PANTHEON

Community-Based Smart City Digital Twin Platform

for Optimised DRM operations and Enhanced Community

Disaster Resilience

D4.2

PANTHEON CONCEPTUAL MODELS

The project has received funding from the European Union's Horizon Europe programme under Grant Agreement N°101074008.

DOCUMENT INFO

Deliverable Number	D4.2
Work Package Number & Title	WP4: Design & Development of a Smart City Digital Twin for Community DRM
Lead Beneficiary	SIMAVI
Due date of deliverable	31/08/2024
Deliverable type[1]	R
Dissemination level[2]	PU
Author(s)	Iacob Crucianu (SIMAVI); Otilia Bularca (SIMAVI)
Internal reviewer(s)	Mike Karamousadakis (THL), Cristina Barrado (UPC), Anna Tsabanakis (THL)
Version - Status	**2.0**

TASK ABSTRACT

The deliverable incorporates the results of Task T4.2-Analysis and representation of collected data & Conceptual Models. According with DOA (PANTHEON - Consortia, 2023), "this task investigates methods and approaches for the representation of the data collected in T4.1 to be used by the simulation models and the Machine Learning algorithms adopted for the implementation of the SC Digital Twin framework. The data to be used will be analysed from a statistical point of view, to identify statistical distributions able to represent empirical data and processed to be used by the Machine Learning algorithms and Simulation Models. The main focus of this deliverable will be on Conceptual Models used, considering that the data structures are described in detail in D4.1 Integrated data model for the SCDT.

The conceptual models have been developed based on the outputs of T3.3 and T3.6. Conceptual models are an abstract representation of the systems and modules to be simulated. The development has been guided by the IEEE-1730 DSEEP (Distributed Simulation Engineering and Execution Process) and Agile methodologies since the beginning of the project.

[1] **Please indicate the type of the deliverable using one of the following codes**:
R = Document, report
DEM = Demonstrator, pilot, prototype, plan designs
DEC = Websites, patents filing, press & media actions, videos
DATA = data sets, microdata
DMP = Data Management Plan
ETHICS: Deliverables related to ethics issues.
OTHER: Software, technical diagram, algorithms, models, etc.

[2] **Please indicate the dissemination level using one of the following codes**:
PU = Public
SEN = Sensitive

REVIEW HISTORY

Version	Date	Modifications	Editor(s)
1.0	22/07/2024	First draft / TOC	Iacob Crucianu (SIMAVI); Otilia Bularca (SIMAVI)
1.1	23/08/2024	Models State of the art	Iacob Crucianu (SIMAVI); Otilia Bularca (SIMAVI)
1.2	30/08/2024	Data structures	Iacob Crucianu (SIMAVI); Otilia Bularca (SIMAVI)
1.3	03/09/2024	Models	Iacob Crucianu (SIMAVI); Otilia Bularca (SIMAVI)
1.4	05/09/2024	Tools used to work with models	Iacob Crucianu (SIMAVI); Otilia Bularca (SIMAVI)
1.5	19/09/2024	Input data, recommendations, conclusions	Iacob Crucianu (SIMAVI); Otilia Bularca (SIMAVI)
1.6	24/09/2024	Review remarks	Cristina Barrado (UPC)
1.7	30/09/2024	Review remarks	Mike Karamousadakis (THL)
1.8	1/10/2024	Solve review comments Statistical Models	Iacob Crucianu (SIMAVI); Otilia Bularca (SIMAVI)
1.9	14/10/2024	Review formulas	Ana-Maria Dumitrescu (SIMAVI)
1.10	21/10/2024	Address review comments	Iacob Crucianu (SIMAVI); Otilia Bularca (SIMAVI)
1.11	22/10/2024	Review remarks	Mike Karamousadakis (THL)
1.12	24/10/2024	New approach for conceptual models based on workflows	Iacob Crucianu (SIMAVI); Otilia Bularca (SIMAVI)
1.13	25/10/2024	Added in Annex the ML and deterministic models	Iacob Crucianu (SIMAVI); Otilia Bularca (SIMAVI)
1.14	1.11.2024	Explanations of Use case treatment	Iacob Crucianu (SIMAVI); Otilia Bularca (SIMAVI)
2.0	18/11/2024	Final version	Mike Karamousadakis (THL); Anna Tsabanakis (THL)

DISCLAIMER

Funded by the European Union. Views and opinions expressed are, however, those of the author(s) only and do not necessarily reflect those of the European Union or European Commission. Neither the European Union nor the granting authority can be held responsible for them.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

Abbreviation	Description
AI	Artificial Intelligence
API	Application Programming Interface
CBDRM	Community-based Disaster Risk Management
FR	Functional Requirement
GIS	Geographic Information System
IT	Information Technology
IUC	Intended Use and Classification
DAG	Directed Acyclic Graphs
JSF	Java Server Faces
JMS	Java Messaging System
ML	Machine Learning
NFR	Non-functional Requirement(s)
SCDT	Smart City Digital Twin
UAV	Unmanned Aerial Vehicle

EXECUTIVE SUMMARY

Background	Goals
To enable the definition and usage of Digital Twins, the Conceptual models are analysed and presented in this deliverable. The conceptual models will focus on the management of multi-hazards and critical system interactions, and they will be part of SCDT developments. Systems simulating multi-hazards are very complex, with numerous interacting components. The conceptual model simplifies these complexities by focusing on the key elements and their relationships, allowing us to grasp how the system functions without getting overwhelmed by every detail.	The aim of this Deliverable is to provide the description of the Conceptual Models used, so that the development of individual components can take place in WP4, WP5 and WP6 along with the integration of the platform in WP7. Specifically, the work presented here builds upon the work done in D2.4, D3.1, D3.2, D3.3, D3.4 and D3.6, by further specifying (i) the necessary system requirements to fulfil the user requirements described in D3.2 and D3.6 and (ii) the technical components that participate in the architecture along with the sequence diagrams of their interconnection after analysis on (a) the existing technological landscape (D2.4 & D3.1) (b) the conceptual model of the SCDT (D3.3), (c) the data delivery schemes in D3.4 and (d) the definition of data in D4.1

Approach and course of action
Utilise work achieved in previous work packages and previous tasks: • WP2 PANTHEON Approach for Building Disaster-Resilient Communities • WP3 PANTHEON Requirements, Participatory Design Process and Pilot Use-Cases Specifications First the theoretical background is presented for the models, considering the specific scenarios and use cases. Next the SCDT as the main system approach is described, and the conceptual models are presented.

Findings and results
Four categories of models are described and proposed to be used in PANTHEON: • Wildfire models • Earthquake models • Heatwave models • Man-Made Disaster models

Impact
Conceptual models for multi-hazard management are analysed and the most appropriate for the use cases in PANTHEON are proposed. These models set the groundwork for the technical development of the PANTHEON platform in WP4, WP5, WP6 and WP7.

1. INTRODUCTION

1.1 OVERVIEW & PROPOSED APPROACH

The present deliverable is the main output of task T4.2. As part of this task, methods and approaches for the representation of the data collected in T4.1 are investigated to be used by the simulation models and the Machine Learning (ML) algorithms adopted for the implementation of the Smart City Digital Twin (SCDT) framework. The data to be used for this is analysed from a statistical perspective, to identify statistical distributions able to represent empirical data and processed to be used by the ML algorithms and Simulation Models.

The results from T3.3 Conceptual Model of SCDT for disaster management and T3.6 Project use case specification and scenarios, are taken over and used as input for the development of conceptual models. Conceptual models described in this report are an abstract representation of the systems and modules to be simulated in PANTHEON. The development of the models is guided by the IEEE-1730 DSEEP (Distributed Simulation Engineering and Execution Process) (IEEE, 2022) and Agile methodologies.

The design of the proposed study follows three main steps: (1) statistical data analysis, (2) simulation-based parameter selection, and (3) digital twin modelling.

The first phase includes an in-depth statistical analysis of pre-existing data, identification of key trends and patterns that are relevant to the PANTHEON system. This analysis informs the parameters and distributions that will serve as inputs for the subsequent simulation phase.

Statistical simulation models to replicate the conditions of the system and systematically explore a range of potential parameters are exploited in the second phase. Parameters that best reflect the underlying dynamics of the system, ensuring that they accurately capture both variability and dependencies present in real-world conditions are refined and selected by running these simulations.

The final phase involves constructing and running a digital twin model, which serves as a dynamic and predictive replica of the system. Here, the setup parameters are a mixture of values: some are manually entered by users based on specific scenarios, while others are computationally derived from the statistical and simulation phases. This blended parameter set enables the digital twin to simulate cascading effects and complex interactions within the system. By integrating both manual inputs and statistically optimized parameters, the digital twin achieves a high level of adaptability and precision, allowing for realistic scenario analysis and robust predictive insights.

1.2 DELIVERABLE STRUCTURE

The current report is structured as follows:

Chapter 2 presents the methodologies used:

- AGILE methodology, detailing its principles, practices, and how it is applied within the context of the project. It will cover the iterative process of AGILE, its benefits, and how it facilitates project management and development.
- IEEE-1730 DSEEP (Distributed Simulation Engineering and Execution Process) methodology, focusing on its application in distributed simulation engineering. It covers the process framework, steps, and best practices as outlined by the IEEE standard, and how it integrates with the overall project.

Chapter 3 includes the data representation with a reference to: the PANTHEON architecture and a detailed representation of data:

- PANTHEON architecture. A detailed reference to the architectural framework used in PANTHEON for data representation is provided. It explains how data is structured, managed, and utilised within the architecture, including diagrams or models where applicable.
- Data representation overview. This subsection gives a broad overview of how data is represented in PANTHEON. It covers key concepts, methodologies, and visual representations that are used to ensure accurate and efficient data handling.

Chapter 4 presents the statistical analysis of data in terms of scenarios and use cases. First a few distributions are selected based on the literature and previous works, next the simulation process of the distributions is explained for the selected use cases, and finally the parameters of the distributions are computed based on the data analysis available for the use cases mentioned.

Chapter 5 presents the conceptual models with a reference to the four PANTHEON use cases:

1) **Wildfire models** - delves into the conceptual models developed for simulating wildfires and includes descriptions of model parameters, simulations, and their application in understanding and managing wildfire scenarios.
2) **Earthquake models** - focuses on models designed for simulating earthquake events and covers the conceptual framework, model specifics, and how these models help in predicting and analysing seismic activity.
3) **Heatwave models -** discusses conceptual models used for simulating heatwaves and details the modelling approach, key variables, and how these models contribute to understanding the impact of heatwaves on various systems.
4) **Man-made disaster models**. - explores models developed for simulating man-made disasters, such as industrial accidents or terrorist attacks and includes descriptions of the model structure, inputs, and how these models are utilised for risk assessment and management.

The models proposed refer to the following stages in disaster management:

- **Preparation**: (a) Resource allocation, (b) Risk management processes, (c) Communication flows, (d) Evacuation planning
- **Training**: Training system for emergency scenarios
- **Simulation**: Simulation-based decision support
- **Operation**: (a) Operational structure, (b) Real-time response in actual emergencies.
- **Post event actions**

Specific to this project are the cascading effects modelling of the processes, and their implementation for simulation conducted for each scenario. In the preparatory use cases, simulation based on statistical models is conducted, then the SCDT works based on the setup of the cascading effects for each specific situation.

Chapter 6 concludes the report and summarizes the key findings and insights, while also reflecting on the overall impact of the methodologies, data representations, and conceptual models discussed, and provides final conclusions.

2. METHODOLOGIES

The complexity inherent in the PANTHEON system necessitates the application of well-defined and mature methodologies across various domains, including project management, software development, testing and validation, deployment, and other related activities. Employing comprehensive methodologies throughout the development process allows for a structured approach that ensures all aspects are meticulously planned, executed, and reviewed. This structured methodology facilitates organised workflows, minimises disarray, and maintains a clear focus on project objectives.

Adherence to a methodological framework enables the team to maintain consistency in their work processes, which is crucial for ensuring effective collaboration among the team members. Furthermore, methodologies are essential in promoting and sustaining quality throughout PANTHEON's lifecycle.

For the Pantheon project, we have identified two key categories of methodologies:

- **General methodology for project development**: We have selected AGILE[3], specifically the SCRUM[4] framework (PMI, 2024).
- **Methodologies for the development and execution of distributed simulations**: We have adopted the IEEE-1730 DSEEP (Distributed Simulation Engineering and Execution Process) (IEEE-1730, 2022).

These methodologies will be briefly discussed in the subsequent subchapters.

2.1 AGILE

AGILE methodologies[5], as defined by the AGILE Alliance[6] are grounded in the principles of the Agile Manifesto, which articulates four core values and twelve guiding principles.

The four values are:

1. Individuals and Interactions Over Processes and Tools
2. Working Software Over Comprehensive Documentation
3. Customer Collaboration Over Contract Negotiation
4. Responding to Change Over Following a Plan

The twelve principles are:

1. Our highest priority is to satisfy the customer through early and continuous delivery of valuable software.
2. Welcome changing requirements, even late in development. Agile processes harness change for the customer's competitive advantage.
3. Deliver working software frequently, from a couple of weeks to a couple of months, with a preference to the shorter timescale.
4. Businesspeople and developers must work together daily throughout the project.
5. Build projects around motivated individuals. Give them the environment and support they need and trust them to get the job done.

[3] https://www.agilealliance.org/
[4] https://www.agilealliance.org/glossary/scrum/
[5] https://www.agilealliance.org/agile101/agile-glossary/
[6] https://www.agilealliance.org/

6. The most efficient and effective method of conveying information to and within a development team is face-to-face conversation.
7. Working software is the primary measure of progress.
8. Agile processes promote sustainable development. The sponsors, developers, and users should be able to maintain a constant pace indefinitely.
9. Continuous attention to technical excellence and good design enhances agility.
10. Simplicity--the art of maximizing the amount of work not done--is essential.
11. The best architecture models, requirements, and designs emerge from self-organizing teams.
12. At regular intervals, the team reflects on how to become more effective, then tunes and adjusts its behaviour accordingly.

In concrete terms, in PANTHEON we will implement a SCRUM methodology by making usage of Teams[7] and defining the SCRUM board like in the next figure (Figure 1 SCRUM board).

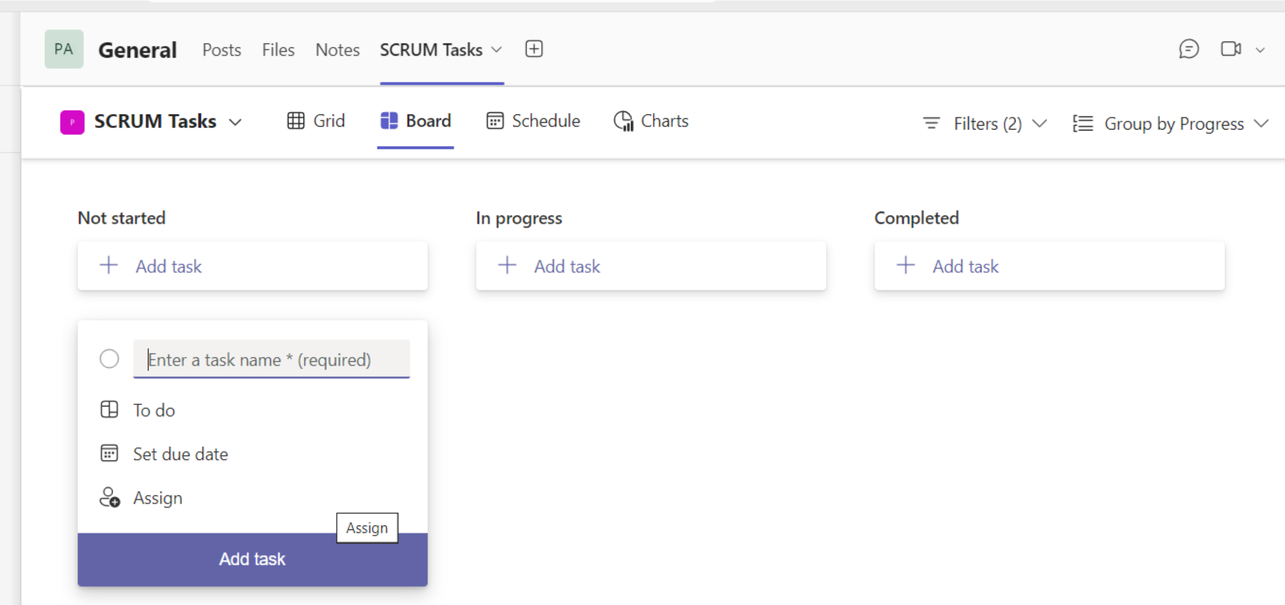


Figure 1 SCRUM board

2.2 IEEE-1730 DSEEP (DISTRIBUTED SIMULATION ENGINEERING AND EXECUTION PROCESS)

Based on (IEEE-1730, 2022) we have:

> *"The DSEEP is a generalized systems engineering process for building and executing distributed simulation environments, independent of the underlying simulation architecture. Intended as a high-level framework for simulation environment construction and execution, the DSEEP is the successor of architecture dependent engineering processes, e.g., concerning HLA (IEEE Std 1516.3™) or DIS (IEEE Std 1278.3™).*

[7] https://teams.microsoft.com/

The DSEEP represents a tailoring of best practices in the systems and software engineering communities to the M&S domain, and to the development and execution of distributed simulation environments. The DSEEP is simulation architecture-neutral, but it does contain annexes that map this architecture-neutral view to DIS, HLA, and TENA terminology"

A short description of each of the seven major steps that DSEEP follows is listed below:

1. **Definition of simulation environment objectives**: Establish the objectives and scope of the simulation environment.
2. **Perform conceptual analysis and design**: Develop a high-level design and identify key components and interactions.
3. **Develop detailed simulation design**: Create a detailed design of the simulation, specifying components, data flows, and interactions.
4. **Develop simulation environment**: Implement the simulation environment according to the detailed design.
5. **Integrate and test simulation environment**: Integrate individual simulation components and conduct tests to ensure proper functionality and interoperability.
6. **Execute simulation**: Conduct the actual simulation runs based on the defined objectives and scenarios.
7. **Analyse data and evaluate results**: Collect and analyse data from the simulation to evaluate its success in meeting the objectives and to identify areas for improvement.

This is also represented in the next figure (Figure 2 DSEEP Processes):

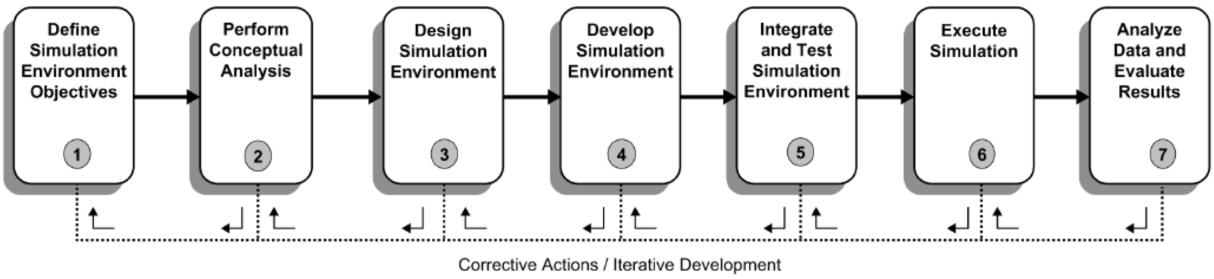


Figure 2 DSEEP Processes

The glossary of terms contained in the DSEEP (IEEE-1730, 2022), and used in PANTHEON project includes:

conceptual model: *An abstraction of what is intended to be represented within a simulation environment, which serves as a frame of reference for communicating simulation-neutral views of important entities and their key actions and interactions. The conceptual model describes what the simulation environment will represent, the assumptions limiting those representations, and other capabilities needed to satisfy the user's requirements. Conceptual models are bridges between the real world, requirements, and simulation design.*

constructive simulation: *Models and simulations that involve simulated people operating simulated systems. Real people stimulate (make inputs) to such simulations but are not involved in determining the outcomes.*

issue: *A concern, such as a situation within a development process or a technical element of an architecture, from which obstacles to achieving the objectives of the simulation environment may arise.*

live simulation: *A simulation involving real people operating real systems.*

live, virtual, and constructive (LVC) simulation: *A broadly used taxonomy describing a mixture of live simulation, virtual simulation, and constructive simulation. Note that live, virtual, and constructive simulations always include a real or synthetic person in the simulation as contrasted with a science-based simulation, which models a phenomenon or process only.*

member application: *An application that is serving some defined role within a simulation environment. This can include live, virtual, or constructive (LVC) simulation assets or can be supporting utility programs such as data loggers or visualization tools.*

objective: *The desired goals and results of the activity to be conducted in the distributed simulation environment expressed in terms relevant to the organization(s) involved.*

requirement: *A statement identifying an unambiguous and testable characteristic, constraint, process, or product of an intended simulation environment.*

simulation data exchange model (SDEM): *A specification defining the information exchanged at runtime to achieve a given set of simulation objectives. This includes class relationships, data structures, parameters, and other relevant information. (IEEE Std 1730)*

simulation environment: *A named set of member applications along with a common simulation data exchange model and set of agreements that are used to achieve some specific objective. Also referred to as a distributed simulation environment.*

stimulation: *Stimulation is the use of simulations to provide an external stimulus to a system or subsystem. An example is the use of a simulation representing the radar return from a target to drive (stimulate) the radar of a missile system within a hardware/software-in-the-loop simulation.*

virtual simulation: *A simulation involving real people operating simulated systems. Virtual simulations inject human-in-the-loop (HITL) in a central role by exercising motor control skills (e.g., flying an airplane), decision skills (e.g., committing fire control resources to action), or communication skills [e.g., as members of a command, control, communications, computers, and intelligence (C4I) team]."*

3. DATA REPRESENTATION

The conceptual models presented in this document constitute a critical component of the overall architecture of the PANTHEON project. These models and the associated data structures underpin the simulations, Decision Support System (DSS), data analysis, and the preparation of output results across all considered scenarios.

The data representations employed in PANTHEON are derived from existing data sources, tailored to meet the input requirements of the simulation, analysis, and DSS algorithms, while also considering the software technologies available for this project.

In this chapter, we will first reference the system architecture, then examine the data formats of existing sources, and finally establish the data structures utilized by the tools developed or employed within PANTHEON.

3.1 REFERENCE TO THE ARCHITECTURE

The architecture of the Pantheon system was previously detailed in D3.7 (Pantheon Consortia, 2024). Within this architectural framework, the conceptual models are positioned in the Functional View to elucidate their role within the system. Additionally, these models are described within the structural layered architecture to illustrate their placement in the implementation of the PANTHEON system.

The Functional View and the Layered Architecture are briefly presented in the following subchapters. For a comprehensive description, refer to D3.7. (Pantheon Consortia, 2024)

3.1.1 FUNCTIONAL VIEW

The functional view of the architecture described in D3.7 is shown in next figure: Figure 3.

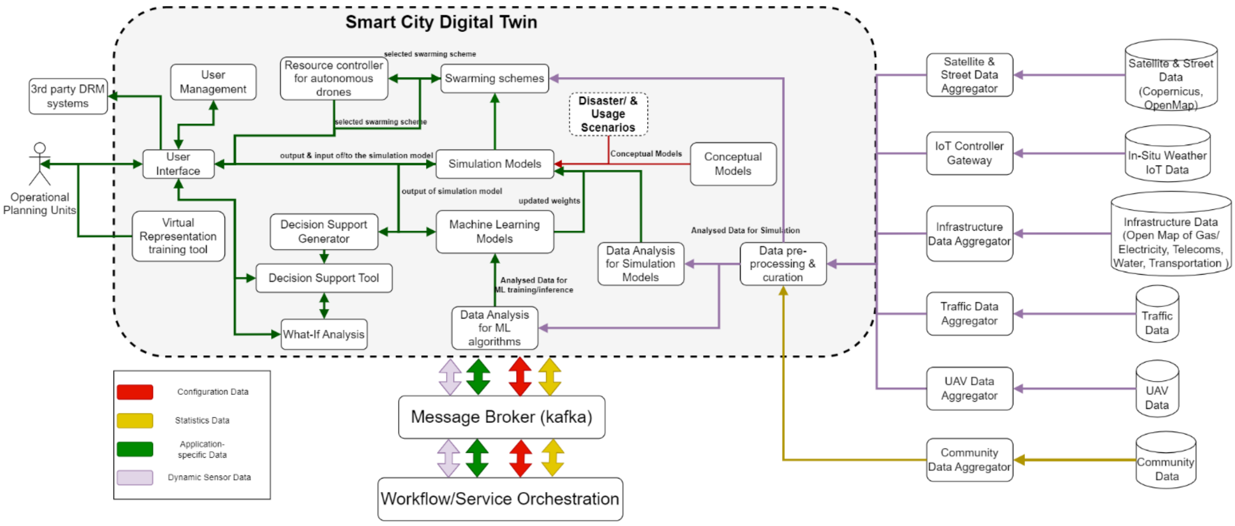


Figure 3 The functional view of the PANTHEON architecture.

Each component of the functional view of the architecture is further described in D3.7 (Pantheon Consortia, 2024) and a summary list is in the next table. In the referenced document, for each component, we provide a name, a brief description, the pilot and the scenario it will run, the module and submodule it has (if applicable), its inputs & outputs (if applicable) from/to the rest of the PANTHEON components, its algorithms (if applicable), its technologies (if applicable), its hardware requirements (if any) and a schematic for its internal architecture (if applicable). In the next table we are just summarizing the list of components referring to the task where it is developed and to the deliverable where it is described.

Table 1: List of main components.

Component	Relevant task(s)	Deliverable(s)
Satellite & Street Data Aggregator	T6.3	D6.2
Weather IoT Data Aggregator	T6.3	D6.2
UAV Data Aggregator	T6.3	D6.2
Traffic Data Aggregator	T4.3	D4.3
Community Data Aggregator	T4.1	D4.4
Data Pre-processing and Curation	T4.1	D4.4
Data Analysis for ML & ML models	T4.2, T4.3	D4.2, D4.3
Data Analysis for Simulation models	T4.2	D4.2
Conceptual Models	T4.2	D4.2
Simulation Models & Decision Support Generator	T4.4, T5.2	D4.3, D5.1
Decision Support & What-If Analysis	T5.5	D5.1
UAV Swarming Schemes	T6.1	D6.1
Resource Controller	T6.2	D6.1
User Interface	T4.5	D4.4
Virtual Representation Training Tool	T4.5	D4.4
Message Broker	T7.1	D7.1
Service & Workflow Orchestrator	T7.1	D7.1
User Management	T7.1	D7.1

3.1.2 LAYERED ARCHITECTURE

The placements of the components in the architecture are further described in D3.7 (Pantheon Consortia, 2024) and a summary is presented now.

