Advanced modelling of a biodiesel production process from *Ricinus Communis* alternative fuel in sub-Saharan African countries

PhD THESIS

Prepared by

Serge Tshibangu

Student No 0512200A

Submitted to

School of Chemical and Metallurgical Engineering, Faculty of Engineering and the Built Environment, University of the Witwatersrand, Johannesburg, South Africa

Supervisor: Dr. Diakanua Nkazi.

November, 2018

DECLARATION

I declare that the thesis titled *"**Advanced modelling of a biodiesel production process from ricinus communis alternative fuel in sub-Saharan African countries**"* is my own unaided work, except as cited and acknowledged in the references. This PhD thesis has not been submitted for any degree.

Name: **Serge Tshibangu**

Student No 0512200A

Signature: …….

Date: ……..

DEDICATION

This work is dedicated to:

My dear family Tshibangu,

My beloved wife Josline Kapinga Kalonji Tshibangu

ACKNOWLEDGEMENTS

In the Name of Jesus-Christ, the Saviour and Lord of all.

First of all, thank you to you my Lord Jesus-Christ, the Most High, Most Powerful, Most Merciful and the One full of favour and grace for assisting me to complete my PhD successfully. Thank you for guiding and protecting me throughout this period and for putting so many good people in my PhD journey. I am grateful to you for standing right next to me in the midst of storms that arose since I had started my PhD studies.

Sincere thank you and real appreciation to everyone who has contributed to the successful completion of this thesis:

I hereby offer my deep gratitude and high consideration to my supervisor, Dr. Diakanua Nkazi, School of Chemical and Metallurgical Engineering at the University of the Witwatersrand, for the patience, guidance, encouragement, constructive criticism and leadership he has given me during my PhD research. I learnt a lot from him, without his commitment and help, I could not have completed this thesis successfully.

Special thanks to my beloved wife, Josline Kapinga Kalonji Tshibangu, for her unconditional support, love and prayers. Thank you for always being next to me and helping me keep focus on my PhD study. Without you, this period of my studies would not be easy. May this work inspires you to make the same dream becomes a reality in your career.

Also, thanks to my Pastor Bryan Coetzee, to my family and my in-laws for the supports, prayers, encouragements and for standing by my side through good and adverse times.

ABSTRACT

Oil and Natural Gas are amongst the depletable resources of fossil energies. These characteristics have major implications in the supply demand equilibrium. For example, in a schematic way, the shortage (or the fear of shortage like during the oil crisis of 1973) automatically triggers a sudden price increase which decreases the demand according to the political economy of energy. The fear of shortage in the Sub-Saharan context is also an old issue in the African oil industry that always seeks to know what alternative solutions Africa has and how can the continent benefit from them. It is in this context that this work constitutes an extensive research that focused on biodiesel produced from *ricinus communis* oil. The goal of this research was the development of an advanced mathematical model of a *ricinus communis* oil based biodiesel production process, taking into account all phases, from planting *ricinus communis* trees, extracting oil, to biodiesel production. All these phases were thoroughly studied and models were developed based on mathematical principles. Firstly, a mathematical model based on the Pareto-Levy law was constructed in order to observe the sizes of plantations and the quantities of *ricinus communis* oil extracted. Secondly, another mathematical model was developed for the biodiesel production phase.

With regards to the first phase, the growth of *ricinus communis* plants was investigated: 65 *ricinus communis* trees were planted in the Limpopo province of South Africa. Plants grew and *ricinus communis* beans were manually harvested within a constant interval of 24 months. Observations of the growth of *ricinus communis* trees suggested that there was neither branch mortality nor amortisation. Results of the mathematical model developed for the first phase indicated that *ricinus communis* plants growth was not stochastic because its probability of ramification was as high as 0.95. Also, the probability of the *ricinus communis* branching process remained constant, at a value of 0.95. For each tree trunk, findings indicated that the probability of its process was reduced to 0.87 and its resulting probability ratio was equal to 0.84.

During the second phase, *ricinus communis* oil extraction through mechanical and chemical operations was performed. The extracted oil was a mixture of fatty ester acids where the main acid was ricinoleic. This acid imparted specific proprieties to *ricinus communis* oil, including the solubility into alcohol. It was also accompanied by oleic, linoleic and stearic acids.

In addition, the problem of *ricinus communis* oil reserves was covered in a qualitative point of view and it justified the use of the Pareto-Levy distribution in order to model the quantities of *ricinus communis* oils stored for biofuels production. And the development of the mathematical model for oil extraction process and reserve storages was performed using the Doehlert experimental plan and response surface methodology (RSM). Thus, the reformulated Derrida's Random Energy Model (REM) of *ricinus communis* oil reserve storages indicated that it was natural to represent those quantities by using the jumps of stable subordinators.

Furthermore, modelling of oil extraction based on *ricinus communis* plantations that had different sizes was designed. For this purpose, one important part of this study consisted on developing the best mathematical model that took into account the size inequality of the plantations. The main constraint for the development of this model was the low quantity of available data because they were whether inexistent, or expensive to obtain. Therefore, a model in which the estimation of parameters required a minimum input data was constructed. The model was constructed around mathematical observations and it was based on the description of *ricinus communis* plantations and oil. Also, in order to address the problem of inequalities observed on the size of plantations and on the quantities of *ricinus communis* oil, it was constructed based on the Pareto-Levy law. And, it also represented the *ricinus communis* oil production process, covered the indication of the *ricinus communis* plantations structures and indicated the types of biodiesel production performed.

In addition, while the stoichiometry of the reaction required 6 methanol moles for 1 *ricinus communis* oil mole (6:1 ratio) in order to obtain 6 esters moles of fatty acid and 1 glycerol mole, the performance was less compared to the performance obtained with the 24:1 ratio that was used and which corresponded to 66% of excess of methanol. Thus, under the model constructed, by increasing the methanol/*ricinus communis* oil molar ratio from 6:1 to 24:1, ester content was increased from 60.5% to 95.2%. This indicated that our model provided a fairly better performance than results observed in the literature. At the beginning of the experiment, both 3:1 and 4:1 molar ratios (33% of excess in methanol) were used. Further experiments were performed with 5:1 molar ratio (66% of excess in methanol) and so on until 24:1 and 25:1 molar ratios. From the excess of alcohol, a 90-95.2% conversion was achieved, which was good. It was also evident that a 3:1 ratio gave a lower performance and made intermediaries lower products such as diglycerides and triglycerides.

During the third phase, biodiesel was produced from an alkaline transesterification reaction of *ricinus communis* seed oil. The transesterification reaction was performed on two different tests without changing the operational conditions (molar ratio and temperature of reaction) and experimental planning was operated to evaluate the impact of temperature in product yield and quality; 13 experiments were performed. During the first test, the duration of the reaction was set to 1 hour and during the second test, the duration of the reaction was set to 4 hours. The calculation of reaction performance of each test indicated that the 1-hour duration was not enough to convert the triglycerides into esters. Meanwhile, the 4-hour duration produced a conversion performance of 95.2%, which was aligned with the optimal performance found in the literature. Also, the coefficient of determination $R^2 = 0.883$ indicated that the model was able to predict more than 88% of the total variance. The optimum performance, with a value of 84%, was obtained for the ethanol/vegetable oil molar ratio of 15:1, a KOH mass concentration of 0.5% and a temperature of 35℃.

Table of Contents

List of Tables

List of Figures

NOMENCLATURE

CO_2	carbon dioxide
FAO	Food and Agriculture organization of the United Nations
ha	hectare
Co,H_2	synthesis gas
ETBE	Ethyl tert-butyl ether
DMF	2,5- Dimethylfuran
BtL	Biomass to liquid
TG	triglyceride
DG	diglyceride
MG	monoglyceride
G	glycerol
E	alkyl ester
C12	lauric acid
C14	myristic acid
C16	palmitic acid
C18	stearic acid
C18:1	oleic acid
C18:2	linoleic acid
C18:3	linolenic acid
Arrhenius equation	$\boldsymbol{k} = AT^n \exp(-E/RT)$
ROH	Alcohol
FFA	Free Fatty Acids
$\bar{X}$	Mean
V	Variance
V_i	Mean internal variance
P	Probability of elongation
μ	Pareto-Levy law

RSM	Response surface methodology
X’	The transposed matrix of X.
X’ X	Matrix of information
$(X'X)^{-1}$	Matrix of dispersion
ρ_i	Density of the liquid
ρ_{ref}	Density of the liquid of reference
μ_i	Viscosity of the liquid studied
μ_{ref}	Viscosity of the liquid of reference
t_i	Time of flow of the liquid studied
t_{ref}	Time of flow of the liquid of reference
$\widehat{H}$	Spline function

CHAPTER 1 INTRODUCTION

1.1 Background and motivation

For the past 20 years, the world is witnessing a significant increase in the primary energy demand, which means an increased use of fossil fuels. Such a development could not happen without bringing a certain number of ecological problems due to greenhouse gases and emissions such as the carbon dioxide (CO_2) resulting from human activities. Meanwhile, CO_2 emissions come from a variety of natural sources; the main activity that emits CO_2 is the combustion of fossil fuels (coal, natural gas, and oil).

The Kyoto protocol (1992) clearly showed specific commitments for reducing emissions of greenhouse gases by 5.2% in 2012 compared to 1990 emission levels. In November 2005, 156 of 192 independent countries approved, ratified and acceded to the protocol. In this context, the search for alternative energy became a major global concern. Among the solutions in the transportation sector, biofuels obtained by direct use of biomass or transformation processes such as pyrolysis, fermentation or gasification become a potential fuel candidate.

Biofuels are renewable fuels produced from vegetable oils, animal fats and waste. Depending on the industry, light biofuels such as alcohol are produced by alcoholic fermentation of sugars or hydrolysis of starch, gaseous fuels (Dihydrogen or methane) from vegetable oils or animal fats, and solid fuels from charcoal. The ecological balance of these green fuels is significantly better than that of the petroleum products, which could favour their usage. In facts, it is renewable energy that releases into the atmosphere only the equivalent CO_2 absorbed by the plants during their growth, provided that they are maintained [1]. By taking into account pollutant emissions from the production chain (seed land with a tractor, fertilizer pollution, etc.), it is estimated that biofuels reduce emissions of greenhouse gases by 80 to 90% compared to petroleum products [2, 3].

A major drawback of the development of first-generation biofuels is that they compete with food crops. New branches of second generation biofuels from lignocellulosic biomass sources (leaves, wood, straw ...), gradually emerge with better returns [4]. Nevertheless, the use of pure vegetable oils in biofuel engines would appear as an attractive solution to enhance biomass energy reserves. But in general, the straight use of vegetable oils on recent vehicle engines, for example where electronic components are present and which require a lot of

precisions, generates combustion problems [5]. In fact, although vegetable oils have the calorific capacity similar to that of standard fuels, their high viscosity at room temperatures often leads to the damage of injection systems [6]. Also, crude vegetable oils are composed of three fatty acid molecules linked to a glycerol molecule and which constitute the glycerine that creates problems when it is used as a substitute for fuel such as diesel. On the other hand, South Africa and other Sub-Saharan African countries are particularly vulnerable to the adverse effects of climate change, which makes them to be sensitive to environmental imbalances caused by global warming affecting the planet. As the transportation sector remains in expansion, this sector constitutes the most important source of greenhouse gas emissions and one of the main sources of global warming. The use of bioenergy as alternative renewable and less polluting fuels seems interesting.

Furthermore, beyond South Africa, eight hundred million people live in Sub-Saharan Africa and a vast majority doesn't have enough energy [7]. It is estimated that by 2050, 1.95 billion people will be trying to live off the land in the Sub-Saharan region of Africa [8]. It is in this perspective that most Sub-Saharan African countries are currently showing interests in the production of biofuel as an alternative to fossil fuel. Since then, the biofuel sector has been one of the key focus points of discussions within the region. And there is a growing consensus that oil crops constitute the basis for biofuel production in the region [9]. Also, in order to develop the energy production and clean fuels in the Sub-Saharan African region, it is not possible to rely on crops that are destined to human or animal consumption. It would be wiser to turn to second-generation biofuels produced from cellulose waste or specific crops. Thus, as part of this work, the choice of biomass that provides natural oil fell on the *ricinus communis* or Castor oil plant. In fact, beside the most dominant oil crops (palm, soybean, rapeseed and sunflower), the Sub-Saharan African region has the potential to use *ricinus communis* beans as a good resource for oil. The use of such oil can rapidly gain momentum in the region, where large volumes of this plant can be used for biofuel production [10]. Furthermore, with 44 African states, including South Africa, members of the Sub-Saharan African region, and according to the Food and Agriculture Organisation of the United Nations (FAO), about 874 million hectares (ha) of Africa's lands are considered suitable for agricultural production [11]. And the FAO estimates that the size of lands available for cultivation in the Sub-Saharan African region alone amounts to more than 700 million ha [12]. More interestingly, tens of millions of agricultural lands in Sub-Saharan Africa are able to produce *ricinus communis* plants [13]. It is in this perspective that there is a strong believe

that the production of *ricinus communis* oil plants for biofuel production has the potential to elevate the socio-economic life of the people of the region.

It is evident from the number of researchers listed in the literature that an important amount of research is being done to model the performance of second generation biofuels production. However, these works are producing results that are going in different directions, making it unclear whether second generation biofuels are sustainable alternative fuels for the Sub-Saharan African region. Some of the noticeable difference in the reported values of key performance factors is due to the distinction in assumptions and input parameter identification made by researchers. This condition of the study of second generation biofuels such as *ricinus communis* triggers a considerable importance for a new, advanced and yet flexible model with which a researcher can modify those assumptions and input parameter selections, so that the key drivers of process performance can be identified and determined.

1.2 Research objective

The aim of this research is to develop a mathematical model for biofuel production from planting of *ricinus communis*, the extraction of its oil to biofuel production. A proposed mathematical model will be derived using specific variables. In order to achieve the aim of this research, the below objectives will be investigated:

- ✓ Modelling of the *ricinus communis* fields plants;
- ✓ Optimization of oil extraction by varying seed pre-treatment temperature, extraction temperature and pressure for different seed load;
- ✓ Optimisation of methyl ester production by varying reaction time, temperature and oil-catalyst ratio;
- ✓ Assessment of existing oil extraction and biodiesel production models on the one based on castor oil;
- ✓ Development of mathematical models from first principle using process parameters such as time, temperature and oil-catalyst ratio and models validation.

These objectives will assist:

- ✓ To improve the understanding of the process;
- ✓ To help in performing experiments;
- ✓ To facilitate qualitative and quantitative interpretations of experimental results;
- ✓ To help with the process of optimisation;
- ✓ To develop estimators and process control loops.

Chapter 2: Literature review

2.1 Introduction: Political, economic and environmental contexts.

Fossil fuels are still used in the Sub-Saharan region of Africa. Due to factors such as the demographic growth and the high demand among African countries, the consumption of oil has exponentially increased which makes African countries to be dependent on oil. According to a 2009 joint study by the African Development Bank and the African Union, the total African consumption of oil in 2006 was close to 3 million barrels per day [16]. Globally, this consumption will reach 107 millions of barrels per day in 2030 [17] and the transport sector will be then the source of 80% of this increase [17]. Another study conducted by the OCDE shows that 95% of global transports depend on oil and its consumption is higher than 60% [18].

The availability of oil resources is an important source of preoccupation in Sub-Saharan Africa as the global production had already been predicted according to the Hubbert's peak theory [19].

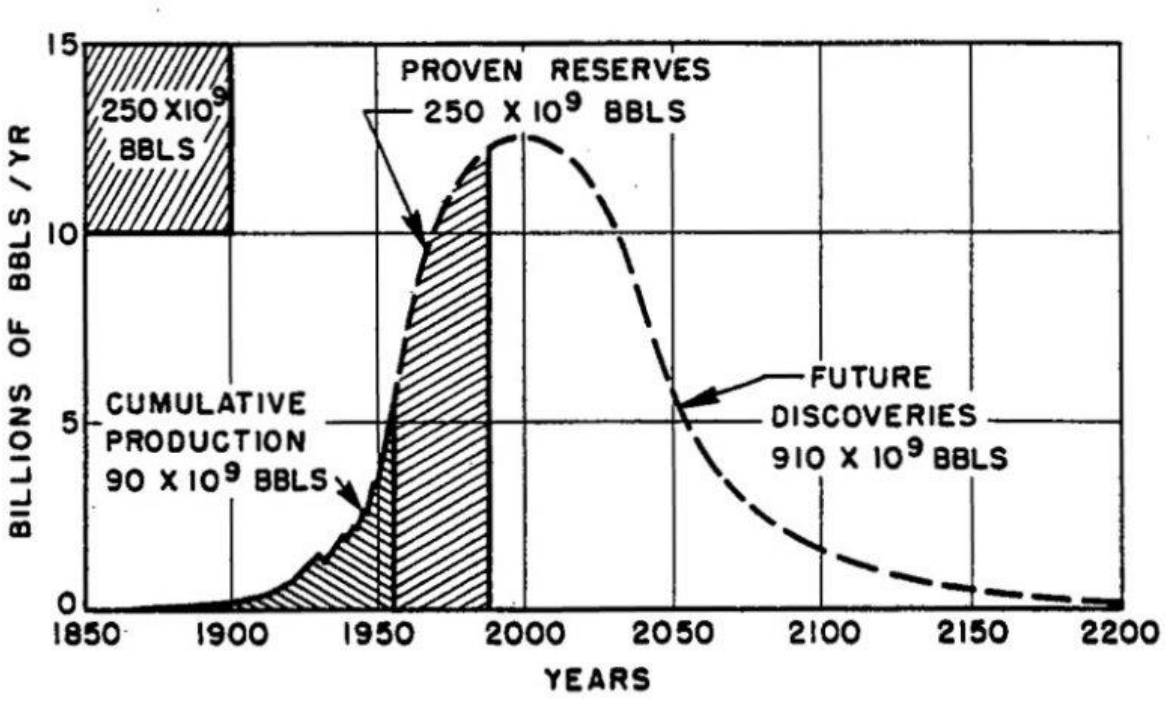


Figure 2.1: Graph of the Hubbert's peak theory [19]

Hubbert showed that the production of oil and gas in the USA will first peak before it declines. It happened that the peak happened in 1971. As it was established that Hubbert was entirely right, his studies were analysed in order to retender global resources, including oil. This was not an easy exercise because it had implied the adaptation to the oil shock that made

the production to drastically decrease. However, the extrapolation of his research performed by other experts enables to determine the peak of the production that varies according to the hypothesis of growth and to the discovery of oil reserves [20]. Meanwhile, the International Energy Agency (IEA) projected that in 2030, in the absence of public policies in the field, the global energy demand will grow by about 50% from 2002 to 2030 [20].

In general, most projections agreed with a global production decline of fossil combustibles at the beginning between 2010 and 2050 as well as an impoverishment of reserves during the next century [19]. The last decade has highlighted the fact that fossil energies are not inexhaustible and that there is a need to think of alternative sources of energy in order to satisfy the global energy demand. In addition, alternative energies will help to reduce the greenhouse effect. The future belongs to the energy sobriety and to the growth of energies that do not emit greenhouse gases. However, nowadays scientists continue to make considerable efforts on the research of alternatives to fossil energies.

In a goal to replace oil fuels, many studies have been conducted on alternative non-renewable and renewable energies. Researches on non-renewable energies are mainly based on coal and natural gas. In 2009, 89% of the energy produced in a country such as Canada (15 exajoules) came from non-renewable sources [21]. For example, coal can be converted into synthesis gas (Co,H_2) through gasification [22], then into fuel by the Fischer-Tropsch reaction, into methanol or synthetic natural gas by catalytic processes [23].

In 2013, the renewable energy sources used in the world have been mainly biomass and waste (58%) [24], hydroelectricity (31%) [25] and other sources (12%) among which we have wind, thermal and solar energies [26, 27].

In the United States, the energy produced from biomass is expected to increase by 300% between 2017 and 2030 in order to reach 153 billion of kWh. Biomass is a source of energy that is widely used to produce biofuels of 1st, 2nd and 3rd generations [27].

It is in this perspective that the industrial production of biofuel is carried out by transesterification of vegetable oils and methanol in homogeneous basic catalysis. And methanol itself is synthesized from fossil resources. Substituting methanol with ethanol has generated a lot of interest because ethanol is produced from renewable agricultural resources and can be locally produced in the region [28].

The biofuel production line from ethyl esters allows the access to better energy independence and has environmental advantages on the methyl esters stream because the carbon of the ethyl esters is entirely of photosynthetic and renewable origin.

The production of esters is dependent on many reaction factors, the main ones being the ethanol/vegetable oil molar ratio, the amount of catalyst and the temperature of the reaction. The performance of ethyl esters can be elevated (>90%) if those factors are optimized. Despite all this, the development of industrial ethyl esters production is still limited because of the difficult separation of the ethyl esters from the medium of reaction, the water content of bioethanol, which may reach 5% according to the production process [29].

2.2 First generation biofuels

First generation biofuels correspond to those obtained from vegetable cultures [30, 31, 32]. They mainly correspond to ethanol-based combustibles which are obtained from the fermentation of sugars (Maize, sugar beet, sugar cane and so on), from oil-based fuels (crude oil, biodiesel, renewable diesel) produced from catalytic hydro deoxygenation of oils [20, 21] and from oilseeds (Colza, palm, etc).

Biogases emitted by raw materials or by landfill sites [22] are also part of the biofuels of first generation. We recall that the use of vegetable oils as fuel is not a new concept per-se. In fact, during the world exposition of Paris in 1900 [33], the inventor of biodiesel engines RCK Diesel demonstrated the possibility of using vegetable oils as substitute to fuel in a compression ignition engine. However, this success was set apart for economics motives because of the low oil price. The major obstacle to the commercialisation of biofuel was and still is its cost of production in which the price of raw materials is about 75-80% of the total operating expenses [34, 35].

Biofuels of first generation can be divided into two big families: The first family of first generation biofuels is diester derived from rapeseed, soya or *ricinus communis* oils and that are commonly used in diesel engines. Here, crude oils are obtained by simple seed presses. The second family is bioethanol which is an alcohol obtained by the fermentation of starch crops such as as sugarcane, sugar beet or sweet sorghum. Bioethanol is commonly used in petrol engines and mostly in flexible-fuel vehicles (FFVs) which are alternative fuel vehicle that run on petrol blended with either ethanol or methanol fuel [36].

2.2.1 Biofuels made from oil

Vegetable oils can be used (up to 100%) as fuels by all diesel engines after minor modifications under specific conditions with the objective of warming the fuels in question. In order to use vegetable oils without modification, they must be mixed with an ordinary gasoil. The main disadvantage of both vegetable oils and animal fats is essentially linked to the high viscosity (Between 11-17 times higher than the diesel fuel) [37], the weak volatility that generates the formation of deposits inside the engines due to the incomplete combustion and the incorrect characteristics of the vaporisation. These problems are associated with the triglyceride grass chains having elevated molecular masses. But these inconvenient factors can be overcome.

Triglycerides that constitute vegetable oils can be transformed into alcohol esters or split up into glycerol by a transesterification reaction with methanol or ethanol. The small molecules of biofuel obtained from transesterification possess proprieties that are almost the same than those of gasoil and can be used as fuel in compression ignition engines or diesel engines. This type of biofuel does not have sulfur, is not toxic and is biodegradable. It is often called diester. It is important to mention that the energy and carbon performance are usually better when the engine is adapted to pure vegetable oil than when it comes to adapt vegetable oil to engines designed to work with products derived from petrol.

2.2.2 Biofuels made from alcohol

A. Bioethanol

Bioethanol is produced through fermentation of whether simple sugars from plants (beets, sugar cane) or starch from yeast cereals (wheat, maize). It can be blended directly with petrol at contents ranging from 5 to 26% and at even higher rates for so-called "flexible" vehicles. A small proportion of ethanol can also be added to diesel, but this practice is infrequent. In Europe, ethanol is most often incorporated into petrol after transformation into ETBE.

B. Ethyl tert-butyl ether (ETBE)

ETBE is a derivative of ethanol and is obtained by reaction between ethanol and isobutene. It is usually used as a fuel additive. Meanwhile, isobutene is obtained when oil is refined. More importantly, ETBE/petrol mixture has a lower vapour pressure than ethanol/petrol mixture and thus meets the petrol specifications. Incorporating the ETBE has the following advantages:

- ✓ No volatility problem;
- ✓ High octane rating;
- ✓ Perfect tolerance to water.

C. Bio-butanol or butyl alcohol

Bio-butanol (or butyl alcohol) is enzymatically obtained from sugars to butanol-1 (acetone-butanol fermentation). The driving force behind the use of bio-butanol as a substitute for petrol is to benefit from its physical advantages to overcome the difficulties associated with the use of ethanol (volatility, aggressiveness to certain plastics used in the automobile industry). Bio-butanol can be used as fuel mixed with petrol up to 10% of the volume without adaptation of the engine. It can also be used in mixture with ethanol and petrol or even gasoil. Bio-ethanol production units can be adapted to produce bio-butanol.

D. Methanol

Methanol obtained from methane, can also be used, partially replacing (under certain conditions) petrol, as an additive in diesel, or eventually for certain types of fuel cells. However, it should be reminded that the methanol is very toxic to humans.

2.2.3 Advantages and disadvantages of 1st generation fuels

First generation biofuels have the advantage that their technical production is well mastered and mature [38]. They are the only alternative to petroleum-based liquid fuels and are directly compatible with current flexible-fuel vehicles (FFVs) that have internal combustion engines [36, 39]. But they are limited at the following points:

- ✓ The potential competition and a remarkable tension on resources linked to high agricultural area needs;
- ✓ The energy and environmental balances of these fuels are modest (especially for greenhouse gases). This is even more visible for the alcohol industry;
- ✓ The balance sheet in terms of greenhouse gas emissions appears less favourable when taking into account the conditions of production and distribution;
- ✓ The probable interference with natural ecosystems and global food systems.

2.3 Second generation biofuels

Second generation biofuels are materials-based cellulosic biofuels obtained from non-edible crops (wood, leaves, etc). The second generation feedstocks are non-edible vegetable oils and are considered the best alternative to edible vegetable oils for biofuels production.

These biofuels are: Bio-alcohols, bio-oils, 2.5-Dimethylfuran (DMF), bio-hydrogens, Fischer-Tropsch diesels, wood diesel [40].

In addition, second generation biofuels can be obtained from the biomass without any competition with the use of foods: cereals, woods, forest residuals and dedicated crops. There are two ways of producing second generation biofuels:

A. Biochemical pathway

It is related to the production of the cellulosic ethanol by fermentation. This process is performed in 3 big steps. Of the three main contents of the lignocellulosic biomass (cellulose, hemicellulose and lignin), only the cellulose is transformable into ethanol. A first step consists of the extraction of the cellulose and then its transformation into glucose by hydrolysis using enzymes. Then glucose is fermented into ethanol using yeast. Finally, ethanol is purified by distillation and dehydration.

B. Thermochemical pathway

One of the ways is the production of BtL (Biomass to liquid) synthetic diesel fuel. For the thermochemical way, the biomass is first conditioned by pyrolysis. Then, it is gaseous at up to 1000 °C in the presence of water vapour or oxygen. Synthesis gas consisting of carbon monoxide (CO) and hydrogen (H_2) is obtained. The next step is the Fischer-Tropsch synthesis, the catalytic chemical transformation of synthesis gas into linear paraffin, which, hydrocracked and isomerized, will produce a synthetic gasoil.

According to the physical state, there are three types of second generation biofuels:

- ✓ Solids: solid biocombustible;
- ✓ Liquids: bio-oil, biofuel, bioethanol;
- ✓ Gaseous: Biogas, biosynthesis gas.

The use of non-edible crops oils is crucial because edible oils are in high demand for food security. In addition, edible oils are more expensive to be used as feedstock for alternative fuel. Thus, biofuel production from non-edible oils such as *ricinus communis* oil is a better way to circumvent all the challenges linked to edible oils. As a non-edible oil, *ricinus communis* oil contains about 50-70% of oil and triglyceride with fatty acid chains composed of 90% ricinoleic acid, 4% linoleic acid, 3% oleic acid. Stearic, palmitic and linoleic acids represent less than 1% each. With this composition, *ricinus communis* oil constitutes a good source of biofuel with a prospect of good productivity for sub-Saharan African countries.

2.4 Third generation biofuels

Third generation biofuels are biofuels produced from algae and combustibles-based micro-organisms (yeast and fungi) such as vegetable oils, biodiesel, renewable diesel [41, 42]. Like second generation biofuels, third generation biofuels are better than those of first generation in terms of sustainable development because they are carbon-neutral and they reduce atmospheric CO_2 concentrations [43]. Third generation biofuels include methane produced by anaerobic digestion of algae [44], biodiesel derived from microalgae oil [45, 46] as well as from photobiological hydrogen production [47, 48]. However, the idea of using micro-algae as a source of fuel is not new but it has started to be taken seriously because of the escalation of oil prices and the global warming which are associated with the consumption of fossil fuels [49].

According to some authors, the cultivation of microalgae seems to theoretically be 30 to 100 times more efficient than the terrestrial oilseeds. In order to obtain an optimal oil performance, the microalgae growth must be carried out with a CO_2concentration of about 13%. This is possible at a very low cost due to the coupling with a CO_2 source. For example, CO_2 produced from coal, natural gas, biogas thermal plant, alcoholic fermentation unit or cement plant. By comparison, for each ton of microalgae biomass produced, some authors estimate that 1.8 tons of CO_2 would be consumed (180% reduction) [50]. It is in this context that this study focused on the use of *ricinus communis* oil for biofuel production.

2.5 Factors of biofuel production

The *ricinus communis* plant based on the seed oil content is able to produce biofuel. Factors to be considered are formulated as follows:

- ✓ Tolerance and absorption of CO_2;
- ✓ The temperature of tolerance;
- ✓ The stability in cultivation inside specific bioreactors;
- ✓ By-products;
- ✓ Vulnerability to infection and to potential herbivores;
- ✓ Harvest and process;
- ✓ The potential manipulation of genetic engineering.

Nowadays, main biofuel produced are: Methyl ester and bioethanol. And the production of biodiesel comprises several steps as shown in Figure 2.2.

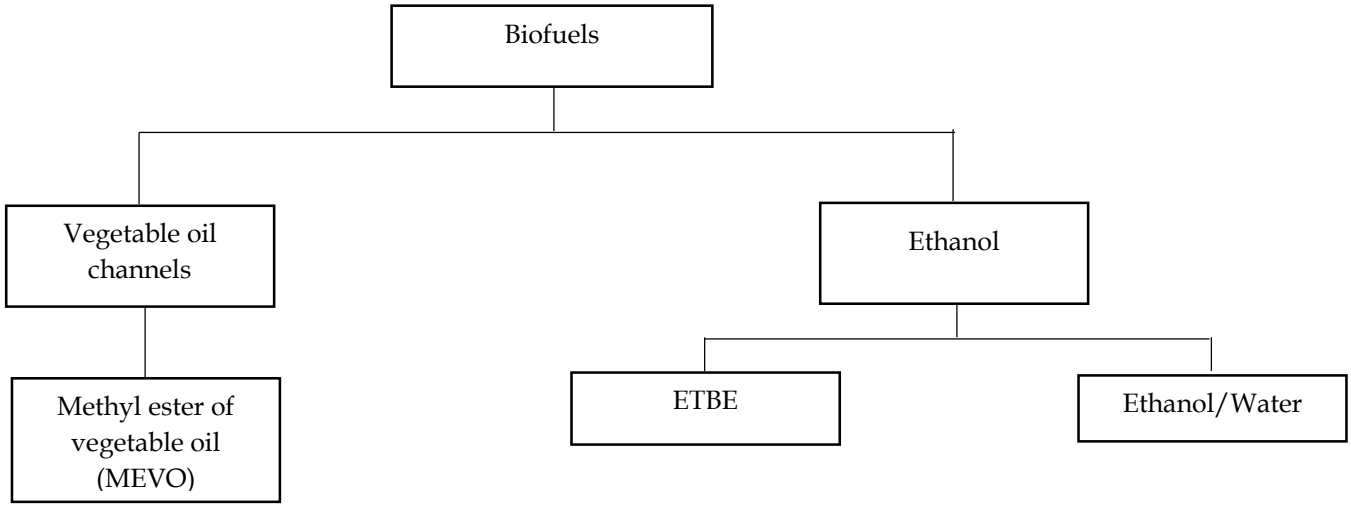


Figure 2. 2: Biofuel channels [51]

In addition, Figure 2.5 is used to display all steps of oil transformation into biofuel.

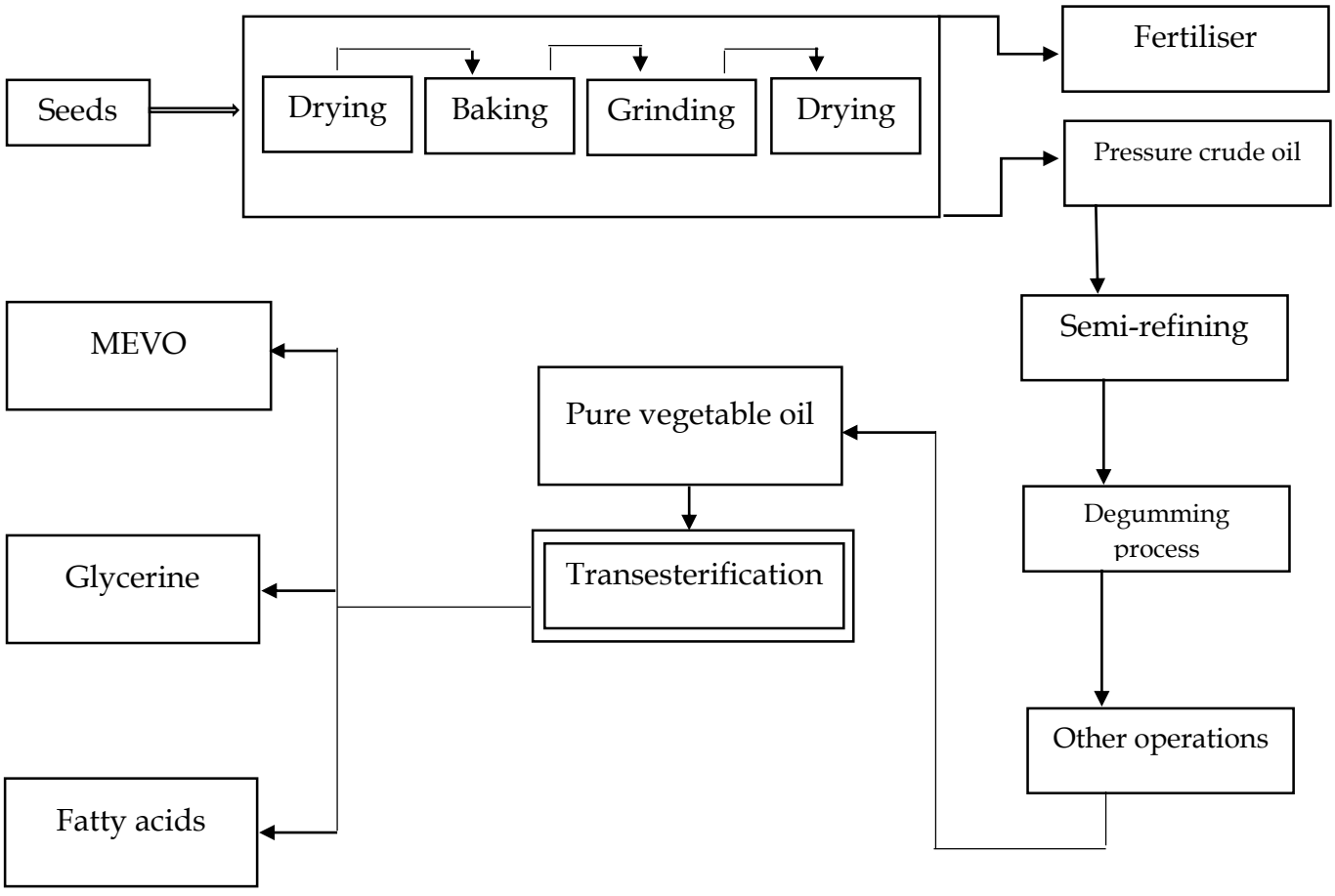


Figure 2. 3: Steps of oil transformation [52]

2.6 Oils: Various sources, criteria of selection, composition and characteristics

2.6.1 Various sources of oil

In principle, any source of fat body can be used to prepare the biodiesel. However, some sources are more preferred than others depending on countries. For example, in the United States, manufacturers prefer the use of soybean oil. USA is the largest producer of soybean oil before Brazil. On the other hand, Brazilians use different sources of oil because of the biodiversity of the country. For example, in the north of Brazil, it is mainly palm oil and soybean oil whereas in the centre-west, it is soybean oil, cotton oil, castor oil, and sunflower oil. In France, producers mainly use rapeseed oil [53]. In the case of South Africa, the biofuel production technology is fairly well established. The preferred crops for biofuels production in South Africa are second generation crops which are not dependent on food crops, including soya, canola and sunflower for biodiesel and sugar cane and sweet beet for bioethanol [54]. Four years ago, South Africa had decided that sorghum should be used as biofuels feedstock and soy beans as biodiesel feedstock in order to reduce food security concerns [55].

2.6.2 Oil selection criteria: Competition with food, price and profitability

The source of biodiesel must as much as possible meet two very important criteria:

- ✓ The low production prices; and
- ✓ A large scale of production.

The price of edible oils, such as Soybean oil is higher than diesel. For this reason, waste vegetable oils and non-edible vegetable oils are preferred as a source of fats for the production of biodiesel. Also, it is possible to use frying oils, rendering oils, various animal oils, such as fish oils and even fats, since for esters formed with these oils, it is possible to have pour points reaching more than 10 °C [56].

On the other hand, in order to choose the source of biodiesel, it is important to consider the percentage of oil in the plant and the oil yield per hectare. Let’s take the example of the US, where replacing diesel transport with biodiesel requires 0.53 trillion cubic meters of biodiesel per year depending on the speed of the current consumption.

The table below shows the oil yield per hectare for different oil sources, the areas required and the percentage of the US growing area [55]. In order to produce 50% of the fuel oil in the US, 24% of the cultivable area of the country must be used for the cultivation of palm oil

(considered the most profitable oil). It is clear that the vegetable oil could not significantly replace petroleum derivatives in the future. However, this scenario dramatically changes using micro-algae as sources. In fact, the area required is reduced between 1 and 3%.

A comparison of a selected number of biofuel sources other *ricinus communis* oil is displayed in Table 2.1.

Table 2. 1: Comparison of biofuel sources other than *ricinus communis* [56]

Plant	Oil yield (liter/ha)	Land area required (mha)	Percentage of the area of cultivation of the US
Maize	172	1540	846
Soy	446	594	326
Canola	1190	223	122
Jatropha	1892	140	77
Coconut	2689	99	54
Palm	5950	45	24
Microalgae 70% of biomass oil	136.900	2	1.1
Microalgae 30% of biomass oil	58.700	4.5	2.5

2.6.3 Chemical characteristics composition of oils

The benefit of knowing the composition of oil is obvious since it will have an influence on the physical characteristics that are essential to evaluate the quality of the resulting biofuel such as the viscosity, the melting point as well as its thermal stability. From a chemical point of view, the composition of oils depends on the variations of cultivation conditions (soil, climate and humidity). Vegetable oils consist mainly of triglycerides, free fatty acids and by-products. By-products with maximum content in the range of 1 to 5%, in a normal crude oil are essentially:

- ✓ Phospholipids (lecithin, cephaline), gum sources
- ✓ Carotenes and xanthophylls (polyunsaturated) very polymerisable
- ✓ Tocopherols (natural antioxidant)
- ✓ Free or esterified steroid alcohols and triterpenic alcohols
- ✓ Monoglycerides and diglycerides
- ✓ Water trace.

In some cases, the wrong artisanal fabrication of oil or an inadequate storage can lead to an important presence of these by-products. Free acids derived from the hydrolysis of triglycerides can be present with various contents from less than 0.5% for comestible oils up to 60% or even more for very acidic oils. Triglycerides are long molecules whose hydrolysis leads to the glycerol and to three fat acids. Figure 2.6 shows the general chemical structure of triglycerides:

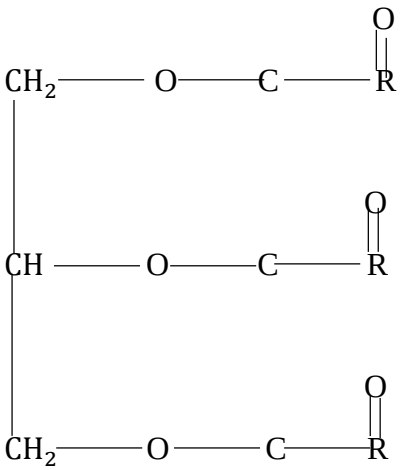


Figure 2. 4: Triglyceride general chemical structure

Fatty acids are divided into three categories: major, minor and unusual.

- ✓ Major fatty acids are the most known and are in big quantities inside the plants: lauric acid (C12), myristic acid (C14), palmitic acid (C16), stearic acid (C18), oleic acid (C18:1), linoleic acid (C18:2), and linolenic acid (C18:3).
- ✓ Minor fatty acids are secondary contents of fatty bodies.
- ✓ Unusual fatty acids are polyunsaturated fatty acids with conjugated double liaisons, acetylic fatty acids (with triple liaisons). Triglycerides fatty acids differ by the length of the chain, the degree of unsaturation and the presence of other chemical functions.
- ✓ Fatty acids are constituted of a linear chain of a pair number of atoms of carbon going from 8 to 22 atoms.

The fatty acid composition of the vegetable oils is summarized in Table 2.2.

Table 2. 2: Fatty acids composition of vegetable oils [57, 58]

	Fatty acids composition (%)									
Vegetable oils	14:0	16:0	18:0	20:0	22:0	24:0	18:1	22:1	18:2	18:3
Maize	0	12	2	Trace	0	0	25	0	6	Trace
Cotton	0	28	1	0	0	0	13	0	58	0
Lin seed oil	0	5	2	0	0	0	20	0	18	55
Nuts	0	11	2	1	2	1	48	0	32	1
Canola	0	3	1	0	0	0	64	0	22	8
Soy	0	12	3	0	0	0	23	0	55	6
Sunflower	0	6	3	0	0	0	17	0	74	0

It was noted that triglycerides of certain oils such as rice oil are composed of various fatty acids, present in diverse proportions. Also, other oils are essentially composed of only one type of fatty acid. It's the case of canola oil essentially composed of triglyceride fatty acids. That justified why canola oil was the most used in literature even though in this work, the literature related *ricinus communis* oil was used. In addition, literature indicated that fatty acids was saturated, monounsaturated or polyunsaturated. Those double liaisons were found in chains of more than 14 carbon atoms but were not present at the end of the chain. They were fragile and easily oxidable in order to form peroxides and then carboxylic acids. That was the reason why oil must always be stored far from light, heat and humidity. In fact, the influence of the length of the chain was the priority when it was related to the degree of unsaturations. It was further noted that some acids possessed functional groups such as unusual triglycerides. In a basic environment, these molecules produced alcoholate. And alcoholate was able to enter into competition with the reactive, which dropped the performance. In this regard, oils were characterised by important basic chemical characteristics: iodine index (II), acidity index (AI) and saponification values (SI) [59].

A. Iodine index (II)

This was the diode mass (I2) (expressed in g) capable of binding itself to the unsaturations (most often double liaison) of fatty chains of 100 g of fat. It is used to measure the number of unsaturations. Oil with a high iodine value is not very resistant to oxidation, which can cause problems in combustion (cracking and polymerization in engine combustion chambers). Saturated, oil can be resistant to oxidation but often solid at room temperature.

B. Acidity index (AI)

Acidity index (AI) expresses the mass of potash (expressed in mg) required to neutralize the free fatty acids contained in 1g of fat. This index is determined at cold state. It gives an idea of the thermal stability of the oil especially the smoke point and the flash point which sharply drop when the acidity of the oil increases. This can also lead to corrosion problems.

C. Saponification Index (SI)

Saponification index (AI) determines the mass of KOH in mg required to saponify the combined fatty acids in one gram of fat. The saponification number is therefore an indirect measure of the molar mass PM of fatty acids and subsequently of oil.

$$PM = \frac{1000 \times 56 \times 3}{AI} \text{ for neutral oil.} \quad (2.1)$$

$$PM = \frac{1000 \times (1 - \frac{AI}{SI}) \times 56 \times 3}{SI - AI} \text{ for acid oil [59]} \quad (2.2)$$

2.6.4 Comparison of physical proprieties of oils, their esters (biodiesel) and gasoil

Table 2.3 displays the comparison of the physical properties of gasoil with those of oils and esters.

Table 2. 3: Compared physical proprieties between oils, their esters and gasoil [60]

Oil or ester	Density (g/ML)	Viscosity @ 37.8°C (mm²/S)	Cetane number	Cloud point (°C)	Flash point (°C)	Lower heating value (MJ/kg)
Gasoil of reference	0.832	1.6-6.0	45.0	-17.8	46	35.3
Nuts oil	0.921	41.2	41.5	+3	237	34.1
Methyl ester of nuts	0.883	4.9	54.0	+5	176	33.6
Soybean oil	0.923	36.8	38.5	-4	219	34.0
methyl ester of Soybean	0.885	4.5	45.0	+1	178	33.5
Palm oil	0.918	39.6	42.0	+31	267	35.0
methyl ester of palm	0.88	5.7	62.0	+13	164	33.3
sunflower oil	0.924	37.1	35.5	-5	232	34.0
methyl ester of sunflower	0.88	4.6	49.0	+1	183	33.5
canola oil	0.92	30.2-40°C	35		decomposition >320	37.7
methyl ester of canola	0.88-0.885	4.5-40°C	51		170-180	39.96

The properties of esters derived from oils depended on the nature of oils used especially the cetane number, the cloud point and the flash point [60]. However, the properties of esters were close to those of gas oil. And the main benefit of using esters instead of their oils was the reduction of viscosity and the limit temperature of filtering. In addition, viscosity decreased by a factor between 7 and 10.

In addition, the cetane number was used to evaluate the auto-inflammation aptitude of gas oil in the scale between 0 and 100 [61]. With regards to oil, it was observed that transesterification clearly improved the cetane number. Therefore, it appeared that the cetane number index was improved with the molecular weight of fatty acid as well as with the molecular weight of alcohol for the same fatty acid; on the other hand, it was decreased with the number of unsaturations [62, 63]. In some parts of the world such as in Europe, the lower cetane number of gasoil was about 45 and the flash point of esters was higher than that of gas oil. This made them less dangerous fuels to be manipulated compared to gas oil. And the stability included the thermal stability at high and low temperatures, resistance to oxidation, the polymerisation, the water absorption and the microbial activity.

Like seen above, the main source of biofuel instability was the chain of unsaturated fatty acids. And the iodine index enabled the measurement of the degree of unsaturation in a biofuel. The unsaturation led to the formation of deposits and of problems of stability in the stored biofuels because this last became less resistant to oxidation. For example, the soy and canola methyl esters had iodine indices of 133 and 97 respectively. Research conducted at Mercedes-Benz concluded that biodiesel whose iodine index was higher than 115 was not acceptable due to the excessive deposits of carbon products [64]. All these physical characteristics were very important and must be respected according to international standards in order to produce biofuels that can be used.

2.7 Overview of *ricinus communis* plant and its potential to be used as source of biofuel.

Ricinus communis is a plant and a flowering specie plant in the spurge family called Euphorbiaceae [65]. It is a native of India and the most yield countries of its seed are China, Brazil and Thailand. The plant is very variable in habit and appearance [65]. Tropical latitudes and dry lands are favourable conditions for the growth of the plant as it naturally grows in dry parts of a country, where it can become a 9 to 12 meters high tree. In warmer Mediterranean regions, the plant is of meagre growth, attaining an average growth of 3 to 4 meters high. The cultivated varieties grow to a height of 60–120 Cm within a year, and several metres in perennial cultivation [66]. The plant grows in the humid tropics to the subtropical dry zones (optimal precipitation 750–1000mm, temperature 15–38°C) which makes it suitable for cultivation in many places of Southern Africa [67].

One of the agronomic advantages of *ricinus communis* plant is its resistance to drought, semi-arid climate and it can grow on poor soils. Its irrigation is not a problem because it is a less exigent specie in terms of water, fertilisers and maintenance. On the contrary of other plants, *ricinus communis* plant can grow anywhere in Sub-Saharan Africa, even in soils with sands as well as in salted soils and it does not necessitate a special irrigation mechanism. The plant can be easily multiplied in laboratory by in vitro culture, seeding or cuttings. It produces seeds at the second and third seasons following its sowing. It also has the particularity of not being eaten by animals because of its toxicity.

Depending on species and environments, the production of *ricinus communis* oil varies between 1 to 2 tons per hectare. However, each hectare of *ricinus communis* planted in semi-arid regions can produce up to 350-900 kg of oil per hectare, making it a good potential substitution for petroleum in the manufacturing of biofuels, plastics and lubricants. This leads to the conclusion that there is no hesitation in the choice of biomass. In addition, *ricinus communis* can be grown on marginal lands, which means it cannot compete with food crops. The plant is poisonous and the natural oil contained in its shiny dark green or reddish leaves is non-edible and is used in the pharmaceutical or cosmetic industry.

2.7.1 Botany of *ricinus communis*

In Latin, the term Ricinus means “tick”: The seed has that name because it resembles to ticks that have markings and a bump [68]. *Ricinus communis* is a species in the Euphorbiacae family containing 8100 other species. This plant is the sole representative of the *ricinus* which is a tree with big palmed leaves [69, 70]. And Euphorbiaceae comes from the Euphorbia family dedicated by Juba II, King of Mauritania, to his personal physician Euphorbus [71]. With its very variable aspect, plants of the Euphorbiacae family are characterised by their white latex, irritating the skin and large with a fruit that has three seeds [72].

As seen in the literature [73], *ricinus communis* plant is originated from Sub-Saharan Africa and it is developed as an ornamental plant in various regions of Asia, North America, Africa and Europe [73]. It is largely cultivated in most tropical and subtropical dry regions as well as in many temperate zones with a hot climate [74, 75]. The plant can be found all around the Sub-Saharan African region, including in South Africa and in the Indian Ocean [7]. However, the big scale cultivation of the *ricinus communis* is done in India, China and Brazil [76].

2.7.2 Taxonomic classification

Ricinus communis seed can produce up to 47% of oil with a production of 1200 liters/ha using artisanal extractors and around 80% with a machine. The *ricinus communis* plant taxonomy is as follows:

Reign: Plantea

Phylum: Spermaphyte (seedling plant)

Sub-branch: Angiosperm (Magnoliophyta: Flowering plants)

Class: Magnoliopsida

Subclass: Rosidae

Order: Euphorbiales

Family: Euphorbiaceae

Genre: Ricinus

Species: Ricinus Communis

On top of being a source of biofuel, *ricinus communis* oil is also used as a traditional medicine product because of its medicine and cosmetic properties. In addition, the exploitation of *ricinus communis* can contribute to the development of rural areas through:

- ✓ Control of landslides;
- ✓ Reduction of poverty by the sales of seeds and derived products;
- ✓ Rural electrification by the use of renewable energy produced by local fuel for generators.

Figure 2.5 shows an extensive use of the *ricinus communis* plant in various field of life. This also highlights the fact that further research should be carried out to determine how various parts of the plant could be used for biofuel production.

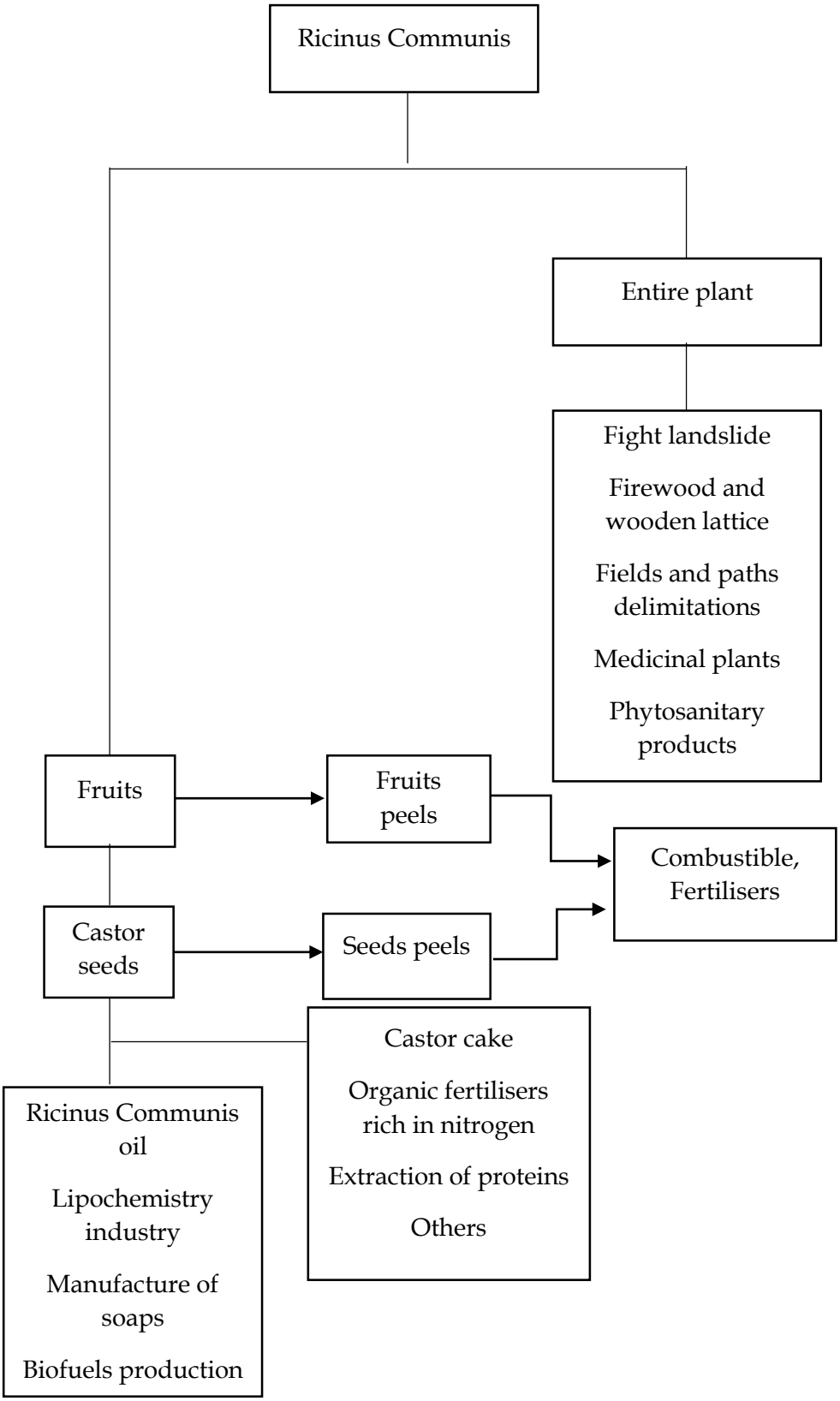


Figure 2. 5: Various uses of the *ricinus communis* plant

The most economically viable use of *ricinus communis* oil before its use as biofuel is the production of soaps with abundant foams that have positive effects on the skin, because of the content of glycerol. For the production of soaps, we will come back to it later in this work in the section related to the saponification reaction. The oilcake resulting from oil extraction constitutes an organic fertiliser of non-negligible performance: 1 ton of *ricinus communis* oilcake values 200 kg of mineral fertilisers. Also, the oilcake has insecticides properties and reduces the quantity of nematodes on the soil because of its residual oil content. The use of this fertiliser is recommended for bios culture that bans the chemical contribution in the fertilisation or the fight against parasites.

2.8. Kinetic models and transesterification process

2.8.1 Kinetic models

Unstructured kinetic models are based on the observation of macroscopic kinetic inside the reactor. They are called unstructured because they do not reflect the intracellular composition of cells. In our case, the biomass is represented by a single variable that is its concentration in the culture medium.

Almost all the growth kinetic expressions described by the authors are derived from the law established by Monod: $\mu = \frac{\mu_{max} \cdot S}{K_{sx}+S}$ [77]. This reflects the influence of limitation in substrate on the specific growth rate. We note that if $S > K_{sx}$ then $\mu = \mu_{max}$. The biomass production rate will decrease when S will be near K.

Papastratos [78], Oliveira et al. [79] and Costa et al. [80] have modelled inhibition by ethanol by a nonlinear function while Lafforgue [81], Groot et al. [82] and Warren et al. [83] have modelled it by a linear function. As this has been shown in the literature, many authors have found an inhibitory effect on the biomass concentration on the kinetics of growth whose magnitude depends on the amount of accumulated cells.

The study of kinetics on the transesterification of vegetable oil for production of biodiesel determines the parameters for predicting the extent of reactions at any given time under certain conditions. The kinetic involves the determination of activation energy, reaction rate equations and reaction rate constants. Even though the significance of biodiesel as an alternative fuel has grown over the last two decades, the chemical kinetics of transesterification remains controversial [84]. In the literature, the kinetic models focus on determining the best fit of empirical data to simple models of the reaction order.

De Lama da Silver [85] carried out a kinetic study of the transesterification reaction. The results showed that the vapor pressure osmometry (VPO) technique was a robust methodology for determining vegetable oil molecular weight presenting lower cost than gas chromatography analysis. However, this method has the disadvantage of being an averaging technique which is almost unable to provide information about the various impurities that may be present in large molecules.

Jain and Sharma [86] performed a kinetics study of two-step acid-base catalyzed transesterification process carried at optimum temperature. As mentioned by the authors, that were the first of its kind which dealt with simplified kinetics of two step, acid-catalyzed, transesterification process, carried under optimum conditions [87].

Back in the 1980's, Freedman et al. [88] reported the transesterification reaction of soybean oil and other vegetable oils with butanol and alcohols, and studied what were the effects of the types of alcohol and molar ratio, the types and amounts of catalyst as well as the effect of temperature on constant reaction rates and kinetic order. Constant reaction rates were determined, and the effect of other reaction parameters was investigated such as alcohol to soybean molar ratio, temperature, catalyst type and concentration. The authors proposed a completely reversible three-step second-order reaction model in order to describe the mechanism. Their findings indicated that at a molar ratio of 6:1, a second-order mechanism with a fourth-order shunt mechanism best described the kinetics. The limitations of the Freedman's kinetic model were in the fact that they only considered one overall reaction which consisted of one molecule of triglyceride (TG) reacting with three molecules of alcohol (ROH) as indicated in the following equation [89]:

$$\mathbf{TG} + 3\mathrm{ROH} \rightleftharpoons \mathrm{G} + 3\mathrm{E} \tag{2.3}$$

The overall reaction in (2.3) occurred in three steps: Firstly, triglyceride (TG) broken down into diglyceride (DG) and monoglyceride (MG) to produce glycerol (G) and alkyl ester (E). This operation in included in equations (2.4) – (2.5):

$$\mathbf{TG} + \mathrm{ROH} \rightleftharpoons \mathrm{DG} + \mathrm{E} \tag{2.4}$$

$$\mathbf{DG} + \mathrm{ROH} \rightleftharpoons \mathrm{MG} + \mathrm{E} \tag{2.5}$$

$$\mathbf{MG} + \mathrm{ROH} \rightleftharpoons \mathrm{G} + \mathrm{E} \tag{2.6}$$

It appeared that Freedman's kinetic model was obtained from the application of the law of mass action to the three steps of the reaction. The forward reactions were considered to be

second order, with regards to the overall order of reactions. The concentration of alcohol was considered to be constant when the alcohol to triglyceride molar ratio was very high. The rate of reaction was only a function of concentration of triglyceride. The reaction order was known as "pseudo-first-order". When the data did not fit the model, a "shunt reaction" was proposed by Freedman [88] whereby three alcohol molecules simultaneously reacted with triglyceride.

Xiao et al. [90] developed an intrinsic kinetic model based on the Eley-Rideal mechanism according to the experimental data, which indicated a new mechanism of transesterification reaction catalysed by solid base. Results indicated that the transesterification reaction occurred between methanol adsorbed on solid base active sites and glyceride from the liquid phase. The authors found that the model calculation agreed well with experimental data. Further, findings indicated that the transesterification reaction was an endothermic reaction.

Al-Zuhair et al. [91] used the experimental determination of the separate effects of palm oil and methanol concentrations on the rate of their enzymatic transesterification to test their generated kinetic model. Their findings showed that at a constant methanol concentration of 300 mol m^{-3}, the initial reaction rate increased as the palm oil concentration increased. And the initial reaction rate dropped sharply at substrate concentrations larger than 1250 mol m^{-3}. The authors observed similar behaviour for methanol concentration effects, where at a constant substrate concentration of 1000 mol m^{-3}, the initial rate of reaction dropped at methanol concentrations larger than 3000 mol m^{-3}. Thus, the authors developed a mathematical model from a proposed kinetic mechanism and the model was used to identify the regions where the effect of inhibition by both substrates arose.

Mittelbach and Trathnigg [92] conducted experiments on the kinetics of sunflower oil using methanolysis. Their experiment did not explore any reaction rate constants, however. They determined the parameter affecting the transesterification process. Some authors discovered that the forward reaction of the conversion of triglycerides is not second order reaction as proposed by Freedman [88]. This was indicated by the first reaction step between methanol and triglyceride forming diglyceride as the rate limiting step, as other reaction steps occurred faster. They further found that the reaction rate was a function of temperature; however, the percentage conversion was a poor function of temperature as long as the reaction had taken place at a minimum duration of 10 minutes. Moreover, according to the authors, the reaction was a two-phase system, which was different to Freedman's finding of a single phase system. Also, they found that for the first two minutes of the reaction, a two-phase system occurred;

thereafter a complete solution took place for a duration of 5 to 10 minutes. After some time, when glycerol was fully formed the two phase-systems occurred again. Experiments show that the reaction followed first-order kinetics with a calculated activation energy, *Ea* = 53.99 kJ/mol while the pre-exponential factor, $A = 2.9 \times 10^7$ min^{-1}. Finally, the authors highlighted the facts that their findings could help improve an environmentally friendly biodiesel process.

Noureddini and Zhu [93] studied the kinetic of the transesterification of soybean oil using the methanol as the solvent and the NaOH as the catalyst. The authors investigated the effect of mixing intensity and temperature on the reaction rates using a molar ratio of 6:1 for methanol to soybean oil.

In their study, measurements were taken at different mixing intensities and the Reynold's number was calculated. The effect of temperature was determined using the Arrhenius equation and modified Arrhenius equation as expressed in equation (2.7) [94].

$$\mathbf{k} = \mathrm{AT^{n}exp(-E/RT)} \qquad (2.7)$$

where, n = experimentally derived parameter

In their study, the authors used a value of n=1. Looking at the results of their study, it was concluded that:

- ✓ The shunt reaction proposed by Freedman was negligibly small;
- ✓ The activation energy changed with the change in Reynold's number.
- ✓ The rate constant for the reverse reactions was larger than that of the forward reactions [94].

On the other hand, Komers' kinetic models [95] investigated the methanolysis of rapeseed oil using a heterogeneous alkaline catalyst, potassium hydroxide. For simplification purposes, Komers' kinetic models were used for any alkyl alcohol, selected ROH and any alkaline (base) catalyst. Knowing that Kinetic modelling requires simplifying assumptions, for the theoretical development of Komers' kinetic models, the following assumptions were made [95]:

- ✓ The concentration of FFAs (free fatty acids) was negligibly small;
- ✓ From the possible theoretical reactions only two reactions proceeded to form products. These two reactions were the alcoholysis of glycerides (TG, DG and MG) and saponification of TG, DG and MG or E;

- ✓ The isomers namely TG, DG, MG and E proceeded at the same rate;
- ✓ Alcoholysis was catalyzed by OH^- or RO^- (alkoxide) ions. The concentration of OH^- and RO^- ions were very small compared to the ions of TG and ROH.

Based on these assumptions, simplifications were represented as possible reactions:

Formation of alkoxide:

$$\mathbf{ROH} + OH^- \underset{k_{1r}}{\overset{k_1}{\rightleftarrows}} RO^- + H_2O \quad (2.8)$$

Alcoholysis:

$$\mathbf{TG} + RO^- \underset{k_{2r}}{\overset{k_2}{\rightleftarrows}} DG^- + E \quad (2.9)$$

$$\mathbf{DG} + ROH \underset{k_{3r}}{\overset{k_3}{\rightleftarrows}} DG + RO^- \quad (2.10)$$

$$\mathbf{DG} + RO^- \underset{k_{4r}}{\overset{k_4}{\rightleftarrows}} MG^- + E \quad (2.11)$$

$$\mathbf{MG} + RO^- \underset{k_{5r}}{\overset{k_5}{\rightleftarrows}} G^- + E \quad (2.12)$$

$$\mathbf{DG} + RO^- \underset{k_{6r}}{\overset{k_6}{\rightleftarrows}} MG^- + E \quad (2.13)$$

$$G^- + ROH \underset{k_{7r}}{\overset{k_7}{\rightleftarrows}} MG^- + E \quad (2.14)$$

Saponification:

$$\mathbf{E} + OH^- \overset{k_8}{\rightarrow} ROH + A \quad (2.15)$$

$$\mathbf{TG} + OH^- \overset{k_9}{\rightarrow} DG + A \quad (2.16)$$

$$\mathbf{DG} + \mathrm{OH}^- \xrightarrow{k_{10}} \mathrm{MG} + \mathrm{A} \tag{2.17}$$

$$\mathbf{MG} + \mathrm{OH}^- \xrightarrow{k_{11}} \mathrm{G} + \mathrm{A} \tag{2.18}$$

Where:

A= soap of the corresponding fatty acid chain.

As the effect of FFA content was studied, the following equation (2.23) needed to be taken into consideration. Equation (2.23) was considered to be a saponification reaction.

$$\mathbf{FFA} + \mathrm{OH}^- \xrightarrow{k_{12}} \mathrm{A} + \mathrm{H_2O} \tag{2.19}$$

The reaction showing the saponification of FFA was an undesirable reaction because it consumed the reaction producing soap instead of biodiesel. Therefore the kinetic model consisted of 10 components namely TG, DG, MG, G, E, M, A, OH, FFA and H_2O. Also, reactions from equations (2.8) – (2.19) were assumed to be elementary and obeyed the rate law called the law of mass action (LMA). Rate laws were helpful in calculating reaction rates by using the product of a rate constant and reactant concentrations. The direct use of LMA produced 13 differential equations. The number and complexity of the differential equations could be reduced by simplifying assumptions. There were two approaches that could be used. The first approach was the rate-limiting step approach which was based on the assumption that both the forward reaction rate and reverse reaction rate of the first reaction step were greater than the rate of reaction of the second reaction step. The second approach was the steady state assumption which was that the rate of change of concentration of all reactive intermediates was negligible.

