Working Paper Series

A keyword selection method for mapping technological knowledge in specific sectors through patent data: the case of biofuels sector.

by

Valeria Costantini, Francesco Crespi, Ylenia Curci

17/2014

The Sustainability Environmental Economics and Dynamics Studies (SEEDS) is an inter-university research centre that aims at developing and promote research and higher education projects in the fields of ecological and environmental economics, with a special eye to the role of policy and innovation in the path towards a sustainable society, in economic and environmental terms. Main fields of action are environmental policy, economics of innovation, energy economics and policy, economic evaluation by stated preference techniques, waste management and policy, climate change and development.

The SEEDS Working Paper Series are indexed in RePEc and Google Scholar.
Papers can be downloaded free of charge from the following websites:
<http://www.sustainability-seeds.org/>.

Enquiries: info@sustainability-seeds.org

SEEDS Working Paper 17/2014
June 2014
by Valeria Costantini, Francesco Crespi, Ylenia Curci.

The opinions expressed in this working paper do not necessarily reflect the position of SEEDS as a whole.

A keyword selection method for mapping technological knowledge in specific sectors through patent data: the case of biofuels sector

Valeria Costantini, Roma Tre University and SEEDS
Francesco Crespi, Roma Tre University, BRICK – Collegio Carlo Alberto and SEEDS¹
Ylenia Curci, Roma Tre University

Abstract

In this paper we propose an innovative methodology that aims to solve drawbacks related to how patent data are allocated and organized in international databases. We propose as a case study the biofuels sector, in order to evaluate the validity of such a method. Starting with a systematic mapping of biofuels production value chain, we have built a comprehensive description of the biofuels technological domain. The resulting list of keywords relies on an iterative selection approach, based on an analysis of recent scientific literature combined with the keyword search tool developed by Scopus. The final patent database, BioPat, has been finalized by a validation procedure with the help of expert interviews, revealing improved accuracy compared with standard IPC-based codes. Collected information in BioPat allows us to derive more intriguing insights on the characteristics and evolution of technological patterns in the biofuels sector with respect to standard classification methods.

J.E.L.: O31, O33, Q42.

Keywords: patents classification; keyword selection; innovation patterns; liquid biofuels, knowledge complexity.

¹Corresponding author: Department of Economics, RomaTreUniversity, Via Silvio D'Amico 77, 00145 Rome, Italy. Tel: 0039 06 57335755; Fax: 0039 06 577335771; francesco.crespi@uniroma3.it.

The authors would like to thank Alessandra Molina and Alessandro Palma for their contribution to the collection of keywords, Federico Caviggioli for his advice on downloading patents, Vincenzo Alfano, Alessandro Blasi and Carlo Pignatelli for the patent validation process. Financial support from the European Union D.G. Research with the Grant number 266959 in the research project "PICK-ME - Policy Incentives for the Creation of Knowledge: Methods and Evidence" and the Italian Ministry of Education, University and Research (Scientific Research Program of National Relevance 2010 on "Climate change in the Mediterranean area: scenarios, economic impacts, mitigation policies and technological innovation") is also gratefully acknowledged. The usual disclaimers apply.

1. Introduction

Measuring innovation is a challenging task especially when specific technological domains have to be analyzed. Innovation in fact depends on a variety of activities ranging from formalised research and development (R&D) to production engineering. Organizational innovations and different forms of soft innovations are also relevant. Moreover, the introduction of innovations does not follow a linear process from R&D activities to the eventual commercialisation of new products (Archibugi and Pianta, 1996). Consequently, the most common innovation input and output indicators have been subject to much criticism (Sirilli, 1999).

In particular, the use of patent data for measuring innovation is widespread in the literature, although it has its pros and cons (Griliches, 1990). A single patent provides information on relevant aspects of the innovative process such as the geographical origin of the innovation, its relevance in terms of technological progress, the previous stock of knowledge that allowed the development of incremental innovation, inventors and owners of the patent and the relevance of the state of the art for future innovations. However, only a limited part of produced innovations is patented (Archibugi and Pianta, 1996) and patent data classifications are not organized according to economic principles with an intrinsic variability in patent value (Jaffe and Trajtenberg, 2002).

Moreover, when patent data are used for the analysis of a specific technological domain, an additional (and to some extent crucial) issue related to the choice of the criterion adopted to define and describe a specific sector has to be considered. In fact, the hierarchical International Patents Classification (IPC) aims to classify the innovative content of the patent (WIPO, 2011) while economic activities are classified according to the domain of goods they produce. In this respect, the IPC is only of limited usefulness when it comes to identifying a specific sector which does not fit the criteria used in the classification itself (Narin, 2000). The different sections in IPC allow distinctions between patents belonging to categories which do not necessarily have economic relevance and, when the focus of the analysis is not on an existing section, the use of IPC is inappropriate.

This study deals with this particular aspect of mapping technological change in specific sectors where partitioning data through patent classification systems is not feasible. We develop an original iterative approach to obtain a selection criterion capable of capturing the information that is supposed to identify a complex sector, reducing the risk of either

considering external elements in the analysis or excluding relevant information which typically characterizes standard IPC code-based analysis.

In order to test the validity of such an approach, we have chosen the biofuels sector as a case study for applying our methodology because of three main reasons. First of all, the measurement and analysis of environmental innovation are receiving increasing interest in the current academic and political debate and much effort has been devoted to investigating the dynamics, characteristics, determinants of eco-innovations and their impact in terms of economic and environmental performances (Arundel and Kemp, 2011; Berkhout, 2011; Borghesi *et al.*, 2013; Horbach, 2008; Kemp and Oltra, 2011; Markard *et al.*, 2012; OECD, 2011).² Second, within the domain of environmental technologies those related to the production of renewable energies including biofuels have been characterized by strong dynamics in recent years (Costantini *et al.*, 2013). Third, biofuels represent an example of a sector characterized by cross-cutting applications of innovations which would be difficult to describe by referring to standard patent classifications. This happens because biofuel-related inventions are usually suitable for a variety of purposes since they are based on very complex domains such as chemical transformation or biological processes (Cherubini *et al.*, 2009).

Our analysis shows that a typical IPC class-based tool, such as the biofuels codes included in Green Inventory (GI) as the official list developed by the World Intellectual Property Organization (WIPO) and widely adopted to study eco-innovation dynamics, is a weaker tool of analysis with respect to our selection approach which results in the BioPat database. Moreover, we claim that this result is mainly related to the complex nature of knowledge structure in this specific technological domain. In this respect, the paper provides a discussion of the relevance and implications of knowledge complexity in the biofuels sector is

The rest of the paper is organized as follows. In Section 2 we discuss our methodology with respect to existing approaches. In Section 3 we sketch the steps of the proposed methodology and we describe how we have applied it to collect patents in the biofuels sector. In Section 4 we offer a description of the dynamics of patenting activity, while in Section 5 we focus on the characteristics of knowledge structure in the biofuels sector. Section 6 discusses the economic and policy implications deriving from the

²Nonetheless, very few contributions have analysed the biofuels sector and its technological development in a comprehensive approach (Costantini and Crespi, 2013; Karmarkar-Deshmukh and Pray, 2009).

characterization of knowledge structure in the biofuels domain. Finally, Section 7 provides some conclusive remarks and some possible future research developments.

2. The use of patents for mapping technological knowledge in complex sectors

A patent usually has a very standard object: a chemical formula, a variation or an improvement in a natural process or a mechanical, artistic or even immaterial device. Once registered, the patent receives a code that classifies its content. During the last century, the increasing amount of patents registered daily worldwide and the great number of interactions among patent offices made the adoption of a uniform system of patent classification necessary.

Classification is fundamentally a technical problem referring to how patent data are allocated and organized in national and international databases. Every patent office provides each patent with an internal code that includes a reference to the object of the invention. An international code (IPC) is associated with the internal code which allows patents to be classified by following hierarchical criteria based on chemical and technological principles, only occasionally related to specific sectors.³

The IPC divided the universe of patents into eight sections, 20 subsections, 118 classes, 624 subclasses and over 67,000 groups (of which approximately 10% are main groups and the remainder are subgroups). The different sections in IPC allow distinctions to be made between patents belonging to classes that not necessarily fit into specific technological domains or economically identifiable categories (such as a specific sector or a specific product). Hence, the resulting classification is only of limited usefulness when a specific sector which does not fit the criteria used in the classification is under scrutiny, so that IPC is not suitable when the focus of the research does not match an existing section.⁴

In order to be able to fully exploit patents information for mapping technological knowledge related to a specific economic activity, it is necessary to associate a patent to

³ According to the 2011 version of the IPC guide, “the Classification, being a means for obtaining an internationally uniform classification of patent documents, has, as its primary purpose, the establishment of an effective search tool for the retrieval of patent documents by intellectual property offices and other users, in order to establish the novelty and evaluate the inventive step or non-obviousness (including the assessment of technical advance and useful results or utility) of technical disclosures in patent applications” (WIPO, 2011, p. 1).

