"Estimating Operations' Efficiencies, through Business Intelligence, as an Input for the Assignation of Machines in Agricultural Processes"

Master of Science Thesis

Written by:

JUAN PABLO BOTERO TERREROS

Tutors:

FRANCESCO LOMBARDI (Academic)

MASSIMO DE MARZIIS (Enterprise)

POLITECNICO DI TORINO

DIPARTIMENTO DI INGEGNERIA INDUSTRIALE E DELLA PRODUZIONE

TORINO, ITALY

2018

Contents

Figures

Tables

Glossary

The following concepts have been taken from 3a S.r.l Drying, 2018

Big Bag: Flexible container of standardized measures used for the transportation and storage of the hazelnut

Bins: Plastic recipients where the hazelnuts are transported.

Cart: vehicle with an identification code that is pulled by tractors and harvesters.

Detail: Groups of acceptations lots characterized by the fact of being dried in the same silo and at the same time.

Line: the part of the drying plant including the transportation lines and the drying silo.

Load: minimum element of picking.

Machines: Tractors and Implements used for a specific operation.

Operation: a sub-operation done with certain resources, that has certain characteristics.

Prescription: Products and their doses.

Tare: Difference between the weight of the container with the load, and without it.

Temporal Warehouse: It could be referred to the one in the field, which is close to the fields where the farming and harvesting is done, and where the hazelnuts are gathered before being sent to the drying plant; or to the acceptation one, where the product is stored temporarily before entering into the processing line. In addition, it could be understood as the final warehouse, where the hazelnuts are stored before being dispatched.

Territorial Unit: Configuration of the division of a company.

Transportation document: Paper document containing all the information related with the content of the camion, the camion itself, and the driver.

Abstract

New Technologies and constant dynamism are key factors to be successful in a specific market. The ability to reinvent themselves in terms of processes and organization, analyzing what could be improved, requires to be a fundamental pillar of the company. Fortunately, 3a S.r.l is aware of this situation, and this is precisely what gave support to the current master thesis.

Starting from an introduction of the environment in which 3a is immersed in, where the context is presented, until the process of designing an application for a better assignation of resources; this work follows a logical line that allows to understand step by step how the objectives were reached.

First, an explanation of the framework is given, in way such that it is possible to recognize the main core of the company, going through its products, its location, its productive process, and the way in which data is entered and handled into the platforms it offers to its clients. Next, the context of the problem is illustrated. In this case, it deals with efficiency estimation for the operations carried out within the Farms of the company's client in Chile, estimating, throughout Business Intelligence, the ratio between quantity processed and duration. In doing so, a data model was created, extracting what was needed from a testing DB and submitting it to an ETL process. The results obtained have been statistically validated and measures regarding them have been discussed with people in charge.

Finally, what was obtained previously regarding efficiencies has been used to create an algorithm (using Visual Basic), that allows a comparison between the current criteria of assignation of resources and a proposed one that follows the logic of a Greedy algorithm. This, with the objective of analyzing if the current one could be improved, subject to some assumptions, and also to evaluate what the behavior of the operations has been throughout the years.

Keywords

Business Intelligence, Greedy Algorithm, Hazelnut, SQL Server, Visual Basic, Operations, Machines, Duration, Fuel Consumption.

1. Introduction

The current thesis starts with the identification of a need from the enterprise 3a S.r.l to analyze its data, stored in its DB, in a systematic way. In addition, some optimization tools were needed to be implemented to find the combination of resources that would allow a better estimation of the budget for the harvesting process of the hazelnut in Chile.

For a clearer contextualization, it is necessary to point out that 3a S.r.l. is the company that provides the Information System services to an agricultural organization based in Chile, called *Frutícola Agrichile*. Working hand by hand, *Agrichile* and 3a allow the input of high quality raw material to *Ferrero.* In doing so, the process is cautiously monitored throughout the whole chain, keeping track of its operations, both, the ones taking place in the field, and in the factory.

For this purpose, 2 (two) applications have been adopted: FMP, which takes care of all the activities related to the farming; and HDS, which manages the drying process. In the former, the user stores information about the operations done, what implements and tractors were used to perform it, what chemical products were necessary for the correct growth of the plant, how long did the operation took, the location of the field, and the unit of measure of the respective operation, among others; whereas the latter stores data related of the drying process, usually taking place within the drying plants, which can be divided into the following major processes: Arrival and approval, quality analysis, cleaning and drying; and storage and transfer.

However, even though data has been stored in the database through the former platforms (It has been stored for 4 (four) years for the case of FMP, and for 1 (one) year for HDS), it has never been analyzed.

Following this line of reasoning, it is clear that the use of information to reach conclusions that will allow the company to take strategic decisions, is of extremely importance. In this sense, the objective is to use Business Intelligence in order to analyze specific information for some metrics of performance; and, in addition, use this analysis to propose a better assignation of machines and implements with respect to the one that has been done historically.

In the process of achieving this complex task, the following structure has been defined:

1. Formal presentation of the problem, and the objectives to achieve.
2. Within the second section, a presentation of 3a S.r.l, including its foundation, its 2 (two) main activities, and its main products. In addition, an introduction of one of its main clients, *Fruticola Agrichile*.
3. Production Process.
4. Description of the Information Flow.
5. Categorizing of the activities.
6. Languages and tools used in the work.
7. Efficiency estimation for each of the sub-operations with their respective implement and machine.
8. Statistical validation of the results from previous point.
9. Algorithm proposal for the assignation of machines and implements, minimizing duration with respect to historical values.
10. Conclusions and future lines of actions.

2. Definition of the Problem

The last section gave a general idea of what is going to be done throughout the present work, and the way in which it will be developed following a logical structure. Nonetheless, delimiting more the scope of the thesis, it could be defined from the two (2) subsequent main problems:

1. Estimation of the Efficiency in terms of hectares per day, or number of plants per day for each set of sub-operations and machines used. This is, for each sub-operation, "machine collection" as an instance, a different combination of Vehicles and implements have been used throughout the years to perform it. Each of this combination has have a different time to complete the activity, related to a quantity. The efficiency is then the ratio between these two.
2. Use the efficiency computed for each of the sets, to understand if the assignation of the machines has been done, throughout the years, in an efficient way. And if not, what combination of resources could give a lower duration to a specific operation. This, to define future lines of actions in the terms of how the assignation of machines is being done.

3. Objectives

Analyzing the information available in the FMP platform, and confronting it with what was discussed with Mr. De Marziis (CEO of 3a S.r.l), the author has come to define the structure of the deliverable through the following main objectives:

1. Extract, Transform, and Load (ETL) process with Power BI application, in order to create a data warehouse with useful information. The former will result as an output which will be used for efficiency computations in a systematic way.
2. Perform an analysis with the data model obtained in the previews point. As a consequence, a global view of the process will be displayed, and strategic decisions regarding the data could be taken. More precisely, comparisons between the current method to estimate efficiencies, and the one proposed in the thesis, will be made.
3. To create an algorithm that will make possible, for the specific case of the Company San Sebastian in Chile, to calculate a near-optimal solution for the assignation of the machines, i.e. tractors and implements. That is, a well-defined combination of productive factors used in the agricultural processes that will allow a flexible estimation of the theorical duration of the activity.

It is important to point out that, although separate objectives, they follow a coherent sequence, which means that the output of one objective will be the input for the next one. It is in fact, for this reason, that all of them have the same weight and are essential for the scope of the work. *Figure 1* illustrates this relation.

Figure 1. Logical sequence of the objectives. Elaborated by author

4. 3a S.r.l

As stated on their webpage (3a S.r.l Green Planet, 2018): "**3a S.r.l.** is an Italian company that has been working in ICT for agriculture for approximately 20 years. The company was created by the integration of the founders' expertise in meteorology, in Agro-management and in IT solutions". It is a recognized member of the ICT cluster in Turin (ICT Pole), which brings together the most innovative companies in the industry. It is participating in several research projects in partnership with universities.

To achieve its strategy, 3a is organized into 2 (two) integrated areas: the first one deals with technical assistance in agriculture, focusing its activities within the sphere of integrated farming, identifying agrometeorological information as a fundamental factor in the rational control of pest management and the use of water resources. The second area of 3a activity provides ICT tools for the food companies. It deals with the design of ERP (enterprise resources planning) solution for the management control of the farms, DSS (decision support system) for farming, and supplying mobile solution, software, and devices, for the field force, and IoT technology: weather stations and GPS device for remote machine control."

Its core products are defined as the ones that differentiate the company of the others in the field, and are not imitable due to numerous factors, such as know-how, client-enterprise relationship, and trust, among others. Those previously mentioned products are listed and explained below.[1]

4.1. GPS Remote Control System for Tractor-Implement Management

This product offers allows the use to register data about the geo-location of agricultural machines and obtain statistics relating to their operation. It consists of an on-board module (gps/gprs locator), a remote control, data acquisition and processing system and information display software (web platform).

[1] The information about the products and the company was obtained from 3a S.r.l Green Planet, 2018.

The solution can be integrated as an additional module of the Green Planet Farm Management Platform (it will be explained latter), providing information relating to the work hours of the machines and their distribution in different company plots (geofencing).

The system is able to provide the following information:

- Geographical position of the tractor or self-propelled machines or power implements
- Average speed, machine status (on, off, on the move)
- Recognition of the operating machine towed by a tractor
- Calculation of engine hours and PTO hours of use
- Statistics on machine hours organized according to closed geographical areas and geo-referenced on the map
- Monitoring of fuel levels
- Status change alerts

4.2.Weather Monitoring Technology for Integrated Agriculture

It is an innovative low-cost monitoring solution, designed as a specific weather station for agricultural application, solid and reliable. Its standard version features a set of sensors able to process and transmit 11 meteorological parameters: 5 recorded directly (Temperature and Humidity, Atmospheric Pressure, Precipitation, Wind Speed) and 6 elaborated from sophistical algorithms (Global Solar Radiation, Net Solar Radiation, Dew Temperature, Wt. Bulb Temperature, Leaf Wetness, Potential Evapotranspiration). The following table shows what the physical sensors measure, and what data it is elaborated to help the process:

Table 1. Measurements for Monitoring

Measurements recorded on physical sensors	Data elaborated from virtual sensors
Air temperature	Wet Bulb Temperature
Air humidity	Dew Temperature
Atmospheric Pressure	Potential Evapotranspiration
Wind Speed	Global Daily Radiation
Wind Direction (option)	Net Daily Radiation
Precipitation (accumulation and intensity)	Leaf Wetness

Available at: http://www.green-planet.it/index.php/en/

The station transmits data to the system with a GPRS and uses a 12V 18Ah battery, recharged by a 20W solar panel. The structure is made in iron galvanized, powder coated white. To display and manage data, two application tools are available: A Web-GIS monitoring platform and a webApp for Android devices (tablet and smartphones).

4.3.Weather and Crop Management Android App

Designed for agricultural activities in order to meet technician's needs, works also as a "way navigator", continuously updating while travelling by car. The current mobile application is intended to work just online.

The following activities are supported by the system:

- viewing of meteorological charts
- data entry about crop observation:
    - phenology
    - phytosanitary survey
    - pest symptoms
    - harmful insect observations and trap capture
    - damages for meteorological events
- data entry about agronomical activity:
    - treatment
    - fertilization
    - irrigation
    - system warning due to meteorological or phytosanitary risk

4.4. Green Planet Farm Management Web Platform (FMP)

For the purpose of the current work, this platform, along with the Hazelnut Drying System (HDS), are the most important to analyze, since the data storage in the data base is entered through them.

This is an Enterprise Resource Planning System focused on the farms. It is designed to answer the needs of food companies in adopting a solution for management control, capable of organizing agricultural activities, follow their course and monitor the progress.

It is a powerful platform, with a geographic basis, that can be powered by external databases while avoiding double jeopardy of the data already available, such as human resources, machinery, and technical means. The platform is multi-user and multi-lingual; which means it can handle several farms of the same property with a working group that uses the same application in different languages and for different companies.

The structure of privileges allows the user to organize the information access of each ser of the Working Group according to its specific business functions.

Advantages

- maintain continuous monitoring of the activities and budget variance of the beginning of the year
- calculate real yields of field activities
- have an early estimate of resource requirements and monitor the actual use during the year
- support the technical staff through the processing of the recorded data subdivided by type of activity, for the plot, for the type of machines used

Reports

- Farm assets list
- Permanent and temporary worker list
- Budget in quantity report
- Input list (chemicals)
- Activity planning report
- Inputs needs and chemical requirement report
- Activity report
- HR worker reports (workers engaged in the daily activities)
- Harvesting report
- Batch traceability reports
- Management control reports

Figure 2. Farm Management Platform Logo. Available at: HDS Technical Manual

4.5. Hazelnut Drying System (HDS)

This platform is used for the process tracking in *Fruticola Agrichile*. It refers to an informative platform formed by a set of elements (applications of software and hardware) that are integrated together and perform an interface with the user for the support in data and operations management in the process of harvesting and first processing of hazelnuts.

As mentioned before, it is of crucial importance for the scope of the work, since it gives support to the management of data and operations characterizing the processing of hazelnuts. All the information is loaded in the database, which constitutes the central nucleus of the infrastructure, and it has been designed to take advantages of cloud technologies. The input, storage and processing of data is done in a centralized way, gathering up in an only point all the information asset, in a way that the security and availability of data is assured, from any part of the world where the user wants to access.

It is crucial to also be assured from data users mistakes, and that is precisely why the system has a data back-up in a testing area, where an exact copy of the data base is provided, reflecting not only the structure, tables, relational logic, but also the work environment (instances, data mass, etc.). In this order of ideas, it would be precisely this copy the one hat will be used for the development of the current work.

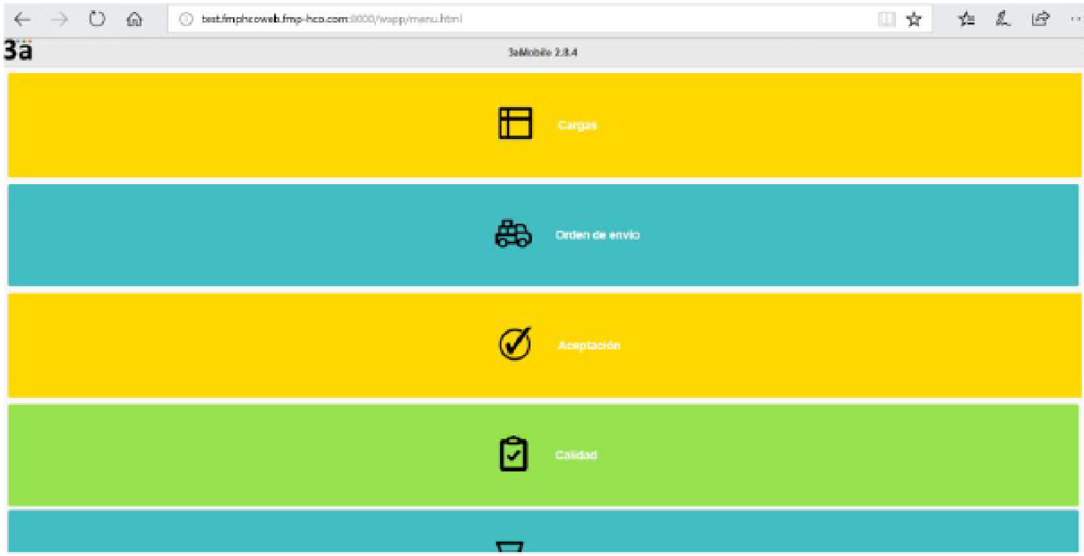


Figure 3. Homepage of the System. Available at 3a S.r.l Drying, 2018.

5. Frutícola AgriChile

In the year 1991 Ferrero arrives to Chile through this subsidiary company of its own, seeking to assure the production and supply of fresh hazelnuts throughout the whole year. For more than 25 years, AgriChile has worked with an innovative model for the industry, showing the high quality of its products, and capacitating external producers of the crops, letting Chile to become an important exporter of hazelnuts, being Italy, Brazil, and United States its main targets.

Currently, Agrichile is located in 3 (three) regions of the country: El Maule, Biobío, and La Araucanía, summing up a total of more than 3,500 Ha of plantations. *Figure 4* shows a geographical location of the regions mentioned above.

Figure 4. Map of Companies locations Available at: agrichile.cl/dónde-estamos

In addition, table no. gives some information about the plantations. This will be useful within the development of the work, since each one of them gives an input of data that will be used in the analysis.

Table 2 Regions and Plantations

Region	Plantation
Maule	Camarico, San Rafael, Los Niches, plus a Husking Plant
Bio Boi	San Gregorio, Cleaning and Drying plant
Araucania	Caracas, Cleaning and Drying plant

Available at: https://www.agrichile.cl/agrichile/

Furthermore, AgriChile is conscious of the importance of social responsibility, and that is why it has generated a positive impact in local communities, guarantying an increment in the labor force. In the same way, the company has developed crucial relationships of collaboration with different actors related to the agricultural industry, such as guilds, and public and private institutions.

On the other side, it is necessary to point out that the product, technically speaking, is a *European Hazelnut*, from the family of the Betulaceae, and *the Corylus* gender, whose crop takes place after the cleaning of the ground and the organization of the fruit. All of these, after the harvesting machine has done its part. In an overview, the Hazelnut process consists in a cleaning phase, drying, peeling, auditive selection, and with laser machines, calibration, and division in ranges. (Frutícola Agrichile, 2018).

6. Production Process

The set of operations that are analyzed throughout the development of the work are those corresponding to the activities in the farm. However, it is necessary to describe, in a general way, how the activities related to the harvesting, acceptance of the hazelnut in the drying plant, cleaning of the hazelnuts, drying, packing, and shipping of the product to the peeling plant, work. [2]

To be more precisely, the main objective is focused on the harvesting process, where some variables regarding the budget are specified. All of the activities mentioned above corresponds to the process in Chile and Georgia, analyzed in a meticulous way to create the information flow, which is explained in a way such that it follows each phase of the process, highlighting some points where the tracking of the product requires a more complex sequence of steps to be done properly.

6.1. Operations and Sub-operations

In addition, it must be clarified that for each operation, there is a lower category called "sub-operations", which corresponds to the activity that is done per se. As an instance, "Weeds control" will be a sub-operation, whereas "Weeds Controls" with Tractor number 5, with implement 3, with a determined duration, workers assigned, code, quantity, and unit of measure, is an operation itself.

Figure 5 represents the input and the output for the Operations related both to Farming, and to Harvesting. It is important to be pointed out, since the analysis of efficiencies and the algorithm are done within these two. Also, it needs to be highlighted that the operations can be divided into manual, and mechanical, as it will be explained later, and that for the later, just one person is registered as being in charge of performing it.

[2] The description of the production process was found in 3a S.r.l Drying, 2018. Here, a paraphrase is made throughout the section.

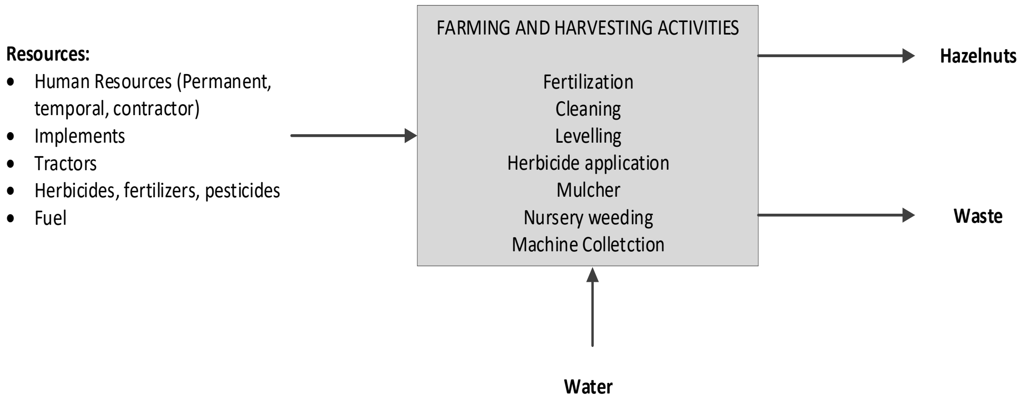


Figure 5. Resources and Activities. Source: Own elaboration.

6.2. Farming

A huge set of activities are required to be done before the plant is ready to be harvested. Those vary from the composition of the soil, and how it is set in a way such that, biologically and chemically speaking, it is suitable for holding the plants; to the processing of growing up the plants, starting from the nursery. Below, a description of the main activities regarding the farming is included.

Table 3. Farm Activities and its Operations.

Activity	**Main Operations**
Fertilization	Application of the fertilizer, transport of the fertilization, miscellaneous, supervising
Fencing	Fence installation and maintenance, fence removal, miscellaneous, supervising.
Soil Operation	Levelling, gathering and removal from field, primary soil operation, secondary soil operation, transport soil operation, supervising, miscellaneous.
Drainage	Cleaning pipes and channels, miscellaneous, supervising.

Harvesting	Collecting, miscellaneous, supervising.
Infrastructure	Cleaning pipes and channels, maintenance, installing pipes and channels, road preparation, supervising, transport infrastructure, miscellaneous.
Irrigation	Irrigation maintenance, flushing, miscellaneous, supervising, watering plants.
Plants management	Pruning, plants care, suckers removal, transport plants management, supervising.
Nursery operations	Planting, plant extraction, plant propagation, miscellaneous, supervising, nursery preparation.
Phytosanitary Treatments	Pesticide application, rodent control, pruning plants with diseases/insects, miscellaneous, supervising.
Weeds control	Cleaning between lines, herbicide application, nursery cleaning, cleaning around plants, cleaning across lines, miscellaneous, supervising.
Planting	Miscellaneous, supervising, transport planting, planting, replacement of trees.

Available at: Yagüe, 2018.

6.3. Harvesting

It could be done manually or mechanically in different modes depending on the place, whether it is in Georgia or in Chile. In the later, the manual picking takes place using containers that are then empty in a harvesting car (*Figure 6*). Right after that, the hazelnuts are gathered in some specific points of picking and are transported to the plant with camions.

Figure 6. Harvesting Car. Available at: HDS Technical Manual

The mechanical harvesting, on the other hand, is made through harvesters that deposit the product directly into the cars or into the camion, for its subsequent transport to the plant. The camion starts its path with a paper document (sending order), that contains all the information relative to the content of the transport. In some fields of Chile, the content of the cars is weighted before the transport.

When there are not enough camions to ship directly the product to the drying plant, the hazelnuts could be loaded into a temporary warehouse called "field warehouse".

6.4. Transportation

Once the product has been collected in the harvesting, it is gathered at a collection point, that acts like a buffer, until a vehicle comes to transport it to the drying plant. In doing so, several varieties of transport could be used:

- Carts: Pulled by tractors. It is used when the harvesting is performed in fields that are close to the drying plant. Carts could transport either jute bags with loose hazelnuts, or hazelnuts gathered from mechanized harvesting. An example of its configuration could be appreciated in *Figure 6.*

- Trucks: Used when the distance between the field and the drying plant is longer. Once the fruit has been collected from the field, it is taken into the trucks that carry the hazelnut either by jute bags or loose.
- External suppliers or Ferrero's group: All the process is done autonomously, but then the product is sent directly to the drying plant.

6.5. Processing in the Drying Plant

The next sequence of steps summarizes the sequence related to the processing of the product within the plant; from the arrival of reception lots up until the shipping. The different stages involve different operators and take place in different parts of the factory: Temporary warehouse, laboratory, plant, warehouses, and zones of elaboration.

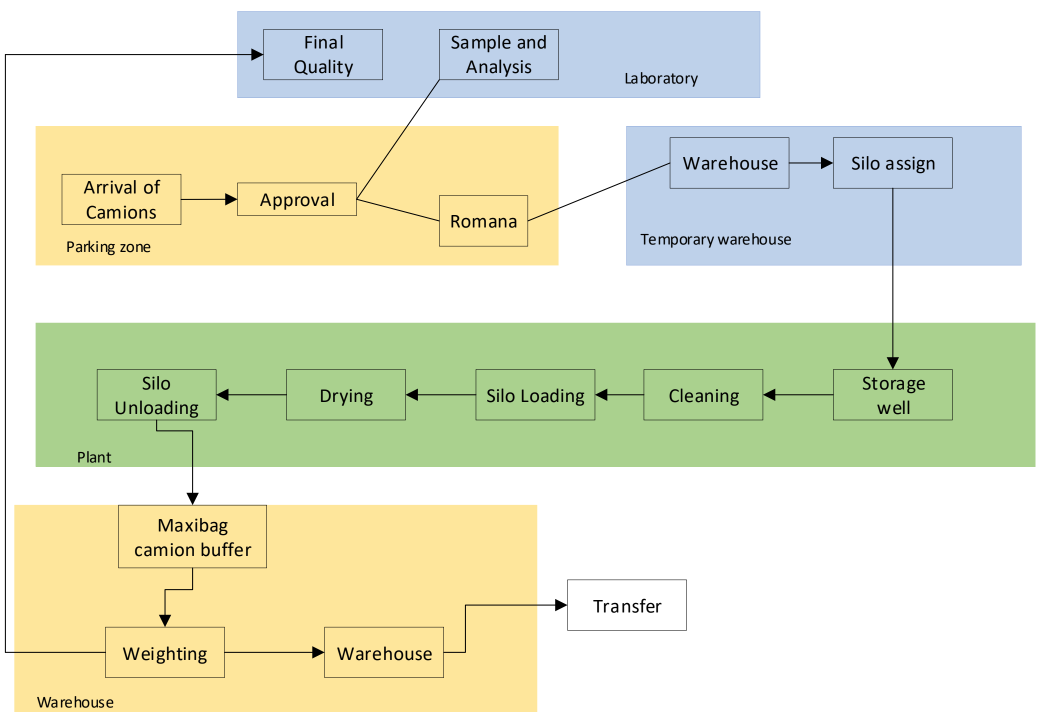


Figure 7. Flowchart of the Process Inside the Drying Plant. Available at: 3a S.r.l Drying, 2018. Adapted by author

6.6. Approval

The hazelnuts, once in the camion, are transported to the drying plant. The approval takes place here, where the lots must be weighted initially. The gross weight and the tare of the camion are identified in order to obtain the net weight. After the two weighting activities, the truck is unloaded. The information related to the weight is of vital importance to decide the next stages of product processing. During this stage, different cases of delivery are registered; as an instance, in bulk cargos and packings. Furthermore, if the picking takes place in a place close to the drying plant, the hazelnuts are transported directly in the cars, pulled by tractors.

In addition, the fruit could come directly from Ferrero's field or from external suppliers (delivering their product to the drying plants).

6.7. Quality Analysis

As soon as the hazelnuts arrives to the plant, the *detail* (group of hazelnuts corresponding to the same sending order) is submitted to a laboratory sample to obtain useful information for statistical means and to determine how to compose the cargo of the silo, in a way such that homogeneity is assured, from a qualitative point of view.

In Chile, the quality analysis also has another purpose: the presence of external suppliers requires a greater emphasis in the quality control of the incoming hazelnut. The fruit delivered from external producer needs to have some qualitative characteristics that determine whether they are accepted or not and its price.

Once all the lots have been identified, unloaded, and approved, they could be loaded directly into the lines or being stocked in a temporary warehouse, waiting to be loaded into the storage wells, which is done by a responsible. This last step is done because, usually, the input volume is higher than the plant capacity.

6.8. Wells/Cleaning Lines

The next step of the process is the one related to the load of the hazelnut into the cleaning lines. The workers perform this activity manually, putting the hazelnuts in the wells, according to their supervisor orders. Once the *detail* is unloaded into the hoppers that are connected to the cleaning lines, it is no longer possible to set a different silo to the product; thus, with the scope of guarantying the optimal tracking of the product and to create cleaning lots as homogeneous as possible by variety and/or quality of the hazelnut, it is preferable to previously determine the final silo and immediately, load the cleaning line, making sure of its availability and its correct connection to the right silo.

6.9. Drying

Done within the drying silo, where, with a vertical screw, the hazelnuts are mixed without interruption. A drying cycle is composed generally by 2 hours of "burning" with hot air injection coming from a boiler, intercalated with some hours of ventilation with cold air injection. The fact of alternating this allows the product to don't get stressed and to keep the hazelnut high-quality level. The number of cycles to which a lot is submitted depends on the initial humidity of the acceptation lots that contain it. The Drying lot is ready to its unloaded in the moment that the humidity within the hazelnut is less than 7% (this is determining through samples directly from the silo after each drying cycle).

6.10. Unload and Shipping

The silo content can be unloaded in Big Bags with a capacity of around 1 Tm, in camions directly for its transport, or the interior of silos designated to storage *(Figure 8).* As the last stage, there is the shipping, where the camion could be loaded with:

- Maxi-Bags: weighted before the load.

- Bulk Hazelnuts: Proceeding directly form the drying silo, or the silo designated to storage. Here, the camion is weighted before and after the load to compute the gross weight, the tare, and the net.

Figure 8. Storage Silos in Chile. Available at 3a S.r.l Drying, 2018.

6.11. Territorial Units

Within the operations performed, one could come to an idea of the importance of a territorial unit in the whole process: Each of this divisions requires more or less the same set of operations than the other, and yet, historically, some require special treatment, because of the characteristics of the land, its position, or how steep it is, among other reasons.

It is precisely this what gave the motivation to perform an analysis of efficiencies among different territorial units, but before going further, let's clarify some concepts about the matter.

As pointed out in the platform manual, the system allows to configure several units of territory among countries, according to its requirements.

Following the *Figure 9*, different divisions within the same territory could be commented, and explained deeply.

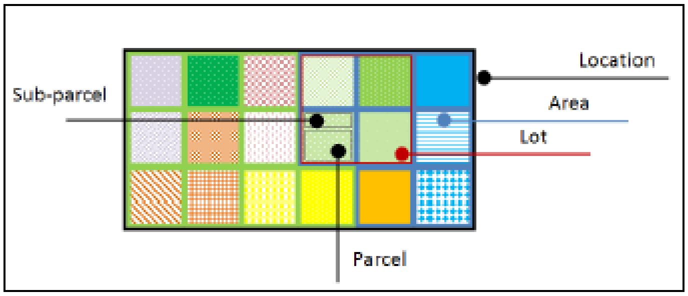


Figure 9. Territorial Units. Available at 3a S.r.l Farming.

- **Area:** represents a grouping of several Lots. In Chile, it is a group or lot powered by the same genset.
- **Lot:** represents a grouping of several parcels according with the irrigation block.
- **Parcel:** Corresponds to the local administrative unit and should be characterized by different varieties, plantation years and spacing.
- **Sub-parcel:** is the smallest for all companies, it is uniform in variety, year of plantation, and spacing.

With this hierarchical relation among subdivisions, it is logic to assign a level, for each territorial unit, in the subsequent way:

Table 4. Territorial Units Levels

Level no.	Territorial Unit
4	Sub parcel
3	Parcel
2	Lot
1	Area

Available at 3a S.r.l Farming.

7. General Description of the Information Flow

The manuals of the company provide a clear explanation of how the information flows throughout the process. This resource is used throughout this section, paraphrasing its content: (3a S.r.l Drying, 2018).

The development of the information system counts with 3 (three) essential elements, named, actors, entities, and attributes.

The *actors* perform a role represented by a certain set of activities that interact with the system (including users, other software systems, hardware devices, etc.). A paper corresponds to a certain family of interactions related that the actor undertakes in the system. It does not refer to a unique user, but to the role or the operations/tasks developed and to all the sets of information related to it. As a consequence, the actors are the responsible of the sectors seen in the description of the process: the responsible of the approval assigns the weight to the reception lots; the responsible of the lab assigns the results of the quality analysis of the product to the reception lots or to the final product; the responsible of the plant determines the drying lot of the hazelnuts, the responsible of the picking crews introduces the information into the platform.

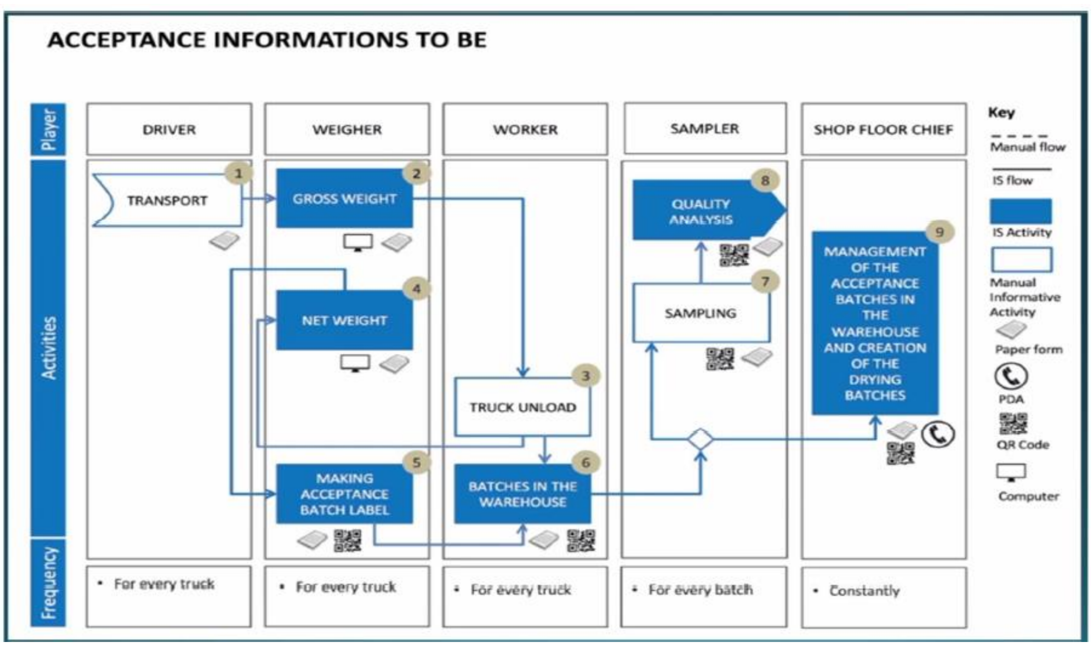


Figure 10. Acceptance Information to Be. Available at (3a S.r.l Drying, 2018).

Establishing actors, entities and attributes allows to put the information flow into a compartment. In the scheme, an actor who interacts with the system introducing, manipulating, and deleting information through different types of devices, corresponds to each column (or swim lane). Another fundamental node of the analysis is the study related to the tracking of the product, for which a specific analysis is done.

