LETTER • OPEN ACCESS

Leveraging patent analysis to measure relatedness between technology domains: an application on offshore wind energy

To cite this article: Yiwen Wang *et al* 2024 *Environ. Res. Lett.* **19** 024045

View the article online for updates and enhancements.

ENVIRONMENTAL RESEARCH
LETTERS

LETTER

Leveraging patent analysis to measure relatedness between technology domains: an application on offshore wind energy

Yiwen Wang[1,*], **Erin Baker**[1] **and Anna Goldstein**[2]

[1] Department of Mechanical and Industrial Engineering, University of Massachusetts Amherst, Amherst, MA 01003, United States of America
[2] Director of Impact, Prime Coalition, Cambridge, MA 02139, United States of America
* Author to whom any correspondence should be addressed.

E-mail: yiwwang@umass.edu

Abstract

As the global energy sector transitions towards a cleaner and more sustainable future, observational evidence suggests that many new energy technologies share a close relationship with well-established technologies. Yet, the topic of how closely technologies are related has not been addressed rigorously, rather it has been the purview of practitioner know-how and informal expert opinion. In this study, we propose a quantitative method to supplement practitioners' subjective understanding of the relatedness between technology domains. The method uses patents to represent the position of a technology in knowledge space and calculates the Hausdorff distance between patent domains to proxy the relatedness between technologies. We apply this method to investigate the relatedness of offshore wind energy technology to two more mature domains: onshore wind energy technology and offshore oil and gas technology. We examine the technological relatedness of individual offshore wind components to these two technologies, and represent the changes in relatedness through time. The results confirm that offshore wind components such as foundations, installation, and maintenance are more related to the offshore oil and gas industry; while other components, such as rotors and nacelles, are more related to onshore wind energy. The results also suggest that many offshore wind energy components are becoming less related through time to both of these domains, possibly indicating increasing innovation. This method can provide quantitative parameters to improve the modeling of technological change and guide practitioners in strategic decision-making regarding the positioning of industries and firms within those industries.

1. Introduction

Advances in technologies will critically influence climate mitigation costs (Popp 2005, Rao *et al* 2006, Sue Wing 2006, Weyant 2017). Understanding the innovation path of an emerging technology is a significant input to models and decision making. Yet, modeling innovation paths is rife with uncertainties, including assumptions about technology novelty. Ferioli *et al* (2009) showed that technological forecasts derived from cost observations are sensitive to assumptions about which components are more or less mature, such that the same historical data can produce very different cost forecasts depending on assumptions that are used. Technology novelty is especially hard

to define for emerging technologies with components that are strongly related to more mature ones; yet this is a common occurrence, especially during the energy transition, where many emerging energy technologies exhibit a close relationship with well-established ones. For example, offshore wind clearly has similarities with onshore wind; electric vehicles use similar mechanical technologies as used by internal combustion vehicles; carbon capture and storage and hydrogen production both share common technologies with oil and gas for gas processing and refinery. However, this relatedness is often ignored when modeling; and discussions tend to rely on practitioner know-how and informal expert opinion. For example, offshore wind clearly has similarities to

onshore wind, yet it is often considered a novel technology (van der Zwaan *et al* 2012, Haas *et al* 2022).

A few previous patent analyses have discussed relatedness when studying the innovation path of technology. Joo and Kim (2010) measured and visualized relatedness among broad technological fields, such as 'agriculture' and 'semiconductors', but did not get down to the individual technology level. Other studies measured the relatedness of a technology to a knowledge base of a given region (Tanner 2016, Li *et al* 2020, Moreno and Ocampo-Corrales 2022). For example, Tanner (2016) compared the relatedness of fuel cell technology to 12 broad technological fields with European samples. Some studies measured relatedness between firms and institutions to identify renewable energy adoption opportunities enabled by spillovers (Breschi *et al* 2003, Punt *et al* 2022). Nakamura *et al* (2015) measures the relatedness between industries, such as automobile and aircraft, to identify future and past technological breakthroughs. Our study differs in that we measure the relatedness between individual technologies. We focus on technologies that are observationally considered related; and investigate how the relatedness to forebears have evolved through time, which can be important information for projecting the trajectory of emerging technologies.

We present a method that calculates technological relatedness using patent data, specifically classification codes, to quantitatively measure relatedness of technologies by the similarity of their patents. There are several advantages of using patent data for understanding emerging and complex technologies. Patent data can capture the technical characteristics as patents represent state-of-the-art inventions as authorized documents for intelligence protection. Patents are widely used as an indicator for invention activities (Braun *et al* 2010, Popp *et al* 2013, Yu 2017). Patents are required to be specific and descriptive and to be examined by experts. Analyzing technologies by patent data adds a dimension of analysis that goes beyond individual expertise about a technical domain. Patents are public data available worldwide; thus the same analysis can be done in different regions and across time. Patent data can enhance the modeling of emerging technologies that have insufficient installation or cost data.

In previous studies, patent similarity has been measured using backward citations, text, or classification codes (Rodriguez *et al* 2015, Aharonson *et al* 2016, Verhoeven *et al* 2016, Yan and Luo 2017, Arts *et al* 2018). We use classification codes to measure

relatedness of patents in this study because we are interested in the technological relatedness of patents revealed by the classification system as a technology-based hierarchical taxonomy (Jaffe 1986). Backward citations work at a higher level, relating the linkage with pre-existing full patents. For example, Rodriguez *et al* (2015) built a citation network in which each node represents a patent and each link represents direct citations. Text similarity can potentially capture more technical details; however, the accuracy of text similarity depends on researcher-identified spelling variants, synonyms and abbreviations of specialized terms (Yoon and Kim 2012, Arts *et al* 2018). As a result, full text similarity requires expertise in a narrow field and is not usually deployed across multiple technology domains (Aharonson *et al* 2016).

We apply a novel measure, the Hausdorff distance, to patent analysis. Hausdorff distance is a metric for measuring the distance between two sets. We use it to measure the distance between technology domains. The Hausdorff distance is a max-min definition: it is defined as the longest of the distances between each point in one set to its closest partner in the other set. Hausdorff distance has not been applied to measure the relatedness or similarity of patents. It has been applied to compare the similarity of graphs in patents, but this application is more about image processing, rather than focusing on the specifics of patents and technological change (Mogharrebi *et al* 2013).

Offshore wind is of particular interest as it has a high potential to contribute to mitigation efforts (Cranmer and Baker 2020, Kanyako and Baker 2021); and a clear relationship to more mature technology. The offshore wind industry was originally developed with experience from onshore wind energy as well as the offshore oil and gas industry, in particular work on platforms in this industry. The historical relationship between offshore wind, onshore wind and offshore oil and gas are noted in the literature. However, the specific relationships are implied and qualitative, rather than formal and quantitative. Moreover, from informal discussions with experts, there is some uncertainty about the relative relatedness of offshore wind with onshore wind and offshore oil and gas. We aim to go beyond the sense of relatedness, by applying a quantitative method to test the degree to which this sense holds. In this study, we use a data-driven method to quantify the relatedness between offshore wind and the other two industries to investigate the observations around this relation. We then extend the discussion to a more detailed level and explored the relatedness of offshore wind components to the other industries.

2. Methods

Our method uses patent data to estimate a technology's position in knowledge space and proxy its relatedness to other technologies by their position similarity. The steps in the methodology are: (1) define patent domains for each technology by a keyword-code-combined query approach; (2) define the position of a patent family by creating position vectors based on classification codes for each patent family; (3) calculate pairwise distances between patent families; (4) calculate the Hausdorff relatedness between sets of patent families. We compare relatedness using classification codes of patent families, rather than individual patents. A patent family is a group of patent applications for an identical invention that has been filed in different countries or regions. Using patent families avoids duplicates when analyzing the relatedness between inventions or technologies. It also reduces missingness if patents from some patent offices do not have CPC information disclosed. In addition, considering the full patent family, rather than just the priority application, allows us to include inventions in different languages.

The knowledge space has been used in patent analyses (Rigby 2015, Whittle 2020) to represent a conceptual framework where knowledge associated with all technologies, concepts, and skills is assumed to have a position that represents its technological composition or ontology. The distance in knowledge space symbolizes the cognitive proximity between two different pieces of knowledge. Knowledge in the same domain is considered to be closer together, with a higher proximity, while knowledge from distinct domains are further apart with lower proximity.

We define a vector space to represent this conceptual knowledge space. Our vector space is defined such that each patent family is represented by a vector, and each element in the vector is related to a classification code. Specifically, each element represents the fraction of the classification codes for that patent family that falls into the specific subgroup. The distance between patent families is the Euclidean distance between vectors, normalized to the scale of 0–1. Two patent families with the exact same vectors have a distance of 0. A patent family with all codes under one code subgroup has a distance of 1 with another patent family that has all codes under a different subgroup.