The assumption was good because the concentrations of reactive intermediates were low since small variables had small time derivatives on condition they did not oscillate quickly. Thus, the assumption meant the following:

$$k_2, k_{2r} \ll k_3, k_{3r} \tag{2.20}$$

$$k_4, k_{4r} \ll k_5, k_{5r} \tag{2.21}$$

$$k_6, k_{6r} \ll k_7, k_{7r} \tag{2.22}$$

$$k_3, k_{3r}, k_5, k_{5r}, k_7, k_{7r} > k_8, k_9, k_{10}, k_{11}, k_{12} \quad (2.23)$$

Based on the above assumption, removing small terms and substituting them into rate equations at equilibrium the following was obtained.

$$\frac{d[H_2O]}{dt} = \frac{d[RO^-]}{dt} = \frac{d[DG^-]}{dt} = \frac{d[MG^-]}{dt} = \frac{d[G^-]}{dt} = 0 \quad (2.24)$$

Komers normalised each species by the initial concentration of triglyceride $[TG]_0$ and alcohol $[ROH]_0$ as follows:

Let a = $[TG]_0$ and b = $[ROH]_0$.

$$\mathbf{TG} = \frac{[TG]}{a} \quad (2.25)$$

$$\mathbf{DG} = \frac{[DG]}{a} \quad (2.26)$$

$$\mathbf{MG} = \frac{[MG]}{a} \quad (2.27)$$

$$\mathbf{G} = \frac{[G]}{a} \quad (2.28)$$

$$\mathbf{A} = \frac{[A]}{a} \quad (2.29)$$

$$\mathbf{OH} = \frac{[OH^-]}{a} \quad (2.30)$$

$$\mathbf{W} = \frac{[H_2O]}{a} \quad (2.31)$$

$$\mathbf{ROH} = \frac{[ROH]}{b} \quad (2.32)$$

$$\mathbf{E} = \frac{[\mathbf{E}]}{\mathbf{b}} \quad (2.33)$$

$$\frac{-dTG}{dt} = b \cdot OH\,(k_2' \cdot TG \cdot ROH - k_{2r}' \cdot DG \cdot E) + a \cdot OH \cdot DG \cdot E \quad (2.34)$$

$$\frac{-dDG}{dt} = b \cdot OH\,(-k_2' \cdot TG \cdot ROH + k_{2r}' \cdot DG \cdot E + k_4' \cdot DG \cdot ROH - k_{4r}' \cdot MG \cdot E) + a \cdot OH(-k_9 \cdot TG + k_{10} \cdot DG) \quad (2.35)$$

$$\frac{-dMG}{dt} = b \cdot OH\,(-k'_4 \cdot DG \cdot ROH + k'_{4r} \cdot MG \cdot E + k'_6 \cdot MG \cdot ROH - k'_{6r} \cdot G \cdot E) + a \cdot OH(-k_{10} \cdot DG + k_{11} \cdot MG) \quad (2.36)$$

$$\frac{-dG}{dt} = b \cdot OH\,(-k'_6 \cdot MG \cdot ROH - k'_{6r} \cdot G \cdot E) - a \cdot OH \cdot k_{11} \cdot MG \quad (2.37)$$

$$\frac{-dROH}{dt} = \frac{dE}{dt} = a \cdot OH\left(\begin{matrix}-k'_2 \cdot TG \cdot ROH + k'_{2r} \cdot DG \cdot E + k'_4 \cdot DG \cdot ROH - k'_{4r} \cdot MG \cdot E + k'_6 \cdot MG \cdot ROH - \\ k'_{6r} \cdot G \cdot E - k_8 \cdot E\end{matrix}\right) \quad (2.38)$$

$$\frac{-dH_2O}{dt} = \frac{dFFA}{dt} = a \cdot k_{12} \cdot FFA \cdot OH \quad (2.39)$$

$$\frac{-dOH}{dt} = \frac{dA}{dt} = b \cdot OH \cdot k_8 \cdot E + a \cdot OH \cdot (k_9 \cdot TG + k_{10} \cdot DG + k_{11} \cdot MG) \quad (2.40)$$

The resulting differential equations for the ten reaction components (TG, DG, MG, G, E, ROH, OH, A, FFA and H_2O). Where: k_1 *to* k_{12} = reaction rate constants ($Lmol^{-1}\ s^{-1}$)

The constants of the above reaction had the following definitions:

$$k'_2 = k_2 \cdot \frac{K_1}{W} \quad (2.41)$$

$$k'_{2r} = k_{2r} \cdot \frac{K_1}{K_3 \cdot W} \quad (2.42)$$

$$k'_4 = k_4 \cdot \frac{K_1}{W} \quad (2.43)$$

$$k'_{4r} = k_{4r} \cdot \frac{K_1}{K_5 \cdot W} \quad (2.44)$$

$$k'_6 = k_6 \cdot \frac{K_1}{W} \quad (2.45)$$

$$k'_{6r} = k_{6r} \cdot \frac{K_1}{K_7 \cdot W} \quad (2.46)$$

And:

$$K_1 = \frac{k_1}{k_{1r}} = \frac{[RO^-][H_2O]}{[ROH][OH^-]} \quad (2.47)$$

$$K_2 = \frac{k_2}{k_{2r}} = \frac{[DG^-][E]}{[TG][RO]} \quad (2.48)$$

$$K_3 = \frac{k_3}{k_{3r}} = \frac{[DG][RO^-]}{[DG^-][ROH]} \quad (2.49)$$

$$K_4 = \frac{k_4}{k_{4r}} = \frac{[MG^-][E]}{[DG][RO^-]} \quad (2.50)$$

$$K_5 = \frac{k_5}{k_{5r}} = \frac{[MG][RO^-]}{[MG^-][ROH]} \quad (2.51)$$

$$K_6 = \frac{k_6}{k_{6r}} = \frac{[G][E]}{[MG][RO^-]} \quad (2.52)$$

$$K_7 = \frac{k_7}{k_{7r}} = \frac{[G][RO^-]}{[G^-][ROH]} \quad (2.53)$$

Reactions rate were represented in a balance equation as follows:

$$\mathbf{1} = TG + DG + MG + G \quad (2.54)$$

$$\mathbf{ROH} + E = 1 \quad (2.55)$$

$$\mathbf{OH} + A = p \quad (2.56)$$

$$\mathbf{p} = [OHJ]_o/[TG]_0 \quad (2.57)$$

$$\mathbf{n} \cdot E + 3 \cdot TG + 2 \cdot DG + MG + A = 3 \quad (2.58)$$

$$\mathbf{n} = [ROH]_0/[TG]_0 \quad (2.59)$$

Initial concentrations from the above equations were as follows:

$$[TG]_0 = 1 \quad (2.60)$$

$$[ROH]_0 = 1 \quad (2.61)$$

$$[OH]_0 = p \quad (2.62)$$

$$[DG]_0 = [MG]_0 = [E]_0 = [A]_0 \quad (2.63)$$

Komers defined new equilibrium constants by combining the product of pairs of original equilibrium constants as the following:

$$K_2' = K_2 \cdot K_3 = \frac{[DG][E]}{[TG][ROH]} = \frac{k_2'}{k_{2r}'} \quad (2.64)$$

$$K_4' = K_4 \cdot K_5 = \frac{[MG][E]}{[DG][ROH]} = \frac{k_4'}{k_{4r}'} \quad (2.65)$$

$$K_6' = K_6 \cdot K_7 = \frac{[G][E]}{[MG][ROH]} = \frac{k_6'}{k_{6r}'} \quad (2.66)$$

The values of the intermediates directly leaded to:

$$\mathbf{DG} = \mathrm{K}_2' \cdot \mathrm{TG} \cdot \frac{(1-\mathrm{E})}{\mathrm{E}} \tag{2.67}$$

$$\mathbf{MG} = \mathrm{K}_4' \cdot \mathrm{DG} \cdot \frac{(1-\mathrm{E})}{\mathrm{E}} = \mathrm{K}_2' \cdot \mathrm{K}_4' \cdot \mathrm{TG} \cdot \left(\frac{(1-\mathrm{E})}{\mathrm{E}}\right)^2 \tag{2.68}$$

$$\mathbf{G} = \mathrm{K}_6' \cdot \mathrm{MG} \cdot \frac{(1-\mathrm{E})}{\mathrm{E}} = \mathrm{K}_2' \cdot \mathrm{K}_4' \cdot \mathrm{K}_6' \cdot \mathrm{TG} \cdot \left(\frac{(1-\mathrm{E})}{\mathrm{E}}\right)^3 \tag{2.69}$$

Combining equations (2.47), (2.50), (2.62) with equations (2.54) – (2.59), a relation between TG and E was formed and represented in the equation:

$$\mathbf{TG} = \frac{1}{1+\mathrm{K}_2'\cdot\frac{(1-\mathrm{E})}{\mathrm{E}}+\mathrm{K}_2'\cdot\mathrm{K}_4'\cdot\left(\frac{(1-\mathrm{E})}{\mathrm{E}}\right)^2+\mathrm{K}_2'\cdot\mathrm{K}_4'\cdot\mathrm{K}_6'\cdot\left(1+\mathrm{K}_6'\cdot\frac{(1-\mathrm{E})}{\mathrm{E}}\right)^3} \tag{2.70}$$

The desired result for the initial alcohol to triglyceride molar ratio was obtained by substituting equation (2.70) and equations (2.64) - (2.66) into equation (2.59). The equation was expressed as follows:

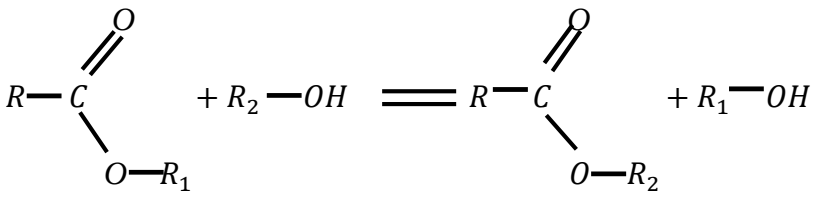
$$\mathbf{n} = \frac{1}{\mathrm{E}}\left[\mathrm{P} + 3 - \frac{3+2\mathrm{K}_2'\cdot\frac{(1-\mathrm{E})}{\mathrm{E}}+\mathrm{K}_2'\cdot\mathrm{K}_4'\cdot\left(\frac{(1-\mathrm{E})}{\mathrm{E}}\right)^2}{1+\mathrm{K}_2'\cdot\frac{(1-\mathrm{E})}{\mathrm{E}}+\mathrm{K}_2'\cdot\mathrm{K}_4'\cdot\left(\frac{(1-\mathrm{E})}{\mathrm{E}}\right)^2+\mathrm{K}_2'\cdot\mathrm{K}_4'\cdot\mathrm{K}_6'\cdot\left(1+\mathrm{K}_6'\cdot\frac{(1-\mathrm{E})}{\mathrm{E}}\right)^3}\right] \tag{2.71}$$

where: P= Base catalyst to triglyceride molar ratio

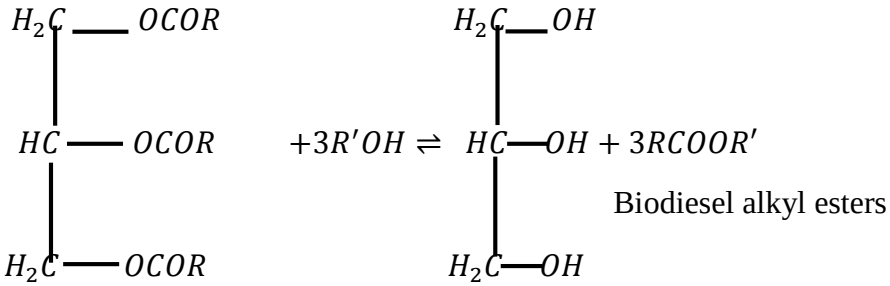
Making use of equations (2.64) – (2.66) and equation (2.70), it was possible to calculate the values of n, DG, TG, G and MG for every alkyl ester selected. The value of alcohol was determined using equation (2.55). The Komers' kinetic models helped for the prediction of the concentration of the end products using analytical expressions. The model also assisted in indicating how the reaction conditions were affected by the concentration of the reaction. An example was the increase in the concentration of ROH which increased the total reaction rate. Talebian-Kiakalaieh, Amin, et al. [96] used the RSM methodology and Artificial Neural Network (ANN) to study the relationship between process variables and free fatty acid conversion and for predicting their optimal parameters. The authors found that both the RSM and the ANN could accurately predict the experimental results, with $R^2 = 0.9987$ and $R^2 = 0.985$, respectively. Furthermore, the authors studied kinetics studies to describe the system.

2.8.2 Transesterification process

Transesterification (also called alcoholysis) is the reaction between ester and alcohol producing a different ester [97].



In this study, transesterification was considered as a reaction between triglyceride in *ricinus communis* oil with alcohol in order to produce glycerol and a mixture of monoesters used to produce biofuel [97].



Triglycerides (TG) Glycerol (G)

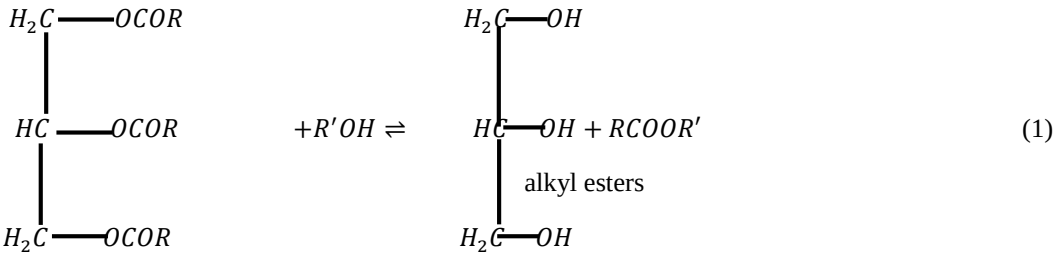
It was catalyzed by both acids and bases [97] and favoured by an augmentation of temperature, an excess of alcohol or when the operational conditions allow the glycerol decantation. The details of the reaction of transesterification, production processes of the biodiesel and various ways in the development (heterogeneous catalysis and supercritical catalysis) were studied in the following section.

2.8.2.1 Oil reaction of transesterification

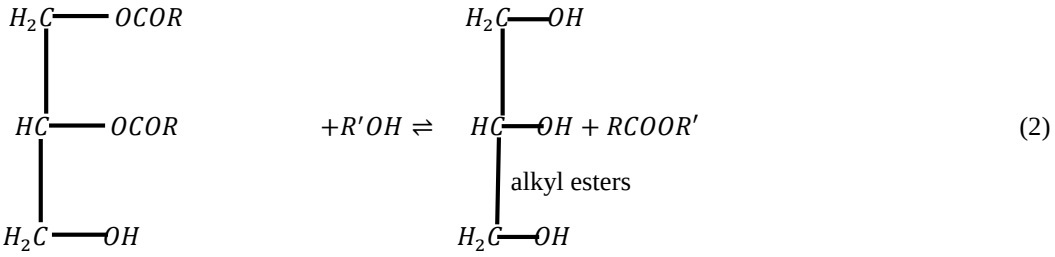
Transesterification of triglycerides is not a new process. It started in 1853 when Patrick and Duffy performed it many years before the first Diesel engine becomes functional [98]. This reaction has been at the center of intensive research because of the diverse use of products including the polyesters synthesis in the polymer industry [99], the intermediates synthesis in the pharmaceutical industry [100], the hardening of resins in the paint industry as well as in the production of biodiesel as an alternative to diesel.

2.8.2.2 Mechanism of reaction

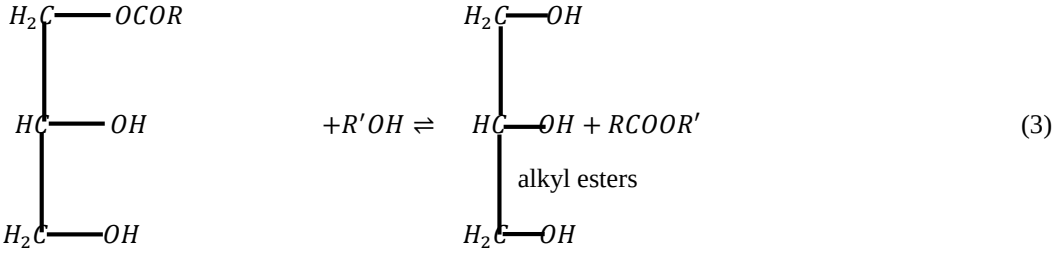
The transesterification reaction is performed in three successive phases:



(TG) (1)

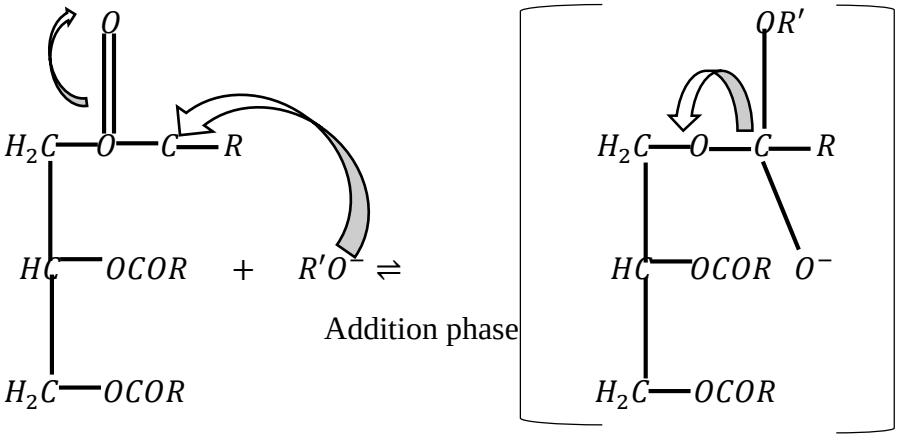


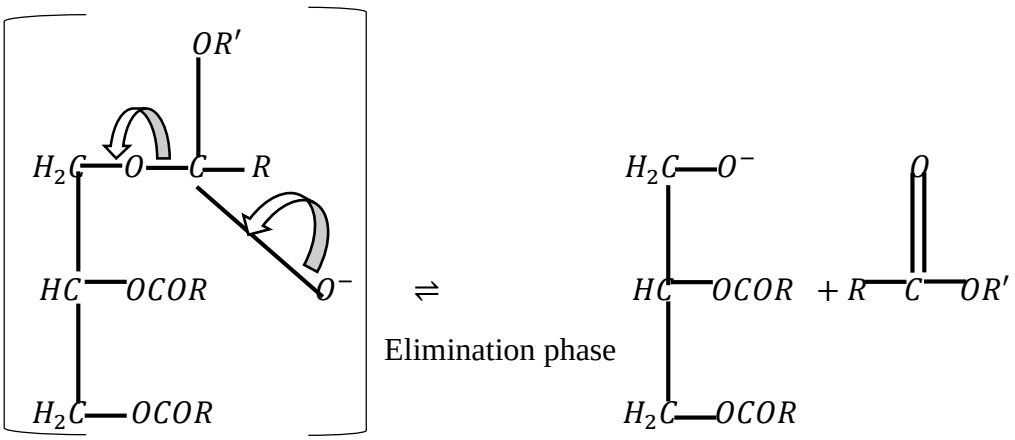
(DG) (2)

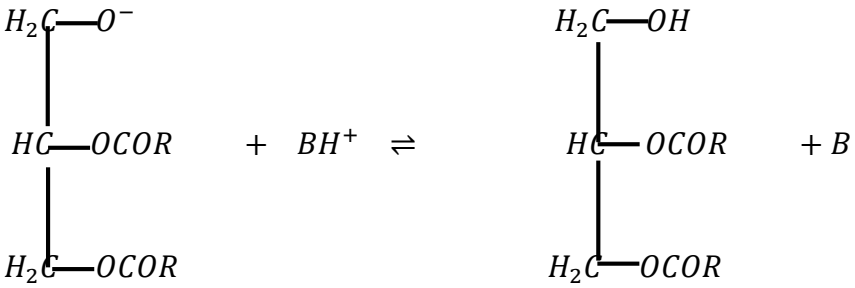


(MG) (3)

Transesterification reactions were chemically balanced. Phases (1) and (2) were quick because the primary ester functions were transesterified first. Phase (3) was slow [101, 102]. In basic catalysis, the mechanism was as follows:







First, it consisted of a nucleophilic attack of carbonyl (triglyceride) by the alkoxide anion to form an intermediate carbanion (addition phase). Then, we had the departure of the nucleofuge when the doublet of oxygen was folded (elimination phase). The alkoxide was regenerated as soon as an alcoholate function of the glycerol appeared.

With NaOH, KOH and K_2CO_3 or other similar catalysts, alkoxide that was formed was then identified as the catalytic specie [103]. The third reaction appeared to be the determining phase in the reaction because the monoglycerides were more stable intermediates than diglycerides [104].

2.8.2.3 Important parameters of the transesterification reaction of oils

The vegetable oil transesterification has been largely studied in the homogeneous environment by Bradshaw [105], Freedman [88], and Hanna [106]. These studies explain that the reaction is strongly influenced by the nature of the catalyst (acid or basic), the nature of the alcohol and the oil, the triglycerides/alcohol molar ratio, the temperature, the presence of free acids and water, the speed of the agitation and the reaction time. However, a lot of research are actually orientated towards the catalysis in the heterogeneous environment that we will see in some of the following sections.

Types of catalysts

In the case of the transesterification reaction, it was admitted that the basic catalysis was quicker than acid catalysis. Thus, the reaction time varied between 3 and 48 hours with acid catalyst, and reactions were complete after one hour with a basic catalyst [88]. However, in the case of acid catalyst, a high temperature allowed the acceleration of the speed of the reaction [107].

A. Basic catalysts

At the beginning, works were done in the presence of a basic catalysis. The most used catalysts were NaOH, KOH and NaOMe in proportions around 0.5 to 1% of the total mass and Na_2CO_3 [88]. Results obtained were between 95 and 99% after one hour reaction to the reflux temperatures of alcohol.

According to Freedman, NaOMe was a more efficient catalyst than the NaOH. The comparative study between NaOMe and NaOH shows that 0.5% of NaOMe gives the same performance than 1% of NaOH when the alcohol/oil molar ratio is 3/1 and it gives performances clearly superior when the ratio is 6/1 [88]. The principal cause of the superiority of the NaOMe compared to NaOH is that it generates less water. In facts, water will be able to hydrolyse or saponify the esters formed and diminishing the reaction performance.

Chen et al. [108] also found that both NaOH and NaOMe offered a good catalytic activity in the methanolysis of palm oil (99% of conversion) at 60 °C for alcohol/oil molar ratio of 10/1. But they observed that for the same concentration of catalyst (0.125/kg), the beginning of the reaction evolved more quickly with NaOMe than with NaOH. Furthermore, in the case of NaOMe, they obtained the same performance even with a 6/1 ratio [108].

Vicente et al [109] compared homogeneous basic catalysts (NaOH, KOH, NaOMe and KOMe) in the methanolysis of sunflower oil. The purity of biodiesel obtained was 100% for all catalysts. However, only methoxide offered a performance of around 100% in biodiesel. According to Vicente, the losses in performance were due to the saponification reaction of triglycerides and to the dissolution of methyl esters in glycerol [109]. In 2006, Meneghetti et al. [110] confirmed Vicente's results in a study that compared *ricinus communis* oil methanolysis and ethanolysis [110].

In 1998, Ma et al. [111] demonstrated that NaOH was more efficient than NaOMe. This was contradictory with the previous Freedman et al report. The authors attributed these results to the difference in the reaction system used.

B. Acid catalysts

The most used acid catalysts were mineral acids such as hydrochloric acids (HCI), sulfuric acid (H_2SO_4), phosphoric acid (H_3PO_4), etc. [88, 112].

Freedman found that transesterification in acid catalysis was 4000 times slower than in basic catalysis [88] and that reaction times were longer. Thus, for a conversion greater than 99%, the soybean oil methanolysis (at 65 °C) was completed in the presence of 1% of H_2SO_4 after 50 hours of reaction; butanolysis (at 117 °C.) and ethanolysis (at 78 °C.), with the same amounts of catalyst and alcohol, were carried out after 3 and 18 hours respectively. He noticed that the reaction was faster when the chain of alcohol was longer. The fact that butanolysis was faster than ethanolysis and methanolysis results not only from the higher temperature but also from a higher solubility of butanol in oil at the beginning of the reaction. Therefore, it was preferable to take stoichiometry alcohol/triglycerides greater than 3. Also, despite the fact that the basic catalysis was faster than acid catalysis, many studies observed that the choice of catalyst was judicial when vegetable oil or fatty acid contained a strong proportion of free acids. And those free acids had the potential to destroy the basic catalyst by reacting with it in order to form soaps; however, acid catalyst could perform the triglycerides transesterification and free acids esterification by increasing the conversion [113]. Severe conditions of temperature and pressure were then called in order to accelerate the reaction speed [114]. Acids catalysts were rarely used because of their low reactivity and the high risk of corrosions of industrial infrastructures.

C. Enzymes

The transesterification reaction of vegetable oils by enzymatic catalysis flourished in recent years. In fact, enzymes have several advantages: They are biodegradable, selective [115], which makes it possible to increase the performance of the reaction while reducing the production of by-products of the reaction. The reaction conditions are relatively mild (low temperature and pressure), which reduces the price in terms of energy and equipment and tends towards lower waste costs.

In the case of enzymatic catalysis there is neither the problem of free acids nor the sensitivity to water; it is possible to work whether in an aqueous or non-aqueous environment. With this, it was noted that enzymes may be immobilised on a support [115]. However, the enzymatic catalysis has certain disadvantages: enzymes are too expensive (for example, the lipase is more expensive than a basic catalyst) and has an unstable activity. The reaction is too slow [115] and is therefore industrially limited as it lasts between 4-16 hours to reach a conversion of 95%. However, studies were carried out in order to recycle enzymes given its high price.
In particular, research performed on the immobilization of enzymes using a support are mentioned. The activity of enzymes was inhibited by methanol and glycerol present in the reaction of the environment. Hence the idea of adding a solvent, tert-butanol in order to gradually eliminate glycerol that comes to be adsorbed by enzymes. Methanol was added stepwise to the reaction medium. This reduced the inhibitory effects thereby increasing the profitability of the process. The enzymatic production was possible by using intra and extracellular lipases.

The nature of alcohol

Because the reaction was balanced, an excess of alcohol was recommended in order to push the reaction in the direct sense. Alcohol used was a primary or secondary mono aliphatic alcohol having 1-8 carbon atoms [116]. Methanol, ethanol, propanol, butanol and amyl alcohol were among alcohols that could be used for the transesterification process. Methanol and ethanol were the most used alcohols, especially methanol because of its low cost and its physical and chemical advantages (short and polar alcohol chain) [117]. Fatty acids methyl esters obtained were known under the name of biodiesel.

In 1996, Gokhan Kildiran [118] worked on the transesterification of the in-situ soybean oil with different alcohols as well as the extraction of the residual oil of those alcohols. He showed that the performance of ethanol, propanol and butanol was higher than methanol. In

fact, it was about the influence of the length of the carbon chain of alcohol. The longer the chain was, the more oil was mixed with alcohol. Alcohol was not miscible with oil at ambient temperature. It formed the emulsion with oil and the reaction intermediates (mono and diglycerides). The longer the alcohol chain, the stable the emulsion.

Dasari et al [119] studied the transesterification reaction of the soybean oil with methanol, ethanol and isopropanol at 150 °C without catalyst. In fact, the authors mentioned that methanol and ethanol indicated a similar activity at 150 °C and higher than that of isopropanol. They explained the weak activity of isopropanol by its steric gene, which caused repulsion between the groups. Ethanol possessed a weaker intrinsic speed but a better solubility than methanol. This showed a comparable activity in the conditions used.

However, ethanol possessed the advantage of being derived from a renewable source by sugar fermentation issued from sugar cane or from beet. Thus, biodiesel obtained was 100% bio and was less harmful than methanol. Ethanol improved better the cetane number than methanol [120]. The flowing point of esters formed with a longer alcohol than methanol was lower than that of methyl esters. The benefit was sometimes 10 °C, which allowed the use of saturated oils at the beginning.

Influence of alcohol/oil molar ratio

According to the majority of studies, the alcohol/oil molar ratio as well as temperature were the two most important factors that governed the vegetable oils transesterification reaction speed [121]. The stoichiometry of the reaction implied the use of three moles of alcohol for one mole of triglyceride in order to obtain a mole of glycerol and 3 moles of fatty acid esters. Thus, the transesterification reaction was a reaction of balance in which a great excess of alcohol was requested in order to favour the reaction in the sense of the formation of esters.

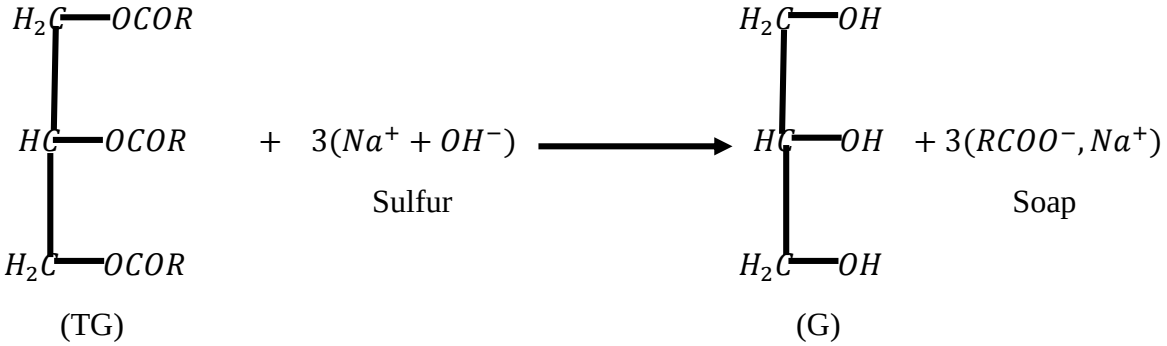
In 1984, Freedman et al [88] tried the optimisation of the conditions of the vegetable oils transesterification reaction. In 1987, Schwab et al [122] showed that at least a molar ratio of 6:1 must be used to maximise the performance in basic catalysis. They also mentioned that a molar ratio higher than 6:1 rendered the glycerol decantation and its esters separation more difficult.

Eucinar et al [123] studied the transesterification of cynara oil with ethanol for a ratio between 3:1 and 15:1. The best results were obtained for a molar ratio 9:1 and 12:1. For a molar ratio of 15:1, glycerol separation became difficult. They further mentioned that when glycerol stayed in the reactional environment, it contributed to the displacement of the

equilibrium towards the formation of triglycerides by lowering the esters performance. However, in acid catalysis, a higher ratio was recommended in order to increase the performance of the reaction [88]. For example, Zheng et al [124] studied the french fries oil transesterification with H_2SO_4. They found that at 70 °C, the oil/methanol/avid ratio of 1:245:3.8 was necessary.

Influence of water and free acids.

The presence of water and free acids constitute a parameter to take into consideration during the transesterification reaction. In the presence of the sulfur or the potash, a parasite reaction will happen with the transesterification: it's the saponification according to the following reaction:



(TG) (G)

This reaction was favoured in the case of acid oil or in the presence of water. In fact, in the presence of water, triglycerides were easily hydrolysed, producing glycerol and free acids. With sulfur and potash, soaps were obtained from these acids.

Wright et al. [125] observed that in the presence of water, there was a decrease of alkalinity due to the formation of soaps and a significant diminution of the performance reaction. They showed that a quantity of water as low as 0.3% could drop the conversion rate at more than 30% and can cause the jellification of the reactional environment. They concluded that all the products must be anhydrous and triglycerides must have had an acidity of less than 1. A supplementary quantity of NaOH must have been added in order to neutralise free acids when the value of the acidity was greater than 1.

Ma et al. [111], demonstrated that the negative effect of water was more pronounced than that of free acids during transesterification reaction. Thus, for good results, the content of water should have been less than 0.06 % and that of acid should have been less than 0.5%.

Graille et al [126] showed that triglycerides ethanolysis was faster than the reaction of secondary esters hydrolysis obtained. They observed that longer reaction times favoured the saponification without improving esters performance in an important manner and that the

high temperature favoured saponification. However, a large excess of ethanol improved esters performance. Their conclusion was that ethanol wa not adapted to the production of esters.

The storage of certain oils leaded to the hydrolysis of most triglycerides. It was for example the case of rice oil where 76% of fatty acids were obtained after 6 months of storage [127]. The homogeneous basic catalysis of such oils was delayed by the formation of soaps. That was the reason why acid catalyst was recommended during the use of acid oil [128].

Other original ideas was put in place to resolve the problem of free acids. For example, the two-steps transesterification can be named: First, esterification of free acids was proceeded for oils that were very acidic, having up to 20%, 30% or more in free acids weights, in order to reduce the acidity of the fatty body to 1%, with an acid catalyst. Secondly, triglycerides transesterification was proceeded with a basic catalyst. The produced biodiesel responded well to the world standards [127, 129, 130].

Influence of temperature and reaction time.

The literature explained that the transesterification reaction time was influenced by temperature. It was completed at ambient temperature but with longer reaction times [131]. In basic catalysis, transesterification reactions were implemented at the neighbouring temperature of ebullition of the alcohol used at atmospheric pressure. But these moderated conditions implied the use of a non-acidic or refined oil. This pre-treatment of oil was not necessary when the reaction happened at high temperature (240 °C) and pressure (9000 kPa) because in those conditions, esterification and transesterification happened at the same time [132].

Freedman and Pryde studied the methanolysis of nuts oil, cotton oil, sunflower oil and soybean oil refined at 60 °C for a methanol/oil molar ratio of 6:1 and 0.5% of NaOMe. They found that the high temperature had mostly an impact at the beginning of the reaction. For both soybean and sunflower oils, 80% of the conversion was obtained after one minute. But after one hour, the ester performance was very close (93-98 % for the four oils) [88]. In acid catalysis, the reaction was slower. That was the reason why a high temperature was often recommended in order to increase the ester performance.

2.8.2.5 The effect of co-solvent

Boocock et al. [133] reported that the methanolysis of soybean oil in the presence of NaOMe at 40 °C evolved 15 times lower than butanolysis at 30 °C. They interpreted it by the formation of two phases in the reactional environment where oil methanolysis happened during oil methanol mixture. The slow oil dissolution in methanol created an initiation period. Mono and diglycerides preferred the methanolic phase and reacted faster, which derived the kinetic of second order. In order to conduct a reaction in one phase, the authors suggested to add one co-solvent such as tetrahydrofuran (THF), 1,4-dioxane or diethyl ether. Thus, by addition of 1.25 THF volume by the methanol volume, only one phase was created in the system, the reaction speed spectacularly accelerated and it happened as fast as butanolysis. In particular, THF was chosen based on its ebullition point which was 67 °C. Thus, at the end of the reaction, methanol and THF, in excess, were co-distilled and recycled [134].

2.8.3 Kinetic of the transesterification reaction

As previously indicated, the vegetable oil transesterification consisted of three successive reactions: TG-DG-MG. The kinetic of the transesterification reaction was complex because each of these three successive reactions were equilibrated. Some kinetic studies were done in order to find the optimal conditions for the reaction [135].

Foon et al. [108] observed that both NaOH and NaOMe possessed high constant kinetics for the methanolysis reaction of palm oil. The reaction happened at 60 °C with a methanol/oil ratio of 10:1 and a catalyst concentration of 0.125 mol/kg of oil.

2.8.4 Thermodynamic of the transesterification reaction

Meneghetti et al. [110] compared methanolysis with ethanolysis in acid catalysis (HCI and H_2SO_4) and basic catalysis (NaOH, KOH, NaOMe and KOMe) have found that the thermodynamic of this reaction was affected by the nature of alcohol and of the catalyst used. Meanwhile the transesterification reaction was a reaction of equilibrium between esters and alcohols. It could have been under kinetic control before the thermodynamic equilibrium was achieved. This favoured the formation of esters. This behaviour was very clear in the case of ethanolysis catalysed by HCI, NaOMe or KOMe where the thermodynamic equilibrium was only achieved after 7 hours of reaction. This situation was not been observed in the case of methanolysis where the thermodynamic equilibrium was performed very quickly (< 1 h).

Figure 2.6 shows the influence of sodium methoxide concentration on the evolution of ethyl esters yields with time. Also, the effect of ricinus communis oil on the yield of methyl esters during basic catalysis is shown in Figure 2.7

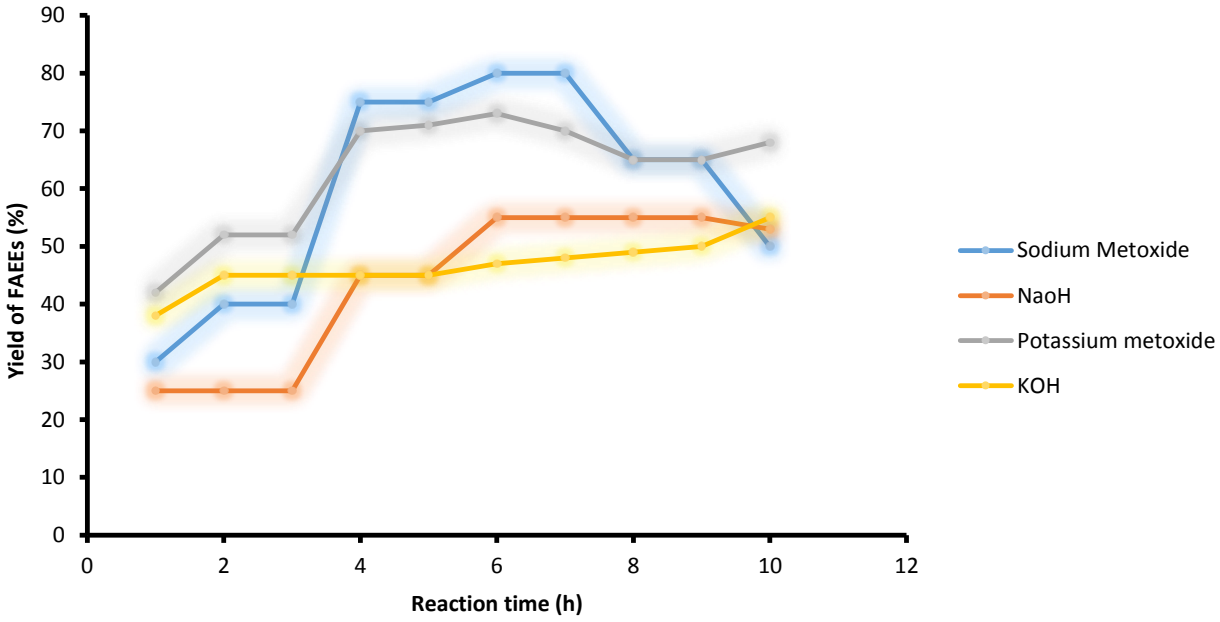


Figure 2. 6: Production of ethyl esters by Ricinus Communis oil ethanolysis by basic catalysis, ethanol/oil/catalyst 60:12:2 at 80 °C [110]

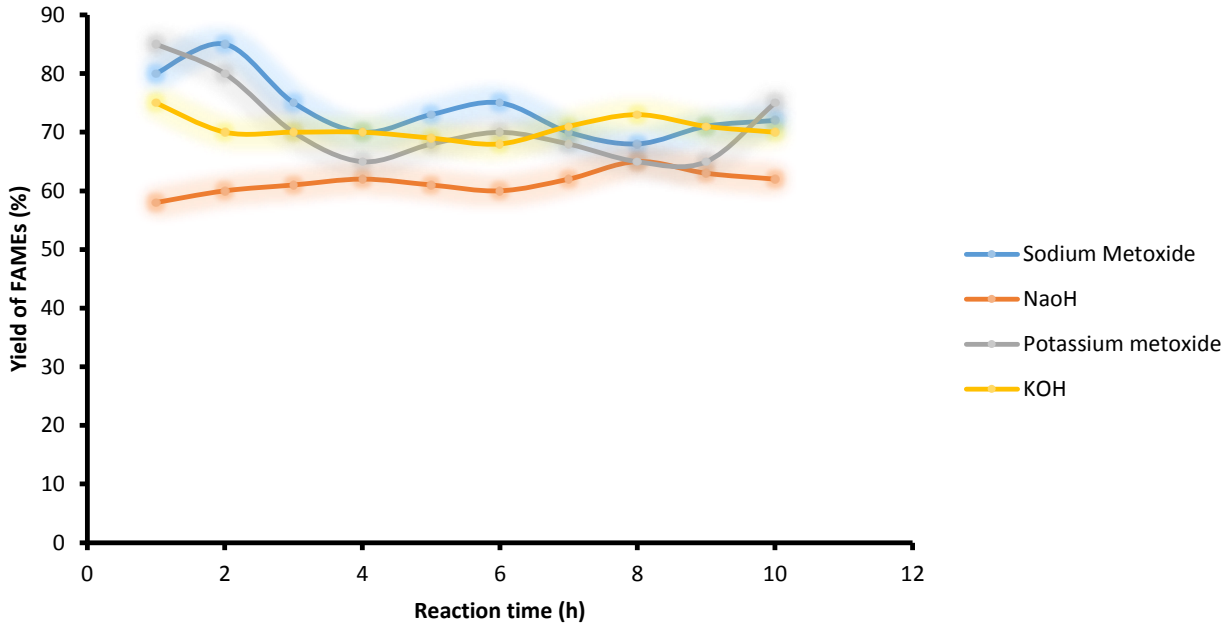


Figure 2. 7: Production of methyl esters by Ricinus Communis oil methanolysis by basic catalysis, methanol/oil/catalyst 60:10:2 at 60 °C [110]

2.8.5 Transesterification by heterogeneous acid and basic catalysis

Several research studied transesterification in heterogeneous environments. The number of publications in heterogeneous catalysis increased in an exponential manner. Since 2009, magazines published on heterogeneous catalysts used in the vegetable oil transesterification [136, 137, 138]. Some researchers used well known catalysts such as CaO, cation-doped alumina, grafted zirconia, hydrotalcites, zeolites ... others developed novel solids for this reaction or even naturally occurring catalysts. This section covers the bibliography in order to look at the most used and original catalysts.

2.8.5.1 Comparison between homogeneous and heterogeneous catalysts

The industrial process of transesterification generally used basic homogeneous catalysts because of their low cost, their high speed of reaction in smooth conditions of temperature and pressure. But these processes were mostly adequate with the oil sources that contained very low water concentrations and free acids such as vegetable oils and refined animals fats. However, refined oils were in high demand in the food industry, which limited their use for biodiesel production [139, 140]. Thus, the need to recycle oils that were already used or to make use of non-edible oils [141]. Non-edible oils are often rich in free acid and in water. This constituted an inconvenience for the homogeneous process because of the formation of soap as well as the washing steps that resulted from it. On the other hand, the enrichment of glycerol obtained (sub product of the reaction) was crucial for the equilibrium of biodiesel. Applications of enrichment depended on its degree of purity. This was not simple because to obtain a very pure glycerol required many phases of purification and a very high cost. If we take the example of the most used alkyl catalysts, there was a need to eliminate soaps and alcoholates present, filter the salts that were formed, evaporate glycerol after the elimination of water. Finally, the excessive alcohol had to be evaporated and often distilled by avoiding that this evaporation would lead to the reaction of ester present with the partially dissolved glycerine. This might had led to the formation of monoglycerides. All those phases increased the cost of transformation. That was the reason many research are orientated towards the heterogeneous catalysis. This catalysis presented many advantages. In fact, it was a process that was cleaner than the homogeneous process because the formation of soaps and of emulsion was avoided. This facilitated the separation and the purification of the products of the reaction. The recycling of the catalyst was also possible and in the future, a continued process should be considered.

2.8.5.2 Heterogeneous acid catalysts

The heterogeneous catalysis created a great interest because of its capacity to perform the transesterification of acid oils without the constraint of soaps formation. In fact, acid catalysis can do both the transesterification of triglycerides and the esterification of free acids, which increases the performance [142]. Recently, several acid solids have been used in the vegetable oils transesterification reaction. Some works performed in the field are listed below:

A. Modified zirconium and tin oxide

Among solid acid catalysts, sulfated zirconia received much attention in the last 20 years because of its high acid properties [143, 144, 145, 146]. The idea of its use came from the difficulty of recovering sulfuric acid in homogeneous catalysis. However, despite its high catalytic properties in many reactions, zirconia appeared to be inactive under mild conditions. Thus, Vicente et al. [143] studied sulfated zirconia and silicon-doped zirconia for transesterification of sunflower oil with methanol under mild conditions (60 °C and 1 atm) but no methyl ester was formed. However, at high temperature, sulfated zirconia suffered from a very strong leaching problem. Sulfate groups were hydrolysed as H_2SO_4or HSO_4^- [147]. As alternatives to sulfated zirconia, Furuta et al. [148] used sulfated tin oxide, sulfated zirconia-alumina, and tungsten zirconia-alumina in the methanolysis of soybean oil and the esterification of caprylic acid by methanol in a fixed bed reactor. Tungsten zirconia-alumina appeared to be a promising catalyst because it maintained its activity (> 90%) up to 100 hours and was capable of carrying out both reactions but at high temperature (> 200 °C for transesterification and > 175 °C for esterification). Two years later, for the same reactions, Furuta et al. [148] observed that an isomerisation of methyl linoleate, which was the major expected ester, took place in the presence of tungsten zirconium. This was not the case when using a doped amorphous zirconia with metals such as alumina or titanium and which appeared to be more efficient.

Also, Tungsten zirconia was studied in transesterification [149]. It was demonstrated that the monoclinic phase of zirconia was inactive but the tetragonal phase alone was not sufficient for a good tungsten zirconia activity. In fact, it was the coexistence of the tetragonal phase of zirconia with WO_3 groups in the amorphous phase that generated this catalyst activity in transesterification [149].

Lopez et al. [149] demonstrated that the optimum calcination temperature was 800 °C where the WO_x density was 6.6 tungsten atom per nm^2. A selective poisoning of active sites

showed that these were Bronsted acids mostly contributing to the catalytic reactivity [149]. The authors also studied titanium, sulfates or tungstates-modified zirconia for two simultaneous reactions: the tricaprylin transesterification and the oleic acids esterification with ethanol.

In transesterification, the doped-titanium zirconia showed a greater activity than tungsten zirconia but in esterification, it was the inverse. The authors explained that the poisoning of the doped-titanium zirconia basic sites in the presence of free acids during esterification decreased the catalyst activity. Tungsten zirconia appeared to be the most suitable: In fact, it was capable of performing both reactions. Also, it was more efficient than sulfated zirconia [150]. Furuta et al. [151] demonstrated that the sulfated tin oxide was more acidic than sulfated zirconia and with a smaller distribution of the force of the sites. Ferreira et al. [152] studied the tin-based complexes and showed that the performance was depending on the chemical structure of the complex and their solubility in the reactional environment.

B. Heteropolyacids

Heteropolyacids are interesting because they have strong Bronsted acid sites. Those acids are homogeneous catalysts in the reactional environment. However, they behave like solids once they are transformed in salts or supported. Their acidity largely depends on the nature of the metal and on the number of protons that they have. Recently, many studies used heteropolyacids in biodiesel production because they were often described as water and free acids tolerant catalysts, therefore there was no limitation on the quality of oil used [153, 154, 155].

Morin et al. [156] used the ammonium salt of phosphotungstic acid (PTA) in the esterification of palmitic acid by methanol. They showed that a partial exchange of protons by the ammonium offered a strong acid more important than when the exchange was total. Also, Narasimharao et al. [157] studied the preparation and the characterisation of the cesium salt of the PTA $C_{s_x}H_{3-x}PW_{12}O_{40}$. They have noticed that the density of the sites decreased with the exchange of cesium iodine but that one $x = 2 - 2.3$ was necessary in order to have a great number of accessible acid sites.

Chai et al. [155] found that $C_{s_{2.5}}H_{0.5}PW_{12}O_{40}$ was a very efficient catalyst for the vegetable oil transesterification by methanol and in the presence of THF as co-solvent. Biodiesel performance was 99% in the relatively mild conditions (60°C, alcohol/oil ratio = 5.3/1 and

cata/oil ratio = 1.85*10-3 during 45 min). According to the authors, the catalyst was more water tolerant than sulfuric acid and sulfated zirconia.

Bokade et al. [158] studied the vegetable oil transesterification at 170 °C in the presence of supported hPa at an increasing humidity. They used various supports and alcohols. The authors found that clay was the most efficient support because it had the lower specific surface. According to them, supports that had a big pore surface and pore size increased the reactive speed of absorption on the catalyst but decreased the reaction surface and reduced the mass transfer between the two reactive which leaded to the diminution of the conversion. However, they found that the length of alcohol chain did not really have an influence on the reaction conversion. Catalyst seemed to tolerate free acids up to 5-6% for oil transesterification. Two years later, other authors [159] studied the influence of many parameters of the reaction. They reported that tungsten based-heteropolyacid was more active than molybdenum based-heteropolyacid and that a 20% content in hPa on the clay was necessary. The catalyst activity was around 90-95% even with a high content of free acids (32%) and water (5%) or even with an alcohol with a long chain (84% of conversion with octanol). The kinetic of this reaction indicated that it was a mechanism of second order.

Alsalme et al. [160] focused on tungsten based-heteropolyacid, both in homogenous and heterogeneous, as catalyst for esterification of hexanoic acid and the transesterification of propionate and ethyl hexanoate. They noticed that their activity was more important than that of usual homogeneous acid catalysts. According to them, this situation was based on their acid force:

$$\mathrm{H_3PW_{12}O_{40}} = \mathrm{C_{s_{2.5}}H_{0.5}PW_{12}O_{40}} > \mathrm{H_4SiW_{12}O_{40}} > 15\%\ \mathrm{H_3PW_{12}O_{40}/Nb_2\ O_5},$$

$$15\%\frac{\mathrm{H_3PW_{12}O_{40}}}{\mathrm{Zr_2O_2}}, 15\%\frac{\mathrm{H_3PW_{12}O_{40}}}{\mathrm{TiO_2}} > \mathrm{H_2SO_4} > HY, H-Beta > \ Amberlyst-15$$

However, the activity per mass unit of catalysts decreased in the other direction:
$\mathrm{H_2SO_4} > \mathrm{H_3PW_{12}O_{40}} - \mathrm{H_4SiW_{12}O_{40}} > Amberlyst - 15 > \mathrm{C_{s_{2.5}}H_{0.5}PW_{12}O_{40}} >$
supported $\mathrm{H_3PW_{12}O_{40}} > HY, H-Beta.$

Pesaresi et al. [161] evaluated a family of silicotungstic acid cesium salt in the C_4 and C_8triglyceride transesterification as well as esterification of C_{16} acids. An exchange level higher than 1.3 C_s per Keggin unit made the heterogeneous catalyst insoluble and active in the two reactions.

2.9 Classification of biofuels

2.9.1 Biofuels gases

2.9.1.1 Hydrogen

Hydrogen is currently produced from fossil hydrocarbons. It is used in the chemical industry as well as in the refinery of petroleum products. Thereby, hydrogen could be used as an energetic vector in the transport sector, via its transformation into electricity in combustible batteries. And hydrogen production from renewable resources is also under study: bioethanol and syngas extracted from biomass gasification could be sources of hydrogen [162].

From bioethanol and syngas

Clean and non-toxic liquid, bioethanol was easily transformed into hydrogen with a good thermal efficiency (>80%). There exist three transformation processes from ethanol to H_2:

- ✓ The most used industrial process was the steam reforming process that consisted of processing steam water with a catalyser;
- ✓ Another method of production of hydrogen was the partial oxidation which consisted of provoking hydrocarbon (ethanol), air and water steam reactions;
- ✓ But it was the catalytic partial oxidation that seemed to lead to the better performance of hydrogen. A purification of the obtained H_2 was generally required. The choice of this process from ethanol depended on the targeted application.

In light of this, syngas contained H_2. It was simply purified where CO was converted into H_2 following the water gas shift:

$$CO + H_2O \rightleftharpoons CO_2 + H_2 \qquad (2.72)$$

Also, this exothermal reaction was favoured by the presence of steam water and low temperatures [163].

B. Via Biochemical processes

Various biochemical processes allow hydrogen production:

- ✓ Anaerobic fermentation of organic compounds by microorganisms (Heterotrophic production of H_2);
- ✓ Photolysis of water in the presence of light energy (autotroph process)
- ✓ The transformation of organic matter in the presence of light energy (Heterotrophic process);
- ✓ By combining all the above processes [163].

2.9.1.2 Biogas

Biogas is one of the products of methanation of animal or vegetable organic matters rich in sugars by anaerobic methanogenic bacteria (absence of oxygen). The wastewater treatment plant sludge, manure, food industry and household waste are the main substrates. Several cultures of bacteria are activated in order to achieve the various steps of methane fermentation; it ends with the production of methane (CH_4) by methanogenic bacteria following the reactions:

$$CH_3COOH \rightarrow CH_4 + CO_2 \quad (2.73)$$

$$CO_2 + 4H_2 \rightarrow CH_4 + 2H_2O \quad (2.74)$$

Table 2.4 shows the typical composition of biogas. The calorific value of biogas is based on its content of methane according and varies from 20 to 30MJ/kg. When biogas contains a high CH_4 volume, it is considered as a potential source of biofuel.

Table 2. 4: Composition of biogas [163]

Gas	Volume %
CH_4	$55 - 75$
CO_2	$25 - 45$
H_2S	$0 - 1.5$
NH_3	0.05

In order to be used as biofuel or to be injected into the natural gas network, methane must be purified with contaminants (for the discharge of gas) and dust, and other particles must be removed, dried and concentrated to more than 95%. Also, in order to be used as fuel, methane must be compressed to a pressure between 200 and 250 bars. All types of vehicles that are designed to run on natural gas or adapted to run on gas and petrol, can use biomethane. Heavy vehicles are usually equipped for their original diesel injection system and an ignition system of gas injection of a small amount of diesel. Depending on the mode of production of electricity used for the purification and methane compression, the use of this gas as a fuel for the road can reduce CO_2 emissions by 95 to 99% compared to fossil fuels. Particles and soot emissions, NOx and carbohydrates other than methane, are also greatly reduced [164].

2.9.2 Liquid biofuels

2.9.2.1 Bioethanol

Bioethanol is obtained by fermenting and distilling sugars contained in sugar crops such as sugar canes, sugar beets and the Jerusalem artichokes. When bioethanol is produced from grains such as from corn or wheat, an intermediate hydrolysis step is required in order to transform the starch contained in the sugar cereal. Also, ethanol is blended with gasoline for an amount of 10% in order to give the commonly known automotive fuel called E10. And this fuel is suitable for all gasoline engines. Beyond a concentration of 20%, adaptations of engines are necessary.

Recently, researchers recommended the production of bioethanol from non-food biomass which is called second generation biofuel [165]. They have the particularity of being from lignocellulosic matter, abundant and cheap, and do not compete with food crops. However, its conversion into fermentable sugars is difficult. The growing interest that is created around this field comes from the fact that the processing of lignocellulosic matter into biofuel has a very positive environmental impact. That is to say that the energy generated less the energy consumed is higher than the ethanol produced using corn.

Conclusion of the Chapter

This review of the literature highlighted the fact that a well exploited *ricinus communis* plant will be of benefit to rural communities. The choice of this plant among others was not made by chance. The next section will focus on the production of biofuel from *ricinus communis* oil. Research conducted on the transesterification reaction of vegetable oil has significantly evolved. The literature showed that the transesterification reaction happened more often in basic homogeneous catalysis (Sulfur, potash, alcoholate). It was very fast but it had some disadvantages at the level of products separation of the reaction and at the level of the recovery of glycerol with a high purity. This reaction also happened in basic or heterogeneous acid catalysis. Therefore, a high performance, a pure glycerine and the possibility of recycling the catalyst were obtained. But both temperature and pressure used in heterogeneous catalyst were very severe. The most used alcohol in both cases was methanol. The following table summarises the advantages and disadvantages of homogeneous and heterogeneous catalysis of the transesterification reaction of vegetable oils. Table 2.5 presents the advantages and disadvantages of homogeneous and heterogeneous processes. It shows

that the heterogeneous process has a big advantage over the homogeneous process in the sense that it is a green process and it produces less waste.

Table 2. 5: Advantages (+) and disadvantages (-) of homogeneous and heterogeneous processes.

Homogeneous	Heterogeneous
Quality of glycerol (-)	Quality of glycerol (+)
Separation and more complex purification phases (-)	Separation and more complex purification phases (+)
Aqueous effluents (-)	Less waste (+)
Consumable (mineral acid) (-)	No consumable (+)
	No catalyst consummation (+): Green process

This review of the literature revealed that the increasing cost of fossil fuels and the prospect of the depletion of this natural and non-renewable energy source, coupled with the phenomenon of climate change attributable to greenhouse gas (GHG) emissions in the atmosphere led sub-Saharan African countries to think about new sources of ecologically clean and economically profitable energy. Among these renewable energy sources, bioenergy in general and biofuels in particular were the ones that attracted the attention of African decision makers because of the particularly high level of consumption of petroleum products. Sub-Saharan Africa, characterized by an energy situation that is dependent on fossil fuels has not been left out of this reality. Much of its oil consumption is imported in such a way that its energy cost weighs heavily on its economic and financial equilibrium.

This situation needed to be corrected!

To contribute to the search for alternative solutions, in the context of this work, biodiesel production from *ricinus communis* was chosen. *Ricinus communis* was a plant that stood out because of its properties that will be detailed later. Hence, the aim of this research which was to develop an advanced mathematical model of biofuel production from the planting of *ricinus communis* to the extraction of its oil. The goal being to better study and analyse the parameters influencing the biofuel production process in order to agree (or disagree) with the experimental results found in the literature.

Chapter 3: Modelling of the *ricinus communis* plant growth

3.1 Introduction

Modelling the *ricinus communis* plant growth is of great importance for the understanding of the optimisation of the plantations destined to biofuel production. This section introduces a mathematical model of the *ricinus communis* plant growth based on a stochastic approach. The experiment was performed on 65 *ricinus communis* trees. For each tree, ten records of their size, their leaves, the beans that the trees can bear as well as the internode lengths were recorded. The operation was performed within a constant interval of 24 months. And, it was assumed that the formation of both the tree stems and branches followed a binomial law of probability.

The biomass production performed by the leaves was the double mechanism resulting to the complexity of the architecture of a plant. The architecture of *ricinus communis* plant is the arrangement of the structure of the branches that compose the *ricinus communis* plant. In other words, the architectural development of the plant is the precise and ordered sequence of events based on the repetition and the integration of various levels of the structure of the plant: phytomers, unit of growth, architectural unit and leaves axis [174].

In this work, the criteria of selection for the optimal growth of the *ricinus communis* plant was defined. The growth itself represents a dynamic model that functions in ring of retroaction and generates well-established quantitative effects. However, the modelling is the mathematical adaptation of the knowledge of the botany, the agronomy and the physiology so that they can provide the necessary parameters of the dynamic equations that describe the growth of the plant. The objective here was to model the architecture of the *ricinus communis* plant by a novel method based on the calculation of probabilities. It is a probabilistic design of the description and the modelling of various processes that will lead to the formation of the architecture of the plant, especially the growth of the stem and leaves, as well as the internodes already formed.

3.2 Model development

The mathematical model development was based on the following: A vegetable axis was formed by a succession of internodes resulting from the functioning of the terminal meristem. Identical axes of the same age were considered and the increase of the number of internodes over time was observed. And it was observed that the said number was a variable.

Thus, the obtained population possessed a specific distribution of the number of internodes per axis and this distribution evolved over time. The observed variation resulted from the fact that the meristems of various axes did not have a continuous and synchronous growth. In fact, there were lags in the elongations of the internodes that came from local conditions that a given meristem suffered. Then, it was considered that the meristem was regularly submitted to a succession of "tests of growth" and that during each of test, the elongation event would happen or not according to a certain probability. Thus, the process of meristem was assimilated to a succession of draws, where each of them was affected by a certain probability to give a positive result. Then, the rules of calculation of probabilities were used in order to model this process. Based on that, a new notion resulted, that of the "dimension" of the leave formed by a meristem: It was the number of tests of growth that the meristem suffered. Therefore, the real size of the leave, which was counted in terms of internodes, was less or equal to its dimension. It was equal to it only if all the tests of growth had given a positive result.

Also, the growth of a population of identical plants axes was described by a simple random process. To each test of growth "i", a probability "b_i" of the elongation of the internode formed was related. It was interesting to observe that this approach was necessarily linked to the time. Thus, if the local conditions were relatively good (climate, nutrition), the test of growth had a relatively long duration. However, if there was a minimum time for the formation and the elongation of an internode, the maximum time was not defined. At each phase of the population growth was associated the corresponding mean and variance of the instant distribution of the number of internodes. According to De Reffye et al. [174], this process could perfectly be described if we knew the mean-variance relation. The distribution of the number of phytomers observed on the stems was a binomial law of parameters N, b, because this number was obtained by a succession of random and independent draws and of constant probability equal to b. When N was the number of tests and b the probability of the elongation, the mean $\bar{X}$ was:

$$\bar{X} = Nb \tag{3.1}$$

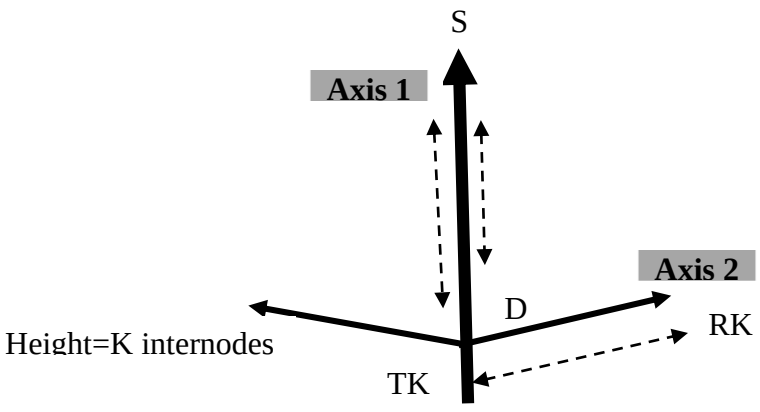
And the variance V:

$$\mathrm{V} = (1 - \mathrm{b})\bar{X} = Nb(1 - b) \tag{3.2}$$

Then the mean-variance relation was represented by a straight line. As this will be seen later, this situation was formally the one that happened on the *ricinus communis*.

Calculation of the treetop

The calculation of the number of tests and of the probabilities of the elongation necessitated a follow up on the population growth. The same result was reached by using the architecture of the treetop through its observation from the highest point of the tree to the base, starting from the point where the internodes became visible. It was assumed that there was a T-treetop resulting from the same growth process, and which was constituted of boughs of order 1 and 2. It was further assumed that the probability of the elongation of axes of order 1 was constant and had the value P, and that the probability of the elongation of the axes of order 2 was constant and had the value b. This meant that the treetop boughs were identical. But when the probabilities were the same, the sequence of formation of internodes was different. Thus, while axis 1 went through N tests of growth between the treetop and the internode that had axis 2, the last went through $W \times N$ tests, and this, when W was the parameter that defined what was called the sequence ratio. W was less, equal or greater than 1. Finally, the T-treetops was described in a dynamic manner by four parameters: three constants P, b and W as well as one variable N. The three parameters P, B, W were unknown at the beginning. Therefore, we need thee relations in order to calculate them. To do that, on the various plants, all boughs that were connected at the same distance K (internodes) from the treetop were observed: their heights were measured, and the mean and the variance of the obtained values were calculated. TK was the branch point of the boughs (Figure 3.1). Therefore, because of the random process of the main stem, the level TK chosen did not correspond to only one dimension but in theory, it corresponded to an infinity of possible values.



D = Dimension = TK-RK and/or D = Dimension TK-S

Figure 3. 1: Diagram of the branch point of an axis of order 2 connected to an axis of order 1

In this case, the dimension varied according to the negative binomial law (K, P). The mean value of the number of units of the dimension corresponding to the number K of internodes really observed will be:

$$\mathrm{N} = \frac{\mathrm{K}}{\mathrm{P}} \tag{3.3}$$

With a variance:

$$\mathrm{V} = \frac{K(1-P)}{P^2} \tag{3.4}$$

Knowing the mean dimension N corresponding to the number K of levels that separated the treetop from the branch point, the distribution of height of the axes 2 of level TK was determined, because they had necessarily suffered the same number of units of dimension.

We have:

$$\overline{\mathrm{X}} = \frac{WKb}{P} \tag{3.5}$$

$$\mathrm{V} = \frac{WKb(1-b)}{P} + \frac{W^2Kb^2(1-P)}{P^2} \tag{3.6}$$

This variance V is called total variance of the population of the axes of order 2 branched to K levels of the treetop. In practice, it will be estimated by the following classical value:

$$\mathrm{V} = \frac{\sum(X_i-\bar{X})^2}{T-1} \tag{3.7}$$

In order to calculate the three unknown values, a relation was missing. To find it, we will calculate the mean internal variance of the various couples each constituted of a bough of the same order but L branched internodes at a lower level, those two boughs being taken on the same plant.

Let X_1 and X_2 be the two boughs. For a plant, the internal variance can be of the following form:

$$\mathrm{V_i} = \frac{(\mathrm{X_1}-\overline{\mathrm{X}})^2+(\mathrm{X_2}-\overline{\mathrm{X}})^2}{2-1} \tag{3.8}$$

Or of a more simple form:

$$\mathrm{V_i} = \frac{(\mathrm{X_1}-\mathrm{X_2})^2}{2(2-1)} \tag{3.9}$$

The mean internal variance for T plants was as follows:

$$\mathrm{V_i} = \frac{\sum(\mathrm{X_1}-\mathrm{X_2})^2}{2\mathrm{T}} \tag{3.10}$$

This variance was weaker than the one given by two boughs taken in the same conditions from two different stems. In fact, from the treetop to the internode K, the succession of tests of growth is the same for the two boughs. We can demonstrate that:

$$V_i = \frac{(K+L/2)}{P} Wb(1-b) + \frac{b^2W^2L(1-P)}{2P^2} + \frac{b^2W^2L^2}{2P^2} \quad (3.11)$$

A particular case is that where L=0 (whorl of branches). Then we have:

$$V_i = \frac{WKb(1-b)}{P} \quad (3.12)$$

This relation presented the particularity of possessing whorls. It was observed that in the second term of the total variance, V disappeared. Finally, $\overline{X}, V, V_i$ were expressed according to K, L, P, b and W and L fixed. Therefore, the three unknown parameters P, b and W were calculated based on $\overline{X}, V, V_i$. In order to facilitate the calculations, the following was considered:

$$U = \frac{Wb}{P} = \frac{\overline{X}}{K}$$

$$1 - b = \frac{1}{\overline{X}}\left[V_i - \frac{L}{2}\left(\frac{V}{K} + LU^2\right)\right]$$

$$1 - P = \frac{1}{\overline{X}^2}\left[K\left(V - V_i + \frac{L^2U^2}{2}\right) + \frac{VL}{2}\right]$$

$$W = \frac{\overline{X}P}{Kb} \quad (3.13)$$

Therefore, we can calculate the parameters of the growth of the treetops from the observation of the variability if this growth follows the proposed model.

3.3 Model validation

3.3.1 Parameters of growth of the stems and the branches of the Ricinus Communis plant

A. Mean-variance liaison

According to the 10 parameters records taken during the experiment, while counting the couple (0,0) before the growth stops, a significant linear regression ($R^2 = 0.968$) was observed. Between the mean and the variance, the liaison $V = 0.0543\ \bar{X}$ was found.

Therefore, considering the equation (3.13), by replacing b by P, then:

$$\frac{V}{\overline{X}} = 1 - P \quad (3.14)$$

Therefore, the main stem probability of elongation had the following value:

$$P = 1 - \frac{V}{\bar{X}} := 1 - 0.0543$$

$$P = 0.9457$$

In fact, Table 3.1 shows the numerical relationship that exists between the mean and variance distribution of the plant stem until the time it dies. On the other hand, Table 3.2 displays the same relationship but during the cycle of the plant growth.

Table 3. 1: Mean-variance relationship of the distribution of the main stem number of internodes until when the plant dies.

$\bar{X}$	V
0	0
3.2	0.2
4.2	0.2
5	0
5.4	0.3
5.6	0.3
5.6	0.3
5.6	0.3
5.6	0.3
5.6	0.3
5.6	0.3
5.6	0.3
5.6	0.3
5.6	0.3
5.6	0.3

Table 3. 2: Mean-variance of the distribution of the number of internodes of the main stem during the growth period

$\bar{X}$	V
0	0
3.2	0.2
4.2	0.2
5.4	0.3

Figure 3.2 shows the variance and mean in correlation over the time series for the plant stem during the growth period.

