Athens Institute for Education and Research
ATINER

ATINER's Conference Paper Series
ENV2015-1750

Identification of Representative Installations in the Best Available Techniques Determination Process: A Methodological Proposal

Damien Evrard
PhD Student
Mines Saint-Etienne
France

Valerie Laforest
Professor
Mines Saint-Etienne
France

Jonathan Villot
Lecturer
Mines Saint-Etienne
France

Rodolphe Gaucher
Head of Unit "Clean Technologies"
INERIS
France

Natacha Darmon
Head of Group Regulation & Installations
EDF-CIDEN
France

Sofia Bouhrizi
Engineer
EDF-CIDEN
France

An Introduction to ATINER's Conference Paper Series

ATINER started to publish this conference papers series in 2012. It includes only the papers submitted for publication after they were presented at one of the conferences organized by our Institute every year. This paper has been peer reviewed by at least two academic members of ATINER.

Dr. Gregory T. Papanikos
President
Athens Institute for Education and Research

This paper should be cited as follows:

Evrard, D., Laforest, V., Villot, J., Gaucher, R., Darmon, N. and Bouhrizi, S. (2015). **"Identification of Representative Installations in the Best Available Techniques Determination Process: A Methodological Proposal",** Athens: ATINER'S Conference Paper Series, No: **ENV2015-1750.**

Athens Institute for Education and Research
8 Valaoritou Street, Kolonaki, 10671 Athens, Greece
Tel: + 30 210 3634210 Fax: + 30 210 3634209 Email: info@atiner.gr URL: www.atiner.gr
URL Conference Papers Series: www.atiner.gr/papers.htm

ISSN: 2241-2891
14/12/2015

Identification of Representative Installations in the Best Available Techniques Determination Process: A Methodological Proposal

Damien Evrard

Valerie Laforest

Jonathan Villot

Rodolphe Gaucher

Natacha Darmon

Sofia Bouhrizi

Abstract

Best Available Techniques (BATs) in Europe have become crucial for the industrial sectors concerned with the Industrial Emission Directive (IED) and for other sectors like the nuclear industry. In the IED, the creation and revision process to define a list of BATs which will be used as a reference by operators and environmental authorities is known as the “Sevilla Process”. The decision process is based upon exchange of information between stakeholders, data analysis and expert judgement. Several methods to assess technique performances compared to BATs from reference documents or to support the determination and the application of BATs have been developed. However, existing tools integrate either the European or the local level but not both levels. This situation makes the application of BATs highly dependent on expert judgements. In order to contribute to strengthen the robustness of BAT determination and application, the authors argue that there is a need for an integrated method to help define a list of BATs and to support the decision for their application. This paper presents the first results of an on-going research aiming to answer these questions. It first introduces the context in which the BAT concept is used. In a second section, a method is proposed to select installations considered as representative of these issues in a view of illustrating the current situation in the sector and fuelling the discussions between experts. Finally, the proposed method is discussed and perspectives for selecting BATs are given.

Keywords: Best Available Technique, Industrial Emission Directive, Statistical analysis

Acknowledgments: Our thanks to CNIEL for allowing us to use the data of the French dairy industry they gathered in our study.

Introduction

A growing number of industrial sectors must apply Best Available Techniques (BATs). Since 1996, BATs have been applied in the industrial sectors within the scope of the Integrated Pollution Prevention and Control (IPPC) directive (European Commission, 1996). The role of these BATs has then been reinforced and strengthened in the Industrial Emission Directive (IED) (European Commission, 2010)

BATs have implied obligations at two levels: (1) the European Commission has had to set the appropriate process in order to provide sector-specific reference documents describing these techniques (called BREF); (2) local environmental authorities and operators have had to compare local installation performances with the information contained into these reference documents to define emission limit values.

In the IED, "Best Available Technique" is defined in article 3 as "*the most effective and advanced stage in the development of activities and their methods of operation which indicates the practical suitability of particular techniques for providing the basis for emission limit values and other permit conditions designed to prevent and, where that is not practicable, to reduce emissions and the impact on the environment as a whole*". Considered individually, the three words have the following meaning:

- "Technique" includes both the production technology used and the way in which the installation is designed, built, maintained, operated and decommissioned. In other words, a technique is not limited to a pollution abatement device but also encompasses management approaches such as environmental management systems.
- "Availability" means the technique considered is developed on a scale which allows an implementation in the industrial sector, under economically and technically viable conditions. These conditions take into consideration the costs and advantages, whether or not the technique is used or produced inside the Member State in question, and if it is reasonably accessible to the operator (Laforest, 2008).
- "Best" means that the technique considered is the most effective for achieving a high general level of protection of the Environment as a whole.