We use the Hausdorff distance as it captures the characteristics of domain-to-domain relatedness that conventional methods may not. Unlike methods that average pairwise distance to estimate domain relatedness, a Hausdorff distance of X (implying relatedness of 1-X) implies that any patent family in one set can find at least one patent family in the other set with a distance not greater than X. From another perspective, each patent family in the first set has at least one patent family in the other set with a relatedness greater than or equal to 1-X. We note that the Hausdorff distance between two patent domains will not necessarily change with the addition of new patents to either set; thus a change in the Hausdorff distance may be a signal of significant and transformative changes. To our knowledge, the Hausdorff method has never been applied to patent analysis.
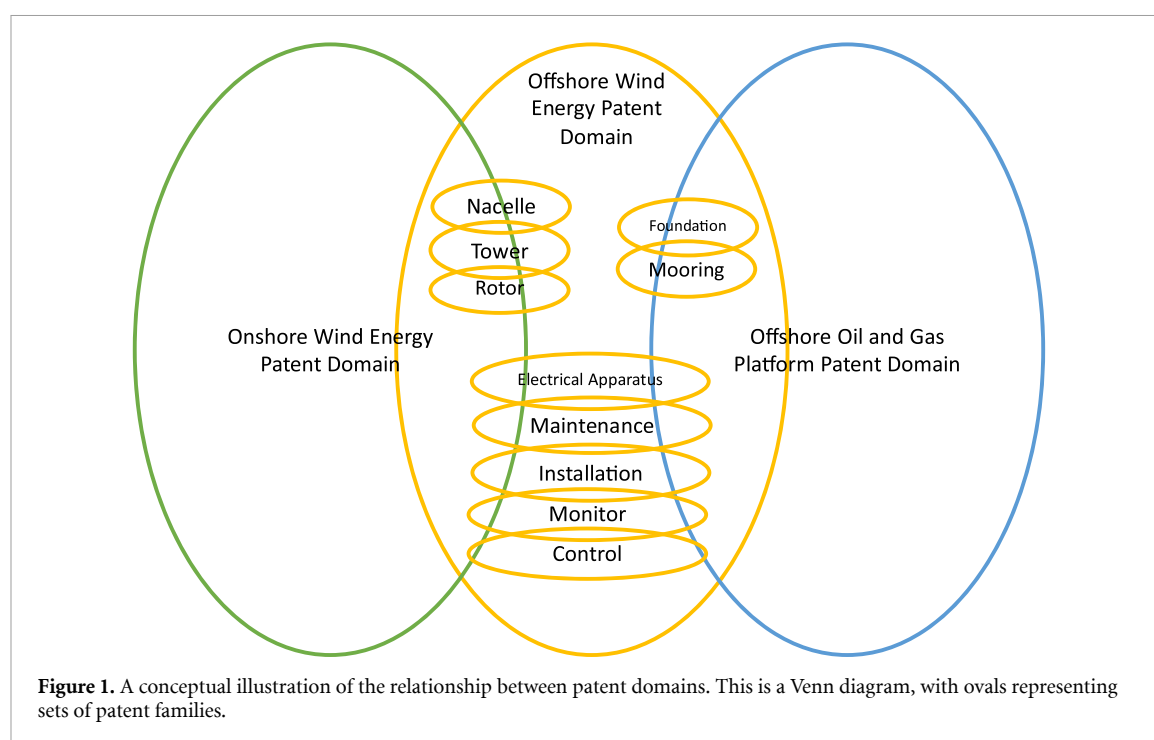
We focus on the relatedness of individual offshore wind components with the entire domains of onshore wind and offshore oil and gas platforms, rather than using component subdomains in the comparison industries, for two reasons. First, the process of identifying similar components in a second domain introduces errors. It is ambiguous to a priori determine exactly what the similar component is in another domain for every component we want to assess. Second, there may be technical relatedness between two patent families even though they are directly related to different components. To investigate the impact of this methodological choice, we repeated the analysis using component subdomains in onshore wind and offshore oil and gas platforms, using four components as an example. The subdomain is retrieved by the same definitions as used to define the offshore wind component. The results are very close (appendix G).

Patent data for this study is sourced from PATSTAT Global, a downloadable database of patent applications sourced from dozens of patenting authorities worldwide. PATSTAT contains the title, abstract, assignee(s), classifications, and other key attributes of over 100 million patent records.

2.1. Define patent domains
A preliminary search of the Espacenet online database yields 657 981 patent applications with 'wind' in the title or abstract, and 119 998 applications classified under 'F03D wind motors'; and 137 467 patent applications with 'offshore oil' or 'offshore gas' in text fields, and 45 827 patent applications with 'offshore oil platform' or 'offshore gas platform' in text fields. However, Fung *et al* (2023) showed that simple searches using keywords alone or codes alone are not sufficient for defining clear technology domains. In this section we build on Fung *et al* (2023), using an iterative method to define technology domains that have better coverage and are more precise than the domains resulting from simple searches.

We start with the wind energy domain defined by Fung *et al* (2023). This paper presented an iterative method for defining technology domains, using combinations of classification codes, keywords, and expert review. The paper expanded the domain for

Figure 1. A conceptual illustration of the relationship between patent domains. This is a Venn diagram, with ovals representing sets of patent families.

wind energy patents by 7.5%, providing a 6% increase in recall, and only a 1% decrease in precision.

Here, we define partially overlapping domains for offshore wind energy, onshore wind energy, and offshore oil and gas platforms, and then go on to define sub-domains for a selection of components in the offshore wind energy domain. Based on discussions with experts in the fields of onshore and offshore wind, as well as the visual similarity, we started with the anecdotal opinions that the 'above the water' portions of the technology, such as nacelle, tower, and rotors are highly related to the same components used for onshore wind energy; while the 'below the water' components are related to the platforms used for offshore oil and gas; and finally technology electrical apparatus, maintenance, installation, monitoring and control techniques might plausibly be related to both. We therefore consider onshore wind energy and offshore oil and gas platforms as the two key mature technology forbears in the innovation path of offshore wind energy.

Figure 1 uses a Venn-diagram to conceptually illustrate the relationship between the three industry-level patent domains as well as a potential positioning of the sub-domains for components. In figure 1, each large colored oval represents an industry-level technology patent domain, where a domain is a set of patent families. The smaller ovals represent sub-domains for illustrative components within the offshore wind energy patent domain. An overlap indicates that there may be patent families that belong to both domains. For example, there may be particular

patent families for nacelles can be applied to both onshore and offshore wind. There may also be particular patent families for nacelles that can be applicable to towers and rotors.

Within the expanded wind energy patent domain defined in Fung *et al* (2023), we define patent sub-domains for offshore and onshore wind using a keyword-code-combined query approach. We distinguish offshore-only and onshore-only wind domains using keywords and classification codes specific to each industry. The remaining patent families in the wind energy patent domain are patent families that can be applied to both. Tsai *et al* (2016) also identifies offshore-only wind energy patents, but used only keywords. We start by using a set of keywords and codes specific to offshore wind, such as 'offshore', 'ocean,' and the code 'Y02E 10/727', which refers to offshore wind turbines. We use these to identify patent families that are related to offshore wind only, but not to onshore wind. For onshore wind energy we use keywords such as 'onshore', 'land' and the code 'Y02E 10/728', which refers to onshore wind turbines. This categorization was validated by expert reviews (table B1). The complete sets of keywords and codes used to distinguish offshore and onshore wind patent families are listed in table 1. Detailed queries are provided in table A1.

For offshore oil and gas, we focus on innovations related to offshore oil and gas platforms, in order to reduce computational complexity, since this domain is observationally the closest technology to offshore wind energy. This simplification will not impact the

Table 1. Keywords and classification codes for identifying offshore-only and onshore-only wind energy patents out of the general wind energy patent domain; and for retrieving offshore oil and gas platform patents.

	Codes	Keywords
Offshore-only wind energy	B63H, B63B, E02B 17/027, E02D 15/06, E02D 27/52, E02D 29/06, E21B 7/12, F03D 13/25, F05B2240/95, Y02E 10/727	Offshore, ocean, lake, marine, aquatic anchor, float, undersea, at sea
Onshore-only wind energy	F03D 9/48, Y02E 10/728	Onshore, tree, mountain, road, land
Offshore oil and gas platforms	B63B 35/00, B63B 35/003, B63B 35/44, B63B 35/4413, B63B 77, E02B 17, E02B2017, E21B	Offshore, platform

relatedness analysis because we are using a one-sided Hausdorff distance measure, which finds the closest distance from items in offshore wind to items in oil and gas, thus excluding the less-relevant patent families will not change the distance measure. The patent domain for offshore oil and gas platforms is defined in a similar manner to the above, using a keyword-code-combined query approach. Relevant keywords and codes, as listed in table 1, are selected through patent searching and literature review. Detailed queries are provided in table A1. We did not perform expert validation for offshore oil and gas platform patents.

