Critical Action Planning over Extreme-Scale Data

D2.3 - Final Use Case Evaluation, Pilots, Demonstrators and Simulation Models and Tools

Version 1.1

Documentation Information

Contract Number	**101092749**
Project Website	https://crexdata.eu/
Contractual Deadline	M36, 12.2025
Dissemination Level	Public
Nature	Report
Author	Arnau Montagud (BSC), Miguel Ponce de Leon (BSC), Thalia Diniaco (BSC), Isabel Martínez (BSC), Giannis Spiliopoulos (UAEGEAN), Dimitris Zissis (UAEGEAN), Marcel Ebel (UPB), Jens Pottebaum (UPB)
Contributors	Piergiorgio Castioni (URV), Hadrien Calmet (BSC), Georgios Grigoropoulos (KPLER), Taina Hanhikoski (MoIFi), Elias Xidias (UAEGEAN), Christina Tsikou (UAEGEAN), Jenny Papaziou (UAEGEAN), Thomas Kogias (UAEGEAN), Georgia Spyridaki (UAEGEAN), Nikos Sapountzis (UAEGEAN), Iris Gräßler (UPB), Niklas Döhner (UPB), Deniz Özcan (UPB), Jasmina Schmidt (DCNA), Enes Derin (FDDO), Ilona Lang-Ritter (FMI), Lisa Haga (FMI)
Reviewer	Elias Alevizos, Charilaos Akasiadis, George Katsimpras (DEMOKRITOS)
Keywords	Use case evaluation, pilots, demonstrator, forecasting, online data analysis

CREXDATA has received funding from the European Union's Horizon Europe programme under grant agreement number 101092749.

Change Log

Version	Author	Date	Description Change
V0.1	Miguel Ponce de León, Isabel Martínez (BSC)	17/09/2025	First Structure, TOC
V0.2	Miguel Ponce de León, Arnau Montagud, Isabel Martínez, Thalia Diniaco (BSC)	22/10/2025	Adjustment of the structure & partner responsibilities
V0.3	Giannis Spiliopoulos, Dimitris Zissis (UAEGEAN), Georgios Grigoropoulos (KPLER)	29/10/2025	Adjustment of the structure of the Maritime use case
V0.4	Miguel Ponce de León, Isabel Martinez (BSC)	10/11/2025	Add the Epidemiological Scenario of the Health Crisis Use Case
V0.5	Giannis Spiliopoulos, Elias Xidias, Dimitris Zissis, Christina Tsikou, Jenny Papaziou, Nikos Sapountzis, Thomas Kogias, Georgia Spyridaki (UAEGEAN), Georgios Grigoropoulos (Kpler)	17/11/2025	Added maritime use case
V0.6	Arnau Montagud, Thalia Diniaco (BSC)	20/11/2025	Add the Multiscale infection Scenario of the Health Crisis Use Case
V0.7	Marcel Ebel (UPB)	23/11/2025	added Weather Emergency Use Case
V0.8	Thalia Diniaco (BSC)	24/11/2025	Finishing Conclusions
V0.9	Elias Alevizos, Charilaos Akasiadis, George Katsimpras (DEMOKRITOS)	09/12/2025	Revisions done
V0.95	All authors	15/12/2025	Revisions corrected
V0.99	Marcel Ebel (UPB), Giannis Spiliopoulos (UAEGEAN), Arnau Montagud (BSC)	19/12/2025	Format of tables, figures and references
V1.0	Antonios Deligiannakis (TUC)	25/12/2025	Final approval by the Coordinator
V1.1	Arnau Montagud (BSC), Marcel Ebel (UPB), Georgios Grigoropoulos (Kpler), Giannis Spiliopoulos (UAEGEAN)	15/03/2026	Adding Section 5

Contents

Executive Summary

Deliverable D2.3 presents the comprehensive evaluation of CREXDATA use cases, pilots, demonstrators, and simulation models and tools using a harmonized methodology to ensure consistency across applications.

The evaluations have affirmed the practical relevance and impact of the CREXDATA platform. Pilots and demonstrators have validated its applicability in real-world contexts, while simulation models and analytical tools have been employed for hypothesis testing and solution validation under controlled conditions. Final scenario definitions for each use case have clarified the operational challenges addressed by CREXDATA, guiding technology development accordingly.

The Weather Emergencies Use Case presents results in two scenarios: urban flooding and large-scale wildfires. For the final evaluation of the integrated system, use cases were specified per CREXDATA technology based on the results of the first evaluation phase in the first half of the project. Domain-specific simulators were combined in a Simulator HyperSuite and made available for use by end users. The technologies and simulators were integrated in the EmCase Demonstrator System, tested at lab-scale and evaluated at the pilot sites in Dortmund and Innsbruck together with experts from emergency management. In addition, the transferability of the technologies to other scenarios was investigated in a large-scale emergency management exercise in Kuopio (Finland). The technologies and simulators support decision-makers in critical situations in the decision-making process.

The Health Crisis use case presents results in both scenarios: epidemiological and mechanistic infection, modelling to support pandemic response and patient-specific treatment optimization. It leverages advanced simulators and federated learning to improve clinical decision-making. In the Epidemiological scenario we combine simulation-based and data-driven approaches to improve epidemic preparedness and response. It includes pilots on epidemic parameter calibration, vaccination and mobility intervention optimization, and the use of information-theoretic and visual analytics methods to uncover the relationship between population mobility and epidemic spread. In the multiscale infection scenario, we used Rapidminer's AI platform to integrate two simulators and a supercomputer to allow for better parameter characterisation, the early classification of simulation trajectories and the modification of parameters at runtime. This architecture allowed us to identify potential drug targets that avoid the further infection of epithelial cells by SARS-CoV-2.

The Maritime use case presents results of hazardous weather rerouting (HWR) and collision avoidance (CA) scenarios that enhance maritime safety. Services are implemented to scale for global execution. A small autonomous surface vehicles (ASV) fleet was used to create realistic validation conditions in a constrained environment. Services leverage a voyage data streamer IoT device for real-time telemetry data, a Multi-Modal data fusion toolbox for fusing data from various modalities, and Rapiminer's AI platform for workflow design to create reliable digital twins within the CREXDATA platform. Technologies were validated through a three-phase pilot progression during the Aegean Ro-Boat Races, progressing from passive data acquisition to active remote control and autonomous navigation. Context enrichment was demonstrated through technology demonstrators in data fusion, augmented reality, advanced visual analytics, and wave analysis. The final evaluation used an ASV fleet that executed autonomous validation scenarios at sea, demonstrating sub-second latency and COLREG-compliant rerouting in collision avoidance scenarios exceeding 80% accuracy. An

expert mariners group confirmed the system’s ability to accurately forecast and provide correct rerouting solutions in more than 80% of hazardous weather cases. These capabilities support Vessel Traffic System operators and remote crews by providing early detection of maritime events and timely, safe rerouting solutions before incidents occur.

Results from the earlier evaluations were used as critical feedback to improve CREXDATA solutions further, addressing both technical and user-centric aspects. This deliverable consolidates the project’s progress and presents its innovative technologies as they meet the needs of emergency and crisis management.

1 Introduction

This deliverable D2.3 concludes the comprehensive evaluation and refinement of the CREXDATA project's three primary use cases and marks the culmination of development efforts initiated in Deliverable D2.2. This final report synthesizes results from the Health Crisis, Weather Emergencies, and Maritime use cases, documenting the maturation of pilots and demonstrators from initial prototypes to operationally relevant systems. The deliverable addresses the complete lifecycle of use case implementation, encompassing simulation models and tools, pilot site deployments, expert user validation, lessons learned from field trials, and prospects for post-project sustainability and exploitation.

The CREXDATA project represents a notable contribution in emergency management and critical action planning, transitioning from reactive, offline decision-making approaches grounded in historical practices to proactive, online methodologies informed by extreme-scale data analytics and predictive forecasting. Across its three complementary domains, weather-induced emergencies, health crisis management, and maritime safety, CREXDATA has developed integrated technological ecosystems that couple sophisticated simulation models with advanced machine learning, complex event forecasting, and visual analytics capabilities. The overarching vision is to enable decision-makers, including civil protection authorities, health agencies, and maritime professionals, to anticipate critical situations, evaluate intervention strategies, and execute coordinated responses with confidence under conditions of profound uncertainty and data complexity.

The project consists of three use cases that use different technologies developed in the project (Table 1). In the domain of the Weather Emergency Use Case, CREXDATA delivers tools for situational awareness, rapid risk assessment, and action planning under conditions of flooding, wildfires, and other extreme events. Through integration with physical simulators, data from distributed sensors, and real-time social media feeds, the system generates and tests hypotheses about evolving threats, enabling stakeholders to evaluate alternative intervention strategies and resource allocations. Augmented reality visualizations and uncertainty-aware dashboards provide actionable insights to field operators and command centres, while federated learning on decentralized sensor networks bolsters prediction accuracy and system robustness despite heterogeneous and incomplete data streams.

The Health Crisis Use Case demonstrates CREXDATA’s ability to bridge epidemiological and molecular scales using advanced computational tools. It combines two scenarios, one with large-scale epidemic models (integrating mobility and incidence data) and another with multiscale mechanistic models of viral infection in the lung, producing patient-specific digital twins that simulate both disease spread and individual response to therapies. This multiscale approach supports the optimization of interventions, such as vaccination rollout, drug treatment, and ventilation strategies, and provides clinicians and policymakers with interactive analytics and transparent model explanations.

For the Maritime Use Case, CREXDATA enhances operational safety at sea by forecasting hazardous situations such as collisions, severe weather, or navigational anomalies. The platform streams and fuses data from onboard sensors, automatic identification systems (AIS), and environmental forecasts, producing real-time digital twins of vessel movements and situational risk. Simulation-driven forecasts and explainable AI tools deliver timely alerts, recommended evasive maneuvers, and adaptive routing, supporting crew, fleet operators, and coastal authorities in making informed decisions under evolving maritime conditions.

This report consolidates findings from 18-36 months of project execution, including results from pilot deployments at designated sites, expert user evaluations using standardized methodologies, technical validation of demonstrator systems against predefined success criteria, and iterative refinement informed by stakeholder feedback. Cross-cutting themes include the effectiveness of graphical workflow design tools in reducing complexity, the value of interactive learning and visual analytics in enabling expert users to explore high-dimensional parameter spaces, and the importance of transparent, explainable AI methods in building stakeholder confidence in algorithmic recommendations.

The deliverable is organized into sections addressing the final status of use case scenarios, descriptions of advanced simulators and their integration, characterization of pilots and demonstrators deployed at operational sites, results of expert user evaluations using both quantitative (System Usability Scale, KPI achievement) and qualitative methods, and synthesis of lessons learned regarding technical feasibility, user adoption, and pathways toward post-project exploitation and sustainability.

The report furthermore outlines recommendations for technology transfer, identifies prioritized opportunities for commercialization and open-source community contribution, and discusses integration pathways with existing operational systems and future EU research initiatives.

A glossary of terms has been added for the benefit of readers (Table 2).

Table 1: Uptake of technologies in Use Cases (cf. [D2.1])

	Emergency Case		Health Case	Maritime Case
	Dortmund	Austria		
T2.4 Simulation and Tools	X[3]	X[3]	X	X
T3.2 Graphical Workflow Specification	X[2]	X[2]	X	X
T4.1 Complex Event Forecasting	X[3]	X	X	X
T4.2 Interactive Learning for Simulation Exploration	X[2]	X[2]	X	X
T4.3 Federated Machine Learning	X	X	X	
T4.4 Optimized Distributed “Analytics as a Service”	X[2]	X[2]		X
T4.5 Text Mining for Event Extraction	X	X		
T5.1 Explainable AI	X[2]	X[2]	X	
T5.2 Visual Analytics supporting XAI	X[2]	X[2]	X	
T5.3 Visual Analytics for Decision Making under Uncertainty	X[2]	X[2]	X	X[1]
T5.4 Augmented reality at the field	X	X		X[1]
T5.5 Uncertainty Visualization in Augmented Reality	X[2]	X[2]		X[1]

X[1] support of potential TUC, FR contribution, X[2] scheduled for M19-M36, X[3] initial exploration in first trials

Table 2: Glossary

Abbreviation	Expression	Explanation
UC	*Use Case*	Applications of CREXDATA technology resp. the CREXDATA system in real-world scenarios. Within the project three use cases are defined: weather induced emergencies, health and maritime.
	Pilot (site)	Conceptual term to describe a set of stakeholders within their context like spatial environment, equipment, data sources etc. For each Use Case, several Pilot (sites) can be specified (for instance, Dortmund and Austria in the emergency case).
	Application Scenario	Procedural and structural description of potential uptake of CREXDATA technologies

Abbreviation	Expression	Explanation
		in Use Cases (for instance, flooding and forest fires in the emergency case).
	CREXDATA system	Output of WP3, integrating technologies created in WP4 and WP5 without use case-specific customizations. It includes customization and configuration functionality, esp. through graphical workflow management.
	Demonstrator (system)	Technical system based on the CREXDATA system, which is customized and configured for specific Use Cases, Pilots and/or Application Scenarios. The Demonstrator might include additional components both as data sources and sinks (for instance, legacy systems of end users or the ARGOS system in the emergency case).

2 Weather-Induced Emergencies Use Case

In line with the case-study methodology outlined in Deliverable D2.1 (p. 23, see Figure 1), the overarching research questions were further specified within the conceptual design of the Weather-Induced Emergencies Use Case (EmCase), and more precisely for implementation at the pilot sites. EmCase integrates the developed technologies into the disaster response and control environment for extreme weather events through visualization in ARGOS and Augmented Reality. The specific procedure in the EmCase was presented in Figure 1 of Deliverable 2.2. In accordance with all preparations, three guiding research questions were developed and described in Deliverable D2.2 (p. 11):

a. Does information generated by artificial intelligence algorithms need to be visualized differently from information generated by traditional situational awareness and assessment?
b. Does it make a difference whether the uncertainty of information provided to enable situational awareness is based on algorithms or witnesses on the ground?
c. What influence does an algorithm-based recommendation have on action planning and decision?

Section 2.1 introduces a general overview, setting the frame with regard to the context of evaluation activities at all three pilot sites Aligned with the other use cases, Section 2.2 outlines domain-specific assumptions and environmental parameters, which help translate the research questions into concrete test scenarios and technology-specific test cases. Section 2.3 then focusses on the scenario descriptions and the demonstrator setup within the EmCase, which combines CREXDATA components with domain tools such as ARGOS, focusing on the environments of Innsbruck, Dortmund and Finland. Section 2.4 focuses on the Simulators used in the EmCase, while the results of the final use case evaluation and its interpretation are described in Section 2.5. The use case is concluded with lessons learned in Section 2.6.

2.1 Introduction

Extreme weather events such as urban floodings and wildfires are becoming more frequent, partly due to effects of climate change. For example, urban flooding is a significant threat to densely populated regions in Europe. During episodes of extreme precipitation, water rapidly accumulates within urban drainage systems, often resulting in river overflows, flooded streets and the flooding of basements and underpasses. Within minutes, entire neighborhoods can become inaccessible, posing considerable challenges for first responders. These challenges include evacuating vulnerable populations, maintaining access to critical infrastructure such as schools, hospitals and nursing homes and ensuring the safety of responders themselves in the presence of unstable ground, hidden currents or obstructed streets. An Emergency Case demonstrator system was set up to evaluate the CREXDATA technologies that had been developed. The ARGOS system forms the basis for this. ARGOS is both a data source (e.g., for providing weather data) and a data sink for visualizing all technology outputs in the EmCase. The demonstrator system was used to conduct a number of small and mid-size lab tests as well as the final evaluation together with experts at the pilot sites in Dortmund, Innsbruck, and Finland. The following Table 3 lists the evaluation activities in the second half of the project, which were carried out in development iterations and final evaluations together with end users. In preparation for the final evaluations in Dortmund, Innsbruck, and Kuopio, smaller system and subsystem tests were carried out in the form of lab tests. The aim was to gather specific feedback for optimizing individual subsystems and components of the Emergency Case Demonstrator System.

Table 3: Evaluation activities regarding the EmCase

evaluation dates	activity	location	category
10 + 11 Jun 24	M18 Trial Innsbruck	Innsbruck	trial
13 + 14 Jun 24	M18 Trial Dortmund	Dortmund	Trial
27 May 25	Text Mining / Gazebo online test	Dortmund	end-user lab test
11 + 12 Jun 25	ESAF R&D days	Kuopio	end-user lab test
7 Jul 25	AR workshop	Dortmund	end-user lab test
3 + 4 Oct 25	Trial Innsbruck	Innsbruck	field trial
27 + 30 + 31 Oct 25	Trial Dortmund	Dortmund	field trial
18 + 19 Nov 25	ESAF exercise	Kuopio	field trial

2.2 Scenario Description

Urban flooding and wildfires are addressed as complex socio-technical challenges rather than purely physical hazards. Both scenarios combine environmental dynamics with human, infrastructural, and organizational dimensions, illustrating the multifaceted nature of modern disaster management. The urban flooding scenario integrates weather events, critical infrastructure (transport networks, stadiums, public areas), communication flows (including social media), and operational decision-making processes to reflect the practical and organizational challenges of coordinating a large-scale response under uncertainty. Similarly, the wildfire scenario focuses on mountainous and rural terrains, where topography, vegetation, wind conditions, and limited accessibility necessitate the coordination of multiple stakeholders such as air services, volunteer and professional fire brigades, and mountain rescue units. It highlights the interplay between rapid situational assessment, air-ground cooperation, and dynamic resource allocation. The specific scenarios addressed in the CREXDATA EmCase are described in D2.2.

In both scenarios, almost every CREXDATA technology is tested through assigned phases and specific test cases, enabling participants to engage directly with the system, explore its functionalities, and provide structured feedback on usability, reliability, and operational impact. This comprehensive approach ensures that each pilot trial serves not only as a demonstrator of technological innovation, but also as an empirical evaluation of the system's practical applicability in real-world emergency response contexts-spanning from densely built urban environments to remote alpine firefighting operations.

2.3 Pilot Definition and Demonstrator

2.3.1 Pilot Definition

The following subsections describe the variety of test scenarios in Dortmund and Innsbruck for evaluating the CREXDATA technologies and EmCase demonstrator system for critical action planning in urban flooding and wildfire scenario. In addition to that, an existing large-scale exercise of rescue services in Finland is used to evaluate the transferability of CREXDATA to other emergencies beside floods and fires.

2.3.1.1 Urban Flooding Scenario Innsbruck

The test scenario (see Table 4) models a severe rainfall event in the urban area of Innsbruck, designed to evaluate the functionality and interoperability of the CREXDATA system under realistic emergency conditions. The environmental setting involves intense pluvial flooding resulting from sustained heavy rainfall. Among the affected sites is a shopping mall operating during regular business hours, where an unknown number of individuals become trapped due to the rapid onset of flooding. In response to the situation, all participating emergency response organizations are alerted and dispatch personnel and equipment to the affected locations. Based on early meteorological forecasts, civil protection authorities had anticipated the potential for such an event and assessed its criticality as corresponding to a 50-year rainfall probability. As a result, predefined response and coordination plans were activated in advance of the event. At the initiation of the test scenario, rainfall begins at the western periphery of Innsbruck, gradually moving toward the city center. The water level of the River Inn subsequently rises west of the city, contributing to localized flooding. Concurrently, Tyrol Control Center experiences sharp increase in emergency calls related to the heavy rainfall. As the situation increases, management responsibility is transferred to the Innsbruck Fire Department. To further challenge situational awareness and decision-making, the scenario introduces emergent and unanticipated conditions, including:

- rainfall exceeding initial forecasts
- obstructions within the municipal drainage and sewer systems
- inundation of critical buildings and infrastructure

The stakeholders tested the CREXDATA EmCase demonstrator system within a coordinated operational environment. The trial aimed to assess the system's effectiveness in supporting situational analysis, risk identification, tactical planning, and resource allocation in a complex, data-rich crisis scenario.

Table 4: Test Scenario Innsbruck Heavy Rain

Test_Scenario_Name	Test Scenario Innsbruck Heavy Rain
Test_Scenario_ID	TS_003
Test_Scenario_Owner	DCNA (UPB)
Test_Case_IDs	TC_001 ARGOS system TC_016 SHS GUI TC_019 FloodWaive TC_004 CEF flooding TC_005 Parameter exploration in urban flooding simulation model – Barrier Optimizer TC_006 Complex Event Forecasting TC_007 Text mining
Test_Item(s)	Several CREXDATA sub-systems: ARGOS-system, FloodWaive Simulation, Parameter exploration (barrier optimizer), complex event forecasting, FloodWaive
Test_Scenario_ Activities	• View situational map and available forecasts for an expected emergency

	• View situational map of flooding situation, view options in urban flooding models → create multiple simulations with current forecast parameters as well as best and worst case scenario • Measures in the form of barriers are defined and implemented → Parameter exploration → Calculation of the optimal height of barriers through a large number of simulations • Sewer system is overloaded due to excess water in its system → CEF to detect potential points where excess water gets pushed out on the streets • Provide feedback: Complete SUS questionnaire + joint discussion → capture usability, perceived value, and impact on decision-making
Test_Scenario_ Environmental_Parameter	Derived from precipitation data of heavy rain event in July 2016 in the Amras district in Innsbruck

2.3.1.2 Urban Flooding Scenario Dortmund

The test scenario of a two-day event reproduces a complex urban flooding scenario in a controlled but realistic setting, combining both strategic decision-making and field operations. Additionally, a workflow modelling workshop complements the trials, ensuring that all technologies are integrated within coherent emergency management procedures.

On the first day, the focus is on command room technologies tested within a simulated command room replicating an actual operations center. Representatives from designated staff functions - S2 (Situation), S3 (Operations), S5 (Media/Press) and the Head of Staff - use a combination of predictive analytics, text mining, eXplainable AI and pre-developed workflows. The scenario simulates an urban flooding event coinciding with a European Championship match in Dortmund based on data of the EM2024 match between Germany and Denmark, creating additional complexity as large crowds gather at public venues and transportation hubs. Initially, the situation seems manageable; however, as heavy rainfall intensifies, emergency services face a surge of incidents including flooded basements and blocked roads. This setup allows a systematic evaluation of decision-support technologies under time-constrained conditions, emphasizing situational awareness, information management and coordinated planning. Key assessment metrics include usability, comprehension of uncertainty and the influence of system outputs on operational decisions.

The second day focuses on field-based validation of the flooding scenario. Operational stakeholders of the Dortmund Fire Service including firefighting teams, incident commanders and drone operators engage directly with the technologies deployed in the field. Augmented Reality tools enable firefighters to navigate flooded areas, visualize water levels, flow directions and potential worst-case developments, and receive dynamic routing information that adapts to blocked streets. In parallel, robotics and drones are tested to scout inaccessible areas, transmit live video and sensor data, and communicate critical information such as hazards or safe entry points directly to the AR devices used by field teams.

Carefully designed to align with and validate the core objectives of CREXDATA, the Dortmund trial demonstrates how predictive analytics, explainable AI, visual analytics and augmented reality tools can be effectively integrated into the operational workflows of emergency management authorities. The urban flooding test scenario (see Table 5), unfolding alongside a large-scale public event, provides a realistic stress test of these technologies and their contribution to more informed, coordinated and adaptive disaster response. To replicate real-world complexity, the trial incorporates advanced features such as uncertainty visualization, displaying confidence levels and explanatory details regarding data reliability and fatigue prediction systems, which allow incident commanders to monitor the physical and cognitive resilience of their teams. These elements support evaluation of not only technical functionalities but also their impact on decision-making, operational efficiency and responder safety under realistic stress conditions.

Table 5: Test Scenario Flooding Dortmund

Test_Scenario_Name	Test Scenario Flooding Dortmund
Test_Scenario_ID	TS_006
Test_Scenario_Owner	FDDO (UPB)
Test_Case_IDs	TC_002 ARGOS TC_016 SHS GUI TC_019 FloodWaive TC_008 Text Mining TC_018 Gazebo TC_004 CEF TC_005 Interactive Learning Optimizer TC_009 XAI TC_006 NeRF and FML TC_020 Mental Fatigue Assessment TC_013 Augmented Reality routing Dortmund TC_014 AR visualizing points of interests Dortmund
Test_Item(s)	CREXDATA system in general with its several CREXDATA sub-systems
Test_Scenario_ Activities	Command Room: • ARGOS • SHS GUI • FloodWaive • Text Mining • Gazebo • CEF • Interactive Learning Optimizer • XAI On-Site: • ARGOS • NeRF and FML • Mental Fatigue Assessment with AR • AR drone data visualization • AR flooding system
Test_Scenario_ Environmental_Parameter	Derived from precipitation data and operation data of heavy rain event in June 2024 in Dortmund

2.3.1.3 Wild Fire Scenario Innsbruck

The test scenario (see Table 6) is closely aligned with a real-life air service exercise carried out by multiple fire brigades of Innsbruck and surrounding villages. It assumes a wildfire in steep alpine terrain in the Stubaital region. Due to the topography, the fire spreads quickly uphill and threatens forested slopes above Telfes as well as nearby touristic and residential infrastructure. Early reconnaissance indicates that the fire line is expanding toward critical areas, with the potential to endanger hiking trails, cable car infrastructure, and isolated farm buildings.

Responding to this situation, the district fire brigades of Stubaital, the Innsbruck fire service, the Tyrol air service, and the mountain rescue service are alarmed. Helicopter operations are activated according to the special alarm plan for wildfires. A dedicated operations base is set up at the Schlickerboden area, serving as take-off point, water supply location, and coordination hub.

The exercise assumes that initial fire suppression attempts from the ground have been challenged by steep terrain and limited accessibility. Therefore, aerial firefighting operations are prioritized, including reconnaissance flights, bucket drops, transport of equipment to the fire line, and shuttling of crews to the fire site via helicopter sling load operations.

Table 6: Test Scenario Innsbruck Wildfire

Test_Scenario_Name	Test Scenario Innsbruck Wildfire
Test_Scenario_ID	TS_004
Test_Scenario_Owner	DCNA (UPB)
Test_Case_IDs	TC_001 ARGOS TC_002 Text Mining TC_016 SHS GUI TC_018 Gazebo TC_017 SPARK TC_009 Physical Fatigue TC_012 AR visualizing points of interest TC_011 AR routing & navigation
Test_Item(s)	CREXDATA system in general with its several CREXDATA sub-systems: ARGOS-system, text mining, Gazebo, mental fatigue assessment, SPARK, physical fatigue assessment, Augmented Reality (hardware/software-application)
Test_Scenario_ Activities	• Receive alarm call: Smoke reported in alpine area, exact location unclear → set incident category/level, initiate scenario • Analyze social media reports: Use CREXDATA Text Mining to identify potential fire-related → evaluate added value for situational awareness, list potential POIs • Decide on drone reconnaissance: Select POIs based on text-mining results • Simulate drone flight: Use GAZEBO to plan reconnaissance flight route → verify POIs • Perform live drone reconnaissance: Fly over POIs, collect visual confirmation • Simulate wildfire progression: Run SPARK model with local conditions (wind, topography, vegetation) • Make tactical decision: Combine drone input + simulation results to decide on suppression strategy • Monitor physical fatigue: Collect vital data of responders, determine when to rotate staff • Conduct evaluation / hot wash: Complete SUS questionnaire and join structured debrief → capture user experience, improvement suggestions, and operational relevance Augmented Reality:

	• Assume initial situation: Firefighting units dispatched to wildfire in alpine terrain → set incident context, brief participants on scenario • View situational picture in AR: Fire hotspots, vulnerable areas (critical infrastructure, settlements) and positions of other units displayed • Use AR navigation: Receive routing guidance to key POIs (water sources, danger zones, assembly points)
Test_Scenario_ Environmental_Parameter	Wildfire in an alpine environment with vulnerable touristic infrastructure in the vicinity – dynamic scenario closely aligned with the real-life wildfire exercise

2.3.1.4 Wildfire Scenario Dortmund

In Dortmund, a real forest fire scenario (cf. test scenario in Table 7) in the surrounding area was reenacted shortly before the final evaluation. In June 2025, a forest fire broke out at Lake Möhnesee after several days of extreme heat and persistent drought had dried out the vegetation. Strong winds quickly drove the flames across a difficult-to-access slope, causing the fire to spread rapidly. The fire department had to use ground forces and drones to assess the situation, as smoke temporarily severely restricted visibility. For safety reasons, nearby hiking trails and a campground were evacuated. After several hours of intensive firefighting, the emergency services managed to stabilize the fire and prevent it from spreading further towards residential areas.

Table 7: Test Scenario Wildfire Dortmund

Test_Scenario_Name	Test Scenario Wildfire Dortmund
Test_Scenario_ID	TS_005
Test_Scenario_Owner	FDDO (UPB)
Test_Case_IDs	TC_002 ARGOS TC_016 SHS GUI TC_017 Spark TC_018 Gazebo TC_002 ARGOS
Test_Item(s)	CREXDATA system in general with its several CREXDATA sub-systems: ARGOS, Gazebo, SPARK
Test_Scenario_ Activities	• Receive alarm call: Smoke reported in wild area, exact location unclear → visualization in ARGOS • Simulate wildfire progression: Run SPARK simulation with local conditions (wind, topography, vegetation) • Visualize simulated fire spread in ARGOS • Decide on drone reconnaissance: Select POIs based on critical points in the area effected • Simulate drone flight: Use GAZEBO to plan reconnaissance flight route for verifying POIs • Visualize best routing options in ARGOS
Test_Scenario_ Environmental_Parameter	Wildfire in with vulnerable touristic infrastructure – dynamic scenario closely aligned with the real-life wildfire exercise

2.3.2 Emergency Case Demonstrator System

As a result of the evaluation activities carried out in month 18, various measures for optimizing the system were defined. To increase usability for the end user, the ARGOS system from HYDS is used as a central component of the EmCase demonstrator system. Figure 1 below shows the system architecture of the EmCase demonstrator system. Based on the project results reported in D2.1 and D2.2, the final system architecture is explained below. In the demonstrator system, various data sources act as suppliers for simulation models and CREXDATA WP4 and WP5 technologies. For example, ARGOS reads various types of data, such as radar, weather data, forecasts, or sensor data from weather stations, and injects it into the demonstrator system via defined real-time workflows and interfaces. In addition, it is also possible to retrieve and inject historical data via ARGOS. Mobile sensor systems such as UAVs represent another data source. In addition, posts from social networks are used and analyzed as a data source. Furthermore, devices that collect biometric data from first responders are used and data streams are generated. The aforementioned input data streams are then fed into the simulators of the EmCase (FloodWaive, Gazebo robotic simulation, SPARK wildfire simulation, and the tool for emergency tasks and traffic risks by FMI) and the various CREXDATA technologies.

The detailed applications of the technologies in the EmCase are described technically in Deliverables 4.2 and 5.2. In the Complex Event Forecast (T4.1), data streams from weather data and sensor data from the sewer network are merged to identify complex events in the form of overflowing sewer manholes. The Interactive Learning for Simulation Exploration (T4.2) technology is used to decide on parameters of barriers set to influence water distribution in urban areas during the FloodWaive flood simulation. Federated Machine Learning (T4.3) is performed in the Emergency Case through the application of Neural Radience Field (NeRF). Drones are used to capture images of the operational environment. The drone images are then transferred to the computer, where the camera location and camera angle are calculated in the pose calculation. This creates a data set that is used to generate the NeRF. To speed up the calculation of the NeRF, the data set is divided into several areas using federated machine learning technology, calculated individually, and then recombined for visualization. Text Mining for Event Extraction (T4.5) is used in EmCase to classify relevant social media posts related to flooding and forest fires. The first step is to perform a relevance classification. Using a specially developed GUI, the end user can view and select the relevant posts. The selected posts can then be transmitted to ARGOS via a Kafka topic connection and visualized. For more in-depth information, a question answering model (Q&A model) has also been developed, which is based on a language model and allows users to ask specific questions that are answered based on the content of the social media posts. Explainable AI (T5.1) is used in EmCase for fatigue assessment based on biometric data. For this purpose, biometric data was collected during realistic respiratory protection exercises at FDDO. The data was used together with CNR to train a model that outputs a fatigue level (see deliverable D5.2). Based on the output, the explainability of the fatigue assessment is then provided by XAI. Augmented reality (T5.4/T5.5) is the other option in EmCase, alongside ARGOS, for visualizing simulation and technology results. The use cases in Augmented Reality (AR) can be divided into specific test scenarios. In the AR flooding system, mission-relevant information such as manholes in the ground or points of interest (POIs) are displayed. In addition, the results of the flood simulation can be augmented in the form of water level, water flow speed, and direction. In the AR fire system, wildfire-relevant information is augmented into the real world. Here, the results of the wildfire simulation are visualized in the form of fire line, arrival time, or flame height. In addition, a component in AR supports the use of drones. Members of a drone unit can track live data from the drone, such as position, images, or videos. A specific role for the drone pilot was also developed. The drone pilot can set POIs in AR and report hazards, which are then displayed to team members to speed up the flow of information during the operation. Within the EmCase, the demonstrator system was first developed conceptually (cf. Figure 1).

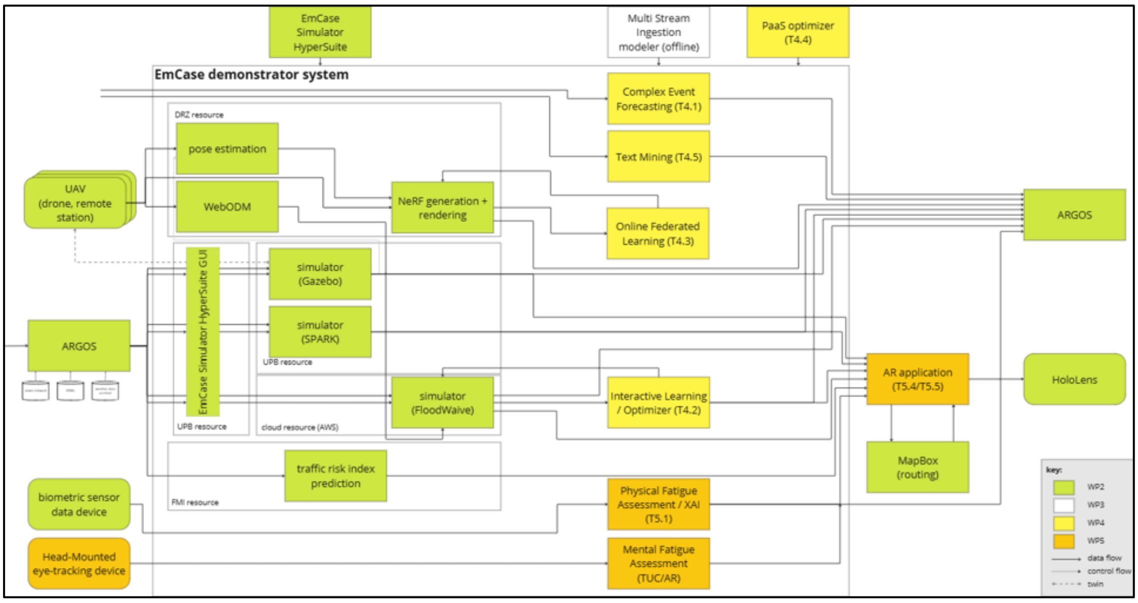


Figure 1: Final system architecture of EmCase demonstrator system

Conceptual workflows were then derived from the system architecture. These workflows were set up in the RapidMiner Altair AI Studio modeling environment within the CREXDATA extension and the EmCase extension; they were modeled according to the applications described (cf. Figure 2 below with exemplary Gazebo workflow). All EmCase workflows are documented in the Appendix 1 (Section 9.2). All workflows listed in the appendix have been integrated into interoperable processes. The processes are stored in the repository and include data collection, processing, and output of the processed data. The EmCase system architecture has been implemented into various extensions (.jar files). Key results were incorporated into various RapidMiner Custom Extensions, examples of which include the EmCase Simulator HyperSuite Extension and the EmCase Operator Extension. The CREXDATA Simulator HyperSuite Extension comprises various custom operators based on the defined workflows. The CREXDATA extension offers a simple and intuitive level of abstraction for using the various simulators. The EmCase operators include, for example, the various technologies from WP4 and WP5, like Text Mining, Biometric Data, and Fatigue Assessment. The three EmCase simulators and the Data Injector are embedded in the 'EmCase Simulator HyperSuite' custom extension.

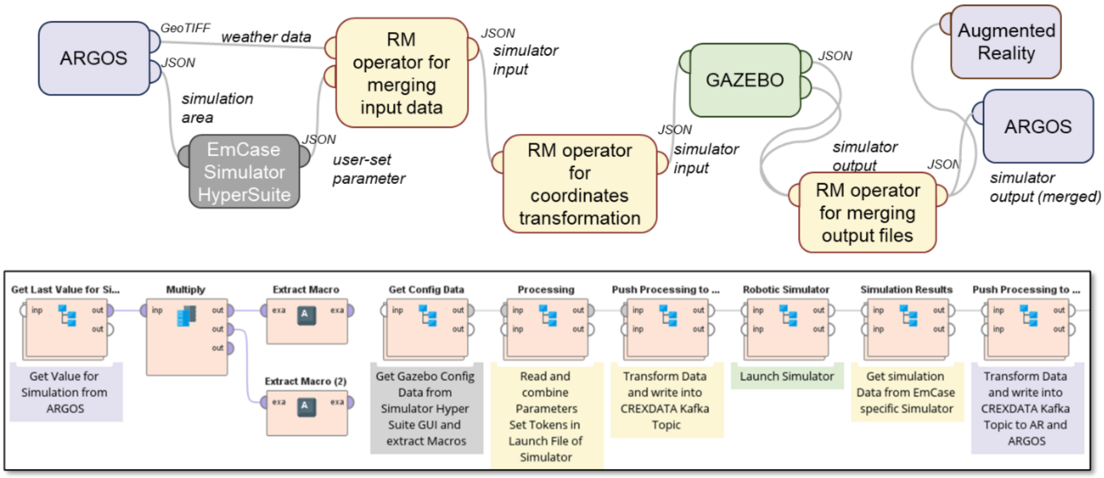


Figure 2: Conceptual Workflow transferred to RapidMiner Workflow (example: Gazebo)

The operators shown in the extension include executable simulators and underlying workflows. The various operators from the custom extensions can be executed in RapidMiner AI Studio and on the RapidMiner AI Hub. On the left side of Figure 3, the CREXDATA Simulator HyperSuite Custom Extension from RapidMiner AI Studio is shown. Various extensions (folders (.jar files)) can be seen.

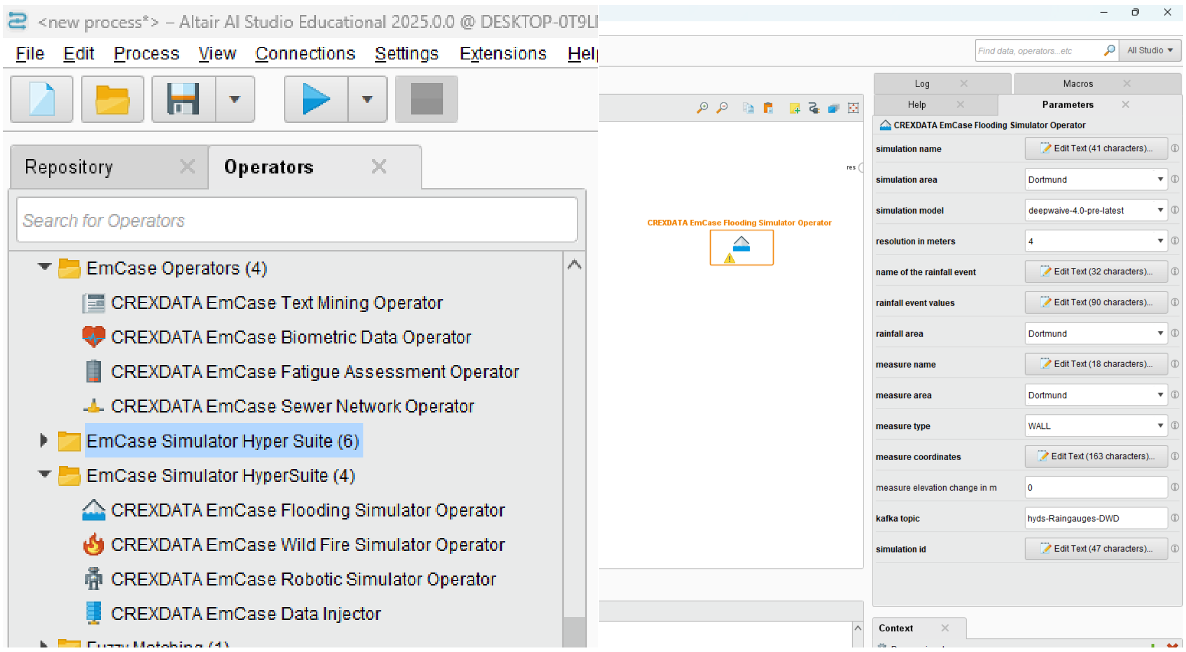


Figure 3: CREXDATA Custom Extensions

On the right-hand side of Figure 3, a custom operator ‘CREXDATA EmCase Flooding Simulator Operator’ from the ‘EmCase Simulator HyperSuite’ and its editable parameters are shown. After executing the workflow on the AI Hub, simulation results are automatically retrieved and visualized in ARGOS and AR. These custom operators are implemented for all EmCase-compatible technologies.

2.4 Simulation tools and Multi Model Data Fusion

Various simulation models are part of the EmCase Demonstrator System. Some models must be configured individually for each simulation to be performed. The configurable simulation models include the FloodWaive flood simulation, the Spark wildfire simulation, and the Gazebo robotic simulation. Detailed descriptions of the simulation models are provided in Sections 2.4.1, 2.4.2, and 2.4.3 below. There are basically three different ways to parameterize the simulation models: The user can enter the parameters directly in the simulation model and trigger a simulation (1). However, since the aim is to develop a generic data platform based on workflows, the simulation models can be parameterized in the workflow within the Simulator HyperSuite (SHS) in RapidMiner (2). This enables the user to set relevant parameters when creating the workflow and adapt the simulation model to the operating conditions. The SHS in RapidMiner is a directory within the extension that contains a collection of relevant simulator workflows and underlying processes. Another option for parameterizing the simulation models is to use the Simulator HyperSuite Graphical User Interface (SHS GUI) (3). Appendix 1 (Section 9.3) shows a screenshot of the SHS GUI web application. The user can define metadata for the simulation and select a simulation model. The SHS GUI simplifies the use of the connected simulation models by only requiring the relevant parameters to be defined and providing realistic parameter ranges. The use of such a GUI ensures that end users in the fire department can use it without having to possess domain-specific knowledge or extensive IT skills.

2.4.1 FloodWaive Flooding Simulation

FloodWaive is a German start-up company that uses an AI surrogate model to simulate water distribution in real time. FloodWaive provides a generalized, physics-informed deep learning model that acts as a surrogate model for classical hydrodynamic numerical 2D simulations (solving shallow water equations). The product is called DeepWaive. Unlike conventional, area-specific AI approaches, which require retraining for each new catchment area, FloodWaive uses a specialized architecture that allows spatio-temporal dependencies in hydrological systems to be abstracted. This makes the model applicable to unknown topographies and catchment areas (generalization) without the need for recalibration or retraining. The model processes complex geodata (digital terrain models, infrastructure data, land use, infiltration properties) as well as dynamic meteorological and hydrological input data (precipitation grids, inflow hydrographs). It simulates unsteady, areal 2D flooding processes and provides time-resolved raster data for water depths, flow velocities and flow vectors as output. A key feature of the model is the drastic reduction in computing time while maintaining a high degree of physical plausibility. DeepWaive is provided as a cloud service for integration into the CREXDATA architecture. Technical integration is achieved via a REST API, which enables complete parameterization of external applications.

Simulations are created based on historical, current, future or free configurable weather data. A detailed Digital Elevation Model (DEM) of the region to be simulated serves as the basis for the simulation. The simulation is started using the weather data based on the simulation area, the area of interest. The simulator requires precipitation values, precipitation durations, resolution of simulation, surrogate model used, and other configuration data. The model calculates water distribution using the previously specified parameters based on a stored

DEM. The result is the water distribution in the area of interest. The simulator specifies water level as a function of the existing DEM structure per coordinate. The water level per point is visualized using a pre-defined scale. The reduction in simulation computing time opens up new areas of application for simulation models for decision-makers. The simulation can be started in three different ways. To do this, the configuration data for the simulator can be entered from the ARGOS system, the SHS GUI or in the RapidMiner AI Studio Operator. The level of abstraction of the configuration data input varies depending on the type. The data entered can be taken from the Input-Process-Output (IPO) diagram (see Figure 4).

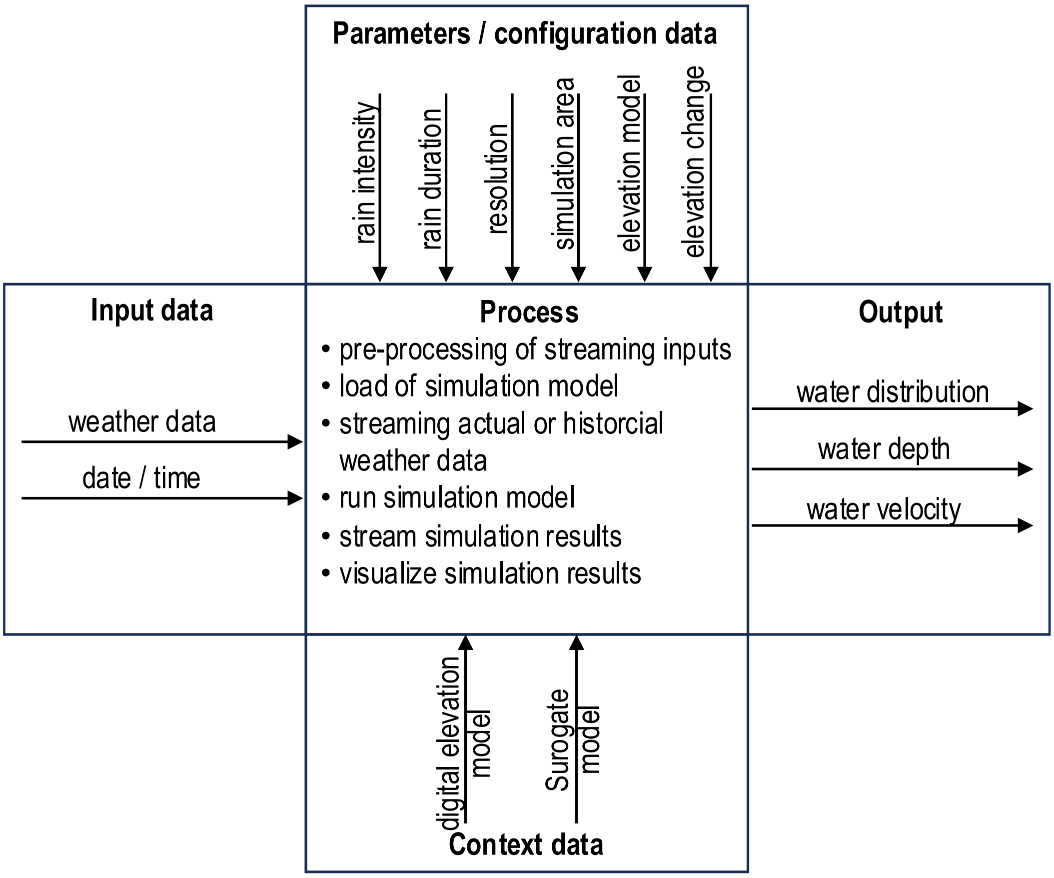


Figure 4: IPO diagram FloodWaive flooding simulation

The DeepWaive flood simulation model is integrated via REST endpoints using an API provided by FloodWaive. There are three ways to start the simulator. First, the SHS GUI can capture the necessary parameters and trigger the simulator. The second opportunity is via parameterization directly in RapidMiner AI Studio. The third way is that ARGOS collects all relevant information and starts the simulation. The latter is closely linked to the optimizer, as the process is mainly set up by the optimizer to optimize the barrier height. Basically, the status of simulation is reported back to the respective initiator. However, the simulation time is so short that the result is already complete before the user can view the status. The simulation results can be retrieved after completion using a simulation ID. The simulation ID is written to a Kafka topic, from which it can be read. Among other things, ARGOS System reads from this Kafka topic and visualizes the result in the correct target map. The target map can be Innsbruck or Dortmund. The simulation ID is used to retrieve and process the extreme large GeoTIFF files. The files are processed in ARGOS and the water distribution per point is visualized. The resolution can be 1x1, 2x2, 4x4, 8x8 or 16x16 meters. Data is only processed once a simulation is complete, which prevents incomplete data from being called up by ARGOS. Water distribution based on the simulation enables decision-makers to analyze various scenarios and make decisions based on different simulations. The operators created are illustrated in Figure 5. RapidMiner Operators enable all possible API calls with the simulator. The operators have been combined into intuitive operators with reduced granularity for improved usability by users. By reducing input options, user is intuitively guided through parametrization process without having to fill in unnecessary parameters. A significant advantage is the use of reduced content, as the end user is not shown irrelevant parameters. The screenshots in Figure 5 illustrate different levels of granularity within the CREXDATA Simulation Hyper Suite. On the left, all API-integrated operators are visible in RM AI Studio, allowing detailed parametrization of the DeepWaive model, including map visualizations, rainfall event creation, simulation setup, and result monitoring. On the right, a higher-level Custom Operator for end users encapsulates all background processes, exposing only four essential parameters—flood area, simulation name, rainfall values, and resolution enabling users to control simulation detail while simplifying operation and reducing computation time.

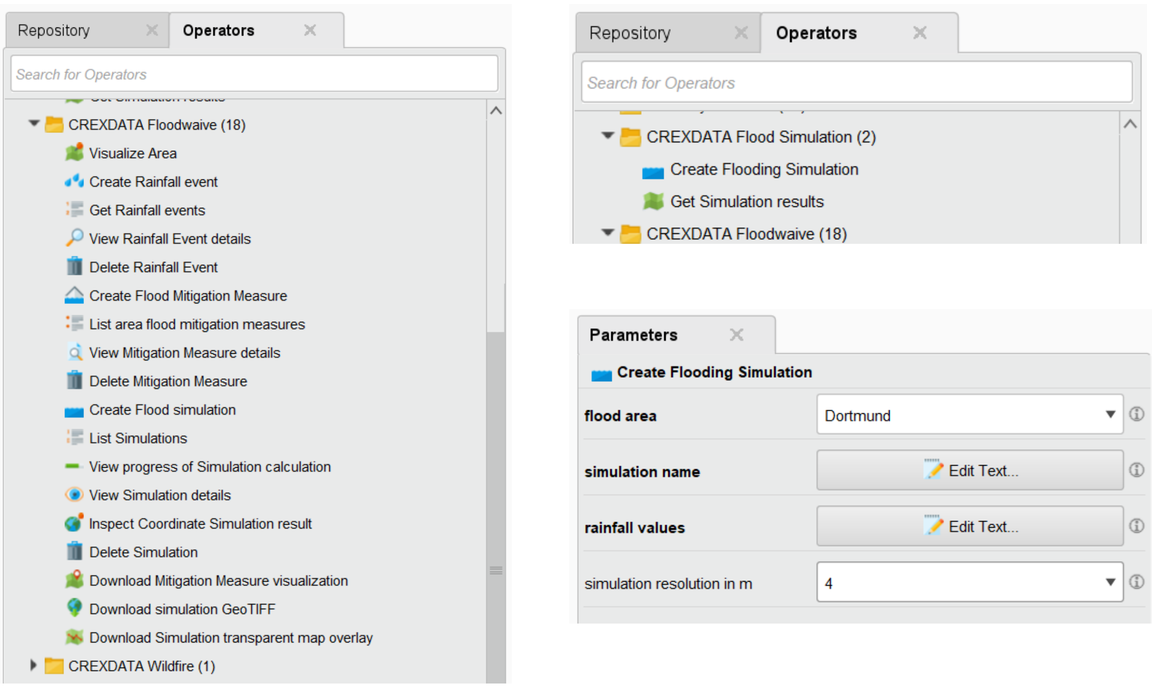


Figure 5: RapidMiner integration

2.4.2 Spark Wildfire Simulation

Spark[1] is a wildfire simulation model for researchers and experts in the disaster resilience field. Spark is a comprehensive toolkit for the complete processing, simulation, and analysis of wildfires. It enables the development of customized spread models based on a GPU-

[1] https://research.csiro.au/spark/

supported computational solver and offers numerous components tailored to wildfire modeling for input, processing, and visualization. The toolkit supports extensive processes with many integrated operations, allows the definition of different vegetation classes with their own spread functions, the import of compatible GDAL and NetCDF data, the performance of extensive ensemble analyses, and the visualization, output, and reproducible documentation of all results. Within the EmCase, Spark is utilized to simulate and calculate the spread of a fire in a wildfire scenario in order to provide decision support to emergency responders. The Figure 6 below illustrates the IPO diagram of the utilization of Spark in CREXDATA. The Spark wildfire simulation process begins with the acquisition of a variety of different input data formats, including text and CSV files, images, and geographic raster and vector data such as ESRI grids, ESRI shapefiles, NetCDF datasets, or GeoTIFF files. If necessary, these are converted to an internal format using GDAL. The initialization model then processes this data and uses it to generate the starting conditions required for the simulation, such as topographical information, fuel distributions, weather parameters, and the starting position of the fire. CREXDATA input data is data such as the wind direction, wind speed, temperature or relative humidity.

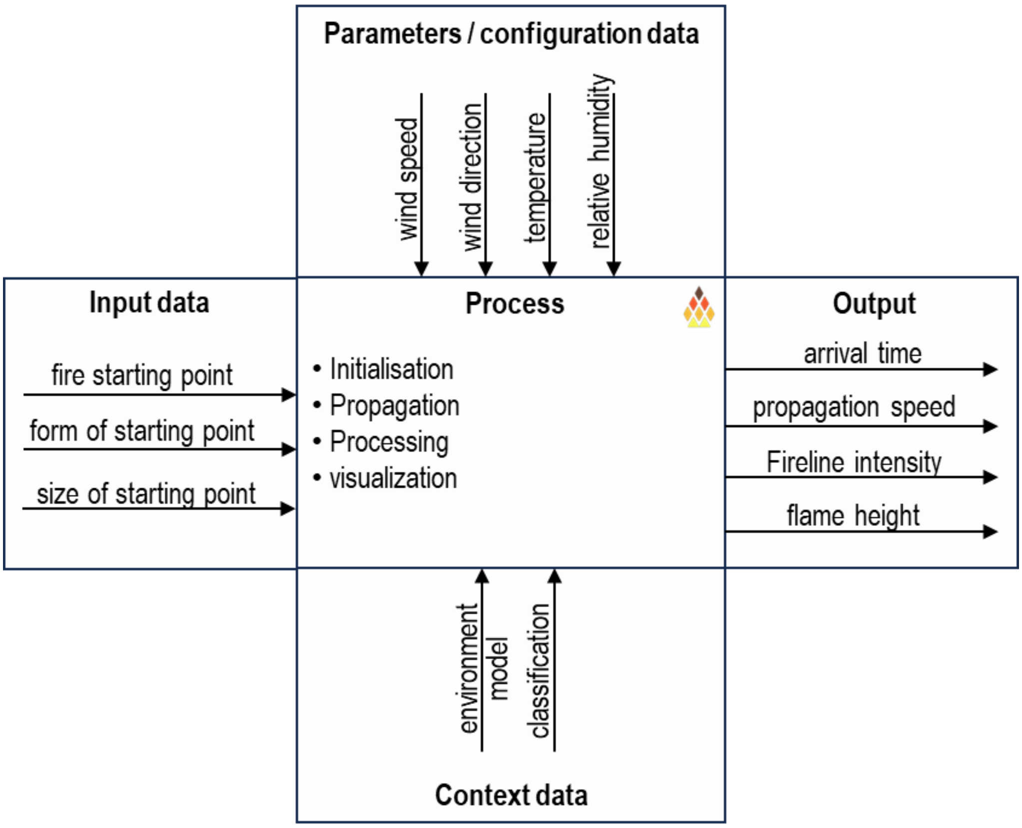


Figure 6: IPO diagram Spark wildfire simulation

On this basis, the propagation model performs the actual simulation of fire spread. It calculates time-dependent changes, such as how fast and in which direction the fire is moving based on wind, terrain, and fuel. The results of this propagation calculation are transferred to the analysis system, which performs statistical evaluations, comparative analyses, or further processing steps. Both intermediate and final results can be displayed visually, for example in a 2D or 3D scene or via web-based map systems such as OpenLayers. At the end of the process, the simulation and analysis results generated are provided in various output formats. These include text/CSV files, images, ESRI-based raster and vector data, and NetCDF or GeoTIFF files. These structured outputs allow the simulation results to be reused in other GIS systems, analysis tools, or visualization platforms like ARGOS. During the course of the project, the provider Csiro developed an API for the Spark wildfire simulation model that allows the simulator to be integrated and connected to the CREXDATA EmCase demonstrator system. However, as the API was not released until the final evaluation phase, it was not possible to complete the integration. Nevertheless, all the necessary steps have been completed and the API can be connected in a similar way to the FloodWaive integration. The workflow is as follows: The user in ARGOS aims to simulate a reported wildfire. To do this, the EmCase Simulator HyperSuite GUI is opened. The user sets the simulation area or adopts the area from ARGOS. The other input data is also set. Once the parameters have been completed, they are written to a configuration file by the workflow. The configuration file is used via the integrated API to trigger the simulation. Once the simulation has been carried out, the output files are returned and used for visualization in ARGOS or AR.

2.4.3 Gazebo Robotic Simulation

Gazebo[2] is a robotic simulation environment providing open-source development libraries. It is used to simulate robotic behavior and the interaction between robots and their environments [1]. The simulator allows individual robots and environments to be added. Simulated robots are controlled just like real robots based on the open-source Robotic Operating System (ROS) [2]. As a result, code for drone control can be applied to both simulated and real drones. In emergency response, this can be used to simulate routes for unmanned aerial vehicles (UAVs) before the code is sent to real drones [3]. This is applied in the EmCase. Robotic simulation is used to simulate and compare different route alternatives for drone flights. Compared to classic routing problems, dynamic influences on the drone, such as wind direction and wind speed, are considered. Wind influences have an impact on energy consumption and thus on the drone's operating time. The simulation is therefore carried out based on current or forecast wind speeds and directions. This supports two decision points during operations. On the one hand, it gives incident commanders an impression of how successful the drone will be and how many drones need to be requested to the operational site to explore the relevant area. On the other hand, drone pilots use this information to determine the most optimal route alternative according to their optimization criteria, such as the shortest flight time, the shortest distance, or the highest efficiency. Configuration and input data can be entered using the SHS GUI or by directly changing settings in the custom operators. The IPO diagram for the Gazebo workflow is visualized in Figure 7. In the next step, the n shortest routes are calculated based on the number of routes n to be simulated, which must be entered in the SHS GUI. This step is performed to limit the simulation duration, as the number of simulations to be performed increases significantly with the increase in waypoints. The calculated routes and other configuration and input data are then transmitted to the virtual machine via a REST endpoint call. Subsequently, the bash script that initializes the simulation is called via another REST endpoint. For each route to be simulated, the bash script calls a Python script that coordinates drone control along the

[2] https://gazebosim.org/home

respective route and the storage of drone status information. The simulation uses drone models from the open-source PX4-Autopilot[3] repository. Status information stored during the flights is post-processed after each flight through the post-processing workflow. First, the data is retrieved, and KPIs such as the flight duration and energy consumption are calculated.

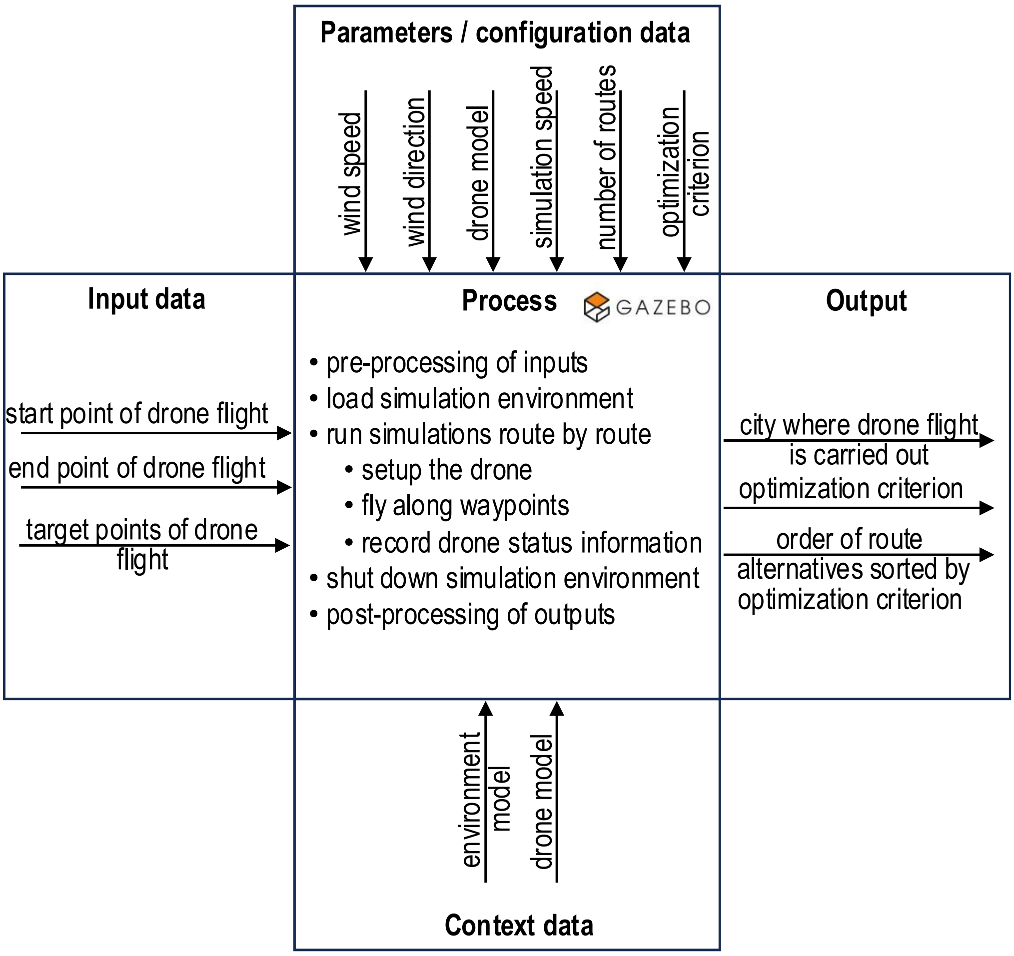


Figure 7: Gazebo workflow IPO diagram

These KPIs are used to sort route alternatives according to the given optimization criterion, e.g., the flight duration. Following this example, all routes are sorted according to flight duration, with the fastest route and the corresponding sequence of waypoints at the beginning. This sequence of routes is published into the Kafka topic. The post-processing workflow is running in parallel while further simulation runs are carried out. Thus, every time a new simulation run is completed, outputs are post-processed and the sequence of routes is updated.

[3] https://github.com/PX4/PX4-Autopilot

2.4.4 Risk index by Gradient Boosting machine learning technique

FMI provides meteorological expertise for weather-related hazards and develops a Gradient Boosting machine-learning model (see Figure 8) that links real-time weather forecasts with observed impacts in Finnish municipalities. The model uses various inputs, including hourly impact data (alert tasks, fire-fighting operations, traffic and vehicle accidents) as well as high-resolution weather prediction data from ECMWF HRES and the MEPS ensemble system, complemented by newly derived meteorological parameters that better capture events such as thunderstorms. These inputs reflect Finland's diverse seasonal hazards, from winter storms and poor road conditions to summer heatwaves, drought, wildfires and severe thunderstorms. Based on these data, the outputs of the system are municipality-level impact and risk predictions that support emergency services in anticipating storm damage or other weather-related challenges. Model performance is evaluated through extensive cross-validation, including historical datasets, and continuously monitored particularly for extreme events to ensure reliable and accurate predictions.

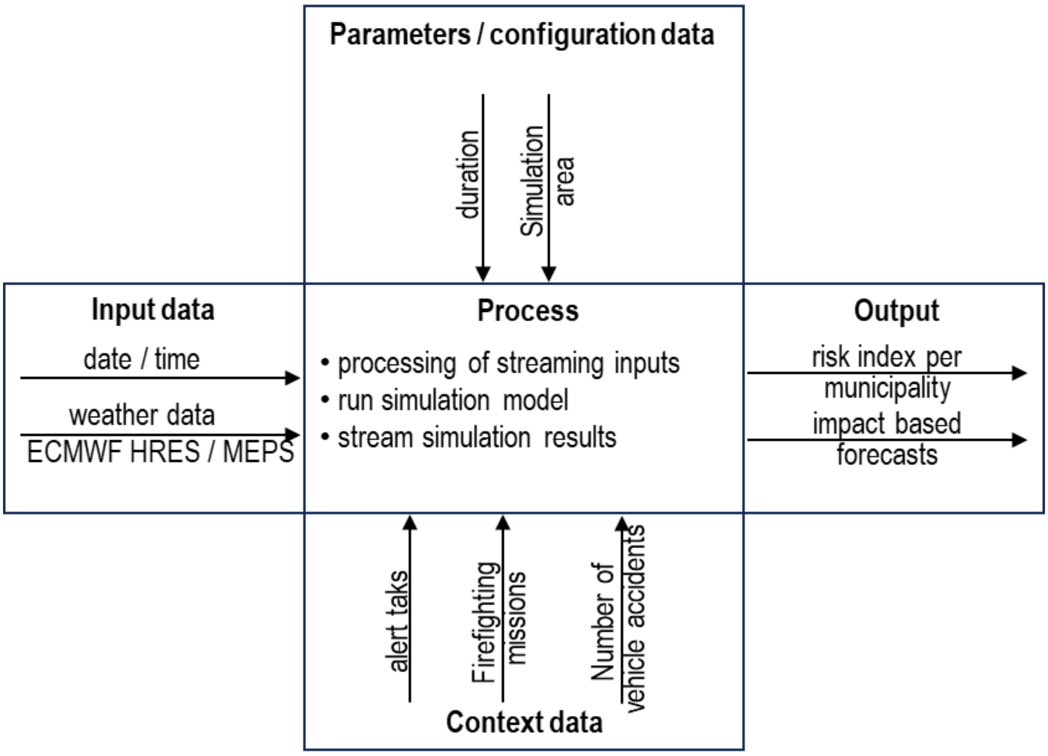


Figure 8: IPO diagram of tool for emergency tasks by FMI

2.4.5 EmCase Data Injector

Workflows developed in the Emergency Case are designed to process data from the field, e.g., sensor streams from sewer network sensors, biometric devices or drone data, or data from other services, such as social media posts and weather forecasts. The Data Injector allows entering the data into the workflows even without real data sources. For example, system tests, demonstrations, and user training are usually conducted outside real emergency situations. Even in an incident there is the explicit use case to run "what-if" scenarios in a way that past situations are considered to happen similarly again. What if the forecasted rain event would evolve like a similar event in 2016 did? Thus, the Data Injector enables the execution of the workflows based on stored data. To do this, the data is first retrieved from the corresponding source. There are then two modes of data injection, as depicted in Figure 9. In the first mode, the data is injected at the same frequency and with the same timestamps as it was originally recorded. The second mode allows adjusting the frequency at which the data is injected and updating the timestamp of datapoints. The timestamp is increased based on a defined start time in accordance with the set injection frequency.

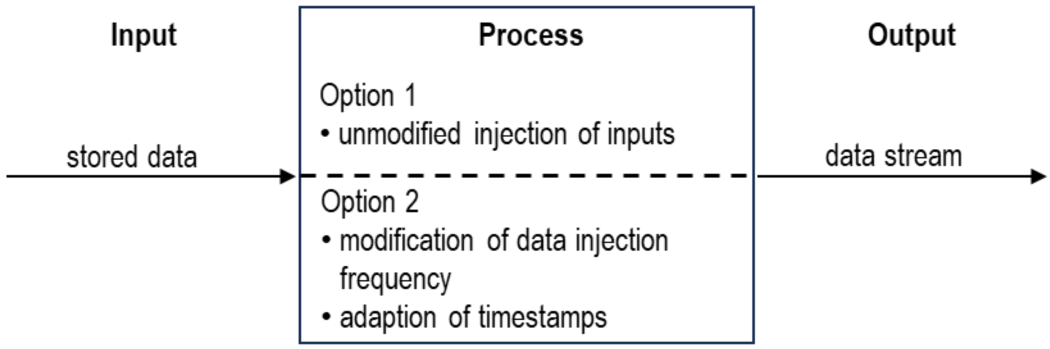


Figure 9: IPO diagram of the Data Injector

Within the Data Injector operator, the step size in which the timestamp is increased, and the according frequency as well as the start timestamp are selected. Based on this, a new timestamp is created for every retrieved data point. The attribute containing the old timestamp is deleted and the updated data points are written into a Kafka topic one by one according to the initially chosen frequency. Datasets can be stored by any data processor and simulator. In the setup of the demonstrator system, it is done for Text Mining and XAI, i.e., social media posts and biometric data. Stored biometric datasets include recordings from firefighting operations in Dortmund as well as synthetically created datasets, which are based on ranges occurring in measurements from Dortmund to provide realistic data. The social media datasets include posts and meta data. The posts contain content typical of social media, such as natural language statements in German, tags, hashtags, special characters, and smileys. A large proportion of the posts contain statements that are irrelevant to an incident and therefore situation independent. In addition, there are some scenario-specific posts. These describe observations of wildfires or flooding events as they might be published by citizens in a crisis.

2.4.6 Multi-Model Data Fusion

Timely and visual representation of information is critical on emergency case scenarios. In the Multi-Model Data fusion (MMDF) demonstrator, we utilize a small client server application to facilitate end-users' interpretation of the real time fusion results, we developed a small client-server application. The technical architecture is described in D3.2. The server application directly consumes data from Kafka topics and emits records via web sockets to the client application (leaflet-map). The server application implements a Kafka consumer on node.js, that is registered to the 'mmdf-single-responder-incident-feed', 'mmdf-responder-team-status' topics. The first topic emits records of social media posts related to a flood event and fused to the responder in proximity that reported last its biometric data. The second topic dynamically organizes responders into teams based on their location, focusing on individuals in proximity rather that organizational units of responders. Each part of the workflow serves a different need. The first can assign incidents in real time to the nearby responder, and the second aims to monitor collective team load of responders on a larger incident such as a flood incident. Figure 10 illustrates a synthetic scenario of a flood event in Dortmund. The client-server application leverages MMDF EmCase workflow results to visualize in real time links (blue dashed lines) of first responders (markers) with recent flood incidents (blue dots), and to organize first-responders into teams based on their proximity.

