EO4EU
Earth Observation Data & Service Accessibility

D5.1 - Pilot Implementation methodology and release of evaluation guidelines (a)

Work Package	WP 5, EO Data Uptake Demonstration
Lead Author (Org)	Stefano Natali (SISTEMA)
Contributing Author(s) (Org)	Alexandra Bojor, Clemens Rend (SISTEMA) Maria Luisa Quarta, Marco Folegani (MEEO) Kaklis Dimitris (DANAOS) Roberta Paduano, Monia Santini, Tammam Suliman (CMCC) Mikhail Sofiev (FMI) Fabrizio Ferrucci, Massimo Cristaldi (IES)
Due Date	31.05.2023
Date	31.05.2023
Version	V1.0

Dissemination Level

X	PU: Public
	PP: Restricted to other programme participants (including the Commission)
	RE: Restricted to a group specified by the consortium (including the Commission)
	CO: Confidential, only for members of the consortium (including the Commission)

Versioning and contribution history

Version	Date	Author	Notes
0.1	14.04.2023	Stefano Natali, Clemens Rendl (SISTEMA)	TOC and V0.1
1.0	31.05.2023	Alexandra Bojor, Clemens Rend (SISTEMA) Maria Luisa Quarta, Marco Folegani (MEEO) Kaklis Dimitris (DANAOS) Roberta Paduano, Monia Santini, Tammam Suliman (CMCC) Mikhail Sofiev (FMI) Fabrizio Ferrucci, Massimo Cristaldi (IES)	First release of the document finalized

Terminology

Terminology/Acronym	Description
AI	Artificial Intelligence
ADAM	Advanced geospatial DAta Management
ADS	Copernicus Atmosphere Data Store
CAM	Copernicus Atmosphere Monitoring Service
CC	Climate Change
CDAS	Copernicus Space Component Data Access
CDS	Copernicus Climate Data Store
CMCC	Euro-Mediterranean Center on Climate Change
DDS	Data Delivery System
DIAS	Data and Information Access Services
EC	European Commission
ECMWF	European Centre for Medium-Range Weather Forecasts
EO	Earth Observation
EO4EU	Horizon Europe project called: AI-augmented ecosystem for Earth Observation data accessibility with Extended reality User Interfaces for Service and data exploitation
EU	European Union
EUMETSAT	European Organisation for the Exploitation of Meteorological Satellites
FAO	Food and Agriculture Organization
FEMs	Forest Ecosystem Models

Terminology/Acronym	Description
FMI	Finnish Meteorological Institute
FTP	File Transfer Protocol
GA	Grant Agreement to the project
GeoTIFF	Geographic Tagged Image File Format
GIS	Geographic Information System
GPU	Graphics processing unit
GRIB	GRIdded Binary or General Regularly-distributed Information in Binary form
HPC	High-performance computing
HTTP	Hypertext Transfer Protocol
ISTAT	Italian National Institute of Statistics
ISRIC	World Soil Information
IVI	Fraunhofer Institute for Transportation and Infrastructure Systems
KEMEA	Kentro Meleton Asfaleias
KPI	Key Performance Indicator
LMCS	Copernicus Land Monitoring Core Service
LU	University of Latvia
ML	Machine learning
NASA	National Aeronautics and Space Administration
NDVI	Normalized difference vegetation index
NetCDF	Network Common Data Form
NKUA	National and Kapodistrian University of Athens
NOAA	National Oceanic and Atmospheric Administration
PASYFO	Personal Allergy Symptom Forecasting System
PU	Public
UC	Use Case
UI	User Interface
URL	Uniform Resource Locator
US	United States
VM	Virtual Machines
VR	Virtual Reality
VU	Vilnius University
WP	Work Package

Table of Contents

List of Figures

List of Tables

Executive Summary

A vast amount of Earth Observation data is produced daily and made available through online services and repositories. Contemporary and historical data can be retrieved and used to power existing applications, foster innovation, and improve EU citizens' lives. However, the audience for this information is relatively small, resulting in vast amounts of valuable data remaining unexplored.

The EO4EU project aims to provide innovative tools, methodologies, and approaches that can assist a broad range of users, from domain experts and professionals to ordinary citizens, in benefiting from accessing EO data. It aims to deliver dynamic data mapping and labelling AI-augmented modules that add fairness to the data and introduce an ecosystem for the holistic management of EO data. EO4EU envisions bridging the gap between domain experts and end-users while bringing technological advances to the forefront to address the market's demand for wider usage of EO data.

The project will support the wider exploitation of EO data by delivering: (i) Machine Learning (ML) methodologies for Semantic Annotation of existing and growing data sources, (ii) semantically enhanced knowledge graphs that will enable structuring of content around diverse topic areas and building step by step journeys from different sources into a unified approach, (iii) data fusion techniques to extend the scalability of existing distributed systems, (iv) Augmented and Virtual Reality for interactive user experience, and (v) advanced data analytics visualizations for improved learning and evidence-based interpretations of environmental observations. Its operational and technical capacity will be demonstrated within seven distinct pilots that cover different thematic areas, such as personalized health care, sea route planning, ocean monitoring, food security, food ecosystems, soil erosion, environmental pest, and crisis management. These thematic areas will engage a wide spectrum of involved stakeholders, from EO providers, policy makers and actors, researchers and academics to citizen scientists and the general public to join efforts and provide their multidisciplinary expertise to support the Commission's strategic goals towards further exploitation of EO data.

In this document, each of the seven use cases is analysed in detail to identify their data management needs, software component needs, and computational needs. Additionally, a summary of the deployment plan for each use case is presented, to consolidate it at a later stage against the data and processing infrastructure provider for technical alignment. These two chapters, namely the deployment plan and technical alignment, will be briefly described in this document, with more details to be provided in version B at a later stage.

1 Introduction

1.1.1 Purpose of the document

This document, the Pilot Implementation methodology and release of evaluation guidelines (a), provides a detailed analysis of the management needs, software component needs, and computational needs for each of the seven use cases. This analysis aims to support the consolidation against the data and processing infrastructure provider for technical alignment.

1.1.2 Mapping EO4EU outputs

EO4EU GA Component Title	EO4EU GA Component Outline	Respective Document Chapter(s)	Justification
T5.1	This task deals with the UCs definition and technical implementation guidelines. In this task all the proposed UCs will be analysed in detail to identify in terms of data management needs, SW components needs and computational infrastructure needs. Each UC will be required to generate a deployment plan that will be consolidated against the data and processing infrastructure providers for the technical alignment. The data management needs must include input data needs such as spatial and temporal domains, data sources, data format, data transfer rate, disk size, intermediate and/or temporary data products and output data description. The SW needs include AI/ML components description in terms of scope, configuration and pipeline. The Computational needs must include a description of virtual resources, scalability requirements, resource occupation for development purposes and for the operational phase of the service. The technical alignment will consist of the critical analysis of the UCs deployment plans to support the coordinated exploitation and occupation of the computational infrastructure.	Section 2: Chapter Introduction Section 3: Use Case I -Health Section 4: Use Case II – Ocean Section 5: Use Case III – Crop Section 6: Use Case IV – Forest Section 7: Use Case V – Soil Section 8: Use Case VI – Food Section 9: Use Case VII – Fires Section 10: Deployment Plan Section 11: Technical alignments	The document is based on a complete survey performed on the use cases to collect their needs. On this basis, data management needs, SW components needs and computational infrastructure needs were identified for each of the use cases. Also, a description of the deployment plan for each of the seven use cases and a detailed analysis of the technical alignment will be provided here.
D5.1	A public report on the (7) demonstrators of EO4EU. Involved actors and organisations, tested modules, schedule, participants, scenarios, lessons learned and results will be part of this document. Version a.		

1.1.3 Deliverable overview and report structure

The document is structured as follows:

- Section 2 provides a summary of the different chapters, such as use cases workflow, data management needs, software components need, computational infrastructure needs and deployment plan
- Section 3 to Section 9 provide the detailed description per use case of the workflow, data management needs, software components need, computational infrastructure needs and deployment plan (it will be provided in version B of the document)
- Section 10 contains the technical alignment which will be provided in the version B of the document
- Section 11 summarizes the document's conclusions

2 Document Structure

This section gives an overview of the methodology used to describe the project use cases, as provided in chapters 3 to 9.

The base of the document are the collected use case requirements accessible in Annex A, collected during task T2.1 – Multidisciplinary landscape analysis & Best of Breed Technologies, and delivered as deliverable D2.1 – Research and Innovation Landscape analysis report.

The requirements were elaborated based on a questionnaire, which was drafted by EBOS and SISTEMA. It contains four sections, which cover input data specification needs, data preparation needs, data processing needs and means of presenting and visualizing results. To fill the individual questionnaires, it was first provided online to the use case leaders (seven teams involved), enabling them to familiarise themselves with the presented sections. In a second round, one-to-one meetings/ telcos were held by SISTEMA with the corresponding use case conductors, to complete the questionnaires and eventually get a complete overview of the workflow conducted within each use case.

Based on the work done in T2.1, additional interaction and exchange were done to complete and finalize this deliverable. The use case leaders were asked to refine all information previously provided with more updated information on data needs, processing steps and ICT resources needs. Moreover, given the final version of the document to be released at month 18, a preliminary definition of the deployment plan has been provided by each service provider.

2.1 Introduction

For each use case, this section aims at introducing the involved partners and the scope of the use case.

2.2 Use case workflow

The scope of this chapter is to introduce and describe the workflow of each use case using tables. The chapter contains four dedicated sub-chapters, covering the input data needs, pre-processing and processing steps and the generated results.

2.2.1 Input data

The table in the chapter "Input data" provides an overview of the required input data for the dedicated use case and includes specific information for each dataset. Table 1 shows the structure of the input data table. It includes the following columns:

- **Input data (Dx)**

  A numerical indication of each dataset, e.g., dataset one (D1), dataset two (D2), etc.
- **Data**

  The name of the individual dataset
- **Geographic extent**

  The geographic extent of the individual dataset, which corresponds to the area of interest (AOI) of the use case

- **Spatial resolution**

  The spatial resolution of the dedicated input dataset

- **Time range**

  The temporal resolution of the individual dataset

- **Data source**

  The source of the individual dataset

- **Data format**

  The format of each dataset

- **Disk size**

  An estimation of the disk space needed to store the dedicated dataset

Table 1. Generic input data table

Input data (Dx)	**Data**	**Geographic extent**	**Spatial resolution**	**Time range**	**Data source**	**Data format**	**Disk size**
Dx							

2.2.2 Pre-processing

The chapter "Pre-processing" provides an overview of the required pre-processing steps, if applicable. As seen in

Table 2, it includes the following elements:

- **Pre-processing (PPx)**

  Indicating the pre-processing step through numbering, e.g., pre-processing step one (PP1)

- **Input Data**

  The required input data to perform the dedicated pre-processing step, derived from Table 1

- **Step description**

  A description of the dedicated pre-processing step

- **Outcome (preservation y/n)**

  An indication of the outcome of the pre-processing step should be preserved (y) or can be deleted (n)

- **Processing tool (open source or commercial)**

  The software used for the dedicated pre-processing step, including an indication if the software is open-source (o) or commercial (c)

- **Processing interface**

  The mean used to trigger the execution of the processing module: Command Line Interface (CLI), Application Programming Interface (API), manual execution, …

- **Facility**

Whether the pre-processing is done locally or remotely

- **Infrastructure requirements / resource needs (development)**

  An indication of the infrastructure requirements during the development phase

- **Infrastructure requirements / resource needs (operational)**

  An indication of the infrastructure requirements during the operational phase

Table 2. Generic pre-processing table

Pre-processing (PPx)	**Input Data**	**Step description**	**Outcome** (preservation y/n)	**Processing tool** (open source or commercial)	**Processing interface**	**Facility**	**Infrastructure requirements / resource needs** (development)	**Infrastructure requirements / resource needs** (operational)
PPx	Dx							

2.2.3 Processing

The chapter "Processing" provides an overview of the required processing steps. Table 3 displays a generic table including the following columns:

- **Processing step (Sx)**

  Indicating the processing step through numbering, e.g., processing step one (S1)

- **Input Data**

  The required input data to perform the dedicated processing step, derived from Table 1

- **Step description**

  A description of the dedicated processing step

- **Outcome (preservation y/n)**

  An indication of whether the outcome of the processing step should be preserved (y) or can be deleted (n)

- **Processing tool (open source or commercial)**

  The software used for the dedicated processing step, including an indication if the software is open-source (o) or commercial (c)

- **Processing interface**

  The mechanism through which the processing step is executed: Command Line Interface (CLI), Application Programming Interface (API), Graphic User Interface (GUI), and manual triggering.

- **Facility**

  Whether the processing is done locally or remotely

- **Infrastructure requirements / resource needs (development)**

An indication of the infrastructure requirements during the development phase

- **Infrastructure requirements / resource needs (operational)**

  An indication of the infrastructure requirements during the operational phase

Table 3. Generic processing table

Processing step (Sx)	Input Data	Step description	Outcome (preservation y/n)	Processing tool (open source or commercial)	Processing interface	Facility	Infrastructure requirements / resource needs (development)	Infrastructure requirements / resource needs (operational)
Sx								

2.2.4 Results

The chapter "Results" provides an overview of the outcome of the processing to be delivered. Table 4 displays the content of the generic results table, including the following columns:

- **Result (Rx)**

  Indicating the result through numbering, e.g., result one (R1)

- **Result description**

  A description of the dedicated result

- **Result delivery**

  Means of delivering the processing outcome/result, e.g., GUI

- **Infrastructure requirements / resource needs (development)**

  An indication of the infrastructure requirements during the development phase

- **Infrastructure requirements / resource needs (operational)**

  An indication of the infrastructure requirements during the operational phase

Table 4. Generic results table

Result (Rx)	Result description	Result delivery	Infrastructure requirements / resource needs (development)	Infrastructure requirements / resource needs (operational)
Rx				

2.3 Data management needs

This chapter covers and describes for all individual use cases the characteristics and resource needs for:

- Input data
  - Intermediate data generated during
    - Pre-processing (if applicable)
    - Processing, and
  - Output data

The content of the chapter corresponds to the individual tables in the chapter "Workflow" of each use case.

2.4 Software component needs

This chapter covers and describes for all individual use cases the software components utilized to perform data pre-processing and data processing. The section shall give an overview of the software used and the steps involved to produce the results of each use case. If applicable, the use of AI/ML is highlighted. If currently not used, the potential use of AI/ML is elaborated. The content of the chapter corresponds to the individual tables in the chapter "Workflow" of each use case.

The following elements are considered:

- AI/ML components description
    - Scope
    - Configuration
    - Pipeline

2.5 Computational infrastructure needs

This chapter includes a description of the computational infrastructure needs for the individual use cases. The content of the chapter corresponds to the individual tables in the chapter "Workflow" of each use case and summarizes the following infrastructure needs regarding pre-processing, processing and results. The following elements are included:

- Description of virtual resources
- Scalability requirements
    - Fixed vs dynamic scalability
    - Horizontal vs. vertical scalability
- Resource occupation for development purposes
- Resource occupation for the operational phase of the service

2.6 Deployment plan

The deployment plan will be provided in more detail within each of the seven use cases at a later stage, in the version B of the document. It will be consolidated against the data and processing infrastructure providers for the technical alignment and will contain the following points:

- Resources allocation (Storage, computation etc.);
- Development phase (information on activities vs. dates and allocation plan);
- Operational phase (information on activities vs. dates and allocation plan).

3 Use Case I – Health

3.1 Introduction

UC1, led by the Finnish Meteorological Institute (FMI), with the participation of the Vilnius University (VU) and the University of Latvia (LU), focuses on further expanding the capacity of the PASYFO model. The Personal Allergy Symptom Forecasting System (PASYFO) model is an operational symptom forecasting model that includes a mobile application, which provides personal pollen allergy symptoms forecasting. Currently available for Latvia and Lithuania, the scope of UC1 is to expand the coverage of the PASYFO model to the globe, making it universally available. This will significantly increase the geography features of users and thus contribute to raising public awareness and help prevent chronic diseases (such as respiratory allergies and consequent asthma).

3.2 Workflow

The scope of this chapter is to introduce and describe the workflow of use case I using tables. The chapter contains four dedicated sub-chapters, covering the input data needs, pre-processing and processing steps and the generated results.

