

Report

Report no. 15/22

Estimating the CO₂ intensities of EU refinery products: Statistical Regression Methodology



Estimating the CO₂ intensities of EU refinery products:

Statistical Regression Methodology

This report was prepared by:

Damien Valdenaire, Science Executive, Concawe
Alireza Tehrani, Member Company Expert (TotalEnergies, STF2-RT Chairman)
Serge Mennecier, Independent Consultant

With the support: STF2-RT Members (Refinery Technology)

For RETMG (Refinery into the Energy Transition Management Group)

Reproduction permitted with due acknowledgement

KEYWORDS

Linear Programming, refinery modelling, CO₂ emissions, CO₂ intensity, refinery product, EU refining, restricted regression

INTERNET

This report is available as an Adobe pdf file on the Concawe website (www.concawe.eu).

NOTE

Considerable efforts have been made to assure the accuracy and reliability of the information contained in this publication. However, neither Concawe nor any company participating in Concawe can accept liability for any loss, damage or injury whatsoever resulting from the use of this information.

This report does not necessarily represent the views of any company participating in Concawe.

CONTENTS		Page
SUMMARY		VII
1.	BACKGROUND	1
1.1	OBJECTIVE	2
1.2	LIMITATIONS OF THE GENERALIZED MARGINAL APPROACH	2
1.2.1	Identified limitations of Generalized Marginal Approach	2
1.2.2	Main challenges identified by external stakeholders	3
1.3	LIFE CYCLE ASSESSMENT METHODOLOGY	3
1.4	AVERAGE VERSUS MARGINAL CO ₂ FOOTPRINTS	4
1.5	MAIN MODELING ASSUMPTIONS	7
1.5.1	Crude Slate	7
1.5.2	Other Refinery feedstocks	7
1.5.3	Refining finished products demand	8
1.5.4	Finished product specifications	8
1.5.5	Process unit capacities	9
1.5.6	Pricing	9
1.6	LP MODEL	9
1.6.1	Concawe LP model main features	10
1.6.2	Calibration	11
2.	NEW METHOD	13
2.1	BENEFITS	13
2.2	BASIC PRINCIPLE	14
2.3	IMPLEMENTATION	14
2.3.1	Data driven approach	14
2.3.2	Design Of Experiment (DOE)	15
2.3.3	Latin Hypercube Sampling (LHS)	15
2.3.4	Ordinary Least Squares regression (OLS)	16
3.	NUMERICAL APPLICATION	17
3.1	STEP 1 - DOE USING LHS METHOD	17
3.2	STEP 2 - DOE RESULTS - REFINING CO ₂ EMISSIONS	17
3.3	STEP 3 - RESTRICTED OLS REGRESSION - CO ₂ INTENSITIES	18
3.3.1	CO ₂ Intensity results	18
3.3.2	Industrial interpretation of the results	19
3.4	VERIFICATION OF REGRESSION ASSUMPTIONS	20
3.4.1	DOE cases	20
3.4.2	Regression results	22
4.	ADAPTATION TO LCA STANDARD REQUIREMENTS	25
4.1	STEP A - DECOMPOSITION OF BOTTOM OF THE BARREL PRODUCTS	25
4.2	STEP B - CO ₂ BALANCE	26
4.3	STEP C - ALLOCATION OF CO ₂ IMBALANCE	27
4.4	STEP D - FINAL SET OF POSITIVE CO ₂ INTENSITIES	30

5.	GASOLINE AND DIESEL WELL-TO-WHEELS LCA	32
	GLOSSARY	37
	APPENDIX 1 - DETAILED ASSUMPTIONS	38
	APPENDIX 2 - DOE ASSUMPTIONS	45
	APPENDIX 3 - OLS REGRESSION RESULTS	46
	APPENDIX 4 - STATISTICAL CONCEPTS	47

LIST OF FIGURES

Figure 1. A refining joint production system	5
Figure 2. Latin Hypercube Sampling (LHS) vs. random sampling	15
Figure 3. Kernel density curves of the DOE cases for the main products	21
Figure 4. Pearson correlation matrix of the DOE cases	22
Figure 5. Behavior of the residuals of the Restricted OLS model	23
Figure 6. Residuals distribution	23
Figure 7. Homoscedasticity of residuals	24
Figure 8. OLS regression results	46

LIST OF TABLES

Table 1. 2010 EU refinery product CO ₂ intensities (Concawe, 2017).....	1
Table 2. Carbon footprints of refined petroleum products, grams of CO ₂ equivalent per MJ LHV	6
Table 3. DOE - Product demand 2020, Million tons per annum.....	17
Table 4. DOE - Product demand 2019 and corresponding refining CO ₂ emissions, Million tons per annum.....	18
Table 5. CO ₂ intensities of refining finished products, ton of CO ₂ per ton of product	19
Table 6. CO ₂ balance verification.....	20
Table 7. Finished product mass balance, million tons per annum	25
Table 8. Categories 7 and 8 split, weight percent.....	26
Table 9. Categories 7 and 8 positive CO ₂ intensity calculation, tons of CO ₂ per tons of product	26
Table 10. CO ₂ balance	27
Table 11. Allocation keys of CO ₂ allocation methodologies - Weighted CO intensities, mass, and energy	28
Table 12. CO ₂ intensities resulting from allocation methodologies, ton of CO ₂ per ton of product	29
Table 13. Positive CO ₂ intensities of refining finished product categories, ton of CO ₂ per ton of product.....	30
Table 14. Positive CO ₂ intensities of refining finished product categories, grams of CO ₂ per megajoule of product	31
Table 15. Well-to-wheels CO ₂ intensities of Gasoline and Diesel, grams of CO ₂ equivalent per megajoule of product	32
Table 16. Countries included in the LP model.....	38
Table 17. Simplified BFD of the LP model.....	39
Table 18. Registration of Crude Oil Imports and Deliveries in the European Union, 2019.....	40
Table 19. LP Model crude slate	41
Table 20. EU crude slate main properties (estimated)	42
Table 21. LP model demand - EU 2019 adjusted refining system production, millions tons per annum.....	42
Table 22. Summary of LP model main finished products specifications	43
Table 23. LP model main process units capacities, millions tonnes per annum	44
Table 24. Finished products allocation to DOE categories - Base case mass balance, 2019 adjusted reference year	45

SUMMARY

This study aims at estimating the CO₂ intensity of petroleum products at the gate of an average EU refinery (refining step of the production process, excluding upstream and downstream emissions) using a novel approach, which reconciles economic relevance with accounting conventions in joint production industries. It revisits and finds a fair and logical way to distribute total CO₂ emissions from the EU refineries over the most significant finished products, selected from the refinery slate.

The refinery linear programming model of Concawe is used to generate randomly 500 consistent production level data around the calibrated reference year 2019.

Firstly, an innovative restricted regression methodology is applied to determine the linear coefficients which are equal to the CO₂ intensities of the EU finished refined products (see column “CO₂ intensities” in the table below).

Then, this novel data-driven framework, based on supervised machine learning technics, enables, firstly, to assess causal-based, as it captures the interactions between the process units and the products (necessary for co-production process), and is extended via the restricted regression to provide positive dataset. Secondly the figures are statistically valid (highly robust results) in the area of analysis (+5% of the demand). Lastly, similar to results derived from applying an LCA methodology (both attributional and consequential), this method gives positive CO₂ intensities for all products. Therefore, the last column in the table below is labeled as “LCA compliant, Positive CO₂ intensities”.

The final set of CO₂ intensities is summarized in the table below:

CO ₂ Intensities (Concawe LP model & OLS regression)	Base production Mt / a	CO ₂ Intensities t CO ₂ / t	LCA compliant Positive CO ₂ intensities t CO ₂ / t
1 LPG	24.3	0.3731	0.3501
2 NAPHTHA	37.4	0.2218	0.2082
3 GASOLINE	127.5	0.2687	0.2522
4 KEROSENE	51.9	0.2718	0.2551
5 DIESEL	174.5	0.3012	0.2827
6 HO MARINE DSL	80.6	0.2026	0.1901
7 DMF RMF 0.5%S	35.6	0.0116	0.0705
8 HSFO	30.3	-0.1905	0.0182
9 OTHERS	37.4	0.0602	0.0565
TOTAL CO ₂ calculated from CO ₂ intensities and Base production, Mt / a		131.5	131.5

Various statistical hypotheses are verified to ensure the credibility and robustness of the results.

It can be noted that the high sulphur fuel oil (HSFO) shows negative CO₂ intensity (2nd column). In a refinery, if the production of high sulphur fuel oil has to be

reduced, it would be converted into lighter products, through CO₂ intensive process units.

As described in the Concawe report n° 1/17 “Estimating the marginal CO₂ intensities of EU refinery products”, this Generalized Marginal approach captures **crucial interactions between processes and products**.

The results of this report cannot be applied to individual refineries and are **only valid for the EU refining on its global average**. Refineries in Europe are all different in their process unit’s configuration and level of complexity, and consequently in their emission intensity and in the product mix they produce.

Furthermore, the innovative methodology presented in this report is adapted to generated positive numbers for carbon intensity of the refinery products. These “LCA compliant Positive CO₂ intensities can be used as part of the “Eco-profiles” dataset for the **carbon footprint** of the **selected refined finished products** (Eco-profiles assess the environmental impacts along the production chain from «cradle-to-gate»).

In LCA world, it is common to report the CO₂ intensities in grams of CO₂ per megajoule of product. The table below shows the same CO₂ intensities as above, but in g CO₂/MJ.

CO ₂ Intensities from OLS regression	CO ₂ Intensities g CO ₂ / MJ	LCA compliant Positive CO ₂ intensities g CO ₂ / MJ
1 LPG	8.14	7.64
2 NAPHTHA	4.96	4.66
3 GASOLINE	6.33	5.94
4 KEROSENE	6.23	5.85
5 DIESEL	6.95	6.52
6 HO MARINE DSL	4.70	4.41
7 DMF RMF 0.5%S	0.27	1.67
8 HSFO	-4.70	0.45
9 OTHERS	1.64	1.54

The estimated CO₂ intensities, as well as their confidence intervals, confirm that diesel is more CO₂ intensive than kerosene and gasoline. These marginal CO₂ intensities are causal-based and, therefore, relevant for policy making such defining the fossil fuel comparators for GHG saving computations.

Oil refining is a complex interactive production system for which allocation-based average data are both meaningless and misleading. In this report, we reconcile, for the first time, the causal relevance with the accounting conventions such as additivity and nonnegativity. The computation principles are transparent and readily replicable by all stakeholders.

1. BACKGROUND

One of the key characteristics of refineries is its structure with process units organized in parallel and series leading to the co-production of a wide range of products. In other words, estimating the CO₂ emissions associated with production of individual oil products is challenging inasmuch as they are produced simultaneously through a combination of interrelated processes. Allocation methods, although providing for a mostly straightforward set of calculations, often fail to capture crucial interactions between processes and products and can lead to unrealistic or misleading results.

Utilizing the specific features of the Linear Programming (LP) techniques (tool widely used to model refineries) enables to produce a consistent set of CO₂ intensities for all refinery products (Tehrani Nejad, 2007). Concawe has applied this methodology (referred to as Generalized Marginal Approach) on its LP model (Concawe, report n° 1/17, “Estimating the marginal CO₂ intensities of EU refinery products”), so as to generate a full set of refinery product CO₂ intensities for a study case representing 2010 EU refining system. Results of this study are summarized in the following table.

Table 1 2010 EU refinery product CO₂ intensities (Concawe, 2017)

	Product demands	Product demand constraints	Reallocated process unit constraints	Total allocation to final products	Total allocation to final products
Products	Mt / a	t CO ₂ / t	t CO ₂ / t	t CO ₂ / t	g CO ₂ / MJ
Chemicals	54.6	1.46	-0.10	1.36	31.1
LPG	13.0	0.28	-0.04	0.24	5.2
Gasoline	126.3	0.27	-0.03	0.24	5.5
Kerosene	56.6	0.31	-0.05	0.26	6.1
Diesel Fuel	207.1	0.35	-0.04	0.31	7.2
Heating Oil	72.5	0.26	-0.06	0.20	4.7
Marine Gasoil	7.0	0.20	-0.08	0.13	2.9
Heavy Fuel Oil	88.6	-0.24	0.09	-0.15	-3.7
Bitumen	19.2	-0.45	0.05	-0.40	-10.1
Petroleum Coke	5.0	-0.89	0.01	-0.88	-25.0
Lubes & Wax	5.1	0.39	0.21	0.60	14.1
Sulphur	3.2	-	-0.02	-0.02	-1.3
Fuel & Losses	60.5	-	-	-	-

Main results of the Concawe publication have been integrated in JRC comparison of refining CO₂ allocation methods (JRC, 2017) and JEC Well-To-Wheels report v5 (JEC 2020, p23).

1.1 OBJECTIVE

The new method developed in this report is primarily motivated by the necessity to update Concawe's results and to define an easier way to assess the CO₂ intensities of refining products. It also demonstrates the relevance and necessity of a Linear Programming Model in the assessment of carbon intensity for products originated from co-production processes.

In order to comply with LCA standards and to further develop the potential of this innovative methodology, further developments have been carried out to generate only positive CO₂ intensities.

Being easier to understand and a higher enabler to reproduce results than the Concawe Generalized Marginal Approach, the new method is more likely to be accepted in the Life Cycle Analysis (LCA) community. The credibility and recognition will be increased with the submission for publication of an article in a scientific Journal.

1.2 LIMITATIONS OF THE GENERALIZED MARGINAL APPROACH

1.2.1 Identified limitations of Generalized Marginal Approach

Tehrani Nejad's (2007) theory was applied to Concawe LP model (Concawe, 2017). However, as Concawe model is a lot more complex than the example presented in the article, the implementation of the Generalized Marginal Approach required a specific post LP program, which had to be developed on purpose. Developing and maintaining such tool involves a high-level of expertise in LP modelling. This expertise is available within Concawe and its Member Companies, but outside the refining industry, for the other stakeholders involved in LCA, this is not generally the case. Consequently, the major challenge is to ensure that the methodology is correctly applied and understood by the potential users.

