D5.2 TAPAS Validation report

Deliverable ID:	**D5.2**
Dissemination Level:	**PU**
Project Acronym:	**TAPAS**
Grant:	**892358**
Call:	**H2020-SESAR-2019-2**
Topic:	**SESAR-ER4-01-2019 Digitalisation and Automation principles for ATM**
Consortium Coordinator:	**CRIDA**
Edition date:	**17 June 2022**
Edition:	**00.03.00**
Template Edition:	**02.00.03**

Authoring & Approval

Authors of the document

Name/Beneficiary	Position/Title	Date
Ian CROOK / **ISA Software**	WP5 Leader	14/06/2022
Sandrine MOLTON / **ISA Software**	Project Member	14/06/2022
Natividad VALLE / **CRIDA**	Project Member	14/06/2022
María Florencia LEMA / **CRIDA**	Project Member	14/06/2022

Reviewers internal to the project

Name/Beneficiary	Position/Title	Date
Natividad VALLE / **CRIDA**	Project Member	04/05/2022
María Florencia LEMA / **CRIDA**	Project Member	04/05/2022
Gennady ANDRIENKO / **Fraunhofer**	Project Member	04/05/2022
George VOUROS / **UPRC**	Project Member	04/05/2022
Rubén RODRÍGUEZ / **CRIDA**	Project Coordinator Alternate	17/06/2022

Approved for submission to the SJU By - Representatives of all beneficiaries involved in the project

Name/Beneficiary	Position/Title	Date
Jose Manuel CORDERO / **CRIDA**	Project Coordinator	17/06/2022
Sandrine MOLTON / **ISA**	Project Member	06/05/2022
Gennady ANDRIENKO / **Fraunhofer**	Project Member	06/05/2022
George VOUROS / **UPRC**	Project Member	06/05/2022
Antonio GARCIA / **Boeing**	Project Member	06/05/2022
Hugo SALINAS / **INDRA**	Project Member	06/05/2022

Rejected By - Representatives of beneficiaries involved in the project

Name/Beneficiary	Position/Title	Date
N/A		

Document History

Edition	Date	Status	Author	Justification
00.00.01	31/07/2021	Draft	ISA Software	Draft for review

EUROPEAN PARTNERSHIP

Co-funded by the European Union

00.00.02	02/09/2021	Draft	ISA Software	Pre-release review by partners prior to submission
00.01.00	08/09/2021	Final	ISA Software	Final version for submission to SJU with partner comments included
00.01.01	14/12/2021	Final	ISA Software and CRIDA	Integration of SJU review comments
00.01.02	30/03/2022	Draft	ISA Software and CRIDA	Update to new VALR template from SJU. D5.3 – inclusion of CD&R Validation Experiment results
00.01.01	04/05/2022	Draft	ISA Software and CRIDA	Pre-release review by partners prior to submission
00.02.00	06/05/2022	Final	ISA Software and CRIDA	Final version for submission to SJU
00.03.00	14/06/2022	Final	ISA Software and CRIDA	Update of the document to integrate last SJU review comments

TAPAS

TOWARDS AN AUTOMATED AND EXPLAINABLE ATM SYSTEM

This document is part of a project that has received funding from the SESAR Joint Undertaking under grant agreement No 892358 under European Union's Horizon 2020 research and innovation programme.

Abstract

This document presents the final version of the TAPAS deliverable D5.2 Validation Report. It contains the results of the ATFCM and CD&R experiments carried out under the umbrella of TAPAS project.

Several Human-in-the-Loop (HITL) Real Time Simulation (RTS) were performed for both ATFCM and CD&R use cases involving operational experts to validate eXplainable Artificial Intelligence (XAI) decision support components working in scenarios at different levels of automation (level 1, 2 and 3). Th HITL validation experiments were used to extract conclusions and principles for transparency and explainability when deploying automation based on these types of AI tool. However, as the TAPAS activity remains an exploratory research project with a low TRL, the experiments were not designed to evaluate the performance of the support tools and focused only on the aspects of explainability needed to help operational users comprehend solutions at the different automation levels.

Different simulation sessions were performed for scenarios at each of the automation levels. In the first series of experiments these focused on FMP tasks carried out to support ATM network Demand and Capacity Balancing activities in the pre-tactical planning phase. In the second series of experiments the focus was in supporting the Conflict Detection and Resolution tasks carried out by Air Traffic Controllers in the execution phase. The results obtained from the various simulations are described in this VALR at an exercise level, and then aggregated into the global results according to the different levels of automation.

EUROPEAN PARTNERSHIP

Co-funded by the European Union

Deviations from D5.1 TAPAS VALP, as well as conclusions and recommendations for future activities are also provided.

Table of Contents

List of Tables

List of Figures

1 Executive summary

The TAPAS research aims at using *explainability* and *transparency* components to demonstrate how Artificial Intelligence (AI) based Decision Support Tools (DST) can be used to assist operators in the execution of Air Traffic Flow and Capacity Management (ATFCM) and Air Traffic Control (ATC), Conflict Detection and Resolution (CD&R) tasks in a manner that can be understood by a human operator.

In support of this objective, a set of validation experiments were performed that included prototype DST working at varying levels of automation, ranging from the provision of advisory information to the automated execution of actions identified by the DST. These tools were accompanied by a dedicated set of transparency tools, which provided interactive Visual Analytics (VA) and explanatory information, designed to help the human operator to understand the proposed decisions.

The deployment of AI-based tools is becoming commonplace in many aspects of our daily lives, and Air Traffic Management (ATM) is no exception to this phenomenon. In practice, as the ATM system is becoming increasingly saturated, enhanced tools which employ AI techniques are being considered to help to increase the capacity and resilience of the system through higher levels of automation.

In this scenario, a fundamental change in the automation approach from classical human-machine interfaces (HMI) to potentially richer solutions supported through AI and Machine Learning (ML) techniques is proposed. However, a significant challenge related to AI/ML solutions is the fact that these types of tools tend to be based on complex mathematical and highly recursive, deep searching, pattern matching algorithms to support the 'learning' process. As a result, this can render them difficult to comprehend by human users.

Using '*eXplainable* AI' (XAI) techniques, supported by enhanced Visual Analytics, it is expected that the reasons why certain solutions are being proposed by the DST can be presented in an understandable way to the human operator and that the associated explanations can help build '*trust*' in the new technology as the level of automation increases. We recall that trust in these new AI-based systems is paramount if the decisions being made are going to be widely accepted, and a potential lack of explainability would be detrimental to their future deployment or certification.

As a part of the validation process, subjective qualitative and objective quantitative data has been collected from a series of Human in the Loop (HITL) simulation experiments**.** These have been analysed to assess the TAPAS concept and research goals as expressed in the validation plan, with a strong focus on identifying principles and recommendations relating to transparency needs when using AI solutions.

In the initial scenarios, the focus was on the ATFCM domain with XAI based DST providing automated recommendations at levels of automation ranging from Level 1 to Level 3. A Visual Analytics support tool and a prototype FMP client application were also used to assist the human operators for explainability and transparency for simulation scenarios that were executed using the INnovative Network Operations Validation Environment (INNOVE) [2].

The second set of experiments provided automated AI based support in the Conflict Detection and Resolution activity provided by the ATC Radar/Executive controller with DST also running at the three different levels of automation. The XAI based CD&R automation was used to identify conflict situations, provide recommended actions, and support conformance monitoring to help users solve ATC problems in real-time experiments conducted using the ENAIRE/CRIDA SACTA ATC simulator platform.

2 Introduction

2.1 Purpose of the document

This document provides the Validation Report (VALR) for experiments performed to support explainability and transparency concepts when using XAI automation tools to support Demand-Capacity Management in the ATFCM domain and the detection of aircraft separation conflicts, provision of potential solutions and monitoring of traffic conformance for Radar Controllers during the execution phase (CD&R). It describes the results of validation exercises defined in the Validation Plan [3] and documents how the objectives included in that plan were achieved.

Based on additional post-simulation analysis, a set of relevant conclusions, lessons learned, and recommendations relating to the principles expressed in the TAPAS ATFM Transparency Requirements [4] are provided.

2.2 Intended readership

This document is intended to be used by:

- SJU programme managers.
- TAPAS project members, in particular partners from WP5 dealing with the execution of the validation exercises and the validation report, WP3 related to transparency principles and WP4 for the implementation of the transparent AI/ML components and VA support.
- SESAR2020 and the international research community addressing automation in Air Traffic Management, Artificial Intelligence, Machine Learning and transparency/explainabiity principles.

2.3 Background

The TAPAS research will determine how the use of XAI techniques combined with supporting Visual Analytics can help to deliver better transparency about how decisions have been reached by the AI/ML components supporting automation in ATM. It is expected that these explanations will allow domain experts to better understand why those decisions were made. In turn, it is expected that sufficient trust and confidence can be established in this type of tool to allow AI-based systems to be certified as reliable solutions which can be deployed to support ATFCM and ATC operators in their daily activity.

It should be noted, however, that as TAPAS remains an exploratory research activity executing at a low TRL, *no attempt is made to try to identify 'how' AI/ML based solutions can be verified and/or certified*. TAPAS only focuses on identifying methods by which the process can be *explained* to human operators to help them understand why (and possibly how) decisions have been made and why the proposed actions are appropriate.

TAPAS experiments have been designed in two selected ATM/ATC domains – the Air Traffic Flow and Capacity Management (ATFCM) domain and Air Traffic Control (ATC) domain. These experiments were executed in two distinct sets of validation experiments in June 2021 (ATFCM) and March 2022 (CD&R).

To execute the ATFCM validation experiments, XAI and VA prototypes were integrated in the INNOVE platform - which is a human in the loop simulation and gaming platform, and a prototype FMP client working position was developed to allow operators to perform Demand-Capacity Management activities in real-time.

During the execution of the ATFCM experimental scenarios, operators were required to:

- Monitor and manage air traffic demand against the available capacity for all airspace sectors located in the Madrid ACC
- Identify overload periods (*Hotspots*) where demand exceeds the defined capacity thresholds
- Investigate the characteristics of the traffic that contributes to the overloads
- Determine suitable ATFCM solutions that can be applied to mitigate the identified problems
- Apply and test those solutions and assess the results of the mitigation actions

In the case where the XAI components were used to automatically identify Hotspots in the region and optionally propose or automatically implement solutions, the operators were required to:

- Review and validate the Hotspots that were identified by the XAI
- At automation level 2:
    - Select one or more of the proposed solutions and implement them
- At automation level 3:
    - Review the solutions that had been automatically applied by the platform
- Review the explanations provided and the associated Visual Analytics which help to understand the choices and solutions being proposed
- Report on their level of understanding of the solutions and their confidence that those solutions have been based on reliable reasoning

For the CD&R validation experiments, the XAI and VA prototypes were connected to the ENAIRE/CRIDA SACTA real-time ATC simulator platform, from where they consumed the necessary data to detect conflicts, elaborate solutions and present all this information through visual aids to the air traffic controllers.

SACTA is the system that manages ATC in all enroute, approach, and terminal centres in Spain. A replica of this ATC platform, installed at the CRIDA premises, was used as a human in the loop simulation platform that both manages how aircraft operate in one or more ATC sectors and which provides a highly realistic ATC Controller Working Position (CWP). The CWP provides all the same features as the actual ATC system used to manage the Spanish airspace. Flight trajectories are calculated using the SACTA flight models and control of flights in the region is performed by support staff operating via a pseudo pilot interface. In addition, to emulate fully automated execution of conflict resolution clearances proposed by the AI tool, a 'ghost controller' is included in the platform who can implement the changes being recommended by the DST without the need for the ATC user to intervene.

EUROPEAN PARTNERSHIP

Co-funded by the European Union

Using the CWP and other SACTA tools, users can perform all the functions that are carried out by both the Planner and Executive/Radar Controller for the selected airspace in real-time, including any communication that is required with pseudo aircraft pilots that are also connected to the simulator platform.

During the execution of the CD&R experimental scenarios, operators were required to:

- Monitor and manage air traffic as it flies through one of two upper airspace sectors (Toledo and/or Domingo upper) located in the Madrid ACC
- Acquire and hand-off flights as they arrive into / depart from the managed sector
- Observe the radar picture and flight plans as traffic progresses through the region
- Monitor aircraft-aircraft separation in the controlled sector and the bordering region
- Identify potential separation issues (with the help of the conflict alerting tool provided by the AI component) and prevent collisions between aircraft in flight
- Identify of potential solutions to predicted separation issues, with or without the help of the AI automation tool
- Provide of instructions/clearances to traffic to avoid separation losses
- Monitor the traffic compliance to proposed flight plans and any separation management instructions provided (manually, using proposed solutions and/or automatically)
- Provide of instructions/clearances to allow traffic to recover its original plan following separation management actions towards, or at, the sector exit / transfer point
- Select and execute efficient solutions to traffic separation issues, e.g. through the use of direct to solutions or manoeuvres that are as efficient as possible
- Timely deliver conflict avoidance instructions to ensure the safe and efficient conduct of flight operations in the region
- Select and implement the most suitable solution being recommended by the AI DST (automation level 2)

  Monitor and understand the solutions that have been automatically implemented by the AI DST and recovery of control whenever the automatic XAI system fails partially or completely to resolve conflicts (automation level 3)

2.4 Structure of the document

The document is structured as follows:

- **Section 1 – Executive Summary**

  Provides a short summary of the document.

- **Section 2 (*this section*) – Introduction**

  Describes the purpose of the document, the intended readership, the background, and provides explanations of the acronyms used throughout the document.

- **Section 3 – Context of the validation**

  Presents the context of the validation and a short description of the experiments, validation aspects, objectives, assumptions, etc.

- **Section 4 – Validation results**

  Provides the results and achievements of the exercises.

- **Section 5 – Conclusions and Recommendations**

  Presents the conclusions of the validation exercise and from the analysis of the results.

- **Section 6 – References**

  Provides a list of references.

- **Appendix**

  Appendix A includes a description of the validation platform and the connected components.
  Appendix B provides detailed results from the XAI automation
  Appendix C contains a description of the VA features.
  Appendix D provides information about the real-time ATC simulation platform used to perform the TAPAS CD&R validation exercises
  Appendix E offers an insight into the VA information display provided to support Radar Controllers when performing the CD&R process
  Appendix F contains details on how the XAI algorithms were validated

2.5 Acronyms

Term	Definition
AI	Artificial Intelligence
ANSP	Air Navigation Service Provider
ATC	Air Traffic Control
ATCo	Air Traffic Controller
ATFCM	Air Traffic Flow and Capacity Management
ATM	Air Traffic Management
CD&R	Conflict Detection and Resolution
CWP	Controller Working Position
DCB	Demand and Capacity Balance
DDR2	Demand Data Repository
DST	Decision Support Tool
FMP	Flow Management Position
HITL	Human In The Loop
HMI	Human Machine Interface
INNOVE	INnovative Network Operations Validation Environment
iTEC	Interoperability Through European Collaboration
KPI	Key Performance Indicator
LTM	Local Traffic Manager
ML	Machine Learning
NM	Network Manager
OCVM	Operational Concept Validation Methodology
SESAR	Single European Sky ATM Research
SJU	SESAR Joint Undertaking
TAPAS	Towards an Automated and Explainable ATM System

Term	Definition
TBO	Trajectory Based Operations
TRL	Technology Readiness Level
VA	Visual Analytics
VALP	Validation Plan
VALR	Validation Report
XAI	Explainable AI

Table 1. Acronyms and terminology

3 Context of the Validation

3.1 Explainable AI/ML automation and VA: a summary

3.1.1 eXplainable AI (XAI)

With the advances in computing power that have been seen in the last 5-10 years, the application of AI is becoming commonplace for solutions where automated support is concerned. The ATM domain is no exception to this.

AI techniques rely on high powered statistical methods that are usually deeply recursive, which can make the understanding and verification of these algorithms a challenge. This is particularly true if solutions are to be widely adopted and need to work in safety-critical applications. However, the concept of trust in AI solutions still presents a significant challenge which is magnified when safety is a concern. Therefore, as the number of AI/ML applications increases, so does the need for them to provide additional information to render them trustworthy.

In response to these challenges, and the need for humans to understand how the AI came to a given solution, organisations around the world have set up expert groups to help to elaborate a strategy on explaining AI and to consider the societal and trust elements of the technology.

A remarkable initiative is underway at the US DARPA (Defence Advanced Research Projects Agency), which is progressing in creating an XAI framework addressing a variety of techniques and methods covering the effectiveness-versus-explainability trade space. In Europe, the EUROCAE and SAE WG-114 and G-34 working groups have also been developing a set of guidelines for the development, certification, and deployment of AI-based tools, particularly in safety critical applications. These guidelines all agree that explainability is a major axis for the certification effort for AI-based automation.

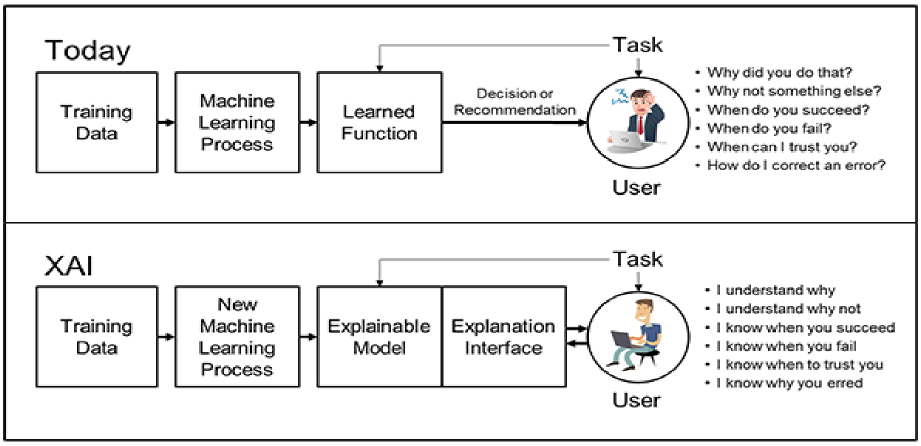


Figure 1. Explainable Artificial Intelligence (XAI)

3.1.2 Visual Analytics (VA) to support transparency

In the context of TAPAS, Visual Analytics is the use of analytical reasoning, supported by interactive visual interfaces to help synthesize complex data and provide a better picture to human operators of why certain solutions have been selected by an automated process.

The VA modules that were developed for the ATFCM use case in TAPAS work in combination with the ML/(X)AI components as an integrated prototype. The ML/(X)AI components develop solutions for demand-capacity imbalance problems while the VA component provides additional support to allow users to explore the problems being addressed and the solutions that have been proposed.

In this integrated solution, the VA component is used to help ATFCM operators to:

- Understand why problems, such as demand-capacity imbalances, have been identified by the automated system.
- Explore the reasons, through visual representation, why the proposed solution is appropriate for the problem
- Help the human to understanding how solutions developed by the ML/AI component were determined

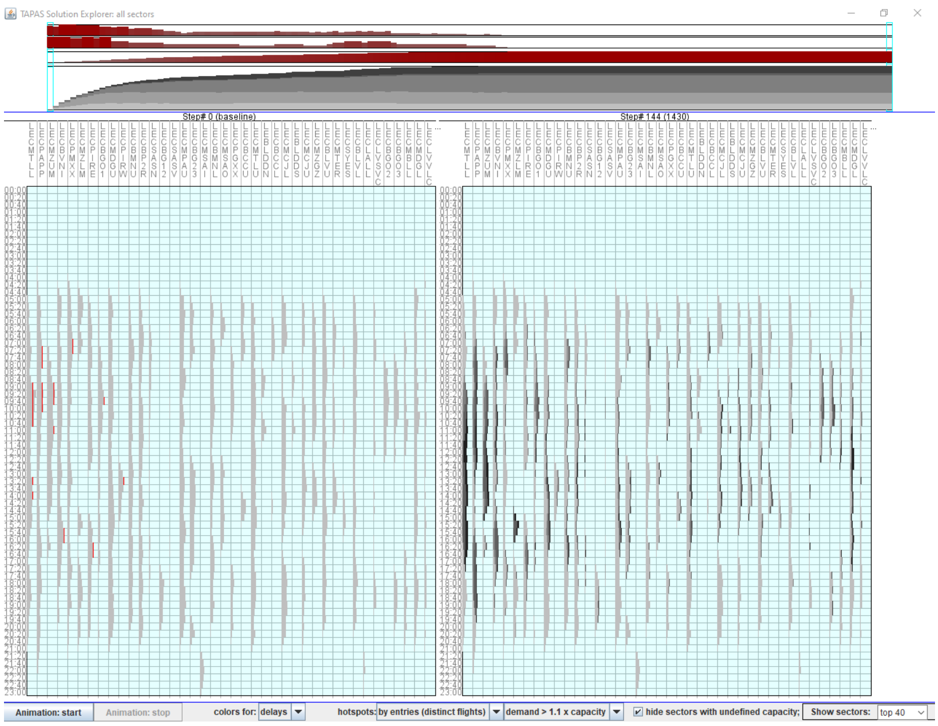


Figure 2. Snapshot of the VA tool for the ATFCM use case

For TAPAS, the ATFCM VA component supports the following analytical tasks:

- Gain an overview of a single scenario.
- Compare two or more scenarios that involve the same set of flights.

- Understand the process of solution development, i.e., see how the ML component modifies the flight plans and resolves the hotspots, with each step of the solution development represented by a scenario based on the modified flight plans.
- Investigate details for selected scenarios, sectors, time intervals, and subsets of the flights.
- Explore solutions at high level of abstraction and low levels of detail to obtain overviews of scenarios, to identify major differences between scenarios, or to track major changes along the process of solution development.
- Explore sectors at lower levels of abstraction and higher level of detail, including information about individual flights

For the CD&R exercises, the VA tool is also integrated with the ML/XAI component and connected to the ATC platform. In this case, the VA component supports the ATCO in the following analytical tasks:

- Awareness of potential conflicts in the following 7-10 minutes between flights inside the sector of focus and in the immediate exit area of that sector.
- Understanding of the conflicts detected, including their severity and main characteristics:
    - Flight IDs and attitude of the aircraft involved in the conflict.
    - Horizontal and vertical separation at the start of the infringement point, closest point (CPA) and at the end point.
    - Time of the potential conflict, at the start, CPA and at the end of conflict.
    - Conflicts associated with the new detected conflict, whenever it comes from a previous detected one or other resolution action proposed.
    - Severity scores of the conflict and Measure of Compliance with the separation minima required.
    - 2D visualisation of the conflicts together with a representation in the Z axis.
- Explore solutions proposed by the XAI algorithm for the specific conflicts. The solutions (direct to a WP, change of HDG/FL/speed) are presented through a tab and arranged in a ranking list from best to worst solution.
- Impact metrics relating to each solution proposed are also included in the listed solutions.
- Awareness of non-conformance monitoring events through a pop-up window showing the aircraft and related condition/s that has/have been violated according to the last available FPL.

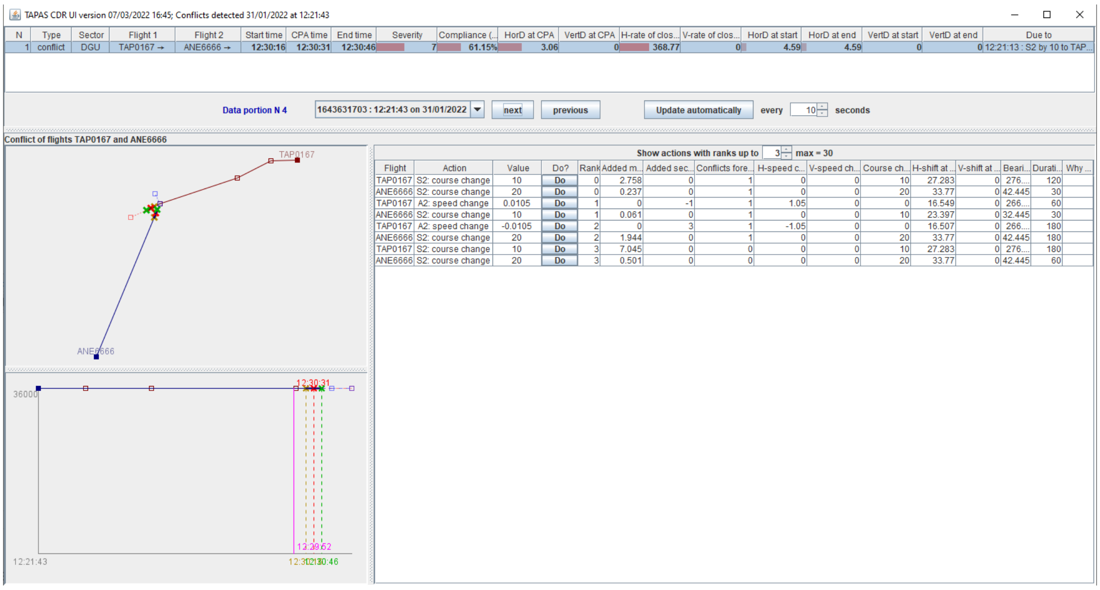


Figure 3. Snapshot of the VA tool for the CD&R use case

3.2 Summary of the Validation Plan

3.2.1 Validation Plan Purpose

The objective of the TAPAS Validation Plan was to set the framework for all the research activities and exercises that have been performed by the members of the TAPAS project in two main domains:

- ATFCM – with a focus on Demand-Capacity Management activities to solve load imbalances in the planning phases of ATM network management, using XAI-based automation with VA to support transparency and explainability.
- ATC– applying XAI and VA techniques to support automated Conflict Detection and Resolution processes during the execution of aircraft flight plans.

The objectives of the validation, along with any assumptions, and a description of the exercises to be performed are provided in the VALP.

The different domains are addressed using two independent sets of validation experiments, held at different times during the research project.

The ATFCM experiments aimed to allow human operators to work interactively in a realistic pre-tactical network capacity planning environment at different levels of automation (levels 1 – 3). The operators were asked to identify and solve overloads during the ATFCM planning phase either manually (level 1), with partial automation (level 2) where possible solutions are identified but the operator maintains the control of whether to apply them or not, and full automation (level 3) where proposed solutions are automatically executed by the tools.

The focus of the experiments was on the regional Flow Management Position - FMP (also known as the Local Traffic Manager - LTM) and, in line with the VALP, the objectives were to understand if situational awareness and understanding of the problems identified or being solved by the AI tool was sufficiently supported through the explainability and visual analytics components.

The TAPAS ATFCM VALP also provides traceability between the validation and the research objectives, by setting the relationship between the TAPAS ATFCM Use Cases [3] and the corresponding Validation Exercises.

The ATC CD&R experiments offer a much more challenging environment within which to deploy the AI-based automation support tools. The aim of the CD&R to allow human operators to work interactively in a realistic ATC environment where the DST provides conflict alerting and, according to the level of automation (levels 1 – 3), recommendations for potential clearances that would solve those issues, or at the highest level, solve those issue automatically without ATCo intervention. The challenge in this role was the short lead time between detection of a problem and the necessary actions needed to resolve those issues - which needed to be performed in a very short time-window leading up to the identified separation problems. With such a short timeframe, beyond which safety could be compromised, the content of the VA 'explainability' elements needed to be very concise and provide information rapidly to the user, in a manner that could be quickly and easily understood. Furthermore, with such a short timeframe between the identification of the conflict and the need to provide suitable clearance(s) little or no time was available to allow the user to *drill down* into the VA support tools to discover more detailed information.

During the exercises, operators were asked to either identify and solve conflicts in the controlled sector, either manually (level 1), with partial automation (level 2) where possible solutions are identified but the operator maintains the choice of which solution(s) to consider/apply or not, and full automation (level 3) where proposed solutions are automatically executed by the tools.

The focus of the experiments was on the Executive/Radar Controller Position, and in line with the VALP, the objectives were once again to understand if situational awareness and understanding of the problems identified or being solved by the AI tool was sufficiently supported through the explainability and visual analytics components but in the CD&R case, with significant time constraints on the overall process.

The TAPAS VALP also provides traceability between the validation and the research objectives, by setting the relationship between the TAPAS ATFCM and CD&R Use Cases [3] and the corresponding Validation Exercises.

3.2.2 Summary of Validation Objectives and success criteria

No changes were made to the validation objectives and success criteria that were expressed in the TAPAS D5.1 Validation Plan Ed 00.01.00 (see VALP section 3.5 [3]).

3.2.3 Validation Assumptions

The following table summarises the validation assumptions that have been made regarding the ATFCM validation exercises (see TAPAS D5.1 Validation Plan section 4.2 [3]):

Identifier	Title	Description	Justification	Impact on Assessment
ASS-TAPAS.ATFCM-001	ATFCM automation tasks allocation	The allocation of tasks between human and machine is done as indicated in VALP section 3.2.3 [ref]	Exercise characteristics	High
ASS-TAPAS.ATFCM-002	Familiarisation of human operators with ATFCM process	The operational actors involved in the execution of the experiments are fully familiar with NM Pre-Tactical Planning and DCB process	Exercise performance	High
ASS-TAPAS.ATFCM-003	Human operators Training	Sufficient training/ briefing has been provided to the human operator in regard to the scenario and available tools to ensure no lack of familiarity has a negative factor in the understanding or acceptance of the proposed solution.	Exercise performance	High
ASS-TAPAS.ATFCM-004	Realistic environment	The operational scenario is modelled in a realistic environment using a validation platform (INNOVE) that supports all of the B2B ATFCM planning services available from NM.	Representative of actual environment	Medium
ASS-TAPAS.ATFCM-005	XAI Training	XAI components have been suitably trained using historic datasets	To guarantee the applicability of XAI algorithm to the	Medium

Identifier	Title	Description	Justification	Impact on Assessment
		for the analysis region	validation scenarios	
ASS-TAPAS.ATFCM-006	Exercise execution Data	Data used for the validation exercise is unseen data for the XAI training process	Using different data to avoid over-fitting of the XAI algorithm	Low
ASS-TAPAS.ATFCM-007	Integration of XAI and validation platform	The XAI component is integrated (loosely) with the validation platform and uses the same scenario data	Exercise performance	High
ASS-TAPAS.ATFCM-008	FMP client prototype	It is assumed the FMP prototype will support human operator similarly to real operational environment	Exercise performance	High
ASS-TAPAS.ATFCM-009	VA prototype	A co-located visual analytics and explanation display included in the FMP Position	Exercise performance	High
ASS-TAPAS.ATFCM-010	VA data consumption and visuals for explainability	VA components consume data from the XAI automation component to provide suitable visual and information and scenario drill down functionality	VA component is responsible of providing the explanations to extract main conclusions from the validation exercises.	High

Table 2. Validation Assumptions overview for the ATFCM use case

The XAI CD&R prototype, together with the appropriate VA techniques, provides features to monitor the current situation, detect possible conflicts, propose resolution measures for those conflicts, monitor traffic compliance to any clearances that are provided and, for automation level 3, implement those solutions automatically.

However, unlike the ATFCM solution, which executed in the D-1 time frame, the CD&R solution must focus on the timely provision of solutions and associated explanatory information due to the short lead time between the identification of a problem and the need to solve it.

To respond to the different requirements in the CD&R validation exercises the following assumptions have therefore been made (see TAPAS D5.1 Validation Plan section 5.2 **[3]**):

Identifier	Title	Description	Justification	Impact on Assessment
ASS-TAPAS-CDR-001	CD&R automation task allocation	The allocation of tasks between human and machine is done as indicated in VALP section 3.2.4	Exercise characteristics	High
ASS-TAPAS-CDR-002	Familiarisation of human operators with ATC process	The operational actors involved in the execution of the experiments are fully familiar with the tasks carried out by ATC Radar controllers and are certified on the airspace sectors used in the scenario	Exercise performance. Relevance and utility of results.	High
ASS-TAPAS-CDR-003	Human operators Training	Sufficient training/ briefing has been provided to the human operator in regard to the scenario and available tools to ensure no lack of familiarity has a negative factor in the understanding or acceptance of the solutions being proposed	Exercise Performance	High
ASS-TAPAS-CDR-004	Realistic Environment	The operational scenarios are modelled in an environment that is both realistic	Representative of the real working environment	Medium

Identifier	Title	Description	Justification	Impact on Assessment
		and has an environment that is similar to the real working situation (using SACTA)		
ASS-TAPAS-CDR-005	XAI Training	XAI components have been suitably trained using historic datasets	To guarantee the applicability of the AI models for the validation scenarios and tasks to be performed	Medium
ASS-TAPAS-CDR-006	Exercise Execution Data	The data used for the validation is unseen data for the AI DST support tools	Use different data to avoid over-fitting of the XAI algorithm	Low
ASS-TAPAS-CDR-007	Integration of the XAI DST and the validation platform	The XAI components are well integrated with the SACTA simulation platform	The XAI components are integrated with the SACTA platform using pub-sub messaging and can share key data easily and in a timely manner	High
ASS-TAPAS-CDR-008	Emulation of full automation	A 'ghost' controller position will be used to emulate the automated execution of clearances provided by the XAI	Users need to experience 'full automation' to evaluate how the associated information helps them understand what has been performed by the XAI	Medium
ASS-TAPAS-CDR-009	Available resolution clearances	The XAI will be limited in scope to include Lateral, Speed, Altitude	The scope of the simulation is to focus on understanding explainability and not to evaluate the	Low

Identifier	Title	Description	Justification	Impact on Assessment
		and Direct-To solutions	efficiency of the XAI CD&R solutions	
ASS-TAPAS-CDR-010	Conflict Alerting	Existing conflict alerting (e.g. STCA) will be disabled in favour of using the conflict alerts provided by the XAI	The exercises should use a consistent alerting mechanism to avoid any confusion on the part of the operator	Low
ASS-TAPAS-CDR-011	Real-Time execution	The scenarios will be executed in synchronised real-time mode with regular track/flight plan updates being provided to the XAI to support the CD&R decision making process	The XAI needs to work on up-to-date information to identify conflict situations in real-time and to provide solution options in a timely and safe manner	Medium
ASS-TAPAS-CDR-012	Late provision of resolution action	It the XAI is unable to identify a problem and recommend solutions in a suitable time to solve the conflict, the operator is able to rationalise the situation and provide suitable solutions themselves	The human remains capable of identifying and solving problems even when automation is running at the highest level to ensure the safe execution of traffic across the entire scenario	Medium
ASS-TAPAS-CDR-013	Controller Working Position	The CWP provided to the users during the validation exercises is as realistic as possible using the SACTA CWP and associated simulator features	The human operator is working in a realistic and up-to-date Radar control environment	Low

Identifier	Title	Description	Justification	Impact on Assessment
ASS-TAPAS-CDR-014	VA Support	A co-located visual analytics and explanation display is provided for use by the controller	Appropriate information is provided to allow the user to understand and implement the proposed solution(s) as needed	Medium
ASS-TAPAS-CDR-015	Interoperable components	The VA and XAI components can exchange key data in a fully integrated and timely manner during the exercise	Key information from the XAI should be provided to the use in a clear, concise, and timely manner to support decision making during the validation exercises	Medium

Table 3: Validation Assumptions overview for the CD&R use case

3.2.4 Validation Exercises List

As previously stated, TAPAS project consists of two main experiments:

- ATFCM validation, the focus of the Interim version of the document
- CD&R validation, added to produce the latest TAPAS deliverable D5.3 Validation Report (this version of the document).

3.2.4.1 AFTCM exercises

Regarding the TAPAS ATFCM validation, three different scenarios were considered based on the levels of automation tackled by the project, namely automation level 1, 2 and 3 (section 4.6 of VALP [3]).

For each scenario, several runs were executed using different traffic dates and studied sectors as summarised below.

Exercise Day / Execution Date	TAPAS Scenario	Automation Level	Traffic Sample Date	Summary of the Exercise
Day 1 14/06/21	Scenario#1	Level 1	22/06/2019	Initial exercise for 2 FMP users working manually using the FMP client application.

				The exercise was primarily aimed at the provision of training and platform familiarisation.
	Scenario#2	Level 2	03/07/2019	Initial exercise for 2 FMP users working with the FMP client application and supported by XAI/VA prototypes to automatically detect Hotspots, provide suggestions for potential solutions and offer explanations supported by the VA application on a co-located display. The exercise was primarily aimed at the provision of training and familiarisation with the XAI and VA components.

Table 4. Summary of exercises - Day 1

Exercise Day / Execution Date	TAPAS Scenario	Automation Level	Traffic Sample Date	Summary of the Exercise
Day 2 16/06/21	Scenario#1	Level 1	05/07/2019	Interim exercise for a single FMP user working manually using the FMP client application. The exercised aimed at improving familiarity with the platform to help gain more confidence in its use to identify solve DCB problem scenarios using Regulation and/or Flight Level Capping measures as well as the gathering of results and FMP user feedback
	Scenario#3	Level 3	04/07/2019	Initial exercise for a single FMP user working with the FMP client application and full automation from the XAI prototype to automatically detect Hotspots and implement solutions without human intervention. The VA application was available on a co-located display and could be

				interrogated as needed by the FMP user. The exercise was primarily aimed at the provision of training and platform / VA-component familiarisation in full automation mode

Table 5. Summary of exercises - Day 2

Exercise Day / Execution Date	TAPAS Scenario	Automation Level	Traffic Sample Date	Summary of the Exercise
Day 3 17/06/21	Scenario#1	Level 1	05/07/2019	Additional exercise for a single FMP user working manually using the FMP client application. The exercised provided an additional opportunity for users to work manually with the platform to continue to gain more confidence in its use to identify and solve DCB problem scenarios using Regulation and/or Flight Level Capping measures as well as the gathering of results and FMP user feedback
	Scenario#2	Level 2	22/06/2019	Exercise for a single FMP user working with the FMP client application and partial automation from the XAI prototype to automatically detect Hotspots, and provide proposals for potential solutions. The VA application was available on a co-located display and could be interrogated as needed by the FMP user. The exercise was a repeat of the initial exercise, performed previously on day 1, but with a higher level of automation, with the aim of acquiring more feedback from the FMP user

Exercise Day / Execution Date	TAPAS Scenario	Automation Level	Traffic Sample Date	Summary of the Exercise
	Scenario#3	Level 3	04/07/2019	Additional exercise for a single FMP user working with the FMP client application and full automation from the XAI prototype to automatically detect Hotspots, and implement solutions without human intervention. The VA application was available on a co-located display and could be interrogated as needed by the FMP user. The exercise was aimed at enhancing FMP user familiarity with the fully automated scenario as well as in the interactive use of the VA-component to help understand the solutions that had been implemented. FMP user feedback was gathered including information on workload, situational awareness, levels of trust and confidence in the solutions.
	Scenario#3	Level 3	14/07/2019	Additional scenario running with full automation from the XAI component and the FMP user interacting with the VA component to investigate and to help understand how and why solutions had been implemented by the DST. The exercise was aimed at further enhancing FMP user familiarity with the fully automated scenario as well as in the interactive use of the VA-component to help understand the solutions that had been implemented. FMP user feedback was gathered including information on workload, situational awareness, levels of

Exercise Day / Execution Date	TAPAS Scenario	Automation Level	Traffic Sample Date	Summary of the Exercise
				trust and confidence in the solutions.

Table 6. Summary of exercises - Day 3

This resulted in a total of eight (8) individual exercises that were performed using the platform. It should be noted that the first two exercises were executed with a heavy focus on training purposes, to ensure that the FMP experts could acquire sufficient knowledge and familiarity with the DSTs and the platform that was used during the different sessions.

Interim sessions were aimed at enhancing the FMP familiarity with the platform and the associated DST running at different levels of automation.

The final day of the experiment included execution of the platform at each of the levels of automation to allow the FMP users to operate in scenarios with increasing levels of automation, and since time permitted, included a second scenario running at full automation to allow the FMP to delve deeper into the VA support component and its capabilities.

During the exercise, facilitators, that is, the validation team involved during the execution of the exercises, provided interactive training and advice as well as performing observations on how the FMP were interacting with the various tools. First of all, the training was delivered intensively during the first day of the executions and through the use of presentations, demonstrations of the tools and by performing different runs with those tools to familiarise the FMP users with the new FMP Client and XAI/VA tool. Additionally, the validation team accompanied the operational staff involved in the tests throughout all the process and executions, supporting them in any doubts or questions they may have on the functioning of the tools and algorithms behind, as well as preparing the needed material for the executions (airspace scenarios and traffic demand, presentations, facilities, etc.), making 'observations over the shoulder' at the same time on the experts comments about the performance of the prototypes and other related topics included in this VALR.

At the end of each exercise a debrief session was held and the FMP completed a questionnaire about the experience.

TAPAS Scenario	ATFCM Scenario #1: DCM with low (Level 1) automation
Scenario description	**Scenario#1** was a series of Human-in-the-Loop, Demand Capacity Balancing gaming exercises running at Automation Level 1 for traffic that was planned across the Iberian peninsula for various 24-hour traffic samples from the June-July 2019 time period. The ATFCM environment data was extracted from the EUROCONTROL Demand Data Repository (DDR2) for the corresponding AIRAC cycle (1907). Traffic data was provided in ALLFT+ V5 format for traffic that was planned to execute in the region at any time for each of the selected validation exercise dates (detailed below).

TAPAS Scenario	ATFCM Scenario #1: DCM with low (Level 1) automation
	The ATFCM scenario was loaded into the INNOVE Network Management gaming platform which emulates the majority of the NM B2B (SWIM) services and simulates the majority of the NM system features, as well as a number of new prototype services in support of on-going SESAR research projects. INNOVE was executed on a dedicated Amazon Web-Service (AWS) instance as a cloud-based service-oriented gaming platform, working interactively in real-time with the human operator and connected client applications to provide a highly realistic Network Management environment in which the FMP users can operate (see Appendix A). To provide a familiar environment, a dedicated FMP client interface was developed through which the FMP were able to interact with the INNOVE platform to perform demand monitoring for any airspace in the Spanish ACC. This interface allows the FMP operator to identify overload issues, create hotspots, interrogate flight counts and lists, and create ATFCM Measures/Regulations to help manage imbalances. Using the same application, the FMP is also able to immediately evaluate the results of any proposed action, as well as to undo previously requested measures/regulations if required. The focus of the various exercises performed for scenario #1 was to: • Provide training to the FMP users in the first rounds of exercise, in particular for the use of the FMP client interface that had been specifically developed by ISA Software to support the exercises • Use the available functions in the FMP client to interrogate traffic demand / load in Madrid ACC sectors • Create SIMULATION snapshots of the current situation to support 'what-if' assessment of potential issues and solutions • Identify demand-capacity imbalances and declare Hotspots in either the OPERATIONAL dataset or in one or more SIMULATIONs • Interrogate traffic lists for the Hotspot periods • Proposed suitable Regulations and Level Capping measures to traffic and 'Cherry-Pick' those flights which should be delayed to solve the problem • Review the impact of the Regulations and Level Capping measures when submitted to the INNOVE platform to

EUROPEAN PARTNERSHIP

Co-funded by the European Union

TAPAS Scenario	ATFCM Scenario #1: DCM with low (Level 1) automation
	validate if the proposed actions had successfully resolved the identified issues As Scenario #1 was executed at Automation Level 1 –no additional automated support was provided, other than the standard functionality available from the INNOVE platform and its B2B services and the support features that were available via the FMP client interface. This allowed the FMP users to focus on the different DCB Measures that could be applied based on their own working experience but did not provide any proposals for potential solutions.
Exercises performed	Three exercises were performed at Scenario #1 (Level 1) automation during the TAPAS ATFCM validation: **Exercise 1:** ***[Day 1]*** Simulation of all traffic planned to operate in the Iberian peninsula on 22nd June 2019 [3728 flights] – manual identification and solving of Hotspots using Regulations and/or Flight Level Capping Measures for Madrid ACC sectors. Primarily focused on training and familiarisation with the validation platform. **Exercise 3:** ***[Day 2]*** Simulation of all traffic planned to operate in the Iberian peninsula on 5th July 2019 [3736 flights] – manual identification and solving of Hotspots using Regulations and/or Flight Level Capping Measures for Madrid ACC sectors. Designed to allow users gain more confidence in the use of the platform to identify and solve DCB problem scenarios using Regulation and/or Flight Level Capping measures and gathering of results and FMP feedback. **Exercise 5:** ***[Day 3]*** Simulation of all traffic planned to operate in the Iberian peninsula on 5th July 2019 [3736 flights] – manual identification and solving of Hotspots using Regulations and/or Flight Level Capping Measures for Madrid ACC sectors. Designed to be part of a sequence of exercises with increasing levels of automation where users could identify and solve DCB problem scenarios manually using Regulation and/or Flight Level Capping measures as well as the gathering of results and FMP feedback.
Expected Achievements	At Automation Level 1, the expected achievements were: • Train the operational experts involved in the simulations on the new FMP Client tool so they become familiar with the new DST available, avoiding a negative impact on the results due to ignorance lack of familiarity with the tools to be used. (in particular for exercises 1 and 3)

TAPAS Scenario	**ATFCM Scenario #1: DCM with low (Level 1) automation**
	• Contribute to a series of consecutive exercises with increasing levels of automation support (exercise 5) • Allow users to become comfortable with the functionality and features available in the validation platform (all exercises) • Obtain observational ('over the shoulder') data on how the users worked with the platform and the available features to identify and solve DCB related issues (all exercises) • Debrief users after completion of each exercise and gather feedback / complete scoring questionnaires (all exercises)
V Phase	FO-AO
Use Cases	ATFCM Use case
Validation Technique	HITL Gaming
KPA Considered	<Human Performance> <Efficiency> <Safety>
Start Date	14/06/2021
End Date	17/06/2021
Validation Coordinator	ENAIRE/CRIDA
Validation Platform	INNOVE / FMP Client
Validation Location	Madrid (ENAIRE/CRIDA Premises)
Status	Complete
Dependencies	None

Table 7. Description of TAPAS Scenario #1

TAPAS Scenario	**ATFCM Scenario #2: DCM with partial (Level 2) automation**
Scenario description	**Scenario#2** was also a series of Human-in-the-Loop, Demand Capacity Balancing gaming exercises but running with partial automation (Level 2) using traffic that was planned across the Iberian peninsula traffic from the June-July 2019 time period.

