Exploring technology in the biofuels sector through patent data: the BioPat database[1]

Valeria Costantini

Department of Economics, Roma Tre University

Francesco Crespi

Department of Economics, Roma Tre University

BRICK, Collegio Carlo Alberto

Ylenia Curci

Department of Economics, Roma Tre University

Abstract

This paper describes the methodology, characteristics and potential use of BioPat, a dataset containing patents in the field of biofuels. The innovative methodology we use aims to solve drawbacks related to how patent data are allocated and organized in international databases. In order to create a database which includes patents strictly related to the investigated field, we propose an original method based on keywords, rather than on International Patent Classification (IPC) codes. Starting with a systematic mapping of biofuel production processes, we built a simplified but comprehensive description of the technological domain related to the production of biofuels by applying so-called process analysis. The keyword selection relies on an iterative approach, based on an analysis of recent scientific literature. The database was finalized with a series of interviews with experts in the biofuels sector, and compared with IPC-based biofuel codes, revealing improved accuracy when selecting data using our methodology.

J.E.L.: O310; O340; Q420; Q550

Keywords: Biofuels, Patents, Keywords selection method, Innovation patterns

[1] The authors would like to thank Alessandra Molina and Alessandro Palma for their initial contribution during the collection of keywords, Federico Caviggioli for his advice on downloading patents, Vincenzo Alfano, Alessandro Blasi and Carlo Pignatelli for patent validation and Carlotta Greci and Corrado Pollastri for assistance in data management. The support of the Regione Lazio, University Roma Tre, INEA, ENEA and European Union D.G. Research with the Grant number 266959 to the research project "PICK-ME - Policy Incentives for the Creation of Knowledge: Methods and Evidence" is also gratefully acknowledged. The usual disclaimers apply.

1. Introduction

The last decade has been a period of intense instability in oil prices and there has been growing concern about the environmental costs of carbon emissions from fossil fuels in the transport sector. As described in the “Energy, Transport and Environment Indicators” published by Eurostat (2007), in 2005, the transport sector accounted for about 31% of total energy consumption in the European Union (EU-27 Members), representing 19% of total Greenhouse Gases (GHG) emissions. Due to high oil prices and the need to reduce GHG emissions, biofuels for transport use such as ethanol and biodiesel, which are the only suitable substitutes for fossil fuels, have gained importance in many countries.

In 2005, the US Energy Bill established a mandate requiring minimum levels of biofuel consumption from 11.9 million tons in 2006 up to 22.1 million tons in 2012. The European Union (EU) is fostering the use of biofuels, and bioenergy in general, in several forms. There are various documents in place settled by the European Commission (EC) to promote the use of bioenergy such as Directives 2001/77/EC, 2003/30/EC, 2003/96/EC, the EU “Biomass Action Plan” (EC, 2005) and the “European Union Biofuel Strategy” (EC, 2006). According to the EU biofuels directive 2003/30/EC, EU Member States should ensure a minimum amount of biofuels and other renewable fuels in their total consumption of transport fuel. In the “Renewable Energy Roadmap” (EC, 2007), the EC proposed binding minimum targets of 10% for biofuels in each Member State. On 23 January 2008, the EC put forward an integrated proposal for Climate Action, including a directive that sets an overall compulsory target for the European Union of 20% renewable energy by 2020 and a 10% minimum target for the market share of biofuels by 2020, to be observed by all Member States.

Despite the fact that the US mandate had almost been reached by 2007, and despite the very recent change in petroleum consumption among OECD countries which is showing a slow decrease, the past ten years demonstrate that current European policies for a sustainable energy system are inadequate in the transport sector and highly dependent on fossil fuels, thus requiring further efforts to expand alternative energy sources.

The global production of biofuels amounted to 59,261 ktoe in 2010, which represents around 1-2% of total fuel consumption in transportation. The projections of future market shares shape a huge increase reaching around 13% of global fuel consumption in 2050 (IEA, 2007). The size of such an increase will depend critically on the rate of technological change and the diffusion rate of new technologies in the biofuels sector. It is worth mentioning that the OECD-FAO (2010) projection for 2010-2019 on bioethanol and biodiesel production pointed out that the 13% growth rate is probably underestimated. In 2009, alternative energy sources to fossil fuels account for more than 50 % of installed capacity in US and above 60% in the EU (UNEP, 2010), remaining almost resilient against economic turbulence.

Among renewable energy sources, investments in biofuel plants declined in 2009, whereas waste-to-energy investment increased from 9 to 11 billion dollars. In 2008, the biofuels sector had a total investment of 18 billion dollar whereas in 2009 it ended up with just 7 billion dollars. The UNEP Energy Finance Initiative Report suggests that investment in first generation biofuels is declining due to the fact that most firms are not operating at full capacity: “investment in new biofuel plants declined from 2008 rates, as corn ethanol production capacity was not fully utilized in the United States and several firms went bankrupt. The Brazilian sugar ethanol industry also faced economic troubles, with no growth despite ongoing expansion plans. Europe faced similar softening in biodiesel, with production capacity only half utilized” (UNEP, 2010, p. 6).

The recent evolution in the biofuels sector has been characterized by strong price volatility and a mismatch between demand and supply. Part of the responsibility for the current situation can be attributed to the confusion created by governmental policies that conflict with one another and a lack of knowledge of the biofuels production system (Costantini and Crespi, 2012). However, the increased price of fossil fuels as well as a need for environmental-friendly and cost-effective technologies for the production of clean energy, made us support the idea that these changes must be reflected in evolution of the sector’s technological regime.

The measurement of innovative activities is a rather challenging task and a great number of different science and technology indicators have been identified in the literature (Sirilli, 1997). The main input indicator relies on research and development (R&D) expenditure, while the most used innovation output indicators are based on patent data. Both types of indicators have strong limitations since not all research efforts translate into the introduction of innovations and not all innovations are patented. For our purposes, specific and systematic information on private R&D expenditures in the biofuels sector are not available while access to patent data makes it possible to collect information on the evolution of the innovative performance of economic systems by looking at the volume of patents registered and granted (Johnstone et al., 2010).

As already mentioned, the use of patents has its pros and cons. The advantages of using patents as a proxy of innovation are manifold. 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 new technological knowledge, the inventors and the owners of the patent and the usefulness of patented knowledge for subsequent innovations. On the other hand, using patents as a proxy for innovation presents several relevant issues (Griliches, 1990). In particular, only a limited part of produced innovations are patented (Archibugi and Pianta, 1996) and there is an intrinsic variability of patents' value (Jaffe and Trajtenberg, 2002).

