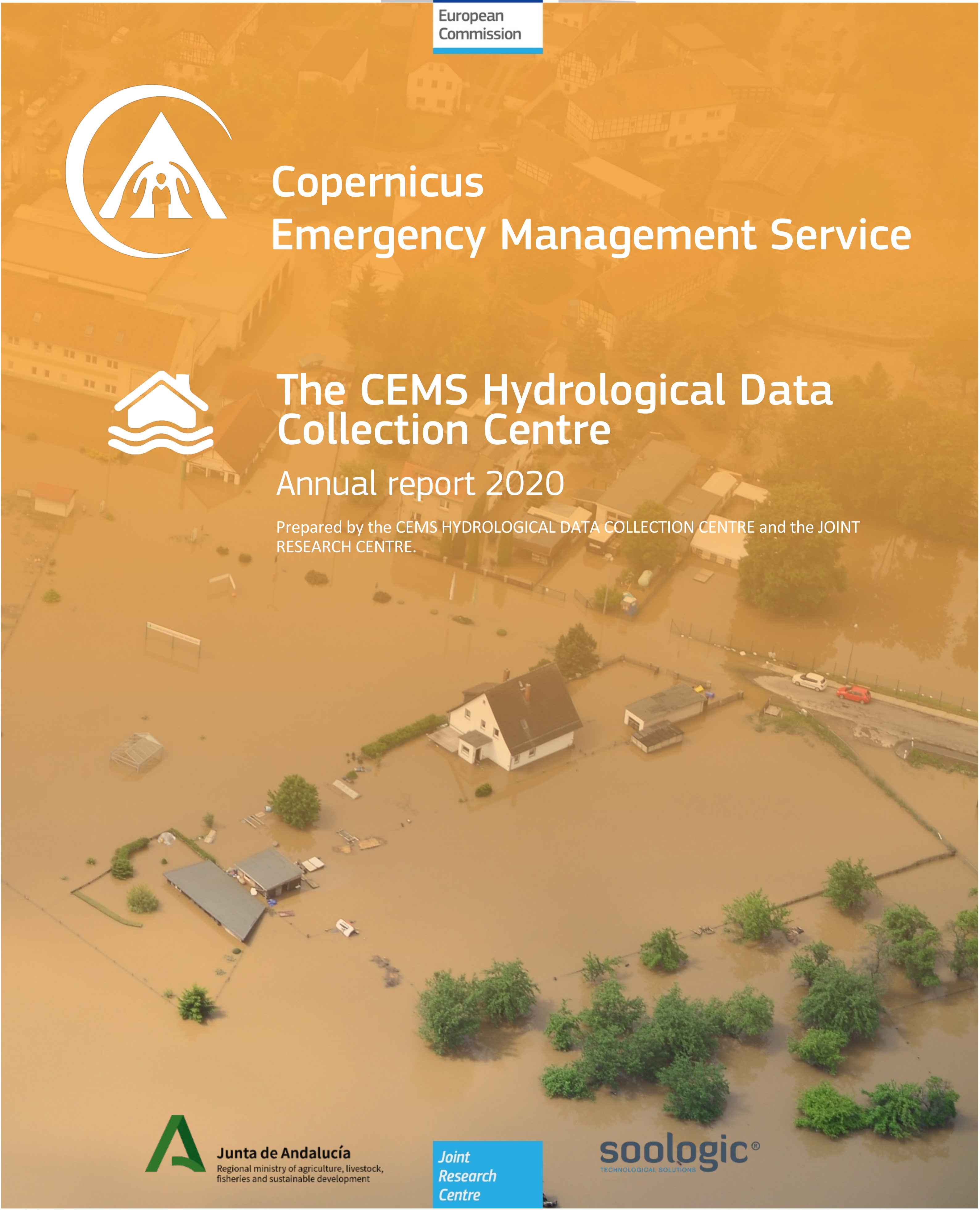
European
Commission

Copernicus
Emergency Management Service



The CEMS Hydrological Data Collection Centre

Annual report 2020

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

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

Joint
Research
Centre

soologic®
TECHNOLOGICAL SOLUTIONS

Copernicus Emergency Management Service

This publication is a Technical report by the Copernicus Emergency Management Service - Hydrological Data Collection Centre (operated by the Agencia de Medio Ambiente y Agua de Andalucía and Soologic Technological Solutions S.L) 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 scientific output expressed does not imply a policy position 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://ec.europa.eu/jrc

JRC126323

Ispra: European Commission, 2021

© European Union, 2021

The reuse policy of the European Commission 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). Except 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. For any use or reproduction of photos or other material that is not owned by the EU, permission must be sought directly from the copyright holders.

All content © European Union, 2021

How to cite this report: García Padilla, M., Santiago Nieto, F., Lara-Fernández, A., Montoya-Manzano, G., 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 2020, European Commission, 2021, JRC126323.

Contents

1 Acknowledgement

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.

2 Introduction

2.1 Background

This report contains an analysis of the hydrological data received by the CEMS Hydrological Data Collection Centre (HDCC) for the year 2020. 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 2020, 48 data providers contributed with near real-time hydrological data at 1,990 stations to the CEMS

Hydrological Data Collection, covering 32 countries and 50% of all the European water basins.

Figure 2.1: Spatial distribution of data providers to the CEMS (full list in Annex 1).

In the following section we first highlight the growth of the HDCC database in 2020, 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.

2.2 Update the HDCC database in 2020

During 2020, two additional hydrological data providers contributed with their hydrological data to the HDCC. Those are:

- the Icelandic Meteorological Office, Iceland with 28 stations
- the Confederación Hidrográfica del Júcar, Spain with 63 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. That are:

- Slovenian Environment Agency with 102 stations.

This makes a total of 193 new stations in the HDCC database since 2019. In addition, some existing EFAS data providers uploaded new historic data sets during 2020. An overview is given in Table 2.1.

Table 2.1: Historic data received during 2020

Country	Hydrological data provider	Dataset received during 2020. (year/s)
Poland	Institute of GeoSciences Energy Water and Environment	1990-2018
Ukraine	State Emergency Service of Ukraine - Ukrainian Hydrometeorological Center	2019, 2020
Serbia	Republic Hydrometeorological Service of Serbia	2019
Ireland	Office of Public Works	Downloaded
Norway	Norwegian Water Resources and Energy Directorate, Hydrology Department	2016 to 2019
Romania	Institutul Național de Hidrologie Si Gospodarire A Apelor	2018

Table 2.2 provides the most important statistics summarizing all the changes to the HDCC database in 2020.

Table 2.2: Number of data providers, stations and values managed during 2020.

	Before 2020	In 2020	End of 2020	Increment
Data Providers	64	3	67	4,5%
Active Data Providers (Portugal provides historical data)	46(+1)	2	48(+1)	4,3%
No Of Stations Registered	3,253	280	3.533	8,6%
No Of Active Stations	1,792	198	1.990	11,0%
No Of Near Real-Time Values	34,215.577	73,142,406	421,357,983	19,3%
No Of Stations with defined threshold levels	1,276	204	1.480	16,0%
No Of Historic Values	113,588,724	12,357,356	125,946,080	10.9%

2.3 Analysis of the data in the HDCC database

The hydrological data received by the CEMS HDCC for the year 2020 is analysed in the following three chapters, each focusing on a certain aspect:

- Chapter 3: An analysis on the general hydrological conditions across Europe, focusing on important deviations of average discharge.

- Chapter 4: An assessment of the HDCC Data Collection in terms of gaps and outliers, including a classification according to causes, duration, length and distribution.
- Chapter 5: An evaluation on the threshold level exceedances, looking at the duration, magnitude, number and distribution of exceedances according to the threshold levels.

The analysis presented in this report is based on 1790 (out of 1900 hydrological stations that the HDCC currently collects. This is due to the fact that only stations that actively delivered data throughout the entire year 2020 and that had a stable data provision to the HDCC before January 1 2020 were selected. Out of these 11790 stations, 456 deliver exclusively discharge data, 402 only water level data and 932 stations provide discharge and water level data. Figure 2.2 shows the geographical distribution of those stations.

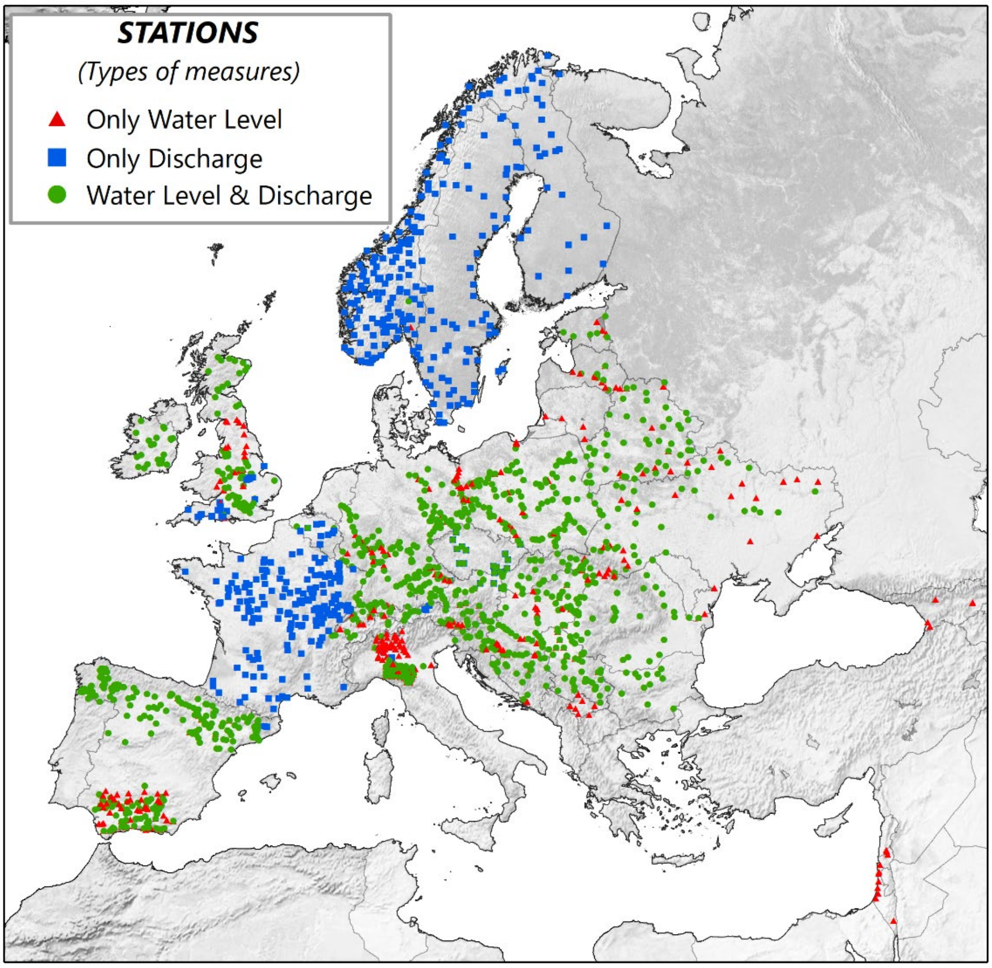


Figure 2.2: Spatial distribution of the 1790 selected stations and variables measured.