⁴ As an example, IPC contains a whole section for weapons. When interested in the investigation of innovation in the weapons industry, researchers should refer to the IPC code F41. On the contrary, IPC does not contain an *ad hoc* code for the cosmetic industry (that would probably fall under several categories in the chemistry section).

one industry. However, patents and economic classifications follow different criteria: while the IPC system has been established to provide an organic classification for novel technical characteristics of an invention, industry classifications seek to disaggregate goods into commercially and economically meaningful categories.

The innovative content of a patent is more often ascribed to production factors rather than final goods. In the case of univocal applicability of the content of a patent and clear definition of the domain of the industry, patent data and economic data can communicate to each other (e.g. a patent about new loader for guns do not have many cross-cutting applications and can be clearly matched with the weapon industry). On the contrary, when the patent content refers to an invention that may find several applications in the production of different final goods (goods belonging to different sectors but sharing one or more production inputs), industry and patent classifications do not coincide. An obvious example is represented by inventions related to chemical components which are suitable for the production of a variety of final goods belonging to different economic sectors, such as the case of lipids to produce tractors' fuel as well as baby-care products.

The difficulty of linking patents to specific economic activities varies according to the nature of the sector under scrutiny and to the complexity of technological knowledge related to it. Knowledge complexity entails two main aspects. First, the generation of new knowledge is conditional on the integration of different bits of complementary knowledge that are inputs into the knowledge production process (Antonelli, 2003, 2008; Gibbons *et al.*, 1994; Loasby, 1999; Nooteboom, 2000). Secondly, the downstream complementarity of any bit of knowledge implies that specific elements of technological knowledge may apply to a great array of new products and processes in different and new technological fields (Antonelli, 2013). Hence, the greater is the cross-cutting nature of technological knowledge and its complexity, the more difficult is to correctly identify patents related to specific sectors.

Several attempts have been made to deal with this issue. Here we analyse three main categories representing alternative classification methods. The first category of attempts consists of a top-down approach that relies on the IPC class system and aims to define its content. Accordingly, this approach consists in the exploitation of the linkages between classes assigned to the same patent by considering those appearing together as a class family where IPC codes are seen as technologies. So, different IPC codes can refer to sub-technologies that a patent covers and they can be interpreted as a class family when they

frequently appear together. An empirical application of this approach is represented by the IPC co-occurrence method (Antonelli *et al.*, 2010; Krafft *et al.*, 2011). However, co-occurrence could be efficiently used when the investigated sector is already defined *per se* (as in weapons) or when it is relatively easy to identify.⁵ In other cases, such as biofuels or other cross-cutting sectors, the use of this method may result in rough and unpredictable results.

A second category gathers attempts which try to identify the classes that are suitable for containing a patent related to a specific economic sector (OECD, 2011). Several attempts to define a link between patents and different industries have been carried out in the last 50 years, even before the creation of the IPC system. Schmookler (1966) classified US patents according to the industry where the patent would find a reasonable application. Other important attempts are represented by: the MERIT Concordance that matches IPC subclasses to 22 industrial classes based on ISIC codes (Verspagen *et al.*, 1994); the Yale Technology Concordance (Kortum and Putnam, 1997), which links the IPC code to the Canadian Standardized Industrial Classification system; the OECD Concordance (Johnston, 2002), that elaborates more on the translation of the IPC into the ISIC; the European Commission Concordance (Schmoch *et al.*, 2003) which assigns 625 IPC subclasses to 44 manufacturing sectors associated with ISIC codes.

However, when economic analysis refers to highly detailed sectors, a more refined concordance is needed. To this aim, a step ahead within this second category is represented by the so called patinformatics, a new method to analyse patents data in order to disclose relationships between patents and technological fields by mean of text analysis and text clustering software (Trippe, 2003). For instance, following this line of research, Lybbert and Zolas (2012) proposed the Algorithm Link with Probabilities concordance which uses data mining and indexing as well as a probabilistic matching approach to provide links between IPC and industrial classification. According to this study, inventions described in patents may have a very cross-cutting application in several sectors and one IPC class may contain patents which can be useful for numerous economic activities falling into different

⁵ This is the case for sectors that are sufficiently wide to cover an entire section of the IPC such as information and communication technologies, biotechnologies or nanotechnologies, where ad hoc selection tools are also available in search engines such as OECD Patstat. Patents related to these industries are scattered in more than one section but could be easily identifiable. Another example is given by the fertilizer industry. Fertilizers belong to class C05, an *ad hoc* category in the chemistry section, but fertilizer distributors or fertilizing practices are included in the A01 class which contains patents related to agricultural activities. Thus, a patent referring to a new fertilizer and its distribution practice would fall under both classes, ideally making a class family that identifies the fertilizer industry.

industrial categories. In this sense, concordance between IPC and industrial classification systems may be of a poor usefulness when it comes to analyse sectors which use inputs or processes suitable for the production of a variety of goods.

With respect to environmental technologies, a classification example falling into this second category is represented by the GI classification, developed by the IPC Committee of Experts in order to facilitate searches for patent information related to Environmentally Sound Technologies (ESTs), as listed by the United Nations Framework Convention on Climate Change (UNFCCC).⁶ ESTs are currently scattered widely across IPC in numerous technical fields. The GI allows all ESTs to be collected in a single, organized repository. Following the IPC system, the ESTs are presented in a hierarchical structure. According to the WIPO website, two steps were followed to create the GI. First, a list of technologies was completed by the UNFCCC as a basis for the work of the IPC Committee of Experts who identified the related IPC classes. In order to identify the IPC classes correctly, the experts used the IPC Catchword Index, the IPC term search and their expertise in the relevant technical areas in order to collect all the green-related IPC classes under the specific category. Hence, GI consists of a list of IPC classes characterized by the fact that they are suitable for containing patents related to a green technology.⁷

A third category of attempts for patent selection and classification is a keyword-based methodology for the identification of patents strictly related to the specific sector under scrutiny. This method, by directly focusing on patent documents tries to go beyond the boundaries of the IPC codes to identify patents related to a specific economic sector. Within this framework of analysis, several researchers have developed different methodologies essentially based on the exploitation of catchword tools and literature scrutiny. The last decade's literature on keyword analysis basically consists of selections of words from already existing keyword lists or the extraction of keywords from titles and, at least, abstracts of patents and scientific publications.

⁶ Personal correspondence with WIPO Green Inventory Project staff.

⁷ The most recent effort in classifying environmental-friendly technologies is given by the Cooperative Patents Classification (CPC), consisting in a new common classification scheme initiated as a joint partnership between USPTO and EPO. The CPC is operative at the two offices by 1st January 2013 and replaced ECLA at the EPO from that date; it will replace the USPC at the USPTO during a period of transition to 2015. Although CPC seems to be promising since it is more focused on punctual technologies and it has an ad hoc environmental-friendly technologies class, named Y02, it is also true that it is still based on classes mainly, but more importantly it is still under validation and it is a reliable search tool only for recent patents.

More specifically, the literature followed three main approaches so far: i) co-word study based on the keywords proposed by experts (Looze and Lemarie, 1997); ii) use of descriptors chosen by professional indexers employed in patent offices and search engines (Coulter *et al.*, 1998); iii) extraction of keywords from titles and abstracts of patents (Corrocher *et al.*, 2007).

These three approaches are characterized by strong differences. The first two are based on an attempt to describe the sector using words that are commonly considered sector specific, whereas the last one seeks to eliminate the arbitrariness of the selection process. In fact, Corrocher *et al.* (2007) pointed out that the *ex-ante* selection of the keyword procedure may reflect preconceptions, different backgrounds and points of view of the words' selectors and differences in the training and backgrounds of professional indexers. As a result, the authors proposed to use the most frequent sequential triples of words without imposing a priority constraint on the selection of keywords to identify technological domains that can be compared with the existing IPC technological classes. Unfortunately, the method that looks *ex-post* for the triples of words is more appropriate when it comes to investigating a sector that is sufficiently wide to cover an entire section of the IPC or when patent novelty is based on engineering contents which are more likely to fit into *ad hoc* classes, two characteristics that do not always fit with the investigated sector.

In order to solve such criticisms, we propose here an original keyword-based selection method, precisely developed for complex and cross-cutting domains, which are far from being well defined and described in their technological and dynamic features by already existing selection and classification methods. As a test for the validity of this original method, we implement it in the biofuels sector as a case study, developing a patent database here called BioPat.