The General Marginal Approach being based on marginal results of a LP model solution, generated CO₂ intensities that can be negative. A typical example is heavy fuel oil. If the production of heavy fuel oil had to be reduced, it would be converted into more valuable products through CO₂ intensive process units. Consequently, in a marginal approach, heavy fuel oil is granted for saving this CO₂, hence given a negative CO₂ intensity. For the LCA specialists, this is invalid as they assume negative values only for CO₂ which is actually removed from the atmosphere.

In addition, the marginal values of a LP solution can be very sensitive to modelling assumptions. Furthermore, sensitivity analysis on input parameters (e.g. product demand, product pattern, refining system configuration) can be difficult to perform.

However, it has to be acknowledged that the General Marginal Approach developed by Concawe is additive (i.e. the total refining CO₂ emissions are described by the individual products CO₂ intensities), as requested by LCA standards, which is not always the case with marginally based methodologies (in this report, we are proposing a method to address this issue). For this type of model that predicts total emissions of a refinery, in which the products are produced simultaneously through a combination of interrelated processes, it is key to capture crucial interactions between processes and products. The link between total emissions and product related emission is demonstrated via the regression and the statistical robustness of the coefficients. The Design of Experiment (DOE) defines the space (incremental variation of the products) in which the solution is valid.

1.2.2 Main challenges identified by external stakeholders

The main challenges received when the Concawe 2017 study was published confirm the limitations listed in the previous section:

- Results come from a “black box” and require LP expert knowledge
- Difficulty to reproduce the results
- Negative CO₂ intensities values (though being “somehow” understood) are not compatible with LCA standard methodology, which aim at providing so-called “average” values and negative values meaning CO₂ extracted from the atmosphere.

Though some of these challenges are subject to discussion, it appears that clarification and improvement of the Generalized Marginal Approach are required to gain confidence of the potential users. The new methodology presented in this report addresses all the questions above.

1.3 LIFE CYCLE ASSESSMENT METHODOLOGY

Life Cycle Assessment (LCA) is the broadly accepted scientific approach to quantify potential environmental impacts of products during their life cycle. The LCA results can be reported for Well-to-Wheel (WTW) or cradle-to-grave which is composed of Well-to-Tank (WTT) and Tank-to-Wheel (TTW). The lifecycle of WTT includes the crude oil extraction, transportation, and petroleum refining up to the final fuel use in vehicles. TTW evaluates the final combustion in vehicle. This report addresses only carbon footprint of the oil refining production step which is the main responsible for controversial data in the literature¹.

There are two main approaches depending on the type of questions we want to answer: attributional and consequential CO₂ intensity.

- The attributional CO₂ intensity of a refined product accounts for all flows physically linked to its production within the given state of the examined refinery irrespective of economic or policy context. It tends to estimate what share of the global CO₂ emissions of the refinery belongs to that product according to some normative rules.
- The consequential CO₂ intensities are context-specific and capture impacts beyond direct physical relationships assessed by the attributional approach. They include the expected impacts of a change in the product system on other economic sectors.

Consequential WTT uses marginal data whereas attributional WTT requires average data resulting from normative allocation methods.

The LCA researchers usually recommend using the consequential approach for decision-making and the attributional approach for normative analyses where no decision is to be made based on the results. One argument in favor of this recommendation is that decision making must rely on causal mechanism. In an oil refinery, allocating the total emissions over the individual refined products in proportion to their mass or energy content suffer from the absence of any causal

¹ JRC publication, 2017 :

<https://www.sciencedirect.com/science/article/pii/S0306261917312102?via%3Dihub>

connections. This is mainly explained. Within refineries, the high degree of interactions among process units and final products is creating this causal connection.

In joint production industry, allocation based on causal relationships is possible if and only if the internally production function is marginally affected. A suitable analytical framework for performing marginal analysis is Linear Programming (LP). Within some adjustments, LP models can become a powerful tool for computing the CO₂ footprints of refined products. It has been demonstrated how LP models can depict causality between physical flows and assign accordingly the CO₂ emissions to refined products on a marginal basis. This fits well into the framework of a WTT methodology that is designed to quantify the GHG intensity of fossil fuels on a physical basis (representing real plant operation).

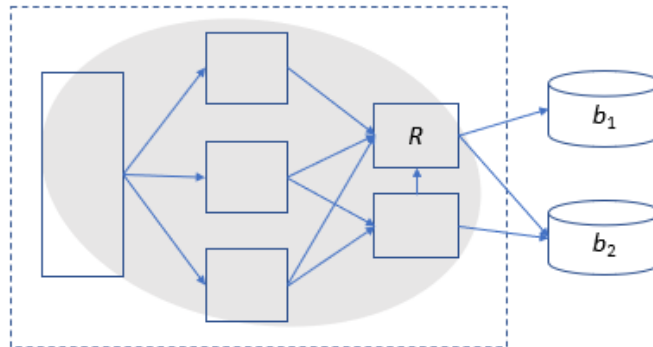
In general, short-run LP models do not respect the accounting convention that requires the sum of the allocated emissions to be equal to the total emissions of the refinery. This is a valid objection to LP as a support tool to contribute to the computation of fossil fuel comparators (in this report, we are proposing a method to address this issue as well). By extending the simplex mechanism, Tehrani Nejad (2007) proposed an additive CO₂ intensity pattern based on a generalized marginal analysis. This method was officially used by Concawe to estimate the CO₂ intensity of the European refinery products (Concawe, 2017). The results were integrated in the JEC WTW version 5 (JEC, 2020). DG Climate (2020) reported that its refining footprints were in line with Concawe model used in JEC for conventional crude chains. The JRC considered Concawe's results as reference values for decision-making (Moretti et al., 2017).

1.4 AVERAGE VERSUS MARGINAL CO₂ FOOTPRINTS

The concept of average CO₂ footprint is essentially meaningless in a multiproduct industry. This is a well-known issue among economists who have successfully discredited the use of average data for policy analysis for over 60 years ago. Before discussing the appropriate emissions measure, it will be of value to recall the following fundamentals.

First, in a multifunctional process, marginal and average data are both results of a change variation. Second, the change variation required to properly compute average data is generally inexistent for a joint production process. To make this clear, and without tedious mathematical notations, let $y = y(b_1, b_2)$ denote the total emissions function of a refinery that jointly produces b_1 and b_2 . This means, the production of b_1 (say gasoline) simultaneously and necessarily involves the production of b_2 (say diesel) in a causally coupled manner. This situation is simply presented in the figure below where the subsystem of process units contributing to the life cycle of joint products is highlighted in grey.

Figure 1 A refining joint production system



All process units and their respective CO₂ emissions are affected by the totality of products. Figure 1 shows that increasing the production of b_2 would also increase the production of b_1 even if there are no clients for the extra produced volumes of b_1 . The process unit R corresponds to the Reformer, the main gasoline process unit in a refinery, which produces 3 to 4% of hydrogen necessary to produce diesel (b_2). The feed of Reformer is connected to the process units that jointly produce for gasoline and diesel blends. These complex interactions lead to the non-separability of the total emissions even after subdividing the refinery into individual process units. To estimate what share of total emissions truly belongs to b_1 one should isolate the contribution of b_1 from the total emissions as,

$$S = [y(b_1, b_2) - y(0, b_2)] / b_1 \quad (1)$$

Where S is the average CO₂ footprint of b_1 , $y(b_1, b_2)$ the total emissions related to the production of b_1 and b_2 and $y(0, b_2)$ the emission related to the production of b_2 without producing b_1 .

The critical point to this change variation requirement is the impossibility to calculate the numerator. In effect, the stand-alone emissions for the first product, i.e., $y(0, b_2)$ where only b_2 is produced, does not exist neither in theory nor in practice. Since total emissions $y(b_1, b_2)$ are not traceable to individual products, any allocation measure which attempts to approximate the inexistent S , such as $y(b_1, b_2) / b_1$, is groundless by construction. The meaningless of this partitioning, which is systematically calculated in attributional approach, is less about the arbitrary selection of the allocation keys, e.g., mass, energy, economic value, etc. It is essentially due to the absence of any casual effects between total emissions and individual products. The greater the degree of non-traceable emissions, the more inappropriate is the use of such allocation rules as a guide for policy making. As argued by Baumol et al. (1962), “*The obvious stumbling block is that a multiproduct possesses no natural scalar quantity over which (emissions) may be averaged (...) without committing the sin of adding apples and oranges*”.

In a recent publication, Johnson and Vadenbo (2020) gathered the CO₂ intensity for refined products from multiple databases and publications. The following table summarizes the CO₂ intensity values that are computed based on attributional allocation rules.

Table 2. Carbon footprints of refined petroleum products, grams of CO₂ equivalent per MJ LHV

Source	Allocation key	Gasoline	Jet Fuel	Diesel	Fuel Oil
Ecoinvent	Energy	7.7	4.7	4.7	2.4
Sphera / Thinkstep	Energy	9.6	2.5	3.4	4.1
EU JRC ²	Added Value	6	7.9	10.3	-29.8
EU JRC	Energy	5.8	6.1	7.2	-4.3

The authors observed that the reported CO₂ intensities for the main products vary by a factor of three. This range of variation, within the same attributional framework, makes the results useless for policy decisions. Besides, half of the values concludes that gasoline is more CO₂ intensive than diesel. The other half contradicts. The reasonableness of the basis of allocation makes absolutely no difference except to the success of the advocates of the figures in deluding others, and perhaps themselves, about the defensibility of the results.

In decision science, relevance relies on mechanisms of causality. In an oil refinery, allocation based on causal relationships is possible, if and only if, the production function is marginally affected. In our previous example, the marginal contribution of b_1 to the total emissions is,

$$M = [y(\Delta b_1, b_2) - y(b_1, b_2)] / \Delta b_1 \quad (\text{eq 2})$$

Where M is the marginal emission linked to the production of the last ton of b_1 , Δb_1 the variation of the product b_1 (could also be written $b_1 + \epsilon$) and $y(\Delta b_1, b_2)$ the total emissions for different productions of b_1 .

Relation (2) is simply the difference between two LCAs. Within a given crude slate (not a single crude barrel), it is possible to vary the output of one individual product, within some ranges, without impacting the level of the other products. Hence, $y(\Delta b_1, b_2)$ and M are computable in a causal manner. If the production variation tends to be very small (ϵ), then the marginal CO₂ intensity would be equal to the partial derivative at the optimal point of operation. Relation (2) includes the CO₂ emissions of all the processes that are identified to truly contribute to the production of the last tons of the first product.

Two concluding remarks are in order. First, relations (1) and (2) have the same cradle-to-gate boundary, i.e., the refinery gate-to-gate process (excluding upstream and downstream emissions). Second, the change step variation used for the marginal intensity is very small: it doesn't correspond neither to new production plants nor to large scale-effects on the production system. Therefore, the marginal CO₂ data from relation (2), can be used as the best proxy for attributional studies.

² <https://www.sciencedirect.com/science/article/pii/S0306261917312102?via%3Dihub>

1.5 MAIN MODELING ASSUMPTIONS

This section presents the main assumptions that are most likely to affect the LP modelling results. More detailed information can be found in Appendix 1.

The refining system data for 2020 are not relevant due to the perturbations of the oil market linked to the Covid-19 pandemic. However, the introduction of the Marine fuel specification change (maximum Sulphur content down from 3.5% to 0.5%) applied from January 2020 onwards, is an important element to consider. Consequently, 2019 reference year data is selected for this study, but with 2020 Marine Fuel specifications implemented. This mix of assumptions is referred as to “2019 adjusted” reference year.

1.5.1 Crude Slate

The crude slate is derived from EU Commission data³ (EU crude supply data are available per country). First, EU crude slate (composed of around 60 crude types) from EU Commission data is matched with crude assays from Concawe database, so as to estimate the main properties of the average EU crude slate (API gravity, Sulphur content, main products yield). Then a combination of the LP model crude assays (6 crudes and 2 residues) is determined to match the average EU crude slate properties. Final LP model crude slate for reference year 2019 can be found in Appendix 1.

1.5.2 Other Refinery feedstocks

Refinery feedstocks other than crude oil can be imported in the LP model, as in actual refining operation. Eurostat also provides global data on these feedstocks, which are reworked to match LP model environment.

Eurostat ‘Refinery feedstocks’ category amount is considered to be Heating Oil type and/or Russian M100 type (ratio being a result from the LP model optimization). This amount is set as a maximum availability for the LP model, hence giving some flexibility on the refinery supply. This Refinery feedstocks category originally contains returns from Petrochemical plants, but for simplification the amount of Petrochemical returns is deduced from the Chemical Naphtha demand.

Biofuels and oxygenates are considered for Gasoline, but not for Diesel and Jet. So, for Diesel and Jet, demand is related to fossil demand only, the biofuels having very little impact on the optimization, and added directly to the finished products. For Gasoline, several types of oxygenates are considered: Ethanol and ETBE are used to represent bio oxygenates and MTBE to represent non-bio oxygenates. These oxygenates imports are driven by the bio energy and oxygenate content of the gasoline grades (refer to 1.5.4 section below) and for ethanol, also by ETBE/TAEE production within refineries.

Natural gas amount imported by the EU refineries (as energy) is also available in Eurostat and is implemented as a maximum availability in the LP model.

³ https://energy.ec.europa.eu/data-and-analysis/eu-crude-oil-imports-and-supply-cost_en

1.5.3 Refining finished products demand

Historical products demand is extracted from Eurostat database. More specifically for the LP model, EU refining production is considered to be equal to the products demand because it corresponds to the actual constraint supported by the refining system. On one hand this means that finished products imports are not accounted because they have no impact on the refining system. On the other hand, finished products net exports are accounted because they are actually produced by the EU refining system.

As mentioned at the top of this chapter, the definition of 2019 adjusted demand is based on 2019 actual demand data and corrected to consider the 2020 IMO Marine Fuel specification change (from 3.5% to 0.5 wt.% Sulphur content). Final LP model demand is available in Appendix 1.

1.5.4 Finished product specifications

Standard finished products specifications are used for the main products. However, for some specific grades, some assumptions and intermediate calculations are required.