3.2.1 Input data

Table 5: Use case 1 input data specifications

Input data (Dx)	Data	Geographic extent	Spatial resolution	Time range	Data source	Data format	Disk size
D1	ECMWF high-resolution forecast dew point	Europe	0.1 degrees	0+3 days	ECMWF	nc	~0.1GB
D2	ECMWF high-resolution forecast air temperature	Europe	0.1 degrees	0+3 days	ECMWF	nc	~0.1GB
D3	CAMS forecast O3	Europe	0.1 degrees	0+3 days	Copernicus (CAMS)	nc	~0.1GB
D4	CAMS forecast CO	Europe	0.1 degrees	0+3 days	Copernicus (CAMS)	nc	~0.1GB
D5	CAMS forecast SO2	Europe	0.1 degrees	0+3 days	Copernicus (CAMS)	nc	~0.1GB
D6	CAMS forecast NO2	Europe	0.1 degrees	0+3 days	Copernicus (CAMS)	nc	~0.1GB
D7	CAMS forecast PM10	Europe	0.1 degrees	0+3 days	Copernicus (CAMS)	nc	~0.1GB
D8	CAMS forecast PM2.5	Europe	0.1 degrees	0+3 days	Copernicus (CAMS)	nc	~0.1GB
D9	CAMS forecast birch	Europe	0.1 degrees	0+3 days	Copernicus (CAMS)	nc	~0.1GB
D10	CAMS forecast olive	Europe	0.1 degrees	0+3 days	Copernicus (CAMS)	nc	~0.1GB
D11	CAMS forecast grass	Europe	0.1 degrees	0+3 days	Copernicus (CAMS)	nc	~0.1GB
D12	CAMS forecast ragweed	Europe	0.1 degrees	0+3 days	Copernicus (CAMS)	nc	~0.1GB
D13	CAMS forecast alder	Europe	0.1 degrees	0+3 days	Copernicus (CAMS)	nc	~0.1GB
D14	CAMS forecast mugworth	Europe	0.1 degrees	0+3 days	Copernicus (CAMS)	nc	~0.1GB

3.2.2 Pre-processing

Table 6: Use case 1 pre-processing steps specifications

Pre-processing (PPx)	Input Data	Step description	Outcome (preservation y/n)	Processing tool (open source or commercial)	Processing interface	Facility	Infrastructure requirements / resource needs (development)	Infrastructure requirements / resource needs (operational)
PP1	D1	conversion of dew point to relative humidity	data array (y)	Open source	PythonCLI script	local	N/A	N/A

3.2.3 Processing

Table 7: Use case 1 processing steps specifications

Processing step (Sx)	Input Data	Step description	Outcome (preservation y/n)	Processing tool (open source or commercial)	Processing interface	Facility	Infrastructure requirements / resource needs (development)	Infrastructure requirements / resource needs (operational)
S1	D1, D2	Relative humidity calculation	Data array (y)	Python (o)	CLI	local	1 CPU, 1GB RAM	1 CPU, 1GB RAM estimation of the resource needs, once service is potentially moved to the EO4EU system
S2	D9, D10, D11, D12, D13, D14	Pollen load calculation	birch, olive, grass, ragweed, alder, mugworth maps (y)	Python (o)	CLI	local	1 CPU, 1GB RAM	1 CPU, 1GB RAM
S3	D3, D4, D5, D6, D7, D8	Pollutant load calculation	O3, CO, SO2, NO2, PM10, PM2.5 maps	Python (o)	CLI	local	1 CPU, 1GB RAM	1 CPU, 1GB RAM

Processing step (Sx)	Input Data	Step description	Outcome (preservation y/n)	Processing tool (open source or commercial)	Processing interface	Facility	Infrastructure requirements / resource needs (development)	Infrastructure requirements / resource needs (operational)
			(y)					
S4	D2	Humidex	Humidex map (y)	Python (o)	CLI	local	1 CPU, 1GB RAM	1 CPU, 1GB RAM
S5	D1, D2, D3, D4, D5, D6, D7, D8, D9, D10, D11, D12, D13, D14	Air quality index for persons sensitive to pollen allergens	Index map (y)	Python (o)	CLI	local	1 CPU, 1GB RAM	1 CPU, 1GB RAM

3.2.4 Results

Table 8: Use case 1 results specifications

Result (Rx)	Result description	Result delivery	Infrastructure requirements / resource needs (development)	Infrastructure requirements / resource needs (operational)
R1	Birch, olive, grass, ragweed, alder, mugworth maps	online	1 CPU, 1GB RAM, server	1 CPU, 1GB RAM
R2	O3, CO, SO2, NO2, PM10, PM2.5 maps	online	1 CPU, 1GB RAM, server	1 CPU, 1GB RAM
R3	Humidex map	online	1 CPU, 1GB RAM, server	1 CPU, 1GB RAM

3.3 Data management needs

The system development consists of two steps, which are the following:

1. The first step is an update of the current release of PASYFO core modules, implementation of the new algorithm, improvements in parallelization, etc.
2. The second step is a replacement of the input data delivery from the current ADS/ECMWF resources to the developments of the EO4EU.

The PASYFO core update is ongoing, with the new algorithm practically developed and tested in a different problem of fire intensity forecasting. Its completion is expected within 2023. It will be accompanied by the expansion of the system coverage towards the European countries available for PHFD and MASK-Air users – and further expanding it by the inclusion of new regions, owing to the work of two subcontractors.

The PASYFO system receives several kinds of input data. The main source, provided via the technological solutions of EO4EU, is CAMS AQ and weather forecasts for Europe, with the forecasting horizon reaching up to 96 hours.

The other data sources are the SILAM high-resolution AQ forecasts for Northern Europe and Central Europe, as well as the last-day updates of symptom self-reporting files from the PHFD and MASK-Air online apps. All forecasts are received as NetCDF files, whereas the retrospective symptom reports a scrambled text file.

The main volume of input data is the AQ and meteorological forecasts, with a daily exchange volume of about 100GB. The exchange is time-critical since the output forecasts need to be produced within a few hours after midnight, and the work cannot be started before the new forecast arrival. A solution exploited this-far is to utilize the previous-day forecast, thus substantially releasing the limitation. However, the simulations still have to be started after midnight to allow the users to fill in their last symptoms.

The system output includes personalized allergy symptom forecasts for every app user with a sufficient amount of data, and an overall allergy risk map, which presents the aggregated individual forecasts. The updated system will replace the current prototype in the FMI operational computers in early 2024. After this, operational activities will continue with the above-described resource allocation.

3.4 Software component needs

The system is built on Python with the ML component consisting of a non-linear data transformer and multilinear regression steps. Classical AI components are not used since they require far too extensive input, in particular, long symptom time series in retrospect to be trained. However, such time series are never at hand. Firstly, the vast majority of the app users enter the data for just a few days during the main flowering season. Secondly, the season of many aeroallergens lasts just a couple of weeks, and the corresponding allergy symptom forecasting must pick-up practically with a few days of history at the beginning of the season.

The system is run independently for each app user, with the same list of input data. Pre-processing and non-linear transformations aim at reducing the non-linearity in the predictor-prediction relations, and also to filter out the variables, which do not correlate with the retrospective symptoms, i.e., should not be considered in the main regression.

The processing starts from the extraction of the forecasting time series for each app user. The retrospective AQ and pollen forecasts, as well as past-time symptom reports, are kept in the protected PASYFO system storage in the form of scrambled files. Therefore, only the last-day update of these data needs to be communicated with the PHFD/MASK-Air central servers and CAMS/SILAM data portals. Upon extraction of the time series, the system processes the symptom and AQ data for each user and stores the output predictions as scrambled interface files, to be sent back to the PHFD/MASK-Air servers.

3.5 Computational infrastructure needs

Computation-wise, the PASYFO system is moderate as well and all intermediate data are produced within the system and not stored on disks. Therefore, the largest consumed time is from the input pre-processor module, which deals with large data volumes from CAMS and SILAM and is poorly parallelizable due to numerous disk operations. Experience shows that a powerful workstation or one node of a supercomputer is sufficient to perform all necessary computations within the needed time slot (about 30 minutes). Larger resource allocation may be required only if the number of the PASYFO users becomes 10x of the current level.

The challenge is only the large amount of input information, which needs to be processed to obtain the time series for every user. After that, individual-case processing is a computationally cheap task performed with a single core within a few seconds.

The initial step of the time series extraction scales quite poorly since it requires sequential work with the data. In PASYFO, it is arranged time-step-by-time-step and variable-by-variable extracting data for all needed locations at once. The timing of this step is acceptable and depends on the amount of input information rather than on the number of users. Upon completion of the data extraction, the user cases are computed one by one, fully independently and thus perfectly parallelizable and scaling practically infinitely. The final step of formation of the output file is again sequential with very limited options for parallelization. However, the amount of data is very small (megabytes), hence, even a single-core processing is sufficient.

3.6 Deployment plan

To be provided at a later stage in the document version B.

4 Use Case II – Ocean

4.1 Introduction

UC2, led by DANAOS Shipping, with contribution from the National and Kapodistrian University of Athens (NKUA), aims to incorporate cutting-edge EO technologies and data to optimize ship routing in terms of fuel consumption, safety and arrival time precision. The scope of the use case is to integrate the EO4EU capability of handling extreme volumes of data by fusing the meteorological data collected from EO data sources with on-board vessel sensory information to perform route optimization during the voyage of the ship.

4.2 Workflow

The scope of this chapter is to introduce and describe the workflow of use case II using tables. The chapter contains four dedicated sub-chapters, covering the input data needs, pre-processing and processing steps and the generated results.

4.2.1 Input data

Table 9: Use case 2 input data specifications

Input data (Dx)	Data	Geographic extent	Spatial resolution	Time range	Data source	Data format	Disk size
D1	draft	Mediterranean, Atlantic Ocean, Pacific	0.1 degrees	2019- today	Danaos	time-series	N/A
D2	wind direction	Mediterranean, Atlantic Ocean, Pacific	0.1 degrees	2019- today	Danaos	time-series	N/A
D3	speed through water	Mediterranean, Atlantic Ocean, Pacific	0.1 degrees	2019- today	Danaos	time-series	N/A
D4	fuel oil consumption	Mediterranean, Atlantic Ocean, Pacific	0.1 degrees	2019- today	Danaos	time-series	N/A
D5	trim	Mediterranean, Atlantic Ocean, Pacific	0.1 degrees	2019- today	Danaos	time-series	N/A
D6	steaming hours	Mediterranean, Atlantic Ocean, Pacific	0.1 degrees	2019- today	Danaos	time-series	N/A
D7	swell wave height	Mediterranean, Atlantic Ocean, Pacific	0.1 degrees	2019- today	Danaos	time-series	N/A
D8	swell wave direction	Mediterranean, Atlantic Ocean, Pacific	0.1 degrees	2019- today	Danaos	time-series	N/A
D9	combined wind/wave height	Mediterranean, Atlantic Ocean, Pacific	0.1 degrees	2019- today	Danaos	time-series	N/A
D10	combined wind/wave direction	Mediterranean, Atlantic Ocean, Pacific	0.1 degrees	2019- today	Danaos	time-series	N/A
D11	wave height	Mediterranean, Atlantic Ocean, Pacific	0.1 degrees	2019- today	Danaos	time-series	N/A
D12	wave direction	Mediterranean, Atlantic Ocean, Pacific	0.1 degrees	2019- today	Danaos	time-series	N/A
D13	Lat	Mediterranean, Atlantic Ocean, Pacific	0.1 degrees	2019- today	Danaos	time-series	N/A
D14	Lon	Mediterranean, Atlantic Ocean, Pacific	0.1 degrees	2019- today	Danaos	time-series	N/A
D15	wind speed	Mediterranean, Atlantic Ocean, Pacific	0.1 degrees	2019- today	Danaos	time-series	N/A
D16	current speed	Mediterranean, Atlantic Ocean, Pacific	0.1 degrees	2019- today	Danaos	time-series	N/A
D17	current direction	Mediterranean, Atlantic Ocean, Pacific	0.1 degrees	2019- today	Danaos	time-series	N/A
D18	speed overground	Mediterranean, Atlantic Ocean, Pacific	0.1 degrees	2019- today	Danaos	time-series	N/A

Input data (Dx)	Data	Geographic extent	Spatial resolution	Time range	Data source	Data format	Disk size
D19	rounds per minute of the main engine	Mediterranean, Atlantic Ocean, Pacific	0.1 degrees	2019- today	Danaos	time-series	N/A
D20	power	Mediterranean, Atlantic Ocean, Pacific	0.1 degrees	2019- today	Danaos	time-series	N/A
D21	weather data NOAA	Globe	0.1 degrees	2018- today	NOAADanaos	Time-series (3h intervals)	N/A
D22	weather data NOAA	Globe	0.1 degrees	2018- today	NOAADanaos	Time-series (3h intervals)	N/A

4.2.2 Pre-processing

Table 10: Use case 2 pre-processing steps specifications

Pre-processing (PPx)	Input Data	Step description	Outcome (preservation y/n)	Processing tool (open source or commercial)	Processing interface	Facility	Infrastructure requirements / resource needs (development)	Infrastructure requirements / resource needs (operational)
PP1	operational	Filtering noise	"Cleansed" version of data	Statistical analysis tools (DBSCAN, KMEANS)	API/ DANAOS ANALYTICS SUITE	Local	N/A	N/A

4.2.3 Processing

Table 11: Use case 2 processing steps specifications

Processing step (Sx)	Input Data	Step description	Outcome (preservation y/n)	Processing tool (open source or commercial)	Processing interface	Facility	Infrastructure requirements / resource needs (development)	Infrastructure requirements / resource needs (operational)
S1	all	synchronize with the latest data coming from the vessel	y	Oracle servers, airflow, spark	CLI	N/A	N/A	CPU
S2	all	map sensor data with weather data	y	SQL servers, python APIs	CLI	N/A	N/A	CPU
S3		apply custom or state-of-the-art algorithms for cleaning the bulk of	y	python, sklearn, spark	CLI	N/A	N/A	CPU

Processing step (Sx)	Input Data	Step description	Outcome (preservation y/n)	Processing tool (open source or commercial)	Processing interface	Facility	Infrastructure requirements / resource needs (development)	Infrastructure requirements / resource needs (operational)
		data.						
S4		train a set of multidisciplinary models depending on the exact outcome (safety, cost, charter party agreement, ETA,)	y	python, sklearn, keras, mlflow,	CLI/GUI	N/A	N/A	CPU/GPU
S5		validate models on historical data	y	python, sklearn, keras, mlflow,	CLI	N/A	N/A	CPU
S6		trigger any refinements on the models based on the validation procedure	y	python, sklearn, keras, mlflow,	CLI	N/A	N/A	CPU
S7		incorporate models into the routing algorithm. Calibrate.	y	API, python - c# comm. OR wrap the models as libraries/ store them in the model registry	CLI/API	N/A	N/A	CPU
S8		run a set of experiments on historical data validating the WR routing scheme in terms of	y		GUI/API	N/A	N/A	CPU

4.2.4 Results

Table 12: Use case 2 results specifications

Result (Rx)	**Result description**	**Result delivery**	**Infrastructure requirements / resource needs** (development)	**Infrastructure requirements / resource needs** (operational)
R1	alternative route suggestion	API	N/A	N/A
R2	KPIs on alternative route (Overall Cost, time,	API	N/A	N/A

	etc) Computational – Complexity			

4.3 Data management needs

To realise the use case of OCEAN MONITORING and consequently, that of Voyage Planning we will need operational data coming from the vessels' data acquisition system (sensor instalments) or AIS (Automatic Identification System) data. We also need a variety of weather features, with high granularity (at least ~ 0.1 degrees spatial resolution) that outline the sea state. These features will be exploited accordingly to employ a novel weather routing module leveraging the exploration capabilities of path-finding traditional algorithms like A*, Dijkstra and the exploitation of AI incorporating Reinforcement Learning.

4.4 Software component needs

As previously mentioned, we will employ a versatile ecosystem of evaluating different routes that incorporates into the workflow traditional path-finding algorithms from pertinent literature. This streamlined procedure will be facilitated by SOTA-supporting mechanisms and framework regarding scheduling (e.g. Airflow) parallel processing (e.g. Spark) and visualisation through a web-based integrated environment (e.g. Flask).

The platform will generate a variety of blueprints, by utilising a rule-based language (YAML), to vastly automate and standardise the procedures concerning the Voyage Planning module. These configuration files will be consumed in turn by the platform to generate the appropriate workflows that need to be adopted in each case (trim/speed optimization, weather avoidance, ETA, emission control, etc).

An Extract- Load-Transform (ETL) approach will be realised through an Agile development procedure. The streamlined workflow employed will support the full spectrum of procedures concerning a SOTA ETL pipeline from data extraction-filtering and analysis to model deployment as well as KPI identification and validation through a versatile Decision Support System (DSS).

4.5 Computational infrastructure needs

Computation resources to realise the Ocean Monitoring use case have already been defined and can be categorised as local (on-premises) and ON-EDGE (vessel).

- Local: Bare metal provision (Servers on HQ)
- ON-EDGE: Virtual Machines & Servers on board to synchronise and analyse and map data. This is of high importance if we eventually resort to real-time solutions.

The Voyage Planning module is developed and will be refined in alignment with SOTA scalable approaches concerning software development (horizontal and vertical approach). Any new modules added will be incorporated into the broader framework as isolated stand-alone components following the best practices of SaaS (System as a Service) architecture.

4.6 Deployment plan

To be provided at a later stage in the document version B.

5 Use Case III – Crop

5.1 Introduction

UC3, led by SISTEMA, with a contribution from the Euro-Mediterranean Center on Climate Change (CMCC), will perform crop productivity estimations based on EO data and AI. The scope of the use case is the use of specific climate indicators to evaluate the impact of extreme climate events and other adverse phenomena on crops. The indicators are used to estimate risk, make forecasts, and issue alerts for potential production losses. This way, it is possible to investigate feedback loops of environmentally damaging food systems on the climate and food production. As a result, transformative adaptation is enabled, promoting long-term resilience by continually shifting the geographical locations where specific types of crops and livestock are produced, aligning agricultural production with changing landscapes and ecosystems, and/or introducing resilience-building production methods and technologies across value chains.

5.2 Workflow

The scope of this chapter is to introduce and describe the workflow of use case III using tables. The chapter contains four dedicated sub-chapters, covering the input data needs, pre-processing and processing steps and the generated results.

