy, FDI	N11 countries	Non-linear: ECI-based EKC; Pollution Haven Hypothesis not rejected
A17 Stojkoski et al. (2023)	GDP growth, Income inequality, CO ₂ intensity	ECI/Fitness	World countries	Predicting CO ₂ intensity: Multidimensional EC (trade + technologies + publications) outperforms trade-based; Fitness outperforms ECI

focus, and main results. Contrarily to what early contributions seemed to suggest (for a review, see Hidalgo 2021), the last column in Table 3 hints towards aggregate findings that are far from unanimous, due to differing samples, geographical focus, and estimation techniques. This despite the fact that most of this evidence uses the Economic Complexity Index (Fig. 1) as explanatory variable; nevertheless, estimation techniques adopted to retrieve the effect of EC on the environment – and the way in which environmental impacts are measured – differ widely.

One strand of this literature assumes a linear relationship between EC and environmental impacts, producing mixed evidence. In a rather comprehensive study, Romero and Gramkow (2021) analyze the relationship between 67 countries’ complexity levels and their CO₂ emissions, measured as aggregate, per capita, and product-specific. The latter is proxied by the Product Emission Intensity Index (PEI), which averages the emissions of countries exporting a product with comparative advantage – following a methodology similar to Hartmann et al. (2017). The study finds that lagged ECI is associated with a reduction of both emissions intensity and per capita, and that lower emissions are associated with more interconnected (complex) products. Looking exclusively at OECD countries, Dogan et al. (2021) also find a positive relationship between EC and the reduction of CO₂ emissions between 1990-2014. In addition, the authors show that the complexity of exports interacts positively with the consumption of renewable energy, contributing to mitigating environmental impacts in high income countries. One

possible underlying mechanism is the emergence of greener occupations – and therefore greener economic activity – in regions with higher complexity, although the spatial dynamics of evolution towards the green occupations space prove to be rather sticky (Dordmond et al., 2021).

Although the studies mentioned above seem to agree on the positive contribution of EC to the environment, comparing their results is not always straightforward. For instance, Boleti et al. (2021) rely on a measure of environmental performance – the Environmental Performance Index¹¹ – to show that increased complexity is associated with better environmental performance across 88 low- and high-income countries over a (short) period spanning between 2002 and 2012. However, in the same period, higher complexity is also associated with worse air quality¹², higher emissions of CO₂ (kg per 2011 PPP USD of GDP), methane (kg of CO₂ equivalent), and nitrous oxide (thousand metric tons of CO₂ equivalent).

Some studies also find that EC is linked to an increase in CO₂ emissions. For instance, Neagu and Teodoru (2019) examine the heterogeneous effects of EC on GHG emissions in European countries. Their work shows that higher EC is positively linked to the growth of GHG emissions by countries, although this happens faster for countries with relatively lower levels of EC. These results are consistent with other studies focusing exclusively on the most complex countries: using time-series estimation techniques, Martins et al. (2021) find that higher EC is positively associated, in an unidirectional way, the levels of CO₂ emissions in the top 7 countries in the EC ranking. Nevertheless, integration in international trade has mitigated such negative effects, allowing early industrializers to shift towards knowledge-intensive, less polluting tasks. Similarly, Rafique et al. (2022) resort to dynamic panel data

¹¹ The Environmental Performance Index used in this study is obtained aggregating 40 indicators across 11 issues, such as Climate Change Mitigation; Air Quality; Sanitation and Drinking Water; Heavy Metals; Waste Management; Biodiversity and Habitat; Ecosystem Services; Fisheries; Acid Rain; Agriculture; Water Resources. For more information, see the latest EPI report: <https://epi.yale.edu/downloads/epi2022report06062022.pdf>

¹² Air quality is a composite index used to compute the EPI, and which in turn aggregates different indicators that quantify the number of age-standardized disability adjusted life-years lost per 100,000 persons (DALY rate) due to exposure to: PM_{2.5} exposure, household solid fuels, ozone exposure. Moreover, the air quality indicator includes measures of population-weighted annual average concentration at the ground level for: NO_x, sulfur dioxide, carbon monoxide, and volatile organic compound.

estimation techniques to find that ECI – along with urbanization and export growth – is positively linked to the ecological footprint¹³ in the top 10 complex countries.

The mixed evidence on the relationship between EC and environmental sustainability could also be explained by the fact that such linkage may be non-linear. It has been argued that countries increase their environmental impact in terms of GHG (with CO₂ being the most important) and pollutant emissions as they industrialize, reaching a peak in their emissions per capita. However, as they move towards more sophisticated activities and services, emissions per capita start to decrease, while GDP keeps growing. The reversed U-shape of the relationship between GDP and emissions has been named Environmental Kuznets Curve (EKC) (Grossman and Krueger, 1991; Selden and Song, 1994; Grossman and Krueger, 1995), building on the work by Kuznets (1955), who observed the same reversed U-shaped relationship between inequality and economic growth unfolding along with the process of structural transformation¹⁴. A growing body of literature has attempted to identify a complexity-based EKC; the intuition behind a reversed U-shaped relationship between EC and CO₂ emissions lies in the fact that, as countries become 'fitter' and accumulate productive capabilities, they shift their specialization towards more knowledge-intensive goods, with the latter being greener.

An ECI-based EKC has been identified especially for high-income countries, which find themselves at a mature post-industrial stage – that is, beyond the peak of emissions that coincided with higher intensity of employment manufacturing industries. In particular, ECI-based EKCs have been empirically tested for France (Can and Gozgor, 2017) and the US (Pata, 2021), as well as for a sample of 25 European countries (Neagu, 2019) and for leading exporters (Zheng et al., 2021). The ECI-based EKC hypothesis holds also for a sample of emerging economies

¹³ The authors rely on an index – the Ecological Footprint Index – which includes factors such as area occupied by forests, cropland, grazing, built-up land, fishing, and CO₂ emissions.

¹⁴ The EKC hypothesis suggests that the link between economic growth and its environmental impact depends on two channels: the change in composition of economic activities, and in the techniques of production (Grossman and Krueger, 1991). As countries shift from agricultural to industrial activities, there is a rise in carbon-intensive economic activities, leading to a positive correlation between GDP per capita and emissions. Thereafter, as modern techniques of production are introduced – leading to more efficient energy use and a shift to less energy-intensive services – the relationship between GDP per capita and emissions becomes negative. This results in a reversed U-shaped relationship between GDP per capita and emissions.

(Esmaeili et al., 2023)¹⁵, after controlling for FDI and renewable energy consumption. However, the same cannot be said for emerging economies when they are considered individually, as many of them are still intensive in activities associated with higher GHG emissions. For instance, empirical tests of the EKC in China – both at the aggregate (Yilanci and Pata, 2020) and regional level (Akadiri et al., 2022) – find no evidence of a reversed U-shaped relationship between complexity and CO₂ emissions. On the contrary, higher complexity appears to be associated with higher CO₂ emissions in both cases, despite the negative association between regional complexity and coal consumption in China (Dong et al., 2020). In the case of Brazilian regions (Swart and Brinkmann, 2020), the EKC hypothesis is validated using a quadratic fit between ECI and several indicators of environmental quality. The hypothesis holds for waste generation, but not for forest fires, deforestation, and air pollution. It is however important to distinguish that, in this case, the EKC tested does not refer to carbon emissions – which have global consequences on climate change – but to factors related to local environmental impacts.

It is important to mention that both the theoretical and empirical grounds of the EKC hypothesis – as well as its normative implications (Savona and Ciarli, 2019) – are not uncontroversial. First, the future ability of emerging countries – like China and Brazil – to decouple economic growth from GHG emissions is all but certain. This casts a doubt on whether we can expect low- and middle-income countries to go down the same path as currently high-income countries. Even if this will be the case, it must be taken into account that the polluting industries in which countries were once specialized may be simply outsourced to other countries, lending support to the pollution haven hypothesis¹⁶ (Cole, 2004). Moreover, the right part of the EKC – where the decoupling of economic growth and CO₂ emissions is observed – is populated only by a few developed countries (Csereklyei et al., 2016). Even if the same relationship was to hold also for developing countries in the future, waiting for all countries to move beyond the plateau may irremediably compromise

¹⁵ This article focuses on the Next-11 group, namely Iran, Egypt, Bangladesh, Mexico, Nigeria, South Korea, Turkey, Philippines, Vietnam, and Pakistan.

¹⁶ The pollution haven hypothesis states that regulations in high-income countries aiming at reducing GHG emissions may lead firm to relocate in countries with looser environmental regulations, thus leading to an outsourcing of CO₂-intensive activities rather to their effective reduction worldwide. For cases linking EC to the pollution haven hypothesis, see Dong et al. (2020) and Esmaeili et al. (2023).

the environment, as highlighted by the IPCC 2022.

Moreover, as noted by Stern (2004) in his critical review, the EKC literature tends to treat GDP per capita as an exogenous variable, ruling out any possible feedback from the environmental dimension to the productive sphere. Additionally, focusing on individual pollutants or GHG emissions has neglected the changing composition of emissions associated to the change in the productive structure. More recently, it has been noted that convergence approaches may be more relevant in describing the relationship between economic development and environmental impact, accounting also for non-industrial GHG emissions (Stern, 2017).

Another issue worth discussing revolves around the data used in the empirical analyses discussed above. While most of the empirical literature reviewed in this section relies on exported products data, the trade dimension is not the only relevant aspect in terms of environmental sustainability. As shown by Stojkoski et al. (2023), the CO₂ intensity across countries is explained more fully by a combination of trade, technological, and scientific complexity, computed respectively using data on exported products, patent applications, and scientific publications. Moreover, the authors show that not all complexity algorithms work equally well in predicting CO₂ intensity, showing that the Fitness algorithm (Tacchella et al., 2012) outperforms competing measures. As suggested by Fig. 1, however, only a small minority of authors of the papers reviewed here opts for the EFC algorithm (in this review, Boleti et al. 2021; You et al. 2022).

