CERESiS: ContaminatEd land Remediation through Energy crops for Soil improvement to liquid biofuel Strategies

D1.6: Requirements and specifications of DSS

H2020-LC-SC3-2018-2019-2020

Contract No: 101006717

January 2022

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101006717

This project has received funding from the Brazilian Fundação de Amparo à Pesquisa do Estado de Goiás under grant number 202110267000220

This project has received funding from the Canadian New Frontiers in Research Fund under grant number NFRFG-2020-00148 and the Canadian Fond de recherche Société et culture – Québec under grant number 308509

Document control sheet

Project	ContaminatEd land Remediation through Energy crops for Soil improvement to liquid biofuel Strategies
Call identifier	H2020-LC-SC3-2018-2019-2020
Grant Agreement N°	101006717
Coordinator	National Technical University of Athens
Work package	WP1 – Definition of requirements for determining integrated solution pathways
Work package leader	NTUA
Related tasks	T1.6 Requirements and specifications of the CERESiS Decision Support System (DSS)
Deliverable title	D1.6 Requirements and specifications of DSS
Deliverable nature	Report
Dissemination level	PU
Lead Beneficiary	INTRA
Contributing partners	NTUA, EXE, KF
Authors	INTRA: Despina Anastasopoulos, Ioannis Christou, Athanasios Papadakis NTUA: Athanasios Rentizelas, Paraskevas Georgiou, Fragkoulis Psathas.
Reviewer(s)	- D. Giannopoulos (NTUA) - N. Stathopoulos (EXERGIA S.A)
Version	1.0
Total number of pages	36
Issue date	31/01/2022

REVISION HISTORY			
Version	Date	Author(s)	Changes made
0.01	10/12/2021	INTRA	ToC
0.02	16/12/2021	NTUA	ToC additions
0.03	17/12/2021	INTRA	Final ToC
0.04	13/01/2022	NTUA & INTRA	INPUT
0.1	18/01/2022	INTRA	submitted to CERESiS reviewers
0.2	27/01/2022	NTUA, EXE	collected reviewers' feedback
0.3	30/01/2022	INTRA	Final version submitted to the CO
1.0	31/01/2022	NTUA	Final review of the CO

Contents

LIST OF FIGURES

ABBREVIATIONS

Advisory Board: AB
Decision Support System: DSS
Fast Pyrolysis: FP
Functional Requirement: FR
Key Performance Indicator: KPI
Machine Learning: ML
Mathematical programming: MP
Supply Chain: SC
Supply Chain Management: SCM
Supercritical Water Gasification: SCWG
Use Case: UC

1 EXECUTIVE SUMMARY

The purpose of Deliverable 1.6 Requirements and Specifications of Decision Support System (DSS) is to present the preliminary DSS specifications, based on the analysis of the stakeholders and the end user requirements. More specifically, the following chapters explain the work done to collect the user stories by reaching out, not only to all the partners but also extending this exercise by engaging external stakeholders as potential DSS end users as described in chapter 2.

Chapter 2 focuses on providing a more general description on the CERESiS DSS by referring to the DSS perspective, functions, user characteristics, general constraints and assumptions. In a few words, the CERESiS DSS is going to be a cloud-based tool that will enable large user groups with distinct characteristics to assess the viability of an investment in biofuel producing crops and related technologies. This DSS will allow the assessment of various performance indicators (economic, environmental and social KPIs) together with a complex supply chain (SC) design interface.

The collected user stories are translated in specific requirements as presented in Chapter 3. It is worth highlighting that the collected user stories were related to the following topics:

- *Investing in contaminated land*
- *Soil improvement*
- *Choose best process to apply*
- *What can I grow?*
- *Biofuels from contaminated biomass*
- *New conversion technologies*
- *Supply chain network optimization*
- *Input Legal information*
- *Validate / measure the impact of policy / regulation*
- *Provide information on contaminants expected per crop after thermal processes*

In the same chapter the use cases as they derived from the user stories are presented accompanied with the relevant Sequence Diagrams. The Functional requirements have been documented in order to allow us to design the architecture, thus the actual DSS. A Block-and-arrows diagram illustrating the major sub-systems of the entire system at the highest level possible (level-0) of the CERESiS DSS is also presented in this chapter. This high-level architecture depicts the major components of the CERESiS DSS and will feed the work of CERESiS in WP 4 where the DSS is being developed and tested. Additionally, a preliminary evaluation of the performance of alternative modelling and solution approaches for the DSS is performed.

Last but not least, Chapter 4 focuses on the required data input for the DSS. In more detail, it is important to understand the data exchange between the partners and what the DSS and the supply chain model requires.

2 INTRODUCTION

2.1 Purpose of the requirements document

The aim of this document is to present the preliminary DSS specifications, based on stakeholders' analysis and end user requirements. More specifically, this report depicts the process followed for the definition of the use cases and user stories for the CERESiS DSS and also the functional requirements that emerged from this process.

In order to facilitate the process of collecting the user stories from the CERESiS partners, members of the Advisory Board (AB) and other potential end users a template was created (Annex 1) and also an online open survey https://www.linkedin.com/feed/update/urn:li:activity:6867004088479948800/. The methodological approach to collect the requirements and the specifications for the DSS was as follows:

First step: At first the template was presented and explained to the consortium, then all partners, depending on their expertise provided input, and then INTRA, with the support of NTUA, conducted one-on-one sessions with partners in order to elaborate the collected input.

Second step: The consortium reached out to the AB inviting them to one-on-one short interviews. During those interviews with the AB members, INTRA had an open discussion with them.

We wish to highlighted that the collected input from the AB matched with the existing collected user stories/requirements. The discussions with the AB members were extremely helpful since due to their expertise and being closer to the end used profile, they validated the existing collected input and highlighted that everything was on the right track. In an effort to further populate the collection process, EXE took the lead by sending targeted emails to external audience which consist of partners of the value chain and potential users of this DSS, in order to collect their insight on this matter by providing their input, thoughts, and suggestions.

It is worth mentioning that these potential end users were requested to fill the statement "As a < type of user >, I want < some goal > so that < some reason >". The User stories are short, simple descriptions of a feature told from the perspective of the person who desires the new capability, usually a user or customer of the system.

2.2 Scope of the CERESiS DSS

The scope of the CERESiS DSS is to facilitate end-users such as: Farmer associations, External investors, Land owners, Land operators, Regional authorities, Plant Owners or/and Biofuel producers to take informed decisions on topics related to i.e. i) Investing in contaminated land, ii)Soil improvement, iii) Choose best process to apply iv) biofuel production, see more in chapter 3. It may also be used by Policy makers in order to assess the potential impact of policies on the feasibility of biofuel value chains. This will be done by using the CERESiS cloud based tool that will be equipped with KPI calculator for computing various economic, environmental and social indicators (KPIs).

2.3 DSS perspective

The CERESiS Decision Support System (DSS) is going to be a cloud-based tool that will enable large user groups with distinct characteristics to assess the viability of an investment in biofuel producing crops and related technologies. It is going to allow interested parties to estimate what process conversion technology is most appropriate for an area given the characteristics of the area such as the type of contaminants present in the area lands, the extent of contamination, the climatic conditions of the area, the soil characteristics etc. It is also going to allow potential investors to estimate economic as well as environmental indicators of a possible investment in a process technology in an area. Finally, it is going to support the very ambitious goal of optimizing an entire supply network design so as to ensure the viability of an investment in biofuel related technologies for land decontamination.

2.4 DSS functions

At its heart, the DSS provides a KPI calculator for computing various economic, environmental and social indicators (KPIs) together with a complex SC design interface that allows interfacing with state-of-the-art optimization solvers that solve large-scale NP-Hard problems to optimality so as to provide an optimal solution to a complex location/allocation network design problem with complex objectives and constraints. The full list of functionalities of the DSS is specified in section 4 below.

