European Commission

Early warning and monitoring

Floods

Copernicus Emergency Management Service

The CEMS Hydrological Data Collection Centre

Annual report 2021

Prepared by the CEMS HYDROLOGICAL DATA COLLECTION CENTRE and the JOINT RESEARCH CENTRE.

soologic®
TECHNOLOGICAL SOLUTIONS

Junta de Andalucía
Regional ministry of agriculture, livestock, fisheries and sustainable development

Joint Research Centre

EU Science Hub

https://ec.europa.eu/jrc

JRC129075

Ispra: European Commission, 2022

How to cite this report: 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.

Contents

1 Introduction ... 3

1.1 Background ... 3

1.2 Updates in the HDCC database in 2021 ... 4

1.3 Analysis of the data in the HDCC database ... 6

2. Hydrological conditions of EFAS gauging stations ... 8

2.1 Assessing stations and data for analysis ... 8

2.2 Hydrological conditions in 2021 ... 9

2.2.1 Comparative analysis ... 9

2.2.2. Variation of hydrological conditions ... 11

2.2.3. Minimum and maximum value analysis ... 11

3. Data quality of HDCC data collection ... 14

3.1 Rules applied for data quality ... 14

3.2 New data quality rules developed in 2021 ... 15

3.2.1 Monthly MIN/MAX (QL6) ... 15

3.2.2 Negative discharge (QL7) ... 15

3.2.3 Repetitive values (QL8) ... 15

3.2.4 Visual check (QL10) ... 16

3.3 Quality control statistics for 2021 ... 16

4. Gaps Analysis on the CEMS hydrological data base ... 18

Initial considerations ... 18

4.1 Gap analysis ... 18

4.1.1 Gap classification by duration ... 19

4.1.2 Gap classification by status ... 19

4.1.3 Other aspects to be considered ... 20

4.1.4 Gap typology and proposal for future data collection strategy ... 21

4.2 Outliers analysis ... 23

5. Analysis of Exceedance events ... 25

5.1 General description ... 25

5.2 Duration of Exceedances ... 26

5.2.1 Duration per station ... 26

5.3 High threshold level exceedances ... 29

6. Summary ... 32

6.1 Hydrological Conditions ... 32

6.2 Gaps and Outliers ... 32

6.3 Exceedances Events ... 33

List of figures ... 34

List of tables ... 35

References ... 36

Acknowledgements

The EFAS team, and particularly the 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 the questions and solving issues. Without their collaboration, the delivery of this report would not be possible.

Abstract

The CEMS Hydrological Data Collection Centre (HYDRO) is a fully operational and automated hydrological data collection service of the Copernicus Emergency Management Service. HYDRO collects real-time and historical hydrological observations from 48 data providers across Europe and supports in particular the European Flood Awareness System (EFAS) with the provision of quality-controlled and standardised hydrological observations used mainly for the calibration of the hydrological model as well as for the real-time monitoring of the hydrological conditions across Europe.

In this report the collected hydrological data for 2021 are analysed with respect to a) the general hydrological conditions across Europe, focusing on important deviations of average discharge; b) gaps and outliers, including a classification according to causes, duration, length, and distribution; and c) high flow events, looking at the duration, magnitude, number, and distribution of exceedances according to the threshold levels.

The analysis has shown that the water contribution in 2021 is higher than in 2020 but still it was slightly lower than it was in the historical period 1991-2019. The number of data gaps in the hydrological time series have decreased in 2021 with respect to the previous year (606,407 vs 525,936) even when the total number of received data has increased by 22%. With regards to extreme events, the analysis has shown that the total number of exceedance events increased in 2021 compared to the year before, 2020, having had many events but shorter than the events in 2020.

This steadily expanding hydrological data collection is contributing to an improved flood forecasting capacity (European Flood Awareness System) and drought indicators (European Drought Observatory), and has a large potential for other applications, still to explore.

1 Introduction

1.1 Background

This report contains an analysis of the hydrological data received by the CEMS Hydrological Data Collection Centre (HDCC) for the year 2021. The HDCC is contracted by the European Commission and operated by the Agencia de Medio Ambiente y Agua de Andalucía in collaboration with Soologic Technological Solutions S.L. The HDCC is responsible for the collection, quality control, harmonisation and internal distribution of hydrological observations to various components of the Copernicus Emergency Management Service (CEMS), mostly to the European Flood Awareness System (EFAS).

By the end of 2021, 49 data providers contributed with near real-time hydrological data at 2,386 stations to the CEMS Hydrological Data Collection, covering 34 countries and 51% of all the European water basins.

Figure 1: Spatial distribution of data providers to the CEMS.

In the following section, we first highlight the growth of the HDCC database in 2021, before introducing the hydrological analysis of data within the EMS HDCC in the next section, which will in turn occupy the rest of the document.

1.2 Updates in the HDCC database in 2021

During 2021, three additional hydrological data providers contributed with their hydrological data to the HDCC. Those are:

- Republic HydroMeteorological Service of the Republic of Srpska – BA: 8 stations
- Water management Agency of Luxembourg (AGE): 23 stations
- Agenzia Regionale di Protezione Civile (Protezione Civile Lazio): 88 stations

In addition to those new data providers and stations, a number of existing data providers (DP) increased the number of stations providing real-time hydrological data to the HDCC (Table 1) as a response to the EFAS data collection campaign for the next calibration of the hydrological model used by EFAS, which is OS LISFLOOD (https://github.com/ec-jrc/lisflood-code). As a result of this campaign the total number of stations in the HDCC database increased by 19 %.

Table 1: Additional stations per CEMS HDCC data provider.

Data provider	**Nr. of station added (increase in %)**
Bundesanstalt fuer Gewaesserkunde - DE	11 (+14,6 %)
Saxon State Agency for Environment and Geology - DE	1 (+2,7 %)
Hessian Agency for Nature Conservation, Environment and Geology - DE	2 (+50,0 %)
Landesamt für Umwelt Rheinland-Pfalz - DE	3 (+33,3 %)
Confederación Hidrográfica del Ebro - ES	78 (+125,0 %)
Slovak Hydrometeorological Institute - SK	29 (+116,0 %)
Czech Hydro-Meteorological Institute - CZ	2 (+7,9 %)
Republic Hydrometeorological Service of Serbia - RS	11 (+25,0 %)
Institute of Meteorology and Water Management Wroclaw Branch - PL	11 (+6,8 %)
Office of Public Works - IE	26 (+130,0 %)
Government of Andalusia - Regional Ministry of Agriculture, Fisheries and Environment - ES	1 (+0,9 %)
Meteorological and Hydrological Service - HR	28 (+71,0 %)
Catalan Water Agency - ES	15 (+93,7 %)
Confederación Hidrográfica del Duero - ES	12 (+52,0%)
Service Public de Wallonie - BE	10 (+166,0 %)
Hydrometeorological Institute of Kosovo - XK	15 (+75,0 %)
Confederación Hidrográfica del Guadiana - ES	65 (+100,0 %)
Confederación Hidrográfica del Segura - CHS	81 (+100,0 %)

In addition, some EFAS data providers uploaded new historic data sets during 2021. An overview is given in Table 2.

Table 2: Statistics summarizing all the changes to the HDCC database in 2021.

Country	Hydrological data provider	Years
Romania	Institutul National de Hidrologie Si Gospodarire A Apelor	2011-2019
Norway	Norwegian Water Resources and Energy Directorate, Hydrology Department	2016-2019
Switzerland	ch-bafu	2018
Germany	Bundesanstalt fuer Gewaesserkunde	2015-2019
Slovenia	Slovenian Environment Agency	2017-2019
Germany	Saxon State Agency for Environment and Geology	1998-2020
Germany	Hessian Agency for Nature Conservation, Environment and Geology	1984-2021
Germany	Landesamt für Umwelt Rheinland-Pfalz	1992-2019
Hungary	Hungarian Hydrological Forecasting Service, General Directorate of Water Management	2019
Czech Republic	Czech Hydro-Meteorological Institute	2016-2019
France	Ministère de l'Ecologie et du Développement Durable Service Central d'Hydrométéorologie et d'Appui à la Prévision des Inondations	1910-2021
Belgium	Hydrological Information Centre	2018
Poland	Institute of Meteorology and Water Management Wroclaw Branch	2019-2020
Estonia	Estonian Environmental Agency	1902-2020
Spain	Government of Andalusia - Regional Ministry of Agriculture, Fisheries and Environment	2015-2018
Croatia	Meteorological and Hydrological Service	2013-2019
United Kingdom	Flood Forecasting Centre	1991-2021
Ukraine	State Emergency Service of Ukraine - Ukrainian Hydrometeorological Center	2019-2020
Albania	Institute of GeoSciences Enery Water and Environment	2002-2008
Spain-Jucar	Confederación Hidrográfica del Júcar	2010-2020
Lithuania	Lithuanian Hydrometeorological Service	2010-2021
Italy	ARPA Lombardia	1997-2020
Iceland	Icelandic Meteorological Office	2005-2020
Luxembourg	Water management Agency of Luxembourg	2002-2020
Italy	Agenzia Regionale di Protezione Civile - Protezione Civile Lazio	1995-2019
Republic of Srpska	Republic HydroMeteorological Service of the Republic of Srpska	2003-2021

In table 3 it is possible to visualize the number of data providers, stations and values managed during 2021. The difference between "Data Providers" and "Active Data Providers", "Stations" and "Active Stations" is that the "Active", both Data Providers and Stations, are already providing data to EFAS, as opposed to those that are still in process of doing it.

Table 3: Number of data providers, stations and values managed during 2021.

	Before 2021	In 2021	Total	Increment
Data Providers	67	3	70	4%
Active Data Providers	48	1	49	2%
Nr. Of Stations Registered	3,533	416	3,949	12%
Nr. Of Active Stations	1,990	379	2,386	19%
Nr. Of Stations with defined threshold levels	1,480	281	1,761	19%
Nr. Of Near Real-Time Values	421,357,983	90,495,211	511,853,194	21%
Nr. Of Historic Values	125,946,080	77,615,375	203,561,455	62%

1.3 Analysis of the data in the HDCC database

The hydrological data received by the CEMS HDCC for the year 2021 is analysed in the following chapters, each focusing on different aspects:

- Chapter 2 presents the hydrological conditions of EFAS gauging stations.
- Chapter 3 describes how quality checks are done in discharge and water level data.
- Chapter 4 provides an analysis on the general hydrological conditions across Europe, focusing on important deviations of average discharge
- Chapter 5 provides an assessment of the HDCC Data Collection in terms of gaps and outliers, including a classification according to causes, duration, length and distribution.
- Chapter 6 prevides an evaluation on the threshold level exceedances, looking at the duration, magnitude, number and distribution of exceedances according to the threshold levels.

In 2021, CEMS HDCC has collected hydrological data from 2,386 stations across Europe. In this report only 1,963 stations will be analysed, as only stations that were actively delivering data throughout the entire year 2021 and that had a stable data provision to the HDCC since 01st February of 2021 were selected. Out of those 1,963 stations, 490 deliver exclusively discharge data, 390 only water level data and 1,083 stations provide discharge and water level data. Figure 2 shows the geographical distribution of those stations.

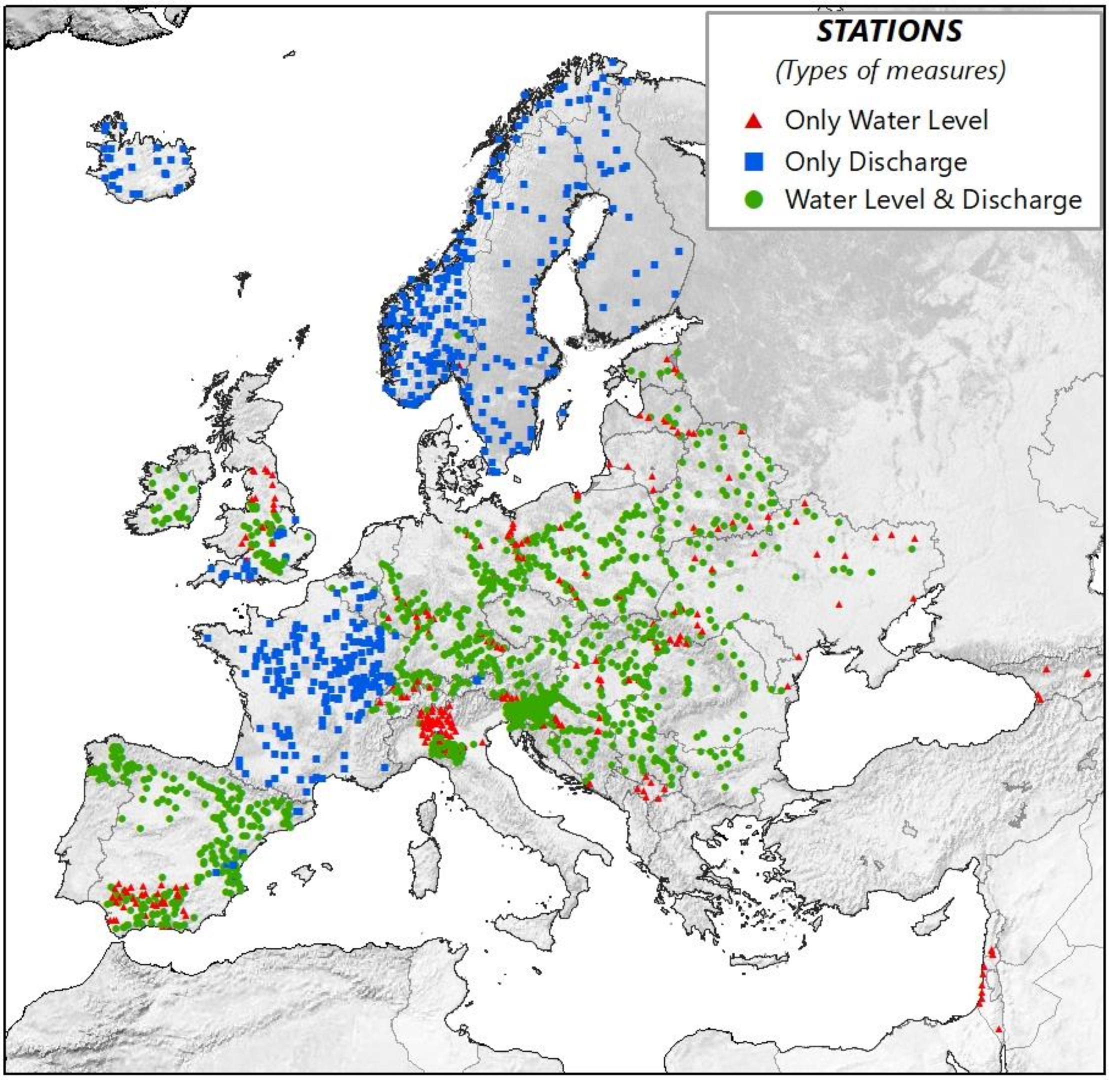


Figure 2: Spatial distribution of the 1963 selected stations and variables measured

2. Hydrological conditions of EFAS gauging stations

This chapter describes the hydrological conditions for the year 2021 across the entire EFAS domain, by comparing near real-time data of 2021 with near real-time data from 2020 and historical data (1991-2019) respectively. The reference period was from 2016 to 2019.