Finally, we would like to raise some methodological issues regarding the evidence canvassed in this section. Apart from rare exceptions (i.e Can and Gozgor 2017), the empirical papers that examine the relationship between EC and indices of environmental performance, use a measure – the Economic Complexity Index (Hidalgo and Hausmann, 2009) – which should only be treated as relative within each year (as explained in Section 2.2.1). As the term of reference for the measure changes every year, changes in ECI over time do not have any longitudinal interpretation, as the scale with which the index is measured changes every year. In order to address this, the use of EC metrics within longitudinal regressions should rely on the projection of product complexity in a given year of the series, upon countries' Revealed Comparative advantages in every year (Sbardella et al., 2018a; Operti et al., 2018) – i.e. an "exogenous" complexity, as explained at the end of Section 2.2.1. Alternatively, comparability over time is ensured by measures that maintains an invariance of scale, as suggested by Mazzilli et al. (2024) and explained in the

paragraph following Eq. 4.

One of the advantages offered by complexity methods in studying the sustainability transition is that these allow to observe productive structures and trajectories at a finely disaggregated level, for instance identifying single products or technologies towards which diversification should be steered in order to facilitate the transition towards greener activities. The role of products and technologies will be discussed in the following two subsections.

3.2. Economic Complexity for assessing green productive capabilities

Product-level data is the most commonly used data source in the EC framework. However, it is less frequently employed in studies linking EC and the sustainability transition, as compared to data on green patenting activity. This is mainly due to the difficulty of defining the 'green products' outlined in Section 2.1. Table 4 summarizes our selection of articles for this line of research.

Few attempts have been made to assess the green complexity/potential of national industrial systems. So far, the literature focused on the following research questions: Are green products more complex? How related to new green products is a given productive structure? Is there a significant difference in the dynamics of the product space for green and non-green products? Thus, on the one hand, scholars have tried to characterize green products as a particular subset of products in international trade networks by using standard EC tools. On the other hand, the main question – 'how prepared is a country to a green transition?' – is tackled mainly via network tools such as the products space, not introducing specific tools or frameworks for green commodities. Most contributions in the literature in fact apply the standard EC methodology and restrict the analysis to the subset of green products.

For instance, Fankhauser et al. (2013) assess the starting point of the green race for eight countries (China, Germany, UK, USA, France, Italy, Japan, and South Korea) for the period 2005-2007. They analyze three main factors of the green transition: green conversion, estimated using the Green Innovation Index (GII), an index constructed using patent data; change in RCAs; and green production at the outset, which, due to lack of specific data on this dimension, is assumed to be proportional to total production. Their findings reveal different areas of competitiveness among countries, detected from a correlation analysis of their Green

Table 4: Literature for EC based green products classifications

Article	EC indicator	EC methods	Geo	Main Results
B1 Fankhauser et al. (2013)	Green Innovation Index	Descriptive statistics	China, Germany, UK, USA, France, Italy, Japan, South Korea	Characterization of areas of green potential/weakness
B2 Hamwey et al. (2013)	Product proximity	Product space	Brazil	Full description of Brazilian product space and opportunities for new green production
B3 Fracascia et al. (2018)	Green product proximity	Product space and linear regression	141 countries	Products proximity is significant for new new green product export at 4-year horizon
B4 Pérez-Hernández et al. (2021)	Green Complexity Index and Potential	Product space	Regions of Mexico	Full characterization of Mexican green potential. Confirmed positive correlation between GCI and GCP. Possible scale-variant effects.
B5 Mealy and Teytelboym (2022)	ECI, Green Complexity Index and Potential	Product space	141 countries	Green products are more complex, negative correlation between GCI and CO ₂ /cap emission
B6 Müller and Eichhammer (2023)	PCI, GCI	Product space	141 countries, special focus on EU and MENA regions	Extended list of product relevant for green hydrogen production via proximity in the product space. Measure of GCI for green hydrogen-related product of EU and MENA countries

Innovation Index and changes in RCAs. They find no actual correlation between patenting GII and RCA competitiveness, which could mean a lack of policies for the green conversion in most of these countries.

Hamwey et al. (2013) apply the product space technique to identify opportunities for countries in green production by taking into consideration an arbitrary set of eleven environmental groups of products (in the SITC 4 classification at 4 digits) drawn from the list proposed by the WTO Committee on Trade and Environment (CTE) in 2011. In particular, their analysis focuses on Brazil, finding few selected products as potential opportunity of green diversification. However, the authors recognize the weakness of the green product definition, classification, and selection, as well as the limitation of neglecting the dynamics of the RCA time series they use to construct the products space.

Building on Hamwey et al. (2013), Fraccascia et al. (2018) apply the product space description for green products and measure the proximity between green products – as defined by the environmental goods and service sector classification by EUROSTAT (2009) – and non-green products in which world countries display an $RCA > 1$. Using panel regression analysis, they find this proximity to be significant for the development of the export of green products in a 4-years horizon.

In Mealy and Teytelboym (2022) two lists of green products are defined: a broader list comprising 293 green products, obtained by merging the WTO Core list, OECD lists, and the APEC list, and a shorter list of 57 renewable energy products. Using all products available in COMTRADE, they first compute the PCI of all products and construct a product space. Then, they focus on green products extracting the Green Complexity Index (GCI), the sum of green products' complexity, and the Green Complexity Potential (GCP), estimated by the proximity to green products in the product space. They reveal the complexity of green products to be significantly higher than the entire products baskets. Notably, they find a negative correlation between GCI and CO_2 emissions per capita, as well as a positive correlation between GCP and an increase of export in green products.

Pérez-Hernández et al. (2021) follow Mealy and Teytelboym (2022) and apply GCP and GCI at a sub-national level, studying the green potential of Mexican regions. They use an adapted version of the CLEG classification to characterize the green product space of regions. While their findings confirm an overall positive correlation between GCI and GCP, some interesting outliers arise are highlighted. For example, the state of Jalisco shows a high GCP and a comparative advantage on few high complexity products, but a low diversification in green products. This may be explained by some geographical or fine-grained factors, not captured in these indexes, that are less important at larger geographical scale.

Müller and Eichhammer (2023) apply the product space methodology focusing on the production of green hydrogen. They define a list of 36 goods 'that are potentially relevant for the production of green hydrogen from solar and wind power in a stand-alone configuration'. They also define a larger list by selecting 434 products with a high proximity, in the product space, to the first 36 goods. In agreement with the literature, they find higher PCI of products in these two lists compared to the entire HS products classification. Then they focus their study on European and Middle East and North African (MENA) countries, analyzing how their GCI evolved in the period 1995-2019 and assessing their readiness for the upcoming increasing

demand of green hydrogen, and their potential role in this specific supply chain.

3.3. Economic Complexity for assessing green technological capabilities

In recent years, the EC literature exploring the intricate relationship between green technological development, regional specialization dynamics, and policies has grown substantially. By drawing insights from evolutionary economic geography, these studies have revealed valuable insights into the dynamics and drivers of green technology advancement. This has been pursued by looking at the preparedness of European NUTS2 regions or national economic structures to the green technology race, and by highlighting the potential complementarities between green and non-green knowledge bases and/or productive structures.

The majority of the papers reviewed in this section rely on climate change mitigation and adaptation patent data as a proxy for low-carbon innovation, using mainly the ENV-TECH classification and the Y02/Y04S tagging scheme presented in Section 2.1. This is increasingly becoming the gold standard to measure green innovative activities.

Acknowledging the limitations of using patent data for technology development research (refer to Arts et al., 2013; Griliches, 1998; Lanjouw et al., 1998), it is important to note that patent information is readily accessible and offers a wealth of quantitative insights into inventions and their applicants or inventors. This includes geographical details, facilitating easy localization of patents on both national and local scales (Dechezleprêtre et al., 2011). Furthermore, patent data can be disaggregated into increasingly detailed technical domains, enabling the identification of highly specific green technologies (Haščič and Migotto, 2015). This granularity is particularly helpful in the use of EC techniques to study technological specialization (Boschma et al., 2015; Napolitano et al., 2018; Pugliese et al., 2019a). The application of EC approaches to reveal technological advantages in each technology field rests on the idea that the criteria to assign patent applications to specific domains are based on the identifying characteristics of the expertise that is necessary to introduce successful inventions. As a matter of fact, complex technologies appear almost exclusively in the portfolios of high-complexity countries and less diversified countries operating in less complex sectors.

As can be appreciated in Table 5, where we summarize the main results, methodologies, and data employed in the papers included in this section, some of

Table 5: Literature on green technologies

Article	Subtopic	Dep. var.	Indep. var.	EC def.	Relatedness def.	Geo	Main Results	Dataset; Geoloc.; Classification
C1 Sbardella et al. (2018a)	Green innovative capacity across countries			EFC (Green Technology Fitness)		72 countries	Global set of green technological capabilities is sticky with stable leaders, catch-ups (especially in East Asia) and laggards. Direct relationship between income and green innovative capacity, with complexity resonating with green technology life cycle	PATSTAT 2016 (applicants); GeoNames, Google Maps API; ENV-TECH
C2 Montresor and Quatraro (2020)	Green and non-green technology and key enabling technologies (KETs)	New RTAs in green technologies	Relatedness of green technology to all, non-green, green technologies; new RTAs in KETs, interaction between new RTAs in KETs and relatedness		Proximity-based relatedness of tech. z at time t in region i to all technologies at t ; green or non-green technologies at $t - 5$	EU NUTS2 regions	Relatedness to non-green (rather than green) knowledge makes new green-tech specialisation more likely; KETs favour transition to green technologies and moderate relatedness to pre-existing technologies' impact	REGAPT (applicants); REGPAT NUTS3; ENV-TECH
C3 Barbieri et al. (2020)	Green technology life cycle, related and unrelated variety	Green technology patent families (all and at different life cycle stages)	Green technology related, semirelated, unrelated variety	Technology ubiquity	Related, semirelated, unrelated variety of green technologies computed with Shannon entropy if they share IPC code at different aggregation levels	US states	Unrelated variety positive predictor of green-tech specialisation; unrelated variety main driver of green technology development over the life cycle	PATSTAT 2016 (applicants); GeoNames, Google Maps API; ENV-TECH