Oxygenates and bio components in Gasoline

Several oxygenates are available in Concawe LP model: MTBE, ETBE, ethanol and TAAE. MTBE is used to represent the non-bio oxygenates, ethanol is considered 100% bio, ETBE 37% bio (according to RED directive) and TAAE bio content is calculated based on the ethanol content.

According to Fuels Quality Monitoring report (EU Commission, 2021), 2019 bio ethanol equivalent content of the gasoline consumed in Europe is around 6.1 vol%. Bioenergy content of the gasoline is derived from this figure and fixed to 4.1 % in the LP model for the year 2019. From Concawe market survey (Concawe, 2017), an ETBE / ethanol ratio in the gasoline pool is assumed (approximately 30/70 vol%), allowing to estimate the ethanol content in the pool. This ethanol content (5.4 wt.% or 5.1 vol%) is fixed in the LP model and the ETBE is floating, but indirectly fixed by the bio energy content of the gasoline pool.

From Concawe market survey (Concawe, 2017), 2015 average EU gasoline oxygen content is estimated to be around 2.4 wt.%. For this study based on 2019 data, it is assumed to be around 2.6 wt.%. Knowing the oxygen brought by bio ethanol and bio ETBE, MTBE content of the gasoline pool is estimated to be around 2 wt.% (2 vol%).

These average calculations are applied to both 95 and 98 gasoline grades. Oxygenates are not allowed in export gasoline grades (gasoline specification for exports to US or Africa do not contain oxygenates).

Bio components in Road Diesel and Jet

Bio-components in Road Diesel and Jet are not available in Concawe LP model. Consequently, any demand or production data related to these products correspond to the fossil part.

Marine Fuels

As mentioned above, 2019 is taken as reference year for the demand, but the IMO Marine Fuel specification change occurred in 2020. Therefore, the following

assumption has been taken: Eurostat “International maritime bunkers” demand is assumed to be at 0.5 wt.% Sulphur content, 50%/50% of the demand to be supplied by Distillate/Residual Marine Fuels (Concawe report 21/20⁴, “Producing low Sulphur marine fuels in Europe - 2020-2025 vision”).

Diesel Marine fuel is a separate category in Eurostat, which is assumed to have a Sulphur content of 0.1 wt.%.

Heavy fuels

Eurostat high/low Sulphur content fuel oils are assumed to have respectively 3.5/1 wt.% of Sulphur content.

A summary of Finished products specifications is presented in Appendix 1.

1.5.5 Process unit capacities

Concawe owns a detailed database on process units’ types and capacities of European refineries (including expansion projects and closures). These capacities are aggregated to determine the LP model process unit capacities. Main capacities considered in the LP model are summarized in Appendix 1.

1.5.6 Pricing

Feed and product pricing are mainly based on Wood Mackenzie data (Concawe subscription). For some specific quotations not available (natural gas for example), public sources are used. 2019 average yearly prices are used.

CO₂ pricing has an impact on the LP optimization, as it drives the effort of the optimization to produce the demand with minimum CO₂ emitted. For this study, 49 USD/t CO₂ (44 EUR/t) has been used, being the CO₂ price used in the ‘MIX’ scenario of EU Commission 2030 Climate Target Plan Impact Assessment (European Union, 2020). A sensitivity test has been performed with 100 USD/t CO₂.

1.6 LP MODEL

The purpose of an oil refinery is to turn crude oil into marketable products in the most efficient and economical way. A particular refinery generally serves a particular market which sets the quality of the products to be supplied and to an extent the amount of each grade. Depending on the geographical location of the refinery, there can also be opportunities to export to other markets. The refinery has access to certain crude oils and other feedstocks, the range of which is a function of its location and the way it is supplied (e.g. ships or pipelines). Finally, the refinery features a certain combination of process units (generally referred to as its configuration).

Refinery operation is thus characterized by multiple real constraints arising from feedstock supply, product demand (quantity and quality) and process unit limitations. Yet there are many ways of operating within these constraints and refiners have always strived to optimize their operation in order to maximize profit or minimize costs to supply a given market demand within a given set of product prices and input costs. The tool used to that end by refiners worldwide is known as Linear Programming (LP), a mathematical technique which, given a quantity to be

⁴ https://www.concawe.eu/wp-content/uploads/Rpt_20-21.pdf

optimized, aims at identifying the optimum solution amongst the myriad of possible solutions to a complex problem.

In an LP model the refinery constraints are represented by a system of linear equations linking the different variables. Because there are more degrees of freedom (or variables) than there are constraints, the system has an infinite number of possible solutions. Provided that appropriate cost factors are defined as model inputs (i.e. cost of feedstocks, energy, additional plant capacity, price of products etc.), a so-called “objective function” can be derived, describing the quantity to be optimized (maximum profit or minimum cost). The LP solver then provides a pathway towards the optimum solution.

For a given set of desired products, the LP solution tells the refiner how much of each available feedstock should be processed, the level at which each refinery process units will be utilized and, more generally, which amongst all the constraints will actually be binding. Crucially it also provides information on the impact on the objective function of a marginal change in each of the binding constraints (the so-called “marginal values”). This last property of an LP solution was used to assess the CO₂ intensities of refining products in the Concawe 2017 study.

1.6.1 Concawe LP model main features

Since the mid-90s Concawe has operated a refinery LP model representing the combination of all refineries operating in the EU. This model was originally developed to estimate the cost to EU refiners of EU legislation (mostly affecting product quality) and of expected changes in EU market demands.

Model structure

The model features a full library of refinery process units represented by a number of operating modes including feedstock type, product yield structure and all relevant quality parameters. From this a refinery can be modelled with any combination of process units.

A range of crude oils is available, representing the diversity of grades available to EU refiners.

A blending module allows finished products to be prepared according to the required quality specifications from selected intermediate streams.

In the Concawe master LP model, the EU is divided into 9 regions, each represented by a single refinery having the aggregated capacity, crude intake, process configuration and product demand of all physical refineries in that region. Due to the specific requirements of the methodology (marginal values extraction for additivity) developed to generate product CO₂ intensities in the Concawe 2017 study, a **single-region model** has been derived from the original 9-region model. This model then consists of a single refinery representing the aggregation of all European refineries (capacities, crude intakes, process configurations and product demands). Countries accounted for in the LP model, are listed in Appendix 1.

The single region LP model used for this study has approximately 2500 rows (equations) and 10 000 columns (variables), which is a rather large model for the industry.

CO₂ emissions modelling

In response to the CO₂ emissions challenge, the model was adapted in the early years of the last decade so that it could estimate the CO₂ emissions associated with a particular operating case. This primarily requires CO₂ emission factors for combustion of refinery fuels (t of CO₂ emitted per GJ of fuel burnt), combined with fixed unit-specific energy consumption factors (GJ per t unit feed). It must also consider structural emissions arising from specific chemical reactions (notably the production of hydrogen by reforming or partial oxidation of hydrocarbons), expressed in ton of CO₂ per ton of each feedstock type processed.

Because a refinery generally uses its own feeds or products for a proportion of its fuel needs the carbon contained in the input to the refinery is apportioned between products and fuels. In order to avoid any spurious carbon gain or loss as a result, it is therefore essential that the model be strictly carbon-balanced (i.e. that the amount of carbon entering the refinery in the form of feedstocks and possibly fuels equals the amount that leaves the refinery in the form of CO₂ and products). In order to achieve this, the model is actually carbon, hydrogen, nitrogen, oxygen and Sulphur balanced across each process unit, from the crude oil to the finished products.

In order for the model to optimize CO₂ emissions, these must have an impact on the objective function. In other words, CO₂ emissions must be assigned a monetary value. Although the actual number used is not crucial to the outcome, it must be sufficiently high to ensure small changes have a discernible impact on the objective function. As mentioned in the 1.5.6. Pricing section, 49 USD/t CO₂ has been used ('MIX' scenario of EU Commission 2030 Climate Target Plan Impact Assessment).

LP model main variables for optimization

Crude oil throughput: total amount processed in the EU refining system is free, but the crude oil ratios are fixed to match average EU feedstock quality (refer to section 1.5.1. Crude Slate).

Oxygenates (ETBE, MTBE, and Ethanol): total amount of imported oxygenates is open, but is indirectly limited by gasoline pool specifications (refer to section 1.5.2. Other Refinery feedstocks).

Other intermediate product imports as Kerosene or Gasoil: as for crude oils, total amount processed is free, but ratios compared to crude oils are fixed to match actual EU imports level (refer to section 1.5.2. Other Refinery feedstocks).

Refining operations: routings of intermediate streams, severities and utilizations of process units are the major optimization variables of the LP model. They represent the flexibility of the refineries to respond to market demands and constraints.

1.6.2 Calibration

This type of model, where individual refinery data is aggregated into one single large refinery model, is subject to over-optimization (under-constrained system), because all process unit capacities are available simultaneously, significantly increasing the degrees of freedom of the refining system. To mitigate this risk of over-optimization, a calibration run is performed at the beginning of each Concawe study where LP is involved.

A reference year is selected, 2019 for this study, for which all the input data are publicly available (Eurostat). 2019 product demand is then corrected to consider

IMO 2020 (Marine Fuel new Sulphur specification at 0.5 wt.%). The year 2020 has not been chosen, as deemed not representative because of the atypical demand and refinery operation due to the Covid-19 pandemic.

The main LP model inputs required for the calibration include the European average crude oils slate, the other imported products, the finished products demand (and qualities) and the exported products amounts. In addition, the energy efficiency of the process units is tuned to match the overall CO₂ emissions of the refining system as measured for the reference year.

A calibration run is performed with the LP model to ensure that the optimization can meet the demand pattern with similar amounts and types of refinery feedstocks as observed in real operation. In addition, it is essential that the optimization faces the same constraints as the actual refineries to achieve the products demand. For this purpose, marginal values associated to demand constraints have been tuned to reflect the actual market. For example, the European market is a net importer of middle distillates because the actual refining system has issues to meet the demand, this should be reflected in the optimization results (i.e. marginal values associated to middle distillates demand should be higher than the products prices). Tuning the marginal values is performed through tuning of the process units' capacities. More constrained/relaxed specific process units capacities result in more constrained/relaxed specific finished products production, hence impacting the marginal values of these specific finished products demand in both direction (increase or reduction).

This artificial way of tuning process unit's availability allows to reduce the over-optimization inherent to aggregated models. This can be seen as a conservative approach, as it may imply that the refining assets would not adapt to the future constraints. Decades of experience in regional LP modelling showed us that it remains the best way to proceed to get results as realistic as possible. Refer to Appendix 1 for further details on the calibrated capacities.

2. NEW METHOD

2.1 BENEFITS

The new method described below overcomes the drawbacks listed in section 1.2. Here are the main advantages of this approach compared to the original methodology (Concawe, 2017).

Implementation

The Generalized Marginal Approach (Concawe 2017) requires a LP model specifically designed for this method, so as to be able to generate and extract numerous marginal coefficients from the LP solution. The additivity is ensured through a rigorous, but complex reallocation process, involving manipulation of marginal rates of technical substitution (Concawe, 2017).

The new method requires a refining model (Linear Programming is advised but not mandatory) and the capability to generate a consistent Design Of Experiment (DOE) for the Demand data (model inputs). This DOE and the ability of the LP model to react accurately to demand changes will then ensure robustness of the regression applied on the results. The regression allowing to compute the CO₂ intensities requires a statistical software (Microsoft® Excel® can be sufficient for basic results), but is independent from the refining model used.

Confidence

The new method can be understood by a larger audience than the Generalized Marginal Approach. It requires only generic knowledge in refinery modelling and statistics, whereas the original method requires to understand expert level LP optimization notions. The generic concepts of the new method help gaining confidence of the audience.

Reproducibility

In the Generalized Marginal Approach, extraction of marginal coefficients from the LP solution requires complex post-processing computation. This makes it difficult to reproduce the published results, even if a LP model is available for the experiment. With the new method, if the model results are provided, it is possible to exactly reproduce the calculations of finished products CO₂ intensities.

Validation of the Generalized Marginal Approach

The new method shows similar results than the Generalized Marginal Approach in terms of CO₂ intensity values and ranking of the finished products CO₂ intensities. This is expected as it follows the same “marginal causal-related” approach, which is a key element when considering co-production processes (e.g. any change in a product quantity has an impact on the whole refining system).

Like the Generalized Marginal Approach, the new method is also additive, meaning that all the emissions from the refinery are integrated in the CO₂ intensities of the products.

In addition, the new method provides confidence intervals directly from regression results, whereas in the original method additional cases are required.

2.2 BASIC PRINCIPLE

From the LP theory and the settings of Concawe model, it has been shown that the total CO₂ emissions can be expressed as a linear function of finished products amount and saturated process unit capacity constraints (Concawe 2017):

$$Total\ CO_2 = \sum_i (CI_i \times Product_i) + \sum_j (CI_j \times Process\ Unit_j)$$

Where CI_i is the CO₂ intensity of Product i (equal to M_i in the theoretical equation 2 detailed in paragraph 1.4) and CI_j the CO₂ intensity associated to saturated capacity constraint of Process Unit j .

The non-product CO₂ emissions can be regarded as residuals of the above equation (Tehrani Nejad, 2007):

$$Total\ CO_2 = \sum_i (CI_i \times Prod_i) + Residual(\varepsilon)$$

These residuals follow the requirements for an Ordinary Least Square (OLS) statistical regression (as non-systematic pattern, orthogonality versus explanatory variables, no autocorrelative pattern). The OLS regression resolution leads to Residual(ε) minimization and consequently resulting CO₂ intensities (regression coefficients) contain the Process unit capacity constraints impact. The Residual(ε) minimization replaces the complex reallocation procedure of CO₂ associated to process unit capacity constraints used in the Generalized Marginal Approach.