Next, we identify the patent sub-domains for offshore wind energy components within the offshore wind patent domain. Component domains are subsets of the patent domain of a technology. We combine keywords and classification codes that are exclusive to a component to define a component domain. Patent families related to a component can be retrieved by a code specific to the component. For example, 'F03 7/00' is a code defined as 'controlling wind motors'. We use it to retrieve patent families related to control techniques within the offshore wind patent domain. If no such code exists for a component, we find the most relevant codes and use exclusive keywords to refine the search results. For example, 'F03D 13/10' and 'F03D 13/20' are codes for 'arrangements for erecting, mounting, or supporting wind motors, including towers'; and 'E04H 12/00' refers to towers of all kinds. In addition to these codes, we add the keyword 'tower' to refine the results from these three codes as our component domain for towers. In table 2, we list the combinations of codes and keywords we use for each component.

The table is formatted such that codes and keywords that are in the same column are combined with an 'AND' Boolean operator; whereas codes in the same row are combined with an 'OR' Boolean operator. As illustrated by figure 1, component domains will have overlaps when a patent family involves innovation on more than one component. We will then measure the distance from individual offshore components to the entire domains of onshore wind and offshore oil and gas platforms.

2.2. Defining the position of a patent family

The position of patent families in the patent domain can be defined by classification codes. Classification codes are identifiers assigned to patents that reflect the technological fields that the patent is related to (Tanner 2016). Classification systems are hierarchical schemes of classification codes that have descriptions of technological fields with different levels of details. cooperative patent classification (CPC) and international patent classification (IPC) are two of the most widely adopted hierarchical classification systems. CPC is based on IPC and provides additional details. We use both CPC and IPC for the retrieval process to reduce missingness. We use only CPC for defining the position of a patent family for consistency. Using classification codes to calculate patent similarity is highly relevant to discussions of innovation path, as it implies how a patent relates to the basis of 'prior art' (Jaffe 1986). Patent classifications have been used previously in this way, as coordinates identifying a patent's position in a technological landscape, in order to measure the relative distance/similarity between patents (Jaffe 1989, Aharonson *et al* 2016).

We categorize the classification codes of patent families at the subgroup level, specifically the subgroup representing the eighth level in the CPC hierarchy. By using CPC classification, our intention is to measure the relatedness between domains under an established technical framework. The subgroup levels of CPC hierarchy are meticulously defined with detailed technical information.

We define a position vector for each patent family that represents the proportion of classification codes at this subgroup level. As a first step, we map all codes to this eighth level, using CPC scheme obtained as bulk data in XML format from the version of February 2023[3]. The elements in the position vector represent the count of classification codes within each subgroup; the vector is normalized so that the sum of all vector elements is one. Equation (1) shows a typical position vector, P_i, for patent family i. Each element c_{ik} in the vector represents how

---

Table 2. Keywords and classification codes for retrieving offshore wind energy components.

Components	Combinations of codes and keywords		Description
Nacelle	F03D 15/00 or F03D 80/70 or H02K 7/183		Mechanical transmission; bearing; lubricating arrangements.
Rotor	F05B2240/21 or F03D 3/06 or F03D 1/06 Rotor or blade or pitch		Rotors for wind turbines
Tower	F03D 13/10 or F03D 13/20 or E04H 12/00 Tower		Arrangements for erecting, mounting, or supporting wind motors that are like a tower.
Electrical apparatus	H02J 3/386 or F03D 9/25 or Y02E 10/76		Power transfer, power conversion, electrical generators for wind energy.
Control	F03 7/00		Controlling wind motors
Maintenance	F03D 80/50		Maintenance or repair of wind motors
Installation	F03D 13/40	F03D 13/00 Install or lift or transport or assemble or logistics or handle	Assembly, mounting, commissioning, or transporting of wind motors.
Monitor	F03D 17/00		Monitoring or testing of wind motors
Foundation	E02B 17/00 or E02D 27/00 or E02D 29/00		Artificial islands on piles or like supports; foundations or underwater substructures
Mooring	B63B 21/00		Anchoring or mooring

frequently subgroup k is found in the patent family i. For example, patent families that have classification codes in only a single subgroup will have a 1 for that subgroup and zeros elsewhere. We exclude subgroups under class 'Y02E 10/70', which stands for wind energy, to reduce bias. An example of calculating a position vector is presented in table C1 in the appendix,

$$P_i = (c_{i1}, c_{i2}, \ldots, c_{ik}, \ldots, c_{iK}) \qquad (1)$$

where c_{ik} is the frequency of the subgroup-lel code k in patent family i; K is the total number of unique subgroup-level codes among all the patent families considered. This is calculated as shown in equation (2),

$$c_{ik} = \frac{n_{ik}}{\sum_{k=1}^{K} n_{ik}} \qquad (2)$$

where n_{ik} is the number of CPC codes assigned to patent family i that belong to subgroup-level code k, and $\sum_{k=1}^{K} n_{ik}$. is the total number of CPC codes assigned to this patent family.

2.3. Calculate pairwise distance between patent families

We use the Euclidean distance to measure the distance between two vectors. There are many potential methods that can be applied to calculate the pairwise distance. For example, the pairwise distance can be the cosine distance between position vectors of patent families; or it can be calculated from the Jaccard Index based on full-text similarity, which is the size
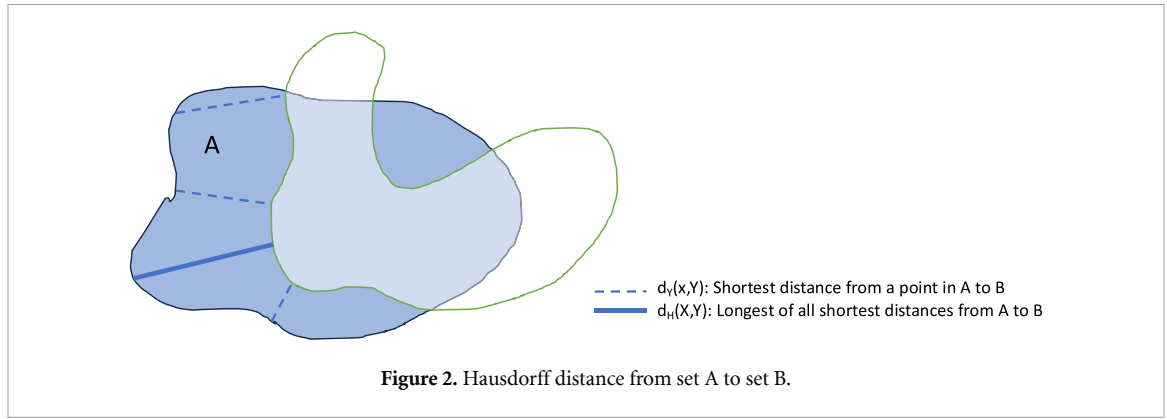
of the intersection of classification codes divided by the size of the union. Any of these methods could be used for the Hausdorff Measure described in the next section. We use Euclidean distance for its simplicity and ease of understanding.

The distance between two patent families x and y, $d(P_x, P_y)$, is calculated using Euclidean Distance and normalized as shown in equation (3). We normalized the standard Euclidean distance to be between 0 and 1. The distance is zero between two patent families if they have the exact same frequencies of each CPC subgroup. The maximum distance between two patent families is achieved when each family has all CPC codes in a single subgroup, and the two subgroups are different. We use this distance to define the distance between two sets in the next section,

$$d(P_x, P_y) = \frac{\sqrt{\sum_{k=1}^{K} (c_{xk} - c_{yk})^2}}{\sqrt{2}}. \qquad (3)$$

2.4. Calculate relatedness by Hausdorff distance

We define the relatedness of two technologies to be the opposite of the Hausdorff distance. Hausdoff distance is a metric for measuring distance between two sets. The original Hausdorff distance between set A and B is defined as the maximal distance between any point in set A to its closest point in set B, and between any point in set B to its closest point in set A. Instead of considering a bilateral distance as the original definition does, we calculate a one-way Hausdorff distance, the maximal distance between any point in the technology of interest (e.g. offshore

Figure 2. Hausdorff distance from set A to set B.

wind energy) to the related mature technology (e.g. onshore wind energy). Figure 2 shows an example of Hausdorff distance between two sets, where the dashed lines represent the shortest distance between pairs and the solid line represents the longest distance among all the shortest distances. It can be observed that for any sets with overlaps, the elements in the overlap have themselves as the nearest pair. Their shortest distance to the other set is zero.