Table 5 continued from previous page

Article	Subtopic	Dep. var.	Indep. var.	EC def.	Relatedness def.	Geo	Main Results	Dataset; Geoloc.; Classification
C4 Bergamini and Zachmann (2020)	Potential for green technology development and socio-economic regional characteristics	RTAs in green technologies at time t	Potential technological advantage (Hausmann et al., 2022); labour market participation, STEM employment, education, R&D expenditures etc.		Proximity-based relatedness of green tech. z in which region i not specialised at time t to all technologies in which i already specialised at t	EU NUTS2 regions	Green techs RTAs concentrated in FR, DE, Northern IT, increase over time, but depend on technology type. Future regional green RTAs positively associated with labour market participation rate, employment duration, STEM employment, R&D exp., higher education	REGPAT (inventors and applicants); REGPAT NUTS3; Fiorini et al. (2017) classif. based on Y02/Y04S
C5 Perruchas et al. (2020)	Green technological specialisation and life cycle across countries	New RTAs in green technologies (all/only if country has RTA >1 at $t-1$)	Relatedness, technology life cycle, environmental policy stringency	ECI (green technology complexity)	Proximity-based relatedness of green tech. z in which country i specialised at time t to technologies in which i already specialised at t	63 countries	Diversification in green technologies related to pre-existing competences and more associated to green technology maturity than income, with green techn complexity not hampering further specialisation	PATSTAT 2016 (applicants); GeoNames, Google Maps API; ENV-TECH
C6 Santoalha and Boschma (2021)	Green technology and national or regional environmental policy support	New RTAs in green technologies	National and regional environmental policy support, green to non-green relatedness		Proximity-based relatedness of green tech. z in which region i not specialised at time t to technologies in which i specialised at t	EU NUTS2 regions	Related technological capabilities more important than regional environmental policies, while national political support for environmental policies mitigates the importance of related technologies	REGPAT (applicants); REGPAT NUTS3; ENV-TECH
C7 Santoalha et al. (2021)	Green technology and digital skills	New RTAs in green or non-green technologies	Digital literacy, overall, green to non-green relatedness		Proximity-based relatedness of green tech. z in which region i not specialised at time t to technologies in which i already specialised at t	EU NUTS2 regions	Digital skills have a positive impact on a region's ability to specialize in new technologies, especially in green domains, with e-skills moderating the effect of relatedness	REGPAT (applicants); REGPAT NUTS3; ENV-TECH

Table 5 continued from previous page

Article	Subtopic	Dep. var.	Indep. var.	EC def.	Relatedness def.	Geo	Main Results	Dataset; Geoloc.; Classification
C8 de Cunzio et al. (2022)	Green technology and exported goods	RCAs in exported products	RTAs in green technologies	EFC (Green Technology Complexity)	Multi-layer based relatedness of green technology z at time t or $t - 5$ to product p at t	48 countries	Strong connection between green technologies and exported products related to the processing of raw materials; with time, more complex green knowledge trickles down into more complex goods	REGPAT (applicants); REGPAT countries; Y02/Y04S
C9 Barbieri et al. (2022)	Green and non-green know-how in EU regions	Green Technology Fitness	Non-green Technology Fitness, Non-green to green relatedness	EFC (Green and Non-green Technology Fitness)	Multi-layer based relatedness of non-green tech. z in which region i specialised at time $t - 5$ to green tech. z in which i already specialised at t	EU NUTS2 regions	Non-green and green competitiveness geographically heterogeneous and stable over time (persistent between Central-Eastern EU dichotomy). Complementarity between non-green and green tech know-how, but green specialisation depends on regional patent portfolios' composition	PATSTAT (inventors); PATSTAT inventor residence, de Rassenfosse et al. (2019)'s geocoding, EUROSTAT GIS NUTS 3; Y02/Y04S
C10 Napolitano et al. (2022)	Green innovative capacity and income inequality	Green Technology Fitness	Income inequality	EFC (Green Technology Fitness)		57 countries	Green innovative capacity negatively correlated to income inequality, but inequality irrelevant for high-GDP countries, while for middle-income countries lower inequality favours green specialisation	PATSTAT 2016 (applicants); GeoNames, Google Maps API; ENV-TECH
C11 Grashof and Basilico (2023)	Regional divergence, relatedness between brown and green know-how in EU regions	New RTAs in green technologies (from time $t - 5$ to t)	Relatedness to green or brown technologies (at time t), economic strength ($= 1$ if region i 's GDP per capita $\geq 75\%$ of EU average, $= 0$ otherwise)	ECI (brown and green Technology Complexity)	Proximity-based relatedness of brown/green technology z in which region i not specialised at time t to technologies in which i specialised at t	EU NUTS2 regions	Related technological know-how more important for green specialisation than economic strength; high-income regions higher green relatedness, but low-income regions with high green relatedness can successfully diversify into complex green tech	REGPAT (inventors); REGPAT NUTS2; WIPO IPC Green/Brown Inventory

these contributions focus on the interplay between green capabilities and the maturity of green technologies in shaping the way in which the technological portfolios of regions or countries grow and evolve over time. Other contributions explore instead the relation between income inequality and innovative capacities (both green and non-green) at the country level, revealing the influence of socioeconomic and policy factors (e.g. smart specialization strategies, political support, digital literacy) . Overall, these contributions offer a cohesive narrative, that emphasizes the important role played by local capabilities and relatedness to non-green knowledge bases or productive capacity in fostering new green technological innovation, providing valuable guidance for policymakers seeking to promote sustainable models of economic development.

Perruchas et al. (2020) investigate the relationship between country characteristics and knowledge structures in the progression of green ENV-TECH patenting, and national specialization or diversification patterns from the end of the 1970s. The authors focus on the life cycle of green technologies and propose a 'ladder of green technology development'. Their evidence emphasizes that not only countries diversify towards green technologies related to their existing competencies, but also that green specialization follows a cumulative path towards more mature technologies. Technology maturity appears to be more relevant than a country's economic development, while technology complexity – computed through the ECI algorithm introduced in Eq. 3 applied to green patent data – does not prevent further specialization. By focusing on a panel of US states from the early 1980s, the same authors (Barbieri et al., 2020) examine the role of related/unrelated variety in green technology development and its heterogeneous effect over the technology life cycle. The authors observe that unrelated variety plays a positive role in fostering green innovation at early stages of the technology life cycle, while as technologies mature, related variety becomes a more important driver.

Sbardella et al. (2018a) introduce Green Technological Fitness (GTF), a measure of green innovative potential based on green patenting as defined in Eq. 4 – here identified within PATSTAT through the ENV-TECH catalogue. Taking a geographical approach, the authors identify heterogeneous global patterns in green technological competitiveness, with the United States, France, and Germany as stable leaders; Eastern and Southern European countries gradually gaining importance; and East Asian countries starting from the periphery and rapidly establishing themselves as key actors. Finally, by analyzing the distribution of countries’ innovation capacity across areas of specialization, they document that innovation in green technology has become more horizontal, with bigger efforts being observed in cross-domain, or enabling technologies.

Finally, focusing only on the technological dimension and not looking at geographical dynamics, de Cunzio et al. (2022) explore the connection between green technological innovation capacity and productive capabilities. In order to do so, the authors define a statistically validated network – as defined in Eq. 6 – connecting comparative advantages in Y02 green technologies to the contemporaneous or subsequent comparative advantages in exported products. When looking at same-year co-occurrences of single products at technologies, their findings emphasize a large number of significant links between green technologies and raw materials, especially critical minerals such as cobalt. In contrast, when selecting ten-year time-lagged green technology-product links, they suggest that green technology is better integrated into manufacturing, and that more complex spillover effects emerge with new additional links of products and technologies with higher complexity.

Evidence of different dimensions of complementarity between green technologies and non-green pre-existing innovative capacity in European NUTS2 regions is investigated by five articles (Bergamini and Zachmann, 2020; Montresor and Quatraro, 2020; Barbieri et al., 2022; Sbardella et al., 2022; Grashof and Basilico, 2023) using different approaches to assess relatedness. Firstly, Bergamini and Zachmann (2020) use regional patent data for Europe sourced from the REGPAT database (Maraut et al., 2008) to predict the potential of European regions to acquire (or maintain) competitive advantages in developing green technologies. To this aim, the authors first leverage a network-based measure of relatedness (as in Eq. 5) between non-green technologies to estimate the potential technological advantage (RTA) in green technologies of each region. In a second stage, they identify a set of socioeconomic variables that hold a statistically significant association with

observed/estimated RTAs and derive possible policy implications.

Secondly, Montresor and Quatraro (2020) explore the relationship between ENV-TECH green patenting and smart specialization strategies in European regions, paying attention to the role of key enabling technologies.¹⁷ With the aim of understanding if the development or acquisition of new sustainable technologies is related to regional technological innovative capacity, the study shows that there is a strong connection between green technologies and both green and non-green pre-existing knowledge bases. Moreover, key enabling technologies support the shift towards green technologies also mitigating the impact of relatedness to pre-existing technologies.

Thirdly, Barbieri et al. (2022) and Sbardella et al. (2022) investigate the nexus between non-green (A-H PATSTAT patents) and green (Y02-Y04S) innovative capacity in and the green development potential of European regions, computing measures of *exogenous* fitness as introduced in Section 2.2.1, regional Non-Green Technology Fitness (NGTF), and Green Technology Fitness (GTF). Using the statistically validated network approach introduced in Section 2.2.2, they define a green potential metric quantifying the relatedness (as defined in Eq. 6) between non-green and green regional knowledge bases. The authors document a heterogeneous but stable distribution of non-green and green innovative competitiveness, and reveal a degree of complementarity between non-green and green knowledge capabilities, albeit dependent on the patent portfolio composition – especially for regions that have not fully developed the entire set of Y02-Y04S technologies.

Finally, by investigating the nexus between EU NUTS2 regions's knowledge base, economic performance, and green technology specialization patterns (identified through WIPO IPC Green Inventory) between 2000 and 2017, Grashof and Basilico (2023) argue that the sustainability transition will not widen regional disparities. Their analysis, where a proximity-based relatedness is computed according to Eq. 5, suggests that both high- and low-income regions can diversify into green technologies. While high-income regions are more specialized in technologies related to green innovations, they do not necessarily have more entries in green technologies. Instead, related technological capabilities appear to be more important for green

¹⁷ The term "key enabling technologies" denote six crucial technological domains identified by the European Commission: 2021 industrial biotechnology, nanotechnology, micro- and nanoelectronics, photonics, advanced materials, and advanced manufacturing technologies.

diversification than economic strength. Notably, low-income regions with high green relatedness can successfully diversify into new, and even more complex, green technologies.