7.1. Tracking and Location

Tracking means the possibility to find and follow the path of a material or object in all of the production, transformation, and distribution stages. The concept of tracking is strictly related with the procedures of location. It is, in fact, the set of procedures of registration, transfer, and analysis of the information related with the product, its displacements and its different transformations.

Figure 11. Tracking Product's Flow. Available at (3a S.r.l Drying, 2018).

The operators, as an instance, continuously register information that allows to track the flow of the hazelnut. In the field are registered, for example, the origin, the territorial unit where they were picked, the variety, the day and the kind of picking, the container, and the transport mean (car or camion). In the next steps, more details concerning the characteristics of the product, the containers of transportation, and the procedures followed for its processing, up until the unloading of the finished product, are introduced in the data base. In the sending stage, the platform prepares documents and reports summarizing the records of the content of the sending lot. The procedures of tracking mentioned above, if they are carefully backed up by an information system well studied, allow to locate the product at the end of the process.

7.2. Detail

Each transport is an entity defined as sending order, and it contains a variable number of loads known as collecting unit. More concretely, the load could represent a unique jut bag for the collecting in Georgia, or a picking car for the case of Chile. For logistic and tracking reasons, within the same transport, an aggregation of loads is done based on some criteria, such as uniformity of variety, which leads to the creation of details.

In this way, a sending order, for example, can contain in its interior a variety of jug bags and, for instance, a variety of loads, and be divided in 2 (two) or more lots of acceptation as a function of this criteria.

On the other hand, the criteria for separation of the transport into details in Georgia are *variety, day of picking, mode of picking and territorial unit*. In Chile, only the first 3 (three) are taken into account. After the definition of the loads and the creation of a sending order, the system can create automatically, in function of the required criteria, the reception lots.

7.3. Load of the Lines

In the stage of load of the lines, the normal operational procedure requires that the individual acceptation lot goes through a unique line in order to end up in a unique silo. The procedure, however, could have its variants. It is possible that 2 (two) or more lots of acceptation are introduced into the lines simultaneously, in a way such that the hazelnuts ended up mixed. Besides, it could happen that an acceptation lot is divided and loaded into 2 (two) different lines or that the content of the line after the cleaning makes it to end up in 2 (tow) different silos. All of these variants make the record of the acceptation lot and its tracking to cross with a different acceptation lot.

7.4. Unload of the Silos

In the case that the unload is done in the big bags or directly into a camion, the tracking of the product is guaranteed. The content of the big bag keeps the same record of the drying lot and it is identifiable with precision. However, if the unload is done inside a storage silo, the content of the drying lots will be mixed, and for that reason, it will be impossible to surely know the record of what is inside the storage silo at that moment.

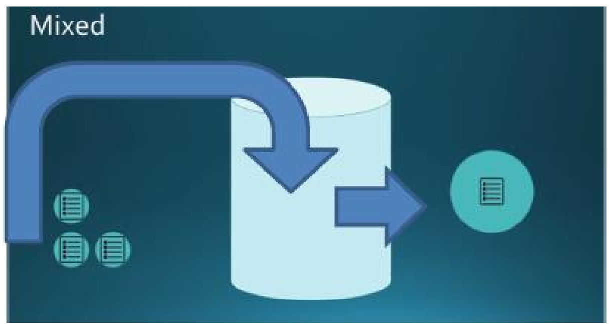


Figure 12. Tracking Product Flow, different acceptance lots loaded are transformed into one identity . Available at (3a S.r.l Drying, 2018).

While the hazelnuts are in separate containers and they do not mix, the tracking of the product corresponds with that of its acceptation. Introducing a lot of jut bags in the interior of the same sending order is not, for instance, a problem for the tracking, because the information related to each element of the Sending order stays in the system and it is recoverable. In this sense, if in the warehouse someone scans a jut bag, it is possible to surely know the record of the product.

The situation changes in the moment that the non-elaborated hazelnuts are mixed in the interior of a line, of a silo, or in a storage silo. In this case, the records of this components are put together to create the collective record and the tracking of each element is defined more clearly.

8. Activity

The activities performed in the field cold be divided into 2 (two) major categories, named Manual and Mechanical Operations. Both differ in the usage or lack of usage of a tractor, and in the number of workers that are intended to complete each operation, being always one for the Mechanical ones. On the other hand, it is precisely the information contained within this section what supports the creation of an algorithm that pursues the third objective of this work. Bearing in mind this, it is extremely crucial to understand the logic in which the data is inserted.

8.1. Manual Operations

The user first needs to set the operation to register the Date, the sub-operation, and the level in which this one is done. Besides, some other fields are taken into account, among which, there are:

- Weather: the default is “sunny”. However, it could be selected from the list.
- Form Number: Optional data that is only active in case the information of the field operations are registered on forms.
- Territorial Units: the TU where the work is/was done.
- Variety and year: both of them will be automatically set if there is just one variety per year; if not, and if the user does not select anything, the system will consider all varieties and all years;
- Workers: one or more workers, or a team.
- Quantity done person: The user has to introduce the quantity done per each worker; nonetheless, it could be also done in the agronomic detail.
- Agronomic detail: Amount of product applied in the operation. If it is a diluted treatment, the user will have to insert the total applied volume and the system will compute in the detail the real sprayed volume per hectare. If not (prescription in kg/ha), the user must add the total amount (kg-L) applied for each product, and the system will calculate the real dose in kg/ha for each product. As final information, the area covered needs to be inserted as well, filling the Qty (Ha) column.

- Number of hours spent to do this operation. If more than one person is selected, the system will attribute to each worker the number of hours.
- Implement: type and the number of the implement used.
- Fuel: the total amount used for small implement motor eventually employed in the sub-operation.

(3a S.r.l Farming, 2018)

8.2. Mechanical Operations

The logic of data insert is alike. However, some differences are pointed out in the process.

- Form Number: Optional data, active only in case the information of the field operations are registered on forms.
- Variety: both of them will be automatically set if there is just one variety per year; if not, and if the user does not select anything, the system will consider all varieties and all years.
- Number of passes: where the number of times the operation is repeated within the same TU is specified.
- Number of hours: the time the operator spends for this operation.
- Implement: All the implements needed.
- UM: This should be the value indicated by the odometer, the default value is “hour”.
- Start-End: Value of the odometer when the user started the operation, and when he/she finished it.
- Fuel (L): daily fuel consumption.

(3a S.r.l Farming, 2018)

9. Languages and tools used

Through the work, several tools and languages were used to reach the goals defined. Some of them are languages with interfaces well characterized, whereas others are applications with graphical support, being more user-friendly. Whatever their nature, they were indispensables for the analysis, and that is precisely why this section has the objective of giving a brief explanation of their concepts.

DAX

States for Data Analysis Expressions, and it is a collection of functions, operators, and constants that are used in a formula or expression, to calculate and return one or more values. It helps create new information from data already defined in a model. (Docs.microsoft.com [DAX], 2018).

PowerBI

As explained in (Docs.microsoft.com [PowerBI], 2018)., It is a collection of software services, apps, and connectors that work together to turn unrelated sources of data into coherent, visually immersive, and interactive insights. The application supports a wide range of data source, from Excel spreadsheets to a collection of cloud-based and on-premises hybrid data warehouses, letting connect this data sources, visualize, discover, and share with anyone.

Although simple and fast, it is also robust and enterprise-grade, allowing real-time analytics and custom development. In other words, it is suitable for being a personal report and visualization tool and can also serve as analytical and decision engine behind group projects, divisions, or corporations.

Microsoft SQL Server

According to (SearchSQLServer,2018), Microsoft SQL Server is a relational database management system (RDBMS) that supports a wide variety of transaction processing, business intelligence and analytics applications in corporate IT environments. Built on top of SQL, it is tied to Transact-SQL (T-SQL), an implementation of SQL from Microsoft that adds a set of proprietary programming extensions to the standard language.

Visual Basic

It is defined, by Christensson (2007), as a programming language and development environment created by Microsoft and being an extension of the BASIC programming language that combines BASIC functions and commands with visual controls. It provides a graphical user interface GUI that allows the developer to drag and drop objects into the program as well as manually write program code.

It is intended to make software development easy and efficient, while still being powerful enough to create advanced programs. In addition, the Visual Basic program also includes features like "IntelliSense" and "Code Snippets", which automatically generate code for visual objects added by the programmer.

RStudio

As stated in RStudio, 2018, it is an integrated development environment (IDE) for R. It includes a console, syntax-highlighting editor that supports direct code execution, as well as tools for plotting, history, debugging and workspace management. It is available in open source and commercial editions and runs on the desktop (Windows, Mac, and Linux), or in a browser connected to RStudio Server or RStudio Server pro (Debian/Ubuntu, RedHat/CentOs, and SUSE Linux).

10. Efficiency Estimation

Up until this point, 3a S.r.l has worked along Agrichile with two core platforms, named Farm Management Web Platform (FMP) and Hazelnut Drying System (HDS). Both of their operations started from 2015, and data has been loaded ever since. Nonetheless, there has not been any analysis performed so far.

Bearing in mind this current condition, the author has proposed a method to evaluate what they currently input to the forecasts as efficiency, which is an information known *a priori* for each sub-operation, using a specific set of resources. The goal is then to contrast what is being used at the moment (explained right after, under the section of "Budget Estimation") with feasible computation of the efficiency, done through these two general formulas:

- $Manual\ \ Efficiency = \frac{sum_{qta}(manual)}{hours\ x\ number\ of\ workers}\ x\ 9\ or\ 8$
- $Mechanical\ \ Efficiency = \frac{sum_{qta}(mechanic)}{end\ time - start\ time} x\ 9\ or\ 8$

The above will have the same units of measure as the current input. Throughout this chapter, the process of obtaining the respectively values, starting from the point of obtaining the data, will be explain. But first, let's define how the estimation of the budget is performed.

10.1.Budget Estimation

The importance of the Budget estimation is clear: it is the basis for defining what needs to be done in a certain period, and, as so, this was precisely what gave rise to the conception of the present work.

This specific functionality of the Platform is found under the Management Control section in FMP, and it is thought as one of the most useful tools, since it is an instrument to define targets and goals, and to program the monthly needs in term of human resources, machines, and products.

The input for the budget comes from the completion of two lists: the prescription list (products and their doses) and the efficiencies register. After done, the user can also create a new budget, or modify an existing one, fill it with tasks, schedule them, and compare the programmed activities with the work actually done.

Manually, the budget is filled with tasks, where each task corresponds to a sub-operation done within a specific territorial unit level (area, lot, Parcel, or sub-Parcel). As soon as the data is specified, some automatic calculated data, along with other editable one appears in the interface. When it happens, the user has to select the Operations to perform, and the prescription (if any) that corresponds to it. In addition, some other information about where to perform the task, with what efficiency, how many resources and machines the task needs, has to be entered as well.

For the next step, the specific territorial unit within the company needs to be selected. The default is that all the territorial units appear to be marked. After this, it is needed to associate the efficiency field with the operation. However, if the operation efficiency is not measured in terms of hectares or number of plants, the user needs to complete the information about how much work this task is planning.

As the last step to be performed, information about workers and machines has to be completed (3a S.r.l Farming, 2018):

- Nº P: number of permanent workers.
- No T: number of temporary workers.
- No C: number of contractor workers.
- HRi [no]: the number of workers that help during the operation but are not involved in it. For this reason, the number of HRi workers doesn't change the number of days needed to complete the task.
- HR Tot: automatically calculated (sum of the Permanent with the temporary and the contractors).
- Qty: automatically calculated if the efficiency is measured in hectare or number of Plants. Otherwise, it is manually set.
- No Pass: number of passage in this area, doing the operation.
- Tot. Machines: number of machines allocated for this operation.

On the other hand, the following values are automatically completed by the platform:

- Tot. machine/days: Total days multiplied by the number of machines.
- Tot. man/days: Total days multiplied by the number of HR Tot.
- Tot. P: Total days multiplied by No Permanent workers.

- Tot T: Total days multiplied by No Temporary workers.
- Tot C: Total days multiplied by No Contractors.
- Total days: Depends of the operation. If mechanical operation (Quantity divided by Efficiency), divided by the number of machines allocated. If manual operations (Quantity divided by efficiency), divided by the number of men allocated

In the following figure, an illustration of the process described above can be appreciated. It is clear that the input is both, the prescription, and the efficiency, being the former a function of the kind of activity, the type (if manual or mechanic), the human resources, both helping and required for the operation, and the consumption (measured in fuel per day).

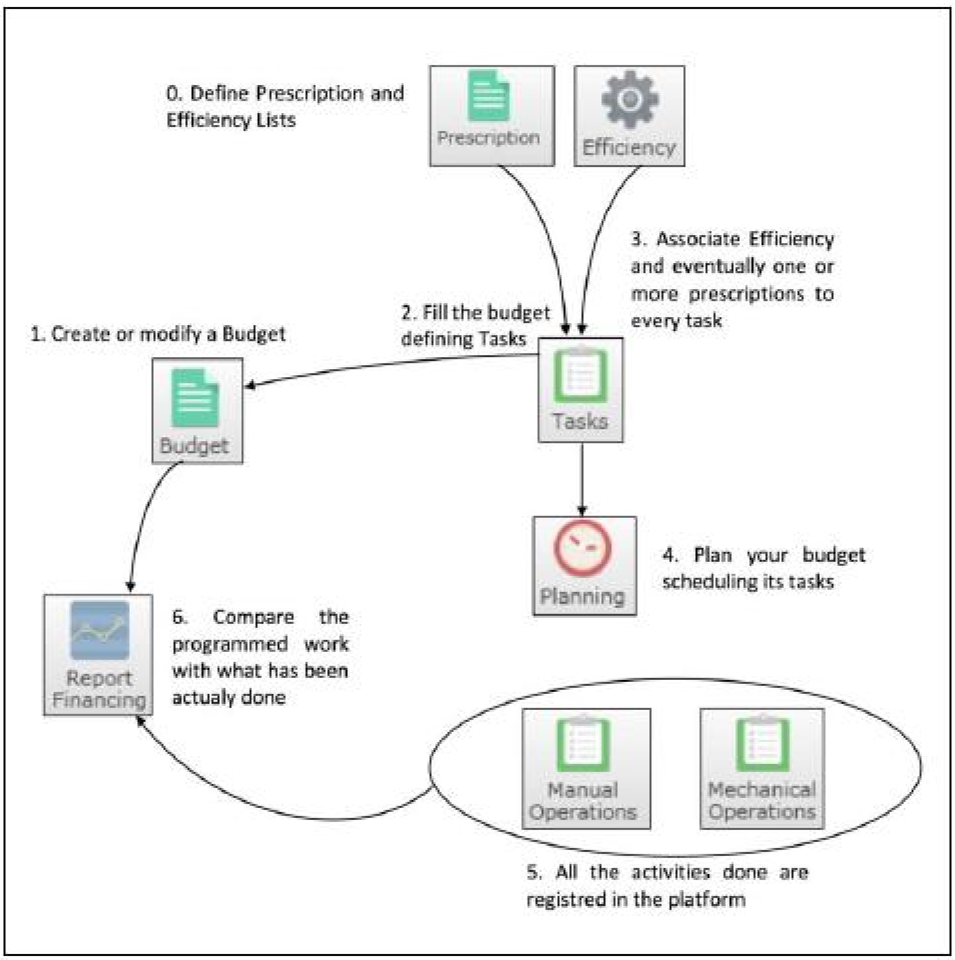


Figure 13. Management Control. Available at 3a S.r.l Farming, 2018.

Then, the tasks are associated to the Budget in matter, allowing them to be flexible and comparable throughout the time (In the sense of the planning, and in the status of the work done with respect to that scheduled).

Afterwards, the Manual and Mechanical operations are registered in the platform. The process is then repeated for other instances in the same way it has been explained in this subsection.

10.2. General Concepts [3]

With a clear view of how the scope of this chapter is related to the Budget Estimation, and how this will make a contribution for more accurate values, it is then important to clarify some concepts that are used throughout the reach of the enterprise.

Starting from the basis that this is a process of decision making, it is intuited that the definition of Business Intelligence needs to come out. It was first coined in 1958 by Hans Peter Luhn in an article for an IBM publication (as stated by van der Lans, 2012), where it was defined as “the ability to apprehend the interrelationships of presented facts in such a way as to guide action toward a desired goal” .

It is to be noted from the above definition that the concept establishes a way of supporting the process of making decisions, and the decision making itself. In addition, it must also be said that there are systems that let this process to be carried on in an effective way, known as business intelligence systems. Those could be divided into 2 (two) categories*: reporting tools*, and *analytical tools*. The former allow users to study, filter, aggregate, summarize data and so on, presenting what has happened in the organization to the users. The latter, on the other hand, is based on statistics, data mining, and operations research, supporting algorithms for forecasting, predictive analysis, and optimization.

It is obvious that, in order to operate, the analytical tools need data as input. Although, there are many types of Data Stores (named Data warehouse, Data mart, Data staging area, Operational data store, and Personal data store), the one used in the current work was Data

[3] The general ideas were obtained from van der Lans, 2012. For some parts, paraphrasing from the main article was used.

warehouse, since it deals with the problem of not accessing the database directly to create reports, solving some of the following potential problems:

- Data integration: data needed by reports might be stored in multiple production databases.
- Defective data: Data stored in production systems might be faulty or missing. Reports written directly from them will be handling with defective data.
- Data consistency: Different integration logic with different reporting tools.
- Historical data: Not all production databases keep track of history, and for certain reports and forms of analytics, historical data is needed.
- Interference. Queries executed by the reporting tools on the production databases might cause too much interference on the production systems.
- Query performance: Running complex queries directly on the database can lead to very poor performance, making the users wait hours before seeing results.
- External data: If data is taken directly from the database, without creating a data warehouse, there will be no room for analyzing external data.

Formally, a data warehouse is then a separate data store designed specifically for reporting and analytics. Here, data is periodically copied from production databases to the warehouse, implementing and ETL process (it will be explained latter), meaning that the data warehouse is constantly being refreshed.

Moreover, the concept of Data scheme takes relevance in the matter. Despite the fact that there are many in the literature; the 4 (four) most well-known are the Normalized, Denormalized, Star, and Snowflake schema.

Normalized Schema

Here, columns are assigned to tables in such a way that each business fact is stored only one, i.e. the table should not contain duplicate data. The goal is then to avoid storage of duplicate data, so that stored data can't become inconsistent, being highly suitable for supporting transactions in which data is inserted, updated, and deleted. The reason is simple: because there is no duplicate data, each insert, update, and delete involves only one row.

Denormalized Schema

It is the opposite from the Normalized ones. When tables do not adhere to the former rules and contain duplicate data, they have a denormalized schema.

Star Schemas

This kind of schema owes its name to its graphical representation, where the tables within it, classified as *fact* and *dimension* tables, are set in such a way that the fact table forms the center, and the dimension tables are drawn as rays originating from the center, together forming a star. Every dimension table features a primary key of one column and a set of attributes describing the dimension, each row within it represents some business object, whereas all the nonprimary key columns of a dimension table contain data describing that business object.

Also, dimension tables don't have relationships between each other, only relationships with fact tables, which contain the primary key that consists of all the primary keys of the dimension tables it refers to .

In addition, a row in a fact table represents a business event; and because of the particular primary key structure, the relationship of a fact table with each dimension table is always ONE-to-MANY.

Another important thing to point out is that the key columns in dimension and fact tables of star schemas are filled with *surrogate key values*, having no meaning to the business users. In most cases these are just plain numbers, that are used to obtain never-changing key values that forever represent business objects (dimensions) or business events (facts). These key values used to identify the objects and events are necessary because there have to be some values that never change.

Almost every star schema contains a date dimension, capturing some business event, that took place somewhere in time; being the moment in which the event occurred a necessary part of the description.

The primary goal of arranging tables as a star schema is to limit the number of tables that have to be accessed and joined when a query is processed. Another advantage is that it becomes much easier to write queries and to present the end user with a set of options from which a tool can generate a query. The fact that duplicate data can lead to inconsistent data is not considered a significant disadvantage either, which makes sense in a data warehouse environment where all the inserts and updates are executed in a very controlled fashion.

A datastore can contain many fact tables and thus many star schemas. If fact tables share the same dimension tables, these dimension tables are called *conformed dimension tables.*

Snowflake Schemas

Like star schemas, both organize the tables around a central fact table and use surrogate keys. Although, dimension tables in a snowflake schema are normalized. A fact table in a snowflake schema has only relationships with dimension tables, just as in a star schema. Dimension tables, on the other hand, can have relationships with one another, and the relationships that exist are all ONE-to-MANY relationships.

Also, the dimension tables form a hierarchy, meaning that some will have lower granularity level than others. The advantage of a snowflake is that is fewer duplicate data is stored than in an equivalent star schema; furthermore, a snowflake schema can support queries on the dimension tables on a lower granularity level.

Data Transformation

Getting the data from production databases to various data stores implies copying the data, taking into account that the format and contents of the data stored is quite different from how users want to see it in their reporting and analytical tools. As an instance, in production systems, customer data might be spread out over multiple databases, while users want to have an integrated view; data in the production systems might be heavily coded, while users want to see meaningful values; historical data might be missing from the source systems, while users need it for trend analysis; or even there might me defective data, when users want to work with correct one. All this process is what is commonly known as *Data Transformation.* Even though there are 3 (three) well known methods to do so (ETL, ELT, and Replication), the one used during the development of the thesis has been the ETL method.

Extract Transform Load

ETL process is one of the most popular methods for this purpose; here, data is retrieved from one or more source data stores. The data is then transformed, cleansed, and integrated, and finally stored in a target data store, varying in simple process, such as the concatenation of two strings, to highly complex ones, such as the deduplication of rows. Throughout this chapter, the author is going to explain the process deeply to illustration, and step by step solutions.

10.3. Database Description

As stated in the company's manual (3a S.r.l Drying, 2018): The DB is the central nucleus of the infrastructure, allowing the user interacts with the platform through the applications. It has been designed in a way such that it is possible to take advantages of cloud technology, gathering, archiving, and processing the data in a centralized way, where all the information asset is contained within one point, guaranteeing security and full availability of data from everywhere.

The Cloud Computing service is in charge of Amazon (Amazon Web Service- AWS), more precisely through the kind commonly known as IaaS (Infrastructure as a Service). In addition, it is important to point out that there is a backup of the DB made every certain time, established previously, to assure more safety.

Particularly, the current thesis dealt with a database called "DBMeteo", supported by Windows Server 2008 R2 and with the characteristic of having an exact copy, where developers and responsible of testing can have an environment to perform any kind of activity without risk of damaging official data.

This DB is a relational database with a total of 210 queries, storing data about Farming process and first elaboration of the hazelnut, responding to SQL logic, and being supported by a Relational Database Management System known as Microsoft SQL server.

Furthermore, because of its nature, it responds to the standard ISO/IEC 9075-1:2016, where the conceptual framework is described to specify the grammar of SQL and the result of processing statements in that language by an SQL-implementation is defined. (ISO 9075-1:2016, 2018).

11. ETL: The step by step of the process

As mentioned before, under the subsection *Description of the Process*, a particular part deals with efficiency, and it is the one taking place within the farms. Right now, it is done with the expertise of the workers, meaning that for each operation, a set of resources is used, and an estimation on **hectares** or **number of plants per day is done**, which is the input for the computation of the budget. Within the database, it is displayed based on tractor or implement family, and operation. So, if a particular sub operation, 1st hand collection as an instance, is done with a particular family of implements, it will have an efficiency beforehand.

Although it has been useful so far, the client has explicitly expressed its concern about this part being crucial for its forecasts, and, bearing in mind the fact that data has been historically loaded into the data base (in terms of production and resources used), obtaining the real efficiency is possible, and it is, in fact, considered of critical importance at the moment.

For this purpose, the database has been submitted to a process of homologation, so data was correctly organized to deploy the information needed, all of this through a set of steps described below.

Generally speaking, the ETL process could be summarized through the following logical sequence applied to data.

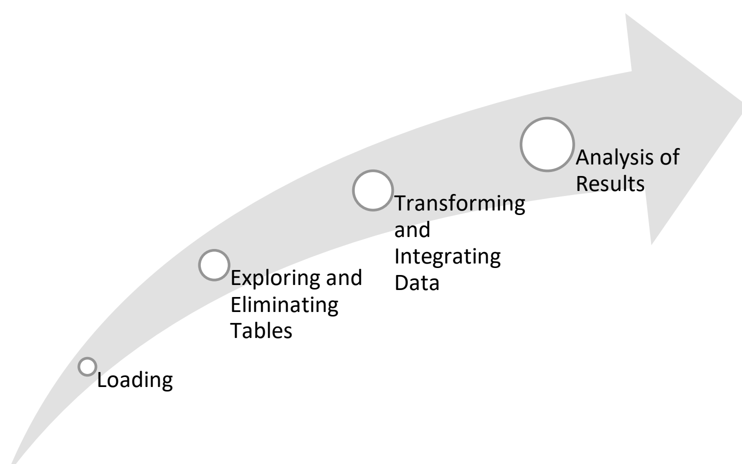


Figure 14. General Steps for the ETL process. Source: Own Elaboration.

11.1.Loading the Data.

Even though there are multiple ways of loading data into Power BI, because of the server in which 3a S.r.l keeps it, the option Database SQL server was chosen to correctly load the data. Credentials were provided by the enterprise, and the access to the servers was possible just within the office.

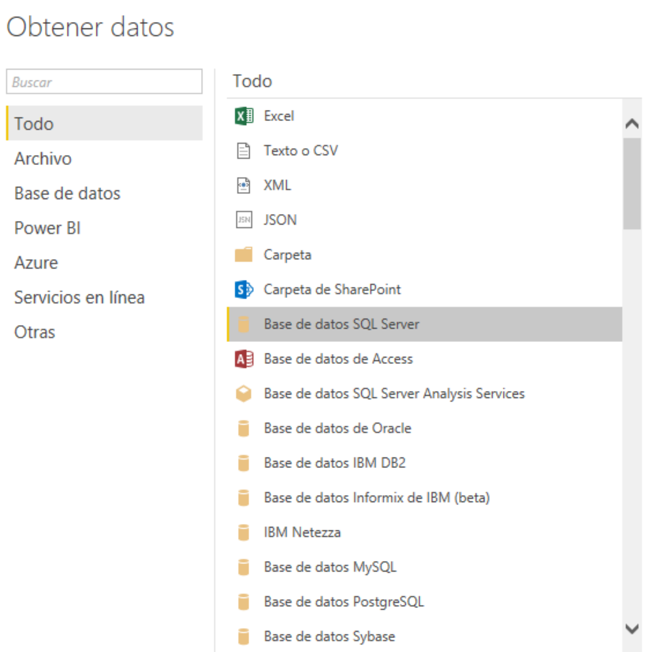


Figure 15. Power BI Data Load Options. Source: Screenshot of the Process

In addition, the queries were imported directly from the test area, so that, the model created will be automatically and constantly being fed of real time changes in the server, and, in the same way, any change will not have an impact in the real database.

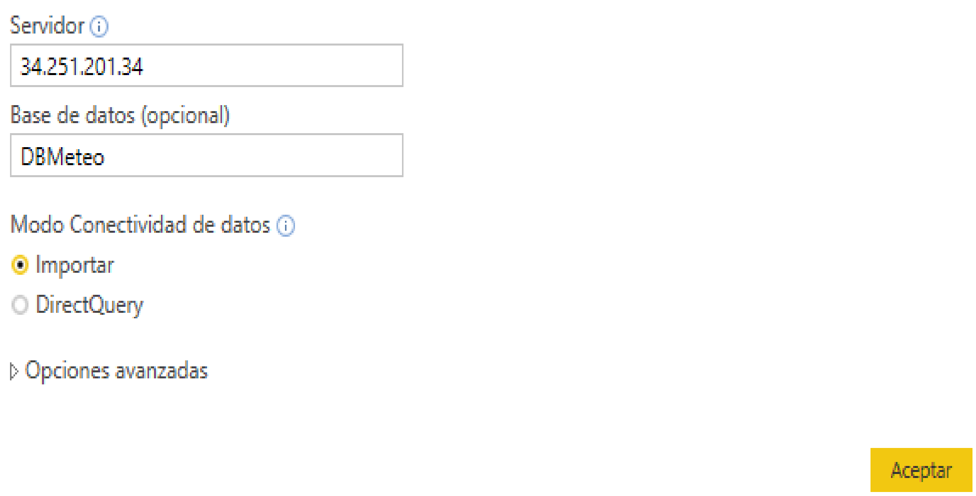


Figure 16. Login to the Server and DB. Source: Power BI Screenshot

With this first approach, the author had a clearer idea of the divisions of the tables within the database, and where to find the information needed. However, the great number of queries represented a problem in terms of relations for the model. In addition, not all of them were needed for the computation of the real efficiency. *Figure 17* gives a picture of the cubic relation between queries that took place as soon as all the tables were downloaded from the database.

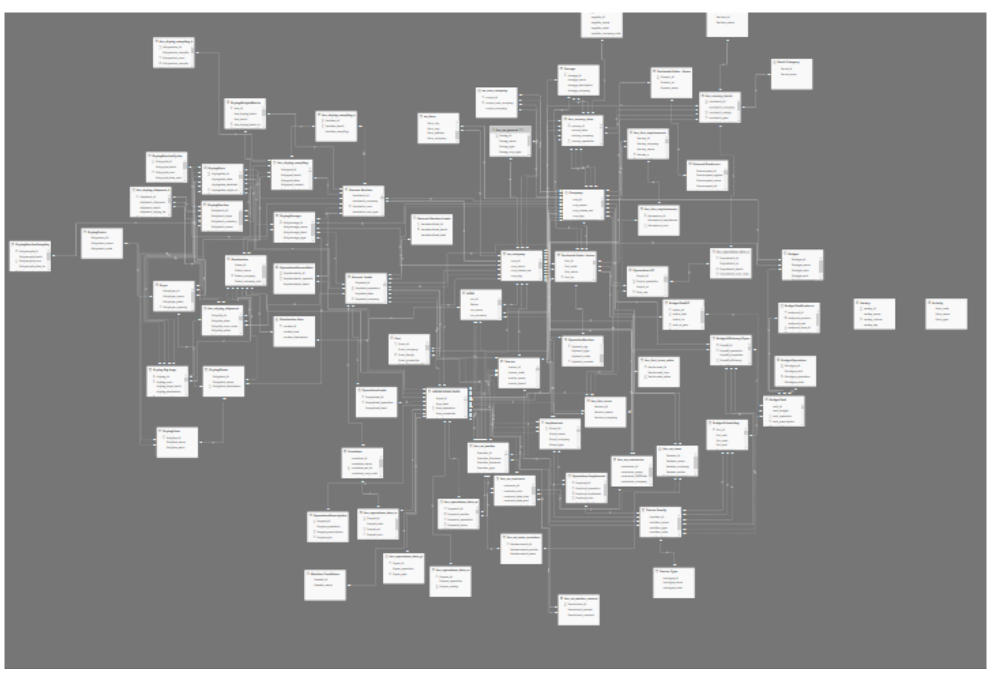


Figure 17. Initial Queries of the Model. Source: Power BI Screenshot

11.2.Exploring and Eliminating Tables.

Once the database was loaded, a filtering of tables were made . This means, tables that were not giving useful information to the analysis, were deleted. With less tables to go through, some groups were created based on type of activity or process of the tables' information, as shown in *Figure 18*.

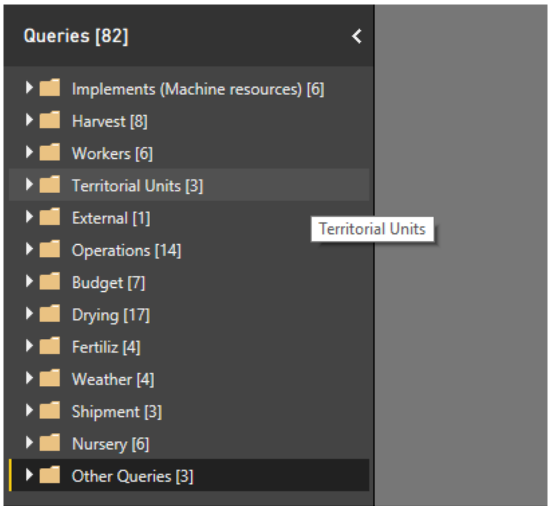


Figure 18. Grouping by Category. Source: Power BI Screenshot

11.3.Transforming and Integrating Data.

As it could be seen in *Figure 17* ,multiple relations, as well as multiple paths exist between queries, so in this order of ideas, the first step was to clearly understand what were the main tables to work with for the objective; and after a deep analysis, the author came to the conclusion that, from all the existing queries, there were needed just 7 (seven) of them with a proper setting, for the correct computation of the real efficiencies. These tables are numerated below with an explanation of their content. In addition, for the child queries, there are figures showing the column that relates them to the parent table (Operations Data).

11.3.1. Set of Queries for the Analysis.

The procedure for establishment the relations through the tables explained above was to initially delete all the links between queries, and then select, with and understanding of the data, how to connect the tables with the foreign and the primary keys.

Operations Data

Downloaded as ferr_operations data. This is the main table, which contains the primary keys to relate its content with the other tables. Here, a big quantity of information about the harvesting process could be found; however, what was interesting for the objective, was the information about the quantities in hectares or plants, the worker who did the activity, what machine did he/she used and with what implement (if any), the number of hours it took to perform that operation, in what company did it take place (see *Frutícola Agrichile* section for the company names), in what territorial unit, and what was the ID of the operation.

As said before, some settings had to be done to be able to work with this query. Those were:

- Deploy the implements and the machines that were used to perform the activity. When the table was downloaded, each operation ID was associated with the ID of the tractor and the implement, both at first in different tables. However, the ID was duplicated, and the foreign key was not appropriated for a correct search, so a relation for code of the implement or machine was created, which assured its uniqueness.
    - The approach had an additional problem though, and it was that for some machines (i.e. tractors), there was no code associated, so, whenever a search will take place for a null value, the program would associate the code with the first null value that appeared. To solve this issue, the author filtered the null tractors, and personalized a column through the program, so that, when a null case occurred, the name of the tractor would be displayed on the summary table (this concept will be explained later).
- Filter data for the years 2015, 2016, 2017 and 2018.