In this analysis, we consider patent domains as the sets, and each patent family as a point in one or more of the sets. We then define the distance from a patent family to a patent domain as in equation (4). This finds the distance between an individual point P_x in patent domain X and its closest neighbor in patent domain Y,

$$d_Y(P_x) = \min_{P_y \in Y} d(P_x, P_y) . \tag{4}$$

The original Hausdorff distance measures the maximal shortest distances between partners, as illustrated in equation (5). In our analysis, we define a modified Hausdorff distance, $\tilde{D}_X(Y)$, as the mean of the top 1 percentile of all $d_Y(P_x)$. Using the entire top 1 percentile rather than the maximum provides a measure that is more robust to the addition of new patents and to retrieval errors. The modified Hausdorff distance is defined by equation (6). The numerator of equation (6) sums up the $d_Y(P_x)$ of patent families that rank the highest 1% among all patent families in patent domain X,

$$D_X(Y) = \max_{P_x \in X} d_Y(P_x) \tag{5}$$

$$\tilde{D}_X(Y) = \frac{\sum_{P_x \in X, \, r_x \leqslant T} d_Y(P_x)}{T} \tag{6}$$

$$T = \frac{N_X}{100} \tag{7}$$

where X, Y are patent domains, which are ss of patent families; P_x, P_y are the position vectors of patent families within patent domain X, Y, defined by equation (1); $\tilde{D}_X(Y)$. is the modified Hausdorff distance from patent domain X to Y; n_X is the number of points in patent domain X; r_x. is the rank of $d_Y(P_x)$ when ranked from the largest distance to Y down; the r_x are integers between 1 and N_X; T is the largest integer that is smaller than $\frac{N_X}{100}$, which gives the 1 percentile.

Note that our modified Hausdorff distance fropatent domain X to Y may be different than the distance from Y to X. We use the one-y distance because this ignores patent families in the other domain Y that have no relationship to X. Finally, we define relatedness as $1 - \tilde{D}_X(Y)$, since the maximum distance is 1,

$$R_X(Y) = 1 - \tilde{D}_X(Y) . \tag{8}$$

We modified the Hausdorff distance definition so that it calculates the mean of the top 1 percentile instead of the maximum for two reasons. First, this modified Hausdorff distance reflects the magnitude of overlap between domains. For any patent family shared by two domains (which 'locate' in the overlap if visualized by a Venn diagram), the closest neighbor in the other domain is itself, which has a distance $d_Y(P_x)$ as defined by equation (4) of zero. All else being equal, the more overlap two domains have, the more patent families will have zero $d_Y(P_x)$, thus decreasing the modified Hausdorff distance. For example, imagine that new patent families all fall in the overlap between two domains. This means more zeros are added to the end of the $d_Y(P_x)$ series. The top 1 percentile will get closer and closer as this happens. However, the magnitude of overlap is not the decisive factor of the Hausdorff distance; this is how it differs from the Jaccard Similarity index between two sets. Two pairs of patent domains may have different levels of modified Hausdorff distance even though they have the same magnitude of overlap, depending on the patent families that have

the furthest closest neighbors in the other domain. Second, the modified Hausdorff distance is less sensitive to retrieval errors. The top 1 percentile is more representative of the space as a whole, as opposed to driving the results by the single most differentiated pair. By definition, the modified Hausdorff distance should be slightly smaller than or equal to the original Hausdorff distance that is defined by the maximum. In appendix D, we repeat the analysis using the maximum and present the difference between two methods.

2.5. Change rate of relatedness

In order to investigate the dynamics of innovation, we create subsets based on the earliest filing year of each patent family. We consider subsets with an end year of 1980, 1990, 2000, 2010, and 2020 from each component domain and from onshore wind and offshore oil and gas platform technology domains. For example, the subset with the end year of 1980 contains all patent families in the domain with an earliest filing year no later than 1980; thus each later subset contains the earlier subsets.

We calculate the average annual change rate in relatedness from subsets by the following steps. First, we calculate the relatedness between each component domain and a technology domain for each subset. The relatedness of domain X to domain Y at the time step of year n is denoted as $R_X(Y_n)$. Second, we calculate the change rate of relatedness between two consecutive timesteps, as in the numerator of equation (9). Third, we divide the change rate from the last step by number of years, as defined by ACR_{n-m} in equation (9):

$$ACR_{n-m} = \frac{\frac{R_X(Y_n)}{R_X(Y_m)} - 1}{n - m} \times 100\% . \qquad (9)$$

Then, we average this ACR across the four 10 year timesteps:

$$AverageAnuualChangeRate = \frac{ACR_{1980-1990} + ACR_{1990-2000} + ACR_{2000-2010} + ACR_{2010-2020}}{4} . \qquad (10)$$

A negative average annual change of relatedness indicates a decreasing relatedness through time, while a positive change indicates the component evolves to be more related to the given technology.

3. Results

3.1. Domain definitions

We retrieved 46 745 patent families of offshore wind, 47 752 patent families of onshore wind and 4011 patent families of offshore oil and gas platforms. We excluded patents that have no classification code fields assigned in PATSTAT, which resulted in a decrease of 18%, 16% and 22% in data size, respectively. This shows a limitation of our method if the data missingness in classification codes is high. This results in 38 194 patent families of offshore wind, 40 053 patent families of onshore wind, and 3148 patent families of offshore oil and gas platforms. We find a large overlap between offshore and onshore wind energy, with 82% of the patent families in this intersection (figure 3). There are 3098 patent families that are exclusive to offshore wind (8% of all offshore wind patent families), and 4957 patent families exclusive to onshore wind (12% of all onshore wind patent families).

We can compare this to Tsai *et al* (2016), the only paper in the literature that has attempted to define an offshore-only wind energy domain. That paper used keywords only and manually reviewed each patent. They found 381 granted patents in the offshore-only domain from the European Patent Office (EPO) or United States Patent and Trademark Office (USPTO) with publication dates between January 1, 1976, and June 30, 2015. Over this same period, our offshore-only domain has 2,610 patent families worldwide, including 669 patent families with applications either under EPO or USPTO. Our offshore-only domain is larger than that in Tsai et al., due to our usage of both classification codes and keywords for definitions. On the other hand, our results were not manually examined one by one, thus, we will have some false positives, as reported in table B1.

We focus on technology components that have significant impacts on the total costs of wind energy according to component-level levelized cost of electricity (LCOE) contribution for wind energy projects (Gonzalez-Rodriguez 2017, Stehly and Duffy 2021). We identified three 'above the water' components believed to be closely related to onshore wind energy (nacelle, tower, rotor), two 'below the water' components closely related to offshore oil and gas platforms (foundation, mooring) and a few important cost components of offshore wind energy whose relations to the two technologies
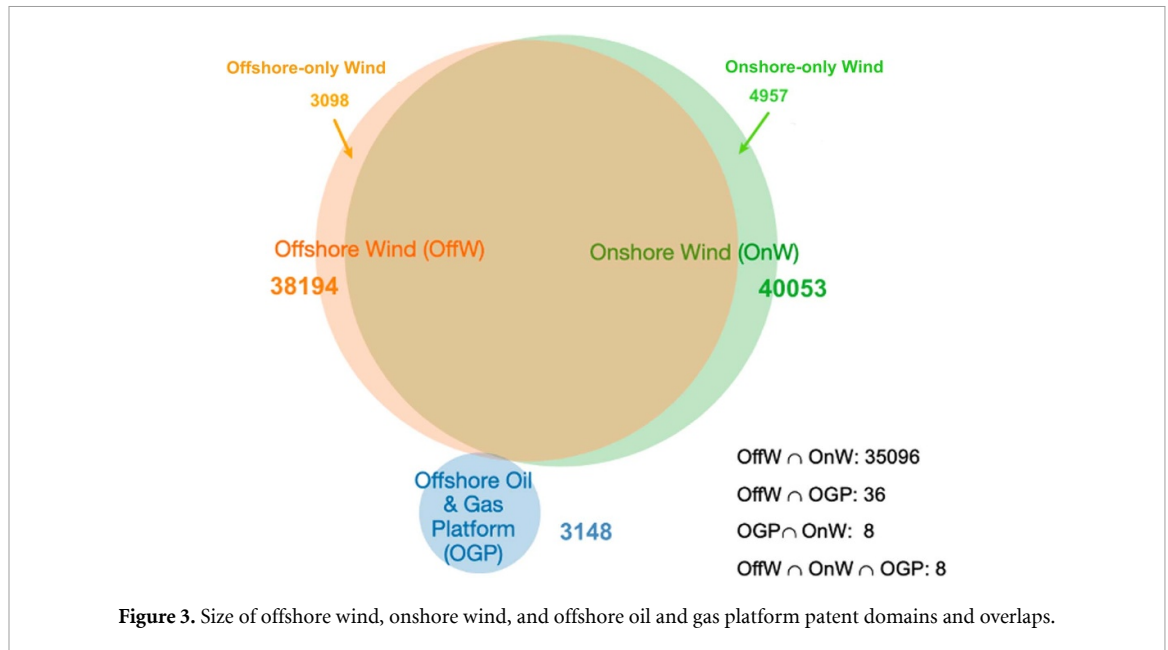
Figure 3. Size of offshore wind, onshore wind, and offshore oil and gas platform patent domains and overlaps.

Table 3. Patent family count of offshore wind component domain.

