waterschap limburg

Royal Netherlands Meteorological Institute
Ministry of Infrastructure and Water Management

Interreg
Euregio Meuse-Rhine
EM Flood Resilience

EUROPEAN UNION
European Regional Development Fund

Deliverable D.T1.2.2

Central rain gauge data and quality control

The EMfloodResilience project is being carried out within the context of Interreg V-A Euregio MeuseRhine and is 90% funded from the European Regional Development Fund.

Authors
Mats Veldhuizen (KNMI)
Aart Overeem (KNMI)
Ely Deckers (KNMI)

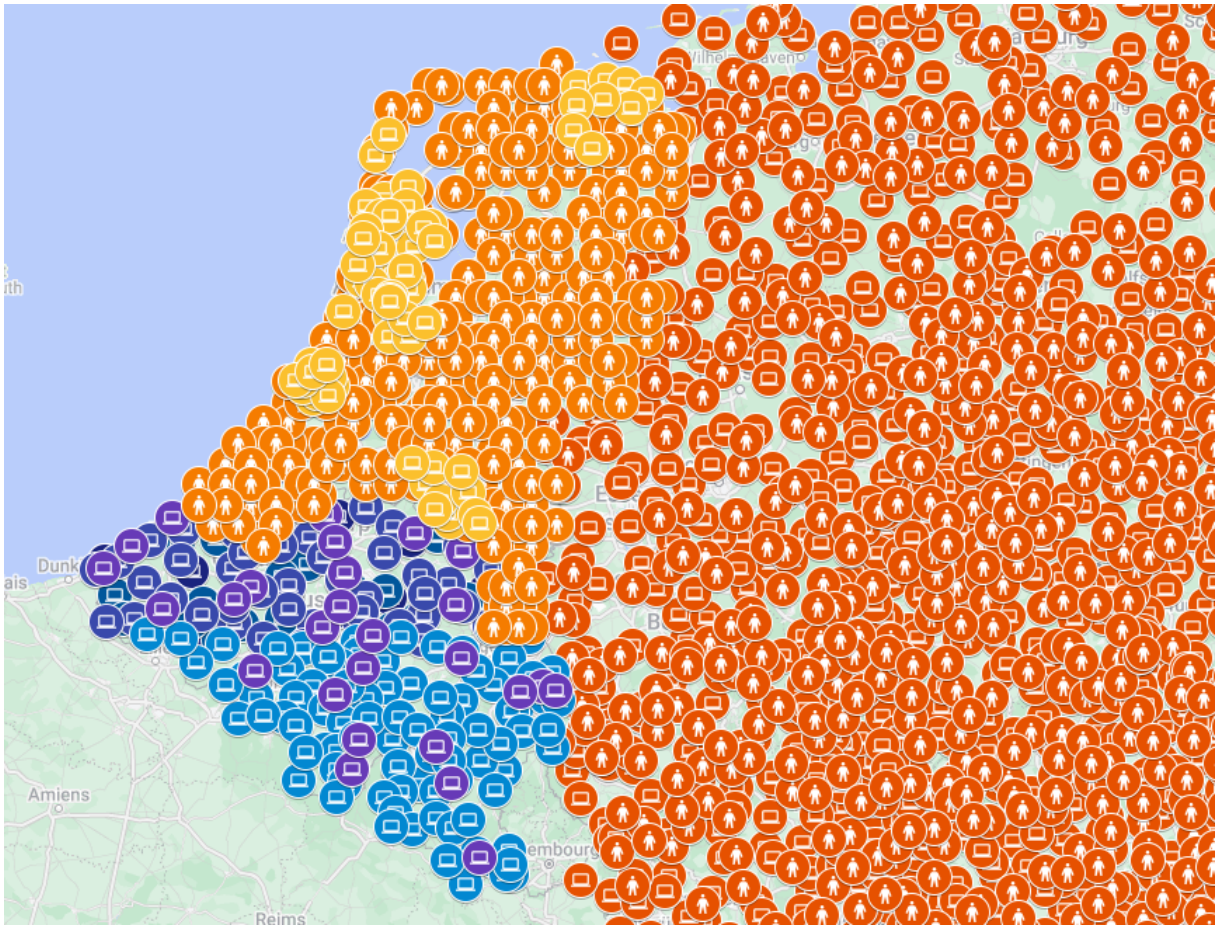
TABLE OF CONTENTS

Introduction

For this report we have gathered access to and information about rain gauge data, originating from rain gauges in Germany, Belgium and the Netherlands. Rain gauge data is important for many applications. Here, we focus on the availability of rain gauge data for the adjustment of radar precipitation accumulations. Typically, national meteorological services only use rain gauge data from their own nationwide networks to improve radar precipitation products. The (near) real-time version of these products are used as input for nowcasting and are hence important for early warnings, such as the extreme rainfall that occurred in the Belgian-Dutch-German border region mid-July 2021. The inventory of rain gauge datasets in this study is therefore important to improve radar precipitation products from all national weather services active in the Interreg region of this project, namely the Royal Netherlands Meteorological Institute (KNMI) for the Netherlands, the Royal Meteorological Institute (RMI) for Belgium, and the Deutscher Wetterdienst (DWD) for Germany. For every rain gauge dataset we have access to, we tried to answer the following questions:

- What is the timeliness of this data? I.e., how fast after the measurement is taken does it arrive and what is the temporal resolution of this data?
- What is the spatial resolution of this dataset? I.e., how many rain gauges does it contain inside the IRC products area?
- What quality control is performed on this dataset? Is this dataset later republished with a higher quality level after additional quality control is performed?
- What is the usability of this dataset for the IRC products?

The IRC products mentioned above are KNMI's operational quantitative precipitation products, colloquially labelled as International Radar Composite (IRC) for external users, e.g., in water management. These are described in more detail below.
To best visualize the second question, we have created an interactive map in addition to the static images presented in this report. This map can be explored here, and looks like this.



To answer the last question, we first need to specify what the IRC products currently need and what we want to extend them to in the future. Similar radar products in terms of timeliness of data, real-time versus (more) climatological data, are produced by RMI and DWD. Hence, the inventory of available rain gauge data geared towards its potential use for improving radar products, is also relevant for the RMI and DWD radar precipitation products.

1 IRC products

1.1 Real-time

Currently creates a consequent adjustment factor field every clock hour based on precipitation sum over the past clock hour from automatic rain gauges. This is then used the following hour for the real-time IRC product. This requires hourly rain gauge sums that are available ASAP after every clock hour.
In the future we wish to change this so that it creates an adjustment factor field every 5 minutes, with the precipitation sums of the past 60 minutes. This would require 5-minute rain gauge sums that are available ASAP.

1.2 Early reanalysis

Currently creates a consequent adjustment factor field every day (in the afternoon) based on the precipitation sum over the previous day from 8:00 to 8:00 UTC (using the manual KNMI rain gauges, typically ~200 out of 319 are available). This requires daily sums from manual rain gauges that measure from 8:00 to 8:00 UTC or hourly sums from automatic rain gauges, available the next day.
In the future we wish to use disaggregation to turn the daily sums into hourly sums or even 5-minute sums, so that we can create an adjustment factor field every clock hour or every 5 minutes, just as for the real-time product.

1.3 Final reanalysis

This is the same product as the early reanalysis, but it is run a couple of weeks later. This requires the same data as the early reanalysis, with a higher quality level. Moreover, the manual rain gauge data are complete (319 gauges).

2 Rain gauge data sources

2.1 Germany

We have access to 2 data sources from the Deutscher Wetterdienst (DWD), namely publicly available open data and closed real time data for which we have received an account.

2.1.1 DWD open data

The open data can be freely explored here and contains data for all DWD rain gauges (automated and manual). The open data arrives 40 minutes later than the closed real time data (see further in this document). Therefore, the open data is not suitable for the real time IRC product, but it could be used in the early reanalysis product.

2.1.1.1 Quality codes (QN)

The open data from DWD uses a QN number to indicate the quality of the data the extensive description of what the most used codes mean is as follows:

Q1 (QN=1) = includes the individual test steps virtually directly at the measuring station

- Completeness check
- Limit value testing
- Checking temporal consistency
- Checking the internal consistency in the data set

Q2 (QN=2) = After the station messages have been retrieved, decrypted and imported into the database, the check is made that

- all stations have been successfully retrieved
- were able to enter all retrieved data into the corresponding database tables be provided
- All required messages were generated
- the saved data records and messages are complete
- the reported data is correct

Q3 (QN=3)=data testing with the QualiMET monitoring system. The data from the station go through the following five sub-tests:

- Completeness check
- Climatological consistency check
- Temporal consistency check
- Internal consistency check
- Spatial consistency check

The QN value of 3 means that the QualiMET monitoring system performed automatic corrections, which are continuously applied to the 1-, 5-, 10-minute dataset for the first 75 days after the measurements are first made. This means that during these 75 days this data can be updated as more accurate data becomes available. After these 75 days the QN=3 data is at its best quality and this quality flag is also assigned to the hourly and daily datasets. There are also QN values of greater than 3, but they are only assigned after manual validation of historical (more than a year old) data from the hourly and daily datasets. This data is interesting for climatological applications, or a historical run of the IRC products, but not for the operational IRC products.

2.1.1.2 Dataset: 1-, 5-, 10-minute sums

Description/metadata: The 1-,5-,10-minute datasets contain the same 1-minute rain gauge measurements summed in different intervals. The data can be explored in the following links.