2.3 IMPLEMENTATION

2.3.1 Data driven approach

The data used for the statistical regression are generated by the LP model (meta-modelling). A high number of cases (500 runs done with the LP Model) is generated by varying the finished products demand. Each case provides the optimal quantity of CO₂ that needs to be emitted to satisfy a given demand of finished products. The cases results (CO₂ emitted corresponding to a specific products demand) can then be used as inputs for a multi-linear regression described in section 2.2. In addition, the high number of generated cases allow to explore a large range of solutions and consequently to depict the underlying technical pattern of the LP model.

Concawe's carbon balanced LP model (refer to section 1.6.1) is a clear motivation for such a data-driven approach, as it ensures that product demand impact on refining CO₂ emissions is accurately captured in each of the cases.

It could be argued that historical time series could be used instead of generated cases. However, such an approach is not recommended, as historical time series contain major structural changes that lead to confusion in the definition of the products and the refining system itself. For example, gasoline specifications have changed over time, hence gasoline in 1990 cannot be considered to be the same product as of today. Similarly refining system has evolved by step changes (shutdown of units or entire plants, addition of new process units and/or revamping existing process units). These multiple structural changes require adopting adequate statistical technics in regression model that may jeopardize the additivity of the method. In addition, a major advantage of the generated data is that they

are nearly orthogonal by design, thereby overcoming the inevitable multicollinearity issue caused by the joint production times series data.

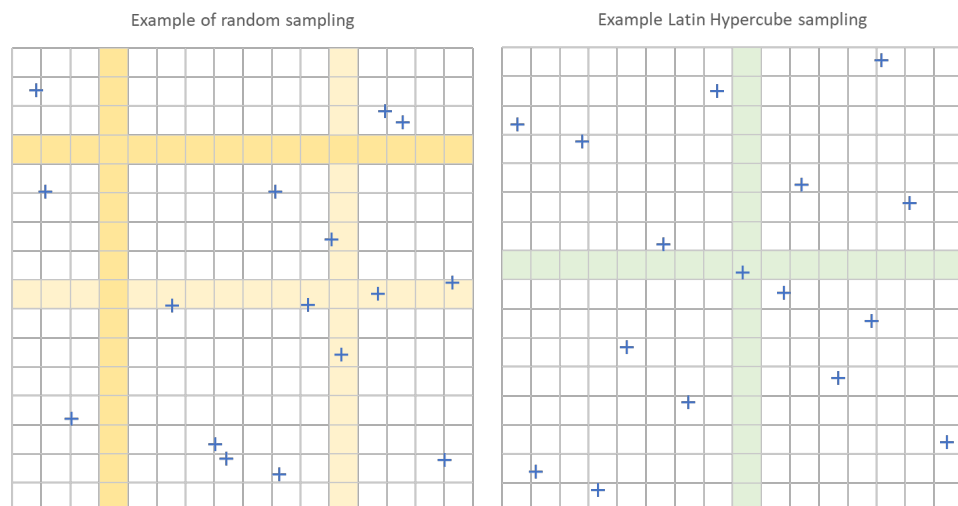
2.3.2 Design Of Experiment (DOE)

In order to generate the high number of cases mentioned in the previous section, a design of experiment (DOE) is prepared (definition in Appendix 4). For this objective, 500 cases are generated using the Latin Hypercube Sampling (LHS) method described in the next section. Cases are centered around 2019 adjusted reference year products demand, within a $\pm 5\%$ range. This arbitrary range variation is a compromise between expected validity of the results and ensuring that the refining system response remains representative of the given period.

2.3.3 Latin Hypercube Sampling (LHS)

The design of experiment for the production demand is a crucial step. Among the existing literature, it has been selected to use the well-known Latin Hypercube Sampling (LHS). Its basic idea is to cover evenly and consistently the range of the production distribution (more details in Appendix 4). The figure below illustrates the difference of random sampling versus Latin Hypercube sampling for a 2-dimensional problem. By definition, in LHS there is 1 (and only 1) random point per row or column, allowing a more distributed sampling than with a random sampling, which can have empty rows and/or columns and/or multiple random points per row and/or column.

Figure 2 Latin Hypercube Sampling (LHS) vs. random sampling



LHS algorithm is thus used to generate different combinations of demand levels, within the $\pm 5\%$ range around 2019 adjusted reference year, as stated in the previous section. For each case generated, the sum of the finished products demand is constant (equal to 2019 adjusted reference year value), so as to ensure that only variation in the products pattern is evaluated. It has been decided to generate 500 cases in order to assess the robustness of the results.

The next step is then to solve each of the cases with the LP model in order to get the optimal refining CO_2 emissions corresponding to each case.

2.3.4 Ordinary Least Squares regression (OLS)

Ordinary Least Squares (OLS) is a type of linear least squares method for estimating the unknown parameters in a linear regression model (more details in Appendix 4). This method is used here to determine the following relation between the finished products demand and the related refining CO₂ emissions:

$$Total\ CO_2 = \sum_i (CI_i \times Prod_i) + Residual(\varepsilon)$$

Which can be expressed in more generic terms as:

$$Y = Z\hat{\beta}^{OLS} + \hat{u}$$

Where Y is the vector of CO₂ emissions, Z is the product matrix defined by the LHS design of experiment, $\hat{\beta}^{OLS}$ is the vector of unknown CO₂ intensities to be estimated by the Ordinary Least Squares method (OLS), and \hat{u} is the vector of error term.

An additional constraint is added to the OLS regression: residual has to be zero for the 2019 reference year demand (required to satisfy the additivity of the CO₂ allocation). The method is therefor called “Restricted Ordinary Least Square” regression.

This type of regression can be implemented in generic tools as Microsoft® Excel® for example, or with more dedicated statistical tools. It has to be noted that several statistical tests are required to assess the validity of the regression: linearity of the relation, normality and homoscedasticity of \hat{u} . They are developed in the numerical application presented in the next section.

3. NUMERICAL APPLICATION

This section describes the implementation of the new method to determine the CO₂ intensities of EU refining products for the 2019 adjusted reference year.

3.1 STEP 1 - DOE USING LHS METHOD

A DOE of 500 cases is generated using LHS sampling method. The table below presents an extract of the cases generated. The Base row shows the 2019 adjusted reference year production.

It can be observed that the 17 finished products of the LP model have been distributed to 9 main categories. Each category contains similar finished products in terms of refining components, except category OTHERS, which contains various specialty products (Bitumen, Lubes, Waxes, Coke and Sulphur). The modelling of these products would require specific development (due of the low product demand and specific production routes should be isolated) to appear individually in the DOE. Composition of categories are presented in Appendix 2.

Table 3 DOE - Product demand 2020, Million tons per annum

	1 LPG	2 NAPHTHA	3 GASOLINE	4 KEROSENE	5 DIESEL	6 HO MARINE DSL	7 DMF RMF 0.5%S	8 HSFO	9 OTHERS
Base	24.3	37.4	127.5	51.9	174.5	80.6	35.6	30.3	37.4
1	23.9	36.6	128.6	51.4	177.1	79.1	35.6	29.6	37.7
2	24.9	37.2	130.4	52.8	170.3	79.8	36.0	30.4	37.5
3	23.9	37.3	129.6	51.4	172.8	81.6	35.7	29.9	37.5
4	23.9	37.0	129.0	51.8	171.4	81.7	36.0	30.6	37.9
5	24.5	37.2	127.9	52.3	170.6	83.3	36.7	29.0	38.1
...
496	24.4	37.2	126.1	50.8	172.2	83.6	37.0	30.7	37.5
497	24.7	36.4	128.0	51.8	174.0	79.6	37.5	29.9	37.6
498	23.9	37.0	127.4	53.0	173.8	80.9	36.7	29.7	37.4
499	24.0	36.6	128.2	52.3	174.0	81.3	35.4	30.1	37.5
500	24.1	36.4	125.7	52.8	176.6	81.1	35.3	30.5	37.2

Note that refinery fuel (auto-consumption) and losses are not represented in these mass balances as they are not products (outputs) of the refineries.

3.2 STEP 2 - DOE RESULTS - REFINING CO₂ EMISSIONS

The 500 cases generated during step 1 are run sequentially in the calibrated Concawe LP model. The corresponding total refining CO₂ emissions are extracted for each of the cases. The table below shows an extract of the results.

Table 4 DOE - Product demand 2019 and corresponding refining CO₂ emissions, Million tons per annum

	1 LPG	2 NAPHTHA	3 GASOLINE	4 KEROSENE	5 DIESEL	6 HO MARINE DSL	7 DMF RMF 0.5%S	8 HSFO	9 OTHERS	Total refining CO ₂ emissions (LP model result)
Base	24.3	37.4	127.5	51.9	174.5	80.6	35.6	30.3	37.4	131.5
1	23.9	36.6	128.6	51.4	177.1	79.1	35.6	29.6	37.7	132.1
2	24.9	37.2	130.4	52.8	170.3	79.8	36.0	30.4	37.5	131.3
3	23.9	37.3	129.6	51.4	172.8	81.6	35.7	29.9	37.5	131.5
4	23.9	37.0	129.0	51.8	171.4	81.7	36.0	30.6	37.9	130.9
5	24.5	37.2	127.9	52.3	170.6	83.3	36.7	29.0	38.1	131.6
...
496	24.4	37.2	126.1	50.8	172.2	83.6	37.0	30.7	37.5	130.8
497	24.7	36.4	128.0	51.8	174.0	79.6	37.5	29.9	37.6	131.3
498	23.9	37.0	127.4	53.0	173.8	80.9	36.7	29.7	37.4	131.5
499	24.0	36.6	128.2	52.3	174.0	81.3	35.4	30.1	37.5	131.6
500	24.1	36.4	125.7	52.8	176.6	81.1	35.3	30.5	37.2	131.7

3.3 STEP 3 - RESTRICTED OLS REGRESSION - CO₂ INTENSITIES

3.3.1 CO₂ Intensity results

An OLS regression is applied to the data generated at the previous step. As a result, CO₂ emissions are expressed as a linear function of the product demand. The regression coefficients are then the CO₂ intensities of the products.

$$Total\ CO_2 = \sum_i (CI_i \times Prod_i) + Residual(\epsilon)$$

With the additional constraint that Residual term equals zero for the 2019 adjusted reference year, the following CO₂ intensities (Restricted OLS regression coefficients) are obtained.

Table 5 CO₂ intensities of refining finished products, ton of CO₂ per ton of product

CO ₂ Intensities	t CO ₂ / t
1 LPG	0.3731
2 NAPHTHA	0.2218
3 GASOLINE	0.2687
4 KEROSENE	0.2718
5 DIESEL	0.3012
6 HO MARINE DSL	0.2026
7 DMF RMF 0.5%S	0.0116
8 HSFO	-0.1905
9 OTHERS	0.0602

This table can be read as:

$$Total\ CO_2 = 0.3731 \times Prod_{LPG} + 0.2218 \times Prod_{Naphtha} + \dots + 0.0602 \times Prod_{Others}$$

From a statistical perspective, the results of this Restricted OLS regression are very robust and stable as proved by the detailed statistical results available in section 3.4. and in Appendix 3.

3.3.2 Industrial interpretation of the results

From an industrial perspective, and focusing first on the main transport fuels (gasoline, Kerosene and Diesel), then ranking of the CI's is explained as follows.

Gasoline: the refining system having excess of gasoline (export parity product), the refining system is not constrained by gasoline production. The CI confirms this situation with a lower figure than Diesel and Kerosene.

Diesel and Kerosene are on import parity and are the main driving force for the refining system (production maximized and on purpose). This constraint is reflected by high CI's with Kerosene lower than Diesel as it requires less processing to meet the required specification.

LPG: the smallest production (24.3 Mt/a), but the highest CI. It is caused by a high market demand relative to the production capability of the refining system. The model having less options than for other products to meet the demand, the constraints are stronger and logically reflected in the CI.

Naphtha is similar to gasoline as export parity product, its CI is lower than gasoline as it is a straight run product requiring minimum processing to meet the final specification.

The Marine fuel 0.50%S (DMF RMF 0.5%S) is positive as it requires a significant amount of fluxant and components that are desulphurised and produced on purpose. On the other hand, the HSFO is negative because its components require a minimum of processing (no cracking and minimum quality improvement). In practice, it means that the more a refiner is producing HSFO, the less it operates conversion and upgrading units, leading to a net reduction of total refinery CO₂ emissions.

As mentioned in section 3.1, Others category contains various products and the average CI of this category does not necessarily reflect individual CI of its components (for example, it is most likely that Sulphur CI is a lot lower than Lubes CI, because Lubes production require more CO₂ intensive processing than Sulphur production). It can be mentioned that for Bitumen, the trade Association published their assessment through a Life Cycle Inventory 5 (3rd edition published in 2020).

It can easily be checked that the additivity of CO₂ emissions is preserved: the summation of each product demand (Mt / a) times its CO₂ intensity (t CO₂ / t) gives 131.5 Mt CO₂ / a, which is the total refining CO₂ emissions of the Base case (refer to the Table 4 above).