2.5 User characteristics

In the process of extracting user-stories, the following different user roles (the "who" part of the user-story) were identified:

1. Farmer association

2. External investor

3. Landowner

4. Land operator

5. Regional authority/policy makers

6. Plant Owner

7. Biofuel producer

2.6 General constraints

The major constraints of the DSS are listed as “Non-Functional Requirements” in section 4.4.

2.7 Assumptions and dependencies

The development of the DSS assumes that the process technologies being investigated within the project will not be obsoleted in the near future; belief in this assumption has been reinforced after discussions between the developers and consortium partners and expert members of the advisory board that confirmed their belief in these technologies as well.

Development of the DSS also takes for granted that all users that will be using the DSS are computer literate in the sense that they know how to use a browser from a desktop computer, and that in fact they have access to computers connected to the Internet and are able to understand basic geographical information and input this information in the DSS.

3 SPECIFIC REQUIREMENTS

3.1 User stories

1. **Investing in contaminated land**

   Description: DSS user wants to invest in contaminated land in order to grow energy crops for biofuel feedstock.
   Who: Farmer association/Problem owner/External investor
   What: Using the results of the project, the end-user must be able to enter data regarding their plot and see a list of the best options for appropriate crops to cultivate in their field. The list should be accompanied with associated KPIs for each proposed choice, including expected yield, expected cost structure and components, estimated total production cost.
   Why: User needs to be able to see the best available solutions to assess the economic viability of potential investment. Environmental aspect is of less importance.
   Inputs: Land cost, Land plot coordinates, land plot area, potential crops to be used. Production cost components: seeding, cultivation, harvesting, (optional: crops value/selling price).
   Outputs: Total production cost. Expected NPV/IRR of investment (if crop value/selling price is known).

2. **Soil improvement**
   Description: DSS end-user (farmer or regional authority) needs to know how the contaminated land can best be decontaminated/managed and reclaimed.
   Who: Land operator/Problem Owner (farmer, land agent, etc.), Regional Authority
   What: Ensure that the optimum land decontamination/management strategy can be identified for the particular land characteristics, contaminant and climatic conditions.
   Why: To understand how and when land can be decontaminated/managed.
   Inputs: Land location, soil contaminants, climatic conditions.
   Outputs: Information regarding the ideal energy crops to use and optimum crop management strategy to adopt. When land will reach a particular desired level of decontamination.
   Functional requirements: Choose the best crop to decontaminate/manage your land and estimate time to reach a specific level of contaminant concentration.

3. **Choose best process to apply**
   Description: The stakeholder needs to choose which is the best process to produce bio-fuels based on techno-economic and environmental constraints.
   Who: Plant owner

What: Using the results of the project, the stakeholder must be able to enter data regarding the crop location and characteristics (production rate, basic characterization, contamination) and see the list with associated KPIs for both the thermochemical processes considered in the project including expected bio-fuel mass and energy yields, expected capital and production costs for the chosen technology, revenues from products and by-products.

Why: Stakeholders need to be able to see the best results recommended by the DSS so they can make an informed choice

Inputs: Data regarding the crop location and characteristics, potentially based on the project results for energy crops investigated, or external input if not investigated within the project (production rate, basic characterization, contamination).

Outputs: Expected biofuel mass and energy yields, expected capital and production/operational costs for the chosen technology, revenues from by-products considering different valorization routes.

4. **What can I grow?**

Description: The farmer needs to be able to easily enter information that he/she normally has for the land and in a streamlined way display the options for deciding which biomass to use and which processing system to feed it into.

Who: Farmer

What: The farmer should enter data on the soil characteristics of his/her land, the climate and, if known, the pollutants present. They should be able to choose between different crops, with information on production and prices, but also the different possibilities of the commercial product to be able to direct them towards the supply chain closest to their location.

Why: To be able to make the most appropriate decision on biomass type selection and match with existing supply chains.

Inputs: Geographical coordinates, Soil characteristics, Climate parameters, Plot size, existing supply chains in the region.

Outputs: List of species (also with pictures), with growth characteristics, productivity and ecology, crop recommendations, pollutant absorption capacity. Processing system to feed biomass into. Possible supply chains to deliver the collected biomass to.

5. **Biofuels from contaminated biomass**

Description: DSS user wants to utilize contaminated biomass feedstock for biofuel production.

Who: Biofuel producers

What: Using the results of the project, the end-user must be able to enter data regarding the quality and quantity of contaminated biomass feedstock and see a list of the best options for biofuel production and safe management of contaminants. The list should be accompanied with associated KPIs for each proposed choice, including expected cost structure and components, estimated total production cost and environmental impacts.

Why: User needs to be able to see the best available solutions in order to assess the economic viability and the environmental aspect of producing biofuel from contaminated feedstocks.

Inputs: Pretreatment type (loose, bale, pellets etc.) time, contaminants present on the land plot and cost of biomass feedstock. Candidate facilities locations. Facilities fixed costs (construction/purchase) and facilities operating costs. Handling and storage, labor, maintenance & repairs and production processing. Number of mobile processing facilities and mobile facility relocation cost. Time needed to install/uninstall mobile units, average mobile unit velocity and total number of trucks. Truck purchase/rental price, transportation cost between facilities and average truck velocity. Storage unit cost and storage capacity (storage price per unit). Capacity of static and mobile facilities, bioenergy plant conversion rate of biomass unit. Price of converting biomass, biofuel selling price and Taxes & insurance.

Outputs: Total production cost (composed by feedstock cost, pre-processing, transportation, storage, conversion, post-processing towards existing transport fuel quality). Expected NPV/IRR of investment. Corresponding environmental impact indicators of production chain (feedstock-to-fuel).

6. New conversion technologies

Description: DSS user wants to invest in new biomass conversion technologies, able to utilize contaminated biomass

Who: Conversion system providers

What: Using the results of the project, the end-user must be able to enter data regarding the quality and quantity of contaminated biomass feedstock (after pretreatment) and see a list of the best conversion options for biofuel production and safe management of contaminants. The list should be accompanied with associated KPIs for each proposed choice, including expected cost structure and components, estimated total conversion cost and environmental impacts.

Why: User needs to be able to assess the economic viability and the environmental aspect of a range of contaminated feedstocks conversion technologies for biofuel production.

Inputs: Contaminants which are present in the biomass feedstock (ready for conversion), the cost of biomass feedstock after pretreatment. Candidate facilities locations, conversion facilities fixed costs (construction/purchase) and conversion facilities operating costs. Labor, maintenance & repairs and production processing. Capacity of static and mobile facilities. Bioenergy plant conversion rate of biomass unit, price of converting biomass, biofuel selling price and finally Taxes & insurance.

Outputs: Total conversion cost (composed by pretreated feedstock cost, conversion, - not including post-processing towards existing transport fuel quality). Expected NPV/IRR of investment. Corresponding environmental impact indicators of production chain (gate-to-gate).

7. Supply chain network optimization

Description: DSS end-user needs to be able to define the optimal design of the supply chain network to decontaminate land with the best economic outcome.
Who: Problem owner (land owner or manager), Policy maker.
What: The end-user must be able to enter data related to the supply chain network and to see the best option recommended by the optimization model.
Why: User needs to be able to see the optimal solution under specific constraints to make an informed choice about how to design the supply chain network.
Inputs: Land cost, land plot area and coordinates. Type of crop used, biomass volume/mass(density) available on the land plot, rate of the crop harvesting. Pretreatment type (loose, bale, pellets etc.) time, contaminants present on the land plot. Candidate facilities locations, facilities fixed costs (construction/purchase) and facilities operating costs. Handling and storage, maintenance & repairs and production processing. Number of mobile processing facilities, mobile facility relocation cost, time needed to install/uninstall mobile units. Average mobile unit velocity, total number of trucks, truck purchase/rental price. Transportation cost between facilities, average truck velocity. Storage unit cost, storage capacity (storage price per unit) and capacity of static and mobile facilities, bioenergy plant conversion rate of biomass unit, price of converting biomass, biofuel selling price and finally Taxes & insurance.
Outputs: Supply chain network visible to authorized end-user. Number, type, location and size of processing plants and mobile processing facilities. Flow rates of materials. Amount of biofuel produced per time period. Allocation & relocation of mobile processing facilities at certain time periods. System's profit/cost.

