Climate scenarios

Times series information

Theo Brandsma

Rudmer Jilderda

KvR 041/11

Times series information

Authors

Theo Brandsma
Rudmer Jilderda

KNMI

Royal Netherlands
Meteorological Institute
Ministry of Infrastructure and the Environment

KvR report number KvR 041/11
ISBN ISBN/EAN 978-90-8815-033-3

This project (CS08; Time series information) was carried out in the framework of the Dutch National Research Programme Climate *changes* Spatial Planning. This research programme is co-financed by the Ministry of Infrastructure and the Environment.

Contents

Summary

Summary in Dutch

Het CS8 project richt zich op het digitaliseren van databronnen met daarin lange tijdreeksen van meteorologische variabelen. De resulterende datasets spelen een belangrijke rol bij de inrichting van Nederland en vormen de basis voor studies naar klimaatverandering en de gevolgen daarvan. Zo worden de ontwerpnormen van rivierdijken, ontwatering- en afwateringsystemen en rioleringen voor een belangrijk deel bepaald door het optreden van extreme meteorologische gebeurtenissen in het verleden. Daarnaast is voor het detecteren en begrijpen van klimaatverandering en de gevolgen daarvan, de beschikbaarheid van digitale meteorologische datasets essentieel. Daarbij zijn inbegrepen de calibratie, verificatie en tuning van (impact) modellen.

Nederland is een van de weinige landen waar meteorologische data met voldoende lengte en resolutie, in ruimte en tijd, beschikbaar zijn. Een groot deel van deze data is echter alleen beschikbaar op papier. In het CS8 project is een belangrijk deel van deze data gedigitaliseerd en opgenomen in digitale datasets. Het gaat om de volgende vier databronnen: (1) alle tijdseries met dagneerslagen over de periode 1850-1950, (2) hoge resolutie (5-minuten) neerslagsommen uit pluviograafstroken van de stations De Bilt (1897–1993), Eelde (1954–1993), Den Helder/De Kooy (1954–1993), Vlissingen (1954–1993), Beek (1954–1993) en Amsterdam (1920–1983), (3) Meerdaagse waarnemingen van het weer voor verschillende locaties in de 18e en 19e eeuw, en (4) uurtemperaturen van Batavia/Jakarta in de periode 1866-1980. De databronnen zijn op kwaliteit gecontroleerd en zijn vrij beschikbaar via Internet. Verschillende voorbeelden van toepassingen zijn nader uitgewerkt.

Voor nadere informatie kunt u contact opnemen met Theo Brandsma, theo.brandsma@knmi.nl.

Summary

The CS8 project focuses on the digitization of data sources of long time series of meteorological variables. The resulting datasets play an important role in the infrastructural planning of the Netherlands and form the basis for studies into climate change and its consequences. For instance, the design standards of river dikes, drainage and sewerage systems are to a large extent determinded by the occurrence of extreme meteorological events in the past. Furthermore, the availability of digital meteorological data sets is essential for detecting and understanding climate change and its consequences, including the calibration, verification and tuning of (impact) models.

The Netherlands is one of the few countries where meteorological data with sufficient length and resolution, in space and time, are available. However, much of this data is only available on paper. The CS8 digitized a large amount of this data and entered it into digital data sets. It concerns the following four data sources: (1) all time series with daily precipitation for the period 1850-1950, (2) high-resolution (5-minute) precipitation totals from pluviographs of the stations De Bilt (1897-1993), Eelde (1954-1993), Den Helder/De Kooy (1954-1993), Vlissingen (1954-1993), Beek (1954-1993) and Amsterdam (1920-1983), (3) multi-day weather forecasts for several locations in the 18th and 19th century, and (4) hourly temperatures of Batavia / Jakarta in the period 1866-1980. The data sources are checked for quality and are freely available via the Internet. Several examples of applications are presented.

Extended Summary

Background

Datasets including long time series of meteorological variables form the basis for infrastructural design and studies into climate change and its effects. Unfortunately those datasets are scarce. This especially holds when one is interested in extremes because these require high quality data. The Netherlands is one of the few countries for which data of sufficient length and spatial and temporal resolution are available. However, a large proportion of the data is only available as hard copy and therefore not readily available to researchers and the wider public. The CS8 project aims to rectify this situation. In CS8 four data sources have been digitized, checked for quality and, where necessary, homogenized. For some sources, relevant statistical parameters are also determined and used to describe the present climate, including extremes.