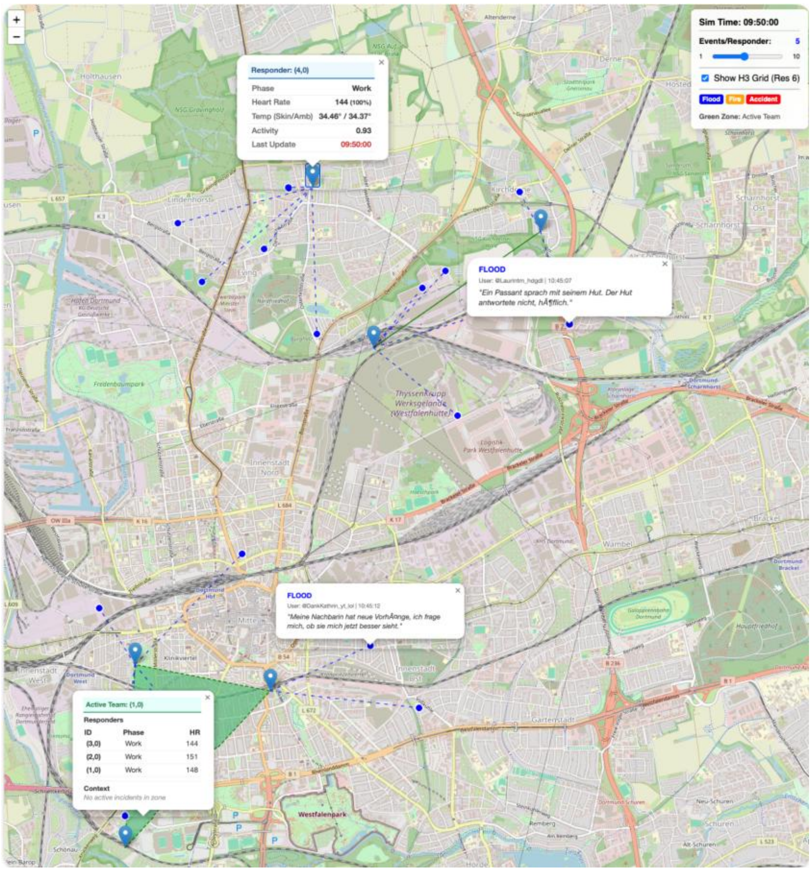


Figure 10: Synthetic scenario of a flood event in Dortmund.

2.5 Final Use Case Evaluation

The CREXDATA EmCase Trials follow a mixed-method research design, integrating quantitative, qualitative, combined approaches to evaluate both technological performance and human–system interaction. This design allows for a holistic assessment of how predictive analytics, simulation tools, AR interfaces, dynamic data streams influence decision-making, situational awareness and responder resilience under high-pressure urban flooding scenarios. The evaluation is structured around scenario phases and aligns with CREXDATA main objectives, ensuring that participants' engagement with the integrated demonstrator environment can be systematically monitored and analyzed.

A forest fire scenario and a flood scenario were carried out at the two pilot sites in Dortmund and Innsbruck (cf. section 2.3.1). The field trial consists of testing activities in Dortmund and Innsbruck. In Innsbruck, a forest fire scenario was carried out during a large-scale exercise, and a flood scenario was evaluated in a workshop that included graphical workflow modeling. In Dortmund, a three-day workshop series included a workshop on graphical workflow modeling using a forest fire scenario from nearby Lake Möhnesee and a large-scale flood scenario based on the heavy rainfall event at UEFA Euro 2024.

2.5.1 Research Design

The mixed-method approach combines quantitative and qualitative evaluation to assess both technological performance and human–system interaction. The quantitative component applies objective metrics such as Likert-scale questionnaires to measure usability, trust, and decision quality, System Usability Scores (SUS) to evaluate the user-friendliness and efficiency of interactive tools, and prediction accuracy measures to assess the reliability of flood simulations, incident forecasts, and decision-support outputs. Complementing these metrics, the qualitative evaluation captures participant experiences and decision-making strategies under realistic, high-pressure conditions. Behavioral observations and cognitive walkthroughs are used to document how participants navigate the system and to identify usability challenges, while think-aloud protocols reveal reasoning processes and difficulties in interpreting visualizations and predictive analytics. Post-trial surveys further reflect perceptions of usefulness, trust in system recommendations, and operational relevance. Integrating these methods ensures a holistic assessment that links objective performance data with user perceptions and observable behaviors. Whereas quantitative results reveal accuracy, speed, and usability, qualitative insights illuminate cognitive processes, trust formation, and contextual demands. This mixed-method evaluation helps determine whether technological outputs genuinely influence decision-making and operational outcomes. Overall, the research design enables the evaluation of system effectiveness and reliability in realistic urban flooding and wildfire scenarios, the assessment of human–system interaction factors such as situational awareness and trust, and insights into the operational applicability of CREXDATA tools in dynamic emergency contexts. It also supports the identification of system improvement needs and future opportunities for integration into civil protection strategies. By combining scenario-based trials with quantitative measurement and qualitative inquiry, the design ensures a comprehensive understanding of both technological capabilities and their real-world impact.

2.5.2 End-User Lab Tests

In preparation for the final evaluation activities, successive lab tests with end-users were carried out to test individual components and develop them as close to user requirements as possible. Parts of the EmCase demonstrator system were demonstrated to users and tested live. The feedback was incorporated into the further development of the EmCase demonstrator system in order to prepare for the final evaluations in the best possible way. The lab tests performed are briefly described below and results are presented in the Appendix 1 (Sections 9.4 - 9.6).

2.5.2.1 Text Mining & Gazebo online lab test

The scenario for the text mining and Gazebo online lab tests is based on the heavy rainfall event in the Amras district of Innsbruck in July 2016. The scenario is described in detail in

D2.2. In a small test, the components text mining, Gazebo robotic simulation, and ARGOS visualization were tested with experts from the Dortmund fire department and feedback was collected. First, ARGOS was explained and the participants had the opportunity to try out the visualization of the various data layers in ARGOS using their own access data. During the monitoring of the rainfall, the situation report was to be expanded with information from the population, which is why the relevance classification of text mining was used. After visualizing the relevant social media posts, they were further analyzed using the Q&A model. With the information gained from this, the decision was made to verify important information from social media with a drone flight, so the Gazebo robotics simulation was used to plan the drone route in order to identify the most efficient route along the specified waypoints. Following the visualization of the results, tactical operational planning was discussed in order to make the online lab test even more realistic and closer to actual operations. The results are summarized and listed in the Appendix 1 (Section 9.4).

2.5.2.2 Emergency Service Academy Finland R&D Days

The Research & Development Days of the Emergency Service Academy Finland (ESAF) in Kuopio provided another opportunity to evaluate subsystems of the EmCase Demonstrator System. Various subsystems were demonstrated in a trade fair-like setting in collaboration with the TEMA project (see cohesion activities in Deliverable D6.2). ARGOS, the tool for emergency tasks developed by FMI, eXplainable AI, and the augmented reality applications were tested from the EmCase Demonstrator System. In a two-hour session, the approximately 80 participants of the ESAF R&D Days had the opportunity to experience functionalities such as ARGOS or Augmented Reality by themselves. To increase the practical relevance, a realistic scenario was also developed for this test, which provided the emergency services with the necessary context for applying the technologies in the field. Afterwards, some of them gave their feedback in a questionnaire. The results are summarized and listed in the Appendix 1 (Section 9.5).

2.5.2.3 Augmented Reality Workshop

The AR application includes the visualization of robotic data. In a specially constructed workshop at the facilities of the DRZ, emergency personnel and drone experts tested the visualization possibilities of robotic data from the perspective of a drone operator and a team member or first responder in AR. The workshop began with an example-supported and guided discussion on the visualization of uncertainties in ARGOS and AR. This was followed by a theoretical introduction to the use of augmented reality devices. Participants were then each given an AR device to test the application from the two different roles. Each participant was given individual instruction and had the opportunity to experience both the drone operator's and the first responder's perspectives. The goal was for each participant to learn about the functionalities of AR robotics data visualization and to work together as a team. For example, the drone operator marked specific danger points that were identified during the virtual drone flight and sent his team members there to assess the situation. Afterwards, the AR application was evaluated by experts using a questionnaire. The results are summarized and listed in the Appendix 1 (Section 9.6).

2.5.3 Urban Flooding Scenario

2.5.3.1 Dortmund

CREXDATA in the command room environment

The first day urban flooding scenario combined real precipitation data from past Dortmund flood events (see Figure 11) with a fictional escalation set during the Germany vs. Denmark EURO 2024 match, creating a high-pressure environment involving large crowds, transport disruptions and severe flooding. Although only thunderstorms affected the actual matchday, the trial amplified the situation using historical 2008 flood data and synthetic inputs to simulate critical urban impacts on the stadium, public transport and emergency services. The scenario unfolded different phases, each introducing new data, challenges and decisions, providing a realistic yet structured setting to evaluate CREXDATA's predictive analytics, simulation tools and AI-supported decision-making under uncertainty. All phases were supported by realistic situational reports from the real operation in 2024.

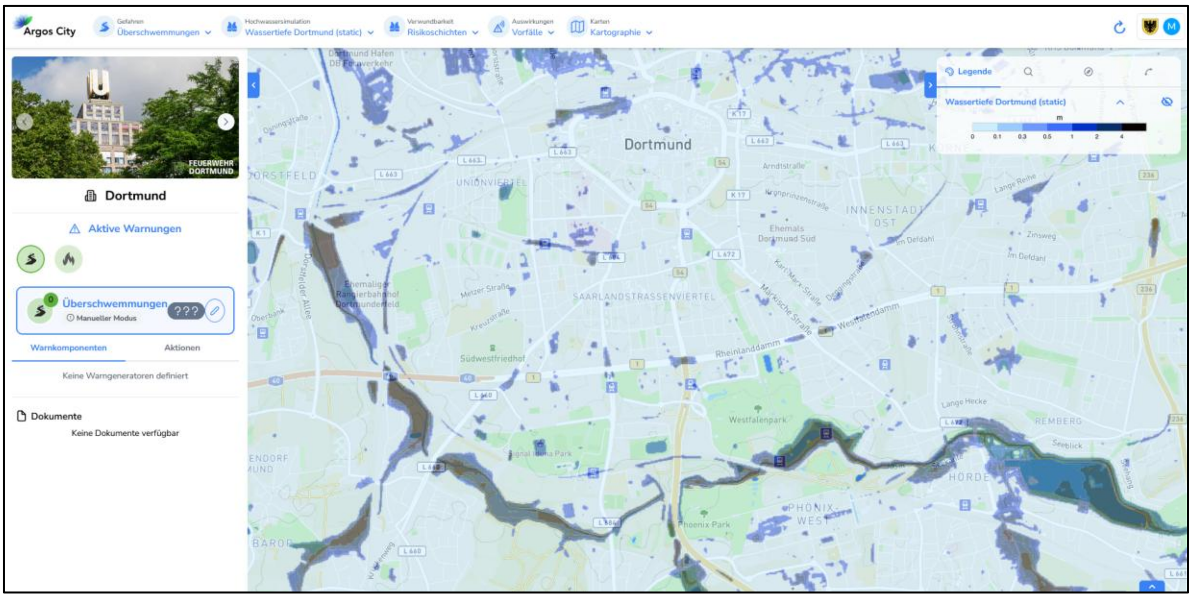


Figure 11: simulated water distribution within the Dortmund scenario

The Dortmund trial involved carefully selected emergency service professionals with substantial staff experience to ensure realistic evaluation of CREXDATA technologies. Key participants included representatives from staff functions S2 (Situation), S3 (Operations), and S5 (Media/Press), along with a Head of Staff. S2 personnel focused on situational analysis and digital information management, assessing tools like FloodWaive, ARGOS and visual analytics systems, especially regarding predictive simulations and uncertainty displays. S3 representatives evaluated how simulation outputs and predictive analytics support field operations, resource allocation and decision-making under pressure. S5 participants assessed the usefulness of CREXDATA tools for public communication, including how social media monitoring can inform timely messaging. The Head of Staff had an overview for coordination across all functions, testing how well the technologies support leadership decisions.

Before the trial, all stakeholders joined a workshop on graphical workflow modelling to familiarize themselves with the CREXDATA Workflow Design environment. Their operational insights and feedback ensured a realistic, practice-oriented assessment of usability, added value and overall relevance.

The trial design followed a structured phase model, mapping directly onto real emergency response workflows. Each phase was designed to test the EmCase demonstrator system and its components. The general narrative followed a phased progression that began with relatively minor incidents and escalates into a large-scale urban flood crisis. Each phase was associated with a time window and introduces additional stressors such as growing numbers of incidents, transport disruptions and the need to make fast and well-informed decisions under uncertainty. The trial thereby enabled a stepwise testing of technologies, workflows and decision-making support under increasingly complex and resource-intensive conditions.

By combining a realistic public event, historical flood data, phased escalation and integrated CREXDATA technologies, the Dortmund trial provided a robust platform for testing predictive analytics, AI-assisted decision-making, human-machine interaction under complex and high-pressure conditions.

CREXDATA on-site

The second day of the Dortmund trial took place at the German Rescue Robotics Center, using a realistic urban flood simulation with rising water, blocked streets, and a building where civilians are assumed trapped. Responders needed to conduct a full rescue mission from assessing the situation to executing an evacuation under worsening conditions. CREXDATA tools were embedded into the exercise: AR devices displayed flood and uncertainty data, UAVs survey inaccessible zones, and predictive fatigue monitoring supports command decisions. The scenario was designed to escalate in complexity, providing a safe yet realistic environment to test how well CREXDATA improves decision-making, situational awareness, and responder safety during urban flood operations.

On the second day of the CREXDATA Field trial, operational stakeholders from the Dortmund Fire Service tested the technologies in a realistic, practice-oriented setting. Firefighters with attack-team experience acted as primary AR users. They navigated flooded areas, assessed points of interest, evaluated system-suggested routes, including those marked with uncertainty and carried out evacuations. Their feedback focused on usability, robustness, and operational relevance. The incident commander oversaw tactical coordination, validated route decisions, used AR and simulation data for situational assessment, and monitored responder fatigue predictions to ensure operational safety. The communication unit and drone operators provided aerial reconnaissance. They deployed UAVs, transmitted live data to AR devices, marked hazards and POIs, and assessed how well UAV data integrates with AR visualizations. Their role was key to evaluating the added value and interoperability of robotic support in inaccessible environments.

2.5.3.2 Innsbruck

The field trial of the CREXDATA system, focusing on heavy rainfall and flooding, was conducted in a workshop format with representatives of public authorities responsible for decision-making during weather-induced emergencies. Participating organizations included the Tyrol Warning Centre, the Centre for Crisis and Disaster Management Tyrol, the Department of Water Management, the Subject Area of Hydrography and Hydrology, and the Municipal Department of Waste Water Management. Depending on the specific test case, different stakeholders were involved in the evaluation. Representatives from the aforementioned public authority departments took part, particularly those responsible for crisis management in heavy rain and flooding scenarios. The geographical setting of the field trial varied across the two days due to the different testing arrangements. The first day, which consisted of the workshop with public authorities, was held at the University of Innsbruck (Innrain 52a, 6020 Innsbruck, Austria). The corresponding scenario was based on a real heavy rain event that occurred on 2 July 2016 in Innsbruck, with particular attention given to the Amras quarter, which was significantly affected during that event. Innsbruck served as an ideal pilot area due to its combination of alpine topography, urban density, and critical infrastructure. Situated in a narrow valley surrounded by steep slopes and crossed by several small rivers, the city is particularly prone to flash floods and debris flows during intense rainfall. Such events can endanger key infrastructure, including transport routes, the hospital district, and historic city areas. The city's urban vulnerability is comparatively high, not only because of its critical infrastructure—such as the hospital complex, university campus, train station, and highway crossings—but also due to its narrow streets, historic buildings in the old town, and a large transient population consisting of students and tourists. This combination of physical and social factors significantly increases the complexity of evacuation planning and information management during flood-related emergencies.

2.5.4 Wild Fire Scenario

2.5.4.1 Innsbruck

The wildfire scenario of the CREXDATA project was tested in a real-life exercise (see Figure 12) conducted together with multiple fire brigades from Innsbruck and the surrounding municipalities of the Innsbruck-Land district.

Figure 12: Field trial as part of an exercise in Stubaital close to Innsbruck photos: DCNA

Participating units included the air services of Innsbruck-Land, the Fire Brigade of the City of Innsbruck, several volunteer fire departments from the Stubaital section, and the Mountain

Rescue Service Vorderes Stubai. In total, more than 140 participants were involved in the district's helicopter exercise. Professional and volunteer firefighters with experience as incident commanders were specifically targeted to take part in evaluating the CREXDATA system's applicability to wildfire management. The geographical setting for testing the system components related to wildfire management was implemented on the second day of the Innsbruck trial, leveraging a large-scale helicopter wildfire exercise organized by local fire brigades and mountain rescue services. The exercise was based in the mountainous region near the village of Telfes, where the mobile incident command post and air service command center were located at Schlickerboden/Speicherteich-Panoramasee. A nearby slope below the Krinnenköpfe peak served as the simulated fire area, where smoke was released to replicate realistic wildfire conditions (see Figure 13). The exercise involved several coordinated components, including helicopter-based water drops, ground firefighting operations, radio communication and coordination using standardized phraseology, and aerial surveillance by drones to detect hotspots and remaining embers.

This integrated exercise provided an authentic and challenging environment for evaluating the CREXDATA wildfire management tools under real emergency conditions, ensuring that system functionalities such as situational awareness, data fusion, and decision support could be thoroughly tested in cooperation with operational stakeholders.

Figure 13: Geographical setting of alpine wildfire exercise, source: Google Earth

2.5.4.2 Dortmund

The wildfire scenario followed a structured sequence that combines meteorological data, simulation tools, and situational awareness technologies to support operational decision-making. The exercise began with the ARGOS overview map, which displayed weather forecasts and visualized parameters such as temperature, wind direction, and wind speed, indicating dry conditions on the day of the operation. Based on these inputs, a wildfire simulation model was initiated to calculate the spread of fire while accounting for prevailing weather conditions. The model could be started through a simplified input interface, allowing the command unit to explore potential fire propagation dynamics. Subsequently, a drone simulation was used to identify relevant locations for reconnaissance flights and to determine the most efficient route for data collection. Drones were deployed virtually to gather situational information and feed it back into the system. In the final phase, results from the simulations were visualized again on the ARGOS overview map, supporting the incident command in deriving tactical response measures and enhancing overall coordination during wildfire management operations. Figure 14 illustrates the area affected by the wildfire as well as the simulated fire spread as result of the Spark wildfire simulation.

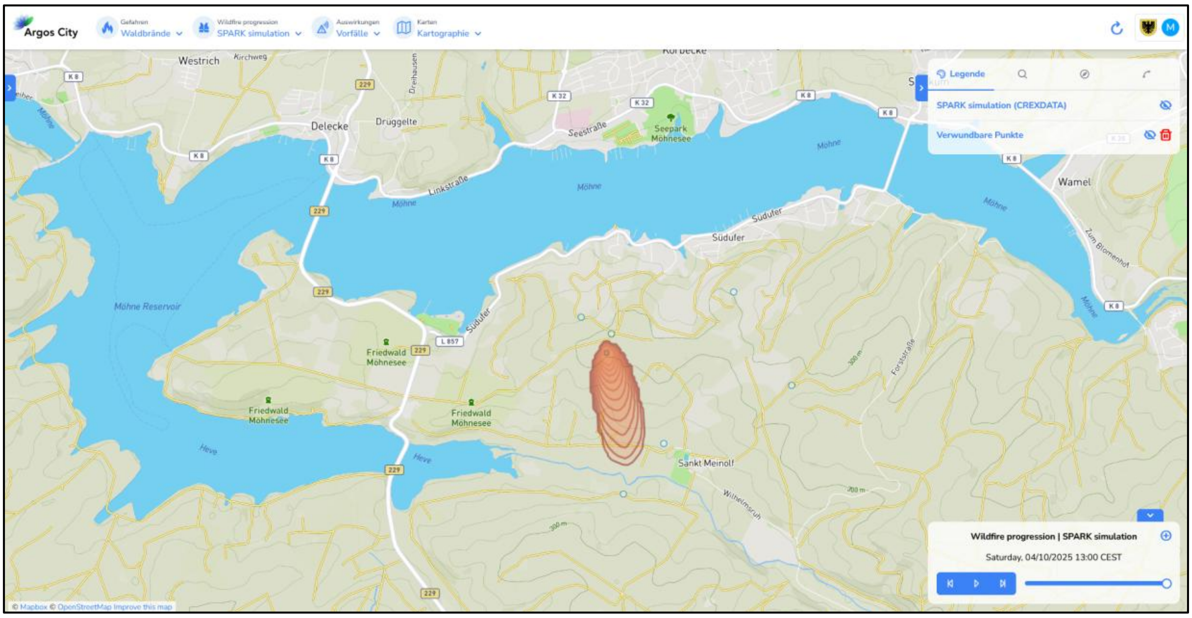


Figure 14: simulated fire spread in Möhnesee wildfire visualized in ARGOS

2.5.5 KPIs and Objectives

The trials alignment with CREXDATA objectives is further reinforced through the Key Performance Indicators (in accordance with the Description of Action):

KPI (a): 80% accuracy in critical event prediction.

The aim that participating end-users report that the system advice matches 80% of the real situation in the context of real-world crises, such as severe urban flooding, wildfire or comparable high-pressure situations. This KPI directly addresses the system's perceived usefulness and trustworthiness among professional stakeholders, including staff officers (S2, S3, S5, head of staff) and decision-makers with operational responsibilities.

KPI (b): System Usability Score (SUS) of interactive exploration tools for XAI

Staff members used interactive visual analytics tools during the trial to explore predictions, compare model outputs and interpret uncertainties. Their feedback, collected through standardized SUS questionnaires, assesses usability and ensures that the tools are not only technically robust but also operationally intuitive.

KPI (c): SUS score of uncertainty visualization in AR and control center.

The scenario's staged escalation, beginning with light rainfall and culminating in severe flooding, provides multiple opportunities to test whether users understand the uncertainty visualizations presented by the system. Both staff officers in the control room and field users with AR devices evaluate whether these representations are clear, trustworthy and actionable.

KPI (d): Observable impact of system's output on action planning.

Observations and short interviews document whether and how the system's outputs influenced these decisions, thereby validating the tangible impact of CREXDATA technologies on operational planning.

KPI (e): Users following system advice under risk in 80% of situations.

As the scenarios involve both perceived risks and real uncertainty, it allows for systematic observation of user behavior. The trial captured whether decision-makers trust and act upon the system's recommendations even under high-risk conditions and under what circumstances alternative strategies are chosen.

Since the KPIs are described in very abstract terms, various statements were formulated for the individual CREXDATA system components, which, when answered, allow conclusions to be drawn about the fulfillment of the underlying KPIs. Questionnaires were developed for the individual technologies and simulators. In addition, the System Usability Score was used to evaluate the CREXDATA EmCase demonstrator system. The various questionnaires are included in the Appendix 1 (Section 11) to this deliverable.

2.5.6 Evaluation of individual technologies

The descriptions of the results are standardized for each technology and include the classification of the average values per statement as well as an overall classification of the evaluation.

EmCase demonstrator platform ARGOS

The evaluation results show that participants generally experienced the ARGOS system as intuitive and user-friendly, while also identifying some areas for improvement in transparency and the handling of uncertainty. With an average score of 7.14, users largely agreed that the operation of the system is intuitive, indicating that the interface and workflow could be navigated without major difficulties. The visualization of uncertainties received a lower score of 6.00, suggesting that – while users recognized its potential value – the current representation does not always support decision-making as effectively as intended. The clarity of the system's predictions and their explanations were rated positively with an average of 7.30, reflecting that most participants considered the provided information to be understandable and well-presented. However, the ability to comprehend why the system displayed certain results or recommendations scored 6.16, implying users sometimes lack full insight into the system's reasoning processes.

EmCase simulation models

The evaluated statements about the **EmCase SHS GUI** indicate generally positive feedback. Users agree that the web application (Simulator HyperSuite) facilitates the parameterization of simulators even without deep technical knowledge (average rating around 6.2). The integration of simulators into workflows is seen as particularly valuable, as it encourages users to make greater use of the application by automating the execution of different simulation models and parameter settings (rating about 7.5). Furthermore, the graphical interface is perceived as a clear advantage, making it easier to apply various simulation models in practical contexts (rating about 7.6). Overall, the results suggest that the EmCase SHS GUI is seen as user-friendly and effective in supporting simulation processes.

The evaluated statements for the **FloodWaive** urban flooding simulation show an overall very positive perception of the system's performance and usefulness. The plausibility of the system's forecasts received a moderately good rating (around 6.3), suggesting that users generally found the simulated developments convincing, though with some room for improvement. The correspondence between displayed endangered locations and the actual situation was rated higher (about 7.7), indicating strong trust in the accuracy of spatial predictions. The short calculation time achieved the highest rating (around 9.1), emphasizing that rapid simulations are a major strength, allowing the results to be effectively used in real-time operations. The ability to dynamically adjust the simulation model (8.2) and the clarity of the visualization (8.0) were also rated very positively, reflecting that users appreciate both flexibility and comprehensibility. Overall, the results highlight FloodWaive's effectiveness as a reliable, fast, and operationally valuable tool for flood management.

The evaluated statements for the **Spark** wildfire simulation reveal a strong overall confidence in the system's reliability and usefulness. Users rated the plausibility of the situation forecasts highly (around 7.9), suggesting that the predicted developments are largely seen as realistic and credible. Similarly, the correspondence between identified endangered areas and their actual locations was rated equally high (about 7.93), indicating good spatial accuracy of the model. The visualization received the best evaluation (around 8.3), showing that users particularly value the clarity and comprehensibility of the visual output, which effectively supports quick understanding and informed decision-making in the field. Overall, these results suggest that Spark performs well in providing accurate, transparent, and operationally helpful wildfire simulations.

The evaluated statements for the **Gazebo** robotic simulation show overall positive feedback regarding trust and usefulness. Users expressed moderate confidence in following the system's flight planning recommendations under challenging conditions such as strong winds or uncertainties (rating around 5.9), indicating some reservations in fully relying on the system in critical situations. However, they rated the general inclusion of the simulation's recommendations in flight planning much higher (about 7.5), reflecting that the system's input is considered valuable under normal circumstances. Similarly, the route selection recommendations were seen as sensible and suitable for integration into decision-making processes (around 7.4). Overall, the results suggest that users perceive the Gazebo robotic simulation as a helpful tool that provides meaningful support for planning, though further trust-building may be needed to ensure adoption in more uncertain or high-risk scenarios.

WP4 and WP 5 technologies

The evaluation of **Complex Event Forecasting** indicates an overall positive perception of the system's performance. The accuracy of both short-term and long-term forecasts for identifying critical points in the sewer network was rated moderately high (6.3), suggesting that users generally trust the model's predictions for decision-making but still see potential for improvement. The slightly higher score (6.73) for the system's usefulness in supporting emergency planning highlights that the forecasts are particularly valuable in operational contexts where timely action is needed. Finally, the assessment of practical relevance (6.33) shows that the forecasts are largely aligned with users' needs and help them make informed decisions, though there remains room to further enhance reliability and applicability in practice.

The evaluation of **Interactive Learning for parameter space exploration** shows a strong positive response, with an average rating of 7.7 out of 10. This indicates that users find the automatic parameter optimization—such as determining the optimal height of a mobile barrier—highly beneficial for operational efficiency. The system is perceived as an effective tool for utilizing available resources and personnel in a targeted and efficient manner during operations. The relatively high score suggests that users appreciate the practical value and support this functionality provides in real-time decision-making.

The evaluation of **Federated Machine Learning** with NeRFs reveals a very positive overall perception of the technology's usefulness. Respondents rated the impact of faster computation with preserved quality the highest (9.14), emphasizing that efficiency and performance significantly enhance decision-making processes. The visualization of the surveyed area on the ARGOS overview map also received strong approval (8.29), showing that contextual integration of 3D models supports better spatial understanding. The availability of a 3D model itself was considered highly beneficial (7.83), as it helps users assess situations more accurately. Finally, the ability to continuously build and update the 3D model (7.29) was valued for enabling early integration of information into decisions, even before the model's completion. Overall, the results indicate that combining efficient computation, contextual visualization, and iterative updates strongly supports informed and timely decision-making.

The Gazebo robotics simulation was used to evaluate the principles described in D4.2 in relation to **Workflow Optimization** (T4.4). To this end, Real-Time Scoring Agents (RTSA) were used to distribute the various drone simulations across different simulation units in order to arrive at an overall result more quickly. The principle was explained and presented to end users before they gave their feedback. The results show that participants rated the usefulness of RTSAs highly across all evaluated aspects. The highest score (9.0) was given to the statement that RTSAs help evaluate the current operational situation in the simulation more quickly and accurately. The integration of RTSAs into simulations was also seen as very beneficial for practical applications in real-life firefighting operations, receiving a score of 8.0. Slightly lower, but still positive, was the rating of 7.4 for making more informed decisions about the further use of drones during the simulation. Overall, the findings indicate that RTSAs significantly enhance both the efficiency and practical value of drone-based simulations.

The evaluation of **Text Mining** shows a strong and positive user perception overall. Participants rated the automated analysis and visualization of social media content the highest (8.17), indicating that this functionality is seen as a clear improvement over traditional, manual information-gathering methods. The system's ability to help users capture relevant information and gain a better situational understanding also scored well (7.48), demonstrating its practical value in decision support. The inclusion of relevant social media posts in the decision-making process (7.31) further highlights the perceived usefulness of integrating real-time data sources. The slightly lower score (6.83) regarding the transparency of data selection criteria suggests that while users recognize the benefits of Text Mining, there is still room to improve the clarity and interpretability of how relevance is determined. Overall, the results reflect strong appreciation for the efficiency and enhanced situational awareness provided by Text Mining tools, coupled with a desire for greater transparency in automated data processing.

The evaluation of **eXplainable AI** shows generally positive perceptions of its usefulness and comprehensibility. The usefulness of the predictions for estimating fatigue conditions received an average rating of 5.92, indicating moderate confidence in their practical value. The comprehensibility of the system's explanations was rated higher, at 6.71, suggesting that users generally found the explanations clear. Similarly, the statement about always understanding why the system produced certain results achieved a score of 6.29, reflecting a rather good but not complete transparency. Finally, the highest rating of 6.75 was given to the statement that explanations of technological results support understanding and decision-making, implying that users particularly valued the explanatory support provided by the system.

The evaluation of the **Augmented Reality** application reveals a consistently positive user perception. The efficiency of accessing relevant information through AR visualizations was rated 8.38, showing that users found AR very effective in supporting information retrieval. The ability to better assess on-site situations received a similarly high score of 8.25, indicating that AR significantly enhances situational awareness. The clarity and comprehensibility of forest fire information visualizations also achieved 8.38, underlining the strong support AR provides for orientation in complex scenarios. Trust in the presented AR information was rated somewhat lower at 7.38, suggesting that while confidence is generally good, there remains room for improvement. Finally, the helpfulness of uncertainty visualization reached 7.75, reflecting that users appreciate transparency in data representation and find such features useful for interpretation.

The **Handheld Augmented Reality Usability Scale** (HARUS) is a special instrument to evaluate Augmented Reality applications running on mobile devices. The HARUS results indicate that users generally experienced the Augmented Reality application as clear, responsive, and appropriately designed. The perceived mental effort required (5.50) suggests a moderate cognitive load manageable but not effortless. Users rated the amount of information displayed as appropriate (6.88), meaning that the visual density was well balanced. The readability of the information was fairly good, as difficulty in reading scored low (4.50) and readability of words and symbols was evaluated positively (6.00). The system responsiveness received a high score of 6.75, implying that users found the AR interface quick and reactive. In contrast, confusion caused by the display was rated low (4.13) and flickering even lower (3.63), both pointing to a stable and comprehensible visual presentation. Lastly, the consistency of the displayed information, rated 6.38, reinforces the impression of a well-structured interface. Overall, the results reflect a good usability and visual clarity of the AR application, with minor room for improvement in reducing mental workload.

The calculated **SUS score for Augmented Reality** of 75.33 indicates above-average usability. In the commonly used interpretation framework, scores above 68 are considered good, and those above 80 reflect excellent usability. This result shows that users found the AR application intuitive, well-structured, and easy to learn, while only few aspects may still benefit from finetuning. Overall, the AR system demonstrates a strong level of user satisfaction and usability.

The calculated **SUS score for Uncertainty Visualization** in ARGOS and AR of 56.5 indicates below-average usability. In general, SUS scores above 68 are considered good, while scores below 60 suggest usability challenges. This means that although the uncertainty visualization in ARGOS and AR appear somewhat functional and moderately clear, users still experience difficulties possibly related to complexity, need for support, and learning effort. Improving intuitiveness, consistency, and ease of learning could significantly enhance the user experience.

The evaluation of **graphical workflow modeling** with RapidMiner Altair AI Studio in the EmCase extension for fire service deployment planning reveals a very positive user perception overall. The statement that graphical workflow modeling can improve the execution of defined processes received a high rating of 8.92, while the potential to accelerate process execution was rated almost equally high at 8.83. These results indicate that users see clear efficiency and performance benefits from the visual modeling approach. The intuitiveness and comprehensibility of creating a graphical workflow achieved 6.92, showing that the modeling process was generally understandable, though it may require some experience. Lastly, the integration of complex simulation models via RapidMiner operators was rated 7.00, suggesting that the integration is perceived as largely straightforward and applicable in practice. Overall, the findings point to strong usability and high perceived value of the graphical workflow modeling approach for operational planning in emergency scenarios.

The evaluation results for the **EmCase demonstrator system** indicate an overall positive perception of its usefulness for decision-making. Participants reported that the system information influenced their actions with a mean score of 7.22, and that they prioritized or changed decisions based on the data with a slightly lower score of 6.78. The highest rating, 8.35 was given to the statement that the information displayed can support real operational decisions, emphasizing the system's strong practical relevance. Decisions made directly based on the information provided were rated with 7.16, suggesting consistent integration of the system's output into user actions. Participants also demonstrated awareness of data uncertainty, scoring 7.57 for considering visualized uncertainties during decision-making. Finally, the statement about using the system's information even in critical situations received 7.53, highlighting a generally high level of trust and acceptance of the EmCase demonstrator in operational contexts.

2.5.7 Usability evaluation of the integrated Demonstrator System

The usability evaluation of the EmCase demonstrator system resulted in a System Usability Scale score of 68.25 (see Figure 15), which aligns almost exactly with the widely accepted average benchmark of 68. This indicates that the system demonstrates good usability and is generally well accepted by its users. Firefighters working with the demonstrator in both flood and wildfire scenarios reported that they would like to use the system frequently and found its functions to be well integrated. These results suggest that the system is relevant for operational decision-making and that its core functionality supports practical tasks in emergency management effectively. At the same time, some participants pointed to areas that could be improved. A strong effect is implied by the fact that CREXDATA is mostly on data processing technologies which are executed on hidden layers of the system architecture, and only some technologies include a stronger focus on the user interface with visible effect for users.

Ratings related to the need for technical support and the amount of learning effort required indicate that the system may still be somewhat complex for first-time users. A few respondents also mentioned that certain interactions felt cumbersome. Consequently, while the EmCase system provides an overall positive user experience and can be considered suitable for field deployment, further optimization particularly through clearer design interfaces, simpler workflows, and more intuitive training materials could enhance its usability further. Overall, the score of 68.25 shows that the system already achieves a satisfactory level of usability, with promising potential to reach a good to excellent range through targeted refinements. Of particular note is the Technology Readiness Level (TRL) of ARGOS, which is expected to reach 6 (prototype in operational environment). ARGOS is a finished product that has been tested in field trials in the operational environment of the staff room. The CREXDATA technologies developed, on the other hand, are research prototypes and are made usable for end users through visualization in ARGOS.

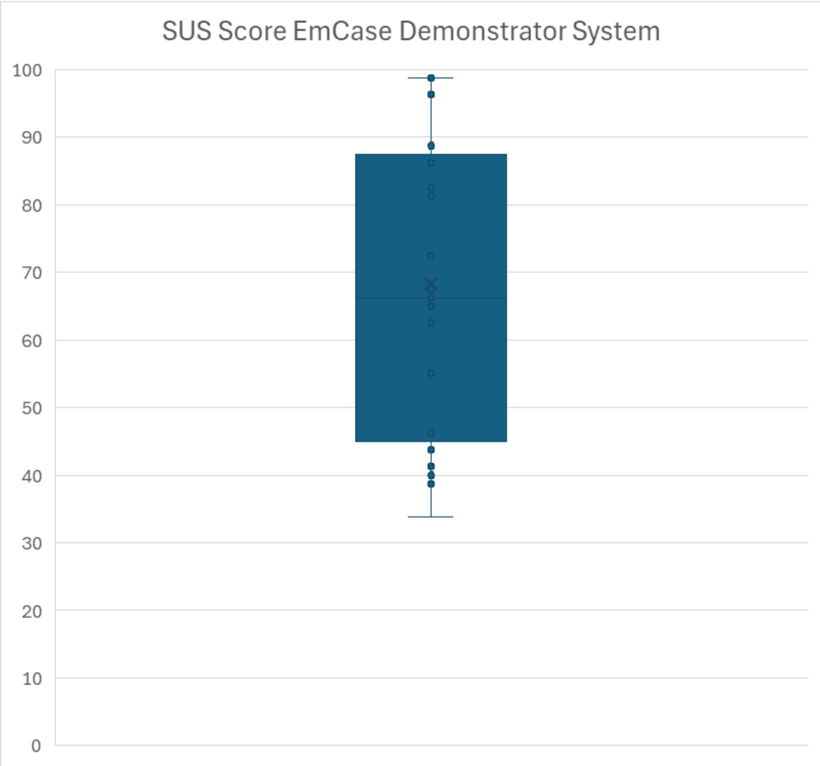


Figure 15: SUS Score boxplot of the EmCase demonstrator system

2.5.8 Investigation of transferability in the Finland trial

A second training event at ESAF in Kuopio was used to evaluate CREXDATA technologies. As part of a large-scale exercise, the transferability of the demonstrator system to other application scenarios in the field of hazard prevention by rescue services was tested. The scenarios were varied, focusing on different types of accidents and scales. In Tampere, prolonged rainfall led to subsoil oversaturation and the formation of a slip surface, causing a building collapse. Residents reported vibrations and visible structural damage, including foundation subsidence, twisted balcony railings, hairline façade cracks, and tilting retaining walls. Power outages, damage to water and wastewater networks, and localized smoke and dust were also reported. In Mikkeli, partial ground movement caused structural vibrations and minor foundation subsidence, leading to the partial collapse of two stairwells and the lakeside end of an apartment building. Several injuries were presumed, with some people trapped due to jammed doors or collapsed structures. Power outages, water network damage, road blockages, slippery conditions, and limited emergency access further complicated the response. Aviation accidents were simulated at Helsinki-Vantaa and Oulu airports. At Helsinki-Vantaa, a plane landed short of the runway, causing an engine fire and fuel spill, with passengers sustaining various injuries. A similar scenario occurred at Oulu airport. Another aviation accident scenario involved an aircraft making an emergency landing on a road between two regional command centers. The fuselage broke apart without fire, and some passengers were injured. In all cases, emergency services were alerted, and initial search and rescue operations were critical. The scenarios described were included in a two-day exercise with a fixed schedule. Three hours were scheduled for each scenario. During this time, the commanders had to perform their usual tasks and were expected to utilize new tools that had been presented and tested in advance. To minimize the influence on the exercise and learning success of the commanders and not to overwhelm them with further new applications from CREXDATA, the CREXDATA system was not used by them, but was operated by the project partners UPB, MoIFI, and FMI. Data from the real situation was entered into the CREXDATA system and processed. The results, in the form of aggregated information on maps, were presented to the commanders together with an explanation to support their decision-making.

To this end, the scenarios and current measures taken by the commanders were tracked and the visualized information from the CREXDATA system was submitted at appropriate moments. Under consideration of the scenarios, text mining, drone simulation, and fatigue assessment using XAI were presented. To demonstrate text mining, the posts classified as relevant were visualized on a map. Due to a lack of real data on the scenarios, tweets with appropriate content were generated. The geo-referenced posts were given to the commanders in early phases to enhance situational awareness and the situational picture. At a time when the first responders were at the scene and could therefore provide reconnaissance results on the extent of the damage, information from the drone simulation was submitted. To demonstrate drone simulation, routes along selected POIs around the incident sites were simulated. In the case of collapses, these routes led along the surrounding buildings to check them for possible damage. In the case of aviation accidents, the areas that were flown over by the aircraft shortly before the accident were to be checked with a drone to identify possible damage and debris. Route alternatives and respective parameters, such as flight duration and energy consumption, were visualized and made available to the commanders. As the mission progresses, it becomes increasingly important to consider the fatigue levels of emergency personnel and thus to evaluate when replacement personnel need to be dispatched to the scene. For that, the fatigue assessment was presented after measures had been initiated in the respective scenario. Since no emergency personnel were active in the field during the exercise, archived biometric data was used.

After completing each scenario, commanders were asked to evaluate the technologies based on their agreement with the given statements. In total, nine commanders participated in the scenarios and evaluated the technologies. The mean value was calculated from all individual ratings. The evaluation shows that the text mining functions are rated very positively overall. All four statements scored between 7.0 and 7.78 points. The transparency in the selection of relevant social media posts (7.78) and the added value of automated analysis and map visualization (7.67) were rated particularly highly. Overall, participants clearly support the usefulness of text mining for understanding situations and making decisions. The overall assessment of the drone simulation is positive. Participants would generally incorporate the system recommendations into their flight planning, although their willingness to follow recommendations in difficult weather conditions was rated slightly lower (6.00). However, the usefulness of the route recommendations was assessed as clearly comprehensible and relevant to decision-making (up to 7.11). The XAI-based fatigue analysis received very high approval ratings. The fatigue predictions (8.56) and the explanatory presentations of the results (8.22) were rated as particularly helpful. The comprehensibility of the system explanations and the reasons for specific recommendations were also rated positively, which effectively supports decision-making. The corresponding diagrams of the results presented are attached in the Appendix 1 (Section 9.9). The evaluations show that text mining significantly supports decision-making. The questions and discussions about text mining also highlighted the commanders' keen interest in this technology. Geo-referenced visualization offered the advantage of a quick overview of the situation. In the case of an aviation accident, for example, where the location of the aircraft was initially unclear and an extensive search had to be launched, this meant that the search area for the rescue services could be narrowed down based on the posts, enabling the emergency services to reach the actual incident site more quickly. Due to the tendency of the population to publish disaster scenarios on social media promptly, text mining was considered by all commanders involved to be highly relevant and helpful for situational awareness and decision-making in operations. The feedback on drone simulations shows a similar trend, even though the quantitative assessment is slightly below the text mining values. The slightly lower rating is due to the fact that the commanders participated already have extensive experience in the use of drones and are therefore less reliant on simulation results. Nevertheless, the visualization of the drone routes with the respective characteristic values was found to be helpful in confirming the assessment with little effort and in estimating at an early stage how many drones are required to cover a specific area. The fatigue assessment using XAI received the highest average approval rating for the statements given. The evaluation shows that both the fatigue levels and the explanation of the assessment support decision-making during operations. In discussions, commanders described the relevance of the technology in both long-term and short-term operations, as rapid fatigue can also occur in short-term operations due to unknown pre-existing stress. Due to the intrinsic motivation of emergency personnel, fatigue is usually hardly noticed, so monitoring and objective assessment help commanders to assess the actual level of fatigue and request replacement personnel to the scene at an early stage. The discussions showed that it is useful to summarize fatigue levels of individual personnel into a team-level rating, which then represents the highest individual rating, since the entire unit is always relieved when individual personnel become fatigued. This illustrates the importance of the multimodal data fusion toolbox for simplified presentation of fatigue assessments and thus better decision support.

The FMI tool was integrated into the annual exercise (Krisu) organized by ESAF, where command-level officers practiced managing various emergency situations in Finland. During the exercise, the FMI tool supported weather-related scenarios, including a forest fire case starting from 15 June 2023 and a windstorm scenario (Storm Jari), which occurred on 20–21 November 2024. In the forest fire scenario, the FMI tool was presented through the ARGOS system using its historical view (see Figure 16). The forest fire model simulation was run for the selected dates starting at Thursday 15/06/2023 03:00 EEST and updated every 3 hours until Saturday 17/06/2023 03:00 EEST. The forest fire simulation enabled the command centers to utilize platform information to estimate the risk level of forest fires, the duration of the risk, and the municipalities most affected. The command centers were able to access and use this information effectively.

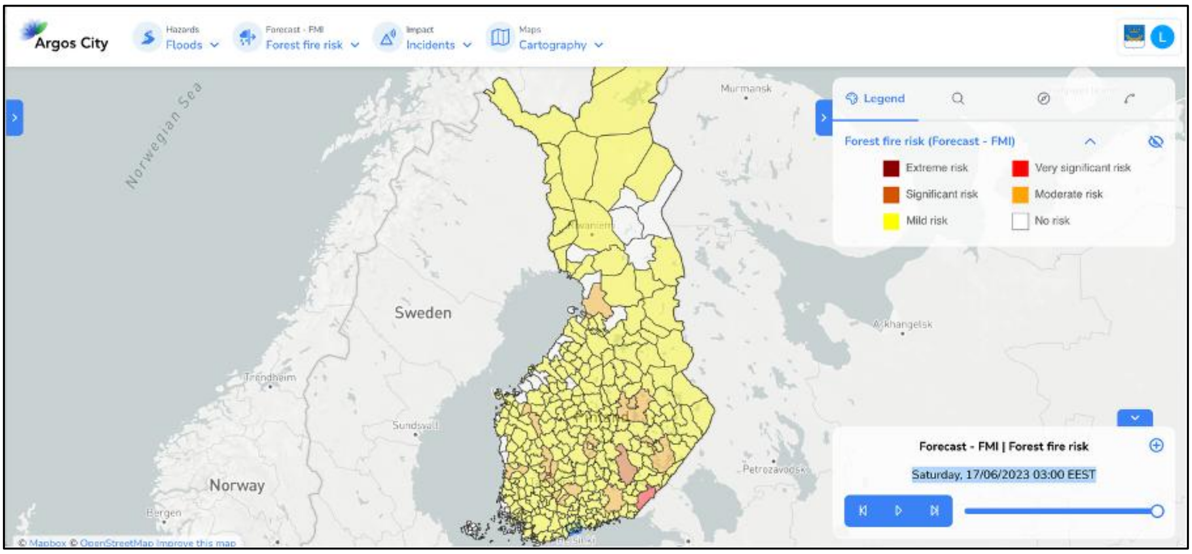


Figure 16: FMI Forest fire model simulation view 15-17.6.2023 in ARGOS

For the windstorm scenario, the FMI tool outputs were delivered as PDF forecasts sent via email to the command centers. In this case, the tool provided forecasts on traffic accident risk caused by snow blizzard conditions and storm damage clearance risk due to intense wind speeds. The qualitative feedback from end users was highly positive. They emphasized that the FMI tool's ability to forecast impacts and risk levels, such as traffic accidents, storm damage clearance, and forest fires, was significantly more useful than traditional tools that focus only on weather parameters. Users found the time-series view particularly valuable for estimating how long severe conditions would persist, which helped them plan resource allocation, including both official and voluntary teams. Additionally, the tool enabled them to determine safe time windows for storm damage clearance, a capability currently missing from their operational systems. Also, in the future this type of impact-based forecasts could be used in resource management and planning purposes. The lessons learned from this exercise highlight that even high-quality tools must be integrated into existing systems or platforms to be truly effective, given the large volume of information and numerous parallel systems in use. Building trust in new tools takes time, but clear explanations of their benefits and practical applications make end users more interested, hopeful, and eager to adopt them.

In summary, the trials in Finland show that CREXDATA EmCase technologies support decision-making not only in wildfire and flood scenarios, but also in other types of operations. While drone simulation demonstrates its added value particularly in large-scale operations, where drones are operated close to their maximum range due to long flight times, text mining also provides support in smaller operations with high public visibility. Similarly, fatigue assessment based on XAI is helpful in various types of operations to detect fatigue resulting from pre-stress caused by other operations or everyday work among volunteer emergency personnel, even during short operations. This demonstrates the transferability of the technologies beyond the scenarios considered in EmCase.

2.6 Lessons Learned

The evaluation and testing activities carried out across the CREXDATA EmCase provided extensive insights into the system's usability, performance, and operational relevance in emergency situations. A key lesson learned is that while the individual technological components are diverse ranging from simulation models and machine learning modules to augmented reality interfaces and workflow tools their effectiveness depends greatly on user-centered design, transparency of results, and the clarity of information visualization.

One of the most important findings concerns the EmCase demonstrator platform ARGOS. Users appreciated its intuitiveness and overall user-friendliness, but they also identified transparency and uncertainty visualization as critical aspects for improvement. Although users found the system's predictions and explanations generally clear, they sometimes lacked a deep understanding of why specific recommendations were generated. This underlines the necessity of integrating more explainability features, particularly in decision-support environments where human trust and comprehension are essential for adoption.

The assessments of the EmCase simulation models revealed good user satisfaction, particularly regarding their operational usefulness and efficiency. The EmCase SHS GUI proved intuitive and effective for the parametrization of domain-specific simulation models without requiring deep technical knowledge. The FloodWaive urban flooding simulation was highlighted as especially powerful due to its rapid computational performance and realistic predictions, enabling reliable real-time application. Similarly, the Spark wildfire simulation excelled in providing clear visualizations and trustworthy forecasts, underlining the importance of visual transparency in aiding rapid decision-making. The Gazebo robotic simulation demonstrated that users valued its planning recommendations, but some hesitation remained when facing uncertain or high-risk conditions an important reminder that technological trust must be built gradually through consistent system reliability and intuitive interface feedback.

Complex Event Forecasting proved to be a valuable decision-support tool, but users still desired higher predictive accuracy and broader demonstrable reliability. Interactive Learning, on the other hand, was seen as highly effective, particularly in optimizing resource allocation in real-time operations showing that automation and adaptive algorithms can significantly enhance operational efficiency when they are interpretable and easy to use. The results for Federated Machine Learning with NeRFs were especially encouraging: users appreciated the combination of fast computation, contextual 3D visualization, and iterative updating. This indicates how future decision-support systems can benefit from tightly integrating high-performance computing with intuitive spatial representation.

With Text Mining and eXplainable AI the importance of transparency became evident. Automated text analysis of social media content was widely praised for improving situational awareness, yet users wished for greater clarity regarding data selection criteria. Similarly, explainable AI approaches were appreciated for their interpretability, but full trust was still hampered by incomplete transparency of individual predictive outcomes. These findings emphasize that future development should focus not only on algorithmic performance but also on improving the interpretability, traceability, and communication of results to non-technical users.

The results related to Augmented Reality applications consisted of notably positive feedback, highlighting AR's strong potential to enhance situational awareness and facilitate access to relevant information in dynamic field environments. Users valued the clarity and responsiveness of the AR interface and its ability to visualize complex information intuitively. The moderate cognitive effort reported suggests a well-balanced information density; however, some users noted potential improvements in trust and in the representation of uncertainties. These insights, supported by a good AR SUS score, confirm that AR is a mature and valuable tool for decision support, particularly when combined with careful management of cognitive workload and data reliability.

At system level, the EmCase demonstrator demonstrated satisfactory good usability, aligning closely with general usability benchmarks. Fire service personnel responded positively, finding the system relevant and well-integrated into their workflows. Practical engagement with the demonstrator, especially in flood and wildfire contexts confirmed that its visualizations, simulations, and predictive capabilities effectively support real operational decisions. At the same time, feedback indicated that some interactions remain complex, and that first-time users may require additional support and training. This shows the value of iterative design and user co-creation, ensuring that systems developed for high-pressure operational contexts are tailored to the working methods and expectations of emergency responders.

An important overarching lesson of the Finnish field trials is that the transferability of CREXDATA EmCase technologies extends beyond the original wildfire and flood use cases. Drone simulations provide strong value in large-scale operations, especially when flight ranges are stretched to their limits. Text Mining, on the other hand, offers tangible benefits in smaller-scale, highly visible events where situational communication is central. Additionally, fatigue assessment using explainable AI proved valuable in both long and short operations, helping to identify early signs of operator strain or overexertion. This demonstrates the adaptability and scalability of the EmCase system components for various crisis management scenarios.

In summary, the lessons learned from the EmCase evaluations highlight several cross-cutting themes. First, user trust and transparency are as important as algorithmic precision that technologies clearly communicate their reasoning foster confidence and effective adoption. Second, visualization and usability play decisive roles in determining whether complex simulation and prediction outputs can be understood and acted upon in real time. Third, modular integration and interoperability across components enhance overall value, as users can combine simulations, forecasts, and analyses fluidly during decision-making. Finally, systematic user involvement, iterative refinement, and consistent emphasis on training and intuitiveness will be key to achieving widespread acceptance and operational success of the next-generation CREXDATA decision-support technologies.

3 Health Crisis Use Case

3.1 Epidemiological Scenario

3.1.1 Introduction

The COVID-19 pandemic demonstrated how rapidly infectious diseases can spread in an interconnected world, where global mobility, dense urbanization, and frequent social interactions facilitate exponential transmission. These challenges underscore the need for computational frameworks capable of anticipating epidemic dynamics, evaluating intervention strategies, and supporting evidence-based decision-making during health crises.

The Epidemiological Scenario of the Health Crisis Use Case addresses this challenge by developing a computational toolbox that integrates epidemiological simulations, high-performance computing (HPC), artificial intelligence (AI), machine learning (ML), and visual analytics (VA) to enhance model calibration, epidemic forecasting, and intervention design. This toolbox enables researchers and decision-makers to integrate multimodal data into the simulator, explore large parameter spaces, quantify uncertainty in the model's forecasting, and optimize response strategies. This scenario constitutes one of the Health Crisis Use Case demonstrators within CREXDATA, focusing specifically on epidemic modelling.

At the core of the developed toolbox lies *EpiSim.jl*, a multi-agent simulator of epidemic processes on metapopulations. The model integrates multimodal data, including demographic structures, population mobility, and epidemiological data to simulate the spatiotemporal evolution of infectious diseases across interconnected populations. Spain was selected as the reference case, leveraging high-resolution datasets on mobility, demographics, and COVID-19 incidence to construct detailed national and regional models. Because metapopulation models are realistic yet highly parameterized, they require calibration of numerous uncertain or unknown parameters and repeated simulations to quantify uncertainty or evaluate interventions. To efficiently address these challenges, CREXDATA developed a suite of scalable HPC workflows that combine simulation-based learning with visual analytics for model exploration and optimization.

The use case was structured around a pilot study of the COVID-19 pandemic in Spain, serving as a realistic testbed for data integration and validation, large-scale simulation, and decision support. The following demonstrators showcase the toolbox's capabilities for: i) parameter calibration and uncertainty estimation; ii) controlling disease outbreaks through non-pharmaceutical interventions; and iii) optimizing vaccination campaigns. In addition to the three simulation-based pilots, we also developed data-driven approaches to complement these workflows. These include methods based on Transfer Entropy and visual analytics to uncover directional relationships between mobility and epidemic spread, enhancing the interpretability and validation of simulation results.

Together, these developments demonstrate how CREXDATA technologies provide a unified, data-driven framework for simulating and managing epidemic scenarios, offering a foundation for the broader multi-scenario framework developed within the project.

3.1.2 Methods

3.1.2.1 Data

We used publicly available datasets from national health authorities and the Spanish Ministry of Transport (MITMA) gathered in the COVID-19 Flow-Maps project [4]. Epidemiological data included daily and weekly reports of COVID-19 cases, hospitalizations, and deaths, disaggregated by age group and geographic level (country, province, and MITMA mobility zones). Demographic data were derived from census-based population distributions, and recurrent origin-destination matrices were reconstructed from anonymized mobile phone data describing daily mobility patterns between 2850 different regions, corresponding to municipalities or districts. Integrating the different data sources we have implemented a set of ready to use *EpiSim.jl* reference models [5].

We also investigated the possibility of having comparative data from Finland. Together with the Finnish project partners (FMI and MoIFi). The Finnish data would have covered the capital area (Helsinki, Espoo, Vantaa, Kauniainen) on two time periods. However, in the end, it was decided that adding a separate data set would have not provided sufficient added value for the project and thus the data was not acquired.

3.1.2.2 Simulators

We developed *EpiSim.jl* [6], a modular high-performance simulator for modelling epidemic dynamics across spatially structured populations. Built upon the *Microscopic Markov Chain Approach* (MMCA) introduced by Arenas et al. (2020) [7], it enables realistic simulations of infectious disease spread within a metapopulation—a network of spatially connected subpopulations (e.g., provinces or districts) linked by daily mobility flows. The model supports multiple age groups with distinct mobility and contact patterns, allowing the specification of group-specific epidemiological parameters such as infection probability, hospitalization risk, and mortality rate. Disease progression follows an extended SEIRD framework including hospitalization and fatality compartments, with parameters derived from literature and calibrated against real epidemiological data. All inputs, parameters, and datasets are defined in a single JSON configuration file, ensuring transparency, reproducibility, and ease of use.

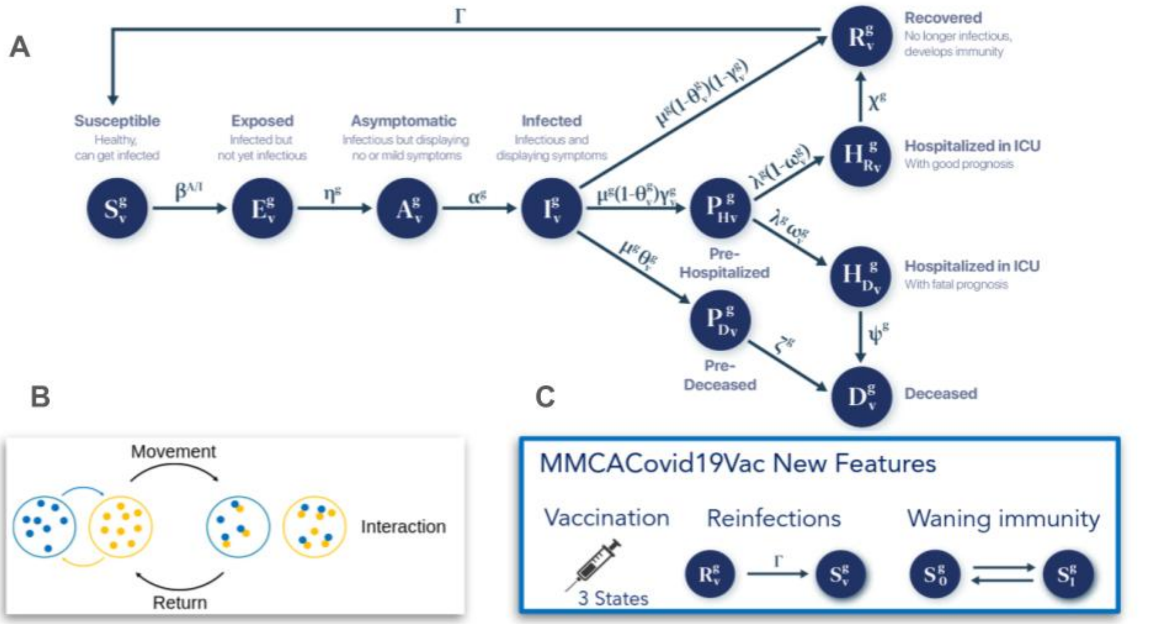


Figure 17: Outline of the MMCAcovid19 and MMCAcovid19Vac models. Panel A illustrates the compartmental structure of the model, where each node represents a disease stage, and arrows denote transition probabilities between states. Superscripts and subscripts g and v indicate dependence on age group and vaccination status, respectively. Panel B depicts the metapopulation dynamics, while Panel C shows the additional features introduced to support vaccination modelling.

Within CREXDATA, the base *MMCAcovid19* model was extended to *MMCAcovid19Vac* [8], which introduces vaccination and immunity dynamics (Figure 17). This enhanced version models i) the effect of vaccination on transmission and severity, ii) the accumulation of herd immunity, and iii) reinfection due to partial immunity loss. Each regional population is subdivided into three vaccination states, non-vaccinated, vaccinated, and post-vaccinated, capturing the temporal and epidemiological effects of vaccine rollout.

Both models were implemented as independent Julia modules integrated into *EpiSim.jl*, ensuring scalability across local and HPC environments. The simulator supports multiple spatial resolutions, from national to regional levels (e.g., Catalonia, Madrid), using mobility networks derived from the MITMA dataset of anonymized phone-based commuting data. The standardized configuration structure and modular engine design make *EpiSim.jl* a central asset of the CREXDATA Health Crisis toolbox. It provides a robust foundation for large-scale, data-driven epidemic simulations and underpins the project's workflows for calibration, optimization, and uncertainty analysis.

3.1.2.3 Workflows

HPC-Based Model Exploration Workflow

We developed *episim-emews* [9] , a high-performance exploratory workflow that integrates the *EpiSim.jl* simulator with the EMEWS [10] framework for model exploration and optimization at scale. The workflow is designed to exploit parallel computing resources efficiently, enabling the simultaneous execution of thousands of epidemic simulations on HPC infrastructure. It supports two primary modes of operation:

- Parameter sweeping, where simulations are run over predefined parameter grids or a custom list of candidate parameters.
- Optimization-based exploration, where parameters are iteratively refined using machine-learning algorithms such as Genetic Algorithms (GA) or Covariance Matrix Adaptation Evolution Strategy (CMA-ES).

The overall workflow architecture is illustrated in Figure 18A which depicts the Swift/T orchestration responsible for distributing parameter sets to worker instances and collecting results for iterative refinement. Each worker instance executes an independent *EpiSim.jl* simulation, including the pre- and post-processing, as shown in the bottom panel, which details the sequence of computational steps in the evaluation of a single set of parameters.

The model exploration algorithm generates new parameter sets, which are queued and dispatched to multiple worker instances for evaluation. Each worker runs an independent simulation and returns evaluated parameters that feed back into the learning step for iterative optimization. Input data and configuration files are prepared for each instance, and simulation results are stored in standardized NetCDF format. Visual analytics modules can generate epidemic curves and uncertainty maps, while evaluation scripts compute objective function values to guide the parameter learning process. Each simulation cycle follows six main steps (Figure 18B).

The *episim-emews* workflow is fully configurable through a single JSON file that defines every step of a model exploration run—from simulation to post-processing, visualization, and evaluation—without requiring code changes. Users can specify custom post-processing operations (e.g., aggregation or data transformation), configure automated visual analytics for diagnostic or uncertainty plots, and select or define evaluation metrics such as RMSE. This modular configuration ensures flexibility, reproducibility, and adaptability across diverse epidemiological modelling and optimization experiments.

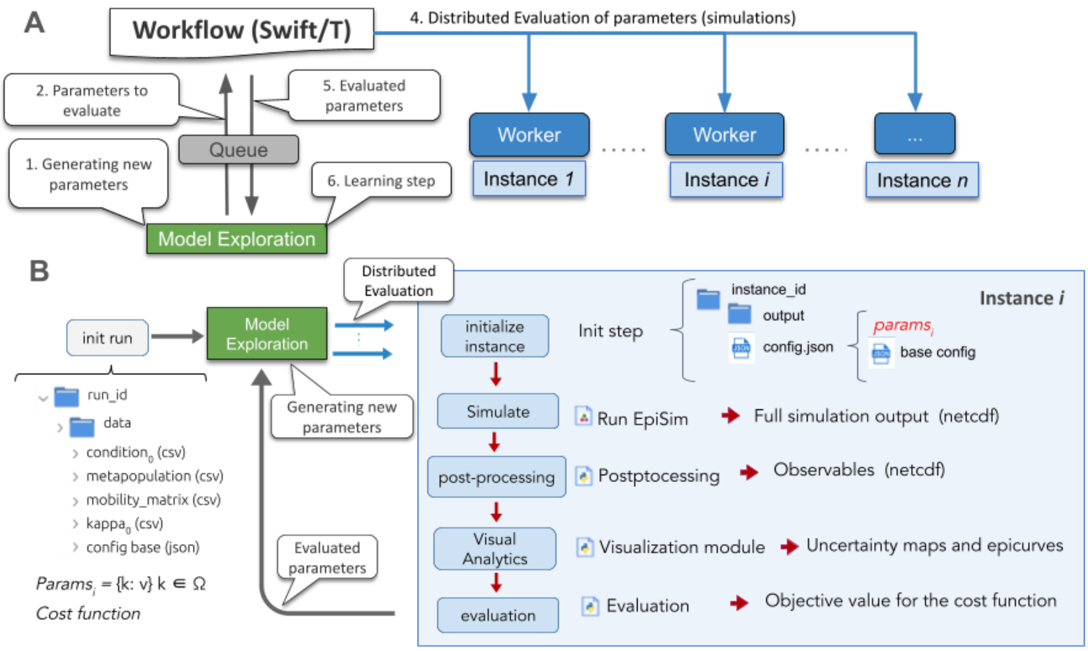


Figure 18: Architecture of the episim-emews high-performance workflow for epidemic model exploration and calibration. Panel A depicts the orchestration layer, which manages the model exploration module and the distributed execution of simulations on high-performance computing (HPC) infrastructure. Panel B shows how a worker instance executes a complete EpiSim.jl simulation workflow, consisting of initialization, simulation, post-processing, visualization, and evaluation steps.

Reinforcement Learning workflow for policy optimization

To extend the decision-support capabilities of our simulator, we developed *episim-rl* [11], a reinforcement learning (RL) workflow that couples *EpiSim.jl* with a Q-learning agent in collaboration with WP4. We used our *episim-rl* to identify optimal dynamic non-pharmaceutical intervention (NPI) policies. In this closed-loop setup, the agent interacts with the epidemic model through iterative simulation–feedback cycles. Each episode represents a full epidemic trajectory, divided into shorter decision intervals (e.g., biweekly). At each step, the agent observes key indicators—such as ICU occupancy, infection levels, mortality, and the effective reproduction number (R_0)—and selects the confinement intensity (parameters κ_0, δ) to apply (Table 8). More technical details on the *episim-rl* implementation can be found in D4.2.

Table 8. Mobility Reduction parameters.

Symbol	Description
t^c	Time steps when the containment measures will be applied
κ_0	Mobility reduction
δ	Social distancing factor

After each simulation step, a reward function evaluates the trade-off between health outcomes and socio-economic impact, penalizing hospitalizations, deaths, and excessive restrictions. This feedback updates the agent's policy, which progressively improves over successive episodes until converging on an optimal balance between epidemic control and societal cost. The modular architecture of *episim-rl* allows users to redefine the reward structure, learning parameters, and observables, enabling flexible adaptation to different epidemic contexts and decision-making scenarios.

3.1.2.4 Visual Analytics

To evaluate the outcomes of our simulation workflows, we developed visual analytics to assess and communicate uncertainty in model predictions. The calibration workflows explore the parameter space to identify sets of parameters that reproduce observed epidemiological data. However, multiple distinct parameter combinations can yield similarly good fits—an effect arising from model complexity and from the limited spatial and temporal resolution of available data.

To address this, and in collaboration with WP5, we designed a visual analytics approach that compares simulation trajectories generated using different candidate parameter sets that all fit the data with comparable accuracy. For each region in the modelled metapopulation, we compute the mean and variance of the arrival time of the first infection across these simulation trajectories and represent these measures as maps using colour scales to convey both the expected value of the forecast and its associated uncertainty. This approach provides an intuitive representation of spatiotemporal epidemic dynamics, facilitating visual comparison of trajectories from multiple calibrated parameter sets. It supports the identification of parameter uncertainty regions, improves understanding of model sensitivity, and enhances confidence in the robustness of simulation-based forecasts.

In addition, we contributed to the development of a visual analytics pipeline for detecting stable cross-impact patterns in time-dependent data, in collaboration to T5.3. In this pipeline, we combine a tolerance-aware variant of Kendall's tau and a new developed interactive Ikat plot for exploring lag-by-time dependencies, and event extraction. This framework enables users to identify and interpret persistent relationships between epidemiological indicators and external drivers such as mobility, supporting interactive exploration of lag structures and parameter tuning. The approach was applied to the analysis of epidemic and mobility data, demonstrating its usefulness for uncovering robust dynamic relationships and informing subsequent modelling workflows.

3.1.2.5 Data-Driven Analysis of Epidemic Dynamics

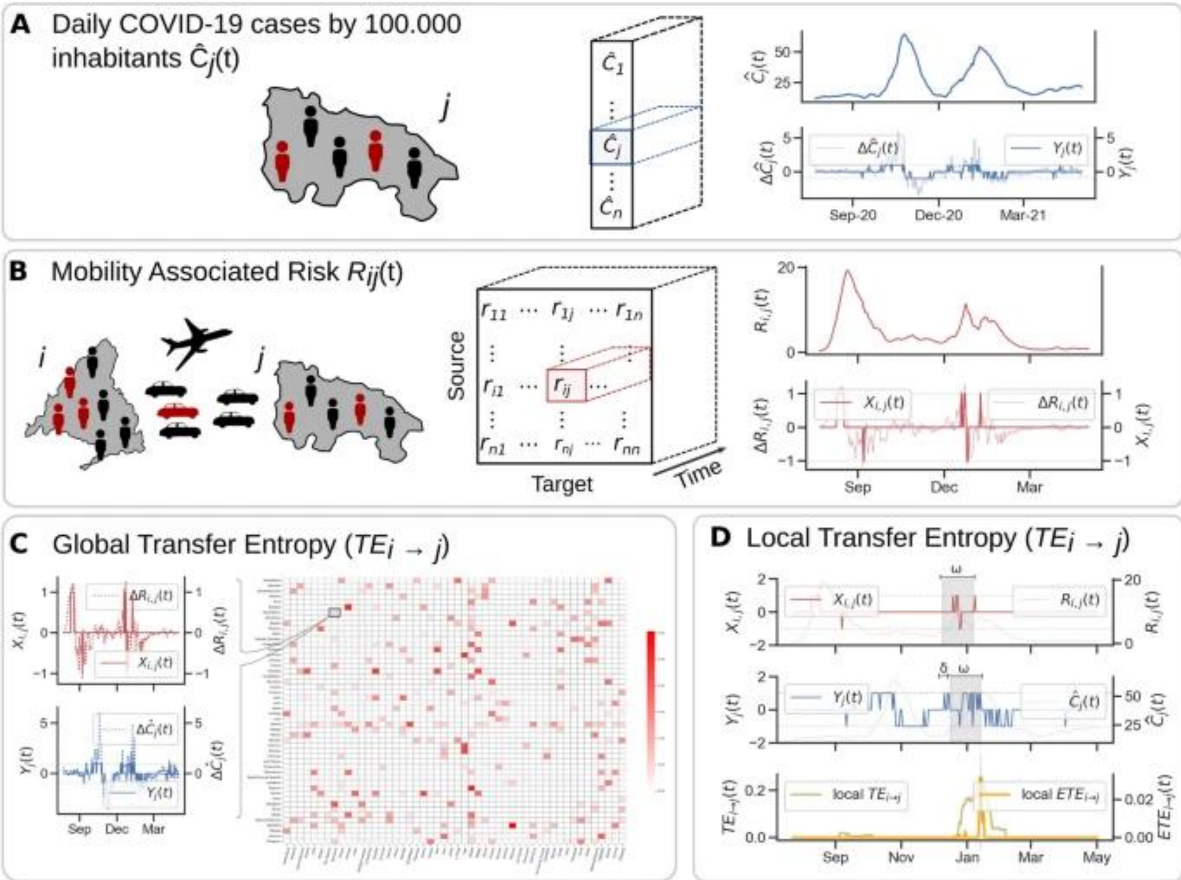


Figure 19: Measuring transfer entropy between time-series of phone-based mobility data and incidence to explore patterns of disease spread. Source: [12].