For our purposes, another important problem has to be taken into account. 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. 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 named IPC (International Patent Classification) is associated with the internal code which allows the classification of patents by following a hierarchical criterion (from 8 main fields to almost 70.000 subgroups) based on chemical and technological principles, only occasionally related to manufacturing sectors. In particular, the resulting classification is only of limited usefulness when it identifies a specific sector which does not fit the criteria used in the classification, as in the biofuels sector.

The aim of this work is therefore to illustrate a possible methodology for building a sector-specific patent database and showing how it can be potentially used for economic analysis. Despite the well-known limitations related to the use of patent data in innovation studies, in order to draw a picture of sectoral technological patterns, a valuable option is to build a database that tries to identify precisely the entire universe of patents strictly related to the biofuels sector. To do this, we must first adopt an early approach suggested by Hekkert *et al*. (2007) in order to map the actors which participate in the Biofuels Innovation System systematically by means of a process analysis. In the following, we first describe the IPC system and the Green Inventory database. We then provide details of the adopted keyword methodology and after that we give first descriptive results drawn from the collected database. The conclusions provide a synthetic discussion of the reached objectives and future research developments.

2. The IPC system and the Green Inventory database

During the last century, the increasing amount of patents registered daily worldwide and the great number of interactions among patents offices made the adoption of a uniform system of patent classification necessary.

The first attempt to create a global market for patents came with the founding of the World Intellectual Property Organization (WIPO), as a United Nations agency. WIPO was established by the WIPO Convention in 1967 with a mandate from its Member States to promote the protection of intellectual property (IP) throughout the world through cooperation among states in collaboration with other international organizations.

The will to foster closer international cooperation in the industrial property field, and to contribute to the harmonization of national legislation in that field led in 1971, after 15 years of international cooperation, to the Strasbourg Agreement concerning International Patent Classification (which entered into force on October 7th 1975). The huge number of patents (and related documents) created two main problems the treaty had to deal with: the administrative processing of the patent applications and the maintenance of the search files containing the published patent documents.

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" (IPC Guide, 2011, p. 1).

The International Classification divided the universe of patents into 8 sections, 20 subsections, 118 classes, 624 subclasses and over 67,000 groups (of which approximately 10% are main groups and the remainder are subgroups). Each of the sections, classes, subclasses, groups and subgroups has a title and a symbol, and each of the subsections has a title. Each classification term consists of a sequence of symbols: the first one is a capital letter which represents the section. The letter is followed by a two digit number which represents the class and then by another capital letter that stands for the subclass. The subclass is then followed by a 1 to 3 digit "group" number, an oblique stroke and a number of at least two digits representing a "main group" or "subgroup". Hence, the IPC is a hierarchical system, with layers of increasing detail. The following represents an example of the classification: A01B1/00 symbolizes Human Necessities (Section A); Agriculture (Subsection title); Agriculture, Forestry, Animal Husbandry, Hunting, Trapping, Fishing (Class A01); Soil working in agriculture or forestry, Parts, details or accessories of agricultural machines or implements in general (Subclass A01B); Hand tools (Group A01B1); subgroup not specified (A01B1/00).

These different sections allow distinctions to be made between patents belonging to categories which sporadically present an economic importance (such as the case presented above, hand tools used in agriculture). On the contrary, the IP classification is not suitable when the focus of the research does not match an existing section (e.g., harvest tools). Several attempts have been made to provide a crosscutting interpretation of the standard classification.

The first category of attempts is a top down approach that relies on the IPC class and aims to define its content:

- a rough and unpredictable method consists in the exploitation of the linkages between classes assigned to the same patent by considering those appearing together as a "class family";
- a more advanced technique tries to identify the classes which are appropriate for containing a patent related to the investigated object.

The "IPC Green Inventory" database (GI) falls into the latter category and was developed by the IPC Committee of Experts in order to facilitate searches for patent information relating to Environmentally Sound Technologies (ESTs), as listed by the United Nations Framework Convention on Climate Change (UNFCCC).

ESTs are currently scattered widely across IPC in numerous technical fields. The GI allows all ESTs to be collected in one place. Following the IPC system, the ESTs are presented in a hierarchical structure. According to the WIPO web site, two steps were required 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 identify the related IPC places. In order to identify the IPC places correctly, the experts can use 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 places under the specific category. Hence, the inventory consists of a list of IPC classes characterized by the fact that they are appropriate for containing patents related to a green technology.

Among the ESTs, for our purpose, we considered 44 IPCs (40 subgroups and 4 subclasses) that identify the biofuels sector.

In Table 1 we list the IPC subgroups and subclasses, the number of patents included in them (accordingly to Thomson Routers as of February 2011) and the technology associated with the different IPC codes.

As already mentioned, the classes above are suitable for containing patents related to the object specified in the GI (last column). It is worth remembering that these objects, which refer to the related IPC class, are not the IPC class object. For example, the first class (first row) A01H, which, according to GI, is appropriate for containing patents related to liquid biofuels obtained by genetically engineered organisms, can actually contain, according to the IPC, all the patents that fall into the category (subclass title) "new plants or processes for obtaining them, plant reproduction by tissue culture techniques".

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 (as a sort of validation).

Table 1 - Green Inventory classes related to biofuels

IPC Subgroup and subclass	Number of patents	Object (hierarchical definition) defined by the 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	In the previous	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	In the previous	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	In the previous	Biofuels - Solid fuels - Torrefaction of biomass
C10L 5/42	In the previous	Harnessing energy from manmade waste – agricultural waste - Fuel from animal waste and crop residues
C10L 5/44	In the previous	Harnessing energy from manmade waste – agricultural waste - Fuel from animal waste and crop residues
C10L 5/46	In the previous	Harnessing energy from manmade waste - Landfill gas – Municipal waste
C10L 5/48	In the previous	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

Hence, in order to shed light on the accuracy of the GI databases, we validated a sample of patents included in the IPC classes indicated above by asking a team of experts from the Italian National Agency for New Technologies, Energy and Sustainable Economic Development (ENEA) to check their coherence. Additionally, we asked the group of experts to distinguish between patents with a direct application in the biofuel production process and an indirect one. We downloaded the description field of the whole universe of patents belonging to these classes for USPTO, WIPO and EPO and eliminated the duplicates (each patent can fit in more than one class) ending up with 107,161 elements from which we selected a 1% sample.

The results of the expert validation showed that on average, only 25% of the patents included in the sample have a direct application in the biofuels sector. This percentage significantly varies among the patent offices. Such a result confirmed our intuition regarding the limits associated with the identification of patents through the IPC system in the biofuels sector.