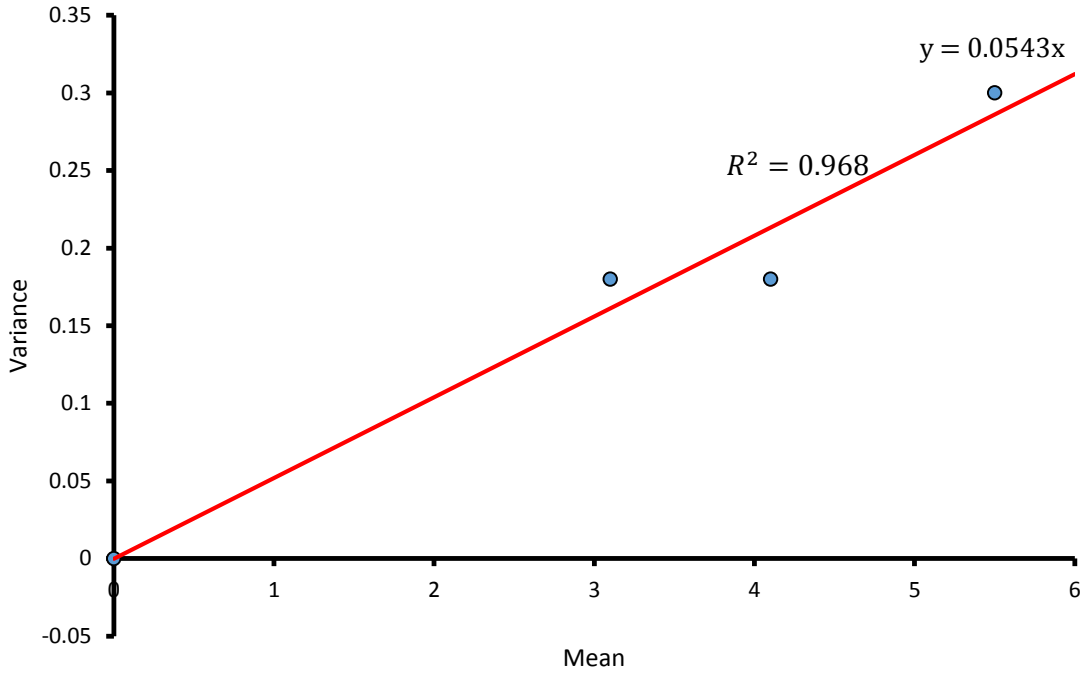


Figure 3. 2: Mean-variance relationship found on the height values (number of internodes) of the *ricinus communis* plant stem during the growth period.

B. Distribution of the number of internodes per stem at the end of the growth

Table 3.3 indicates the distribution obtained in relation with the number of internodes per stem after the growth.

Table 3. 3: Mean-variance liaison found on the height values (number of internodes) of the Ricinus Communis plant stems during the growth period.

Nbr. of internodes	5	6
Nbr. of individuals	15	20

This distribution presents a mean value $\overline{X} = 5.5714$ and a variance V=0.2521.

We can consider a binomial law of parameters P and N:

$$P = 1 - \left({}^{0.2521}/_{5.5714}\right) = 0.954$$

And:

$$N = \frac{5.5714}{0.954} = 5.84$$

Considering six tests of growth, the dimension of the stems was also six. And the validity of the adjustment to the binomial law (6; 0.954). According to this law, the probability to obtain a value x of the number of internodes, in other words, the probability to obtain a succession of positive x after N tests of growth was:

$$P_{(x)} = C_N^x P^x (1-P)^{N-X} \quad (3.15)$$

Multiplying this probability by 35 (which was the size of the sample of *ricinus communis* plants), the theoretical number of *ricinus communis* plants presenting x internodes was obtained. Table 3.4 indicates the comparison of this value with the value really observed. And Figure 3.6 presents the corresponding histograms.

C. Update of the number of branches of the Ricinus Communis tree to a binomial law

Table 3. 4: Updated number of branches

Class	Observed number	Calculated theoretical number
4	0	2.7
5	15	11.6
6	20	20.3

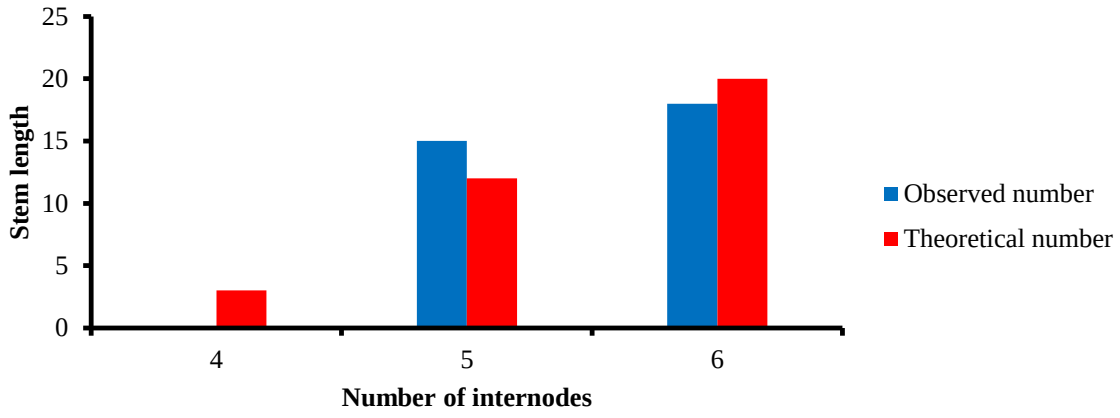


Figure 3. 3: Histogram of observed number and the theoretical number of internodes for 35 plants

The growth of the main axis of the *ricinus communis* tree was performed with the probability of elongation constant and equal to 0.954 over time. It was considered that all was happening as if the terminal meristem process was submitted to a succession of tops of regular clocks, number of six in total. A random draw was done for each of those tops, which decided if yes or no an internode will be formed, and this, with a probability of success of 0.954 at each time.

3.3.2 *Ricinus communis* treetop after the end of the growth

On the treetop of the *ricinus communis* tree, the number of internodes of branches at each level from the treetop to the base was measured. And in this case, L=0 (whorl branch).Therefore, solving the system constituted of equations (3.23), (3.24) and (3.25) enabled the calculation of P, b and W of the *ricinus communis* plant. According to the analysis of the treetops, the following were obtained:

For the treetop of the Axis 2 on the axis 1:

- ✓ The probability of elongation of the stem P= 0.87
- ✓ The probability of elongation of the branches b= 0.95
- ✓ The stem/branches ratio W=0.84

Therefore, it was observed that the probability of elongation of branches was stronger than the probability of stems. However, the growth process of branches was not too different than the growth process of the stem. Thus, the validity of the comparison of values of P calculated by the following three methods were:

- ✓ The mean-variance liaison method: P=0.9457
- ✓ The distribution of the height of the stems at the end of the growth: P=0.954
- ✓ And the calculation of the treetops: P=0.87

3.4 Discussion

The growth-development interactions were taken into consideration and organogenesis were dependent on photosynthesis. The case of the development of the *ricinus communis* tree was completely different from the growth of other trees because it had a simple process: There was no branch mortality, nor an amortisation phenomenon. Therefore, the growth of the plant was not stochastic because its probability of ramification was elevated (0.95), which meant that branches were formed at each internode. However, it was observed that in other conditions of cultivation (climate, organic matter, density, atmospheric pressure, etc), the growth of the *ricinus communis* plant was divergent by producing both the quantity and the quality of the harvest, including in its stochastic component. Also, the optimal control of the model enabled the optimisation of itineraries of the operations such as irrigation and process. Furthermore, the use of a mathematical method in the modelling of the architecture of a *ricinus communis* tree gave coherent results and revealed itself to be fruitful. The probability of the set of branches process was constant, with a value of 0.95. For the main axis, the probability of the process was reduced to 0.87 and the ratio process had the value of 0.84.

Finally, results indicated that in all cases, the probability was calculated with a good precision and with concordant results using three different methods including: the mean-variance relationship, the updated number of internodes of the stem and the treetop analysis. Therefore, the probability of growth as well as the probability of the formation of branches were quasi deterministic. In other words, the *ricinus communis* plant was not stochastic because to each internode, there were always branches associated to it.

Chapter 4: Modelling of the *ricinus communis* oil extraction process.

4.1 Introduction

In this chapter, a probabilistic model was applied to the oil extraction process as well as to various *ricinus communis* oil storage containers. Because of that, a qualitative approach that consisted of finding a simple probabilistic model was used. Specifically, a model with less parameters was sought after. It described the general structure of oil storage containers with respective quantities of *ricinus communis* oil within the same platform. In this perspective, the platform was defined as the set of all oil containers having small and big quantities of *ricinus communis* oil. Therefore, the Pareto-Levy distribution was used in order to model quantities of the *ricinus communis* oil distributed in the storage containers of three simulated platforms (oil depots).

The first two sections of this chapter cover the description of the extraction of lipids and the introduction to the model process of probability distributions. This enabled the construction of a probabilistic model of the *ricinus communis* oil extraction process. And its related mathematical tools were described in the second section. Then, the Bolthausen-Sznitman fragmentation model is introduced in section 3 in order to describe the progressive fragmentation of *ricinus communis* oil extracted from the plants. Therefore, the extracted oil constituted the ideal reserves for the production of biofuel.

Section 4 of this chapter covers the optimisation of the extraction process as well as the construction of its model using the Bolthausen-Sznitman model: Then, the optimisation of the process was performed in an empirical way. Therefore, the process was tested by performing different experiments under various operating conditions. Results were expressed in a series of correlations from which the performance was estimated and extrapolated for new operating conditions.

It is in this context that the model was formulated, solved and the predicted values were compared with experimental values. Then, the model was readjusted until a good correlation with the experiment was obtained. Then, the model was validated based on new experiments and it was used for optimization and process control. Experimental data were needed in order to confirm the model but the amount of data required in comparison with the empirical approach were considerably reduced.

4.2 Probability distribution of *ricinus communis* oil containers

This section describes the quantitative determination of oil content in small and big containers which translates the technical proprieties of a selected platform.

4.2.1 *Ricinus communis* lipid extraction techniques and modelling

Ricinus communis seeds could be harvested through centrifugation, flocculation, sedimentation, gravity, filtration, flotation or other techniques of electrophorese. In order to produce biofuel from *ricinus communis* lipids, oil extraction was performed. Various *ricinus communis* lipid extraction techniques were considered including physico-chemical and biochemical methods. The direct transesterification of *ricinus communis* oil into biofuel was also one of the major methods considered. Even though at laboratory-scale, lyophilisation remains the most popular method, drying through atomisation, the drying through oven, and sometimes vacuum evaporation were used in order to dry *ricinus communis* seeds. However, the process of drying seeds before the extraction of lipids necessitated 2.5 times more energy than the normal drying process. This rendered the process non-productive in energy. And even if the chloroform-methanol mixture was largely used with up to 83% of extraction productivity, less polar solvents such as n-hexane were often preferred due to their lower level of toxicity and their affinity for non-lipid contaminants.

Thereafter, the modelling of oil extraction was performed using the response surface methodology (RSM) with the Doehlert experimental plan. To achieve that, the following parameters were considered: Temperature, time and solvent.

Therefore, a time-parameter model was developed because of its significant influence on the oil extraction process. The remaining two parameters were those that influenced the oil extraction process. They were:

- ✓ The temperature of extraction (X_1)
- ✓ The time of extraction (X_2)
- ✓ The solvent rate (X_3)

Following the identification of parameters, the Doehlert four-parameter experimental process was used for the execution of manipulations. The Doehlert's plan was chosen because of the small number of experiments, the uniform distribution of experiments as well as its ability to evolve after results were obtained [175].

And, after the choice of the experimental plan, the equation of the model was defined and the parameters of this model were provided. Because the model used in the RSM is generally a quadratic equation, the second degree model was written as follows:

$$y = \beta_0 + \sum_{j=1}^{k} \beta_j X_j + \sum_{j=1}^{k} \beta_{jj} X_j^2 + \sum \sum_{i<j} \beta_{ij} X_j X_i + \varepsilon \quad (4.1)$$

where β_0: Constant; ε: error; β_j, β_{jj} and β_{ij}: coefficients of the model.This mathematical model considered the coded parameters that were as follows: $Y(X_1, X_2, X_3) =$ Mathematical model for the *ricinus communis* oil extraction, with X_1= temperature (℃), $X_2 =$ time of production/extraction (min), $X_3 =$ the catalyst amount (ml). The assumption was that the model parameters were linear(X_1, X_2, X_3), quadratics (X_1^2, X_2^2, X_3^2) and with interaction (X_1X_2, X_1X_3, X_2X_3).

Application of the Surface Response Method for oil extraction

The choice of the experimental design for the optimization of the operating conditions took into account the parameters to be optimized. They were three continuous variables dependent on each other: temperature, time and catalyst rate. The objective here was to optimize the three parameters and model their influence on the synthesis efficiency by carrying out a minimum of experiments. Therefore, the first criterion for choosing the plan was the number of experiments that it required, considering that the Doehlert's plan required few trials, with $N = k^2 + k + 1$, where k was the number of parameters.

Let $k = 3, N = 13$ experiments.

For the optimisation of the three parameters, a Doehlert matrix was constructed. Then, after describing the procedure applied to N=13 experiments, results were presented in detail and comments were attached. The main characteristic of the Doehlert's plan was the need to have a uniform distribution of the experimental points [176]. For a three parameters' plan, the experimental matrix consisted of 13 experiments with one parameter of 5 levels, one with 7 levels and one with 3 levels. As a result of that, the limits of the domain as well as the number of levels of each parameter were found to be as follows:

Time was the most important parameter because, ideally, the continuous reactor should make it possible to work at a minimum time. Because of that, time was found to be the parameter with 7 levels. Also, the lower limit of the reactor was set to 30 minutes and for the upper limit, it was observed that the preliminary tests performed within 240 minutes, yields did not exceed 20% for a temperature of 130 ℃ and a catalyst rate of 1%. Then the upper limit was set to 240 minutes to have a time variation range of 250 minutes.

Secondly, *temperature* was the least precise parameter to be set and therefore was the parameter with 3 levels. It was controlled by the thermostat and its value was the temperature set point of the thermostat. The upper limit was determined by the limit of the equipment, i.e. 158 ° C. The lower limit was set in such a way that the viscosity of the reaction medium was not increased.

Thirdly, the *catalyst* level was the parameter with 5 levels. Its minimum was chosen lower than that of the batch reactor (1%) and was set at 0.5%. Its maximum was set to maintain a catalytic amount of 5.5%. In addition, as the influence of each of the three parameters was studied, their variations were limited within two values: The lower bound was set to -1 and the higher bound was set to +1.

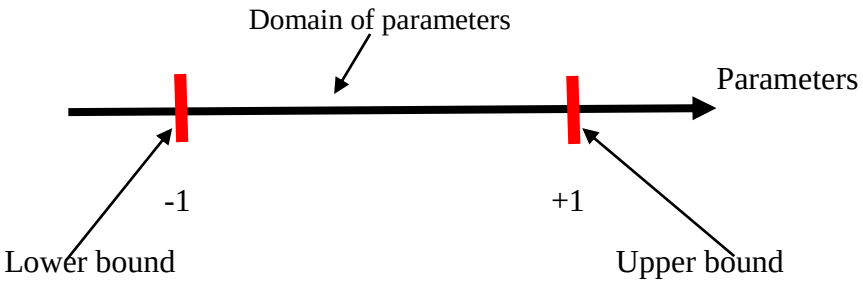


Figure 4. 1: Domain of parameters with lower and upper bounds

4.2.2 Definition of the experimental model

With the pre-analysis, it was possible to identify 3 parameters that had a statistically significant influence on the observed responses As sown in Table 4.1, the model designed was a second degree polynomial; therefore, each parameter was endowed with at least 3 modalities. The table further indicates the boundaries used for each parameter.

Table 4. 1: Boundaries of parameters used for response surface.

Parameters	Minimum boundary	Maximum boundary
Time (1)	30 minutes	240 minutes
Temperature (2)	40 °C	160 °C
Catalyst rate (3)	1%	5.50 %

. In this context, it was easy to calculate the number of coefficients to be identified in each of the models from the formula: $p = \frac{(k+1)(k+2)}{2}$, where k was the number of parameters. In this case, the formula indicated that the number of unknown values p to be determined in a polynomial model of second order was calculated. In fact, when k was equal to 3 parameters the application of this formula produced a total of 10 coefficients p to be determined.

The objective of the algorithm was the calculation of N experiments out of the number of points that provided the best results in terms of quality. Results indicated that 13 was this number N of experiments that constituted the experimental Matrix. For material reasons, the maximum size of the Matrix X was limited to 25 experiments. Each experimental point enabled to obtain a value of the response which was modelled by the polynomial function where the 10 coefficients p were the unknowns. At the end of the experimental plan, a system of n equations in p unknowns was produced and was written as follows:

$$y = X\,a + e \tag{4.2}$$

where: y is the vector of responses; X is the matrix of calculation, or Matrix of the model that depends on the experimental points chosen to execute the plan. "a" is the vector coefficients vector and e is the error.

Furthermore, it was found that the number of equations (n) was less than the number of unknowns (p). In other words, there were 10 equations in p unknowns. Therefore, in order to solve the system, the least squares regression method was used and the following estimation of regression coefficients was obtained:

$$\hat{a} = (X'X)^{-1}X'y \tag{4.3}$$

Where: X' is the transposed matrix of X. It was further noted that two matrices constantly intervened in the theory or plan of experience: X' X, the information matrix and $(X'X)^{-1}$ the dispersion matrix.

4.3 Modelling of *ricinus communis* oil containers

Prior to the design of a comprehensive mathematical model for the production of biofuel, it was essential to make a good choice of the probability distribution of the model of the quantity of oil that was stored. This section focuses on the proposed probabilistic model for the *ricinus communis* oil storages destined to the biofuel production. In order to describe the random repartition of various quantities of oil in the containers, the concept of satellite structure was used. In the field of petroleum engineering, the following approach describes the spatial repartition of oil deposits in an oil field: Next to a very big oil deposit, there usually are smaller oil deposits. Similarly, there can be a container with a very big quantity of *ricinus communis* oil as well as a number of smaller containers around the big one. This type of structure was compatible with the stochastic invariance property that characterises the Pareto-Levy distribution [177] and which was a suitable model to characterise the quantity of *ricinus communis* oil in a platform. The main limit of this approach was that the Pareto-Levy

distributions could only describe containers with a quantity of oil below a certain level ε and that also needed to be estimated. Therefore, the level ε was interpreted as the quantity of *ricinus communis* oil extracted and stored in containers for the production of biofuel. The assumption was that there was a big platform containing a certain number of *ricinus communis* oil containers. And the platform was sub divisible into a big number of small containers $d_1 \ldots . d_M$. Also, the action of a certain chemical factor d_p on the platform of castor oil was to multiply the quantity d_p by the coefficient $\exp \varepsilon_i^p$ where it was further assumed that the variables $(\varepsilon_i^p; 1 \leq i \leq N, 1 \leq p \leq M)$ were independent and of the same law $\left(\varepsilon_i^p = -v\right) = P\left(\varepsilon_i^p = v\right) = 1/2$, where v corresponds to the amplitude of the chemical event. The cumulated action of N chemical factors was the multiplication of d_p by the coefficient $exp\left(\sum_{i=1}^{N} \varepsilon_i^p\right)$ .By only considering chemical factors that helped for the distinction and the determination of the subdivision of M castor oil storage containers, it indicated that N was the order of $\log M$.

Furthermore, considering $M = 2^N$, the *ricinus communis* oil quantity d_p admitted the expression $\langle t_v, N\rangle \, exp\left(v\sqrt{N}G(d_p)\right)$, where $G(d_p)$ and $\langle t_v, N\rangle$ are a standard centered Gaussian random variable and a common quantity of *ricinus communis* oil respectively. Then, it was indicated that a critical amplitude v_c existed such that, for $v > v_c = \sqrt{2\log(2)}$, and for a good selection of $\langle t_v, N\rangle$, the quantities of castor oil converged, following the law of decreasing order, towards the distribution $PD(\alpha, 0)$. Therefore, $PD(\alpha, 0)$ distributions were used in order to model the quantities of castor oil. And, in order to consider that v was subjected to change during the extraction of *ricinus communis* oil (with $v > v_c$), the Bolthausen and Sznitman models were used [178]. Their models describe the coalescence operations on stable particles. And the dual process obtained by reversing the time is called the Bolthausen-Sznitman fragmentation and this dual process has the remarkable propriety of being Markovian. The evolution of castor oil quantities submitted to successive operations of fragmentation and coalescence was qualitatively described by the following model: The dual process was such that at all time, the distribution of *ricinus communis* oil was $PD(\alpha, 0)$ where only α changed over time. This remarkable propriety fitted very well in the selection of Pareto-Levy laws for modelling the quantities of castor oil stored. Thus, in order to design a model for biofuel production from *ricinus communis* oil, each phase that intervened in biofuel production was described. Therefore, biofuel production from *ricinus communis* was defined as the aggregate production of existing *ricinus communis oil* at a given time. In a

platform containing n containers of castor oil, B_a *and* C_a were the quantities of *ricinus communis* oil and the date of the production of biofuel a respectively. Thus, the production of biofuel at the time t had the following expression:

$$\mathrm{Prod(t)} = \sum_{\mathrm{a=1}}^{\mathrm{n}} \mathrm{prod_a(t - C_a)}, \tag{4.4}$$

where $prod_a(\tau)$ designated the castor oil production of container a after a certain time τ of production.

4.3.1 The Pareto-Levy law and the distribution of income

It was possible to define the Pareto-Levy distribution by the following propriety of change of scale invariance: A variable X follows a Pareto-Levy law μ if and only if there exists $\varepsilon > 0$ such that the support of μ is $[\varepsilon, \infty]$. $\forall\, \eta_1, \eta_2 \in [\varepsilon, \infty]$, the $\eta_1^{-1}X$ law under $\mathbb{P}(.\,|X > \eta_1)$ is the same than the $\eta_2^{-1}X$ law under $\mathbb{P}(.\,|X > \eta_2)$. This definition immediately implied that $\log(\varepsilon^{-1}X)$ satisfies the propriety of absence of memory that characterises exponentials; therefore, a variable X follows a Pareto-Levy law Iff $X = \varepsilon \exp(Y)$, where Y is a random variable of the exponential law of parameter α. The density distribution of Pareto-Levy has the following expression:

$$\mathrm{f(x)} = \alpha \frac{\varepsilon^{\alpha}}{\mathrm{x}^{\alpha+1}} 1_{\mathrm{x} \geq \varepsilon} \tag{4.5}$$

where α and ε are the index (or the exponent) and the level of the Pareto-Levy distribution respectively. Let's note that this distribution doesn't allow a moment of order 1 when $\alpha \leq 1$ (heavy-tailed distribution). The definition of the Pareto-Levy distribution suggests that there exist a threshold ε under which the extracted castor oil was no more existing in a storage.

With regards to the distribution of the extracted *ricinus communis* oil, the following hypothesis of distribution was validated: $G_c$was a container with a quantity of *ricinus communis* oil higher than ε follows a Pareto-Levy law on the $[\varepsilon, \infty[$ interval.

The log-log graph

The choice of the model was graphically verified using the log-log graph: Let $\{z_1, \ldots, z_p\}$ be the series of real data strictly positive [3]. Let σ be the permutation of the set $\{1, \ldots, p\}$ which, at $\{z_1, \ldots, z_p\}$, associates the series $\{z_{\sigma(1)}, \ldots, z_{\sigma(p)}\}$, where $z_{\sigma(1)} \geq \cdots \geq z_{\sigma(p)}$. The log-log graph of series $\{z_1, \ldots, z_p\}$ was the graph of double logarithmic scale with points $(i, z_{\sigma(i)})_{1 \leq i \leq p}$. It has been demonstrated that the log-log graph of a sample that followed a Pareto-Levy law of exponent α presented a linear regression of slope $-\frac{1}{\alpha}$ [3].

4.3.2 Model of random fragmentation

The model of stable subordinators of index α was used to model the extracted oil. Subordinators are Levy processes of values comprise in the interval $[0, +\infty[$.This type of modelling approach has many advantages:

Firstly, it is coherent with the Pareto-Levy sample because the jumps of a stable subordinator with a size higher than the threshold ε are random variables of Pareto-Levy distribution; the G_c hypothesis is therefore verified. Also, we will see that this modelling approach is compatible with the idea that the current repartition of castor oil is the result of a random fragmentation of the initial quantity of castor beans.

The G_a hypothesis suggested that the fragmentation process followed a Markovian behaviour, and the $G_b$hypothesis precised the type of fragmentation mechanism of the extracted castor oil. And, the Bolthausen-Sznitman coalescent reversed back in time described the progressive fragmentation of stable partitions.

4.3.3 Review of the concept of stable partitions

4.3.3.1 Definitions

This section focuses on random combinatorics models. The same probability space (Ω, F, P) was considered. Therefore, $[n]$ was defined as the set of natural numbers $\{1, 2, \dots n\}$. Thereby, the partition of the set $[n]$ was any collection of mutually disjoint non-empty sets $\{A_1, \dots, A_k\}$ whose union was $[n]$ [179]. In addition, $P_{[n]}$ was the set of all partitions of $[n]$.

Furthermore, a composition of n was a sequence of natural numbers $(n_1, \dots, n_k)$whose sum is n. Thereby, because the partition of $[n]$ was not ordered, this raised the problem of blocs enumeration. One of the solutions was to index them by order of appearance, particularly in an ascending order of their smallest element. The bloc with element 1 was the first in the list; the second was the bloc with element 2, etc. And, $\left(\overline{N}_{n,1}, \dots, \overline{N}_{n,k_n}\right)$ was the sequence of bloc sizes ranged this manner. Another way of enumerating the blocs was to range them in descending order of bloc sizes. And $\left(N^{\downarrow}_{n,1}, \dots, N^{\downarrow}_{n,k_n}\right)$ was the composition of n resulting from the partition when blocs were ordered in descending order of their size. Generally, for a partition Π, it was denoted $i \sim \pi j$ when $i \; and \; j$ were in the same bloc of Π.

Moreover, a random partition π_n of the set $[n]$ was a random variable with a value in the set $P_{[n]}$. The symbol P_∞ was defined as the tribe $TR = \sigma(\{\Pi_\infty \in P_\infty : i \sim_{\Pi_\infty} j\}, i, j \in N)$. And a

random partition of N was a TR-measurable random variable. Because the numbering had to remain arbitrary, partitions were exchangeable: Therefore, a random partition of Π_n of $[n]$ was called exchangeable when its distribution was invariant under the action of the group of permutations of $[n]$. Also, a random partition Π_∞ of N was exchangeable if its distribution was invariant under the action of the group of permutations with finite support. For a random partition of $[n]$, this definition meant that the distribution did not depend on natural numbers that were in each bloc, but it depended on the random composition of n deduced from the random partition:

$$P(\Pi_n = \{A_1, \dots, A_k\}) = p_n(|A_1|, \dots, |A_k|) \quad (4.6)$$

Where p_n was a symmetric function of compositions of n and is called the probability function of the exchangeable partition.

With regards to *ricinus communis* oil, these random partitions represented the various quantities of *ricinus communis* oil that were extracted and stored in a platform. It was assumed that all various quantities of extracted oil over time had the same propriety no matter what the size of the set to be partitioned was. This propriety is called consistence: An exchangeable sequence of random partitions (Π_n) is consistent in distribution if

$$\forall m < n, \quad \Pi_{m,n} \overset{d}{=} \Pi_m$$

Where $\Pi_{m,n}$ is the restriction at $[n]$ of Π_n: if $\Pi_n \coloneqq \{A_1, \dots, A_K\}$ then

$$\Pi_{m,n} = \{A_1 \cap [m], \dots, A_K \cap [m]\}$$

And for a sequence (Π_n) that is consistent, then:

$$\mathrm{p_n}(\mathrm{n_1}, \dots, \mathrm{n_k}) = \sum_{\mathrm{j}=1}^{\mathrm{k}} \mathrm{p_{n+1}}(\dots, \mathrm{n_j} + 1, \dots) + \mathrm{p_{n+1}}(\mathrm{n_1}, \dots, \mathrm{n_k}, 1) \quad (4.7)$$

Where $(n_1, \dots, n_k)$ is a composition of n *and* p as defined in (1.1). In fact, if the restricted random partition at $[n]$ contains blocs of sizes $(n_1, \dots, n_k)$, then the restriction at $[n+1]$ contains blocs of sizes to be chosen among the compositions $(n_1, \dots, n_k, 1), \dots, (n_1, \dots, n_k + 1), (n_1, \dots, n_k, 1)$. Also, if Π_n is the restriction of a partition Π_∞ of N at the n first natural numbers, then π_n is exchangeable for all n, and the sequence (Π_n) is consistent. Inversely, by the Kolmogorov theorem [180], a consistent sequence of exchangeable partitions defines, in a unique manner, a partition Π_∞ of N.

4.3.3.2 The Poisson-Kingman partitions

Let's consider a sequence of random variables $Y = (Y_n,\ n \geq 0)$ defined on (Ω, F), with values in $[0, 1]$ that are exchangeable. This means they satisfy the following conditions:

An infinite sequence $(Y_0, Y_1, \dots)$ of random variables is exchangeable if

$$(\mathrm{Y}_0, \mathrm{Y}_1, \dots) \overset{\mathrm{d}}{=} \left(\mathrm{Y}_{\sigma(0)}, \mathrm{Y}_{\sigma(1)}, \dots\right) \tag{4.8}$$

For any permutation $\sigma\ of\ N$ we are only moving a limited number of natural numbers. Let $M_1([0,1])$ be the set of probabilities defined on the Borel set $[0,1]$, equipped with the weak topology of convergence and Borelians associated to this topology. A random probability measure M is a function of Ω in $M_1([0,1])$ that is measurable.

And, let M be a random probability measure. Let $(Y_1, \dots, Y_n, \dots)$ be random variables of the conditional law M, i.e: $\forall$ sequence of functions F_n: $[0,1] \to R$ measurable boundaries.

$$\forall n, E[F_1(Y_1), \dots, F_n(Y_n)]$$
$$= \int_\Omega P(d\omega) \int_{[0,1]} M(\omega, dy_1) F_1(dy_1) \times \dots \times \int_{[0,1]} M(\omega, dy_n) F_n(dy_n)$$

The sequence $(Y_i)_{i\geq 0}$ is exchangeable and $(Y_1, \dots, Y_n, \dots)$ is a M-Mixture and the De Finetti theorem says that any exchangeable sequence is a mixture.

Let M be a random probability measure and a M-Mixture $Y = \left(Y_{n,}, n \geq 0\right)$. We define a random partition $\Pi_\infty(Y)$ associated with the exchangeable sequence $(Y_1, Y_2, \dots)$ by:

$$i \sim \Pi_\infty(Y) \leftrightarrow Y_i = Y_j$$

It is obvious that $\Pi_\infty(Y)$ is exchangeable and the distribution of Π_∞ is determined by the random atoms of the measure M. Now, let's introduce the following set:

$$T_1 \coloneqq \left\{p^\downarrow = \left(p_n^\downarrow, n \geq 0\right);\ \textstyle\sum_{n=1}^{\infty} p_n^\downarrow \leq 1\right\} \tag{4.9}$$

For each sequence of atoms $p^\downarrow$ and T_1 , we can define a measure of a random probability:

$$\mu_{p^\downarrow}(dy) = \textstyle\sum_{i\geq 1} p_i^\downarrow\ \delta_{\frac{1}{i}}(dy) + \left(1 - \sum_{i\geq 1} p_i^\downarrow\right) 1_{[0,1]} dy, \tag{4.10}$$

And the function $p^\downarrow \to \mu_p^\downarrow$ is measurable.

Let $\Pi_\infty \coloneqq (\Pi_n)$ be an exchangeable random partition on N. And let $N_{n,i}^\downarrow$ be a sequence of bloc sizes of Π_n classified in descending order. $\exists$ a random variable $P = \left(P_i^\downarrow\right)$ in T_1 such that $\forall i, \frac{N_{n,i}^\downarrow}{n} \to n \to \infty\ P_i^\downarrow$.

In the context of Ricinus Communis oil extraction, blocs of random repartition represent various quantities of the extracted castor oil as well as the total quantity of extracted oil available for the production of biofuel. In order to study the distribution of quantities of castor oil, it is sufficient to study the quantities of oil related to the blocs of $\Pi_\infty(Y)$.

The Kingman theorem for exchangeable partitions is equivalent to the De Finetti theorem for exchangeable variables. The above theorem establishes the correspondence between the random partition and the frequency of quantities of extracted oil with the atoms of a random measure on T_1.

4.3.3.3 Stable partitions

A finite subordinator S is a positive process of independent homogenous growth. Also, a subordinator $(S_s)_{s\geq 0}$ is ascendant such that $S_0 = 0. \forall\, t > 0,$ the S_t law is infinitely divisible and characterised by its Laplace transform. Its Φ Laplace exponent verified the following Khintchine-Levy formula:

$\Phi(\lambda) = \mathrm{d}\lambda + \int_{(0,\infty)}\left(1 - e^{-\lambda y}\right)\Lambda(dy)$, where Λ was the Levy measure. Let $(\Delta_s)_{t\geq 0}$ be the process of jumps associated to S, the measure $M = \sum_{t:\Delta_t>0} \delta_{(t,\Delta_t)}$ was a Poisson measure on $(0,\infty)^2$ of intensity $dt \otimes \Lambda(dy)$. This resulted to the Levy-Ito decomposition $S_t = dt + \sum_{t\geq 0} \Delta_t$. For this work, a drift function X was added to the model. This made the model to be a nonstationary process that is null. Both the mean and the variance of the function increase over time. For a real positive k_0, the random measure $M_{k_0} = \sum_{\substack{t:\Delta_t>0 \\ t\in[0,k_0]}} \delta_{\Delta_t}$ was a Poisson measure on (0,∞) of intensity $k_0\Lambda(dy)$.

Proposition 1

Two independent subordinators $S = (S_t, t \geq 0)$ and $T = (T_t, t \geq 0)$ of Laplace exponents $\Psi\, and\, \Phi$ respectively were considered. Then $SoT = \left(S_{T_t}, t \geq 0\right)$ was a subordinator of Laplace exponent $\Phi o \Psi$. In addition, a subordinator T, where the Levy measure verified the above proprieties, allowed the definition of an exchangeable random partition by considering blocs of partition for the sequence of frequencies of the subordinator:

$$\mathrm{P_i^{\downarrow}} = \frac{\Delta_{\mathrm{i}}^{\downarrow}}{\mathrm{T_{S_0}}}, \tag{4.11}$$

where $\Delta_i^{\downarrow}$ was the sequence of jumps $\Delta_s \neq 0$ for $s \leq S_0$. By hypothesis, the set of all $s \geq s_0$ was discrete because $\Lambda([\varepsilon,\infty[) < \infty$ and it was possible to align the jumps in a descending order. For the modelling of biofuel extraction and production, T_{S_0} the total quantity of castor oil that was used for biofuel production and the jumps $\Delta_i^{\downarrow}$ corresponded to the ieth biggest quantity of castor oil of the platform.

4.3.3.4 Stable subordinator

A stable subordinator is a subordinator T of the Levy measure $\Lambda(y) = s\,\rho(y)dy$ where $s > 0$ [181], and:

$$\rho(y) = \frac{\beta}{\Gamma(1-\beta)}\frac{1}{y^{1+\beta}} \tag{4.12}$$

For this particular subordinator, the expression of the Laplace exponent was: $\Phi(\lambda) = \lambda^{\beta}$ and T verified the scaling propriety $(T_s)_{s\geq 0} = \left(\lambda^{-1/\beta}T_{\lambda s}\right)_{s\geq 0}$. Now, let $\varepsilon > 0$ and $(T_s)_{s\geq 0}$ a subordinator β-stable. Therefore:

$$C_{\varepsilon,\beta} := \Lambda\big((\varepsilon,\infty)\big) = \frac{\varepsilon^{-\beta}}{\Gamma(1-\beta)} < \infty \tag{4.13}$$

And the subordinator T verified all the sufficient conditions to define an exchangeable random partition from the normalised jumps of T. The stable partition was therefore the random partition associated to such subordinator. Therefore, the stable subordinator was the best to describe the quantities of castor oil available for the production of biofuel. The Bolthausen and Sznitman fragmentation model [182] introduced in the next section described the total quantity of the extracted castor oil in a platform, which was compatible with the jumps of a stable subordinator.

4.4 Modelling of the extracted Ricinus Communis oil using the Bolthausen-Sznitman model

This model was introduced by Bolthausen and Sznitman [174] in order to study a model called Random Energy Model (REM) introduced by Derrida [175]. In this thesis, we use this model to study the normal biofuel distributed in the heavy crude Ricinus Communis oil extracted for the biofuel production.

4.4.1 REM and extraction of Ricinus Communis oil mathematical model.

In order to model the oil extraction process and its storage into the platform, a toy model was used with a goal to qualitatively explain the form of laws of storage sizes (Quantities of oil in storages). The model was based on three hypotheses:

(i) *Hypothesis of subdivision into basic equivalent storages.*

The assumption was that the selected storages (observed) inside the same platform were subdivided into a great number of small basic storages $g_1, \ldots, g_M$ that were equivalently filled in a statistical point of view. The content of those small storages varied over time depending on the sequence of extraction. In addition, it was observed that at each sequence of extraction,

the variation of the quantity of oil in a basic storage affected its size in proportion than in the absolute value. For example, the fusion with a similar storage multiplied the quantity by two.

(ii) *Hypothesis of the middle-size storage*.

Here, the assumption was that an important number of factors N contributed to the extraction of basic storages and determined their size and position within the platform. There was no detailed knowledge of those factors at the level of basic storages. It was further assumed that those N factors played the same role, were independent and acted randomly in the same way and independently to each basic storage. And this constituted the hypothesis of the middle-size storage. Various interests were on the fluctuation created by the N factors on the quantity of *ricinus communis* oil of the storages. More precisely, the following type of simplified action was selected: The factor $i, 1 \leq i \leq N$ provoked the fluctuation of the quantity of oil in the storage g_p by a factor $exp(\epsilon_i^p)$, where ϵ_i^phad two possible values –v and v. The cumulated action of the N factors on the storage g_p was to multiply the quantity of oil inside the storage g_p by the coefficient: $\exp\left(\sum_{i=1}^{N} \epsilon_i^p\right)$.

The above hypotheses of the middle-size containers were traduced by assuming that the variables $\left(\epsilon_i^p, 1 \leq i \leq N, 1 \leq p \leq M\right)$ were independent and of the same law:

$$\mathbb{P}\left(\epsilon_i^p = -v\right) = \mathbb{P}\left(\epsilon_i^p = v\right) = \frac{1}{2} \tag{4.14}$$

(iii) *Hypothesis of the economy of subdivision*.

In this system, the N factors contributing to the extraction of oil in a storage were the N factors that were permitted to distinguish and determine the subdivision into M basic storages. This leaded to assume that N was close to the minimal number of factors that were necessary to discriminate the M storages. As $\{-v, v\}^N$ represented the 2^N possible values of ϵ_i^p. It was coherent to choose N in the order of log M. For the sake of clarification, N was chosen great and: $M = 2^N$. And for all containers g_p, the following was applied:

$$H_N^*(g_p) = \sum_{i=1}^{N} \epsilon_i^p \tag{4.15}$$

As N was assumed to be great, the law $(N\, v^2)^{-1/2} H_N^*(g_p)$ was close to a centered Gaussian standard that was noted $G(g_p)$. In the future, the following expression was considered:

$$H_N^*(g_p) \approx v\sqrt{N}\, G(g_p) \tag{4.16}$$

The quantity of oil in the container which defined its size was:

$$\langle t_v, N\rangle \exp\left(v\sqrt{N}\, G(g_p)\right) \tag{4.17}$$

Where $\langle t_v, N\rangle$ was a typical size common to all basic storages that fixed the scale in which the basic storages were placed. Therefore, the relative size of the basic storage $g_p$was:

$$\mu N, v\left(g_p\right) = \frac{\exp\left(v\sqrt{N}\, G(g_p)\right)}{\sum_{q=1}^{2^N} \exp\left(v\sqrt{N}\, G(b_q)\right)} \tag{4.18}$$

Thus, the question was to know how those relative sizes behaved when N was big. Also, when v was big, meaning when the phenomena amplitude of fluctuation was great, the basic storages really observed were those for which the variable $v\sqrt{N}\, G(b_q)$ took important values. However, when v was small, this didn't happen: the storage was split up when N was very big. There was a critical value v_c on top of which the first asymptotic behaviour prevailed and under which the second prevailed too.

A quick heuristic explaining what happened when v was small is provided. Therefore, the total amount of quantity of oil in the storage was given by: $Z_{v,N} = \langle t_v, N\rangle \exp\left(v\sqrt{N}\, G(g_p)\right)$
A simple calculation indicated that:

$$\mathbb{E}[Z_v, N] = \langle t_v, N\rangle 2^N \exp(N\, v^2/2) = \langle t_v, N\rangle \exp\left(N\left(\frac{v^2}{2} + \log(2)\right)\right)$$

And

$$\mathrm{var}(Z_v, N) = \langle t_v, N\rangle\left(\exp\left(N(2\, v^2 + \log(2))\right)\right) - \exp\left(N(v^2 + \log(2))\right)$$

Thus

$$\frac{\mathrm{var}(Z_v, N)}{(\mathbb{E}[Z_v, N])^2} \sim N \to \infty \exp\left(N(v^2 - \log(2))\right)$$

Thus, when $v < \sqrt{\log(2)}$, $\frac{Z_v,N}{\mathbb{E}[Z_v,N]}$ approached 1 in probability, a more precise argument using the Lindeberg criterion showed that a central limit theorem happened. In this case, none of the basic storage was very big. This came from the fact that distributions queues of log-normal variables $exp\left(v\sqrt{N}\, G(g_p)\right)$ satisfied a central limit theorem. However, when v was very big, these distributions queues became big and some storages were abnormally big and influenced a lot on the μ_v, N distribution. There were precise mathematical results on the behavioural limit of μ_v, N which had the same law than the Gibbs measure derived from the Derrida's REM. On works related to the REM, v played the role of inverse of the β temperature and in the traditional normalisation, it was convenient to take $\sqrt{2}.v = \beta$.

The critical value v_c was not $\sqrt{\log(2)}$ as this seemed to be suggested by the succinct analysis by variance, but it was very close. Therefore, $\mathrm{v_c} = \sqrt{\log(2)}$.

This numerical value did not have a real physical significance for our model because of the many choices done in order to manage less possible parameters in our qualitative description. But the advantage of v_c is its existence.

Case where $v < v_c$. The following measure was considered:

$$\mathrm{m_{v,N}} = \sum_{\mathrm{p=1}}^{2^{\mathrm{n}}} \mu_{\mathrm{v}}, \mathrm{N}(\mathrm{g_p})\delta_{\mathrm{p^{2-N}}} \tag{4.19}$$

Which was a measure in [0,1]. Then, surely when N approached the infinity, $\mathrm{m_{v,N}} \rightarrow \mathrm{N}$
For the convergence of measures in [0,1], where λ designated the Lebesgue measure. It was observed that the limit did not depend on v. One way of interpreting this result was to say that no basic storage prevailed and that the storages were not structured as bigger clusters. Thus, when $v < v_c$, the storage presented many storages regularly distributed with a size that followed the normal-log law. The bigger N was, the more storages there were.

Case where $v > v_c$.We order the relative sizes $\left(\mu_{v,N}(g_p), 1 \leq p \leq 2^N\right)$ by decreasing order:

$$\mu_{\mathrm{v,N}}^{\downarrow} = \left(\mathrm{P_1^{\downarrow,N,v}} > \mathrm{P_2^{\downarrow,N,v}} > \cdots > \mathrm{P_{2^N}^{\downarrow,N,v}}\right) \tag{4.20}$$

where, $\left\{\mathrm{P_n^{\downarrow,N,v}}; 1 \leq \mathrm{n} \leq 2^{\mathrm{N}}\right\} = \left(\mu_{\mathrm{v,N}}(\mathrm{g_p}), 1 \leq \mathrm{p} \leq 2^{\mathrm{N}}\right)$. When $N \rightarrow \infty$, then:

$$\mu_{\mathrm{v,N}}^{\downarrow} \rightarrow \mathrm{P}^{\downarrow,\alpha} = \left(\frac{\Delta_{\mathrm{n}}^{\alpha\downarrow}}{\mathrm{T_1}}; \mathrm{n} \geq 0\right) \tag{4.21}$$

Where $\left(\frac{\Delta_n^{\alpha\downarrow}}{T_1}; n \geq 0\right)$ represented the reordered suite of jumps of a stable subordinator $(T_t; t \in [0,1])$ of index

$$\alpha = \frac{\sqrt{2\log(2)}}{\mathrm{v}} \tag{4.22}$$

In other words, the basic storages were grouped in order to form clusters that were the storages that were observed in the platform and the relative were distributed as a $PD(\alpha, 0)$ law. The convergence (4.4) gave the form of the statistics of the relative sizes but the study of its proof showed that the sizes of the basic storages were ordered by decreasing order converging towards the reordered jumps of a stable subordinator of index $\alpha = \frac{v_c}{v}$.

Furthermore, the proof used a result on the extreme Gaussian values. In order to give an idea about it, the convergence for big containers was demonstrated as follows:

$$\mathrm{T_1^{v,N\downarrow}} \rightarrow \Delta_1^{\alpha\downarrow} \tag{4.23}$$

Proof of (4.23)

Later on in this work, it was established that $\left(\Delta_1^{\alpha\downarrow}\right)^{-\alpha}$ followed an exponential law. Thus, to prove (4.23) meant to prove $\left(T_1^{v,N\downarrow}\right)^{-\alpha}$ converged by law towards an exponential.

Therefore:

$$M_N = \max_{1\le p\le 2^N} G(g_p)$$

And it was observed that:

$$T_1^{v,N\downarrow} = \langle t_{v,N}\rangle \, exp\left(v\sqrt{N}\,(M_N)\right)$$

Because α and v were linked, therefore:

$$\left(T_1^{v,N\downarrow}\right)^{-\alpha} = \left(\langle t_{v,N}\rangle\right)^{-\alpha} exp\left(-\sqrt{2\log(2^N)}\; M_N\right)$$

Furthermore, $x > 0$was fixed. Then, we have the following equalities:

$$\mathbb{P}\left(\left(T_1^{v,N\downarrow}\right)^{-\alpha} \ge x\right) = \mathbb{P}\left(M_N \le \frac{\alpha \log\langle t_{v,N}\rangle}{\sqrt{2\log(2^N)}} - \frac{\log x}{\sqrt{2\log(2^N)}}\right)$$

$$= \mathbb{P}(M_N \le a(N,x))$$

where:

$$a(N,x) = \sqrt{2\log(2^N)} - \frac{\log x}{\sqrt{2\log(2^N)}} - \frac{1}{2}\frac{\log(4\pi\log(2^N))}{\sqrt{2\log(2^N)}}$$

Thus, we have:

$$\mathbb{P}\left(\left(T_1^{v,N\downarrow}\right)^{-\alpha} \ge x\right) = \left[1 - u\left(a(N,x)\right)\right]^{2^N}$$

Where we recall that:

$$u(y) \coloneqq \frac{1}{\sqrt{2\pi}}\int_y^{\infty} e^{-z^2/2}\,dz \sim y\to\infty \frac{1}{y\sqrt{2\pi}} e^{-y^2/2}$$

A basic calculation showed that in order to fix x, and N approaching the infinity, we have:

$$u\left(a(N,x)\right) = 2^{-N}x\left(1 + o(1)\right),$$

Which leaded to $\left[1 - \mathrm{u}\left(\mathrm{a(N,x)}\right)\right]^{2\mathrm{N}} \sim \mathrm{N}\to\infty\, \mathrm{e}^{-\mathrm{x}}$.

And this concluded the proof of (4.23).

This qualitative model enabled the conciliation of the two types of laws observed for the quantity of oil inside the respective storages. In fact, the distribution queue of the size of storages was considered in a specific platform and through a log-log graph, and it was observed that this distribution decreased approximately by $x^{-\alpha}$.

When $\alpha < 1$, than most equations enabled to conclude the existence of the $v > v_c$ regime and that the distribution of Dirichlet Poisson prevailed. The sizes of containers were similar to the jumps of a stable subordinator of index α that were truncated at a certain level (of extraction or rentability) and they were distributed as Pareto-Levy laws.

Secondly, when $\alpha > 1$, then the existence of the $v < v_c$ regime was deduced and the sizes of containers were distributed according to the normal-log laws.

When the model pretended to qualitatively explain this phenomenon, it was necessary to consider it in a dynamic manner, which in other words meant to consider that it was susceptible to vary within various time scales. Therefore, the amplitude v became a time function v(t). In fact few examples of such situations are highlighted:

The function $t \longmapsto v(t)$ remains below the level v_c. In this case, the amplitude v(t) has a weak impact on the form of the size laws of the storages which follow a normal-log law. The increase or decrease of v(t) does not bring anything that is irreversible in the statistics of the storages.

The function $t \longmapsto v(t)$ crosses the level v_c. In this case, the storages are brutally organised in clusters and this brings an irreversible change on the statistics of the sizes of the storages. In fact, if after crossing v_c, the function $t \longmapsto v(t)$ crosses v_c later on in descending, there is a brutal fragmentation and if N is very big, it is reasonable to interpret it as there is only very small residual storages remaining.

The function $t \longmapsto v(t)$ evolves by remaining higher than v_c. In this case, we can say that when v(t) increases, the basic storages have the tendency to aggregate and when v(t) decreases, they fragment, and this, without causing a total disappearance of the oil. In terms of coalescence-fragmentation of the statistics of the relative sizes of the storages, this evolution is justified by the $PD(\alpha, 0)$ Poisson-Dirichlet distributions, in which, according to the analysis, constitute the limit models.

The study of variations of $v \longmapsto \mu_{v,N}$ has motivated the introduction of the Bolthausen-Sznitman coalescent/fragmentation that defines the mechanism of evolution in terms of the

coalescence/fragmentation reversible on the limit laws. Therefore, it is a $\alpha \longmapsto P^{\downarrow,\alpha}$ process that is reversible. The definition of this process is the content of section 4.4.3.

4.4.2 The two-parameter Poisson-Dirichlet distribution

$\forall\, \alpha \in]0,1[$ and $\forall\, \theta > -\alpha, \exists$ a random partition Π_∞ of $\mathbb{N}$ that verifies the following expression:

$$p_{\alpha,\theta}(n_1, \dots, n_k) = \frac{(\theta+\alpha)_{k-1\uparrow\alpha} \prod_{i=1}^{k} (1-\alpha)_{n_{i-1\uparrow 1}}}{(\theta+1)_{n-1\uparrow 1}}$$

The sequence of asymptotic frequencies of blocks ranged by appearing order of their smallest element admits the following v representation:

$$\left(\tilde{P}_1, \tilde{P}_2, \dots\right) = (w_1, \bar{w}_1 w_2, \bar{w}_1 \bar{w}_2 w_3, \dots)$$

Where $\bar{w}_i = 1 - w_i$ and w_i are independent random variables of beta-law $(1-\alpha, \theta + i\alpha)$.

By ranging the sequence $\left(\tilde{P}_i\right)_{i\geq 1}$ of random frequencies of the partition N in descending order, a sequence $\left(P_i^{\downarrow}\right)_{i\geq 1}$ in which the distribution was called the Poisson-Dirichlet was obtained, and it was denoted $PD(\alpha, \theta)$ [13, Pitman and co,]. Also, it was possible to define the two-parameter model using the stable subordinators which allowed the verification of the distribution $PD(\alpha, 0)$ that corresponded to the distribution of the jumps of stable subordinators. The correspondence between the two point of view was detailed by Pitman [176]. Thus, it was possible to define the Bolthausen-Sznitman fragmentation model in order to propose the modelling of *ricinus communis* reserves destined to biofuel production.

4.4.3. The Bolthausen-Sznitman fragmentation model

In order to describe the progressive fragmentation of initial and/or harvested *ricinus communis* oil reserves, the notions of fragmentation and coagulation of N partitions were defined. On this, readers are referred to Pitman [183] who largely tackled these questions.

Definition 1

A. Let $\Pi = (A_1, A_2, \dots)$ be a partition of $\mathbb{N}$, and let $\Pi^{(.)} = \left(\Pi^{(i)}, i = 1, \dots, n\right)$ be a suite of partitions of $\mathbb{N}$. We note $A_j^{(i)}$the j$^{\text{th}}$ block of the $\Pi^{(i)}$partition. For all natural numbers i, we consider the $\Pi_{|A_i}^{(i)}$ of A_i induced by the i$^{\text{th}}$ partition $\Pi^{(i)}$ of the suite $\Pi^{(.)}$. This means:

$$\Pi_{|A_i}^{(i)} = \left(A_j^{(i)} \cap A_i, j \in \mathbb{N}\right).$$

The partition obtained by assembling the set of all blocks of all $\Pi_{|A_i}^{(i)}$ partitions form of a partition of $\mathbb{N}$ that we note Frag$(\Pi, \Pi^{(.)})$.

B. Let $\Pi = (A_1, A_2, \ldots)$ and $\Pi' = (A_1', A_2', \ldots)$ be two partitions of $\mathbb{N}$. We call coagulation of Π by Π' the partition $\mathrm{Coag}(\Pi, \Pi')$ of the blocks B_j defined by:

$$B_j := \bigcup_{i \in A_j'} A_i$$

The coalescence and fragmentation processes are Markovian processes with values in the P_∞ space.

Definition 2

Let $\Pi = (\Pi(\mathrm{t}), \mathrm{t} \geq 0)$ be a Markovian process with values in P_∞ which is continuous in probability. Π was an exchangeable coalescence process when its semi-group was described in the following manner: $\forall$ positive t and t', the conditional distribution of $\Pi(\mathrm{t} + \mathrm{t}')$ knowing $\Pi(\mathrm{t}) = \pi$ was the same than that of $\mathrm{Coag}(\pi, \Pi')$, where Π' was an exchangeable random partition whose law only depended on t'. One of the well-known example of exchangeable random partition was the Bolthausen-Sznitman coalescent.

Lemma 1

α and β were fixed in $]0,1[$. And, let Π be the random partition of the $PD(\alpha, 0)$ distribution, and Π' be the random partition of $PD(\beta, 0)$. Then, the $\mathrm{Coag}(\Pi, \Pi')$ partition was a random partition (exchangeable) of $PD(\alpha\beta, 0)$. Thereby, the Π^{BS} Bolthausen-Sznitman coalescent was defined by its probabilities of transition P_t^{BS} by considering the following operator on the continuous functions space: $\Phi: P_\infty \to \mathbb{R}$,

$P_t^{BS}(\Phi)(\Pi) = \mathrm{E}\left[\Phi\left(Coag\left(\Pi, \Pi^{(e^{-t})}\right)\right)\right], \Pi \in \Pi_{\mathbb{N}}$, where $\Pi^{(e^{-t})}$ is a random partition of the $\mathrm{PD}^{(e^{-t},0)}$ distribution.

The abstract representation of a partition in $P_\infty$was less significant than the frequencies that were associated to it by the Kingman theorem because its frequencies represented the relative quantities of *ricinus communis* oil. Thus, it was necessary to reformulate the Bolthausen-Sznitman mechanism of coalescence as follows:

Let $(T_s^1, \geq 0)$ and $(T_s^2, \geq 0)$ be two independent subordinators of exponents $\alpha_1 = e^{-t_1}$

and $\alpha_2 = e^{-t_2}$ respectively. $T_{1_0}T_2$ is a stable subordinator with $\alpha_1\ \alpha_2 = e^{-(t_1+t_2)}$. And the following notations were posed:

1. The expression $\left(\Delta_n^{\alpha_1,\downarrow}; n \geq 0\right)$ was the jumps $\{T_s^1 - T_{s^-}^1; s \in [0, T_1^2]\}$ indexed in a decreasing order; and,

2. The expression $\left(\Delta_n^{\alpha_1\alpha_2,\downarrow}; n \geq 0\right)$ was the jumps $\left\{(T^1{}_0\, T^2)_s - \left(T^{1\,0\,T^2}\right)_{s^-}; s \in [0,1]\right\}$ indexed in a decreasing order. A Bolthausen-Sznitman coalescent $(\Pi_t^{BS}, t \geq 0)$ whose initial value was $\Pi^{BS}(0) = e$ was considered. Therefore:

$P^{t_1\downarrow} = \left(P_n^{t_1\downarrow}; n \geq 0\right)$ and $P^{t_1+t_2\downarrow} = (P_n^{t_1+t_2\downarrow}; n \geq 0)$

The associated frequencies to Π_t^{BS} and $\Pi_{t_1+t_2}^{BS}$ by the Kingman theorem.

Then, it can be shown that:

$$\left[\left(\frac{\Delta_n^{\alpha_1,\downarrow}}{T_{T_1^2}^1}; n \geq 0\right); \left(\frac{\Delta_n^{\alpha_1\alpha_2,\downarrow}}{T_{T_1^2}^1}; n \geq 0\right)\right] = \left[P^{t_1\downarrow}; P^{t_1+t_2\downarrow}\right]$$

The process was also considered by inversing the time in order to observe the progressive fragmentation of reserves. In general, there was no perfect duality between the fragmentation process and the coalescence process. In the case of the Bolthausen-Sznitman coalescent, the following results were obtained:

Proposition 2

The Π_f^{BS}process defined by $\Pi_f^{BS}(u) \coloneqq \Pi^{BS}(-\ln u)$ obtained by reversing the time in the Bolthausen-Sznitman coalescent is a Markov process on P_∞ inhomogeneous in time. Its probabilities of transition are such that for $0 < u \leq u' \leq 1$, conditionally to $\Pi_f^{BS}(u) = \pi$, the partition $\Pi_f^{BS}(u')$ is distributed as Frag(π, $\Pi^{(.)}$), where $\Pi^{(.)} = (\Pi^{(1)}, \Pi^{(2)}, \dots)$ was a suite of random partitions of distribution PD(u', -u).

Conclusion of the chapter

The Bolthausen-Sznitman fragmentation model provided a process such that at all instant $\in$]0,1] , $\Pi_f^{BS}(u)$, was the distribution PD(-ln u, 0). The process was not defined at 0, and not in a manner to complete the model. Therefore, $\Pi_f^{BS}(0)$ was the partition of $\mathbb{N}$ that was equal to the set $\mathbb{N}$ itself. In fact, at time 0, all *ricinus communis* oil reserves were stored in only one initial container. The time u that indexed the Bolthausen-Sznitman fragmentation process was considered like a specific time that elapsed in a linear manner. And this time allowed the structuration of the remaining successive fragmentations of the initial resources. Also, it was verified that the Bolthausen-Sznitman fragmentation model validated the two hypotheses H_a and H_b that were imposed. In mathematical terms, the H_a hypothesis meant that the fragmentation of reserves had a discrete-time Markov process behaviour, which was really the Π_f^{BS}process. The H_b hypothesis meant that all the partitions were fragmented according to the random partitions $(\Pi^{(1)}, \Pi^{(2)}, \dots)$ that were of the same distribution. And this constraint was well respected by the Π_f^{BS}process. The partitions that guided the fragmentation between the times u and u' were all of PD(u', -u) distribution. It would had been ideal if the only process of fragmentation that would verify H_a, H_b and H_c was the Markov process described here, but to the best of our knowledge, such result has never been demonstrated. Nevertheless, it appeared simple to start by describing things in the fragmentation point of view, which was easier to study mathematically. However, because $v > v_c$, the model authorised the dual mechanisms of coalescence and fragmentation to happen successively. This was more realistic than to suppose that a big *ricinus communis* oil quantity was progressively fragmented over time. Therefore, successive processes of fragmentation and coalescence happened and this, without any change of the type of law describing the size of the container: $(\alpha, 0)$, where only α varied with time. At the end, other *ricinus communis* storages remained distributed like the jumps of stable subordinators and they were considered like Pareto-Levy random variables. Finally, the problem of *ricinus communis* oil reserves was covered in a qualitative point of view and it justified the use of the Pareto-Levy distribution in order to model the quantities of *ricinus communis* oil stored for the production of biofuels. The adapted REM model of *ricinus communis* oil reserves storages indicated that it was natural to represent those quantities by using the jumps of stable subordinators. Also, in the context of the Bolthausen-Sznitman model, this type of distribution was not changed by coalescence or fragmentation operations of the reserves (for $v > v_c$).

Chapter 5: Probabilistic model for biofuel production from *Ricinus Communis* oil

This chapter focuses on all intervening factors in the modelling of biofuel production based on *ricinus communis* oil. Each factor will be modelled. This will render the model of production usable in practice.

5.1 Introduction

Many existing models for biofuel production are essentially econometrics models. Moroney and Berg [184], and Kemp and Kasim [185] studied biofuel production using temporal series. Cleveland and Kaufmann [186] work reconciled the econometrics point of view and the adjustments suggested by Hubbert [186]. The ultimate objective of their methods was to highlight the relationship that exists between production and explanatory variables. However, none of those models took into consideration the probability distribution of the quantity of the extracted oil. Interestingly, from extraction to production, all phases intervening in the production of biofuel are influenced by the amount of oil extracted. In this chapter, a complete model of biofuel production taking into consideration the probability distribution of the quantity of *ricinus communis* oil extracted is proposed. And biofuel production was the aggregate of the production of small quantities of the extracted oil.

5.2 Modelling of the transesterification process for biofuel production

The objective of this section is to present a modelling technique that combined the Komers model [95] with the model of Noureddini [93] in order to obtain a general chemical kinetic model of the transesterification process able to use various types of *ricinus communis* oils. Various types of oils were used in the same reactor. The section also highlights the result of experiments from new reactions and catalysts as well as the optimisation of the reactor and finally, the prediction of the behaviour of the reactional system based on operational conditions. This helped for the reduction of the number and the cost of experiments while mastering the synthesis and the use of biofuels within the Sub-Saharan African context. The definition of transesterification suggests that it is the classic technique of biofuel production. It is a process in which vegetable oils, animal biomasses or microalgae oils are mixed with alcohol (ethanol or methanol) in the presence of a catalyst (hydroxide of sodium or potassium). Figure 5.1 shows the chemical structure of transesterification process.

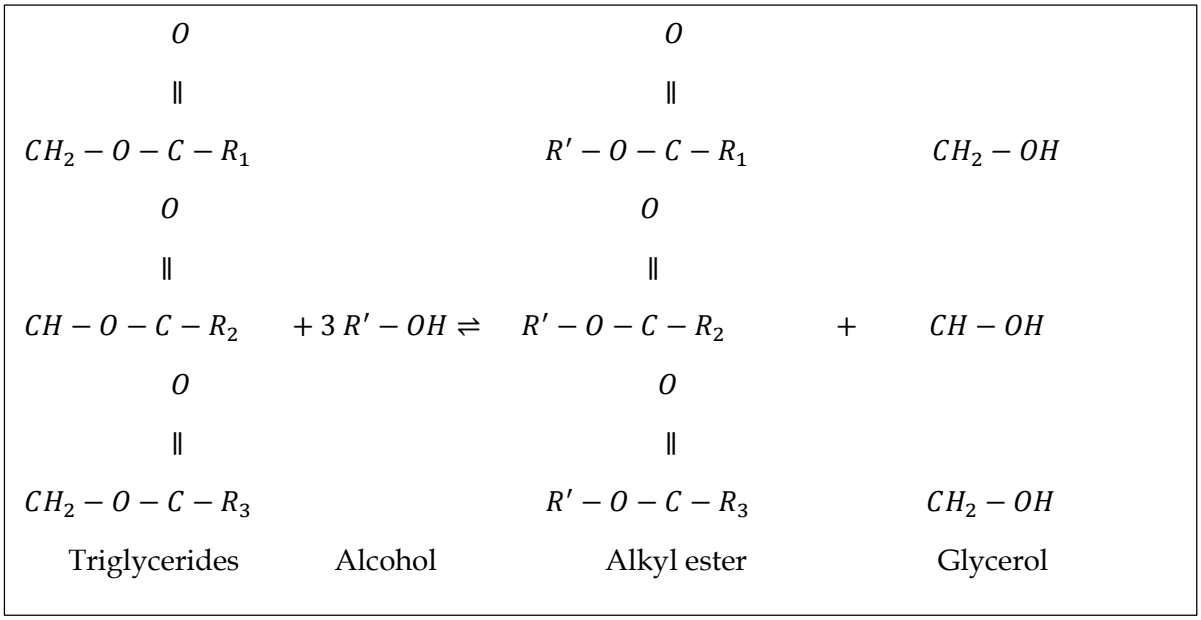


Figure 5. 1: Chemical equation of transesterification

Vegetable oils are essentially triglycerides. Their reaction with alcohol leads to alkyl esters of vegetable oils and to glycerol. The transformation of oil into alkyl esters or methyl enables the reduction of the molecular mass by a third of the one of the oil. It also enables the reduction of the viscosity of a factor eight, the reduction of the density and increases the volatility. The physical properties of ethyl esters and methyl obtained from the reaction of transesterification are close to those of diesel. The transesterification model used in this section is based on the analysis and the combination of both the Komers and the Noureddini models [93, 95]. In fact, the Noureddini model takes into consideration the intensity of the agitation during the reaction and enables the variation of temperature. However, the model does not take into account saponification. On the other hands, the Komers model takes into account the type of catalyst, but all the coefficients of the model are only valid for a temperature of ≈ 23 ℃, even though the reaction environment can reach a temperature of up to 85 ℃. Hence, the necessity to introduce a model that considers factors that influence the reaction. This ensures a more complete simulation and a more pertinent optimisation of the transesterification process. Furthermore, the objective of the model is to transfer the results of the optimisation to various experimental tests of the reactor. This enables the reduction of the number and the cost of experiments, especially the high cost of methanol. Therefore, the model introduced took into account the energy of activation, which means the possibility to perform a simulation at various temperatures, the effect of agitation intensity and the type of

alcohol (MeOH) or vegetable oil (VO) used as well as the molar ratio $\frac{\mathrm{MeOH}}{\mathrm{VO}}$. This resulted to a system of ten differential equations with coefficients [187]:

$$-\frac{dY_{TG}}{dt} = b.\frac{Y_{OH^-}}{YH_2O}.(k'_1.Y_{TG}.Y_{ROH} - k'_2.Y_{DG}.Y_{BD}) + b.Y_{OH^-}.k_{16}.Y_{TG} \quad (5.1)$$

$$-\frac{dY_{DG}}{dt} = b.\frac{Y_{OH^-}}{YH_2O}.(-k'_1.Y_{TG}.Y_{ROH} + k'_2.Y_{DG}.Y_{BD} + k'_3.Y_{DG}.Y_{ROH} - k'_4.Y_{MG}.Y_{BD}) + b.Y_{OH^-}.(-k_{16}.Y_{TG} + k_{17}.Y_{DG}) \quad (5.2)$$

$$-\frac{dY_{MG}}{dt} = b.\frac{Y_{OH^-}}{YH_2O}.(-k'_3.Y_{DG}.Y_{ROH} + k'_4.Y_{MG}.Y_{BD} + k'_5.Y_{MG}.Y_{ROH} - k'_6.Y_{GL}.Y_{BD}) + b.Y_{OH^-}.(-k_{17}.Y_{DG} + k_{18}.Y_{MG}) \quad (5.3)$$

$$-\frac{dY_{GL}}{dt} = b.\frac{Y_{OH^-}}{YH_2O}.(k'_5.Y_{MG}.Y_{ROH} - k'_6.Y_{GL}.Y_{BD}) + b.Y_{OH^-}.(k_{18}.Y_{MG}) \quad (5.4)$$

$$-\frac{dY_{ROH}}{dt} = b.\frac{Y_{OH^-}}{YH_2O}.(k'_1.Y_{ROH} - k'_2.Y_{DG}.Y_{BD}) + (k'_3.Y_{DG}.Y_{ROH} - k'_4.Y_{MG}.Y_{BD} + k'_5.Y_{MG}.Y_{ROH} - k'_6.Y_{GL}.Y_{BD}) - b.Y_{OH^-}.k_{15}Y_{BD} \quad (5.5)$$

$$\frac{dY_{BD}}{dt} = -\frac{dY_{ROH}}{dt} \quad (5.6)$$

$$-\frac{dY_{OH^-}}{dt} = b.Y_{OH^-}(k_{15}Y_{BD} + k_{16}Y_{TG} + k_{17}Y_{BD} + k_{18}Y_{MG} + k_{19}Y_{AGL}) \quad (5.7)$$

$$\frac{dY_{SV}}{dt} = -\frac{dY_{OH^-}}{dt} \quad (5.8)$$

$$-\frac{dY_{AGL}}{dt} = k_{19}Y_{OH^-}Y_{AGL} \quad (5.9)$$

$$-\frac{dYH_2O}{dt} = -\frac{dY_{AGL}}{dt} \quad (5.10)$$

With $b = [TG]_0$ [187]

5.2.1 Validation of the model

In order to validate the model, experiments were performed using the biofuel transesterification machine which comprised a primary tank where resistors brought oil to the reaction temperature level, a reservoir with an alkaline solution, sodium/potassium hydroxide, alcohol, and pipes. A pump placed in a tubing loop that enabled the mixture of glycerides with the strong base mixture and alcohol. A transparent gauge placed on the side of the tank that enabled the visual inspection of each phase. After transesterification, the bottom of the tank containing glycerol and other residues was evacuated. Then, water washing took place before drying which was the final step to recover biofuel.