TAPAS Scenario	ATFCM Scenario #2: DCM with partial (Level 2) automation
	As with Scenario #1, ATFCM environment data was extracted from the EUROCONTROL Demand Data Repository (DDR2) for the corresponding AIRAC cycle (1907). Traffic data was provided in ALLFT+ V5 format for traffic that was planned to execute in the region at any time for each of the selected validation exercise dates (detailed below). The ATFCM scenario was loaded into INNOVE and executed on a dedicated Amazon Web-Service (AWS) instance, working interactively in real-time with the human operator and connected client applications. The FMP client interface was further supported by the XAI component which used the same scenario data to automatically identify DCB issues, define Hotspots and provide proposals for potential solutions using Regulations and/or Flight Level Capping Measures. FMP users were also able to interrogate the VA component, using a dedicated co-located interactive display, to help understand the Hotspots that had been identified and how/why the proposed solutions had been determined by the DST. The focus of the various exercises performed for scenario #2 was to: • Provide training to the FMP users in the first rounds of exercise at level #2, in particular for the use of the VA tools provided in the FMP workstation • Use the available functions in the FMP client and the VA display to interrogate traffic demand / load in Madrid ACC sectors, understand Hotspots that had been identified by the XAI, and to help understand the proposals for mitigation that the tools provided. • Create SIMULATION snapshots of the current situation to support 'what-if' assessment of potential issues and solutions • Evaluate/verify the Hotspots that had been identified by the DST in either the OPERATIONAL dataset or in one or more SIMULATIONs • Interrogate traffic lists for the Hotspot periods • Review the Regulations / Flight Level Capping measures that were being suggested by the automation tools,

TAPAS Scenario	ATFCM Scenario #2: DCM with partial (Level 2) automation
	• Investigate reasons and explanations of the proposed solutions using the interactive features made available in the VA display to hep understand why those solutions had been proposed • Review the impact of the Regulations and Flight Level Capping measures when submitted to the INNOVE platform to validate if the proposed actions had successfully resolved the identified issues Since Scenario #2 was executed at Automation Level 2, additional automated support was provided, along with the standard functionality available from the INNOVE platform, its B2B services and the support features that were available via the FMP client interface. These included the automated identification of Hotspots by the XAI component (which were consolidated by the FMP client to conform to the NM definition of Hotspots – see Appendix A for more details), provision of a series of potential solutions (Regulation/FLC) and interactive display of explanatory data using the dedicated VA component. This allowed the FMP to validate DCB issues identified by the automation and to use the VA features to help to understand the DCB Measures being proposed by the DST and the reasoning that lead to those solutions being proposed.
Exercises performed	Two exercises were performed for Scenario #2 (Level 2) automation during the TAPAS ATFCM validation: **Exercise 2:** ***[Day 1]*** Simulation of all traffic planned to operate in the Iberian peninsula on 3rd July 2019 [3734 flights] –Hotspots were identified by the XAI component and automatically consolidated/added to INNOVE by the FMP client application. Proposals for possible solutions were provided by the XAI (Regulations and/or Flight Level Capping Measures) for Madrid ACC sectors and the user was able to interact with the VA component through its dedicated display features to help understand what was being proposed and why. The exercise, held on the first day, was mainly focused on training and familiarisation with the validation platform and in particular, how the partial solutions being proposed by the connected XAI component were able to be integrated with the scenario being simulated in the INNOVE platform. It was also used to introduce and provide training on the accompanying VA component and the features it provides through

TAPAS Scenario	ATFCM Scenario #2: DCM with partial (Level 2) automation
	its dedicated display in support of transparency, explanations and drill-down understanding of the DST processes. **Exercise 6: *[Day 3]*** was the second in the series of exercises performed at increasing levels of automation that were executed on the final day of the validation. Traffic planned to operate in the Iberian peninsula on 22nd June 2019 [3728 flights] was simulated with the XAI using the same scenario data to identify Hotspots and providing a set of proposals (using Regulations and/or Flight Level Capping Measures) on how to solve them for Madrid ACC sectors. Users were able to consult the FMP client interface to review demand charts created using the data provided by the XAI, as well as the accompanying VA display to investigate why hotspots had been identified, and how the solutions being proposed had been developed by the XAI algorithms. Users were also able to use the VA 'drill down' features to help investigate the proposals in more detail and to develop a better understanding why those solutions were being suggested by the XAI tool.
Expected Achievements	At Automation Level 2, the expected achievements were: • Train the operational experts involved in the simulations on the new FMP Client and VA tool so they become familiar with these two new DST, therefore avoiding a negative impact on the results due to a lack of familiarity with the tools to be used. (In particular for exercise2) • Contribute to a series of consecutive exercises with increasing levels of automation support (exercise 6) • Allow users to become comfortable with the functionality and features available in the validation platform and the DST features being provided by the XAI (all exercises) • Use the features available in the VA display to understand the decisions being proposed by the XAI and why/how those decisions were made (all exercises) • Use the validation tools to understand what the impact of single or combined solutions may be on the overall demand-capacity issues that have been identified (all exercises) • Obtain observational ('over the shoulder') data on how the users worked with the platform and the available features and VA support to identify and solve DCB related issues (all exercises)

TAPAS Scenario	**ATFCM Scenario #2: DCM with partial (Level 2) automation**
	• Debrief users following the completion of each exercise and gather feedback / complete scoring questionnaires (all exercises)
V Phase	FO-AO
Use Cases	ATFCM Use case
Validation Technique	HITL Gaming
KPA Considered	<Human Performance> <Efficiency> <Safety>
Start Date	14/06/2021
End Date	17/06/2021
Validation Coordinator	ENAIRE/CRIDA
Validation Platform	INNOVE / FMP Client / XAI DST / VA Display
Validation Location	Madrid (ENAIRE/CRIDA Premises)
Status	Complete
Dependencies	None

Table 8. Description of TAPAS Scenario #2

TAPAS Scenario	**ATFCM Scenario #3: DCM with full (Level 3) automation**
Scenario description	**Scenario#3** included a set of Demand Capacity Balancing gaming exercises running with full automation (Level 3) for traffic that was planned across the Iberian peninsula traffic from the June-July 2019 time period. As with the other scenarios, ATFCM environment data was extracted from the EUROCONTROL Demand Data Repository (DDR2) for the corresponding AIRAC cycle (1907). Traffic data was provided in ALLFT+ V5 format for traffic that was planned to execute in the region at any time for each of the selected validation exercise dates (detailed below). The ATFCM scenario was loaded into INNOVE and executed on a dedicated Amazon Web-Service (AWS) instance, working interactively in real-time with the human operator and connected client applications.

TAPAS Scenario	ATFCM Scenario #3: DCM with full (Level 3) automation
	The FMP client interface was further supported by the XAI component which used the same scenario data to automatically identify DCB issues, define Hotspots and provide solutions using Regulations and/or Flight Level Capping Measures. These solutions were automatically published by the FMP client for implementation in INNOVE. To help understand those solutions, FMP users were able to interrogate the VA component, using the co-located interactive display, to review the Hotspots that had been identified and any/all of the proposed solutions had been provided by the XAI. In particular the sector and flight-based drill down features available in the VA tool were able to be used to help users understand how and why those solutions had been determined. Impacts of the solutions were able to be seen in traffic demand charts available in the FMP client as well as through advanced features in the VA display. The platform also provided the opportunity (if required) for the user to cancel any of the solutions that had been published to INNOVE. The focus of the various exercises performed for scenario #3 was to: • Provide training to the FMP users in the early runs of exercises at level #3, in the use of the VA tools provided in the FMP workstation to help understand what the XAI solutions included and why. • Use the available functions in the FMP client and the VA display to interrogate traffic demand / load in Madrid ACC sectors, understand Hotspots that had been identified by the XAI, and to help understand the chosen mitigation actions. • Create SIMULATION snapshots of the current situation to support 'what-if' assessment of potential issues and solutions • Evaluate/verify the Hotspots that had been identified by the DST in either the OPERATIONAL dataset or in one or more SIMULATIONs • Evaluate/verify/assess the impact of Regulations/Flight Level Capping Measures that had been automatically implemented from the XAI • Interrogate demand and traffic lists before/after the solutions had been published

TAPAS Scenario	ATFCM Scenario #3: DCM with full (Level 3) automation
	• Review the Regulations / Flight Level Capping measures that were being suggested by the automation tools, • Investigate reasons and explanations of the proposed solutions using the interactive features made available in the VA display to hep understand why those solutions had been proposed • Review the impact of the Regulations and Flight Level Capping measures when submitted to the INNOVE platform to validate if the proposed actions had successfully resolved the identified issues • Cancel XAI actions that were automatically implemented by the platform (optional) • Add new Regulations / Measures to solve any problems that remained after the automated solutions from the XAI had been implemented (optional) Since Scenario #3 was executed at Automation Level 3, additional automated support was provided, along with the standard functionality available from the INNOVE platform. This included the automated identification of Hotspots by the XAI component (which were consolidated by the FMP client to conform to the NM definition of Hotspots – see Appendix A for more details), automatic implementation of solutions (Regulation/FLC) coming from the XAI and interactive display of explanatory data using the dedicated VA component including full sector and flight-based drill down features. FMP were able to validate DCB issues identified by the automation and to use the VA features to help to understand the Measures being proposed as well as the reasoning behind those solutions.
Exercises performed	Three exercises were performed for Scenario #3 (Level 3) automation during the TAPAS ATFCM validation: **Exercise 4:** ***[Day 2]*** Simulation of all traffic planned to operate in the Iberian peninsula on 4th July 2019 [3787 flights] –Hotspots were identified by the XAI component and automatically consolidated/added to INNOVE by the FMP client application. Proposals for possible solutions provided by the XAI (Regulations and/or Flight Level Capping Measures) for Madrid ACC sectors were consumed by the FMP client application and automatically published for implementation in the INNOVE platform. The FMP user was able

TAPAS Scenario	ATFCM Scenario #3: DCM with full (Level 3) automation
	to interact with the VA component through the dedicated display features to help understand what was being proposed and why. The exercise, which was held on the day two of the process, focused on training and familiarisation with the validation platform and how the solutions proposed by the connected XAI component were able to be processed to allow them to be published for implementation in the scenario simulated in INNOVE. Users were also introduced to many of the advanced features available in the VA component and in particular the features it provides to support transparency, explanations, and drill-down understanding of the DST processes – especially given that solutions were automatically published and implemented without operator action. This allowed them to become familiar with the functionality of the VA tool prior to the final Day 3 experiments. **Exercise 7:** ***[Day 3]*** was the final exercise in the series that were performed at increasing levels of automation during the final day of the validation. Traffic planned to operate in the Iberian peninsula on 4th July 2019 [3787 flights] was simulated once again with the XAI using the same scenario data to identify Hotspots and providing solutions (using Regulations and/or Flight Level Capping Measures) for Madrid ACC sectors that were automatically converted and published/implement in INNOVE by the FMP client with no operator intervention. Users were able to consult the FMP client interface to review demand charts created using the data provided by the XAI, as well as the accompanying VA display to investigate why hotspots had been identified, and how the solutions that had been developed by the XAI algorithms and automatically implemented had impacted the scenario. Users used the VA 'drill down' features to help investigate the solutions in more detail and to develop a better understanding why those solutions were being suggested by the XAI tool. **Exercise 8:** ***[Day 3]*** was the final exercise of the validation and was used to execute an additional Level 3 (full) automation exercise with a different traffic scenario. Traffic planned to operate in the Iberian peninsula on 14th July 2019 [3724 flights] was simulated with the XAI using the same scenario data to identify Hotspots and automatically provide solutions to any issues identified. The main objective of the final exercise (#8) was to offer the FMP user an additional opportunity to work with the full automation in

TAPAS Scenario	ATFCM Scenario #3: DCM with full (Level 3) automation
	place and to try to use more of the advanced features available in the VA support tool.
Expected Achievements	At Automation Level 3, the expected achievements were: • Train the operational experts involved in the simulations on the FMP Client, XAI and VA tools so they become familiar with these all the available DST capabilities, therefore avoiding a negative impact on the results due to a lack of familiarity with the tools to be used. (In particular for exercise 4) • Contribute to a series of consecutive exercises with increasing levels of automation support (exercise 7) • Allow users to become comfortable with the functionality and features available in the validation platform and the DST features being provided by the XAI • Use the features available in the VA display to understand the which solutions had been implemented and what they were solving as well as why/how those decisions were made • Obtain observational ('over the shoulder') data on how the users worked with the platform and the available XAI features and VA support to identify and understand how DCB related issues were actually solved by the automation • Debrief users following the completion of each exercise and gather feedback / complete scoring questionnaires
V Phase	FO-AO
Use Cases	ATFCM Use case
Validation Technique	HITL Gaming
KPA Considered	<Human Performance> <Efficiency> <Safety>
Start Date	14/06/2021
End Date	17/06/2021
Validation Coordinator	ENAIRE/CRIDA
Validation Platform	INNOVE / FMP Client / XAI DST / VA Display
Validation Location	Madrid (ENAIRE/CRIDA Premises)

TAPAS Scenario	ATFCM Scenario #3: DCM with full (Level 3) automation
Status	Complete
Dependencies	None

Table 9. Description of TAPAS Scenario #3

3.2.4.2 CD&R exercises

Regarding the TAPAS CD&R validation, three different scenarios were considered based on the levels of automation tackled by the project, namely automation level 1, 2 and 3 (section 5.4 of VALP [3]).

The validation took place over a period of three days, with exercises on Day 1 focused on the provision of training to the users as well as to ensure they became familiar with the various XAI and VA solutions provided as part of the validation process.

This included

1) getting to know and understand the exercise objectives,
2) familiarisation with the TAPAS operational concept in a CD&R context, and
3) training on the support tools provided to support the CD&R process.

Days 2 and 3 were used to run the full validation scenarios and to assess the effect on human performance and understanding when using the XAI and VA tools at different levels of automation. One of the two selected sectors from the Madrid ACC – Toledo Upper (TLU) and Domingo Upper (DGU) was used as the measured sector in each scenario.

The original experimental planning included:

- 3 scenarios executed at Level 1 automation (*one per day*),
- 3 scenarios at Level 2 automation (*one per day*) and,
- 5 exercises running at Level 3 automation with the ATCO working in only a *monitoring mode*
    - 1 exercise on day 1 used for training purposes
    - 2 exercises on day 2, one with and one without unexpected failure of the tool
    - 2 exercises on day 3, one with and one without unexpected failure of the tool

However, following the verification phase of the CD&R activity, which included running scenarios with ATC experts from ENAIRE and getting their feedback on the proposed exercises, it was decided to reduce the number of exercises running in Level 3 ('full automation') for a variety of reasons, including the performance of the XAI tool in 'full' automation mode and time limitations when the ENAIRE ATC experts were not available.

From a technical perspective, the XAI solutions did not always solve all the conflicts that occurred in the scenarios. Therefore, this meant that all the Level 3 scenarios that were executed already included some type of "*system failure/ATCO recovery of control*" component, which eventually resulted in the scenario reverting to automation Level 2.

Furthermore, all the controllers who helped to verify the platform and assess the suitability of the automated solutions thought that Level 3 (full automation) was not a realistic/feasible solution at this stage of the research (for the CD&R activity) – that is scenarios where the ATCO is only acting in a monitoring role whilst the system implements solutions automatically were unrealistic in the scope of the experiments.

Hence due to this view of the automation, and the time constraints on the availability of domain experts, it was decided to reduce number of Level 3 runs from (from 5 to 3) rather than impacting scenarios at the other levels and more focus was placed on Level 2.

The adapted experimental planning that was finally used included:

- 3 scenarios executed at Level 1 automation
    - 1 scenario with 2 ATCo managing the sector and 2 scenarios with a single ATCO, resulting in 4 sets of questionnaire responses,
- 5 scenarios at Level 2 automation
    - 2 scenarios with 2 ATC experts managing the airspace and 3 scenarios with a single ATCo, resulting in 7 questionnaire responses and,
- 2 scenarios executed at Level 3 (*including unexpected degradation of the automation*)
    - both with only a single ATCo monitoring the scenario resulting in 2 questionnaire responses.

Scenarios were then executed using different traffic levels and studied sectors as summarised in the table below.

Exercise Day / Execution Date	Scenario / Sector	Automation Level	Traffic	Summary of the Exercise
Day 1 07/03/2022	TS1.1 - TLU	Level 1	Low density	The TS1.1 scenario provided a baseline reference scenario against which other results could be compared if required. The TS1.1 scenario, executed at the start of the first day was used as a training and familiarisation exercise. In this scenario, 2 ATCo worked with the system providing Radar/Executive controller

				functions for the Toledo Upper sector (TLU) between FL345 and UNL. A low-density traffic sample was used to allow the users to familiarise themselves with the working position and available tools provided by the SACTA simulator platform. Conflict alerts were provided by the XAI DST component which was consuming all traffic and flight plan data on a 30-second cycle and assessing the information received for potential conflict situations. TS1 traffic samples presented low complexity, low traffic demand (average OCC of less than 5 flights in windows of 5 minutes) and a low number of conflicts (2 conflicts per 15 minutes). For these traffic samples night hours were selected (between 2:00 to 3:00) from the 30th of June 2019. Alerts were provided via the connected VA support display which was co-located on a small screen placed next to the main Radar view. This allowed the users to familiarise themselves with the information being provided in relation to any conflict that was detected by the tools. However, all resolution decisions and resulting actions were made and implemented by the human in this scenario without help from the AI tools.

	TS2.1 - DGU	Level 2	Medium to low density	The TS2.1 scenario was executed for the Domingo Upper (DGU) sector with partial automation support being provided by the XAI tools with the 2 ATCo providing support for the Tactical/Radar controller functions during the exercise. TS2.1-DGU on day 1 was primarily used as a training and familiarisation exercise. Traffic in the TS2 level samples presented medium to low complexity, medium traffic demand (with average OCC of more than 5 flights and less than 10 flights in windows of 5 minutes). A low number of conflicts (3 conflicts per 15 minutes) was included in the traffic sample. For these traffic samples morning hours were selected (from 8:00 to 8:30) from the 4th of July 2019. Users received conflict alerts and a set of proposed solutions that had been determined by the XAI tool. Using the VA support, the users could review the actions proposed and decide to apply one or more of them to try to solve the conflict. Actions were based on clearances for one or other of the flights involved and were *prioritised* to offer users an idea of which may be the most effective clearance(s) to attempt.

				Clearances were still given by the users through the ATC management features available in the SACTA CWP (e.g. using the HMI and voice-based instructions) and clearances were implemented by the pseudo pilot on request. Following the issue of the clearance(s) the users were responsible for monitoring traffic to ensure it complied with the instructions given. The monitoring process was also supported by the XAI/VA components which provided visual alerts if traffic was identified to be 'off-track' or non-compliant with the instruction.
Day 2	TS4 - DGU	Level 1	Medium Density	The TS4-DGU scenario was the first of the *measured* validation scenarios to be executed for the CD&R study and focused on Radar control at Level 1 automation in the DGU sector. The airspace was managed by 1 ATCo who provided support for Tactical/Radar Control functions via the SACTA CWP. TS4 level traffic presents very high complexity, medium traffic demand (average OCC of more than 5 flights and less than 10 flights in windows of 5 minutes) and a medium number of conflicts (6 conflicts per 15 minutes). For these traffic samples morning hours were selected (from 8:00 to 8:30) from the 25th of June 2019. Basic levels of automation that are the same as features that

				are available in current day operations were used. The XAI component was used to identify conflicts and to provide alerts which were shown in the associated VA display component. No potential resolution actions were provided, and the user was requested to solve the conflicts that the XAI identified with no additional assistance from the automation components. Conformance monitoring (both against the proposed flight plan and any clearance provided by the Radar controller) was also active in the scenario.
	TS1.2 - TLU	Level 2	Low Density	TS1.2-TLU was executed with 1 ATCo providing the Tactical/Radar control function on the TLU radar position. TS1 traffic samples presented low complexity, low traffic demand (average OCC of less than 5 flights in windows of 5 minutes) and a low number of conflicts (2 conflicts per 15 minutes). For these traffic samples night hours were selected (between 2:00 to 3:00) from the 30th of June 2019. Partial automation support at Level 2 was provided to the user. The automation performed conflict alerting and provided a prioritised list of potential actions for flights that could be used to solve those issues.

				The ATCo users was able to review the conflicts that were reported using both features available in the CWP and information provided by the co-located VA display tool. At level 2, the user was asked to select and apply solutions from the proposed list but the implementation of any clearances chosen had to be performed by the Radar controller using the facilities available in the SACTA CWP HMI and using voice-based instructions to the pseudo pilot. Conformance monitoring (both against the proposed flight plan and any clearance provided by the Radar controller) was also active in the scenario.
	TS2.1 - DGU	Level 3	Medium Density	TS2.1-DGU provided users with a more complex, higher density traffic scenario running with Level 3, full automation. 1 ATCo was responsible for monitoring the DGU sector and intervening if needed. In this scenario, the user was requested to allow the XAI components to perform all the conflict detection, alerting, and to initiate the preferred resolution actions without help from the Radar controller. The human monitored the scenario, looking at why conflicts had been identified, and assessed how and why the proposed solutions were chosen as well as their

				suitability in solving those issues. The user had to maintain a good situational awareness as well as to monitor the conformance of traffic to the automated resolution clearances – with the assistance of the conformance monitoring functionality in the XAI components. In addition, the user was requested to *intervene and recover the situation when a sudden and unexpected failure of the automation occurred*. This included identifying conflicts which were not captured by the automation, or which were captured but not solved. The scenario used TS2 level traffic samples which present low to medium complexity and medium traffic demand (average OCC of more than 5 flights and less than 10 flights in windows of 5 minutes) and a medium number of conflicts (3 conflicts per 15 mins). For these traffic samples morning hours were selected (from 8:00 to 8:30) from the 4th of July 2019
	TS3 - DGU	Level 2**	Medium Density	The TS3 – DGU scenario performed on Day 2 provided users with a more complex, higher density traffic scenario running at level 2, partial automation. The Radar/Executive controller function was provided by 2 ATCo during the simulation exercise.

EUROPEAN PARTNERSHIP

Co-funded by the European Union

				Users were requested to allow the XAI components to perform all the conflict detection and alerting, as well as to provide a set of resolution actions that could be used to solve conflicts using one or more of the recommended clearances. The human role was then both monitor traffic in the scenario using the CWP, looking at why conflicts had been identified, and assessing the proposed solutions to determine which were suitable to help solve those issues. Users were not required to use the highest priority solution(s) and were free to choose other alternatives being suggested, or to implement their own solution(s). Tasks also included identifying conflicts which were not captured by the automation, or which were captured but no suitable solution was offered. To support this task the users had to maintain a good situational awareness as well as to monitor the conformance of traffic any resolution clearances that they had selected – with the assistance of the conformance monitoring functionality in the XAI components. Traffic for the TL3 scenario presents medium complexity, medium traffic demand (average OCC of more than 5 flights and less than 10 flights in windows of 5 minutes) and a medium number of conflicts (4 conflicts per 15 minutes). For

				these traffic samples morning hours were selected (from 8:00 to 8:30) from the 25th of June 2019.
Day 3	TS4 – DGU	Level 1	Medium Density	The TS4-DGU scenario performed on Day 3 of the validation was a repeat of the same scenario from Day 2 with a different Radar controller operator performing the exercise. 1 ATCo was responsible for providing Radar/Executive controller functions during the execution of the scenario with limited automation support at Level 1.
	TS1.2 - TLU	Level 2	Low Density	The TS1.2-TLU scenario performed on Day 3 of the validation was a repeat of the same scenario from Day 2 with a different Radar controller operator performing the exercise. 1 ATCo provided support for the Radar/Executive controller function with the assistance of the XAI-based automation and VA information support tool. Suggestions offered by the automation could be selected by the ATCo then appropriate clearances delivered to traffic with automation running at Level 2.
	TS3 - DGU	Level 3	Medium Density	The TS3-DGU scenario performed on Day 3 of the validation was a repeat of the same scenario from Day 2 with a different Radar controller operator performing the exercise.

				1 ATCo was responsible for monitoring conflicts identified by the automation and the solutions that were being automatically applied to try to solve them. The user was also requested to indicate when conflicts were not correctly identified and/or if solutions were inappropriate or did not solve the problems.
	TS2.1 - TLU	Level 2	Medium to Low density	The TS2.1-TLU scenario performed on Day 3 of the validation was a repeat of the same scenario from Day 2 with a different Radar controller operator performing the exercise.

Table 10. Summary of exercises - Day XX

At the end of each exercise a debrief session was held and the Controller completed a questionnaire about the experience.

The three main types of scenarios, based on the level of automation made available to the users in each set, are described in the following tables:

Scenario	**ATC CD&R Scenario #1: low (Level 1) automation**
Scenario description	**CD&R Scenario#1** provided a series of Human-in-the-Loop simulation exercises running at Automation Level 1 using the SACTA real-time simulation platform with the XAI and VA support tools connected. Simulations were performed with users working the Radar/Executive control position for either the Toledo or Domingo Upper airspace sectors (TLU/DGU) and depending on the exercise they were supported by a planning controller. As the scenarios were automation level 1, the connected tools only provided conflict alert information. No proposals were made for actions that might help solve those conflicts, and the Radar controller had to determine their own resolutions and provide clearances to traffic using the standard ATC features available in the SACTA system (i.e. using the CWP HMI features and voice instructions). Human operators on the SACTA pseudo-pilot position implemented any instructions provided to traffic by the user.

Scenario	ATC CD&R Scenario #1: low (Level 1) automation
	The XAI also provided some conformance monitoring for flight tracks against both the original flight plans and when following ATC clearances. Low density traffic data was used for the baseline reference scenario (TS1.1-TLU) which also formed part of the training sessions that were provided to users at the start of the validation exercises. Medium to High density traffic scenarios were used to assess the effect of the XAI/VA support tools on the performance of the operators at level 1 (TS4-DGU). The focus of the various exercises performed for scenario #1 was to: • Provide training to the ATC users in the first rounds of exercise in the use of the SACTA CWP and other ATC features along-side the VA support display that was co-located with the CWP. • Use this level 1 scenario as a baseline of the current operating methods and as a means of comparison between other scenarios with higher levels of automation (level 2/3). • Use the available functions in the SACTA CWP to manage traffic and provide ATC services for the selected Madrid ACC sectors • Help the users to become familiar with the additional information being provided by the XAI and VA display for conflict alerts and conformance monitoring • Let the users perform ATC services for the target sectors and based on the conflict alerts provided determine and implement suitable resolution actions using SACTA ATC features (HMI/Voice) • Monitor the compliance of traffic to flight plans and any ATC clearances provided for separation management • Provide additional instructions to traffic to recover the original flight plan and coordinate the hand-off to downstream sectors in accordance with the original flightplan / any localised agreements • Review the impact of the clearances provided on flight efficiency, safety, and ATC workload
Exercises performed	Three validation exercises were performed at level 1 automation: TS1.1-TLU –baseline scenario for training/familiarisation (Day 1) TS4-DGU – measured scenario with high density traffic (Day 2)

Scenario	ATC CD&R Scenario #1: low (Level 1) automation
	TS4-DGU – measured scenario with high density traffic (Day 3)
Expected Achievements	At automation level 1, the expected achievements were: • Train the operational experts involved in the simulations on the SACTA platform and ATC support features (CWP / Voice etc), the XAI DST and the co-located VA support display so that they become familiar with the available tools. This would avoid a negative impact on the results due to ignorance or a lack of familiarity with the tools being used. • Contribute to a series of consecutive exercises with increasing levels of automation support • Allow users to become comfortable with the functionality and features available in the validation platform (all exercises) • Obtain observational ('over the shoulder') data on how the users worked with the platform and the available features to identify and solve DCB related issues (all exercises) • Debrief users after completion of each exercise and gather feedback / complete scoring questionnaires (all exercises)
V Phase	FO-AO
Use Cases	Air Traffic Control – CD&R use case
Validation Technique	ATC Real-time HITL Simulation
KPA Considered	<Human Performance> <Efficiency> <Safety>
Start Date	07/03/2022
End Date	09/03/2022
Validation Coordinator	ENAIRE/CRIDA
Validation Platform	SACTA ATC Real-Time Simulator platform
Validation Location	Madrid (ENAIRE/CRIDA Premises)
Status	Complete
Dependencies	None

Table 11. Description of TAPAS CD&R Scenario #01

EUROPEAN PARTNERSHIP

Co-funded by the European Union

Scenario	ATC CD&R Scenario #2: medium (Level 2) automation
Scenario description	**CD&R Scenario#2** provided a series of Human-in-the-Loop simulation exercises running at Automation Level 2 using the SACTA real-time simulation platform with the XAI and VA support tools connected. Simulations were performed with users working the Radar/Executive control position for either the Toledo or Domingo Upper airspace sectors and depending on the exercise they were supported by a planning controller. At automation level 2, the connected tools provided both conflict alert information and a set of proposals for actions that might help solve those conflicts. These were listed on a flight-by-flight basis and included an assessed priority which helped to inform the user about which actions the XAI believed were the most suitable. The Radar controller had to review the proposed actions and decide which of them to use to try to solve the conflict. No constraints were placed on the users regarding which, or how many, solutions to try, nor whether the highest priority ones should be selected or not. This choice was left to each user for each conflict situation that was identified. Once one or more solutions had been chosen, the Radar controller had to provide clearances to traffic using the standard ATC features available in the SACTA system (i.e. using the CWP HMI features and voice instructions). As previously, human operators on the SACTA pseudo-pilot position implemented any instructions provided to traffic by the user. The XAI also provided some conformance monitoring for flight tracks against both the original flight plans and when following ATC clearances. Low to medium density traffic data was used for the training scenario (TS2.1-DGU) that was run on the first day of the validation to allow to users to become familiar with the information provided by the VA display support tool based on the XAI decisions. Similar traffic scenarios were used to assess the effect of the XAI/VA support tools on the performance of the operators at level 2 on the second and third days (TS1.2-TLU). The focus of the various exercises performed for scenario #2 was to:

Scenario	ATC CD&R Scenario #2: medium (Level 2) automation
	• Provide training to the ATC users on the first day of exercises in the use of the SACTA CWP and other ATC features and advanced features of the VA support display that was co-located with the CWP. • Use the available functions in the SACTA CWP to manage traffic and provide ATC services for the selected Madrid ACC sectors • Allow users to interrogate the additional information being provided by the XAI and VA display for conflict alerts and proposed solutions • Let the users preform ATC services for the target sectors and based on the conflict alerts provided determine and to select and implement resolution actions from the list being proposed by the XAI • Transmit appropriate clearances to traffic using the SACTA ATC features (HMI/Voice) based on the actions being recommended by the XAI • Verify that the selected action(s) were able to successfully resolve the identified conflict(s) • Monitor the compliance of traffic to flight plans and any ATC clearances provided for separation management • Provide additional instructions to traffic to recover the original flight plan and coordinate the hand-off to downstream sectors in accordance with the original flightplan / any localised agreements • Review the impact of the clearances provided on flight efficiency, safety, and ATC workload
Exercises performed	Five validation exercises were performed at level 2 automation: TS2.1-DGU – baseline scenario for training/familiarisation (Day 1) TS1.2-TLU – measured scenario with low-med density traffic (Day 2) TS3-DGU – measured scenario with medium complexity traffic (Day 2) TS1.2-TLU – measured scenario with low-med density traffic (Day 3) TS2.1-TLU – measured scenario with medium complexity traffic (Day 3)
Expected Achievements	At automation level 2, the expected achievements were:

Scenario	ATC CD&R Scenario #2: medium (Level 2) automation
	• Train the operational experts involved in the simulations on the SACTA platform and ATC support features (CWP / Voice etc), the XAI DST and the co-located VA support display so that they become familiar with the available tools. This would avoid a negative impact on the results due to ignorance or a lack of familiarity with the tools being used. • Contribute to a series of consecutive exercises with increasing levels of automation support • Allow users to become comfortable with the functionality and features available in the validation platform (all exercises) • Let users select and implement one or more of the solutions being proposed by the XAI and verify that they successfully resolved the identified conflict(s) • Obtain observational ('over the shoulder') data on how the users worked with the platform and the available features to identify and solve DCB related issues (all exercises) • Debrief users after completion of each exercise and gather feedback / complete scoring questionnaires (all exercises)
V Phase	FO-AO
Use Cases	Air Traffic Control – CD&R use case
Validation Technique	ATC Real-time HITL Simulation
KPA Considered	<Human Performance> <Efficiency> <Safety>
Start Date	07/03/2022
End Date	09/03/2022
Validation Coordinator	ENAIRE/CRIDA
Validation Platform	SACTA ATC Real-Time Simulator platform
Validation Location	Madrid (ENAIRE/CRIDA Premises)
Status	Complete
Dependencies	None

Table 12. Description of TAPAS CD&R Scenario #02

Scenario	ATC CD&R Scenario #3: full (Level 3) automation
Scenario description	**CD&R Scenario#1** provided a series of Human-in-the-Loop simulation exercises running at Automation Level 3 using the SACTA real-time simulation platform with the XAI and VA support tools connected. Simulations were performed with users working the Radar/Executive control position for either the Toledo or Domingo Upper airspace sectors (TLU/DGU) and depending on the exercise they were supported by a planning controller. As the scenarios were automation level 3, the connected tools provided conflict alert information and automatically issued clearances to traffic that should resolve those conflicts. The Radar controllers were requested to monitor the situation without intervening on any reported conflicts and allowing the XAI to provide a solution. Users were also tasked with understanding the actions that had been taken and to verify that those actions were suitable for each of the identified conflict situations. Information related to the conflict and associated actions was provided to the Radar controller via the co-located VA display. To emulate the automated execution of resolution actions, a 'ghost' controller provided the clearances to the SACTA system and human operators on the SACTA pseudo-pilot position then implemented those instructions without communication to the ATC radar controllers – to emulate that the automation had performed the task. The XAI also provided some conformance monitoring for flight tracks against both the original flight plans and when following ATC clearances which allowed the Radar controller to review how flights were responding to the automated solutions. Medium complexity, medium density traffic scenarios were used for both the training and the measured scenarios in the level 3 full automation experiments (TS3-TLU for training and TS3-DGU for the two measured exercises). Users were also required to monitor the situation to respond to any unexpected *failure* of the automation tool (e.g. conflicts not detected or missing/inappropriate resolution actions). On recognising the failure, users were required to recover the situation and intervene manually to resolve any on-going conflicts in a safe and timely manner.

Scenario	ATC CD&R Scenario #3: full (Level 3) automation
	This allowed the team to further evaluate the level of situational awareness that the Radar controllers were able to maintain in situations where automated solutions were being executed. The focus of the various exercises performed for scenario #3 was to: • Provide training to the ATC users in the first rounds of exercise in the use of the SACTA CWP and other ATC features along-side the VA support display that was co-located with the CWP. • Allow the users to become familiar with the additional information being provided by the XAI and VA display when identified conflicts were being automatically resolved by the XAI tool and to support conformance monitoring of those solutions. • Let the users evaluate whether the resolution actions being performed were suitable for the conflict situations that occurred and resolved them successfully • Monitor the compliance of traffic to flight plans and the clearances provided by the automation for separation management • Provide additional instructions to traffic to recover the original flight plan and coordinate the hand-off to downstream sectors in accordance with the original flightplan / any localised agreements following an automated resolution action • Determine whether users could maintain a sufficient degree of situational awareness to be able to recover in situations when the automation fails unexpectedly • Review the impact of the automated clearances provided by the XAI on flight efficiency, safety, and ATC workload
Exercises performed	Two validation exercises were performed at level 3 automation: TS2.1-DGU – measured scenario with low-med density traffic and unexpected failure of automation/situation recovery by ATC (Day 2) TS3-DGU – measured scenario with med density traffic and unexpected failure of automation/situation recovery by ATC (Day 3)
Expected Achievements	At automation level 3, the expected achievements were: • Train the operational experts involved in the simulations on the SACTA ATC support tools and the XAI and VA tools operating in fully autonomous mode. This allows users to

Scenario	ATC CD&R Scenario #3: full (Level 3) automation
	become familiar with these the available DST capabilities and how conflicts are solved by the tools. This helps to avoid a negative impact on the results due to a lack of familiarity with the tools to be used. • Contribute to a series of consecutive exercises with increasing levels of automation support • Allow users to become comfortable with the functionality and features available in the validation platform and the DST features being provided by the XAI • Use the features available in the VA display to understand the solutions that were implemented, what they were solving and why/how those decisions were made • Obtain observational ('over the shoulder') data on how the users worked with the platform and the available XAI features and VA support to identify and understand how DCB related issues were actually solved by the automation • Debrief users following the completion of each exercise and gather feedback / complete scoring questionnaires
V Phase	FO-AO
Use Cases	Air Traffic Control – CD&R use case
Validation Technique	ATC Real-time HITL Simulation
KPA Considered	<Human Performance> <Efficiency> <Safety>
Start Date	07/03/2022
End Date	09/03/2022
Validation Coordinator	ENAIRE/CRIDA
Validation Platform	SACTA ATC Real-Time Simulator platform
Validation Location	Madrid (ENAIRE/CRIDA Premises)
Status	Complete
Dependencies	None

Table 13. Description of TAPAS CD&R Scenario #03

3.3 Deviations

3.3.1 Deviations with respect to the Validation Plan

The following deviations with respect to TAPAS D5.1 Validation Plan were identified:

- During the ATFCM exercise execution, due to time and resources limitations not all the opened sectors of Madrid ACC were analysed. Only some of the most interesting in terms of number of hotspots were finally studied by the FMP expert.

- For ATFCM automation level 1, the developed simplified FMP Client did not have a system integrating algorithms like CASA to automatically allocate delays to flights once a regulation is created by the user. This led to a slightly lower level of automation than is currently available in the live operational system (which uses CASA to automatically assign delays to traffic for a user requested Regulations). However, this should have no impact on the results since TAPAS aims to extract principles and requirements on explainability not to validate the tools developed to support said aim, and since the operators were still able to solve some of hotspots manually.

- For ATFCM automation level 3, non-nominal situations were tested. However, none of the runs performed for this automation level were used to ask the user to try to solve the remaining hotspots once the XAI implemented its solution. Instead, they focused their work into understanding the solutions and problems detected using the available tools.

- For the CD&R exercise execution, the original workstation design was to provide a co-located display for information related to the conflicts identified by the AI support tool and depending on the automation level, solutions proposed to help solve them. However, as this required that the ATCo had to *turn away from the standard CWP to look at the display* and since the display was relatively small in comparison to the CWP, it became evident that operators were reluctant to use the view when issues needed to be solved rapidly. To address this issue, the VA information was integrated into the CWP display itself during some of the later exercises and this gave the operator much more opportunity to use the information provided.

- In response to limited availability of ATC experts to perform the experiments combined with some technical difficulties and a somewhat limited ability for the Level 3 ('full') automation to identify and successfully solve conflicts that was observed by controllers during the verification of the platform, the number of Level 3 exercises was reduced to 2 instead of the planned 5, with Level 2 (partial automation) exercises being increased to 5 instead of 3.

The reported deviations have not altered the meaningfulness of the exercises and the corresponding results, since, as previously stated, TAPAS main objective is to define principles and criteria for explainability. The deviations identified above have no major impact in this aim.