	Nacelle	Rotor	Tower	Electrical Apparatus	Control	Maintenance	Installation	Monitor	Foundation	Mooring
Patent family count	2711	3775	469	6546	4507	561	1089	1059	580	174

are less clear *a priori* (electrical apparatus, installation, control, monitor, maintenance). The results for each component are listed in table 3. We identify 15 648 patent families in the union of the components we consider after dropping duplicates, which makes up 41% of all offshore wind patent families.
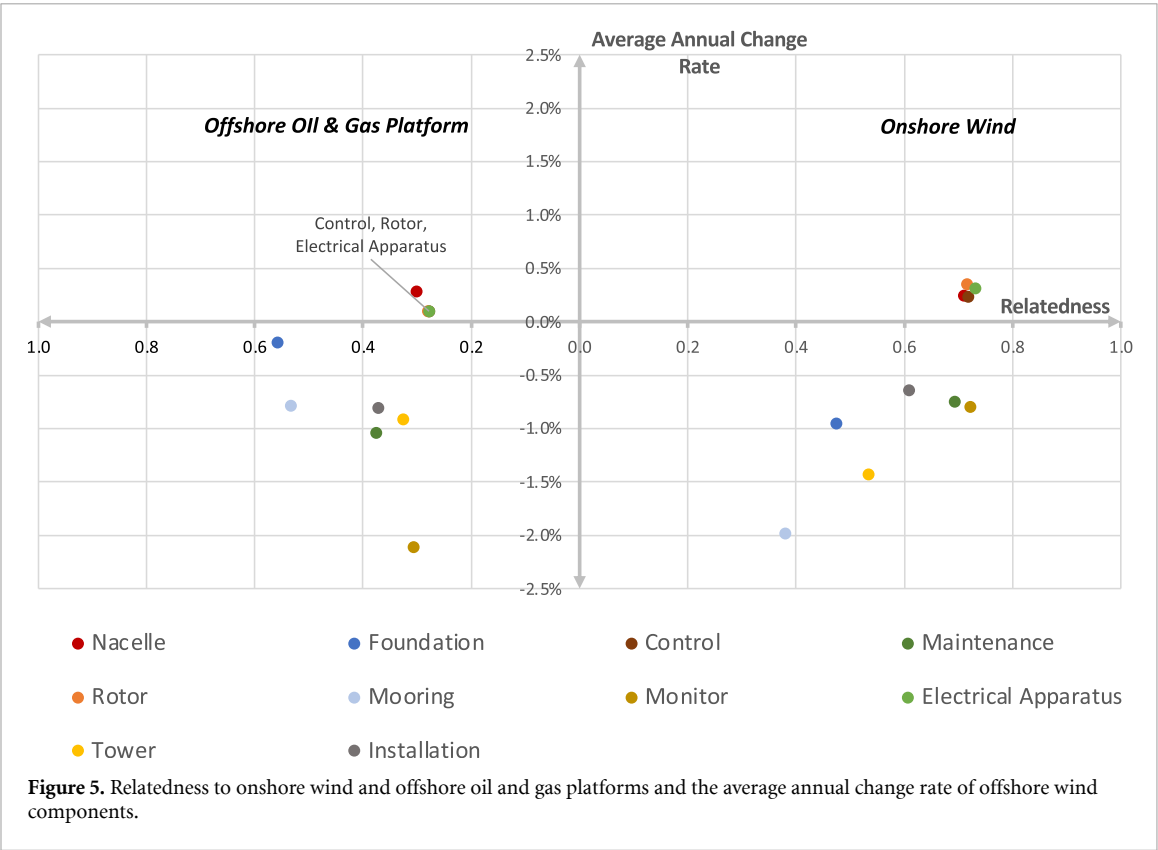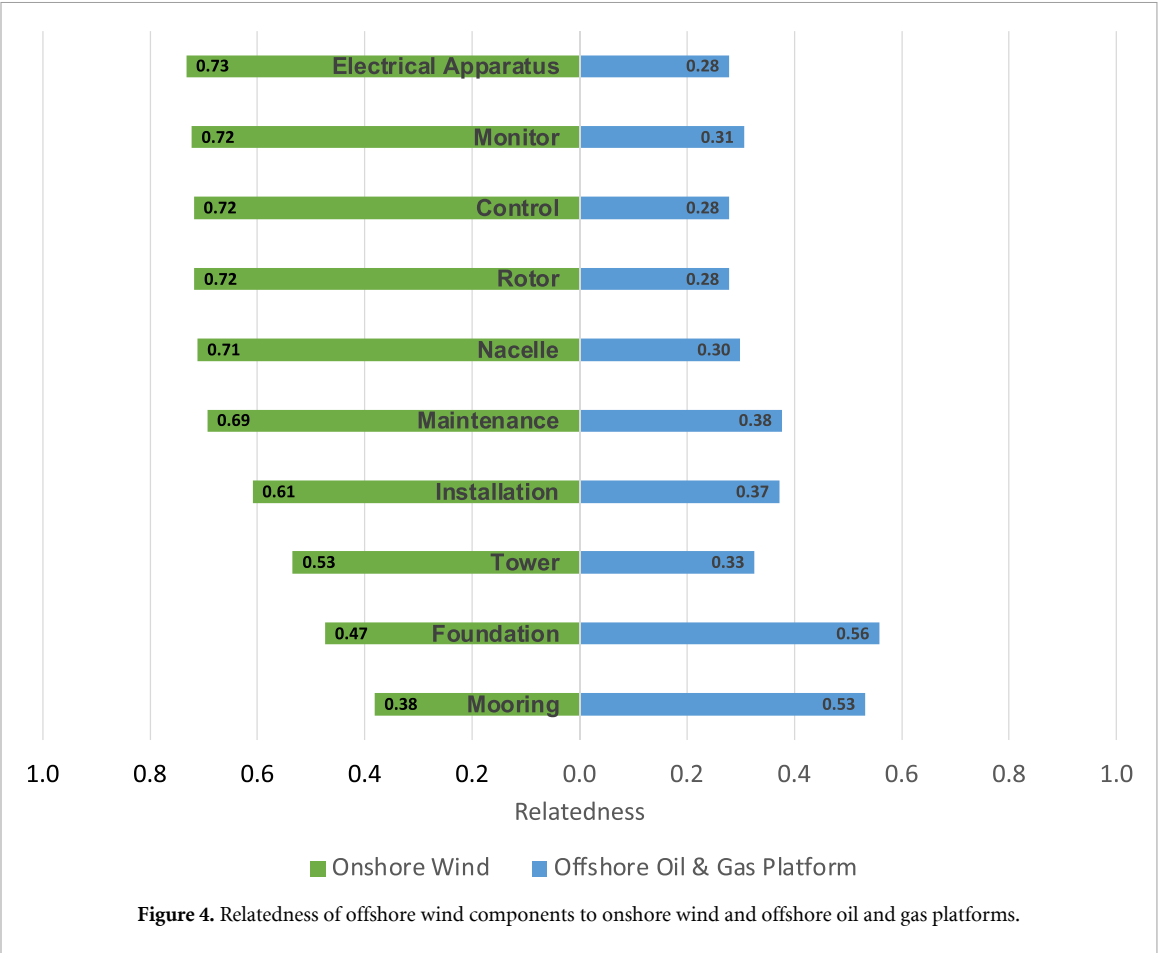
3.2. Relatedness
Figure 4 shows the relatedness calculated for the ten offshore wind energy technology components with onshore wind energy and offshore oil and gas. The horizontal axis to the left represents relatedness of offshore wind components to onshore wind; the horizontal axis to the right represents relatedness of offshore wind components to offshore oil and gas platforms. The longer the bars, the more related the offshore wind component is to the other technology domain.

We calculate our Hausdorff relatedness for each timestep and compute the average annual change in relatedness as defined by equation (10). In figure 5, the horizontal axis shows the relatedness calculated for the ten offshore wind energy technology components with onshore wind energy (to the right) and offshore oil and gas platforms (to the left). A patent family is more related to onshore wind or offshore oil and gas platforms if it is located farther to

the right (for onshore wind) or left (for offshore oil and gas platforms). The vertical axis shows the results of Average Annual Change Rate since 1980 calculated by equation (10). Components above the horizontal axis have shown increasing relatedness, those below decreasing relatedness.

4. Discussion

Figure 4 shows that rotors and nacelles are highly related to onshore wind, as expected, with a low relatedness to offshore oil and gas among 'above the water' components. Towers, however, are relatively less related to onshore wind. This difference may be explained by the large size and weight of offshore towers and the importance of the marine environment. In addition, erecting and mounting such towers in the offshore environment may require extra knowledge from offshore oil and gas platform technology; we see that towers are slightly more related to offshore oil and gas. As for 'below the water' components, we find foundations and mooring are the components least related to onshore wind energy and most related to offshore oil and gas. This is in accordance with the observational similarity between the offshore oil and gas platforms and fixed, submersible and floating platforms for offshore wind turbines. Maintenance and installation are the next two most

Figure 4. Relatedness of offshore wind components to onshore wind and offshore oil and gas platforms.