3. The BioPat methodology

Setting a proper methodology to select patents in a rather specific sector is not an easy task. As shown by the experts' validation on the GI, the IPC class selection fails to extrapolate the classes that are supposed to identify a single economic sector, maintaining a high risk of considering external elements. Moreover, considering the huge variety of raw material and processes available for biofuel production that often overlap with other manufacturing sectors, it is highly probable that the GI classification does not catch all the patents that have a direct or an indirect application in the investigated field. Moreover, the method usually adopted by several international organizations, which considers all patents directly or indirectly linked with each other in a single family, is not appropriate when it comes to working on a small sector (or on a limited number of patents) because the smaller the sector, the higher the likelihood of catching external elements.

In order to tackle the lack of specificity from an economic point of view, 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 in selections of words from already existing keyword lists or the extraction of keywords from titles and, at least, abstracts of patents and scientific publications.

The literature followed three main approaches:

- co-word study based on the keywords proposed by experts (Looze and Lemarie, 1997);
- use of descriptors chosen by professional indexers employed in patent offices and search engines (Coulter et al., 1998);
- 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 arbitrarity of the selection process. In fact, Corrocher et al. (2007) pointed out that the ex-ante selection of the keyword procedure might reflect preconceptions, different backgrounds and points of view of the words' selectors and differences in the trainings and backgrounds of professional indexers. As a result, the authors decided to identify the most frequent sequential triples of words without imposing any priority constraint on the selection of keywords. The authors argue that triples of words within patent abstracts can identify technological domains that can be compared with the existing IPC technological classes.

Unfortunately, the method which 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 (which is not the case for biofuels). Moreover, it is also more appropriate when the novelty of patents is based on engineering contents, which are more likely to fit into *ad hoc* classes.

On the contrary, the patents related to biofuels are spread across several IPC classes because the technology that characterizes the sector basically consists of thermo/bio-chemical processes and very common raw materials that can find applications in several fields.

Since we realized that the subjectivity of the selection process could represent a big challenge for the research outcome, we tried to make the process as objective as possible. We then decided to consult technical experts in the field of biofuels. We interviewed exponents of ENEA who helped us describe the process of biofuel production. This team of technical experts completed and validated the list of keywords derived from the scrutiny of a large number of scientific publications and the keyword list extracted by Scopus, a powerful search tool which provides access to a large number of scientific publications and patents office databases.

The choice and classification of keywords derives from recent scientific literature which gives us the empirical basis of the process analysis. The search for keywords was divided into 2 different steps: the first one was dedicated to a search for "raw material" keywords, where a relevant number of technical and scientific papers was analysed in order to pick out the terms describing the biomass used (or potentially used) to produce biofuels. The second step consisted in an accurate description of the "transformation process" currently known in biofuel production, including pre-treatment processes, chemical agents involved in the process and technical instrumentation used in it. Keywords were then tested on Scopus (www.scopus.com). At the same time, Scopus allows you to check if patents exist containing the selected keywords. Hence, the final selection of the keywords comes from an iterative procedure which allows results from scientific articles to be compared with patent results. This first step led to selecting several keywords which showed positive results both in patents and articles via Scopus. These keywords were submitted to the ENEA experts (see Appendix).

Finally, we improved the traditional keyword methods that look for keyword matches only in the patent's titles and abstracts. 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 is true for the biofuels sector and represents the explanation of the crosscutting shape that it assumes in the IPC classification. 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 hypothetical, and not necessarily direct, function in the biofuel production process.

The patents were downloaded using Thomson Innovation, a single, integrated solution that combines intellectual property, scientific literature, business data and news with analytic, collaboration and alerting tools in a robust platform. With Thomson Innovation, we were able to export up to 30,000 records into csv formats in one single operation. Thomson Innovation has the world's most comprehensive collection of patent data from major patent authorities, specific nations and proprietary sources exclusive to Thomson Reuters.

All process-specific and raw material keywords were used in the Thomson innovation jointly with a more general keyword (such as bio-diesel, bio-ethanol, bio-gas, bio-fuels) in order to exclude patents that share the same raw materials or transformation processes (in particular pharmaceutics and cosmetics are strongly related to the biofuels sector). Afterwards, some testing searches were implemented with a few selected keywords in order to verify the response of the Thomson database to the inputs. The Thomson 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 sugarcane" and so on.

Furthermore, we carried out a special search using general keywords in the "applicant" field, hypothesizing that a firm called "The Biofuel Company" deals with patent inventions related to biofuels.

Using Thomson Innovation, patents can be downloaded from national and international patent data offices. We focused our research on the European Patent Office (EPO), World Intellectual Property Organization (WIPO) and United States Patent and Trademark Office (USPTO) as described in Table 2.

Table 2 - Data available on Thomson Innovation

WIPO APPLICATIONS	
Published international patent applications, fully searchable, language: 70% English, 15% German, 5% French, 1% Spanish	1978 – present
UNITED STATES	
US Granted, fully searchable, language: English	1836 – present
US Applications, fully searchable, language: English	2001 – present

EUROPE	
European Granted, potentially 31 countries, fully searchable, language: 60% English, 30% German, 10% French	1980 – present
European Applications, potentially 31 countries, fully searchable, language: 60% English, 30% German, 10% French	

With regard to raw material keywords, the search on Thomson was carried out as follows: by using Boolean operators “OR” and “AND” we selected all the 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.[2]

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 transesterification, Fischer-Tropsch, anaerobic digestion and so on.[3]

An important advantage of the adopted methodology is that by selecting patents related to previously classified keywords, specific categories can be assigned to patents derived from each keyword.

According to the IEA classification method (IEA, 2008), in order to improve building and management of the dataset, production stages, “generations” and final product were used in order to classify patents (raw materials and transformation process; old and new generation, fat, alcohol and gas).

IEA classifies biofuels as follows: first generation biofuels, which are mainly produced from agricultural crops and traditional oleaginous plants (such as palm and colza), are characterized by mature commercial markets and well known technologies. On the contrary, second generation biofuels are represented by non-food crops, especially from forestry residues (that we classify as Ligno and Waste) or dedicated energy crops (LIGNO). Third generation biofuels are mainly related to Algae and genetically modified plants.

Unfortunately, IEA classification is not always suitable for the entire production process and any final biofuel products (bio-ethanol, above all) because most of the definitions are overlapping. Main shortcomings of the IEA classification were reduced by repeated interviews with a panel of experts in agro-bio-technologies. Their responses helped us define a logical structure model that focused more on our search attempts.