Different contributions in this field are in agreement in suggesting that relatedness is a driving force behind diversification in green technologies. However, this new field of analysis has not yet fully emphasized the role of the socioeconomic fabric and institutional set-up in sustaining the sustainability transition. The sustainability transition literature has paid greater attention to the policy and socioeconomic dimensions, however, putting forward an important but not systematic collection of case studies, has often failed to provide generalizable or scalable evidence on the role of environmental policies or local characteristics in shifting towards a more sustainable economy. The three following articles go in this direction and analyze how green technological competitiveness, diversification in green technologies, and/or relatedness to non-green knowledge interact with a number of socioeconomic characteristics of countries or regions.

Looking at the regional dimension, Santoalha and Boschma (2021) examines the relationship between political support for environmental policies and the diversification of green technologies in European regions. By employing a measure of relatedness similar to the one encapsulated in Eq. 5, they find evidence of a stronger relationship between related capabilities and green diversification, as compared to regional policies. However, national political support appears to mitigate the importance of capabilities.

Building on a similar metric for relatedness, Santoalha et al. (2021) operationalize the notion of capabilities by studying whether and to what extent digital literacy – as a proxy for the competencies embedded in ICT infrastructures – foster diversification in green and non-green technologies across European regions. The level of digital skills in the workforce has a positive impact on a region’s ability to specialize in new technologies, especially in green domains, with e-skills moderating the effect of relatedness.

Shifting to a country-level analysis, Napolitano et al. (2022) investigate whether and to what extent income inequality is a barrier to a country’s environmental innovative capacity, proxied by a measure of Green Technological Fitness based on ENV-TECH technologies. To this end, and differently from Sbardella et al. (2018a), they define a measure of *sectoral* fitness, to provide a more realistic assessment of green technology complexity: first, they account for the full technological spectrum

in computing the EFC algorithm (Eq. 4); then, they select only the complexities of ENV-TECH technologies in computing GTF. A negative and significant relationship between income inequality and GTF is observed; by contrast, no significant association is found when all technologies are considered. However, while for high-income countries inequality does not appear to be a barrier, there is an income threshold below which it is unlikely to develop a sufficient number of complex technologies to achieve high green fitness. Low inequality reduces such thresholds allowing middle-income countries to achieve greater green innovative capacity.

4. Conclusions

In this review we have provided an encompassing account of the empirical literature implementing EC methods and metrics to understand the sustainability transition. In this fast growing literature, this work will help the reader to navigate through the vast array of methods and data used in the framework. With more and more scholars entering this field, this review will serve as a rapid introduction to the most relevant papers in terms of methodology, applications and results.

First, the paper has summarized the most used product- and patent-level data sources to compute EC metrics, with a particular focus on green economic activities. Secondly, the review has attempted to harmonize the most relevant methods adopted in the literature on EC and environmental sustainability, in order to identify similarities and differences across methods relying on a common framework and language. Third, the three main blocks of empirical literature linking EC methods and metrics have been reviewed, looking at i) the relationship between countries' and regions' EC and the environment, and the role of ii) green products and iii) green technologies in fostering the sustainability transition.

The growing literature on EC and the environment suggests that EC approaches can be a useful lens to better understand how productive structures and technological capabilities can be steered into the sustainability transition. The empirical literature reviewed here offers some evidence that can be summarized as follows. With respect to the first block of literature, which examines how trade-based EC measures link to environmental outcomes such as GHG emissions, ecological footprint, and environmental impacts, we conclude that the evidence produced is mixed. On the one hand, high-income countries – which are more specialized in knowledge-intensive activities – seem to exhibit a positive association between the complexity of their

export basket and their ability to preserve the environment. However, the same relationship cannot be observed in less mature, emerging economies, where increasing complexity is associated to higher environmental impact. Empirical test of the EKC hypothesis have shown that countries become able to decouple economic growth from carbon emissions only after passing the industrialization phase. Having this said, the extant literature on green EC has some important limitations. First, the literature on country/regional complexity and sustainability relies largely on aggregate measures of EC. This approach can be limiting if confined to examining the relationship between complexity and environmental variables, without taking advantage of the high level of disaggregation made possible by EC methods. Secondly, as we have argued in Section 3.1, the quest for empirical validation of the EKC hypothesis can be of little normative utility, considering that not all countries will be able to emulate the diversification and specialization patterns of today’s high-income countries, some of which are specialized in knowledge-intensive activities. Moreover, as shown by the IPCC (2022), the current efforts to curb carbon emissions also in high-income countries may be insufficient to prevent a climatic catastrophe.

Looking at the second block of literature, one of the main goals of these studies is to steer production systems toward a green, feasible transition. While the nature of the trade-off between these elements and the socio-economic outcome of this shift is left to more theoretical studies, EC has the potential to identify, characterize, and measure potential paths of such transformation. From the reviewed attempts, it is clear that a greener production of commodities requires higher capabilities. Green products show higher complexity and are intertwined to non-green production. As stressed in the Data section of this paper, the studies relying on green classifications for products are subject to the problem of how to define a green product: different classifications based on inconsistent definitions may undermine the replicability of the results. A green sub-classification in the Harmonized System would be desirable for the upcoming updates, so to have a single and exhaustive database for green products. On the methodological front, most contributions in the EC literature studying green products apply the standard product space description to approach the problem of identifying the green potential of productive structures. While the product space is indeed useful to visualize the distances between the current production and green products, it has been shown that this approach has low accuracy in quantitative forecasting of future diversification (Tacchella et al., 2023).

The third block of studies on green technological development yields several key

conclusions. Firstly, there is evidence of countries engaging in a dual strategy of diversification and specialization, entering green technologies aligned with existing competencies while specializing in mature technologies with accumulated experience. Geographically, there is a mix of stable leaders and emerging players in green technological competitiveness, with Eastern and Southern European countries gaining prominence. Furthermore, economic factors – such as income inequality – affect a country’s environmental innovative capacity. For instance, lower income inequality in middle-income countries seems to lower barriers to the successful development of complex green technologies. Finally, the complementarity between non-green and green innovation capabilities, the influence of regional factors and policies, and the significance of digital literacy and e-skills in promoting green technology adoption are highlighted. These findings provide valuable insights into the complex dynamics of green technological development, emphasizing the interplay between capabilities, regional factors, policies, and socioeconomic aspects.

Whilst the studies on technology-based EC can shed light on various aspects of green technological development, it is important to acknowledge their limitations. Firstly, the analyses often rely on patent data as a proxy for technological innovation, which may not capture the full spectrum of green technologies, or account for innovations that are not patented. This could lead to a potential underrepresentation of certain sectors or technologies. Secondly, the studies primarily focus on regional or national levels of analysis, which may overlook the dynamics at the firm or individual levels. The role of specific organizations or entrepreneurs in driving green innovation is not extensively explored, potentially limiting the understanding of micro-level factors influencing technological development. Lastly, the studies primarily analyze the relationships between different variables and identify correlations rather than establishing causal relationships. While the associations observed are informative, further research is needed to delve deeper into the underlying mechanisms and causality. These limitations highlight the need for further research that encompasses a broader range of data sources, considers multiple levels of analysis, and employs rigorous methodologies to better understand the complexities of green technological development and its implications for environmental sustainability.

The discussion on the empirical literature presented in Section 3 uncovers a number of areas that require further exploration. First, the productive and technological structures of countries are always accounted for separately in relation

to their linkage with environmental technologies. Whilst there have been previous efforts linking the productive, technological and scientific capabilities of countries with each other Pugliese et al. (2019a), and one contribution focused on the green know-how-productive capacity nexus de Cunzo et al. (2022), such effort should be extended to better understand which specific capabilities are the most conducive to green specialization across fields. Second, and related to the previous point, scientific capabilities have not yet been included in the discourse on the sustainability transition from a complexity perspective. Their relationship with the development of green capabilities should be examined in depth. In fact, being cognizant of the broad consensus on the key role of green technologies for the sustainability transition, these cannot be viewed as a one-size fits all solution, and should be integrated in a broad policy agenda targeting wider dimensions of socioeconomic sustainability (Parkinson, 2010; Sarewitz and Nelson, 2008).

Third, ensuring a green and fair transition will require to take into account its interrelation with the process of structural change, labour reallocation, and their distributive consequences. Future research will have to assess the job creation/destruction potential of the green transition, and to identify the transversal skills required to ensure a seamless reallocation of workers across jobs. By the same token, not all territories and geographies are equally well equipped with the necessary knowledge base to diversify away from carbon-intensive production and technologies. It will be paramount to identify the most viable ways for regions and countries to enter green economic activities, bearing in mind the potentially negative (or positive) impact on local labor markets.

Finally, green technologies and products may create further pressure on the environment due to their dependency on natural commodities and minerals (Marín and Goya, 2021), such as rare earth elements, lithium, cobalt, and others (de Cunzo et al., 2023).

Precisely, the World Bank (Hund et al., 2023) estimates that meeting the 2°C scenario by 2050 for energy storage alone will require a 450 percent increase in the production of graphite, lithium, and cobalt. Therefore, while pursuing the green transition may contribute to reducing global dependence on fossil fuels, keeping up with current levels of energy demand will shift the pressure towards the extraction and trade of raw materials, with increasing risks of creating new supply chain vulnerability (European Commission, 2021; Hund et al., 2023; International Energy Agency, 2021; Herrington, 2021; Kowalski and Legendre, 2023) and moving polluting

and exploitative practices in the Global South (Berman et al., 2017; Church and Crawford, 2018; Christensen, 2018; Sovacool et al., 2020), thus exacerbating social and environmental pressures in an already globally uncertain outlook (Norgate and Haque, 2010; Wanger, 2011; Romare and Dahllöf, 2017; Azadi et al., 2020). For this reason, future research will need to examine the production processes and recycling potential associated to each green product, the CO₂ production incorporated in each value chain, its raw material content, the safety (also concerning toxicity and pollutant exposure) and workplace condition of the labor force employed in its production, and the environmental impact of production and use (e.g. life cycle emissions, energy content, and waste management).