The layers of the proposed system are:

- L1- Data and Persistence Layer. This layer is responsible for data storage and maintenance
- L2-Models Layer. This layer is responsible for model management. Simulation and ML models are part of this layer. Model storage is in L1, but their management is considered in this layer.
- L3-Backend Services Layer. It provides the homogeneous access of all services in the backend (L1 and L2) to the upper layers. Backup layer also organises the backend services by their semantics.
- L4-Analysis and Optimizations. This layer is placed on top of L3 and is also a backend layer but is specialized in analysis and optimization of data provided at the level of Layer 3.
- L5-Frontend. This layer is placed on top of the architecture and is responsible for presenting data to final users. The presentation can be in a user interface or exposed as APIs.

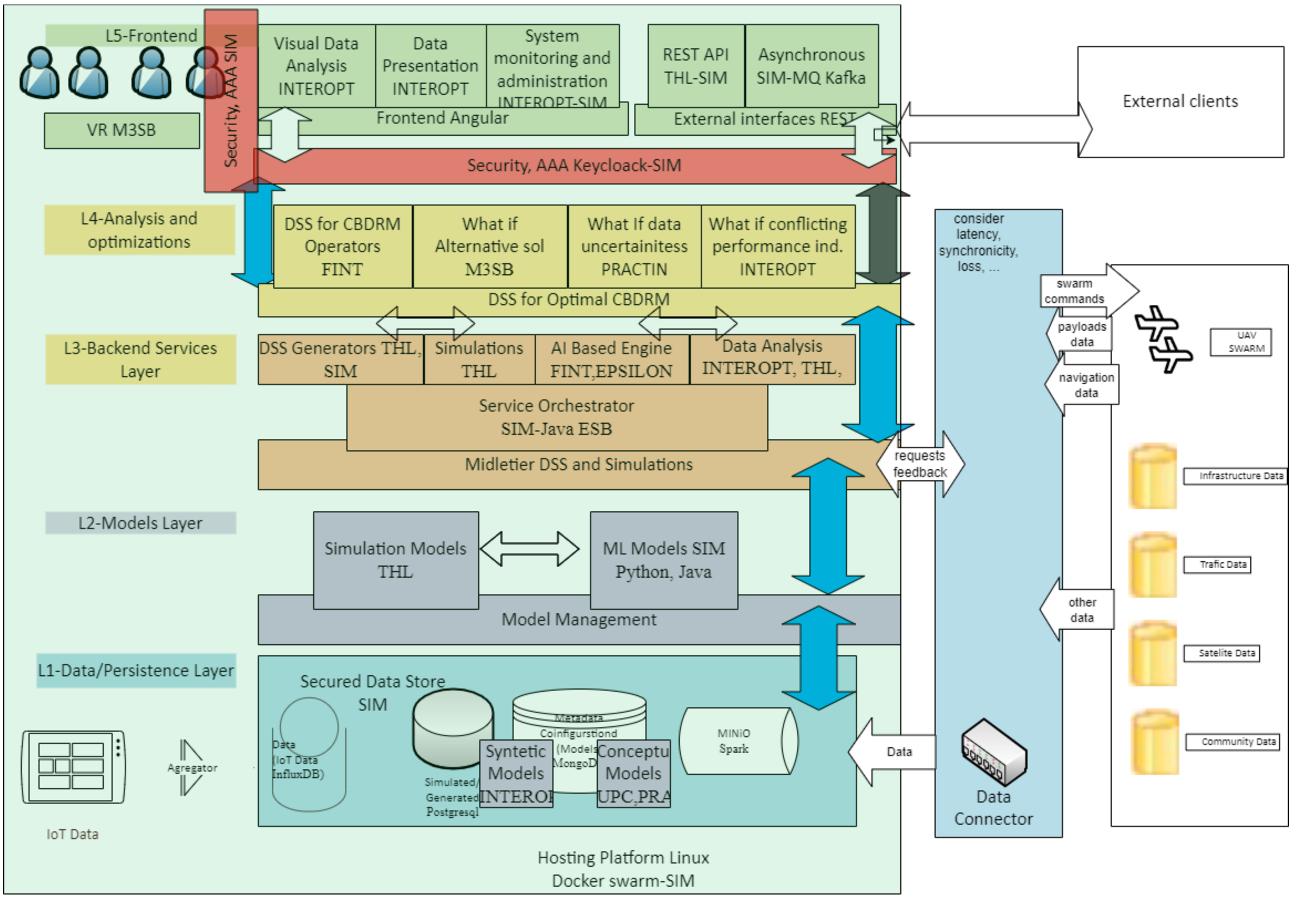


Figure 4 The Layered architecture of PANTHEON.

In the present document, our focus is primarily on Layer L1 which pertains to data persistence, and Layer L2 which concerns the processing and treatment of the models.

3.2 DATA REPRESENTATION OVERVIEW

We are considering data representation specific to the available data sources for each scenario. This approach is necessary because the scenarios are executed in various locations, involving different organizations and institutions, with information originating from diverse sources. The descriptions of these scenarios are drawn from Deliverable D3.6 (Pantheon Consortia, 2024)

3.2.1 SCENARIO 1 | DS-ATH-B: WILDFIRE IN THE REGION OF ATTICA

For this scenario, we consider data sources grouped in the following categories:

- **Local GIS and topographical data**: Geographic Information Systems (GIS) and topographical maps of the NW suburbs of Athens, the Parnitha mountain, Fyli town, and surrounding areas. This data can help in accurately modelling the fire's spread and its impact on the region.
- **Weather data** for simulating the behaviour of the wildfire under strong NE winds.
- **Infrastructure and facilities data:** This can include infrastructure maps, network layouts, and vulnerability assessments to understand the potential impact and dependencies between these infrastructures.

- **Emergency response protocols**: As described by Hellenic Fire Service, Hellenic Police, National Centre for Emergency Assistance, and other organizations involved in disaster response and management.
- **Historical wildfire and disaster data**: for creating realistic simulations and understanding potential cascading effects.
- **Population and vulnerable groups data**: Information about the population density, demographics, and locations of vulnerable groups in the affected area can help in assessing the potential impact on communities and planning for evacuation and medical assistance.
- **Risk analysis and dependency data:** Data sources for analysing dependencies between affected infrastructures, identifying critical nodes, and understanding the potential cascading effects, as mentioned in the scenario, can be derived from infrastructure risk assessments, network analysis, and relevant studies.

The location of the main sources of data as described in D3.6, is presented in the next table (Table 2):

Table 2: City of Athens Open Data main sources.

City of Athens Open Data main sources	**https://data.gov.gr/** **JSON, CSV** **https://geodata.gov.gr/en/dataset** **Proprietary, ZIP**
Datasets related with Transportation	map-of-the-main-road-network_greece_vector_shapefile Road traffic in Attica**:** 2023-10-24. Measurements of number of crossings and average speed per measurement station of the traffic monitoring network in Attica Traffic and itineraries of shipping companies: 2023-11-12. Shipping companies enter information about their itineraries (departure times, ports of call, passenger details, details of issued tickets, vehicles, etc.) Passenger public at OASA: 2023-11-10. Travel/transfer endorsements on fixed transit and bus routes Railway Network of Greece [EL]: Railway Network of Greece. This file does not include descriptive information, it contains only the routes of the railway network.
Datasets related with Infrastructure	Schools [EL]**:** This dataset contains all schools (nursery schools, primary schools, secondary schools, high schools, technical vocational schools, etc.) in Greece.
Datasets related with Environment	List of Forest Fires: 2018-12-31 Data concerning the Forest events in which the PS intervenes. Stations for the Measurement of Air Pollution [EL]**:** This dataset contains the data and positions of the stations for the measurement of air pollution in Greece. In part of the records the geographical position is absent. Wind map of Greece [EL] The wind map comprises the wind potential of Greek territory (except for Crete and parts of the prefectures of Kavala and Ksanthi) on a 150 x 150 m grid. Wildfire in North Attica, Greece (2023-08-22) Wildfire in western Attica, Greece (2023-07-17) Wildfire in Attica, Greece (2023-07-17) Forest fire in Lavrio, Eastern Attica, Greece (2021-08-17) Wildfires in Greece (2021-08-04) Fire in Western Attica, Greece (2021-08-17) Forest Fires in Attika, Greece (2018-07-24)
Datasets related with Demography	04. Estimated Population by sex and five-year age groups on 1st January for the years (2001 - 2021)
	09. Estimated Population by sex, on the 1st of January and in the middle of the year (2001 - 2021)

	10. Estimated Population on the 1st of January for the years, Hellas Total, Region, Departments (2002 - 2021)
	17. Estimated Population by Sex, Group of Citizenship and Age Group At 1st January (2009 - 2021)
	18. Estimated Population by sex, Group of Country of Birth and Age Group at 1st January (2009 - 2021)

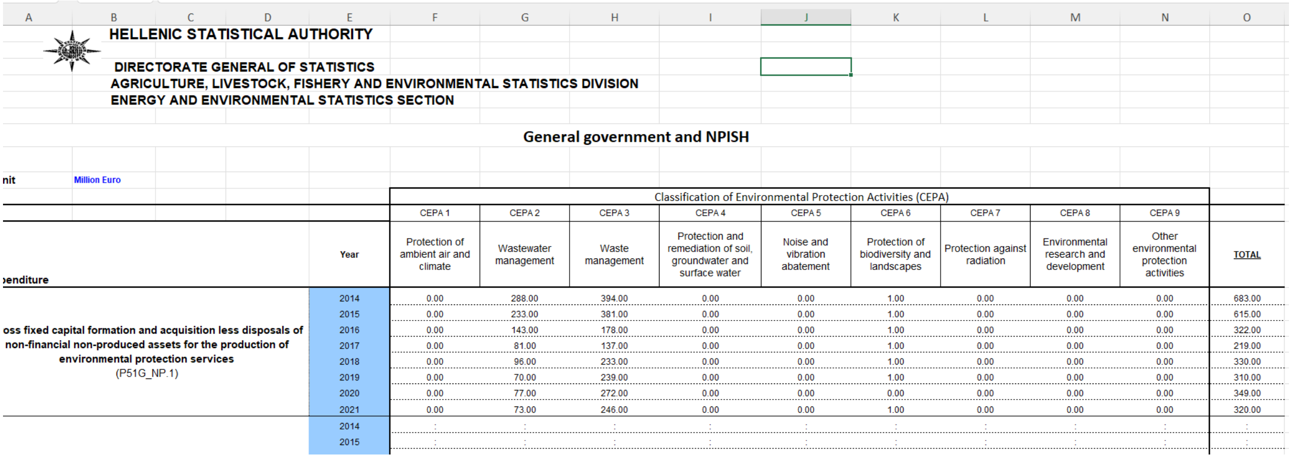
Besides the main sources already mentioned, we are also considering some additional data sources:

Table 3. Additional data sources

Data source	Details
https://effis.jrc.ec.europa.eu/	The data is obtained from Copernicus[8] - ECMWF[9] Data is in HTML format and is displayed in browser by a specific application.
https://gwis.jrc.ec.europa.eu/apps/gwis_current_situation/index.html	The data is obtained from Copernicus[10] - ECMWF[11] Data is in HTML format and is displayed in browser by a specific application.
https://www.cityofathens.gr/	Data is in HTML format and is displayed in browser by a specific application.
https://deddie.gr/en/	Data is in HTML format and is displayed in browser by a specific application.
https://www.desfa.gr/en/	Data is in HTML format and is displayed in browser by a specific application.
https://www.cosmote.gr/cs/otegroup/en/ote_ae.html	Data is in HTML format and is displayed in browser by a specific application.
https://www.gov.gr/en/upourgeia/upourgeio-upodomon-kai-metaphoron	Data is in HTML format and is displayed in browser by a specific application.
https://www.fireservice.gr/	Data is in HTML format and is displayed in browser by a specific application.
https://epadap.web.auth.gr/?lang=en	QGIS[12] plugin

[8] Homepage | Copernicus
[9] Forecasts | ECMWF
[10] Homepage | Copernicus
[11] Forecasts | ECMWF
[12] https://qgis.org/

https://www.preventionweb.net/national-platform/greece-national-platform	Data is in HTML format and is displayed in browser by a specific application.
https://civilprotection.gov.gr/	Data is in HTML format and is displayed in browser by a specific application.
https://www.astynomia.gr/?lang=en	Data is in HTML format and is displayed in browser by a specific application.
https://www.statistics.gr/en/home/	Data is presented in tabular format.



HELLENIC STATISTICAL AUTHORITY

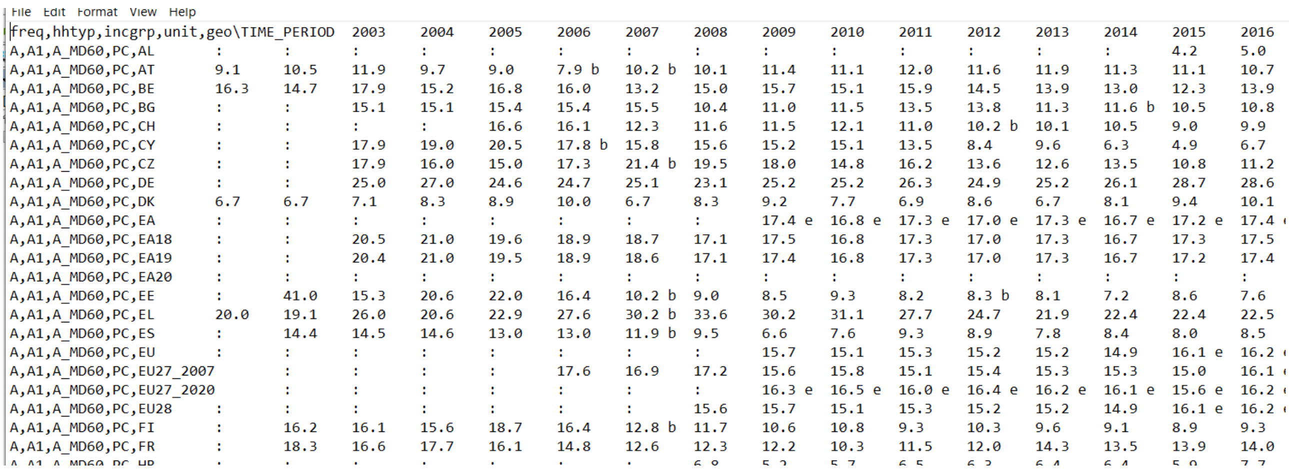
DIRECTORATE GENERAL OF STATISTICS
AGRICULTURE, LIVESTOCK, FISHERY AND ENVIRONMENTAL STATISTICS DIVISION
ENERGY AND ENVIRONMENTAL STATISTICS SECTION

General government and NPISH

nit Million Euro

		Classification of Environmental Protection Activities (CEPA)									
		CEPA 1	CEPA 2	CEPA 3	CEPA 4	CEPA 5	CEPA 6	CEPA 7	CEPA 8	CEPA 9	
ɔenditure	Year	Protection of ambient air and climate	Wastewater management	Waste management	Protection and remediation of soil, groundwater and surface water	Noise and vibration abatement	Protection of biodiversity and landscapes	Protection against radiation	Environmental research and development	Other environmental protection activities	TOTAL
oss fixed capital formation and acquisition less disposals of non-financial non-produced assets for the production of environmental protection services (P51G_NP.1)	2014	0.00	288.00	394.00	0.00	0.00	1.00	0.00	0.00	0.00	683.00
	2015	0.00	233.00	381.00	0.00	0.00	1.00	0.00	0.00	0.00	615.00
	2016	0.00	143.00	178.00	0.00	0.00	1.00	0.00	0.00	0.00	322.00
	2017	0.00	81.00	137.00	0.00	0.00	1.00	0.00	0.00	0.00	219.00
	2018	0.00	96.00	233.00	0.00	0.00	1.00	0.00	0.00	0.00	330.00
	2019	0.00	70.00	239.00	0.00	0.00	1.00	0.00	0.00	0.00	310.00
	2020	0.00	77.00	272.00	0.00	0.00	0.00	0.00	0.00	0.00	349.00
	2021	0.00	73.00	246.00	0.00	0.00	1.00	0.00	0.00	0.00	320.00
	2014	:	:	:	:	:	:	:	:	:	:
	2015	:	:	:	:	:	:	:	:	:	:

Figure 5 Example of data from www.statistics.gr



File Edit Format View Help

```
freq,hhtyp,incgrp,unit,geo\TIME_PERIOD 2003 2004 2005 2006 2007 2008 2009 2010 2011 2012 2013 2014 2015 2016
A,A1,A_MD60,PC,AL : : : : : : : : : : : : 4.2 5.0
A,A1,A_MD60,PC,AT 9.1 10.5 11.9 9.7 9.0 7.9 b 10.2 b 10.1 11.4 11.1 12.0 11.6 11.9 11.3 11.1 10.7
A,A1,A_MD60,PC,BE 16.3 14.7 17.9 15.2 16.8 16.0 13.2 15.0 15.7 15.1 15.9 14.5 13.9 13.0 12.3 13.9
A,A1,A_MD60,PC,BG : : 15.1 15.1 15.4 15.4 15.5 10.4 11.0 11.5 13.5 13.8 11.3 11.6 b 10.5 10.8
A,A1,A_MD60,PC,CH : : : : 16.6 16.1 12.3 11.6 11.5 12.1 11.0 10.2 b 10.1 10.5 9.0 9.9
A,A1,A_MD60,PC,CY : : 17.9 19.0 20.5 17.8 b 15.8 15.6 15.2 15.1 13.5 8.4 9.6 6.3 4.9 6.7
A,A1,A_MD60,PC,CZ : : 17.9 16.0 15.0 17.3 21.4 b 19.5 18.0 14.8 16.2 13.6 12.6 13.5 10.8 11.2
A,A1,A_MD60,PC,DE : : 25.0 27.0 24.6 24.7 25.1 23.1 25.2 25.2 26.3 24.9 25.2 26.1 28.7 28.6
A,A1,A_MD60,PC,DK 6.7 6.7 7.1 8.3 8.9 10.0 6.7 8.3 9.2 7.7 6.9 8.6 6.7 8.1 9.4 10.1
A,A1,A_MD60,PC,EA : : : : : : : : 17.4 e 16.8 e 17.3 e 17.0 e 17.3 e 16.7 e 17.2 e 17.4
A,A1,A_MD60,PC,EA18 : : 20.5 21.0 19.6 18.9 18.7 17.1 17.5 16.8 17.3 17.0 17.3 16.7 17.3 17.5
A,A1,A_MD60,PC,EA19 : : 20.4 21.0 19.5 18.9 18.6 17.1 17.4 16.8 17.3 17.0 17.3 16.7 17.2 17.4
A,A1,A_MD60,PC,EA20 : : : : : : : : : : : : : : : :
A,A1,A_MD60,PC,EE : 41.0 15.3 20.6 22.0 16.4 10.2 b 9.0 8.5 9.3 8.2 8.3 b 8.1 7.2 8.6 7.6
A,A1,A_MD60,PC,EL 20.0 19.1 26.0 20.6 22.9 27.6 30.2 b 33.6 30.2 31.1 27.7 24.7 21.9 22.4 22.4 22.5
A,A1,A_MD60,PC,ES : 14.4 14.5 14.6 13.0 13.0 11.9 b 9.5 6.6 7.6 9.3 8.9 7.8 8.4 8.0 8.5
A,A1,A_MD60,PC,EU : : : : : : : : 15.7 15.1 15.3 15.2 15.2 14.9 16.1 e 16.2
A,A1,A_MD60,PC,EU27_2007 : : : : 17.6 16.9 17.2 15.6 15.8 15.1 15.4 15.3 15.3 15.0 16.1
A,A1,A_MD60,PC,EU27_2020 : : : : : : : 16.3 e 16.5 e 16.0 e 16.4 e 16.2 e 16.1 e 15.6 e 16.2
A,A1,A_MD60,PC,EU28 : : : : : : : 15.6 15.7 15.1 15.3 15.2 15.2 14.9 16.1 e 16.2
A,A1,A_MD60,PC,FI : 16.2 16.1 15.6 18.7 16.4 12.8 b 11.7 10.6 10.8 9.3 10.3 9.6 9.1 8.9 9.3
A,A1,A_MD60,PC,FR : 18.3 16.6 17.7 16.1 14.8 12.6 12.3 12.2 10.3 11.5 12.0 14.3 13.5 13.9 14.0
```

Figure 6 Example of data from https://ec.europa.eu/eurostat

Summarising the existing sources of data, we can conclude that:

- There is JSON data available on several sources
- There is **Copernicus** geographic data available in one of the formats:
  - **GeoTIFF (.tif): A common format for geospatial raster data.**
  - **NetCDF (.nc):** Used for multi-dimensional scientific data, such as atmospheric or oceanographic datasets.
  - **Shapefile (.shp, .shx, .dbf):** A popular format for geographic information system (GIS) data.

- GeoJSON (.geojson): A format for encoding a variety of geographic data structures.
- GML (Geography Markup Language): XML-based format used for expressing geographic features.

- There is **statistics** data in tabular format (CSV)
- There is data presented in **web pages** as html.

For the purposes of this project, **JSON and CSV data will be stored as JSON columns and tabular data, respectively, within a** (Postgresql[13]) **relational database**. Geographic data will be managed as objects within an object storage system (Minio[14]). Additionally, HTML data will be parsed and stored in either tabular or object formats, depending on the specific requirements of each case.

3.2.2 SCENARIO 2 | DS-ATH-A: EARTHQUAKE IN THE REGION OF ATTICA

Given the location of this scenario, the proximity of the primary data sources, and the similarity of all information sources, the setup closely mirrors that of Scenario 1.

For this scenario, we consider data sources grouped in the following categories:

- **Infrastructure mapping and vulnerability assessments:** This data can include infrastructure maps, vulnerability assessments, and risk analysis reports.
- **Historical earthquake data:** Understanding the patterns of cascading effects following earthquakes will help in accurately simulating and planning for this scenario.
- **Real-time seismic monitoring and early warning systems:** This data can aid in rapid response coordination.
- **Emergency response protocols and stakeholder responsibilities:** Hellenic Fire Service, Hellenic Police, National Centre for Emergency Assistance, and other organizations involved in disaster response and management.
- **Population density and vulnerable groups data:** for assessing the potential impact on communities and planning for effective search, rescue, and medical evacuation operations.
- **Soil composition and stability:** This data can be obtained from geological surveys and soil analysis reports.
- **Geological information:** can aid in predicting the areas that are most vulnerable to seismic impact.
- **Seismic data and geological surveys.** This data is typically provided by geological and seismological institutions. We are using European-Mediterranean Seismological Centre (EMSC) (https://www.emsc-csem.org/) and European Integrated Data Archive (EIDA) (https://www.orfeus-eu.org/data/eida/)
- **NOAA national centres for environmental information (NCEI): website:** https://www.ncei.noaa.gov/ NCEI provides access to environmental and climate data sets, which can be valuable for analysing natural disasters such as severe weather events, hurricanes, and climatological data.

[13] https://www.postgresql.org/
[14] https://min.io/product/enterpriseoverview

- **EM-DAT: the international disaster database:** Website: https://www.emdat.be/ EM-DAT is a global database providing essential data on the occurrence and effects of over 22,000 natural and technological disasters from 1900 to the present.

- **USGS earthquake hazards program: website:** https://earthquake.usgs.gov/ The USGS provides access to earthquake data sets, seismic event information, and geospatial data related to seismic hazards and monitoring.

In the same way as for Scenario 1, summarizing the existing sources of data we can conclude that:

- There is JSON data available on several sources
- There is **Copernicus** geographic data available in one of the formats:
  - **GeoTIFF (.tif): A common format for geospatial raster data.**
  - **NetCDF (.nc):** Used for multi-dimensional scientific data, such as atmospheric or oceanographic datasets.
  - **Shapefile (.shp, .shx, .dbf):** A popular format for geographic information system (GIS) data.
  - **GeoJSON (.geojson):** A format for encoding a variety of geographic data structures.
  - **GML (Geography Markup Language):** XML-based format used for expressing geographic features.
- There is **statistics** data in tabular format (CSV)
- There is data presented in **web pages** as html.

For the purposes of this project, **JSON and CSV data will be stored as JSON columns and tabular data, respectively, within a** (Postgresql[15]) **relational database**. Geographic data will be managed as objects within an object storage system (Minio[16]). Additionally, HTML data will be parsed and stored in either tabular or object formats, depending on the specific requirements of each case.

3.2.3 SCENARIO 3 | DS-VIE-A: HEATWAVE IN VIENNA

For this scenario, we consider data sources grouped in the following categories:

- **Meteorological data:** Historical and real-time meteorological data, including temperature records, humidity levels, and weather forecasts, is essential. The Austrian central meteorological institute ZAMG and the Central Institute for Meteorology will provide this data.

- **Geospatial and urban data:** data on rapid urbanization, population growth, land use, vegetation cover, and infrastructure details. City urban planning and infrastructure departments, as well as geographical and environmental research institutes, have such data.

- **Demographic and health data:** This includes vulnerability assessments for different groups, hospital occupancy rates during heatwaves, ambulance call-out statistics, and community health profiles. Access to health ministries, hospitals, health research organizations, and public health agencies is crucial for obtaining this data.

- **Emergency operational data**: Data on ambulance dispatches, first responder resource allocation, and emergency response plans will be required. This should cover human resource planning, equipment assessments, and the impact of heatwaves on operational functions.

[15] https://www.postgresql.org/
[16] https://min.io/product/enterpriseoverview

- **Satellites and in situ sensors** offer indispensable tools for comprehensive weather monitoring, forecasting, urban heat island analysis, environmental assessment, and health impact evaluations, ultimately supporting proactive and effective responses to heatwave events in urban areas like Vienna.
- **Simulation input data:** To run simulations and decision support, the system will need parameters such as temperature ranges, humidity levels, and duration of the heatwave. This data input is critical for generating models, predicting impacts, and identifying preventive actions.

The location of the main data sources, as described in D3.6, is presented in the next list.