The other classification method adopted is based on the following assumption: the actual technology used to produce biofuels, which includes raw materials, techniques knowledge, tools and machineries, is considered the current technological knowledge stock. Within this knowledge stock, two main technological categories can be discerned: “old generation” and “new generation”, both for raw material and process keywords, which are related and include the entire supply of technologies for biofuel production. Making use of the exclusion principle, it is easy to define everything that is not in the old category as belonging to the new category.

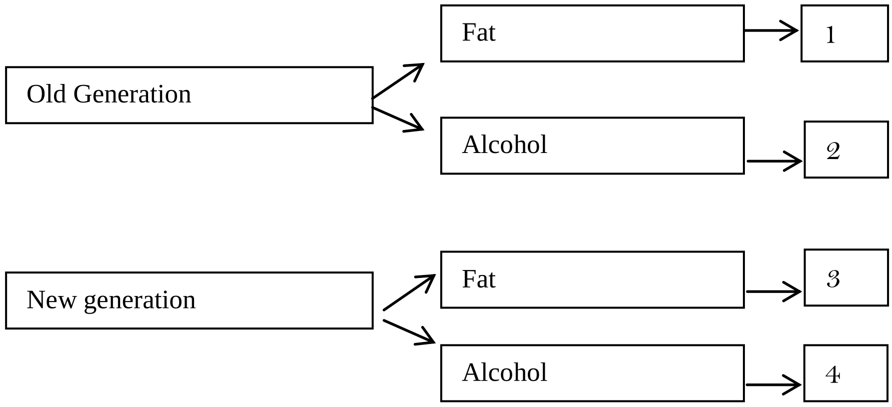
The raw material keywords can be divided into several categories which help to identify the patent’s content: chemical agents, agricultural waste/crop, agricultural waste/ligno, algae, crops, GMO, ligno, livestock, oleaginous, sugar, urban waste and non-urban waste. Some keywords can overlap with more than one category. Obviously, different combinations are possible and numerous categories can be created. As an example, in Figures 11.1 and 11.2, we provide more than one possible combination of keywords and categories.

---

[2] For example: Nannochloropsis (an alga) AND “renewable *ethanol” OR “green *diesel” OR *methanol OR *buthanol OR biomethane OR biomethiletere OR “Synthet* fuel*” OR biodiesel OR “renewable fuel*” OR biofuel* OR etc.

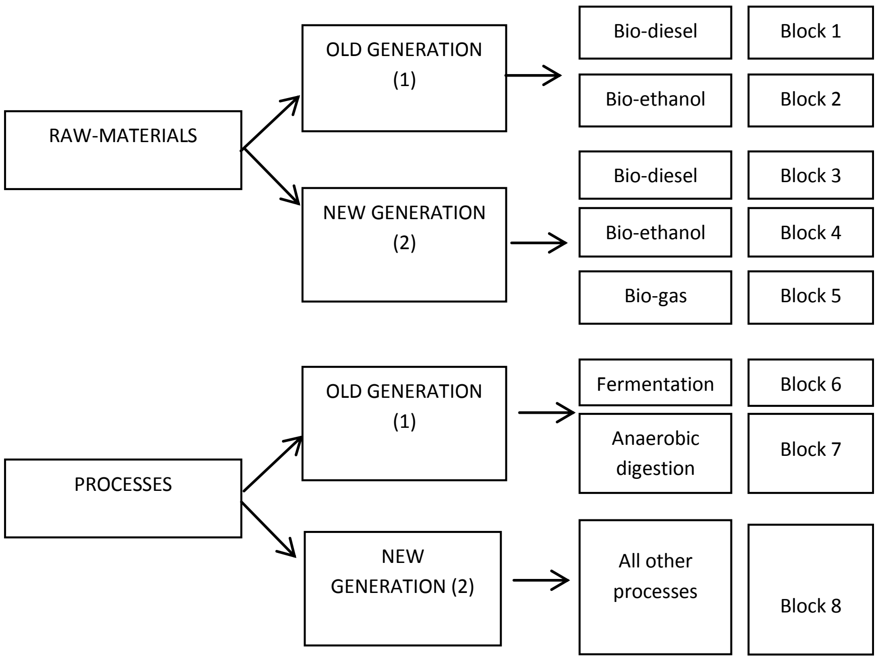
[3] After that, we verified if the downloads could represent a significant part of the whole universe achieved using only the general keywords. The huge specific outcome obtained by using the general keywords strongly reinforces the choice of working with selected specific keywords rather than working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

Figure 1 - Exemplificative alternative structures of database and classifications using keywords (case a)



TYPE	KEYWORD	BLOC	GEN.	FAT	ALCOOL
algae	Chlorella vulgaris	3-4	2	1	1
food	Corn	2	1	0	1
algae	Dunaliella tertiolecta	3-4	2	1	1
Food	Maize	2	1	0	1
sugar	Sorghum	2	1	0	1
ligno	Miscanthus	4	2	0	1
oleaginous	Jatropha	3	2	1	0
sugar	Bagasse	2	1	0	1

Figure 2 - Exemplificative alternative structures of database and classifications using keywords (case b)



TYPE	KEYWORD	BLOC	GEN	DIES	ETHA	GAS
algae	Chlorella vulgaris	3-4	2	1	1	0
algae	Dunaliella tertiolecta	3-4	2	1	1	0
livestock	Anaerobic digestion	8	1	0	0	1
crop	Corn	2	1	0	1	0
crop	Maize	2	1	0	1	0
crop	Colza	1	1	1	0	0
crop	Soybean	2	1	0	1	0
ligno	Switchgrass	4	2	0	1	0
ligno	Miscanthus	4	2	0	1	0
ligno	Poplars	4	2	0	1	0
livestock	edible tallow	3-5	2	1	0	1
livestock	animal manure	3-5	2	1	0	1
oleaginous	palm oil	1	1	1	0	0
oleaginous	vegetable oil	1	1	1	0	0
oleaginous	coconut oil	1	1	1	0	0
oleaginous	Jatropha	3	2	1	0	0
sugar	Sugarcane	2	1	0	1	0
sugar	Sorghum	2	1	0	1	0
sugar	Bagasse	2	1	0	1	0

4. Database structure and preliminary descriptive statistics

The database was obtained using Thomson Innovation, which provides access to all the available information on patents. The collected information consisted of the 72 different fields listed in Table 3 that can be classified as follows:

1. Patent identification (international, national and office codes, patents' class)
2. Patent object (title, description, claims, abstract)
3. Patent owners (applicants, inventors, assignee, buyers)
4. Patentability process stages and dates (from the application to granted patent)
5. Patent opposition (other claims on the invention)
6. Patent quality (citation)

Table 3 - Information available in the BioPat database

Publication Number, Title (Original), Title (English), Abstract, Abstract (English), Claims, Claims Count, Claims (English), Description, Assignee/Applicant, Assignee/Applicant First, Assignee – Standardized, Assignee – Original, Assignee - Original w/address, Assignee Count, Inventor, Inventor First, Inventor – Original, Inventor - w/address, Inventor Count, Publication Country Code, Publication Kind Code, Publication Date, Publication Month, Publication Year, Application Number, Application Country, Application Date, Application Year, Priority Number, Priority Country, Priority Date, Priority Year(s), Related Applications, Related Application Number, Related Application Date, Related Publication Number, Related Publication Date, PCT App Number, PCT App Date, PCT Pub Number, PCT Pub Date, IPC – Current, IPC Class, IPC Class Group, IPC Section, IPC Subclass, IPC Subgroup, IPC Class First, IPC Class Group First, IPC Section First, IPC Subclass First, IPC Subgroup First, ECLA, US Class, US Class – Main, US Class – Original, Locarno Class, Cited Refs – Patent, Count of Cited Refs – Patent, Cited Refs - Non-patent, Count of Cited Refs Non-patent, Citing Patents, Count of Citing Patents, Citing Pat 1st Assignee, Litigation (US), Opposition (EP), Opposition (EP) – Opponent, Opposition (EP) - Date Filed, Opposition (EP) – Attorney, Language of Publication

The information provided by the database can be used to study the impact of technological change on biofuel production, which is supposed to be large considering the weight of innovation effort on biotechnological sectors. It will also be possible to study the evolution of the sectoral innovation system using indicators that capture the dynamics of innovations, their concentration in terms of geographical location, holding companies and inventors.

The information collected can help to solve the problem of defining and measuring the magnitude of inventions and the problematic distinction between the cost of producing invention and the value it creates, containing many items of information such as the identity and the location of applicants and inventors, the technological area of the invention and citation of previous patents. The latter is a fundamental part of the total amount of information contained in the database. It follows a cumulative view of the process of technological change (Weitzman, 1996 and 1998) so that each inventor benefits from the work of colleagues before, and in turn contributes to the base of knowledge upon which future inventors build.

All information provided by the "patent opponent" section can be qualitatively exploited to verify if, due to existing connections between biofuel production and plants and, moreover, due to inter-linkages between biofuel raw materials and pharmaceutical raw materials, limitations to the patentability of living materials affect the innovation process of the sector. Starting from the TRIPs' model (Art. 27),[4] two main trends can be distinguished: a moderately liberal pattern represented by the U.S. patent system, and a more restricted system as designated by the European directive and, to some extent, by the EPO practice. "Since the adoption of the Agreement, the differences in the treatment of biotechnological inventions among developed countries have been reduced, but not eliminated", noting "plant varieties and animal races are not patentable in Europe, while they are eligible for protection in the USA" (UNCTAD-ICTSD, 2005, p. 388).

Differences in USA and EU patentability limitations and exclusions are just one of the aspects that can be studied. Patent applications can be viewed as a noisy indicator of the success of the innovation process, with the "propensity to grant a patent" possibly varying over institutions[5] (de Saint-Georges and van Pottelsberghe de la Potterie, 2011). Nevertheless, different regimes in patenting procedure are strongly reflected in the number of patents, the length of patentability *iter* and the scientific quality of the patents (that can be effortlessly tested by using information on citation). Finally, comparing patents from different institutions can reveal which organization manages the possessed information better, making this information clear and available to everyone.

Patents citations represent a useful tool to skip over the variability problem in terms of patent value by quantifying the impact of knowledge contained in a specific patent on subsequent innovation through the analysis of citation data (Narin et al., 1997; Jaffe and Trajtenberg, 2002). A patent can be weighted with the number of received citations. The number of patent citations can be used to characterize the technological and economic impact of a given invention providing a more meaningful measure of inventive output than a simple patent count. Moreover, patent citations can also represent an important instrument for studying some aspects of knowledge diffusion and technological spillovers such as the geographical distribution of citations, inventors and patentees (Jaffe et al., 1993).

All the patents downloaded using our methodology amount to 1,293,197records, including duplicates (21% EPO, 59% USPTO, 20% WIPO, considering both applications and grants). Then, using this initial information, we tried to make the database suitable for our purposes. First of all, 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 numerous countries. This allowed us to create a panel database that raises the number of studied countries, listed in Table 4, to a total of 37.[6]

[4] The Trade-Related Aspects of Intellectual Property Rights (TRIPS) Agreement is Annex 1C of the Marrakesh Agreement Establishing the World Trade Organization, signed in Marrakesh, Morocco on 15 April 1994. The TRIPS agreement introduced intellectual property law into the international trading system. In 2001, the Doha declaration clarified the scope of TRIPS, stating, for example, that TRIPS can and should be interpreted in light of the goal "to promote access to medicines for all" and should respect the traditional knowledge of tribal communities. The declaration also mentioned the patentability of living materials. TRIPS also specifies that the protection and enforcement of all intellectual property rights shall meet the objectives of contributing to the promotion of technological innovation and the transfer and dissemination of technology, to the mutual advantage of producers and users of technological knowledge and in a manner conducive to social and economic welfare and a balance of rights and obligations.

[5] In fact, the USPTO is often criticized for its propensity to grant many low quality patents. See The Economist (March 17, 2011) and Lemley and Sampat (2008).

[6] Figure 37 represents the highest number of countries considered so far in a environmental technology field. For instance, Johnston et al. (2009) considered 25 countries.

Table 4 - Selected countries in BioPat for descriptive statistics

US (United States of America), TH (Thailand), SG (Singapore), SE (Sweden), RU (Russia), PT (Portugal), NZ (New Zeeland), NO (Norway), NL (Holland), MY (Malaysia), MX (Mexico), LU (Luxemburg), KR (South Korea), KP (North Korea), JP (Japan), IT (Italy), IN (India), ID (Indonesia), HK (Hong Kong), GR (Greek), GB (Great Britain), FR (France), FI (Finland), ES (Spain), DK (Denmark), DE (Germany), CN (China), CH (Switzerland), CA (Canada), BR (Brazil), BE (Belgium), AU (Australia), AT (Austria), AR (Argentina), AE (Arab Emirates).