Although the CEMS Hydrological Data Collection Centre (HDCC) collects water level and discharge values, the analyses in this chapter is done based on discharge data only. This is because, unlike water level, discharge does not depend on the river's geometry and hence allows for a comparison of the hydrological behaviour between stations.

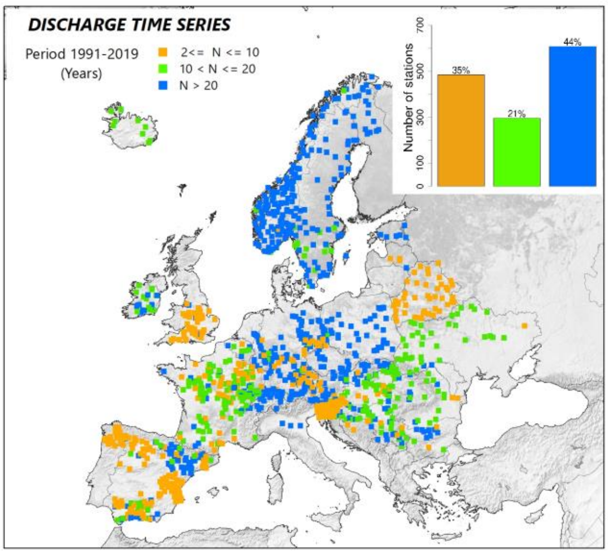
The mean daily values have been used to calculate all the statistics for the analyses: the annual mean minimum and maximum for 2021, as well as the percentiles of the year 2020 and the period 1991-2019 respectively. The average of the annual mean is an indicator of the annual water contribution at the gauging points, whereas the percentiles allow comparing the annual minima and maxima in 2021 to the reference periods in order to determine their variations.

We would like to point out that the analysis covered by this section is based only on discharge measures collected from gauging stations. As an increasing number of stations are strongly regulated upstream by hydraulic infrastructures, many of these stations show discharge values that are not according to their natural discharge regimes. For this reason, it's not uncommon to find discharge variations that are not caused by meteorological factors, and the potential impact of anthropic interference on natural river flow should be kept in mind while interpreting the results.

2.1 Assessing stations and data for analysis

In order to guarantee a good quality analysis, only stations with good temporal coverage have been selected for the analyses. For 2020 and 2021 only stations that were fully operational and active throughout the reference period, and received more than 75% of their expected annual discharge observations were selected. For the 1991-2019 period, only stations with at least two years of data were included. As a result, a total of 1512, 1211, and 1194 stations were chosen for 2021, 2020 and 1991-2019, respectively.

Figure 3 (left) shows the spatial distribution of the hydrological gauging stations selected for this analysis, including the length of their historical time series. More than 40% of the stations have more than 20 years of historical data. The longer the time series, the more representative are the derived statistical parameters. Henceforth, we expect the accuracy of the assessment to be higher in areas



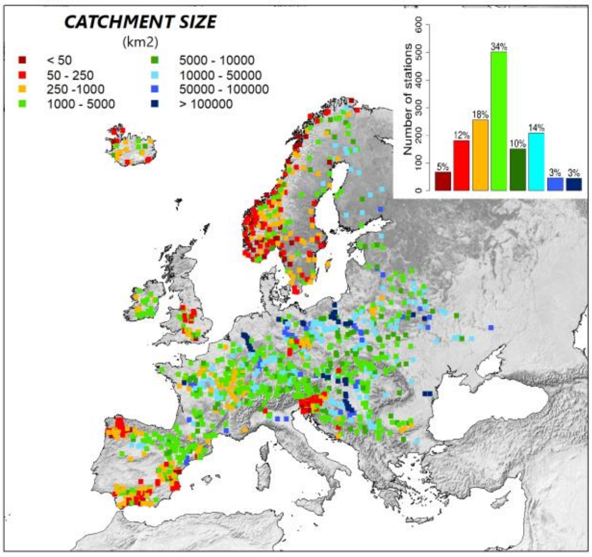


Figure 3: Spatial distribution of stations according to the length of their historical time series (left) and catchment size (right).

with long historical time series such as Norway, Sweden, Poland, Estonia, the Ebro River basin in Spain, and stations across the Rhine and Danube River basins.

Figure 3 (right) shows the upstream areas of all the selected stations. Many of the stations from the Scandinavian peninsula, Spain, Slovenia, England, Iceland and across the Elbe River basin have small catchment areas (< 250 km^2), whereas many of the stations from the Danube, Vistula, Ebro and Rhine River basins have large upstream areas (>1000 km^2). The distribution of catchment areas of the stations depends on three main factors: firstly, the catchments' hydrological features, secondly, it is a result of where hydrological services want to monitor the hydrological situation, lastly, it depends on which of the observations the hydrological services are willing to share. We have normalized the discharge values by the upstream area, this index allows comparisons between stations. Nevertheless, differences in catchment areas are still likely to influence the minimum and maximum values (smaller catchments typically have a larger difference between minimum and maximum specific discharge than larger catchments) and on annual variability (smaller catchments typically have larger annual variability). The units for this index are millimetres of water per year (mm/year), which is the same as litres per square meter and per year [l/(m^2·year)].

2.2 Hydrological conditions in 2021

Figure 4 shows the normalized mean discharge values for 2021. 12% of the studied stations present values below 100 mm/year. These are mostly present in Spain, Elbe, Oder, Vistula, Dnieper, Neman, Daugava and the Northern, Eastern and Central Danube River basins. These values usually belong to dry meteorological regimes and/or regulated or overexploited streams. The highest values (over 1000 mm/year) occur for stations in Norway, Iceland, Ireland, Northern and Eastern Spain, the upper Rhine and Danube basins and usually occur in relatively small catchments with high precipitation.

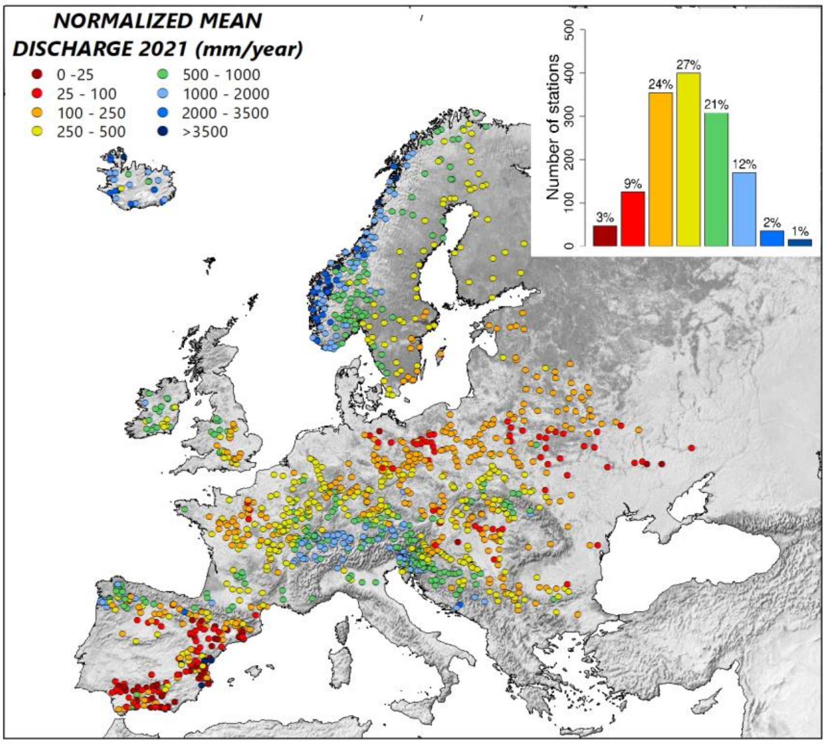


Figure 4: Spatial distribution of normalized mean discharge values in 2021

2.2.1 Comparative analysis

In this section, the hydrological situation of 2021 is compared to the previous year (2020) and to the historical reference period (1991-2019). This is to assess if and how the hydrological conditions of 2021 differ from the past. The comparison of the relative variation of the average values is done through two indices: the Streamflow Variation Index (SVI) and the Normalized Variation Index (NVI).

SVI is applied comparing 2021 and the period 1991-2019. It is adapted from the Streamflow Drought Index (SDI) (Nalbantis, 2009):

$$SVI_H = \frac{\acute{X}_{2021} - \acute{X}_H}{S_H}$$

$\acute{X}_{2021}$ and $\acute{X}_H$ are the mean discharges for 2021 and 1991-2019, respectively. S_H is the standard deviation of the annual mean discharge for the period 1991-2019. This index is a standardization of annual mean discharge in 2021 according to the annual mean and the standard deviation of the annual mean discharge in the period 1991-2019.

NVI is a normalized difference index widely used for comparing two measures taken in a particular location but in two different moments. NDWI and NDVI are cases of use of NVI index (Zhangyan, 2006) (Bo-cai, 1996)

The Normalized Variation index (NVI) is used when comparing the 2021 and 2020 mean discharges. The SVI is not applicable when the reference period covers only one year:

$$NVI_H = \frac{\acute{X}_{2021} - \acute{X}_{2020}}{\acute{X}_{2021} + \acute{X}_{2020}}$$

Where $\acute{X}_{2021}$ and $\acute{X}_{2020}$ are the mean discharges for 2021 and 2020 respectively.

Table 4 defines quality classes based on the distribution of the resulting SVI and NVI values.

Table 4: SVI and NVI classes. Positive/Negatives indices indicates 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
Midly negative	−1 ≤ SVI < -0,25	-0,25 ≤ NVI < -0,02
Moderately negative	−1.0 ≤ SVI < -2	−-0,5, ≤ NVI < -0,25
Extreme negative	SVI < -2	NVI < -0,5

On the other hand, the percentile of the minimum and maximum daily mean values of 2021 is calculated according to the time series of daily mean values from 2020 and the period 1991-2019 respectively. These percentiles are used to indicate how close the minimum and maximum river flows of 2021 are to the minimum and maximum for those periods. The extreme values of 2021 are then classified according to their percentile in the periods 2020 and 1991-2019. The percentiles intervals are shown in Table 5.

Table 5: Classification based on percentiles.

Classes	*Minimum*	*Maximum*
Below / Exceeded	*	*
Very Low / High	P < 1%	P > 99%
Low / High	1% ≤ P < 2,5%	97,5% ≥ P ≤ 99%
Medium	2,5% ≤ P < 5%	95% ≤ P < 97,5%
High / Low	5% ≥ P ≤ 10%	90% ≤ P < 95%
Very High / Low	P > 10%	P < 90%

** The percentile is 0 for values lower than the minimum and 1 for a value greater than the maximum. Separated classes have been added for such extremes.*

2.2.2. Variation of hydrological conditions

The spatial distribution of Normalized Variation Index for annual averages between 2021 and 2020, Figure 5 (left), shows clearly a dominance of low variations, both positive (44%) and negative (30%), in stations across Europe. Stations with the lowest annual mean discharge for 2021 compared to 2020 are mostly located in basins of Eastern Spain and Southern Norway. This situation also occurs in some stations in Sweden, England, Ireland, Southern France, Finland, Ukraine, Romania and Bosnia and Herzegovina. On the other hand, the stations that registered the highest increases of discharge in 2021 compared to 2020 are located in Elbe River in Germany, Jucar, Guadalquivir and Minho-Sil river basins in Spain, Rhône and Seine rivers in France, Vistula river basin in Western Poland, Dnieper river basin in Belarus and Ukraine, Western Danube river basin. There are also a few stations with high increases in Southern England, Iceland, Sweden and Norway. In summary, most of the stations Europe had a similar annual mean discharge in 2021 to what they had in 2020 although the balance is slightly more positive in 2021.

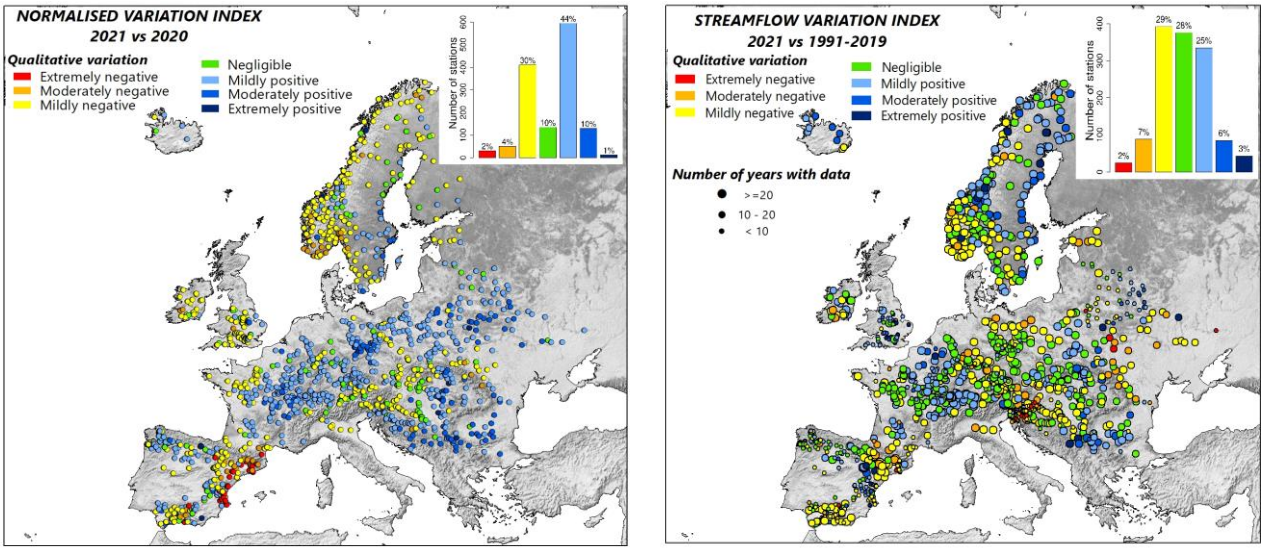


Figure 5: Spatial distribution of Normalized Variation Index in 2021 with respect to 2020 (left) and Streamflow Variation Index the period 1991-2019 (right).