8. Input Legal information

Description: As a responsible for parametrization, I want a template for local legal aspects so the system can be updated.
Who: Responsible for parametrization.
What: legal aspects template information.
Why: automate system parameters updates
Inputs: template document with local legal aspects
Outputs: legal report

9. Closing the fertilizer cycle - recovering nutrients

Description: End-user needs to see which nutrients (P) can be recovered during hydrothermal conversion of different crops.
Who: agricultural sector
What: Using the results of the project, the end-user should be able to check to what extent it is possible to recover nutrients from different crops by entering data on the plants (species, harvest time, soil etc.).
Why: Users needs to see whether it is worthwhile to aim at nutrient recovery in a specific use case.

Inputs: Average composition of biomass depending on crop data, separation efficiencies for different plants.
Outputs: The amount of phosphorus that can be recovered depending on crop data.

10. **Validate / measure the impact of policy / regulation**

Description: DSS end-user needs to validate / measure the impact of policy or regulation.
Who: Policy-maker / Regional & municipal authority.
What: Using the tool, the end-user must be able to acquire information for various scenarios of phytoremediation and biofuel production. The user can provide initial information (contaminated land and energy crop) and expects data on produced biofuel and decontamination achieved.
Why: User needs to be able to quantify the impact of proposed policy/regulation, so that they can choose the best policy/regulation among alternative options according to local conditions.
Inputs: Surface and characteristics of contaminated land, type of contaminant, type of energy crops, policy/regulation-related parameters.
Outputs: Type and quantity of produced biofuel, cost of produced biofuel, other economic, environmental and social KPIs.

11. **Provide information on contaminants expected per crop after thermal processes**

Description: The design of decontamination processes needs to be customized per case (crop, contaminated soil), also considering the conditions/specifications of thermal treatment processes and final product (fuel) requirements.
Who: The designer of decontamination processes of gaseous & liquid products.
What: Using the results of the project, the end-user must be able to obtain data to identify best options for appropriate customization (e.g. choice of most appropriate materials/key components combinations, enrichment in terms of cleaning/decontamination steps if needed) of decontamination processes. The list should be accompanied with associated KPIs in terms of critical contaminants removal.
Why: User needs to be able to see a representative range of products compositions derived by the thermal treatment processes provided to them by the DSS so they can make an informed choice in terms of (re-)design/customization of decontamination steps.
Inputs: Soil conditions, range of suitable crops, estimated suitable operating conditions of thermal treatment processes. Mathematical formulas/models from combined experimental and simulation results.
Outputs: Estimated expected range of key compounds' concentrations to be utilized by the user for the design/optimization of decontamination processes/steps.

3.2 Main Use Cases (UCs)

The Use Cases derived from the abovementioned user stories that were collected via interviews and surveys.

3.2.1 Overall (level-0) Use Case diagram

The level-0 Use Case diagram is shown below:

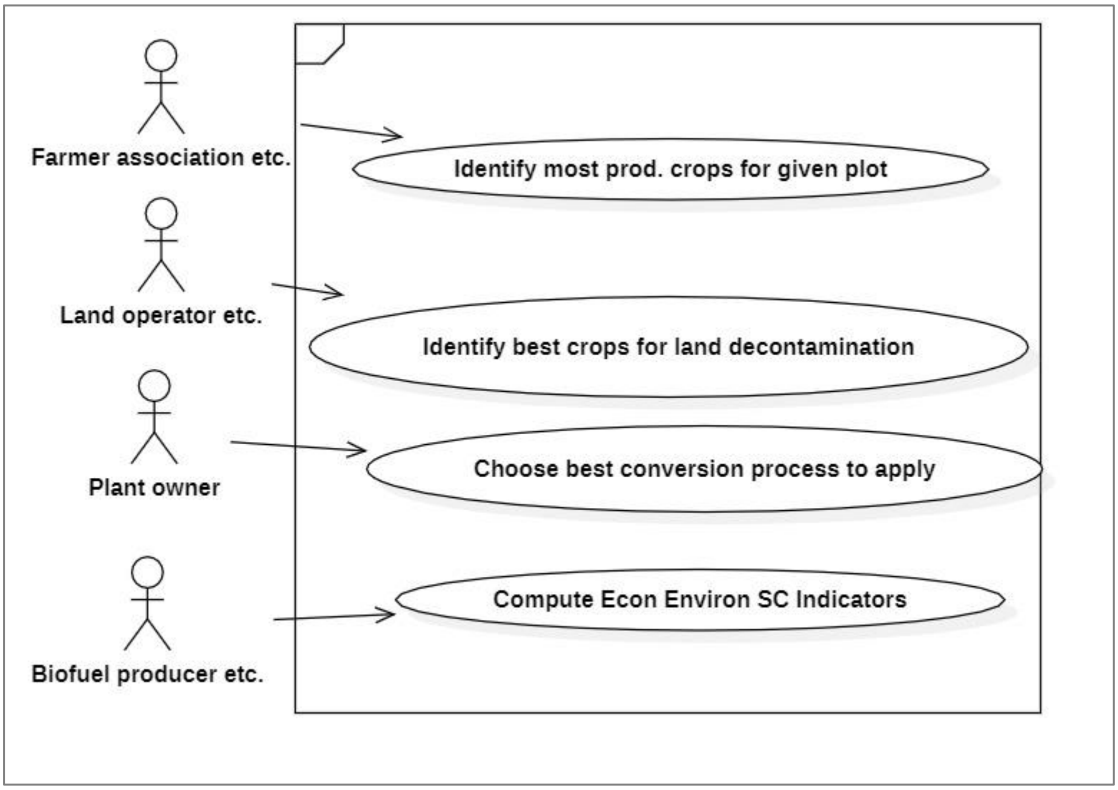

Figure 1 level-0 Use Case diagram

3.2.2 Analysis of major UCs

Upstream (Biomass production) KPI calculation related UCs

Use Case 1: Identify most productive crops for given plot

Actor: Farmer association/Problem owner/External investor

Main Sequence of Events:

1. User enters plot coordinates defining land as a polygon (coordinates may be entered either as a series of GPS longitude, latitude pairs, or with the help of a GIS system showing ortho-photos on which the user can click to define the area of their intended land)

2. System computes the area to check, highlights the surface on the screen, and asks the user to verify the land is correctly shown.

3. User replies positively.

4. System computes for each available crop in its database, the expected yield given the soil characteristics, and presents the results sorted in descending order of expected yield.

Alternate Scenarios:

Step 3: User replies negatively. In this case, the system asks user to repeat step 1 and continues from there.

Step 4: System databases do not include land characteristics for computing expected yield for any crop in the user-specified land. In such an event, the system asks the user to enter land characteristics before proceeding. Land characteristics are saved into the system database.

Use Case 2: Identify best crops for land decontamination

Actor: Land operator/Problem Owner (farmer, land agent, etc.), regional authority

Main Sequence of Events:

1. User enters plot coordinates defining land as a polygon (coordinates may be entered either as a series of GPS longitude, latitude pairs, or with the help of a GIS system showing ortho-photos on which the user can click to define the area of their intended land)

2. System computes the area to check, highlights the surface on the screen, and asks the user to verify the land is correctly shown.

3. User replies positively.

4. System asks user to enter pollution levels in area for each pollutant known to the user.

5. User enters list of pollutant levels in the selected land area.

6. System computes decontamination rates for given soil for every suitable biomass type available in the database and presents the results sorted in descending order of decontamination performance expected.

Alternate Scenarios:

Step 3: User replies negatively. In this case, the system asks the user to repeat Step 1 and continues from there.

Step 6: System does not have formula for computing decontamination rates for given soil pollutant for a crop: it is assumed that the given crop growing in this land cannot help removing pollutants from land.