5.2.1 Input data

Table 13: Use case 3 input data specifications

Input data (Dx)	Data	Geographic extent	Spatial resolution	Time range	Data source	Data format	Disk size
D1	AgERA5 2m temperature	Italy	10 km	10 years	CDS	NetCDF-4	3GB
D2	AgERA5 Precipitation flux	Italy	10 km	10 years	CDS	NetCDF-4	3GB
D3	AgERA5 2m relative humidity	Italy	10 km	10 years	CDS	NetCDF-4	3GB
D4	AgERA5 10m wind speed	Italy	10 km	10 years	CDS	NetCDF-4	3GB
D5	AgERA5 Solar radiation flux	Italy	10 km	10 years	CDS	NetCDF-4	3GB
D6	Sentinel-1 GRD	Italy	10 m	From 2016 to present	Sentinel Open Access Hub / DIAS	*.zip files in SAFE format	1GB
D7	Sentinel-2 L1C - L2A	Italy	10 m	From 2016 to present	Sentinel Open Access Hub / DIAS	*.zip files in SAFE format	1GB for L2A 600MB for L1C
D8	PRISMA	Italy	30 m	From March 2019 to 2024	PRISMA mission portal	HDF-EOS5 format	1 GB ca. for L2C per single tile
D9	Corine Land Cover (CLC)	Italy	10 m	2021	Copernicus CLMS	TIFF	805 MB ca.
D10	ESA World Cover	Italy	N/A	2023		GeoTIFF (COG)	8.8 MB ca.
D11	Sicily CTR (Regional Technical Chart)	Italy	N/A	10 years	ISTAT Italy Statistical Institute	.shp	Few MB
D12	ISTAT	Italy	0.11 degree	10 years (historical)	ISTAT database	.csv	7 MB ca. for 1 year (compressed zip file)
D13	EURO-CORDEX (historical simulations	Italy	0.11	20-30 years	CDS	nc (zipped)	7 MB ca. for 1 year

Input data (Dx)	Data	Geographic extent	Spatial resolution	Time range	Data source	Data format	Disk size
	and projections) -– Temperature		degree	(projections)			(compressed zip file)
D14	EURO-CORDEX (historical simulations and projections) -– Precipitation	Italy	0.11 degree	10 years (historical)	CDS	nc (zipped)	7 MB ca. for 1 year (compressed zip file)
D15	EURO-CORDEX (historical simulations and projections) - Relative humidity	Italy	0.11 degree	20-30 years (projections)	CDS	nc (zipped)	7 MB ca. for 1 year (compressed zip file)
D16	EURO-CORDEX (historical simulations and projections) - Wind speed	Italy	0.11 degree	10 years (historical)	CDS	nc (zipped)	7 MB ca. for 1 year (compressed zip file)
D17	EURO-CORDEX (historical simulations and projections) - Solar radiation	Italy	0.02 degree	20-30 years (projections)	CDS	nc (zipped)	7 MB ca. for 1 year (compressed zip file)
D18	CMCC VHR (historical simulations and projections) - Temperature	Italy	0.02 degree	10 years (historical)	CMCC-DDS	nc	<15 GB per year
D19	CMCC VHR (historical simulations and projections) - Precipitation	Italy	0.02 degree	20-30 years (projections)	CMCC-DDS	nc	<15 GB per year
D20	CMCC VHR (historical simulations and projections) - Dew point temperature	Italy	0.02 degree	10 years (historical)	CMCC-DDS	nc	<15 GB per year
D21	CMCC VHR (historical simulations and projections) - Wind speed	Italy	0.02 degree	20-30 years (projections)	CMCC-DDS	nc	<15 GB per year
D22	CMCC VHR (historical simulations and projections) - Solar radiation	Italy	N/A	10 years (historical)	CMCC-DDS	nc	<15 GB per year
D23	CREA-Rica crops production data	Italy	N/A	10 years	Council for Agricultural Research and Economics (CREA)	.xls	Few KB

Input data (Dx)	Data	Geographic extent	Spatial resolution	Time range	Data source	Data format	Disk size
D24	Emilia-Romagna Region soil map	Emilia-Romagna region	Scale: 1:50000	2021	Emilia-Romagna Region Environmental database	.shp	32 MB (zip file)
D25	Campania soil map	Campania region	Scale: 1:250000	2014	Sistema Informativo Territoriale (S.I.T.) Regione Campania	.shp	6 MB (zip file)
D26	Puglia soil map	Puglia region	Scale: 1:50000	N/A	Sistema Informativo Territoriale (S.I.T.) Regione Puglia	.shp	6.6 MB (zip file)

5.2.2 Pre-processing

Table 14: Use case 3 pre-processing steps specifications

Pre-processing (PPx)	Input Data	Step description	Outcome (preservation y/n)	Processing tool (open source or commercial)	Processing interface	Facility	Infrastructure requirements / resource needs (development)	Infrastructure requirements / resource needs (operational)
PP1	D1	Derivation of bioclimatic indicators feeding the model	n	Open source	API	local (MEEO)	N/A	N/A
PP2	D2	Derivation of bioclimatic indicators feeding the model	n	Open source	API	local (MEEO)	N/A	N/A
PP3	D3	Derivation of bioclimatic indicators feeding the model	n	Open source	API	local (MEEO)	N/A	N/A
PP4	D4	Derivation of bioclimatic indicators feeding the model	n	Open source	API	local (MEEO)	N/A	N/A
PP5	D5	Derivation of bioclimatic indicators feeding the model	n	Open source	API	local (MEEO)	N/A	N/A

Pre-processing (PPx)	Input Data	Step description	Outcome (preservation y/n)	Processing tool (open source or commercial)	Processing interface	Facility	Infrastructure requirements / resource needs (development)	Infrastructure requirements / resource needs (operational)
PP6	D6	Subsetting, orbit file calibration, speckle filtering, range terrain correction, thermal correction, (decibel conversion), conversion to TIFF	n	Open source	CLI / API	Remote	N/A	N/A
PP7	D7	Conversion to TIFF, atmospheric correction (L1C to L2A)	n	Open source	CLI / API	Remote	N/A	N/A
PP8		Conversion to TIFF, atmospheric correction	n	Open source	request of user account and manual download	local (MEEO)	N/A	N/A
PP9	D9	Subsetting	n	Open source		Remote	N/A	N/A
PP10	D10	Subsetting	y	Open source	API	local (MEEO)	N/A	N/A
PP11	D11	data storing, data extraction at province level, data aggregation	y	Open source	GIS software	local (MEEO)	N/A	N/A
PP12	D12	extraction of data at province level, data aggregation (province/crop of interest)	n	Open source	GUI/MO sw interface	remote	N/A	N/A
PP13	D13	Derivation of bioclimatic indicators feeding the model	n	Open source	API	remote	N/A	N/A
PP14	D14	Derivation of bioclimatic indicators feeding the model	n	Open source	API	remote	N/A	N/A
PP15	D15	Derivation of bioclimatic indicators feeding the model	n	Open source	API	remote	N/A	N/A
PP16	D16	Derivation of bioclimatic indicators feeding the model	n	Open source	API	remote	N/A	N/A

Pre-processing (PPx)	Input Data	Step description	Outcome (preservation y/n)	Processing tool (open source or commercial)	Processing interface	Facility	Infrastructure requirements / resource needs (development)	Infrastructure requirements / resource needs (operational)
PP17	D17	Derivation of bioclimatic indicators feeding the model	n	Open source	API	remote	N/A	N/A
PP18	D18	Derivation of bioclimatic indicators feeding the model	n	Open source	API	remote	N/A	N/A
PP19	D19	Derivation of bioclimatic indicators feeding the model	n	Open source	API	remote	N/A	N/A
PP20	D20	Derivation of bioclimatic indicators feeding the model	n	Open source	API	remote	N/A	N/A
PP21	D21	Derivation of bioclimatic indicators feeding the model	n	Open source	API	remote	N/A	N/A
P22	D22	Data storing, extraction at regional and province level and data aggregation (province/region/crop of interest).	y	Open source	GUI/MO sw interface	local (MEEO)	N/A	N/A
P23	D23	Data storing, data extraction at province level, data aggregation	y	Open source	GIS software	local (MEEO)	N/A	N/A
P24	D24	data storing, data extraction at province level, data aggregation	y	Open source	GIS software	local (MEEO)	N/A	N/A
P25	D25	data storing, data extraction at province level, data aggregation	y	Open source	GIS software	local (MEEO)	N/A	N/A

5.2.3 Processing

Table 15: Use case 3 processing steps specifications

Processing	Input	Step description	Outcome (preservation	Processing tool (open source or	Processing	Facility	Infrastructure requirements /	Infrastructure requirements /

step (Sx)	Data		y/n)	commercial)	interface		resource needs (development)	resource needs (operational)
S1	D6, D7	data load and splitting; training, validation and testing	y	gdal library; fast AI (Torch) Python library	Jupyter notebook / CLI	remote	N/A	GPU /CPU
S2	D6, D7	choice of application (super-resolution); choice of architecture, choice of model (discriminative/generative models), (SISR, MISR) choice of loss function	y	gdal library; fast AI (Torch) Python library	Jupyter notebook / CLI	remote	N/A	GPU
S3	D6, D7	identification and monitoring of metrics (fine-tuning)	y	gdal library; fast AI (Torch) Python library	Jupyter notebook / CLI	remote	N/A	GPU
S4	D6, D7	to measure the efficiency of the model	y	gdal library; fast AI (Torch) Python library	Jupyter notebook / CLI	remote	N/A	GPU
S5	D6, D7	creation of an application of the model	y	gdal library; fast AI (Torch) Python library	Jupyter notebook / CLI	remote	N/A	GPU /CPU
S6	D13-D22	conduction of projections	y	gdal library; fast AI (Torch) Python library	Jupyter notebook / CLI	remote	N/A	PU /CPU

5.2.4 Results

Table 16: Use case 3 results specifications

Result (Rx)	**Result description**	**Result delivery**	**Infrastructure requirements / resource needs** (development)	**Infrastructure requirements / resource needs** (operational)

R1	Map showing the future change in crop production in correlation with relevant climate indicators	online (Data Lake/basic explorer)	N/A	4 CPUs, 16 GB RAM
R2	High-resolution maps showing vegetation indexes in correlation with weather parameters from reanalysis	online (Data Lake/basic explorer)	N/A	4 CPUs, 16 GB RAM
R3	Maps showing areas with the best climate conditions for specific crops and cultivation potential values (different classes, from Low=0 to High=100)	online (Data Lake/basic explorer)	N/A	4 CPUs, 16 GB RAM
R4	Map showing climate projections with crops classified differently depending on their suitability in each specific area	online (Data Lake/basic explorer)	N/A	4 CPUs, 16 GB RAM

5.3 Data management needs

The data management needs of UC3 are visualised tabularly in Tables 13, 14 and 15.

Table 13 covers the input data needs. The table provides information on the twenty-two input datasets relevant to UC3. The datasets include AgERA5 2m Temperature, Precipitation flux, 2m relative humidity, 10m wind speed and Solar radiation flux which are retrieved from the CDS in nc file format. The native resolution of these datasets is 0.1° x 0.1°, with spatial coverage at the country level (Italy) and 10 years of time coverage. The Sentinel-1 and -2 data are retrieved through the Sentinel Open Access Hub in *.zip files in SAFE format with global coverage from 2016 to the present.

Data from the PRISMA mission portal are available in HDF-EOS5 format through the PRISMA/ASI portal with 30 m spatial resolution, with coverage over Italy from March 2019 to 2024.

The input data also include the ESA World Cover with a spatial resolution of 10 m and a global geographical coverage for the year 2021.

The Regional Technical Chart from the Italy Statistical Institute (ISTAT) is available through the ISTAT web interface, it has a country spatial coverage (Italy) and the last update refers to the year 2023.

The EURO-CORDEX historical simulations and projections include the following variables: temperature, precipitation, relative humidity, wind speed and solar radiation which are retrieved, with daily time step, through the CDS upon request. The native resolution of these datasets is 0.11° x 0.11°, with spatial coverage at the country level (Italy) and 10 years of time coverage for historical/evaluation runs and 20-30 years for future runs.

The CMCC VHR historical simulations and projections include the following variables: temperature, precipitation, dew point temperature, wind speed and solar radiation, which are retrieved through the CMCC-DDS with hourly time step and at a spatial resolution of 0.02 degrees over Italy and with a 10 years' time coverage for historical/evaluation runs and 20-30 years from future runs.

The Rica crops production data from the Council for Agricultural Research and Economics (CREA) are available through the CREA web interface at country geographical coverage (Italy) and a 10 years' time range will be considered for the present use case.

Finally, the UC3 input data include the soil maps for the following Italian regions: Emilia-Romagna (year 2021), Campania (year 2014) and Puglia, all in shapefile format and retrieved through the respective regional information system.

The size of the AgERA5 data is 3GB per day for the global geographical coverage. The size of the Sentinel-1 GRD and Sentinel-2 L2A data is about 1GB per image (raw data) and per SAFE file respectively, while it is about 600 MB for the Sentinel-2 L1C per SAFE file. The ESA WorldCover 10m 2021 V200 product is delivered in 3x3 degree tiles grouped in 60x60 degree tiles. The approximate size for the tiles covering the geographical area of Italy is 805 MB. The PRISMA data have an approximate size of 1 GB ca. for L2C per single tile, while the ESA World Cover size for the area over Italy is 805 MB ca.

The size for the ISTAT Regional Technical Chart is 8.8 MB ca. while for the ISTAT crops production data is a few MB and for the CREA crops production data is a few KB.

The size for the EURO-CORDEX data is between 200 and 300 MB approximately per year per variable.

The respective size for the Emilia-Romagna, Campania and Puglia soil maps are 32, 6 and 6.6 MB.

Pre-processing of the input data includes the derivation of bioclimatic indicators feeding the model starting from the AgERA5, the EURO-CORDEX and the CMCC VHR data. Sentinel-1 data foresee sub-setting, orbit file calibration, speckle filtering, range terrain correction, thermal correction, (decibel conversion), and conversion to TIFF. Sentinel-2 as well as PRISMA data pre-processing includes conversion to TIFF and atmospheric correction. The Italian Administrative boundaries data from ISTAT as well as the soil maps for the Emilia-Romagna, the Campania and the Puglia regions foresee data storing, data extraction at regional and province level, and data aggregation. Both the ISTAT and CREA crops production data include storing, extraction at regional and province level and data aggregation (province/region/crop of interest).

The pre-processing of each input dataset is visualized in Table 14.

Table 16 gives an overview of the outcome of the process implemented in UC3. The results include maps showing the future change in crop production in correlation with relevant climate indicators (R1); high-resolution maps showing vegetation indexes in correlation with weather parameters from reanalysis (R2); maps of areas with best climate conditions for specific crops and cultivation potential areas (R3); and climate projections maps with crops classification depending on their suitability in each specific area (R4).

All results shall be made available online through a Data Lake/basic explorer.

5.4 Software component needs

The software component needs of UC3 are reported in section s5.2.2. and 5.2.3. The software components are substantially based on open-source tools for what concerns the exploratory analysis of the datasets and the application of Machine Learning techniques and partially based on the ADAM platform for what concerns the pre-processing and the data preparation. The scope of the processing is the training of a model to predict crop productivity under given climate conditions. The model is then applied to decadal climate scenarios to estimate the potential productivity I the future and to identify different climatic areas that might result favourable for a given type of crop.

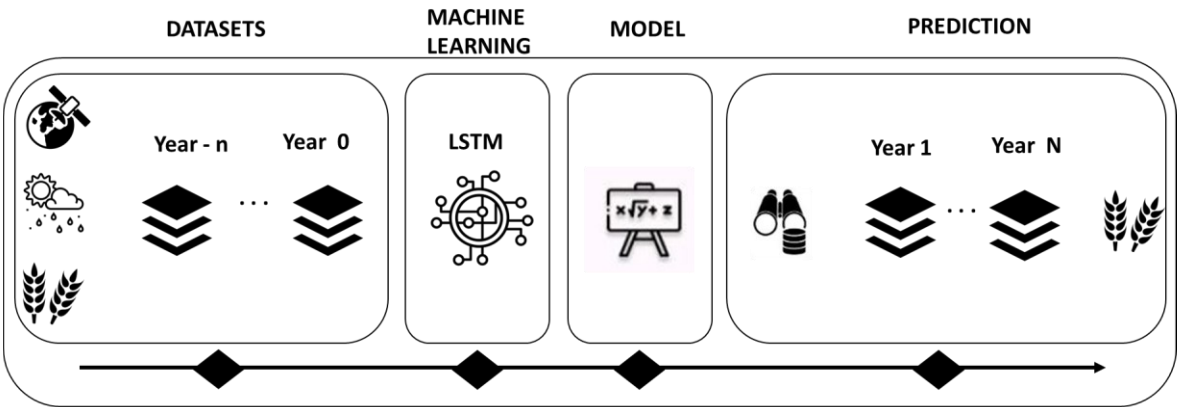


Figure 1: Pipeline for use case III

The configuration of the training datasets listed in the previous section consists of a subset of historical climate data and production data about a given crop type over different climate areas. The exploratory analysis of these datasets includes a comparison of data at the district scale (statistical data) with observed (local) production data and related vegetation time series. Once the data have been consolidated the resulting dataset is used for model training. The crop productivity model is applied to the identified regional datasets for verification and validation. The productivity model is then applied to long-term climate forecast data (climate scenario) to identify areas with a predicted loss of crop productivity and vice-versa to identify those areas that could present favourable climate conditions for a given crop type.

The AI component of this workflow is based on the fast AI Python library.

5.5 Computational infrastructure needs

The virtual environment will be based on e.g. Jupyter notebooks for the following steps:

- pre-processing phase for data preparation
- exploratory analysis
- local data analysis
- LSTM model training
- predictive analysis

These tools are not supposed to be used in parallel thus the computation user needs can be estimated in 4 CPUs; 16 GB RAM; 500 GB storage. In terms of scalability, it might result necessary to extend the training over more areas within the four regions identified as areas of interest, while for the prediction of estimated productivity, the analysis will be performed over Italy. The need for resources for the operational phase could increase a factor concerning the development phase.

5.6 Deployment plan

To be provided at a later stage in the document version B.

6 Use Case IV – Forest

6.1 Introduction

UC4, led by CMCC, with the support of SISTEMA, will evaluate the impacts of current and future climates on forests using Forest Ecosystem Models (FEMs). These models simulate the fluxes of water, energy, and carbon, and can be used to project the effects of modified climates on forest growth and carbon sequestration. This information can be used to enhance sustainable forest management in the face of climate change and the demand for forest ecosystem services. Climate change may cause forest species to adapt to new conditions or migrate to areas with more suitable conditions for survival. Changes in climate dynamics can impact the quantity and quality of goods and functions provided by forests, known as forest ecosystem services. The scope of this use case is to generate information which can be used to identify and validate a business model for forestry companies that exchange ecosystem services produced by sustainably managed forests.