1. ZAMG - Central Institute for Meteorology and Geodynamics:
   Website: https://www.zamg.ac.at/cms/en

Klimawerte Jahr 2023 (1981-2010) GeoSphere Austria

			Lufttemperatur						Kenntage				Niederschlag					Schnee		Sonnenschein	
			Monatsmittel		Maximum		Minimum		Eistage	Frosttage	Sommert.	Tropent.	Monatssumme		Max. Tages-summe		Tage mit Niederschlag	Schnee-deckentage	Maximale Schneehöhe	Monatssumme	
Station	BL	Seehöhe	°C	Abw. in K 1981-2010	°C	Tag	°C	Tag	Anz. Tage Tmax < 0°C	Anz. Tage Tmin < 0°C	Anz. Tage Tmax ≥ 25°C	Anz. Tage Tmax ≥ 30°C	mm	% v. Mittel 1981-2010	mm	Tag	Anz. Tage ≥ 0.1 mm	Anz. Tage	cm	h	% v. Mittel 1981-2010
Bregenz	V	424	11.8	2.2	35.8	15.7.	-7.0	4.12.	1	45	92	20	1844	122	77.0	28.8.	190	13	23	1790	104
Feldkirch	V	438	11.6	2.1	36.2	11.7.	-8.2	4.12.	2	63	98	24	1563	115	69.9	4.8.	182	12	18	1910	109
Innsbruck-Flgh.	T	578	10.6	1.9	36.7	11.7.	-11.6	4.12.	6	88	90	30	964	109	57.0	27.8.	177	30	21	1907	97
Kufstein	T	490	10.5	2.1	36.5	15.7.	-7.9	4.12.	6	69	66	21	1411	106	37.0	16.5.	188	38	32	1585	94
Lienz	T	661	9.7	2.3	33.9	21.6.	-12.7	30.1.	5	122	77	23	1013	113	71.3	28.8.	134	33	28	1966	96
Patscherkofel	T	2251	1.8	1.4	22.4	24.8.	-16.5	20.1.	113	192	0	0	1114	127	52.9	27.8.	181			1816	93
Reutte	T	842	8.6	1.9	33.5	22.6.	-12.0	3.12.	16	106	44	5	1811	126	42.3	22.12.	198	80	60	1666	93
St. Anton/Arl.	T	1304	6.7	1.7	31.4	24.8.	-14.1	4.12.	25	131	40	3	1475	130	57.6	28.4.	202	120	98	1452	
Bad Gastein	S	1092	7.2	1.5	33.4	22.6.	-14.4	8.2.	29	137	46	9	1419	118	88.3	28.8.	207			1310	98
Bischofshofen	S	550	9.5	1.8	34.5	22.6.	-12.2	4.12.	9	91	68	20	1149	110	68.7	28.8.	192			1529	97

Figure 7 Example of data from ZAMG

2. City of Vienna Open Data Portal:
   Website: https://www.data.gv.at/katalog/organisation/ddbffc36-78ae-455d-8c1f-59b2c7b344e8

A	B	C	D	E	F	G	H	I	J	K
NAME	RÃ„UMLICHE FLEXIBILITÃ„T	ZEITLICHE FLEXIBILITÃ„T	BESTELLNUMMER	HOMEPAGE	BUNDESLAND	GEMEINDE	BETRIEBSZ	EINSCHRÃ	EINZELTICI	KOORDINATEN
60plusTaxi Burgenland	TÃ¼r-TÃ¼r	nach Bedarf		https://www.wko.at/bgld/tran	Burgenland	Donnerskii	24-Jul	eingeschrÃ	2,5	47Â° 50' 46.93" N,
ALSOLE-Mobil	TÃ¼r-TÃ¼r	nach Bedarf	-5020	http://alsole.dellach.at/ALSO	KÃ¤rnten	Dellach,Ki	Mo-Fr 08:0	lange Vorb	2,5	46Â° 36' 19.58" N,
AST Hollabrunn	Haltestelle-TÃ¼r	mit Fahrplan nach Bedarf	-824	https://www.vor.at/fileadmin/	NiederÃ¶sterreich	Hollabrunr	Mo-Fr 08:0	eingeschrÃ	3	48Â° 34' 8.80" N, 1
AST Katzelsdorf	Haltestelle-TÃ¼r	mit Fahrplan nach Bedarf	-824	https://www.vor.at/fileadmin/	NiederÃ¶sterreich	Katzelsdor	Mo-Fr 08:30-20:00		2,5	
AST Kirchschlag - HellmonsÃ¶dt - !	Haltestelle-TÃ¼r	mit Fahrplan nach Bedarf	-1321	https://www.kirchschlag.net/	OberÃ¶sterreich	HellmonsÃ	Sa,So,PH (	eingeschrÃ	12	
AST Krems	Haltestelle-Haltestelle	mit Fahrplan nach Bedarf	-179	https://www.krems.at/leben/	NiederÃ¶sterreich	Krems an c	Mo-We 00:00-00:30, (		2,5	48Â° 24' 35.96" N,
AST Linz (Nacht-AST)	Haltestelle-TÃ¼r	mit Fahrplan nach Bedarf	-833	https://www.linzag.at/portal/	OberÃ¶sterreich	Linz,Ansfel	Mo-Su 20:(	eingeschrÃ	4,6	48Â° 18' 24.98" N,
AST Linz (Tages-AST)	Haltestelle-TÃ¼r	mit Fahrplan nach Bedarf	-661955	https://www.linzag.at/portal/	OberÃ¶sterreich	Linz	Mo-Su 5:30-19:30		2,5	48Â° 18' 24.98" N,
AST Mitterndorf	Haltestelle-TÃ¼r	mit Fahrplan nach Bedarf	-826	https://www.mitterndorf.at/U	NiederÃ¶sterreich	Ebreichsdc	Mo-Fr 04:40-23:15; Su		3	
AST Moorbad Harbach	Haltestelle-Haltestelle	mit Fahrplan nach Bedarf	-824	https://www.vor.at/fileadmin/	NiederÃ¶sterreich	Moorbad H	Mo-Fr 06:00-17:30; PH off			
AST Moosbrunn	Haltestelle-TÃ¼r	mit Fahrplan nach Bedarf	-824	https://www.moosbrunn.gv.a	NiederÃ¶sterreich	Mitterndor	Mo-Sa 05:00-01:30		2,8	
AST Schwechat	Haltestelle-TÃ¼r	mit Fahrplan nach Bedarf	-737	https://www.schwechat.gv.at	NiederÃ¶sterreich	Schwecha	05:00-00:20		3	48Â° 8' 27.34" N, 1

Figure 8 Example of data from City of Vienna Open Data Portal

3. Public data on data.gv.at
   Website:data.gv.at/katalog/api/3/action/package_show?id=1f4a62eb-6f1a-4bfe-b7cd-e6dede0be00a

Refresh (Ctrl+R)


```
{
    "help": "https://www.data.gv.at/katalog/api/3/action/help_show?name=package_show",
    "success": true,
    "result": {
        "attribute_description": "",
        "author": null,
        "author_email": null,
        "begin_datetime": "2011-05-17T00:00:00",
        "categorization": "{http://publications.europa.eu/resource/authority/data-theme/SOCI,http://publications.europa.eu/
theme/HEAL,http://publications.europa.eu/resource/authority/data-theme/AGRI,http://publications.europa.eu/resource/authorit
theme/REGI,http://publications.europa.eu/resource/authority/data-theme/TRAN,http://publications.europa.eu/resource/authorit
        "creator_user_id": "1ef312ee-5bd3-42ed-a704-5eb22d67fbbe",
        "en_title_and_desc": "All metadata of the Austrian metadata records published on data.gv.at",
        "end_datetime": "",
        "geographic_bbox": "",
        "geographic_toponym": "",
        "id": "1f4a62eb-6f1a-4bfe-b7cd-e6dede0be00a",
        "isopen": false,
        "license_citation": "Datenquelle: Cooperation OGD Österreich -  https://data.gv.at",
        "license_id": null,
        "license_title": null,
        "license_url": "https://creativecommons.org/licenses/by/4.0/deed.de",
        "lineage_quality": "",
        "maintainer": "Cooperation OGD Österreich",
        "maintainer_email": "coopogd-l@wien.gv.at",
        "maintainer_link": "https://www.data.gv.at/infos/cooperation-ogd-oesterreich/",
        "metadata_created": "2015-11-05T07:45:05.751810",
        "metadata_identifier": "1f4a62eb-6f1a-4bfe-b7cd-e6dede0be00a",
        "metadata_linkage": "http://vmdev9004.adv.magwien.gv.at:8002/html/index.html?package=add66f20-d033-4eee-b9a0-47019E
        "metadata_linkage_name": "",
        "metadata_modified": "2024-07-01T06:24:03.795382",
        "metadata_original_portal": "",
        "name": "metadaten-von-ogd-osterreich",
        "notes": "Alle Metadaten der auf data.gv.at publizierten Metadatensätze Österreichs",
        "num_resources": 1,
        "num_tags": 2,
        "openapi_schema": "https://www.data.gv.at/katalog/schema.yml",
        "organization": {
            "id": "f8382f20-b9d1-459b-82fe-b42b31078c8f",
            "name": "cooperation-ogd-oesterreich",
            "title": "Cooperation OGD Österreich",
            "type": "organization",
            "description": "",
            "image_url": "https://www.data.gv.at/wp-content/uploads/logos-organisationen/cooperation-ogd-oesterreich.png",
            "created": "2017-12-01T10:19:32.361539",
            "is_organization": true,
            "approval_status": "approved",
            "state": "active"
        },
        "owner_org": "f8382f20-b9d1-459b-82fe-b42b31078c8f",
        "private": false,
        "publisher": "Cooperation OGD Österreich",
        "publisher_email": "",
        "publisher_link": "",
        "schema_characterset": "utf8",
        "schema_language": "ger",
        "schema_name": "OGD Austria Metadata 2.3",
        "state": "active",
        "terms_url": "",
        "title": "Metadaten von OGD Österreich",
        "type": "dataset",
        "update_frequency": "monthly",
        "url": null,
        "version": null,
        "resources": [
            {
                "cache_last_updated": null,
                "cache_url": null,
                "characterset": "",
                "created": "2015-11-18T00:00:00",
                "datastore_active": true,
```


Figure 9 Example of metadata published on data.gv.at

4. Central Institution for Meteorology and Geodynamics (ZAMG):
   Website: https://www.zamg.ac.at/cms/en/climate

Klimawerte Jahr 2023 (1981-2010) GeoSphere Austria

			Lufttemperatur						Kenntage				Niederschlag					Schnee		Sonnenschein	
			Monatsmittel		Maximum		Minimum		Eistage	Frosttage	Sommert.	Tropent.	Monatssumme		Max. Tages-summe		Tage mit Niederschlag	Schnee-deckentage	Maximale Schneehöhe	Monatssumme	
Station	BL	Seehöhe	°C	Abw. in K 1981-2010	°C	Tag	°C	Tag	Anz. Tage Tmax < 0°C	Anz. Tage Tmin < 0°C	Anz. Tage Tmax ≥ 25°C	Anz. Tage Tmax ≥ 30°C	mm	% v. Mittel 1981-2010	mm	Tag	Anz. Tage ≥ 0.1 mm	Anz. Tage	cm	h	% v. Mittel 1981-2010
Bregenz	V	424	11.8	2.2	35.8	15.7.	-7.0	4.12.	1	45	92	20	1844	122	77.0	28.8.	190	13	23	1790	104
Feldkirch	V	438	11.6	2.1	36.2	11.7.	-8.2	4.12.	2	63	98	24	1563	115	69.9	4.8.	182	12	18	1910	109
Innsbruck-Flgh.	T	578	10.6	1.9	36.7	11.7.	-11.6	4.12.	6	88	90	30	964	109	57.0	27.8.	177	30	21	1907	97
Kufstein	T	490	10.5	2.1	36.5	15.7.	-7.9	4.12.	6	69	66	21	1411	106	37.0	16.5.	188	38	32	1585	94
Lienz	T	661	9.7	2.3	33.9	21.6.	-12.7	30.1.	5	122	77	23	1013	113	71.3	28.8.	134	33	28	1966	96
Patscherkofel	T	2251	1.8	1.4	22.4	24.8.	-16.5	20.1.	113	192	0	0	1114	127	52.9	27.8.	181			1816	93
Reutte	T	842	8.6	1.9	33.5	22.6.	-12.0	3.12.	16	106	44	5	1811	126	42.3	22.12.	198	80	60	1666	93
St. Anton/Arl.	T	1304	6.7	1.7	31.4	24.8.	-14.1	4.12.	25	131	40	3	1475	130	57.6	28.4.	202	120	98	1452	
Bad Gastein	S	1092	7.2	1.5	33.4	22.6.	-14.4	8.2.	29	137	46	9	1419	118	88.3	28.8.	207			1310	98
Bischofshofen	S	550	9.5	1.8	34.5	22.6.	-12.2	4.12.	9	91	68	20	1149	110	68.7	28.8.	192			1529	97

Figure 10 Example of metadata published on ZAMG-climate

5. Austrian Cyber Security Center:
   Website: https://www.acsc.gv.at/

Data is in HTML format.

6. City of Vienna Official Website:
   Website: https://www.wien.gv.at/english/

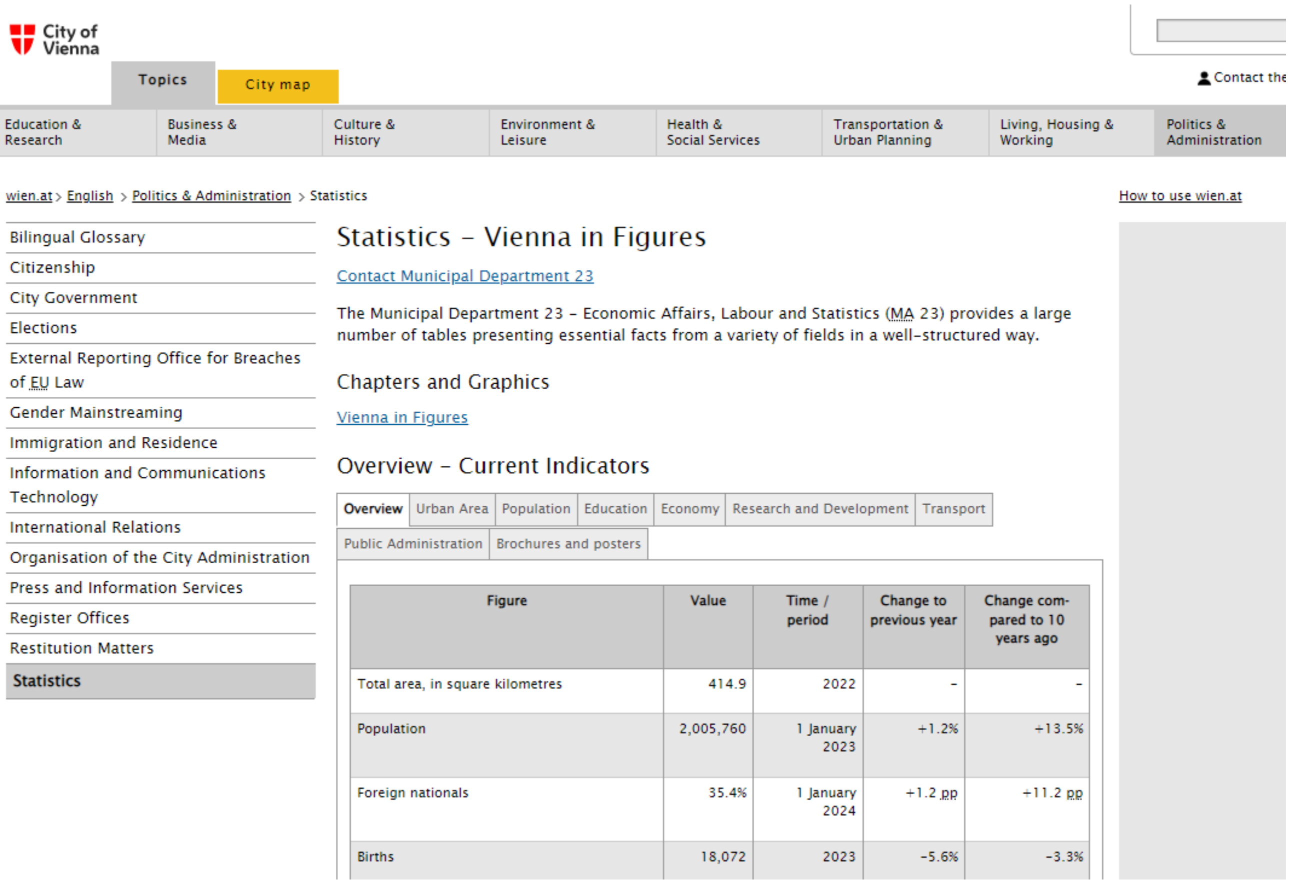


Figure 11 Example of metadata published on City of Vienna Official Website

Summarising the existing sources of data, we can conclude that:

- There is JSON data available on several sources
- There is data from different sources mainly in tabular format
- There is **statistics** data in tabular format (CSV)
- There is data presented in **web pages** as html.

For the purposes of this project, **JSON and CSV data will be stored as JSON columns and tabular data, respectively, within a** (Postgresql[17]) **relational database**. Additionally, HTML data will be parsed and stored in either tabular or object formats, depending on the specific requirements of each case.

3.2.4 SCENARIO 4 | DS-VIE-B: MAN-MADE DISASTER IN VIENNA

Given the location of this scenario, the proximity of the primary data sources, and the similarity of all information sources, the setup closely mirrors that of Scenario 3.

For this scenario, we consider data sources grouped in the following categories:

[17] https://www.postgresql.org/

- **Historical disaster data:** Access to comprehensive historical disaster data, including records of cyber-attacks, incidents of industrial infrastructure compromise, and previous forest fires in the region.
- **Geographic information system (GIS) data:** Detailed GIS data that includes infrastructure mapping, forested areas, transportation networks, and vulnerable community locations in Vienna and its surroundings.
- **Climate and weather data:** Information on local weather patterns, wind conditions, temperature, humidity, and historical heatwave and fire weather indices for Vienna.
- **Cybersecurity threat intelligence data:** Data sets related to cybersecurity threat intelligence, including information on known cyber-attack tactics, techniques, and procedures, and historical attack patterns.
- **Environmental monitoring data:** Real-time and historical environmental monitoring data, which includes information from in-situ IoT sensors for local environmental conditions, such as temperature, air quality, and sudden changes indicative of hazards, including potential fire outbreaks.
- **Urban planning and infrastructure data:** Data sets related to urban planning, critical infrastructure locations, transportation networks, and population density to assess vulnerability and cascade effects.
- **First responder and emergency services data:** Information on first responder protocols, emergency response plans, and available resources for specific organizations involved in disaster response and management.

The sources of data are the same as for Scenario3.

In the same way as for Scenario 3, summarising the format of existing sources of data we can conclude that:

- There is JSON data available on several sources
- There is data from different sources mainly in tabular format
- There is **statistics** data in tabular format (CSV)
- There is data presented in **web pages** as html.

For the purposes of this project, **JSON and CSV data will be stored as JSON columns and tabular data, respectively, within a** (Postgresql[18]) **relational database**. Additionally, HTML data will be parsed and stored in either tabular or object formats, depending on the specific requirements of each case.

[18] https://www.postgresql.org/

4. STATISTICAL DATA ANALYSIS

Data available from different sources was studied from statistical point of view to identify the statistical parameters and distributions.

Base on the literature ((Ogata, 1988), (Maxwell B. Joseph, 2019), (Liu, 2018)), we have selected the most appropriate distribution for each scenario, and then estimated the distribution parameters based on the input data obtained from the sources presented in the previous chapter.

The statistical data analysis is presented grouped on scenarios. The four scenarios are addressed: Wildfire, Earthquake, Heatwave and Man-made disaster.

The use cases are considered for each scenario, and the statistical analysis refers to each particular use case. The statistical analysis is used for the preparatory use cases (setup or initiation) when simulation based on statistical models is run. These simulations give inputs for the main simulation use case, which is then used by following the description of the cascading effects.

The use cases are taken from deliverable D3.6 (Pantheon Consortia, 2024).

4.1 WILDFIRE MODELS

4.1.1 USE CASES

The use cases specific to wildfire scenario are presented in the next figure extracted from D3.6

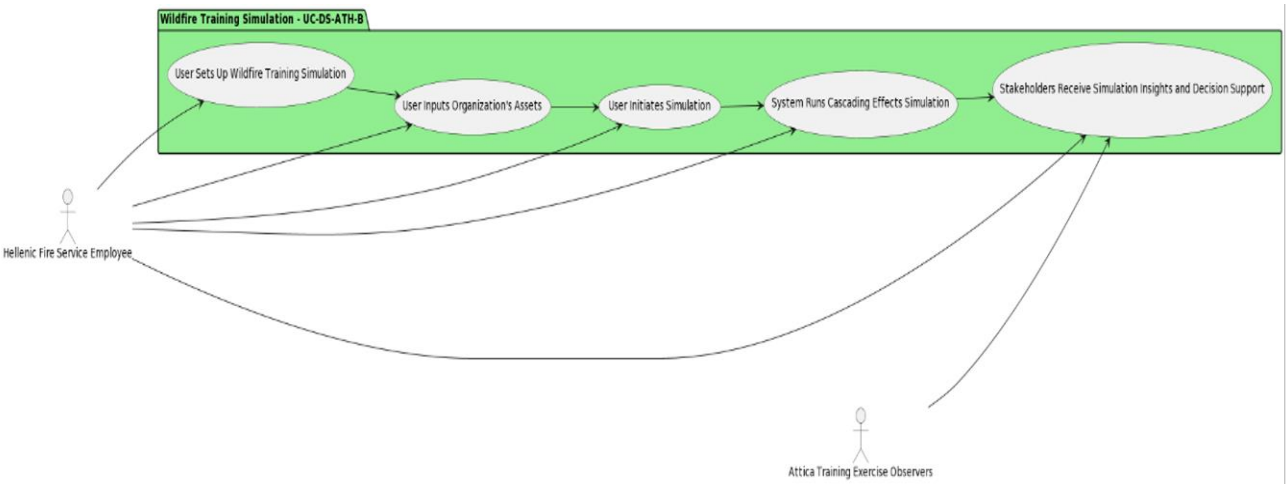


Figure 12 Wildfire simulation use cases (as in D3.6)

The way the statistical analysis of data and statistical simulations are applied is described below:

- UC-DS-ATH-B-1-**User Sets Up Wildfire Training Simulation:**

This initial use case involves a user (specifically a Hellenic Fire Service Employee) configuring the wildfire simulation environment. The setup includes defining simulation parameters such as location, severity, expected weather conditions, and other relevant inputs to create a realistic training scenario.

The simulation parameters are obtained by analysing existing data, and by selecting the simulation type.

The statistical analysis of existing data applies to all sort of simulations. Hoerver, the selection of distribution applies only to statistical simulation.

The parameters analysed and used in statistical analysis refers to:

- Occurrence (in a period)
- Time between events
- Duration of events
- Fire size
- Cumulative time of events
- Extreme or rare events
- Spatial distribution

- UC-DS-ATH-B-2-**User Inputs Organisation's Assets:**

The user then provides details about the assets available for the simulation. These assets could include firefighting vehicles, personnel, equipment, and any infrastructure the organisation intends to use in response to the wildfire. This step ensures that the simulation accurately reflects the organisation's capacity to respond to a wildfire.

This use case does not involve any statistical analysis of existing data.

- UC-DS-ATH-B-3-**User Initiates Simulation:**

After configuring the setup and inputting organizational assets, the user starts the simulation. This action triggers the system to begin the simulated wildfire scenario, putting the predefined assets and conditions into action. It can select one of the statistical simulations presented for this scenario. This use case uses the analysis of statistical data done for UC-DS-ATH-B-1. (see next subchapters 4.1.3-4.1.9)

System Runs Cascading Effects Simulation:

Once the simulation is initiated, the system processes the wildfire scenario, taking into account factors such as wind, terrain, and vegetation. This "cascading effects" component likely refers to how the fire spreads and interacts with the environment, impacting nearby assets and possibly creating secondary events (such as smoke impacting nearby towns or fire threatening infrastructure). This use case (the simulation) is based on the parameters entered in previous use cases. To define the right parameters, the user makes use of the statistical analysis of data, and then of the results of statistical simulations.

Stakeholders Receive Simulation Insights and Decision Support:

The final use case involves stakeholders, including Attica Training Exercise Observers and other relevant authorities, receiving insights from the simulation. These insights might include data on wildfire spread, asset utilisation, and decision support for handling similar real-life incidents. This output provides critical information to guide training, improve decision-making, and refine response strategies. This use case is indirectly affected by the statistical analysis of existing data, as this is an input for running the simulations.

4.1.2 INPUT DATA(FORMAT)

The data used refers to images (satellite images) and Vector data referring to maps.

- Images from Satellite using Copernicus[19] (GeoTIFF format).
- Vector files which describe elements on the map like lakes, rivers, streets, buildings (shp format).
- User input related to the simulations.

For possible sources of data see section 3.2.1 inside this document.

4.1.3 OCCURRENCE

We have analysed data from GWIS[20] for several years between 2010 and 2023.

An example statistic for years 2022 and 2023 is in the next figures (Figure 13 Wildfire occurrences in Greece 2022 and 2023):

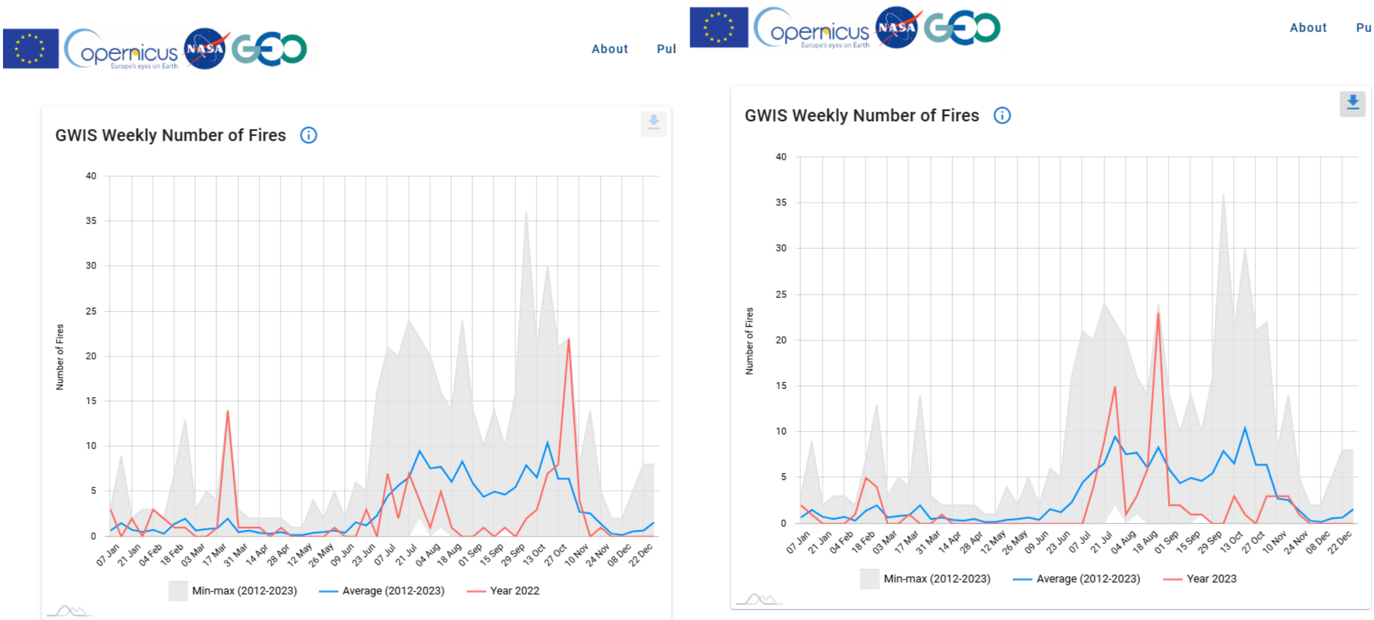


Figure 13 Wildfire occurrences in Greece 2022 and 2023

Based on the data studied, it appears that the reference of the week number in a year is important when doing simulations. This has been taken into account for simulations presented.