When comparing 2021 with the 1991-2019 period, the analysis shows a balance with slight negative variations (Figure 5 right). However, 28% of the stations have negligible variations. They are distributed uniformly across Europe except in Southern Spain, Eastern Europe and Northern Scandinavian Peninsula. 9% of stations suffer moderate or extreme negative variations. Most of them are located in upper Danube river basin and basins of Ukraine, Spain and Norway. A number of stations in the Rhine river basin, basins in South-Eastern France, Ireland, and the rivers Elbe, Oder and Narve show a moderate drought as well. On the other hand, 11% of the stations present a severe or moderate surplus of mean discharge in 2021 compared with the period 1991-2019. They are mostly located in basins in England, Spain, Norway, Sweden, Iceland, Northern Meuse and Scheldt river basins and across Danube river basin, but they can also be found in stations of the Dnieper river basin in Belarus and isolated stations in France.

As summary Figure 5 (left and right) shows, that the water volume in the rivers was slightly larger in 2021 than in 2020; conversely, the water volume in the rivers in 2021 was slightly smaller compared to the historical period 1991-2019.

2.2.3. Minimum and maximum value analysis

In 2021, 42% of the stations recorded minimum mean daily discharge values that were lower than the ones in 2020 (or the river flow was zero), as it's shown in Figure 6 (left). We can see that these stations are located all across Europe but the higher concentration is in the Southern Rhine, Dnieper and

Danube river basin, British Isles, Spain, and Scandinavian Peninsula. On the other hand, around 26% of the stations recorded minimum mean daily values in 2021 that were considerably higher than the minimum values in 2020. This mainly occurred in stations located basins in Spain, France, Germany (Rhine and Elbe) and Eastern Europe (Oder, Vistula and Dnieper) and Sweden. High minimum values were also found in basins of Southern England, Belarus, Eastern Danube river basin, Iceland, Norway and Latvia.

The minimum mean values in 2021 are predominantly higher than the ones in period 1991-2019. We found that only 10% of the stations recorded a lower minimum value than in the reference period (or the river flow was zero) (Figure 6 right). Most of these stations are located in the Danube and Dieper basins, and Norway. We also found a number of these stations in basins of Spain, Ireland, England, Italy, France and Estonia. Contrastingly, 28% of the stations had discharge minimum values considerably higher than the minima in the historical period. This mostly occurred in basins in France, Spain, Norway, Sweden and Meuse, Rhine, Elbe and Oder river basins but also occurs in isolated stations of Ireland, Iceland, Italy, Finland as well as in the Daugava and Dnieper river basins. The minimum values of the rest of the stations are equally distributed according to the different degrees of closeness to the minimum for the period 1991-2019.

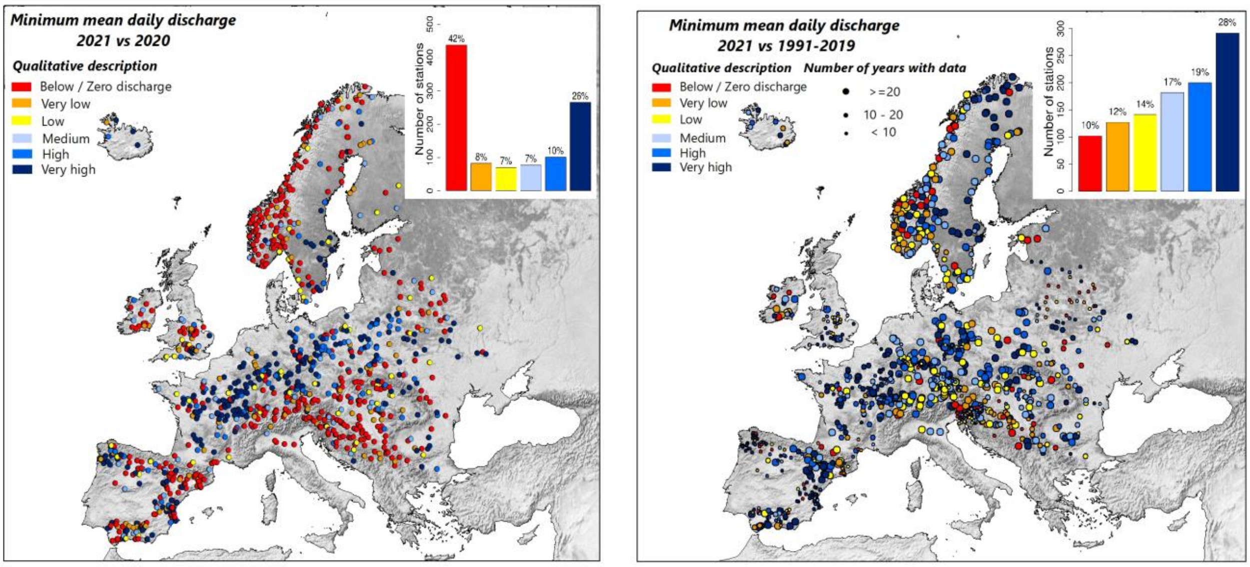


Figure 6: Spatial distribution of stations and the minimum values in 2021 with respect to 2020 (left) and the period 1991-2019 (right)

Figure 7 (left) shows a comparison of the maximum mean daily discharge for 2021 and 2020 and show that the maximum values were higher in 2021 for 53% of the stations across Europe, which are almost equally distributed. However, around 10% of the stations recorded maximum mean daily values considerably below the maximum value in 2020. These stations are mainly located in basins of Eastern Spain, Norway, Estonia, Northern Sweden and Finland, Oder, Seine, Loire and Danube river basins. Considerably lower extremes also occurred more locally for some stations in England, Ireland, Dniester and Po river basins.

Figure 7 (right) shows that 52% of the stations across Europe recorded maximum values for 2021 that were just below their historic maxima from the period 1991-2019. Moreover, 13% of the stations exceeded in 2021 the maximum mean daily value of the period 1991-2019. These exceedances took place in stations of Spain, France, England Southern Scandinavian Peninsula, Iceland and Meuse, Rhine and Danube river basins. There were also exceedances at stations located in stations of Ireland and Dnieper and Vistula river basins. On the other hand, around 12% of the station recorded maximum mean daily values in 2021 considerably below the maximum historical values. These stations are mainly located in Spain, Italy, the Elbe, Oder, Vistula, Dnieper, Neman, Narva, Danube river basins and isolated stations in the Rhine, Daugava and Po river basins, Sweden, England and Norway.

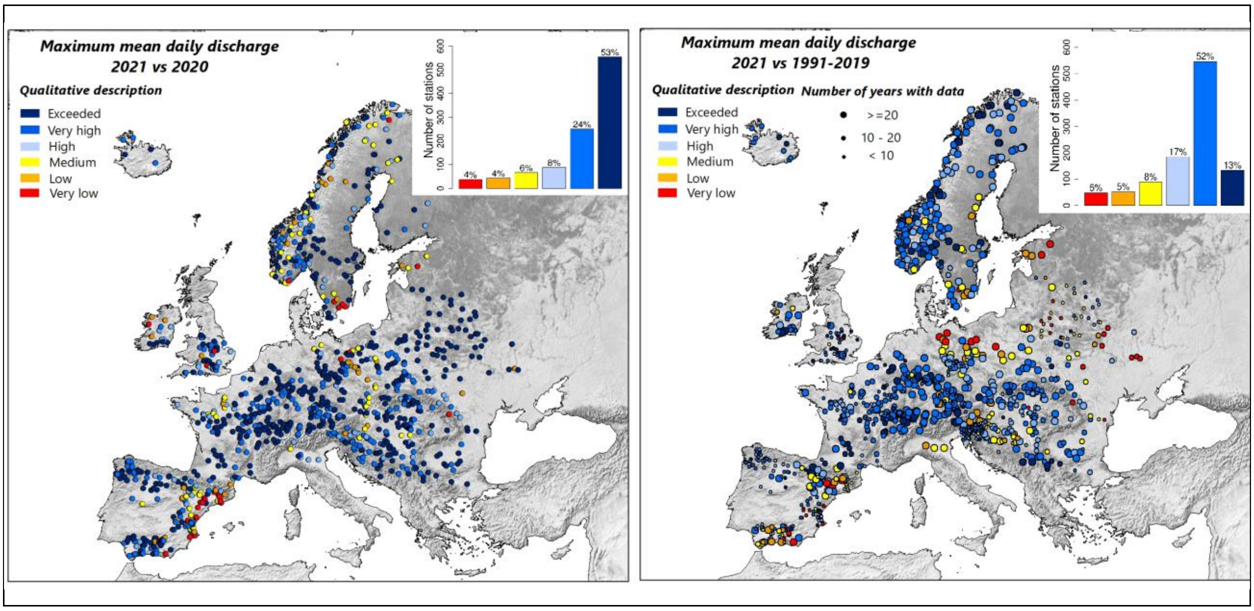


Figure 7: Spatial distribution of stations and the maximum values in 2021 compared with 2020 (left) and the historical period 1991-2019 (right)

3. Data quality of HDCC data collection

3.1 Rules applied for data quality

In EFAS, data plays a fundamental role, and the absence of the required quality can seriously affect the quality of results from applied models. Thus, quality evaluation and control is a very important activity for the quality of results.

Inappropriate quality levels can appear due to some obstacles during the life cycle of data. This section describes the data quality checks implemented within this project. The criteria for the implementation of the quality checks are based on ISO/IEC 25012 standard.

The implementation of quality checks requires a series of threshold parameters connected with measurement quality. The definition of these thresholds are based on the near real time series which are provided by the Data Providers. To ensure the threshold values calculated are truly representative, the time series provided should satisfy the following requirements:

(i) Validated data. To operate with the received data, they must go through a validation process. Usually the data received from Data Providers should be validated. Although this is done only for historic data series, not often is not applied to real time data series.

(ii) Complete range of measurements. The time series must cover periods of at least two years to make sure they include low and high flow seasons. Therefore, to fully characterize a gauging station, it must account for a measurement scope. It covers values ranging from the minimum to the maximum discharge or water level values registered by the station. Once defined thresholds of the maximum relative variation checks are defined for water level and discharge ranges. The greater the variability in the discharge or water level data, the more stable and reliable will be the thresholds calculated from their values.

(iii) Regular time intervals. The time interval between each measurement is crucial for a station. It is used in the calculus processes. It must remain constant during the complete time series. In case a Data Provider communicates changes in the frequency, the system will carry out the interpolation required to homogenize the previous data series according to the new ones. If this harmonization is not possible, a new independent series will be created.

(iv) No gaps. Time series must have an adequate number of consecutive measurements to properly characterize relative and absolute variations regarding their previous measured value. This allows more consistent statistical analysis. The presence of gaps in the series require the use of interpolation methods to estimate the missing values. If there is need to fill large gaps an accurate numerical modelling should be employed. For this reason, only gaps shorter than 3 days are interpolated.

(v) Homogeneous measurement units. Each measurement has a reference unit. This measurement unit must remain the same during the whole time series. For example, water level data can be reported in meters, centimetres or according to the station's local reference stage level.

Quality checks are carried out based on calculations of threshold values for either discharge or water level measurements so that a specific value will be tagged as 'suspicious' if the threshold value is exceeded. These quality checks occur in real time, meaning that once a datum enters the system it is immediately evaluated.

Thus, the quality of water level or discharge measurement is subject to the verification of a series of conditions:

Range test (QL1): Checks if the value is included in the station's measuring range.

Rating Curve test (QL2): This is carried out on stations that provide water level and discharge values. With this check we verify if the value represented by each data (Water Level – Discharge) follows the station's rating curve characteristics.

Maximum relative variation test (QL4-QL5): The data must have a relative variation compared to the previous measurement below a certain threshold. The previous measurement can either be the immediately previous one, or any other valid measurement. The best reference value is the previous one. The further away in time the data considered, the more restrictive will be the control carried out. . This test is implemented using two reference values: the immediately previous value, and the value measured 24hours before.

3.2 New data quality rules developed in 2021

A set of new quality checks has been developed in 2021. These quality checks will become part of the routine quality checks.

3.2.1 Monthly MIN/MAX (QL6)

This check is done comparing each value with the maximum and minimum values received during a month of observation, for the whole set of years. The maximum and minimum historical values for each month, from January to December, is regularly updated.

Each value is compared with the table of maximum and minimum registers for that month. In the case of exceeding the registered values, the quality flag change to suspicious and the Min/Max table is updated with this value (after verifying that the value is not an outlier). In other case, quality flag change to pass.

3.2.2 Negative discharge (QL7)

Negative discharge values are not always errors, some stations upstream of reservoirs or dams can have negative discharge (backwater flow). In other cases, negative discharges are considered errors and must be discarded.

A new check will be applied to selected stations that provides negative discharge values when not upstream of reservoirs or dams.

This control needs more information about spatial location of the stations. A backwater flag, will be included in the set of information requested to data providers when new stations are included in the system.

For existing stations that have provided negative discharge, providers will be contacted to clarify if it could be an error or not.