6.2 Workflow

The scope of this chapter is to introduce and describe the workflow of the use case IV using tables. The chapter contains four dedicated sub-chapters, covering the input data needs, pre-processing and processing steps and the generated results.

6.2.1 Input data

Table 17: Use case 4 input data specifications

Input data (Dx)	Data	Geographic extent	Spatial resolution	Time range	Data source	Data format	Disk size (Aafter tailoring to the UC area)
D1	ERA5-Land; CERRA (TBD) - Temperature	global (ERA5-Land); Europe (CERRA)	ca. 9 km (ERA5-Land); 5.5 km or 11 km based on the reanalysis system (CERRA)	1950-present (ERA5-Land); 1984 (Sep) - 2021 (Jun) (CERRA, soon operational)	CDS	grib/NetCDF	ERA5-Land (105KB per 24 hours in Austria): <1.3 GB for 1991-2022; CERRA: Very rough estimate based on hor.res. proportion from ERA5-land above and considering 3-hourly resolution: < 300 MB or < 1.1 GB, function of the reanalysis system (11km o 5.5 km)
D2	ERA5-Land; CERRA (TBD) - Precipitation	global (ERA5-Land); Europe (CERRA)	ca. 9 km (ERA5-Land); 5.5 km or 11 km based on the reanalysis system (CERRA)	1950-present (ERA5-Land); 1984 (Sep) - 2021 (Jun) (CERRA, soon operational)	CDS	grib/NetCDF	ERA5-Land (105KB per 24 hours in Austria): <1.3 GB for 1991-2022; CERRA: Very rough estimate based on hor.res. proportion from ERA5-land above and considering 3-hourly resolution: < 300 MB or < 1.1 GB, function of the reanalysis system (11km o 5.5 km)
D3	ERA5-Land; CERRA (TBD) - Radiation	global (ERA5-Land); Europe (CERRA)	ca. 9 km (ERA5-Land); 5.5 km or 11 km based on the reanalysis system (CERRA)	1950-present (ERA5-Land); 1984 (Sep) - 2021 (Jun) (CERRA, soon operational)	CDS	grib/NetCDF	ERA5-Land (105KB per 24 hours in Austria): <1.3 GB for 1991-2022; CERRA: Very rough estimate based on hor.res. proportion from ERA5-land above and considering 3-hourly resolution: < 300 MB or < 1.1 GB, function of the reanalysis system (11km o 5.5 km)
D4	ERA5-Land;	global (ERA5-	ca. 9 km (ERA5-	1950-present	CDS	grib/NetCDF	ERA5-Land (105KB per 24 hours in

Input data (Dx)	Data	Geographic extent	Spatial resolution	Time range	Data source	Data format	Disk size (Aafter tailoring to the UC area)
	CERRA (TBD) - VPD	Land); Europe (CERRA)	Land); 5.5 km or 11 km based on the reanalysis system (CERRA)	(ERA5-Land); 1984 (Sep) - 2021 (Jun) (CERRA, soon operational)			Austria): <1.3 GB for 1991-2022; CERRA: Very rough estimate based on hor.res. proportion from ERA5-land above and considering 3-hourly resolution: < 300 MB or < 1.1 GB, function of the reanalysis system (11km o 5.5 km)
D5	EURO-CORDEX (historical simulations and projections) - Temperature (mean, min, max)	pan-European	0.11°	1951-2100 (1989-2008 for evaluation)	CDS	nc (zipped)	Very rough estimate based on hor.res. proportion from ERA5-land above and considering daily resolution for each 30-year period: < 50 MB
D6	EURO-CORDEX (historical simulations and projections) - Precipitation	pan-European	0.11°	1951-2100 (1989-2008 for evaluation)	CDS	nc (zipped)	Very rough estimate based on hor.res. proportion from ERA5-land above and considering daily resolution for each 30-year period: < 50 MB
D7	EURO-CORDEX (historical simulations and projections) – Radiation	pan-European	0.11°	1951-2100 (1989-2008 for evaluation)	CDS	nc (zipped)	Very rough estimate based on hor.res. proportion from ERA5-land above and considering daily resolution for each 30-year period: < 50 MB
D8	EURO-CORDEX (historical simulations and projections) - VPD	pan-European	0.11°	1951-2100 (1989-2008 for evaluation)	CDS	nc (zipped)	Very rough estimate based on hor.res. proportion from ERA5-land above and considering daily resolution for each 30-year period: < 50 MB
D9	SOILGRIDS	global	250 m	Latest dataset/version	ISRIC	GeoTIFF	Per property: Austria 0.07% of 120 GB= 84MB. N.B. Each global map occupies circa 5 GB. The full

Input data (Dx)	Data	Geographic extent	Spatial resolution	Time range	Data source	Data format	Disk size (Aafter tailoring to the UC area)
				available			collection of maps for a single property, with six standard depth intervals and four quantiles per depth requires about 120 GB.
D10	ERA5-Land; CERRA; ESDB (TBD) - Available Soil Water	global (ERA5-Land); Europe (CERRA); Europe (ESDB)	ca. 9 km (ERA5-Land); 5.5 km or 11 km based on the reanalysis system (CERRA); 1 km (ESDB)	1950-present (ERA5-Land); 1984 (Sep) - 2021 (Jun) (CERRA, soon operational); Not Applicable (ESDB)	CDS (ERA5-Land, CERRA); ESDAC (ESDB)	grib/NetCDF (ERA5-Land, CERRA) ESRI-GRID (ESDB)	ERA5-Land (420KB per 24 hours in Austria): <5 GB for 1991-2022 CERRA: Very rough estimate based on hor.res. proportion from ERA5-land above and considering 3-hourly resolution: < 1.2 GB or < 4.5 GB, function of the reanalysis system (11km o 5.5 km) ESDB: < 0.8% (surface Austria vs. Europe) of 40MB (each parameter) x2 layers (top/sub) < 1MB
D11	site latitude	local	Based on local info source and the number of sites	Not Applicable	Unknow, possibly from local surveys/campaigns	Not Applicable	Not relevant
D12	site soil fertility rating	local	Based on local info source or spatialized input dataset to derive	Unknown	Unknow, possibly from local surveys/campaigns on behalf of forest owners/managers or derived from soil databases (e.g., additional attributed from D9)	Unknown	Not relevant
D13	EU-DEM	pan-European	25 m	Latest dataset/version available	CLMS	GeoTIFF	< 6 GB the tile covering Italy
D14	CO_2	global	Not applicable	1765 to future	PROFOUND DATABASE	relational SQLite made of independent	Not relevant

Input data (Dx)	Data	Geographic extent	Spatial resolution	Time range	Data source	Data format	Disk size (Aafter tailoring to the UC area)
						tables	
D15	Soil class	as D5	as D5	as D5	Parameterized from D5 and fao.org	as D5	Derived from operating property of D9 into a new property: Austria 0.07% of 5 GB=3.5MB
D16	Seedling mass (species-specific)	Not Applicable (species-specific)	Not Applicable (species-specific)	Referring to the initial state	info from local surveys/campaigns?	Unknown	Not relevant
D17	Stand mass (species-specific)	Not Applicable (species-specific)	Not Applicable (species-specific)	Referring to the initial state	info from local surveys/campaigns?	Unknown	Not relevant
D18	Initial WF (species-specific)	Not Applicable (species-specific)	Not Applicable (species-specific)	Referring to the initial state	info from local surveys/campaigns?	Unknown	Not relevant
D19	Initial WR (species-specific)	Not Applicable (species-specific)	Not Applicable (species-specific)	Referring to the initial state	info from local surveys/campaigns?	Unknown	Not relevant
D20	Initial WS (species-specific)	Not Applicable (species-specific)	Not Applicable (species-specific)	Referring to the initial state	info from local surveys/campaigns?	Unknown	Not relevant
D21	Initial stocking (species-specific)	Not Applicable (species-specific)	Not Applicable (species-specific)	R Referring to the initial state	info from local surveys/campaigns?	Unknown	Not relevant
D22	Initial ASW	local OR as D10	Not Applicable (site specific) OR as D10	Unknown OR as D10	info from local surveys/campaigns or derived from D10	Unknown OR as D10	Not relevant
D23	Silvicultural "events": Fertilising data	Based on forest management	Based on forest management	Forest management time	Info from management plans	Unknown	Not relevant

Input data (Dx)	Data	Geographic extent	Spatial resolution	Time range	Data source	Data format	Disk size (Aafter tailoring to the UC area)
		info	spatialization	scale			
D24	Silvicultural "events": Min ASW data	Based on forest management info	Based on forest management spatialization	Forest management time scale	Info from management plans	Unknown	Not relevant
D25	Silvicultural "events": Irrigation data	Based on forest management info	Based on forest management spatialization	Forest management time scale	Info from management plans	Unknown	Not relevant
D26	Silvicultural "events": Thinning data	Based on forest management info	Based on forest management spatialization	Forest management time scale	Info from management plans	Unknown	Not relevant
D27	Silvicultural "events": Defoliation data	Based on forest management info	Based on forest management spatialization	Forest management time scale	Info from management plans	Unknown	Not relevant
D28	species-specific parameters	Not Applicable (species-specific)	Not Applicable (species-specific)	Not Applicable (species-specific)	References and values in the 3PGmix_Parameters sheets in the 3PGmix package	.xlsx	Not relevant

6.2.2 Pre-processing

Table 18: Use case 4 pre-processing steps specifications

Pre-processing (PPx)	Input Data	Step description	Outcome (preservation y/n)	Processing tool (open source or commercial)	Processing interface	Facility	Infrastructure requirements / resource needs (development)	Infrastructure requirements / resource needs (operational)
PP1	D1	convert Kelvin to Celsius and calculate Monthly mean daily maximum temperature, Monthly mean daily minimum temperature and Monthly mean	Y, needed for processing	Open source	API	remote	See disk size in 6.2.1	See disk size in 6.2.1

Pre-processing (PPx)	Input Data	Step description	Outcome (preservation y/n)	Processing tool (open source or commercial)	Processing interface	Facility	Infrastructure requirements / resource needs (development)	Infrastructure requirements / resource needs (operational)
		temperature, Calculate frost days per month (e.g., days with < 0°C minimum temperature)						
PP2	D2	convert to Monthly rainfall (mm month -1)	Y, needed for processing	Open source	API	remote	See disk size in 6.2.1	See disk size in 6.2.1
PP3	D3	calculate Monthly mean daily solar radiation (MJ m-2 day-1)	Y, needed for processing	Open source	API	remote	See disk size in 6.2.1	See disk size in 6.2.1
PP4	D4	convert to relative humidity (ERA5-Land) then calculate VPD (in mbar)	Y, needed for processing	Open source	API	remote	See disk size in 6.2.1	See disk size in 6.2.1
PP5	D5	convert Kelvin to Celsius and calculate Monthly mean daily maximum temperature, Monthly mean daily minimum temperature and Monthly mean temperature, Calculate frost days per month (e.g., days with < 0°C minimum temperatures)	Y, needed for processing	Open source	API	remote	See disk size in 6.2.1	See disk size in 6.2.1
PP6	D6	convert to Monthly rainfall (mm month -1)	Y, needed for processing	Open source	API	remote	See disk size in 6.2.1	See disk size in 6.2.1
PP7	D7	calculate Monthly mean daily solar radiation (MJ m-2 day-1)	Y, needed for processing	Open source	API	remote	See disk size in 6.2.1	See disk size in 6.2.1
PP8	D8	Calculate VPD (in mbar)	Y, needed for processing	Open source	API	remote	See disk size in 6.2.1	See disk size in 6.2.1
PP9	D9	Texture parameterization from available layers	Y, needed for processing	Open source	API	remote	See disk size in 6.2.1	See disk size in 6.2.1
PP10	D10	ASW parameterization from available layers or AWC	Y, needed for processing	Open source	API/other	remote	See disk size in 6.2.1	See disk size in 6.2.1
PP11	D11	Formatting according to model requirements	Y, needed for processing	Open source	other/CLI	local	See disk size in 6.2.1	See disk size in 6.2.1

Pre-processing (PPx)	Input Data	Step description	Outcome (preservation y/n)	Processing tool (open source or commercial)	Processing interface	Facility	Infrastructure requirements / resource needs (development)	Infrastructure requirements / resource needs (operational)
PP12	D12	Spatial disaggregation and data formatting according to model requirements	Y, needed for processing	Open source	other/CLI	local	See disk size in 6.2.1	See disk size in 6.2.1
PP13	D13	Spatial disaggregation and data formatting according to model requirements	Y, needed for processing	Open source	other/CLI	local	See disk size in 6.2.1	See disk size in 6.2.1
PP14	D14	Formatting according to model requirements	Y, needed for processing	Open source	other/CLI	local	See disk size in 6.2.1	See disk size in 6.2.1
PP15	D15	Parameterized from D9 and https://www.fao.org/fishery/docs/CDrom/FAO_Training/FAO_Training/General/x6706e/x6706e06.htm; Spatial disaggregation and data formatting according to model requirements	Y, needed for processing	Open source	other/CLI	local	See disk size in 6.2.1	See disk size in 6.2.1
PP16	D16	Spatial disaggregation and data formatting according to model spatialization and requirements	Y, needed for processing	Open source	other/CLI	local	See disk size in 6.2.1	See disk size in 6.2.1
PP17	D17	Spatial disaggregation and data formatting according to model spatialization and requirements	Y, needed for processing	Open source	other/CLI	local	See disk size in 6.2.1	See disk size in 6.2.1
PP18	D18	Spatial disaggregation and data formatting according to model spatialization and requirements	Y, needed for processing	Open source	other/CLI	local	See disk size in 6.2.1	See disk size in 6.2.1
PP19	D19	Spatial disaggregation and data formatting according to model spatialization and requirements	Y, needed for processing	Open source	other/CLI	local	See disk size in 6.2.1	See disk size in 6.2.1
PP20	D20	Spatial disaggregation and data formatting according to model spatialization and requirements	Y, needed for processing	Open source	other/CLI	local	See disk size in 6.2.1	See disk size in 6.2.1
PP21	D21	Spatial disaggregation and data formatting according to model spatialization and requirements	Y, needed for processing	Open source	other/CLI	local	See disk size in 6.2.1	See disk size in 6.2.1
PP22	D22	Spatial disaggregation and data formatting according to model spatialization and requirements	Y, needed for processing	Open source	other/CLI	local	See disk size in 6.2.1	See disk size in 6.2.1

Pre-processing (PPx)	Input Data	Step description	Outcome (preservation y/n)	Processing tool (open source or commercial)	Processing interface	Facility	Infrastructure requirements / resource needs (development)	Infrastructure requirements / resource needs (operational)
PP23	D23	Spatial disaggregation and data formatting according to model spatialization and requirements	Y, needed for processing	Open source	other/CLI	local	See disk size in 6.2.1	See disk size in 6.2.1
PP24	D24	Spatial disaggregation and data formatting according to model spatialization and requirements	Y, needed for processing	Open source	other/CLI	local	See disk size in 6.2.1	See disk size in 6.2.1
PP25	D25	Spatial disaggregation and data formatting according to model spatialization and requirements	Y, needed for processing	Open source	other/CLI	local	See disk size in 6.2.1	See disk size in 6.2.1
PP26	D26	Spatial disaggregation and data formatting according to model spatialization and requirements	Y, needed for processing	Open source	other/CLI	local	See disk size in 6.2.1	See disk size in 6.2.1
PP27	D27	Spatial disaggregation and data formatting according to model spatialization and requirements	Y, needed for processing	Open source	other/CLI	local	See disk size in 6.2.1	See disk size in 6.2.1
PP28	D28	Data formatting according to the model requirement	Y, needed for processing	Open source	other/CLI	local	See disk size in 6.2.1	See disk size in 6.2.1

6.2.3 Processing

Table 19: Use case 4 processing steps specifications

Processing step (Sx)	Input Data	Step description	Outcome (preservation y/n)	Processing tool (open source or commercial)	Processing interface	Facility	Infrastructure requirements / resource needs (development)	Infrastructure requirements / resource needs (operational)

S1	D1-D4; D13 & others	Machine learning for climate downscaling	Y	Languages available on the platform for ML	launch by the platform	remote	Unknown
S2	D1-D4; D9-D22	Machine learning for model parameterization	Y	Languages available on the platform for ML	launch by the platform	remote	Unknown
S3	D1-D4; D9-D22	simulation excluding climate change and management	Y, only Biomass pools, production, allocation and Water Use outputs from the list in 3PGmix.Data.xls, sheet "3PGmixOutputs" from 3PGmix	3PG-mix code (re-written in a programmable environment)	launch by the platform	remote	The running time for 2 species and 13 years is 0.3500199 secs on a common pc (x64 based) Intel i7 3.4 GHz, RAM 8Gb. This duration has to be rescaled by several species, years of simulation and representative stands/grid points.
S4	D1-D4; D9-D28	simulation excluding climate change but including management					
S5	D5-D22	simulation excluding management but including climate change					
S6	D5-D28	simulation including management and climate change					

6.2.4 Results

Table 20: Use case 4 results specifications

Result (Rx)	Result description	Result delivery	Infrastructure requirements / resource needs (development)	Infrastructure requirements / resource needs (operational)
R1	Forest growth and carbon dynamics with no forest management or climate change	web GIS (visualization), direct download	Storage<150MB/map/time step (assuming 1 ha grid cell)	Storage<150MB/map/time step (assuming 1 ha grid cell)
R2	identify the effects of possible scenarios of CC on carbon dynamics and forest growth			
R3	identify the effects of alternative management on carbon dynamics and forest growth			
R4	suggest adaptative management to maximize forest growth and carbon sequestration under CC			
R5	Forest water use with no management or climate change			
R6	identify the effects of possible scenarios of CC on forest water use			
R7	identify the effects of alternative management on forest water use			
R8	suggest adaptative management to maximize the forest water use under CC			
R9	suggest adaptative management to maximize the forest water use, growth and carbon sequestration under CC			

6.3 Data management needs

Use case 4 aims to simulate the 3pgmix model utilizing input data consisting of various environmental variables and soil properties, with different combinations of climate change and management scenarios. The datasets used in this project include information on climate, soil properties, atmospheric CO2 levels, and species-specific initial conditions. Table 17 includes information such as the geographic extent, spatial resolution, time range, data source, data format, and disk size, when quantifiable, for each dataset.