Complementing the simulation-based workflows, we developed a data-driven analytical framework to quantify and visualize the directional relationship between interregional mobility and epidemic spread. This workflow combines information-theoretic analysis, real-world mobility data, and metapopulation simulations to examine how population movement shaped the spatiotemporal propagation of COVID-19 across Spain [12].

The framework employs Transfer Entropy (TE) to measure the flow of information from mobility-associated risk—defined as the estimated number of potentially infected individuals travelling between regions—to local infection incidence (Figure 19). This allows for the detection of directional coupling between mobility and disease spread, identifying regions acting as drivers or receivers of infection transmission. By computing TE dynamically through a sliding-window approach, the method captures temporal variations in mobility-driven transmission strength and supports the identification of periods when specific connections dominate the epidemic dynamics.

3.1.3 Results and evaluation

3.1.3.1 Pilots

The pilots were designed using the COVID-19 pandemic in Spain as a reference case to demonstrate and evaluate the technologies developed for the Epidemiological Scenario. Each pilot addresses a specific modelling or decision-support challenge, i.e. parameter calibration, adaptive intervention control, and vaccination strategy optimization, showcasing how the developed workflows integrate simulation, machine learning, and visual analytics to improve epidemic forecasting and management.

The first pilot focuses on calibrating unknown epidemiological parameters by fitting simulations to observed incidence, hospitalization, and mortality data. Using the *episim-emews* workflow, in this pilot we combine ML and VA into an *interactive learning* workflow to efficiently explore the parameter space and quantify uncertainty in epidemic forecasts, addressing *KPI 3 Calibration of epidemiological parameters* and *KPI 4 Characterization of the parameter space with 50% fewer simulations*.

The second pilot targets the control of epidemic outbreaks through adaptive non-pharmaceutical interventions (NPIs) such as mobility reduction and social distancing. The *episim-rl* workflow couples reinforcement learning with simulation to identify dynamic intervention policies that minimize hospitalizations and deaths while limiting the socio-economic impact, addressing *KPI 2 Use the runtime adaptation of simulation trajectories to improve the outcomes of 5 scenarios or patients*.

The third pilot focuses on designing effective vaccination campaigns by optimizing key decision variables, including the start time, duration, and prioritization of vaccination across age groups. Using *episim-emews*, large-scale scenario exploration was conducted to identify Pareto-optimal strategies that minimize mortality and hospitalizations during the second wave, directly contributing to KPI 1 (Forecasting intervention parameter sets that reduce infections).

In addition to the simulation-based pilots, we also developed data-driven approaches to complement and validate the modelling workflows. These include a mobility-informed information-theoretic framework based on Transfer Entropy to quantify directional relationships between population movement and epidemic spread, and visual analytics tools for exploring dynamic correlations and lagged dependencies in epidemic data. Together, these developments extend the use case beyond simulation-based analysis, providing empirical insight and interactive tools for understanding and communicating the drivers of epidemic dynamics. Together, these pilots demonstrate the potential of CREXDATA technologies to translate advanced epidemic modelling into actionable insights for policymakers, enabling transparent and evidence-based decision support during health emergencies.

3.1.3.2 Demonstrators

Parameter Calibration and Uncertainty Estimation using Interactive Learning

This demonstrator shows how the *EpiSim.jl* simulator and *episim-emews* workflow can be used for efficient parameter calibration and uncertainty estimation through interactive learning. The goal was to demonstrate that the workflow effectively explores complex, high-dimensional parameter spaces and iteratively improves calibration using expert feedback and visual analytics.

We defined epidemiological and confinement parameter ranges, along with initial conditions for the first reported COVID-19 cases in Spain, mapped to the metapopulation model. The calibration employed the CMA-ES algorithm for parameter exploration and used the Root Mean Square Error (RMSE) between simulated and reported hospitalizations and deaths as the cost function. Initial runs rapidly converged to parameter sets reproducing national-level epidemic trends, demonstrating the workflow’s efficiency and ability to approximate real dynamics with relatively few simulations (Figure 20A).

Through interactive learning, users iteratively analyse calibration results and uncertainty maps to refine parameter ranges, relax dependencies, and adjust initial infection seeds. This revealed compensatory relationships (e.g., between infection probability and infectious period), indicating that the parameters are not identifiable. Figure 20 panel B and C, show the improvement from the initial (blue) to the final (orange) iteration. Allowing age-specific variations and adding the rate of hospitalization of younger individuals into the calibration improved fits across age groups and provinces, achieving more realistic spatio-temporal patterns (Figure 20B-C). Furthermore, a comparison with a brute-force parameter sweep confirmed the advantages of the evolutionary calibration: CMA-ES converged to accurate solutions after evaluating only ~50,000 parameter sets versus 10^6 for random sampling, fully meeting KPI 3 (Calibration of epidemiological parameters to fit incidence time series) and KPI 4 (Characterization of the parameter space with 50% fewer simulations).

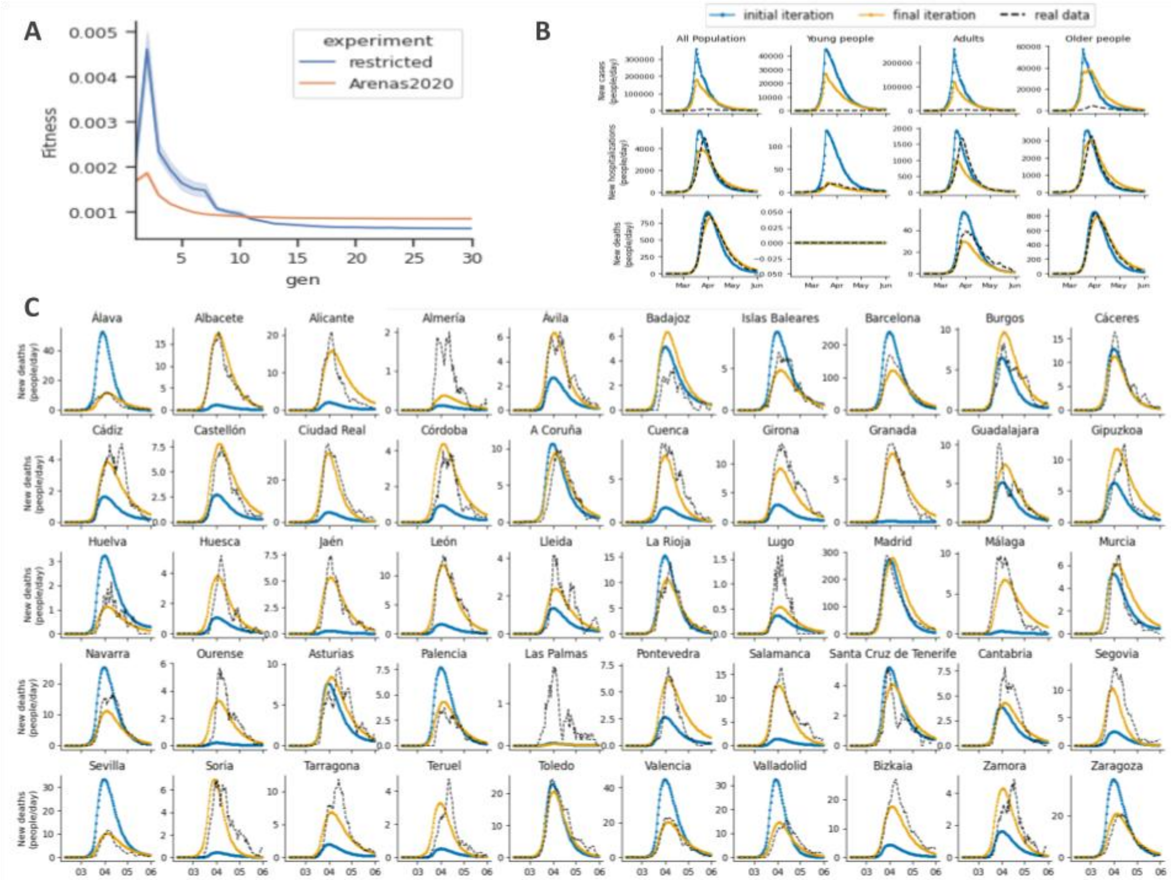


Figure 20: Parameter calibration and uncertainty refinement using the episim-emews. Panel A shows the convergence of the CMA-ES algorithm for two iterations. Panels B and C show the comparison of reported and simulated epidemic curves for Spain across age groups and provincial-level daily deaths, respectively.

Finally, to address forecast uncertainty, we analysed multiple well-fitted parameters sets that, despite differing substantially in value, produced similar epidemic dynamics. As part of the workflow, we applied the developed visual analytics approach to assess a key indicator forecasted by the simulation, the arrival time of the first infected individual (ATFI) in each region, and its associated uncertainty.

The resulting maps (Figure 21) reveal that regions with low population density or greater distance from the initial infection foci tend to experience later epidemic arrival times, reflecting slower disease diffusion through the mobility network. In contrast, densely populated and highly connected urban areas show early infection onset and lower variance, indicating more predictable dynamics. The variance map also highlights areas where uncertainty is higher, suggesting regions where small changes in parameters can substantially alter local epidemic timing. Comparing these maps provides an intuitive summary of the simulated trajectories, showing that most spatiotemporal patterns remain consistent despite parameter variability.

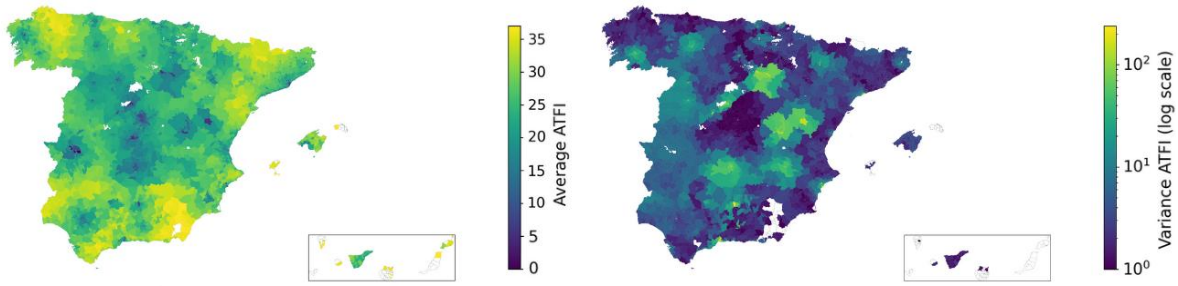


Figure 21: Maps of the average (left panel) and variance (right panel) of the Arrival Time of the First Infected over different calibrated experiments.

Together, these results demonstrate the value of this approach in identifying spatial patterns of predictive uncertainty and enhancing confidence in the robustness of the calibrated forecasts.

Reinforcement Learning for Dynamic NPI Optimization

To demonstrate the capabilities of the *episim-rl* workflow in designing adaptive non-pharmaceutical intervention (NPI) policies, we trained a Q-learning agent coupled with the *EpiSim.jl* simulator. The agent iteratively interacts with the epidemic environment, adjusting two control parameters—κ_0 (mobility reduction) and δ (social distancing)—every two weeks based on epidemic indicators such as hospital load and fatalities. A total of 5,000 training episodes were conducted under three initial epidemic conditions (low, medium, and high severity), allowing the agent to explore the full parameter space. Each episode represented a simulated epidemic wave, and the reward function balanced health outcomes (hospitalizations and deaths) against the socio-economic costs of restrictions. The training process converged toward stable control policies that effectively minimized the overall epidemic burden.

The mobility reduction observed during the first wave of COVID-19 (Figure 22A), derived from the anonymized mobility data, remained at moderate levels for most of the period, decreasing only toward the end. In contrast, the best strategies uncovered by the workflow exhibit a different pattern, characterized by an initial strong reduction in mobility, followed by alternating phases of relaxation and reinforced social distancing. Compared to a baseline simulation fitted to the real pandemic response, these adaptive strategies achieved a more efficient containment of the outbreak, significantly reducing both hospitalizations and total deaths.

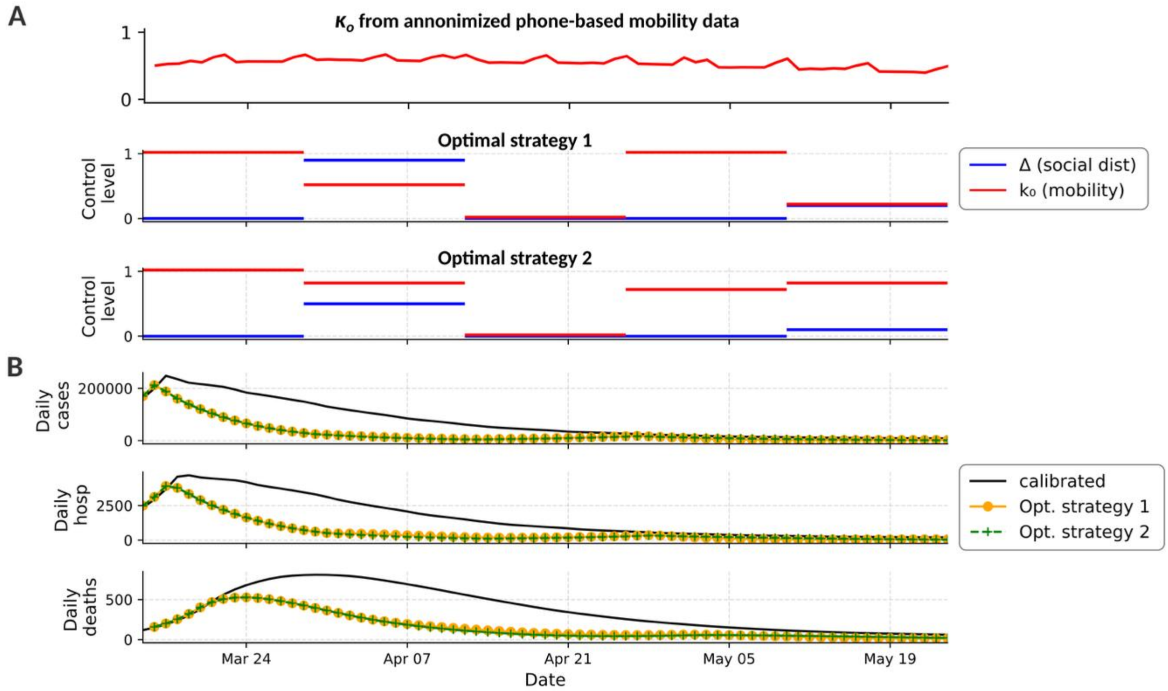


Figure 22: Summarizes the outcomes for the medium-severity scenario. Panel A shows the level of mobility reduction observed during the COVID-19 pandemic alongside two representative optimal control strategies identified by the RL agent, which dynamically adjust mobility and contact reduction levels throughout the epidemic course. Panel B compares the resulting epidemic trajectories against the baseline calibrated model.

These results demonstrate that the *episim-rl* framework can autonomously learn effective, interpretable, and adaptive epidemic control policies. By coupling reinforcement learning with mechanistic simulation, the system identifies strategies that balance public health impact and socio-economic cost without requiring predefined intervention schedules. This demonstrator fully achieves KPI 2 (runtime adaptation of simulation trajectories to improve scenario outcomes), confirming the system's ability to dynamically optimize interventions in real time during epidemic evolution.

Designing Effective Vaccination Campaigns

This demonstrator aims to evaluate the pilot *episim-emews* workflow coupled with the extended MMCACovid19Vac model to explore the role of re-infection, herd immunity, and to design effective vaccination strategies for epidemic control. The simulator incorporates vaccination as an additional compartment within the MMCA-based epidemiological framework, allowing different epidemiological parameters for vaccinated and non-vaccinated individuals. The formalism enables testing alternative vaccination strategies by varying parameters such as the start date, duration, and intensity of the vaccination campaign, as well as prioritization by age group. These parameters are summarized in Table 9.

Table 9. Parameters to define vaccination strategies

Symbol	Description
ϵ^g	Total vaccinations per age strata
start_vacc	Start of the vaccination
dur_vacc	Duration of the vaccination
end_vacc	End of the vaccination
t^v	[start_vacc, end_vacc, T]

To assess the effectiveness of vaccination strategies, we conducted large-scale simulation experiments using high-performance computing (HPC) infrastructure. The exploration process systematically varied vaccination parameters across thousands of simulations to identify strategies that minimize two key outcomes:

- The relative variation in the second peak of new infections or hospitalizations
- The relative variation in cumulative deaths

Both are compared to a baseline scenario without vaccination. In the initial phase (reported in the midterm deliverable), these analyses were conducted using the Spain-wide model at the provincial level, revealing that early vaccination targeting the elderly population substantially reduced total mortality (by 20–50%) and lowered the epidemic peak (by 5–17%).

For this final evaluation, we extended the experiments to include higher spatial resolution models (MITMA level) and a regional model for Catalonia, confirming the robustness and generalizability of the findings. Across all configurations, early-start vaccination campaigns focusing on high-risk (older) groups consistently achieved the best trade-offs between reducing mortality and mitigating epidemic intensity, forming the Pareto front of optimal solutions. Figure 23 shows the outcome of this optimization, where each point represents a unique vaccination strategy simulated under different combinations of timing, duration, and prioritization. The Pareto-optimal strategies (black line) highlight two dominant solution classes: i) early interventions that maximize mortality reduction and ii) later interventions that effectively suppress secondary peaks. These results demonstrate that *episim-emews* can efficiently explore large, multidimensional parameter spaces to identify policy-relevant intervention strategies under uncertainty.

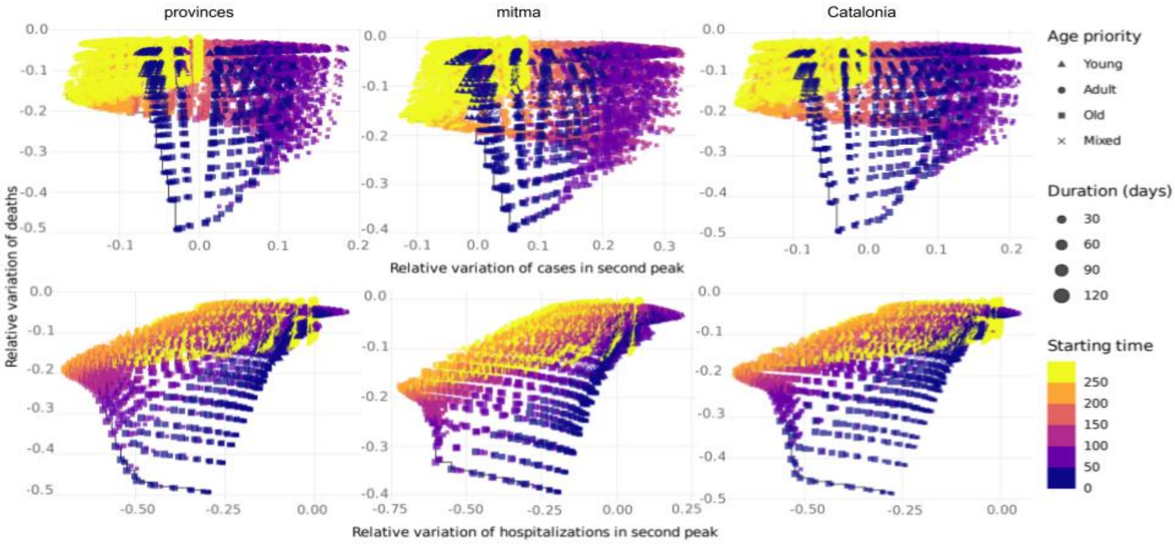


Figure 23: Results from the optimization of vaccination strategies using the models of Spain (MITMA and provinces) and Catalonia. Across all cases, the Pareto-optimal solutions consistently indicate that prioritizing elderly populations yields the most effective outcomes in reducing mortality and mitigating epidemic peaks.

Altogether, this demonstrator successfully meets KPI 1 (Forecasting 7 parameter sets that reduce the COVID infection), and the results confirm that the developed computational workflow can be used to design, evaluate, and optimize vaccination campaigns in realistic epidemic settings, supporting data-driven public health planning and decision-making.

Data-Driven Analysis of Epidemic Dynamics

Finally, we complemented the simulation-based workflows by developing two data-driven analysis to empirically assess the relationship between population mobility and epidemic dynamics. On the one hand, we contributed to task T5.2 in the development of a visual analytics pipeline that integrates a tolerance-aware variant of Kendall's tau, an interactive lkat plot for exploring lag-by-time relationships. We use this to explore data-driven relationships between mobility patterns and epidemic spread, helping to detect stable cross-impact intervals where changes in mobility precede shifts in infection trends (further details can be found in D5.2).

In the second data-driven workflow we assessed the impact of interregional mobility by measuring information flow between time series of mobility and cases using Transfer Entropy. We applied this novel approach to uncover spatio-temporal patterns of mobility-driven transmission using real-world data of the COVID19 pandemic in Spain. The main results of the analysis revealed interpretable spatial and temporal patterns consistent with epidemiological evidence, such as the Lleida-Zaragoza outbreak chain during summer 2020 and the impact of non-pharmaceutical interventions in Catalunya, where bar and restaurant closures led to a marked reduction in mobility-related transmission. The framework was further validated using synthetic data generated using the epidemic simulator, which confirmed that mobility-informed TE recovers the true network of infection spread, while case-only analyses introduce spurious dependencies. This analysis provided complementary evidence supporting the simulation outcomes, highlighting the directional influence of mobility on regional infection trends and validating the predictive value of mobility data for early outbreak detection [12].

These data-driven workflow thus complements CREXDATA's simulation pipelines by providing an empirical method for inferring causal structures and validating model assumptions. It enables the integration of observed mobility and epidemiological data into the decision-support environment, reinforcing the robustness and interpretability of epidemic use-case analyses.

3.1.3.3 Expert User Evaluation and Usability Results

Performance and Outcomes

The integrated demonstrators were successfully executed on CREXDATA's HPC infrastructure, confirming both scalability and computational efficiency (see Table 10). The *episim-emews* workflow reduced computational cost by characterizing the parameter space with up to 50% fewer simulations while maintaining calibration accuracy and it also supports finding effective vaccination campaigns.

Table 10. Summary of achieved KPIs for the Health Crisis Use Case.

KPI	Description	Completion
1	Forecasting 7 parameter sets that reduce the COVID infection	**100%**
2	Runtime adaptation of simulation trajectories	**100%**
3	Calibration of the epidemiological parameter to fit incidence time series	**100%**
4	Characterising the space of parameters with 50% fewer simulations	**100%**

The visual analytics tools effectively supported uncertainty assessment and interpretation of multi-scenario results, providing actionable insights into epidemic dynamics. In parallel, the reinforcement learning demonstrator proved the feasibility of automatically optimizing confinement strategies that balance epidemiological control and socio-economic impact. Overall, the outcomes demonstrate that integrating HPC-based simulation, AI-driven optimization, and visual analytics substantially enhances epidemic modelling and decision support. These results validate the toolbox as a powerful framework for data-driven preparedness against future health crises.

Evaluation of CREXDATA Objectives

A structured evaluation involving three expert users assessed the usefulness, relevance, and usability of the developed tools, particularly the *EpiSim.jl* simulator and the different workflows. Th expert users with backgrounds in epidemiological modelling, computational physics, and public health policy assessment evaluated how effectively the Health Emergency Use Case addressed the overarching objectives of CREXDATA.

Experts consistently rated the project’s main goals, real-time forecasting, dynamic modelling, predictive analytics under uncertainty, and visual explainability, as highly relevant and useful (average 4–5 out of 5). They emphasized that CREXDATA’s capacity to integrate multimodal data, perform large-scale dynamic simulations, and provide interpretable visual analytics directly supports data-driven decision-making in epidemic management. These capabilities were seen as essential for enhancing preparedness and improving the design of public health interventions.

The ability to perform large-scale scenario analysis (e.g. testing alternative vaccination strategies or non-pharmaceutical interventions) was regarded as a key strength. One expert noted that “the system has strong potential for policy and decision-making applications,” while another highlighted its value for modelling complex, multi-scale health systems. Less relevant components, such as augmented reality or multilingual data handling, were rated lower (average 2–3 out of 5), reflecting their limited applicability to epidemiological modelling. Overall, experts agreed that CREXDATA’s scientific and technical direction aligns well with real-world user needs, while recommending continued focus on usability and accessibility to ensure that its advanced analytics can be more broadly applied in operational contexts.

Evaluation of the Health Crisis Use Case – Epidemiological Scenario

The expert evaluation of the Health Emergency Use Case confirmed that the integrated demonstrators, combining epidemic simulation, HPC workflows, AI-driven optimization, and visual analytics, provide a coherent and scientifically solid framework for epidemic modelling and decision support. Experts validated the robustness of the simulator and associated workflows, recognizing their strong analytical foundation and effectiveness for model calibration, uncertainty quantification, and scenario exploration. These features were seen as particularly relevant for understanding the evolution of infectious diseases and evaluating policy responses under different assumptions and data conditions.

However, experts also noted that while the workflows are technically mature and powerful, their accessibility could be improved. Non-technical users such as public health officials would benefit from simplified configuration procedures, built-in templates, and guided setup options to reduce reliance on technical assistance. Despite these challenges, all evaluators acknowledged the system’s potential for broader application beyond epidemic crises, including the long-term simulation of chronic diseases or health system dynamics.

Overall, the evaluation demonstrates that the Health Emergency Use Case successfully integrates simulation, AI/ML, and visual analytics into an operational decision-support framework that meets scientific, technical, and policy-oriented objectives. While this scenario focuses on population-level epidemic forecasting, the complementary multiscale infection scenario extends these capabilities to patient-level modelling, together representing CREXDATA’s full spectrum of health crisis management.

System Usability Scale (SUS)

The System Usability Scale (SUS) evaluation at Month 36 assessed the usability of the developed toolbox, including the simulator, the workflows. The three experts' individual SUS scores were 65, 72.5, and 65, with an average score of 67.5, corresponding to Good/Acceptable usability according to standard benchmarks [13].

Compared with earlier evaluations, this marks a notable improvement, reflecting the positive impact of workflow modularization, standardized configuration files, and integrated visual analytics. Experts rated system integration highly (average 4.3/5) and expressed moderate-to-strong confidence in using it (3.3/5). They identified transparency, reproducibility, and scientific soundness as key strengths of the platform.

However, they also pointed out that usability remains a challenge for users without programming experience. The onboarding and configuration processes could be streamlined through graphical interfaces, embedded guidance, and ready-to-use templates. As one expert summarized, the system is *"the system has strong potential for policy and decision-making applications"*.

Overall, the *Epi-Sim* ecosystem demonstrates technical maturity and scientific rigor. Future development should focus on enhancing learnability and user guidance to ensure that its analytical power translates into accessible and actionable decision-support tools for diverse end users. Improving usability through guided workflows, embedded tutorials, and simplified configuration interfaces will be key to transitioning CREXDATA from a research-grade platform to an operational decision-support system.

3.2 Multiscale Infection Scenario

3.2.1 Introduction

The COVID-19 pandemic has revealed substantial gaps in global health system preparedness and emphasized the necessity for computational models that address disease mechanisms at multiple biological scales. Traditional epidemiological models enable tracking at the population level, but do not explain why COVID-19 manifests so differently across individuals, a challenge that requires mechanisms operating from cellular to organ level. The multiscale infection scenario in CREXDATA tackles this by integrating molecular, cellular, and tissue processes within patient-specific digital twins for SARS-CoV-2 lung infection.

This approach is vital, as the full spectrum of COVID-19 severity arises from complex interactions among viral characteristics, host immune responses, and individual variability. Mathematical modelling provides an important supplement to purely experimental work, by integrating multimodal data, omics, imaging, clinical records, into multifaceted simulations. Agent-based models, especially when combined with Boolean networks of intracellular signalling, provide detailed simulations of cell, virus, and immune interactions. Coupling these with fluid dynamics models of airflow yields insights into how ventilation and anatomy affect infection patterns and disease severity.

CREXDATA's scenario leverages two advanced simulators: PhysiBoSS, which replicates the infection and immune response in lung tissues while incorporating omics data for patient-specific modelling; and Alya, which simulates airflow and viral particle movement dictated by patient anatomy. Together, these allow a multiscale view from molecules to organs, facilitating the development of personalized intervention strategies, including both drugs and ventilation, based on digital twin models. This use case scenario was drafted in Deliverable 2.1 in M6 and its final status was described in Deliverable 2.2.

The integration of these models with CREXDATA's broader platform enables handling of extreme-scale data, real-time forecasting, and user-friendly visual analytics for clinicians and researchers. Results guide the optimisation of different treatments, supported by Altair's RapidMiner Studio for event forecasting and federated machine learning. CREXDATA's AI technologies enhance interpretability, predictive capability, and decision-making from large-scale data. The ultimate aim is to inform therapeutic and management decisions, optimize interventions before clinical application, and lay groundwork for rapid response to future pandemics.

3.2.2 Methods

3.2.2.1 Simulators

The multiscale infection scenario integrates two computational frameworks to capture the spatiotemporal dynamics of SARS-CoV-2 infection across biological scales.

Alya [14] is a high-performance computational mechanics code for solving complex multi-physics and multiscale problems, mainly in engineering. It handles flows, solid mechanics, heat transfer, particle transport, chemistry, and electrical propagation. Designed for massively parallel supercomputers, it uses advanced high-performance computing techniques for distributed and shared memory systems. Alya is proprietary and accessible via its official website.

In the present project, a patient-specific 3D lung geometry was reconstructed from computed tomography (CT) scans of a healthy adult, covering the trachea to the terminal bronchioles and including up to 14 airway generations with 128 outlets. The resulting mesh comprised approximately 4 million elements, refined towards the smaller bronchioles to capture complex flow structures. Airflow was simulated under an inspiratory flow rate of 20 L/min, using a Large Eddy Simulation (LES) approach with the Wall-Adapting Local Eddy-viscosity (WALE) model to resolve turbulent structures with optimal computational efficiency. These simulations were executed on MareNostrum-5 using 300 CPU cores and approximately 20 hours of wall-time, achieving a detailed and physiologically realistic velocity field across all lung lobes.

PhysiBoSS [15] is an open-source integration of stochastic Boolean and agent-based modelling frameworks for simulating intracellular signalling within individual cells. Version 2.0 extends PhysiCell to model cell signalling and microenvironment interactions, useful in studying disease dynamics and cell population behaviour. Its repository is available on GitHub.

In the multiscale infection scenario, PhysiBoSS models SARS-CoV-2 interactions within alveolar tissue, capturing viral entry, replication, immune activation, and cell fate decisions. Each cell is represented as an agent equipped with a Boolean network describing host–virus signalling pathways such as interferon, and apoptosis, whose probabilistic state transitions feed back into macroscopic behaviours like proliferation, death, and cytokine secretion.

For this project, PhysiBoSS was extended to better reproduce physiological infection conditions. Instead of random viral seeding, infections occur in controlled spatial rounds, each loading virion positions from a structured input file that mirrors Alya's deposition maps, allowing reproducible and iterative infection waves. An explicit oxygen field was also added to represent hypo-, hyper- and normoxia, dynamically influencing viral replication and cell fate via oxygen-dependent signalling transitions. Hypoxic areas promote viral persistence, while hyperoxic zones enhance oxidative stress and apoptosis. These additions create a mechanistic link between Alya's macroscopic airflow and PhysiBoSS's microscopic infection dynamics, enabling exploration of how oxygen heterogeneity, viral load, and immune responses jointly determine tissue outcomes.

EMEWS [10], another open-source framework, enables large-scale exploration and optimization of parameters in multiscale models using HPC resources. Within this project, it explores parameters in PhysiBoSS simulations to improve model accuracy and identify optimal conditions for airflow and COVID-19 infection modelling, supporting better treatment and prevention strategies.

3.2.2.2 Workflows

The multiscale infection scenario integrates user-centric visual workflows with secure, distributed HPC orchestration to democratize access to extreme-scale simulation technologies. Implemented in Altair RapidMiner Studio, two complementary workflow classes operationalize the end-to-end simulation pipeline: local workflows enable biomedical researchers to configure and explore PhysiBoSS simulations through intuitive drag-and-drop interfaces with real-time monitoring via Kafka, early time-series classification for rapid outcome prediction, and interactive intervention testing without requiring HPC access. Remote HPC workflows extend this capability to coupled Alya-PhysiBoSS simulations by connecting securely through a dedicated virtual machine gateway that manages authentication, file system mounting, and SLURM job orchestration, thereby transforming RapidMiner into a transparent front-end for high-performance distributed computing. This architecture bridges the gap between exploratory local modelling and large-scale digital twin simulation, enabling non-expert users to launch extreme-scale analyses while enforcing robust security and computational efficiency.

The simulator coupling between Alya and PhysiBoSS implements a bidirectional, iterative feedback mechanism linking macroscopic airflow dynamics with microscopic cellular infection and immune response processes (Figure 24). First, Alya simulations compute patient-specific airflow and nanoparticle deposition patterns across lung lobes. Second, these deposition results initialize five lobe-specific PhysiBoSS simulations that model epithelial and immune cell dynamics, viral replication, and alveolar damage at the cellular scale. Third, alveolar health status is then fed back to parameterize the next Alya round, adjusting airflow and particle distribution to reflect ventilation impairment in damaged lobes, creating a dynamic closed-loop interaction.

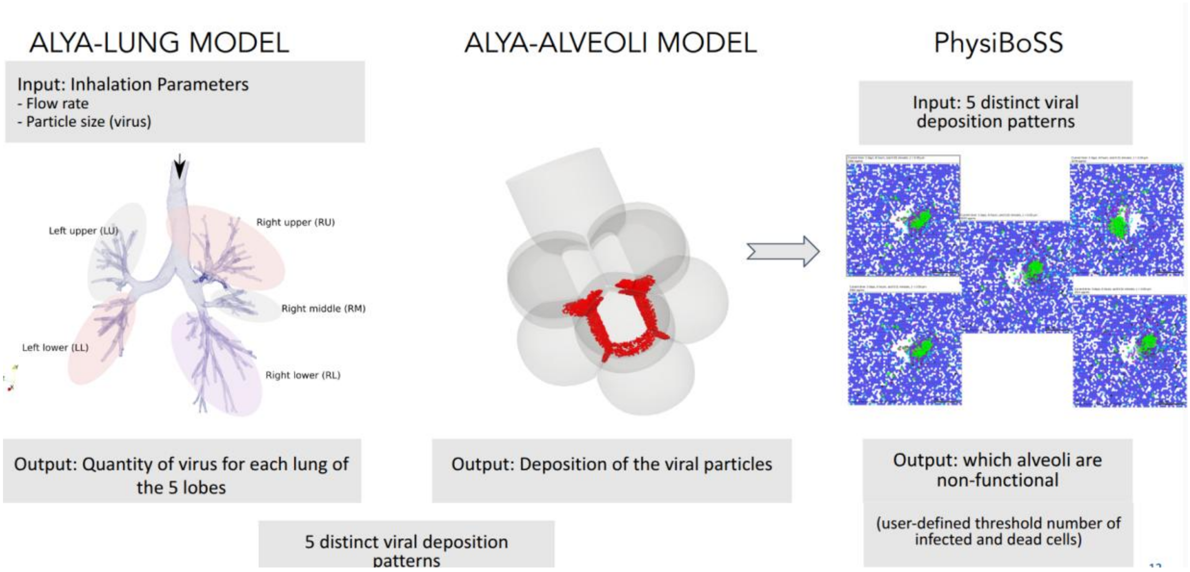


Figure 24: Schema of the workflow connecting the simulators. The Alya-Lung model estimates viral load across the five lung lobes, which is then downscaled by the Alya-Alveoli model into five viral deposition patterns. These patterns seed PhysiBoSS simulations that determine which alveoli become non-functional based on infection and cell death thresholds.

The EMEWS framework orchestrates large-scale parameter exploration across patient phenotypes (mild, severe, vaccinated), drug treatments, and vaccination status to identify biologically meaningful thresholds governing alveolar state transitions. The feasibility of this iterative approach is enabled by the tools' computational performance: conducting-zone simulations require approximately 20 hours on 300 cores, while alveolar-zone simulations complete in under 20 minutes, ensuring interactive learning and real-time forecasting can operate on high-resolution multiscale models without compromising HPC runtime viability.

Architecture

RapidMiner Studio operates as a local desktop application, ideal for workflow design, data visualization, and user interaction. However, this local runtime environment limits direct execution of large-scale coupled simulations. Consequently, only PhysiBoSS-only simulations can be run locally for rapid prototyping or parameter testing. In contrast, Alya-coupled simulations, which require distributed parallel computation, are executed remotely on MareNostrum5 (MN5) within the BSC infrastructure.

To unify these two operational environments, the RapidMiner workflows were extended to communicate with a dedicated virtual machine (VM) deployed within the BSC network. This VM serves as a secure gateway between the RapidMiner environment and the HPC infrastructure. It mounts the MareNostrum file system, manages authentication, and handles controlled job submission and result retrieval. A custom orchestrator ensures that each submitted job is securely executed through the SLURM scheduler, with progress and output logs transmitted back to RapidMiner in real time.

This architecture effectively abstracts the complexity of the HPC backend, providing a low-barrier access point for ad-hoc and non-coding users who need to reproduce or launch large-scale simulations without manual command-line interaction. It also addresses security requirements, ensuring that external user workflows can interact with MareNostrum resources through a controlled intermediary layer rather than direct cluster access.

Together, these workflows extend the interactive learning and runtime adaptation framework described in Deliverable 4.2 (Interactive Learning for Simulation Exploration) by adapting it to the Health Crisis Use Case. The result is an enhanced, automated, and accessible infrastructure that connects local experimentation with remote HPC execution, ensuring a seamless transition between small-scale model exploration and full-scale digital twin simulations.

Local RapidMiner Workflows for PhysiBoSS-Only Executions

The first group of workflows operates entirely within RapidMiner Studio and is intended for local experimentation. These workflows enable parameter exploration, monitoring, and forecasting for PhysiBoSS simulations, allowing users to prototype infection scenarios before scaling them up to HPC execution.

a) XML Editing and Simulation Launch

The following workflow provides a configurable interface for setting up PhysiBoSS experiments. Using dropdown menus, users can select the patient type (mild, severe, vaccinated, wild type), adjust viral load or infection rate, choose between epithelial or macrophage cell types, and toggle Boolean signalling models or gene knockouts (e.g., MAPK14, AKT1, FADD, P38) (Figure 25).

Once the parameters are selected, the workflow automatically edits the PhysiBoSS XML configuration and launches the simulation locally. The user can also decide whether they desire to use KAFKA to stream the results, select the broker and the topics that the PhysiBoSS simulation will be listening to. This enables fast, small-scale tests to verify model behaviour or test biological hypotheses without accessing HPC resources.

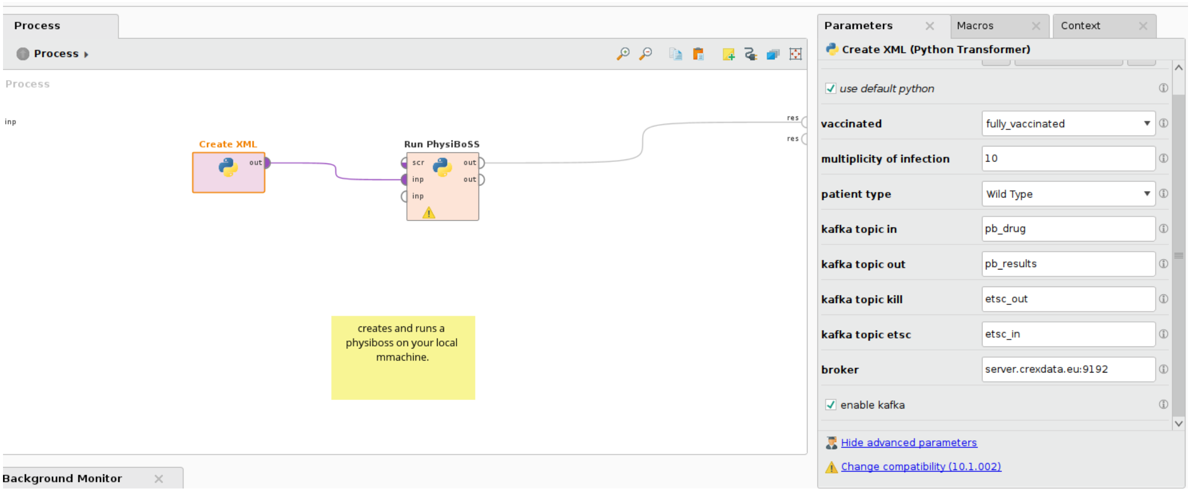


Figure 25: RapidMiner Workflow that initiates a PhysiBoSS simulation.

b) Monitoring via Kafka Integration

This workflow is the responsible for communicating the simulation statistics to the user by enabling real-time streaming of simulation outputs. Users can subscribe to specific Kafka topics corresponding to active simulations and visualise evolving data such as counts of infected, apoptotic, or necrotic cells, and microenvironmental variables (oxygen, virion, cytokine concentrations).

This workflow illustrates how data-streaming pipelines can be integrated into visual analytics environments, allowing users to monitor model behaviour without direct access to log files or cluster consoles (Figure 26).

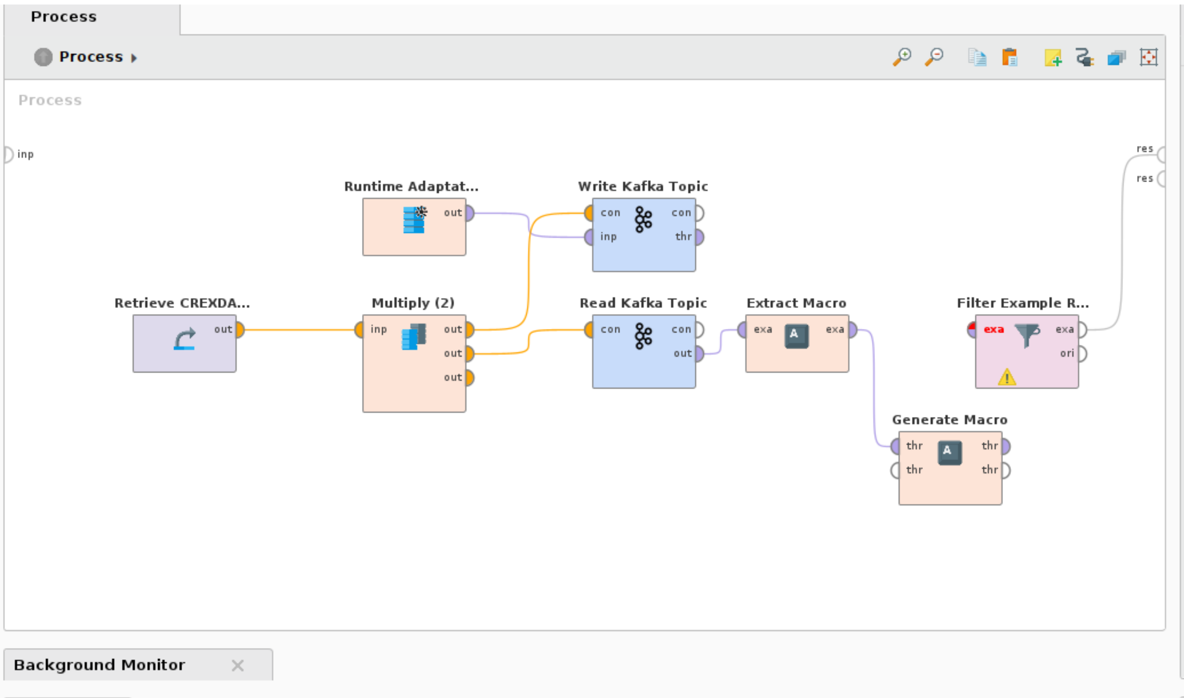


Figure 26: RapidMiner Workflow that analyses a simulation and allows for online monitoring.

c) *Early Time-Series Classification (ETSC) Prediction*

This workflow implements the Early Time-Series Classification model trained on prior PhysiBoSS simulation data. By analysing the first portion of the trajectory, ETSC predicts the likely infection outcome, efficient or inefficient alveolar recovery, before the simulation completes.

d) *Intervention Testing (as introduced in Deliverable 4.2)*

This workflow corresponds to the interactive intervention and re-execution pipeline originally presented in Deliverable 4.2 – Interactive Learning for Simulation Exploration allowing users to introduce drug interventions directly from the RapidMiner interface.

Remote HPC Workflow via the BSC Virtual Machine (Alya-PhysiBoSS Coupling)

While RapidMiner Studio can handle PhysiBoSS simulations locally, the coupled Alya-PhysiBoSS digital twin requires large-scale computation that cannot be executed on local machines. To enable this within the same workflow environment, the project integrated RapidMiner with the VM infrastructure, which provides a secure interface between the external user environment and the MN5 supercomputer.

This design was driven by both computational requirements and security constraints: direct external access to MN5 is restricted, so the VM serves as an intermediary layer that authenticates users, mounts shared directories, and manages job submission.

Secure HPC Orchestration

The central workflow connects to the BSC VM, which runs a dedicated orchestrator responsible for:

1. Establishing a secure SSH session and mounting the MN5 file system;
2. Submitting the Alya-PhysiBoSS coupled simulation job via the SLURM scheduler;
3. Tracking job identifiers and runtime status;
4. Returning results to RapidMiner for visualization and follow-up actions by using Kafka connections

This architecture transforms RapidMiner from a local modelling tool into a front-end for secure HPC orchestration, enabling users to launch extreme-scale digital twin simulations without direct cluster access. In this mode, users can select the relevant PhysiBoSS parameters and alveolar compartment from the RapidMiner interface. The workflow mounts the corresponding XML configuration on the VM, triggers the HPC job, and collects the output once execution is complete.

3.2.2.3 Data

The multiscale infection scenario leverages diverse data sources spanning publicly-available molecular and cellular data, patients' CT scans of lungs and clinical data, as described in D2.1 and D2.2.

The molecular data with which we personalised the models came from a publicly-available dataset that also had clinical information [16]. These data was analysed and integrated into the models using our PROFILE tool [17], [18]. The pulmonary 3D positional data patients' CT scans were used to have diverse geometries to test Alya's simulations. The molecular interactions databases used to inform the cellular interaction networks were obtained from publicly-available resources such as SIGNOR [19].

3.2.3 Results and evaluation

3.2.3.1 Pilot

The Multiscale Infection Pilot was developed to demonstrate the integration of macroscopic and microscopic simulations within a unified biomedical digital twin of the human lung. Using COVID-19 as a reference case, the pilot aims to showcase how airflow dynamics and cellular infection models can be coupled to capture the complex, multiscale processes governing disease progression and immune response in the respiratory system.

The main goal of this pilot is to provide a computational framework that merges mechanistic modelling, agent-based simulations, high-performance computing (HPC), extreme-scale data transfer, and online AI-driven interactive learning, enabling the exploration of infection dynamics and treatment strategies under uncertainty.

At M36, the pilot demonstrates feasibility, scalability, and interoperability of a bidirectionally coupled Alya-PhysiBoSS digital twin of the lung. In this setup, airflow simulations computed by Alya determine the spatial distribution of viral deposition, which initializes corresponding PhysiBoSS simulations of epithelial infection and immune response. The resulting tissue alterations can then be fed back to Alya to modify local airflow and diffusion dynamics, closing the loop between the macroscopic and microscopic scales. This coupling served as a proof of concept for the digital-twin integration, enabling continuous transfer of deposition data from the macroscopic Alya domain to the microscopic PhysiBoSS cellular model.

Through this pipeline, this pilot validates the CREXDATA vision of merging physics-based, agent-based, and AI-driven approaches into a single orchestrated environment capable of supporting predictive, interpretable, and adaptive health-modelling workflows.

3.2.3.2 Demonstrators

The demonstrator of the multiscale infection scenario showcases how the CREXDATA infrastructure combines advanced multiscale simulation technologies with accessible workflow interfaces that can be operated by non-expert users. A suite of interconnected workflows was implemented in Altair RapidMiner Studio to provide an intuitive and modular environment for configuring, launching, monitoring, and analysing both PhysiBoSS-only and Alya-PhysiBoSS coupled simulations.

Forecasting parameter sets that reduce the COVID infection

A comprehensive parameter sweep was performed to evaluate how immune preparedness and intracellular regulatory logic shape the ability of an alveolus to control SARS-CoV-2 infection. More than 5,000 PhysiBoSS simulations were executed through the CREXDATA multiscale workflow, systematically varying vaccination-like immune parameters, intracellular Boolean models for epithelial and macrophage cells, environmental oxygen levels, and the presence of viruses (termed multiplicity of infection or MOI). To ensure a consistent comparison, the analysis focused on a well-defined and highly informative region of the parameter space, selecting all simulations executed under identical conditions (O_2 = 100, MOI = 100). This subset contained sufficient variation in intracellular signalling and immune configurations to reveal clear determinants of infection control and tissue preservation.

Across this controlled setting, the combined effect of vaccination parameters and intracellular logic produced three distinct phenotypic regimes. Fully vaccinated simulations consistently achieved *complete suppression of infection*, retaining nearly all epithelial cells, while mildly vaccinated simulations exhibited strong protection with negligible infection and high tissue viability. In contrast, non-vaccinated simulations displayed a broad distribution of high infection levels, with final infected epithelial ratios ranging from 20% to over 30%. Importantly, this variability was strongly modulated by the state of the intracellular Boolean models: non personalised epithelial–macrophage logic maintained partial antiviral capability even without vaccination, whereas severe signalling networks, particularly the Severe patient and the Severe patient lacking p38 functional gene variants, produced the highest viral burdens. These findings demonstrate that both vaccination status and intracellular model outcomes jointly determine the infection outcome, with strong nonlinear interactions between the two scales.

The synthesis of these results is captured in Figure 27, which maps the mean infected epithelial ratio across all combinations of vaccination-like immune settings and intracellular logic. Fully vaccinated and waning vaccination scenarios eliminate infection completely across all models, while non-vaccinated outcomes depend critically on signalling logic. The worst outcomes arise from a Severe patient's intracellular pathways combined with the absence of immune priming, whereas non-personalized epithelial and macrophage logic offers partial intrinsic protection even in a non-vaccinated context. Altogether, the parameter sweep robustly identifies the regions of the parameter space where infection is minimised, fulfilling the use case's KPI 1 on reducing COVID-19 infection. These results validate the multiscale workflow’s capability to differentiate functional and dysfunctional alveolar responses and highlight mechanistically meaningful combinations of immune and intracellular parameters that maximise tissue-level protection.

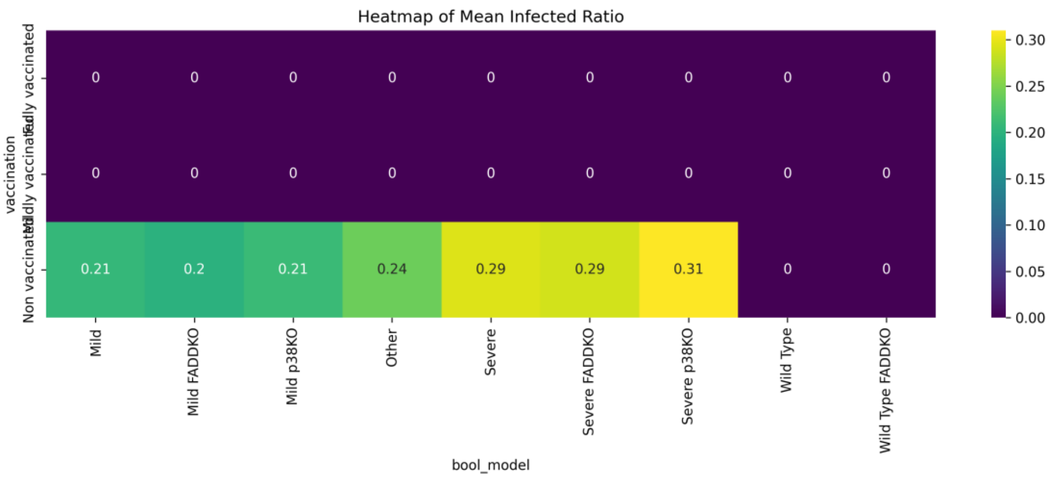


Figure 27: Mean infected epithelial ratio across vaccination and Boolean signalling configurations (O_2 = 100, MOI = 100). Rows correspond to vaccination status (non-vaccinated, waning vaccination, fully vaccinated), while columns represent the epithelial–macrophage intracellular logic used in each simulation (The Mild and Severe models represent personalised epithelial–macrophage signalling logics corresponding to patients with mild or severe infection, the Wild Type is the non-personalised baseline network, and the FADDKO/p38KO variants introduce targeted gene knockouts to mimic drug-like inhibitory interventions). Colour indicates mean infected epithelial ratio. Fully and mildly vaccinated configurations achieve infected ratio = 0 across all Boolean models. In non-vaccinated simulations, infection increases substantially for severe and knockout intracellular networks, while wild-type logic retains partial antiviral protection. This heatmap identifies which combinations of immune and intracellular parameters minimise COVID-19 infection.

Runtime adaptation of simulation trajectories

To demonstrate the CREXDATA platform’s capacity for real-time intervention, we implemented a closed-loop system that couples multiscale infection simulation with online analysis and response capabilities. Using the RapidMiner's AI Studio pipeline in conjunction with a Kafka event stream listener, we monitored outputs from active simulations in real time (identified by their *sim_id*).

The ETSC algorithm, developed in D4.2, evaluated these outputs continuously to identify critical or actionable states. Upon detection, a control message was sent back to the simulator to execute a specific biological intervention, in this case, a gene knockout as a surrogate of a drug targeting the protein coded by that gene.

As illustrated in Figure 26, the user interface allows for interactive monitoring of cell state trajectories over time. In simulation *sim_2*, the ETSC module detected a condition requiring modulation and triggered the administration of a drug targeting MAPK14 protein at timestep 270. The system then executed the intervention, demonstrating the ability to influence biological outcomes dynamically and stopped the infection of more cells. Interestingly, there are commercially available drugs targeting MAPK14 such as dasatinib [20] and fostamatinib [21].

With this looped decision-support system we contribute to KPI 2, where we aimed at having a mechanism for the runtime adaptation of simulation trajectories to improve outcomes in real-time, and further validates the usability of CREXDATA for expert-driven, exploratory control of multiscale processes.

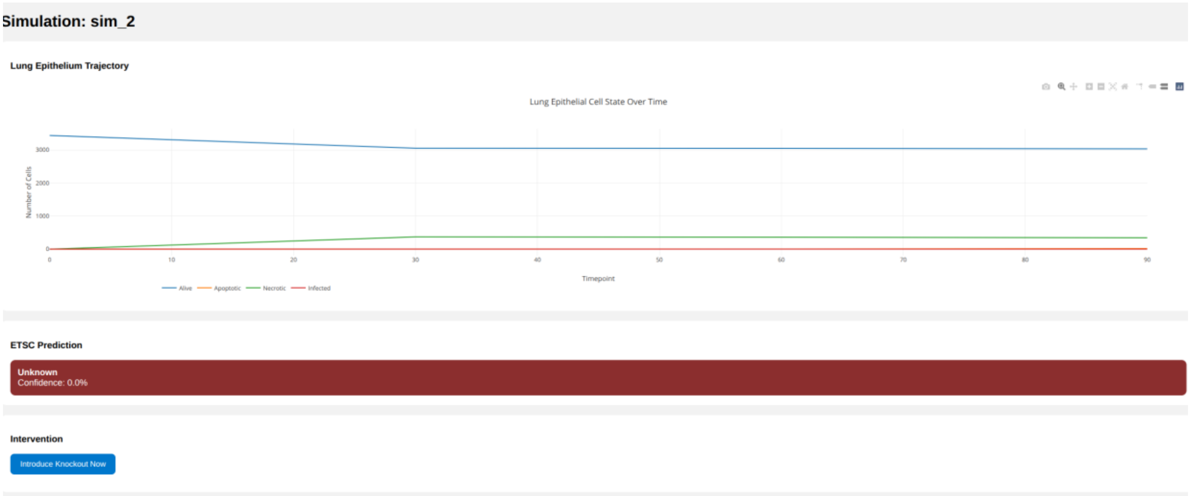


Figure 28: Live simulation interface showing lung epithelial cell states over time. Colours meaning are: Alive in blue, Apoptotic in red, Necrotic in black, Infected in green. Upon ETSC evaluation, a gene knockout intervention is triggered in real time (MAPK14 at t=270) to modulate infection progression.

To further assess the effectiveness of the runtime adaptation mechanism, we applied the MAPK14 knockout intervention across a broad parameter sweep covering different immune configurations and physiological settings (Figure 28). For each simulation where the intervention was triggered, we matched it to its corresponding non-intervention baseline using the complete PhysiCell parameter vector, ensuring a like-for-like comparison. We then evaluated the epithelial alive ratio at the final simulation time point, which serves as a direct proxy for alveolar integrity and overall infection severity.

Across these explorations, we identified several scenarios where the runtime MAPK14 knockout produced a measurable improvement in epithelial survival compared to the corresponding baseline trajectory. These improvements were observed across fully vaccinated, mildly vaccinated, and non-vaccinated profiles, demonstrating that the intervention can influence outcomes across a spectrum of immune backgrounds. Four representative examples are presented in Table 11. In all of them, the alive ratio at the end of the simulation was markedly higher in the MAPK14 KO trajectory than in the non-intervention equivalent, indicating that the intervention successfully redirected the infection progression toward a more favourable state.

Table 11. Results of the different simulation scenarios. Scenarios differ in vaccination status and interventions on the Boolean models. Alive Ratio is the variable of study and is characterised by counts of alive cells in the different scenarios compared to the no-intervention simulations.

Scenario	Vaccination Status	Baseline Alive Ratio	KO Alive Ratio	Alive Ratio	Outcome interpretation
Scenario A	Fully vaccinated	0.697	0.980	**+0.283**	Epithelial preservation markedly improved following MAPK14 inhibition.
Scenario B	Fully vaccinated	0.476	0.703	**+0.227**	Substantial increase in alveolar viability under MAPK14 knockout.
Scenario C	Waning vaccination	0.695	0.960	**+0.265**	Runtime intervention significantly shifts the trajectory under moderate immune preparedness.
Scenario D	Non-vaccinated	0.677	0.929	**+0.252**	Noticeable improvement even in immune-naïve conditions.

3.2.3.3 Expert User Evaluation and Usability Results

The month 36 evaluation of the CREXDATA Life Science use case demonstrates substantial improvements across key performance dimensions compared to the month 18 assessment. Three expert users evaluated the system's progress toward project objectives, the implemented health crisis use case capabilities, and overall system usability. The results reveal a multifaceted evolution of the platform, with notable enhancements in user interface integration and workflow capabilities, though challenges remain in accessibility for non-technical users.

Performance and Outcomes

The multiscale infection scenario within CREXDATA has successfully achieved the key performance indicators (KPIs) defined at the outset of the project. The integrated computational framework combining PhysiBoSS for agent-based infection and immune dynamics in lung epithelia and Alya for patient-specific airflow and viral particle transport has provided accurate simulations across multiple biological scales.

The scenario has demonstrated the ability to forecast more than 7 parameter sets that reduce infection severity (KPI 1), use the runtime adaptation of simulation trajectories to improve the outcomes of more than 5 scenarios or patients of the multiscale infection scenario (KPI 2) and characterizing the space of parameters with 62.5% fewer simulations (KPI 4) (Table 12).

These achievements show that CREXDATA technologies support clinical decision-making under uncertainty. The models have shown alignment with patient-specific data and clinical markers, underpinning their utility as personalized digital twins for COVID-19 management. This success paves the way for further refinements and potential extension to other infectious diseases, ultimately contributing to improved patient care and pandemic preparedness.

Table 12. Summary of achieved KPIs for the Health Crisis Use Case.

KPI	Description	Completion
1	Forecasting 7 parameter sets that reduce the COVID infection	**100%**
2	Runtime adaptation of simulation trajectories	**100%**
4	Characterising the space of parameters with 50% fewer simulations	**100%**

Evaluation of CREXDATA Objectives

The three expert evaluators provided quantitative assessments of the project's core objectives across nine discrete technical components. The evaluations revealed strong consensus on several strategic pillars:

Extreme-scale data ingestion and exploitation received consistently high ratings, particularly for multimodal data ingestion and dynamic modelling. All three evaluators rated "Using dynamic modelling to predict the systems' behaviour" as 5 (Very useful), indicating alignment between project capabilities and user needs. Multimodal data ingestion was rated 5 by two evaluators, and 4 by another. However, multilingual social data handling received lower ratings (1-4 range), suggesting this specific component has limited perceived relevance to the expert user cohort.

Real-time predictive knowledge and forecasting demonstrated variable relevance depending on user context. Optimization techniques for Prediction-as-a-Service (PaaS) received ratings of 5, 4, and 2, reflecting differential applicability across user backgrounds. Complex event forecasting under uncertainty was rated highly by one user (5) but lower by another (2), indicating that the practical utility of this capability depends on specific research domains.

Reducing perceived complexity emerged as a particularly critical objective. Graphical workflow design received ratings of 5, 4, and 2, demonstrating that while visual workflow interfaces represent a significant advancement, their effectiveness varies substantially by user technical background. Visual analytics coupled with explainable artificial intelligence was rated as 5 and 3, suggesting this feature requires continued refinement.

Evaluation of the Health Crisis Use Case, Multiscale infection scenario

The health crisis use case evaluation demonstrated notably strong and consistent endorsement across multiple technical components. All three expert users awarded maximum or near-maximum ratings for the foundational platform features:

Graphical user interface received ratings of 5 and 4, confirming that the transition from the month 18 partial implementation to a more comprehensive visual interface has addressed one of the primary identified mitigation strategies.

Parameter calibration functionality achieved uniform 5-point ratings across all three evaluators, indicating that this capability successfully meets expert user requirements. This represents a critical capability for systems biology modelling, as parameter fitting is essential for predictive accuracy.

Online model exploration was highly rated by experts (5 and 4), confirming that users value the ability to interactively investigate model behaviour during execution.

Real-time/online forecasting simulation trajectories received ratings of 5 and 3, reflecting strong endorsement from computationally sophisticated users but more cautious assessment from another user profile.

The Key Performance Indicators (KPIs) targeting specific multiscale infection scenario objectives received positive evaluation. KPI 1 (forecasting parameter sets for intervention optimization) achieved ratings of 5 and 4. KPI 2 (Use the runtime adaptation) was rated 5, 4 and 3. KPI 4 (parameter space characterization with 50% computational reduction) was rated 5 and 4. Note that KPI 3 is specific to the epidemiological scenario.

Multiscale simulation capabilities constituted a distinctive strength. The ability to seamlessly launch Alya-PhysiBoSS simulations directly from the workflow received universal 5-point ratings, reflecting the project's successful integration of previously decoupled computational tools. This technical achievement directly addresses the month 18 strategy of creating an integrative system of different CREXDATA technologies.

System Usability Scale (SUS)

The month 36 System Usability Scale evaluation reveals a stratified pattern of user experience marked by substantial divergence in perceived usability across the three expert evaluators. Two expert users (score: 60.00 and 62.50) provided assessments indicating acceptable, though not exceptional, system usability, while another user (score: 52.50) reported some usability challenges. The two more positive users' evaluations identify the platform's principal strengths as system consistency (both rated the inconsistency as minimal), well-integrated functions (4/5), and the absence of excessive cumbersome design (both rated cumbersome interface at 2/5 or lower). These users, both with advanced computational modelling expertise, demonstrated moderate confidence in system use despite acknowledging the requirement for technical support. Conversely, the other user's substantially lower score reflects critical weaknesses in accessibility and learnability: he/she reported that the system is unnecessarily complex (rated 2/5 on ease of use) and would require technical support (rated 5/5), indicating that the platform remains inaccessible for users without strong computational backgrounds. The incomplete responses from he/she on 3 of ten SUS questions further suggests potential confusion or difficulty.

The System Usability Scale assessment reveals mixed but generally positive trends relative to the month 18 baseline of 39.167. The month 36 evaluations yielded three distinct SUS scores: 60.00, 62.50, and 52.50. The aggregate mean SUS score across the three expert users is 58.33, representing a 19.16-point improvement from the month 18 baseline. This improvement reflects the project's progress in addressing the identified mitigation strategies, although the average score remains below the 68-point threshold considered acceptable for above-average usability.

Nevertheless, the aggregate improvement from the month 18 baseline demonstrates meaningful progress, yet the heterogeneity of responses underscores the platform's limitations regarding universal accessibility, particularly for domain experts without computational training. This usability profile indicates that the system successfully serves computationally sophisticated model developers, though more training needs to be done to enable use by people lacking computational backgrounds.

3.3 Lessons learned

3.3.1 Epidemiological Scenario

The evaluation of the Health Crisis Use Case, combining pilot demonstrators and expert user feedback, revealed key lessons for the continued development and operationalisation of CREXDATA's tools for managing health crises. Experts confirmed the strong analytical and computational foundation of the simulator and model exploration workflows, highlighting their effectiveness for parameter calibration, scenario exploration, and uncertainty analysis. These capabilities position the platform as a valuable decision-support resource for public health authorities responding to emerging infectious disease outbreaks.

At the same time, while the underlying workflows (*episim-emews* and *episim-rl*) are scientifically and technically mature, the evaluation identified the need for enhanced usability and accessibility. Specifically, experts recommended developing a more intuitive user interface, streamlined onboarding, and guided configuration options to support users without technical backgrounds. Future development should therefore incorporate user-centred design principles, including default templates, step-by-step setup wizards, and contextual help, to reduce reliance on expert assistance. Iterative usability testing and continuous engagement with end-users will be essential to ensure that technical innovation aligns with operational needs.

Additionally, while focusing on COVID-19 provided a rich and data-driven test case, it also limited exploration of other disease types. As noted by one expert, the framework could be extended to simulate a wider range of health challenges, including vector-borne or chronic diseases. Expanding the modelling capabilities and integrating additional simulation engines would significantly enhance the system's versatility and long-term impact.

Overall, the Health Crisis Use Case successfully demonstrated how HPC, AI, and visual analytics can be integrated into a unified, data-driven decision-support framework for epidemic management. The next development phase should focus on improving user experience, interactivity, and model generalisation, ensuring that CREXDATA's scientific capabilities translate into accessible and actionable tools for public health preparedness and crisis response.

3.3.2 Multiscale Infection Scenario

The development and evaluation of the multiscale infection scenario revealed several methodological and operational insights that shaped the maturation of the demonstrator. First, the integration of agent-based models with mechanistic intracellular signalling highlighted the importance of clearly defining model interfaces: small inconsistencies in variable scaling, lattice resolution, or unit conventions had disproportionate effects on downstream infection and immune-response dynamics. Establishing strict harmonization rules for model coupling significantly improved numerical stability and interpretability of the results.

Second, results showed that high-dimensional parameter spaces—particularly those involving viral kinetics, immune activation thresholds, and tissue-level oxygen constraints—often exhibit compensatory behaviours that generate similar macroscopic trajectories. This non-identifiability required the systematic use of interactive learning, uncertainty quantification, and visual analytics to detect robust vs. fragile regions of the parameter space. These tools proved essential to characterizing outcomes and identifying biologically meaningful intervention levers.

Third, the scenario demonstrated that runtime adaptation of simulations is feasible but sensitive to the selection of early-time biomarkers. The effectiveness of adaptive strategies depended heavily on which early signals (infection load, epithelial viability, cytokine levels) were monitored, and how rapidly these indicators diverged under different parameterizations. This informed the refinement of early-forecast metrics and improved the design of the adaptive workflow.

Finally, expert evaluation emphasized the need for clearer visualization of multi-scale outputs. Representing relationships between viral load, cell fate transitions, tissue function, and intervention effects required visualization techniques that communicate uncertainty, spatial heterogeneity, and temporal dynamics simultaneously. Improvements in this area increased the usability of the demonstrator for both modelling experts and domain scientists. Collectively, these insights will guide the consolidation of the multiscale scenario and support the definition of stable, reproducible workflows for future rounds of simulation-based decision support.

Regarding the evaluation by expert users, the month 36 evaluation demonstrates significant progress in the platform development. The successful integration of previously independent simulation tools and adoption of visual workflow interfaces represent meaningful technical achievements validated by expert evaluation. Computational model developers and experienced researchers provided consistently positive assessments of the platform's capability for multiscale modelling and extreme-scale data handling. The 19.16-point improvement in System Usability Scale scores from month 18 (39.167) to month 36 (58.33 average) reflects genuine advancement through targeted technology integration and workflow visualization. Key lessons include the importance of domain-specific workflow templates for reducing cognitive load while maintaining flexibility, the demonstrated value of abstracting high-performance computing resources through intuitive graphical interfaces, and recognition that continued refinement of user interfaces will enhance adoption across diverse researcher backgrounds. The evaluation confirms that deliberate mitigation of identified weaknesses produces measurable improvements, supporting the technical direction and implementation strategy pursued throughout the project's second half.

In parallel, the implementation and validation of the multiscale infection scenario revealed several methodological and architectural lessons that extend beyond usability. The coupling of Alya and PhysiBoSS highlighted the need for robust orchestration mechanisms capable of coordinating heterogeneous simulation scales, secure HPC execution, and iterative data exchange without compromising performance. The use of early trajectory forecasting (ETSC) demonstrated that surrogate analytical models are essential for accelerating exploration of expensive multiscale configurations, enabling researchers to identify promising intervention strategies before full simulation completion. The introduction of runtime intervention loops further showed the feasibility of adaptive modelling, while underscoring the requirement for well-defined intervention points when working with simulations that cannot support continuous real-time steering. These experiences confirmed that modular workflow design, progressive automation of configuration steps, and clear separation between user-facing logic and HPC-level execution are critical for ensuring scalability, reproducibility, and long-term maintainability of multiscale digital-twin applications.

4 Maritime Use Case

4.1 Introduction (Scenario description)

The maritime use case involves testing components from the CREXDATA platform and related algorithms for forecasting hazardous events like collisions or hazardous weather. These forecasts help reroute vessels to safety. This research impacts various stakeholders including Vessel Traffic System operators deck officers and crew coastal authorities' remote operators and others as outlined in D2.1. With the rise of autonomous ships and related technologies these components are becoming increasingly crucial especially for remote operators. The Maritime Use Case aims to enable early detection and forecasting of maritime events allowing for proactive measures to prevent incidents.

In the context of this use case, the following objectives will be pursued:

a) Development of a novel IoT device to access real-time data from the "black box" of a vessel.
b) Fusion of vessel data with global data sources (e.g., AIS and Copernicus data) to create reliable digital twins of vessels.
c) Development of highly scalable route forecasting algorithms.

The end-users have the capability to view forecast vessel motion and predicted routes, along with confidence levels for each route. Key performance indicators (KPIs) were achieved, exceeding the 80% accuracy threshold in route forecasting/weather routing and hazardous event detection/forecasting, achieving sub-second latency in streaming data, and forecasting maritime hazardous weather-related events 15 minutes before they occur.