Conversion Process related UCs

Use Case 3: Choose best conversion process to apply

Actor: Plant owner

Main Sequence of Events:

1. User enters crop location, potential production technologies to consider, their production rate, basic characterization, and field contaminants.

2. System computes and displays the following:

1. expected biofuel mass and energy yield
2. Expected capital and production (variable) costs for input production technology
3. Expected revenues from main process outputs and from main process by-products

Alternate Scenarios:

Step 2: If the user requires an investigation as to whether it is worthwhile to aim at nutrient recovery during hydrothermal conversion, a calculation of nutrient recovery amounts and value is performed

Supply chain & Economic performance related UCs

Use Case 4: Compute Economic and Environmental Performance Indicators at supply chain level

Actor: Biofuel producer/Investor/Policy maker/Regional authority/Problem owner

Main Sequence of Events:

1. User inputs the following data into the system: Potential biomass types for consideration and harvesting times, Potential Biomass pretreatment type (loose, bale, pellets etc.), Geographical location of available land plot(s) for biomass cultivation, of candidate facilities locations and of biofuel end users (optional: if known), Contaminants present on the land plot(s), Cost of biomass feedstock , Facilities fixed costs (construction/purchase), Facilities operating costs, Indication on whether to consider mobile facilities, Mobile facility relocation cost and time, Average mobile unit velocity, Truck purchase/rental price, Transportation cost (or distances) between facilities, Average truck velocity, Storage unit cost and capacity (storage price per unit), Capacity of static and mobile facilities, Biofuel selling price, Taxes & insurance, ultimate objective of user (cost minimization or profit maximization).

2. The system solves a corresponding location/allocation optimization problem similar in nature to the p-Median problem.

3. The system, having obtained the optimal supply chain design, computes and displays the following (indicative) supply-chain wide economic & environmental performance information: Total production cost and cost breakdown per supply chain stage and key cost category, expected NPV/IRR of investment, corresponding environmental impact indicators of production chain (feedstock-to-fuel), details on system design (facilities numbers, sizes, locations and flows of materials between them).

Alternate Scenarios:

Step 1: If the user is interested in investigating a particular conversion technology, they can limit the conversion technology data input to a single technology of interest to obtain the performance outputs (User story 6)

Step 4: If the user requires to run 'what if scenarios' on the optimal supply chain design obtained at step 3, an additional UI will offer this opportunity for assessing the performance of the optimal solution by changing some key problem parameters (User story 10).

If the user requires to run 'what if scenarios' and identify the optimal supply chain design for each scenario, the steps 1 –3 can be performed iteratively for each scenario to produce the required results (User story 10).

3.2.3 Mapping User Stories to UCs

User Story 1 is mapped exactly on Use Case 1 "Compute Best Crops for Land".

User Story 2 is mapped exactly on Use Case 2 "Compute Best Crops for Land Decontamination"

User Story 3, 9 & 11 is mapped exactly on Use Case 3 "Choose Best Process to Apply"

User Story 4 is mapped exactly on Use Case 2: "Compute Best Crops for Land Decontamination"

User Story 5, 6, 7 & 10 are mapped exactly on Use Case 4: "Compute Economic and Environmental Performance Indicators at supply chain level"

3.2.4 Sequence diagrams per UC

Use Case 1 Sequence Diagram:

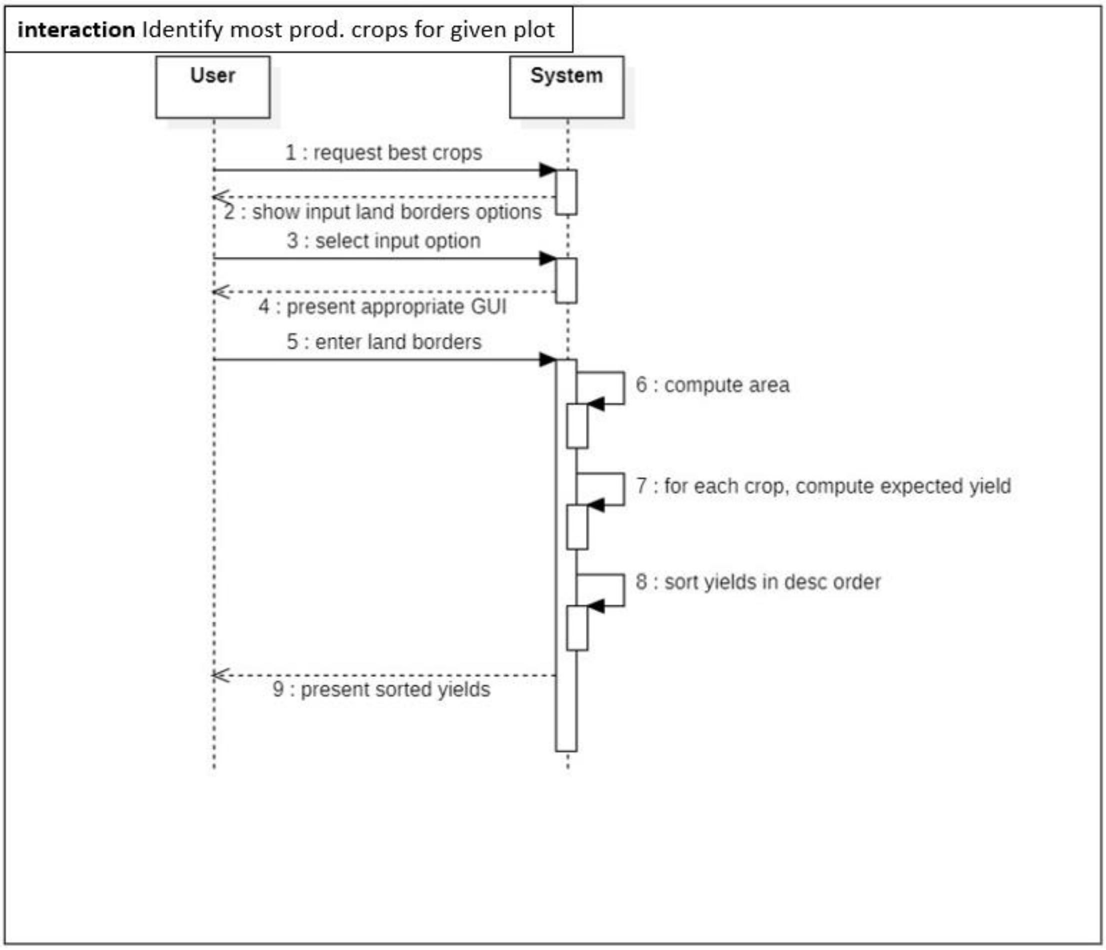

Figure 2 Use Case 1 Sequence Diagram

Use Case 2 Sequence Diagram:

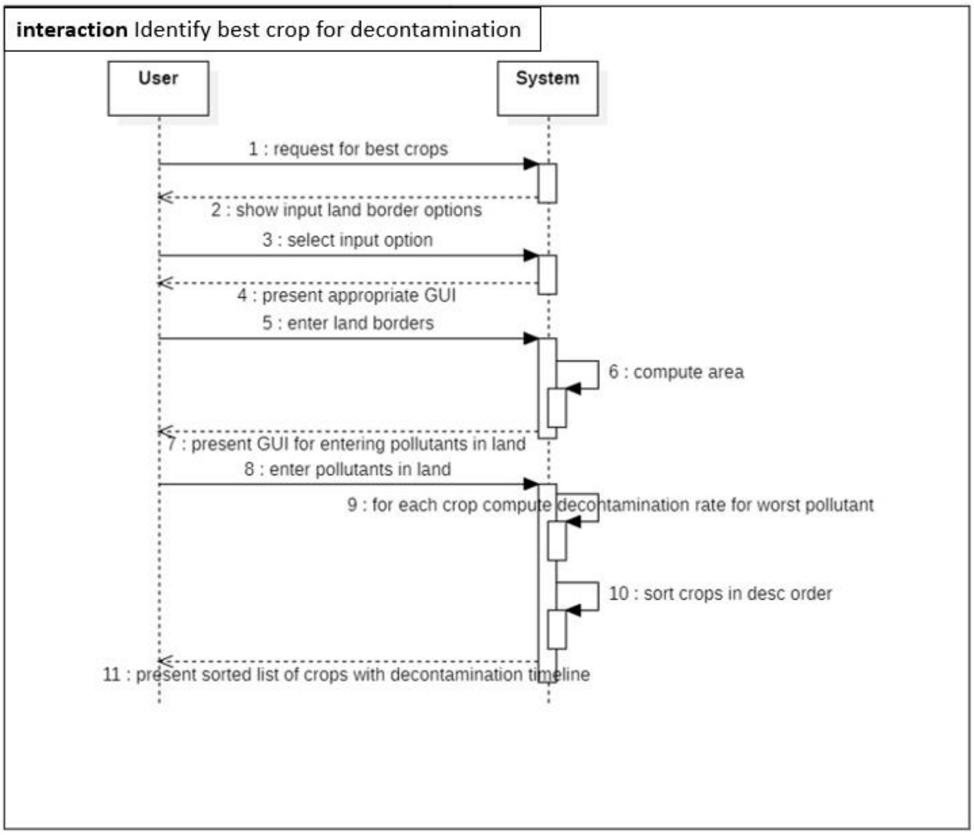

Figure 3 Use Case 2 Sequence Diagram

Use Case 3 Sequence Diagram:

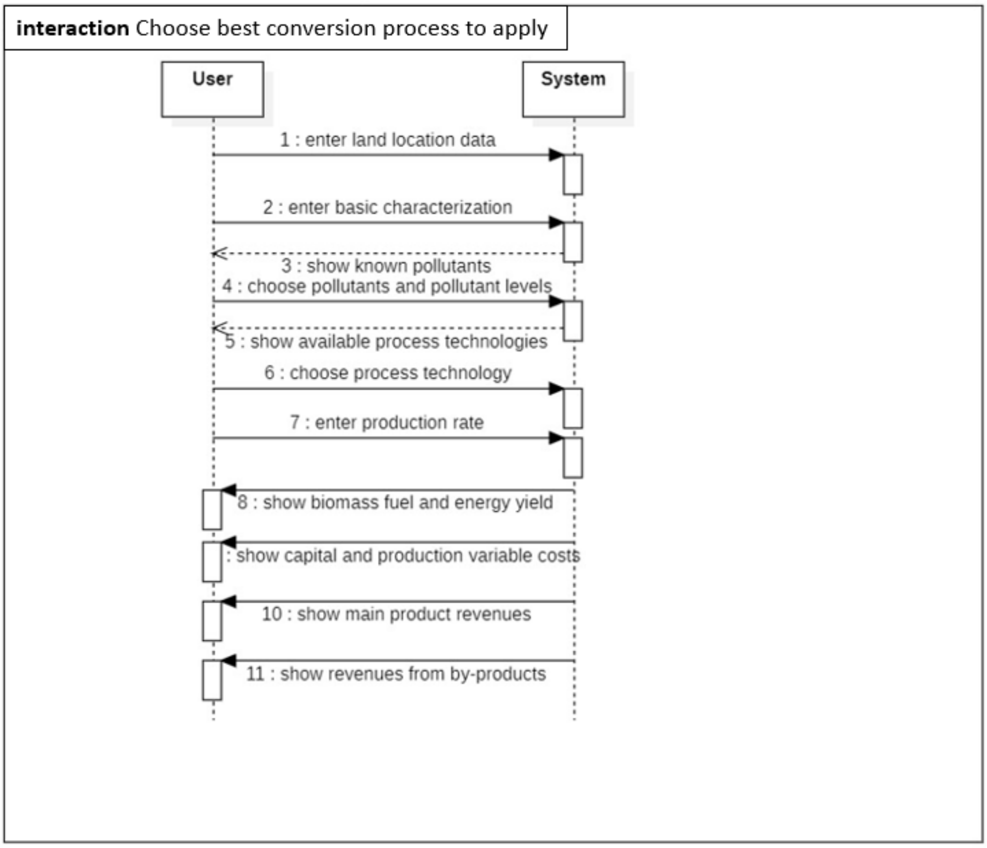

Figure 4 Use Case 3 Sequence Diagram

Use Case 4 Sequence Diagram:

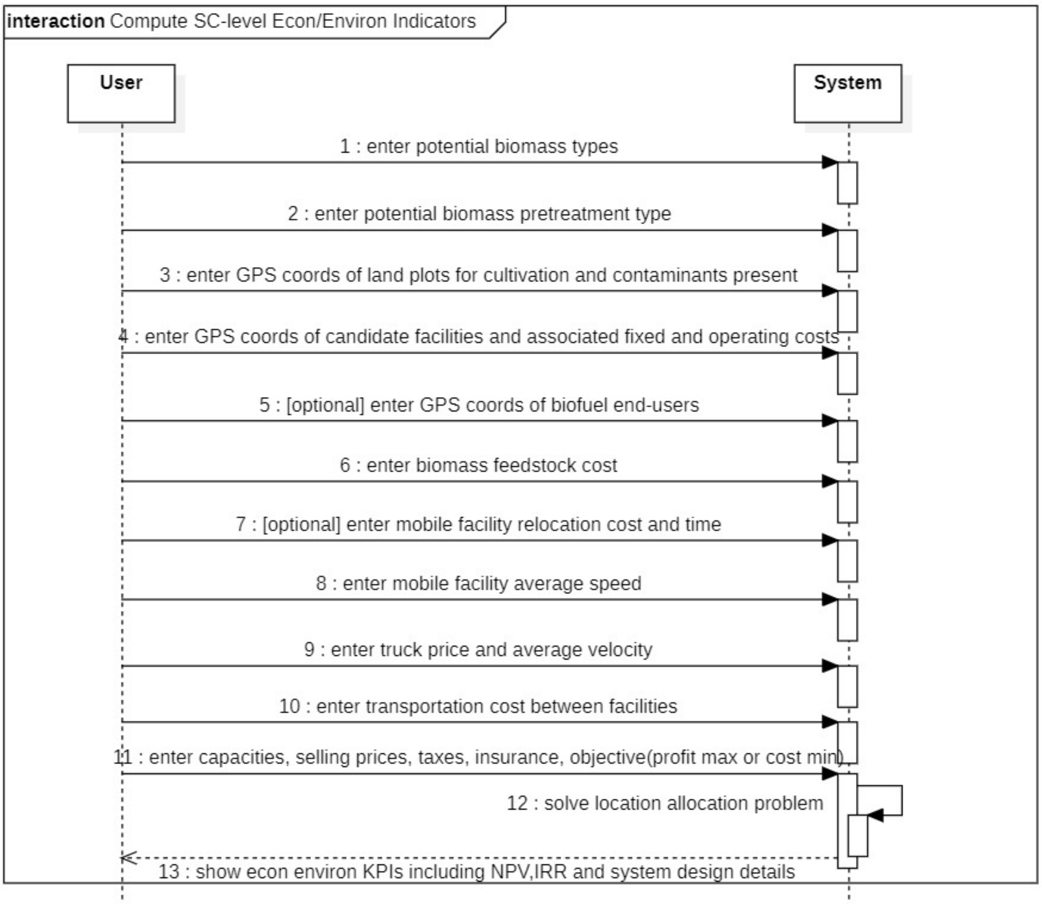

Figure 5 Use Case 4 Sequence Diagram

3.3 Functional requirements specifications

FR1: The system shall compute expected yield for any crop type for any given land area. The land area will be provided as a polygon with a fixed set of (lon,lat) GPS measurement pairs. Given the characteristics of the general area where the land is located, the system shall have in its database a formula for the expected yield of a crop per unit of land area per season. Applying this formula, and multiplying by the total area of the land, the expected yield for this type of crop shall be computed.