5.2.2 Experimental part

Figure 5.2 shows a schematic diagram of the biofuel production process where *ricinus communis* oil was treated with sodium metasilicate; methanol was purified and residual solids were removed from the distillation bottoms liquid and fed to a reactor; then glycerol was separated from esters and produced glycerol destined for other uses on the one side, and produced methyl esters on the other side. Then methyl esters purification took place to produce biodiesel that was separated from waste.

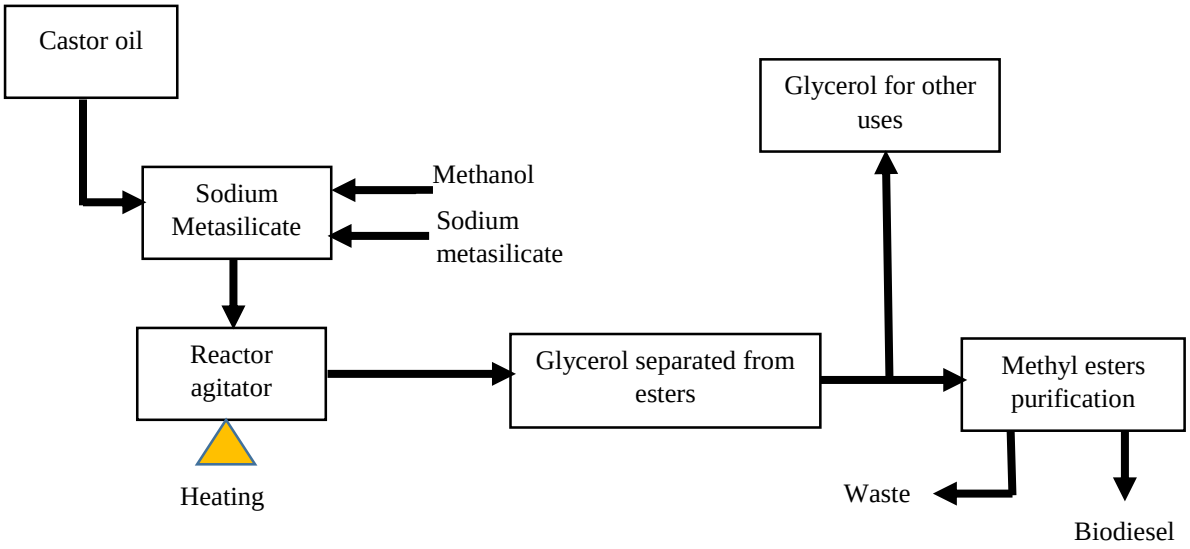


Figure 5. 2: Diagram of biofuel production

5.3 Viscosity

5.3.1 The kinetic viscosity

In this section, some methods for oil characterisations were presented. And the Poiseuille law-based method was used because in the case of Newtonians liquids, the flow through the capillary tube followed the Poiseuille law that links viscosity with the flow of the fluid [188].

5.3.2 Operational mode

Viscosity was measured with an "OSTWALD" viscometer [189]: A capillary tube of length l and of radius r links two tanks A and B. The liquid which was initially in B was introduced in A by aspiration above the position X_1. The time that the liquid put to flow between the two marks A and B were timed, which corresponded to a swept volume V. Then, the capillary was calibrated under exactly reproducible liquid loading and at exactly a controlled temperature. Thus, the Poiseuille equation was written as follows:

$$\mu = \frac{\pi r^2 pt}{8lV}$$ [190] (5.11)

Where p was the hydrostatic energy that forced the liquid to flow:

$$p = h\rho g \quad (5.12)$$

Also, h was the height of the slope and it varied during the experiment but always the same way from one experiment to another. ρ was the density of the liquid studied and g was the acceleration due to gravity (9.81 $m.s^{-2}$).

Thus:

$$\mu = k\rho t \quad (5.13)$$

Where k was a constant of the viscometer used.

Thereby, the more viscous the liquid was, the higher the time t.

Absolute determinations were rarely made, but rather relative measurements were made by calibrating the viscometer with a liquid of reference, generally water.

$$\mu_i = \mu_{ref} \times \frac{\rho_i}{\rho_{ref}} \times \frac{t_i}{t_{ref}} \quad (5.14)$$

ρ_i : Density of the liquid studied

ρ_{ref} : Density of the liquid of reference

μ_i : Viscosity of the liquid studied

μ_{ref} : Viscosity of the liquid of reference

t_i, t_{ref}: Time of flow of the liquid studied and of the liquid of reference.

5.3.3 Flow time measurements

- ✓ A suitable amount of the solution was inserted into the viscometer up to the filling line using a pipette.
- ✓ The solution was forced to pass through the capillary of the viscometer by blowing dry air with a rubber bulb.
- ✓ The operation was stopped when the liquid level was just above the top reference point.

- ✓ The viscometer was immersed in the 25 ℃ ± 0.1 ℃ bath.
- ✓ The viscometer was adjusted in such way that on the one hand, the bath level was at 20 mm above the upper graduation mark, and on the other hand, the tube axis was perfectly vertical.
- ✓ The liquid was allowed to flow by measuring the time taken by the meniscus to go down from the higher reference point to the lower reference point. That was the flow time of the solution.

The Tables 5.1, 5.2 and 5.3 indicate the various measurements performed at a temperature of 25℃:

Table 5. 1: Flow time of distilled water at 25 ℃.

Measurements	**1**	**2**	**3**	**4**
Time (seconds)	106” 39	106”42	106” 41	106” 38
Time (Milliseconds)	9639	9642	9641	9638
$t_{mean} = 9640'''$				

Table 5. 2: Flow time of acetic acid at 25 ℃.

Measurements	**1**	**2**	**3**	**4**
Time (seconds)	126” 76	126” 54	126” 02	125” 65
Time (Milliseconds)	12676	12654	12602	12563
$t_{mean} = 12624$				

Table 5.3 indicates oil densities and viscosities for temperatures other than **25**℃.

Table 5. 3: Density and viscosity of the Ricinus Communis oil

$T°$ (℃)	20	30	35	40	45	50	55	60
Density	.92	.910	.905	.901	.895	.890	.889	.889
Viscosity	69.91	44.87	37	30.21	28	25	19.27	19

Figure 5.3 displays the variations in ricinus communis oil density as function of temperature. In this Density against temperature figure, the blue curve represents castor oil.

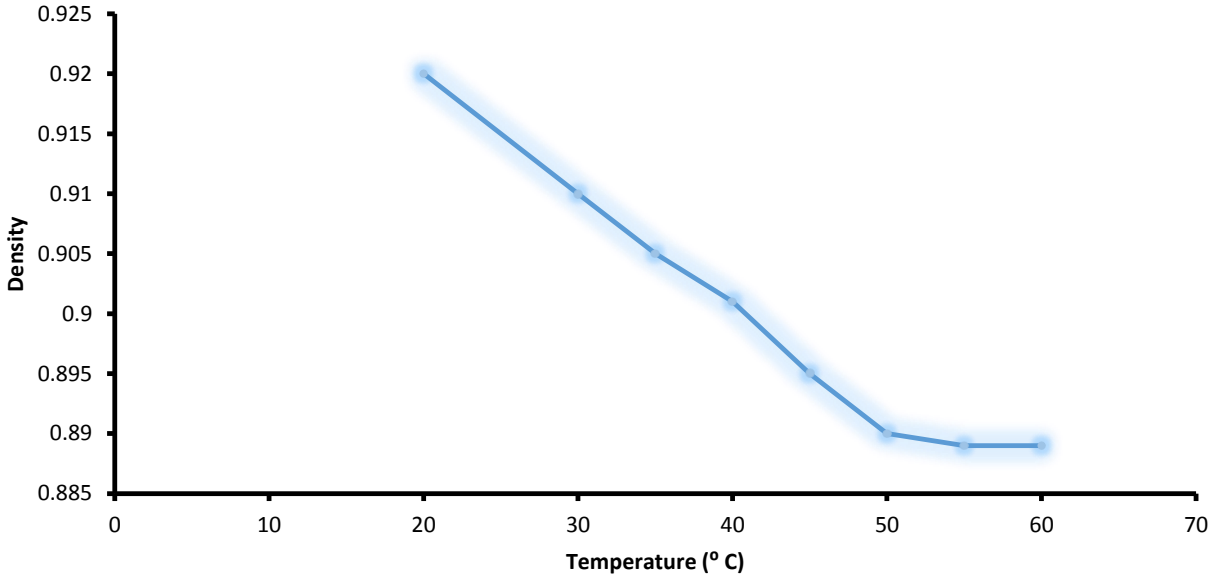


Figure 5. 3: Density vs temperature Graph

In addition, Figure 5.4 is a typical viscosity curve at container temperature. It shows the *ricinus communis* oil viscosity at various temperature.

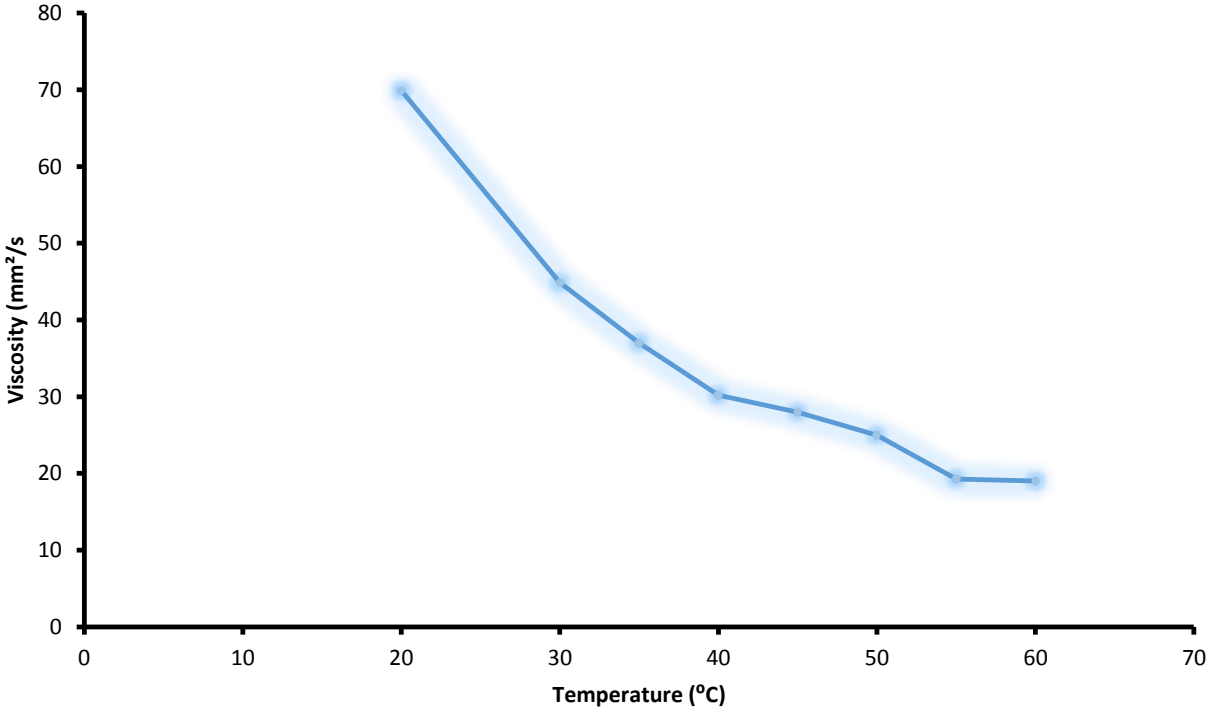


Figure 5. 4: Plot of *ricinus communis* oil viscosity as function of the container temperature

5.4 The acidity index

The acidity index of a fatty substance was determined by neutralising its free acidity with a certain amount of potash. The amount of potassium hydroxide (KOH) required to neutralise one gram of oil was weighed in mg. It was called the degree of acidity and is expressed by the parameters of the following formula:

$$A_i = \frac{V \times M \times N}{10 \times m} \times [\%] \quad (5.15)$$

Where:

A_i: The oil acidity index

M: molar mass of the oleic acid (282.47 g/mol)

m: mass in gram of the oil sample

N: Normality of the KOH solution used

During the experiment, the following proportions were used:

N=0.5 m=10g V(KOH)=10.80 ml

We found $A_i = 15.253\ mg$

By hypothesis this indicated that 15.253 mg of KOH were needed in order to neutralise 1000 mg of oil. The acidity was expressed in % of the major acid which was the ricinoleic acid in the case of *ricinus communis* oil.

5.5 Flash point

In the liquid considered, vapours appeared at a certain temperature. Therefore, vapours combined with a fuel (hydrogen for example), produced a deflagration. A simple definition would be the lowest temperature where a liquid can form a flammable mixture at its surface. There exist several devices that have been developed by manufacturers to measure flash point. The most used instrument is the Pensky Martens [191]. However, with our limited resources, it was not possible to determine that. The principle was long and necessitated to have a nomenclature of the instrument, which we will not detail here. But, after determining the temperature with the instrument, it must be corrected with the following relation:

$$T_c = T_0 - \frac{(760 - D)}{30} \quad (5.16)$$

With:

T_c: Corrected value of the temperature

T_0: Value of the temperature measured with the instrument

$D_{[mm\ of\ Hg]}$: Density during the tests.

According to the literature, the Ricinus Communis oil flash point is between $229\ ℃ - 448\ ℃$. The flash point high values it possesses show good characteristics for the storage and that one can plan its secure transportation as well as its secure manipulation compared to the gasoil.

5.6 The saponification index

The saponification reaction consists of doing the hydrolysis of an ester. In terms of vegetable oils, this reaction gives the formation of soap as a result. The saponification index is the number of milligrams of potash caustic (KOH) needed in order to transform fatty acids and one gram of triglycerides into soap.

Principle

An excess of potassium hydroxide was taken. It was heated with the oil sample that had a known mass until complete saponification. Then the excess of alkaline was titrated with an acid solution. The reaction happened according to the following equation:

$$Oil + \text{Potassium Hydroxide} \rightarrow \text{Glycerol} + \text{Soap} \qquad [192]$$

The saponification index had the following formula:

$$S_I = \frac{(V_0 - V) \times 56.1 \times N}{W} \qquad (5.17)$$

Where: V_0 is the volume of the hydrochloric acid solution at 0.5N for the blank titration (in ml); V is the volume of the hydrochloric acid solution at 0.5N for the titration of the sample (in ml); W is the weight of the sample (in mg); N is the Normality of the hydrochloric solution; S_I is the saponification index (mg of KOH/g). However, even though the saponification index was obtained after a series of experiments, the author restrained himself to one experiment because of the time constraint.

During the single experiment, the following propositions were used:

N=0.5 $\quad V_0 = 24.40\ ml \quad V = 10.5\ ml \quad W = 2.1\ g$

Therefore, by (5.17), $S_I = 185.664$ mg.

Analogically, in the case of acidity, results indicated that 185.664 mg of KOH were needed in order to perform saponification of 1000 mg of oil.

5.7 Protocol of transesterification

5.7.1 Vegetable oils and fatty esters

In a catalysed reaction with alcohol, vegetable oils and animal acids produced a correspondent alkyl ester. This reaction is known as transesterification. It completely changed the chemical composition of oil that it transformed into ester. During the reaction, vegetable oil reacted with methanol molecules (or ethanol) in order to form ethyl monoesters (or ethyl) and glycerol. Thus, oil proprieties were changed and it became comparable to a diesel combustible: that was biofuel. Further to that, the transesterification reaction contributed to the refining of *ricinus communis* oil as illustrated in Figure 5.5:

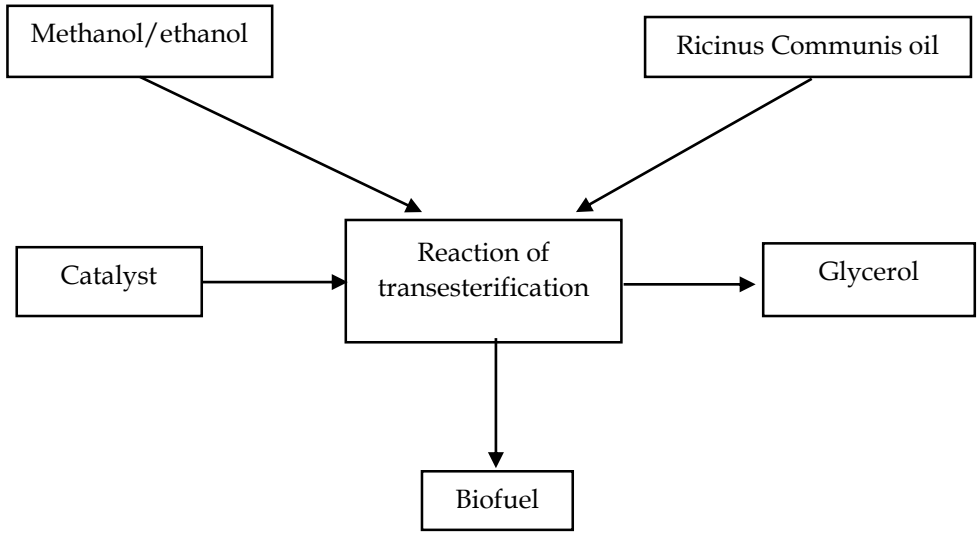


Figure 5. 5: Catalysed reaction of transesterification of Ricinus Communis oil

5.7.2 Separation phase

Transesterification was carried out using different masses of castor oil. Sodium metasilicate catalyst was dissolved in 30 g of methanol by slowly heating and agitating the mixture. Thereafter, the solution obtained was added to oil in order to vigorously agitate the mixture that was maintained to 40, 45, 50, 55, 60, 65 and 70°C for different period of time from 1 hour. After the agitation, the reaction was stopped and the mixture was put into a separating funnel for a period of at least half a day. The ester phase was decanted from the balance of the mixture.

5.7.3 Washing

The excess alcohol and catalyst residues were removed from the ester phase by slowly pouring water into the separating funnel. This operation was delicate and had to be done slowly with the least agitation since agitation caused the formation of an emulsion which reduced the performance of the synthesis. In that case, absorbent was added to biodiesel for a period of six hours, thereafter biodiesel was filtrated.

5.7.4 Calculation of the conversion rate

Chemically, the transesterification conversion rate was written under the following form [193]:

$$X = \frac{n_h^0 - n_h}{n_h^0} \tag{5.18}$$

And the number of moles of triglycerides at time t was defined by:

$$n_h = n_h^0 + \xi v_h \tag{5.19}$$

The number of moles of esters at time t is defined by:

$$n_{es} = n_{es}^0 + \xi v_{es} \tag{5.20}$$

With:

ξ being the degree of progress of the reaction, v_h and v_{es} were the stoichiometric coefficients of triglycerides and esters produced. Those parameters were deduced from the transesterification chemical reaction where they took the values of 1 for triglyceride and 3 for ester respectively. Equations (5.18) and (5.19) enabled the deduction of the following reaction:

$$\xi = -\frac{\mathrm{X n_h^0}}{\mathrm{v_h}} \tag{5.21}$$

Substituting the expression of ξ into (5.20), and considering that at time t_0 the number of moles of esters is null, the conversion rate of the reaction will be written as follows:

$$\mathrm{X} = \frac{\mathrm{n_{nes}}}{\mathrm{3n_h^0}} \tag{5.22}$$

Or:

$$\mathrm{X(\%)} = 100\left[\frac{\mathrm{m_{es}/M_{es}}}{\mathrm{3m_h/M_h}}\right] \tag{5.23}$$

Considering the results of the *ricinus communis* ester basic analysis, the molar mass of the M_{es} ester was $927 g/mol$. Because of that, there was a need to determine the ester mass m_{es} at the end of the experiment, as well as the oil mass $m_h$used during the experiment in order to calculate conversion rates. The transesterification reaction was performed within two different durations of agitation without changing other operational conditions (molar ratio and temperature of reaction). During the first experiment, the duration of the reaction was set to 1 hour and during the second experiment, it was set to 4 hours. The calculation of the performance of the reaction of each experiment indicated that 1-hour duration was not enough to convert all triglycerides into ester. Meanwhile, 4-hours duration enabled a conversion performance of 95.2%, which was aligned with the optimal performance found in the literature.

The following Table 5.4 shows the values of performance of transesterification.

Table 5. 4: Performance values of the transesterification reaction

	$\boldsymbol{m_{es}(g)}$	$\boldsymbol{m_h(g)}$	t(hours)	X (%)
Test 1	47	74.79	1	60.5
Test 2	72.46	74.79	4	95.2

Table 5.5 compares the mass composition of biodiesel using ricinus communis oil and the mass composition of biodiesel from other feedstocks.

Table 5. 5: Mass composition of *ricinus communis* biodiesel and other biodiesel

	C (%)	***H*** (%)	O (%)	N (%)	S (%)
Soy	77.0	12.18	10.82	0	0
Ricinus Communis	77.45	12.23	10.32	0	0

The basic analysis of *ricinus communis* ester obtained gave the mass composition of carbon, azote, hydrogen and sulfur (Table 5.5). Results indicated that biodiesel kept the same C, H and O compositions than crude oil.

Results displayed in Tables 5.4 and 5.5 provided the basis for comparison of results published in other studies as shown in Table 5.6

Table 5. 6: Comparison of results with other published studies

	Element	Fraction	Atomic mass	Nbr. Of atoms	Formula
Gasoil	C	86	12	7.17	$C_{10.65}H_{20.8}$
	H	14	1	14	
Ricinus Communis	C	77.45	10.32	14	0
	H	12.23			
	O	10.32			

Chapter 6: In-lab biofuel production from *ricinus communis* oil and mathematical interpretation

6.1 History

The process of making combustible from biomass used as raw material in 1800 is essentially the same than the one used today. The history of biofuel is more political and economic than technological [193]. At the beginning of the 20th century, massive amount of fuel cars were introduced. Since then, petroleum companies have been obliged to refine a lot of crude oil destined to fuel the fuel supply of those vehicles.

Today, the exhaustion of fossil energies is announced because of this massive exploitation which has generated the emergence of the use of vegetable oil as fuel. The production of fuel from vegetable oil is not a new process. As seen in previous chapters, the transformation of vegetable oils or animal fats into mono alkyl ester or biofuel is called the transesterification reaction. This process is known since the end of the 19th century [194].

In 1853, J. Patrick and E. Duffy had performed a transesterification reaction [195]. In 1893, the life of the Diesel engine started when the famous German inventor Rudolph Diesel published a document titled “the theory and construction of a rational heat engine” [196]. The document described a revolutionary engine in which the air was compressed by a very high pressure piston causing a very high temperature. Dr. Rudolph Diesel designed the first Diesel engine able to work with vegetable oil [196]. In 1900, during the world exposition in Paris, a small Diesel engine that was working with peanut oil was shown. The engine was working so well that the change of fuel was not noticed by few visitors. Due to the high temperature created, the engine was able to work with a variety of vegetable oils.

In 1911, during the universal celebration of Paris, Diesel also made his engine work with peanut oil [197] and declared: “the diesel engine can be fuelled with vegetable oils and will contribute to the development of Agriculture in the countries that will use it”.

The term “biofuel” was given to vegetable oil that went through transesterification in order to bring its characteristics close to the diesel fuel [198]. Thus, the use of vegetable oil as alternatives of renewable fuel, in concurrence to oil, was proposed at the beginning of 1980, but the production only started in the 1990s.

6.2 Materials used

Essentially, the following materials were used:

- ✓ Pyrex glass reactor 1000 ml
- ✓ Magnetic stirrer bars
- ✓ Mercury thermometer
- ✓ Separation funnel
- ✓ Condensers

6.3 Proportions of reagents

In order to perform the following experiments, oil with the following properties was randomly selected: Volume used: 600 ml; Density: 0.95 g/ml

Knowing that:

$$\rho = \frac{m}{V} \Longrightarrow m_{oil} = \rho \times V = 600 \times 0.95 = 570g \qquad [199]$$

Methanol in excess (molar ratio 5:1) was used, meaning 5 moles of methanol for one of oil. Thus:

Molar mass of methanol: $M_{methanol} = 32.04\ g/mol$

$$n_{methanol} = r.n_{oil} = 5 \times \frac{570}{887} = 3.213\ moles \Longrightarrow m_{methanol} = 102.944\ g$$

$$Density_{methanol} = \frac{mass_{methanol}}{volume_{methanol}} = \frac{102.944}{792} \times 1000 = 129.979\ ml$$

For the catalyst (Sodium Hydroxide) 1.5% of the mass of oil was used, which represented 8.55 g. Table 6.1 resumes quantities for all products:

Table 6. 1: Quantities of various reagents

Oil	Methanol	NaOH
570 g	65.71g (33% excess)	1.5% of the oil weight
	102.944g (66% excess)	1.5% of the oil weight

6.4 Operations

The process is described as follows:

- ✓ After the determination of various reagents proportions (oil, methanol, etc), the reaction mixture was raised to a temperature between $55℃ - 70℃$ in the reactor, which corresponded approximately to the temperature of ebullition of the methanol.
- ✓ The temperature of mixture was maintained within the range in order to avoid the evaporation of alcohol during the reaction; and the excess was evacuated.
- ✓ After the complete evacuation of alcohol in excess, the temperature was stabilised at a constant value of $67℃$.

Therefore, the reaction ended after 4 hours.

6.4.1 Reaction time effect

The duration of the transesterification reaction of *ricinus communis* oil was very long, even though, the formation of ester only lasted few minutes. Thus, the reaction was not economic because the duration (4 hours) and the cost of energy necessary for the heat of the reaction mixture were high.

6.4.2 Effect of reaction temperature on transesterification

The temperature of the reaction medium was closely monitored because it turned around the boiling point of methanol ($55 - 70℃$). However, for energy purpose, researchers supported that the reaction of transesterification could be done at the ambient temperature but could take more time. That's the reason it was judicious to find a compromise between the benefit of energy and the cost for heating.

6.4.3 The catalyst effect

Catalyst was one of the most important factors during the reaction; it increased its kinetic. Sodium hydroxide (1.5% of the mass of oil used) was used because it had the advantage of being cheap and was widely used in alcoholysis works.

6.4.4 The separation of phases

The success of a reaction of transesterification is materialised by the presence of two phases. That is why after the reaction, there was a mixture of esters with glycerol inside the reactor. And it was crucial to stop the agitation during the separation. Glycerol which had a higher density than ester was located below the funnel after several hours of decanting [200], [201].

Figure 6.1 shows the separation between ester and glycerol. The rate of separation of ester and glycerol was closely monitored.

Figure 6. 1: Separation between ester and glycerol after the reaction

6.4.5 The rinsing process

Ester obtained was washed in order to eliminate excess alcohol and excess of catalyst. Biofuel was put in a funnel and water was slowly poured for rinsing (about 100 ml). This phase was carefully performed as it required a lot of precautions including less agitation. In fact, any agitation could lead to emulsions that would had negatively affected the performance of the synthesis because the formation of emulsions can represent 15 to 20% of the mass of ester. Thus, decanting carried on for about 24 hours, and biofuel was recovered. The residual alcohol and catalyst dissolved during the aqueous phase.

6.5 *Ricinus communis* oil reserves

The set of oil containers available for biofuel production was called ultimate resources of a storage. In the following sections, the term "size" was used to designate the quantity (or resources) of *ricinus communis* oil in a container. The model suggested in the previous chapter justified the use of the Pareto-Levy law in order to represent the quantities of *ricinus communis* oil stored in the containers. During the experiment, it appeared convenient to use the "sample" point of view at that stage. During simulation, three different platforms were used: Platform A, platform B and platform C. Then, it was assumed that in platform B, the lower level of the Pareto-Levy distribution reserves was set to 1Mb.

However, this didn't mean the negligence of containers of size lower than 1Mb. When the level was set to ε, it was possible to go back to the standard distribution $Par(\alpha)$ provided that the sizes of the containers were divided by the minimal level ε because the Pareto-Levy distributions were invariant by change of scale. In terms of the estimation of the Pareto-Levy distribution α parameter, Lopez works [202], [203] were used in order to completely address this question and there was no need for supplementary mathematical work. Because results were based on this method, a short summary of the fundamental principles of these works is presented. According to Lopez [202], [203] the population of containers extracted at date t^* of the present was the result of a draw without replacement and biased by the size of the storages within the complete population of all storages known or not within the platform. More precisely, storage of reserves X was extracted or not at t^* following the value of a variable of censure ε of the Bernoulli law of ω(X) parameter, conditionally to X. And storage i was extracted at date t^* if and only if $\varepsilon_i = 1$. Also,

$$\begin{cases} \mathbb{P}(\varepsilon = 1|X) = \omega(\mathrm{X}) \\ \mathbb{P}(\varepsilon = 0|X) = 1 - \omega(\mathrm{X}) \end{cases} \tag{6.1}$$

Where ω was the probability of inclusion of the storage in the population of storages extracted at date t^*. This model is called "static" because the only information it brought about the platform was related to the present time. If the law of sizes of the storages was a Pareto-Levy distribution, then the sizes-log sample was an exponential distribution of α parameter and the sample $(Y_1, \dots, Y_n)$ of sizes-log of the storages extracted at date t^* was i.i.d of density.

$$q_m(y) = \exp\left(\sum_{I \in m}(\theta_1 + \log\alpha - \alpha y)1_{I(y)}\right) \tag{6.2}$$

Where m was a partition that defined the storage classes of the sizes, and θ_I was expressed based on the bias of the draw on class I. Likelihood equations that were derived from (6.2) were solved in $(\theta_I)_{I \in m}$ and in α, in an explicit way.

In order to choose the best ω function, the author considered a collection of models corresponding to a collection M of partitions. In order to select a model, meaning a partition m, Lopez referred to works performed on the selection of exponential models. Therefore, there was a procedure to estimate the α parameter and the biased function ω. In what follows, it was possible to assume that the index α of the Pareto-Levy distribution was known. Table 6.2 gives the assumed values corresponding to the reserves of the three platforms. Results indicated that the habitat in platform B was more concentrated than that of the two other platforms.

Table 6. 2: Pareto-Levy parameters for 3 platforms (Values are set by convention).

Platforms	$\hat{\alpha}$
Platform A	0.75
Platform B	0.92
Platform C	0.90

6.5.1 Results of the simulations performed

A. Influence of temperature on the transesterification reaction

Figures 6.2, 6.3, 6.4, 6.5, 6.6 and 6.7 show the influence of temperature on the transesterification process: The best performance was obtained in the range between 50°C and 90°C, with an optimum around 60°C, which corresponded well to most experimental results found in the literature.

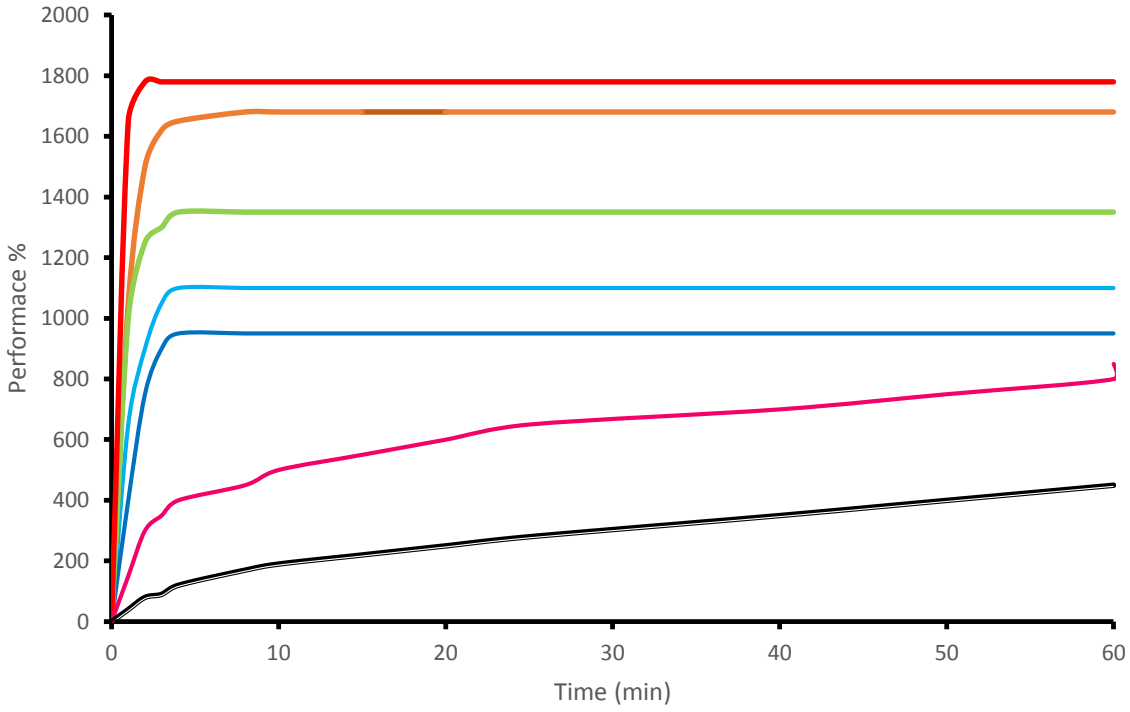


Figure 6. 2: Influence of the temperature on the transesterification reaction

Legend of Figure 6.2

Lines	T (°C)	R_60 (%)	R(%)
	10	25	76
	30	68	80
	50	84	84
	60	85	85
	70	86	86
	90	89	89
	235	93	93

Figure 6.2 indicates better the influence of temperature on transesterification. The best performances were obtained between 125°C and 225°C, with an optimum around 150°C. This corresponded well to most experimental results found in the literature [134], [135].

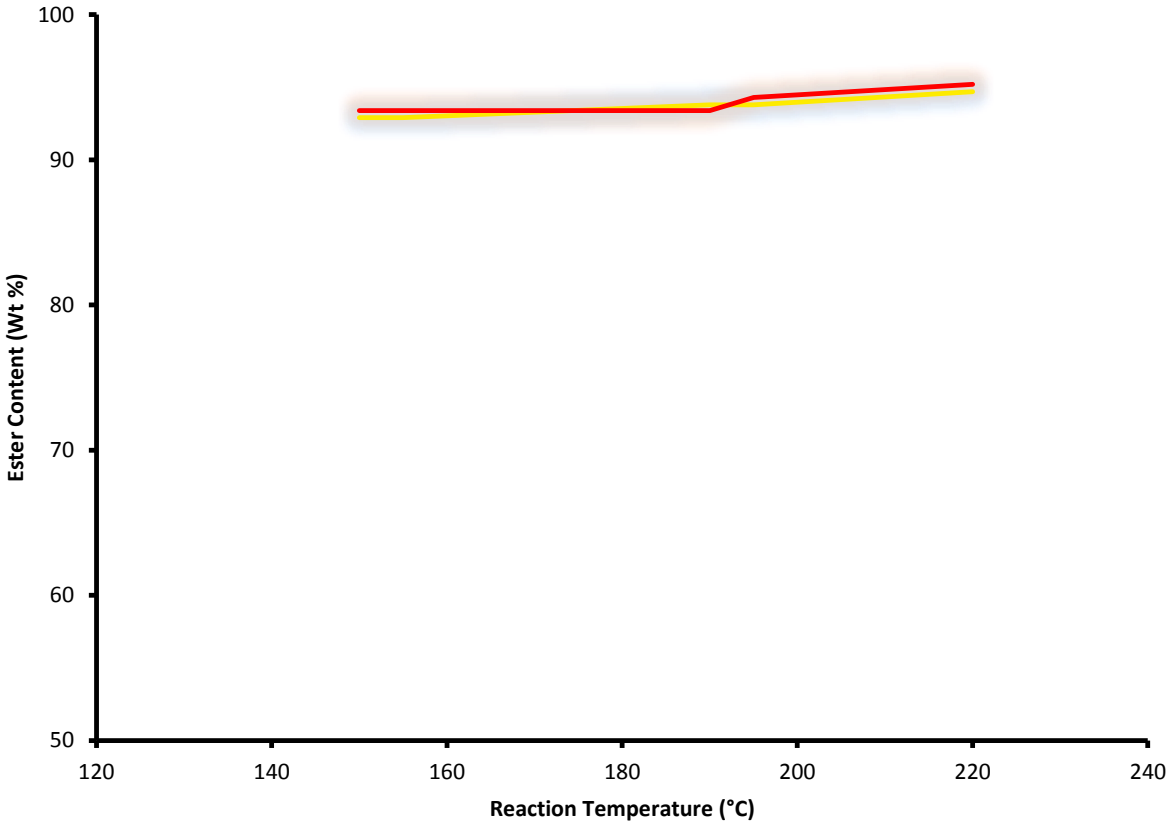


Figure 6.3: Comparison of ester content during the reaction temperature of the transesterification reaction model with the literature

Like in the literature, with the model, when the temperature was high, it implied a high ester content. As shown in figure 6.3, there was a slight difference between the plot of the curve resulting from our model and the curve in the literature, which validated our model. The model also indicated that when the temperature was increased between 20 ºC-40 ºC, the ester content also increased by one unit.

6.5.2 MeOH/VO molar ratio effect on the transesterification reaction

One of the most influential variables on the formation of ester was the methanol-oil molar ratio. The following figure indicates the effect of the MeOH/VO molar ratio on the transesterification reaction. Finding an optimum value of this ratio enabled the reduction of the cost of production while increasing productivity. The recommended optimal value for this ratio was of the order of 6.

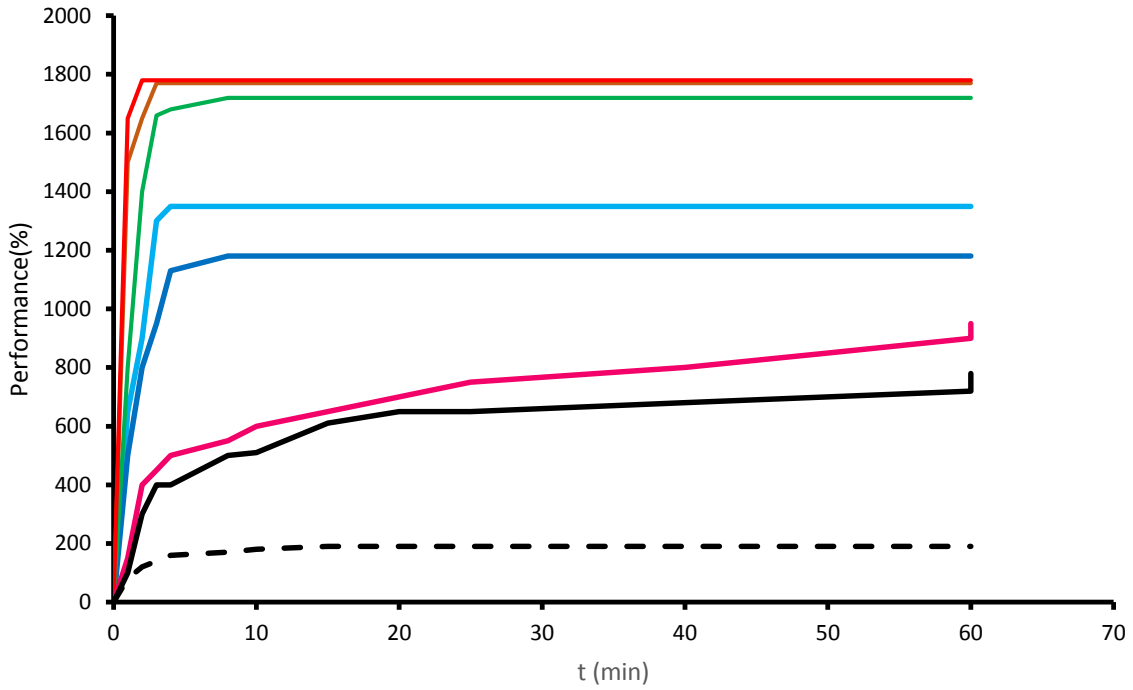


Figure 6. 4: Influence of the n molar ratio on the transesterification reaction

Legend of Figure 6.4

Lines	n	R_60 (%)	R(%)
- - -	1	22	23
———	3	57	57
———	5	78	78
———	6	85	85
———	7	90	90
———	12	98	98
———	21	99.4	99.4
———	30	99.7	99.7

6.5.3 Kinetic characteristics of the reaction based on the model

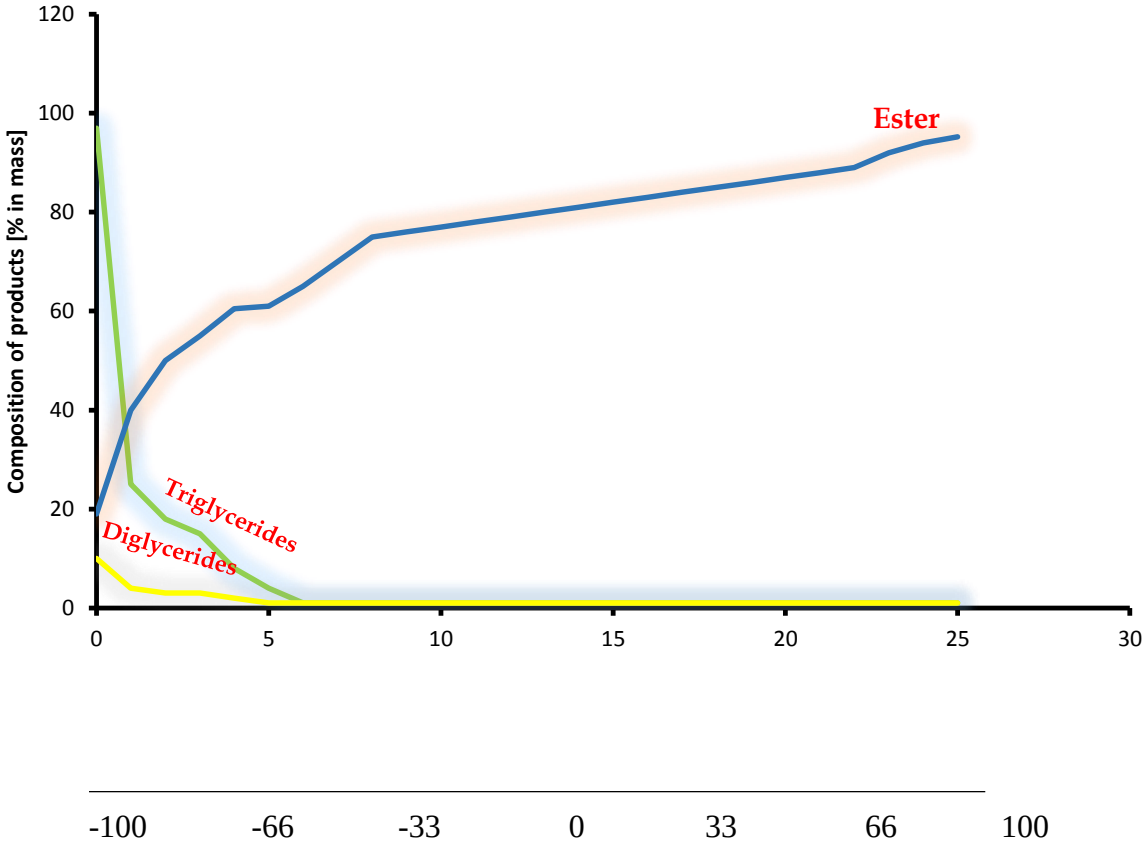


Figure 6. 5: Composition of products during the transesterification reaction

In chemistry, the stoichiometry of the reaction would require 6 methanol moles for 1 *ricinus communis* oil mole (6:1 ratio) in order to obtain 6 esters moles of fatty acid and 1 glycerol mole. In this case, the performance would be less as this was indicated in the ester curve of Figure 6.5, compared to the performance obtained with the 24:1 ratio that was used and which corresponded to 66% of excess of methanol.

Under the model, increasing the methanol/*ricinus communis* oil molar ratio from 6:1 to 24:1, ester content was increased from 60.5% to 95.2%. This indicated that our model provided a fairly better performance/result than results observed in previous works as this can be observed in the following figure that combines the ester content and the ester content found after applying our model.

Observations of figure 6.6 indicated that the two curves are similar under experimental conditions. In fact, at the beginning of the experiment, both 3:1 and 4:1 molar ratios (33% of excess in methanol) were used.

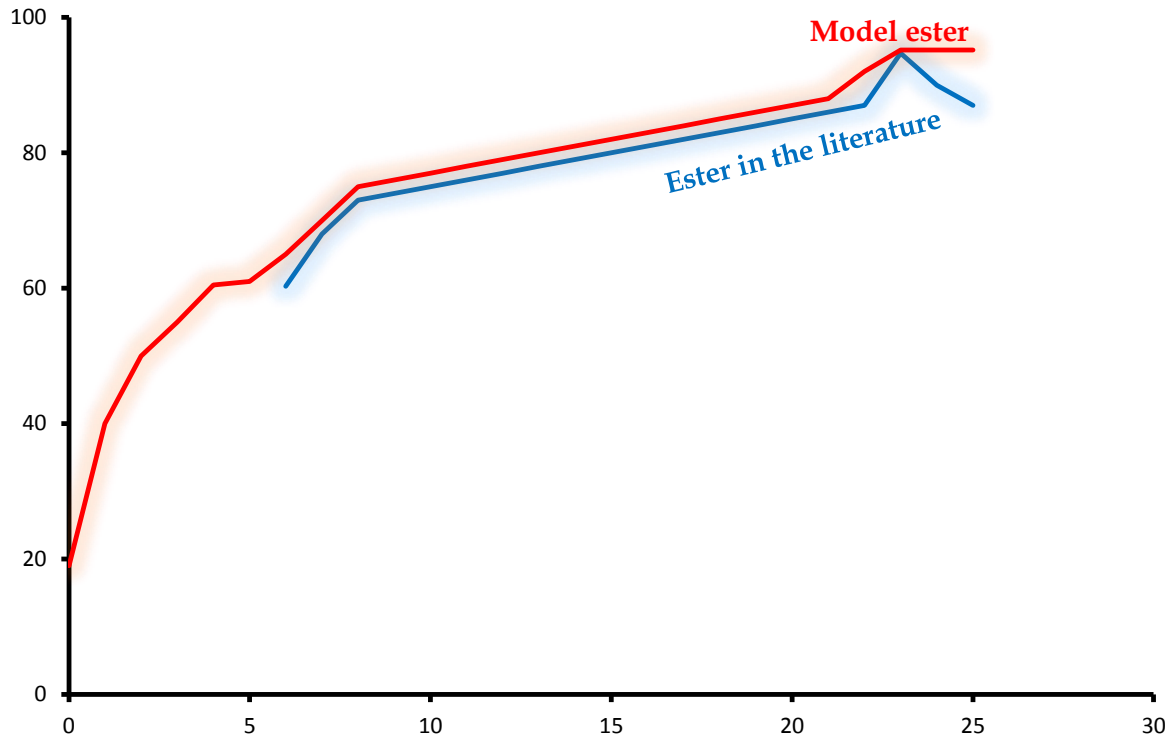


Figure 6. 6: Comparison of ester produced during the transesterification reaction model with the literature

Other experiments were performed with 5:1 molar ratio (66% of excess in methanol) and so on until the 24:1 and 25:1 molar ratios. From the excess of alcohol, a 90-95.2% conversion was achieved, which was good. It appeared clear that a 3:1 and less theoretical ratio would have a consequence of giving a lower performance and would make intermediaries lower products such as diglycerides and triglycerides. This would make the conversion incomplete.

The Triglyceride/catalyst molar ratio effect on the transesterification reaction

Figure 6.7 indicates the triglyceride/catalyst molar ratio effect on the transesterification reaction. The catalyst content was a factor that had a positive impact on the reaction speed meanwhile this impact was negative on the performance. This was due to the fact that, when there are more foundations, a more important saponification was obtained. The reaction performance increased with the ratio p, until when it reached an optimum with an approximate value of p=0.0785 (or a MeOH=0.5% of the VO mass), then it fell after this value because of the saponification.

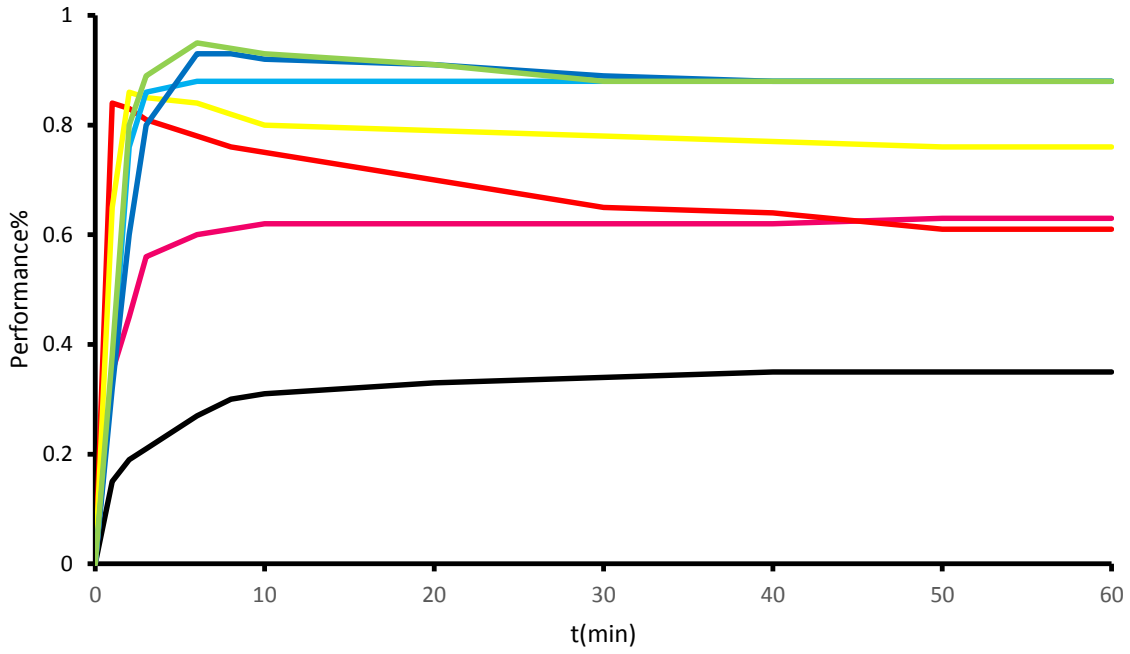


Figure 6. 7: Molar ratio p on the transesterification reaction

—— : $\boldsymbol{m_{MeOH} = 0.1\%}$ or $\boldsymbol{p = 0.0157}$

—— : $\boldsymbol{m_{MeOH} = 0.2\%}$ or $\boldsymbol{p = 0.0314}$

—— : $\boldsymbol{m_{MeOH} = 5\%}$ or $\boldsymbol{p = 0.7849}$

—— : $\boldsymbol{m_{MeOH} = 0.4\%}$ or $\boldsymbol{p = 0.0628}$

—— : $\boldsymbol{m_{MeOH} = 0.5\%}$ or $\boldsymbol{p = 0.0785}$

—— : $\boldsymbol{m_{MeOH} = 0.6\%}$ or $\boldsymbol{p = 0.0942}$

—— : $\boldsymbol{m_{MeOH} = 2\%}$ or $\boldsymbol{p = 0.314}$

6.6 Experimental model: Modelling of the *ricinus communis* oil extraction

In this section, a model was formulated to describe the *ricinus communis* oil extraction and a mathematical model was deduced for the dates of extraction that happened before the present day. The difference with the model of extraction introduced in the previous chapter is that the following model can be applied to real biofuel platforms in order to suggest the extensions of the production processes observed. Nevertheless, from the model presented in the previous chapter, it can be attested that the distribution of the time of extraction for a specific container can be conditionally defined by the size X of the container in question. More precisely, the dates of extraction were independent between them and of the same conditional law:

$$(D|X) \sim E(h(X)) \tag{6.3}$$

Where h is an increasing function of $\mathbb{R}^+$ in $\mathbb{R}^+$called function of visibility and that was needed to be determined.

In addition, it was possible to verify that this new model was coherent with the one proposed by Lopez [202], [203]. By using the notations of the previous section, a container was observed if its variable of censure ε verified $\varepsilon = 1$, which corresponded to $D \leq t^*$. Thus, $\varepsilon = 1_{D \leq t^*}$ and $\omega(x) = 1 - \exp(-h(x)\, t^*)$.

In reality, containers of small size were less visible, which enabled the introduction of a simplified model of the dynamic of successive extractions for this particular category of containers. This simplification was based on the model of the Poisson approximation (6.3). Which led to the following stratifying model: The small containers were extracted according to an homogeneous process of intensity μ_0. Let $I_0 = [1, x_0]$ be the class of corresponding size. For a container of size $X \geq x_0$, its time of extraction had the conditional distribution $(D|X) \sim E(h(X))$, where h was an affine and increasing function by pieces.

6.7 Mathematical interpretation of containers with small amount of oil.

The I_0 class corresponded to the set of small storages of sizes in the $[1, x_0]$ interval. Even when those storages contained *ricinus communis* oil, their extraction was considered as a failure in terms of biofuel production. These "failed" extraction happened over time with a constant rate r. It was assumed that the next extraction performed within the platform had to be modelled by an homogeneous process of binomial law of intensity λ.

Then, the dates of extraction in the I_0 category also formed an homogeneous process of Poisson of intensity $\mu_0 = r\lambda$. According to this simple presentation, it was natural to use the homogeneous Poisson process for the model of the extraction of this category of containers. Let's note that this reasoning could be defaulted if there was a risk of exhaustion of the reserves of this class. In the platform B, it was assumed that there were at least one thousand containers of sizes less than 20Mb. Assuming that about 6 per year were extracted, thus, there was no risk of exhaustion at the scale of tens of years.

Therefore, in order to justify the model, it was demonstrated that the conditional law defined in (6.3) was a limited situation of its dynamic. Firstly, because the size of containers in this category was approaching the minimal level, their visibility was very weak too. Inside this class, it was assumed that all containers possessed the same visibility h_0. Therefore, containers of this category were in big numbers within the platform.

Let N_0 be the total number of containers of this category in the platform. According to the model in (6.3), the number $n_0(t)$ of containers of this class extracted before the date t

followed a binomial law with n_0 and $\omega(t) = 1 - \exp(-h_0(t)) \approx h_0(t)$. When $h_0(t)$ was small, then $\omega(t)$ was also small. And by a classical approximation, it was possible to assume that $n_0(t)$ followed a binomial distribution of parameter $N_0 h_0(t)$.

Models formulation

Let S_m be the set of densities g^*defined from a partition m of $[x_0, x_{max}]$.

$$S_m = \left\{ g^* : (x,t) \longmapsto \alpha \frac{h(x) \exp\{-h(x) t\, x^{-\alpha-1}\}}{P_{dec}(\alpha, t^*, h)\, x_0^{-\alpha} - x_k^{-\alpha}} 1_{0 \le t \le t^*, x_1 \le x \le x_k}, \alpha > 0, h \in \mathcal{H}_m \right\}$$

In fact, there was a possibility to construct an estimator of the likelihood maximum that was based on both the $X_i^{(p)}$and the X_i^*. A simple alternative consisted of an explicit resolution of a system of the likelihood equation which renders the estimation more reliable. Therefore, the following sets were considered:

$$S_m(\alpha) = \left\{ g^* : (x,t) \longmapsto \alpha \frac{h(x) \exp\{-h(x) t\, x^{-\alpha-1}\}}{P_{dec}(\alpha, t^*, h)\, x_0^{-\alpha} - x_k^{-\alpha}} 1_{0 \le t \le t^*, x_1 \le x \le x_k}, \alpha > 0, h \in \mathcal{H}_m \right\}$$

That was abusively wrote S_m.

Proposition

The S_m model was identifiable with regards to the parameterisation in h and α.

Demonstration

The demonstration consisted of the parameterisation of the model that was injective. It was supposed that t^* was fixed as well as a partition m of $[x_0, x_{max}]$ of size k. Let α_1 and α_2 be two positive real numbers and let h_1 and h_2 be two visibility functions of the set $\mathcal{H}_m$ such that for all $x \in [x_0, x_{max}]$ and for all $t \in [0, t^*]$,

$$\alpha_1 \frac{h_1(x) \exp\{-h_1(x) t\}}{P_{dec}(\alpha_1, t^*, h_1)} \frac{x^{-\alpha_1 - 1}}{x_0^{-\alpha_1} - x_{max}^{-\alpha_1}} = \alpha_2 \frac{h_2(x) \exp\{-h_2(x) t\}}{P_{dec}(\alpha_2, t^*, h_2)} \frac{x^{-\alpha_2 - 1}}{x_0^{-\alpha_2} - x_{max}^{-\alpha_2}} \qquad (6.4)$$

For t=0, and for $x \in [x_0, x_1]$,which gave:

$$\alpha_1 \frac{a_j^{(1)} x + b_j^{(1)}}{P_{dec}(\alpha_1, t^*, h_1)} \frac{x^{-\alpha_1 - 1}}{x_0^{-\alpha_1} - x_{max}^{-\alpha_1}} = \alpha_2 \frac{a_j^{(2)} x + b_j^{(2)}}{P_{dec}(\alpha_2, t^*, h_2)} \frac{x^{-\alpha_2 - 1}}{x_0^{-\alpha_2} - x_{max}^{-\alpha_2}}$$

The identification of power functions implied that $\alpha_1 = \alpha_2 = \alpha$. For $x \in [x_0, x_{max}]$ and for $t \in [0, t^*]$, the identity (6.4) could simplify itself under the following form:

$$\frac{h_1(x)\exp\{-h_1(x)t\}}{P_{dec}(\alpha_1,t^*,h_1)}\frac{x^{-\alpha_1-1}}{x_0^{-\alpha_1}-x_{max}^{-\alpha_1}}=\alpha\frac{h_2(x)\exp\{-h_2(x)t\}}{P_{dec}(\alpha_2,t^*,h_2)}\frac{x^{-\alpha_2-1}}{x_0^{-\alpha_2}-x_{max}^{-\alpha_2}}$$

The two exponential terms as functions of t had necessarily the same exponent, which gave $h_1(x)=h_2(x)$. And this was true for all $x\in[x_0,x_{max}]$, then, $h_1=h_2$. It was seen that $h\in\mathcal{H}_m$ only admitted one notation, which allowed the conclusion.

Form of the partitions considered

This section explains why the level x_0 was not included in the procedure of selection of a partition. Even when it meant the consideration of the joint law of complete sample:

$$\left(\left(X_1^{(p)},D_1^{(p)}\right),\ldots,\left(X_{n_0}^{(p)},D_{n_0}^{(p)}\right),(X_1^*,D_1^*),(X_n^*,D_n^*)\right)$$

A family of partitions $\widetilde{m}$ of $[1,x_{max}]$ that enabled the variation of x_0 was suggested. The optimal level was selected in the sense of a statistical criteria. However, one of the main objectives was to introduce the extensions of the extraction process. But it was indicated that the binomial approximation presented in the previous section was only valid when the class of small containers was not exhausted. Therefore, there was a danger that the procedure of selection of the model chose a level x_0 such that the binomial approximation was no more valid in the future. That is why it was preferable that the level x_0 was chosen once and for all users. Therefore, the challenge was to choose the best partition m of the set $[1,x_{max}]$ in order to better estimate the dynamic of extraction applied to it. Also, the mathematical process of this problem was performed in the next chapter.

Proposition

Let h_l be a suite of positive reals that approached 0 when l approached infinity. For all l, let Z_l be a random variable such that $Z_l\sim\varepsilon(h_l)$. Then, for all >0, $(Z_l|Z_l\leq t)\ \underset{l\to+\infty}{\mathcal{L}}\ U$, where $U\sim u([0t])$.

Demonstration

Let $F_{t,l}$ be the repartition function of $(Z_l|Z_l\leq t)$. Then: $F_{t,l}(u):=\frac{1-\exp(-h_l u)}{1-\exp(-h_l t)}1_{0\leq u\leq t}$

Therefore, when $\to+\infty$, $F_{t,l}\to F_U(u):=\frac{u}{t}1_{0\leq u\leq t}$, where F_U was the repartition function of the random variable U.

Knowing $n_0(t)$, the dates of extractions of containers of class I_0 were upgraded before the date t could be modelled by an i.i.d. sample of uniform law in [0,t]. As long as the probability of inclusion remained low, $n_0(t)$ followed a binomial distribution of parameter $N_0 h_0(t)$. Therefore, it was justified to consider the suite of extraction times of small storages as the realisation of an homogeneous binomial process of intensity μ_0.

Let $\left(D_i^{(p)}\right)_{i\geq 1}$ be the suite of extractions. To each of these dates, the size of the extracted container was associated and was noted $X_i^{(p)}$. Random variables $X_i^{(p)}$ were i.i.d. from Pareto-Levy distribution law $Par(\alpha, 1, \ x_0)$ restrained to $[1, x_0]$. Also, the extraction process of small containers was modelled using the binomial process marked on $\mathbb{R}^+$, in which the marks law was the distribution $Par(\alpha, 1, x_0)$. The suite of extractions happened before the date of the present t^* corresponding to the restriction of this process on [0, t^*].

6.7.1 Selection of a model

Instead of focusing on the dimension of models to define an acceptable penalty, results obtained by Massart in this context were based on the notion of entropy that also enabled to give the measure of the size of a set [204], [205]. Here, some definitions were recalled before enouncing the general theorem that was used:

The standard $\left\|\sqrt{f}-\sqrt{g}\right\|_2$between two positive functions f and g of $\mathbb{L}_1$n is noted $d_H(f,g)$, where f and g are densities with regards to the measure of Lebesgue on $\mathbb{R}^Q$, $d_H^2(f,g)$ corresponding to the double of the squared Hellinger distance between f and g. In what follows, $d_H(f,g)$ will be called the Hellinger distance even when f and g are not densities [206].

$$\leq (1+\delta)^v \prod_{q=1}^{v}\left(1+\frac{\beta v}{\sigma_q^2}\right)$$

$$\leq (1+\delta)^v \left(1+\frac{\beta v}{\sigma_m^2}\right)^v$$

$$\leq (1+2\delta)^v$$

Therefore, $\frac{f(x)}{u(x)} \leq (1+2\delta)^{-\frac{v}{2}} \exp\left(\frac{\|\mu-\nu\|_2^2}{2\beta v}\right)$

6.7.2 $\mathcal{M}[LB_k]$ entropies

For the density mixture t of a $\mathcal{L}_{(K,v)}$ model, the matrices of covariance that intervened in the mixture t were all of the same volume. Therefore, it was not possible to decompose the set $\mathcal{L}_{(K,v)}$. The following $\mathcal{F}_{(K,v)}$sets were considered:

$$\mathcal{F}_{(K,v)} = \left\{ \left(\Phi(.|\mu_1, \lambda\Sigma_1), \dots, \Phi(.|\mu_k, \lambda\Sigma_k)\right); \begin{matrix} \sigma_M^{-2} \leq \lambda \leq \sigma_M^2, k = 1, \dots, K \\ \mu_k \in [-a, a]^v, \Sigma_k \in \Delta_{(v)}^1(\sigma_m^2, \sigma_M^2) \end{matrix} \right\}$$

Let l and u be two functions of $\mathbb{R}^v$ in $\mathbb{R}^K$ such that $l_k \leq u_k$for $k = 1 \dots K$, with the notations $l(x) = (l_1(x), \dots, l_k(x))$ and $u(x) = (u_1(x), \dots, u_k(x))$. The set $[l, u]$ was composed of all functions $f = (f_1, \dots, f_k)$ of $\mathbb{R}^v$ in $\mathbb{R}^K$ such that $l_k \leq f_k \leq u_k$ for k=1…K.

Therefore, the entropy number of $\mathcal{F}_{(K,v)}$ is the minimal cardinal of families of the form $[l, u]$ recovering the $\mathcal{F}_{(K,v)}$ space.

Proposition 1

Let $\sum \in \Delta_{(v)}^1(\sigma_M^{-2}, \sigma_M^2)$ and $\lambda \in [\sigma_m^2, \sigma_M^2]$. Then, $\exists\, J \in \{2, \dots, r\}^{v-1}$ and $z \in \{0, \dots, r'\}$ such that $\forall\, q \in [1, \dots, v]$, $\lambda_{z+1} B_{J,q}^l \leq \lambda \sum_{qq} \leq \lambda_z B_{J,q}^u$.

Demonstration

Let z be the unique natural number in $\{0, \dots, r'\}$ such that $\lambda_{z+1} < \lambda \leq \lambda_z$. By definition, there exists a vector J such that $B_{J,q}^l \leq \sum_{qq} \leq B_{J,q}^u$ for $q = 1, \dots, v-1$; which gave the wanted direction for the first v-1 coordinates. Because $\prod_{q=1}^v \sum_{qq} = 1$, then $\sum_{vv} = \prod_{q=1}^{v-1} \sum_{qq}^{-1}$. From this, it was deduced that:

$$\prod_{q=1}^{v-1} \left\{B_{J,q}^u\right\}^{-1} \leq \sum_{vv} \leq \prod_{q=1}^{v-1} \left\{B_{J,q}^l\right\}^{-1}$$

Which gave the said directive for the last coordinate. Covariance matrices associated with the couple (z, J) were useful because it helped for the definition of the extremity functions. However, the centering parameters also needed to be defined: For a fixed couple (z, J), the regular mesh of centering parameters was considered: $v^{(zJ)} = \left(v_1^{(zJ)}, \dots, v_v^{(zJ)}\right) \in [-a, a]^v$ such that for $\in \{1, \dots, v-1\}$, $v_q^{(zJ)} = (1+\delta)^{-\frac{j(q)+1}{4} - \frac{z}{2}} \sigma_M^2 \sqrt{c_1}\, \delta s_q$,

With $s_q \in \{-N_q, \dots, N_q\}$ where $N_q = \left[\frac{a(1+\delta)^{\frac{j(q)+1}{4}+\frac{z}{2}}}{\sqrt{c_1}\, \sigma_M^2\, \delta}\right]$, and,

With $s_v \in \{-N_v, \dots, N_v\}$ where $N_v = \left[\frac{a(1+\delta)^{\frac{v+z}{2}-\frac{S_J}{4}}}{\sqrt{C_1}\,\sigma_M^{2-v\delta}}\right]$, and

$$C_1 \coloneqq \frac{1-2^{-\frac{1}{4}}}{2}.$$

For a given couple (z,J), this guaranteed that for all $\mu \in [-a, a], \exists$ a vector $v^{(zJ)}$ of the mesh such that:

$$\left\{\sum_{q=1}^{v-1}\left(v_q^{(zJ)}-\mu_q\right)^2(1+\delta)^{\frac{j(q)+1}{2}+z}\,\sigma_M^{-4}\right\}+\left(v_v^{(zJ)}-\mu_v\right)^2(1+\delta)^{-\frac{S_J}{2}+v+z}\,\sigma_M^{2v-4} \le c_1 v\delta^2 \qquad (6.5)$$

Furthermore, for a couple (z, J), and a parameter $v^{(zJ)}$ of the above family, the following two functions were finally considered:

$$\begin{cases} l(x) = (1+\delta)^{-2v}\Phi\left(x|v^{(zJ)}, (1+\delta)^{-\frac{1}{4}}\lambda_{z+1}\,B_J^l\right) \\ u(x) = (1+\delta)^{2v}\Phi\left(x|v^{(zJ)}, (1+\delta)\,\lambda_z\,B_J^u\right) \end{cases} \qquad (6.6)$$

In addition, let $\in \left[\sigma_m'^2,\ \sigma_M'^2\right], \sum \in \Delta_{(v)}^1\left(\sigma_m'^2,\ \sigma_M'^2\right), \mu_k \in [-a, a]^v$, then:

$$\Phi \coloneqq \Phi(.|\mu, \lambda\textstyle\sum)$$

There existed λ_z and J such that for all $q \in \{1, \dots, v\}$,

$$\lambda_{z+1}\,B_{J,q}^l \le\ \lambda\textstyle\sum_{qq} \le\ \lambda_z\ B_{J,q}^u$$

There existed also a vector $v^{(zJ)}$ of the mesh associated to the couple (z, J) such that the constraint (6.5) was verified for the vector μ. For z, J and this vector $v^{(zJ)}$, u and l were considered as two functions defined like in (6.6).