The pilot was validated through a successful sea trial experiment. The purpose of this use case was to develop a comprehensive solution that combines hardware and software development to utilise vessel data, fused with global data, to create reliable digital twins of vessels. Additionally, it aimed to develop weather and emergency routing solutions that can be performed for all vessels of a fleet simultaneously, leveraging big data and AI technologies.

Early in the project, it was decided to use the autonomous vessels from the Ro-Boat Race as the primary testbed, offering a realistic yet safe maritime environment for validating new components. This allowed the team to test technologies under real sea conditions while maintaining full control over the operational setting. The IoT "black box" device developed for the Maritime Use Case was designed to be fully NMEA-2000-compatible[4], mirroring the behaviour and data interfaces of standard vessel systems. The standard describes a low-cost moderate capacity bi-directional, multi-transmitter/multi-receiver instrument network to interconnect marine electronic devices. Equipment designed to this standard will have the ability to share data, including commands and status with other compatible equipment over a single channel. Before broader deployment, the device was seamlessly tested on our 6.5-metre high-end RHIB "AIGAIO I"[5], confirming stable communication, real-time data access, and full compatibility with professional-grade onboard electronics. This approach reduced

[4] https://www.nmea.org/nmea-2000.html

[5] https://www.marinetraffic.com/en/ais/details/ships/shipid:9827754#overview

integration risks and ensured a smooth transition toward multi-vessel pilots during the Aegean Races.

Figure 29. The "AIGAIO I" RHIB of University of the Aegean in action.

The development and validation of the Maritime Use Case builds on the experience of the Ro-Boat Race [22], an annual autonomous vessel competition organised in the Aegean Sea. The first Ro-Boat Race established a unique real-world testbed where student-built autonomous surface vessels operated in challenging marine conditions, completing navigation, obstacle-avoidance, and mission-execution scenarios. This event provided an invaluable foundation: it created a safe but realistic environment for experimenting with multi-vessel communication, sensor integration, remote operations, and autonomous decision-making. Over the years, the Ro-Boat Race[6] evolved into a mature, reproducible, and instrumented maritime experimentation framework. Its rich telemetry, operational diversity, and the presence of multiple vessels in a confined but dynamic maritime space made it an ideal early sandbox for testing CREXDATA technologies before transitioning into large-scale pilot activities.

[6] https://smartmove.aegean.gr/events/1st-aegean-roboat-race/

Building on this foundation, the Maritime Use Case adopted a three-phase pilot progression, each phase tested during the Aegean Races, allowing for incremental validation of technologies under real operational maritime conditions. Pilot 1 leveraged the Ro-Boat Race methodology and focused exclusively on achieving reliable communication and data exchange with vessels at sea. During this phase, the project validated the operation of the novel IoT device, real-time extraction of “black box” data, and stable streaming of AIS and onboard sensor information to the CREXDATA platform. Pilot 2 advanced from passive data acquisition to active interaction, testing the system’s capability to send commands, alerts, and routing recommendations back to vessels. This enabled evaluation of latency, system responsiveness, message reliability, and crew/remote-operator workflows within a controlled but authentic maritime environment. Pilot 3 represented the first full-scale integration test, combining all major components—digital-twin creation, multi-vessel route forecasting, hazardous-event prediction, weather-aware routing, and AR visualisation—across several vessels simultaneously participating in the Aegean Races.

This structured progression—from the foundational lessons of the first Ro-Boat Race, through three increasingly complex pilots—ensured that each subsystem was validated rigorously and safely. By the time of the final sea-trial experiment, the project had tested all hardware and software components under escalating levels of operational complexity, ensuring robustness, resilience, and end-user confidence in real-world maritime environments.

During the initial phase of the pilot, various components were tested to ensure reliable connectivity. Tools were developed to access on-board telemetry from the motion controller in real time via the IoT device. Tests were conducted to confirm that the IoT device could safely process the required volume of logs on edge and reliably transfer them through the internet when offshore. Onshore, the necessary infrastructure and control room were established to support real-time logging of multiple Autonomous Surface Vehicles (ASVs) and sensor detections through Kafka topics. At the conclusion of the first phase, mission plans were either hard-coded or manually transferred through remote control to the ASVs by operators. Telemetry logging was achieved both on-board and via Kafka.

In the second phase, the focus shifted to controlling the ASVs through the on-board IoT device. Mechanisms were developed and extensively tested to manipulate one mission plan at a time. Mission plans, stored in the memory of the motion controller, were converted from well-known text (WKT) line strings into a series of Micro Air Vehicle Link[7] (MavLink) commands to control the motion controller. Onshore, REST APIs were developed in the control room to send commands in JSON format to each ASV through the MQTT protocol. By the end of the second phase, manual execution of remote commands for the ASVs was possible.

In the final phase of the project, the objective was to eliminate the human factor from the execution loop while ensuring safety mechanisms were in place to bypass autonomous execution of complex scenarios. Multiple improvements were necessary to support real-time re-routing of missions, both within the IoT device at the edge and in the control room onshore. It was ensured that the original mission plan was preserved on the IoT device during re-routing and that it continued execution after collision avoidance. Onshore, the MMDF tool,

[7] MAVLink is a very lightweight messaging protocol for communicating with drones (and between onboard drone components (https://mavlink.io/en/)

developed within the project, was included to efficiently process data in real time and communicate with external services. Necessary data bridges from internal to CREXDATA planform were deployed. Extensive testing was conducted using Software in the Loop in the lab before heading to sea, focusing on multiple collision scenarios. The emphasis was on providing COLREG-compliant re-routing solutions and enhancing system responsiveness.

4.2 Methods

4.2.1 System architecture

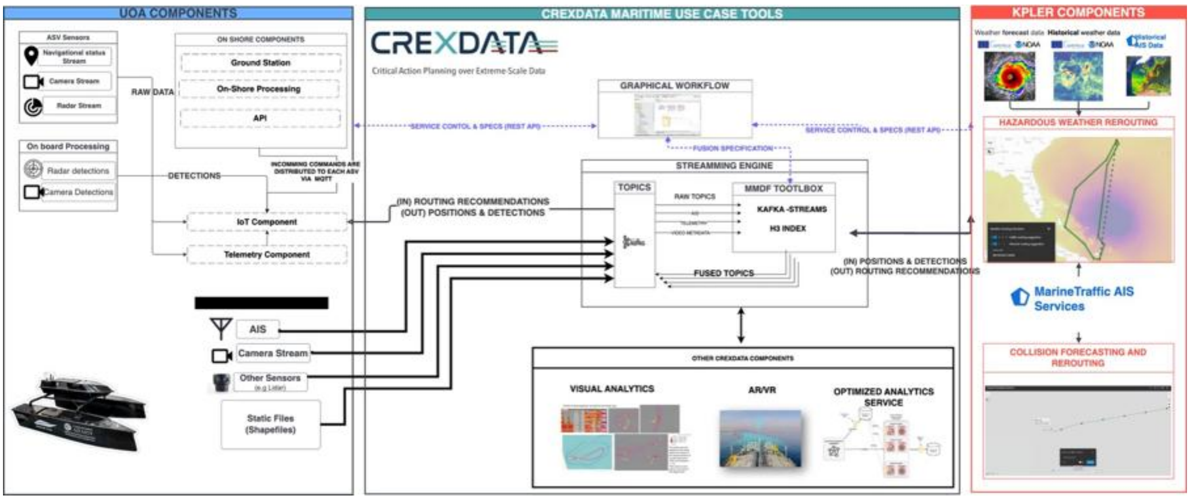


Figure 30 Maritime Use Case Architecture

The system developed for the maritime use case comprises of several components. The overall architecture is illustrated in Figure 30. At the core of the maritime use case system are technology demonstrators such as the graphical workflow and the Multi-Modal Data Fusion engine (MMDF), which have been made available within the CREXDATA project. The vessel collision avoidance service and the hazardous weather routing service are integrated into the architecture through RM operators and API endpoints respectively.

During sea trials of the maritime use case, UAEGEAN provides injects into Kafka topics real time data from different sources. This data consists of vessel traffic data (local AIS messages), ASV telemetry data and sensor detections. Sensor detections are gathered either on-board ASVs or by sensors onshore.

To execute any scenario, each ASV must receive a mission plan and initiate its execution. Simultaneously, the corresponding collision avoidance and hazardous weather routing services must start execution on Kpler's servers.

University of the Aegean servers, located in university premises, host the necessary API endpoints, a Kafka and an MQTT cluster, that are required services for monitoring and controlling the ASVs. In addition, at the marina control room a local installation of AI studio is used to initiate execution, gather information and communicate with CREXDATA platform services.

This information is transferred on-board by sending JSON messages to each Autonomous Surface Vehicle (ASV) via the MQTT protocol. During its initialisation phase, each ASV's companion computer subscribes to a dedicated command MQTT topic. Upon receiving a new command, the companion computer processes it in real time. Once the JSON message is consumed, the companion computer initiates processing at the edge. First, it converts the message to a MavLink command and then transmits it to the onboard motion controller. The companion computer also monitors mission execution and provides telemetry back to Kafka topics via the control room server.

In CREXDATA servers, the Multi-Modal Data Fusion (MMDF) streaming service processes records of ASVs telemetry and near-by vessels, acquired through AIS, to identify proximity events and in case of possible collision events it publishes fused records in a dedicated Kafka topic. To monitor execution, a web application visualises updates on a map from mission plans, ASV telemetries and nearby vessels.

4.2.1.1 Voyage Data Streamer (VDS)-IoT device

At the core of the maritime use case is the concept of data acquisition and processing at the edge. In this context, we deployed on-board of our ASVs a "voyage data streamer", an universal edge computing IoT device, to support on-board sensor data collection, streaming operations and vessel control.

The selected device model is a general-purpose raspberry pi v5 model that is a low-consuming power device, it has small size and a multitude of connectivity interfaces, both wired (ethernet, USB ports, on-board pins) and wireless (wifi, bluetooth). The selected device runs the Linux Operating system and runs a python service.

This service communicates via MavLink[8] protocol with the motion controller device to pool GPS and telemetry data from it and through the same channel it sends back new updated execution plans. In addition, it collects data from the onboard camera sensor connects to the control station over internet through the MQTT[9] protocol.

Importantly, the system has been designed to be fully compatible with NMEA 2000, ensuring seamless integration with standard maritime navigation and sensor networks and enabling deployment on conventional vessels without modification. In larger vessels, multiple companion computers can be deployed onboard to collect data from remote locations on board.

[8] https://mavlink.io/en/about/overview.html

[9] https://mqtt.org/

Figure 31. Example NMEA 2000 electronic systems for a RHIB, illustrating data flow between GPS, AIS, engine gateway, and onboard sensors.

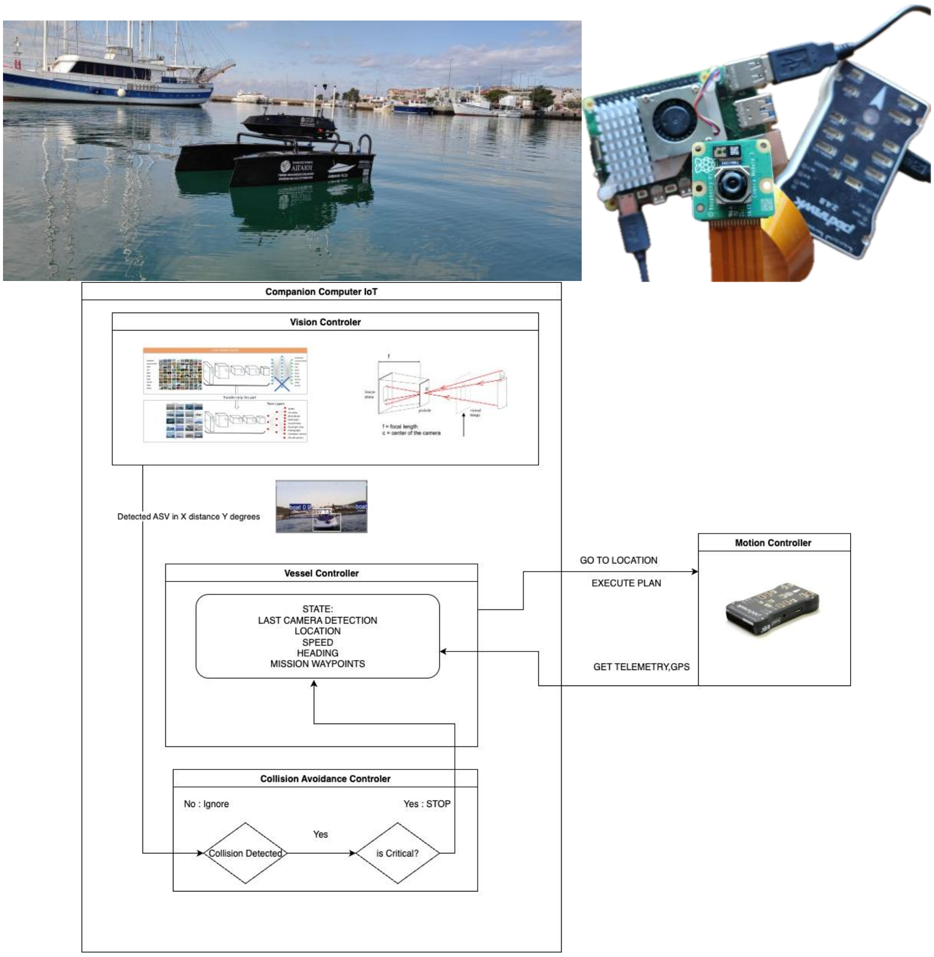


Figure 32 Components of the autonomous vessel. The AEROSEA ASV hull l(top-left), the hardware components the voyage data streamer consists of and communicates with within our ASV setup (top-right), Software architecture of the voyage data streamer.

4.2.1.2 Simulators – SITL

The maritime domain presents a dynamic and challenging environment for field testing. Factors such as maritime traffic and weather conditions may impede the smooth execution of experiments at sea and put at risk both equipment and personnel. Consequently, even though digital twin through AIS data has been presented in terms of monitoring and event forecasting [23], with regards to sea trails digital equivalents are used for prototyping and early problem identification and for accelerating experimentation cycles.

Figure 33 Software in the loop working an ASV simulator within the Maritime Use case Architecture.

In this context, we selected a software component, Software in the Loop (SITL)[10], to emulate the behaviour of the motion controller on our vessels. This allows us to communicate with the companion computer and the Voyage Data Streamer (VDS) software in a laboratory environment. The combination of the SITL and VDS provides a digital twin of our vessels enabling us to test the entire architecture and the execution of maritime use case scenarios in the lab before deployment at sea (T2.4 Simulation and Tools).

Figure 34 Use of SITL along real traffic experiment.

In a laboratory setting, we were able to initiate and send specific mission plans to the companion software. Upon receipt of the mission, the companion computer transcribed the plan into MavLink commands and sent them to the SITL. The SITL component then emulates movement by providing updated telemetry back to the companion software, enabling us to test all components developed within the project.

For visual inspection and monitoring of executed scenarios we have developed a visual analytics component, i.e. a live map, that consumes and presents AIS data from University of the Aegean antennas and along with telemetry feeds, that can originate both from ASVs

---

[10] https://ardupilot.org/dev/docs/sitl-simulator-software-in-the-loop.html

offshore or their digital equivalent SITL in lab. Allowing us to combine real world data with our simulations. In addition, assigned mission plans for the ASV and their short-term forecasts are also displayed.

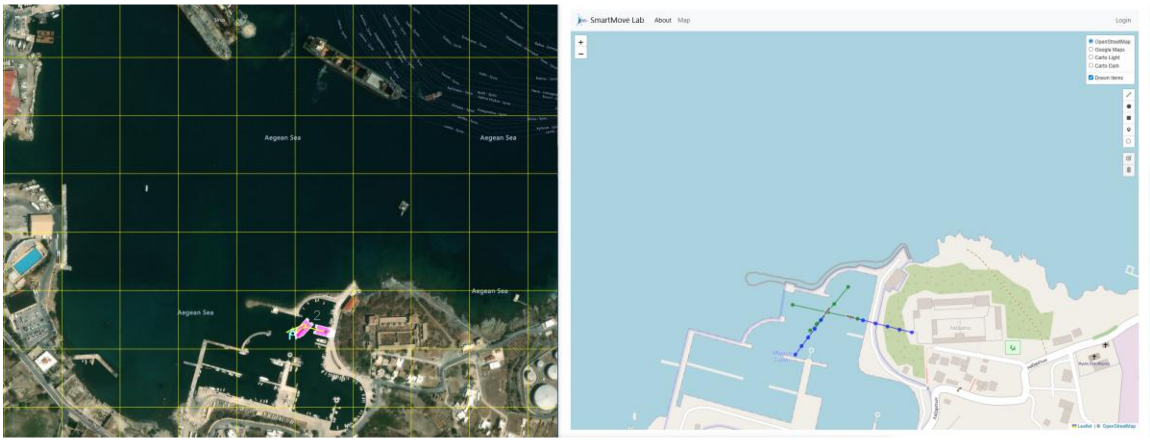


Figure 35. SITL (left) and Visual Analytics component (right) simulating and inspecting a crossing scenario in the 2025 sea-trials area. The assigned mission plans for each ASV as are depicted with green lines while the corresponding short-term forecast of each ASV is also shown with blue coloured lines.

4.2.1.3 Workflows

Graphical workflows are at the core of the maritime use case. As presented later in System architecture in WP3 deliverable, maritime use case relies heavily on the MMDF toolbox and its extension to design and execute its data flow (T3.2 Graphical Workflow Specification).

4.3 Results and evaluation

4.3.1 Scalability and Performance Evaluation of the Maritime Use Case Services in Large Maritime Data Processing

4.3.1.1 Distributed System Architecture

The Collision Avoidance and Hazardous Weather Routing Services of the Maritime-Use Case are built on a scalable, distributed computing framework that integrates heterogeneous streams of Automatic Identification System (AIS) messages with static vessel and environmental data. The Kpler MarineTraffic AIS services supply near-real-time AIS telemetry, while the NOAA and Copernicus services periodically ingested weather data, which are fused and cached in memory to maximize throughput. The architecture adopts the *actor model* and is implemented using the Pekko framework; asynchronous message passing and light-weight actors enable concurrent processing and automatic scaling across many cores. Incoming data are partitioned by vessel identifier to create a set of independent vessel actors each of which maintains the state of a single vessel and hosts deep-learning models for short-term and long-term route forecasting. The forecasts generated at the vessel level feed higher-level *cell actors* for proximity event detection and *collision actors* for collision prediction, both operating on a spatial H3 index grid. A *writer actor* periodically persists in the actor states in a Redis database, from which the user interface and an external CREXDATA platform can retrieve aggregated situational awareness and route-planning information. Figure 1 illustrates the overall architecture.

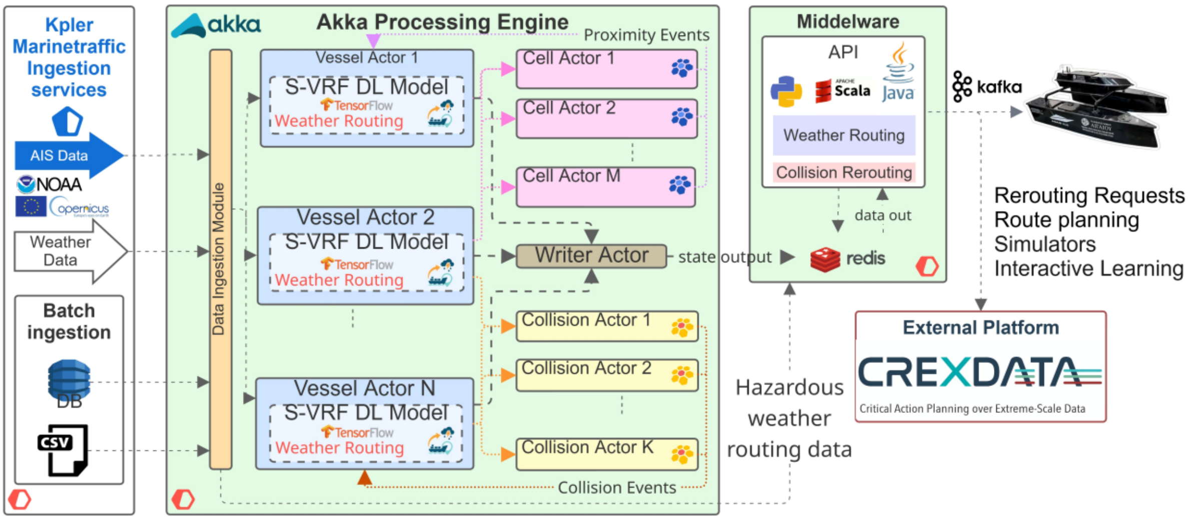


Figure 36 System architecture of the maritime use case. AIS and weather data are ingested, partitioned by vessel, processed in parallel actor and persisted in a database for visualisation and route planning. APIs are deployed to expose the services to external users.

The architecture addresses two classical shortcomings of maritime situational awareness systems—scalability and forecasting capability—by combining actor-based concurrency with prediction algorithms and models (kinematic models, short-term deep learning models, collision detection algorithm, collision avoidance algorithm (see D4.2), weather routing algorithm (see D4.2)) and distributed state management. The following sections detail the experiment setup and evaluate the performance of the weather-routing and collision-avoidance services.

4.3.1.2 Experiment setup and data

The evaluation employed a real-world dataset of historical AIS messages streamed from 24 June 2024 to 26 June 2024, covering the entire global fleet including terrestrial and satellite AIS data. In total 64.230.097 AIS messages were processed in real time from the entire globe. The processing engine operates on a high-performance system with 128 CPU cores and 256 GB of RAM. These resources allowed the Akka processing engine to instantiate tens of thousands of actors concurrently and maintain in-memory caches of vessel and environmental data.

4.3.1.3 Weather-routing evaluation

Weather routing computes navigation of safe vessel routes that avoid hazardous meteorological conditions. For each vessel, the processing engine integrates within the actor level the weather routing algorithm that, in combination with weather forecast and historical traffic data, generates rerouting suggestions according to the predicted sea states. Over the evaluation period, 12.682.140 vessel routes were generated based on the reported vessel speed and destination port in the AIS message. The significant difference with the number of total AIS messages relies on both to data quality and logic restrictions of the weather routing algorithm. In case the AIS message does not include a reported destination port or the destination port is included but is incorrectly spelled or is not included in the MarineTraffic Port Registry then the port of destination is not recognized, and a route won't be generated. The second limitation has to do with the algorithm logic where vessels that are inside the H3 cell area of their reported destination port won't generate a route suggestion as they are considered to have reached their destination. This area corresponds to approximately 253km^2 for H3 cells of resolution 5. As a significant amount of vessel traffic along with high and consistent AIS network coverage is found along the shoreline and port areas, this is the most significant factor influencing the resulting number of routing outputs.

The histogram in Figure 37 reveals that the vast majority of weather-routing computations completed in less than 10 milliseconds, with a long tail reaching up to 1125 milliseconds. The median processing time was 3.0 ms and the mean was 5.64 ms, indicating that occasional outliers had little impact on average performance. The minimum computation time was effectively negligible because many routes did not require rerouting. The sub-millisecond processing times observed for most routes demonstrate that the actor-based architecture can efficiently handle extreme-scale routing request on streaming AIS data.

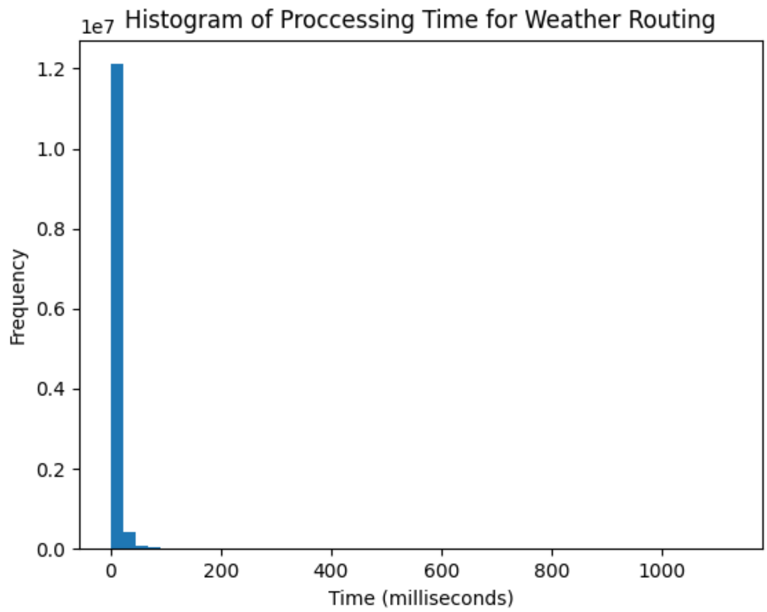


Figure 37 Histogram of processing time for weather routing shows that most routes are processed within a few milliseconds. The distribution is highly skewed with a small number of outliers above 100 ms.

4.3.1.4 Collision-avoidance evaluation

Collision avoidance operates on two actor levels. *Cell actors* monitor proximity events within individual H3 grid cells and identify vessel pairs that might collide based on their predicted trajectories. *Collision actors* then compute the near-term collision probabilities using a high-frequency time step. An external python client is called to handle the collision avoidance requests from the collision actors. During the evaluation, 13.213 collision-avoidance events were processed.

Figure 38 depicts the distribution of processing times for collision-avoidance events. The median elapsed time was 0.38 seconds and the mean was 0.43 seconds, with a minimum of 0.0296s and a maximum of 4.37s. The right-skewed distribution indicates that most events were evaluated under a second, while a minority of complex scenarios required more than three seconds of computation. The longer durations are attributable to dense traffic situations in which multiple vessel trajectories intersect, necessitating additional iterations of the collision-prediction model. Even in these cases, processing times remained below five seconds, enabling near-real-time alerts.

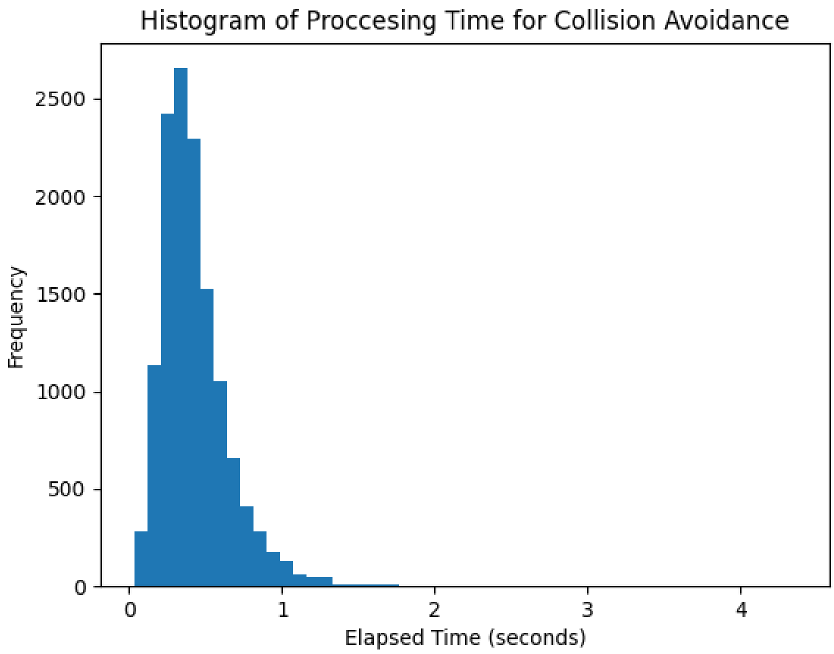


Figure 38 Histogram of processing time for collision avoidance shows that most events are evaluated in less than one second, with a small tail of longer computations.

The quantitative metrics from the experiment are presented in Table 13. Weather-routing times are reported in milliseconds (ms) and collision-avoidance times in seconds (s). The table highlights the system's ability to process millions of AIS messages and thousands of maritime risk events in real time while maintaining low latency.

Table 13: Production system specifications and quantitative performance metrics for the on-stream evaluation of the weather routing and collision avoidance services on the Pekko distributed framework

System Setup / Metric	Value
System cores	128 cores
RAM	256 GB
Data period	24–26 June 2024
Global AIS messages processed	64.230.097 messages
Weather-routing (Pekko): total routes processed	12.682.140 routes
Weather-routing (Pekko): min time	0ms
Weather-routing (Pekko): max time	1125ms
Weather-routing (Pekko): mean time	5.64ms
Weather-routing (Pekko): median time	3.0ms
Collision avoidance (Pekko): total events	13.213 events

System Setup / Metric	Value
Collision avoidance (Pekko): min time	0.0296 s
Collision avoidance (Pekko): max time	4.37s
Collision avoidance (Pekko): mean time	0.426s
Collision avoidance (Pekko): median time	0.38s

The results demonstrate that the distributed computing framework meets the requirements for real time maritime situational awareness. By partitioning the real-time data stream across vessel, cell and collision actors, the system scales linearly with the number of cores and maintains low processing latency even when confronted with extreme volumes of AIS messages. Weather-routing computations rarely exceed ten milliseconds on stream, making rerouting decisions effectively instantaneous for the global fleet. Collision-avoidance evaluations typically complete in under a second, providing sufficient lead time for course adjustments and enabling integration with autonomous navigation systems.

The formations of the performance distributions observed are important for operational planning. In weather routing, the outliers correspond to situations where long routing paths with multiple bad-weather areas require several alternative time dependent paths during the path finding step of the A* algorithm to be explored before an optimal (safe) route is identified. For collision avoidance, long durations arise in cases where the algorithm must generate routes for two interacting vessels (head-on collision) or in edge cases of the frenet frame algorithm optimal path search step, which serves as the basis of the definition of the optimal collision avoidance path (see D4.1). These cases highlight scenarios where additional computational optimization or early warning heuristics could further reduce latency. Nevertheless, the relatively small number of outliers and the resilience of the actor framework illustrate that the system can handle extreme-scale streaming data and deliver real-time decision support for maritime operations.

4.3.2 Demonstrators

4.3.2.1 MMDF

At the core of the Maritime Use Case is the use of the Multi-modal Data Fusion Toolbox and its AI Studio extension. We rely on MMDF to process data in real time and fuse data sources and records based on their spatial and temporal proximity, common tasks in the mobility data mining [24], [25]. In the AI Studio workflow designer, a two-directional workflow has been designed to support the maritime use case. The first direction is the organisation of ASV telemetries AIS and other sources to fused records based on vessel spatial proximity and short-term forecasts. These records are formatted and sent to the Collision Avoidance Service. The second direction is the recommendation from the Collision Avoidance Service to reroute ASVs, which is sent back to them via Kafka topics with appropriate format. A detailed presentation of the use of the MMDF in the maritime domain is presented in WP3.2 deliverable in the corresponding MMDF section.

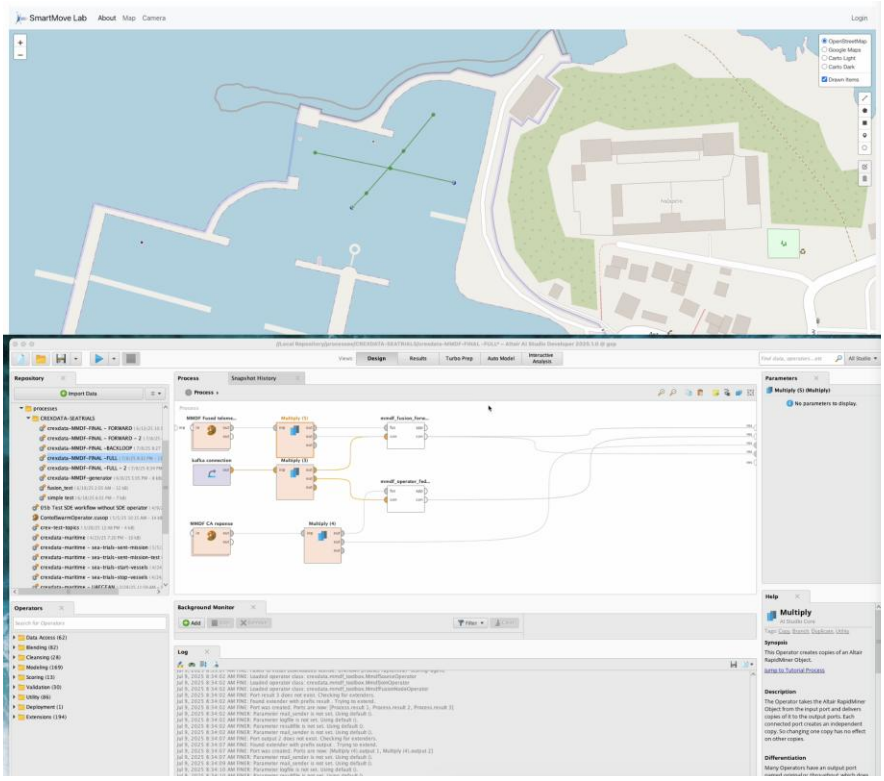


Figure 39 Use of MMDF AI studio extension for the maritime use case.

4.3.2.2 Augmented Reality

Within the maritime use case, the augmented reality demonstrator supports real-time decision making for ship captains during navigation, especially in congested or restricted waters. Traditional bridge tools like radar ECDIS and AIS displays demand frequent head-down interaction and cognitive integration of multiple sources. The demonstrator explores whether head-worn augmented reality (AR) can present crucial situational information directly to the captain's forward field of view without obstructing visibility or increasing workload. Building on the initial prototype reported in the previous reported period in deliverable D5.1, the final version features a head-worn AR navigation assistant on Microsoft HoloLens 2.

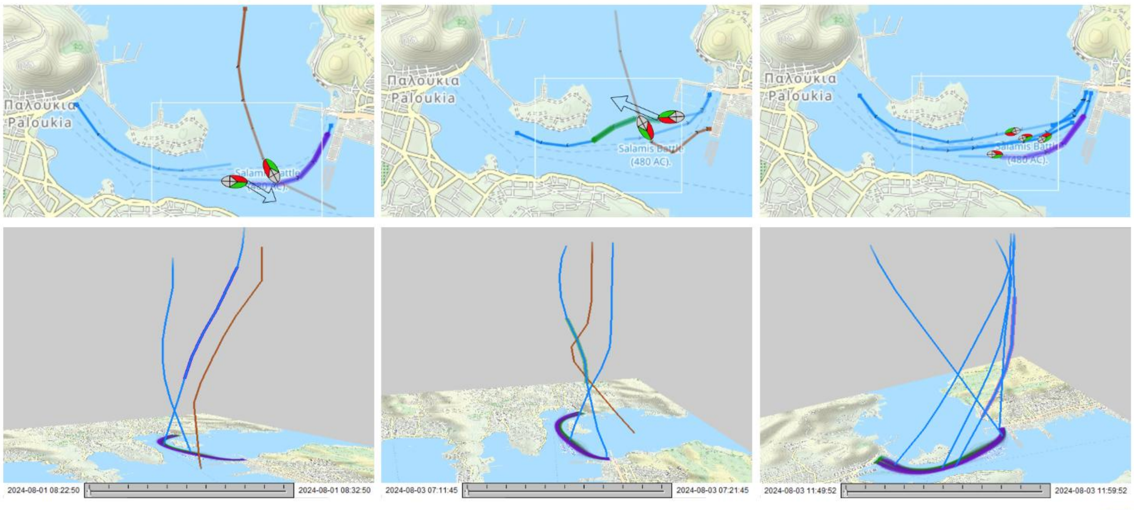
The design objectives for the final maritime prototype are threefold: enhance situational awareness by presenting relevant navigational and risk information directly through the bridge windows. Minimise interaction overhead through hands-free gaze and gesture-based controls. Demonstrate robustness in realistic conditions by validating the system on two ships.

This demonstrator integrates with CREXDATA data and prediction pipeline, ingesting live GPS and AIS streams. It also incorporates route-forecast ensembles from the platform's collision avoidance and risk assessment modules. This information is then overlaid as spatially registered holograms on the ship's bridge, including recommended routes, predicted collision points, points of interest and uncertainty regions. To test the demonstrator technology, a 44-metre cruise ship and a tugboat were used. Expert user feedback iteratively refined the system's calibration visualisation and interaction. The final architecture and implementation, along with the results from the ship trials, are presented in deliverable D5.2.

4.3.2.3 Visual Analytics and XAI

Computer-based visualisation systems help vessel operators analyse movement characteristics, detect anomalies, and assess stability. These systems aid in data exploration and analysis, particularly for detecting potential collisions and interactions between vessels. In this context within deliverable D2.2 an initial analysis of the of maritime data collected from the 1st Aegean RoBoat Race (Autonomous Robotic Vessels Competition) race has been presented [26].

This analysis has been extended [27] and applied to a historical AIS dataset from a narrow heavy traffic passage near Piraeus port transferring knowledge gained from field testing into a real-world scenario. In the figure below a visual exploration of details of individual deviations from standard routes id depicted.



The upper images are map fragments showing the spatial contexts of the deviations. Below them, the spatio-temporal contexts are represented in a space-time cube. The ellipsoid markers at the origin of each deviation on the map indicate each vessel's direction. The colored segments indicate sailing navigational lights that are located in the front part of each vessel, according to the International Regulations for Preventing Collisions at Sea (COLREGs) (International Maritime Organization (1972)). The vessels with arrows in the top-center and top-left figures indicate clear application of the COLREGs for powerboats crossing each other's routes. In crossing situations, a vessel approaching another vessel's left/red side (port side in nautical terms) must avoid crossing ahead of her, while the other vessel has to maintain her course. In the top-right figure, multiple vessels approach the port area simultaneously restricting significantly their manoeuvrability, forcing vessels to reduce

4.3.2.4 Integrated services

Users are able to access the both the collision avoidance and the hazardous weather routing services developed by Kpler on the Maritime Situational Awareness platform. Through the integration of the services on the Actor level in Pekko possible vessel collision events and collision avoidance paths are calculated in close to real time while weather optimal routes are generated for incoming AIS messages towards the reported destination ports. At the same time for all detected events and generated routes on stream simulator functions are available to the users for increased situational awareness and solution space exploration. The following figures present UI views of real predicted events and routes on the MarineTraffic AIS Stream. Figure 40 presents a predicted collision event between two vessels at the port of Algeciras in Spain. Figure 41 presents the resolution of the collision avoidance path generated on stream for the predicted collision event in Figure 40 only for the overtaking vessel along with an example on the use and output of the simulator function. Figure 42 depicts the hazardous weather routing UI and solutions for a voyage plan in the case of a hazardous weather event. Finally, Figure 43 depicts a hazardous weather routing example on streaming vessel AIS positions from the MarineTraffic platform.

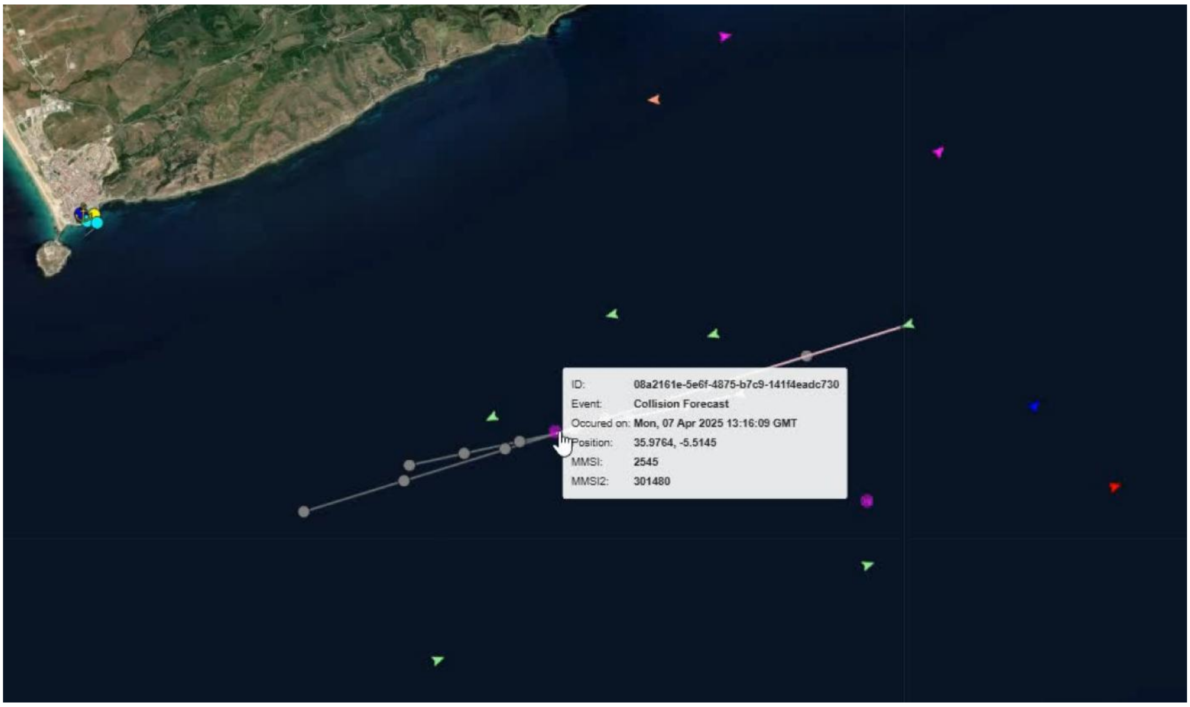


Figure 40 Collision detection prediction example from the AIS stream on the MarineTraffic platform

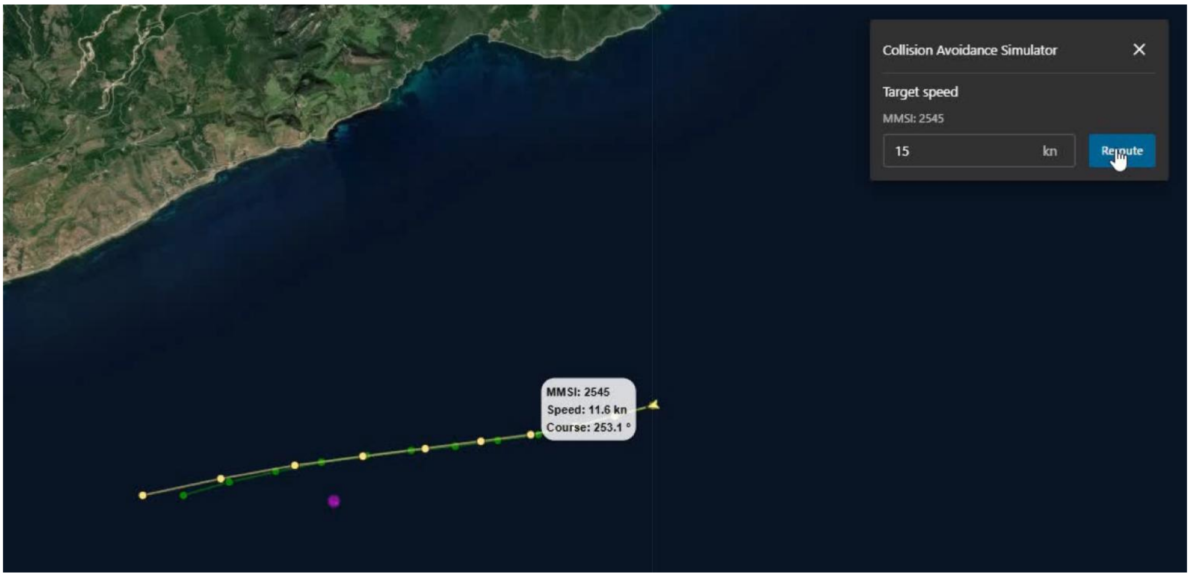


Figure 41 Collision avoidance path generated on stream for the collision detection example in Figure 40. Result is generated only for the vessel tasked to perform the avoidance according to the COLREG regulations (green line). Yellow line showcases the alternative route with a user selected speed.

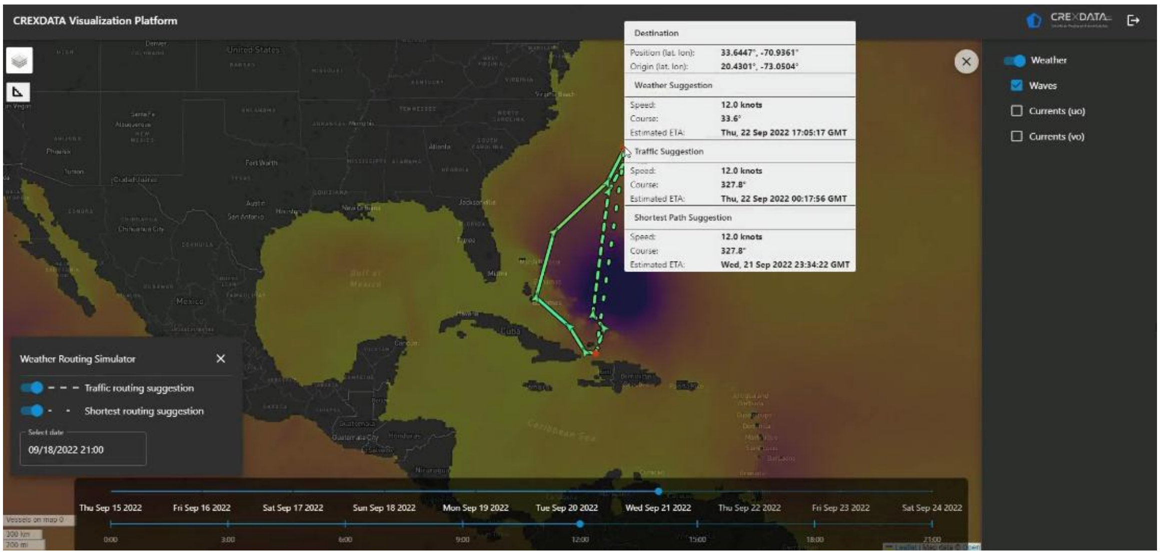


Figure 42 The hazardous weather routing GUI

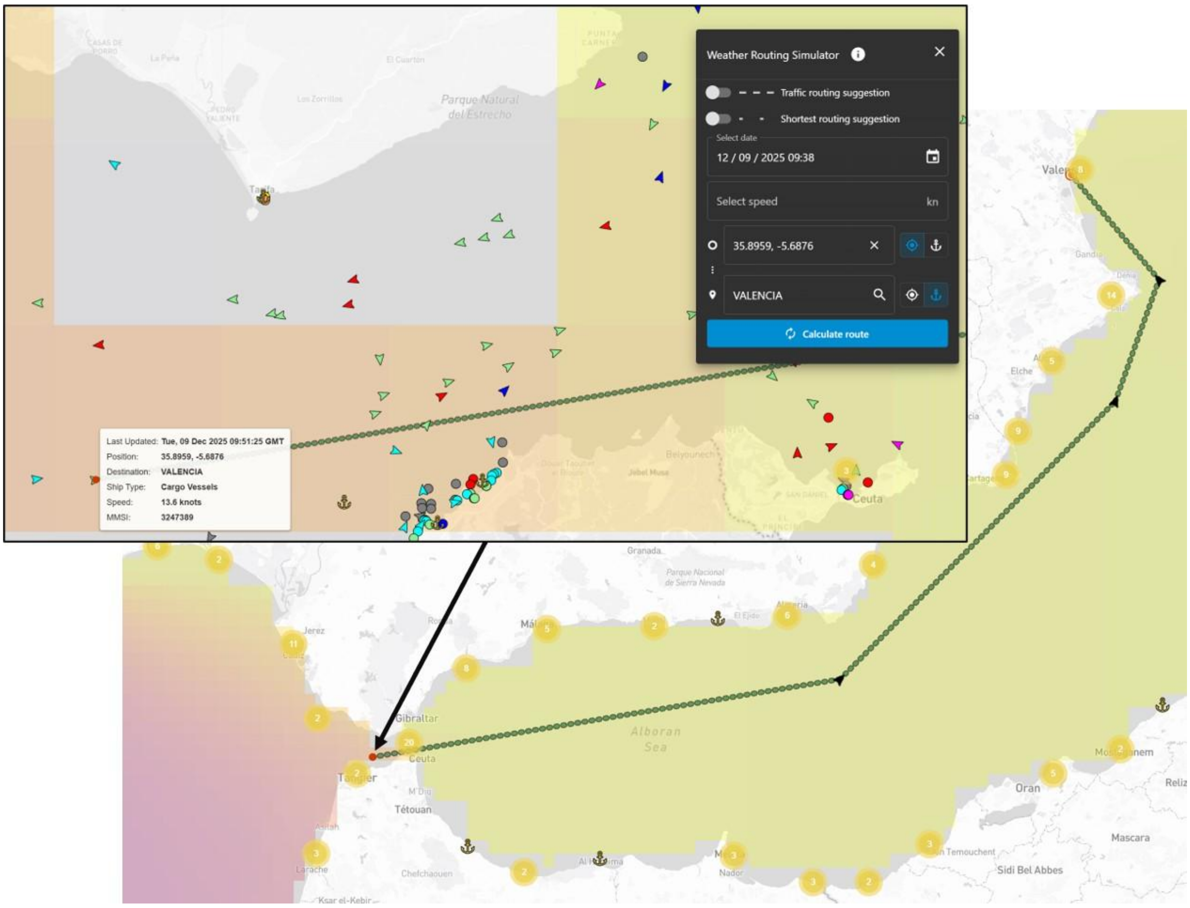


Figure 43: Hazardous weather routing example on streaming vessel AIS positions from the MarineTraffic platform. Image depicts the weather optimal route for a vessel (MMSI: 3247389) towards its destination port VALENCIA.

4.3.2.5 Wave Analysis (FMI)

Accurate weather forecasts are of extreme importance to ensure safety of equipment during sea trials. Even though there are some publicly available oceanographic forecasts for example in the Copernicus Marine Service (marine.copernicus.eu), those products typically have coarse resolution considering the coastal regions and archipelagos and cannot be used to estimate surface waves in areas that typically small ASVs can operate, i.e. near coastline, or in ports. In addition, limited capabilities and size of smaller vessels, make them sensitive in sea conditions that larger vessels typically sail.

In this context, we explored the option of using an event-based method to identify severe, potentially dangerous wave conditions. This method estimates the exceedance of predefined thresholds for significant wave height. It assesses the duration, magnitude, and location of the exceedance based on gridded model data when wave conditions meet predefined criteria. For the sea trials, the method was applied to estimate exceedances of a 2 m significant wave height using the Copernicus Marine Service Mediterranean Monitoring and Forecasting Centre's wave forecast. Figure 44 shows an example of the event-based check in the Aegean Sea near Syros Island, one day before the sea trial. The method visualizes on a map the areas where wave conditions exceed safe limits and indicates how long the exceedance is expected at different locations. For example, waves were forecasted to remain above 2 m for 25 hours in the northern regions of the islands, but only about half as long in the south. Combined with information on the start time of the exceedances and other specific details such as wave period and dominant wave direction, operators at sea can better plan the timing of different operations.

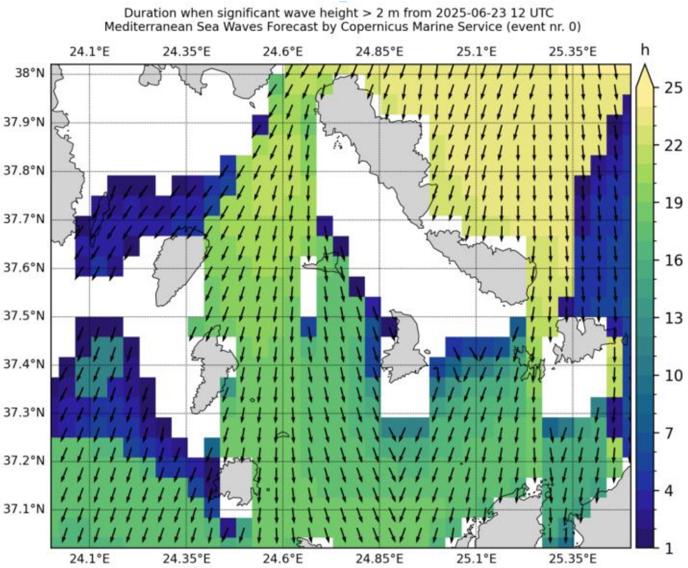


Figure 44: Example of event-based exceedance analysis showing the regions (coloured areas) where significant wave height exceeds the given threshold of safe navigation (2 m in this example) are forecasted for the first time and for how long (colour range) around the study area of Syros Island. The dominant direction of the waves (arrows)

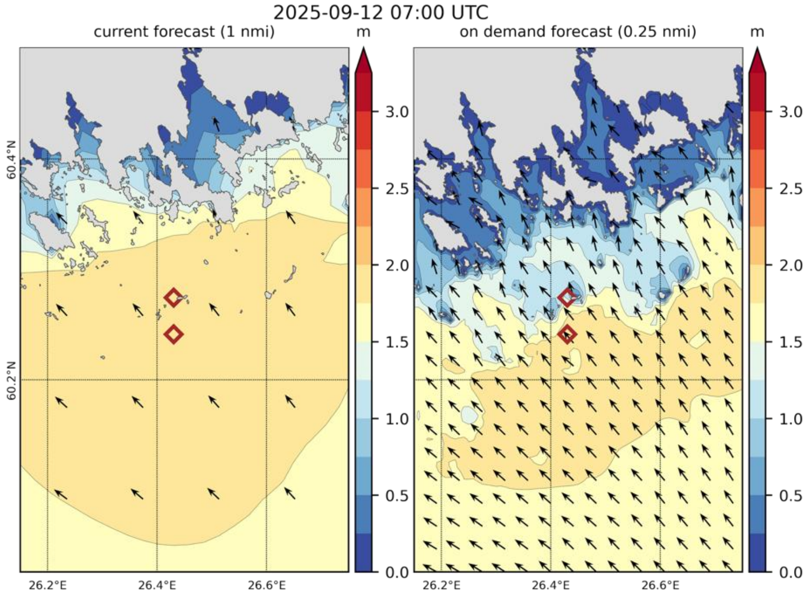


Figure 45: Wave forecast for the coastal area of eastern Gulf of Finland from regional 1nmi resolution forecast (on the left) and on-demand high resolution, 0.25 nmi (on the right). Color scale shows significant wave height and arrows the dominant wave direction on every 5th model grid point.

We also used the event-based method to design a coastal on-demand wave forecasting service. Due to the high computational cost of high-resolution coastal grids, it is beneficial to apply them only when a more detailed forecast is needed, for example, during severe wave conditions. We built a demo service where high-resolution wave forecasts are generated for a coastal area in the eastern Gulf of Finland, Baltic Sea. The production of the high-resolution forecast (Figure 45) is triggered when the coarser regional wave forecast predicts that the threshold of 2 m significant wave height will be exceeded. An example of the event-based check (Figure 46), similar to those used during the sea trial, shows that waves exceeding 2 m are predicted over a large area of the central Gulf of Finland for forecast lead times between 46 and 64 hours, and this exceedance is expected to last up to 24 hours.

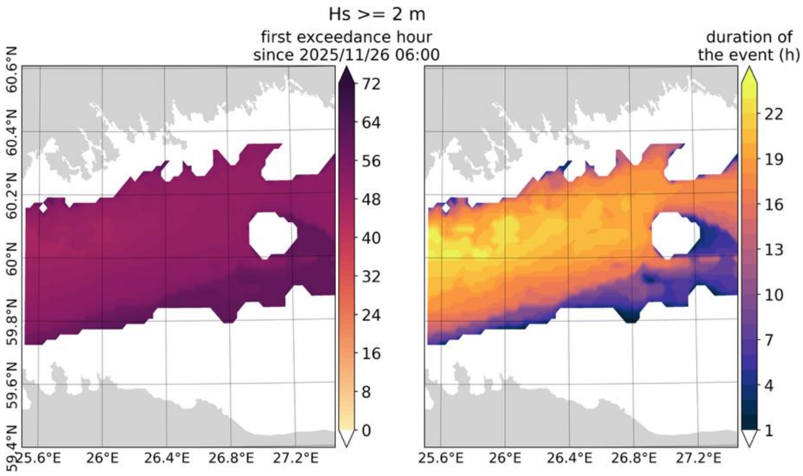


Figure 46: Example of an exceedance check for the November 26, 2025, 06:00 UTC forecast from the Copernicus Marine Service Baltic Sea Wave Analysis and Forecast product. The left panel shows the hour of the first exceedance of 2 m significant wave height from the start of the forecast, and the right panel indicates the expected duration of this event.

The demo service has gone through internal validation during Oct-Nov 2025 at FMI and is planned to be published during Dec 2025. An example of the forecast given on the website[11] is given in Figure 47.

[11] https://marine.fmi.fi/aallokko/

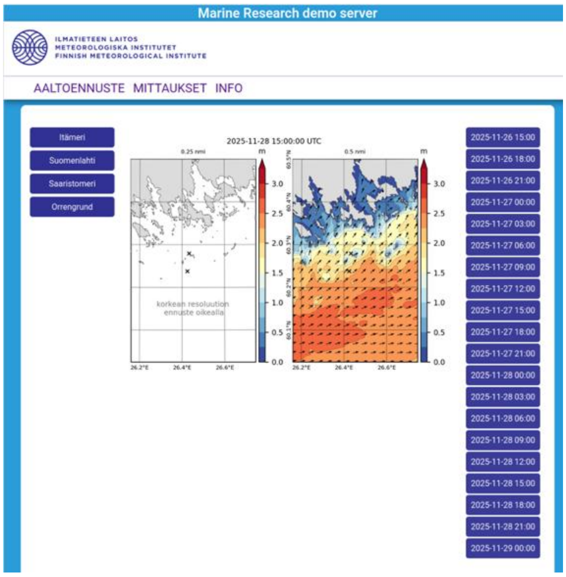


Figure 47: Snapshot of a demo website showing a high-resolution wave forecast for the eastern part of the Gulf of Finland (right panel). On the maps, the colour scale represents significant wave height and arrows indicate the main wave direction. When high-resolution forecast is not provided, the left panel shows the coarser resolution forecast. The website is in Finnish as it is primarily intended for the local users

4.3.3 Pilots

The sea trials of the Maritime Use Case were designed as a **multi-year, iterative validation campaign, gradually increasing** in complexity and realism. This approach built directly on the long-standing experimentation framework of the **Ro-Boat Race**, an annual competition of autonomous vessels in the Aegean Sea. The Ro-Boat Race provided a robust, safe, and repeatable test environment where multiple vessels operate simultaneously in real sea conditions—ideal for assessing communication systems, autonomy behaviors, and multi-vessel interaction before deploying technologies onto larger, crewed vessels.

The first phase of trials took place in 2023 [28], during which the project team focused on establishing foundational capabilities. These early tests concentrated on achieving stable communication between the universal IoT “companion computer,” the vessel’s motion controller, and the onshore control room. Using the Ro-Boat platform allowed the team to experiment with real NMEA navigation data, GPS streams, and telemetry under authentic sea dynamics while maintaining full operational control and safety.

In 2024, the trials progressed to field testing of two-way command and control. During the Aegean races, the autonomous vessels were used to validate the reliable reception and execution of remote commands, such as switching operational modes, uploading updated mission plans, and initiating fully autonomous navigation sequences. These trials demonstrated that the IoT device—already proven on our 6.5 m high-end RHIB—performed seamlessly on the ASVs, confirming its ability to function as a real-time edge computing node in a multi-vessel operational setting.

By 2025, the project reached an advanced stage of experimentation focused on evaluating complex autonomous behaviors. The vessels executed dynamic, COLREG-compliant collision-avoidance maneuvers in scenarios involving multiple ASVs moving in intersecting trajectories. These tests reproduced realistic maritime encounters, enabling the team to assess prediction accuracy, responsiveness, and behavior under uncertainty. The combination of autonomous vessels, high-fidelity data streaming, and real-time route forecasting created a unique operational testbed for validating the maritime components of the CREXDATA platform before deployment on full-scale vessels.

The final service evaluation took place in a small marina near Ermoupolis on Syros Island, Greece, offering a controlled yet fully authentic maritime environment. This location allowed consistent repetition of scenarios involving communication, command execution, hazard forecasting, and autonomous maneuvering. Across these multi-year trials, the iterative use of the Ro-Boat Race ecosystem ensured safe, cost-effective, and realistic validation before transitioning to the full-scale sea-trial experiment planned for the end of the project.

Figure 48: Glimpses from the Aegean Ro-boat race 2023. The registration on-site(top-left),the UAegean ASV(top-right), the University of Ioannina ASV (bottom)

4.3.3.1 Sea Trials – Collision avoidance (CA)

Three different collision scenarios, namely head-on (ca-1), crossing (ca-2) and overtaking scenarios (ca-3) were tested during sea-trials. The tests are conducted with a pair of small autonomous surface vessels executing each scenario. The ASVs are equipped with the universal IoT box device that ensures communication with control rooms via Kafka and MQTT services and ensures real-time monitoring. In each scenario the initial mission plan was sent to the vessels through kafka with the aid of an AI studio workflow, designed specifically for the maritime use case.

Within each scenario, the ASV's raw telemetry data (location, speed and heading) is transferred via the onboard internet connection to the University of the Aegean's MQTT cluster mechanism. A bridge logger application then pushes this information to two Kafka clusters: one hosted internally to the University of the Aegean and the other in the CREXDATA server. The MMDF (multi-modal data fusion) toolbox is deployed on a virtual machine to execute the workflow designed through AI Studio.

This workflow fuses ASV consumes telemetry Kafka topics and calculates for each ASV positional records a short-term kinematic forecast. Records that appear in proximity are fused together in a unified record internally. In case that two kinematics forecasts of a fused record intersect then a collision detection record materialized into a new Kafka topic. The collision detection topic is consumed by the Collision Forecasting and Rerouting service that generates records containing revised path plans that are sequentially pushed back to the ASVs for execution. This is a streaming operation, meaning that as new data are available new records of detections and mitigation actions are generated.

In the following sections we will present results for rerouting plans based on the logged data collected during sea trials. All logs of collision rerouting recommendations are included in the analysis regardless of their actual execution from the ASVs. A qualitative evaluation of the results and video recordings will be presented to experts for further evaluation.

4.3.3.2 Scenario Definitions

For the sea trials, four to five Autonomous Surface Vessels (ASVs) from the UAEGEAN fleet were used, each approximately 1–1.5 meters in length and equipped with electric propulsion, GNSS/IMU navigation sensors, and a MAVLink-enabled motion controller for autonomous guidance. The vessels carried an onboard camera and featured a modular electronics bay that supported integration of the compatible companion computer developed within the project. Their lightweight hulls, compact form factor, and reliable telemetry made them ideal platforms for safely testing multi-vessel interactions, COLREG-compliant manoeuvres, and real-time data-streaming functionalities under real maritime conditions.

Operating alongside the ASVs, a 6.5-metre high-end RHIB served as the support vessel, providing safety oversight, deployment and recovery assistance, and live validation of communication links under real maritime conditions. This combined setup allowed safe, repeatable testing of multi-vessel interactions and COLREG-compliant behaviours at sea. At all points in time, safety was ensured through strict operational procedures, visual oversight from the RHIB, geofenced mission areas, and immediate manual-override capabilities. The control center was hosted in the marina bay at the University of the Aegean's facilities, providing stable connectivity, mission monitoring infrastructure, and direct coordination with the field team.

This combined setup enabled safe, repeatable testing of multi-vessel interactions and COLREG-compliant behaviours under real maritime conditions.

In the following section we provide a pictorial representation for each one the scenarios tested during sea trials.

Rules 13, 14, and 15 were selected because together they represent the three most common and operationally critical encounter situations in maritime navigation. A full list of rules can be found in the International Maritime Organization page[12]. These rules cover overtaking (Rule 13), head-on encounters (Rule 14), and crossing situations (Rule 15)—the full set of scenarios that account for the majority of collision risks at sea. By ensuring that the autonomous vessels were able to detect, classify, and respond to these situations in a COLREG-compliant manner, the project validated the essential behaviours required for safe navigation in mixed human–autonomous traffic environments. These rules also provide clean, well-defined manoeuvring requirements, making them ideal for algorithmic implementation and for progressive testing during sea trials. Collectively, they form the minimum operational set needed to demonstrate the viability of autonomous collision-avoidance behaviours in realistic maritime conditions.

A summary table is provided below:

Table 14. A Summary table of scenarios' definitions.

Scenario / COLREG	**Specifications**	**Mitigation actions**
Head-on / RULE 14	• Two vessels from opposing directions • Approaching angle < 30degress	• Safe distance • Both vessel turn right
Crossing / RULE 15	• Two vessels from approaching with angle >30 degrees	• Safe distance • Vessel approaching the other vessels left side must give way/stand-on
Overtaking /RULE 13	• Fast vessel approaching another leading from behind (<10degrees)	• Safe distance • Vessel overtaking from either side.

[12] https://www.imo.org/en/ourwork/safety/pages/preventing-collisions.aspx

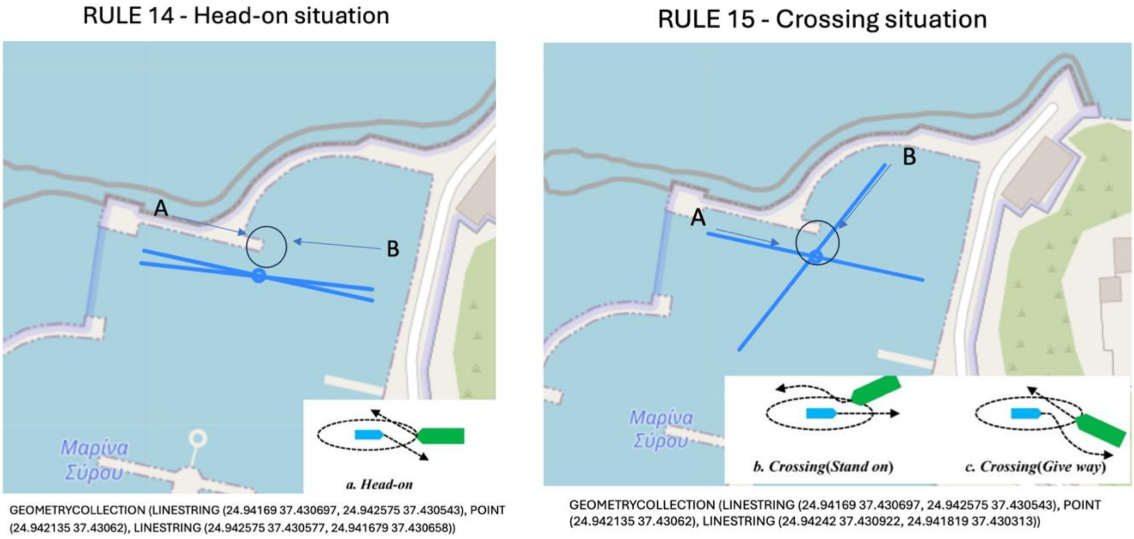


Figure 49: Indicative mission plans for the head-on and crossing scenario in the area of the sea trials.

The **head-on scenario** occurs when two vessels approach each from opposing directions, with their respective projected courses crossing each other in less than 30 degrees angle. We consider a safe buffer zone of 2m diameter around the intersecting point. In this scenario both vessels must follow the COLREG - rule 14 that expects vessels to turn right as a mitigation action.

The **crossing scenario case** and mitigation actions are described by the COLREG rule 15. In this scenario two vessels course cross one another, both approaching the intersection point at an angle greater or equal to 30 degrees. In this case the vessel approaching the intersection point having the other vessel on its right side must stand on or give way. The other vessel must continue its course.

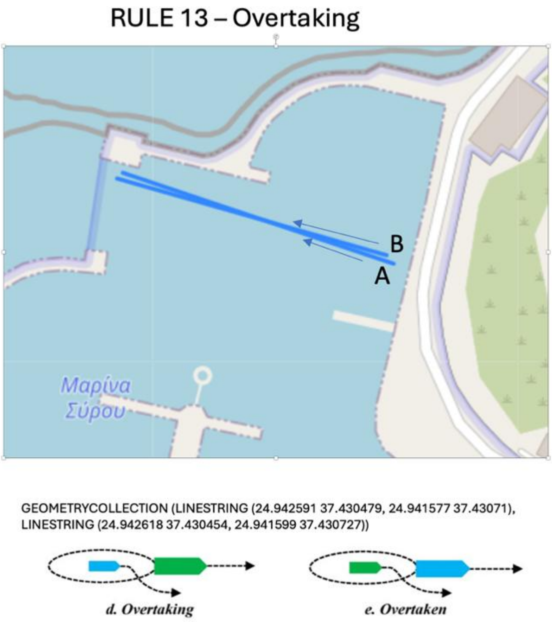


Figure 50: Indicative mission plan for the overtaking scenario in the area of the sea trials.

Finally, the **overtaking scenario** occurs when a vessel approaches fast from behind (less or equal to 10 degrees) a leading vessel. In this scenario according to COLREG rule 13 the fast-approaching vessel can pass the leading vessel keeping safe distance from either left or right side.

Conditions of Success

The evaluation of the sea trials will be based on specific quantitative and qualitative performance criteria. These include real-time monitoring of ASV positional data, generation of short-term kinematic forecasts for all operating USVs, and timely detection of potential collision events. Success is further defined by the system's ability to produce effective collision avoidance maneuvers maintaining a safe separation radius greater than 1 m, while demonstrating low-latency responsiveness throughout the data streaming and decision-making processes. In addition, all rerouting recommendations must adhere to COLREG-compliant navigation behavior to ensure operational safety and regulatory conformity.

4.3.3.3 Hazardous Weather routing Tests

The second service evaluated is the Hazardous Weather Routing service. Developed by Marinetraffic (Kpler) in the context of the CREXDATA project, it consumes daily updates of publicly available weather products from the Copernicus Service to identify weather conditions that could potentially pose a threat to maritime navigation. The end user responsible for navigation plans, whether onboard or onshore, can request a route plan that avoids the hazardous weather event. To evaluate the service, we selected 40 major Cross Atlantic routes shown in Table 15. Upon those routes, we executed a series of sequential service requests to examine service availability and response times. Consequently, the outcome along with additional cases were presented to experts were presented to an experts group during a workshop organized by the University of the Aegean.

Table 15. Cross Atlantic routes for HWR performance evaluation.

