

Deliverable 4.1b

Catalogue of observed and forecast data suitable for Open Data Portal - Part b: Catalogue of forecast data

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



Document Control Page

Title	D4.1b Catalogue of observed and forecast data suitable for Open Data Portal - Part b: Catalogue of forecast data
Creator	ARPAE
Publisher	Mistral Consortium
Contributors	Cinzia Caroli (CINECA), Davide Cesari (ARPAE) Luca Delli Passeri (DPC), Estibaliz Gascón (ECMWF)
Type	Report
Language	en-GB
Rights	copyright "Mistral Consortium"
Audience	<input checked="" type="checkbox"/> public <input type="checkbox"/> restricted
Requested deadline	28/02/2019

Contents

Contents	3
Executive Summary	4
1. Catalogue of forecast data	5
1.1. Lami operational model suites	5
1.2. SPHERA reanalysis.....	6
1.3. ECMWF Point Rainfall	6
1.4. DPCN Radar composite.....	6
1.5. Summary of the datasets	6
1.6. Other NWP model suites	8
Annex I. Glossary	9

Executive Summary

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

For this purpose Mistral platform will collect the meteorological observed and forecast data at Italian level coming from the Regional Agencies (including ARPAE and ARPAP and those of the Civil Protection) in a centralized digital archive. The access procedures will be ensured to the weather community and to external users with different grants according to their role. This centralised archive will be the base to build a National Meteorological Open Data Portal.

In this part b of the deliverable, the meteorological forecast data that will be possibly archived in the Mistral platform are described, including the Numerical Weather Prediction data produced under LAMI agreement and by other providers in Italy, as well as data produced by ECMWF Point Rainfall post-processing.

Furthermore, the Annex I provides a glossary.

1. Catalogue of forecast data

The main source of meteorological forecast data in the Mistral portal is planned to be the set of Numerical Weather Prediction (NWP) modelling suites run under the LAMI agreement. This section examines the LAMI data, together with other sources of data suitable for inclusion in the data portal.

1.1. Lami operational model suites

LAMI is an agreement between ARPA Piemonte, Arpa-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. This activity is financed by DPCN; the configuration of the model and modelling suites are decided by the LAMI partners, Arpa-SIMC is in charge of setting up and managing part of the LAMI suites; Arpa-SIMC has assigned to CINECA Supercomputing Centre, by means of a public tender, the task to provide the computing and human resources for implementing, running and monitoring part of the LAMI NWP suites.

The LAMI NWP suites are currently based on the COSMO model and include the following model runs:

- **COSMO-5M** performs a forecast on the Mediterranean area up to 72 hours with a grid step of 5 km, initialised twice a day by AM analysis and using ECMWF IFS boundary conditions. An additional continuous assimilation run with nudging technique is also performed in order to have a continuous analysis dataset.
- **COSMO-2I** analysis and forecast on the Italian area with a grid step of 2.2 km. It is split into four main runs:
- **COSMO-2I-assim** performs a quasi-continuous data assimilation with the LETKF technique using the KENDA ensemble data assimilation system of COSMO. It includes 40 ensemble members and it runs in forecast/update steps of one hour. It uses COSMO ensemble of AM and COSMO-5M as boundary conditions.
- **COSMO-2I-fcast** performs a forecast up to 48 hours, initialised twice a day by COSMO-2I-assim and using COSMO-5M boundary conditions
- **COSMO-2I-fcruc** performs a forecast up to 18 hours, initialised eight times a day by COSMO-2I-assim and using COSMO-5M boundary conditions, for very short range prediction and now-casting

- **COSMO-2I-fcens** performs a daily ensemble forecast with 20 ensemble members, up to 51 hours, initialised by COSMO-2I-assim and using COSMO ensemble of AM as boundary conditions.

1.2. SPHERA reanalysis

In addition to LAMI operational runs, Arpae-SIMC is producing a long term reanalysis dataset, named SPHERA, based on a model configuration analogous to COSMO-2I, using ECMWF ERA5 reanalysis as boundary conditions and the nudging technique in order to assimilate observational data. It is planned that the final dataset will cover 25 years.

1.3. ECMWF Point Rainfall

Another noteworthy dataset which is going to be hosted in the Mistral Portal is the result of ECMWF Point Rainfall post-processing applied to ECMWF IFS and the blending of this post-processed output with post-processed COSMO-2I-fcens ensemble system output, as described in the corresponding use case. The outputs will consist of a representation of forecast rainfall probabilities.

1.4. DPCN Radar composite

Strictly speaking, this is an observational dataset, however, being it defined on a rectangular grid, its characteristics are similar to NWP model data. The dataset includes estimates of instantaneous rain rate and of accumulated rainfall, based on real time data from the Italian weather radar network. The original source for this dataset is in hdf format following ODIM convention, which is a common format in the Weather Radar community, on a projected Mercator grid. As an aid to the numerical modelling community, it is considered to store a version of this dataset in GRIB format, re-gridded on a regular lat-lon projection. This may be useful for the purposes of assimilating radar data, as well as for forecast verification and calibration.

1.5. Summary of the datasets

The following table summarizes the datasets produced by the different procedures and underlines the storage required for each of them in view of their possible archiving in the Mistral infrastructure.

Name	hor. grid step	grid size	full daily output	long term yearly output	emission interval	interval of forecast steps	forecast range
COSMO-5M	5km	1083x559x45	170GB	1.3TB ¹	12h	1h	72h
COSMO-2I-assim	2.2km	576x701x65	24GB	- ²	1h	-	-
COSMO-2I-fcast	2.2km	576x701x65	70GB	3.8TB ³	12h	1h	48h
COSMO-2I-fcruc	2.2km	576x701x65	120GB	-	3h	1h	18h
COSMO-2I-fcens	2.2km	576x701x65	19GB	3.5TB ⁴	24h	1h	51h
SPHERA	2.2km	622x747x65	10GB	4TB	1h	-	-
Point rainfall IFS	18km	~530x330	22GB (European grid output)	8TB (European grid output)	12h	3h&6h&12h	240h
Point rainfall blended with COSMO	2.2km	~500x700	2GB	730GB	24h	1h	51h
DPCN Radar grib	1km	~1200x1400	300MB ⁵	110GB	15'	-	-

The output has been divided into a full output, and a long term output. The full output, which includes all the raw model output, is meant to be kept for a period of about one year in order to allow verification of all the aspects of model results, reproduction of all the post-processed products and downstream model procedures, testing of new products for a long enough period.

The long term output is meant to be conserved for at least ten years or for the time during which the modelling system producing the data is considered to be reasonably state of the art and has not been superseded by another modelling system with at least a ten year record of reanalysis.

For most modelling runs, this long term output includes only the main surface fields and, possibly, fields on a few isobaric levels. This allows a verification of the main model fields, a reconstruction of the synoptic situation and the generation of a restricted amount of post-processed data for feeding a

¹ Assuming to keep only main data on surface levels and data on isobaric levels restricted on an Italian domain.

² No long term storage assuming that this can be replaced by the SPHERA dataset for climatic or scenario studies.

³ Assuming to keep only main data on surface levels on the full model domain.

⁴ Assuming to keep only surface fields of each ensemble member and discard fields on isobaric levels and probability products.

⁵ Estimated by assuming 50% of missing data in every fields and 16 bit coding.

subset of downstream modelling procedures. An exception is the reanalysis dataset, which is by definition a long term dataset and shall contain the full model output for all the period covered.

This long term archive is also important for forensic reasons, in case competent bodies may ask the forecasters to explain why some particular decision had been taken at a certain time in the past.

All the indicated data will be presumably in GRIB edition 1 or 2 format on a regular or rotated lat-lon projection.

1.6. Other NWP model suites

Other Italian operational modelling suites, currently executed in real time and made available to DPCN, based on Moloch, Bolam and WRF models and run by LAMMA Toscana, ARPA Liguria and ISAC-CNR, have been examined from the technical point of view. Also these models produce their final output on a regular or rotated lat-lon grid in GRIB 1 or 2 format, thus any tool developed for the LAMI models will be applicable, with minor adaptations, also to the output of these modelling suites.

Annex I. Glossary

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.

ARPA Liguria: It is the Regional Agency for the Environmental Protection of Liguria region. See <https://www.arpal.gov.it/>

ARPA Piemonte: It is the Regional Agency for the Environmental Protection of Piemonte region, that owns a huge monitoring station network, remote sensing instruments together with the deep expertise in weather data interpretation, data processing and model forecast verification

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

GRIB: 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>

ISAC-CNR: See <http://www.isac.cnr.it/en>

LAMMA: See <http://www.lamma.rete.toscana.it/en>