Some of the sectors which must apply BATs are outside the scope of the IED and, therefore, they have to refer to twelve criteria given in annex III of the IED to propose techniques corresponding to these definitions. This is the case, for example, in France for the nuclear industry (Arrêté, 2012). Nevertheless, these criteria are a general guidance for the Industry and are difficult to use on-site (Laforest, 2014).

In the IED, the process of BAT selection is formalised in an implementing decision (European Commission, 2012). This process, called the Sevilla Process, is an iterative process in which the selection of techniques in an industrial sector is carried out through expert judgement, and based on the collection of data from actual European sites. It is therefore the best basis for the determination of BATs, even in sectors outside the scope of the directive.

Furthermore, considering the high amount of information to consider in this process, the authors argue that a statistical approach may help experts in their decision-making process: (1) to identify key issues specific to a given industrial sector; (2) to classify installations according to their level of performance and peculiarities; (3) to select in each group, a sample of installations which can illustrate the whole.

This paper presents the first developments of a method for the selection of representative sites within an industrial sector based on a statistical analysis. These sites are meant to illustrate the sectoral issues and performances in order to select those which are likely to apply to BATs. Firstly, the context of this research work is drawn considering the Sevilla Process and the previous works on BAT selection. Secondly, the steps of the method are presented. Then, the limits and perspectives of the results are discussed. The paper ends with a conclusion about further studies to answer the limits identified.

Context of Application of Best Available Techniques

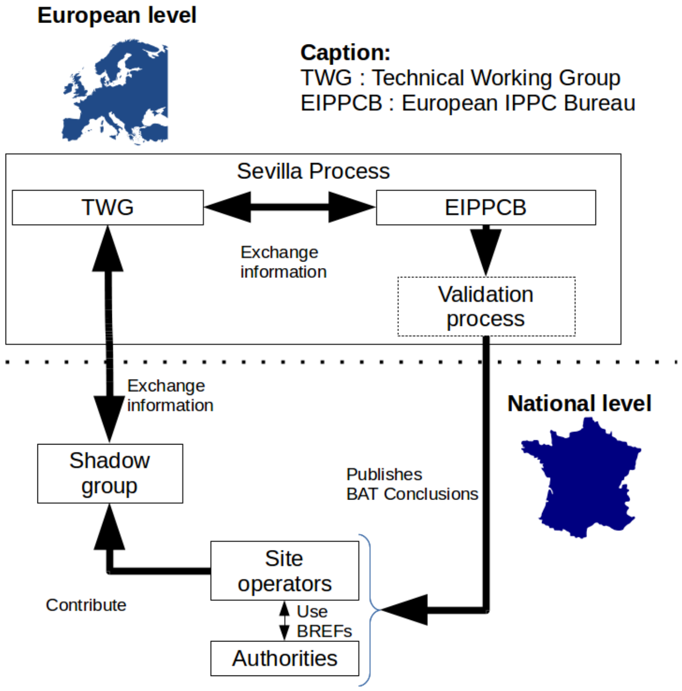
The Sevilla Process

The Sevilla Process is the official process to create or revise BREFs. It is based on information exchanges between Member States, the European Commission, the Industry, and environmental NGOs (European Commission, 2012). With the legal requirements associated with BAT-associated emission levels (BATAELs), the outcome of their revision has become strategic.

Figure 1 illustrates the interactions between the groups involved in the technical work within the Sevilla Process in order to publish a BREF at European and national levels, with the example of France. Thus, a Technical Working Group (TWG) composed of representatives from the organisations listed previously is formed in order to define the scope and plan the work. An important step is the collection of data about the performances of sites currently operating in Europe. The extensive analysis of information is requiring an increasingly complex data processing phase. The transition from this step of information analysis to the definition of BATs highly relies on the expertise of its actors with a risk of a biased assessment.