The input data for the forest management system includes various data sources and formats. The first four data sources, D1 to D4, are from the CDS and provide temperature, precipitation, radiation, and vapour pressure deficit information from ERA5-Land and CERRA products, at the global to European level, with a data format of grib or NetCDF. The next four data sources, D5 to D8, are from EURO-CORDEX and provide historical simulations and projections for temperature (mean, min, max), precipitation, radiation, and vapour pressure deficit information in a zipped NetCDF format. Climate datasets are of the order of 5.5-11 km of horizontal resolution. D9 provides soil information from SOILGRIDS in GeoTIFF format, available for the whole globe at 250 m grid size, while D10 provides available soil water information still from CDS, dynamically from ERA5-Land and CERRA, in grib/netcdf formats, and static from ESDB in ESRI-GRID format at 1 km grid size. D11 provides site latitude information from local information, while D12 could potentially derive from local surveys/campaigns or can be secondarily derived from D9. D13 provides EU-DEM data in GeoTIFF format with a 25 m grid size derived from Copernicus Land Monitoring Service. D14 provides CO2 information from the Profound database in relational SQLite format, while D15 provides soil class information in the same format as D5. The remaining data sources, D16 to D28, are species- or site-specific input/initialization data and could be obtained from local surveys/campaigns or management plans, with various and for now unknown formats. The species-specific parameters are provided in an Excel file with references and values in the "3PGmix_Parameters sheets" in the 3PGmix package. The Table also reports the indicative disk size needed to store input data after tailoring to the UC domain.

There are four use cases for this simulation project. The first scenario involves simulating without climate change and management, using D1-D4 and D9-D22 datasets. The focus will be on biomass pools, production, allocation, and water use outputs, with a baseline simulation. The second scenario involves simulating without climate change but with different management alternatives, using D1-D4 and D9-D28 datasets. Similar to the first scenario, the focus will be on biomass pools, production, allocation, and water use outputs. The third scenario involves simulating different climate projections but without management, using D5-D22 datasets. The same outputs as the previous scenarios will be studied. Finally, the fourth scenario involves simulating both management and climate change scenarios, using D5-D28 datasets. The outputs of interest remain the same as in the previous scenarios.

The storage is expected to cover, for maps at 100m resolution over Austria, less than 150MB per time step (regardless of month or year) if in GeoTIFF format.

6.4 Software component needs

The software components needed for this project include exploiting programming languages and tools provided by the platform for data preparation, parameters' optimization, model evaluation and sensitivity analysis and also interfacing with programmable environments to run the re-parametrized 3-PGmix model. The AI/ML components used include Boosted Regression Tree Model

(BRT), Random Forest Model (RF), and Multivariate Adaptive Regression Splines Model (MARS) (or at least one of them) for model parameterization. The scope of the project includes predicting forest growth response to future climatic conditions and management treatments and providing insights on the types of trees that can be used to map land cover and detect soil properties. The configuration of the AI/ML components includes using the input data such as site characteristics, stand characteristics, and weather data, and deriving tree/stand variables such as basal area, volume, and average values for input data, etc. The pipeline includes data pre-processing, parameterization, model evaluation, and simulation of the re-parametrized 3-PGmix model coupled with climate scenarios.

6.5 Computational infrastructure needs

In terms of computation, the running time in case of a forest mix of only 2 species over 13-year simulations at monthly time step is around 0.35 secs on a common pc (x64 based) Intel i7 3.4 GHz, RAM 8Gb. This duration must be rescaled by the number of species, years of simulations and representative stands/grid points.

As a reference for the UC, climatological studies as the one proposed in the UC are expected to consider at least 30 (better 40) years per time horizon and 11 scenarios (1 historical plus at least 10 are for ensemble uncertainty evaluations in the future). Supposing to apply the model at 1 ha (100 m x 100 m) resolution over the whole Austrian forest area this will mean 3.9Mil points simulated by 440 years (if looking at a one-time horizon in the future). In terms of forest species, a mix of around five species (Picea Abies, Abies Alba, Pinus Sylvestris, Fagus Sylvatica, Quercus sp.) could be considered credible in each 1 ha area.

6.6 Deployment plan

To be provided at a later stage in the document version B.

7 Use Case V – Soil

7.1 Introduction

UC5, led by CMCC, with a contribution from SISTEMA, aims to integrate datasets on rainfall and soil susceptibility to assess the risk of water erosion. Soil erosion occurs when detached soil is transported and deposited away due to rainfall, runoff, snow melting, or irrigation. If the soil erosion rate is higher than the soil formation rate, the soil becomes depleted and the land's productivity is reduced. The economic costs of soil erosion are high and affect many sectors. The developed service will provide information on water-induced soil erosion to a variety of end-users, considering the potential impacts of climate change on erosion. This information can help land management actors and territorial planners make informed decisions on farming practices, forest management, and soil recovery after disturbances, and can also be used to design more resilient infrastructure. Investments can also be more appropriate if a range of future outlooks is considered when evaluating modified risks for infrastructure.

7.2 Workflow

The scope of this chapter is to introduce and describe the workflow of the use case V using tables. The chapter contains four dedicated sub-chapters, covering the input data needs, pre-processing and processing steps and the generated results.

7.2.1 Input data

Table 21: Use case 5 input data specifications

Input data (Dx)	Data	Geographic extent	Spatial resolution	Time range	Data source	Data format	Disk size
D1	Point scale gauges (retrieved observations)	Italy	N.A. (Point level)	various	Regional Authorities in Italy	txt, xls, csv	Not Applicable
D2	Point scale gauges (processed observations)	Italy	N.A. (Point level)	2002-2011	CMCC	xls	4.2 MB (current version)
D3	Point scale gauges (digitized observations)	Italy	N.A. (Point level)	various	Hydrological Yearbooks	pdf	Not Applicable
D4	ERA5/ERA5-Land (reanalysis)	global (ERA5), global land (ERA5-Land)	(ca. 28 km (ERA5; native ca. 31 km); ca. 9 km ERA5-Land)	1959-present (ERA5), 1950-present (ERA5-Land)	CDS	grib/NetCDF	< 100 MB (after clipping)
D5	CERRA (reanalysis)	pan-European	5.5 km or 11 km based on the reanalysis system	1984 (sep) - 2021 (jun)	CDS	grib/NetCDF	ERA5 (130KB per 24 hours in Italy): <1.6 GB for 1991-2022; ERA5-Land (792KB per 24 hours in Italy): <9.3 GB for 1991-2022 (after clipping)
D6	E-OBS (interpolation)	pan-European	0.10° or 0.25°	1950 - present	CDS	nc (zipped)	Very rough estimate based on hor.res. proportion from ERA5-land above and considering 3-hourly resolution: < 2.5 GB or < 8.5 GB, function of the reanalysis system (11km o 5.5 km) (after clipping)
D7	VHR-Rea (downscaled	Italy; 36°N-48°N	0.02° (≃2.2 km)	1981-2020	CMCC-DDS	nc	Very rough estimate based on hor.res. proportion from ERA5-land above and

Input data (Dx)	Data	Geographic extent	Spatial resolution	Time range	Data source	Data format	Disk size
	reanalysis)	- 5°W-20°E					considering daily resolution: < 500 MB (after clipping)
D8	EURO-CORDEX (evaluation, historical simulations and projections)	pan-European	0.11°	1951-2100 (1989-2008 for evaluation)	CDS	nc (zipped)	13GB/y per 1 hourly file for 1 variable 1 year; x 30 years ≈ 400 GB (after clipping)
D9	VHR-Pro (downscaled historical simulations and projections)	Italy; 36°N-49°N - 3°W-20°E	0.02° (≃2.2 km)	1981-2070	CMCC-DDS	nc	Very rough estimate based on hor.res. proportion from ERA5-land above and considering a daily resolution for each 30-year period: < 500 MB (per simulation member) (after clipping)
D10	Land cover products (TBD which ones)	global (LC, WC), pan-European (HRL)	LC: 100 m CLC: 100 m HRL: 10-20 m WC: 10 m	Annual from 2015 to 2019 (LC); 1990, 2000, 2006, 2012, 2018 (CLC); 2012, 2015, 2018 (HRL) 2020 (WC)	CLMS (LC, CLC, HRL); ESA (WC)	GeoTIFF	13GB/y per 1 hourly file for 1 variable 1 year; x 80 years ≈ 1000 GB (per simulation member) (after clipping)
D11	Land cover product	Italy		2000, 2006, 2012, 2018	SINANet	shapefile	LC = estimated 400 MB for Italy (Italy land area=0.25% of global land area), based on <200 GB for a globe for all maps: 14 for BASE years (2015), 15 for other years (2016, 2017, 2018, 2019); HRL = estimated < 5 GB for tiles including Italy: imperviousness (6 maps x 170 MB); tree cover (3 maps x 1 GB); dominant leaf (3 maps x 250 MB); forest type (3 maps x 15 MB). WC = estimate < 1.5 GB for tiles including Italy

Input data (Dx)	Data	Geographic extent	Spatial resolution	Time range	Data source	Data format	Disk size
							(All after clipping)
D12	soil products (SOILGRIDS and/or ESDB)	global (SOILGRIDS); pan-European (ESDB)	250 m (SOILGRIDS); 1 km (ESDB)	Latest dataset/version available	ISRIC (SOILGRIDS), JRC (ESDB)	GeoTIFF (SOILGRIDS), ESRI grid/vector (ESDB)	250 MB (after clipping)
D13	EU-DEM	pan-European	25 m	Latest dataset/version available	CLMS	GeoTIFF	Per property: Italy 0.25% of 120 GB=300 MB. N.B. Each global map occupies circa 5 GB. The full collection of maps for a single property, with six standard depth intervals and four quantiles per depth requires about 120 GB (after clipping)
D14	OpenStreetMap - roads and railways	Italy	Not applicable	updated to May 2023	OpenStreetMap (OSM) project (www.openstreetmap.org)	shapefile	< 7 GB the two tiles covering Italy (after clipping)
D15	Support Practices factor (P-factor) for the EU	pan-European	1 km	Latest dataset/version available	JRC	GeoTIFF	<11 GB (after clipping)
D16	Vegetation products	global	300 m	2014 - present	CLMS	NetCDF	> 1 GB (after clipping)

7.2.2 Pre-processing

Table 22: Use case 5 pre-processing steps specifications

Pre-processing (PPx)	Input Data	Step description	Outcome (preservation y/n)	Processing tool (open source or commercial)	Processing interface	Facility	Infrastructure requirements / resource needs (development)	Infrastructure requirements / resource needs (operational)
PP1	D1	Selection of rain gauges and native	y	manual due to	other	local	N/A	N/A

Pre-processing (PPx)	Input Data	Step description	Outcome (preservation y/n)	Processing tool (open source or commercial)	Processing interface	Facility	Infrastructure requirements / resource needs (development)	Infrastructure requirements / resource needs (operational)
		time coverage; also extract period 1991-present		different formats and coverage				
PP2	D2	Selection of rain gauges and native time coverage	y	automatic thanks to the previous harmonization	other	local	N/A	N/A
PP3	D3	Selection of rain gauges and native time coverage; also extract the most recent period	y	OCR/manual	other	local	N/A	N/A
PP4	D1, D3	Unification of D1 and D3 (raw data) in a unique database of observations; conversion to daily and monthly	y	automatic (code by licensed or open software)	other	local	N/A	N/A
PP5	D4	conversion to daily and monthly, conversion in mm, clipping in space and selection of multiple time windows	Y	automatic (code by licensed or open software)	other/CLI/API	remote	N/A	N/A
PP6	D5	conversion to daily and monthly, regular lat/lon grid, conversion in mm, clipping in space and selection of multiple time windows	Y	automatic (code by licensed or open software)	other/CLI/API	remote	N/A	N/A
PP7	D6	conversion to monthly, clipping in space and selection of multiple time windows	Y	automatic (code by licensed or open software)	other/CLI/API	remote	N/A	N/A
PP8	D7	conversion to daily and monthly, conversion in mm, clipping in space and selection of multiple time windows	Y	automatic (code by licensed or open software)	other/CLI/API	remote	N/A	N/A
PP9	D8	conversion to monthly, regular lat/lon grid, conversion in mm, clipping in space and selection of multiple time	Y	automatic (code by licensed or open	other/CLI/API	remote	N/A	N/A

Pre-processing (PPx)	Input Data	Step description	Outcome (preservation y/n)	Processing tool (open source or commercial)	Processing interface	Facility	Infrastructure requirements / resource needs (development)	Infrastructure requirements / resource needs (operational)
		windows		software)				
PP10	D9	conversion to monthly, conversion in mm, clipping in space and selection of multiple time windows	Y	automatic (code by licensed or open software)	other/CLI/API	remote	N/A	N/A
PP11	D10	conversion to netcdf, regular lat/lon grid, clipping in space	Y	automatic (code by licensed or open software)	other/CLI	remote	N/A	N/A
PP12	D11	conversion to netcdf, regular lat/lon grid	Y	automatic (code by licensed or open software)	other/CLI	remote	N/A	N/A
PP13	D12	conversion to netcdf, regular lat/lon grid, clipping in space	Y	automatic (code by licensed or open software)	other/CLI	remote	N/A	N/A
PP14	D13	conversion to netcdf, regular lat/lon grid, clipping in time and space	Y	automatic (code by licensed or open software)	other/CLI	remote	N/A	N/A
PP15	D14	conversion to netcdf, clipping in space	Y	automatic (code by licensed or open software)	other/CLI	remote	N/A	N/A
PP16	D15	conversion to netcdf, clipping in space	Y	automatic (code by licensed or open software)	other/CLI	remote	N/A	N/A
PP17	D16	clipping in space	y	automatic (code by licensed or open software)	other/CLI	remote	N/A	N/A

7.2.3 Processing

Table 23: Use case 5-processing steps specifications

Processing step (Sx)	Input Data	Step description	Outcome (preservation y/n)	Processing tool (open source or commercial)	Processing interface	Facility	Infrastructure requirements / resource needs (development)	Infrastructure requirements / resource needs (operational)
S1	D1 to D9; D13	Set-up of methodologies enabling the disaggregation of spatialized hourly or daily rainfall data based on observations.	spatialized rainfall products at sub-hourly resolution; uncertainty analysis to identify the best technique (y)	N/A	source (pseudo-)code from licensed or open software	locally tested and then remote	N/A	N/A
S2	D4 to D7; D13	Comparison among different precipitation datasets remapped on the EURO-CORDEX grid	uncertainty analysis (not data, keep locally)	licensed software such as Matlab/Arcgis		local	N/A	N/A
S3	D4 to D7; D13; outcome of S1	Computation of rainfall erosivity relying on different datasets for current climate conditions on native and remapped (on EURO-CORDEX) grids	rainfall erosivity maps for the current period (y)	N/A	source (pseudo-)code from licensed or open software	remote	N/A	N/A
S4	obtained at step S3	Comparison among different rainfall erosivity maps (current period) remapped on the EURO-CORDEX grid	uncertainty analysis (not data, keep locally)	licensed software such as Matlab/Arcgis		local	N/A	N/A
S5	D1 to D6; D8	Bias correction at the monthly level using different aggregated datasets as reference conditions on	Precipitation maps for future horizons (y)	N/A	source (pseudo-)code from licensed or open	remote	N/A	N/A

		remapped grids			software			
S6	D7 and D9	Bias correction at a monthly level using VHR-Rea on its native grid	Precipitation maps for future horizons (y)	N/A	source (pseudo-)code from licensed or open software	remote	N/A	N/A
S7	obtained at steps S5 and S6 (and D13)	Application of models relying on different projections on their native grids	rainfall erosivity maps for future horizons (y)	N/A	source (pseudo-)code from licensed or open software	remote	N/A	N/A
S8	obtained at step S7	Comparison (for future horizons) between rainfall erosivity map based on EURO-CORDEX and VHR-Pro projections	uncertainty analysis (not data, keep locally)	licensed software such as Matlab/Arcgis		local	N/A	N/A
S9	D10, D11	Identification of proper land classification comparing existing products with project results from AI use	Land use/cover map (y)	licensed software such as Matlab/Arcgis		local	N/A	N/A
S10	D10, D11, D16, outcomes of S9	Calculation of (dynamic) soil susceptibility C-factor for current/recent/ future period	C-factor maps for recent/current period, possibly accounting for seasonality (y)	N/A	source (pseudo-)code from licensed or open software	remote	N/A	N/A
S11	D12	Calculation of soil susceptibility K-factor for current/recent period	K-factor maps for recent/current period (y)	N/A	source (pseudo-)code from licensed or open software	remote	N/A	N/A
S12	D13, D14	Calculation of soil susceptibility LS-factor for	LS-factor maps for	N/A	source (pseudo-)code	remote	N/A	N/A

		current/recent period	recent/current period (y)		from licensed or open software			
S13	Outcomes of S3, S7, S10, S11, S12, D15	Calculation of soil loss for all time horizons	Soil Loss maps for recent/current/future period (y)	N/A	source (pseudo-)code from licensed or open software	remote	N/A	N/A

7.2.4 Results

Table 24: Use case 5 results specifications

Result (Rx)	Result description	Result delivery	Infrastructure requirements / resource needs (development)	Infrastructure requirements / resource needs (operational)
R1	(dynamic) Rainfall erosivity maps building on ERA5, ERA5-Land, CERRA, E-OBS, VHR-Rea for current period	web GIS (visualization), direct download	N/A	CPU, ≤ 1GB
R2	(dynamic) Rainfall erosivity maps building on bias-corrected (possibly new) EURO-CORDEX and VHR-Pro projections for future (short to long-term, TDB) horizons	web GIS (visualization), direct download	N/A	CPU, ≤ 100GB
R3	(dynamic) maps of soil susceptibility components based on D15, S10, S11, S12 for current period	web GIS (visualization), direct download	N/A	CPU, ≤ 2GB
R4	Soil susceptibility factors maps for future periods (month-to-season aggregation)	web GIS (visualization), direct download	N/A	CPU, ≤ 2GB
R5	Soil Loss maps building on the overlap of susceptibility factors and erosivity with different options for both (based on input datasets and processing alternatives) (in the past/current period)	web GIS (visualization), direct download	N/A	CPU, ≤ 2GB
R6	Soil Loss Maps building on the overlap of susceptibility factors and erosivity with different options for both (based on input datasets and processing alternatives) in the future (short to long-term, TDB) horizons	web GIS (visualization), direct download	N/A	CPU, ≤ 50GB
R7	Maps of anomalies of rainfall erosivity for future (short to long term, TBD) horizons	web GIS (visualization),	N/A	CPU, ≤ 100GB

Result (Rx)	Result description	Result delivery	Infrastructure requirements / resource needs (development)	Infrastructure requirements / resource needs (operational)
		direct download		
R8	Maps of anomalies of soil loss for future (short to long term, TBD) horizons	web GIS (visualization), direct download	N/A	CPU, ≤ 50GB

7.3 Data management needs

The data management needs of UC6 are visualized tabularly in Table 21, Table 22, Table 23 and Table 24.