- Analyze that the column "quantity" was variable across the years, and as so, it could not be used as an input to estimate the real efficiency. After some simulations with Power BI, and pointing out that before the year 2017, the conclusion reached was that the quantities for manual and mechanical operations were separated for the harvesting and for the other activities. Besides, the data for mechanical operations was not recorded for any year.

  To deal with the problem, an extensive research through different tables and some interviews with the personnel of 3a lead to the solution of finding a unique column where the quantities were registered in a systematic and unique way for every type of operation, manual or mechanical, and for the years in concern. It is to say, however, that NULL values for the quantity were not taken into account. This column is named frop_sum_qta.
- Back to the argument of implements and tractors; how the efficiency is currently displayed in the database depends on the ID of the sub operation, and in separate columns if it was done with one implement or two implements, and with a specific tractor. In spite of this, the data is deployed in a different way, historically, in the tracking of the process; all of the implements were just loaded into one column. As a solution, a list of all the group of implements and tractors used for each operation was created, to later separate that list into columns, so that the information was alike, and comparisons between historical data to calculate efficiencies, and current efficiencies (based on expertise) could be done.
- The column quarter was extracted from the date. This to take into account the fact that for the quarters related to autumn and summer, the working time per day is 9 (nine) hours, instead of the usual 8 (eight) hours.
- Columns for date, hour, minute, and second for each operation ID were deployed, so that potential duplicate of information could be found and taken care of.

Machine

Downloaded as ferr_tractor. It contains the information associating the code of the tractor or implement, to the family, which is used to compare the efficiency. The idea, and how this table was handled, was to have all the machines (i.e. tractors and implements) in just one table, so that the number of queries in the model were reduced. Operations data and this one, have a MANY to ONE relation, which means that many operations could have one machine.

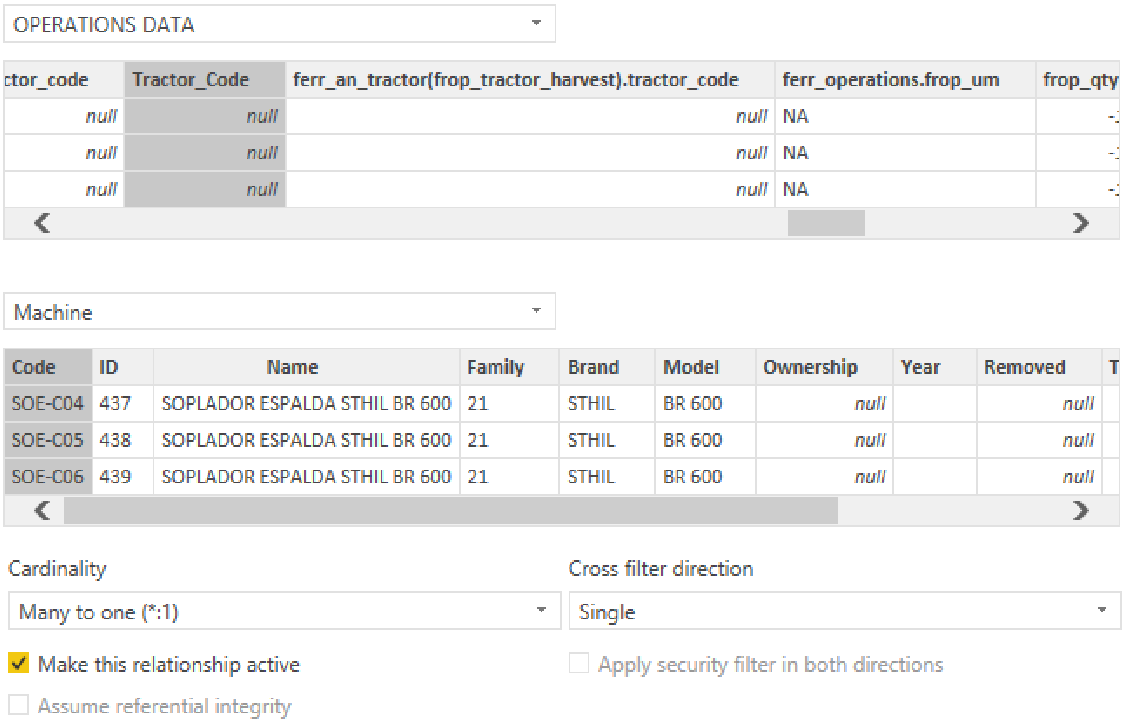


Figure 19. Relationship Between Operations Data and Tractor Queries. Source: Power BI Screenshot

Machine Family

Relates the code of the tractor or the machine to the family. The first approach to deal with this was to relate Operations Data with machine, and then Machine to Machine Family; however, a simpler way to link this table with machine Family was to deploy the code of the machine in Operations Data, and then connect it directly with that information with this table.

OPERATIONS DATA

code	FamilyImplement1	QuantityImp1	FamilyImplement2	QuantityImp2	FamilyName-Imp1	FamilyNa
null	null	null	null	null	null	
null	null	null	null	null	null	
null	null	null	null	null	null	

Machine Family

tractfam_id	tractfam_name	tractfam_type	tractfam_code	tractfam_fuel	tracttype_code
1	Tractor - general	1	TRACTOR	DIESEL	SELF_PROPELLED
2	Car	1	VEHICLE	DIESEL	SELF_PROPELLED
3	Selfpropelled Harvester	1	HARVEST	DIESEL	SELF_PROPELLED

Cardinality: Many to one (*:1)

Cross filter direction: Single

☑ Make this relationship active

☐ Apply security filter in both directions

☐ Assume referential integrity

Figure 20. Relation between tractor and Implement Family 1. Source: Power BI Screenshot

OPERATIONS DATA

ment-Mechanical-ID	Implement-Mechanical-code	FamilyImplement1	QuantityImp1	FamilyImplement2	Qu
null	null	null	null	null	
null	null	null	null	null	
null	null	null	null	null	

Machine Family

tractfam_id	tractfam_name	tractfam_type	tractfam_code	tractfam_fuel	tracttype_code
1	Tractor - general	1	TRACTOR	DIESEL	SELF_PROPELLED
2	Car	1	VEHICLE	DIESEL	SELF_PROPELLED
3	Selfpropelled Harvester	1	HARVEST	DIESEL	SELF_PROPELLED

Cardinality: Many to one (*:1)

Cross filter direction: Single

☐ Make this relationship active

☐ Apply security filter in both directions

☐ Assume referential integrity

Figure 21. Relation between tractor and Implement Family 2. Source: Power BI Screenshot

This particular case requires a deeper explanation. Here, there are two relations, of which just one is active, and the other one is handled through a function in DAX language that "activates" the relation when needed. Following what was said before about the existence of different columns for the implements, it is easy to see that multiple relations make perfectly sense: One column of implements is related to the family, and it works the same way for the other column of implements. In addition, the type of relation is many to one, meaning that many operations could have one family.

Company

Through this query, it is assured that the companies that are included in the analysis are the ones in Chile, for which the codes are: 46, 51,52,53,54, 55, 80, 84,85 and 86. A wider definition of those was given in the section *Frutícola Agrichile.*

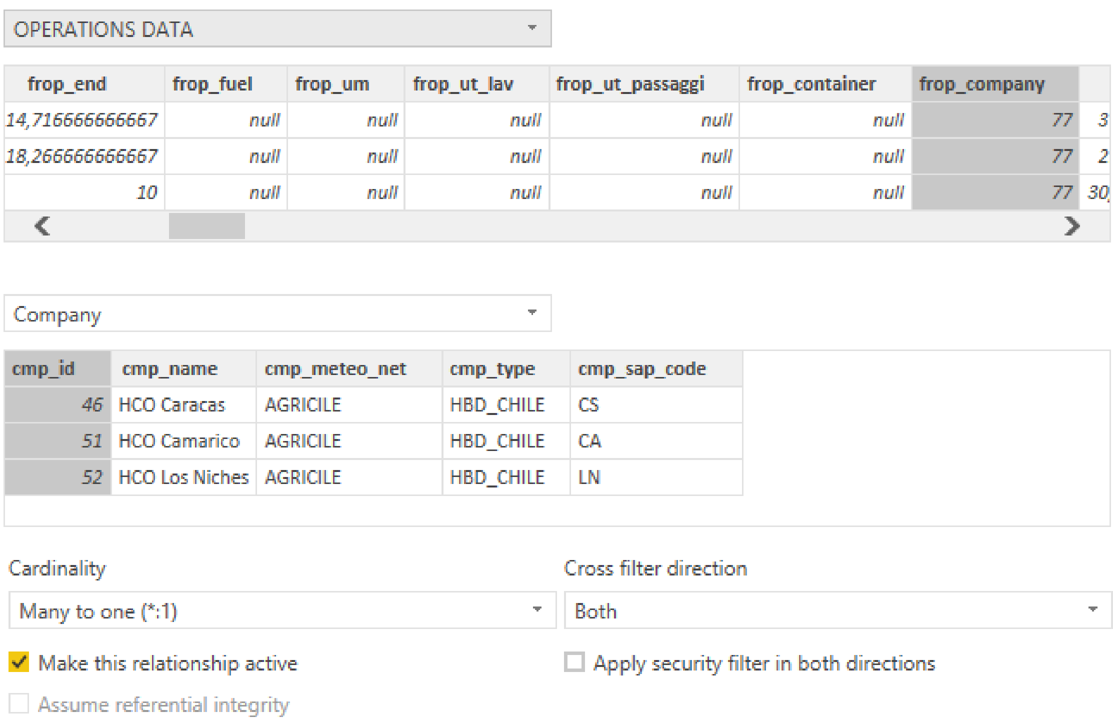
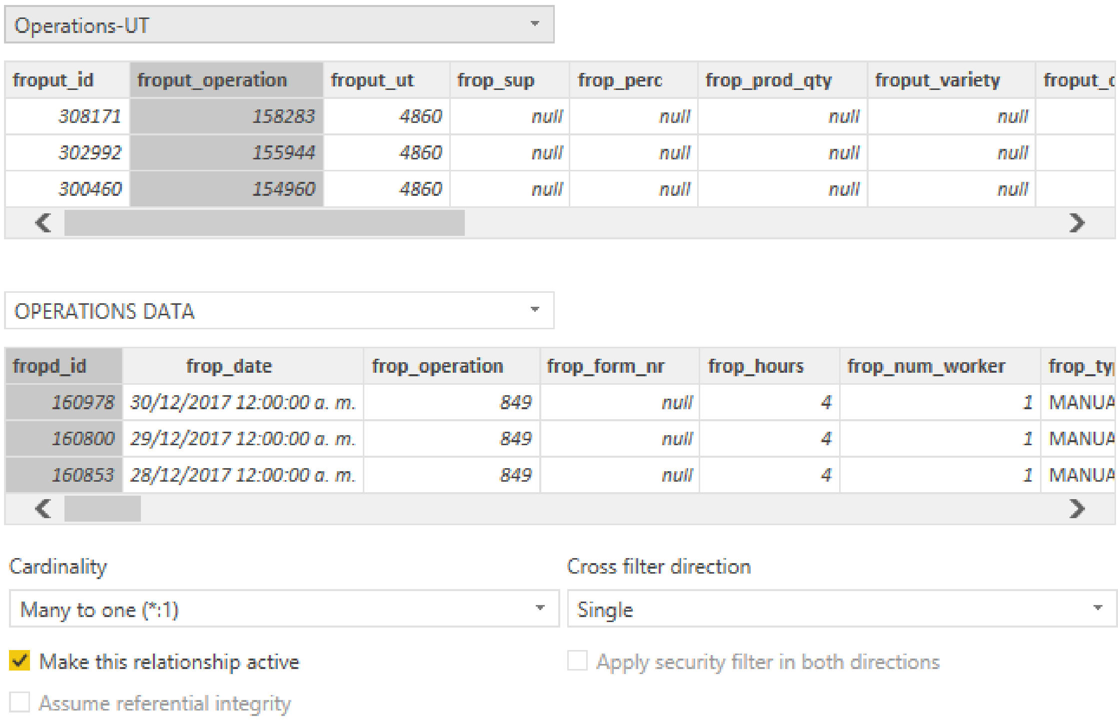


Figure 22. Relation Between Operations Data and Company. Source: Power BI Screenshot

It is perfectly reasonable that the relation is many to one, since many operations could take place in a Company.

Operations UT

Through this table, it is possible to communicate the operation to the Territorial Unit where it took place. In addition, more information is deployed, such as sum of quantity, the date, and if the type of operation is either Manual or Mechanic.


Operations-UT

froput_id	froput_operation	froput_ut	frop_sup	frop_perc	frop_prod_qty	froput_variety	froput_(
308171	158283	4860	null	null	null	null	
302992	155944	4860	null	null	null	null	
300460	154960	4860	null	null	null	null	

OPERATIONS DATA

fropd_id	frop_date	frop_operation	frop_form_nr	frop_hours	frop_num_worker	frop_ty
160978	30/12/2017 12:00:00 a. m.	849	null	4	1	MANUA
160800	29/12/2017 12:00:00 a. m.	849	null	4	1	MANUA
160853	28/12/2017 12:00:00 a. m.	849	null	4	1	MANUA

Cardinality: Many to one (*:1)

Cross filter direction: Single

Make this relationship active

Apply security filter in both directions

Assume referential integrity

Figure 23. Relation Between Operations Data and Operations within Territorial Units. Source: Power BI Screenshot

As it could be seen, the relation is Many to one, meaning that many operations could take place in one Territorial Unit. To solve this problem, however, what was done was to think of the relation backwards, i.e. define the table when needed from the part of Operations UT, and not from the part of Operations Data (as usual).

Territorial Units

It contains the set of territorial units per company, with their corresponding information about hectares, fruit type (variety), "*equipo*", machines, and also the codes necessary for their identification.

Operations-UT

froput_id	froput_operation	froput_ut	frop_sup	frop_perc	frop_prod_qty	froput_variety	froput_c
308171	158283	4860	null	null	null	null	
302992	155944	4860	null	null	null	null	
300460	154960	4860	null	null	null	null	

Territorial Units

frut_id	frut_code	frut_name	frut_ha	frut_company	frut_type	frut_dependency	frut_propertie
8597	LNVPO1_1.36	null	0,007536232	52	ROW	1516	
8598	LNVPO1_1.37	null	0,007536232	52	ROW	1516	
8599	LNVPO1_1.38	null	0,007536232	52	ROW	1516	

Cardinality: Many to one (*:1)

Cross filter direction: Single

- [x] Make this relationship active
- [ ] Apply security filter in both directions
- [ ] Assume referential integrity

Figure 24. Relation Between Operations in Some Territorial Units and Territorial Units. Source: Power BI Screenshot

The same situation than Operations-Ut happens here, and hence, the same procedure to solve it was applied.

To conclude this part, *Figure 25* shows how was the final state of the relations between tables, clearly demonstrating a snowflake structure.

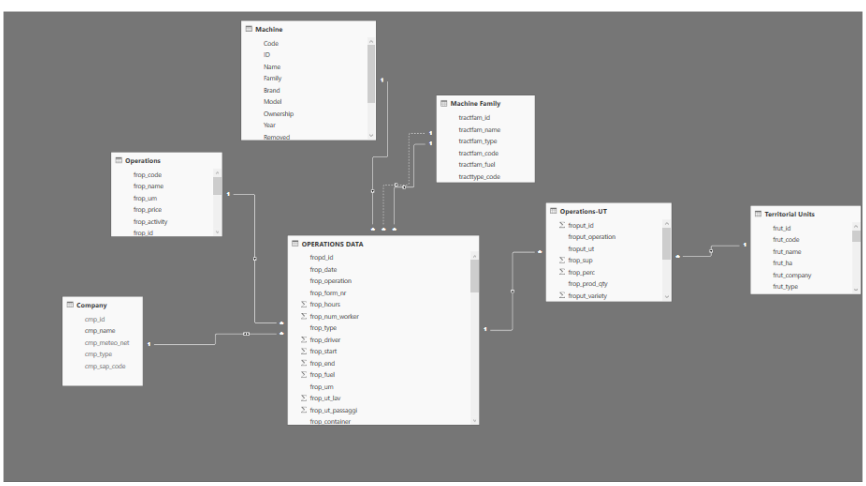


Figure 25. Snowflake Scheme. Source: Power BI Screenshot

11.3.2. Simple Efficiency Table.

This table was not originally in the database, but it has been created by the author to analyze the real efficiencies. It is not other but the outcome of some computations, applying filters and taking information from several tables, exploiting the advantages of the relations defined previously. It contains a total of 28 columns, enunciated and explained below:

- Family Machine: Set this way to have a common measure of comparison between current efficiencies, and real ones.
- Machine Family Name: In the report, it will be displayed by a name and not just a number.
- Total Hours Per Worker: Number of hours per operation, times the number of workers (if the case were that a worker has to perform the operation by himself or herself, how long it will take him to finish it).
- Simple Efficiency: Allows to show a value even though the Manual efficiency was set as Null (when there is a Mechanical operation with just one worker, the current procedure is to set the time as null, even though there is actually a worker performing the activity).
- T/F: It allows to evaluate certain cases were the computation will not be taken into account, i.e. when the start time is null for a mechanic work, when the number of hours worked appears as zero in a manual work, when there are not workers shown and the activity is manual, and when the start time and end time are the same in a mechanic operation.
- Year-OperationData
- Quarter-OperationData: This way, the number of working hours are defined as 9, if it is in the first or second quarter, or 8 for the remaining quarters.
- Frop_date
- FamilyImplement1: If one implement was used.
- Tracfam_name: Currently, the efficiencies are also computed on a tractor family basis, so if the information is presented in this way, the results would be comparable.

- Family Implement2: If the operation has more than one implement.
- Implement1-Family Name: The names of the families of the implements if any.
- Implement2-Family Name: The names of the families of the implements if any.
- Fropd_id: the number of the operation.
- Frop_name: Name of the sub operation.
- Frop_id: Number of the sub operation.
- Tractor Code: It is a reference for the tractor family basis.
- Frop-type: whether the operation is Manual or mechanic.
- Frop_um: because how data is stored in the database, the unit of measure is just in hectares or number of plants.
- Cmp_name: The name of the company (plantation).
- Frop_end: time of finishing the operation.
- Frop_start: time of starting the operation.
- Frop_sum_qta: Column where the quantity done within the time interval is shown.
- Frop_hours: time interval in which a manual activity took place.
- Frop_num_worker: Number of workers for performing the operation.
- Machine Hours per operation: end time minus start time of the operation.
- Manual Simple Efficiency: Quantity done in a manual operation divided by the number of workers times the total hours of work. In addition, the result is multiplied by 8 or by 9 to give it on a daily basis.
  - $Manual\ Simple\ Efficiency = \frac{sum_{qta}(manual)}{hours\ x\ number\ of\ workers}\ x\ 9\ or\ 8$
- Mechanical Simple Efficiency: Quantity done in a mechanical operation divided by the duration of the operation. As for the case above, the result is also multiplied by 8 or by 9 to give it on a daily basis.
  - $Mechanical\ Simple\ Efficiency = \frac{sum_{qta}(mechanic)}{end\ time - start\ time} x\ 9\ or\ 8$

It is also important to highlight the fact that the table was created using DAX language. The code is shown below:

```
Simple Efficiency =
CALCULATETABLE(
 SUMMARIZE( 'OPERATIONS DATA',
        'OPERATIONS DATA'[Year-OperationData],
        'OPERATIONS DATA'[Quarter-OperationData],
        'OPERATIONSDATA'[frop_date],
        'OPERATIONS DATA'[FamilyImplement1],
         'Machine Family'[tractfam_name],
         'OPERATIONS DATA'[FamilyImplement2],
        Operations[frop_name],
        Operations[frop_id],
        'OPERATIONS DATA'[Tractor_Code],
        'OPERATIONS DATA'[frop_type],
        Operations[frop_um],
        'Company'[cmp_name],
        'OPERATIONS DATA'[frop_sum_qta] ,
         'OPERATIONS DATA'[frop_end],
        'OPERATIONS DATA'[frop_start],
        'OPERATIONS DATA'[frop_hours],
        'OPERATIONS DATA'[frop_num_worker],
        'OPERATIONS DATA'[fropd_id] ,
       "Machine Hours per operation", SUMX('OPERATIONS DATA', -
       1*'OPERATIONS DATA'[frop_start] + 'OPERATIONS DATA'[frop_end]),
       "Manual Simple Efficiency",
           IF(OR('OPERATIONS DATA'[frop_type]="MANUAL",'OPERATIONS
          DATA'[frop_type]="HARVEST_MANUAL"), IF(OR('OPERATIONS
          DATA'[Quarter-OperationData]=1 ,'OPERATIONS DATA'[Quarter-
          OperationData]=2) ,(DIVIDE('OPERATIONS
```

```
        DATA'[frop_sum_qta],('OPERATIONS DATA'[frop_hours]*'OPERATIONS
        DATA'[frop_num_worker])))*9 , (DIVIDE('OPERATIONS
        DATA'[frop_sum_qta],('OPERATIONS DATA'[frop_hours]*'OPERATIONS
        DATA'[frop_num_worker])))*8),BLANK()),
      "Mechanical Simple Efficiency",
         IF(OR('OPERATIONS DATA'[Quarter-OperationData]=1 ,'OPERATIONS
        DATA'[Quarter-OperationData]=2),(DIVIDE('OPERATIONS
        DATA'[frop_sum_qta], ABS(SUMX('OPERATIONS DATA', -
        1*'OPERATIONS DATA'[frop_start] + 'OPERATIONS
        DATA'[frop_end])))*9),(DIVIDE('OPERATIONS DATA'[frop_sum_qta],
        ABS(SUMX('OPERATIONS DATA', -1*'OPERATIONS DATA'[frop_start] +
        'OPERATIONS DATA'[frop_end])))*8)
) ) ,
 'OPERATIONS DATA'[frop_type] IN { "MECHANIC", "MANUAL",
"HARVEST_MANUAL", "HARVEST_MECHANIC"},
  Operations[frop_um] IN {"ha","pl"},
 'OPERATIONS DATA'[Year-OperationData]>2014 ,
 'OPERATIONS DATA'[frop_sum_qta]<> BLANK(),
  'OPERATIONS DATA'[frop_operation]<> BLANK(),
  'OPERATIONS DATA'[frop_company] IN {46 ,51 ,52 ,53 ,54 ,55 ,80 ,84 ,85 ,86}
        )
```

11.3.3. Simple Efficiency Final.

This table was not originally in the database, as happened with the table of simple efficiency. Here, all the particular cases analyzed in the T/F columns of the former table were not taking into account, and for instance, only the values for which is possible and makes sense to compute the efficiency are taken into account.

11.3.4. Simple Group by Values.

For each combination of variables, i.e. each combination of sub operations within a company, with a particular implement/s and/or tractor, the average value was computed and deployed in this table. This is the first approach used, since it is a point estimation. Later, however, an estimation based on confidence intervals will be introduced.

11.3.5. Broken Down Efficiency.

The logic used for this query was quite similar to the one used in *Simple Efficiency*. Nonetheless, in addition of containing information regarding Territorial Units, it deals with the problem described before with the query of the model: since the relation is 1 to many, the calculate Table function was called from *Operations-UT*, and not from Operations Data, as it is usually done.

11.3.6. Broken Down Efficiency Final.

Like the case of *Simple Efficiency Final*, all the particular cases analyzed in the T/F columns of the former table were not taking into account, and for instance, only the values for which is possible and makes sense to compute the efficiency are taken into account.

11.3.7. Creating tables: Broken Down Group by Values.

This is the first deliverable of the computations. There are two versions of it though: the one dealing with point estimations, which is this one, and another one dealing with confidence interval estimations. Although, the structure in terms of columns and comparisons is the same. A general description makes realized that the query contains 7 (seven) columns:

- Brokendown_efficiency Final_frop_id: The id of the operation.
- Brokendown_Efficiency Final_cmp_name: The company name where it took place.
- Brokendown_Efficiency Final_frut_code: The code of the territorial unit.

- Brokendown_Efficiency Final_Code(Tractor): The code of the tractor used for the operation (if any).
- Brokendown_Efficiency Final_FamilyImplement1: The family of the first implement used, if any.
- Brokendown_Efficiency Final_FamilyImplement2: The family of the second implement used, if any.
- Average Simple Efficiency: The point estimate simple efficiency for the different combination of the variables.

12. Analysis of Results.

During the subsequent subsections, an analysis of the results is done. However, it is important to highlight the fact that this analysis takes place using statistical methods that assure that the comparison was done in a right and suitable way.

12.1.Comparison between the Database and Tables created

As a clarification, the creation of this set of tables had the scope of building a structure that was comparable with what the enterprise currently has in its data base in terms of efficiency. Right now, following the premise that was introduced before in the section Database Description, what they use as a base for efficiency budget estimation has the following format:

fropeff_id	fropeff_operation	fropeff_efficiency	fropeff_machine	fropeff_impl1	fropeff_impl2
717	533	10			
792	533	1			
533	536	200			

Figure 26. DB Format for Efficiencies. Source: Power BI Screenshot

As it could be seen, the data is organized in a way such that for each combinations of sub operation, company, machine family, and family of implements, there is an efficiency in terms of **number of plants**, or **number of hectares** per day. In this order of ideas, all the steps done before where directed towards the presentation of the information in this way, bearing in mind that for the same combination of variables (i.e. columns), the activity being analyzed has been done one or many times, so, the efficiency could be estimated throughout the years. However, statistically speaking, the formal way to present the efficiency is not with just an average value, but with the interval of confidence of that average value.

Since the query obtained and what was already in the database are in different positions, i.e. there are two different queries containing the information, there was necessary to follow a process for the homologation of data to make the comparison between the table Budget Efficiency with BrokenDownEfficiencyGroupValues possible, and it was the following:

The later query was exported to Excel, and then reloaded to the Query Editor in PowerBI as a local database source (this procedure will be done once more at the moment when the final report of the work will be presented, this to keep the information as updated as possible. Connections with all the tables, but this one, will be done automatically, through the server of 3a).

Once reloaded, it has been merged with the former table through the option *Merge Queries* of the editor. What this will do is what is known in MySQL as *Outer Join*, merging the data for which the fields of interest coincide, plus the extra one. Creating, at the end, a table with all the possible combinations of operations with other variables (Machines, companies, etc.) with their corresponding efficiency. More details regarding the comparison of what was computed, and what was already used for estimating efficiencies is given in the subsection "*Analysis of the Data in R studio*". Furthermore, a smaller version of the table of *Simple Efficiency* and *Broken_Down Efficiency by Group*s used for an Excel deliverable to the company is displayed in *Exhibits 1 and 2*. However, the one corresponding to Broken_Down Efficiency cannot be compared, because there is nothing to be compared with at the moment. [4]

12.2. Restrictions of the model

The following restrictions have been taking into account due to the fact of anomalies in the database:

- Values related to operations that appeared as having Implements, but without quantity in the filed "Number of implements", for Manual Operations
- Start time null for a mechanical work,
- number of hours is zero in a manual work,

---

[4] Analysis of efficiency is done by two families of implements, even though for some historical cases, there are operations with even 4 implements associated to it. The reason for this is to make the model comparable in terms of what is registered right now in the database; besides, there will be a bias in the analysis with operations of more than 2 (two) implements, since it is not the same to perform an activity with a different number of operations (in terms of times).

In addition, the usage of more than two implements are few and very peculiar.

- Not workers shown, and the activity is manual,
- When the start time and end time are the same in a mechanic operation,
- Some groups biased by human mistakes,
- Some combinations have been found in the Database that have been not performed historically.
- To all of the above, what was explained in the subsection *Simple Efficiency Table*, under the calculation of the T/F column, has to be added.

13. Analysis of the Data in R studio

Once the data has been deployed as it was explained in the last section[5], it has been loaded to R studio to estimate the interval of confidence for each combination of the variables. It is of extremely importance to point out that it is not necessary to know in advance the distribution of the data, since the Bootstrap principle is implemented, and it could be no parametric. According to the Cambridge Dictionary of Statistics (Everitt and Skrondal, 2018) a Confidence Interval is

> A range of values, calculated from the sample observations, that are believed, with a particular probability, to contain the true parameter value. A 95% confidence interval, for example, implies that were the estimation process repeated again and again, then 95% of the calculated intervals would be expected to contain the true parameter value….

As so, the estimation of the mean through confidence intervals, for the combination of variables explained above, will assure the expectation, with high probability, to find the true parameter value of simple efficiency (and a Broken_Down efficiency, depending on the case) within it. The reason to work with a 5% significance level (alpha) is that conventionally it is set this way (Everitt and Skrondal, 2018), and because with a higher alpha, the risk of incurring in Type II-Error increases (the probability not reject the null hypothesis, giving the fact that it is false, hence, the power of the test will be lower).

With this in mind, the sense of using confidence intervals, instead of just point estimates should be clearer: it is all terms of probabilities. However, because of how the data set exists in the database of 3a S.r.l, some operations have just been done rarely throughout the years for which data is being analyzed, and this is precisely why the estimation of the confidence intervals is just done for observations that are repeated more than 5 (five) times. In addition, this was the minimum number of repetitions that was agreed with the enterprise to consider an operation as meaningful to them

The estimation explained above was be done in R studio, through a method called Bootstrap, which is basically a simulation for statistical inference that can be used to study the variability of estimated characteristics of the probability distribution of a set of observations and

[5] The procedure was done for Simple Efficiency and for Broken-Down Efficiency as well. In this subsection, just the first one is explained. Although, it applies the same for the second one.

provide Confidence intervals for parameters. The basic idea of the procedure involves sampling with replacement (resampling) to produce random samples of size n from the original data x_1,x_2, …., x_n; each of these is known as a *bootstrap sample* and each provides an estimate of the parameter of interest. (Everitt and Skrondal, 2018).

In particular, the bootstrap principle states that the empirical distribution of the resampling is, by the law of large numbers and with enough data, a good approximation of the true distribution (Orloff and Bloom, 2018); and it is set up considering a sample x1,x2, …., xn from a distribution F, where a desired statistic (t) is to be computed from the sample. Also, we have, F*, which is the empirical distribution of the data (the resampling distribution), x1*,x2*, …., xn* ,named, the resampling data of the same size as the original sample, and a statistic t*, computed from the resample.

So, the principle states that F will be well approximated by F*, and that the variation of t is well-approximated by the variation of t* (this is precisely why the resample size has to be the same size as the sample size). (Stat.cmu.edu, 2018).

Assuming that the true value is set as t_0, the true and that a confidence interval C is calculated, being Cl the Inferior Confidence interval limit, and Cu the upper one, what we want to calculate is

$$\frac{\alpha}{2} = Pr(t - Cl \leq t - t0) \qquad (1)$$ (Stat.cmu.edu, 2018).

$$\frac{\alpha}{2} = Pr(t - Cu \geq t - t0) \qquad (2)$$ (Stat.cmu.edu, 2018).

The point is that the distribution of t*-t is approximately the same as the distribution of t-t_0. This is what is called the **basic bootstrap** confidence interval.

The logic behind this calculation is quite interesting. How R studio works is that the number of times a resampling is done is defined by the user, being it usually a big number (i.e. 1000), once done, for each resampling, the statistical is computed, organized, and the percentile 5^{th} and 95^{th} are used to compute the difference between the sampling statistical, and the resampling one, so that the limit values of the intervals are found.

The code for the bootstrap in particular require the following lines to be executed:

```
library(boot)
myStat <- function(x, idx) {
        boot.sample <- x[idx]
        theta.star <- mean(boot.sample)
        return(theta.star)
        }
```

On the other hand, it the structure of the function within the program was defined as follows:

```
for(i in 1:633){
  objetoBoot<-boot(data = datos$SimpleEfficiency[datos$IndexR==i], statistic = myStat,
R = 1000)
  IntervaloBoot<-boot.ci(objetoBoot, conf = 0.95, type = "basic")
  MediaGrupoBoot<-IntervaloBoot$t0
  LimInfBoot<-IntervaloBoot$basic[4]
  LimSupBoot<-IntervaloBoot$basic[5]

  VectorPromedios<-c(VectorPromedios,MediaGrupoBoot)
  VectorLimInf<-c(VectorLimInf,LimInfBoot)
  VectorLimSup<-c(VectorLimSup,LimSupBoot)
}
```

In the part of the function boot, the first field tells the program with which data to work with (it goes from 1 to 633, which is the number of combination of variables in the database), then it is giving the instruction to run a number of simulations equals to 1000 for each ones of the groups (logic of bootstrap function explained before), finally, it is using a level of confidence of 95%, and is also using a *basic* method to run the bootstrap.

At the end, all of the information is saved in three vectors: two for the limits, and one for the mean. The deliverable, however, is an excel document with all the possible combination of variables that could be estimated based on what is in the database, with its confidence intervals respectively. The idea is then to load the data using Power BI, so that it could be compared with what is currently being used.

13.1. Heuristic for R analysis

Generally speaking, the procedure followed for the computation of the confidence intervals could be summarized in the following steps:

- Download the table from Power BI where all the operations appear (without grouping).
- Create a list for each unique combination of variables, each one will be assigned to a group called group 1
- Count the number of operations done for each row of the list
- Clear the data for values in which the number of repetitions is lower than 5
- Perform the analysis in R for values obtained in the step before. Since low repetitions have not been taken into account for the analysis, a new group have been created, group2.
- For values in group 1 not included in group 2, just the average have been computed, not the confidence interval.
- Coming back to the R analysis, some groups will have exactly the same values for the operations within them, so the confidence interval will not be computed. Instead, just the average will be taken. This will create a group 3.
- As a final step, the confidence intervals obtained for Group2, plus the average for groups with observations lower than 3 (group 1), plus exact same values per groups (group 3), are summarized in a table.