In addition to the information presented in this report, the HDCC has analysed the spring floods in Scandinavia in 2020 from a hydrological point of view contributing to the detailed assessment of those floods, which has been carried out in a collaborative effort among all EFAS Centres. The full report on this assessment can be accessed under the EFAS website.

3 Hydrological conditions of EFAS gauging stations

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

Although the CEMS Hydrological Data Collection Centre (HDCC) collects water level and discharge values, the analyses in this chapter have been carried out on the 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 2020, as well as the percentiles of the year 2019 and the period 1991-2016 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 2020 to the reference periods in order to determine their variations

We 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. Any interpretation of the results presented in this section should consider this point.

3.1 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 2019 and 2020 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 1991-2016 period, only stations with at least two years of data were included. As a result, a total of 1337, 1165 and 999 stations were chosen for 2020, 2019 and 1991-2016, respectively.

Figure 3.1 (left) shows the spatial distribution of the hydrological gauging stations chosen for this analysis, including the length of their historical time series. More than 50% 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 with long historical time series (such as Norway, Sweden, the Ebro River basin in Spain, and stations across the Rhine and Danube river basins).

Figure 3.1 (right) shows the upstream areas of all the selected stations. Many of the stations from the Scandinavian peninsula, Spain, England and across the Elbe river basin have small catchment areas (< 250 km2), whereas many of the stations from the Danube, Vistula, Ebro and Rhine river basins hold large upstream areas (>1000 km2). The distribution of catchment areas of the stations is partly a result of hydrological features, and partly a result of where hydrological services want to observe and which of the observations they are willing to share. We have normalized the discharge values with the upstream area to get a normalised discharge, as this index allows comparisons between stations. Nevertheless, differences in catchment areas is still likely to have an effect on 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/(m2· year)].

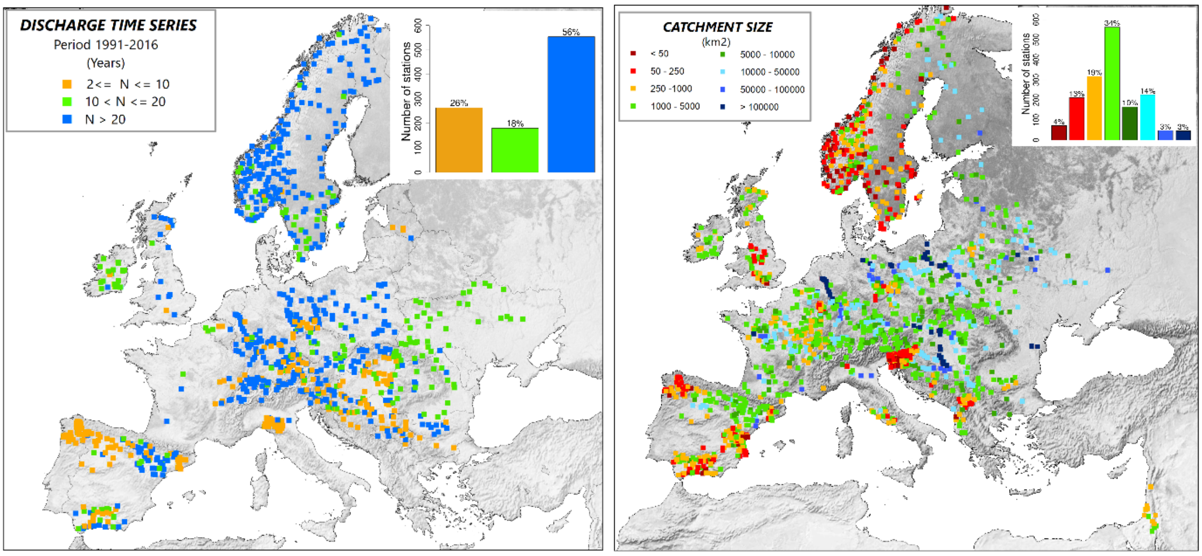


Figure 3.1: Spatial distribution of stations according to the length of their historical time series (left) and catchment size (right).

3.2 Hydrological conditions in 2020

Figure 3.2 shows the normalized mean discharge values for 2020. 15% of the studied stations present values below 100 mm/year. These are mostly located 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 over exploited streams. The highest values (over 1000 mm/year) occur for stations in Norway, Scotland, Ireland, the upper Rhine and Danube basins and usually occur in relatively small catchments with high precipitation.

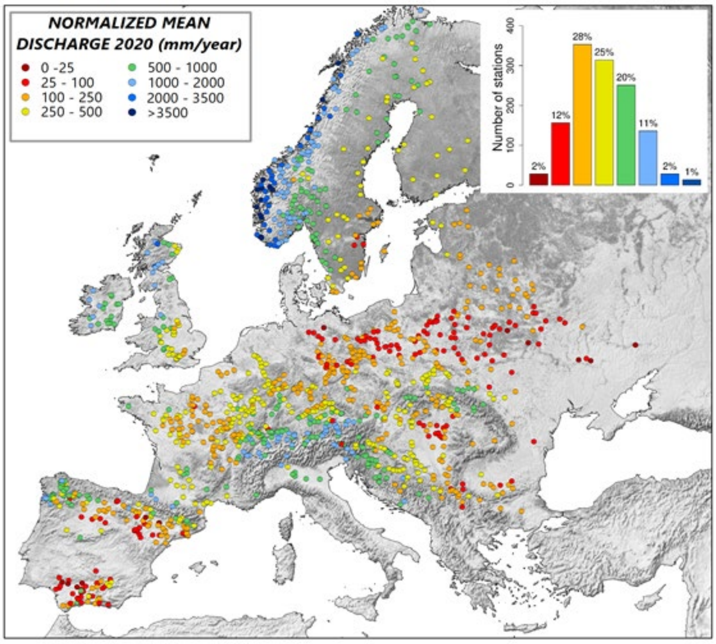


Figure 3.2: Spatial distribution of normalized mean discharge values in 2020

3.2.1 Comparative analysis

In this section the hydrological situation of 2020 is compared to the previous year (2019) and to the historical reference period (1991-2016). This is to assess if and how the hydrological conditions of 2020 differ from the past. The comparison of the relative variation of the average values is done through two indexes: the Streamflow Variation Index (SVI) and the Normalized Variation index (NVI).

SVI is applied when comparing 2020 and the period 1991-2016. It is adapted from the Streamflow Drought Index (SDI) (Nalbantis and Tsakiris, 2009):

$$SVI_H = \frac{\bar{X}_{2020} - \bar{X}_H}{S_H}$$

X_{2020} and X_H are the mean discharges for 2020 and 1991-2016, respectively. SH is the standard deviations of the annual mean discharge for the period 1991-2016. This index is a standardization of annual mean discharge in 2020 according to the annual mean and the standard deviation of the annual mean discharge in the period 1991-2016.

The Normalized Variation index (NVI) is applied when comparing the 2020 and 2019 mean discharges as the SVI is not applicable when the reference period covers only one year:

$$NVI_H = \frac{\bar{X}_{2020} - \bar{X}_{2019}}{\bar{X}_{2020} + \bar{X}_{2019}}$$

Where X_{2020} and X_{2019} are the mean discharges for 2020 and 2019 respectively.

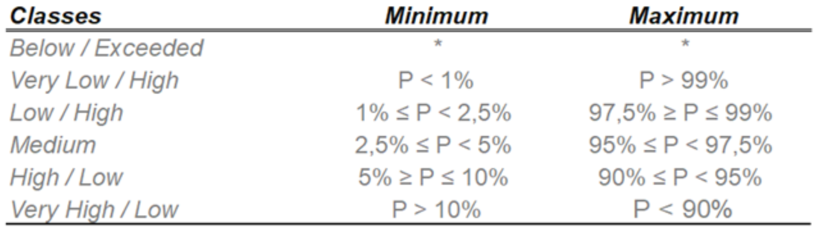
Table 3.3 defines quality classes based on the distribution of the resulting SVI and NVI values.

Table 3.3: SVI and NVI classes

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 2020 are calculated according to the time series of daily mean values from 2019 and the period 1991-2016 respectively. These percentiles are used to indicate how close the minimum and maximum river flows of 2020 are to the minimum and maximum for those periods. The extreme values of 2020 are then classified according to their percentile in the periods 2019 and 1991-2016. The percentiles intervals are shown in the Table 3.4.

Table 3.4: 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 values greater than the maximum. Separated classes have been added for such extremes.*

3.2.2 Variation of hydrological conditions

The spatial distribution of Normalized Variation Index for annual averages between 2020 and 2019, Figure 3.3 (left), shows clearly a dominance of low variations, both positive and negative, in stations across Europe. Stations with the lowest annual mean discharge for 2020 compared to 2019 are mostly located in Spain in Guadalquivir, Minho-Sil and Ebro river basins. This situation also occurs in some stations in Sweden, Oder river basin in Poland, across Danube river basin in Hungary, Germany, Romania and Bosnia and Herzegovina, and Dnieper river basin in Ukraine and Belarus. On the other hand, the stations that registered the highest increases of discharge in 2020 compared to 2019 are located in Llobregat, Ebro Guadalquivir and Douro river basins in Spain, Oder river in Western Poland, Danube river basin in Czech Republic, Slovakia, Hungary and Ukraine and Norway. There are also a few stations with high increases in Southern England, Sweden, Finland and Dniester and Dnieper river basins in Ukraine. In summary, most of the stations Europe had a similar annual mean discharge in 2020 to what they had in 2019.

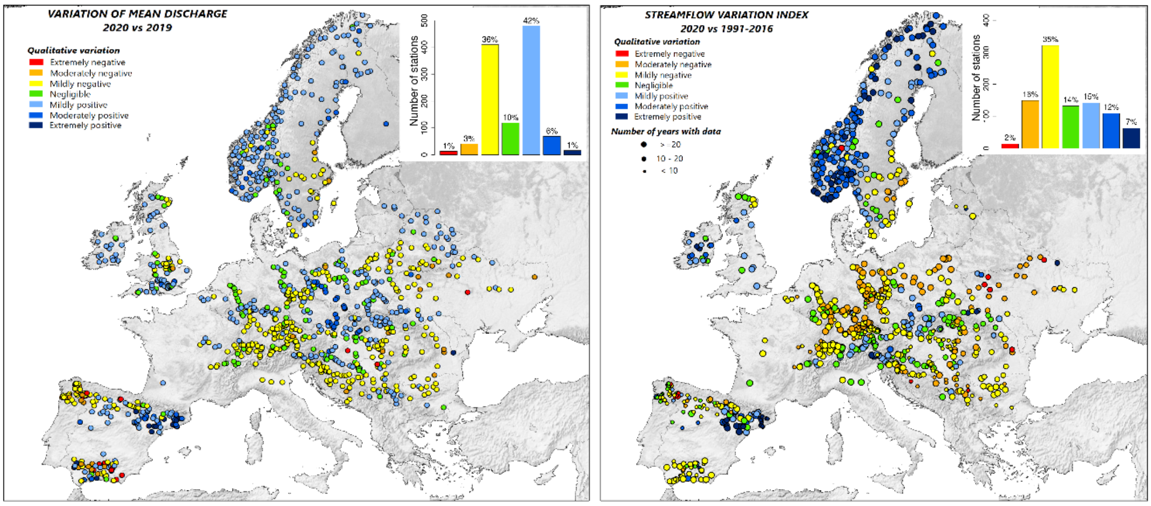


Figure 3.3: Spatial distribution of Streamflow Draught Index in 2020 with respect to 2019 (left) and the period 1991-2016 (right)

When comparing 2020 with the 1991- 2016 period, an increased number of stations with lower mean discharge is notable. However, 14% of the stations have negligible variations. These are mostly located in the Danube river basin, Sweden and basins in Northern Spain. 18% suffer moderate or extreme negative variations. Most of those stations are located in Rhine, Danube, Elbe, Oder, Vistula and Dnieper river basins. A number of stations in the Rhone river basin, South-Eastern Sweden and Norway show a moderate drought as well. On the other hand, 19% of the stations present a severe or moderate surplus of mean discharge in 2020 compared with the period 1991-2016. They are mostly located in basins in Norway and Northern Sweden and Finland, and North-

Western Spain, but they can also be found in Ireland and isolated stations of the Dnieper, Danube and Oder river basins.