Table 6 CO₂ balance verification

CO ₂ balance	Base production Mt / a	CO ₂ Intensities t CO ₂ / t	CO ₂ Intensities x Base production Mt CO ₂ / a
1 LPG	24.3	0.3731	9.1
2 NAPHTHA	37.4	0.2218	8.3
3 GASOLINE	127.5	0.2687	34.3
4 KEROSENE	51.9	0.2718	14.1
5 DIESEL	174.5	0.3012	52.6
6 HO MARINE DSL	80.6	0.2026	16.3
7 DMF RMF 0.5%S	35.6	0.0116	0.4
8 HSFO	30.3	-0.1905	-5.8
9 OTHERS	37.4	0.0602	2.3
TOTAL CO ₂ calculated from OLS regression, Mt / a			131.5

3.4 VERIFICATION OF REGRESSION ASSUMPTIONS

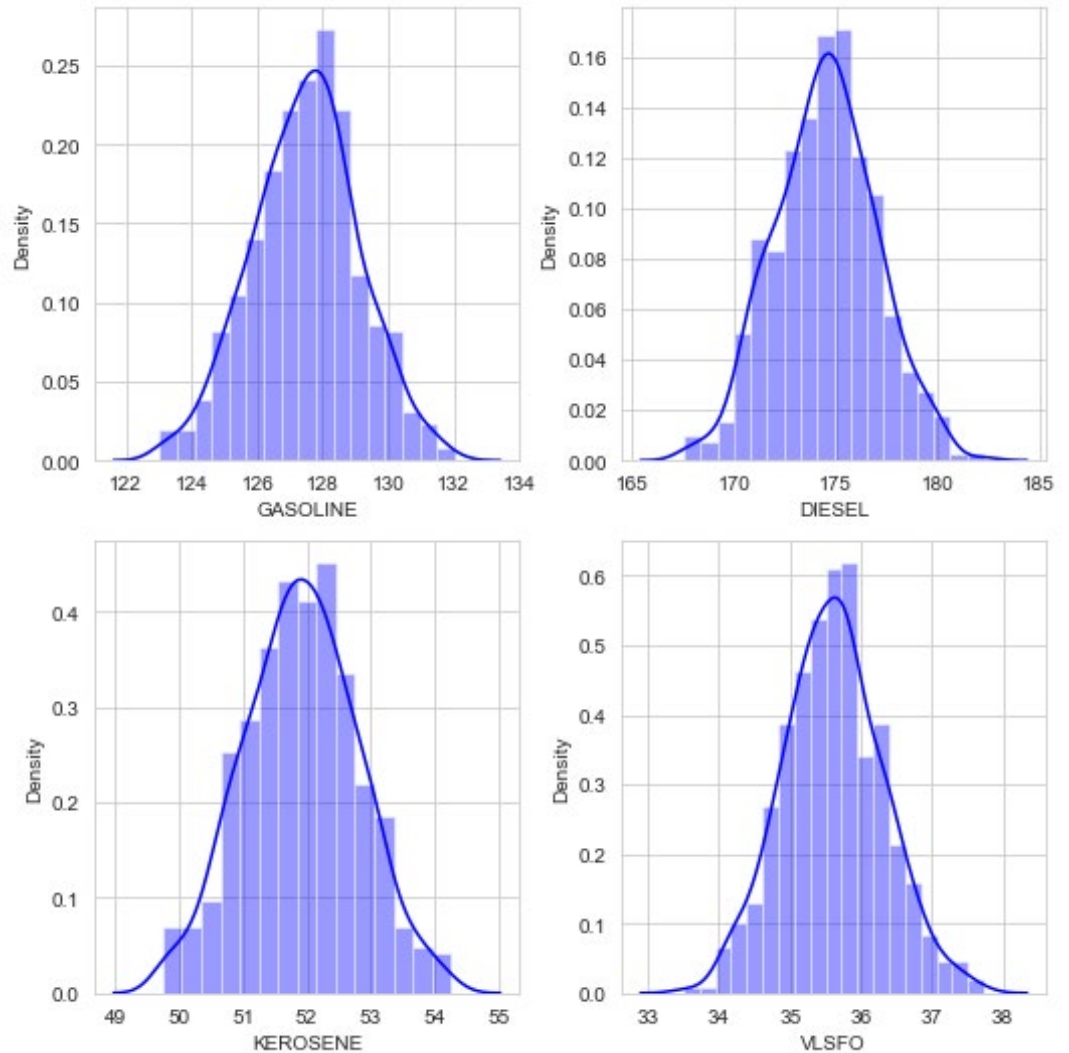
Standard tests have been performed to assess the robustness of the data and the validity of the results. They confirm the ability of the LP model to generate results representative of the refining system and the reliability of the CO₂ intensities calculated from the Restricted OLS regression.

3.4.1 DOE cases

The kernel density curves (definition in Appendix 4) in the figure below show that the 500 cases generated are reasonably normally distributed around the reference year levels.

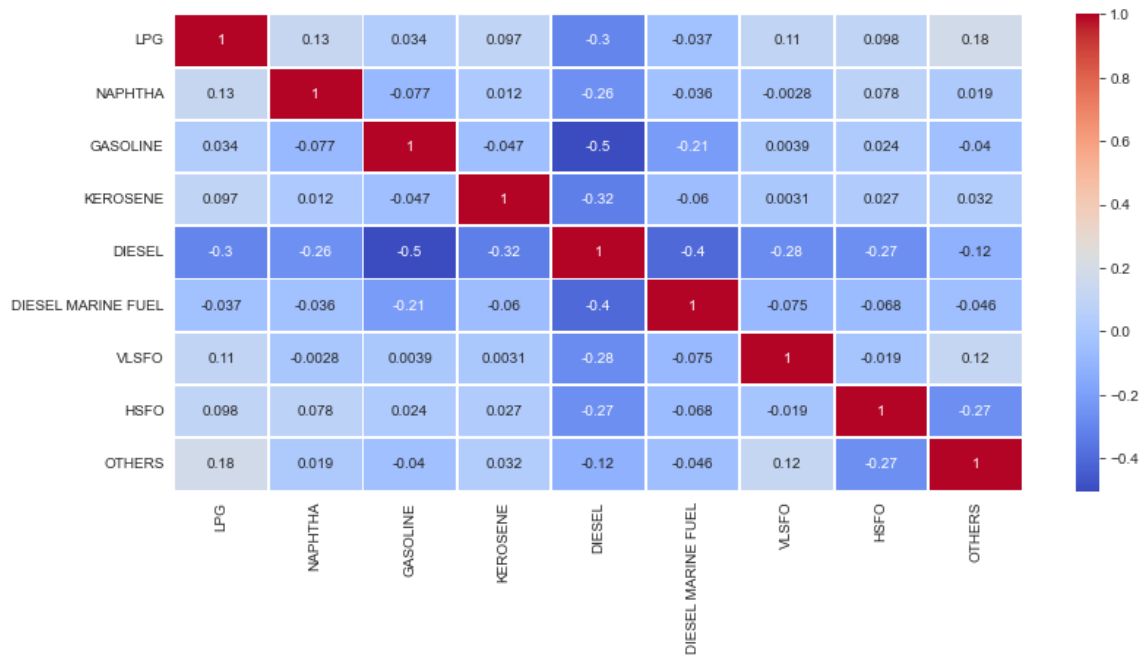
⁵ Bitumen Life Cycle Inventory report:
https://www.eurobitume.eu/fileadmin/Feature/LCI/EUB2975.001_LCI_Update_2020_01_LR_pages.pdf

Figure 3. Kernel density curves of the DOE cases for the main products



The following figure corresponds to the Pearson correlation coefficients (definition in Appendix 4) ranging between -1 and +1. Following the commonly accepted rule of thumb, the Pearson coefficients are all lower than |0,7| indicating the absence of linear correlation between any of the two variables.

Figure 4 Pearson correlation matrix of the DOE cases



3.4.2 Regression results

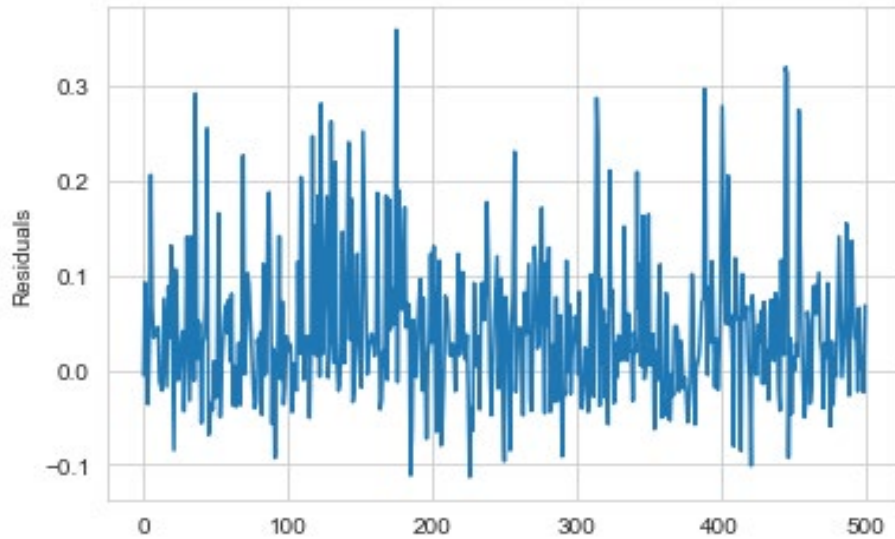
Several statistical tests on the Restricted OLS regression have been evaluated (definitions in Appendix 4). The following points can be highlighted:

All p-values are lower than 0.05, meaning that all the estimated CO₂ intensities are statistically significant at 95% (see Appendix 3).

The coefficients standard errors are very small. This brings a very high confidence on the ranking of the CO₂ intensities (see Appendix 3).

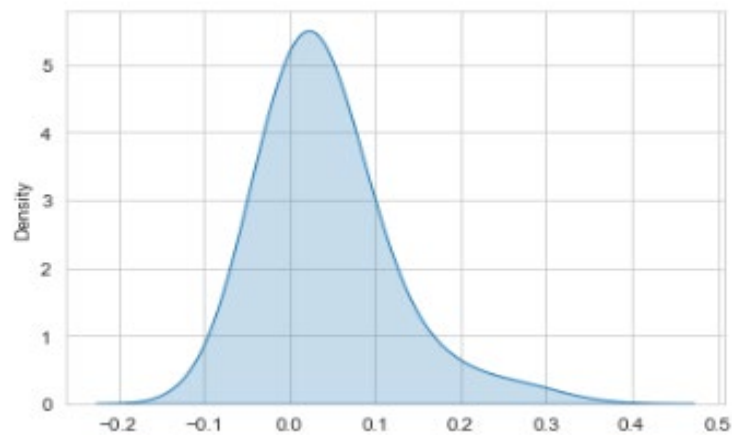
The adjusted R-squared (0.99) and the p-value of the F-statistic (0.0) confirm that the joint effect of all the variables together is fully significant. The Durbin-Watson value (1.56, value below 2.0) also indicates the absence of any autocorrelation between residuals as it was theoretically expected. The figure below confirms this important feature, as there is no systematic pattern of the residuals, there is no autocorrelation. It can also be noted that the residuals values are very small, between -0.1 and +0.35 MTPY of CO₂, below 0.3% of the reference year CO₂ emissions.

Figure 5 Behavior of the residuals of the Restricted OLS model



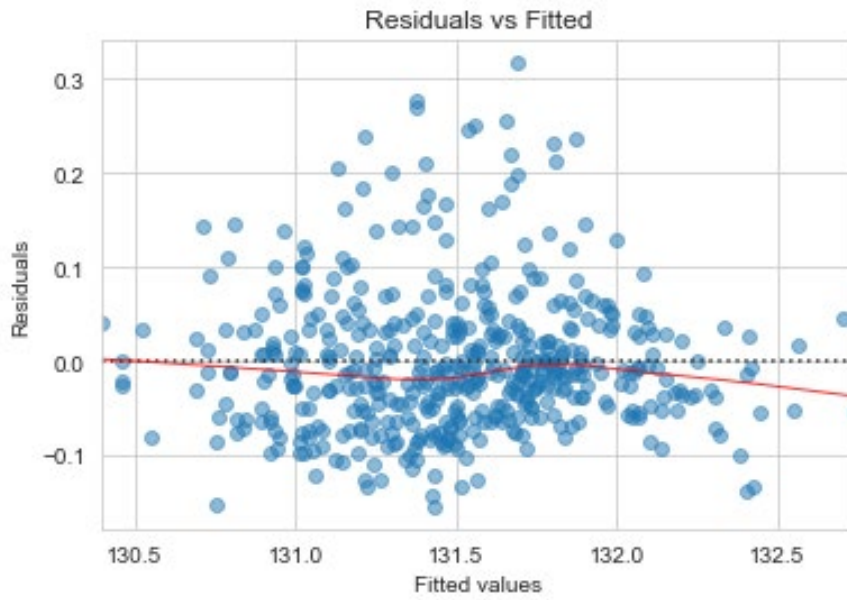
The figure below illustrates that the residuals of the restricted regression are very close to a normal distribution. A violation of this hypothesis can inflate the confidence intervals.

Figure 6 Residuals distribution



The figure below doesn't show any heteroscedasticity issue since the red line is far from a sharp V shape. The Breusch-Pagan statistical test also validates this observation. The violation of homoscedasticity can affect the significance of the CO₂ intensities and their confidence intervals.

Figure 7 Homoscedasticity of residuals



4. ADAPTATION TO LCA STANDARD REQUIREMENTS

In LCA applications, it is generally admitted that a negative CO₂ intensity value applies on processes actually removing CO₂ from the atmosphere or achieving a potential net reduction versus a business as usual case (when a marginal or counterfactual approach is followed). As it is not the case for refinery products, CO₂ intensity values complying with LCA standard must be positive.

This section describes a suggested methodology based on the Restricted OLS regression in order to generate LCA compliant positive CO₂ intensities for refinery products. For this purpose, more granularity on bottom of the barrel categories is added.

The numerical application is based on the results presented in section 3.

4.1 STEP A - DECOMPOSITION OF BOTTOM OF THE BARREL PRODUCTS

Bottom of the barrel finished products (heavy fuels) are blends of residues (e.g. residues from Crude distillation, Vacuum distillation, FCC and Visbreaking units) with lighter products in order to satisfy mainly the viscosity, density and Sulphur specifications (main constraints). Based on the results presented in section 3.3 and in Concawe 2017 study, it is anticipated that the lighter products (kerosene or gasoil types) would have positive CO₂ intensities if they were considered as finished products.

The objective of this first step “A” is then to split the bottom of the barrel categories (7 and 8) into fluxants (kerosene and gasoil types) and Residues, in order to segregate the negative CO₂ intensity components from the positive ones. The table below presents the original finished products mass balance, for the 2019 adjusted reference year.

Table 7 Finished product mass balance, million tons per annum

Categories	Mt / a
1 LPG	24.3
2 NAPHTHA	37.4
3 GASOLINE	127.5
4 KEROSENE	51.9
5 DIESEL	174.5
6 HO MARINE DSL	80.6
7 DMF RMF 0.5%S	35.6
8 HSFO	30.3
9 OTHERS	37.4
TOTAL	599.5

In the table below, categories 7 and 8 are split into 3 categories: 2 fluxants (Kerosene, Gasoil) and 1 residue (total Residues).