The objective of this on-going research is to look for a way to help, with a scientific method, the national experts in the shadow group (Figure 1) during their collection of the most relevant information to provide to the Sevilla Process. Shadow groups are meant to assist TWG members in their preparatory work for the Sevilla Process. They can be composed of the national TWG members, representatives from national authorities and industrials. In order to accomplish this, data about national installations which may illustrate the performances reachable thanks to a good application of BATs are sought. The Sevilla Process is used as a reference framework but the developments of this research are aimed to be applicable to other industrial sectors outside the scope of the IED, in particular to the nuclear industry.

Figure 1. *Exchange between National and European Actors in the IED*



The present on-going research follows the development of the revision of the reference document for the Food, Drink and Milk industry (EIPPCB, 2015). It especially focuses on the dairy sub-sector. All the following developments were experimented on data provided by the French dairy industry. For confidentiality reasons, details on the data used cannot be published in this article.

Previous Works on Bat Selection

Several existing methods to determine sectoral BATs were found in the scientific literature. The first one is a local BAT determination method which mainly relies on expert judgement supported by mostly qualitative data (Dijkmans, 2000; Polders et al., 2012) and was developed in Flanders. Then, two other methods aiming to reduce subjective elements are described. The first one by Geldermann and Rentz (2004) relies on more factual parameters to assess techniques, and the second one by the National Observatory of Athens (2006) uses mathematical approaches to guide the decision.

To sum up, these approaches to determine BATs are greatly influenced by their area of application (from Europe-wide or a given region) and their context (actors, end-user and data sources). Thus, either a method starts from a given sector to consider all relevant pollutants, or it begins from one particular environmental issue to select the relevant sectors which could contribute to an improvement thanks to the use of BATs. Moreover, they differ in the number of installations considered. When this number is too high to consider every single installation, a sampling is used (European Commission, 2012; Geldermann and Rentz, 2004).

The Sevilla Process and Dijkmans' method seem to focus on emissions, though also consider larger performance considerations, whereas Geldermann and Rentz focus on environmental impacts, and the National Observatory of Athens on economical valuation of impacts. Other mathematical approaches can be found in the literature, although in articles about the selection of techniques on-site.

Two other methods for the determination of sectoral BATs were found (Nicholas et al., 2000; Halog et al., 2001) with a focus on one very specific sector. Halog et al. (2001) have a concern for the assessment and comparison of techniques by expert judgement, whereas Nicholas et al. (2000) use Life Cycle Assessment. Moreover, in Halog et al. (2001), a preliminary step of determination of the key impacts is first carried out and then, the issues for BAT determination are considered as stakeholders' requirements. Therefore, these two methods only illustrate the issue of the "BAT candidate" assessment and the comparison of technique performance without detailing any representativeness assessment of the candidate techniques. In this aspect, they might be more appropriate if considered as tools at the installation level.

Overall, all the methods found focus on the selection of techniques and performance levels associated to techniques without explaining the upstream decision process based on the selection of installations. Such a preliminary process would be essential firstly to identify sectoral key environmental issues, then to select the installations which are bound to use techniques that might contribute to these issues, whether the level of performance is high or not. Another limit of these methods is that they are aimed to be applied in very specific contexts and do not consider both European and local applications illustrated in Figure 1.

Applied Methodological Approach

Data and Tools Used

Data Used

In parallel to the work undertaken by the French representatives in the Sevilla Process and the French Industry, a data collection was carried out by the dairy national federation in order to have an overview of the sector's performances and specific issues. Thus, the data on production, consumptions and emissions have been collected, in addition to a few elements of the applied techniques for 117 IED installations located in France. Therefore the data are either qualitative or quantitative, with levels of completeness varying from one variable to another or for one site to another. To sum up, the gathered data concern:

- **Generalities:** activities, production capacity, classification in the French regulation, date of construction, production of small series, seasonal activities;

- **Inputs:** nature and mass of milk and other dairy products, ingredients, water, and energies.
- **Outputs:** nature and mass of products (e.g. cheese, milk powder, whey, etc.).
- **Production and mitigation techniques:** curing, drying, filtration system, etc.
- **Emissions:** volume of wastewater, characteristics of wastewater, dust emissions, and waste.