Proposition 2

Under the previous hypotheses, for all $x \in \mathbb{R}^d$, $l(x) \le \Phi(x) \le u(x)$.

Demonstration

By applying proposition 2 that enables the increase of the gaussian densities quotients, then:

$$\frac{\Phi(x)}{u(x)} \le (1+\delta)^{-2v}\sqrt{\frac{\left|(1+\delta)\lambda_z\,B_J^u\right|}{|\lambda\textstyle\sum\ \ |}}\exp\left[\frac{1}{2}\left(v^{(zJ)}-\mu\right)'\left\{(1+\delta)\lambda_z\,B_J^u-\lambda\textstyle\sum\right\}^{-1}\left(v^{(zJ)}-\mu\right)\right]$$

$$\le (1+\delta)^{-\frac{v+1}{2}}\exp\left\{\frac{1}{2\delta}\left(v^{(zJ)}-\mu\right)'\left(\lambda_z\,B_J^u\right)^{-1}\left(v^{(zJ)}-\mu\right)\right\}$$

For $\Phi \le$ u, it is sufficient to have:

$$\left(v^{(zJ)}-\mu\right)'\left(\lambda_z\, B_J^u\right)^{-1}\left(v^{(zJ)}-\mu\right) \le 2\delta\,\frac{v+1}{2}\ln(1+\delta)$$

Or

$$\left(v^{(zJ)}-\mu\right)'\left(\lambda_z\, B_J^u\right)^{-1}\left(v^{(zJ)}-\mu\right) \le \frac{\delta^2}{2}(v+1) \tag{6.7}$$

In applying the new proposition to Φ and l, therefore:

$$\frac{l(x)}{\Phi(x)} \le (1+\delta)^{-2v}\sqrt{\frac{|\lambda\sum|}{\left|(1+\delta)^{-\frac{1}{4}}\,\lambda_{z+1}\,B_J^l\right|}}\; exp\left[\frac{1}{2}\left(v^{(zJ)}-\mu\right)'\left\{\lambda\sum-(1+\delta)^{-\frac{1}{4}}\,\lambda_{z+1}\,B_J^l\right\}^{-1}\left(v^{(zJ)}-\mu\right)\right]$$

By using the concavity of the function $\delta \mapsto 1-(1+\delta)^{-\frac{1}{4}}$. In order to have $l \le \Phi$, it was sufficient that $\left(v^{(zJ)}-\mu\right)'\left(\lambda_{z+1}\,B_J^u\right)^{-1}\left(v^{(zJ)}-\mu\right) \le \left(\frac{7}{8}v+\frac{1}{2}\right)\ln(1+\delta)(2\delta)(1-2^{-\frac{1}{4}})$

Or,

$$\left(v^{(zJ)}-\mu\right)'\left(\lambda_{z+1}\,B_J^u\right)^{-1}\left(v^{(zJ)}-\mu\right) \le \delta^2(1-2^{-\frac{1}{4}})^{\frac{v}{2}} \tag{6.8}$$

Finally, it was verified that condition (6.6) implied the (6.7) and (6.8) inequalities.

Proposition 3

There exists a constant A_2 that only depends on σ_m, σ_M and a such that:

$$card\ \mathcal{R}(\mathcal{E},K,v) \le \left(A_2\,\frac{v}{\mathcal{E}}\right)^{K(2v-1)+1} \tag{6.9}$$

Demonstration

The definition of the set $\mathcal{R}(\mathcal{E},K,v)$ enabled to increase its cardinal in the following manner:

$$card\ \mathcal{R}(\mathcal{E},K,v) \le \sum_{z=0}^{r'}\left[\sum_J\left\{\prod_{q=1}^{v-1}1\vee\frac{2a(1+\delta)^{\frac{j(q)+1}{4}+\frac{3}{2}}}{\sqrt{c_1}\,\sigma_M^2\,\delta}\right\}\left\{1\vee\frac{2a(1+\delta)^{\frac{v+z}{2}}}{\sqrt{c_1}\,\sigma_M^{2-v}(1+\delta)^{\frac{S_J}{4}}\delta}\right\}\right]^K$$

If $2a \ge \sqrt{c_1}(\sigma_M^2 \vee \sigma_M^{2-v})$, then:

$$card\ \mathcal{R}(\mathcal{E},K,v) \le r'\left[(r-1)^{v-1}\left\{\frac{2a(1+\delta)^{\frac{r+1}{4}+\frac{r'}{2}}}{\sqrt{c_1}\,\sigma_M^2\,\delta}\right\}^{v-1}\left\{\frac{2a(1+\delta)^{\frac{v+r'}{2}}}{\sqrt{c_1}\,\sigma_M^{2-v}\delta}\right\}\right]^K$$

$$\le \frac{r'(r-1)^{K(v-1)}}{\delta^{vK}}(1+\delta)^{\frac{Kv(r'-1)}{2}+\frac{K(v-1)(r-1)}{4}+Kv-\frac{K(v-1)}{2}}\left(\frac{2a}{\sqrt{c_1}\,\sigma_M}\right)^{Kv}$$

We note that $\delta \le 1$. Then, the definitions of both r and r' indicated that on the one side, $r-1 \le \frac{4(1+\delta)\sigma_M^2}{\delta\sigma_m^2}$ and $r' \le \frac{2\,\sigma_M^2}{\delta\,\sigma_m^2}$

And on the other side, $(1+\delta)^{r'} \leq 4\left(\frac{\sigma_m}{\sigma_M}\right)^4$ and $(1+\delta)^{r} \leq 8\left(\frac{\sigma_m}{\sigma_M}\right)^4$

Therefore, we can find a constant c_2 that only depends on σ_m, σ_M and a such that

$$card\ \mathcal{R}(\mathcal{E}, K, v) \leq \frac{{c_2}^{Kv}}{\delta^{K(2v-1)+1}} \leq \left(A_2 \frac{v}{\mathcal{E}}\right)^{K(2v-1)+1}$$

With A_2 that only depended on σ_m, σ_M and a. Furthermore, if $2a \leq \sqrt{c_1}\ (\sigma_M^2 \vee \sigma_M^{2-v})$ then the increment in (6.30) was also verified on the condition that the constant c_2 must change. Lemmas 3 and 4 indicated that the family $\mathcal{R}(\mathcal{E}, K, v)$ formed a $\mathcal{E} - recovery$ of $\mathcal{F}_{(K,v)}$. The lemma in 5.2.17 gave the increase of the number of entropy stated in proposition 4 by using $v \leq Q$.

End of proof of proposition for $\mathcal{M}[\mathcal{LB}_k]$.

In addition, it was clear that:

$$H_{[.]}(\mathcal{E}, S_{(K,v)}, d_H) = H_{[.]}(\mathcal{E}, \mathcal{L}_{(K,v)}, d_H) \leq H_{[.]}\left(\frac{\mathcal{E}}{3}, \tau_{K-1}, d_H\right) + H_{[.]}\left(\frac{\mathcal{E}}{3}, \mathcal{F}_{(K,v)}, d_H\right)$$

Then propositions 2 and 3 were used:

$$H_{[.]}(\mathcal{E}, \mathcal{L}_{(K,v)}, d_H) \leq \ln K + \frac{K}{2}\ln(2\pi e) + (K-1)\ln\frac{3}{\mathcal{E}} + [K(2v-1)+1]\left(\ln A_2 + \ln v + \ln\frac{3}{\mathcal{E}}\right)$$
$$\leq D(K, v)\left[\ln\frac{1}{\mathcal{E}} + A_1 \ln v\right]$$

6.7.3 Rational supporting the extraction of big size containers

Here, the interest was on the extraction of storages higher than x_0. Let N be the total number of storages of the platform with sizes higher than x_0. The model was based on the two following hypotheses:

1. The size of containers of this category were represented by an i.i.d. sample $(X_1, \ldots, X_N)$ of common distribution $Par(\alpha, x_0, +\infty)$;

2. The extraction dates D_i were independent. And for all i=1,…,N,

$$(D_i | X_i) \sim \varepsilon(h(X)) \tag{6.10}$$

However, this distribution did not correspond to the sample of storages observed, meaning extracted before the current day.

Sample of containers observed before the time t^*

The platforms in which the extraction process was processed were mature platforms and from which the most important containers were extracted long time ago.

In this case, containers of this category were all of sizes found in the interval $[x_0, x_{max}]$, where x_{max} was the size of the biggest container of the platform. Therefore, for a storage of size X of this category, we have:

$$X \sim Par(\alpha, x_0, x_{max}) \tag{6.11}$$

From the above, the conditional density of a couple (X,D) knowing that $X \in [x_0, x_{max}]$ was given by:

$$g: (x,t) \longmapsto \alpha h(x) \exp\{-h(x)t\} \frac{x^{-\alpha-1}}{x_0^{-\alpha} - x_k^{-\alpha}} 1_{t \geq 0, x_0 \leq x \leq x_k} \tag{6.12}$$

Let's recall that t^* designated the present date. At the date t^*, an observer only saw the storages such that $D \leq t^*$. Storages distribution of size in $[x_0, x_{max}]$ and extracted before t^* corresponded to the distribution of density (6.12), truncated beyond t^*.

Let (X^*, D^*) be the couple associated to one of the storages observed; this has the following distribution:

$$(X^*, D^*) = \left((X, D) | D \leq t^*\right)$$

For all $t \geq 0$ and all $x \in \mathbb{R}^+$, the density g^* of (X^*, D^*) was defined by:

$$g^*(x,t) = \frac{g(x,t)}{P_{dec}(\alpha, t^*, h)} 1_{t \leq t^*} \tag{6.13}$$

Where P_{dec} represented the probability that a storage of size $[x_0, x_{max}]$ was extracted before t^*. The quantity was calculated as follows:

$$P_{dec}(\alpha,\ t^*, h) = P(D \leq t^* | X \in [x_0, x_{max}])$$

$$= \frac{\alpha}{x_0^{-\alpha} - x_{max}^{-\alpha}} \int_{x_0}^{x_{max}} \int_0^{t^*} h(x) \exp\{-h(x)t\} x^{-\alpha-1}\, dx dt$$

$$= 1 - \frac{\alpha}{x_0^{-\alpha} - x_{max}^{-\alpha}} \int_{x_0}^{x_{max}} \exp\{-h(x)t\} x^{-\alpha-1}\, dx$$

In addition, let n be the number (random) of storages extracted at date t^*. Conditionally to n, the sample $\left((X_1^*, D_1^*) \ldots (X_n^*, D_n^*)\right)_{i=1,\ldots,n}$ was i.i.d. and each couple (X_n^*, D_n^*) had the joint law of density given in (6.13). In order to construct the estimators of h, all the probabilities were conditional to n, which were considered as a deterministic quantity.

Major results

The following two results were obtained in both the ordered and the unordered collection models.

A. Ordered collection models

In this case, the bloc of variables of classification was of the form $\{1, \dots, v\}$. For the considered three types of collections of gaussian mixtures, the models $S_{(K,v)}$ were such that:

$$S_{(K,v)} = \left\{x \in \mathbb{R}^Q \longmapsto f(x_1, \dots, x_5)\Phi\left(x_{v+1}, \dots, x_Q | 0, I_{Q-v}\right); f \in \mathcal{L}_{(K,v)}\right\}$$

Some results were obtained for the collections of mixtures that were defined by the families of densities $\mathcal{L}_{(K,v)}$ highlighted earlier. Each collection corresponded to a certain form of variance-covariance matrix for components of mixtures chosen in $\mathcal{L}_{(K,v)}$. Also in terms of definitions suggested in the previous chapter, supplementary hypotheses were added in order to limit the parameters of mixture:

$\mathcal{M}[LB_k]$ collection:

$$\mathcal{L}_{(K,v)} = \left\{\sum_{k=1}^{K} p_k \Phi(.|\mu_k, \lambda\sum_k); \begin{matrix} \lambda \in [\sigma_m^2, \sigma_M^2], \sum_k \in \Delta_{(v)}^1(\sigma_m^2, \sigma_M^2) \\ \mu_k \in [-a, a]^v, 0 < p_k < 1, \sum_{k=1}^K p_k = 1 \end{matrix}\right\}$$

Where $\Delta_{(v)}^1(\sigma_m^2, \sigma_M^2)$ designated the set of diagonal matrices defined positive of determinant 1 and whose proper values belonged to the interval $[\sigma_m^2, \sigma_M^2]$.

$\mathcal{M}[L_k B_k]$ collection:

$$\mathcal{L}_{(K,v)} = \left\{\sum_{k=1}^{K} p_k \Phi(.|\mu_k, \sum_k); \begin{matrix} \sum_k = diag\ (\sigma_{k1}^2, \dots, \sigma_{kv}^2), \sigma_{k1}^2, \dots, \sigma_{kv}^2 \in [\sigma_m^2, \sigma_M^2] \\ \mu_k \in [-a, a]^v, 0 < p_k < 1, \sum_{k=1}^K p_k = 1 \end{matrix}\right\}$$

$\mathcal{M}[L_k C_k]$ collection:

$$\mathcal{L}_{(K,v)} = \left\{\sum_{k=1}^{K} p_k \Phi(.|\mu_k, \sum_k); \begin{matrix} \sum_k \in \mathcal{D}_{(v)}^+(\sigma_m^2, \sigma_M^2), \mu_k \in [-a, a]^v \\ 0 < p_k < 1, \sum_{k=1}^K p_k = 1 \end{matrix}\right\}$$

Where $\mathcal{D}_{(v)}^+(\sigma_m^2, \sigma_M^2)$ designated the set of symmetric matrices defined positive whose proper values belonged to the interval $[\sigma_m^2, \sigma_M^2]$.

In the next theorem, $\mathcal{M}$ designates one of the three collections $\mathcal{M}[LB_k]$, $\mathcal{M}[L_k B_k]$ or $\mathcal{M}[L_k C_k]$.

Proposition 4

There exist $k, A\ and\ C$ constants such that if

$$pen(K,v) \geq k\frac{D(K,v)}{n}\left\{2A\ln v + 1 - \ln\left(1\wedge\left[\frac{D(K,v)}{n}A\ln v\right]\right)\right\}$$

Then the $(\widehat{K},\hat{v})$ index model that minimised

$$crit\ (K,v) = \gamma_n\left(\hat{s}_{(K,v)}\right) + pen\ (K,v)$$

On $\mathcal{M}$ exists, and

$$\mathbb{E}\left[d_H^2\left(s,\hat{s}_{(\widehat{K},\hat{v})}\right)\right] \leq C\left[\inf_{(K,v)\in\mathcal{M}}\left\{KL\left(s,S_{(K,v)}\right) + pen(K,v)\right\} + \frac{1}{n}\right] \qquad (6.14)$$

In addition, the constants $k\ and\ C$were absolute and the constant A depended only on $\sigma_m$and σ_M and a. The proof of proposition 4 is given in one of the next sections and is based on the technical calculations of the entropy.

B. Unordered collection models

It was recalled that the unordered collections models were noted $S_{(K,v)}$ where v designated the set of variables of classification. For $v = |v|$, a model $S_{(K,v)}$ was obtained from the $S_{(K,v)}$ model of the ordered collection that was associated to it by posing

$$S_{(K,v)} = \left\{x \in \mathbb{R}^Q \longmapsto f\circ\tau(x), f \in S_{(K,v)}\right\}, \qquad (6.15)$$

Where τ was a permutation such that $(\tau(x)_1, \ldots, \tau(x)_v)'$=$x_{[v]}$.

The unordered collection models were noted $\mathcal{M}'[LB_k]$, $\mathcal{M}'[L_kB_k]$ or $\mathcal{M}'[L_kC_k]$. In the next theorem, the set $\mathcal{M}$ corresponded to one of those three collections.

Proposition 5

There exist $k, A\ and\ C$ constants such that if:

$$pen(K,v) \geq k\frac{D(K,v)}{n}\left(2A\ln v - \ln\left(1\wedge\left[\frac{D(K,v)}{n}A\ln v\right]\right) + \frac{1}{2}\ln\left[\frac{8eQ}{\{D(K,v)-1\}\wedge(2Q-1)}\right]\right) \qquad (6.16)$$

Then the $(\widehat{K},\hat{v})$ that minimises the criterion

$$crit(K,v) = \gamma_n\left(\hat{s}_{(K,v)}\right) + pen\ (K,v) \qquad (6.17)$$

On the collection of models $\mathcal{M}'$ exists, and

$$\mathbb{E}\left[d_H^2\left(s,\hat{s}_{(\widehat{K},\hat{v})}\right)\right] \leq C\left[\inf_{(K,v)\in\mathcal{M}'}\left\{KL\left(s,S_{(K,v)}\right) + pen(K,v)\right\} + \frac{2}{n}\right] \qquad (6.18)$$

6.7.4 $\mathcal{M}[LB_k]$ entropies

Proposition 1

Under the previous hypotheses, $d_H(u,l) \leq \mathcal{E}$.

Demonstration

Consider:

$$d_H^2(l,u) = (1+\delta)^{-2v} + (1+\delta)^{2v} - \left\{ \prod_{q=1}^{v-1} \frac{(1+\delta)^{-\frac{1}{8}} \sqrt{\lambda_{z+1}}\, b_{j(q)+1} (1+\delta)^{\frac{1}{2}} \sqrt{\lambda_z}\, b_{j(q)}}{(1+\delta)^{-\frac{1}{4}}\, \lambda_{z+1} b_{j(q)+1^2} + (1+\delta)\, \lambda_z b_{j(a)^2}} \right\}$$

$$\times \left\{ 2 \frac{(1+\delta)^{-\frac{1}{8}} \sqrt{\lambda_{z+1}}\; B_{J,v}^{l}\, (1+\delta)^{\frac{1}{2}} \sqrt{\lambda_z}\; B_{J,v}^{u}}{(1+\delta)^{-\frac{1}{4}}\, \lambda_{z+1} \left[B_{J,v}^{l} \right]^2 + (1+\delta)\, \lambda_z \left[B_{J,v}^{u} \right]^2} \right\}^{\frac{1}{2}}$$

$$= (1+\delta)^{-2v} + (1+\delta)^{2v} - 2 \left\{ \frac{2}{(1+\delta)^{-\frac{11}{8}} + (1+\delta)^{\frac{11}{8}}} \right\}^{\frac{v-1}{2}} \left\{ \frac{2}{(1+\delta)^{-\frac{5+v}{4}} + (1+\delta)^{\frac{5+v}{4}}} \right\}^{\frac{1}{2}}$$

$$= 2\cosh(2v \ln(1+\delta)) - 2 \left\{ \cosh\left(\frac{11}{8} \ln(1+\delta) \right) \right\}^{-\frac{v-1}{2}} \left\{ \cosh \frac{5+v}{4} \ln(1+\delta) \right\}^{-\frac{1}{2}}$$

$$= [2\cosh(2v \ln(1+\delta)) - 2] + \left[2 - 2 \left\{ \cosh\left(\frac{11}{8} \ln(1+\delta) \right) \right\}^{-\frac{v-1}{2}} \right]$$

$$+ 2 \left\{ \cosh\left(\frac{11}{8} \ln(1+\delta) \right) \right\}^{-\frac{v-1}{2}} \left[1 - \left\{ \cosh\left(\frac{5+v}{4} \ln(1+\delta) \right) \right\}^{-\frac{1}{2}} \right]$$

Therefore,

$$d_H^2(l,u) \leq 4\sinh(1) v^2 \delta^2 + 2 \frac{v-1}{2} \frac{11}{8} \delta^2 + 2 \frac{5+v}{4} \frac{1}{2}\, \delta^2$$

$$\leq 9\, v^2 \delta^2$$

$$\leq \mathcal{E}^2$$

At this stage, it was possible to construct some elements of the set $\mathcal{F}_{(K,v)}$.

Let $(\Phi_1, \ldots, \Phi_K) \in \mathcal{F}_{(K,v)}$, with $\Phi_k = \Phi(.|\mu_k, \lambda \sum_k)\ for\ k = 1, \ldots, K$.

Also, let $\lambda_z, J_1, \ldots, J_k$ such that for all $k = 1 \in \{1, \ldots, K\}$ and for all $q \in \{1, \ldots, v\}$,

$\lambda_{z+1}\, B_{k,q}^{l} \leq \lambda \sum_{k,qq} \leq \lambda_z\, B_{k,q}^{u}$, where $\sum_{k,qq}$ designated the (q, q) coefficient of the matrix $\sum_k$. For all k, there existed a vector $v^{(z\, J_k)}$ of the mesh associated with the couple (z, J_k).

Proposition 2

Let $K \geq 2$ and $v \geq 1$ be natural numbers. Then, it was assumed that $\{[a_i, b_i], i \in I\}$ was a family that formed a simplex recovery$\mathcal{T}_{K-1}$, where $a_i \coloneqq (a_{1i}, \ldots, a_{Ki})$ and $b_i \coloneqq (b_{1i}, \ldots, b_{Ki})$. It was further assumed that there existed a family $\{[l_j, u_j], j \in J\}$ that formed a recovery of $\mathcal{F}_{(K,v)}$, with $l_j \coloneqq (l_{1j}, \ldots, l_{Kj})$ and $u_j \coloneqq (u_{1j}, \ldots, u_{Kj})$.

Therefore, $\forall\, (i,j) \in I \times J$, let $L_{ij} = \sum_{k=1}^{K} a_{ki} l_{kj}$ and $U_{ij} = \sum_{k=1}^{K} b_{ki} u_{kj}$. Then the family,

$\{[L_{ij}, U_{ij}], (i,j) \in I \times J\}$ formed a recovery of the $\mathcal{L}_{(K,v)}$ set.

The demonstration of this result was identical to the result that was demonstrated in [185]. And the increment of the $\mathcal{L}_{(K,v)}$ (or of the $S_{(K,v)}$) metric entropy was reduced to the metric entropy of $\mathcal{F}_{(K,v)}$.

Proposition 3

There exists a constant A_2 such that for all $\mathcal{E} \in\,]0,1]$,

$$N_{[.]}(\mathcal{F}_{(K,v)}) \leq \left(A_2 \frac{v}{\mathcal{E}}\right)^{K(2v-1)+1}$$

And the constant A_2 only depends on σ_m, σ_M and a. This equation will be demonstrated using many successive lemma techniques. The proof of proposition 1 was given at the end of this section. First, when $v = 1$, entropies of models $S_{(K,1)}$ were calculated by imitating the method used to calculate $\mathcal{M}[L_k B_k]$. Therefore, $v \geq 2$.

Also, let $\mathcal{E} \in\,]0,1]$, $K \geq 2$ and $v \geq 2$ were fixed, by posing $\delta = \frac{\mathcal{E}}{3v}$. For $j = 1 \ldots r$, let

$$b_j^2 = (1+\delta)^{1-\frac{j}{2}} \sigma_M^2$$

Where $r = \left[2\, \frac{\ln\left\{\frac{\sigma_M^2}{\sigma_m^2}(1+\delta)\right\}}{\ln(1+\delta)}\right]$ such that $b_r^2 \leq \sigma_m^2$ and $b_2^2 = \sigma_M^2$.

Also, for $z = 0 \ldots r'$, let $\lambda_z = (1+\delta)^{-z} \sigma_M^2$

Where $r' = \left[\frac{\ln\left\{\frac{\sigma_M^2}{\sigma_m^2}\right\}}{\ln(1+\delta)}\right]$ such that $\lambda_{r'} \leq \sigma_m^2$ and $\lambda_0 = \sigma_M^2$. For all vector

$$J = (j(1), \dots, j(v-1)) \in \{2, \dots, r\}^{v-1}$$

B_J^l and B_J^u matrices were defined by the following:

$$B_J^l = diag\left(b_{j(1)+1}^2, \dots, b_{j(v-1)}^2, \sigma_M^{-2v+2}(1+\delta)^{\frac{S_J}{2}-(v-1)}\right),$$

And

$$B_J^u = diag\left(b_{j(1)}^2, \dots, b_{j(v-1)}^2, \sigma_M^{-2v+2}(1+\delta)^{\frac{S_J}{2}-\frac{(v-1)}{2}}\right),$$

With $S_J = \sum_{q=1}^{v-1} j(q)$

In what follows, for $q \in \{1, \dots v\}$, the notations $B_{J,q}^l$ and $B_{J,q}^u$ designated the (q, q) coefficients of B_J^l and B_J^u matrices.

Proposition 4

Let $\sum \in \Delta_{(v)}^1(\sigma_M^{-2}, \sigma_M^2)$ and $\lambda \in [\sigma_m^2, \sigma_M^2]$. Then, $\exists\, J \in \{2, \dots, r\}^{v-1}$ and $z \in \{0, \dots, r'\}$ such that $\forall\, q \in [1, \dots, v]$, $\lambda_{z+1} B_{J,q}^l \leq \lambda \sum_{qq} \leq \lambda_z B_{J,q}^u$

Demonstration

Let z be the unique natural number in $\{0, \dots, r'\}$ such that $\lambda_{z+1} < \lambda \leq \lambda_z$. By definition, there exists a vector J such that $B_{J,q}^l \leq \sum_{qq} \leq B_{J,q}^u$ for $q = 1, \dots, v-1$; which gave the wanted direction for the first v-1 coordinates. Because $\prod_{q=1}^{v} \sum_{qq} = 1$, then $\sum_{vv} = \prod_{q=1}^{v-1} \sum_{qq}^{-1}$. From this, it was deduced that:

$$\prod_{q=1}^{v-1} \left\{B_{J,q}^u\right\}^{-1} \leq \sum_{vv} \leq \prod_{q=1}^{v-1} \left\{B_{J,q}^l\right\}^{-1}$$

Which gave the said directive for the last coordinate. Covariance matrices associated with the couple (z, J) were useful because it helped for the definition of the extremity functions. However, the centering parameters also needed to be defined: For a fixed couple (z, J), the regular mesh of centering parameters was considered: $v^{(zJ)} = \left(v_1^{(zJ)}, \dots, v_v^{(zJ)}\right) \in [-a, a]^v$ such that for $\in \{1, \dots, v-1\}$, $v_q^{(zJ)} = (1+\delta)^{-\frac{j(q)+1}{4}-\frac{z}{2}} \sigma_M^2 \sqrt{c_1}\, \delta s_q$,

With $s_q \in \{-N_q, \dots, N_q\}$ where $N_q = \left[\frac{a(1+\delta)^{\frac{j(q)+1}{4}+\frac{z}{2}}}{\sqrt{c_1}\, \sigma_M^2\, \delta}\right]$, and,

With $s_v \in \{-N_v, \dots, N_v\}$ where $N_v = \left[\frac{a(1+\delta)^{\frac{v+z}{2}-\frac{S_J}{4}}}{\sqrt{c_1}\, \sigma_M^{2-v\delta}}\right]$, and

$$C_1 \coloneqq \frac{1-2^{-\frac{1}{4}}}{2}.$$

For a given couple (z,J), this guaranteed that for all $\mu \in [-a,a], \exists$ a vector $v^{(zJ)}$ of the mesh such that:

$$\left\{\sum_{q=1}^{v-1}\left(v_q^{(zJ)}-\mu_q\right)^2(1+\delta)^{\frac{j(q)+1}{2}+z}\,\sigma_M^{-4}\right\}+\left(v_v^{(zJ)}-\mu_v\right)^2(1+\delta)^{-\frac{S_J}{2}+v+z}\,\sigma_M^{2v-4} \le c_1 v\delta^2 \tag{6.19}$$

Furthermore, for a couple (z, J), and a parameter $v^{(zJ)}$ of the above family, the following two functions were finally considered:

$$\begin{cases} l(x) = (1+\delta)^{-2v}\Phi\left(x|v^{(zJ)},(1+\delta)^{-\frac{1}{4}}\lambda_{z+1}\,B_J^l\right) \\ u(x) = (1+\delta)^{2v}\Phi\left(x|v^{(zJ)},(1+\delta)\,\lambda_z\,B_J^u\right) \end{cases} \tag{6.20}$$

In addition, let $\in \left[\sigma_m^{,2},\ \sigma_M^{,2}\right], \sum \in \Delta_{(v)}^1\left(\sigma_m^{,2},\ \sigma_M^{,2}\right), \mu_k \in [-a,a]^v$, then:

$$\Phi \coloneqq \Phi(.|\mu,\lambda\sum)$$

There existed λ_z and J such that for all $q \in \{1,\dots,v\}$,

$$\lambda_{z+1}\,B_{J,q}^l \le\ \lambda\sum_{qq} \le\ \lambda_z\ B_{J,q}^u$$

There existed also a vector $v^{(zJ)}$ of the mesh associated to the couple (z, J) such that the constraint (6.19) was verified for the vector μ. For z, J and this vector $v^{(zJ)}$, u and l were considered as two functions defined like in (6.20).

Proposition 5

Under the previous hypotheses, for all $x \in \mathbb{R}^d$, $l(x) \le \Phi(\mathrm{x}) \le \mathrm{u(x)}$.

Demonstration

By applying the lemma 2 that enables to increase the gaussian densities quotients, we obtain

$$\frac{\Phi(\mathrm{x})}{u(x)} \le (1+\delta)^{-2v}\sqrt{\frac{\left|(1+\delta)\lambda_z\,B_J^u\right|}{|\lambda\sum\quad|}}\exp\left[\frac{1}{2}\left(v^{(zJ)}-\mu\right)'\left\{(1+\delta)\lambda_z\,B_J^u-\lambda\sum\right\}^{-1}\left(v^{(zJ)}-\mu\right)\right]$$

$$\le (1+\delta)^{-\frac{v+1}{2}}\exp\left\{\frac{1}{2\delta}\left(v^{(zJ)}-\mu\right)'\left(\lambda_z\,B_J^u\right)^{-1}\left(v^{(zJ)}-\mu\right)\right\}$$

For $\Phi \le$ u, it is sufficient to have:

$$\left(v^{(zJ)}-\mu\right)'\left(\lambda_z\,B_J^u\right)^{-1}\left(v^{(zJ)}-\mu\right) \le 2\delta\,\frac{v+1}{2}\ln(1+\delta)$$

Or

$$\left(v^{(zJ)} - \mu\right)'\left(\lambda_z\, B_J^u\right)^{-1}\left(v^{(zJ)} - \mu\right) \leq \frac{\delta^2}{2}(v+1) \tag{6.21}$$

In applying the new lemma to Φ and l, therefore:

$$\frac{l(x)}{\Phi(x)} \leq (1+\delta)^{-2v}\sqrt{\frac{|\lambda\sum|}{\left|(1+\delta)^{-\frac{1}{4}}\lambda_{z+1}\, B_J^l\right|}}\; exp\left[\frac{1}{2}\left(v^{(zJ)} - \mu\right)'\left\{\lambda\sum - (1+\delta)^{-\frac{1}{4}}\lambda_{z+1}\, B_J^l\right\}^{-1}\left(v^{(zJ)} - \mu\right)\right]$$

By using the concavity of the function $\delta \longmapsto 1-(1+\delta)^{-\frac{1}{4}}$. In order to have $l \leq \Phi$, it was sufficient that $\left(v^{(zJ)} - \mu\right)'\left(\lambda_{z+1}\, B_J^u\right)^{-1}\left(v^{(zJ)} - \mu\right) \leq \left(\frac{7}{8}v + \frac{1}{2}\right)\ln(1+\delta)(2\delta)(1-2^{-\frac{1}{4}})$

Or,

$$\left(v^{(zJ)} - \mu\right)'\left(\lambda_{z+1}\, B_J^u\right)^{-1}\left(v^{(zJ)} - \mu\right) \leq \delta^2(1-2^{-\frac{1}{4}})^{\frac{v}{2}} \tag{6.22}$$

Finally, it was verified that condition (6.20) implied the (6.21) and (6.22) inequalities.

Discussion

On the contrary to classical situations in Mathematics and Statistics, the number of variables Q was fixed meanwhile n approached infinity. The two previous results enabled to consider situations where Q increased with n. For the specific problems where the number of variables was of order n, or even greater than n, the two oracle inequalities indicated that the penalised criterion were still pertinent.

Even though the LB_k models were sub-models of the $L_kB_k$family, which were sub-models of the L_kC_k family, results of the two previous theorems required a specific demonstration for each of the three collections. In fact, in order to have the penalty term $D(K,v)$ to correspond to the number of free parameters in each of the three situations, it was necessary to perform the calculations of the metric entropy for each of the three types of models.

In the discussion related to proposition 4, it was possible to rely on a risk limit that only involved the Hellinger distance. Therefore, for $\|x\| \to +\infty$, the gaussian mixture densities models had all the same behavior than the gaussian density. In order to guarantee that:

$$\underset{t \in S_{(K,v)}}{sup}\left\|\frac{s}{t}\right\|_\infty \leq M,$$

It was natural to suggest that s was as compact support, in this case, the limit M depended on σ_m, σ_M, a and $\|s\|_\infty$, but it was not necessary to know it in order to use the penalised criterion in practice.

As for proposition 4, results obtained must be interpreted in a qualitative way because their principal benefit was to provide the form of the penalty to be minimised in the criterion. Once adapted to this context, the slope method enabled to calibrate the penalty. A description of the corresponding procedure is given later on.

And the term $\frac{D(K,v)}{n}\ln\left\{1/\left(1A\frac{D(K,v)}{n}\ln v\right)\right\}$ that was found in the ordered and unordered collections was probably not necessary in order to define the optimal penalties, meaning those leading to estimators minimising the estimation risk. The presence of this term was directly linked to the fact that we manage to globally control the entropy of models and not locally as it would be necessary in order to apply theorem 1. In fact, it was difficult to perform fine recoveries of the sets

$$S_{(K,v)}(u,\xi) \coloneqq \left\{t \in S_{(K,v)}; d_H(t,u) \leq \xi\right\}$$

The entropy of those sets was majored by that of the entire $S_{(K,v)}$. In practice, it will be seen that the factor ln v in the first term was not necessary either in order to define the optimal penalties. A discussion was entirely consecrated to the practice of the results in the upcoming sections.

One way to highlight that an estimator was of good quality was to consider its maximal risk on a class $\mathcal{F}$, of densities functions. The performances of the estimator were evaluated by comparing this maximal risk to the minimax risk on the class $\mathcal{F}$,

$$R_{minimax} \coloneqq \inf_{\hat{s}} \sup_{s \in \mathcal{F}} \mathbb{E}(\|s - \hat{s}\|^2)$$

Where the lower limit covered all possible estimators $\hat{s}$. Generally, for the class $\mathcal{F}$, functional classes of Hölder, Sobolev, Besov or ellipsoids of $\mathbb{L}^2$ [207] were taken. When the penalised estimator reached the minimax risk limit on a large functional classes family and without using the knowledge of the particular class to which it is belonged to, it was called adaptive. The fact that the oracle inequalities obtained in theorems proposition 1 and proposition 2 were valid for all density s enabled to major the maximal risk $\hat{s}_{(\hat{K},Q)}$ for the Hellinger distance, on a functional class $\mathcal{F}$ by the following quantity:

$$C \sup_{s \in \mathcal{F}}\left[\inf_{(K,Q) \in \mathcal{M}}\left\{KL\left(s, S_{(K,Q)}\right) + pen(K)\right\} + \frac{1}{n}\right]$$

For a collection of ordered models. In order to evaluate the performances of the penalised estimator in terms of the minimax risk, the major challenge faced was not in the mathematical

context but it was related to the functional approximation. For a chosen function s in the classes cited above, there was no results enabling to quantify the bias $KL\left(s, S_{(K,Q)}\right)$ in a precise manner between the density s and the functional space composed of gaussian mixture densities with K components. Even in dimension 1, a result could not be found.

6.7.5 Choice of the form of the visibility function

In the previous chapter, we have noted that the ideal model that was based on a visibility function proportional to the size was not sufficiently realistic. In the Lopez model, ω was a constant function by pieces, the visibility function associated to ω was therefore constant by pieces too. For the subject that matters in this section, namely the modelling of the extraction, the use of constant visibility functions by pieces was confronted to the following challenges: On the one hand, it was known that levelled reserves of a platform followed the Pareto-Levy law. On the other hand, each class of the partition used had to contain a sufficient number of storages for evident reasons of quality of estimation. The Pareto-Levy distribution leaded to the definition of a class of big storages that had a big amplitude. Inside this class of big storages, the size effect on the extraction times was observed as well. In reality, the five biggest storages were extracted more quickly than other storages of this class. Choosing a constant visibility function by pieces led to a systematic under-estimation of the extraction dates of these storages. However, at the beginning of the extraction of the platform, the total production was essentially oriented by this small population of giants' storages. By being modelled, the beginning of the production would also be under-estimated because of the systematic delay of extraction dates of bigger storages. For this reason, in what follows, we cannot use visibility functions corresponding to the Lopez model rigorously.

In order to consider precedent remarks, the use of a visibility function h affined by pieces and increasing was suggested. This model allowed to conserve the non-parametric point of view of Lopez, while authorising the visibility to grow inside the partition classes in which h was defined [204].

Let $m = (I_1, \dots, I_k)$ be a partition of size k of $[x_0, x_{max}]$. The partition was composed of the following intervals:

$$I_1 =]x_0, x_1], \dots, I_j =]x_{j-1}, x_j], \dots, I_k =]x_{k-1}, x_{max}],$$

With $1 < x_0 < x_1 < \cdots < x_j \dots < x_k = x_{max}$. Associated with the partition m, let $\mathcal{H}_m$ be the set composed of continuous functions and affined by pieces.

$$\mathcal{H}_m = \{\text{h: x} \in [x_0, x_{max}] \longmapsto \sum_{j=1}^{k}\{a_j(x - x_{j-1}) + b_j\}1_{x\in I_j}, (A,B) \in (\mathbb{R}^+)^{2k}\} \bigcap C^+([x_0, x_{max}])$$

Where $C^+([x_0, x_{max}])$ designated the set of strictly positive continuous functions on $[x_0, x_{max}]$. A and B vectors designated $(a_1, \dots, a_k)$ and $(b_1, \dots, b_k)$ respectively. For all function $h \in \mathcal{H}_m$, the condition of continuity imposed that:

$$\forall\, j \geq 2, b_j = \sum_{u=1}^{j-1} a_u(x_u - x_{u-1}) + b_1 \tag{6.23}$$

In addition, the condition of positivity was equivalent to $b_1 > 0$. Therefore, the set $\mathcal{H}_m$ was entirely parametered by the slope vector $A \in (\mathbb{R}^+)^k$ and the only coefficient $b \coloneqq b_1 > 0$. It was noted that a function $h \in \mathcal{H}_m$ admitted only one notation of the form:

$$\sum_{j=1}^{k}\{a_j(x - x_{j-1}) + b_j\}1_{x\in I_j}$$

With all b_j verifying (6.8), which allowed the assimilation of the function h to A and B coefficients. In what follows, it was useful to write h(A,b) or h(A,B) as the visibility function h defined by (A,b) or (A,B).

6.7.5 Identification of a model for the computation of the density

At the beginning of this chapter, an overview of the statistical context for the identification of models was provided. This was done in order to compute the density and a general theorem established by Massart [205] and that enabled the determination of the penalised criteria for the collection of models that were applied in biofuel production based on *ricinus communis* oil.

Let $X_1, \dots, X_n$ be an i.i.d. sample, with $X_i \in \mathbb{R}^d$ of probability density s and unknown for the measure of Lebesgue on $\mathbb{R}^d$. Let S be the set of all densities for the measure of Lebesgue on $\mathbb{R}^d$. The method of maximum likelihood that consisted of finding the parameters of a model that maximised the likelihood of observations was reinterpreted as a method of contrast minimisation. Thereby, the contrast $\gamma(t,.) = -\ln\{t(.)\}$ was considered. Let $\gamma_n(t) = -\frac{1}{n}\sum_{i=1}^{n}\ln\{t(X_i)\}$ be the empirical contrast associated to γ. In this context, the loss function defined by:

$$l(s,t) = \mathbb{E}[\gamma_n(t)] - \mathbb{E}[\gamma_n(s)],$$

Was exactly the Kullback-Leibler information [208] and was defined for the two densities f and g of S by:

$$KL(f,g) = \int \ln\left\{\frac{f(x)}{g(x)}\right\} f(x)\, dx,$$

Where f dx was absolutely continuous with regards to g dx, and $+\infty$ otherwise. The density s was the unique function of S such that $s = \underset{t\in S}{argmin} \int \gamma(t,x)s(x)dx$.

Let $\mathcal{S}$ be a subset of S, the estimator of the maximum likelihood (EML) of s on $\mathcal{S}$ was defined by:

$$\hat{s} \coloneqq \underset{t \in S}{argmin}\ \gamma_n(t)$$

Thus, by replacing γ by γ_n and S by S_n, it was expected that the estimator obtained was close to the real density s, at least in cases where s was not too far from the S model and from n sufficiently great.

In the context of models selection, there was a collection of models $(S_m)_{m\in\mathcal{M}_n}$and an estimator of the likelihood maximum $\hat{s}_m$ for each of them. As suggested in the notations, the collection of the model was enabled to depend on the size n of the sample observed. In addition, it was not necessary to suppose that the density s belonged to one of the models of the collection.

Thereafter there was a need to use the EML associated to the best model in the sense of a certain statistical criteria. In the context of the computation of the density, a natural criteria was the minimisation of the mean estimation error (or estimation risk), that was defined for an estimator $\hat{s}_m$ by:

$$\mathcal{R}(\hat{s}_m) = \mathbb{E}[KL(s,\hat{s}_m)]$$

Ideally, there was a need to select the model minimising this quantity. However, in practice, this was impossible because the risk depended on the density s that was unknown; the density $\widetilde{m}$ that minimised the estimation risk for the collection $(S_m)_{m\in\mathcal{M}_n}$was called the oracle. The oracle inequality enabled to highlight such proprieties in a non-asymptotic manner [209]:

$$\mathbb{E}[KL(s,\hat{s}_m)] \leq CE\left[\underset{m\mathcal{M}_n}{inf}\ \{KL(s,\hat{s}_m) + R(m,n)\}\right]$$

Where C was a constant and R(m,n) was a term that should not be very big compared to the term of risk.

6.8 Method of calibration and applications

Theoretical results obtained in section 6.7 and which were related to the selection of gaussian mixture models for oil extraction models were not directly applied. In fact, the form of penalties was determined but there was a need to calibrate the constants in order to use the associated penalised criteria in practice. In this section, we base ourselves on a method of calibration of the penalty called slope method in order to determine the optimal penalties from the data.

6.8.1 Discrete transformation in wavelets for normalised profiles of production

In order to obtain information about the time and the frequency, a discrete wavelets transformed of normalised profiles of production was used. Here is the recall of the definitions and results in order to rigorously present decompositions in wavelets performed on all curves: Let Y be a temporal series of length $N = 2^J$. Then, $\mathcal{W}$ was the $N \times N$ matrix associated to the transformation in wavelets and W was the vector of $\mathbb{R}^n$ of the coefficients of the wavelets of Y. Then, $W = \mathcal{W}$ Y. Therefore, the matrix $\mathcal{W}$ was written under the form:

$$\mathcal{W} = \begin{bmatrix} \mathcal{V}_J \\ \mathcal{W}_J \\ \mathcal{W}_{J-1} \\ \vdots \\ \mathcal{W}_1 \end{bmatrix}$$

Also, for all j $\in \{1, \dots, J\}$, the sub-matrix $\mathcal{W}_j$ was of size $\frac{N}{2^j} \times N$. It was possible to describe the lines of $\mathcal{W}_j$ from the wavelets functions: $\Psi_{j,k} = 2^{-j/2}\,\Psi(2^{-j}x - k2^j)$, $j = 1, \dots, J, k = 0, \dots, 2^{J-j} - 1$, where Ψ was called the mother wavelet of the family. However, in this work, the Haar wavelet was used with $\Psi = 1_{[0,1/2]} - 1_{[1/2.1]}$.

Then, the k[th] line of $\mathcal{W}_j$ was defined by the vector of values taken by $\Psi_{j,k}$ at points $\{0, \dots, N - 1\}$. Thus, it was possible to indicate that the transformation in wavelets was an orthogonal transformation and therefore: $Y = \mathcal{W}'W$.

Also, the vector W was decomposed the same way than $\mathcal{W}$ by posing: $W = (V_J', W_J', \dots, W_1')'$,

Where $W_j = \mathcal{W}_j Y$ contained the set of coefficients of wavelets associated to the level j. W_j coefficients corresponded to the variations of Y, at the 2^{J-j} scale, while the V_J coefficient was equal to $\frac{\bar{Y}}{\sqrt{N}}$. In the present case, the profiles of production were of size N=64. Each curve was described by a vector $W_{.i}$ of length 64, formed with coefficients of wavelets.

This vector was decomposed into 6 blocs corresponding to the different levels of resolution, for the curve I, such that: $W_{.i} = (V'_{6i}, W'_{6i}, \dots, W'_{1i})$.

It was necessary to indicate that the families of gaussian mixture models used in this method were based on the distinction between the variables vector of classification that had the distribution of a multivariate gaussian of a mixture and the vector of the rest of the variables that had the distribution of a reduced and centered gaussian vector. Thereby, in order to facilitate the adjustment of a distribution of a gaussian vector on the second bloc, all the variables of wavelets coefficients were centered and reduced beforehand.

Technical variables and assumptions

In order to determine what types of technical variables were coherent with the obtained classification, the following three descriptive variables were considered:

1. The reserves of the three platforms: The assumption was that the estimations were available in the same base than the ones providing the production curves.

2. The depth of the container: The assumption was that some storages were composed of many reservoirs. In this case, the variable corresponded to the less deep reservoir.

3. The oil density in the container: Which was measured in API degrees. The more the oil was light, the more its density was low, and the more the API index was high. The API degree varied inside the same platform, especially when it was composed of many storages or reservoirs. The API index used here corresponded to a mean index on the set of the platform. Therefore, the following notations described above were used, considering that they were centered and reduced: $y_{\text{res i}}, y_{\text{dep i}}, y_{\text{den i}}$

6.8.2 Introduction to the concept of slope method

The objective of this sub-section is to present both the theoretical and the practical aspects of the slope method. This method was introduced by Massart [205] in order to calibrate all the known penalties forms of data. In fact, in many situations of selection of models, the collection of models that was considered had many models of the same dimension. In the current context, the collection $(S_D)_{D\in\mathcal{D}}$ was obtained by grouping a series of models of the same dimension into one model. In what follows, this collection of models was assumed to be indexed by the dimension of the models.

Let $X_1, \dots, X_n$ be a sample of the same unknown distribution s. Like in the previous sections, γ and γ_n designated the contrast associated to the Kullback-Leibler information and the empirical contrast associated to it respectively.

Furthermore, it was assumed that for all D in $\mathcal{D}$, there existed two densities s_D and $\hat{s}_D$ that minimised $KL(s,.)$ and $\gamma_n(.)$ on S_D. The density $\hat{s}_D$ was therefore the estimator of the likelihood maximum of s on S_D. The model considered to select in the collection was the one where the risk $\mathbb{E}[KL(s_D, \hat{s}_D)]$ was minimal, which was not possible to realise directly because this quantity depended on the unknown density s. For each estimator $\hat{s}_D$, the following decomposition of risk was considered:

$$\mathbb{E}[KL(s_D, \hat{s}_D)] = b_D + \mathbb{E}(V_D)$$

Where $V_D \coloneqq \int \ln(s_D, \hat{s}_D) s \, dx$ was a term of the variance and $b_D \coloneqq KL(s, s_D)$ was a term of the bias. Then the quantity V_D increased and the bias b_D decreased when the dimension D increased. The selection of models by penalised criterion consisted on the selection of the model of collection $(S_D)_{D\in\mathcal{D}}$ that minimised a criterion of the form:

$$crit(D) = \gamma_n(\hat{s}_D) + pen\,(D) \tag{6.24}$$

Also: $\widehat{b_D} \coloneqq \gamma_n(s_D) - \gamma_n(s)$ and $\hat{V}_D = \gamma_n(s_D) - \gamma_n(\hat{s}_D)$

Therefore, it was clear that according to the criterion in (6.8.1), the selected model was a minimizer of:

$$\gamma_n(\hat{s}_D) - \gamma_n(s) + pen(D) = \hat{b}_D - \hat{V}_D + pen(D) = KL(s_D, \hat{s}_D) + \left(\hat{b}_D - b_D\right) - \left(V_D + \hat{V}_D\right) + pen(D) \tag{6.25}$$

According to the law of big numbers,$\hat{b}_D - b_D \approx 0$. In addition, the arguments of the concentration enabled to show that the quantity $KL(s_D, \hat{s}_D)$ was close to its Esperance that was the risk of $\hat{s}_D$. In order for the criterion in (6.32) to be close to the risk $\mathbb{E}[KL(s_D, \hat{s}_D)]$, the optimal penalty was chosen such that $pen_{opt}^{(D)} = V_D + \hat{V}_D$

Also, the main hypothesis of the slope method was about assuming that $\hat{V}_D \approx V_D$. In order to justify this, it was observed that in the expressions V_D and $\hat{V}_D$, the measure of probability and the empirical measure played a symmetric role. When the two measures were exchanged in the expressions of V_D and $\hat{V}_D$ as well as in the expressions of s_D and $\hat{s}_D$, then V_D became $\hat{V}_D$ and reciprocally. This hypothesis led to consider a penalty of the form $pen(D) = 2\hat{V}_D$ and there was a need to determine $\hat{V}_D$ from the data. And by the definition of $\hat{V}_D$:

$$\hat{V}_D = \hat{b}_D + \gamma_n(s) - \gamma_n(\hat{s}_D)$$

However, for the models of big dimension, the term bias ended up being stabilised and the function $D \longmapsto \hat{V}_D$ was known through the quantity $-\gamma_n(\hat{s}_D)$. In many situations in statistics, the theorems of the selection of non-asymptotic models led to the consideration of the penalties proportional to $\frac{D}{n}$. In this case, it was easy to verify that the function $-\gamma_n(\hat{s}_D)$ was a linear function of $\frac{D}{n}$ for D sufficiently big. For a fixed η, the estimation $\hat{\eta}$ of the slope of this function led to the choice of the penalty: $pen(D) = 2\hat{\eta}D$

6.8.2 The slope method and the jumps across dimensions

In order to complete the description of the slope method, the overview of its classical definition is provided here in the form suggested by Massart [205].

Let's assume there was an interest to a penalty of the form $\eta - pen$ where *pen* was a function of D and n, and where ηa parameter to be fixed was. The slope method was based on the two following assertions:

1. There existed a minimal penalty $pen_{min} = \eta_{min}$in the family of the selected functions such that each lower penalty selected the models of big dimensions, and each higher penalty selected the models of reasonable dimensions. A chosen penalty of pen_{min} 2 order enabled to select an estimator of risk comparable to the oracle risk.

2. For penalties proportional to $\frac{D}{n}$, it was verified that this presentation of the slope method and the previous description were coherent by considering $pen_{min} = \hat{\eta}D$ as the minimal penalty, where $\hat{\eta}$ was the estimation of the slope defined above.

The two assertions were verified by Massart in the context of the regression on the white homoscedastic gaussian. Also, important progress was realised in this field few years ago and it was probable that this method was still available in many other situations.

6.8.3 Curves classification by gaussian mixture models

The method of selection of gaussian mixture models that was not totally described enabled to perform an unsupervised curves classification. To do that, data of the initial curves were projected in the space corresponding to the coefficients of a decomposition of Fourier, a decomposition in wavelets, or even a spline decomposition. This space was chosen for sufficient dimension Q so that the procedure selection of variables was really useful. In the context of the curves classification, the unordered collections were not used, if not in order to

downscale the number of considered variables to a dozen. Decompositions of Fourier and in wavelets presented the advantage of providing a schedule of variables and to compare the general trend of curves. It was natural to privilege lower frequencies or information of lower resolution. In practice, the use of those two types of decomposition was preferred. Therefore, in order to position this work in the context of the $S_{(K,v)}$ models collections defined earlier, the coefficients obtained were then centered and reduced in each variable. The slope method was applied to a collection of the mixture models in order to determine a classification of the initial curves.

6.8.4. Construction of oil platforms for biofuel production

In this section, the procedure of classification of curves was used to study a sample of profiles of oil extraction based on the *ricinus communis* oil destined to biofuel production. The first section was consecrated to the details of the construction of the database.

Some countries that had the experience in biofuel production from any type of oil published the monthly productions of the hundreds of oil storages they possessed. This situation was exceptional because for platforms already exploited, it was necessary to buy those information from companies that commercialise them.

In order to compare the forms of profiles without considering their amplitude, all curves were initially normalised by an estimation of the reserves of each storage. In practice, this operation was delicate. For example, a database oil storages dedicated for biofuel production did not always provide the estimations of the reserves of the platforms. However, oil storages of various bases, where they existed, were not always corresponding because some storages were not grouped in the same entity. Thus, an initial base of homogeneous curves representing the rates of production of each storage was obtained. Ideally, the procedure of classification was to be performed on the profiles of production of platforms whose exploitation was completed and this, in order to study the complete production curves. In reality, this was impossible to perform because the big majority storages in some platforms were still in production. Many small platforms only started to produce few years ago. In terms of big oil storages, even when they were exploited for many years, their production was not terminated. However, the beginning of the profiles was known and the first months of the production revealed important differences of the form between the curves.

For this work, only the 64 first months of the exploitation were considered. Also, all storages not producing during a long period were withdrawn from the initial sample. Some storages

whose production presented some anomalies (for example, stop of production due to accidents) were also eliminated from the base. This resulted to a family of curves homogeneous between them and defined for a period of 64 months.

6.8.5 Classification of production curves

In order to obtain a classification of production curves that was not influenced by technical variables, only a database constituted of wavelets coefficients was considered. However, there was no use of the complete wavelets vectors $\widetilde{W}_{.i}$ to apply the slope method. In fact, the use of all the wavelets coefficients would had led to consider the models of great dimensions related to the size of the sample, and in the estimation phase, many models could not be estimated. Wavelet coefficients corresponding to the two finest resolutions (j=1 and j=2) were removed. Because of the desire to classify curves based on their general form, it was natural to retain the wavelet coefficients associated to the largest scales for the analysis.

Furthermore, it was observed that in order to construct the classification, 6 variables out of the 16 selected for the analysis were sufficient, what comforted us in our choice to forget about the other coefficients of the decomposition. Lastly, it was noted that by reducing the observations of the dimension Q=16, there was a significant diminution of the necessary calculation times in order to perform the complete procedure of the slope method. In the procedure, the observed vectors were used:

$$\tilde{y}_{.i} = \left(\tilde{V}'_{6i}, \widetilde{W}'_{6i}, \dots, W'_{3i}\right)'$$

In this work, y_q designated the qth variable of the centered and reduced wavelet coefficient. Ideally, we would prefer to consider a collection of unordered models so that all the choices of blocs of variables of classification would be possible. Unfortunately, the collection would be then too rich for the procedure of selection of the model to be performed in a realistic calculation time. Therefore, we need to find a way to order the variables in order to go back to a collection of ordered models. A natural idea is to order the variables according to their resolution first, then according to the equal resolution time, meaning by keeping the order of variables inside the $\tilde{y}_{.i}$ vectors. We will use a second solution that consists of ranging the variables by decreasing order of their adequacy to a centered and reduced gaussian variable, and that enables to select the models of lower dimension. To do that, we perform an adequacy test for each wavelets variable to a centered and reduced gaussian variable. And finally, we

order the variables by decreasing order of statistics of test. We note by $y_{.i}$ the vectors obtained by permuting the variables in this manner in each of the $\tilde{y}_{.i}$ vectors.

The collection of $\mathcal{M}[LB_K]$ models was used, and this choice enables to have a family of distributions sufficiently rich while limiting the risks of decay during the procedure of the estimation. Let's note the hypothesis of the independence based on the wavelets coefficients is partly justified because the wavelets coefficients of the same level of resolution are less correlated. Let's also remember that the results demonstrated in previous sections do not necessitate that the true density should be in the collection of models used.

Figure 6.8 below illustrates the application of the slope method on the sample of vectors $(y_{.1}, \ldots, y_{.180})$. Let's also note that the cloud of points in figure 6.1 present the expected linear behavior in the big dimensions. The robust regression enables to calibrate the penalty, and the model finally selected is the one having 6 informative variables and 3 components in the mixture. The following table gives the effectives of the classification deduced from the rule of the posteriori maximum applied to the mixture part of the density $\hat{s}_{(\hat{K},\hat{v})}$.

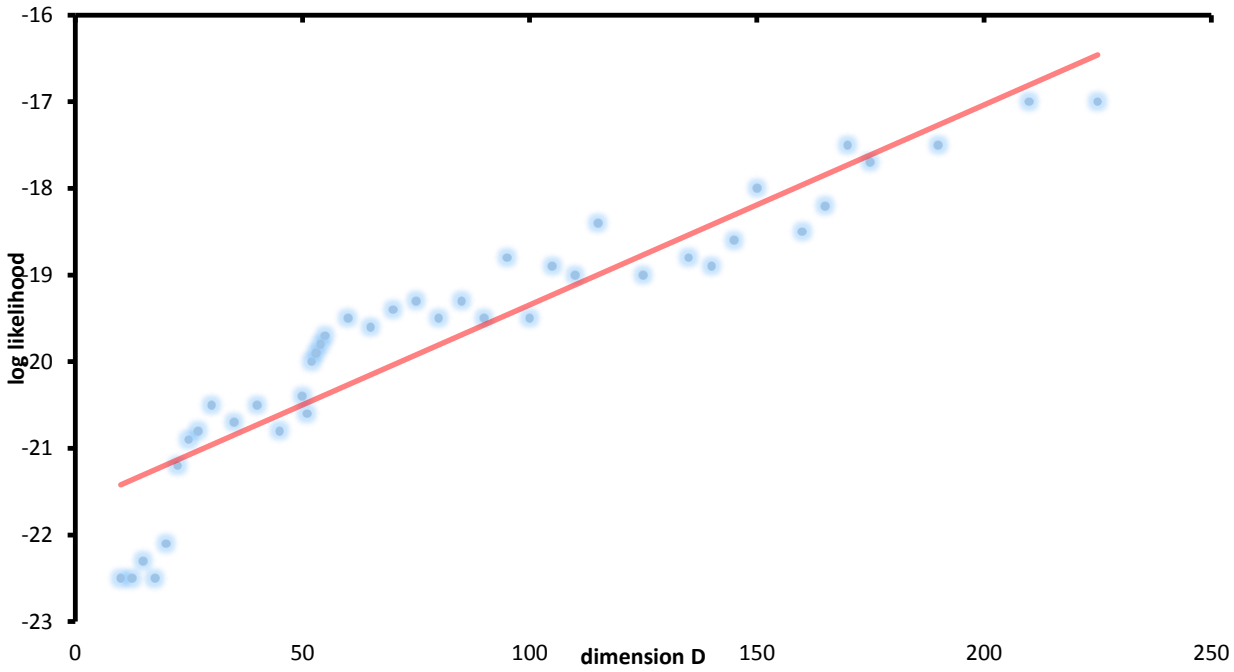


Figure 6.8: Slope method applied and composed of wavelet variables (without explanative variables).

Groups	1	2	3
Numbers	147	8	25

Then, it was possible to find which form of the mean profile corresponded to each of the classes obtained. Also, there was a possibility to compare the mean normalised profiles of each of the 3 classes. Those groups naturally corresponded to the rates of different productions. It was further assumed that the first class corresponded to big platforms, the second class contained storages of small sizes and the third class represented intermediary sizes. Thus, this reminded us of the principle raised in the presentation of the problem earlier in this work.

6.8.6 Classification of curves and selection of technical variables

The technical variables that explained the curves classification were determined. To do that, procedure used was so in order to simultaneously perform a curves classification, a selection of wavelet variables and a selection of technical variables. The 3 technical variables were added to the 16 wavelet variables, which were also centered and reduced. Therefore, there were 180 individuals and 19 variables. The vector that corresponded to the observation of the platform i was of the form $\left(y'_{.i}, y_{res\,i}, y_{dep\,i}, y_{den\,i}\right)$'. Then, a new collection of models was defined as follows: For a model, the set of indices of classification variables was noted by v. The collection of models used was composed of models for which the variables of classification were highlighted by the set v of indices of the form: $\mathrm{v} = \mathrm{v}_{tec} \cup \{1, \dots, v\}$, where v_{tec} was chosen in the set of parts $\{res, dep, den\}$. Thus, with regards to technical variables, all possible models were considered. In addition, v was the cardinal of v. For a given couple (K, v), there existed several models with K components and v variables of classification in the family. The application of the slope method was illustrated in figure 6.2 and figure 6.3 which enabled to verify that the range of parameters $(K, v) \longmapsto D(K, v)$ was well adjusted on the range of $(K, v) \longmapsto \gamma_n\left(\hat{s}_{(K,v)}\right)$ log-likelihood. The procedure led to the selection of a model with 4 components and 7 informative variables including one variable "Reserves" and 6 wavelets variables. This confirmed that the variable "Reserves" was coherent with the global classification meanwhile the variables depth and gravity were not selected. Thus, the principle in which the form of the normalised profile was validated based on the quantity of oil it contained. This time, the following numbers for the classification were obtained:

Groups	1	2	3	4
Numbers	23	37	119	1

The class 4 only contained one profile whose production was very quick. It was noted that the withdrawal of the unique curve of class 4 of the studied sample did not allow to diminish the number of selected classes. The incorporation of the variable “Reserves” into the procedure had the effect of reinforcing the grouping by classes of sizes as this was confirmed in Figure 6.4.

Discussion

The study of biofuel profiles as well as the details of the simulations in the annex illustrated the good proprieties of this new method of classification and of selection of simultaneous variables, which reinforced the theoretical results obtained in the previous sections. It was noted that it would be possible to demonstrate similar results for other types of gaussian mixtures than the three forms used here. Therefore, there was a large choice of collections of models and it was easy to verify that the range of parameters could be well adjusted on the log-likelihood range in the big dimensions. The example of waveforms presented in the annex illustrated this. The application of the slope method necessitated to sufficiently estimate models for which the linear “regime” of $\gamma_n(\hat{s}_D)$ is achieved. When a big number of components or variables of classification is necessary to observe the linear behavior $\gamma_n(\hat{s}_D)$ it is advised to use a richer collection of models in order to apply the slope method easily.

Furthermore, it was noted that the method used in this chapter could also be applied to the unsupervised classification without the selection of variables. The theoretical results obtained in the previous chapter suggested to consider the penalties proportional to the dimension, which in that case was uniquely a function of K. By considering the collection of models $\left(S_{(K,Q)}\right)_{K\geq 2}$, the proposed criterion enabled to select the number of classes. With regards to the interpretation of the model selected by our method, it was important to remember that the optimisation of the penalised criterion did not really correspond to the simultaneous research of a classification of data and of the bloc of all its coherent variables. In fact, it was decided to privilege the minimisation of the risk of estimation as the objective. Thus, the optimisation of the penalised criterion did not always reach the selection of all the coherent variables with

the classification selected. For example, in a situation where Q was of the order of n, some variables were not selected in order to limit the error of estimation in the selected model even though they were coherent with the classification. In the case of a unordered collection of models, it was confirmed that there was no bloc of variables v of the same size than $\hat{v}$ such that the likelihood of $s_{(\hat{K},v)}$was greater than the likelihood of $\hat{s}_{(\hat{K},\hat{v})}$. In this way, the bloc of variables $\hat{v}$ was optimal. In the ordered case, it was not always the case and variables were not ranged in a random manner in order to facilitate the interpretation of the selection of variables.

In terms of the curves classification, they constituted a privileged field of application of our method. In fact, the number of wavelets coefficients could be greater than n and in that case, the selection of variables took all its sense. Also, the fact that the wavelets variables were not able to explain the obtained classification moderated the problem of interpretation of the selection of variables that was presented in the previous section. Indeed, the main problem faced in the application of this new method of classification of curves was the ordering of the variables, which was compulsory because of the calculation times that were unrealistic in the case of unordered collections. The adequacy to a centered and reduced normal distribution was a possible criterion in order to order the variables: Therefore, the first variables were those whose empirical distributions moved away from the centered and reduced normal distribution.

With regards to the study of biofuel profiles, the principle by which the normalised profiles had a form that depended on the quantities of biofuel contained in the platform was validated. Only the variable “Reserves” was selected among the technical variables, which meant that the variable “Reserves” explained better the obtained classification than other technical variables.

6.9 Selection of a model of oil extraction

In this section, the slope method was applied in order to select a partition on which the visibility function of platforms was estimated. The first sub-section detailed the procedure of estimation on a fixed partition and the second was dedicated to the application of the slope method.

Proposition

The function q was strictly concave and admitted a unique maximum on its field of definition.

The proof of this proposition is given in the next sections. The calculation of the partial derivatives of q indicated that there was no explicit solution to the problem of the maximisation of this function. Even though the domain $[0,+\infty[^k \times]0,+\infty[$ was not compact, the previous proposition enabled to rigorously define the estimator of the maximum likelihood $\hat{g}_m^*$ in the S_m model. The concavity of the function q mostly guaranteed that it did not admit the local maximum. Therefore, the function h was regular on (A,b), which placed us on a favorable position to evaluate $\hat{g}_m^*$ by using a gradient based descent method. This method of probability was estimated by using an estimation by the Monte Carlo method. The performances of this procedure of estimation for a fixed partition were studied on the simulated samples in the annex section.

6.8.6.2 Slope method applied to *ricinus communis* oil extraction

In this section, the slope method was applied by using detection of the jumps of the dimension corresponding to the minimal penalty. In the situation of H_1 and H_2, it was suggested earlier to use a minimal penalty of the form:

$$pen_{min}(m) = \eta \frac{D_m}{n} \ln n \qquad (6.26)$$

Adapted to this context of this work, the algorithm occurred in the following manner:

1. For all η, to determine the model $\hat{m}(\eta)$ that minimised the criterion:

$$crit(m) = \gamma_n \, \hat{g}_m^* + \eta \frac{D_m}{n} \ln n \qquad (6.27)$$

2. To determine the constant η_{min} such that the selected models were of big dimensions when $\eta < \eta_{min}$ and of reasonable dimensions when $\eta > \eta_{min}$.
3. To choose the model $\hat{m}$ minimising the criterion in (6.5) for the penalty $2\,\eta_{min}$.

In order for this method to take the jumps of the dimension corresponding to the minimal penalty, it was preferable that the family contained some models of big dimension. In practice, this procedure was illustrated by a graph on which the constant of penalty η was marked on the x-axis and the dimension of the selected model $D(\eta)$ on the y-axis. The simulations presented in the annex indicated that the procedure enabled to select a model close to the oracle model.