The behaviour of this quality flag will be as follows:

a) if data value of the variable is positive the flag check is always PASS.

b) if data value of the variable is negative, the quality flag is suspicious and data will be evaluated:

b1) if station has the backwater metadata active, the station could have negative values in the variable, so even if value is negative the flag check will be PASS

b2) if the backwater metadata is not activated, the data will be considered incorrect because it is an unexpected value and therefore the flag will be FAIL.

In case this check flags an observation as FAIL, the summary flag (L1) will be set to FAIL regardless the results of the rest of the checks.

3.2.3 Repetitive values (QL8)

When comparing value received against its previous ones, it is possible to identify sensor issues that result in the provision of the same measurements for some time. The causes can be due to a water in the sensor, or presence of branches or vegetation in the station.

To implement this check, it will be necessary to define the period in which the constant value of the observation will be checked. In some cases, sensors provide data in a small of interval time and there is small data variations. Those cases must not be confused with erroneous or suspicious values.

This check will be applied to near real-time and historic water level and discharge observations.

3.2.4 Visual check (QL10)

This check allows the hydrologist to manually flag data observation as FAIL or PASS regardless the results of the automated checks.

Summary flag L1

Each received data is assessed using the quality checks QL1 to Ql10. The results (i.e. PASS, SUSPICIOUS, UNCHECKED, or FAIL) for each quality check are combined to define the result of the analysis. The conjunct analysis of the quality flags has the purpose to minimize the risk to introduce erroneous values to the system, According to this conservative approach, the manual check by the expert hydrologist (QL10) overrides all the other quality checks, the final outcome of the protocol in then the summary flag L1. The L1 flag can have the following values: PASS, SUSPICIOUS, UNCHECKED, or FAIL.

3.3 Quality control statistics for 2021

The following pictures show statistics summaries of the quality checks applied, for all the data providers. The quality checks analysed are directly related with the outlier detection (Ql1 to QL10, as described above). Specifically, the graphs show the percentage of values that received the flags PASS, SUSPICIOUS, or FAIL.

The flag UNCHECKED means that for some reason (check parameters are not available, the previous value for comparison is missing, etc.) the validation could not be calculated. It happens mainly with new stations. The calculation of the parameters of the checks is done only when a long time series is available (usually, the data provider needs to be active for more than one year). For stations recently added, in a first stage the result of the check remains UNCHECKED. When the volume of data allows, the parameters are calculated, the flag is then PENDING, until all the checks are applied and then all the quality marks are produced.

Summary of QL1 results, for all the data providers (Period: May 2021 - Feb 2022).

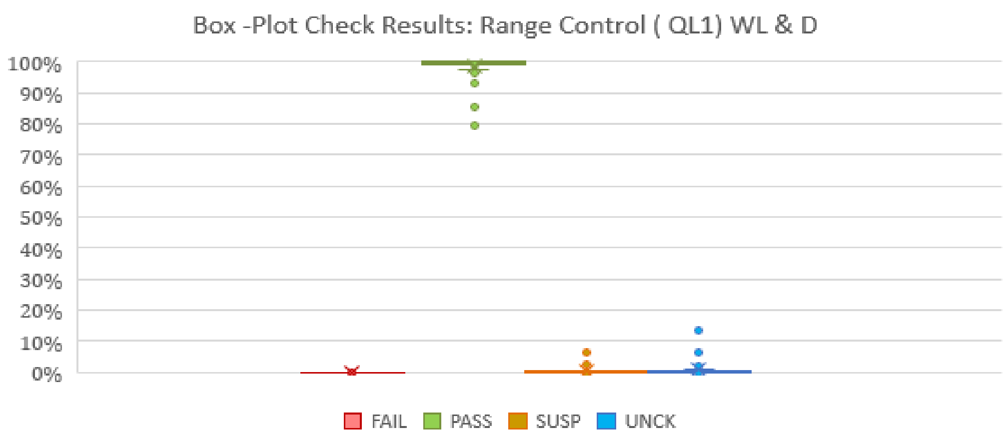


Figure 8: Box-plot range control (QL1).

Summary of QL4 (variation from previous value) results, for all the data providers (Period: May 2021 - Feb 2022):

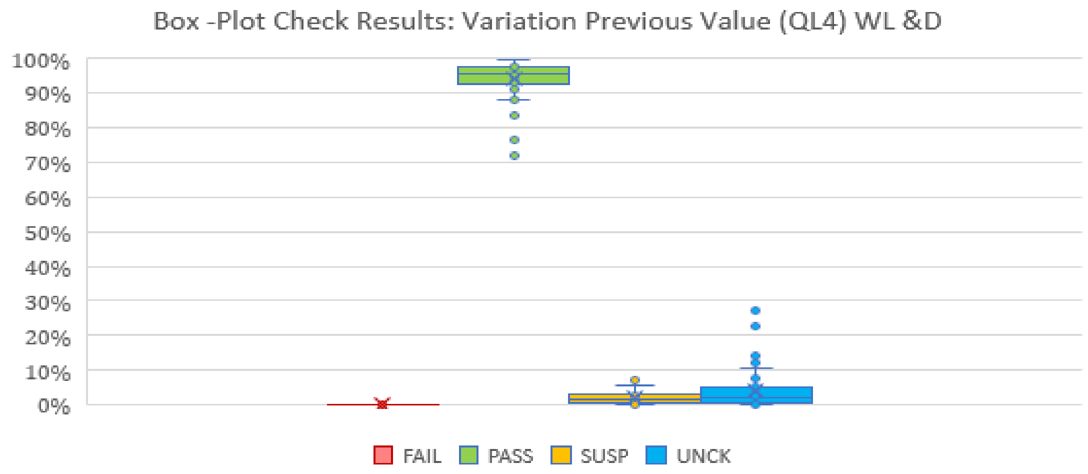


Figure 9: Checks results for all the data providers: variation vs previous data, same time (QL4).

Summary of QL5 (variation from the previous 24 hours) results, for all the data providers (Period: May 2021 - Feb 2022):

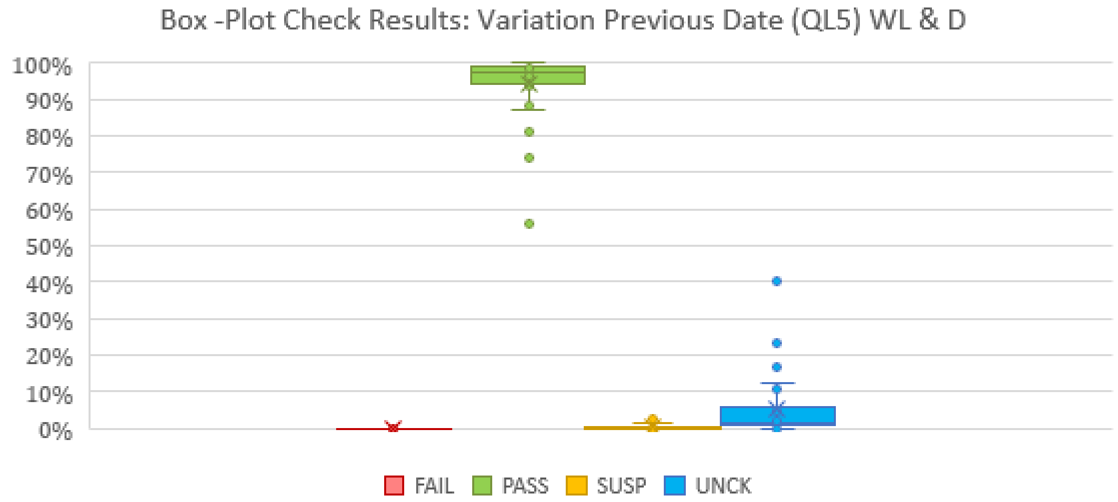


Figure 10: Checks results for all the data providers: variation vs previous date, same time (QL5).

4. Gaps Analysis on the CEMS hydrological data base

Initial considerations

This chapter analyses the gaps in the CEMS hydrological data collection for the year 2021, collected by the Hydrological Data Collection Centre (HDCC).

The CEMS hydrological data collection is continuously growing with hydrological data from 2,386 gauging stations across Europe. The data observation frequency among those varies from every minute to daily (see Figure 11). A gap occurs when either no data is received for a specific period of time or if the data received fails the quality control criteria and is considered as missing. The basic gap unit considered is a single missing value. A gap ends once the data delivery is resumed, and the missing values are not uploaded. The importance of a gap will depend on its length.

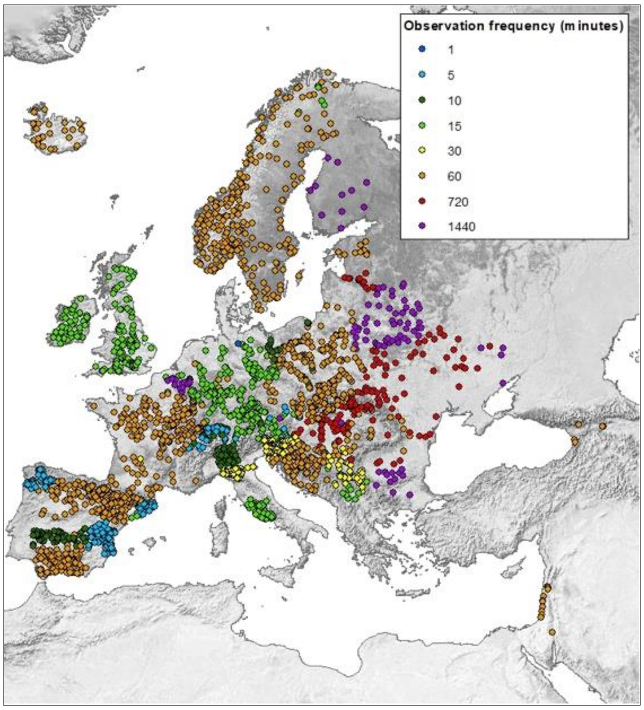


Figure 11: Observation frequency by station.

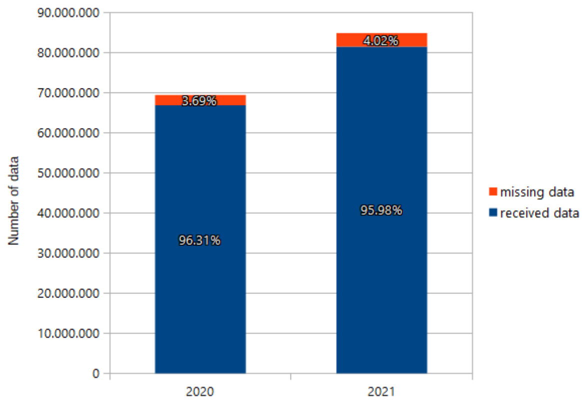


Figure 12: Reception rate comparison between 2020 and 2021.

4.1 Gap analysis

We analysed data from 1,963 stations providing water level and/or discharge values (see Introduction), from 44 data providers (DPs). 1,821 of these stations had problems with data transmission between January 1st 2021 and December 31st 2021 on at least one occasion.

In total 4.02% of all the data values expected for 2021 were not received, which is slightly higher in percent compared to 2020 (3.7%). 99% of all the 606,407 gaps lasted less than 1 day and 69% lasted less than 1 hour. To select only gaps relevant for HDCC operations, gaps of 1 hour or less are discarded as those do not interfere with the data processing tasks of the HDCC.

This filtering reduced the number of gaps to be analysed to 143,399, coming from 1,787 stations and for 2,802 station-variable combinations. This number is higher than the number of stations, as each station can provide up to two variables (water level and/or discharge values).

4.1.1 Gap classification by duration

We define five classes of duration:

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

Figure 13 (left panel) shows the number of gaps according to their duration. 95% of the gaps have a duration of ≤ 1 day, resulting mostly from changes in the data observation frequencies and/or delays in data transmissions. 3.7% last between 1 and 3 days, whereas 1.3% (1,841 gaps) lasted more than three days and required a follow up by the HDCC. Figure 13 (right panel) shows the distribution of those gaps longer than three days.

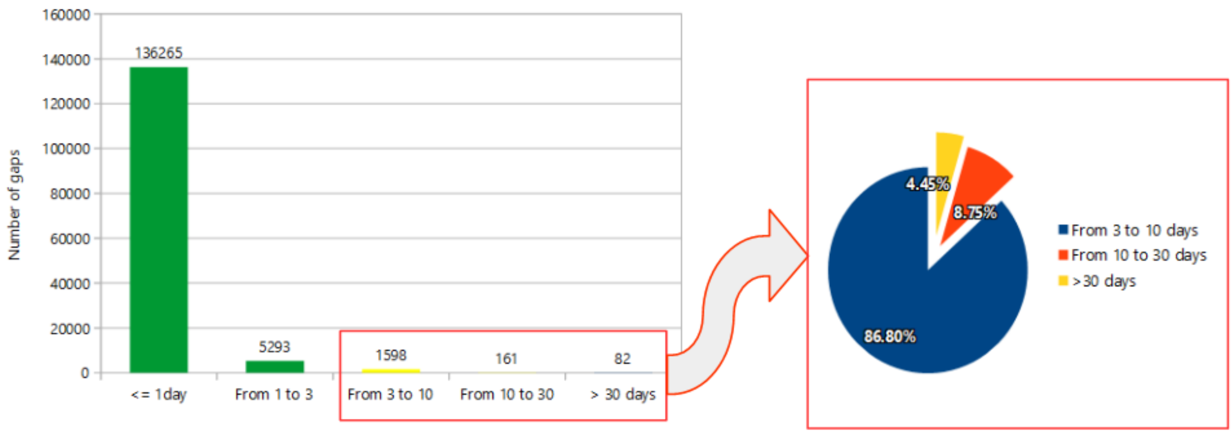


Figure 13: Number of gaps by duration (left panel) and distribution in percentage of those longer than three days (right panel).

Most gaps longer than 3 days last ≤ 10 days as the HDCC establishes contact with the respective data provider after three days of failed delivery. Usually, the data providers can solve the issues within a couple of days. Gaps longer than 30 days are less frequent as all parties involved have been notified and, in most cases, had time to solve the issues.