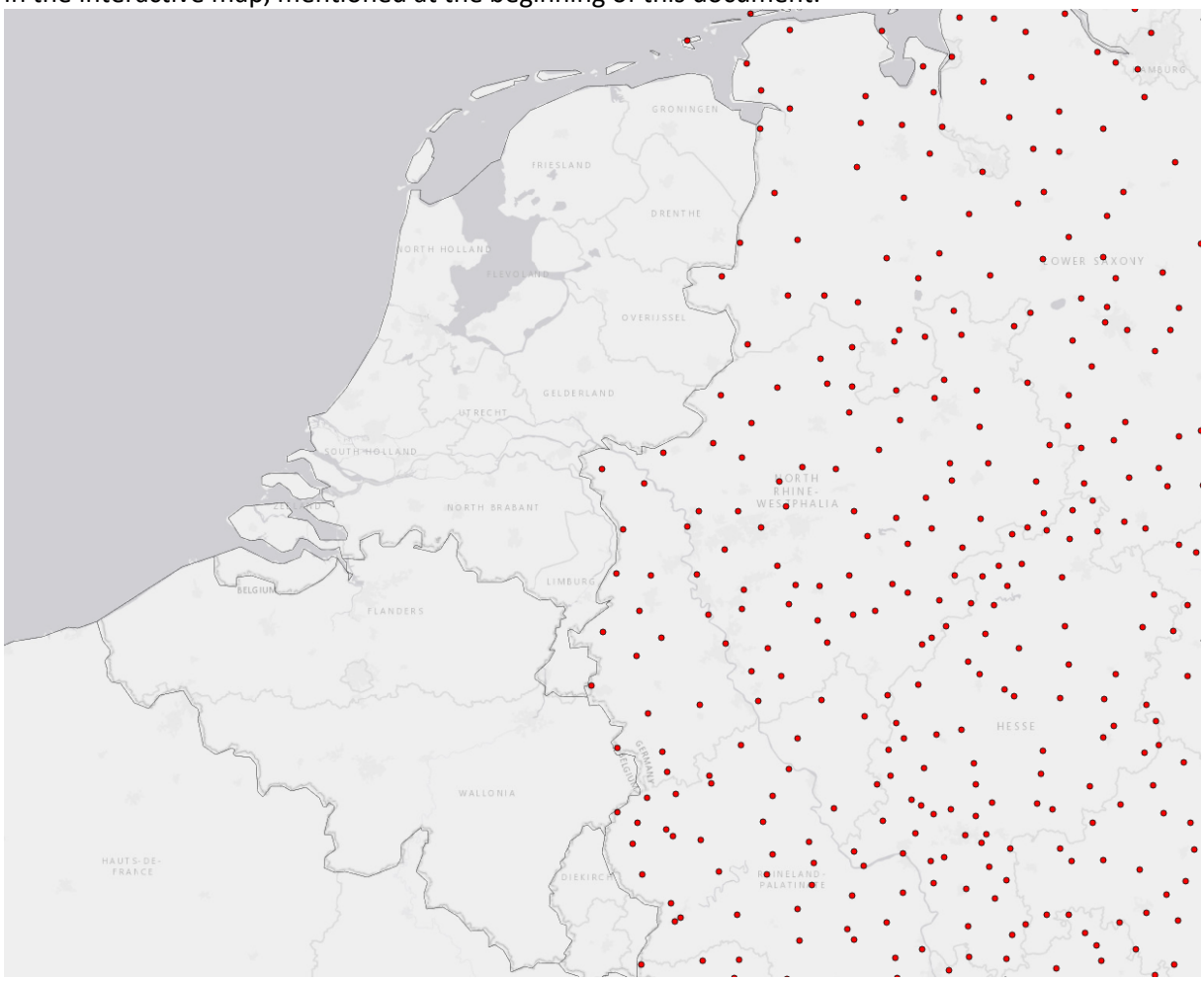
- 1-minute sums
- 5-minute sums
- 10-minute sums

The datasets are described in this document, which seems to be a little bit outdated, but still touches on the core. Note that in this document the following is mentioned:
The measurements relate to a time stamp (in UTC). The meaning of the time stamp is instrument dependent. For the instrument rain[e]H3, the time stamp denotes the end of the measurement inverval. In case of instrument Pluvio Ott a time delay of approximately 5 minutes is introduced due to the inertia of the instrument response. The 10min values are calculated from the 1min values with the respective time stamps. For instance, from November 2008 onward, the 10min value from Pluvio Ott with time stamp 11:50 UTC corresponds to the actual 10min intervall ending 11:45:00 UTC.
This would mean that to retrieve the actual interval to which the received data applies we would need to check the metadata to see if the rain gauge in question is a "rain[e]H3" or a "Pluvio Ott" model. This meta data is delivered in a separate zip file on the same open data server, and it would be possible (if a bit clunky) to automatically read in. However, this is only the case for the open data. In the case of the closed real time data, we do not seem to currently receive this level of station metadata and the stations have completely different IDs, so this metadata is not easily matched. This statement also imposes an interesting question on all data from Pluvio OTT rain gauges as also some Dutch and Belgian institutes use these models. This could be investigated further, but it might not be a big issue for the IRC products as they already use hourly sums to mitigate similar timing issues between radars and rainfall.
The data is delivered in raw text files and contains a semicolon separated table of which for the IRC the following columns would be interesting (unfortunately column naming does not seem to be consistent between datasets).

- The measurement values, usually denoted by something like "RS" or "RWS". This is rain accumulation height (precipitation sum) over the interval in mm, given with an accuracy of 0.01 mm.
- The measurement timestamps, usually denoted by "MESS_DATUM" in UTC.
- The quality indicator, usually denoted by something like "QN". This is a number that is extensively described above, but for these datasets you will only see the numbers 2 and 3. The most recent data already has a QN of 2 and the data of more than a couple of days old will have a QN of 3.

In addition to the text files containing the measurements there is a collection of metadata semicolon separated plaintext files present for each station, but those are contained in a separate zip file per station.
Lastly there is a separate text file that contains a tab separated table with, for each rain gauge, the location (latitude, longitude, height in meters) and some metadata like human readable name and time it was/is active.
Quality: This data has received initial automatic validation when it is initially published and receives additional validation over time, by the QualiMet system.
Access: This data can be accessed through a publicly accessible FTP server.
File format: Plain text semi colon separated.

Timeliness: New data for these datasets is uploaded every half hour (at 5 past every half and full hour) and it contains data that is (depending on the rain gauge) up to 60 minutes old.
Spatial resolution: These datasets contain ~1000 automatic rain gauges spread across Germany. Of these ~1000 rain gauges ~300 are contained in the IRC domain. You can view the locations below or in the interactive map, mentioned at the beginning of this document.



Possible applications: This dataset can be used in the early reanalysis product as well as the final reanalysis product. As discussed later, this dataset is even preferable to the hourly sums dataset as that dataset does not have quality controlled for the first 75 days. This data is too late for the real time product unless we are willing to delay that product by 2 hours.

2.1.1.3 Dataset: Hourly sums

Description/metadata: The hourly dataset contains the data from the automatic rain gauges summed to hourly sums. This dataset is described in this document, and can be explored here. The data is structured similarly to the minute data, with a few exceptions.

- There does not exist a "real-time" directory for this data
- The measurement is given with an accuracy of 0.1 mm
- The most recent data only has a QN of 1 instead of 2 for the 1-,5-,10-minute datasets.
- The quality indicator goes from 1 to 3 much later. The quality indicator is set to 3 for data older than 75 days instead of after only a couple of days for the minute data.

Quality: Unvalidated for the first 75 days, after this it is validated by QualiMet.

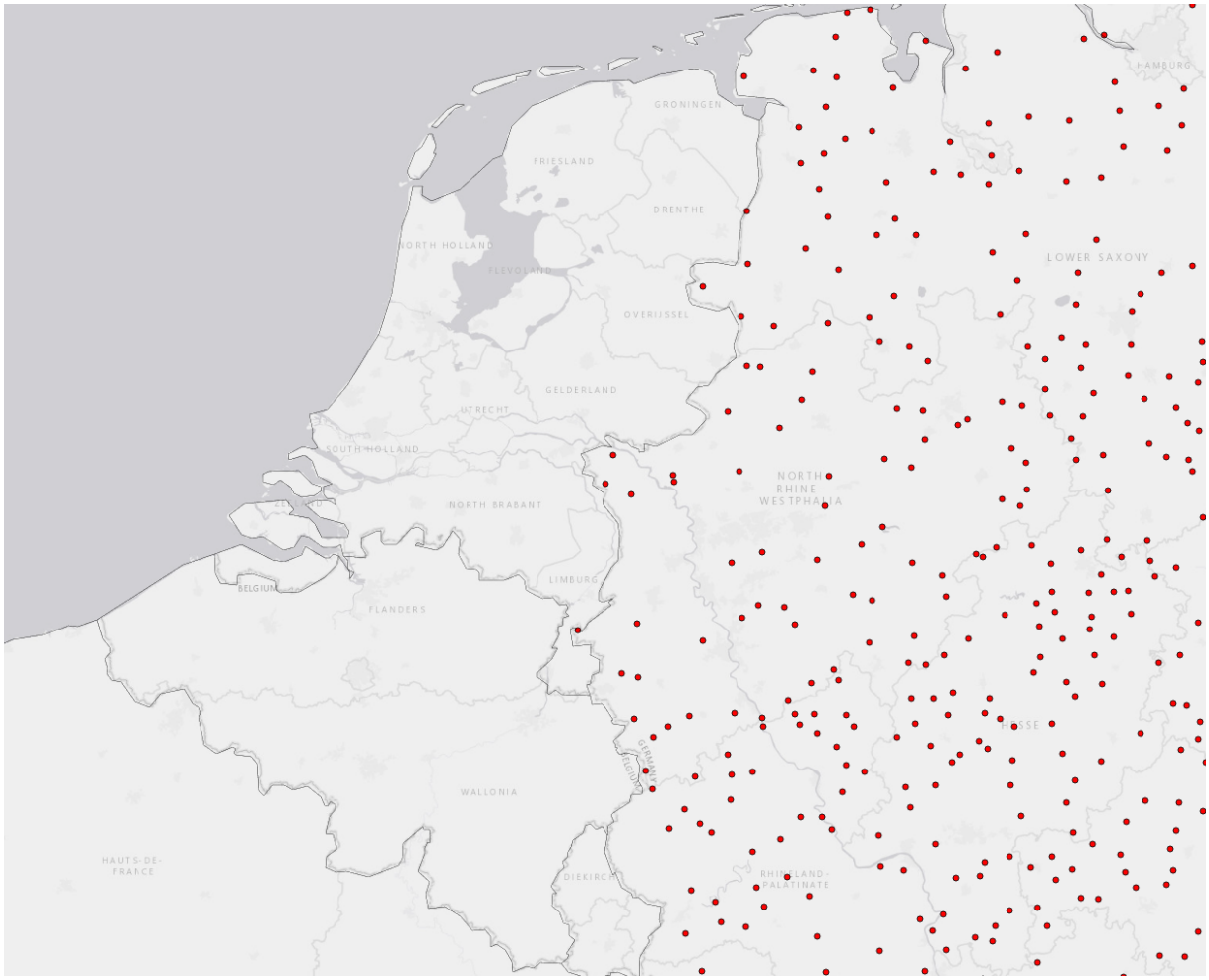
Access: This data can be accessed through a publicly accessible FTP server.
File format: Plain text semi colon separated.
Timeliness: New data for these datasets is uploaded every day at around 09:05 UTC and contains data for up to and including the entire day before.
Spatial resolution: This dataset contains the same ~1000 rain gauges as the previous datasets.
Possible applications: Since this data will have a lower QN number than the 1-,5-,10-minute data for the first 75 days the data is available, we can probably not use this in any of the IRC products unless we delay the final reanalysis by a lot. Also, the 1-,5-,10-minute data already has a QN of 3 after a couple of days, so it would typically be of high enough quality already after this time.