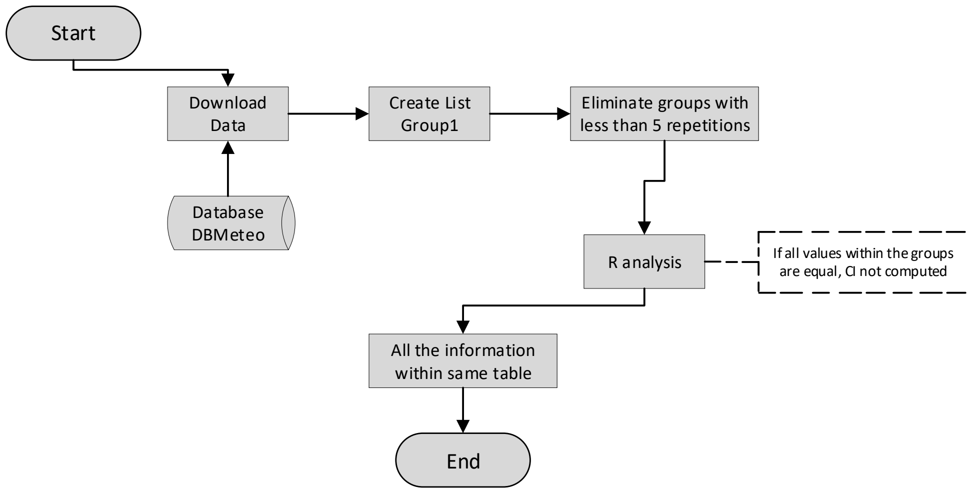


Figure 27. Process Diagram for R Analysis. Source: Own Elaboration.

13.2. Outliers

Within the process of exploration of the data, the author noticed that for certain groups, there were some values that were particularly strange compared to those seeing in their same group. This observation was shared with the people in charge in the enterprise, and the conclusion was that they were probably outliers. As an instance, a method was chosen in order to remove them, and allowing, at the same time, to have less noise in the computation of the mean and confidence interval for each group.

The method used was the one described by Tukey (1977) and it states some basic definitions for its use:

- The Inter Quartile Range (IQR) is known as the distance between the lower Quartile (Q1) and the upper Quartile (Q3).
- Inner fences are located at a distance 1.5 IQR below Q1 (Q1-1.5IQR), and above Q3 (Q4+1.5IQR).
- Outer fences are located at a distance 3IQR below Q1 and above Q3 (Q1-3IQR, Q3+3IQR)
- A value beyond the inner fences is a probable outlier.

The advantage of this method is that it makes no distributional assumptions, so it does not depend on a mean or standard deviation. However, one must be careful because it may not be appropriate for small sample size (Seo, 2006, cites the work by Tukey, 1977 to clear this point).

Once the outliers were eliminated, with the consent of the enterprise, the analysis in R was done, as described before.

13.3. Validation of Results

Once the confidence intervals were obtained, the outcome needed to be evaluated with respect to what was already uploaded in the data base (the theoretical table of efficiencies). Although surprising for some cases, for others, the results were well behaved. For a clearer explanation of this point, the results were divided into 5 (five) major groups:

1. Values with a low percental difference with respect to the historically calculated Efficiency: These are the desired values and corresponds to the operations that have been specified precisely in the budget until now.
2. Values with a high percental difference with respect to the historically calculated Efficiency: These operations need to be looked at to evaluate what has happened throughout the years, and to understand if the computation of the budget has had major deviations with respect of what has indeed happened.
3. Values with a really high percental difference with respect to the historically calculated Efficiency: Most of them correspond to mistakes done while inserting data into the database. There is no need to alarm by the huge gap that exists between what has been calculated historically, and what was already in the database as predetermined.
4. Groups that appeared historically but were not in the database: Operations that had a specific combination of machines and implements, that were not taken into account at the time of defining the table of efficiencies.
5. Groups that did not appear historically and were in the database: Operations with a specific combination that were taking into account when defining the efficiencies, but have not been performed; or if performed, some inconsistency in the data was made.

Each of these particular cases were discussed and measures are being taken. In the future, the enterprise expects to show the results to an agronomist, so that he/she could evaluate in the field what is actually happening.

In addition, for statistical validation of the data that was already in the database, with respect to the one found historically, the value for each specific combination was compared to the limits of the confidence limit computed. If the value was within the limits, there was no reason to statistically say that they were different (this was marked in the Excel document as Validated). However, if it was beyond the limits, then the opposite hypothesis would be true.

As expected, there were some values, corresponding to groups 2 (two) to 5 (five) that were not within the intervals.

Exhibit 3 illustrates what is explained above for the case of Simple Efficiency, and *Exhibit 4* does the same for the case of Broken_Down Efficiency.

14. Algorithm

14.1. Context

This is where the scope of the third objective defined in the work is reached. The idea of the algorithm is to assign the machines (i.e. tractors and implements) to use in an operation to be done, knowing in advance the type of operation, and the date in which it needs to be done (including time), but also taken into account that each combination of machines has a specific duration, which was already defined in the previous section, and a specific fuel consumption, measured in liters per units of time. Moreover, it also considers restrictions of time and availability of the machine and the implement.

Generally speaking, the *Figure 28* explains broadly what the procedure does: it takes as an input the efficiency, the machines available, how much of fuel it consumes, and the date in which the operation needs to be done, including hour and minute of the day.

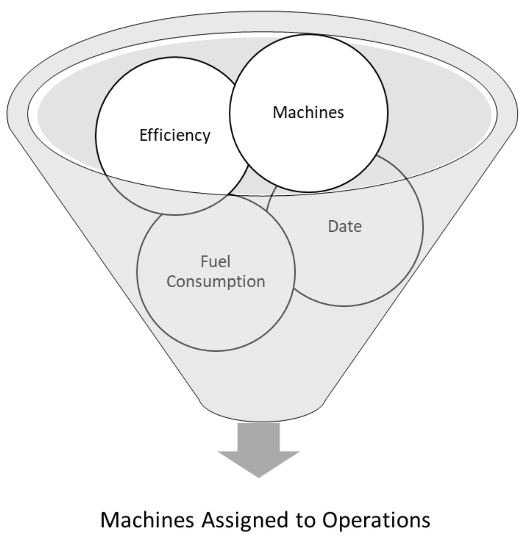


Figure 28. Input and Output for the Algorithm. Source: Own Elaboration

It is important to state that all the coding will take place in Visual Basic, which is described under the section of Languages and Tools used.

14.2. Restrictions

Although the algorithm is a good replica of what is actually happening; it cannot be 100% precise in terms of real-life scenarios. In this sense, for a correct comprehension of its performance, the following restrictions need to be considered:

- Assignation of resources is made on a daily basis.
- Operations entered have to be real, i.e. they have to exists within the DB.
- If the quantity to be processed is set as 0 (zero), then the duration will be zero. Although there have been some inconsistencies historically speaking between this quantity, and the time taken to perform the operation. This means, that, even though they have set the quantity as zero, the duration appears different than this number.
- No-feasible solutions in terms of time intervals to be respected for an operation to be accomplished are not taken into account.
- All the operations analyzed are performed by 1 (one) worker, named the driver of the machine., i.e. the number of workers to be assigned for each operation will not be analyzed.
- Free circulation of machines among enterprises in Chile is assumed. This because there is no a clear stock for each of the company. However, data was analyzed historically throughout 4 (four) years and a half for San Sebastian, so all the machines used were considered.
- No set up times are taken into account.
- No preemption.
- There is no specification for when the machines are being repaired or under maintenance, so they will all be assumed to work well for the period of analysis.
- The performance will be the same throughout the whole shift. There are no hours when some machine is more or less productive than usual.
- Drivers are supposed to have the same skills with the machines, so there is no division between “experts” and “rookies”.
- Zones are assigned randomly.

- Implements and tractors are together when they need to be used.
- At the end of the working day, the machines are let in the last place they worked, and do not go back to a Warehouse.

14.3.Linear Programming for the Problem

Even though the problem has been solved without Linear Programming, it is of extreme importance to formulate it in these terms for a better understanding of the approach used during the process of reaching the solution.

What this technique does is to find the global optimum, delimiting, through linear functions, the area of iteration to figure out what the maximum or the minimum is, depending of the nature of the problem.

In this particular instance, the following code explains, in a different way, what the algorithm does. However, no solution is available because of the number of indexes that the Decision variable has.

```
/*Linear Programming: Assignation of Machines*/
/*Assumptions:
 There was no capacity analysis for implements
 Data in which the operation starts could not be modified
 One implement is the maximum (the cases in when there are 2 are weird)*/

/*Sets*/
set I;/*machines*/
set J; /*implements*/
set K;/*Sub-operations. Thought as possibilities (elements of the efficiency list)*/
set L; /*Operations*/

/*Decision Variables*/
var x{I,J,K,L} binary; /*If assigning machine i, implement j, to operation l, with
suboperation k*/
```

```
var y{I,K} binary; /*It keeps track of the machines used in the date*/
var w{I,K,L} binary /*Restarts the variable y when the machine is available again*/


/*Parameters*/
param Pos{K,L};/*Binary parameter that states the possibility to assign, for each operation, a suboperation (an element of the list)*/
param EM{I,K};/*Existence of the use of the machine in the suboperation K*/
param EI{J,K};/*Existence of the use of implement j in the suboperation k*/
param Dates{L};/*Date in which the operation needs to start*/
param Machines{I};/*Number of times machine i appears in the list of efficiencies*/
param QuantityEfficiency{I,J,K}; /*Efficiency of number of plants, or hectares per unit of time*/
param Quantity {L}; /*quantity to be done per operation*/


/*Objective Function*/
minimize z: sum{i in I, j in J, k in K, l in L} x[i,j,k,l]*QuantityEfficiency[i,j,k]*Quantity[l]*Pos[k,l];

/*Restrictions*/
s.t.Assingnation{l in L}: sum{i in I, j in J, k in K} x[i,j,k,l]=1; /*For each operation, one assignation must be done*/
s.t.Possibility{i in I,k in K:EM[i,k]=0}: sum{j in J, l in L} x[i,j,k,l]=0 ; /*If the machine cannot be assigned to suboperation, then the variable is zero*/
s.t.Possibility2{j in J,k in K:EI[j,k]=0}: sum{i in I, l in L} x[i,j,k,l]=0 ; /*If the Implement cannot be assigned to suboperation, then the variable is zero*/
s.t.AlreadyAssigned{l in L}: sum{i in I, j in J, k in K} x[i,j,k,l]*Machines[i]=sum{i in I, k in K} y[i,k]; /*If assigned, all the machines will be marked as being used*/
```

```
s.t.Reestart{k in K, l in L: l>=2}:sum{i in I, j in J} (x[i,j,k,l-1]*(QuantityEfficiency[i,j,k]*Quantity[l-1]*Pos[k,l-1]+Dates[l-1]))<=sum{i in I} (Dates[l]*w[i,k,l]);
/*It will be 1 if the condition of usability of the machine is met*/
s.t.Reestart2{i in I, l in L, k in K}: (1-w[i,k,l])<=y[i,k]; /*If condition of reset is 1 (machines is able to use again), then, Y[i,k], which is the restriction for the machines, will be available again*/
```

14.4. Getting the Data

Although what was used before for getting and retrieving the data was Power BI, from this point forward, Microsoft SQL Server Management Studio was implemented. The reason is simple: the computer where all the work was done did not have enough RAM memory to run Power BI in a good way; any simple computation was taking a lot of time, and most of the queries, when designed, were not running either.

The change implied the use of a different language and logic to get the work done, but the principle was the same: tables were accessed through the database, and foreign keys were used for joining tables.

From all the data that is stored into the DB, there were some particular tables that came to the interest of the thesis, and those are precisely the ones used as input for what was done in this section, named:

- [DBMeteo].[dbo].[ferr_an_implements]: Table where all the implements were listed.
- [DBMeteo].[dbo].[ferr_an_tractor]: Every tractor for all the companies was found here.
- [DBMeteo].[dbo].[ferr_budget_anag_efficiency]: tractor, implements and efficiencies associated to a sub-operation.
- [DBMeteo].[dbo].[ferr_budget_task]:total days, human resources, and machines used for a task with its respective efficiency.
- [DBMeteo].[dbo].[ferr_operations]: sub-operations.

- [DBMeteo].[dbo].[ferr_operations_data]: Operations done with their duration, consumption, and quantities.
- [DBMeteo].[dbo].[ferr_operations_data_implement]: Relates the sub-operation with the implement.

14.4.1. Delimitations of the Queries.

Since the algorithm was dealing with the scheduling of machines, among which tractor is including, the operations done needed to be filtered just for those mechanic, and for those code named by 3a S.r.l. as harvest mechanic. In addition, the company chosen for the analysis was San Sebastian, corresponding to number 53 of companies' list in Chile. The reason for the former was because 3a has in high esteem this particular company, considering it as the biggest one in Chile in terms of processes and resources used, and if a new tool was to be implemented (the algorithm), it made sense that the pilot needed to work well in San Sebastian.

Moreover, other restrictions were taken into account, so that, data was congruent, and the procedure will not drop strange results, among them are:

- The unit of measure needed to be hectares or number of plants. This comes directly from the logic of how the data is entered in the database.
- Frop_sum_qta is not null: There was the need to be a quantity defined. In this way, a computation was possible.
- Tractor_name is not null: Also, a tractor was needed for analyzing the operation
- Frop_fuel is not null AND Frop_fuel>0
- Frop_end>frop_start AND (frop_end>0 OR frop_start>0): The time in which the operation ends needs to be greater than the time on which it starts, and one or both have to be greater than zero.
- Frop_um_code='HOURS' this way, it is assured that the difference between start and end is in hours.
- Just some particular operations that filled the requirements were taken into account.

For the last restriction, an analysis was done throughout the time to understand how many times a particular operation was done in the year. Below, the results for the operations, applying the filters mentioned above, for the years 2015, 2016, 2017 and 2018, are shown:

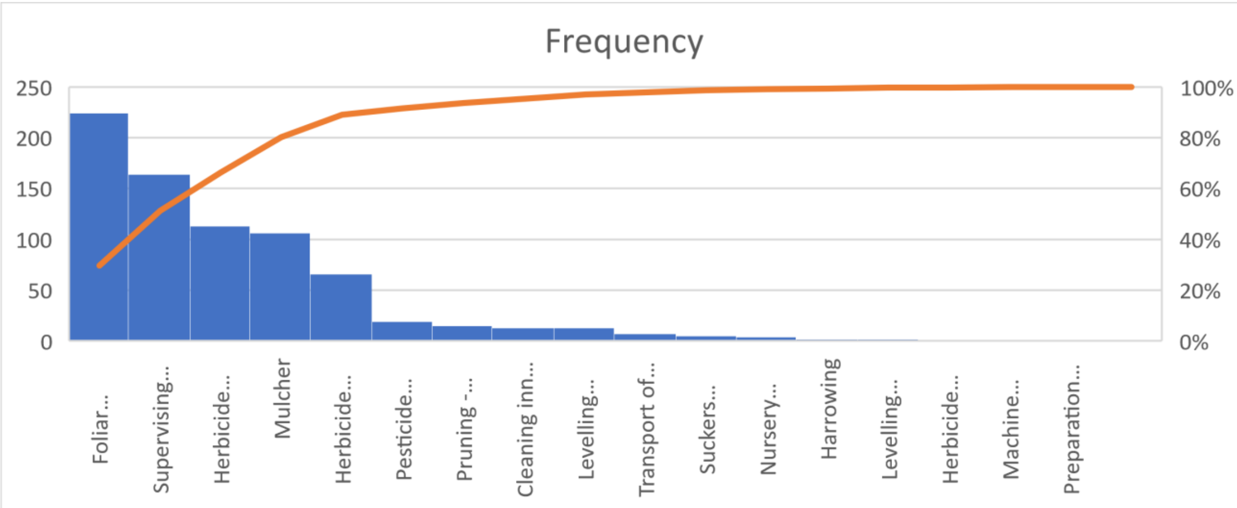


Figure 29. Pareto Diagram for the Number of Times an Operation is Done in 2015. Source: Own Elaboration

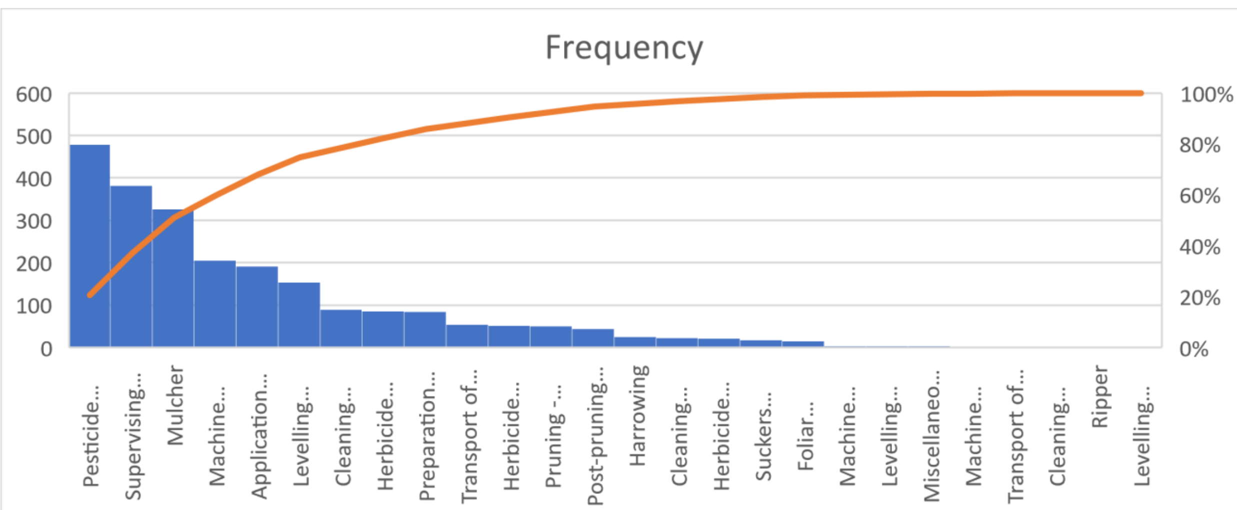


Figure 30. Pareto Diagram for the Number of Times an Operation is Done in 2016. Source: Own Elaboration

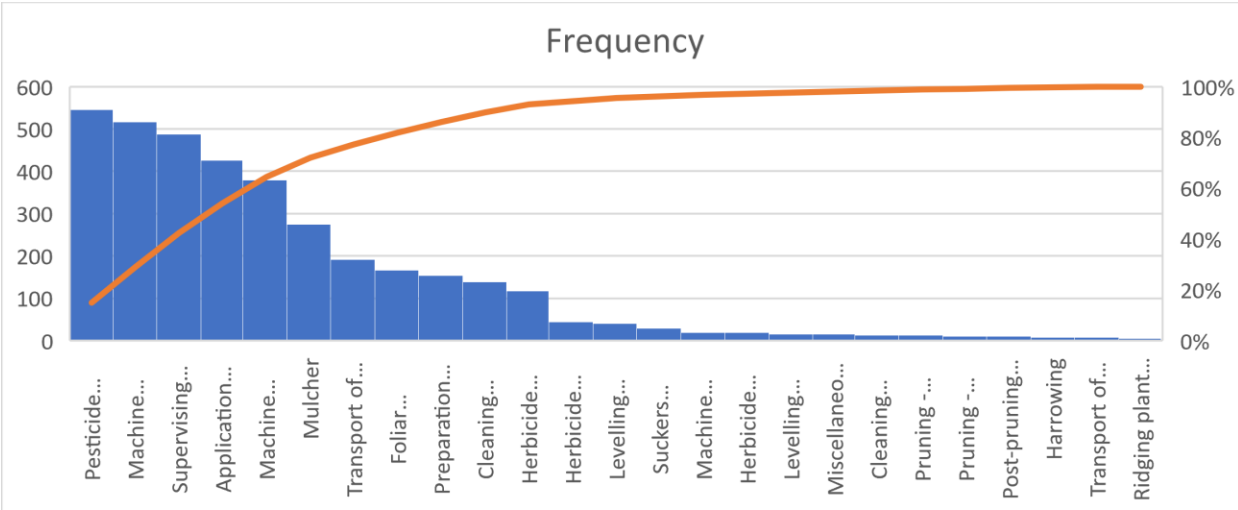


Figure 31. Pareto Diagram for the Number of Times an Operation is Done in 2017. Source: Own Elaboration

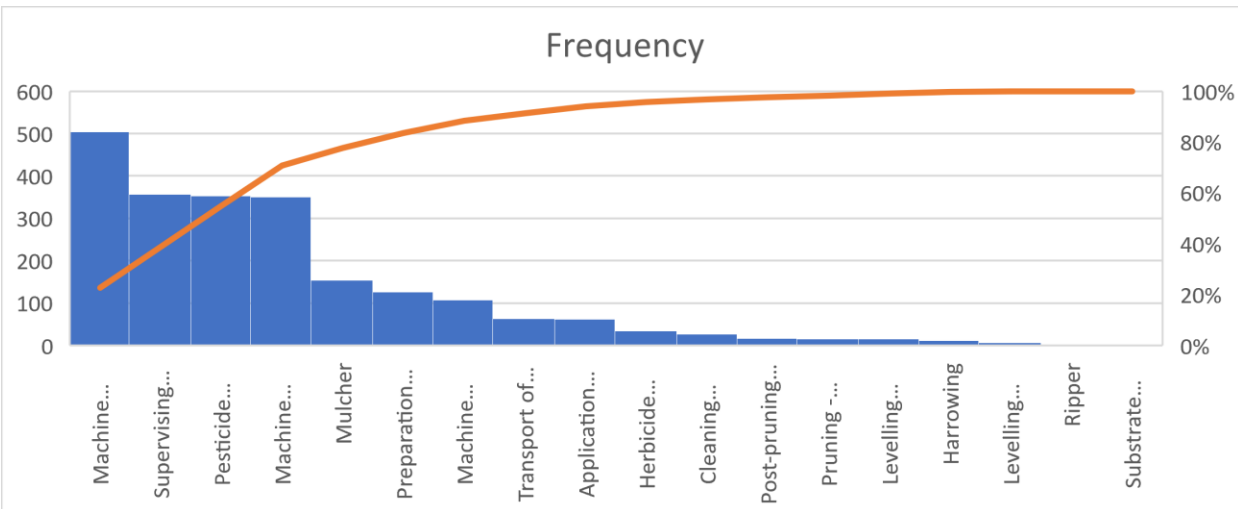


Figure 32. Pareto Diagram for the Number of Times an Operation is Done in 2018. Source: Own Elaboration

The idea for this analysis is to point out what operations are done more frequently throughout the years, and to pay special attention to them, treating them as priority in the time of scheduling. Although there are some with low frequencies that can be considered also as important. However, the algorithm deals with all of them; it is the decision of the chief to decide which ones are worth to be submitted to a deeper analysis.

14.4.2. Finding the Efficiencies.

As stated before, the efficiencies in terms of quantities were found historically in the previous section. What is new here is that also the fuel efficiency was taken into consideration, computed as the ratio between what was consumed and the difference in the time of finishing and starting the operation.

$$\text{Fuel Efficiency} = \frac{frop\_sum\_qta}{frop\_end - frop\_start}$$

The procedure was the same than last time, this time applied for fuel, and with the great clarification that the algorithm takes punctual values and no the confidence interval, to iterate. Something else to clarify is the fact that sometimes, for the fuel, the data is not entered in an ideal way, in the sense that, usually, the amount of fuel consumption per liters corresponds to an operation; but how the workers insert the data in some occasions, is just writing in the database the date and time in which they pumped up the tank, an associate it with the last operation done. What happens then is that the fuel consumption is being higher for some operations, when it should be lower, or lower and even zero for some others where the consumption should be higher. Bearing in mind this particular situation, the analysis with respect to Fuel consumption is biased.

With this explained, the next procedure is to show how the efficiencies were estimated using Microsoft SQL Server Management Studio: From all the tables mentioned above, two of them, merged with others were the key ones to find the information needed. This procedure was done following the subsequent code:

```
SELECT TOP (1000000)

a.[frop_date], a.[frop_operation] , b.frop_name, a.[frop_tractor],
e.tractor_name, e.tractor_family , a.[frop_start], a.[frop_end],
(a.[frop_end]-a.[frop_start]) AS No_Hours,
a.[frop_fuel],Implements.frimpl_family, a.[frop_sum_qta], b.frop_um,
(a.frop_fuel/(a.[frop_end]-a.[frop_start])) AS FuelEfficiencyHours,
(a.frop_sum_qta/(a.[frop_end]-a.[frop_start])) AS QuantityEfficiencyHours


FROM [DBMeteo].[dbo].[ferr_operations_data] a
       LEFT JOIN [DBMeteo].[dbo].[ferr_operations] b
       ON a.frop_operation= b.frop_id
        LEFT JOIN (SELECT d.frimpl_family, c.fropimpl_operation,
                         c.fropimpl_implement, c. fropimpl_num
                   FROM [DBMeteo].[dbo].[ferr_operations_data_implement] c
```

```
                        LEFT JOIN [DBMeteo].[dbo].[ferr_an_implements] d
                         ON c.fropimpl_implement=d.frimpl_id) AS Implements
       ON a.fropd_id= Implements.fropimpl_operation
       LEFT JOIN [DBMeteo].[dbo].[ferr_an_tractor] e
       ON a.frop_tractor= e.tractor_id
WHERE a.frop_type IN ('MECHANIC','HARVEST_MECHANIC') AND
 b.frop_um IN ('ha', 'pl')
  AND a.frop_company=53 AND YEAR(a.frop_date)=2017
  AND a.frop_sum_qta IS NOT NULL
  AND e.tractor_name IS NOT NULL
  AND a.frop_fuel IS NOT NULL
  AND a.[frop_end]>a.[frop_start]
  AND (a.frop_end>0 OR a.frop_start>0)
  AND a.frop_fuel>0
  AND a.[frop_um_code]='HOURS'
  AND b.frop_name IN ('Application of fertilizer on Hazelnuts trees',
  'Cleaning across lines',
  'Cleaning around plants',
  'Cleaning collector channels',
  'Cleaning inner channels',
  'Foliar fertilization ',
  'Harrowing',
  'Herbicide application between hazelnuts lines',
  'Herbicide application inner & outer channels',
  'Herbicide application on hazelnut line',
  'Levelling across the rows',
  'Levelling before planting',
  'Levelling between the rows',
  'Machine collection 1st',
  'Machine collection 2nd',
  'Machine collection last pass',
  'Miscellaneous Harvesting',
  'Mulcher',
  'Nursery weeding',
  'Pesticide application',
  'Post-pruning application',
  'Preparation with blowers',
  'Pruning - Mainteinance - High Pruning (European varieties)',
  'Pruning - Mainteinance - High Pruning (Young trees)',
  'Ridging plants in nursery',
  'Ripper',
  'Substrate mixture ',
  'Suckers removal spraying',
  'Supervising irrigation work',
  'Transport of pruning material',
  'Transport of pruning material with Xanthomonas')

 ORDER BY b.frop_name ASC;
```

This first code outputs the date of the operation, its number, its name, the tractor used on it, when did it start and finish, as well as its duration, the amount of fuel consumed, the implement used for it, and the efficiencies in terms of quantities and fuel. It was designed using the typical SELECT, FROM, WHERE, GROUPBY, HAVING, ORDERBY, LIMIT structure of the SQL language.

In the select part, the columns that are written (both calculated and from a table), are the ones that will appear on the resultant table; in the FROM clause, the table(s) where the data will come is presented. In this case, Left Inner Joins were made, also, with the so called Derived Table. In other words, what the code here is telling is to choose from operations data, where there is a sub-operation name, but also from the derived table that relates the sub-operation with and implement, and with the tractor.

In the WHERE part, all the restrictions (i.e. filters) that were discussed previously are taken into account; and, finally, ORDER BY is ordering in an ascendant way the name of the operation.

On the other hand, the code used for establishing the efficiencies that were already in the database was the following:

```
SELECT TOP (1000000)

a.[fropeff_operation] , c.frop_name, a.[fropeff_efficiency]
b.task_efficiency_real, b.task_qta, c.frop_um, a.[fropeff_machine],
a.[fropeff_impl1], a.[fropeff_impl2] ,a.[fropeff_consump],
a.[fropeff_fueltype],b.task_tot_days ,b.task_num_perm, b.task_num_temp,
b.task_num_contr, b.task_tot_man_days, b.task_tot_machine_days,
b.task_ut_type, b.task_tot_machines

FROM [DBMeteo].[dbo].[ferr_budget_anag_efficiency] a
     LEFT JOIN [DBMeteo].[dbo].[ferr_budget_task] b
     ON a.fropeff_id= b.task_efficiency
     LEFT JOIN  [DBMeteo].[dbo].[ferr_operations] c
     ON a.fropeff_operation=c.frop_id
WHERE c.frop_um in ('ha','pl') AND a.fropeff_company=53
       AND b.task_tot_days IS NOT NULL
       AND  b.task_tot_machines IS NOT NULL
       AND c.frop_name IN ('Application of fertilizer on Hazelnuts trees',

'Cleaning across lines',
  'Cleaning around plants',
  'Cleaning collector channels',
  'Cleaning inner channels',
  'Foliar fertilization ',
  'Harrowing',
  'Herbicide application between hazelnuts lines',
  'Herbicide application inner & outer channels',
  'Herbicide application on hazelnut line',
  'Levelling across the rows',
```

```
   'Levelling before planting',
   'Levelling between the rows',
   'Machine collection 1st',
   'Machine collection 2nd',
   'Machine collection last pass',
   'Miscellaneous Harvesting',
   'Mulcher',
   'Nursery weeding',
   'Pesticide application',
   'Post-pruning application',
   'Preparation with blowers',
   'Pruning - Mainteinance - High Pruning (European varieties)',
   'Pruning - Mainteinance - High Pruning (Young trees)',
   'Ridging plants in nursery',
   'Ripper',
   'Substrate mixture ',
   'Suckers removal spraying',
   'Supervising irrigation work',
   'Transport of pruning material',
   'Transport of pruning material with Xanthomonas')

ORDER BY a.fropeff_operation;
```

The logic is the same than what was explained for the past code, so no explanation in particular is given for this one.

Once the data was obtained, the efficiencies were computed and the Maximum number of workers, estimated. For the former, there was some criteria taken into account:

- Since the same combination of sub-operation/ machine appeared multiple times in Budget Efficiency, with different quantities of hectares or number of plants, then the maximum number of workers for the combination was set as the max. between them.
- Historically, from operations data, each operation is done by one worker (this is true if the operation is mechanic). However, information about which machine did the sub-operation use, was known. Then, the maximum number of workers was defined as the maximum number of machines for that particular family of machine.
- If the combination of Budget Efficiency matched with the one obtained from operations data, the maximum number of workers between the two of them, was taken (this happen in very little cases).

Nonetheless the Maximum number of workers was displayed, this information is just seen as an approximative representation of the reality, and **was not** used in the algorithm, since it intends to create a scenario of data input as the one of operations data, and for that scenario, the quantity of workers for mechanical activities is always one.

14.4.3. Tractors and Implements.

The database contains all the machines for all the companies in different countries where 3a S.r.l has operations. In this sense, a filter for the stock of machines in Chile had to be made. However, there was no way of finding the number of vehicles and instruments per single company; and, in addition, the companies in this particular country lend them between each other, so there is no a clear view of where the machines belong. To overcome this issue, historical data from operations data was analyzed for San Sebastian (from year 2014 to year 2018), so that, all the machines used with its respective code and name were known, and an approximation to the actual stock became feasible.

Furthermore, different queries were written (as shown below), to know the tractors, the implements, and the maximum number of them, per family in the database.

```
SELECT TOP (1000000)

frimpl_family, COUNT (frimpl_family) AS NumberImplementsFamily

FROM [DBMeteo].[dbo].[ferr_an_implements]

WHERE frimpl_company_code='HBD_CHILE' AND frimpl_id IS NOT NULL

GROUP BY frimpl_family;
```

This code displays the number of implements per family in the database. In order to look at the complete table, what appears between SELECT TOP (1000000) AND FROM, should be replaced by *.

```
SELECT TOP (100000)

[tractor_family], COUNT(tractor_family) AS NumberTractorsFamily

FROM [DBMeteo].[dbo].[ferr_an_tractor]

WHERE tractor_company_code='HBD_CHILE' AND tractor_code IS NOT NULL;
```

On the other hand, the aforementioned query shows the number of tractors per family. To look at the complete table, like before, what appears between SELECT TOP (1000000) AND FROM, should be replaced by *.

14.5. Input Analysis

From the picture seen in the *Context* subsection, it is more or less clear how data has to be entered. However, the way it is entered, and to what time period does it corresponds to, needs to be analyzed deeply, since it is considered as the core for the assignation process in the program.

Because of how the algorithm was conceived, its scope does not go further in time. Instead, it deals with what has historically took place in the form of operations from the years 2015, until now. (Nonetheless, the current research intends to be used as a basis for future lines, where the scheduling of the activities would be done in a higher scale by the agricultural team, and the program would assign the machinery in an efficient way). With this in mind, the input information for the operations and their dates, looks as in *Figure 33*.

OPERATIONS TO BE SCHEDULED					
frop_date	frop_operation	frop_name	frop_um	frop_sum_qta	Zone
03/01/2017 10:48	607	Pesticide application	ha	6,6	1
03/01/2017 10:52	650	Mulcher	ha	7,21	4
03/01/2017 10:53	650	Mulcher	ha	6	4
04/01/2017 13:44	607	Pesticide application	ha	6,5	3
04/01/2017 13:54	607	Pesticide application	ha	7,68	4
04/01/2017 13:56	607	Pesticide application	ha	7,68	4
04/01/2017 13:57	607	Pesticide application	ha	6,5	4

Figure 33.Input Data. Source: Own Elaboration.

The fact that the operation corresponds to a single day, with the exact time, and with a quantity such that, it will be completed in a maximum of one day, does not escape to the eye. In fact, this responds to the logic behind which the tables in the database were created: Everyday, the workers will type in the start time, the quantity done, the duration, and the machines used. In this sense, the assignation of resources is done on a daily basis and bearing in mind the fact that the quantities have been done historically, no unfeasible solutions are taken into account in terms of time intervals to be respected for an operation to be accomplished. In addition, and because of the daily basis, the work done and inserted into the database, corresponds to the one of one worker, named the driver of the machine.

On the other hand, the table that contains information about the possible combination of machines that have been used or could be used to perform a certain operation is the following one.