Poisson distribution (See Annex I-74)

Wildfire occurrences are often modelled as rare events happening randomly in time or space. The Poisson distribution is commonly used to describe the number of events in a fixed interval (time, area, etc.) when the events happen independently of each other.

A simulation with Poisson formula can be seen in the next figure (Figure 14 Simulated wildfire).

[19] Copernicus satellite data access | Copernicus

[20] https://gwis.jrc.ec.europa.eu/apps/gwis.statistics/seasonaltrend

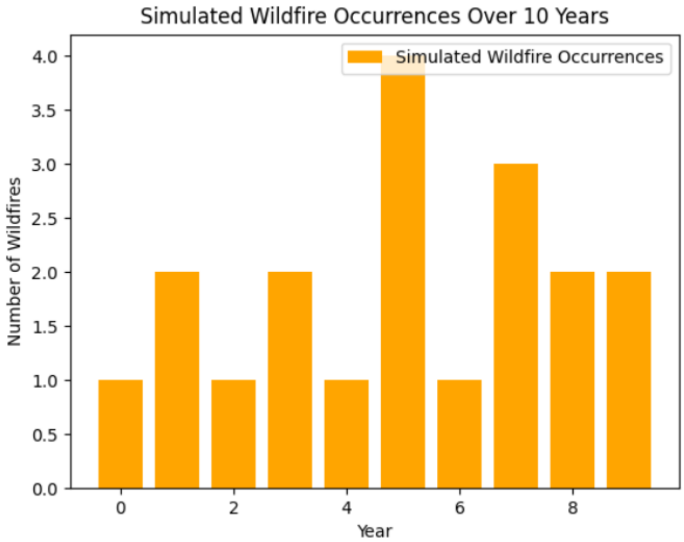


Figure 14 Simulated wildfire occurrences

Plot Interpretation:

x-axis represents the number of years (in our case 10, but can be changed as necessary)

y-axis represent the estimated number of wildfires in the period.

average_wildfire_rate = 2. In our case 2 per week. It can be changed according to the data in a region.

4.1.4 TIME

The **exponential distribution** is a continuous probability distribution that is often used to model the time between independent events that happen at a constant rate. It is particularly useful in scenarios where events occur randomly and independently over time, such as the time until a wildfire occurs in a region.

Exponential Distribution Formula (See Annex I-74)

It is used to get:

- The time between consecutive wildfires
- **The time until a fire reaches a certain size** or grows out of control
- The time between periods of weather conditions conducive to wildfires (e.g., dry, hot, and windy conditions)
- The time until a wildfire is successfully suppressed or extinguished
- The time between **lightning strikes** (a common natural ignition source for wildfires)

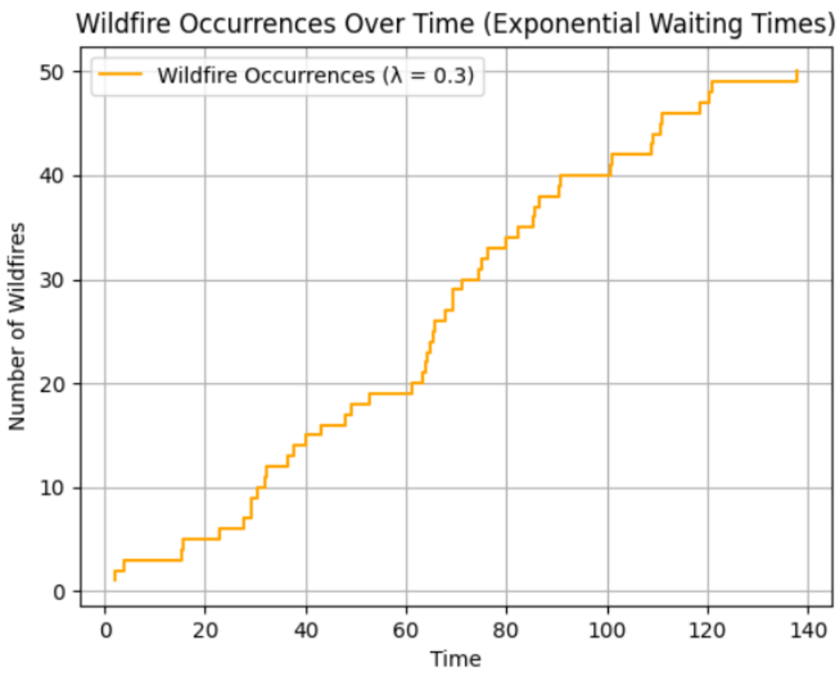


Figure 15 Simulated wildfire cumulated occurrences

Plot Interpretation:

The x-axis represents time (units can be days, months, years, etc., depending on your context). Our example presents days (150)

The y-axis represents the number of wildfires that have occurred up to a given time.

The step plot shows how the number of wildfires increases over time, with random intervals between each wildfire based on the exponential distribution.

Rate (λ): The rate of wildfire occurrence (in this example, $\lambda = 0.3$), which means on average there are 0.3 wildfires per unit time (day). This implies the expected waiting time until the next wildfire is 3-time units. It can be changed based on the data registered for a specific region.

Time to next wildfire

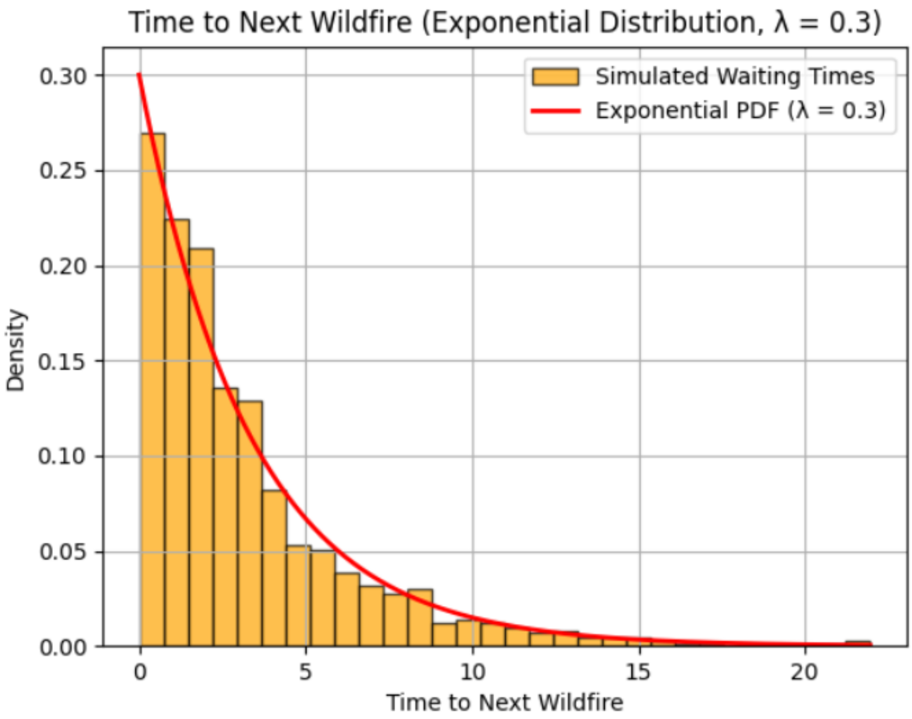


Figure 16 Time to next wildfire

The x-axis represents the waiting time until the next wildfire (e.g., in days, months, years, etc.). In our case, days are presented.

The y-axis represents the density of those waiting times.

The histogram shows how the simulated waiting times are distributed, and the red curve is the theoretical exponential distribution PDF for the same rate λ.

Time to extinguish

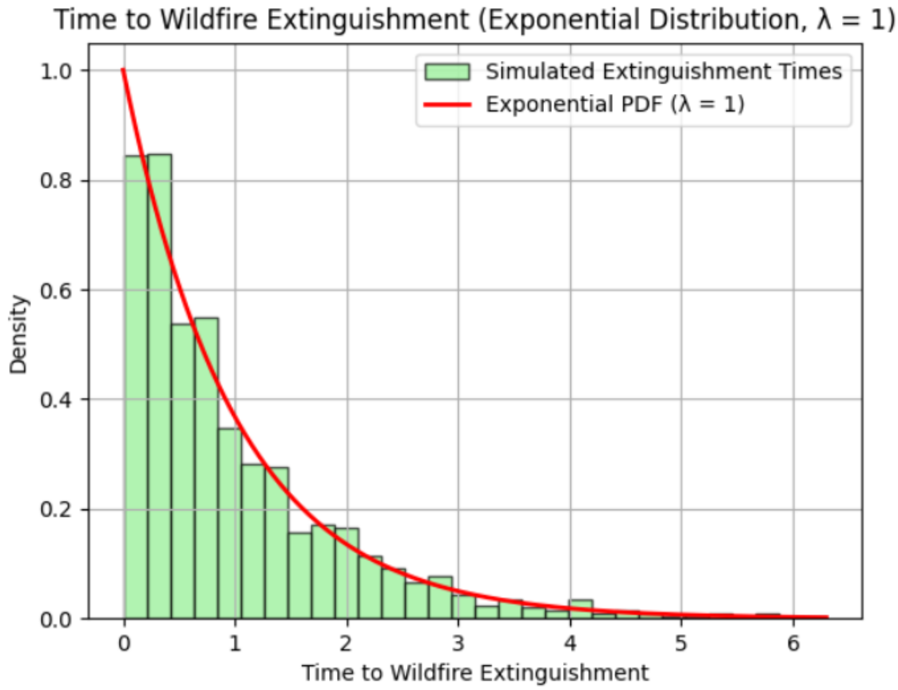


Figure 17 Time to wildfire extinguishment

Plot Interpretation:

The **x-axis** represents the **time to wildfire extinguishment** (e.g., in hours, days, or any relevant time unit).

The **y-axis** represents the **density** of those times.

The **histogram** displays the simulated data for the time it takes to successfully extinguish the wildfires, and the **red curve** is the theoretical PDF for an exponential distribution with the same rate.

The exponential distribution provides a natural model for waiting times, particularly in situations where the likelihood of an event (such as wildfire suppression) occurring is constant over time.

The rate is 1 (one wildfire in the selected period of time)

4.1.5 DURATION OF EVENTS

The **Weibull distribution** (See Annex I- 75) is used to model the **duration of events**, especially when the rate of occurrence is not constant over time. Unlike the exponential distribution, which assumes a constant hazard rate (i.e., a fixed probability of an event occurring at any given time), the **Weibull distribution** allows the hazard rate to change over time, making it more flexible for modelling wildfire durations.

Wildfire duration simulated with Weibull distribution is presented in the next figure (Figure 18 Wildfire duration)

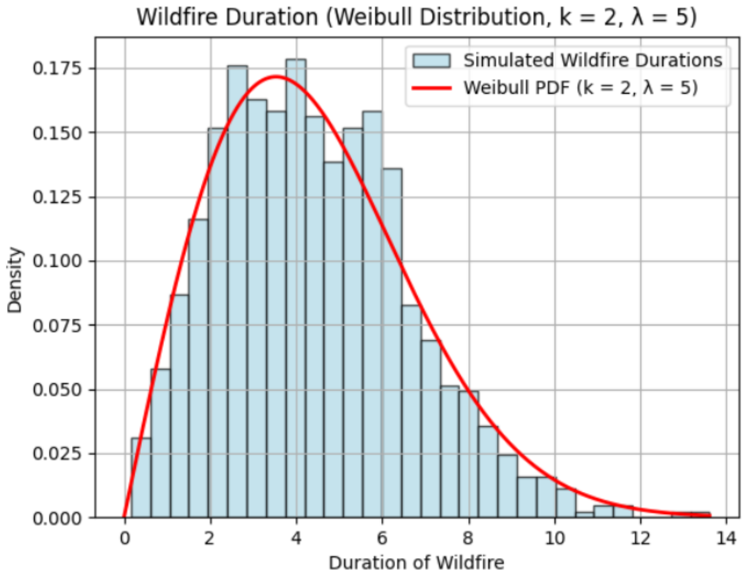


Figure 18 Wildfire duration

Plot Interpretation:

The **x-axis** represents the **duration of the wildfire** (in hours, days, etc.).

The **y-axis** represents the **density** of those durations (i.e., the likelihood of different wildfire durations).

The **histogram** shows the simulated wildfire durations, while the **red line** represents the theoretical probability density function (PDF) of the Weibull distribution with the specified shape and scale parameters.

Time until the next wildfire

To model the time until the next wildfire using the Weibull distribution, we can assume that the likelihood of a wildfire occurring changes over time rather than being constant. The Weibull distribution provides flexibility in modelling events with changing hazard rates, making it suitable for scenarios where the probability of a wildfire occurring might increase or decrease over time based on environmental conditions, such as increasing dryness or accumulating fuel loads

***Shape parameter** k:*

$k > 1$: Increasing hazard rate. This means the longer the time passes, the more likely a wildfire is to occur (e.g., due to fuel buildup, increased dryness, or worsening environmental conditions).

$k = 1$: Constant hazard rate. This case is equivalent to the exponential distribution.

$k < 1$: Decreasing hazard rate. This means that as time passes, the likelihood of the next wildfire occurring decreases (perhaps due to effective fire prevention measures).

Scale parameter λ: The scale of the distribution, affecting the typical time to the next wildfire. Larger values of λ result in longer times between wildfires on average. It will be setup considering the measure unit used for time (days, weeks, etc.)

Plot Interpretation:

The x-axis represents the time to the next wildfire (e.g., in months, years, or any other appropriate time unit).

The y-axis represents the density of those times.

The histogram shows simulated data for the time to the next wildfire, while the red line represents the theoretical probability density function (PDF) for the Weibull distribution.

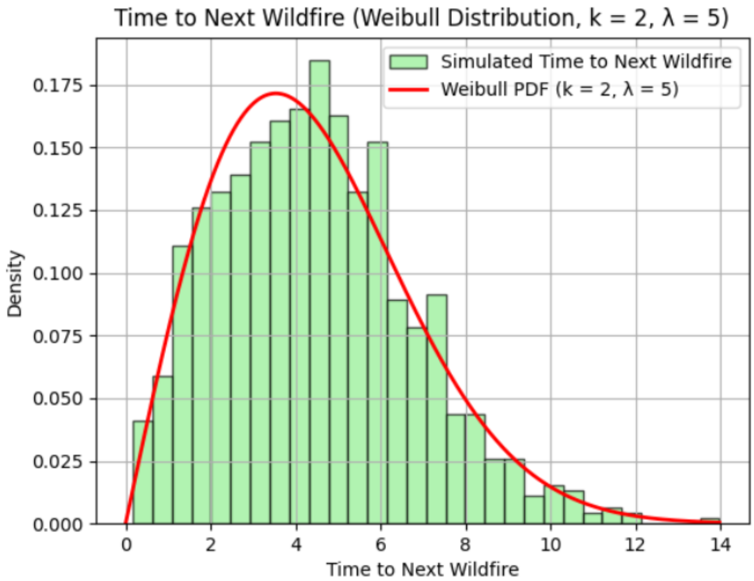


Figure 19 Time to next wildfire (Weibull)

4.1.6 FIRE SIZE

Areas burnt provide a global view about the dimension of a wildfire. The time to extinguish is dependent on the dimension of areas and the duration of the events.

We have analysed data from GWIS[21] for several years between 2010 and 2023.

An example statistic for years 2022 and 2023 is in the next figure:

[21] https://gwis.jrc.ec.europa.eu/apps/gwis.statistics/seasonaltrend

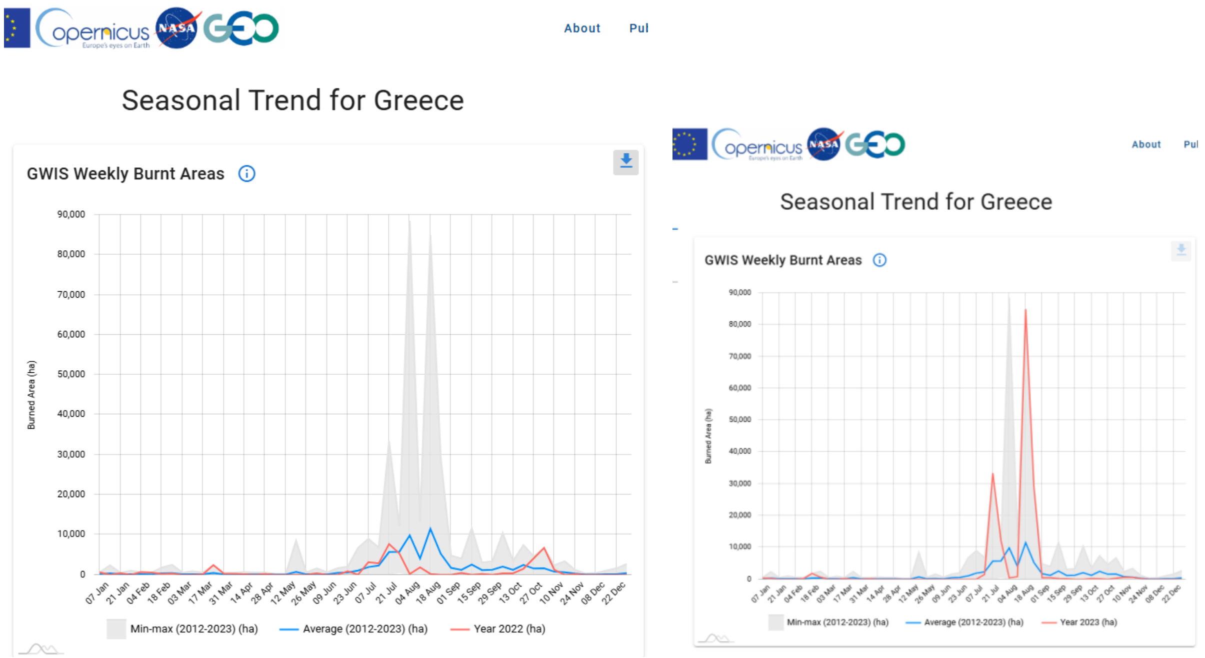


Figure 20 Wildfire burnt areas in Greece 2022 and 2023

The **log-normal distribution (See Annex I - 76)** is commonly used to model variables that are positively skewed and cannot be negative, such as wildfire sizes. In this context, wildfire size often follows a log-normal distribution because most fires are small, but a few can be extremely large, leading to a long tail in the distribution.

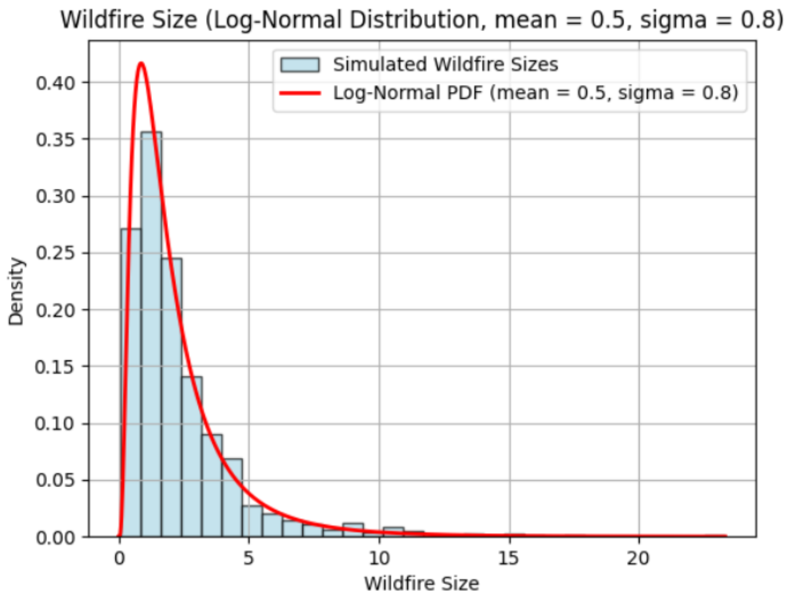


Figure 21 Wildfire size

Plot Interpretation:

The x-axis represents the wildfire size (in hectares, acres, or any other appropriate unit of area). This controls how many wildfire sizes are simulated.

The y-axis represents the density of those sizes.

The histogram shows simulated data for wildfire sizes, while the red line represents the theoretical probability density function (PDF) for a log-normal distribution.

4.1.7 CUMULATIVE TIME OF EVENTS

Let's take the case of modelling the time until the next wildfire occurs, where the hazard rate changes over time in a non-linear fashion (e.g., the probability of a fire increases initially, then levels off, and finally decreases as conditions change).

The **Generalized Gamma** distribution (See Annex I -77) allows to capture this complexity better than simpler distributions such as the exponential or Weibull.

In this case:

We assume that the occurrence of a wildfire depends on multiple independent factors (e.g., weather, fuel buildup, humidity, wind speed, human activities, etc.).

The shape parameter k of the Gamma distribution represents the number of these independent factors contributing to the wildfire's ignition.

The scale parameter θ adjusts the average time between wildfires based on these factors.

Explanation of the Parameters:

Shape Parameter k: Represents the number of factors that need to happen before a wildfire is triggered. For example, in this case, we assume there are three key factors (fuel buildup, weather, and human activity), so $k=3$.

If $k=1$, this would be equivalent to the exponential distribution, meaning that only one key factor is influencing the event, and the hazard rate is constant.

If k is greater than 1, this means that multiple stages or events need to occur before the wildfire.

Scale Parameter θ: This controls the spread or scale of the waiting times. Larger values of θ result in longer average times to the next wildfire.

θ adjusts how long it takes on average for these stages (represented by k) to be completed.

Size: This controls how many wildfires are simulated in this example (1,000 simulated wildfires).

Plot Interpretation:

The x-axis represents the time to the next wildfire (in days, months, or any other relevant unit of time).

The y-axis represents the density (i.e., the probability of different waiting times).

The histogram shows the simulated data, while the red line is the theoretical Gamma PDF, showing how the Gamma distribution models the waiting time for the next wildfire.

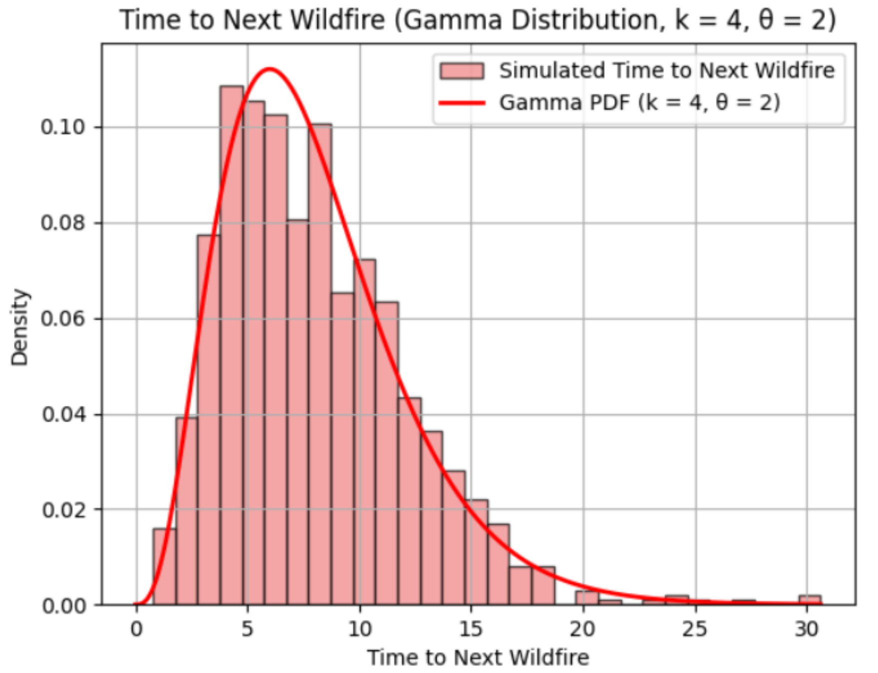


Figure 22 Time to next wildfire (Gamma)

4.1.8 EXTREME OR RARE EVENTS

The **Pareto distribution** (See Annex I-78) is commonly used to model extreme events or situations where a small number of events account for most of the total impact, which is often referred to as the "80/20 rule" (i.e., 80% of the effects come from 20% of the causes). In the context of wildfires, the Pareto distribution can be used to model extreme wildfires where a few large wildfires contribute to most of the affected area.

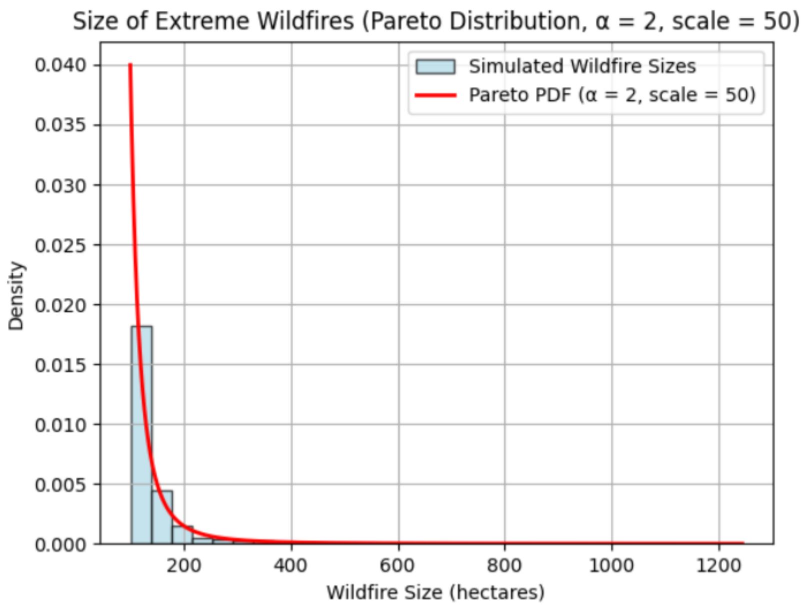


Figure 23 Size of Extreme Wildfires (Pareto)

The **Gumbel distribution** (See Annex I -78) is often used to model the distribution of the maximum (or minimum) of extreme values, such as the maximum size of a wildfire or the time until the next extreme wildfire. It is particularly useful when dealing with extreme events and is commonly applied in fields like hydrology, meteorology, and risk analysis (e.g., flood levels, extreme temperatures).

There are two types of Gumbel distributions:

- Gumbel for maxima: Used for modelling the largest values (e.g., maximum wildfire size, extreme temperatures).
- Gumbel for minima: Used for modelling the smallest values.

In this approach, we'll use the Gumbel distribution for maxima to model the size of extreme wildfires or time to the next extreme wildfire. The Gumbel distribution is flexible and allows for modelling extreme events in scenarios where the largest (or smallest) values are of interest.