Data source 1: Daily precipitation amounts 1850-1950

KNMI measures daily precipitation since about 1850. In the 1850-1950 period, the network gradually increased to its present density of about 325 rain gauges. The measurements are taken by voluntary observers every morning at 8:00 UTC. In the first decade of the 20th century a standardization of the measurements was implemented. In contrast to the post-1950 period, where all observations are digitially available, only about 10% of the observations were digitally available in the 1850-1950 period. Within the CS8 project, all pre-1951 precipitation were digitized and quality controlled.

The new dataset extends the post-1950 data a hundred years back in time with about 12300 station years (4.5 million measurements). Special attention is given to the homogenization and analysis of this dataset. The data is freely available from the KNMI website including the digital images of the data source, which allows the user to go back to the original data when needed/desired. Digital images were made also of all reports made by station inspectors. These images may contain relevant metadata and are available upon request for each station.

The dataset can serve several purposes, such as: the calculation of design standards for dikes, canals and drainage systems by ministries, waterboards, municipalities and consultancies; the calculation of a base-line climate in climate impact studies; the validation of (regional) climate models; and the assessment of precipitation trends and variability and their areal distribution.

Data source 2: High resolution precipitation from rainfall strip charts

Self-recording rain gauges have been applied for continuous rainfall measurements at a selected set of KNMI stations since the end of the 19th century. At first, rainfall was recorded on daily and sometimes weekly rainfall strip charts. Thereafter, from about 1980 through 1993, paper rolls were used containing the rainfall registration for about 10-20 days. Finally, from 1994 onwards rainfall measurements are transferred electronically and operationally stored at 10-minutes resolution. Until now, the strip charts and paper rolls have been used mostly for extracting hourly values. In infrastructural design (e.g. sewer systems, tunnel drainage) there is, however, a need for long high-resolution rainfall series with resolutions much higher than 1 hour. Fortunately, the charts and rolls can be used to extract this kind of information with a time resolution of about 5 to 10 minutes.

In CS8 we created a dataset with 321 station years of 5-minute precipitation from 20th century rainfall strip charts and paper rolls. The pluviograph registrations of the following stations have been digitized: De Bilt (1897-1993), Eelde (1954 - 1993), Den Helder/De Kooy (1954-1993), Vlissingen (1954-1993), Beek (1954-1993) and Amsterdam (1920-1983). Besides their use for the design of fast responding systems like sewerages and tunnels, such high-resolution precipitation values are important for validation of climate models, the calibration of stochastic rainfall models and the assessment of trends and variability of high resolution precipitation.

Because of the labor-intensive extraction work, we developed an automatic curve extraction (ACE) algorithm for processing bulk quantities of rainfall strip charts and rolls. The framework largely automates the processing of rainfall curves on the strip charts and paper rolls and can be applied to other elements as well.

The quality controlled dataset is freely available via the KNMI website. A set of high-resolution digital images of the rainfall charts and paper rolls (1.4 Tb in total) is available upon request for each station. The source code of the ACE software is also freely available and may be adapted for other applications.

Data source 3: Daily pre-1850 meteorological observations

CS8 also digitised part of the old pre-1850 measurements in the Netherlands. For the reconstruction of the pre-1850 climate, it is important to have as many as possible parallel time series available. In previous projects a large part of these data has already been digitized and published on the KNMI website. The CS8-project adds a signicant number of stations to these data. The digitization of this kind of data is labor-intensive because of the mostly hand-written base material and the non-standard way of observing like the use of old units and standards, irregular observation times and (sometimes) the use of Latin as writing language.

Data files have been created with (sub)daily meteorological measurements/observations of 13 stations in the Netherlands in the pre-1850 period. In total it concerns about 150 station years and 110.000 measurements. The data (corrected for typing errors) is freely available via the KNMI website, along with pre-1850 data digitized in previous projects. A set of digital images of the hardcopy source material of the pre-1850 stations in the Netherlands is also available from the KNMI website and enable the user of the data to go back to the original data source.

The main application of this type of data is mainly in climate reconstruction but the observations may also be used to study the impact of weather on society, like agricultural production, diseases, floods, etc.

Data source 4: Hourly observations of Batavia/Jakarta (1866-1970)

CS8 digitized the hourly temperature measurements of Batavia/Jakarta in Indonesia in the 1866-1980 period. Indonesia is situated in a data-sparse area where long time series are scarce. These series are, however, needed to validate climate models for that area, to construct a baseline climate and to develop climate scenarios. The KNMI archives contain a large amount of hardcopy meteorological observations and measurements for Indonesia for the 1850-1980 period. The remainder of the CS8 budget has been used to digitize a part of this unique set of data. The larger part of the data is being digitized within in a cooperation between KNMI and the Indonesian meteorological office BMKG (see www.didah.org).