Table 21 covers the input data needs of U5. The first datasets (D1 to D3) refer to local rainfall observations, which are made available by the local authorities in pdf format (D3) and, for several rain gauges, data have already been digitized and are available as spreadsheets as raw (D1) or post-processed by CMCC (D2) values. Many spatialized precipitation datasets are also considered, including reanalyses (ERA5, ERA5-Land, CERRA, VHR-Rea), interpolation products (E-OBS), and projections (EURO-CORDEX, VHR-Pro). ERA5 and ERA5-Land are both available with high spatial and temporal coverage (global and 1959-now – 1950 for ERA5 – respectively) with the hourly resolution, only differing for the horizontal resolution, coarser for ERA5. CERRA, instead, is only available with a pan-European coverage and for a narrower time window (1984-2021): it is characterized by two possible horizontal resolutions of 5.5 and 10 km, depending on the chosen version, and two alternative time paces (3 or 6 hours). E-OBS is the only interpolation product considered, as it is available at the daily scale (i.e., a coarser time resolution concerning the reanalyses) with a pan-European coverage and two different horizontal resolutions according to the chosen version. All these precipitation products are made available by the Climate Data Store of C3S, where they can be all found in NetCDF format and, for some of them, also as grib files. One additional precipitation product is VHR-Rea, which is a downscaling of ERA5 at a horizontal resolution of about 2 km and hourly time step, covering the time window 1981-2020 and limited to Italy. This dataset is made available by the Data Delivery System of CMCC. As concerns projections, D9 includes a relevant number of EURO-CORDEX climate simulations (from 1971 to 2100), all with the same temporal (native hourly) and horizontal (about 12 km) resolution, with a pan-European coverage. Each projection is available for different concentration scenarios, including RCP 2.6, 4.5 and 8.5. It is still to be defined which specific EURO-CORDEX members will be adopted among the several dozen available. All EURO-CORDEX simulations are available as NetCDF files in the CDS of C3S. VHR-Pro is a downscaling of the Italy_8 climate projection by CMCC to the resolution of about 2 km with an hourly time step, made available by the DDS of CMCC as NetCDF files. Land-related data (D10) are available with different horizontal resolutions (but significantly higher than precipitation products) and temporal coverage, as they usually consist of "instantaneous" pictures released every bunch of years starting from about ten years ago. The dataset contains different fields, that can be used as an alternative to the land classification product that is going to be produced by the Project. Potentially, D10 includes Land Cover (LC) and World Cover (WC), both with global coverage and High-Resolution Layers (HRL) with pan-European coverage. LC and HRL are made available by CLMS and WC by ESA, all as GeoTIFF files. Also, as a potential land cover alternative, D11 represents the IV level of CORINE Land Cover, made available for Italy by a national authority (SINANet). D12 potentially includes pH, soil organic carbon content, bulk density, coarse fragments, sand, silt, clay content, cation exchange capacity (CEC), total nitrogen, soil organic carbon density, soil organic carbon stock, and total available water content. Part of these variables is released by ISRIC (SOILGRIDS initiative) with global coverage, as GeoTIFF files others are released by JRC (ESDB initiative) with a pan-European coverage as ESRI grid or vector files.

D13 and D14 are ancillary datasets providing elevation information (D14, released by CLMS as a GeoTIFF file with a horizontal resolution of 25 m with a pan-European coverage) and linear infrastructures (D15, freely available as a shapefile covering Italy). D15 provides the RUSLE P-factor, released by JRC as a GeoTIFF file, and finally, D16 provides vegetation products (in particular LAI and NDVI) allowing for the enhancement of the RUSLE C-factor estimation.

In terms of pre-processing (Table 22), most of the envisaged operations concern homogenization: first, precipitation units from all the listed datasets will be converted in mm, and all the temporal resolutions will be aggregated to daily (if the raw resolution is hourly) and monthly. Also, EURO-CORDEX irregular grids will be converted into regular longitude/latitude coordinates. Finally, GeoTIFFs will be converted into NetCDF files. The storage resources listed in Table 21 refer to the datasets after clipping to the domain and time windows of interest.

All the input datasets will be manipulated and processed (Table 23) to obtain relevant results which are described in Table 24. Results mainly consist of dynamic (in space and, for some variables, in time) maps of some of the parameters of the RUSLE equation (R-factor, C-factor, LS-factor, K-factor) and the final results in terms of Soil Loss. From a perspective of climate change, results for R-factor and Soil Loss will be also delivered in terms of expected variations for relevant future horizons.

7.4 Software component needs

Table 23 shows the main requirements in terms of processing and software components. In general terms, all the operations needed are performed locally (for testing purposes) using licensed software such as ArcGIS and MatLab, or open software such as CDO. To feed the platform for computations, codes will be provided written in the language of the mentioned software.

In UC5, AI/ML tools will be used for two main purposes. On one hand, the workflow will leverage the enhanced land cover map that will be produced by the Project making use of self-supervised classification (S9). This novel map can be used to obtain a better estimate of the C-factor, which is the RUSLE parameter that describes the different types of land use/land cover. Attempts will be also made, in this sense, to account for the seasonality of specific land covers (e.g., crops) and year-by-year changes, where possible, to obtain dynamic C-factor maps.

On the other hand, AI/ML can be used to temporally downscale precipitation products from the hourly (or daily, according to the specific product) to the sub-hourly (at least 30 min) scale (S1). This will allow for the estimation of the R-factor at the event scale by applying the canonical equation instead of simplified approaches relying on coarser aggregations. In turn, this enables not only a more reliable estimation of canonical annual and monthly R-factors but also an unprecedented dynamical mapping of rainfall erosivity that can be used, for example, for immediate post-event evaluations. Disaggregated rainfall can be also of more general use for different applications, and useful to store in the Platform to satisfy possible future user needs in terms of EO (although post-processed).

The downscaling of precipitation can be obtained by leveraging sub-hourly precipitation observations collected at a relevant number of rain gauges over Italy (D1 and D3 in Table 21), which allows for the calibration of the process to be applied to the selected spatially distributed precipitation products (D4 to D7 in Table 21). Different tools can be applied to statistically downscale precipitation information, such as Genetic Programming, Artificial Neural Network, Support Vector Machine or Relevance Vector Machine (with the last two particularly indicated for analyses focusing on high flows). Using an ensemble of different techniques can be useful to compare performance in terms of under/overestimation of specific percentiles (the highest percentiles in particular) as well as the effect of elevation and climatology.

7.5 Computational infrastructure needs

Virtual resources are needed not only to store Use Case results but also to allow their dynamic exploration, giving the possibility to the users of selecting different areas and time windows.

All the operations described in the workflow will be performed in Italy. The scalability in space is however possible but some additional elements are needed. As concerns the enhancement of the C-factor, the improved land cover map will be provided by the Project for the whole of Europe: its conversion in a C-factor map will be only performed for Italy, however, results are exportable to the pan-European domain for all those land cover classes in Italy that are also present in the rest of Europe; for those missing in Italy, additional analysis has to be done to associate C-factor values.

As concerns the enhancement of the LS-factor, although the computation will be limited to Italy, the EU-DEM will be used, so the overall framework is directly applicable to the pan-European domain. Instead, for the role of linear infrastructure as a possible obstacle to flow, OpenStreetMap will be used which is limited to Italy. However, the framework is transferrable nonetheless, as long as some vectorial information about roads and railways is available.

For the enhancement of the R-factor, the downscaling of precipitation is only possible where sub-hourly rainfall observations are available. This information is collected for Italy in the Project: however, the general framework can be potentially applied to Europe as long as a sufficient number of rain gauges with suitable data (i.e., with sufficient temporal resolution and coverage) is available.

7.6 Deployment plan

To be provided at a later stage in the document version B.

8 Use Case VI – Food

8.1 Introduction

UC6, led by SISTEMA, with a contribution from CMCC, aims at providing information services for assessing and predicting the impact of locust plagues. The service combines Earth Observation and climate data using AI and machine learning techniques to improve the reliability and effectiveness of the monitoring and prediction service. The use case aims to improve the information service by introducing new climate models for prediction and using high-performance computing infrastructure for computational aspects. Desert locust plagues are strongly influenced by climate conditions and vegetation status. Previous studies have separately analysed Earth Observation techniques and climate predictions, but this service combines both to provide a more comprehensive approach.

8.2 Workflow

The scope of this chapter is to introduce and describe the workflow of the use case VI using tables. The chapter contains four dedicated sub-chapters, covering the input data needs, pre-processing and processing steps and the generated results.

8.2.1 Input data

Table 25: Use case 6 input data specifications

Input data (Dx)	Data	Geographic extent	Spatial resolution	Time range	Data source	Data format	Disk size
D1	ERA5 Land Temperature	global	0.1° x 0.1°; Native resolution is 9 km.	1985 - 2021	CDS	nc	~3TB
D2	ERA5 Land Precipitation	global	0.1° x 0.1°; Native resolution is 9 km.	1985 - 2021	CDS	nc	~3TB
D3	ERA5 Land Volumetric soil water layer 1	global	0.1° x 0.1°; Native resolution is 9 km.	1985 - 2021	CDS	nc	~3TB
D4	ERA5 Land LAI (leaf area index)	global	0.1° x 0.1°; Native resolution is 9 km.	1985 - 2021	CDS	nc	~3TB
D5	FAO locust location reference swarms	global	n/a	1985 - 2021	FAO	gjson	~10MB
D6	FAO locust location reference hoppers	global	n/a	1985 - 2021	FAO	gjson	~10MB
D7	Short-term weather forecast	global	0.1° x 0.1°	0+10 days	ECMWF	grib	~12GB
D8	Seasonal weather forecast	global	1° x 1° or better (depending on the model)	0+6 months	ECMWF	grib	~200GB

8.2.2 Pre-processing

Table 26: Use case 6 pre-processing steps specifications

Pre-processing (PPx)	Input Data	Step description	Outcome (preservation y/n)	Processing tool (open source or commercial)	Processing interface	Facility	Infrastructure requirements / resource needs (development)	Infrastructure requirements / resource needs (operational)
PP1	D1	conversion to tiff; conversion to daily avg.	tiff (y)	Python (o)	CLI / API	possibly remote	4 CPU/16GB RAM / 500 GB Storage	16 CPU/64GB RAM / 2TB Storage
PP2	D2	conversion to tiff; conversion to	tiff (y)	Python (o)	CLI / API	possibly remote	4 CPU/16GB RAM / 500 GB Storage	16 CPU/64GB RAM / 2TB Storage

Pre-processing (PPx)	Input Data	Step description	Outcome (preservation y/n)	Processing tool (open source or commercial)	Processing interface	Facility	Infrastructure requirements / resource needs (development)	Infrastructure requirements / resource needs (operational)
		daily avg.						
PP3	D3	conversion to tiff; conversion to daily avg.	tiff (y)	Python (o)	CLI / API	possibly remote	4 CPU/16GB RAM / 500 GB Storage	16 CPU/64GB RAM / 2TB Storage
PP4	D4	conversion to tiff; conversion to daily avg.	tiff (y)	Python (o)	CLI / API	possibly remote	4 CPU/16GB RAM	16 CPU/64GB RAM
PP5	D5	image rasterization	tiff (y)	Python (o)	CLI / API	possibly remote	4 CPU/16GB RAM / 500 GB Storage	16 CPU/64GB RAM / 2TB Storage
PP6	D6	image rasterization	tiff (y)	Python (o)	CLI / API	possibly remote	4 CPU/16GB RAM / 500 GB Storage	16 CPU/64GB RAM / 2TB Storage
PP7	D7	conversion to tiff	tiff (y)	Python (o)	CLI / API	possibly remote	4 CPU/16GB RAM / 500 GB Storage	16 CPU/64GB RAM / 2TB Storage
PP8	D8	conversion to tiff	tiff (y)	Python (o)	CLI / API	possibly remote	4 CPU/16GB RAM / 500 GB Storage	16 CPU/64GB RAM / 2TB Storage

8.2.3 Processing

Table 27: Use case 6 processing steps specifications

Processing step (Sx)	Input Data	Step description	Outcome (preservation y/n)	Processing tool (open source or commercial)	Processing interface	Facility	Infrastructure requirements / resource needs (development)	Infrastructure requirements / resource needs (operational)
S1	D1, D2, D3, D4	four-dimensional raster file	tiff (y)	gdal library	script	remote (e.g., CINECA, if	4 CPU/16GB RAM / 500 GB Storage	16 CPU/64GB RAM / 2TB Storage

Processing step (Sx)	Input Data	Step description	Outcome (preservation y/n)	Processing tool (open source or commercial)	Processing interface	Facility	Infrastructure requirements / resource needs (development)	Infrastructure requirements / resource needs (operational)
						possible)		
S2	all	trained model for image segmentation and forecasting	y	gdal library; fast AI Python library	Jupyter notebook	remote (e.g., CINECA, if possible)	4 CPU/16 GPU / 16GB RAM / 500 GB Storage	16 CPU/64GB RAM / 2TB Storage
S3	D7, D8	model efficiency	y	gdal library; fast AI Python library	Jupyter notebook	remote (e.g., CINECA, if possible)	4 CPU/16 GPU / 16GB RAM / 500 GB Storage	16 CPU/64GB RAM / 2TB Storage
S4	D7, D8	application	y	gdal library; fast AI Python library	script, CL	remote (e.g., CINECA, if possible)	4 CPU/16GB RAM / 500 GB Storage	16 CPU/64GB RAM / 2TB Storage

8.2.4 Results

Table 28: Use case 6 results specifications

Result (Rx)	Result description	Result delivery	Infrastructure requirements / resource needs (development)	Infrastructure requirements / resource needs (operational)
R1	risk map, layer / mask showing estimation of the probability of appearance; binary output	web GIS (visualization) / direct download	4 CPU/16GB RAM / 500 GB Storage	16 CPU/64GB RAM / 2TB Storage
R2	information about the reliability of the model output	web GIS (visualization) / direct download	4 CPU/16GB RAM / 500 GB Storage	16 CPU/64GB RAM / 2TB Storage
R3	the density of the estimated appearance of locusts	web GIS (visualization) / direct download	4 CPU/16GB RAM / 500 GB Storage	16 CPU/64GB RAM / 2TB Storage

8.3 Data management needs

The data management needs of UC6 are visualized tabularly in Table 25Table 25: Use case 6 input data specifications, Table 26, Table 27 and Table 28.

Table 25 covers the input data needs. The table provides information on the eight input datasets relevant to UC6. The datasets include ERA5 Land temperature, precipitation, volumetric soil water layer 1, and leaf area index, which are retrieved from the CDS in nc file format. The native resolution of these datasets is 9 km, with global coverage from 1985 to 2021. The reference data related to locust location swarms and hoppers are provided by FAO in geoJSON file format, with a global extent and coverage from 1985 to 2021. The input data also includes global short-term and seasonal weather forecasts covering the years 1985 to 2021, which are retrieved from ECMWF in grib file format. The resolution of the short-term forecasts is 0.1° x 0.1° and a coverage range of ten days in the future. The resolution of the seasonal forecasts is 1° x 1° or better, depending on the model, providing information on weather patterns over a longer time range, up to six months. The FAO reference data have a size of 7.1 MB. The size of the ERA5 Land model data is 12TB in sum, whereas the weather forecast data uses up to 220GB in disk space.

Pre-processing of the input data includes the conversion of ERA5 Land data to tiff file format and the conversion to daily averages. FAO reference data is rasterized, and the weather forecasts are also converted to tiff file format. The pre-processing of each input dataset is visualized in Table 26.

Table 28 gives an overview of the outcome of the process implemented in UC6. The results of the use case are threefold: binary risk maps mask showing the estimation of the probability of appearance are generated (R1). The output also includes information about the reliability of the model output (R2), as well as the density of the estimated appearance of locusts (R3). All results shall be made available through a web GIS for visualization and made available for direct download.

