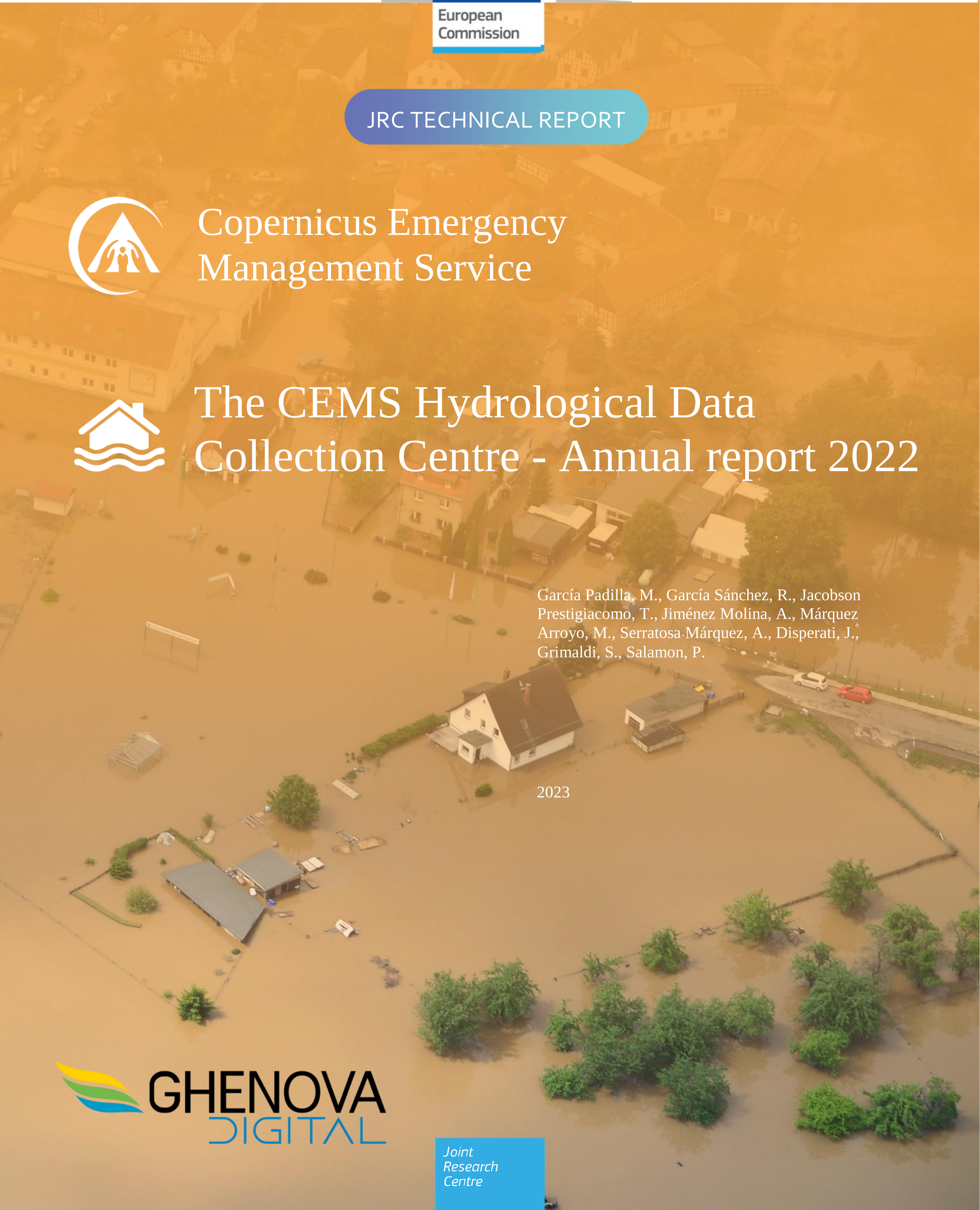
European Commission



JRC TECHNICAL REPORT

Copernicus Emergency Management Service

The CEMS Hydrological Data Collection Centre - Annual report 2022

García Padilla, M., García Sánchez, R., Jacobson Prestigiacomo, T., Jiménez Molina, A., Márquez Arroyo, M., Serratosa Márquez, A., Disperati, J., Grimaldi, S., Salamon, P.

2023

GHENOVA DIGITAL

Joint Research Centre

This publication is a technical report by the Copernicus Emergency Management Service – Hydrological Data Collection Centre (Operated by GHENOVA Digital) in collaboration with the Joint Research Centre (JRC), the European Commission's science and knowledge service. It aims to provide evidence-based scientific support to the European policymaking process. The contents of this publication do not necessarily reflect the position or opinion of the European Commission. Neither the European Commission nor any person acting on behalf of the Commission is responsible for the use that might be made of this publication. For information on the methodology and quality underlying the data used in this publication for which the source is neither Eurostat nor other Commission services, users should contact the referenced source. The designations employed and the presentation of material on the maps do not imply the expression of any opinion whatsoever on the part of the European Union concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries.

EU Science Hub
https://joint-research-centre.ec.europa.eu

JRC134937

Ispra: European Commission, 2023

© European Union, 2023

The reuse policy of the European Commission documents is implemented by the Commission Decision 2011/833/EU of 12 December 2011 on the reuse of Commission documents (OJ L 330, 14.12.2011, p. 39). Unless otherwise noted, the reuse of this document is authorised under the Creative Commons Attribution 4.0 International (CC BY 4.0) licence (https://creativecommons.org/licenses/by/4.0/). This means that reuse is allowed provided appropriate credit is given and any changes are indicated.

How to cite this report: García Padilla, M., García Sánchez, R., Jacobson Prestigiacomo, T., Jiménez Molina, A., Márquez Arroyo, M., Serratosa Márquez, A., Disperati, J., Grimaldi, S. and Salamon, P., *The CEMS Hydrological Data Collection Centre - Annual report 2022*, European Commission, Ispra, 2023, JRC134937.

Contents

Abstract

The Copernicus Emergency Management Service (CEMS) Hydrological Data Collection Centre (HDCC) is responsible for the collection, quality control, harmonisation and internal distribution of hydrological observations (river discharge and water level, reservoirs variables) for the CEMS European Flood Awareness System (EFAS) and Global Flood Awareness System (GloFAS). The Joint Research Centre (JRC) of the European Commission is the entrusted entity responsible for CEMS EFAS and GloFAS in terms of management, technical implementation and evolution. GHENOVA Digital is the designated contractor to implement the operational functionalities of the CEMS HDCC.

This report contains an analysis of the hydrological data collected during the year 2022, as well as an updated description of the data collection and quality control protocols. During 2022, the CEMS HDCC added two new data providers and increased the number of stations providing real-time and historical hydrological data for both the European extended domain (CEMS EFAS) and the global domain (CEMS GloFAS). In total, 49 data providers contributed to CEMS EFAS with near real-time and historical discharge and water level data, making a total of 2,382 hydrological stations. 46 data providers and 2,588 stations with daily historical discharge data were included in the CEMS GloFAS database.The quality control algorithm was updated with the introduction of the Summary Quality Flag. Moreover, the analysis of the near real-time data series for the European extended domain led to interesting conclusions about the hydrological conditions in 2022.

Finally, this report introduces the developments foreseen in 2023: the collection of reservoir data for the European extended domain, and the further extension of the collection of historical discharge data for the global domain.

Acknowledgements

The CEMS Hydrological Data Collection Centre (HDCC), would like to thank the EFAS Partners and Data Providers that contributed to the CEMS hydrological data collection. We would like to acknowledge their dedication to the EFAS project, their commitment and the sharing of their hydrological data. We thank them for their cooperation with the HDCC, both in the provision of data and for their proactive role in responding to questions and solving issues. Without their collaboration, the delivery of this report would not be possible.

1 Introduction

This report contains an analysis of the hydrological data collected by the Copernicus Emergency Management Service (CEMS) Hydrological Data Collection Centre (HDCC) in the year 2022. The CEMS HDCC is responsible for the collection, quality control, harmonisation and internal distribution of hydrological observations (river discharge, water level, and reservoirs data) to the CEMS European Flood Awareness System (EFAS) and Global Flood Awareness System (GloFAS). The European Commission’s Joint Research Centre is the entrusted entity responsible for CEMS EFAS and GloFAS in terms of management, technical implementation and evolution. Ghenova Digital is the designated contractor to implement the operational functionalities of the CEMS HDCC.

This report is structured as follows: Chapters 2 to 5 focus on the activity concerning the hydrological data collection for the extended European domain (CEMS EFAS); Chapter 6 presents the activities commenced in 2022 by the HDCC; and Chapter 7 presents the conclusions.

More specifically, Chapters 2 to 5 present the updates to the data collection for the extended European domain (CEMS EFAS) in 2022; the workflow for quality checks, gap analysis and post-processing; the analysis of the hydrological conditions in Europe in 2022; and the analysis of large flow conditions in Europe in 2022.

The main new developments explained in Chapter 6 are the expansion of the discharge data collection to the global domain, as used in CEMS GloFAS, and the collection of reservoir data for CEMS EFAS. It should be noted that these activities now form part of the regular tasks of the HDCC; a more detailed analysis and more comprehensive results will be included in the Annual Report for the reference year 2023.

2 Updates in the HDCC database in 2022 for the European extended domain (CEMS EFAS)

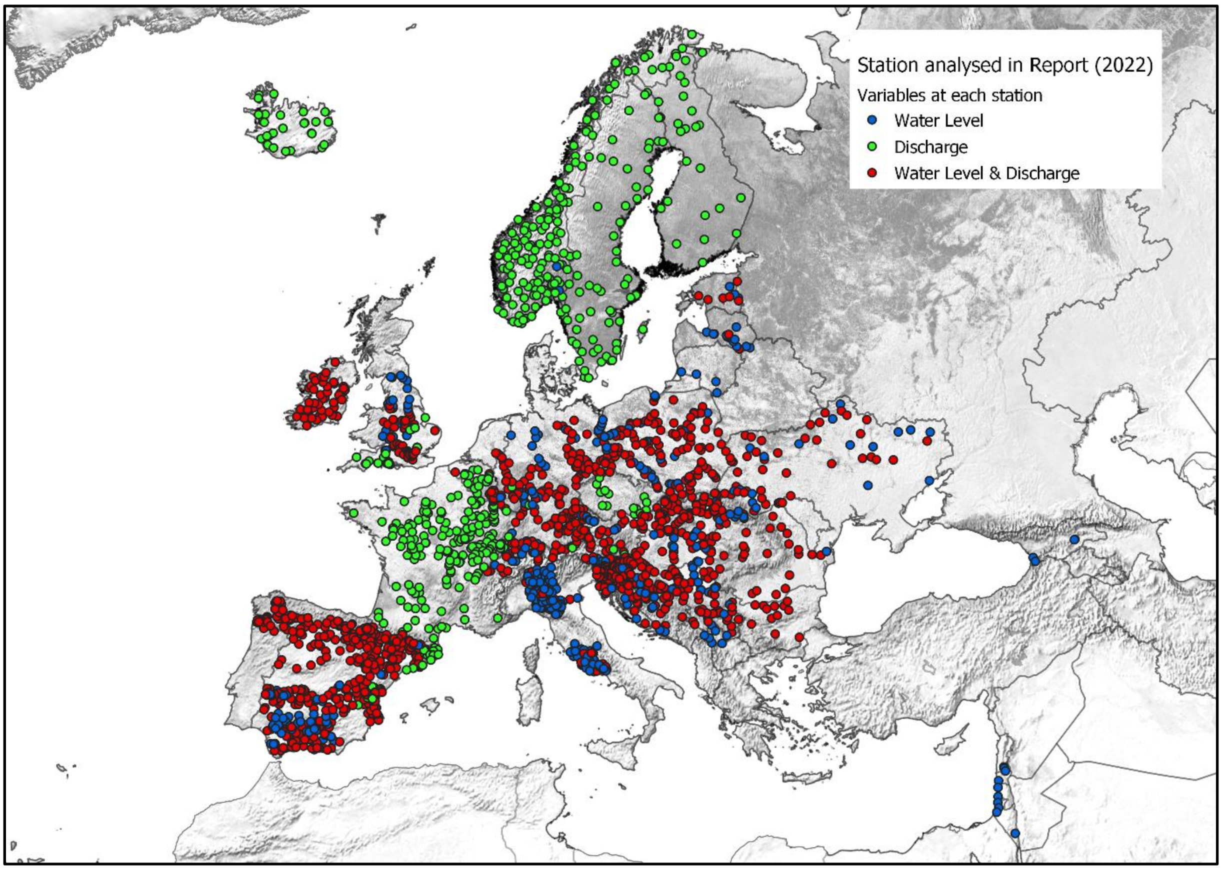
CEMS HDCC has been collecting hydrological in-situ observations for the European extended domain since 2012. The EFAS Extended domain covers the European continent and it includes also parts of North Africa and Middle East (to include more river basins located near the southern and eastern borders of Europe). Data providers can be national or regional institutions and they share their data with CEMS HDCC on voluntary basis. The CEMS HDCC strives to constantly increase the spatial and temporal coverage of the database. For this purpose, the CEMS HDCC actively participates to CEMS outreach events (e.g. the EFAS Annual Meetings) and it organizes dedicated outreach activities. Furthermore, essential information for new data providers are available from the EFAS web site.

In 2012, the CEMS HDCC database benefitted of the contribution from 30 data providers. By December 2022, 49 data providers had contributed to EFAS with near real-time (NRT) hydrological data, comprising a total of 2,382 hydrological stations, covering 33 countries and 54% of all European water basins.

Since January 2022, 2,270 stations have been active and will be analysed in this report. Of these stations, 460 offer exclusively water level data, 503 offer exclusively discharge data, and 1,307 stations provide data for both variables.

Figure 1 shows the spatial distribution and variables of the 2,270 stations analysed in this report.

Figure 1. Spatial distribution of the variables analysed at each station.