Table 5 displays the number of patents divided by patent office for the main countries considered here.[7]

Table 5 - Count of records and share of patents by main country and patent office

Country	Count	Share	EPO	WIPO	USPTO	EPO %	WIPO %	USPTO %
US	272,234	21.1	81,038	103,124	88,072	30.5	39.6	11.5
JP	129,683	10.0	79,158	5,465	45,060	29.8	2.1	5.9
DE	84,675	6.6	20,693	6,882	47,100	7.8	6.5	6.1
CA	55,348	4.3	3,100	7,528	44,720	1.2	2.9	5.8
GB	40,288	3.1	15,481	17,717	7,090	5.8	6.8	0.9
CH	28,633	2.2	11,153	10,787	6,693	4.2	4.1	0.9
FR	26,715	2.1	8,405	5,827	12,483	3.2	2.2	1.6
NL	18,433	1.4	8,937	5,802	3,694	3.4	2.2	0.5
Others	535,224	41.4	7,150	49,761	478,313	2.7	19.1	62.4

At the present stage, given the difficulty of managing data deriving from different patent offices at the same time, we decided to start with an analysis of data collected from the EPO source since it significantly reduces data management problems compared with other sources.

With regard to EPO patents, we subsequently asked the team of experts from ENEA to validate our database. We started validating the same classes indicated in the GI filtered with our keywords. The sample was built as follows: we took the EPO patents in our database; we selected the patents that shown at least one IPC class indicated by the GI, eliminated the duplicates and delivered 1% of the selected patents to the experts from ENEA.

The results of the validation are summarized in Table 6 which shows that our methodology allowed the percentage of patents actually related to the sector to be doubled. Additionally, the share of patents directly related to the investigated sector also increased.

Table 6 - Validation of BioPat for EPO patents: percentage of patents related to the biofuels sector

	Green Inventory	Share of biofuels related patents between direct and indirect application	Green Inventory filtered by keywords	Share of biofuels related patents between direct and indirect application
Direct application in biofuels	5%	28	15%	40
Indirect application in biofuels	14%	72	23%	60
Total	19%		38%	

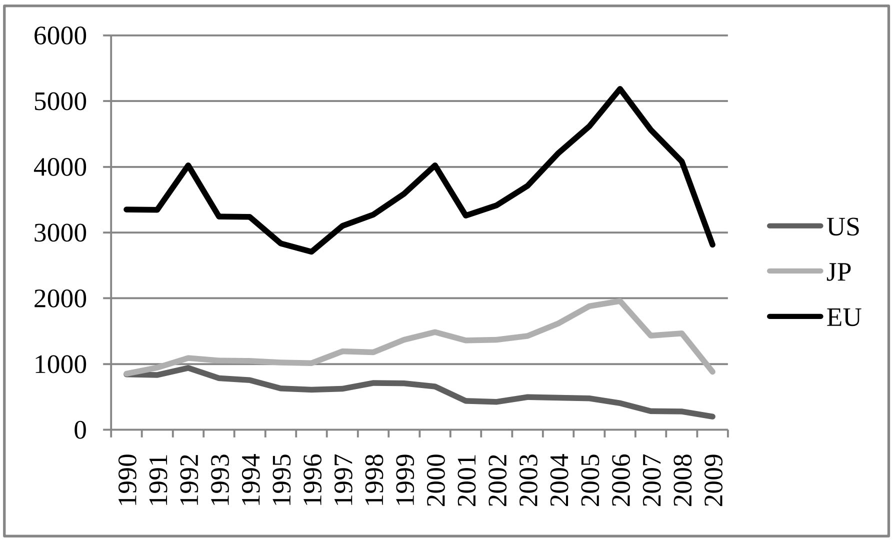
[7] Our methodology results particularly effective for EPO because the address contained in the variable is consistent in all records. As shown by Table 11.5, the variable "assignee address" is not exploitable for USPTO.

In order to provide some preliminary descriptive evidence deriving from the collected information, Figures 2 and 3 show the evolution of patenting activity registered at the EPO since 1990 for US, Japan and EU countries. As a common practice in literature (Johnstone et al., 2010; Picci, 2010), we opted to cut the series (4 years) considering the lag between the innovation efforts to be transformed into an output innovation measures as patents.

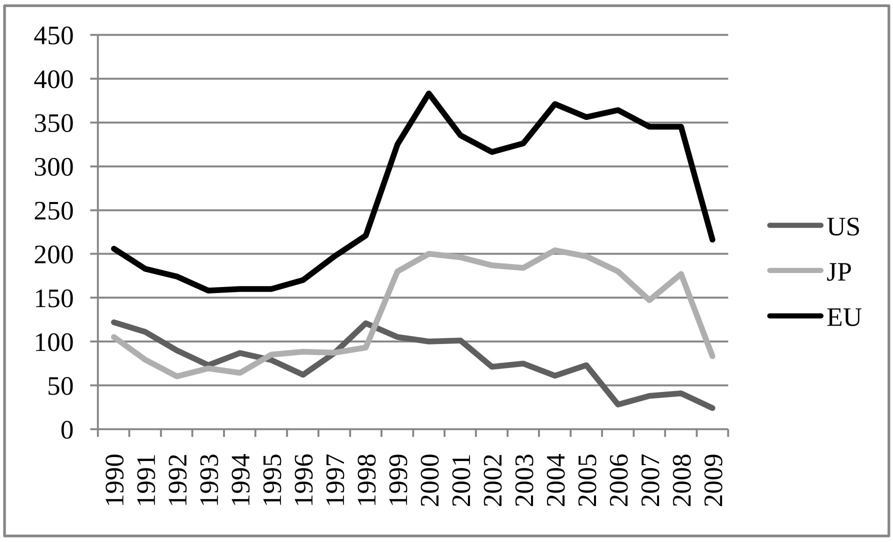
Figure 3 shows the evolution of patenting activity for EU, Japan and US from 1990 to 2009 as captured by the BioPat database and the subsample referring to patents in the GI classes which are present in BioPat. Although the number of patents differs significantly, the trend of the two series shows similar results. In particular we can observe in both patents count an increase of patenting activity at the beginning of the second decade for European countries and Japan and a constant slow decrease for US (consistently with previous findings shown in Johnstone et al. 2010 for other green technology domains). Moreover, the effect of the recent economic crisis is clearly visible in the two series.

