



# Copernicus Emergency Management Service

## New data validation rules for hydrological data

**17th EFAS ANNUAL MEETING**

Presented by Rafael García

26 09 2022



Emergency  
Management

## HDCC: new rules for hydrological data

### Hydrological Data Collection Centre main tasks:

- Near real-time data collection
- Historical data collection
- Harmonisation
- Post processing → quality checks
- Storage
- Data sharing



Emergency  
Management

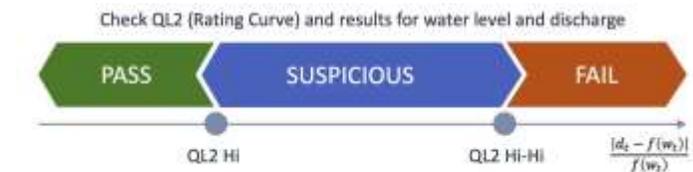
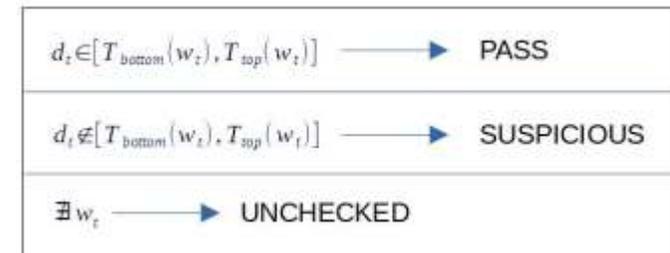
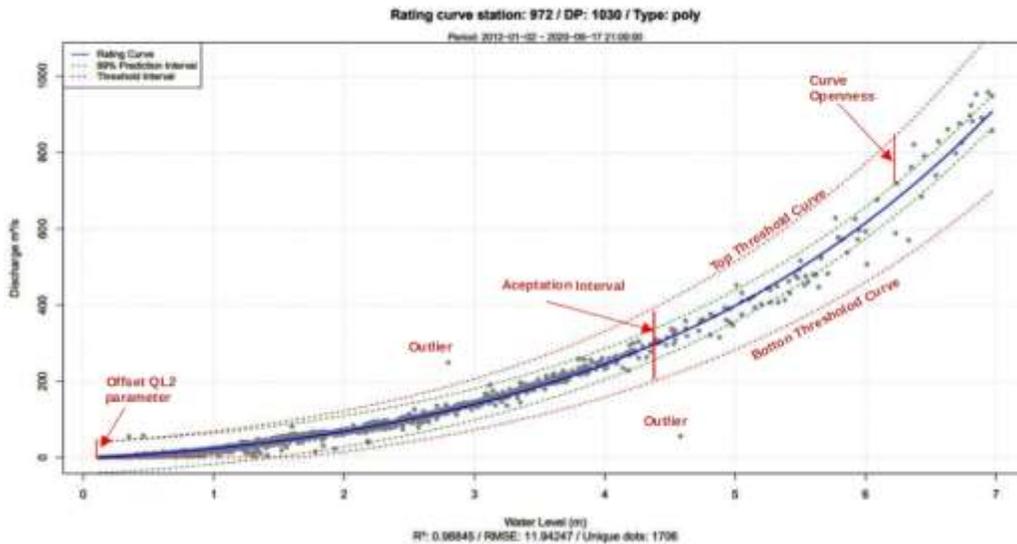
## HDCC: new rules for hydrological data - Background

- Near real-time data considered as raw data
  - Gaps
  - Outliers
  - Odd values
  - ...
- Historical data: reviewed and postprocessed by the data owner/provider (gap filling, error correction...)
- Checks are applied to arise possible bad data.
  - Nevertheless, results do not provide direct conclusions → Visual check and also further developments would be needed



# HDCC: new rules for hydrological data – rules in the previous version

- Range Control (HiHi, Hi, Lo, LoLo)  
Range limits are defined using the metadata provided by the station owner (max/min historic values, mainly)
- Rating Curve (for discharge)