2.1.1.4 Dataset: Daily sums

Description/metadata: The daily dataset contains the data from the automatic rain gauges summed to daily intervals combined with the daily measurements from the manual rain gauges. This dataset is described in this document, and can be explored here.
The data is structured similarly to the hourly data, with a few exceptions.

- This data also contains the manual rain gauges

Quality: Unvalidated for the first 75 days, after this it is validated by QualiMet.
Access: This data can be accessed through a publicly accessible FTP server.
File format: Plain text semi colon separated.
Timeliness: New data for these datasets is uploaded daily and contains data for up to and including the entire day before.
Spatial resolution: This dataset contains ~1000 automatic rain gauges spread across Germany as well as ~1000 manual rain gauges spread across Germany. Of these ~2000 rain gauges ~600 are contained in the IRC domain. Below is an image of just the manual rain gauges as the automatic ones are in the same location as the image above. Again, these locations can also be explored in the interactive map, mentioned at the beginning of this document.



Possible applications: For the same reason as for the hourly data, this data might not be suitable for any of the IRC products. The manual rain gauges might provide extra datapoints for the early/final reanalysis products, but these manual rain gauges measure from 6:00-6:00 UTC instead of 8:00-8:00 UTC, so without proper disaggregation these data sources cannot be mixed.

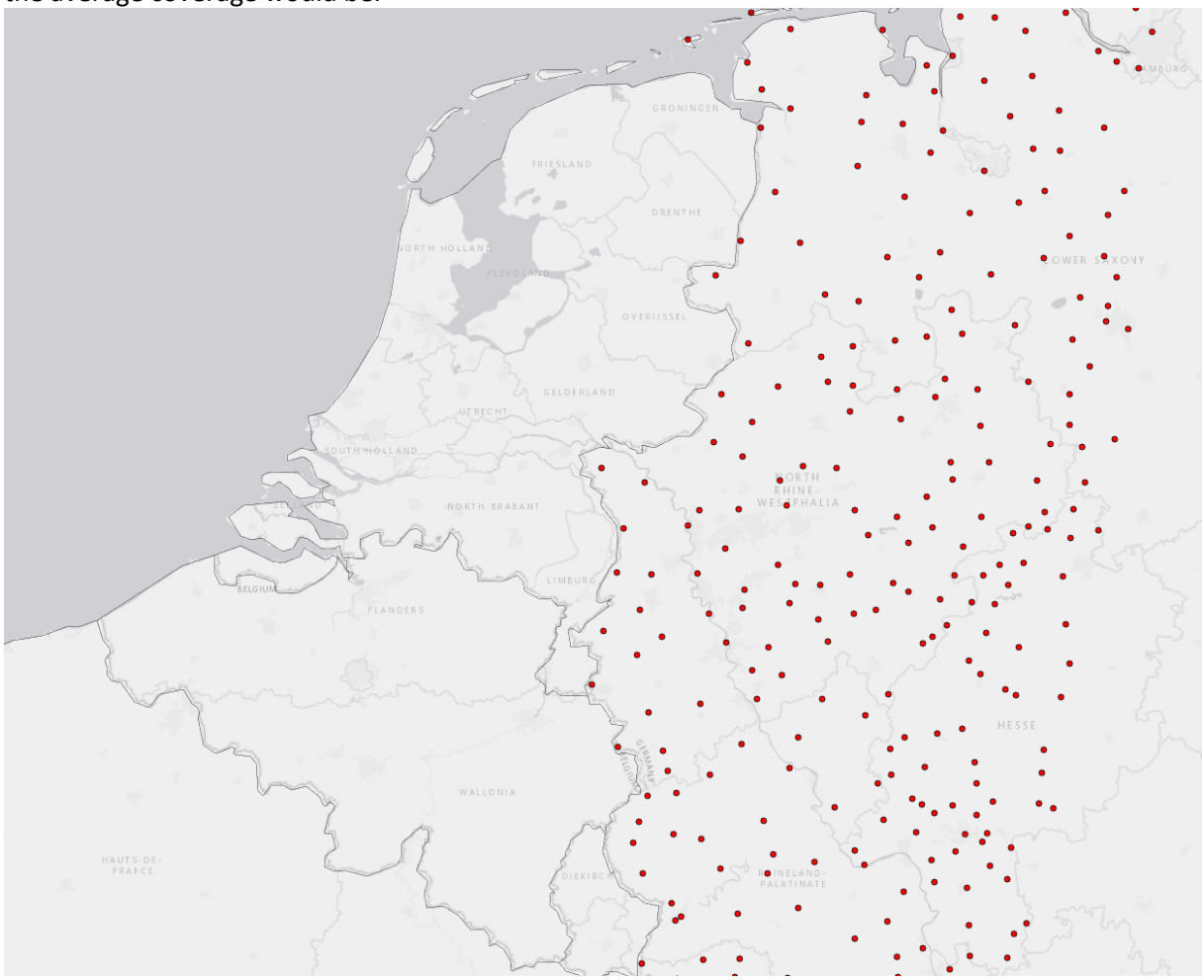
2.1.2 DWD closed real time data

2.1.2.1 Dataset: Real time rain gauge measurements

Description/metadata: This dataset is accessible here with username and password.
This dataset contains exclusively rain gauge data from the same automatic network as the open data, with no quality control. This data is formatted in xml as described in the linked pdf above instead of the plain text format used in open data. Also, the latitude and longitude are contained in the xml files and not delivered in a separate metadata file. Unfortunately for some (~5-10%) stations this location information is missing and since the station IDs are completely different from the ones used in open data it is impossible to use that metadata to fill these in. The metadata that indicates what type (model) of rain gauge is used, which might be of influence on what the timestamps really mean (see earlier in this document), is not present at all and again cannot be derived from the open data.
Quality: Unvalidated
Access: Protected SFTP server
File format: XML (AustauschXML)
Timeliness: New data for these datasets is uploaded at a varying frequency per rain gauge and it contains data up to the minute it was uploaded. Some gauges update every 10 minutes, but most

have an update frequency of 30 minutes. It looks like about half of the gauges deliver data at 5 before the full or half hour, but the other half seems to be at either a different update cycle or is just delayed by around 10 minutes. This means that if you pull all the real time data at 10:37 UTC, you will get almost full coverage from 10:22 onwards. Practically this indicates an update frequency of 30 minutes combined with a delay of 15 minutes.

Spatial resolution: This dataset contains the same ~1000 automatic rain gauges spread across Germany, but only the subset of ~300 of them that are contained in the IRC domain. This means that we get the same data only faster and with less quality control than for the open data sources. Some stations don't report their coordinates and sometimes some are missing, so below is an image that represents the stations that reported at a random moment in the past to give an indication of what the average coverage would be.



Possible applications: This data might be usable in the real time product, but it will mean we need to delay it by 30-45 minutes and maybe change the update frequency. However, since this data has no quality control and limited metadata it is questionable if this data is even worth waiting for, especially since the open data is only another half hour later. If the DWD can increase the update frequency of this data then it would be usable in the real-time product.

2.2 Belgium

From Belgium we have access to rain gauge data from four different sources.

- Hydrologisch InformatieCentrum (HIC)
- Vlaamse Milieumaatschappij (VMM)

- Service public de Wallonie (SPW)
- Koninklijk Meteorologisch Instituut van België (RMI)

The first 2 have their own public APIs, backed by the same KiWIS software, so they are very similar in use. These APIs are free to use for exploring, but in an operational system an authorization key is needed. These keys are already in KNMI's possession. These APIs are hosted on different domains and serve the data in a slightly different way, so we will be discussing them separately. The last 2 are available on the same private API supplied to KNMI by RMI.

2.2.1 HIC/VMM

2.2.1.1 Quality codes

The KiWIS backed data sources report the following table for the quality codes. However, using the "Quality Code Description" parameter we can see that this table is not completely accurate, namely the codes reported between 110 and 179 are described as automatically validated measurements (described by our contact as a possible min/max values validation). The codes below 80 mean that the data has been manually validated (and possibly rejected) by the RMI.