Table 8 Categories 7 and 8 split, weight percent

Composition, wt.%	7 DMF RMF 0.5%S	8 HSFO	TOTAL
Fluxant - Kerosene	7%	7%	7%
Fluxant - Gasoil	27%	1%	15%
Residues	66%	92%	78%
TOTAL	100%	100%	100%

It is then assumed that because of their nature, Kerosene and Gasoil fluxants have positive CO₂ intensities, respectively equal to categories 4 *Kerosene* and 6 *HO Marine DSL*. For LCA compliance, according the EU LCA community, Residues type is considered as “waste” and consequently a zero CO₂ intensity is assigned to this category. Based on these assumptions, it is possible to calculate a positive CO₂ intensity for category 8 and to re-calculate the category 7. This “DMF-RMF 0.5%S” is originally positive but contain Residue with negative CO₂ intensity, which needs to be corrected to zero (as for cat.7 “HSFO”). Results are shown in the table below.

Table 9 Categories 7 and 8 positive CO₂ intensity calculation, tons of CO₂ per tons of product

CO ₂ Intensity calculation	7 DMF RMF 0.5%S	8 HSFO	CO ₂ Intensities t CO ₂ / t
Fluxant - Kerosene	7%	7%	0.2718
Fluxant - Gasoil	27%	1%	0.2026
Residues	66%	92%	0
CO ₂ Intensities, t CO ₂ / t	0.0751	0.0194	

4.2 STEP B - CO₂ BALANCE

From previous step, there is no more product categories having a negative CO₂ intensity. However, because the new method is additive, if the negative CO₂ intensities are set at zero, the CO₂ balance is no longer ensured, as shown in the following table

Table 10 CO₂ balance

CO ₂ Intensities from OLS regression	Base production Mt / a	CO ₂ Intensities t CO ₂ / t	Positive CO ₂ Intensities t CO ₂ / t
1 LPG	24.3	0.3731	0.3731
2 NAPHTHA	37.4	0.2218	0.2218
3 GASOLINE	127.5	0.2687	0.2687
4 KEROSENE	51.9	0.2718	0.2718
5 DIESEL	174.5	0.3012	0.3012
6 HO MARINE DSL	80.6	0.2026	0.2026
7 DMF RMF 0.5%S	35.6	0.0116	0.0751
8 HSFO	30.3	-0.1905	0.0194
9 OTHERS	37.4	0.0602	0.0602
TOTAL CO ₂ calculated from CO ₂ intensities and Base production, Mt / a		131.5	140.1

There are 8.6 Mt / a of CO₂ imbalance. In the next step, we suggest a methodology to allocate this imbalance to the other categories, in order to satisfy the CO₂ mass balance.

4.3 STEP C - ALLOCATION OF CO₂ IMBALANCE

The CO₂ imbalance resulting from Step B has to be allocated to product categories, so as to preserve the additivity of the CO₂ intensities.

The method “Weighted CO₂ intensities” has been chosen. It consists in allocating the CO₂ imbalance in proportion of the CO₂ allocation (in tons) resulting from the Restricted OLS regression. This methodology has the major advantage of being consistent with the methodology used to generate the CO₂ intensities, it does not require any additional data.

In order to present the possibilities of reallocation, this methodology is compared with more common mass and energy allocations. Firstly, allocation keys are calculated:

- In mass, the allocation keys are the percentages in mass of production.
- In energy, the allocation keys are percentages in mass times lower heating value.
- In mass of CO₂ (“Weighted CO₂ intensities”) the allocation keys are percentages in mass times CO₂ intensities obtained by the Restricted OLS regression.

The table below summarizes the allocation keys of the three different methodologies compared.

Table 11 Allocation keys of CO₂ allocation methodologies - Weighted CO₂ intensities, mass, and energy

Allocation keys	Weighted CO ₂ intensities CO ₂ wt.%	Mass wt.%	Energy energy%
1 LPG	6%	4%	4%
2 NAPHTHA	6%	6%	7%
3 GASOLINE	24%	21%	21%
4 KEROSENE	10%	9%	9%
5 DIESEL	38%	29%	30%
10 MARINE DSL	12%	13%	14%
7 DMF RMF 0.5%S	2%	6%	6%
8 HSFO	0%	5%	5%
9 OTHERS	2%	6%	5%
TOTAL	100%	100%	100%

The relevance of the “Weighted CO₂ intensities” is demonstrated from the category “OTHERS”. With this methodology, these products (Bitumen, Waxes, Lubes, Coke and Sulphur) receive a lower re-allocation (2% versus 6% and 5%) of the CO₂ associated to Residues. From a refinery operation perspective, it makes more sense as their production modes are either fatal (Sulphur, Coke) or dedicated (Bitumen, Waxes, Lubes), hence not correlated to the CO₂ emitted by the production phase of Residues (mainly components for Marine Fuels).

The CO₂ intensities obtained from these allocation keys are summarized in the table below. CO₂ additivity is then recovered: in all cases, total CO₂ recalculated from CO₂ intensities give 131.5 Mt/a of CO₂, which is the total refining CO₂ emissions of the Base case (refer to the Table 4 above).

Table 12 CO₂ intensities resulting from allocation methodologies, ton of CO₂ per ton of product

CO ₂ Intensities from OLS regression	Base production Mt / a	Allocation methodology		
		Weighted CI t CO ₂ / t	Mass t CO ₂ / t	Energy t CO ₂ / t
1 LPG	24.3	0.3501	0.3587	0.3577
2 NAPHTHA	37.4	0.2082	0.2074	0.2068
3 GASOLINE	127.5	0.2522	0.2543	0.2544
4 KEROSENE	51.9	0.2551	0.2574	0.2571
5 DIESEL	174.5	0.2827	0.2868	0.2866
6 HO MARINE DSL	80.6	0.1901	0.1882	0.1881
7 DMF RMF 0.5%S	35.6	0.0705	0.0607	0.0609
8 HSFO	30.3	0.0182	0.0051	0.0058
9 OTHERS	37.4	0.0565	0.0458	0.0479
TOTAL CO ₂ calculated from CO ₂ intensities and Base production, Mt / a		131.5	131.5	131.5

From this step, CO₂ intensities are all positive and additivity of CO₂ emissions is recovered. It can be noted that ranking of the CO₂ intensities is preserved after adaptation to LCA standard requirements, except a switch between DMF RMF 0.5%S and OTHERS categories. This is not a concern at this stage, as OTHERS category CO₂ intensity should not be used for individual product composing the category (refer to sections 3.1 and 3.3.2).

4.4 STEP D - FINAL SET OF POSITIVE CO₂ INTENSITIES

The final set of CO₂ intensities is summarized in the table below.

Table 13 Positive CO₂ intensities of refining finished product categories, ton of CO₂ per ton of product

CO ₂ Intensities from OLS regression	Base production Mt / a	CO ₂ Intensities t CO ₂ / t	LCA compliant Positive CO ₂ intensities t CO ₂ / t
1 LPG	24.3	0.3731	0.3501
2 NAPHTHA	37.4	0.2218	0.2082
3 GASOLINE	127.5	0.2687	0.2522
4 KEROSENE	51.9	0.2718	0.2551
5 DIESEL	174.5	0.3012	0.2827
6 HO MARINE DSL	80.6	0.2026	0.1901
7 DMF RMF 0.5%S	35.6	0.0116	0.0705
8 HSFO	30.3	-0.1905	0.0182
9 OTHERS	37.4	0.0602	0.0565
TOTAL CO ₂ calculated from CO ₂ intensities and Base production, Mt / a		131.5	131.5

In LCA world, it is common to report the CO₂ intensities in grams of CO₂ per megajoule of product. The table below shows the same CO₂ intensities as above, but in g CO₂/MJ.

Table 14 Positive CO₂ intensities of refining finished product categories, grams of CO₂ per megajoule of product

CO ₂ Intensities from OLS regression	CO ₂ Intensities g CO ₂ / MJ	LCA compliant Positive CO ₂ intensities g CO ₂ / MJ
1 LPG	8.14	7.64
2 NAPHTHA	4.96	4.66
3 GASOLINE	6.33	5.94
4 KEROSENE	6.23	5.85
5 DIESEL	6.95	6.52
6 HO MARINE DSL	4.70	4.41
7 DMF RMF 0.5%S	0.27	1.67
8 HSFO	-4.70	0.45
9 OTHERS	1.64	1.54

5. GASOLINE AND DIESEL WELL-TO-WHEELS LCA

This report provides updated refining CO₂ intensities for Gasoline and Diesel. These values can then be combined with reference upstream and downstream values to determine updated Well-to-Wheels (WTW) Gasoline and Diesel CO₂ intensities.

European Council Fuel Quality Directive (EU, 2015) shows a 2010 WTW fossil fuel baseline standard of **94.1 gCO₂eq / MJ**. More precisely it gives **93.2 gCO₂eq / MJ** for Gasoline and **95.0 gCO₂eq / MJ** for Diesel (both based on conventional crude route).

In their WTT and WTW reports (JEC, 2020), JEC provides detailed calculations to determine the WTW CO₂ intensities for Gasoline and Diesel (JEC COG1 and COD1 cases, “Crude oil from typical EU supply, transport by sea, refining in EU (marginal production), typical EU distribution and retail.”). These values are used in the table below, in order to calculate a WTW based on the refining values provided in this report.

Table 15 Well-to-wheels CO₂ intensities of Gasoline and Diesel, grams of CO₂ equivalent per megajoule of product

WTW CO ₂ Intensities	Gasoline g CO ₂ / MJ	Diesel g CO ₂ / MJ
Crude Oil production	9.8	10.0
Crude Oil transport	0.8	0.8
Refining*	5.94	6.52
Distribution	0.6	0.5
Dispensing at retail site	0.4	0.4
Combustion	73.4	73.2
Total WTW*	90.9	91.4
JEC WTW	90.4	92.1
FQD WTW	93.2	95.0

* Calculated in this study

CONCLUSIONS

The key outcome from this innovative development is a unique combination of specific known tools (Linear Programming Model with statistical methods (LHS and OLS) and Restricted Regression) with the fundamental concept of refinery operation, from the process units to the blending composition of the finished products (composition of heavy fuel oils made of intermediate streams both positive and negative in terms of Carbon intensity).

The new method suggested for estimating CO₂ intensities of refining products is consistent with the Generalized Marginal Approach published by Concawe in Report 1/17. A major advantage of this new method is its implementation. It is based on a multi-linear regression of refining CO₂ emissions as a function of finished products quantity. Any entity having a refining model capable of calculating the refinery CO₂ emissions resulting from a finished products demand, can implement this new method.

A numerical application is presented to detail the different steps of the new method. A DOE is generated using a LHS, then the cases are run with the model and finally a Restricted OLS regression is applied on the results. The validity of the CO₂ intensities generated are validated through rigorous statistical tests.

An industrial significance of the results is provided to show the consistency of this new method with the actual constraints of the refining system. The figures are statistically valid (highly robust results) in the area of analysis (+/- 5% of the demand)

Finally, to comply with LCA standards, the new methodology has to be additive and must provide positive CO₂ intensities for all products. Additivity is obtained through a Restricted OLS regression. However, as demonstrated in this report and the 2017 Concawe report, negative CO₂ intensities make sense for a joint production system as refining. To comply with LCA approach, a methodology is suggested to reallocate the negative CO₂ emissions to other finished products, while keeping the additivity of the final set of positive CO₂ intensities.

The resulting datasets are relevant for both attributional and consequential considerations and meet the requirements for most LCA databases (see column “LCA compliant, Positive CO₂ intensities”).

The results of this report cannot be applied to individual refineries and are only valid for the EU refining on its global industry average. Refineries in Europe are all different in their process unit’s configuration and level of complexity, and consequently in their emission intensity and in the product mix they produce.

Based on this new methodology, the updated results for the “2019 adjusted” demand data and for the average EU refining industry are:

CO ₂ Intensities from OLS regression	CO ₂ Intensities g CO ₂ / MJ	LCA compliant Positive CO ₂ intensities g CO ₂ / MJ
1 LPG	8.14	7.64
2 NAPHTHA	4.96	4.66
3 GASOLINE	6.33	5.94
4 KEROSENE	6.23	5.85
5 DIESEL	6.95	6.52
6 HO MARINE DSL	4.70	4.41
7 DMF RMF 0.5%S	0.27	1.67
8 HSFO	-4.70	0.45
9 OTHERS	1.64	1.54

REFERENCES

Tehrani Nejad M. A., 2007. Allocation of CO₂ emissions in petroleum refineries to petroleum joint products: A linear programming model for practical application, *Energy Economics*, Vol. 29, 974-997.

Concawe, 2017. Estimating the marginal CO₂ intensities of EU refinery products, Report 1/17, <https://www.concawe.eu/publication/estimating-the-marginal-co2-intensities-of-eu-refinery-products-report-117/>.

DG Climate, 2020, European Commission; Ricardo Energy & Environment; IFEU. Determining the Environmental Impacts of Conventional and Alternatively Fueled Vehicles through LCA. Final Report. <https://op.europa.eu/en/publication-detail/-/publication/1f494180-bc0e-11ea-811c-01aa75ed71a1>

JRC, 2017. Analysis of standard and innovative methods for allocating upstream and refinery GHG emissions to oil products, *Applied Energy*, ISSN 0306-2619, 206, JRC105041, <https://publications.jrc.ec.europa.eu/repository/handle/JRC105041>.

JEC, 2020. JEC Well-To-Wheels report v5, EUR 30284 EN, Publications Office of the European Union, Luxembourg, 2020, ISBN 978-92-76-20109-0, doi:10.2760/100379, JRC121213, <https://publications.jrc.ec.europa.eu/repository/handle/JRC121213>.

JEC, 2020. JEC Well-To-Tank report v5, EUR 30269 EN, Publications Office of the European Union, Luxembourg, 2020, ISBN 978-92-76-19926-7, doi:10.2760/959137, JRC119036, <https://publications.jrc.ec.europa.eu/repository/handle/JRC119036>

European Union, 2014. 2014/746/EU: Commission Decision of 27 October 2014 determining, pursuant to Directive 2003/87/EC of the European Parliament and of the Council, a list of sectors and subsectors which are deemed to be exposed to a significant risk of carbon leakage, for the period 2015 to 2019 (notified under document C(2014) 7809), Document 32014D0746, <https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=CELEX:32014D0746>.