FR2: The system shall compute decontamination time to specific contaminant concentration levels for any crop type for any given land area for which it is known the level of pollution with any possible pollutant type. Given pollution levels in a given land area, the system shall be able to compute the time it will take to remove pollutants from the land when cultivated with any given energy crop/biomass to a given level (e.g. 5 years to reduce presence of contaminant to less than 1 ppm).

FR3: The system shall compute expected biofuel mass and energy from a land area given a process technology. Using the functionality in FR1, the system shall compute crop yield, and will then compute mass of biofuel and energy yield extracted from crop using the particular process technology.

FR4: The system shall compute by-products of a given process technology for given crops.

FR5: The system shall compute economic indicators arising from the solution of location/allocation problems involving fixed as well as mobile facilities and fleets. The location/allocation problem will include all stages of the supply chain, from energy crops production to biofuel delivery to end customers. The economic indicators are based on the standard NPV, IRR and other formulae, but the total production costs are heavily influenced by the ability of the system to obtain the globally optimal solution to the supply chain design optimization problem.

FR6: The system shall allow the investigation of 'what if' scenarios, based on different values of key parameters. It shall use the functionality in FR5

3.3.1 Mapping UCs to functional requirements

- UC1 maps directly to FR1.
- UC2 maps directly to FR2.
- UC3 maps directly to FR3 and FR4. FR3 depends on FR1.
- UC4 maps directly to FR5 and FR 6.

3.4 Non-functional requirements specifications

3.4.1 Performance constraints

The system is expected to run in a cloud environment, so it is expected that a high-end server with significant CPU and RAM resources will be available; in particular it is expected that at least 8 CPU cores running at a base frequency of at least 2.5 GHz will be available, and at least 64 GB of DDR4 memory will be available. On the other hand, the system should not assume the presence of high-end GPU (or TPU) accelerators for linear algebra or other rendering operations. Similarly, the system should not assume the availability of multiple servers or distributed computing infrastructure.

3.4.2 Environmental constraints

There is no constraint regarding the implementation languages, but Java is usually a good fit for cloud native applications. Further, Python is a language that fits well as a glue language for sub-systems integration (replacing Tcl/Tk) as well as interface to optimization software (GUROBI etc.) and has bindings to many Machine Learning libraries (Tensorflow etc.).

3.4.3 Architectural constraints

The DSS software needs to be implemented as a cloud-native application, according to the description of the work of the project.

3.5 High-level system architecture

The following diagram is a Block-and-arrows diagram illustrating the major sub-systems of the entire system at the highest level possible (level-0).

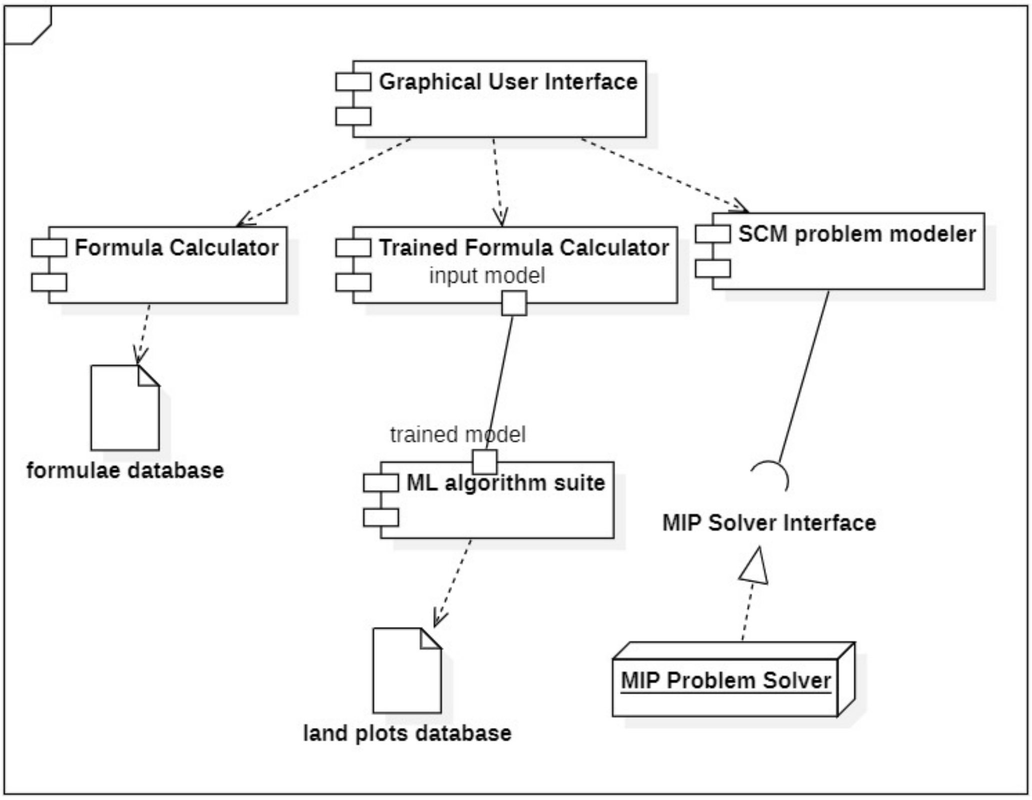

Figure 6 CERESiS high-level architecture (level-0)

The major components of the architecture are as follows:

1. GUI: the Graphical UI of the system that will be implemented as a web-based interface since the entire DSS will be a cloud-native application.

2. The Formula Calculator will perform direct calculations for the cases where a quantity of the form "$y = f(x)$" with known function f is sought.

3. The Trained Formula Calculator will perform calculations for cases where there is no known function $f(x)$ but for which a dataset of inputs/outputs is available as artifact "land plots database" and from which certain outputs are sought for lands with characteristics not exactly fitting those of lands for which equations exist in the "formulae database".

4. The SCM problem modeler is a component that allows (expert) users to enter their model representing an optimization problem, and once initialized with appropriate inputs (values for data and the various parameters of the model), it can submit the model for solution to a MIP (Mixed Integer Programming) solver.

5. The MIP Solver Interface is the mechanism by which the DSS communicates with a 3rd party MIP solver.

6. The MIP Problem Solver is a 3rd party software (examples: GUROBI, CPLEX, XPress, SCIP etc.) that can solve general Mixed Integer Programming Problem models given in a high-level programming language (examples: GAMS, AMPL, ZIMPL etc.) or simply in MPS format.

7. The ML algorithm suite is a collection of Machine Learning algorithms for classification and/or regression (classical supervised ML).

8. The Formulae Database is a collection of formulae for various KPI calculations given inputs directly from the user. It includes formulae for computing KPIs related to process conversion technologies (pyrolysis etc.)

9. The Land Plots Database is a collection of land plots together with their characteristics some of which can be used as inputs and others as target outputs for regression/classification purposes.

3.5.1 Mapping of functional requirements to top-level system components

- FR1 is implemented in components 1, 2, 3 and 9 above.
- FR2 is implemented in components 1, 2, 3 and 9 above.
- FR3 is implemented in components 1, 2, 3 and 9 above.
- FR4 is implemented in components 1, 2, 3 and 9 above.
- FR5 is implemented in components 1, 2, 3, 4, 5, 6, 8 and 9 above.
- FR6 is implemented in components 1, 2, 3, 4, 5, 6, 7, 8 and 9 above.

3.6 Performance evaluation of alternative approaches

In a decision-making environment, model formulation is crucial to represent in the most accurate way the system and the related problem studied. There are various methodological approaches to solve supply chain models depending on several characteristics. To choose an appropriate solution method for the CERESiS biomass supply chains, a brief overview and performance evaluation of alternative approaches is carried out.