State of Value	Declaration
6-8	External validated data (6=good, 7=estimated, 8=suspect)
10-19	Good measurements
20-29	Good calculations
30-39	Estimate measurements
40-49	Estimate calculations
60-69	Suspect measurements
70-79	Suspect calculations
110-179	Unchecked
221-223	Unknown (import)
255 / -1	Missing

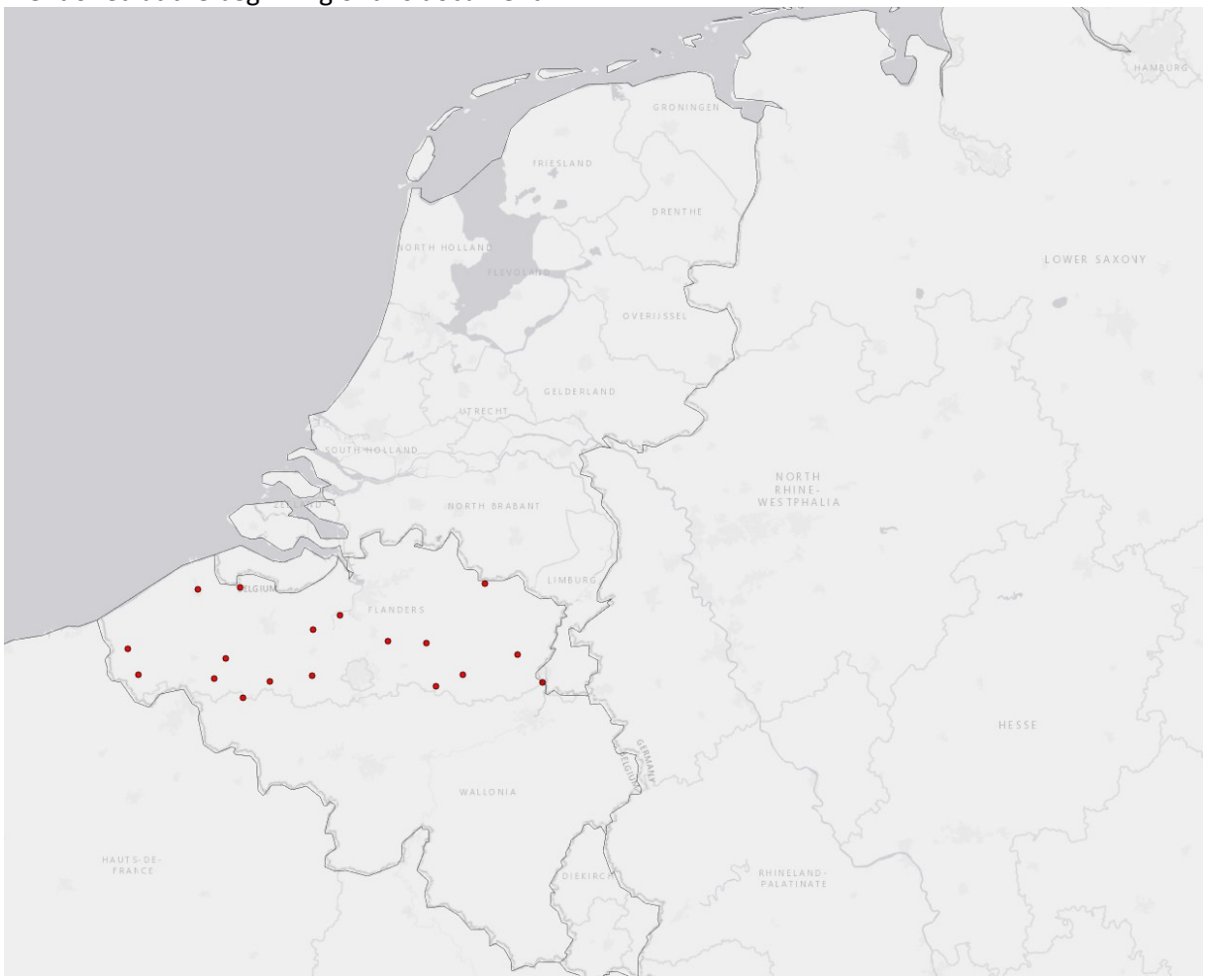
2.2.1.2 Dataset: HIC

Description/metadata: This dataset can be queried through an API. The API manual can be found here. The API can be queried per station with a supplied time-range. There are a number of temporal resolution endpoints, but they all contain the exact same data quality, so the 5-minute temporal resolution endpoint will serve any use case. For a supplied time-range and station, the API will return the following fields of interest.

- The stations latitude and longitude.
- The unit of the measurement, which is mm for all stations.
- A list of datapoints with for each datapoint the following fields of interest.
    - The timestamp of the measurement. This can be returned in any specified time zone, by supplying the timezone query parameter with the request. This timestamp indicates the end of the measurement interval, however since pluvio instruments are used here as well, the same 5 minutes delay as for the DWD pluvio sidenote might hold here.
    - The measured value given with an accuracy of 0.01 mm

- The quality code (see section above), which is 110-179 for real-time data and 10-79 for data older than 6:00 UTC yesterday.

Quality: This data has received initial automatic validation when it is initially published and is further validated the next day by RMI.
Access: Through a publicly accessible API, that requires a key for operational use.
File format: The API supports a number of different response types, one of which is json.
Timeliness: If you query this API every 5 (whole) minutes, it will return all data for up to 5 minutes before, for all stations.
Spatial resolution: The dataset from HIC consists of data for 19 automatic rain gauges in Flanders. The list of stations is found here. You can view the locations below or in the interactive map, mentioned at the beginning of this document.

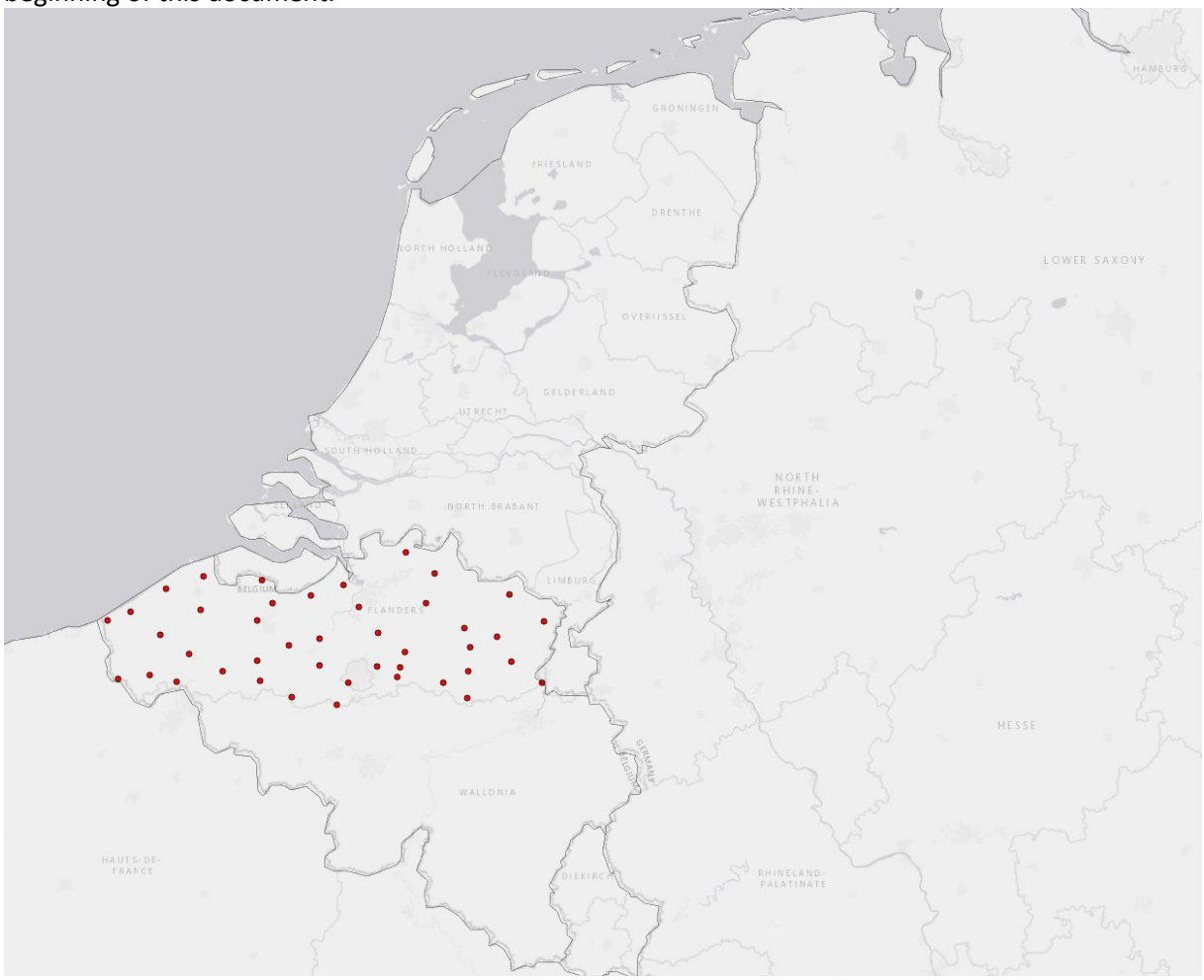


Possible applications: This data should be usable in every product of the IRC. Even the stricter possible future requirements are met.

The VMM also has an API. The API manual can be found here. This API works similarly to the HIC API. For this API there are 2 endpoints of interest. These are the 1-minute precipitation intensity endpoint, which is the timeliest one and the 15-minute precipitation sum one which includes (a small number of) ALMC stations which are not included in the other endpoint. Since for each of these endpoints there are only some slight differences to the HIC API endpoint, only the differences are stated below.