Figure 3 - EPO patents count by country and year, Green Inventory in BioPat and BioPat total

EPO patents count by country and year, BioPat



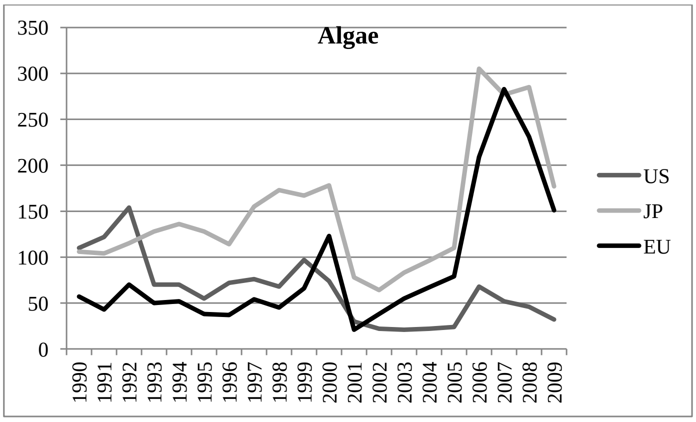
EPO patents count by country and year, Green Inventory in BioPat

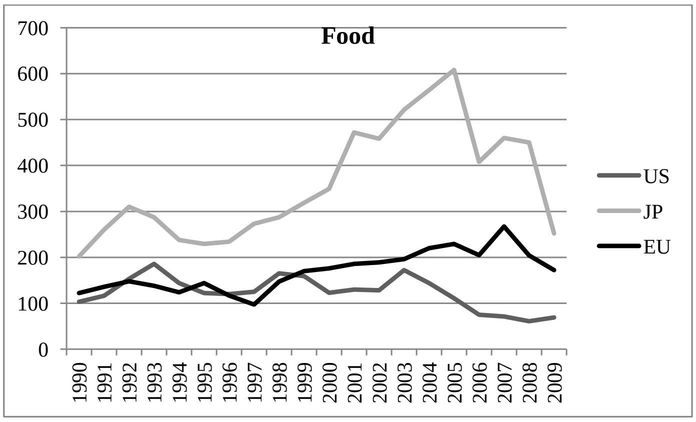


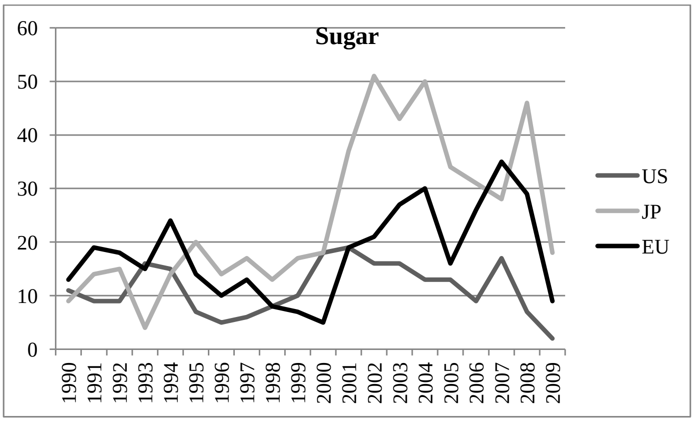
Finally, Figure 4 shows the patterns of innovation output in the biofuels sector by using all the keywords referring to specific types of raw materials. The food series confirm that old generation biofuels represent more mature technologies, with a high number of patents and more regular performance. In these fields, Japanese patenting activity shows a peak in the 2004 year and a significant decrease later on, whereas US and EU show a more regular trend and a recent slow decrease In the sugar series, the three countries seem to have a pretty common trend, with an increase of the patenting activity in the second decade, especially for Japan. In this regard, it worth reminding that the sugar-based biofuel industries rely on very traditional production process, and that the main innovation activity in this field consists in irrigation and agricultural best practices. On the other hand, the two less mature technologies, algae and ligno, show a clear increasing trend after the period 2006/2007, in particular for EU, consistenly with the European Biofuels Policy oriented towards a strong promotion of environmental sustainability standards to be respect in biofuels production process.

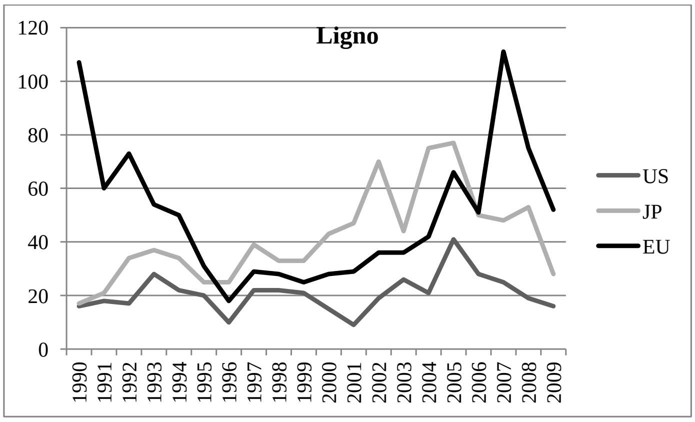
Hence, we can conclude that the trend identified in Figure 3 is mainly driven by techologies related to old generation raw material (food), while strong heterogeneity in terms of trends and patents number exists in the dynamics of patenting activities associated with different technology generations.

Figure 4 - Patterns of innovation in the biofuels sector by using all keywords referring to specific raw materials









5. Conclusions

This paper has analysed issues associated with the measurement of innovation activities through patents in a narrow economic sector such as the biofuels sector. 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 which includes patents strictly related to the investigated field, we developed an original method based on keywords, rather than on International Patent Classification (IPC) codes. Starting with a systematic mapping of biofuel production processes, we built a simplified but comprehensive description of technological domains related to the production of biofuels by applying so-called process analysis. The keyword selection is based on an iterative approach based on the analysis of recent scientific literature. The construction of the database allows a distinction to be made between innovations in raw materials and transformation processes. Moreover, both materials and processes were divided into first generation and new generation, as well as according to the biofuel type. The database was finalized by a series of interviews with experts in biofuels, and compared with IPC-based biofuel codes, revealing improved accuracy when selecting data using our methodology.

Our preliminary descriptive findings show that the distinction between different technology generations can provide interesting insights into the evolution of technologies in the biofuels sector. Moreover the information contained in the database will allow in depth scrutiny of the characteristics, determinants and effects of innovative activities in this sector. In particular, the possibility of constructing indicators that capture the dynamics of patenting activities, their value, their concentration in terms of geographical location, holding companies and inventors will allow better comprehension of the sectoral innovation system that is being examined.

References

Archibugi, D., Pianta, M. (1996). Measuring technological through patents and innovation surveys. *Technovation*, Vol. 16 (9), pp. 451-468.