European Union, 2020, SWD/2020/176 final: Commission Staff Working document impact assessment accompanying the document communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions stepping up Europe's 2030 climate ambition investing in a climate-neutral future for the benefit of our people, Document 52020SC0176, <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52020SC0176>.

European Union, 2015. Council Directive (EU) 2015/652 of 20 April 2015 laying down calculation methods and reporting requirements pursuant to Directive 98/70/EC of the European Parliament and of the Council relating to the quality of petrol and diesel fuels, Document 32015L0652, <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32015L0652>

European Commission, 2021. Fuel quality monitoring in the EU in 2019, ETC/CME Eionet Report, <https://www.eionet.europa.eu/etcs/etc-cme/products/etc-cme-reports/etc-cme-report-1-2021-fuel-quality-monitoring-in-the-eu-in-2019>

Concawe, 2017. Market Fuels Survey: 2015 - 2016, Report 10/17, <https://www.concawe.eu/publication/concawe-market-fuels-survey-2015-2016/>

Moretti C, Moroa A, Edwardsa R, Vincenzo Roccob M, Colombo E (2017) Analysis of standard and innovative methods for allocating upstream and refinery GHG emissions to oil products. *Applied Energy* 206, 372-381, <https://www.sciencedirect.com/science/article/pii/S0306261917312102>

Baumol, William, Behling B, Bonbright J, Brozen Y, Dean J, Edwards F, Hoover C, Pegrum D, Roberts M, Williams E (1962) The Role of Cost in the Minimum Pricing of Railroad Services. *Journal of Business*, 35(4)357-366, <https://www.jstor.org/stable/2350611>

Johnson E, Vadenbo C (2020) Modelling Variation in Petroleum Products' Refining Footprints. *Sustainability*. 12, 9316, 1-15. doi:10.3390/su12229316, <https://www.mdpi.com/2071-1050/12/22/9316/htm>

GLOSSARY

CO ₂ eq	CO ₂ equivalent
DMF	Diesel Marine Fuel
DOE	Design of experiment
DSL	Diesel
ETBE	Ethyl Tert-Butyl Ether
EU	European Union
g	Gram
HO	Heating Oil
HSFO	High Sulphur Fuel Oil
JEC	JRC-EUCAR-Concawe
JRC	Joint Research Centre of the European Commission
kbbl	Kilo barrel
LCA	Life Cycle Analysis
LHS	Latin Hypercube Sampling
LP	Linear Programming
LPG	Liquefied Petroleum Gases
LSFO	Low Sulphur Fuel Oil
MJ	Mega Joule
Mt / a	Millions tonnes per annum
MTBE	Methyl Tert-Butyl Ether
OLS	Ordinary Least Square
RMF	Residual Marine Fuel
t	Metric tonne
TAAE	Tert-Amyl Ethyl Ether
WT%	Weight percent
VOL%	Volume percent
WTT	Well to Tank
WTW	Well to Wheels

APPENDIX 1 - DETAILED ASSUMPTIONS

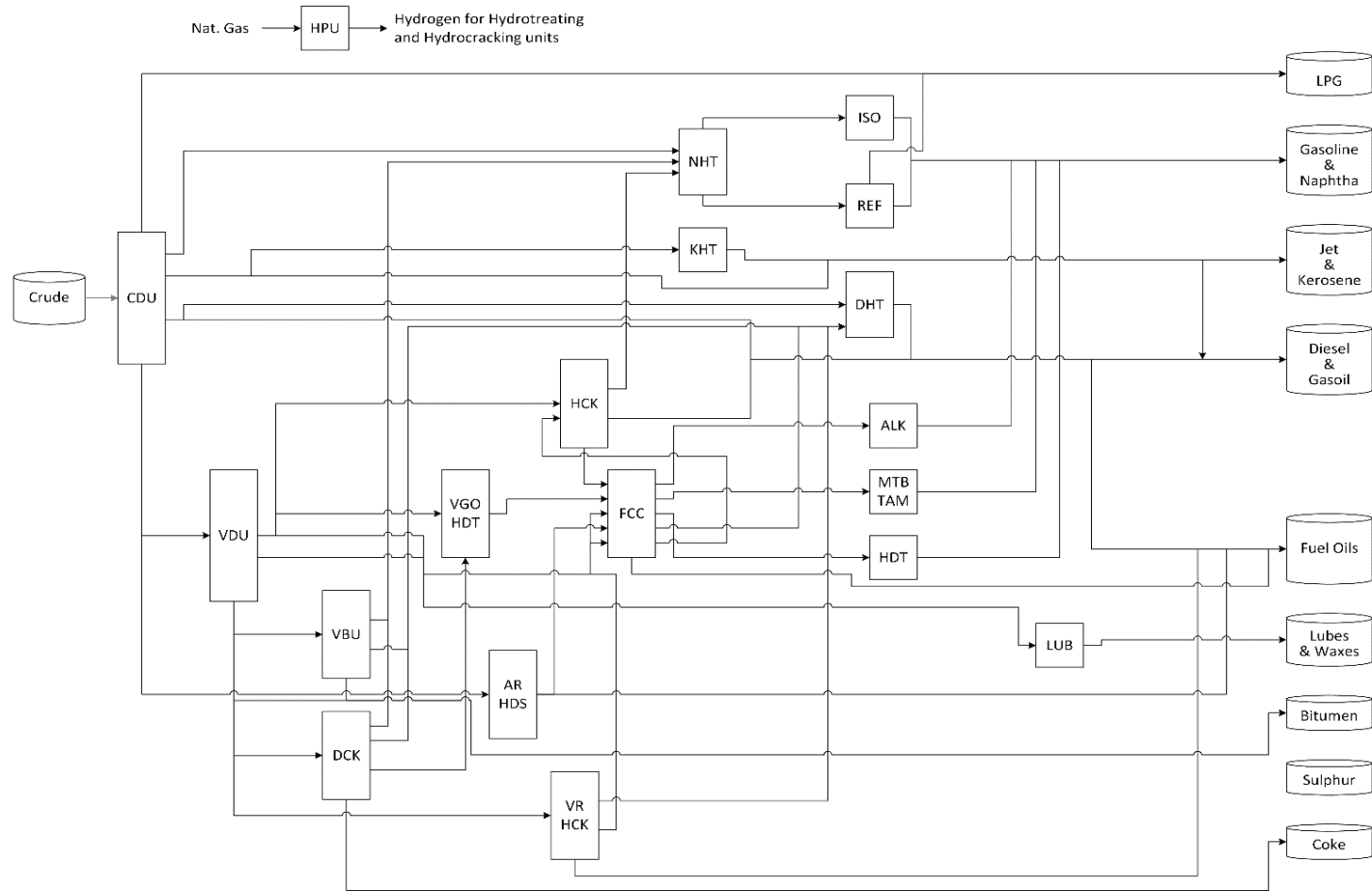
Countries included in the LP modelling

Table 16 Countries included in the LP model

European Union + 4	European Union + 4	European Union + 4
Austria	Italy	Iceland
Belgium	Latvia	Norway
Bulgaria	Lithuania	Switzerland
Croatia	Luxembourg	United Kingdom
Cyprus	Malta	
Czechia	Netherlands	
Denmark	Poland	
Estonia	Portugal	
Finland	Romania	
France	Slovakia	
Germany	Slovenia	
Greece	Spain	
Hungary	Sweden	
Ireland		

Simplified Block Flow Diagram of the LP model

Table 17 Simplified BFD of the LP model



Crude slate

The first table shows the EU Commission crude supply data. The second table is the LP model crude slate matching the EU average crude slate properties (API gravity, Sulphur content, main products yield). The third table gives the main crude slate properties.

Table 18 Registration of Crude Oil Imports and Deliveries in the European Union, 2019

Region	Country of Origin	Type of crude oil	kbbl
Africa	Algeria	Saharan Blend	124 283
Africa	Algeria	Other Algeria Crude	11 833
Africa	Angola	Cabinda	6 679
Africa	Angola	Other Angola Crude	45 559
Africa	Cameroon	Cameroon Crude	12 601
Africa	Congo	Congo Crude	2 775
Africa	Congo (DR)	Congo (DR) Crude	1 852
Africa	Egypt	Medium/Light (30-40o)	34 521
Africa	Gabon	Rabi/Rabi Kounga	621
Africa	Gabon	Other Gabon Crude	5 252
Africa	Libya	Medium (30-40o)	154 982
Africa	Libya	Heavy (<30o API)	13 017
Africa	Libya	Light (>40o)	78 861
Africa	Nigeria	Medium (<33o)	125 516
Africa	Nigeria	Light (33-45o)	176 646
Africa	Nigeria	Condensate (>45o)	7 010
Africa	Other African Countries	Other Africa Crude	36 414
Africa	Tunisia	Tunisia Crude	3 939
America	Argentina	Argentina Crude	935
America	Brazil	Brazil Crude	32 650
America	Canada	Canadian Heavy (<33° API)	15 574
America	Canada	Light Sweet (>30o API)	15 691
America	Colombia	Other Colombia Crude	8 808
America	Mexico	Maya	69 460
America	Other Latin America countries	Other Latin America Crude	773
America	United States	Alaska	1 240
America	United States	Other US Crude	282 570
America	Venezuela	Medium (22-30o)	1 913
America	Venezuela	Heavy (17-22o)	3 613
America	Venezuela	Extra Heavy (<17o)	28 398
Europe	Denmark	Denmark Crude	12 587
Europe	Norway	Statfjord	21 416
Europe	Norway	Ekofisk	62 915
Europe	Norway	Other Norway Crude	235 938
Europe	Norway	Oseberg	31 852
Europe	Norway	Gullfaks	44 038
Europe	Other European countries	Other Europe Crude	8 711
Europe	United Kingdom	Flotta	18 374
Europe	United Kingdom	Forties	16 186

Region	Country of Origin	Type of crude oil	kbbbl
Europe	United Kingdom	Brent Blend	26 946
Europe	United Kingdom	Other UK Crude	126 615
FSU	Azerbaijan	Azerbaijan Crude	163 506
FSU	Kazakhstan	Kazakhstan Crude	292 236
FSU	Other FSU countries	Other FSU Crude	21 909
FSU	Russian Federation	Other Russian Fed. Crude	391 041
FSU	Russian Federation	Urals	578 878
FSU	Ukraine	Ukraine Crude	581
Middle East	Iraq	Basrah Light	151 686
Middle East	Iraq	Kirkuk	54 928
Middle East	Iraq	Other Iraq Crude	127 004
Middle East	Kuwait	Kuwait Blend	38 515
Middle East	Saudi Arabia	Arab Light	208 988
Middle East	Saudi Arabia	Arab Medium	3 288
Middle East	Saudi Arabia	Other Saudi Arabia Crude	7 889
Middle East	Saudi Arabia	Arab Heavy	42 905
Middle East	Saudi Arabia	Berri (Extra Light)	24 759
Middle East	Syria	Syria Light	107
Middle East	Yemen	Other Yemen Crude	1 425
TOTAL			4 019 210

Table 19 LP Model crude slate

Composition, wt%	7 DMF RMF 0.5%S
Brent	24.1%
Forcados	20.0%
Russian Export	20.0%
Iranian Light	20.0%
Kuwait	12.5%
Algerian Condensate	3.4%
Russian Long Residue	0.0%
Brent Short Residue	0.0%
TOTAL	100%

Table 20 EU crude slate main properties (estimated)

Property	Value
API gravity	34.2
Specific Gravity	0.854
Sulphur content, wt%	0.98
C5-180C TBP cut, wt%	22.1%
180-350C TBP cut, wt%	32.0%
350+ C TBP cut, wt%	44.3%
550+ C TBP cut, wt%	14.9%

LP model demand

Table 21 LP model demand - EU 2019 adjusted refining system production, millions tons per annum

Product	Mt / a
LPG	24.3
Chemical Naphtha	37.4
Premium 95 Gasoline	76.0
Super 98 Gasoline	18.7
US export Gasoline	17.9
Other export Gasoline	14.9
Jet and Kerosene	51.9
Road Diesel	174.5
Diesel Marine Fuel	10.3
Heating Oil	70.3
DMF 0.5%S	17.8
RMF 0.5%S	17.8
HSFO 3.5%S	20.3
LSFO 1.0%S	10.0
Bitumen	18.3
Lubes	6.5
Waxes	0.8

Finished products specifications
Table 22 Summary of LP model main finished products specifications

		95 Gasoline	98 Gasoline	US Gasoline	Other Gasoline	Jet	Road Diesel	Heating Oil	Diesel Marine Fuel	DMF 0.5%S	RMF 0.5%S	LSFO 1.0%S	HSFO 3.5%S
Specific gravity	-	0.720-0.775	0.720-0.775	0.725-0.775	0.7-0.8	≥ 0.775	0.820-0.845	0.800-0.850	0.8-0.9	0.8-0.9	0.900-0.991	0.900-0.995	0.900-0.991
Sulphur content	wt%	≤ 0.0010	≤ 0.0010	≤ 0.0010	≤ 0.0500	≤ 0.3	≤ 0.0010	≤ 0.1	≤ 0.1	≤ 0.5	≤ 0.5	≤ 1.0	≤ 3.5
Olefins content	vol%	≤ 18.0	≤ 18.0	≤ 10.0									
Aromatics content	vol%	≤ 35.0	≤ 35.0	≤ 28.0									
Benzene content	vol%	≤ 1.0	≤ 1.0	≤ 0.6	≤ 1.5								
Oxygen content	wt%	≤ 3.7	≤ 3.7										
RVP	kPa	≤ 60.0	≤ 60.0	≤ 58.0	≤ 60.0								
Evaporated @70°C	vol%	24-50	24-50	20-45	20-45								
Evaporated @100°C	vol%	46-72	46-72	47-65	47-65								
Evaporated @150°C	vol%	≥ 75	≥ 75										
Evaporated @250°C	vol%						≤ 65	≤ 65					
Evaporated @350°C	vol%						≥ 85	≥ 85					
RON	-	≥ 95.0	≥ 98.0	≥ 92.0	≥ 91.0								
MON	-	≥ 85.0	≥ 88.0	≥ 82.0	≥ 81.0								
Cetane index	-						≥ 49	≥ 41	≥ 40	≥ 40			
Viscosity @40°C	cSt						2.0-4.5	≤ 4.5	≤ 11	≤ 6			
Viscosity @100°C	cSt										≤ 35	≤ 35	≤ 35
Poly Aromatics content	wt%						≤ 8						
Cloud point	°C						≤ -5	≤ -10	≤ 0	≤ 0			
Pour point	°C								≤ 0	≤ 0	≤ 30		≤ 30
Conradson carbon	wt%										≤ 18	≤ 18	≤ 18