2.2.1.3 Dataset: VMM 1-minute precipitation intensity

Description/metadata: The unit of measurement for this dataset is mm/h as it returns rain intensity. To convert to a 1-minute sum, this value can be divided by 60. The measured value is still given with an accuracy of 0.01 mm, so 0.60 mm/h. This data is not checked by RMI every day, so the quality does not increase from the automatically validated quality it is given initially.
Quality: This data has received initial automatic validation when it is initially published.
Access: Through a publicly accessible API, that requires a key for operational use.
File format: The API supports a number of different response types, one of which is json.
Timeliness: If you query this endpoint every 5 (whole) minutes, it will return all data for up to 5 minutes before, for most stations. Some (~5) can be a bit delayed, but no more than an extra 5 minutes.
Spatial resolution: This endpoint consists of data for 43 automatic rain gauges in Flanders. The list of stations is found here. You can view the locations below or in the interactive map, mentioned at the beginning of this document.

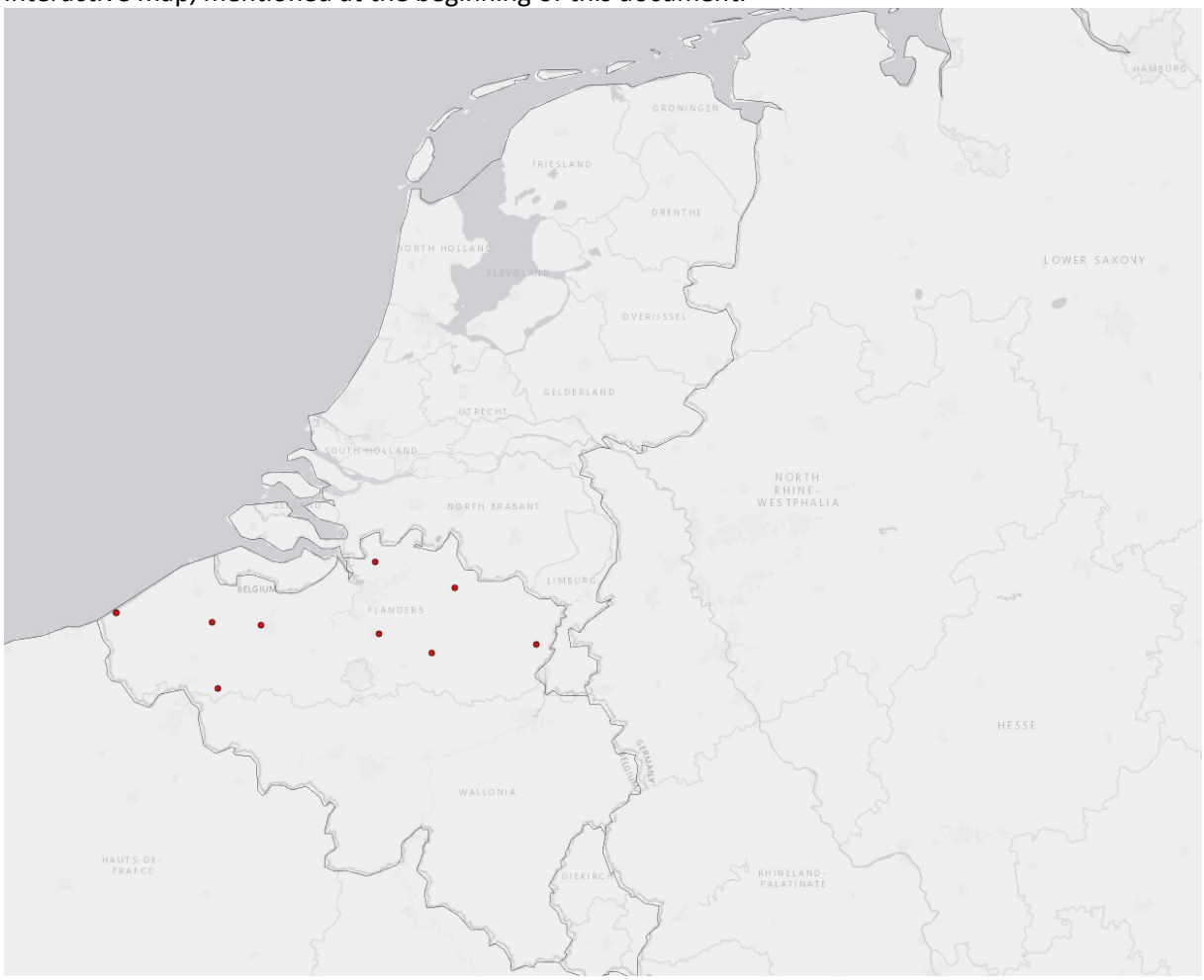


Possible applications: This data should be usable in every product of the IRC. Even the stricter possible future requirements are met. For the final reanalysis product, it might be an issue that the quality does not increase after the initial automatic quality control.

2.2.1.4 Dataset: VMM ALMC 15-minute sums

Description/metadata: This data is also not checked by RMI every day, so the quality does not increase from the automatically validated quality it is given initially.

Quality: This data has received initial automatic validation when it is initially published.
Access: Through a publicly accessible API, that requires a key for operational use.
File format: The API supports a number of different response types, one of which is json.
Timeliness: If you query this endpoint 10 minutes after each quarter clock hour, it will return all data for all stations, up that exact quarter clock hour.
Spatial resolution: This dataset consists of 10 additional ALMC stations. The list of stations is found here.The additional ALMC stations are shown in the image below and can be explored in the interactive map, mentioned at the beginning of this document.



Possible applications: This data should be usable in every product of the IRC. However, the stricter possible future requirements are not met as this data only consists of 15-minute sums and not 5-minute sums. This is not feasible to handle with disaggregation, given the very short duration for which it is difficult to match radar and gauge observations to disaggregate from 15 min to 5 min. For the final reanalysis product, it might be an issue that the quality does not increase after the initial automatic quality control.

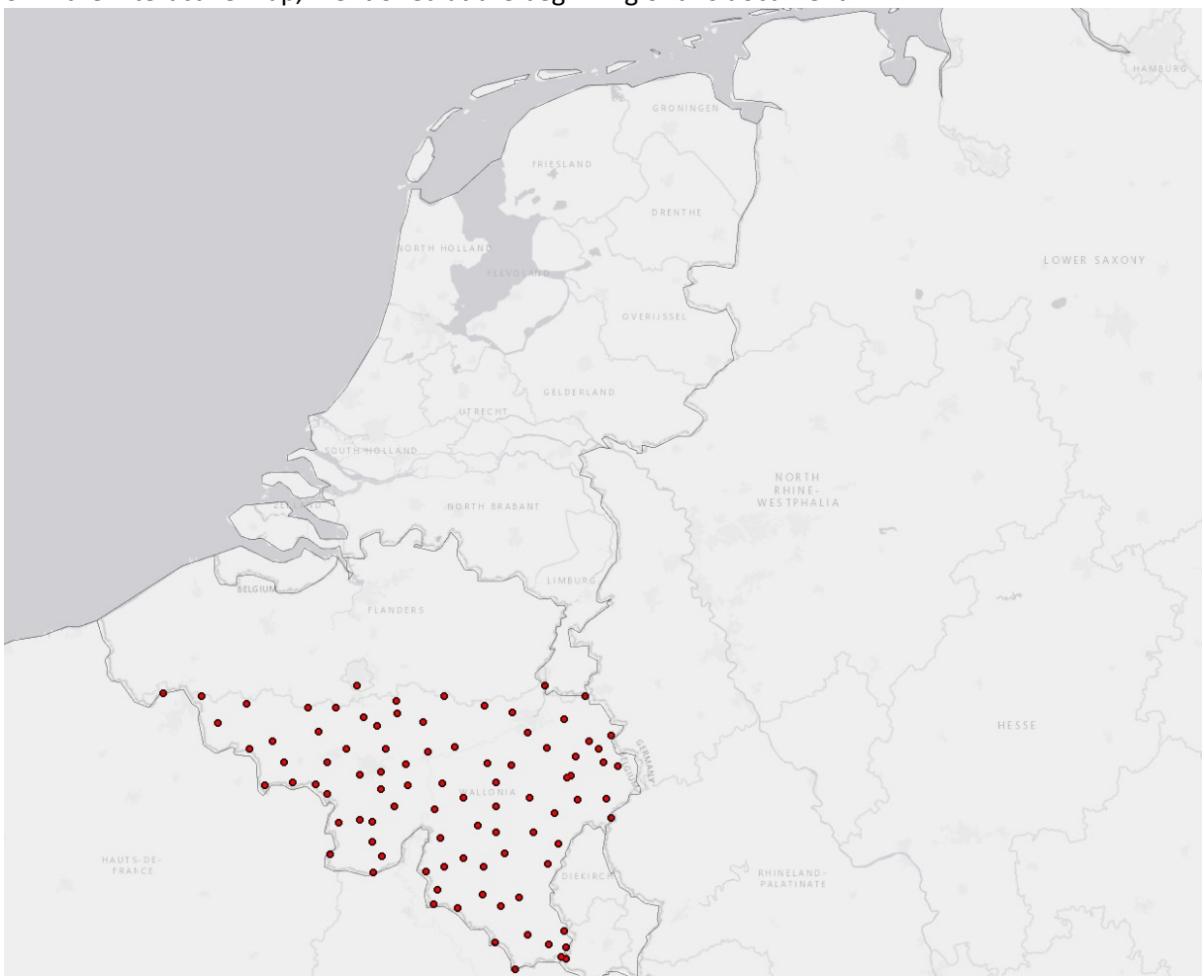
2.2.2 RMI/SPW

The data delivery process of these datasets is still being set up between the KNMI and RMI. We have had some early access, which provided initial insights, but we don't yet have the full picture. For the datasets described here, the only information we currently have on quality control is the following. At least the RMI manual rain gauges are redelivered a few weeks later with quality control, but this may apply to all datasets. The initial data from the automatic rain gauges may have had some

automatic quality control, but this is not known at this time. The data that is delivered by RMI can be split into 3 datasets.