3. A keyword-based methodology for patent analysis: an application to the biofuels sector

3.1. General description of the methodological framework

We first provide a brief description of the overall methodology and the consequential steps followed to identify and classify relevant keywords for patent selection through the analysis of the specific sector under scrutiny. We then describe keyword validation, patent collection and final database validation. The whole procedure has been summarized in Figure 1.

Figure 1 - The keyword-based methodology for cross-cutting technologies

In order to build a sector specific patent database, it is necessary to define a meaningful keywords list which is at the same time narrow enough to confine patents in the analysed sector and sufficiently complete to gather all required information. Hence, the first step of

the proposed methodology consists in an in depth analysis of the sector under scrutiny in order to identify relevant activities that take place in the specific technological innovation system (Hekkert *et al.*, 2007). Such a sector analysis is aimed at qualifying the relevance of the sector under scrutiny, identifying the main trends at the international level, the dynamics of market supply and demand, the characteristics of the sectoral policy framework, and all the information needed to depict the global context surrounding it.

Building on this analysis, the second step is designed to identify the main final goods produced in the investigated sector, in order to define a fix set of keywords used to query the patent database. The fix set of keywords are general keywords unequivocally linked to the sector, mainly including commercial and scientific names of the final products.

The third step allows the definition of the relevant aspects that characterise the sector, allowing for a logic categorization of keywords in meaningful groups and leading to the identification of the database structure. Once identified the specific object of the analysis and the database structure, this step allows for the construction of the non-fixed list of keywords, which mainly relies on the scrutiny of relevant literature with a specific focus on single phases of the production value chain.

Since the subjectivity of the selection process could represent an issue it is important to design the selection process as objectively as possible. Hence, the fourth step consists in the validation of the set of keywords originated from literature scrutiny by means of catchword instruments and experts' opinion.

In the subsequent step, in order to collect data, the fixed and the non-fixed set of keywords are used to query a patent search engine. Finally, the outcome of the downloading procedure is validated by a panel of experts, by comparing the relevance of the downloaded patents in terms of coherence with the investigated sector and with respect to patents downloaded with already existing classification methods.

In the following Sections we describe the methodology in single details for each step by applying it to the biofuels sector.

3.2. Sector overview

The global production of biofuels has been growing steadily over the last decade (IEA, 2011). Today, biofuels provide around 3% of total road transport fuel globally (on an energy basis) and considerably higher shares are achieved in some countries. Brazil, for instance, met about 23% of its road transport fuel demand in 2009 with biofuels. According

to British Petroleum statistics (BP, 2012) and EIA-DOE (2013), the major biofuel producers are the US with a 44% of global production in 2011, followed by Brazil with 27% and the EU with a share of 18% of world biofuel production. Figure 2 provides a description of the biofuel market concentration for biodiesel and bioethanol respectively, where the US and Brazil are leaders in producing bioethanol whereas the EU is specialized in the production of biodiesel, obtained mainly from rapeseed, sunflowers and soya bean.

Figure 1 – Biodiesel and bioethanol production distribution (2011)

Source: own elaboration on BP (2012) and EIA-DOE (2013)

Biofuel production costs vary significantly across the main producing countries. Brazil has the highest competitive advantage for bioethanol and is the only producer, based on the current state of technology, which can compete with fossil fuel production costs without public support. All other producing countries have to adopt some form of policy intervention in order to push production volumes.

By looking at recent IEA forecasts, biofuels can provide up to 27% of world transportation fuel by 2050 (IEA, 2011). With the transportation sector growing considerably and demand for transport fuels raising globally, the IEA assesses biofuels as one of the key technologies for reducing GHG emissions and reducing dependency on liquid transport fuels in the medium run.

Moreover, biofuels represent a key point in countries' strategies for energy security, with both production and consumption characterized by pervasive and large subsidies all over the world. The massive increase in biofuel production is controversial because of three

major aspects: the effective competitiveness of their production, the energy vs. food conflict issue and the transition towards new technological domains for biofuel generation.

While in a limited number of favourable circumstances producing energy from biofuels can be cost competitive today, in many cases policy incentives are currently needed to offset cost differences between biofuels and fossil fuels. Such support is justified by the environmental, energy security and socio-economic advantages associated with biofuels, but should be introduced as transitional measures leading to cost competitiveness in the medium term (IEA, 2011).

Furthermore, current production processes and certain conventional biofuels have been criticised for causing deforestation and adding pressure to agricultural land needed for food and fodder production. The initial interest in biofuels based on agricultural crops, commonly defined as first generation biofuels, is now partially reduced by several drawbacks of these technologies (above all, biofuels based on agricultural crops suffer from production uncertainty due to possible adverse climate conditions).

Finally, as recommended by IEA (2011), there is an urgent need to ensure that domestic biofuel public support systems foster the transition towards fully sustainable biofuels, including advanced biofuel technologies, providing support mechanisms to ensure that the new technologies reach full market deployment. According to Carriquiry *et al.* (2011), the production of advanced generation biofuels from various feedstocks could significantly contribute to the future energy supply mix, where production cost is a major barrier to its commercial production in the near to medium term. Depending on the type of biofuel, the cost of cellulosic ethanol is found to be two to three times higher than the current price of gasoline on an energy equivalent basis. The average cost of biodiesel produced from microalgae is seven times higher than the current price of diesel. While significant cost reductions are needed for all types of advanced generation biofuels, the critical barriers are at different steps of the production process.

As already mentioned, while the biofuels industry has shown a positive trend over the last decades, concerns remain as to whether this industry would survive in the absence of fiscal incentives. Furthermore, biofuels are considered environmental-friendly, but the global impact is still ambiguous. Advanced generation biofuels seem to address some of the most outstanding issues and technological improvements are gaining importance in the process of transition towards sustainability. In light of this, it appears to be crucial to

properly map the technological evolution of such a complex domain in order to correctly identify different paths leading to new generations of technologies.

3.3. *Definition of the main products within the sector*

Biofuels cover a wide range of fuels derived from biomass which is material of biological origin excluding material embedded in geological formation and/or transformed to fossil fuels. Biofuels include:

- liquid biofuels: fuels and bio additives such as bioethanol, biodiesel, biobutanol, biomethanol, bioETBE (ethyl tert-butyl ether), bioMTBE (methyl tert-butyl ether), biogasoline and combustible oils produced by plants;
- gaseous biofuels such as biogas consisting mainly of methane and carbon dioxide produced by the process of anaerobic digestion of biomass;
- solid biofuels such as wood pellets, wood chips and charcoal, including char-briquettes.

As previously discussed, liquid biofuels have been acknowledged as the main substitute to fossil fuels in the transport sector and have been characterized by pervasive and large subsidies all over the world. Hence, this study focuses on this sub-category as the most relevant one within this sector. More specifically, in this work we rely on the definition provided by the European Union (EU) Directive 2003/30/EC (European Parliament, 2003) focusing on the two most commercially available liquid biofuels: biodiesel and bioethanol.

3.4. *Definition and classification of relevant sector categories and database structure*

According to the EU Directive, biodiesel and bioethanol production is essentially based on the conversion of a set of *raw materials* into fuel (for biodiesel, vegetable oils, waste oils and animal fat; for bioethanol, starch or sugar-rich biomass) through specific *transformation processes* (for biodiesel, transesterification by using alcohols; for bioethanol, fermentation).

As early evidenced, the evolution of different *technological generations* represents a crucial aspect in the analysis of the sector. For this purpose, two major categories can be envisaged: i) conventional biofuels (first generation) are produced on a commercial scale following well-understood technologies; ii) advanced biofuels (from second to fourth

generation) are comparatively immature and there is much room for further deployment (IEA, 2008; Janda *et al.*, 2012).⁸

As a consequence, technologies in this complex sector can be classified according to multiple criteria. In this work we have selected three main categories able to map and classify technologies according to the state of the art and the sector analysis, i.e. “raw materials”, “transformation process” and “technological generations”.

The combination of the multiple classes where one patent might be classified allows us to design the structure of the database that can be exemplified as in Figure 3. The classification of each keyword and consequently each patent in different categories allows creating a structure of the whole technological system, according to the description obtained through the sectoral analysis. In this way we obtain a patent classification method which is rather more flexible than a standard IPC codes-based approach and which may be useful for several purposes of analysis. This classification method is also valid for highlighting those technologies which might present a transversal use. As a mere example in Figure 3 the keyword “*Dunaliella Tertiolecta*” which has been selected for algae-related biofuels belongs to two categories, three and four, because it is suitable for producing both bioethanol and biodiesel.⁹

Once identified the specific object of the analysis and the database structure it is possible to identify the list of keywords, which mainly relies on the scrutiny of relevant literature with a specific focus on production value chain. We started the selection process by identifying keywords to be classified in relevant categories from recent scientific literature related to the biofuels sector and in particular to the value chain production of biodiesel and bioethanol. The search for keywords was divided into two different fields. The first one was dedicated to a search for “raw material” keywords where a significant number of technical and scientific papers were analysed in order to pick out the terms describing the biomass used (or potentially used) to produce biofuels. The second one consisted of an accurate description of the transformation process currently known in biofuels production,

⁸Conventional biofuels are: bioalcohols (especially ethanol) and biodiesel derived from food crops rich in sugar (such as sugar cane and sugar beet) or starch (such as wheat and corn) or vegetable oil (such as palm oil, soya bean and rapeseed). They also include products obtained from a chemical combination of biological molecules with fossil molecules, such as ETBE. Advanced biofuels are: bioalcohol based on ligno-cellulosic biomass, HVO (Hydrotreated Vegetable Oil), BTL diesel and Bio-SG (bio-synthetic gas), algae-based biofuels, sugar conversion into diesel-type biofuels, artificial photosynthesis reactions (solar-to-fuel) and hydrocarbons obtained from genetically modified organisms (GMO).