EUROPEAN PARTNERSHIP

Co-funded by the European Union

4 TAPAS Validation Results

4.1 Summary of ATFCM Validation Results

The table below summarises the validation results against the stated objectives that were obtained from the ATFCM validation scenarios:

TAPAS ATFCM Validation Objective	Sub-Focus	Success Criterion	Validation Results	Status
OBJ 1: Identify principles for Transparency of AI-based solutions	1.1 Determine how much additional information is needed at automation levels 2 & 3 to ensure that the human operator is able to make informed decisions to help solve ATM problems.	VA and explanatory support information is clear and understandable and the tools are able to provide the required information at the right time.	Users were able to access and consult the VA display and associated information very easily and in an interactive manner. Information was provided in a timely fashion via the VA display and helped users to maintain their Situational Awareness. Explanations provided via the VA component were generally clear, but some issues to understand solutions being proposed were highlighted.	POK
	1.2 Identify when support information is required, what level of detail is needed and how should it be provided in a timely manner.	Key data that can be easily understood by the human has been identified that supports transparency needs and is provided in the required time frame and at an appropriate frequency	During the initial training scenarios held on day 1 (level 2 partial automation) and day 2 (level 3 full automation) and for the final day 3 exercises, users reported that the ***effort to scan both the FMP and VA displays was low***. However, ***understanding the XAI solutions was considered poor***, particularly at level 2 (partial automation).	POK

Co-funded by the European Union

TAPAS ATFCM Validation Objective	Sub-Focus	Success Criterion	Validation Results	Status
			Effort required to gather and interpret information was initially considered high but improved as the users became familiar with the VA display and its available features.	
	1.3 Evaluate areas where the levels of transparency may need to be improved.	Information that is unavailable but could help during the use of the proposed XAI has been identified and catalogued for future analysis.	General comments from users suggested that ***some aspects of the system were very intuitive*** and both tools used provide all the information they could need. However, ***the way the information was present in some cases were somewhat obscure could be improved.*** Additionally since the methodology applied by the XAI (solve all of the problems at once) was quite different to how FMP work today (solve problems one by one) – it was identified that the ***provision of aggregated information describing the overall effect of the solutions*** would be advantageous in future – this need has been captured and documented as unavailable information at that moment (the impact of the solution was included but not in aggregated way) and will be taken into consideration as an input for the other TAPAS use case of CD&R.	OK
	1.4 Propose suitable methods by which the level of understanding and trust in the AI automation can be measured	Questionnaires, 'over-the-shoulder' observation and debriefing analysis metrics have been identified to support the necessary measures.	***Over the shoulder observation*** and the use of ***debriefing*** and ***questionnaires*** at the end of each exercise ***proved to be very useful***, in particular to capture how the scores tended to improve as users became more familiar with the platform. In general, levels of trust could be seen to improve as users became more familiar with the tools. This was reflected in	OK

TAPAS ATFCM Validation Objective	Sub-Focus	Success Criterion	Validation Results	Status
			the scores which became visibly ‘more green’ in the later exercises.	
OBJ2: Develop prototype XAI/VA methods for ATM use cases to address transparency at various levels of automation	2.1 Produce customised VA views to support transparency and explanatory information to the human operator at different levels of automation.	VA display tools are able to consume data provided by the XAI component to support interactive drill down views for the human operator	The VA ***display was able to consume data from the XAI*** component and to present a series of visual aids to help users to interrogate information related to Hotspots and solutions being identified by the tool. Detailed capabilities that are ***Sector based and Traffic based allowed users to drill down and interactively*** obtain more detailed information from the VA tools whenever this was required.	OK
	2.2 Assess how the VA methods can help enhance operator understanding and trust in AI-based automation	Elements provided in the VA provide clear visual evidence related to the actions being performed by the XAI tools	The ***elements provided in the VA support tool were designed to help users understand the issues*** that had been identified, as well as the reasons why the proposed solutions were selected Users felt that solutions were sometimes slightly difficult to understand. However additional comments from the users suggest that the difference in the paradigm that the XAI uses to create solutions (solve all in one go) contributed to these issues as it was unfamiliar to users who normally solve issues one by one.	POK
	2.3 Evaluate the effectiveness of the transparency solutions being deployed	Human operators are able to use the visualisation to interrogate the on-going scenario and solutions being considered	Using the ***VA display and features available combined with the FMP client interface, users were able to easily interrogate*** the scenario and investigate solutions being proposed.	OK

Co-funded by the European Union

TAPAS ATFCM Validation Objective	Sub-Focus	Success Criterion	Validation Results	Status
			Additional explanations about how, and why, solutions were selected were provided through the VA sector and traffic exploration features.	
	2.4 Determine the different needs for transparency at different automation levels	Human operators classify the information being provided and confirm that it is sufficient to explain the decisions being made	At automation level 1 (manual) the users were able to easily classify information being provided (sector load/demand charts, overloads, flight lists etc.). At higher levels of automation, available information could be easily found but users identified missing information which would have helped to improve the explanations being provided	POK
	2.5 Evaluate the level of understanding and situational awareness of the human as the automation proposes / implements solutions	Human operators are able to describe what the automation is doing and why solutions have been proposed	In general, using the available features in the FMP client and VA display, ***users were able to maintain Situational Awareness at all times***, and remained neutral about things being under control. At higher levels of automation, users reported some difficulty to focus on the problem 'as a whole' rather than single problems and were sometimes 'surprised' by unexpected events. This was probably due to the differences of the new operating method used (solve all in one go).	OK
	2.6 Verify that the human can successfully take over and recover control of the situation if	The human was able to either take over and complete the current task when automation failed,	Not tested during the validation exercise	**Not tested**

TAPAS ATFCM Validation Objective	Sub-Focus	Success Criterion	Validation Results	Status
	the automation fails for any reason			
	2.7 Ensure that the human is able to identify and resolve any remaining issues at the end of the XAI process, if present.	The human operator was able to identify and to complete any remaining issues that were not successfully solved at the end of the process	Not tested during the validation exercise	**Not tested**
	2.8 Demonstrate how transparency can promote operational and social acceptance of 'black-box' AI solutions	The operator confirms that the solutions provided by the XAI were fit for purpose	Users responded either neutrally or slightly negatively when asked whether the system was useful, reliable, accurate and understandable when automation was in place. Information provided to help explain Hotspot identification was reported as being good, but explanations related to solutions were reported as being a little difficult to use in some cases. Scores related to overall confidence in the solutions remained relatively neutral or slightly negative. However, in discussions during the exercises and the debrief sessions, users seemed to be willing to accept that if the solutions that were implemented by the automation were reasonable across all airspace users, then acceptance and trust would be able to be achieved.	**POK**

Co-funded by the European Union

TAPAS ATFCM Validation Objective	Sub-Focus	Success Criterion	Validation Results	Status
	2.9 Assess shortfalls and areas where transparency can be improved in future solutions	Operational experts identify areas where information was insufficient to support understanding	The ***operational experts were able to review the Hotspots that the XAI identified with relative ease, together with the solutions that were included in the FMP tool and their explanations in the VA display***. However sometimes the information was difficult to understand, some explanation features were no auto explanatory and more focused on how the XAI algorithm worked. In this aspect, users identified that providing information in an 'operational context' would help improve transparency.	POK
	2.10 Identify opportunities for additional training	Additional training or processes to enhance the ability for the XAI/VA to assist the human in understanding the process at different automation levels has been identified by the team	Following the completion of the various exercises, the analysis team performed a general review of the design, execution and results of the validation scenarios and results. Key points were presented to the TAPAS project partners to provide feedback on Explainability, Tooling and Lessons Learned.	OK

Table 14. Summary of ATFCM Validation Exercises Result

4.2 Summary of CD&R Validation Results

The table below summarises the validation results against the stated objectives that were obtained from the CD&R validation scenarios:

TAPAS CD&R Validation Objective	Sub-Focus	Success Criterion	Validation Results	Status
OBJ 1: Identify principles for Transparency of AI-based solutions	1.1 Determine how much additional information is needed at automation levels 2 & 3 to ensure that the human operators can make informed decisions to help solve conflicts identified by the system at various levels of automation.	VA and explanatory support information that is clear and understandable is provided in a short timeframe and the tools provide the required information to allow the user to rapidly understand the situation being managed and context of the proposed solution.	As regular and certified users of the Spanish ATC system, which the SACTA simulator platform replicates, users were highly familiar with the tools and features that the CWP provided. For this reason, it was difficult for them to move away from using those features when analysing predicted conflict situations and their possible solutions during the TAPAS CD&R exercises. Nevertheless, when instructed to use the available information provided in the co-located VA display, users were able to quickly assimilate the information that was provided for a conflict and use the detailed information to review the solutions being proposed. Some features of the display were considered less useful than others. For example, Users indicated that the graphical display of the conflict trajectories provided at the side of the textual information in the VA display was less useful than capabilities already available in the SACTA CWP. When observing the users work at levels 2 and 3, it was clear that they preferred to use existing ATC/traffic monitoring features over that view. However, other information that was provided via the co-located display, related to conflict alerts and the proposed actions, was considered very useful and allowed the users	OK

TAPAS CD&R Validation Objective	Sub-Focus	Success Criterion	Validation Results	Status
			to quickly understand the conflict and traffic involved, as well as the solutions that were proposed, even if they did not always agree with the priorities given by the XAI tool. In cases where the users disagreed with the clearances being proposed, or those implemented automatically at level 3, they still tended to accept the solutions if they solved the conflicts – even though the solutions differed from those that they would have applied themselves. However, if the proposed solution only partially solved, or failed to solve the issue, users questioned why those actions had been proposed. In such cases, there was no further information available to explain why those actions had been selected. In conclusion, most users indicated that due to the often very short lead times for conflicts to be identified, solved and instructions given to traffic, offering more information than was already provided would not necessarily have changed the understanding that they could usually acquire in a very short time due to their own experience and expertise in the domain. Therefore, users generally agreed that the level of information provided was sufficient for their needs in the CD&R use case. Users indicated the need to integrate the VA display information into the SACTA CWP HMI. Having the display located on a smaller screen that was next to the main CWP display required the Radar controller to change from	

TAPAS CD&R Validation Objective	Sub-Focus	Success Criterion	Validation Results	Status
			viewing the Radar screen to look at information on the co-located display which was a distraction and could lead to loss of awareness about the evolving traffic conditions in the sector, especially when traffic loads are high and complex. To further evaluate this effect, during some of the later exercises some modifications were included in the CWP to allow the VA-display information to be seen in pop-up dialogues within the main CWP HMI. When these changes were included in the scenario it was very clear that the controller made more use of the information that was being provided than when it was located on the adjacent display.	
	1.2 Identify when support information is required, what level of detail is needed and how should it be provided in a timely manner.	Key data that can be easily understood by the human has been identified that supports transparency needs and is provided in the required time frame and at an appropriate frequency. Additional information providing more detailed information that can help explain more complex situations and the decisions that were made is available for consultation by the user in an 'on-demand' mode if required	Integration of the XAI and VA information display components was carried out using a publish-subscribe Message Queue solution (RabbitMQ). However, the refresh rates at which key flight plan and track data was shared with the XAI/VA was set at 30-seconds for the experiments. This setting sometimes resulted in delays occurring between a conflict being identified and the related information being provided to the users. Additionally, users needed to wait on some occasions for the information to be updated with the set of proposed resolution action, while on other occasions, the information being displayed was modified or over-written as new or changing conflict situations manifested which caused users to be a little disoriented from time to time.	OK

TAPAS CD&R Validation Objective	Sub-Focus	Success Criterion	Validation Results	Status
			The consequence of these technical issues did have a noticeable impact on the users at times, and they were sometimes waiting for information that was expected (based on their own experience/expertise) when it was not already provided. From the perspective of validation objective 1.2, which was looking at both the content and delivery time for key support information, these technical issues provided a good insight into what controllers would need in an eventual operational deployment of such tools. Information was found to be useful and suitably presented in a way that allowed users to quickly comprehend the situation and understand any actions that were being proposed/implemented. However, the delays in presenting information (usually due to the selected data sharing frequency of 30 seconds) did result in some negative effects due to both unexpected delay in the provision of that data as well as sudden updates that caused the user to be a little disoriented at times.	
	1.3 . Evaluate areas where the levels of transparency may need to be improved	Information that is unavailable but could help during the use of the proposed XAI has been identified and catalogued for future analysis.	ATC were generally satisfied with information that was made available by the XAI/VA components. Information provided was easy to use and complemented the expertise of ATCO as well as existing SACTA tool suite.	OK
	1.4 Propose suitable methods by which the level of understanding	Questionnaires, 'over-the-shoulder' observation and	***Over the shoulder observation*** and the use of ***debriefing*** and ***questionnaires*** at the end of each exercise ***proved to***	OK

TAPAS CD&R Validation Objective	Sub-Focus	Success Criterion	Validation Results	Status
	and trust in the AI automation can be measured	debriefing analysis metrics have been identified to support the necessary measures.	***be very useful***, and in general how the scores tended to improve as users became more familiar with the platform.	
OBJ2: Develop prototype XAI/VA methods for ATM use cases to address transparency at various levels of automation	2.1 Produce customised VA views to support transparency and explanatory information to the human operator at different levels of automation.	VA display tools consumed data provided by the XAI component to support interactive views for the human operator in a timely and concise manner.	In practice, due to the short timeframe between conflicts being identified and the need to implement a suitable and safe clearance to one or more flights, the usefulness of the co-located view was limited. Over the shoulder observation and discussions held during the Level 2 exercises confirmed that the ATCo tended to use the existing functions and features within the SACTA CWP (features that they are highly familiar with already) over any new features that were provided by the VA support display. Furthermore, as the amount and type of information that was able to be provided was somewhat limited, by design to keep it as concise as possible, in-depth 'explanatory' information was not available from the VA tool. Furthermore, users indicated that little time is available to use information of this type in a CD&R situation since actions need to be performed and verified rapidly, before moving on to the next task. Other issues were identified in discussions with the users that related to the set-up of the working position. Since the VA support display was located next to the main CWP and was only provided on a small laptop screen, consultation of information on that display resulted in the user losing focus on the main CWP and on-going traffic	OK

TAPAS CD&R Validation Objective	Sub-Focus	Success Criterion	Validation Results	Status
			situation, which in turn could risk a loss of situational awareness or worse. Similarly, the update frequency (30 sec) and lack of clear indication about when data was last refreshed could lead to users being unclear whether the information was up to date (or remaining from a previous situation) and in some cases new data appearing on the display could lead to other information being hidden or potentially lost. A clear improvement in how the information was provided and used by the ATCo was visible when the CWP interface was enhanced to include an additional pop-up window for the VA information. In this case it was clearly observable that the ATCo made greater use of the information than when it was co-located on the smaller laptop screen. Most, if not all ATCo also indicated that they did not really require the additional graphical view of conflict situations that accompanied conflict details in the VA display. In general, the existing tools in the CWP HMI, combined with the expertise and experience of the ATCo were considered sufficient to understand all of the conflict situations rapidly without the need to consult the additional graphical view, and '*measuring*' tools in the CWP HMI that the ATCo were highly experienced in using were seen as the 'go-to' function when review conflict situations and potential clearances.	
	2.2 Assess how the VA methods can help enhance operator	Elements provided in the VA provide clear visual evidence	Users had mixed opinions when responding to question about visual evidence for conflict resolution proposals.	POK

TAPAS CD&R Validation Objective	Sub-Focus	Success Criterion	Validation Results	Status
	understanding and trust in AI-based automation	related to the actions being performed by the XAI tools	Low confidence in proposed actions from the automation which did not match how ATCO would have solved the issue themselves influenced some of the answers. Questions that were directly related to the information provided about the proposed solution, rather than its quality were well reviewed.	
	2.3 Evaluate the effectiveness of the transparency solutions being deployed	Human operators are able to use the visualisation to interrogate the on-going scenario and solutions being considered	User were able to easily access the information provided in the VA support to understand solutions being proposed and using features in the CWP they could measure the impact/applicability easily.	OK
	2.4 Determine the different needs for transparency at different automation levels	Human operators classify the information being provided and confirm that it is sufficient to explain the decisions being made. Optional detailed views are able to support more complex situations and can provide additional detailed understanding in an acceptable timeframe	CD&R users indicated that information provide was sufficient for their needs and no additional data would be necessary to support transparency. In CD&R little/no time is available for drill down actions	OK
	2.5 Evaluate the level of understanding and situational awareness of the human as the automation proposes / implements solutions	Human operators are able to describe what the automation is doing and why solutions have been proposed.	Many responses were in agreement or neutral on many questions relating to understanding what the automation was doing/proposing. However, when trying to understand why some solutions were proposed users were split 40:60 between strongly agreeing and strongly disagreeing suggesting more work is required on this topic.	POK

TAPAS CD&R Validation Objective	Sub-Focus	Success Criterion	Validation Results	Status
	2.6 Verify that the human can successfully take over and recover control of the situation if the automation fails for any reason	The human was able to either take over and complete the current task when automation failed,	Technical issues with the XAI automation resulted in the operator needing to take over control when running at Level 3. However, while ATCO could intervene in most cases, it was not always able to be carried out in a suitable time.	POK
	2.7 Ensure that the human is able to identify and resolve any remaining issues at the end of the XAI process, if present.	The human operator was able to identify and to complete any remaining issues that were not successfully solved at the end of the process	For automation level 3 scenarios the XAI only detected some (but not all) of the conflicts. Others that were identified were solved in a less than efficient manner or using 'open loop' manoeuvres (e.g. a heading change) without additional instructions to recover the original plan. Hence the ATCO was required to identify those missing conflicts, enhance some of the solutions and/or include additional clearances to recover the flight plan following a resolution action.	POK
	2.8 Demonstrate how transparency can promote operational and social acceptance of 'black-box' AI solutions	The operator confirms that the solutions provided by the XAI were fit for purpose	ATCO were able to understand the solutions being proposed using the available information but on some occasions those solutions were not considered suitable to solve the problem. Scoring suggests that ATCO confidence in the solutions remained low and this impacted some of the operational acceptance of the automation proposals.	POK
	2.9 Assess shortfalls and areas where transparency can be improved in future solutions	Operational experts identify areas where information was insufficient to support understanding	In the CD&R scenarios ATC indicated that all required information was already provided. Some issues were reported regarding conformance monitoring, but these were solved once users became familiar with how this feature was supported by the tool.	OK

TAPAS CD&R Validation Objective	Sub-Focus	Success Criterion	Validation Results	Status
			ATC suggested that conformance monitoring in primarily a human task.	
	2.10 Identify opportunities for additional training	Additional training or processes to enhance the ability for the XAI/VA to assist the human in understanding the process at different automation levels has been identified by the team	Users responded that the system was easy to use and understand with little or no assistance from technical support personnel. Users also indicated that little or no additional training was considered necessary	OK

Table 15. Summary of CD&R Validation Exercises Result

4.3 Detailed analysis of SESAR Solution Validation Results per Validation Objective

4.3.1 Analysis approach

The various high-level objectives identified in the TAPAS validation plan and summarised in the previous section were analysed during the execution of the different ATFCM and CD&R scenarios, and detailed questionnaires were completed by the participants to support analysis.

Results for each of the operational scenarios have been categorised by validation objective and sub-focus topics for each operational domain as described below. These are then compared/contrasted for each of the high-level objectives/sub focus topics in the different validation domains.

4.3.2 TAPAS VAL_OBJ_1 – Principles for Transparency

Objective 1 of the TAPAS activity aimed to *identify principles for transparency of decision support solutions that have been implemented using AI and ML techniques*.

The two scenarios focused on different ATM functions, one in the ATFCM pre-tactical planning domain and the other for the provision or air navigation services by the ANSP for ATC Conflict Detection and Resolution (CD&R) activities. For each domain, scenarios were executed that considered the AI-based DST working at different levels of automation:

Level 1 with limited automation (the current working environment),

Level 2 with partial automated support that allowed the user to review proposed actions and select one or more of them for implementation and,

Level 3, where the system automatically publishes and implements the solutions without any user action.

Using low automation scenarios (level 1) the users were able to obtain a fundamental understanding of the issues relating to the identification of ATFCM demand-capacity issues (Hotspots) and ATC CD&R separation issues (Conflicts). At level 1, the integrated systems used to support the validation exercises supported the investigation of the traffic involved as well as the manual development of suitable solutions to help resolve the problems. In this way, the users were able to become familiar with the validation platform and support tools: For AFTCM - the INNOVE simulator and the FMP client application. For the CD&R experiments, the SACTA platform, and its CWP/ATC support features.

At higher levels of automation, XAI decision support tools combined with visual analytic information displays were used to help operators understand the ATFCM/CD&R problems being identified and solutions that were proposed or automatically implemented.

To measure whether the principles for transparency have been achieved, three different methods were applied including:

- 'Over the shoulder' observation and discussion during the execution of each validation exercise
- Debriefing sessions held immediately after each exercise

- Completion of scoring questionnaires for each exercise.

The TAPAS validation OBJ 1 is made up of four sub-criteria whose assessment results are summarised in the following sections. Results for each of the validation domains (ATFCM then CD&R) are provided for each of the sub-criteria below. Note that when evaluating these objectives, as for the other validation objectives considered in the TAPAS validation, all success criteria involved are considered equally and no weight is applied for their consideration.

On this basis, from the ATFCM perspective the overall OBJ-1 objective result is considered *partially achieved* – since two out of four success criteria were totally achieved but the other two were only partially achieved,

From the CD&R perspective, the overall OBJ-1 objective is considered as having been *fully achieved* for Level 2 since all the sub criteria were considered as OK.

Results for each of the four sub-criteria for each ATM domain are detailed below:

4.3.2.1 ATFCM Scenario – Objective 1.1 assessment results

Sub-Focus	Success Criterion	Success Criterion Verdict
1.1 Determine how much additional information is needed at automation levels 2 & 3 to ensure that the human operator is able to make informed decisions to help solve ATM problems.	VA and explanatory support information is clear and understandable and the tools are able to provide the required information at the right time.	POK

Table 16. TAPAS ATFCM OBJ1.1 assessment result

The verdict of this first success criterion was based on the feedback received from the FMP during the execution of the ATFCM validation experiments.

The figures below present the answers received in the questionnaires on situational awareness and explanations that the tools provided.

As described previously, during the three-day exercise execution there were two different FMP experts involved in the tests. Each one of them participated during two different days, but due to time limitations not all of them performed the same number of runs, therefore the number of answers may vary according to the automation level tackled due to this reason.

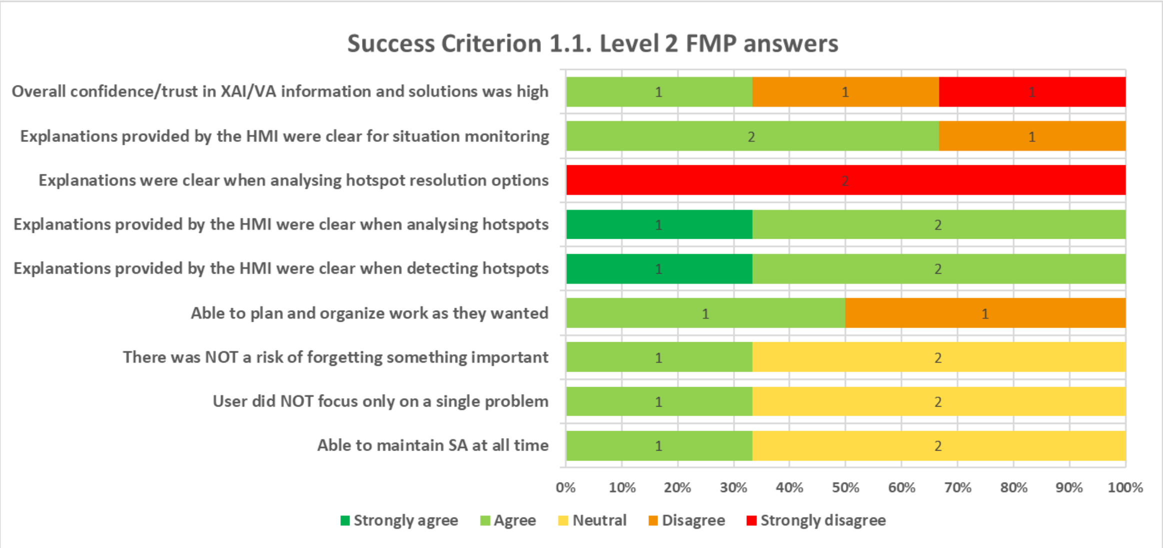


Figure 4. Success Criterion 1.1. FMP answers. Automation level 2

It must be noted that level 2 answers also include results from Day 1, whose runs were focused on training purposes. Answers provided differed, especially related to the last four questions shown on the graph related to situational awareness. For these questions, on the first day of validation the general answer was neutral, but once the knowledge on the tool improved in the last run, the FMP experts agreed they were able to maintain the situational awareness.

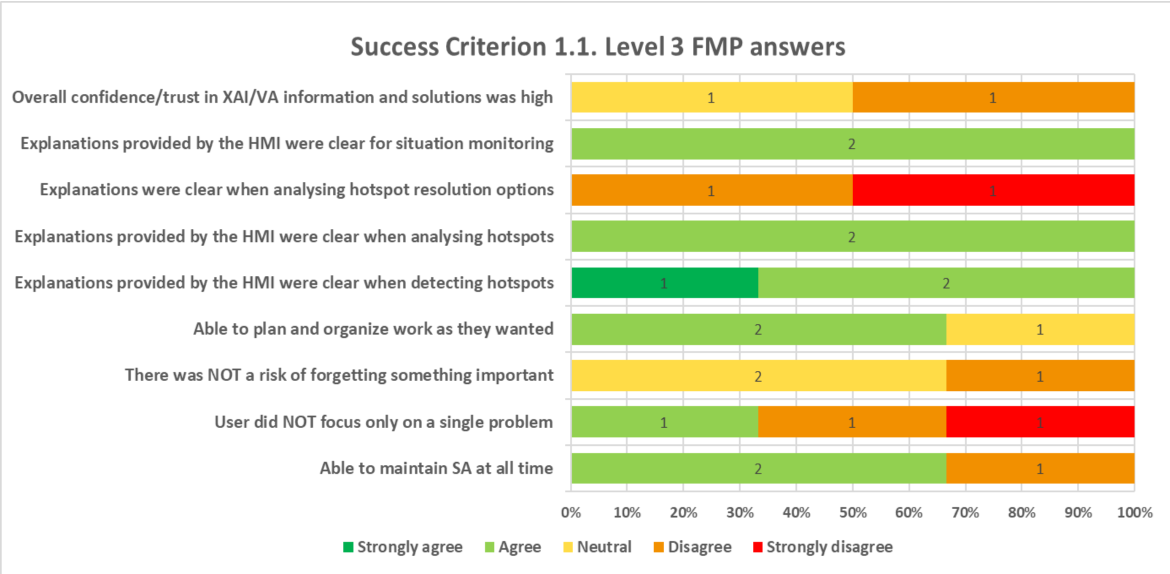


Figure 5. Success Criterion 1.1 FMP answers. Automation level 3

4.3.2.2 CD&R Scenario – Objective 1.1 assessment results

Sub-Focus	Success Criterion	Success Criterion Verdict
1.1 Determine how much additional information is needed at automation levels 2 & 3 to ensure that the human operators can make informed decisions to help solve conflicts identified by the system at various levels of automation.	VA and explanatory support information that is clear and understandable is provided in a short timeframe and the tools provide the required information to allow the user to rapidly understand the situation being managed and context of the proposed solution.	OK

Table 17. TAPAS CD&R OBJ1.1 assessment result

In the CD&R case, success criterion for Objective 1.1 were based on the feedback received from the expert users that worked the Radar/Executive position during the execution of the real-time ATC simulations for upper sectors in the Madrid ACC during the ATC validation experiments carried out using the SACTA real-time simulation platform. These experiments were performed at automation levels 2 and 3 with integrated XAI automation components providing conflict alerts, resolution proposals and providing conformance monitoring services. Key information relating to the conflicts, associated resolution actions and conformance were shared through the co-located VA information display.

The figures below present the answers received in the questionnaires on situational awareness and explanations that the tools provided.

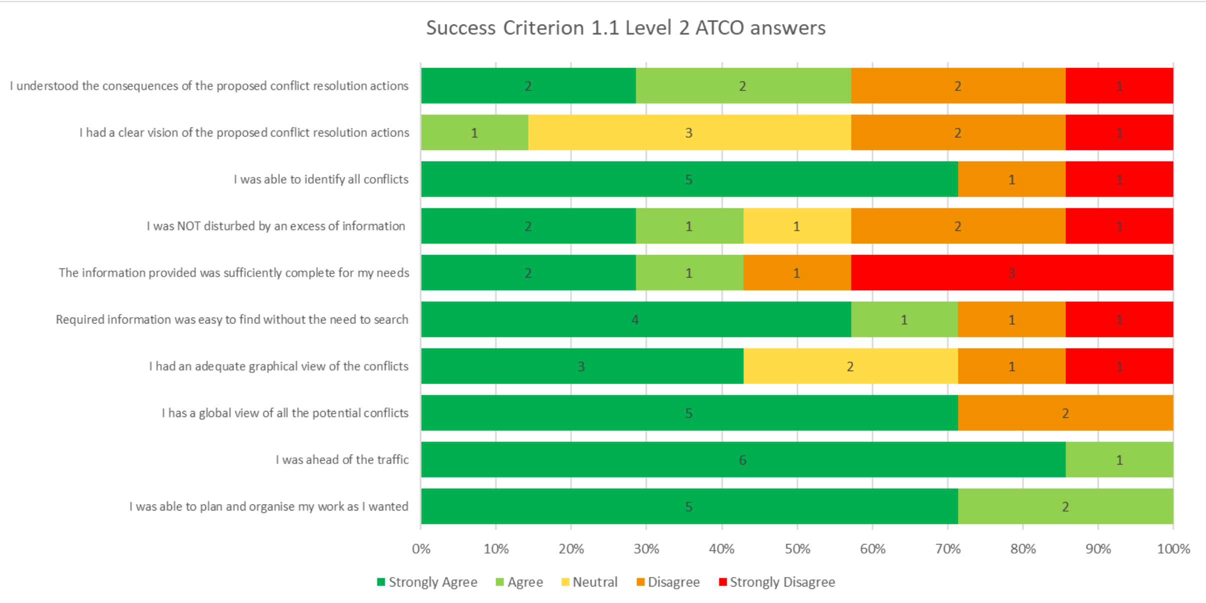


Figure 6: Success Criterion 1.1 ATCO answers. Automation level 2

Responses confirmed that information was readily available, easy to access, and that it was clear and able to be understood in a timely manner. As seen in the selection of questions shown for Level 2 above, a high number of answers to questions relating to Objective 1 sub criterion 1.1 *agreed* or *strongly agreed* with the statements.

Users considered that the information provided a clear vision of the conflict resolution actions being proposed, they were able to easily identify all the conflicts and were not disturbed or overloaded by too much information. Moreover, users indicated that they had a good global overview and were ahead of the traffic and fully capable of planning and organising the work that they needed to do.

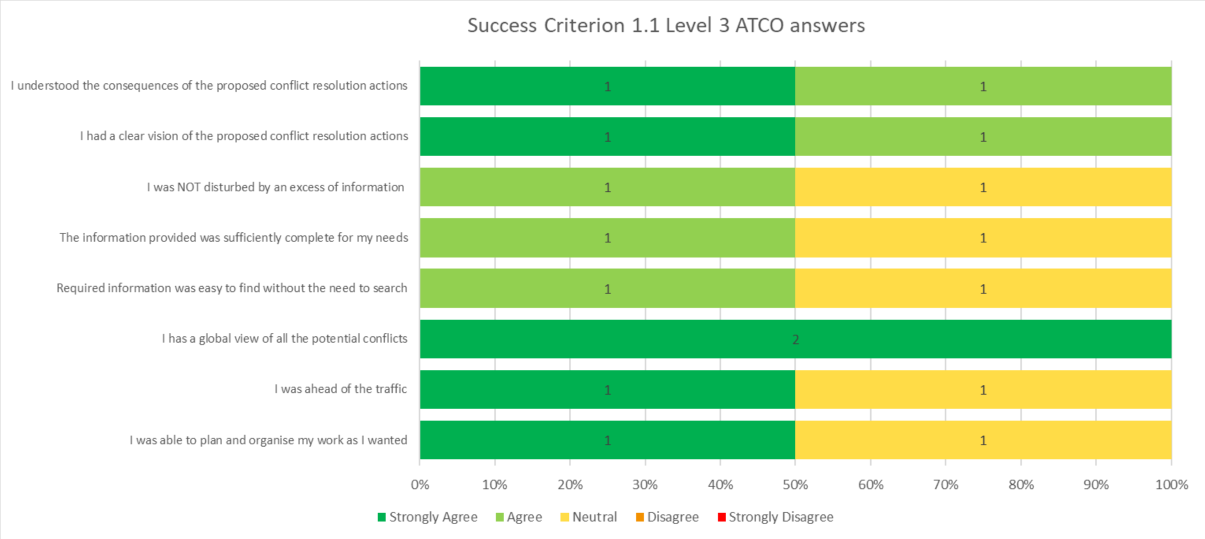


Figure 7: Success Criterion 1.1 ATCO answers. Automation level 3

With Level 3 automation users also scored positively for Objective 1.1, with responses that agreed with the proposed observations, or which remained neutral.

4.3.2.3 ATFCM Scenario – Objective 1.2 assessment results

In general, the supporting tools provided all the necessary information. However, the explanations were not always easy and clear to understand. Nevertheless, FMP understanding on explanations did improve when they had gained more knowledge on the tools.

Sub-Focus	Success Criterion	Success Criterion Verdict
1.2 Identify when support information is required, what level of detail is needed and how should it be provided in a timely manner.	Key data that can be easily understood by the human has been identified that supports transparency needs and is provided in the required time frame and at an appropriate frequency	POK

Table 18. TAPAS ATFCM OBJ1.2 assessment result

The second success criterion refers to human understanding of the data presented by the tools and the support that those tools provide to ease human workload. This criterion is considered to be *partially achieved*, since the data was easy to access, but the FMP expert had difficulties when understanding the explanations shown in the VA tool especially for automation level 2.

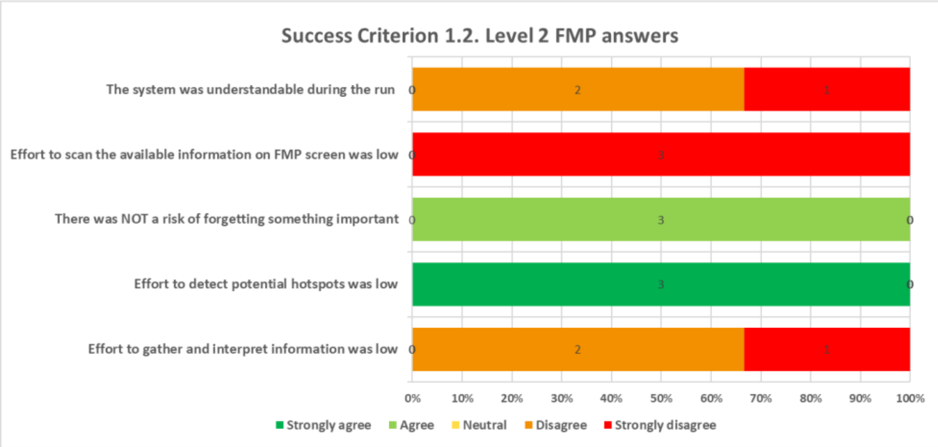


Figure 8. Success Criterion 1.2 FMP answers. Automation level 2

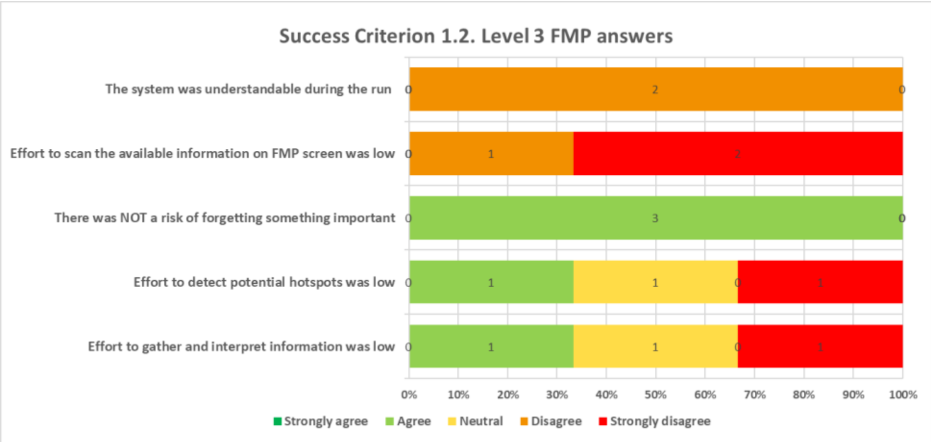


Figure 9. Success Criterion 1.2 FMP answers. Automation level 3

4.3.2.4 CD&R Scenario – Objective 1.2 assessment results

In the CD&R scenario, objective 1.2 considered when support information is required, and which level of detail was appropriate. This included how quickly information should be provided to ensure that the ATCO has sufficient time to respond and deliver a suitable clearance(s) to solve the identified conflict(s).

Sub-Focus	Success Criterion	Success Criterion Verdict
1.2 Identify when support information is required, what level of detail is needed and how should it be provided in a timely manner.	Key data that can be easily understood by the human has been identified that supports transparency needs and is provided in the required time frame and at an appropriate frequency. Additional information providing more detailed information that can help explain more complex situations and the decisions that were made is available for consultation by the user in an 'on-demand' mode if required.	OK

Table 19. TAPAS CD&R OBJ1.2 assessment result

Given the short timeframe available between the identification of conflict situations and the need to identify then deliver clearances to resolve those conflicts, during the exercises most controllers indicated that considering the information provided by the VA support, combined with their own experience and the information and tools available on the standard CWP, the need for additional 'drill down' information was low. Furthermore, if additional drill-down details existed, they would be unlikely to consult it in an operational situation (although this may be useful in a simulation exercise and/or to help validate, certify or enhance the automation tools – which was outside the scope of the TAPAS analysis).

As seen in the responses to questions relating to objective 1.2 topics, responses were predominantly in strong agreement with the statements provided and ATCO were happy with the information being provided, its content and ability to support transparency and understanding of the situation and proposed actions. Similarly, the information allowed users to maintain a good global overview of the situation and it was provided with sufficient time to allow them to understand the situation, review the options being proposed and select a solution.

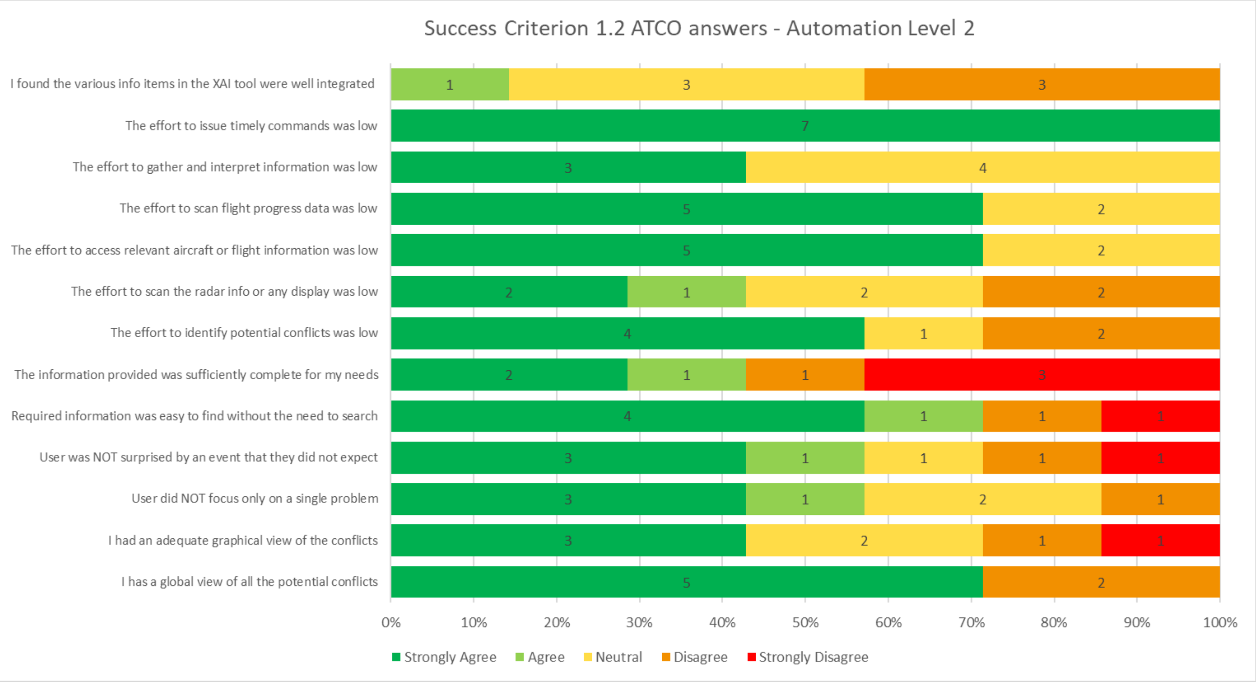


Figure 10. Success Criterion 1.2 ATCO answers. Automation level 2

For automation Level 3, responses to some of the questions that were related to objective 1.2 were more varied. However, only 2 scenarios were executed at this level so it remains difficult to draw distinct conclusions from these scores.

Nevertheless, the answers collected did reflect those seen at Level 2 in the sense that users considered that low effort was needed to establish a good understanding about what the automation was trying to do. However, some disagreement was seen in responses that related to how that information should or could be interpreted and how to cross reference that information with the traffic situation shown in the radar display.

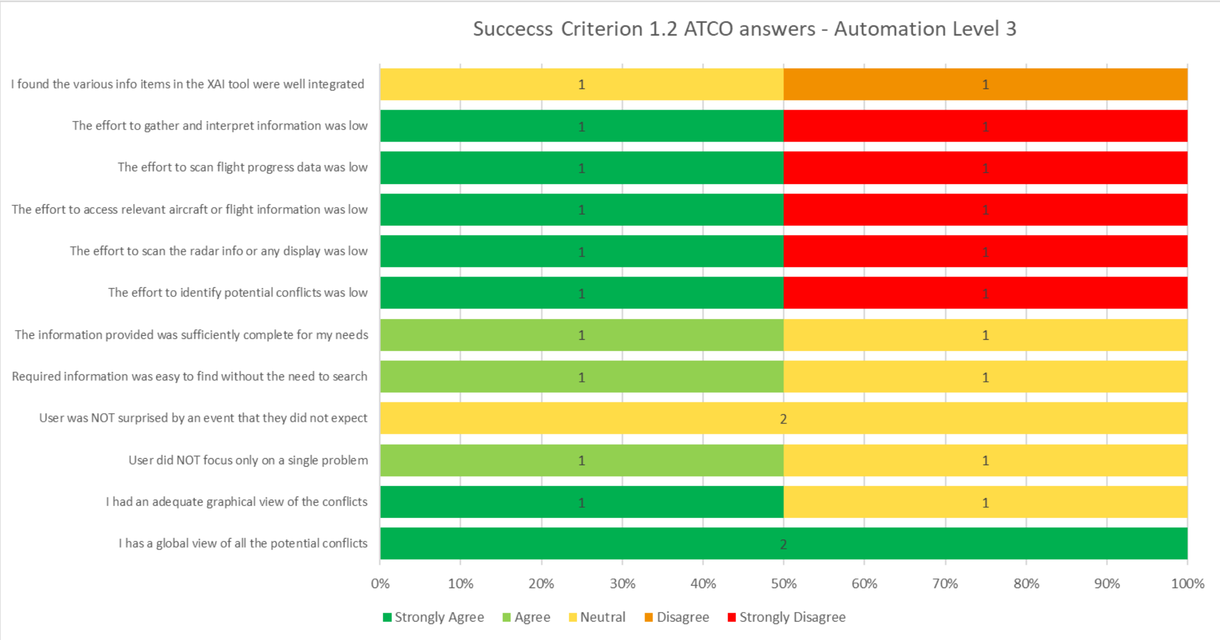


Figure 11. Success Criterion 1.2 ATCO answers. Automation level 2

4.3.2.5 ATFCM Scenario – Objective 1.3 assessment results

Sub-Focus	Success Criterion	Success Criterion Verdict
1.3 Evaluate areas where the levels of transparency may need to be improved.	Information that is unavailable but could help during the use of the proposed XAI has been identified and catalogued for future analysis.	OK

Table 20. TAPAS ATFCM OBJ1.3 assessment result

In general, regarding the data provided by the supporting tools, the FMP experts comments agreed that all the information they needed was included in some way and was presented in the tools. In some aspect the systems were very intuitive to use (e.g.: information on declared hotspots, traffic counts charts, etc.). Nonetheless, in some cases the access to information was somewhat obscure and could be improved.

Users specifically mentioned that the visualisation of aggregated information describing the impact of the proposed measures could be very useful in the future when assessing the effectiveness of the XAI solutions. This was considered especially necessary since the methodology applied by the XAI algorithm consisted in a global technique, solving all the hotspots in Madrid ACC at once, in contrast with the way the FMP currently works, solving hotspot locally, focusing on a single TV at a time.

Therefore, during the exercises execution this aggregated information on the impact of the proposed measures was identified as unavailable (the information on the impact was there but not in an aggregated way) and very useful to improve the use and transparency levels presented by the tool. In consequence, this new founding will be taken into consideration and included when tackling the other TAPAS use case, the CD&R.

4.3.2.6 CD&R Scenario – Objective 1.3 assessment results

EUROPEAN PARTNERSHIP
Co-funded by the European Union

Sub-Focus	Success Criterion	Success Criterion Verdict
1.3 .	Information that is unavailable but could help during the use of the proposed XAI has been identified and catalogued for future analysis.	OK

Table 21. TAPAS CD&R OBJ1.3 assessment result

During the CD&R experiments, ATCO were generally satisfied with the information made available which was easy to use and complemented their own expertise and the existing tools available in the SACTA CWP.

They indicated that no additional information relating to conflicts and the resolution clearances being proposed was required, even if on some occasions the specific actions were not necessarily the action that they would take without the help of the automation tool.

Some concerns were expressed relating to the additional information provided to support the conformance monitoring task, particularly following a resolution clearance. However, once users became familiar with how that information was provided, they found it sufficient for purpose.

Additionally, some confusion arose when information was not refreshed quickly enough, or when new information overlaid or caused existing information of interest to be partially hidden. This situation was linked to the technical choices that were made to allow the XAI prototype and VA support to update information coming from SACTA (which was a 30-second refresh frequency) and had no consequence on the information that was available and hence no impact on the Objective 1.3 success criterion.

One suggestion of interest (not captured in the questionnaire responses) was that during the exercises where higher levels of automation were active, it was sometimes unclear if the machine was currently processing an action or if the view was not changing due to the 30-second refresh. To improve this an additional feature that shows the status of the automation (e.g. pending, in-progress, applied …) would be useful to keep users informed.

4.3.2.7 ATFCM Scenario – Objective 1.4 assessment results

Sub-Focus	Success Criterion	Success Criterion Verdict
1.4. Propose suitable methods by which the level of understanding and trust in the AI automation can be measured	Questionnaires, 'over-the-shoulder' observation and debriefing analysis metrics have been identified to support the necessary measures.	OK

Table 22. TAPAS ATFCM OBJ1.4 assessment result

Over the shoulder observation, debriefing and questionnaires were used during all the execution. Especially at the end of each exercise, where a questionnaire was delivered to each FMP and a debriefing session was held. These methods of capturing information were very useful.

The scores tended to improve as users became more familiar with the platform. In general, levels of trust could be seen to improve as users became more familiar with the tools. This was reflected in the scores which became visibly 'more green' in the later exercises.

4.3.2.8 CD&R Scenario – Objective 1.4 assessment results

Sub-Focus	Success Criterion	Success Criterion Verdict
1.4 Propose suitable methods by which the level of understanding and trust in the AI automation can be measured	Questionnaires, 'over-the-shoulder' observation and debriefing analysis metrics have been identified to support the necessary measures.	OK

Table 23. TAPAS CD&R OBJ1.4 assessment result

The application of over the shoulder observations, debriefing and questionnaires were used during all the execution of CD&R validation experiments as was the case in the previous ATFCM scenarios. During the execution of the scenarios, several TAPAS team members were able to both observe how ATCO managed with the system and explanatory support tools as well as to discuss specific conflict/resolution situations with the users as they happened.

As previously, at the end of each exercise, questionnaires were completed by the operator and a debriefing session was held. These methods of capturing information were very useful.