Figure 5. Relatedness to onshore wind and offshore oil and gas platforms and the average annual change rate of offshore wind components.

related to offshore oil and gas; these also have relatively high relatedness with onshore wind energy. This implies that these techniques and technologies are where the knowledge from both technologies combine: the onshore wind industry provides knowledge specific to wind turbines while the offshore oil and gas industry provides knowledge relevant to the offshore environment. Of interest are electrical apparatus, control, and monitoring techniques, as there was a less clear prior opinion for these components. These show high relatedness to onshore wind energy and low relatedness with the offshore oil and gas, which indicates the innovation in these fields for offshore wind turbines are very similar for onshore wind.

As shown by figure 5, nearly all offshore wind components have negative or very small average annual change of relatedness with both onshore wind energy and offshore oil and gas platforms. This indicates that most offshore wind energy components are innovating away from the two more mature technologies. Five of the components (those located in the top right of figure 5) have small positive annual changes in relatedness to onshore wind, including five of the top six of those most related to onshore wind. These components would be amenable to spillovers, and therefore benefit from overall growth in wind energy; but are less likely to see large cost reductions as offshore wind grows. Conversely, the four components least related to onshore wind, mooring, tower, foundation, and installation techniques, have been evolving even further away. These components are more promising for offshore-specific innovation, and more likely to reduce in cost as the offshore wind energy industry grows. These implications come from the model presented in Hernandez-Negron *et al* (2023). The idea is that the cost of a technology drops at a fairly constant rate with every doubling of cumulative capacity. If a component has a high relatedness to a more mature technology, it is building on a larger initial cumulative capacity, thus increases in capacity in the emerging industry will have a relatively small impact on overall cumulative capacity, and thus a smaller impact on cost reductions.

We are seeing that offshore wind is evolving away from offshore oil and gas. In particular, the components that started more related to offshore oil and gas, such as mooring, maintenance, towers, and monitoring, are trending away from this domain. It suggests that while these offshore wind components started out relatively related to offshore oil and gas techniques, innovation is moving in a different direction.

There are at least four applications for this method and its results. First, for modelers, our method can help improve models of technological change through time. For example, technological forecasting models, such as learning curve models, usually ignore the relatedness of offshore wind energy to these more mature industries and consider it as a purely novel technology (van der Zwaan *et al* 2012, Haas *et al* 2022). Our method produces a quantitative value for relatedness that can be used in learning curves as in the framework introduced in (Hernandez-Negron *et al* 2023). Second, this analysis can inform future patent analyses aimed at understanding technological change and its impacts through time. One question in patent analysis is how to define the boundaries between technologies. When technologies are related, such as the case for onshore and offshore wind, the a priori boundaries may be misleading. Our method helps to clarify when technology domains are significantly related, thus requiring more nuanced domain definitions. As mentioned in Sun *et al* (2021), a narrow technology domain definition may miss potential breakthrough patents and evidence for future innovation trends. Our method can help modelers identify technology domains with high relatedness to the focal technology so that are necessary to be included in the analyses. Third, understanding relatedness is relevant to decisions around R&D investments, whether for government policy-makers or individual firms. It helps us understand how and where related technologies are diverging, thus highlighting where R&D investments will be more impactful in an emerging industry. Finally, for practitioners, the understanding of component-level technological relatedness can be informative to strategic firm positioning. For example, firms that work in fields that are highly related across two technology domains can position themselves to support both domains, whereas firms that work on components that are differentiated or diverging, can position themselves as focused on the emerging technology.

5. Conclusions

We have illustrated a method for understanding the relatedness of technology domains through the use of patent data. We propose a method that builds on the Hausdorff distance to calculate relatedness between technology domains. We use offshore wind energy as an example and look into how offshore wind energy components are related to onshore wind energy and offshore oil and gas. The results are

fairly consistent with experts' opinions that 'above the water' components such as the rotors and nacelles are highly related to onshore wind; while 'below the water' components such as mooring and foundations are more related to offshore oil and gas. The results also shed light on components for which there was less prior opinion, indicating that techniques related to electrical apparatus, monitoring, and control techniques are highly related to onshore wind energy, even more than nacelles and rotors. Moreover, these components do not show movement away from onshore wind. This implies that these components are less likely to see significant cost reductions with the growth of offshore wind. In addition, maintenance and installation show significant relatedness to both onshore wind and offshore oil and gas. Considering the dynamics, the 'below the water' components are evolving away from both mature technologies, implying accelerating innovation in these areas. This may promise more significant potential for cost reductions in these components with the growth of the offshore wind industry, compared to those that are becoming more similar to the mature technologies. Moreover, this information can be useful for guiding and fostering R&D investment and firm positioning in the field of offshore wind energy.

Measuring relatedness by patent data has limitations. First, patents may not represent all innovation in a field, particularly in areas that tend to be unpatented (Popp 2005). Second, it is possible the innovation has not been fully transferred to patents in this area due to filing and examination delays. However, our methodology using patent data may be particularly helpful for understanding an emerging and complex technology that includes components highly related to other technologies.

These results contribute to a topic of emerging interest around defining ontologies or hierarchies of distinction among technologies. In making use of the rich data available in patent datasets, we harness knowledge that can be combined with expert judgment and data science to improve the understanding of technologies and their innovation trajectories. This is particularly important and timely to support technological change in the energy transition, where many promising newer technologies have evolved from more mature technologies.

Data availability statement

All data that support the findings of this study are included within the article (and any supplementary files).

Appendix

A. Technology domain definitions

Table A1. Retrieval queries for offshore-only wind, onshore-only wind, and offshore oil and gas platforms.

	Retrieval query
Offshore-only wind	Within the wind energy domain defined by Fung *et al*: Any(B63H, B63B, E02B 17/027, E02D 15/06, E02D 27/52, E02D 29/06, E21B 7/12, F03D 13/25, F05B2240/95)@(CPC, IPC) OR Y02E 10/727@CPC OR Any(offshore, ocean, lake, marine, aquatic anchor, float, undersea, at_sea)@(title, abstract)
Onshore-only wind	Within WEDD defined by Fung *et al.* F03D 9/48@(CPC,IPC) OR Y02E 10/728@CPC OR Any(onshore, tree, mountain, road, land))@(title, abstract)
Offshore oil and gas platforms	B63B 35/4413@CPC OR Any(E02B 17, E02B2017, B63B 35/003, B63B 77)@CPC AND Any(oil, gas))@(title, abstract) OR Any(E02B 17, B63B 35/00, B63B 35/44, B63B 77)@IPC AND Any(oil, gas))@(title, abstract) OR Any(E21B)@(CPC,IPC) AND All(offshore, platform)@(title, abstract)

B. Expert validation of offshore and onshore wind patent retrieval

We used expert judgment to validate our patent sets. We recruited three wind energy experts to validate our results. Experts were given 91 patents chosen randomly from the wind energy patent domain. They were asked to mark patents as 'applicable only to offshore wind energy', 'applicable only to onshore wind energy' or not. The validation results are summarized in a 2 × 2 table called the confusion matrix for each patent domain. The two dimensions of a confusion matrix are predictions and actual values as shown in figure B1, both having true and false rows or columns. Take the confusion matrix for the offshore patent domain as an example. A patent is actually true if it is marked by experts as 'applicable only to offshore wind energy'; and vice versa. It is predicted true if it is found in our offshore wind energy patent domain; and vice versa. Furthermore, a patent is

	Actually Positive	Actually Negative
Predicted Positive	True Positives	False Positives
Predicted Negative	False Negatives	True Negatives

Figure B1. Confusion matrix definitions.

Table B1. Confusion matrix of offshore and onshore wind patents.

Offshore	Predicted true	Predicted false
Actual true	9	1
Actual false	3	78

Onshore	Predicted true	Predicted false
Actual true	8	3
Actual false	3	77

'true' if it is actually true and 'false' if it is not. It is 'positive' if it is predicted true and 'negative' if it is not. From the confusion matrix, we calculate the precision and recall of our patent retrieval for both offshore and onshore domains. Precision refers to the percentage of true positives over the sum of all positives. Recall refers to the percentage of true positives over the sum of all trues. Both offshore and onshore patent domains have good precisions and recalls of at least 0.7.

$$Precision_{\text{offshore}} = \frac{9}{12} = 0.75, Recall_{\text{offshore}} = \frac{9}{10} = 0.9$$

$$Precision_{\text{onshore}} = \frac{8}{11} = 0.72, Recall_{\text{onshore}} = \frac{8}{11} = 0.72$$

C. An example of defining patent position vectors by CPC code frequency

Table C1 shows an example of a set of three patent families, under the assumption that there are a total of 7 relevant subgroups. The position vectors for patent family 1, 2 and 3 are:

$$P_1 = (0.67, 0, 0.33, 0, 0, 0, 0)$$
$$P_2 = (0, 0.2, 0, 0.2, 0.2, 0.4)$$
$$P_3 = (1, 0, 0, 0, 0, 0, 0).$$

The elements in the vector are the frequency c_k of codes B63B 21/50, B63B 21/56, B63B2035/44, B64D 17/02, B63G 8/42, and B63H 9/04, respectively.