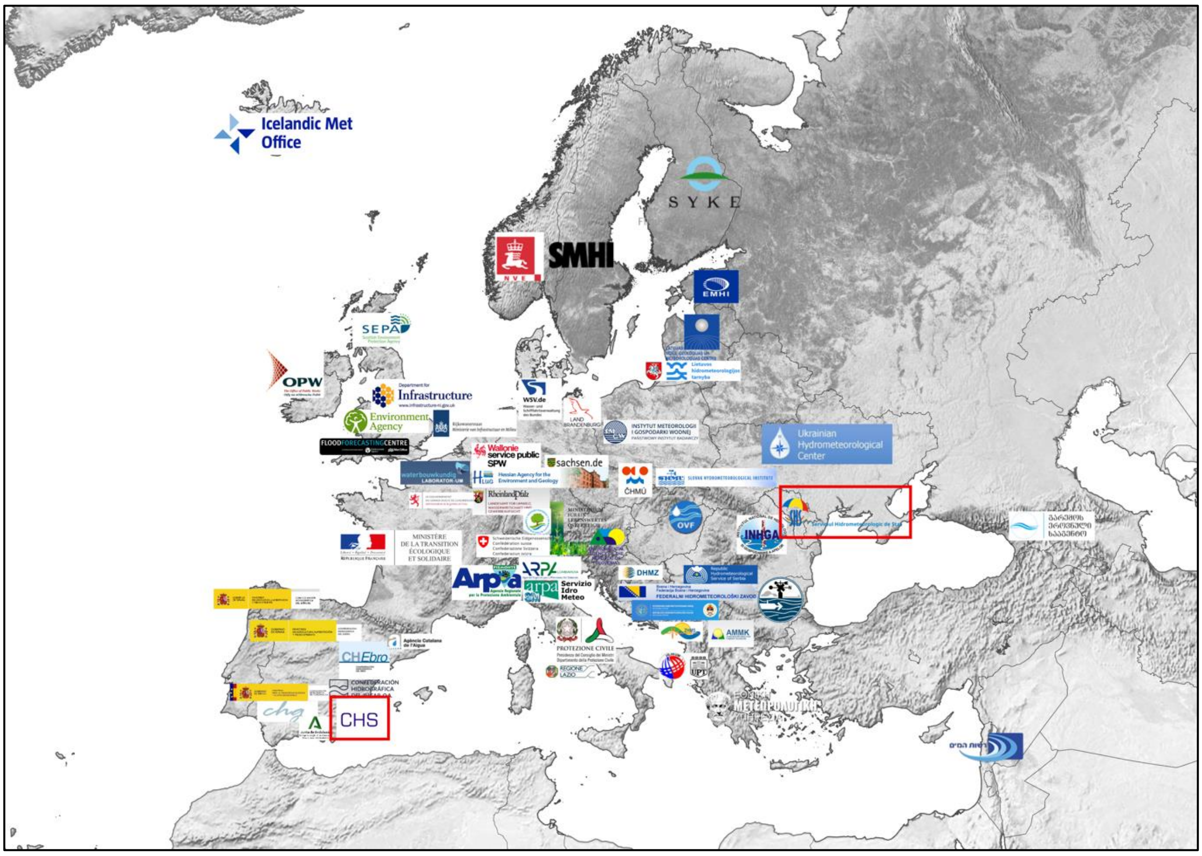


The 49 data providers that delivered near real-time hydrological data are shown in Figure 2.

Two new providers started contributing to EFAS hydrological data collection in 2022:

- Confederación Hidrográfica del Segura, CHS (ES, 81 stations).
- State Hydrometeorological Service of Moldova (MD; 3 stations).

Figure 2. Spatial distribution of data providers (new providers shown in a red box).



Moreover, in 2022, 97 new stations from existing data providers were included in the EFAS database. Table 1 describes the data providers that activated new stations. The table in Annex 1 provides the number of stations for each data provider contributing to the EFAS extended domaain.

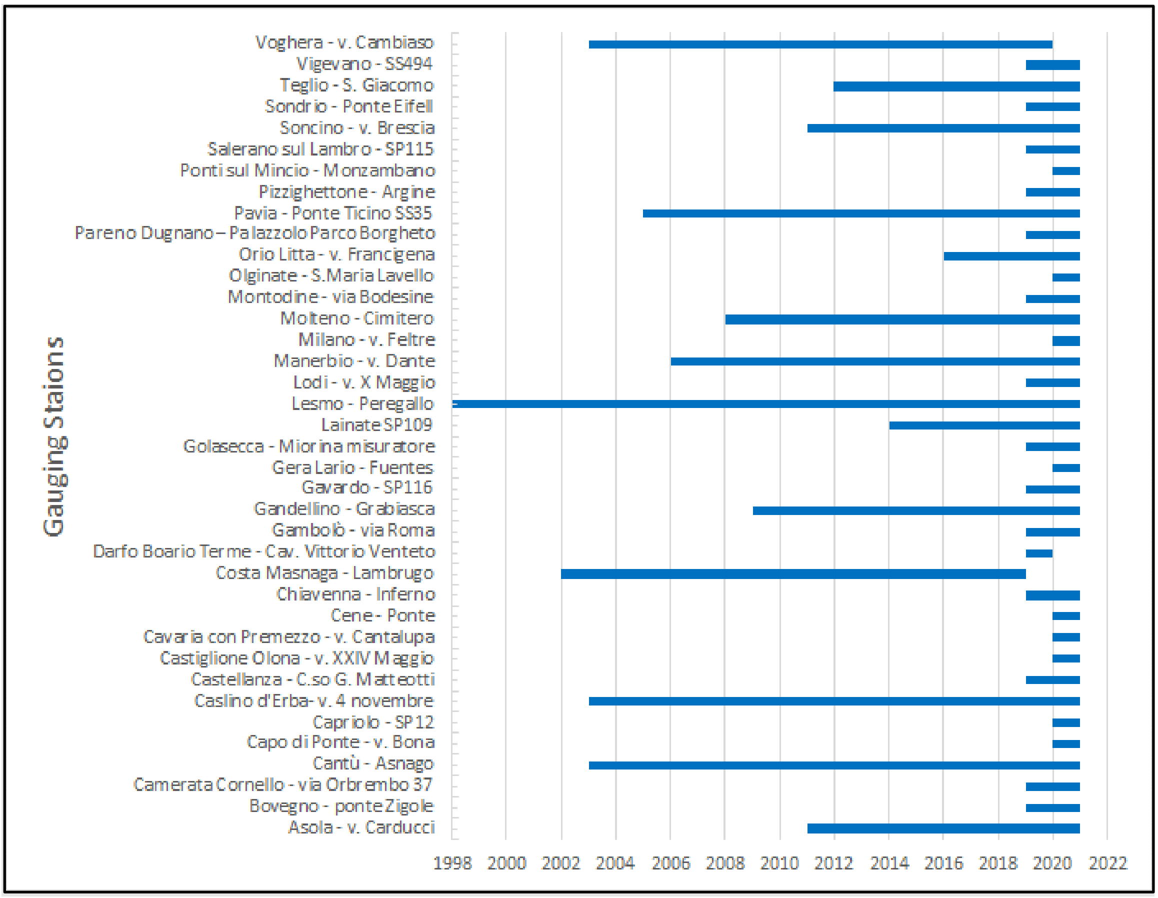
Table 1. Additional stations per CEMS HDCC data provider

Data provider	No. of NRT stations added (% increase)	Variables	Total number of active stations
ARPA Lombardia (IT)	3 (5.45%)	2 stations with water level 1 station with discharge	55
Bayerisches Landesamt für Umwelt (DE – BV)	1 (1.45%)	Water level	69
Confederación Hidrográfica del Guadiana (ES)	2 (3.12%)	Discharge	64
Confederación Hidrográfica del Segura, CHS (ES)	81 (100%)	81 stations with water level and discharge	73 (Not all the stations included are active because thedataset not extensive enough to allow the definitions of the range control quality check)
Federal Ministry of Agriculture, Forestry, Environment and Water Management (AT)	1 (1.58%)	Water level and discharge	63

Hydrometeorological Institute of Kosovo (XK)	2 (25%)	Water level	9
Institute of Meteorology and Water Management Wroclaw Branch (PL)	1 (0.69%)	Discharge	145
Croatian Meteorological and Hydrological Service (HR)	1 (1.75%)	Discharge	57
Republic Hydrometeorological Service of Serbia (RS)	1 (1.85%)	Water level and discharge	54
Ukrainian Hydrometeorological Center, State Emergency Service of Ukraine (UA)	1 (1.37%)	Discharge	73
State Hydrometeorological Service of Moldova (MD)	3 (100%)	Water level and discharge	3

In terms of the historical data table, 49 providers make up a total of 3,144 stations. The timeframe for the historical data is from 1806 to December 2021. In 2022 new data from the data provider ARPA Lombardia (Italy) were uploaded in the system. The ARPA Lombardia provider had previously shared data with HDCC, however historical data were later added for 38 stations. See Figure 3 for more details about the temporal range of the historical data provided.

Figure 3. Range of new historical data uploaded for ARPA-Lombardia (IT) in 2022.



3 Data flow

The data flow between the data providers and the CEMS HDCC consists of some important steps, which are described below. Hydrological data is stored in two main tables based on the delivery protocol: near real-time (NRT) data and historical data.

Near real-time (NRT) data implies that there are no significant delays in data collection, only the time requested for electronic communication and automatic data processing. Data values are not previously reviewed by the data providers. The frequency of the data can vary from minutes to 24 hours.

The second main table shows the historical data. This table contains those data values collected in previous years, which are often reviewed and normalised by the provider before being shared with the CEMS HDCC. Once a year, the CEMS HDCC sends an email to the data providers requesting that they share revised historical data from previous years. Historical data are usually shared and stored in yearly batches.

Even though there are two different tables, the main steps of the workflow are similar. The main difference is found in the first step:

1. **Data providers share their data with the CEMS HDCC**. Data providers can share collected data with the CEMS HDCC through different channels. NRT data are usually shared via FTP servers, HTTP servers, FTP clients or API. Historical data transmission is usually performed in batches, once a year, and typically involves the direct transfer of files via email or data repositories.

2. **Collection and processing**. Once stations and time frequency of data transmission are agreed with the provider, each station's metadata (station name, coordinates, river name, presence of backwater effects, etc.) are verified and included in the database. The accurate validation of metadata is strictly required before commencing data transmission. The data are then ingested into the CEMS HDCC's database, the Hydrological Data Management System (hDMS).

3. **Quality and quantity controls**. The data added to the database go through a set of quality and quantity control algorithms that will be described in the sections 3.1 and 3.2. The automatic algorithms for quality and quantity checks allow to identify anomalous behaviours in data due to sensor failures not indicated by the data providers. In addition to the automatic checks, manual reviews are performed so as to have a better view of the situation. For instance, when data are lost due to transmission failures, the authomatic system detects the issue and the CEMS HDCC technician performs a visual inspection. If needed, the CEMS HDCC interacts with the data providers to solve the issues.

4. **Operational data computation.** Once data have been stored and have passed the quality and a quantity controls, the values are aggregated to the desired aggregation interval. There are 3 different types of aggregation tables: hourly, 6-hourly and daily (24 hours). The aggregation methodology depends on the time resolution of the incoming data. Data delivered with hourly or lower aggregation intervalcan be aggregated to 1, 6 or 24 hours. The aggregation period is always greater than or equal to the reporting period. Therefore, data incoming daily data are not disaggregated. The operational data are the average values within each aggregation period, considering the quality controls performed previously.

5. **Share data with the CEMS system network.** The operational data table, providing a clean dataset aggregated by 1, 6 or 24 hours with quality controls, is the main product offered by the CEMS HDCC to the other Copernicus services.

3.1 Protocol to quality check the data

Since 2012, the CEMS HDCC has had the important task of collecting and normalising hydrological data. Many quality checks have been implemented over the years in order to increase the reliability of the database. Quality checks are performed in two phases: first, upon reception of the data (raw data) and second, when data are aggregated (aggregated data) to provide the operational values.

3.1.1 Quality flags associated with each value

A quality flag is computed for each value that enters the CEMS HDCC database (for both water level and discharge data). This quality flag describes the reliability of the data. There are four different flags that can be assigned to a data value:

- PASS: observation value has passed the quality controls.
- FAIL: observation value has not passed the quality controls.
- SUSPICIOUS: observation value is not reliable.
- UNCHECKED: quality check could not be applied for various reasons, including undefined check parameters, missing reference value, and observation value outside the defined range of applicability of the check.

A summary of the quality checks applied to each incoming data is provided below:

- Range test (QL1): the range quality check verifies whether a measurement value ‘v’ is included between a minimum value Vmin and a maximum value Vmax. Vmin and Vmax are specific for each station and variable. These threshold values may be defined from the historic maximum and minimum values registered by a station and as communicated by the relevant data provider. In case the data provider cannot share this information, the CEMS HDCC, based on their experience, can set these thresholds depending on different variables such as the minimum or maximum historical value, the minimum critical alert value or maximum return period, the interpolated value from discharge rating curve, and the minimum or maximum value extracted from the time series.

- Rating curve test (QL2): quality control applies to discharge data only. Furthermore, QL2 can be applied exclusively to stations where the rating curve is defined and validated. The rating curve is the numerical relationship used to compute discharge values from measured water level values. QL2 verifies whether the discharge value follows the station’s characteristic rating curve. The rating curve is often defined by the provider; however, if this is not the case, the CEMS HDCC analyses the data series and estimates the rating curve for stations where both water level and discharge data are available.

- Maximum relative daily variation test (QL4): this control verifies whether the relative variation of a measurement with respect to the measurement made 24 hours before is lower than a specific threshold. These thresholds are determined after analysing the hydrograph daily relative variation of each variable and station.

- Maximum relative variation test with respect to previous value (QL5): this control verifies whether the relative variation of a measurement with respect to the previous measurement is lower than a specific threshold. These thresholds are determined for each variable and station analysing the hydrograph intervalue relative variation.

- Negative discharge (QL7): negative discharge values can be errors. Stations close to reservoirs, controlled channels or the coast can produce negative discharge (backwater effects). In other cases, negative discharges are errors and must be discarded. QL7 is linked to the station metadata indicating whether a station is influenced by backwater. During the metadata validation phase (see Section 3) the operator can activate the backwater flag so that the system will not remove negative values because they are considered backwater. Only those stations that share discharge values undergo this quality check.

- Repetitive values (QL8): the comparison of repeated data values helps to identify potential sensors issues. QL8 flags the constant data values as suspicious when no variation is found in a predefined period. The time period is defined by the reporting interval of the data. If the reporting interval is higher than or equal to 720 minutes (12 hours), QL8 will flag the values as suspicious after 3 days of repeated values. If the interval is less than 12 hours, the period will be 24 hours.

- Manual check (QL10): visual inspections overrule any other quality check. After the automatic quality check, if data are suspicious or fail, an alert email is sent to the technician and data are manually reviewed.

These checks are carried out meticulously because if flagged as FAIL, data values are discarded and are not employed in the aggregation table to produce the operational values.