Max Number of Workers (Taken from Merging Budget_anag_efficiency with Budget_task_efficiency + Historical Data)

frop_operation	frop_name	frop_tractor	tractor_name	tractor_family	frimpl_family1	frimpl_family2	FuelEfficiencyHours	QuantityEfficiencyHours	frop_um	Zone
541	Application o	601	TRACTOR PAT.	25	0		2,753	0,517	ha	2
541	Application o	612	Tractor PAT. YS	31	0		4,517	0,654	ha	4
541	Application o	589	Moto Yamaha J	5	0		0,182	0,455	ha	4
541	Application o	533	MOTO ENDURO	5	0		0,053	0,310	ha	3
814	Cleaning aro	1304	MOTO 4X4 HON	5	0		0,588	0,018	ha	1
706	Cleaning coll	650	RETROEXCAVA	95	0		8,518	0,635	ha	4

Figure 34. Machines and Implements per Sub-operation, with their Efficiency in Terms of Duration and Fuel Consumption. Source: Own Elaboration.

This information was obtained using the first query explained in the subsection *Finding the Efficiencies,* but this time for all the years of operation. Once all the operations were deployed, they were categorized by group, and the mean value was used for Fuel Efficiency and Quantity Efficiency. Moreover, some combinations were already defined in the database, so they were also used to complete the list.

In different words, for the Sub-operation 541, which corresponds to *Application of Fertilizer on Hazelnut Trees*, the only combinations allowed, both from historical retrieval and from the database, are those performed with the tractor 601, of the tractor family 25; or with the tractors 612, 589, 533, 1304, from the tractor family 31, 5, 5, and 5, respectively, with no Implements assigned.

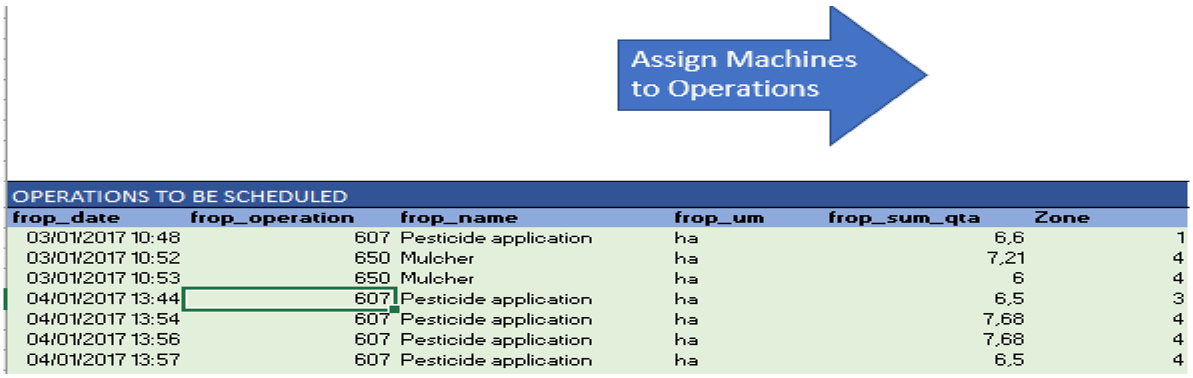
It is also important to highlight that right here, there is a list of the stock of resources to use. If one is already assigned, it will be marked as "on use" for the duration of its usage, and once it is free, it could be re-assigned to a different operation.

Additionally, the total list of the tractors available and implements, with their respective quantity within a family is known in advanced as well. However, the screenshots will not be shown here, since they are already in the Excel File.

In addition, the zones where the operation needs to be performed is also taken into account, assigned with a layout discussed with the enterprise and setting their distances in terms of time of transportation of the machines. Since the distances are impossible to know in advanced, different values were taken to evaluate performance, and the median one (60 minutes) was used for comparisons.

14.6. Logic Behind the Algorithm

Having a better understanding of how the input will be entered into the algorithm, it is then fair enough to analyze the logic behind it, step by step. But before, just to clarify what has been describing just with words so far, the interface of the deliverable looks like this:



OPERATIONS TO BE SCHEDULED					
frop_date	frop_operation	frop_name	frop_um	frop_sum_qta	Zone
03/01/2017 10:48	607	Pesticide application	ha	6,6	1
03/01/2017 10:52	650	Mulcher	ha	7,21	4
03/01/2017 10:53	650	Mulcher	ha	6	4
04/01/2017 13:44	607	Pesticide application	ha	6,5	3
04/01/2017 13:54	607	Pesticide application	ha	7,68	4
04/01/2017 13:56	607	Pesticide application	ha	7,68	4
04/01/2017 13:57	607	Pesticide application	ha	6,5	4

Figure 35. Interface of the Application. Source: Own Elaboration.

When clicking in the arrow, the macro (Visual Basic) will start to run. What users have to type in are the dates when the operation was done, the operation itself, the unit of measure, and the quantity of that unit of measure to perform.

Once done , the program starts with a main Sub, named *Greedy,* since this category was the base of how the algorithm works. A Greedy algorithm makes the optimal choice at each small stage, picking the best solution at the moment regardless of the consequences. In other words, it picks the best immediate output, but does not take into account the whole picture. The logic is that it hopes to take the optimal path, leading it to the end. As a drawback, it might not be the best method, and a large amount of memory is required, compared to other alternatives (Techopedia.com, 2018).

This Sub is the structure for everything, since from here, all the other Subs are called to do the computations and iterations required. The only thing that is valid throughout the whole program and is not declared here, are the Global variables. Below, the Global variables are shown, and then the architecture of the Greedy is deployed.

```
Option Explicit
Option Base 1
```

```
Dim DatesOperations(), DatesOperationsCopy() As Variant 'Copies are created so that the data can be organized by date
Dim OperationNumber(), OperationNumberCopy() As Integer
Dim OperationName(), OperationNameCopy() As String
Dim UnitOfMeasure(), UnitOfMeasureCopy() As String
Dim SumQuantity(), SumQuantityCopy() As Double
Dim ZoneOperation(), ZoneOperationCopy() As Integer 'keeps track of the zone where the tractor and the implements are
Dim Distance() As Double 'Matrix that reflects the time from going to one zone to another
Dim aux, aux1 As Variant
Dim NewDates() As Variant 'Keeps track of the dates of the operations that are rescheduled due to not availability of the tractor

Dim Wage As Double
Dim MaxImplements() As Integer 'per family
Dim MaxTractors() As Integer 'per family

'Dim Tractors() As Integer 'matrix that contains the family number and the number of the tractor corresponding to the family
'Dim Implement() As Integer 'Array with the implements

Type Result
   TractorResult As Integer
   TractorNameResult As String
   TractorFamilyResult As Integer
   Implement1FamilyResult As Integer
   Implement2FamilyResult As Integer
   Duration As Double 'Reflects the duration of the task in hours
   FuelResult As Double
```

```
    Zone As Integer
    'NumberOfWorkers As Integer
    'NumberOfMachines As Integer
End Type

Dim Results() As Result

Type Efficiencies 'Designed for storing what historically has happened in terms of the combination of workers, machines, and implements
    OperationEfficiency As Integer
    OperationNameEfficiency As String
    TractorEfficiency As Integer
    TractorNameEfficiency As String
    TractorFamilyEfficiency As Integer
    Implement1FamilyEfficiency As Integer
    Implement2FamilyEfficiency As Integer
    'MaxNumberWorkers As Integer
    FuelEfficiency As Double
    QuantityEfficiency As Double
    ZoneTractor As Integer
    NumberTimes As Double
End Type

Dim Efficiency() As Efficiencies
```

The first 5 lines correspond to the variables containing the data that was shown at the beginning of this section (Dates of Operations, Operations, Unit of Measure, and their Quantities). In addition, the Type "Results", reflects how the outcomes of the algorithm will be arranged. Although the variable "Results" is just an array, it is like if it contained more variables within it. The Type "Efficiencies", on the other hand, contains the data that was discussed in the section "Finding Efficiencies".

```
Sub Greedy()

Dim i, j As Integer
Dim NumberOperationsDone As Double
Dim NumEfficiencies As Integer
Dim Available() As Integer 'Array that checks availability of the machine
Dim DateOfAssignation() As Variant 'array that will keep track of when the machine and
the implements were assigned
Dim TrackDuration() As Double 'Array that keeps track of the duration of use of the
machine and the implements
Dim TotalDuration As Double ' Sum of All Durations
Dim TotalFuelConsumption As Double 'Sum of Fuel consumption

NumberOperationsDone = Range("Input").Rows.Count
NumEfficiencies = Range("MaxNumWorkers").Rows.Count

ReDim Results(1 To NumberOperationsDone)
ReDim Available(1 To NumEfficiencies)
ReDim DateOfAssignation(1 To NumEfficiencies)
ReDim TrackDuration(1 To NumEfficiencies)
ReDim NewDates(1 To NumberOperationsDone)

Call ReadData
Call OrganizedByDate(NumberOperationsDone)

For i = 1 To NumEfficiencies
  Available(i) = 0 'at first, all the machines are available. 0 Means available, 1 not available
Next i

For i = 1 To NumberOperationsDone
   If DatesOperationsCopy(i) <> "" Then
```

```
        Call FillResults(i, Available, DateOfAssignation, TrackDuration)

    Else
        Exit For
    End If

Next i

'Computing for total

For i = 1 To NumberOperationsDone
    If DatesOperationsCopy(i) <> "" Then
    TotalDuration = TotalDuration + Results(i).Duration
    TotalFuelConsumption = TotalFuelConsumption + Results(i).FuelResult

    End If
Next i

'================================
'           Printing
'================================



Application.ScreenUpdating = False ' Avoids updating everytime it prints

For i = 1 To NumberOperationsDone
    If DatesOperationsCopy(i) <> "" Then
        Range("NewDate").Offset(i, 0).Value = NewDates(i)
        Range("frop_tractor").Offset(i, 0).Value = Results(i).TractorResult
        Range("tractor_name").Offset(i, 0).Value = Results(i).TractorNameResult
```

```
        Range("tractor_family").Offset(i, 0).Value = Results(i).TractorFamilyResult
        Range("frimpl1_family").Offset(i, 0).Value = Results(i).Implement1FamilyResult
        Range("frimpl2_family").Offset(i, 0).Value = Results(i).Implement2FamilyResult
        Range("Duration").Offset(i, 0).Value = Results(i).Duration
        Range("frop_fuel").Offset(i, 0).Value = Results(i).FuelResult

    Else
        Exit For
    End If
Next i

Range("Time").Offset(1, 0).Value = TotalDuration
Range("Consumption").Offset(1, 0).Value = TotalFuelConsumption

End Sub
```

After declaring variables that will exist only in the sub, and, if passed as parameters, in the subs that are called from inside the “Greedy”, the reading of the data is performed, and this is done throughout the sub “ReadData”.

```
Sub ReadData()
Dim i, j As Integer
Dim NumberOperationsDone, NumEfficiencies As Double
Dim NumberOfZones As Integer

NumberOperationsDone = Range("Input").Rows.Count
NumEfficiencies = Range("MaxNumWorkers").Rows.Count
NumberOfZones = Range("NumberOfZones").Cells.Value
```

```
ReDim DatesOperations(1 To NumberOperationsDone), OperationNumber(1 To
NumberOperationsDone)
ReDim OperationName(1 To NumberOperationsDone), UnitOfMeasure(1 To
NumberOperationsDone), SumQuantity(1 To NumberOperationsDone)
ReDim Results(1 To NumberOperationsDone)
ReDim Efficiency(1 To NumEfficiencies)
ReDim MaxImplements(1 To Range("MaxNumImplements").Rows.Count,
Range("MaxNumImplements").Columns.Count)
ReDim MaxTractors(1 To Range("MaxNumTractors").Rows.Count,
Range("MaxNumTractors").Columns.Count)
ReDim Distance(1 To NumberOfZones, 1 To NumberOfZones)
ReDim ZoneOperation(1 To NumberOperationsDone)



For i = 1 To NumberOperationsDone
    DatesOperations(i) = Range("Input").Cells(i, 1).Value
    OperationNumber(i) = Range("Input").Cells(i, 2).Value
    OperationName(i) = Range("Input").Cells(i, 3).Value
    UnitOfMeasure(i) = Range("Input").Cells(i, 4).Value
    SumQuantity(i) = Range("Input").Cells(i, 5).Value
    ZoneOperation(i) = Range("Input").Cells(i, 6).Value

Next i



For i = 1 To NumEfficiencies
    Efficiency(i).OperationEfficiency = Range("MaxNumWorkers").Cells(i, 1).Value
    Efficiency(i).OperationNameEfficiency = Range("MaxNumWorkers").Cells(i, 2).Value
    Efficiency(i).TractorEfficiency = Range("MaxNumWorkers").Cells(i, 3).Value
    Efficiency(i).TractorNameEfficiency = Range("MaxNumWorkers").Cells(i, 4).Value
    Efficiency(i).TractorFamilyEfficiency = Range("MaxNumWorkers").Cells(i, 5).Value
    Efficiency(i).Implement1FamilyEfficiency = Range("MaxNumWorkers").Cells(i, 6).Value
    Efficiency(i).Implement2FamilyEfficiency = Range("MaxNumWorkers").Cells(i, 7).Value
    'Efficiency(i).MaxNumberWorkers = Range("MaxNumWorkers").Cells(i, 8).Value
```

```
    Efficiency(i).FuelEfficiency = Range("MaxNumWorkers").Cells(i, 9).Value
    Efficiency(i).QuantityEfficiency = Range("MaxNumWorkers").Cells(i, 10).Value
    Efficiency(i).ZoneTractor = Range("MaxNumWorkers").Cells(i, 12).Value
    Efficiency(i).NumberTimes = Range("MaxNumWorkers").Cells(i, 13).Value
Next i

For i = 1 To NumberOfZones
    For j = 1 To NumberOfZones
        Distance(i, j) = Range("Distance").Cells(i, j).Value
    Next j
Next i

For i = 1 To Range("MaxNumImplements").Rows.Count
    For j = 1 To Range("MaxNumImplements").Columns.Count
        MaxImplements(i, j) = Range("MaxNumImplements").Cells(i, j).Value
    Next j
Next i

For i = 1 To Range("MaxNumTractors").Rows.Count
    For j = 1 To Range("MaxNumTractors").Columns.Count
        MaxTractors(i, j) = Range("MaxNumTractors").Cells(i, j).Value
    Next j
Next i

End Sub
```

The variables are re-dimensioned to the size of the number of operations done, and to the one of the number of the recorded cases where the efficiency was found. Usually it is done with a loop, if it is an array, or with a nested loop, if it is a matrix.

After the initialization, the program turns back to the “Greedy”, where another sub is called, this time the “OrganizedByDate” one, passing the variable “NumberOperationsDone” as a parameter.

```
Sub OrganizedByDate(ByVal NumberOperationsDone As Double)

Dim i, j As Integer

ReDim DatesOperationsCopy(1 To NumberOperationsDone)
ReDim DatesOperationsOrganized(1 To NumberOperationsDone)
ReDim OperationNumberCopy(1 To NumberOperationsDone)
ReDim OperationNameCopy(1 To NumberOperationsDone)
ReDim UnitOfMeasureCopy(1 To NumberOperationsDone)
ReDim SumQuantityCopy(1 To NumberOperationsDone)
ReDim ZoneOperationCopy(1 To NumberOperationsDone)

For i = 1 To NumberOperationsDone 'A copy of the date is made so it will not modify the
original array of dates
   DatesOperationsCopy(i) = Range("Input").Cells(i, 1).Value
Next i

For i = 1 To NumberOperationsDone
   DatesOperationsOrganized(i) = i 'An array that keeps the position of the Original Date
Next i



For i = 1 To NumberOperationsDone
   For j = 1 To NumberOperationsDone
      If DatesOperationsCopy(i) < DatesOperationsCopy(j) And i <> j And
DatesOperationsCopy(i) <> "" _
         And DatesOperationsCopy(j) <> "" Then

         aux = DatesOperationsCopy(i)
         DatesOperationsCopy(i) = DatesOperationsCopy(j)
         DatesOperationsCopy(j) = aux
```

```
        aux1 = DatesOperationsOrganized(i)
        DatesOperationsOrganized(i) = DatesOperationsOrganized(j)
        DatesOperationsOrganized(j) = aux1
      End If
   Next j
Next i

For i = 1 To NumberOperationsDone

OperationNumberCopy(i) = OperationNumber(DatesOperationsOrganized(i))
OperationNameCopy(i) = OperationName(DatesOperationsOrganized(i))
UnitOfMeasureCopy(i) = UnitOfMeasure(DatesOperationsOrganized(i))
SumQuantityCopy(i) = SumQuantity(DatesOperationsOrganized(i))
ZoneOperationCopy(i) = ZoneOperation(DatesOperationsOrganized(i))

Next i

'=================================
'          Printing
'=================================

'Deletes the data (if any) before printing

Range("G14:T2513").Select
Selection.ClearContents

Application.ScreenUpdating = False ' Avoids updating everytime it prints

For j = 1 To NumberOperationsDone
   If DatesOperationsCopy(j) <> "" Then
```

```
        Range("Sequence").Offset(j, 0).Value = j 'CountDown of the Number of Operations
Scheduled
        Range("frop_date").Offset(j, 0).Value = DatesOperationsCopy(j)
        Range("Operation").Offset(j, 0).Value = OperationNumberCopy(j)
        Range("OperationName").Offset(j, 0).Value = OperationNameCopy(j)
        Range("frop_um").Offset(j, 0).Value = UnitOfMeasureCopy(j)
        Range("frop_sum_qta").Offset(j, 0).Value = SumQuantityCopy(j)
        Range("Zone").Offset(j, 0).Value = ZoneOperationCopy(j)
    Else
        Exit For
    End If
Next j



End Sub
```

What is going on here is that the operations are sorted in an ascending order by date, following the logic that, if in the list inserted manually of operations to be scheduled with machines, two or more operations could be done with the same machine, then, the machine will be assigned to the one earlier, hopping that, once the other operation needs to be done, the machine will be available. Also, the copy for the first five variables declared under the Global Declarations is filled (the copy corresponds to the Operations organized by dates). Furthermore, the organized vectors are printed into the Excel Worksheet.

Next, the program goes back to the "Greedy", to first initialize the vector "Availability", composed by ones and zeros, and that will tell the algorithm, what combination for efficiency is available: zero meaning Available and one Not Available, and then, it goes into a Loop with i=NumberOperationsDone times, as long as the data in the table is not zero (this because there was a limit defined of 2500 operations to enter. Of course, this limit can be replaced; it was just a big number seeking to avoid size problems). In the loop, the Sub "FillResults" is called, having as parameters the variables "i", "Available", "DateOfAssignation", and "TrackDuration".

```
Sub FillResults(ByVal num As Double, ByRef Available() As Integer, ByRef DateOfAssignation() As Variant, _
            ByRef TrackDuration() As Double)

Dim i, j As Integer
Dim NumEfficiencies As Double
Dim Objective() As Result 'Variable that stores the FO in the position it is computed for each operation
Dim KeepTrack As Integer 'keeps track of the position of the tractor in the list efficiencies

NumEfficiencies = Range("MaxNumWorkers").Rows.Count

ReDim Objective(1 To 100)

For i = 1 To 100
   Objective(i).Duration = 0
   Objective(i).FuelResult = 0
   Objective(i).TractorResult = 0
   Objective(i).TractorNameResult = ""
   Objective(i).TractorFamilyResult = 0
   Objective(i).Implement1FamilyResult = 0
   Objective(i).Implement2FamilyResult = 0
   Objective(i).Zone = 0
Next i
j = 1
KeepTrack = 0

For i = 1 To NumEfficiencies

   If OperationNumberCopy(num) = Efficiency(i).OperationEfficiency Then
      Call computeFO(Objective, num, i, j)
```

```
    j = j + 1
  Else
    If j = 1 Then
      KeepTrack = KeepTrack + 1
    End If
  End If
Next i

Call BestOption(Objective, num, j, NumEfficiencies, Available, DateOfAssignation, TrackDuration, KeepTrack)

End Sub
```

The variable “Objective” is created as a Type “Result” to stores the result in an array, once all the possible machines able to perform the operation are found within the list of “Efficiencies”. It is set with a dimension of 100, because, in the last list, there is not a same group of possible combinations for operations to be done that surpasses this value. Also, the variables “j” and “keepTrack” are created. The first one will start counting from the time the operation is found within the list “Efficiencies”; in other words, if the operation 706 is in the list for the first time in the position 6, then, the variable “j” will start counting from this point, until the operation is no longer found. On the other hand, the variable “KeepTrack” will count the number of positions it takes to reach for the first time the j=1, i.e. 5 for this example.

The sub “ComputeFO” will start to iterate within a cycle of i=NumEfficiencies times. This guaranties that all the members of the list “Efficiencies” are analyzed. It has as parameters, “Objective”, “num”, “i”, “j”.

```
Sub computeFO(ByRef Objective() As Result, ByVal num As Integer, ByVal i As Integer, ByVal j As Integer)

Dim TemporaryDuration() As Double
```

```
Dim TemporaryFuel As Double 'stores the FO comparing with a bigger number of workers. It takes the best
Dim TemporaryMachine, TemporaryImplement As Integer
Dim TransportDuration As Double 'How much it takes to transport the machine to where it is needed

'locates the duration of the transportation between where the tractor is, and where the operation needs to be done
TransportDuration = Range("Distance").Cells(ZoneOperationCopy(num), Efficiency(i).ZoneTractor).Value



Objective(j).Duration = (SumQuantityCopy(num) / Efficiency(i).QuantityEfficiency) + TransportDuration
Objective(j).FuelResult = (Efficiency(i).FuelEfficiency) * Objective(j).Duration
Objective(j).TractorResult = Efficiency(i).TractorEfficiency
Objective(j).TractorNameResult = Efficiency(i).TractorNameEfficiency
Objective(j).TractorFamilyResult = Efficiency(i).TractorFamilyEfficiency
Objective(j).Implement1FamilyResult = Efficiency(i).Implement1FamilyEfficiency
Objective(j).Implement2FamilyResult = Efficiency(i).Implement2FamilyEfficiency
Objective(j).Zone = ZoneOperationCopy(num)



End Sub
```

This particular sub does not require a lot of explanation: it computes the array of results as many times, as the operation appears in the list of “Efficiencies”. These computations are stored in the variable “Objective” to then go back to the sub “FillResults”. Afterwards, the Sub “BestOption” is called, having the parameters “Objective”, “num”, “j”, “NumEfficiencies”, “Available”, “DateOfAssignation”, “TrackDuration”, and “KeepTrack”.

```
Sub BestOption(ByRef Objective() As Result, ByVal num As Double, ByVal num2 As Integer, ByVal NumEfficiencies As Double, _
                ByRef Available() As Integer, ByRef DateOfAssignation() As Variant, ByRef TrackDuration() As Double, _
                 ByRef KeepTrack As Integer)

Dim i, j, k, l As Integer
Dim ObjectiveCopy() As Result
Dim aux As Double
Dim aux1, OrganizedObjective() As Integer
Dim AvailableCopy() As Integer 'Created for the case when all the tractors are not available

ReDim ObjectiveCopy(1 To 100), OrganizedObjective(1 To 100)
ReDim AvailableCopy(1 To NumEfficiencies)

ObjectiveCopy = Objective

For j = 1 To 100
   OrganizedObjective(j) = j 'keeps track of the position (just for the portion of the array where the tractor is)
Next j

'=========================
'Ordering by duration
'========================



For j = 1 To 100
   For k = 1 To 100
      If ObjectiveCopy(j).Duration < ObjectiveCopy(k).Duration And j <> k _
        And ObjectiveCopy(j).Duration <> 0 And ObjectiveCopy(k).Duration <> 0 Then
```

```
            aux = ObjectiveCopy(j).Duration
            ObjectiveCopy(j).Duration = ObjectiveCopy(k).Duration
            ObjectiveCopy(k).Duration = aux

            aux1 = OrganizedObjective(j)
            OrganizedObjective(j) = OrganizedObjective(k)
            OrganizedObjective(k) = aux1



        End If
    Next k
Next j

'It just fills what has values. Otherwise, is just zero
For i = num2 To 100
    OrganizedObjective(i) = 0
Next i

For j = 1 To num2 - 1
    ObjectiveCopy(j).FuelResult = ObjectiveCopy(OrganizedObjective(j)).FuelResult
    ObjectiveCopy(j).TractorResult = ObjectiveCopy(OrganizedObjective(j)).TractorResult
    ObjectiveCopy(j).TractorNameResult =
ObjectiveCopy(OrganizedObjective(j)).TractorNameResult
    ObjectiveCopy(j).TractorFamilyResult =
ObjectiveCopy(OrganizedObjective(j)).TractorFamilyResult
    ObjectiveCopy(j).Implement1FamilyResult =
ObjectiveCopy(OrganizedObjective(j)).Implement1FamilyResult
    ObjectiveCopy(j).Implement2FamilyResult =
ObjectiveCopy(OrganizedObjective(j)).Implement2FamilyResult
    ObjectiveCopy(j).Zone = ObjectiveCopy(OrganizedObjective(j)).Zone
```

```
Next j




'=============================================
'Assignation of the Machines and implements
'=============================================



For i = 1 To NumEfficiencies
   If OrganizedObjective(i) > 0 Then 'checks if all the tractors are assigned
      If Available(KeepTrack + OrganizedObjective(i)) = 0 Then
         Results(num).Duration = ObjectiveCopy(i).Duration
         Results(num).FuelResult = ObjectiveCopy(i).FuelResult
         Results(num).TractorResult = ObjectiveCopy(i).TractorResult
         Results(num).TractorNameResult = ObjectiveCopy(i).TractorNameResult
         Results(num).TractorFamilyResult = ObjectiveCopy(i).TractorFamilyResult
         Results(num).Implement1FamilyResult = ObjectiveCopy(i).Implement1FamilyResult
         Results(num).Implement2FamilyResult = ObjectiveCopy(i).Implement2FamilyResult
         Results(num).Zone = ObjectiveCopy(i).Zone

         Available(KeepTrack + OrganizedObjective(i)) = 1 'Mark as Machine being used
         DateOfAssignation(KeepTrack + OrganizedObjective(i)) = DatesOperationsCopy(num)
'The date of assignation is set
         TrackDuration(KeepTrack + OrganizedObjective(i)) = ObjectiveCopy(i).Duration

         Exit For
      End If
   Else 'If all the tractors are assigned, the operation will be moved by 1 minute, until one tractor
is available again
```

```
        AvailableCopy = Available

        Call Availability(AvailableCopy, TrackDuration, DateOfAssignation,
OrganizedObjective, KeepTrack, Available, _
                            ObjectiveCopy, num, NumEfficiencies)

        Exit For

    End If
Next i

'==========================================================================
================================================
'                       All the tractors in the list marked as being used
'==========================================================================
=================================================

For i = 1 To NumEfficiencies
 If Results(num).TractorResult = Efficiency(i).TractorEfficiency Then
    Available(i) = 1
    DateOfAssignation(i) = DatesOperationsCopy(num)
    TrackDuration(i) = Results(num).Duration
    Efficiency(i).ZoneTractor = Results(num).Zone
 End If
Next i


'==========================================
'    Restart of the Availability
'==========================================
```

```
For i = 1 To NumEfficiencies
   If Available(i) = 1 Then
         If DateOfAssignation(i) + (TrackDuration(i)) / 24 <= DatesOperationsCopy(num) _
           And DatesOperationsCopy(num) <> DateOfAssignation(i) Then

            Available(i) = 0

         End If

   End If
Next i



End Sub
```

This is probably the core sub of the whole program. First, the array of "Objective" is organized according to the criteria of having Minimum Duration. This way, the combination of the "Efficiencies" table that will guarantee this condition are positioned first. This is all done with the variable "ObjectiveCopy". Also, the original position is kept in the variable "OrganizedObjective". As stated before, the variable "Objective", and therefore "ObjectiveCopy" has a size of 100. Here, for all the numbers beyond the size of the number of times that the operation was found in the list, a value of zero or null is assigned.

The next step is the assignation of the machines: it will be done if and only if the value for the position is neither zero, nor null (that is why the precedent step was so important). And it will work if the machines for that particular position are available. If not, the next best combination in terms of duration will be evaluated and then assigned, if available. The process will keep going until an available one is found. Once assigned, the position will be marked, and the machine there will be set as "Not Available", also, the "DateOfAssignation" in the correspondent position will be set as well as the "Duration" in that position.

Following the above, the algorithm makes sure that, if a particular tractor is assigned, and it corresponds to a position in the list of "Efficiencies", then all the same tractors will be marked as "Not available" for as long as the Operation assigned previously is being done. However, once this time finishes, the availability has to be restarted, so that the assignation process could be done once more, this time for a new operation.

To restart the availability, a comparison is made between the date of start of the operation under study, and the time elapsed since the machine has been assigned.

Nonetheless, if at some point, the number of iterations finish, and no machines are available (even the worse one) for a particular operation, starting in a specific date, then, the algorithm reschedules the operation to start in another time. This is done minute by minute, until the first machine is available. This particular scenario is evaluated under the sub Availability, which has as parameters "AvailableCopy", "TrackDuration", "DateOfAssignation","OrganizedObjective", "KeepTrack", "Available", "ObjectiveCopy", and "NumEfficiencies".

```
Sub Availability(ByRef AvailableCopy() As Integer, TrackDuration() As Double, ByRef DateOfAssignation(), _
                ByRef OrganizedObjective() As Integer, ByVal KeepTrack As Integer, ByRef Available() As Integer, _
                ByRef ObjectiveCopy() As Result, ByVal num As Double, ByVal NumEfficiencies As Double)



Dim i, j As Integer



For i = 1 To NumEfficiencies

    If OrganizedObjective(i) > 0 Then 'checks if all the tractors are assigned
        If AvailableCopy(KeepTrack + OrganizedObjective(i)) = 0 Then

            Results(num).Duration = ObjectiveCopy(i).Duration
```

```
            Results(num).FuelResult = ObjectiveCopy(i).FuelResult
            Results(num).TractorResult = ObjectiveCopy(i).TractorResult
            Results(num).TractorNameResult = ObjectiveCopy(i).TractorNameResult
            Results(num).TractorFamilyResult = ObjectiveCopy(i).TractorFamilyResult
            Results(num).Implement1FamilyResult =
ObjectiveCopy(i).Implement1FamilyResult
            Results(num).Implement2FamilyResult =
ObjectiveCopy(i).Implement2FamilyResult
            Results(num).Zone = ObjectiveCopy(i).Zone

            Available(KeepTrack + OrganizedObjective(i)) = 1 'Mark as Machine being used
            DateOfAssignation(KeepTrack + OrganizedObjective(i)) =
DatesOperationsCopy(num) 'The date of assignation is set
            TrackDuration(KeepTrack + OrganizedObjective(i)) = ObjectiveCopy(i).Duration

            Exit For
        End If
    Else

   DatesOperationsCopy(num) = DatesOperationsCopy(num) + (1 / 1440)
   NewDates(num) = DatesOperationsCopy(num) 'variable just to keep track of when the
operations were rescheduled

    For j = 1 To NumEfficiencies
        If AvailableCopy(j) = 1 Then
             If DateOfAssignation(j) + (TrackDuration(j)) / 24 <= DatesOperationsCopy(num)
_
              And DatesOperationsCopy(num) <> DateOfAssignation(i) Then

               AvailableCopy(j) = 0
```

```
            End If
        End If
    Next j

    i = 0
    End If
Next i


End Sub
```

Finally, the program goes back to the “Greedy” to print the results.

The following Flowchart is presented for a better understanding of the code above, so that a graphical sequence will clarify what was already explained by words.

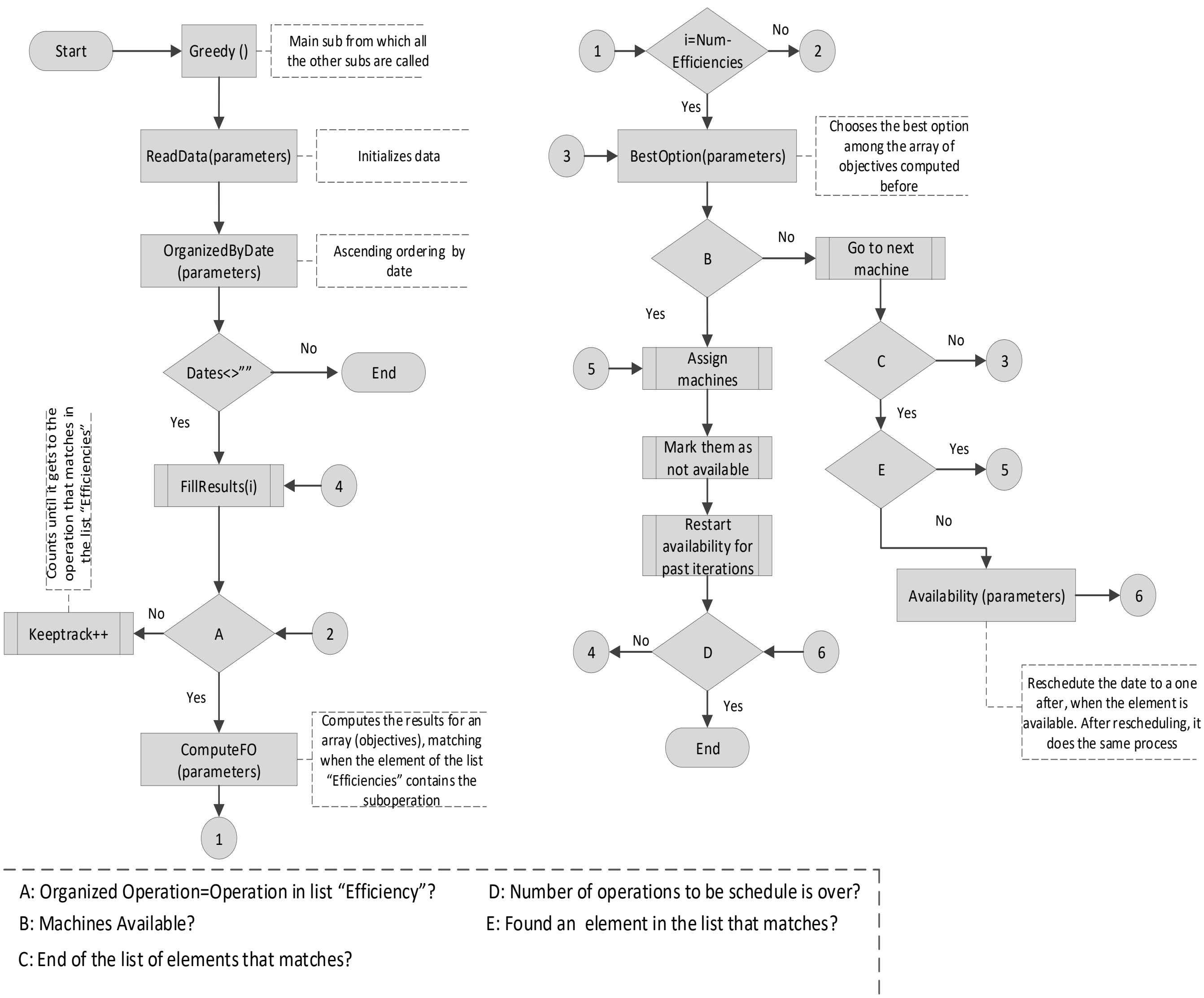


Figure 36. Flowchart of the Logic Behind the Algorithm. Source: Own Elaboration.