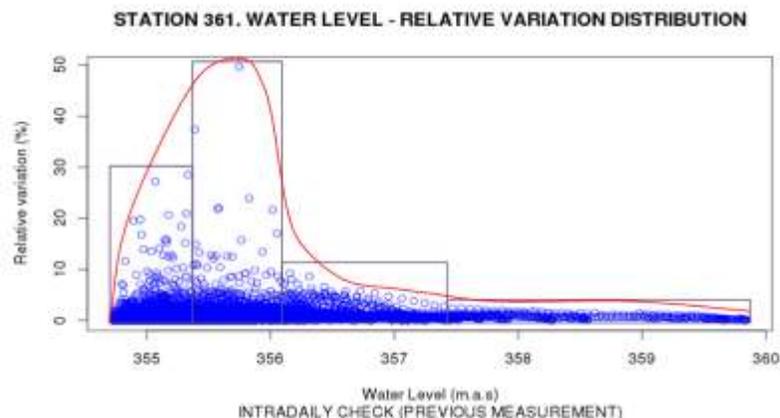


- Relative variation checks

1. From the previous value
2. From the value of the previous day at the same time

Clustering of relative differences using the time series stored in the HYDRO DB.

Classes are defined for data ranges. If relative difference is greater than the limit for the class where the observation is, it is flagged as suspicious





- Repetitive values

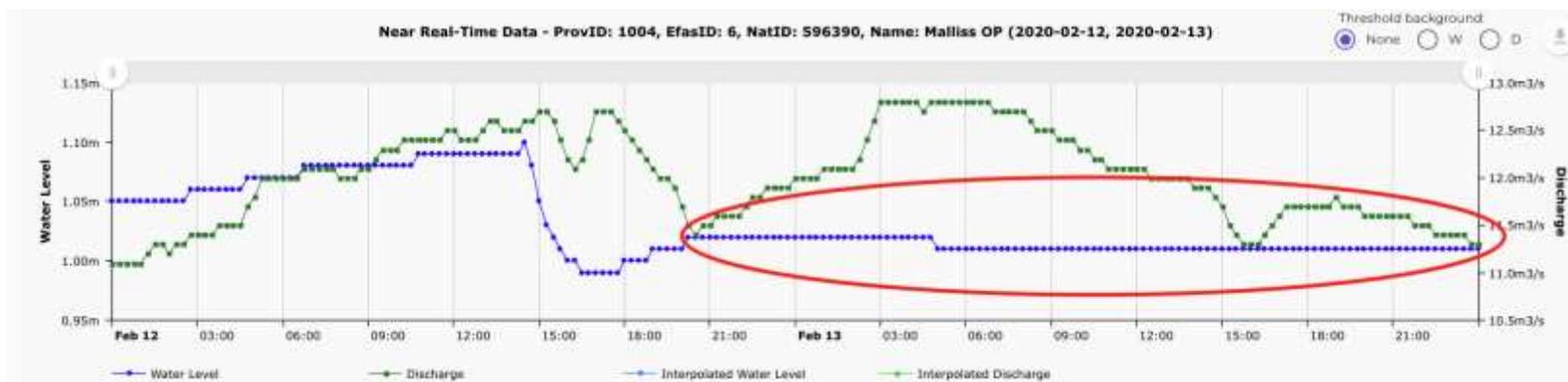
It is useful for detecting malfunctioning sensors for any reason.

Criteria depends on the reporting interval:

- For reporting interval  $\geq 12$  hours, for more than 3 days with the same value, observations are flagged as SUSPICIOUS

- For reporting interval  $< 12$  hours, for more than 1 day with the same value, observations are flagged as FAIL

- Otherwise, observations are flagged as PASS





Emergency  
Management

## HDCC: new rules for hydrological data – new rules

- Negative discharges

Negative discharge values are not always errors, as some stations close to reservoirs or dams can produce negative discharge. In the other cases, negative discharges are errors and must be discarded.

When the stations is not flagged as *has backwater*, this check is applied and observations of discharge with negative values are flagged as FAIL.

- Monthly min/max

Each observation value is compared with the maximum and minimum values for that month. Those observation out of range are flagged as SUSPICIOUS.