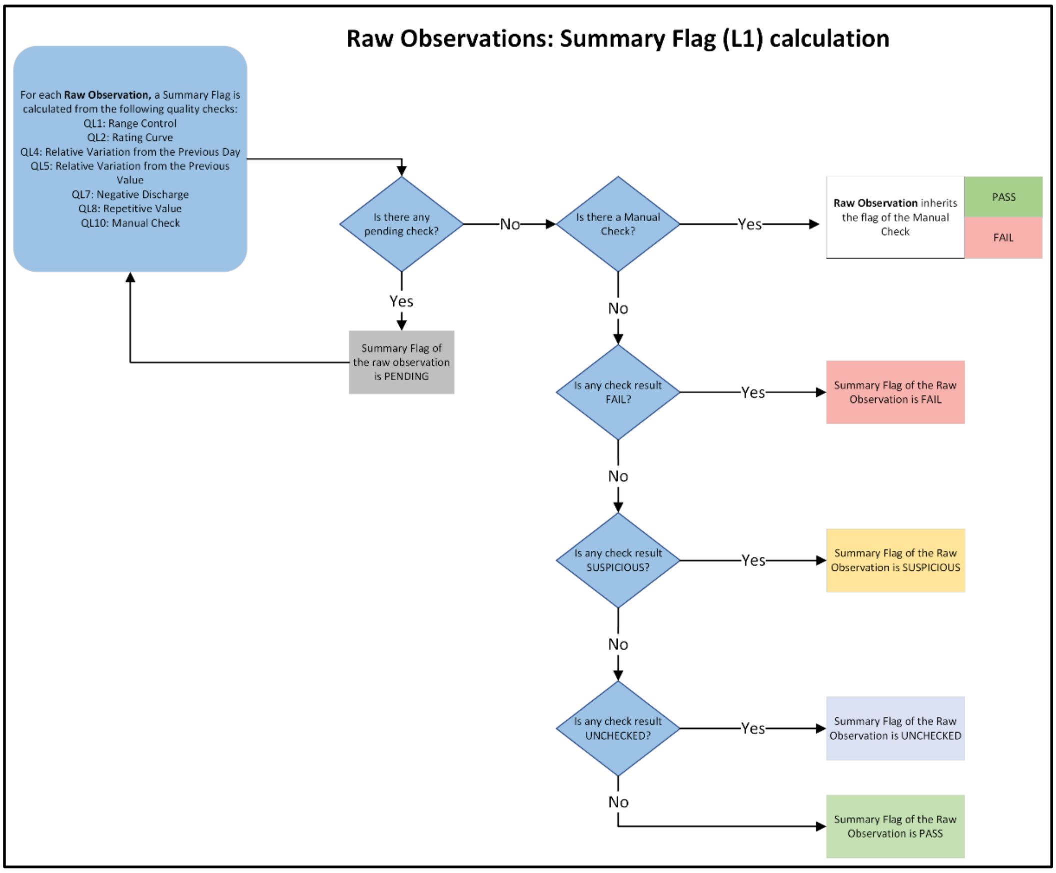
Table 2 shows the possible flags for each quality control, since not all flags can be applied to all quality controls.

Table 2. Quality checks and flags: "✓" means that the quality check can return the identified flag; conversely, "X" means that the quality check cannot return the identified flag (e.g. the manual check QL10 can only return "pass" or "fail").

Quality checks/Flags	Pass	Fail	Suspicious	Unchecked
Range test (QL1)	✓	✓	✓	✓
Rating curve test (QL2)	✓	X	✓	✓
Maximum relative daily variation test (QL4)	✓	X	✓	✓
Maximum relative variation test with respect to previous value (QL5)	✓	X	✓	✓
Negative discharge (QL7)	✓	✓	X	✓
Repetitive values (QL8)	✓	X	✓	✓
Manual check (QL10)	✓	✓	X	X

Once all the quality checks have been processed for each value, the results of the quality checks are combined to compute the Summary Flag (L1). L1 is calculated based on the union of the quality checks described above, using the workflow described in Figure 4. The quality flag L1 is therefore the result of the first phase of checks. Each value is assigned an L1 flag.

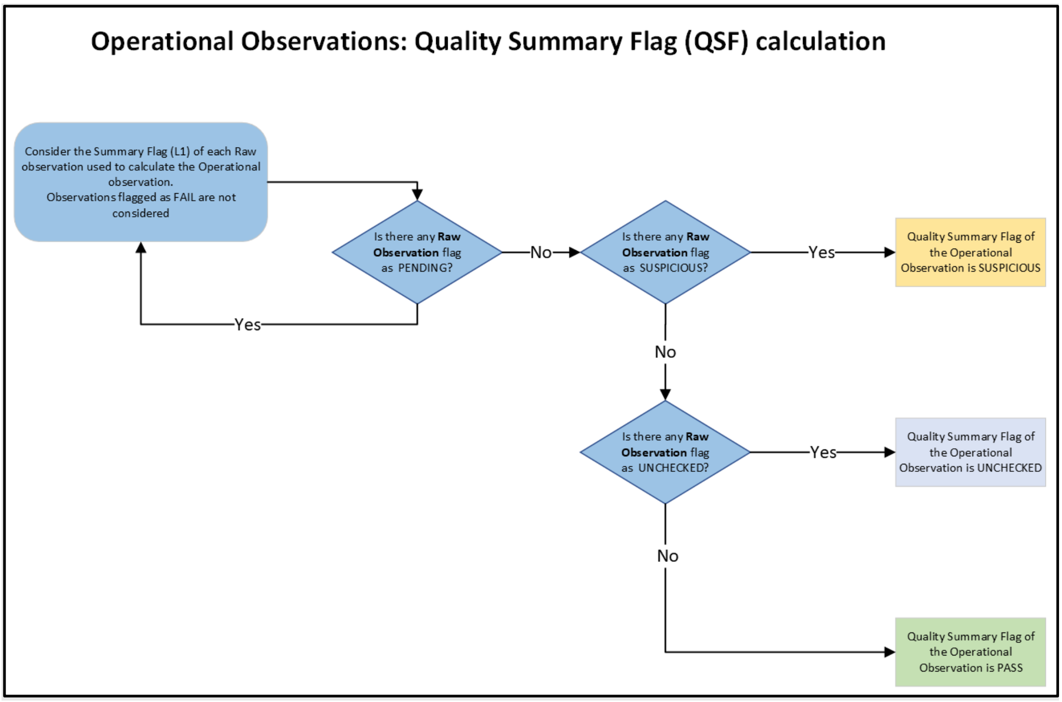
Figure 4. Algorithm for Summary Flag L1 calculation.



3.1.2 Quality Summary Flag (QSF) for aggregated data

After the L1 quality checks are computed, the data are aggregated to 1 hour, 6 hours or 24 hours. When computing the aggregation, the combined analysis of the L1 flag for each data leads to the Quality Summary Flag (QSF). In other words, the process to define the quality of the aggregated data is based on the Summary Flags (L1) of each value used in the aggregation. It is important to highlight that values with Summary Flag (L1) equal to FAIL are not considered for the calculation of the aggregated data or their quality rate. The algorithm to compute the QSF is shown in Figure 5.

Figure 5. Algorithm for Quality Summary Flag QSF calculation.



3.1.3 Results for the year 2022 on near real-time data for the European extended domain (CEMS EFAS)

The figures below (Figure 6 and Figure 7) show the distribution of the Summary Flags (L1) for the data received in 2022, for each data provider, distinguished by water level or discharge variable.

Most of the data received in 2022 has been flagged as PASS (87.1%), while the amount of data that are not valid (Fail) is small (0.2%). 10.5 % was flagged as suspicious and 2.2% was unchecked.

Values are flagged as "unchecked" when one or more quality checks could not be applied. As an example, anomalies in the data transfer interval (compared to the planned transfer interval) can lead to a large number of unchecked values for quality checks QL4 and QL5 (as explained in section 3.1.1, these checks require information for the previous day or time step).

Figure 6. Percentage for Summary Flag L1 for each data provider on data received in 2022: water level variable.

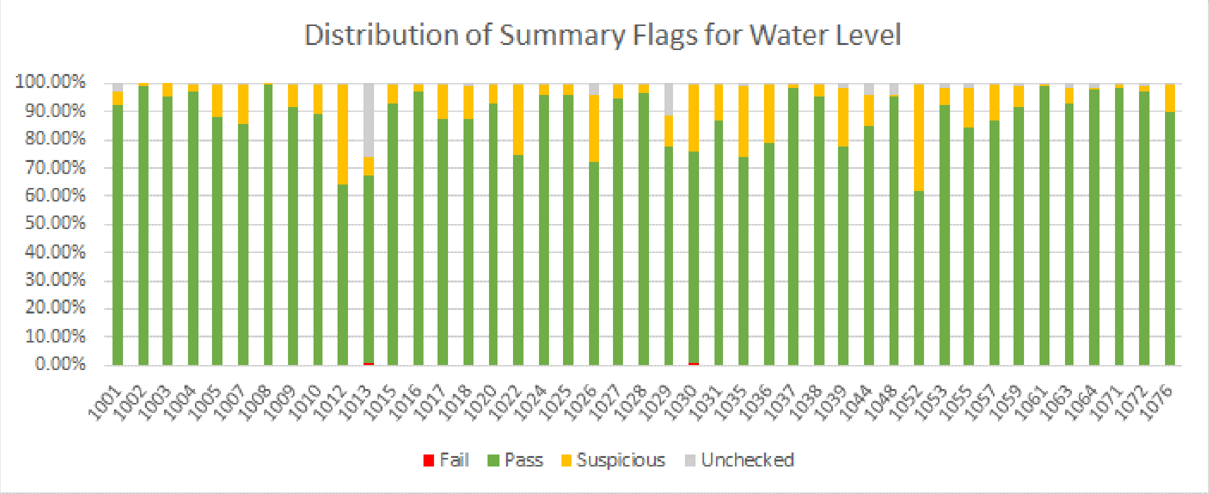
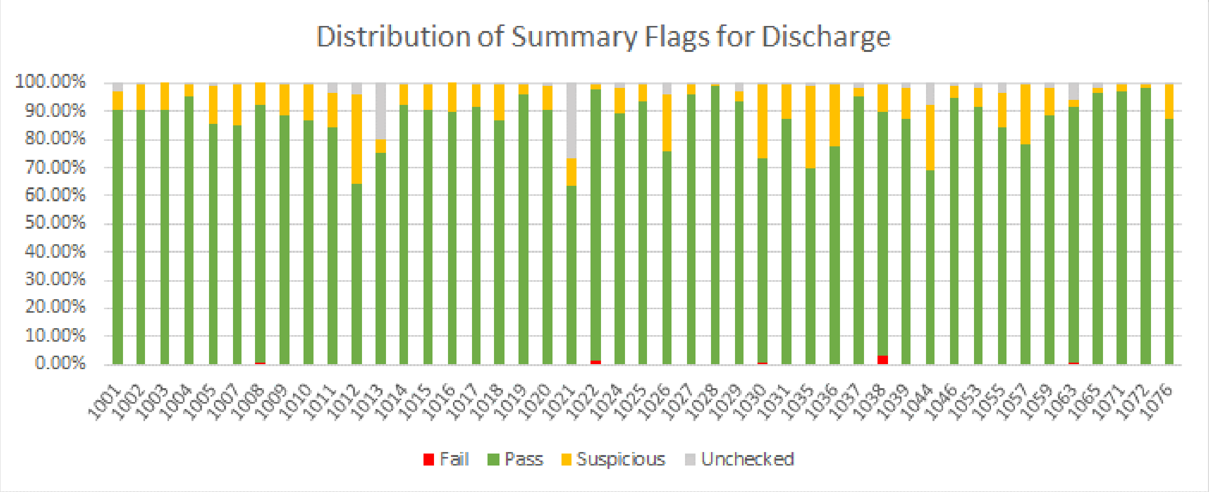


Figure 7. Percentage of Summary Flag L1 for each data provider on data received in 2022: discharge variable.



The following figures show the results of the main quality checks applied for all the data providers, and considering both variables (discharge and water level).

Figure 8. Boxplot range control (QL1).

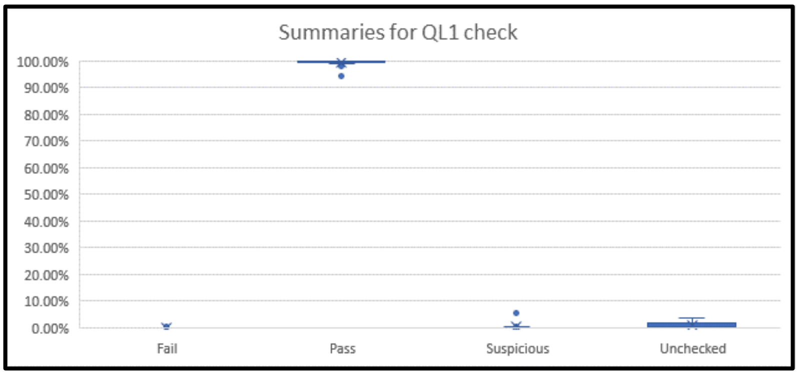


Figure 9. Left: boxplot relative variation compared to previous day value control (QL4). Right: boxplot relative variation compared to previous value control (QL5).

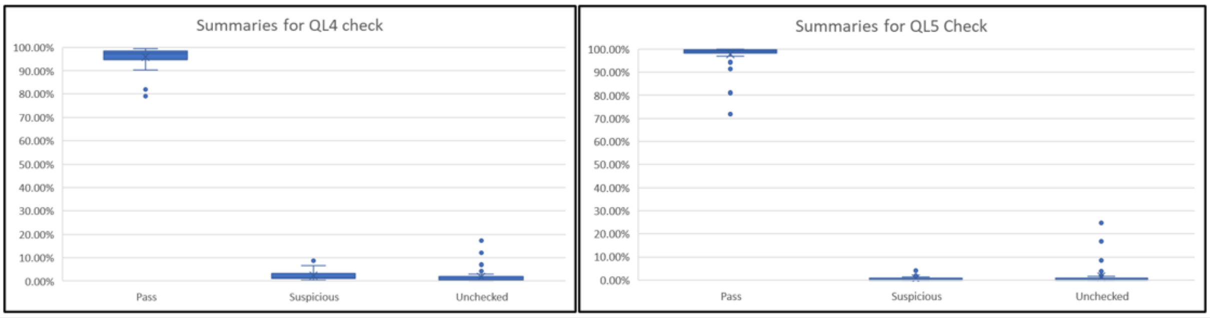
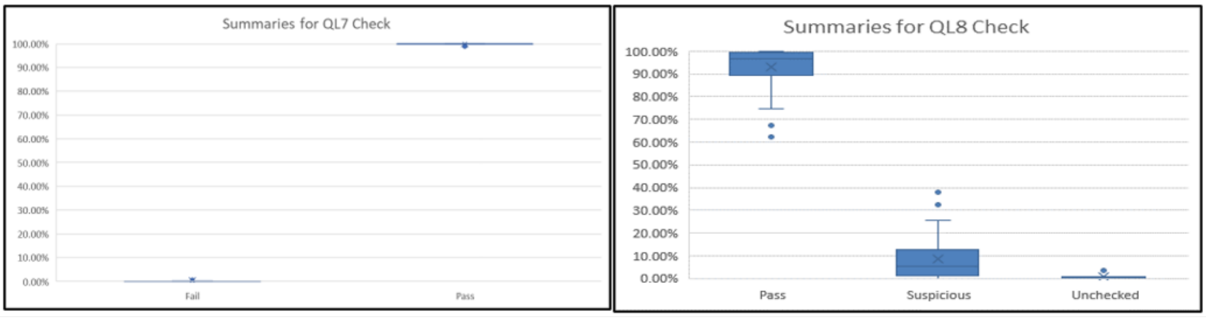


Figure 10. Left: boxplot negative discharge control (QL7). Right: boxplot repetitive value control (QL8).