14.7.Comments on the Results

It was well known from the conception of the problem that the algorithm was not a perfect reflection of the reality of the operations; instead, it was just an approximation that could be used to help in the scheduling of the machines (tractors and implements). This, for multiple reasons, among which there is the impossibility to know by heart the real nature of the activities performed in the field, as well as some inconsistencies in the way data is entered into the database. However, to overcome this situation, some assumptions have been made through dialogs with the people in 3a in the coding of the algorithm, while others are made in the analysis of results in the enterprise (i.e. sitting side by side in 3a to compare and comment on the outcomes, and possible future lines of action).

Those assumptions and comments are enlisted below:

- It has been decided to optimize by duration, due to inconsistencies in the data related to Fuel consumption.
- For the operations that have a big difference between the outcome of the algorithm, and historical data, being that of the algorithm way better, it was found that different sub-activities are performed under the same code of operation, even though they differ, and as so, its performance (i.e. efficiency). However, it is not a common denominator, and the cases are weird and to be evaluated by the enterprise. In other words, taking as an instance the operation "cleaning collector channels", it could be done with a tractor, with a motorcycle, or with an excavator, because every machine corresponds to a different sub-task, so the efficiency will be different, and the algorithm will choose that with higher one, in search for an optimal result. This situation is not reflecting reality.
- If the date is rescheduled, it will be to the first moment when the implement and/or tractor are/is available; not taking into account if the daily shift is over. If this happens, the result needs to be evaluated by the company, so that the actual start of the operation is in the first available working date instead.
- When the algorithm doesn't print the tractor, but it does print the family, it means that any tractor of that family can perform the operation

14.8. Results

All the instances for the application described above have to be validated, and in addition, compared to the values obtained historically in terms of performance. As an instance, throughout this section, the algorithm is executed for the years 2015 to 2018, and the results obtained are contrasted with those obtained with the current assignation criteria, both for duration of operation and fuel consumption.

Furthermore, and as discussed before with the sub-activities, some operations were not considered, due to inconsistencies in the data insertion. Those were:

- Levelling between rows
- Cleaning collector channels
- Miscellaneous harvesting
- Preparation with blowers
- Supervising

Keeping track of this, the following table is shown year by year, in a way such that it is possible to make conclusions on the matter.

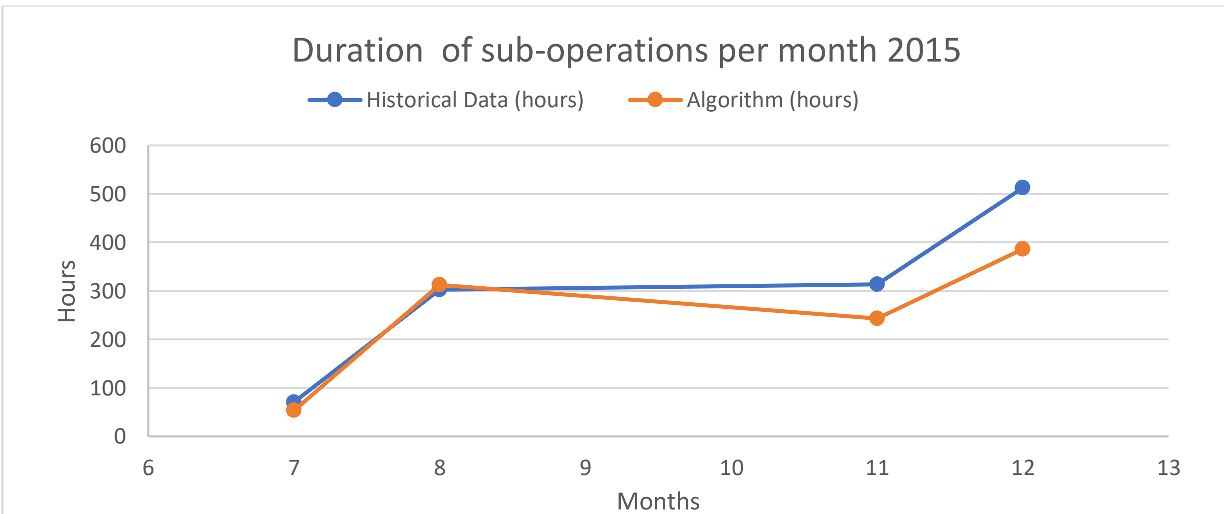


Figure 37. Duration per month, year 2015. Source: Own Elaboration.

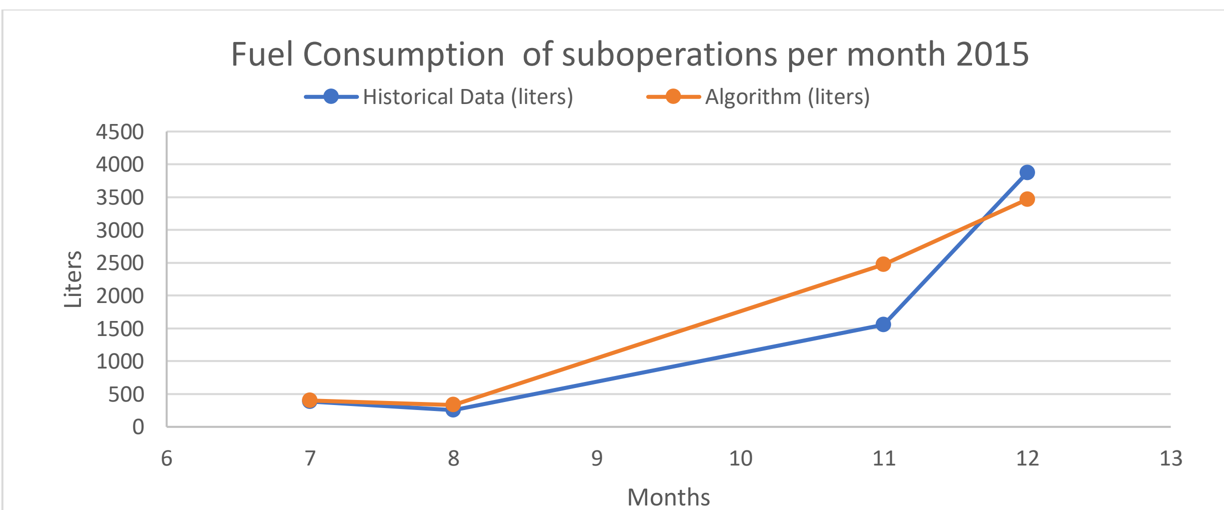


Figure 38. Consumption per month, year 2015. Source: Own Elaboration

The first two graphs displayed are for the year 2015. Here, and with the filters applied to the operations, there is just data available for the months 7,8, 11, and 12. It is clear that the y-axis for them are not in the same scale, since the first one corresponds to a duration in hours, whereas the second one to the consumption, measured in liters. However, what is important to see is that both of the lines, in all two cases, follow the same trend; and, as expected, the line for the algorithm is, overall, lower.

What this means is that the program is being coherent with the tendency , and, in addition, is allowing the operations to be performed in a lower time, subject to the restrictions.

Following 2015, the same table is shown, this time corresponding to the year 2016.

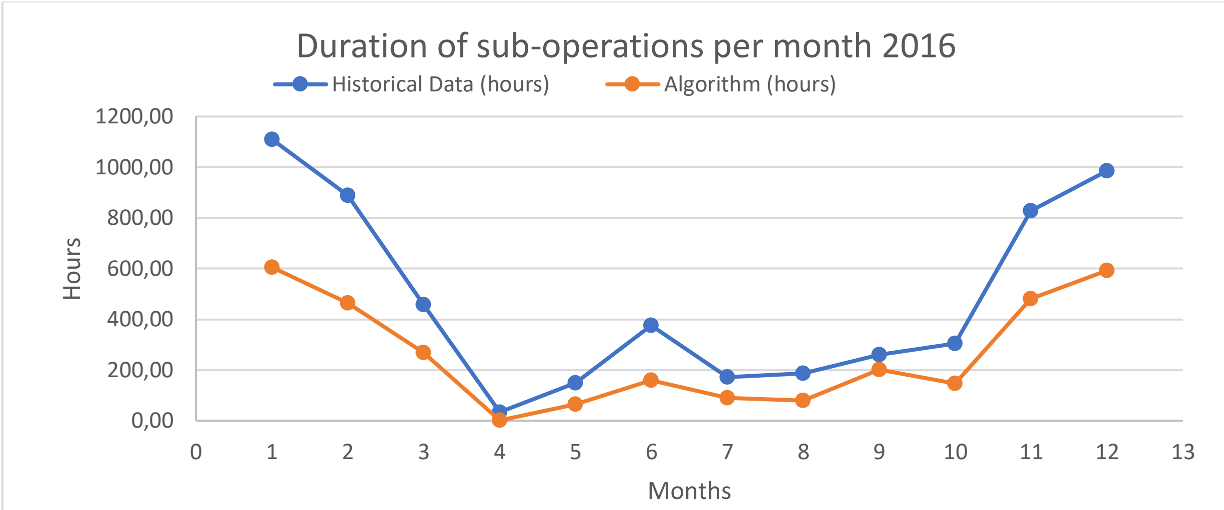


Figure 39. Duration per month, year 2016. Source: Own Elaboration.

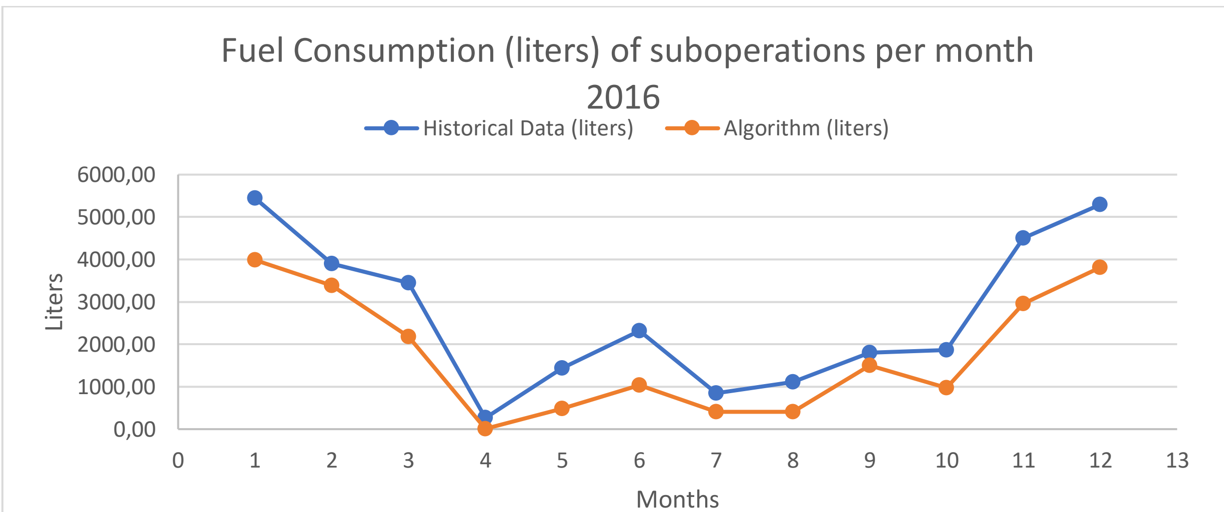


Figure 40. Consumption per month, year 2016. Source: Own Elaboration.

Although the analysis is very similar to the one done in 2015, there are some things that need to be pointed out as well: Here, there is data for all the months, which is consistent with the fact that the FMP platform was better adapted in this year. Also, it is more notable that the consumption and the duration trends are alike, since the former is a function of the latter.

Furthermore, the behavior of the date throughout the months could be appreciated in a better way, compared to the previous year. In month four, there is a low duration, due to a low quantity processed mechanically, and with the filters applied. This case of low values is replicated in some cases of the following years.

Next, tables for the year 2017 are presented.

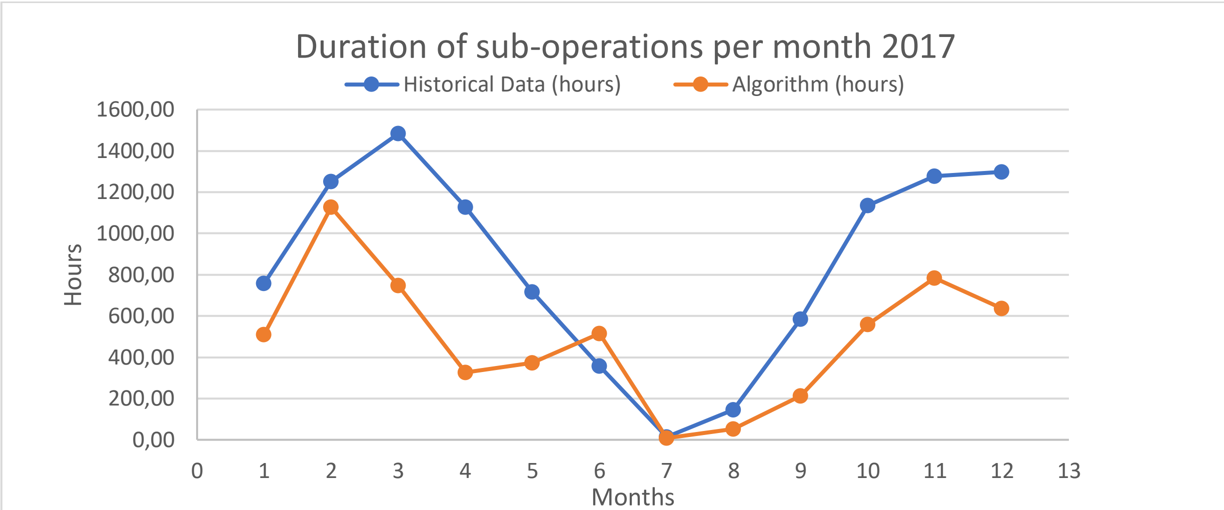


Figure 41. Duration per month, year 2017. Source: Own Elaboration.

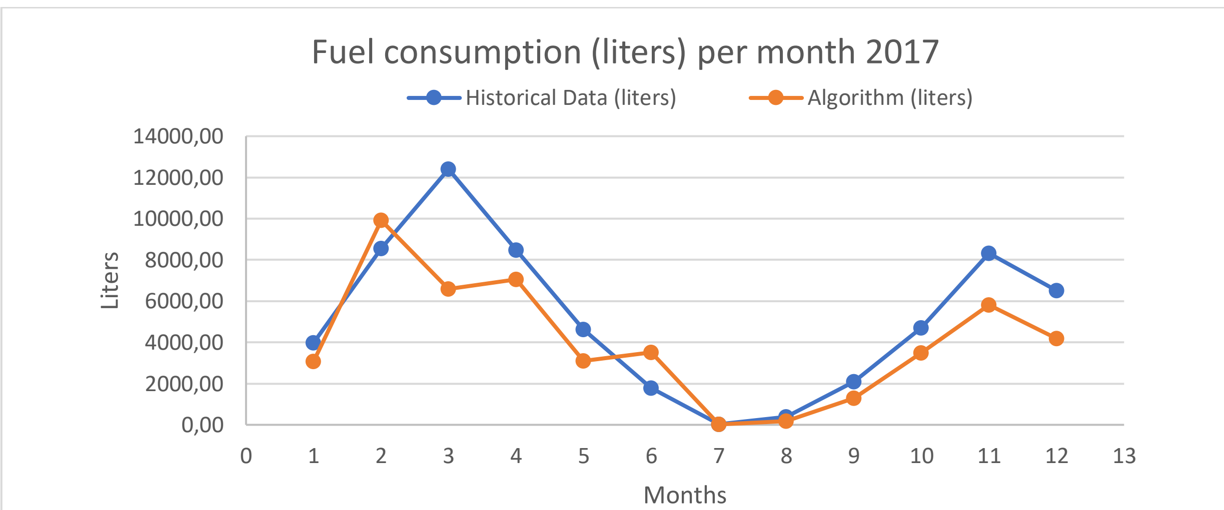


Figure 42. Consumption per month, year 2017. Source: Own Elaboration.

The reasoning is quite similar to the years before. However, it is important to highlight the situation for the months 6 and 7. In the former, the result for the historical data is better to the one found in the algorithm; this is interesting to see, and it is coherent, since the combination of machines is different, and even though the program seeks to minimize duration, because of the restriction of availability, it is not always possible.

In the latter, as an instance, the duration is close to zero. As the previous case, the number of sub-operations is low compared to the other months.

Lastly, the tables corresponding to the year 2018 are also set out.

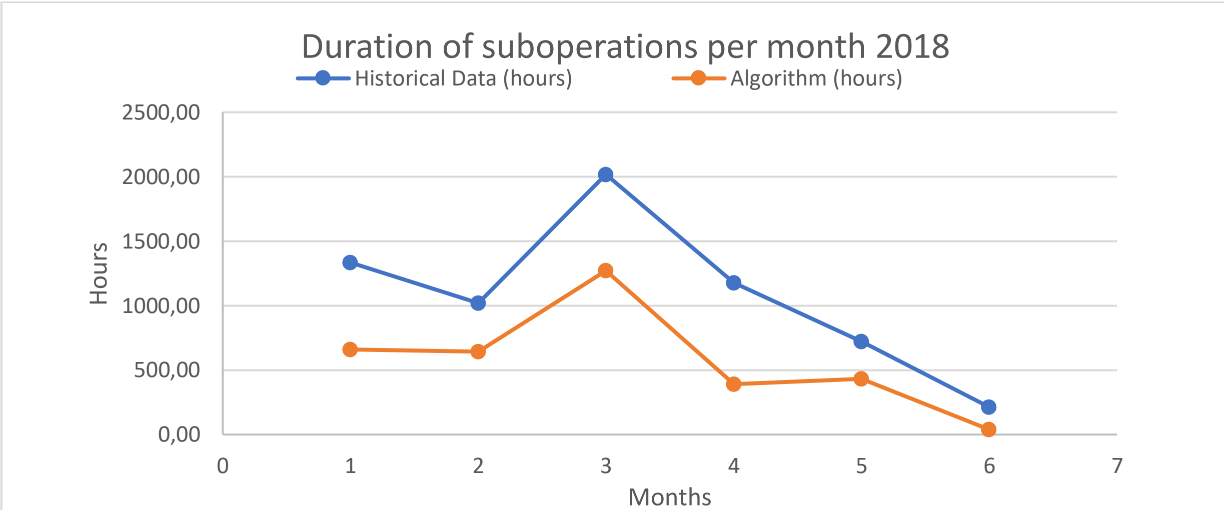


Figure 43. Duration per month, year 2018. Source: Own Elaboration.

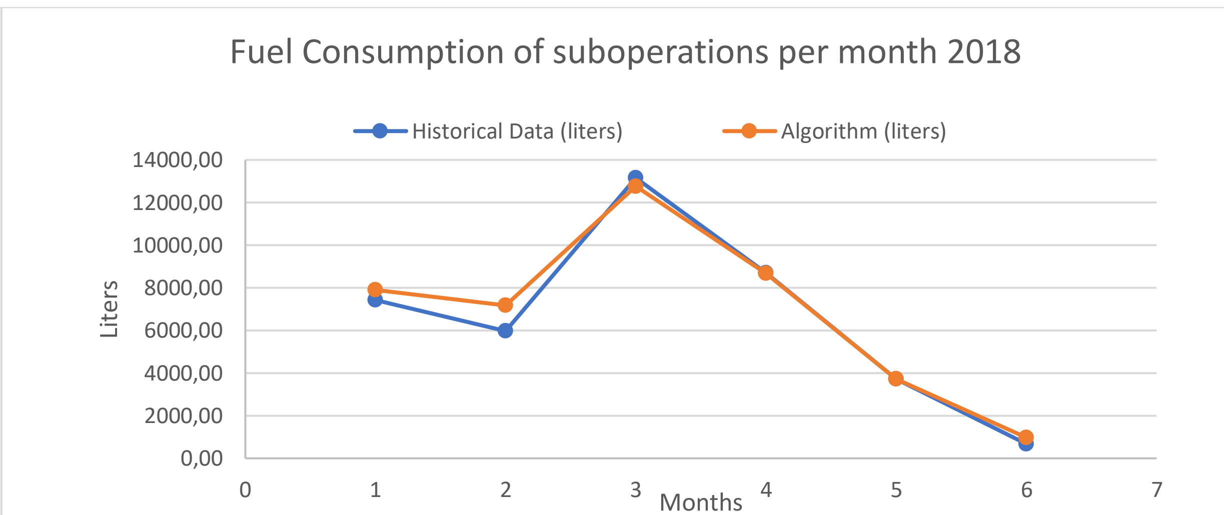


Figure 44.Consumption per month, year 2018. Source: Own Elaboration.

No further comments need to be done for this year, with the exception that there is data just for six months, because the year is not over, and that was the last month for which data was obtained. Also, the duration is low due to the fact of performing operations with low quantity. As a consequence, fuel consumption is low as well.

Furthermore, it makes sense that the behavior of the line throughout the months follows more or less the same shape than the one from the quantity, since it is in a way a function of the latter. For a closer look refer to *Exhibit 5*.

14.9. Daily Calendar

Once the results were discussed with the head of the company, it came to his interest the possibility to design a calendar in which, knowing in advance the sub-operations that needed to be done, and their respective quantity, it was possible to have a template with the tractors and implements to use, and for how long, in a daily basis. This way, they can assure the availability of the machines ahead of time.

The objective of the above is simple: future lines imply a scheduling of the machines, based on some program already organized by the agricultural team, instead of just a historical comparison (which is what is done in this deliverable). Keeping this in mind, the whole maintenance, repairing, and circulation schedule has to be consistent with the results. *Exhibit 6* show the format in which it would be displayed.

14.10. Other Approaches

Although the results were satisfactory in terms of performance, different approaches with new criteria needed to be also evaluated and compared to the Algorithm previously explained. For this purpose, 2 (two) other methods based on 2 (two) principles for optimization were evaluated, named Company Decision, and Random assignation.

The first one takes into account the number of times a set of tractor and implements has been assigned historically to a certain sub-operation. So, following the same logic of the code already explained, instead of choosing an answer by duration, it does so by the set that has been used the most for that sub-operation. In case of not availability at the moment, it takes the second most used, and so on. The other approach, on the other hand, does not consider any principle for the scheduling. Instead, it is based on random number generation, which follows a uniform distribution, with every value having the same probability to be chosen, to make a list with the machines and implements that can perform the sub-operation and does the assignation according to which one is the first element available of the list. Like the others, if it is not available, it starts looking for the other one. In addition, the algorithm is executed a number of times (value that can be set by the user) and takes the best solution out of this iterations. For the case of the current work, that number was set to 20.

Exhibit 7 illustrates a comparison of the performance by years and with different time distances between zones, following a square layout. The scenarios were ran for different instances of time, starting from 10 (ten) minutes, up to 120 minutes (two hours). Each activity takes place within a certain zone. If the machine is required in a different zone for another activity, it will be transported from one place to another, adding this time to the duration of performing the activity.

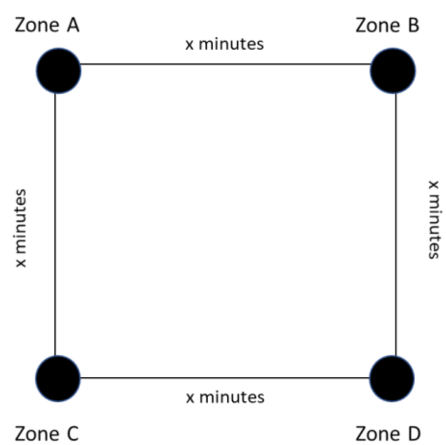


Figure 45. Time Distances between zones. Source: Own Elaboration.

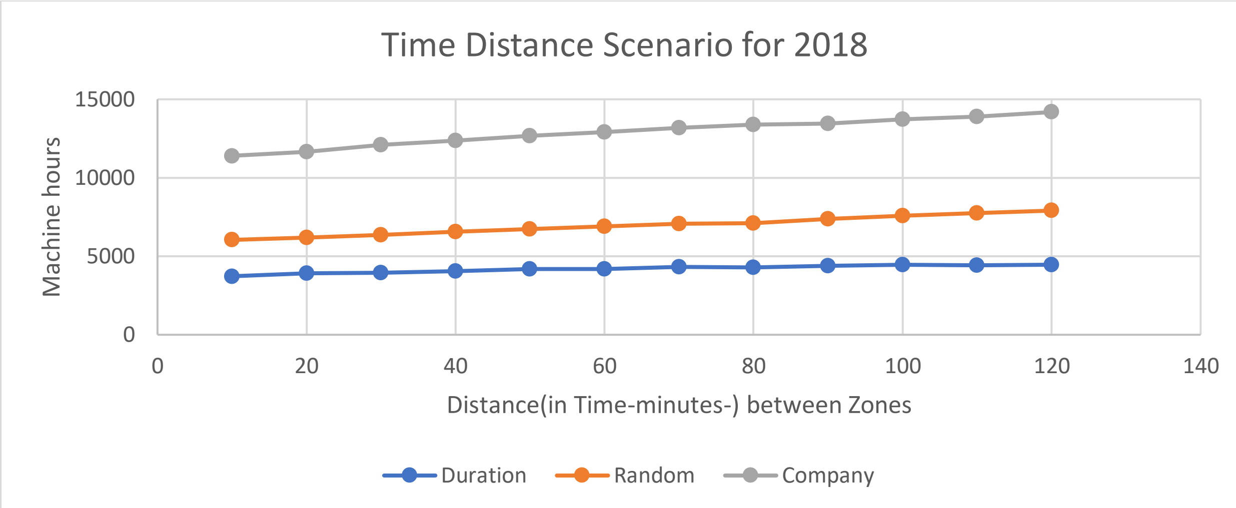


Figure 46. Machine hours per Time scenario for 2018. Source: Own Elaboration

Linearity can be appreciated in the figure. Also, within this graph, and going back to the Exhibit already mentioned during this subsection, it is clear that the Random algorithm is never better than the one used in the company, instead, it performs equally or better. On the other hand, the one using the duration criteria always beats the others.

For the subsequent analysis, a time distance of *60 minutes* is used, so that, a comparison in terms of cumulative quantity, cumulative time, and cumulative fuel consumption could be made. As usual, the reasoning goes for every year of the horizon (2015 to 2018), however, just one of them will be shown within this section. The others, instead, are displayed within the exhibit section.

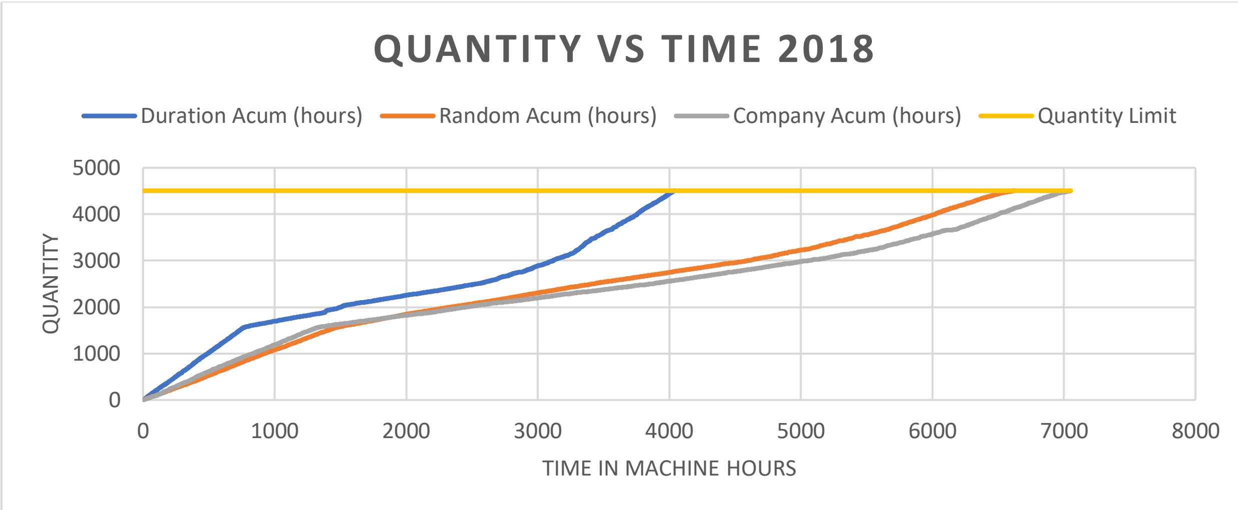


Figure 47. Cumulative Quantity vs Cumulative Time for 2018. Source: Own Elaboration.

This first graph displayed is in terms of cumulative quantity and time, showing for each of the methodologies, how long it takes in machine hours to process the total quantity of the year. In this order of ideas, the horizontal distance between the lines corresponds to the difference in times. As it could be appreciated, the algorithm with the duration criteria for optimization performs better; being the others really close in terms of performance[6].

Once the relation between quantity and time is treated, it becomes interesting to also analyze how time and fuel consumption interact with each other. For this purpose, a similar graph is plotted, this time comparing Cumulative Quantity with Cumulative Time for the same year.

[6] The same behavior is shown for the other years. See Exhibit 8 for more information.

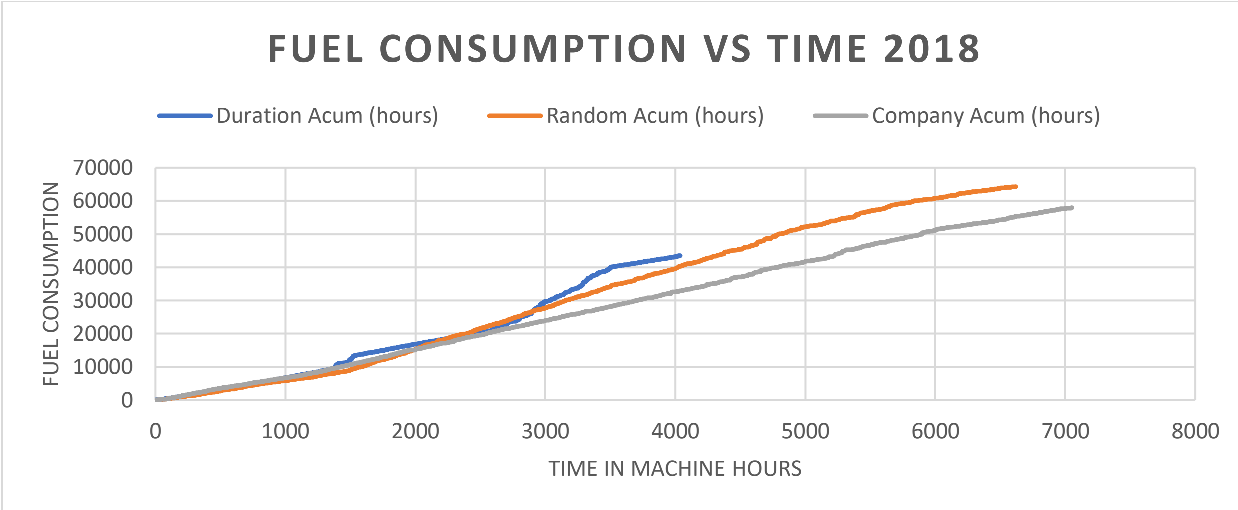


Figure 48. Fuel Consumption Cumulated vs Time Machine in Hours for 2018. Source: Own Elaboration.

What is interesting to interpret from this graph is the fact that fuel consumption does not depend entirely on the duration of the activity. As an instance, the Random criteria performs better in terms of time, but poorly on fuel consumption. However, for a big difference on time performance, it is unavoidable to notice that there will also be a big difference in consumption.

In addition, another fact does not escape to the eyes. Taking as a reference any value between 3000 and 4000 hours in time machine, the consumption of the blue algorithm is higher, even though it stops here. From this scenario, it makes sense to consider both criteria, consumption, and time, to implement a methodology that could minimize both.

In this sense, and for this last analysis, a modification to the original algorithm was made, seeking to choose among the options those machines and implements with a high performance in terms of quantity to be processed over time, and with a low value for fuel consumption over time. The criteria was then the following:

$$Criteria\ by\ machine = \frac{Quantity\ the\ machine\ processes\ for\ a\ certain\ activity\ \ in\ an\ hour}{Fuel\ consumption\ in\ an\ hour\ of\ performing\ a\ certain\ activity}$$

What this proportion does is exactly what was explained above. Since it is the ration between the performance by quantity and by consumption, it punishes those values with low numbers in the numerator, and those with high numbers in the denominator, turning the problem into a maximization problem, instead of a minimization, like it has been done before.

Figures 49 and 50 show the comparison, for the year 2018 of the cumulative time with cumulative quantity and fuel consumption, between the new approach and the other ones.