⁹The full list of keywords used for BioPat is available upon request from the authors.

including pre-treatment processes, chemical agents involved in the process and technical instrumentation used in it.

Figure 3 - Exemplificative structure of BioPat

3.5. Keywords validation

Then we tested if the selected keywords matched the keywords automatically suggested by Scopus, as the Scopus research tool allows checking if patents that contain the selected keywords actually exist. In this way the outcome of the selection process derives from an iterative procedure, which allows results from scientific articles to be compared with patent results in Scopus.

Finally we submitted the keywords list to the Italian National Agency for New Technologies, Energy and Sustainable Economic Development (ENEA) experts, thus obtaining information on the consistency of keywords with the specific production phase and also with the quality of the innovative contents of the selected technology (conventional or advanced generation).

3.6. Patent collection

According to the IPC terms of reference, patent novelty is usually classifiable following two main principles: a patent can be characterized by engineering content or by bio-chemical

content. The latter mainly applies for the biofuels sector and represents a key explanation of the cross-cutting nature of the sector in terms of dispersion in different IPC classes. In light of this, we decided to expand the use of keywords to the “patent descriptions” and “patent claims” fields in order to exploit the possibility of catching all patents that have a potential, and not necessarily direct, function in the biofuel production process.

The patents were downloaded using Thomson Innovation (TI), a single, integrated solution that combines intellectual property, scientific literature, business data and news with analytic, collaboration and alerting tools in a robust platform.¹⁰

All process-specific and raw material keywords were jointly used in TI with a more general keyword (such as bio-diesel, bio-ethanol, bio-fuel) in order to exclude patents that share the same raw materials or transformation processes (in particular, pharmaceuticals and cosmetics are strongly related to the biofuels sector). Furthermore, we carried out an additional search using general keywords related to biofuels in the “applicant” field. We focused our research on EPO, USPTO and WIPO.

With regard to raw material keywords, the search on TI was carried out as follows: by using Boolean operators “OR” and “AND” we selected all patents (kind code A1 and B1 from 1/01/1990 to 31/12/2010) containing the keywords among a fixed set of general keywords introduced with the Boolean operator OR (at least one of the term must appear) and a more specific one (added one by one to the fixed set), with the Boolean operator AND. Multiple words were added in quotation marks.¹¹

With regard to the transformation process, keywords were used with the same sequence of fixed terms representing the general name of biofuel products (with Boolean OR, kind code A1 and B1 from 1/01/1990 to 31/12/2010) and a second level containing all general terms (added one by one with the Boolean AND) for production process such as

¹⁰The list of keywords was built on different levels. First, we tested searches with selected keywords in order to verify the response of the TI database to the inputs and exclude the possibility that the keyword list was inadequate. Second, we created sub-groups of terms standing for the same keyword. For example, raw material terms derived from literature scrutiny were usually in English, while some inventors tend to use scientific Latin or Greek names. Scientific names were attributed to the same original (English) keyword. Moreover, the TI search engine also allows symbols to be used as a means of catching variations of the same word as well as plurals. For instance “fermented sugar” was entered as “ferment* sugar*”, catching in this way a combination of different words such as “fermenting sugars” or “ferment sugar cane” and so on.

¹¹ For example: *Nannochloropsis* (an alga) AND “renewable *ethanol” OR “green *diesel” OR *methanol OR *butanol OR biomethane OR bio-methyl-ether OR “Synthet* fuel*” OR biodiesel OR “renewable fuel*” OR biofuel* OR etc.

transesterification, Fischer-Tropsch, anaerobic digestion and so on.¹²

Following our methodology, the patents downloaded through TI amount to 1,293,197 patents across three patent offices(EPO, 62% USPTO, 18% WIPO)in the period 1980-2011.¹³Then, using this initial information, we tried to make the database suitable for our purposes. In order to link each patent with the nationality of a specific applicant, we looked for country codes in the variable “assignee address” obtaining information on 35 countries.¹⁴ This procedure is particularly effective for EPO patents because the address contained in the variable is consistent in all records. The procedure allowed us to identify the nationality of the 97.3% of BioPat patents in EPO, but only 47.6% in USPTO.

3.7. Database validation

In order to compare the accuracy of BioPat with respect to GI, we asked a team of experts from ENEA to check the coherence of patents classified by BioPat methodology and GI with regard to EPO patents.Among the ESTs, for our purpose, we considered 44 IPC codes (40 subgroups and four subclasses) that identify the whole biofuels sector (see Table A1 in the Appendix for a complete list of IPC codes) within the GI.

At present, the GI website does not display any statistics on the effective number of patents in each class that are also coherent with the object assigned. Hence, in order to shed light on the accuracy of the GI classification, we validated a sample of patents included in the IPC classes indicated in GI. Additionally, we asked the group of experts to distinguish between patents with a direct application in the biofuels production process and an indirect one. We downloaded the description field of the whole universe of patents belonging to these classes, from which we randomly selected a 1% sample.

¹²After that, we verified if the downloads could represent a significant part of the whole universe achieved using only the general keywords. The huge specific outcome obtained by using general keywords strongly reinforces the choice of working with selected specific keywords rather than working on a broader definition of biofuels (e.g. Karmarkar-Deshmukh and Pray, 2009) or on IPC codes.

¹³ Note that 1,293,197 patents refer to the sum of all patents downloaded (general set +specific keyword for each category) for all kind codes. Such methodology allows for several repetitions: in fact one patent might be downloaded more than once if it contains more than one keyword.

¹⁴ To the best of our knowledge, 35 represents the highest number of countries considered so far in an environmental technology field in international scientific contributions. Our sample includes: AE (Arab Emirates), AR (Argentina), AT (Austria), AU (Australia), BE (Belgium), BR (Brazil), CA (Canada), CH (Switzerland), CN (China), DE (Germany), DK (Denmark), ES (Spain), FI (Finland), FR (France), GB (Great Britain), GR (Greece), HK (Hong Kong), ID (Indonesia), IN (India), IT (Italy), JP (Japan), KP (North Korea), KR (South Korea), LU (Luxembourg), MX (Mexico), MY (Malaysia), NL (The Netherlands), NO (Norway), NZ (New Zealand), PT (Portugal), RU (Russia), SE (Sweden), SG (Singapore), TH (Thailand), US (United States of America).

Then, the experts were asked to validate the same GI classes filtered with our keywords. The sample was built as follows: we took the EPO patents in our database, selected the patents that showed at least one IPC class belonging to GI, eliminated the duplicates and randomly extracted 1% of the selected patents.

The results of the validation are summarized in Table 1, which shows that the keywords selection method applied to GI classes allowed us to double the percentage of patents related to the sector. Moreover, the results of the validation procedure also showed that the share of patents directly related to the investigated sector increases when GI classes are filtered by the selected keywords.

Table 1 - Validation of BioPat for EPO

Type	GI	% of biofuel-related patents	GI filtered by keywords	% of biofuel-related patents
Direct application	5%	28%	15%	40%
Indirect application	14%	72%	23%	60%
Total	19%		38%	

Source: own elaborations on BioPat

4. Technological patterns in the biofuels sector: evidence from BioPat

According to the purpose of this paper, we investigate differences in technological patterns of the selected case study by comparing the only IPC class-based methodology already available in the literature for the biofuels sector (GI) with our keyword-based classification method.¹⁵ Very broadly, the use of BioPat allows for a description of the evolution of patenting activity registered at the EPO for 35 countries.¹⁶

The first comparison refers to the total number of patent applications to EPO for all available countries in the period 1990-2008 deriving from three classifications (Figure 4): i) the total number of patents available in BioPat (without double counting); ii) the total number of patents available in the biofuels classes given by GI; iii) the overlapping measure

¹⁵The exact class codes indicated in GI describing the biofuels domain are reported in Table A1 in the Appendix.