References

- Akadiri, S. S., Adebayo, T. S., Asuzu, O. C., Onuogu, I. C., and Oji-Okoro, I. (2022). Testing the role of economic complexity on the ecological footprint in China: a nonparametric causality-in-quantiles approach. *Energy & Environment*, 0(0):0958305X221094573.
- Angelucci, S., Hurtado-Albir, F. J., and Volpe, A. (2018). Supporting global initiatives on climate change: The EPO’s “Y02-Y04S” tagging scheme. *World Patent Information*, 54:S85–S92.
- APEC (2012). ANNEX C - APEC List of Environmental Goods. Technical report, APEC.
- Arts, S., Appio, F. P., and Van Looy, B. (2013). Inventions shaping technological trajectories: do existing patent indicators provide a comprehensive picture? *Scientometrics*, 97(2):397–419.
- Atmar, W. and Patters, B. (1993). The measure of order and disorder in the distribution of species in fragmental habitat. *Oecologia*, 96.
- Azadi, M., Northey, S. A., Ali, S. H., and Edraki, M. (2020). Transparency on greenhouse gas emissions from mining to enable climate change mitigation. *Nature Geoscience*, 13(2):100–104.
- Balassa, B. (1965). Trade liberalisation and “revealed” comparative advantage. *The Manchester School*, 33(2):99–123.
- Balland, P.-A., Broekel, T., Diodato, D., Giuliani, E., Hausmann, R., O’Clery, N., and Rigby, D. (2022). The new paradigm of economic complexity. *Research Policy*, 51(3):104450.
- Balland, P.-A. and Rigby, D. (2017). The geography of complex knowledge. *Economic Geography*, 93(1):1–23.
- Barbieri, N., Consoli, D., Napolitano, L., Perruchas, F., Pugliese, E., and Sbardella, A. (2022). Regional technological capabilities and green opportunities in europe. *The Journal of Technology Transfer*, pages 1–30.
- Barbieri, N., Perruchas, F., and Consoli, D. (2020). Specialization, diversification, and environmental technology life cycle. *Economic Geography*, 96(2):161–186.
- Bergamini, E. and Zachmann, G. (2020). Exploring EU’s regional potential in low-carbon technologies. *Sustainability*, 13(1):32.

References

- Berman, N., Couttenier, M., Rohner, D., and Thoenig, M. (2017). This mine is mine! how minerals fuel conflicts in africa. *American Economic Review*, 107(6):1564–1610.
- Bistline, J., Mehrotra, N., and Wolfram, C. (2023). Economic implications of the climate provisions of the inflation reduction act. Working Paper 31267, National Bureau of Economic Research.
- Boleti, E., Garas, A., Kyriakou, A., and Lapatinas, A. (2021). Economic complexity and environmental performance: Evidence from a world sample. *Environmental Modeling & Assessment*, 26(3):251–270.
- Boschma, R., Balland, P.-A., and Kogler, D. F. (2015). Relatedness and technological change in cities: The rise and fall of technological knowledge in US metropolitan areas from 1981 to 2010. *Industrial and Corporate Change*, 24(1):223–250.
- Boschma, R. and Frenken, K. (2018). Evolutionary economic geography. *The new Oxford handbook of economic geography*, pages 213–229.
- Boschma, R. A. and Frenken, K. (2006). Why is economic geography not an evolutionary science? Towards an evolutionary economic geography. *Journal of Economic Geography*, 6(3):273–302.
- Breschi, S., Lissoni, F., and Malerba, F. (2003). Knowledge-relatedness in firm technological diversification. *Research Policy*, 32(1):69–87.
- Bruno, M., Mazzilli, D., Patelli, A., Squartini, T., and Saracco, F. (2023). Inferring comparative advantage via entropy maximization.
- Caldarelli, G., Cristelli, M., Gabrielli, A., Pietronero, L., Scala, A., and Tacchella, A. (2012). A network analysis of countries’ export flows: firm grounds for the building blocks of the economy. *PLoS ONE*, 7.
- Can, M. and Gozgor, G. (2017). The impact of economic complexity on carbon emissions: evidence from France. *Environmental Science and Pollution Research*, 24(19):16364–16370.
- Christensen, D. (2018). Concession stands: How mining investments incite protest in africa. *International Organization*, 73(1):65–101.
- Church, C. and Crawford, A. (2018). Green Conflict Minerals: The fuels of conflict in the transition to a low-carbon economy. Technical report, International Institute for Sustainable Development (IISD).

References

- Cimini, G., Squartini, T., Saracco, F., Garlaschelli, D., Gabrielli, A., and Caldarelli, G. (2019). The statistical physics of real-world networks. *Nature Review Physics*, 1:58–71.
- Cimoli, M. and Dosi, G. (1995). Technological paradigms, patterns of learning and development: An introductory roadmap. *Journal of Evolutionary Economics*, 5(3):243–268.
- Cole, M. A. (2004). Trade, the pollution haven hypothesis and the environmental kuznets curve: examining the linkages. *Ecological Economics*, 48(1):71–81.
- Common, M. and Stagl, S. (2005). *Ecological economics: an introduction*. Cambridge University Press.
- Cristelli, M., Gabrielli, A., Tacchella, A., Caldarelli, G., and Pietronero, L. (2013). Measuring the intangibles: A metrics for the economic complexity of countries and products. *PloS one*, 8(8):e70726.
- Cristelli, M., Tacchella, A., and Pietronero, L. (2015). The heterogeneous dynamics of economic complexity. *PloS one*, 10(2):e0117174.
- Csereklyei, Z., Rubio-Varas, M. d. M., and Stern, D. I. (2016). Energy and economic growth: the stylized facts. *The Energy Journal*, 37(2).
- de Cunzio, F., Consoli, D., Perruchas, F., and Sbardella, A. (2023). Mapping critical raw materials in green technologies Technical Report 23.2, Utrecht University.
- de Cunzio, F., Petri, A., Zaccaria, A., and Sbardella, A. (2022). The trickle down from environmental innovation to productive complexity. *Scientific Reports*, 12(1):22141.
- de Rassenfosse, G., Kozak, J., and Seliger, F. (2019). Geocoding of worldwide patent data. *Scientific Data*, 6(1):1–15.
- Dechezleprêtre, A., Glachant, M., Haščič, I., Johnstone, N., and Ménière, Y. (2011). Invention and transfer of climate change-mitigation technologies: A global analysis. *Review of Environmental Economics and Policy*, 5(1):109–130.
- Dogan, B., Driha, O. M., Balsalobre Lorente, D., and Shahzad, U. (2021). The mitigating effects of economic complexity and renewable energy on carbon emissions in developed countries. *Sustainable Development*, 29(1):1–12.
- Dong, Z., Chen, W., and Wang, S. (2020). Emission reduction target, complexity and industrial performance. *Journal of Environmental Management*, 260:110148.

References

- Dordmond, G., de Oliveira, H. C., Silva, I. R., and Swart, J. (2021). The complexity of green job creation: An analysis of green job development in Brazil. *Environment, Development and Sustainability*, 23(1):723–746.
- Dosi, G. and Nelson, R. R. (1994). An introduction to evolutionary theories in economics. *Journal of Evolutionary Economics*, 4(3):153–172.
- Esmaeili, P., Balsalobre Lorente, D., and Anwar, A. (2023). Revisiting the environmental kuznetz curve and pollution haven hypothesis in n-11 economies: Fresh evidence from panel quantile regression. *Environmental Research*, 228:115844.
- European Commission (2020a). Critical raw materials resilience: Charting a path towards greater security and sustainability. Technical report, European Commission.
- European Patent Office (2013). EPO–Finding sustainable technologies in patents.
- European Patent Office (2018). EPO worldwide patent statistical database data catalog, climate change mitigation technologies.
- EUROSTAT (2009). The environmental goods and services sector.
- Fankhauser, S., Bowen, A., Calel, R., Dechezleprêtre, A., Grover, D., Rydge, J., and Sato, M. (2013). Who will win the green race? In search of environmental competitiveness and innovation. *Global Environmental Change*, 23(5):902–913.
- Fiorini, A., Georgakaki, A., Pasimeni, F., and Tzimas, E. (2017). Monitoring R&I in low-carbon energy technologies. *Publications Office of the European Union: Luxembourg*.
- Fraccascia, L., Giannoccaro, I., and Albino, V. (2018). Green product development: What does the country product space imply? *Journal of Cleaner Production*, 170:1076–1088.
- Grashof, N. and Basilico, S. (2023). The dark side of green innovation? green transition and regional inequality in europe. Technical report, Utrecht University, Department of Human Geography and Spatial Planning
- Griliches, Z. (1998). Patent statistics as economic indicators: A survey. In *R&D and productivity: The econometric evidence*, pages 287–343. University of Chicago Press.
- Grossman, G. M. and Krueger, A. B. (1991). Environmental impacts of a north

References

- american free trade agreement. Working Paper 3914, National Bureau of Economic Research.
- Grossman, G. M. and Krueger, A. B. (1995). Economic Growth and the Environment. *The Quarterly Journal of Economics*, 110(2):353–377.
- Hamwey, R., Pacini, H., and Assunção, L. (2013). Mapping green product spaces of nations. *The Journal of Environment & Development*, 22(2):155–168.
- Hartmann, D., Guevara, M. R., Jara-Figueroa, C., Aristarán, M., and Hidalgo, C. A. (2017). Linking economic complexity, institutions, and income inequality. *World Development*, 93:75–93.
- Haščič, I. and Migotto, M. (2015). Measuring environmental innovation using patent data. OECD Environment Working Papers 89, OECD Publishing, Paris.
- Hausmann, R. and Hidalgo, C. A. (2011). The network structure of economic output. *Journal of Economic Growth*, pages 309–342.
- Hausmann, R., Hidalgo, C. A., Bustos, S., Coscia, M., Chung, S., Jimenez, Juan Simoes, A., and Yildirim, M. A. (2011). *The Atlas of economic complexity: Mapping paths to prosperity*. Puritan Press, New Hampshire.
- Hausmann, R., Hidalgo, C. A., Bustos, S., Coscia, M., and Simoes, A. (2014). *The Atlas of economic complexity: Mapping paths to prosperity*. MIT Press.
- Hausmann, R., Hwang, J., and Rodrik, D. (2007). What you export matters. *Journal of Economic Growth*, 12:1–25.
- Hausmann, R. and Rodrik, D. (2003). Economic development as self-discovery. *Journal of Development Economics*, 72(2):603–633.
- Hausmann, R., Rodrik, D., Benavente, J. M., and Rodríguez, F. (2005). Self-discovery in a development strategy for El Salvador [with comments]. *Economía*, 6(1):43–101.
- Hausmann, R., Stock, D. P., and Yıldırım, M. A. (2022). Implied comparative advantage. *Research Policy*, 51(8):104143.
- Herrington, R. (2021). Mining our green future. *Nature Reviews Materials*, 6(6):456–458.
- Hidalgo, C. A. (2009). The Dynamics of Economic Complexity and the Product Space over a 42 year period. *CID Working Paper Series*, 189.
- Hidalgo, C. A. (2021). Economic complexity theory and applications. *Nature Reviews Physics*, 3(2):92–113.