LP model main process units' capacities

Table 23 LP model main process units' capacities, millions tonnes per annum

Process unit	Original Capacities Mt/a	Calibrated Capacities Mt/a
Crude Distillation	668.1	668.1
Vacuum Distillation	314.5	314.5
Visbreaker	78.2	78.2
Delayed Coker	31.2	31.2
De-Asphalting	4.4	4.4
Fluid Cat Cracker	112.3	112.3
Hydrocracker	80.6	80.6
FCC VGO Pre-treatment	51.0	51.0
Atm. Resid. Hydrodesulphurisation	2.8	2.8
Vac. Resid. Hydrodesulphurisation	1.9	1.0
Resid. Conversion	14.1	7.7
Naphtha Hydrotreatment	151.8	151.8
Reformers	88.4	88.4
Alkylation	9.4	9.4
Isomerisation	22.5	22.5
Kerosine Hydrodesulphurisation	50.9	28.0
Distillate Hydrodesulphurisation	212.7	117.0
Sulphur recovery	5.8	5.8
Hydrogen manufacture	2.2	2.2

APPENDIX 2 - DOE ASSUMPTIONS

Table 24. Finished products allocation to DOE categories - Base case mass balance, 2019 adjusted reference year

Product	Mt / a
1 LPG	24.3
<i>LPG</i>	24.3
2 NAPHTHA	37.4
<i>Chemical Naphtha</i>	37.4
3 GASOLINE	127.5
<i>Premium 95 Gasoline</i>	76.0
<i>Super 98 Gasoline</i>	18.7
<i>US export Gasoline</i>	17.9
<i>Other export Gasoline</i>	14.9
4 KEROSENE	51.9
<i>Jet and Kerosene</i>	51.9
5 DIESEL	174.5
<i>Road Diesel</i>	174.5
6 HO MARINE DSL	80.6
<i>Diesel Marine Fuel</i>	10.3
<i>Heating Oil</i>	70.3
7 DMF RMF 0.5%S	35.6
<i>DMF 0.5%S</i>	17.8
<i>RMF 0.5%S</i>	17.8
8 HSFO	30.3
<i>HSFO 3.5%S</i>	20.3
<i>LSFO 1.0%S</i>	10.0
9 OTHERS	37.4
<i>Bitumen</i>	18.3
<i>Lubes</i>	6.5
<i>Waxes</i>	0.8
<i>Coke</i>	8.2
<i>Sulphur</i>	3.6

APPENDIX 3 - OLS REGRESSION RESULTS

Figure 8 OLS regression results

Generalized Linear Model Regression Results						
=====						
Dep. Variable:	CO2	No. Observations:	486			
Model:	GLM	Df Residuals:	478			
Model Family:	Gaussian	Df Model:	7			
Link Function:	identity	Scale:	0.0075267			
Method:	IRLS	Log-Likelihood:	502.53			
Date:	Thu, 17 Mar 2022	Deviance:	3.5978			
Time:	15:01:35	Pearson chi2:	3.60			
No. Iterations:	1	Pseudo R-squ. (CS):	1.000			
Covariance Type:	nonrobust					
=====						
	coef	std err	z	P> z	[0.025	0.975]

LPG	0.3731	0.012	30.959	0.000	0.349	0.397
NAPHTHA	0.2218	0.005	40.441	0.000	0.211	0.233
GASOLINE	0.2687	0.002	124.680	0.000	0.264	0.273
KEROSENE	0.2718	0.004	63.084	0.000	0.263	0.280
DIESEL	0.3012	0.001	231.159	0.000	0.299	0.304
HO MARINE DSL	0.2026	0.002	85.927	0.000	0.198	0.207
DMF RMF 0.5%	0.0116	0.006	2.064	0.039	0.001	0.023
HSFO	-0.1905	0.007	-29.246	0.000	-0.203	-0.178
OTHERS	0.0602	0.012	5.133	0.000	0.037	0.083
=====						

APPENDIX 4 - STATISTICAL CONCEPTS

This section contains the basic definition for the main concepts developed and used in this report.

Design of Experiment (DOE)

The design of experiments is the design of any task that aims to describe and explain the variation of information under conditions that are hypothesized to reflect the variation. The term is generally associated with experiments in which the design introduces conditions that directly affect the variation, but may also refer to the design of quasi-experiments, in which natural conditions that influence the variation are selected for observation.

In its simplest form, an experiment aims at predicting the outcome by introducing a change of the preconditions, which is represented by one or more independent variables, also referred to as "input variables" or "predictor variables." The change in one or more independent variables is generally hypothesized to result in a change in one or more dependent variables, also referred to as "output variables" or "response variables." The experimental design may also identify control variables that must be held constant to prevent external factors from affecting the results. Experimental design involves not only the selection of suitable independent, dependent, and control variables, but planning the delivery of the experiment under statistically optimal conditions given the constraints of available resources. There are multiple approaches for determining the set of design points (unique combinations of the settings of the independent variables) to be used in the experiment. Main concerns in experimental design include the establishment of validity, reliability, and replicability.

Latin Hypercube Sampling (LHS)

Latin hypercube sampling (LHS) is a statistical method for generating a near-random sample of parameter values from a multidimensional distribution.

In the context of statistical sampling, a square grid containing sample positions is a Latin square if (and only if) there is only one sample in each row and each column. A Latin hypercube is the generalisation of this concept to an arbitrary number of dimensions, whereby each sample is the only one in each axis-aligned hyperplane containing it.

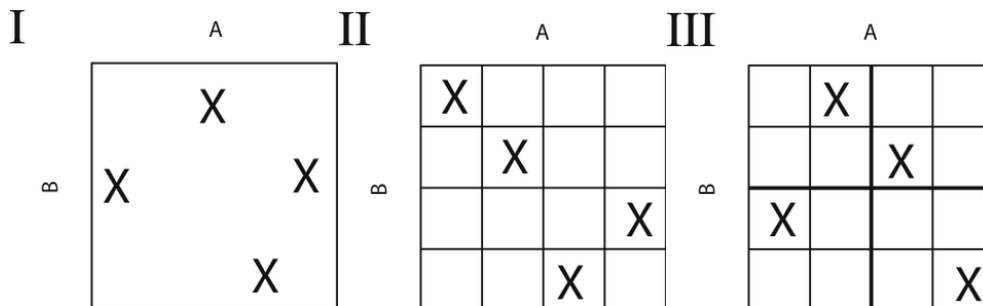
When sampling a function of N variables, the range of each variable is divided into M equally probable intervals. M sample points are then placed to satisfy the Latin hypercube requirements; this forces the number of divisions, M , to be equal for each variable. This sampling scheme does not require more samples for more dimensions (variables); this independence is one of the main advantages of this sampling scheme. Another advantage is that random samples can be taken one at a time, remembering which samples were taken so far.

In two dimensions the difference between random sampling, Latin hypercube sampling, and orthogonal sampling can be explained as follows:

In random sampling new sample points are generated without taking into account the previously generated sample points. One does not necessarily need to know beforehand how many sample points are needed.

In Latin hypercube sampling one must first decide how many sample points to use and for each sample point remember in which row and column the sample point was taken. Such configuration is similar to having N rooks on a chess board without threatening each other.

In orthogonal sampling, the sample space is divided into equally probable subspaces. All sample points are then chosen simultaneously making sure that the total set of sample points is a Latin hypercube sample and that each subspace is sampled with the same density.



Thus, orthogonal sampling ensures that the set of random numbers is a very good representative of the real variability, LHS ensures that the set of random numbers is representative of the real variability whereas traditional random sampling (sometimes called brute force) is just a set of random numbers without any guarantees.

Ordinary Least Squares regression (OLS)

In statistics, ordinary least squares (OLS) is a type of linear least squares method for estimating the unknown parameters in a linear regression model. OLS chooses the parameters of a linear function of a set of explanatory variables by the principle of least squares: minimizing the sum of the squares of the differences between the observed dependent variable (values of the variable being observed) in the given dataset and those predicted by the linear function of the independent variable.

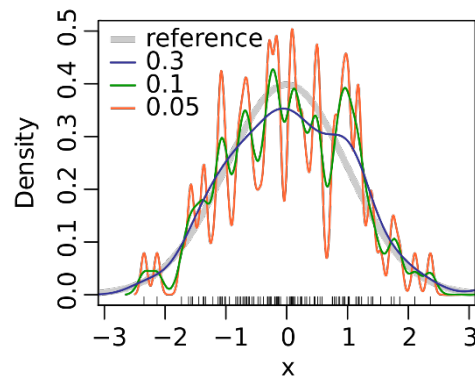
Kernel density curve

In statistics, kernel density estimation (KDE) is a non-parametric way to estimate the probability density function of a random variable. Kernel density estimation is a fundamental data smoothing problem.

Conceptually, a smoothly curved surface is fitted over each point. It can be considered as a smooth representation of an histogram.

The bandwidth of the kernel is a free parameter which exhibits a strong influence on the resulting estimate. The bandwidth must be chosen carefully. A small bandwidth will create a very overfitted curve while a too big bandwidth will create an oversmoothed curve.

To illustrate its effect, we take a simulated random sample from the standard normal distribution (plotted at the blue spikes in the rug plot on the horizontal axis). The grey curve is the true density (a normal density with mean 0 and variance 1). In comparison, the red curve is under smoothed since it contains too many spurious data artifacts arising from using a bandwidth $h = 0.05$, which is too small. The green curve is oversmoothed since using the bandwidth $h = 2$ obscures much of the underlying structure. The black curve with a bandwidth of $h = 0.337$ is considered to be optimally smoothed since its density estimate is close to the true density.



Pearson correlation matrix

In statistics, the Pearson correlation coefficient is a measure of linear correlation between two sets of data. It is the ratio between the covariance of two variables and the product of their standard deviations; thus, it is essentially a normalized measurement of the covariance, such that the result always has a value between -1 and 1 . As with covariance itself, the measure can only reflect a linear correlation of variables, and ignores many other types of relationship or correlation. As a simple example, one would expect the age and height of a sample of teenagers from a high school to have a Pearson correlation coefficient significantly greater than 0 , but less than 1 (as 1 would represent an unrealistically perfect correlation).

The Pearson correlation matrix is simply a table of correlations, which compares side-by-side Pearson's correlation coefficient of two variables.

Adjusted R-Squared

R-squared is the coefficient of determination indicating goodness-of-fit of the regression. This statistic will be equal to one if fit is perfect, and to zero when regressors X have no explanatory power whatsoever. This is a biased estimate of the population R-squared, and will never decrease if additional regressors are added, even if they are irrelevant.

Adjusted R-squared is a slightly modified version of R-squared, designed to penalize for the excess number of regressors which do not add to the explanatory power of the regression. This statistic is always smaller than R-squared, can decrease as new regressors are added, and even be negative for poorly fitting models.

p-value

In null-hypothesis significance testing, the p-value is the probability of obtaining test results at least as extreme as the result actually observed, under the assumption that the null hypothesis is correct (i.e., parameter not significant). A very small p-value means that such an extreme observed outcome would be very unlikely under the null hypothesis. Reporting p-values of statistical tests is common practice in academic publications of many quantitative fields.

F-Statistic (F-test)

F-statistic tries to test the hypothesis that all coefficients (except the intercept) are equal to zero. This statistic has $F(p-1, n-p)$ distribution under the null hypothesis and normality assumption, and its p-value indicates probability that the hypothesis is indeed true.

The F-test of overall significance indicates whether a linear regression model provides a better fit to the data than a model that contains no independent variables. If the p-value for the F-test is less than the significance level, the sample data provide sufficient evidence to conclude that the regression model fits the data better than the model with no independent variables.

Durbin-Watson

Durbin-Watson statistic tests whether there is any evidence of serial correlation between the residuals. A Durbin-Watson (d) of 2 indicates no autocorrelation. The value of d always lies between 0 and 4. If d is substantially less than 2, there is evidence of positive serial correlation. As a rough rule of thumb, if Durbin-Watson is less than 1.0, there may be cause for alarm. Small values of d indicate successive error terms are positively correlated. If $d > 2$, successive error terms are negatively correlated.

Homoscedasticity and heteroscedasticity

In statistics, a sequence (or a vector) of random variables is homoscedastic if all its random variables have the same finite variance. This is also known as homogeneity of variance. The complementary notion is called heteroscedasticity.

Assuming a variable is homoscedastic when in reality it is heteroscedastic results in unbiased but inefficient point estimates and in biased estimates of standard errors, and may result in overestimating the goodness of fit as measured by the Pearson coefficient.

The existence of heteroscedasticity is a major concern in regression analysis and the analysis of variance, as it invalidates statistical tests of significance that assume that the modelling errors all have the same variance.

A scatterplot of residuals versus predicted values is good way to check for homoscedasticity. There should be no clear pattern in the distribution; if there is a cone-shaped or a v-shaped pattern, the data is heteroscedastic.

Concawe
Boulevard du Souverain 165
B-1160 Brussels
Belgium

Tel: +32-2-566 91 60
Fax: +32-2-566 91 81
e-mail: info@concawe.org
<http://www.concawe.eu>

ISBN 978-2-87567-163-9



9 782875 671639 >