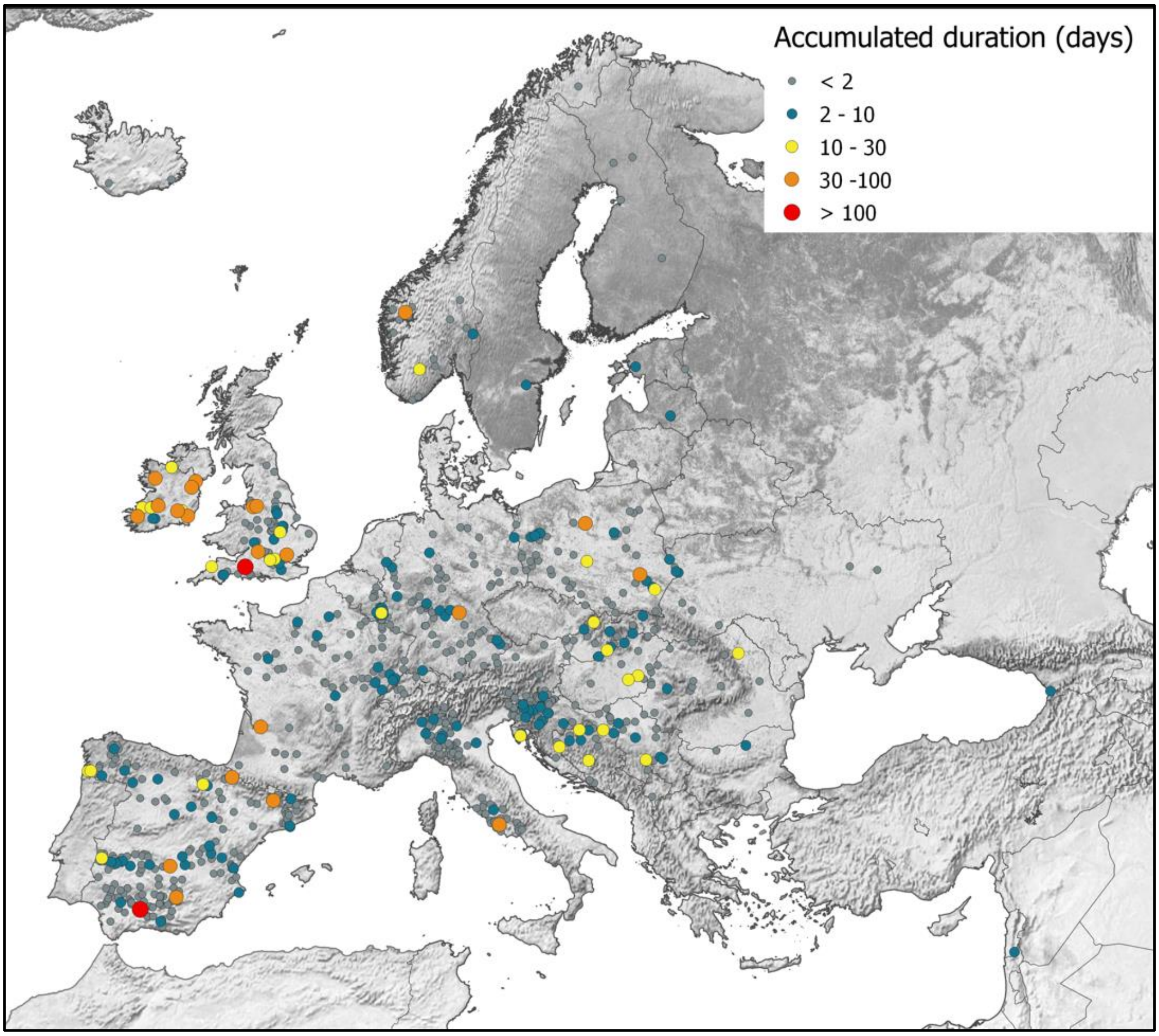


The range control quality check QL1 allows to identify outliers. Minimum and maximum values of discharge and water level are defined based on provider information or on the analysis of historical data. Outliers are defined as values that are beyond their minimum or maximum (threshold) levels. Outliers are automatically marked for further visual inspection. This step is necessary to establish whether the value is actually an error or a valid value produced by a natural event. If several consecutive outliers are detected, these are defined as a set of erroneous data values.

A total of 392,096 outliers were detected in data from 809 stations, which represent 0.4% of the total data received. This ratio is higher than the previous year (0.13%).

Figure 11 shows the spatial distribution of the stations which reported at least one outlier, and the cumulative duration of the reported outliers. The stations with the higher number of outliers usually report invalid data. The attribute “quarantine” is then used to mark stations with unreliable data.

Figure 11. Stations with outliers in 2022 and their accumulated duration (days).



3.2 Quantity analysis for the European extended domain (CEMS EFAS), year 2022

This chapter analyses the gaps in the CEMS hydrological data collection for NRT data in the extended European domain during 2022.

3.2.1 Methodology for gap analysis

One of the main attributes of a gauging station is the reporting interval, which corresponds to the time interval between the reception of one value and the next one. This reporting interval depends on the data provider; moreover, a data provider may share stations with different reporting intervals.

Based on this interval, a gap can be defined as the absence of a number of values in the expected timestamp. These “missing values” create a gap. The basic gap unit defined in this document is considered a single missing value. A gap starts with the first missing value and ends once the data delivery is resumed.

The gap can be filled if the data provider sends the missing values. For this purpose, the CEMS HDCC sends a specific request to the data provider. In case the data provider does not deliver the missing data, a gap can be filled by interpolation when its length is less than or equal to three days. Interpolation is calculated by taking the value prior to the gap and the first value after the gap.

This chapter analyses the gaps that occurred in the 2,270 stations that were actively delivering data for 2022 since 1st January. Of the expected data, 2% is missing, a quantity lower than the previous year, when this ratio was 4%.

Specifically, 2,129 stations (47 providers) had gaps due to data transmission issues for 2022. Of all the 707,875 gaps, 98.8% lasted less than one day and 88.3% lasted less than one hour. Gaps that lasted less than one hour were discarded in the subsequent analysis because they did not affect the quality of the aggregated data. This filtering reduced the number of gaps to be analysed to 82,907, with data coming from 1,990 stations and for 3,181 station-variable combinations, since each station can provide up to two variables (water level and/or discharges values).

In this dataset, five classes of gaps are defined, based on their duration:

- More than 30 days
- From 10 to 30 days (inclusive)
- From 3 to 10 days (inclusive)
- From 1 to 3 days (inclusive)
- Less than or equal to 1 day (inclusive)

Gaps in the NRT data are managed according to three scenarios:

- Filled: The gap is filled later with new data sent by the data provider.
- Filled Interpolated: The gap is filled by the CEMS HDCC data interpolation process. Gaps with a duration of 3 days or less are automatically filled. However, interpolated data are replaced if the data provider sends the real values later.
- Not Filled: No interpolation or filling is carried out.

3.2.2 Results for the year 2022 on near real-time data for the European extended domain (CEMS EFAS)

Figure 12 and Figure 13 show the number of gaps according to their duration and the percentage of each group. The duration of 89.9% of the gaps is 1 day or less. The gaps with a duration between 1 and 3 days is 7.5% of the total analysed. Just 2.8% of the gaps last more than 3 days: these gaps require follow-up actions by the CEMS HDCC.

In 2022, the percentage of gaps that could be filled using data delivered by the providers was 6.7% for gaps with a duration less than or equal to 1 day; 5.5% for gaps between 1 and 3 days; 20.5% for gaps between 3 and 10 days; 11.7% for gaps between 10 and 30 days; and 11.9% for gaps longer than 30 days.

Figure 12. Total gaps analysed for 2022.

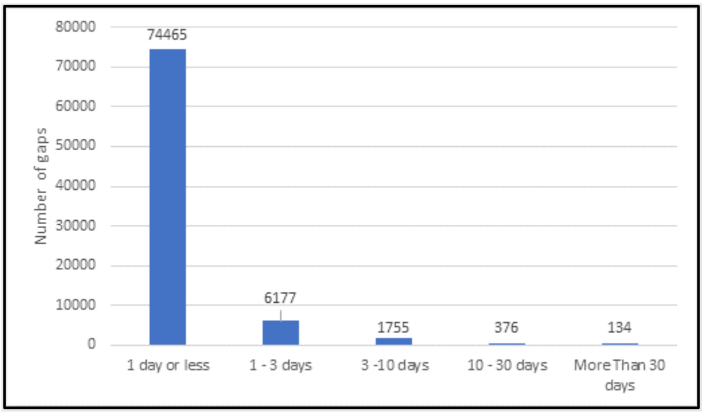


Figure 13. Percentage of groups based on the gap duration for 2022.

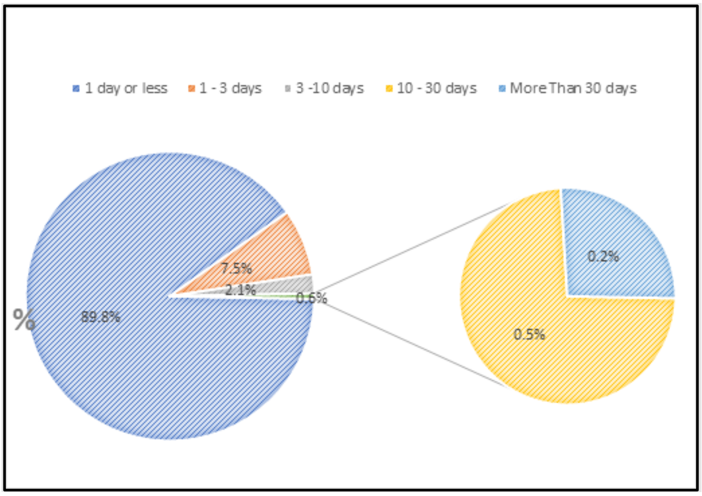


Figure 14 shows monthly boxplots with the average percentage of received data against expected data for each data provider. The monthly mean values range between 95.3% and 99.4%, with a total average of 97.5%. Comparing this value with the previous year, the average percentage is slightly higher (97.5% in 2022 versus 95.3% in 2021), however this year the monthly delivery rates of some data providers were lower than 50%. There are stations that encountered transmission problems over the course of the year and were therefore set as non-operative (these stations are considered in the monthly ratio). The title “non-operative” helps the technical group know that data from a variable is not going to be received temporarily or definitively.

Figure 14. Boxplot showing the monthly percentage of data received, out of the expected number of data records to be delivered from EFAS data providers.

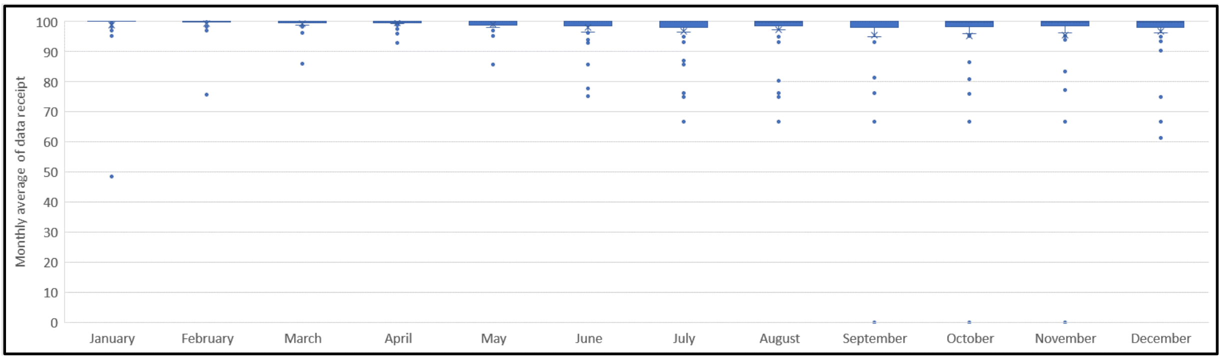


Figure 15 and Figure 16 show the distribution of the gaps, related to the average gap duration and maximum gap length (both in days), respectively.

Figure 15. Average gap length in days per station.

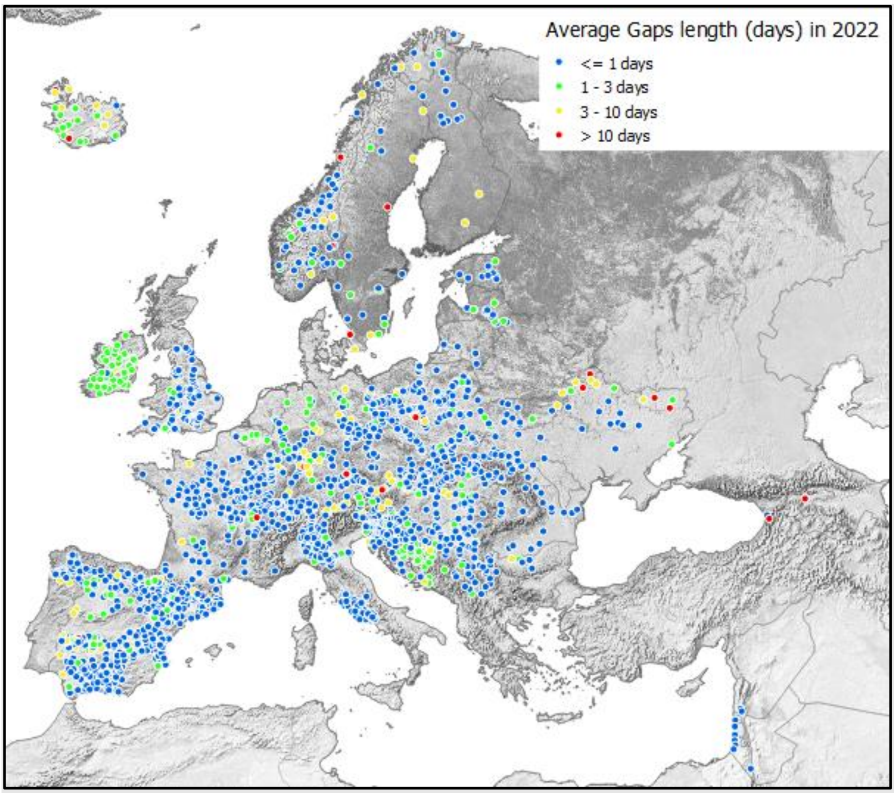
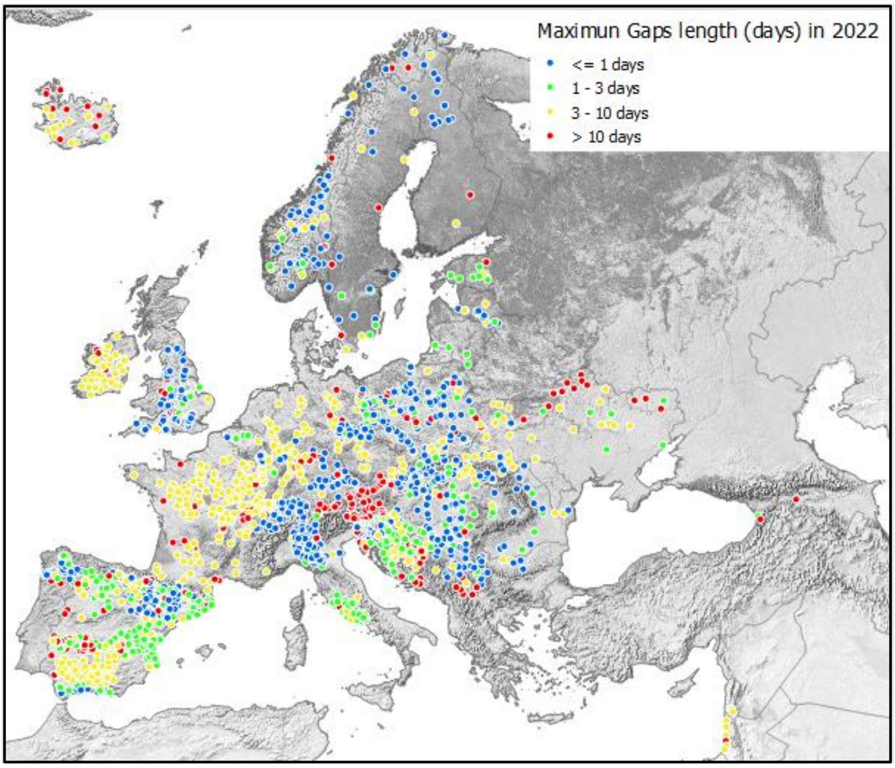


Figure 16. Maximum gap length in days per station.