D. Comparing results from the modified and the original Hausdorff method

We will hereafter refer to these two methods as 'TOP1' and 'MAX'. The numerical results are listed in table D1. In figure D1, we plot the relatedness measured by TOP1 on the *X* axis and by MAX on the *Y* axis. Data points closer to the diagonal of the figure indicate the relatedness from the two methods are more similar.

As the figure shows, the TOP1 and MAX relatedness are very close. For offshore wind and offshore

Table C1. An example of calculating position vectors for three patent families.

Patent family	CPC classification codes	CPC codes at subgroup level	Frequency of codes
1	B63B 21/50, B63B 21/505 B63B2035/446 Y02E 10/727 (excluded)	B63B 21/50 B63B2035/44	0.67 0.33
2	B63B 21/66 B64D 17/025 B63G 8/42 B63H 9/069, B63H 9/10	B63B 21/56 B64D 17/02 B63G 8/42 B63H 9/04	0.2 0.2 0.2 0.4
3	B63B 21/50	B63B 21/50	1

oil and gas platform, the difference is extremely small. Components with lower relatedness can be better differentiated by the TOP1 relatedness. As for offshore wind and onshore wind, the difference between the MAX and TOP1 relatedness is small for most components, but components with higher TOP1 relatedness are better differentiated by the MAX relatedness. We added the best linear fit line to the onshore wind datapoints to show the deviation from the diagonal. The difference is greater for electrical apparatus (0.23) and rotors (0.22). The reason might be their large domain sizes, such that the top 1 percentile and the maximum are very different. While we see some differences in the results, they do not strongly qualitatively or ordinally change things, and we continue to believe that using the top 1 percentile is more representative of the space as a whole, as opposed to driving the results by on the single most differentiated pair.

E. Average annual change rate with varying timesteps

In order to check our results we repeat the process used to create figure 5, but use timesteps with approximately equal patent families. Specifically we created subsets with an end year of 1985, 2000, 2010, 2015, 2020. The years were chosen in a way that each timestep has fairly same number of patent families. We have this consideration because wind patents started to increase rapidly starting in 2000.

As shown by figure E1, the results are similar to figure 5. There are four components that have a positive average annual change rate with both onshore wind and offshore oil and gas: control, rotors, electrical apparatus, and nacelle. The remaining six components have negative average annual change rates with the two domains.

F. Descriptive analysis of CPC-based position vectors

We conduct descriptive analysis on the position vectors of patent families and subgroups across domain.

Table D1. TOP1 and MAX relatedness with onshore wind and offshore oil and gas platform of components.

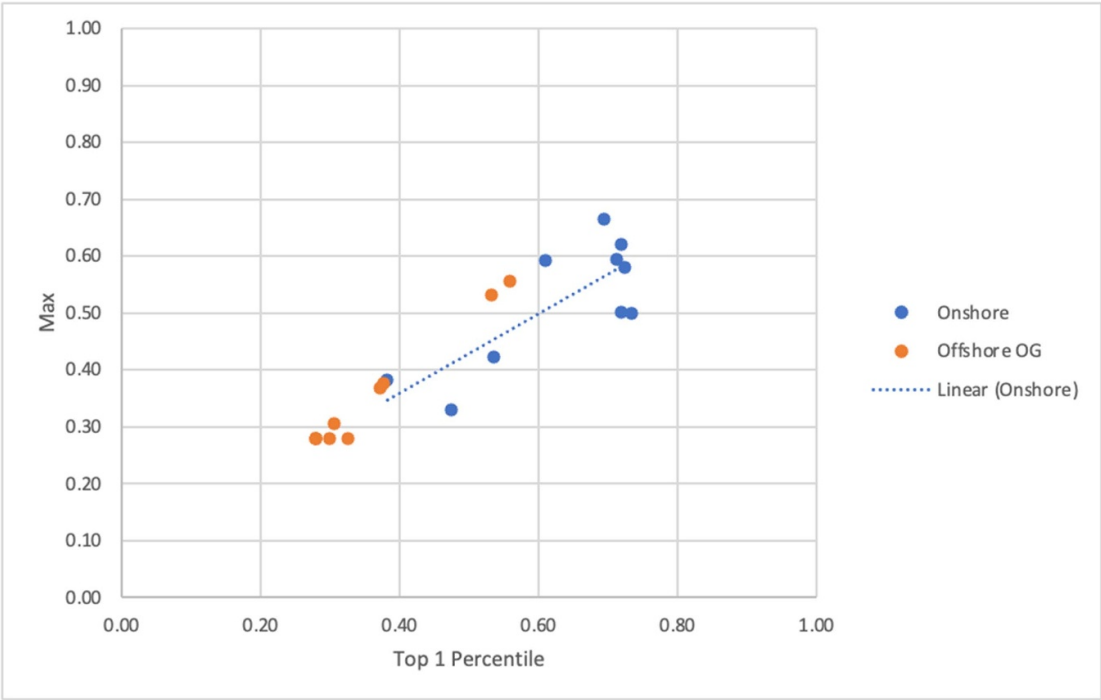|		Nacelle	Rotor	Tower	Electrical apparatus	Control	Maintenance	Installation	Monitor	Foundation	Mooring
Onshore Wind	TOP1	0.71	0.72	0.53	0.73	0.72	0.69	0.61	0.72	0.47	0.38
	MAX	0.59	0.50	0.42	0.50	0.62	0.66	0.59	0.58	0.33	0.38
Offshore Oilgas Platform	TOP1	0.30	0.28	0.33	0.28	0.28	0.38	0.37	0.31	0.56	0.53
	MAX	0.28	0.28	0.28	0.28	0.28	0.38	0.37	0.31	0.56	0.53

Figure D1. *X–Y* plot of TOP1 and MAX relatedness with onshore wind and offshore oil and gas platform.
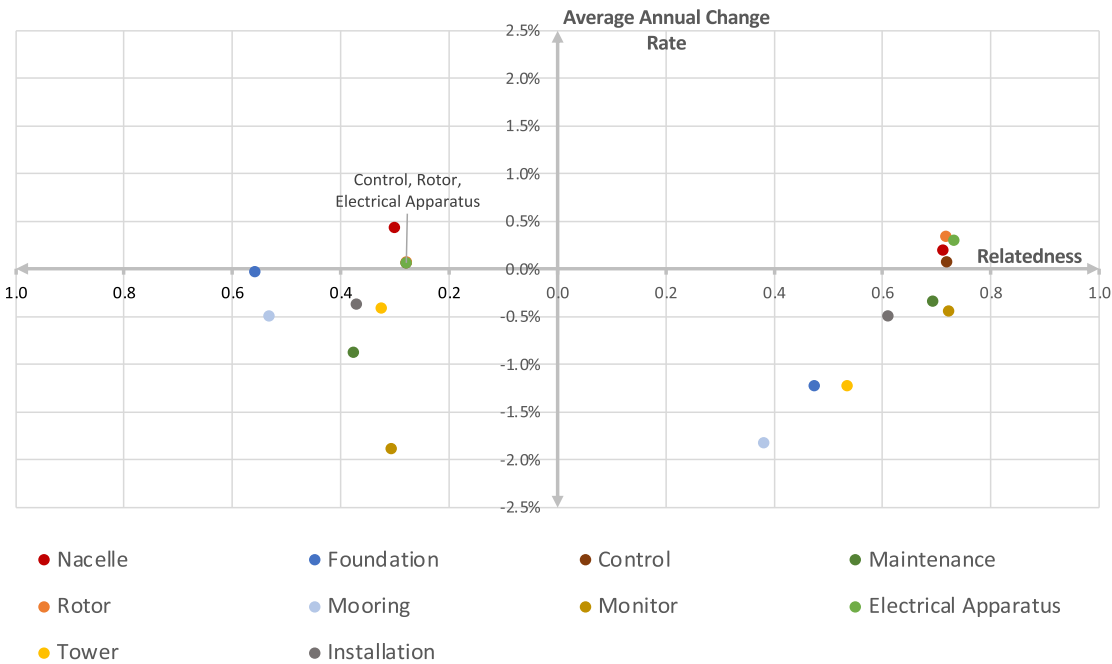

Figure E1. Relatedness to onshore wind and offshore oil and gas platforms and the average annual change rate of offshore wind components (using varying timesteps).

Tables below show that more than 60% of patent families in each technology domain have unique position vectors, indicating they are located at unique positions in the vector space of that domain. We notice there exist some patent families sharing the same position vector, but even the largest group only accounts for less than 5% of patent families in all technology domains.

	Offshore	Onshore	Offshore oil & gas platform
% of Patent families with unique position vectors	65.5%	65.2%	62.3%
Largest share of patents families sharing the same position vector	3.5%	4.5%	2.9%
Number of subgroups	5904	5886	1393
Longest length of position vectors (excluding zeros)	44	44	31
Average length of position vectors (excluding zeros)	4	4	4

We also look at different offshore wind components, using rotors, platforms and nacelles as an example. It can also be concluded that most patent families are 'spread-out' in its domain according to the percentage of patent families with unique position vectors.