In summary, we can say that 2020 was slightly dryer compared to the historical data but also slightly more rainy than 2019.

3.2.3 Minimum and maximum value analysis

In 2020, 37% of the stations recorded minimum mean daily discharge values that were lower than the ones in 2019 (or the river flow was zero), as it's shown in Figure 3.4 (left). We can see that these stations are found all across Europe but the concentration was higher in the Rhine, Rhone, Southern Danube river basin, British Isles and Southern Scandinavian Peninsula. On the other hand, around 23% of the stations recorded minimum mean daily values in 2020 that were considerably higher than the minimum values in 2019. This mainly occurred in stations located basins in Eastern Europe (Danube, Elba, Oder, Vistula and Dnieper), Western Norway and North-Eastern Spain. High minimum values were also found in basins of England, Southern Spain, Sweden, Finland and Ireland.

The minimum mean values in 2020 are predominantly higher than the ones in period 1991-2016. We found that only 13% of the stations recorded a lower minimum value than in the reference period (or the river flow was zero) (Figure 3.4, right). Most of these stations are located in the Elbe, Oder, and Vistula basins and South-Eastern Sweden. We also found a number of these stations in basins of Spain (Llobregat Douro and Minho), Norway, Ireland, England and France. Contrastingly, 25% of the stations had discharge minimum values considerably higher than the minima in the historical period. This mostly occurred in basins in Norway, Northern Sweden and Finland, across the Danube river basin, Ebro, Guadalquivir and Mediterranean basins in Spain, and isolated stations in Ireland, England and Daugava and Dnieper river basins. The minimum values of the rest of the stations are almost equally distributed according the different degrees of closeness to the minimum for the period 1991-2016.

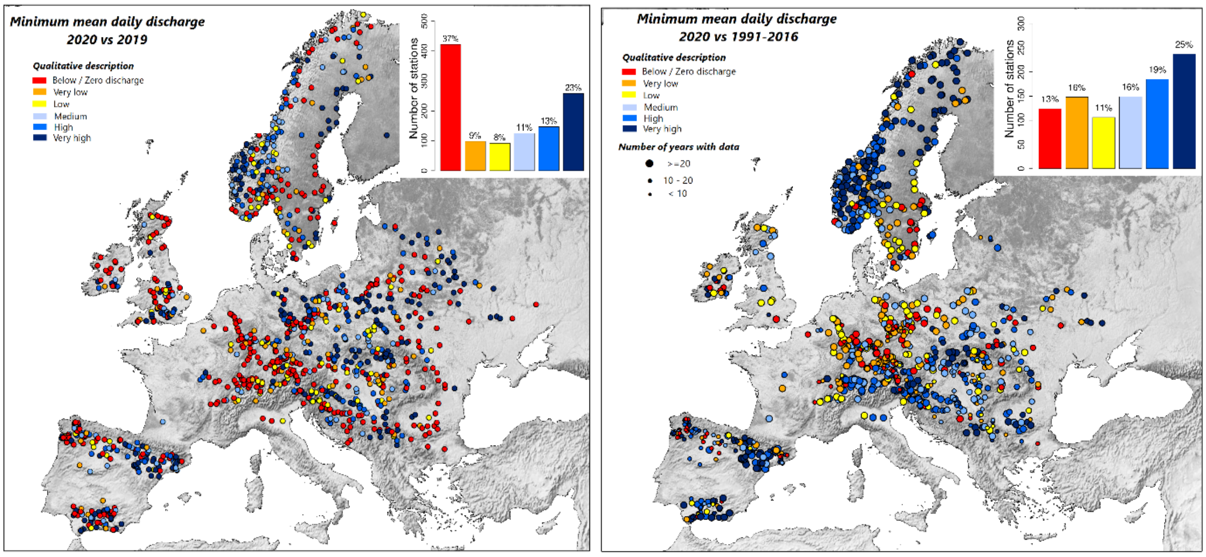


Figure 3.4: Spatial distribution of stations and the minimum values in 2020 with respect to 2019 (left) and the period 1991-2016 (right).

Figure 3.5 (left) shows a comparison of the maximum mean daily discharge for 2019 and 2020 and show that the maximum values were higher in 2020 for 47% of the stations across Europe. However, around 11% of the stations recorded maximum mean daily values considerably below the maximum value in 2019. These stations are mainly located in the Vistula, Dnieper and Neman river basins. Considerably lower extremes also occurred more locally for some stations in Danube, Oder, Rhine, Guadalquivir, Welland rivers and Eastern Sweden. Between the high and low maximum values, we find 5% of the stations that recorded lower maximum discharge in 2020 than in 2019. These are found in Spain, Belarus and some stations in the Elbe, Rhine and Dniester river basins and Sweden and Finland as well.

Figure 3.5 (right) shows that 49% of the stations across Europe recorded maximum values for 2020 that were just below their historic maxima from the period 1991-2016. Moreover, 11% of the stations exceeded in 2020 the maximum mean daily value of the period 1991-2016. These exceedances took place in stations of Scandinavian Peninsula, Spain, and Danube river basin. There were also exceedances at stations located in the Rhine, Dniester river basins, Ireland and Scotland. On the other hand around 16% of the station recorded maximum mean daily values in 2020 considerably below the maximum historical values. These station are mainly located in the Elbe, Oder, Vistula, Dnieper, Danube river basins, Spain and isolated stations in Sweden.

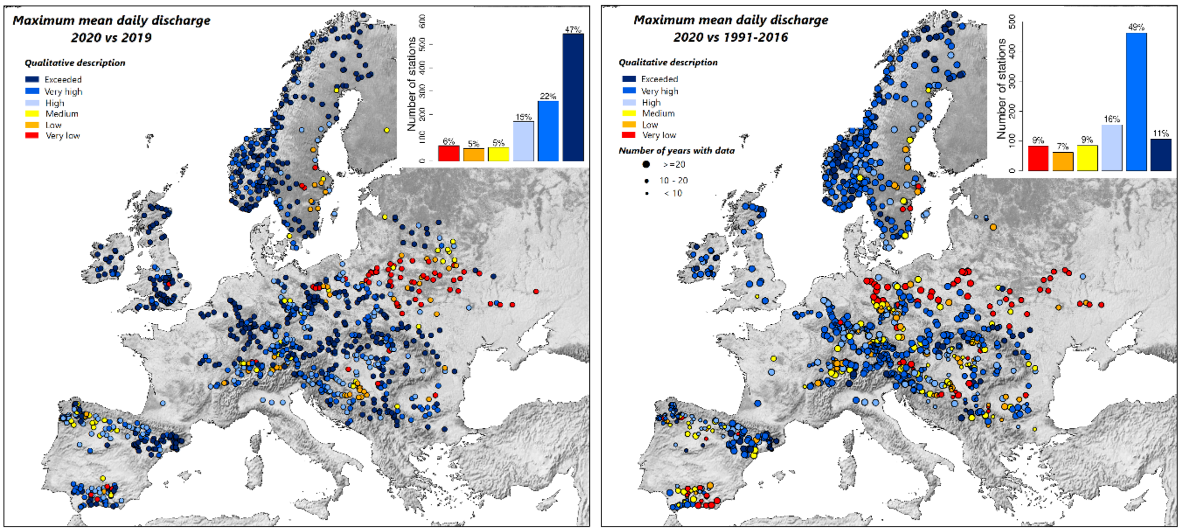


Figure 3.5: Spatial distribution of stations and the maximum values in 2020 compared with 2019 (left) and the historical period 1991-2016 (right).

4 Gaps Analysis on the CEMS hydrological data base

4.1 Initial considerations

This chapter analyses the gaps in the CEMS hydrological data collection for the year 2020, collected by the Hydrological Data Collection Centre (HDCC).

The CEMS hydrological data collection is continuously growing with hydrological data from 1,990 gauging stations across Europe. The data observation frequency among those vary from every minute to daily (see Figure 4.1). 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.

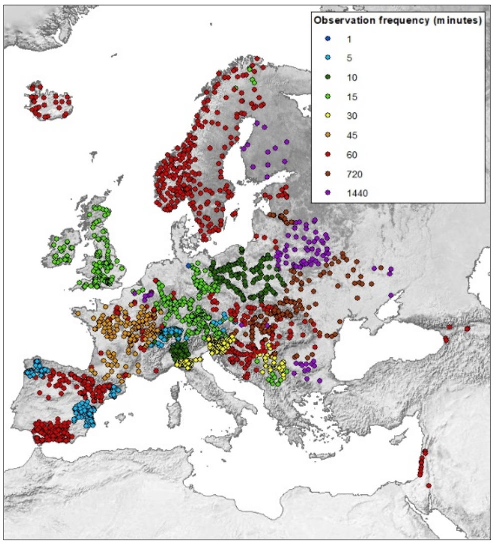


Figure 4.1: Observation frequency by station.

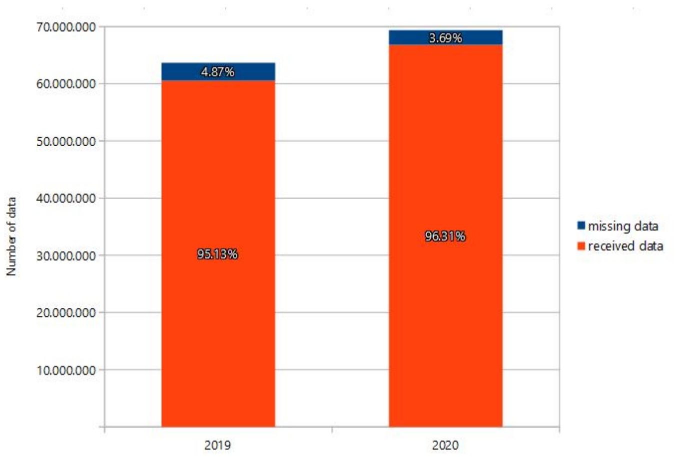


Figure 4.2: Reception rate comparison between 2019 and 2020.