Identified Specificities of the Sector

The main peculiarity of the sector is that most installations produce several products. Therefore, a site which is identified as a cheese manufacturer may also produce other products like whey. However, another one may also make whey but without cheese. This possibility of having various products made in the same installation makes the comparison between installations difficult.

Tools Used

Spreadsheet

The first tool to be used in this study is spreadsheet software in order to calculate basic statistics on the available quantitative data such as the averages, standard deviations, medians, minima, and maxima.

Statistical Analysis

In order to consider qualitative data such as the application of certain techniques, a statistical analysis software was used ("Coheris Analytics SPAD - Data Scientist - Logiciel Data Mining - Coheris," 2015). The method used in this study is the factorial analysis method of the Multiple Correspondence Analysis, which enables to group individuals according to their similarities and according to various characteristics.

Proposed Method

Overview

This study was decided to focus on the *key environmental issues* which had been defined during the kick-off meeting of the revision process for the FDM BREF (EIPPCB, 2015). Several of them were discarded due to a lack of available data: Total Organic Carbon and ammonium nitrogen. Thus, nine variables were considered to classify the installations: energy consumption, water consumption, chemical oxygen demand (BOC), biological oxygen demand (BOD), total suspended solids (TSS), chlorides (Cl^-), total nitrogen (TN), total phosphorus (TP), and dust emissions.

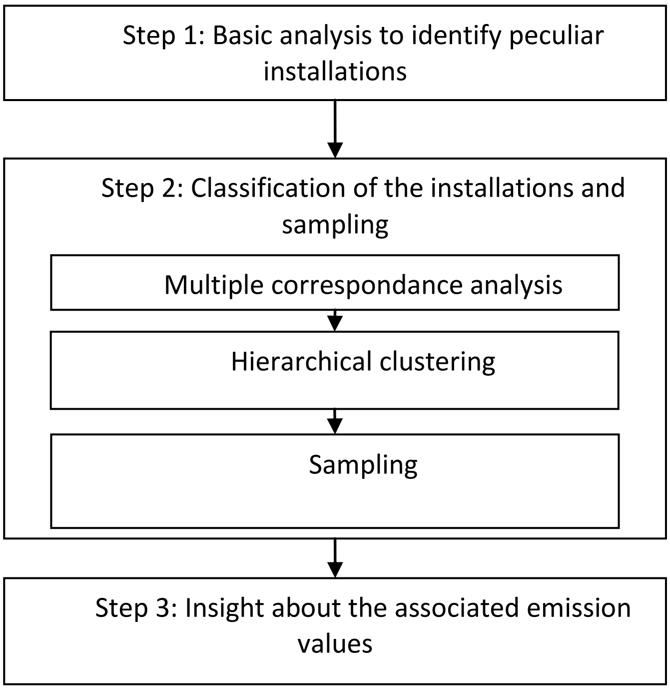
Due to the large amount of datasets to consider (9 variables for 117 installations, i.e. a maximum of 1,053 datasets), a statistical approach had seemed relevant to handle simultaneously so much information.

Besides, the installations differed in sizes and activities. Therefore, a common denominator appeared to be necessary to be able to compare all the sites in spite of their differences. Because of the lack of an official and clear rule to normalise the installations, three common denominators have been considered in this study:

- **Input of processed milk in tons:** this is the preferred approach in Europe (EIPPCB, 2015); all the products are considered as undistinguishable (e.g. 1 T of cheese = 1 T of milk powders). This approach is relevant for installations which make only one kind of product, but seem to be a source of over- or under- estimation when considering several products at once.
- **Input of dairy matter in tons, assessed from milk-equivalent:** in this approach, the various products are converted into milk-equivalent weight according to conversion factors. The factors used are the one stated in the French regulation (INERIS, 2015).
- **Output of dairy products in tons:** it focuses on the product but does not consider the differences between products.

Without prejudice about which denominator is the most relevant, the three of them have been tested in order to find out the differences in classification that they may imply. Considering these three denominators to calculate ratios for each of the nine variables, the statistical analysis itself was carried out according to the three steps presented in Figure 2.