4 Hydrological conditions of EFAS gauging stations

This chapter describes the hydrological conditions for the year 2022 across the entire European extended (EFAS) domain by comparing near real-time data from 2022 against near real-time data from 2021 and historical data (1991–2019), respectively. The historical reference period (1991–2019) used for this report (as well as for the report of the reference year 2021: García Padilla et al., 2022) will be maintained in the future reports, thus enabling a relative comparison.

Although the HDCC collects water level and discharge values, the analysis presented in this chapter is based on discharge data only. Unlike water level, discharge does not depend on the river's geometry and hence allows for a comparison of the hydrological behaviour between stations.

The mean daily values used to calculate the statistics for the analyses are the annual mean minimum and maximum for 2022 (the minimum/maximum value of each day, averaged for each month, and then averaged for the year); percentiles of the year 2021; and percentiles of the period 1991 to 2019, respectively. As mentioned above, the reference period 1991–2019 was used in this Annual Report for consistency with the previous Annual Report (García Padilla et al., 2022). The average of the annual mean is an indicator of the annual water contribution at the gauging points, whereas the percentiles enable comparing the annual minimum and maximum in 2022 with the reference periods.

The measurements from stations located downstream of hydraulic infrastructures often show variations that are not caused by meteorological factors. However, the impact of anthropogenic influence on natural river flows is not accounted for by the results of this chapter.

4.1 Gauging stations: drainage area and length of the time series

To ensure a reliable analysis, only stations that met the following conditions were used:

1. 2021 and 2022: only stations that were fully operational and active January through December and that received more than 75% of their expected discharge observations over the year.
2. 1991 to 2019: only stations with at least two years of data.

As a result, a total of 1,703, 1,593 and 1,419 stations were used for 2022, 2021 and 1991–2019, respectively.

Figure 17. Spatial distribution of stations according to the length of their historical time series (left) and catchment size (right).

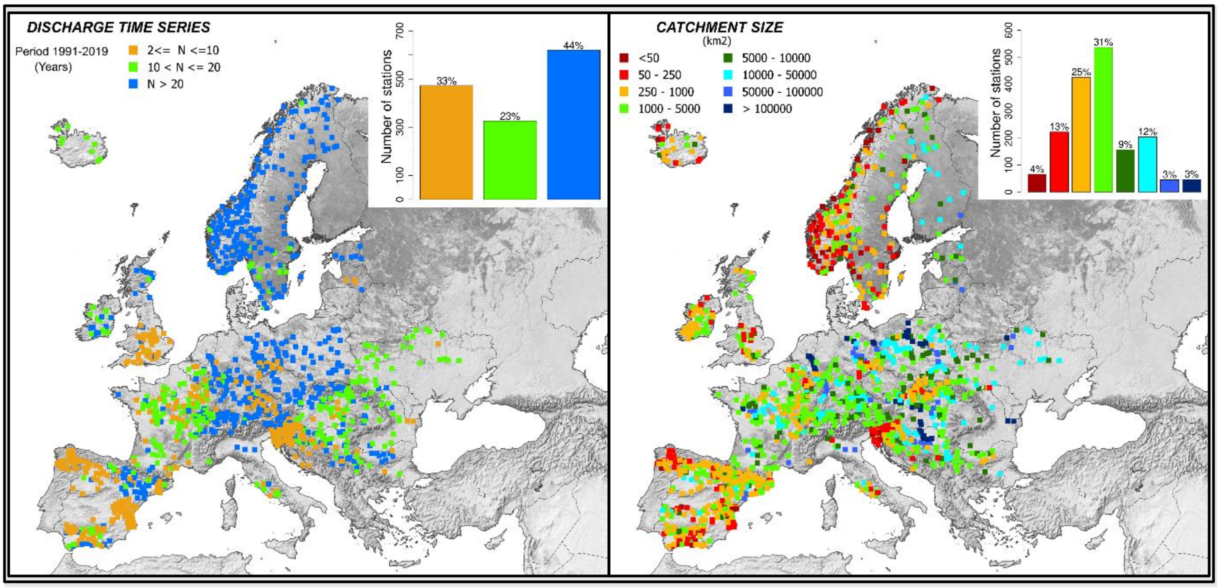


Figure 17 (left) shows the spatial distribution of the hydrological gauging stations with at least 2 years of data in the time interval 1991–2019; the figure also shows the length of the historical time series. The longer the time series, the more representative the derived statistical parameters. Remarkably, over 40% of the stations have more than 20 years of historical data. It is expected that the analysis of this chapter has higher information content in areas with longer historical time series such as Norway, Sweden, Poland, Estonia, Scotland, the Ebro River Basin in Spain, and stations across the Rhine, Elbe and Danube River Basins.

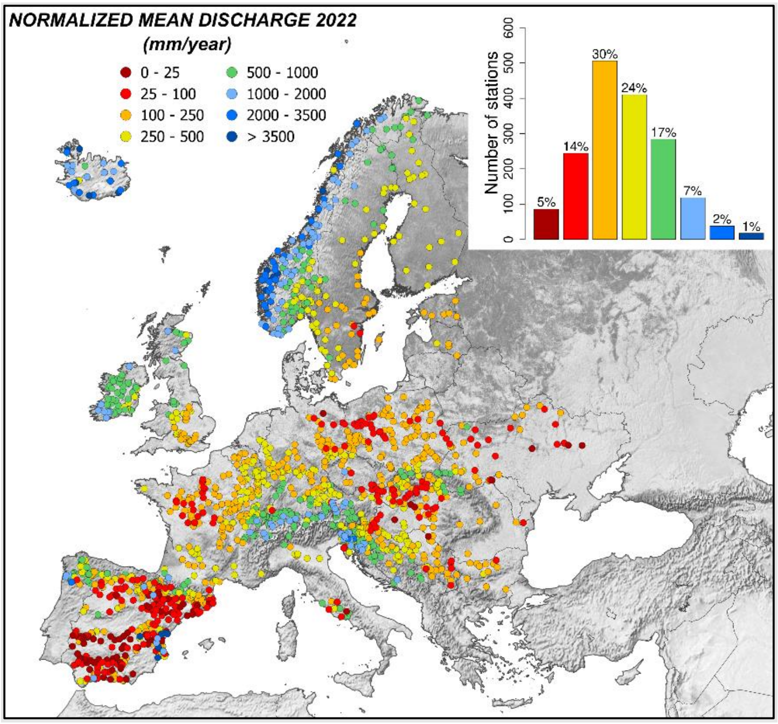
Figure 17 (right) shows the upstream areas of all the selected stations to be studied. Many of the stations from the Scandinavian peninsula, Spain, Slovenia, England, Iceland and Luxemburg have small catchment areas (< 250 km^2), whereas many of the stations from the Danube, Oder, Elbe, Vistula, Ebro and Rhine River Basins and Finland have large upstream areas (> 10,000 km^2). The distribution of the station catchment areas depends on three main factors: firstly, the catchments' hydrological features; secondly, the location where hydrological services monitor the hydrological situation; and lastly, the observations the hydrological services are willing to share.

For the purpose of the analysis of this chapter, discharge values have been normalised using the upstream area: the use of area-specific discharge values (or discharge per unit contributing area) facilitates the comparison between stations. Nevertheless, differences in catchment areas are still likely to influence the minimum and maximum values and annual variability. Smaller catchments typically have a larger difference between minimum and maximum specific discharge values and larger annual variability than larger catchments. The units of the normalized discharge (discharge per unit contributing area, as explained above) are millimetres of water per year (mm/year), which is the same as litres per square meter per year [l/(m^2·year)].

4.2 Normalised mean discharge in 2022

Figure 18 shows the normalised mean discharge values for 2022. 19% of the stations present values below 100 mm/year. These values usually belong to dry meteorological regimes and/or regulated or overexploited streams. These are mostly located in Spain, central Italy, Loire, Seine, Elbe, Oder, Dnieper, Vistula and the northern and central Danube River Basins. The highest values (over 1000 mm/year) occur for stations in Norway, Iceland, Ireland, Scotland, north-western and eastern Spain, Slovenia, Switzerland and the northern Rhône and western Danube River Basins, and usually occur in relatively small catchments.

Figure 18. Spatial distribution of normalised mean discharge values in 2022.



4.3 Historical analysis

4.3.1 Streamflow variation index and normalised variation index

In this section, the hydrological situation of 2022 is compared to the previous year (2021) and to the historical reference period (1991–2019). The purpose of this analysis is to identify and investigate anomalies in the 2022 hydrological conditions. The comparison of the relative variation of the average values is done by means of two indices: the Streamflow Variation Index (SVI) and the Normalised Variation Index (NVI). Both the indices are computed for each single station.

SVI is applied comparing 2022 and the period 1991–2019. It is adapted from the Streamflow Drought Index (SDI) (Nalbantis, 2009):

$$SVI \quad = \frac{\acute{X}_{2022} - \acute{X}_{H}}{S_{H}}$$

where $\acute{X}_{2022}$ and $\acute{X}_{H}$ are the mean discharges for 2022 and 1991–2019, respectively. S_H is the standard deviation of the annual mean discharge for the period 1991–2019. This index is a standardisation of annual mean discharge in 2022 according to the annual mean and the standard deviation of the annual mean discharge in the period 1991–2019.

The SVI is not applicable when the reference period covers only one year; in this case, the Normalised Variation Index (NVI) is used. The NVI is a normalised difference index that allows the comparison of two measurements taken in the same location but at two different times.

The NVI is used to compare the 2022 and 2021 mean discharges:

$$NVI \quad = \frac{\acute{X}_{2022} - \acute{X}_{2021}}{\acute{X}_{2022} + \acute{X}_{2021}}$$

where $\acute{X}_{2022}$ and $\acute{X}_{2021}$ are the mean discharges for 2022 and 2021, respectively.

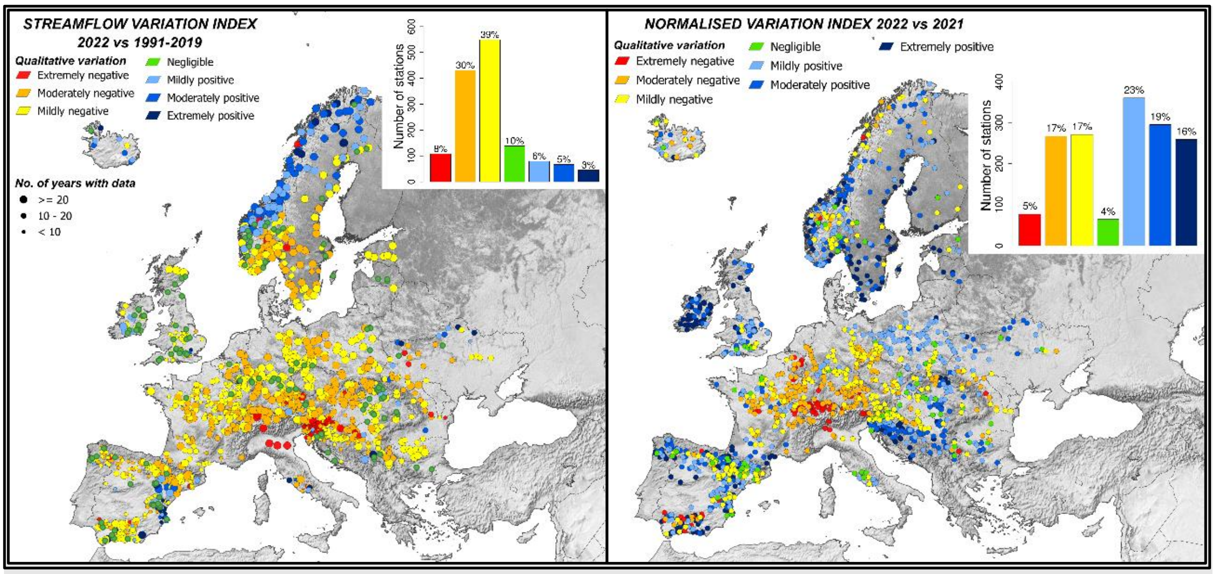
Table 3 defines eight classes based on the distribution of the resulting SVI and NVI values. Positive SVI means that the mean discharge for the year of study was higher than the mean discharge for the 1991–2019 period, and negative SVI means that the mean discharge was lower than the mean for the reference period. Positive NVI means that the mean discharge for 2022 was higher than the mean discharge for 2021, while negative NVI means that the mean discharge for the year of study was lower than the previous one.

Table 3. SVI and NVI classes. Positive/negative indices indicate a larger/smaller mean discharge in 2021 compared to 1991–2019 (SVI) or 2020 (NVI)

Classes	**SVI interval**	**NVI interval**
Extremely positive	SVI > 2	NVI > 0.5
Moderately positive	2 ≥ SVI > 1	0.5 ≥ NVI > 0.25
Mildly positive	1 ≥ SVI > 0.25	0.25 ≥ NVI > 0.02
Negligible	0.25 ≥ SVI ≥ -0.25	0.02 ≥ NVI ≥ -0.02
Mildly negative	-0.25 > SVI ≥ -1	-0.02 ≥ NVI > -0.25
Moderately negative	1 > SVI ≥ 2	-0.25 ≥ NVI > -0.5
Extremely negative	SVI < -2	NVI < -0.5

When comparing 2022 with the 1991–2019 period, the analysis highlights a majority of negative variations, with 39% mild, 30% moderate, 8% extreme (Figure 19, left); less than 25% of the stations had negligible or positive variations; and 8% of stations showed extremely negative variations. Most of these stations are located in the upper Danube River Basin and small basins in Slovenia, the upper Rhine and Po River Basins, and basins in Ukraine, Spain, Norway and southern England. Conversely, 8% of the stations present a severe or moderate surplus of mean discharge in 2022 compared with the period 1991–2019. They are mostly located in basins in Iceland and Scandinavia, but they can also be found in some stations in southern and eastern Spain, Luxemburg, central Italy northern Ukraine and southern England, and in the Dnieper, Dniester and south of the Danube River Basins.

Figure 19. Streamflow Variation Index for the period 1991–2019 (left) and spatial distribution of Normalised Variation Index in 2022 with respect to 2021 (right).