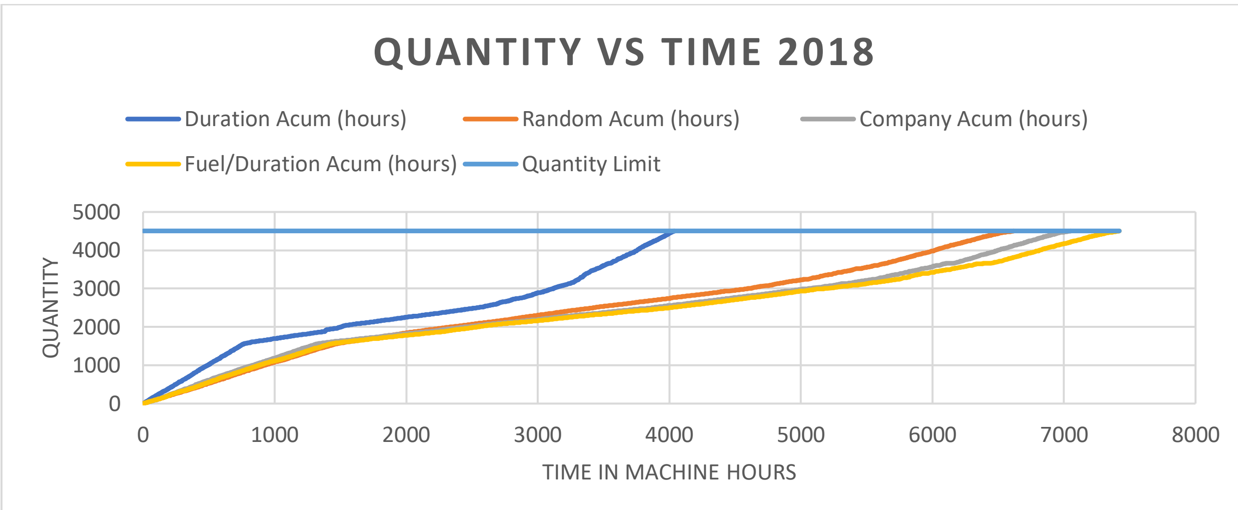


Figure 49. Cumulative Quantity vs Cumulative Time for 2018 with All Criteria. Source: Own Elaboration.

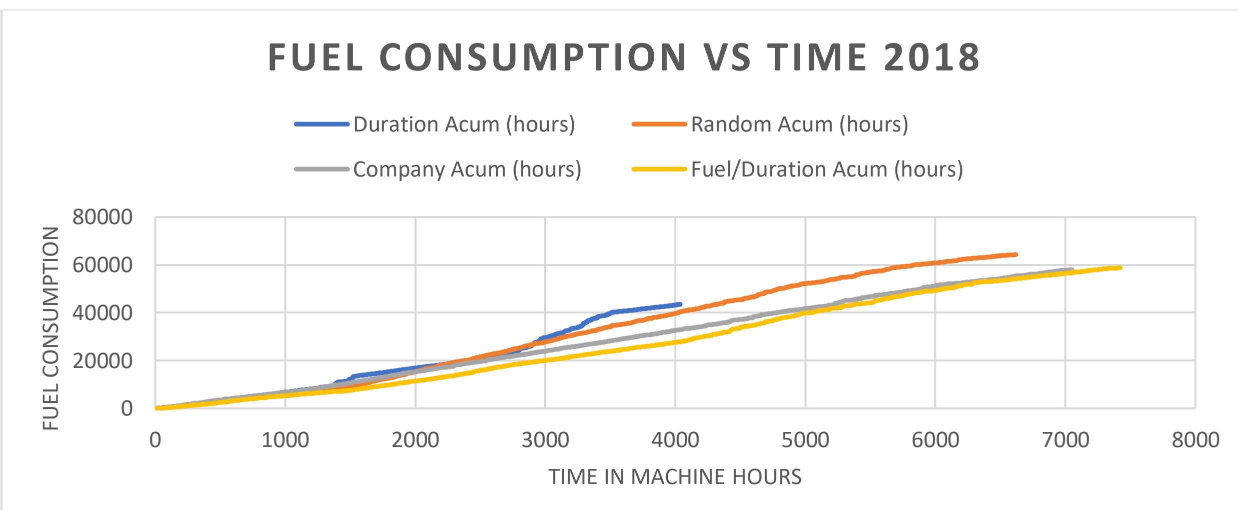


Figure 50. Fuel Consumption Cumulated vs Time Machine in Hours for 2018 with All Criteria. Source: Own Elaboration.

The first figure reflects the same situation than in former cases, the algorithm with the minimization criteria for Duration takes lower time to complete the annual quantity; whereas the others perform more or less the same on this terms.

However, what is interesting to point out is that for each level of time, it consumes less fuel for activity. Nevertheless, overall, the duration criteria continues to perform better due to a big difference in time for completing the activities.

Another important thing that needs to be highlighted as well is that the closest one in terms of full consumption to the last proposal is the company methodology, even outperforming it for some point in the year 2017 (Exhibit 9). This situation gives the idea that currently the company does the assignation taking into account, to a large degree, the fuel consumption.

15. Conclusions

Continuous change creates the need of quick response and adaptation. In doing so, strategical decisions have to be taken regarding the internal and external environment, and for that, data transformation has to be carried out, in a way such that it could be molded into useful information.

The fact that this work was developed as a first approach to analyze the data loaded into the DB (specifically for what was stored in the FMP platform, corresponding to Farm and Field activities), has led to multiple conclusions regarding the current treatment of information and possible improvements on how the registration is made. In addition, throughout the process, several things were proposed in order to make future analysis friendlier and faster for the stakeholders in both sides: the makers, and the implementers.

More concretely, and going in line with the structure of the current work in terms of objectives and scope, the following conclusions could be listed:

- Data is inserted manually by secretaries at the end of the day, it is highly recommended to be more careful in doing so, since there are things that do not correspond to each other and cause a lot of noise. It is clear though, that this current way of aggregating content is subjected to human error; however, a stricter control has to be carried on.
- Out of 1,616 combinations of operations with resources found, just 315 were included beforehand within the DB, i.e., there were efficiencies already taken into account. This is a signal for the company for widen its scope in terms of scenario analysis, and it does reinforce the idea that what was done within the work of analyzing historical efficiencies, needed to be done.
- A data model has been created to be manipulated by the users, so that, computations regarding a factor of interest not done within the present work, could be done latter.

- The assignation method used so far can indeed be improved in terms of duration and fuel consumption, meaning that a better usage of resources for the activities within the company is possible.
- As expected, the lower the duration, the lower the Fuel consumption for activity. However, there are some variables that are not taken into account and that are difficult to count, such as steepness of the field, kind and consistency of the soil within the same company, and skill of the driver, among others. Ergo, Fuel consumption does not depend entirely on duration of the activity.
- The results obtained within this work are a good approach to what happens in reality; nevertheless, it does not reflect it in a hundred percent. The assignation of machines is just a suggestion for the company, based on data analysis, to carry on its activities. However, it is its decision to follow what is shown here, or to use a different criteria for the assignation.

16. Future lines of action

Responding to the needs in the market requires exploration of new horizons. In doing so, flexible, and scalable systems are required. However, because of the scope of this thesis, there are many things that still need to be done to reach to that point. In particular, improvement of the model created within Power BI, and continuous learning of the platform is now in hands of the stakeholders; and, in addition, it has been demonstrated that the current assignation method can be overcame, so that, a scheduling of the machines, based on farm and field activities conceived beforehand by the agricultural team, could be elaborated and it will be of extremely interest for the company.

17. References

[1] Christensson, P. (2007). Visual Basic Definition. [online] Available at: https://techterms.com [Accessed 12 Aug. 2018].

[2] Docs.microsoft.com. (2018). DAX basics in Power BI Desktop - Power BI. [online] Available at: https://docs.microsoft.com/en-us/power-bi/desktop-quickstart-learn-dax-basics [Accessed 12 Aug. 2018].

[3] Docs.microsoft.com. (2018). What is Power BI? - Power BI. [online] Available at: https://docs.microsoft.com/en-us/power-bi/power-bi-overview [Accessed 12 Aug. 2018].

[4] Everitt, B. and Skrondal, A. (2018). The Cambridge Dictionary of Statistics. [online] Stewartschultz.com. Available at: http://www.stewartschultz.com/statistics/books/Cambridge%20Dictionary%20Statistics%204th.pdf [Accessed 16 Jun. 2018].

[5] Fresco, A. (2017). Gestione delle Informazioni e Simulazione di un Processo di Essiccazione delle Nocciole (Master Thesis). Politecnico di Torino, Torino.

[6] Frutícola Agrichile. (2018). Sobre Agrichile. [online] Available at: https://www.agrichile.cl/agrichile/ [Accessed 21 May. 2018].

[7] ISO 9075-1:2016, I. (2018). ISO/IEC 9075-1:2016 - Information technology -- Database languages -- SQL -- Part 1: Framework (SQL/Framework). [online] Iso.org. Available at: https://www.iso.org/standard/63555.html [Accessed 1 Aug. 2018].

[8] Luhn, H.P. (1958). A Business Intelligence System. IBM Journal, pp.314-319.

[9] Orloff, J. and Bloom, J. (2018). Bootstrap Confidence Intervals. [online] Math.mit.edu. Available at: https://math.mit.edu/~dav/05.dir/class24-prep-a.pdf [Accessed 8 Aug. 2018].

[10] RStudio. (2018). RStudio - RStudio. [online] Available at: https://www.rstudio.com/products/rstudio/ [Accessed 12 Aug. 2018].

[11] SearchSQLServer. (2018). What is Microsoft SQL Server? - Definition from WhatIs.com. [online] Available at: https://searchsqlserver.techtarget.com/definition/SQL-Server [Accessed 12 Aug. 2018].

[12] Seo, S. (2006). A Review and Comparison of Methods for Detecting Outliers in Univariate Data Sets (Master Thesis). University of Pittsburgh, Pittsburgh. [online] Available at http://d-scholarship.pitt.edu/7948/1/Seo.pdf [Accessed 22 Jul. 2018].

[13] Stat.cmu.edu. (2018). The Bootstrap. [online] Available at: http://www.stat.cmu.edu/~cshalizi/402/lectures/08-bootstrap/lecture-08.pdf [Accessed 8 Aug. 2018].

[14] Techopedia.com. (2018). What is a Greedy Algorithm? - Definition from Techopedia. [online] Available at: https://www.techopedia.com/definition/16931/greedy-algorithm [Accessed 12 Aug. 2018].

[15] Tomaiuolo, D. (2018). (Master Thesis). Politecnico di Torino, Torino.

[16] Tukey, JW (1977). Exploratory Data Analysis. Addison-Wesley.

[17] van der Lans, R. (2012). Business Intelligence and Data Warehousing. Data Virtualization for Business Intelligence Systems, pp.27-57.

[18] Yagüe, A. (2018). Data Analyzing by the Development of a Self-Service Business Intelligence Solution Applied to an Agro-Food Organization (Master Thesis). Politecnico di Torino, Torino.

[19] 3a S.r.l Drying. (2018). HDS Technical Manual.

[20] 3a S.r.l Farming. (2018). FMP Technical Manual.

[21] 3a S.r.l Green Planet. (2018). Green Planet Home. [online] Available at: http://www.green-planet.it/index.php/en/ [Accessed 18 May. 2018].

18. Exhibits

Exhibit 1: Excel Deliverable Results for Simple Efficiency

Subop_ID	Subop Name	Company	frop_type	FamilyMachine	FamilyImplement1	FamilyImplement2	Unit	MAX	MIN	MEDIAN	Average Simple efficiency	FMP Efficiency	%Error	If in DB
691	Extracting fror	HCO Los Nic	MANUAL				ha	0,01	0,01	0,01	**0,01**	**0,01**	**0,00**	normal
544	Hand collectio	HCO Camari	MANUAL				ha	0,05	0,02	0,05	**0,04**	**0,04**	**0,00**	normal
550	Machine collec	HCO Caracas	MECHANIC	27			ha	4,5	0,26	1,625	**1,86**	**1,80**	**3,23**	normal
591	Harrowing	HCO S.Greg	MECHANIC	1	48		ha	18	5,4	10,8	**10,56**	**10,00**	**5,28**	normal
551	Machine collec	HCO Caracas	MECHANIC	30			ha	6,15	0,23	2,37	**2,55**	**2,40**	**5,76**	normal
541	Application of	HCO Caracas	MECHANIC				ha	17,78	4,57	14,4	**12,77**	**12,00**	**6,02**	normal
607	Pesticide appli	HCO Los Nic	MECHANIC	25	44		ha	18	1	8,61	**8,55**	**8,00**	**6,40**	normal
607	Pesticide appli	HCO Camari	MECHANIC	25	45		ha	13,5	2,25	5,25	**6,42**	**6,00**	**6,51**	normal
547	Preparation wi	HCO Caracas	MANUAL		21		ha	2	0,11	1	**0,93**	**0,87**	**6,74**	normal
550	Machine collec	HCO Caracas	MECHANIC	30			ha	6	0,4	2,25	**2,65**	**2,44**	**7,78**	normal
543	Foliar fertilizat	HCO Camari	MECHANIC	25	44		ha	20,16	3,47	10,4	**11,00**	**10,00**	**9,09**	normal
607	Pesticide appli	HCO Camari	MECHANIC	25	44		ha	21,33	2	12	**11,34**	**10,00**	**11,84**	normal
562	Inspecting the	HCO Los Nic	MANUAL				ha	5	0,03	2	**2,28**	**2,00**	**12,24**	normal
562	Inspecting the	HCO Caracas	MANUAL				ha	13	0,07	1	**2,67**	**3,00**	**12,39**	normal
548	Preparation wi	HCO Los Nic	MECHANIC	25	14		ha	14,18	0,84	6,3	**6,85**	**6,00**	**12,44**	normal
607	Pesticide appli	HCO Caracas	MECHANIC	25	44		ha	21,82	0,01	11,33	**11,46**	**10,00**	**12,74**	normal
550	Machine collec	HCO S.Sebas	MECHANIC	28			ha	10	1	4	**4,30**	**3,75**	**12,86**	normal
687	Weeds control	HCO Camari	MANUAL				ha	1	0,01	1	**0,58**	**0,50**	**13,95**	normal
687	Weeds control	HCO Los Nic	MANUAL				ha	0,06	0,01	0,01	**0,02**	**0,02**	**14,86**	normal
547	Preparation wi	HCO Los Nic	MANUAL		21		ha	2	0,09	1	**1,03**	**0,87**	**15,36**	normal
550	Machine collec	HCO Camari	MECHANIC	28			ha	10,5	0,07	4	**4,45**	**3,75**	**15,75**	normal
547	Preparation wi	HCO Camari	MANUAL		21		ha	2	0,25	1	**1,04**	**0,87**	**16,57**	normal
607	Pesticide appli	HCO Los Nic	MECHANIC	25	44		ha	18	1	8,61	**8,55**	**10,00**	**17,00**	normal
581	Pruning - Main	HCO S.Sebas	MECHANIC	25	68		ha	10,29	2,67	8,305	**7,24**	**6,00**	**17,10**	normal
650	Mulcher	HCO S.Sebas	MECHANIC	25	36		ha	27	3	11,52	**12,08**	**10,00**	**17,25**	normal
607	Pesticide appli	HCO S.Sebas	MECHANIC	25	44		ha	22,03	2	12	**12,14**	**10,00**	**17,63**	normal
543	Foliar fertilizat	HCO S.Sebas	MECHANIC	25	44		ha	21,68	3,63	12,47	**12,14**	**10,00**	**17,66**	normal
547	Preparation wi	HCO S.Sebas	MANUAL		21		ha	2	0,09	1	**1,08**	**0,87**	**19,07**	normal
607	Pesticide appli	HCO Caracas	MECHANIC	1	41		ha	10	10	10	**10,00**	**8,00**	**20,00**	normal
551	Machine collec	HCO Camari	MECHANIC	28			ha	10	1	4,375	**4,76**	**3,75**	**21,14**	normal
562	Inspecting the	HCO Camari	MANUAL				ha	3,18	0,04	1	**1,28**	**1,00**	**21,64**	normal
562	Inspecting the	HCO Camari	MANUAL				ha	3,18	0,04	1	**1,28**	**1,00**	**21,64**	normal
548	Preparation wi	HCO Camari	MECHANIC	25	14		ha	31,63	0,23	12,75	**13,23**	**10,33**	**21,94**	normal
634	Miscellaneous	HCO S.Greg	MANUAL				ha	0,33	0,03	0,09	**0,12**	**0,09**	**22,58**	normal
551	Machine collec	HCO Los Nic	MECHANIC	26			ha	7,38	0,01	3	**3,26**	**2,50**	**23,32**	normal
627	Transport of pr	HCO Los Nic	MECHANIC	25	38		ha	8,2	1,33	4,1	**4,05**	**5,00**	**23,36**	normal
540	Application of	HCO Caracas	MANUAL				ha	0,82	0,35	0,56	**0,57**	**0,70**	**23,89**	normal
650	Mulcher	HCO Caracas	MECHANIC	25	36		ha	35,29	0,37	13,33	**13,17**	**10,00**	**24,08**	normal
545	Hand collectio	HCO Caracas	MANUAL				ha	0,23	0,04	0,08	**0,11**	**0,08**	**24,53**	normal
624	Transport of pr	HCO S.Sebas	MECHANIC	25	38		ha	6,4	2,58	5,33	**4,81**	**6,00**	**24,68**	normal
576	Pruning - Form	HCO Camari	MANUAL				pl	144	144	144	**144,00**	**180,00**	**25,00**	normal
544	Hand collectio	HCO Camari	MANUAL				ha	0,05	0,02	0,05	**0,04**	**0,03**	**25,00**	normal
562	Inspecting the	HCO Caracas	MANUAL				ha	13	0,07	1	**2,67**	**2,00**	**25,07**	normal
687	Weeds control	HCO Los Nic	MANUAL		86		ha	1,23	0,04	0,7	**0,54**	**0,40**	**25,35**	normal
606	Pruning plants	HCO Camari	MANUAL				ha	0,79	0,15	0,22	**0,40**	**0,30**	**25,74**	normal
607	Pesticide appli	HCO S.Greg	MECHANIC	25	44		ha	27	0,73	13,2	**13,54**	**10,00**	**26,12**	normal
550	Machine collec	HCO Camari	MECHANIC	30			ha	8	1	3	**3,32**	**2,44**	**26,56**	normal

Average Simple Efficiency is the one computed historically, FMP Efficiency, on the other hand is the one within the Platform. In addition, the column "If in DB" evaluates if the set of resources within the operations were taken into account for in the DB, and historically. Normal means it was considered in both, DB, just in the Database, and historical, historically.

Exhibit 2: Excel Deliverable Results for Broken_Down Efficiency

Index	frop_id	frop_name	frop_type	cmp_name	frop_un	froput_ut	FamilyMachine	FamilyImplement1	FamilyImplement2	Max	Min	Median	Average Broken down efficiency
1	534	Levelling bef	MECHANIC	HCO S.Gregori	ha	1601	31			14,46	7,89	11,18	**11,18**
2	534	Levelling bef	MECHANIC	HCO S.Sebasti	ha	996	25	19		1,20	1,20	1,20	**1,20**
3	534	Levelling bef	MECHANIC	HCO S.Gregori	ha	1601	25	19		9,00	9,00	9,00	**9,00**
4	534	Levelling bef	MECHANIC	HCO Santa Ani	ha	13476				0,90	0,60	0,72	**0,74**
5	534	Levelling bef	MECHANIC	HCO Camarico	ha	6475				0,45	0,41	0,41	**0,42**
6	534	Levelling bef	MECHANIC	HCO Camarico	ha	185	95			0,58	0,32	0,45	**0,45**
7	534	Levelling bef	MECHANIC	HCO Caracas	ha	827	25	4		11,25	11,25	11,25	**11,25**
8	534	Levelling bef	MECHANIC	HCO Caracas	ha	817	25	4		2,50	2,50	2,50	**2,50**
9	534	Levelling bef	MECHANIC	HCO Caracas	ha	817	1	4		1,13	1,13	1,13	**1,13**
10	534	Levelling bef	MECHANIC	HCO S.Gregori	ha	1601	1	19		6,00	6,00	6,00	**6,00**
11	534	Levelling bef	MECHANIC	HCO Caracas	ha	824	25			7,71	7,71	7,71	**7,71**
12	534	Levelling bef	MECHANIC	HCO Caracas	ha	817	25			5,56	1,55	3,55	**3,55**
13	540	Application c	MANUAL	HCO Los Niche	ha	1591				1,23	1,23	1,23	**1,23**
14	540	Application c	MANUAL	HCO Camarico	ha	202				1,68	1,68	1,68	**1,68**
15	540	Application c	MANUAL	HCO Camarico	ha	173		59		2,60	2,60	2,60	**2,60**
16	540	Application c	MANUAL	HCO Caracas	ha	1697				0,82	0,82	0,82	**0,82**
17	540	Application c	MANUAL	HCO Camarico	ha	160		132		2,00	2,00	2,00	**2,00**
18	541	Application c	MECHANIC	HCO Caracas	ha	815	31			40,00	40,00	40,00	**40,00**
19	541	Application c	MECHANIC	HCO Caracas	ha	826	31			64,50	50,00	57,25	**57,25**
20	541	Application c	MECHANIC	HCO Caracas	ha	809	31			28,93	28,93	28,93	**28,93**
21	541	Application c	MECHANIC	HCO Caracas	ha	814	31			44,00	34,10	39,05	**39,05**
22	541	Application c	MECHANIC	HCO Caracas	ha	810	31			32,00	17,39	24,70	**24,70**
23	541	Application c	MECHANIC	HCO Caracas	ha	811	31			22,86	16,00	19,43	**19,43**
24	541	Application c	MECHANIC	HCO Caracas	ha	826	1			55,38	40,91	45,92	**47,40**
25	541	Application c	MECHANIC	HCO Caracas	ha	818	1			27,69	27,69	27,69	**27,69**
26	541	Application c	MECHANIC	HCO Caracas	ha	827	1			113,68	113,68	113,68	**113,68**
27	541	Application c	MECHANIC	HCO Caracas	ha	818	31			27,00	27,00	27,00	**27,00**
28	541	Application c	MECHANIC	HCO Caracas	ha	824	1			60,00	28,07	51,68	**47,86**
29	541	Application c	MECHANIC	HCO Caracas	ha	809	1			96,43	23,68	37,50	**54,43**
30	541	Application c	MECHANIC	HCO Camarico	ha	185	1	44		10,40	10,40	10,40	**10,40**
31	541	Application c	MECHANIC	HCO Caracas	ha	810	25	7		1,80	1,80	1,80	**1,80**
32	541	Application c	MECHANIC	HCO Camarico	ha	187	25	44		5,20	5,20	5,20	**5,20**
33	541	Application c	MECHANIC	HCO Camarico	ha	187	1	44		11,89	8,32	10,10	**10,10**
34	541	Application c	MECHANIC	HCO Camarico	ha	178	1	44		7,06	7,06	7,06	**7,06**
35	541	Application c	MECHANIC	HCO Camarico	ha	186	1	44		10,40	10,40	10,40	**10,40**
36	541	Application c	MECHANIC	HCO Caracas	ha	828	31			35,00	21,33	28,17	**28,17**
37	541	Application c	MECHANIC	HCO Caracas	ha	829	31			32,73	32,73	32,73	**32,73**
38	541	Application c	MECHANIC	HCO Caracas	ha	827	31			145,45	30,86	30,86	**69,06**
39	541	Application c	MECHANIC	HCO S.Sebasti	ha	986	5			2,79	2,79	2,79	**2,79**

Differently from Average Simple Efficiency, it cannot be compared. Instead, this information is useful for a more detailed analysis of the efficiencies by Territorial Unit within a field.

Exhibit 3: Confidence Intervals and Statistical Validation Simple Efficiency

IndexR	Index	frop_id	frop_name	cmp_name	frop_type	FamilyMachine	FamilyImplement1	FamilyImplement2	SimpleEfficiency	Simple efficiency Final	Max	Min	Median	LCL	UCL	%error	Validation
1	4	534	Levelling bef	HCO Camarico	MECHANIC				0,45	0,4191	0,45	0,41	0,41	0,4121	0,4264	7,38%	Not Validated
1	4	534	Levelling bef	HCO Camarico	MECHANIC				0,45	0,4191	0,45	0,41	0,41	0,4121	0,4264	7,38%	Not Validated
1	4	534	Levelling bef	HCO Camarico	MECHANIC				0,41	0,4191	0,45	0,41	0,41	0,4121	0,4264	2,17%	Not Validated
1	4	534	Levelling bef	HCO Camarico	MECHANIC				0,41	0,4191	0,45	0,41	0,41	0,4121	0,4264	2,17%	Not Validated
1	4	534	Levelling bef	HCO Camarico	MECHANIC				0,41	0,4191	0,45	0,41	0,41	0,4121	0,4264	2,17%	Not Validated
1	4	534	Levelling bef	HCO Camarico	MECHANIC				0,45	0,4191	0,45	0,41	0,41	0,4121	0,4264	7,38%	Not Validated
1	4	534	Levelling bef	HCO Camarico	MECHANIC				0,41	0,4191	0,45	0,41	0,41	0,4121	0,4264	2,17%	Not Validated
1	4	534	Levelling bef	HCO Camarico	MECHANIC				0,41	0,4191	0,45	0,41	0,41	0,4121	0,4264	2,17%	Not Validated
1	4	534	Levelling bef	HCO Camarico	MECHANIC				0,41	0,4191	0,45	0,41	0,41	0,4121	0,4264	2,17%	Not Validated
1	4	534	Levelling bef	HCO Camarico	MECHANIC				0,41	0,4191	0,45	0,41	0,41	0,4121	0,4264	2,17%	Not Validated
1	4	534	Levelling bef	HCO Camarico	MECHANIC				0,41	0,4191	0,45	0,41	0,41	0,4121	0,4264	2,17%	Not Validated
1	4	534	Levelling bef	HCO Camarico	MECHANIC				0,41	0,4191	0,45	0,41	0,41	0,4121	0,4264	2,17%	Not Validated
1	4	534	Levelling bef	HCO Camarico	MECHANIC				0,41	0,4191	0,45	0,41	0,41	0,4121	0,4264	2,17%	Not Validated
1	4	534	Levelling bef	HCO Camarico	MECHANIC				0,41	0,4191	0,45	0,41	0,41	0,4121	0,4264	2,17%	Not Validated
1	4	534	Levelling bef	HCO Camarico	MECHANIC				0,41	0,4191	0,45	0,41	0,41	0,4121	0,4264	2,17%	Not Validated
1	4	534	Levelling bef	HCO Camarico	MECHANIC				0,45	0,4191	0,45	0,41	0,41	0,4121	0,4264	7,38%	Not Validated
1	4	534	Levelling bef	HCO Camarico	MECHANIC				0,41	0,4191	0,45	0,41	0,41	0,4121	0,4264	2,17%	Not Validated
1	4	534	Levelling bef	HCO Camarico	MECHANIC				0,41	0,4191	0,45	0,41	0,41	0,4121	0,4264	2,17%	Not Validated
1	4	534	Levelling bef	HCO Camarico	MECHANIC				0,41	0,4191	0,45	0,41	0,41	0,4121	0,4264	2,17%	Not Validated
1	4	534	Levelling bef	HCO Camarico	MECHANIC				0,41	0,4191	0,45	0,41	0,41	0,4121	0,4264	2,17%	Not Validated
1	4	534	Levelling bef	HCO Camarico	MECHANIC				0,45	0,4191	0,45	0,41	0,41	0,4121	0,4264	7,38%	Not Validated
1	4	534	Levelling bef	HCO Camarico	MECHANIC				0,41	0,4191	0,45	0,41	0,41	0,4121	0,4264	2,17%	Not Validated
2	5	534	Levelling bef	HCO Santa An	MECHANIC				0,81	0,7422	0,9	0,6	0,72	0,7166	0,7694	9,13%	Not Validated
2	5	534	Levelling bef	HCO Santa An	MECHANIC				0,75	0,7422	0,9	0,6	0,72	0,7166	0,7694	1,05%	Validated
2	5	534	Levelling bef	HCO Santa An	MECHANIC				0,81	0,7422	0,9	0,6	0,72	0,7166	0,7694	9,13%	Not Validated
2	5	534	Levelling bef	HCO Santa An	MECHANIC				0,81	0,7422	0,9	0,6	0,72	0,7166	0,7694	9,13%	Not Validated
2	5	534	Levelling bef	HCO Santa An	MECHANIC				0,75	0,7422	0,9	0,6	0,72	0,7166	0,7694	1,05%	Validated
2	5	534	Levelling bef	HCO Santa An	MECHANIC				0,81	0,7422	0,9	0,6	0,72	0,7166	0,7694	9,13%	Not Validated
2	5	534	Levelling bef	HCO Santa An	MECHANIC				0,6	0,7422	0,9	0,6	0,72	0,7166	0,7694	19,16%	Not Validated
2	5	534	Levelling bef	HCO Santa An	MECHANIC				0,81	0,7422	0,9	0,6	0,72	0,7166	0,7694	9,13%	Not Validated
2	5	534	Levelling bef	HCO Santa An	MECHANIC				0,72	0,7422	0,9	0,6	0,72	0,7166	0,7694	2,99%	Validated
2	5	534	Levelling bef	HCO Santa An	MECHANIC				0,9	0,7422	0,9	0,6	0,72	0,7166	0,7694	21,26%	Not Validated
2	5	534	Levelling bef	HCO Santa An	MECHANIC				0,63	0,7422	0,9	0,6	0,72	0,7166	0,7694	15,12%	Not Validated
2	5	534	Levelling bef	HCO Santa An	MECHANIC				0,63	0,7422	0,9	0,6	0,72	0,7166	0,7694	15,12%	Not Validated
2	5	534	Levelling bef	HCO Santa An	MECHANIC				0,63	0,7422	0,9	0,6	0,72	0,7166	0,7694	15,12%	Not Validated
2	5	534	Levelling bef	HCO Santa An	MECHANIC				0,75	0,7422	0,9	0,6	0,72	0,7166	0,7694	1,05%	Validated
2	5	534	Levelling bef	HCO Santa An	MECHANIC				0,72	0,7422	0,9	0,6	0,72	0,7166	0,7694	2,99%	Validated
2	5	534	Levelling bef	HCO Santa An	MECHANIC				0,84	0,7422	0,9	0,6	0,72	0,7166	0,7694	13,17%	Not Validated
2	5	534	Levelling bef	HCO Santa An	MECHANIC				0,72	0,7422	0,9	0,6	0,72	0,7166	0,7694	2,99%	Validated
2	5	534	Levelling bef	HCO Santa An	MECHANIC				0,8	0,7422	0,9	0,6	0,72	0,7166	0,7694	7,78%	Not Validated
2	5	534	Levelling bef	HCO Santa An	MECHANIC				0,6	0,7422	0,9	0,6	0,72	0,7166	0,7694	19,16%	Not Validated
2	5	534	Levelling bef	HCO Santa An	MECHANIC				0,75	0,7422	0,9	0,6	0,72	0,7166	0,7694	1,05%	Validated
2	5	534	Levelling bef	HCO Santa An	MECHANIC				0,72	0,7422	0,9	0,6	0,72	0,7166	0,7694	2,99%	Validated
2	5	534	Levelling bef	HCO Santa An	MECHANIC				0,72	0,7422	0,9	0,6	0,72	0,7166	0,7694	2,99%	Validated
2	5	534	Levelling bef	HCO Santa An	MECHANIC				0,72	0,7422	0,9	0,6	0,72	0,7166	0,7694	2,99%	Validated
2	5	534	Levelling bef	HCO Santa An	MECHANIC				0,72	0,7422	0,9	0,6	0,72	0,7166	0,7694	2,99%	Validated
2	5	534	Levelling bef	HCO Santa An	MECHANIC				0,72	0,7422	0,9	0,6	0,72	0,7166	0,7694	2,99%	Validated
2	5	534	Levelling bef	HCO Santa An	MECHANIC				0,72	0,7422	0,9	0,6	0,72	0,7166	0,7694	2,99%	Validated
2	5	534	Levelling bef	HCO Santa An	MECHANIC				0,72	0,7422	0,9	0,6	0,72	0,7166	0,7694	2,99%	Validated
2	5	534	Levelling bef	HCO Santa An	MECHANIC				0,72	0,7422	0,9	0,6	0,72	0,7166	0,7694	2,99%	Validated
2	5	534	Levelling bef	HCO Santa An	MECHANIC				0,72	0,7422	0,9	0,6	0,72	0,7166	0,7694	2,99%	Validated
2	5	534	Levelling bef	HCO Santa An	MECHANIC				0,72	0,7422	0,9	0,6	0,72	0,7166	0,7694	2,99%	Validated
2	5	534	Levelling bef	HCO Santa An	MECHANIC				0,72	0,7422	0,9	0,6	0,72	0,7166	0,7694	2,99%	Validated
2	5	534	Levelling bef	HCO Santa An	MECHANIC				0,9	0,7422	0,9	0,6	0,72	0,7166	0,7694	21,26%	Not Validated
2	5	534	Levelling bef	HCO Santa An	MECHANIC				0,9	0,7422	0,9	0,6	0,72	0,7166	0,7694	21,26%	Not Validated

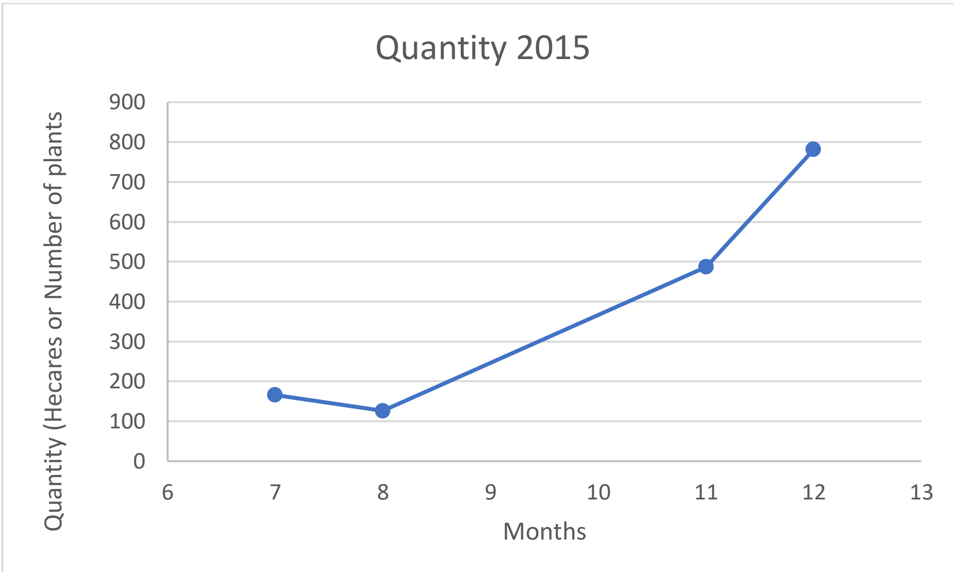
As explained throughout the work, for each set of resources, this tables contains the efficiency stored in the platform, the historical one, Max, Min, Median, Lower and Upper control Limits values, and the validation: if the value is within the confidence interval.