Corrocher, N., Malerba, F., Montobbio, F. (2007). Schumpeterian patterns of innovative activity in the ICT field. *Research Policy*, Vol. 36 (3), pp. 418-432.

Costantini, V., Crespi, F. (2012). Public policies for a sustainable energy sector: regulation, diversity and fostering of innovation. *Journal of Evolutionary Economics*, DOI:10.1007/s00191-010-0211-3.

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*, Vol. 49 (13), pp. 1206-1223.

EC. (2005). *Biomass action plan*. Brussels 7.12.2005, COM(2005) 628 final. http://europa.eu/legislation_summaries/energy/renewable_energy/l27014_en.htm

EC. (2006). *European Union biofuel strategy*. COM(2006) 34 final, Brussels 8.2.2006. http://europa.eu/legislation_summaries/internal_market/single_market_for_goods/motor_vehicles/interactions_industry_policies/l28175_en.htm.

EC. (2007). *Renewable energy roadmap*. COM(2006) 848 final, Brussels 10.1.2007.http://europa.eu/legislation_summaries/energy/renewable_energy/l27065_en.htm

EUROSTAT. (2007). *Energy, transport and environment indicators.* Eurostat Pocketbooks, Luxemburg, Eurostat.

Griliches, Z. (1990). Patent statistics as economic indicators: a survey. *Journal of Economic Literature*, Vol. XXVIII, p.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*, Vol. 74, pp. 413–432.

IEA. (2007). *World energy outlook 2007*. Paris.

IEA. (2008). *From 1st to 2nd generation biofuels technologies. An overview of current industry and R&D activities.*

IPC Guide. (2011). Available at http://www.wipo.int/export/sites/www/classifications/ipc/en/guide/guide_ipc.pdf.

Jaffe, A.B., Trajtenberg, M., Henderson, R. (1993). Geographic localisation of knowledge spillovers as evidenced by patent citations. *Quarterly Journal of Economics*, Vol. 108, pp. 577-598.

Jaffe, A.B., Trajtenberg, M. (2002). *Patents, citations and innovations*. The MIT Press, Cambridge, MA.

Johnstone, N., Haščič, I., Popp, D. (2010). Renewable energy policies and technological innovation: evidence based on patent counts. *Environmental & Resource Economics*, Vol. 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*, Vol. 12 (1), pp. 141-148.

Lemley, M., Sampat, B. N. (2008). Examiner characteristics and the patent grant rate. *Stanford Law and Economics Olin Working Paper*, No. 369, http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1329091

De Looze, M-A., Lemarie, J. (1997). Corpus relevance through co-word analysis: an application to plant proteins. *Scientometrics*, Vol 39 (3), pp. 267-280.

Narin, F., Hamilton, K. S., Olivastro, D. (1997). The increasing linkage between U.S. technology and public science. *Research Policy*, 26(3), pp.317-330.

OECD-FAO. (2010). *Agricultural Outlook 2010-2019.* OECD-FAO, Paris.

Picci, L. (2010). The internationalization of inventive activity: a gravity model using patent data. *Research Policy*, Vol. 39, pp. 1070-1081.

de Saint-Georges, M., van Pottelsberghe de la Potterie, B. (2011). A quality index for patent systems. Working Papers ECARES. ECARES 2011-010, ULB. Universite Libre de Bruxelles.

Sirilli, G. (1997). Science and technology indicators: the state of the art and prospects for the future, in: Antonelli G. and De Liso N. (eds) *Economics of structural and technological change.* London, Routledge.

The Economist. (March 17, 2011). Patent reform: the spluttering invention machine – America's patent system has problems; a new law would fix only a few.

Weitzman, M. L. (1996). Hybridizing growth theory. *American Economic Review,* Vol. 86, pp. 207-212.

Weitzman, M. L. (1998). Recombinant growth. *Quarterly Journal of Economics,* Vol. 113, pp. 331-360.

UNCTAD-ICTSD. (2005). *Resource Book on TRIPS and Development*. Cambridge university press, New York.

UNEP. (2010). *Global trends in green energy 2009: new power capacity from renewable sources tops fossil fuels again in US, Europe*. Press release, July 2010

Appendix

Table A1 - Examples of keywords

fame	eicosapentaenoic acid scenedesmus	peanut
fatty acid methyl esters	corn	oil-bearing organisms
fatty acid ethyl esters	maize	jatropha curcas
free fatty acid	cassava	jatropha
lipids as feedstock	grain	babassu coconut
lipids microbial organisms	soybean	helianthus tuberosus
fatty acyl-acp thioesterase	genetically engineered microbes	oleaginous microorganisms
fatty acyl-coa/aldehyde reductase	genetically modified crops	rhodotorula glutinis
fatty aldehyde decarbonylase	ligno-cellulosic	medicago sativa l.
acyl carrier protein	perennial grasses	nut shells
volatile fatty acids	forest	sugarcane
microbial lipids	panicum virgatum l	beet
microbial hosts	perennial plant	sorghum
trichosporon	phalaris	sugar esters
agricultural feedstocks	alfafa	bagasse
starch	reed canarygrass	fermentable sugars
corn cobs	fibrous plant materials	cooking oil
corn stover	switchgrass	wet organic wastes
cereal straw	bark	monosodium glutamate wastewater
forest harvest residues	wood shavings	urban wood residues
husks	chip boards	ammonium
chlorella vulgaris	garden mulch	animal waste
spirulina maxima	vegetative grasses	anlage
nannochloropsis sp.	miscanthus	excreta
scenedesmus obliquus	prairie grass	feed mixture
dunaliella tertiolecta	short rotation forest species	fibrobacter succinogenes
scenedesmus dimorphus	eucalyptus	kalium
chlorella emersonii	poplars	lignocellulose

chlorella protothecoides	lignin	liquid manure
chlorella minutissima	cellulose	microorganisms
dunaliella bioculata	hemicellulose	ruminococcus albus
dunaliella salina	wood process residues	sewage
microalgae oil	wheat chaff	siloxane
phaeodactylum tricornutum	animal fat	sulphide
vegetable oil	edible tallow	digested sludge
soya oil	animal manure	fibrous material
untreated raw oils	granular sludge	hydrolysate
oilseed rape	porcine pancreatic lipase	liquid manure
coconut oil	rapeseed	mesophilic bacteria
jojoba (limited to biodiesel)	palm oil	microbial consortia
canola oil (limited to biodiesel)	organic material	sludge
methanogenic bacteria	animal slurries	treated wastewater