¹⁶ We acknowledge that EPO has a very high rate of application refusals and patent applications would not properly reflect the production of new knowledge. Nevertheless, patent applications offer a clear picture of the productive innovative effort. Moreover, working with granted patents obliges us to cut the series considering the years required for the patents acceptance process (the lag between application date and grant date), as suggested by EPO online forum (Johnstone *et al.*, 2010; Picci, 2010), losing about five years of information. Moreover, patenting procedures are usually quite slow and patents granted in 2010 do not precisely reflect innovation activities in the earlier years, especially since applicants tend to apply for patents in a national office before applying to an international one (De Rassenfosse *et al.*, 2012).

between BioPat and GI, given by the patents that belong to the two databases. We have considered this sub-period with respect to the overall available time span (1980-2011) for two reasons: first, number of patents included in GI classification in the first decade (1980-1989) is very low with respect to BioPat; second, the last three years (2009-2011) present statistics which are incomplete with respect to the effective number of applications due to structural delay in patent offices database updating.

As represented in Figure 4, the patents contained in GI which are part of BioPat (BIOPAT-GI line) are simply a subset of GI, showing a common trend and a distance in number of patents which is quite stable over time. On the contrary, the overall set of patents included in BioPat follows a similar trend only in the time frame 1995-2001, while it shows a positive and high growth rate after 2001, exactly in the opposite direction of GI. This specific result deserves some attention, since it perfectly coincides with the introduction of the most pervasive public supports to biofuels production across main biofuels producing countries. When market opportunities increased thanks to the introduction of mainly demand-pull public policies, private firms started to explore how to exploit them by developing complementary activities well revealed by the growing outputs in terms of technological achievement. This strong dynamics is well caught by the keyword-based methodology which is rather more flexible than a fix set of IPC classes.

Figure 4– Patent applications in EPO for different classifications (1990-2008)

Source: own elaborations on BioPat

Diversity in number of patents is also reflected by the geographical distribution of applications. Figure 5 and 6 represent the share of the three main countries (European

Union, Japan, the United States) and the Rest of the Word (RoW) in patent applications in GI filtered by keywords of BioPat and in total BioPat, respectively.

The most diverging result is given by the stable share of the three main countries over time in the case of BioPat (Figure 6), compared to a quite higher share of the RoW, especially in the first decade in GI filtered (Figure 5). It is also intriguing to notice that the lion share in BioPat is taken by Japanese applications, which are above the sum of applications by EU and the US. The increase in shares of EU and Japan is more pronounced after 2001, revealing that these two regions are those with the highest growth rates in applications.

In addition, the categorization of patents using specific groups of keywords developed within the methodology here proposed allows us to shed light on the specific composition of these trends. In particular, here we present the dynamics of total patent applications of EU, Japan and the US for two sub-categories (Figure 7): i) raw materials which belong to the food sector which are also classified as conventional generation biofuels; ii) algae-related biofuels which are to be considered as applications belonging to the advanced generation category.

Figure 5– Share of patent applications in EPO for BIOPAT-GI- (1990-2008)

Source: own elaborations on BioPat

Solid lines refer to BioPat while dashed lines to the overlapping subset (GI in BioPat). If we compare conventional with advanced generations in the two samples, it is worth noting that while difference in patent numbers between the two categories is relatively small in the

overlapping subset (dashed lines), the divergence is rather higher when comparing these categories into the full BioPat database.

Figure 6— Share of patent applications in EPO for BIOPAT - (1990-2008)

Source: own elaborations on BioPat

Figure 7— Comparison between BIOPAT and GI applications to EPO for selected categories and countries (1990-2008)

Source: own elaborations on BioPat

Note: number of patents refers to the sum of application to EPO by European Union, Japan and the US.

Considering that the food category corresponds to the sector branch with more consolidated knowledge in a wide range of well-established production activities, the

keywords methodology allows us to better capture those complementary bits of knowledge that characterize the more complex sub-domain. On the contrary, the relatively newer and narrower branch represented by algae-related patents is still well described by a class-based methodology. Finally, the divergence for the food sub-domain is particularly marked after 2001, when pervasive support policies have involved the whole sector and sustained mainly conventional biofuels. As a result, the faster is the expansion of an economic sector towards complementary activities, the more appropriate appears to be a flexible keyword-based methodology with respect to a fixed class-based tool.

5. Knowledge complexity in the biofuels domain

The analysis of the characteristics of knowledge structure in the biofuels sector can be further developed on the basis of the information contained in BioPat. In particular, the implementation of an articulated methodology for patent selection was motivated by recognizing that the greater is the cross-cutting nature of technological knowledge induced by knowledge complexity, the greater is the difficulty to correctly identify patents related to specific sectors. Such a difficulty is associated with the fact that relevant technological knowledge may be dispersed on a great variety of patent classes and that it may be applied to a great and even increasing array of different purposes (Antonelli, 2003, 2008). With respect to these issues the analysis of BioPat data can provide interesting insights.

First, starting with the analysis of EPO applications in BioPat included in the GI list, it is possible to confirm the cross-cutting nature of the biofuels sector by looking at the number of IPC classes associated with biofuel patents. In this respect, we observed that 6,835 patents have been placed in more than 6,000 IPC subgroups over 30 years (1980-2010), that is to say, only few biofuel patents have been classified in the same subgroup. To the extent that an IPC class identifies a specific technological domain, this preliminary evidence suggests that technological knowledge in the biofuels sector exhibits a cross-cutting nature. This result is confirmed when we adopt a less detailed class specification (4 digit categorization) since we found that biofuel patents have been placed in 233 IPC subclasses since 1980.

Second, the gathered information allows us to verify if the identified cross-cutting characteristic displays a stable or an increasing pattern over time. In particular, considering the increasing patenting activity in the biofuels sector after the year 2000, we are interested in understanding whether the increase in number of patents occurred in the same patent

classes in which biofuel-related patents were identified up to the year 2000 (which would suggest an intensification of the inventive activity in closely related technologies), or if it occurred mainly in IPC classes in which biofuel-related patents were not selected until the year 2000 (suggesting a further increase in the cross-cutting nature of technological knowledge in this sector). In so doing we followed Strumsky *et al.* (2012) who suggested that technological change can be analysed by studying the dynamics of technology codes (in our case the IPC classes) over time. Changes in the dimension of each code (i.e. number of patents granted to each class) and in the number of codes assigned to a common technological domain (number of IPC classes) allow revealing how technological capabilities and functionalities modify in a selected field. In this respect, we find that starting from January 2000, 47 additional IPC subclasses (referring to 42 different patents) appear in the database, which shows that the cross-cutting nature of technological knowledge in the biofuels sector increases over time.

As a robustness check, we have also considered the possibility that this increase in IPC class range may be merely related to the introduction of new classes in the IPC system which is periodically updated by WIPO.¹⁷ In order to be sure that the increase in IPC classes *per se* does not entirely explain the increase in IPC classes in BioPat, we have checked whether the 47 new classes existed in the International Patent Classification before the year 2000, finding that 44 of the 47 classes were already present in IPC Version 6 (in force before 2000), thus confirming the expansive dynamics of the biofuels sector in terms of the increasing variety of patent classes involved. This finding is more generally confirmed by comparing the variation of BioPat IPC classes with the variation of IPC classes in general for three sub-periods covering the years 1995-2009 (Figure 8).¹⁸ From bars comparison in Figure 8 it is worth noting that the former has been systematically higher in the whole period, suggesting that the observed pattern of increasing diversification in patent class groups in BioPat is mainly related to the peculiar evolution of the knowledge structure in the sector under scrutiny and not merely induced by the growth in IPC groups *per se*.

¹⁷ The IPC Versions have been modified seven times from September 1968 to December 31, 2005 (labelled Version 1, Version 2, etc.). Since 2006, the IPC has been frequently revised (sometimes more than once per year) and each edition is indicated by the year and month of its entry into force.

¹⁸ We computed variation in the number of IPC classgroup, which is the higher detailed class specification available for the whole period (subgroups did not exist in previous versions of the IPC system). IPC version 7, version 8 and the annual review for the period 2006-2009 have been considered.

Third, in order to characterize the dynamics of technological knowledge in the biofuels sector, in addition to quantitative considerations related to the amount of new IPC subclasses, we propose a more qualitative analysis by looking at the IPC subclass objects.

Figure 8 - Variation in class group number in the IPC system and in BioPat

Source: our elaboration on BioPat and WIPO statistics

According to our sectoral analysis, the observed increasing cross-cutting nature of knowledge in the biofuels sector can be in fact related to a transition towards large scale production of energy through biofuels which calls for a number of technical issues to be solved if biofuels are to become a reliable, sustainable and relevant energy source. These go from the identification of alternative feedstock production to instruments for boosting yields (pesticides, fertilizers. etc.), the efficiency of fuel processing, distribution and final use. The number of technological domains involved in solving these problems is thus increasing.