8.4 Software component needs

The software component needs of UC6 correspond to the processing steps visible in Table 27 making use of open-source tools. First, a four-dimensional raster file of the ERA5 Land data is generated (S1). All input datasets are utilized to generate a trained model for image segmentation and forecasting (S2). To verify the model efficiency, the FAO reference datasets are used (S3). Step four (S4) includes the compiling of the application, which makes use of the weather forecast data from ECMWF. The AI component of this workflow is based on the fast AI Python library. The steps are executed through either scripts or a Jupyter notebook interface.

The objective of the AI for UC6 is to determine the breeding location of the locust from environmental information obtained with ERA5. This objective involves multiple dimensions concerning the data. Firstly, a multi-dimensional aspect concerning environmental factors. It is necessary to combine several elements to explain the appearance of the locust. Additionally, a temporal dimension must be added to explain the evolution and development of the locust in a specific area. In this way, the dataset for the AI model is configured, as shown in S1 Table 27, as a temporal series of multi-dimensional raster data containing all environmental factors to explain locust breeding.

This specific configuration is to be used in a type of model named Recurrent Neural Network (RNN) and, more precisely, Long-Short Term Memory (LSTM). This architecture is made to understand the temporal structure/trend of the data and extract the most meaningful information across all dimensions to answer the objective. The modelling is presented in Figure 2.

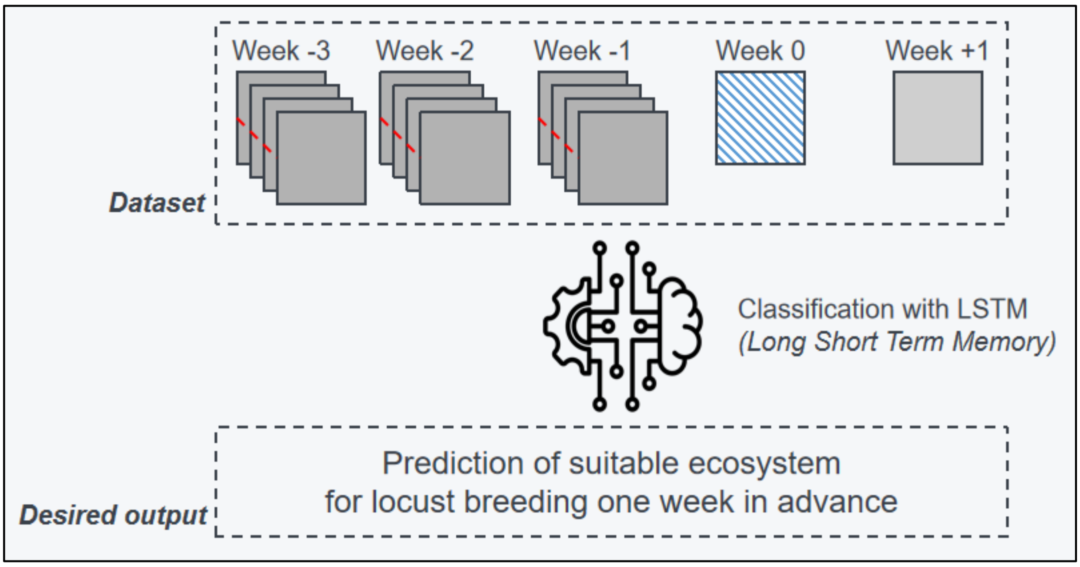


Figure 2. UC6 LSTM modelling

As previously mentioned, the dataset is in a multi-dimensional structure and can be analysed with models such as LSTM. The desired output to answer the objective is a prediction of a suitable ecosystem for locust breeding. This output can be seen as a segmentation problem, where the different classes for the images correspond to a level of confidence for breeding. For each location provided by the FAO, a distance threshold can be performed, corresponding to a level of confidence in the presence of a locust. During the training, the AI model will search for the most suitable area.

The scope of using AI here is to prevent the appearance of locusts by defining "risk" areas where they can breed. The advantage of AI is that it can handle a large amount of diverse information to retrieve similarities and trends to provide the best solution. Nevertheless, some limitations can be identified. The first one remains on the creation of the dataset itself. The information used to perform the analysis comes from various sources, including earth observation modelling data (ERA5), manual verification of locust presence on the ground, and seasonal forecast. Moreover, these different data are structured differently in raster/vector formats. Therefore, it is necessary to identify how to structure them properly to make it easier for the machine learning part. Finally, the complexity of the task is added, as the final objective is to identify a suitable location for insect breeding based on data at several kilometres of spatial resolution, involving multiple dimensions of analysis.

8.5 Computational infrastructure needs

Table 26, Table 27, and Table 28 include the computational infrastructure needs of UC6, covering pre-processing steps as well as processing and results delivery and visualization. From the elaboration of the input data (see 8.3), a total disk size of 12.5TB is requested. Pre-processing steps of UC6 are implemented using Python through CLI or API, possibly in a remote environment. The processing steps as seen in Table 27 require HDD space, GPU and GPU or CPU (S4). The requirements remain the same for the use case development purposes and the use case operation. For the delivery of the results, the requirements are as follows: 4 CPUs; 16 GB RAM; 1 TB storage.

For development purposes, the use case is applied to Sub-Saharan Africa. For operations, it can be scaled up to other locations in the Asian and African continents. To achieve scalability, the following points need to be taken into consideration:

- resources
- model re-training
    - potential model re-training over Asia

8.6 Deployment plan

To be provided at a later stage in the document version B.

9 Use Case VII – Fires

9.1 Introduction

UC7 is led by the Intelligence for Environment & Security (IES), with the support of the Center for Security Studies (Kentro Meleton Asfaleias, KEMEA), the Fraunhofer Institute for Transportation and Infrastructure Systems (IVI) and NKUA. This use case will improve the use and exploitation of European Union-observed datasets in Civil Protection operations. The goal is to improve Civil Protection activities by providing timely, targeted reactions to dynamic events such as earthquakes, forest fires, landslides, and floods. The use case aims to deliver updated imagery with automatically detected changes caused by these events, using Earth Observation for Europe services to discover and present relevant satellite-derived information to end users.

9.2 Workflow

The scope of this chapter is to introduce and describe the workflow of use case VII using tables. The chapter contains four dedicated sub-chapters, covering the input data needs, pre-processing and processing steps and the generated results.

9.2.1 Input data

The input data listed below are needed by EO4EU services to process the API calls from the VAA. Use case VII expects EO4EU services will make use of them.

Table 29. Use case 7 input data specifications

Input data (Dx)	Data	Geographic extent	Spatial resolution	Time range	Data source	Data format	Disk size
D1	RGB NIR (preferably Near Red and Green channels) CAT1	Sicily/Europe	3m x 3m	daily	Planet.com, (8-band) Planetscope	GeoTIFF, sentinel-2 harmonised	2 ~ 20 GB
D2	VNIR + SWIR	Sicily/Europe	20m x 20m	5 days	Sentinel-2, level 2A	JPEG2000	1 ~ 3 GB
D3	ENDVI10_metop LST	Europe	1km x 1km	10 days	Eumetsat LSA-SAF	ENVI (img, hdr)	to be defined
D4	ENDVI10_metop NDVI	Europe	1km x 1km	10 days	Eumetsat LSA-SAF	ENVI (img, hdr)	to be defined
D5	ENDVI10_metop NIR SWIR	Europe	1km x 1km	10 days	Eumetsat LSA-SAF	ENVI (img, hdr)	to be defined
D6	Eumetsat LSA-SAF FRMv2	global	3.3 km x 3.3 km	daily	Eumetsat LSA-SAF	HDF5	to be defined
D7	Corine Land Cover (CLC)	Sicily/Europe	100m	upd 2018 (IV)	Copernicus CLMS	SQLite Database	3,5 GB
D8	Static regional vegetation fire danger	regional	20m	seasonal	Sicily Regional Forest Corps	TIFF	to be defined
D9	Administrative boundaries	regional		upd 2022	ISTAT Italy Statistical Institute	shp	few Mbs
D10	FIRMS Fire HotSpots	global	1km x 1km (and/or) 0.375km x 0.375km	2000-to-present	NASA FIRMS-LANCE	shp, csv	to be defined

Input data (Dx)	Data	Geographic extent	Spatial resolution	Time range	Data source	Data format	Disk size
D11	archived ENDVI10_metop NDVI	Europe	1km x 1km	10 years	Eumetsat LSA-SAF	ENVI (img, hdr)	to be defined
D12	archived ENDVI10_metop NIR SWIR	Europe	1km x 1km	10 years	Eumetsat LSA-SAF	ENVI (img, hdr)	to be defined

9.2.2 Pre-processing

Not applicable in this use case, as the system will pass a CAP message to the EO4EU services.

9.2.3 Processing

The Data Processing will be performed within the EO4EU services.
On the base of the CAP category, CAP coordinates and interpreting the CAP text the EO4EU services will identify the correct location and type of needed processing. This processing should be performed daily.

Table 30. Use case 7 Processing steps specifications

Processing step (Sx)	Input Data	Step description	Outcome (preservation y/n)	Processing tool (open source or commercial)	Processing interface	Facility	Infrastructure requirements / resource needs (development)	Infrastructure requirements / resource needs (operational)
S1	D9, D10	Compute NDVI 10-year average. Compute NDWI from Metop channel 3-4 reflectance values. Compute NDVI 10-year average. This may be done once daily.	Yes, data may be stored for further re-use or forensic data checks.	Many tools can be used for processing NDVI, such as Matlab, ENVI, Microsoft Excel or a dedicated script	Not Applicable	EO4EU cloud infrastructure	4 CPUs / 64 RAM	to be defined
S2	D1, D2, D3, D5, D6, D9, D10	rate based on current NDVI, LST and NDWI values and % differences between current and long-term averaged values. This	Yes, data may be stored for further re-use or forensic data checks.	Script	Not Applicable	EO4EU cloud infrastructure	4 CPUs / 64 RAM	to be defined

Processing step (Sx)	Input Data	Step description	Outcome (preservation y/n)	Processing tool (open source or commercial)	Processing interface	Facility	Infrastructure requirements / resource needs (development)	Infrastructure requirements / resource needs (operational)
		may be done once daily.						
S3	D4	FWI risk rate (+24h, +48h, +72h) This may be done once daily.	Yes, data may be stored for further re-use or forensic data checks.	Script	Not Applicable	EO4EU cloud infrastructure	4 CPUs / 64 RAM	to be defined
S4	all	merge of the dynamic vegetation and meteorological danger rate at the Municipality level This may be done once daily.	Yes, data may be stored for further re-use or forensic data checks.	Script	Not Applicable	EO4EU cloud infrastructure	4 CPUs / 64 RAM	to be defined
S5	all	multitemporal change detection on very high-resolution data (D1 & D2). This may be done once daily.	Yes, data may be stored for further re-use or forensic data checks.	Script	Not Applicable	EO4EU cloud infrastructure	4 CPUs / 64 RAM	to be defined

9.2.4 Results

Results are the direct outcomes of each EO4EU API call.

Table 31. Use case 7 results specifications

Result (Rx)	Result description	Result delivery	Infrastructure requirements / resource needs (development)	Infrastructure requirements / resource needs (operational)
R1	Daily fire risk map: 1 Region, 7 Provinces, 391 Municipalities	direct delivery to VAA (possible: WMS, WFS, formatted JSON, etc.) Results may be persistent on the	Delivery of the results will not imply specific infrastructure requirements unless EO4EU requires data persistence to exploit data visualisation on different software components.	Delivery of the results will not imply specific infrastructure requirements unless EO4EU requires data persistence to exploit data visualisation on different software components.

Result (Rx)	Result description	Result delivery	Infrastructure requirements / resource needs (development)	Infrastructure requirements / resource needs (operational)
		EO4EU services as well as on the VAA.		
R2	Detected changes in the target area according to the CAP event category. This will include images and pixel count of detected changes (and, hence, an estimation of the area changed)	direct delivery to VAA (possible: WMS, WFS, formatted JSON, etc.) Results may be persistent on the EO4EU services as well as on the VAA.	Delivery of the results will not imply specific infrastructure requirements unless EO4EU requires data persistence to exploit data visualisation on different software components.	Delivery of the results will not imply specific infrastructure requirements unless EO4EU requires data persistence to exploit data visualisation on different software components.

9.3 Data management needs

The Common Alerting Protocol (CAP) provides an open, non-proprietary digital message format for all types of alerts and notifications. The CAP format is used by crisis responders, weather prediction agencies and emergency management stakeholders for broadcasting data about crisis conditions and locations over various media, including SMS, police radio and weather radio[1]., while offering enhanced capabilities such as:

- Flexible geographic targeting using latitude/longitude shapes and other geospatial representations in three dimensions;
- Multilingual and multi-audience messaging;
- Staggered and delayed effective times and expirations;
- Enhanced message update and cancellation capabilities;
- Template support for framing complete messages and effective warning messages;
- Compatibility with digital signature capability;
- Digital image and audio capability.

In Europe, CAP is used by various organisations and systems involved in emergency management and public safety:

- European Emergency Number Association (EENA): EENA promotes the implementation and use of CAP across European emergency services. They work towards enhancing the effectiveness of emergency response by facilitating the exchange of information through CAP.
- National Emergency Management Agencies: many European countries have their national emergency management agencies or civil protection agencies that utilise CAP for issuing alerts and warnings to the public. These agencies may include the respective national disaster management organisations or ministries responsible for emergency response.
- Weather Forecasting and Meteorological Agencies: meteorological agencies across Europe, such as Meteo France, Met Office (UK), Deutscher Wetterdienst (Germany), and others, use CAP to disseminate weather warnings, severe weather alerts, and forecasts to the public and emergency responders.
- Public Warning Systems: several countries in Europe have implemented public warning systems that utilise CAP for delivering alerts to the population during emergencies. These systems can include sirens, mobile applications, emergency broadcasting networks, and other notification platforms.

In use case VII, the user belonging to a civil protection or first responder organisation, may start using its CAP-based[2] command and control system (referred to hereafter as Value Added

[1] https://joinup.ec.europa.eu/collection/rolling-plan-ict-standardisation/emergency-communication-and-public-warning-systems

[2] http://docs.oasis-open.org/emergency/cap/v1.2/CAP-v1.2-os.html

Application, VAA in the following[3]) and create an event that deals with forest fire (step 0 in the picture below).

In Italy CAP is used by the Italian Civil Protection and Fire Brigades corps, as well as by some regional civil protection agencies, such as the one of Sicily Region.

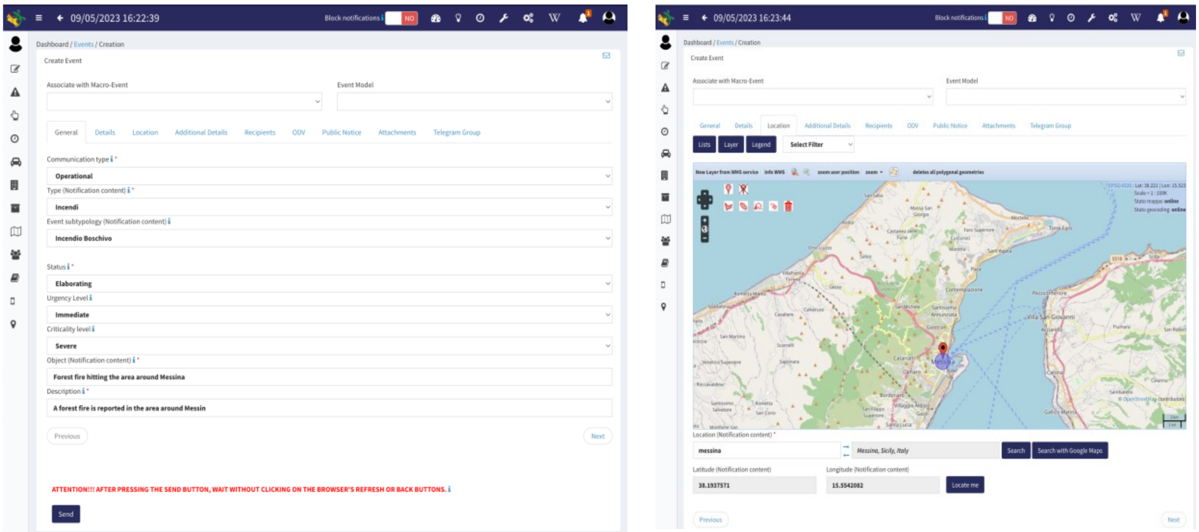


Figure 3. Step 0 of the CAP based command and control system

After the event creation, the command-and-control system will be used to make a call to the EO4EU API. (step 1 in the picture below).