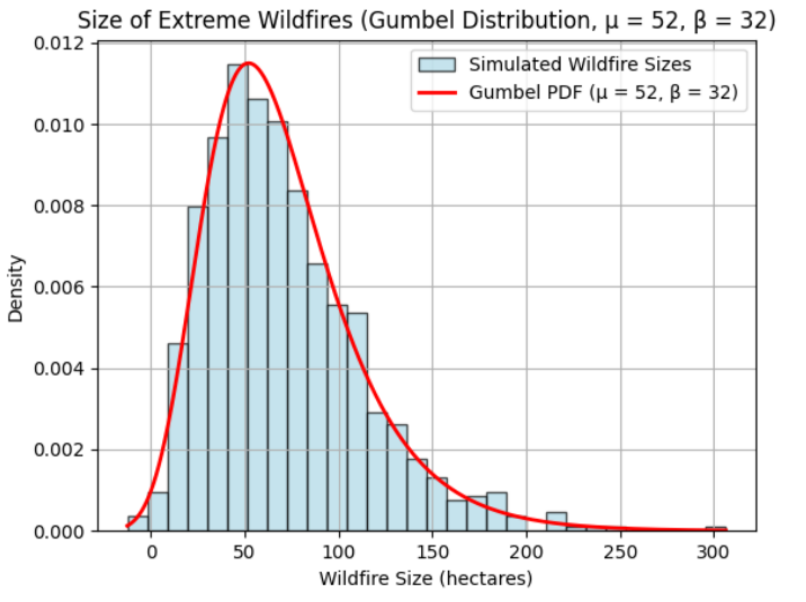


Figure 24 Size of Extreme Wildfires (Gumbel)

Plot Interpretation:

The x-axis represents the size of the wildfires (e.g., in hectares).

The y-axis represents the density of these wildfire sizes (i.e., the likelihood of a wildfire being a certain size).

The histogram shows simulated data for wildfire sizes, while the red line represents the theoretical Gumbel PDF.

4.1.9 SPATIAL DISTRIBUTION AND DIRECTIONS OF SPREAD

Geometric distribution (see Annex I - 79)

This is useful in **fire detection systems**, where sensors are deployed in a grid, and we want to estimate how many sensors (or cells) will need to be checked before detecting a wildfire. It is possible to model how many cells (discrete spatial units) we need to check before finding a wildfire ignition, assuming each cell has an independent and constant probability of containing a wildfire.

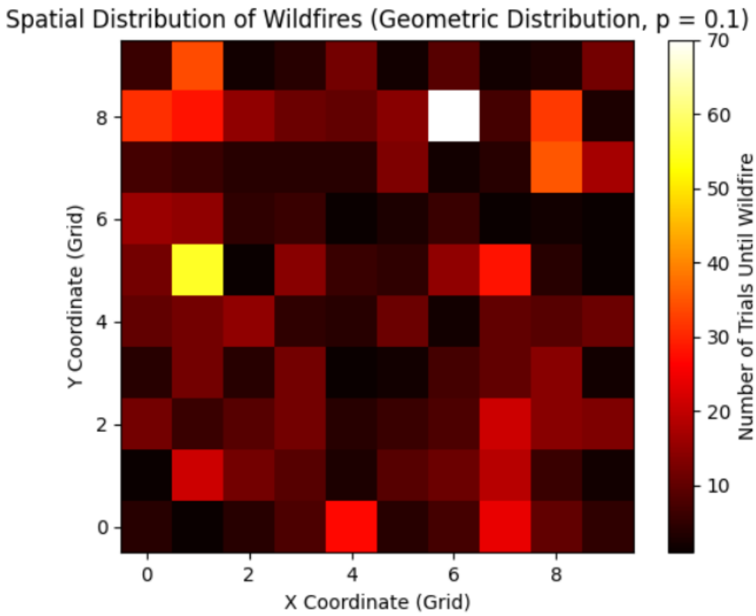


Figure 25 Spatial distribution of Wildfires

Plot Interpretation:

The x-axis and y-axis represent the grid coordinates of the region being checked for wildfires.

The heatmap shows how many grid cells need to be checked before a wildfire is found. Darker regions represent areas where more cells were checked (more trials) before finding a wildfire, while lighter regions represent areas where fewer cells were needed.

4.2 EARTHQUAKE MODELS

4.2.1 USE CASES

The use cases specific to earthquake scenario are presented in the next figure extracted from D3.6

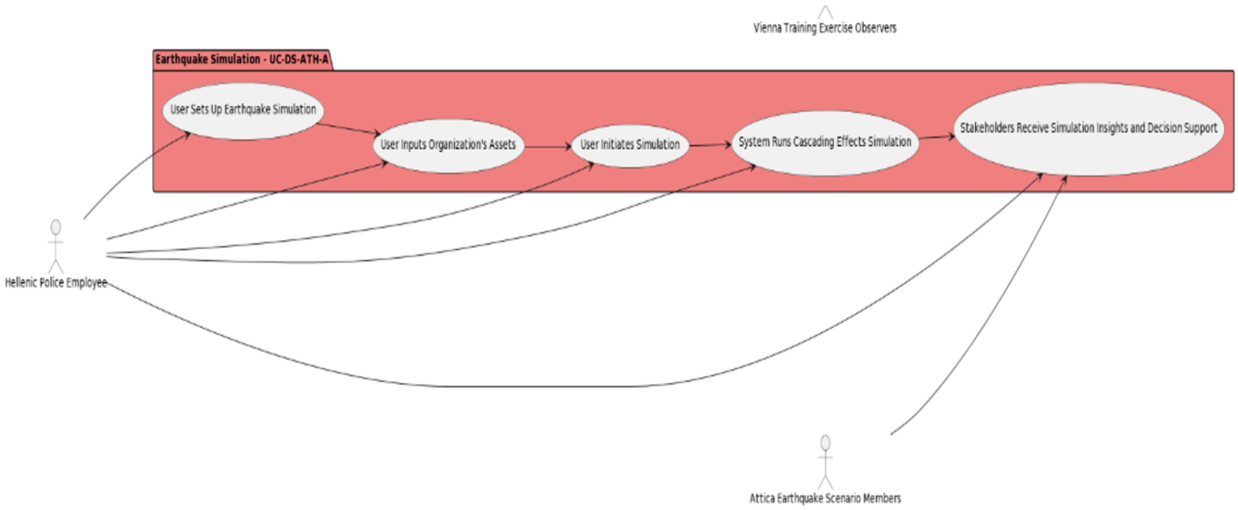


Figure 26 Earthquake simulation use cases (as in D3.6)

This diagram describes an earthquake preparedness simulation, allowing users to configure a realistic earthquake scenario, initiate the simulation, and analyse its cascading effects. Stakeholders receive detailed insights and decision support to improve their response strategies and enhance readiness for actual earthquake incidents. This setup provides a comprehensive tool for training, decision-making, and evaluating resource allocation in earthquake scenarios.

UC-DS-ATH-A-1-User Sets Up Earthquake Simulation:

In this step, a user, specifically a Hellenic Police Employee, configures the earthquake simulation. The setup would include defining parameters relevant to an earthquake scenario, such as magnitude, epicentre location, potential aftershocks, and affected infrastructure.

The statistical analysis of existing data addresses this step. This analysis gives information to the users about how to launch the simulation (using what type of simulations and what parameters).

The statistical simulations using the parameters computed refer to:

- Occurrence
- Time between earthquakes
- Extreme effects
- Magnitude
- Complex events

UC-DS-ATH-A-2-User Inputs Organization's Assets:

The user provides details about the resources and assets available to respond to the earthquake, such as police units, emergency personnel, vehicles, and medical supplies. These inputs help the system simulate a realistic response scenario based on the organization's available assets. This use case is not affected by the statistical analysis of existing data.

UC-DS-ATH-A-3-User Initiates Simulation:

Once the earthquake scenario is set up and assets are input, the user initiates the simulation, triggering the system to start the simulated earthquake event. The statistical based simulations are run. They will use the parameters selected in the previous use case, and the algorithms specific to the target values simulated. (see next subchapters 4.2.3-4.2.7). The statistical analysis of data give input to this step, considering the parameters of the statistical models.

UC-DS-ATH-A-4-System Runs Cascading Effects Simulation:

The system processes the earthquake scenario, simulating the cascading effects of the earthquake. This may include aftershocks, structural damage, potential fires, or other secondary crises that could arise following an earthquake. The cascading effects simulate the broader impact on infrastructure, populations, and available assets. This use case is indirectly affected by the statistical analysis of existing data, as the inputs may consider the statistical parameters or the result of simulations using statistical models.

UC-DS-ATH-A-5-Stakeholders Receive Simulation Insights and Decision Support:

The final step involves stakeholders, including Vienna Training Exercise Observers and Attica Earthquake Scenario Members, receiving insights and decision support from the simulation. These insights may cover the impact on affected areas, asset deployment efficiency, and recommendations for handling similar real-world incidents. This information is critical for improving preparedness, refining response plans, and enhancing training for future earthquake events. This use case is indirectly affected by the statistical analysis of existing data, considering that the statistical analysis gives inputs for the whole simulation process.

4.2.2 INPUT DATA (FORMAT)

For possible sources of data see chapter 3.2.2 inside this document. The formats of the different sets follow:

- Seismic Model representations (XML)
- Vulnerabilities (CSV)
- Geospatial data (GeoJSON)
- Hazard and risk data (Natural hazard Risk Markup Language -NRML)

4.2.3 OCCURRENCE

For occurrence, Poisson and exponential distribution are commonly used.

Poisson distribution (See Annex I - 74)

An example is in the next figure:

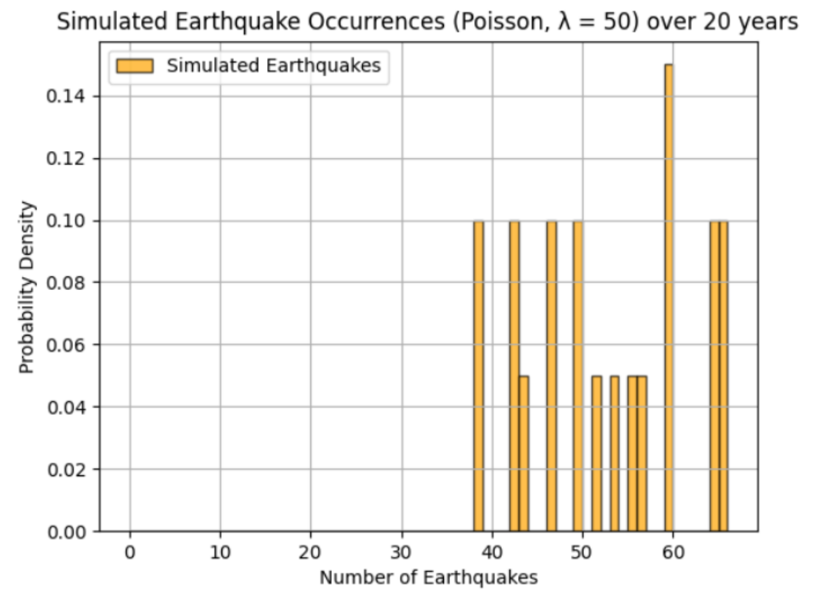


Figure 27 Simulated earthquake occurrences (Poisson)

Interpretation:

The x-axis of the histogram represents the number of earthquakes that occurred in a year.

The y-axis represents the density, or the probability of observing a certain number of earthquakes each year.

It shows how likely it is to observe a certain number of earthquakes per year based on the Poisson distribution, with the average number centred around the value of λ=50.

Customisation:

The number of years can be increased to simulate a longer time by changing the years variable λ.

4.2.4 TIME BETWEEN EARTHQUAKES

The **Exponential distribution (See Annex I -74)** is often used to model the **waiting time between events** in a Poisson process.

An example is in the next figure:

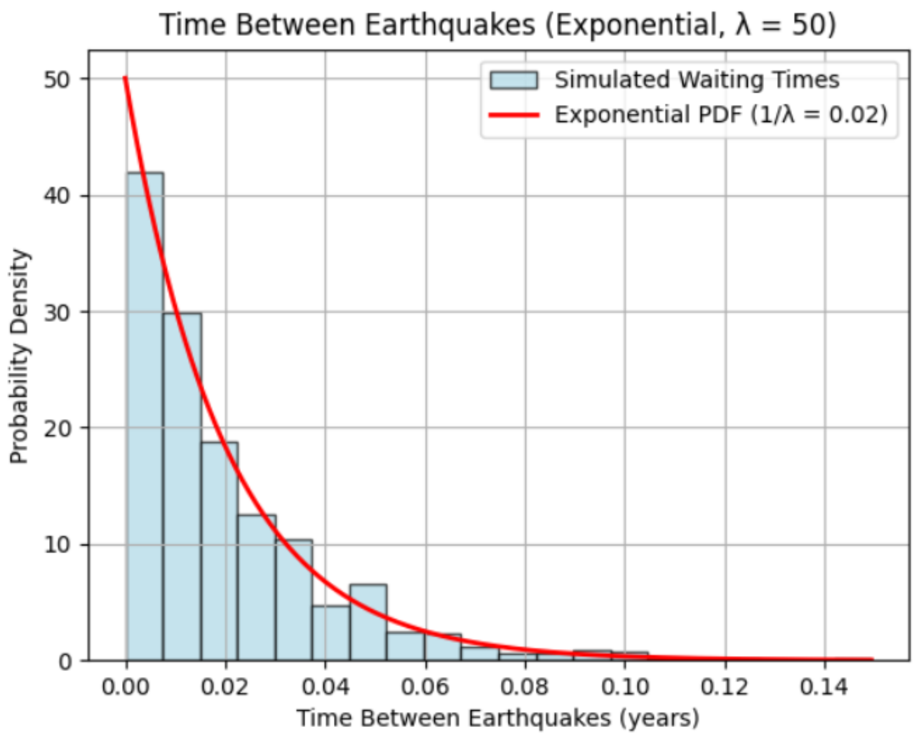


Figure 28 Time between earthquakes occurrences (Exponential)

Interpretation:

The x-axis of the histogram represents the time between consecutive earthquakes (in years).

The y-axis represents the density, or the probability of observing certain waiting times between consecutive earthquakes.

The red curve shows the theoretical Exponential distribution for comparison. It indicates that short waiting times between earthquakes are more likely than long waiting times.

Customisation:

The rate λ can be adjusted to simulate different average earthquake occurrence rates. For example, setting λ=50 models a region with an average of 50 earthquakes per year.

4.2.5 EXTREME EFFECTS

Gumbel Distribution (See Annex I -78)

The example of usage in our use case is presented below.

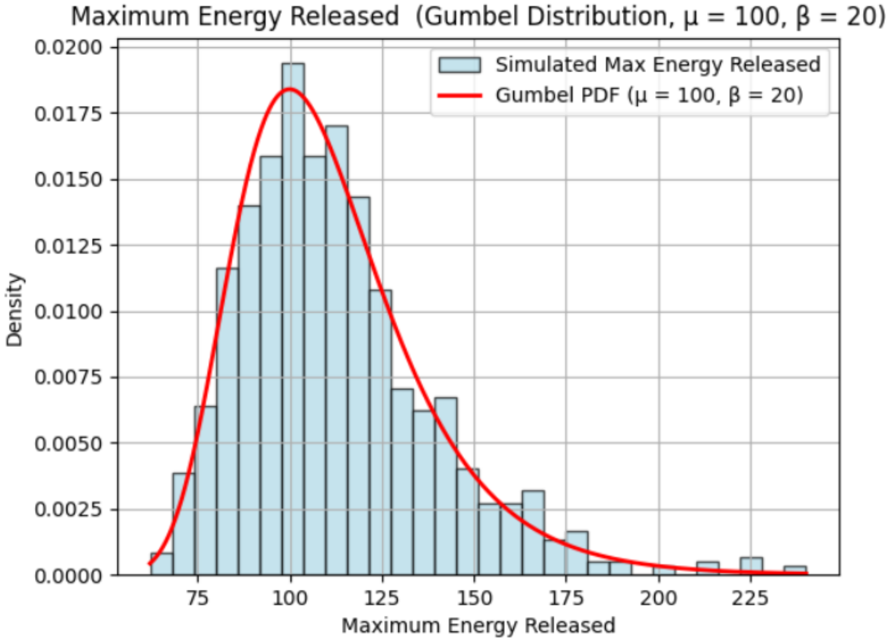


Figure 29 Maximum energy released Gumbel)

Plot Interpretation:

The x-axis represents the maximum energy released by earthquakes.

The y-axis represents the density, or the probability of observing a certain amount of energy released.

The histogram shows the simulated maximum energy values, and the red line shows the theoretical Gumbel PDF.

Customisation:

Location Parameter μ can be adjusted to reflect the average or typical maximum energy release by the largest earthquakes in a region.

For example, μ=150 means that the largest earthquakes in the region typically release more energy.

Scale Parameter β can be adjusted this to change the spread of the distribution. Larger values of β indicate more variability in the maximum energy released by earthquakes.

4.2.6 MAGNITUDE

The **Exponential distribution** is often used to model the **waiting time between events** in a Poisson process (See Annex I -74).

An example is in the next figure:

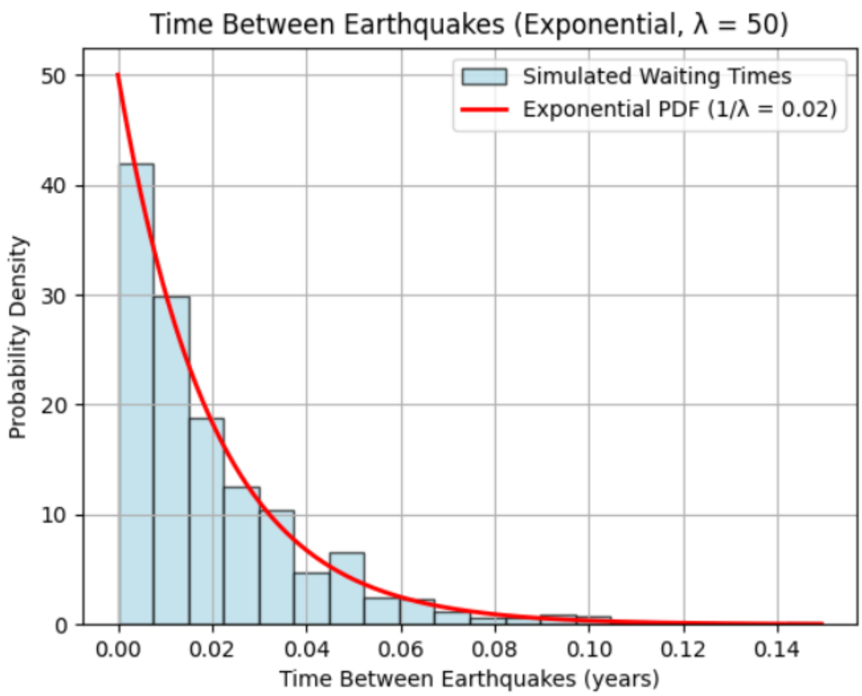


Figure 30 Time between earthquakes occurrences (Exponential)

Interpretation:

The x-axis of the histogram represents the time between consecutive earthquakes (in years).

The y-axis represents the density, or the probability of observing certain waiting times between consecutive earthquakes.

The red curve shows the theoretical Exponential distribution for comparison. It indicates that short waiting times between earthquakes are more likely than long waiting times.

Customisation:

The rate λ can be adjusted to simulate different average earthquake occurrence rates. For example, setting λ=50 models a region with an average of 50 earthquakes per year.

4.2.7 CUMULATIVE TIME AND COMPLEX EVENTS

This is modelled using **Gamma distribution**.

The example of usage in our use case is presented below.

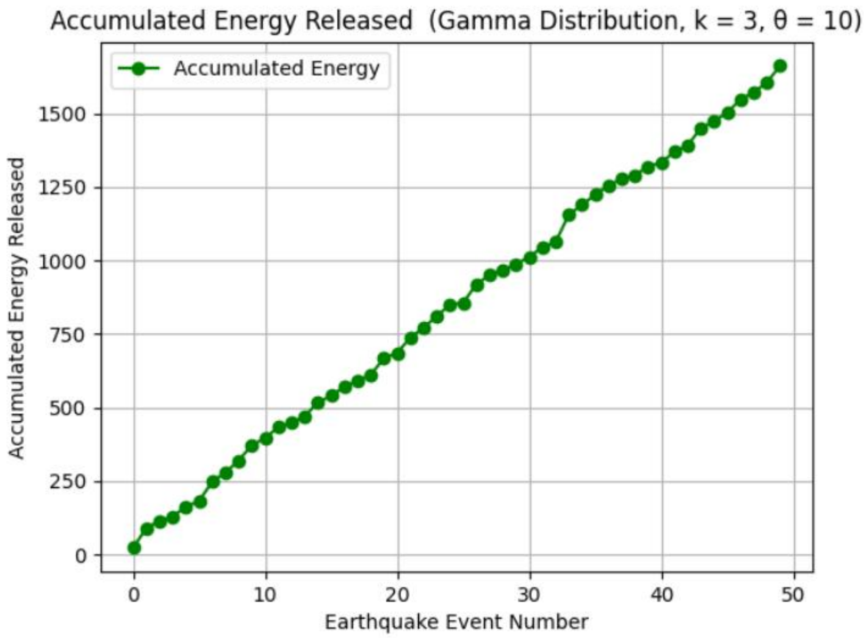


Figure 31 Accumulated energy (Gamma)

Plot Interpretation:

The x-axis represents the earthquake event number, which corresponds to the number of earthquakes in the sequence.

The y-axis represents the accumulated energy released.

The plot shows the total energy released by all the earthquakes in sequence. The energy release increases over time as more earthquakes occur.

Customisation:

Shape Parameter *k* can be adjusted to represent how many independent processes contribute to the energy release in each earthquake. A larger *k* indicates more factors influencing energy release.

For example, k=3 might indicate that the energy release depends on three independent factors such as fault stress, slip, and tectonic pressure.

Scale Parameter θ controls the average energy released by each earthquake. A higher θ value means that each earthquake releases more energy on average.

4.3 HEATWAVE MODELS

4.3.1 USE CASES

The use cases specific to earthquake scenario are presented in the next figure extracted from D3.6

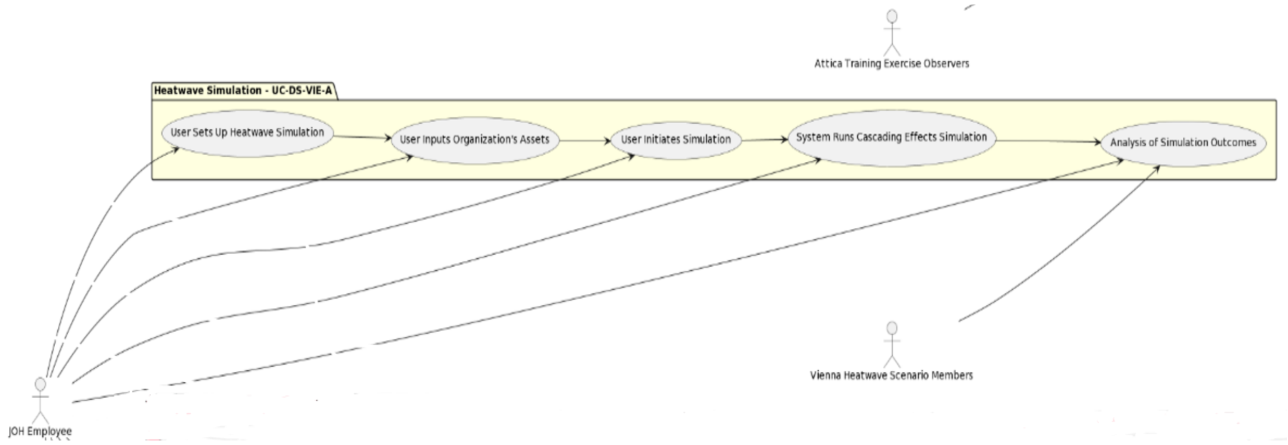


Figure 32 Heatwave simulation use cases (as in D3.6)

This diagram describes a heatwave preparedness simulation, allowing users to configure a realistic heatwave scenario, run the simulation, and analyse the cascading effects. Stakeholders gain valuable insights into the impacts of heatwaves on resources, infrastructure, and public health, enabling them to refine response strategies and enhance readiness for future incidents. This simulation provides a comprehensive tool for training, resource allocation, and decision-making in heatwave response scenarios.

UC-DS-VIE-A-1-User Sets Up Heatwave Simulation:

In this first step, a user, specifically a JOH Employee, configures the heatwave simulation environment. This setup likely includes defining parameters for the simulation, such as temperature thresholds, geographic scope, duration, and affected populations or regions. The parameters defined uses the results of the statistical analysis of existing data. The statistical parameters are identified by visualizing existing data. There is no automated process, considering the nature of data. The data studied is from https://www.zamg.ac.at/cms/en/climate/climate-overview. Considering that for this specific use case there aren't significant statistical data, general purpose statistical models might be selected to obtain more information.

UC-DS-VIE-A-2-User Inputs Organisation's Assets:

The user inputs details about the assets available to respond to the heatwave. This may include cooling centers, medical staff, transportation, water distribution systems, and any other resources that could be deployed in response to extreme heat conditions. These inputs help the system simulate a realistic response based on the organization's capacity. This use case is not affected by statistical analysis.

UC-DS-VIE-A-3-User Initiates Simulation:

After setting up the scenario and inputting organisational assets, the user initiates the simulation, which triggers the system to start the simulated heatwave event. Statistical analysis of data is indirectly used, when parameters for simulation are established.

UC-DS-VIE-A-4-System Runs Cascading Effects Simulation:

The system processes the heatwave scenario, taking into account the effects of prolonged high temperatures on infrastructure, public health, and resources. Cascading effects might include increased health emergencies, demand for water and power, strain on medical resources, and other secondary impacts due to the heatwave. This use case is not affected by statistical analysis.

UC-DS-VIE-A-5-Analysis of Simulation Outcomes:

The final step involves analysing the simulation outcomes. Stakeholders such as Vienna Heatwave Scenario Members receive insights from the simulation. These insights may include data on asset utilisation, effectiveness of response strategies, impact on affected populations, and potential improvements for future real-world scenarios. This analysis supports better preparedness and decision-making for heatwave-related incidents. This use case is indirectly affected by statistical analysis, considering that input parameters when established may use the results of statistical analysis.

4.3.2 INPUT DATA(FORMAT)

The data used refers to images (satellite images) and Vector data referring to maps.

- Images from Satellite using Copernicus (GeoTIFF format).
- Vector files which describe elements on the map like lakes, rivers, streets, buildings (shp format).
- Historical data in CSV format

For possible sources of data see chapter 3.2.3 of this document.

For this type of events, we do not select a specific distribution, but rather consider the usage of general-purpose models to filter data for outliers, or to detect min, max and average values.