Figure 2. ***Steps of the Proposed Method***



Step 1: Basic Analysis to Identify Peculiar Installations

The objective of this step was to identify the installations whose characteristics were peculiar, i.e. to highlight those whose performances were different from the rest of the French dairy sector. The study first applied to the raw data in order to compare the values of each installation to target-values, e.g. existing regulations. The aim here was to alert in case of deviation whether a site was out of bound. Then, a second analysis was carried out on the normalised data to search for installations deviating from the average of the sector. The aim here was also to alert about "extreme" installations. Therefore, experts' attention could be drawn on some of the sites to help them decide if they present peculiar issues or if they should be excluded from the analysis.

Further work should include the definition of cut-off rules in order to exclude odd installations from the beginning of the study. Nevertheless, exclusion would require ensuring that there are good reasons for it to prevent excluding singular but interesting installations. Therefore, this decision would be critical in a full-scale application and the decision may only be taken by expert judgment.

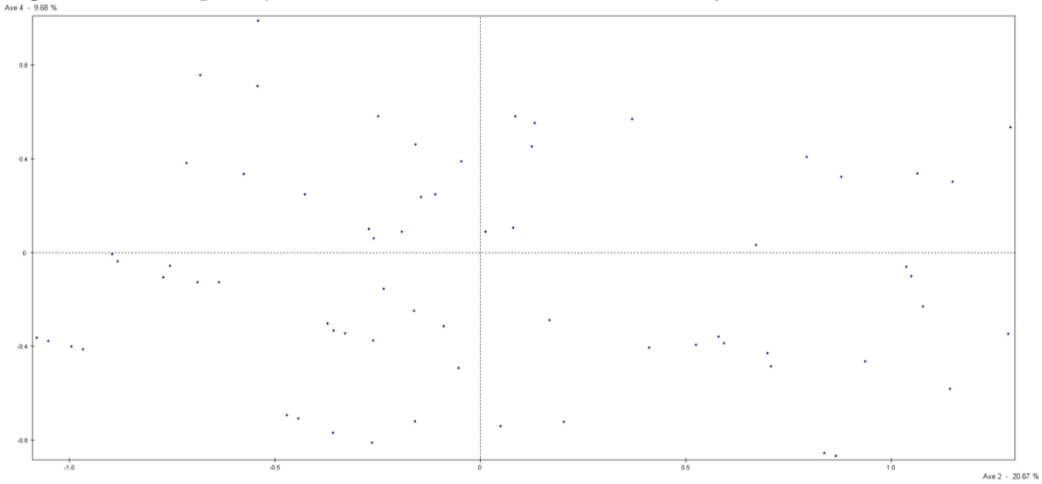
Step 2: Classification of the Installations and Sampling

The objective of this step was to classify the installations according to their performances described through the nine variables previously considered. Then, to deduce from the sub-groups identified one or several installations which illustrated their group as best as possible. In the study presented in this article, the approach was to focus on key environmental issues in order to suggest a short-list of installations to experts who would make the final decision. The final goal was to provide a decision-support tool to guide the experts' choice by considering simultaneously a large number of parameters and installations. Several operations were carried out on the normalised data (Figure 2).

Multiple Correspondance Analysis (MCA)

This factor analysis method enables us to study the relation between several nominal variables in a view to establish a classification based on their coordinates in a Euclidian space (Saporta, 2011). Its outcome is a graphical display which can be used to visualise the proximity between nominal categorical variables. However, considering the number of variables taken into account in this study, and the dispersion of the data, the graphical display was not enough to clearly identify groups of installations (dots in Figure 3). Therefore a formal classification method had to be used.

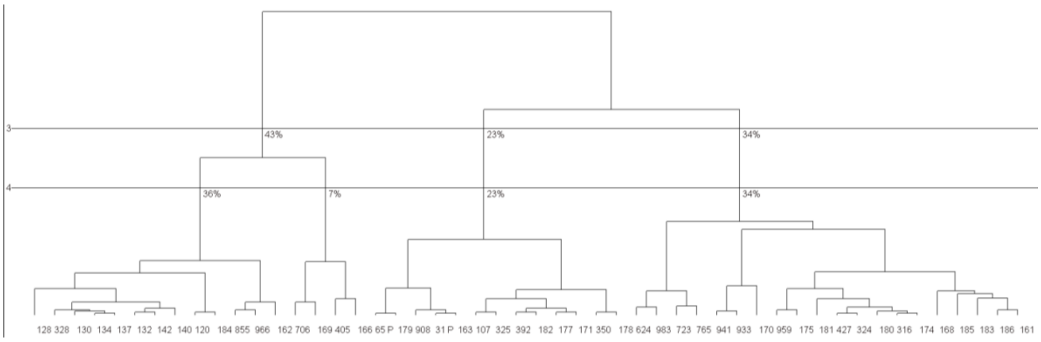
Figure 3. *Example of MCA Carried out on French Dairy Installations*