4.3.3 TAPAS VAL_OBJ_2 – Development of XAI/VA DST Prototype Methods to support Transparency in Automation

The second objective of the validation experiments was to develop and deploy XAI automation tools with varying levels of automation being emulated to help identify and solve demand-capacity issues (Hotspots) for ATFCM and propose suitable actions using either Regulations, Flight Level Cap measures or a combination of both. In the CD&R use case, the XAI tools would identify separation problems (conflicts) in the controlled airspace sector and propose clearances that could be used to resolve the issue(s) while monitoring traffic compliance.

Associated with the XAI component, a dedicated VA display was included in the FMP and ATC working positions. This was used to provide explanatory information and interactive visualisation features to support transparency at various automation levels, particularly for the ATFCM use case. It should be noted however that, In CD&R use case, it was more difficult to provide highly detailed explanatory information or optional 'drill-down' features due to the very short time frame between the detection of a problem and the need to resolve it.

Overall, ten separate validation objectives along with appropriate success criteria had been identified in the validation plan. These are reported in detail in the sections below.

For the ATFCM experiments, five (5) of the criteria were considered as *OK* and a further three (3) as *Partially OK*. This suggests that from an overall perspective, the general feeling was that the success criteria were OK and hence the Objective 2 can be considered *partially achieved* for the ATFCM use case. The remaining two (2) criteria, which were indicated as *Not OK*, were not included in the validation exercises and have therefore been assigned NOK by default as they were not tested.

For the CD&R use-case, at Level 2 the OBJ-2 criteria is considered to be *Partially Achieved* since 5 of the 10 sub-criteria that were scored by the users were OK, 5 and the other 5 were considered Partially OK.

Sub-criteria for both ATFCM and CD&R use cases are detailed below:

4.3.3.1 ATFCM Scenario – Objective 2.1 assessment results

Sub-Focus	Success Criterion	Success Criterion Verdict
2.1 Produce customised VA views to support transparency and explanatory information to the human operator at different levels of automation.	VA display tools are able to consume data provided by the XAI component to support interactive drill down views for the human operator	OK

Table 24. TAPAS ATFCM OBJ2.1 assessment result

This success criterion is considered to be met, since a VA display was developed appropriately to support FMP actions and to present them relevant data from the XAI algorithm or component. The VA tool presents a series of visual aids to the users, so they could further investigate the situation, further analyse a hotspot and flights involved in the hotspots, as well as to consult the explanations provided by the XAI component on the different propose/implemented solutions.

The VA tool has different views available:

- A general view that shows information about the overall state of the sectors in terms of traffic demand (entries and occupancies) highlighting the detected hotspots by the XAI algorithm.
- A more specific view to analyse a certain sector, through which the user can see the flights crossing that sector in a 2D representation and analyse the declared hotspot and flights involved during that time.
- Another series of views exclusively to present to the user the explanations giving by the XAI algorithm on the different solutions (flight list, hotspots detected per flight, time duration inside a hotspot, etc.).

All those views allowed users to drill down and interactively obtain more detailed information from the VA tools whenever they wanted.

4.3.3.2 CD&R Scenario – Objective 2.1 assessment results

Sub-Focus	Success Criterion	Success Criterion Verdict
2.1 Produce customised VA views to support transparency and explanatory information to the human operator at different levels of automation.	VA display tools consumed data provided by the XAI component to support interactive views for the human operator in a timely and concise manner.	OK

Table 25. TAPAS CD&R OBJ2.1 assessment result

In practice, due to the short timeframe between conflicts being identified and the need to implement a suitable and safe clearance to one or more flights, the usefulness of the co-located view was found to be limited in the CD&R use case.

Over the shoulder observation and discussions held during the Level 2 exercises confirmed that the ATCO tended to *use the existing functions and features within the SACTA CWP* (features that they are highly familiar with already) over any new features that were provided by the VA support display.

Furthermore, as the amount and type of information that was able to be provided was limited, by design, to keep it as concise as possible, in-depth 'explanatory' information was not available from the VA tool. Users also indicated that *since little time is available to use information of this type in a CD&R*

EUROPEAN PARTNERSHIP

Co-funded by the European Union

situation, decisions and actions must be performed and verified rapidly, before moving on to the next task.

Some minor issues were identified during discussions with the users that related to the set-up of the working position:

- Since the VA support display was located *next to* the main CWP and was only provided on a small laptop screen, *consultation of information on that display resulted in the user losing focus on the main CWP and on-going traffic situation, which in turn could risk a loss of situational awareness or worse*.
- Similarly, the update frequency (30 sec) and lack of clear indication about when data was last refreshed could lead to users being unclear whether the information was up to date (or remaining from a previous situation).
- In some cases, new data appearing on the display could lead to other information being hidden or potentially lost.

A clear improvement in how the information was provided and used by the ATCo was observed when the CWP interface was enhanced to include an additional pop-up window to allow the VA information to be consulted directly from within the CWP traffic display.

In this case it was evident that the ATCO made greater use of the information coming from the VA support tool than when it was co-located on the smaller laptop screen.

Most, if not all ATCO also indicated that they did not really require the additional graphical view of conflict situations that accompanied conflict details in the VA display. In general, the existing tools in the CWP HMI, combined with the expertise and experience of the ATCO were considered sufficient to understand all the conflict situations rapidly without the need to consult the additional graphical view. Measuring tools and other features available in the CWP HMI were seen as the '*go-to*' function when reviewing conflict situations and potential clearances so the need for additional support via the VA display was low.

Given that the only issues identified were directly linked to artificial technical constraints for integrating the TAPAS prototype tools with the SACTA real-time simulation platform and were not related to this the actual data being shared by the tools, this objective is considered to have been successfully *achieved*.

4.3.3.3 ATFCM Scenario – Objective 2.2 assessment results

Sub-Focus	Success Criterion	Success Criterion Verdict
2.2 Assess how the VA methods can help enhance operator understanding and trust in AI-based automation	Elements provided in the VA provide clear visual evidence related to the actions being performed by the XAI tools	POK

Table 26. TAPAS ATFCM OBJ2.2 assessment result

The VA support tool was designed accordingly so it can provide the users with elements to help them to understand the situation, the reasons behind the different actions and decisions the XAI is taking.

From questionnaires and debriefing sessions, it was concluded that sometimes the solutions and reasons proposed were slightly difficult for the user to understand as they were too complex and

EUROPEAN PARTNERSHIP
Co-funded by the European Union

disruptive with the current operating method the FMP follows. This was reflected in the scores of the questionnaires of level 2 and level 3 runs as it can be seen in the figures below. However, in those scores it can also be shown a change from day 1 to the final day. The scores slightly improved as the knowledge on the tool increased, from an overall neutral score to more positive scores.

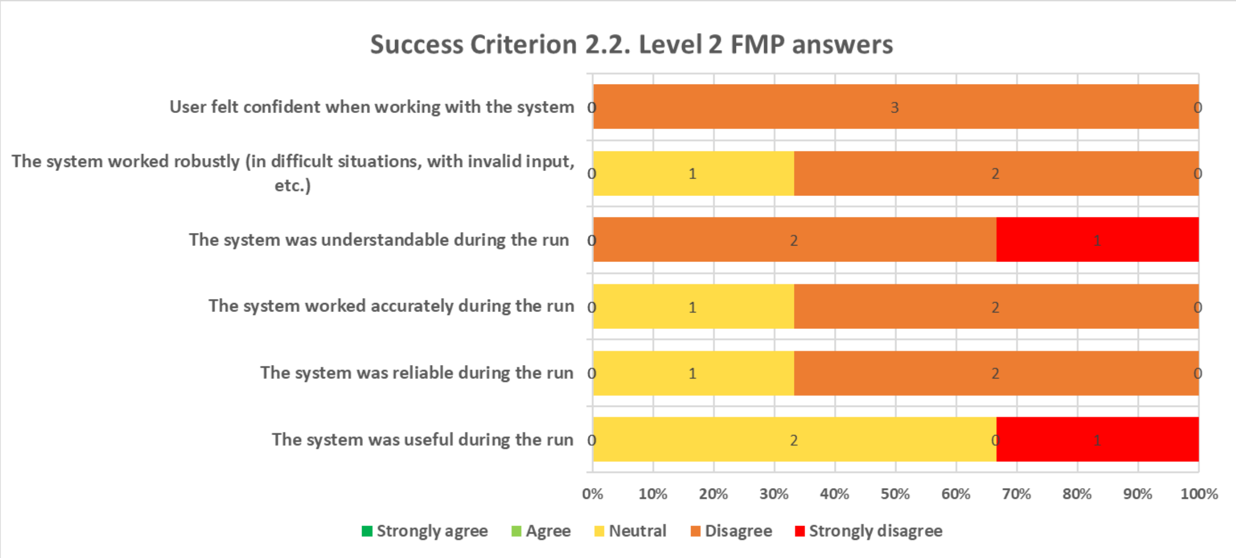


Figure 12. Success Criterion 2.2 FMP answers. Automation level 2

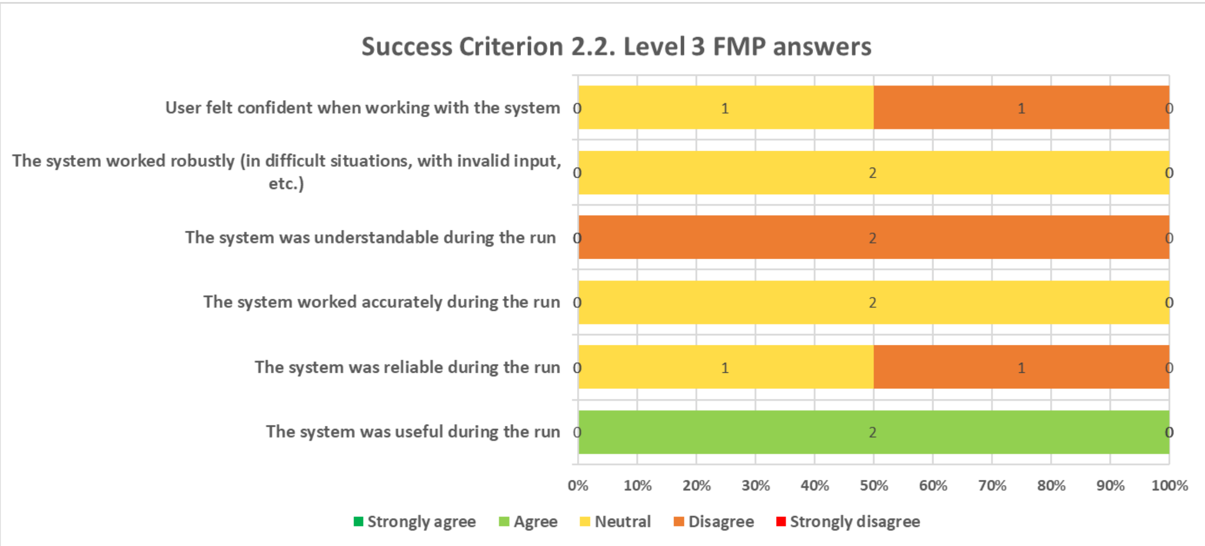


Figure 13. Success Criterion 2.2 FMP answers. Automation level 3

In the debriefing sessions, additional comments from the users suggest that the difference in the paradigm that the XAI uses to create solutions (solve all in one go) contributed to these issues as it was unfamiliar to users who normally solve issues one by one, looking at one sector at a time.

As shown in the questionnaires results above, this factor has a particular impact on level 2 runs, where the FMP must decide whether to implement the proposed solutions, or not.

4.3.3.4 CD&R Scenario – Objective 2.2 assessment results

Sub-Focus	Success Criterion	Success Criterion Verdict
2.2 Assess how the VA methods can help enhance operator understanding and trust in AI-based automation	Elements provided in the VA provide clear visual evidence related to the actions being performed by the XAI tools	POK

Table 27. TAPAS CD&R OBJ2.2 assessment result

During the CD&R experiments, user sentiments were somewhat mixed when responding to questions concerning the visual evidence relating to conflict resolution proposals or actions being provided by the XAI automation tool, as shown in the responses below.

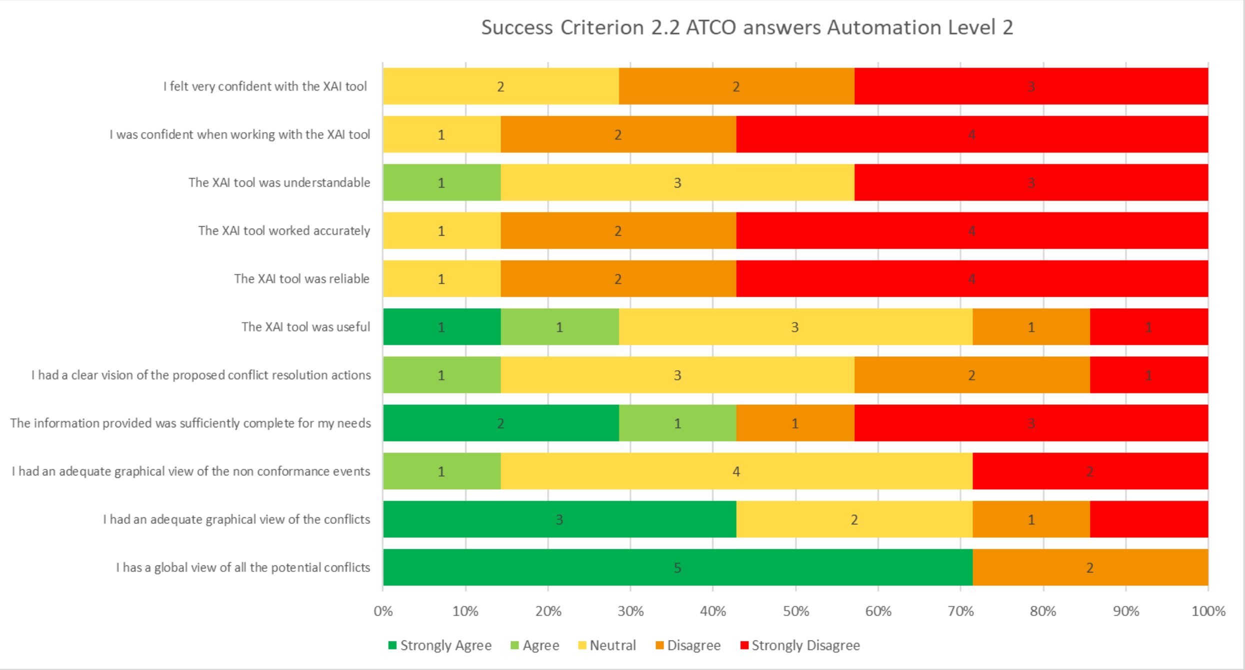


Figure 14 Success Criterion 2.2 FMP answers. Automation level 2

Confidence in the XAI tool and solutions it was proposing were lower than hoped. Discussions held during the execution of the exercises and in the debrief sessions suggested that this may have been due to solutions being offered which were not necessarily the same as those that an ATCO would have chosen by themselves. Moreover, on some occasions, the solutions could lead to additional issues which the ATCO would have been able to avoid based on their own experience working this airspace region.

Additionally, some ATCOs suggested that for some type of conflicts the solutions to be applied are limited and usually recurring. That is, XAI should adapt the “type of solution” (speed change, heading change, FL change…) to the type of conflict being analysed. Not all types of solutions are used for all types of conflicts. For instance, when solving a conflict in an enroute environment involving two aircraft that are at the same flight level, it is preferable to *descend* one of the flights rather than to make one of them to *climb* to higher flight levels.

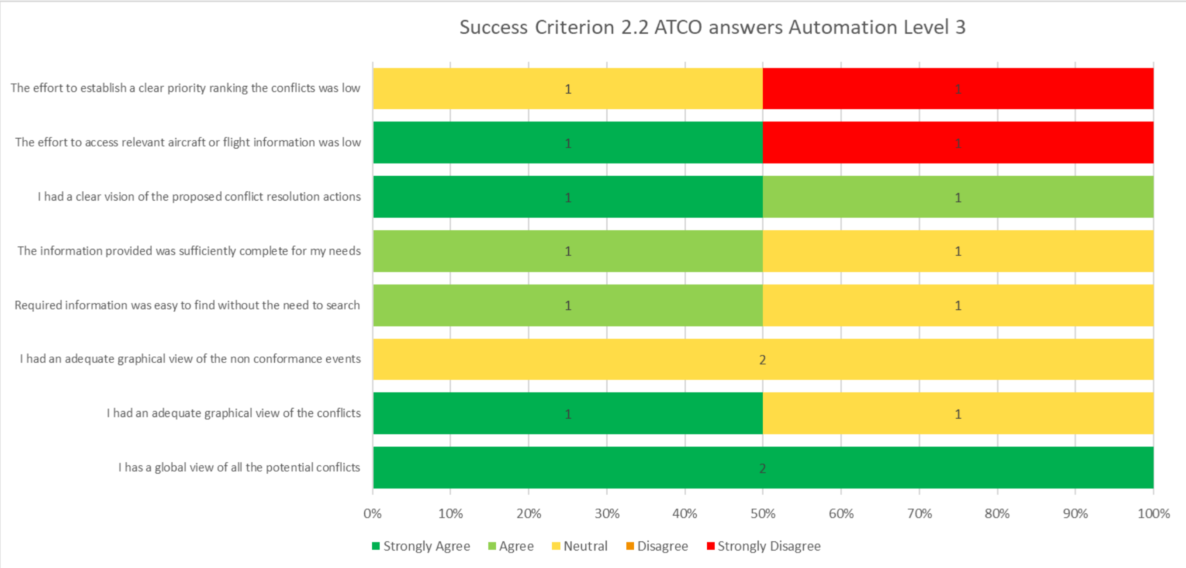


Figure 15. Success Criterion 2.2 FMP answers. Automation level 3

Nevertheless, for questions that were directly related to the information being provided, rather than the performance of the XAI automation and its solutions, responses were more positive, tending to be in the green/neutral region as seen in the lower half of the results shown above.

For this reason, it is considered that objective 2.2 was *partially achieved* for the CD&R scenarios.

4.3.3.5 ATFCM Scenario – Objective 2.3 assessment results

Sub-Focus	Success Criterion	Success Criterion Verdict
2.3 Evaluate the effectiveness of the transparency solutions being deployed	Human operators are able to use the visualisation to interrogate the on-going scenario and solutions being considered	OK

Table 28. TAPAS ATFCM OBJ2.3 assessment result

The DSTs used, VA display and FMP Client interface, allowed the user to easily interrogate the scenario and investigate the proposed solutions. Therefore, this success criterion is considered *fully achieved*.

In the case of level 2 runs, the solutions were published and declared by the FMP. However, before implementing them, the solutions were presented in the FMP Client interface, in particular the proposed demand measures and impacted flights were shown. Through this interface the user also could perform a what-if simulation to evaluate the impact of applying the presented measures. In addition to this functionality, the VA display also provided the FMP with views through which to consult the different solutions and the reasons behind those.

Similarly, to level 2, for automation level 3 runs the user could interrogate the tool to further analyse and explore the implemented solutions and features provided by the VA tool. In this case the difference consisted in that the solutions were automatically published in the FMP Client tool.

4.3.3.6 CD&R Scenario – Objective 2.3 assessment results

Sub-Focus	Success Criterion	Success Criterion Verdict
2.3 Evaluate the effectiveness of the transparency solutions being deployed	Human operators are able to use the visualisation to interrogate the on-going scenario and solutions being considered	OK

Table 29. TAPAS CD&R OBJ2.3 assessment result

Despite the VA display and CWP being provided on separate screens, during the simulation exercises the information provided allowed the user to easily interrogate the scenario, to understand the conflicts that had been identified, and to investigate the proposed solutions – often supported by the other tools provided by the SACTA HMI. Users were able to comprehend and evaluate proposed solutions, even if they did not necessarily agree that they would take such actions using their own experience, which allowed them to choose solutions and issue clearances to traffic to implement them.

In general, users were able to maintain very high levels of situational awareness during the exercises and had no issue in understanding what was happening during the execution of the scenarios.

Therefore, this success criterion was considered fully achieved.

4.3.3.7 ATFCM Scenario – Objective 2.4 assessment results

Sub-Focus	Success Criterion	Success Criterion Verdict
2.4 Determine the different needs for transparency at different automation levels	Human operators classify the information being provided and confirm that it is sufficient to explain the decisions being made	POK

Table 30. TAPAS ATFCM OBJ2.4 assessment result

The success criterion defined for this sub objective was considered to have been partially accomplished. This verdict relies in the feedback collected during the debriefing sessions, which shows varying results depending on the automation level considered.

In automation level 1, the closest automation level to the current operating method of the FMP, the users were able to classify the information provided by the FMP Client Tool (in this level they do not have support of other visuals from the VA component). They highlighted that the information provided was enough to assess the situation: sector load, demand charts (entries and occupancies charts), overloads, flight lists, etc.

However, for automation level 2 and 3, although the FMP were provided with information that they considered useful, understandable, and easy to access, in some cases they mentioned that the way information was presented could be enhanced by providing more aggregated data, like some statistics about the number of impacted flights by the solutions being proposed, most impacted airlines, minutes of delay saved, etc.

4.3.3.8 CD&R Scenario – Objective 2.4 assessment results

EUROPEAN PARTNERSHIP

Co-funded by the European Union

Sub-Focus	Success Criterion	Success Criterion Verdict
2.4 Determine the different needs for transparency at different automation levels	Human operators classify the information being provided and confirm that it is sufficient to explain the decisions being made. Optional detailed views can support more complex situations and can provide additional detailed understanding in an acceptable timeframe	OK

Table 31. TAPAS CD&R OBJ2.4 assessment result

In the CD&R scenario, users indicated that information provided was sufficient for their purpose and no additional detailed information is necessary.

In some cases, ATCO suggested that on many occasions they do not have a need for additional explanations - provided that the solutions are reasonable. This is because in the majority of cases they are already fully capable of understanding conflict situations rapidly due to their own experience and are able to evaluate/understand the possible impact of proposed solutions with relative ease.

At Level 3, however, some concerns were expressed regarding conflicts not being captured or resolutions being ineffective and not solving the conflict. However, this situation did not affect the transparency needs at different automation levels which are considered to be low by ATC.

As the issues concerned the performance of the automation tool itself, which can be put down to technical constraints, this objective was considered to be achieved.

4.3.3.9 ATFCM Scenario – Objective 2.5 assessment results

Sub-Focus	Success Criterion	Success Criterion Verdict
2.5 Evaluate the level of understanding and situational awareness of the human as the automation proposes / implements solutions	Human operators are able to describe what the automation is doing and why solutions have been proposed	OK

Table 32. TAPAS ATFCM OBJ2.5 assessment result

During the execution of the exercises, the operational experts could use the available features in the FMP Client and VA display, which allow them to keep their situational Awareness at all the time. This was particularly true once the operational staff participating in the experiments gained more knowledge on the tools.

At higher automation levels, sometimes the users reported some difficulties when trying to focus on the problem as a whole, rather than single problems. Nonetheless, this is mainly due to the fact that they work nowadays solving one problem at a time, therefore on some occasions they were surprised by induced events that they did not expect. For example, they missed to detect some problems appearing in the VA display (identified by the XAI algorithm) in sectors they were not focused on at that time, as well as some problems and solutions proposed to certain sectors and flights in order to solve a hotspot in another different sector.

The risk of forgetting something or being surprised by an unexpected event can be also seen in the questionnaires results as shown in the figures below. Especially for automation level 3, the FMP experts declared they were missing some of the detected problems and focusing on one single problem.

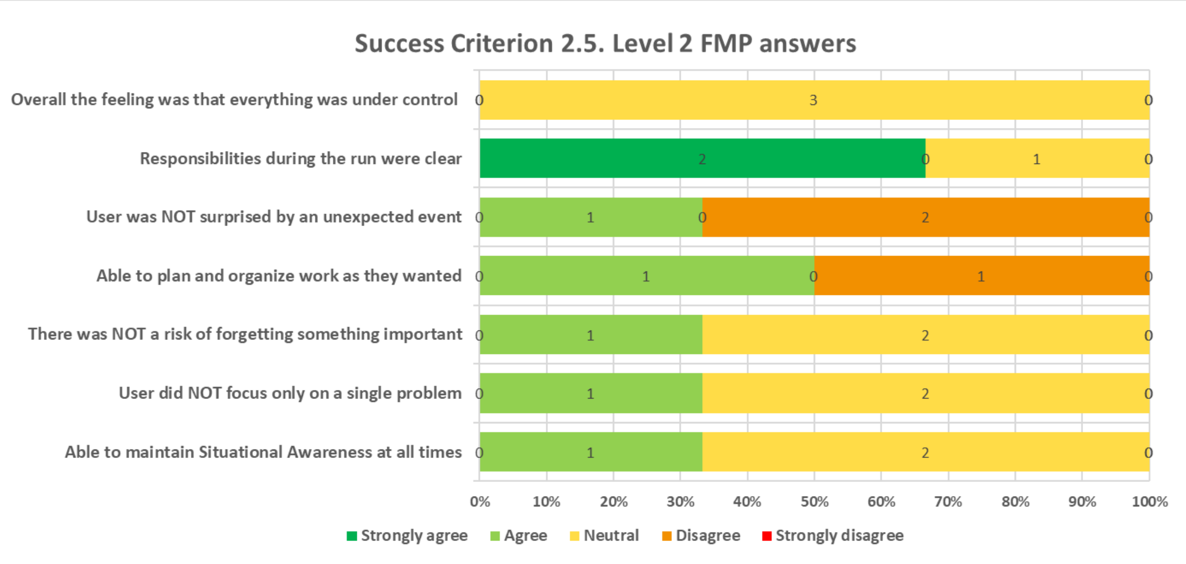


Figure 16. Success Criterion 2.5 FMP answers. Automation level 2

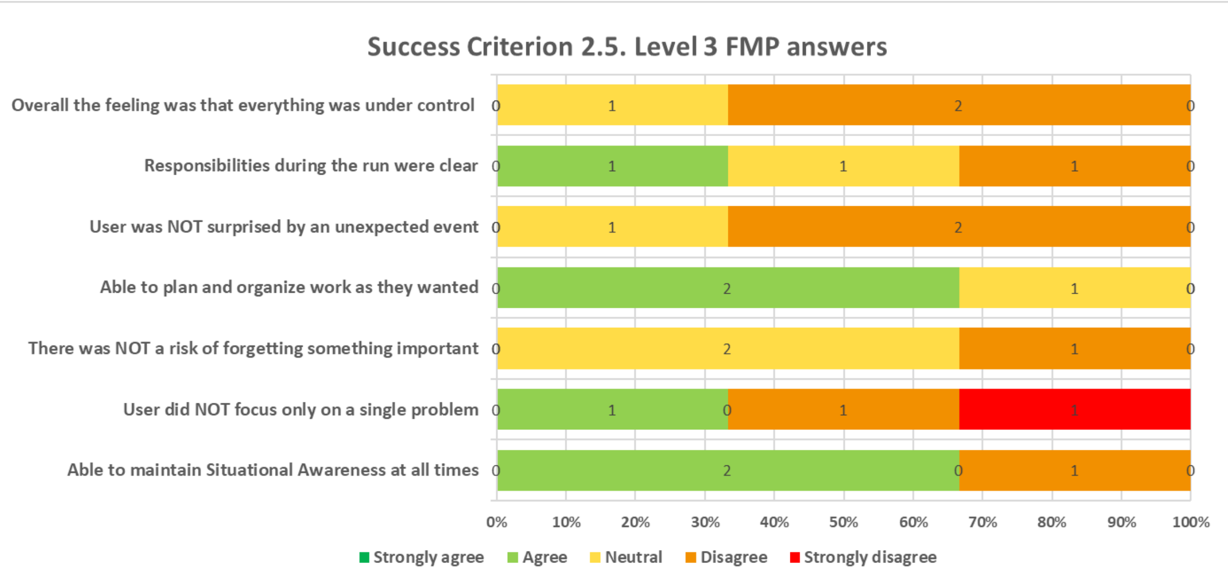


Figure 17. Success Criterion 2.5 FMP answers. Automation level 3

However, in discussions that were held during the exercise and subsequent debrief sessions, this appears to be related to the difference in the methodology applied by the XAI (solving all problems in one go) and in particular to the impact of Hotspots that did not exist at the start of the exercise but were 'induced' as a consequence of earlier solutions to other issues that caused downstream Hotspots to appear.

4.3.3.10 CD&R Scenario – Objective 2.5 assessment results

Sub-Focus	Success Criterion	Success Criterion Verdict
2.5 Evaluate the level of understanding and situational awareness of the human as the automation proposes / implements solutions	Human operators are able to describe what the automation is doing and why solutions have been proposed.	POK

Table 33. TAPAS CD&R OBJ2.5 assessment result

For the CD&R case, users agreed that understanding of most of the actions being proposed by the XAI was able to be achieved to a reasonable level with many answers relating to Objective 2.5 being neutral or in agreement as shown below.

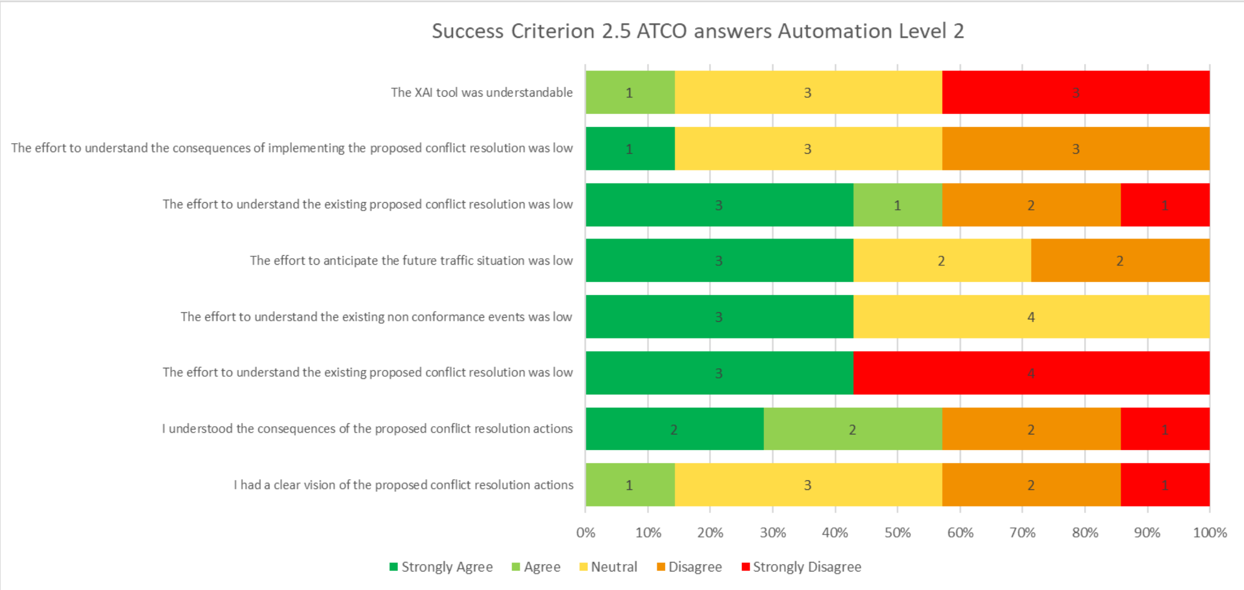


Figure 18 Success Criterion 2.5 ATCO answers. Automation level 2

However, for one specific question, relating to the understanding of proposed resolution actions, users were split 40:60 between strongly agreeing and strongly disagreeing. While this may be a consequence of the actual clearances being proposed by the XAI, and the lack of confidence in the applicability of those solutions by the ATCO, it is still a matter of concern to see such a split in views.

At automation Level 3, users responded more negatively for Objective 2.5 related questions than they had during the Level 2 scenarios.

Responses related to the understanding of why the tool was providing certain solutions appear to be those where users reported more difficulties to understand at Level 3. This may have been further influenced by the choices that were made by the automation, some of which were considered to be unrealistic or inefficient by the users, which in turn may have influenced the responses to questions related to the understanding of the XAI tool and the resolutions being performed. However, despite this, users generally agreed that even if the resolutions were not ones that they would have chosen themselves, the information provided by the system was sufficient to allow them to easily understand the consequence of those actions and to maintain a good overall situational awareness.

EUROPEAN PARTNERSHIP

Co-funded by the European Union

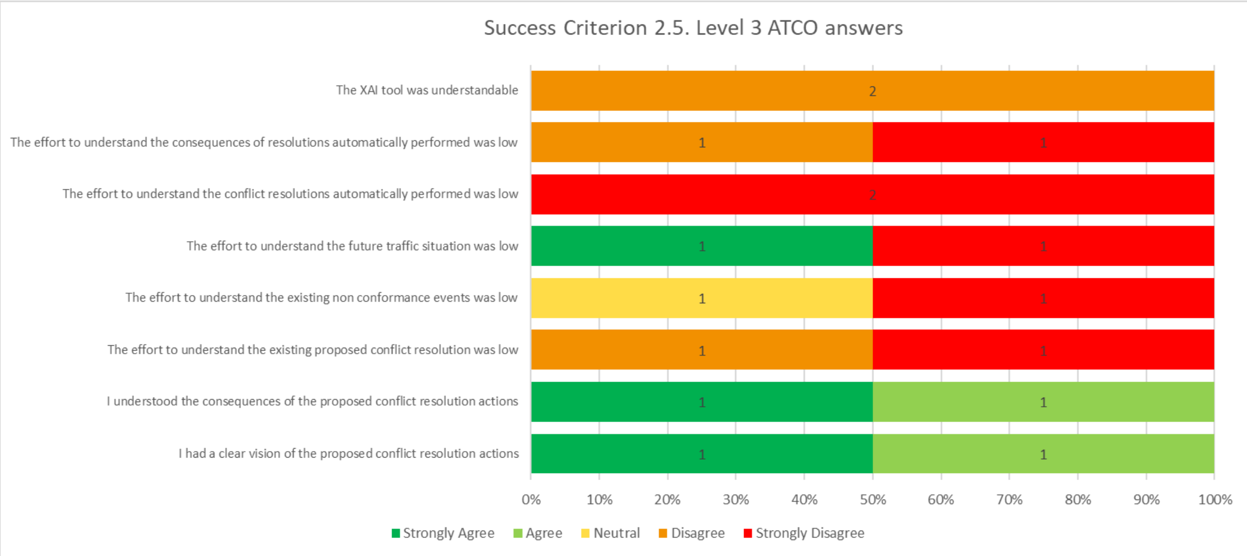


Figure 19 Success Criterion 2.5 ATCO answers. Automation level 3

For these reasons we consider the objective in only *partially achieved* which may justify further improvement or investigation (also given the large number of neutral responses).

4.3.3.11 ATFCM Scenario – Objective 2.6 assessment results

Sub-Focus	Success Criterion	Success Criterion Verdict
2.6 Verify that the human can successfully take over and recover control of the situation if the automation fails for any reason	The human was able to either take over and complete the current task when automation failed,	**Not tested**

Table 34. TAPAS ATFCM OBJ2.6 assessment result

The scenario where the XAI algorithm failed, and then the human tried to solve the remaining issues or hotpots was included in the VALP, but it was not finally tested. The main reason to not simulate this case was because recovering control from a failure in the system required to the human to solve all the problems in different sectors. This would have been a particularly hard task to do manually (even nowadays the delays are not allocated manually, since the FMPs have automated algorithms such as EUROCONTROL CASA algorithm to aid them) and also, since the different measures proposed/implemented by the XAI algorithm are interacting with each other (the algorithm solves all problems at once) it was difficult for the experts to first understand what measures are failing and then select those that they should remove or maintain.

4.3.3.12 CD&R Scenario – Objective 2.6 assessment results

Sub-Focus	Success Criterion	Success Criterion Verdict
2.6 Verify that the human can successfully take over and recover control of the situation if the automation fails for any reason	The human was able to either take over and complete the current task when automation failed,	POK

Table 35. TAPAS CD&R OBJ2.6 assessment result

As explained in the section on variations in the validation scenarios, the ability for the human to recover control of a situation following an unexpected failure of the automation was not specifically tested during the validation experiments due to technical issues with the XAI automation tool and limited availability of the ATC experts.

Nevertheless, as the Level 3 scenarios that were tested included some situations where the automation did not capture all the conflicts or could not successfully provide a suitable solution, users were invited to intervene to provide suitable clearances of their own (or to identify conflicts that have not been found).

In these cases, it was observed that the ATC controllers were fully aware of the situation and were able to provide suitable actions that would resolve the issue when requested and their situational awareness was not degraded, even when performing a monitoring role at Level 3.

However, all the ATC experts that were consulted on this objective agreed that if automation of this type was deployed in the real world, there would be a large risk that over a sustained period, ATCO may become 'de-skilled' which could lead to issues when trying to recover following unexpected failures of the automation.

For this reason, it is only concluded that this objective is *partially achieved* in the CD&R use case.

4.3.3.13 ATFCM Scenario – Objective 2.7 assessment results

Sub-Focus	Success Criterion	Success Criterion Verdict
2.7 Ensure that the human is able to identify and resolve any remaining issues at the end of the XAI process, if present.	The human operator was able to identify and to complete any remaining issues that were not successfully solved at the end of the process	**Not tested**

Table 36. TAPAS ATFCM OBJ2.7 assessment result

Objective 2.7 was not tested during the ATFCM use case due to the difficulties of the human to solve all the remaining issues or hotspots and the time and effort it would have needed to do so.

4.3.3.14 CD&R Scenario – Objective 2.7 assessment results

Sub-Focus	Success Criterion	Success Criterion Verdict
2.7 Ensure that the human is able to identify and resolve any remaining issues at the end of the XAI process, if present.	The human operator was able identify and to complete any remaining issues that were not successfully solved at the end of the process	POK

Table 37. TAPAS CD&R OBJ2.7 assessment result

Objective 2.7 was tested during the CD&R use case automation level 3 scenarios. As described in the previous Objective 2.6, the Level 3 scenarios that were tested included some situations where the automation did not capture all the conflicts. Therefore, there were some remaining conflicts at the start and end of the CD&R processes during the HITL simulations that were detected only by the ATCO and that needed to be tackled by them once the first iteration of the XAI ended.

The ATCO could detect through the CWP radar display those unaddressed conflicts and provide the appropriate instructions to the pseudo pilot to try to solve those conflicts.

Another important aspect that was highlighted by the users was the ability of the proposed solution to subsequently identify additional clearances to traffic to allow them to recover the original route. In practice, ATCO do not consider a conflict to be solved by simply making a manoeuvre that keeps the aircraft suitably separated, which was the approach taken by the XAI automation. The conflict is only fully solved once additional instructions have been provided and actioned that put the traffic back onto the original planned route – especially when it concerns the handover of that traffic to downstream sectors. When using the solutions proposed by the automation tool (at level 2) or monitoring solutions at level 3, ATCO still needed to identify and submit new clearances to bring traffic back to the original plan (if possible), which implies that the resolution processes provided by the automation were not complete. The conformance monitoring feature of the automation tools did help ATCO to capture the need to re-establish traffic on its former route, however, the way of processing that information was not ideal. Overall ATC users would have preferred to receive a more clear and straightforward notification of the exact moment when instructions could be given to allow flights to resume their original plan (or go back to it). This could also have been achieved through the creation and presentation/monitoring of a set of actions that are needed in the form of a check list or other types of action planning feature in the system.

4.3.3.15 ATFCM Scenario – Objective 2.8 assessment results

Sub-Focus	Success Criterion	Success Criterion Verdict
2.8 Demonstrate how transparency can promote operational and social acceptance of 'black-box' AI solutions	The operator confirms that the solutions provided by the XAI were fit for purpose	POK

Table 38. TAPAS ATFCM OBJ2.8 assessment result

While operators indicated that their levels of Situational Awareness remained acceptable for all exercises and that the situation seemed to be under control, general comments regarding the solutions were less positive.

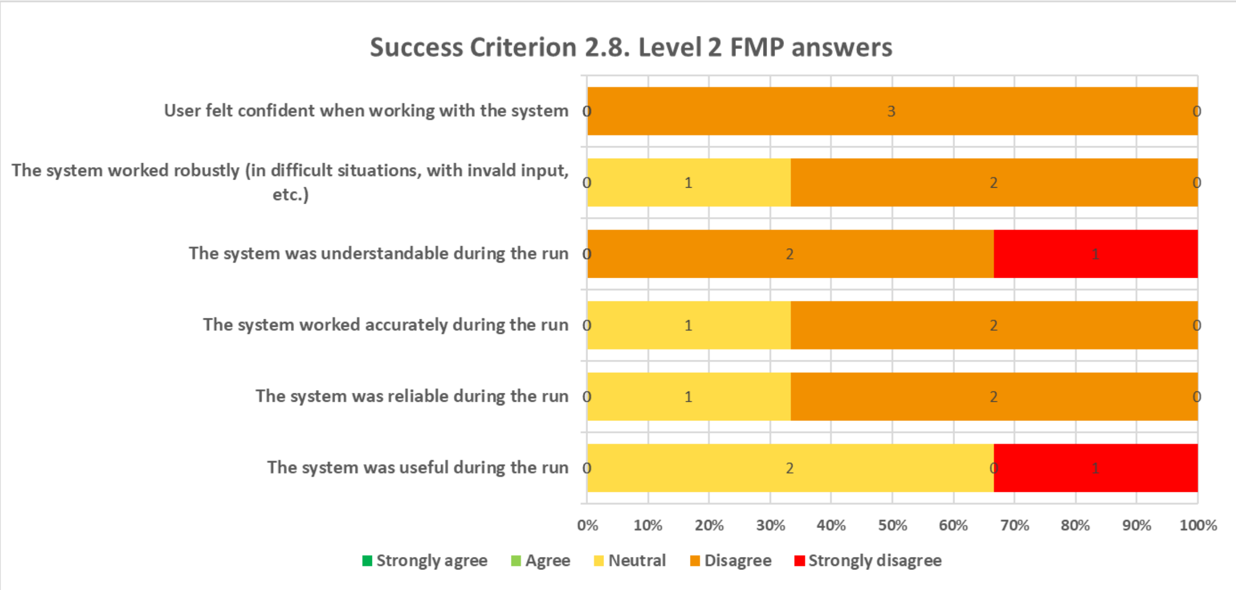


Figure 20. Success Criterion 2.8 FMP answers. Automation level 2 (I)

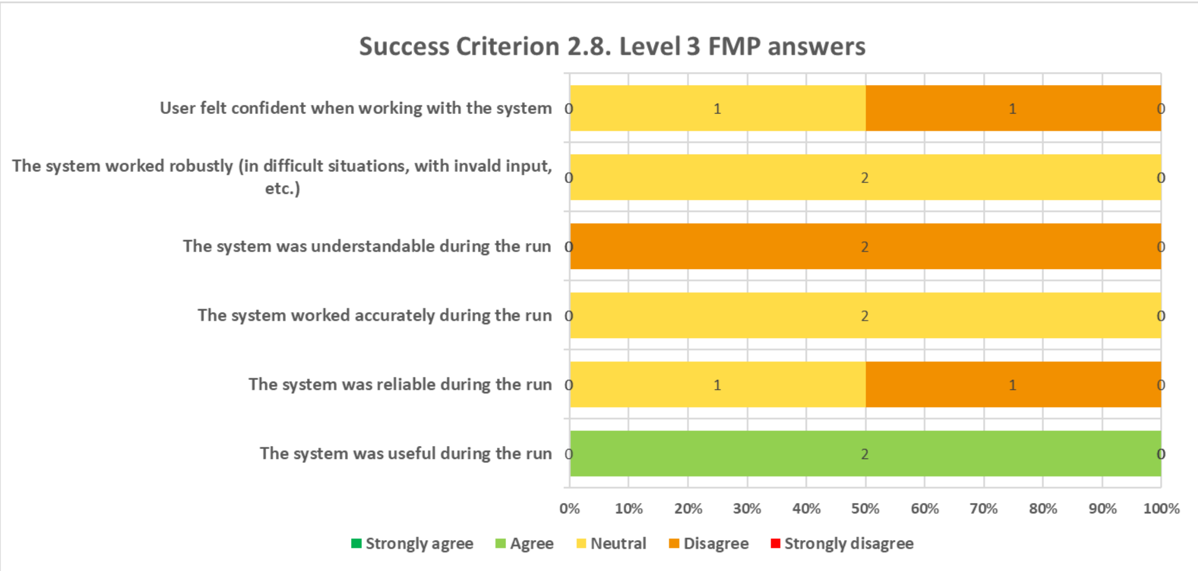


Figure 21. Success Criterion 2.8 FMP answers. Automation level 3 (I)

Users responded either neutrally or slightly negatively when asked whether the system was useful, reliable, accurate and understandable when automation was in place as can be seen in the above figures, which represent the FMP answers and their frequency. Nevertheless, it must be noted that this general view on the system did improve with time and familiarisation with the tools and XAI algorithm.