#id	source	target	#id	source	target
0	New York, USA	Lisbon, Portugal	20	Fort Lauderdale, USA	Cadiz, Spain
1	Norfolk, USA	Rotterdam, Netherlands	21	Freeport, Bahamas	Lisbon, Portugal
2	Boston, USA	Dublin, Ireland	22	San Juan, Puerto Rico	Valencia, Spain
3	Baltimore, USA	Antwerp, Belgium	23	Kingston, Jamaica	Las Palmas, Canary Islands
4	Philadelphia, USA	Bristol, UK	24	Port of Spain, Trinidad	Dakar, Senegal
5	Charleston, USA	Hamburg, Germany	25	Bridgetown, Barbados	Las Palmas, Canary Islands
6	Savannah, USA	Bilbao, Spain	26	Havana, Cuba	Lisbon, Portugal
7	Portsmouth, USA	Southampton, UK	27	Boston, USA	Reykjavik, Iceland
8	Jacksonville, USA	Le Havre, France	28	St. John’s, Canada	Reykjavik, Iceland
9	Halifax, Canada	Plymouth, UK	29	Reykjavik, Iceland	Bergen, Norway
10	Houston, USA	Brest, France	30	Boston, USA	Bergen, Norway
11	New Orleans, USA	Las Palmas, Canary Islands	31	New York, USA	Barcelona, Spain
12	Mobile, USA	Valencia, Spain	32	Miami, USA	Genoa, Italy
13	Tampa, USA	Porto, Portugal	33	Houston, USA	Marseille, France
14	Corpus Christi, USA	Tangier, Morocco	34	Charleston, USA	Naples, Italy
15	Montreal, Canada	Le Havre, France	35	Norfolk, USA	Piraeus, Greece
16	Quebec City,Canada	Liverpool, UK	36	Port Everglades, USA	Lagos, Nigeria
17	St. John’s, Canada	Belfast, UK	37	San Juan, Puerto Rico	Casablanca, Morocco
18	Halifax, Canada	Copenhagen, Denmark	38	New Orleans, USA	Algeciras, Spain
19	Miami, USA	Casablanca, Morocco	39	Houston, USA	Dakar, Senegal

4.3.4 Evaluation

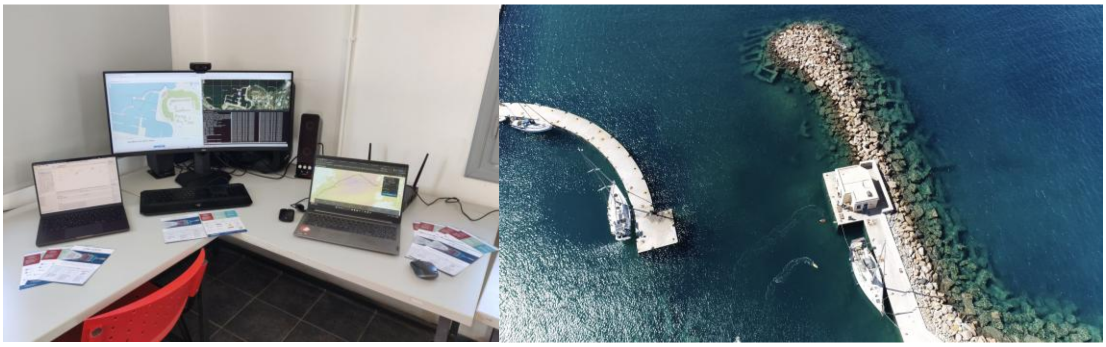


Figure 51: Control room setup (left) and aerial view of the control room and sea trials site (right).

4.3.4.1 Scenario Execution

During the final use case evaluation, sea trials were conducted in two phases. The first took place on 25 June 2025 and was co-located with the Maritime Informatics & Robotics summer school[13] and the IEEE symposium on Maritime Informatics & Robotics[14]. The second phase occurred on 19 September 2025. The first phase focused on system responsiveness and scenario provisioning while the second phase concentrated on data collection and COLREG-compliant execution.

A video summary for the first[15] and the second[16] trial is available online.

13 https://summerschool.eitdigital.eu/maritime-informatics-robotics

14 https://maritimesymposium.eu/

15 https://youtu.be/Q4CiqIZTjo0

16 https://youtu.be/WcXZZu1e8PA

Figure 52: A head on scenario execution during sea trials.

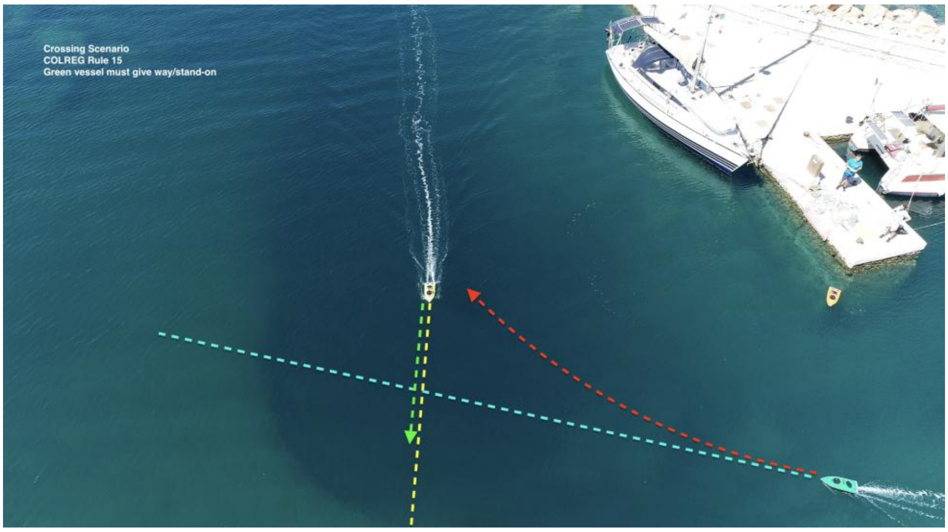


Figure 53: A crossing scenario execution during sea trials.

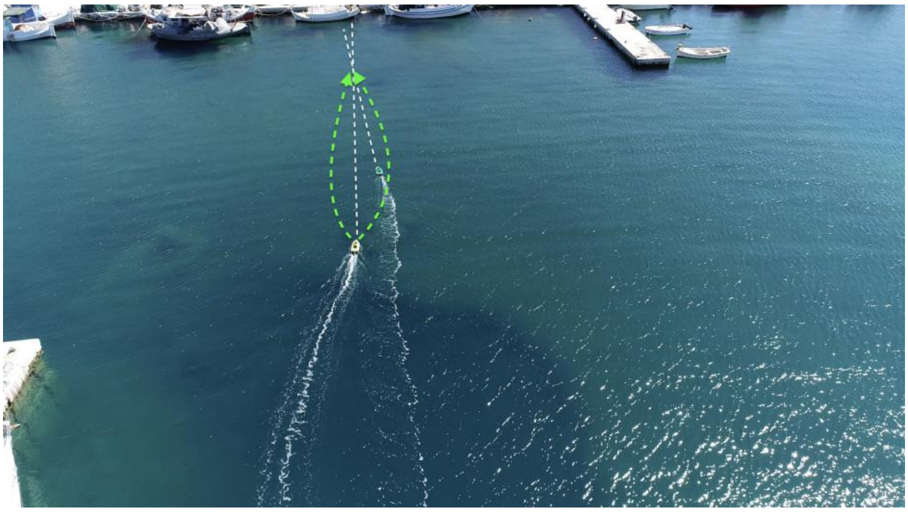


Figure 54: An overtaking scenario execution during sea trials.

4.3.5 Analysis of Sea trials data (CA)

4.3.5.1 Safe Zone avoidance:

One of the most challenging aspects to evaluate in real maritime conditions is whether two vessels on a potential collision course will successfully avoid one another while maintaining a safe passing distance. Environmental factors—such as surface currents, wind gusts, and small disturbances in vessel dynamics—can cause Autonomous Surface Vehicles (ASVs) to deviate from their planned trajectories, making on-field assessment particularly demanding.

To assess the performance of the rerouting service, two evaluation criteria were used. First, we examined whether the recommended trajectory ensured that each vessel remained at least 1 metre away from the predicted collision point, effectively creating a 2-metre safety buffer zone around the encounter. Second, we calculated the separation distance at the collision point if both ASVs were to travel at the speed suggested by the service. For safety reasons, this speed-adjustment feature was disabled during sea trials, and the vessels maintained their original speeds; however, the computation still provides an indication of expected separation under full system operation.

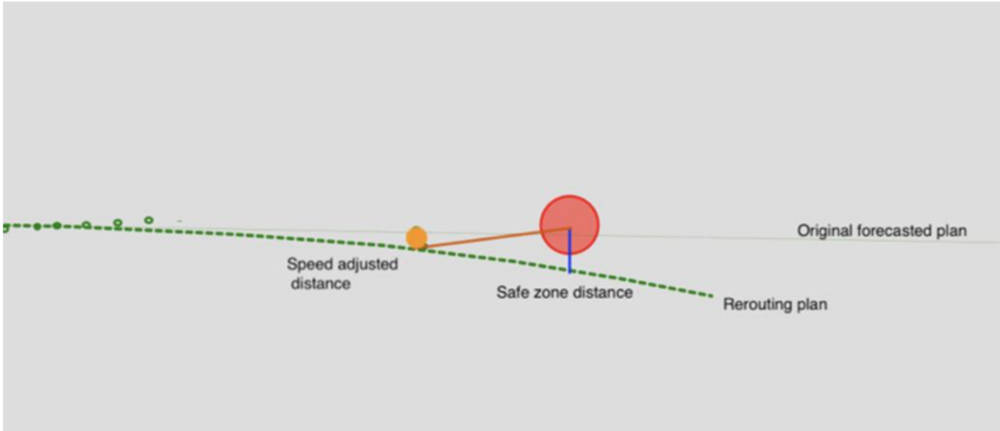


Figure 55: Pictorial representation of the safe zone, rerouting plan and originally forecasted plan.

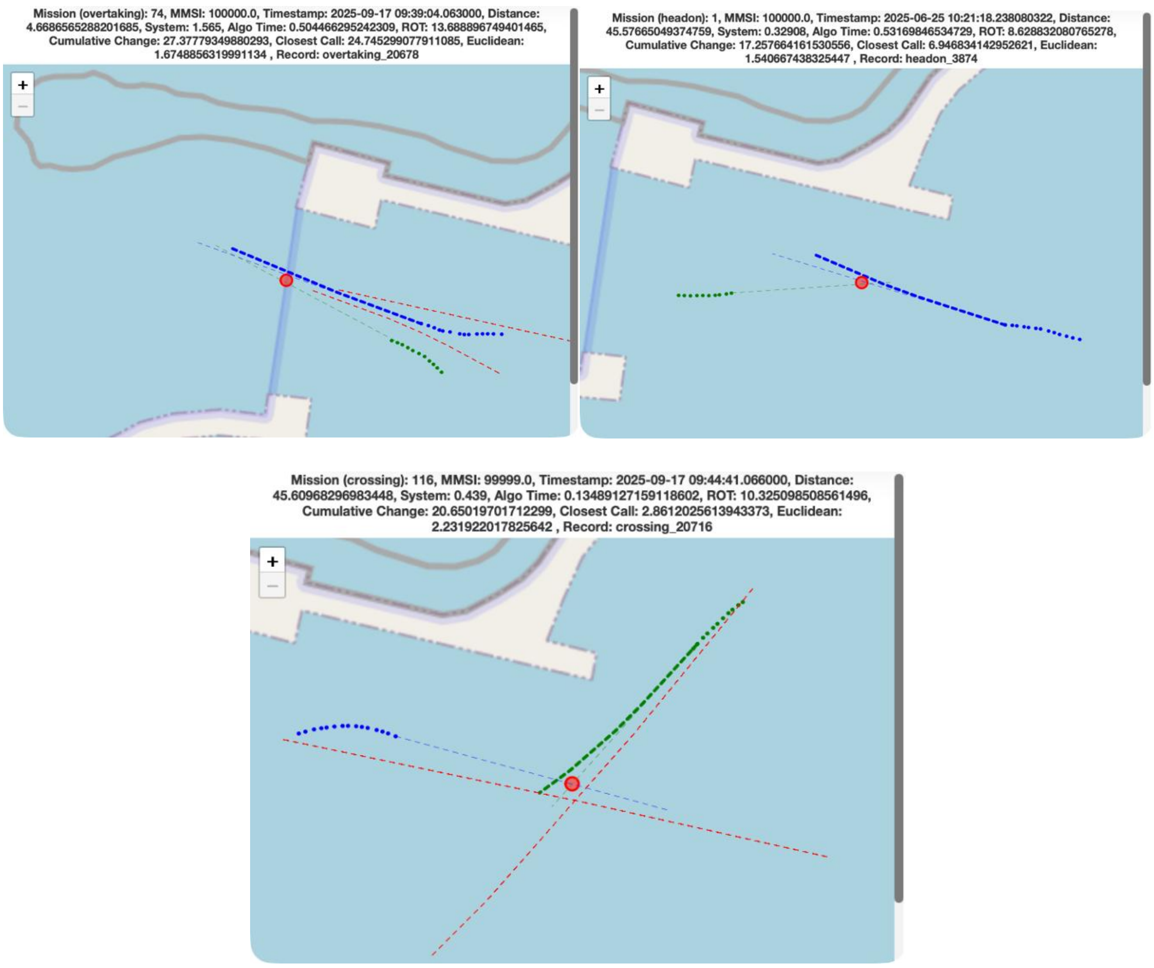


Figure 56: Selected examples from data analysis for overtaking (top left) , head-on(top-right) and crossing(bottom) scenarios. Light dashed lines correspond to the

original mission plan (red) and current forecasts (blue & green). Bold dashed blue and green lines are re-routing recommendations and dots (green& blue) are telemetry logs.

Table 16. Collected incident per case as they have been processed by the CA service.

	Incidents	Observed				Acceptance Rate			
	Valid	Collision	Safe Zone Intersection	Colreg Error	Rerouting error	Collision Avoided (%)	Safe Zone Intersection Avoided (%)	Colreg Error Avoided (%)	Rerouting Error Avoided (%)
crossing	98	0	0	11	17	100	100	88.75	**82.65**
head-on*	3	0	0	0	0	100	100	100	**100**
Overtaking *	36	0	0	5	5	100	100	86.11	**86.11**
Total	**137**	**0**	**0**	**40**	**31**	**100**	**100**	**88.32**	**83.94**

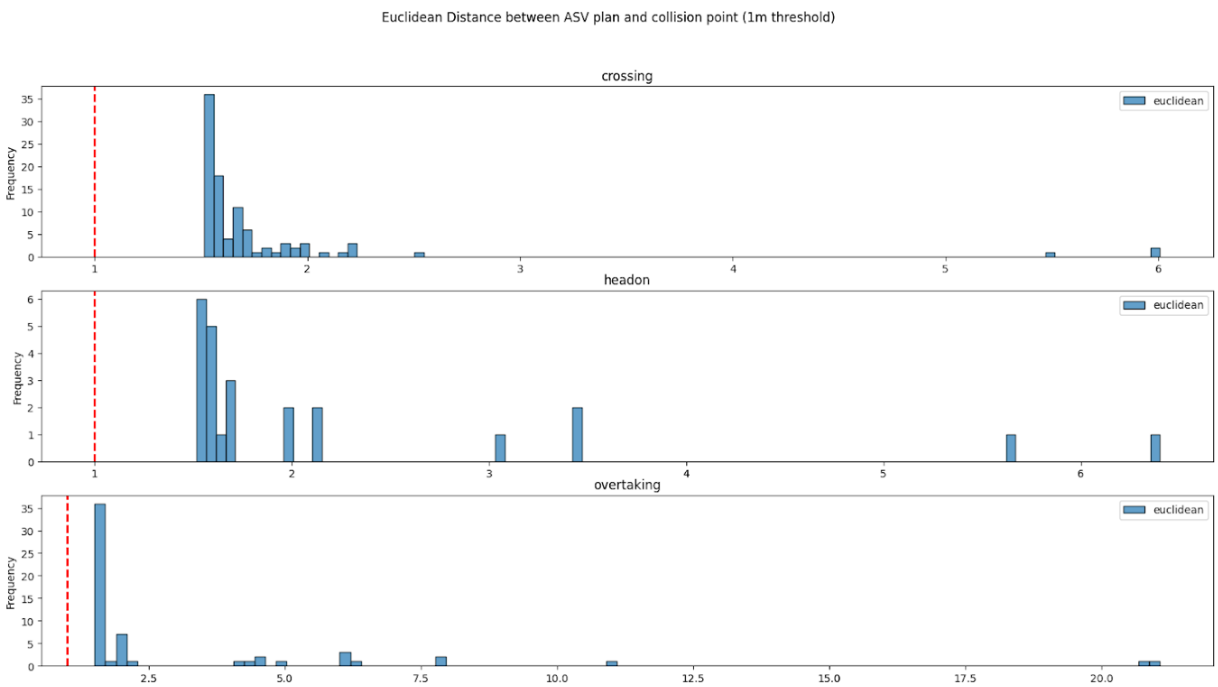


Figure 57: Histogram of distances between collision point and ASV rerouting plan per collision avoidance scenario type. The threshold marked with a red dashed line corresponds to the collision zone of a 1-metre buffer zone indicating that the suggested plan does not intersect with the collision point.

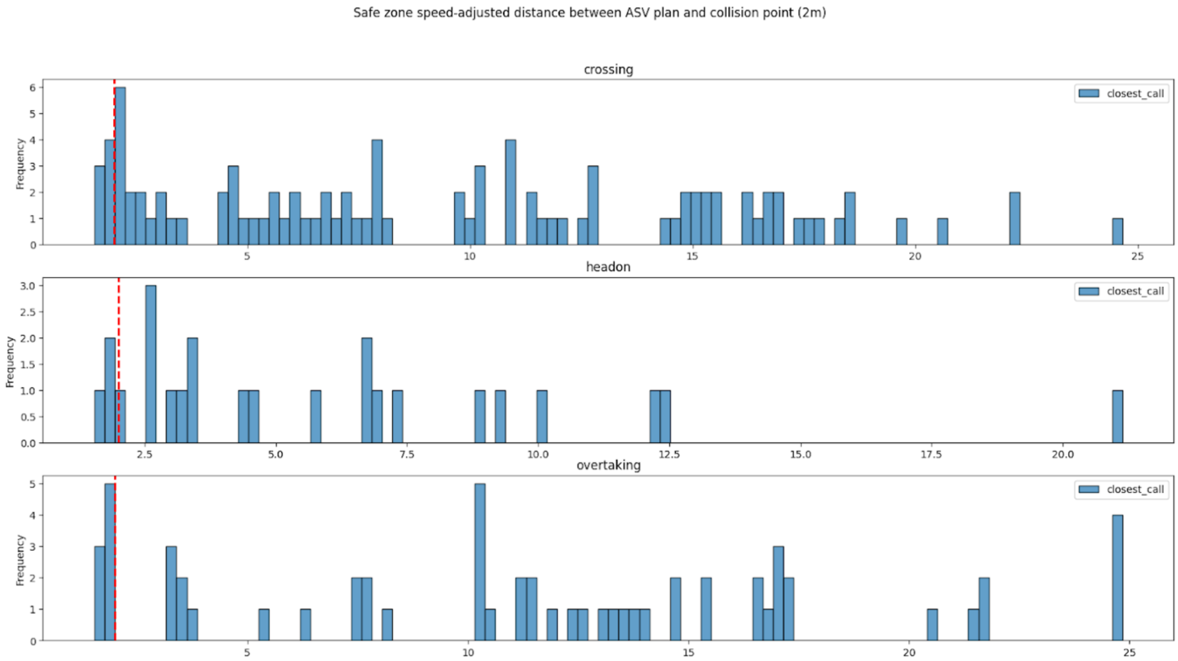


Figure 58. The histogram of distances with respect to speed adjustments between the collision avoidance rerouting plan and collision point is shown. The red dashed line represents a safe zone of 2 metres indicating that respecting the suggested speed limits the suggested plan will maintain a clear distance of two meters.

A total of 137 encounter incidents were recorded involving 4-5 ASVs operating within geofenced areas near Ermoupolis marina.

COLREG compliance was evaluated for each encounter based on Rules 13, 14, and 15.

- Crossing: 88.75% of COLREG errors avoided
- Head-on: 100% avoided
- Overtaking: 86.11% avoided
- Overall COLREG compliance: 88.32%

Rerouting performance reflects how reliably ASVs executed the recommended avoidance paths:

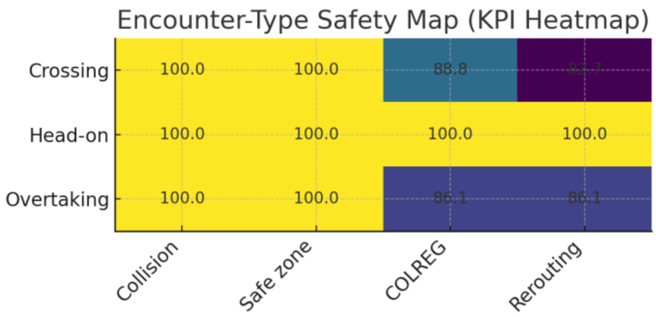
- Crossing: 82.65% of routing errors avoided
- Head-on: 100% avoided
- Overtaking: 86.11% avoided
- Overall rerouting accuracy: 83.94%

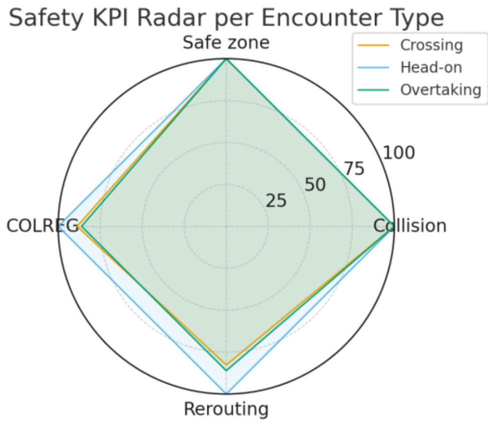
All routing deviations were minor (e.g., small offset in tracking) and did not compromise safety. Performance was strongest in head-on encounters, where geometry is simplest.

The 2025 pilot confirms that the Maritime Use Case components are robust, safe, and operationally mature:

- 100% safety performance (no collisions, no breaches)

- High COLREG-compliance (≈88%)
- Strong rerouting accuracy (≈84%)
- Consistent safe-distance margins under real environmental disturbances





4.3.5.2 Analysis of HWR tests

In contrast to collision avoidance service where there are specific metrics to be calculated, a qualitative assessment of the HWR service requires a good understanding of the weather forecasts, local conditions, and each vessel’s unique characteristics. Weather routing is an extremely challenging task, so we relied on experts’ knowledge to assess the qualitative characteristics of the service as presented later in the Expert User Evaluation sections of this use case.

In a workshop event held on 21-10-2025 in UAegean premises we invited faculty members and senior cadets of the Merchants Marine Academy of Syros to assist with the evaluation. Workshop participants we demonstrated the integrated version of the service and were presented videos of 4 different scenarios, three of them during extreme weather events in Atlantic.

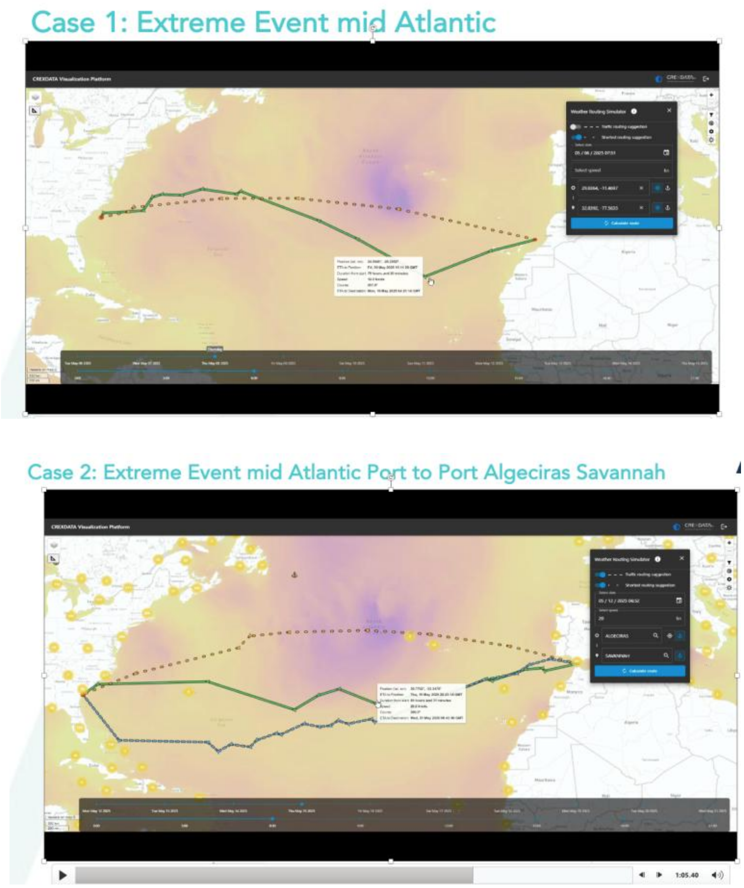


Figure 59: Two extreme scenarios presented to experts during evaluation event

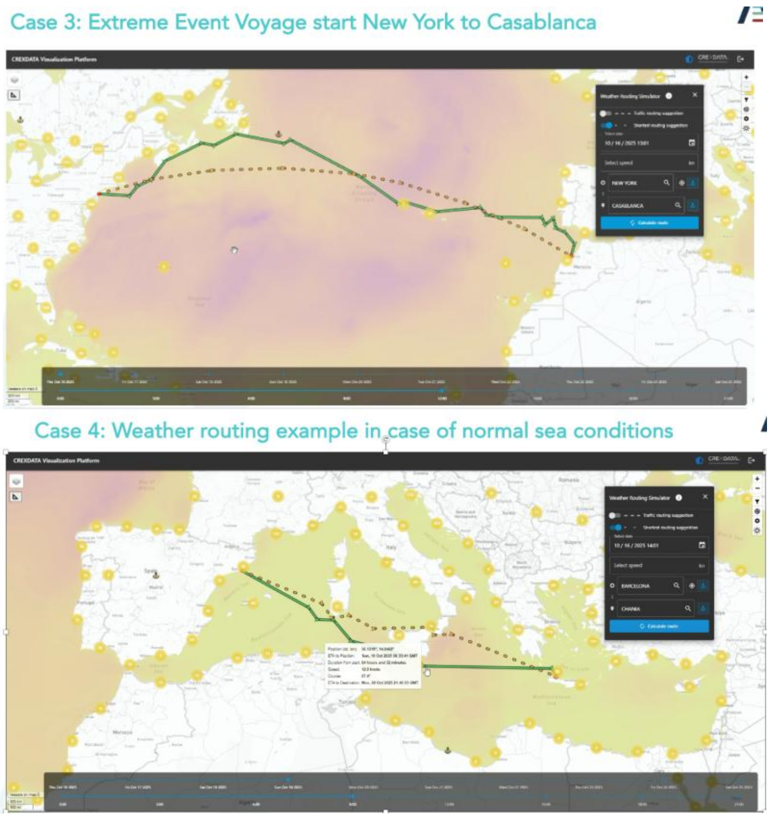


Figure 60: Two weather routing events presented to experts. A extreme case and a normal one.

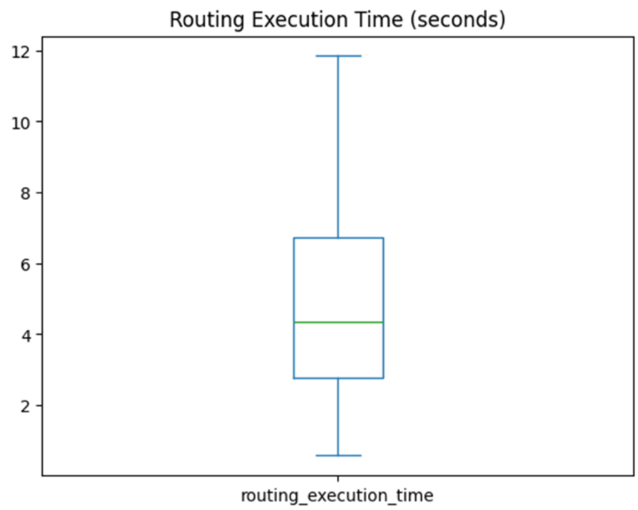


Figure 61 HWR responsiveness tests

To evaluate the service from an information systems perspective, we performed some tests regarding service responsiveness and capability to produce reasonable results in a challenging set of cross-Atlantic routes presented earlier in Table 15. Performing the tests, multiple API requests were generated for the cross-Atlantic routes. The average response time was 4 seconds with a maximum of 12 seconds. Visualising the results we consistently observe that in most cases the biggest part of the weather route aligns with the most frequent route as presented by the service, altering an approximate 2-3-day part of the route. In some cases, an entirely different route was suggested (Reykjavoik-Bergen).

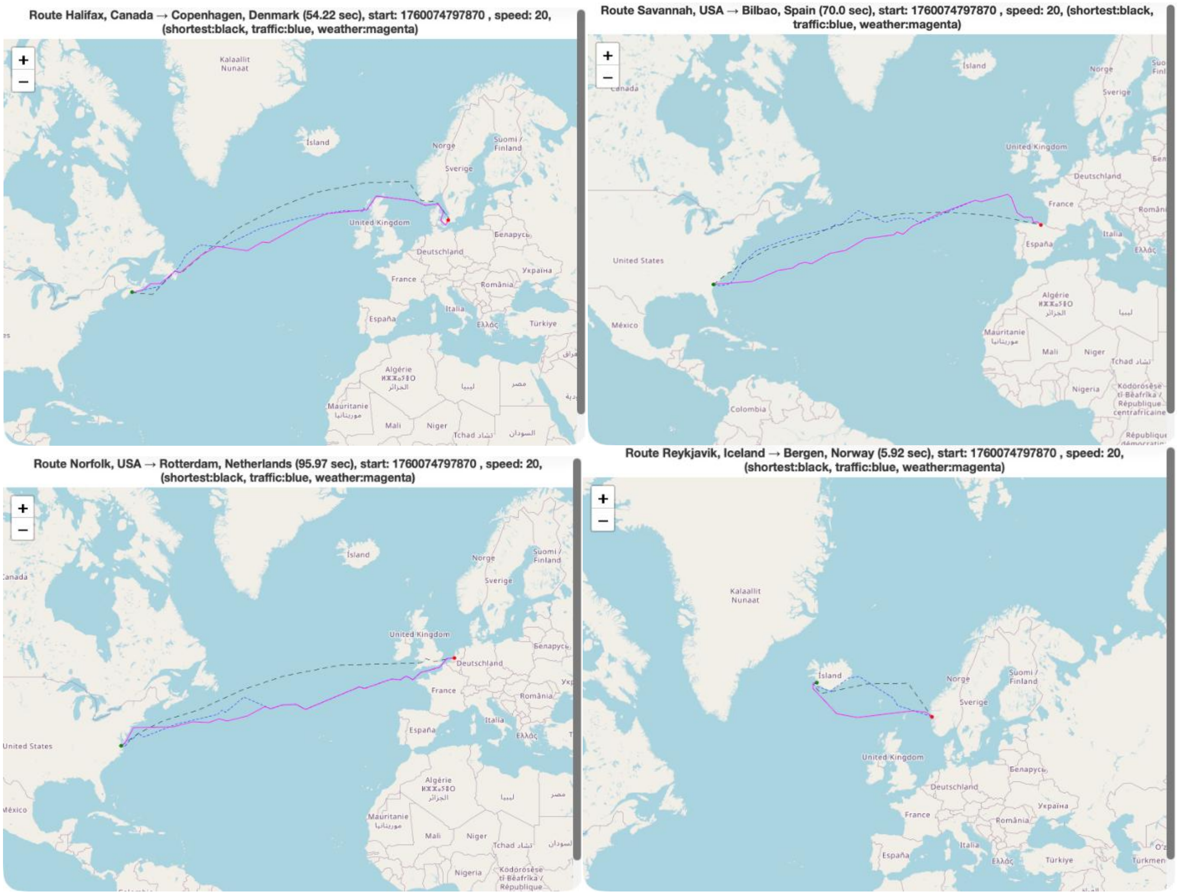


Figure 62 Examples of results produced during HWR performance evaluation on 12-10-2025. Magenta is weather routing result, Blue is the most frequent route and black is the shortest according to the results returned from the service.

4.3.5.3 Expert User Evaluation and Usability Results

The Maritime Use Case of the CREXDATA project represents a complex integration of real-time data acquisition, digital-twin generation, autonomous manoeuvre planning, and intelligent forecasting. Although the system has been validated extensively through quantitative means, particularly during the multi-stage pilots conducted from 2023 to 2025, **qualitative assessment by expert users** is equally essential. Maritime operations rely heavily on human judgment, situational awareness, and accumulated experience; therefore, the acceptance, interpretability, and operational relevance of advanced AI-driven technologies must be evaluated directly by those who navigate and manage vessels.

To this end, the project team engaged with a broad community of **industry practitioners, academic experts, maritime instructors, professional captains, VTS operators, and Merchant Marine students**. Across the project's duration, the architecture and services of the Maritime Use Case were presented in several forums, including:

- **DronePort Horizon Workshop**
- **Merchant Marine Academy of Syros (AEN Syros)**
- **University of the Aegean events and open demonstration sessions**
- **Specialised stakeholder briefings with maritime technology experts**

These engagements ensured that the Maritime Use Case was assessed not only for its technical robustness but also for its practical relevance, usability, and alignment with maritime operational norms.

The culmination of the qualitative assessment was the **Final Expert User Evaluation**, held at the **Merchant Marine Academy of Syros**.

This session gathered a diverse group of:

- **Professional captains and deck officers**
- **Instructors from AEN Syros**
- **Maritime safety and navigation experts**
- **Final-year Merchant Marine cadets**

The presence of instructors and cadets was particularly valuable. The students had previously observed the autonomous vessels during sea-trial operations and were therefore able to comment from a position of familiarity. The instructors, carrying decades of seafaring and bridge experience, brought the practical judgment required to evaluate the operational credibility of autonomous decision-support tools.

During the session, the project team demonstrated key elements of the Maritime Use Case, including:

- The architecture and data flow
- Real examples of COLREG-compliant manoeuvres
- Multi-vessel avoidance scenarios
- Hazardous weather rerouting
- Real-time system responses and latency characteristics
- Visualisation of predicted motion and confidence indicators

These demonstrations triggered lively discussions regarding the role of intelligent systems aboard modern vessels, the evolution of autonomous navigation, and the balance between automation and human expertise.

Figure 63: Final Expert User evaluation with Marine Merchant academy of Syros Instructors and final year students.

Instructors and senior cadets from AEN Syros raised insightful and operationally grounded questions regarding the maturity and reliability of such systems. Their feedback emphasised several points:

Advisory vs. Autonomous Roles

Experts agreed that, at this stage, intelligent systems such as those developed in CREXDATA should serve primarily as advisory and decision-support tools. While the algorithms demonstrated strong performance, they cannot yet fully replicate the depth of human situational awareness, especially in congested waters or rapidly changing conditions.

Importance of Maritime Experience

The instructors highlighted that the "seafarer's eye"—intuitive judgment shaped by years at sea—remains central to maritime safety. They stressed that the value of autonomous tools lies in supporting, not replacing, human decision-making.

Predictability and COLREG Transparency

Captains emphasised the necessity of predictable, transparent behaviour. Systems must not only execute COLREG-compliant manoeuvres but also explain or visualise their decisions in a manner immediately recognisable to a human officer.

Perspectives from Cadets

Senior cadets provided a future-looking viewpoint. As the generation who will operate alongside increasingly automated systems, they stressed:

- the usefulness of clear AR/visual projections of predicted paths,

- the importance of simplicity in interface design,
- the need for systems that integrate naturally into existing bridge workflows, and
- their openness to adopting AI-driven support tools if reliability is proven.

Their familiarity with modern navigation systems made their insights particularly forward-looking.

At the end of the process the evaluation questionnaire available in Appendix 1 was completed by the attendants on-site. Questions 1-2 of the questionnaire relate to demographic and presence information. Questions 3-4 of the questionnaire ask experts to what extend the presented services (collision avoidance and hazardous weather routing) reached their respected KPI from an expert's viewpoint, providing their qualitive assessment for the services. Question 5 examines system usability and question 6 is a free text comments section. An anonymised summary table of the answers is presented in Table 17.

Looking into the results, we note that on average experts agree (>4) or Strongly Agree (>=4.5) that collision avoidance was able to avoid collisions, keep safe distance and provide a COLREG compliant solution for all tested scenarios. In addition, on average experts agree that the system response was real time. Regarding Hazardous Weather Routing service on average experts based on the information provided agree that the system correctly redirected vessels due to weather conditions, the proposed re-routing suggestion could be accepted and that all responses were provided in a timely manner. Experts marginally agree that they would also accept the full length of the suggested route as recommendation.

Table 17. Experts user questionnaire responses and SUS scores.

Question	R1	R2	R3	R4	R5	R6	R7	R8	R9	R10	R11	R12	R13	R14	R15	R16	R17	R18	R19	R20	AV.USER
3.1 Did the Collision Avoidance Service perform as expected (i.e., avoided collisions effectively)?	5	4	2	4	4	5	4	5	3	4	4	5	4	4	5	5	5	4	5	5	4
3.1 HEAD-ON	5	4	1	4	3	5	4	5	3	4	4	5	4	4	5	5	5	4	5	5	4.2
3.1 CROSSING	4	3	5	4	4	5	4	5	3	4	4	5	4	5	5	5	5	4	5	5	4.4
3.1 OVERTAKING	5	4	1	4	4	5	4	5	2	4	4	5	4	4	5	5	5	4	5	5	4.2
3.2 DID the Collision Avoidance Service consistently maintain a safe distance (greater than 2 m) when avoiding danger zones? -HEAD-ON	4	4	3	4	3	5	5	5	2	5	5	5	5	4	5	5	5	5	5	5	4.45
3.2 HEAD-ON	4	4	5	4	3	5	5	5	4	5	5	5	5	4	5	5	5	5	5	5	4.65
3.2 CROSSING	4	3	2	4	3	5	5	5	2	5	5	5	5	5	5	5	5	5	5	5	4.4
3.2 OVERTAKING	4	4	3	4	4	5	5	5	1	5	5	5	5	4	5	5	5	5	5	5	4.45
3.3Did the vessels consistently demonstrate COLREG-compliant	4	5	3	4	3	5	5	5	3	5	5	5	5	4	5	5	5	4	5	5	4.5
3.3 HEAD-ON	4	5	3	4	3	5	5	5	2	5	5	5	5	4	5	5	5	4	5	5	4.45
3.3 CROSSING	5	5	2	4	2	5	5	5	4	5	5	5	5	5	5	5	5	4	5	5	4.55
3.3 OVERTAKING	4	5	5	4	3	5	5	5	3	5	5	5	5	4	5	5	5	4	5	5	4.6
3.3. The LATENCY OF THE SYSTEM WAS REAL -TIME (SUB-SEC)	5	5	2	4	2	5	4	5	2	4	4	5	4	4	5	5	5	4	5	5	4.2
3.4 HEAD-ON	5	5	1	4	2	5	4	5	2	4	4	5	4	4	5	5	5	4	5	5	4.15
3.4 CROSSING	5	5	4	4	2	5	4	5	1	4	4	5	4	5	5	5	5	4	5	5	4.3
3.4 OVERTAKING	5	5	2	4	3	5	4	5	2	4	4	5	4	4	5	5	5	4	5	5	4.25
4. Hazardous Weather Routing Service																					
4.1 Are re-routing suggestions necessary?	4	5	1	3	3	5	5	5	2	5	5	5	5	5	5	5	5	5	5	5	4.4
4.2 Would you accept the re-routing suggestions?	5	5	5	3	2	5	4	5	4	4	4	5	4	5	5	5	5	3	5	5	4.4
4.3. WHOULD YOU ACCEPT THE FULL LENGTH OF THE SUGGESTED route?	4	4	3	3	3	5	3	5	2	3	3	5	3	5	5	5	5	3	5	5	3.95
4.4. the service WAS able to provide responses in a timely manner (less than an hour)?	4	4	2	3	4	5	4	5	3	4	4	5	4	5	5	5	5	5	5	5	4.3
5. System Usability (SUS)																					
5.1 I think that I would like to use this system frequently	4	4	3	3	3	5	4	5	3	4	4	5	4	5	5	5	5	5	5	5	4.3
5.2 I found the system unnecessarily complex	3	2	2	3	2	1	3	4	4	2	2	2	1	2	4	2	2	1	2	2	2.3
5.3. I thought the system was easy to use	4	3	2	3	3	5	3	4	3	3	3	4	3	5	4	5	5	4	4	4	3.7
5.4. I think that I would need the support of a technical person to be able to use this system	4	5	4	3	4	3	4	5	2	3	4	5	4	3	5	3	3	5	5	5	3.95
5.5 I found the various functions in this system were well integrated	4	4	1	3	3	5	4	4	4	4	4	4	4	5	4	5	5	4	4	4	3.95
5.6 I thought there was too much inconsistency in this system	2	3	4	3	2	1	5	3	1	4	3	3	3	1	3	1	1	3	2	2	2.5
5.7 I would imagine that most people would learn to use this system very quickly	2	4	5	3	4	5	3	4	3	3	4	4	4	5	4	5	5	4	4	4	3.95
5.8 I found the system very cumbersome to use	2	1	3	3	3	5	3	4	4	3	3	4	3	5	4	5	5	4	4	4	3.6
5.9 I felt very confident using the system	3	5	2	3	2	5	4	3	2	3	3	3	4	5	3	5	5	4	4	4	3.6
5.10 I needed to learn a lot of things before I could get going with this system	4	5	3	3	4	5	5	5	3	3	3	5	3	5	5	5	5	5	5	5	4.3
SUS SCORE	80	85	68	75	75	100	70	73	78	80	83	78	88	98	73	98	98	83	83	83	82.13

Regarding usability of the system the lowest score was 70 and the maximum score was 100. On average experts produce an 82.13 SUS score, depicting that the services demonstrated would provide useful information without putting too much additional strain on the on-board mariner.

At this point we must note some comments from the experts group mentioning the importance of human factor on-board and the necessity of mariners to be in control. According to them the existence of recommendation systems on-board even though are useful at the same time put pressure to mariners responsible for decision making, especially during a crisis handling situation.

Expert User Evaluation – Summary

——Collision Avoidance Service (CAS)

Experts strongly agreed that the CAS performed reliably across all scenarios.

- CAS avoided collisions effectively (avg. 4.2–4.4 / 5)
- Safe passing distance (>2 m) consistently maintained (avg. 4.4–4.6 / 5)
- Manoeuvres were COLREG-compliant in all tested situations (avg. 4.4–4.6 / 5)
- System responded in real time (sub-second latency) (avg. 4.1–4.2 / 5)

- Feedback aligns with quantitative pilot results (0 collisions, 0 violations)

Conclusion:

Experts validated that the CAS behaves safely, predictably, and in accordance with navigational rules.

——Hazardous Weather Routing Service

Experts agreed that the rerouting logic was correct, useful, and timely.

- Rerouting suggestions were necessary (4.4 / 5)
- Experts would accept the rerouting suggestions (≈4.0 / 5)
- Willingness to accept the full suggested route was positive but lower (3.95 / 5)
- Responses were timely (<1 hour) (4.3 / 5)

Conclusion:

The routing service is considered operationally sound; further refinements should target clarity of full-route suggestions.

—— System Usability (SUS)

Overall usability rated positively, with expected complexity for a research prototype.

Strengths:

- Experts would like to use the system frequently (4.3)
- System easy to use (4.0)
- Functions well integrated (3.95)
- Easy to learn for most users (3.95)
- Users felt confident with the system (3.6)

Areas for improvement:

- Perceived unnecessary complexity (2.3)
- Some inconsistency observed (2.5)
- Occasional cumbersome interactions (2.2)
- Moderate learning effort required (3.0)

Conclusion:

Usability is strong for the system; simplification and UI consistency should be prioritised for future iterations.

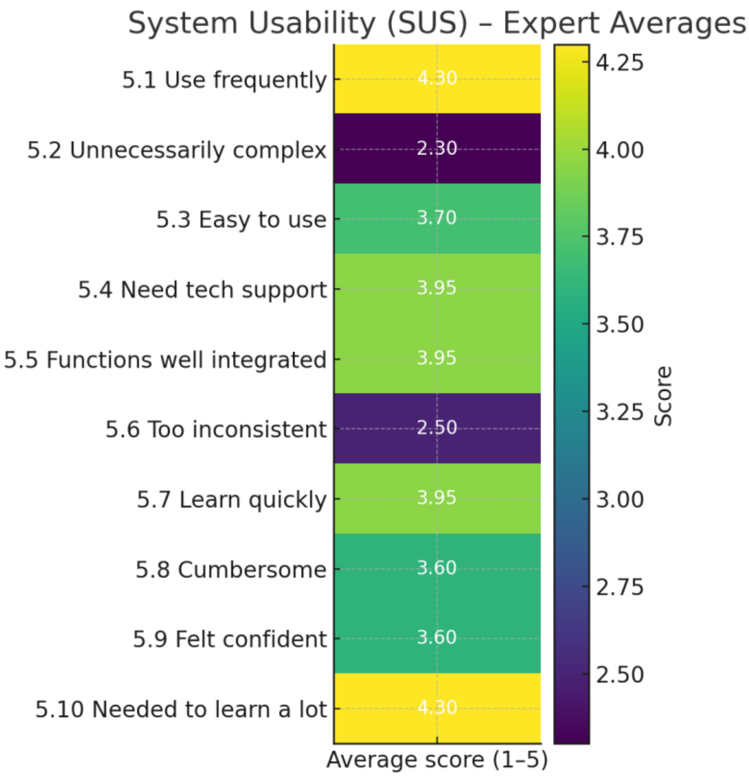
—— Overall Assessment

Expert consensus confirms the Maritime Use Case is mature, safe, and operationally credible.

- High trust in collision avoidance and COLREG-compliance
- Strong acceptance of rerouting logic and hazard-detection behaviour
- Positive user experience with clear opportunities for refinement
- Validation strengthened by participation of captains, instructors, VTS-related personnel, and senior Merchant Marine cadets

—— **Final conclusion:**

The system is viewed as a reliable, real-time, decision-support tool with clear potential for deployment on larger vessels and future integration into autonomous-shipping operations.



4.4 Lessons learned

In this section, we present key results achieved and conclusions reached for the maritime use case within the project.

Comparison Between Required KPIs and Achieved Results

—— **KPI (a): ≥ 80% accuracy in route forecasting / weather routing**

Achieved:

✔ 100% collision avoidance (all 137 incidents)

- ✔ 100% safe-zone avoidance
- ✔ COLREG compliance: 88.32%
- ✔ Rerouting accuracy: 83.94%

Conclusion:

The system exceeds the 80% accuracy requirement for route forecasting and weather routing.

Both quantitative results (CAS performance) and expert evaluation (>4/5 on all routing questions) confirm high accuracy and operational acceptance.

—— KPI (b): ≥ 80% accuracy in hazardous event detection / forecasting

Achieved:

- ✔ 0 collisions
- ✔ 0 safe-zone violations
- ✔ Early detection of collision risk in all 137 encounters
- ✔ Expert rating on collision avoidance effectiveness: 4.2–4.4/5
- ✔ Expert rating on safe-distance maintenance: 4.4–4.6/5

Conclusion:

The system achieves 100% hazard avoidance in all scenarios, outperforming the ≥80% requirement.

Both system data and expert feedback confirm consistent hazard identification, forecasting, and mitigation.

—— KPI (c): Sub-second latency in forecasting and detection

Achieved:

- ✔ CAS latency consistently sub-second during the 2025 pilot
- ✔ Expert assessment of real-time response: 4.1–4.2 / 5
- ✔ All rerouting and avoidance recommendations generated within real-time constraints

Conclusion:

The system meets the sub-second latency requirement, confirmed both instrumentally and by expert users.

——KPI (d): Forecast hazardous events ≥15 minutes before occurrence

Achieved:

- ✔ Collision-risk predictions generated well ahead of CPA (Closest Point of Approach)
- ✔ Routing predictions and hazard forecasts computed in advance
- ✔ In the weather-routing evaluation, experts rated timeliness of alerts: 4.3 / 5
- ✔ Speed-adjusted distance modelling shows hazards resolvable far in advance

Conclusion:

The behaviour-prediction components meet the KPI for early hazard forecasting, providing ample time-window for avoidance (>15 minutes in typical scenarios).

5 Extreme-Scale Performance of the Use Cases

Extreme data has been defined by Petcu and co-authors as “an incarnation of the Big Data concept distinguished by the massive amounts of data that must be queried, communicated and analysed in near real-time” [29]. In parallel, Big Data itself has been widely characterized as exhibiting five key aspects, namely Volume (the overall size of the data), Velocity (the speed at which data are generated and ingested), Variety (the heterogeneity of data types and formats), Veracity (issues of data quality and uncertainty), and, in distributed settings, Distribution (the way in which these data are spread across storage and compute resources) [30]. CREXDATA explicitly targets extreme-scale data along all these axes, but in different combinations across its three emergency management use cases.

In the Weather Emergencies Use Case, the emphasis is on Variety, Veracity, Velocity and Distribution. The Emergency Case demonstrator integrates heterogeneous live and historical sources (weather radar and forecast data, drone imagery, biometrics from first responders, and multilingual social media streams), all processed through distributed Kafka topics and RapidMiner workflows rather than a single centralized pipeline. Dynamic simulators (FloodWaive, Spark, Gazebo) are driven via REST APIs and produce extremely large GeoTIFF outputs at resolutions down to 1×1 m, which must be ingested, fused and visualized in ARGOS and Augmented Reality in near real time during flood and wildfire scenarios. The Data Injector further allows archived data streams (e.g., social media and biometrics) to be replayed either at original timestamps or accelerated frequencies, enabling realistic stress-testing of the platform under high event rates without depending on live crises.

Data processing is associated with considerable challenges resulting from data volume, even though it may be debatable whether this volume can already be classified as extreme. ARGOS processes approximately 25 GB of data for every ECWMF forecast received twice daily. The archive, available for all technologies involved, grows continuously. The FMI tool produces 22 GB of output with every single simulation run, leading to an archive size of several Terabytes. Furthermore, UAV imagery contributes significantly to the data load, as a single flight of half an hour by involving only one UAV generates approximately 11 GB of images for input into the NeRF processing pipeline.

Nevertheless, the relevance of the Emergency Case does not primarily result from the volume, but from the highly heterogeneous nature of the data sources, the extreme variety, and the extreme veracity of the available information. The integrated datasets comprise global data from services such as Copernicus and ECMWF as well as social media channels, complemented by diverse local data sources, including UAV sensor data, weather station measurements, sewer network data, and information obtained from biometric devices.

The datasets also exhibit pronounced variability in temporal and spatial resolution. Forecasts differ in temporal granularity, ranging from minute- and hour-based outputs to daily intervals or specialized time steps of 15 minutes. These are combined with sensor data gathered at frequencies as high as one image per second. Spatial harmonization represents an additional challenge, as context data such as terrain models with a resolution of 1x1 meter, sewer network data with precise geolocations, rainfall predictions with 1x1 kilometer resolution, and wind fields with 5x5 kilometer resolution must be integrated consistently.

Furthermore, the data exists in numerous formats that must be assimilated into streaming-based processing technologies. Although GeoTIFF serves as a standard for geospatial data exchange, it often proves insufficient for maintaining the semantic integrity of complex datasets across processing layers. The dimension of veracity constitutes an even more significant challenge than variety. The pre-processing of sewer network data exemplifies this issue, as data heterogeneity is compounded by missing values sensors located in flooded areas fail to transmit data while weather forecasts, inherently uncertain by nature, must nevertheless be transformed into predictive models for specific temporal windows. In the context of interactive learning and parameter optimization for flood scenario modeling, dynamic environmental factors such as parked vehicles or temporary construction sites can significantly influence the reliability of model outputs. Moreover, the volatility of datasets, including updates to sewer networks, variations in elevation models, or changes in rainfall predictions, imposes continuous adaptation requirements on the underlying processing frameworks.

The strengths of the ARGOS system lie in its capacity to handle variety and scalability while accommodating frequent updates, data homogenization and continuous integration. The ARGOS implementations for Dortmund and Innsbruck, in addition to the CREXDATA technologies, incorporate a wide spectrum of heterogeneous data sources. These include water level observations from the HYDRO Tyrol network and meteorological parameters precipitation, wind, and temperature from Tyrol weather stations and the German Weather Service, all updated on an hourly basis. Regional official warnings from national weather services are aggregated via MeteoAlarm with hourly refresh rates. Rainfall estimations are provided through Opera Radar from the European radar network every 15 minutes. Forecast data derived from the global GFS model encompass precipitation, temperature, wind speed and wind direction with hourly resolution for a five-day horizon and are updated every twelve hours. Similarly, forecasts based on the Austrian NWP model supply precipitation, temperature, and relative humidity data with equivalent temporal resolution and refresh rate, whereas GSA nowcasts deliver these variables on an hourly update cycle. Fire incident data for Austria are provided by Boku University, and supplementary information is obtained through the Copernicus Emergency Management Services. Within this, EFAS updated with each new forecast, typically twice daily provides reporting points, ERIC reporting points, and total precipitation at the 80th percentile. EFFIS, updated on the same schedule, offers the Fire Weather Index together with seven variables based on both ECWMF and MeteoFrance forecasts.

By contrast, the Health Crisis Use Case stresses mainly Volume and computational intensity, coupled with uncertainty-aware forecasting. In the multiscale infection scenario, CREXDATA executed approximately 5,400 coupled PhysiBoSS-Alya simulations representing patient-specific SARS-CoV-2 lung infection dynamics. Each simulation leveraged patient-derived 3D lung geometries (~4M mesh elements from CT scans) with different initial conditions on 64 full nodes of MareNostrum 5 (7,168 CPU cores and 14 TB RAM) over 1h 48min 27s modelling 10 days of lung infection dynamics and generating about 9.5 TB of multiscale output data per single parameter exploration simulation.

To make this tractable, early time-series classification (ETSC), developed in WP4, was used to terminate non-promising trajectories after detecting divergence from favorable outcomes, reducing the number of full simulations required by 62.5% while still satisfactorily characterizing the parameter space. This efficiency enabled exploration of vaccination status, intracellular signalling and virus distribution interactions that identified MAPK14 inhibition as a drug target improving epithelial survival by 48% in vaccinated virtual patients.

The epidemiological scenario processed high-resolution Spanish COVID-19 datasets (daily cases/hospitalisations/deaths by age/province; MITMA mobility matrices across 2,850 regions) through the episim-emews workflow. Covariance Matrix Adaptation Evolution Strategy (CMA-ES) calibration converged after 50,000 parameter evaluations versus 1,000,000 required for brute-force grid search, a 20-fold computational reduction. Visual analytics of Arrival Time of First Infection (ATFI) uncertainty across calibrated parameter ensembles revealed spatially coherent prediction patterns, with urban regions showing <5% variance versus >20% in rural areas.

Finally, the Maritime Use Case combines Velocity, Variety and Distribution at operational scale. The platform must process continuous AIS streams, on-board sensor feeds through the VDS device, high-resolution weather forecasts, and simulation outputs from Software-in-the-Loop (SITL) vessel digital twins to assess large numbers of potential collision and hazardous-weather situations in real time. These streams originate from the global AIS infrastructure of Kpler that includes more than 13,000 receivers across terrestrial, satellite and roaming systems deployed in over 190 countries, enabling continuous tracking of global maritime traffic from coastal zones to remote oceans. The Kpler AIS network monitors more than 350,000 vessels and processes over one billion raw AIS position messages every day, delivering vessel-position updates with latencies of only a few seconds and ensuring near real-time visibility of maritime activity worldwide. AIS messages include dynamic positional information (latitude, longitude, speed and course), static vessel characteristics, and voyage information such as destination ports and ETAs. The continuous broadcasting of large AIS data across global shipping lanes generates extremely high event rates that have to be ingested, filtered, processed and analysed in close to real-time to support safety critical maritime operations and services. Large-scale stress tests of the CREXDATA distributed architecture of Kpler operating in the Pekko framework confirmed that collision-avoidance and hazardous-weather routing services maintain sub-second end-to-end latency while delivering COLREG-compliant rerouting with over 80% accuracy, and expert mariners verified that the system proposed correct rerouting solutions in more than 80% of hazardous-weather cases during sea trials.

The vessel collision avoidance operates on two actor levels. Cell actors monitor proximity events within individual H3 grid cells and identify vessel pairs that might collide based on their predicted trajectories. In case of a forecasted collision, collision actors compute near-term collision probabilities using a high-frequency time step, invoking an external Python client to process collision-avoidance requests. During the evaluation for the same AIS dataset, 13,213 forecasted collision-avoidance events were processed. It is important to note that this number represents only the subset of cases where a valid collision risk was confirmed after applying the prediction and filtering logic of the collision detection and avoidance services. In practice, the number of active collision-monitoring actors created by the system is orders of magnitude larger, because collision actors are dynamically instantiated whenever the short-term trajectory prediction of a vessel intersects with the predicted trajectory of any other vessel. For each ingested AIS message, the system evaluates potential interactions between vessels located within the same spatial cell or neighbouring H3 cells. As a result, the number of candidate collision checks grows combinatorially with vessel density, since vessels can form multiple pairwise combinations with nearby traffic. This means that the actual number of actor instantiations and collision-monitoring computations is orders of magnitude larger than the 13,213 confirmed events, which correspond only to the cases where the algorithm determined that a collision probability exceeded the predefined threshold and therefore triggered a full collision-avoidance evaluation. Thus, the number of potential collision checks grows according to the pairwise interaction complexity $O(n^2)$, where *n* denotes the number of vessels present in a monitored region. In dense maritime traffic environments such as coastal approaches, ports, or narrow shipping lanes, even a modest increase in vessel density can therefore produce a quadratic growth in candidate vessel interactions that must be evaluated.

In this context, each decoded AIS message that holds the necessary fields (vessel identifier, longitude, latitude, timestamp, speed over ground, heading, course) for short-term forecast consists of approximately 70 bytes for deserialised records. The scenario at its extreme case where no message filtering is happening beforehand, a global AIS network captures every AIS message transmitted by all vessels around the globe and the collision avoidance detection follows a brute force processing approach (no spatial indexing or early elimination of distant pairs), would require handling pairwise combinations of 350K vessels every 2 seconds. This results in 61.2 billion combinations of pairs of records (140 bytes each pair), which corresponds to a data flow of 3.89 TB/second data for centralised processing. In modern vessels, a significant volume of sensor data is also handled at the edge (on board) that is not publicly accessible, and it is not included in the volumes mentioned earlier. Nevertheless, the system presented in the maritime use case handles data from both the AIS system and small autonomous vessels (ASVs). ASVs are scale models of larger vessels, scale boats used to safely assess the performance of the system in real-world conditions. ASVs' telemetry provides a gross estimate of processing requirements at the edge. In our setup, the VDS handles data from the onboard motion controller (Pixhawk) and the on-board camera (Raspberry Pi Camera module 3). The default specification for the motion controller produces approximately 4KB/s data that includes (GPS information, gyroscope information, battery level, ground control system parameters, and other sensor data directly connected). The onboard camera produces a 3750 KB/s data rate at a 720p120 configuration. The VDS output is configured to generate a similar to the AIS positional report record that corresponds to a 70 bytes/second data rate for each ASV. Incorporating edge processing data volume estimates can increase processing requirements by at least an order of magnitude.

The CREXDATA vessel weather-routing service relies on the integration of AIS traffic data with historical and forecast meteorological information. To support real-time routing decisions, the system performs extensive pre-processing of large historical datasets combining vessel trajectories with environmental conditions. In practice, the processing engine maintains in memory the results of preprocessing steps applied to multi-year datasets of AIS and weather information. One day of raw AIS data corresponds to 1–1.2 TB, while the associated environmental data integrated for maritime routing—consisting of wave, wind, and ocean-current forecasts of the NOAA/Copernicus services—amounts to roughly 1.5 GB per day, or about 550 GB per year. The preprocessing steps aggregated AIS positions with time-synchronised meteorological conditions and generate intermediate routing graphs used by the routing algorithm, enabling the system to perform routing computations with millisecond latency on stream for the entire global fleet of ca. 350.000 vessels.

Through evaluations in a large historical AIS streaming dataset consisting of 64.230.097 messages 12,682,140 vessel weather routing suggestions were generated for each valid ingested AIS position based on the reported vessel speed and destination port with an average 5.64ms processing time. The difference between the number of generated routes and the total number of AIS messages reflects both data-quality constraints and algorithmic restrictions. For example, routing suggestions cannot be produced when AIS messages lack a valid destination port or when vessels are already located within the spatial cell corresponding to their reported destination.

In conclusion the CREXDATA maritime use case, delivers the observed performance operationalizes Petcu's extreme data framing by converting continuous AIS emissions into time-critical collision and hazardous-weather decisions. Volume and Distribution are evident both upstream (Kpler AIS derives positions from 13,000+ receivers, tracks 350,000+ vessels daily, and Kpler reports >1 billion AIS messages processed per day) and downstream, where yearly preprocessing and in-memory/cached artefacts are on the order of TB are maintained as fused spatial indices and fast lookups for low-latency actor access. Variety and Veracity are addressed by a daily ETL step that projects NOAA/Copernicus forecasts onto an H3 grid (and merges fleet-intelligence traffic patterns) while on stream data-quality processes suppress invalid/noisy AIS attributes. Finally, although candidate vessel collision checks grow as $O(n^2)$ with local vessel density due to pairwise interactions, the actor-based Pekko processing engine delivers high performance through partitioning messages per vessel and per H3 cell, instantiating collision actors only when trajectory intersections are detected, and distributing large numbers of lightweight actors and manages to contain quadratic growth and preserve near-real-time responsiveness.

Across all three domains, these quantitative results on data volumes, event rates, and compute usage substantiate CREXDATA's claims to operate at extreme scale, complementing the qualitative expert evaluations and System Usability Scale scores reported elsewhere.

6 Conclusions

This report presents the final evaluation of the use cases, pilots, demonstrators, and simulation models and tools developed in the CREXDATA project. The assessment was carried out using a common methodology, ensuring coherence across the Weather Emergencies, Health Crisis, and Maritime Safety use cases.

Across all domains, the final evaluations confirm that CREXDATA's technologies have evolved from early prototypes into mature demonstrators with validated real-world applicability. Each use case has demonstrated the practical value of the developed tools: advanced forecasting and situational awareness in weather-induced emergencies, scalable epidemiological and patient-specific modelling in the health crisis domain, and enhanced navigation and hazard avoidance capabilities in maritime operations. These pilots and demonstrators substantiate the project's impact by showing how extreme-scale data analytics and simulation workflows can support complex decision-making processes.

The simulation models and interactive learning tools developed within the project have been central to this progress. They have enabled rigorous testing of hypotheses, calibration of predictive models, and systematic exploration of high-dimensional parameter spaces. Their integration into the demonstrator systems has ensured that decision-support outputs are grounded in validated, transparent computational methods.

The final scenario definitions delivered for each use case provide a consolidated description of the operational challenges addressed by CREXDATA and the technological solutions deployed. These scenarios were evaluated by with expert users and have guided the refinement of workflows, user interfaces, and data pipelines, ensuring alignment and stakeholder needs and real operational practices.

Overall, the results presented in this deliverable represent a significant milestone for the CREXDATA project. They demonstrate the successful implementation and validation of the project's technologies and provide a consolidated view of their performance at the end of the project. As CREXDATA moves towards its post-project phase, the insights gained from this evaluation will support future exploitation, ensure long-term sustainability of the tools, and pave the way for continued research at the intersection of HPC, AI, and crisis management.

7 Acronyms and Abbreviations

Each term should be bulleted with a definition.

Below is an initial list that should be adapted to the given deliverable.

- CA – Consortium Agreement
- D – deliverable
- DoA – Description of Action (Annex 1 of the Grant Agreement)
- EB – Executive Board
- EC – European Commission
- GA – General Assembly / Grant Agreement
- HPC – High Performance Computing
- IPR – Intellectual Property Right
- KPI – Key Performance Indicator
- M – Month
- MS – Milestones
- PM – Person month / Project manager
- WP – Work Package
- WPL – Work Package Leader

8 References

[1] N. Koenig and A. Howard, 'Design and use paradigms for gazebo, an open-source multi-robot simulator', *2004 IEEERSJ Int. Conf. Intell. Robots Syst. IROS IEEE Cat No04CH37566*, vol. 3, pp. 2149–2154, 2004, doi: 10.1109/IROS.2004.1389727.

[2] M. Quigley, J. Faust, T. Foote, J. Leibs, and others, 'ROS: an open-source Robot Operating System', in *ICRA Workshop on Open Source Software*, 2009.

[3] I. Gräßler, N. Döhner, M. Ebel, and J. Pottebaum, 'Shifting boundaries from preparedness to response: Using simulation of rescue robots in weather-induced emergencies', presented at the Mensch und Computer 2025 - Workshopband, Gesellschaft für Informatik e.V., 2025, p. 10.18420/muc2025. Accessed: Dec. 18, 2025. [Online]. Available: https://dl.gi.de/handle/20.500.12116/46841

[4] M. Ponce-de-Leon *et al.*, 'COVID19 Flow-Maps Daily Cases Reports'. Zenodo, Aug. 18, 2021. doi: 10.5281/zenodo.5217386.

[5] Miguel Ponce-de-Leon and Isabel Martinez, 'episim-models collection CREXDATA Release', Nov. 06, 2025, *Zenodo*. doi: 10.5281/ZENODO.17544616.

[6] Miguel Ponce-de-Leon, Lewis Knox, and Isabel Martinez, *EpiSim.jl Simulator CREXDATA Release V0.2.3*. (Nov. 07, 2025). Zenodo. doi: 10.5281/ZENODO.17552713.

[7] A. Arenas *et al.*, 'Modeling the Spatiotemporal Epidemic Spreading of COVID-19 and the Impact of Mobility and Social Distancing Interventions', *Phys. Rev. X*, vol. 10, no. 4, p. 041055, Dec. 2020, doi: 10.1103/PhysRevX.10.041055.

[8] P. Castioni, S. Gómez, C. Granell, and A. Arenas, 'Rebound in epidemic control: how misaligned vaccination timing amplifies infection peaks', *Npj Complex.*, vol. 1, no. 1, p. 20, Nov. 2024, doi: 10.1038/s44260-024-00020-0.

[9] Miguel Ponce-de-Leon, IsaMtz16, and Pedro Velho, *Epi-Sim/episim-emews: EpiSim-EMEWS CREXDATA Release*. (Nov. 07, 2025). Zenodo. doi: 10.5281/ZENODO.17553481.

[10] J. Ozik, N. T. Collier, J. M. Wozniak, and C. Spagnuolo, 'From desktop to large-scale model exploration with SWIFT/T', *Proc. Winter Simul. Conf. Winter Simul. Conf.*, vol. 2016, pp. 206–220, Dec. 2016, doi: 10.1109/WSC.2016.7822090.

[11] Miguel Ponce-de-Leon, Isabel Martinez, and Charilaos Akasiadis, *Epi-Sim/episim-rl: EpiSim-RL CREXDATA Release*. (Nov. 07, 2025). Zenodo. doi: 10.5281/ZENODO.17553313.

[12] M. Ponce-de-Leon, C. Pontes, A. Arenas, and A. Valencia, 'Assessing the impact of interregional mobility on COVID19 spread in Spain using transfer entropy', *Sci. Rep.*, vol. 15, no. 1, p. 31504, Aug. 2025, doi: 10.1038/s41598-025-17218-4.

[13] A. Bangor, P. Kortum, and J. Miller, 'Determining what individual SUS scores mean: adding an adjective rating scale', *J Usability Stud.*, vol. 4, no. 3, pp. 114–123, May 2009.

[14] M. Vázquez *et al.*, 'Alya: Multiphysics engineering simulation toward exascale', *J. Comput. Sci.*, vol. 14, pp. 15–27, May 2016, doi: 10.1016/j.jocs.2015.12.007.

[15] M. Ponce-de-Leon *et al.*, 'PhysiBoSS 2.0: a sustainable integration of stochastic Boolean and agent-based modelling frameworks', *Npj Syst. Biol. Appl.*, vol. 9, no. 1, Art. no. 1, Oct. 2023, doi: 10.1038/s41540-023-00314-4.

[16] M. Liao *et al.*, 'Single-cell landscape of bronchoalveolar immune cells in patients with COVID-19', *Nat. Med.*, vol. 26, no. 6, pp. 842–844, June 2020, doi: 10.1038/s41591-020-0901-9.

[17] J. Béal, A. Montagud, P. Traynard, E. Barillot, and L. Calzone, 'Personalization of logical models with multi-omics data allows clinical stratification of patients', *Front. Physiol.*, vol. 9, p. 1965, Jan. 2019, doi: 10.3389/fphys.2018.01965.

[18] A. Montagud *et al.*, 'Patient-specific Boolean models of signalling networks guide personalised treatments', *eLife*, vol. 11, p. e72626, Feb. 2022, doi: 10.7554/eLife.72626.

[19] P. Lo Surdo *et al.*, 'SIGNOR 3.0, the SIGnaling network open resource 3.0: 2022 update', *Nucleic Acids Res.*, vol. 51, no. D1, pp. D631–D637, Jan. 2023, doi: 10.1093/nar/gkac883.

[20] M. I. El-Gamal, H. S. Anbar, H. Tarazi, and C.-H. Oh, 'Discovery of a potent p38α/MAPK14 kinase inhibitor: Synthesis, in vitro/in vivo biological evaluation, and docking studies', *Eur. J. Med. Chem.*, vol. 183, p. 111684, Dec. 2019, doi: 10.1016/j.ejmech.2019.111684.

[21] M. G. Rolf *et al.*, 'In vitro pharmacological profiling of R406 identifies molecular targets underlying the clinical effects of fostamatinib', *Pharmacol. Res. Perspect.*, vol. 3, no. 5, p. e00175, Oct. 2015, doi: 10.1002/prp2.175.

[22] D. Zissis *et al.*, 'The Aegean Ro-Boat Race: A Regatta to Accelerate Innovation in Marine Robotics and Autonomous Vessels', *IEEE Aerosp. Electron. Syst. Mag.*, vol. 39, no. 9, pp. 36–45, Sept. 2024, doi: 10.1109/MAES.2024.3425317.

[23] G. Spiliopoulos *et al.*, 'GMSA: A Digital Twin Application for Maritime Route and Event Forecasting', 2024.

[24] A. Troupiotis-Kapeliaris, D. Zissis, S. Kohlbrenner, and C. Kastrisios, 'Mobility Data Mining: the Maritime Use Case', in *2024 11th International Workshop on Metrology for AeroSpace (MetroAeroSpace)*, Lublin, Poland: IEEE, June 2024, pp. 207–212. doi: 10.1109/MetroAeroSpace61015.2024.10591577.

[25] A. Troupiotis-Kapeliaris, C. Kastrisios, and D. Zissis, 'Vessel Trajectory Data Mining: A Review', *IEEE Access*, vol. 13, pp. 4827–4856, 2025, doi: 10.1109/ACCESS.2025.3525952.

[26] N. Andrienko, G. Andrienko, D. Zissis, and G. Spiliopoulos, 'Techniques for interactive visual examination of autonomous vessel performance', *Proc. VLDB Endow. ISSN*, vol. 2150, p. 8097, 2024.

[27] N. Andrienko, G. Andrienko, D. Zissis, A. Troupiotis-Kapeliaris, and G. Spiliopoulos, 'Techniques for interactive visual examination of vessel performance', *Big Data Res.*, vol. 43, p. 100575, Feb. 2026, doi: 10.1016/j.bdr.2025.100575.

[28] A. Troupiotis-Kapeliaris *et al.*, 'Collection: The Aegean Ro-Boat Race 2023', *IEEE Data Descr.*, vol. 1, pp. 87–94, 2024, doi: 10.1109/IEEEDATA.2024.3475332.