4.2 Gap analysis

We analysed data from 1,790 stations providing water level and/or discharge values (see Introduction), from 43 data providers (DPs). 1,603 of these stations had problems with data transmission between January 1st 2020 and December 31st 2020 on at least one occasion.

In total 3.69% of all the data values expected for 2020 were not received, which is less in percent and absolute number compared to 2019 (Figure 4.2). 99% of all the 525,936 gaps lasted less than 1 day and 72% 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 147,407, coming from 1,568 stations and for 2,407 variables. This number is higher than the number of stations, as each station can provide up to two variables (water level and/or discharge values).

4.2.1 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 4.3 (left panel) shows the number of gaps according to their duration. 96.4% of the gaps have a duration of ≤ 1 day, resulting mostly from changes in the data observation frequencies and/or delays in data transmissions. 2.9% last between 1 and 3 days, whereas 0.6% (961 gaps) lasted more than three days and required a follow up by the HDCC. Figure 4.3 (right panel) shows the distribution of those gaps longer than three days.

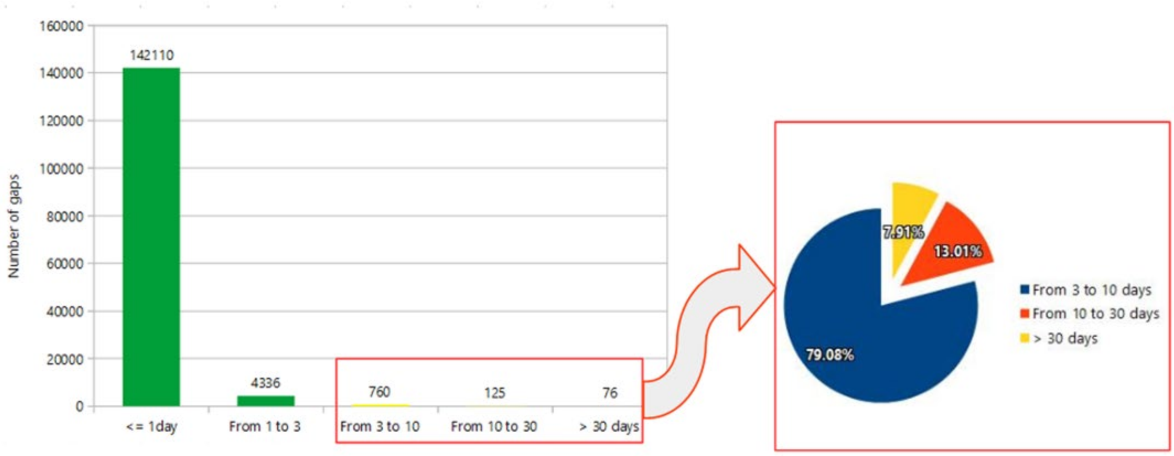


Figure 4.3: Number of gaps by duration (left panel) and distribution in percentage of those longer than three days (right panel).

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.

4.2.2 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 5 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 4.4 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, nearly a third with data from the DP and nearly a 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.

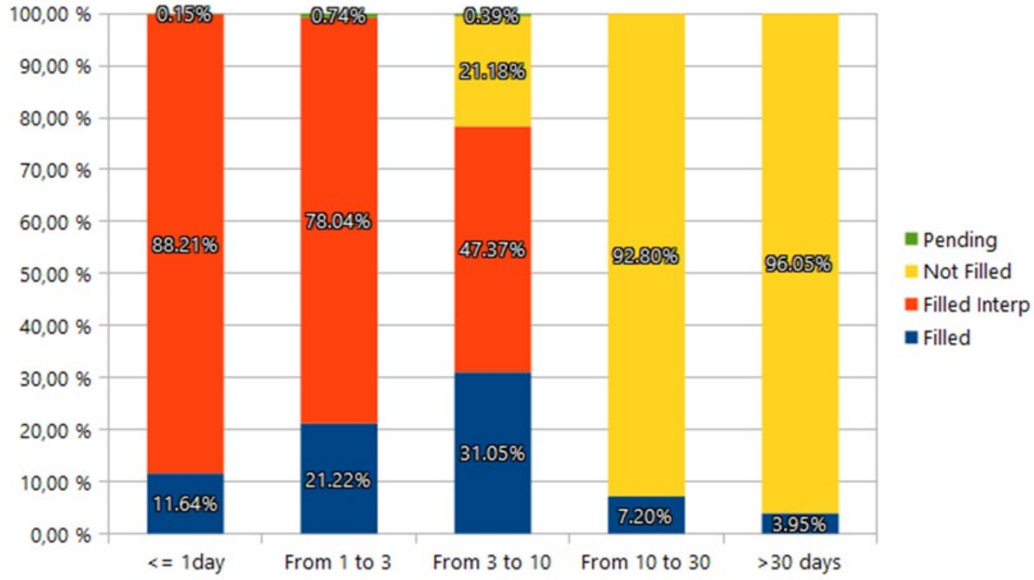


Figure 4.4: Percentage of gap status by gap length.

4.2.3 Other aspects to be considered

The 147,407 gaps analysed add up to 1,888,869 missing values covering a total of 39,137 accumulated days. The average length per gap is 0.26 days (about 6 hours), whereas the average number of gaps per station and variable is about 61.2; hence an average of 16 days of gaps for each data variable.

Figure 4.5 presents monthly boxplots with the percentage of received data against expected data for each data provider. The mean value ranges between 94.7 and 97.2 %, although some data providers provide lower ratios.

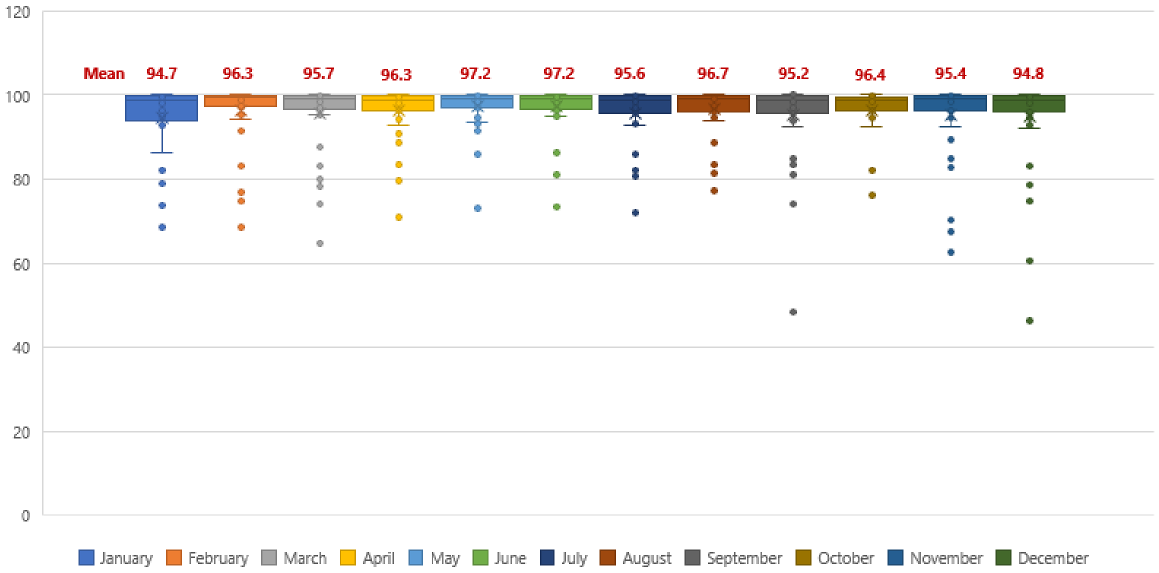


Figure 4.5: Box plot showing the monthly percentage of data received, out of the expected number of data records to be delivered from EFAS data providers.

When comparing these values to 2019 the average percentage for 2020 is slightly higher (96 against 92.4 %) although for two data providers there was a lower than 50% reception rate in September and December.

The maps in Figure 4.6 and Figure 4.7 show the spatial distribution of gaps, with respect to the average gap duration (days) and maximum gap length (days), respectively.

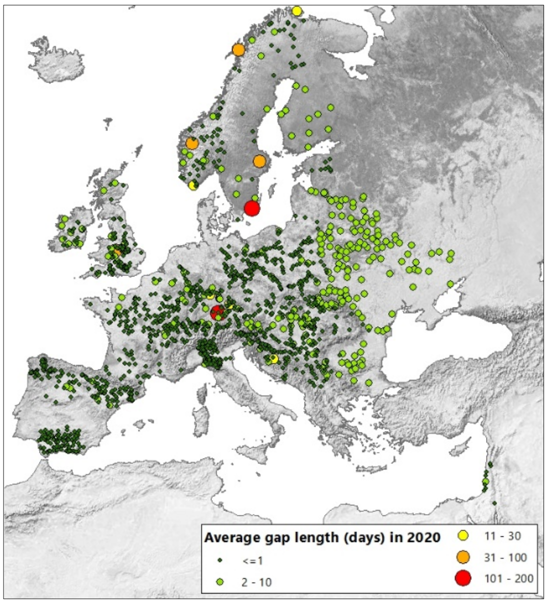


Figure 4.6: Average gap length in days per station.