Application to *ricinus communis* oil platforms

The biofuel platforms that we will study here are presented in annex A. Here, the proposition was that data of platforms A, B and C corresponded exactly to data used to estimate the α index of the Pareto-Levy distributions. For each platform, the estimations of the Pareto index α were modelled using the Lopez method of estimation. The visibility function was estimated for storages of sizes higher than x_0 using the slope method and this, in order to select a convenient partition on which h was estimated. Table 6.5 below provides the estimated visibility functions of the three platforms:

Table 6. 3: Estimations of the visibility function of storages for three platforms of biofuel production

Platforms	Selected partitions	Estimations of h A	b
A	[12 17 40.5 72.5 4482]	$2.63\ 10^{-3}$ $2.69\ 10^{-10}$ $2.4\ 10^{-3}$ $4.72\ 10^{-2}$	$4.26\ 10^{-3}$
B	[20 63.5 110 4139]	$1.15\ 10^{-11}$ 0.0012 $5.39\ 10^{-5}$	$1.99\ 10^{-2}$
C	[20 350 1250]	$4.35\ 10^{4}$ $1.16\ 10^{-5}$	$6.89\ 10^{-3}$

1. To start, it was shown that the function q was strictly concave. Because a strictly concave sum of function was also a strictly concave function itself, it was sufficient to show that for all $x \in [x_0, x_{max}]$ and all $\tau \in [0, t*]$, $(A,b) \longmapsto \ln\{h(A,b)(x)\} - \tau h(A,b)(x) - \ln P_{dec}(A,b)$ was strictly concave.

Now, let $x \in [x_0, x_{max}]$ and $\tau \in [0, t*]$. The function h(A,b)(x) was linear on (A,b), and it was deduced that $(A,b) \longmapsto \ln\{h(A,b)(x)\} - \tau h(A,b)(x)$ was strictly concave. It was sufficient to verify that $(A,b) \longmapsto -\ln P_{dec}(A,b)$ was concave. In addition, let (A,b) and (A',b') be two couples in $[0,+\infty[^k \times]0,+\infty[$ and let $\lambda \in [0,1]$. According to the convexity of the exponential function, then:

$$\exp\{-h(\lambda A + (1-\lambda)A', \lambda b + (1-\lambda)b')(x)t^*\} \leq \lambda \exp\{-\lambda(A,b)(x)t^*\} + (1-\lambda)\exp\{-h(A',b')(x)t^*\}$$

Thus:

$$\frac{\alpha x^{-\alpha-1}}{x_0^{-\alpha}-x_{max}^{-\alpha}}[1-\exp\{-h(\lambda A+(1-\lambda)A',\lambda b+(1-\lambda)b')(x)t^*\}] \geq \lambda\frac{\alpha x^{-\alpha-1}}{x_0^{-\alpha}-x_{max}^{-\alpha}}[1-\exp\{-h(A,b)(x)\}]$$

$$+(1-\lambda)\frac{\alpha x^{-\alpha-1}}{x_0^{-\alpha}-x_{max}^{-\alpha}}[1-\exp\{-h(A',b')(x)\}]$$

By integrating this inequality on $x \in [x_0, x_{max}]$, we obtained:

$$P_{dec}\,(\lambda A+(1-\lambda)A',\lambda b+(1-\lambda)b') \geq \lambda\, P_{dec}\,(A,b)+(1-\lambda)v\, P_{dec}\,(A',b')$$

Finally, the convexity of the logarithmic function was used to show that $-\ln P_{dec}$ was a concave function of (A, b), which terminated the proof of the strict concavity of q.

2. In order to show that q accepted a maximum, it was demonstrated that q was majored at 0. The function q was decomposed in the form: $q(A,b)=\sum_{j=1}^{q} q_j(a_j,b_j)$, Where b_j was defined by $b_j=\sum_{u=1}^{j-1} a_u(x_u-x_{u-1})+b_1$ and $q_j(a_j,b_j)=\sum_{i|X_i^*\in I_j}\{\ln h(X_i^*)-h(X_i^*)D_i^*-\ln(P_{dec})\}$. Furthermore, for all $h \in S_m$: $P_{dec}\,(h)=\frac{\alpha}{x_0^{-\alpha}-x_{max}^{-\alpha}}\int_{x_0}^{x_{max}}[1-\exp\{-h(x)t^*\}]x^{-\alpha-1}dx =$

$\frac{\alpha}{x_0^{-\alpha}-x_{max}^{-\alpha}}\sum_{j=1}^{k}\int_{x_{j-1}}^{x_j}[1-\exp\{-h(x)t^*\}]x^{-\alpha-1}dx$

And by increase of the function h: $P_{dec}\,(h) \geq \frac{\alpha}{x_0^{-\alpha}-x_{max}^{-\alpha}}\sum_{j=1}^{k}[1-\exp\{-h(x_{j-1})t^*\}]\int_{x_{j-1}}^{x_j} x^{-\alpha-1}dx \geq$

$\sum_{j=1}^{k}\frac{x_{j-1}^{-\alpha}-x_j^{-\alpha}}{x_0^{-\alpha}-x_{max}^{-\alpha}}[1-\exp\{-h(x_{j-1})t^*\}]$. In addition, by using the inequality $1-\exp u \geq u\exp(-u)$, for $u \geq 0$, we have: $P_{dec} \geq \sum_{j=1}^{k} c_j h\,(x_{j-1})\exp\{-h(x_{j-1})t^*\}$,

Where the c_j are strictly positive constants and that do not depend on a_j and b_j. This enabled that for all j, $q_j$was increased in the following manner:

$$q_j(a_j,b_j) \leq \sum_{i|X_i^*\in I_j}\left\{\ln\frac{h(X_i^*)}{\sum_{r=1}^{k} c_r \ln(x_{r-1})\exp\{-h(x_{r-1})t^*\}}-h(X_i^*)D_i^*\right\}$$

$$\leq \sum_{i|X_i^*\in I_j}\left\{\ln\frac{h(X_i^*)}{c_j h(x_{j-1})\exp\{-h(x_{j-1})t^*\}}-h(X_i^*)D_i^*\right\} \leq \sum_{i|X_i^*\in I_j}\left\{\ln\frac{h(X_i^*)}{h(x_{j-1})}+h(x_{j-1})t^*-h\,c_j-h(X_i^*)D_i^*\right\}$$

$$\leq \sum_{i|X_i^*\in I_j}\left\{\ln\left(1+\frac{X_i^*-x_{j-1}}{x_{j-1}}\right)+h(x_{j-1})t^*-\ln c_j\right\}.$$

The last inequality was obtained by using the fact that $h(x)=a_j(x-x_{j-1})+b_j$ in the interval I_j. In addition, h(x) approached 0 when (A, b) approached 0, which enabled to show that q_j was majored at 0. Also, it was verified that for all j, the function q_j approached $-\infty$ when $\|(A,b)\| \to +\infty$. In fact, for all I such that $X_i^* \in I_j$:

$\ln[a_j(X_i^*-x_{j-1})+b_j]-[a_j(X_i^*-x_{j-1})+b_j]\|(A,b)\| \to +\infty-\infty$

$h \to +\infty[x_0,x_{max}]\|(A,b)\| \to +\infty \lim_{\|(A,b)\|\to+\infty} P_{dec}=1.$

Hence the limit q_j announced earlier.

The function q was continuous and majored at 0 and approached $-\infty$ when $\|(A,b)\|$ approached $+\infty$. Therefore, it was deduced that the function accepted a maximum at $[0,+\infty[^k \times]0,+\infty[$. The unicity of the maximum came from the strict concavity of the q function.

Conclusion of the chapter

In this chapter, procedures enabling the exploitation in practice of the theoretical results obtained in the previous chapter were developed. In terms of biofuel profiles, it has been demonstrated that the normalised profiles can be deformed based on the ultimate reserves of the platform, using the original method of classification of curves. And this justified the introduction of a model for the individual production of storages in which the production directly depended on the quantity of *ricinus communis* oil contained on the platform. In terms of oil extraction, the estimation of the visibility which constituted the main challenge allowed the introduction of extensions of the extraction process.

Chapter 7: Modelling of the industrial biofuel production

In order to model the complete biofuel production at an industrial level, and based on the extracted/stored *ricinus communis* oil, a basic model of biofuel production based on individual *ricinus communis* oil storages was proposed. The second section of the chapter focuses on the definition of the production model as well as on the study of the impact of various factors under the form of the production curves. The final section focuses on the biofuel industrial production.

7.1 Modelling of the biofuel production based on individual Ricinus Communis oil storages

In this section, a simple parametric model of biofuel production based on individual *ricinus communis* oil storages as time function and reserves variables was proposed.

7.1.1 Presentation of the model

By hypothesis, it was suggested that the *ricinus communis* oil storages that had big oil quantities were produced in a slower manner compared to those that had small quantities.

Figure 7.1 illustrates this phenomenon. For a sample of production curves, the durations of production of biofuels, and the pic levels of production were presented in the two graphics based on the quantities contained in each storage, and according to the scales of logarithm.

It was noted that 90% of the corresponding storages were always in a production state, and the durations of production considered were mainly estimations. However, the pic level of production based on a specific *ricinus communis* oil storage was quickly achieved, and very few pics of production necessitated to be estimated in this sample. It was easy to adjust a line on the cloud of points of the lower graphic, which suggested that the pic level was a power of the reserves variable. The adjustment was of less quality on the higher graphic but like it was explained earlier, data used in this case were estimated.

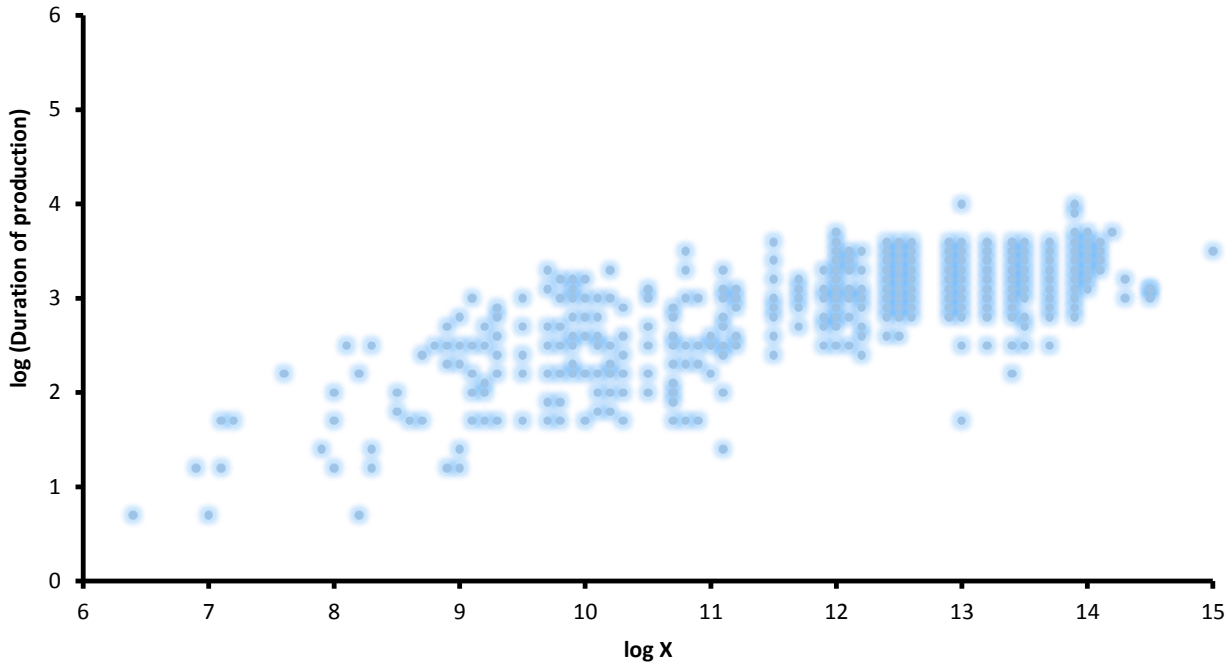


Figure 7.1(a): Logarithms of duration of production based on the quantities of ricinus communis oil stored, for a sample of 250 storages.

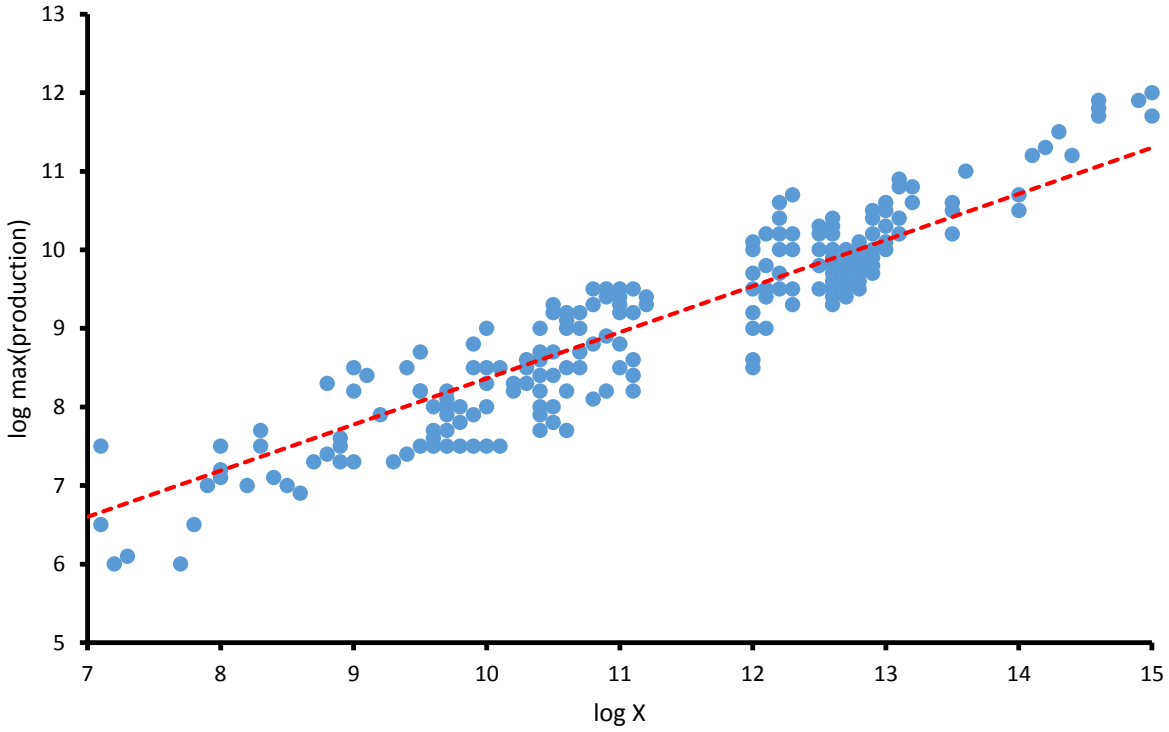


Figure 7. 1(b): Logarithms of maximal productions based on the quantities of *ricinus communis* oil stored, for a sample of 250 storages

The quantity of *ricinus communis* oil, noted x, as well as the time t were deterministic variables.

Let prod(x,t) be the random variable of production at time t of a quantity x of Ricinus Communis oil. The previous remarks lead us to suppose that:

$$\mathrm{E}\left(\frac{1}{x}prod(x,t)\right) = \frac{1}{x^{\beta}}K\left(\frac{t}{x^{\beta}}\right) \quad (7.1)$$

Where the function K was called basic profile. The coefficient β, called coefficient of inertia controlled the speed of biofuel production based on the *ricinus communis* oil quantities. Figure 7.2 allowed the validation of this model. On this model, the ${x_i}^{\beta-1}prod(x_i,t_{ij})$ quantities based on the $\frac{t_{ij}}{x_i}$ points were considered. The quantity $prod(x_i,t_{ij})$ corresponded to the production of the i[th] *ricinus communis* oil storage at time t_{ij}. The coefficient of inertia β was randomly estimated by using the linear regression [212] performed on the cloud of points of the lower graphic of figure 7.1, which gave β=0.25. The form K of the basic profile clearly designed itself through the cloud f points $\left({x_i}^{\beta-1}prod(x_i,t_{ij}),\frac{t_{ij}}{x_i}\right)$. Figure 7.2 suggested that noise acted in a multiplicative manner on the prod variable. Thus, the following model of production was reached:

$$\frac{1}{x}prod(x,t) = \frac{1}{x^{\beta}}K\left(\frac{t}{x^{\beta}}\right)\tilde{\varepsilon}_{x,t} \quad (7.2)$$

Where $\tilde{\varepsilon}_{x,t} := \log(\tilde{\varepsilon}_{x,t})$ was a centered Gaussian variable of variance σ^2.

Supposing that noise acted additively and not multiplicatively, this slightly modified the β estimation. However, the obtained residuals for the additive model presented a stronger heteroscedasticity than for the multiplicative case.

7.1.2 Principle of estimation

Per the logarithm in (7.2), the model is written as the following form:

$$\ln\frac{prod(x,t)}{x} = -\beta\ln x + \mathrm{H}\left(\frac{t}{x^{\beta}}\right) + \varepsilon_{x,t} \quad (7.3)$$

With: H:=ln K. Productions of *ricinus communis* oil storage i (of quantity x_i) were known at times $t_{i1},\dots,t_{ij},\dots,t_{iq(j)}$, where q(j) corresponded to the number of years during which *ricinus communis* oil storage produced. And to each possible (i,j) couple, a new index u=u(i,j) was associated. For the available observations, equation (7.3) was written for available observations under the form:

$$z_u = -\beta\ln x_u + H\left(\tau_u^{\beta}\right) + \varepsilon_u \quad (7.4)$$

Where $z_u := \ln\frac{prod(x_i,t_{ij})}{x_i}, x_u := x_i, \quad \tau_u^{\beta} = \frac{t_{ij}}{x_i^{\beta}}$

Random variables ε_u were i.i.d of centered Gaussian law of variance σ^2. The indexing did not make a distinction between various *ricinus communis* oil storages. In fact, it allowed the repositioning in a more classical statistical estimation context.

In addition, the spline interpolation method [209] was used to estimate H in the regression model of (7.3). A cubic spline function was a piecewise polynomial continuous function of degree of at least 3 defined on a nodes delimited interval partitioning, such that the partition would had its first and second continuous derivatives. An occurring problem in the use of cubic splines was their tendency to explode at the two boundaries of their interval of definition. The natural cubic spline functions allowed the resolution of this problem by imposing to the function to have a linear behavior beyond the two boundaries. For a set of k fixed nodes, the set of natural cubic spline functions defined on the nodes formed a vector space of dimension k.

Thus, for the remaining of this work, the number of nodes k was supposed fixed and the nodes ξ were regularly spaced on $[0, \tau_{max}]$, where τ_{max} was a maximal value for the observed τ_u. Also, let $m = (m_r)_{r=1,\dots,k}$ be a base of the space $S_{cn}(\xi)$ for the natural spline functions built on the ξ nodes. A function $H \in S_{cn}(\xi)$ was therefore written under the form: $H = \sum_{r=1}^{k} \theta_r m_r$, with $\theta \in \mathbb{R}^k$. In order to avoid that the estimated function over-adjusts the observations, the following least squares criteria is minimised in $H \in S_{cn}(\xi)$ and in β:

$$crit(H, \beta, \lambda) = \sum_{u=1}^{N} \left\{ Z_u + \beta \ln x_u - H\left(\tau_u^{(\beta)}\right) \right\}^2 + \lambda \int_0^{\tau_{max}} \{H(\tau)\}^2 \, d\tau \quad (7.5)$$

This transformation excluded null productions that were observed when the exploitation was finished. For a sample of *ricinus communis* oil storage used here, it was considered out of twenty storages, and the impact on the estimation of K was very weak. In order to avoid the exclusion of data, the model with the additive noise was used by keeping in mind that the modelling was less realistic. Then, the H functions that presented more pronounced oscillations were penalised by the second term; in general, the function chosen by this criteria was called smoothing splines [210]. In the following, a detailed procedure was adopted from [179] into this particular case by mixing the H and β estimations. Let's define the matrix $M^{(\beta)} \mathrm{X}\, M_{ur}^{(\beta)} = m_r\left(\tau_u^{(\beta)}\right)$

Let's pose: $W_u \coloneqq Z_u + \beta \ln x_u$. The criteria in (7.5) was expressed under the following matrix form:

$$\mathrm{crit}(\theta, \beta, \lambda) = \left(\mathrm{w}^{(\beta)} - \mathrm{M}^{(\beta)}\theta\right)^{\mathrm{t}}\left(\mathrm{w}^{(\beta)} - \mathrm{M}^{(\beta)}\theta\right) + \lambda^{\mathrm{t}}\theta\Omega^{(\mathrm{m})}\theta \quad (7.6)$$

where the matrix $\Omega^{(m)}$ was defined according to the base m by:

$$\Omega_{rl}^{(m)} = \int m_r''(t) m_l''(t)\ dt$$

For λ and β parameters, the criteria in (7.6) was a minimum in:

$$\hat{\theta}(\beta,\lambda) = \left(M^{(\beta)}M^{(\beta)} + \lambda\Omega^{(m)}\right)^{-1} M^{(\beta)} w^{(\beta)} \qquad (7.7)$$

Minimising (7.6) in β was performed in a numerical way by calculating the $crit\ (\hat{\theta}(\beta), \beta, \lambda)$ on a fine grid of the β values; $\hat{\beta}(\lambda)$ the corresponding optimal β. Also, in order to adjust the λ parameter, it was preconized to perform a cross-validation in order to minimise the risk of prediction. The sample used here had a length of approximately 2000. And it was sufficient and simple to randomly separate the sample into two distant blocks of data: The first block $B_1$was used to calculate the $\hat{\theta}$ and $\hat{\beta}$ estimators for various values of λ. The risk corresponding to each choice of λ was estimated by using the second block of data B_2. The optimal smoothing parameter that minimised the estimated risk was chosen:

$$\hat{R}(\lambda) = \sum_{u\in B_2}\left[Z_u + \hat{\beta}x_u - \hat{H}\left(\tau_u^{(\hat{\beta})}\right)\right]^2, \text{ where: } \hat{H} = \sum_{r=1}^{k}\hat{\theta}(\hat{\beta},\lambda)m_r \qquad (7.8)$$

7.1.3 Results

It was assumed that the number of nodes was fixed at k=50. It was not useful and was more expensive to rely on a fine mesh for the time calculation. Therefore, the smooth splines function of the R software was used to estimate the fixed H to β. This R procedure presented the advantage of giving the user the possibility to address the number of nodes on which the natural cubic splines relied on. The proposed linear regression of Figure 7.1 allowed to precise a reasonable interval in which research $\hat{\beta}$ was to be researched in order to minimise the criteria that was performed for β between 0.1 and 0.4.

As indicated in Figure 7.2, $\hat{\beta} = 0.27$, and the spline function $\hat{H}$, corresponding to $\hat{\beta}$ was plotted on the top right of the graph. Residuals corresponding to the selected $(\hat{\beta}, \hat{\theta})$ couple were also plotted. When the temporal zone corresponding to the beginning of the production was excluded, the observation of residuals indicated that this simplistic model explained well the data observed. According to the model, here the noise was multiplicative and when the production was low, the variance was theoretically low as well.

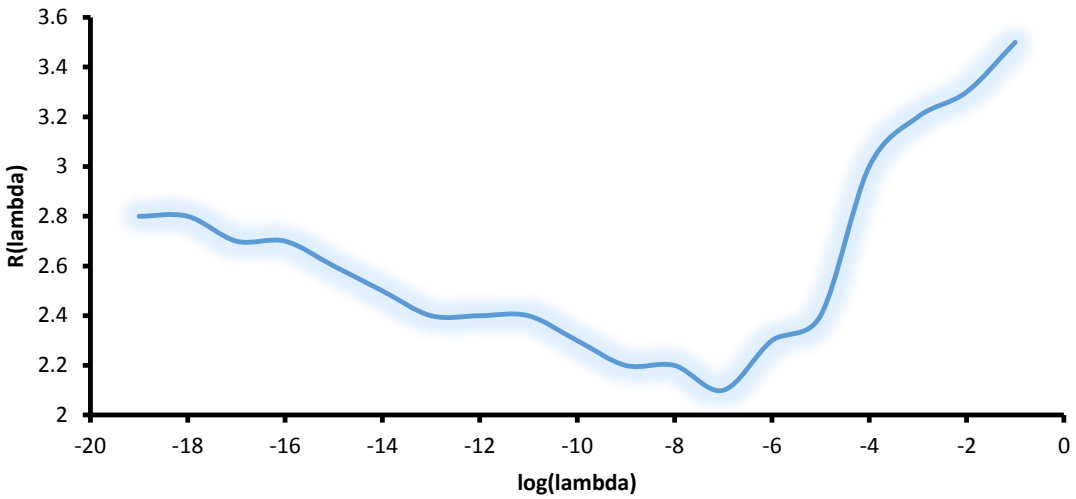


Figure 7. 2: $x_i^{(\beta-1)}$ prod(x_i,t_{ij}) quantities based on $\frac{t_{ij}}{x_i}$ points for the set of available *ricinus communis* oil storages. The prod(x_i,t_{ij}) quantity corresponds to the production of a storage of quantity x_i at time t_{ij}.

The observed quantities here were cumulated per year of production. And the exploitation of *ricinus communis* oil storage took place at the beginning or at the end of the first year of production. This explained the important variance of the residuals at 0.

Also, the basic profile was naturally estimated by $\widehat{K} = \exp(\widehat{H})$. This function was represented by Figure 7.4. It was interesting to note that beyond the pic, the decrease of this function was attenuated by a slight rebound at the beginning of the production. This can be explained by the techniques of secondary and tertiary recuperations put in place in one part of the storage of the sample which allowed to access new reserves in the storage. It was satisfying that the estimation of the basic profile K integrated this phenomenon by this rebound of production. In addition, according to the model defined in (7.1), the basic function was an integral function of 1. Meanwhile, it was known that the procedure detailed above did not impose this constraint which would complicate the estimation of H. To address this challenge, it was necessary to restraint the support of the function $\widehat{K}$ such that it verified the constraint. Finally, the couple $(\widehat{K}, \hat{\beta})$ that was used in the following sections for the production modelling of the set of *ricinus communis* oil storages.

7.2 Profile of production of r*icinus communis* oil total storage (platform)

In this section, the model of production of oil total storage to be used was completely defined. Let $(X_1, \dots, X_{N_B})$ be the sample of the quantities of *ricinus communis* oil storages; and let $(D_1, \dots, D_{N_B})$ be the sample of the dates of the corresponding extraction, where N_B designated the total number of storages contained in the platform. For $u \geq 1$, $(L_u)_{u\geq 1}$ was the suite of dates of production. Also, let Y_u be the reserves of the storages of the platform where the production was activated at the date L_u. The whole number u(i) designated the rank of the activation of production of the platform i; and if it had never been produced, then by convention $u(i) = +\infty$ with $L_\infty = +\infty$. If $u(i) < \infty$, then $Y_{u(i)} = X_{i(u)}$. The stock of storages available at date t had the following mathematical expression:

$$\mathrm{Stock_t} = \left\{\mathrm{i} \in \{1, \dots, \mathrm{N_B}\} | \mathrm{D_i} \leq \mathrm{t} < \mathrm{L_{u(i)}}\right\} \tag{7.9}$$

In general, we call politics of production, a function of L_u and of $Stock_t$ (and eventually other external parameters) that determine the storages to be selected within the $Stock_{L_u}$ at date L_u [211]. The following synthesis enumerated the set of hypotheses based on which the model of the production of the platform relied on.

1. Reserves X of a *ricinus communis* oil storage of the platform followed a Pareto-Levy law $Par(\alpha, 1, x_{max})$.

2. The storages were selected according to the following model:
The dates of selection of storages of quantity higher than a certain level x_0 are of the conditional law $(D|X) \sim \varepsilon(h(X))$ where h was the function of visibility of the storages. Storages of quantity lower than x_0 were selected according to a process of Poisson homogeneous in time.

3. The storages of the selected stock are developed according to a politics of production.
4. A storage that was exploited produced its reserve according to the model:

$$\mathrm{prod(x,t)} = \mathrm{x^{1-\beta} K}\left(\frac{\mathrm{t}}{\mathrm{x^{\beta}}}\right) \tag{7.10}$$

Where the basic profile K and the inertia coefficient β were estimated in the previous section. Under this set of hypotheses, the production of the total platform had the following mathematical expression [211]:

$$\mathrm{Prod(t)} = \sum_{\mathrm{u=1}}^{\infty} \mathrm{Y_u^{1-\beta} K}\left(\frac{\mathrm{t-L_u}}{\mathrm{Y_u^{\beta}}}\right) \tag{7.11}$$

Also, it was assumed that X followed a non-restraint Pareto-Levy law $Par(\alpha)$. In order to ensure that the production Prod(t) was integrable on all t, additional hypotheses on the law of (X,D) and on β were needed. The following simulations aimed at determining the mean production curves. In order to avoid any discussion about the integrability of Prod(t), it was assumed that X was increased. For the extensions of the proposed production curves in the next section, this hypothesis was natural because for a mature platform, the quantity of the biggest storage was known. In the objective to determine the mean production curves, it was not necessary to incorporate the multiplicative noise in the model of individual production of the storages that were used here. Thereby, the impact of the politics of production of the storages and production curves of *ricinus communis* oil storages was studied in the following sections.

7.2.1 Impact of the politics of production

In order to study the influence of the stock management on the form of the production curves of the whole platform, many scenarios of management were considered in the protocols of simulation that follow. For each of the protocols, 100 simulations were performed in order to evaluate a mean production profile of the platform. In addition, at each date L_u, a storage was chosen in the population of $Stock_{L_u}$ according to a certain criterion. In fact, it was difficult to faithfully model the politics of section of a storage within the same stock. Therefore, decisions related to the production were influenced by many economical parameters meanwhile, in the model, only the quantities of *ricinus communis* oil reserves as well as their date of extraction were known. However, for evident economic reasons, it was evident that there was a great preference to produce storages of the stock that possessed the greater quantities of reserves.

Protocol 1: The suite of dates L_u was modelled by a Poisson process of intensity λ, constant over time. Table 7.5 (a) details the characteristics of the simulated process that were common for all simulations performed. Indeed, at each date L_u, a storage was selected in the stock according to the following modes of draws:

- ✓ Unbiased draw;
- ✓ Biased draw according to a power γ of the quantity (here, γ=1 and γ=3).

Selection of the storage with the biggest quantity of *ricinus communis* oil.

Mean production curves corresponding to each of the situations were represented in Figure 7.3 (c). First, it was observed that the more the draw within the stock advantaged storages

with big quantities, the more the pic appeared earlier and with a great amplitude. However, if the bias was low, the production was globally better distributed the first years. Also, it was observed that for powers $\gamma \geq 1$, the platform profiles were identical.

Protocol 2: To start, many scenarios of intensity constant over time that were considered, for the same distributions of reserves and exploitation (See table 7.6 (a)). Thereby, Table 7.6 and figure 7.4 (c), presented the results obtained in each scenario. It was observed that for these simulations, the visibility and the intensity μ_0 were sufficiently elevated so that the stock will never be exhausted during the exploitation of the platform. The increase of the intensity had an important effect on the curve as long as the levels of intensity remained low. For high levels of intensity, the impact of the intensity on the production curves was much less important and the cumulated productions in table 7.1 confirmed this observation. The distribution of reserves was at the origin of this phenomenon. In fact, for $\alpha = 0.75$, about 70% of reserves were contained in the set of storages of quantities higher than 100 Mb. These storages were only a few dozen and the assumption was that most of them had been extracted the first 30 years. The fact that the intensity of the production of storages was increased, it could only slowly increase the reserves available for the production. Also, Figure 7.4(c) indicated that the increase of λ had the effect of producing most reserves earlier, which naturally accentuated the decline of biofuel production beyond the pic.

$N_B = 2500$

$\alpha = 0.75$

h: $\begin{cases} Partition: [20, 70, 110, 5000] \\ A = [10^{-3}, 210^{-3}, 710^{-5}] \\ b = 10^{-3} \end{cases}$

$\mu_0 = 6$

Intensity of the production: $\lambda = 6$

(a) Parameters of the protocol

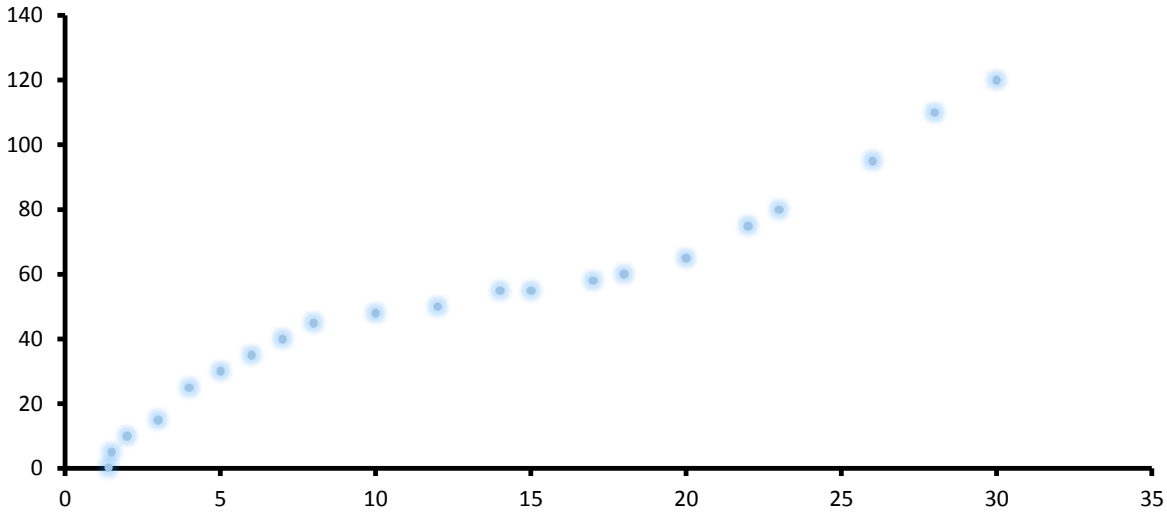


Figure 7.3(a): Impact of the politics of selection (Protocol 1): Associated visibility function.

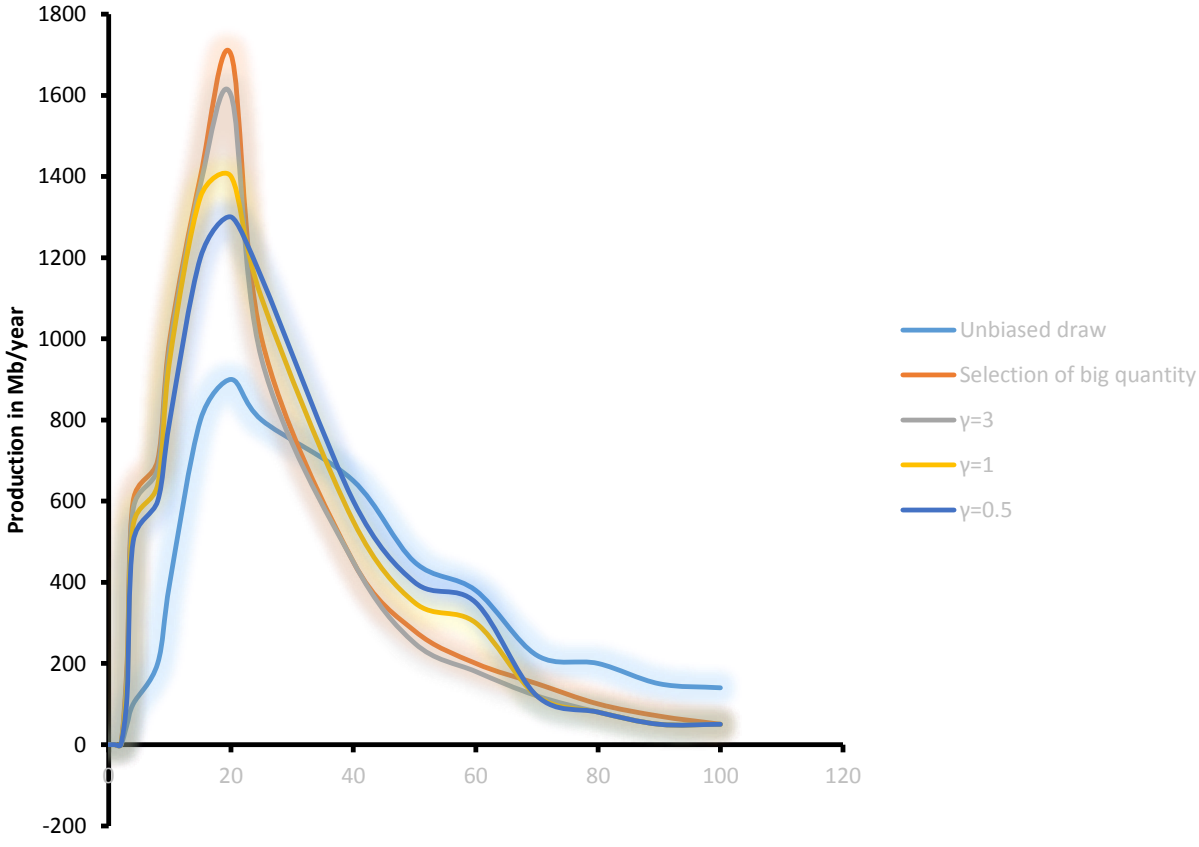


Figure 7.3(b): Mean production curves for various draws within the stock

$N_B = 2500$

$\alpha = 0.75$

h: $\begin{cases} Partition: [20, 70, 110, 5000] \\ A = [10^{-3}, 10^{-3}, 10^{-4}] \\ b = 2\ 10^{-4} \end{cases}$

$\mu_0 = 6$

Draw in the stock: $\gamma = 2$

(a) Parameters of the protocol

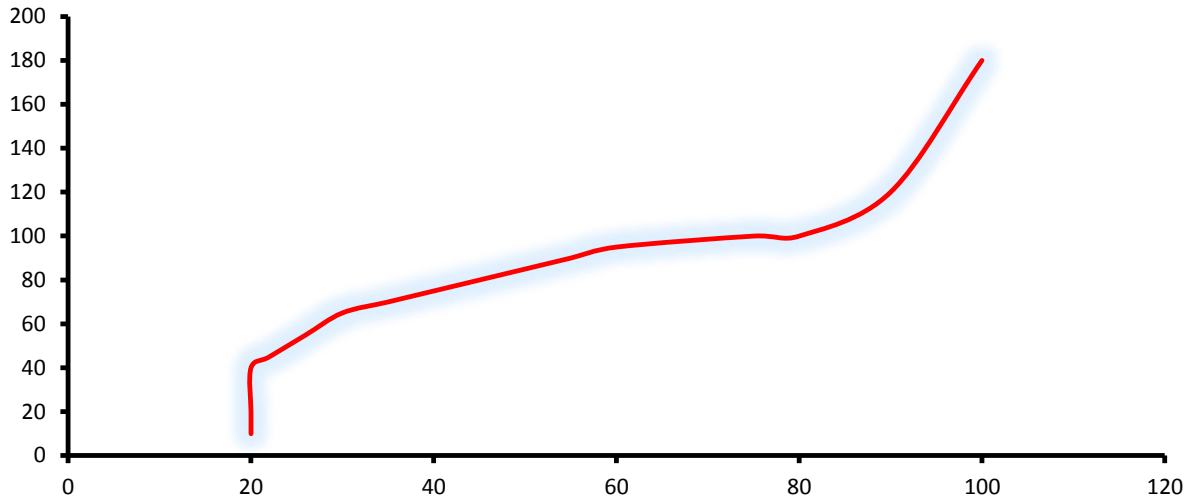


Figure 7.3(c): Associated visibility function derived from stock: $\gamma = 2$

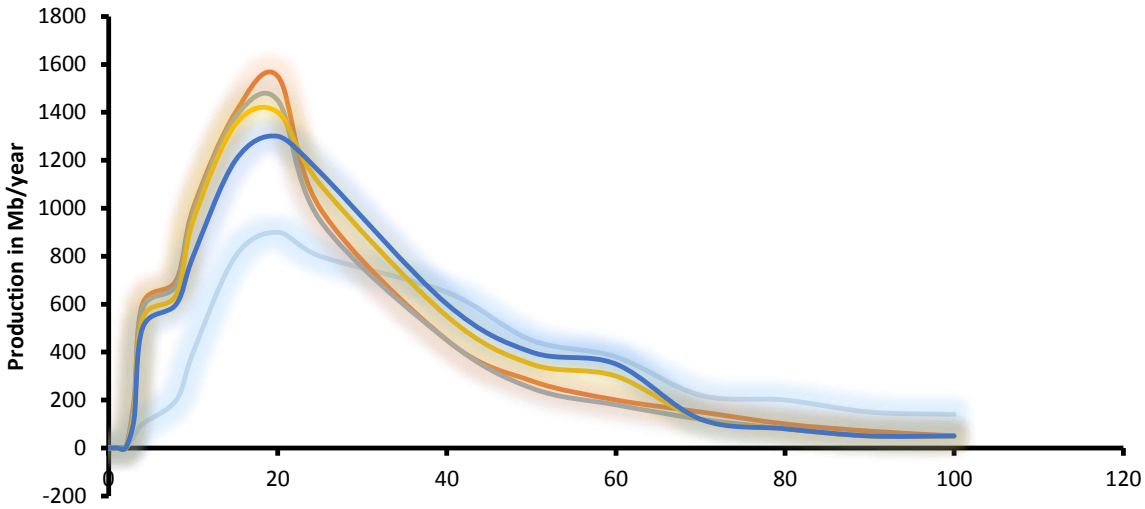


Figure 7.4(a): Impact of the production intensity (Protocol 2): Mean production curves for various hypotheses of production intensity.

Table 7. 1: Description of productions of the platform for scenarios of various intensity. The period of production considered was [0, 70].

λ	Mean quantity of storages produced	Mean cumulated production ($\mathbf{10^9\ b/year}$)
2	140	373
6	420	440
10	700	455
15	1050	465
20	1400	469

Protocol 3: In this protocol, the number of storages increased over time in order to study the way the decay can be attenuated beyond the pic of production. This scenario was modelled by an inhomogeneous intensity $(t) = \tilde{\lambda} + \frac{t}{2}$. Here, $\tilde{\lambda} = 6$. And table 7.7 (a) gave the values of the other parameters for this example. Under these hypotheses, the number of storages selected each year for production increased up to 40 storages per year at the end of 70 years. In total, more than a thousand of supplementary storages were activated for production with regards to the situation where the intensity remained at $\tilde{\lambda} = 6$. Graph 7.7 (c) allowed the comparison of two mean production curves corresponding to two situations λ and $\tilde{\lambda}$. In reality, this important effort of production managed to increase the cumulated production by approximately 7%, and the speeds of decay were comparable to the levels of production slightly higher in the case of the intensity λ. Therefore, it was not sufficient to increase the effort of production of storages in order to compensate the decay of the production beyond the pic.

Protocol 4: In this experiment, four types of simulations were performed in order to measure the effect of the increase of visibility of storages combined with an increase of effort of production. The detail of the protocol of simulation is presented in Figure 7.8. The effect on the curve of production was more important than in the situation of protocol 2 where the effort of production was identical for all the intensities. But the increase of biofuel activity (production) had the effect of advancing the pic date and accentuating the speed of the decay beyond it, without the production of many more storages. Scenario 4 mobilised four times the quantity of storages than scenario 1 but produced only 20% more than the first scenario.

7.2.2 Impact of the law of reserves

In order to illustrate the influence of the distribution of reserves under the form of the profile of the platform, the following simulation schema was introduced:

Protocol 5: The parameters of the protocol were given in table 7.9 (a), the parameter α of the Pareto-Levy varied between 0.3 and 0.9 in order to describe the situations of concentrations of the habitat very differently. Figure 7.9 (c) presents the mean production curves obtained.

$N_B = 2500$

$\alpha = 0.75$

h: $\begin{cases} Partition: [20, 70, 110, 5000] \\ A = [5.10^{-4}, 5.10^{-4}, 5.10^{-5}] \\ b = 10^{-4} \end{cases}$

$\mu_0 = 20$

Draw in the stock: $\gamma = 2$

(a) Parameters of the protocol

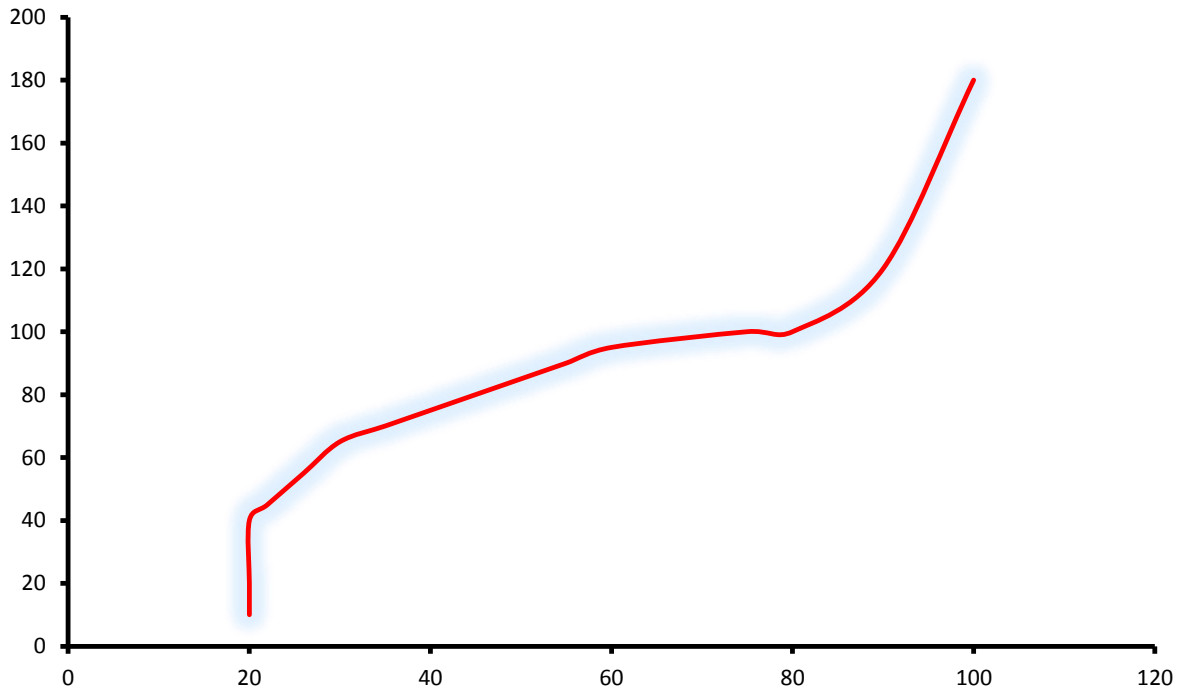


Figure 7.4(b): Associated visibility function

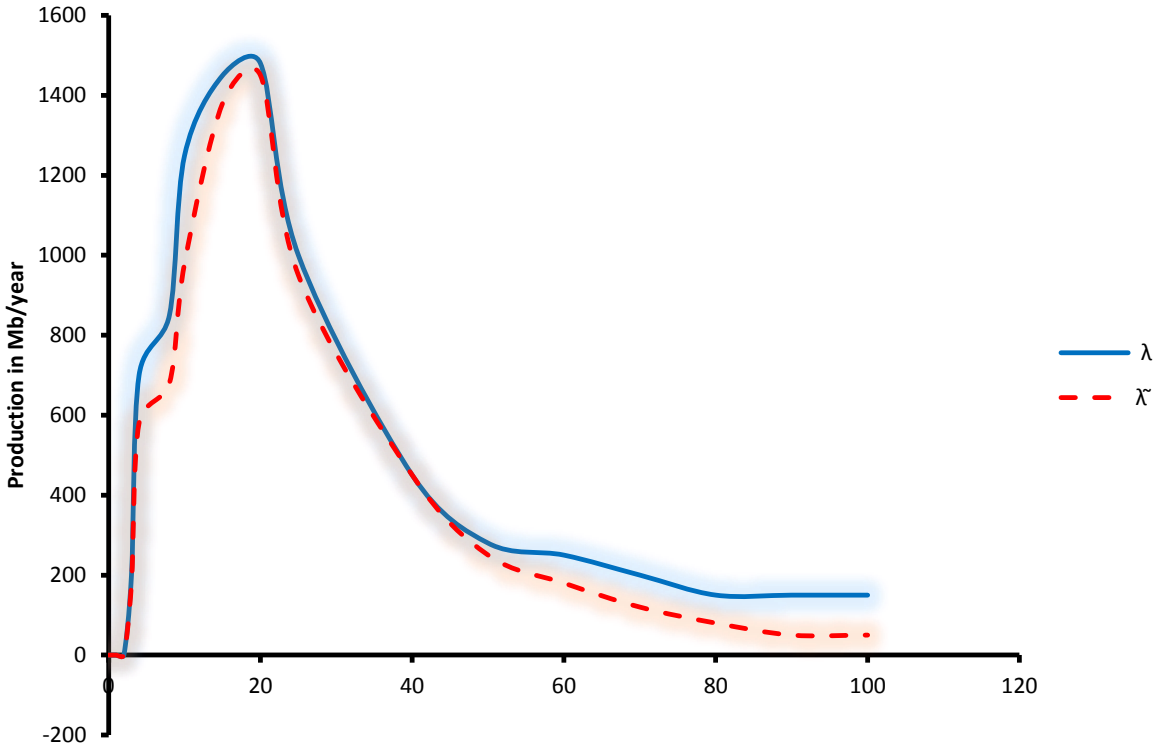


Figure 7.4(c): Mean production curves of the platform for the two hypotheses of intensity

$N_B = 2500$

$\alpha = 0.75$

h: $\begin{cases} Partition: [20, 70, 110, 5000] \\ A = [5.10^{-4}, 5.10^{-4}, 5.10^{-5}] \\ b = 10^{-4} \end{cases}$

Draw in the stock: $\gamma = 2$

(a) Parameters of the protocol

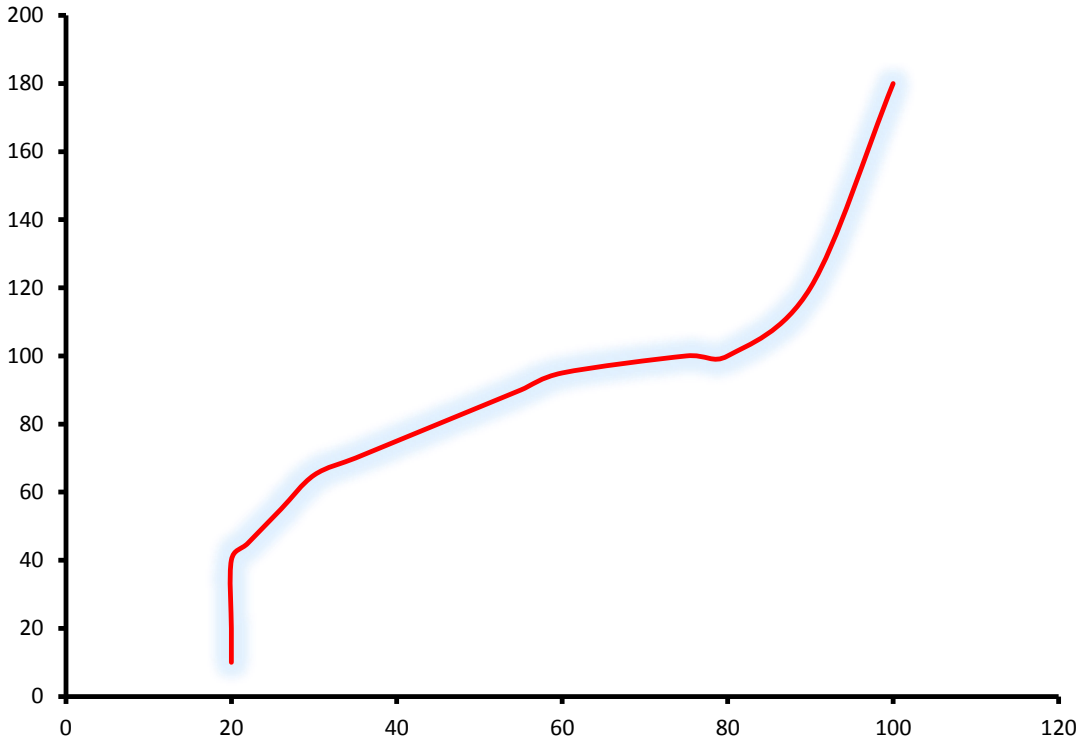


Figure 7. 5(a): Impact of the intensity of production (Protocol 3): Associated visibility function

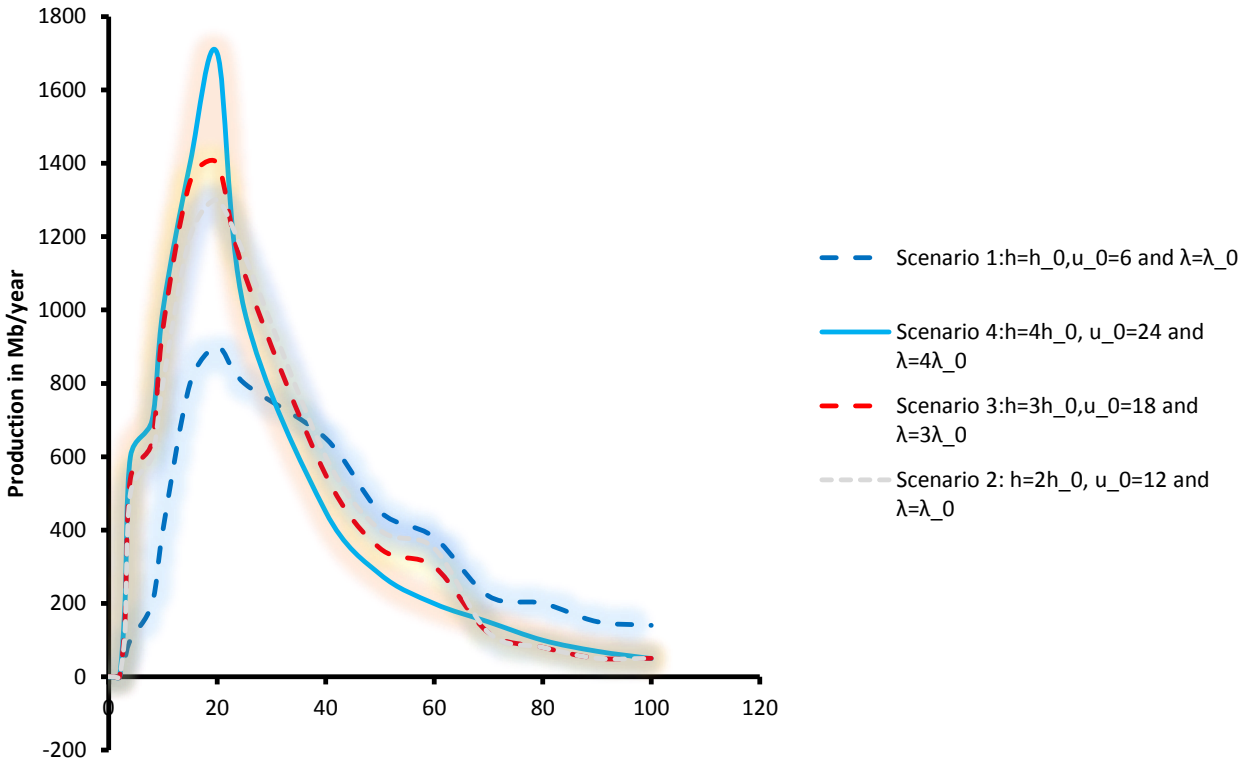


Figure 7.5(b): Mean production curves of the platform for production scenarios of various intensities

Firstly, it was observed that the amplitude of the production curve increased with the concentration of the habitat, meaning when α decreased. This was natural because the Esperance of the random variable of the Pareto Levy distribution of parameter α restrained to $[1, x_{max}]$ was a decreasing function of α. However, once those curves were normalised by the value of their maximum, they globally had the same allure, as this was shown in Figure 7.10. Finally, it was observed that the pic of the production appeared later for a concentrated habitat. This phenomenon was natural because there were many *ricinus communis* oil storages with big quantities when α was low, and the production of these oil storages with big quantities was extended more and more over time only when there were some of them, and this had the effect of delaying the passage of the pic.

7.3 Extension of the production of a platform

In this section, we propose scenarios for the production of biofuels in the platforms of current production. In order to be able to introduce this extension, we must first estimate the intensity of the process of extractions of small quantities *ricinus communis* oil storages, as well as the number of storages with quantities higher than x_0 remaining in the platform.

7.3.1 Estimation of the intensity of the process of extraction of small quantities storages

Let Q_p be the number of storages with quantities in $[1, x_0]$ and extracted during the year p. We note t^* the current year, then $(Q_1, \dots, Q_{t^*})$ forms a sample of random variables of Poisson distribution of parameter of intensity μ_0 of the process of extraction. The estimation of the maximum likelihood of μ_0 has the following expression:

$$\hat{\mu}_0 = \frac{1}{t^*} \sum_{p=1}^{t^*} Q_p \qquad (7.12)$$

Table 7.2 presented the estimations obtained for the zones of production. Let's recall that a description of these platforms is available in Annexure A.

Table 7. 2: Estimations of the intensities of the process of Poisson in the class of small quantities of extraction for the simulated platforms

Ricinus Communis oil platforms	I_0	t^*	$n_0(t^*)$	$\hat{\mu}_0$
Platform A	$[1, 12]$	42	104	2.36
Platform B	$[1, 20]$	31	226	6.85
Platform C	$[1, 20]$	39	103	2.5

7.3.2 Number of storages with quantities higher than x_0

Let $(X_1, \dots, X_N)$ be a sample of random variables of Pareto Levy distribution restrained to $[x_0, x_{max}]$. $\forall\, i \in \{1, \dots, N\}$, or D_i the random variable $(D_i|X_i) \sim \varepsilon\big(h(X_i)\big)$ with $h \in \mathcal{H}_m$of the conditional law, where $m \in (I_1, \dots, I_k)$ is a partition of intervals of $[x_0, x_{max}]$. As previously, X_i and D_i represent the quantity and the date of extraction of the storage i respectively. Let n be the number of storages of the sample extracted before the date t^*. Then:

$$n = \sum_{i=1}^{N} 1_{D_i < t^*} \tag{7.13}$$

Let N_j be the total number of storages of class I_j buried in the platform, and n_j designates the number of storages of this category extracted before t^*. Similarly, we can write:

$$n_j = \sum_{i|X_i \in I_j}^{N} 1_{D_i < t^*} \tag{7.14}$$

From (7.14), we deduce the following relation: $\mathbb{E}\big(n_j\big) = N_j \mathbb{P}(D < t^* | X \in I_j$

In the right term, the probability corresponds to the probability of inclusion ω_j of a storage of category I_j. For $\big[x_{j-1}, x_j\big]$, the probability has the following expression:

$$\omega_j = 1 - \int_{x_{j-1}}^{x_j} \exp\{-h(x)t^*\} \frac{\alpha x^{-\alpha-1}}{x_j^{-\alpha-1} - x_{j+1}^{-\alpha-1}} dx \tag{7.15}$$

This probability of inclusion can be estimated by replacing the function of visibility h by the estimator $\hat{h}$ that is associated to the penalised estimator $\hat{g}^*_{\hat{m}}$ in (7.15). Let's note $\hat{\omega}_j$ the estimator of ω_j thus obtained. Let I_j be an interval of partition $\hat{m}$, the number of storages N_j of this category in the platform can be estimated by the quantity:

$$\hat{\mathrm{N}}_{\mathrm{j}} = \frac{\mathrm{n_I}}{\hat{\omega}_{\mathrm{I}}} \tag{7.16}$$

The total number of storages with a quantity higher than x_0 is finally estimated by

$$\hat{\mathrm{N}} = \sum_{\mathrm{j}=1}^{\mathrm{k}} \hat{\mathrm{N}}_{\mathrm{j}} \tag{7.17}$$

$N_B = 2500$

h: $\begin{cases} Partition: [20, 70, 110, 5000] \\ A = [5.10^{-4}, 5.10^{-4}, 5.10^{-5}] \\ b = 10^{-4} \end{cases}$

$\mu_0 = 8$

Draw in the stock: $\gamma = 2$

Intensity of production: $\lambda = 8$

(a) Parameters of the protocol

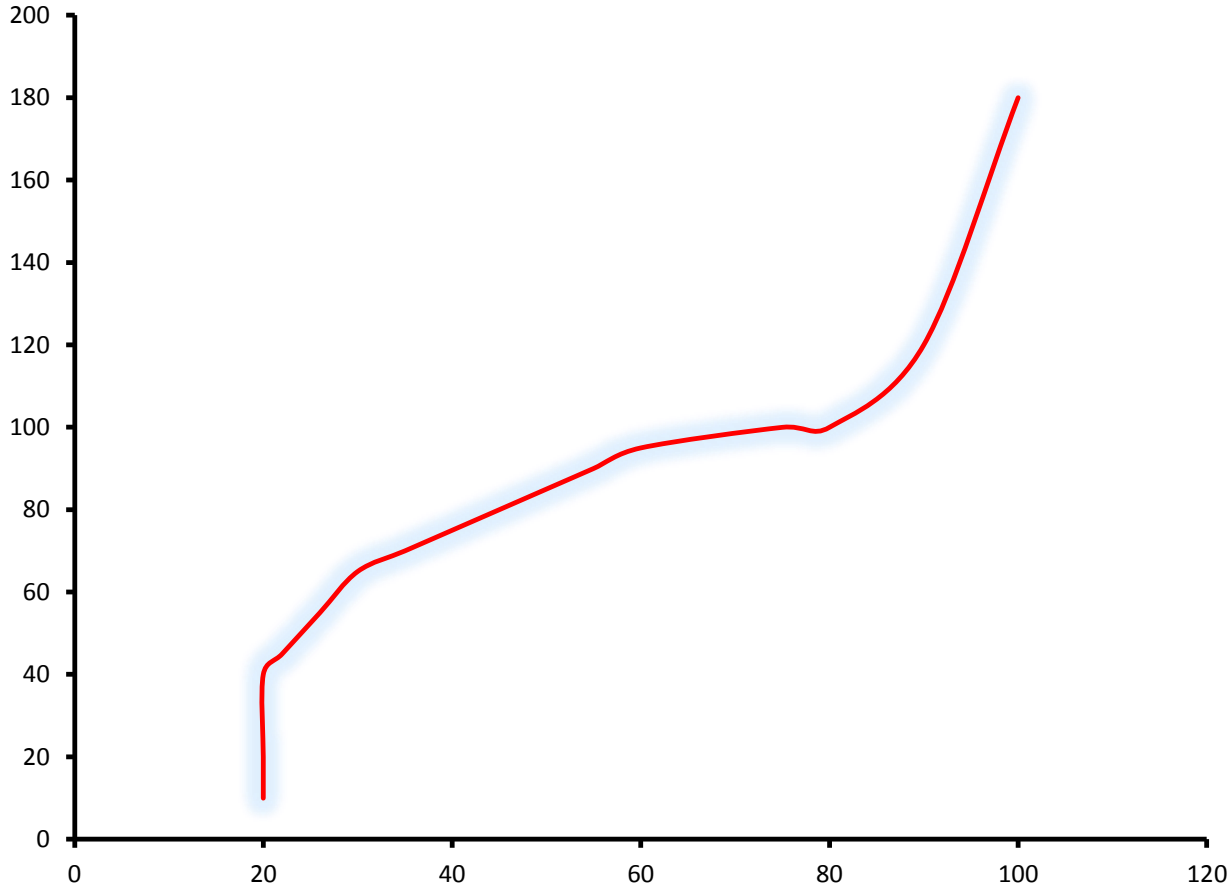


Figure 7.6: Associated visibility function

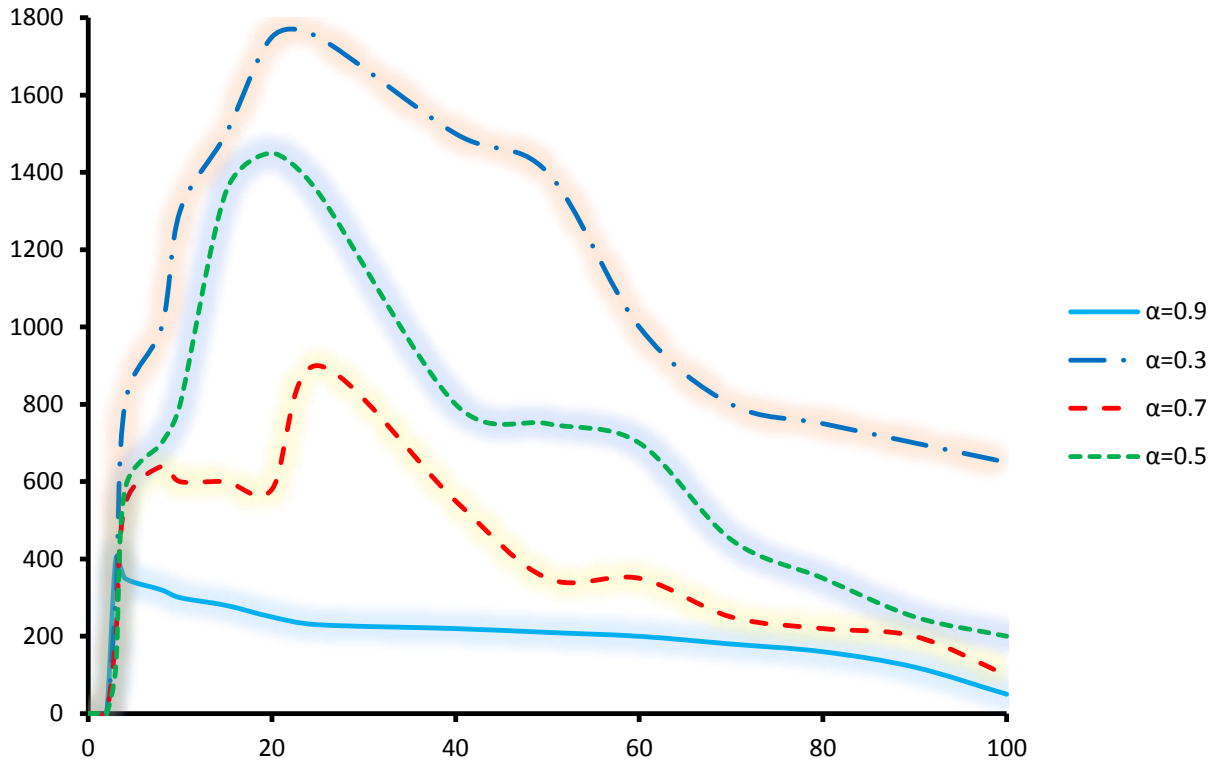


Figure 7.7: Impact of the politics of reserves law (Protocol 5): Mean production curves for various coefficients α for the Pareto-Levy law

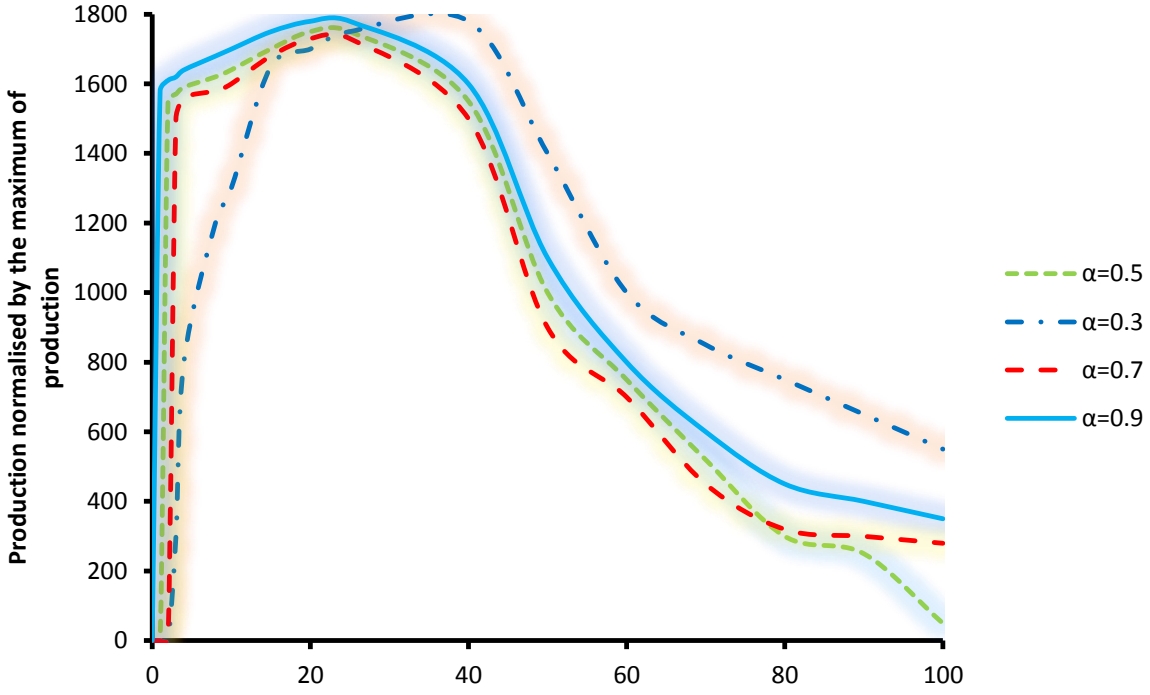


Figure 7. 8: Mean production curves of figure 7.9c normalised by the maximum value of production.