4.4 MAN-MADE DISASTER MODELS

4.4.1 USE CASES

The use cases specific to earthquake scenario are presented in the next figure extracted from D3.6

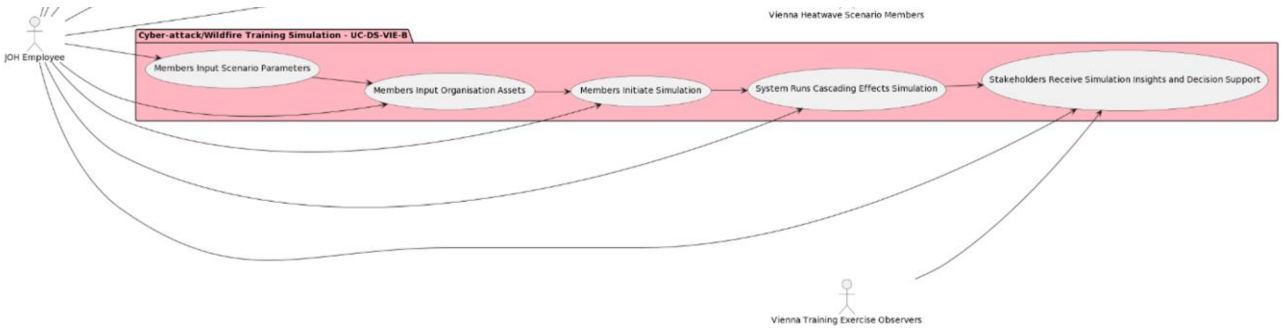


Figure 33 Man-made disaster simulation use cases (as in D3.6)

4.4.2 INPUT DATA (FORMAT)

The data used refers to images (satellite images) and Vector data referring to maps.

- Images from Satellite using Copernicus (GeoTIFF format).
- Vector files which describe elements on the map like lakes, rivers, streets, buildings (shp format).
- Historical data in CSV format

The man-made disaster models will use also data generated by Monte Carlo simulations.

For the data sources please refer to chapter 3.2 inside this document

For this type of events we do not select a specific distribution, but rather consider the usage of general purpose models to filter data for outliers, or to detect min, max and average values.

Also, there is no statistical analysis of data, as the existing data is not relevant for this scenario.

5. CONCEPTUAL MODELS

5.1 OVERVIEW

Conceptual models guide the digital twin's ability to simulate scenarios**,** predict risks, and optimize resource allocation, significantly improving disaster management. By continuously refining the models based on real-world data, the SCDT becomes an ever-evolving system capable of handling both natural and man-made disasters with greater precision and resilience.

The conceptual models for the different disaster scenarios—wildfire**,** earthquake**,** heatwave, and man-made disasters—play a critical role in guiding the **SCDT** by structuring the system's approach to understanding, simulating, and responding to various hazards. Each conceptual model provides a framework that allows the SCDT to tailor its operations based on the unique characteristics of each disaster type.

Conceptual models are essential for ensuring that the SCDT functions effectively by defining the structure and dynamics of each disaster scenario. These models:

- Outline key variables (e.g., weather conditions, infrastructure status, population movement) that the SCDT must monitor in real-time.
- Provide the underlying logic for predictive simulations, guiding how the disaster might evolve and identifying high-risk areas.
- Ensure resource allocation, risk management, and response strategies are tailored to the specific demands of each type of disaster.
- Enable training and learning, refining the system over time through feedback loops and post-event analysis.

5.2 SMART CITY DIGITAL TWIN CONCEPT

The Smart City Digital Twin (SCDT) is an important part of the PANTHEON project, and one of its main objectives. In this chapter we present the definitions used, the possible approaches and the selected approach.

5.2.1 DIGITAL TWIN IMPORTANCE

The core of the PANTHEON project is developed around the concept of a Smart City Digital Twin. According to the DoA (PANTHEON - Consortia, 2023):

> *"The SCDT will monitor the community conditions to reflect the impacts of decisions and actions to be taken. The main operations of the proposed platform will be: to prevent and reduce the impact of the disasters; to respond to the events in a timely, fast and effective manner; to create a new and safer environment for victims; to determine vulnerable regions that exist on the fault line based on local disaster models; Inform and raise awareness of the society about disaster hazard and risk; improvement of disaster management capacity and capability; utilization of all sensing sources and devices for situation awareness; Impact planning and assessment and optimized decision-making through simulations and intelligent data processing; To accomplish disaster risk reduction. Simulations, Data and conceptual models will focus on the management of multi-hazards and critical system interactions. "*

Based on the objective mentioned, and considering the inputs from deliverable D3.7 (Pantheon Consortia, 2024), we can also mention that the SCDT is beneficial because it offers a real-time, data-driven, predictive, and adaptive system that integrates various aspects of disaster management. It helps improve situational awareness, optimize resource allocation, provide decision support, and reduce operational costs. Additionally, it enhances risk management, supports collaboration, and enables effective training.

A digital twin offers a dynamic, real-time virtual replica of a physical environment, asset, or system. This model can continuously receive real-world data, simulate outcomes, optimize resource usage, and inform decision-makers.

The PANTHEON project's aim is to use the SCDT for several important actions such as:

- **Real-time monitoring and awareness.** By continuously receiving live data from sensors and other data sources, the digital twin keeps an up-to-date digital model of the real world.
- **Predictive analytics and simulations.** The SCDT can run simulations and predictive models based on real-time data and historical information. This allows decision-makers to explore what-if scenarios and predict the future behaviour of systems under different conditions.
- **Risk management and mitigation.** The SCDT will serve as a central tool for risk assessment and management, allowing continuous monitoring of current risks and simulating the impact of potential future risks.
- **Optimizing resource allocation.** This will be possible by providing real-time visibility into resource availability and requirements,
- **Enhanced decision-making support.** It will integrate data from multiple sources and simulate future outcomes. Decision-makers can rely on the digital twin to offer data-driven insights and recommendations in real-time.
- **Continuous learning and system improvement**. The SCDT will learn over time using ML components, by continuously integrating new data and outcomes from past simulations or real-life events. This helps improve the accuracy of predictions and models.
- **Collaboration and communication.** The SCDT will display visualizations and simulations that are easy to understand and share.
- **Testing and training.** The SCDT allows emergency responders and decision-makers to train in virtual environments, testing different strategies without real-world consequences.

5.2.2 POSSIBLE APPROACHES

For complex systems, the creation and usage of digital twins can be done in different ways. A DT can be the central point in a system or can be considered itself as the whole system.

We are presenting here the aspects of two possible implementations:

- SCDT as the whole system
- SCDT as the central point

Based on advantages and challenges of the two approaches, we have selected SCDT as the whole system. The motivation will be presented in this chapter.

5.2.2.1 *Digital Twin as the Whole System*

In this approach, the entire system represents the digital twin, and all the components (data flows, simulations, resource allocation, communication, risk management, etc.) are parts of the SCDT.

It is continuously updated with real-time data and providing actionable insights.
The most important advantages of this approach are:

- **Holistic representation**: Treating the entire system as the digital twin makes it clear that the digital twin is a comprehensive system that integrates all the various components - data aggregation, simulation, decision support, and resource management.
- **All components as parts of the twin**: Since the digital twin mirrors reality, it encompasses all the functional components (real-time data, simulations, AI decision support, resource allocation) in its structure. By treating the entire system as the digital twin, we emphasize that the twin is a complete, interconnected model of the physical world.
- **No duplication**: We avoid redundancy or the potential confusion of having a separate "digital twin" component inside the diagram, because all parts are inherently part of the twin.
- **System thinking**: It highlights the integrated nature of the system and shows how every element (data aggregator, simulations, decision support, etc.) is equally important in forming the digital twin.

This approach has also disadvantages, such as**:**

- **Conceptual overlap**: By treating the whole system as the digital twin, it may be harder to distinguish between different functional roles within the system. For example, the data aggregator or decision support components might be seen as operational layers rather than parts of a distinct digital twin architecture.
- **Complexity**: If every part is considered part of the digital twin, it might complicate the architecture by making the twin seem overly complex, which could obscure the distinct functionalities of different modules (e.g., simulation versus decision-making).

5.2.2.2 *Digital Twin as the central component in a complex system*

In this approach, the **SCDT** is treated as a central component in the system, interacting with other system components such as simulations, decision support, communication, and resource allocation.

The advantages of this approach could be:

- **Clarity**: By having a clear central component labelled as the digital twin**,** it becomes easier to explain that the digital twin is a specific virtual representation of the physical world, and interacts with other system components like simulations, decision support systems, and resource allocation.
- **Modularity**: This approach clearly distinguishes between different modules (e.g., data aggregators, AI decision support**,** resource allocation) and the digital twin itself. This modular approach helps in understanding how each component feeds into or depends **on** the digital twin.
- **Simplifies Understanding**: If stakeholders are unfamiliar with the concept of a digital twin, this method simplifies the explanation by showing the twin as a central entity that collects data from other components and sends outputs (e.g., predictions, visualizations, decisions) back to those components.
- **Easier Maintenance and Scalability**: Conceptually separating the digital twin from other parts of the system makes it easier to update or modify specific aspects of the twin without affecting the entire system.

The disadvantages of this approach could be:

- **Over-simplification**: Placing the digital twin as a central component could **oversimplify its role**, presenting it as just another module rather than the central system that connects all other components.
- **Potential Duplication**: It is possible to end up duplicating functionality in the diagram.

5.2.3 PANTHEON APPROACH

Considering the advantages and the disadvantages enumerated before, we have selected the approach with the SCDT as the main global system. This is reflected also in the proposed architecture D3.7 (Pantheon Consortia, 2024) and illustrated in Figure 3 The functional view of the PANTHEON architecture.

5.2.3.1 *Motivation*

The main reasons for selecting SCDT as the whole system are:
We want to emphasize that every component in the system (e.g., simulations, decision support, data aggregation) is part of the digital twin's virtual representation.

We are presenting to an audience to which we want to showcase how all components work together to create a unified system. The system is deeply interconnected, and the real-time data, simulations, and resource management are closely integrated to the point where it's difficult to separate them conceptually.

The objectives of the project all conduct to this approach.

5.3 CONCEPTUAL MODELS COMMON TO ALL SCENARIOS

In this chapter we present the conceptual models, which are common to all scenarios. We'll start with the explanation of what conceptual models are in the scope of this project, and how we are treating them.

5.3.1 OVERVIEW

The Models used and presented in the next diagrams are primarily conceptual or workflow models that help in visualizing processes, systems, and interactions between components in a disaster management framework. They are used to represent:

- Preparation
    - Resource allocation
    - Risk management processes
    - Communication flows
    - Evacuation planning
- Training
    - Training system for emergency scenarios
- Simulation
    - Simulation-based decision support
- Operation
    - Operational structure
    - Real-time response in actual emergencies.
- Post event

This approach is complementary with the **Mathematical Models** which are subject to next more technical deliverables like D4.3 - Enhanced Intelligence & Self-adaptive Simulations, D4.4 – Virtual Representation, Visualisation and User Interfaces, D5.1 PANTHEON DSS & Assets Management, D5.2 - Evaluation, Risk Mitigation and AI Governance.

The **Mathematical Models** are used to simulate physical phenomena related to disasters, such as the spread of a heatwave, its impact on infrastructure and population, or the progression of other hazards like earthquakes, fires, or man-made disasters. Mathematical models use equations, statistical methods, and algorithms or machine learning systems to describe how a system behaves over time.

Example mathematical models include heat transfer models, population health models, and energy consumption models during heatwaves.

Their purpose is to predict outcomes based on data, provide quantitative analysis, and simulate possible scenarios for informed decision-making.

Presented conceptual models can help organize and coordinate how teams will respond to the outputs of mathematical models. For example, once a mathematical model predicts the heatwave's intensity, a communication flow diagram can be used to ensure that warnings reach vulnerable populations.

The Mathematical models which will be subject of future work, can provide precise predictions about the likely severity of a heatwave or a surge in energy demand, which can then be used in the presented models to refine evacuation plans, communication strategies, and resource allocation workflows.

The generic diagram for a conceptual model used in the PANTHEON use cases is presented in the next figure:

(Figure 34 Conceptual models' diagram)

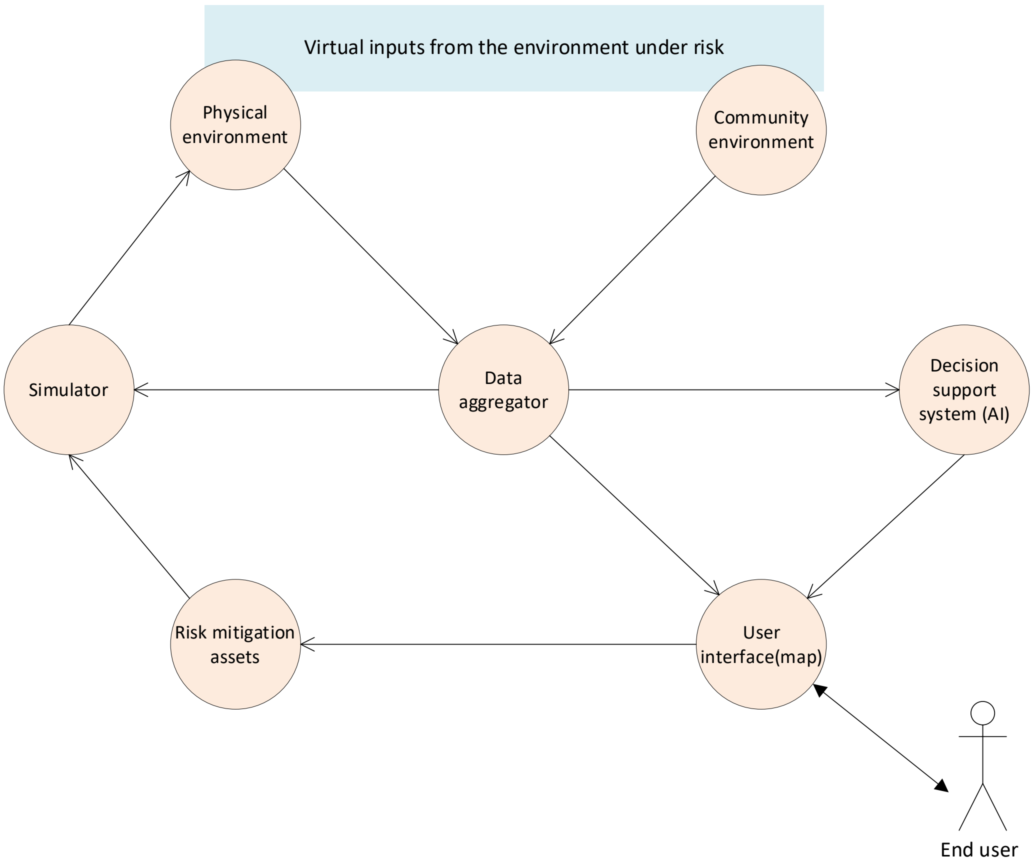


Figure 34 Conceptual models' diagram

In this figure we have as main components:

- **Physical environment.** It includes physical, tangible factors (e.g., infrastructure, geographical data, weather).
- **Community environment.** represents the social factors, including populations, behavioural data, and community responses.
- D**ata aggregator**. The data from both the physical and community environments flows into the data aggregator, which is central to the system. The aggregator processes and combines these diverse inputs into a format usable for decision-making. It ensures that the virtual model is continuously updated with the latest data from sensors, field reports, and external systems.
- **Simulator** part of the predictive modelling capabilities of the digital twin allow the system to simulate possible future states (based on current conditions).
- **Decision support systems (AI)** shown in the diagrams is also part of the predictive modelling capabilities of the digital twin. This is an artificial intelligence-driven system that helps in making decisions by analysing data, possibly offering predictions, insights, or recommendations to mitigate risks. It allows the decision-makers assess what-if scenarios or get advises on the future actions.
- **User Interface (Map)**: The processed data, simulations, and AI insights are presented through a user-friendly interface, typically a map that provides visual context to the risk scenario, showing affected areas and resources. It shows how the digital twin interacts with humans, including emergency responders, decision-makers, and the public. These tools visualize the virtual twin and allow users to interact with the system in a meaningful way.

The disaster management using a SCDT has a specific flow including:

- Early Preparation
  - Resource allocation
  - Risk management processes
  - Communication flows
  - Evacuation planning
- Training
  - Training system for emergency scenarios is used to prepare users (such as emergency managers, responders, or policy makers) for real-life emergencies by exposing them to simulated scenarios. The outcome refers to users (emergency managers, responders) who gain familiarity with potential emergency situations and decision-making strategies before they face real-world scenarios. The flow considers:
    - Trainees are exposed to different emergency scenarios (e.g., flooding, fire, mass evacuation).
    - The decision support system (AI) provides real-time feedback based on their decisions.
    - The system logs how users respond, helping them learn from mistakes and successes.
- Simulation
  - Simulation-based decision support is used to simulate disaster scenarios for planning, analysis, and risk assessment in a non-emergency context. This stage helps refine strategies and optimize resources for potential future disasters. It prepares organizations by helping

them create disaster response strategies, identify critical risk points, and optimize resource deployment in a proactive manner. The flow is:

    - Emergency planners and analysts input real-world data into the system (such as geographic features, population density).
    - The system simulates disaster impacts and potential responses, helping decision-makers identify weaknesses and opportunities in their existing plans.
    - Risk mitigation strategies are developed and tested before real disasters occur.

- Operation
  - Operational structure
  - Real-time response in actual emergencies is used to manage and respond to real-time emergencies using live data inputs. The real-time emergency response system helps manage actual disaster scenarios, making it easier for decision-makers to allocate resources, coordinate evacuation efforts, and save lives during a crisis. The flow of actions is:
    - When an emergency occurs, the system starts pulling in real-time data (from environmental sensors, community feedback, and available assets).
    - Decision-makers rely on the AI decision support system and the map interface to make immediate, data-driven decisions about where to allocate resources and how to mitigate risks.
    - The system provides continuous feedback, allowing for adjustments to be made in real time.
- Post event actions

5.3.2 CONCEPTUAL MODELS USAGE FLOW:

The systems presented function together in a feedback loop. Training provides readiness, simulations optimize planning, and real emergencies refine future iterations, forming a cyclical approach to improving disaster management over time

The models enumerated are used for the following reasons:

- **Early preparation.** (Resource allocation, Risk management, Communication planning)
- **Training** ensures personnel are ready and familiar with the systems before an emergency.
- **Simulation** tests and improves disaster preparedness strategies, ensuring that the system is optimized for a real-world scenario.
- **Real-time emergency response** acts as the operational component during an actual disaster, using the data and strategies developed in the previous stages to save lives and mitigate damage.
- **Post event actions**

Now in detail:

Early preparation allows the user to study the past events and be prepared for future events and actions.

Training Informs Simulation:

The **training** system allows users to practice responding to **simulated** emergencies. The data collected during training can be used to refine disaster modelling efforts by showing how responders interact with different scenarios. For instance, human decisions made during training may reveal common response patterns or potential flaws in strategies, which can then be addressed in simulated planning.

Simulation optimizes real emergency response:

The **simulation** phase during which planners run hypothetical scenarios, directly informs the **real-time response** system. By identifying key risk factors, resource allocation plans, and mitigation strategies in advance, the real emergency response system can leverage those insights. Simulation results provide the groundwork for making quick, informed decisions when an actual emergency happens.

Real emergencies provide data for future training and simulation:

After a real emergency occurs, data collected by the **real emergency response** system can be used to improve both **training** and **simulation**. Real-world data helps refine the accuracy of future **simulated scenarios** and updates the decision-making models that trainees use during **training**. Continuous learning from real events helps the entire system evolve and become more effective over time.

Post event actions are considered after the event is produced. The following actions can be considered:

- **Damage assessment and impact analysis**: After the event, the digital twin can be used to assess the full extent of the damage to infrastructure, resources, and the environment. By integrating real-time data from sensors and field reports, it provides a comprehensive overview of which areas are most affected and require the most urgent attention.
  The digital twin can also compare pre-event and post-event conditions, providing an immediate understanding of the scope of the impact (e.g., infrastructure damage, changes in land use, population displacement).
- **Resource recovery and allocation for reconstruction**: The digital twin helps plan and optimize the allocation of resources for recovery and reconstruction. It can simulate various recovery strategies, ensuring that resources (e.g., construction materials, medical supplies, skilled labour) are used efficiently and allocated where they are most needed.
- **After-action reporting and event analysis**: The digital twin can generate after-action reports (AARs) based on the data collected during the disaster. These reports summarize the decisions made, the actions taken, and the outcomes achieved, allowing responders and decision-makers to reflect on what went well and what could be improved.
- **Learning and continuous improvement**: The digital twin enables learning from past events by analysing what happened and comparing it to predicted outcomes. These insights can be used to update simulation models and decision-making algorithms for future events, ensuring continuous improvement in disaster preparedness and response.
- **Community recovery planning**: The digital twin can assist in community recovery by simulating various rebuilding strategies, helping planners make decisions about where and how to rebuild. For example, the twin can simulate the future resilience of infrastructure designs, allowing communities to rebuild in ways that reduce future risks.
- **Risk re-assessment and updated risk models**: Based on the outcomes of the event, the digital twin can update risk models to reflect the latest understanding of vulnerabilities and hazards. For example, it might highlight weaknesses in infrastructure or population behaviour that were not previously accounted for. These updated risk models can be integrated into future risk assessments and simulation scenarios, ensuring that future disaster response strategies are built on improved understanding of risks.
- **Collaboration and stakeholder engagement**: The digital twin provides a platform for collaboration between different stakeholders during the recovery phase. By sharing post-event data and simulation outputs, the digital twin helps government agencies, non-profits, and private companies work together on recovery efforts.

5.3.3 PREPARATION

5.3.3.1 *Resource allocation*

The digital twin as a whole system can greatly enhance resource allocation by utilising the same components that are part of the simulation framework. Key components like the data aggregator, simulator, decision support system (AI), and risk management systems play dual roles in both running simulations and optimizing resource allocation.

By continuously monitoring real-time data, predicting future scenarios, and optimizing resources dynamically, the digital twin ensures that resources are deployed efficiently and reallocated as needed based on evolving disaster conditions. In this way, the simulation infrastructure directly supports real-time resource management.

The tasks considered refer to:

- Predictive simulation for future resource needs
- Optimization algorithms for resource efficiency
- Feedback from resource usage
- Visualization and decision support for resource deployment
- Simulation for long-term resource planning

The Conceptual model is similar with the one presented in the Simulation chapter (5.3.5)

5.3.3.2 *Risk management processes*

The digital twin as a whole system is extremely valuable for risk management because it integrates real-time data, predictive simulations, and decision support systems to continuously assess, prioritize, and mitigate risks. Many components that are used for simulation also play a direct role in risk management, particularly the data aggregator, simulator, decision support system (AI), and risk assessment models. These components ensure that the digital twin can dynamically monitor risks, predict future hazards, and optimize resource allocation and mitigation strategies to reduce risk in real-time.

The following components of the digital twin are used for the risk assessment:

- **Data aggregation**: Continuous real-time data is critical for both risk monitoring and simulations.
- **Simulation**: Predicting future risks and evaluating mitigation strategies relies on simulation models.
- **AI and Decision Support**: Helping prioritize and mitigate risks through data-driven recommendations.
- **Visualization**: The user interface helps decision-makers understand the current and predicted risk landscape, making it easier to take proactive steps.

The Conceptual model is similar with the one presented in the Simulation chapter (5.3.5)

5.3.3.3 *Communication flows*

The **digital twin** as a whole system significantly improves **communication flows** by acting as a **centralized hub** for real-time data, predictive insights, and decision-making information. Components that are part of the simulation system—such as the **data aggregator**, **simulator**, **decision support system (AI)**, and **user interface**—play crucial roles in enhancing **communication efficiency**, ensuring that **relevant information is shared** in a timely and accurate manner.

The Conceptual model is similar with the one presented in the Simulation chapter (5.3.5)

The following components of the digital twin are used for the risk assessment:

- **Data aggregation**: Continuous real-time data collection supports both communication flows and simulations. It ensures that everyone receives the same real-time information, enabling more informed decision-making and better coordination across teams.
- **Simulation**: Predictive modelling enhances communication by informing stakeholders of potential future risks. It allows the digital twin to communicate possible future outcomes to relevant stakeholders, enabling them to prepare ahead of time.
- **AI and Decision Support**: Automated alerts and recommendations help streamline communication between teams. It assures that clear, actionable information is communicated to the right stakeholders at the right time, reducing response time and improving coordination.
- **Visualization** (User Interface): The interface acts as a central platform for sharing insights and collaborating in real-time. It acts as a visual platform where decision-makers can see the current state of the disaster, monitor resources, and track ongoing risks. It helps communicate complex data in a way that is easy to understand. This component also hosts the feedback mechanism where field teams provide real-time updates,

5.3.4 TRAINING

5.3.4.1 *Training system for emergency scenarios*

This conceptual model refers to the training platform for emergency response, integrating simulations of assets, physical environments, and community dynamics. It allows users to practice in a controlled, risk-free environment using real-time data aggregation, decision support, and visualizations, improving their readiness for real emergencies. The conceptual model considered during the training process is presented in the next figure (Figure 35 Conceptual models for training).

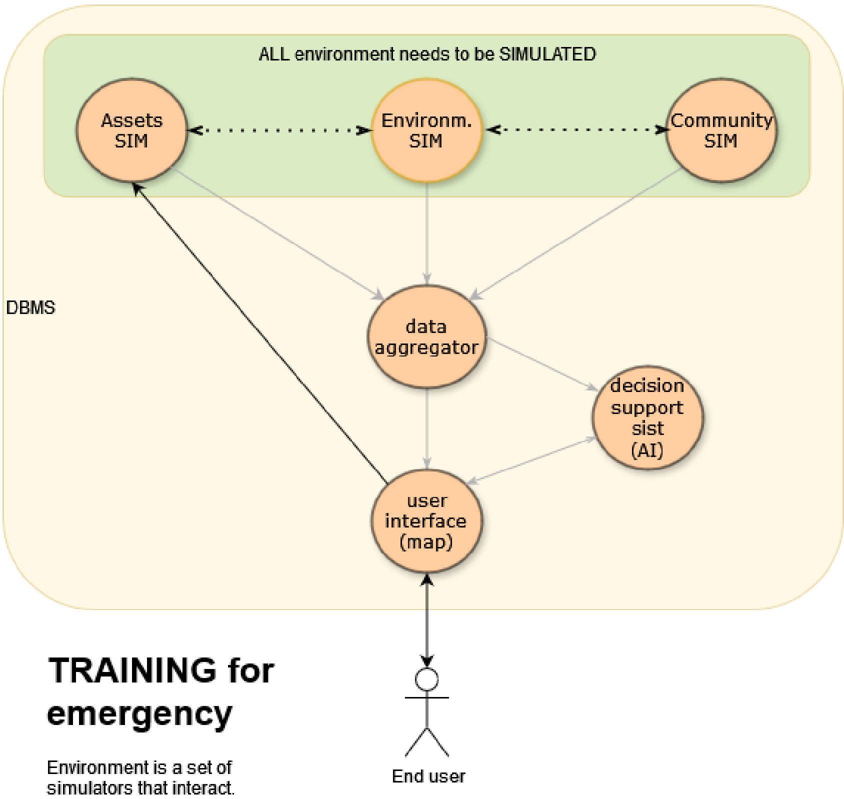


Figure 35 Conceptual models for training

The diagram emphasizes that **everything is simulated** to create a training environment, meaning all components (assets, the physical environment, and community dynamics) are artificially generated to imitate real-world emergencies. This helps users practice decision-making and response strategies without real-world consequences.

