Meteo Italian SupercompuTing poRtAL

Deliverable 3.1

Definition of Mistral Use cases and Services

Deliverable Lead:	DPC
Deliverable due date	28/02/2019
Status	FINAL
Version	V1

This project has received funding from *European Commission*
Connecting Europe Facility 2014-2020
AGREEMENT No INEA/CEF/ICT/A2017/1567101

Document Control Page

Title	D3.1 Definition of Mistral Use cases and Services
Creator	DPC
Publisher	Mistral Consortium
Contributors	Davide Cesari (ARPAE), Emanuele Di Giacomo (ARPAE), Luca Delli Passeri (DPC), Estíbaliz Gascón (ECMWF), Timothy Hewson (ECMWF), Andrea Montani (ECMWF) , Martina Forconi (Dedagroup), Stefano Pezzi (Dedagroup), Gabriella Scipione (CINECA), Giuseppe Trotta (CINECA), Cinzia Caroli (CINECA)
Type	Report
Language	en-GB
Rights	copyright “Mistral Consortium”
Audience	☒ public ☐ restricted
Requested deadline	28/02/2019

Contents

Executive Summary

The aim of the Mistral project is to facilitate and foster the re-use of datasets by the Italian meteorological community, as well as its cross area communities including the private providers and the citizens.

This deliverable describes some possible access methods to the meteorological data could be useful both for institutional and non-institutional stakeholders.

The institutional stakeholders are the Italian Regional Agencies for the Environmental Protection (ARPA), the Italian Civil Protection Department (DPC) and in general all Italian public administrations.

The non-institutional stakeholders are citizens, national and international private organizations.

In Chapter 1 the possible modalities of access to the data are expressed as the informal description of the system features from the perspective of the user.

On the base of the user stories, in Chapter 2 the stakeholders' needs are formalised in the form of use cases. The use cases allow to identify the functionalities the Mistral system could provide.

The valuable results outlined in this deliverable are:

- the description of the stakeholders' needs
- the formalisation of the behaviour of the system to meet the users' needs, through the definitions of the use cases' actors, the UML diagrams of the use cases and the events flows tables.

Furthermore, the Annex I provides a glossary.

This deliverable represents the requirements of the potential users of the Mistral system and is the basis for the next steps that are the design of the architecture and of the components and the development.

Based on the requirements specified in the present D3.1, the technical partners will perform an evaluation of the effort required to implement each single use case proposed. Based on the priorities, the technical partners will assess which of them can be implemented within the scope and the foreseen budget of the Mistral project and which will be done in a possible evolution of the project itself.

1. The stakeholders' needs

In this chapter the stakeholders' needs are expressed as the informal description of the system features from the perspective of the user.

1.1. Observed meteorological data in real time for hydraulic monitoring purposes

Arpae-SIMC monitors the observed precipitation and river level from stations placed in Emilia-Romagna and in the surrounding area.

The data managed by Arpae-SIMC and locally stored are integrated with data from other institutional or private entities or from amateur networks remotely available in one ore more Mistral instances. For performance and fault-tolerance reasons, the remote data are locally stored.

The data from Mistral instances are sent as soon as they are available and without data loss nor duplication; the data can be retransmitted in case of attributes update (e.g. invalidation, computation of confidence, consistency check).

In case of connection loss, Arpae-SIMC can restore the data transfer from Mistral server at the point where it was left. The user of this application must be able to view data backward up to 5 days.

The Figure 1**Errore. L'origine riferimento non è stata trovata.** is an example of the visualisation of observed data in graphical form for hydraulic monitoring.

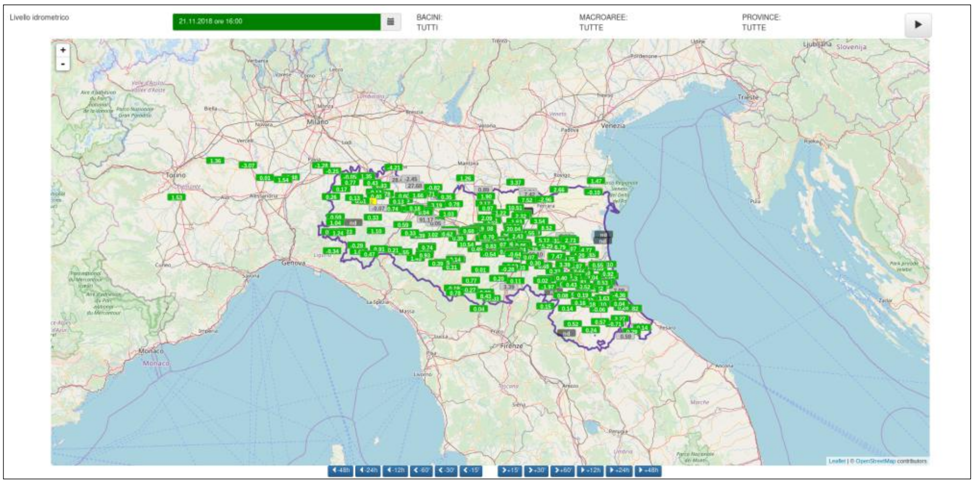

FIGURE 1 - EXAMPLE OF OBSERVED DATA IN GRAPHICAL FORM

1.2. Observed meteorological data in real time for data assimilation with an atmospheric model

Running an atmospheric data assimilation procedure for initialising a numerical weather prediction model involves the access in real time to the latest atmospheric observations available on a short interval in the recent past, with the purpose of creating a picture of the state of the atmosphere at the current time.

Defining as *t* the time at which the state of the atmosphere has to be computed, and *t-dbt* the initial time, in the past, at which observations are required, where *dbt* may vary from about 2 to 12 hours or more, depending on the application, the common practise is to wait for a predefined time *dft* after *t*, so-called "cut-off", where *dft* could typically vary between 1 and 8 hours, then at *t+dft* all the available atmospheric observations taken after *t-dbt* are retrieved and the data assimilation procedure starts.

After the start of the assimilation procedure, usually no more observations can be fed to the system, so the best approach could be the access, at time *t+dft*, to an observational archive updated without appreciable delay with respect to the arrival of data in real time.

The typical archive query criteria could thus be:

- reference time between *t-dbt* and *t+dft* (or simply after *t-dbt*)
- a specified (short) list of datasets (also known as observational networks)
- geographical coordinates lying in a specified lon/lat bounding box.

It could be however kept in mind that, in the near future, data assimilation systems may be capable of receiving runtime updates of the observational dataset during the assimilation process itself, thus a continuous real time access as described in the previous use case may be considered also for this case.

1.3. Verification of an atmospheric model on a long period

Meteorological centres routinely running operational numerical weather prediction models use to periodically verify the quality of the model forecasts with respect to observations on long time intervals. This is usually done on a seasonal (3 month) basis, with a delay of some months with respect to the real time, but it could also be done on a multi-year basis (e.g. all the springs of the last 5 years) or on longer continuous periods, such as 1 or 5 years.

This type of data access is not time critical, so a relatively long waiting time is acceptable, however, when long time intervals of data are desired, it may be useful to split the query in many shorter queries, in order to reduce the data transferred during a single download and reduce the damage in case of transfer failure. For this reason it could be useful to have a command-line client application capable of performing unattended queries, to be repeatedly called by a user script iterating over different time intervals.

Observed data to be used for verification should have undergone quality control.

In order to take into account the different requirements, verification case has been divided into surface verification and upper-air verification.

1.3.1. Surface verification

Meteorological model verification involves the need to retrieve separately observed data and gridded forecast data for long intervals. Since surface verification is usually done separately and with different techniques for different groups of observed meteorological parameters (e.g. temperature at 2 meters above surface, dew-point temperature or relative humidity at the same level, wind at 10 meters, solar radiation at surface, total precipitation at surface), a common approach is to separately retrieve the whole period of forecast fields (including all forecast ranges) for each group of variables to be verified, although this splitting by variable could also be done at a further stage. For observations, due to the nature of weather reports, it is usually not possible to split the query result by variable, thus all the surface observations covering the verification period, for a set of observational networks, have to be retrieved. For the verification of some quantities, it may be necessary to compute derived variables, both from forecast and observed data, e.g. computing speed and direction of wind from x and y components, or computing relative humidity from specific humidity or dew-point temperature (plus pressure or temperature if required).

The typical archive query criteria and post-processing for forecast data could thus be:

- a specified dataset (model run)
- reference time between t1 and t2
- a subset of forecast time ranges (e.g. time ranges modulo 3 hours)
- a specified combination of variables and vertical levels
- full model grid (no space transformations)
- optional time post-processing (e.g. accumulate precipitation on a specified time interval)
- optional derived variable computation (e.g. relative humidity or wind speed).

The typical archive query criteria and post-processing for observed data could be:

- a specified list of datasets (observational networks)
- reference time between t1 and t2
- geographical coordinates lying in a specified lon/lat bounding box

- optional time post-processing (e.g. accumulate precipitation on a specified time interval)
- optional derived variable computation (e.g. relative humidity).

1.3.2. Upper-air verification

For upper-air verification, usually done with respect to vertical atmospheric profiles measured by a few radio-sounding balloons (so called "TEMP"), in principle a big amount of 3-dimensional forecast data is required for reconstructing the model analogue of the observations. For this reason, in order to reduce the data transferred from the archive, it may be useful to perform a data reduction on the archive itself, by horizontally interpolating there all the 2-dimensional sections retrieved from the archive on a set of predefined points provided by the user, representing the coordinates of the few fixed radio-sounding stations. The user will then receive only a set of sparse data, in a format similar to the format of observed data, representing the forecast on the points of interest. This will not reduce the amount of data to be retrieved from the archive, but it reduces the amount of data to be transferred to the user and the working storage required by the user for temporarily keeping the data. A similar data reduction is not usually desired for surface data, firstly because surface model data are much less voluminous than upper-air model data and, secondly, because for surface data, unlike for upper-air data, it is often desired to perform special customised horizontal interpolations which are not suitable for being performed on the data server with generic interpolation tools.

The typical archive query criteria and post-processing for forecast data could thus be:

- a specified dataset (model run)
- reference time between t1 and t2
- a subset of forecast time ranges (e.g. time ranges modulo 6 hours)
- a specified list of variables
- a specified list of vertical levels
- space post-processing, horizontal interpolation of the data on a list of user-requested geographical points.

The typical archive query criteria for observed data could be:

- a specified list of datasets (observational networks)
- reference time between t1 and t2.

The list of points (order of tenths) for model data interpolation may be specified either manually or by uploading a georeferenced vector file.

This use case may benefit from HPC resources in the phases of space/time post-processing and computation of derived variables.

1.4. Redistribution of observed data to external users

Arpae-SIMC has a large number of external users accessing observed data. To minimize the development costs, the users can be forwarded to a proper service in one of the Mistral instances.

The data access can be divided into two different services: real time data and archived data.

1.4.1. Real time data

The user connects to a stream of real time data. The channel must support persistence: on reconnection, the client starts the stream from where it left the previous time with a limited persistence window (e.g. 15 days). A user may have a special authorisation level, (e.g. official institutions or paying customers) and have thus access to a personalised channel collecting a number of desired datasets and a longer persistence window.

On the other hand the user could be an anonymous public user and have thus access only to a shared public stream of data without authentication. The public real time stream doesn't support server side filters, but the user can use a multiplatform client software, distributed in Mistral portal, for locally filtering and formatting the data after they have been received.

1.4.2. Archived data

When accessing archived data, which include both historical data and recent "near-real time" data, the user may select data based on the following criteria: dataset (observational network), reference time (e.g. data observed between time *t1* and *t2*), geographical bounding box and "last modification time" attribute.

The typical archive query criteria for observed data could be:

- a specified list of datasets (observational networks)
- reference time between *t1* and *t2*
- geographical coordinates lying in a specified lon/lat bounding box
- "last modification time" attribute greater than *t3*.

The common use case for the "last modification time" is the following: a user periodically executes some climatological statistics over 30 years. Instead of downloading every time 30 years of data, the user keeps a local archive and downloads only the data that have been recorded or updated after the previous data access.

1.5. Gridded model forecast data in numerical form or aggregated on sparse points

Forecast data are typically distributed in GRIB format to institutional and private clients, either using an interactive web interface or through web services for automated processes. The access is usually done in near-real time, i.e. as soon as a model dataset has been completely filled after a successful model run, this implies that there should be a mechanism for informing the user that/when a certain dataset has been filled for a specific reference time, or for scheduling a data query as soon as this has happened. There should be however the need to perform the same query for a longer period in the past, e.g. for training of a machine learning process, with an access pattern similar to the one shown in the verification use case.

The typical archive query criteria and post-processing could thus be:

- a specified dataset (model run)
- reference time equal *tlast* or between *t1* and *t2*
- a subset of forecast time ranges (e.g. time ranges modulo 3 hours)
- a specified combination of variables and vertical levels
- full model grid (no space transformations)
- optional time post processing (e.g. accumulate precipitation or computation of maximum and minimum temperatures on specified time intervals)
- optional derived variable computation (e.g. relative humidity or wind speed)
- optional spatial post processing (e.g. a bounding box delimitation or a re-gridding of the output).

Forecast data could also be aggregated on sparse points through a postprocessor, that permits the interpolation of gridded data on sparse point with the possibility to choose the methodology (e.g. bilinear, nearest point, area average) and the output format (e.g. BUFR, JSON, CREX).

This use case may benefit from HPC resources in the phases of space/time post-processing and computation of derived variables.

1.6. Gridded model forecast and observations data in graphical form

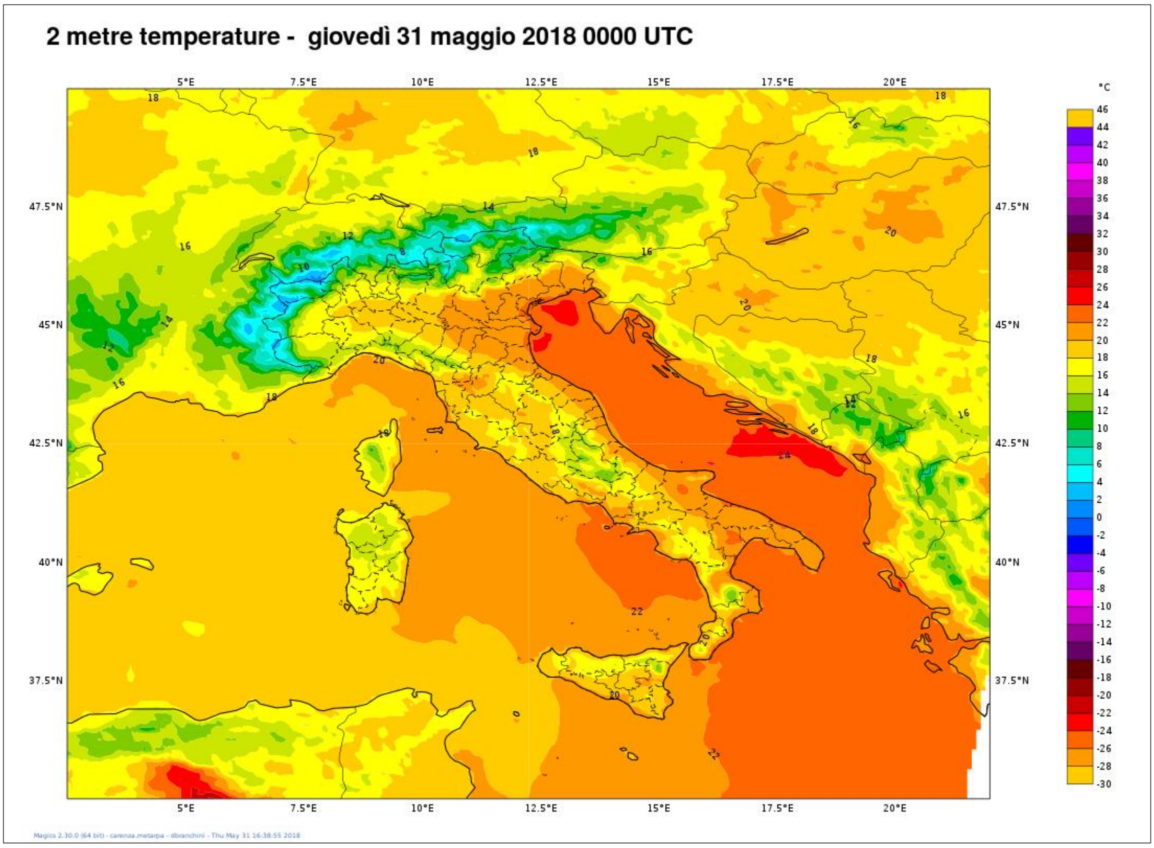

FIGURE 2 - EXAMPLE OF FORECAST DATA PLOT

Observation and forecast data plots, in the form of field contour and shading, wind vector or barbs, other point symbols, are fundamental for internal uses, such as subjective routine weather forecasts or model performance evaluation as well as for external users. In this use case a user chooses the data to be plotted and the plotting style. The result may be either in the form of raster maps or in a georeferenced vector format and should be available either for visualisation online, on a dedicated temporary web space, or for download as a bundle of maps.

As for other use cases, the request may be done interactively or via web services and require near-real time data or past archived data.

The typical archive query criteria and post-processing for forecast data are similar to the ones indicated in the gridded model data use case:

- a specified dataset (model run)
- reference time equal *t0*
- a subset of forecast time ranges (e.g. time ranges modulo 3 hours)
- a specified combination of variables and vertical levels
- full model grid (no space transformations)
- optional time post-processing (e.g. accumulation of precipitation or computation of maximum and minimum temperatures on specified time intervals)
- optional derived variable computation (e.g. relative humidity from temperature and dew-point temperature).

The archive query criteria for observational data could be:

- a specified list of datasets (observational networks)
- reference time between *t1* and *t2*
- geographical coordinates lying in a specified lon/lat bounding box.

After the retrieval from the archive, the data should undergo a graphical post-processing, possibly customised by a style file uploaded by the user, finally obtaining a set of two dimensional georeferenced graphical maps.

This use case may benefit from HPC resources in the phases of space/time post-processing and computation of derived variables and in the generation of the graphical output.

1.7. Italy flash flood

The Italy Flash Flood use case is being implemented with the collaboration of the Centre of Excellence ECMWF that is developing an application for Italy based on recent advances in HPC-based post-processing of global ensemble forecasts, specifically forecasts of rainfall. This application will include blending of this post-processed output with rainfall output provided by the Italian high-resolution 2.2km COSMO ensemble. Also included in this activity will be case-study example(s), in which the output is applied to flash-flood episodes. This use case is a demonstration of the use of supercomputer architecture to make real-time product delivery tractable.

The use of HPC resources is crucial for this use case, because:

- Large amounts of computational and high-quality I/O resources are required for operational production of numerical model weather forecasts. To provide predictions with great accuracy the meteorological models have high resolution grids and need to be run with short time steps. This implies a need for elevated supercomputing power and performance.

- HPC is also necessary for ensemble forecasting, where the meteorological model is executed several times starting from slightly different initial conditions. In this use case the ensemble forecasts will be transformed into much more useful information, in real time, by using the supercomputing resource to analyse and post-process the raw forecast data relatively quickly. For example, in the ECMWF point-rainfall forecast post-processing, the methodology enables us to generate a new 100 member ensemble for each of the 51 ensemble members, to highlight all possible realisations of point rainfall within each ensemble grid box, for each lead time interval. This implies the overall generation of 5100 ensemble members for each of 40 time windows, which amounts to creating 1.6 Tb of data (for just the rainfall totals) at an intermediate stage. The computational demands this creates are very high, and in tests at ECMWF, it has been found that the only viable option for operational production is to use HPC resources.

Several tasks are planned to develop the final Flash Flood use case:

- The first task, now in progress, involves porting pre-existing ECMWF point-rainfall code to the new national supercomputer portal at CINECA, and adapting and optimising the code to fit the specific requirements of the new platform, such as its file system, I/O constraints and memory availability. Currently the ECMWF point-rainfall forecast production, for 12-h precipitation totals, is running on a semi-operational basis on the ECMWF supercomputer in Reading (UK). The development, testing and verification processes for 6-hourly point-rainfall have begun, and the related code will be activated and optimized on the CINECA HPC this year. Once this 6-h point-rainfall has been successfully implemented, 12-h and 24-h point-rainfall will be also considered for implementation on the Italian supercomputer.
  Offline calibration and verification activities will also be a fundamental part of this activity during this year. Code construction is such that computational complexity can be expanded according to what supercomputer capacity allows. Tests with the current system have shown that expansion in this way has clear potential to reap rewards in terms of improving the accuracy of this ECMWF point-rainfall product. This expansion will be investigated further and implemented where beneficial and cost-effective.
- The second task will involve post-processing of a second input source, specifically the recently-developed 2.2km COSMO ensemble, and blending that output with the ECMWF point-rainfall output in such a way that the most skilful aspects of the two systems can be best exploited to provide products for the end users. This blending requires some investigation (currently underway), although there is considerable previous experience at other nearby centres to build upon. A new variable-scale neighbourhood post-processing technique, developed in Reading University meteorology department, is being trialled in the expectation

that this will likely be the best way to obtain maximum benefit from the COSMO probabilistic forecasts.

- The final stage will be to deliver a fully-operational real-time system for forecasting rainfall probabilistically, based on the model output blending. Compared to using raw model output, we expect to deliver gains across the full range of rainfall severity, from small totals right through to the extreme amounts which can lead to devastating flash floods. Verification of the ECMWF 12-h point-rainfall output has already illustrated the substantial improvements that can be realised. Indeed, compared to other published methods, these improvements are very noteworthy, especially for large precipitation amounts. For example, using the standard ROC area verification metric, forecasts for day 10 in the post-processed global ensemble output for totals>50mm/12h are as good as are raw model forecasts for days 1-2. Furthermore, by combining with the COSMO ensemble output we expect to improve our forecasts even more. It is planned that COSMO will provide very high temporal resolution very early on, which will evolve into 6-hourly output when ECMWF point-rainfall input is first used, and forecasts for overlapping 6-hourly intervals will then be delivered up to a lead time limit of 120 to 240h. The final product will comprise probability percentiles, for accumulated rainfall, for each COSMO grid point (percentiles 1,2,..99). These can be used by customers as is, or simply converted into exceedance probabilities for user-defined thresholds.

2. Actors

The actors are users that interact with a system. An actor can be a person, an organization, or an internal or external application. Actors could be external objects that produce or consume data.

The actors of Mistral system are:

Anonymous data consumer	The anonymous data consumer is a user that can download or visualize on a map meteorological observation data or forecast data
Authorized data consumer	The authorized data consumer is a user that can download private data (observational and forecast) depending on his authorization. The user could belong to a "weather institution" (e.g. DPC, ARPA) or could belong to private sector or public institutions; e.g. INPS (National Social Welfare Institution), multi-utility, RFI (Italian rail network), energy company.
Data assimilation system	The assimilation system retrieves observed meteorological data for data assimilation.
Data consumer system	The data consumer system receives observational data in real time for general purpose with no transformations. It is a system that "exchanges" data with Mistral. It may also be another instance of Mistral system.
Data provider system	The data provider system provides observed and forecast data to Mistral. For observed data it could be an official provider (e.g. DPC, ARPA) or a non-official (e.g. MeteoNetwork); in both cases is authenticated and authorized to provide observations. For forecast data is an official system authorized to provide forecast data (e.g. Lami). It may also be another instance of Mistral system.
Flash flood consumer	The flash flood consumer is an authorized user that get the real time forecast rainfall probabilities based on model output blending of ECMWF point-rainfall output with 2.2 KM COSMO ensemble.
Forecast verification user	The forecast verification user retrieves observed meteorological data and gridded forecast data or sparse points for verification of an atmospheric model.

Plot provider system	The Plot provider system provides predefined plots of forecast data to the Mistral system (e.g. Lami).

3. Use cases

In this chapter the use cases are formalized with UML diagram and with event flow tables. The shared use cases are described at the end of the chapter.

3.1. UC1 - Send/Receive observed meteorological data in real time and quality controlled

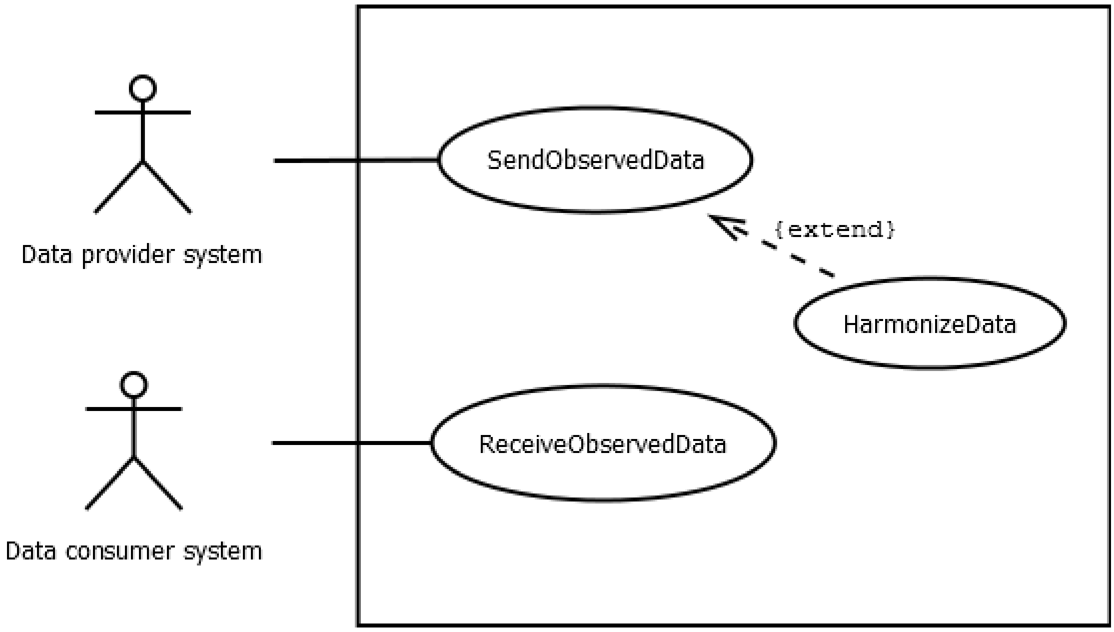

FIGURE 3 - UML DIAGRAM OF UC1

3.1.1. UC1.1 - SendObservedData

Use case: SendObservedData
ID: UC1.1
Actor(s): Data provider system
Pre-conditions: 1. the Data provider is an authenticated user of the system and authorized to send the

observed data

Description:
• the Data Provider sends the observed data of the network of competence
  ○ data can be expressed in various allowed formats (e.g. BUFR, JSON)
  ○ data will contain information such as: station identification, sensor quality, position, recording time, measurement of observed variables (one or more), etc.
• the system validates the data before saving them. It is a structural validation (i.e. JSON or BUFR schema rules); no content validation is applied
• the system optionally applies some transformations to the incoming data (see UC1.3).
• in case of "duplicated" data (i.e. previously sent) the system replaces the existing data with the new ones. Quality controlled data could be re-transmitted and the system replace previous data.
• the observed data will be saved and stored in the system

Post-conditions: 1. the system accepted and saved the data

3.1.2. UC1.2 - HarmonizeData

Use case: HarmonizeData
ID: UC1.2
Actor(s): Data Provider System
Pre-conditions: 1. Data are not fully compliant with mistral data model

Description:
• User access the system to send data with some slight differences from the specifications (temperature in °F instead of °C or different rainfall period of time)
• The system apply the following transformation to make them compliant to mistral:
  ○ units of measure (e.g. temperature in °F instead of °C)
  ○ coordinate reference systems transformations
• The system calculate new value to make data comparable (e.g. cumulated rainfall in a period of time multiple of the others)

Post-conditions: Data are harmonized with mistral

3.1.3. UC1.3 - ReceiveObservedData

Use case: ReceiveObservedData
ID: UC1.3
Actor(s): Data consumer system
Pre-conditions: 1. A new observed data has been saved in the system
Description: • The use case starts every time a new data is saved in the system • The observed data are sent from the system to the data consumer as they are (no data transformation) • Data transmission takes place in "real-time" mode, i.e. re-transmitted as soon as possible with "push" mode • Each data consumer will receive data based on the defined permissions (data policy) • There must be no data loss or duplication in the transmission • The data can be re-transmitted when a quality attribute is updated (e.g. invalidation, computation of confidence, consistency check) • In case of loss of connection, the transmission must start again from the point where it was interrupted
Post-conditions: 1. All data expected for the data consumer have been delivered

3.2. UC2 - Retrieve of observed meteorological data for data assimilation with an atmospheric model

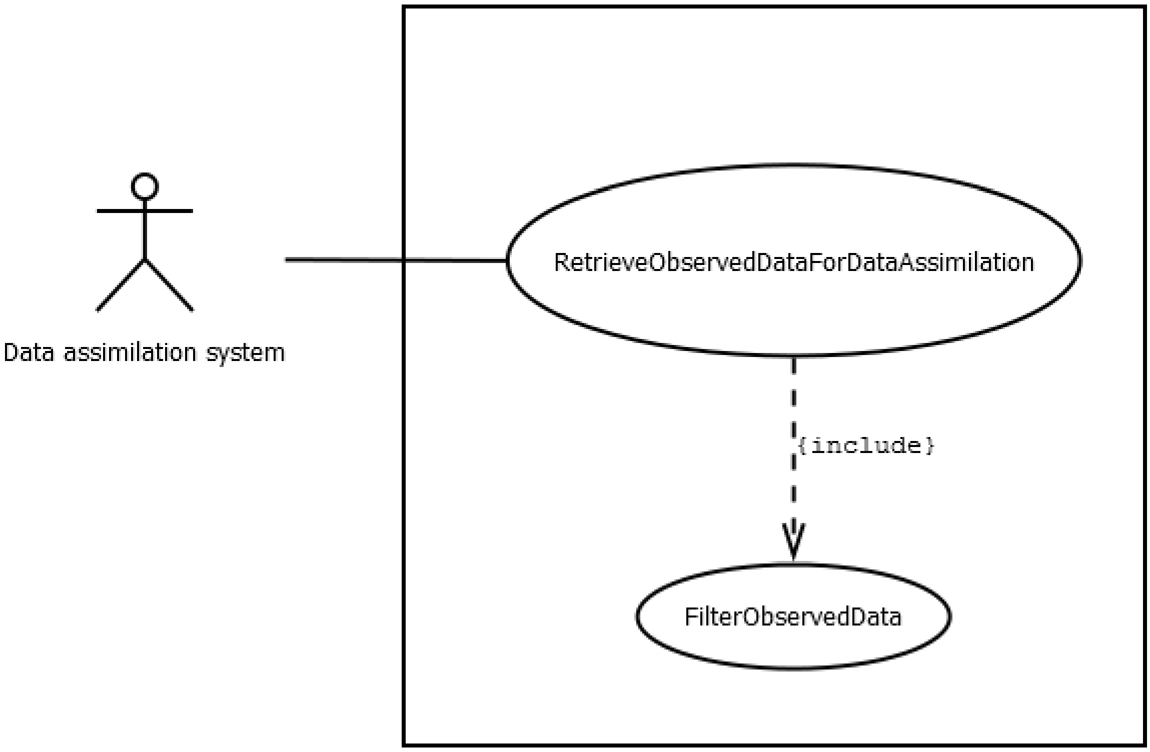

FIGURE 4 - UML DIAGRAM OF UC2

3.2.1. UC2.1 - RetrieveObservedDataForDataAssimilation

Use case: RetrieveObservedDataForDataAssimilation
ID: UC2.1
Actor(s): Data assimilation system
Pre-conditions: 1. the Data assimilation system is an authenticated user of the system and authorized to download observed data
Description: • The use case starts when the user sends a query for near real time observed data to the

system with mandatory filter parameters (see UC9.1) • The system puts the user's request into an elaboration queue with a specific priority • The system takes the request from the queue and executes the query on the local database, applying filter specified by the user • The system sends back to the user the data resulting from the query
Post-conditions: All data have been delivered

3.3. UC3 - Retrieve of observed meteorological data and gridded forecast data for verification of an atmospheric model

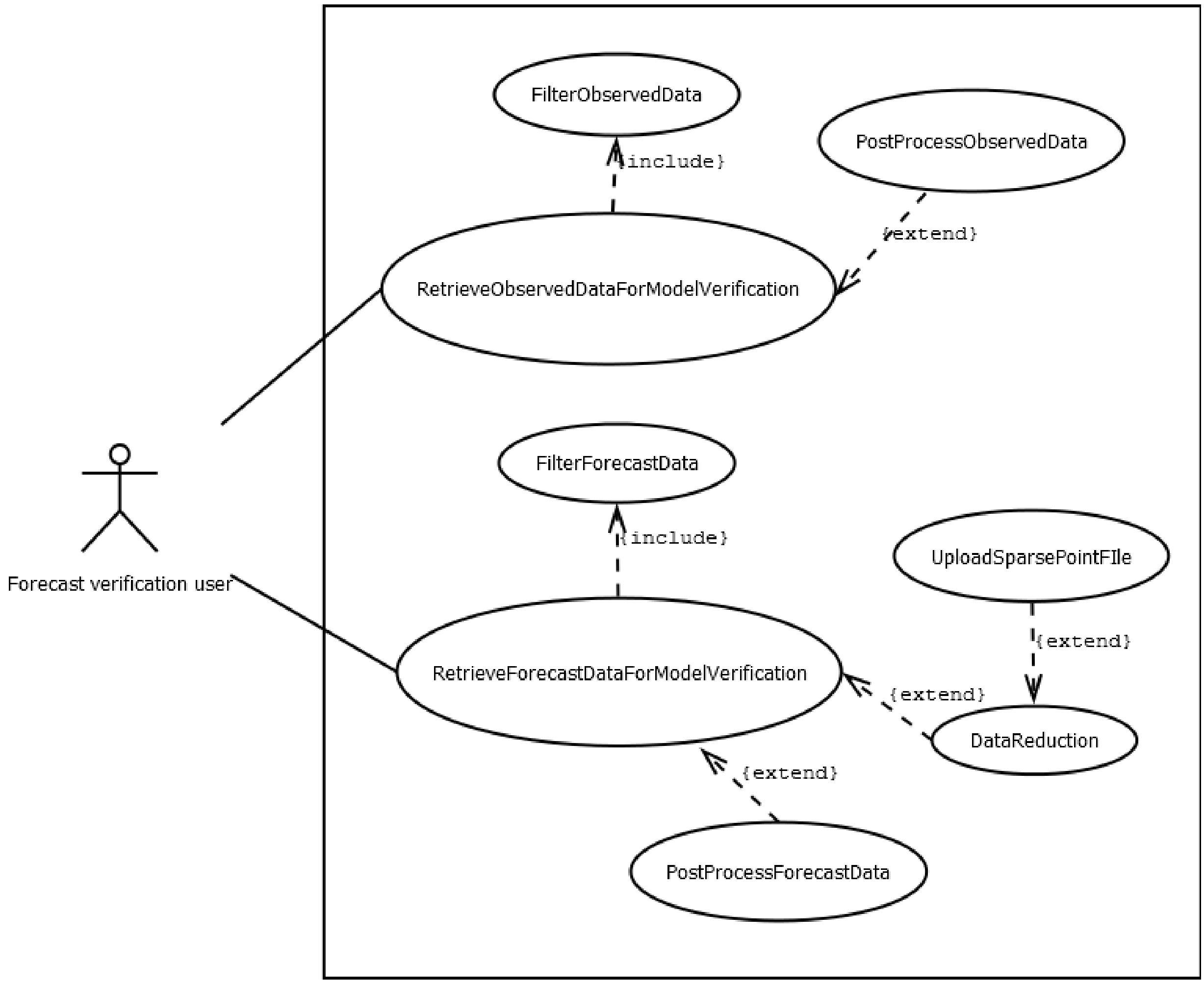

Figure 5 - UML diagram of UC3

3.3.1. UC3.1 - RetrieveObservedDataForModelVerification

Use case: RetrieveObservedDataForModelVerification
ID: UC3.1
Actor(s): Forecast verification user
Pre-conditions: 1. the Forecast verification user is an authenticated user of the system and authorized to download observed data
Description: • The use case starts when the user sends a query for quality controlled observed data for model verification to the system with mandatory filter parameters (see UC9.1) and optionally requires a post-processing (see UC9.3) • The system puts the user's request into an elaboration queue with a specific priority • The system takes the request from the queue and executes the query on the local database, applying filter. The system also applies post-processing, if requested by the user. • The system sends back to the user the data resulting from the query
Post-conditions: All data have been delivered

3.3.2. UC3.2 - RetrieveForecastDataForModelVerification

Use case: RetrieveForecastDataForModelVerification
ID: UC3.2
Actor(s): Forecast verification user
Pre-conditions: 1. the Forecast verification user is an authenticated user of the system and authorized to download gridded forecast data
Description: • The use case starts when the user sends a query for gridded forecast data for model verification to the system with mandatory filter parameters (see UC9.2)

and optional post-processing parameters for surface verification (see UC9.4). For upper-air verification the user could require a post processing data reduction (see UC9.5) • The system puts the user's request into an elaboration queue with a specific priority • The system takes the request from the queue and executes the query on the local database, applying filter specified by the user. The system also applies post-processing, if requested by the user. • The system sends back to the user the data resulting from the query
Post-conditions: All data have been delivered

3.4. UC4 - Redistribution of observed data to external users

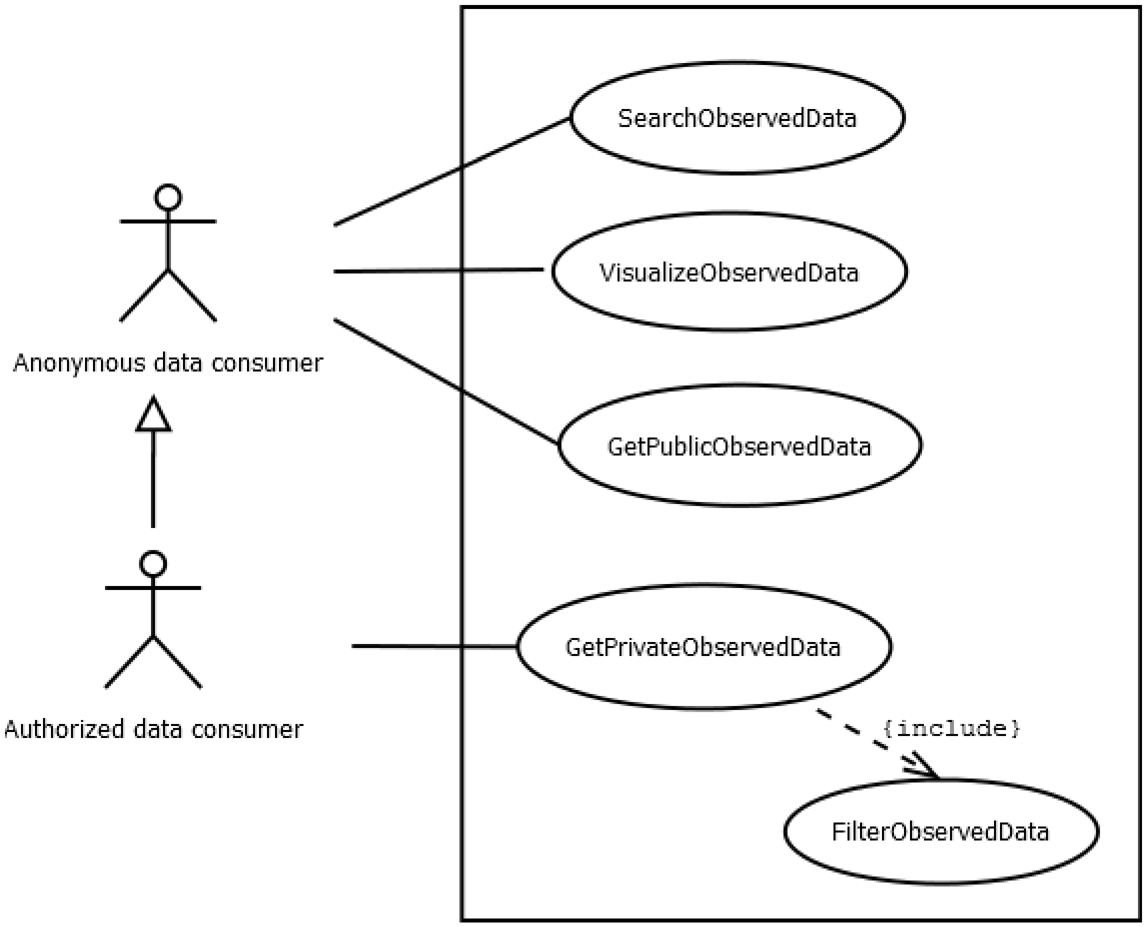

FIGURE 6 - UML DIAGRAM OF UC4

3.4.1. UC4.1 – SearchObservedData

Use case: SearchObservedData
ID: UC4.1
Actor(s): Anonymous data consumer
Pre-conditions: 1. The open data catalog is available in the system

Description:
• User accesses the catalog
• The system shows available datasets and facets to filter data by:
  ○ Organizations
  ○ Categories
  ○ Tags
  ○ Formats
  ○ Licenses
• User chooses one or more facets or free search dataset
• The system shows the results
• User visualizes the dataset (metadata) and accesses one or more resource (link do data). If data are public, the user can also download/visualize the data.

Post-conditions: The dataset and resources are visualized in the system

3.4.2. UC4.2 – VisualizeObservedData

Use case: VisualizeObservedData
ID: UC4.2
Actor(s): Anonymous data consumer
Pre-conditions: 1. The observed data is available in the system
Description: • The user accesses the map of observed data

• The system shows a map with marker clustering regarding a default parameter (e.g. temperature)
• The user can
  ○ change parameter
  ○ navigate the map (move, zoom in / out)
  ○ select a single marker
• The system shows a "time series" of a certain observed parameter selected by the user; the system shows a graph of one day values
• The user can change the interval of time (more/less value, shift the interval before/after)

Post-conditions: Values for an observed data parameter are visualized on a map/graph

3.4.3. UC4.3 – GetPublicObservedData

Use case: GetPublicObservedData
ID: UC4.3
Actor(s): Anonymous data consumer
Pre-conditions: 1. The observed data are available in the system
Description: • User accesses the system • The system shows public available streams of real time observed data with a limited persistence window (e.g. 15 days) • User chooses a stream and gets the data (locally filtering and formatting the data after they have been received)
Post-conditions: All data expected for the data consumer have been delivered

3.4.4. UC4.4 – GetPrivateObservedData

Use case: GetPrivateObservedData
ID: UC4.4
Actor(s): Authorized data consumer

Pre-conditions: 1. The archived observed data are available in the system 2. User is authenticated
Description: • User accesses the system • The system shows available data, which include both historical data and recent "near-real time" data • User sets mandatory filter parameters (see UC9.1) • The system shows the estimated size of the result • The user confirms the request. • The system puts the user's request into an elaboration queue • The system executes the requested elaboration • The user, optionally, checks the progress of the elaboration • The user downloads data
Post-conditions: All data expected for the data consumer have been delivered

3.5. UC5 - Send/Download gridded or sparse points forecast data

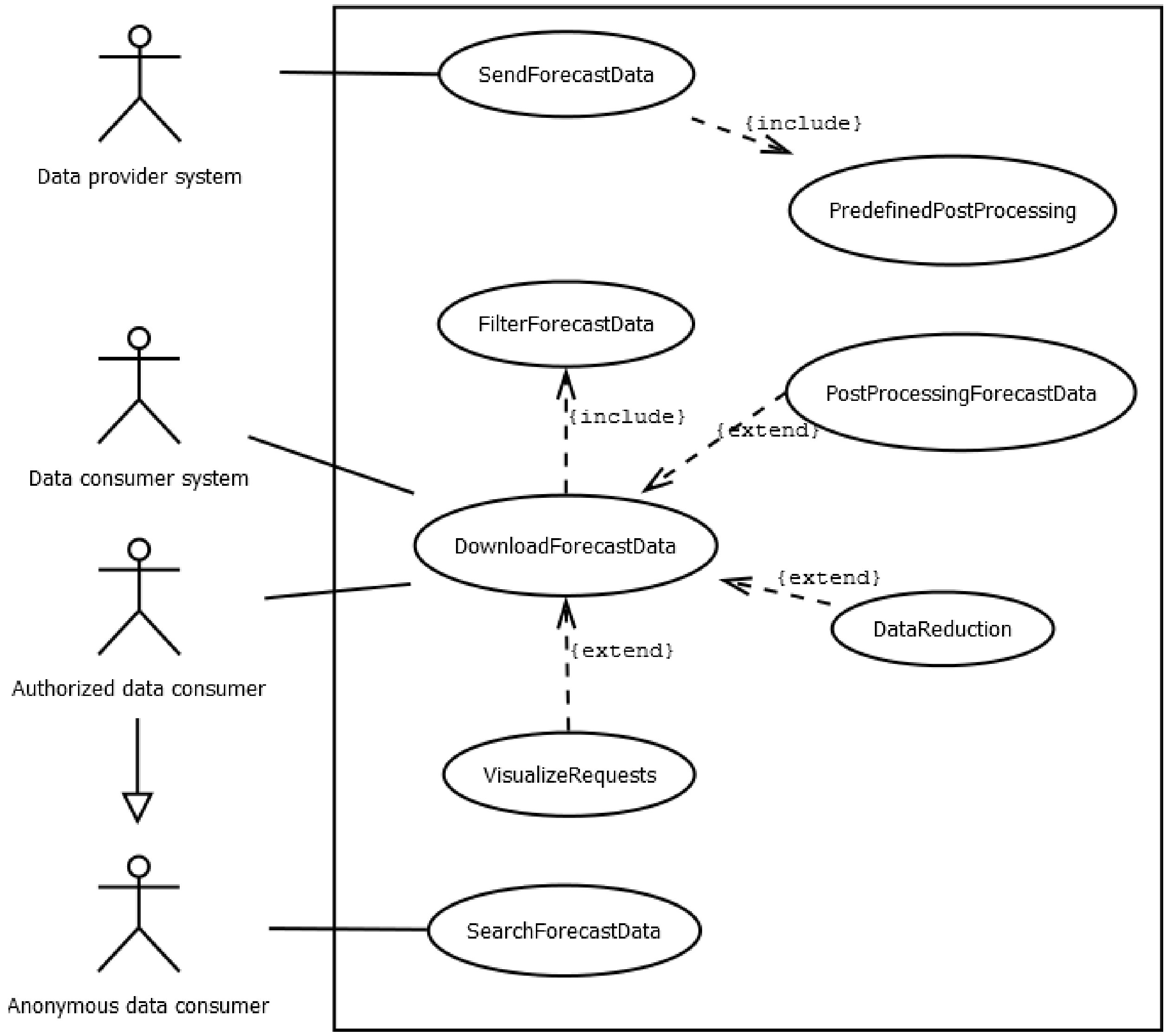

FIGURE 7 - UML DIAGRAM OF UC5

3.5.1. UC5.1 – SendForecastData

Use case: SendForecastData
ID: UC5.1
Actor(s): Data provider system
Pre-conditions: 1. The data provider is authenticated and authorized to send the forecast data
Description: • Data provider system sends the forecast data at regular intervals. Data are expressed in GRIB format • The system checks if it received all the data for the specific run • The forecast data will be saved in the system • The system run a predefined post-processing (see UC5.2).
Post-conditions: 1. The forecast data for that run are available

3.5.2. UC5.2 – PredefinedPostProcessing

Use case: PredefinedPostProcessing
ID: UC5.2
Actor(s): Data provider system
Pre-conditions: 1. forecast “raw” data fully transmitted

Description:
• The use case starts when the system receives completion notification of that particular "run"
• The post-processing could be:
    ○ time post-processing (e.g. accumulate precipitation or computation of maximum and minimum temperatures on specified time intervals)
    ○ derived variable computation (e.g. relative humidity)

Post-conditions:

3.5.3. UC5.3 – DownloadForecastData

Use case: DownloadForecastData
ID: UC5.3
Actor(s): Authorized data consumer
Pre-conditions: 1. The forecast data have been saved in the system
Description: • User accesses the system • The system shows the available filters • The user sets a filter (see UC9.2) • The user, optionally, chooses a post-process elaboration on the result (see UC9.4). • The system shows the estimated size of the result. • The user confirms the request. • The system puts the user's request into an elaboration queue • The system executes the requested elaboration • The user, optionally, checks the progress of the elaboration (see UC5.4) • The user downloads data • The extracted data will be maintained for a finite period of time
Post-conditions:

3.5.4. UC5.4 – VisualizeRequests

Use case: VisualizeRequests
ID: UC5.4
Actor(s): Authorized data consumer
Pre-conditions:

Description:
• The use case starts when the user asks for download requests
• The system shows a list of requests with the following info
  ○ date
  ○ query as a list of filters
  ○ status

○ size • The system allows to download the data related to the request (if status is completed)
Post-conditions:

3.5.5. UC5.5 – SearchForecastData

Use case: SearchForecastData
ID: UC5.5
Actor(s): Anonymous data consumer
Pre-conditions: 2. The open data catalog is available in the system
Description: • User accesses the catalog • The system shows available datasets and facets to filter data by: ○ Organizations ○ Categories ○ Tags ○ Formats ○ Licenses • User chooses one or more facets or free search dataset • The system shows the results • User visualizes the dataset (metadata) and accesses one or more resource (link do data). If data are public, the user can also download/visualize the data.
Post-conditions: The dataset and resources are visualized in the system

3.6. UC6 - Plot gridded forecast and observed data

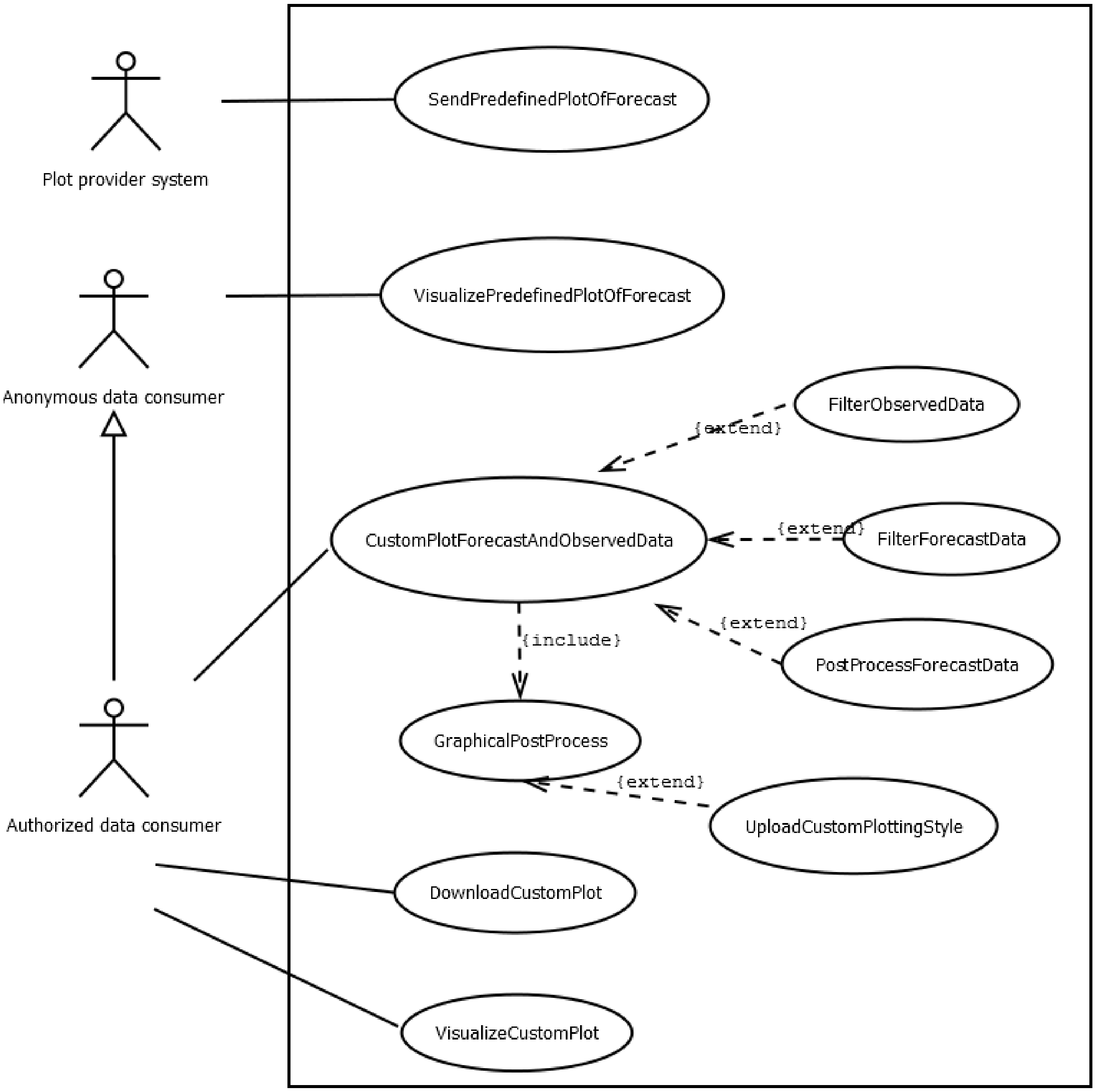

FIGURE 8 - UML DIAGRAM OF UC6

3.6.1. UC6.1 – SendPredefinedPlotOfForecast

Use case: SendPredefinedPlotOfForecast
ID: UC6.1
Actor(s): Plot provider system
Pre-conditions: 1. The plot provider system is authorized to send the predefined plots
Description: • The use case starts when the plot provider system sends the predefined plot of forecast data to Mistral system • The system saves the predefined plots
Post-conditions: All plots have been saved

3.6.2. UC6.2 – VisualizePredefinedPlotOfForecast

Use case: VisualizePredefinedPlotOfForecast
ID: UC6.2
Actor(s): Anonymous data consumer
Pre-conditions: The predefined plots of forecast data are available
Description: • The use case starts when the user accesses the map application to visualize the predefined plots of forecast • The system shows a map (predefined plot) and the following parameters: ○ Fields: ▪ accumulated total prec. 3h and 6h ▪ Temperature at 2m ▪ Wind at 10 meters ▪ Cloud coverage (%) ▪ Cloud coverage, high, medium, low (%) ▪ Relative humidity ▪ Accumulated total snow prec. 3h and 6h ○ Run (00, 12) ○ Resolution (2.2 km, 5 km) ○ Area:

▪ Italy ▪ Mediterranean Region ▪ Northern Italy ▪ Central Italy ▪ Southern Italy ▪ (Regional maps) • The system show also a slideshow • The user can click on slideshow or can change the parameters • The system show the corresponding map
Post-conditions: The system displays the plots requested

3.6.3. UC6.3 – CustomPlotForecastAndObservedData

Use case: CustomPlotForecastAndObservedData
ID: UC6.3
Actor(s): Authorized data consumer
Pre-conditions: The forecast and observed data saved in the system
Description: • The use case starts when the user sends a query for forecast or observed data (near real-time or past archived data) to be plotted specifying: ○ mandatory filter parameters (see UC9.1 and UC9.2) ○ optional post-processing parameters (see UC9.4) ○ plotting style (see UC6.4) • The system puts the user's request into an elaboration queue with a specific priority • The system takes the request from the queue and executes it (applies filter and executes the post-processes specified by the user) • The system informs the user that the resulting files are ready
Post-conditions: All custom plots are available for download or visualisation online

3.6.4. UC6.4 – GraphicalPostProcess

Use case: GraphicalPostProcess
ID: UC6.4
Actor(s): Authorized data consumer
Pre-conditions: 1. The set of required forecast map tipologies are available
Description: • The user chooses a plotting style to be applied (field contour and shading, wind vector or barbs, other point symbols) or upload a custom style file (see UC6.5) • The user chooses the output format ○ raster map ○ two dimensional georeferenced vector maps
Post-conditions: The graphical post process (style and format) are applied to data

3.6.5. UC6.5 – UploadPlottingStyle

Use case: UploadPlottingStyle
ID: UC6.5
Actor(s): Authorized data consumer
Pre-conditions:
Description: • The user upload a plotting style file • The system validate the file and save it
Post-conditions: The custom plotting style is applied

3.6.6. UC6.6 – DownloadCustomPlot

Use case: DownloadCustomPlot
ID: UC6.6
Actor(s): Authorized data consumer
Pre-conditions:
Description: • The user download the custom plot and after ask the system to delete it • The system delete the custom plot. After a period of time configured in the system the plot will be removed automatically
Post-conditions: The custom plotting has been downloaded and deleted from the system

3.6.7. UC6.7 – VisualizeCustomPlot

Use case: VisualizeCustomPlot
ID: UC6.7
Actor(s): Authorized data consumer
Pre-conditions:
Description: • The user visualizes the custom plot in the map application • The system shows a map with the custom plot requested by the user and shows the parameters (fields, run, resolution, area). The custom plot is a new field in addition to the one predefined (see UC6.2) • The system shows also a slideshow • The user can click on slideshow or can change the parameters • The system shows the corresponding map • The user asks the system to delete the custom plot • The system deletes the custom plot. After a period of time configured in the system the plot will be removed automatically