Hierarchical Clustering

Hierarchical clustering is an iterative method which calculates the dissimilarity between individuals, then pairs them according to an aggregation criterion to minimise cluster inertia. This operation is repeated until all individuals are in a cluster. The result is presented as a dendrogram as illustrated in Figure 4.

Figure 4. *Example of Dendrogram of Clustering of the French Dairy Installations*



This dendrogram enables us to select the most relevant number of classes. In Figure 4, two partitions would be the most relevant, statistically speaking. However, in order to better discriminate the installations, 4 groups were considered in order to take into account the specificities of the small second group. Thus, the first cluster would encompass 36% of the installations, the second 7%, the third 23%, and the fourth 34%.

Sampling of the Classes

The objective is to select a small number of installations which could illustrate their class in order to have a short-list of installations embodying the whole sector and its specificities. In this study, the closest individuals to the centre of each class were chosen, while introducing a few selection rules based on expert judgement. Thus:

- The list of the installations is ordered in ascending order by class and distance to its centre.
- Installations are considered according to their main activity.
- Once a main activity has been selected in a class, it cannot be selected a second time.
- If an installation is the only representative of its main activity, it will always be selected if it is the closest to the centre of its class.
- The selection within a class ends when the number of installations of a given main activity reaches a certain limit. For example, if 30% of the installations in a class are cheese-making installations and the following site to select is a milk powder installation whose activity is done by 20% of the installations in the class, then 50 % of the installations in the class should be covered and the selection in this class ends.
- If installations with the same main activity and distance to the centre of the class should be selected, all of them are selected.

In the end, a total of 19 installations were selected. In a real application, they would be submitted to experts for validation and possibly a new selection according to different selection criteria could be undergone.

Step 3: Insight about the Associated Emission Values

Thanks to the selected sites, it was possible to consider the emission values associated with them to assess the number of installations which would be excluded if they were used to create national emission values.

Thus, the percentage of installations whose values for the key environmental issues were below each selected site were assessed. Another use for this percentage could be to state a level of conformity for the selected sites. For example by deciding that 80% of the existing installation should be within the chosen limit value.

Limits and Perspectives

Common Denominator

The same method applied to the three common denominators showed that there were important differences in the classification of installations and sampling. Therefore, the impact on the clustering step should be further studied.

This could be answered by translating each qualitative variable (k discrete values) into k binary variables in each denominator. This should show invariant installations, i.e. installations which remain grouped together whatever the common denominator used.

Relevance of the Statistical Approach

General Considerations

This first exploration of the use of a statistical approach for installation classification and selection raised several questions about the depth of its relevance. Indeed, how far can it go to highlight installation specificities and to find out what technical characteristics influence their level of performance?

Another general issue is the management of missing data from the information collected on actual installations. Indeed, for some key environmental issues, 50% of the installations did not provide any values. How can such a situation be considered? What is its impact on the classification? What are the limits of applicability of the MCA?

Answers may be found thanks to sensitivity analyses, with or without their exclusion, or data reconstruction.

For Installation Classification

The first step of the proposed method implies to compare values according to their distance to a sectoral average but in cases where the values are very different from one installation to the next, the standard deviation may be too large to discriminate installations in a relevant manner. Therefore, are there any other approaches which may be more interesting such as the median and deviation to the median? How can expert judgment be used to target relevant sub-population without creating too important a bias?

These issues also concern the choice of the number of classes to consider. Is there a method which could help select the most appropriate number of classes to discriminate installations effectively? The meaning of the classes and the significance of a sub-division may bear part of the answer.

For Installation Selection

In the same class, several installations may appear to be at the same distance from its centre. So far, it is impossible to choose one rather than another and the decision seems to be dependant from experts. Is there a method to guide this decision or highlight the right information to help the decision-making?