[29] D. Petcu *et al.*, 'On Processing Extreme Data', *Scalable Comput. Pract. Exp.*, vol. 16, no. 4, pp. 467–490, Jan. 2016, doi: 10.12694/scpe.v16i4.1134.

[30] I. Emmanuel and C. Stanier, 'Defining Big Data', in *Proceedings of the International Conference on Big Data and Advanced Wireless Technologies*, in BDAW '16. New York, NY, USA: Association for Computing Machinery, Nov. 2016, pp. 1–6. doi: 10.1145/3010089.3010090.

[31] J. Brooke, 'SUS: A "Quick and Dirty" Usability Scale', in *Usability Evaluation In Industry*, CRC Press, 1996.

[32] J. Sauro, 'SUStisfied? Little-known system usability scale facts. User experience', *Mag. User Exp. Prof. Assoc.*, vol. 10, no. 3, 2011.

9 Appendix 1: Expert users' questionnaire of the Weather-Induced Emergencies Use Case

9.1 Test Cases and Test Scenarios

The following Test Cases are added to the list of Test Cases introduced in D2.2.

Test_Case_Name	**Simulation Parametrization with EmCase SHS GUI**
Test_Case_ID	TC_016
Test_Item	SHS GUI
Test_Procedure	Physical Test
CREXDATA_system_level	sub-system
Test_Case_Owner	UPB
Use_Case_underlaying	Emergency_UC_01
Detailed_Description	
Environmental_Conditions	
Test_Case_ Environmental_Parameter	simulator specific input and context data
System_State	SHS GUI active and connected to kafka
User	B level
Input Parameter	Simulator specific parameter • FloodWaive: weather data, date/time, rain intensity, rain duration, resolution, simulation area, elevation model, elevation change • Spark: fire starting point, form of starting point, size of starting point, wind speed, wind direction, temperature, relative humidity • Gazebo: start point, end point, target points, wind speed, wind direction, drone model, simulation speed, number of routes, optimization criterion
Test_Case_Sequence	1. Starting ARGOS system 2. Choose ARGOS project (city of Dortmund) 3. Select SHS GUI 4. Select simulation meta data 5. Select simulator 6. Define parameter 7. Send parameter
Expected_Result	triggered simulation model

Test_Case_Name	**Spark Wildfire Simulation**
Test_Case_ID	TC_017
Test_Item	Spark
Test_Procedure	Physical Test
CREXDATA_system_level	sub-system
Test_Case_Owner	UPB
Use_Case_underlaying	Emergency_UC_01
Detailed_Description	
Environmental_Conditions	
Test_Case_ Environmental_Parameter	simulator specific input and context data

System_State	Spark active and logged in
User	B level
Input Parameter	fire starting point, form of starting point, size of starting point, wind speed, wind direction, temperature, relative humidity
Test_Case_Sequence	1. Select SHS GUI 2. Select simulation meta data 3. Select simulator 4. Define parameter 5. Send parameter 6. Run simulation 7. Send results to ARGOS 8. Visualize simulation results in ARGOS
Expected_Result	visualized simulation

Test_Case_Name	**Gazebo Robotic Simulation**
Test_Case_ID	TC_018
Test_Item	Gazebo
Test_Procedure	Physical Test
CREXDATA_system_level	sub-system
Test_Case_Owner	UPB
Use_Case_underlaying	Emergency_UC_01
Detailed_Description	
Environmental_Conditions	
Test_Case_ Environmental_Parameter	simulator specific input and context data
System_State	Virtual Machine running, gazebo active and logged in
User	B level
Input Parameter	start point, end point, target points, wind speed, wind direction, drone model, simulation speed, number of routes, optimization criterion
Test_Case_Sequence	1. Select SHS GUI 2. Select simulation meta data 3. Select simulator 4. Define parameter 5. Send parameter 6. Run simulations 7. Rank drone routes 8. Send results to ARGOS 9. Visualize simulation results in ARGOS
Expected_Result	visualized simulations

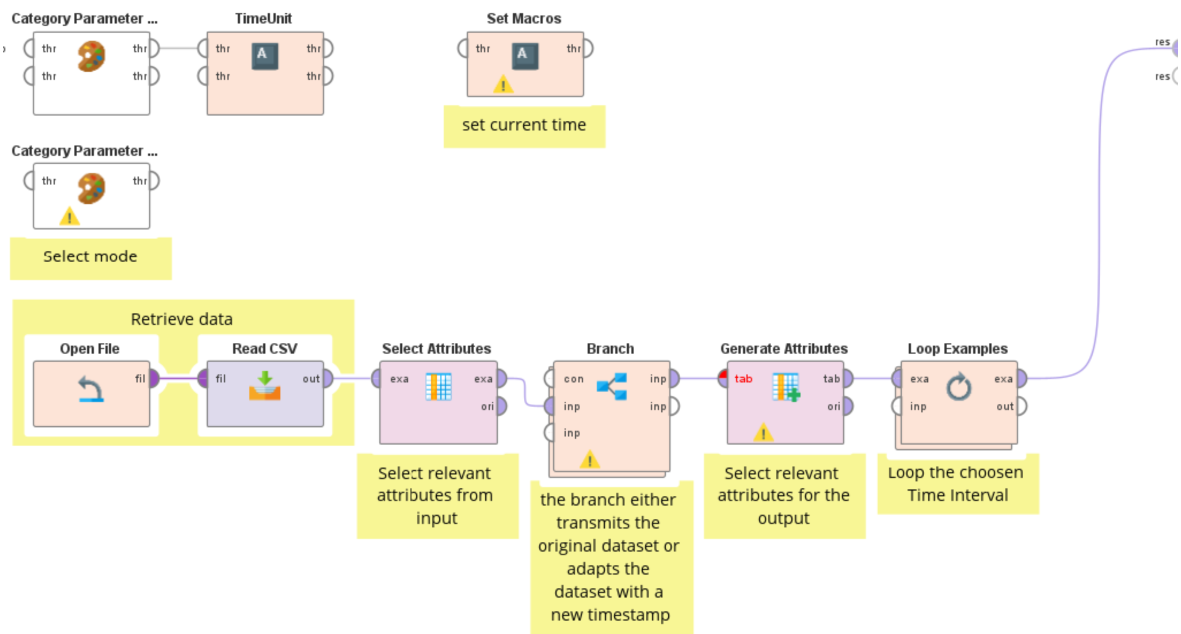
Test_Case_Name	**FloodWaive Flooding Simulation**
Test_Case_ID	TC_019
Test_Item	FloodWaive
Test_Procedure	Physical Test
CREXDATA_system_level	sub-system
Test_Case_Owner	UPB
Use_Case_underlaying	Emergency_UC_01
Detailed_Description	
Environmental_Conditions	

Test_Case_ Environmental_Parameter	simulator specific input and context data
System_State	FloodWaive API connected
User	B level
Input Parameter	weather data, date/time, rain intensity, rain duration, resolution, simulation area, elevation model, elevation change
Test_Case_Sequence	1. Select SHS GUI 2. Select simulation meta data 3. Select simulator 4. Define parameter 5. Send parameter 6. Run simulation 7. Send simulation ID to ARGOS 8. ARGOS requests simulation results 9. Visualize simulation results in ARGOS
Expected_Result	visualized simulations

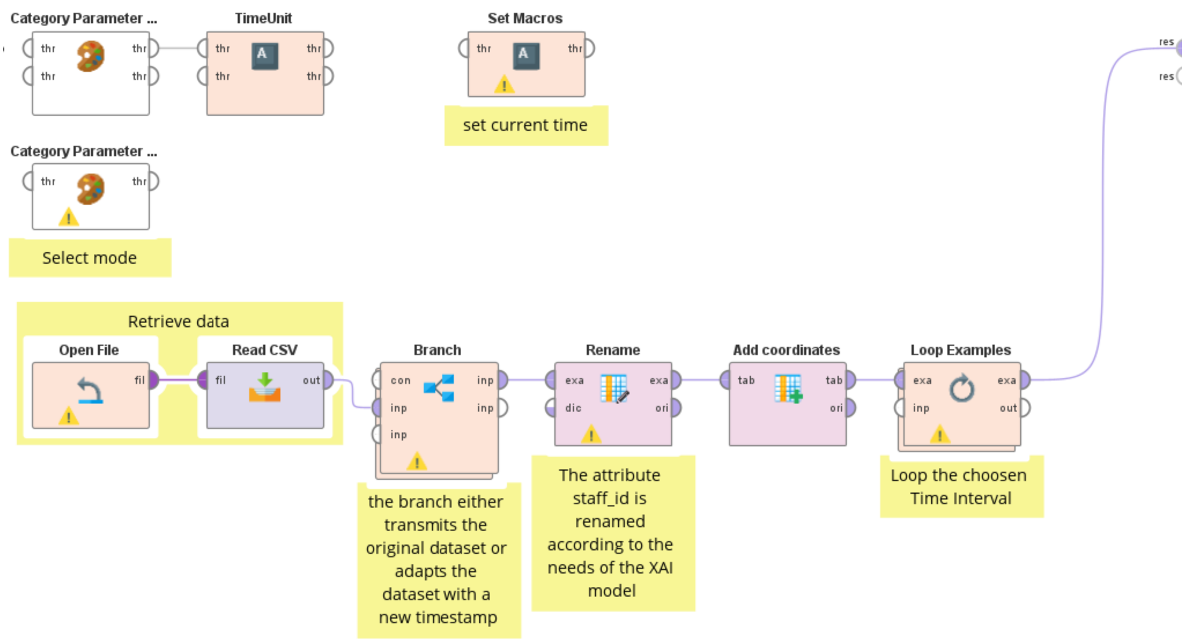
9.2 Workflows within the EmCase

9.2.1 Data Injector

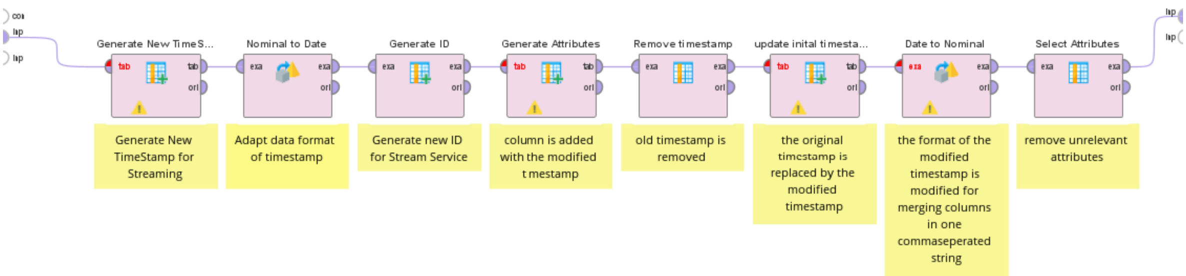
Data Injector for Social Media posts



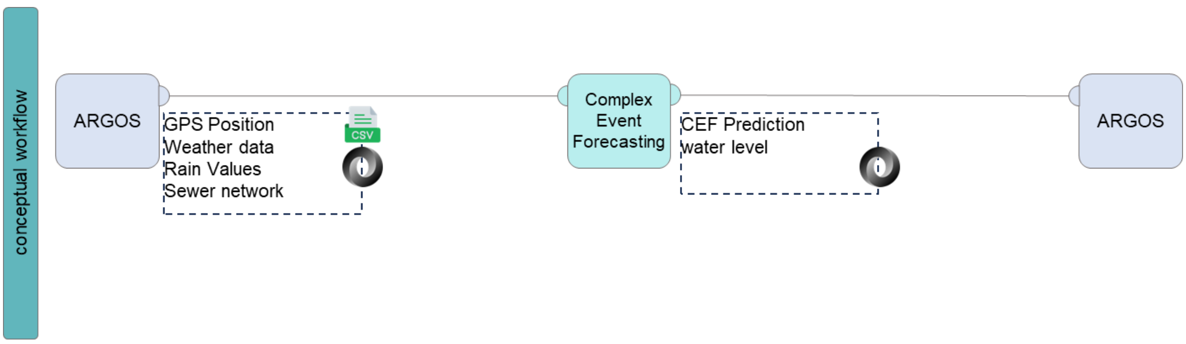
Data Injector for biometric data



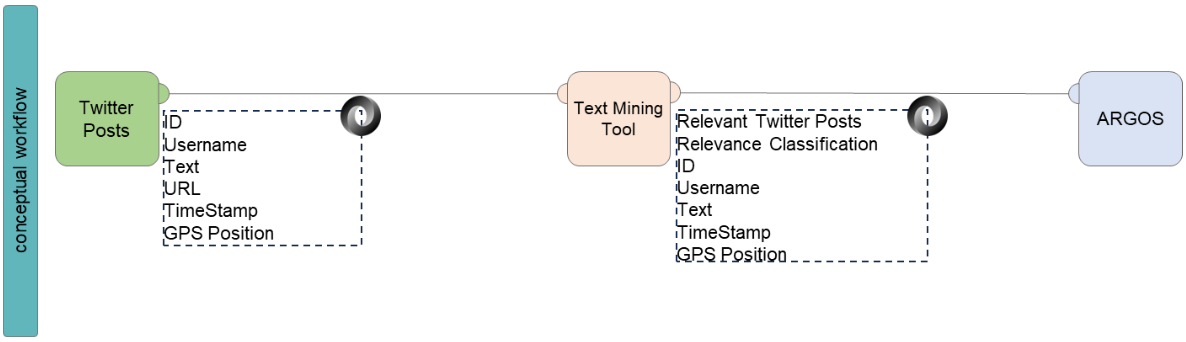
Insights into the branch: the input is either directly transmitted in case the original timestamp should be kept, or the following subprocess is executed within the else branch.



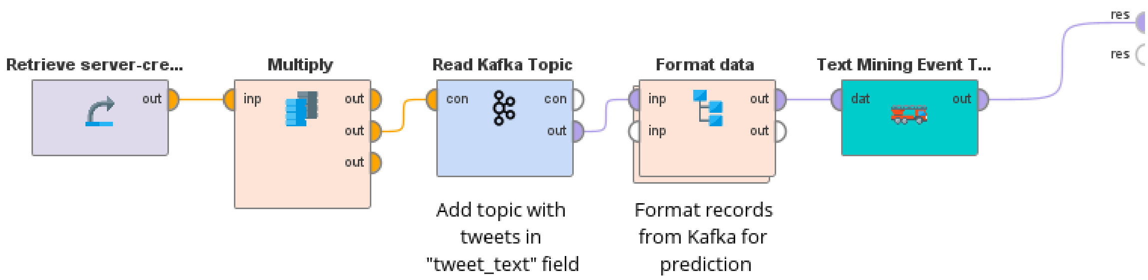
9.2.2 Complex Event Forecasting



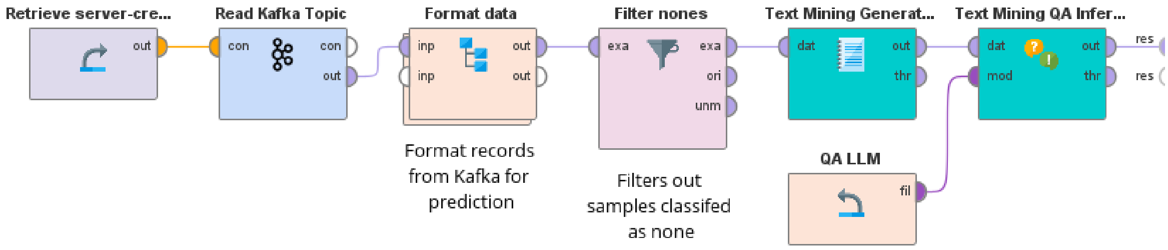
9.2.3 Text Mining



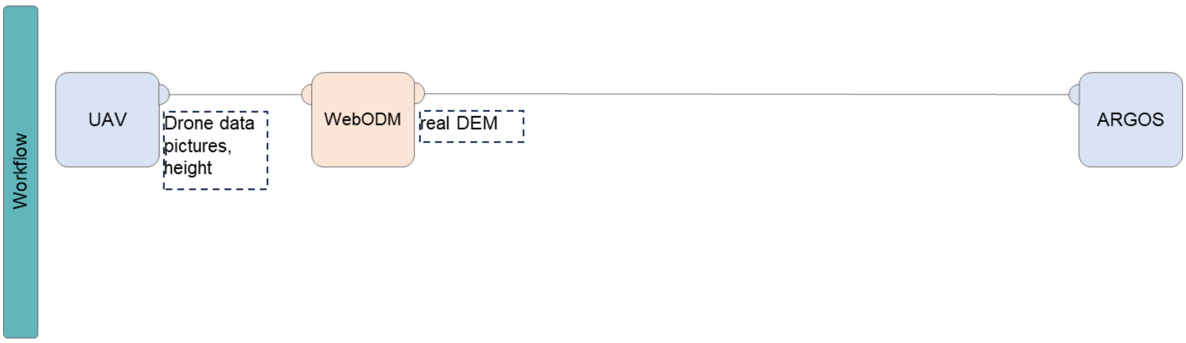
Workflow for event type prediction for social media posts



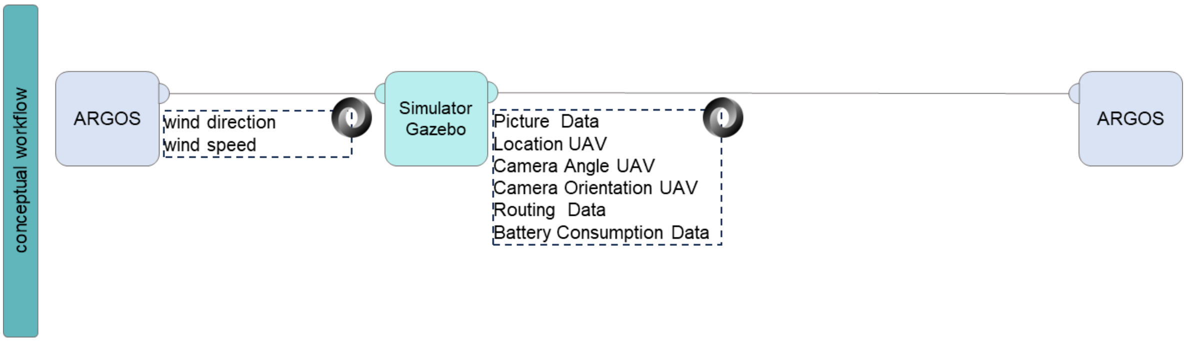
Workflow for question-and-answering functionality based on classified posts

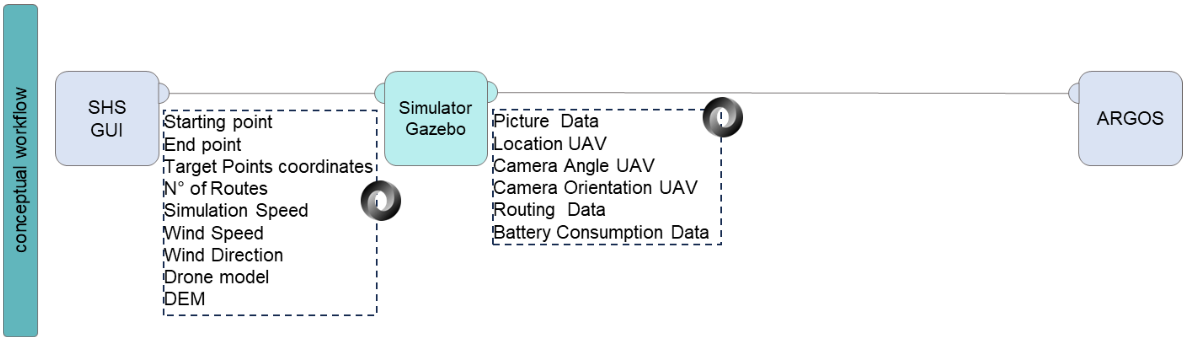


9.2.4 WebODM

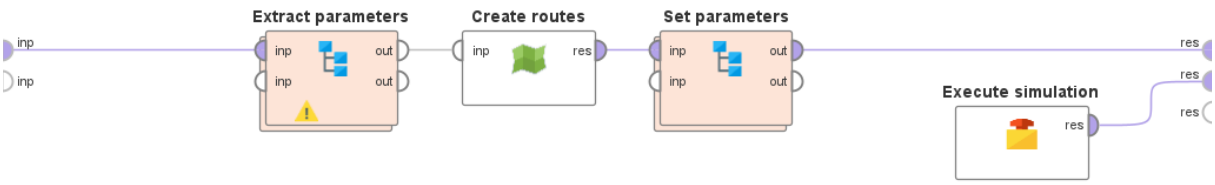


9.2.5 Gazebo Simulator

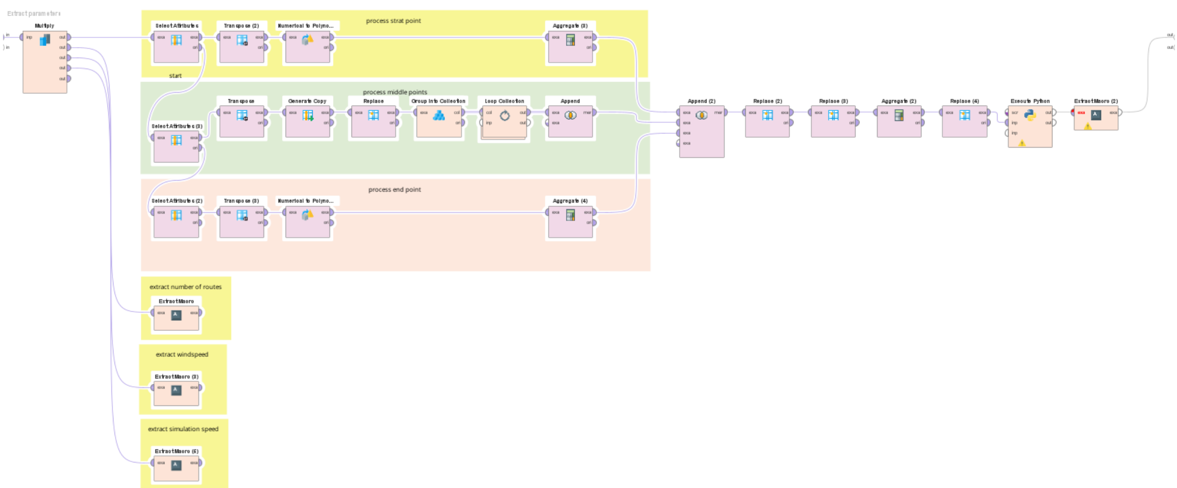




Gazebo workflow for parameterization and simulation execution

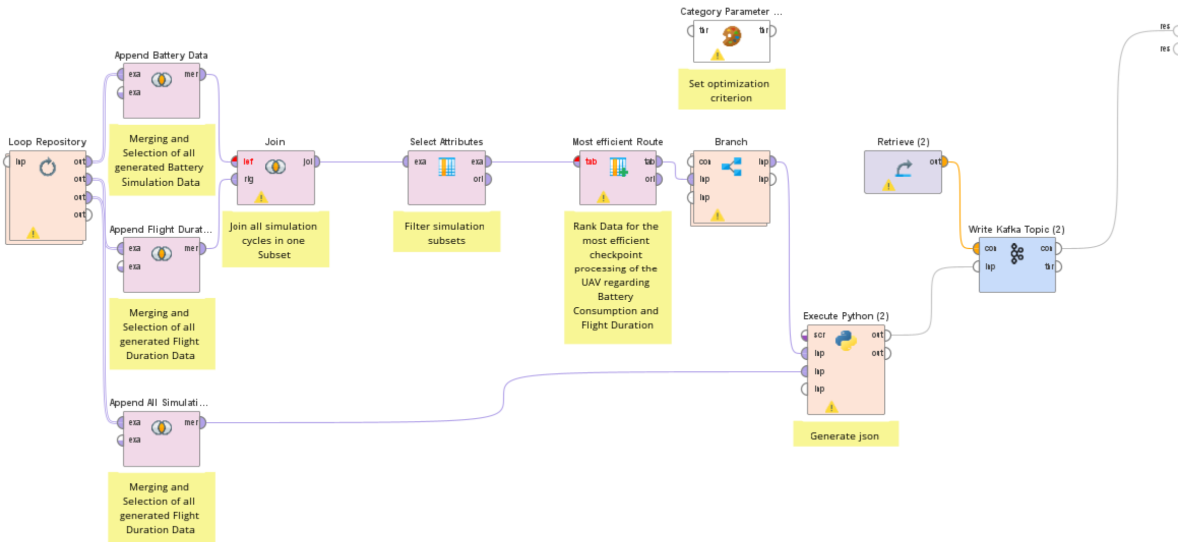


Subprocess « extract parameters »

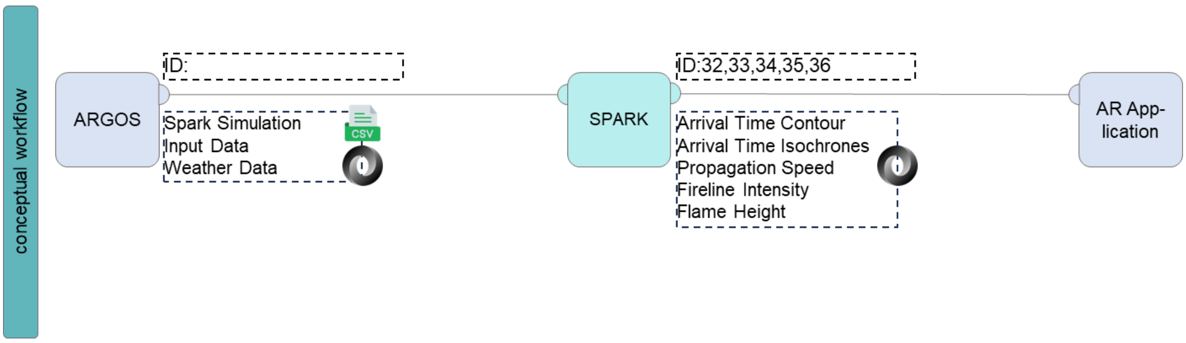


Gazebo workflow for retrieving simulation results

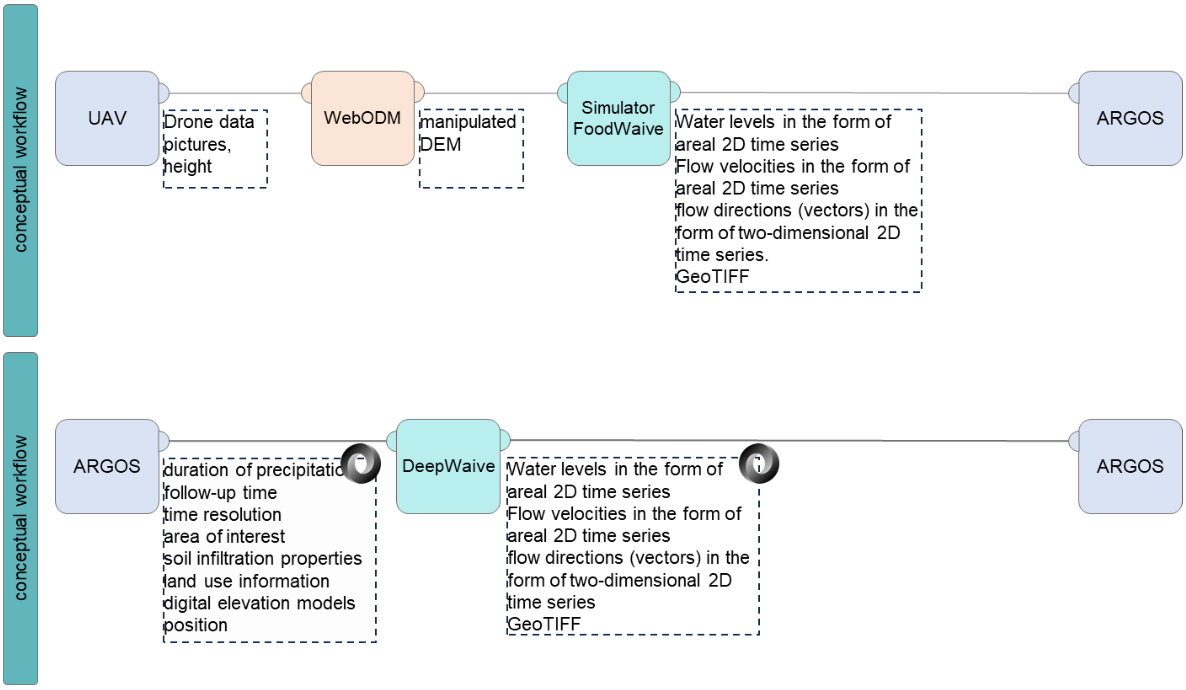
Gazebo workflow for simulation results processing



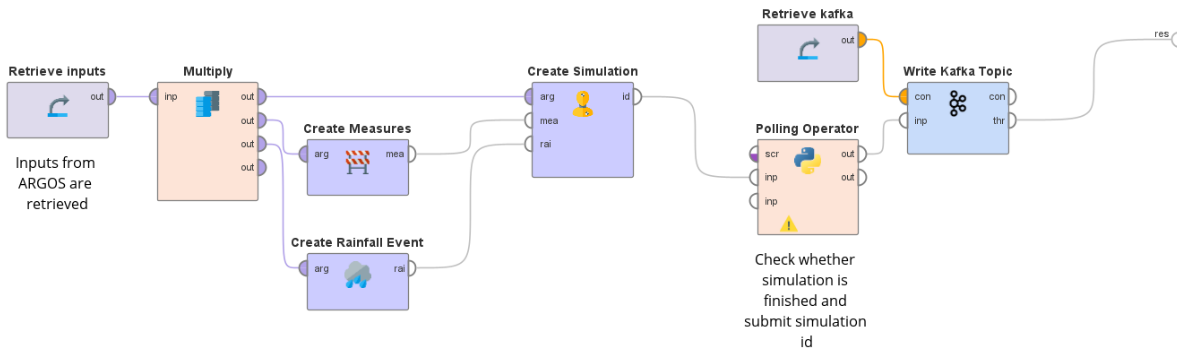
9.2.6 SPARK Simulator



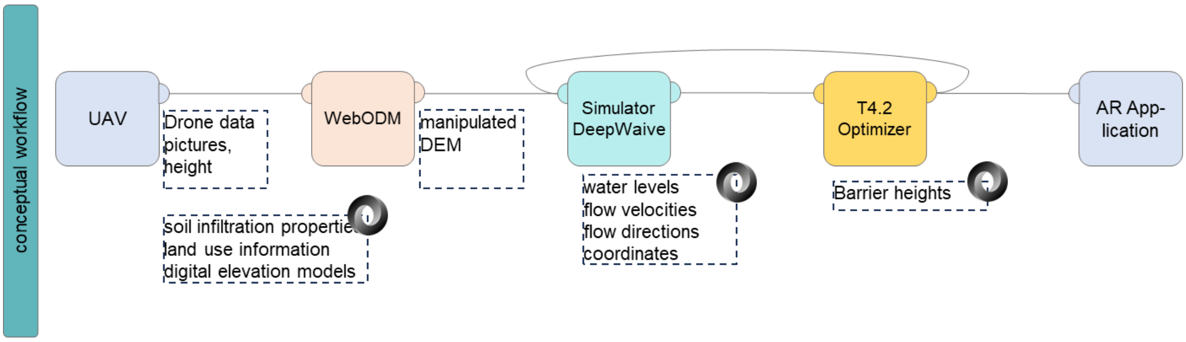
9.2.7 FloodWaive Simulator



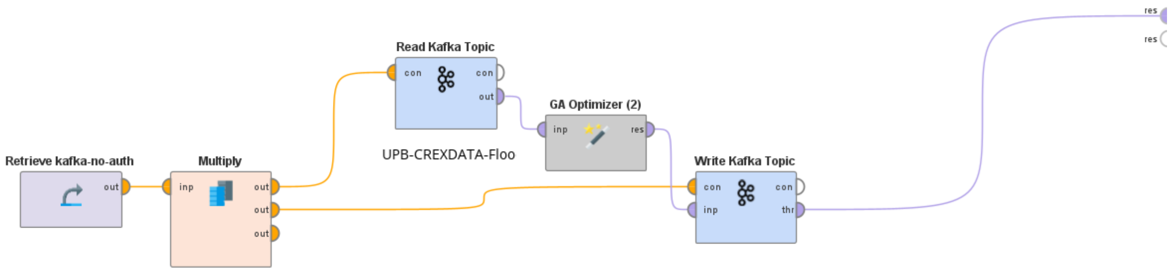
Workflow for creating a flood simulation with rainfall data and measures from ARGOS



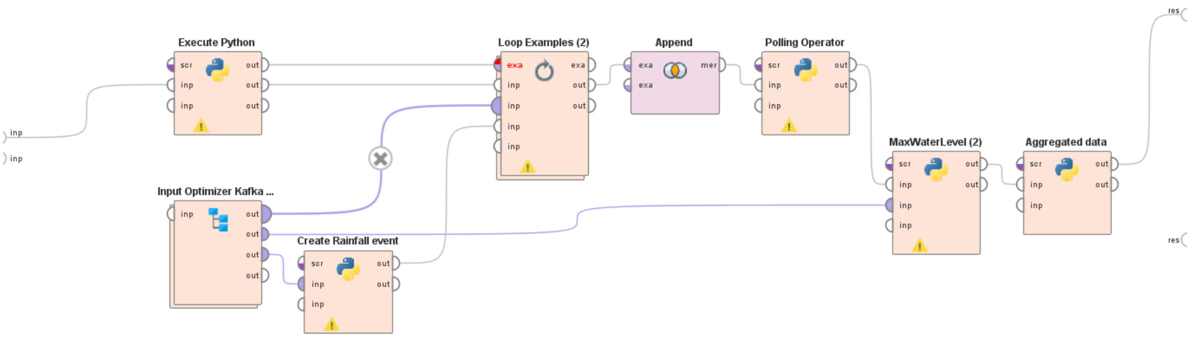
9.2.8 Barrier Height Optimizer



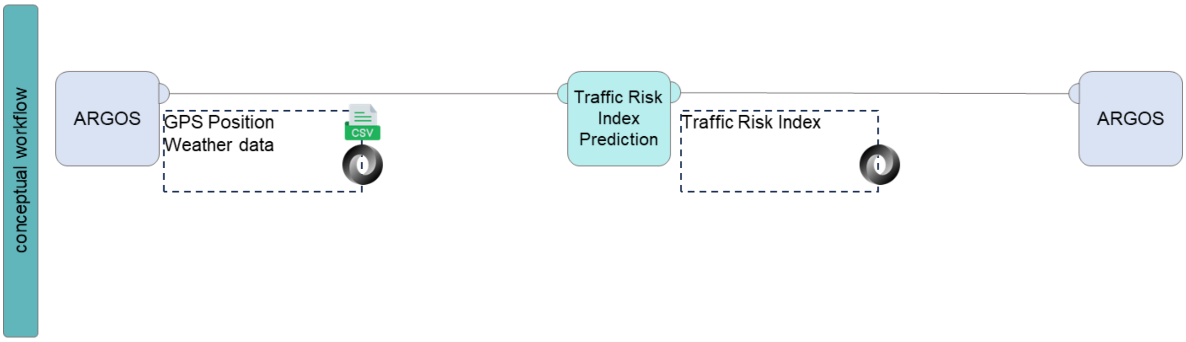
Workflow to configure and start the optimizer



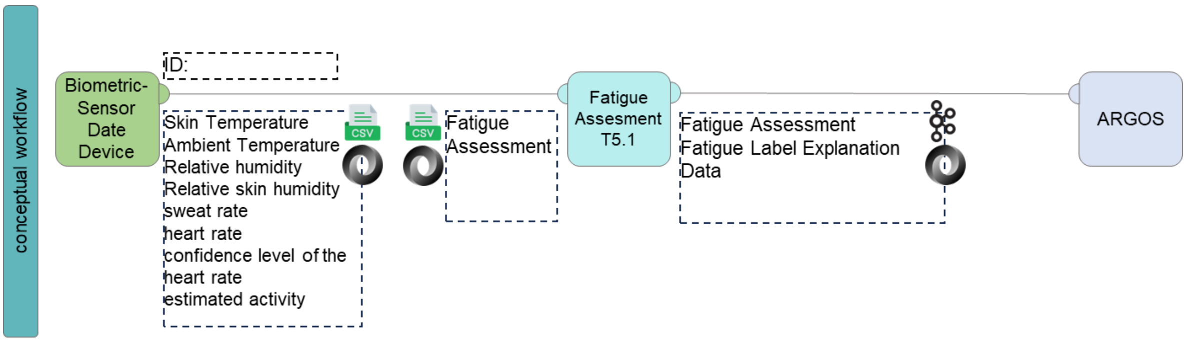
Workflow to start flood simulations and extract the maximum water level



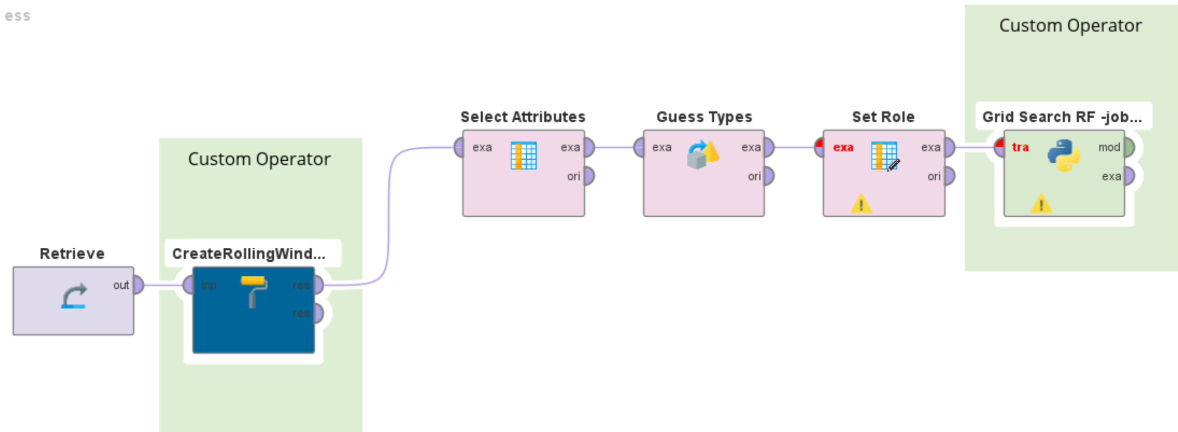
9.2.9 Traffic Risk Index Prediction



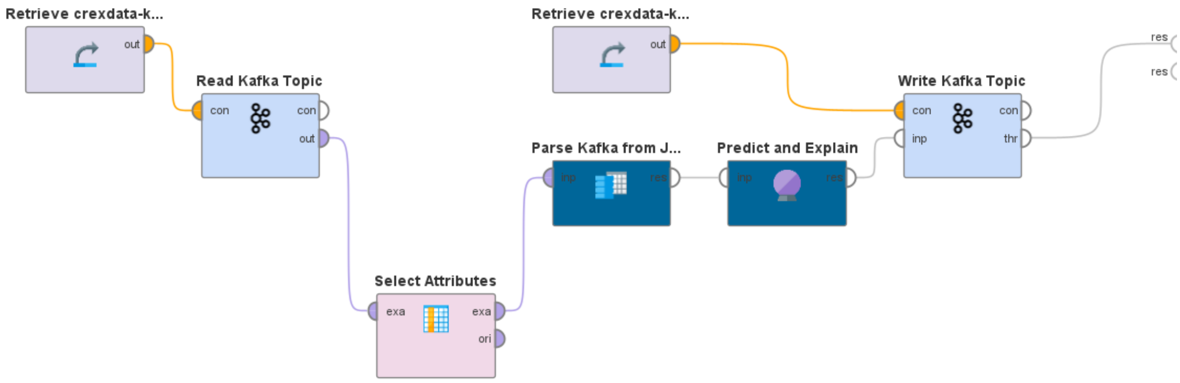
9.2.10 Fatigue Assessment xAI



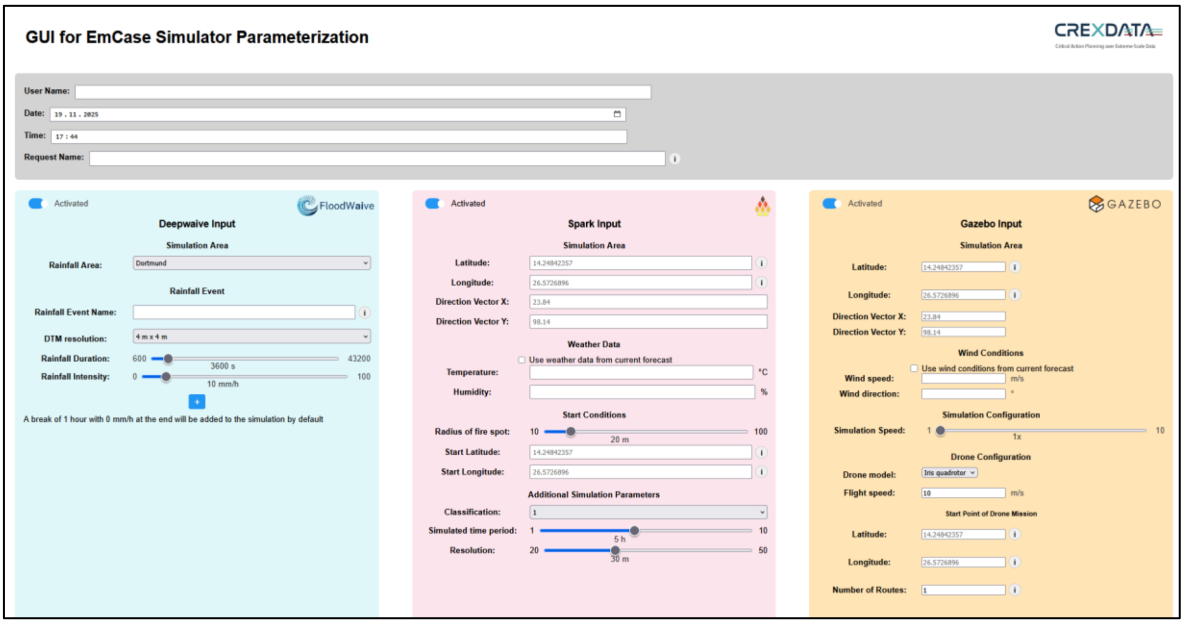
Workflow for model training based on biometric data



Workflow for fatigue assessment based on biometric streaming data, applying the trained model



9.3 EmCase Simulator HyperSuite GUI



9.4 Results of Text Mining & Gazebo online lab test

expert #	1	2	3	4	main
The interactions offered by text mining helped me analyze the situation.	5	3	-	4	**4,00**
The functions offered by text mining helped me gather relevant information in a targeted manner.	4	4	-	3	**3,67**
I could clearly see which data or assumptions were used to decide whether a post was relevant.	4	1	-	3	**2,67**
The information provided in ARGOS would influence my decisions during an operation.	5	4	4	4	**4,25**

The system's warnings and recommendations are helpful for mission planning.	4	4	5	4	**4,25**
I had confidence in the forecasts and data provided by the system.	4	4	5	4	**4,25**
I would recommend the use of ARGOS for real heavy rainfall events.	4	4	5	4	**4,25**
I would regularly follow the system's flight planning recommendations even in strong winds or uncertain conditions.	3	2	-	4	**3,00**
I found the system's route recommendations difficult to understand in difficult conditions (e.g., crosswinds, multiple destinations).	2	4	-	3	**3,00**
I would rely on the system's recommendations for flight planning.	3	3	-	4	**3,33**
The route adjustment recommendations seemed reasonable and can be incorporated into the decision-making process.	4	3	-	4	**3,67**
I found the system's suggestions in critical weather situations to be contradictory or confusing.	2	3	-	2	**2,33**
I found it difficult to trust the system's flight decisions in the event of significant environmental changes.	2	3	-	2	**2,33**
I would feel confident in accepting the suggested route despite existing uncertainties.	2	3	-	3	**2,67**

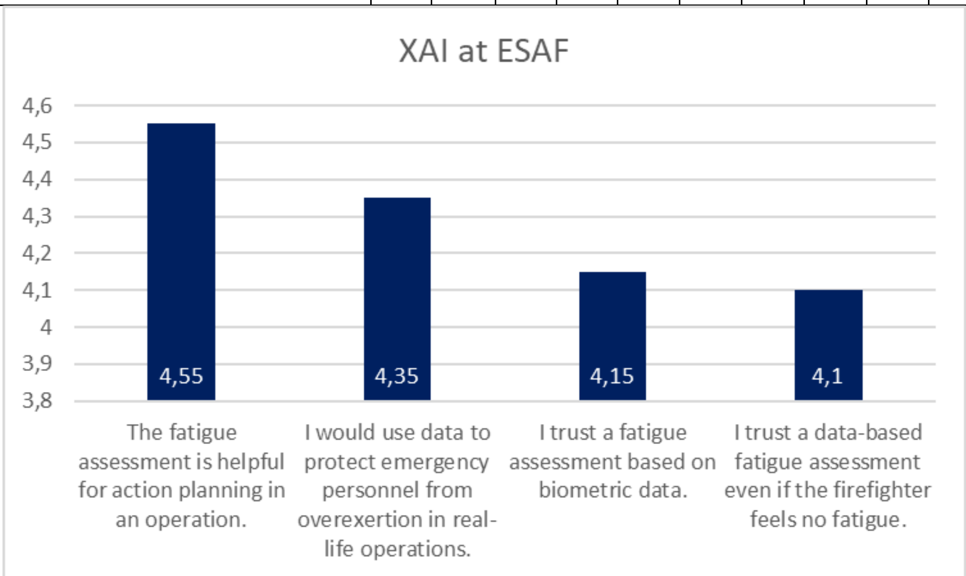
9.5 Results of ESAF R&D Days

FMI tool for emergency tasks (scale from 1 – strongly disagree to 5 – strongly agree)

expert #												**main**
The impact forecast from the tool for emergency tasks would help my decision-making in weather-related emergencies and would add value compared to traditional weather forecasts.	3	4	5	5	5	5	5	5	5	4	5	4,64
The tool for emergency tasks would be useful support for planning the operational work.	5	5	5	5	5	5	5	5	5	4	4	4,82
The tool for emergency tasks is trusthworfty and easily understandable.	4	5	5	3	5	5	5	4	5	4	3	4,36
I would recommend the tool for emergency tasks for other users as well.	5	5	5	5	5	5	5	5	5	5	5	5,00

eXplainable AI based on fatigue assessment with biometrical data (scale from 1 – strongly disagree to 5 – strongly agree)

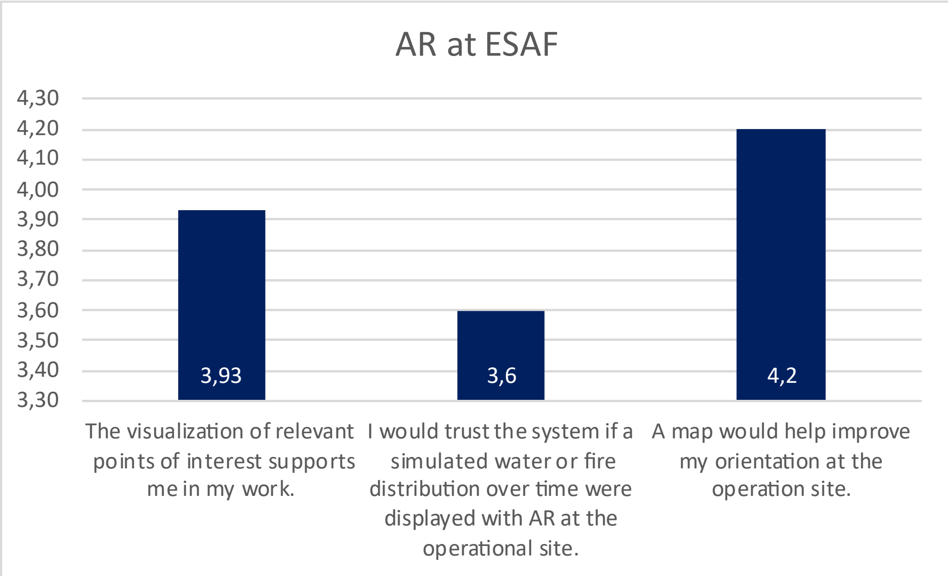
expert #											**main**
The fatigue assessment is helpful for action planning in an operation.	5	5	5	5	4	4	4	4	5	5	4,55
I would use data to protect emergency personnel from overexertion in real-life operations.	5	5	5	4	3	5	4	4	5	5	4,35
I trust a fatigue assessment based on biometric data.	4	5	5	4	4	4	4	4	3	5	4,15
I trust a data-based fatigue assessment even if the firefighter feels no fatigue.	4	4	5	4	4	5	4	4	3	4	4,1



Augmented Reality applications (scale from 1 – strongly disagree to 5 – strongly agree)

expert #																**main**
The visualization of relevant points of interest supports me in my work.	3	4	3	3	5	4	4	5	1	5	5	5	3	5	4	3,93
I would trust the system if a simulated water or fire distribution over time	2	4	5	4	5	4	4	4	2	3	3	4	2	3	5	3,6

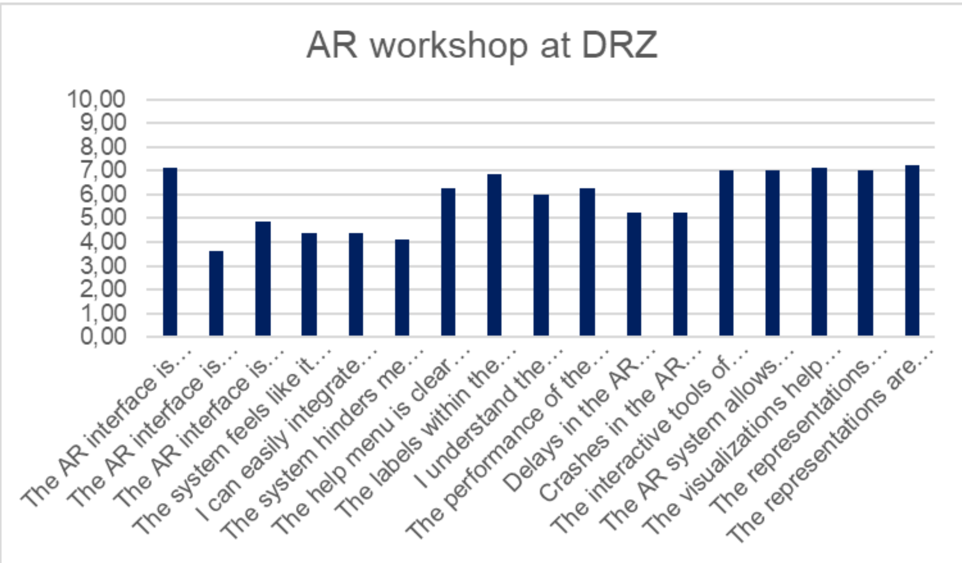
were displayed with AR at the operational site.																
A map would help improve my orientation at the operation site.	4	4	5	5	5	2	4	4	5	4	4	5	4	4	4	4,2



9.6 Results of AR Workshop at DRZ

expert #									**main**
The AR interface is intuitive.	5	8	7	5	7	7	8	10	7,13
The AR interface is confusing.	6	4	3	5	3	3	3	2	3,63
The AR interface is difficult to use.	6	6	3	5	7	5	4	3	4,88
The system feels like it is part of my usual equipment.	7	3	2	4	4	5	3	7	4,38
I can easily integrate the system into my usual workflows.	8	5	2	4	3	7	3	3	4,38
The system hinders me in typical tasks such as situation assessment or navigation.	4	6	2	4	5	2	5	5	4,13
The help menu is clear and helpful.	4	8	3	6	7	5	8	9	6,25

The labels within the system are clear and helpful.	5	9	7	5	8	6	6	9	6,88
I understand the purpose of each function without needing additional instructions.	4	5	10	5	5	4	5	10	6,00
The performance of the AR system is responsive and stable.	4	6	5	4	7	6	9	9	6,25
Delays in the AR system prevent me from completing tasks effectively.	8	3	7	3	5	6	2	8	5,25
Crashes in the AR system prevent me from completing tasks effectively.	8	5	4	5	4	6	5	5	5,25
The interactive tools of the AR system help me to better assess the situation.	7	8	7	6	6	5	9	8	7,00
The AR system allows me to access relevant information efficiently (quickly and easily).	6	9	7	4	5	7	9	9	7,00
The visualizations help me to assess risk and uncertainty.	8	8	7	5	6	7	9	7	7,13
The representations increase my confidence in my decisions.	6	8	8	5	6	7	9	7	7,00
The representations are realistic and helpful.	5	7	9	7	7	7	7	9	7,25

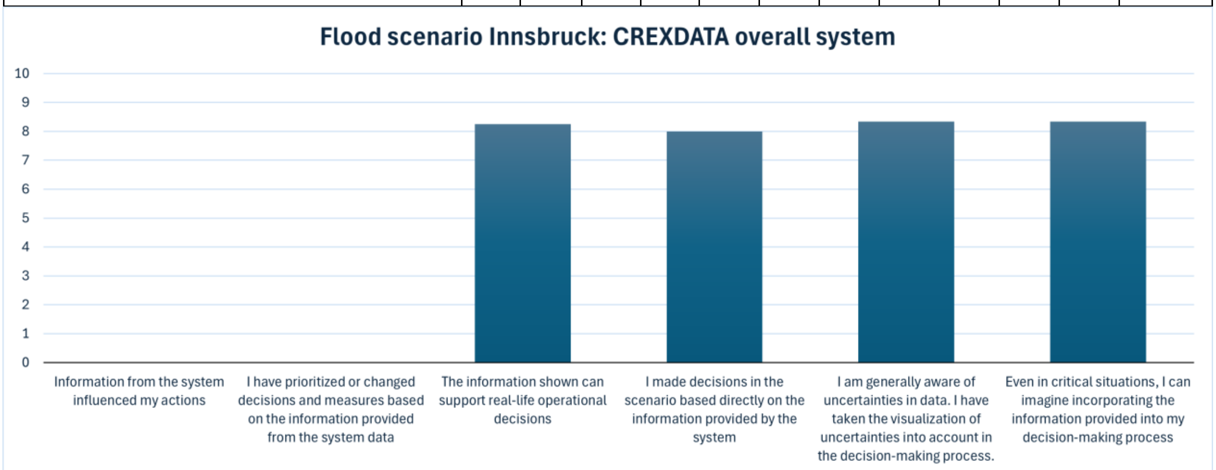


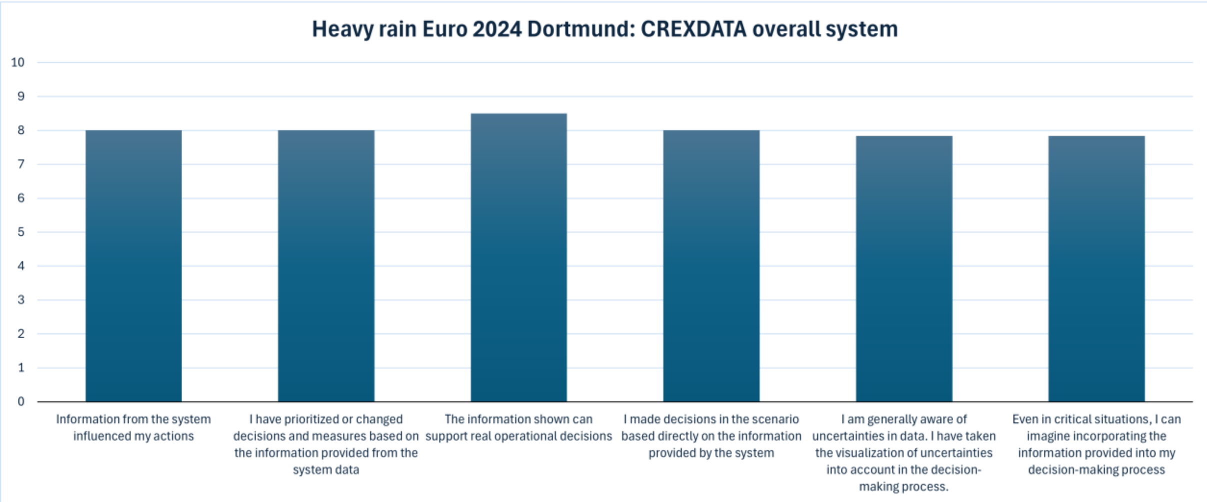
9.7 Questions and results per technology in the EmCase Demonstrator System in the Field trials in Innsbruck and Dortmund

9.7.1 EmCase Demonstrator System

Results CREXDATA overall system (Innsbruck & Dortmund Flood Scenario)

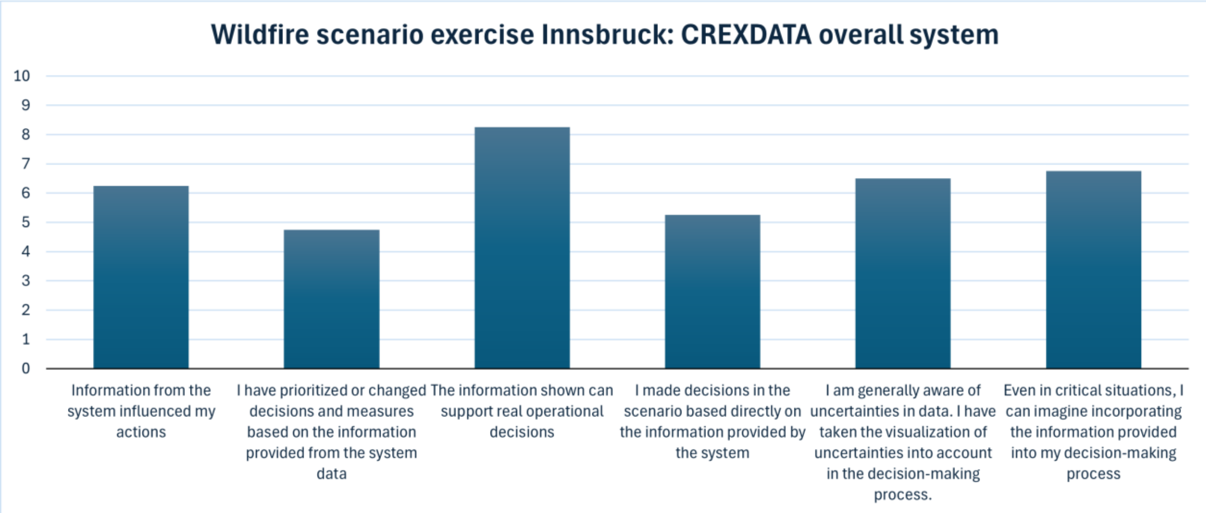
expert #	11	12	13	14	70	71	72	73	74	75	77	main
Information from the system influenced my actions	-	-	-	-	10	8	8	7	-	6	9	**8**
I have prioritized or changed decisions and measures based on the information provided from the system data	-	-	-	-	10	8	8	8	-	6	8	**8**
The information shown can support real-life operational decisions	8	9	9	7	10	9	8	8	-	7	9	**8,38**
I made decisions in the scenario based directly on the information provided by the system	-	8	-	-	9	9	7	8	-	7	8	**8**
I am generally aware of uncertainties in data. I have taken the visualization of uncertainties into account in the decision-making process.	8	8	9	-	9	9	7	8	-	7	7	**8,08**
Even in critical situations, I can imagine incorporating the information provided into my decision-making process	-	8	9	8	8	9	7	7	-	7	9	**8,08**

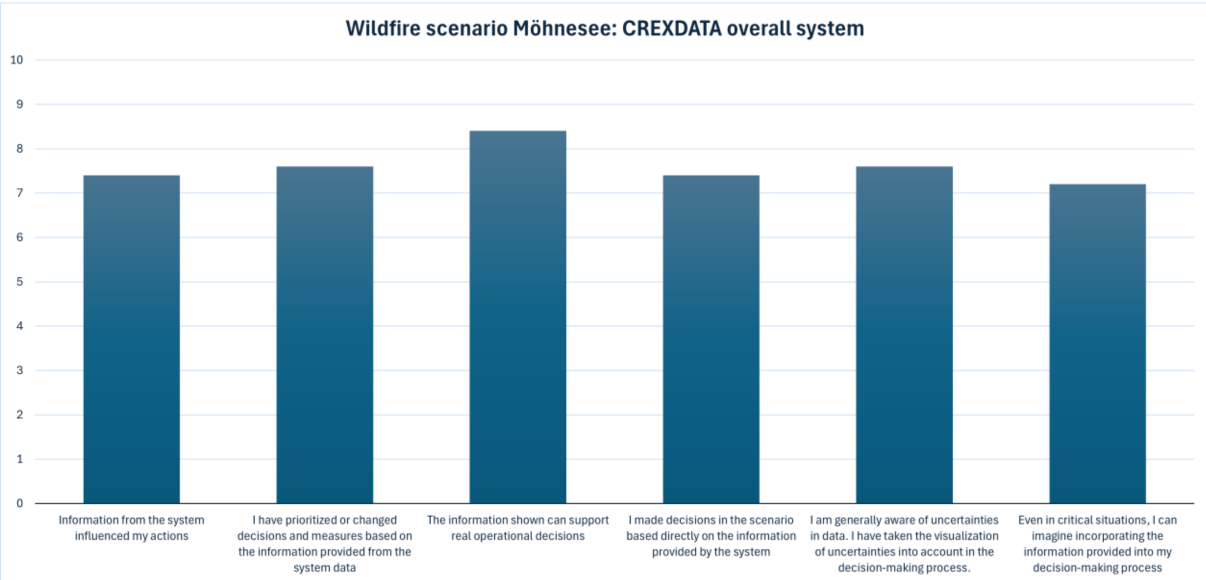




Results CREXDATA overall system (Innsbruck & Dortmund Wildfire Scenario)

expert #	**21**	**23**	**24**	**25**	**26**	**60**	**61**	**62**	**63**	**64**	**65**	**main**
Information from the system influenced my actions	6	5	7	7	6	8	7	9	-	8	6	**6,825**
I have prioritized or changed decisions and measures based on the information provided from the system data	5	5	3	6	5	8	8	9	-	8	5	**6,175**
The information shown can support real operational decisions	8	10	8	7	8	9	9	10	-	8	8	**8,325**
I made decisions in the scenario based directly on the information provided by the system	7	5	3	6	7	9	7	10	-	8	7	**6,325**
I am generally aware of uncertainties in data. I have taken the visualization of uncertainties into account in the decision-making process.	8	8	5	5	8	8	6	10		8	8	**7,05**
Even in critical situations, I can imagine incorporating the information provided into my decision-making process	8	8	5	6	8	8	8	10	-	8	8	**6,975**



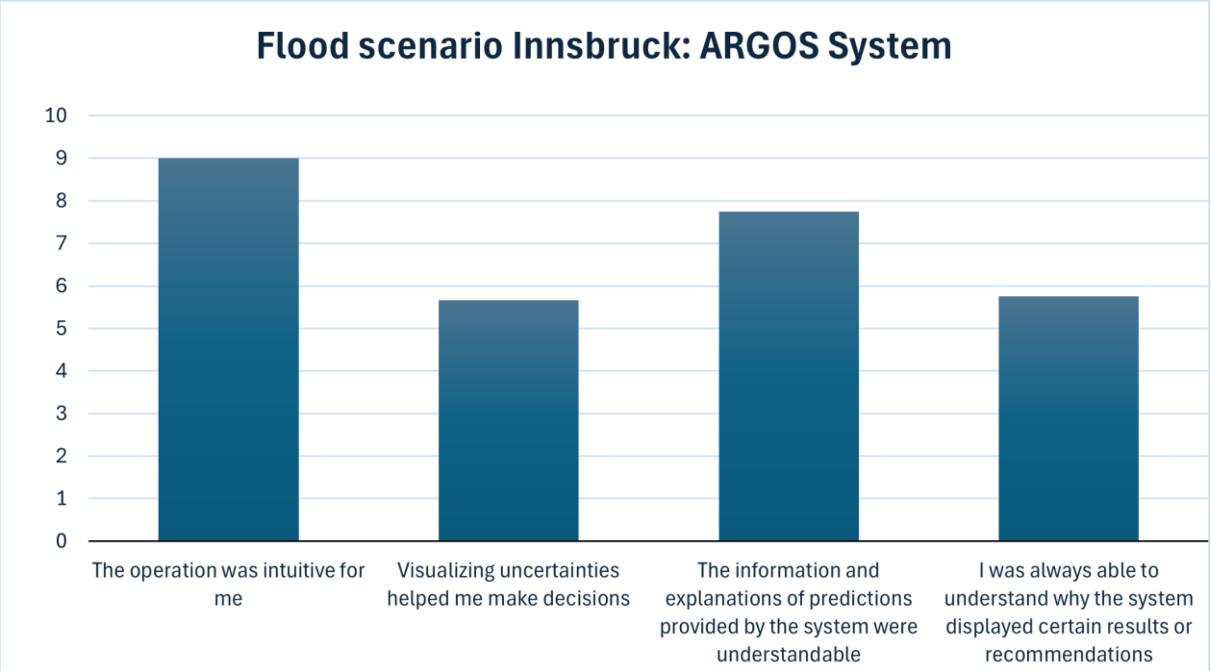


9.7.2 ARGOS

Results ARGOS System (Innsbruck)

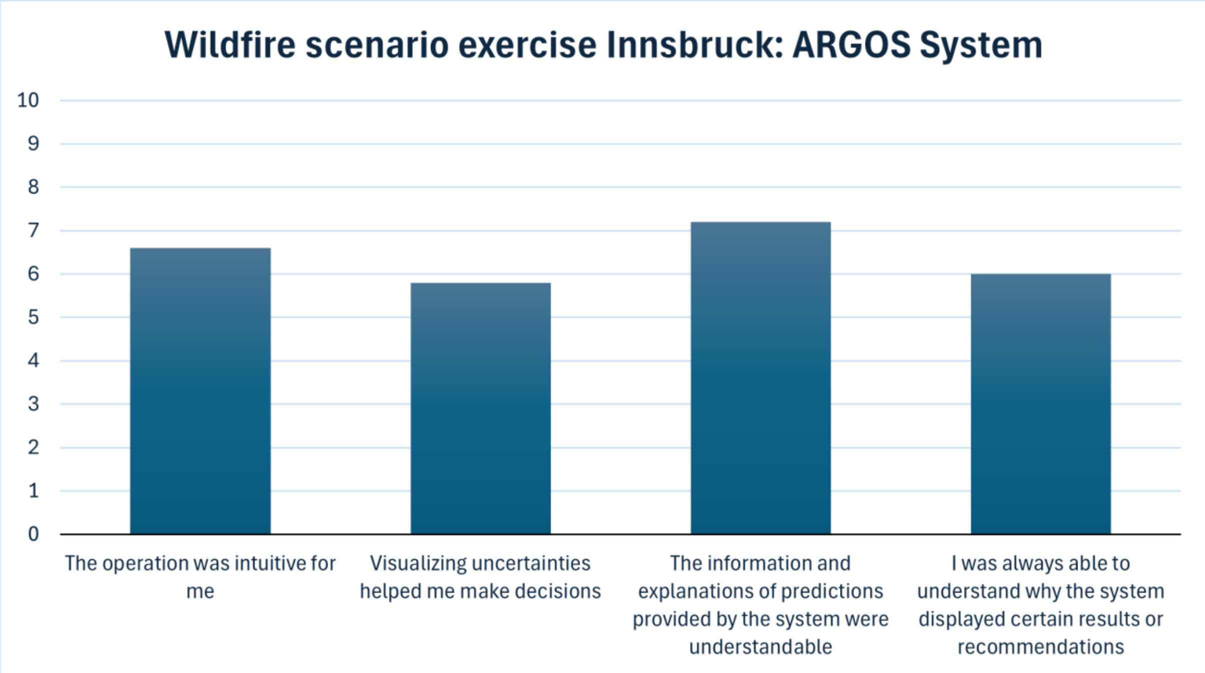
expert #	11	12	13	14	main
The operation was intuitive for me	10	9	9	8	**9**
Visualizing uncertainties helped me make decisions	2	-	8	7	**5,67**
The information and explanations of predictions provided by the system were understandable	6	9	9	7	**7,75**

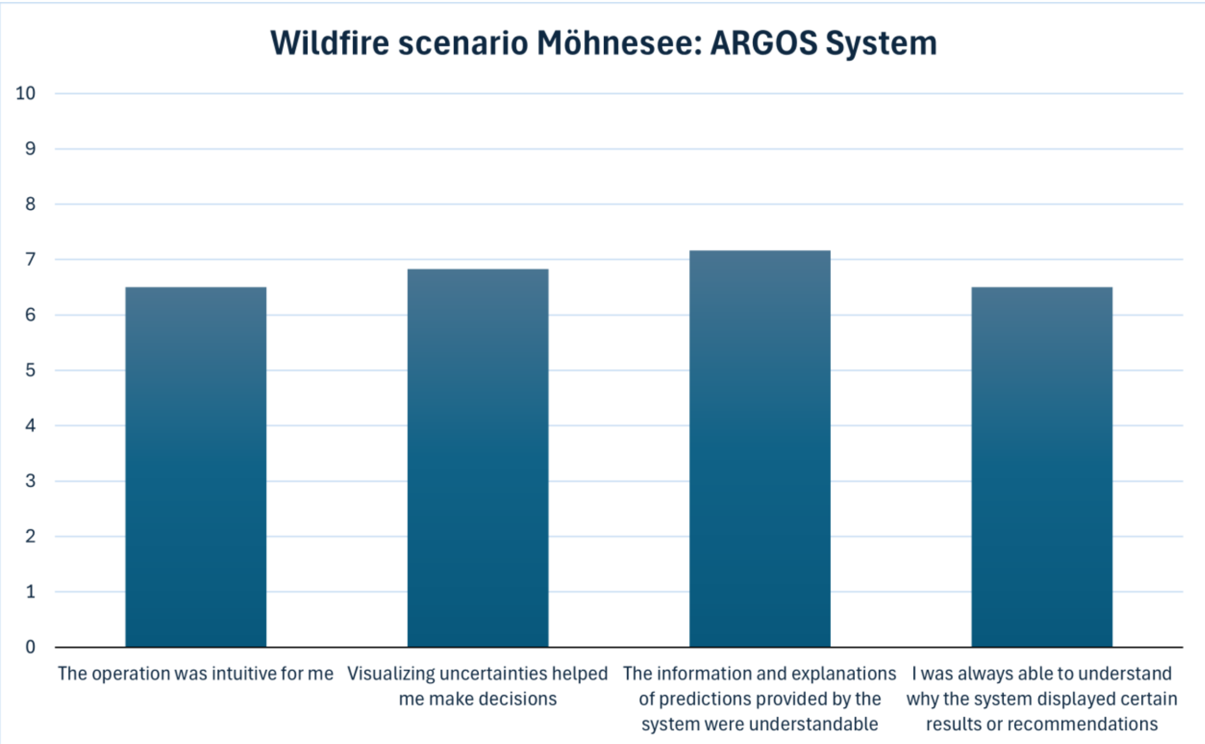
I was always able to understand why the system displayed certain results or recommendations	2	7	8	6	**5,75**