4.1.2 Gap classification by status

Once a gap occurs, 4 scenarios may unfold:

- FILLED: The gap is filled at a later stage, with the missing data sent by the data provider.
- FILLED INTERPOLATED: The gap is filled by the HDCC data interpolation process. Gaps with a duration of less than 3 days are filled by an automatic interpolation process.
- PENDING: Pending action, this applies to gaps recently detected.
- NOT FILLED: No interpolation or filling is carried out. It usually happens for gaps longer than 5 days. The gap remains.

In the case of interpolated data, if the missing data from the data provider is received at a later stage, the new data replaces the interpolated data.

Figure 14 shows the distribution of gaps status for each duration interval. Most gaps (of less than 3 days) are filled either through interpolation or data from the DPs. About a half of the gaps between 3 and 10 days are filled by interpolation, a quarter with data from the DP and nearly the last quarter remain not filled. The vast majority of gaps longer than 10 days are not filled and will be permanent unless the data providers deliver the missing values.

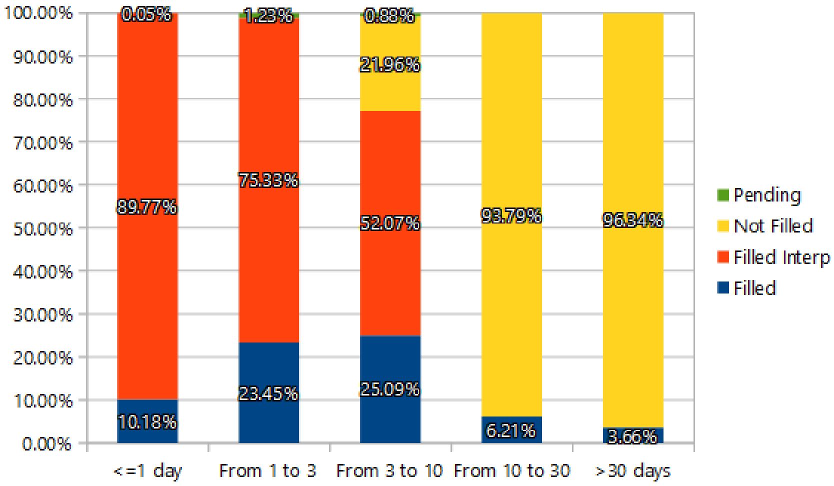


Figure 14: Percentage of gap status by gap length.

4.1.3 Other aspects to be considered

The 143,399 gaps analysed add up to 2,660,823 missing values covering a total of 45,441 accumulated days. The average length per gap is 0.32 days (about 8 hours), whereas the average number of gaps per station and variable is about 51.2; hence an average of 16 days of gaps for each data variable.

Figure 15 presents monthly boxplots with the percentage of received data against expected data for each data provider. The mean value ranges between 94.0 and 96.5 %, although some data providers provide lower ratios.

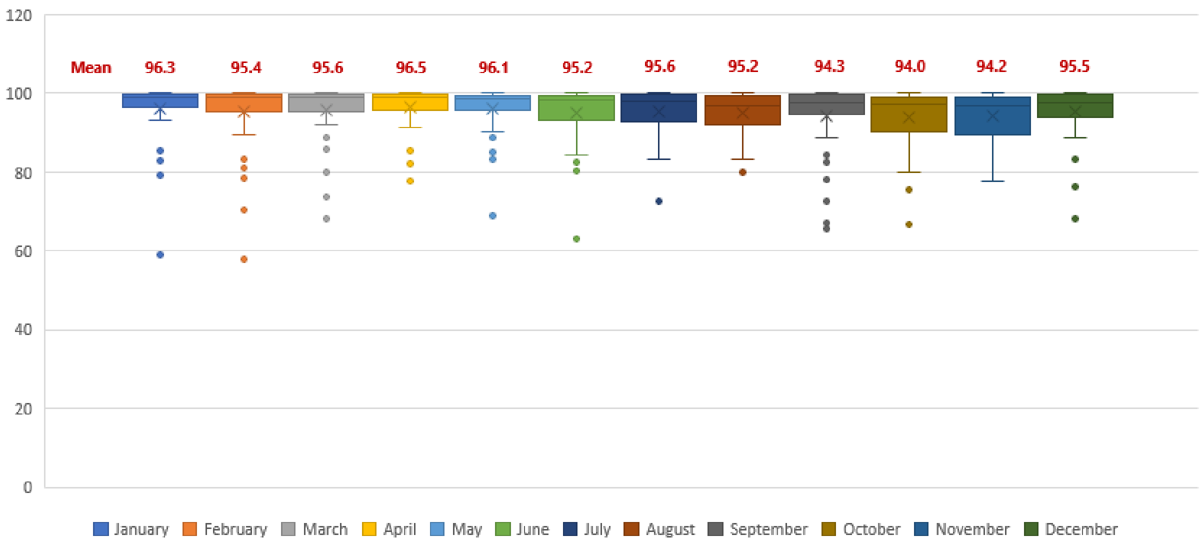


Figure 15: Box plot showing the monthly percentage of data received, out of the expected number of data records to be delivered from EFAS data providers.

When comparing these values to 2020 the average percentage for 2021 is slightly lower (95.3 against 96 %), even though the annual reception rate by data provider was always higher than 75%, and the monthly reception rate was never below 50 %.

The maps in Figure 16 and Figure 17 show the spatial distribution of gaps, with respect to the average gap duration (days) and maximum gap length (days), respectively.

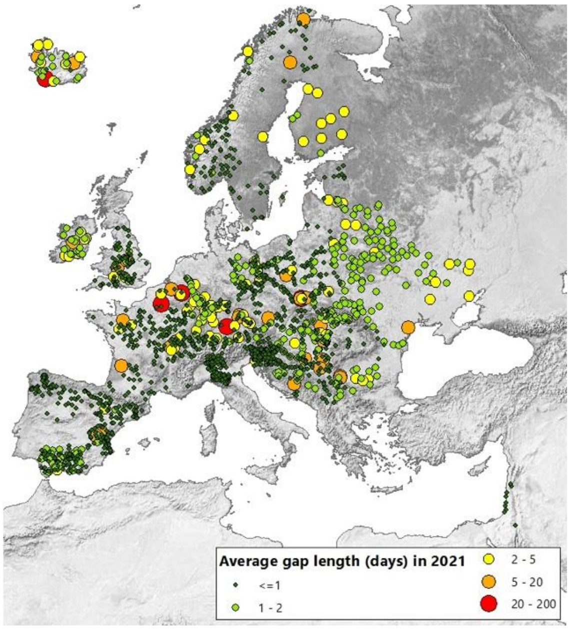


Figure 16: Average gap length in days per station.

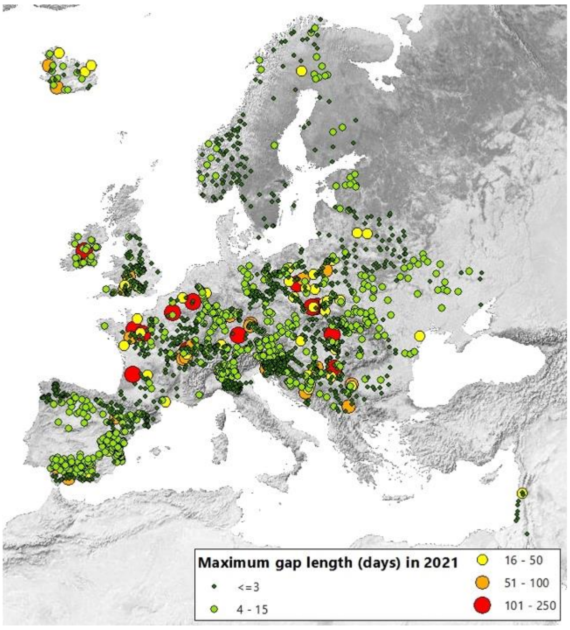


Figure 17: Maximum gap length in days per station.

4.1.4 Gap typology and proposal for future data collection strategy

In 1,841 cases (1.3% of all gaps), gaps were longer than 3 days and required the HDCC to communicate with the data provider. Based on communication with the data providers and their replies, it was possible to establish a gap classification system based on gap causes (see Table 6). This classification helps to develop and propose a series of measures to improve the data collection strategy both quantitatively and qualitatively.

Table 6: Gap classification with possible solutions.

Gap typology	Further info	Occurrence	Recommendation / possible solution
Technical issues between DP and HDCC	Issues concerning the Data Collection service between Data Providers and the HDCC: Delays in data transfers from DP to HDCC, changes in IP directions, problems with the servers, etc.	28.03%	Improving communications with DPs to achieve a more efficient and faster solution. (Prompt communications when missing data is detected or when IP directions are changed)
Lack of reply from Data Provider.	DP usually reply to HDCC communications, but on certain occasions we don´t receive replies.	17.33%	These issues rely entirely on the DP. A meeting between HDCC and DP to analyse the situation is highly recommended.
Data Sensor Failure	Sensor malfunction that causes data transmission failures, or wrong/unexpected data to be sent (i.e -9999 values).	16.35%	The solution is repairing the sensor or replacing it with a new one. This solution depends directly on the DP.
To be determined.	No information on this type of gap	15.48%	These are the cases of the smallest gaps (from 3 to 5 days). As it is not known which might be the cause of these gaps, it is difficult to propose a solution.
Limited resources	Lack of technical personnel	6.79%	No easy solution exists as it does not

of DP to attend data gap requests	availability to attend data gap requests on behalf of HDCC.		depend on the HDCC. Some agreement between the HDCC and Data Providers might help minimize the effort needed (data services access, etc.). Otherwise, this could result in the number of gaps to grow, issue should be discussed with EFAS.
Communications Failure between Sensor and DP	Communications Failure between Sensor and the facilities responsible for the data collection and transmission.	5.92%	This relies on the DP (data collection and transmission personnel). Quick communications help minimize the impact of missing data.
Readings taken only during specific hydrological conditions	Data values only obtainable under specific conditions (i.e. above a certain water level).	5.92%	For this kind of issues our proposal, whenever possible, is to look for an alternative station. In case this is not possible, it would be advisable to find out if missing values can be calculated from the station rating curve in order to complete the data series.
Gauging Station out of order	Usually caused by breakdown, maintenance, repairs, etc. because of lightning, floods, sensor replacement, long term breakdown	2.72%	If the station has any alternative sensor with identical characteristics, those data could be an alternative.
Extreme meteorological conditions beyond sensor capacity.	Extreme meteorological conditions that obstruct the correct functioning of the sensor. Frozen rivers are the most common cases in this category.	0.92%	We need to consider if the extreme meteorological conditions are odd and very rare situations or if they occur on a regular or frequent basis. If the events are regular and frequent, either an alternative station or a different placement would be advisable.
Delay due to stations requiring a manual intervention	Delay in the data collection of stations which require a manual intervention of personnel as well as lack of personnel to obtain the data.	0.54%	HDCC always procures to maintain quick communications with DPs when no data is being received.

Figure 18 shows the number/percentage of gaps by/per duration and typology.

The following considerations may be useful as well:

- If a specific station presents gaps repeatedly, an alternative station (located nearby) could be proposed as replacement or the station could be removed from the system.
- When transmission data often delays for a specific data provider or station, a possible solution to avoid the unnecessary communication between HDCC and DP could be to increase the response time for the data collection process, i.e., increase the time before considering the data as missing. This measure would reduce the need for HDCC to intervene when the missing data is likely to be automatically updated in the following data transfer.
- For gaps that are of less than 1 hour or apparent gaps in time series with irregular observation frequency, these could be avoided by normalizing the data series (i.e., aggregating data to 1-hour operational tables).

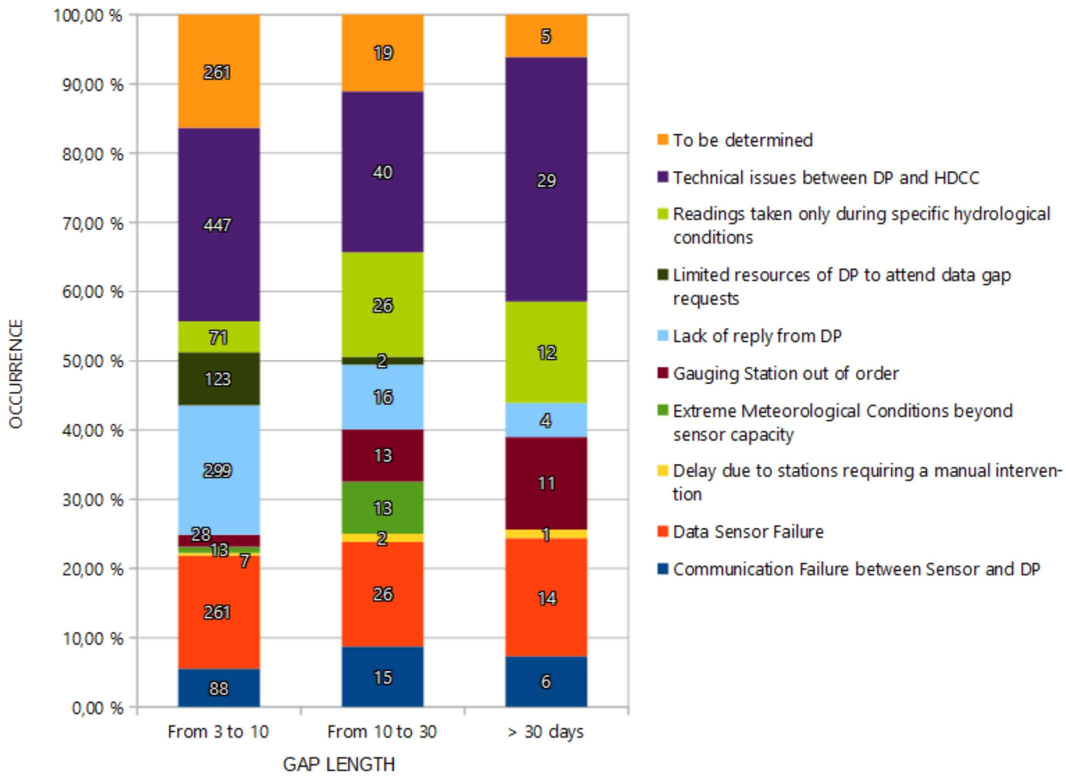


Figure 18: Number/percentage of gaps by/per duration and typology.