A more fundamental issue is: what is a “representative” installation? In this article, it is presented as the closest to the centre of its class. In the Sevilla Process, “good-performers” are sought. In any cases, the criteria to select representative installations are still unclear. Therefore, a clear definition and indicators should be sought to assess the representativeness of installations.

Conclusions

To conclude, BAT determination is a complex issue even where a framework exists such as the Sevilla Process. Considering the large amount and wide variety of information to provide to decision makers, a statistical approach seems relevant although the availability of data makes it difficult to relate to every variable. Nevertheless, a performance-oriented approach seems to be relevant since the purpose of BATs, as stated in the IED directive is to *"achiev[e] a high general level of protection of the Environment as a whole"*.

Further study in this project shall include a more critical analysis of the contribution of the statistical approach to the whole issue of representative and BAT identification. In addition, application to other sectors shall help identify where a generic approach might be possible.

References

Arrêté du 7 février 2012 fixant les règles générales relatives aux installations nucléaires de base, 2012.

Coheris Analytics SPAD - Data Scientist - Logiciel Data Mining - Coheris, 2015.

Dijkmans, R., 2000. Methodology for selection of best available techniques (BAT) at the sector level. Journal of Cleaner Production 8, 11–21. doi:10.1016/S0959-6526(99)00308-X.

EIPPCB, 2015. Kick-off Meeting Report for the review of the Food, Drink and Milk Industries (FDM) BREF [WWW Document]. URL http://bit.ly/1P1LWAS [accessed 4.15.15], 45p.

European Commission, 1996. Council directive 96/61/EC of 24 September 1996 concerning integrated pollution prevention and control, 22p.

European Commission, 2010. Directive 2010/75/EU of the European Parliament and of the Council of 24 November 2010 on industrial emissions (integrated pollution prevention and control). Off J Eur Union 2010, L334/17-118.

European Commission, 2012. Commission implementing decision of 10 February 2012 laying down rules concerning guidance on the collection of data and on the drawing up of BAT reference documents and on their quality assuranc referred to in Directive 2010/75/EU of the European Parliament and of the Council on industrial emissions. Off J Eur Union 2012, L63/1-39.

Geldermann, J., Rentz, O., 2004. The reference installation approach for the techno-economic assessment of emission abatement options and the determination of BAT according to the IPPC-directive. Journal of Cleaner Production 12, 389-402. doi:10.1016/S0959-6526(03)00032-5.

Halog, A., Schultmann, F., Rentz, O., 2001. Using quality function deployment for technique selection for optimum environmental performance improvement. Journal of Cleaner Production 9, 387–394. doi:10.1016/S0959-6526(00)00080-9.

INERIS, 2015. Rubrique ICPE 2230. Réception, stockage, traitement, transformation etc. du lait [Reception, storage, treatment, processing, etc. of milk] [WWW Document]. AIDA. URL http://bit.ly/1m0uWiN [accessed 4.15.15].

Laforest, V., 2008. Applying best available technologies in environmental management accounting: from the definition to an assessment method. In: Environmental Management Accounting for Cleaner Production. 5th EMAN book, 29–47.

Laforest, V., 2014. Assessment of emerging and innovative techniques considering best available technique performances. Resources, Conservation and Recycling 92, 11–24. doi:10.1016/j.resconrec.2014.08.009.

National Observatory of Athens, 2006. Decision Aid Framework for the assessment of impacts from the introduction of Best Available Techniques in industry (DAF-BAT) (Final Project Report Executive Summary), 39p.

Nicholas, M. J., Clift, R., Azapagic, A., Walker, F. C., Porter, D. E., 2000. Determination of "Best Available Techniques" for Integrated Pollution Prevention and Control: A Life Cycle Approach. Process Safety and Environmental Protection 78, 193-203. doi:10.1205/095758200530637.

Polders, C., Van den Abeele, L., Derden, A., Huybrechts, D., 2012. Methodology for determining emission levels associated with the best available techniques for industrial waste water. Journal of Cleaner Production 29–30, 113–121. doi:10.1016/j.jclepro.2012.02.008.

Saporta, G., 2011. Probabilités, analyse des données et statistique [Probability, data analysis and statistics]. Editions TECHNIP, 656p.