As shown in Table 2, more and more distant bits of knowledge are contributing to this process. Knowledge related to vessels for containing or storing liquefied or solidified gases or compounds containing metals is needed to solve shipping difficulties linked to the different organic nature of biofuels or to improve engines adaptability.

Technological knowledge in the domain of underground or underwater structures emerges as relevant for algae cultivation as shown by the appearance in 2004 of the associated IPC class. Knowledge related to computer systems based on specific

computational models is exploited to simulate biochemical reactions in order to discover new raw materials for the production of biofuels or to improve the efficiency of fuel processing. Knowledge in the chemical-based sectors such as food supplements, pharmaceuticals and cosmetics is increasingly exploited in the hunt for new feedstock suitable for the production of biofuels.¹⁹ In parallel, complementary bits of knowledge are shown to be needed for the development of recent techniques adopted for the co-production of sugar and bioethanol, confirming that the full exploitation of by-products or co-products (for instance in feed production) is one of the key elements of the biofuels industry as indicated by the international biofuels technology roadmap (IEA, 2011). This result might constitute food for thought for further research and for policy design, revealing a strong and growing linkage between the biofuels market and other already existing markets, thus suggesting that policy design as well as policy assessment should necessarily take into account the nature and the direction of innovation patterns in such a dynamic sector.

Table 2 – Selection of new IPC classes included in BioPat/GI from 2000

Year	Patent	IPC	Class object	Patent object	Issue
2000	EP1167861A1	F17C	Vessels for containing or storing liquefied or solidified gases	Shipping facilities for biofuels	Distribution of fuels
2000	EP1048614A1	C01G	Compounds containing metals	Additives for biodiesel	Adaptability to existing infrastructure
2002	EP2226739A1	G06G	Computer systems based on specific computational models	Design biochemical reactions	Improving production efficiency
2004	EP1695956A1	E02D	Underground or underwater structures	Algae cultivation	Production of new generation biofuels
2005	EP1778851A1	C13J	Extraction of sugar from molasses	Producing ethanol and sugars for food purposes	Co-production
2008	EP2245166A1	A61K	Preparations for medical, dental, or toilet purposes	Cross-cutting use of lipids	Expanding production possibilities

These results may have important implications in terms of the appropriateness of the methodology adopted in this study. The increasing cross-cutting nature of knowledge in the biofuels sector as evidenced by the high rate of appearance of new technological classes may make the use of a limited set of patent classes problematicas in the case of the GI

¹⁹ Extract from patent EP2245166A1: The composition of the invention comprises, in various embodiments, a dietary supplement, a food product, a pharmaceutical formulation, humanized animal milk, an infant formula, a cosmetic item and a biodiesel fuel, for example. A pharmaceutical formulation can include, but is not limited to, a drug for the treatment of neurodegenerative disease, a retinal disorder, age-related maculopathy, a fertility disorder, particularly regarding sperm or testes, or a skin disorder.

repository. On the contrary, a keyword-based approach is able to account for the expansion of the sector that could be observed outside the GI classes. In order to further investigate the relevance of this issue, we performed a search on TI, using a short keywords list including only general and broad terms such as biofuel or biodiesel etc., and distinguishing between patents belonging to GI classes or not. This allowed us to isolate the non-GI patents included in BioPat and to compare them with GI patents selected using the same criteria (same keywords searched for in same fields of the patent document and same publication date range). These patents are very likely to be directly linked to the biofuels sector thanks to the highly selective nature of the keywords used.

The procedure's outcome is particularly interesting: by excluding GI classes from the search, we found 5,056 patents while, when limiting the search to GI classes, we found 3,395 patents. This result implies that GI classes may be those where the probability of finding biofuel-related patents is lower, and that a relevant portion of biofuel-related patents is dispersed in a great array of different patent classes outside GI. This suggests that the cross-cutting nature induced by knowledge complexity in the biofuel sector is actually much relevant than what can be appreciated from an analysis based on GI classes only.

This result is confirmed when the dynamic analysis of new IPC classes (starting from 2000) is performed, showing a significant emergence in the database of new IPC classes both within and outside GI classes (Figure 9).

Figure 9 – New IPC classes in BioPat, out and in GI from year 2000

Source: our elaboration on BioPat and EPO

6. Implications for economic analysis and policy design

The evidence provided in the analysis of the characteristics and dynamics of knowledge in the biofuels sector as given by Figures 4-8 reveals how crucial is to develop methods able to capture such increasing complexity, where standard class-based tools fail to catch the emergence of growing interactions with complementary activities not directly focused on the sector under scrutiny. This result suggests the importance of cumulability and complementarity effects in the examined sector. The first refers to the relevance of previous technological knowledge in the generation of new knowledge, the second is related to its horizontal indivisibility, i.e. the possibility to use knowledge generated in specific domains for a broad array of technological applications (Antonelli, 2003 and 2013).

In this context, a single company can be hardly able to master all the needed technological and organizational competencies to develop new knowledge on its own. In contrast, in order to introduce innovations it is increasingly required to mobilise a wider array of technological capabilities which are distributed across a range of firms and other knowledge generating organizations within and outside the biofuels sector (Coombs and Metcalfe, 2002; Granstrand *et al.*, 1997). In fact, when we scattered the BioPat database by looking at patent ownership, we found a significant prevalence of firms mainly producing food and chemicals products, rather than biofuel products, which are exploring for new market opportunities.

In complex technological domains such as the biofuels sector, knowledge can be viewed as a collective process made possible by the development of continuous accumulation of highly differentiated but complementary competences and technological knowledge. Hence, to fuel the generation of new knowledge, the integration and recombination of external and internal stocks of knowledge and the systemic interactions of a variety of agents is essential (Antonelli, 2013; Antonelli *et al.*, 2013; Foray, 2004; Quatraro, 2012).

A number of business strategies can be settled to integrate different technological capabilities and knowledge sources including vertical integration, strategic alliances and collaborative agreements (Pisano, 1991). According to Hu and Philips (2011) two major constraints are limiting the expansion of the biofuel sector: availability of raw materials, and the development of fermentation and bio-refinery technologies. Technological advancements related to these issues, which are not necessarily purposely developed in the biofuel industry, will crucially interplay in creating favourable economic and technological

conditions for the full deployment of biofuels. Therefore they suggest that inter-industrial collaboration on these aspects will be central for the future dynamics of the biofuels sector.

According to our findings, biofuels industry is facing much more than two constrains and a mere process of industry vertical integration would probably fail in fostering the consolidation of a biofuels system. Firstly, when knowledge complexity is relevant, vertically integrated corporation are losing their primary position in the introduction of innovations as a single company has increasing difficulties in possessing and managing all the competencies needed for the process of the generation of new knowledge. Secondly, as emerged from our previous analysis and summarized in Figure 10, the tree of the biofuels system is composed by at least three different vertically integrated logs: production, distribution and utilization of biofuels products. Each log is in turn constituted by several roots that play a vital role in feeding the system. The higher is the number of roots for each log, the larger is the range of different bits of knowledge necessary to complement the core technological domain.

Figure 10 -Thecross-cutting nature of the biofuels system

Hence, while in the biofuels system it is hard to imagine a fully vertically and horizontally integrated company able to completely dominate all the technological and organizational competencies needed to overcome structural constraints in the evolution of the biofuels sector, the establishment of networks and collaborations between different organizations appears to be the only way to facilitate the coordination and integration of complementary technological competencies in a context characterized by a high degree of complexity, uncertainty and the dispersion of competencies among heterogeneous sources, as evidenced in other complex systems such as the car industry (Patrucco, 2011, 2014).

However, distinct firms operating in different industries may well not share the same goals, or direct their innovative efforts towards a common direction, so that economic systems and the spontaneous interaction of heterogeneous agents acting in different contexts typically are not able to generate and implement a perfectly efficient network architecture (Ozman, 2009). In such cases, a new form of knowledge governance may emerge, which the literature identified as innovation platforms. Innovation platforms consist of hierarchical networks, in which the interactions do not emerge and evolve spontaneously, but in which the key players guide the behaviour of the other actors, selecting the members of the platform and directing the evolution of the system as a whole (Consoli and Patrucco, 2008; Patrucco 2012). In the case of the biofuels system, given the pervasiveness of sectoral specific public policies (Costantini and Crespi, 2013; Suurs and Hekkert, 2009) the main candidate to assume the role of key player in the biofuels innovation platform is probably the public sector, which should design and implement a policy mix that favours the interactive behaviour of agents operating in the system with the deliberate strategy to facilitate the circulation and the actual exploitation of complementary knowledge inputs dispersed in a variety of different agents in private and public organizations.²⁰

In this context, the use of detailed information on technological knowledge broadly related to the biofuels sector as those contained in BioPat may help policy makers to identify the interactions between different sectors and to unveil the existence of horizontal

²⁰So far, the majority of developed countries have focused their policy interventions on specific roots as highlighted in Figure 10, without paying attention to the whole functioning and interconnections within the system. An interesting exception is represented by the case of the Swedish pro-biofuels policy framework, whose success can be traced into its comprehensive approach. In 2011 there were in fact 12 different pro-biofuels policies in Sweden, including incentives for biofuels pumps and import tools (Holmgren, 2012).

and vertical bottlenecks in order to activate specific inter-sectoral policies to favour the technological evolution and the economic development of this sector.