4.2 Outliers analysis

One of the data quality control procedures of the HDCC checks for outliers. Outliers are defined as values that are beyond their minimum or maximum threshold level. Those threshold levels usually correspond to the historical minimum and maximum value recorded by that station. Hence, they are station and variable specific and are usually provided by the respective DP.

Once a data value exceeds its threshold level it is marked in the database for further visual inspection. This is a necessary step for deciding if this outlier is an actual erroneous value or merely the consequence of a natural event. If an outlier is confirmed to be an erroneous data value, then it is flagged as such. If several consecutive outliers are detected, these are defined as a set of erroneous data values. A total of 108,469 outliers were detected in data from 436 stations out of the 1,963 stations studied in this report. Considering that the total number of values received is 81,431,744, the rate of outliers is approximately 0.13 %. Figure 19 illustrates the different types of outliers according to their aggregation and frequency.

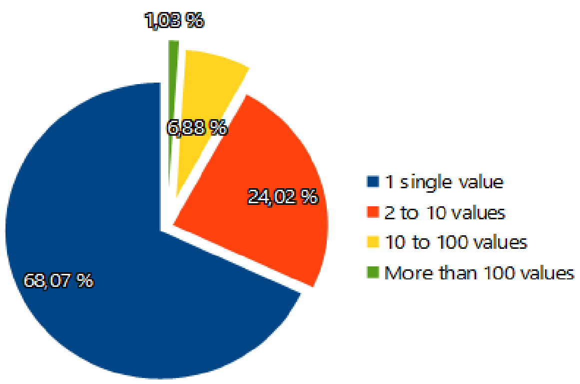


Figure 19: Sets of outliers and their frequencies.

Most outliers detected are single values while aggregations of subsequent outliers are the least frequent.

The following figures show the stations that registered outliers in 2021, the total outlier's duration in days per station and the rate of outliers relative to received data for each station.

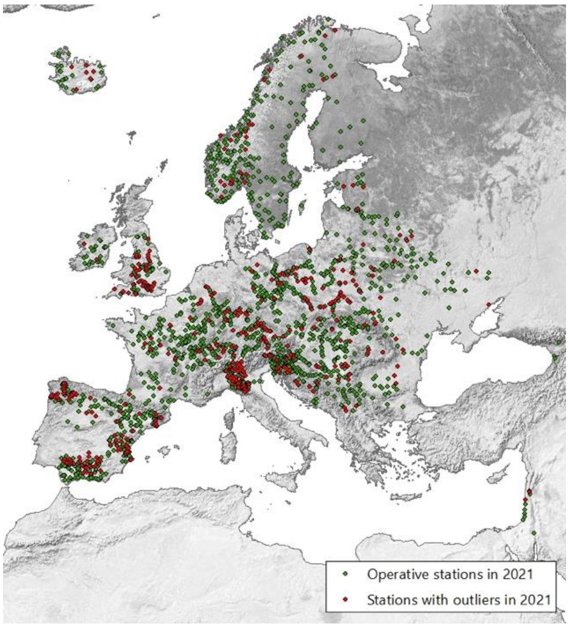


Figure 20: Stations that registered outliers in 2021.

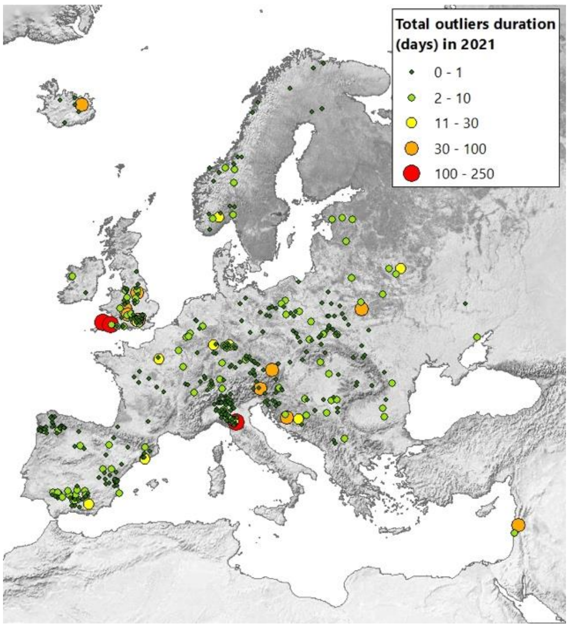


Figure 21: Total outliers duration in days in 2021.

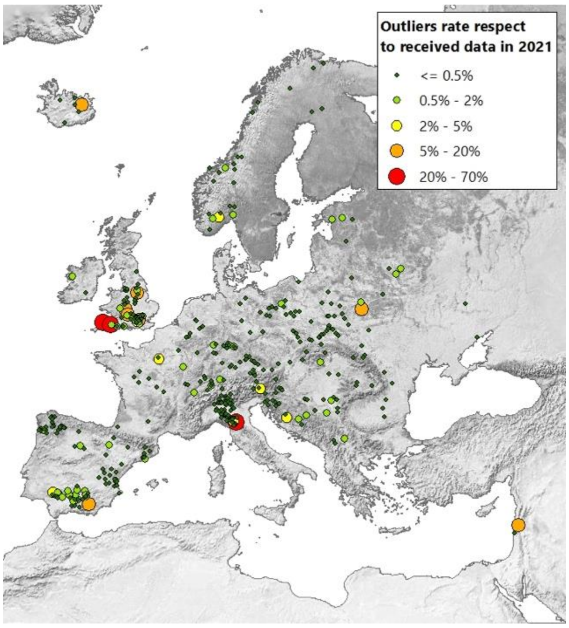


Figure 22: Percentage of outliers occurrence relative to the total amount of data received per station in 2021.

5. Analysis of Exceedance events

In this section, the hydrological stations that exceeded their threshold level during 2021 are analysed. A threshold level is a gauging-station specific value, usually discharge or water level, provided by the national/regional authorities responsible for a gauging station network. The number of threshold levels varies from 0 to 4 for each station. These levels help the authorities in assessing the current hydrological situation, and in case of a threshold exceedance they can start to plan and implement mitigation measures.

The analysis focuses on the exceedance of threshold levels for high river flows. An event is defined as each time a measured discharge or water level value exceeds any of the station's threshold levels. The event duration is considered from the first level exceedance until the values again drop below the lowest threshold level.

All near real-time observations (water level and discharge) are displayed on the EFAS website in the "National flood monitoring" layer. Where available, also the national/regional threshold levels are shown and exceedances of those are highlighted by the HDCC.

5.1 General description

Out of the 2,386 active stations initially selected for this report, threshold levels are available for 1,760 stations (74%). Compared to 2020, the number of stations with at least one threshold level has increased by 224. This increase has allowed EFAS to monitor new countries and areas: Ireland, Luxembourg, Northern Spain, Belgium, Germany and Slovakia (see maps in Figure 23 and Figure 24).

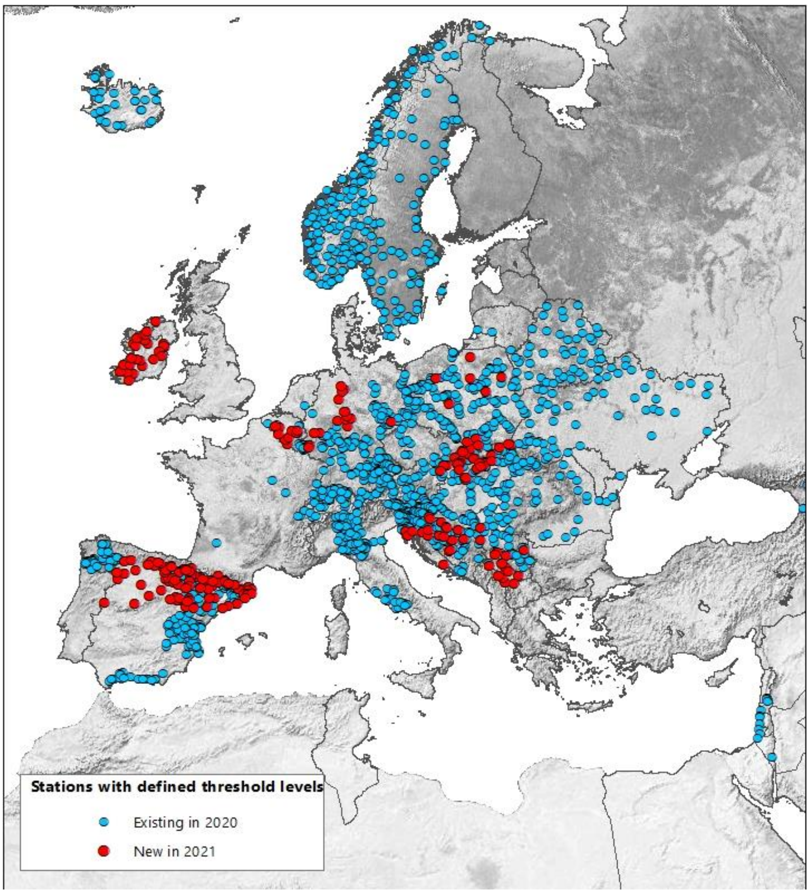
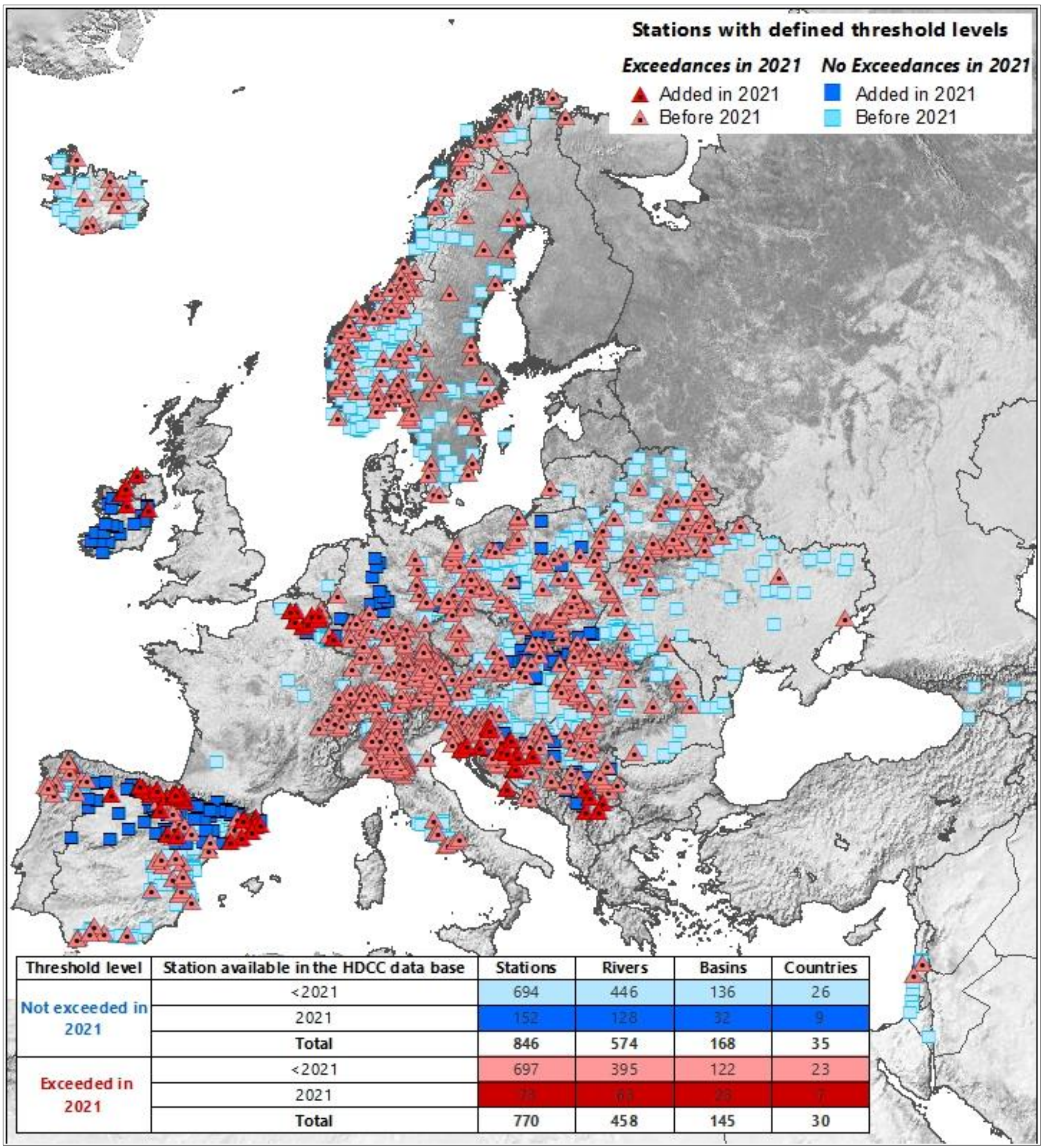


Figure 23: Distribution of the stations with threshold levels defined in 2020 and comparison with the new stations with threshold levels included in 2021.



Threshold level	Station available in the HDCC data base	Stations	Rivers	Basins	Countries
Not exceeded in 2021	<2021	694	446	136	26
	2021	152	128	32	9
	Total	846	574	168	35
Exceeded in 2021	<2021	697	395	122	23
	2021	73	63	23	7
	Total	770	458	145	30

Figure 24: Stations with no exceeded threshold levels in 2021 (in blue tones), differentiating between stations already existed in 2020 and stations added in 2021. Stations with threshold levels exceeded in 2021 in red tones (stations existed in 2020 and stations added in 2021). The table shows a summary of threshold levels (exceeded and no exceeded) in 2021 by spatial aggregation levels (station, river, basin and country), providing the total number of stations, rivers, basins and countries in 2021 (all).

During 2021, a total of 770 stations from the 1761 stations with threshold levels defined have registered exceedances of at least the first one, happening in 458 rivers, 145 basins and 30 countries, as can be seen in Figure 24. Red tones triangles represent stations that had threshold levels exceedances. 73 of the 770 stations were added to the HDCC data collection during 2021. In blue can be seen the areas without exceedances during the last year (mainly Ukraine, Belarus, Poland, East of Romania, Norway, Austria and Hungary.