	Offshore rotor	Offshore platform	Offshore nacelle
% of Patent families with unique position vectors	67.3%	85.7%	91.7%
Largest share of patents families sharing the same position vector	3.3%	5.0%	0.4%
Number of subgroups	1317	399	1424
Longest length of position vectors (excluding zeros)	37	31	28
Average length of position vectors (excluding zeros)	5	6	7

G. Relatedness if comparing to a subdomain and to the entire domain

Below figure G1 shows the comparison between two methods: comparing an offshore wind component domain to a similar subdomain, or to the entire domain. We found the results are very close. The reason might be the similar component in the other domain is largely the closest neighbor. Indeed, comparing to the entire domain may include some unrelated patent families. However, the Hausdorff method can reduce the impacts of less related patents by finding the closest neighbors.

The subdomain in onshore wind and offshore oil and gas platforms are retrieved using the same definitions for corresponding offshore wind components. However it is possible the definitions for offshore wind are not the most accurate for the onshore wind and offshore oil and gas platforms. A more precise definition can only affect the results by introducing more related patents, which is also the outcome of comparing to the entire domain.
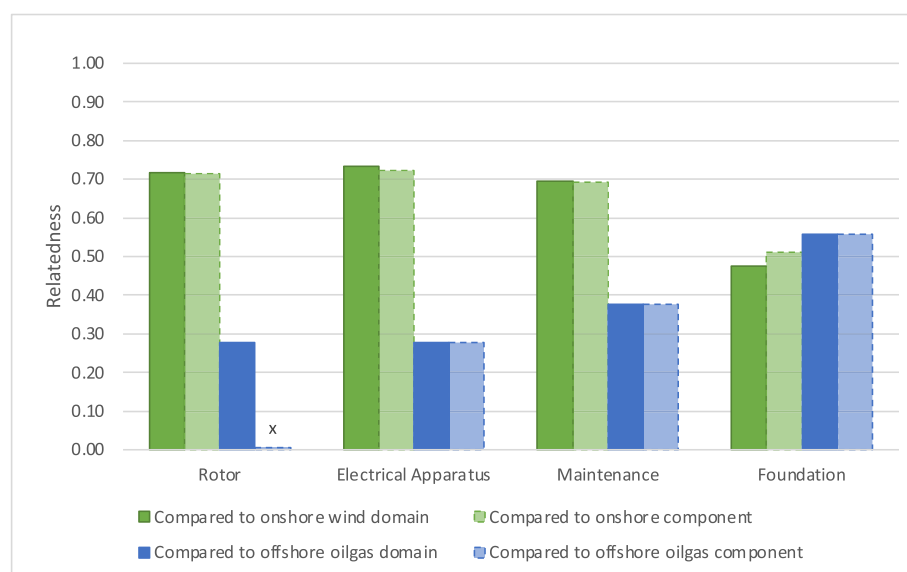
Figure G1. Relatedness for four offshore components comparing to a subdomain and to the entire domain.
Note: 'x' indicates there are no patent results for 'offshore oil and gas "rotors"'.

References

Aharonson B S, S. A B and Schilling M A 2016 Mapping the technological landscape: measuring technology distance, technological footprints, and technology evolution *Res. Policy* **45** 81–96

Arts S, Cassiman B and Gomez J C 2018 Text matching to measure patent similarity *Strateg. Manage. J.* **39** 62–84

Braun F G, Schmidt-Ehmcke J and Zloczysti P 2010 *Innovative Activity in Wind and Solar Technology: Empirical Evidence on Knowledge Spillovers Using Patent Data* Discussion Paper No. 993 (DIW Berlin) (https://doi.org/10.2139/ssrn.1633875)

Breschi S, Lissoni F and Malerba F 2003 Knowledge-relatedness in firm technological diversification *Res. Policy* **32** 69–87

Cranmer A and Baker E 2020 The global climate value of offshore wind energy *Environ. Res. Lett.* **15** 054003

Ferioli F, Schoots K and van der Zwaan B C C 2009 Use and limitations of learning curves for energy technology policy: a component-learning hypothesis *Energy Policy* **37** 2525–35

Fung K, Goldstein A, Baker E and Wang Y 2023 Rethinking the patent domains: an application to wind energy *World Pat. Inf.* **74** 102209

Gonzalez-Rodriguez A G 2017 Review of offshore wind farm cost components *Energy Sustain. Dev.* **37** 10–19

Haas R, Sayer M, Ajanovic A and Auer H 2022 Technological learning: lessons learned on energy technologies *WIREs Energy Environ.* **12** e463

Hernandez-Negron C G, Baker E and Goldstein A P 2023 A hypothesis for experience curves of related technologies with an application to wind energy *Renew. Sustain. Energy Rev.* **184** 113492

Jaffe A B 1986 Technological opportunity and spillovers of R & D: evidence from firms' patents, profits, and market value *Am. Econ. Rev.* **76** 984–1001

Jaffe A B 1989 Characterizing the "technological position" of firms, with application to quantifying technological opportunity and research spillovers *Res. Policy* **18** 87–97

Joo S H and Kim Y 2010 Measuring relatedness between technological fields *Scientometrics* **83** 435–54

Kanyako F and Baker E 2021 Uncertainty analysis of the future cost of wind energy on climate change mitigation *Clim. Change* **166** 10

Li D, Heimeriks G and Alkemade F 2020 The emergence of renewable energy technologies at country level: relatedness, international knowledge spillovers and domestic energy markets *Ind. Innov.* **27** 991–1013

Mogharrebi M, Ang M C, Prabuwono A S, Aghamohammadi A and Ng K W 2013 Retrieval system for patent images *Proc. Technol.* **11** 912–8

Moreno R and Ocampo-Corrales D 2022 The ability of European regions to diversify in renewable energies: the role of technological relatedness *Res. Policy* **51** 104508

Nakamura H, Suzuki S, Sakata I and Kajikawa Y 2015 Knowledge combination modeling: the measurement of knowledge similarity between different technological domains *Technol. Forecast. Soc. Change* **94** 187–201

Popp D 2005 Lessons from patents: using patents to measure technological change in environmental models *Ecol. Econ.* **54** 209–26

Popp D, Santen N, Fisher-Vanden K and Webster M 2013 Technology variation vs. R&D uncertainty: what matters most for energy patent success? *Resour. Energy Econ.* **35** 505–33

Punt M B, Bauwens T, Frenken K and Holstenkamp L 2022 Institutional relatedness and the emergence of renewable energy cooperatives in German districts *Reg. Stud.* **56** 548–62

Rao S, Keppo I and Riahi K 2006 Importance of technological change and spillovers in long-term climate policy Energy J. **27** 123–39 (available at: www.jstor.org/stable/23297059)

Rigby D L 2015 Technological relatedness and knowledge space: entry and exit of US cities from patent classes *Reg. Stud.* **49** 1922—-37

Rodriguez A, Kim B, Turkoz M, Lee J-M, Coh B-Y and Jeong M K 2015 New multi-stage similarity measure for calculation of pairwise patent similarity in a patent citation network *Scientometrics* **103** 565–81

Stehly T and Duffy P 2021 *2020 Cost of Wind Energy Review* NREL/TP-5000-81209 (National Renewable Energy Laboratory) (available at: www.nrel.gov/docs/fy22osti/81209.pdf)

Sue Wing I 2006 Representing induced technological change in models for climate policy analysis *Energy Econ.* **28** 539–62

Sun B, Kolesnikov S, Goldstein A and Chan G 2021 A dynamic approach for identifying technological breakthroughs with an application in solar photovoltaics *Technol. Forecast. Soc. Change* **165** 120534

Tanner A N 2016 The emergence of new technology-based industries: the case of fuel cells and its technological relatedness to regional knowledge bases *J. Econ. Geogr.* **16** 611–35

Tsai Y-C, Huang Y-F and Yang J-T 2016 Strategies for the development of offshore wind technology for far-east countries—a point of view from patent analysis *Renew. Sustain. Energy Rev.* **60** 182–94

van der Zwaan B, Rivera-Tinoco R, Lensink S and van den Oosterkamp P 2012 Cost reductions for offshore wind power: exploring the balance between scaling, learning and R&D *Renew. Energy* **41** 389–93

Verhoeven D, Bakker J and Veugelers R 2016 Measuring technological novelty with patent-based indicators *Res. Policy* **45** 707–23

Weyant J P 2017 Some contributions of integrated assessment models of global climate change *Rev. Environ. Econ. Policy* **11** 115–37

Whittle A 2020 Operationalizing the knowledge space: theory, methods and insights for Smart Specialisation *Reg. Stud. Reg. Sci.* **7** 27–34

Yan B and Luo J 2017 Measuring technological distance for patent mapping *J. Assoc. Inf. Sci. Technol.* **68** 423–37

Yoon J and Kim K 2012 Detecting signals of new technological opportunities using semantic patent analysis and outlier detection *Scientometrics* **90** 445–61

Yu N 2017 Innovation of renewable energy generation technologies at a regional level in China: a study based on patent data analysis *Int. Econ. Econ. Policy* **14** 431–48