The spatial distribution of the Normalised Variation Index for annual averages between 2022 and 2021, Figure 19 (right), shows that the variations were predominantly positive (58%), with the negative variations mainly located in Central Europe. However, most of the variations (both positive and negative) were classified as mild or moderate: mildly and moderately positive variations entailed a total of 42%, while mildly and moderately negative variations entailed a total of 34%. The stations with the lowest annual mean discharge for 2022 compared to 2021 are mostly located in basins in southern Spain, France, Belgium and Switzerland. This situation also occurs in some stations in eastern Spain, Norway, Austria, northern Italy, Romania and Serbia. Conversely, the stations that registered the highest increases of discharge in 2022 compared to 2021 are located in Iceland, Ireland, Scotland, central England, Norway, Sweden, the Guadalquivir, Ebro, Minho Jucar and Turia River Basins, and small river basins in eastern Spain, both sides of the Pyrenees and the Rhône River Basin in France, eastern Poland, the Daugava River Basin in Latvia, Estonia, the south-western and central Danube River Basin and other Slovenian river basins, as well as the Vistula and Dnieper River Basins. There are also a few stations with high increases in southern England, France, Ukraine and the south-eastern Danube River Basin.

In summary, Figure 19 (left and right) shows that, comparing 2022 and 2021, the water volume in the rivers was smaller in Western and Central Continental Europe, while it was larger in the rest of Europe; conversely, the water volume in the rivers in 2022 was considerably smaller compared to the historical period 1991–2019.

4.3.2 Minimum and maximum value analysis

The analysis of minimum and maximum values enables the comparison of the extreme discharge values from 2022 with the extreme discharge values from 2021 and from 1991–2019. The percentiles of the daily values are calculated according to the time series of daily values for 2021 and the period 1991–2019, respectively. These percentiles are used to indicate how close the minimum and maximum river flows of 2022 are to the minimum and maximum of the reference periods. The minimum and maximum values of 2022 are compared to the percentiles for the reference periods 2021 and 1991–2019. The percentiles intervals are shown in Table 4.

Table 4. Classification based on percentiles. **The percentile is 0 for values lower than the minimum and 1 for values greater than the maximum. Separated classes have been added for such extremes*

Classes	**Minimum**	**Maximum**
Below / Exceeded	*	*
Very Low / High	P < 1%	P > 99%
Low / High	1% ≤ P < 2.5%	99% ≥ P ≤ 97.5%
Medium	2.5% ≤ P < 5%	97.5% > P ≤ 95%
High / Low	5% ≤ P < 10%	95% > P ≤ 90%
Very High / Low	P > 10%	P < 90%

In 2022, the minimum mean daily discharge values are predominantly lower than those in the period 1991–2019 (Figure 20, left). 20% of the stations recorded a lower minimum value than in the reference period (or the river flow was zero). Most of these stations are located in the Danube, Rhine, Elbe and Oder River Basins, France, England and Norway. A number of these stations were also found in basins in Spain, Ireland, Ukraine, Sweden, Latvia and Italy. On the contrary, 15% of the stations had minimum discharge values considerably higher than the minima of the historical period. This mostly occurred in basins in Spain, Italy, Iceland, the Scandinavian Peninsula and the Dnieper, Danube and Rhine River Basins, but also in isolated stations in Ireland and England, as well as in the Rhône, Elbe and Daugava River Basins. The minimum values for the rest of the stations are equally distributed according to the different degrees of proximity to the minimum for the period 1991–2019, but are predominantly low.

Compared to the previous year, 72% of the stations recorded minimum mean daily discharge values that were lower than those in 2021 (or the river flow was zero), as shown in Figure 20 (right). These stations are located all across Europe, except for the Dnieper and southern Danube River Basins and the Scandinavian peninsula and the Baltic region. Conversely, 14% of the stations recorded minimum mean daily values in 2022 that were considerably higher than the minimum values in 2021 (classes high or very high in Table 4 and in Figure 22). This mainly occurred in stations located in basins in Iceland, Spain, Norway, Sweden, Finland and Ukraine, and in the southern Danube River Basin, as well as in isolated stations in Ireland, Scotland, England, France Italy, Switzerland and Germany. High minimum values were also found in the southern Rhine, Vistula and Rhône River Basins.

Figure 20. Minimum values in 2022 with respect to the period 1991–2019 (left) and to the year 2021 (right).

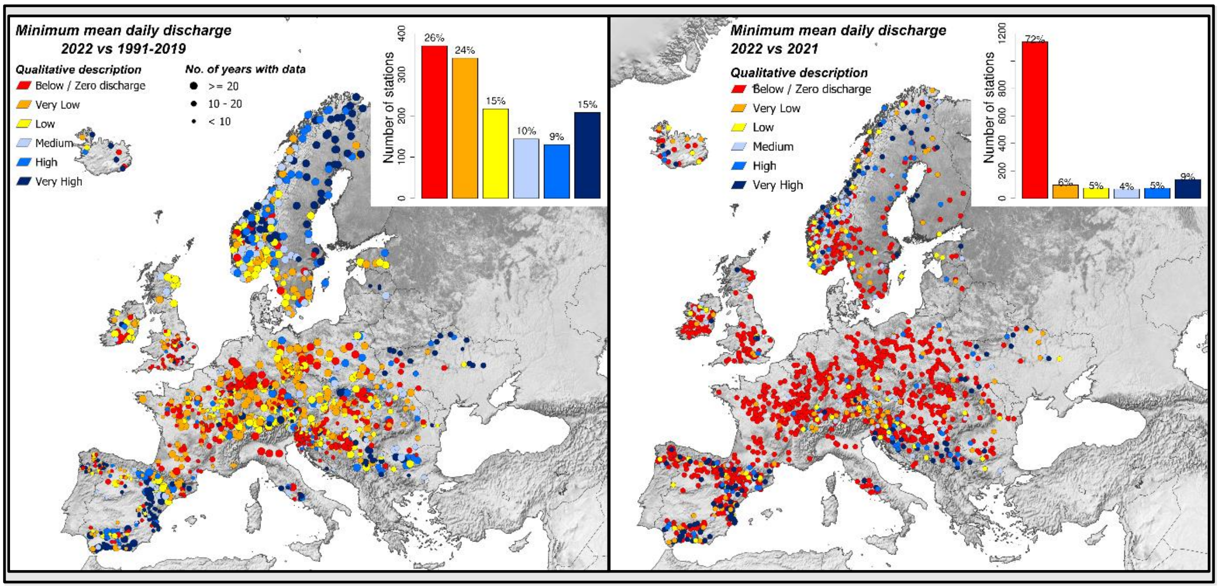
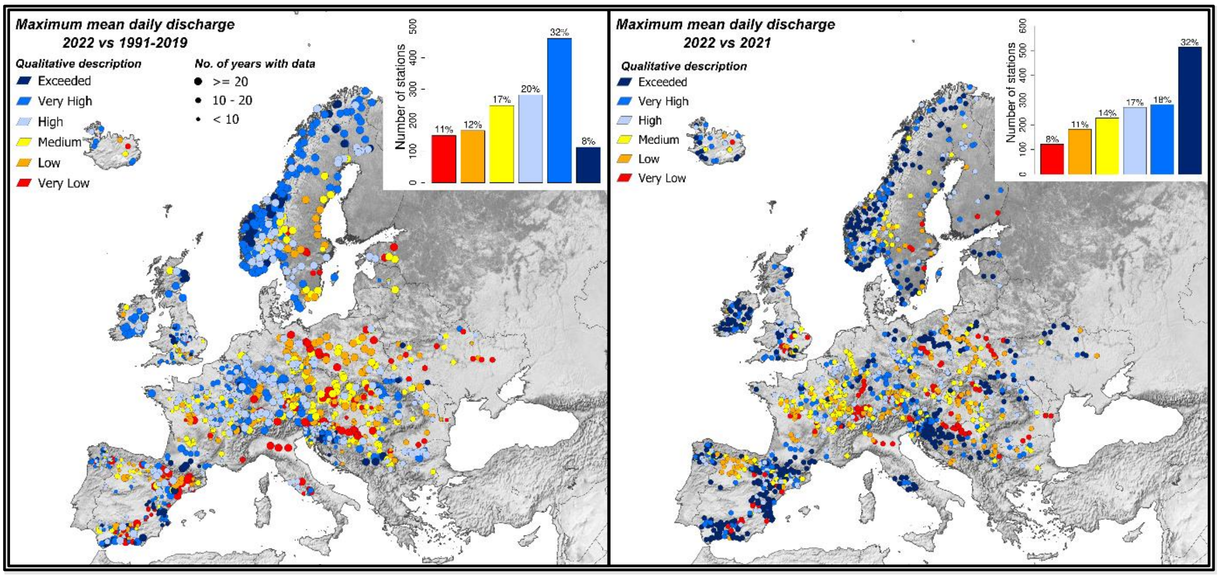


Figure 21 (left) shows that 32% of the stations across Europe recorded maximum values for 2022 that were just below their historic maxima from the period 1991–2019. Moreover, in 2022 8% of the stations exceeded the maximum mean daily value of the period 1991–2019. These exceedances occurred in stations in Spain, southern France, England, Scotland, Norway, the northern Scandinavian peninsula, Slovakia, western Ukraine and the Rhône, Rhine and southern Danube River Basins. On the other hand, around 23% of the stations recorded maximum mean daily values in 2022, considerably below the maximum historical values. These stations are mainly located in Spain, France, Italy, Sweden, the Loire, Elbe, Rhine, Oder, Vistula, Dnieper, Narva and Danube River Basins, and isolated stations in Iceland, Ireland, southern England, Norway, Slovenia, Croatia, and the Elbe and Dniester River Basins.

Figure 21 (right) shows a comparison of the maximum mean daily discharge for 2022 and 2021 and reveals that the maximum values were higher in 2022 for 32% of the stations across Europe, which are mainly located in Spain, Ireland, Iceland, southern England, Scotland, central Italy, Poland, the Scandinavian peninsula, the Baltic countries and the south-western Danube, Dniester, Dnieper, Rhine, Weser and Elbe River Basins, along with some isolated stations in France. However, 23% of the stations recorded maximum mean daily values considerably below the maximum value in 2021. These stations are mainly located in the basins in northern and eastern Spain, southern England, France, Scandinavia, and the Vistula, Dnieper, Oder, Elbe, Rhine, Po and Danube River Basins. Considerably lower extremes also occurred more locally for some stations in Iceland, Ireland, Slovenia and the Neretva River Basin.

Figure 21. Maximum values in 2022 with respect to the period 1991–2019 (left) and the year 2021 (right).



5 Analysis of threshold exceedance

5.1 Stations with available thresholds

The CEMS HDCC collects information about “local threshold levels” of discharge and/or water level values. These are supplied by EFAS partners and data providers. Four threshold levels are defined (TL1 through TL4, with TL1 being the lowest threshold), ranging from warning to critical values for flood events. However, some organisations use fewer thresholds; consequently, not all the stations have four threshold levels. The system orders the provided thresholds levels as follows:

- Only one level: TL1
- Two levels: TL1 and TL4
- Three levels: TL1, TL3, TL4
- Four levels: TL1, TL2, TL3, TL4

The analysis presented in this section compares the observed water level and discharge values during 2022 with the thresholds provided by EFAS partners and data providers.

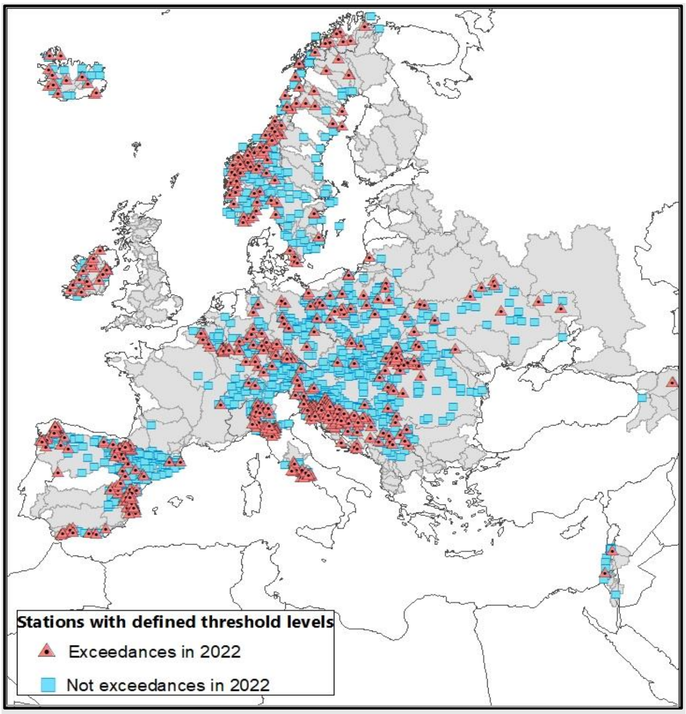
For this report, *exceedance* is considered when the values of water level or discharge in a station exceed its first threshold level at least once. In the same way, the word *event* is used here to highlight the exceedance of a threshold value during a period; therefore, this analysis does not report the number of flood events.

Of the 2,270 stations which were active from 2021-12-31, threshold levels are available for 1,552 stations (68%).

5.2 Stations that exceeded at least one of the thresholds

During 2022 a total of 512 of the 1,552 previously mentioned stations registered at least one exceedance. These stations were distributed in 372 rivers, 139 basins and 25 countries, as can be seen in Figure 22.

Figure 22. Map with the spatial distribution of the stations with defined threshold levels. Stations that did not exceed threshold levels in 2022 are shown in blue; stations that exceeded the threshold level at least once are shown in red.



The spatial distribution of the number of exceedances registered for each station can be seen in Figure 23. Of the stations, 40% registered one single exceedance event (204 of the 512 mentioned previously), 23% registered 2 exceedances, 20% between 3 and 4, 9% between 5 and 8, 7% between 9 and 40, and finally, only 1.54% registered more than 40 exceedances. The analysis of the number of exceedances must be completed by analysing the duration of these events. Adding up the time in which a station water level or discharge exceeded its first threshold level (accumulated time with exceedances for a station) reveals that 45% of the stations accumulated events of less than 1 day, 42% between 1 and 10 days, 11% between 10 and 100 days, and only 0.38% more than 100 days.

Figure 23. Map with the accumulated duration of the events and number of occurrences in stations with exceedances during 2022.

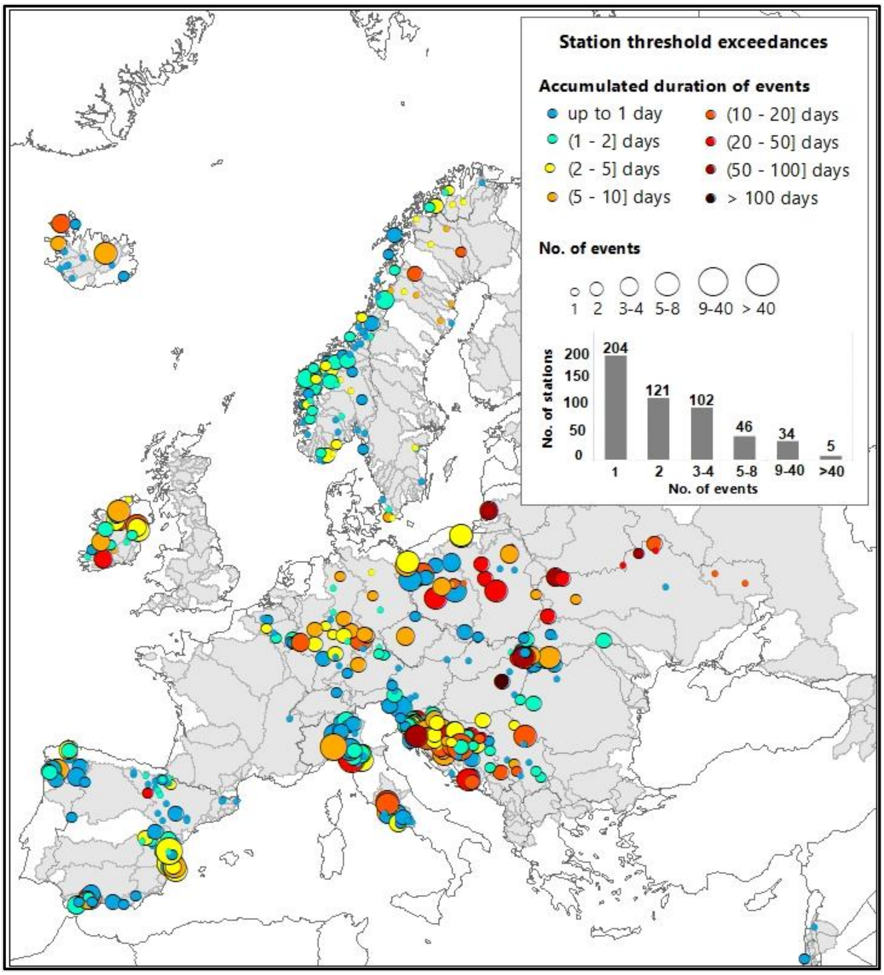


Table 5 shows those stations with more than 50 days of accumulated duration of events.