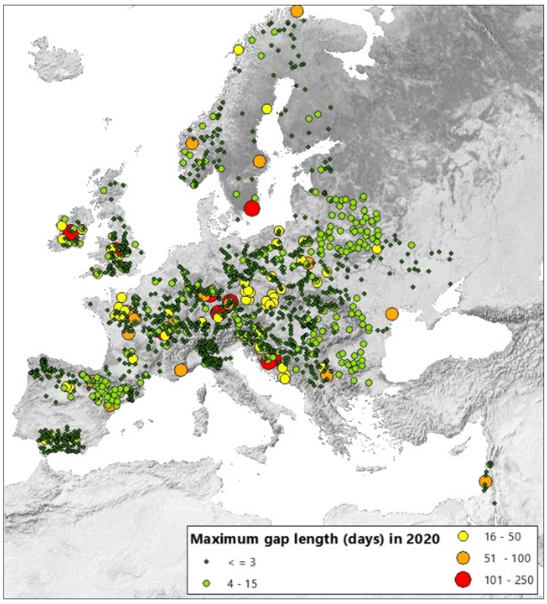
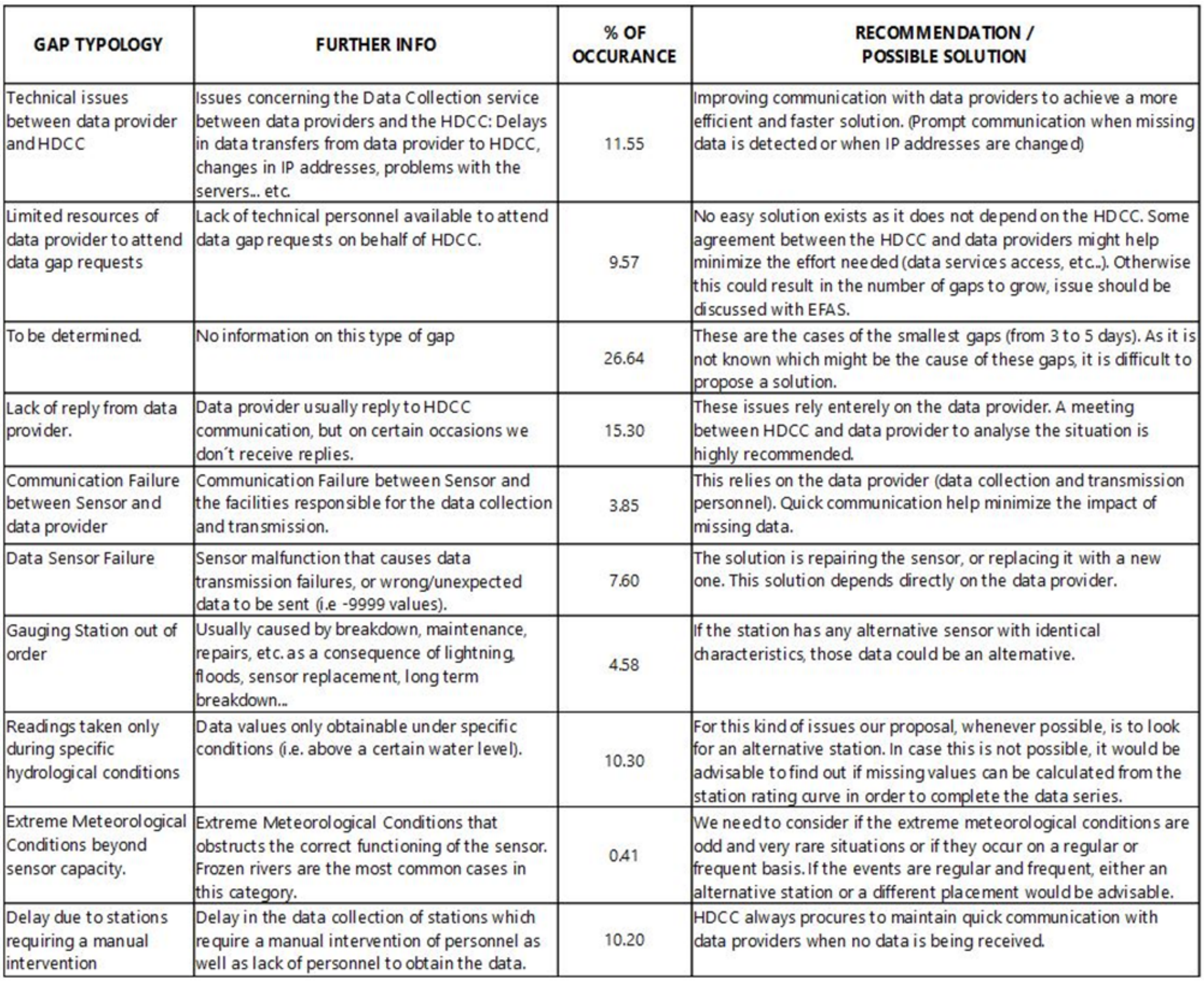


Figure 4.7: Maximum gap length in days per station.

4.2.4 Gap typology and proposal for future data collection strategy

In only 961 cases (0.6% 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 4.5). This classification helps to develop and propose a series of measures to improve the data collection strategy both quantitatively and qualitatively.

Table 4.5: Gap classification with possible solutions



GAP TYPOLOGY	FURTHER INFO	% OF OCCURANCE	RECOMMENDATION / POSSIBLE SOLUTION
Technical issues between data provider and HDCC	Issues concerning the Data Collection service between data providers and the HDCC: Delays in data transfers from data provider to HDCC, changes in IP addresses, problems with the servers... etc.	11.55	Improving communication with data providers to achieve a more efficient and faster solution. (Prompt communication when missing data is detected or when IP addresses are changed)
Limited resources of data provider to attend data gap requests	Lack of technical personnel available to attend data gap requests on behalf of HDCC.	9.57	No easy solution exists as it does not 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.
To be determined.	No information on this type of gap	26.64	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.
Lack of reply from data provider.	Data provider usually reply to HDCC communication, but on certain occasions we don´t receive replies.	15.30	These issues rely enterely on the data provider. A meeting between HDCC and data provider to analyse the situation is highly recommended.
Communication Failure between Sensor and data provider	Communication Failure between Sensor and the facilities responsible for the data collection and transmission.	3.85	This relies on the data provider (data collection and transmission personnel). Quick communication help minimize the impact of missing data.
Data Sensor Failure	Sensor malfunction that causes data transmission failures, or wrong/unexpected data to be sent (i.e -9999 values).	7.60	The solution is repairing the sensor, or replacing it with a new one. This solution depends directly on the data provider.
Gauging Station out of order	Usually caused by breakdown, maintenance, repairs, etc. as a consequence of lightning, floods, sensor replacement, long term breakdown...	4.58	If the station has any alternative sensor with identical characteristics, those data could be an alternative.
Readings taken only during specific hydrological conditions	Data values only obtainable under specific conditions (i.e. above a certain water level).	10.30	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.
Extreme Meteorological Conditions beyond sensor capacity.	Extreme Meteorological Conditions that obstructs the correct functioning of the sensor. Frozen rivers are the most common cases in this category.	0.41	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.	10.20	HDCC always procures to maintain quick communication with data providers when no data is being received.

Figure 4.8 shows the number of gaps for each category classified by gap duration.

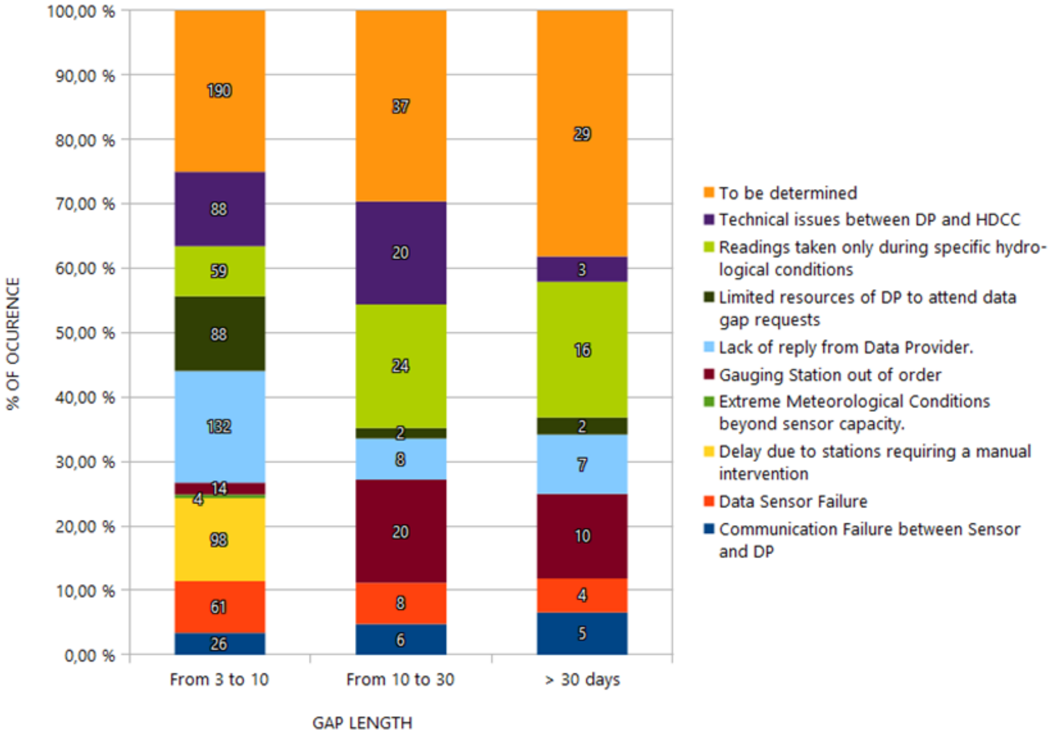


Figure 4.8: Number/percentage of gaps by/per duration and typology.

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).

4.3 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 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 89,377 outliers were detected in data from 375 stations out of the 1,790 stations studied in this report. Considering that the total number of values received is 66,831,520, the rate of outliers is approximately 0.13 %.

Figure 4.9 illustrates the different types of outliers according to their aggregation and frequency.

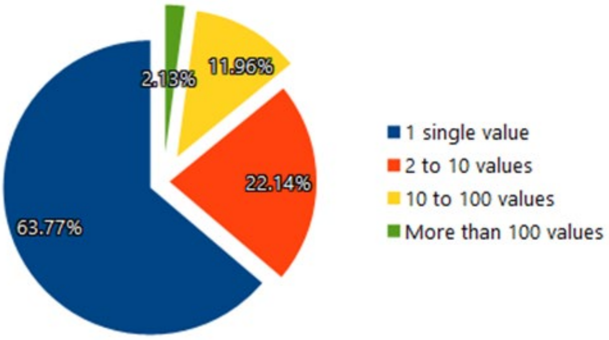


Figure 4.9: Sets of outliers and their frequencies.

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

The following figures show the stations that registered outliers in 2020, the total outlier's duration in days per station and the rate of outliers relative to received data for each station.

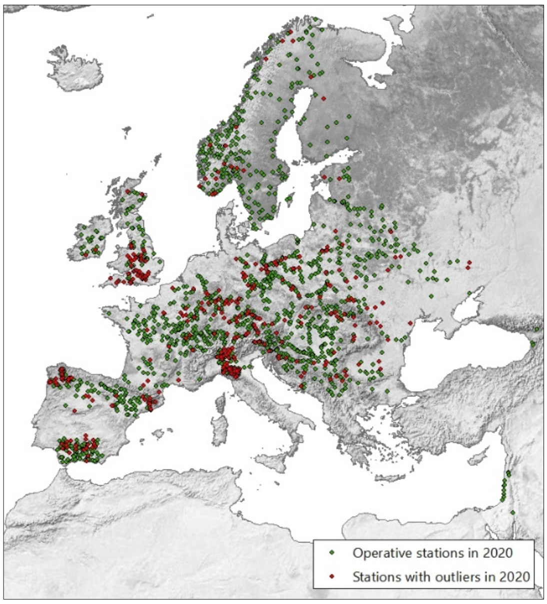


Figure 4.10: Stations that registered outliers in 2020.

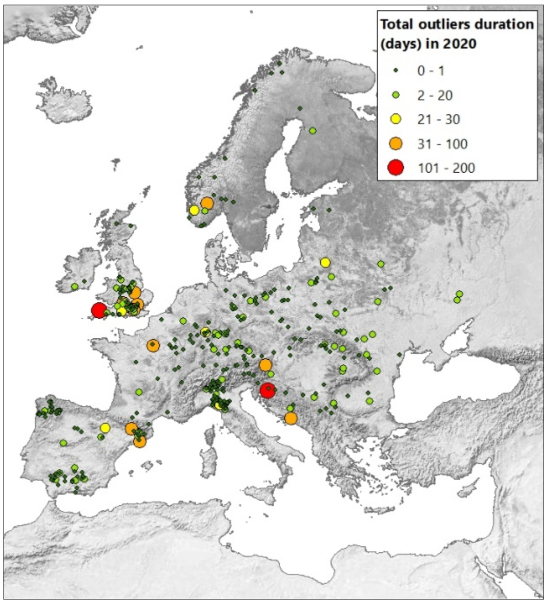


Figure 4.11: Total outliers duration in days in 2020.