It focuses on simulating all aspects of the environment to prepare users (such as emergency responders or decision-makers) for real-life emergency situations.

Here's the breakdown of the components:

- **Simulated environment** (green section at the top): In contrast to the previous diagrams dealing with real or emergency situations, here **all aspects of the environment are simulated** to create a training scenario. This includes:
  - **Assets SIM**: The simulation of risk mitigation assets (such as emergency response resources, personnel, equipment) is part of the training. This allows users to practice deploying and utilizing resources in a simulated setting.
  - **Environ. SIM**: This simulates the physical environment, like natural disasters, infrastructure, weather, etc., creating realistic scenarios for emergency response.
  - **Community SIM**: This component simulates community responses, human behaviour, or social conditions in a disaster setting. It can involve population movement, communication patterns, and community needs during a crisis.
- **Data aggregator**: The data aggregator, as in the previous diagrams, serves as the central processing unit, gathering simulated data from all parts of the environment (simulated assets, physical environment, and community) and preparing it for the next steps.
- **User Interface (Map)**: The results of the simulated data are displayed to the end user through a visual interface, most likely a map. This is where users interact with the training system, making decisions based on the scenario presented to them.
- **Decision Support System (AI)**: Similar to the other diagrams, the AI-based decision support system helps guide decision-making by providing insights and recommendations based on the simulated scenario. It can serve as a feedback mechanism during training, suggesting optimal courses of action or evaluating the user's decisions.
- **End user**: The user in this case is a trainee or a group of trainees undergoing the emergency response training. They interact with the simulated environment, use the interface, and make decisions as if they were in a real emergency.
- **DBMS (Database Management System)**: The DBMS is responsible for managing the simulated data, storing, and retrieving it as necessary during the training session.

5.3.5 SIMULATION

5.3.5.1 *Simulation-based decision support*

The following diagram represents a system that integrates real-time environmental inputs, simulations, AI-driven decision support, and visualization tools to enable more effective disaster response

The flow of the system indicates that data from both the physical and community environments are continuously fed into the simulator and decision support system, and the results are displayed to the user, who can then make real-time decisions to mitigate disaster impacts. It focuses on simulating all aspects of the environment.

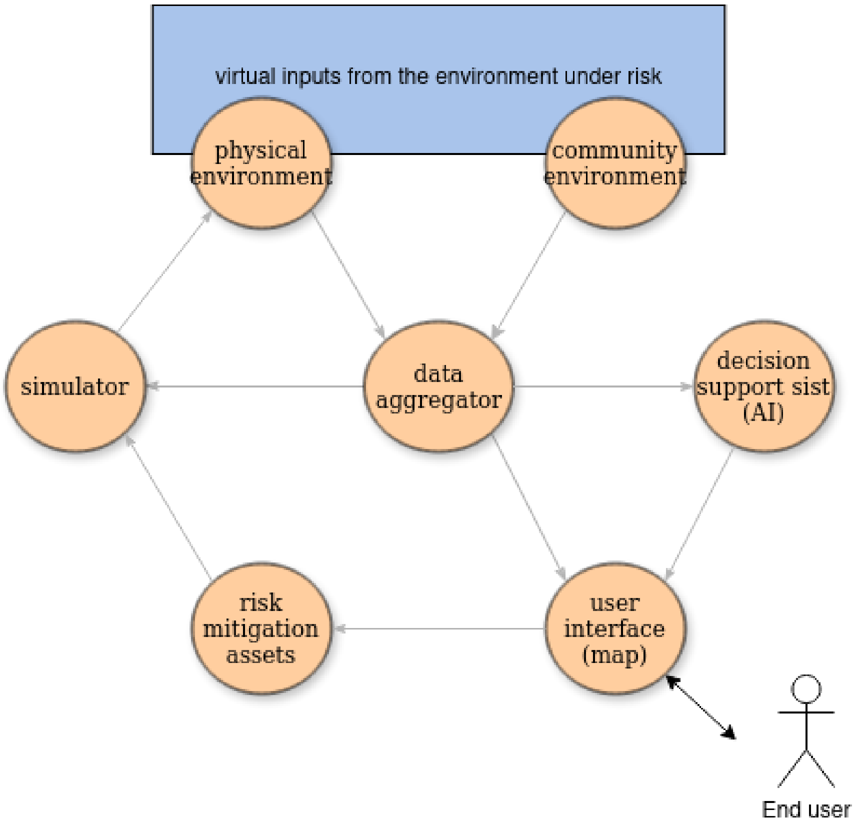


Figure 36 Conceptual models for simulation

Here's the breakdown of the components:

- **Virtual inputs from the environment under risk** (top box): This refers to data inputs coming from the real-world environment that is at risk, such as data from physical sensors, social systems, or environmental changes (e.g., a flood zone or an area experiencing a fire).
- The **physical environment** would include physical, tangible factors (e.g., infrastructure, geographical data, weather).
- The **community environment** represents the social factors, including populations, behavioral data, and community responses.
- **Data aggregator**: The data from both the physical and community environments flows into the data aggregator, which is central to the system. The aggregator processes and combines these diverse inputs into a format usable for decision-making.
- **Simulator**: The simulator uses the aggregated data to simulate possible disaster scenarios, testing various outcomes and responses.
- **Risk mitigation assets**: These are tangible resources (e.g., emergency supplies, personnel, evacuation routes) that can be activated or utilized based on simulations and real-time data.
- **Decision support system (AI)**: This is an artificial intelligence-driven system that helps in making decisions by analyzing data, possibly offering predictions, insights, or recommendations to mitigate risks.
- **User interface (Map)**: The processed data, simulations, and AI insights are presented through a user-friendly interface, typically a map that provides visual context to the risk scenario, showing affected areas and resources.

- **End user**: The system is ultimately designed for an end user, likely a decision-maker (e.g., emergency management personnel) who interacts with the user interface to make informed decisions during a disaster event.
- **DBMS (Database Management System)**: The DBMS is responsible for managing the simulated data, storing, and retrieving it as necessary during the training session.

5.3.6 OPERATION

5.3.6.1 *Operational structure*

The **digital twin** should be placed centrally in the operational structure, representing a real-time virtual model that is continuously updated to reflect the current state of the environment, assets, and ongoing operations. It supports the entire operational structure by serving as the foundation for real-time decision-making**,** coordination, and resource management across all units involved in disaster management. The **digital twin** will enable each operational unit (e.g., resource management, command centres, field operations) to have a unified, up-to-date view of the ongoing situation, ensuring they are all aligned in their actions.

The role of the operational structure is to give real-time response in actual emergencies. This is explained in the next subchapter.

5.3.6.2 *Real-time response in actual emergencies.*

The next conceptual model refers to the real-time response.

This diagram represents a system that operates during a **real emergency** situation, contrasting with the earlier diagrams focused on disaster modelling and simulation. Here, the system is actively responding to an ongoing emergency event, using real-time data from the environment.

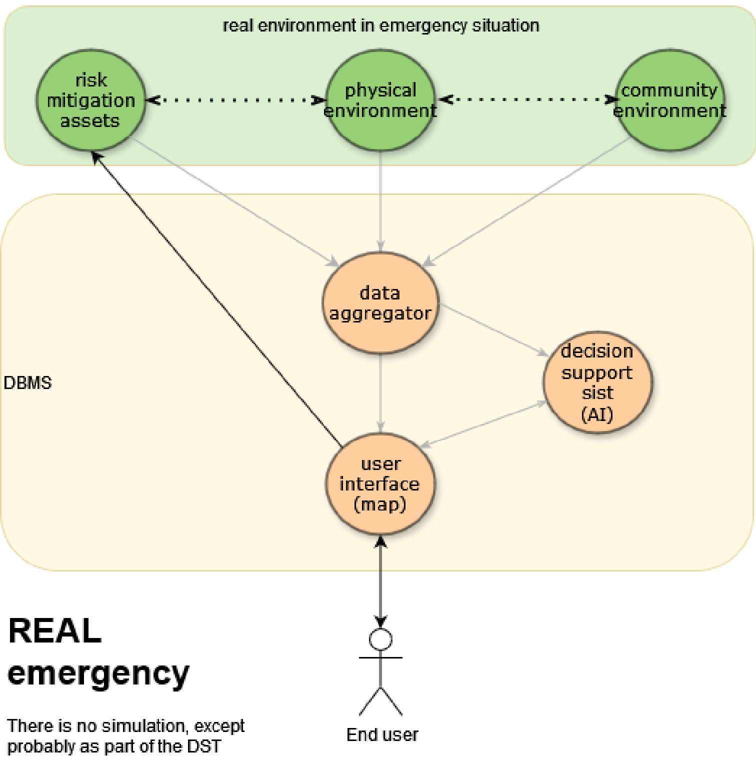


Figure 37 Conceptual models for real-time response

The key components presented in the diagram are:

- **Real environment in emergency situation** (top green area): This part of the diagram refers to the real-world conditions during an emergency event. It includes:
    - **Physical environment**: Real-time data about physical conditions, such as weather, infrastructure damage, and environmental hazards.
    - **Community environment**: Data concerning social factors, such as population movement, communication, and local community conditions.
- **Risk mitigation assets**: Resources that are actively deployed to mitigate the risk, such as emergency services, evacuation plans, or physical resources like food, medical supplies, and shelters.
  These components are interacting with each other and providing inputs to the system through the **data aggregator**.
- **Data aggregator**: The data aggregator gathers information from the physical and community environments, as well as risk mitigation assets. It processes the data and makes it available for further analysis and decision-making.
- **User Interface (Map)**: The processed data from the aggregator is presented visually to the end user through a map or other visualization tools. This is the critical interface where real-time information is displayed, enabling decision-makers to respond effectively.
- **Decision Support System (AI)**: The AI-powered decision support system provides insights and recommendations based on the real-time data. It helps users analyse the situation and choose the most effective actions during the emergency.
- **End user**: The end user interacts with the system through the user interface. This person (or group) is responsible for making real-time decisions based on the information provided by the system.
- **DBMS (Database Management System)**: The entire system relies on a database management system (DBMS), which is essential for storing and retrieving real-time data during the emergency. The DBMS likely serves as the backbone for the data aggregator and user interface.

5.3.7 POST EVENT

The **digital twin** serves an important role in **post-event actions**, offering a comprehensive platform for damage assessment, resource recovery, after-action analysis, learning, and future risk planning. Several components that were initially developed for simulation can be adapted and reused to support these post-event processes. These components include the data aggregator, simulator, decision support system (AI), risk assessment models, and the user interface.

- **Data aggregation**: Real-time data collection continues during recovery to assess damage and resource allocation.
- **Simulation**: Simulations can be used to replay the event and test different recovery strategies.
- **AI and Decision Support**: Post-event recommendations help optimize recovery planning and resource management.
- **Risk Models**: Updating risk models based on the event's outcomes ensures that future preparedness is improved.
- **Visualization (User Interface)**: Visualizing damage, recovery progress, and after-action reports helps decision-makers, and the public stay informed.
- By integrating these components, the digital twin enables **efficient recovery efforts**, continuous **learning from past events**, and better **preparation for future disasters**.

The Conceptual model is similar with the one presented in the Simulation chapter (5.3.5)

5.4 WILDFIRE SPECIFIC MODELS

Wildfires are part of the natural cycle of life in vegetated regions. The apparent wildfire increase in size and frequency of recent years reflects land management legacy, expansion of human activity, and changes in climatic conditions.

The conceptual models previously presented are applicable to wildfire use cases and should cover all the steps mentioned before. There are some particularities for these use cases, which will be hereby presented.

The **digital twin** concept in wildfire management provides a highly dynamic and accurate representation of the wildfire's real-time behaviour**,** incorporating real-world data, running simulations, and delivering actionable insights to decision-makers. This system significantly enhances the responsiveness and effectiveness of wildfire mitigation efforts, making it a powerful tool for disaster management.

Some of the benefits of using digital twin with the specific conceptual models, are:

- Real-time, adaptive response
- Better predictive modeling
- Improved decision making
- Enhanced public safety

5.4.1 STEP-BY-STEP USAGE IN A WILDFIRE SCENARIO

The steps used by a digital twin for a wildfire scenario are:

1) Real-Time Data Integration into the Digital Twin:

   The **digital twin** continuously collects real-time data from various sources such as:

   - **Satellite imagery**: Provides a real-time view of the wildfire's location, size, and spread.
   - **Ground sensors**: Measure factors like wind speed**,** air humidity**,** temperature, and fuel conditions (vegetation moisture levels).
   - **Drones and firefighting units**: Collect aerial imagery and thermal data on the ground to provide high-resolution, localized views of the fire.
   - **Community reports**: Provides crowdsourced data from local communities reporting fire sightings, evacuation needs, or smoke levels.

   The digital twin uses this data to create a real-time, continuously updated model of the wildfire, including its current boundaries, behaviour, and projected path.

2) Simulator running what-if scenarios**:**

   The **simulator** leverages the digital twin's up-to-date environment and runs multiple **what-if scenarios** based on possible changes in weather or firefighting actions:

   - **Fire spread simulation**: Models how the fire might spread under different wind speeds or directions, simulating the movement of the fire front based on terrain, vegetation, and weather.
   - **Fire intensity simulation**: Predicts the intensity of the fire in various areas, highlighting hotspots where temperatures are highest, and vegetation is most flammable.

- **Resource allocation simulation**: Tests different firefighting strategies, such as the deployment of firebreaks**,** water bombers**,** or ground crews, to find the most effective methods for controlling the fire.

3) Feedback Loop to the Digital Twin**:**

After running simulations, the outcomes are fed back into the digital twin, updating the virtual model based on the most likely scenario:

- If the simulator predicts that the fire will change direction due to high winds, the digital twin adjusts the fire's projected path and updates the environmental conditions in the model to reflect this.
- If the simulation suggests that a firebreak or controlled burn successfully halted the fire's advance in one area, the digital twin reflects these changes.

4) Decision Support System (AI) Recommendations**:**

The **decision support system (AI)** processes the updated model from the digital twin and provides recommendations to emergency managers and firefighters**:**

- Evacuation orders
- Resource deployment
- Community alerts**.**

5) Visualization for End Users:

The **end users** (firefighters, emergency responders, and local authorities) interact with a visual interface (e.g., a map or dashboard) that represents the digital twin of the wildfire:

- Real-time map of the fire**.**
- Evacuation routes and safe zones**.**
- Resource tracking**.**

6) Continuous Adaptation**:**

As conditions change (e.g., a sudden shift in wind direction or a break in the fire line), the **digital twin** updates in real time and triggers new simulations:

End users receive new visualizations and action plans based on the updated model.

5.5 EARTHQUAKE MODELS

Earthquake Models refers to models used to simulate or describe earthquake phenomena, typically focusing on the prediction of earthquake occurrences, ground motion, and their effects on structures and the environment.

In the context of **earthquake disaster management**, using a **digital twin** can provide real-time insights and dynamic simulations of how an earthquake affects the built environment, infrastructure, and populations. By integrating real-time data from sensors, geological monitoring systems, and structural health data, the digital twin offers a continuously updated model that supports decision-making before, during, and after an earthquake.

The conceptual models presented before are applicable to the earthquake use case and should cover all the steps mentioned before. Some particularities of these use cases are hereby presented.

Some of the benefits of using digital twin with the specific conceptual models, are:

- Real-time situational awareness
- Accurate predictive modeling
- Enhanced decision support
- Efficient response coordination
- Improved public safety

5.5.1 STEP-BY-STEP USAGE IN A EARTHQUAKE SCENARIO

The steps used by a digital twin for an Earthquake scenario are:

1) Real-time data integration into the digital twin**:**

The **digital twin** is continuously updated with real-time data from a range of sources:

- **Seismic sensors**: Provide data on the magnitude, depth, and location of the earthquake. These sensors are deployed across regions prone to seismic activity and are part of global earthquake monitoring networks.
- **Structural health monitoring systems**: Sensors embedded in buildings, bridges, and other critical infrastructure send real-time data on building stress**,** vibration levels, and potential structural damage.
- **Transportation networks**: Sensors on roads, railways, and transit systems monitor infrastructure integrity and track disruptions in services (e.g., collapsed bridges, damaged highways).
- **Utility systems**: Sensors and monitors in gas lines**,** water systems**,** and power grids detect leaks, failures, and disruptions caused by the earthquake.

2) Simulator running what-if scenarios**:**

The **simulator** works with the digital twin to run multiple what-if scenarios based on the earthquake's parameters and potential aftershocks:

- **Building collapse simulation**: The simulator models the likelihood of building collapse based on real-time data from structural health sensors, analysing which buildings are at the highest risk.
- **Infrastructure stress simulation**: It simulates how bridges, roads, and tunnels might fail under the earthquake's stress, providing predictions about which transportation routes will remain functional, and which will be blocked.
- **Aftershock simulations**: The system simulates the impact of potential aftershocks, which could further damage already weakened buildings and infrastructure. It predicts where aftershocks might cause secondary damage, allowing responders to prepare.

3) Feedback loop to the digital twin**:**

The outcomes of the simulations feed back into the **digital twin**, keeping the virtual model updated and dynamically reflecting potential outcomes:

- If the simulator predicts the collapse of specific buildings, the digital twin immediately reflects these structural changes, updating the virtual environment in real time.

- If a bridge or road collapses, the transportation system is updated in the digital twin, allowing emergency services to reroute traffic and responders.

4) Decision Support System (AI) Recommendations**:**

The **decision support system (AI)** processes data from the digital twin and the simulator to offer real-time **recommendations** to emergency managers and response teams:

- Evacuation orders**.**
- Resource allocation**.**
- Public safety alerts**.**

5) Visualization for End Users**:**

Emergency responders, city planners, and government agencies interact with a **visual interface** that represents the **digital twin** and shows real-time earthquake impact data:

- Real-time map of damage**.**
- Evacuation routes and safe zones**.**
- Resource and personnel tracking**.**

6) Continuous Adaptation**:**

As the situation evolves (e.g., an aftershock occurs or new damage is reported), the **digital twin** adapts:

- If an **aftershock** strikes, the simulator runs a new scenario to predict additional damage and collapse, updating the digital twin accordingly.
- The **digital twin** immediately reflects any new changes in the environment, whether they come from structural collapses, infrastructure failures, or new sensor data.
- Emergency response teams and decision-makers receive updated recommendations from the **decision support system** based on the new data.

5.6 HEATWAVE MODELS

Heatwave models are tools or systems used to simulate, predict, and understand the occurrence, intensity, duration, and impacts of heatwaves. These models are important for weather forecasting, public health, agriculture, infrastructure management, and urban planning, as they help predict extreme heat events and assess their potential consequences. Heatwave models typically combine climate data, weather patterns, and geographical information to provide short-term forecasts or long-term projections.

Heatwaves are different from other disasters like earthquakes in that their impacts are often gradual and widespread, affecting vulnerable populations more severely over time rather than through a single event. However, many of the principles of disaster preparedness, response, and recovery still apply.

5.6.1 STEP-BY-STEP USAGE IN A HEATWAVE SCENARIO

The step-by-step usage considers two phases:

- Pre event
- During the event

5.6.1.1 *Pre event steps*

1) Real-time environmental data collection

The **data aggregator** collects real-time information from weather sensors**,** climate data sources**,** and urban infrastructure (e.g., temperature, humidity, energy consumption).

The digital twin monitors vulnerable regions (urban heat islands, elderly populations, areas with limited access to air conditioning) and assesses the likelihood of extreme temperatures based on weather forecasts.

2) Simulation of Heatwave Scenarios

The digital twin runs simulations to predict the intensity and duration of the heatwave. It uses historical climate data and weather forecasts to model how the heatwave will affect different parts of the region.

Risk models are updated to estimate the likelihood of power outages due to high energy demand (air conditioning usage) and infrastructure strain (e.g., water shortages, health system overload).

3) Early warning systems

The **decision support system (AI)** analyses the simulation outcomes and provides early warnings to the public and authorities, indicating when the heatwave is expected to peak.

5.6.1.2 *During the event steps*

1) Real-time monitoring of temperature and power demand

2) Vulnerability and risk assessment

- The digital twin identifies high-risk populations and locations.
- Predictive models simulate how the heatwave will evolve.
- Health system capacity is assessed by tracking hospital admissions and ambulance availability, with predictive simulations showing the likelihood of overcrowding or resource shortages.

3) Resource allocation and risk mitigation

- **Cooling centres** are pre-activated based on risk models that predict the most vulnerable regions.
- Resources such as water supply**,** portable cooling units, and medical teams are dynamically allocated to areas with the greatest need**.**
- The system tracks hospital capacity and healthcare resource availability in real-time. It advises health authorities to redistribute medical personnel and supplies based on the heatwave's projected peak.

4) Public communication and alerts

The digital twin continuously communicates updates to the public, advising people to avoid outdoor activities during peak heat hours, stay hydrated, and find cooling centres.

5.7 MAN-MADE DISASTER MODELS

Man-made disaster models are tools or frameworks designed to predict, simulate, assess, and manage the impact of disasters caused by human activities. These disasters can result from industrial accidents, environmental pollution, terrorism, wars, infrastructure failures, and technological accidents. Man-made disaster models help in risk assessment, mitigation planning, and response strategies by simulating various scenarios and evaluating their potential impacts on society, the environment, and infrastructure.

The conceptual models presented before are applicable to earthquake use cases and should cover all the steps mentioned before. Particularities for these use cases are presented hereby.

Some of the benefits of using digital twin with the specific conceptual models, are:

- real-time, dynamic response
- accurate predictive modeling
- improved decision support
- effective risk mitigation
- public safety and communication.

5.7.1 STEP-BY-STEP USAGE IN A MAN-MADE SCENARIO

The steps used by a digital twin for a man-made scenario are:

1) Real-time data integration into the digital twin**:**

The **digital twin** integrates real-time data from various sources, creating a detailed, real-time virtual representation of the industrial facility, the chemical spill, and the surrounding environment:

- **Industrial sensors**: Sensors within the facility continuously monitor critical processes, such as temperature, pressure, and chemical storage conditions. During the incident, these sensors detect failures (e.g., a leaking tank or pipeline rupture).
- **Environmental sensors**: Monitors surrounding the facility collect real-time data on air quality**,** water contamination levels**,** and wind direction, helping to track the spread of hazardous substances.
- **Drones and robots**: Autonomous systems deployed to the site capture visual data and thermal **imaging** from dangerous areas where humans cannot safely enter.
- **Emergency responders**: Data from on-the-ground teams feeds into the system, updating the digital twin with situational awareness (e.g., status of fire containment efforts or access restrictions due to high toxicity).

2) Simulator running what-if scenarios**:**

The **simulator** interacts with the digital twin to model various **what-if scenarios** related to the chemical spill:

- **Toxic cloud dispersion simulation**: The simulator models how the leaked chemicals disperse into the air, factoring in wind speed, temperature, and terrain. This model helps predict which areas will be affected by the toxic plume.
- **Water contamination spread**: If chemicals enter nearby water sources (e.g., rivers, groundwater), the simulator predicts how contaminants will spread, and how long it will take to reach downstream communities or critical water supply points.
- **Infrastructure failure simulation**: The simulator runs failure tests on critical infrastructure (e.g., storage tanks, pipelines) to determine if additional leaks or explosions could occur due to structural degradation.

- **Explosion and fire spread scenarios**: For industrial accidents involving explosions, the simulator models how the fire might spread through the facility, potentially triggering secondary explosions or fires in adjacent buildings.

3) Feedback loop to the digital twin**:**

The outcomes of the simulations are fed back into the **digital twin**, which dynamically updates to reflect the likely evolution of the disaster**.**

This feedback loop ensures that the digital twin continuously adapts to the evolving disaster, providing real-time insights into how the situation is likely to progress.

4) Decision support system (AI) recommendations**:**

The **decision support system (AI)** processes data from the digital twin and simulator to offer real-time recommendations**:**

- Evacuation Planning**.**
- Containment and Mitigation**.**
- Emergency Resource Allocation**.**
- Public Health**.**

5) Visualization for end users**:**

Emergency responders, industrial operators, and local authorities interact with the digital twin through a **visual interface**, typically a real-time map or dashboard:

- Real-time hazard map**.**
- Infrastructure status**.**
- Resource tracking**.**
- Population movement**.**

6) Continuous adaptation**:**

The **digital twin** continuously updates as new data flows in from the sensors, drones, and responders:

As the situation changes, the AI system provides **updated recommendations** for containment, evacuation, and resource allocation.

5.7.2 INPUT DATA (FORMAT)

- Historical data in CSV format

The sources of data include all kind of data available as described in chapter 3.2 inside this document.

6. CONCLUSIONS

In conclusion, the deliverable D4.2 presents the methodologies, data representations, and conceptual models that are essential for developing the Smart City Digital Twin framework. Its objectives were to investigate methods for representing data collected from task T4.1, preparing it for integration with simulation models and machine learning algorithms and the presenting conceptual models used.

The application of the AGILE and IEEE-1730 DSEEP methodologies provided a robust framework for the effective management of the project's inherent complexity, ensuring adaptability and iterative development while maintaining rigorous simulation standards. AGILE facilitated efficient project management by providing a mechanism for continuous feedback, while the IEEE DSEEP framework structured the development and execution of distributed simulation systems.

In the domain of data representation, the deliverable provides an effective statistical treatment of empirical data, ensuring its suitability for both machine learning models and simulations. The architectural reference for data representation provides a clear framework for the structuring and management of data within the PANTHEON system, ensuring its efficient and accurate use in simulations.

The development of conceptual models, including those for wildfires, earthquakes, heatwaves, and man-made disasters, evinces the project's extensive capabilities. These models serve as critical tools for understanding and managing both natural and man-made disaster scenarios. The models are used by the SCDT which is described as the whole system offered. Specific for this project is the treatment of the cascading effects for the simulation process. Furthermore, the inclusion of general-purpose models offers flexibility, thereby ensuring the framework's applicability to a variety of situations beyond the specific disaster scenarios.

Overall, this deliverable provides a comprehensive framework for integrating data representation and conceptual modelling into the Smart City Digital Twin. It establishes the foundations for future work, ensuring that the developed models can be continuously refined, tested, and applied to real-world scenarios with accuracy and efficiency. The contributions from all partners have significantly enhanced the quality and depth of the findings, setting a strong precedent for subsequent tasks and deliverables.

7. REFERENCES

Anand Deshpande, M. K. (2018). *Artificial Inteligence for Big Data.* Packt.