Additionally, the main objective of the ATFCM exercises execution was not to provide the FMP with perfect solutions, but to extract from them feedback on the principles and how the explanations of the solutions should be provided. In this sense during the debriefing sessions, they stated several times that they do not need long explanations or explanations at all during the operations, but for the solutions to be effective and to see their impact.

As for analysing that impact, the users had the possibility to perform simulations using the *what-if* functionality that the FMP Client tool incorporated. Specific observations were made by users in this regard that included the need to support what-if assessment of proposed solutions (especially for level 2 solutions that required their decision on the implementation) by adding in this assessment the possibility to have higher level aggregated information relating to the impact of the solutions – this information was considered to be important to allow users to evaluate the suitability of solutions, which was otherwise quite difficult to estimate.

Regarding the identification of the hotpots, this identification was reported as being good and useful, but the explanations related to the solutions were reported as being a little difficult to use in some cases. As shown in the figures below, scores related to overall confidence in the solutions remained relatively neutral or slightly negative.

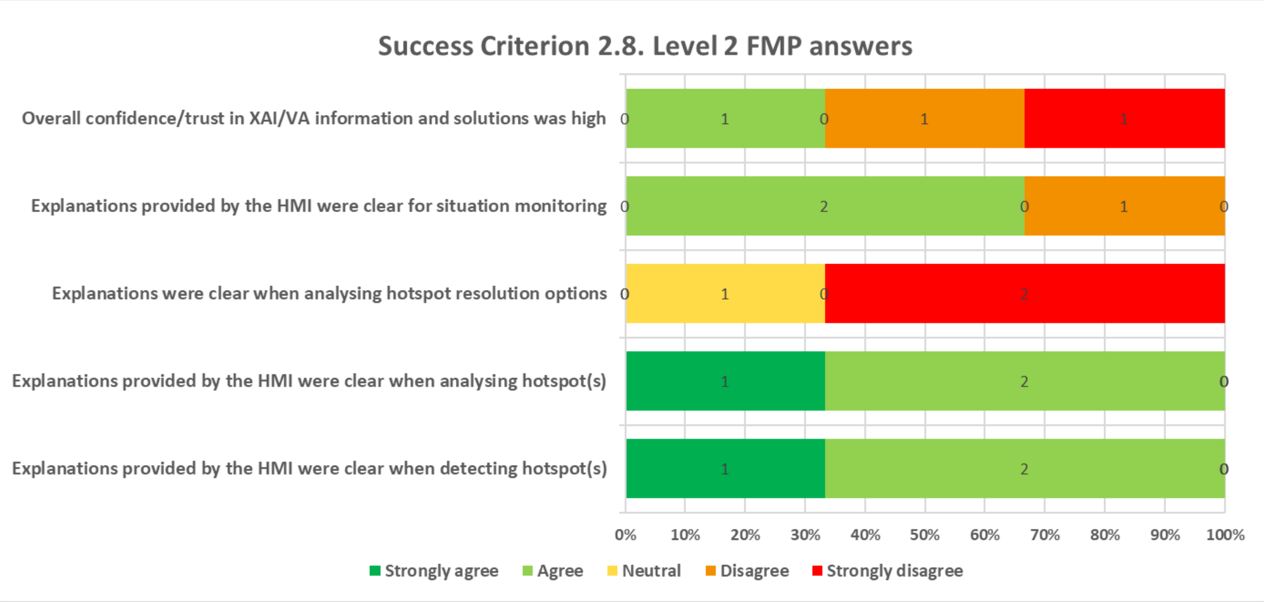


Figure 22. Success Criterion 2.8 FMP answers. Automation level 2 (II)

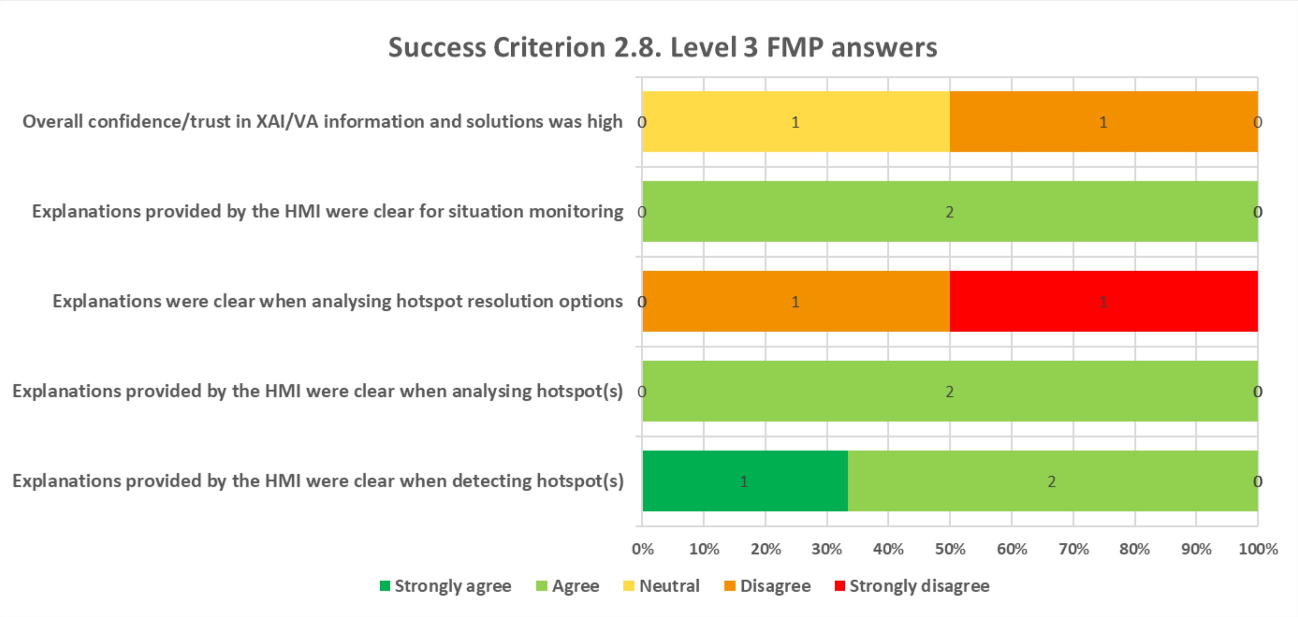


Figure 23. Success Criterion 2.8 FMP answers. Automation level 3 (II)

However, in discussions during the exercises and the debrief sessions, users seemed to be willing to accept that if the solutions that were implemented by the automation were reasonable across all airspace users, then acceptance and trust would be able to be achieved. Additionally, although this is not their current operating method, if the system was accepted and no bias was included in the algorithm when solving all problems at one go, they had no problem with working with the system in such a way.

Nevertheless, on some occasions, especially at level 2 (partial automation), the proposed solutions were not able to resolve the demand capacity issues in the way that users would have expected. Further investigation for level 2 revealed inconsistencies between the XAI paradigm (solve all issues in one go) and the selection and implementation of specific solutions (but not necessarily all) during the level 2 process. The selection of just 'a few' solutions that were indicated as addressing a specific Hotspot did not necessarily solve it, since flights involved in the Hotspot could be solved by actions applied for other Hotspots that were identified elsewhere by the XAI process. Therefore, for this specific automation level all the solutions proposed by the algorithm should be proposed at once to the user.

Also at automation level 2, the FMP experts stated that they expected to have more than one solution proposed by the system so they could compare those solutions, assessing their impact and finally selecting the most appropriate one, however this was not supported.

4.3.3.16 CD&R Scenario – Objective 2.8 assessment results

Sub-Focus	Success Criterion	Success Criterion Verdict
2.8 Demonstrate how transparency can promote operational and social acceptance of 'black-box' AI solutions	The operator confirms that the solutions provided by the XAI were fit for purpose	POK

Table 39. TAPAS CD&R OBJ2.8 assessment result

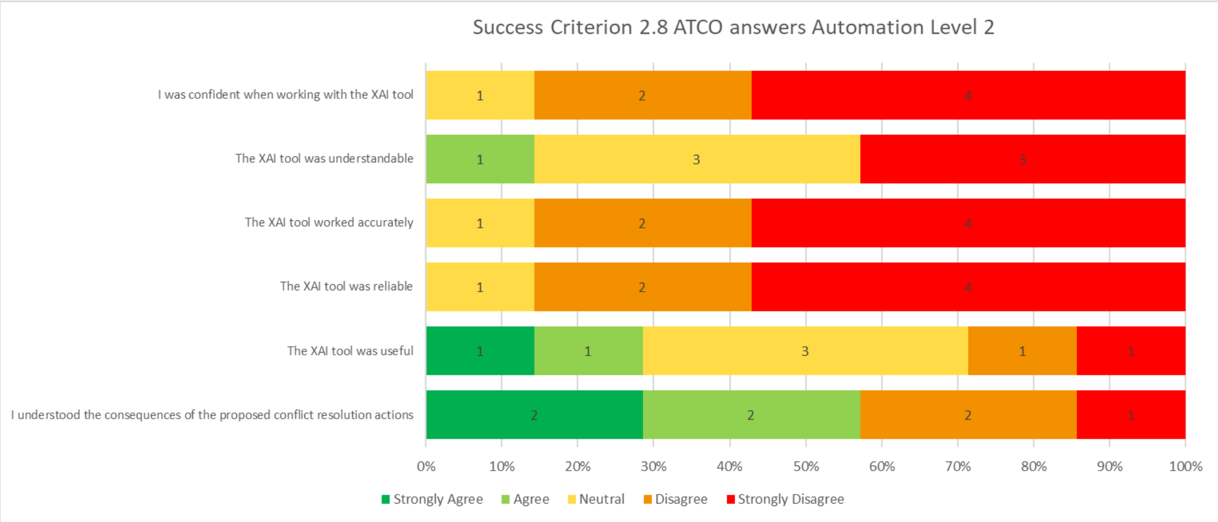


Figure 24 Success Criterion 2.8 ATCO answers. Automation level 2

Although the ATC operators generally found the automation and VA display support information to be easy to use and understandable, solutions that were proposed were often considered to be insufficient for solving the conflicts that were identified. As a result, ATCO confidence in the automation and solutions being proposed was observed to be low, and this is reflected in the scoring for objective 2.8 shown above and is further supported in responses gathered at Level 3 shown below.

Specific reasons provided by ATC experts included clearances that resulted in additional and sometimes more critical downstream conflicts or solutions that requested traffic to climb to solve a conflict when a descent would have been better, or the type of aircraft might not have been capable of flying any higher due to its performance limitations or other constraints. Similarly, the lack of additional instructions to help traffic to resume its original planned route also led to a reduction in the confidence established by the users in the solutions being proposed/implemented by the tool.

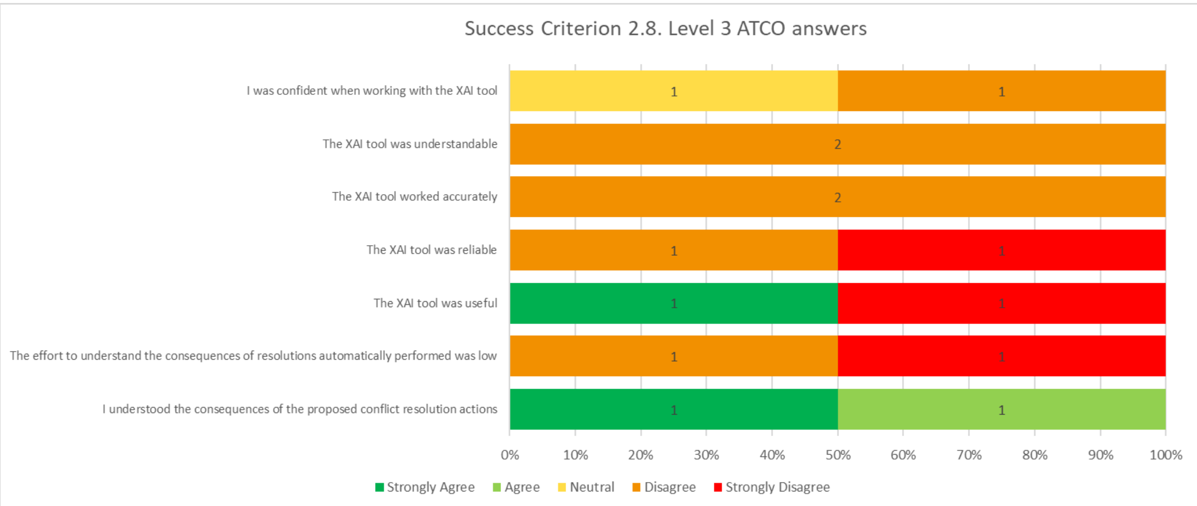


Figure 25 Success Criterion 2.8 ATCO answers. Automation level 3

The overall impact of resolutions being inefficient, not solving issues in a reasonable manner or potentially causing other issues downstream was clearly reflected in ATCO responses to questions regarding the confidence, accuracy, and reliability of the automation tool. Consequently, the users generally indicated that the solutions could not be considered as fit for purpose. However, since this was not directly due to the transparency features, and users indicated that despite the inappropriate solution, they could understand the consequences, we consider that this objective is only *partially achieved*.

4.3.3.17 ATFCM Scenario – Objective 2.9 assessment results

Sub-Focus	Success Criterion	Success Criterion Verdict
2.9 Assess shortfalls and areas where transparency can be improved in future solutions	Operational experts identify areas where information was insufficient to support understanding	POK

Table 40. TAPAS ATFCM OBJ2.9 assessment result

When working at level 3 (full automation) the operational experts were able to review the Hotspots that the XAI identified with relative ease. Similarly, from the overall perspective the effect of the solutions that had been identified and automatically implemented could also be seen and the VA display functions provided sufficient sector and flight-based drill down capabilities to help users understand the decision-making process that led to the solutions that had been implemented.

However, in the post exercise scoring questionnaires, users indicated that the information being presented was still difficult to understand, and that the consequence of the solutions was not as clear as it could be.

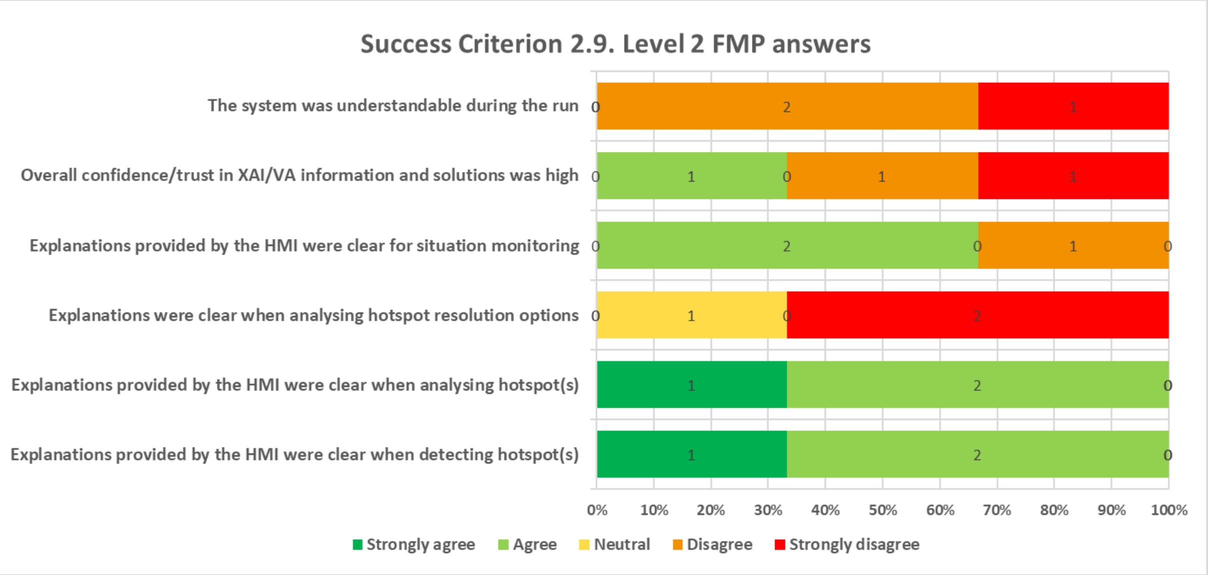


Figure 26. Success Criterion 2.9 FMP answers. Automation level 2

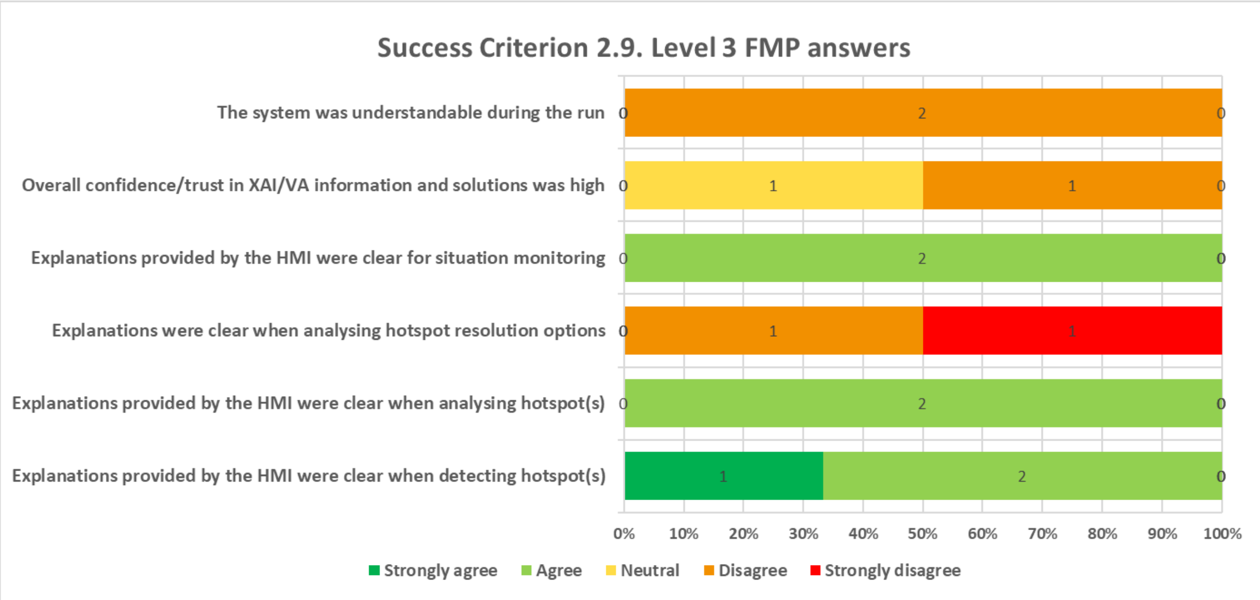


Figure 27. Success Criterion 2.9 FMP answers. Automation level 3

Similarly, scores and general post exercise comments suggested that while there was a great deal of information available from the VA and explanatory features, this seemed to be more focused on 'how the XAI algorithm worked'.

In this aspect, users identified that providing information in an 'operational context' would help improve transparency.

4.3.3.18 CD&R Scenario – Objective 2.9 assessment results

Sub-Focus	Success Criterion	Success Criterion Verdict
2.9 Assess shortfalls and areas where transparency can be improved in future solutions	Operational experts identify areas where information was insufficient to support understanding	OK

Table 41. TAPAS CD&R OBJ2.9 assessment result

During the CD&R scenarios, ATC indicated that all the information they required was easily obtainable from the existing system, with additional information that was provided by the XAI and VA support tool being able to provide enough information about the conflicts and proposed solutions.

This was reflected in the responses related to transparency of solutions being proposed as illustrated below.

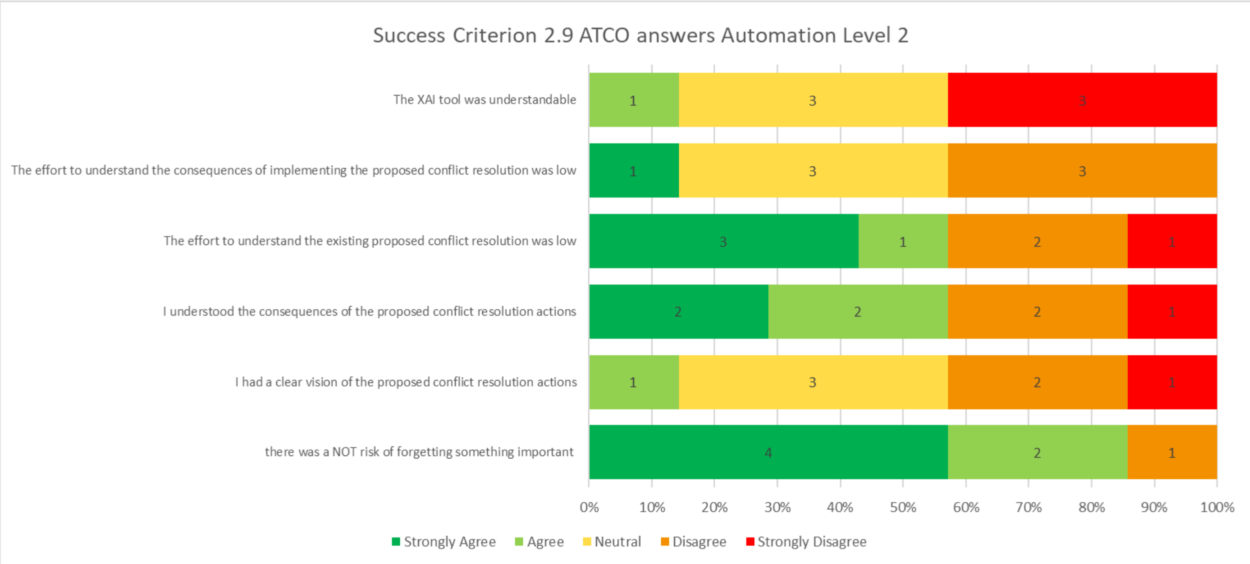
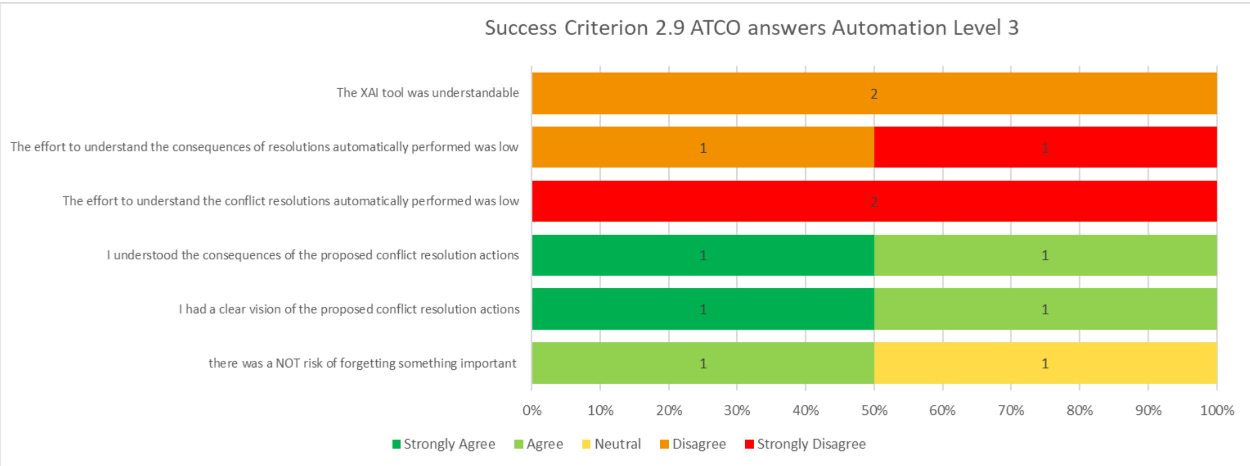


Figure 28 Success Criterion 2.9 ATCO answers. Automation level 2

The only shortfall that was discussed regarding information provided concerned the conformance monitoring function which was not initially easy to comprehend. However, once the ATCO became familiar with the way that this information was provided, they accepted that no further information would be required.



Following the Level 3 scenarios, users responded negatively when answering questions about the way that the XAI was automatically resolving problems, as shown above. When solutions were applied that were not necessarily the same solution that they would use without the help of automation, and which sometimes made matters worse rather than better, users lost faith in the manner that the tool was trying to solve issues. Nevertheless, despite this short coming, users confirmed that the information provided was suitable to allow them to understand the consequence of those solutions, maintain a clear visibility of the actions being proposed and to maintain good situational awareness, and that no additional information was needed for this purpose.

For these reasons we consider that this objective was *achieved* since all the information required was already available and no additional information was needed.

4.3.3.19 ATFCM Scenario – Objective 2.10 assessment results

Sub-Focus	Success Criterion	Success Criterion Verdict
2.10 Identify opportunities for additional training	Additional training or processes to enhance the ability for the XAI/VA to assist the human in understanding the process at different automation levels has been identified by the team	OK

Table 42. TAPAS ATFCM OBJ2.10 assessment result

Opportunities for additional training were identified. After the execution of the simulations, the validation team analysed all the data collected from the questionnaires, debriefing sessions and data retrieved from the FMP Client and VA tool, in order to extract conclusions on three main aspects: explanations and transparency, tooling and lessons learnt to be applied in subsequent use cases, and in particular for the CD&R use case planned in phase 2 of the TAPAS validation.

All these lessons and conclusions are described in more detail later in this document.

4.3.3.20 CD&R Scenario – Objective 2.10 assessment results

Sub-Focus	Success Criterion	Success Criterion Verdict
2.10 Identify opportunities for additional training	Additional training or processes to enhance the ability for the XAI/VA to assist the human in understanding the process at different automation levels has been identified by the team	OK

Table 43. TAPAS CD&R OBJ2.10 assessment result

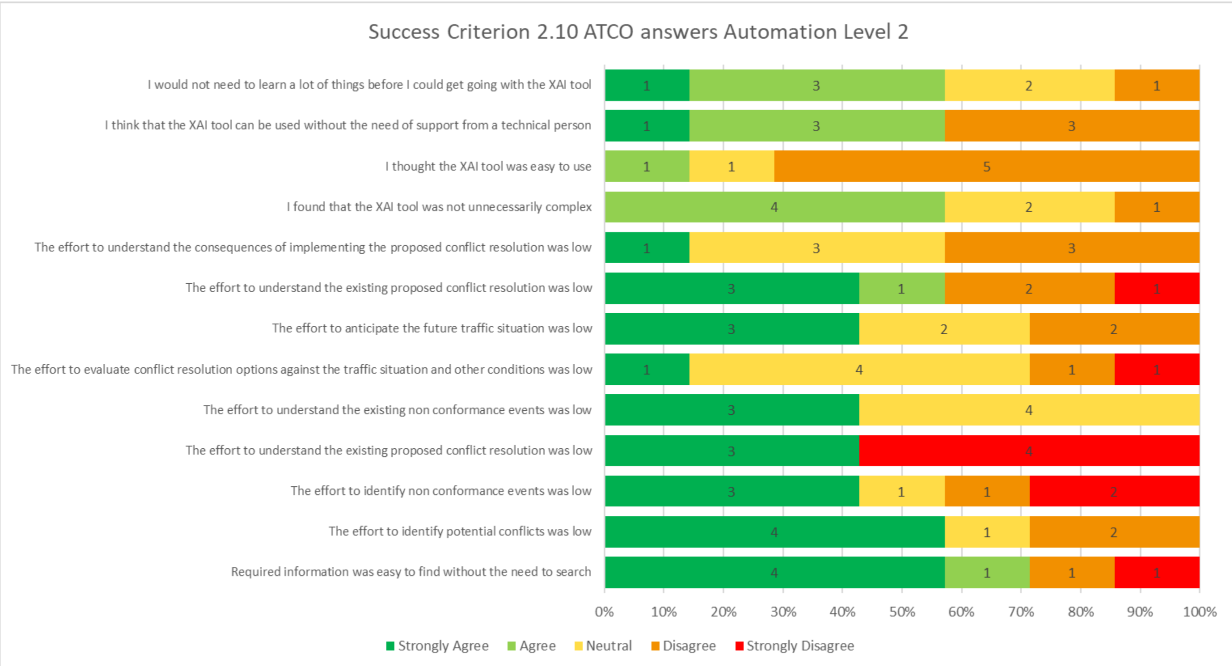
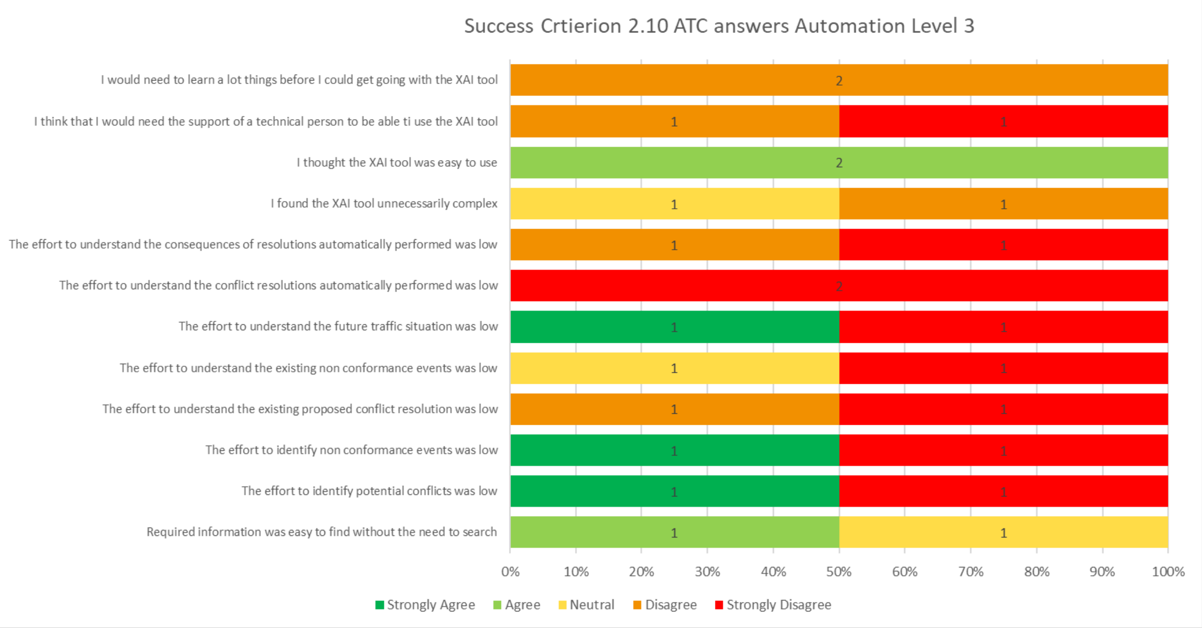


Figure 29 Success Criterion 2.10 ATCO answers. Automation Level 2

Regarding opportunities for additional training or processes to enhance the ability of the XAI/VA to assist the human in the CD&R process, ATCO indicated that the platform and tools were easy to understand, quick to learn and did not need the help of technical support personnel to use. These comments were fully reflected in the answers to questionnaires relating to objective 2.10 shown above.



Responses at Level 3 tended to be a little less positive relating to the need for some additional training or the support of technical personnel when using the tool. Following discussions with the users during the debrief it became clear that these feelings mainly came from the users attempting to rationalise the solutions being applied by the automation tools, which was sometimes confusing, and the lack of additional actions that helped flights to return to their original plan (which was not included as part of the conflict resolution mechanism).

However, overall users indicated that they were happy with the information being provided and how the tools could be used, and indicated that if these issues could be rectified, responses to these questions (at level 3) would be more positive.

As little or no additional training was considered necessary by the ATC experts regarding the XAI/VA display components we consider this objective to be *achieved*.

4.4 Confidence in Validation Results

4.4.1 Limitations of Validation Results - AFTCM

There are some limitations regarding the validation results. First of all, the results of the ATFCM validation exercises are based on a number of realistic simulation exercises in which users can operate as if they are in a real-world environment. However, due to limitation on human resources used (only two operational staff from ENAIRE) the significance of the results is limited. Nonetheless, the operational staff used were highly qualified FMP experts with many years of expertise in the ATFCM field. More participants will be beneficial to enrich the comments and results of the experiment, but the overall status of the validation objectives as they are today most probably will remain the same with the use of more human resources.

During the exercise other capacity measures that could be used to respond to overload issues were not considered, for example, the opened sectors were fixed. Therefore, there was an inconsistency compared to the current way of operation, where the opening and closing of different sector configurations to adapt the available capacity can be considered alongside other demand management measures. Additionally, only two types of demand measures were considered: Regulations and Level Capping.

There was also a limitation regarding data integration between the two different tools used during the validation, the FMP Client and VA tool. Since those two tools were not connected, even though they consumed the same information from the XAI algorithm for automation level 2 and 3, there was a potential risk of having a slightly different flight lists in each tool for the same TV.

Finally, the notion of a Hotspot and corresponding Regulation / Flight Level Capping measures was implemented in a slightly different manner in the XAI tool to that used in NM. The XAI allocated a Hotspot to any 1-hour period that was deemed to be overloaded by more than 10% above the declared traffic entry count threshold using a 20-min sliding window. If two consecutive periods were found to be overloaded by the XAI, two Hotspots were created. In reality the FMP will declare a single Hotspot for the entire period that a region is overloaded which can often be considerably greater than 1-hour, and typically, FMP and NM favour the use of Occupancy metrics over Demand counts. To cope with this difference, algorithms were included in the FMP client application that were able to convert the 1-hour based Hotspots (and solutions) into longer periods more commonly used in NM.

4.4.1.1 Quality of Validation Results - ATFCM

Training sessions with the operational staff were very valuable, as shown in the results retrieved. The perception on the tools and feasibility of integrating more automation improved considerably once users became more familiar with the validation systems, scenarios, and tools.

The participant experts considered the scenario realistic and, because of their background as FMP experts, they were familiar with the simulated airspace, having knowledge on the specific constraints and peculiarities of the region.

For each scenario considered (level 1, 2, 3) several runs were executed with a different FMP and using different traffic samples. Hence, having more runs to gather information helped to improve the quality, accuracy, and significance of the results.

4.4.1.2 Significance of Validation Results - ATFCM

The results obtained during the exercises execution can be extended to other scenarios apart from the simulated ones. Although the airspace simulated is the Spanish Madrid ACC, the operational environment is similar to other airspaces in Europe with similar traffic flows and density. Moreover, the XAI algorithm was firstly trained with information relating to all the flights operating in the ECAC area.

However, as there was limited access to operational experts and as only three morning sessions were able to be arranged to execute the different validation scenarios, results are somewhat limited in scope and size and a larger set of extended experiments with more participants would probably improve the significance of the results obtained.

As for the assessment of benefits and impact in terms of Safety, Efficiency and Human Performance, the used technique was based on a qualitative and quantitative assessment.

4.4.2 Limitations of Validation Results - CD&R

In the CD&R experiments, operational experts that were brought in to perform the Radar/Executive control functions were all highly experienced. This is the situation in any ATC centre where all ATCO must be certified for managing the airspace(s) in the ACC and must maintain that certification and keep their expertise *up to date.*

For this reason, it was clear that ATC experts did not require a significant amount of additional VA/explanatory information to understand where conflicts would occur and what actions could be applied to resolve them.

However, as a result of the technical solution that was adopted to integrate the automation with the simulation platform, when using the XAI tools, ATCO often commented that they were uncertain if the automation had completed its task, or if it was still working on it as no indications were provided to them (and the refresh frequency chosen was 30-seconds).

Other issues relating to the efficiency or suitability of solutions being proposed resulted in the users having low confidence in the automation. This lack of confidence did not impact the view on how results were explained but could lead to artificially lower scores in some cases.

The set-up of the Radar working position, which used a large CWP screen with a co-located VA display on which the conflicts were displayed along with potential actions for traffic that could solve them, was also a limitation during the initial validation exercises. This was mainly due to the need for ATCO to turn away from the main traffic display, to consult the considerably smaller XAI/VA information display screen which could result in a loss of awareness concerning the ongoing traffic situation.

In some of the later exercises, changes to the actual CWP (in SACTA) allowed key information from the XAI/VA tools to be displayed using pop-up windows in the SACTA HMI. When exercises were performed with this adaptation available, it was clear that ATCO used the XAI/VA information much more than previously - suggesting that the information is of great value. However, this kind of information needs to be presented as part of a truly integrated system rather than on a co-located view.

Limitations on the availability of human resources (operational staff from ENAIRE) can also impact the significance of the results. However, the operational staff that were used were all ATC experts with many years of expertise in the ATC field and CRIDA were able to also call on additional experts from their training unit who were also able to perform scenarios during the experiment. In addition, during the platform verification phase leading up to the validation experiments, the CRIDA team were also able to make use of other ATC experts (a total of 8 ATCOs that were consulted by the team during the platform/scenario verification phase) who could also provide their opinions on the automation tool and explanatory information being provided.

Additional technical issues resulted in the number of exercises performed at Level 3 being reduced.

To compensate for this however, additional exercises were included at Level 2 automation which was considered a good alternative since most of the ATC experts considered that partial automation, which recommends potential solutions to problems, is currently considered to be more realistic in the mid-term than a fully 'hands-off' monitoring solution for ATC.

4.4.2.1 Quality of Validation Results – CD&R

Through the inclusion of a reasonable number of validation exercises at Level 2 with single or multiple ATCO working the Radar/Executive control functions and responding to proposals provided by the XAI tool through the VA interface, results for Level 2 are considered good quality.

Due to the technical issues with the tools and restricted access to ATC experts however, only a limited set of scenarios with higher automation (Level 3) could be performed. However, ATC experts indicated that at this stage, mid-term scenarios where automation may begin to support the ATC function are likely to feature partial automation that is similar to the Level 2 implementation used in TAPAS. So reducing the number and scope of the Level 3 experiments is not considered to be a major disadvantage at this point of the research.

The use of the SACTA platform with its highly realistic ATC and traffic modelling environments, combined with the CWP which provides all the same functions that are currently used by active controllers in the Spanish airspace also added value to the experiment outcomes – not least due to them being performed in a very realistic environment.

Finally, the organisation by the CRIDA team, and the clean interoperable integration of the XAI and VA display components with the SACTA simulation platform allowed the components to work in real-time to provide clear and timely indications of predicted conflict situations, potential solutions (to one or more flights) and to provide conformance monitoring functionality as needed. For this reason, ATC experts were able to work in a highly realistic environment when performing the TAPAS CD&R experiments.

4.4.2.2 Significance of Validation Results – CD&R

Access to additional operational experts and the ability to execute multiple scenarios, particularly at Level 2, helped to enhance the significance of the results obtained during the CD&R experiments. During the 3-days of validation using the SACTA real-time simulation platform with the XAI and VA components closely integrated and operating interactively in real-time, 10 distinct exercises were able to be performed that involved 13 operators.

An enhanced set of questionnaires was produced to accompany the experiments which allowed the team to investigate how black-box AI-based automation solutions can be deployed in a real-time safety critical environment and which additional information is necessary to help users understand and gain confidence in how those tools are working.

4.5 Traceability of TAPAS Validation Results with the Requirements

4.5.1 Traceability of the results with Transparency Requirements

The following tables present a summary of the traceability between the results obtained during the exercise execution and the requirements developed regarding transparency. For more details on this, please consult the deliverable D3.2 Principles for Transparency in AI-ML automation in ATM, Edition 00.03.00., [7]. This last document aims at providing an analysis of the proposed requirements fulfilment according to the validation results and other information extracted from the prototypes available during these trials.