5.2 Duration of Exceedances

5.2.1 Duration per station

Figure 25 shows the number of events and their total accumulated duration per station during 2021. A total of 6427 exceedances were recorded during 2021 at 770 stations, twice more than in 2020 when

3635 exceedances were registered. The average number of events by station in 2021 is 8.3 (only 3.8 in 2020) and the average of the accumulated duration per station is 10.5 days in 2021 and 6.2 days in 2020. So, on average there are more events registered in 2021 and with a longer duration.

Considering the number of events in more detail, during 2021, the highest percentage of stations (34%) registered only a single event. In a second place, the percentage of stations with 3-4 events was 18%, and only 1.3% of the stations registered more than 40 events. This last occurred mainly in Spain, Italy, Austria, Bosnia & Herzegovina and Switzerland, concretely in Llobregat, Po, Ter, Danube, Rhine and Neretva river basins. The main cause could be the definition of the thresholds by each data provider. In some cases they can be very low.

Regarding the accumulated duration of the events, 32% of the stations summarised less than 1 day of accumulated time, 46% summarised events during less than 10 days and 2% registered events during more than 100 days. This last happened in Spain (Llobregat basin), Austria (Danube basin), Germany (Rhine basin), Bosnia and Herzegovina (Neretva basin), Belarus (Dnieper basin), Hungary (Danube basin), Slovadia (Danube basin), Ukraine (Dnieper basin) and Belarus (Dnieper and Vistula basins).

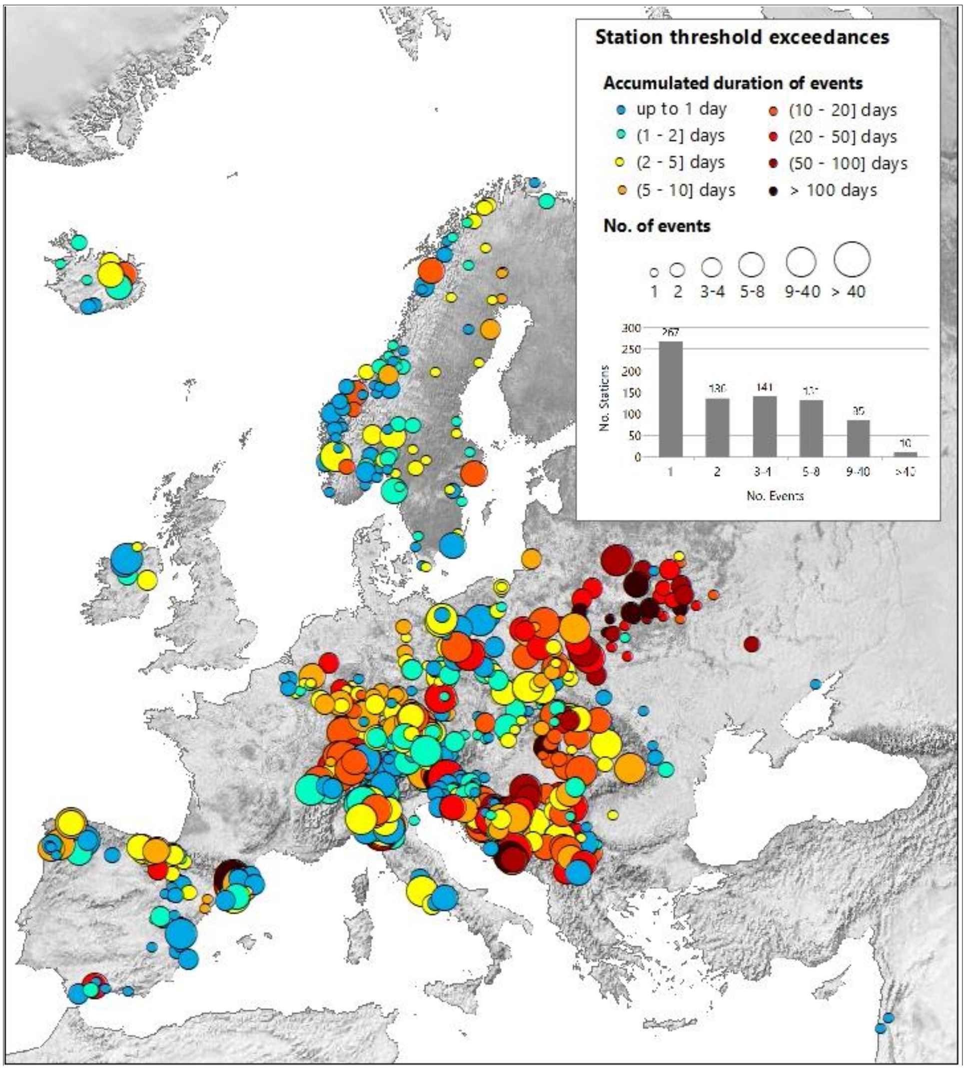


Figure 25: Station classification according to number of events (symbol size) and total accumulated duration of exceedance events (colour coding). Bar chart on the upper left corner shows the number of stations per event frequency

Figure 26 shows the classification of the stations during 2021 considering the average event duration. The average duration of the events was 1.25 days in 2021, compared to 1.6 days in 2020. For 85% of the stations the average duration is less than 1.5 days and less than 5 days for 95% of the stations. Only 12% of the stations recorded events lasting more than 20 days.

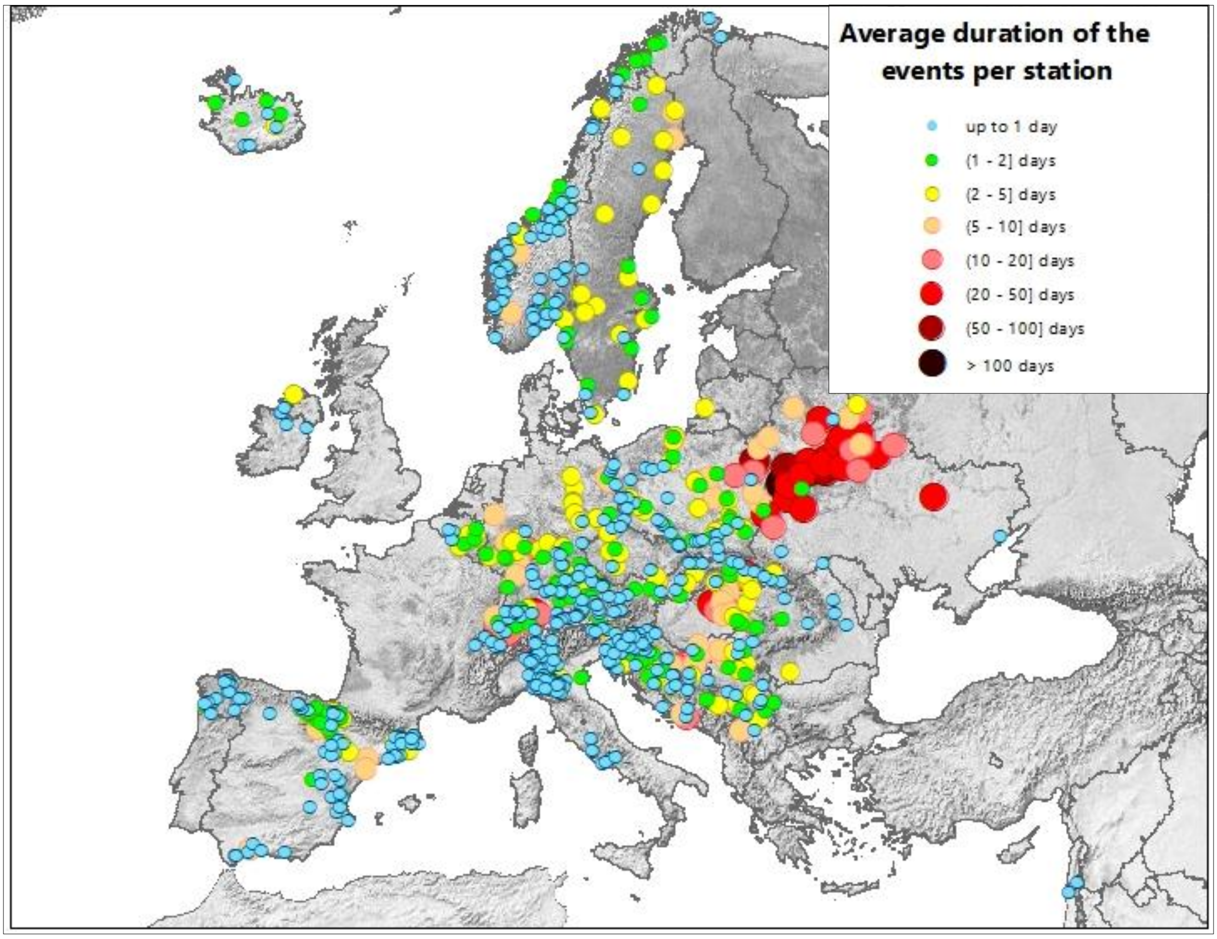


Figure 26: Average duration of events for EFAS stations in 2021.

Events lasting less than 1.3 days were the most common and occurred in 86% of the cases and in 365 out of 451 rivers with any registered event (80% of the rivers with events had a duration of less than 1.3 days). The ones with the largest number of these short events are Llobregat and Ter in Spain, Cedra, Seveso and Oglio in Italy, Aare in Switzerland and Pland, Bug in Poland, Neretva in Bosnia & Herzegovina, and Lake Ossiach, Lake Worth and Inn river in Austria. This kind of events took place majority in Summer (38% of the cases), during July and August.

The longest 69 events (over 20 days) were located across:

- Austria: Lake Faak and Lake Ossiach in the Danube river basin and Lake Constance in the Rhine river basin.
- Belarus: Dnieper basin (Berezina, Besed, Dnieper, Goryn, Iput, Pripyat, Pthich, Slutch, Sozh, Stviga, Styr, Svisloch and Ubort rivers), and Vistula basin (Lesnaya and Nerw rivers).
- Bosnia & Herzegovina: Danube basin (Sava river) and Neretva basin (Krupa and Trebizat rivers).
- Croatia: Danube basin (Sava river).
- Germany: Rhine basin (Rhine river).
- Hungary: Danube basin (Bodrog, Tisza and Zagyva rivers).

- Kosovo: Drin-Bojana basin (White Drin river)
- Poland: Vistula basin (Bug river).
- Serbia: Danube basin (Nera river).
- Slovakia: Danube basin (Bodrog and Latorica rivers).
- Spain: Guadalhorce and Llobregat basins (Guadalhorce and Llobregat rivers respectively).
- Switzerland: Rhine basin (Obersee river).
- Ukraine: Danube basin (Latorica river), Dnieper basin (Horyn, Snov, Stokhid, Styr and Vorskla rivers) and Vistula basin (Western Bug river).

These longest events are distributed throughout the year as follows: 80% of the events took place between Spring and Winter, mainly during February, March and April. The stations with events of more than 20 days for the longest time (8 months) is located in Hungary, on the Risza river (Danube basin), and its duration was 206 days. With 6 months of exceedances, there are three stations, one of them in Ukraine (Stokhid river in Dnieper basin), other in Belarus (Dnieper river) and the last one in Spain (Llobregat river). It is necessary to take into account that these events only overpassed the first threshold level, except for the first one (in Hungary) which overpassed the third level for 4 days at the end of February.

5.3 High threshold level exceedances

This section identifies the most severe events. As mentioned above, the number of threshold levels per station varies across Europe between 0 and 4. In this subsection we will treat the following cases as high level events:

1. stations with more than 1 threshold level, and the highest threshold has been exceeded
2. stations that have only 1 threshold level, and the level (discharge or water level) has been exceeded by at least 50%.

There were 191 high level events in 2021, distributed across 126 stations that exceeded 1) or 2). Figure 27 shows the spatial distribution and duration of the high level events for 2021.

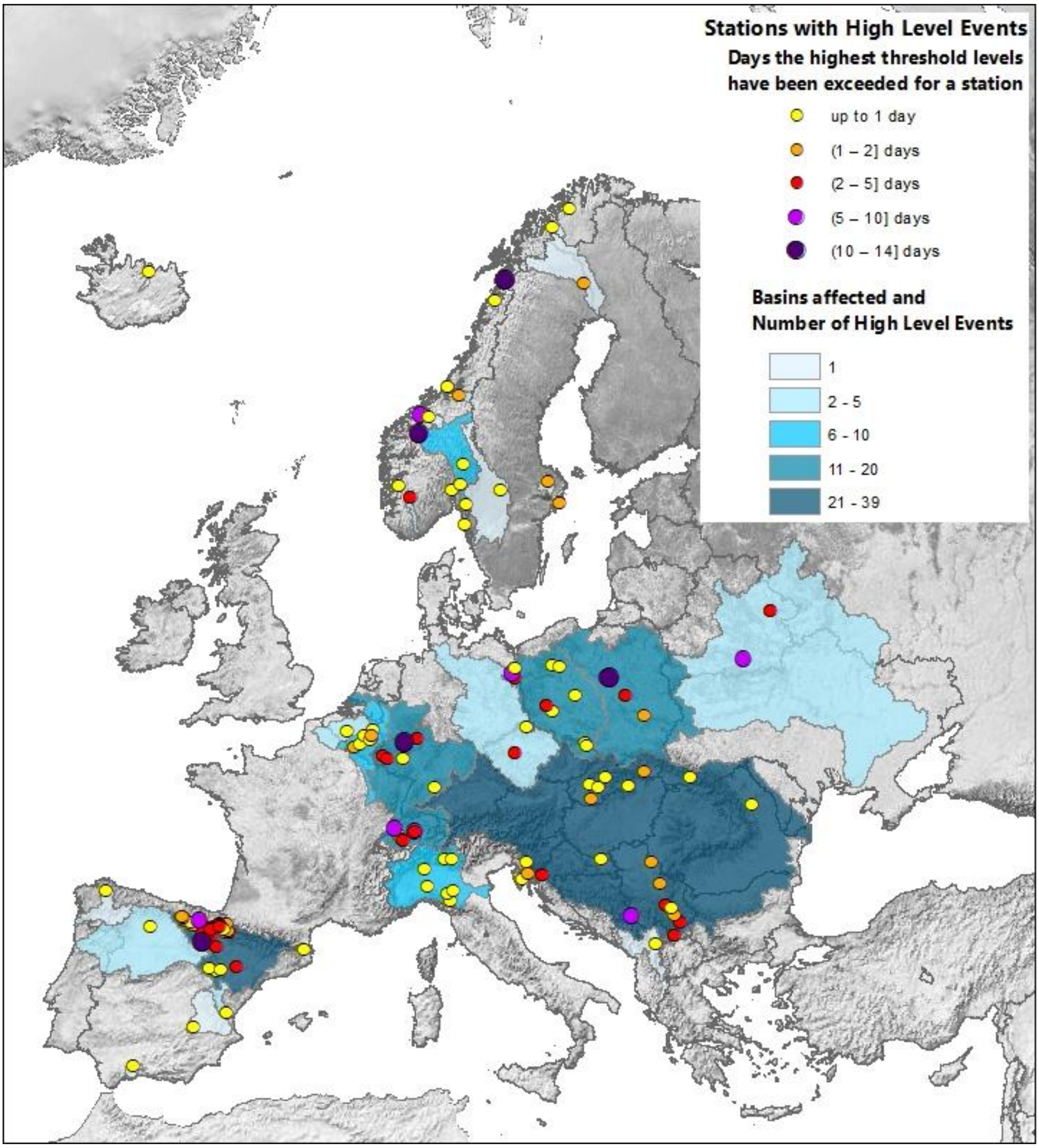


Figure 27: Duration of the high level events and river basins where the stations are located

Nearly 25% of these stations are located in the Ebro basin (32 stations), with 2.3 days of average time above the highest threshold. A station with 12.3 days of exceedance can be seen in the map. For another station the event lasted 8.5 days. The events took place in December 2021.