It was possible to suggest another estimator of the number. In the context of the extraction under unequal probabilities of inclusion, the Horvitz-Thompson estimator [181] is more often used. This estimator is mostly used for the estimation of petroleum reserves. Hence the reason to apply it to biofuel production from the extracted Ricinus Communis oil. And its expression is as follows:

$$\widehat{\mathrm{N}}_{\mathrm{j}}^{\mathrm{HS}} = \sum_{\mathrm{i}|\mathrm{X}_{\mathrm{i}}^{*} \in \mathrm{I}_{\mathrm{j}}} \frac{1}{\widehat{\omega}(\mathrm{X}_{\mathrm{i}})} \tag{7.18}$$

Where $\widehat{\omega}(x) = 1 - \exp\{-\hat{h}(x)t^{*}\}$ is the probability of inclusion of the storage of quantity x. As previously, we define the estimator of the total number of storages N by:

$$\widehat{\mathrm{N}}^{\mathrm{HS}} = \sum_{\mathrm{j}=1}^{\mathrm{k}} \widehat{\mathrm{N}}_{\mathrm{j}}^{\mathrm{HS}} \tag{7.19}$$

Application

Tables 7.3, 7.4 and 7.5 present the estimations obtained for the number of the three platforms proposed in this work.

Table 7.3: Estimations of the number of remaining storages to be extracted by class of the quantity for platform A.

j	$\boldsymbol{I_j}$	$\boldsymbol{\widehat{\omega}_j}$	$\boldsymbol{n_j}$	$\boldsymbol{\widehat{N}_j - n_j}$
1	$[20, 63.5]$	.49	111	115
2	$[63.5, 110]$	.73	37	13
3	$[110, 4139]$	.95	74	4

Table 7.4: Estimations of the number of remaining storages to be extracted by class of the quantity for platform B.

j	$\boldsymbol{I_j}$	$\boldsymbol{\widehat{\omega}_j}$	$\boldsymbol{n_j}$	$\boldsymbol{\widehat{N}_j - n_j}$
1	$[12.1, 17]$	.35	23	43
2	$[17, 40.5]$	.53	43	38
3	$[40.5, 72.5]$	.81	22	5
4	$[72.5, 4482]$	.98	44	1

Table 7.5: Estimations of the number of remaining storages to be extracted by class of the quantity for platform C.

j	$\boldsymbol{I_j}$	$\boldsymbol{\widehat{\omega}_j}$	$\boldsymbol{n_j}$	$\boldsymbol{\widehat{N}_j - n_j}$
1	$[20.5, 350]$	.52	89	81
2	$[350, 1250]$	.99	43	0

7.3.3 Extensions of production

In the class of small quantities of storages, the future extractions were modelled by naturally extending the process of Poisson whose intensity was estimated. The following proposition allows to model the suite of extractions for the class of storages of quantity higher than x_0.

Proposition

Let X be a random variable of Pareto Levy distribution restrained to $[x_0, x_{max}]$ and let D be a random variable of conditional law $(D|X) \sim \varepsilon(h(X))$ with $h \in H_m$ where m was a partition of $[x_0, x_{max}]$. For an interval $I = [x_{j-1}, x_j]$ of the partition m, let $(\tilde{X}, \tilde{D})$ be a random vector of the same law than $(X, D|D > t^*, X \in I)$. We note $P_{dec}(I) := \mathbb{P}(D \leq t^*|X \in I)$. The density of the joint law $(\tilde{X}, \tilde{D})$ was the function defined in $I \times \mathbb{R}_+$ by the expression:

$$(x, t) \longmapsto \frac{1}{1 - P_{dec}(I)} \frac{\alpha x^{-\alpha-1}}{x_{j-1}^{-\alpha} - x_j^{-\alpha}} h(x) \exp\{-h(x)u\} 1_{x \in I, t > t^*} \qquad (7.19)$$

In particular, the marginal density of $\tilde{X}$ was:

$$x \mapsto \frac{\exp\{-h(x)t^*\}}{1-P_{dec}(I)} \frac{\alpha x^{-\alpha-1}}{x_{j-1}^{-\alpha}-x_j^{-\alpha}} 1_{x\in I} \qquad (7.20)$$

And the conditional law of $\widetilde{D}$ knowing that $\widetilde{X}$ was of exponential distribution of $h(\tilde{X})$ parameter left-truncated on t^*.

Demonstration

Let z be a measurable and bounded function defined in $I \times \mathbb{R}_+$ and with values in $\mathbb{R}$. Then:

$$\mathbb{E}\left(z(\tilde{X},\tilde{D})\right) = \frac{1}{\mathbb{P}(D > t^* | X \in I)} \int_{t>t^*} \int_{x\in I} \frac{\alpha x^{-\alpha-1}}{x_{j-1}^{-\alpha} - x_j^{-\alpha}} h(x)\, exp\{-h(x)t\} z(x,t) dx\, dt$$

$$= \frac{1}{1 - P_{dec}(I)} \int_{x\in I} \int_{u\geq 0} \frac{\alpha x^{-\alpha-1}}{x_{j-1}^{-\alpha} - x_j^{-\alpha}} h(x) \exp\{-h(x)(u+t^*)\}\, z(x,t)\, dx\, dt$$

The marginal law $\tilde{X}$ was obtained directly by integrating the joint density with u, and the conditional law of $\widetilde{D}$ was read on the expression of the joint law.

The law of remaining reserves or storages to be extracted was therefore not of Pareto Levy type. For fixed class I, the exponential term in most mathematical expressions of this section indicated that the density of the $\tilde{X}$ law takes more important values on the small quantities of storages of I. This is natural because inside the class I, the visibility of a storage increases with the quantity of its oil, and storages of this class that have big quantities have been easily extracted compared to others.

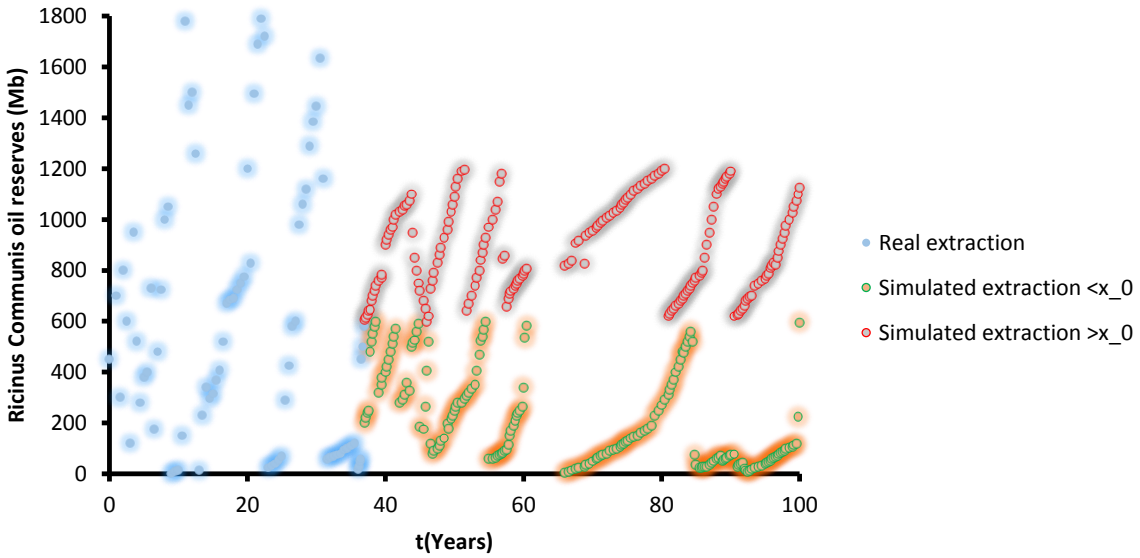


Figure 7. 9: Extension of the process of extraction in platform B. Ricinus Communis oil reserves are represented in a logarithmic scale in order to better visualize small quantities extracted.

Let $S_j(t) = \sum_{i|X_i \in I_j} X_i 1_{t^* < D_i \leq t}$ be the random variable of the total reserves extracted and cumulated during the period between $[t^*, t]$, for storages of category I_j. According to the previous proposition: $\mathbb{E}(S_j(t)|N_j - n_j) = (n_j - N_j) \frac{1}{1-P_{dec}(I_j)} \int_{x \in I_j} \int_{t \geq t^*} \frac{\alpha x^{-\alpha}}{x_{j-1}^{-\alpha} - x_j^{-\alpha}} h(x)\, exp\{-h(x)t\}\, dx\, dt$

In order to study the evolution of the quantities of oil extracted over time, the mean rate of extraction (in Mb per year) was defined by the following function τ, for all $t > t^*$:

$$\tau(t) \coloneqq \frac{1 - x_0^{-\alpha+1}}{1 - x_0^{-\alpha}} \frac{\alpha}{1-\alpha} \mu_0 + \sum_{j=1}^{k} \frac{d}{dt} \mathbb{E}[S_j(t)] = \mu_0 + \sum_{j=1}^{k} (n_j - N_j) \frac{1}{1 - P_{dec}(I_j)}$$

$$\int_{x \in I_j} \frac{\alpha x^{-\alpha}}{x_{j-1}^{-\alpha} - x_j^{-\alpha}} exp\{-h(x)t\}\, dx$$

This quantity allowed the measurement of the decline of quantities of reserves extracted over time.

Applications

Figure 7.11 presents a simulation of the extension of platform B. Storages extracted in class $[x_0, x_{max}]$ were simulated by using the estimation of the visibility function $\hat{h}$ on the partition $\hat{m}$ selected by the slope method and proposition 7.3.1.

Also, storages extracted in the category $[1, x_0]$ were obtained by extending the process of Poisson marked with small quantities extracted. And, during the extended period, very small quantities of oil (From 1Mb to 3Mb) were extracted in big numbers than during the known period of production of the platform. In reality, this situation was not inconvenient because the quantities of reserves considered were negligible. Also, it was probable that in the future, those very small quantities of oil will be needed by biofuel producers. In the next section, we use such extensions of the production process in order to propose the production extensions of the platform.

In addition, figure 7.12 represents the mean rate of extraction τ that corresponded to the area of platform B. And, figure 7.13 represents trust boundaries for the random variable $S_x(t)$ from a bigger number of simulations of extensions. Beyond the vertical line corresponding to the present, the function $t \longmapsto \mathbb{E}(S_x(t))$ was governed by 5% and 95% of the $S_x(t)$ quintile obtained for 1000 simulations. Figures 7.14 and 7.15 represent the mean rate of extraction and the cumulated extractions for both platform A and platform C respectively. On these figures, it was evident that the production from platform A brought significantly less

extraction than in platform B or in platform C. Therefore, it was necessary to highlight that the model was designed to describe the production cycle corresponding to a certain type of biofuel. The model did not pretend to consider important extractions of better quality *ricinus communis* oil. However, the obtained extensions tended to respect the dynamic of the production process for the three platforms. Particularly, scenarios related to the mean of extraction rates extended the history of quantities of oil extracted in the past.

7.3.4 Extensions of production profiles

In order to suggest a scenario of the production extension of a platform, there was a need to use an extension of the production of the platform and define the politics of stock management. Each new storage extracted was activated according to the politics of stock management and produced according to the model presented in section 7.1. It was difficult to infer which politics of production had to govern the production of the platform, because the draw performed on the population of storages evolved over time. However, in the previous section, it was observed that for draws biased by a power of a quantity sufficiently important, the production profiles of the platform had globally the same production curve. Then, a draw biased was used by the quantity ($\gamma = 1$) as politics of stock management.

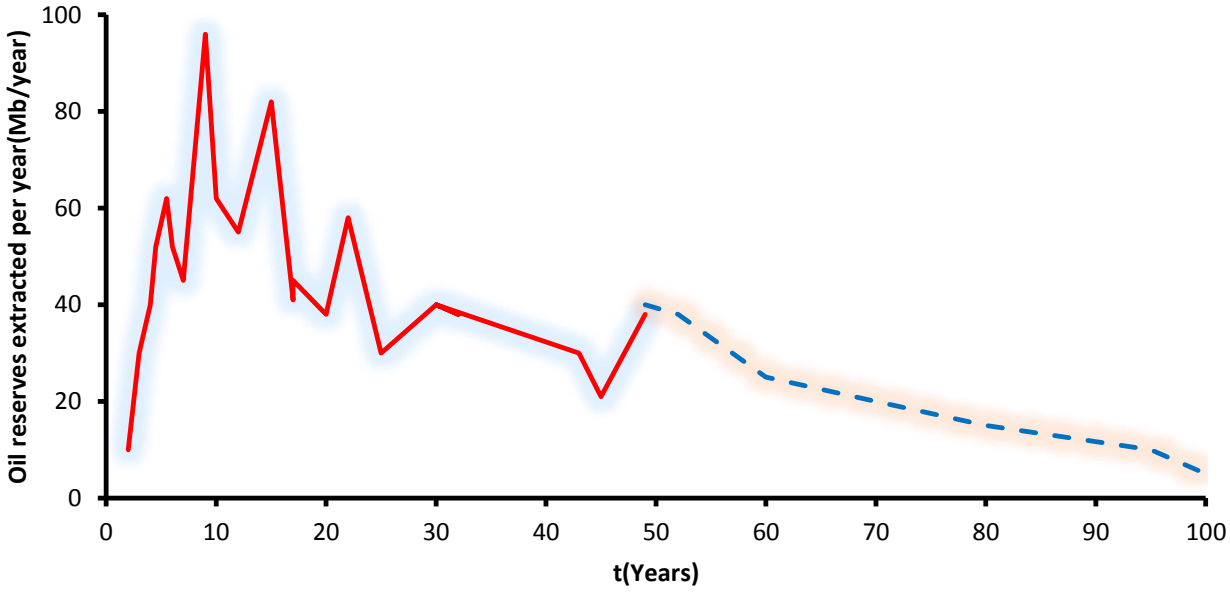


Figure 7. 10: Annual Ricinus Communis oil quantities extracted and stored in platform B. The curve with the continuous line corresponding to the oil extracted in the past is extended by the estimated mean extraction rate function τ.

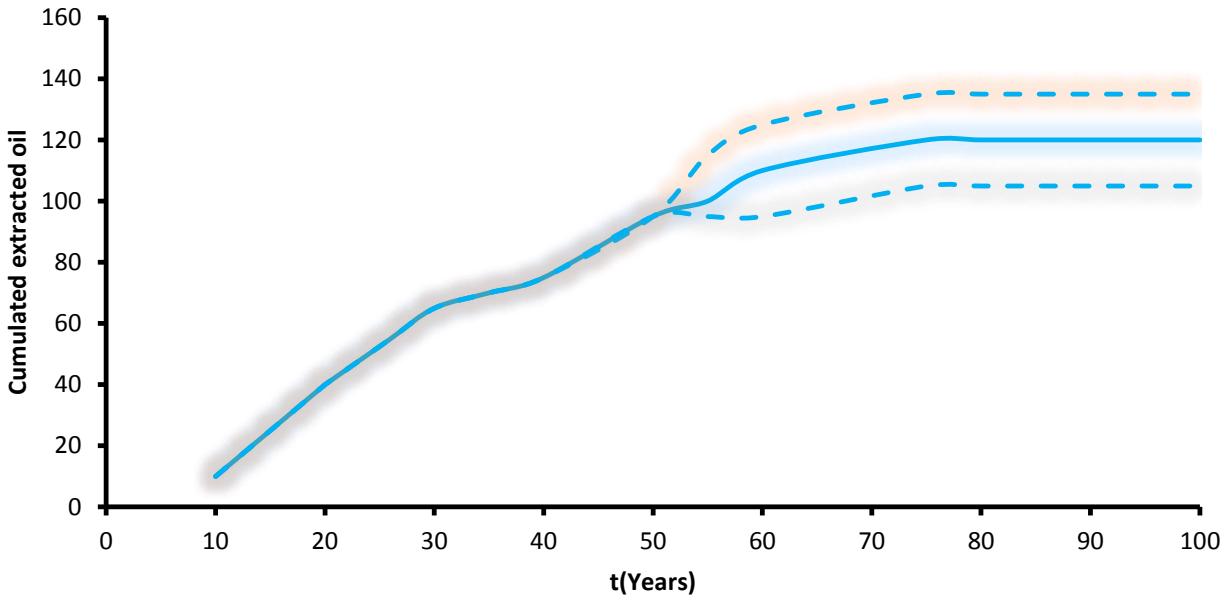


Figure 7. 11: Cumulated Ricinus Communis oil extracted from platform B. Beyond the present time, the extended curve corresponds to the mean of S_x^t, governed by 5% and 95% quintiles for 1000 simulations of S_x^t.

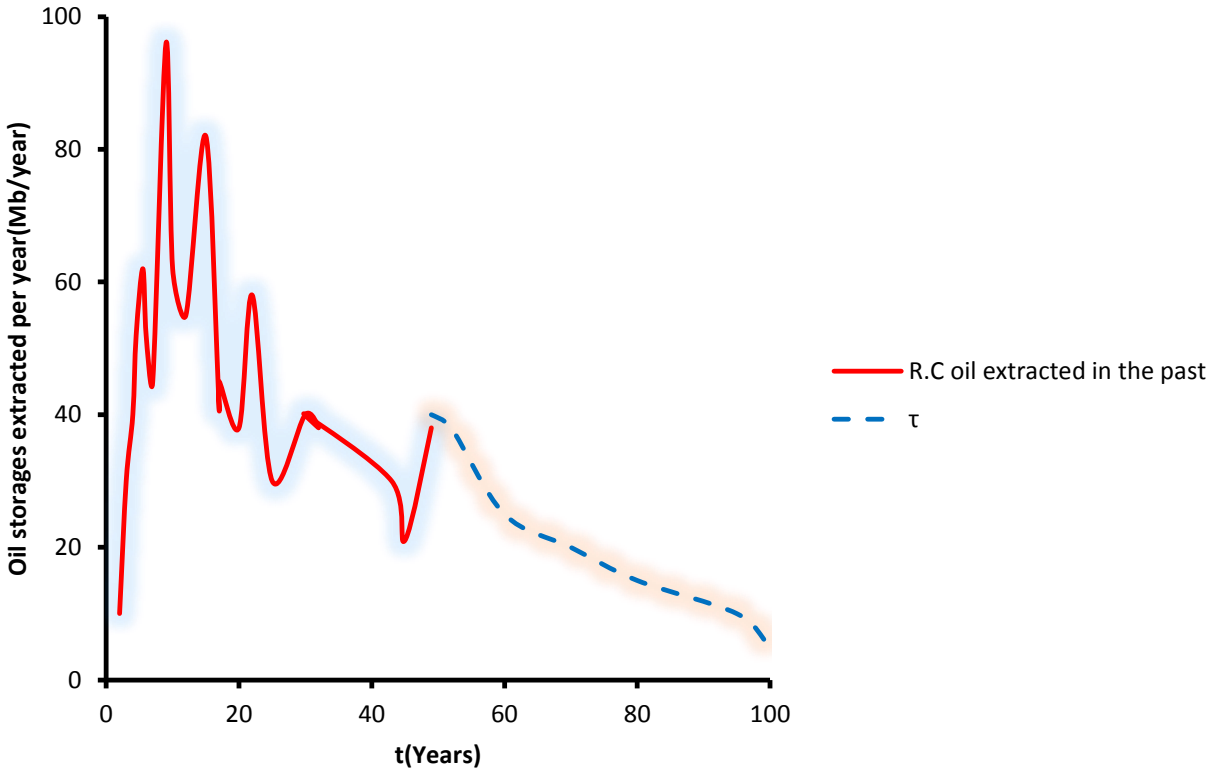


Figure 7.12(a): Annual quantities extracted from platform A.

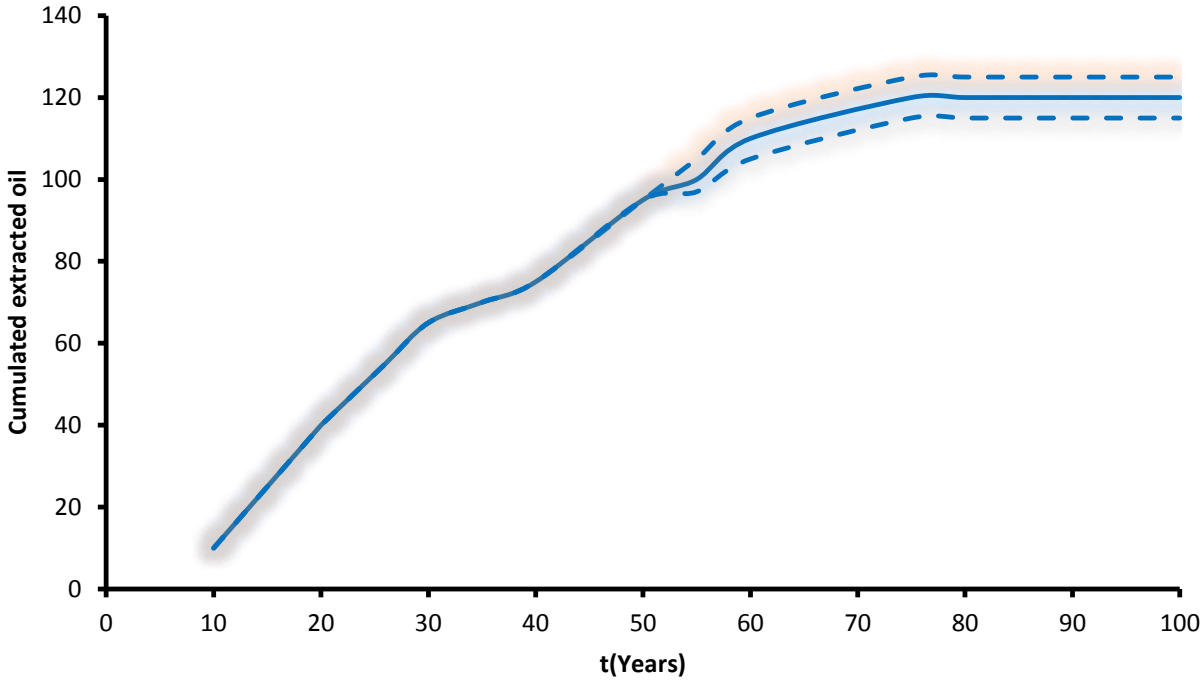


Figure 7.12(b): Cumulated ricinus communis oil extracted from platform A.

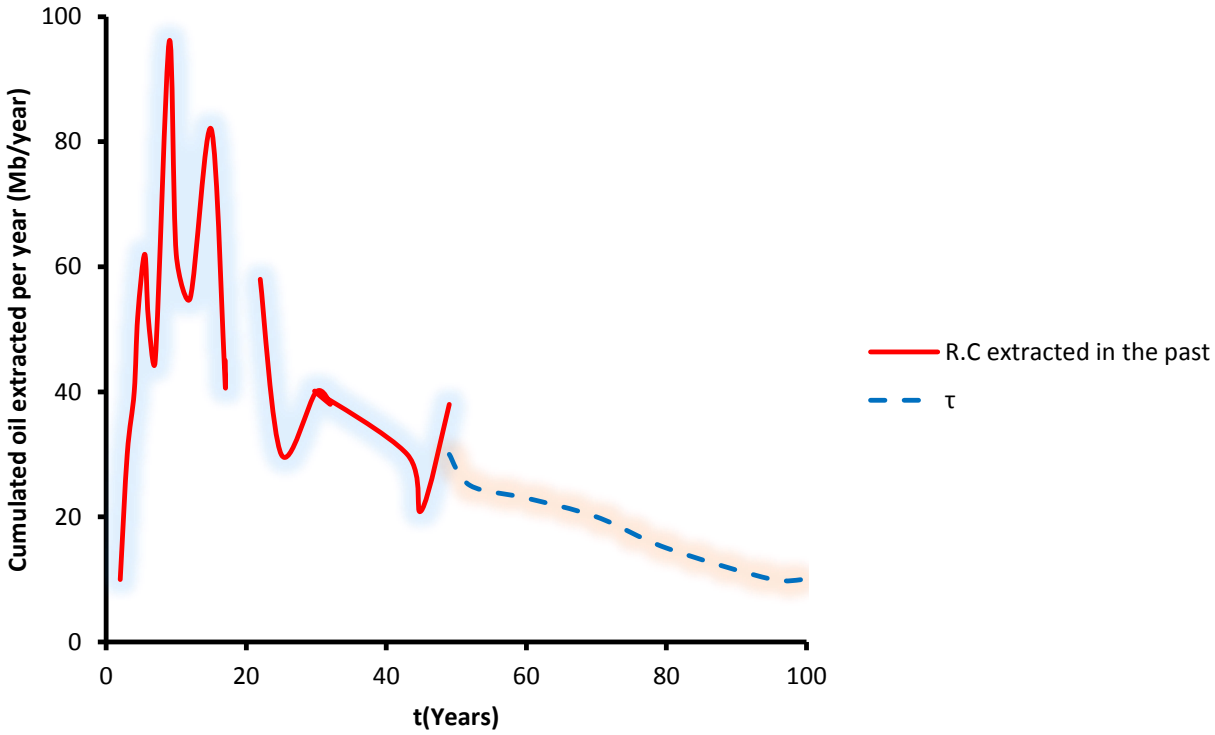


Figure 7.13(a): Annual quantities of *ricinus communis* oil extracted from platform C.

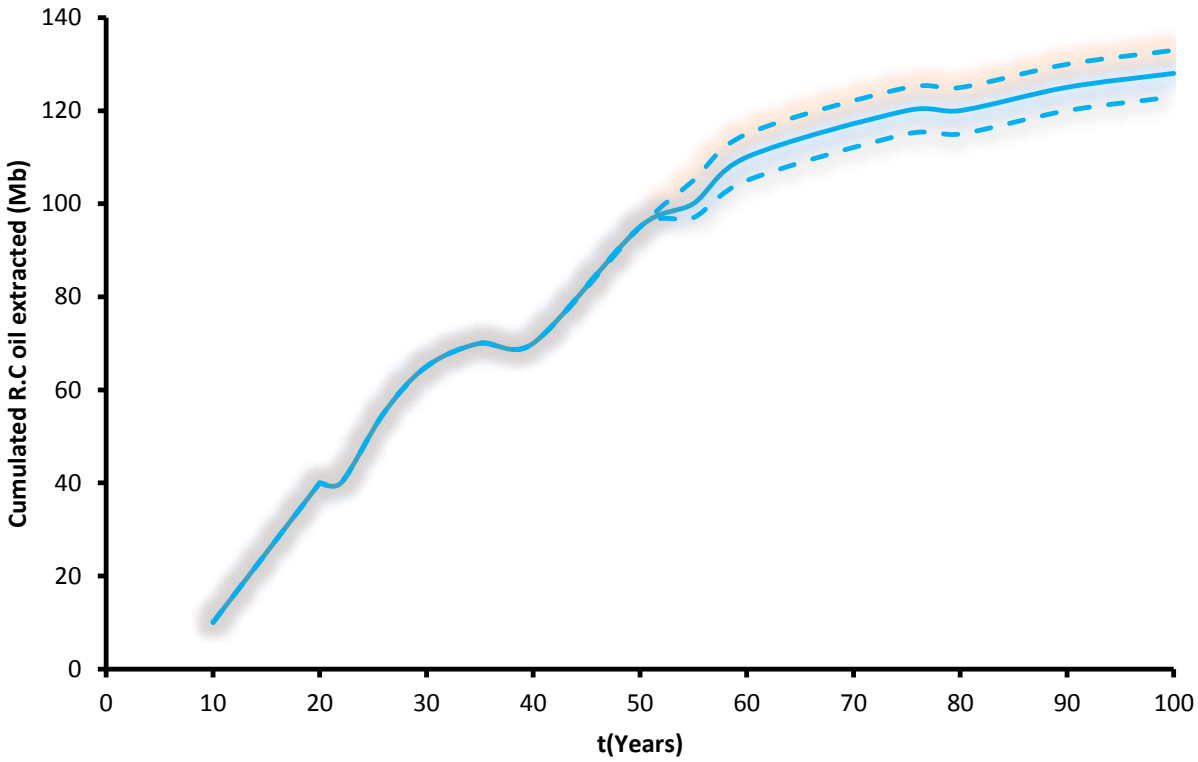


Figure 7. 13(b): Cumulated *ricinus communis* oil extracted from platform C.

In reality, a storage could not produce right after the oil extraction. Therefore, for simulations, a comprehensive delay of three years between the date of completion of the extraction and the date of production was imposed. The past production of platform A was not isolated because it was assumed that existing *ricinus communis* oil storages were grouped by oil producers. And this did not allow the isolation of the production restrained to platform A. Figure 7.16 presents the evolution of the stock in platform B. A first scenario that corresponds to the rhythm of production of the last years was proposed, meaning by considering an intensity $\lambda = 8$.

In addition, the second scenario was based on an increasing intensity $\tilde{\lambda}(t) = 0.7t + 8$. The corresponding extensions were traced in figure 7.12. And the evolution of the stock of storages for platform C was presented in figure 7.13(a). The same scenarios of production than those in platform B were considered in order to extend the production in figure 7.13(b).

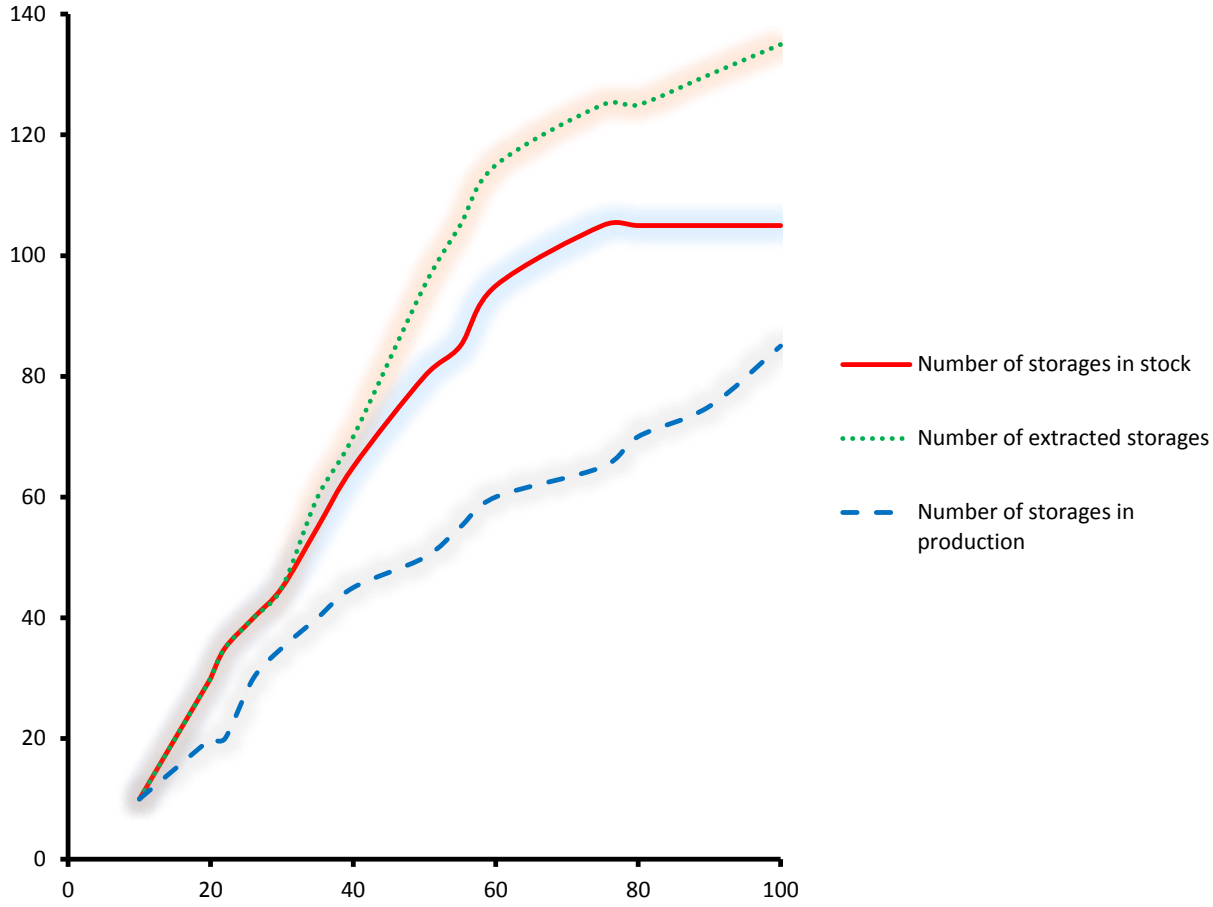


Figure 7. 14: Evolution of the stock of storages available in any of the three platforms

Comments

The industrial production of biodiesel required basic technical equipment, including a reactor, centrifuges, pumps, distillation columns, funnels and storage containers. It was at the level of the basic technical equipment that the reaction of transesterification happened. The amount of each reagent before its introduction into the reactor was determined. And the chemical composition of the reactional mixture changed with time. The key parameters that determined the reaction of conversion were: temperature, pressure, reaction time and the degree of mixture. In general, the increase of temperature increased the speed of reaction, and accelerated the reaction of conversion. After that, the reactor was closed after defining the conditions of reaction. The separation of biodiesel and glycerin was obtained by using a tray settling. In case the tray settling was expensive, a centrifuge was used.

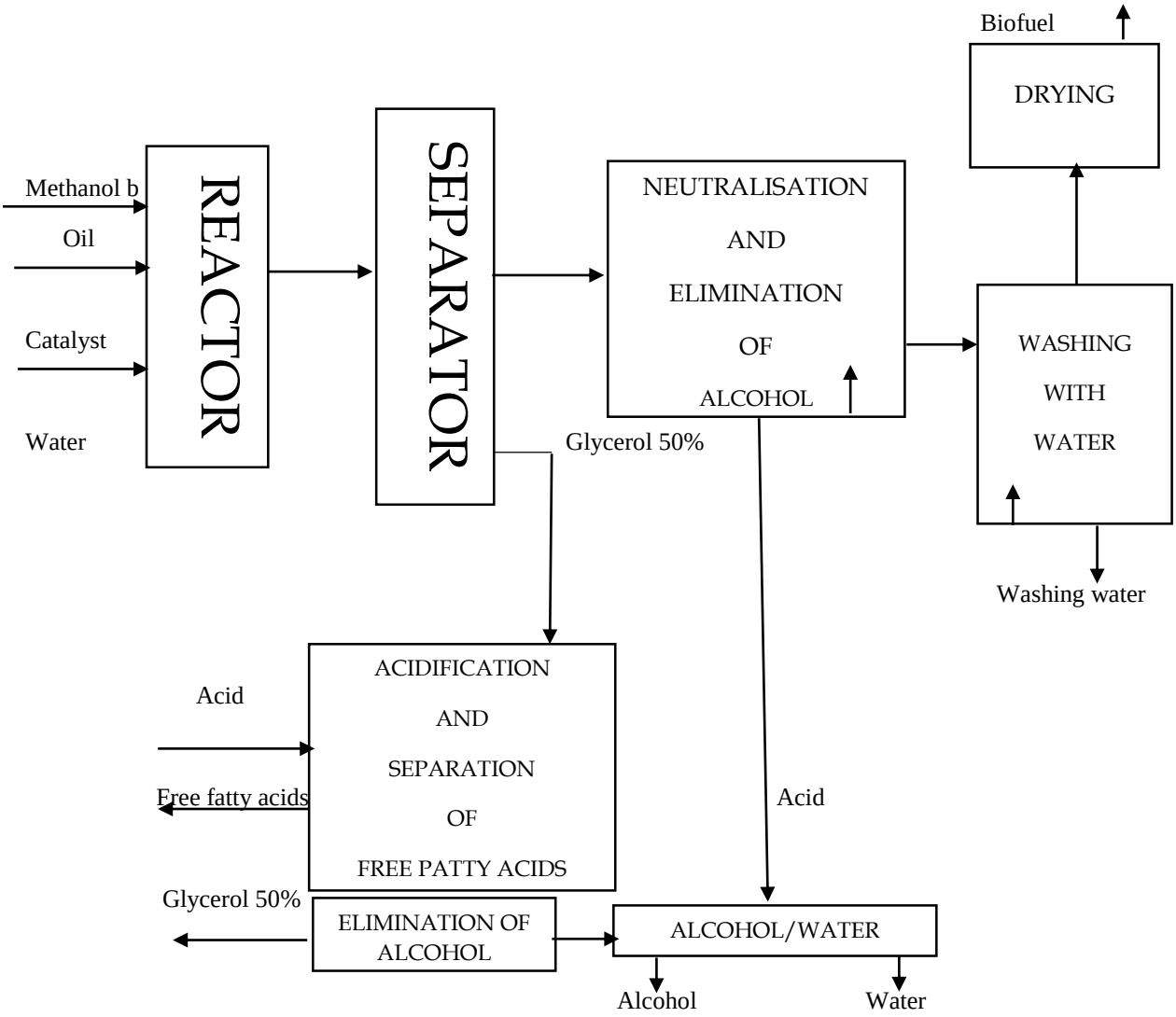


Figure 7. 15: Principle of biofuel production [3]

7.5 Industrial biofuel production

The simplest method to produce alcohol esters was the mixture of alcohol, oil and catalyst in a reactor that had a container of agitation. The reactor had a condenser and the temperature was approximately at $340\,°K$. Sodium hydroxide was the most used catalyst for the transesterification of triglycerides. Oil was first added to the system, then the catalyst and finally, the methanol. Then the system was agitated during the reaction time which was approximately equal to 4 hours. Then the agitation was stopped. During some of the processes, the reactional mixture was left in the reactor in order to obtain a separation of esters and glycerol. And in some of the other processes, the reactional mixture was pumped in the funnel or separated using a centrifuge. Alcohol was separated from glycerol and esters using an evaporation. Next, esters were neutralised, then washed using vapor. Then, fine quantities of residual methanol were eliminated: it was the drying phase. After all those

phases, biofuel was stored. After that, glycerol was also neutralised and washed with the water, then sent to the refinery. Free fatty acids contained in vegetable oils reacted with the catalyst. This was already described earlier in this work. The acceptable fatty acid rate in those systems was $< 2\%$ but preferably $< 1\%$. The process is shown as a reactor in two phases: The first phase consisted of recovering the glycerin. A rectification column led to the separation of the excess of methanol and glycerin. In the second phase, a washing column allowed to extract the methyl ester by purifying it from glycerin and the residual methanol.

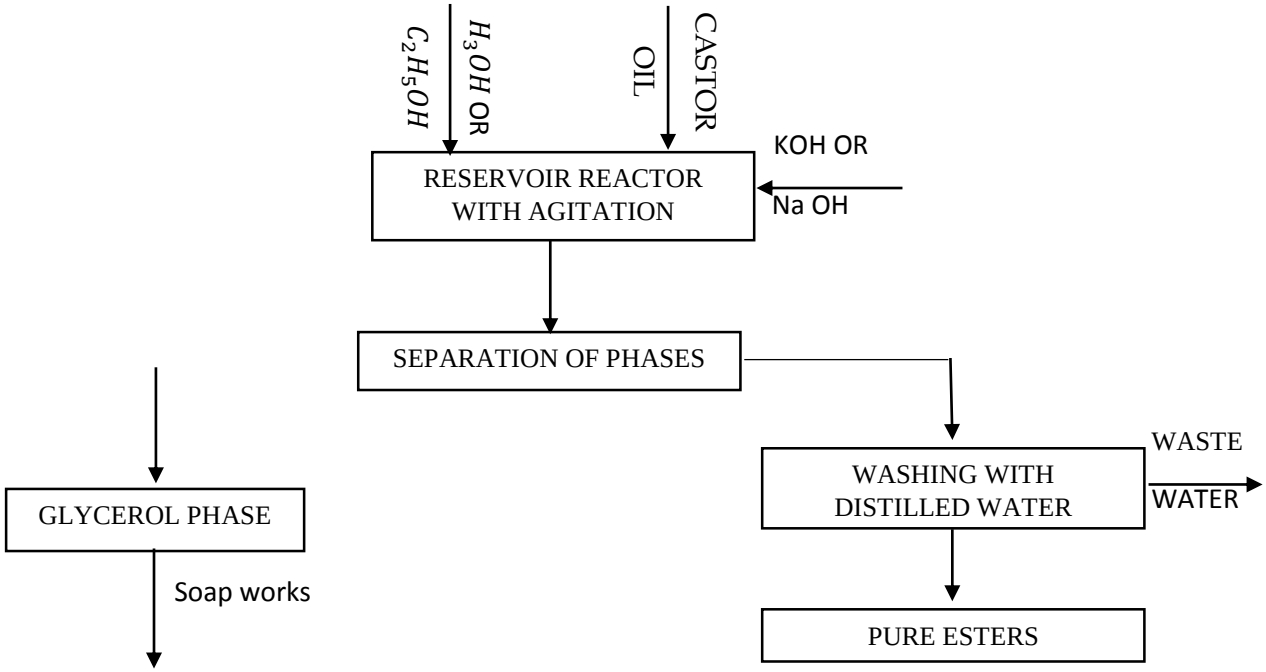


Figure 7. 16: Principle of separation of esters and *ricinus communis* oil glycerol [11]

Partial Conclusion

Despite many challenges faced during the transesterification reaction, a small quantity of biofuel was produced in laboratory. It was important to know all the characteristics of this biofuel. This will be the main focus of the next chapter.

7.6 Characterisation of biofuel from *ricinus communis* oil

Biofuels are characterised by many different parameters including viscosity, density, cetane number, flash point, index of acidity, alcohol composition, malar mass and calorific value. In this section, we will study some of them.

7.6.1 Example of calculation of the average molar mass

Ricinus communis oil methyl ester contains:

90% of ricinoleic acid: $C_{19}H_{38}O_3$. M: 298 g/mol

4% of oleic acid: $C_{19}H_{36}O_2$. M: 282 g/mol

3% linoleic acid: $C_{19}H_{34}O_2$. M: 280 g/mol

Other acids that are present as traces.

$$\begin{cases} 90\%\ of\ C_{19}H_{38}O_3 : M = (19 \times 12) + 38 + 48 = 314\ g/mol & (1) \\ 4\%\ of\ C_{19}H_{36}O_2 : M = (19 \times 12) + 36 + 32 = 296\ g/mol & (2) \\ 3\%\ of\ C_{19}H_{34}O_2 : M = (19 \times 12) + 34 + 32 = 294\ g/mol & (3) \end{cases}$$

Let m_E be the ester mass. Then, the total number of moles are:

$$n_\tau = m_E \left(\frac{0.9}{314} + \frac{0.04}{296} + \frac{0.03}{294}\right)$$

$$n_\tau = 31 \times 10^{-4}\ m_E$$

The molar fractions are:

$$X_{(1)} = \frac{0.9}{314 \times 31 \times 10^{-4}} = 92.46\%$$

$$X_{(2)} = \frac{0.04}{296 \times 31 \times 10^{-4}} = 4.4\%$$

$$X_{(3)} = \frac{0.03}{294 \times 31 \times 10^{-4}} = 3.3\%$$

$$\bar{M}_E = \sum_{i=1}^{3} X_i M_i$$

$$\bar{M}_E = (92.46 \times 314) + (4.4 \times 296) + (3.3 \times 294) \times \frac{1}{100}$$

$$\bar{M} = 313.05\ g/mol$$

7.6.2 Density

Densities and viscosities of ester and gasoil were determined for temperatures from 20℃ – 60℃. Then:

Table 7.6: Viscosities and Densities of *ricinus communis* methyl ester

Temperature	20	25	30	35	40	45	50	55	60
Density	0.921	0.910	0.904	0.900	0.890	0.883	0.875	0.860	0.860
Viscosity	11.31	10.43	7.93	4.5	4	4	4	4	4

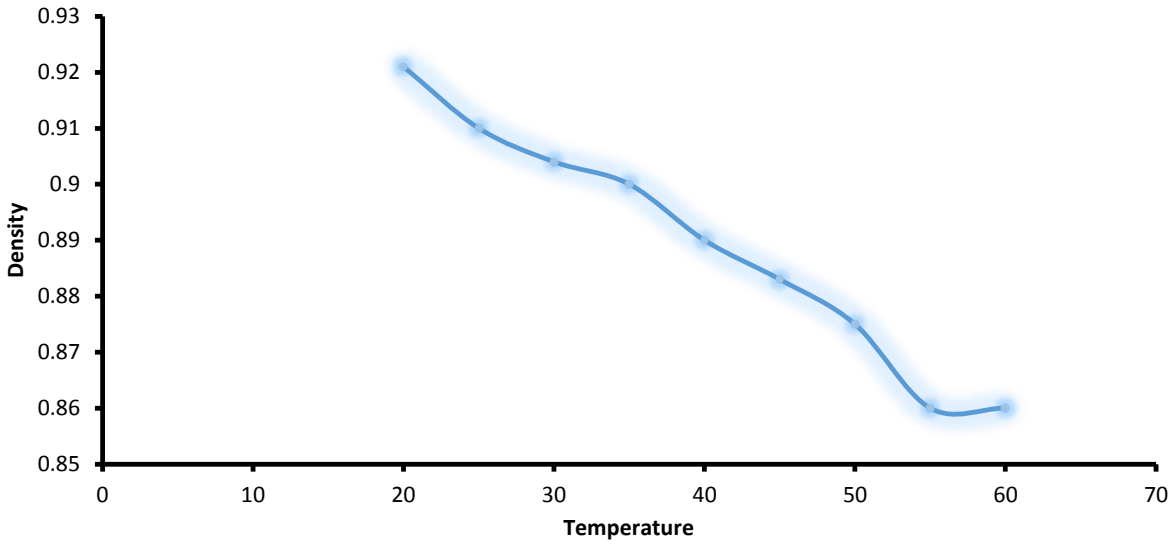


Figure 7. 17: Methyl ester density of *ricinus communis* oil in relation with temperature

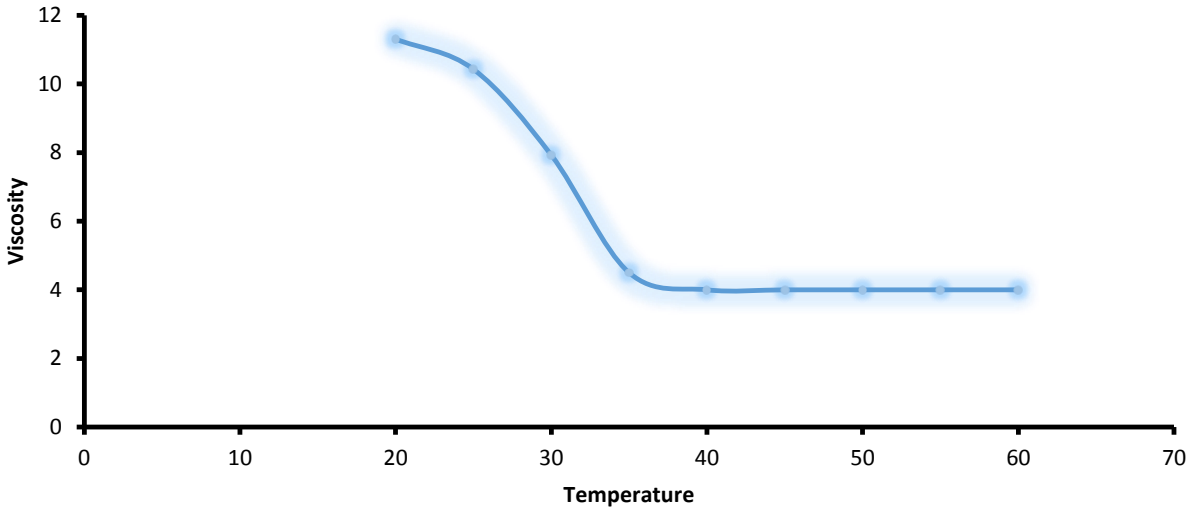


Figure 7.18: Methyl ester viscosity of ricinus communis oil in relation with temperature

Table 7. 7: Viscosities and Densities of gasoil

Temperature	20	25	30	35	40	45	50	55	60
Density	0.820	0.810	0.805	0.800	0.800	0.800	0.800	0.800	0.800
Viscosity	4.03	4.03	3.5	3	3	3	3	3	3

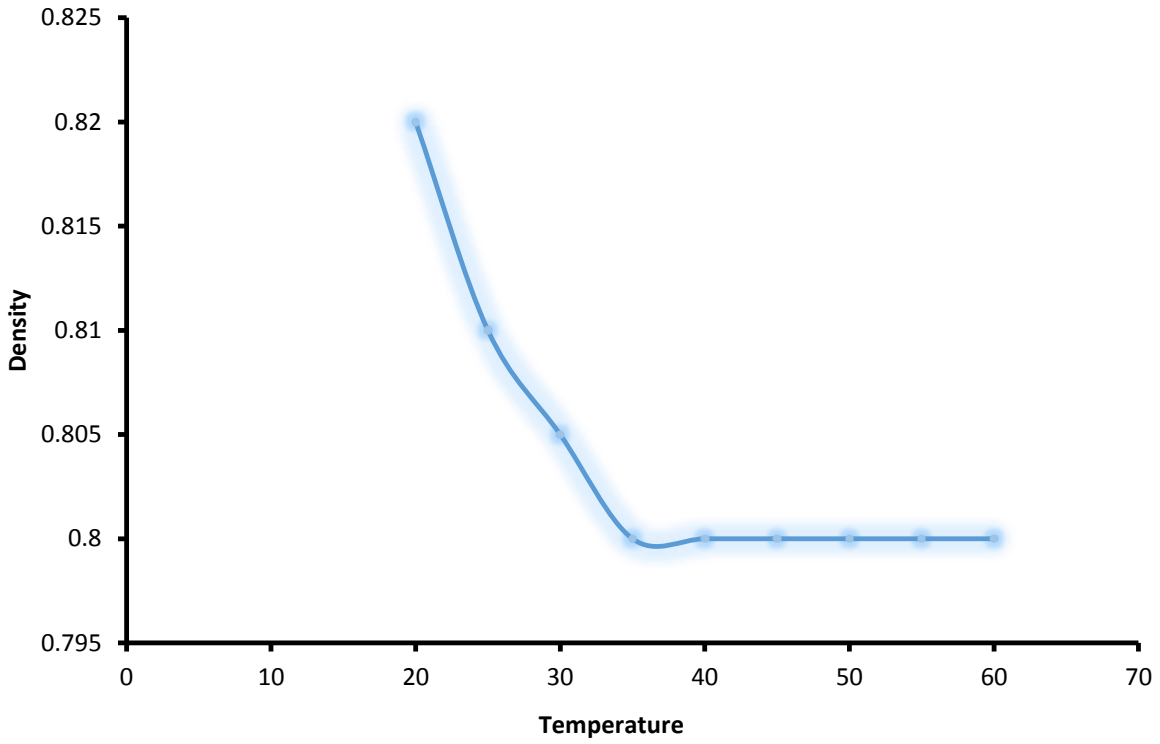


Figure 7.19: Gasoil density in relation with temperature

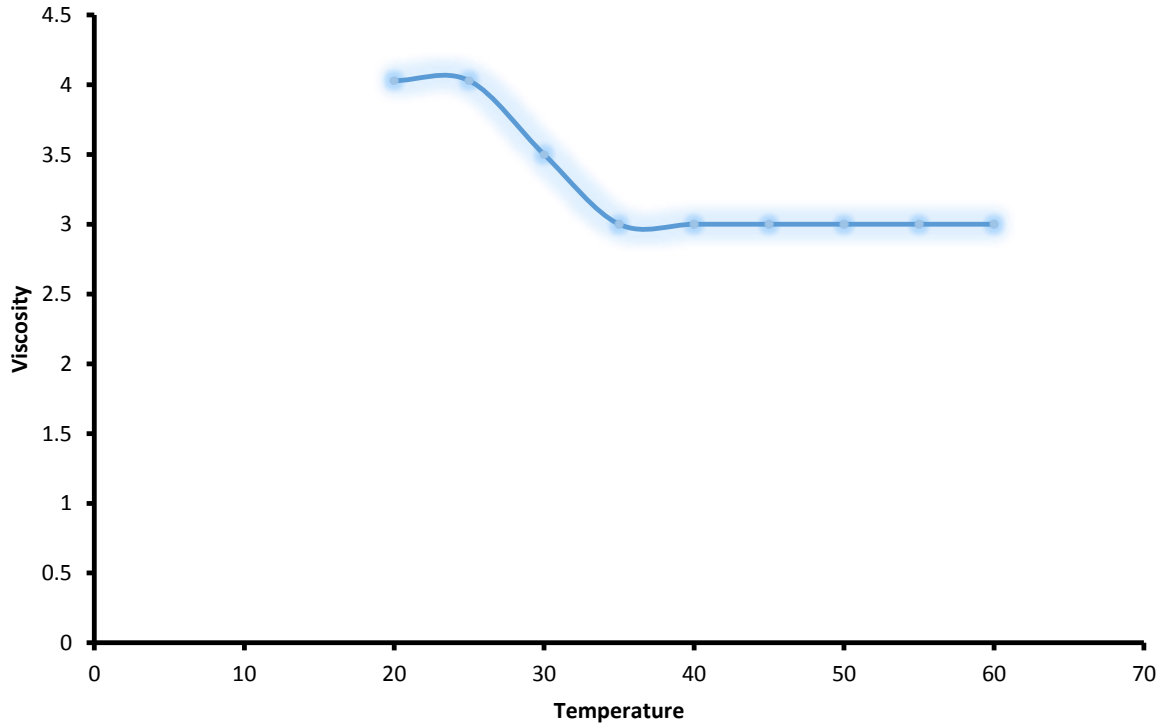


Figure 7.20: Gasoil viscosity in relation with temperature

Comments

Ricinus communis methyl esters have the advantage of having a density and viscosity much lower than those of oil determined earlier in this work. As the above curves have shown, viscosity and density of esters inversely varied with temperature. Compared to gasoil curve, the *ricinus communis* curve had a much more pronounced slope. That what justified its approximation above 60°C. Considering the density curve, it was observed that from 40°C, the density of gasoil did not vary. This was very favorable for the quality of the injection of oil and of its esters at the level of the combustion chamber.

7.6.3 Cetane number

Cetane number was used to appreciate the auto inflammation aptitude of gasoil. The simplest way to determine the cetane number was the application of the Klopfenstein equation formula. This equation considers the number of double liaisons and the carbon number. This gives:

$$I_{cetane} = 58.1 + 2.8 \times \frac{n-3}{2} - 15.9\,N$$

with:

n: Carbon number

N: Number of double liaisons

Because the *ricinus communis* methyl ester contained 3 double liaisons and 57 carbons, we obtained a number equal to 51. This value was very close to the cetane number of the normal diesel fuel, which was 52, and was higher than some diesel fuels.

7.6.4 Acidity index

After the transesterification reaction, esters with a good conversion (90-96%) were obtained. The index of acidity of those esters was determined the same way than in the case of oil. Experimental results indicated that the index of acidity was between 0.03 and 0.05. This confirmed what was said earlier about the transesterification reaction: it refined the free fatty acids oil. It decreased its acidity rate, which justified values that were very different.

7.6.5 Calorific quantity consumed by one methyl ester liter

For one methyl ester mole, we need 667.8 kJ of energy, or $Q_m = \frac{667.8}{M}\ kJ/g$

Now, let ρ be the density of ester at 20°C. ($\rho = g/l$).

We have: $Q_v = \rho \frac{667.8}{M}\ (kJ/g)$

Also, we need a maximum of one hour for a complete reaction.

In this case, $P = \frac{667.8\,\rho}{3600 \times M}.10^3(w/l) = \frac{185.5}{M}\rho(W/l)$

In watt-hour, this power corresponds to:

$$P' = \frac{185.5}{M}\rho \times 1h = 185.5\ \frac{\rho}{M}\ Wh/l$$

$$= 185.5\ \frac{\rho}{M}\ Wh/l$$

For methyl ester of *ricinus communis* oil, we have:

$$P = \left(\frac{185.5 \times 921}{313.05}\right) = 545.7450\ Wh/l$$

7.6.6 Heating values

The massic heating value represented the fuel quantity of energy by units of mass or of volume during the chemical reaction of complete combustion leading to the formation of $CO_2\ and\ H_2O$. Except when the temperature of reference was specified, the fuel was always taken at 25℃. The air and the combustion of products were considered at this same temperature. A distinction was made between higher heating values (HHV) and lower heating values (LHV) depending on if the water obtained by combustion was at a liquid or gaseous state. The most significant measure was the HHV because water was rejected under the form of vapor in the engine products of combustion.

7.6.6.1 Lower heating value (LHV)

Determination method: To determine the massic LHV, we must know the mass content of hydrogen of the fuel. If W_H designated that content in %, the mass of water formed during the fuel combustion was written as follows:

$$\mathrm{m_{H_2O}} = \frac{\bar{\mathrm{M}}}{2}\mathrm{W_H}.\mathrm{M_{H_2O}} \qquad (7.21)$$

$$\bar{\mathrm{M}} = \frac{(\rho_i \mathrm{V_i} + \rho_{i'} \mathrm{V_{i'}}) \times \mathrm{M_i} \times \mathrm{M_{i'}}}{\mathrm{M_i}\rho_i \mathrm{V_i} + \mathrm{M_{i'}}\rho_{i'} \mathrm{V_{i'}}} \qquad (7.22)$$

With i and i' for a mixture of two given bodies.

M_{H_2O} =molar mass of water (=0.018 kg)

By considering the massic enthalpy of evaporation Δh of water at 25℃, or 2358 kJ/kg, we obtain (6.3):

$$\mathrm{LHV_m} = \mathrm{HHV_m} - \frac{\bar{\mathrm{M}}}{2}\mathrm{W_H}.\mathrm{M_{H_2O}}.\Delta\mathrm{h} \quad (7.23)$$

$$\mathrm{LHV_m} = \mathrm{HHV_m} - 21.22\ \bar{\mathrm{M}}\ \mathrm{W_H} \quad (7.24)$$

$LHV_m and\ HHV_m$ are expressed in kJ/kg.

7.6.6.2 Higher Heating Value (HHV)

A. Measuring principle of a bomb calorimeter

A known mass of the substance or of the material was introduced in the bomb calorimeter. We install in it a firing device constituted of electrodes and we fill the bomb with the oxygen under high pressure. Then, it is placed in a calorimeter filled with water whose temperature can be measured. The firing device is subsequently connected and the agitation is started. The ignition is started and the variation of temperature is read. The benzoic acid is used as standard and, by comparison, the higher heating value of the unknown sample is deduced.

B. Process steps

- ✓ Place few grams of the combustible in a small container prepared for that;
- ✓ Connect the electrodes with a Chromium-Nickel alloy wire or even a platinum wire;
- ✓ Place 10 ml of distilled water in the bomb calorimeter and close it;
- ✓ Fill the bucket with 2l of water and start the agitator;
- ✓ For 5 minutes, record the change in water temperature;
- ✓ Ignite and, from the sixth minute to the fifth minute after the first constant value, record the change in water temperature.

C. Calculation of Higher Heating Value (HHV)

Evolution of the temperature:

Let:

a= time of firing

b= time at the end of which the rise of temperature reached 60% of its total value

c= time from the beginning of the period in which the temperature begins to become constant

ta= temperature at the time of firing

tc= time at time c

r'_1= rate of increase in temperature during the first five minutes before firing

r'_2 = rate of increase in temperature for five minutes after time c

On the contrary, if the temperature drops, r'_2 becomes negative.

C_1=volume (in ml) of the standard alkaline solution used to draw the acid

C_2= percentage of sulfur in the sample

C_3= length of the burnt fuse wire (in cm) during the firing

W= equivalent energy of the calorimeter

m= mass of the burnt sample

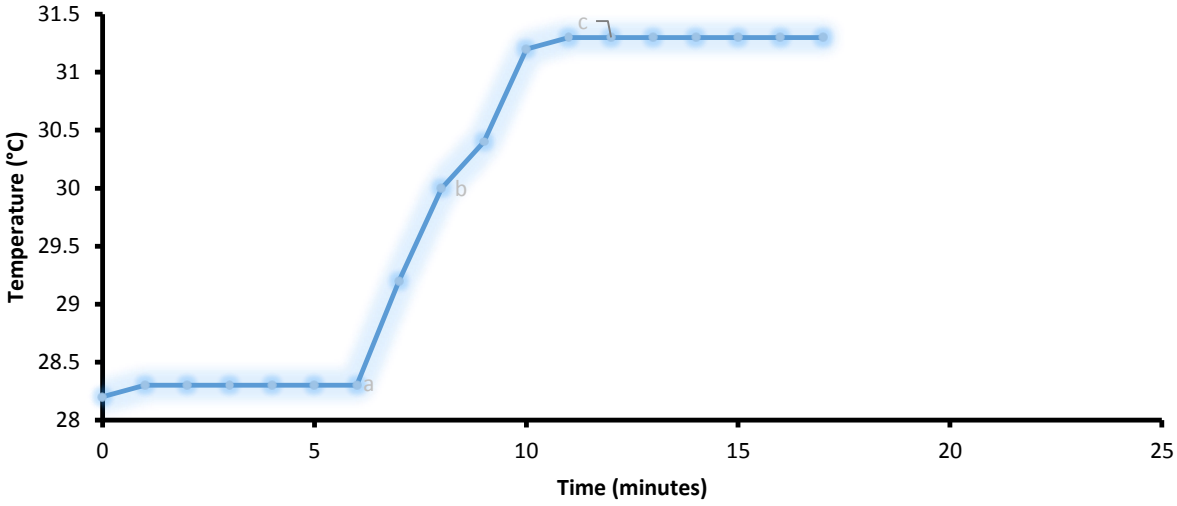


Figure 7. 21: Heating values curve: example of benzoic acid

A. Increase of temperature

The net increase of the corrected temperature is given by the following equation:

$$\Delta T = T_c - T_a - r'_1(b - a) - r'_2(c - d) \qquad (7.25)$$

B. Thermochemical corrections

e_1 = Correction in calories for the heat formation of nitric acid (HNO_3)

$e_1 = C_1$ if 0.075N is used for the titration

e_2 = Correction in calories for the heat formation of sulfuric acid (H_2SO_4)

$$e_2 = (14 \times C_2 \times m)$$

e_3 = Correction in calories for the fuse combustion of heat

= 2.3 C_3 if we use the chromium-nickel

= 2.7 C_3 if we use the galvanized steel.

The HHV

$$\mathrm{HHV_1} = \frac{\Delta \mathrm{T} \times \mathrm{W} - \mathrm{e_1} - \mathrm{e_2} - \mathrm{e_3}}{\mathrm{m}} \quad (7.26)$$

D. Standardisation of the calorimeter

The standardisation was about determining the specific heat W of the calorimeter knowing the evolution of the temperature after the combustion of a benzoic acid mass m.

$$\mathrm{W} = \frac{\Delta \mathrm{h} \times \mathrm{m} + \mathrm{C_1}}{\Delta \mathrm{T}} \quad (7.27)$$

Δh = Combustion heat of the benzoic acid (6318 cal/g=26452.2024 kJ/kg)

m= mass of the burnt sample

C_1 =correction due to the combustion heat of the wire: $C_1 = 2.3\ cal/cm$ (Chromium-nickel)

To do that, 1.1553 g of benzoic acid was burnt and the evolution of temperature was recorded in the calorimeter. Combustion of 1.4302 g of ester was performed in the same conditions and the following results helped for the calculation of HHV_s and of LHV_s of this ester.

Table 7.8 shows the records and calculated values for bomb calorimeter experiment. The bomb and its calorimeter were set up in an identical manner for each of the experiments performed. On the other hand, values of temperature and calculation of the variation of the temperature for B100 are recorded at Table 7.9 [181].

Results

Benzoic acid

m=1.1534 g

$L_f = 10\ Cm$

$\boldsymbol{L_c = 6\ Cm}$

Table 7. 8: Values recorded with experiments performed with a bomb calorimeter.

Time	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
T^0	26.73	26.73	26.73	26.73	26.73	26.73	26.73	26.75	26.8	26.85	26.86	26.9	26.92	26.93	26.95	26.95	27.03

B100

Masses (g)	L_{Cons_fus} (Cm)	V KOH (ml) 0.0725N	T_a (°C)	T_c (°C)	b (mn)	r_1 (°C/mn)	r_2 (°C/mn)	c (mn)	ΔT (°C)
1.4302	6	9.10	28.25	31.18	7.6375	0.008	0	12	2.92

m= 1.5590 g

$L_f = 10\ Cm, L_c = 6\ Cm$

Time	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
T^0	27.09	27.09	27.09	27.09	27.09	27.09	27.09	27.40	27.40	27.50	27.60	27.69	27.75	27.78	27.79	27.80	27.80

Table 7. 9: Record of measures of temperature and calculation of ΔT for B100 [181]

Mixture B_i	Masses (g)	L_{Cons_fus} (Cm)	V KOH (ml) 0.0725N	T_a (°C)	T_c (°C)	b (mn)	r_1 (°C /mn)	r_2 (°C /mn)	c (mn)	ΔT (°C)	W_h	HHV (kJ/kg)	LHV (kJ/kg)
B_{100}	1.5590	7	14.214	27.32	32.88	7.3348	0.016	-0.004	12	5.55734	11.7155	41700.337	28695.753

7.7 Optimisation of biofuel production from *ricinus communis* oil and ethanol

The objective of this section was to establish a model of biofuel production process from *ricinus communis* oil and methanol/ethanol. At the time the model was designed, ethanol was available. Hence, the reason experiments were performed with ethanol. In fact, production optimisation was achieved by changing the variables which included ethanol/ricinus communis oil ratio, KOH catalyst concentration, reaction time, reaction temperature, and rate of mixing to maximise biofuel production. Response surface methodology (RSM) technique was used. In addition, a first-order model was developed to predict the biofuel production if the production criteria is known. The model was validated using further experimental testing.

Therefore, the first step was the evaluation of potentialities and technical limits of the use of bioethanol as raw material for the production of ethyl esters in homogeneous basic catalysis in the Sub-Saharan region of Africa. *Ricinus communis* oil was selected as raw material because it was produced in Sub-Saharan African countries such as South Africa. And the pertinence of its use as potential source of biofuel was highlighted by many authors. To achieve this objective, the following methodology was used:

i) Firstly, the performance of ethyl esters was optimised according to three reaction factors: The ethanol/vegetable oil molar ratio, the amount of catalyst and the reaction temperature;
ii) Secondly, the effect of the water content of bioethanol was evaluated on the ethyl esters performance and their purity.

7.7.1 Materials and methods

In order to determine the optimal conditions for ethyl esters production, the statistical modelling of Y (%) performance was used, assuming a first degree model. The effects of the three reaction factors were evaluated at two distinct levels by a complete 2^3 factorial design: The MR (3:1 and 15:1) ethanol/vegetable oil molar ratio, the catalyst mass concentration C (0.5% and 2.0% KOH) and the reaction temperature T ($35\,℃\ and\ 80\,℃$). 23 tests were performed according to the 3 reaction factors. Other factors including the duration of the reaction and the speed of agitation were 2 hours and 360 rpm respectively.

To avoid bias, the 23 experimental tests were performed in random order as shown in Table 7.9. Design-Expert 8.0 for response surface methodology (RSM) was used for analyses as well as for the optimisation of the process. This version of the software has 3D surface plots

for category factors and a t-value effects Pareto chart as functionalities. Those feature are suitable for our work.