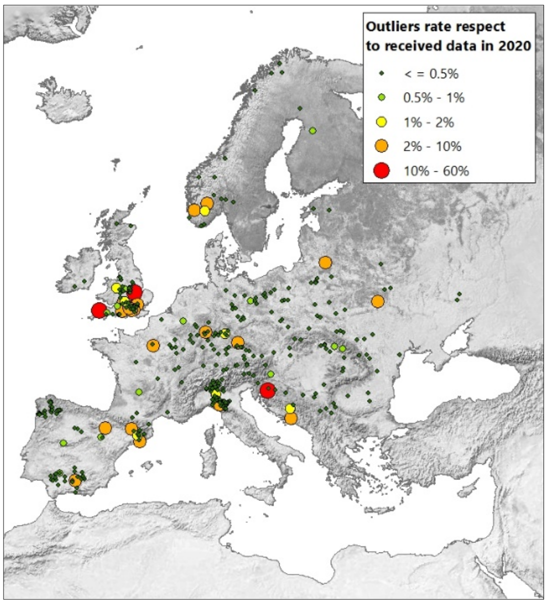


Figure 4.12: Percentage of outliers occurrence relative to the total amount of data received per station in 2020.

5 Analysis of Exceedance events

In this section the hydrological stations that exceeded their threshold level during 2020 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.

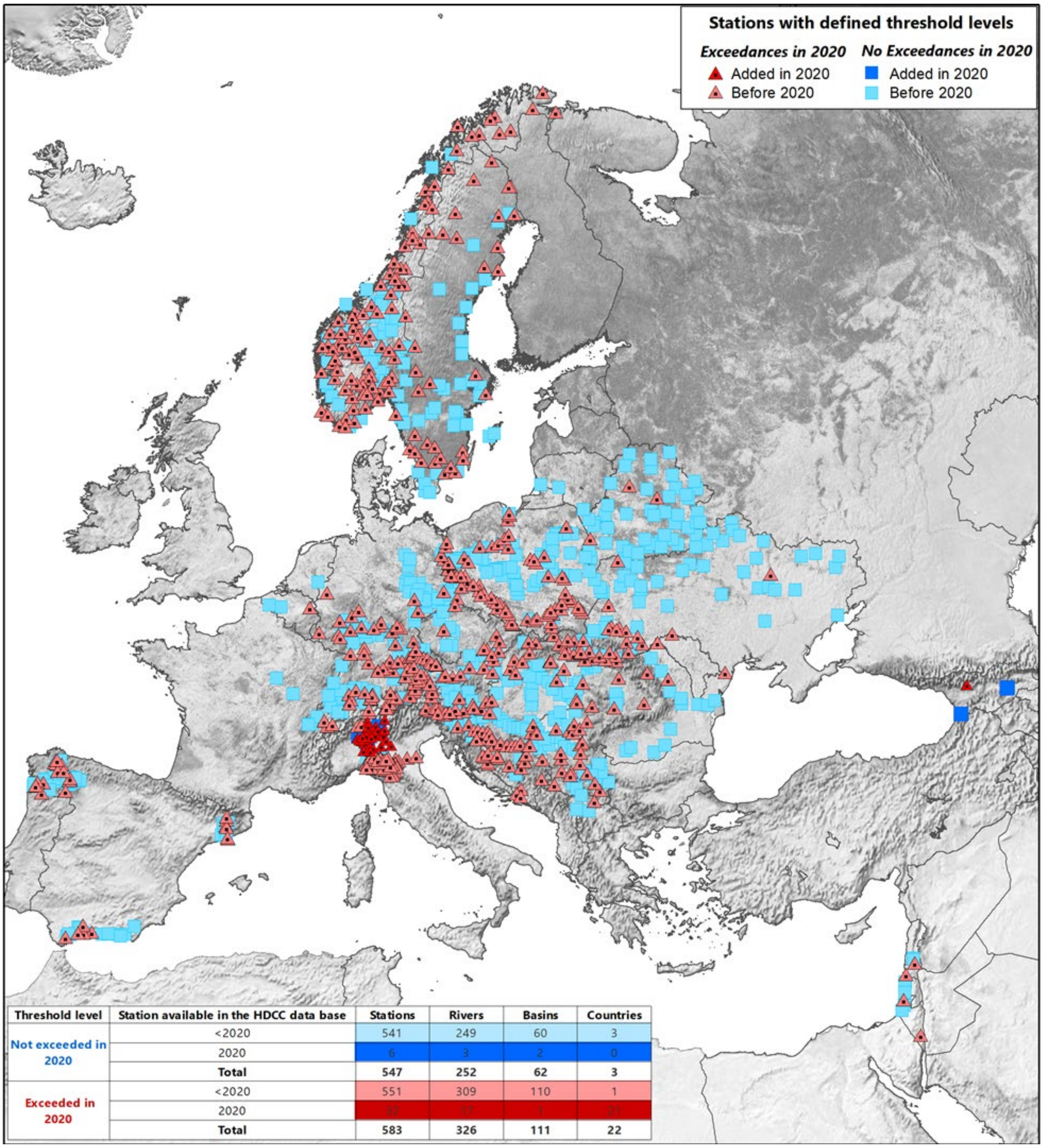
All near real-time observations (water level and discharge) are displayed on the EFAS website in the "National flood monitoring" layer. Where available, also the national/regional threshold levels are shown and exceedances of those are highlighted by the HDCC.

5.1 General description

Out of the 1790 active stations initially selected for this report, threshold levels are available for 1130 stations (63%) (light and dark blue stations in Figure 5.1). Compared to 2019, the number of stations with at least one threshold level has increased by 38 and these stations now cover a total of 578 rivers, 173 basins and 25 countries (see table in Figure 5.1), rising the number of rivers, basins and countries by 20, 3 and 1 respectively, in 2020.

The new country that has been included in 2020 into the EFAS stations threshold level monitoring system is Georgia, with four stations located in the Rioni, Chorokhi and Kura basins.

In addition, in Italy, the Po basin has increased its number of stations with threshold level, monitoring 16 new rivers: Adda, Arno, Bevera, Brembo, Chiese, Eupilio, Lambro, Liro, Lura, Mallero, Mella, Mincio, Oglio, Olona, Serio and Seveso.



Threshold level	Station available in the HDCC data base	Stations	Rivers	Basins	Countries
Not exceeded in 2020	<2020	541	249	60	3
	2020	6	3	2	0
	Total	**547**	**252**	**62**	**3**
Exceeded in 2020	<2020	551	309	110	1
	2020	32	17	1	21
	Total	**583**	**326**	**111**	**22**

Figure 5.1: Stations with no exceeded threshold levels in 2020 (in blue tones), differentiating between stations already existed in 2019 and stations added in 2020. Stations with threshold levels exceeded in 2020 in red tones (stations existed in 2019 and stations added in 2020). The table shows a summary of threshold levels (exceeded and no exceeded) in 2020 by spatial aggregation levels (station, river, basin and country), providing the total number of stations, rivers, basins and countries in 2020 (all).

Red triangles in Figure 5.1 represent stations that had threshold levels exceedances; 583 stations (52%) had at least one of their threshold level exceeded in 2020 (84% of the stations added in 2020). This covers 56% of European rivers and 64% of European basins. Out of 25 countries that share data, only 3 (Belgium, France and Lithuania) did not register any threshold exceedances. However, these countries have very few stations with threshold levels (2, 3 and 3 respectively).

5.2 Duration of Exceedances

5.2.1 Duration per station

Figure 5.2 shows the number of events and their total duration per station. A total of 2,167 exceedances were recorded during 2020 at 583 stations, a little over three quarters of the events in 2019 (2747 exceedances at 552 stations).

In 2020 most stations recorded between 1 and 7 events. The average number of events per station has decreased from 5 in 2019 to 4 in 2020, also the average accumulated duration of the events per station has also diminished from 7.6 days in 2019 to 6.2 days in 2020.

For 80% of the stations, the accumulated duration of all events lasted less than 7 days. Longer accumulated durations (between 13 and 209 days) were found among stations in the Danube, Po, Neman, Helge a, Dnieper, Oder, Morrumsan, Neretva, Ronneby, Rhine, Alsteran, Nattrabyan, Ume, Torne and Vistula basins (see Figure 5.2).