2.2.2.1 Dataset: SPW

Description/metadata: For this dataset we have had a non-real time delivery to explore already, and a real time version will be coming soon. This dataset does not include a lot of metadata, the only relevant fields in the data delivered by the API are date/time (which seem to be local time) and measurement value in mm, given with an accuracy of 0.1 mm. The resolution of the dataset is 10 minutes, so the values given are 10-minute sums. The latitude and longitude are not included in this dataset, but a csv with this information can be downloaded from this page. The station codes are comparable between the 2 datasets so they can be easily matched.
Quality: Not yet known
Access: The dataset can be accessed through a protected http endpoint.
File format: CSV
Timeliness: As stated before, we do not yet have access to the real time dataset, so no conclusion on timeliness can be made.
Spatial resolution: This dataset consists of 90 stations in Wallonia. You can view the locations below or in the interactive map, mentioned at the beginning of this document.



Possible applications: We cannot be sure yet if this data is in time for the real time product as we do not yet have access to the real time data stream, however the data will almost certainly be in time for the early reanalysis.

2.2.2.2 Dataset: RMI automatic rain gauges

Description/metadata: For this dataset we do not yet have anything to explore, but the APIs from RMI all seem to work the same, so the same amount of meta data is expected. This means that we would again need another data source to link the measurements to a latitude and longitude coordinate. In this case there is no csv to be downloaded, but the station latitudes and longitudes can be found here.

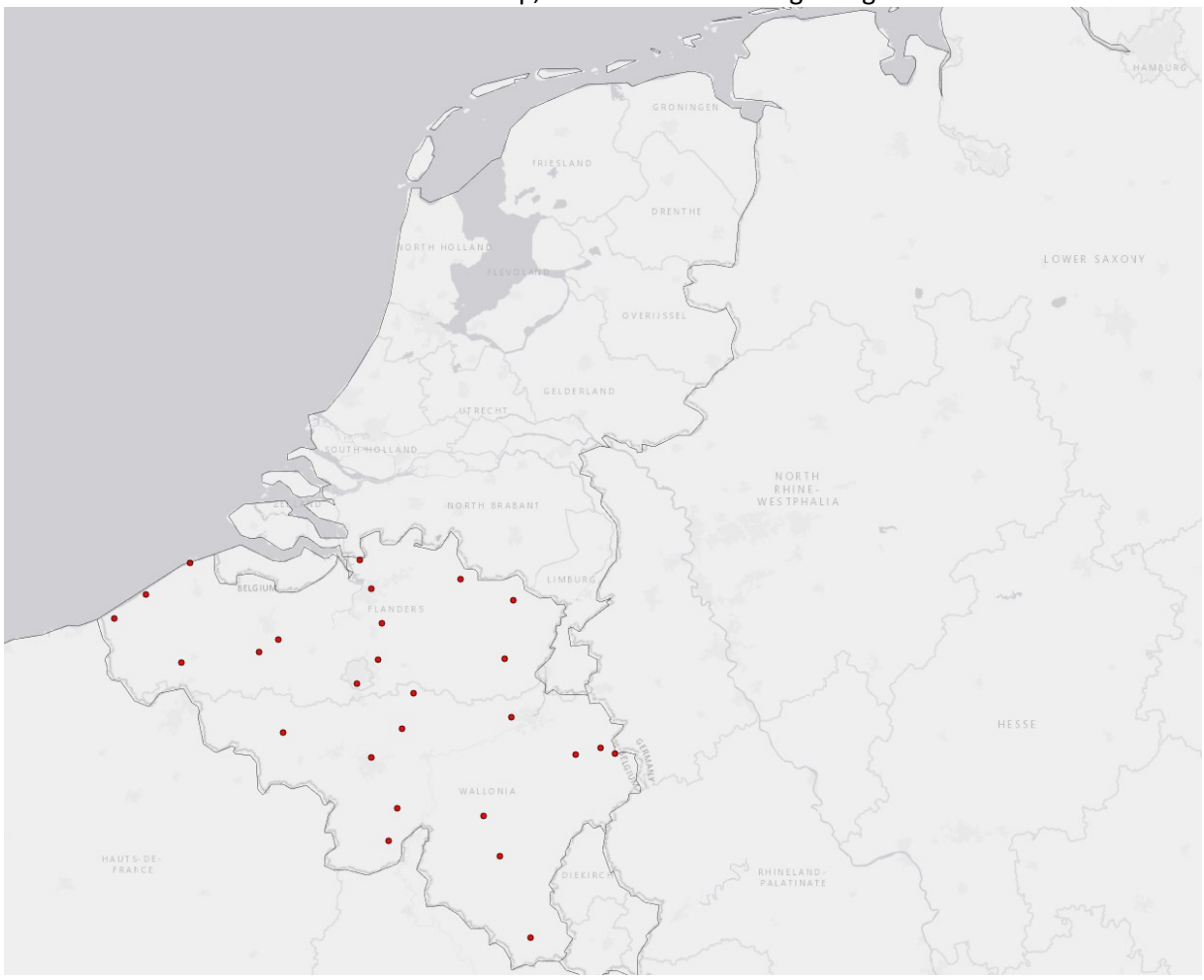
Quality: Not yet known

Access: The dataset can be accessed through a protected http endpoint.

File format: CSV

Timeliness: Since we do not yet have access to this dataset, we cannot make any conclusions on the timeliness.

Spatial resolution: The dataset consists of 26 automatic stations spread across Belgium. You can view the locations below or in the interactive map, mentioned at the beginning of this document.

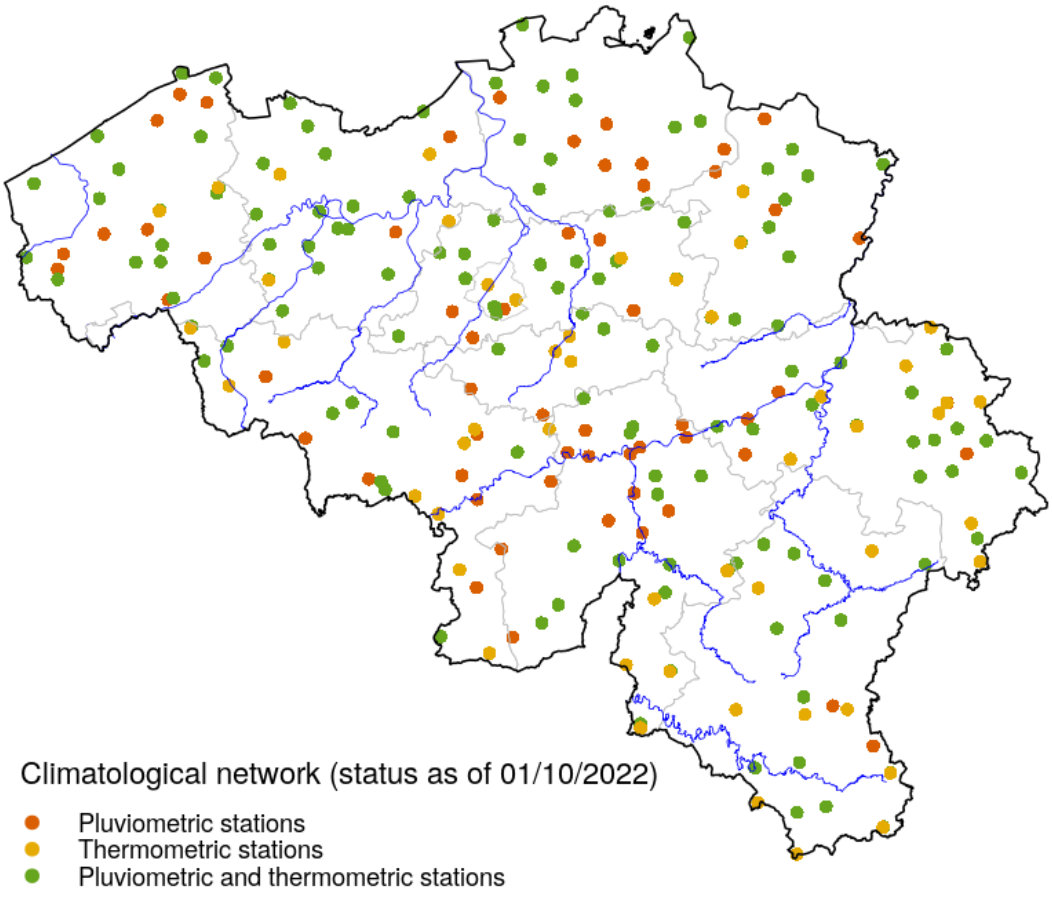


Possible applications: Same as for the SPW dataset.

2.2.2.3 Dataset: RMI manual rain gauges

Description/metadata: This dataset consists of daily sums measured at 08:00 (uncertain if this is local time or UTC). We don't have full access to this dataset yet and since it is delivered in a similar way as the other 2, we also don't have the stations latitude and longitude included in the dataset. In this case we were also not able to find the latitude and longitude on the public RMI page, so a different image is included below, and this data is not included in the explorer webpage for now. For

this dataset we know for sure that there is a redelivery after a few weeks that includes validated data for these stations.
Quality: Not yet known if initial validation is performed, but we do know that this dataset is republished after a few weeks with validated data.
Access: The dataset can be accessed through a protected http endpoint.
File format: CSV
Timeliness: Again, we cannot be sure as we don't have full access yet.
Spatial resolution: The dataset consists of ~200 manual rain gauges spread across Belgium as seen below (only the red and green dots are rain gauges). These are not included in the interactive map, mentioned at the beginning of this document, as we don't have a list of coordinates for these stations.