- Visual check

By regular inspection, hydrologists can flag observations as FAIL values.



## Checks vs variables

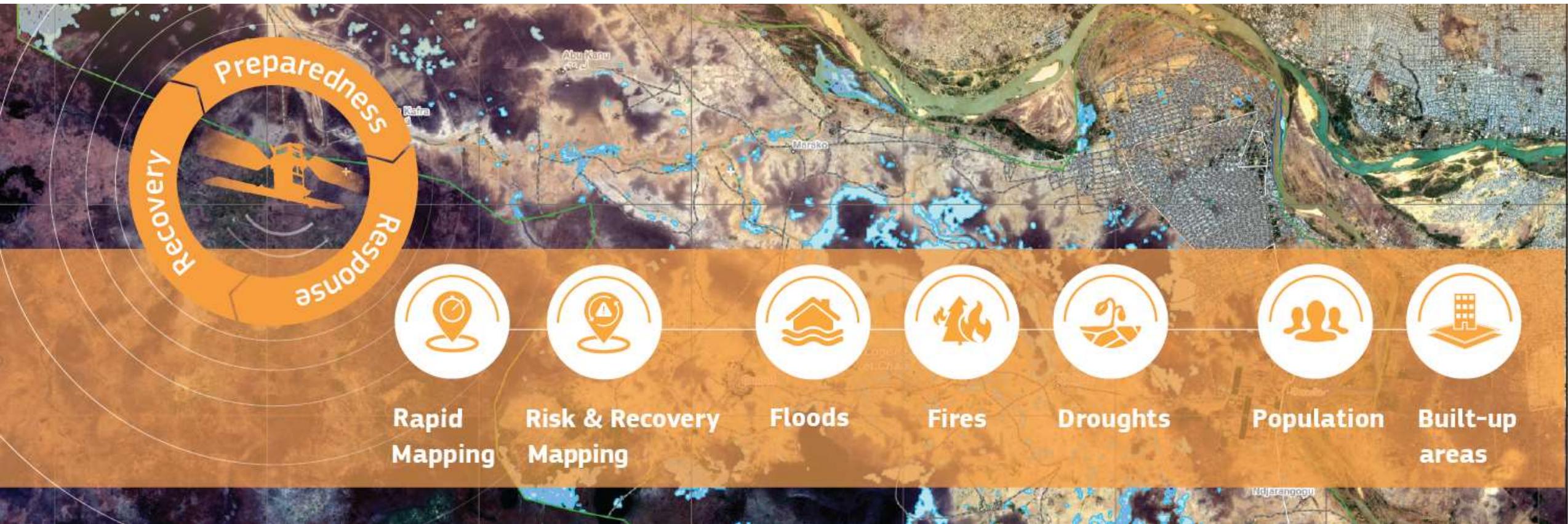
QUALITY CHECK	Water Level	Discharge	Reservoir Outflow	Reservoir Level	Volume
Range Control	Yes	Yes	Yes	Yes	Yes
Rating Curve	No	Yes	No	No	No
Variation Previous Day	Yes	Yes	Yes	Yes	Yes
Variation Previous value	Yes	Yes	Yes	Yes	Yes
Repetitive values	Yes	Yes	No	No	No
Negative discharges	No	Yes	No	No	No
Monthly Min-Max	Yes	Yes	Yes	Yes	Yes
Visual Check	Yes	Yes	Yes	Yes	Yes



Emergency  
Management

## HDCC: new rules for hydrological data

- Most checks do not produce definitive conclusions on their own.
- Some of them produce better results for certain circumstances. For example, Rain Control check is not useful for repetitive values but it is useful for outliers.
- Next step: using the existing time series in the HYDRO database, apply analytics techniques to find the best combination of checks results to establish an unique quality flag for the observation with a high level of confidence for detecting outliers or odd values.
- Meanwhile, results of the checks help hydrologists to identify bad values.



preparedness

Recovery

Response



Rapid Mapping



Risk & Recovery Mapping



Floods



Fires



Droughts



Population



Built-up areas

UdJatenggo