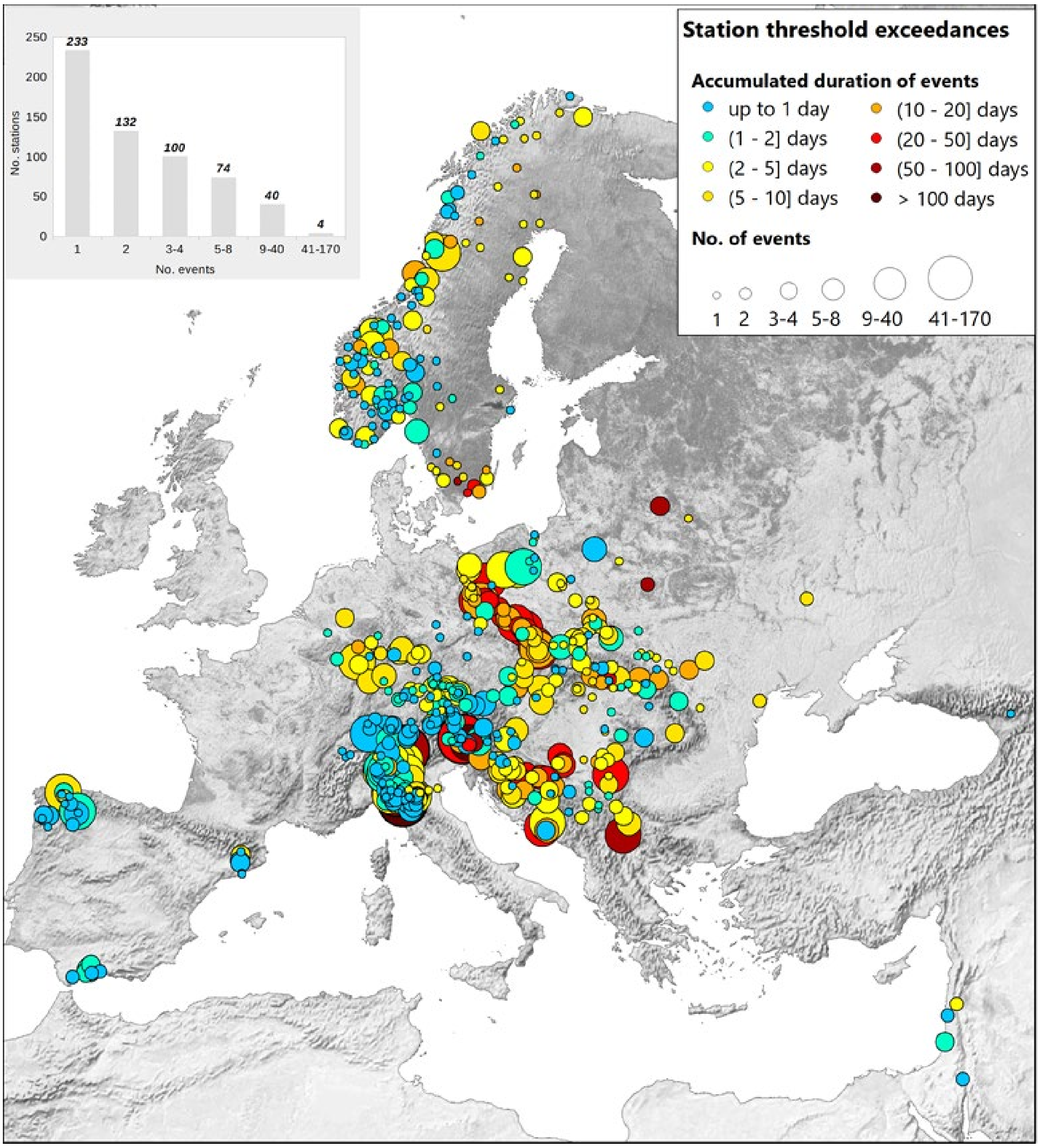


Figure 5.2: Station classification according to number of events (symbol size) and total accumulated duration of exceedance events (colour coding). Bar chart on the upper left corner shows the number of stations per event frequency.

Figure 5.3 shows the average event duration per station (2.6 days), which is very similar to the average of 2019 (2.9 days). For 70% of the stations the average duration is less than 2 days. These stations are mainly located in Norway, Italy, Germany, Spain, Switzerland, and Austria and all the stations in Slovenia, Israel, Netherlands and Georgia had average events lower than 2 days.

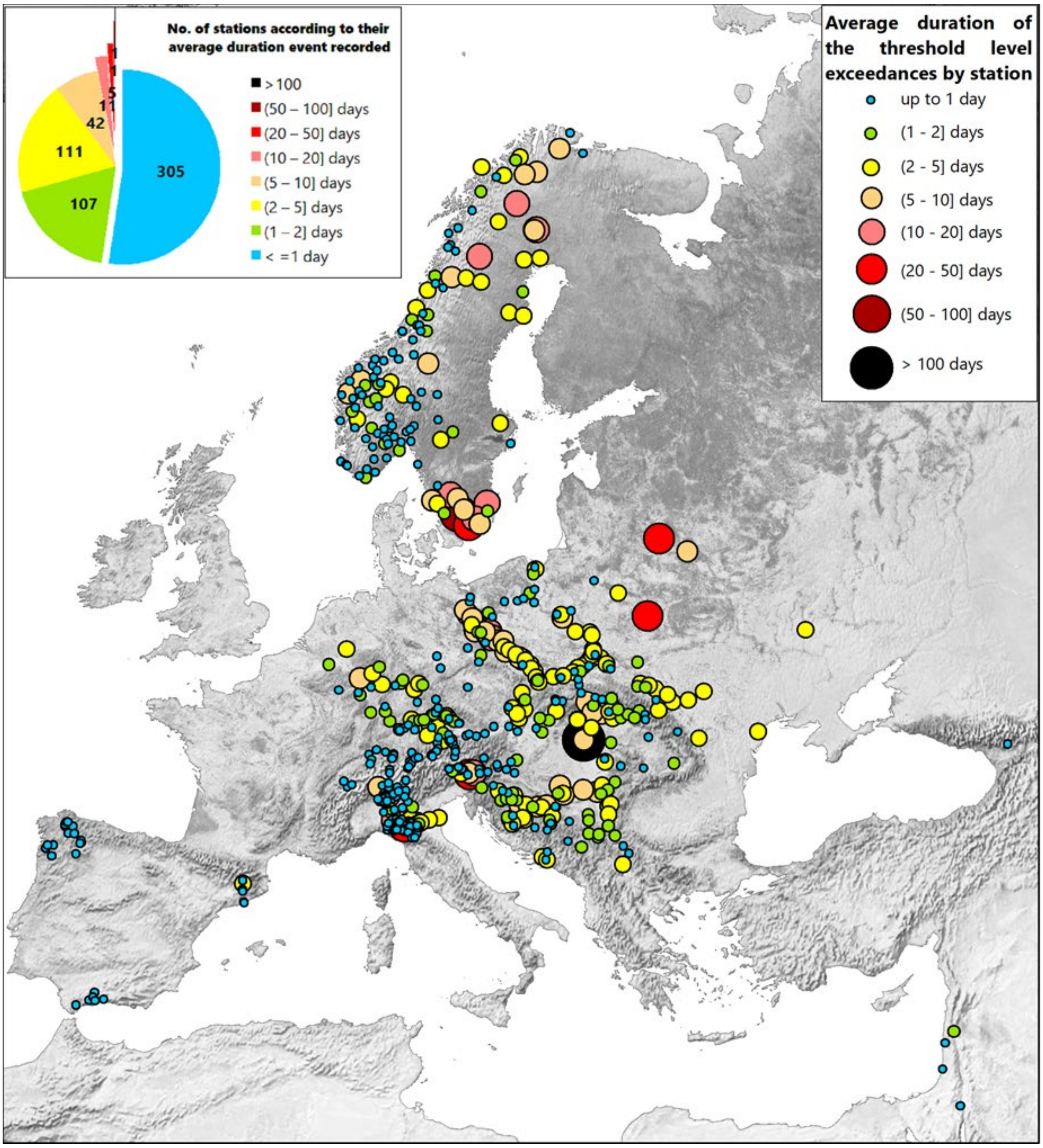


Figure 5.3: Average duration of events for EFAS stations in 2020 and number of stations according to the average duration event recorded in each station (pie chart).

On the other side, the longest average duration (10 to 209 days) were recorded at isolated stations across:

- The Danube basin, on the Tisza River through Hungary (209 days), Lakes Faak (22 days) and Worth (16 days) in Austria and Latorica river (17 days), in Ukraine.
- The Swedish basins, on the rivers: Helge a (65 days), Morrumsan (47 days), Alsteran (19 days), Muonio (16 days), Tjeggelvas (13 days), Idijoki (12 days), Ronneby (11 days) and Nissan (11 days).

- The Po basin (Italy), on the Parma and Riglio rivers with a duration of 43 and 14 days.
- The Dnieper basin, on the Stokhid river in Ukraine (30 days).
- The Neman basin, on the Neris river in Belarus (27 days).
- The Rhine basin, on the Untersee river in Switzerland (24 days).
- The Oder river basin in Poland (two stations with 14 and 12 days).

5.2.2 Duration of events

Despite the fact that there have been less exceedance events in 2020 than in 2019 and the average accumulated duration per station has decreased slightly, the average duration in 2020 considering all the events without group by station is 1.7 days, a bit higher than the average in 2019. 75% of the exceedance events in 2020 lasted less than 1.3 days, very similar to the 2019 figures (1.1 days). The remaining 25% of the events are distributed as follows: 15% of the events lasted between 1.3 days and 3.6 days, 5% lasted from 3.6 to 7.1 days and 5% lasted from 7.1 to 209 days. Stations with short events often have more frequent events, which explains the difference between the average of all events here and the average length for each station above.

Events lasting less than 1.3 days were the most common and occurred across 75% of the rivers. Out of the 326 rivers with exceedances, the ones with the largest number of short events are located on the Cedra, Stirone, Serio, Lambro, Seveso, Oglio, Secchia, Adda, Taro, Lura and Brembo rivers, in the Po basin (Italy), on the Oder and Notec rivers, in the Oder Basin through Czech Republic and Poland, on the Lake Millstatt, Inn and Danube river, in the Danube Basin (Austria), on the Brda and Vistula rivers, in the Vistula basin (Poland) and on the Burbia river in the Minho basin (Spain). The vast majority of these short events took place in June, during the later summer and in the early autumn, reaching the highest number of events in October. The fewest events were registered in May and July.

The longest 21 events (over 20 days) were located across:

- Danube basin, on the Austrian Lakes Faak, Ossiach and Worth and on the rivers Latorica (Ukraine), Binacka Morava (Kosovo) and Tizsa (Hungary).
- Po basin (Italy) on the Parma and Riglio rivers.
- Oder basin, on the Oder river (Poland).
- Swedish basins: Helge, Morrumsan and Ronneby rivers.
- Neman basin: Neris river (Belarus)
- Neretva basin: Trebizat river (Bosnia and Herzegovina)
- Dnieper basin: Stokhid river (Ukraine)

These longest events are distributed throughout the year as follows: 50% started in October, 25% in winter and only the event located in the Stokhid river began in the summer months. This last event does correspond to a real flood event, however, a case by case analysis showed that not all threshold level exceedance events correspond to actual flood events, but rather with relatively low first national threshold levels. This seems to be the case for Tisza River in the Danube basin, which holds the longest event record since this analysis has been carried out, surpassing its threshold level for 209 days in 2020 and the case for some of the longest events in the Po basin.

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 176 high level events in 2020, distributed across 93 stations that exceeded 1) or 2). Figure 5.4 shows the spatial distribution and duration of the high level events for 2020.

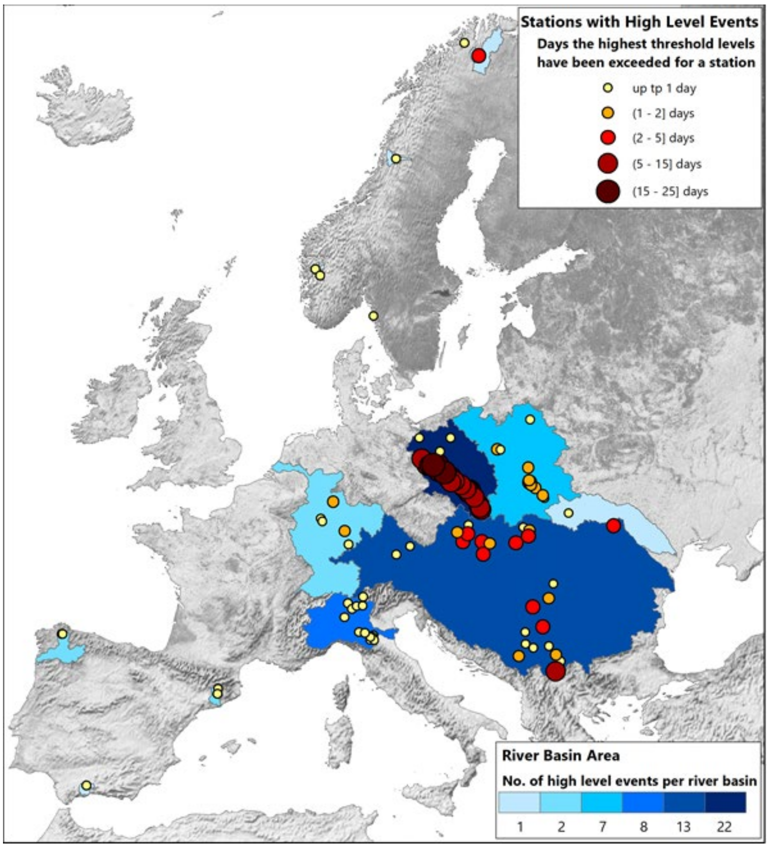


Figure 5.4: Duration of the high level events and river basins where the stations are located.