Results ARGOS System (Innsbruck & Dortmund Wildfire Scenario)

expert #	**21**	**23**	**24**	**25**	**26**	**60**	**61**	**62**	**63**	**64**	**65**	**main**
The operation was intuitive for me	5	5	8	8	7	8	5	4	8	8	6	**6,55**
Visualizing uncertainties helped me make decisions	4	5	8	5	7	8	6	6	9	8	4	**6,32**
The information and explanations of predictions provided by the system were understandable	6	6	8	7	9	7	6	8	10	8	4	**7,18**
I was always able to understand why the system displayed certain results or recommendations	4	4	7	7	8	8	5	8	9	6	3	**6,25**

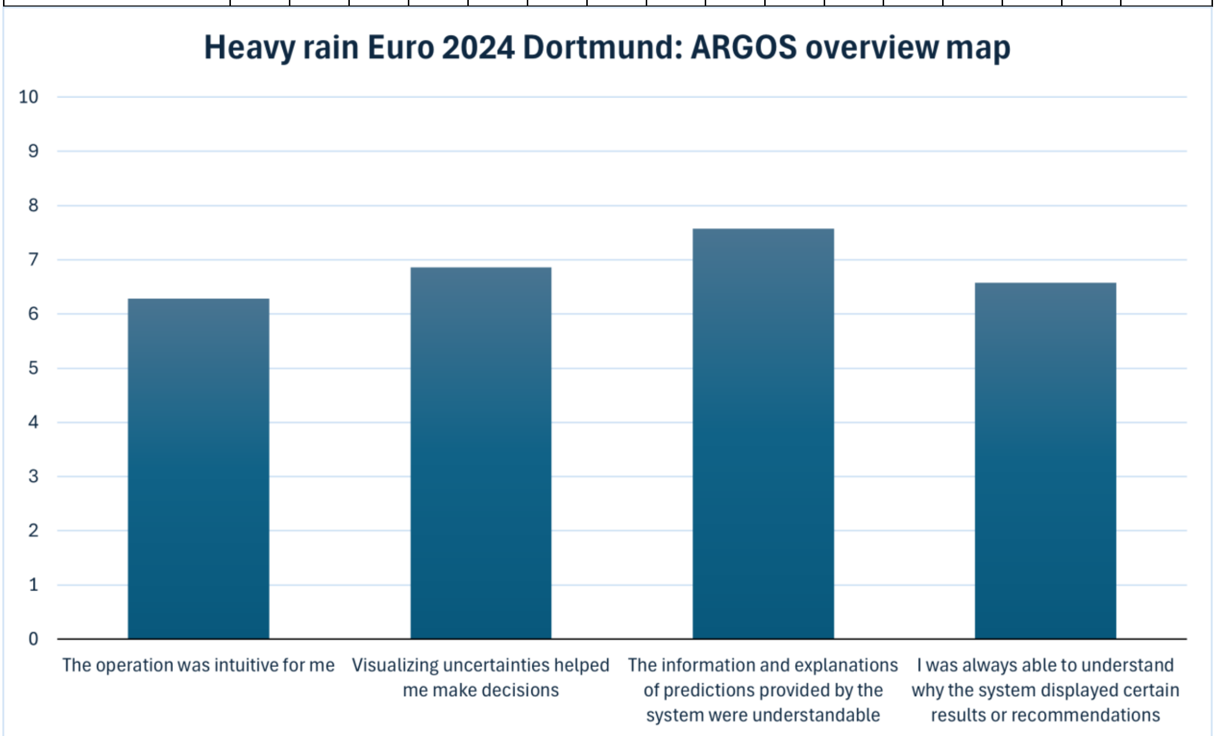


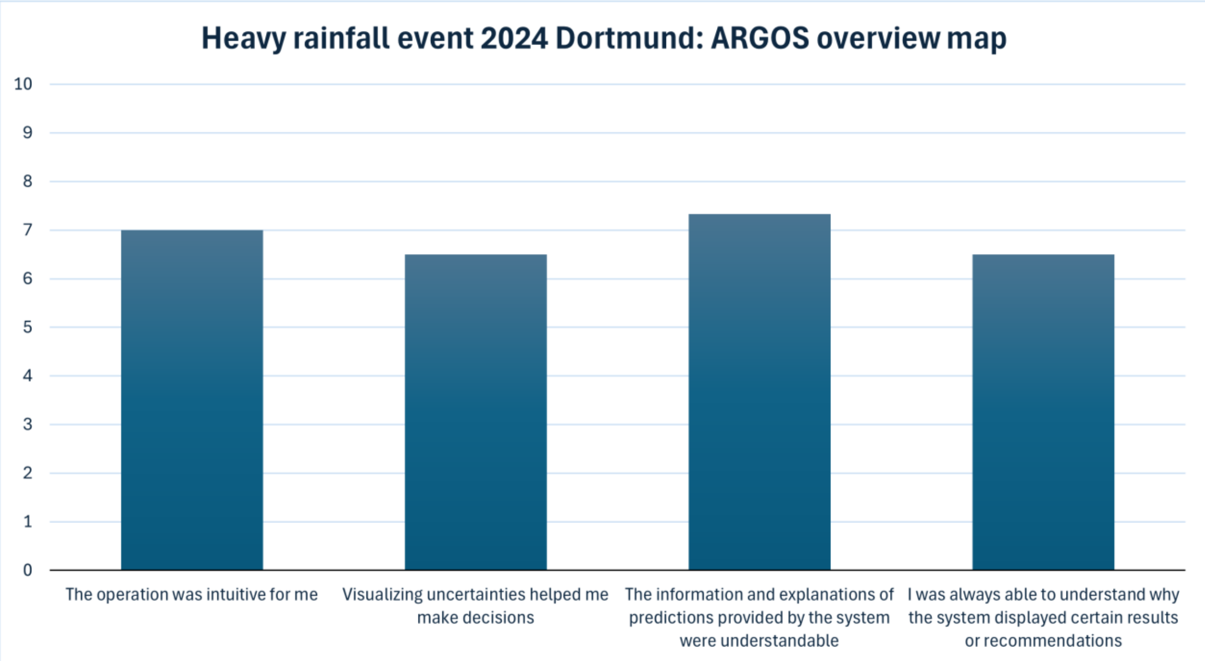


Results ARGOS overview map (Dortmund)

expert #	70	71	72	73	74	75	77	90	91	92	93	94	95	96	97	main

The operation was intuitive for me	4	7	9	8	4	7	5	8	-	8	9	5	4	-	8	**6,64**
Visualizing uncertainties helped me make decisions	4	6	7	9	7	6	9	7	-	7	9	6	4	-	6	**6,68**
The information and explanations of predictions provided by the system were understandable	5	8	7	8	10	6	9	9	-	10	8	5	5	-	7	**7,45**
I was always able to understand why the system displayed certain results or recommendations	5	7	7	7	10	4	6	4	-	7	9	7	5	-	7	**6,54**

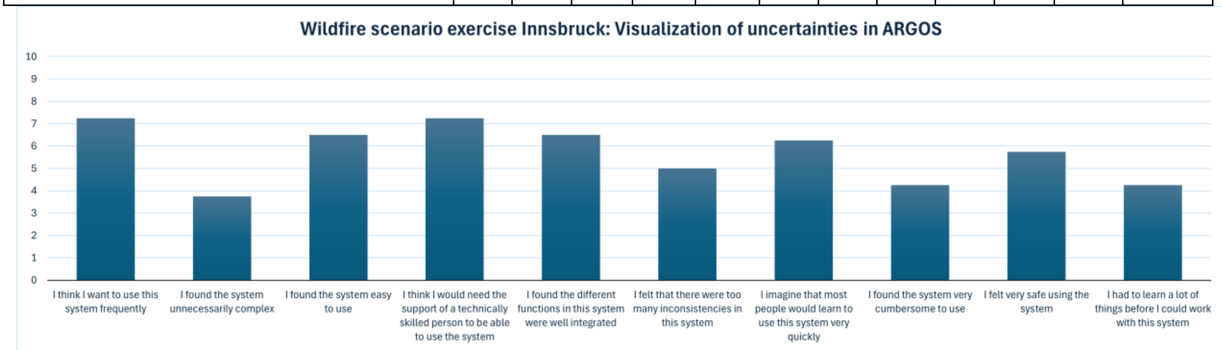


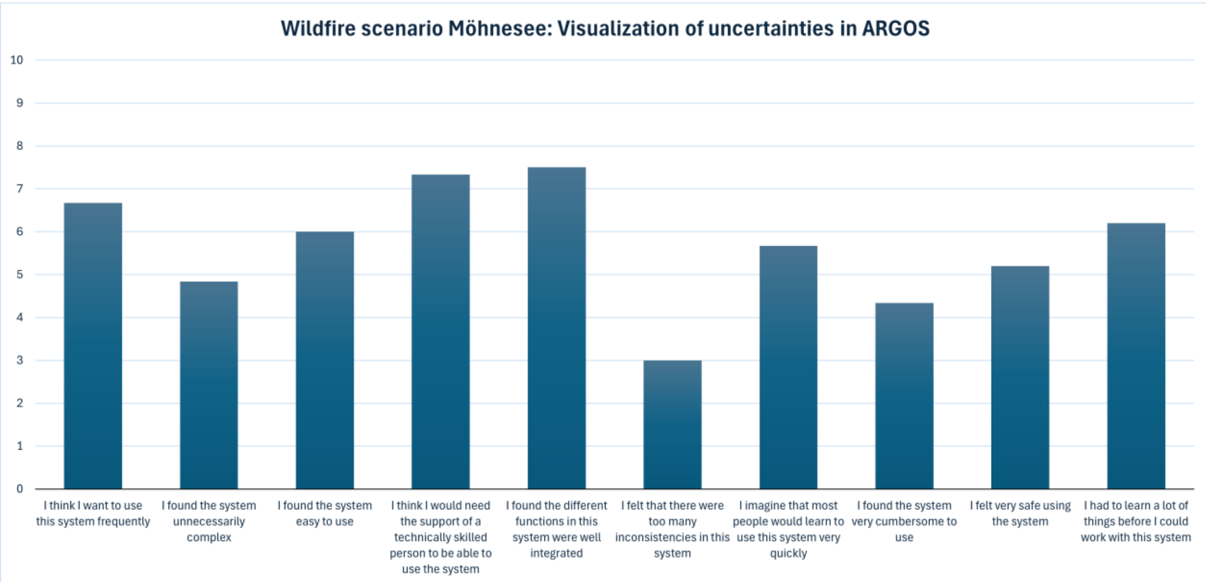


Results Visualization of uncertainties in ARGOS (Innsbruck & Dortmund Wildfire Scenario)

expert #	**21**	**23**	**24**	**25**	**26**	**60**	**61**	**62**	**63**	**64**	**65**	**main**
I think I want to use this system frequently	8	8	8	5	8	6	6	10	8	7	3	**6,96**
I found the system unnecessarily complex	3	1	4	7	3	4	5	4	4	4	8	**4,29**
I found the system easy to use	5	8	8	5	5	6	4	7	7	7	5	**6,25**
I think I would need the support of a technically skilled person to be able to use the system	5	10	7	7	5	2	9	10	8	6	9	**7,29**
I found the different functions in this system were well integrated	5	10	7	4	5	8	7	8	8	8	6	**7**
I felt that there were too many inconsistencies in this system	6	2	5	7	6	3	4	2	2	2	5	**4**
I imagine that most people would learn to use this system very quickly	7	5	8	5	7	8	5	7	6	7	1	**5,96**
I found the system very cumbersome to use	5	2	3	7	5	2	6	6	3	2	7	**4,29**
I felt very safe using the system	5	5	7	6	5	8	4	3	-	7	4	**5,48**

I had to learn a lot of things before I could work with this system	7	1	1	8	7	3	7	9	-	5	7	**5,23**



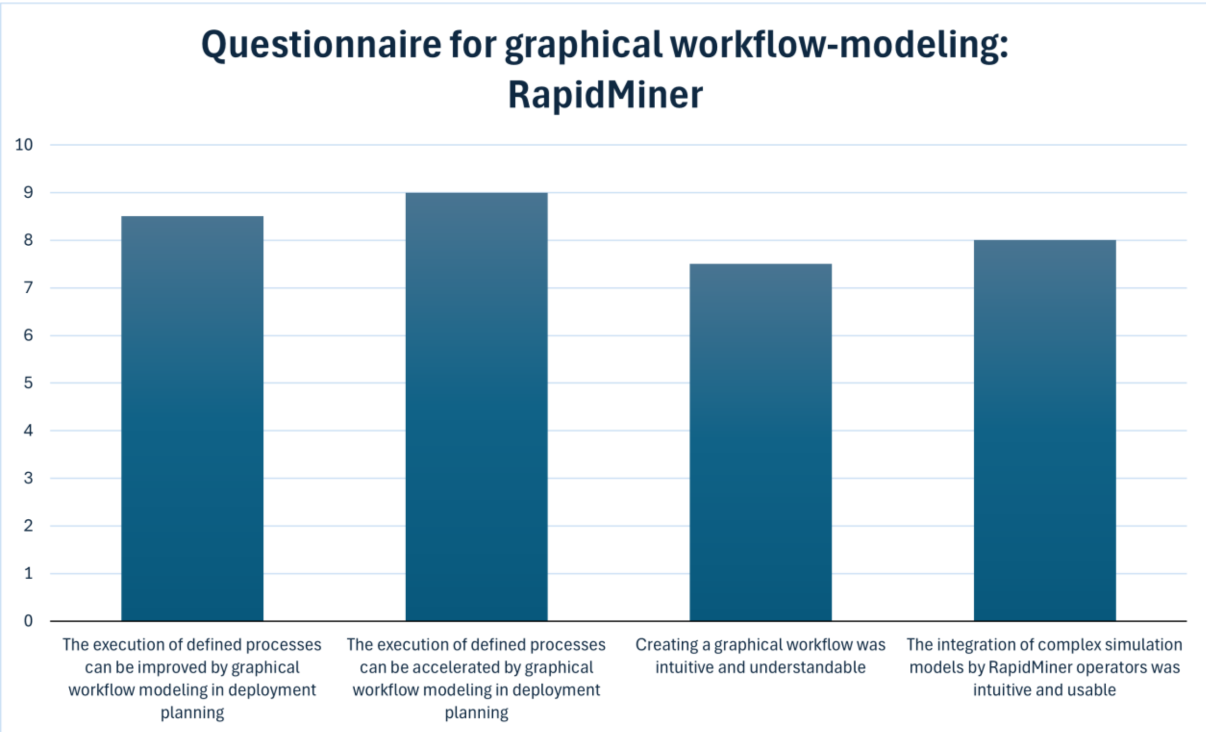


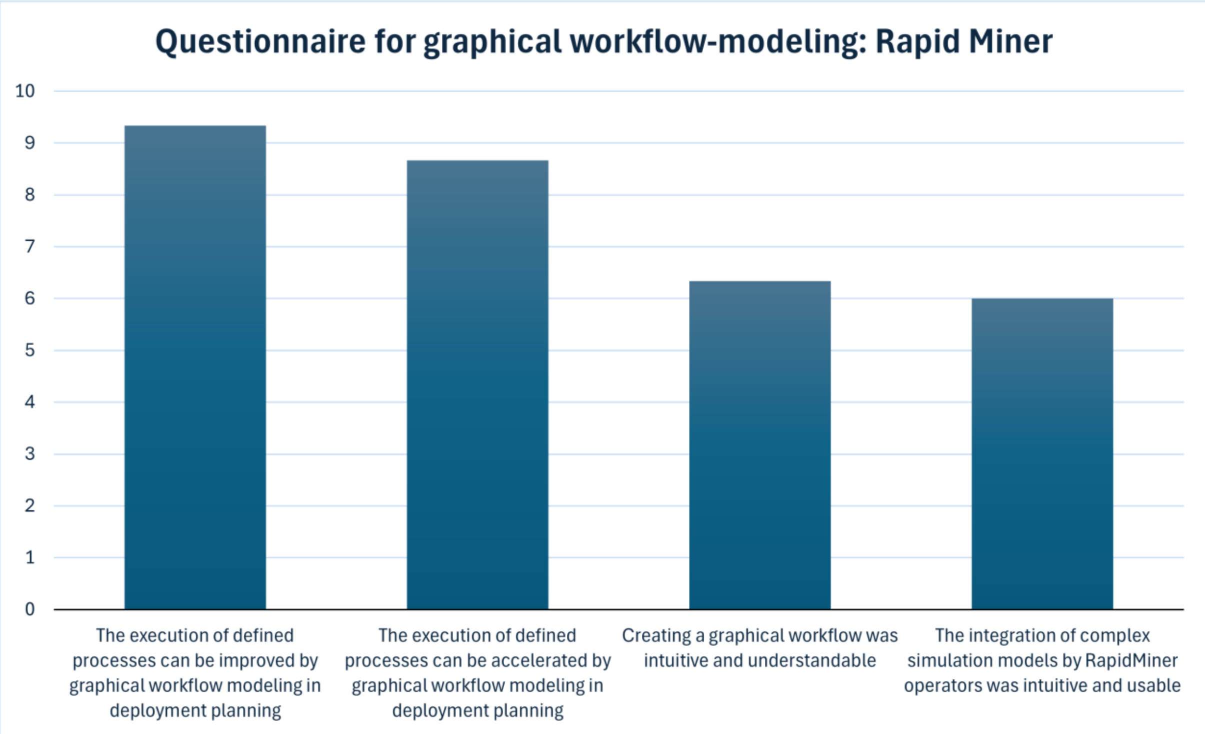
9.7.3 RapidMiner Graphical Workflow Modelling

Results Rapid Miner

expert #	**11**	**12**	**13**	**14**	**60**	**61**	**62**	**63**	**64**	**65**	**main**
The execution of defined processes can be accelerated by graphical workflow modeling in deployment planning	9	8	8	9	10	8	10	10	10	8	**8,92**
Creating a graphical workflow was intuitive and understandable	10	10	7	9	9	9	10	7	10	7	**8,83**

The integration of complex simulation models by RapidMiner operators was intuitive and usable	7	7	9	7	7	6	4	7	5	9	**6,92**
The execution of defined processes can be improved by graphical workflow modeling in deployment planning	7	8	9	8	7	7	6	7	4	5	**7**



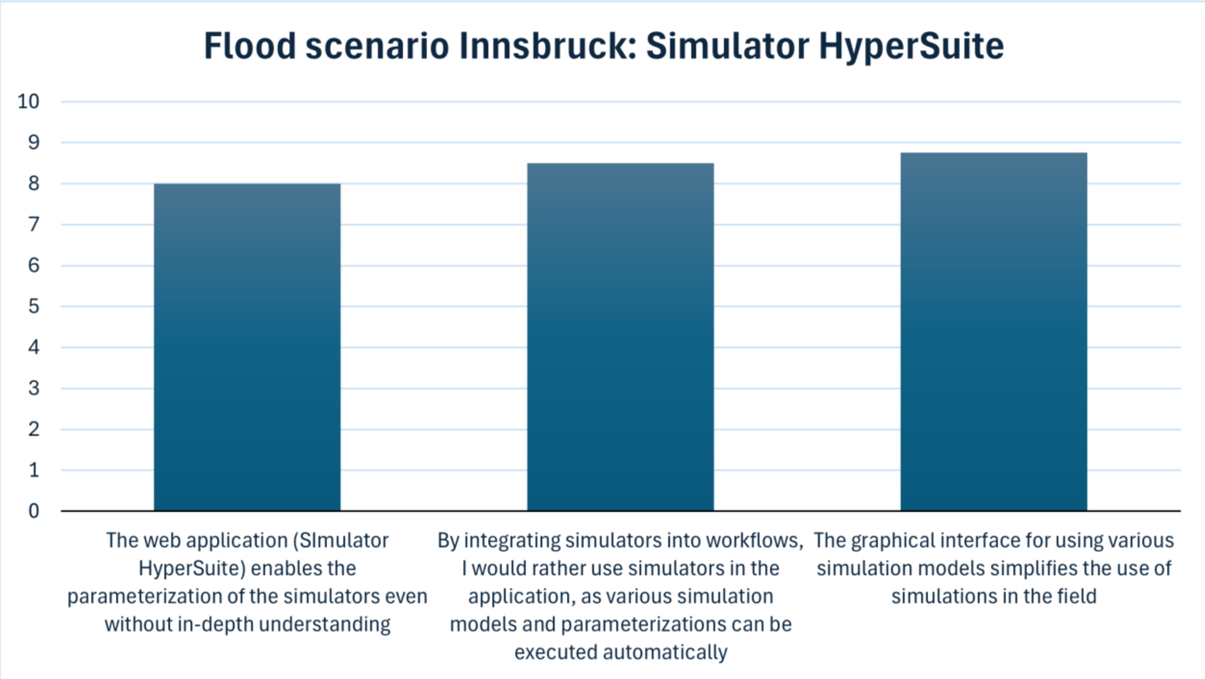


9.7.4 Simulators

9.7.4.1 EmCase Simulator HyperSuite

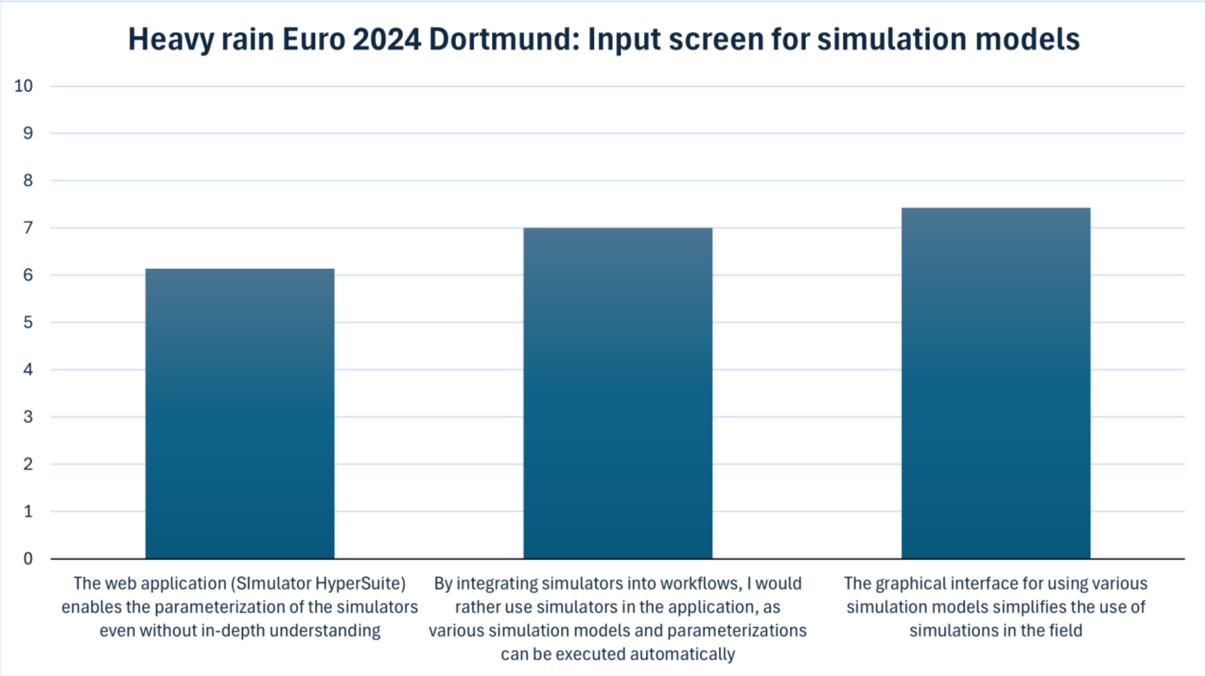
Results Simulator HyperSuite (Innsbruck flooding scenario)

expert #	**11**	**12**	**13**	**14**	**main**
The web application (SImulator HyperSuite) enables the parameterization of the simulators even without in-depth understanding	8	8	9	7	**8**
By integrating simulators into workflows, I would rather use simulators in the application, as various simulation models and parameterizations can be executed automatically	7	9	10	8	**8,5**
The graphical interface for using various simulation models simplifies the use of simulations in the field	8	9	10	8	**8,75**



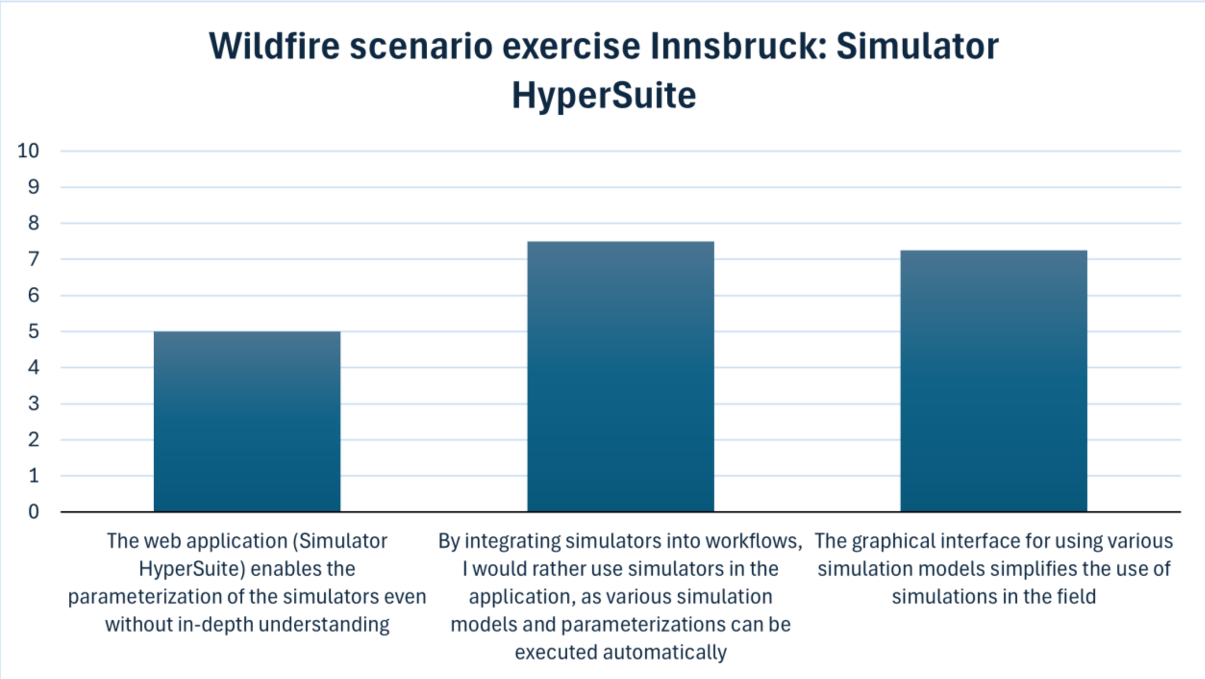
Results Input screen for simulation models (Dortmund)

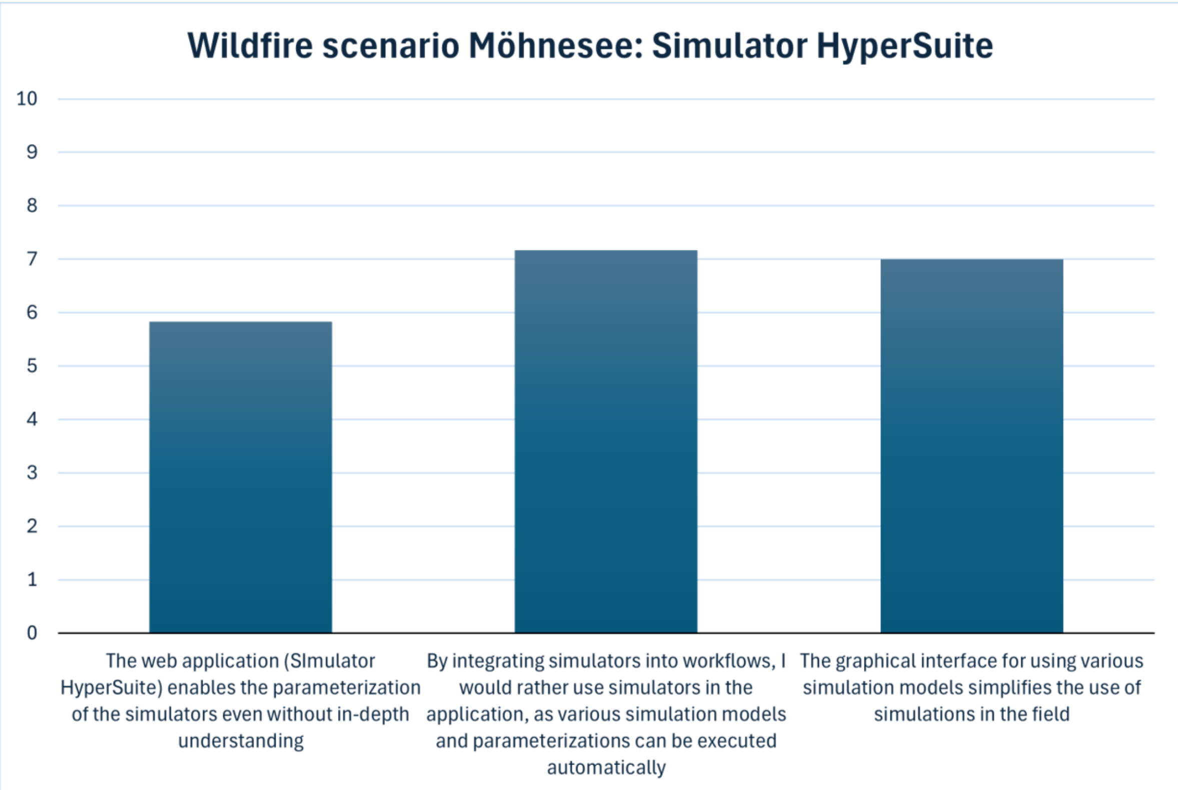
expert #	**70**	**71**	**72**	**73**	**74**	**75**	**77**	**main**
The web application (SImulator HyperSuite) enables the parameterization of the simulators even without in-depth understanding	6	8	5	8	9	3	4	**6,14**
By integrating simulators into workflows, I would rather use simulators in the application, as various simulation models and parameterizations can be executed automatically	9	8	6	7	8	2	9	**7**
The graphical interface for using various simulation models simplifies the use of simulations in the field	6	8	8	8	10	4	8	**7,43**



Results Simulator HyperSuite (Innsbruck & Dortmund Wildfire scenario)

expert #	**21**	**23**	**24**	**25**	**26**	**60**	**61**	**62**	**63**	**64**	**65**	**main**
The web application (Simulator HyperSuite) enables the parameterization of the simulators even without in-depth understanding	3	5	-	-	7	5	5	5	8	7	5	**5,42**
By integrating simulators into workflows, I would rather use simulators in the application, as various simulation models and parameterizations can be executed automatically	7	7	10	-	6	7	7	10	9	9	1	**7,33**
The graphical interface for using various simulation models simplifies the use of simulations in the field	5	7	9	-	8	8	7	5	8	10	4	**7,13**

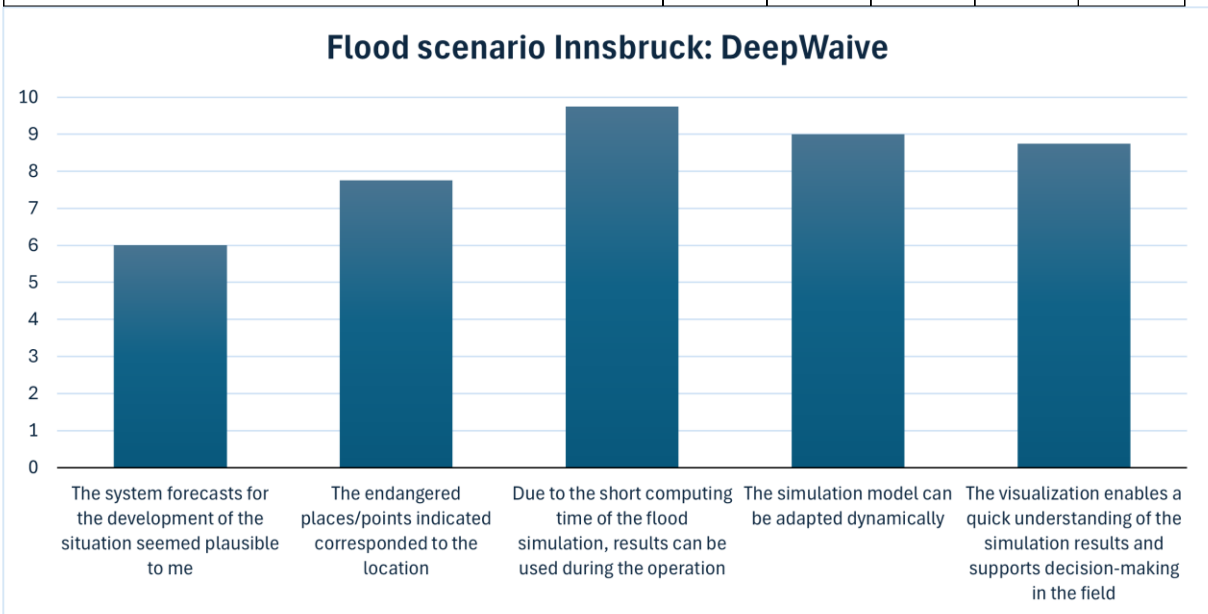

Wildfire scenario exercise Innsbruck: Simulator HyperSuite
10
9
8
7
6
5
4
3
2
1
0
The web application (Simulator HyperSuite) enables the parameterization of the simulators even without in-depth understanding
By integrating simulators into workflows, I would rather use simulators in the application, as various simulation models and parameterizations can be executed automatically
The graphical interface for using various simulation models simplifies the use of simulations in the field


Wildfire scenario Möhnesee: Simulator HyperSuite
10
9
8
7
6
5
4
3
2
1
0
The web application (SImulator HyperSuite) enables the parameterization of the simulators even without in-depth understanding
By integrating simulators into workflows, I would rather use simulators in the application, as various simulation models and parameterizations can be executed automatically
The graphical interface for using various simulation models simplifies the use of simulations in the field

9.7.4.2 FloodWaive

Results FloodWaive (Innsbruck)

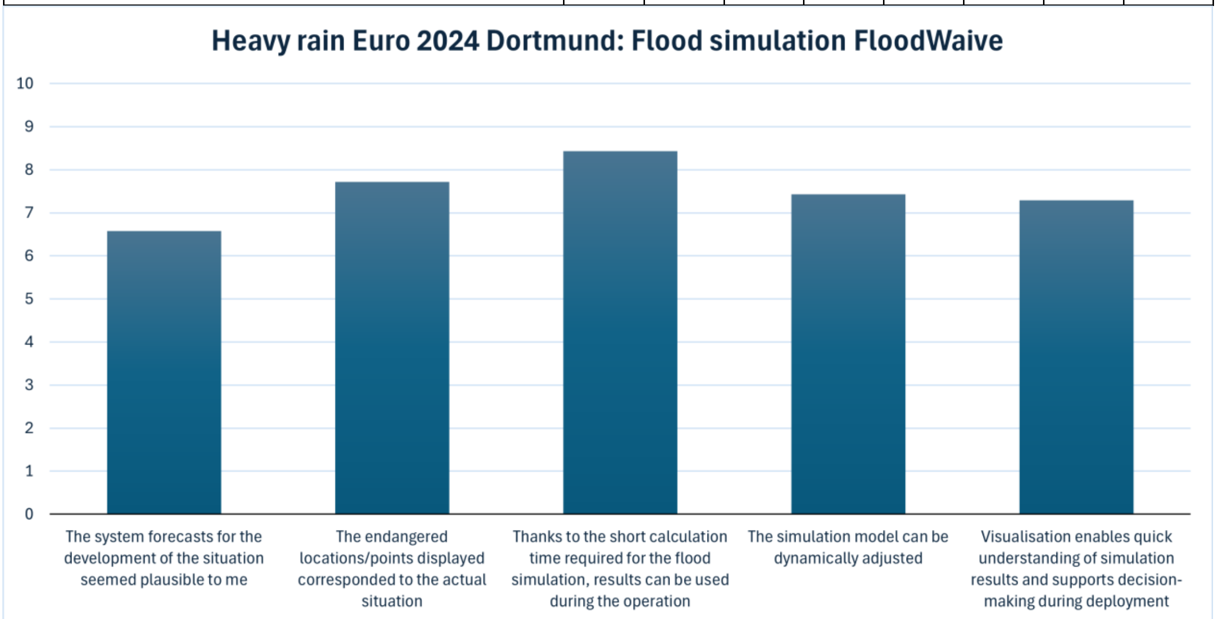
expert #	11	12	13	14	main
The system forecasts for the development of the situation seemed plausible to me	2	8	8	6	**6**
The endangered places/points indicated corresponded to the location	5	10	9	7	**7,75**
Due to the short computing time of the flood simulation, results can be used during the operation	10	10	10	9	**9,75**
The simulation model can be adapted dynamically	10	8	9	9	**9**
The visualization enables a quick understanding of the simulation results and supports decision-making in the field	10	9	9	7	**8,75**



Results Flood simulation FloodWaive (Dortmund)

expert #	70	71	72	73	74	75	77	main
The system forecasts for the development of the situation seemed plausible to me	9	6	8	5	5	4	9	**6,57**
The endangered locations/points displayed corresponded to the actual situation	8	9	9	5	8	6	9	**7,71**

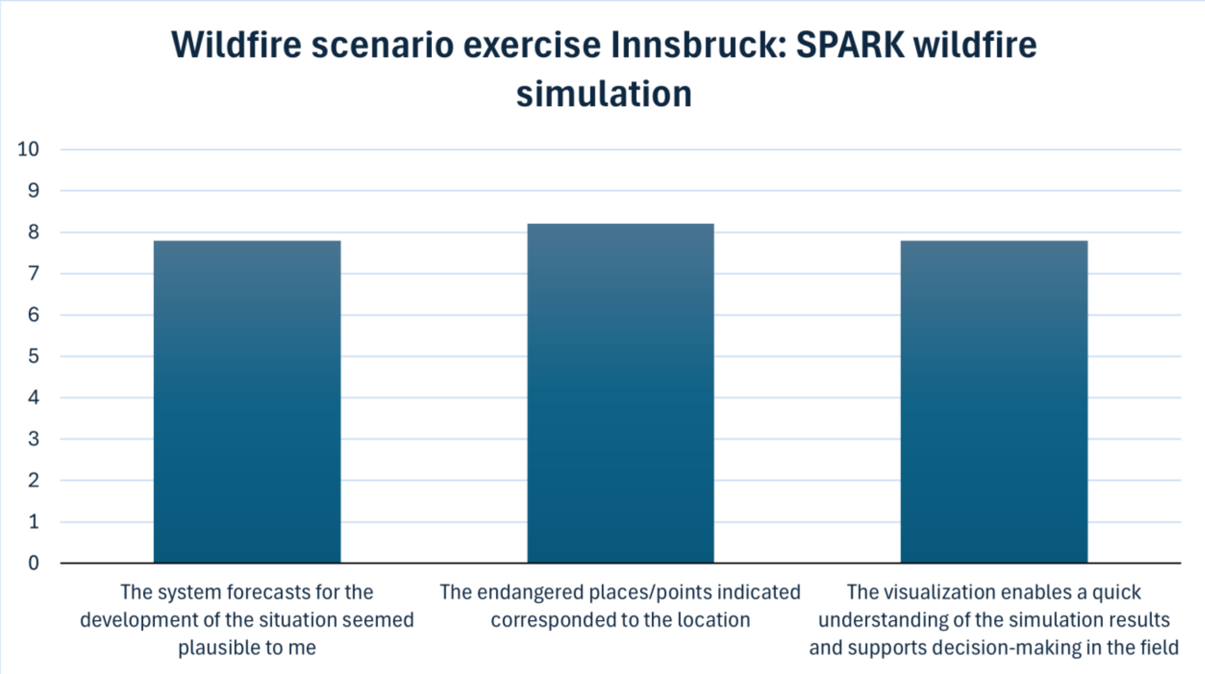
Thanks to the short calculation time required for the flood simulation, results can be used during the operation	10	10	9	7	10	4	9	**8,43**
The simulation model can be dynamically adjusted	7	8	8	7	10	4	8	**7,43**
Visualisation enables quick understanding of simulation results and supports decision-making during deployment	9	10	8	5	6	4	9	**7,29**

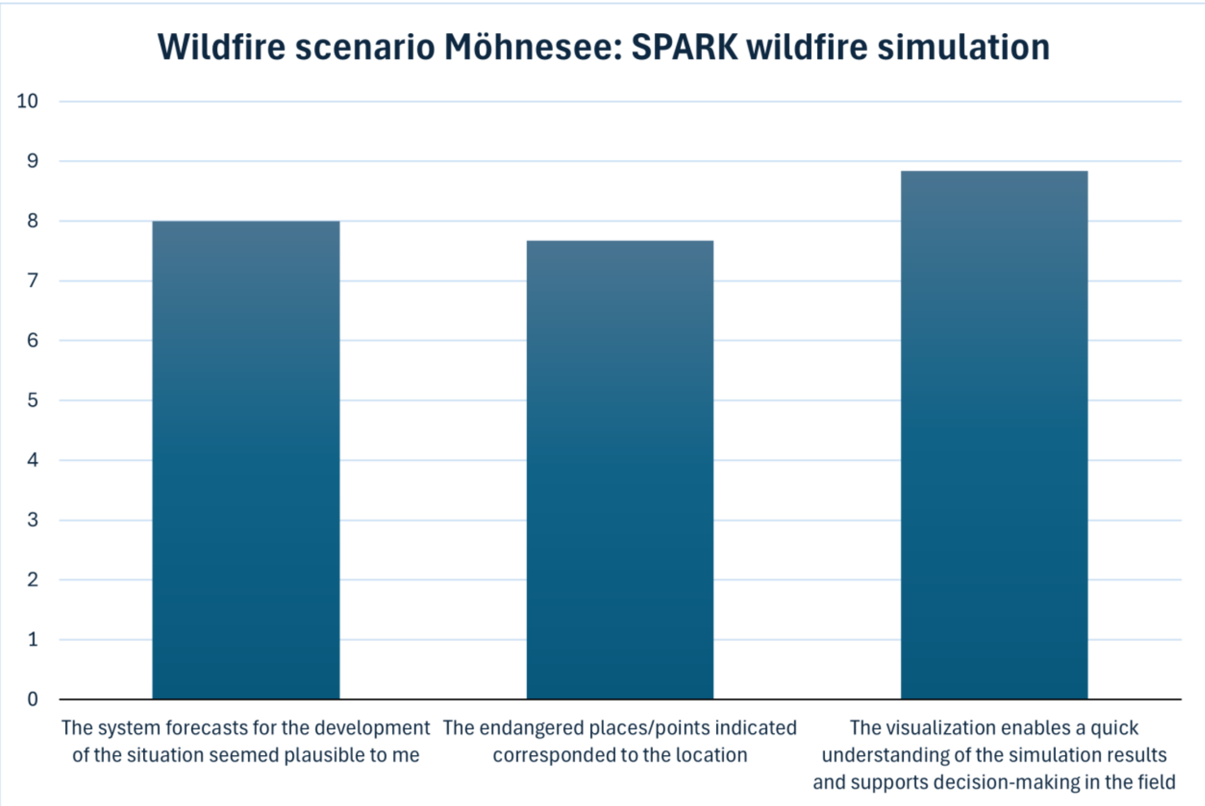


9.7.4.3 Spark

Results Spark Wildfire Simulation (Innsbruck & Dortmund)

expert #	**21**	**23**	**24**	**25**	**26**	**60**	**61**	**62**	**63**	**64**	**65**	**main**
The system forecasts for the development of the situation seemed plausible to me	6	8	10	8	7	9	9	8	9	8	5	**7,88**
The endangered places/points indicated corresponded to the location	9	8	10	8	6	8	9	8	9	8	4	**7,2**
The visualization enables a quick understanding of the simulation results and supports decision-making in the field	6	8	10	8	7	10	9	10	10	9	5	**8,45**

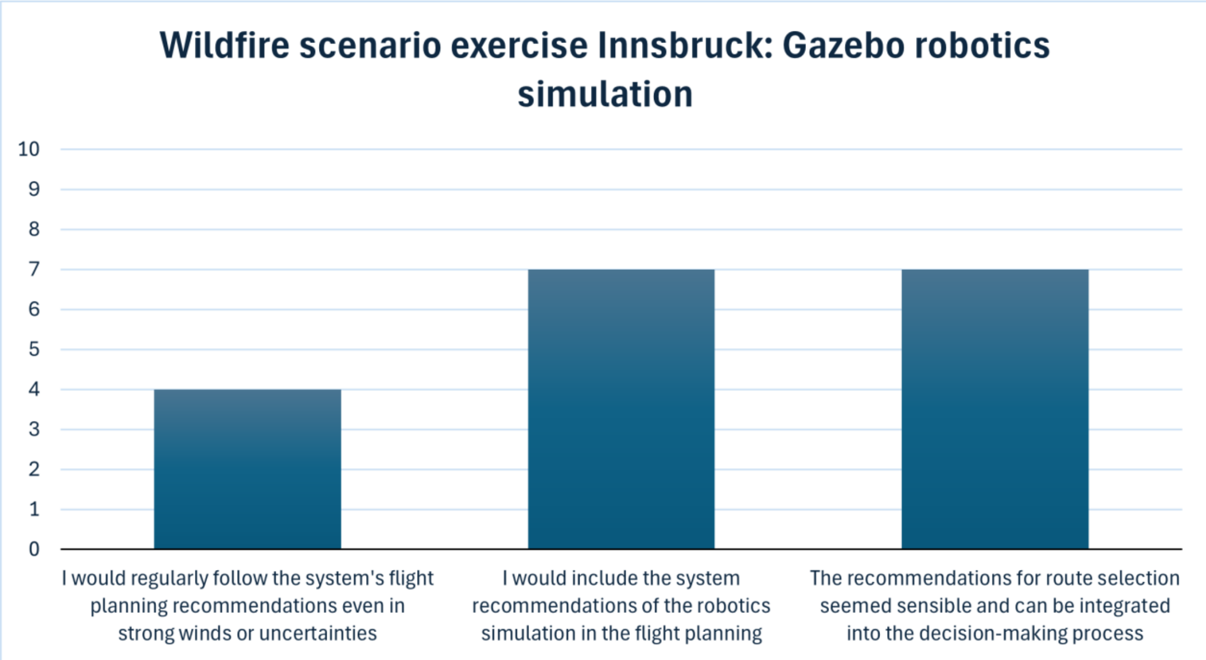

Wildfire scenario exercise Innsbruck: SPARK wildfire simulation
10
9
8
7
6
5
4
3
2
1
0
The system forecasts for the development of the situation seemed plausible to me
The endangered places/points indicated corresponded to the location
The visualization enables a quick understanding of the simulation results and supports decision-making in the field

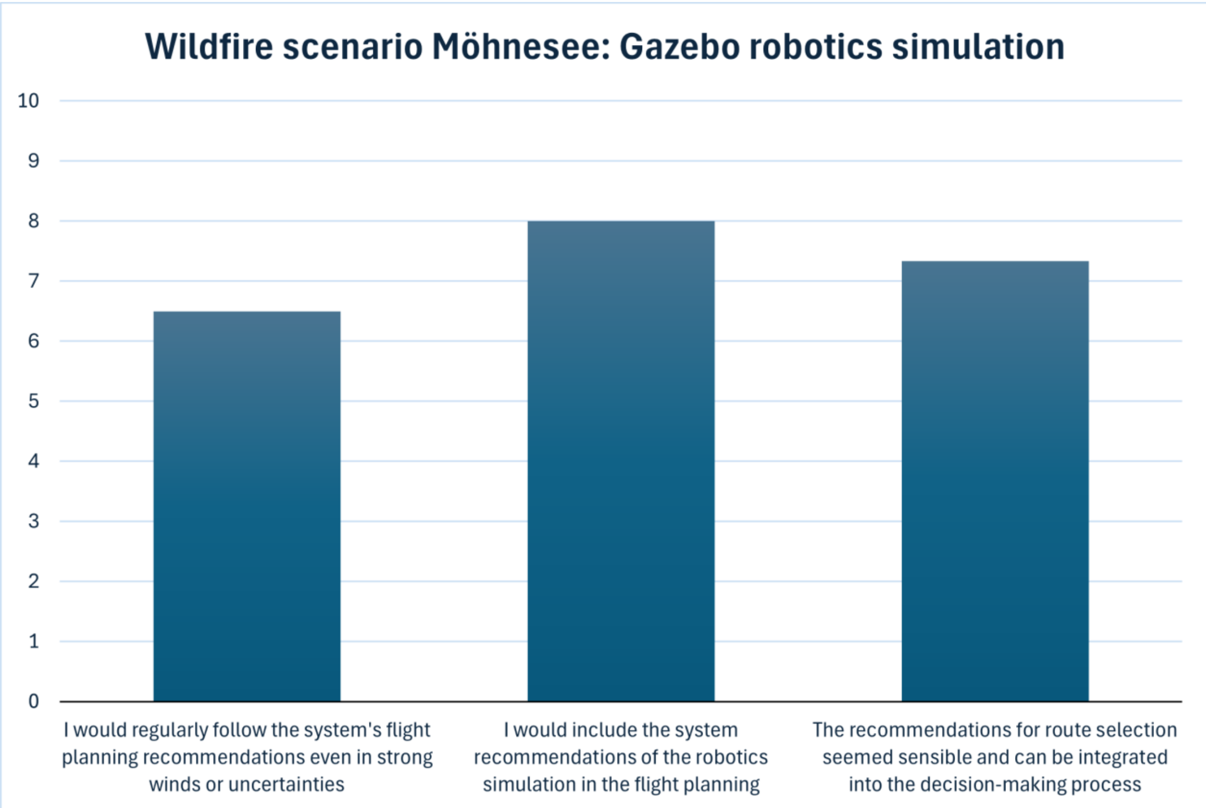

Wildfire scenario Möhnesee: SPARK wildfire simulation
10
9
8
7
6
5
4
3
2
1
0
The system forecasts for the development of the situation seemed plausible to me
The endangered places/points indicated corresponded to the location
The visualization enables a quick understanding of the simulation results and supports decision-making in the field

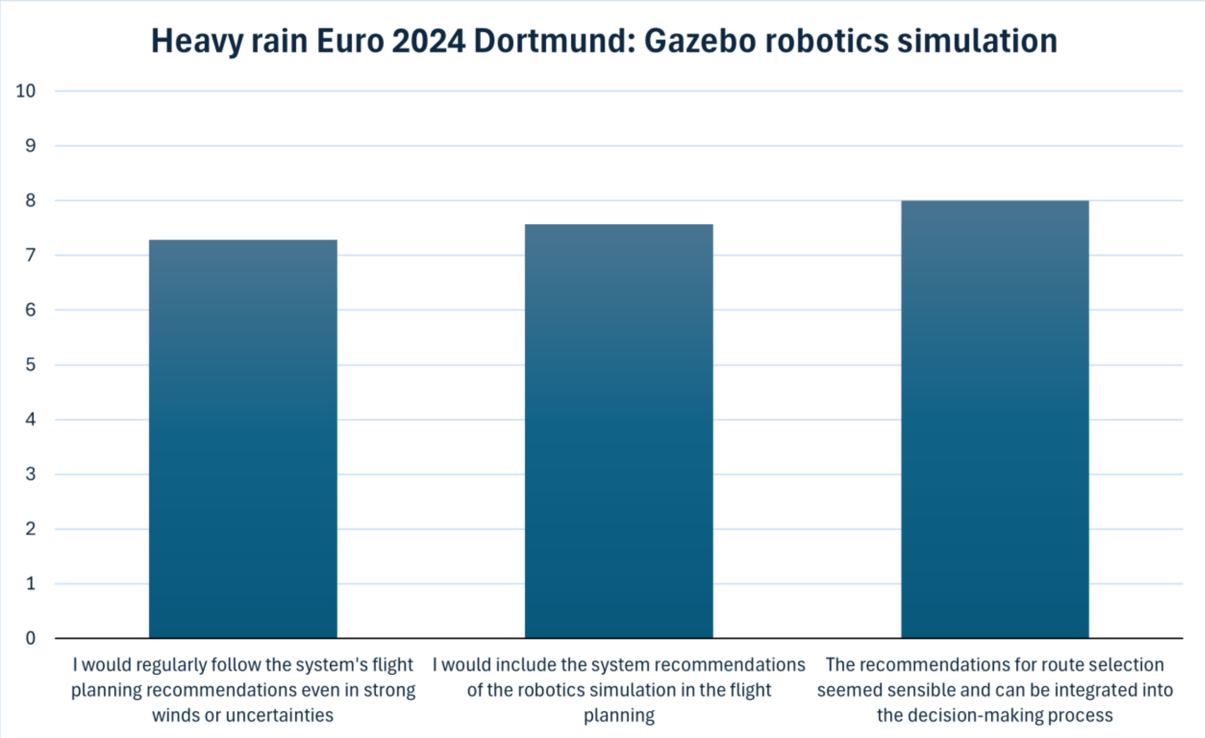
9.7.4.4 Gazebo

Results Gazebo robotics simulation (Innsbruck & Dortmund)

expert #	**21**	**23**	**24**	**25**	**26**	**60**	**61**	**62**	**63**	**64**	**65**	**70**	**71**	**72**	**73**	**74**	**75**	**77**	**main**
I would regularly follow the system's flight planning recommendations even in strong winds or uncertainties	-	4	2	3	7	9	8	10	7	4	1	9	6	8	7	7	5	9	**5,93**
I would include the system recommendations of the robotics simulation in the flight planning	-	8	10	5	5	9	8	10	10	10	1	8	7	7	7	10	4	10	**7,52**
The recommendations for route selection seemed sensible and can be integrated into the decision-making process	-	7	10	5	6	7	9	10	10	7	1	9	9	8	8	7	5	10	**7,44**


Wildfire scenario exercise Innsbruck: Gazebo robotics simulation
10
9
8
7
6
5
4
3
2
1
0
I would regularly follow the system's flight planning recommendations even in strong winds or uncertainties
I would include the system recommendations of the robotics simulation in the flight planning
The recommendations for route selection seemed sensible and can be integrated into the decision-making process


Wildfire scenario Möhnesee: Gazebo robotics simulation
10
9
8
7
6
5
4
3
2
1
0
I would regularly follow the system's flight planning recommendations even in strong winds or uncertainties
I would include the system recommendations of the robotics simulation in the flight planning
The recommendations for route selection seemed sensible and can be integrated into the decision-making process

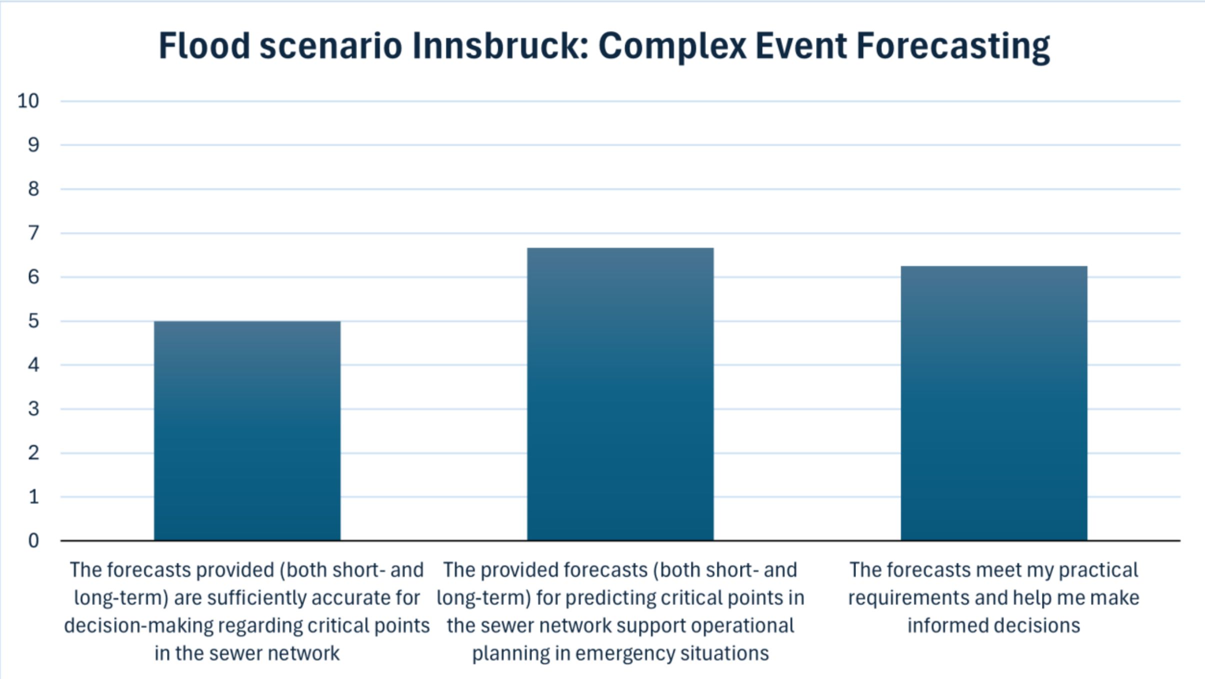


9.7.5 WP4 / WP 5 technologies

9.7.5.1 Complex Event Forecasting

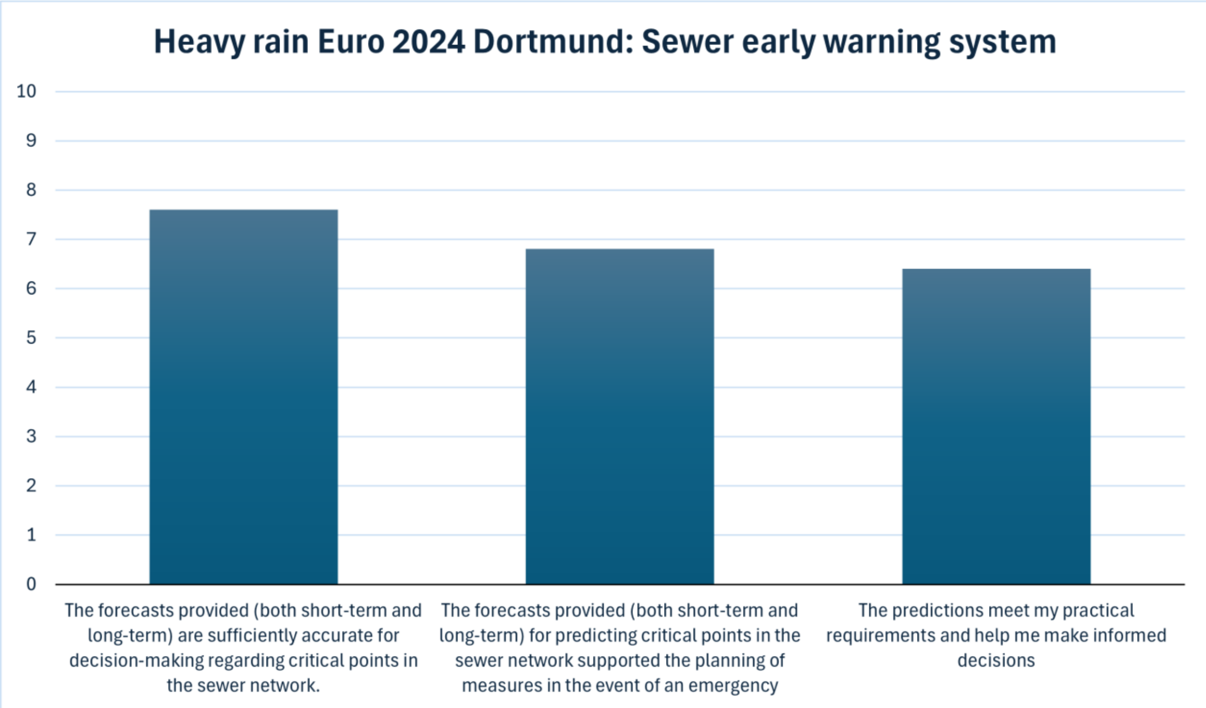
Results Complex Event Forecasting (Innsbruck)

expert #	**11**	**12**	**13**	**14**	**main**
The forecasts provided (both short- and long-term) are sufficiently accurate for decision-making regarding critical points in the sewer network	2	-	7	6	**5**
The provided forecasts (both short- and long-term) for predicting critical points in the sewer network support operational planning in emergency situations	4	-	9	7	**6,67**
The forecasts meet my practical requirements and help me make informed decisions	2	9	8	6	**6,25**



Results Sewer network early warning system (Dortmund)

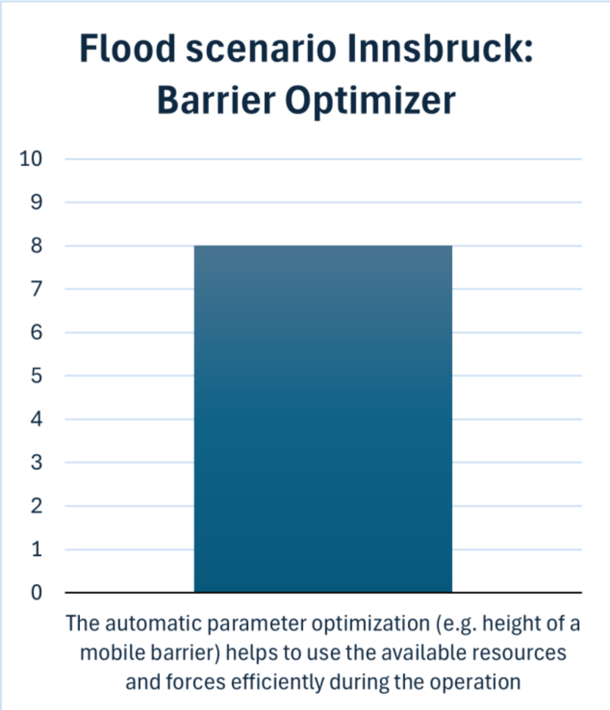
expert #	**70**	**71**	**72**	**73**	**74**	**75**	**77**	**main**
The forecasts provided (both short-term and long-term) are sufficiently accurate for decision-making regarding critical points in the sewer network.	7	7	6	8	10	-	-	**7,6**
The forecasts provided (both short-term and long-term) for predicting critical points in the sewer network supported the planning of measures in the event of an emergency	9	7	5	8	5	-	-	**6,8**
The predictions meet my practical requirements and help me make informed decisions	8	8	5	8	3	-	-	**6,4**



9.7.5.2 Interactive Learning

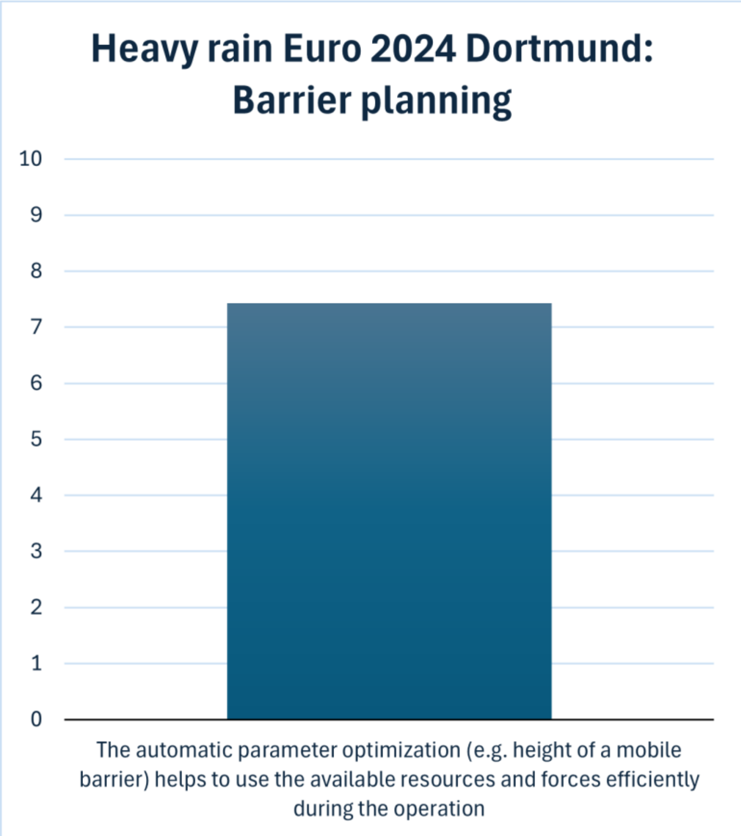
Results Barrier Optimizer (Innsbruck)

expert #	**11**	**12**	**13**	**14**	**main**
The automatic parameter optimization (e.g. height of a mobile barrier) helps to use the available resources and forces efficiently during the operation	8	10	7	7	**8**



Results Barrier planning (Dortmund)

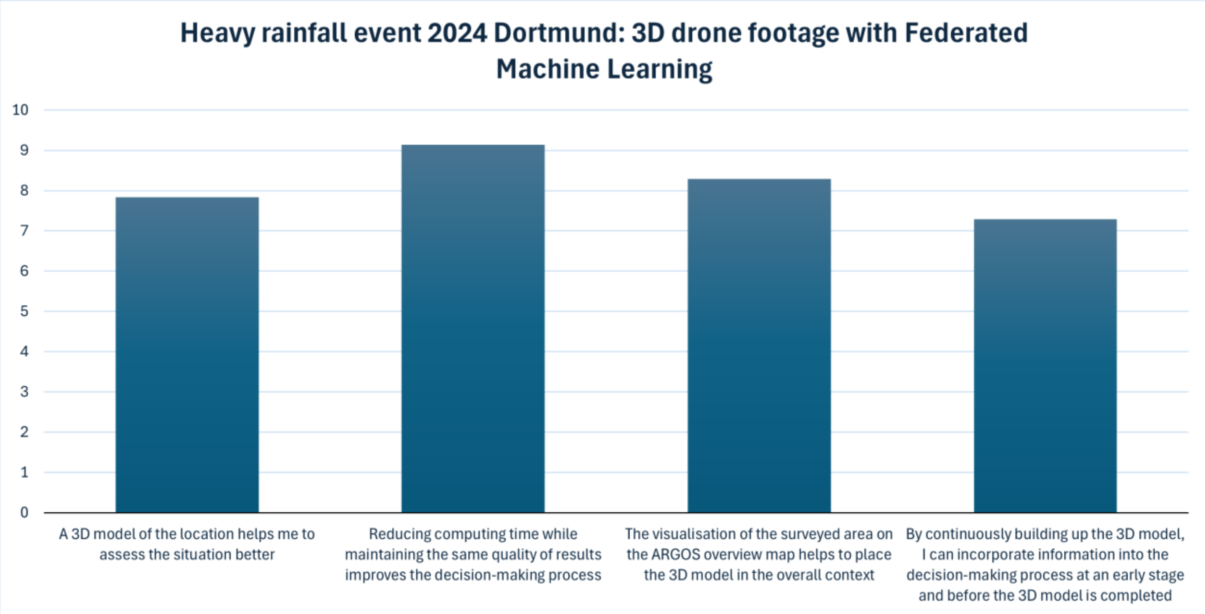
expert #	**70**	**71**	**72**	**73**	**74**	**75**	**77**	**main**
The automatic parameter optimization (e.g. height of a mobile barrier) helps to use the available resources and forces efficiently during the operation	8	10	7	4	10	7	6	**7,43**



9.7.5.3 Federated Machine Learning

Results 3D drone footage with Federated Machine Learning (Dortmund)

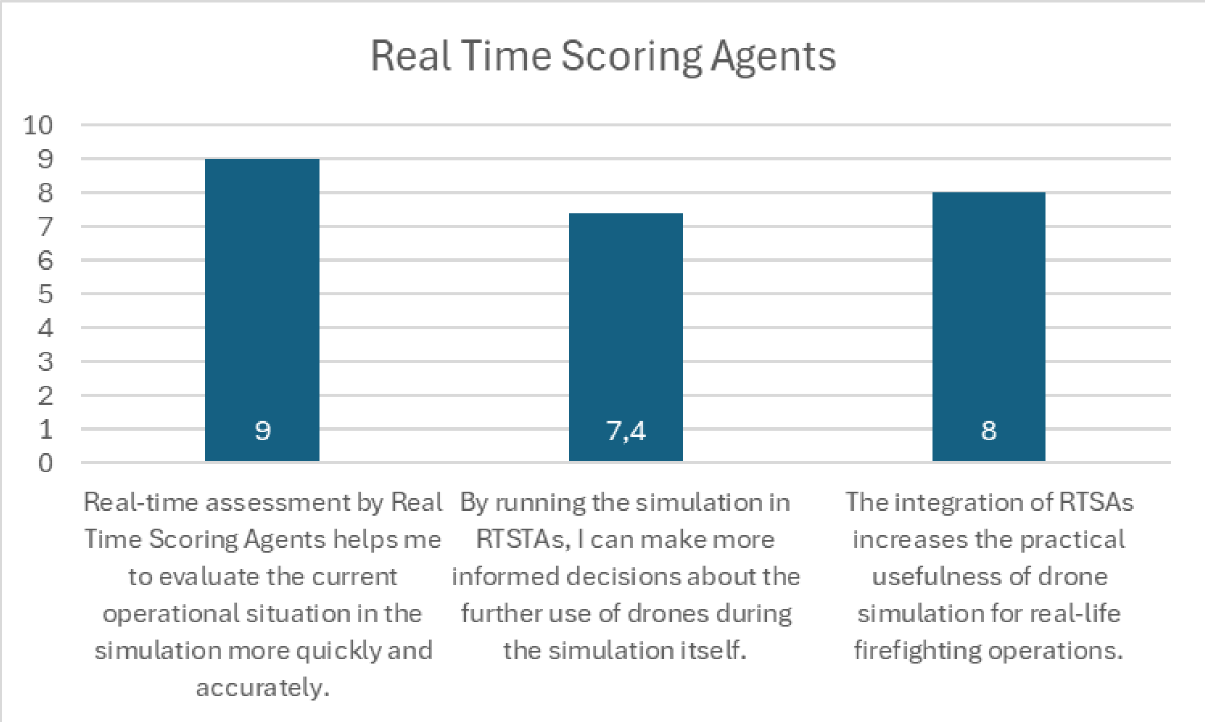
expert #	**90**	**91**	**92**	**93**	**94**	**95**	**96**	**97**	**main**
A 3D model of the location helps me to assess the situation better	10	-	8	9	7	4	-	9	**7,83**
Reducing computing time while maintaining the same quality of results improves the decision-making process	9	10	10	10	10	5	-	10	**9,14**
The visualisation of the surveyed area on the ARGOS overview map helps to place the 3D model in the overall context	10	8	8	10	8	5	-	9	**8,29**
By continuously building up the 3D model, I can incorporate information into the decision-making process at an early stage and before the 3D model is completed	7	8	8	10	6	5	-	7	**7,29**