In the Danube basin there is a total number of 19 stations (15% of the stations with high events). The averaged time over highest threshold for all the stations is 1.6 days. The maximum number of days with exceedances is 5.9 days for one of the stations. The more remarkable events took place in the first place in February, but there are events in January and October too, and shorter in May.

The third basin with the most stations with exceedances is the Rhine basin, with 12 stations (9.5%). The average duration of the events is 3.4 days and in this basin we find the station with the most days above the highest thresholds: 13.3 days. The longest event was during February, but this one affected to a more reduced area. The longest event in the Rhine basin took place in the summer, in July, when heavy rains severely affected this area. This event has been analysed in detail in the "Detailed Assessment Report of EFAS performance during the July 2021 floods", where the HDCC collaborates with a hydrological analysis that evidences the severity and extension of the event, even two stations located in the Rhine basin exceeded their Historical Maximum value. These stations are: Hattingen and Bollendorf, on the Ruhr and Saar rivers, respectively. The Meuse basin was affected too in this event and is mentioned two paragraphs below.

In fourth place, the Oder river basin includes 11 stations with exceedances, 1.2 days of average duration and 6.53 days of maximum duration. The main events occurred during February in four of the stations. It is remarkable the event registered in May too.

The next two basins with the same number of stations with high events are the Meuse and the Po, with 7 stations each one, but in the Meuse we can find stations with more than 2 days of events, and in the Po the maximum duration is 0.3 days. The event in the Meuse basin took place in July, related to the same event in the Rhine basin. In the Po basin the events were registered in January, August and September.

Figure 28 shows the monthly distribution of the events during 2021 and its comparison with the previous year. Considering the high events only, as said before, the month with more events during 2021 is February, followed by December, so winter can be considered the period of the year with more exceedances. There are big differences compared with 2020 when the greater number of events were registered in Jun

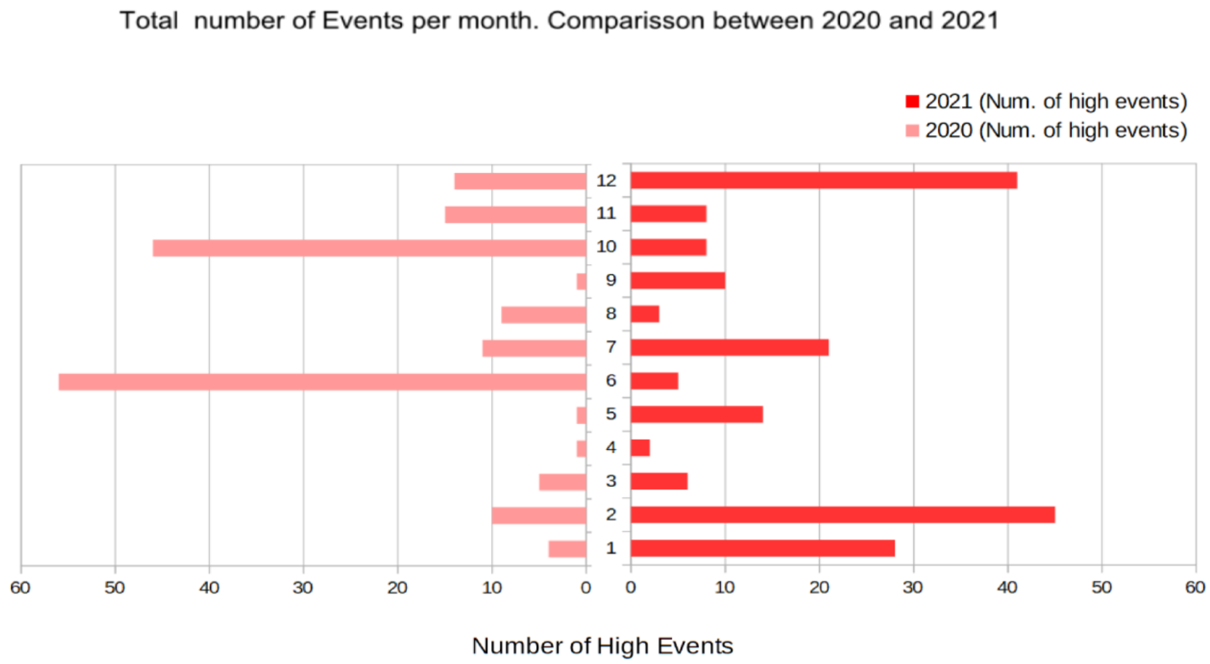


Figure 28: Total number of High Events for each month in 2021 and the comparison with 2020.

6. Summary

During 2021, the CEMS Hydrological Data Collection Centre welcomed the Republic Hydro Meteorological Service of the Republic of Srpska, the Water management Agency of Luxembourg, and the Agenzia Regionale di Protezione Civile Lazio as three new hydrological data providers to the CEMS hydrological data collection network.

In addition, a number of existing data providers (DP) increased the number of stations providing real-time hydrological data to the HDCC as a response to the EFAS data collection campaign for the next LISFLOOD model calibration. As a result of this campaign 395 new real-time stations were received by the HDCC. This brings the number of data providers and stations actively providing data by the end of 2021 to a total of 49 and 2369 respectively, with an increase of 2% and 19% compared to 2020.

In addition, 26 partners provided in 2021 new historic data sets. With a total of 77,615,375 new historical data values, the volume of the historical database has increased in 2020 by 62% compared to 2020.

In the following, the key findings of the various analyses are summarized.

6.1 Hydrological Conditions

According to the data collected, the hydrological conditions of the stations in 2021 present some particularities that are worth mentioning:

- The water contribution in 2021 is higher than in 2020 but still it was slightly lower than it was in the historical period 1991-2019. Drier conditions were more notable in the Upper Danube river basin and stations of Ukraine, Spain and Norway.
- The maximum and minimum mean daily values of discharge in 2021 followed a slightly more extreme behaviour than 2020. Minimum and maximum mean discharge were reached in a moderate number of stations, almost all of them in Spain, British Isles, Scandinavian Peninsula, the Danube, upper Rhine, Vistula, Dnieper and Dniester river basins.
- When comparing the maxima in 2021 to the period 1991-2019 133 stations of Spain, France, England Southern Scandinavian Peninsula, Iceland and Meuse, Rhine and Danube river basin exceeded the maximum mean daily discharge, together with other stations in Ireland and Dnieper and Vistula river basins.
- The hydrological conditions in upper Danube, Elbe, Oder, Rhine river basins and stations of Ukraine, Spain and Norway were was much drier compared to the historical average.

6.2 Gaps and Outliers

Regarding data gaps, the majority (69%) of them had a duration of less than 1 hour and were due to time interval variations (irregular data observation frequencies) or delays in data transmissions. Gaps that have a duration less than 3 days are filled by the HDCC data interpolation process. Gaps of longer duration are only filled if the data is provided by the authorities responsible of the hydrological data provision upon request from the HDCC.

Comparing 2021 with 2020, we see that the rate of received data vs expected data has slightly decreased in 2021 (95.98%) with respect to 2020 (96.31%). The number of gaps has increased in 2021 with respect to the previous year (606,407 vs 525,936), although it should be taken into account that the total number of received data has increased by 22%. The cause of data gaps was identified in 85% of the cases and solutions have been proposed accordingly. However, for the remaining 15% of the cases the causes remain unknown.

The analysis reveals that the percentage of outliers in 2021 is really low compared to the annual amount of data received (0.13%). Most outliers are isolated data values.

6.3 Exceedances Events

224 additional stations with threshold levels were uploaded in the HDCC data collection in 2021 and has allowed to expand the analysis covering more areas of Europe. 73 out of the 224 newly collected stations overpassed their threshold level. T

he high number of new stations is the main cause of the increase in the number of exceedances in 2021, together with the duration of the events, which have been more oscillating, breaking levels for less time but repeatedly, that is, many events but shorter in general terms.

With regards to high level events, 126 of the 770 stations (16.3% of the stations) exceeded their highest threshold. 36 of them corresponds to new stations included in 2021. During 2020 there were 583 stations with exceedances and 93 with high events (15.95%). The percentages are almost the same in both years.

The most affected basin with the highest events is the Ebro. 20% of the highest events took place in this area. In second place is the Danube basin (13 % of the highest events), followed by the Rhine, Vistula and Oder basins with 6% of the highest events.

List of figures

Figure 1: Spatial distribution of data providers to the CEMS 3

Figure 2: Spatial distribution of the 1963 selected stations and variables measured 7

Figure 3: Spatial distribution of stations according to the length of their historical time series (left) and catchment size (right). 8

Figure 4: Spatial distribution of normalized mean discharge values in 2021 9

Figure 5: Spatial distribution of Normalized Variation Index in 2021 with respect to 2020 (left) and Streamflow Variation Index the period 1991-2019 (right). 11

Figure 6: Spatial distribution of stations and the minimum values in 2021 with respect to 2020 (left) and the period 1991-2019 (right). 12

Figure 7: Spatial distribution of stations and the maximum values in 2021 compared with 2020 (left) and the historical period 1991-2019 (right). 13

Figure 8: Box-plot range control (QL1) 16

Figure 9: Checks results for all the data providers: variation vs previous data, same time (QL4) 17

Figure 10: Checks results for all the data providers: variation vs previous date, same time (QL5) 17

Figure 11: Observation frequency by station. 18

Figure 12: Reception rate comparison between 2020 and 2021. 18

Figure 13: Number of gaps by duration (left panel) and distribution in percentage of those longer than three days (right panel). 19

Figure 14: Percentage of gap status by gap length. 20

Figure 15: Box plot showing the monthly percentage of data received, out of the expected number of data records to be delivered from EFAS data providers. 20

Figure 16: Average gap length in days per station. 21

Figure 17: Maximum gap length in days per station. 21

Figure 18: Number/percentage of gaps by/per duration and typology. 23

Figure 19: Sets of outliers and their frequencies. 23

Figure 20: Stations that registered outliers in 2021. 24

Figure 21: Total outliers duration in days in 2021. 24

Figure 22: Percentage of outliers occurrence relative to the total amount of data received per station in 2021. 24

Figure 23: Distribution of the stations with threshold levels defined in 2020 and comparison with the new stations with threshold levels included in 2021. 25

Figure 24: Stations with no exceeded threshold levels in 2021 (in blue tones), differentiating between stations already existed in 2020 and stations added in 2021. Stations with threshold levels exceeded in 2021 in red tones (stations existed in 2020 and stations added in 2021). The table shows a summary of threshold levels (exceeded and no exceeded) in 2021 by spatial aggregation levels (station, river, basin and country), providing the total number of stations, rivers, basins and countries in 2021 (all). 26

Figure 25: Station classification according to number of events (symbol size) and total accumulated duration of exceedance events (colour coding). Bar chart on the upper left corner shows the number of stations per event frequency 27

Figure 26: Average duration of events for EFAS stations in 2021 28

Figure 27: Duration of the high level events and river basins where the stations are located 30

Figure 28: Total number of High Events for each month in 2021 and the comparison with 2020. 31

List of tables

Table 1: Additional stations per CEMS HDCC data provider. 4

Table 2: provides the most important statistics summarizing all the changes to the HDCC database in 2021. 5

Table 3: Number of data providers, stations and values managed during 2021. 6

Table 4: SVI and NVI classes. Positive/Negatives indices indicates larger/smaller mean discharge in 2021 compared to 1991-2019 (SVI) or 2020 (NVI).. 10

Table 5: Classification based on percentiles.. 10

Table 6: Gap classification with possible solutions.. 21

References

Bo-cai, G. (1996). NDWI—A normalized difference water index for remote sensing of vegetation liquid water from space. *Remote Sensing of Environment, 58*(3), 257-266. doi:https://doi.org/10.1016/S0034-4257(96)00067-3

Nalbantis, I. T. (2009). Assessment of Hydrological Drought Revisited. *Water Resources Management , 23*, 881–897. doi:https://doi.org/10.1007/s11269-008-9305-1

Zhangyan, J. H. (2006). Analysis of NDVI and scaled difference vegetation index retrievals of vegetation fraction. *Remote Sensing of Environment, 101*(3), 366-378. doi:https://doi.org/10.1016/j.rse.2006.01.003.

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.

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.

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