TAPAS ATFCM Transparency Requirement	TAPAS ATFCM Validation result	Fulfilment verdict
	Automation level 2 requirements	
REQ-ATFCM-T001-BD. The FMP shall be provided with the set of decision rules that led to the identification of a hotspot or optispot.	Rules for the identification of hotspots are deterministic and very simple in the present form of the prototype. The FMP client presented the identified hotspots in a clear way.	MET
REQ-ATFCM-T002-BD. Should a decision tree be built as part of the rule-based system, the FMP shall be provided the complete decision tree containing the rules regarding the identification of a hotspot or optispot.	The FMP was provided with explanations on the identified hotspot, but the cause of the hotspot was not always clear. Finally, no decision tree existed for this purpose.	PARTIALLY MET
REQ-ATFCM-V003-BD. In order to alleviate the congestion of sectors that are cluttered and overloaded with multiple individual aircraft trajectories, the FMP should be provided with effective visualization means in order to make the visualization clearer and more understandable.	The FMP was able to consult both FMP client and XAI/VA tool, where he could extract relevant information, such as identified hotspots, proposed demand measures, TV charts, explanations on the proposed solutions, etc.	MET
REQ-ATFCM-V004-BD. Trajectories should be clustered by a proper similarity measure and, depending on different density (number of trajectories) thresholds, a group of similar trajectories will be represented by one single, wider, and bigger trajectory flow.	The requirement is considered unmet as it has not been included into the demonstrator as it was originally defined. However, a more appropriate technique is provided. Although tools for clustering of trajectories based on a variety of similarity measures are available, it was found during the trials that these methods are hardly applicable in the DCB scenario. Instead, grouping of trajectories that cross similar sequences of sectors is provided within the sector explorer component. In this tool, the grouped flight trajectories are presented in combination with time histograms displaying demand in involved sectors.	UNMET
REQ-ATFCM-T005-BD. The FMP shall be provided with the most effective DCB measure (or combination of DCB measures) to solve the hotspot.	The best DCB solutions were provided to the user through the FMP client and explanations through the XAI/VA tool.	MET

Co-funded by the European Union

TAPAS ATFCM Transparency Requirement	TAPAS ATFCM Validation result	Fulfilment verdict
REQ-ATFCM-T006-BD. The FMP shall be provided with the most relevant output parameters that influence the most in determining the proposed DCB measure (or combination of DCB measures).	XAI provides, for any single decision taken, all parameters that influence the decision. However, it should be point out that during the debriefing sessions the FMP experts stated that some aggregated information, statistics about solutions impact would be of interest, apart from the information already included in the tool	**MET**
REQ-ATFCM-T007-BD. Should any kind of importance be given to some input parameters used in the determination of the DCB measures, this information shall be explicitly provided to the FMP in form of explanations.	The XAI prototype provides all parameters that affect a decision, but an explanation is not provided as required. During the exercises FMP experts declared that explanations were not always easy to understand as it involved different hotspots (the problem was considered from a more global perspective, rather than the local approach currently used).	**PARTIALLY MET**
REQ-ATFCM-V008-OD. Additionally to REQ-ATFCM-T005-BD, the FMP shall be provided with a preliminary visualization on the display (s) of alternative and backup solutions representing other 'n' most effective DCB measures to solve the hotspot.	Internally, multiple types of alternative solutions (i.e., with different mixtures of measures) are considered, but only one solution is finally provided to the FMP via VA. Three types of solutions and one solution per type are given. For instance, one solution using regulations, one solution using level capping and one using level capping and regulations. In Level 2 the aim of the requirement was to provide several backup and promising solutions for the FMP to decide based on his/her judgment and experience.	**PARTIALLY MET**
REQ-ATFCM-V009-OD. This preliminary visualization should include a motion-based representation of the specific action to be taken (e.g., level-capping or a delay), complemented with a textual display containing the following parameters: • A metric that measures the effectiveness of that DCB measure to solve the hotspot • On-load Areas • Off-load Areas	The most important functionality of this requirement is met by using visualisation means; except On-load Areas and Off-load, that are not supported. Additionally, during the debriefing sessions the FMP experts stated that some aggregated information and statistics about solutions impact would be of interest, in addition to the information already included in the tool.	**MET**

TAPAS ATFCM Transparency Requirement	TAPAS ATFCM Validation result	Fulfilment verdict
• Flights involved and related flight information		
REQ-ATFCM-T010-BD. The FMP shall be provided with explanations regarding the inner reasons of proposing some DCB measures over others.	The system provides all the parameters that affect agent's decisions, also providing the parameters for counter-decisions. The prototype describes the reasons for a specific measure decided. In addition, it provides the arguments that justify not to take that measure. The difference between them could be exploited so view the more important reasons for making a specific decision for a measure, but this is not presented.	MET
REQ-ATFCM-T011-BD. The FMP shall be provided with explanations containing detailed information on how the proposed DCB measure (or combination of DCB measures) impact (s) on other sectors.	The XAI/VA tool provided all the explanations and parameters considered by the algorithm to propose the different decisions/solutions.	MET
REQ-ATFCM-T012-BD. The FMP shall be provided with a list of the most impacted sectors by the proposed DCB measure (or combination of DCB measures).	A list of most impacted sectors by the proposed solution was not provided. However, most of the time, the FMP could infer which potential TVs will be the most affected ones by the proposed solutions.	UNMET
REQ-ATFCM-T013-BD. The FMP shall be provided with explanations and parameters in order to reinforce the trust of the FMP in the AI system in potential counterintuitive situations (to be identified during the validation phase) for the FMP.	The XAI/VA tool provided all the explanations and parameters considered by the algorithm to propose the different decisions/solutions.	MET
REQ-ATFCM-T014-BD. The FMP shall be provided with explanations for her/him to understand the details of the inner process followed during the trade-off process.	The FMP was provided with explanations to understand the inner process of the XAI algorithm, even though it is not proved that the 'FMP understand the details of the inner process', as this must be tested in the validation activities, and it is a lesson learned for the future definition of transparency requirements.	MET
REQ-ATFCM-T015-BD. The FMP shall be provided with all the parameters for her/him to understand	The FMP was provided with explanations to understand the inner process of the XAI algorithm, even though it is not proved that the 'FMP understand the details of the inner process', as	MET

Co-funded by the European Union

TAPAS ATFCM Transparency Requirement	TAPAS ATFCM Validation result	Fulfilment verdict
the details of the inner process followed during the trade-off process.	this must be tested in the validation activities. This is a lesson learned for the future definition of transparency requirements.	
REQ-ATFCM-T016-BD. The FMP shall be provided with a list of the candidate flights or flows impacted by the DCB measure (or combination of proposed DCB measures), along with explanations informing about the key parameters involved in the decision for the selection.	Although it was stated that 'The XAI does not finally work in that way, but we can conclude that the requirement is met, since measures per flight are explained.'	**MET**
REQ-ATFCM-T017-BD. The FMP shall be provided with several alternative solutions to the candidate flights or flows impacted by the DCB measure (or combination of proposed DCB measures), along with explanations informing about the key parameters involved in the decision for the selection	Internally, multiple types of alternative solutions (i.e., with different mixtures of measures) are considered, but only one solution is finally provided to the FMP. Three types of solutions and one solution per type are given. For instance, one solution using regulations, one solution using level capping and one using level capping and regulations. In Level 2 the aim of the requirement was to provide several backup and promising solutions for the FMP to decide based on his/her judgment and experience.	**PARTIALLY MET**
REQ-ATFCM-T018-BD. The FMP shall be provided with the expected benefits obtained, in the sector and other possible sectors, by the different proposed solutions.	Although the FMP user was provided with the benefits/impact of the proposed solutions, for example through the use of what if functionalities. They also stated during the validation exercises that they would also prefer to see aggregated information and statistics about solutions impact would be of interest, in addition to the information already included in the tool.	**PARTIALLY MET**
REQ-ATFCM-T019-BD. The FMP shall be provided with relevant values informing about the total number of affected flights for the different proposed solutions.	XAI/VA tool showed the impact from the flight perspective. Through different drill down functionalities, they could consult the different states of the aircraft. Additionally, through the FMP tool they were informed on the flights affected by the solutions.	**MET**
REQ-ATFCM-T020-BD. The FMP shall be provided with relevant values informing about how long those flights will be affected by the DCB measure (or combination of proposed DCB	All this information was available through the FMP Client as well as presented in the XAI/VA tool.	**MET**

TAPAS ATFCM Transparency Requirement	TAPAS ATFCM Validation result	Fulfilment verdict
measures) for the different solutions.		
REQ-ATFCM-T021-BD. The FMP shall be provided with relevant values informing about how many different airlines are impacted in the different solutions.	The information is not explicitly shown by the prototype, which does not consider airlines. This information could be extracted from the tools but it was not shown directly.	UNMET
REQ-ATFCM-T022-BD. Should there be any other type of valuable information that can help the FMP make accurate comparisons between the different solutions to make final choices, this information shall also be identified and provided by the system.	The FMP was provided with a 'what-if' functionality that allow them to test the solutions proposed by the prototype, in Level 2. The system provides decision rules triggered for decision making, as well as arguments for taking counter-decisions. Tools for running the system simulation towards producing the final solution are provided, also allowing the visualization of all involved flights and joint decision making.	MET
REQ-ATFCM-T023-BD. The FMP shall be provided with a PM that allows the implementation of hypothetical DCB measures (or combination of DCB measures) in order to project, beforehand, their consequences, as well as analysing and assessing the expected impact on the network and potential benefits obtained, in the sector and other possible sectors, by the DCB measure.	The FMP user was provided with: (a) the parameters that affect the decision taken (b) the parameters that drive alternative decisions The impact of alternative decisions is provided by the XAI system. However, no interaction exists with operators for identifying the effects of hypothetical DCB measures.	MET
REQ-ATFCM-V024-BD. The PM shall provide a suitable visualization depicting the expected effect of the implemented DCB measure so that the FMP can quickly and intuitively comprehend the possible consequences of his/her actions in the resolution of the active hotspot (s).	The system provides the functionality even though the FMP 'can NOT quickly and intuitively comprehend the possible consequences of his/her actions in the resolution of the active hotspot(s)'. Mixing technical and interpretability aspects in a requirement is not correct, and it must be merely assessed from a technical side. This certainly indicates an inaccuracy in the definition of the requirement and needs rewording.	MET

TAPAS ATFCM Transparency Requirement	TAPAS ATFCM Validation result	Fulfilment verdict
REQ-ATFCM-T025-BD. In the case of not solving the hotspot, the PM shall provide explanations informing about the underlying reasons of the failure in solving the spot.	No explanations informing about the underlying reasons for unresolved hotspots are provided. Information about a non-resolved hotspot can provide reasons for failure, but no direct explanations are provided.	**UNMET**
REQ-ATFCM-T026-BD. In the case of not solving the hotspot, the PM shall provide explanations suggesting alternative DCB measures which might help solve the hotspot.	In case of not solving the hotspot, no explanations containing and suggesting alternative DCB measures which might help solve the hotspot are provided. Potential solutions are provided, but in cases they cannot solve it, no explanations on failure are provided	**UNMET**
REQ-ATFCM-T027-BD. The PM shall provide a list of the most important parameters involved in the resolution state of the hotspot.	The XAI prototype provides taken all parameters that influence the decision for any single decision. However, the decision of an agent does not consider a single hotspot, but all the hotspots to which it is involved during its flight	**MET**
REQ-ATFCM-T028-BD. The PM shall allow setting up different parameters (e.g., hyperparameter tuning) to be used as input to the ML algorithms, and re-running simulations in order to evaluate the diversity of the possible solutions, as well as their potential impact on the network.	The parameters were fixed without offering the possibility to change them	**UNMET**
REQ-ATFCM-T029-BD. In the case of implementing ATFCM regulations involving specific Regulation Period, Regulation Width and Regulation Rate, the PM shall provide the FMP with the following parameters so that he/she can understand the real magnitude of the expected impact and make pertinent decisions. • Occupancy Counts	All the parameters (with the exception of additional flown distance) were provided.	**MET**

Co-funded by the European Union

TAPAS ATFCM Transparency Requirement	TAPAS ATFCM Validation result	Fulfilment verdict
• Hourly Entry Counts • Total delay • Average delay per flight • Estimated additional flown distance		
REQ-ATFCM-T030-BD. The HRMM shall provide the FMP with up-to-date and relevant parameters in real-time to help the FMP univocally understand what the current resolution state of all the active hotspots is (e.g., number of flights over the declared capacity).	All relevant parameters considered by the algorithm were provided to the FMP in real-time. Nonetheless, **it was not proved** that the FMP unequivocally understands what the current resolution state of all the active hotspots.	MET
REQ-ATFCM-T031-BD. Should the HRMM detects potential deviations, a list possible corrective actions shall be provided to the FMP for further assessment.	Deviations could not be detected. However, the system can proceed the simulation from any state	UNMET
REQ-ATFCM-T032-BD. Should the resolution of a hotspot be not resolved within D-1, the HRMM shall provide accurate and explanations stating the underlying reason (s) of why it has not been resolved.	This feature was not provided through any of the developed prototypes.	UNMET
REQ-ATFCM-T033-BD. Should the resolution of a hotspot be not resolved within D-1, the HRMM shall textually suggest initiate action(s) for preparation of alternative DCB measures (or combination of DCB measures) in order to solve the hotspot, taking into account the current situation of the network.	Alternatives are provided, but not in the same way that the requirement originally states.	PARTIALLY MET

TAPAS ATFCM Transparency Requirement	TAPAS ATFCM Validation result	Fulfilment verdict
	Automation Level 3 requirements	
REQ-ATFCM-T034-BD. The FMP shall be provided with the set of decision rules that led to the declaration, or not, of the hotspot to the NM.	XAI does not focus on individual hotspots, but on the global airspace state	**MET**
REQ-ATFCM-T035-BD. Should a decision tree be built as part of the rule-based system, the FMP shall be provided with the complete decision tree containing the rules regarding the declaration, or not, of a hotspot to the NM.	No decision tree is finally built, but the decision process was able to be followed from the graphic display	**MET**
REQ-ATFCM-V036-BD. In order to complement the information contained in the decision rules, the FMP should be provided with enhanced visualization means in order to better understand the analysed hotspot's root cause and complexity. The visualization shall represent key information in relation to complexity and cause factors, such as: • trajectories • impacts with other sectors • number of flights climbing/descending • number of potential trajectory crossings • number of flows interactions • PRU Complexity Index • Others	The XAI/VA tool and FMP client presented all the information needed and the participants considered it was useful. However, they also stated that it would be preferable to also include some aggregated impact on the solutions. Additionally, the information related to flow interactions and complexity was not included in the tool.	**MET**

TAPAS ATFCM Transparency Requirement	TAPAS ATFCM Validation result	Fulfilment verdict
REQ-ATFCM-V037-BD. This information shall be displayed into a map and located where required by using diverse VA tools, along with their numerical values (in form of text). This shall help the FMP understand in a more analytical objective and transparent way the decision of the AI system to declare, or not, the hotspot to the NM.	As a map representation may be overcrowded and often confusing due to 4D nature of data (dynamics over 3D space), appropriate representation for flights in a form of sequences of crossed sectors was designed. This representation in an interactive and dynamic form is provided within the sector explorer component	**MET**
REQ-ATFCM-T038-BD. Should a set of additional parameters (e.g., those that have not been used in the building of the rule-based system for the identification of hotspots/optispots and might possibly have some relevance in the decision of declaring a hotspot to the NM. For example, parameters related to overall network load, or period of the year, etc.) have been used in the decision of the declaration, they will be provided to the FMP so that he/she can clearly understand the final decision.	Not addressed in the developed prototypes.	**UNMET**
REQ-ATFCM-T039-BD. The FMP shall be provided with explanations containing the inner reason (s) about which DCB measure (or combination of DCB measures) has (have) been selected for implementation.	All relevant parameters considered by the algorithm were provided to the FMP in real-time. Additionally, the users could consult the presented explanations from a flight perspective through the XAI/VA tool and using different drill drown functionalities.	**MET**
REQ-ATFCM-T040-BD. The FMP shall be provided with the set of the parameters that	All relevant parameters considered by the algorithm were provided to the FMP in real-time.	**MET**

TAPAS ATFCM Transparency Requirement	TAPAS ATFCM Validation result	Fulfilment verdict
influenced the most in the final decision.		
REQ-ATFCM-T041-BD. Should the AI system identified more than one DCB measure that are very similar in likelihood in terms of solving the hotspot (and so both are highly susceptible for being selected for implementation), the FMP shall be provided with explanations highlighting the differences that led it made the decision.	Differences among sets of parameters considered for alternative decisions are provided. In addition, explanations for different types of solutions are provided and these can be further elaborated	**MET**
REQ-ATFCM-T042-BD. The FMP shall be provided with a list of the flights impacted by the final selected DCB measure (or combination of DCB measures).	This information was made available through the FMP client, where the demand measure (LC, regulation), its parameters and affected flights were presented.	**MET**
REQ-ATFCM-T043-BD. The FMP shall be provided with information regarding how long the flights are to be impacted by the selected DCB measure (or combination of DCB measures).	This information was made available through the FMP client, where the demand measure (LC, regulation), its parameters and affected flights were presented.	**MET**
REQ-ATFCM-T044-BD. The FMP shall be provided with a list of the most impacted airline (s). ('Impacted' means those airlines whose aircraft are directly affected by the selected DCB measure, or combination of DCB measures).	A list on impacted airlines was not provided. Explanations were provided at a flight level, without differentiating the airline. The system does not consider information about airlines	**UNMET**
REQ-ATFCM-T045-BD. The FMP shall be provided with the most relevant parameters that influence the most in determining the ordering of the airline (s) ranking.	The system does not consider information about airlines. Ranking is not available	**UNMET**

TAPAS ATFCM Transparency Requirement	TAPAS ATFCM Validation result	Fulfilment verdict
REQ-ATFCM-T046-BD. The FMP shall be provided with a list of the most impacted sector (s) by the selected DCB measure (or combination of DCB measures).	The most impacted sectors by the selected DCB measures are provided by displaying them on a graph, rather than on a list	**MET**
REQ-ATFCM-T047-BD. The FMP shall be provided with the most relevant parameters regarding the decision of which sectors are the most impacted by the selected DCB measure (or combination of DCB measures).	FMP was provided with the most relevant parameters regarding XAI decisions proposal.	**MET**
REQ-ATFCM-T048-BD. The FMP shall be provided with explanations to the question: 'How the selected DCB measure (or combination of DCB measures) can impact or create new and unexpected future possible hotspots?	Even though the 'what-if' functionality for Level 2 allows to assess the impact of the solutions, **no explicit and clear explanations are provided** which allow getting insight on the creation of new possible hotspots.	**PARTIALLY MET**
REQ-ATFCM-T049-BD. The FMP shall be provided with an explanation confirming that the DCB measure (or combination of DCB measures) has (have) been successfully implemented in the system, as well as providing the flights impacted register.	Unresolved hotspots are not justified. If a DCB measure fails, there is no explanation on what fails nor further insights	**UNMET**
REQ-ATFCM-T051-BD. The FMP shall be provided with explanations regarding what is/are the ATFCM regulation(s) applied to any single flight.	All the explanations, and parameters, considered by the XAI were available to the participants of the trials through the XAI/VA tool.	**MET**
REQ-ATFCM-T052-BD. The FMP shall be provided with explanations regarding what is the joint ATFCM regulation(s) being applied to a subset	The main functionality (explanations) is provided, regardless how difficult is to conduct actions such as tracking the aggregation	**MET**

TAPAS ATFCM Transparency Requirement	TAPAS ATFCM Validation result	Fulfilment verdict
of flights (e.g., those crossing a specific sector within a period).		
REQ-ATFCM-T053-BD. The FMP shall be provided with explanations regarding what are the ATFCM regulations on flights satisfying specific criteria.	No specific functionality providing explanations identifying the flights that are 'most affected' by hotspots is found. However, flights in a hotspot were able to be found, and explanations' parameters include the hotspots that justify a DCB measure for a flight	PARTIALLY MET
REQ-ATFCM-T055-BD. The FMP shall be provided with explanations regarding what are the ATFCM regulations applied on sectors in different periods.	The main functionality (explanations) is provided, regardless how difficult is to conduct actions as tracking the aggregation	MET
REQ-ATFCM-T056-BD. The FMP shall be provided with explanations regarding how hotspots are linked in terms of flights crossing the corresponding sectors.	Manual steps have to be done to get the explanation / information	PARTIALLY MET
REQ-ATFCM-T057-BD. The FMP shall be provided the inner reasoning considered in support of the assessment of the criticality of a hotspot to the overall situation.	Explanations are not that specific, but the parameters for decision making include the hotspots and flights contributing to these.	PARTIALLY MET

Table 44. Traceability matrix between transparency requirements and validation results. ATFCM use case.

TAPAS CD&R Transparency Requirement	TAPAS CD&R Validation result	Fulfilment verdict
	Automation level 2 requirements	
REQ-CDR-T058-OD. The EC shall be provided with all the parameters that had impact during the evaluation step of the planned against the current trajectory profile.	The XAI/VA tool provides alerts whenever the actual trajectory deviates from the planned one through a pop-up window. The window contains information on (including the exact time when this deviation takes place): - Actual vs planned flight level - Actual vs planned speed	MET

TAPAS CD&R Transparency Requirement	TAPAS CD&R Validation result	Fulfilment verdict
	- Actual vs planned 2D route (course deviation)	
REQ-CDR-T059-OD. The EC shall be provided with simple explanations, highlighting and comparing the numerical differences in magnitude produced for a same type of parameter (e.g., cruise requested flight level vs. actual flight level, cruise requested speed vs. actual speed) between the planned and the current trajectory profile.	The XAI/VA tool provides alerts whenever the actual trajectory deviates from the planned one through a pop-up window. The window contains information on (including the exact time when this deviation takes place): - Actual vs planned flight level - Actual vs planned speed - Actual vs planned 2D route (course deviation)	**MET**
REQ-CDR-V060-OD. The EC shall be provided with appropriate visualization means in order to visually quantify the resulting gap between the planned and the current trajectory profile.	The XAI/VA did not provide any 2D/3D map for the conformance monitoring functionality. But it does provide a table with all the parameters above, allowing the ATCO to quantify the gap between planned and actual trajectory. Additionally, this req. is considered to be fully met since the map visualization is also provided to the EC through the ATC platform and radar screen, so there is no need to duplicate all these functionalities in the VA tool.	**MET**
REQ-CDR-T061-OD. The EC shall be provided with explanations informing about the possible reason (e.g., restrictions applied by upstream sectors, etc.) of the gap between the planned and the current trajectory profile.	Note that this can only be provided when the deviation is due to an ATC resolution action. Due to this reason, the VA/XAI tool provides which resolution action proposed by the algorithm is the cause behind this deviation (it shows the flightID and the particular action and value).	**MET**
REQ-CDR-T062-OD. The EC shall be provided with all the relevant parameters that might have impacted/led to the current gap between the planned and the current trajectory profile.	Note that this can only be provided when the deviation is due to an ATC resolution action. Due to this reason, the VA/XAI tool provides which resolution action proposed by the algorithm is the cause behind this deviation (it shows the flightID and the particular action and value).	**MET**
REQ-CDR-T063-OD. The EC shall be provided with an explanation describing what is the flight state (climb, cruise, descent) that has been considered in the conflict identification process (i.e.,	The XAI/VA tool provides the attitude of both flights included in the conflict through an arrow indicating if it is climbing, ascending or in the cruise phase.	**MET**

TAPAS CD&R Transparency Requirement	TAPAS CD&R Validation result	Fulfilment verdict
the flight state expected when the separation minima is breached), as well as possible corresponding parameters of importance.		
REQ-CDR-T064-OD. The EC shall be provided with an explanation describing what is the most critical type of violation and severity corresponding to the identified conflict (horizontal or vertical separation), as well as possible identified corresponding parameters (e.g., how much the horizontal separation, or the vertical separation will be violated).	The XAI/VA tool provides the separation minima achieved (horizontal and vertical), including a percentage of compliance with the separation minima MOC and severity score according to this MOC and ROC. Use of colour code (red coloured bars) to highlight the severity. An explanation that clearly describes the type of violation and the severity is exactly the presentation of these parameters, no further explanation is needed.	MET
REQ-CDR-T065-OD. The EC shall be provided with information stating if the aircraft speed has influenced or caused the identified conflict and to what extent.	The XAI/VA tool provides the rate of closure indicating how close the flights are moving to the collision point. And this is a parameter that affects the final score for severity.	MET
REQ-CDR-T066-OD. Should the EC be unaware of in which specific sector a conflict has been identified, he/she shall be provided with this information.	The name of the sector where the conflict is detected is included in the VA visualisation.	MET
REQ-CDR-T067-OD. The EC shall be provided with metrics that accurately quantify the uncertainty around the flight trajectories.	This was out of the scope for the XAI/VA prototype.	UNMET
REQ-CDR-T068-OD. The EC shall be provided with parameters related to efficiency (flight and time efficiency) along with an explanation that accurately quantifies and	The XAI/VA tool provides the added miles flown, added seconds flown and conflicts foreseen. It also includes a characterisation of the state of the flight after the solution is implemented.	MET

TAPAS CD&R Transparency Requirement	TAPAS CD&R Validation result	Fulfilment verdict
describes to what extent these parameters influence the final decision of the most effective conflict resolution strategies**.**		
REQ-CDR-T069-OD. The EC shall be provided with an explanation that accurately quantifies and describes to what extent the aircraft's ground speed influences the final decision of the most effective conflict resolution strategies.	There is no explanation on why the XAI proposes one solution over other, only through the presentation of a rank showing the likelihood of that solution to solve the problem/conflict according to its AI algorithm.	**UNMET**
REQ-CDR-T070-OD. Should weather data be available, the EC shall be provided with the parameter wind speed along with an explanation that accurately quantifies and describes to what extent this parameter influences the final decision of the most effective conflict resolution strategies.	Weather is out of the scope for this XAI/VA prototype and validation.	**UNMET**
REQ-CDR-T071-OD. The EC shall be provided with adequate metrics informing about the expected accuracy or likelihood of the proposed strategies to resolve the conflict.	There is a rank showing the likelihood of that solution to solve the problem/conflict according to its AI algorithm.	**MET**
REQ-CDR-T072-OD. The EC shall be provided with the flights impacted by the proposed conflict resolution strategies.	If a flight is impacted with the proposed resolution strategy and creates a new conflict, this is shown in the "Conflict Foreseen" column.	**MET**
REQ-CDR-T073-OD. The EC shall be provided with a quantification of the impact caused by the proposed conflict resolution strategies.	XAI/VA tool provides the impact on flight efficiency (added miles), time efficiency (added seconds, duration of the action), subsequent conflicts (foreseen conflicts) and foreseen deviations (state of the aircraft after the solution is applied).	**MET**

TAPAS CD&R Transparency Requirement	TAPAS CD&R Validation result	Fulfilment verdict
REQ-CDR-T074-OD. The EC shall be provided with adequate values informing about how long those flights will be affected by the proposed conflict resolution strategies (e.g., when the flights will resume their planned route).	The duration of the resolution action is provided. However, not how and how long it will take the aircraft to resume its FPL. During the validation exercises at level 2 it was shown that it is necessary to inform the ATCO in a more straight-forward way (not only through the indication of duration of the resolution action) when the aircraft needs to resume its FPL (e.g.: using an alarm, check list or by including additional clearances to support FPL recovery).	**PARTIALLY MET**
REQ-CDR-T075-OD. Should the resolution of a critical conflict be prioritized over another, the EC shall be provided with explanations clearly stating the reason (s) for such priority (e.g., time to conflict, conflict severity), as well as highlighting the possible consequences in case that it is not successfully resolved.	Prioritisations are done using the severity field, together with MOC and ROT fields. Using red colour bars to indicate which one of the conflicts is more critical.	**MET**
REQ-CDR-T076-OD. The EC shall be provided with all the additional information that has been used (if so) to calculate the most effective conflict resolution strategies.	Agent states, conflict information is provided, as well as state after the resolution. Simplified information since it is a safety critical situation where too much information is not useful and time consuming.	**MET**
REQ-CDR-T077-OD. The EC shall be provided with explanations describing the underlying reasons of proposing one conflict resolution over another. These shall cover aspects like the likelihood of triggering secondary conflicts and the expected benefit in terms of flight efficiency, time efficiency, ATCo workload, etc.	A rank is provided according to the likelihood of the resolution to solve the conflict according to the XAI algorithm.	**PARTIALLY MET**
REQ-CDR-T078-OD. The EC shall be provided with explanations containing insights on to what extent the proposed conflict resolution strategies manage to reconcile with	Only the duration of the resolution action is provided.	**UNMET**

TAPAS CD&R Transparency Requirement	TAPAS CD&R Validation result	Fulfilment verdict
original planned trajectory.		
REQ-CDR-T079-OD. Should the proposed conflict resolution strategies might generate additional conflicts (in-sector or downstream), the EC shall be provided with a list containing all the aircraft to be affected.	Resolution actions indicate if they create another conflict (foreseen conflicts) and by clicking on that field more information about the new conflict is provided, for instance flights affected, using a similar view to the first one.	**MET**
REQ-CDR-T080-OD. Should the proposed conflict resolution strategies might generate additional conflicts (in-sector or downstream), the EC shall be provided with explanations describing in detail the potential impact on other aircraft and in the overall traffic situation.	Resolution actions indicate if they create another conflict (foreseen conflicts) and by clicking on that field more information about the new conflict is provided, for instance flights affected, using a similar view to the first one.	**MET**
REQ-CDR-T081-OD. Should the proposed conflict resolution strategies might generate additional conflicts (in-sector or downstream), the EC shall be provided with explanations including parameters about the conditions of the conflicts occurrence (e.g., location, horizontal and vertical separation at CPA, time to CPA).	Resolution actions indicate if they create another conflict (foreseen conflicts) and by clicking on that field more information about the new conflict is provided, for instance flights affected, using a similar view to the first one.	**MET**
REQ-CDR-T082-OD. Should the proposed conflict resolution strategies might generate additional conflicts (in-sector or downstream), the EC shall be provided with their likelihood of occurrence.	Likelihood of occurrence is not provided. However, if a resolution action triggers another conflict the ATCO will be informed on this regard.	**UNMET**
REQ-CDR-T083-OD. The EC shall be provided with information describing	The FPL only changes when updating the cleared FL or WP when giving a direct. These updates/changes are directly shown in the	**MET**

TAPAS CD&R Transparency Requirement	TAPAS CD&R Validation result	Fulfilment verdict
what is the planned flight plan of the different aircraft involved in the conflict, after the potential conflict time and location.	resolution action and need to be updated in the CWP, in not a non-conformance alert will be produced.	
REQ-CDR-T084-OD. Should the proposed conflict resolution strategies might generate additional conflicts (in-sector or downstream), the EC shall be provided with explanations describing alternative solutions and actions in order to mitigate or prevent the occurrence of such conflicts.	The user can consult alternative and different solutions (depending on the situation and complexity), which imply different subsequent conflicts and impact.	**MET**
REQ-CDR-T085-OD. The EC shall be provided with explanations describing why a conflict resolution strategy has failed, as well as the expected location, horizontal and vertical separation at CPA, and time to CPA.	The prototype won't know why it failed, but it will provide the time and separation at CPA even when the collision has happened (loss as the type of conflict).	**PARTIALLY MET**
REQ-CDR-T086-OD. The EC shall be provided with explanations detailing the inner mechanism for generating the most appropriate clearance. The explanation shall give rationale about the automation process in terms of how the proposed resolution strategies and the agreed exit sector conditions influence the decision for the proposed ATC clearance.	Not provided. At the same time, it was concluded that it is not necessary in this operational phase, but in the training phase. Since in this safety critical use case, solutions are self-explanatory.	**UNMET**
REQ-CDR-T087-OD. The EC shall be provided with explanations about how the selected clearance complies with the agreed	Not provided. At the same time, it was concluded that it is not necessary in this operational phase, but in the training phase.	**UNMET**

TAPAS CD&R Transparency Requirement	TAPAS CD&R Validation result	Fulfilment verdict
restrictions and/or coordination conditions.		
REQ-CDR-V088-OD. The EC shall be provided with an appropriate visualisation of unexpected deviations.	This visualisation was available through the ATC platform and radar screen, apart from the parameters of deviation that are shown in the VA tool.	**MET**
REQ-CDR-V089-OD. The EC shall be provided with an appropriate visualisation for conformance monitoring related aspects.	This visualisation was available through the ATC platform and radar screen, apart from the parameters of deviation that are shown in the VA tool.	**MET**
REQ-CDR-V090-OD. The EC shall be provided with an appropriate visualisation of all conflicts at the time of their appearance, focusing on those that are going to be resolved.	Information of the conflict is available for the EC, including time of appearance, 4D representation (through the VA tool and radar screen).	**MET**
REQ-CDR-V091-OD. The EC shall be provided, on demand, with an appropriate 2D visualisation of the expected conflict. This visualization must represent, intuitively, the following parameters: • **Horizontal separation at CPA** • **Vertical separation at CPA** • **Time to CPA** • **A depiction of the trajectories of the different aircraft involved in the conflict, from the time of request up to the expected CPA** • **Flight levels and speeds for the**	The XAI/VA tool provides horizontal separation at CPA, vertical separation at CPA, time at CPA, depiction of the trajectories, including a 4D map, 2D + time and a vertical depiction of the conflict with flight levels of aircraft. Speed and flight levels as well as 2D map was also available through the ATC platform and radar screen.	**MET**

Co-funded by the European Union

TAPAS CD&R Transparency Requirement	TAPAS CD&R Validation result	Fulfilment verdict
different aircraft.		
REQ-CDR-T092-OD. Should the identified conflict, for which a resolution has been already proposed, be caused by a non-conformance of the ATC clearance, this information shall be provided to the EC.	The XAI/VA tool indicates if the conflict comes from another resolution action through the "Due to" column, where information on the previous resolution actions is specified.	**MET**
Automation level 3 requirements		
REQ-CDR-T093-OD. Should a flight be deviated from a given ATC clearance and any conflict (s) is (are) triggered, the EC shall be provided with explanations informing about the potential impact that such conflict (s) might cause in-sector or downstream.	If a flight deviates from its trajectory and creates a new conflict, this new conflict is detected, and new resolution actions are proposed.	**MET**
REQ-CDR-T094-OD. Should a flight be deviated from a given ATC clearance and any conflict (s) is (are) triggered, the EC shall be provided with possible actions to take.	If flight deviates from its trajectory and creates a new conflict, this new conflict is detected and new resolution actions are proposed.	**MET**

Table 45 Traceability matrix between transparency requirements and validation results. CD&R use case.

Please note that after the validation exercises were performed, and due to the feedback obtained during these exercises, a review of the requirements was made, and some requirements were refined. Please refer to D3.2 document for more information on this regard [7].

4.5.2 Traceability of the results with Operational Requirements

The following tables present a summary of the traceability between the results obtained during the exercise execution and the consolidated functional requirements developed within TAPAS project. For more details on these requirements, please consult the deliverable D2.2 TAPAS Consolidated Requirements and Functional Roadmap, Edition 00.02.00, [8]. This last document aims at providing functional requirements and a functional roadmap for the allocation of tasks between the human and machine for the ATFCM and CD&R use case.

TAPAS ATFCM Functional Requirement	TAPAS ATFCM Validation result	Fulfilment verdict
	Automation level 1 requirements	
REQ-ATFCM-F001: The Local Traffic Manager (human operator) shall be able to monitor the traffic demand in terms of Hourly Entry Counts and Occupancy Counts for the traffic volumes under his Area of Responsibility.	Through the HMI of the FMP client the operational experts were able to consult the HEC (60 min/20 min) and OCC (5 min/ 1 min) charts of the opened TVs. Also, they could select another TV that they desired to consult that same information despite the TV not been activated at the moment.	**MET**
REQ-ATFCM-F002: The Local Traffic Manager shall be able to consult the sector opening scheme for his/her Area of Responsibility.	Even though the opening scheme was not available through the tool, the users have available in another secondary screen the configurations planned during the period of time considered. Also, by default only the TVs involved in said configurations were the ones shown in the tool. Please note that no capacity measures were tackled in this specific use case, only demand measures (level capping, regulations).	**MET**
REQ-ATFCM-F003: The Local Traffic Manager shall have access to the local traffic demand monitoring values and demand and capacity imbalances alerts.	The FMP client present to the user HEC and OCC charts, along with different thresholds to consider in each one of the graphs. For instance, it represented the declared capacity (HEC Threshold) and the sustained and peak OCC value. Using this representation, the FMP was able to detect whenever some value of the graph surpassed some of the	**MET**
REQ-ATFCM-F004: The Local Traffic Manager shall be able to monitor the hotspots and optispots declared under his/her Area of Responsibility and to declare new ones.	The FMP client allow the user to declare hotspots. Once the hotspot is declared by the user in this automation level, it will appear in a dedicated table. At any time, the user will have access to all the information presented in the charts, along with the information and parameters identified in the hotspot.	**MET**
REQ-ATFCM-F005: The Local Traffic Manager shall be able to monitor the DCB measures already applied under his/her Area of Responsibility (including the flights impacted), as well as to create new ones (Flight Level Cappings and ATFCM Regulations or Mandatory Cherry-Picking Ground Delays).	The user could propose and implement demand measures (level cappings and regulations). At any given time, the user is able to monitor the situation, charts, hotspots declared, and measures implemented.	**MET**

EUROPEAN PARTNERSHIP

Co-funded by the European Union

TAPAS ATFCM Functional Requirement	TAPAS ATFCM Validation result	Fulfilment verdict
REQ-ATFCM-F006: The Local Traffic Manager shall be able to analyse the impact of the DCB measures proposed, including flights impacted, prior to the final decision on the most appropriate measures.	Some what if functionalities were available for the user to test the measures prior to their implementation. Through these functionalities the user could see the impact of the proposed measures on the different HEC and OCC charts, test different measures and all of it working against a simulation scenario to avoid any interaction with the operational one.	MET
REQ-ATFCM-F007: The Local Traffic Manager shall be able to monitor the hotspot resolution status and shall be aware of any possible deviations in the planned resolution.	Through the FMP tool the participants of the experiments could monitor all the time the status of the opened TVs, as well of the result of the implementation of their proposed measures.	MET
Automation level 2 requirements		
REQ-ATFCM-F008: The Local Traffic Manager shall be provided with the list of hotspots identified under his/her Area of Responsibility in order to discriminate whether they should be declared or not.	For automation level 2, the FMP client presented the hotspots detected by the XAI algorithm to the user. They also differentiate the hotspots declared by the XAI from those declared manually by the FMP expert, using a colour code.	MET
REQ-ATFCM-F009: The Local Traffic Manager shall be provided with the proposed list of DCB measures and flights impacted to be considered to solve the hotspots declared under his/her Area of Responsibility.	For automation level 2, the FMP client presented the DCB measures (level capping and regulations) proposed by the XAI algorithm. It also included the flights selected to be impacted by those measures.	MET
REQ-ATFCM-F010: The Local Traffic Manager shall be able to analyse the local and network performance impact of the DCB measures proposed. The following Key Performance Indicators should be at least considered when analysing the DCB measures: delay, additional distance (i.e.,	Although the impact could be inferred from the FMP and XAI/VA tools, where the information was presented from the flight perspective and delay could be extracted. During the exercise validation, the users mentioned that for them aggregated measures will be more useful. The performance indicators described in the requirement were not available as an aggregated measure for each TV and additional distance was not included.	PARTIALLY OK

TAPAS ATFCM Functional Requirement	TAPAS ATFCM Validation result	Fulfilment verdict
route extensions), and additional fuel burn.		
REQ-ATFCM-F011: The Local Traffic Manager shall be provided with hotspot resolution monitoring alerts in case there is a deviation with regards to the proposed hotspot resolution plan.	The FMP could monitor the situation at any given time. That means if any value surpasses any thresholds they will be aware of it.	**MET**
Automation level 3 requirements		
REQ-ATFCM-F011: The Local Traffic Manager shall be provided with the list of hotspots identified and declared under his/her Area of Responsibility.	For automation level 2 and level 3, the FMP client presented the hotspots detected by the XAI algorithm to the user. They also differentiate the hotspots declared by the XAI from those declared manually by the FMP expert, using a colour code.	**MET**
REQ-ATFCM-F012: The Local Traffic Manager shall be provided with the list of DCB measures and flights impacted implemented to solve the hotspots declared under his/her Area of Responsibility.	For automation level 2 and 3, the FMP client presented the DCB measures (level capping and regulations) proposed by the XAI algorithm. It also included the flights selected to be impacted by those measures. For this specific use case, those measures and flights impacted were automatically implemented, not only proposed to the user.	**MET**
REQ-ATFCM-F013: The set of DCB measures implemented to solve the hotspots declared under the Local Traffic Manager Area of Responsibility shall consider the local and network impact assessment in terms of performance (delay, additional distance, and additional fuel burn).	The implemented measures considered the impact at a network level, not only local level (as it is currently implemented). Those measures pay special attention to reducing delays, but did not considered other additional performance indicators.	**PARTIALLY MET**

Table 46 Traceability matrix between functional requirements and validation results. ATFCM use case.

Please note that although CD&R requirements for the planner controller have been developed in the framework of D2.2 work, this role was not tested during the validation process. The tests were focused on the executive controller tasks and roles.

Co-funded by the European Union

TAPAS CD&R Functional Requirement	TAPAS CD&R Validation result	Fulfilment verdict
	Automation level 1 requirements	
REQ-CDRE-F001: The executive controller shall be provided with the relevant flight information concerning the planned, desired trajectory profile (including radiofrequency and datalink requests), flight constraints and flight actual trajectory.	All this information was available through the ATC platform used during the validation exercises, which is the one currently used in the Spanish ACCs.	**MET**
REQ-CDRE-F002: The executive controller shall be provided with potential encounters between the aircraft within his/her area of responsibility for a 7/10-minute look-ahead timeframe.	The VA/XAI tool running in a secondary screen at first and included in the CWP later on during the validation exercises, presented to the air traffic controllers the detected conflicts. This tool was connected and consuming radar and flight plan information from the ATC platform and its traffic simulator, and each 30 seconds it updated said information and presented the new (if any) detected conflicts to the user.	**MET**
REQ-CDRE-F003: The executive controller shall be able to implement the ATC clearances given to the flight Cabin Crew in order to comply with the conflict resolution strategy adopted by the controller for the potential encounters and to comply with the exit coordination conditions.	The ATCO was able to implement the ATC clearances through the use of a pseudo pilot system. Once they have decided on the specific action to take, the ATCO will communicate the pseudo pilot the clearance and he will implement it through the simulating system. Later the FMP, if needed, will update the FPL through its ATC interface.	**MET**
REQ-CDRE-F004: The executive controller shall be provided with conformance monitoring alerts, indicating deviations of the flight trajectory compared to the given ATC clearances.	For automation level 1, conformance monitoring was not provided through the developed prototypes. This is a modification regarding initial roadmap on the allocation of ATCO and machine tasks.	**UNMET**
	Automation level 2 requirements	
REQ-CDRE-F005: The executive controller shall be provided with the assessment of the gap between the planned trajectory, desired	All this information was available through the ATC platform used during the validation exercises, which is the one currently used in the Spanish ACCs.	**MET**

TAPAS CD&R Functional Requirement	TAPAS CD&R Validation result	Fulfilment verdict
trajectory, and actual trajectory.		
REQ-CDRE-F006: The executive controller shall be provided with conflict resolution strategies for the encounters identified between the aircraft within his/her area of responsibility.	The VA/XAI tool running in a secondary screen at first and included in the CWP later on during the validation exercises, presented to the air traffic controllers the detected conflicts. For this automation level 2 the display also included the potential solutions to said conflicts. The user could navigate through the different proposed solutions. By default, three solutions are shown, but the user could add more solutions to the view if available.	**MET**
REQ-CDRE-F007: The executive controller shall be provided with ATC clearances implementation options to solve the identified conflicts.	Same comment as REQ-CDRE-F006.	**MET**
REQ-CDRE-F008: The executive controller shall be provided with the ATC clearances required to comply with the agreed sector exit conditions.	Exit conditions were not taken into account. However, the XAI algorithm also detected and proposed solutions for conflicts appearing the downstream sector, in the near area of the border.	**UNMET**
REQ-CDRE-F009: The executive controller shall be provided with conformance monitoring resolution actions at least for the resolution of potential conflicts triggered by the aircraft non-conformance to the given ATC clearances.	Although for automation level 2 XAI/VA tool provide alerts whenever a non conformance situation arose, no further action was recommended by the XAI. The information presented included aircraft involved, the parameters that deviates from the required action and the required values.	**PARTIALLY MET**
REQ-CDRE-F010: The proposed conflict resolution strategies shall consider, at least, the performance impact in terms of flight efficiency (horizontal and vertical), time efficiency, and fuel efficiency.	The tool provided the impact on the conflict resolution actions regarding increase on distance flown, new conflicts generated, duration of the resolution action. Fuel efficiency was not provided directly, but it could be inferred by the ATCO through the impact on additional distance flown.	**PARTIALLY MET**
Automation level 3 requirements		
REQ-CDRE-F011: The executive controller shall be provided with the list of ATC clearances	The user was provided with the ATC clearances proposed by the XAI algorithm.	**PARTIALLY MET**

TAPAS CD&R Functional Requirement	TAPAS CD&R Validation result	Fulfilment verdict
implemented to solve the identified conflicts and to comply with the agreed sector exit conditions.	But exit conditions were not taken into account. However, the XAI algorithm also detected and proposed solutions for conflicts appearing the downstream sector, in the near area of the border.	

Table 47 Traceability matrix between functional requirements and validation results. CD&R use case.

5 Conclusions and recommendations

5.1 Conclusions

Conclusions that have been extracted from both the ATFCM and CD&R validation exercises are classified in different categories in this section, depending on the subject they are related to. In this way, general conclusions on explainability, transparency, user trust/confidence, technical feasibility, and performance assessment are presented in this section.

5.1.1 Conclusions on explainability/concept maturity

The most relevant conclusions on the explainability concept are the following:

- **Rather than having explanations the user needs to trust the system.** Through the constant use of the system, especially during the training phase, the human actor was able to develop trust in the system through how it performed and the solutions it was providing. This was considered to be more valuable than the explanations provided by the support tools by the users.
- **Different levels of explainability are necessary according to the time horizon considered.**
    - **During the operation** the user indicated that they do not need to see all information or explanations related to the proposed solutions by the XAI algorithm (it also requires a time that they generally do not have). In particular, the user does not need to see the *intermediate solutions* the XAI algorithm considers and studies before it arrives to the final solution.
    - **During training.** This is the phase were more explanations on the algorithm and solutions provided by the AI are needed and were appreciated by the user, but once the approach being used was understood, users did not really interrogate this information further.
- **Presentation of aggregated information.** When providing explanations, the user wants to see information in an aggregated way, such as statistics on the impact of implementing the proposed solution, for the ATFCM use case this could also include a breakdown by region, airline operator etc. They also indicated that they would like to be able to go from a more general level to a particular one (e.g. in the case of ATFCM use case, from hotspot related information to flight related information and vice versa). However, for the CD&R use case, which is a more safety critical one, the operator would prefer to see the most relevant information only in a straightforward manner by default, as they have a limited amount of time to assess and solve the conflicts.
- **Visual aids play an important role.** Especially in the CD&R exercises which are more restricted in time and safety critical, the visual aids are important to catch the operator attention when detecting a conflict or prioritising the resolution of several conflicts according to their severity and/or immediacy.

- **Complexity of the solutions limits the capacity of the human to understand the explanations in real time.** Although the explanations are provided, in cases where the solution is too complex the human will have neither the time nor the ability to understand the solutions.

  However, as previously stated, more than having explanations the user wants to see the impact of the solution implemented/proposed by the system, which they previously got to trust in the training phase.

- **For real-time applications, where domain experts are highly qualified/experienced in the function being supported by the automation tools, in-depth or intermediate explanations are not required.** Although explanations can be provided there might not be enough time for users to be able to consult all of them (especially in the CD&R use case). In general, the suitability of solutions will lead the user to accept and gain confidence in the tools and can understand why those solutions are suitable without the need for detailed explanations. Indeed on many occasions the users repeated that *seeing solutions which work, even if they are different to those that the user may have chosen themselves is sufficient to develop trust in the system and to accept how it is performing without the need for more explanations.*

- **When multiple solutions are possible for the same issue, clear ranking of the solutions from best to worst is of great value** – this is particularly important in time-constrained, safety critical situations typically seen in the CD&R scenarios.

- **If multiple actions have the same ranking it should be clarified whether both actions must be performed or not** – sometimes two actions are too many (e.g. to communicate and monitor) and other times two actions with less impact to each recipient can be of added value in terms of capacity / efficiency / equity etc.

- **When actions are being automatically performed by the system, the status of those actions should be clearly indicated -** e.g. whether those actions are *pending*, *in progress* or *completed*.

- **For the safety critical use case of the CD&R, automation level 3 seems to be not very feasible to implement since performing a monitoring task alone may result in ATCO loss of expertise in the controlling tasks** and whenever the XAI fails (even though it will supposedly work well most of the times) the ATCOs will not have the capability to recover control in complex situations in a safe manner.

5.1.2 Conclusions on transparency

Depending on the use-case and the complexity of solutions being proposed/implemented by the automation, levels of transparency can vary.

In the ATFCM use-case, particularly for complex solutions, the traceability of explanations is key for transparency. The user needs, not only to see the final explanation of the solutions but have a clear traceability of the elements related to each measure/solution that has been applied. They prefer to see aggregated information, but also appreciate the possibility of following the thread of certain solutions down to the level of the flights to which the solution is related. This gives a clear transparency to the solutions or explanations provided, making it easier for the user to build trust in the system.