References

- Hidalgo, C. A. and Hausmann, R. (2009). The building blocks of economic complexity. *Proceedings of the National Academy of Sciences*, 106(26):10570–10575.
- Hidalgo, C. A., Klinger, B., Barabási, A.-L., and Hausmann, R. (2007). The Product Space Conditions the Development of Nations. *Science*, 317(5837):482–487.
- Hirschman, A. O. (1958). *The Strategy of Economic Development*. Yale University Press, New Haven.
- Hund, K., La Porta, D., Fabregas, T., Laing, T., and Drexhage, J. (2023). Minerals for climate action: The mineral intensity of the clean energy transition. The World Bank Group, Washington, DC.
- International Energy Agency (2021). The role of critical minerals in clean energy transitions. Technical report, International Energy Agency, Paris.
- IPCC (2022). *Climate Change 2022: Impacts, Adaptation, and Vulnerability*. Contribution of Working Group II to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press, Cambridge, UK and New York, USA.
- Jaffe, A. B. (1986). Technological Opportunity and Spillovers of R&D: Evidence from Firms’ Patents, Profits and Market Value. Working Paper 1815, National Bureau of Economic Research.
- Kemp-Benedict, E. (2014). An interpretation and critique of the method of reflections. Working paper, Munich Personal RePEc Archive.
- Kowalski, P. and Legendre, C. (2023). Raw materials critical for the green transition: Production, international trade and export restrictions. *OECD Trade Policy Papers*, 269.
- Kuznets, S. (1955). Economic growth and income inequality. *The American Economic Review*, 45(1):1–28.
- Lanjouw, J. O., Pakes, A., and Putnam, J. (1998). How to count patents and value intellectual property: The uses of patent renewal and application data. *The Journal of Industrial Economics*, 46(4):405–432.
- Laudati, D., Mariani, M. S., Pietronero, L., and Zaccaria, A. (2023). The different structure of economic ecosystems at the scales of companies and countries. *Journal of Physics: Complexity*, 4(2):025011.

References

- Lewis, A. (1954). Economic Development with Unlimited Supplies of Labour. *The Manchester School*, 22(8):139–191.
- Li, J.-F., Gu, A.-L., Ma, Z.-Y., Zhang, C.-L., and Sun, Z.-Q. (2019). Economic development, energy demand, and carbon emission prospects of china’s provinces during the 14th five-year plan period: Application of cmrcge model. *Advances in Climate Change Research*, 10(3):165–173.
- Lin, J. Y., Krueger, A., and Rodrik, D. (2011). New structural economics: A framework for rethinking development. *The World Bank Research Observer*, 26(2):193–223.
- Maraut, S., Dernis, H., Webb, C., Spiezia, V., and Guellec, D. (2008). The OECD REGPAT database: a presentation. OECD Science, Technology and Industry Working Papers 2008/02, OECD Publishing, Paris.
- Martins, J. M., Adebayo, T. S., Mata, M. N., Oladipupo, S. D., Adeshola, I., Ahmed, Z., and Correia, A. B. (2021). Modeling the relationship between economic complexity and environmental degradation: Evidence from top seven economic complexity countries. *Frontiers in Environmental Science*, 9.
- Marín, A. and Goya, D. (2021). Mining—the dark side of the energy transition. *Environmental Innovation and Societal Transitions*, 41:86–88. Celebrating a decade of EIST: What’s next for transition studies?
- Mazzilli, D., Mariani, M. S., Morone, F., and Patelli, A. (2024). Equivalence between the fitness-complexity and the sinkhorn-knopp algorithms. *Journal of Physics: Complexity*.
- Mealy, P., Farmer, J. D., and Teytelboym, A. (2018). A new interpretation of the economic complexity index. *Alexander, A New Interpretation of the Economic Complexity Index (February 4, 2018)*.
- Mealy, P., Farmer, J. D., and Teytelboym, A. (2019). Interpreting economic complexity. *Science advances*, 5(1):eaau1705.
- Mealy, P. and Teytelboym, A. (2022). Economic complexity and the green economy. *Research Policy*, 51(8):103948.
- Montiel-Hernández, M. G., Pérez-Hernández, C. C., and Salazar-Hernández, B. C. (2024). The intrinsic links of economic complexity with sustainability dimensions: A systematic review and agenda for future research. *Sustainability*, 16(1):391.

References

- Montresor, S. and Quatraro, F. (2020). Green technologies and Smart Specialisation Strategies: A European patent-based analysis of the intertwining of technological relatedness and key enabling technologies. *Regional Studies*, 54(10):1354–1365.
- Müller, V. P. and Eichhammer, W. (2023). Economic complexity of green hydrogen production technologies-a trade data-based analysis of country-specific industrial preconditions. *Renewable and Sustainable Energy Reviews*, 182:113304.
- Napolitano, L., Evangelou, E., Pugliese, E., Zeppini, P., and Room, G. (2018). Technology networks: The autocatalytic origins of innovation. *Royal Society Open Science*, 5(6):172445.
- Napolitano, L., Sbardella, A., Consoli, D., Barbieri, N., and Perruchas, F. (2022). Green innovation and income inequality: A complex system analysis. *Structural Change and Economic Dynamics*, 63:224–240.
- Neagu, O. (2019). The Link between Economic Complexity and Carbon Emissions in the European Union Countries: A Model Based on the Environmental Kuznets Curve (EKC) Approach. *Sustainability*, 11(17).
- Neagu, O. and Teodoru, M. C. (2019). The Relationship between Economic Complexity, Energy Consumption Structure and Greenhouse Gas Emission: Heterogeneous Panel Evidence from the EU Countries. *Sustainability*, 11(2).
- Nelson, R. R. and Winter, S. G. (1982). *An Evolutionary Theory of Economic Change*. The Belknap Press of Harvard University Press, Cambridge, Massachusetts and London, UK.
- Norgate, T. and Haque, N. (2010). Energy and greenhouse gas impacts of mining and mineral processing operations. *Journal of Cleaner Production*, 18(3):266–274.
- OECD (1999). The Environmental Goods and Services Industry: Manual for Data Collection and Analysis. OECD Publishing, Paris.
- OECD (2016). Patent search strategies for the identification of selected environment-related technologies (ENV-TECH). <https://www.oecd.org/environment/consumption-innovation/ENV-tech>
- OECD (2020). Territorial grids.
- Operti, F. G., Pugliese, E., Andrade Jr, J. S., Pietronero, L., and Gabrielli, A. (2018). Dynamics in the Fitness-Income plane: Brazilian states vs World countries. *PloS one*, 13(6):e0197616.

References

- Parkinson, C. L. (2010). *Coming climate crisis? Consider the past, beware the big fix*. Lanham: Rowman & Littlefield Publishers.
- Pata, U. K. (2021). Renewable and non-renewable energy consumption, economic complexity, CO2 emissions, and ecological footprint in the USA: testing the EKC hypothesis with a structural break. *Environmental Science and Pollution Research*, 28(1):846–861.
- Patelli, A., Napolitano, L., Cimini, G., and Gabrielli, A. (2023). Geography of science: Competitiveness and inequality. *Journal of Informetrics*, 17(1):101357.
- Patelli, A., Napolitano, L., Cimini, G., Pugliese, E., and Gabrielli, A. (2022). The evolution of competitiveness across economic, innovation and knowledge production activities.
- Penrose, E. T. (1959). *The Theory of the Growth of the Firm*. John Wiley, New York.
- Pérez-Hernández, C. C., Salazar-Hernández, B. C., Mendoza-Moheno, J., Cruz-Coria, E., and Hernández-Calzada, M. A. (2021). Mapping the green product-space in mexico: From capabilities to green opportunities. *Sustainability*, 13(2):945.
- Perruchas, F., Consoli, D., and Barbieri, N. (2020). Specialisation, diversification and the ladder of green technology development. *Research Policy*, 49(3):103922.
- Pietronero, L., Cristelli, M., Gabrielli, A., Mazzilli, D., Pugliese, E., Tacchella, A., and Zaccaria, A. (2017). Economic complexity: "buttarla in caciara" vs a constructive approach.
- Prebisch, R. (1950). *The Economic Development of Latin America and its Principal Problem*, *Economic Commission for Latin America*. Economic Commission for Latin America, United Nations Department of Economic Affairs, Lake Success, New York.
- Pugliese, E., Chiarotti, G. L., Zaccaria, A., and Pietronero, L. (2017). Complex economies have a lateral escape from the poverty trap. *PloS one*, 12(1):e0168540.
- Pugliese, E., Cimini, G., Patelli, A., Zaccaria, A., Pietronero, L., and Gabrielli, A. (2019a). Unfolding the innovation system for the development of countries: Coevolution of Science, Technology and Production. *Scientific reports*, 9(1):16440.
- Pugliese, E., Napolitano, L., Chinazzi, M., and Chiarotti, G. (2019b). The emergence of innovation complexity at different geographical and technological scales. *arXiv preprint arXiv:1909.05604*.