Possible applications: It is unsure if the measurement interval is from 08:00 to 08:00 UTC or local time, so we cannot be sure if it will be combinable with the Dutch manual rain gauge measurements or if disaggregation is necessary here as well. Depending on this, the dataset could be useful for both the early and final reanalysis. Especially since this data is validated after a few weeks it could prove useful for the final reanalysis.

2.3 The Netherlands

For the Netherlands all rain gauge data is offered by the KNMI as we collect this data from several sources and (mostly) offer this as open data on the KNMI data platform (KDP). The datasets are as follows.

- Automatic rain gauges from the KNMI
- Automatic rain gauges from Dutch water boards and water companies.
- Manual rain gauges from the KNMI.

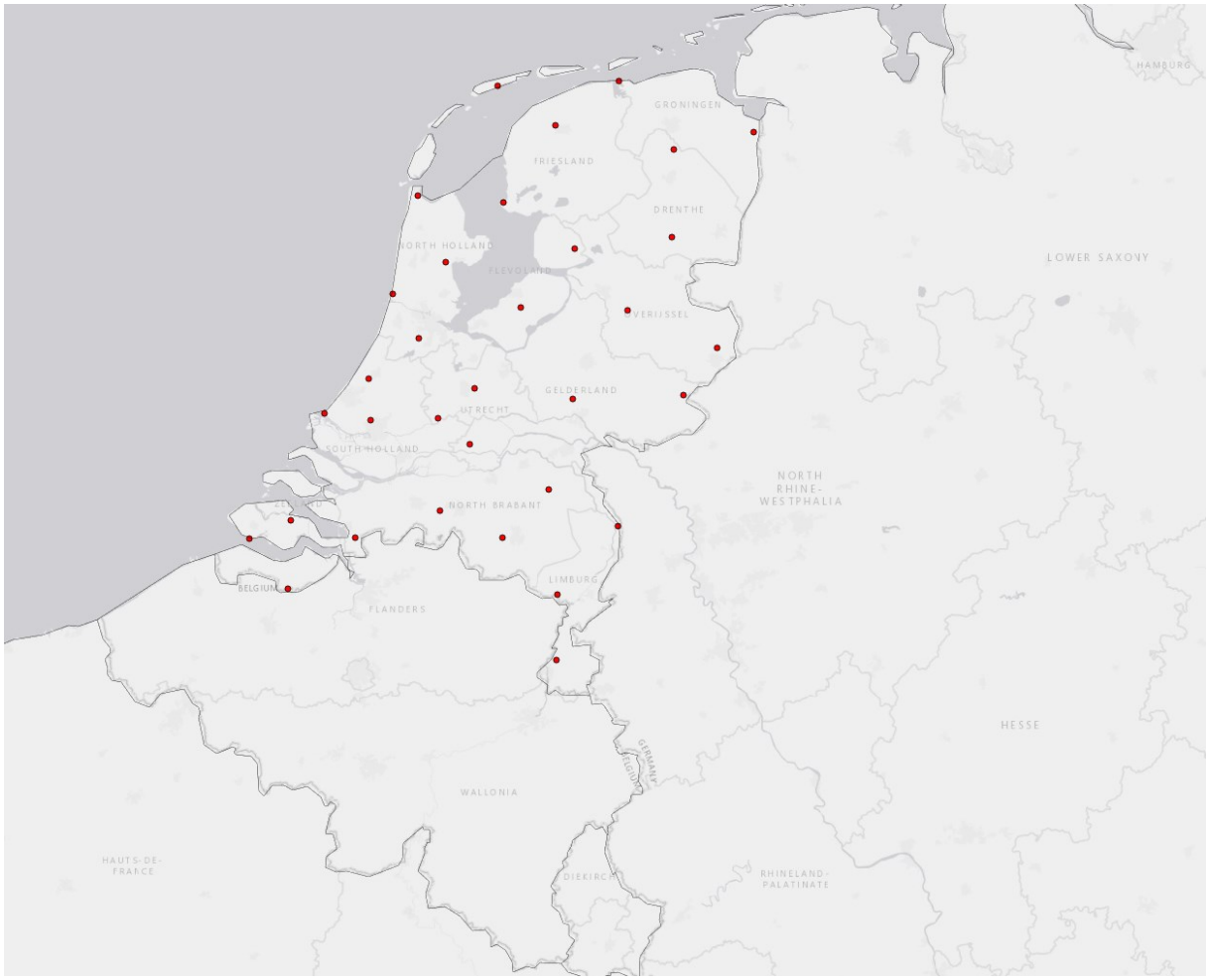
2.3.1 KNMI

2.3.1.1 Dataset: KNMI automatic rain gauges

Description/metadata: This data is available as 10-minute data here. The data is delivered as 1 netcdf file per 10 minutes, with for each data point the following parameters.

- The latitude and longitude.
- The timestamp of the measurement. This timestamp indicates the end of the interval. And is given as seconds since 1950-01-01 00:00:00 UTC.
- The measured value under variable “rg”. This is the precipitation intensity in mm/h averaged over 10 minutes, so to convert this to a precipitation sum in mm, this value needs to be divided by 6.

After multiple months this data is also revalidated and republished here. This data is not relevant for any of the IRC products but could be useful for historical runs.
Quality: This data has received basic automatic validation and incorrect values are removed.
Access: This data can be freely accessed through an API, but for operational use an API key is required.
File format: netCDF
Timeliness: These files become available around 6 minutes after each (whole) 10-minute interval and contain all data for all stations up until the end of that interval.
Spatial resolution: This dataset consists of 33 automatic rain gauges in the Netherlands (and 2 on the BES islands). You can view the locations below or in the interactive map, mentioned at the beginning of this document.



Possible applications: This data is already used in every product of the IRC. However, the stricter possible future requirements are currently not met as this data only consists of 10-minute sums and not 5-minute sums. This could possibly be handled by disaggregation again, but that would mean extra work.

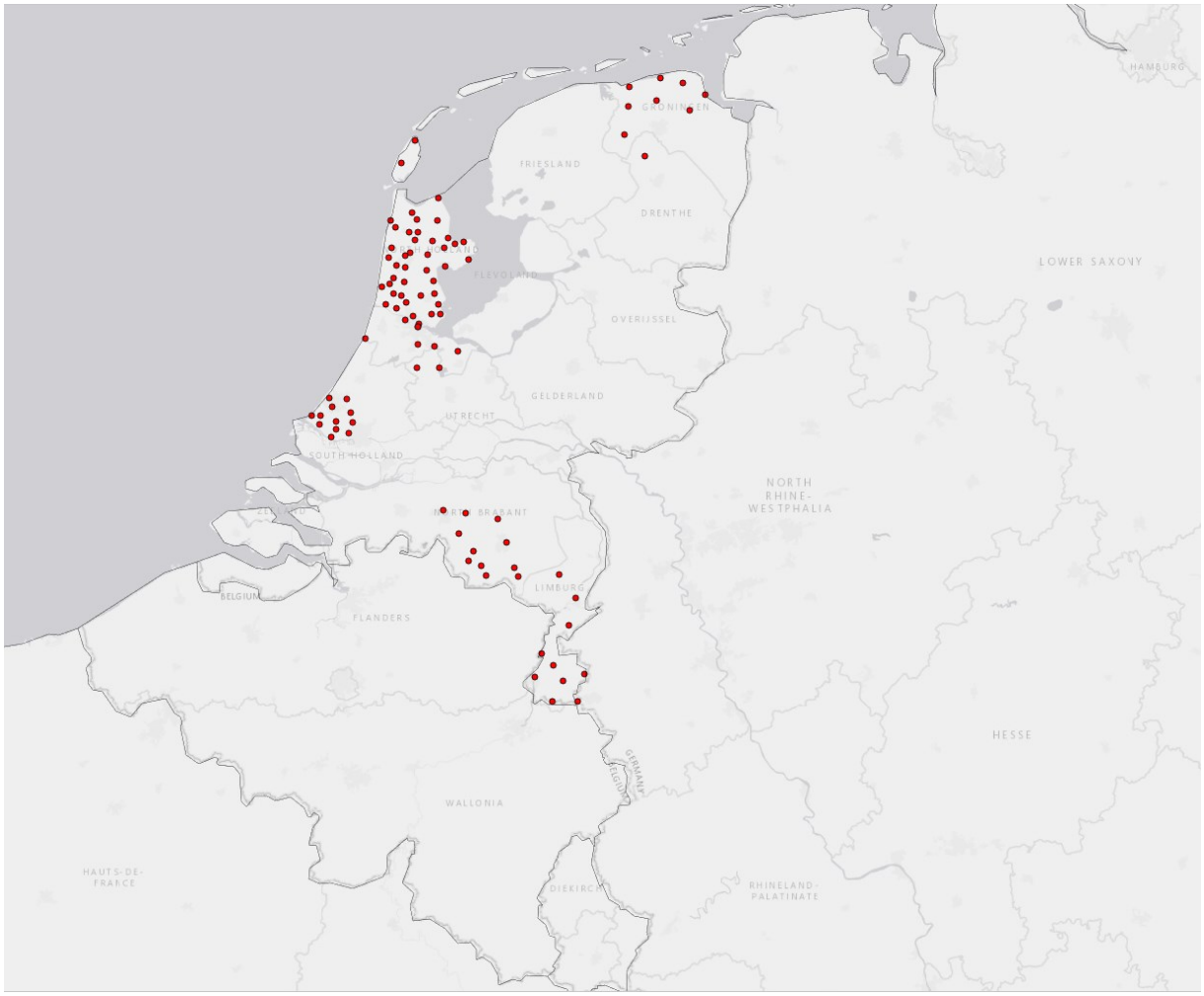
2.3.1.2 Dataset: Automatic rain gauges from Dutch water boards and water companies

Description/metadata: This dataset consists of a collection of rain gauge data from several different Dutch water boards and water companies. Therefore, the timeliness and temporal resolution can vary from rain gauge to rain gauge in this dataset. This data comes from a new product/process only recently started in collaboration between the water boards/companies and the KNMI, the dataset is therefore not yet complete, and we will keep working together to perfect this flow. The data is all collected in the same XML format though. Quality control is performed on all incoming measurements, which currently include algorithms to detect high influx or faulty zero. Additional algorithms are being considered for future implementation. Afterwards this data is republished by KNMI in a single dataset here. (At the time of writing, this dataset is not yet available, but it will be soon). The XML format contains the following fields of interest.