Table 5. Stations with more than 50 days of accumulated duration. The Kiskore felso station is highlighted (*) because the threshold exceedance is likely connected.

Station	**River**	**Basin**	**Country**	**No. of events**	**Accumulated duration (days)**
Chop	Latorica	Danube	Ukraine	11	51.5
Jasenovac	Sava	Danube	Croatia	6	52.6
Rusné	Atmata	Neman	Lithuania	6	56.5
Snovs’k	Snov	Dnieper	Ukraine	2	71.5
Vel’ke Kapusany	Latorica	Danube	Slovakia	6	83
Lubeshiv	Stokhid	Dnieper	Ukraine	5	88
Martinovo Selo uzv.	Rjecina	Rjecina	Croatia	15	93.3
Kiskore Felso (*)	Tisza	Danube	Hungary	3 (*)	157 (*)

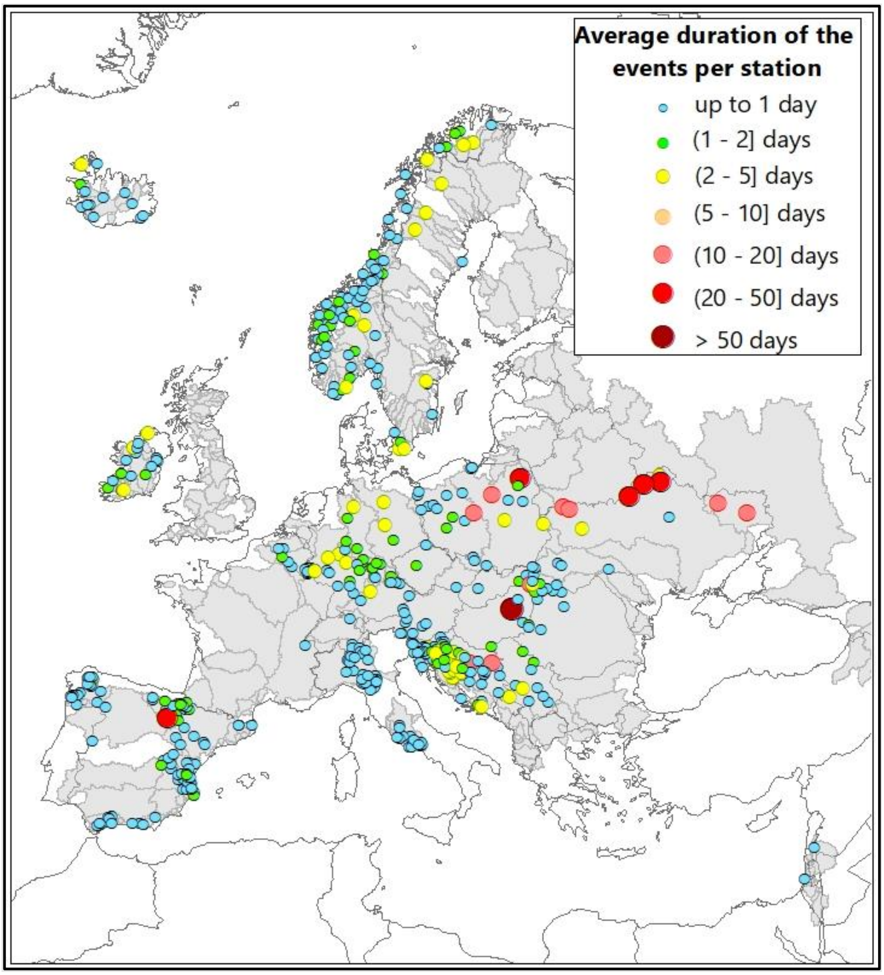
The highest value of accumulated time with exceedance occurred in Hungary at the Kiskore Felso station located on the Tisza River in the Danube Basin. The station recorded three events during 2022, with a total accumulated time of 157 days with exceedance. It should be clarified that this station is located near a reservoir. There is another station named Kiskore Also located just downstream from the dam, this latter station does not present the same behaviour as Kiskore Felso, therefore the long period with events at Kiskore Felso does not correspond to a real flood event. The exceedance at Kiskore Felso is likely related to reservoir management.

Using the average duration of the events for each station instead of the cumulative duration (the results can be seen in Figure 24), the most remarkable station is Nedanchychi, located in the Dnieper River Basin in Ukraine.

Ukraine continues in second and third place, with Snovs’k and Rzl’oty stations on the Snov and Desna Rivers, respectively.

The average event duration taking all the stations into account was 1.72 days. For 68% of the stations, the average duration of events was less than 1 day; for 15%, the duration was 1–2 days; for 10%, 2–5 days; for 5%, 5–10 days; and the remaining stations registered events with an average duration of more than 10 days (only 2%).

Figure 24. Map with the average event duration per station. The only station with an average event duration of more than 50 days is Kiskore Felso, a station connected to reservoir management operations.



Information about the stations with the highest average event duration can be seen in Table 6.

Table 6. Stations with an average event duration exceeding 20 days. The Kiskore Felso station is highlighted by (*) because the threshold exceedance is likely connected to reservoir management operations.

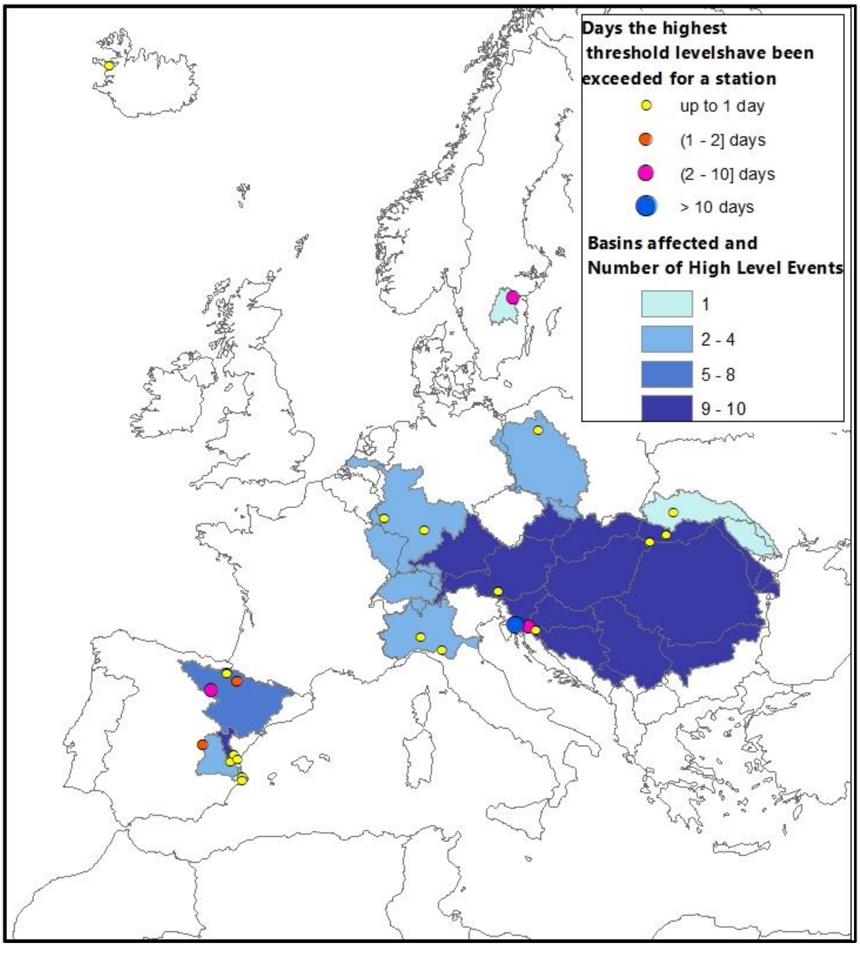
Station	**River**	**Basin**	**Country**	**No. of events**	**Accumulated duration (days)**	**Average duration (days)**
Río Cidacos	Cidacos	Ebro	Spain	2	42.8	21.4
Osowiec	Biebrza	Vistula	Poland	1	28.5	28.5
Rozl'oty	Desna	Dnieper	Ukraine	1	33.5	33.5
Snovs'k	Snov	Dnieper	Ukraine	2	71.5	35.7
Nedanchychi	**Dnieper**	**Dnieper**	**Ukraine**	**1**	**39.5**	**39.5**
Kiskore felso(*)	**Tisza**	**Danube**	**Hungary**	**3 (*)**	**157 (*)**	**52.3 (*)**

Finally, high-level events have been analysed. Specifically, high events are defined as follows:

1. stations with more than one threshold level and surpass the highest level defined, or
2. stations that have only one threshold level, but the discharge or water level is exceeded by at least 50%.

There were 40 high-level events in 2022, distributed across 24 stations that exceeded points 1 or 2. Figure 25 shows the spatial distribution and duration of these events, as well as the basins where they took place.

Figure 25. Stations with high-level events and basins affected.



The station with the highest event (10.3 days over its fourth threshold) is Martinovo Selo uzv, Croatia, on the Rjecina River. The high level event began on 2022-03-31 14:00:00 and ended on 2022-04-10 22:00:00.

The next noteworthy station is Río Cidacos in Yanguas, in the Ebro River Basin, Spain, with 2 high events: one lasting almost 3 days from 2022-04-27 to 2022-04-30 and another lasting 2 days between 2022-03-21 and 2022-03-23. There is only one threshold established for this station, therefore the exceedance corresponds to a value that is 50% higher than the threshold during these dates. This station appeared in the table of stations with exceedances of more than 20 days on average because the station surpasses its first threshold twice: from 2022-03-14 to 2022-04-10 and from 2022-04-22 to 2022-05-09 (16 and 26 days, respectively).

In third place is the station Göstad, in the Motala Strom River Basin, Sweden, with 2.75 days over TL4 from 2022-04-22 to 2022-04-25.

6 New developments to the HDCC database in 2022

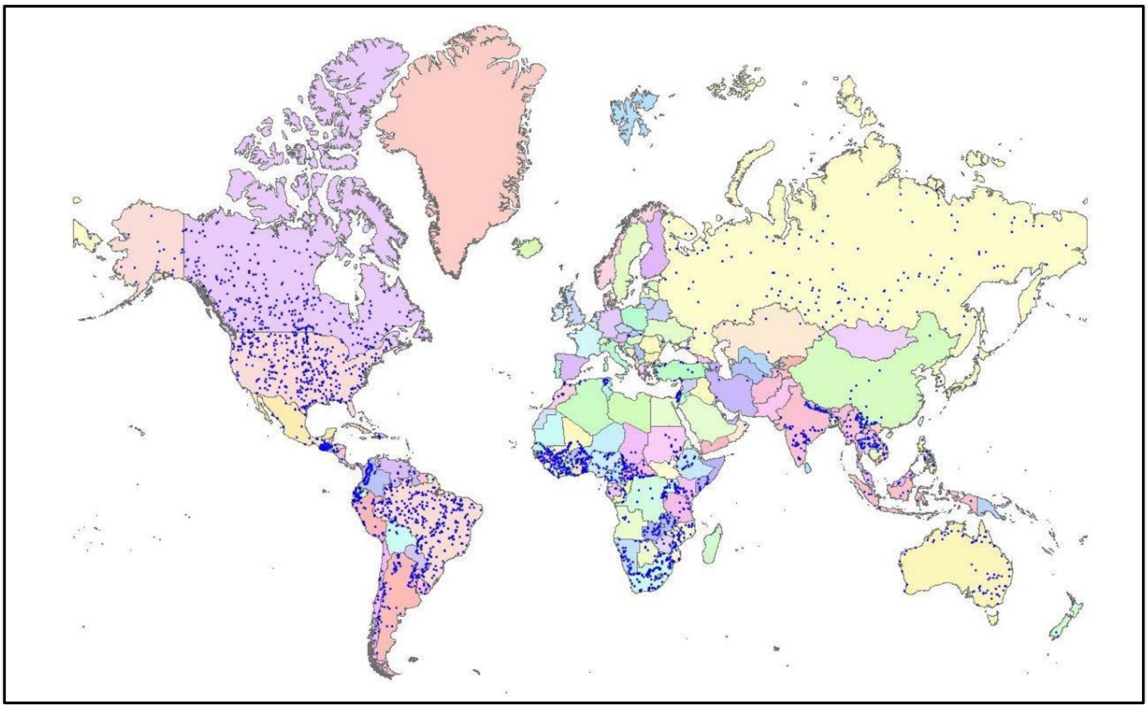
6.1 Extension of discharge collection to the global domain

Since 2011 the Joint Research Centre (JRC, Ispra) has been responsible for the collection of hydrological station data for the global domain. JRC collected historical daily discharge data with the purpose of calibrating and evaluating the CEMS Global Flood Awareness System. Until 2021 the system had a total of 2,906 stations providing discharge values since 1900 from 49 different providers.

Since autumn 2021, the CEMS HDCC has been responsible for the collection, quality control, harmonisation and internal distribution of hydrological observations for the global domain. Specifically, the hydrological observations for the global domain are limited to historical observations of daily discharge values.

In 2022 the global data collection database was transferred to the CEMS HDCC database. All stations were transferred except for those included in the European extended domain, as those stations are already included in the EFAS database. As a result of these efforts, 46 data providers and 2,588 stations with data spanning from 1914 to 2020 were uploaded to the historical table of the database. These stations are distributed across 98 countries (see Figure 26).

Figure 26. Spatial distribution of that stations (blue dots) included in the global historical database.



The data providers currently include international, freely available archives (one prominent example is the Global Runoff Data Centre, https://www.bafg.de/GRDC/EN/Home/homepage_node.html, Burek et al. 2022), national hydrological data collection and meteorological institutions, and research projects (one prominent example is ADHI, Tramblay et al., 2021). Information about data collection protocols and minimum requirements to contribute to the global database are available here: Share your data with GloFAS - Copernicus Emergency Management Service - CEMS - ECMWF Confluence Wiki.