References

- Rafique, M. Z., Nadeem, A. M., Xia, W., Ikram, M., Shoaib, H. M., and Shahzad, U. (2022). Does economic complexity matter for environmental sustainability? Using ecological footprint as an indicator. *Environment, Development and Sustainability*, 24:4623–4640.
- Ramahandry, T., Bonneau, V., Bani, E., Vlasov, N., Flickenschild, M., Batura, O., Tcholtchev, N., Lämmel, P., and Boerger, M. (2021). Key enabling technologies for europe’s technological sovereignty. *Brussels: European Parliamentary Research Service*.
- Raworth, K. (2017). *Doughnut economics: Seven ways to think like a 21st-century economist*. Chelsea Green Publishing.
- Romare, M. and Dahllöf, L. (2017). The life cycle energy consumption and greenhouse gas emissions from lithium-ion batteries.
- Romero, J. P. and Gramkow, C. (2021). Economic complexity and greenhouse gas emissions. *World Development*, 139:105317.
- Santoalha, A. and Boschma, R. (2021). Diversifying in green technologies in European regions: Does political support matter? *Regional Studies*, 55(2):182–195.
- Santoalha, A., Consoli, D., and Castellacci, F. (2021). Digital skills, relatedness and green diversification: A study of european regions. *Research Policy*, 50(9):104340.
- Saracco, F., Di Clemente, R., Gabrielli, A., and Squartini, T. (2015). Randomizing bipartite networks: The case of the World Trade Web. *Scientific Reports*, 5.
- Saracco, F., Straka, M. J., Di Clemente, R., Gabrielli, A., Caldarelli, G., and Squartini, T. (2017). Inferring monopartite projections of bipartite networks: an entropy-based approach. *New Journal of Physics*, 19.
- Sarewitz, D. and Nelson, R. (2008). Three rules for technological fixes. *Nature*, 456(7224):871.
- Sauvage, J. (2014). The stringency of environmental regulations and trade in environmental goods. OECD Trade and Environment Working Papers 2014/03, OECD Publishing, Paris.
- Savona, M. and Ciarli, T. (2019). Structural Changes and Sustainability. A Selected Review of the Empirical Evidence. *Ecological Economics*, 159:244–260.
- Sbardella, A., Barbieri, N., Consoli, D., Napolitano, L., Perruchas, F., and Pugliese,

References

- E. (2022). The regional green potential of the European innovation system. No. JRC124696, Joint Research Centre (Seville site).
- Sbardella, A., Perruchas, F., Napolitano, L., Barbieri, N., and Consoli, D. (2018a). Green technology fitness. *Entropy*, 20(10):776.
- Sbardella, A., Pugliese, E., and Pietronero, L. (2017). Economic development and wage inequality: A complex system analysis. *PloS one*, 12(9):e0182774.
- Sbardella, A., Pugliese, E., Zaccaria, A., and Scaramozzino, P. (2018b). The role of complex analysis in modelling economic growth. *Entropy*, 20(11):883.
- Selden, T. M. and Song, D. (1994). Environmental quality and development: Is there a kuznets curve for air pollution emissions? *Journal of Environmental Economics and Management*, 27(2):147–162.
- Sovacool, B. K., Ali, S. H., Bazilian, M., Radley, B., Nemery, B., Okatz, J., and Mulvaney, D. (2020). Sustainable minerals and metals for a low-carbon future. *Science*, 367(6473):30–33.
- Steenblik, R. P. (2020). Code shift: The environmental significance of the 2022 amendments to the harmonized system. Working paper, International Institute for Sustainable Development.
- Stern, D. I. (2004). The rise and fall of the environmental kuznets curve. *World Development*, 32(8):1419–1439.
- Stern, D. I. (2017). The environmental Kuznets curve after 25 years. *Journal of Bioeconomics*, 19:7–28.
- Stojkoski, V., Koch, P., and Hidalgo, C. A. (2023). Multidimensional economic complexity and inclusive green growth. *Communications Earth & Environment*, 4(1):130.
- Swart, J. and Brinkmann, L. (2020). *"Economic Complexity and the Environment: Evidence from Brazil"*, pages 3–45. Springer International Publishing.
- Tacchella, A., Cristelli, M., Caldarelli, G., Gabrielli, A., and Pietronero, L. (2012). A new metrics for countries' fitness and products' complexity. *Scientific reports*, 2(1):723.
- Tacchella, A., Mazzilli, D., and Pietronero, L. (2018). A dynamical systems approach to gross domestic product forecasting. *Nature Physics*, 14(8):861–865.
- Tacchella, A., Zaccaria, A., Miccheli, M., and Pietronero, L. (2023). Relatedness in the era of machine learning. *Chaos, Solitons & Fractals*, 176:114071.

References

- Teece, D. J., Pisano, G., and Shuen, A. (1997). Dynamic capabilities and strategic management. *Strategic Management Journal*, 18(7):509–533.
- Teece, D. J. and Pisano, G. P. (1994). The Dynamic Capabilities of Firms: an Introduction. *Industrial and Corporate Change*, 3:537–556.
- Teece, D. J., Rumelt, R., Dosi, G., and Winter, S. (1994). Understanding corporate coherence: Theory and evidence. *Journal of Economic Behavior & Organization*, 23(1):1–30.
- Veefkind, V., Hurtado-Albir, J., Angelucci, S., Karachalios, K., and Thumm, N. (2012). A new EPO classification scheme for climate change mitigation technologies. *World Patent Information*, 34(2):106–111.
- Waltman, L. (2016). A review of the literature on citation impact indicators. *Journal of Informetrics*, 10:365–391.
- Wanger, T. C. (2011). The lithium future-resources, recycling, and the environment. *Conservation Letters*, 4(3):202–206.
- WTO (2009). Committee on trade and environment in special session.
- Yang, R. and Wang, W. (2023). Potential of china’s national policies on reducing carbon emissions from coal-fired power plants in the period of the 14th five-year plan. *Heliyon*, 9(9).
- Yilanci, V. and Pata, U. K. (2020). Investigating the ekc hypothesis for china: the role of economic complexity on ecological footprint. *Environmental Science and Pollution Research*, 27(26):32683–32694.
- You, W., Zhang, Y., and Lee, C.-C. (2022). The dynamic impact of economic growth and economic complexity on CO2 emissions: An advanced panel data estimation. *Economic Analysis and Policy*, 73:112–128.
- Zaccaria, A., Cristelli, M., Tacchella, A., and Pietronero, L. (2014). How the taxonomy of products drives the economic development of countries. *PloS one*, 9(12):e113770.
- Zheng, F., Zhou, X., Rahat, B., and Rubbaniy, G. (2021). Carbon neutrality target for leading exporting countries: On the role of economic complexity index and renewable energy electricity. *Journal of Environmental Management*, 299:113558.

Appendix

Appendix A. Patent data for analysing climate change mitigation and adaptation technologies

In the following, we provide additional information on patent data and their potential use for analysing climate change mitigation and adaptation technologies.

As mentioned in the paper, the two main data sources on countries' patenting activity are PATSTAT and REGPAT, on which we offer additional details in the following, especially aimed at a comparison between these two data sources.

In fact, as can be appreciated in Figure A1, PATSTAT and REGPAT present significant differences in their temporal and geographical coverage. The large difference in the number of documents is due to the fact that REGPAT records patents filed only at the EPO, while PATSTAT collects information from most patent offices worldwide. This makes REGPAT less suitable for the analysis of, for instance, smaller countries. However, the EPO tends to receive high-quality applications, making data collected from it more reliable.

While REGPAT is published annually by the OECD and covers the subset of PATSTAT patent applications filed only at the EPO since its inception in 1978, PATSTAT has been published biannually by the European Patent Office (EPO) since 2007 and has grown substantially in coverage over time. As of the latest editions, it records information on more than 100 million patent applications filed since the late eighteenth century, which are collected in over 50 million families. Whenever the information is available for a patent application, PATSTAT records, among other things, the receiving patent office, the filing date, the technologies in which the patent innovates (encoded in standard technology codes), and the residence of the applicants and of the inventors at the time of filing. The geographical information is incomplete, with the coverage varying widely across patent offices. The most commonly available information in this sense is the country of residence of inventors and applicants.

Since, for geographical reasons, the EPO attracts disproportionately more European patent applications, it does not provide a uniform geographical coverage beyond European borders. Furthermore, filing costs at the EPO are higher than in most national offices. This skews the sample towards "high-value" patents and, indirectly, leads to an over-representation of richer European countries. However, REGPAT makes up for this shortcoming thanks to an accurate geocoding across over

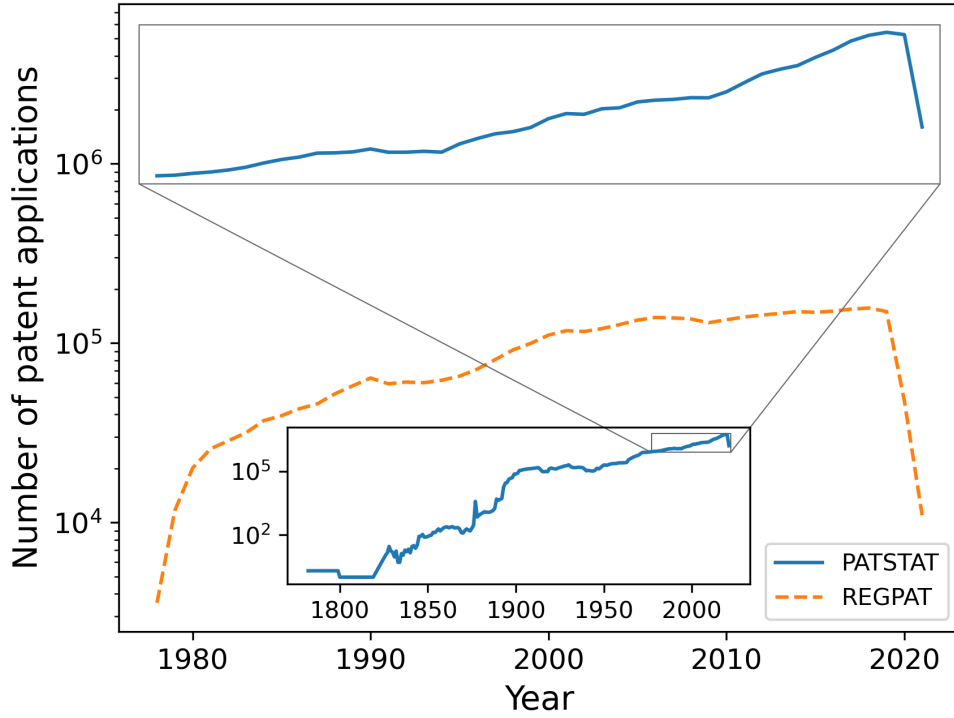


Figure A1: Comparison of the coverage of PATSTAT (blue) and REGPAT (orange). Main plot: time series of the number of patent applications recorded in PATSTAT and REGPAT over the period covered by both databases (1978-2022). Inset: time series of the number of patent applications recorded in PATSTAT over the entire period it covers.