Mathematical programming (MP) is widely used to express and solve supply chain problems through the development of optimization models which consist of decision variables, constraints and objective(s) in order to represent at best possible, the real systems. Besides the classical route of Linear Programming, Mixed Integer Linear Programming (MILP) is that branch of MP which offers the framework to represent complex systems including special features such as set-up costs and times, technology choices, economies of scale, as well as logical conditions (Grossmann, van den Heever, & Harjunkoski, 2002; Pochet & Wolsey, 2006). MILP models meet a variety of optimization problems by including continuous, integer or/and binary variables. Integer variables are necessary to describe integral (non-divisible) quantities of products and resources while 0-1 variables are used to express logical conditions/relations and 'yes or no' decisions. As seen in literature, MILP is widely used for the modelling of facility-location problems (Zahraee et al. 2020). The basic advantage of MP is that it attains the optimal allocation of resources throughout the supply chain, providing confidence and robustness to the decisions the stakeholder makes. This method also enables the stakeholder to monitor the bottlenecks in the supply chain in conjunction with virtualization techniques. Some difficulties that can hinder the decision-making process, concern the need for explicit customization per problem type, making the selection of constrained resources and their relationship with the decision variables important not to have any inaccurate representation of the supply chain into the model. MP's methods require some IT skills to develop a mathematical model into an optimization software package. Another key issue mainly concerning MILP (combinatorial) problems is related to the computational requirements needed to be solved (time, threads and memory). For small scale problems finding the optimal solution is not an issue, but as the problem gets larger and more complex it becomes extremely challenging to obtain an exact optimal solution in reasonable time. Nevertheless, the evolution of modelling languages, the continuous efficiency improvement of solver engines and the availability of high-performance computers can deal with this challenge.

The most well-known commercial solvers used to solve MILP problems are CPLEX, GUROBI and XPRESS which can be used through their standalone optimizers as well as under the GAMS programming environment. All solvers integrate a portfolio of methods to solve MILP problems, of which the most prominent are a) branch and bound method and b) cutting plane algorithms which may show slow convergence and be ineffective for large scale problems. Usually, a combination of them, the so-called branch and cut method exploits their advantages and tends to be significantly efficient as a general-purpose algorithm (NTNU, 2014). It is certain that only one optimization solver is unable to solve all types of problems or for all quality measures. The appropriateness of an optimization solver depends upon the nature of the particular problem to be handled (Anand et al. 2017) (Anand et al. 2017)

After realizing in-house, a series of trials in a couple of MILP problems, and attempting an indicative benchmark of the above-mentioned solvers, some basic conclusions emerged:

- different sets of data (scenarios) used in the same model under the same solver can result to quite different solution times
- some models run faster in a solver while others run faster to a different one

- the number of threads dedicated to parallel optimization mode (crucial for MILP problems) can have diverse effects in solution times (number of iterations) and computational effort in all solvers
- the threshold chosen for the optimality gap (accuracy of optimal solution) can benefit a solver over the other
- the continuous improvement of solver engines can lead to significant impacts on solution times through time

In case of non-linear problems among the CPLEX, GUROBI and XPRESS solvers, only XPRESS can be used but with the drawback of being able to find only local optimum. There are also other non-linear solvers such as SCIP, LINDO, CONOPT, BARON which can be used (Kronqvist et al, 2019) but in most cases their performance is questionable. Nn-linearity can be treated indirectly, with appropriate linearization techniques converting non-linear models to linear on the expense of increased complexity due to the addition of more auxiliary constraints.

There are alternate methods to deal with increased complexity cases. If solving to optimality is infeasible due to time constraints, then there is the possibility of implementing a stopping criterion on the commercial solver at the expense of optimality. Recent research has also provided heuristics to find adequate solutions in a very short amount of time, without the guarantee of optimal solutions. As the problem complexity increases, metaheuristics are used to find a near optimum solution by preventing the algorithm from being trapped in local optima. Well known metaheuristics are genetic algorithms, tabu search, ant colony optimization, artificial immune system, differential evolution, simulated annealing and particle swarm optimization. These techniques are often applied in complex combinatorial problems like vehicle routing (Gogna & Tayal, 2013).

Supply chain problems contain different types of uncertainty, so the application of simulation approaches such Monte Carlo technique, are required to manage uncertain parameters and to generate multiple scenarios for monitoring the supply chain response/behavior. Other methods include robust optimization, in which the supply chain is protected against all scenarios of uncertainty at the expense of cost, which often renders the supply chain extremely costly (Tordecilla et al. 2021). Uncertainty can be combined with the optimization methods mentioned above without much customization to the models, however this requires knowledge of parameter behavior from experiments or historical data which is often unavailable or requires a lot of effort.

When investigating the achievement of multiple goals that are usually in conflict with each other, Multi-Criteria Decision Making (MCDM) models are engaged to produce a set of choices to be evaluated by the decision maker. There two basic categories: the Multi-Attribute Decision Making models where the alternative options are distinct and explicitly predetermined, and the Multiple Objective Mathematical Programming (MOMP) models of which solutions derive after appropriate optimization process of decision variables. In the CERESiS project, the objectives that are going to be investigated are the economic performance, the environmental impact and the social dimension.

Biomass supply chains are similar to the general supply chains where resources are constrained (biomass availability, capacity of facilities) and routing decisions need to be made (mobile facility relocation, transportation). Based on this evaluation the most suitable method would be the development of an MILP model determining the optimal location siting and resource planning. This notion is also validated in a review of Ghaderi et al. (2016), where the majority of papers investigated used MILP models for biomass supply chain network design optimization. If mobile processing facilities are added to the model, then the model's complexity increases and therefore other approaches may be required, such as simplification, decomposition or metaheuristics. Future testing will determine whether other methods will be required.

This high-level evaluation of performance & risks of alternative algorithmic approaches will serve as a guide for WP4.

4 DATA INPUTS FOR DSS

In order to set the foundation for the data used in the DSS, an efficient data collection methodology is important for the smooth data exchange between the partners and to have a complete list of parameters at the end of the project. The significance of this effort is to gradually specify the general data provided by the partners in the DMP in D6.3 and for the partners to get a better understanding of what the DSS and the supply chain model requires.

To this end, a three-phase data collection method is utilized to continuously update the data throughout the project. In the first phase of the data collection, an early version of specific parameters required for the DSS was created after a combination of brainstorming, previous projects and models in the literature. Since the model needs to be customized to the needs of the CERESiS project, the input parameters gathered in the first phase needed to be validated and supplemented in the second phase of the data collection. In this phase, the compiled list of data in spreadsheet form was sent to the project partners, in which the specific list of data was preliminarily matched with the project partners based on the feedback and data availability they provided in the DMP in D6.3. The partners were asked to provide feedback on whether the assumptions made in the first phase were logical and in line with the datasets they expected to provide in the future. Additionally, they were asked to supplement the list with necessary data they thought to be necessary for the operation of the DSS.

Considering this list was generated at an early stage of the project, it would be infeasible to have a complete list of data at this point. For this reason, this method will have multiple iterations in the future as the project progresses and more data becomes necessary and available, primarily at time points where deliverables are submitted and the partners have sufficient input to provide.

The first version of the parameter list is listed in Annex II.

5 REFERENCES

Anupriya Gogna & Akash Tayal (2013) Metaheuristics: review and application, Journal of Experimental & Theoretical Artificial Intelligence, 25:4, 503-526

Ghaderi, H., Pishvaee, M. and Moini, A., 2016. Biomass supply chain network design: An optimization-oriented review and analysis. *Industrial Crops and Products*, 94, pp.972-1000.

Grossmann, I. E., van den Heever, S. A., & Harjunkoski, I. 2002. Discrete optimization methods and their role in the integration of planning and scheduling. American Institute of Chemical Engineers symposium series 98, 150-168.

Kronqvist, J., Bernal, D.E., Lundell, A. *et al.* A review and comparison of solvers for convex MINLP. *Optim Eng* **20,** 397–455 (2019).

NTNU, 2014. MILP algorithms: branch-and-bound and branch-and-cut: Norwegian University of Science and Technology.

Pochet, Y., & Wolsey, L. A. (2006). Production Planning by Mixed Integer Programming: Springer.