Apache. (2024). *ApacheMQ*. Retrieved from https://activemq.apache.org/

Apache Airflow. (2024). Retrieved from Apache Airflow: https://www.geeksforgeeks.org/what-is-apache-airflow/

Apache Kafka. (2024). *Apache Kafka*. Retrieved from https://kafka.apache.org/

Apache Kafka vs RabitMQ. (2024). Retrieved from https://www.upsolver.com/blog/kafka-versus-rabbitmq-architecture-performance-use-case

Apache NIFI. (2024). Retrieved from Apache NIFI: https://nifi.apache.org/

Docker documentation. (2024). Retrieved from Docker documentation: https://docs.docker.com/get-started/overview/

Docker Swarm. (2024). Retrieved from Docker Swarm: https://www.geeksforgeeks.org/introduction-to-docker-swarm-mode/

IEEE. (2022). IEEE Recommended Practice for Distributed Simulation Engineering and Execution Process (DSEEP. Retrieved from https://standards.ieee.org/ieee/1730/10715/

IEEE-1730. (2022). *https://ieeexplore.ieee.org/document/9919118.* IEEE.

ISO/IEC/IEEE 24748-6:2023(en). (2023). Retrieved from ISO/IEC/IEEE 24748-6:2023(en): https://www.iso.org/obp/ui/en/#iso:std:iso-iec-ieee:24748:-6:ed-1:v1:en

Keycloack. (2024). *Keycloack*. Retrieved from https://www.keycloak.org/documentation

Kubernetes. (2024). Retrieved from Kubernetes: https://kubernetes.io/

Kubernetes vs Swarm. (2024). Retrieved from Kubernetes vs Swarm: https://phoenixnap.com/blog/kubernetes-vs-docker-swarm

Liu, H. H. (2018). *Machine Learning - A Quantitative Approach.* PerfMath.

Maxwell B. Joseph, M. W. (2019). Spatiotemporal prediction of wildfire size extremes with Bayesian finite sample maxima. *Ecological Aplications, september 2019*, 1266-1281.

Mulesoft Anypoint. (2024). Retrieved from Mulesoft Anypoint: https://www.mulesoft.com/

Ogata, Y. (1988). Statistical Models for Earthquake Occurrences and Residual Analysis for Point Processes. *Journal of the American Statistical Association, 83(401)*, 9-27.

PANTHEON - Consortia. (2023). *DOA.*

Pantheon Consortia. (2024). *D3.6-Use Case Scenarios.*

Pantheon Consortia. (2024). *D3.7 Pantheon System Architecture.*

PMI. (2024). *Agile Practice Guide.* Retrieved from https://www.agilealliance.org/wp-content/uploads/2021/02/AgilePracticeGuide.pdf

RabitMQ. (2024). *RabitMQ*. Retrieved from https://www.rabbitmq.com/

8. ANNEX I-MATHEMATICAL MODELS REFERED

8.1 STATISTICAL CONCEPTS USED IN MODELS

8.1.1 OCCURRENCE

Poisson distribution

Wildfire occurrences are often modelled as rare events happening randomly in time or space. The Poisson distribution is commonly used to describe the number of events in a fixed interval (time, area, etc.) when the events happen independently of each other.

The probability of observing k events in each interval (where k is a non-negative integer, 0, 1, 2, 3, ...) is given by the formula:

$$P(k;\lambda) = \frac{\lambda^k e^{-\lambda}}{k!}$$

where:

- $P(k;\lambda)$ is the probability of kkk events occurring in the fixed interval.
- λ is the **average rate of occurrence** (the expected number of events in the interval).
- e is Euler's number, approximately 2.71828.
- $k!$ is the factorial of k (i.e., $k \times (k-1) \times (k-2) \times \cdots \times 1$).

8.1.2 TIME BETWEEN EVENTS

The **exponential distribution** is a continuous probability distribution that is often used to model the time between independent events that happen at a constant rate. It is particularly useful in scenarios where events occur randomly and independently over time as is the time until a wildfire occurs in a region.

Exponential Distribution Formula

It is used to get:

The time between consecutive wildfires

The time until a fire reaches a certain size or grows out of control

The time between periods of weather conditions conducive to wildfires (e.g., dry, hot, and windy conditions)

The time until a wildfire is successfully suppressed or extinguished

The time between **lightning strikes** (a common natural ignition source for wildfires)

Formula: $f(t) = \lambda e^{-\lambda t}$, where:

t is the time between events (wildfires),

λ is the rate parameter (e.g., wildfires per month),

$e^{-\lambda t}$ gives the probability that the next wildfire will occur after t time units.

The probability density function (PDF) of the exponential distribution is given by:

$$f(x,\lambda) = \lambda e^{-\lambda x}, x \geq 0$$

where:

- $f(x,\lambda)$ is the probability density function (the likelihood of observing a value close to xxx).
- λ is the **rate parameter**, representing the average rate at which events occur.
    - **Note**: λ is also the inverse of the mean ($\lambda = 1/\mu$), where μ is the average time between events.
- λ is Euler's number, approximately 2.71828.
- $x \geq 0$ since time or distance between events cannot be negative.

The cumulative distribution function (CDF), which gives the probability that a random variable X is less than or equal to x, is:

$$F(x;\lambda) = 1 - e^{-\lambda x}, x \geq 0$$

This function provides the probability that the time until the next event is less than or equal to x.

8.1.3 DURATION OF EVENTS

The **Weibull distribution** is used to model the **duration of events**, especially when the rate of occurrence is not constant over time. Unlike the exponential distribution, which assumes a constant hazard rate (i.e., a fixed probability of an event occurring at any given time), the **Weibull distribution** allows the hazard rate to change over time, making it more flexible for modelling wildfire durations.

The probability density function (PDF) of the Weibull distribution is given by:

$$f(x;k,\lambda) = \frac{k}{\lambda}\left(\frac{x}{\lambda}\right)^{k-1} e^{-\left(\frac{x}{\lambda}\right)^{k}}$$

Where:

x is the random variable (e.g., time to failure).

k (shape parameter) controls the shape of the distribution and the hazard rate.

λ (scale parameter) controls the scale of the distribution (i.e., it stretches or shrinks the distribution).

e is Euler's number (approximately 2.71828).

� Shape parameter $\boldsymbol{k}$: This controls the hazard rate over time:

$k > 1$: Increasing hazard rate (wildfires become more likely to extinguish as time goes on).

$k = 1$: Constant hazard rate (similar to the exponential distribution).

$k < 1$: Decreasing hazard rate (wildfires become less likely to extinguish as time goes on).

Scale parameter λ:

This controls the scale of the distribution and stretches or compresses the duration values. It can be adjusted to reflect how long wildfires typically last on average. Higher values lead to longer average wildfire durations.

Weibull Distribution:

The weibull_min.rvs function generates random wildfire durations using the Weibull distribution, which allows us to model scenarios where the hazard rate changes over time.

Time until the next wildfire

To model the time until the next wildfire using the Weibull distribution, we can assume that the likelihood of a wildfire occurring changes over time rather than being constant. The Weibull distribution provides flexibility in modelling events with changing hazard rates, making it suitable for scenarios where the probability of a wildfire occurring might increase or decrease over time based on environmental conditions, such as increasing dryness or accumulating fuel loads

Shape parameter k:

$k > 1$: Increasing hazard rate. This means the longer the time passes, the more likely a wildfire is to occur (e.g., due to fuel buildup, increased dryness, or worsening environmental conditions).

$k = 1$: Constant hazard rate. This case is equivalent to the exponential distribution.

$k < 1$: Decreasing hazard rate. This means that as time passes, the likelihood of the next wildfire occurring decreases (perhaps due to effective fire prevention measures).

Scale parameter λ: The scale of the distribution, affecting the typical time to the next wildfire. Larger values of λ result in longer times between wildfires on average.

Plot Interpretation:

The x-axis represents the time to the next wildfire (e.g., in months, years, or any other appropriate time unit).

The y-axis represents the density of those times.

The histogram shows simulated data for the time to the next wildfire, while the red line represents the theoretical probability density function (PDF) for the Weibull distribution.

8.1.4 FIRE SIZE

The **log-normal distribution** is commonly used to model variables that are positively skewed and cannot be negative, such as wildfire sizes. In this context, wildfire size often follows a log-normal distribution because most fires are small, but a few can be extremely large, leading to a long tail in the distribution.

The probability density function (PDF) of a log-normal distribution is given by:

$$f(x;\mu,\sigma) = \frac{1}{x\sigma\sqrt{2\pi}} e^{-\frac{(ln(x)-\mu)^2}{2\sigma^2}}, x > 0$$

where:

x is the random variable (must be positive, $x > 0$).

μ is the mean of the natural logarithm of the random variable $ln(X)$, i.e., the mean of the corresponding normal distribution. Adjusting the mean will shift the centre of the distribution, which affects the typical size of the wildfires.

σ is the standard deviation of the natural logarithm of the random variable $ln(X)$, i.e., the standard deviation of the corresponding normal distribution. Increasing the σ parameter will make the distribution more spread out, meaning that there's a wider range of fire sizes, including very large fires.

$ln(x)$, is the natural logarithm of x.

e is Euler's number, approximately 2.71828.

$\sqrt{2\pi}$ is a normalizing constant.

8.1.5 CUMULATIVE TIME OF EVENTS

Suppose we are modelling the time until the next wildfire occurs, and the hazard rate changes over time in a non-linear fashion (e.g., the probability of a fire increases initially, then levels off, and finally decreases as conditions change). The Generalized Gamma distribution allows us to capture this complexity better than simpler distributions like the exponential or Weibull.

In this case:

We assume that the occurrence of a wildfire depends on multiple independent factors (e.g., weather, fuel buildup, humidity, wind speed, human activities, etc.).

The shape parameter *k* of the Gamma distribution represents the number of these independent factors contributing to the wildfire's ignition.

The scale parameter θ adjusts the average time between wildfires based on these factors.

Explanation of the Parameters:

Shape Parameter *k*: Represents the number of factors that need to happen before a wildfire is triggered. For example, in this case, we assume there are three key factors (fuel buildup, weather, and human activity), so *k=3*.

If *k=1*, this would be equivalent to the exponential distribution, meaning that only one key factor is influencing the event, and the hazard rate is constant.

If *k* is greater than 1, this means that multiple stages or events need to occur before the wildfire.

Scale Parameter θ: This controls the spread or scale of the waiting times. Larger values of θ result in longer average times to the next wildfire.

θ adjusts how long it takes on average, for these stages (represented by *k*) to complete.

Size: This controls how many wildfires we are simulating in this example (1,000 simulated wildfires).

Plot Interpretation:

The x-axis represents the time to the next wildfire (in days, months, or any other relevant unit of time).

The y-axis represents the density (i.e., the probability of different waiting times).

The histogram shows the simulated data, while the red line is the theoretical Gamma PDF, showing how the Gamma distribution models the waiting time for the next wildfire.

8.1.6 EXTREME OR RARE EVENTS

The Pareto distribution is commonly used to model extreme events or situations where a small number of events account for most of the total impact, which is often referred to as the "80/20 rule" (i.e., 80% of the effects come from 20% of the causes). In the context of wildfires, the Pareto distribution can be used to model extreme wildfires where a few large wildfires contribute to most of the burned area.

The Pareto distribution is defined by the following formula for the probability density function (PDF):

$$f(x) = \frac{\alpha x_m^{\alpha}}{x^{\alpha+1}} \text{ for } x \geq x_m$$

Where:

α is the shape parameter (also called the "tail index" or "Pareto index"). It controls the heaviness of the tail. Smaller values of α make the tail heavier, meaning more extreme values are likely.

x_m is the scale parameter, which represents the minimum value that x can take. This ensures that the values are larger than x_m, i.e., $x \geq x_m$.

x is the random variable representing the quantity being modeled (e.g., wildfire size).

Shape Parameter α:

Controls the "tail heaviness" of the distribution. The smaller the value of α, the heavier the tail of the distribution, meaning there is a higher likelihood of extremely large values (i.e., very large wildfires).

Larger α values result in fewer extreme values (fewer extreme wildfires).

Scale Parameter: This represents the minimum size of the wildfire, meaning the distribution starts at this value and only produces values larger than the scale. The wildfire sizes will be at least scale.

The Gumbel distribution is often used to model the distribution of the maximum (or minimum) of extreme values, such as the maximum size of a wildfire or the time until the next extreme wildfire. It is particularly useful when dealing with extreme events and is commonly applied in fields like hydrology, meteorology, and risk analysis (e.g., flood levels, extreme temperatures).

There are two types of Gumbel distributions:

- Gumbel for maxima: Used for modelling the largest values (e.g., maximum wildfire size, extreme temperatures).
- Gumbel for minima: Used for modelling the smallest values.

In this approach, we'll use the Gumbel distribution for maxima to model the size of extreme wildfires or time to the next extreme wildfire. The Gumbel distribution is flexible and allows for modelling extreme events in scenarios where the largest (or smallest) values are of interest.

Formula for the Gumbel Distribution (Maxima):

The probability density function (PDF) for the Gumbel distribution (maxima) is:

$$f(x) = \frac{1}{\beta}\left(-\left(\frac{x-\mu}{\beta} + exp\left(-\frac{x-\mu}{\beta}\right)\right)\right)$$

Where:

μ is the location parameter, which shifts the distribution along the x-axis. You can adjust this to reflect the typical size of wildfires. Increasing μ shifts the distribution to the right, meaning larger fires are more likely.

β is the scale parameter, which stretches or compresses the distribution. You can adjust this to reflect the spread of the wildfire sizes. A larger β means more variability in the wildfire sizes, while a smaller β means the sizes are more consistent.

x is the random variable representing the size of the wildfire or the time to the next event. Increasing this number allows for more samples.

8.1.7 SPATIAL DISTRIBUTION AND DIRECTIONS OF SPREAD

Geometric distribution

This is useful in **fire detection systems**, where sensors are deployed in a grid, and you want to estimate how many sensors (or cells) will need to be checked before detecting a wildfire. It is possible to model how many cells (discrete spatial units) you need to check before finding a wildfire ignition, assuming each cell has an independent and constant probability of containing a wildfire.

Geometric distribution formula:

The probability mass function (PMF) of the Geometric distribution is:

$$P(X = k) = (1 - p)^{k-1} \cdot p$$

Where:

p is the probability of success (e.g., the probability that a cell contains a wildfire ignition). You can adjust this to reflect how likely it is that a wildfire is present in each cell. A smaller p means that wildfires are rarer, and more cells need to be checked before one is found.

k is the number of trials (in this case, how many cells were checked until a wildfire was found). You can increase or decrease the number of cells to simulate different regions or areas.

8.1.8 MONTE CARLO SIMULATIONS

8.1.8.1 *Key concepts*

1. **Random sampling**
2. **Probability distributions**
3. **Law of large numbers**
4. **Central limit theorem**

1) Random Sampling

Monte Carlo methods rely on generating random samples from a probability distribution. These samples are then used to approximate the desired quantity.

- **Uniform random sampling:** The simplest form of random sampling, where each sample has an equal probability of being selected.
- **Non-uniform random sampling:** Samples are drawn from non-uniform distributions like normal, exponential, or any custom-defined distribution.

2) *Probability distributions*

Monte Carlo simulations often require sampling from specific probability distributions. The choice of distribution depends on the problem being addressed.

- **Discrete distributions:** For problems where outcomes are finite and distinct (e.g., binomial distribution).
- **Continuous distributions:** For problems with a continuous range of outcomes (e.g., normal, exponential, uniform distributions).

3) *Law of Large Numbers (LLN)*

The Law of Large Numbers is a fundamental theorem in probability theory that justifies the use of Monte Carlo methods. It states that the average of a large number of independent and identically distributed (i.i.d.) random variables converges to the expected value of the distribution.

- **Formally:** Let $X_1, X_2, \dots, X_n$ be i.i.d. random variables with mean μ. As n approaches infinity, the sample average converges to μ:

$$\overline{X}_n = \frac{1}{n}\sum_{i-1}^{n} X_i \rightarrow \mu \text{ as } n \rightarrow \infty$$

4) *Central Limit Theorem (CLT)*

The Central Limit Theorem provides a further justification by describing the distribution of the sample mean. It states that the distribution of the sum (or average) of a large number of i.i.d. random variables, each with finite mean and variance, approaches a normal distribution.

- **Formally:** Let $X_1, X_2, \dots, X_n$ be i.i.d. random variables with mean μ and variance σ^2. Then the standardized sum approaches a standard normal distribution:

$$\frac{\sqrt{n}(\overline{X}_n - \mu)}{\sigma} \rightarrow N(0,1) \text{ as } n \rightarrow \infty$$

8.1.8.2 *Monte Carlo Integration*

Monte Carlo methods are particularly useful for numerical integration, especially in high-dimensional spaces where traditional methods are inefficient.

Estimating an Integral: Suppose we want to estimate the integral of a function f over the interval $[a, b]$:

$$I = \int_a^b f(x)dx$$

Using Monte Carlo, we approximate this integral by sampling x_i uniformly from $[a, b]$ and taking the average value of $f(x)$:

$$I \approx \frac{b-a}{n}\sum_{i-1}^{n} f(x_i)$$

8.1.8.3 *Variance reduction techniques*

To improve the efficiency of Monte Carlo simulations, various variance reduction techniques can be employed:

1. **Antithetic variates:** Use pairs of dependent random variables whose combined variance is less than the sum of their individual variances.
2. **Control variates:** Use known properties of the control variable to reduce the variance of the estimator.
3. **Importance sampling:** Sample from a distribution that gives more weight to important regions of the integrand.

8.2 MACHINE-LEARNING CONCEPTS USED IN MODELS

8.2.1 CART MODEL

CART (Classification and Regression Trees) is a popular algorithm used for both classification and regression tasks in machine learning. It is a type of decision tree algorithm and was introduced by Breiman et al. in 1984[22]. CART works by constructing a binary tree structure, where each node represents a decision based on the input features. The tree splits the data recursively into smaller and more homogeneous subsets (known as "leaves"). At each split, the algorithm selects the best feature and threshold to minimize the impurity or error in the resulting subsets.

1. Key Concepts in CART ((Liu, 2018))

1.1. Decision Tree Structure

Root Node: The top-most node of the tree that contains the entire dataset. This is where the first decision or split is made.

Internal Nodes: Nodes within the tree that split data further based on certain feature values.

Leaf Nodes (Terminal Nodes): The final nodes where no further splitting occurs. These nodes represent the final output, either a class label (for classification) or a numeric value (for regression).

1.2. Splitting Criteria

The main idea behind CART is to split the dataset recursively into subsets that are more homogeneous in terms of the target variable.

For Classification Tasks:

The goal is to split the data into groups where most of the data in each subset belongs to a single class.

Impurity Measures:

[22] https://books.google.ro/books/about/Classification_and_Regression_Trees.html?id=JwQx-WOmSyQC&redir_esc=y

Gini Impurity: Measures the probability of incorrectly classifying a randomly chosen element. The Gini index is calculated as:

$$Gini(t) = 1 - \sum_{i=1}^{C} p(i|t)^2$$

where

$p(i|t)$ is the proportion of class iii observations in node t, and C is the number of classes.

Gini takes a value between 0 (pure) and 0.5 (maximum impurity for two classes).

Entropy/Information Gain: Measures the disorder or uncertainty of a set. It is calculated as:

$$Entropy(t) = -\sum_{i=1}^{C} p(i|t) \log_2\big(p(i|t)\big)$$

Information gain is the reduction in entropy due to splitting. CART generally uses Gini by default, but Entropy is used in other decision tree algorithms (like ID3, C4.5).

For Regression Tasks:

The goal is to minimize the prediction error (such as squared error or mean absolute error) at each split.

Impurity Measures:

Mean Squared Error (MSE): Measures the average of the squared differences between the predicted values and actual target values.

Mean Absolute Error (MAE): Measures the average of the absolute differences between the predicted values and actual target values.

$$MSE(t) = \frac{1}{N_t} \sum_{i=1}^{N_t} (y_i - \hat{y})^2$$

Explanation:

Dataset: A mock dataset is created with features like Temperature, Humidity, WindSpeed, VegetationType, Elevation, and the target variable FireRisk (1 = high risk, 0 = low risk). In practice, we would replace this with actual wildfire data.

Features (X): The independent variables that affect fire risk (e.g., temperature, humidity, wind speed).

Target (y): The dependent variable (FireRisk) which represents whether a location is at high or low risk for wildfires.

Train-Test Split: The data is split into training and testing sets using train_test_split.

Decision Tree Model (CART): The DecisionTreeClassifier is used to build the decision tree, and the model is trained on the training data (fit).

Evaluation: The model is evaluated on the test set using a confusion matrix and classification report to understand how well it performs.

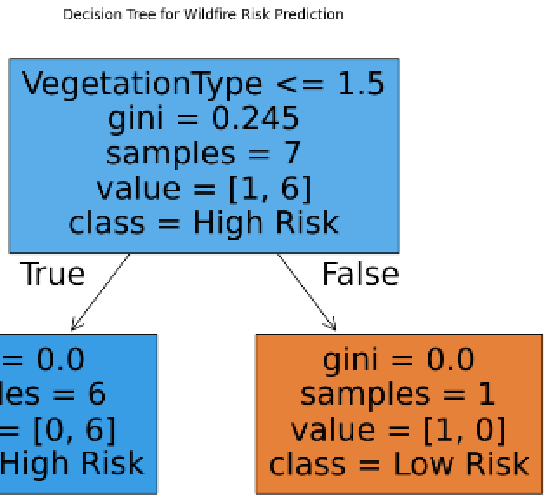


Figure 38 Wildfire risk decision tree

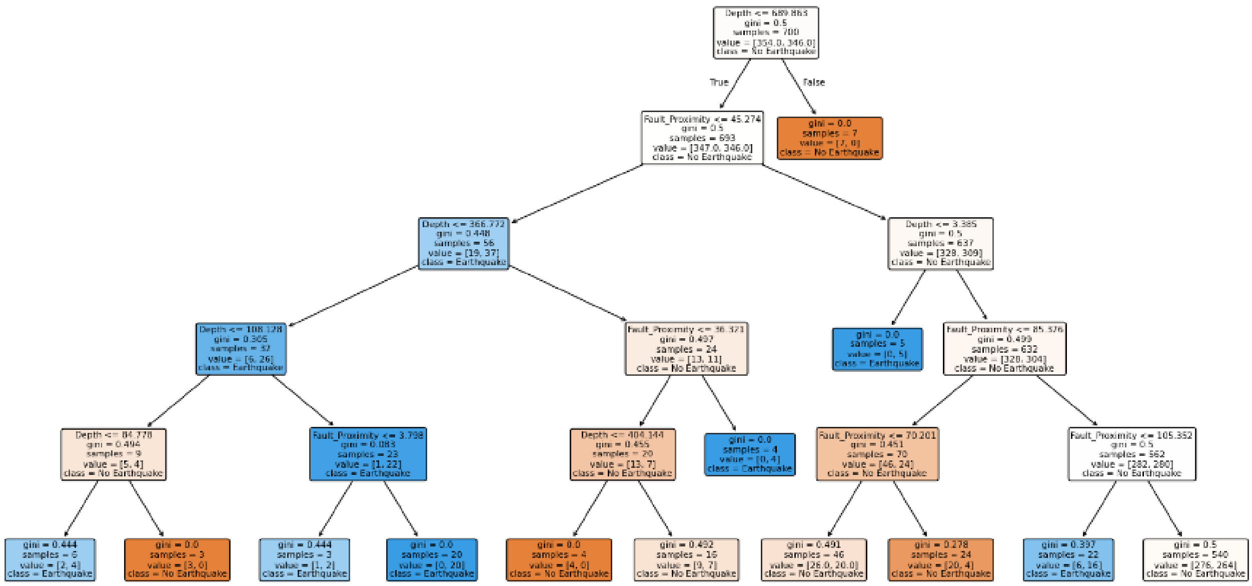


Figure 39 Earthquake risk (Decision Tree)

Tree Plot: We visualize the decision tree to understand how the model makes predictions. The decision tree splits the data based on the most important features (e.g., depth, fault proximity) to predict the earthquake magnitude.

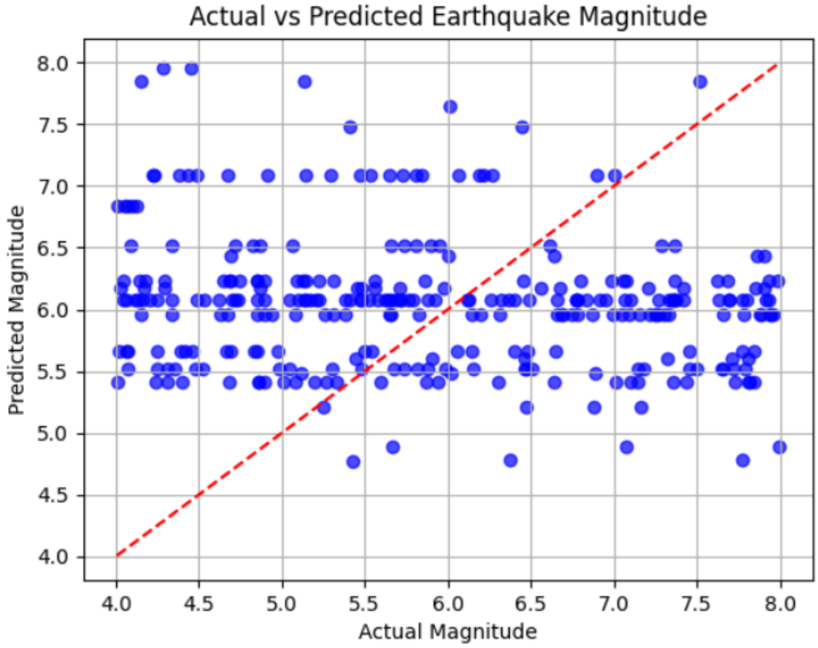


Figure 40 Earthquake predicted vs actual

Actual vs Predicted Plot: This scatter plot shows the relationship between the actual and predicted earthquake magnitudes. Ideally, the points should lie close to the red diagonal line, indicating that the predictions closely match the actual values.

8.2.2 Q-LEARNING (REINFORCEMENT LEARNING) ((ANAND DESHPANDE, 2018))

Q-learning is a classic RL algorithm, to demonstrate how to model a simple wildfire environment. This environment will be grid-based, where each cell represents a section of a forest. The agent's task will be to manage the fire and contain its spread by making optimal decisions.

Q-Learning is a reinforcement learning algorithm where the agent learns to take actions in an environment to maximize cumulative reward. The agent updates its knowledge through a Q-table, which maps states to action values.

Q-Learning Update Formula:

$$Q(s,a) \leftarrow Q(s,a) + \alpha \left(r + \gamma \max_a Q(s',a) - Q(s,a) \right)$$

Where:

- s is the current state,
- a is the current action,
- r is the reward,
- s' is the next state,
- α is the learning rate,
- γ is the discount factor.

The algorithm is applied to a grid where fire can spread, and the agent can take actions to control it. The agent learns through trial and error which actions (e.g., moving, putting out fire) minimize the damage caused by the fire.

Components of the Environment:

State: A 2D grid where each cell represents an area that can be on fire, burnt, or safe.

Actions: The agent can move in four directions (up, down, left, right) and put out a fire in the current cell.

Reward: The agent is rewarded for putting out fires and penalized when the fire spreads.