In the CD&R use-case, where solutions are more time critical, users tend to prefer that the solutions are provided in a rapid manner and in priority order if more than one solution is available. In some

EUROPEAN PARTNERSHIP

Co-funded by the European Union

cases where multiple actions are proposed to solve the same conflict, users need clear indication to distinguish these from other cases where a choice of independent solutions is provided. Additionally, if solutions replicate human behaviour, the operator will take less time to understand those (less training hours as well) and thus during automation level 2 the tool could act as a reinforcement for the operator.

However, in terms of providing additional information to help users to better understand the reason for any given choice above another, users indicated that for CD&R the importance is to provide solutions that work rather than focusing on explanations, and that this is sufficient for users to accept those proposals. In the CD&R use case, users considered that little or no additional explanatory information is needed since the combination of information already provided (usually linked to conflict characteristics) combined with a prioritisation of choices was sufficient to allow them to rapidly understand the proposals being made/implemented and the consequence of those actions.

Users did, however, consider the actions performed in the CD&R use case to be incomplete as the conflict resolution process should also include clearances to allow the traffic to recover its original flight plan. Nevertheless, the information provided via the VA display was considered to be sufficient to promote good levels of transparency.

5.1.3 Conclusions on developing user trust and confidence

In both the ATFCM and CD&R use cases, it was agreed that trust and confidence in the automation should be gained especially during the training phase. During that phase, the user needs to see that the solutions applied/proposed by the system are efficient in terms of the impact that those solutions imposed on the traffic (number of delays, hotspots solved, etc.), and that the solutions are fit for purpose.

Additionally, when dealing with disruptive solutions proposed by the XAI system (more common in the ATFCM scenario), it is important that those solutions do not vary too much when tackling the same type of problem to promote confidence in the system. In this way the user can create a mental pattern on how the system works and thus this will improve their understanding and trust on the system, acquiring said trust and confidence more quickly.

In the CD&R use case, a combination of solutions that were unrealistic, or which may have led to more complex issues further downstream, and the lack of additional clearances to help aircraft recover their original plans, contributed to a reduction in trust and confidence in the automation. However, users indicated that despite this situation, they were still able to comprehend the solutions being proposed/implemented and to rapidly understand the consequence of those actions. However, given that the TRL for the automation used in support of the experiments was low, and that the objectives were to understand the role of the associated explanatory information and how it can promote transparency and understanding for systems of this nature, the overall impact of less efficient solutions on the validation remained low.

Another important remark, made by the operational experts on several occasions during the validation, was that developing trust and confidence in a system does take a long time and relies on that system providing reliable solutions that the user accepts as being a valid response to the problem being considered. However, in the event where something subsequently fails badly, even after trust has been achieved, the confidence in that system can be lost very rapidly and rebuilding it would be a difficult task. Hence for future work, especially in the more time and safety critical CD&R domain, a heavier focus on the reliability and suitability of solutions being proposed would be beneficial.

5.1.4 Conclusions on technical feasibility

An important aspect for the technical feasibility of the introduction of greater levels of automation in the ATM domain is that the algorithms being used in the system should not be biased – especially in the ATFCM planning process. For example, the algorithms cannot systematically benefit or penalise the same airline, type of aircraft, route, etc. They should be impartial to guarantee fairness among the airspace users in this case. Even if such a bias has not been explicitly implemented, it should be ensured that the algorithm has no unintentional bias (this can be a risk with systems based on learning and therefore corrective actions during the training process are necessary to avoid it).

For the CD&R use-case, solutions only focused on actions that would resolve the conflict that was detected, and proposed actions were limited to the establishment of suitable separation in accordance with the minima required for the region in which the aircraft are operating. However, additional actions that are also considered to be an integral and necessary part of the conflict resolution process, such as recovery of the original flight plan, were not included in the automation. Additionally, other aspects of the conflict resolution process such as directional flight level strategies, aircraft types/performance characteristics, proximity to departure/arrival airport etc. were not necessarily incorporated into the automation tool. Hence some of the resolutions actions that were proposed were inconsistent with actual practices in the sectors that were simulated, which could lead to a reduction in confidence in the tool.

To guarantee the technical feasibility of this concept, the tools used in the ATFCM scenarios should be more closely integrated with the planning platform, consuming the same real time information to ensure data consistency. Also, in the actual operating environment and tactical/pre-tactical phase, the developed tools need to consider multiple parameters (weather, flight updates, changes in the opened scheme, etc.) so that the proposed solution is efficient and feasible. By using a loosely integrated solution for ATFCM, the user was not able to select a subset of proposed solutions, implement them using a *what-if* simulation to obtain a partial solution, then re-consult the XAI to evaluate the resulting situation. This was a particular issue when offering level 2 solutions.

In the CD&R scenario, the tools were integrated more closely, and information was exchanged using messaging protocols to allow the XAI to identify conflicts, make decisions and develop solutions in an interoperable manner. This allowed the tools to be used in a more realistic, real-time mode and to respond to unexpected conflicts which may have resulted due to other ATC or flight deck actions. However, better training and use of other constraining factors such as those mentioned previously in this section would help improve the operational and technical feasibility of the CD&R automation in support of the ATC process.

5.1.5 Conclusions on Key Performance Areas (KPA)

As the main objective of the ATFCM validation focuses on the ability of users *to understand issues that have been identified by the XAI component* and *to comprehend the solutions that were being proposed or automatically implemented*, the analysis of KPA related to the *Efficiency* of the solutions were not of lesser importance – that is to say, users were not requested to evaluate whether the solution that was applied (e.g. delays to flight plans) was minimal/optimal, but to confirm that the solution which was chosen was able to be explained in a way that could be easily understood.

Nevertheless, metrics relating to the results of the ATFCM exercises were able to be obtained which illustrate the suitability and efficiency of the solution(s) being proposed.

EUROPEAN PARTNERSHIP

Co-funded by the European Union

Efficiency:

***Level 1 automation**:*

The FMP users were asked to identify overload issues using the flight count graphs in the FMP Client and create Hotspot(s) as considered appropriate.

To solve the issue, corresponding Regulations were then manually created via the interface, and based on the information provided in the corresponding flight lists (available by selecting any given count bar to detail the contained flights) specific flights were added to the regulation using ‘Cherry Picking’ to spread the overload into the lower demand periods following the Hotspot through the application of flight delays at the departure airport.

User experience helped in the identification of suitable flights and the definition of Hotspots and the appropriate Regulation to apply.

To test the solution, users worked in a SIMULATION dataset (shown in light blue below) to perform what-if testing of the regulation and cherry-picked flights.

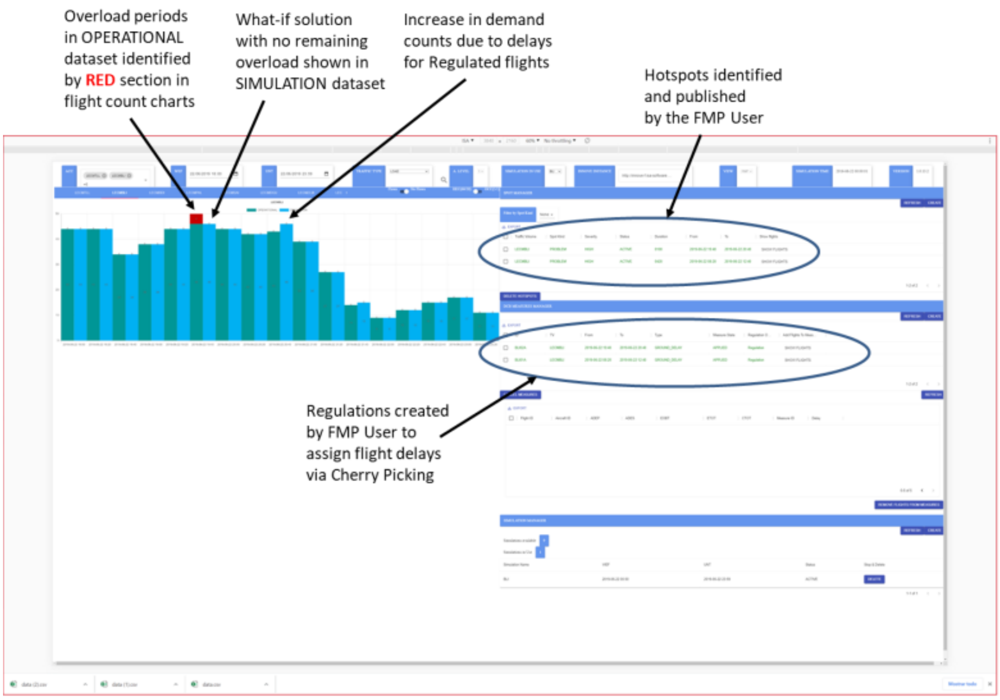


Figure 30. Manual use of the FMP client to Regulate Traffic in SIMULATION dataset

Solutions performed during the level 1 exercises were limited to only consider small overload periods since when larger problems that span several demand count periods it is usual to simply define an extended regulation period which is then solved by the CASA algorithm at NM. Attempting to solve this type of issue manually is too challenging.

Overall, in solving problems using Regulations for the airspace sectors identified in the Level 1 scenarios proved to be efficient in terms of the delays to flights that were applied to remove the Hotspots remained efficient ranging between 3 and 14 minutes with an average delay per flight around 5.8 minutes.

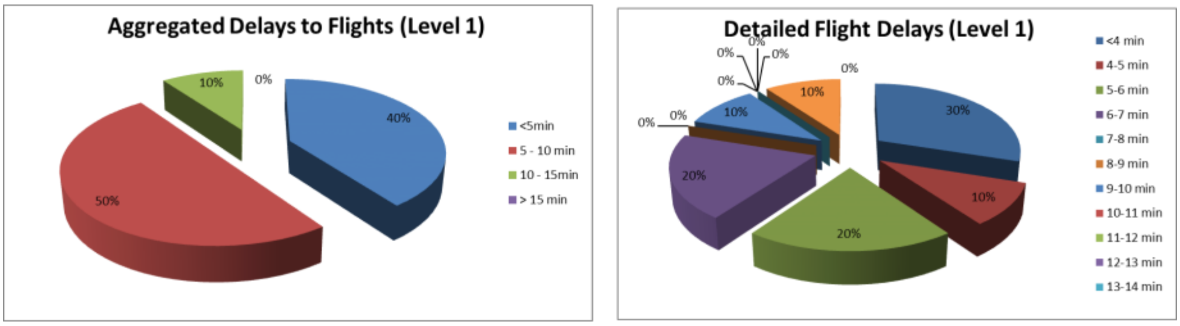


Figure 31. Flight Delays due to FMP Regulation Actions (Level 1)

Level 2 Automation

When running with Level 2 automation, the Hotspots identified by the XAI were automatically adapted to be compatible with the NM definition of a Hotspot and were then published to INNOVE without the need for any user action. Users could then interrogate the sectors with Hotspots and choose from a series of proposed Regulations and flights that had been selected by XAI to allocate delays to solve the selected overloads.

At Level 2, users were able to also solve more complex Hotspots including those which spanned several demand count periods. As a consequence of the more complex issues being solved, overall delays to traffic were higher, ranging from 2 minutes to 46 minutes with an average delay per flight around 14.3 minutes. This was not surprising as solutions also included the larger Hotspots as mentioned previously.

Nevertheless, when considering the aggregated delays to traffic shown below, 61% received 10 minutes or less delay using the proposed solutions from XAI.

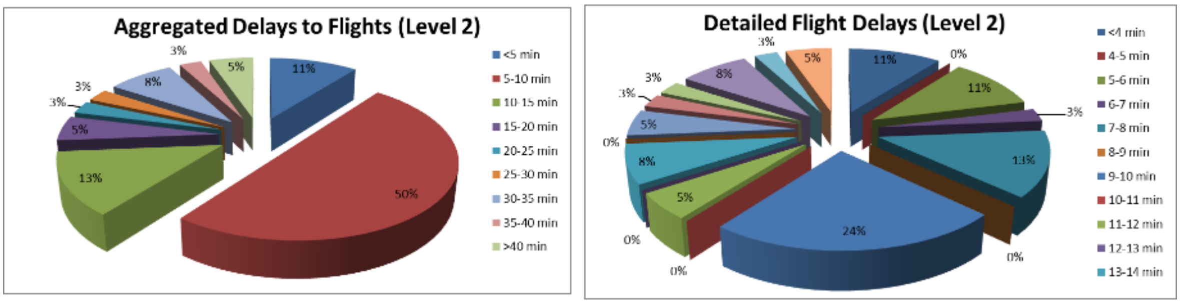


Figure 32. Flight Delays due to XAI Proposed Actions selected by FMP (Level 2)

Level 3 Automation

When running with Level 3 automation, Hotspots are automatically identified by the XAI and are adapted to be compatible with the NM definition of a Hotspot and Regulations plus the associated flight delays are also automatically assigned in the platform. In this way, all of the issues that have been identified in the region are 'solved' by the XAI tool.

However, it is important to recall that, unlike the current approach used by regional FMP/LTM operators, who tend to solve issues 'one by one', the XAI tool attempts to solve all of the issues from the start to the end of the day 'for the entire region' (and not just the Madrid region). This can result in delays to traffic which are a consequence of 'induced' Hotspots which are created by previous solutions that have been applied for earlier problems. These issues can be difficult to identify manually (or even at level 2) using the one-by-one approach or if identified the development of solutions is potentially very time consuming without automated support.

With the XAI automation in place, many more hotspots were able to be identified and solved when running at Level 3 automation than was possible at the lower automation levels. Additionally, on the final day of the validation, it was possible to run two different exercises for traffic days at level 3, allowing more results to be gathered than for the other lower automation scenarios.

For a more complete summary of all the XAI results derived from level 3 please refer to the results presented in Appendix B at the end of the report.

For the 2019-07-04 traffic sample, shown on the left below, around 56% of the delays needed to solve the identified Hotspots required 15 minutes or less. However, 33% of the delays required 100 minutes or more to resolve the issues.

Using the 2019-07-14 traffic, shown on the right, 84% of the delays applied remained at 15 minutes or less, with only 2% needing 100 minutes or higher.

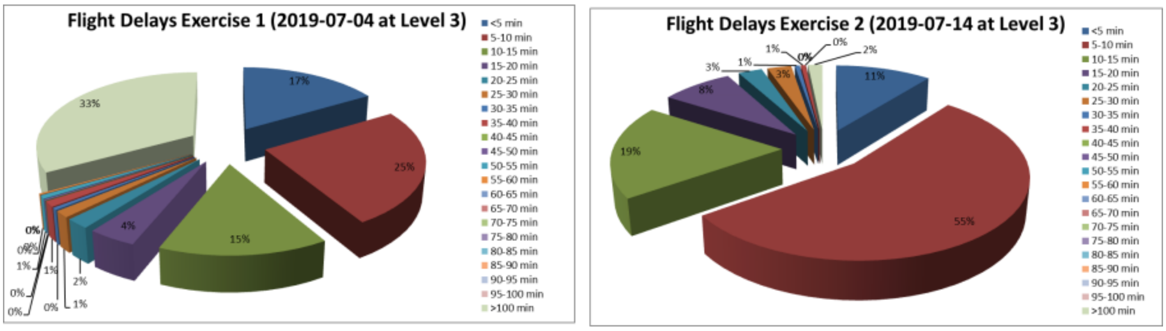


Figure 33. Flight Delays Created from XAI for Level 3 scenarios

Additional analysis of the flights which received very high levels of delay revealed that these were mainly linked to the solution of a very heavily loaded airspace [CJI] in the Madrid ACC in the July 4th exercise.

The demand seen in CJI was heavily overloaded for the entire core period of the day as shown below.

In reality, Regulation would not be applied in such a case, where typically such an overload would result in a sector reconfiguration action, e.g. to split the CJI sector into two separately controlled areas in order to share the overload between to ATC positions.

As this solution was not available in the validation experiments, results without those from the CJI sector were also considered (to reduce the number of very high delays) as shown on the right below.

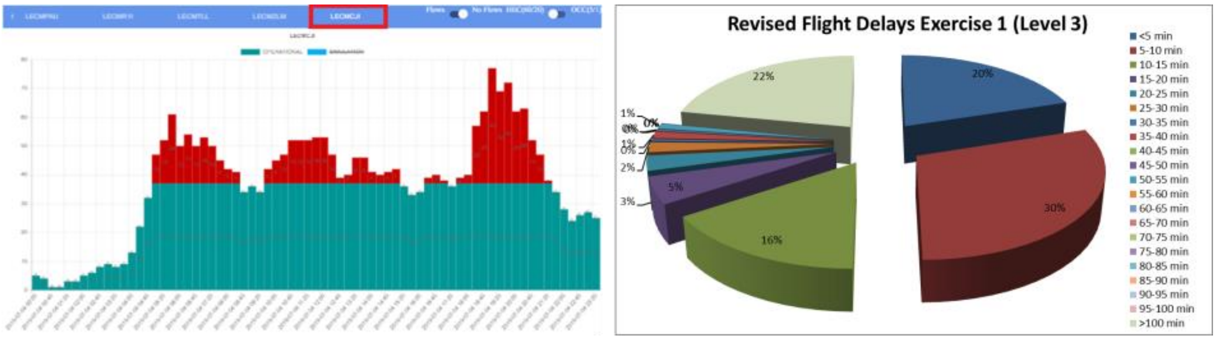


Figure 34. Flight Delays (Level 3) with heavily overloaded sector removed

From the CD&R perspective, exercises that were performed were not specifically designed to *test the performance of the conflict resolution tools*. In this use case, the focus was on the provision of *suitable* solutions in a short amount of time, with a sufficient level of additional information to ensure that the user could either select and apply them or understand what had been automatically applied with a good level of situational awareness.

Once applied the user could rapidly move on to the next problem that was identified while also continuing to support the other ATC services typically provided by the Radar/Executive controller.

For this reason, detailed KPA relating to the efficiency of CD&R solutions were not included as part of the simulation exercises.

Human Performance and Safety:

Conclusions regarding Human Performance and Safety were obtained during the ATFCM validation exercises execution mainly through 'over the shoulder' observation or during the debriefing sessions.

In general, the view of the operational experts was that applying the solutions that were being proposed at level 2 was straight-forward but issues relating to the use of 'what-if' tests and selection of only a sub-set of solutions, combined with the inability to then rerun the XAI to try to solve the remaining issues made use of level 2 more difficult.

Nevertheless, with the additional information available via the co-located VA display and its drill down features, the issues identified, and solutions being proposed could be understood with some effort on the part of the user.

At automation level 3, many more solutions were able to be applied in a very short execution time, and the VA support provided plenty of additional information to help users to understand what actions had been taken. Some additional effort was still required to comprehend the effect of 'induced' Hotspots, created by previous actions, but in general, users were able to consider many more solutions and maintain good levels of Situational Awareness even though those solutions had been 'automatically' applied.

This led the users to consider that the solutions proposed were sufficiently Safe (noting, of course, that Safety when performing a pre-operation planning process in the D-1 time frame is less critical).

With regard to Human Performance, questionnaire responses analysed previously confirm that all the users felt that Situational Awareness was maintained at all levels of automation (with the support of the VA tools), that no 'unexpected events' occurred and through 'over the shoulder' observation, it was clear that the FMP users were able to perform the requested function at all levels of automation in a reasonable time and with an acceptable level of workload.

Hence from the ATFCM perspective, the conclusions from the KPA point of view would be that both Human Performance and Safety KPI were at an acceptable level (qualitatively) for all levels of automation (with the exception of solving complex overloads using a manual approach at level 1 instead of delegating the task to the NM CASA tool).

In the CD&R scenarios, Human Performance and Safety indicators were considered using a specific set of questions relating to both the effort needed to understand what solutions were proposed to help solve ATC conflicts and the consequence that those solutions would have on the aircraft involved. Additional questions to help to determine the level of situational awareness that operators were able to achieve during level 2 and level 3 scenarios as well as the perceived workload during the execution of the scenarios were used – including a series of qualitative responses that were elicited from users about the level of workload they experienced during the runs (using a Bedford scale as shown below).

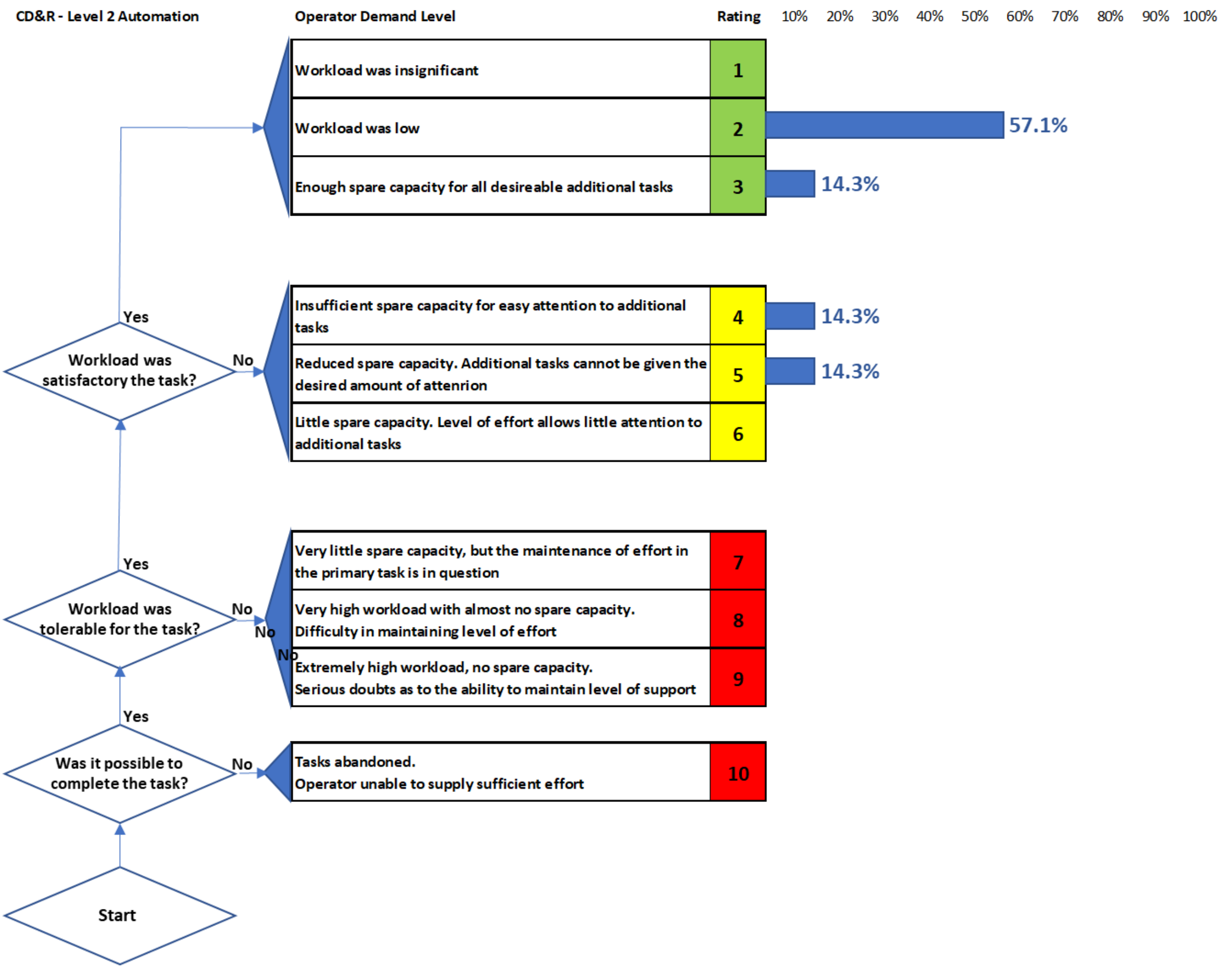


Figure 35: CD&R Workload - Level 2 Automation

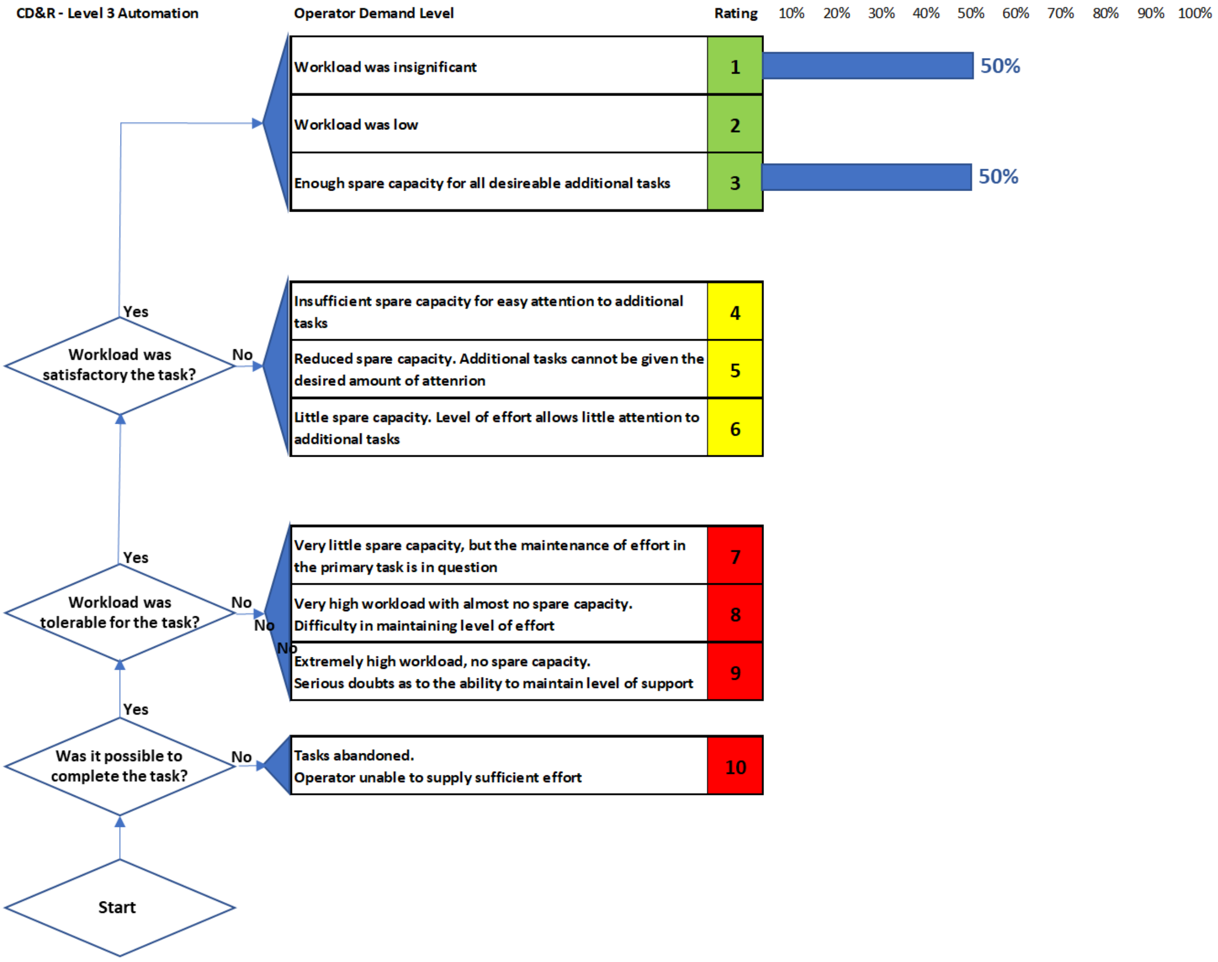


Figure 36 CD&R Workload - Level 3 Automation

5.2 Recommendations

Following the first round of ATFCM validation exercises, recommendations were identified that could be helpful during the preparation of the CD&R use case.

For the CD&R use case, the nature of the conflicts to be resolved are flight-centric therefore explanatory information should also be flight(s) oriented. This means that the explanations provided to support the XAI decisions and recommendations should follow the same flight-centric logic and must be focused on particular flights. This was seen during the validations, as the VA/XAI tool presented the information from the flight perspective (including conflict, time of conflict, aircraft involved, among other features). ATCOs stated throughout the experiments that the information presented was useful and suitably presented to quickly comprehend the situation and to act accordingly. Consequently, this type of view in the tool was the desirable one. Please refer to the results summarised in sections 4.3.2.4 and 4.3.2.6. to consult further details on these results.

However, from the ATFCM experiments it was also concluded that *having aggregated data,* as well as the possibility to have drill down views, was considered very useful by the user and should be considered in the next CD&R use case if possible. As described in the results section 4.3.2.5, the users mentioned that these aggregated measures would be a very useful feature to add to the XAI/VA tool especially regarding the impact of the proposed measures.

In addition, since CD&R is a more safety-critical use case, the *explanations must be significantly easier and quicker to understand*, and it is important for the user to understand the impact of the actions to be applied, as it is more immediate and safety-critical than for the ATFCM use case.

For these reasons, for the CD&R use case the explanations provided should include all the necessary information but structured in a simple way - i.e. *Relevant information must be there* and the *access to it must be clear*.

These two last recommendations were extracted directly from the feedback obtained during the trials, specifically these comments are described in results analysis sections 4.3.2.2 and 4.3.3.2.

Having completed the second validation use-case, with a focus on the need to provide rapid and easy to understand solutions in the CD&R domain, the following recommendations were also identified/reinforced:

- Information should be presented in the simplest and clearest way possible. This should include, by default, only the most relevant information. Operational experts participating in the exercise declared the information presented should be simple and in a straightforward way since they do not have the time to consult the tool in much detail. Please refer to sections 4.3.2.2 and 4.3.3.2.

- Extensive use of visual aids that are available (different colours and other ways to highlight the information) plays an important role in this safety critical use case. The tool used was found useful because it included, along with the ATC platform used, some of these visual features (see section 4.3.3.2).

- Integrating XAI/VA tools with the platforms used is key (refer to sections 4.3.2.2, 4.3.2.4 and 4.3.3.2 for further details on the feedback provided regarding XAI/VA integration). This included

    - Interoperable consumption/exchange of data between the platform and the automation/visualisation components,

    - Update/refresh rates should be carefully designed to avoid confusion or time wasting on the part of the operators when actions had already been taken. This could also be improved using either status information (regarding the actions that are in progress or required) and/or a series of time-based steps that need to be followed by the operator.

    - Integration of visual components in the same display if safety critical views are used by the operator. For example, in the CD&R use case the integration of the VA information with the CWP radar screen during some of the later exercises allowed the operator to maintain their situational awareness and to avoiding unnecessary deviations from the radar display.

- While providing explanations on the proposed solutions was seen to be valuable in both exercises, users considered it was more important for those solutions to be feasible and effective, especially in the CD&R use case (sections 4.3.3.7 and 4.3.3.8).

- During the execution of the validation exercises some constraints were identified which were linked to organisational or logistic based issues and that did not have a major impact on the achieved results. Recommendations how to avoid impacts from such constraints in the future include:

    - Due to restrictions on travel resulting from the COVID pandemic, validation exercises were carried out in a hybrid manner, with a limited amount of *screen-sharing* offering some insight to remote participants (e.g. the developers of the XAI/VA tools, other interested partners etc.) on how the exercises were progressing. While this offered those remote partners the chance to be involved, it did not allow them to observe things that were happening/being discussed during the execution of scenarios in the same way as was possible for those partners who were in the exercise laboratory. If the need to perform such simulations in a hybrid manner were to re-occur it may be advantageous to also include more advanced technical solutions for sharing the experience, such as live video streaming or other multi-media techniques.
    - During the post-exercise de-brief sessions, it would have been more time-effective to gather all the participants in a separate meeting room where a discussion could be held regarding the previously run exercises and the questionnaires could be filled out with exercises organisers monitoring that all questions were being considered and were correctly understood by the users. This was particularly evident during the CD&R exercises, which were further impacted by the fact that CRIDA, who hosted the experiments, were in the middle of moving premises and only the simulation lab and a small adjacent room were available. This resulted in users completing their questionnaires in isolation and in some cases those responses were incomplete.
    - When designing questionnaires, the questions that are presented should be set in the same context (e.g. agree is good, disagree is bad) – while this was not a blocking issue, it did require additional post-processing and interpretation of responses to allow them to be compared between one another.
- In situations where the users were already expert in the specific domain (e.g. CD&R) and timely responses which work are essential for safety reasons, users indicated that once a solution is proposed, the situation is binary – meaning it either solved the problem or it did not. The time needed to solve the problem was considered less critical by users than the ability for the solution to work as needed. Furthermore, when working in a real-time safety critical application, partial solutions which subsequently require additional actions to complete the solution were of less value to users than those which solved the problem immediately then allowed them to move on to the next task in their list (section 4.3.3.8).
- In the CD&R domain, users expressed concern regarding the deployment of *full automation* preferring instead to envisage automation as tools that support human focused solutions such as those seen at level 2 automation in the TAPAS scenarios (consult analysis results sections 4.3.3.12 and 4.3.3.14).
- Higher levels of automation could lead to the '*de-skilling*' of the operators in quite short periods of time following deployment. In this scenario, the risk is that after several months of a *major paradigm change*, and/or for traffic situations beyond a certain threshold/complexity, the human would no longer be able to step back in and regain control. This represents a *disruptive change* to the existing ATC/ATM process and would require that an entirely new set

systems and/or processes are developed to provide a safe and reliable back-up. (please refer to sections 4.3.3.12 and 4.3.3.14).

The last two recommendations were pointed out during the execution of automation level 3 scenarios that involved some type of ATCO recovery control. Therefore, directly related to objectives 2.6 and 2.7. Although these two objectives were not tested in depth, and only partially, initial results already highlighted that this automation level 3 along with the roadmap of tasks allocation, that is, with the human in the loop, seems, at least for the moment, not feasible for this safety critical use case. Additionally, in the ATFCM use case, mainly due to time and resources constraints, it was not possible to test these non-nominal conditions. Therefore, further research on this topic is still recommended as a continuation of the work performed by TAPAS project.

Additionally, in the ATFCM use case, only demand measures were tested, as the capacity ones were not included. However, it should be notice that the project’s objective focused on transparency and explainability requirements. Consequently, how this use case was conceived was sufficient to achieve this purpose. Nonetheless, it is recommended to continue the research on this topic in future projects, in a potential TAPAS continuation or in any other project within and outside the SESAR 3 framework that also tackles the DCB process.

Finally, please refer to TAPAS deliverable D6.3 Final Project Results Report [6] for further details on the recommended steps about the future research or continuation of TAPAS work.

6 References

[1] ED-78A GUIDELINES FOR APPROVAL OF THE PROVISION AND USE OF AIR TRAFFIC SERVICES SUPPORTED BY DATA COMMUNICATIONS.

[2] https://www.eurocontrol.int/simulator/innove

[3] D5.1 TAPAS Validation Plan, Edition 00.01.00, May 2021

[4] D3.1 TAPAS Use Cases Transparency Requirements, Edition 00.01.03

[5] D4.3 TAPAS Visualizations and Visual Analytics methods, Edition 00.01.00.

[6] D6.3 Final Project Results Report, Edition 00.01.01.

[7] D3.2 Principles for Transparency in AI-ML automation in ATM, Edition 00.03.00.

[8] D2.2 TAPAS Consolidated Requirements and Functional Roadmap, Edition 00.02.00

Appendix A ATFCM Validation Platform

To provide a realistic working environment in which the operational experts could work, the experiments were performed using the INNOVE simulation platform. INNOVE is a fully interactive simulation and gaming platform that emulates the SWIM B2B services that are provided by the Network Manager to support ATFCM for all flight operations across Europe. Additionally, INNOVE provides prototype services that have been specifically developed to test new functions and services in support of on-going SESAR research, including Dynamic Airspace Configuration Management (DAC), Free-Route Airspace (FRA), Dynamic Military Areas (DMA), Flight Prioritisation, User Defined Prioritisation Process (UDPP) and many others.

INNOVE has built in collaboration with EUROCONTROL using an existing ATFCM simulation product, CHILL, which was originally developed to support network management simulations for the FAA. The CHILL product was adapted to deploy it as a cloud-based full stack solution, supporting all of the services offered by NM, including all of the SOAP based B2B services, Publish-subscribe services implemented using AMQP and a set of bespoke REST services to support simulation management, synchronisation and control functionality.

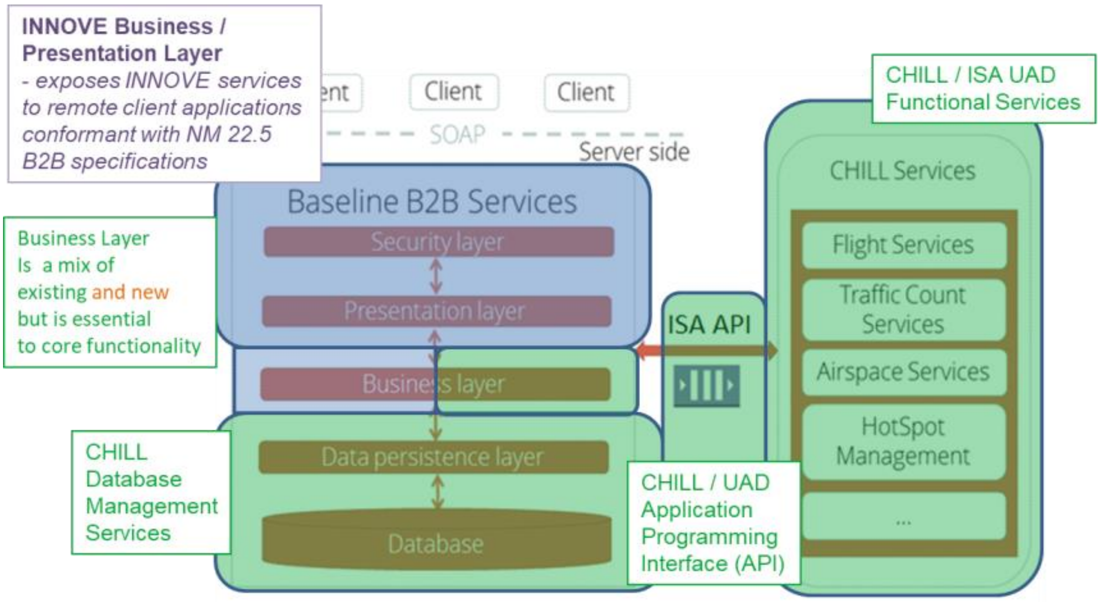


Figure 37. INNOVE Gaming Platform Architecture & Services

Using this architecture, INNOVE is able to provide external users and client applications with an 'identical' service-based environment with API that conform to the 'real-world' NM platform against which it has been built.

The advantage of using INNOVE is that specific scenarios can be captured using a combination of environment data extracted from the EUROCONTROL Demand Data Repository (DDR2) or NEST tool and traffic provided using NM ALLFT+ format. This can then be replayed as often as required.

In the case of the TAPAS experiments, the INNOVE release that was compatible with the NM 24.0 release from June 2020 was deployed using the scenario data described previously to support the exercises described in section **¡Error! No se encuentra el origen de la referencia.**.

To support the FMP/LTM user, a dedicated FMP client application was designed and implemented that provided most of the same functions that are commonly available for FMP when performing Demand-Capacity Management activities.

The FMP client was used during all of the different experimental scenarios to provide interactive demand and occupancy information for any airspace in the Spanish region, although the users were requested to focus on airspace in the Madrid ACC. With this in mind, the application also provided a set of rapid access menus to access and display key information related to Madrid sectors.

Users were also able to identify Hotspots while interacting with the demand/occupancy charts and to manually publish them in the INNOVE platform (level 1 automation) using dedicated functions in the FMP client interface. Regulations and ATFCM measures, including Flight Level Capping could also be created and published by the user via the FMP client application interface. Once created, the impacted flights were updated in the INNOVE scenario and corresponding charts were automatically updated to display the impact of the proposed user actions.

INNOVE, and the FMP client, also allows users to create SIMULATION snapshots of the current active scenario, in the same way that NM supports the SIMULATION dataset. Once created, SIMULATIONS act as sandbox copies of the current situation to allow users to perform what-if analysis of potential actions to see how they would affect the current situation, without actually changing the real traffic situation. Once a user is happy with the outcome of a what-if scenario, the same actions can be published in the real OPERATIONAL scenario as the proposed solution to a given DCB problem.

In order to support higher levels of automation during the experiments, the FMP client was also adapted to allow it to consume data related to traffic demand, Hotspots and proposed ATFCM solutions produced by the XAI component. In this way, the operational users were able to view the information provided by the XAI in exactly the same format as the operational data held in the INNOVE platform, and to utilise all of the interactive charting and flight list features supported by the tool. Additionally, as the paradigm used by the XAI to represent Hotspots and create Regulations was slightly different to that used by NM (e.g. Hotspots in the XAI were declared for any 1-hour period that had <10% overload, whereas in NM a Hotspot can be a longer duration than 1-hour and covers the entire period that an Airspace/Traffic Volume is >100% loaded) the FMP client include algorithms to combine the XAI Hotspots and Regulations and present them using the NM definitions.

To support the explainability and transparency needs for the experiment, a second, co-located display was included in the experimental platform. This additional display was dedicated to the Visual Analytics component. The information provided through the VA display had been carefully designed to support explainability and transparency through a series of interactive views that the user could work with to help understand how Hotspots had been identified and why certain solutions were being proposed.

The VA components were also able to consume data produced by the XAI component to support the various display functions that were supported in the different views.

All of the components in the experimental platform were loosely integrated, with limited interactions based mainly on the use of common data to support the different functionality as illustrated in the overview of the connected platform shown below:

Co-funded by the European Union

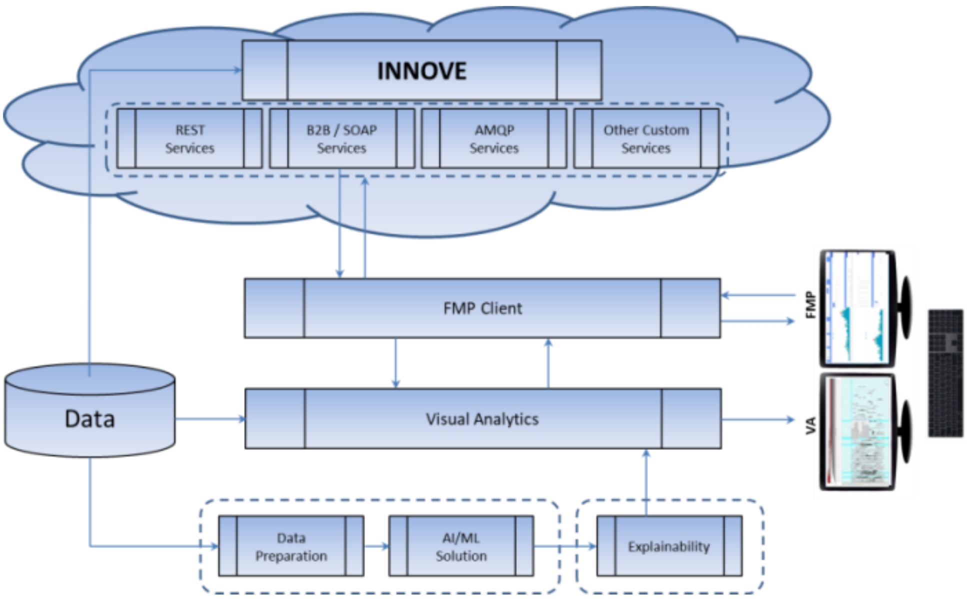


Figure 38. ATFCM Validation Infrastructure

The FMP client provided an operational tool to display the current network situation and allow the user to input new ATFCM actions (e.g., ATFCM regulations or re-routing measures) into the INNOVE simulation in accordance with the level of automation being emulated in the scenario. Functionality included:

- Provide sector / traffic volume load and occupancy charts for any of the ACC in the region
- Support interactive consultation of traffic lists for the flights which make up any of the charted periods
- Identify overloads on the load/occupancy charts in regard to the declared capacity thresholds
- Display information about hotspots identified by the XAI tool
- Allow users to create the same hotspots in INNOVE via the FMP client interface
- Automatically create Hotspots in the INNOVE environment based on data received from the XAI component
    - Without user intervention
- Display proposed solutions to traffic for each Hotspot
    - Delay only
    - FLC only
    - Delay and FLC combined
- Allow users to create a SIMULATION (what-if ‘sandbox’ environment) in the INNOVE scenario
- Allow users to create regulations or measures to apply with the proposed solutions in the INNOVE platform

  - In a SIMULATED 'sandbox' environment to support what-if analysis, or
  - In the OPERATIONAL environment to create the actual measure(s)
- Automatically create measures in the INNOVE scenario using data provided by the XAI component
  - In a SIMULATED 'sandbox' environment to support what-if analysis, or
  - In the OPERATIONAL environment to automatically create the actual solutions
- Perform analysis of the network impact of the implemented/simulated measures
- Record information to support post validation analysis and metrics generation

The experimental VA support information displayed on a second, co-located screen, provided interactive graphical support to assist the user in developing a clear understanding of the criteria and features that were considered by the XAI automation during the decision making process by exploring the data.