7. Conclusions

This paper has analysed issues associated with the measurement of innovation activities through patents in a complex technological domain. The proposed methodology aims to solve some of the drawbacks related to how patent data are allocated and organized in international databases. In order to create a database that includes patents that are strictly related to the investigated field, we have developed an original method based on keywords rather than on IPC codes. The biofuels sector has been taken as an experimental case study.

Starting with a systematic mapping of the biofuels production system, we have built a simplified but comprehensive description of this technological domain by developing an in depth sectoral analysis which represented the basis for the keywords selection obtained through an iterative approach applied to the collection of the recent scientific literature. The construction of the BioPat database was finalized by a series of interviews with experts in the biofuels sector which allowed us to validate the adopted methodology by comparing the results obtained through the keyword-based method with standard class-based codes, revealing improved accuracy when selecting data using our methodology.

Our analysis shows that the information contained in BioPat is more closely linked to innovation in the biofuels sector than that deriving from GI classification. The comparison between a IPC codes-based classification and the keywords method here developed allows underpinning the better performance of our method in providing insightful information concerning the structure, the composition, and more importantly the dynamics of a complex and cross-cutting technological domain. In particular, an important advantage of the adopted methodology is that the use of keywords classified according to the sectoral analysis for selecting patents allows us to assign patents to specific categories which provide useful insights into a description of the characteristics and the evolution of technologies.

The descriptive evidence from patents in BioPat suggests that biofuels are expanding their technological domain and many new applications cover inventions that aim to solve the infrastructural barriers (biodiesel additives, new tanks, new engines, new shipping facilities, etc.). Furthermore, our preliminary descriptive findings show that the distinction

between different technology generations can provide interesting insights for the understanding of the evolution of technologies in the biofuels sector.

The adoption of this keyword-based methodology also allows for depicting the increase in the transversally nature of the biofuels sector, the inter-linkages with several other industries and the enrichment of the knowledge stock from innovations not directly related to the production of biofuels. Building on this evidence, we derived some implications related to the complexity of knowledge associated with the biofuels domain, suggesting the importance of cumulability and complementarity effects in the examined sector and underscoring the potential importance of public sector in the governance of knowledge networks aiming at facilitating the evolution of technologies and the economic development of the biofuels industry.

References

- Antonelli C. (2003). Knowledge complementarity and fungibility: implications for regional strategy. *Regional Studies*, 37(6-7), pp. 595-606.
- Antonelli C. (2008). *Localized technological change, towards the economy of complexity*. London, Routledge.
- Antonelli C. (2013). *Handbook on the economic complexity of technological change*. Cheltenham, Edward Elgar.
- Antonelli C., Crespi F., Scellato G. (2013), Internal And External Factors In Innovation Persistence, *Economics of Innovation and New Technology*, 22(3), 256-280.
- Antonelli, C., Krafft, J., Quatraro, F. (2010). Recombinant knowledge and growth: the case of ICTs. *Structural Change and Economic Dynamics*, 21(1), pp. 50-69.
- Archibugi, D., Pianta, M. (1996). Measuring technological through patents and innovation surveys. *Technovation*, 16(9), pp. 451-468.
- Arundel, A., Kemp, R. (2011). Measuring eco-innovation. *UNU-MERIT Working Paper Series No. 2009-017*, Maastricht, The Netherlands.
- Berkhout, F. (2011). Eco-innovation: reflections on an evolving research agenda. *International Journal of Technology, Policy and Management*, 11, pp. 191-197.
- Borghesi, S., Costantini, V., Crespi, F., Mazzanti, M., (2013), Environmental innovation and socio-economic dynamics in institutional and policy contexts. *Journal of Evolutionary Economics*, 23(2), pp. 241-245.
- British Petroleum (BP). (2012). *Statistical Review of World Energy 2011*, British Petroleum.
- Brouwer, E., Kleinknecht, A. (1999). Innovative output, and a firm's propensity to patent: an exploration of CIS micro data. *Research Policy*, 28(6), pp. 615-624.
- Carriquiry, M. A., Xiaodong, D., Timilsina, G. R. (2011). Second generation biofuels: economics and policies. *Energy Policy*, 39(7), pp. 4222-4234.
- Cherubini, F., Jungmeier, G., Wellisch, M., Willke, T., Skiadas, I., Van Ree, R., de Jong, E. (2009). Toward a common classification approach for biorefinery systems. *Biofuels, Bioproducts & Biorefining*. Wiley InterScience.
- Consoli, D., Patrucco, P.P. (2008). Innovation Platforms and the Governance of Knowledge: Evidence from Italy and the UK. *Economics of Innovation and New Technology*, 17(7-8), pp. 699-716.
- Coombs, R., Metcalfe, J.S. (2002). Organizing for Innovation: Co-Ordinating Distributed Innovation Capabilities. In Foss, N. and Mahnke, V., (Eds.) *Competence, Governance, and Entrepreneurship: Advances in Economic Strategy Research*. Oxford: Oxford University Press.
- Corrocher N., Malerba F., Montobbio F. (2007). Schumpeterian patterns of innovative activity in the ICT field. *Research Policy* 36, pp. 418-432.
- Costantini, V., Crespi, F. (2013). Public policies for a sustainable energy sector: regulation, diversity and fostering of innovation. *Journal of Evolutionary Economics*, 23(2), pp. 401-429.
- Costantini, V., Crespi, F., Curci, Y. (2013). BioPat: an investigation tool for analysis of industry evolution, technological paths and policy impact in the biofuels sector. In Costantini V., Mazzanti, M., (Eds.), *The Dynamics of Environmental and Economic Systems. Innovation, Environmental Policy and Competitiveness*, Springer, Dordrecht.
- Coulter, N., Monarch, I., Konda, S. (1998). Software engineering as seen through its research literature: a study in co-word analysis. *Journal of the American Society for Information Science*, 49(13), pp. 1206-1223.
- Crawshaw, E. H., Lilley, L. C., Van Sluis, B. (2006). Patent No. EP1674553. Netherlands.
- De Rassenfosse, G., Dernis, H., Guellec, D., Picci, L., Van Pottelsberghe de la Potterie, B. (2012). The worldwide count of priority patents: a new indicator of inventive activity. *Research Policy*, 42(3), pp. 720-737.
- EIA-DOE. (2013). *International Energy Statistics*. U.S. Energy Information Administration, Washington DC, USA.
- Foray, D. (2004). *The Economics of Knowledge*, Cambridge MA, MIT Press.
- Gibbons, M., Limoges, C., Nowotny, H., Schwartzman, S., Scott, P., Trow, M. (1994). *The new production of knowledge: the dynamics of science and research in contemporary societies*. London: Sage.
- Grandstand, O., Patel, P., Pavitt, K. (1997). Multi-technology corporations: why they have “distributed” rather than “distinctive” core competencies. *California Management Review*, 39, pp. 8-25.
- Griliches, Z. (1990). Patent statistics as economic indicators: a survey. *Journal of economic literature* 28(4), 1661-1707.
- Hekkert, M. P., Suurs, R. A. A., Negro S.O., Kuhlmann S., Smits R.E.H.M. (2007). Functions of innovation systems: a new approach for analyzing technological change. *Technological Forecasting & Social Change* 74, pp. 413-432.
- Holmgren, C. (2012). Policy promoting biofuels in Sweden. An f3 synthesis report. F3 2012:3. The Swedish Knowledge Centre for Renewable Transportation Fuels.