9.7.5.4 Workflow Optimization

RTSA

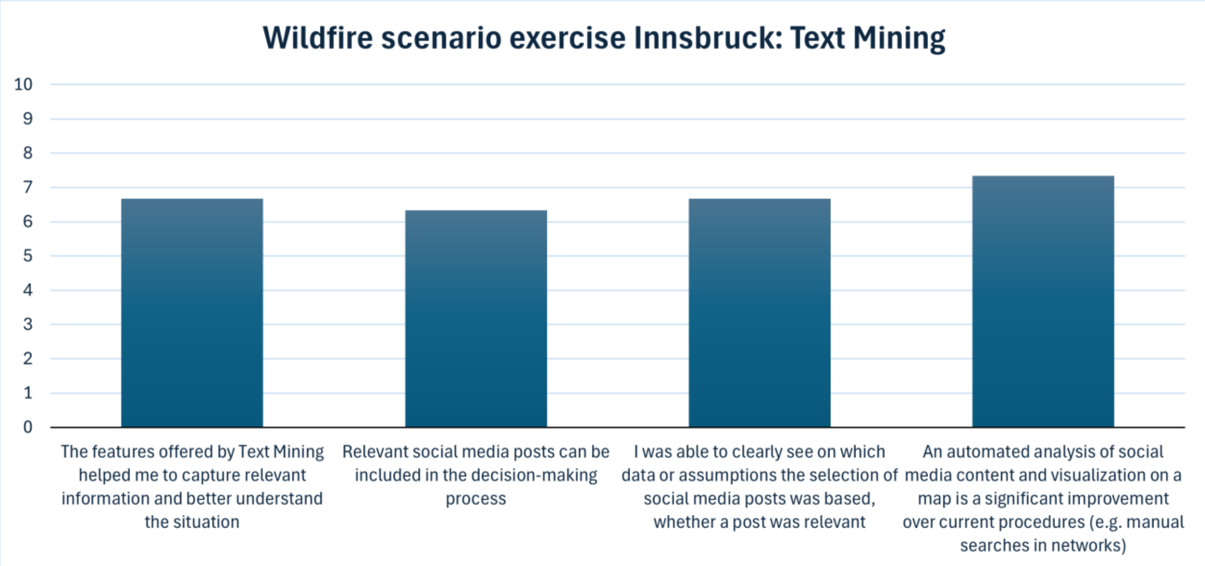
expert #	**60**	**61**	**62**	**63**	**64**	**main**
Real-time assessment by Real Time Scoring Agents helps me to evaluate the current operational situation in the simulation more quickly and accurately.	10	6	9	10	10	**9,00**
By running the simulation in RTSTAs, I can make more informed decisions about the further use of drones during the simulation itself.	8	3	9	7	10	**7,40**
The integration of RTSAs increases the practical usefulness of drone simulation for real-life firefighting operations.	7	5	9	9	10	**8,00**



9.7.5.5 Text Mining

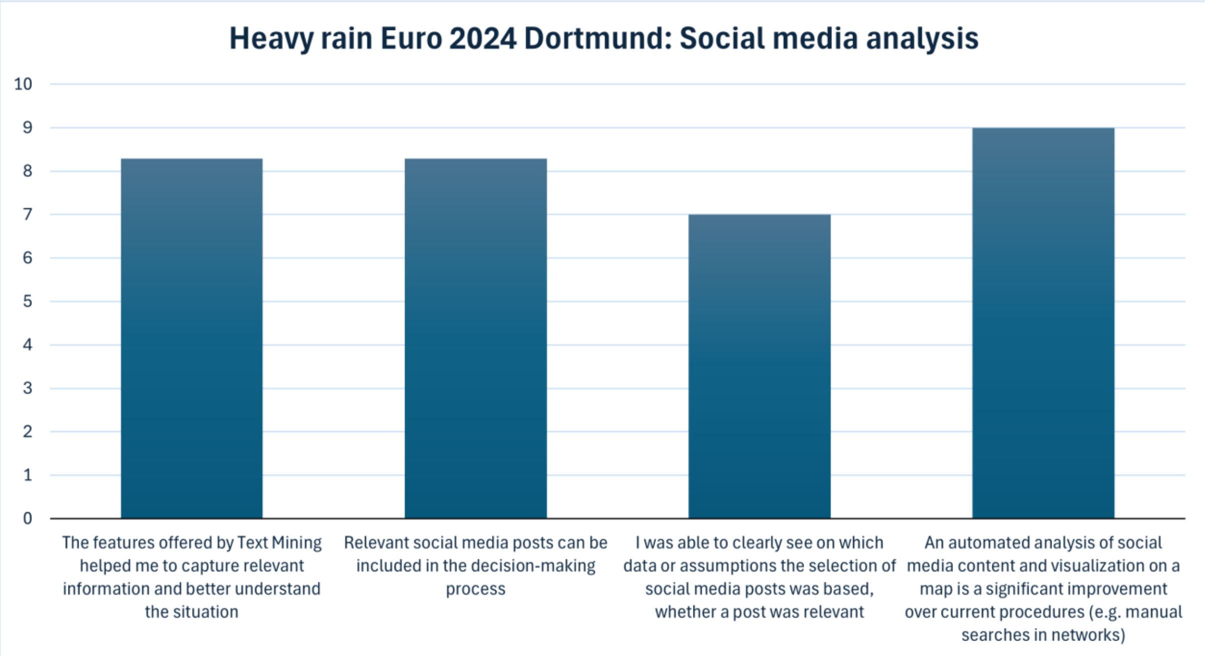
Results Text Mining (Innsbruck)

expert #	**21**	**23**	**24**	**25**	**26**	**main**
The features offered by Text Mining helped me to capture relevant information and better understand the situation	-	7	8	-	5	**6,67**
Relevant social media posts can be included in the decision-making process	-	7	9	-	3	**6,33**
I was able to clearly see on which data or assumptions the selection of social media posts was based, whether a post was relevant	-	6	10	-	4	**6,67**
An automated analysis of social media content and visualization on a map is a significant improvement over current procedures (e.g. manual searches in networks)	-	8	10	-	4	**7,33**



Results Social media analysis (Dortmund)

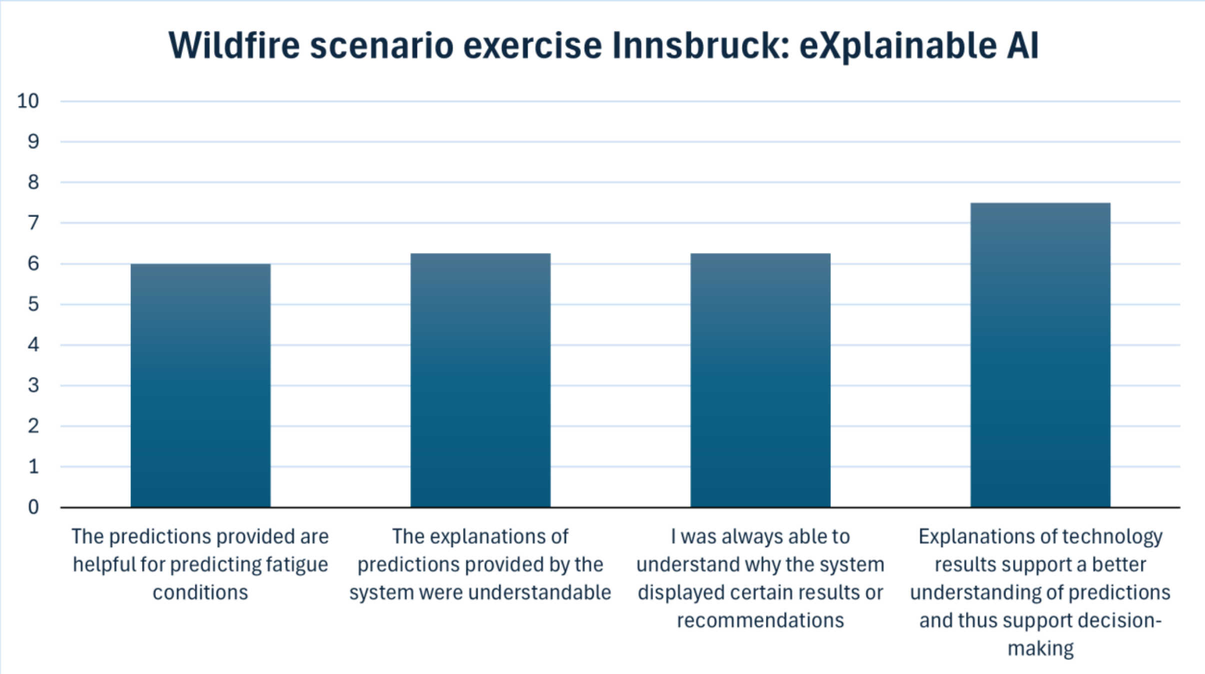
expert #	70	71	72	73	74	75	77	main
The features offered by Text Mining helped me to capture relevant information and better understand the situation	8	10	6	8	10	7	9	**8,29**
Relevant social media posts can be included in the decision-making process	9	10	5	9	10	6	9	**8,29**
I was able to clearly see on which data or assumptions the selection of social media posts was based, whether a post was relevant	6	10	5	7	5	6	10	**7**
An automated analysis of social media content and visualization on a map is a significant improvement over current procedures (e.g. manual searches in networks)	9	10	9	9	10	6	10	**9**



9.7.5.6 eXplainable AI

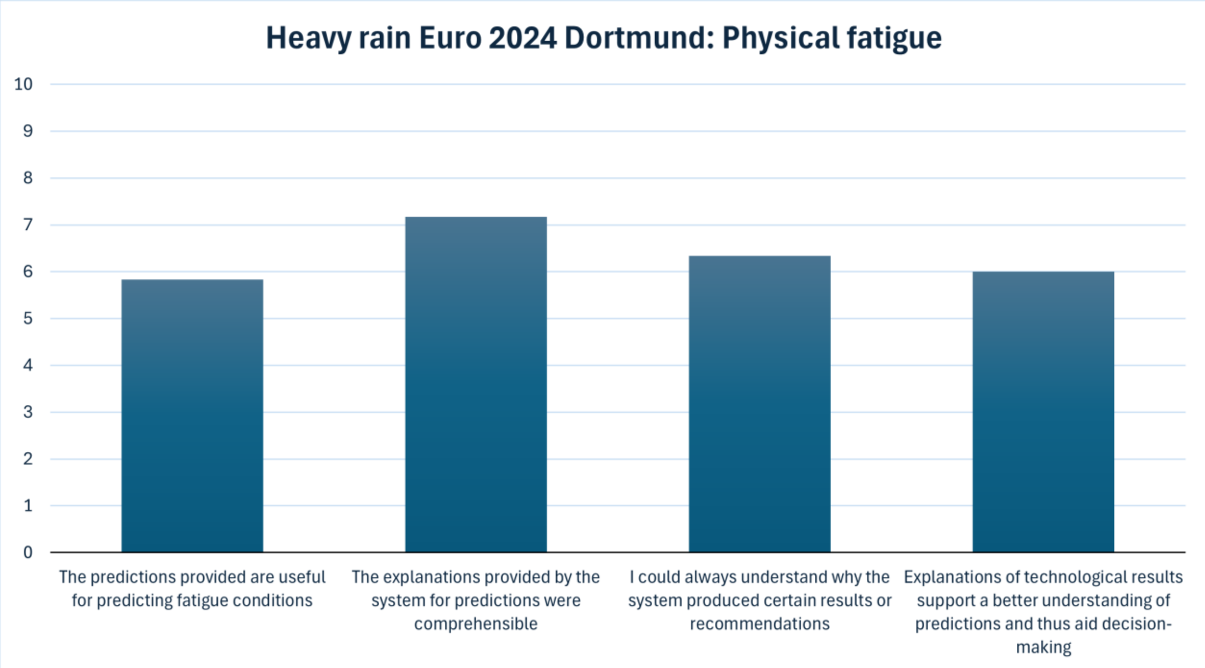
Results eXplainable AI (Innsbruck)

expert #	**21**	**23**	**24**	**25**	**26**	**main**
The predictions provided are helpful for predicting fatigue conditions	8	5	5	6	8	**6**
The explanations of predictions provided by the system were understandable	8	5	5	7	8	**6,25**
I was always able to understand why the system displayed certain results or recommendations	7	8	5	5	7	**6,25**
Explanations of technology results support a better understanding of predictions and thus support decision-making	8	8	7	7	8	**7,5**



Results Physical fatigue (Dortmund)

expert #	70	71	72	73	74	75	77	main
The predictions provided are useful for predicting fatigue conditions	6	5	5	6	-	4	9	**5,83**
The explanations provided by the system for predictions were comprehensible	9	6	5	6	-	7	10	**7,17**
I could always understand why the system produced certain results or recommendations	8	6	5	6	-	6	7	**6,3**
Explanations of technological results support a better understanding of predictions and thus aid decision-making	9	5	3	5	-	5	9	**6**

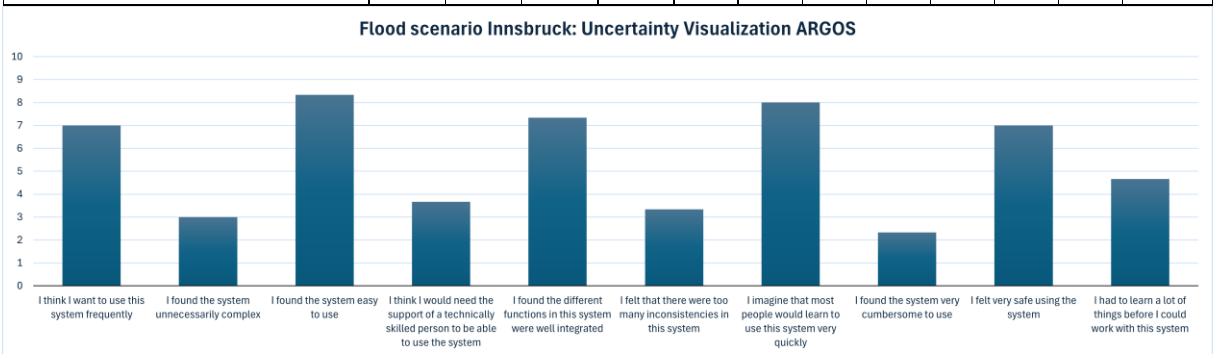


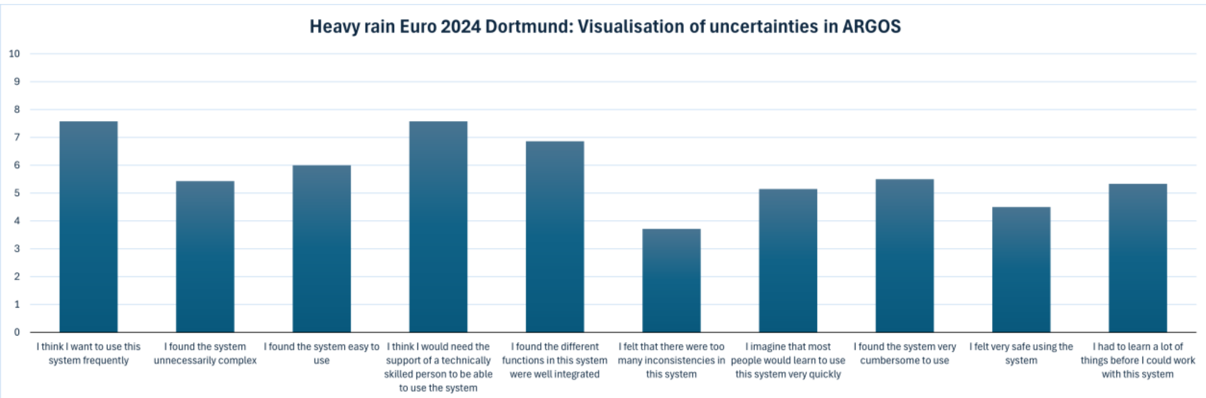
9.7.5.7 Augmented Reality

Results Uncertainty Visualization ARGOS (Innsbruck & Dortmund Flood Scenario)

expert #	**11**	**12**	**13**	**14**	**70**	**71**	**72**	**73**	**74**	**75**	**77**	**main**
I think I want to use this system frequently	4	-	10	7	8	7	7	8	10	5	8	**7,29**
I found the system unnecessarily complex	5	2	1	4	6	6	6	4	3	6	7	**4,21**
I found the system easy to use	-	9	9	7	4	6	7	7	7	7	4	**7,17**
I think I would need the support of a technically skilled person to be able to use the system	-	6	2	3	8	8	8	6	8	5	10	**5,62**
I found the different functions in this system were well integrated	-	9	8	5	6	6	8	6	8	5	9	**7,1**
I felt that there were too many inconsistencies in this system	-	2	3	5	2	6	3	7	2	4	2	**3,52**

I imagine that most people would learn to use this system very quickly	-	9	9	6	3	5	7	6	6	3	6	**6,57**
I found the system very cumbersome to use	-	2	1	4	6	5	4	6	-	4	8	**3,92**
I felt very safe using the system	-	8	7	6	2	5	6	6	-	5	3	**5,75**
I had to learn a lot of things before I could work with this system	-	3	8	3	8	8	3	5	-	4	4	**5**

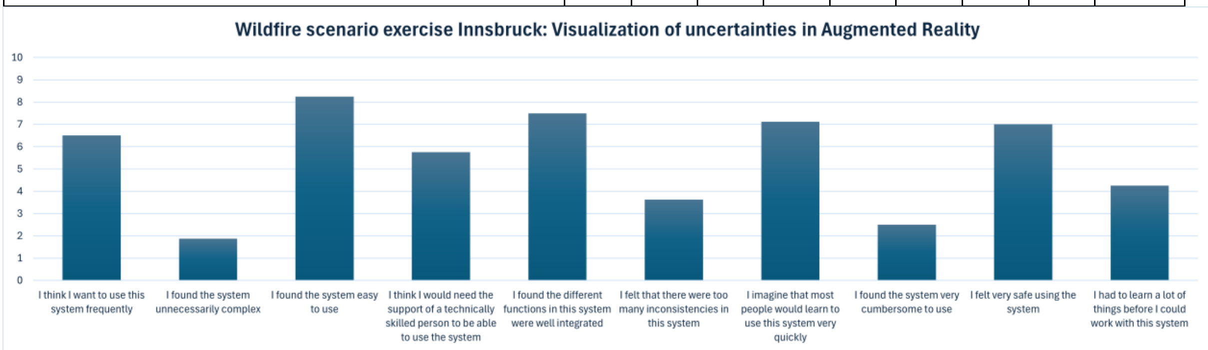




Results Visualization of uncertainties in Augmented Reality (Innsbruck)

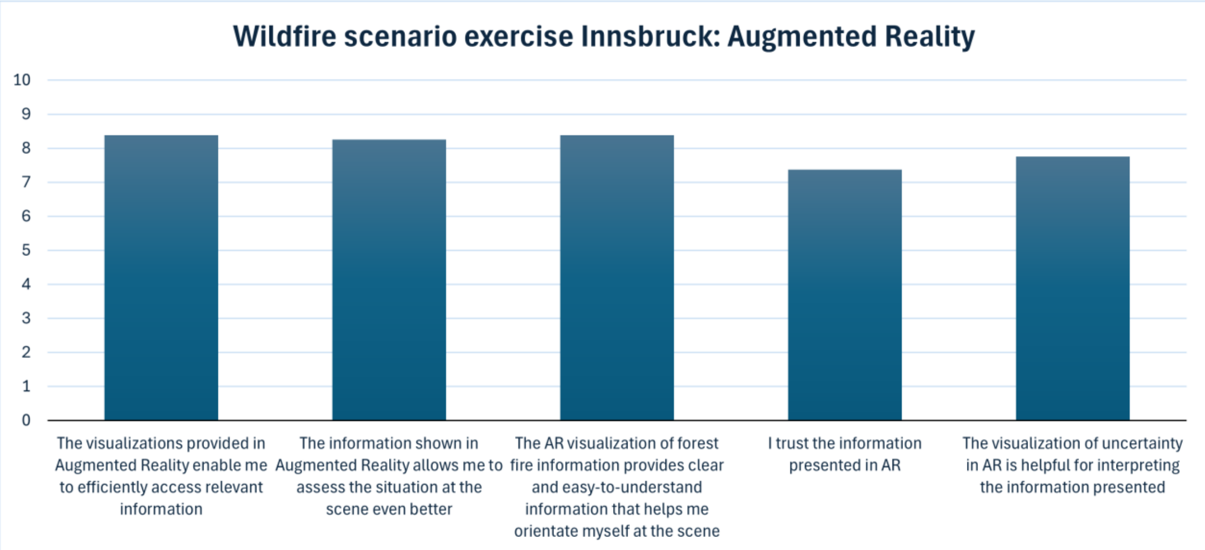
expert #	**51**	**52**	**53**	**54**	**55**	**56**	**57**	**58**	**main**
I think I want to use this system frequently	6	10	6	6	4	4	6	10	**6,5**
I found the system unnecessarily complex	2	1	2	4	2	2	1	1	**1,88**
I found the system easy to use	6	8	8	9	9	8	8	10	**8,25**
I think I would need the support of a technically skilled person to be able to use the system	7	3	9	4	9	4	9	1	**5,75**

I found the different functions in this system were well integrated	8	8	7	8	6	8	7	8	**7,5**
I felt that there were too many inconsistencies in this system	5	1	5	5	6	5	1	1	**3,63**
I imagine that most people would learn to use this system very quickly	8	10	8	7	3	6	9	6	**7,13**
I found the system very cumbersome to use	4	3	3	3	3	2	1	1	**2,5**
I felt very safe using the system	7	10	4	7	5	8	5	10	**7**
I had to learn a lot of things before I could work with this system	3	4	7	3	8	1	7	1	**4,25**



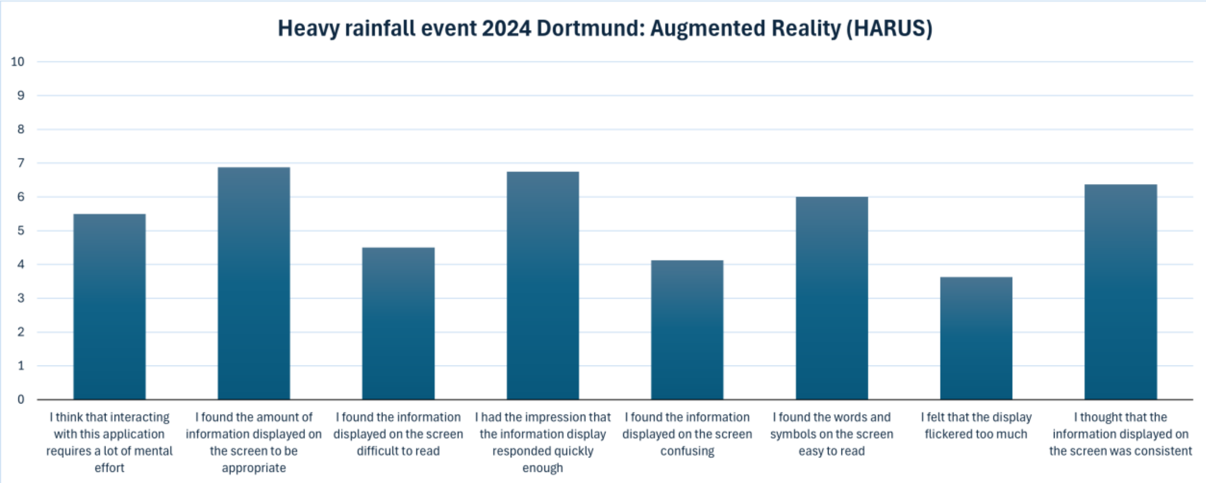
Results Augmented Reality (Innsbruck)

expert #	**51**	**52**	**53**	**54**	**55**	**56**	**57**	**58**	**main**
The visualizations provided in Augmented Reality enable me to efficiently access relevant information	7	10	10	8	8	6	8	10	**8,38**
The information shown in Augmented Reality allows me to assess the situation at the scene even better	8	10	8	9	5	7	9	10	**8,25**
The AR visualization of forest fire information provides clear and easy-to-understand information that helps me orientate myself at the scene	8	10	9	7	8	7	8	10	**8,38**
I trust the information presented in AR	8	10	7	6	8	5	7	8	**7,38**
The visualization of uncertainty in AR is helpful for interpreting the information presented	8	10	8	7	7	5	7	10	**7,75**



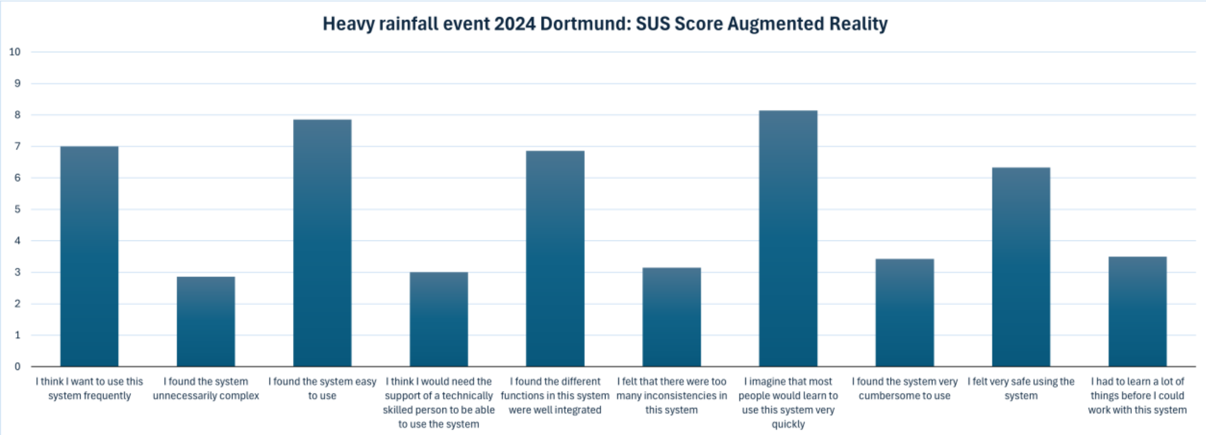
Results Augmented Reality (HARUS) (Dortmund)

expert #	**90**	**91**	**92**	**93**	**94**	**95**	**96**	**97**	**main**
I think that interacting with this application requires a lot of mental effort	3	4	7	7	3	5	8	7	**5,5**
I found the amount of information displayed on the screen to be appropriate	10	7	3	5	9	6	9	6	**6,88**
I found the information displayed on the screen difficult to read	7	6	6	2	3	6	2	4	**4,5**
I had the impression that the information display responded quickly enough	10	8	3	8	7	3	9	6	**6,75**
I found the information displayed on the screen confusing	1	3	7	7	2	4	4	5	**4,13**
I found the words and symbols on the screen easy to read	2	5	5	9	7	4	9	7	**6**
I felt that the display flickered too much	1	1	8	2	6	6	1	4	**3,63**
I thought that the information displayed on the screen was consistent	10	2	7	7	8	5	7	5	**6,38**



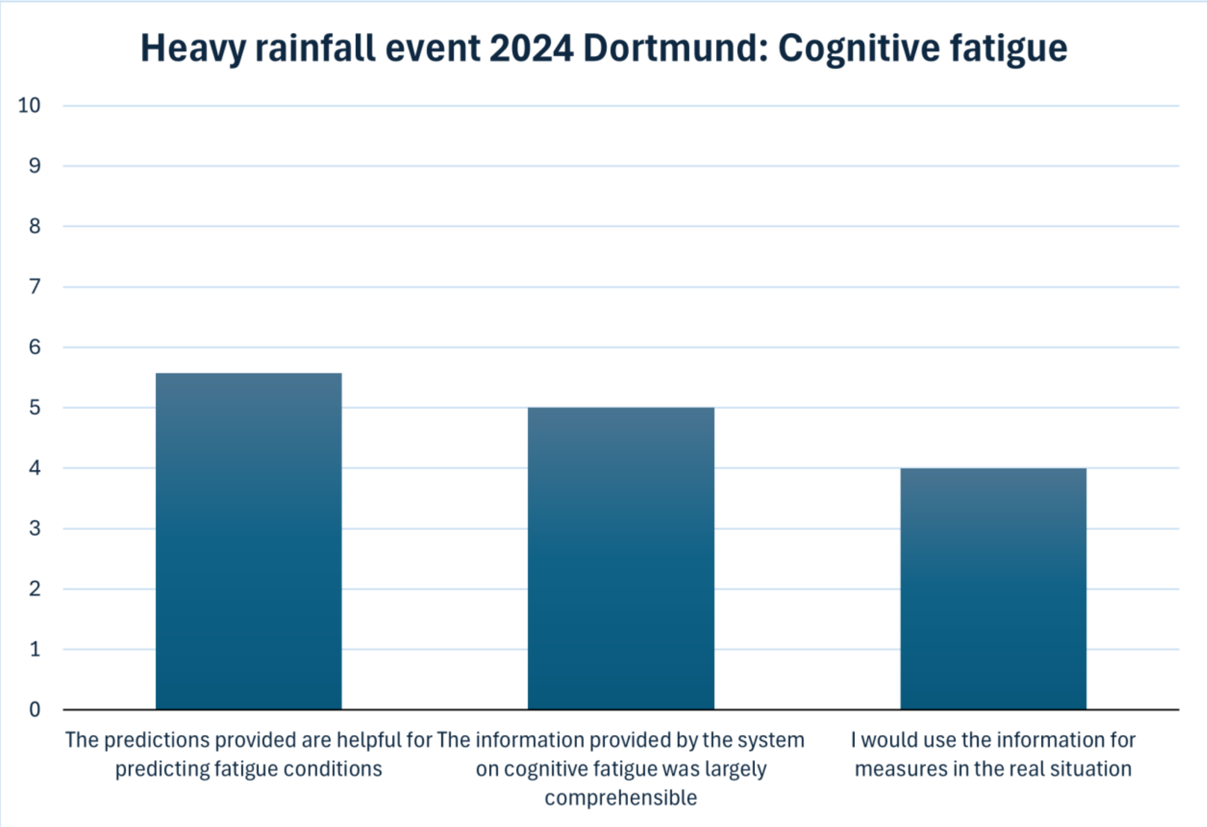
Results SUS Score Augmented Reality (Dortmund)

expert #	90	91	92	93	94	95	96	97	main
I think I want to use this system frequently	8	-	7	8	9	3	7	-	**7**
I found the system unnecessarily complex	1	3	3	3	3	4	3	-	**2,86**
I found the system easy to use	10	6	9	6	10	4	10	-	**7,86**
I think I would need the support of a technically skilled person to be able to use the system	1	2	8	4	2	3	1	-	**3**
I found the different functions in this system were well integrated	5	8	7	6	9	4	9	-	**6,86**
I felt that there were too many inconsistencies in this system	1	2	6	4	1	7	1	-	**3,14**
I imagine that most people would learn to use this system very quickly	10	7	9	7	10	4	10	-	**8,14**
I found the system very cumbersome to use	1	6	3	4	2	7	1	-	**3,43**
I felt very safe using the system	10	3	8	4	9	4	-	-	**6,33**
I had to learn a lot of things before I could work with this system	3	4	3	4	2	5	-	-	**3,5**



Results Cognitive fatigue (Dortmund)

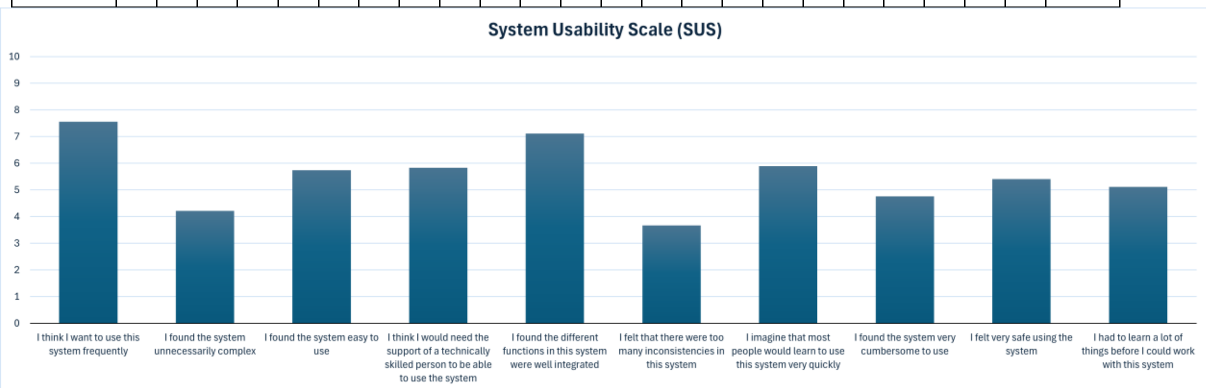
expert #	**90**	**91**	**92**	**93**	**94**	**95**	**96**	**97**	**main**
The predictions provided are helpful for predicting fatigue conditions	6	7	8	6	5	1	-	6	**5,57**
The information provided by the system on cognitive fatigue was largely comprehensible	8	5	6	4	5	1	-	6	**5**
I would use the information for measures in the real situation	6		1	4	5	1	-	7	**4**


Heavy rainfall event 2024 Dortmund: Cognitive fatigue
10
9
8
7
6
5
4
3
2
1
0
The predictions provided are helpful for predicting fatigue conditions
The information provided by the system on cognitive fatigue was largely comprehensible
I would use the information for measures in the real situation

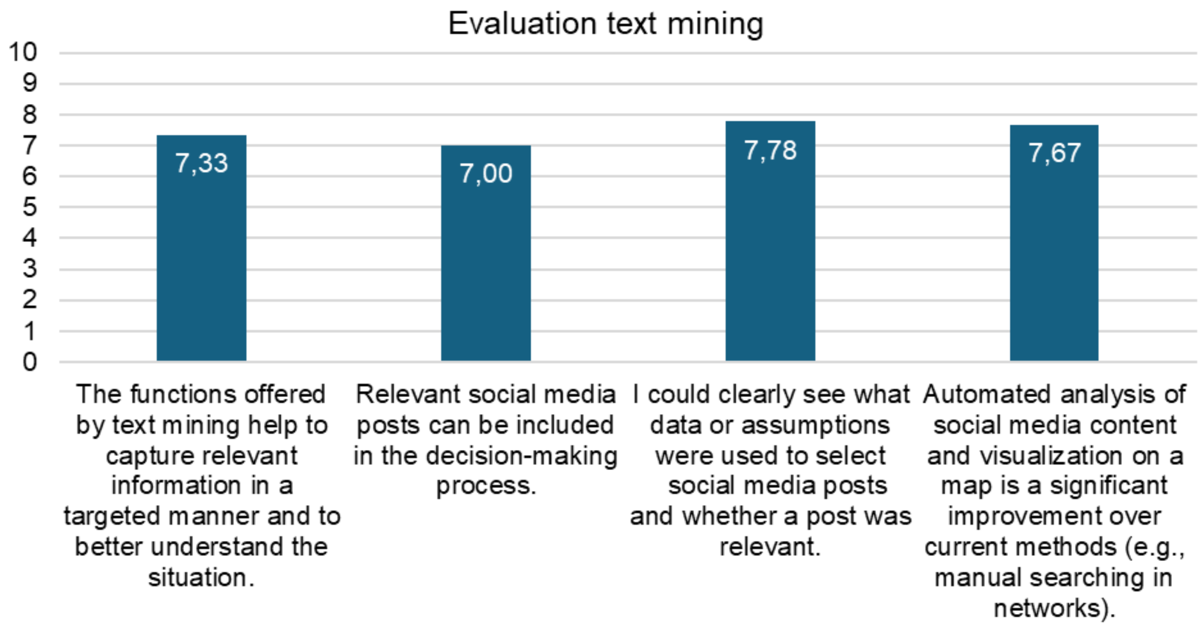
9.8 SUS questionnaires

Results System Usability Scale (SUS)

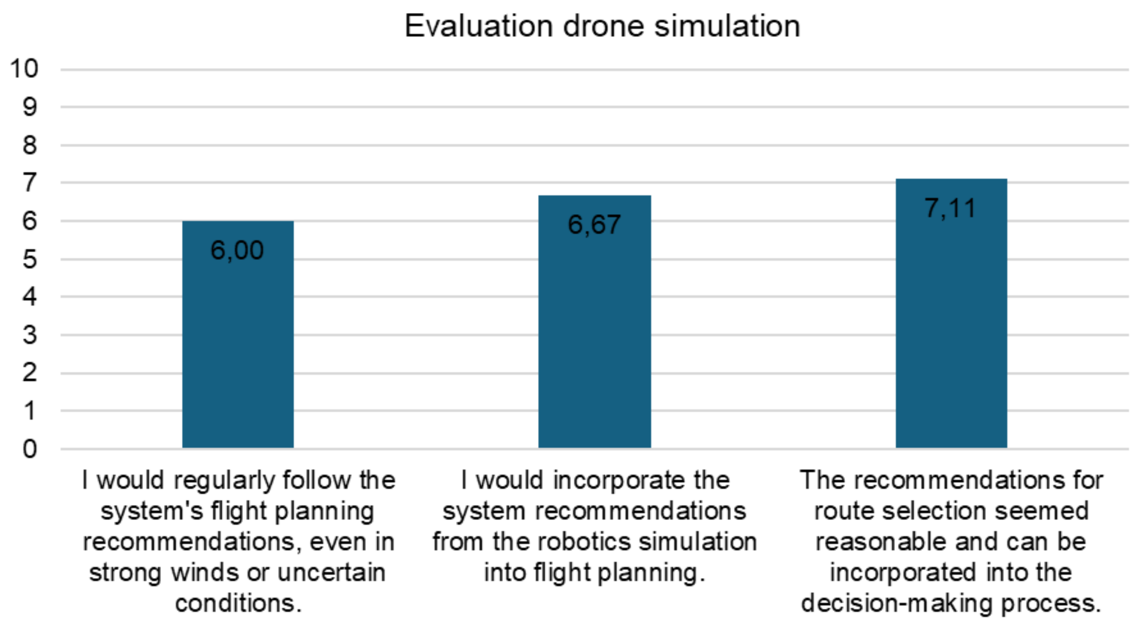
expert #	60	61	62	63	64	65	70	71	72	73	74	75	77	90	91	92	93	94	95	96	97	98	99	main
I think I want to use this system frequently	8	6	10	10	8	4	8	8	7	7	10	6	8	10	9	8	8	7	4	9	10	10	9	8,00
I found the system unnecessarily complex	3	5	3	1	2	6	3	6	3	5	5	7	5	2	4	3	4	7	5	2	1	1	2	3,70
I found the system easy to use	6	4	5	9	7	5	3	6	6	7	5	5	3	8	6	8	7	4	4	10	9	8	10	6,30
I think I would need the support of a technically skilled person to be able to use the system	2	8	10	2	5	6	10	8	7	5	6	5	9	3	8	5	4	6	5	1	2	1	1	5,17
I found the different functions in this system were well integrated	8	6	9	7	8	4	8	7	7	7	7	5	9	9	10	8	8	5	4	9	7	8	9	7,35
I felt that there were too many inconsistencies in this system	3	3	1	1	2	7	2	5	3	5	3	6	3	2	1	3	4	5	7	2	1	3	2	3,22
I imagine that most people would learn to use this system very quickly	8	4	6	9	6	3	5	7	6	7	6	3	4	8	8	9	7	4	4	9	9	10	9	6,57
I found the system very cumbersome to use	2	5	6	1	2	7	7	5	3	4	3	8	8	3	1	2	4	6	7	1	1	2	1	3,87
I felt very safe using the system	8	4	6	10	6	4	2	6	6	6	6	4	4	8	10	10	4	4	4	10	10	9	10	6,57
I had to learn a lot of things before I could work with this system	3	8	4	1	6	7	9	6	3	5	5	5	6	2	3	1	4	7	5	2	1	1	2	4,17



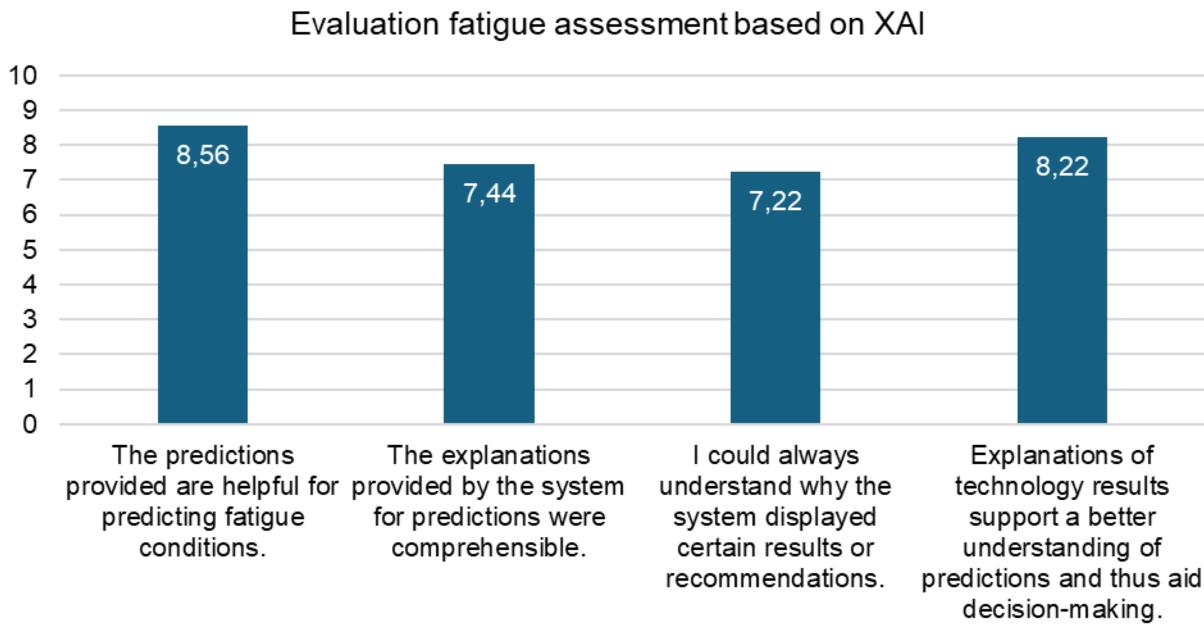
9.9 Results ESAF exercise exploring transferbility



Evaluation results for text mining (n=9)



Evaluation results for drone simulation (n=9)



Evaluation results for XAI-based fatigue assessment (n=9)

10 Appendix 2: Expert users' questionnaire of the Life Sciences Use Case

1. **User background information**

1.1. What is your job title?

Epidemiological Scenario

1.1.1. Full Professor
1.1.2. Posdoctoral researcher
1.1.3. Group leader

Multiscale Infection Scenario

1.1.4. Principal Investigator at CSIC
1.1.5. Associate Professor
1.1.6. Postdoctoral researcher

1.2. Do you work in academia or in industry?

Epidemiological Scenario

1.2.1. Academia
1.2.2. Academia
1.2.3. Academia

Multiscale Infection Scenario

1.2.4. Academia, joint CSIC - University of Valencia research centre
1.2.5. Academia
1.2.6. Academia

1.3. Years of experience

Epidemiological Scenario

1.3.1. 17
1.3.2. 6
1.3.3. 21 (since PhD), 12 in current position

Multiscale Infection Scenario

1.3.4. 15 years
1.3.5. 17
1.3.6. 6

1.4. Background studies (university degree major, etc.)

Epidemiological Scenario

1.4.1. PhD
1.4.2. Biomedical science / Global Health
1.4.3. PhD in Physics

Multiscale Infection Scenario

1.4.4. Bsc Biology, MSc Biodiversity: Conservation and Evolution, PhD Oceanography
1.4.5. Ph.D., D.Eng.
1.4.6. PhD in computational biology

1.5. What are your main job tasks?

Epidemiological Scenario

1.5.1. Lecturing and Research
1.5.2. Health impact modelling of urban and transport planning interventions.
1.5.3. My main task is to manage the group (i.e. arrange and delegate tasks, procure funding, devise strategy).

Multiscale Infection Scenario

1.5.4. Doing research in computational microbial ecology including the mentoring of students.
1.5.5. –
1.5.6. To develop and apply multiscale modeling methods to cancer research

1.6. To whom are you responsible for performing these tasks?

Epidemiological Scenario

1.6.1. University
1.6.2. PI / project coordinator
1.6.3. Department director

Multiscale Infection Scenario

1.6.4. To CSIC directors and the research centre director
1.6.5. Head of Department
1.6.6. Head of Department

2. Modelling background of the user

2.1. Are you a model developer or do you use models already developed?

Epidemiological Scenario

2.1.1. Both (more developer than user)
2.1.2. Both
2.1.3. Both

Multiscale Infection Scenario

2.1.4. I use models already developed
2.1.5. Model developer
2.1.6. Model developer

2.2. How relevant is in your research the use of models and simulation for addressing questions (low / medium / very high)

Epidemiological Scenario

2.2.1. Very high
2.2.2. High (for modelling chronic diseases)
2.2.3. Very high

Multiscale Infection Scenario

2.2.4. Between low and medium
2.2.5. Very high
2.2.6. Very high

2.3. What kind of models and simulations do you use to analyse systems?

Epidemiological Scenario

2.3.1. Epidemiological models
2.3.2. I work with public health models. Specifically, my work is focused on linking urban and transport planning interventions to changes in environmental and lifestyle exposures (e.g., air pollution, noise, green spaces, physical activity) and subsequently to population health, focusing on chronic morbidity and premature mortality.
2.3.3. Some deterministic (ODEs, PDEs) like for fluids or chemistry, and some statistical (ML, DL, RL) based on data for predicting behavior that is hard to model otherwise.

Multiscale Infection Scenario

2.3.4. Time series simulations, machine learning classification and regressions
2.3.5. Finite element method, boundary element method, isogeometric analysis, agent-based models, finite volume method, multiscale / multiphysics
2.3.6. Multiscale mechanistic models, agent-based models, Boolean models, network models, ODEs models

2.4. How much would you regard yourself as a model developer?

Epidemiological Scenario

2.4.1. Very much
2.4.2. I have experience working with models, but I wouldn’t consider myself primarily a model developer. My focus is more on using existing models to inform research and policy rather than building them from scratch.
2.4.3. Not so much anymore that I mostly supervise people

Multiscale Infection Scenario

2.4.4. Not much a developer, I mainly use models
2.4.5. –
2.4.6. yes

2.5. Which modelling and simulation approaches (e.g. ODEs, Stochastic, ABMs) and tools (e.g. programming/modelling languages, libraries, frameworks) do you use in your research?

Epidemiological Scenario

2.5.1. Agent-based models, compartmental, empirical (data-driven). Matlab/Python/R

2.5.2. ABMs, health microsimulation, R, Python.

2.5.3. Mostly PDE models and ABMs. I did most of my models in Fortran, but currently use Python more (either directly for the models or for driving them). For ABMs I use Mesa. For statistical models, I rely on Tensorflow and Pytorch.

Multiscale Infection Scenario

2.5.4. I use ODEs and stochastic models, using scripting languages like Python and R

2.5.5. for simulation approaches see 2.3.1; programming: C, C++, FORTRAN, Matlab; libraries: MPICH, boost, GSL, PETSc, Trilinos, SuperLU, VTK, Tetgen, OpenFOAM, libmesh, BioDynaMo

2.5.6. C++, python, PhysiCell, MaBoSS, NeKo, Copasi, BioMASS

2.6. Do you work with real-time data? If not, would you like to work with this kind of data?

Epidemiological Scenario

2.6.1.1. I work with weekly reported data (not real time).

2.6.1.2. No, but could be interesting.

2.6.1.3. Rarely.

Multiscale Infection Scenario

2.6.1.4. No, but would like to.

2.6.1.5. Yes

2.6.1.6. I do not work with real-time data but I would like to

2.6.2. Are your tools and workflows able to work with such data?

Epidemiological Scenario

2.6.2.1. Currently yes, but it can be further improved.

2.6.2.2. Not at the moment.

2.6.2.3. Yes, but it's not the typical case.

Multiscale Infection Scenario

2.6.2.4. No

2.6.2.5. Partially

2.6.2.6. Yes

2.6.3. Would you be interested in using a data processing workflow that would allow using real-time data?

Epidemiological Scenario

2.6.3.1. Yes, that would be useful.

2.6.3.2. Yes

2.6.3.3. Yes

Multiscale Infection Scenario

2.6.3.4. Yes

2.6.3.5. Yes
2.6.3.6. Yes

2.6.4. Would using this data allow you to address different problems than the ones you are currently addressing?

Epidemiological Scenario

2.6.4.1. Not necessarily, but we would be more efficient analysing the current data we have.
2.6.4.2. Yes
2.6.4.3. Some yes, like anomaly detection or fast response to failures.

Multiscale Infection Scenario

2.6.4.4. Yes, if this was possible.
2.6.4.5. Yes
2.6.4.6. Yes

2.7. Do you currently use forecasting techniques? Are there specific events that you would like to forecast in real-time, which you currently cannot forecast?

Epidemiological Scenario

2.7.1.1. Yes, we apply techniques to perform weekly forecast of seasonal respiratory diseases. Yes, we would like to forecast number of hospitalization or primary care attendances, emergency departments call, to ultimately forecast their impact over the health system.
2.7.1.2. I estimate the long-term impacts on health of urban and transport planning interventions (e.g., low emission zones, sustainable transport, greening interventions). It could be interesting to forecast the real-time impacts of extreme weather events on such interventions (e.g. disrupting the use of active transport or the use of public space).
2.7.1.3. Yes, we forecast performance of machines, and failures.

Multiscale Infection Scenario

2.7.1.4. I do currently use forecasting techniques. There are specific events that you would like to forecast in real-time, but the lack of real time data is preventing me from forecasting more.
2.7.1.5. No + Yes (using RWD for cancer modelling)
2.7.1.6. No

transparent short- and long-term forecasts to end-users, explainable via advanced visual analytics and accurate, real-time, off and on-site augmented reality facilities.

3. **Please rate these objectives of the CREXDATA project according to your background and present and future needs (1: Not useful, 2: Of some use, 3: Average Use, 4: Quite useful, 5: Very useful):**

3.1. Being able to have extreme-scale data ingestion/generation, fusion and exploitation.

3.1.1. Ingesting multimodal data (images, simulations, social media publications, etc).

<u>Epidemiological Scenario</u>

3.1.1.1. 5, 4, 4

<u>Multiscale Infection Scenario</u>

3.1.1.2. 4, 5, 5

3.1.2. Using dynamic modelling to predict the systems' behaviour.

<u>Epidemiological Scenario</u>

3.1.2.1. 4, 5, 5

<u>Multiscale Infection Scenario</u>

3.1.2.2. 5, 5, 5

3.1.3. Handling multilingual social data in real-time.

<u>Epidemiological Scenario</u>

3.1.3.1. 2, 2, 3

<u>Multiscale Infection Scenario</u>

3.1.3.2. 4, 2, 1

3.2. Having real-time predictive knowledge and forecasts.

3.2.1. Using online federated learning.

<u>Epidemiological Scenario</u>

3.2.1.1. 3, 4, 5

<u>Multiscale Infection Scenario</u>

3.2.1.2. 4, NA, 1

3.2.2. Having multiresolution complex event forecasting under uncertainty.

<u>Epidemiological Scenario</u>

3.2.2.1. 4, 4, 5

<u>Multiscale Infection Scenario</u>

3.2.2.2. 5, 4, 2

3.2.3. Using optimization techniques for Prediction-as-a-Service (PaaS).

<u>Epidemiological Scenario</u>

3.2.3.1. 4, 4, 4

<u>Multiscale Infection Scenario</u>

3.2.3.2. 4, 5, 2

3.3. Reducing the perceived complexity.

3.3.1. Using graphical workflow design.

<u>Epidemiological Scenario</u>

3.3.1.1. 4.5, 4, 4

Multiscale Infection Scenario
3.3.1.2. 5, 2, 4
3.3.2. Using visual analytics coupled with XAI for understanding complexity and reasoning under uncertainty.
Epidemiological Scenario
3.3.2.1. 4, 4, 5
Multiscale Infection Scenario
3.3.2.2. 5, NA, 3
3.3.3. Using augmented reality under uncertainty on-site & remotely.
Epidemiological Scenario
3.3.3.1. 1, 2, 3
Multiscale Infection Scenario
3.3.3.2. 3, NA, 1

4. Specific aspects of the health emergency use case (1: Not useful, 2: Of some use, 3: Average Use, 4: Quite useful, 5: Very useful):

4.1. Rate the following components of the health emergency use case according to your interest on the foreseen results

4.1.1. Using a graphical user interphase
Epidemiological Scenario
4.1.1.1. 5, 5, 5
Multiscale Infection Scenario
4.1.1.2. 5, 4, 5
4.1.2. Being able to have parameter calibration
Epidemiological Scenario
4.1.2.1. 5, 5, 4
Multiscale Infection Scenario
4.1.2.2. 5, 5, 5
4.1.3. Using early time series characterisation
Epidemiological Scenario
4.1.3.1. 5, 4, 4
Multiscale Infection Scenario
4.1.3.2. 4, 5, 5
4.1.4. Online model exploration
Epidemiological Scenario
4.1.4.1. 4, 4, 4
Multiscale Infection Scenario
4.1.4.2. 5, 4, 5
4.1.5. Real-time/online forecasting simulation trajectories
Epidemiological Scenario
4.1.5.1. 3, 4, 5
Multiscale Infection Scenario
4.1.5.2. 5, 3, 5

4.2. Rate the following Key Performance Indicators according to your interest on the foreseen results.

4.2.1. KPI 1: Forecasting 7 parameter sets (interventions) that reduce the COVID infection of the simulation (multiscale infection or epidemiologic scenario).

Epidemiological Scenario

4.2.1.1. 3, 5, 4

Multiscale Infection Scenario

4.2.1.2. 5, 5, 4

4.2.2. KPI 2: Use the runtime adaptation of simulation trajectories to improve the outcomes of 5 scenarios or patients of the multiscale infection scenario.

Epidemiological Scenario

4.2.2.1. 4, 3, 3

Multiscale Infection Scenario

4.2.2.2. 5, 3, 4

4.2.3. KPI 3: Calibration of the epidemiological parameter to fit incidence time series.

Epidemiological Scenario

4.2.3.1. 5, 4, 4

Multiscale Infection Scenario

4.2.3.2. NA, NA, 3

4.2.4. KPI 4: Characterizing the space of parameters with 50% fewer simulations.

Epidemiological Scenario

4.2.4.1. 5, 4, 4

Multiscale Infection Scenario

4.2.4.2. 4, 4, 5

4.3. Specific aspects of the epidemiological scenario (1 = Not useful, 5 = Very useful):

4.3.1. Simulating epidemic spread in metapopulations using real mobility and demographic data.

4.3.1.1. 5, 5, 4

4.3.2. Calibrating epidemiological model parameters (e.g., transmission, hospitalization, fatality rates) against real incidence, hospitalization, and mortality data.

4.3.2.1. 5, 5, 4

4.3.3. Exploring and comparing different non-pharmaceutical interventions (NPIs), such as mobility reduction or social distancing policies, and their socio-economic trade-offs.

4.3.3.1. 4, 5, 5

4.3.4. Designing and optimizing vaccination strategies (e.g., prioritization by age, timing, or region) using optimization-via-simulation approaches.

4.3.4.1. 5, 5, 5

4.3.5. Using calibrated simulations to provide short- and medium-term forecasts of epidemic dynamics (cases, hospitalizations, deaths).
4.3.5.1. 5, 5, 4
4.3.6. Integrating simulation outputs and forecasts into decision-making (e.g., selecting interventions under uncertainty, detecting hotspots, early outbreak warnings).
4.3.6.1. 4, 5, 4

4.4. Specific aspects of the multiscale infection scenario

4.4.1. How important is the ability to seamlessly launch Alya–PhysiBoSS simulations on MareNostrum directly from the workflow?
4.4.1.1. 5, 5, 5
4.4.2. How valuable is the possibility to automatically introduce drugs during simulations for exploring therapeutic strategies?
4.4.2.1. 5, 4, 5
4.4.3. Do you consider the Early Time Series Classification (ETSC) early-stopping mechanism a useful feature to save computational resources?
4.4.3.1. 4, 3, 4
4.4.4. As a modeller, what features of the shown workflows appear the most useful and could be incorporated in your work?
4.4.4.1. The coupling of simulation tools with very different time and spatial scales.
4.4.4.2. Intervention of the sims as they are running.
4.4.4.3. GUI, parameter calibration, Alya-PhysiBoSS coupling, automatic drug injection.

5. System Usability Scale questions

The System Usability Scale (SUS) [31] *is a widely used and validated approach to usability assessment. SUS is easy to administer, easy to analyse and quick. The questionnaire results are analysed to obtain an aggregated score for the usability of a product. SUS is a questionnaire of ten items to which participants need to answer using a five-point Likert scale with verbal anchors at the extremes. Following* [32] *the answers are transformed and aggregated to obtain a score between 0-100, where 0 is the minimum usability score possible for the system and scores above 68 are considered above average* [32].

To calculate the SUS score you should follow these steps:

1. *Calculate the single item score contribution which will range between 0 and 4:*
   1.1. *for items 1,3,5,7 and 9 the score contribution is the position marked on the scale by the expert minus 1;*
   1.2. *for items 2,4,6,8 and 10, the contribution is 5 minus the position marked on the scale by the expert;*
2. *Sum the score contributions*

3. *Multiply by 2.5 the sum of the scores (point 2)*

SUS questions (1= Strongly disagree; 5 Strongly Agree):
Totals: 62.5, 52.5, 60. Mean: 58.33

1. I think that I would like to use this system frequently

4, 4, 3

2. I found the system unnecessarily complex

3, 2, 2

3. I thought the system was easy to use

4, 3, 3

4. I think that I would need the support of a technical person to be able to use this system

3, 5, 3

5. I found the various functions in this system were well integrated

4, NA, 4

6. I thought there was too much inconsistency in this system

1, 1, 2

7. I would imagine that most people would learn to use this system very quickly

3, 2, 3

8. I found the system very cumbersome to use

3, NA, 2

9. I felt very confident using the system

3, NA, 3

10. I needed to learn a lot of things before I could get going with this system

3, NA, 3

6. Further comments from expert user

6.1. Expert user 1

I think the system would be complicated to be used by non-technical people without assistance. I would need a high level of understanding of the underlying architecture to build my own infrastructure. However, if building a new architecture is not a requirement, I would have enough confidence to use the workflows presented.

It would be useful to expand the scenario for other virus infections, not only SAR-CoV-2, and also to include upper respiratory parts that are known to have a role in virus infection, like nasal mucosa.

6.2. Expert user 2

Pre-setting up projects in RapidMiner that a user can load up and use.

11 Appendix 3: Expert users' questionnaire of the Maritime Use Case

11.1 Detailed Results of Sea trials

Detection	Incident	Detected	Collision	Safe Zone Intersection	Colreg Error	Rerouting error
0	3874	head-on	0	0	0	0
1	3874	head-on	0	0	0	0
2	3875	crossing	0	0	0	0
3	3875	crossing	0	0	0	0
4	3875	crossing	0	0	0	1
5	3875	crossing	0	0	0	1
6	3875	crossing	0	0	0	1
7	3875	crossing	0	0	0	1
8	3881	overtaking	0	0	0	0
9	3881	overtaking	0	0	0	0
10	3883	crossing	0	0	0	0
11	3884	overtaking	0	0	0	0
12	3884	overtaking	0	0	0	0
13	3884	overtaking	0	0	0	0
16	3889	overtaking	0	0	0	0
17	3889	overtaking	0	0	0	0
18	3889	overtaking	0	0	0	0
19	3892	crossing	0	0	0	0
20	3892	crossing	0	0	0	0
21	3892	crossing	0	0	0	0
22	3892	crossing	0	0	0	0
27	3900	crossing	0	0	0	0
28	3900	crossing	0	0	0	0
29	3900	crossing	0	0	0	0
31	3904	crossing	0	0	0	0
33	3906	crossing	0	0	0	0
34	3906	crossing	0	0	0	0
35	3906	crossing	0	0	0	0
36	3906	crossing	0	0	0	0
37	3910	crossing	0	0	0	0
38	3910	crossing	0	0	0	0

39	3912	overtaking	0	0	0	0
47	3920	crossing	0	0	1	0
48	3920	crossing	0	0	1	0
49	3920	crossing	0	0	1	0
58	3931	overtaking	0	0	0	0
59	20664	overtaking	0	0	0	0
64	20669	crossing	0	0	0	0
65	20669	crossing	0	0	0	0
66	20669	crossing	0	0	0	0
67	20669	crossing	0	0	0	0
68	20669	crossing	0	0	0	0
69	20674	overtaking	0	0	0	0
70	20674	overtaking	0	0	0	0
71	20676	crossing	0	0	0	0
72	20676	crossing	0	0	0	0
73	20678	overtaking	0	0	0	0
74	20678	overtaking	0	0	0	0
75	20680	crossing	0	0	0	0
76	20680	crossing	0	0	0	0
77	20680	crossing	0	0	0	0
78	20680	crossing	0	0	0	0
79	20684	overtaking	0	0	0	0
80	20685	crossing	0	0	0	0
81	20686	overtaking	0	0	0	0
82	20687	crossing	0	0	0	1
83	20687	crossing	0	0	0	0
84	20687	crossing	0	0	0	0
85	20687	crossing	0	0	0	0
86	20687	crossing	0	0	0	0
87	20687	crossing	0	0	0	1
88	20687	crossing	0	0	0	0
89	20694	crossing	0	0	0	1
90	20696	crossing	0	0	0	1
91	20696	crossing	0	0	0	0
92	20697	crossing	0	0	0	1
93	20697	crossing	0	0	0	0
94	20697	crossing	0	0	0	0

95	20697	crossing	0	0	0	1
96	20701	overtaking	0	0	0	0
97	20702	crossing	0	0	0	0
98	20702	crossing	0	0	0	0
99	20702	crossing	0	0	0	0
100	20702	crossing	0	0	0	0
101	20706	crossing	0	0	0	0
102	20707	crossing	0	0	1	0
103	20707	crossing	0	0	1	0
104	20707	crossing	0	0	1	0
105	20707	crossing	0	0	1	0
106	20711	overtaking	0	0	0	0
107	20712	crossing	0	0	0	0
108	20712	crossing	0	0	0	0
109	20714	overtaking	0	0	0	0
110	20715	head-on	0	0	0	0
111	20716	crossing	0	0	0	1
112	20716	crossing	0	0	0	0
113	20716	crossing	0	0	0	0
114	20716	crossing	0	0	0	0
115	20716	crossing	0	0	0	0
116	20716	crossing	0	0	0	0
117	20716	crossing	0	0	0	0
118	20716	crossing	0	0	0	0
119	20716	crossing	0	0	0	0
120	20716	crossing	0	0	0	0
121	20726	crossing	0	0	0	0
122	20726	crossing	0	0	1	1
123	20726	crossing	0	0	1	1
124	20726	crossing	0	0	1	1
125	20730	crossing	0	0	0	0
126	20730	crossing	0	0	0	0
127	20730	crossing	0	0	0	0
128	20730	crossing	0	0	0	1
129	20730	crossing	0	0	0	0
130	20730	crossing	0	0	0	1
131	20730	crossing	0	0	0	1

133	20737	overtaking	0	0	1	1
134	20739	overtaking	0	0	1	1
135	20739	overtaking	0	0	1	1
136	20739	overtaking	0	0	1	1
137	20743	overtaking	0	0	1	1
138	20743	overtaking	0	0	0	0
139	20743	overtaking	0	0	0	0
140	20747	overtaking	0	0	0	0
141	20747	overtaking	0	0	0	0
142	20747	overtaking	0	0	0	0
143	20752	crossing	0	0	0	0
144	20752	crossing	0	0	0	0
145	20752	crossing	0	0	0	0
146	20752	crossing	0	0	0	0
147	20753	crossing	0	0	0	0
148	20753	crossing	0	0	0	0
149	20753	crossing	0	0	0	0
150	20753	crossing	0	0	0	0
151	20753	crossing	0	0	0	0
152	20753	crossing	0	0	0	0
153	20753	crossing	0	0	0	0
154	20753	crossing	0	0	0	0
155	20753	crossing	0	0	0	0
156	20753	crossing	0	0	0	0
157	20753	crossing	0	0	0	0
158	20764	overtaking	0	0	0	0
162	20764	overtaking	0	0	0	0
163	20764	overtaking	0	0	0	0
164	20764	overtaking	0	0	0	0
165	20764	overtaking	0	0	0	0
166	20772	crossing	0	0	1	0
167	20772	overtaking	0	0	0	0
137					16	22
	overtaking	36			88.32%	83.94%
	crossing	98				
	head-on	3				
	TOTAL	137				

11.2 Questionnaire

CREXDATA: Maritime Use Case Evaluation Survey

1.User Background Information

1.1 What is your Job Title?		
1.2 Which sector do you primarily work in?	[ ] Port operations [ ] Shipping/Logistics [ ] Maritime Safety/Security [ ] Environmental monitoring	[ ] Research/Academia [ ] Government/Regulatory Body [ ] Technology Provider [ ] Other :___________________
1.3 Years of experience		
1.4 Background studies (university degree major, etc.)	[ ] University degree [ ] Masters [ ] PhD	
1.5 What are your main job tasks?		
1.6 To whom are you responsible for performing these tasks?		
1.7 Which of the following best describes your role in relation to CrexData's maritime use case?	[ ] End User [ ] Decision Maker [ ] Technical Evaluator [ ] Policy/Regulation Advisor	[ ] Other :___________________

2. Maritime use case - Sea trials evaluation

2.1 Were you present at, or have you seen any footage or received sufficient information about the sea trials for the Collision Avoidance Service conducted in Syros within 2025?	YES	NO
2.2 Have you had the opportunity to view a demonstration—live or recorded—of the Extreme Weather Forecasting Service?	YES	NO

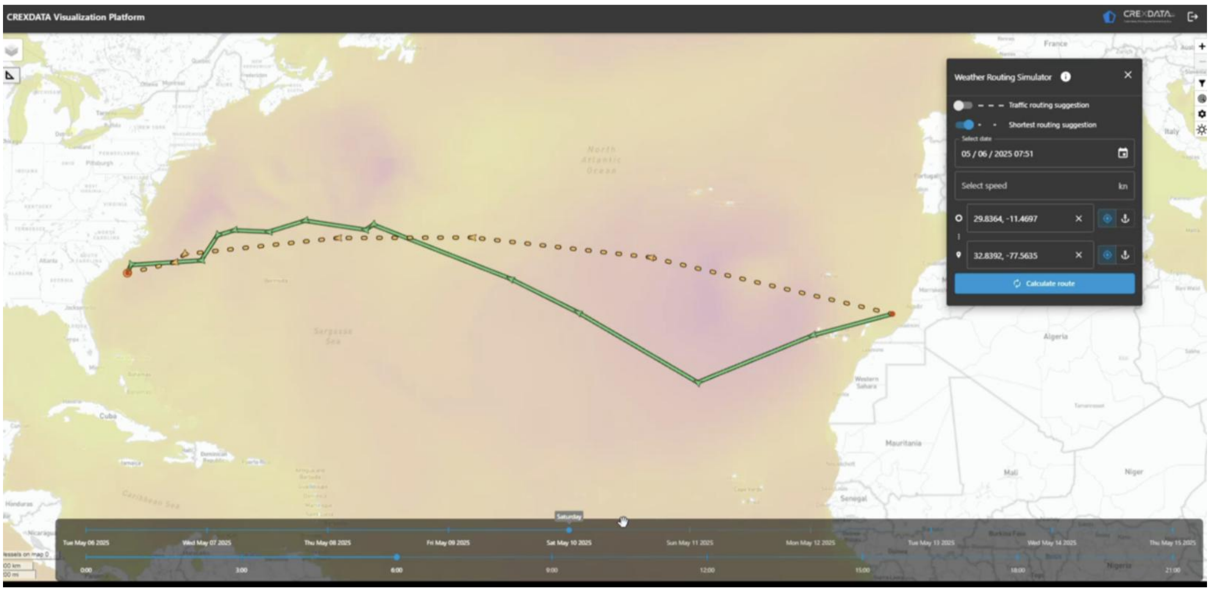
3. Collision avoidance service

COLREG	SCENARIO DEFINITION	CONDITIONS OF SUCCESS
Head on RULE 14	two vessels approach each other from **opposing directions** with relative course difference to be **less than <35 deg**	▪ both vessels turned right ▪ safe distance > 2m ▪ no-contact

Crossing RULE 15	two vessels approach each other from **opposing/crossing directions** with relative course difference to be **more than >35deg**.	▪ Vessel approaching from the left side of the other vessel must give-way ▪ safe distance > 2m ▪ no-contact
Overtaking RULE 13	two vessel move towards the same direction in proximity and the second vessel has higher speed.	▪ Fast vessel passes from either side keeping safe distance ▪ safe distance > 2m ▪ no-contact

QUESTION RATE FOR EACH SCENARIO FROM 1 (STRONGLY DISAGREE) – 5 (STRONGLY AGREE)	**RULE 14 HEAD-ON SCENARIO**					**RULE 15 - CROSSING SITUATION**					**RULE 13 - OVERTAKING SITUATION**				
3.1 DID THE COLLISION AVOIDANCE SERVICE PERFORM AS EXPECTED (I.E., **AVOIDED COLLISIONS EFFECTIVELY**)?	1	2	3	4	**5**	1	2	3	4	**5**	1	2	3	4	**5**
3.2 DID THE COLLISION AVOIDANCE SERVICE CONSISTENTLY MAINTAIN A **SAFE DISTANCE** (GREATER THAN 2 M) WHEN AVOIDING DANGER ZONES?	1	2	3	4	**5**	1	2	3	4	**5**	1	2	3	4	**5**
3.3. DID THE VESSELS CONSISTENTLY DEMONSTRATE **COLREG-COMPLIANT AVOIDANCE BEHAVIOR**	1	2	3	4	**5**	1	2	3	4	**5**	1	2	3	4	**5**
3.4. THE LATENCY OF THE SYSTEM WAS REAL -TIME (SUB-SEC)	1	2	3	4	**5**	1	2	3	4	**5**	1	2	3	4	**5**

4. Hazardous Weather Routing Service



QUESTION RATE FOR EACH SCENARIO 1 (Strongly DISAGREE) – 5(STRONGLY AGREE)	**RULE 14 HEAD-ON SCENARIO**				
4.1 ARE RE-ROUTING SUGGESTIONS NECESSARY?	1	2	3	4	5

4.2 WOULD YOU ACCEPT THE RE-ROUTING SUGGESTIONS?	1	2	3	4	5
4.3. WHOULD YOU ACCEPT THE FULL LENGTH OF THE SUGGESTED ROUTE?	1	2	3	4	5
4.4. THE SERVICE WAS ABLE TO PROVIDE RESPONSES IN A TIMELY MANNER (LESS THAN AN HOUR)?	1	2	3	4	5

5.System Usability Scale Questions

QUESTION	**1 (Strongly DISAGREE) – 5(STRONGLY AGREE)**				
5.1 I think that I would like to use this system frequently	1	2	3	4	5
5.2 I found the system unnecessarily complex	1	2	3	4	5
5.3. I thought the system was easy to use	1	2	3	4	5
5.4. I think that I would need the support of a technical person to be able to use this system	1	2	3	4	5
5.5 I found the various functions in this system were well integrated	1	2	3	4	5
5.6 I thought there was too much inconsistency in this system	1	2	3	4	5
5.7 I would imagine that most people would learn to use this system very quickly	1	2	3	4	5
5.8 I found the system very cumbersome to use	1	2	3	4	5
5.9 I felt very confident using the system	1	2	3	4	5
5.10 I needed to learn a lot of things before I could get going with this system	1	2	3	4	5

6.Comments (Free Text)

Name/ Surname:

Signature: