European Commission

Copernicus Emergency Management Service

The CEMS Meteorological Data Collection

Annual Report 2020

Prepared by the CEMS METEOROLOGICAL DATA COLLECTION CENTRE

GLOBAL PRECIPITATION CLIMATOLOGY CENTRE

Joint Research Centre

KISTERS

EU Science Hub
https://ec.europa.eu/jrc

JRCXXXXX

Ispra: European Commission, 2021

How to cite this report: Rehfeldt, K., Schirmeister, Z., Pichon, D., Rauthe-Schöch, A., Schweim, C., Walawender, J., Ziese M., Gomes, G., Thiemig, V. and Salamon, P., The CEMS Meteorological Data Collection Centre – Annual report 2020, European Commission, 2021, JRCXXXXXX.

Contents

1 Abstract

The Copernicus Emergency Management Service (CEMS) Meteorological Data Collection Centre (METEO) collects, quality-controls and post-processes in situ and ground-based remote sensed meteorological data to provide input data tailored to the needs of the CEMS EFAS (European Flood Awareness System), EFFIS (European Forest Fire Information System) and EDO (European Drought Observatory). By the end of 2020, real-time data delivered by 24 providers for 11 parameters from more than 22.000 stations are collected. On average, 5.500.000 records are being added to the data base each day. Additionally, the data base comprises more than 40.000 stations with historic data.

All data, real-time and historic, are quality-controlled in the same manner during the import to the data base. Only the data that passed the quality control are processed further as reliable input data for the CEMS. This post-processing includes the calculation of minimum, maximum and mean values as well as the aggregation of totals over different accumulation periods. Data were provided as station lists or grids including an uncertainty estimation to the CEMS. The grids are generated by means of a modified SPHEREMAP scheme through interpolation from the quality-controlled station data.

Within 2020, four additional real-time data providers were integrated as well as several historical data. The configuration of the data base was improved to use more of the existing data. As clustered stations can cause unreliable sharp gradients in the grids, an algorithm was developed to merge clustered stations prior to the gridding.

A gap analysis was done regarding the availability of data with a proposal for the future data collection strategy and extended use of existing data.

2 Description of CMES METEO

METEO, the Meteorological Data Collection Centre, was established to provide application tailored quality-controlled surface meteorological data to the Copernicus Emergency Management Services (CEMS). It is operated by the KISTERS AG and Deutscher Wetterdienst (DWD). Current CEMS components served by METEO are EFAS (European Flood Awareness System), EFFIS (European Forest Fire Information System) and EDO (European Drought Observatory). To fulfil this task, METEO collects in situ meteorological observations as well as ground-based remote-sensed data like radar observations of precipitation. The data are received from many data providers using various sources, e.g. ftp-server, APIs or via email, and different file formats. All received data are quality controlled and integrated into a data base. A post-processing is done to make the data usable for the CEMS. This comprises the calculation of minimum, maximum, and mean values, the aggregation and disaggregation of totals. Depending on the component, the data are either provided as station data or gridded fields.

3 Data Providers and Provision

An overview about the current data providers and which data they provide to METEO is given in this section.

3.1 Data Providers, Parameter and Time Resolution

By the end of 2020, 24 data providers deliver real-time data to METEO. The data providers and delivered parameters are summarised in the Table 3.1.

Table 3.1: List of active real-time data providers and delivered parameters (abbreviation). Parameter abbreviations are given in Table 3.2.

Name	Parameter										
	ClCo	**DT**	**Evap**	**Precip**	**ReAiHu**	**SunRad**	**SunD**	**AT**	**VP**	**WDir**	**WSpeed**
Agencia Estatal de Meteorología (Spain)	-	x	-	x	x	x	x	x	-	x	x
Deutscher Wetterdienst (Germany)	x	x	-	x	-	-	x	x	-	x	x
Agenzia Regionale per la Prevenzione e l'Ambiente dell'Emilia-Romagna (Italy)	-	-	-	x	-	-	-	x	-	-	-
Slovenian Environment Agency	-	-	-	x	x	x	-	x	-	x	x
Czech Hydro-Meteorological Institute	-	-	-	x	-	-	-	x	-	-	-
Deutscher Wetterdienst (global)	x	x	-	x	x	x	-	x	-	x	x
Environment Agency (England)	-	-	-	x	-	-	-	-	-	-	-
Federal Hydrometeorological Institute (Bosnia-Herzegovina)	-	-	-	x	-	-	-	x	-	-	-
Finnish Meteorological Institute	x	x	-	x	x	-	-	x	-	x	x
Institute of Meteorology and Water Management (Poland)	x	x	-	x	-	x	x	x	-	x	x
Institute for Ocean and Atmosphere (Portugal)	-	x	-	x	x	x	-	x	-	x	x
Kosovo Hydrometeorological Institute	-	-	-	x	x	x	-	x	-	x	x
Royal Netherlands Meteorological Institute	x	x	-	x	x	x	x	x	-	x	x
National Environmental Agency (Georgia)	x	x	-	x	-	-	-	x	-	x	x
MARS[1] (global)	x	-	-	x	-	x	-	x	x	-	x
MeteoLux (Luxembourg)	x	x	-	x	x	-	x	x	-	x	x
MeteoSchweiz (Switzerland)	x	x	x	x	x	x	x	x	-	x	x
Met Éireann (Ireland)	-	-	-	x	-	-	-	x	-	x	x
Norwegian Meteorological Institute	x	x	-	x	x	-	-	x	-	x	x
Automatic System of Hydrological Information (SAIH) for the Ebro river basin (Spain)	-	-	-	x	x	x	-	x	-	x	x
Slovak Hydro-Meteorological Institute	-	-	x	x	-	-	-	x	-	-	-
Swedish Meteorological and Hydrological Institute	x	-	-	x	x	x	x	x	-	x	x
Zentralanstalt für Meteorologie und Geodynamik (Austria)	-	-	-	x	-	-	-	x	-	-	-

[1] Monitoring Agricultural ResourceS system, JRC, European Commission

Table 3.2: Parameter abbreviation and description.

Parameter abbreviation	Parameter description
ClCo	Cloud cover
DT	Dew point temperature
Evap	Evaporation
Precip	Precipitation
ReAiHu	Relative Air humidity
SunRad	Solar radiation
SunD	Sunshine duration
AT	Air temperature
VP	Water vapor pressure
WDir	Wind direction
WSpeed	Wind speed

In order to provide reliable maps for historical periods, used for example as input datasets to calibrate the hydrological model of EFAS, data providers were asked to deliver also historical data back to 1970. Some data providers provided only historical data. Historical data were also retrieved from research projects and gridded data sets. In case a country or region provided only historic data, these are covered with real-time data by the two global data deliveries from MARS and DWD (global). All the historical data integrated into the METEO data base are summarised in Table 3.3.

Table 3.3: List of data providers with historical data and delivered parameters (abbreviation). Parameter abbreviations are given in Table 3.2.

Name	Parameter type										
	ClCo	DT	Evap	Precip	ReAiHu	SunRad	SunD	AT	VP	WDir	WSpeed
Norwegian Meteorological Institute	x	x	-	x	x	-	-	x	-	x	x
CarpatClim	-	-	-	x	-	-	-	-	-	-	-
ERA-Interim-land	-	-	-	x	-	-	-	-	-	-	-
EURO4M-APGD	-	-	-	x	-	-	-	-	-	-	-
National Environmental Agency Georgia	-	-	-	x	-	-	-	x	-	-	-
MeteoSchweiz	x	x	x	x	x	x	x	x	-	x	x
Institute of Meteorology and Water Management (Poland)	x	x	-	x	x	-	x	x	x	x	x
European Climate Assessment and Dataset (ECAD)	-	-	-	x	-	-	-	-	-	-	-
Euro Synop	x	x	-	x	-	-	-	x	-	x	x
Czech Hydro-Meteorological Institute	-	-	-	x	-	-	-	x	-	-	-
Slowenian Environment Agency	-	-	-	x	x	x	-	x	-	x	x
Met Éireann	-	-	-	x	-	-	-	x	-	x	x
Agenzia Regionale per la Prevenzione e l'Ambiente dell'Emilia-Romagna (Italy)	-	-	-	x	x	x	-	x	-	x	x
MARS	-	-	-	x	-	x	-	x	x	-	x

Name	Parameter type										
Automatic System of Hydrological Information (SAIH) for the Ebro river basin (Spain)	-	-	-	x	-	-	-	x	-	-	-
Slovak Hydro-Meteorological Institute	-	-	x	x	-	-	-	x	-	-	-
Hungarian Meteorological Service	x	x	-	x	x	-	x	x	-	x	x
Danube	-	-	-	x	-	-	x	x	-	-	-
DWD Climatic (Germany)	x	-	-	x	x	-	x	x	-	-	-
Institute for Ocean and Atmosphere (Portugal)	-	x	-	x	x	x	-	x	-	x	x
MeteoConsult	-	-	-	x	-	-	-	-	-	-	-
Zentralanstalt für Meteorologie und Geodynamik (Austria)	-	-	-	x	-	-	-	x	-	-	-
Hellenic National Meteorological Service	x	-	-	x	x	-	x	x	x	-	x

Data are provided with various temporal resolutions and aggregation intervals, depending on the parameter and data provider. The highest received temporal resolution and accumulation period is one minute (air temperature and precipitation from one data provider). Instantaneous parameters, like air temperature or wind speed, are mainly provided with temporal resolution of one, three and six hours, but also with higher and lower temporal resolutions. The majority of provided precipitation totals are accumulated over six and twelve hours, but also daily and one hourly totals are often provided. Minimum and maximum air temperature, are mostly provided on a daily basis.

3.2 New Data Providers in 2020

A continuous task for METEO is the acquisition and integration of additional data providers for real-time and historical data. This is done for two reasons: (1) to get more real-time information into the products and to increase reliability of the grids and (2) to enlarge the data base with additional historical data needed, for example, for the calibration of the hydrological model.

To enlarge the METEO data base, these data providers were added within 2020:

- Israel Meteorological Service (IMS)
- Federal Hydrometeorological Institute (FHMZ, Bosnia-Herzegovina)
- Royal Netherlands Meteorological Institute (KNMI)
- National Environmental Agency (LEPL, Georgia)

The increased station density in comparison to the previous status in shown in Figure 3.1. These four data providers deliver real-time (Table 3.1) and historical (Table 3.3) data.

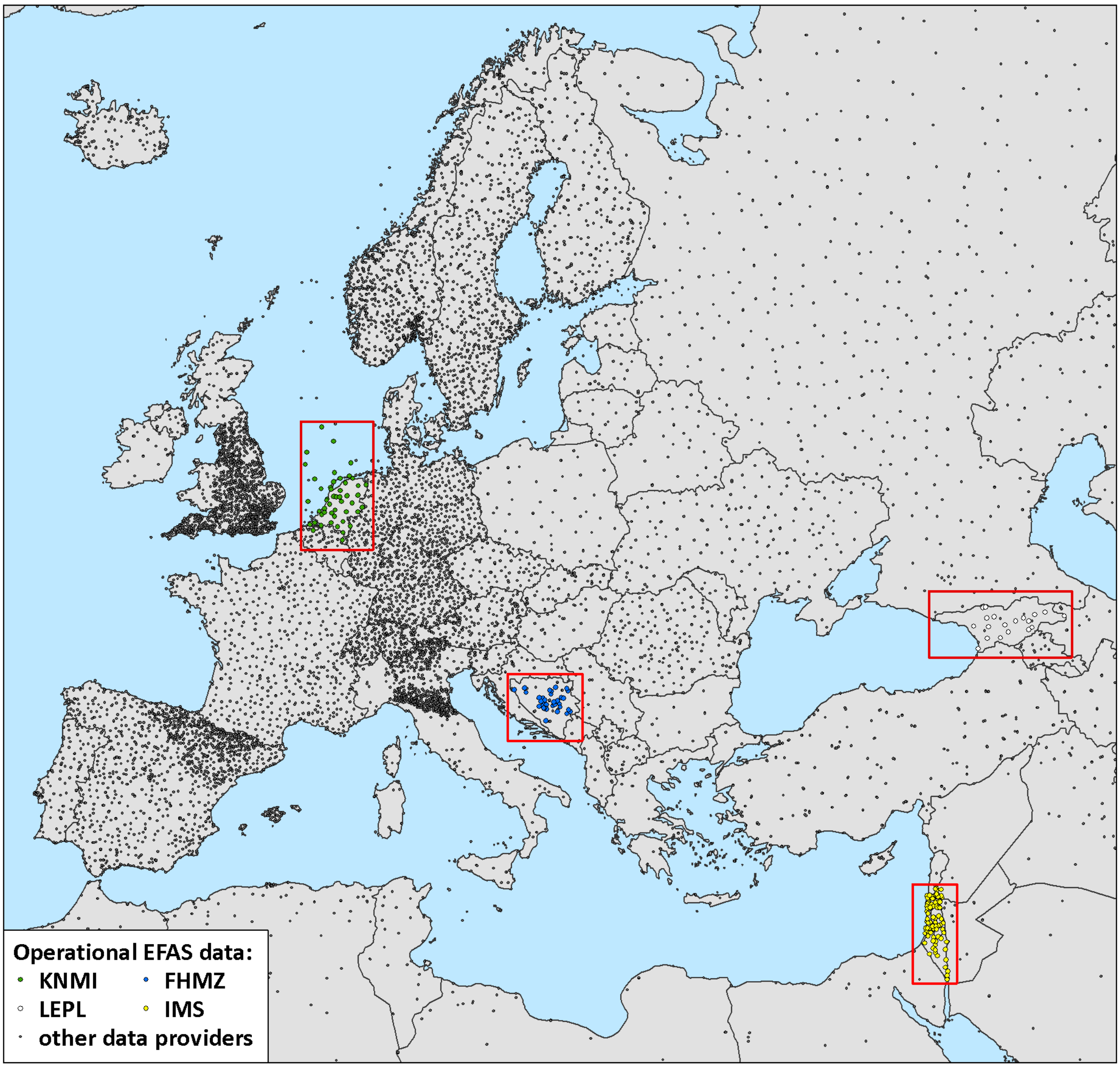

Figure 3.1: Additional stations due to the integration of new data providers in comparison to the existing station network. Red rectangular mark the regions with the new data providers.

4 Database

4.1 Data Flow

The data flow within METEO is illustrated in Figure 4.1. Real-time data are delivered by the data providers via (s)ftp-servers (pulled by METEO/pushed to METEO), web services/APIs and email attachments in provider specific file formats. All files are converted into a uniform and optimized file format for the integration into the METEO data base.

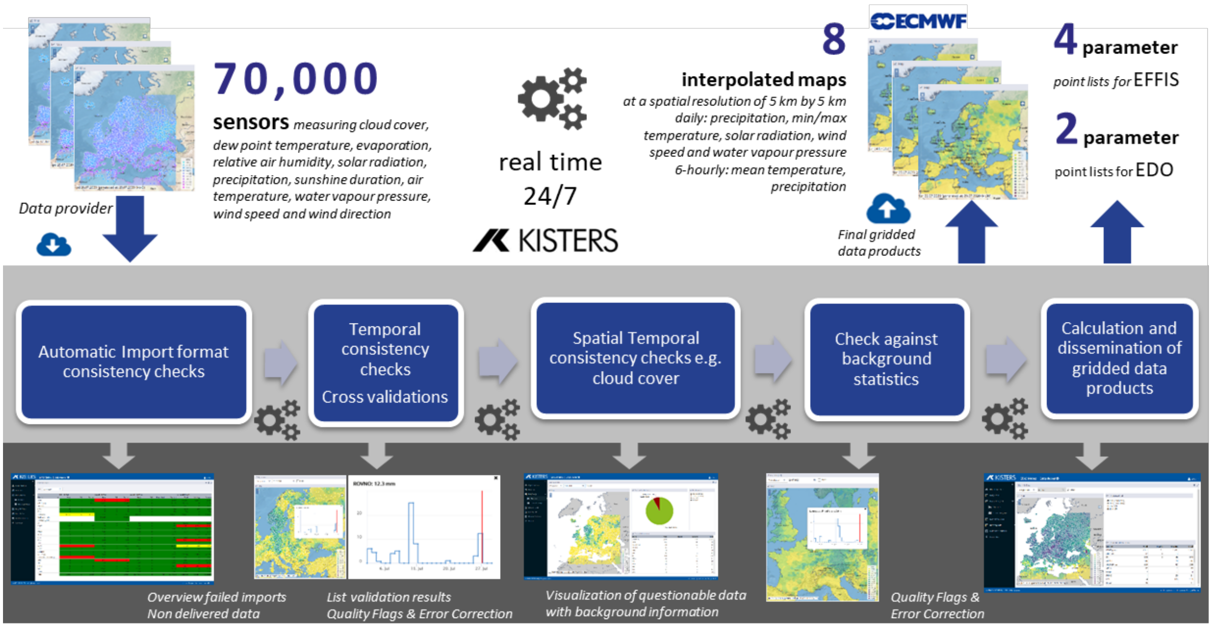

Figure 4.1: Schematic illustration of the data flow within METEO. The amount of input and output data are shown as well as the processing and quality control steps within the data bank system.

The provided data are stored in a data bank super-system WISKI. Within WISKI, the quality control (see section 4.2) of the data and necessary aggregations (see section 5.1) are done. Finally, the data are extracted from the data bank as input for the interpolated maps and station lists provided to the CEMS components. The interpolation procedure is described in detail in section 5.2.

4.2 Quality Control

Although the data is usually quality controlled by the data providers, an own quality control procedure was established based on the experience, that real-time data contains erroneous data points from time to time. This applies to historical data, too. Data providers are regularly informed about detected errors to feed back an added value on the data provision to METEO. As the quality control procedure is triggered by the import of data, also historic data, delayed data and data, which are re-sent by data providers in order to replace the data already existing data in the database, are checked in the same manner. This guarantees the availability of checked data in the data bank used as input for the interpolated maps and station lists. If needed, a quality control is done again on the aggregated data (see section 5.1). This is necessary, as for example the twelve hourly precipitation threshold is not twice the six hourly thresholds (Table 4.2).

Quality flags are added to each data record in the quality control procedure. The following flags are in use: “good” if the value passes all measures, “suspect[2]”, if it is inconsistent with other parameters

[2] The quality flag ‚suspect‘ is also added to data points shifted in time. For example, some stations provide six-, twelve-hourly and daily precipitation totals outside the needed time steps at 00, 06, 12 and 18 UTC. Such totals

(e.g. dew point temperature higher than air temperature) and "rejected", if it didn't pass any threshold. Additionally, a quality flag was defined for missing values in the time series.

The quality control is mainly based on fixed thresholds as shown in Table 4.1. Bear in mind that one threshold value is used for the whole EFAS domain. Additionally, cross-validation procedure against data from other parameters at the same station is carried out.

Table 4.1: Parameters with fixed thresholds. For precipitation see Table 4.2

Parameter	Min. threshold	Max threshold
Cloud cover	0	9 octas
Evaporation	0	15 mm/day or 3 mm/hour
Relative air humidity	5	100 %
Solar radiation	0	1360 cos(lat) W/m^2
Sunshine duration	0	Astronomic max
Wind direction	0	360 deg
Wind speed	0	45 m/s

The thresholds for precipitation depends on the aggregation period (Table 4.2).

Table 4.2: Thresholds for precipitation depending on the aggregation interval

Aggregation interval [min]	Max precipitation threshold [mm]
15	125
30	200
60	250
180	350
360	425
540	475
720	500
900	525
1080	550
1440	600

Temperature data are checked against time-dependent limits taking the annual cycle into account. In winter, only data between -50°C and 25°C are used and in summer between -10°C and 55°C.

Dew point temperature is checked against the air temperature. If the dew point temperature is 30°C below the air temperature or 0.2°C above the air temperature, then the dew point temperature is flagged as "suspect".

4.3 Database Statistics

By the end of 2020, the METEO data base contained more than 62,000 stations, of which more than 22,000 stations delivering real-time data. The other stations provided real-time data in previous times or only historic data. 1324 of the stations with real-time data are so-called 'virtual

are shifted to the nearest needed date, as by doing so the uncertainty is lower compared to splitting such data into hourly totals and aggregate them.

stations', which are extracted from high-resolution gridded data sets (e.g. station adjusted radar quantitative precipitation estimations).

METEO received approximately 60,000 data files per day. All these files are processed, leading to on average 5,500,000 data records added to the data base – per day.

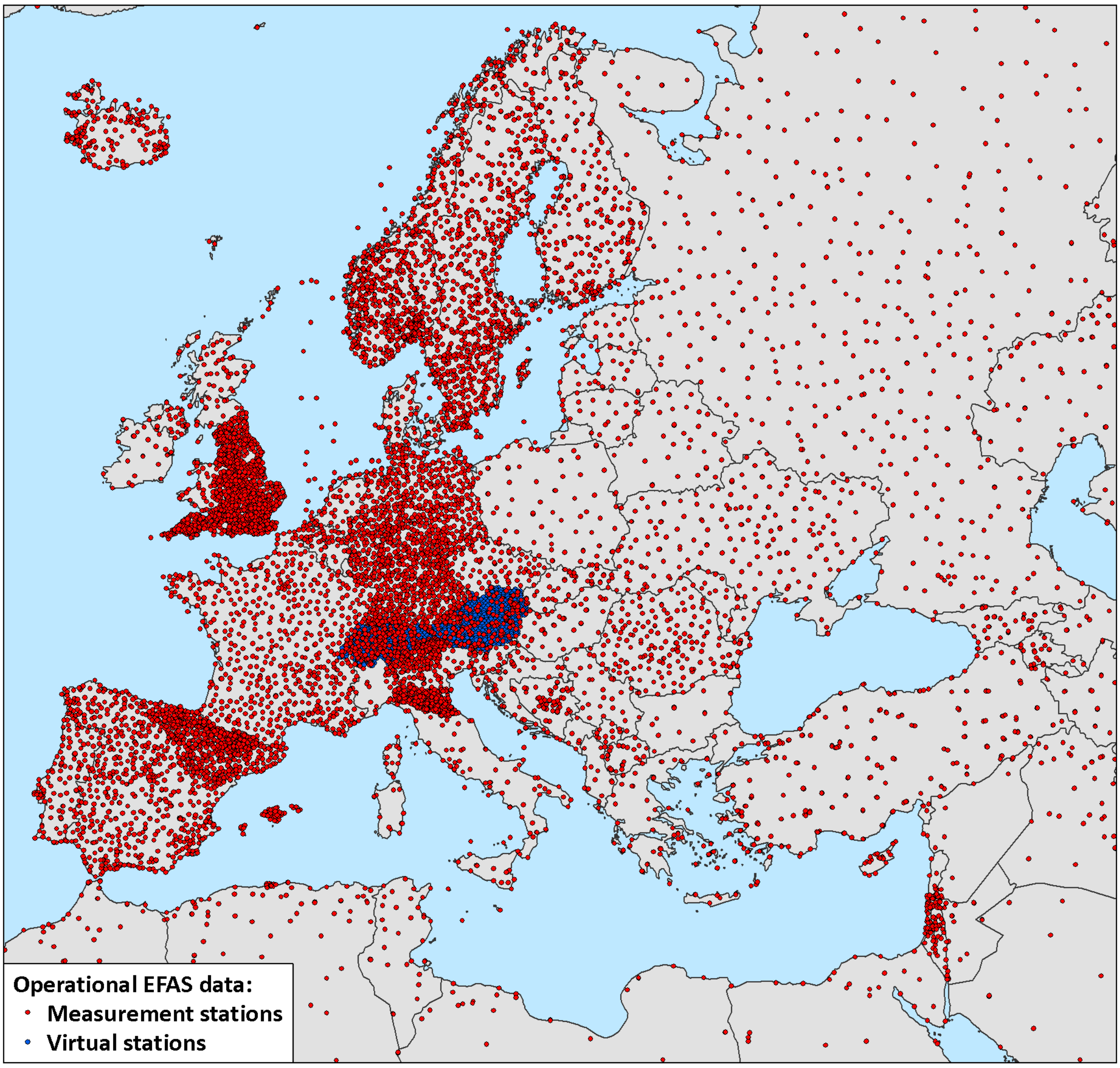

Figure 4.2: Spatial distribution of active stations within the EFAS domain. These stations deliver real-time data of at least one parameter stored in METEO data base. Maps per parameter are shown in Figure 7.1.

The spatial distribution of active stations within the EFAS domain is depicted in *Figure 4.2*. It is obvious that the data coverage varies between countries and even within the countries. Please note that not all active stations deliver all needed parameters, e.g. many stations 'only' deliver precipitation and temperature.

All the stations currently not delivering real-time data are classified as 'inactive'. Even if these stations don't contribute to the real-time grids, they are highly valuable by providing the input for historical grids: From time to time all data from the data base are extracted to compute grids for historical periods. Figure 4.3 shows the spatial distribution of the inactive stations within the EFAS domain. To increase the data coverage of former periods, data sets from research projects were integrated. These are for example CarpatClim or EURO4M-APGD, but also operational data sets like

ERA-Interim land were integrated. If such data sets contain gridded data, those are integrated as so-called ‘virtual stations’ on a regular grid.

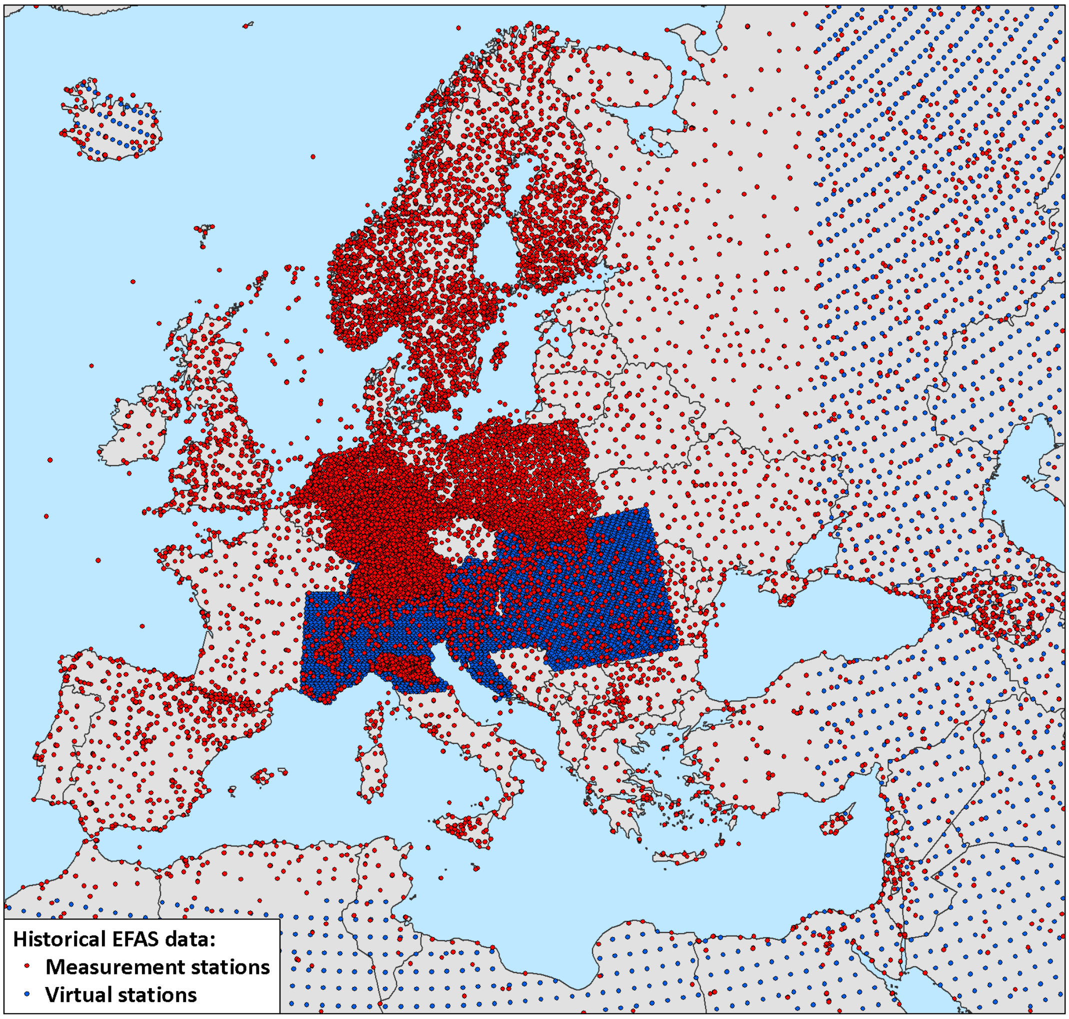

Figure 4.3: Spatial distribution of inactive stations within the EFAS domain. These stations deliver currently no real-time data, but data for former periods.

5 Post-Processing

The received data are post-processed to fulfil the needs of the CEMS, whereas the post-processing is partially different due to the application case of each CEMS. Only the quality-controlled data are used in the post-processing. It is done for the following subset of collected parameters:

- Precipitation;
- Air temperature;
- Wind speed;
- Solar radiation;
- Water vapor pressure;
- Relative air humidity.

Post-processing for all parameters comprises:

- Calculations of 6-hourly means/totals, except for solar radiation and water vapor pressure;
- Calculations of daily minimum, maximum and mean values, expect for water vapor pressure;
- Calculation of daily totals for precipitation and solar radiation;
- With different definitions of the start and end time at a day, depending on the CEMS component;
- Aggregation and disaggregation of precipitation totals;
- Extraction of data from data base;
- Spatial interpolation of station data to generate grids;
- Generation of station lists

5.1 Aggregation and Disaggregation, Calculation of Minimum, Maximum and Mean

Precipitation totals are delivered with various accumulation periods, mainly 6 hours, 12 hours and 24 hours, but also 15 minutes, 30 minutes, 1 hour or 3 hours. Additionally, the reporting behaviour of the stations differs between the data providers and even within some data providers.

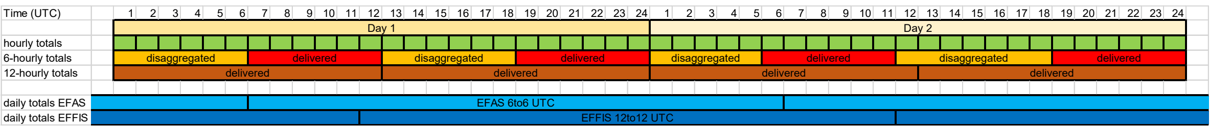

Figure 5.1: Scheme of aggregation and disaggregation of precipitation totals. Disaggregated 6-hourly totals (orange) are the difference between the 12-hourly totals (brown) and the enclosed 6-hourly totals (red). Hourly totals are shown in green and can be aggregated to 6-hourly and daily totals. Daily totals provided to EFAS from 6 UTC to 6 UTC (light blue) and to EFFIS from 12 UTC to 12 UTC (dark blue) are calculated from the enclosed available totals with shorter aggregation periods.

To achieve a high temporal and spatial coverage of 6-hourly precipitation totals, the 12-hourly and 6-hourly totals have to be disaggregated, if the 6-hourly totals are within the 12-hourly total (Figure 5.1). The resulting merged 6-hourly time-series is often the basis for METEO data deliveries to the CEMS. Where original data in higher resolution is available, 6-hourly totals are accumulated from the corresponding time-series, e.g. with 15 minutes temporal resolution.

Daily precipitation totals have to be provided as aggregated values from 6 UTC to 6 UTC of the following day and 12 UTC to 12 UTC of the next day (Figure 5.1). Whereas the daily totals at 6 UTC can be retrieved from the synoptic observations at 6 UTC and the higher temporal resolution time series, the daily totals at 12 UTC can only be computed from the higher time-resolution time-series.

As not all data providers deliver minimum and maximum temperatures according to the definitions of the CEMS, the required values are computed from the delivered instantaneous temperature data (Figure 5.2). The minimum temperature is the lowest temperature between 18 UTC and 6 UTC of the next day, whereas the maximum temperature is calculated from the observations taken between 6 UTC and 18 UTC. Also, the 6-hourly mean air temperature is calculated from the instantaneous data.

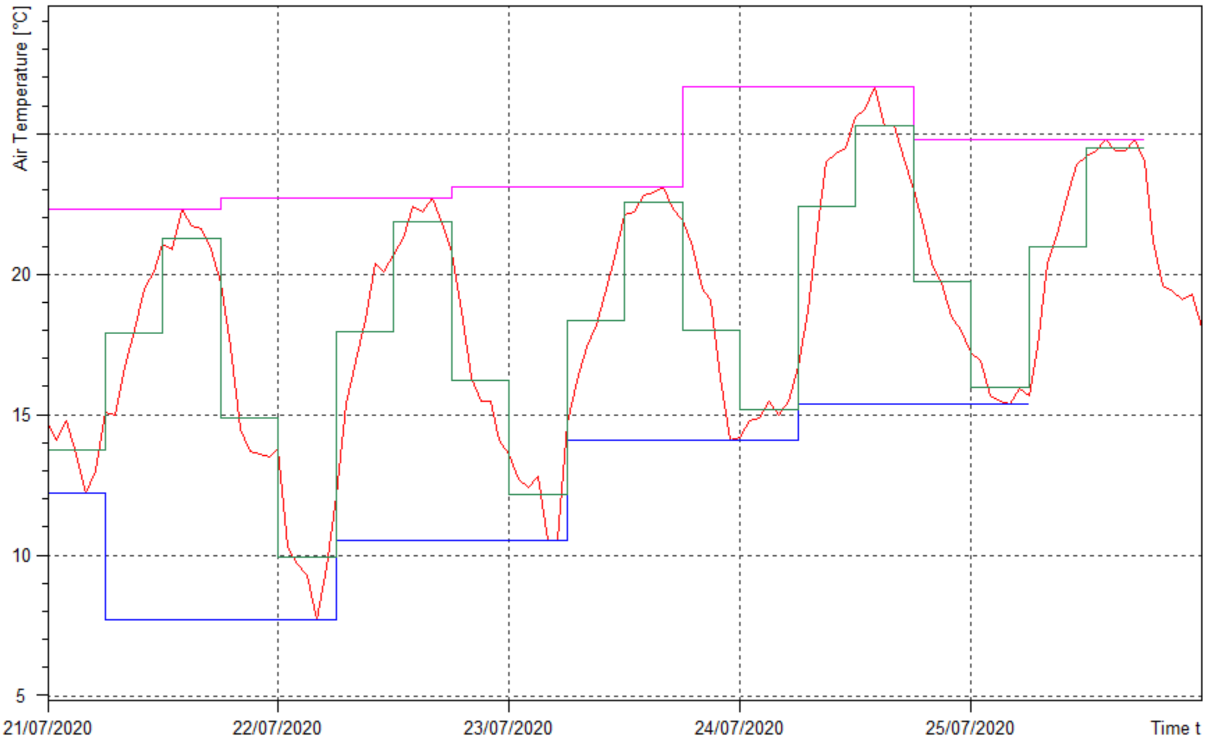

Figure 5.2: Delivered instantaneous air temperature (red) and therefrom calculated daily minimum temperature (blue), daily maximum temperature (magenta) and 6-hourly mean temperature (green).

Additionally, daily means of wind speed and daily accumulated totals of solar radiation are calculated.

A minimum availability of data is required to compute minimum, maximum and mean values as well as aggregated totals – further referred as coverage. Precipitation totals are only computed, if the aggregation period is fully covered by observational data (coverage = 100%). For all other parameters, a coverage of 87% is requested. An exception is made for six and twelve hourly means with a minimum coverage of 66% to consider also stations reporting with a temporal resolution of six hours.

5.2 Gridding

The hydrological model for EFAS relies on gridded input data. Grids are generated by means of the modified SPHEREMAP[3] interpolation scheme. This is a geometric scheme, which considers the distances between the stations and the grid point as well as the clustering of stations. Additionally, an estimation of the grid reliability by means of the standard deviation according to Yamamoto[4]

[3] Willmott, C.; Rowe, C. & Philpot, W.: Small-scale climate maps: A sensitivity analysis of some common assumptions associated with grid-point interpolation and contouring, The American Carthographer, 1985, Vol. 12, No. 1, P. 5-16

[4] Yamamoto, J.: An Alternative Measure of the Reliability of Ordinary Kriging Estimates, Mathematical Geology, 2000, Vol. 32, No. 4, P. 489-509

was implemented. This method depends on the differences between the input data and the interpolated value.

The input data and output of the modified SPHEREMAP scheme are illustrated in Figure 5.3 for the input stations (a), the grid itself (b) and the estimated uncertainty (c). As it can be seen, the uncertainty depends on the observed value and is higher in regions with high precipitation totals and zero in regions without precipitation. If multiple stations are clustered, the estimated uncertainty within the grids is lower than outside the clusters.

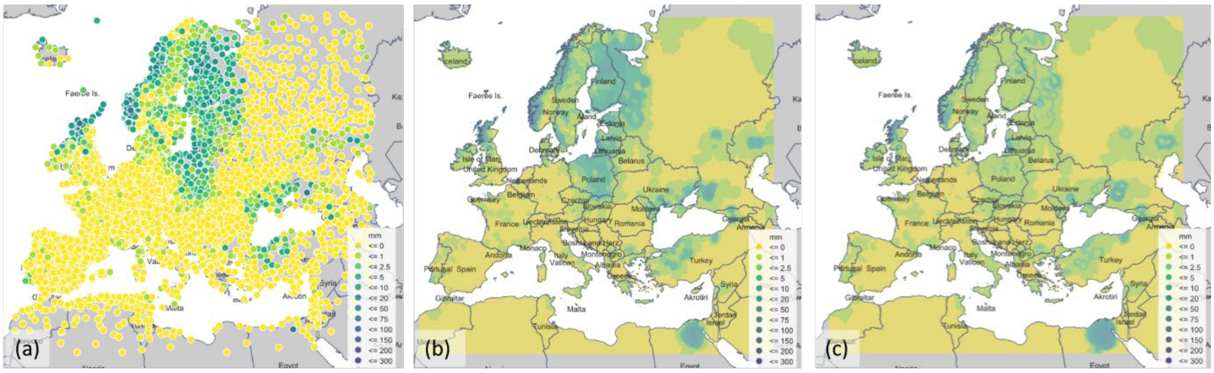

Figure 5.3: Input and output of the gridding: (a) station observations, (b) gridded data, (c) estimated uncertainty of the gridded data. The maps depict the spatial distribution of daily precipitation totals.

5.3 Station Lists

The European Forest Fire Information System (EFFIS) needs observations at station level for one date in a defined region. METEO provides such a list of quality controlled post-processed data from all the relevant stations. The spatial distribution of the station data summarised in a station list is depicted in Figure 5.4.

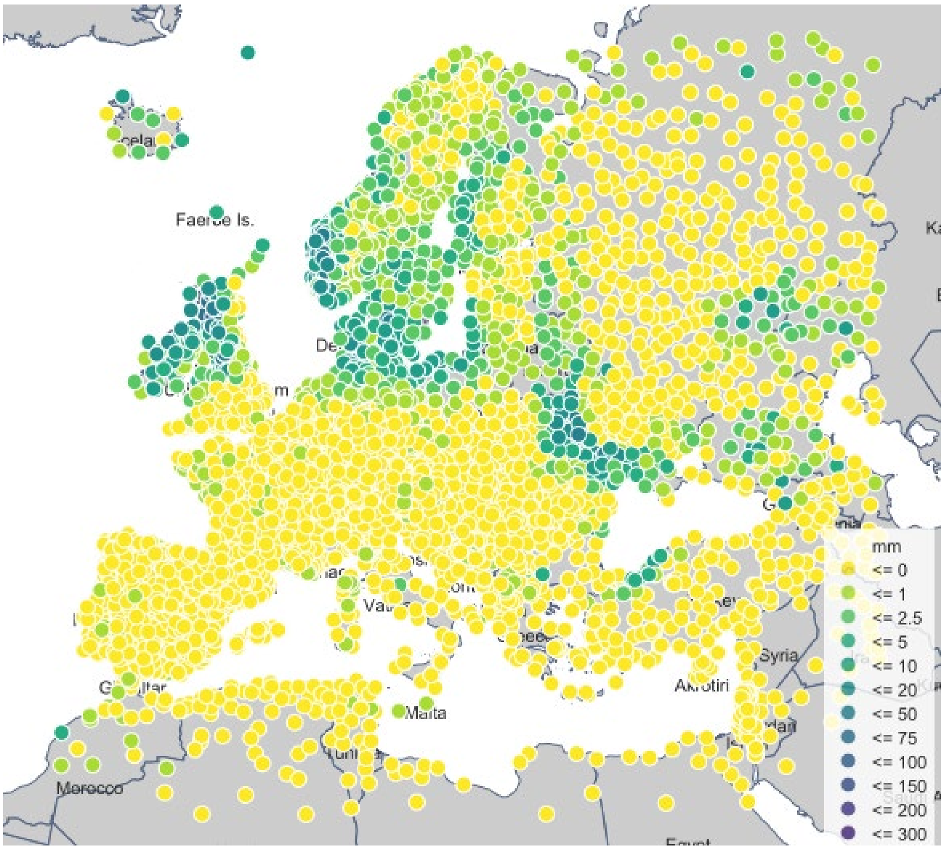

Figure 5.4: Spatial distribution of stations summarised in a station list, as for example provided to EFFIS. Depicted are stations providing such data allowing the calculation of 24-hourly precipitation totals at 12 UTC.

6 Improvements to data flow and post-processing

In order to improve the METEO database and the quality of the products provided for the CEMS, several changes were done in terms of availability of stations, post-processing and data storage.

6.1 Sort-out of duplicate stations

If stations are located close to each other (clustered), the weight of the cluster in the interpolation scheme is higher than it would be in case of a single station. This leads to rather sharp transitions be-tween the cluster and single surrounding stations in case of largely different values. Clustering of stations occurs in our input data for two reasons: 1) Larger cities usually have multiple stations which are very close together (e.g. Oslo in Norway), and 2) distinct providers report data from the same station (e.g. the global data providers DWDSynop & MARS plus regional data providers). As stations in the vicinity of each other are expected to report similar values, the overall approach is to group stations and use a mean value and position of the group for interpolation. The identification of such clustered stations is actually done in a two-stage process using both station- and parameter data and metadata (station id's and position).

In the first stage input as in 2) is treated where possible. Identification is done using the associated 5-digit WMO station number as station ID for the providers applicable. The group of stations with the same ID is removed from the input list and replaced by a "merged" station with that ID. The stations with highest quality code in this group are used to calculate the position of the new, "merged" station, its value (both via mean) and the quality code.

For the second stage effectively a graph is constructed, with the stations as vertices and an edge be-tween two stations if the distance is below a given threshold r. The connected components are then treated as above: Again, only stations with highest quality code are used to calculate the position of the new, final "merged" station, its value (both via mean) and the quality code.

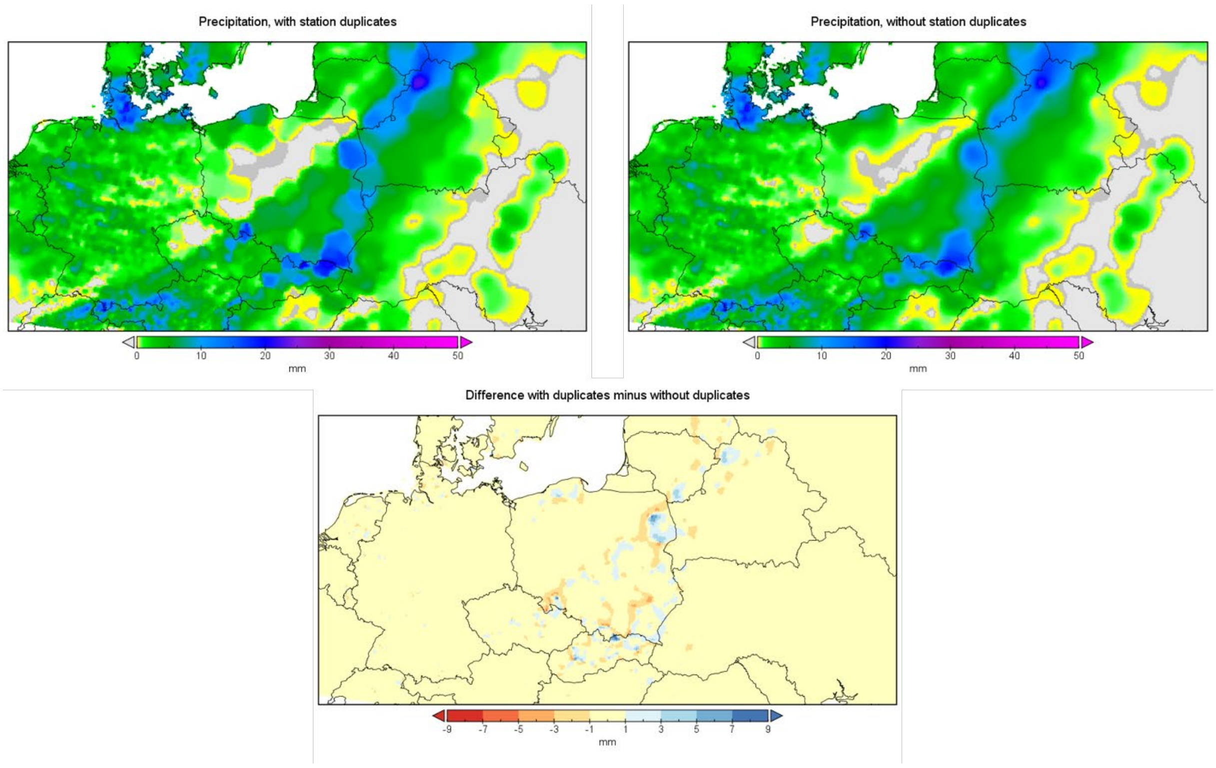

Figure 6.1: Precipitation grids of an arbitrary chosen day to depict the effect of the removal of duplicate stations. Top left: using all stations including duplicated, top right: duplicate stations are replaced by merged stations, bottom: difference between the top left and top right map.

A grid including all stations and one with the removed duplicates is shown in Figure 6.1. Comparing both maps in the top row clearly shown the unrealistic sharp gradients at the edges of the precipitation patterns and the plateau in the centre, if all duplicate stations were used (top left map). This feature vanished as the duplicate stations are replaced by a merged station (top right map). The map in the bottom row depicts the difference between both maps in the top row. It confirms, that the differences are at the edges of the patterns, whereas the values in the centre of the patters are nearly unchanged. Differences occur only in those regions with highly redundant data deliveries. For example, for Poland METEO receives data from the national hydrological and meteorological service IMGW and the both global data providers DWDSynop and MARS for the same stations. On the other hand, in Germany the national meteorological service DWD provides additional stations to them from the global data providers and therefore the impact of the duplicates is rather small.

6.2 Update of temporal coverage criteria and temporal resolution requirements

A systematic review and analysis of the coverage requirement for all near-real time data providers and parameters with the aim to propose adjustments where necessary to increase the amount of data used for the aggregated maps for real-time data was done. In doing so, an analysis of the daily coverage for the period October to December 2019 using data from the DWDSynop stations (3618 temperature stations and 3601 wind stations) was carried out. The calculated daily coverages (corrected for the true temporal resolution of the stations) were visualized as cumulative histogram (Figure 6.2). The results showed that about 11% of the daily data points are not used because the coverage is less than 95%, which was the initial temporal coverage requirement. Almost half of these days would be used if we reduce the coverage requirement to 87%. Based on this analysis, the temporal coverage requirements were adjusted as follows:

Table 6.1: Summary of present and updated temporal coverage requirements

Parameter	**Previous requirement**	**Updated requirement**
Precipitation	100%	100% (no change)
Temperature (Min & Max)	50%	37%
All other parameters	83%/95% (sub-daily/daily, monthly, yearly)	66%/87% (sub-daily/daily, monthly, yearly)

As the reporting behaviour is usually different for historic data compared to real-time data due to changed observing practises and opportunities (e.g. change from manual readings to automated weather stations). This is also considered in the configuration of the time series in the data bank sys-tem. This allows to identify changes in the temporal resolution of the original time series analysed. The analysis was done individually for each time series and the configuration adopted accordingly.

This effort led to an increased amount of input data to the provided analysis to the CEMS, as the data-base was used in a more efficient way.

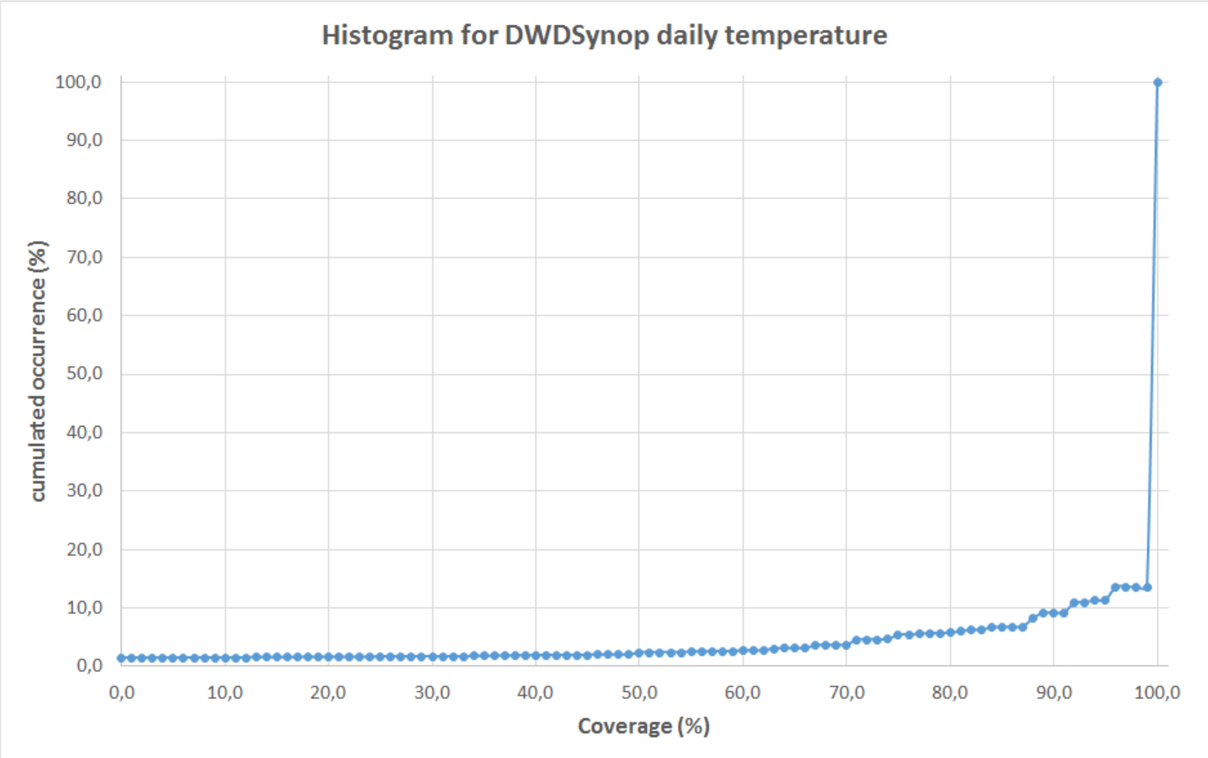

Figure 6.2: Daily coverage histogram for DWDSynop temperature data. The cumulated occurrence is the sum of all counted coverages below the given one (e.g. ~ 8% of all time series have an observed coverage below 87%).

6.3 Acceleration of the interpolation software

The computational time needed to create the grids is critical in terms of a timely provision of the analysis products to CEMS. In order to reduce the computational time in preparation of a higher spatial resolution and a computationally efficient implementation of the interpolation scheme, some changes were implemented after identifying possible improvements. The improvements reduce the computational time for the creation of one grid by around 25% (Table 6.2).

The new code reads the digital elevation model (DEM) raster file as binary file (to save time) and produces binary GeoTIFF files as output, which prevents the need of later re-formatting of the files. The other improvement in terms of calculation time was a parallelisation of the main interpolation loop. The number of processors used can be controlled via an environmental variable. Without setting the environmental variable, all available processor cores are used. Depending on the processing architecture available, it is thus possible to run either the fully parallelised implementation or several non-parallelised interpolations in parallel.

Table 6.2: Comparison of computational times to create grids for daily maximum temperature for the old and improved version of the SPHEREMAP implementation at one core.

SPHEREMAP version	Grid points / spatial resolution	Computational time
Old	950,000 / 5 km	26 s
Improved	950,000 / 5 km	19 s
Improved	13,454,100 / 1 arcmin	194 s

7 Gap Analysis

7.1 Gap Analysis

Even if the interpolation scheme estimates data in un-probed regions, the reliability of grids is higher and uncertainty is lower, if the input stations are spatially homogeneous distributed. Within the EFAS domain, some regions have a very high station density whereas in other regions, large distances be-tween the stations can be observed. Additionally, the station density depends on the parameter, as not all stations measure all parameters of interest nor deliver all data providers all observations at all stations. Meteorological services operate many sensors at a station, but the station density is sometimes low, even if it is sufficient for the legal tasks of the specific service. On the other hand, hydrological services operate often a station network with a higher station density, but focus mainly on precipitation and only at a few stations observe temperature, wind or solar radiation. As currently not all existing meteorological and/or hydrological services are contributing with their data to the METEO data-base, the spatial distribution of available stations and parameters is very inhomogeneous. This section aims to detect gaps in the real-time data availability to give an advice for future data collection activities.

The backbone of the available data are fetched from the World Meteorological Organization (WMO) Global Telecommunication System (GTS). These data (referred to as "Deutscher Wetterdienst (global) in Table 3.1), exchanged from the national hydrological/meteorological services, cover the whole EFAS domain, but are not spatially homogeneous, as some countries exchange more data than other countries. This data exchange, on the other hand, covers nearly all required parameters. Such data can be densified by additional deliveries from national services based on agreements between the service and the Copernicus program. As 'add on', the deliveries from national services offer a redundant data delivery. Hydrological services, on the other hand, provide a high density of mainly precipitation stations, but mostly limited to administrative regions or catchments.

Figure 4.2 depicts the spatial distribution of real-time stations. By the end of 2020, a comparatively low station density is apparent in Algeria, Azerbaijan, Belarus, Bulgaria, Egypt, Greece, Hungary, Iraq, Iran, Ireland, Italy, Jordan, Libya, Lithuania, Morocco, Poland, Russia, Saudi Arabia, Tunisia and Turkey. On a parameter level (Figure 7.1), a comparatively low station density for precipitation occurs additionally in Iceland. Solar radiation is a parameter not observed and distributed at many stations. Countries with the highest station density delivering this parameter are Belgium, Denmark, Italy, Israel, France, Germany, the Netherlands, Portugal, Romania, Switzerland, Tunisia and the United Kingdom of Great Britain and Northern Ireland. In Spain, the river authority for the Ebro river provides solar radiation data with a higher coverage than the other parts of this county. A medium solar radiation station density is in Armenia, Estonia, Iran, Latvia, Lithuania, Russia and Turkey.

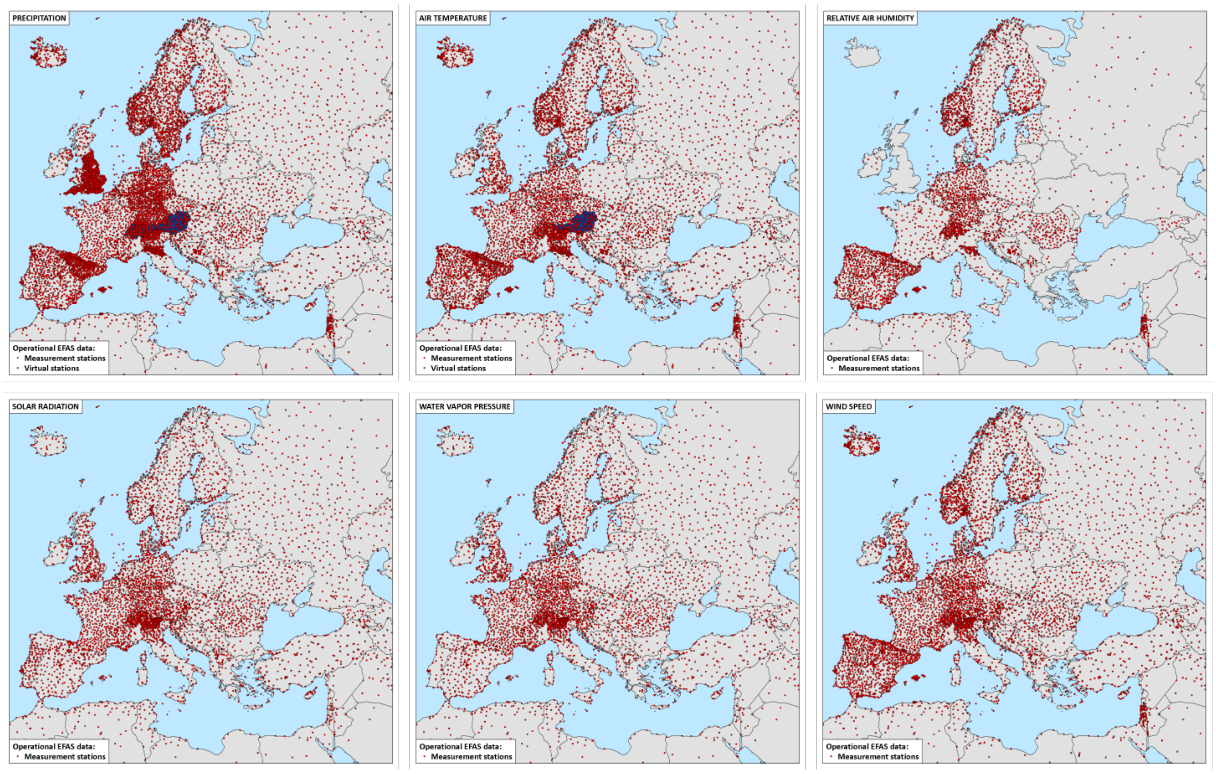

Figure 7.1: Spatial distribution of active stations providing data for the parameter noted above the map. Upper row: precipitation (left), air temperature (centre) and relative air humidity (right). Lower row: solar radiation (left), water vapor pressure (centre) and wind speed (right).

Beside the addition of new stations, an extensive use of existing data would increase the availability of data. This could be done by further improving the configuration of the calculation procedures in the post-processing in order to utilize more uncommon reporting behaviours and aggregation periods for totals (e.g. 9- or 15-hourly precipitation totals). Additionally, the calculation of non-provided parameters from provided parameters, for example the water vapor pressure from air temperature and dew point temperature, increase the amount of available data and provides a redundancy of ready data.

7.2 Proposal for future data collection strategy

Based on the gap analysis, a high priority should be given to add additional stations from countries marked blue in Figure 7.2. These are in alphabetic order Algeria, Azerbaijan, Belarus, Bulgaria, Egypt, Greece, Hungary, Iraq, Iran, Ireland, Italy, Jordan, Libya, Lithuania, Morocco, Poland, Russia, Saudi Arabia, Tunisia and Turkey.

Redundant data deliveries are currently implemented for countries marked in dark green in Figure 7.2. In alphabetic order are these Austria, Bosnia and Herzegovina, Czech Republic, Finland, Georgia, Germany, Ireland, Luxembourg, Kosovo, the Netherlands, Norway, Poland, Portugal, Spain, Slovakia, Slovenia, Sweden and Switzerland. And due to the integration of regional or hydrological data providers regional in Italy, United Kingdom of Great Britain and Northern Ireland. These countries are marked in light green in Figure 7.2.

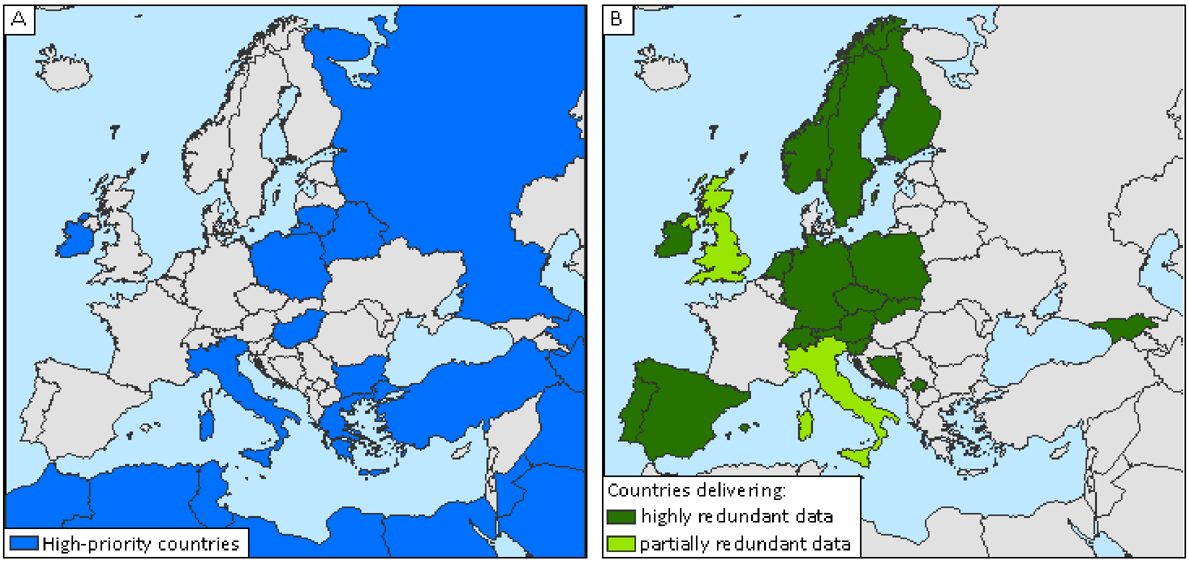

Figure 7.2: Map of the EFAS domain with (A) countries proposed to give a high priority to establish a data provision to METEO and (B) countries with redundant data deliveries (dark green) and redundant data deliveries for parts of the country (light green) to METEO.

Precipitation is of high importance in the hydrological modelling as critical input variable and due to its high spatial and temporal variability. Solar radiation is provided by a limited and inhomogeneous distributed number of stations. For the integration of additional stations, the highest priority should be given to increase the amount of available precipitation observations and secondly to a more homogeneous and redundant delivery of solar radiation data.

Additionally, the calculation of not provided parameters from available data should be implemented, e.g. the computation of water vapor pressure can be calculated from dew point temperature and air temperature values.

List of tables

JRCXXXXXX

The European Commission's science and knowledge service

Joint Research Centre

JRC Mission

As the science and knowledge service of the European Commission, the Joint Research Centre's mission is to support EU policies with independent evidence throughout the whole policy cycle.