Meteo Italian SupercompuTing poRtAL

Deliverable

D4.2 Observed and Forecast Data Harmonization Specifications

Deliverable Lead:	Dedagroup
Deliverable due date	
Status	FINAL
Version	V1.0

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

Document Control Page

Title	D4.2 Observed and Forecast Data Harmonization Specifications
Creator	
Publisher	Mistral Consortium
Contributors	Martina Forconi (DEDAGROUP), Stefano Pezzi (DEDAGROUP)
Type	Report
Language	en-GB
Rights	copyright “Mistral Consortium”
Audience	☐ public ☐ restricted
Requested deadline	

Contents

Executive Summary

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

This deliverable describes the harmonization of observed and forecast data in terms of logical data model, data exchange formats and protocols used to send data to Mistral by data provider. More formats and protocols could be available for data consumer.

The logical data model is the one already adopted by Arpae that has been formalized in UML based on standard ISO 19156:2011 Geographic information Observations and measurements, OGC Sensor Observation Service, OGC SensorThings, Guidelines On Data Modelling For WMO Codes.

Data will be sent to Mistral by means of FTP or with publish and subscribe protocols (MQTT, AMQP), using data formats BUFR, JSON, ASCII DPC, GRIB.

1 Reference standards

The following data models have been useful in modelling Mistral entities:

ISO 19156:2011, Geographic information – Observations and measurements (https://www.iso.org/standard/32574.html)

OGC/IS 08-094r1 SWE Common Data Model (https://www.opengeospatial.org/standards/swecommon)

OGC/IS 15-078r6 SensorThings API (https://www.opengeospatial.org/standards/sensorthings)

Guidelines On Data Modelling For WMO Codes (https://wis.wmo.int/DataModel)

2 Terms and definitions

2.1 Sampling and measurement process

sampling

is the process of obtaining a discretized sequence of measurements of a quantity

measurement

process of experimentally obtaining one or more quantity values that can reasonably be attributed to a quantity (UNI CEI 70099:2008)

NOTE 1 Measurement does not apply to nominal properties.

NOTE 2 Measurement implies comparison of quantities and includes counting of entities.

NOTE 3 Measurement presupposes a description of the quantity commensurate with the intended use of a measurement result, a measurement procedure, and a calibrated measuring system operating according to the specified measurement procedure, including the measurement conditions.

sensor

element of a measuring system that is directly affected by a phenomenon, body, or substance carrying a quantity to be measured (UNI CEI 70099:2008).

observation

an observation (value of a quantity) is the result of the sampling process. In the context of series analysis, an observation is derived from a number of samples[1].

result of measurement

set of quantity values being attributed to a measurand together with any other available relevant information UNI CEI 70099:2008)

instrument calibration

operation that, under specified conditions, in a first step, establishes a relation between the quantity values with measurement uncertainties provided by measurement standards and corresponding indications with associated measurement uncertainties and, in a second step, uses this information to establish a relation for obtaining a measurement result from an indication (UNI CEI 70099:2008).

measuring transducer

device, used in measurement, that provides an output quantity having a specified relation to the input quantity; examples Thermocouple, electric current transformer, strain gauge, pH electrode, Bourdon tube, bimetallic strip.

atmospheric variables

atmospheric variables such as wind speed, temperature, pressure and humidity are four-dimensional functions - two horizontal, one vertical and one temporal. They vary irregularly in all four, and the purpose of the sampling study is to define practical measurement procedures to obtain representative observations with acceptable uncertainties in estimates of averages and variability.

2.2 Data level

In the discussion of the instrumentation associated with the measurement of atmospheric variables, it has become useful to classify the observational data according to data levels. This scheme was introduced in connection with the data-processing system for the Global Atmospheric Research Programme, and is defined in WMO (2010b, 2010c).[2]

Level I data are instrument readings expressed in appropriate physical units and referred to with geographical coordinates. Level I data themselves are in many cases obtained from the processing of electrical signals such as voltages, referred to as raw data. Examples of these data are satellite radiances and water-vapour pressure.

---

[1] https://library.wmo.int/doc_num.php?explnum_id=3206
[2] https://library.wmo.int/doc_num.php?explnum_id=3210

Level II data are recognized as meteorological variables (observations / measurements). They may be obtained directly from instruments (as is the case for many kinds of simple instruments) or derived from Level I data. For example, a sensor cannot measure visibility, which is a Level II quantity; instead, sensors measure the extinction coefficient, which is a Level I quantity.

Level III data are those contained in internally consistent datasets, generally in grid-point form.

Data exchanged internationally are Level II or Level III data.

2.3 Report

A report is a (synchronous) set of Level II observations complete with metadata or with the ability to reconstruct the metadata.

3 Logical Data Model

3.1 Mistral entities

The Logica Data Model is designed based on the ISO/OGC Observation and Measurement (O&M) model [OGC 10-004r3 and ISO 19156:2011]. O&M defines models for the exchange of information describing observation acts, their results as well as the feature involved in sampling when making observations.

The Mistral entities are depicted in Figure 1. The following sections describe the properties of each entity.

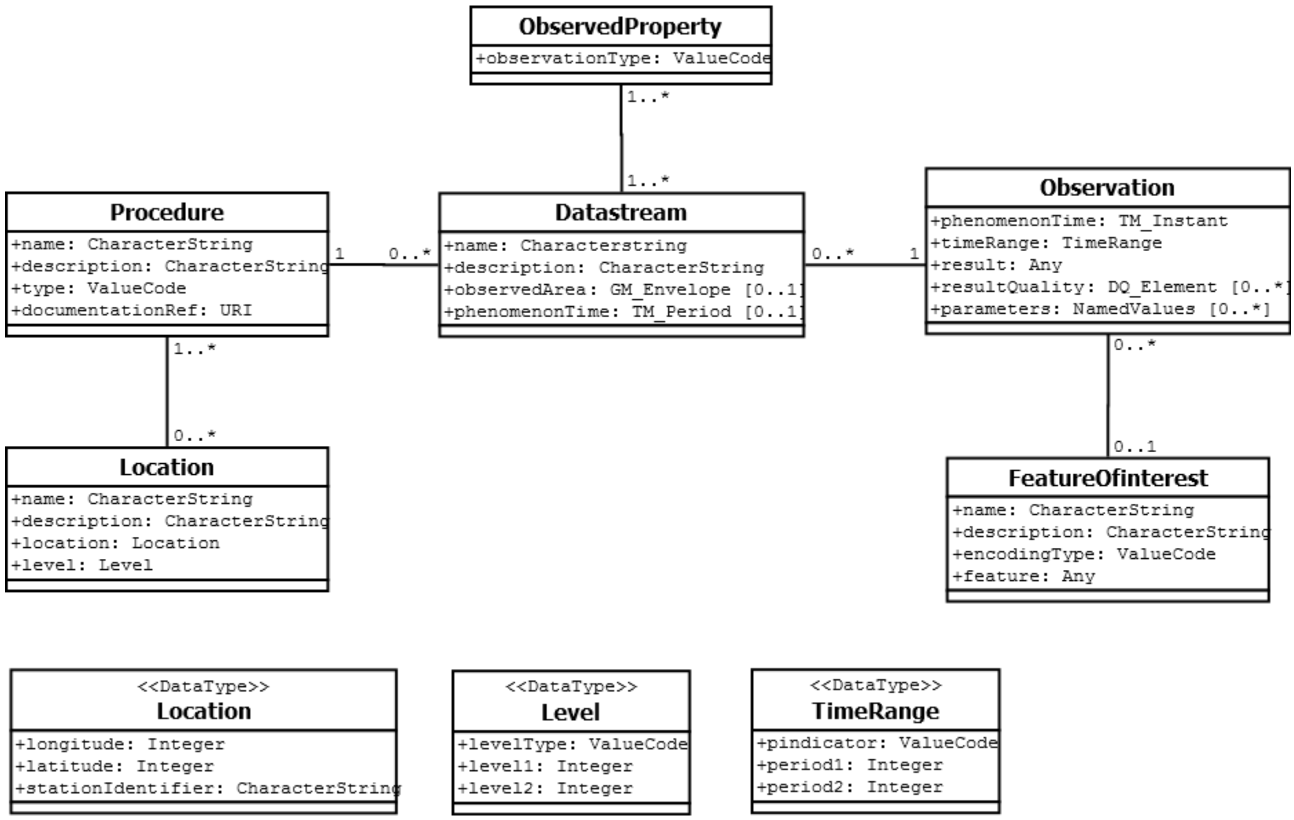


FIGURE 1- MISTRAL LOGICAL DATA MODEL

3.2 Observation

An Observation is the act of measuring or otherwise determining the value of a property [OGC 10-004r3 and ISO 19156:2011].

In O&M, an **observation** is defined as an event that results in an estimation of the value (the **result**) of a property (the **observed property**) of some entity (the **feature of interest**) using a specified **procedure**.

Notes: ISO 19156 "Observations and measurements" provides a conceptual schema for observations and the features involved in sampling when making observations, and specifically designed to support the exchange of information describing both the observing event and the results of the observation between different scientific and technical communities.

Whilst the name of the model invokes a particular concept to meteorologists (e.g. "observation", the measurement of physical phenomena with an instrument or sensor – disjoint from the concept of 'forecast') it is important to consider the semantics of the model. The class OM_Observation is defined as 'an estimate of the value of some property of something using some specified Process'. The process may be an instrument/sensor (directly) measuring some physical parameter or a numerical simulation predicting future values. **Thus, the Observations and measurements" conceptual model may be used to represent both observations and forecasts** [Guidelines On Data Modelling For WMO Codes].

OM_Observation includes three different time attributes. The **phenomenon time** is the time that the result applies to the property of the feature of interest. The **result time** is the time when the result became available. The **valid time** is the period during which the result is intended to be used.

Note: Some of the terminology (in particular, result time, phenomenon time, valid time) established in O&M differs from and conflicts with terminology previously established in WMO publications [Guidelines On Data Modelling For WMO Codes].

The Table 1 describe the time attributes of O&M and Mistral

TABLE 1 - TIME ATTRIBUTE IN O&M AND MISTRAL

O&M time attributes	Mistral time attributes	Comments
phenomenonTime	phenomenonTime	It is the time that the result applies to the property of the parameter of interest. The O&M defines phenomenonTime to be of type TM_Object (it can be either an instant in time or a span of time); Mistral restricts phenomenonTime to be of type TM_Instant
resultTime		Describe the time when the result became available
validTime		The time period during which the result may be used
	validity time	Time at which the data are measured, or at which forecast is valid; for statistically processed data, the validity time is the end of the time interval
	reference time	Nominal time of an observation for observed values, or as the time at which a model forecast starts for forecast values
	timeRange	The timeRange is composed by 3 attributes: a value code (e.g. instantaneous value, average, accumulation), difference between validity time and reference time, duration of statistical processing

TABLE 2 PROPERTIES OF A OBSERVATION ENTITY

Name	Definition
phenomenonTime	The time instant of when the Observation happens. If we are modelling an observation, then this is the time when the physical parameter was observed. If we are modelling a forecast, then this is the time when we expect events will happen. We can expect phenomenonTime to be in the past for observations, and in the future for forecasts. The O&M defines phenomenonTime to be of type TM_Object (it can be either an instant in time or a span of time); Mistral restricts phenomenonTime to be of type TM_Instant. One and only one phenomenonTime must appear in an Observation.
timeRange	The timeRange field is composed by: • pindicator: value code (e.g. instantaneous value, average, accumulation, …) • P1 is defined as the difference in seconds between validity time and reference time. For forecasts it is the positive forecast time. For observed values, the reference time is usually the same as the validity time, therefore P1 is zero • P2 is defined as the duration of the period over which statistical processing is performed and is always nonnegative. Note that, for instantaneous values, P2 is always zero.
result	The estimated value of a property determined through a known observation procedure (method, algorithm or instrument which may be used in making an observation). An Observation results in a value being assigned to a phenomenon. The phenomenon is a property of a FeatureOfInterest of the Observation [OGC and ISO 19156:2001].
resultQuality	Describes the quality of the result. ResultQuality is “instance-specific” and complements the description. ResultQuality should focus on the quality of the specific observation at hand. ISO 19157:2011 Geographic Information—Data quality provides a substantial toolset for modelling data quality. ResultQuality is of type DQ_Element. It may be omitted, and more than one DQ_Element may appear.
parameters	If present, the attributes parameter:NamedValue describe event-specific and/or contextual parameters. This might be an environmental parameter,

	an instrument setting or input, or an event-specific sampling parameter that is not tightly bound to either the feature-of-interest or to the observation procedure. The data type NamedValue provides a mechanism for modelling additional properties; a NamedValue has two attributes, a generic name called "name" and a "value" of type Any.

Parameters might be useful in representing WMO codes. The Table 3 (source: Guidelines on data modelling for WMO Codes) was created by considering Product Definition Template (PDT) 4.1 in GRIB2. Individual ensemble forecast, control and perturbed, at a horizontal level or in a horizontal layer at a point in time. For this illustration, assume we are encoding data for a single point that was generated by an Ensemble Forecast System (EFS).

TABLE 3 - PARAMETER EXAMPLE (SOURCE GUIDELINES ON DATA MODELLING FOR WMO CODES)

parameter.name	Data Type	GRIB2 Contents	Comments
analysis time	TM_Instant	reference time	Not modelled by existing O&M times.
data cutoff time	TM_Instant	Hours/minutes after reference time of data cutoff	Not modelled by existing O&M times.
ensemble forecast type	Code table	Type of ensemble forecast	Could also be modelled as part of om:procedure.
perturbation number	int	Perturbation number	
ensemble size	int	Number of forecasts in ensemble	

3.3 Datastream

A Datastream groups a collection of Observations measuring one or more ObservedProperty and produced by the same Procedure.

TABLE 4 - PROPERTIES OF A DATASTREAM ENTITY

Name	Definition
name	A label for Datastream entity, commonly a descriptive name
description	The description of the Datastream entity
observedArea	The spatial bounding box of the spatial extent of all FeaturesOfInterest that belong to the Observations associated with this Datastream.
phenomenonTime	The temporal interval of the phenomenon times of all observations belonging to this Datastream.

3.4 ObservedProperty

An ObservedProperty specifies the phenomenon of an Observation.

In O&M, an instance of ObservedProperty describe a property that is either assignable or observable, such as "temperature", "height", "colour", "material".

In Mistral will be used the properties listed in the WMO tables (BUFR Table B, GRIB1 Table 2, GRIB2 code table 4.2).

TABLE 5 - PROPERTIES OF A OBSERVEDPROPERTY ENTITY

Name	Definition
observationType	The type of Observation (with unique result type), which is used by the service to encode observations. For observations will be used the WMO BUFR Table B, that defines element description number, name, unit, scale and data width. For forecast will be used GRIB1 Table 2, GRIB2 code table 4.2

3.5 Procedure

The purpose of an observation procedure is to generate an observation result. An instance of Procedure could be an instrument or sensor, but may be a human observer, a simulator, or a process or algorithm applied to more primitive results used as inputs.

There are complex, strategies available for encoding procedure. ISO 19115-2:2009 “Geographic information—Metadata—Part 2: Extensions for imagery and gridded data” provides the following classes that can be useful for modelling the procedure: MI_Instrument, LE_Processing, and LE_Algorithm. The OGC standard named SensorML includes additional tools.

Simple strategies for encoding procedure include encode a URL that points to the description of the observing process; this approach is adopted in Mistral.

TABLE 6 - PROPERTIES OF A PROCEDURE ENTITY

Name	**Definition**
name	A descriptive name of the procedure
description	The description of the procedure
type	For observations defines stations with homogeneous characteristics: class of instruments, representativeness and / or mobile or fixed stations. For stations that do not belong to homogeneous networks with a managing body, the value of "network" follows the following rule: • "fixed" for all fixed stations, whose coordinates do not vary over time • "mobile" for all stations / measurement points whose coordinates change over time For forecast defines the model.
documentationRef	Reference to an external process definition providing information about relevant documentation that describes the associated Procedure.

3.6 *FeatureOfInterest*

Abstraction of real-world phenomena. It is the real-world object whose properties are under observation, or is a feature intended to sample the real-world object,

In the case of remote sensing the FeatureOfInterest can be the geographical area or volume that is being sensed.

TABLE 7 - PROPERTIES OF A FEATUREOFINTEREST ENTITY

Name	**Definition**

name	A label for FeatureOfInterest entity, commonly a descriptive name
description	The description of the FeatureOfInterest
encodingType	The encoding type of the feature property
feature	The detailed description of the feature. The data type is defined by encodingType

3.7 Location

The Location entity locates the Observation.

A Location may be identical to the FeatureOfInterest. For example, the content of the thermostat's location should be the same as the content of the temperature readings' feature of interest.

However, the content of a Location could be different from the content of the FeatureOfInterest (e.g., in the case of remote sensing).

TABLE 8 - PROPERTIES OF A LOCATION ENTITY

Name	**Definition**
name	A label for Location entity, commonly a descriptive name
description	The description of the Location
location	For observations the location is: • longitude: geographic coordinates (ETRF89 - WGS84) • latitude: geographical coordinates (ETRF89 - WGS84) • identifier: data provider / flight / ship identification (mandatory in the case of mobile stations)
level	Vertical coordinates (possible layer) in coded form composed by: • level type: value code from WMO GRIB2 where possible, • level1: first level defined by level type • level2: second level defined by level type

4 Data exchange formats

In this chapter are described the data exchange formats for the mistral data providers; additional formats could be defined for data consumers.

MISTRAL holds meteorological observations mainly using BUFR format and forecasts using GRIB format. CSV format (currently used by DPC) and JSON format are also accepted for observations.

4.1 BUFR

The Binary Universal Form for the Representation of meteorological data (BUFR) is a flexible binary format, mainly used to encode in situ and satellite observations, but can also represent forecast data,

The BUFR format is maintained by the World Meteorological Organization (WMO) and it is a **table-driven code format** (the meaning of data elements is determined by referring to a set of tables that are kept and maintained separately from the message itself).

The table-driven code format BUFR is **self-descriptive**, which means that the form and content of the data in a BUFR message are described within the message itself.

The term "message" refers to BUFR being used as a data transmission format; however, BUFR can, and is, used in a number of meteorological data processing centers as an on-line storage format as well as a data archiving format.[3]

4.1.1 BUFR Sections

For transmission of data, each BUFR message consists of a continuous binary stream comprising six sections as shown in the Figure 2 - BUFR sections.

[3] http://www.wmo.int/pages/prog/www/WMOCodes/Guides/BUFRCREX/Layer1-2-English.pdf

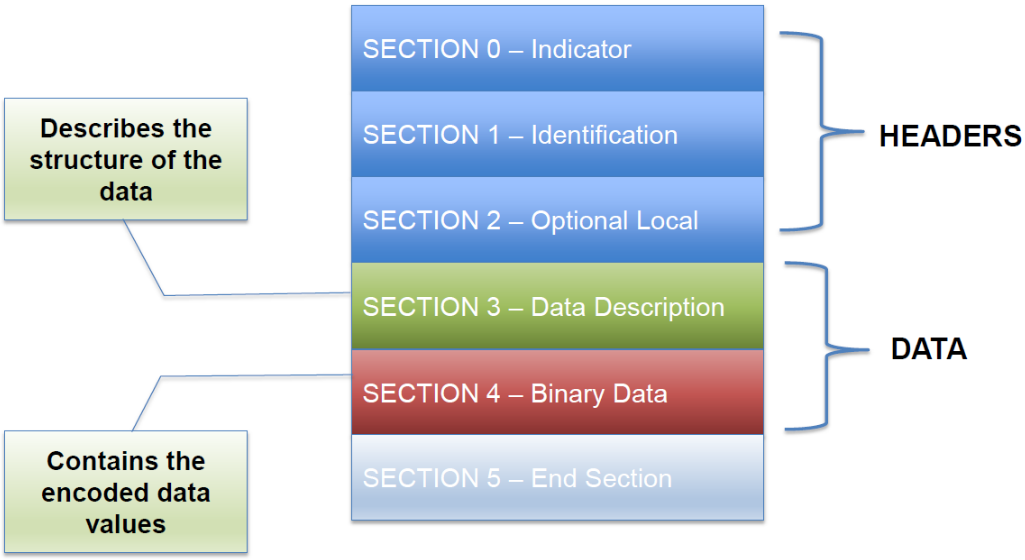


FIGURE 2 - BUFR SECTIONS[4]

The Indicator Section and the BUFR Identification Section are short sections, which identify the message.

The list of descriptors, pointing towards elements in predefined and internationally agreed tables that are stored in the official WMO Manual on Codes, are contained in the Data Description Section 3. These descriptors describe the type of data contained in the Section 4 and the order in which the data appear there. The Optional Section 2 can be used to transmit any information or parameters for national purpose. The End Section 5 contains the four alphanumeric characters "7777" to denote the end of the BUFR message.

4.1.2 BUFR Tables

BUFR is a table-driven system. Tables define how the parameters (or elements) shall be coded as data items in a BUFR message. They are recorded in the WMO PUBLICATION No. 306 - MANUAL ON CODES.

The tables defining BUFR coding are:

- **Table A – data category**: subdivides data into a number of discrete categories (e.g. Surface data – land, Surface data - sea) useful for storage and retrieval of data from a data base.
- **Table B – elements**; describes how individual parameters, or elements, are to be encoded and decoded in BUFR.
- **Table C - operators**; defines a number of operations that can be applied to the elements. Each such operation is assigned an operator descriptor.

[4] https://confluence.ecmwf.int/download/attachments/73011814/eccodes_bufr_in_a_nutshell.pdf

- **Table D – common sequences**; defines groups of elements that are always transmitted together (common sequence) to reduce the amount of space required for a BUFR message.
- **Code-flag table**: code tables define an element based on a code (e.g., Cloud Type). Flag table define an element based on a set of conditions defined by flags (bits set to 0 or 1). The code tables and flag tables are associated with elements in Table B.

4.1.3 BUFR Descriptors

A BUFR descriptor is a set of 16 bits, divided into 3 parts (**F-X-Y structure**).

TABLE 9 - BUFR DESCRIPTOR STRUCTURE

F	X	Y
2 (bits)	6 (bits)	8 (bits)

The F value (0 to 3) determines the **type of descriptor** listed in the following table.

TABLE 10 - BUFR DESCRIPTOR TYPES

F value	Descriptor type	Meaning of X and Y values
0	Element	Element descriptor (**Table B**). X indicates the class of descriptor. Y indicates the entry within a class X.
1	Replication	Repetition of a chosen number of descriptors. The X value specifies the number of following descriptors to be included in the replication; the Y value indicates how many times the replication is to take place. If Y = 0, then the replication is called a "delayed replication" and the number of replications is to be obtained from the value of a special element descriptor.
2	Operator	Operator descriptor (**Table C**). The X value identifies the operator and the Y value is used to control its application.
3	Sequence	Sequence descriptor (**Table D**). The use of the X and Y value is the same as with Element Descriptors

4.1.4 BUFR Table B

Table B describes how individual parameters, or elements, are to be encoded and decoded in BUFR. The descriptor number is **F = 0** (table B), **X = class of descriptor**, **Y = entry within a class**.

As an example, extracts of BUFR Table B for Temperature (class X = 12) is given below.

TABLE REFERENCE			TABLE ELEMENT NAME	BUFR			
				UNIT	SCALE	REFERENCE VALUE	DATA WIDTH (Bits)
F	X	Y					
0	12	001	Temperature/dry-bulb temperature	K	1	0	12
0	12	002	Wet-bulb temperature	K	1	0	12
0	12	003	Dew-point temperature	K	1	0	12
0	12	004	Dry-bulb temperature at 2 m	K	1	0	12
0	12	005	Wet-bulb temperature at 2 m	K	1	0	12
0	12	006	Dew-point temperature at 2 m	K	1	0	12
0	12	007	Virtual temperature	K	1	0	12
0	12	011	Maximum temperature, at height and over period specified	K	1	0	12
0	12	012	Minimum temperature, at height and over period specified	K	1	0	12

FIGURE 3 - BUFR TABLE B CLASS 12 TEMPERATURE

For each element, the table lists:

- reference number (element descriptor number)
- element name
- unit (or code/flag table)
- scale
- reference value
- data width (bits)

Where a code table or flag table is appropriate, code table or flag table, respectively is entered in the UNITS column. The code tables and flag tables associated with Table B are numbered to correspond with the F, X and Y part of the table reference.

To encode values in BUFR, the data (in the units as specified in the UNIT column) must be multiplied by 10 to the power of SCALE and then, the REFERENCE VALUE must be subtracted from them.

For example, a measured latitude is 44,51485 degrees. The high accuracy latitude descriptor is 0 05 001 (Table 11) and the encoded value is:

44,51485 x 100000 (scale 5) – (–9000000) = **13451485**

TABLE 11 - BUFR TABLE B LATITUDE

Reference number	Element name	Unit	Scale	Reference Value	Data width (bits)
0-05-001	Latitude (high accuracy)	Degree	5	-9000000	25

In BUFR table B classes 48 to 63 are reserved for local use; all other classes are reserved for future development. Entries 192 to 255 within all classes are reserved for local use.

4.1.5 Data and meta-data representation

Element descriptors Classes 1 through 9 have the special property of remaining in effect from the moment they appear throughout the remainder of the BUFR template, unless contradicted or canceled. Those descriptors convey data about subsequent data – in other words, meta-data. Classes 1 through 9 descriptors are used for spatial, temporal and other meta-data that is applicable to the data payload of the BUFR message.

TABLE 12 - BUFR ELEMENT DESCRIPTOR CLASSE 1 THROUGH 9

Class	Definition
01	Identification
02	Instrumentation
03	Reserved
04	Location (Time)
05	Location (horizontal – 1)
06	Location (horizontal – 2)
07	Location (vertical)
08	Significance Qualifiers
09	Reserved

Element Descriptors from other Classes do not have the context-setting effect of Classes 1-9 and do not affect the subsequent meaning of the message (although a few do have special properties).

The relation between Element Descriptors that represent basic data and those that represent meta-data is an essential structuring factor in BUFR templates. Class 1-9 descriptors can be reused in the template and given successive values, endowing a BUFR template with a progression in time, space, or even through other meta-properties. The meta-data elements relating to a single data element can be obtained by working backward from that element's position in the template.[5]

4.1.6 BUFR Templates in Mistral

A BUFR template is a sequence of BUFR descriptors in Section 3 that expresses the form and content of a BUFR data product and is recognized by the WMO or by a local authority as a canonical form of the product. Templates are designed to meet the requirements of a specific data product (weather observations, for instance) and to standardize its content and structure. Complex spatial, temporal and statistical relationships, involving any data and meta-data, can be expressed in the template.

[5] http://eumetnet.eu/wp-content/uploads/2017/04/OPERA_BUFR_template_primer_V1.4.pdf

In Mistral the following templates defined by the WMO and ECMWF are adopted:

- acars-ecmwf - ACARS ECMWF (4.145)
- acars-wmo - ACARS WMO
- airep-ecmwf - AIREP ECMWF (4.142)
- amdar-ecmwf - AMDAR ECMWF (4.144)
- amdar-wmo - AMDAR WMO
- buoy - Buoy (1.21)
- metar - Metar (0.140)
- pilot-ecmwf - Pilot (2.91)
- pilot-wmo - Pilot (2.1, 2.2, 2.3)
- pollution - Pollution (8.171)
- ship - Synop ship (autodetect)
- ship-abbr - Synop ship (abbreviated) (1.9)
- ship-auto - Synop ship (auto) (1.13)
- ship-plain - Synop ship (normal) (1.11)
- ship-reduced - Synop ship (reduced) (1.19)
- ship-second - Synop ship (second record) (1.12)
- ship-wmo - Ship WMO
- synop-ecmwf - Synop ECMWF (autodetect) (0.1)
- synop-ecmwf-auto - Synop ECMWF land auto (0.3)
- synop-ecmwf-land - Synop ECMWF land (0.1)
- synop-ecmwf-land-high - Synop ECMWF land high level station (0.1)
- synop-wmo - Synop WMO (0.1)
- temp-ecmwf - Temp ECMWF (autodetect)
- temp-ecmwf-land - Temp ECMWF land (2.101)
- temp-ecmwf-ship - Temp ECMWF ship (2.102)
- temp-radar - Temp radar doppler wind profile (6.1)
- temp-ship - Temp ship (autodetect)
- temp-wmo - Temp WMO (2.101)

4.1.7 Local tables

Since a data processing center may need to represent data conforming to a local requirement, and this data is not defined within Table B, specific areas of Table B and D are reserved for local use.

In BUFR table B classes 48 to 63 are reserved for local use; all other classes are reserved for future development. Entries 192 to 255 within all classes are reserved for local use

Local tables are defined by the originating Centre and can be used to exchange data by bilateral agreement. Local table is intended for local (not international) exchange.

In Mistral will be adoped the following local tables:

004192 [SIM] Time range type

007192 [SIM] First level type

007195 [SIM] Second level type

020193 [SIM] Cloud type (METAR)

4.2 JSON

JSON (JavaScript Object Notation) is a lightweight data-interchange format. It is easy for humans to read and write. It is easy for machines to parse and generate. JSON is "self-describing" and easy to understand.

In Mistral each json object is a report with all the data of a certain station for a certain instant of reference.

The station is uniquely identified by the following fields:

- ident: optional station identification (only necessary if the station is mobile, generally empty for fixed stations).
- lon: longitude
- lat: latitude
- network: name of the network to which the station belongs (lowercase).

The latitudes and longitudes must be written as geodetic coordinates expressed as an integer after multiplying by 10 ^ 5 (thus expressed in 10 ^ -5 decimal degrees).

The reference instant is the datetime field (ISO 8601) which refers to the final instant of the measurement. Consequently, a cumulative precipitation over 30 minutes with reference time "2018-08-05T12: 00: 00Z" is the cumulated precipitation between 11: 30: 00Z and 12: 00: 00Z of 2018-08-05.

The data is in the data field as an array. Each element of the array is an object with the following fields:

- level: vertical coordinates
- timerange: defines the time period and any processing (e.g. instant data, hourly average, etc).
- vars: object whose fields are the codes of the local table B. Each of these is associated with an object with the fields v (the value) and a (object of the data attributes, in which the fields are other codes of the table B to which a value is associated).

Among these, only one element does not have the level and time-range fields. These data relate to invariant data of the station itself (e.g. the name, height, etc.).

Example

Fixed station (ident: null) of the network “rer”located at point (9.15454, 44.51485) (lon, lat) with the following static data (element of the array "data" that has no "level" and "time-range" keys):

- Station name (B01019): Torriglia
- Station height (B07030): 769.0m
- Station barometric height (B07031): 769.0m

And for the reference time "2018-07-30T15: 30: 00Z" has recorded the following data:

On the ground (level: [1, null, null, null]) the following cumulative times (timerange: [1, 0, 3600]):

- Precipitation (B13011): 0.0

At 2m from the ground (level: [103, 2000, null, null]) the instantaneous values:

- Temperature (B12101): 297.15 K. The data has been manually invalidated (B33196: 1 attribute).
- Relative humidity (B13003): 50%


```
{
       "ident": null,
       "network": "rer",
       "lon": 915454,
       "date": "2018-07-30T15:30:00Z",
       "lat": 4451485,
       "data": [
           {
               "vars": {
                   "B01019": {
                       "v": "Torriglia"
                   },
                   "B07030": {
                       "v": 769.0
                   },
                   "B07031": {
                       "v": 769.0
                   }
               }
           },
           {
               "timerange": [
                   1,
                   0,
                   3600
               ],
               "vars": {
                   "B13011": {
                       "a": {
                       },
                       "v": 0.0
                   }
               },
               "level": [
                   1,
                   null,
                   null,
                   null
```

```
            ]
        },
        {
            "timerange": [
                254,
                0,
                0
            ],
            "vars": {
                "B12101": {
                    "a": {
                        "B33196": 1
                    },
                    "v": 297.15
                },
                "B13003": {
                    "a": {
                    },
                    "v": 50
                }
            },
            "level": [
                103,
                2000,
                null,
                null
            ]
        }
    ]
}
```


4.3 ASCII CSV DPC

Observation data of regional ground stations are in proprietary format; most of them in DVD format by CAE[6], some in ETG by ETG[7].

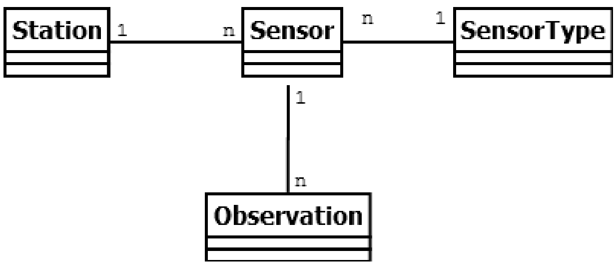
Observation data from regional ground stations are collected by DPC which produces a “national” file in csv format, every 20 minutes, that contains observations for a given day. This file is accompanied by weather stations and sensors files. The following table shows all the files; the number of elements of each file is taken from observations of 12/06/2019 until 10:36.

TABLE 13 - LIST OF DPC OBSERVATION FILES

File name	Description	Format	Elements
<YYYYMMDD>_dvd	Daily observations	CSV	17980

[6] https://www.cae.it/
[7] https://www.etgsrl.it/

romacfc_asn	List of sensors	CSV	17688
romacfc_asp	List of stations	CSV	4980
romacfc_atp	Sensor types	CSV	528



FIGUR 1 – UML DIAGRAM OF DPC OBSERVATIONS FILES

4.3.1 Observations – YYYYMMDD_dvd

The csv file YYYYMMDD_dvd collects all the observations measured in one day and is updated every 20 minutes. There is one row for each sensor and one column for each minute.

The csv file starts with a header that summarize the contents, followed by the columns and effective data. The header is the following.

Ditta	**Formato**	**Release**	**Data Creazione**	**Data Modifica**	**Numero Sensori**	**Data di Acquisizione**	**Ultimo Dato**	**GMT**	**DT minimo**
CAE	DVD	1.02	11/06/2019 23:04	12/06/2019 09:45	17980	20190612	10:36	60	1

In the following table only the first columns are listed; there is one column for each minute. The file regarding observations of 12/06/2019 until 10:36 has 643 columns. For a few sensors the observations are available every minute; most are available every 10/30/60 minutes.

N Sens	DT	Max	Ora Max	Min	Ora Min	00.00	00.01	…	10:36

4.3.2 Sensors - romacfc_asn

The csv file romacfc_asn contains the list of sensors.

The first two rows of the csv file are the following header.

Ditta	**Formato**	**Release**	**Data Creazione**	**Data Modifica**	**# Elementi**

CAE	ASN	1.01	14/02/2012 00:14	07/06/2019 17:30	17688

The columns of the csv file are the following.

Numero Sensore	Numero Stazione	Tipo Sensore	DescrC	DescrB	DescrA	Data Inizio	Data Fine	Parametri DAS
(#)	(#)	(#)	(30c)	(12c)	(6c)	(AAAAMMGG)	(AAAAMMGG)	(100c)

4.3.3 Stations - romacfc_asp

The csv file romacfc_asp contains the list of weather stations.

The first two rows of the csv file are the following header.

Ditta	Formato	Release	Data Creazione	Data Modifica	# Elementi
CAE	ASP	1.01	14/02/2012 00:14	07/06/2019 17:30	4980

The columns of the csv file are the following.

# Stazione	Nome Stazione	Localita'	Comune	Provincia	Regione	Nazione	Bacino	Ente (proprietario)
(#)	(30c)	(25c)	(25c)	(2c)	(25c)	(20c)	(20c)	(6c)

Longitudine	Latitudine	Altezza	GMT	Instradamento
(#gradi*10000)	(#gradi*10000)	(# slm)	(# in min.)	(200c)

4.3.4 Sensort types - romacfc_atp

The csv file romacfc_atp contains the list of sensor types.

The first two rows of the csv file are the following header.

Ditta	Formato	Release	Data Creazione	Data Modifica	# Elementi
CAE	ATP	1.01	14/02/2012 00:14	07/06/2019 17:30	528

The columns of the csv file are the following.

Tipo Sensore	Codice Classe	Descr Classe	DescrC	DescrB	DescrA	Unita'M	Decimali
(#)	(#)	(30c)	(30c)	(12c)	(6c)	(10c)	(#)

DTC	DTM	DTR	Minimo	Massimo	Precisione	Specifico
(# sec)	(# sec)	(# sec)	(#)	(#)	(#)	(100c)

4.4 GRIB

The GRIB format is designed to encode data produced by numerical weather prediction models. It can also represent observations, but on a regularly distributed coverage. GRIB is a table-driven code defined by the World Meteorological Organization (WMO). It is mainly used by national weather services for the exchange of data, but also as data storage format.

The current edition number is 2. However, GRIB edition 1 remains in use. The Edition 2 of GRIB enable the coding of new products, such as the output of ensemble prediction systems, long-range forecasts, climate predictions, ensemble wave forecasts or transport models, cross-sections[8].

4.4.1 GRIB Sections

GRIB messages contain metadata, which describe the data under consideration, and the data itself. GRIB coded data consist of a continuous bit-stream made of a sequence of octets (or bytes: 1 octet = 8 bits). Thus, the representation of the data is independent of any particular machine representation. The octets of a GRIB message are grouped in sections. Most sections are used to specify the metadata, while one special section, the Data Section, contains the data in compressed form.

[8] https://www.wmo.int/pages/prog/www/WMOCodes/Guides/GRIB/Introduction_GRIB1-GRIB2.pdf

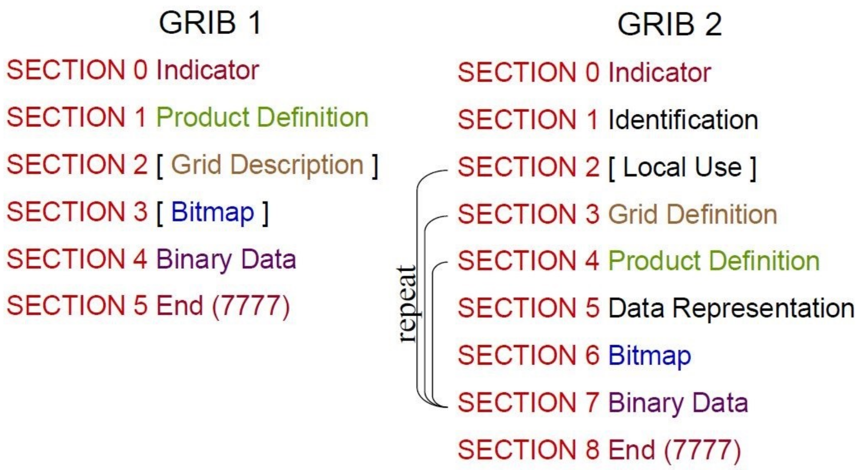


FIGURE 4 - GRIB SECTIONS[9]

Sections 0 in GRIB1 and 0 and 1 in GRIB2 identify the message.

Section 2 in GRIB2 is an optional section that can be used to transmit any information or parameters for national purpose.

In GRIB 1:

Section 1 defines time or time periods, parameter, level(s) and the grid used. It can be a pre-defined grid, just referenced by a number. If the grid is not defined already, the report must contain Section 2 which will define the grid according to predefined templates listed in the WMO Manual (e.g.: Mercator, Polar stereographic). Section 3 contains a "bit" map. If data is present at this grid point (listed in rows), the corresponding bit is set to 1. Section 4 starts by the definition of the packing scheme used (e.g. spherical harmonic coefficient, simple packing, complex packing, second order packing).

The subdivision of the GRIB1 Product Definition Section into the GRIB2 Identification, Product Definition, and Data Representation Sections is one of the fundamental differences between GRIB1 and GRIB2. This, combined with the option for iterating the Sections and expansion of the concept of templates, are what provide GRIB2 with its substantially enhanced flexibility over what is possible in GRIB1.

GRIB uses the concept of Template: "Description of the standardized layout of a set of data entities"[10].

[9] https://confluence.ecmwf.int/download/attachments/45748923/grib_decoding.pdf
[10] https://www.wmo.int/pages/prog/www/WMOCodes/Guides/GRIB/GRIB2_062006.pdf

4.4.2 GRIB Edition 1

In 1985, the World Meteorological Organization (WMO) Commission for Basic Systems (CBS) approved a general purpose, bit-oriented data exchange format, designated FM 92-VIII Ext. GRIB (GRIdded Binary). By packing information into the GRIB code, messages (or records - the terms are synonymous in this context) can be made more compact than character-oriented bulletins.

Changes and extensions to GRIB were approved at the regular meeting of the WMO/CBS. The 1990 changes were of such structural magnitude as to require a new Edition of GRIB, Edition 1. The Sub-Group made further changes to some of the Tables (additions of new parameters or more precise definition of existing ones), resulting a new Version number for them. This brings us now to Table Version 2.

In November 2016, the sixteenth session of the Commission for Basic Systems (CBS-16) decided to remove the GRIB edition 1 from the Manual on Codes, Volume I.2 (WMO-No. 306) and to ensure it remains available so that archived data can still be referenced.

Each GRIB record intended for either transmission or storage contains a single parameter with values located at an array of grid points, or represented as a set of spectral coefficients, for a single level (or layer), encoded as a continuous bit stream. Logical divisions of the record are designated as "sections", each of which provides control information and/or data. A GRIB record consists of six sections, two of which are optional:

Section number	Name	Contents
0	Indicator section	"GRIB", length of message, GRIB edition number
1	Product definition section	Length of section, identification of the coded analysis or forecast
2	Grid description section (optional)	Length of section, grid geometry, as necessary
3	Bit-map section (optional)	Length of section; the bit per grid point, placed in suitable sequence, indicates omission (bit 0) or inclusion (bit 1) of data at respective points
4	Binary data section	Length of section and data values
5	End section	7777

FIGURE 5 GRIB1 SECTIONS[11]

[11] https://apps.ecmwf.int/codes/grib/format/grib1/overview

Message and section lengths are expressed in octets. Section 0 has a fixed length of 8 octets; Section 5 has a fixed length of 4 octets. Sections 1, 2, 3 and 4 have a variable length which is included in the first three octets of each section.

With the exception of the first four octets of the Indicator Section, and the End Section, all octets contain binary values. All sections contain an even number of octets; the variable length sections are padded with zero values as necessary.

The "**Section 0: Indicator section (IS)**" serves to identify the start of the record in a human readable form, indicate the total length of the message, and indicate the Edition number of GRIB used to construct or encode the message. The section is always eight octets long.

The "**Section 1: Product definition section (PDS)**" contains indicators for the Parameter table Version, the originating center, the numerical model (or "generating process") that created the data, the geographical area covered by the data, the parameter itself, the values for the appropriate vertical level or layer where the data reside, the decimal scale factor, and date/time information. The PDS is normally 28 octets long but it may be longer if an originating center chooses to make it so.

The "**Section 2: Grid description section (GDS)**" is optional and it provides a grid description for grids not defined by number in Table 3.

The "**Section 3: Bit-map section (BMS)**" is optional and provides either a bit map or a reference to a bit map pre-defined by the center. The bit map consists of contiguous bits with a bit-to-data-point correspondence as defined in the grid description. A bit set equal to 1 implies the presence of a datum for that grid point in the BDS; a value of zero implies the absence of such. This is useful in shrinking fields where fair portions of the field are not defined. An example would be global grids of sea surface temperature; the bit map would be used to suppress the "data" at grid points over land. One would not want to use the BMS if the data were un-defined at only a small number of grid points as the overhead of adding the bit map array (one bit for each grid point) might add more bits to the overall message that were subtracted by the removal of a few data values.

The "**Section 4: Binary data section (BDS)**" contains the packed data and the binary scaling information needed to reconstruct the original data from the packed data. The required decimal scale factor is found in the PDS. The data stream is zero filled to an even number of octets.

The "**Section 5: End section**" serves a human readable indication of the end of a GRIB record. It can also be used for computer verification that a complete GRIB record is available for data extraction. It should not be used as a search target since a '7777' bit combination could exist anywhere in the binary data stream.

4.4.3 GRIB Edition 2

The Edition 2 (GRIB2) has been approved at the regular meeting of the WMO/CBS in December 2000. A new expanded name was given to the code form: "General Regularly-distributed Information in Binary form".

The GRIB2 involved a basic restructuring of GRIB because of the GRIB1 Code Table 2 (Indicator of Parameter) was completely full. More fundamentally, however, GRIB1 was designed with a single overall structure for identifying products. Although this structure had some flexibility, it was not sufficient to represent many of the new products, such as those from ensemble predictions, that were becoming available by the end of the 1990s.

Each GRIB2 message intended for either transmission or storage contains one or more parameters with values located at an array of grid points or represented as a set of spectral coefficients. Logical divisions of the message are designated as sections, each of which provides control information and/or data. There are nine different types of sections, one of which is optional. The nine sections and their general contents are:

Section	Abbreviation	Name
0	IS	INDICATOR SECTION
1	IDS	IDENTIFICATION SECTION
2	LOC	LOCAL USE SECTION
3	GDS	GRID DEFINITION SECTION
4	PDS	PRODUCT DEFINITION SECTION
5	DRS	DATA REPRESENTATION SECTION
6	BITMAP	BIT-MAP SECTION
7	DATA	DATA SECTION
8	END	END SECTION

FIGURE 6 - GRIB2 SECTIONS[12]

The “**Section 0: Indicator Section**” serves to: identify the start of the GRIB2 message in a human readable form, describe the “Discipline” of the information contained in the message, indicate the

[12] https://apps.ecmwf.int/codes/grib/format/grib2/sections/

Edition Number of GRIB (2 in the case of GRIB2) used to encode the message, and the total length of the message. The section is always 16 octets long.

The "**Section 1: Identification Section**" contains characteristics that apply to all processed data in the GRIB message. These characteristics identify the originating centre and sub-centre, indicate the GRIB Master Table and Local Table versions used, and give the reference time, the production status, and the type of processed data contained in this GRIB message.

The "**Section 2: Local Use Section**" contains information for local use by the originating/generating centre. The originating/generating centre can put anything it desires in this Section.

The "**Section 3: Grid Definition Section**" defines the grid surface and geometry of the data values within the surface for the data contained in the next occurrence of the Data Section.

GRIB2 retains the powerful GRIB1 concept of a template in this Section and extends it to the Product Definition, Data Representation, and Data Sections as well. Use of a template means there are very few values common to all Grid Definition Sections possible in GRIB2. Rather, the number of the Grid Definition Template used is encoded. The values that must follow are those required by that particular Grid Definition Template.

The "**Section 4: Product Definition Section**" describes the nature of the data contained in the next occurrence of the Data Section. The content of this section is defined using the Product Definition Template (code table 4.0[13]).

The "**Section 5: Data Representation Section**" describes how the data values are represented in the next occurrence of the Data Section. The content of this section (e.g. Grid point data or Spectral data, simple or complex packing) is defined using the Data Representation Template (code table 5.0[14]).

The "**Section 6: Bit-Map Section**" indicates the presence or absence of data at each of thegrid points, as applicable, in the next occurrence of the Data Section. Consider, for example, a precipitation field. There will be many data points in a precipitation field with a zero value – no precipitation. In order to conserve space, a bit map can be used to efficiently indicate those data points with a zero precipitation value that does not appear in the Data Section. This is accomplished by generating a bit string with one bit corresponding to each data point. A bit set to one implies the presence of a data value at that data point, while a bit set to zero implies the absence of a data value at that data point. Those data points for which the bit is set to zero will not have a corresponding value in the Data Section.

The "**Section 7: Data Section**" contains the data values themselves.

[13] https://apps.ecmwf.int/codes/grib/format/grib2/ctables/4/0
[14] https://apps.ecmwf.int/codes/grib/format/grib2/ctables/5/0

There are several approaches to data compression used in GRIB2, including simple packing, complex packing, spatial differencing, JPEG2000 and PNG. If only simple packing is used, the data values in octets 6 - nn are in the form of a binary string, filled with bits set to zero to ensure the Section ends on an octet boundary. However, if Complex packing and/or spatial differencing is used, the storage of the data values in the Data Section is more complicated, and is described by the appropriate Data Template. Note the implied link between Data Representation Template 5.X and Data Template 7.X, where X in both cases is the Data Representation Template Number given in octets 10 - 11 of the Data Representation Section.

GRIB2 permits multiple data sets to be encoded in a single GRIB2 message. Multiple data sets implies multiple occurrences of the Grid Description, Product Definition, Data Representation, and Bit-Map Sections. However, this did not require each of these four sections be repeated for each occurrence of the Data Section. Rather, the contents of the most recent occurrences of the Grid Description, Product Definition, Data Representation, and Bit-Map Sections prior to any individual occurrence of the Data Section are those that describe the contents of that occurrence of that Data Section.

The "**Section 8: End Section**" serves to identify the end of the GRIB2 message in a human readable form. The content is "7777" (coded according to the International Alphabet No. 5).

The End Section is always 4 octets long. The Indicator (16 octets long) and End Sections are the only fixed-length sections in a GRIB2 message.

4.4.4 GRIB Templates in Mistral

GRIB2 templates adopted in Mistral:

- **4.0** Analysis or forecast at a horizontal level or in a horizontal layer at a point in time
- **4.1** Individual ensemble forecast, control and perturbed, at a horizontal level or in a horizontal layer at a point in time
- **4.5** Probability forecasts at a horizontal level or in a horizontal layer at a point in time
- **4.8** Average, accumulation, and/or extreme values or other statistically processed values at a horizontal level or in a horizontal layer in a continuous or non- continuous time interval
- **4.9** Probability forecasts at a horizontal level or in a horizontal layer in a continuous or non-continuous time interval
- **4.10** Percentile forecasts at a horizontal level or in a horizontal layer in a continuous or non-continuous time interval. Preliminary Note: This template was not validated at the time of publication and should be used with caution. Please report any use to WMO Secretariat (World Weather Watch - Basic Systems Department) to assist for validation.
- **4.11** Individual ensemble forecast, control and perturbed, at a horizontal level or in a horizontal layer in a continuous or non-continuous time interval

For template description see https://apps.ecmwf.int/codes/grib/format/grib2/templates/

Please note that a formal “Data Governance” procedure is being pursued by ECMWF, together with WMO, to deliver more formal recognition and documentation of the new, innovative post-processed outputs which form an important deliverable of MISTRAL. Such a process can take time to complete and in the meantime "local documentation" will be used to record the exact meaning of new variables being created.