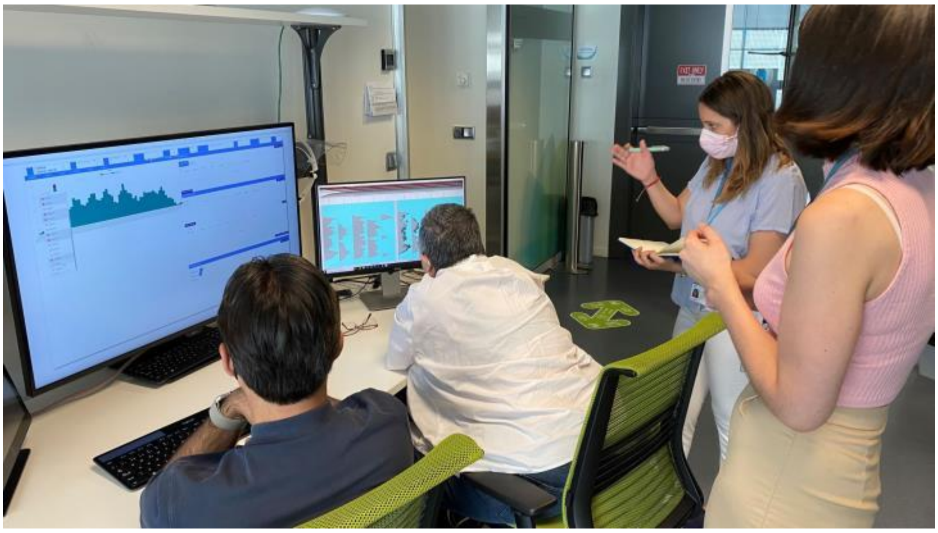


Figure 39. TAPAS Integrated FMP Working Position

Features of the interactive VA display included:

- Analysis of the DCB scenarios being analysed by the AI components
- Visual summaries for the variants of solutions
- Display of the evolution of the solutions over iteration steps of the simulation process
- Provision of details for each hotspot, sector, and time interval, including aggregated information about the flight delays
- Support for on demand: features that justified delay decisions for selected flights, sectors, and time periods
- Display of further relevant information, to be selected in collaboration with project partners.

The data that is supported on the VA screen was also be tailored according to the level of automation being emulated in the scenario and at different levels of automation this included:

- Exploration of baseline scenarios with Hotspots identified but no solutions implemented
- Exploration of scenarios for delay-based solutions only
- Exploration of scenarios for flight level capping solutions only

Exploration of scenarios for combined delay and flight level capping solutions. To emulate different levels of automation, specific features were built into the FMP client tool to automatically carry out actions that had been identified by the XAI with partial or no user intervention, according to the level of automation that had been selected for the specific exercise:

In the Level 1 exercise, the automation identified DCB issues and provided information related to potential hotspots (without publishing them – shown in red below), but no recommendations on how to solve the issues were provided.

Figure 40. FMP Client Showing Potential Hotspots

The Level 2 exercises used the XAI automation to identify hotspots and declare them automatically in INNOVE without user action. Solutions based on either Regulation or Reroute using Flight Level capping were also provided by the XAI and shown in the FMP client interface, but were not published to INNOVE unless the user chose to do so (manually). In this way the human was still be involved in the implementation of the proposed measures, and was able to determine the impact and suitability of the proposed solutions by using a SIMULATION dataset to perform 'what-if' analysis.

At level 2, users were also able to consult the VA display to acquire additional information and insights into what the solution was proposing and why/how the decision had been made.

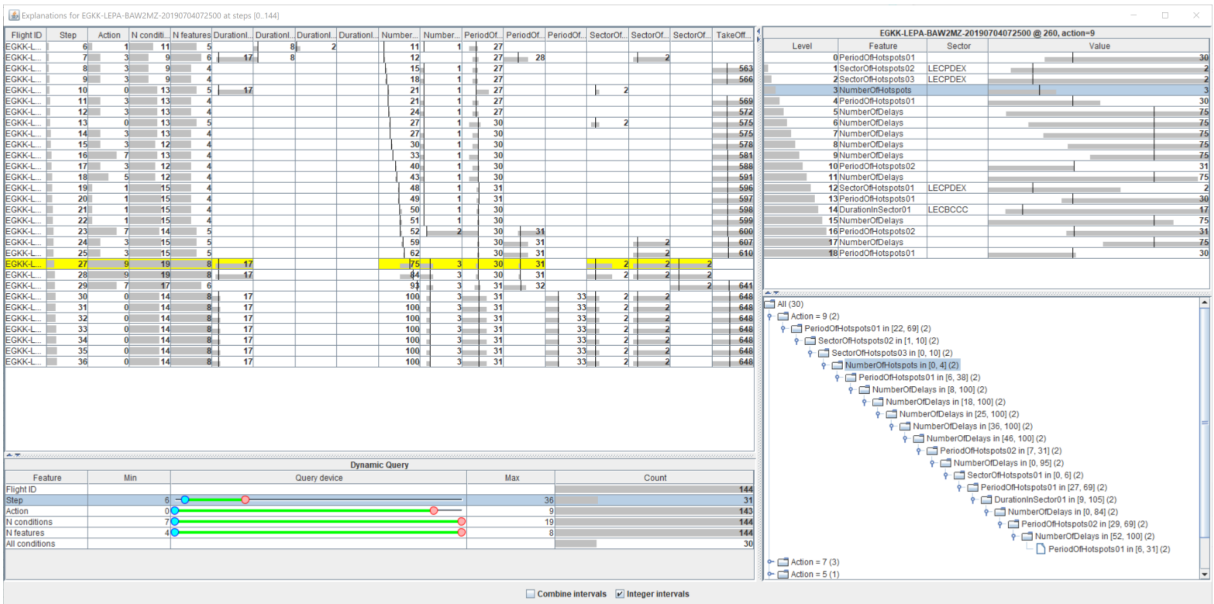


Figure 41. VA display showing the explanations provided at a flight level.

For the Level 3 exercises, all of the proposed solutions produced by the XAI were automatically published to INNOVE and implemented directly in the OPERATIONAL dataset. At Level 3, these solutions were either Regulation only, Flight Level Capping only or a combination of both according to the exercise being performed.

Once published, the users were able to interrogate the outcome using the various charts provided by the FMP client application as well as to search for additional explanations on what the automation had implemented and the reasoning behind that choice, using the extended features available in the VA Display component.

Furthermore, as the users gained more experience in using the platform, they were able to use the advanced features in the VA tool to drill down into solutions and gain better insights to help explain how and why a given problem had been solved.

Appendix B Detailed Results from the ATFCM XAI

This section provides more details on the results obtained from the Level 3 ('full automation') exercises where the XAI was fully responsible for the identification of Hotspots and the proposed mitigation actions.

It should be also noted that the approach that was adopted by the XAI was to solve 'all of the problems encountered' for the entire Iberian Peninsula and not just those concerning the Madrid ACC (as was the case for the exercises that were performed at Levels 1 & 2).

For this reason, it is not a surprise that the numbers of Hotspots (and induced Hotspots coming from previous problem solutions) is considerable higher than for the other experiments performed at lower automation levels and focussed only on Madrid ACC sectors.

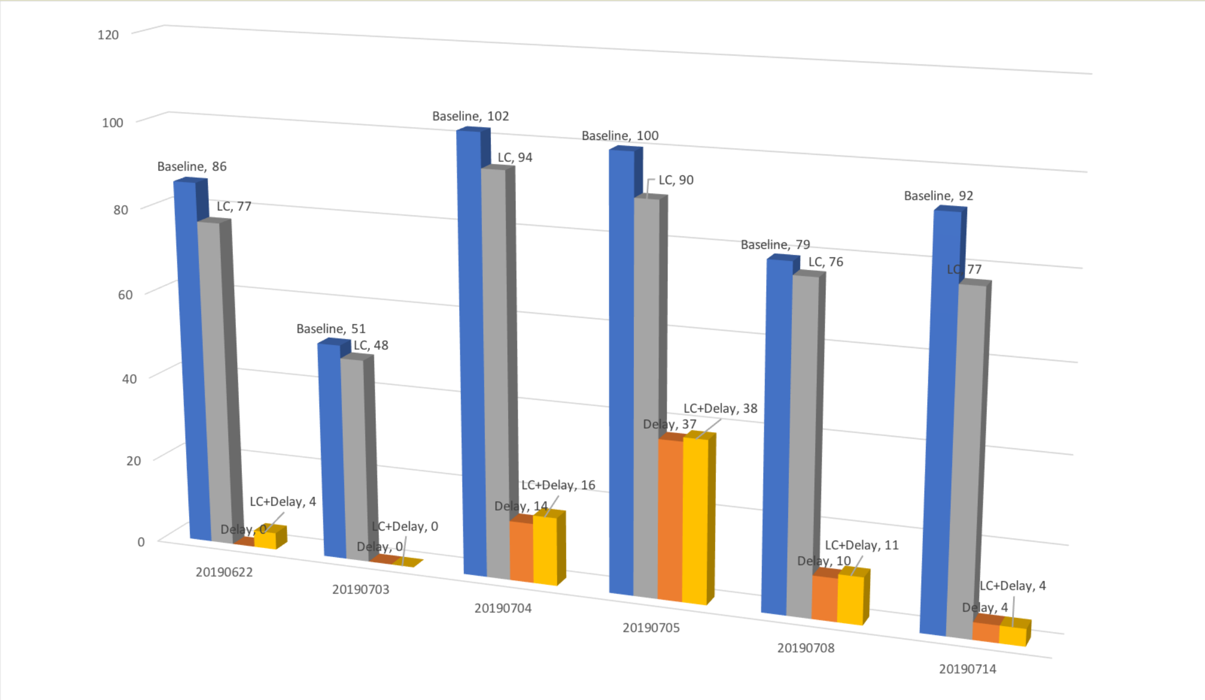


Figure 42. XAI Results from validation scenarios (Hotspots)

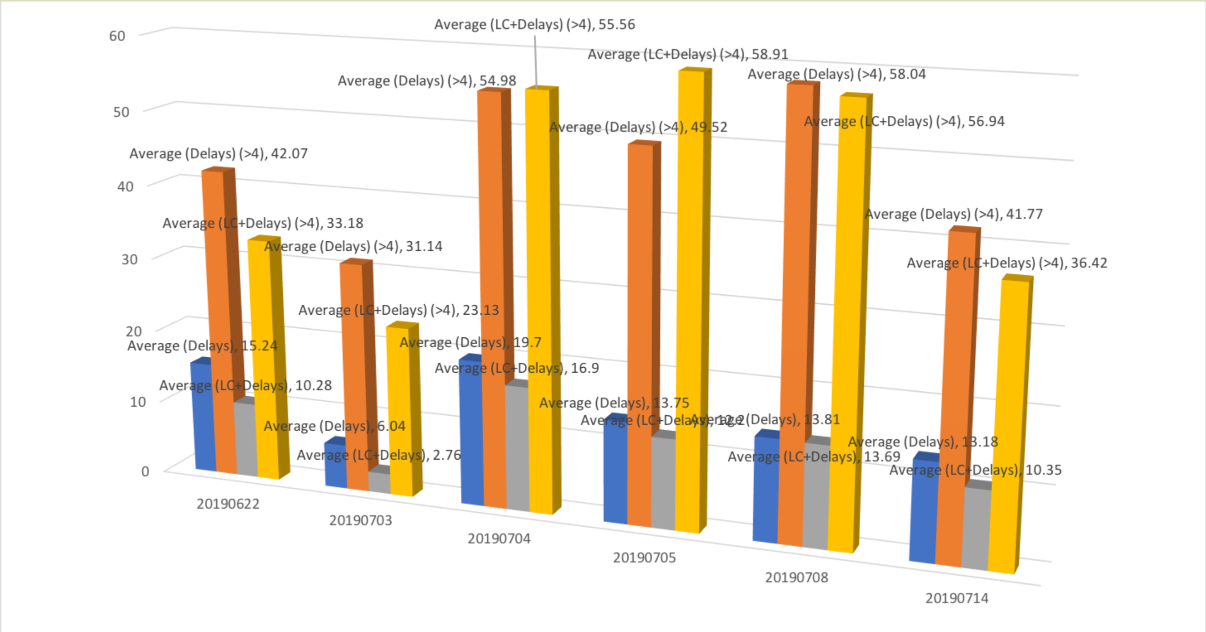


Figure 43. XAI Results from validation scenarios (Av.Delays)

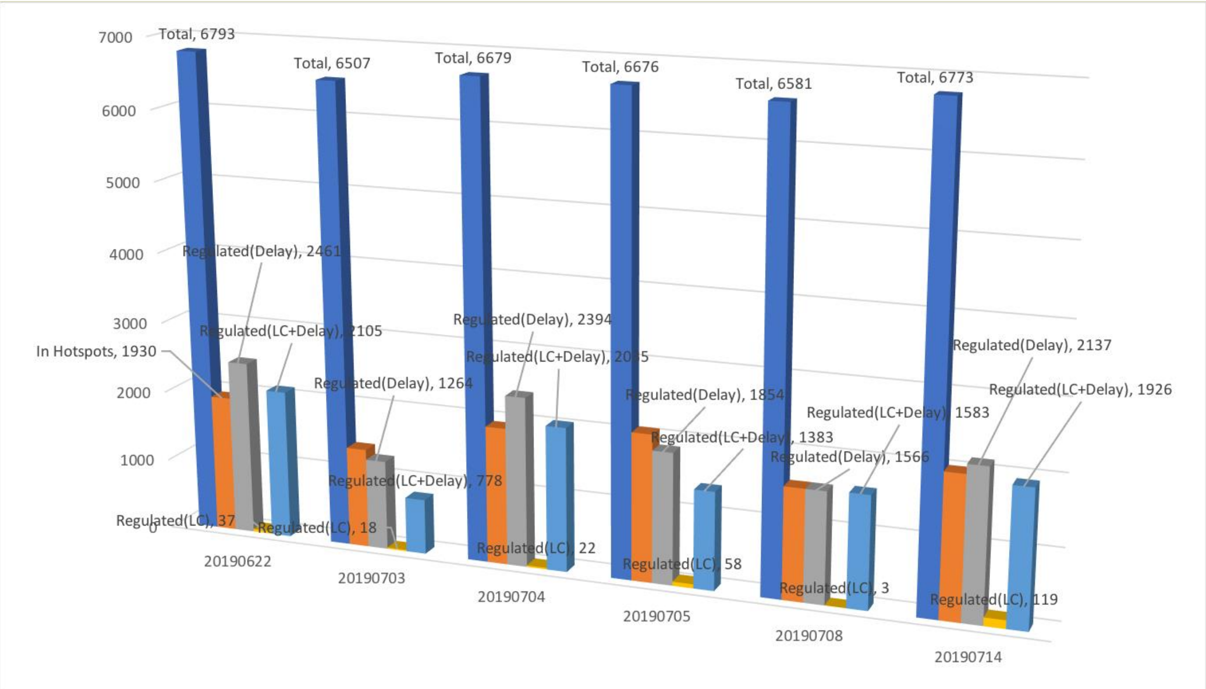


Figure 44. XAI Results from validation scenarios (Flights)

Appendix C Description of the ATFCM VA Features

Visual Analytics tool has been used as a support system to help the human operator to understand the proposed decisions. This section provides a summary of the features provided in support of the ATFCM validation exercises. For additional details, the TAPAS D4.3 - Visualizations and Visual Analytics methods describes the overall design of the Visual Analytics component of the ATFCM prototype system, the functionality of the VA module, and its integration within the TAPAS architecture.

The main features of the VA tool used during the ATFCM validation experiments are related to the analysis of the explanations provided by the XAI component. For this purpose, a first view comparing the baseline scenario and solution scenario is provided (Figure 45. Table views presenting baseline and solution scenarios side by side for enabling high-level comparisons).

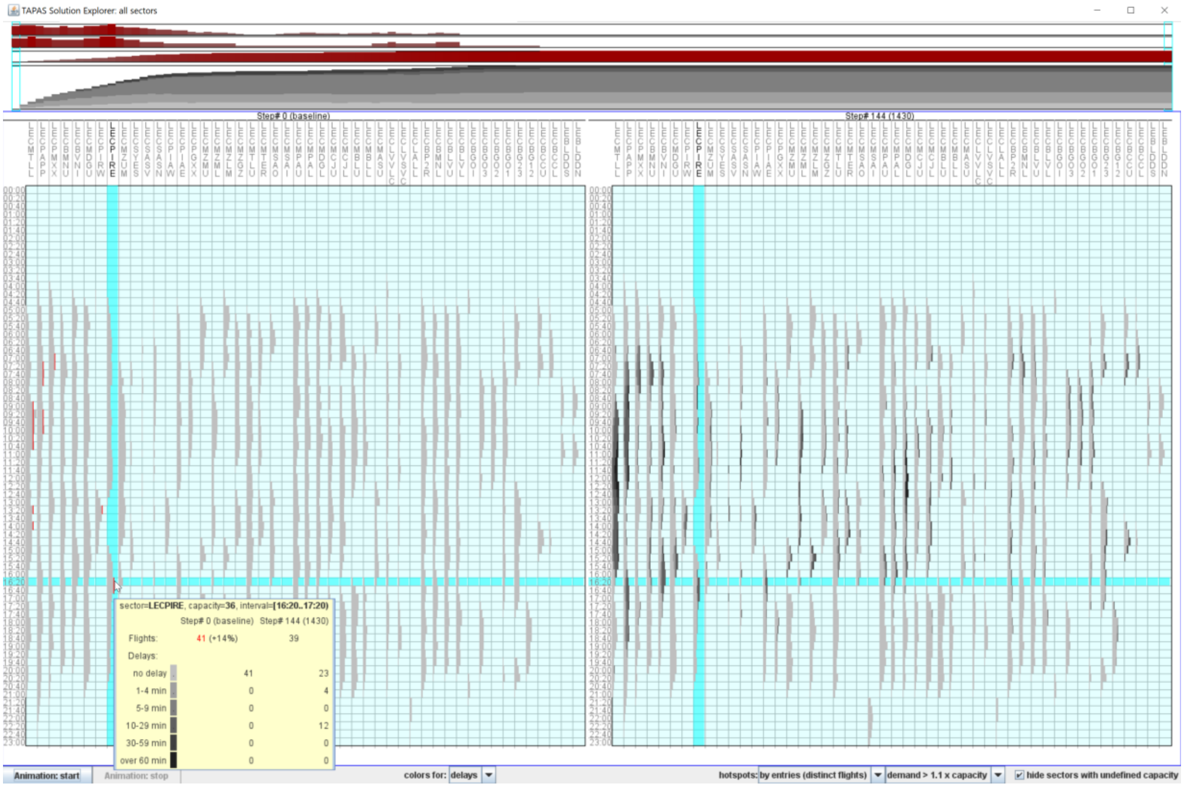


Figure 45. Table views presenting baseline and solution scenarios side by side for enabling high-level comparisons

This view enables a high-level comparison between these two scenarios by the traffic demand distribution in all sectors at every time. A scale of shades of grey represent the delays of the flights, in such a way that darker shades correspond to longer delays. Also, when a sector capacity threshold is exceeded the corresponding cell contains a red vertical line representing such threshold. In this way, capacity and demand can be directly compared. Additional information is displayed in a pop-up window when mouse cursor is positioned over an element of the view. Depending on the element, the information shown would be different. This is further explained in D4.3 Visualizations and Visual Analytics methods.

The next view used during the validation experiments corresponds to the Sector Explorer (Figure 46). It helps the FMP expert to focus on one single sector and the flights going from/to it.

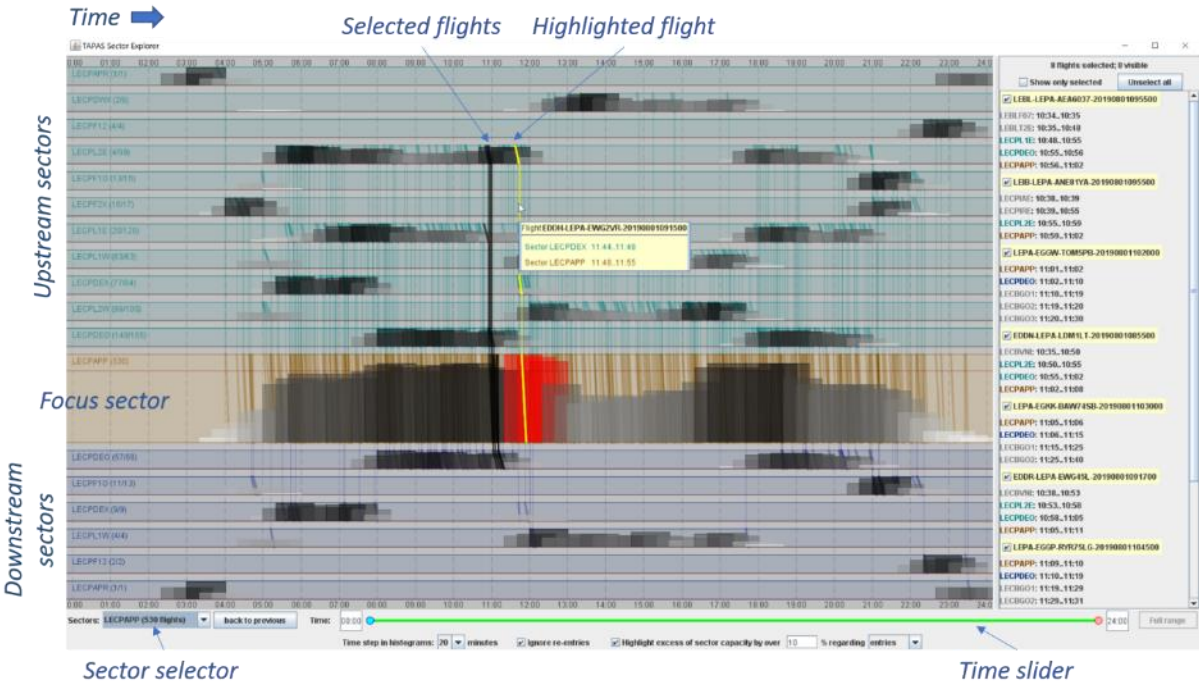


Figure 46. Components and contents of Sector Explorer in the mode of showing a single scenario.

This representation was used during the validation experiments to have a deeper understanding of the situation of one given sector. The entry or occupancy counts are represented by vertical bars and highlighted in red whenever the capacity threshold is exceeded. In this way, the time interval with capacity issues can be analysed along with all the flights involved. Detailed information of the view is represented by pop-up windows when the mouse cursor is positioned over an element of the view

An alternative view of the Sector Explorer would be the comparative mode. This option includes the comparison between baseline and solution scenario (Figure 47), representing their differences.

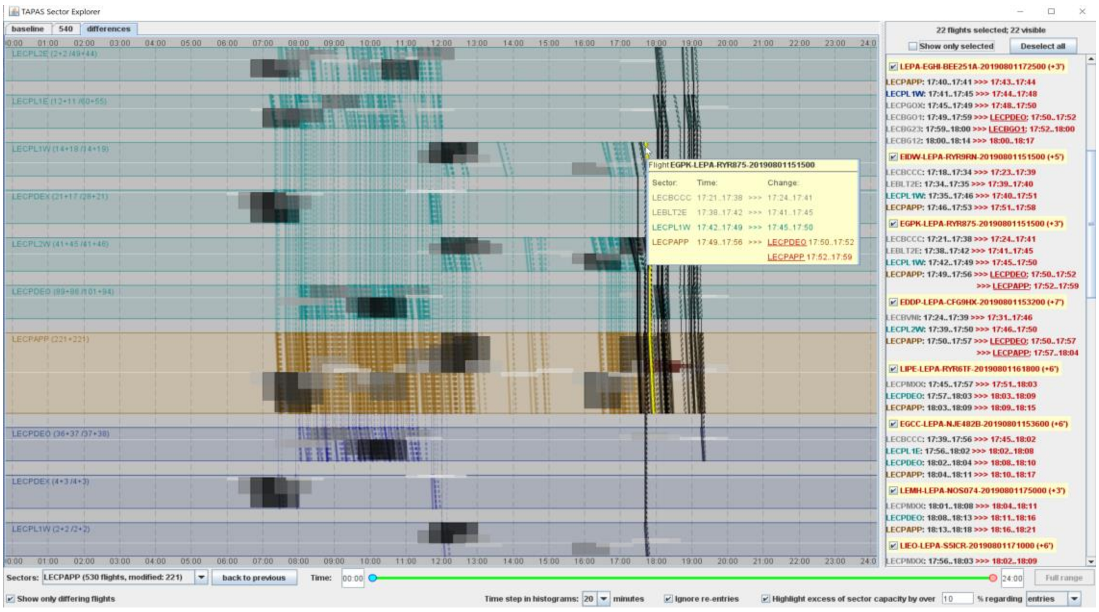


Figure 47. The comparative mode of Sector Explorer.

This was one of the most interesting views used during validation. FMP experts used it to understand the new route some flights had been given with the application of the measured proposed by the prototype related to one focus sector. It also helps to understand the delays there are in the solution scenario in comparison with the initial situation and in which time interval and sector were these delays focused. Positioning the mouse over a flight, user can see the original and new paths it would follow (highlighted in yellow) along with the sectors it crosses (displayed over and under the focus sector).

Finally, the explanations section of the VA component has been the most used during validation experiments. It collects all the explanations provided by XAI component into one view and classify them by flight (Figure 48).

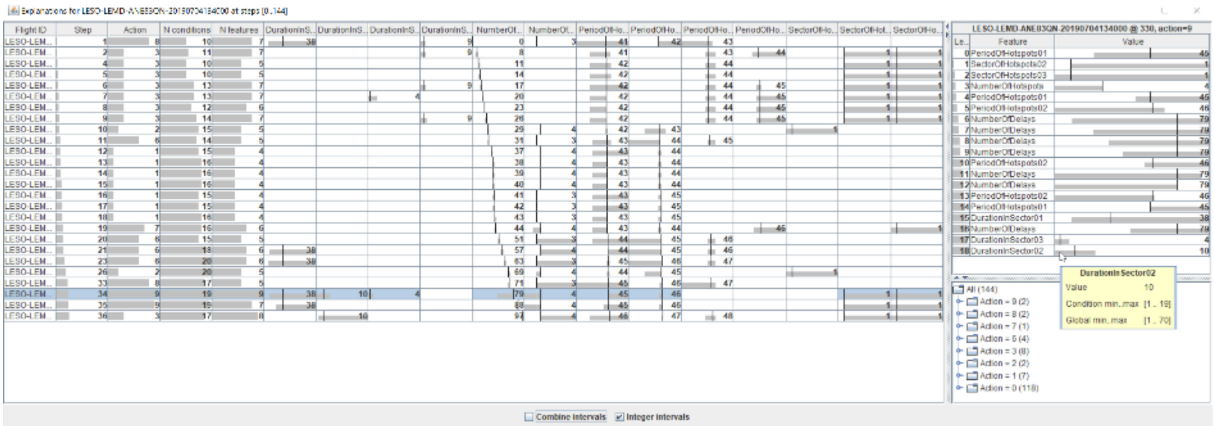


Figure 48. Explanations view for a particular flight

The explanations needed by the FMP experts to allow them to understand the measures proposed by the prototype are contained in this view. The information is classified in elements called features that represent different characteristics of the situation of one flight during time, such as number of hotspots, number of delays, period of hotspot, etc. All this data helps the user to understand how this flight has been affected gradually by the measures until the final solution is reached.

The views explained in this Appendix were the most used during validation experiments. Additional views complementing the ones shown and their description can be found in in D4.3 Visualizations and Visual Analytics methods.

Appendix D CD&R Validation Platform

CD&R exercises were performed using the SACTA Air Traffic Control Real-Time Simulator (RTS) platform at the CRIDA site in Madrid. A dedicated Controller Working Position (CWP) allowed the users to manage traffic in the different scenarios with the XAI automation and VA support tools connected using a solution based on message queues that was implemented using RabbitMQ as illustrated in the schematic shown below.

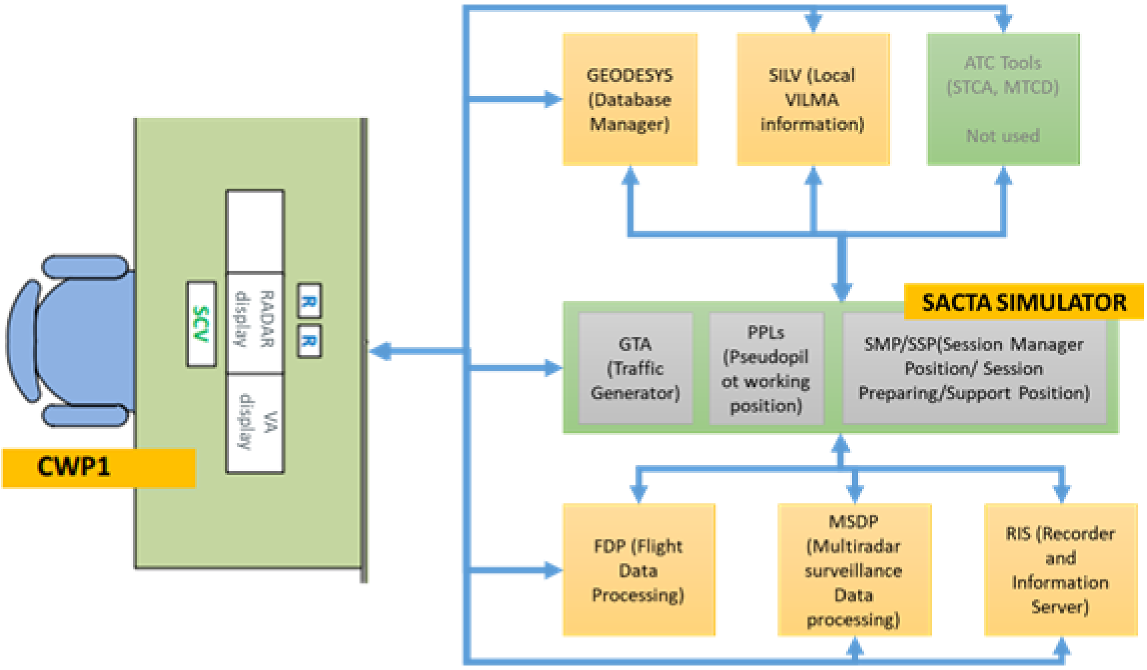


Figure 49: SACTA RTS with TAPAS CWP and co-located VA display

The SACTA RTS platform is primarily composed of two main sub-systems, the *SACTA simulator* and the *Controller Working Position* (iCWP) and the main simulator modules used for the real time simulation exercises included:

- **PPL (Pilot Working position)**, through which the pseudo-pilots have the ability to answer controller voice communications, express their preferences and execute controllers' orders during the simulation exercises
- **PAPO (Supporting position)**. This sub-system allows the preparation of the simulation scenarios and their modelling, including airspace, traffic and aircrafts' capabilities and performances modelling;
- **PCS (Session Control position).** This sub-system was used during the RTS to control the execution of each exercise and to modify it, if necessary (e.g. to generate conflict situations).

The CWP was the tool used by the executive controller to manage the traffic as they would do in a real operational environment. The CWP that was used during the RTS is a replica of the HMI of real Spanish ATC system (SACTA) that is currently in use at all ACC in Spain. It provides all of the same functionalities as the real system using the simulated traffic and airspace that was created for the experimental scenarios.

In addition, each CWP was equipped with a three-button mouse, a digital voice communication system with a headset, a footswitch, and a panel-mounted push-to-talk facility.

When the working position was covered by two ATC controllers, as illustrated in the photo below, one user performed the role of the Radar/Executive controller while the other manned the accompanying planning control interface.

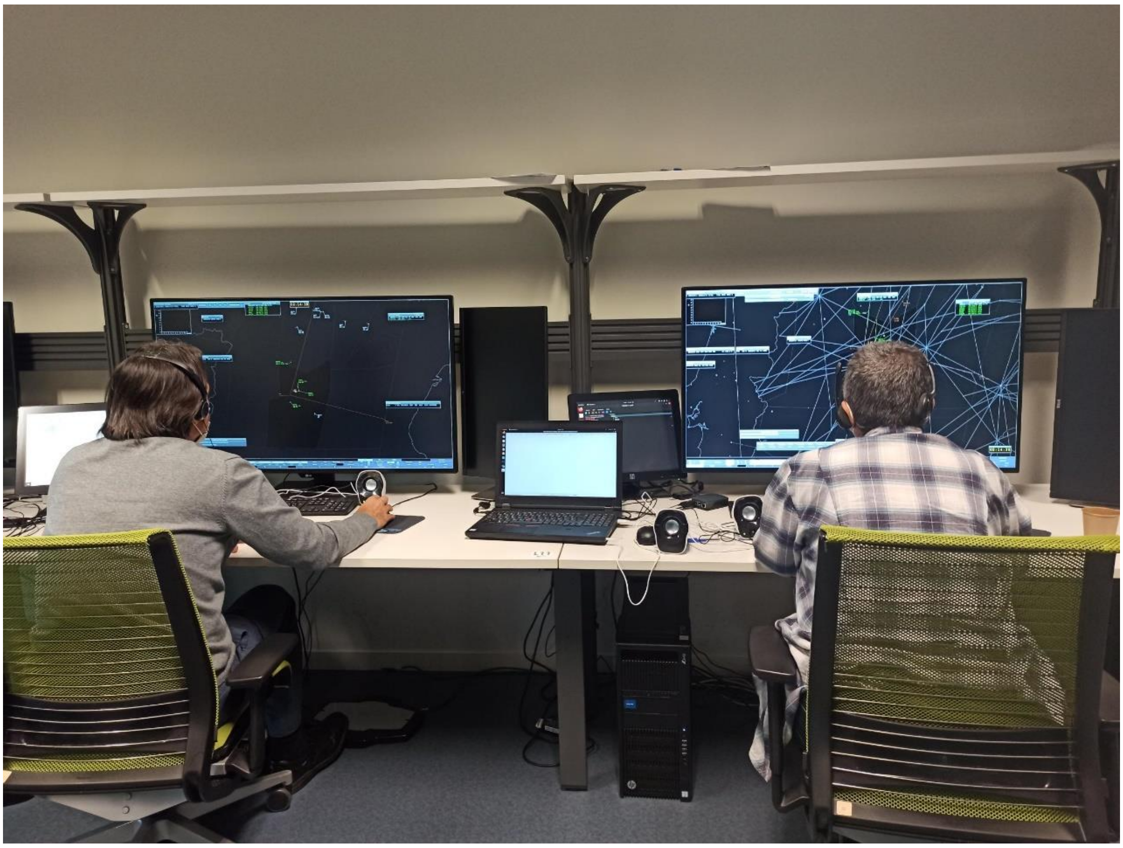


Figure 50: TAPAS - SACTA ATC RTS Working Position

Small differences between the simulated and the operational environment (e.g. absence of strips) were not sufficiently significant to challenge the experience of the users during the execution of the experiment.

A pseudo pilot position was maintained throughout the exercises (located in the second room) and was used to manage all traffic passing through the controlled area, as well as to respond to both voice and automated clearance requests for every aircraft that was included in the simulation scenario.

A second CWP was used during the level 3 automation scenarios to allow a '*ghost*' controller to implement solutions being proposed by the XAI automation tool in order to emulate full automation (i.e. execution of proposed clearances without the need for ATC intervention or instructions).

The VA conflict information was provided for most of the experiments using the smaller laptop computer located between the two controllers as seen in the photo above.

Appendix E Description of the CD&R VA Features

The task of the VA component was to inform the users about imminent conflicts between flights that have been predicted by the XAI component and to present one or more conflict resolution actions that were proposed by the automation tool. This needed to be done in real time, with the current traffic situation changing constantly, which meant that the visual displays were frequently updated to reflect those changes, and to consider any new information that became relevant as the scenario progressed.

For this reason, the visual interface in which the information was presented used displays that were predominantly in the form of tables. In addition, the interface included schematic graphical representations of the conflict geometry to help users to obtain a cognitive picture of the conflict being assessed.

The following is an example of the conflict information that was provided to the users via the VA information support application that was captured live during the execution of one of the validation scenarios:

Figure 51: Example of the VA information support display

As seen in the snapshot above, conflict information that is augmented using key data such as the start, CPA and end times of the conflict were provided in the header information of the conflict situation display. Additional indicators to illustrate the relative criticality of the situation (using a severity score, horizontal and vertical compliance indicators, and rates of closure) were also provided and colour coded highlights helped capture the attention of the viewer if the severity of the situation justified it (seen in the red bars above).

At automation levels 2 & 3, a list of potential actions to flights that could be used to solve the conflict was also provided in the body of the display as a tabular set of records that could be ordered according to any of the associated information contained therein (e.g. the *ranking* of each proposal).

The main window information window provided by the VA display consisted of four sub-windows that were separated by movable dividers. The sizes of any of the sub-windows could be adjusted by moving the dividers through mouse dragging.

At level 2, the user could read the proposed action and provide a clearance to the corresponding flight using the information provided about the manoeuvre to apply and its magnitude (e.g. a heading change of -10 degrees to the flight RYR350M as shown in the first record in the snapshot shown above).

Additional information for a given conflict could be accessed rapidly through pop-up windows by pointing at specific cells in the conflict description table as shown below. This offered users a very quick method by which to access more data if needed.

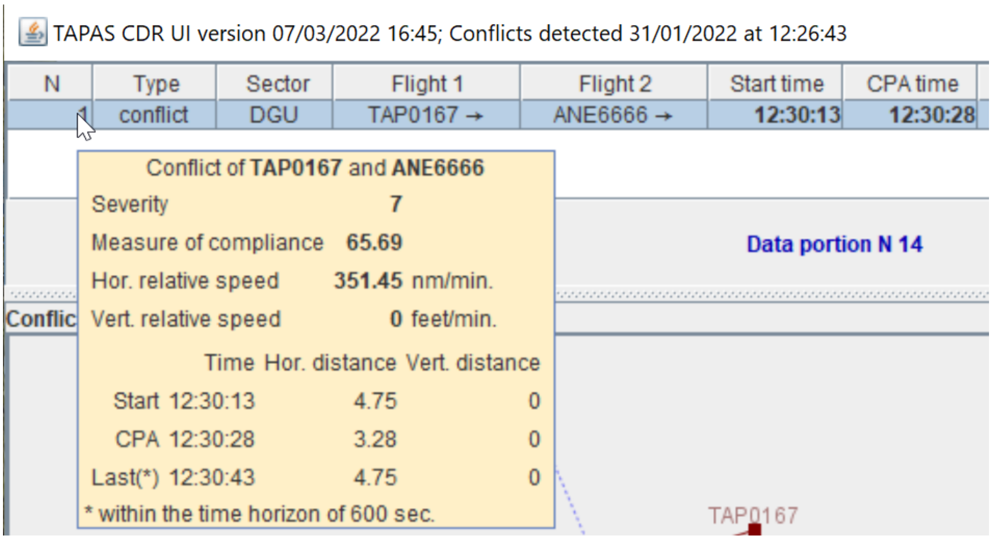


Figure 52. A popup window with information about a conflict re-arranged in a more compact layout

This figure above shows a popup window appearing when the cursor points at a cell containing the sequential number of a conflict. It contains the same information as can be seen in the table, but the information is arranged in a different way, which may be convenient when the table is wide and requires substantial gaze movement for reading values from a long row of the table.

If the cursor is pointed at a cell with a callsign of a flight. The pop-up window provides additional information about the flight itself, specifically, its course, horizontal and vertical speed, altitude, current attitude and destination airport.

 Co-funded by the European Union

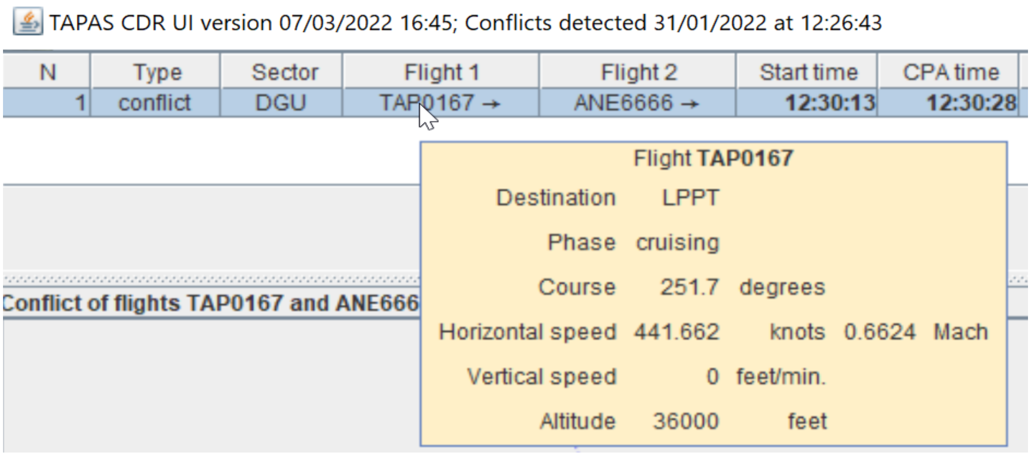


Figure 53. A popup window with additional information about a flight

As the traffic situation in the scenario evolved or following the delivery of a clearance to one or more flights in response to a conflict situation, the automation tools also performed a conformance monitoring function.

Additional popup windows could be consulted upon a user's request or when it was necessary to inform them about one or more non-conformance events.

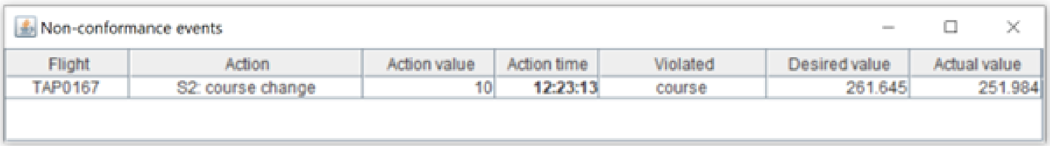


Figure 54. Example of a non-conformance event pop-up

These were used to help bring such anomalies to the attention of the operators and to allow them to take appropriate actions if required.

For a more detailed description about the VA tool as used in the CD&R validation experiments please refer to the TAPAS Validation Report(Section 3).

Appendix F Validation/Verification of the XAI algorithms

The CD&R TAPAS prototype, according to the workplan, is evaluated in two stages:

1. As an AI/ML method to resolve conflicts in an automated way, then
2. As an AI/ML method that can provide sufficient transparency to the air traffic controllers and keep them in the decision-making loop.

Both these stages are designed to fulfil the objectives of automation level 2 and level 3.

It must be pointed out however, that the major emphasis of TAPAS is on the second stage: i.e. Validating explainability and transparency, rather than building a state of the art AI/ML model for CD&R, per se.

Nevertheless, it remains necessary to provide an automated AI/ML method that can resolve conflicts as effectively as possible to ensure that the validation stage has valid objectives. Therefore, the CD&R AI/ML prototype was *trained and tested on real-world scenarios* which are constructed as described in D4.1 TAPAS Integrated Prototype document (Section 3.5).

As specified in D4.2 Reference of XAI Methods document, two models were *comparatively tested* that learned using the following training patterns:

a) **A sequential one** (named ***NSeq***) – where given N scenarios, a model is gradually trained to all scenarios in sequence, starting with a model trained to the 1st scenario, then updating this scenario by training it to the 2nd scenario and so on, until the Nth scenario.

b) A scenario trained to all **N scenarios at the same time** (named ***AllN***).

Using these training patterns (tested with few scenarios, initially) it was concluded that the ***NSeq*** pattern manages to be more effective, but the ***AllN*** approach manages to construct models that generalize more effectively compared to ***NSeq*** models.

Additionally, the training process using the ***NSeq*** pattern is much more computationally efficient compared to the training time required by the ***AllN*** training pattern.

Having these initial results and to ensure that the CD&R models were trained in time to be used during the TAPAS validation phase, it was decided to train one model using a mixture of these training patterns in a set of 36 scenarios: Batches of 6 scenarios (i.e. 6 batches) were created and a model was trained sequentially for these batches. Training the model for each batch, the ***All6*** training pattern was then adopted for the 6 models in a single batch.

Overall, results show that the AI/ML model manages to resolve most of the conflicts in the test scenarios (as it was also the case during the validation exercise), with few ATC instructions (resolution actions) (in average 3) and with few (in average 0) additional NMs.

Additional information describing the complete training and testing process, the distinct scenarios and the corresponding results in greater detail are available in the D4.2 document.

-END OF DOCUMENT-