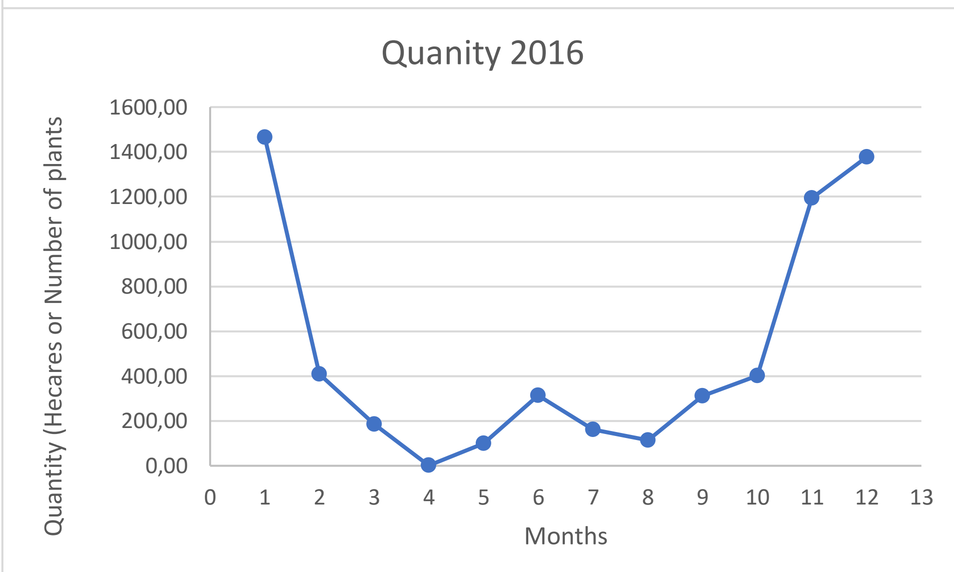
Exhibit 4: Confidence Intervals and Statistical Validation Broken_Down Efficiency

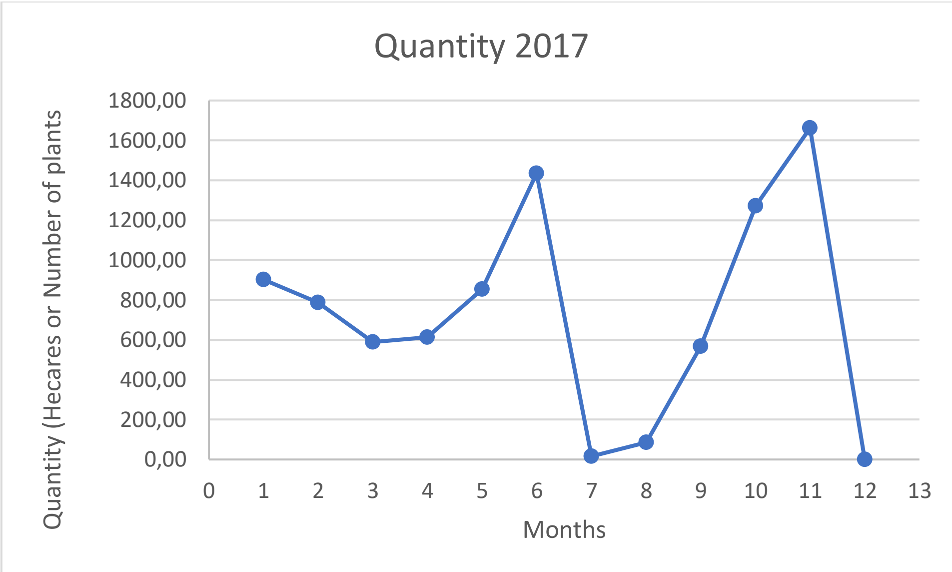
IndexR	Index	frop_name	frop_type	cmp_name	frop_um	froput_ut	FamilyMachine	FamilyImplement	FamilyImplement2	BrokenDownEfficienc	Average Broken down ef	Max	Min	Median	LCL	UCL	Error	Validation
1	4	Levelling be	MECHANIC	HCO Santa A	ha	13476				0,75	0,742	0,9	0,6	0,72	0,717	0,766	0,0106	Validated
1	4	Levelling be	MECHANIC	HCO Santa A	ha	13476				0,8	0,742	0,9	0,6	0,72	0,717	0,766	0,078	Validated
1	4	Levelling be	MECHANIC	HCO Santa A	ha	13476				0,72	0,742	0,9	0,6	0,72	0,717	0,766	0,0298	Validated
1	4	Levelling be	MECHANIC	HCO Santa A	ha	13476				0,72	0,742	0,9	0,6	0,72	0,717	0,766	0,0298	Validated
1	4	Levelling be	MECHANIC	HCO Santa A	ha	13476				0,72	0,742	0,9	0,6	0,72	0,717	0,766	0,0298	Validated
1	4	Levelling be	MECHANIC	HCO Santa A	ha	13476				0,72	0,742	0,9	0,6	0,72	0,717	0,766	0,0298	Validated
1	4	Levelling be	MECHANIC	HCO Santa A	ha	13476				0,72	0,742	0,9	0,6	0,72	0,717	0,766	0,0298	Validated
1	4	Levelling be	MECHANIC	HCO Santa A	ha	13476				0,72	0,742	0,9	0,6	0,72	0,717	0,766	0,0298	Validated
1	4	Levelling be	MECHANIC	HCO Santa A	ha	13476				0,72	0,742	0,9	0,6	0,72	0,717	0,766	0,0298	Validated
1	4	Levelling be	MECHANIC	HCO Santa A	ha	13476				0,72	0,742	0,9	0,6	0,72	0,717	0,766	0,0298	Validated
1	4	Levelling be	MECHANIC	HCO Santa A	ha	13476				0,72	0,742	0,9	0,6	0,72	0,717	0,766	0,0298	Validated
1	4	Levelling be	MECHANIC	HCO Santa A	ha	13476				0,72	0,742	0,9	0,6	0,72	0,717	0,766	0,0298	Validated
1	4	Levelling be	MECHANIC	HCO Santa A	ha	13476				0,72	0,742	0,9	0,6	0,72	0,717	0,766	0,0298	Validated
1	4	Levelling be	MECHANIC	HCO Santa A	ha	13476				0,72	0,742	0,9	0,6	0,72	0,717	0,766	0,0298	Validated
1	4	Levelling be	MECHANIC	HCO Santa A	ha	13476				0,72	0,742	0,9	0,6	0,72	0,717	0,766	0,0298	Validated
1	4	Levelling be	MECHANIC	HCO Santa A	ha	13476				0,72	0,742	0,9	0,6	0,72	0,717	0,766	0,0298	Validated
1	4	Levelling be	MECHANIC	HCO Santa A	ha	13476				0,9	0,742	0,9	0,6	0,72	0,717	0,766	0,2128	Validated
1	4	Levelling be	MECHANIC	HCO Santa A	ha	13476				0,72	0,742	0,9	0,6	0,72	0,717	0,766	0,0298	Validated
1	4	Levelling be	MECHANIC	HCO Santa A	ha	13476				0,9	0,742	0,9	0,6	0,72	0,717	0,766	0,2128	Validated
1	4	Levelling be	MECHANIC	HCO Santa A	ha	13476				0,72	0,742	0,9	0,6	0,72	0,717	0,766	0,0298	Validated
1	4	Levelling be	MECHANIC	HCO Santa A	ha	13476				0,72	0,742	0,9	0,6	0,72	0,717	0,766	0,0298	Validated
1	4	Levelling be	MECHANIC	HCO Santa A	ha	13476				0,6	0,742	0,9	0,6	0,72	0,717	0,766	0,1915	Validated
1	4	Levelling be	MECHANIC	HCO Santa A	ha	13476				0,75	0,742	0,9	0,6	0,72	0,717	0,766	0,0106	Validated
1	4	Levelling be	MECHANIC	HCO Santa A	ha	13476				0,81	0,742	0,9	0,6	0,72	0,717	0,766	0,0915	Validated
1	4	Levelling be	MECHANIC	HCO Santa A	ha	13476				0,75	0,742	0,9	0,6	0,72	0,717	0,766	0,0106	Validated
1	4	Levelling be	MECHANIC	HCO Santa A	ha	13476				0,6	0,742	0,9	0,6	0,72	0,717	0,766	0,1915	Validated
1	4	Levelling be	MECHANIC	HCO Santa A	ha	13476				0,9	0,742	0,9	0,6	0,72	0,717	0,766	0,2128	Validated
1	4	Levelling be	MECHANIC	HCO Santa A	ha	13476				0,63	0,742	0,9	0,6	0,72	0,717	0,766	0,1511	Validated
1	4	Levelling be	MECHANIC	HCO Santa A	ha	13476				0,75	0,742	0,9	0,6	0,72	0,717	0,766	0,0106	Validated

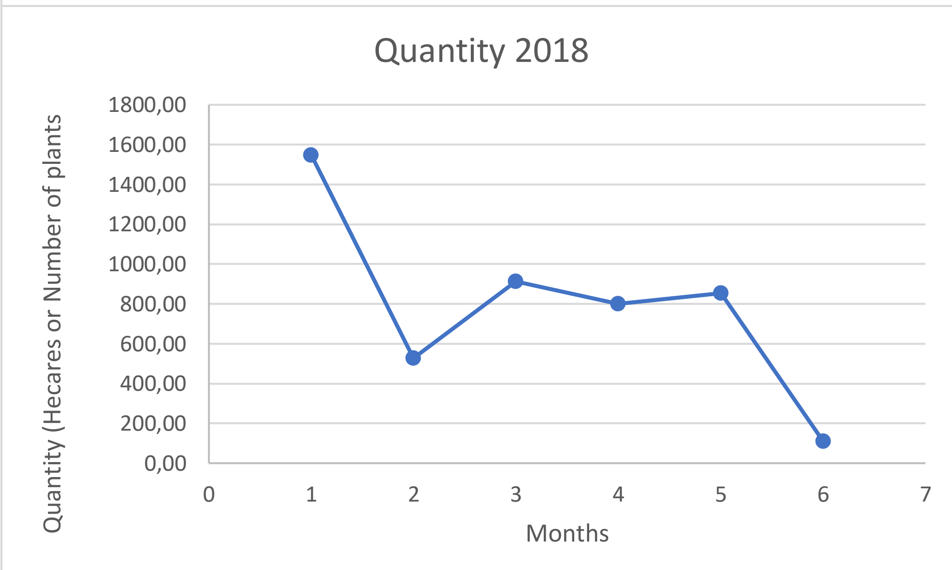
Average Broken down efficiency corresponds to that one computed historically.

Exhibit 5: Historical Behavior of the quantities per month









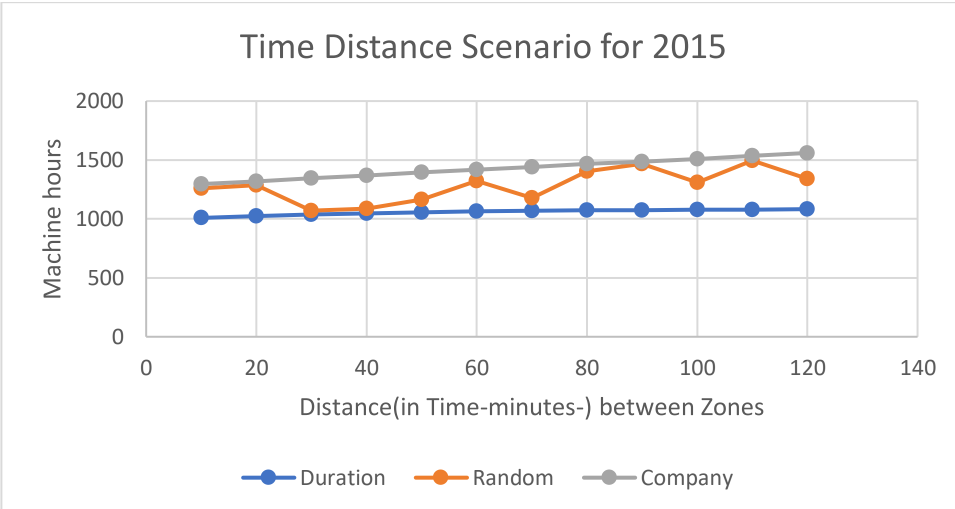
As expected, its behavior is very similar to the one of the Duration. This, because in some sense, the latter is function of the former.

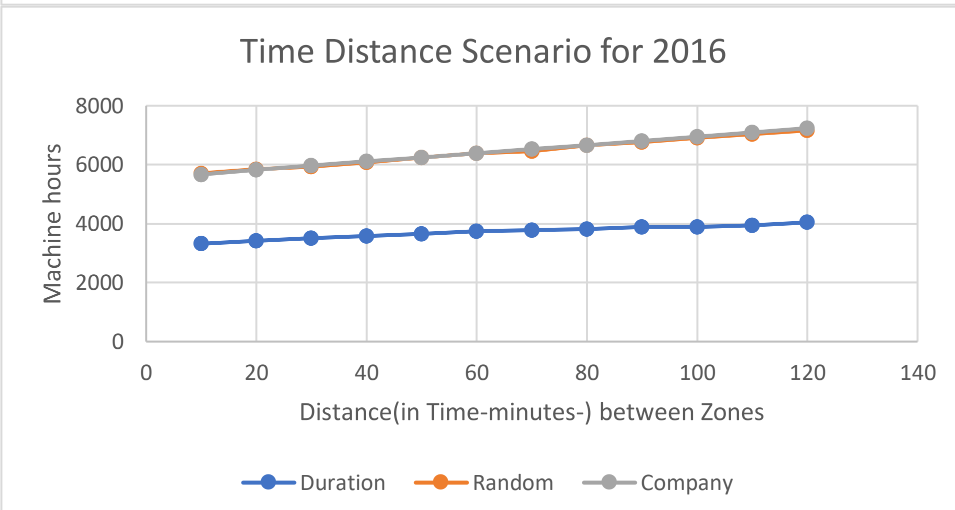
Exhibit 6: Daily Calendar

OPERATIONS SCHEDULED													
Seque	frop_date	Day	frop_operat	frop_name	frop_um	frop_sum_qta	frop_tractor	tractor_name	tractor_family	frimpl1_family	frimpl2_family	Duration	Fuel Consumptio
1	02/01/2018 9:15	02/01/2018	625	Herbicide ap	ha	6,92	772	MOTO 4X4 HO	5	0	0	6	5,5
2	02/01/2018 11:02	02/01/2018	650	Mulcher	ha	7,92	607	LANDINI REX 1	25	36	0	2,9	19,85
3	02/01/2018 11:10	02/01/2018	541	Application	ha	4	0		6	0	0	6,5	28,28
4	02/01/2018 11:11	02/01/2018	650	Mulcher	ha	3,9	1296	TRACTOR LAN	25	36	0	9,2	62,1
5	02/01/2018 11:12	02/01/2018	607	Pesticide ap	ha	1,3	1637	TRACTOR LAN	25	0	0	1	35,27
6	02/01/2018 11:13	02/01/2018	650	Mulcher	ha	6,74	1297	TRACTOR LAN	25	44	0	7	35,76
7	02/01/2018 11:15	02/01/2018	650	Mulcher	ha	6,5	601	TRACTOR PAT	25	33	0	9,9	42,07
8	02/01/2018 11:18	02/01/2018	650	Mulcher	ha	4,5	553	TRACTOR PAT	25	44	0	4	16,2
9	02/01/2018 11:28	02/01/2018	591	Harrowing	ha	5,5	612	Tractor PAT. Y	31	38	0	3,2	26,46
10	02/01/2018 14:48	02/01/2018	607	Pesticide ap	ha	7	1298	TRACTOR LAN	25	44	0	5	54,08
		02/01/2018 Total											
11	03/01/2018 9:10	03/01/2018	607	Pesticide ap	ha	11,85	1637	TRACTOR LAN	25	0	0	6,8	35,44
12	03/01/2018 9:22	03/01/2018	625	Herbicide ap	ha	6,9	772	MOTO 4X4 HO	5	0	0	6	5
13	03/01/2018 9:27	03/01/2018	541	Application	ha	5,44	0		6	0	0	7,8	18,05
14	03/01/2018 9:31	03/01/2018	607	Pesticide ap	ha	11,85	601	TRACTOR PAT	25	36	0	6	39,85
15	03/01/2018 9:35	03/01/2018	607	Pesticide ap	ha	2,45	553	TRACTOR PAT	25	44	0	2,3	54,17
16	03/01/2018 9:36	03/01/2018	607	Pesticide ap	ha	2,45	1298	TRACTOR LAN	25	44	0	3	64,46
17	03/01/2018 9:39	03/01/2018	607	Pesticide ap	ha	2,45	1573	TRACTOR LAN	25	44	0	3	58,17
18	03/01/2018 9:44	03/01/2018	607	Pesticide ap	ha	2,45	1297	TRACTOR LAN	25	0	0	3	41,01
19	03/01/2018 15:14	03/01/2018	650	Mulcher	ha	4,9	607	LANDINI REX 1	25	36	0	3	12,23
20	03/01/2018 15:19	03/01/2018	650	Mulcher	ha	5,5	1296	TRACTOR LAN	25	36	0	9	43,33
21	03/01/2018 15:21	03/01/2018	650	Mulcher	ha	6,12	1297	TRACTOR LAN	25	44	0	3,6	20,02
22	03/01/2018 15:23	03/01/2018	650	Mulcher	ha	5,5	601	TRACTOR PAT	25	33	0	8	37,47
23	03/01/2018 15:27	03/01/2018	650	Mulcher	ha	6	553	TRACTOR PAT	25	44	0	9,1	44,98
		03/01/2018 Total											
24	04/01/2018 9:49	04/01/2018	607	Pesticide ap	ha	9,5	1637	TRACTOR LAN	25	0	0	4	11,24
25	04/01/2018 9:52	04/01/2018	607	Pesticide ap	ha	9,51	601	TRACTOR PAT	25	36	0	5	30,6
26	04/01/2018 9:54	04/01/2018	607	Pesticide ap	ha	3,96	553	TRACTOR PAT	25	44	0	3,2	53,76
27	04/01/2018 10:00	04/01/2018	607	Pesticide ap	ha	3,96	1298	TRACTOR LAN	25	44	0	5	47,76
28	04/01/2018 10:01	04/01/2018	607	Pesticide ap	ha	5,24	1573	TRACTOR LAN	25	44	0	4,1	57,5
29	04/01/2018 10:02	04/01/2018	607	Pesticide ap	ha	5,24	1297	TRACTOR LAN	25	0	0	4	40,66
30	04/01/2018 10:03	04/01/2018	625	Herbicide ap	ha	5,9	772	MOTO 4X4 HO	5	0	0	6	5
31	04/01/2018 14:29	04/01/2018	650	Mulcher	ha	5	607	LANDINI REX 1	25	36	0	3,2	14,34
32	04/01/2018 14:30	04/01/2018	650	Mulcher	ha	5,1	1296	TRACTOR LAN	25	36	0	8	39,07
33	04/01/2018 14:32	04/01/2018	650	Mulcher	ha	5,1	1297	TRACTOR LAN	25	44	0	9,1	36,4
34	04/01/2018 14:34	04/01/2018	650	Mulcher	ha	3	601	TRACTOR PAT	25	33	0	8,1	35,85
		04/01/2018 Total											

This daily calendar has been done for the years 2014 through 2108. The above table is just a snapshot of the document; the original one is much larger.

Exhibit 7: Time Distance Scenarios for years





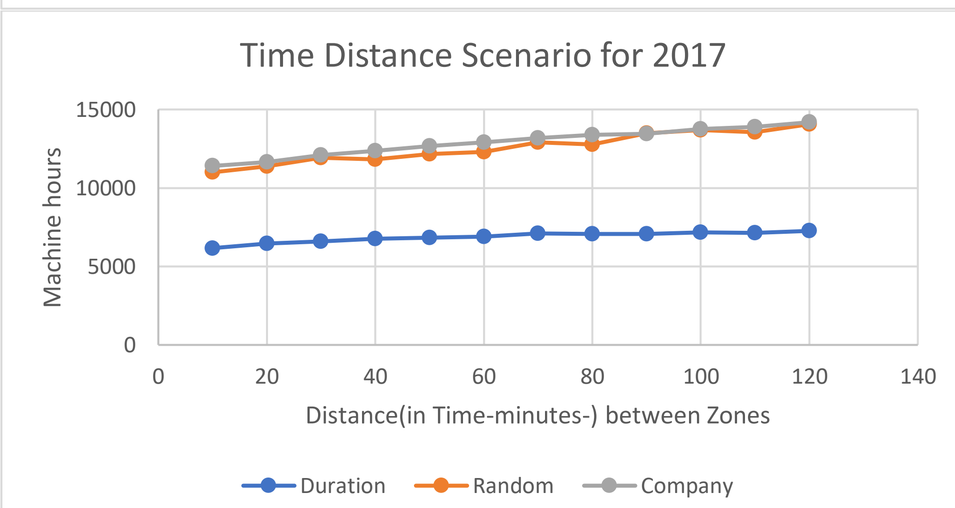
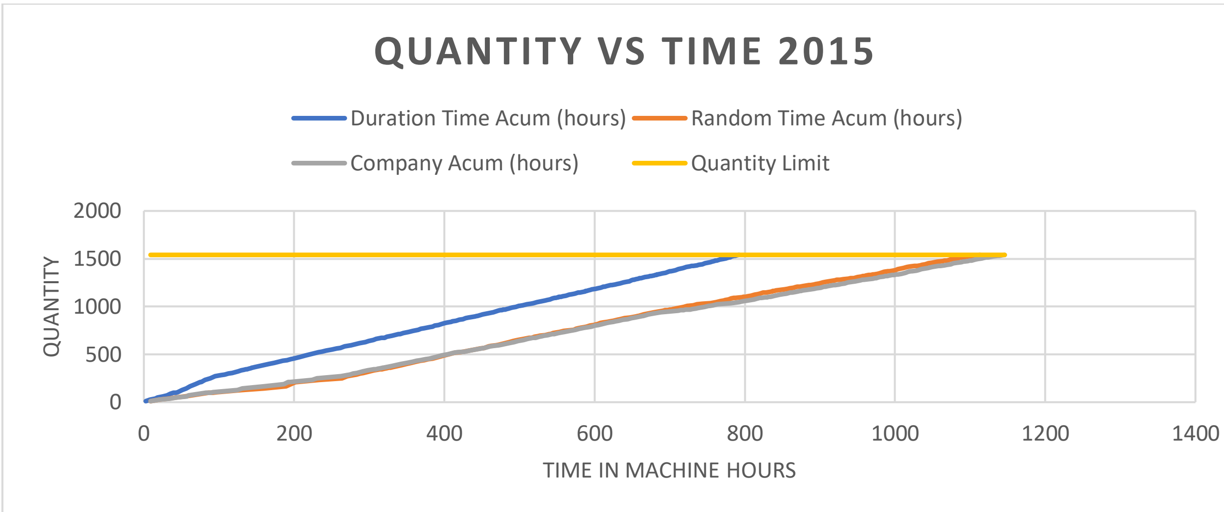
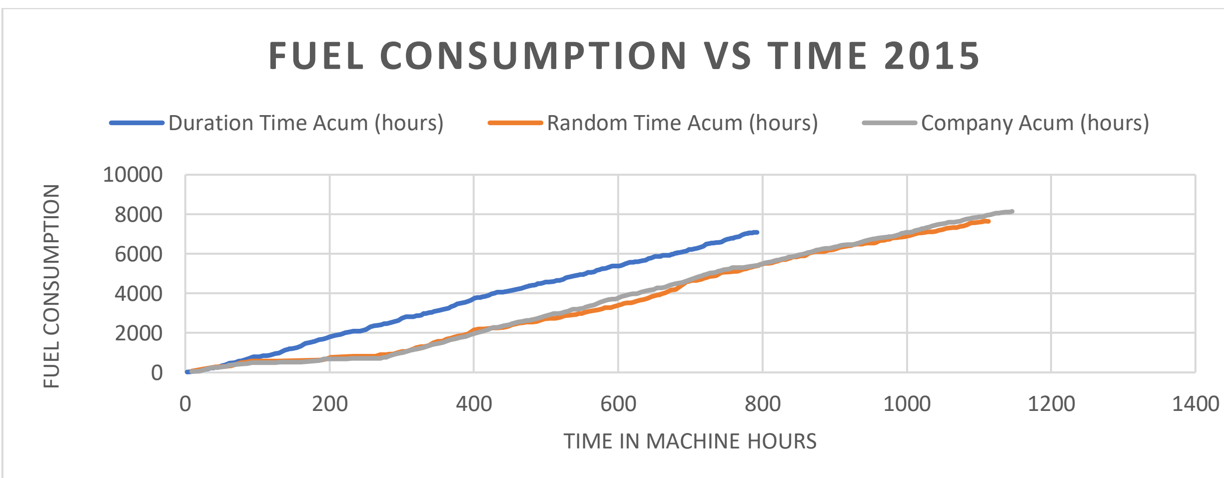
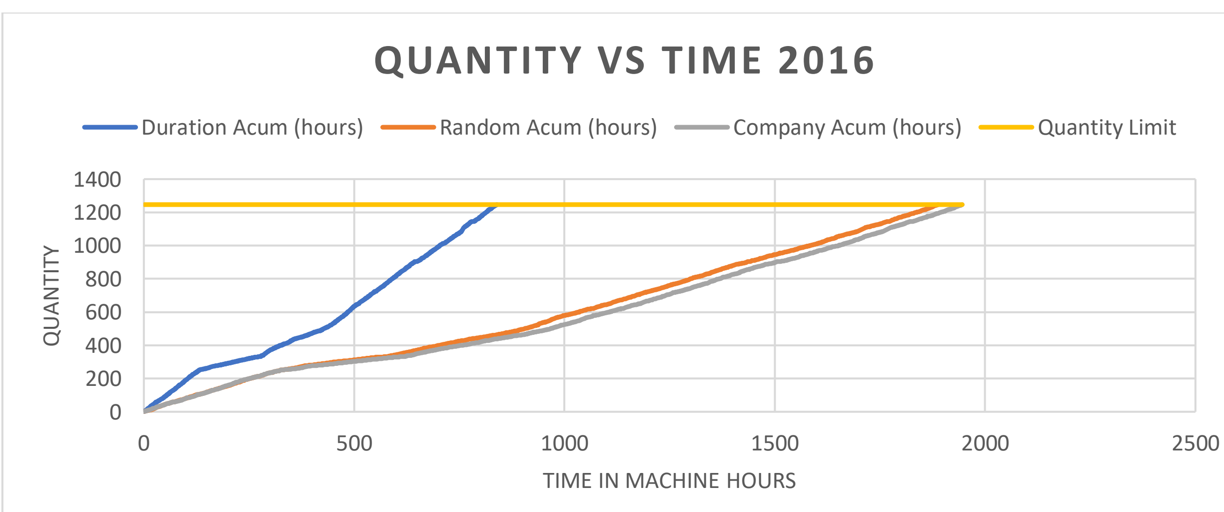
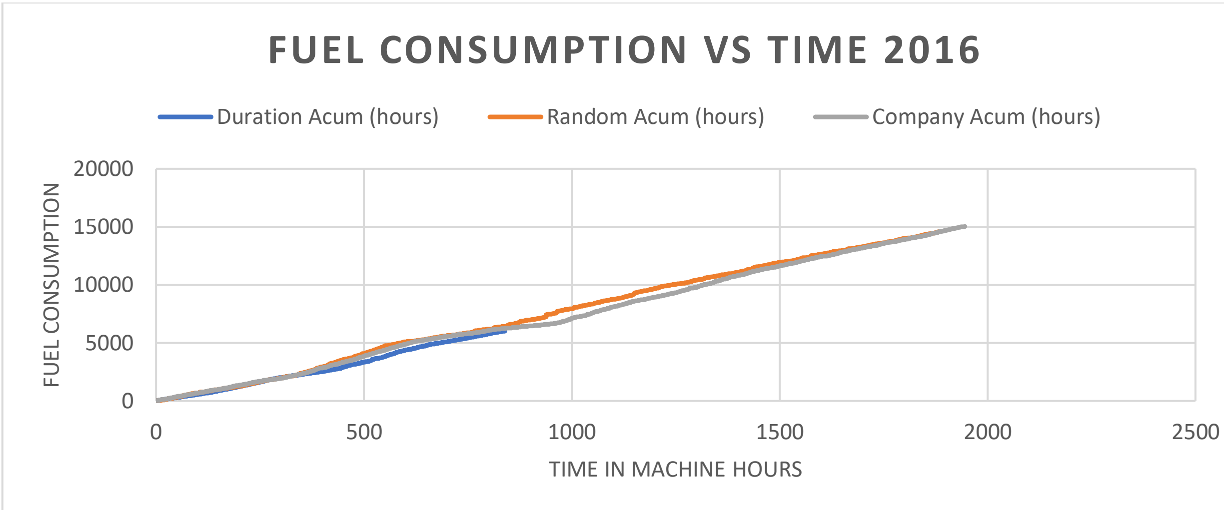


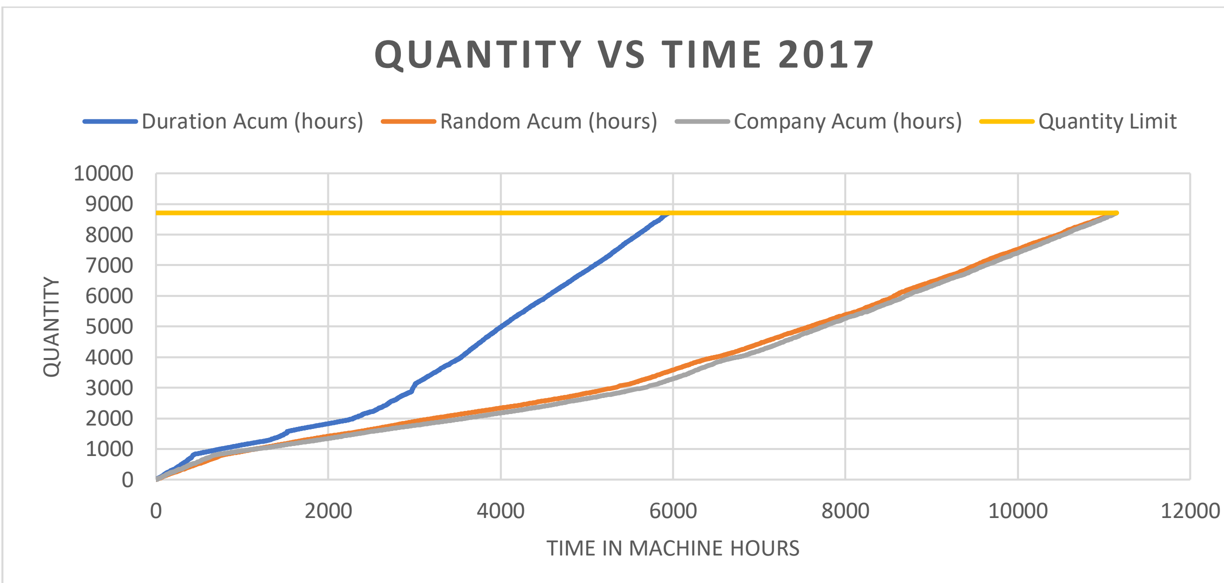
Exhibit 8: Cumulative Comparisons between years for Consumption, Quantity, and Time

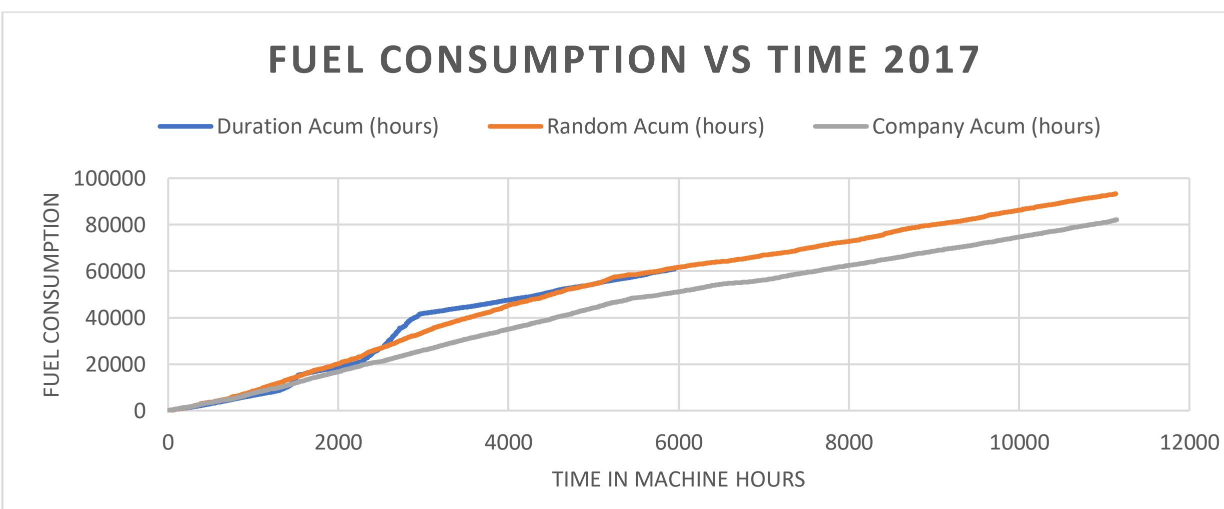







FUEL CONSUMPTION VS TIME 2016
Duration Acum (hours)
Random Acum (hours)
Company Acum (hours)
FUEL CONSUMPTION
0
5000
10000
15000
20000
0
500
1000
1500
2000
2500
TIME IN MACHINE HOURS


QUANTITY VS TIME 2017
Duration Acum (hours)
Random Acum (hours)
Company Acum (hours)
Quantity Limit
QUANTITY
0
1000
2000
3000
4000
5000
6000
7000
8000
9000
10000
0
2000
4000
6000
8000
10000
12000
TIME IN MACHINE HOURS


FUEL CONSUMPTION VS TIME 2017
Duration Acum (hours)
Random Acum (hours)
Company Acum (hours)
FUEL CONSUMPTION
0
20000
40000
60000
80000
100000
0
2000
4000
6000
8000
10000
12000
TIME IN MACHINE HOURS