In 2022, validation tests were carried out on metadata in order to verify coordinates and river names. This preliminary validation also enabled the identification of stations influenced by backwater. All the stations with validated metadata were included into the operational table. The quality checks described in Section 3.1.1 were used to verify each value of each time series However, due to a lack of information, not all quality controls could be completed. The following controls were performed in the global table: range control (QL1), variation versus previous day (QL4), variation versus previous value (QL5), negative discharge (QL7), and repetitive values (QL8).

These comprehensive quality checks were instrumental to assure the accuracy and integrity of the offline data acquired throughout the data acquisition process.

A complete review and update of the global database has been planned for 2023. This review and update entail the following key activities:

- Revisit all the existing data providers to extend the temporal coverage of the time series until the end of 2022.
- Revisit all the existing data providers to include newly available stations with drainage areas larger than 50 km^2.
- Include new data providers.
- Homogenise the time series to UTC 0 standard time.
- Perform quality checks on the homogeneised time series (as described in Section 3.1.2).

6.2 Collection of reservoir data

One of the new aspects under development in the collection of hydrological data for the European extended domain is the integration of reservoir data in the database. These data will provide valuable information for the implementation and calibration of the hydrological model (Open Source Lisflood (ec-jrc.github.io)) at the core of EFAS riverine flood forecasts.

The variables releated to resrvoirs are:

- Water level in the reservoir
- Volume of water stored in the reservoir
- Water inflow to the reservoir
- Water outflow from the reservoir to the downstream watercourse.

Other aspects, such as the purpose of the reservoir (e.g., energy production, irrigation) and other possible outflows are also taken into consideration.

In 2022, some providers were contacted to exchange information to discuss optimal ways for reservoirs data collection and management. The CEMS HDCC test database has therefore been adapted to import reservoir information. Following the conclusion of the tests, the operational system will be activated.

The routine reservoirs data collection from all EFAS data providers will start in 2023.

7 Conclusions

During 2022, the CEMS HDCC welcomed two new data providers for the European extended domain (EFAS): the*Confederacion Hidrografica del Segura*, Spain, and the *State Hydrometeorological Service* of Moldova. Moreover, 97 new stations from existing data providers were included in the EFAS database.

In 2022, the CEMS HDCC inherited from JRC the collection of daily historical hydrological stations on a global scale. The system had over 2,500 stations from 46 providers (outside of the European extended domain). Following an accurate quality check, the information was incorporated into the HDCC database.

The HDCC protocols and algorithms to quality check, homogenise and aggregate the data are constantly updated and improved. In 2022, the major improvement consisted in the refinement and finalisation of the two-step quality control protocol. Specifically, the first step leads to the definition of the L1 quality flag, while the second step leads to the definition of the QSF quality flag. Every value that enters the system goes through a series of quality controls, before being associated to a first quality control, L1.When data are aggregated the system goes through the L1 of each value and performs a final summary quality check (QSF). QSF provides the overall assessment of the quality of each operational value.

Comparing gaps in data collection from 2021 and 2022, the continuity of data transmission is seen to have improved. 97% of the data received fulfilled the conditions shared by providers. 88% of gaps lasted less than one hour. The interpolation process can be done only for gaps with a time interval of up to 3 days, while longer gaps can only be filled by data transmitted by the relevant data provider.

According to the data collected, the hydrological conditions of the stations of the European extended domain in 2022 has some particularities of note:

- Compared to the historical period 1991–2019, the water volume in the rivers in 2022 was considerably smaller.
- In 2022, 19% of the studied stations presented normalised mean discharge values below 100mm/year (in 2021 this ratio was 12%).
- Compared to 2021, almost 60% of the stations in Europe had a higher mean discharge in 2022, while 40%, (mainly located in Central Europe) had a lower mean discharge.
- 72% of the stations recorded minimum mean daily discharge values that were lower than in 2021.On the other hand, 9% of the stations recorded minimum mean daily values in 2022 that were considerably higher than the minimum values in 2021. The minimum mean values in 2022 were predominantly lower than the ones in the period 1991–2019. 20% of the stations recorded a lower minimum value than in the reference period.
- The maximum mean daily discharge values were higher in 2022 than in 2021 for 32% of the stations across Europe, which are mainly located in Spain, Ireland, Iceland, southern England, Scotland, central Italy, Poland, the Scandinavian peninsula, the Baltic countries and the south-western Danube, Dniester, Dnieper, Rhine, Weser and Elbe River Basins. However, 23% of the stations recorded maximum mean daily values that were considerably lower than the maximum value in 2021.
- 32% of the stations across Europe recorded maximum values for 2022 that were just below their historic maxima from the period 1991–2019. Moreover, in 2022 8% of the stations exceeded the maximum mean daily value of the period 1991–2019.

In 2022, 512 out of 1,552 stations registered at least one threshold exceedance event. 40% of these stations had one single exceedance. Conversely, less than 2% registered more than 40 exceedances. The average duration of the exceedance events considering all the stations was 1.72 days. 68% of the stations had events lasting less than 1 day on average, 15% between 1–2 days, 10% between 2–5 days, 5% between 5–10 days and the remaining stations registered events of an average duration of more than 10 days (only 2%). Regarding high-level events, 24 stations exceeded either their maximum threshold or 50% of their first threshold.

The CEMS HDCC constantly strives to expand the hydrological database and improve the quality and quantity control protocols. Planned activities for the year 2023 include further expanding the collection of historical and near real-time discharge and water level data for the extended European domain, implementing routines for the collection of reservoir data for the extended European domain, and expanding the collection of historical discharge data for the global domain. Information about minimum requirements and modalities of data collection are available in two dedicated pages: Share your data with EFAS | Copernicus EMS - European Flood Awareness System for the European extended domain; Share your data with GloFAS - Copernicus Emergency Management Service - CEMS - ECMWF Confluence Wiki for the global domain.

References

García Padilla, M., Santiago Nieto, F., Lara-Fernández, A., Montoya-Manzano, G., Márquez Arroyo, M., Carpintero-Salvo, I., Ruiz Montoro, L., Blanco De Córdova, D., García Sánchez, R., Disperati, J., Thiemig, V. and Salamon, P., The CEMS Hydrological Data Collection Centre - Annual report 2021, European Commission, 2022, JRC129075 https://www.efas.eu/en/report/annual-report-2021-cems-hydrological-data-collection-centre

Burek, P. and Smilovic, M., ‘ The use of GRDC gauging stations for calibrating large-scale hydrological models’, *Earth Syst. Sci. Data*, 2022, preprint, doi: https://doi.org/10.5194/essd-2022-231, in review, 2022.

Nalbantis, I.T., ‘Assessment of Hydrological Drought Revisited’, *Water Resources Management,*Vol. 23, 2009, pp 881-897, doi: https://doi.org/10.1007/s11269-008-9305-1

Tramblay, Y. et al., ‘ADHI: the African Database of Hydrometric Indices (1950–2018)’, *Earth Syst. Sci. Data*, Vol 13, 2021, pp. 1547–1560, doi: https://doi.org/10.5194/essd-13-1547-2021

List of abbreviations and definitions

API	Application Programming Interface
ADHI	African Database of Hydrometric Indices
CEMS	Copernicus Emergency Management Service
COMP	EFAS Forecast Computational Centre
DP	Data Provider
EFAS	European Flood Awareness System
GloFAS	Global Flood Awareness System
FTP	File Transfer Protocol
HDCC	Hydrological Data Collection Centre
hDMS	Hydrological Data Management System. Database fo HDCC
HTTP	Hypertext transfer protocol
JRC	Joint Reseach Centre of the European Commission
L1	Summary Flag
NDVI	Normalised Difference Vegetation Index
NDWI	Normalised Difference Water Index
NRT	Near Real-Time hydrological data
NVI	Normalised Variation Index
QL1	Range Test
QL2	Rating Curve Test
QL4	Daily Variation Test
QL5	Variation Test with respect to Previous Value
QL7	Negative Discharge Check
QL8	Repetitive Values Check
QL10	Manual Check
QSF	Quality Summary Flag
SDI	Streamflow Drought Index
SVI	Streamflow Variation Index

Aggregated data	Set of values on a determinated period.
Daily Variation Test (QL4)	Quality check that verifies whether the relative variation of a measurement with respect to the measurement made 24 hours before is lower than a specific threshold (see par. 3.1.1.).
Gap	The absence of a number of values in a time period.
Manual Check (QL10)	Quality check that overrides others. It is applied by the technicians (see par. 3.1.1.).
Negative Discharge Check (QL7)	Quality check that verifies if a discharge value lower than zero is possible (see par. 3.1.1.).
Normalised Mean Discharge	Annual discharge divided by the upstream area of the gauging station.
Operational data value	The average of the values on each aggregated data set.

Outlier	Value beyond the minimun or maximun thresholds defined for each station and variable.
Variation Testwith respect to Previous Value (QL5)	Quality check that verifies whether the relative variation of a measurement with respect to the previous measurement is lower than a specific threshold (see par. 3.1.1.).
Quality Summary Flag (QSF)	The result of the check process for aggregated data. (see par. 3.1.2).
Range Test (QL1).	Quality check that verifies if a measurement is included between a maximun and a minimum value, which are specific for each station and variable (see par. 3.1.1.).
Rating Curve	Mathematical relationship between the flow through a gauging station and the water level at the gauging station.
Rating Curve Test (QL2)	Quality check that verifies if the discharge value follow the characteristic rating curve of the station (see par.3.1.1.).
Raw data	The data as it is provided.
Repetitive Values Check (QL8)	Quality check that identifies periods when values are repeated (see par. 3.1.1.).
Summary Flag (L1)	The result of the check process for raw data (see par.3.1.1.).

List of figures

List of tables

Annexes

Annex 1. Data providers for the European extended domain: number of stations (water level, discharge)

Table A1. Number of stations (near real time or historical) that provide water level (WL), discharge (Q), water level and discharge (WL & Q) for each provider of the European extended domain.

Provider name	WL	Q	WL & Q
Institutul National de Hidrologie Si Gospodarire A Apelor - RO	30	35	30
Norwegian Water Resources and Energy Directorate, Hydrology Department - NO	2	152	1
Federal Office for the Environment - CH	41	32	32
Bundesanstalt fuer Gewaesserkunde - DE - BfG	86	62	62
Slovenian Environment Agency - SI	119	119	119
Saxon State Office for Environment, Agriculture and Geology - DE - SN	37	37	37
Rijkswaterstaat Institute for Inland Water Management and Waste Water Treatment - NL	3	3	3
Hessian Agency for Nature Conservation, Environment and Geology - DE - HE	6	6	6
Landesamt für Umwelt Rheinland-Pfalz - DE - LRP	9	9	9
Finnish Environment Institute	14	36	14
Confederación Hidrográfica del Ebro - ES	140	138	138
Federal Ministry of Agriculture, Forestry, Regions and Water Management - AT	67	59	55
Swedish Meteorological and Hydrological Institute, core services department - SE	0	71	0
Landesamt für Umwelt Brandenburg - DE - BB	3	3	3
ARPA Emilia Romagna - IT	77	60	59
Hungarian Hydrological Forecasting Service (OVSZ), General Directorate of Water Management (OVF) - HU	61	54	54
Slovak Hydrometeorological Institute - SK	54	54	54
Czech Hydro-Meteorological Institute - CZ	15	28	15
Republic Hydrometeorological Service of Serbia - RS	54	54	54
Ministère de l'Ecologie et du Développement Durable Service Central d'Hydrométéorologie et d'Appui à la Prévision des Inondations - FR	1	174	1
Hydrological Information Centre (HIC) - BE	8	10	7
Institute of Meteorology and Water Management - PL	170	149	149
Bayerisches Landesamt für Umwelt - DE - BV	69	59	59
National Institute of Meteorology and Hydrology - BG	12	12	12
Latvian Enviroment, Geology and Meteorology Centre - LV	14	6	6

Estonian Environmental Agency - EE	17	14	14
Office of Public Works - IE	49	49	49
Government of Andalusia - Regional Ministry of Agriculture, Fisheries and Environment - ES	110	69	69
Croatian Meteorological and Hydrological Service (HR) - HR	67	60	60
State Hydrometeorological Service of Moldova - MD	5	5	5
Confederación Hidrográfica del Miño-Sil - ES	46	46	46
Catalan Water Agency - ES	14	48	14
Flood Forecasting Centre - UK	109	110	109
Confederación Hidrográfica del Duero - ES	35	35	35
State Emergency Service of Ukraine - Ukrainian Hydrometeorological Center - UA	73	53	53
Service Public de Wallonie - BE	0	16	0
Hydrometeorological Institute of Kosovo - XK	17	2	1
Israel Hydrological Service - Water Authority - IL	11	7	7
Confederación Hidrográfica del Guadiana - ES	65	64	64
Confederación Hidrográfica del Júcar - ES	58	63	58
Federal Hydrometeorological Institute - BA	22	19	19
Lithuanian Hydrometeorological Service (LHMT) - LT	4	0	0
ARPA Lombardia - IT	55	40	40
LEPL National Environmental Agency - Ministry of Environmental Protection and Agriculture of Georgia - GE	5	0	0
Icelandic Meteorological Office - IS	0	27	0
Confederación Hidrográfica del Segura - CHS - ES	80	79	78
Water management Agency of Luxembourg (AGE) - LU	23	23	23
Agenzia Regionale di Protezione Civile - Protezione Civile Lazio - IT	75	39	35
Republic HydroMeteorological Service of the Republic of Srpska - BA	8	8	8

GETTING IN TOUCH WITH THE EU

In person

All over the European Union there are hundreds of Europe Direct centres. You can find the address of the centre nearest you online (european-union.europa.eu/contact-eu/meet-us_en).

On the phone or in writing

Europe Direct is a service that answers your questions about the European Union. You can contact this service:

— by freephone: 00 800 6 7 8 9 10 11 (certain operators may charge for these calls),

— at the following standard number: +32 22999696,

— via the following form: european-union.europa.eu/contact-eu/write-us_en.

FINDING INFORMATION ABOUT THE EU

Online

Information about the European Union in all the official languages of the EU is available on the Europa website (european-union.europa.eu).

EU publications

You can view or order EU publications at op.europa.eu/en/publications. Multiple copies of free publications can be obtained by contacting Europe Direct or your local documentation centre (european-union.europa.eu/contact-eu/meet-us_en).

EU law and related documents

For access to legal information from the EU, including all EU law since 1951 in all the official language versions, go to EUR-Lex (eur-lex.europa.eu).

Open data from the EU

The portal data.europa.eu provides access to open datasets from the EU institutions, bodies and agencies. These can be downloaded and reused for free, for both commercial and non-commercial purposes. The portal also provides access to a wealth of datasets from European countries.

The portal data.europa.eu provides access to open datasets from the EU institutions, bodies and agencies. These can be downloaded and reused for free, for both commercial and non-commercial purposes. The portal also provides access to a wealth of datasets from European countries.

Science for policy

The Joint Research Centre (JRC) provides independent, evidence-based knowledge and science, supporting EU policies to positively impact society

EU Science Hub
joint-research-centre.ec.europa.eu

@EU_ScienceHub
EU Science Hub - Joint Research Centre
EU Science, Research and Innovation
EU Science Hub
@eu_science