Post-conditions: The custom plotting has been visualized and then removed

3.7. UC7 - Italy Flash Flood

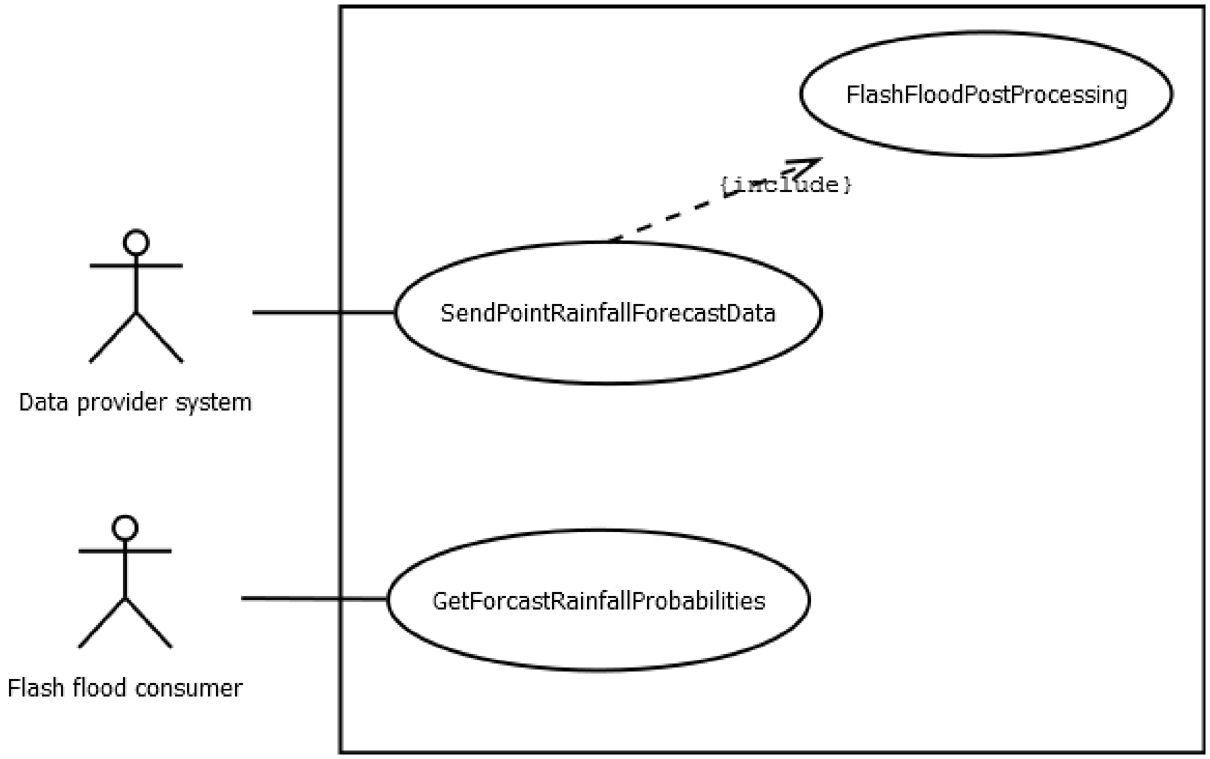

FIGURE 9 - UML DIAGRAM OF UC7

3.7.1. UC7.1 – SendPointRainfallForecastData

Use case: SendPointRainfallForecastData
ID: UC7.1
Actor(s): Data provider system
Pre-conditions: 1. The data provider system is authorized to send rainfall forecast data

Description: • The use case starts when the data provider system sends rainfall forecast data to Mistral system • The system applies a post processing (see UC7.2) • Forecast rainfall probabilities have saved from raw ensemble forecast and after the post-processing.
Post-conditions: Forecast rainfall probabilities have been saved

3.7.2. UC7.2 – FlashFloodPostProcessing

Use case: FlashFloodPostProcessing
ID: UC7.2
Actor(s): Data provider system
Pre-conditions:

Description:
• The system applies a post processing to ECMWF point-rainfall and to 2.2km COSMO ensemble output. The outputs will be blended to provide forecast rainfall probabilities from small totals right through to the extreme amounts which can lead to devastating flash floods.
• The system delivers forecast rainfall probabilities with 6-h temporal resolution, to a lead time limit of 120 to 240h. The final product comprises:
  ○ probability percentiles, for accumulated rainfall, for each COSMO grid point (percentiles 1,2,..99).
  ○ exceedance probabilities for user-defined thresholds.

Post-conditions: Forecast rainfall probabilities have been saved from raw ensemble forecast and after the post-processing.

3.7.3. UC7.3 – GetForecastRainfallProbabilities

Use case:
ID: UC7.3
Actor(s): Flash flood consumer

Pre-conditions: The flash flood consumer is authenticated and authorized to get the forecast rainfall probabilities

Description:
• The flash flood consumer gets the forecast rainfall probabilities:
  ○ probability percentiles, for accumulated rainfall, for each COSMO grid point (percentiles 1,2,..99).
  ○ exceedance probabilities for user-defined thresholds.

Post-conditions: Forecast rainfall probabilities have been delivered

3.8. Shared use cases

3.8.1. UC9.1 – FilterObservedData

Use case: FilterObservedData
ID: UC9.1
Actor(s): Data assimilation system, Forecast verification user, Authorized data consumer
Pre-conditions:

Description:
 	
• The query criteria to filter observed data could be:
  ○ a specified list of datasets (observational networks)
  ○ reference time between *t1* and *t2.*
  ○ geographical coordinates lying in a specified lon/lat bounding box
  ○ "last modification time" attribute greater than t3 (to downloads only the data that have been recorded or updated after the previous data access)
  ○ quality controlled data

Post-conditions:

3.8.2. UC9.2 – FilterForecastData

Use case: FilterForecastData
ID: UC9.2
Actor(s): Forecast verification user, Authorized data consumer
Pre-conditions:

Description:
• The typical archive query criteria for gridded forecast data for model verification could be:
  ○ a specified dataset (model run)
  ○ reference time between *t1* and *t2*
  ○ a subset of forecast timeranges (e.g. timeranges modulo 3 hours, 6 hours)
  ○ a specified list of variables
  ○ a specified list of vertical levels
  ○ full model grid (no space transformations)

Post-conditions:

3.8.3. UC9.3 - PostProcessingObservedData

Use case: PostProcessingObservedData
ID: UC9.3
Actor(s): Forecast verification user
Pre-conditions:

Description:
• The typical post-processing could be:
  ○ time postprocessing (e.g. accumulate precipitation on a specified time interval)
  ○ derived variable computation (e.g. relative humidity).

Post-conditions:

3.8.4. UC9.4 – PostProcessingForecastData

Use case: PostProcessingForecastData

ID: UC9.4
Actor(s): Forecast verification user
Pre-conditions:

Description:
• The typical post-processing for gridded forecast data could be:
    ○ optional time post-processing (e.g. accumulate precipitation on a specified time interval, computation of maximum and minimum temperatures on specified time intervals)
    ○ optional derived variable computation (e.g. relative humidity from temperature and dew-point temperature, wind speed).

Post-conditions:

3.8.5. UC9.5 – DataReduction

Use case: DataReduction
ID: UC9.5
Actor(s): Forecast verification user
Pre-conditions:

Description:
• The user chooses to post-processing the gridded forecast data defining the following parameters:
    ○ list of geographical points (order of tenths). The user inserts them or upload a georeferenced vector file (see UC9.5)
    ○ interpolation methodology (e.g. bilinear, nearest point, area average)
    ○ output format (e.g. BUFR, JSON, CREX)
• The system applies a spatial postprocessing, horizontal interpolation of the data on the list of user-requested geographical points

Post-conditions:

3.8.6. UC9.6 - UploadSparsePointFile

Use case: UploadSparsePointFile
ID: UC9.6
Actor(s): Forecast verification user
Pre-conditions:
Description: • The user upload georeferenced vector file (shapefile/geojson) with points in lat/lonWGS84 (EPSG 4326). • The system validates the file
Post-conditions:

Annex I. Glossary

ARPA: it is the acronym of "Agenzia Regionale per la Protezione dell'Ambiente" (Regional agency for environmental protection). ARPA are institutions of the Italian public administration, managed by the Italian Regions.

Arpae-SIMC: It is the "Servizio Idro-Meteo-Clima" (service of hydrology, meteorology and climatology) of the Regional Agency for the Environmental Protection of Emilia Romagna region, that carries on numerical Weather Predictions activities at national level and owns a huge monitoring station network.

BUFR: It is a binary data format maintained by the World Meteorological Organization (WMO). See https://en.wikipedia.org/wiki/BUFR.

DPC: It is the Italian Civil Protection Department that steers the National Service of Civil Protection and it has a guiding role, in cooperation with regional and local governments, for risk prevention, forecasting and monitoring activities as well as for emergency preparedness and for intervention procedures in cases of ongoing or upcoming crisis events.

COSMO: The COSMO Model is a non hydrostatic limited-are**a** atmospheric prediction model. See http://cosmo-model.org/content/model/general/default.htm

CREX: It is the name of a character code for the representation and exchange of meteorological and other data

ECMWF: The European Centre for Medium-Range Weather Forecasts (ECMWF) is an independent intergovernmental organisation supported by 34 states. ECMWF is both a research institute and a 24/7 operational service, producing and disseminating numerical weather predictions to its Member States.

GRIB: (GRIdded Binary or General Regularly-distributed Information in Binary form) is a concise data format commonly used in meteorology to store historical and forecast weather data. It is standardized by the World Meteorological Organization. See https://en.wikipedia.org/wiki/GRIB

HPC: high performance computing.

JSON: JSON (JavaScript Object Notation) is a lightweight data-interchange format.

LAMI: It is an agreement between ARPA Piemonte, Arpae-SIMC Emilia Romagna and Italian National Air Force Weather Service (AM) for the exploitation of the COSMO model. The Italian National Civil Protection Department (DPCN) assigned to the signers of the LAMI agreement the task to routinely run a state of the art Numerical Weather Prediction model suite to the best of their knowledge and to provide the results in graphical and numerical form to DPCN and to Civil Protection centres in the Italian territory.

MeteoNetwork: no profit association with the task of spreading knowledge in meteorology and climatology, also through an extensive network of amateur weather stations.