- A timezone field, which is UTC.
- A series of rain gauges with for each rain gauge the following fields of interest.
    - The latitude and longitude.
    - The unit and type of measurement, which is precipitation sum in mm.

- A list of measurements with the following fields. In a temporal resolution of either 1 or 5 minutes.
  - The date and time of the measurement in the above mentioned timezone.
  - The measured value given with an accuracy of 0.1 mm.
  - The quality flag, which is 0000 if the measurement is accepted by the validation algorithm and another value if not.

Quality: This data has been automatically validated when it is first published. This is done by comparing it with radar data which is a fairly advanced method of initial automatic validation.
Access: This data can be freely accessed through an API, but for operational use an API key is required.
File format: XML (Delft-Fews)
Timeliness: The timeliness of the measurements depends on how fast the water boards and water companies send this data to the KNMI. Currently, it is observed that within 10 minutes most data from the currently available water boards and water companies is received and processed by the validation algorithm. However, not all of them currently meet this mark. The ambition of KNMI and contributing parties is that all data will be delivered within this time, so they can be used in the real time IRC product.
Spatial resolution: This dataset consists of a continuously growing collection of rain gauges in the Netherlands. Currently it contains 93 rain gauges, but this is expected to grow well beyond 200. You can view the locations below or in the interactive map, mentioned at the beginning of this document.



Possible applications: A short-term goal is to start using suitable rain gauge data, after quality control, in all IRC products. As discussed already, currently this dataset is not yet complete and there are still delivery challenges to overcome. However, for the early and final reanalysis products this data (or at least the part of it that is already available) can already be used. Since the measurement values are always given in a temporal resolution of at least 5 minutes, even the stricter possible future requirements are met. Currently this data is not manually validated after a couple of days, which might be an issue for the final reanalysis. However, since a fairly advanced automatic validation is performed, this might be enough for the reanalysis products as well.

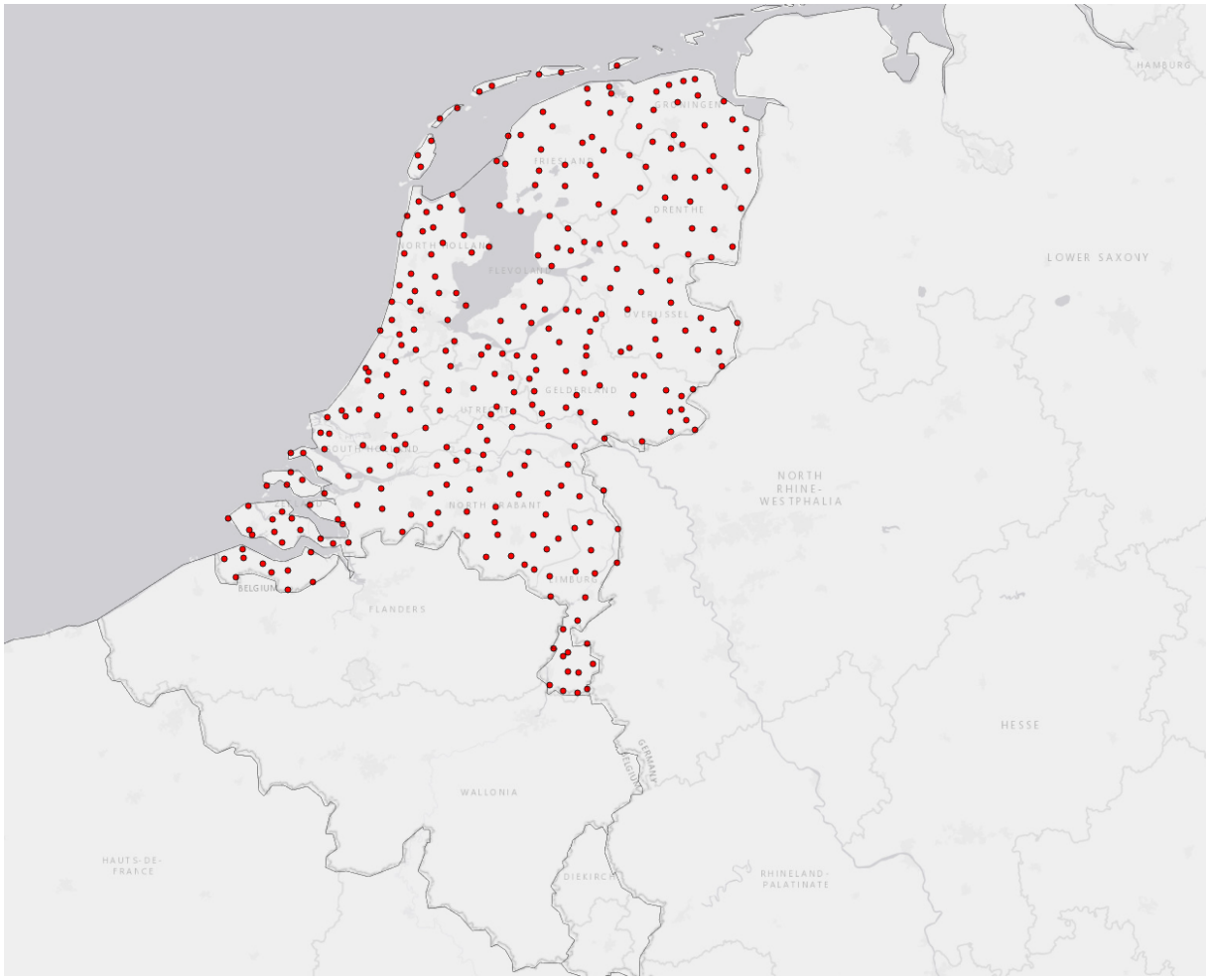
2.3.1.3 Dataset: KNMI manual rain gauges

Description/metadata: This dataset contains a collection of manual rain gauges operated by volunteers spread across the Netherlands. The unvalidated data is available here, and the validated data can be browsed here. At the bottom of the validated data page there are some interesting links. Firstly, a link to a document with the approximate station locations (useful for both the unvalidated and validated data). This document contains approximate locations because of privacy concerns, since these stations are sometimes located in the volunteers' back yards. Internally a more accurate location is known. Secondly, a link to a tutorial on how to retrieve this validated data programmatically. However, this is less usable as KDP, so we would prefer to also have this data available on KDP. Again, internally a better way to retrieve this data is available.

The unvalidated dataset on KDP consists of a comma separated raw text file with on each line a measurement value with the following columns of interest.

- The first column is the station number, this is useful to deduce the approximate coordinates from this list. In that list the latitude can be found in the POS_NB column and the longitude in the POS_OL column.
- The fifth column is the date for which this measurement is valid. The valid time of the measurement is from 08:00 UTC the day before this date until 08:00 UTC of this date.
- The sixth column is the measurement value in tenths of mm (i.e., 15 means 1.5 mm accumulated rain).

Quality: This data has been automatically validated when it is first published and is republished a couple of weeks later after manual validation.
Access: This data can be freely accessed through an API, but for operational use an API key is required.
File format: Comma separated raw text
Timeliness: During the day measurements from the manual stations arrive, with the most complete dataset of all 319 stations usually not even being available within the day. In the afternoon, when the early reanalysis runs, there are usually ~200 stations available. The validated data is usually created around 2 weeks after the unvalidated data becomes available and it contains manually validated and adjusted data for all 319 stations.
Spatial resolution: The early reanalysis runs with ~200 stations distributed across the Netherlands. The final reanalysis uses the updated validated data, with all 319 stations. See below for the precise locations, the approximate locations are in the interactive map, mentioned at the beginning of this document.



Possible applications: This data is already used in the early and final reanalysis product, but since they consist of daily sums, they need to be disaggregated to be useful under the possible stricter future requirements.

3 Conclusions

3.1 Quality

All of the data that is discussed in this report has had some form of validation done already by the official institutes managing the data. This means that additional validation before using it in the IRC products is not necessary.

3.2 Timeliness

We recommend that all participating parties really strive to deliver data every 5 minutes and as real-time as possible (at most 5 minutes delay). The KiWIS data sources from HIC and VMM are a great example of how all data delivery should be. KNMI comes close to reaching this goal, but DWD is still quite far off.

3.3 File format

We recommend that all data be exchanged in a common file format so that they can be more easily combined. Currently there is a big difference between even data sources within the same country. This data should then ideally also be hosted on a central storage so that the entire collection can be easily accessed and used.

3.4 Spatial resolution

The table below contains the approximate number of manual and automatic rain gauges per country.

Country	**Manual rain gauges**	**Automatic rain gauges**
Germany	1000	1000
Belgium	200	200
The Netherlands	300	150

4 Other works

An overview of (sub-daily) rain gauge datasets for Europe is given by van der Schrier et al. (2021). The rain gauge network density would be substantially increased when these datasets would become available for national meteorological services.

4.1.1 References

Gerard van der Schrier, Aart Overeem, Gé Verver. Improving Access to Sub-daily Rain Gauge Data. European Environment Agency. Copernicus programme. EEA/DIS/R0/20/001 Lot 1. 2021. https://insitu.copernicus.eu/library/reports/COINS_Report_WP3_raingauges_v2.pdf