- Horbach, J. (2008). Determinants of environmental innovation-New evidence from German panel data sources. *Research Policy*, 37, pp. 163-173.
- Hu, M.C., Phillips, F., Technological evolution and interdependence in China's emerging biofuel industry. *Technological Forecasting & Social Change*, 78(7), pp. 1130-1146.
- IEA (2008). *From 1st to 2nd generation biofuels technologies. An overview of current industry and R&D activities*. Paris: International Energy Agency.
- IEA. (2011). *Technology roadmap. Biofuels for transport*. Paris: OECD/IEA.
- Jaffe A., Trajtenberg M. (2002). *Patents, citations, and innovations: a window on the knowledge economy*. Cambridge, Mass: MIT Press.
- Janda, K., Kristoufek, L., Zilberman, D. (2012). Biofuels: policies and impacts. *Agric. Econ. Czech*, 58(8), pp. 372-386.
- Johnston, D. (2002). *The OECD Technology Concordance (OTC): Patents by Industry of Manufacture and Sector of Use*. OECD Science, Technology and Industry Working Papers. March.
- Johnstone, N., Hascic, I., Popp, D. (2010). Renewable energy policies and technological innovation: evidence based on patent counts. *Environmental & Resource Economics* 45(1), pp. 133-155.
- Karmarkar-Deshmukh, R., Pray, C. E. (2009). Private sector innovation in biofuels in the United States: induced by prices or policies?. *AgBioForum*, 12(1), pp. 141-148.
- Kemp, R., Oltra, V., (2011). Research insights and challenges on eco-innovation dynamics, *Industry & Innovation*, 18, pp. 249-253.
- Kondamudi, N., Mohapatra, S.K., Misra, M. (2008). Spent coffee grounds as a versatile source of green energy. *Journal of Agricultural Food Chemistry* 56(24), pp. 11757-60.
- Kortum, S., Putnam, J. (1997). Assigning patents to industries: test of the Yale Technology Concordance. *Economic Systems Research*, 9(2), 161-176.
- Krafft, J., Quatraro, F., Saviotti, P.P. (2011). The knowledge-base evolution in biotechnology: a social network analysis. *Economics of Innovation and New Technology*, 20(5), 445-475.
- Loasby B.J. (1999). *Knowledge Institutions and Evolution in Economics*. Routledge, London.
- Looze, M.A., Lemarié, J. (1997). Corpus relevance through co-word analysis: An application to plant proteins. *Scientometrics*, 39(3), pp. 267-280.
- Lybbert, T.J., Zolas, N.J. (2012). *Getting patents and economic data to speak to each other: An "algorithmic links with probabilities" approach for joint analyses of patenting and economic activity*. WIPO Economics & Statistics Series. Working Paper n. 5, October.
- Markard, J., Raven, R., Truffer, B. (2012). Sustainability transitions: An emerging field of research and its prospects. *Research Policy*, 41, pp. 955-967.
- Narin, F. (2000). Tech-Line. Background Paper. In J. Tidd, *Measuring Strategic Competence*. London: Imperial College.
- Noteboom, B. (2000). *Learning and Innovation in Organizations and Economics*. Oxford University Press, Oxford.
- OECD (2011). *Fostering Innovation for Green Growth*, OECD Green Growth Studies, OECD Publishing, Paris.
- Ozman, M. (2009) Inter-firm networks and innovation: A survey of literature, *Economics of Innovation and New Technology*, 18(1), pp. 39-67.
- Patrucco, P.P. (2011). Changing network structure in the organization of knowledge: the innovation platform in the evidence of the automobile system in Turin. *Economics of Innovation and New Technology*, 20(5), pp. 477-493.
- Patrucco, P.P. (2014). The Evolution of Knowledge Organization and the Emergence of a Platform for Innovation in the Car Industry, *Industry and Innovation*, DOI:10.1080/13662716.2014.910892.
- Picci, L. (2010). The internationalization of inventive activity: a gravity model using patent data. *Research Policy*, 39, pp. 1070-1081.
- Pisano, G. (1991). The Governance of Innovation: Vertical Integration and Collaborative Arrangements in the Biotechnology Industry. *Research Policy*, 20, pp. 237-249.
- Quatraro, F. (2012). *The Economics of Structural Change in Knowledge*, London: Routledge.
- Schmookler, J. (1966). *Invention and Economic Growth*. Harvard University Press, Cambridge.
- Schmoch, U., LaVille, F., Patel, P., Frietsch, R. (2003). Linking technology areas to industrial sectors: final reports to the European Commission. *DG Research*. November.
- Sirilli, G. (1999). Innovation Indicators in Science and Technology Evaluation. *Scientometrics*, 45(3), pp. 439-443.
- Strumsky, D., Lobo, J., van der Leeuw, S. (2012). Using patent technology codes to study technological change. *Economics of Innovation and New Technology*, 21(3), pp. 267-286.
- Suurs, R.A.A., Hekkert, M.P., 2009. Cumulative causation in the formation of a technological innovation system: the case of biofuels in the Netherlands. *Technological Forecasting and Social Change* 76, 1003-1020.

- Trippe, A. J. (2003) Patinformatics: Tasks to tools. *World Patent Information*, 25(3), pp. 211–221.
- Verspagen, B., van Moergastel, T., Slabbers, M. (1994). MERIT concordance tables: IPC-ISIC(Rev 2). *MERIT Research Memorandum*. February.
- WIPO. (2011). IPC guide. Retrieved from http://www.wipo.int/export/sites/www/classifications/ipc/en/guide/guide_ipc.pdf.

Appendix

Table A1 – IPC classes in Green Inventory defining the biofuels domain

IPC sub-group and sub-class	Number of patents	Object (hierarchical definition) defined by Green Inventory
A01H	20,189	Biofuels –Liquid fuels -From genetically engineered organisms
A62D 3/02	431	Harnessing energy from manmade waste - Anaerobic digestion of industrial waste
B01D 53/02	3120	Harnessing energy from manmade waste - Landfill gas-Separation of components
B01D 53/04	4423	Harnessing energy from manmade waste - Landfill gas-Separation of components
B01D 53/047	1491	Harnessing energy from manmade waste - Landfill gas-Separation of components
B01D 53/14	2948	Harnessing energy from manmade waste - Landfill gas-Separation of components
B01D 53/22	3498	Harnessing energy from manmade waste - Landfill gas-Separation of components
B01D 53/24	109	Harnessing energy from manmade waste - Landfill gas-Separation of components
B09B	6613	Harnessing energy from manmade waste - Landfill gas
C02F 11/04	576	Harnessing energy from manmade waste - Industrial waste - Anaerobic digestion of industrial waste
C02F 11/14	669	Harnessing energy from manmade waste - Industrial waste - Anaerobic digestion of industrial waste
C02F 3/28	1365	Biofuels – Biogas
C07C 67/00	9671	Biofuels – Liquid fuels - Biodiesel
C07C 69/00	15443	Biofuels – Liquid fuels - Biodiesel
C10B 53/00	1089	Pyrolysis or gasification of biomass
C10B 53/02	Included in C10B 53/00	Biofuels - Solid fuels - Torrefaction of biomass
C10G	17625	Biofuels – Liquid fuels - Biodiesel
C10J	2795	Pyrolysis or gasification of biomass
C10L 9/00	412	Biofuels - Solid fuels - Torrefaction of biomass
C10L 1/00	2713	Biofuels – Liquid fuels
C10L 1/02	Included in C10L 1/00	Biofuels – Liquid fuels – Vegetable oils /Biodiesel / Bioethanol
C10L 1/14	1958	Biofuels – Liquid fuels
C10L 1/182	503	Biofuels – Liquid fuels - Bioethanol
C10L 1/19	672	Biofuels – Liquid fuels – Vegetable oils /Biodiesel
C10L 3/00	1757	Integrated gasification combined cycle (IGCC)/ Biofuels - Biogas
C10L 5/00	759	Biofuels – Solid fuels / Harnessing energy from manmade waste – agricultural waste
C10L 5/40	Included in C10L 5/00	Biofuels - Solid fuels - Torrefaction of biomass
C10L 5/42	Included in C10L 5/00	Harnessing energy from manmade waste – agricultural waste - Fuel from animal waste and crop residues
C10L 5/44	Included in C10L 5/00	Harnessing energy from manmade waste – agricultural waste - Fuel from animal waste and crop residues
C10L 5/46	Included in C10L 5/00	Harnessing energy from manmade waste - Landfill gas – Municipal waste

C10L 5/48	Included in C10L 5/00	Harnessing energy from manmade waste - Industrial waste / Biofuels - Solid fuels
C11C 3/10	925	Biofuels – Liquid fuels - Biodiesel
C12M 1/107	489	Biofuels – Biogas
C12N 1/13	243	Biofuels – From genetically engineered organisms
C12N 1/15	11575	Biofuels – From genetically engineered organisms
C12N 1/21	27080	Biofuels – From genetically engineered organisms
C12N 15/00	16555	Biofuels – From genetically engineered organisms
C12N 5/10	30000	Biofuels – From genetically engineered organisms
C12N 9/24	2754	Biofuels – Liquid fuels - Bioethanol
C12P 5/02	414	Biofuels – Biogas
C12P 7/06	1159	Biofuels – Liquid fuels - Bioethanol
C12P 7/14	104	Biofuels – Liquid fuels - Bioethanol
C12P 7/64	1931	Biofuels – Liquid fuels - Biodiesel
D21C 11/00	983	Harnessing energy from manmade waste - Industrial waste - Pulp liquors