[3] The VAA for use case VII based on Jixel - www.jixel.eu

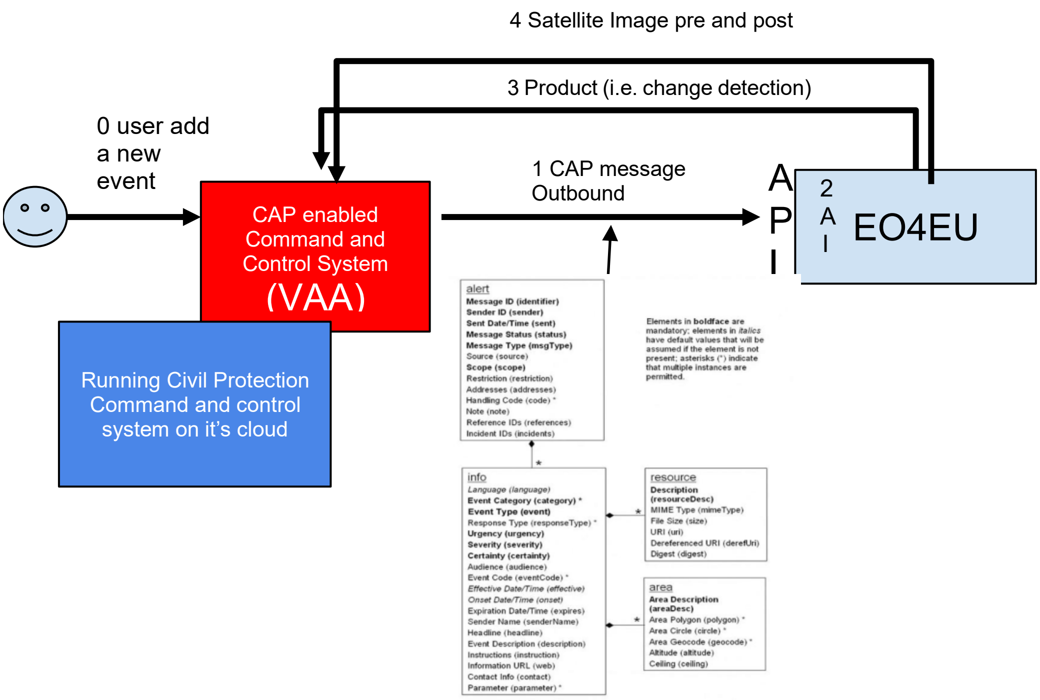


Figure 4: Command and control system overview

The list of information carried by the CAP message for a standard event is given below:

- identifier
- sender (platform)
- datetime (format Y-m-d; h:m:s)
- status
- message Type
- organisation (sender)
- scope
- public/private
- info
- category
- event name
- language
- urgency
- severity
- certainty
- event type
- event su-type

- effective datetime (format Y-m-d; h:m:s)
- instructions
- contact (caller Name)

Alternatively, the user starts from the front end of the EO4EU platform and accesses the VAA through an SSO service. An iFrame of a JIXEL instance within the EO4EU front-end will allow Users to use JIXEL within the EO4EU platform. The VAA component allows users to create an event and, as already discussed above, define the type of event linked to scenarios for which EO data and EO "products" will be retrieved based on the use cases identified by the project (e.g., forest fire, earthquake, floods and flash floods, etc.). In this sense, the VAA, in the red box in the above flow-chart, will be installed on the EO4EU platform **OR**, in case an existing service such as the one used by Sicilian Civil Protection (blue box in the above flow-chart), will remain deployed on its home cloud. For this second case, no SSO functionality is foreseen.

From the end-user perspective, this equals the possibility of defining a new event – belonging to a CAP event category – as follows:

cap. alertInfo. category. code	The code denoting the category of the subject event of the alert message (REQUIRED)	(1) Code Values: "Geo" - Geophysical (inc. landslide) "Met" - Meteorological (inc. flood) "Safety" - General emergency and public safety "Security" - Law enforcement, military, homeland and local/private security "Rescue" - Rescue and recovery "Fire" - Fire suppression and rescue "Health" - Medical and public health "Env" - Pollution and other environmental "Transport" - Public and private transportation "Infra" - Utility, telecommunication, other non-transport infrastructure "CBRNE" – Chemical, Biological, Radiological, Nuclear or High-Yield Explosive threat or attack "Other" - Other events

Figure 5: CAP alert information category code

The scenario selection enables a template (approved and based on user requirements and Common Alerting Protocol specifications) that allows the user to specify the scenario details or related to an event, characterising it.

The VAA component collects the scenario/event-related information provided by the user and generates a CAP message which is automatically sent to the EO4EU services (i.e., ML & AI component) via EO4EU APIs (step 2). The EO4EU services will act, in this use case, as a SaaS component, providing the products needed for the end users and hereafter identified. The use case expects, hence, that based on the information provided in the CAP message, the AI and ML solutions of the EO4EU services shall support in identifying, elaborating and returning products (steps 3 and 4) obtained from available input EO data.

This approach will provide a flexible opportunity for EO data used for different first responder organisations that may, this way, exploit EO4EU services in many EU countries, with the only requirement needed being the use of CAP protocol outbound. The VAA component is open to receive automated inputs in CAP and will be upgraded to accept and visualise the products and geographic data provided by the EO4EU services in WFS/WMS and JSON format. The JSON dataset may be sent to other EO4EU components, e.g. XR/VR or Data Analytics visualisation. For EO4EU we'll be using the "GEO", "MET", and "FIRE" categories.

The data management needs of UC7 are visualised tabularly in Table 29, Table 30 and Table 31.

Table 29 covers the input data needs. The table provides information on the twelve input datasets relevant to UC7. The datasets include RGB NIR (preferably Near Red and Green channels) CAT1, VNIR + SWIR, ENDVI10_metop LST, NDVI, NIR SWIR, Eumetsat LSA-SAF FRMv2, Corine Land Cover (CLC), Static regional vegetation fire danger, Administrative boundaries, FIRMS Fire HotSpots, which are retrieved from the Planet.com, (8-band) Planetscope, Sentinel-2, level 2A, Eumetsat LSA-SAF, Copernicus CLMS, Sicily Regional Forest Corps, ISTAT Italy Statistical Institute and NASA FIRMS-LANCE. All the "input" data are needed by the EO4EU services to perform the above-mentioned steps.

The file formats are ENVI (img, hdr), HDF5, TIFF, SHP and CSV. The native resolution of these datasets is variable, from 1km x 1km of Eumetsat LSA-SAF datasets to 100m of Copernicus CLMS Corine Land Cover, to 20m of Sicily Regional Forest Corps static regional vegetation fire danger TIFF; up to 3.3km x 3.3km of Eumetsat LSA-SAF FRMv2. The data relating to LST, NDVI and NIR SWIR are provided by Eumetsat LSA-SAF in ENVI (img, hdr) file format, with a Europe extent and 10 days coverage. The input data also includes global FRMv2 with 3.3km x 3.3km resolution and daily coverage, which is retrieved from Eumetsat LSA-SAF in HDF5 file format. The Corine Land Cover (CLC) resolution is 100m x 100m, providing information on land cover in 44 classes up to 2018.

Table 31 gives an overview of the outcome of the process implemented in UC7. The results of the use case are threefold: risk fire maps showing the estimation of fire probability for three different levels - regional, provincial, and municipal (R1). The output data will include change detection images (R2).

For these categories, the VAA will expect two results from EO4EU services:

1. For R1:
    a. Daily Fire Risk map valid for the next 24 hours

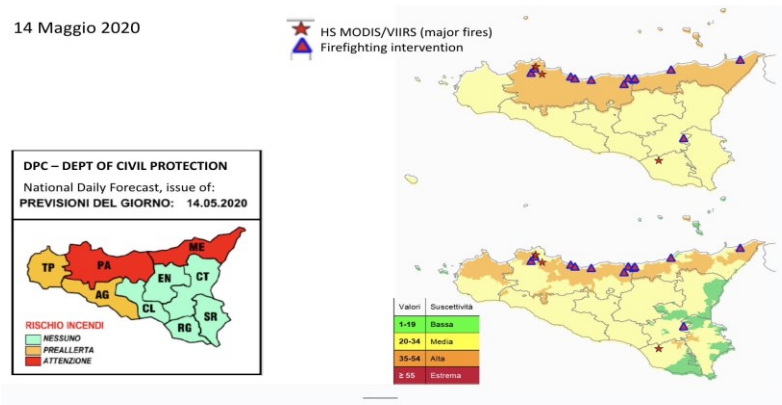


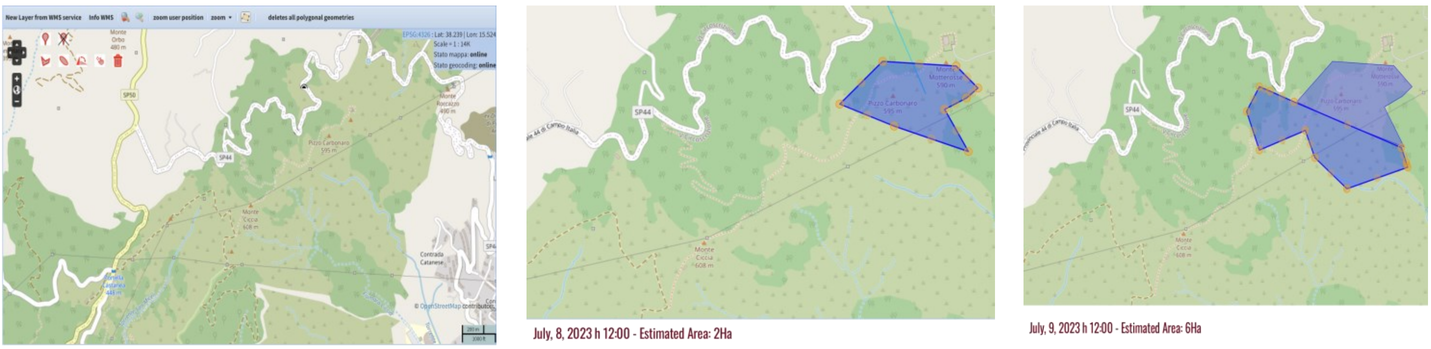
Figure 6: Daily Fire Risk Map

2. For R2:

a. A “product” which, as an example, in the case of FIRE, will be an estimation of the vegetated areas potentially damaged by the ongoing FIRE. The product should have the highest possible fresh rate and space resolution.

b. A “satellite image” of the zone BEFORE the event and the subsequent images as they are made available from EO4EU. Images should have the least possible cloud coverage for allowing a clear vision of the hit area. The images should have the RGB NIR (preferably Near Red and Green channels).

In the following three images, we describe the expected R2 result from the end user's perspective. The first image shows the location of the event (T0). The other two images show the situation as it evolves in time, the day after the event (T1) and two days after it (T2).



(a) (b) (c)

Figure 7: Expected results: (a) Image before the event on July 7th 2023 (T0); (b) Change detection after the event on July 8th 2023 (T1-T0); (c) Change detections after the event on July 9th 2023 (T2-T1)+(T1-T0)

If needed for the specific scenario, geographic data may be returned by the EO4EU services following the scenario/event specification and geographically delimited/restricted by the location (lat. & long.) and buffer specified.

9.4 Software component needs

a. Needs in case of prevention (R1)

The user needs to issue, daily, a geographic fire risk bulletin at a Communal scale, reporting the most prone-to-fire zones based on static and dynamic parameters such as vegetable fuel, hydric stress of vegetation and short-term forecast of relevant meteorological parameters.

The AI component, in this case, should be able to:

- Understand, based on the type of bulletin expected to issue, the relevant data needed;
- Provide the processor with;
- Let the processor ingest data and provide the appropriate output(s);
- Transmit the daily Fire Risk map to VAA

b. Needs in case of response (R2)

The AI should be able to identify, based on the ‘event.category’, the most suitable data and sources of data. This process should be transparent from the end user's perspective.

For instance, should a user select the CAP event.category “**Fire**”, the AI should be able to:

- understand, based on the following CAP information

cap. alertInfo. event. text	The text denoting the type of the subject event of the alert message (REQUIRED)

Figure 8: CAP information

This case requires Very High Resolution (VHR) multispectral images, with pixel footprints <4 metres.

- Identify the most suitable set of data “before the event” and linked to the impact (or likely impacted) area that may include, for instance, the burnt area. The Area identification should be scalable according to the coordinates provided in the CAP message
- process in real-time and detect changes between the post- and pre-event satellite images.

9.5 Computational infrastructure needs

The VAA will be deployed within the EO4EU cloud of services.
Tech-specific requirements for the deployment of the JIXEL Platform, for the development phase, are:

- IT Infrastructure: 4 Linux-based nodes for running Docker containers.
- Jixel is deployed using an orchestrator based on Docker Swarm with the following minimum requirements
- n. 1 for node manager: 2 VCPU, 4GB RAM, 40GB storage
- n. 3 for service nodes: 4 VCPU, 16GB RAM, 160GB storage
- JIXEL is CAP compliant and such protocol will be used for sending messages to the EO4EU platform.
- JIXEL exposes RESTful API for microservices interoperability environment: it accepts output and input standard JSON data format.
- JIXEL is equipped to support Single Sign On authentication via JWT tokens. The flow is already implemented and can be accessed either via RESTful APIs or via browser requests. An authentication server is needed: the user management module could be it, so it will have to store a private key: all frontends will have to share the same public key. A method must be developed to make the JWT token available to the user frontends present in EO4EU Customer Facing Service.
- For SSO support, it is a requirement that the JWT token contains the same UserID as stored in the JIXEL databases.

General requirements:

- Browser compatibility: JIXEL is compatible with browsers such as Chrome, Firefox, Safari and Edge.
- Screen Resolution: To ensure optimal viewing on different devices, JIXEL can adapt to different screen resolutions.
- User Interface: JIXEL provides an intuitive and easy-to-use user interface with elements such as icons, buttons, menus and navigation bars.
- Input device support: JIXEL supports standard input devices such as a mouse and keyboard, as well as touch input devices for mobile devices.
- Accessibility: JIXEL is accessible to users with disabilities, with options to increase the size of text, use a combination of keys to perform actions, and much more.
- Supported languages: JIXEL supports multiple languages, enabling users to use the application in the language they prefer.
- Security: JIXEL guarantees the security of user data by using appropriate security protocols such as HTTPS.
- Develop an API to handle two main functions:
    - Build and issue requests to the EO4EU Platform through our RABBITMQ broker. The request will contain all the fields, organised and filtered from the input user data in compliance with the OASIS Common Alerting Protocol (CAP) standard.
    - Receive, from the RABBITMQ broker, the data structure that will be issued in response by the EO4EU Platform in compliance with the CAP standard after the previous request has been resolved asynchronously.

Table 32: EO4EU platform AI specification

EO4EU platform API specification
a. interoperability, allowing the EO4EU platform to receive requests from other systems already used by end-users
b. API RESTful + JWT authentication
c. Conformance as a CAP V1.2 Message Consumer
d. input: CAP message/output: JSON and/or WMS/WFS where required

For development purposes, the use case is applied to Sicily's regional, provincial and municipal areas. For operations, it may need to be scaled up to other locations in Europe. To achieve scalability, the following points need to be taken into consideration:

- more computational resources
- potential model re-training over Europe

9.6 Deployment plan

To be provided at a later stage in the document version B.

10 Technical alignment

The process of technical alignment involves conducting a thorough evaluation of the deployment plans for a given use case. The objective is to ensure that the computational infrastructure is effectively utilized and occupied in a coordinated manner. Detailed information regarding this alignment will be presented in document version B, along with the deployment plan. The document will encompass the following key aspects:

- Coordinated Exploitation Analysis: This section will provide valuable insights into the development and operational phases of the use case. It will outline the strategies and approaches employed to maximize the potential benefits derived from the computational infrastructure. Additionally, it will highlight any noteworthy considerations related to the use case's implementation and ongoing operation.
- Coordinated Occupation Analysis: This section examines the efficient allocation and utilization of storage and computational resources during the development and operation phases. It assesses how the use case optimally occupies the computational infrastructure, considering its specific requirements and project demands. The analysis highlights strategic resource allocation to support the use case's objectives and operational needs.

11 Concluding remarks

The document titled "Pilot Implementation Methodology and Release of Evaluation Guidelines (a)" offers a comprehensive examination of the management, software component, and computational requirements associated with the seven specific EO4EU use cases. The purpose of this analysis is to facilitate the selection of a data and processing infrastructure provider that aligns with the technical requirements of the project.

The foundation of this document is built upon the use case requirements collected in Annex A, which were obtained during the execution of task T2.1 - Multidisciplinary landscape analysis & Best of Breed Technologies, and delivered as deliverable D2.1 - Research and Innovation Landscape analysis report. Building upon the work accomplished in T2.1, further interaction and exchange took place to enhance and finalize this deliverable. The use case leaders were requested to refine the previously provided information by incorporating more updated data requirements, processing steps, and ICT resource needs. Additionally, as the final version of the document was scheduled for release at month 18, each service provider offered a preliminary definition of the deployment plan.

This version of the document provides a summary of the use case needs that allows providing an initial sizing of the project platform, in terms of computational resources and storage needs. The next and final version of the document will finalise the resources assessment, providing also a clear deployment plan for each service, harmonized to optimise the platform's computational resources.

12 Annex A: Collected use case requirements

Annex A provides links to the spreadsheets containing the collected UC requirements for each of the seven use cases.

- UC1 (Health)
  https://docs.google.com/spreadsheets/d/1tcV_nhp7FD8RIBt52mYl_COtOlBtL1Qo/edit?usp=sharing&ouid=111887094069838382657&rtpof=true&sd=true
- UC2 (Ocean)
  https://docs.google.com/spreadsheets/d/19l9zsMJdf5fx9na40B4cSq_KL_8g77L4/edit?usp=share_link&ouid=111887094069838382657&rtpof=true&sd=true
- UC3 (Crop)
  https://docs.google.com/spreadsheets/d/1B0vd9pX8ANgphYeulLX2-d2t-HXxFXIR/edit?usp=share_link&ouid=111887094069838382657&rtpof=true&sd=true
- UC4 (Forest)
  https://docs.google.com/spreadsheets/d/11Cdrc3rjSHojvquUBC1OP32C_Tt5EKcg/edit?usp=share_link&ouid=111887094069838382657&rtpof=true&sd=true
- UC5 (Soil)
  https://docs.google.com/spreadsheets/d/1CpODYI0w3KZlsiFExHyYTRkoc1INYkKi/edit?usp=share_link&ouid=111887094069838382657&rtpof=true&sd=true
- UC6 (Food)
  https://docs.google.com/spreadsheets/d/1wb7mkKH91xnMDJjOJ1BT0VAtVkCML_8e/edit?usp=share_link&ouid=111887094069838382657&rtpof=true&sd=true
- UC7 (Fires)
  https://docs.google.com/spreadsheets/d/126EIYjF-Da8wm6hanDCsI4UH6EWf4LlG/edit?usp=share_link&ouid=111887094069838382657&rtpof=true&sd=true

13 Annex B: References

None