5500 sub-national regions of the patent documents filed by applicants or inventors residing in one of over 40 countries belonging to the OECD, the EU, the UK, Brazil, China, India, the Russian Federation, and South Africa.¹⁸

Concerning the temporal coverage, Figure A1 shows that REGPAT and PATSTAT differ substantially. However, it also suggests that both databases cover a long enough time window for all practical purposes. The coverage of REGPAT starts only in 1978, while PATSTAT records US and UK patents dating back as far as the

¹⁸ The brochure of an older version of REGPAT, available at http://econ.geo.uu.nl/crespo/tech_div/regpat_201602.pdf, provides more details on the structure of the geographical coverage

late eighteenth century. Nevertheless, large numbers of patents have been recorded only in recent decades. Moreover, both databases are updated to virtually the same date.

It is worth noting that, due to the dynamics of patent offices and the regulations governing patent filing, there is an intrinsic lag of 12-18 months between when an application is received by a patent office and when the corresponding record appears in the database, at the end of the so-called search phase by the patent office. During the search phase, patent offices conduct a primary assessment of the originality of the invention and inform the applicants, who can decide to either withdraw the application thus keeping it confidential, or pursue the grant of a patent and disclose their application to the public. Therefore, patent counts extracted from an edition of PATSTAT or REGPAT published in 2022 are not reliable beyond 2018 or 2019.

An invention can be submitted to different offices — e.g. to cover different geographical regions – by filing different applications at different points in time. The first patent application filed to protect an invention is called the priority application. Subsequent applications that are related to the same invention name the same applications as priorities, allowing to group into the same patent family (there are over 50 million patent families in PATSTAT). Families are a useful way to group together documents referring to the same innovation and are thus frequently used as the basic unit of observation in empirical exercises.

As mentioned above, PATSTAT provides limited information about the geographical location of inventors and applicants for many documents, even though the database records patents filed by applicants and inventors located in more than 200 countries. Instead, REGPAT offers much more detailed information, albeit on a smaller set of countries, by associating patent documents to the OECD Territorial Level (TL) (OECD, 2020) code of the region of residence of applicants and inventors. For European countries the TL follows the hierarchical structure of the 2013 edition of the Nomenclature of Territorial Units for Statistics¹⁹) (NUTS), developed by Eurostat. Instead, in the US, progressively finer levels of the classification identify states, economic areas and counties, that do not follow a nested structure.

To attribute a geographical location to patents one can leverage the inventors' as well as the applicants' residence information. The former is often preferred as a

¹⁹ <https://ec.europa.eu/eurostat/web/gisco/geodata/reference-data/administrative-units-statistical-units/nuts>

proxy for the location of inventive capabilities because it is assumed that inventors (who are always physical persons) tend to live close to where they perform their duties. On the contrary, applicants (who are in many cases companies) may choose to assign to the corporate headquarters a patent that was developed in a subsidiary located in a different country or region for business-related reasons.

Multiple inventors or applicants can be linked to the same patent application or family. In such cases, one may choose between counting the patent fractionally and counting patents in full, when it comes to assigning patents to geographical areas. Full counting counts as a unit of every pair of patent documents and territorial units hosting at least one inventor. As a consequence of this double counting, the weight attributed to a patent (and even more for a family) depends on the number of inventors. Instead, in fractional counting, each patent (or each family) sums to 1, and each territorial unit having an inventor gets a fractional weight inversely proportional to the total number of inventors. There is some debate in the literature concerning the best approach (Waltman, 2016). However, in economic complexity applications, fractional counting is generally the preferred approach.

Appendix B. Identification of green technologies

However, searching for environment-related patent documents was not straightforward at the beginning because a dedicated classification system for sustainable technologies was not available. In fact, before 2011, no specific branch of the IPC or of other technology classification systems covered environment-related inventions.

A first step in this direction was the creation in 2011 of the Y02 class the EPO, in cooperation with the United Nations Environmental Program (UNEP) and the International Centre on Trade and Sustainable Development (ICTSD), to complement the IPC classification System. From the beginning, the purpose of the Y02 class was to tag CCMT patent documents by means of search strategies and algorithms implemented by expert examiners and that can be re-run periodically to update the classes (Veefkind et al., 2012). The Y02 scheme initially covered only patent documents related to CCMTs in the energy sector and was later extended also to other types of mitigation technologies. This effort by the EPO constituted a major advancement for the study of green innovation both from an academic and policy perspective, as it has allowed also non-specialists to easily identify CCMTs.

In 2013, the European Patent Office and the United States Patent and

Trademark Office (USPTO) agreed to harmonize their patent classification practices and developed the Cooperative Patent Classification (CPC). Since then, the CPC has become increasingly popular as a classification standard and has been complementing or substituting the IPC in a growing number of patent offices worldwide. As illustrated in Table 1, two types of codes can be found in the CPC classification: codes starting with the letters A to H – similar to IPC codes and representing the traditional classification of technologies; and codes starting with Y, which are used to tag cross-sectional technologies that are already indexed somewhere else in the classification²⁰ Therefore, the Y classification scheme is used to tag patent documents that are already classified or indexed somewhere else in the classification. In addition to the Y02 class, the new subclass Y04S dedicated to smart grids was integrated in the CPC section Y. As shown in Table 2, the Y02 class consists of more than 1000 tags related to sustainable technologies organized in 9 sub-classes.

With the aim of maximizing the informative content on eco-innovation present both in the IPC and CPC classifications, in 2015 the OECD (Haščič and Migotto, 2015) developed ENV-TECH, an expert-based catalogue of environment-related technologies based on the IPC classification and lastly updated in 2016 (OECD, 2016), which can be used to tag green patent documents in PATSTAT or other patent data-sets. ENV-TECH identifies 94 environment-related technology areas that group 4- to 16-digit IPC and CPC codes, building on the CPC Y02 class whenever possible. The catalogue relies on a keyword search strategy identifying patent documents that correspond to each "target environmental technology field" (Haščič and Migotto, 2015, p.19) covering CCMTs but also environmental management and water-related adaptation technologies (class 1 and 2). When it is not possible to identify single IPC/CPC classes that portray alone the technological field of interest, it employs a combination of different IPC/CPC patent classes, also at different aggregation levels.

While widening the object of analysis is undoubtedly commendable, unfortunately, it makes the catalogue very vulnerable to patent document re-classification. In fact, while the additional information may have been useful for the researchers working with versions of PATSTAT prior or contemporaneous to 2016, the year up to which ENV-TECH is updated, using a combination of different IPC and CPC codes may constitute a drawback when working with more recent versions of PATSTAT. This is because the classification of the codes comprising any technology

²⁰ <https://www.uspto.gov/web/patents/classification/cpc/html/cpc-Y.html>

– especially for finer-grained codes – changes over time, possibly reclassifying past inventions into new codes. Furthermore, being fixed in time, ENV-TECH not only it suffers from the reclassification of previous inventions, but it also fails to consider newer patent applications. In fact, technology codes are revised at least once a year to take into account new technical advances, as well as to improve the search for the prior art that the patent officers use to establish the application’s degree of innovation. Moreover, the Y04S subclass covering smart grids, which at the end of 2016 comprised 54000 patent documents (Angelucci et al., 2018), is not included in the ENV-TECH catalogue. By contrast, the Y02/Y04S codes are part of the CPC section Y and are therefore robust to changes in the CPC classification, are more user-friendly and readily usable for the relevant data can be directly extracted by PATSTAT with no need for intermediate steps. Lastly, another limitation of ENV-TECH is that, due to its limited granularity, it does not allow very detailed studies, an obstacle for carrying out economic complexity analyses that usually rely on very fine-grained information.

Appendix C. List of search keywords

The following table contains the list of keywords used to identify candidate papers to include in the literature. The last three columns in the table broadly classify keywords based on the kind of information they refer to within the papers (respectively the data, analytical methodologies, and topics covered by the papers). We used keywords as search terms in scholarly databases and search engines in isolation or, whenever appropriate, in combination (e.g. a *data* and a *method* and a *topic* keyword in the same search).

Appendix

Table C1: List of keywords used to identify candidate papers to include in the literature review

Keyword	Data	Methods	Topic
Climate change mitigation and adaptation technologies (CCMTs)	x		
Cooperative patent classification (CPC)	x		
Env-tech	x		
Green products	x		
Green technologies	x		
International patent classification (IPC)	x		
Patent data	x		
Y-codes (Y02-Y04S)	x		
Bipartite configuration model (BiCM)		x	
Economic complexity		x	
Economic complexity index (ECI)		x	
Economic fitness		x	
Economic fitness and complexity (EFC)		x	
Green complexity index		x	
Green technology fitness		x	
Multilayer networks		x	
Product progression network		x	
Product space		x	
Proximity		x	
Relatedness		x	
Carbon footprint			x
CO ₂			x
Emissions			x
Evolutionary economic geography			x
Green deal			x
Green economy			x
Green innovative capacity			x
Green technology life cycle			x
Greenhouse gas (GHG)			x
Sustainability			x
Sustainable transition			x

GETTING IN TOUCH WITH THE EU

In person

All over the European Union there are hundreds of Europe Direct information centres. You can find the address of the centre nearest you at: https://europa.eu/european-union/contact_en

On the phone or by email

Europe Direct is a service that answers your questions about the European Union. You can contact this service:

- by freephone: 00 800 6 7 8 9 10 11 (certain operators may charge for these calls),
- at the following standard number: +32 22999696, or
- by electronic mail via: https://europa.eu/european-union/contact_en

FINDING INFORMATION ABOUT THE EU

Online

Information about the European Union in all the official languages of the EU is available on the Europa website at: https://europa.eu/european-union/index_en

EU publications

You can download or order free and priced EU publications from EU Bookshop at:

<https://publications.europa.eu/en/publications>. Multiple copies of free publications may be obtained by contacting Europe Direct or your local information centre (see https://europa.eu/european-union/contact_en).

Science for policy

The Joint Research Centre (JRC) provides independent, evidence-based knowledge and science, supporting EU policies to positively impact society



EU Science Hub

joint-research-centre.ec.europa.eu