Rimmi Anand, Divya Aggarwal & Vijay Kumar, 2017 A comparative analysis of optimization solvers, Journal of Statistics and Management Systems, 20:4, 623-635, DOI: 10.1080/09720510.2017.1395182

Tordecilla, R., Juan, A., Montoya-Torres, J., Quintero-Araujo, C. and Panadero, J., 2021. Simulation-optimization methods for designing and assessing resilient supply chain networks under uncertainty scenarios: A review. *Simulation Modelling Practice and Theory*, 106, p.102166.

Zahraee, S., Shiwakoti, N. and Stasinopoulos, P., 2020. Biomass supply chain environmental and socio-economic analysis: 40-Years comprehensive review of methods, decision issues, sustainability challenges, and the way forward. *Biomass and Bioenergy*, 142, p.105777.

6 ANNEX I

The template used to collect the user requirements.

	Please name your story with a short name (it could be related to a specific functionality)	*User stories are short, simple descriptions of a feature told from the perspective of the person who desires the new capability, usually a user or customer of the system. They typically follow a simple template: As a < type of user >, I want < some goal > so that < some reason >.*	*Please break down the previous User story description by providing a clear answer to the who, what and why.*		*Please describe the required input*	*Please describe the required output (optional for non software providers)*	
ID	**User Story Short Name**	**User Story Description**	**Who**	**What**	**Why**	**Inputs**	**Outputs**
1							

7 ANNEX II

Table A1 shows the specific data gathered after the feedback from partners. The parameters are within the high-level datasets of the DMP, in which DT corresponds to the category in the DMP table in D6.3.

Table A1: data inputs for DSS

Data category	Parameters	Data Type Identifier	Responsible Organizations
Spatial data	Longitude/Latitude and contaminated land area per land plot for all land plots of CERESiS Use Cases, candidate locations of facilities Site names Existing biofuel end users location	1.1	USTRAT, UNITUS, REA, UFG
	Distance between nodes - Distances are calculated and imported from ArcGIS using the coordinates of the nodes (fields, end users, candidate locations of facilities)	4.1	NTUA
Technical specifications of processing facilities	Fast Pyrolysis (FP) processing capacity range of centralized, decentralized and mobile facilities FP operational load range of centralized, decentralized and mobile facilities FP energy requirement for centralized, decentralized and mobile facilities operation (energy source breakdown: e.g. ,diesel, natural gas, electricity, self-consumption) FP capacity/cost curve or data points of centralized, decentralized and mobile facilities FP availability factor of centralized, decentralized and mobile facilities during the day/month/year FP construction or installation time of centralized, decentralized and mobile facilities FP product yield of centralized, decentralized and mobile facilities and how it depends on facility scale Lifetime of FP facilities FP reactor operating conditions and to the type of biomass and contaminants	3.8	CNR
	SCWG processing capacity range of centralized, decentralized and mobile facilities SCWG operational load range of centralized, decentralized and mobile facilities SCWG energy requirement for centralized, decentralized and mobile facilities operation (energy source breakdown: e.g. ,diesel, natural gas, electricity, self-consumption) SCWG capacity/cost curve or data points of centralized, decentralized and mobile facilities	4.4	KIT

	SCWG availability factor of centralized, decentralized and mobile facilities during the day/month/year SCWG construction or installation time of centralized, decentralized and mobile facilities SCWG product yield of centralized, decentralized and mobile facilities and how it depends on facility scale Lifetime of SCWG facilities SCWG operating conditions and to the type of biomass and contaminants		
	Relocation time of FP mobile facilities Relocation time of SCWG mobile facilities	4.1	NTUA
	Energetic contribution of pyrolysis gas to the bio-oil production process	3.9	CNR
	FP filtration removal ratio FP contaminant separation capacity	3.10	CERTH
	SCWG filtration removal ratio SCWG contaminant separation capacity	3.7	CERTH
	Upgrading facility capacity range Upgrading facility yield Upgrading facility energy requirement Upgrading facility availability factor Upgrading facility construction time Upgrading facility capacity/cost curve data Upgrading facility operating conditions	3.6	UNISHERBROOKE
Investment costs	FP installation cost of centralized, decentralized and mobile facilities	3,8	CNR
	SCWG installation cost of centralized, decentralized and mobile facilities	4.4	KIT
	Installation cost of decontamination equipment	3.7	CERTH
	Installation cost of upgrading equipment	3.6	UNISHERBROOKE
Yearly fixed costs	FP yearly operational costs of centralized, decentralized and mobile facilities (labor, overhead, maintenance, insurance, and taxes, etc.)	3.8	CNR
	SCWG yearly operational costs of centralized, decentralized and mobile facilities (labor, overhead, maintenance, insurance, and taxes, etc.)	4.4	KIT

	Annual fixed operating cost of the upgrading facility (this parameter includes labor, overhead, maintenance, insurance, and taxes)	3.6	UNISHERBROOKE
	FP decontamination equipment yearly operation costs	3.10	CERTH
	SCWG decontamination equipment yearly operation costs	3.7	CERTH
	Storage unitary cost Relocation cost of mobile facilities Biomass purchasing cost (in the case of purchasing supplementary non-contaminated biomass) Truck rental/purchasing price trailer trucks Truck rental/purchasing price tanker trucks	4.1	NTUA
Operational costs	Unit conversion cost of FP facilities	3.8	CNR
	Unit conversion cost of SCWG facilities	4.4	KIT
	Unit separation cost of decontamination for FP	3.10	CERTH
	Unit separation cost of decontamination for SCWG	3.7	CERTH
		2.2	USTRAT, UNITUS, REA, UFG
	Seeding unit cost per unit of land per biomass species Cultivation unit cost per unit of land per biomass species Harvesting unit cost per unit of land per biomass species	2.3	USTRAT, UNITUS, REA, UFG
	Biomass drying unit cost (in the case of using ovens)	2.4	USTRAT, UNITUS, REA, UFG
	Upgrading facilities unit conversion cost	3.6	UNISHERBROOKE
Transportation characteristics & costs	Trailer truck dry ton capacity volumetric and per weight capacity Tanker truck volumetric and per weight capacity Trailer truck loading and unloading cost Tanker truck loading and unloading cost Distance-dependent transportation cost of bulk Distance-dependent transportation cost of liquid Travel time dependent transportation cost of bulk Travel time dependent transportation cost of liquid Trailer truck average travel speed Tanker truck average travel speed	4.1	NTUA

Revenues	Selling price of final product (biofuel) Contaminated biomass unitary removal fee	4.1	NTUA
	Selling price of intermediate products	3.8 4.1	CNR, NTUA
Biomass characteristics	Moisture content of biomass Biomass energy content Biomass ash content Biomass loss rate (decomposition rate, maximum storage time, etc.) Contamination uptake per time period per biomass type (trial data series) Biomass feedstock growth function (data points for piecewise linearization) Seeding and harvest months of biomass Processing times of seeding & harvest	2.4	USTRAT, UNITUS, REA, UFG
	Biomass composition and properties, chemical kinetic mechanisms	3.17	NTUA
Soil characteristics	Soil composition & analysis data (trial data series) Climatic/weather conditions Contaminants present in soil Contaminant concentration in soil	2.1	USTRAT, UNITUS, REA, UFG
Demand	Product demand of products of SCWG and FP Product demand of intermediate products of SCWG and FP	4.1	NTUA
Environmental impact (CO2, GHG emissions, etc.)	FP installation energy and environmental impact of centralized, decentralized and mobile facilities SCWG installation energy and environmental impact of centralized, decentralized and mobile facilities FP unit energy and environmental impact of centralized, decentralized and mobile facilities SCWG unit energy and environmental impact of centralized, decentralized and mobile facilities Separation unit energy and environmental impact	4.5	NTUA
	Transportation environmental impact	4.1	NTUA
	Environmental impact of seeding, cultivation and harvesting	2.4	USTRAT, UNITUS, REA, UFG
Social impact	FP social impact of centralized, decentralized and mobile facilities operation	4.5	NTUA

(Social acceptance, job creation, etc.)	SCWG social impact of centralized, decentralized and mobile facilities operation	4.5	NTUA