Nearly 90% of these stations are located across the following basins: Danube (30%), Oder (26%), Po (15%), Vistula (11%), and Rhine (5%). The remaining 10% are found in the Scandinavian basins Kinso, Leirbotn, Tana, Vefsna, Vosso and Anråsa, in the Spanish basins Minho, Guadalhorce and Llobregat and in the Dniester basin, in Ukraine. Most of these basins had high level events also in 2019 (Po, Danube, Vistula, Oder, Minho, Guadalhorce, Rhine and Anråsa).

Although the total number of exceedance events decreased in 2020 compared to 2019, the number of high level events almost doubled in 2020 (with 8% of all events; 171 events) compared to 2019 (see Figure 5.5).

The average duration of these high level events is nearly 2 days and the vast majority of these longest events (over 5 days) occurred in stations on the Eastern Neisse and Oder rivers, in the Oder basin, principally because of floods in western Poland after heavy rain fall in October.

Figure 5.5 illustrates the number of high level events per month and comparing years 2020 with 2019.

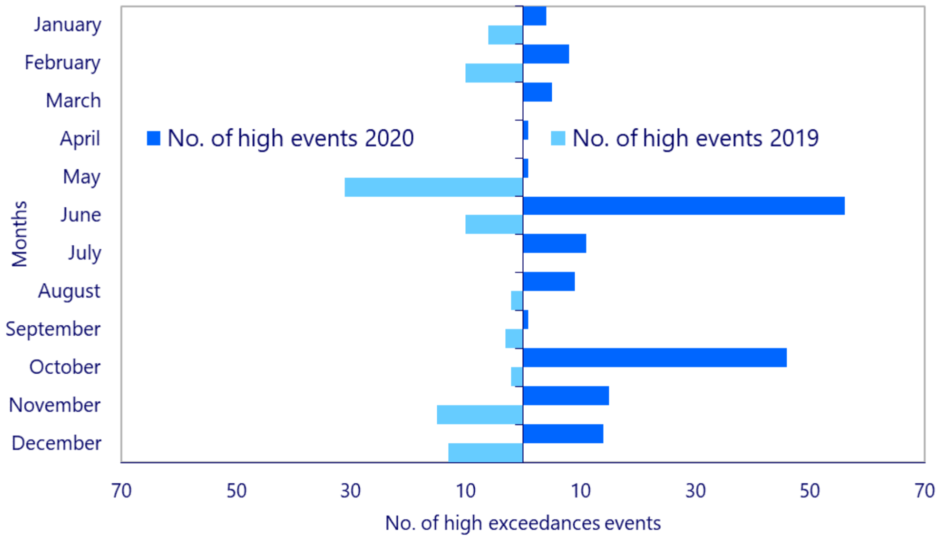


Figure 5.5: Number of high level events per month in 2020 and 2019.

In 2020, most events happened in autumn: October was the month with the largest number of events, while in 2019 nearly 70% of the events were distributed between spring (mainly May) and autumn (November and December). Regarding high level events, autumn and spring stood out both years, the highest number was reached at the end of spring (June) in 2020 and in May in 2019.

On the contrary, there were almost any events in April of 2020 and only three high level events with very small duration in April, May and September. In 2019, no high level events occurred in the spring and almost none of them in the summer months.

6 Summary

During 2020, the CEMS Hydrological Data Collection Centre welcomed the “Icelandic Meteorological Office” and “Confederación Hidrográfica del Júcar” as two new hydrological data providers to the CEMS hydrological data collection network. Together with the “Slovenian Environment Agency”, who increased the number of stations providing real-time hydrological data to the HDCC, a total of 193 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 2020 to a total of 48 and 1990 respectively, with an increase of 4.3% and 11% compared to 2019.

In additional, Poland, Ukraine, Serbia, Ireland, Norway and Romania provided in 2020 new historic data sets. With a total of 12,357,356 new historical data values, the volume of the historical database has increased in 2020 by 10.9% compared to 2019.

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 2020 present some particularities that are worth mentioning:

- The water contribution in 2020 is higher compared to 2019, but lower than the historical average between 1991-2016. Especially in Vistula, Dnieper, Dniester, Elbe, Oder, Rhine river basins and South-Eastern Sweden the drier conditions were more noteworthy, when comparing to the historical period.
- The maximum and minimum mean daily values of discharge in 2020 followed a softer regime than 2019 in most of the stations, excluding the Vistula, Dnieper, Dniester, central area of the Danube river basin and stations in South-Eastern Sweden.
- When comparing the maxima in 2020 to the historical period (1991-2016) a number of stations in basins of Scandinavian Peninsula, northern Spain (Minho, Douro, Ebro and Llobregat) and Danube river basin exceeded the maximum mean daily discharge, together with other stations in the Rhine river basin and British Isles.
- The hydrological conditions in Elbe, Oder, Vistula, Dnieper river basins, South-Eastern Sweden, upper Rhine and lower Danube were was much drier compared to the historical average.

6.2 Gaps and Outliers

Regarding data gaps, the majority (72%) 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 5 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 2020 with 2019, we see that the rate of received data vs expected data has slightly increased in 2020 (96.31%) with respect to 2019 (95.13%). The number of gaps has decreased in 2020 with respect to the previous year (525,936 vs 605,961) even when the total number of received data has increased by 10%. The cause of data gaps was identified in 73% of the cases and solutions have been proposed accordingly. However, for the remaining 27% of the cases the causes remain unknown.

The analysis reveals that the percentage of outliers in 2020 is really low compared to the annual amount of data received (0.13%). Most outliers are isolated data values.

6.3 Exceedances Events

Threshold levels were available for 1130 stations and 25 countries. Since the beginning of 2020, the HDCC incorporated 38 new stations with threshold levels, covering 1 new country (Georgia), 20 new rivers and 3 basins. 52% of all stations had at least one of their threshold levels exceeded during 2020 and registered a total of 2167 exceedance events, a little over a quarter of the exceedances registered in 2019.

Number of events per station and duration of these events have both decreased in 2020.

The average number of events per station decreased from 5 events in 2019 to 4 events in 2020 and the average accumulated duration per station was reduced from 7.6 days in 2019 to 6.2 days in 2020.

The longest events (over 20 days) were located across Austria, Ukraine, Kosovo, Hungary, Italy, Poland, Sweden, Belarus, and Bosnia Herzegovina.

Although the total number of exceedance events decreased in 2020 compared to the year before, 2020 had twice the number of “high level" events as 2019 had: 8% of all the events observed in 2020 were "high level", and they were registered at 93 stations mainly located in the Danube, Oder, Po, Vistula, Rhine, Scandinavian basins (Kinso, Leirbotn, Tana, Vefsna, Vosso and Anråsa), Spanish basins (Minho, Guadalhorce and Llobregat) and Dniester basin. Most of these basins had high level events also in 2019 (Po, Danube, Vistula, Oder, Minho, Guadalhorce, Rhine and Anråsa).

Autumn was the season where most exceedance events occurred, followed by spring. On the contrary, there were almost any events in April and only three high level events with very small duration in April, May and September.

List of tables

Annex 1: Data provider list

Austria	• Federal Ministry of Agriculture, Forestry, Environment and Water Management
Belgium	• Hydrological Information Centre • Service public de Wallonie
Belarus	• Republican Emergency Management and Response Center of the Ministry of Emergency Situations of the Republic of Belarus
Bosnia and Herzegovina	• Federal Hydrometeorological Institute
Bulgaria	• National Institute of Meteorology and Hydrology
Croatia	• Meteorological and Hydrological Service of Croatia
Czech Republic	• Czech Hydro-Meteorological Institute
Estonia	• Estonian Environmental Agency
Finland	• Finnish Environment Institute
France	• Ministère de l'Ecologie et du Développement Durable Service Central d'Hydrométéorologie et d'Appui à la Prévision des Inondations
Georgia	• LEPL National Environmental Agency - Ministry of Environmental Protection and Agriculture of Georgia
Germany	• Bundesanstalt fuer Gewaesserkunde • Saxon State Agency for Environment and Geology • Hessisches Landesamt für Umwelt und Geologie • Landesamt für Umwelt, Wasserwirtschaft und Gewerbeaufsicht Rheinland - Pfalz • Landesamt für Umwelt, Gesundheit und Verbraucherschutz • Bayerisches Landesamt für Umwelt
Greece	• Hellenic National Meteorological service
Hungary	• Hungarian Hydrological Forecasting Service (OVSZ), General Directorate of Water Management (OVF)
Iceland	• Icelandic Metereological Office
Ireland	• Office of Public Works of Ireland
Italy	• Servizio Idro Meteo Clima Agenzia Regionale per la Protezione dell'Ambiente • Regione Lombardia • Agenzia Regionale per la Protezione dell'Ambiente • Regione Piamonte • Presidenza del Consiglio dei Ministri Dipartimento della Protezione Civile • Protezione Civile - Regione Lazio
Latvia	• Latvian Environment, Geology and Meteorology Centre
Lithuania	• Lithuania Hydrometereological Service
Luxembourg	• Administration de la gestion de l'eau
Montenegro	• Administration de la gestion de l'eau
Netherlands	• Rijkswaterstaat Institute for Inland Water Management and Waste Water Treatment

Norway	• Norwegian Water Resources and Energy Directorate, Hydrology Department
Poland	• Institute of Meteorology and Water Management Wroclaw Branch
Republic of Kosovo	• Kosovo Enviromental Protection Agency
Russian Federation	• Hydrometcenter of Russia
Romania	• Institutul National de Hidrologie Si • Gospodarire A Apelor
Serbia	• Republic Hydrometeorological Service of Serbia
Slovakia	• Slovak Hydrometeorological Institute
Slovenia	• Environmental Agency of the Republic of Slovenia
Spain	• Automatic System of Hydrological Information for the Ebro River Basin • Confederación Hidrográfica del Miño - Sil Confederación Hidrográfica del Duero • Confederación Hidrográfica del Guadalquivir • Government of Andalusia - Regional Ministry of Agriculture, Livestock, Fisheries and Sustainable Development • Confederación Hidrográfica del Júcar • Catalan Water Agency
Sweden	• Swedish Meteorological and Hydrological Institute, core services department
Switzerland	• Federal Office for the Environment
Ukraine	• State Emergency Service of Ukraine Ukrainian Hydrometeorological Center
United Kingdom	• UK Met Office - Flood Forecasting Centre • Scottish Environment Protection Agency Departament of Infrastructure

JRC126323

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
ec.europa.eu/jrc

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

Publications Office
of the European Union