Table 7. 10: Complete 2^3 factorial design for ethyl esters production from Ricinus Communis.

Test N^0	Point	MR	C(%m)	T (^{0}C)	Y(%)
1	Central	9:1 (0)	1.25 (0)	57.5 (0)	55.0
2	Central	9:1 (0)	1.25 (0)	57.5 (0)	56.7
3	Central	9:1 (0)	1.25 (0)	57.5 (0)	55.2
4	factorial 1	3:1 (-1)	0.5 (-1)	35 (-1)	76.7
5	factorial 2	3:1 (-1)	0.5 (-1)	35 (-1)	71.4
6	factorial 2	15:1(+1)	0.5 (-1)	35 (-1)	79.2
7	factorial 2	15:1(+1)	0.5 (-1)	35 (-1)	86.4
8	factorial 2	15:1(+1)	0.5 (-1)	35 (-1)	78.0
9	factorial 2	15:1(+1)	0.5 (-1)	35 (-1)	81.2
10	factorial 2	15:1(+1)	0.5 (-1)	35 (-1)	85.5
11	factorial 2	15:1(+1)	0.5 (-1)	35 (-1)	83.5
12	factorial 2	15:1(+1)	0.5 (-1)	35 (-1)	85.0
13	factorial 3	3:1 (-1)	2.0 (+1)	35 (-1)	38.5
14	factorial 4	15:1(+1)	2.0 (+1)	35 (-1)	62.4
15	factorial 4	15:1(+1)	2.0 (+1)	35 (-1)	49.8
16	factorial 4	15:1(+1)	2.0 (+1)	35 (-1)	62.1
17	factorial 5	3:1 (-1)	0.5 (-1)	80 (+1)	71.7
18	factorial 5	3:1 (-1)	0.5 (-1)	80 (+1)	51.5
19	factorial 6	15:1(+1)	0.5 (-1)	80 (+1)	59.5
20	factorial 6	15:1(+1)	0.5 (-1)	80 (+1)	48.3
21	factorial 7	3:1 (-1)	2.0 (+1)	80 (+1)	32.3
22	factorial 8	15:1(+1)	2.0 (+1)	80 (+1)	62.6
23	factorial 8	15:1(+1)	2.0 (+1)	80 (+1)	61.1

In terms of the influence of water content, tests were performed at the optimum level determined earlier, with the same added distilled water of 1% to 12 % in ethanol. The

performance of various tests was determined after the purification of ethyl esters by washing with water. Monoglyceride, diglyceride and triglyceride residual contents were obtained by gas chromatography with internal calibration according to an adapted method of the EN14105 standard (European Committee for standardization, 2003).

7.7.2 Results and discussion

The statistical model of ethyl esters performance indicated that the factor that had the most effect was the catalyst concentration C, with a negative effect and significant interactions at the same time with the molar ratio of ethanol to *ricinus communis* oil and the temperature (Equation 5.3.2), confirming the results authors found in [183]. We have the following equation:

$$y = (57.5 \pm 1.4) + (6.4 \pm 1.5) * MR - (10.3 \pm 1.5) * C - (5.5 \pm 1.5) * T + (6.1 \pm 1.4) * MR * C + (4.9 \pm 1.5) * C * T + (3.3 \pm 1.5) * MR * C * T$$

7.7.3 Y(%) model in relation with the 3 variables of reaction (Equation)

The first degree model assumed was validated by the analysis of the decomposition of the variance and the values of the 3 following factors:

i) The F_{ae} of the factor with a 0.40 value weaker than the Fisher (3,7) $F_{2,14}$, which indicated a good adjustment of the model;

ii) The $F_{reg/res}$ factor of values 25, 9, and 4 times higher than the Fisher (5,1) $F_{6,16}$ which indicated that the set of tested factors were significant;

iii) The coefficient of determination $R^2 = 0.883$ indicated that the model is able to predict more than 88% of the total variance. The optimum performance, with a value of 84%, was obtained for the ethanol/vegetable oil molar ratio of 15:1, a KOH mass concentration of 0.5% and a temperature of 35°C.

In the experimental conditions corresponding to the optimum of the *ricinus communis* oil performance, the water content in ethanol had an unpredicted effect on the ethyl esters performance if we referred to other rare published studies [183] for which performances were less than 50% for water contents of the other 5%.

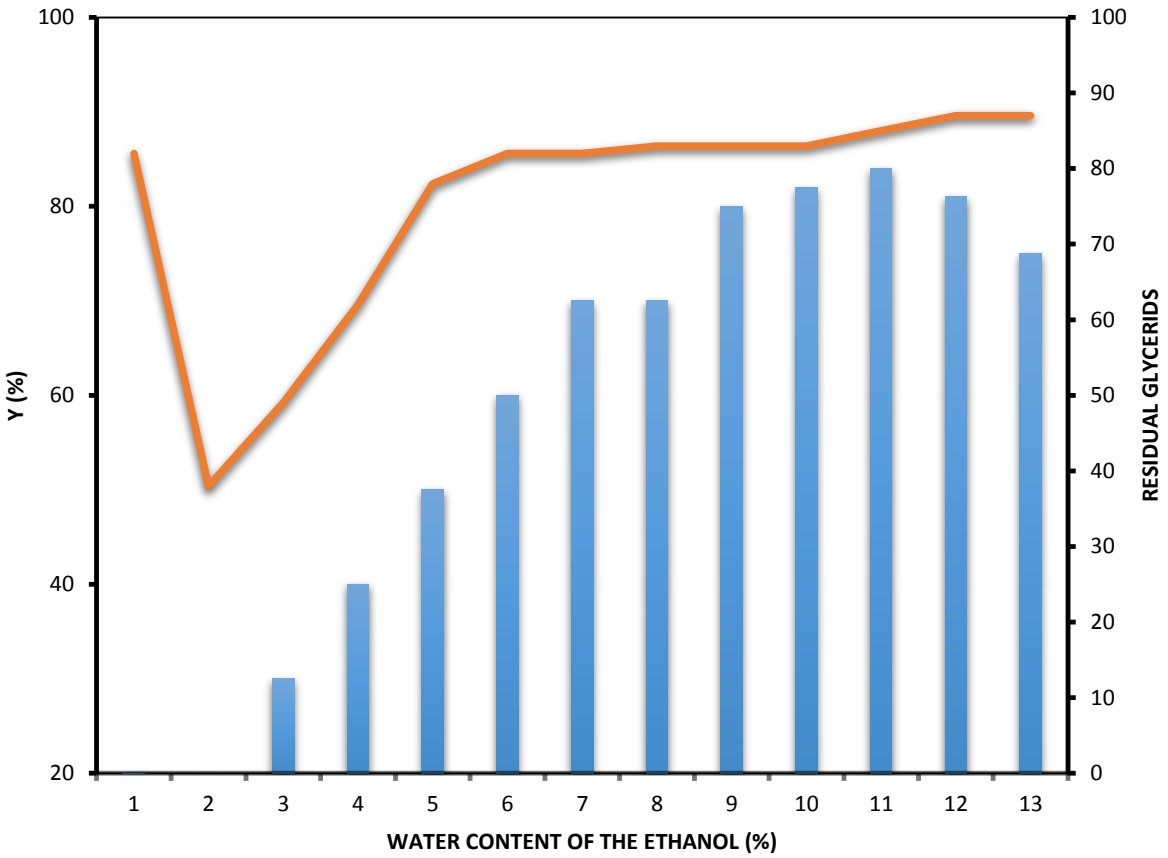


Figure 7. 22: Water content of ethanol and residual glycerides

In this study, for water content between 1%-2%, the performance progressively dropped first until about 49%. This could be explained by the initial formation of objects such as soaps in the presence of water. Then, for water contents between 3% to 5%, and 5% to 13%, the performance increased until when it reached values comparable to those obtained with the anhydrous bioethanol of 87%. It was the first time that such observations were reported. It can be explained by the role played by the ethyl esters performance which depended on a good separation of the phase rich in ethyl esters and in glycerol at the end of the reaction. Also, the excess of ethanol was mainly contained in the polar phase that was rich in glycerol, but had affinities with the phase rich in ethyl esters playing a role of a stabiliser of phases and not facilitating their separation. These affinities had the following consequences:

i) The lowering of the ethyl esters performance in the phase rich in ethyl esters;

ii) In certain cases, the introduction of the existence of one pseudo-homogeneous phase, including the inversions phase.

The presence of water could improve the affinity of ethanol for the phase rich in glycerol rather than for the phase rich in ethyl esters. This contributed to the facilitation of the separation of phases and the improvement of the ethyl esters performance. Moreover, it appeared that the purity of the synthetised ethyl esters was also improved with the hydrated ethanol ($\geq 5\% \, water$) compared to the anhydrous ethanol. It approached the European specifications for biofuel (0.8%, 0.2% and 0.2% for residual mono-, di-, and triglycerides respectively) [184]. The best performances and purity were achieved for a water content of ethanol higher or equal to 5%. Results indicated that the hydrated bioethanol at 87% can be a reactive of interest for the biofuel production, and that even in the presence of low additional quantities of water, the performance and purity of ethyl esters extracted can be optimised.

Chapter 8: General Conclusion and Recommendations for future work

Technically, any vegetable oil can be used to produce biodiesel but considerations related to the cost of production, the performance and ecobalance exclude number of candidates. Our choice of modelling the biodiesel production process from *ricinus communis* oil was based on human and climatic considerations, because the *ricinus communis* is a non-edible plant that cannot compete with foodstuffs destined to human consumption. The *ricinus communis* oil as source of biodiesel production is not new and its physico-chemical proprieties position it as a good product to be used. In a continent where fossil fuels are becoming scarce and expensive, this has triggered an extensive research in the area of biofuel.

A new biodiesel production process model was constructed in this work by using the Pareto-Levy law. This mathematical approach was used in the model to evaluate the biodiesel production process based on *ricinus communis* oil. And the model took into consideration the catalyst, the methanol to triglycerides ratio, the temperature and time of reaction. It was compared to experimental data which produced positive results. And its accuracy provided us with good and encouraging indications about the way the model behaves which is extremely close to experimental data. Also, the model allowed the better study and better analyse parameters influencing the transesterification operation. Simulations performed from this model showed very well the influence of various parameters and fits with experimental data found in the literature. More experimental tests in labs are needed in order to validate the theoretical model developed and this, in order to study the feasibility of the production of biodiesel from other local biomasses, evaluate the various costs of biodiesel production in the Sub-Saharan African context.

In conclusion, the modelling of various operational parameters considered during the all cycle of biodiesel production from *ricinus communis* oil has been performed. The reaction of transesterification has been performed in the presence of 1% of sodium metasilicate (Na_2SiO_3). The influence of operational parameters, including the temperature and the time of reaction, the alcohol/oil molar ratio and the massic quantity of the catalyst have been studied.

The extraction allowed to recover about 13% of lipids (lipid in g/dry material in g). The transesterification by acidic catalysis (Na_2SiO_3) in the presence of methanol leads to performances in biodiesel of 54% (biodiesel in g/lipids in g). A reaction time of 10 minutes (3 times lower than the time suggested in the literature) at a temperature of 67^0C has been

achieved. In operating under 1% mass of catalyst and a 6:1 alcohol/oil molar ratio, an 86% biodiesel performance is obtained at 60^0C within 10 minutes of the reaction time or at an ambient temperature of 26^0C within 60 minutes of reaction time. Thus, when comparing to those in the literature, performances in biodiesel have been obtained but with reaction times or with a temperature of transesterification lower than those cited by various authors.

Recommendations for future work:

Further research are needed in the following areas:

- ✓ Currently, the cost of biofuel is almost three times more expensive than fossil fuel because the largest quantity of biofuel is produced outside the Sub-Saharan region of Africa. Therefore, biofuel is not competitive compared to fossil fuel under the current economic regime of the region as the product is mostly imported from other regions. However, biofuel can be produced from other non-edible oil feedstock crops of low cost oils and fats such as corn stalks and wheat straw which do not compete with food production and could also be converted into biofuel. The main issue with processing these low-cost oils and fats is the high volume of free fatty acids they contain which makes it difficult the conversion into biofuel using common catalysts.
- ✓ In-depth analysis on biofuel is needed in order to obtain good quality properties including cetane number, density and viscosity. Furthermore, more experiments are needed in order to have other properties including flash point.
- ✓ The modelling of the transesterification reaction could also be carried out. This future investigation could cover process factors including the impact of temperature, catalyst and the ratio of methanol to metasilicate. In addition, a comparison between other mathematical models could be explored to determine the seemliness of the proposed mathematical model with the experimental data.
- ✓ And finally, knowing that perhaps the biggest challenge of biodiesel production process remains the high cost of feedstock including good quality edible oils and alcohol. *Ricinus communis* oil shows great potential to provide cheap, good quality and non-edible oil. In order to further decrease the cost of the process, especially the high cost of alcohol, it is important to choose a cheaper option: Taking into consideration that biodiesel can be synthesized using water, it would be beneficial to explore how various water content can influence the biodiesel production using another mathematical model.

REFERENCES

[1] Quadrelli, Elsje A., Centi G., Duplan J-L, and Perathoner S. *Carbon Dioxide Recycling: Emerging Large-Scale Technologies with Industrial Potential*. ChemSusChem 4 no. 9, pp. 1194-1215, 2011.

[2] Campbell P.K, Beer T., Batten D. *Greenhouse gas sequestration by algae energy and greenhouse gas life cycle studies.* Transport Biofuels Stream, CSIRO Energy Transformed Flagship PB 1, 2008.

[3] Elsolh N.E.M. *The Manufacture of Biodiesel from the used vegetable oil.* University of Kassel, 2011.

[4] Naik S. N., Goud V. V., Rout P. K. and Dalai A. K. *Production of first and second generation biofuels: a comprehensive review*. Renewable and Sustainable Energy Reviews, 14(2), pp. 578-597, 2010

[5] Banapurmath N.R., Tewari P.G. and Hosmath R.S. *Performance and emission characteristics of a DI compression ignition engine operated on Honge, Jatropha and sesame oil methyl esters.* Renewable energy, 33(9), pp.1982-1988, 2008

[6] Agarwal A.K. and Rajamanoharan K. *Experimental investigations of performance and emissions of Karanja oil and its blends in a single cylinder agricultural diesel engine*. Applied Energy, 86(1), pp.106-112, 2009

[7] Morrissey J. *The energy challenge in sub-Saharan Africa: A guide for advocates and policy makers*. https://www.oxfamamerica.org/static/media/files/oxfam-RAEL-energySSA-pt2.pdf (Accessed on 23 May 2017).

[8] Foster J.M. *Can Biofuels save Sub-Saharan Africa?* https://green.blogs.nytimes.com/2011/06/28/can-biofuels-save-sub-saharan-africa/?_r=0 (Accessed on 23 May 2017).

[9] Naik, S. N., Goud, V. V., Rout, P. K., & Dalai, A. K. *Production of first and second generation biofuels: a comprehensive review*. Renewable and Sustainable Energy Reviews, 14(2), 578-597, 2010

[10] Santana, G. C. S., Martins, P. F., Da Silva, N. D. L., Batistella, C. B., Maciel Filho, R., & Maciel, M. W. *Simulation and cost estimate for biodiesel production using castor oil*. Chemical engineering research and design, 88(5), 626-632, 2010

[11] Keith D. Wiebe, Meredith J. Soule and David E. Schimmelpfennig. *Agricultural Productivity for Sustainable Food Security in Sub-Saharan Africa.*

http://www.fao.org/docrep/003/X9447E/x9447e06.htm (Accessed on 26 March 2017).

[12] Office of the Director, Agricultural Development Economics Division, FAO. *The special challenge for sub-Saharan Africa*. High level expert forum - how to feed the World in 2050. 2009.

[13] Jumbe, C. B., Msiska, F. B., & Madjera, M. *Biofuels development in Sub-Saharan Africa: Are the policies conducive?* Energy Policy, 37(11), 4980-4986, 2009.

[14] Mutlu H., Meier MAR. *Castor oil as a renewable resource for the chemical industry*. Eur J Lipid Sci Technol; 112 : 10-30; 2010.

[15] HYPERLINK "http://www.kttanzania.org/learning-sharing/appropriate-technology/" http://www.kttanzania.org/learning-sharing/appropriate-technology/ (Accessed on 17 July 2016)

[15] African Development Bank and African Union. *Oil and Gas in Africa*. July 2009.

[16] Houghton, J. *Use of the truncated shifted pareto distribution in assessing size distribution of oil and gas fields*. Mathematical Geology, vol 20: 907-937, 1988.

[17] https://www.eia.gov/outlooks/ieo/transportation.php (Accessed on 11 July 2017)

[18] Mohr, S. H., & Evans, G. M. *Peak oil: testing Hubbert's curve via theoretical modelling*. Natural Resources Research, 17(1), 1-11. 2008.

[19] De Castro, C., Miguel, L.J. and Mediavilla, M. *The role of non conventional oil in the attenuation of peak oil*. Energy Policy, 37(5), pp.1825-1833. 2009.

[20] EIA, Energy Information Administration (EIA) (2009). Energy outlook 2009,

http ://www.eia.doe.gov/oiaf/'ieo/pdf0484(2009).pdf. Accessed on 30 May 2017.

[21] BABUSIAUX, D. and BAUQUIS, P.R. *Que penser de la raréfaction des ressources pétrolières et de l'évolution du prix du brut?* Les cahiers de l'économie, 66. 2007.

[22] Demirbas, A. *Progress and recent trends in biofuels.* Progress in energy and combustion science, 33(1), pp.1-18. 2007.

[23] Demirbas, A. *Biofuels sources, biofuel policy, biofuel economy and global biofuel projections.* Energy conversion and management, 49(8), pp.2106-2116. 2008.

[24] Hossain, A.S., Salleh, A., Boyce, A.N., Chowdhury, P. and Naqiuddin, M. *Biodiesel fuel production from algae as renewable energy*. American journal of biochemistry and biotechnology, 4(3), pp.250-254. 2008

[25] Apergis, N. and Payne, J.E. *Renewable energy consumption and economic growth: evidence from a panel of OECD countries*. Energy policy, 38(1), pp.656-660. 2010

[26] Das, D.C., Roy, A.K. and Sinha, N. *GA based frequency controller for solar thermal–diesel–wind hybrid energy generation/energy storage system.* International Journal of Electrical Power & Energy Systems, 43(1), pp.262-279. 2012.

[27] Pimentel, D. and Patzek, T.W. *Biofuels, solar and wind as renewable energy systems*. Benefits and risks. New York: Springer. 2008.

[28] Ajayi, O.C., Akinnifesi, F.K., Sileshi, G. and Chakeredza, S. *Adoption of renewable soil fertility replenishment technologies in the southern African region: Lessons learnt and the way forward*. In Natural Resources Forum (Vol. 31, No. 4, pp. 306-317). Blackwell Publishing Ltd. 2007.

[29] Bothast, R.J. and Schlicher, M.A. *Biotechnological processes for conversion of corn into ethanol*. Applied microbiology and biotechnology, 67(1), pp.19-25. 2005.

[30] Mohr, A. and Raman, S. *Lessons from first generation biofuels and implications for the sustainability appraisal of second generation biofuels.* Energy policy, 63, pp.114-122. 2013

[31] Lee, R.A. and Lavoie, J.M. *From first-to third-generation biofuels: Challenges of producing a commodity from a biomass of increasing complexity.* Animal Frontiers, 3(2), pp.6-11. 2013

[32] Stöcker, M. *Biofuels and biomass-to-liquid fuels in the biorefinery: Catalytic conversion of lignocellulosic biomass using porous materials*. Angewandte Chemie International Edition, 47(48), pp.9200-9211. 2008.

[33] Knothe, G. *Biodiesel and renewable diesel: a comparison.* Progress in energy and combustion science, 36(3), pp.364-373. 2010

[34] Sekoai, P.T. and Yoro, K.O. *Biofuel development initiatives in Sub-Saharan Africa: opportunities and challenges*. Climate, 4(2), p.33. 2016.

[35] Kashe, K., Kgathi, D.L., Murray-Hudson, M. and Mfundisi, K.B. *Assessment of benefits and risks of growing Jatropha (Jatropha curcas) as a biofuel crop in sub-Saharan Africa: a contribution to agronomic and socio-economic policies*. Journal of Forestry Research, pp.1-12. 2018.

[36] Ambrós, W.M., Lanzanova, T.D.M., Fagundez, J.L.S., Sari, R.L., Pinheiro, D.K., Martins, M.E.S. and Salau, N.P.G. *Experimental analysis and modeling of internal combustion engine operating with wet ethanol.* Fuel, 158, pp.270-278. 2015.

[37] Carriquiry, M.A., Du, X. and Timilsina, G.R. *Second generation biofuels: Economics and policies*. Energy Policy, 39(7), pp.4222-4234. 2011.

[38] Lennartsson, P.R., Erlandsson, P. and Taherzadeh, M.J. *Integration of the first and second generation bioethanol processes and the importance of by-products*. Bioresource technology, 165, pp.3-8. 2014.

[39] Dunn, M.E., Lapetz, J.M. and Welch, A.B., Westport Power Inc. *Apparatus and method for fuelling a flexible-fuel internal combustion engine*. U.S. Patent 9,856,837. 2018.

[40] Eggert, H. and Greaker, M. Promoting second generation biofuels: does the first generation pave the road?. *Energies*, *7*(7), pp.4430-4445. 2014

[41] Singh, A., Olsen, S.I. and Nigam, P.S. *A viable technology to generate third‐generation biofuel*. Journal of Chemical Technology and Biotechnology, 86(11), pp.1349-1353. 2011

[42] Xia, A., Herrmann, C. and Murphy, J.D. *How do we optimize third‐generation algal biofuels?*Biofuels, Bioproducts and Biorefining, 9(4), pp.358-367. 2015

[43] Moncada, J., Tamayo, J.A. and Cardona, C.A. *Integrating first, second, and third generation biorefineries: incorporating microalgae into the sugarcane biorefinery.* Chemical Engineering Science, 118, pp.126-140. 2014.

[44] Alaswad, A., Dassisti, M., Prescott, T. and Olabi, A.G. *Technologies and developments of third generation biofuel production*. Renewable and Sustainable Energy Reviews, 51, pp.1446-1460. 2015.

[45] Milano, J., Ong, H.C., Masjuki, H.H., Chong, W.T., Lam, M.K., Loh, P.K. and Vellayan, V. *Microalgae biofuels as an alternative to fossil fuel for power generation.* Renewable and Sustainable Energy Reviews, 58, pp.180-197.

[46] Zonouzi, A., Auli, M., Dakheli, M.J. and Hejazi, M.A. *Oil extraction from microalgae dunalliela sp. by polar and non-polar solvents.* International Journal of Biological, Biomolecular, Agricultural, Food and Biotechnological Engineering, 10(10), pp.634-37. 2016.

[47] Stephen, A.J., Archer, S.A., Orozco, R.L. and Macaskie, L.E. *Advances and bottlenecks in microbial hydrogen production*. Microbial biotechnology, 10(5), pp.1120-1127. 2017.

[48] Dadak, A., Aghbashlo, M., Tabatabaei, M., Najafpour, G. and Younesi, H. *Exergy analysis as a tool for decision making on substrate concentration and light intensity in photobiological hydrogen production*. Energy Technology, 4(3), pp.429-440. 2016.

[49] Crutzen, P.J., Mosier, A.R., Smith, K.A. and Winiwarter, W. *N_2O release from agro-biofuel production negates global warming reduction by replacing fossil fuels*. In Paul J. Crutzen: A pioneer on atmospheric chemistry and climate change in the anthropocene (pp. 227-238). Springer, Cham. 2016.

[50] Sydney, E.B., Sturm, W., de Carvalho, J.C., Thomaz-Soccol, V., Larroche, C., Pandey, A. and Soccol, C.R. *Potential carbon dioxide fixation by industrially important microalgae*. Bioresource technology, 101(15), pp.5892-5896. 2010.

[51] Pucci, A., Mikitenko, P. and Zuliani, M., IFP Energies Nouvelles. *Process for separating ethyl tert-butyl ether and ethanol*. U.S. Patent 5,348,624. 1994.

[52] Gayubo, A.G., Valle, B., Aguayo, A.T., Olazar, M. and Bilbao. *Pyrolytic lignin removal for the valorisation of biomass pyrolysis crude bio-oil by catalytic transformation*. Journal of chemical technology and biotechnology, 85(1), pp.132-144. 2010.

[53] Pinto, A.C., Guarieiro, L.L.N., Rezende, M.J.C., Ribeiro, N.M., Torres, E.A., Lopes, W.A., de P. Pereira, P.A., de Andrade, J.B., J. *Braz. Chem. Soc*. 16(6B): p. 1313-1330. 2005

[54] South African Department of Minerals and Energy. *Biofuels Industrial Strategy of the Republic of South Africa.* December 2007.

[55] https://www.ensafrica.com/news/biofuels-in-South-Africa (Accessed on 11-Sept-2018).

[56] Gui, M.M., Lee, K.T. and Bhatia, S. *Feasibility of edible oil vs. non-edible oil vs. waste edible oil as biodiesel feedstock*. Energy, 33(11), pp.1646-1653. 2008.

[57] Kostik, V., Memeti, S. and Bauer, B. *Fatty acid composition of edible oils and fats*. Journal of Hygienic Engineering and Design, 4, pp.112-116. 2013.

[58] Teh, S.S. and Birch, J. *Physicochemical and quality characteristics of cold-pressed hemp, flax and canola seed oils*. Journal of Food Composition and Analysis, 30(1), pp.26-31. 2013.

[59] Kartika, I.A., Yani, M., Ariono, D., Evon, P. and Rigal, L. *Biodiesel production from jatropha seeds: solvent extraction and in situ transesterification in a single step*. Fuel, 106, pp.111-117. 2013.

[60] Yoshimoto, Y., Kinoshita, E., Shanbu, L. and Ohmura, T. *Influence of 1-butanol addition on diesel combustion with palm oil methyl ester/gas oil blends*. Energy, 61, pp.44-51. 2013.

[61] Bettahar, Z., Cheknane, B. and Boutemak, K. *Etude de la transestérification d'un mélange des huiles usagées pour la production du biodiesel*. Revue des Energies Renouvelables, 19(4), pp.605-615. 2016.

[62] Song, M., Pei, H., Hu, W. and Ma, G. *Evaluation of the potential of 10 microalgal strains for biodiesel production*. Bioresource technology, 141, pp.245-251. 2013.

[63] Liu, H., Jiang, S., Wang, J., Yang, C., Guo, H., Wang, X. and Han, S. *Fatty acid esters: a potential cetane number improver for diesel from direct coal liquefaction.* Fuel, 153, pp.78-84. 2015.

[64] Rendón Hernández, F.J., Gutiérrez-Díaz, J.L., Silva Martínez, S., Hernández, J.A. and Álvarez-Gallegos, A. *Castor biodiesel-diesel blend to power a diesel engine: evaluation of the bus efficiency and emissions under driving conditions*. Interciencia, 40(1). 2015.

[65] Nemudzivhadi, V. and Masoko, P. *In vitro assessment of cytotoxicity, antioxidant, and anti-inflammatory activities of Ricinus communis (Euphorbiaceae) leaf extracts*. Evidence-Based Complementary and Alternative Medicine. 2014.

[66] Anastasi, U., Sortino, O., Cosentino, S.L. and Patanè, C. *Seed yield and oil quality of perennial castor bean in a Mediterranean environment.* International Journal of Plant Production, 9(1), pp.99-116. 2015.

[67] Dos Santos, C.M., Verissimo, V., de Lins Wanderley Filho, H.C., Ferreira, V.M., da Silva Cavalcante, P.G., Rolim, E.V. and Endres, L. *Seasonal variations of photosynthesis, gas exchange, quantum efficiency of photosystem II and biochemical responses of Jatropha curcas L. grown in semi-humid and semi-arid areas subject to water stress*. Industrial Crops and Products, 41, pp.203-213. 2013.

[68] Ramprasad R. and Bandopadhyay R. *Future of Ricinus communis after completion of the draft genome sequence* .Curr. sci.99(10): 1316-1318. 2010.

[69] Paul C.J, Van R., Lynell K. T. *The contribution of extrafloral nectar to survival and reproduction of the predatory mite Iphiseius degenerans on Ricinus communis*. Exper. Appl. Acarol. 23: 281–296. 1999.

[70] Malathi B ., Ramesh S ., Venkateswara K. R., Dashavantha V. R. *Agrobacterium-mediated genetic transformation and production of semilooper resistant transgenic castor (Ricinus communis L.).* Euphytica. 147: 441–449. 2006.

[71] Riina, R., Peirson, J.A., Geltman, D.V., Molero, J., Frajman, B., Pahlevani, A., Barres, L., Morawetz, J.J., Salmaki, Y., Zarre, S. and Kryukov, A. *A worldwide molecular phylogeny and classification of the leafy spurges, Euphorbia subgenus Esula (Euphorbiaceae)*. Taxon, 62(2), pp.316-342. 2013.

[72] Ali M. S., Ahmed S. and Saleem M. *Spirowallichiione: A rearranged multiflorane from Euphorbia wallichii Hook F.(Euphorbiaceae*). Molecules, 13(2), 405-411. 2008.

[73] Maghuly, F., Vollmann, J. and Laimer, M. *Biotechnology of euphorbiaceae (Jatropha curcas, Manihot esculenta, Ricinus communis)*. In Applied Plant Genomics and Biotechnology (pp. 87-114). 2015.

[74] Cheema, N. M., Muhammad A., Ghulam Q., Malik A. R. *Characterization of castor bean genotypes under various environments using SDS6PAGE of total storage proteins.* Pak. J. Bot. 42(3): 1797-1805. 2010.

[75] Gerard A., Amber W., Pablo, D. R., Agnes, P ., Jacques, R., Paul, K. *Worldwide genotyping of castor bean germplasm (Ricinus communis L.) using AFLPs and SSRs*. Genet. Resour. Crop. Evol. 55:365–378. 2008.

[76] Sujatha M., Reddy T.P. and Mahasi M.J. *Role of biotechnological interventions in the improvement of castor (Ricinus Communis L.) and Jatropha curcas L.* Biotechnology advances, *26*(5), pp.424-435. 2008.

[77] Ding, D. and Benson, D.A. *Simulating biodegradation under mixing-limited conditions using Michaelis–Menten (Monod) kinetic expressions in a particle tracking model*. Advances in water resources, 76, pp.109-119. 2015.

[78] Papastratos. *Modelisation, simulation, dynamique et optimisation d'un procede de fermentation ethanolique base sur un bioreacteur a membrane.* These ecole centrale, Paris; 1996.

[79] Oliveira S.C, Paiva T.C.B, Viscounti A.Z.S, and Giudici R. *Continuous alcoholic fermentation process: model considering loss of cell viability.* Bioproc. Eng., 20: 157-160, 1999.

[80] Costa A.C, Dechechi E.C, Silva F.L, Maugeri F., and Maciel R. *Simulated dynamics and control of an extractive alcoholic fermentation.* Appl. Biochem. Biotechnol. 84: 577-593, 2000.

[81] Lafforgue. *Fermentation alcoolique continue en bioreacteur a membrane.* These INSA Toulouse, 1988.

[82] Groot W.J, Sikkenk C.M, Waldram R.H, Van der lans R.G.J.M, and Muyben K.C.A.M. *Kinetics of ethanol production by baker's yeast in an integrated process of fermentation and microfiltration.* Bioprocess Eng. 8: 39-47, 1992.

[83] Warren R.K, Hill G.A, and Macdonald D.G. *Continuous cell recycle fermentation to produce ethanol.* Trans. Inst. Chem. Eng., 72: 149-157, 1994.

[84] Singh A. and Fernando S. *Reaction Kinetics of Soybean Oil Transesterification Using Heterogenous Metal Oxide Catalysts.* Chemical Engineering & Technology , *30* (12): 1716, 2011.

[85] De Lama da Silver N. *Biodiesel Production from Castor Oil: Optimisation of Alkaline Ethanolysis*. Journal of Energy and Fuels , 23 (11): 3140, 2009.

[86] Jain S. and Sharma M. *Biodiesel Production from Jatropha Curcas Oil*. Journal of Renewable and Sustainable Energy Reviews , 63 (10): 3140, 2010.

[87] Likozar, B. and Levec, J. *Transesterification of canola, palm, peanut, soybean and sunflower oil with methanol, ethanol, isopropanol, butanol and tert-butanol to biodiesel: Modelling of chemical equilibrium, reaction kinetics and mass transfer based on fatty acid composition*. Applied Energy, 123, pp.108-120. 2014.

[88] Freedman, B., Butterfield, R.O. and Pryde, E.H. *Trans-esterification kinetics of soybean oil. In Journal of the American Oil Chemists Society (Vol. 62, No. 4, pp. 662-662).* 1608 BROADMOOR DRIVE, CHAMPAIGN, IL 61821-0489: AMER OIL CHEMISTS SOC. 1985.

[89] Mjalli F. and Hussain, M. *Approximate Predictive versus Self Tuning Adaptive Control Strategies of Biodiesel Reactors.* Industrial and Engineering Chemistry Research , 48 (24):11034. 2009.

[90] Xiao, Yang, et al. *Kinetics of the transesterification reaction catalyzed by solid base in a fixed-bed reactor*. Energy & fuels:5829-5833. 2010

[91] Al-Zuhair, Sulaiman, Fan Wei Ling, and Lim Song Jun. *Proposed kinetic mechanism of the production of biodiesel from palm oil using lipase*. Process Biochemistry 42.6: 951-960. 2007.

[92] Mittelbach M. and Trathnigg B. *Kinetics of alkaline catalyzed methanolysis of sunflower oil*. Lipid/Fett 92.4: 145-148. 1990.

[93] Noureddini H. and Zhu D. *Kinetics of transesterification of soybean oil.* Journal of the American Oil Chemists' Society 74.11: 1457-1463. 1997.

[94] Cheng L. *Study on membrane reactors for biodiesel production by phase behaviour of canola oil methanolysis in batch reactors.* Journal of Bioresource Technology: 101 (17). 1997.

[95] Komers K. *Kinetics and Mechanism of KOH - catalyzed methanolysis of rapeseed oil for biodeisel production.* European Journal of Lipid Science and Technology , 104 (11): 728. 2002.

[96] Talebian-Kiakalaieh, Amin, et al. *Transesterification of waste cooking oil by heteropoly acid (HPA) catalyst: optimization and kinetic model.* Applied Energy 102: 283-292. 2013

[97] Ferdous K., Deb A., Ferdous J., Uddin R., Khan M.R. and Islam M.A. *Preparation of Biodiesel from higher FFA containing castor oil*. Ijser 4: 401-406. 2013

[97] Marchetti J.M., Miguel V.U. and Errazu A.F. *Possible methods for biodiesel production*. Renewable and sustainable energy reviews, 11(6), pp.1300-1311. 2007.

[98] Bisen P.S., Sanodiya B.S., Thakur G.S., Baghel R.K. and Prasad G.B.K.S. *Biodiesel production with special emphasis on lipase-catalyzed transesterification*. Biotechnology letters, *32*(8), pp.1019-1030. 2010.

[99] Meyer U., Palmans A.R., Loontjens T. and Heise A. *Enzymatic ring-opening polymerization and atom transfer radical polymerization from a bifunctional initiator*. Macromolecules, 35(8), pp.2873-2875. 2002.

[100] Pollard D.J. and Woodley J.M. *Biocatalysis for pharmaceutical intermediates: the future is now*. TRENDS in Biotechnology, 25(2), pp.66-73. 2007.

[101] Bondioli P. *The preparation of fatty acid esters by means of catalytic reactions.* Topics in Catalysis, 27(1), pp.77-82. 2004.

[102] Nyce G.W., Lamboy J.A., Connor E.F., Waymouth R.M. and Hedrick J.L. *Expanding the catalytic activity of nucleophilic N-heterocyclic carbenes for transesterification reactions. Organic letters*, *4*(21), pp.3587-3590. 2002.

[103] Bagheri, S., Julkapli, N.M. and Yehye, W.A. *Catalytic conversion of biodiesel derived raw glycerol to value added products.* Renewable and Sustainable Energy Reviews, 41, pp.113-127. 2015.

[104] Demirbaş A. *Biodiesel from vegetable oils via transesterification in supercritical methanol*. Energy conversion and management, *43*(17), pp.2349-2356. 2002.

[105] Rao, P.V. and Ramesh, S. *Optimization of Biodiesel production parameters (Pongamia pinnata oil) by transesterification process*. Journal of Advanced & Applied Sciences (JAAS), 3, pp.84-88. 2015.

[106] Ali Y. and Hanna M.A. *Alternative diesel fuels from vegetable oils*. Bio resource Technology, *50*(2), pp.153-163. 1994.

[107] Wahlen B.D., Barney B.M. and Seefeldt L.C. *Synthesis of biodiesel from mixed feedstocks and longer chain alcohols using an acid-catalyzed method*. Energy & Fuels, 22(6), pp.4223-4228. 2008.

[108] Chen, K.S., Lin, Y.C., Hsu, K.H. and Wang, H.K. *Improving biodiesel yields from waste cooking oil by using sodium methoxide and a microwave heating system*. Energy, 38(1), pp.151-156. 2012.

[109] Vicente G., Martınez M. and Aracil J. *Integrated biodiesel production: a comparison of different homogeneous catalysts systems.* Bioresource technology, 92(3), pp.297-305. 2004.

[110] Meneghetti S.M.P., Meneghetti M.R., Wolf C.R., Silva E.C., Lima G.E., de Lira Silva L., Serra T.M., Cauduro F. and de Oliveira L.G. Biodiesel from castor oil: a comparison of ethanolysis versus methanolysis. *Energy & Fuels*, *20*(5), pp.2262-2265. 2006.

[111] Ma F., Clements L.D. and Hanna M.A. *The effects of catalyst, free fatty acids, and water on transesterification of beef tallow*. Transactions of the ASAE, 41(5), p.1261. 1998.

[112] Su, F. and Guo, Y. *Advancements in solid acid catalysts for biodiesel production.* Green Chemistry, 16(6), pp.2934-2957. 2014.

[113] Marchetti J.M. and Errazu A.F. *Esterification of free fatty acids using sulfuric acid as catalyst in the presence of triglycerides*. Biomass and Bioenergy, 32(9), pp.892-895. 2008.

[114] Van Gerpen J. *Biodiesel processing and production*. Fuel processing technology, 86(10), pp.1097-1107. 2005.

[115] Shibasaki-Kitakawa N., Honda H., Kuribayashi H., Toda T., Fukumura T. and Yonemoto T. *Biodiesel production using anionic ion-exchange resin as heterogeneous catalyst*. Bioresource Technology, *98*(2), pp.416-421. 2007.

[116] Demirbaş A. *Biodiesel fuels from vegetable oils via catalytic and non-catalytic supercritical alcohol transesterifications and other methods: a survey*. Energy conversion and Management, 44(13), pp.2093-2109. 2003.

[117] Demirbas A. *Biodiesel production from vegetable oils via catalytic and non-catalytic supercritical methanol transesterification methods*. Progress in energy and combustion science, 31(5), pp.466-487. 2005.

[118] Kildiran G., Yücel S.Ö. and Türkay S. *In-situ alcoholysis of soybean oil*. Journal of the American Oil Chemists' Society, *73*(2), pp.225-228. 1996.

[119] Dasari M.A., Goff M.J. and Suppes G.J. *Noncatalytic alcoholysis kinetics of soybean oil*. Journal of the American oil chemists' society, 80(2), pp.189-192. 2003.

[120] Yilmaz, N. *Comparative analysis of biodiesel–ethanol–diesel and biodiesel–methanol–diesel blends in a diesel engine*. Energy, 40(1), pp.210-213. 2012.

[121] Kulkarni M.G. and Dalai A.K. *Waste cooking oil an economical source for biodiesel: a review*. Industrial & engineering chemistry research, *45*(9), pp.2901-2913. 2006.

[122] Schwab A.W., Bagby M.O. and Freedman B. *Preparation and properties of diesel fuels from vegetable oils*. Fuel, 66(10), pp.1372-1378. 1987.

[123] Encinar J.M., Gonzalez J.F., Rodriguez J.J. and Tejedor A. *Biodiesel fuels from vegetable oils: transesterification of Cynara c ardunculus L. oils with ethanol*. Energy & fuels, 16(2), pp.443-450. 2002.

[124] Zheng S., Kates M., Dubé M.A. and McLean D.D. *Acid-catalyzed production of biodiesel from waste frying oil*. Biomass and Bioenergy, 30(3), pp.267-272. 2006.

[125] Wright H.J., Segur J.B., Clark H.V., Coburn S.K., Langdon E.E. and DuPuis R.N. *A report on ester interchange*. Journal of the American Oil Chemists' Society, 21(5), pp.145-148. 1944.

[126] Graille J., Lozano P., Pioch D., Geneste P., Finiels A., Moreau C. *Revue Française des Corps Gras*. (8-9): p. 311-316. Août-Sept. 1985.

[127] Zullaikah S., Lai C.C., Vali S.R. and Ju Y.H. *A two-step acid-catalyzed process for the production of biodiesel from rice bran oil*. Bioresource technology, 96(17), pp.1889-1896. 2005.

[128] Hamad B., De Souza R.L., Sapaly G., Rocha M.C., De Oliveira P.P., Gonzalez W.A., Sales E.A. and Essayem N. *Transesterification of rapeseed oil with ethanol over heterogeneous heteropolyacids*. Catalysis Communications, 10(1), pp.92-97. 2008.

[129] Canakci M. and Van Gerpen J. *Biodiesel production from oils and fats with high free fatty acids.* Transactions of the ASAE, 44(6), p.1429. 2001.

[130] Ghadge S.V. and Raheman H. *Biodiesel production from mahua (Madhuca indica) oil having high free fatty acids*. Biomass and Bioenergy, 28(6), pp.601-605. 2005.

[131] Dai, Y.M., Wu, J.S., Chen, C.C. and Chen, K.T. *Evaluating the optimum operating parameters on transesterification reaction for biodiesel production over a LiAlO2 catalyst*. Chemical Engineering Journal, 280, pp.370-376. 2015.

[132] Wang Y., Ou S., Liu P. and Zhang Z. *Preparation of biodiesel from waste cooking oil via two-step catalyzed process*. Energy conversion and management, 48(1), pp.184-188. 2007.

[133] Boocock D.G., Konar S.K., Mao V. and Sidi H. *Fast one-phase oil-rich processes for the preparation of vegetable oil methyl esters*. Biomass and Bioenergy, 11(1), pp.43-50. 1996.

[134] Noureddini H. and Zhu D. *Kinetics of transesterification of soybean oil*. Journal of the American Oil Chemists' Society, 74(11), pp.1457-1463. 1997.

[135] Darnoko D. and Cheryan M. *Kinetics of palm oil transesterification in a batch reactor*. Journal of the American Oil Chemists' Society, 77(12), pp.1263-1267. 2000.

[136] Zabeti M., Daud W.M.A.W. and Aroua M.K. *Activity of solid catalysts for biodiesel production: a review*. Fuel Processing Technology, 90(6), pp.770-777. 2009.

[137] Helwani Z., Othman M.R., Aziz N., Kim J. and Fernando W.J.N. *Solid heterogeneous catalysts for transesterification of triglycerides with methanol: a review*. Applied Catalysis A: General, 363(1), pp.1-10. 2009.

[138] Jothiramalingam R. and Wang M.K. *Review of recent developments in solid acid, base, and enzyme catalysts (heterogeneous) for biodiesel production via transesterification.* Industrial & Engineering Chemistry Research, 48(13), pp.6162-6172. 2009.

[139] Hill J., Nelson E., Tilman D., Polasky S. and Tiffany D. *Environmental, economic, and energetic costs and benefits of biodiesel and ethanol biofuels.* Proceedings of the National Academy of sciences, 103(30), pp.11206-11210. 2006.

[140] Lopez D.E., Goodwin J.G., Bruce D.A. and Furuta S. *Esterification and transesterification using modified-zirconia catalysts.* Applied Catalysis A: General, 339(1), pp.76-83. 2008.

[141] Abbaszaadeh, A., Ghobadian, B., Omidkhah, M.R. and Najafi, G. *Current biodiesel production technologies: a comparative review.* Energy Conversion and Management, 63, pp.138-148. 2012.

[142] Furuta S., Matsuhashi H. and Arata K. *Biodiesel fuel production with solid superacid catalysis in fixed bed reactor under atmospheric pressure.* Catalysis communications, 5(12), pp.721-723. 2004.

[143] Vicente G., Coteron A., Martinez M. and Aracil J. *Application of the factorial design of experiments and response surface methodology to optimize biodiesel production*. Industrial crops and products, 8(1), pp.29-35. 1998.

[144] Osatiashtiani, A., Lee, A.F., Granollers, M., Brown, D.R., Olivi, L., Morales, G., Melero, J.A. and Wilson, K. *Hydrothermally stable, conformal, sulfated zirconia monolayer catalysts for glucose conversion to 5-HMF*. Acs Catalysis, 5(7), pp.4345-4352. 2015.

[145] Jiang, J., Gándara, F., Zhang, Y.B., Na, K., Yaghi, O.M. and Klemperer, W.G. *Superacidity in sulfated metal–organic framework-808*. Journal of the American Chemical Society, 136(37), pp.12844-12847. 2014.

[146] Khan, N.A., Mishra, D.K., Ahmed, I., Yoon, J.W., Hwang, J.S. and Jhung, S.H. *Liquid-phase dehydration of sorbitol to isosorbide using sulfated zirconia as a solid acid catalyst*. Applied Catalysis A: General, 452, pp.34-38. 2013.

[147] Okuhara T., Kimura M., Kawai T., Xu Z. and Nakato T. *Organic reactions in excess water catalyzed by solid acids.* Catalysis today, 45(1), pp.73-77. 1998.

[148] Furuta S., Matsuhashi H. and Arata K. *Biodiesel fuel production with solid amorphous-zirconia catalysis in fixed bed reactor.* Biomass and Bioenergy, 30(10), pp.870-873. 2006.

[149] Lopez D.E., Suwannakarn K., Bruce D.A, Goodwin J.G. *Esterification and transesterification on tungstated zirconia: Effect of calcination temperature*. Journal of Catalysis. 247(1): 43-50. 2007.

[150] Huang Y.Y., Zhao B.Y., Xie Y.C. *Modification of sulfated zirconia by tungsten oxide: Acidity enhancement and structural characterization*. Applied Catalysis A: General. 171(1): 75-83. 1998.

[151] Furuta S., Matsuhashi H., Arata K. *Catalytic action of sulfated tin oxide for etherification and esterification in comparison with sulfated zirconia.* Applied Catalysis A: General. 269(1): 187-191. 2004.

[152] Ferreira D.A., Meneghetti M.R., Meneghetti S.M, Wolf C.R. *Methanolysis of soybean oil in the presence of tin (IV) complexes*. Applied Catalysis A: General. 317(1): 58-61. 2007.

[153] Badday A.S, Abdullah A.Z, Lee K.T. *Optimization of biodiesel production process from Jatropha oil using supported heteropolyacid catalyst and assisted by ultrasonic energy*. Renewable energy, 50:427-432. 2013

[154] Kawashima A, Matsubara K., Honda K. *Acceleration of catalytic activity of calcium oxide for biodiesel production*. Bioresource Technology. 100(2): 696-700. 2009.

[155] Chai F., Cao F., Zhai F., Chen Y., Wang X., Su Z. *Transesterification of vegetable oil to biodiesel using a heteropolyacid solid catalyst.* Advanced Synthesis & Catalysis. 349(7): 1057-1065. 2007.

[156] Morin P., Hamad B., Sapaly G., Rocha M.C, De Oliveira P.P, Gonzalez W.A, Sales E.A. Essayem N. *Transesterification of rapeseed oil with ethanol: Catalysis with homogeneous Keggin heteropolyacids.* Applied Catalysis A: General. 330:69-76. 2007.

[157] Narasimharao K., Brown D.R., Lee A.F., Newman A.D., Siril P.F, Tavener S.J., Wilson K. *Structure-activity relations in C_s-doped heteropolyacid catalysts for biodiesel production.* Journal of catalysis. 248(2): 226-234. 2007.

[158] Bokade V.V, Yadav G.D. *Synthesis of bio-diesel and bio-lubricant by transesterification of vegetable oil with lower and higher alcohols over heteropolyacids supported by clay (K-10*). Process safety and Environmental Protection. 85(5): 372-377. 2007.

[159] Bokade V.V, Yadav G.D. *Transesterification of edible and nonedible vegetable oils with alcohols over heteropolyacids supported on acid-related clay*. Industrial and engineering Chemistry Research. 48(21): 9408-9415. 2009.

[160] Alsalme A., Kozhevnikova E.F., Kozhevnikov I.V. *Heteropolyacids as catalysts for liquid-phase esterification and transesterification.* Applied Catalysis A: General. 349(1): 170-176. 2008.

[161] Pesaresi L., Brown D.R, Lee A. f, Montero J.M, Williams H., Wilson K. *Cs-doped $H_4Si\ W_{12}O_{40}$ catalysts for biodiesel applications.* Applied Catalysis A: General. 360(1): 50-58. 2009.

[162] Persson M. and Wellinger A. *Biogas upgrading and utilisation*. International Energy Agency of Bioenergy. 2006.

[163] Nandi R., Sengupta S. Microbial production of hydrogen: an overview. Critical reviews in microbiology. 24(1): 61-84. 1998.

[164] Cuellar A.D., Webber M.E. *Cow power: the energy and emissions benefits of converting manure to biogas.* Environmental research Letters. 3(3): 034002. 2008.

[165] Taylor M.P., Eley K.L., Martin S., Tuffin M.I., Burton S.G., Cowan D.A. Thermophilic ethanologenesis: future prospects for second-generation bioethanol production. Trends in biotechnology. 27(7): 398-405. 2009.

[166] Cruz J.M, Ogunlowo A.S., Chancellor W.J., Goss J.R. *Vegetable oils as fuels for diesel engines*. Resources and Conservation. 6(1): 69-74. 1981

[167] Knothe G. *Dependence of biodiesel fuel properties on the structure of fatty acid alkyl esters*. Fuel processing technology. 86(10): 1059-1070. 2005.

[168] Ramos M.J., Fernandez C.M., Casas A., Rodriguez L., Perez A. *Influence of fatty acid composition of raw materials on biodiesel properties*. Bioresource technology. 100(1): 261-268. 2009.

[169] Newman M.E. *Power laws, Pareto distributions and Zipf's law*. Contemporary physics. 46(5): 323-351. 2005.

[170] Rosen B. *On sampling with probability proportional to size*. Journal of statistical planning and inference. 62(2): 159-191. 1997.

[171] Bickel P.J, Nair V.N and Wang P.C. Nonparametric inference under biased sampling from a finite population. The Annals of Statistics. 853-878. 1992.

[172] Wang J., Wang W., Huo S., Lee M., and Kollman P.A. Solvation model based on weighted solvent accessible surface area. The journal of Physical Chemistry B. 105(21). pp: 5055-5067. 2001.

[173] Sun J. and Michael Woodroofe. Semi-parametric estimates under biased sampling. Statistica Sinica. pp: 545-575. 1997.

[174] De Reffye P., Houllier F., et al. *Fitting a Functional-structural growth model with plant architectural data*. International Symposium on Plant Growth Modeling, Simulation, Visualization and their Applications, Beijing/China, p.108-117. 2003.

[175] Witek-Krowiak, A., Chojnacka, K., Podstawczyk, D., Dawiec, A. and Pokomeda, K. *Application of response surface methodology and artificial neural network methods in modelling and optimization of biosorption process*. Bioresource technology, 160, pp.150-160. 2014.

[176] Araujo, P. and Janagap, S. *Doehlert uniform shell designs and chromatography*. Journal of Chromatography B, 910, pp.14-21. 2012.

[177] Dagsvik, J.K., Jia, Z., Vatne, B.H. and Zhu, W. *Is the Pareto–Lévy law a good representation of income distributions?* Empirical Economics, 44(2), pp.719-737. 2013.

[178] Schweinsberg, J. *Dynamics of the evolving Bolthausen-Sznitman coalecent.* Electron. J. Probab, 17(91), pp.1-50. 2012.

[179] Jech, T. *Set theory*. Springer Science & Business Media. 2013.

[180] Feller, W. *On the Kolmogorov–Smirnov limit theorems for empirical distributions*. In Selected Papers I (pp. 735-749). Springer, Cham. 2015.

[181] Meerschaert, M.M. and Straka, P. *Inverse stable subordinators*. Mathematical modelling of natural phenomena, 8(2), pp.1-16. 2013.

[182] Ateeq, E.A., Abdelraziq, I.R. and Musameh, S.M. *Biodiesel viscosity and flash point determination* (Doctoral dissertation, Faculty of Graduate Studies Biodiesel Viscosity and Flash Point Determination By Eman Ali Ateeq Supervisor Prof. Issam Rashid Abdelraziq Co-Supervisor Prof. Sharif Mohammad Musameh This Thesis is Submitted in Partial Fulfillment of the Requirements for the Degree of Master of Physics, Faculty of Graduate Studies, An–Najah National University). 2015.

Bolthausen E. and Sznitman A.S. *On Ruelle's probability cascades and an abstract cavity method*. Communications in mathematical physics, 197(2), pp.247-276. 1998.

[183] Pitman J. *Combinatorial Stochastic Processes.* Ecole d'Eté de Probabilités de Saint-Flour XXXII-2002. Springer. 2006.

[184] Moroney, J.R. and Berg, M.D. *An integrated model of oil production*. The Energy Journal, pp.105-124. 1999.

[185] Kemp A. and Kasim S. *A regional model of oil and gas exploration in the UKCS*. Scottish Journal of Political Economy. 53(2): 198-221. 2006.

[186] Cleveland C.J and Kaufmann R.K. *Forecasting ultimate oil recovery and its rate of production: incorporating economic forces into the models of M. King Hubbert*. The Energy Journal. 1:17-46. 1991.

[180] Massart Pascal. *Concentration inequalities and model selection*. Vol. 6. Berlin: Springer, 2007.

[181] Maiti T. *Horvitz-Thompson Estimator*. In International Encyclopedia of Statistical Science. Springer Berlin Heidelburg. 637-638. 2011.

[182] Joshi H., Toler J., Moser B.R., and Walker T. *Biodiesel from canola oil using a 1:1 molar mixture of methanol and ethanol*. European journal of lipid science and technology. 111(5): 464-473. 2009.

[183] Vieitez I., Pardo M.J., Da Silva C., Bertoldi C., De Castilhos F., Oliveira J.V., Grompone M.A. and Jachmanian I. *Continuous synthesis of castor oil ethyl esters under supercritical ethanol*. The journal of Supercritical Fluids. 56(3): 271-276. 2011.

[184] Balat M. *An overview of biofuels and policies in the European Union*. Energy Sources, Part B, 2(2): 167-181. 2007.

[185] Favaro, S. and Walker, S.G. *Slice sampling σ-stable Poisson-Kingman mixture models.* Journal of Computational and Graphical Statistics, 22(4), pp.830-847. 2013.

[186] Bertoin J. *Levy processes*. Cambridge University Press. 1996.

[187] Mattheij, R.M., Rienstra, S.W. and ten Thije Boonkkamp, J.H. *Partial differential equations: modeling, analysis, computation* (Vol. 10). Siam. 2005.

[188] Makinde, O.D. and Aziz, A. *Second law analysis for a variable viscosity plane Poiseuille flow with asymmetric convective cooling*. Computers & Mathematics with Applications, 60(11), pp.3012-3019. 2010.

[189] Gupta, S.V. *VISCOMETRY FOR LIQUIDS*. SPRINGER INTERNATIONAL PU. 2016.

[190] Kalogirou, A., Poyiadji, S. and Georgiou, G.C. *Incompressible Poiseuille flows of Newtonian liquids with a pressure-dependent viscosity*. Journal of Non-Newtonian Fluid Mechanics, 166(7-8), pp.413-419. 2011.

[191] Nikolic, G., Zlatkovic, S., Cakic, M., Cakic, S., Lacnjevac, C. and Rajic, Z. *Fast fourier transform IR characterization of epoxy GY systems crosslinked with aliphatic and cycloaliphatic EH polyamine adducts*. Sensors, 10(1), pp.684-696. 2010.

[192] Sousa, L.L., Lucena, I.L. and Fernandes, F.A. *Transesterification of castor oil: effect of the acid value and neutralization of the oil with glycerol.* Fuel Processing Technology, 91(2), pp.194-196. 2010.

[193] Ramezani, K., Rowshanzamir, S. and Eikani, M.H. *Castor oil transesterification reaction: a kinetic study and optimization of parameters*. Energy, 35(10), pp.4142-4148. 2010.

[194] Delshad, A.B., Raymond, L., Sawicki, V. and Wegener, D.T. *Public attitudes toward political and technological options for biofuels.* Energy Policy, 38(7), pp.3414-3425. 2010.

[195] Gavrilescu, M. *Biorefinery systems: an overview*. In Bioenergy Research: Advances and Applications (pp. 219-241). 2014.

[196] Saha, S., Ahamed, J.U., Razzaq, M.A., Fahadullah, S.M., Barman, H. and Bala, S.K. *Production of Biodiesel from Waste Vegetable Oil*. In Proceedings of the International Conference on Mechanical Engineering and Renewable Energy (pp. 26-29). 2015.

[197] Hassan, M.H. and Kalam, M.A. *An overview of biofuel as a renewable energy source: development and challenges*. Procedia Engineering, 56, pp.39-53. 2013.

[198] Aghbashlo, M. and Demirbas, A. *Biodiesel: hopes and dreads*. Biofuel Research Journal, 3(2), pp.379-379. 2016.

[199] Russo, D., Dassisti, M., Lawlor, V. and Olabi, A.G. *State of the art of biofuels from pure plant oil*. Renewable and Sustainable Energy Reviews, 16(6), pp.4056-4070. 2012.

[200] Klein, A. *Mechanism for compaction and densification of large stover modules*. 2014.

[201] Ramírez-Verduzco, L.F., Rodríguez-Rodríguez, J.E. and del Rayo Jaramillo-Jacob, A. *Predicting cetane number,* kinematic *viscosity, density and higher heating value of biodiesel from its fatty acid methyl ester composition*. Fuel, 91(1), pp.102-111. 2012.

[202] Arnold, B.C. *Pareto Distributions Second Edition*. Chapman and Hall/CRC. 2015.

[203] Figueroa-López, J.E. *Nonparametric estimation for Lévy models based on discrete-sampling*. Lecture notes-monograph series, pp.117-146. 2009.

[204] Van Der Vaart, A. and Wellner, J.A. *A local maximal inequality under uniform entropy*. Electronic Journal of Statistics, 5(2011), p.192. 2011.

[205] Massart, P. *Concentration inequalities and model selection*. 2007.

[206] Oosterhoff, J. and van Zwet, W.R. *A note on contiguity and Hellinger distance*. In Selected Works of Willem van Zwet (pp. 63-72). Springer, New York, NY. 2012.

[207] Sadhanala, V., Wang, Y.X. and Tibshirani, R.J. *Total variation classes beyond 1d: Minimax rates, and the limitations of linear smoothers*. In Advances in Neural Information Processing Systems (pp. 3513-3521). 2016.

[208] Park, S., Rao, M. and Shin, D.W. *On cumulative residual Kullback–Leibler information*. Statistics & Probability Letters, 82(11), pp.2025-2032. 2012.

[209] Pierce, K.M., Kehimkar, B., Marney, L.C., Hoggard, J.C. and Synovec, R.E. *Review of chemometric analysis techniques for comprehensive two dimensional separations data*. Journal of Chromatography A, 1255, pp.3-11. 2012.

[210] Gu, C. *Smoothing spline ANOVA models* (Vol. 297). Springer Science & Business Media. 2013.

[211] Borras Jr, S.M., McMichael, P. and Scoones, I. eds. *The politics of biofuels, land and agrarian change*. Routledge. 2013.

Annexure A

Review of the Poisson process

Here, we review the essential notions of the measures of Poisson related to the Levy process. All demonstrations of the results enounced below are well known and can be found in [185] and [186].

A.1 Poisson measures and punctual Poisson process

Let E be a Polish space and υ a σ −finite measure on E.

Definition A.1.1

Let Φ be a random measure on E. We say that Φ is a Poisson measure of intensity υ if:

For all Borel set A of E, such that $\upsilon(A) < \infty$, Φ(A) follows a Poisson law of parameter $\upsilon(A)$.

For all disjoint Borel sets $B_1, \dots, B_n$, then Φ(B), ..., Φ(B) are independent random variables.

A Poisson measure Φ can be written as a denumerable sum of Dirac masses:

$$\Phi = \sum_{i \in I} \delta_{e_i}(de)$$

By adding a temporal component, meaning by considering the Poisson measure $\bar{\Phi}$ on the product space $[0, \infty[\times E$ of intensity $\mu = dt \otimes \upsilon$, it is possible to define the $\big(e(t)\big)_{t \geq 0}$ by ordering the atoms of Φ. We note:

$$\bar{\Phi} = \sum_{i \in I} \delta_{(t_i, e_i)}(dt, de)$$

And we can also show that all t_i are distinct.

Let γ be a cemetery state. The punctual Poisson process of intensity υ is defined by:

$$e(t) = \begin{cases} e_i \; if \; t = t_i, i \in I \\ \gamma \text{ if } t \notin \{t_i, i \in I\} \end{cases}$$

Therefore, we have: $\{t_i, i \in I\} = \{t \geq 0, e(t) \neq \gamma\}$.

Also, we will use the notation $\sum_{t \geq 0} \delta_{e_{(t)}}(de)$ for $\sum_{i \in I} \delta_{e_i}(de)$

Proposition A.1.1.

Let $\sum_{t\geq 0}\delta_{(t,e(t)}$ be a Poissonian cloud of intensity $\otimes\upsilon(de)$. Then, if $\upsilon(E)=\infty$, $\{t_i, i\in I\}$ is dense in [0,+∞), and if $\upsilon(E)<\infty$, the jumps form a Poisson punctual process of intensity $\upsilon(E)$ and the set of time of jumps is discrete.

Furthermore, waiting times T_n between the jumps are independent variables of the exponential law of parameter $\upsilon(E)$. Finally, the jumps $(e(T_n)_{n\geq 1}$ are i.i.d of $\frac{\upsilon(.)}{\upsilon(E)}$ law. Now, let's cite two important formulas, first the exponential formula is a very useful result in order to manipulate the Poisson punctual processes.

Proposition A.1.2

Let f be a Borel function on $E\cup\{\gamma\}$ with complex values, such that $f(\gamma)=0$ and

$\int_E \upsilon(dE)\left|1-e^{f(E)}\right|<\infty$. Then, $\forall\, t\geq 0$, we have:

$$\mathbb{E}\left(exp\left\{\sum_{0\leq s\leq t} f(e(s))\right\}\right)=exp\left\{-t\int_E \upsilon(d\epsilon)\left(1-e^{f(E)}\right)\right\}$$

Proposition A.1.3

Let f be a Borel function on $E\cup\{\gamma\}$ with positive values, and $\Phi=\sum_{i\in I}\delta_{e_i}(de)$ be a Poisson measure of intensity υ. We note $\mathcal{M}_E$ the set of $\sigma-finite$ measures on E. Let $F\colon \mathcal{M}_E\to\mathbb{R}_+$ be a measurable function. Then:

$$\int \Phi(de)f(e)F[\Phi-\sigma_e]=\mathbb{E}[F(\Phi)]\times\int \upsilon(de)f(e)$$

Or:

$$\mathbb{E}\left[\sum_{i\in I} f(e_i)F\left(\sum_{j\in I-\{i\}}\delta_{e_j}\right)\right]=\mathbb{E}[F(\Phi)]\times\int \upsilon(de)f(e)$$

This formula teaches us that a point randomly drawn from the Poissonian cloud according to Φ(de) is independent from the remaining of the cloud, and the last is of the same law than Φ.

A.2 The Levy process and stable subordinators

Definition A.2.1

The law of a random variable Y is called infinitely divisible if for all n, there exist i.i.d. random variables $Y_{n,1}, \dots, Y_{n,n}$ such that $Y = Y_{n,1} + \cdots + Y_{n,n}$. Also, the characteristic exponent Ψ of the υ law of a random variable Y is defined by:

$$\mathbb{E}\left(e^{i\langle Y,u\rangle}\right) = e^{-\Psi(\mathrm{u})}$$

The starting point of the study of infinitely divisible laws is the Levy-Khintchine formula that characterises the characteristics exponents of the infinitely divisible laws.

Theorem A.2.1

A function Ψ is the characteristic exponent of an infinitely divisible law on $\mathbb{R}^d$ if and only if there exist $a \in \mathbb{R}^d$, Q a quadratic form semi-positively defined on $\mathbb{R}^d$ and Λ a measure on $\mathbb{R}^d - \{0\}$ verifying $\int (1 \wedge |x|^2)\Lambda(\mathrm{dx}) < \infty$ such that for all $u \in \mathbb{R}^d$,

$$\Psi(\mathrm{u}) = \mathrm{i}\langle a, u\rangle + \frac{1}{2}Q(u) + \int_{\mathbb{R}^d}\left(1 - e^{i\langle u,x\rangle} + \mathrm{i}\langle u, x\rangle 1_{|x|<1}\right)\Lambda(\mathrm{dx}) \quad \text{(A.1)}$$

Definition A.2.2

A $(Z_t, t \geq 0)$ Levy process is a process of stationary independent growth.

Also, here we will suppose that $Z_0 = 0$, and that Z is continuous on the right and possesses left limits at all point, which is always possible. Then, for all $t \geq 0$, the Z_t law is infinitely divisible because:

$$\forall\, n \geq 0, Z_t = Z_{\frac{t}{n}} + \left(Z_{\frac{2t}{n}} - Z_{\frac{t}{n}}\right) + \cdots + \left(Z_t - Z_{\frac{(n-1)t}{n}}\right)$$

Let's note Ψ the characteristic exponent of Z_1. By the same argument, we show that:

$$\mathbb{E}\left[e^{iuZ_t}\right] = e^{-t\Psi(u)}$$

The law of Z is therefore entirely characterised by Ψ, and that we will also call it characteristic exponent of Levy. Reciprocally, tout any infinitely divisible law υ, we can associate a Levy process derived from 0 such that Z_1 will have the law υ. The characteristic exponent of a Levy process Z can be written under the (A.1) form, the measure Λ and the quadratic form Q are called Levy measure and Gaussian coefficient of the Levy process respectively. Therefore, the following propriety indicates how the measures of Poisson intervene in the Levy process theory.

Proposition A.2.1

Let$(Z_t, t \geq 0)$ be a Levy process. We pose $X_t \coloneqq Z_t - Z_{t-}$.

Then, the $(X_t)_{t\geq 0}$ jumps process is a punctual process of Poisson of intensity Λ. In other words, $\sum_{t\geq 0} \delta_{(t,X_t)}(dt.dx)$ is a Poissonian cloud of dt$\otimes\Lambda$ intensity.

Annexure B

Study of the penalised criteria for the selection of Gaussian mixture models using simulations

B.1 Notations

This section focuses on the study of the penalised criteria for the selection of Gaussian mixture models using performances simulations.

We show that the criteria, combined with the slope method, allows the definition of a penalised estimator whose prediction error is comparable to that of the oracle. We also compare the penalised estimator with deduced estimators for the Akaike Information Criterion (AIC), the Bayesian Information Criterion (BIC) and the Integrated Completed Likelihood Criterion (ICL).

Let's recall that those criterions are defined by:

$$crit_{AIC}(D) = \gamma_n(\hat{s}_D) + \frac{D}{n}$$

$$crit_{BIC}(D) = \gamma_n(\hat{s}_D) + \frac{Dln(n)}{2n}$$

$$crit_{ICL}(D) = crit_{BIC} - \frac{ENT}{n}$$

With:

$$ENT = \sum_{i=1}^{n}\sum_{k=1}^{n} z_{ik} ln(t_{ik})$$

Where z was given by the Maximum a posteriori Probabilities (MAP) rule.