The dataset delivered by CS8 contains the hourly air temperature measurements of Batavia/Jakarta for the 1866-1980 period (about 102 station years, 894.120 measurements). This quality controlled data is freely available via the KNMI website. A set of digital images of all hardcopy Indonesian data available in the KNMI archives is available from the KNMI website and enable the user of the data to go back to the main hardcopy source of the data.

The digitized data show a warming trend of about 1.5°C in the 1866-1980 period, which is much larger than the global warming trend in that period. A comparison of the mean diurnal cycles in non-overlapping 30-periods show that the warming is mainly visible in the daytime temperatures.

The value of this dataset will increase when the other variables (e.g. wind speed and direction, humidity, rainfall, air pressure) become available (see www.didah.org). The daily values of the digitized data for Indonesia and neighboring countries are also being collected in the Southeast Asian Climate Assessment & Dataset (SACA&D), which enables the analysis of the extreme events in the region (see http://saca-bmkg.knmi.nl).

1. Introduction

Time series of meteorological variables are essential for the design of the infrastructure of a country and for understanding and modeling climate change and variability and its possible impacts. In the Netherlands the design of, for instance, dams, dikes and sewer systems is strongly related to precipitation. With the advance of computers and the awareness of climate change, there is a tendency to model the corresponding systems and use climate time series as input (see e.g. Parmet et al., 1990). The availability of historical climate time series provides the possibility to assess the behavior of these systems for the current climate, including its natural variations. Because violent extremes have a dominating impact, the time series should be of high resolution not only in space but also in time. Also, to determine probabilities of occurrence of such events, the series should be as long as possible.

To understand and model climate change and variability, datasets with climate data are indispensible. For instance, the datasets are needed to validate global and regional climate models with respect to their ability to simulate realistic temperature and precipitation patterns. Climate datasets are also the main tool to assess the magnitude of climate change and variability during the past 150 years. They serve as the baseline climate for any future climate. To understand e.g. the effect of climate on land use, the availability of suitable meteorological datasets is crucial to calibrate and tune both climate models and impact models on the variability of the current climate.

In contrast to what is generally believed, the frequencies of violent weather events are hardly known, even for the present state of the climate. Datasets with high-resolution climate data are still scarce, though the number of datasets with daily resolution is increasing (Klein Tank et al, 2002). In the light of safety design levels for infrastructural works, it is of utterly importance that quantitative information about this becomes available. The Netherlands is among the few countries where data of sufficient length and temporal resolution is available to empirically estimate this information, but the bulk of this data is only in the form of hard copies.

The objective of the CS8-project ‘Time series information’ is to create digital datasets with high-resolution historical climate data, which at present are only available as hardcopy. The following sources have been selected based on their relevance for climate research and impact communities:

1. Daily precipitation amounts in the 1850-1950 period
2. High resolution (5-minute) 20th century precipitation amounts from rainfall strip charts
3. Daily pre-1850 meteorological observations
4. Hourly observations of Batavia/Jakarta in the 1866-1980 period.

The following chapters describe these data sources, their digitization, quality-control and, where applicable, analysis and homogenization. All data are freely available at the KNMI website ([1]http://www.knmi.nl/klimatologie).

2. Daily precipitation amounts 1850-1950

This chapter describes the creation of the dataset of daily precipitation amounts in The Netherland for the 1850-1950 period. Post-1950 data was already digitally available at KNMI. In January 2009 KNMI disclosed this data on the Internet where it can be downloaded freely. The current dataset extends these data a hundred years back in time with about 12300 station years (4.5 million measurements). Special attention is given to the homogenization and analysis of this dataset.

2.1 KNMI manual rain gauge network in the Netherlands

The data for this part of CS8 originates from the KNMI manual rain gauge network in the Netherlands. This network is maintained since 1850. Incidentally rainfall measurements were taken before that time, with the oldest known record the one of Cruquius in Delft (1715-1725).

Figure 2.1 shows the number of rain gauges in the Netherlands since 1850. The figure shows a gradual increase in the number of rain gauges in the period 1850-1950 and an almost constant network size since 1951. Figure 2.2 shows the geographical distribution of the current rain gauges (about 325). The average distance between the station is 9.9 km. The rain gauges are operated by voluntary observers. Each morning the 24 hour (0800-0800 UTC) precipitation sums are measured and since 1995 digitally transferred to KNMI by telephone.

For the 19th century data little is known about the local situation of the rain gauges and the type of rain gauge used. In 1903 a standardization was carried out with the introduction of a standard rain gauge. The rim of this rain gauge was at 1.50 m above ground level and its funnel had an area of 400 cm^2.

1 KNMI disclosed almost all of its climate data freely on the KNMI website during the term of the KvR projects. At the time of writing, the disclosure of the data digitized within CS8 has not been finished yet, but is expected to be ready by the end of 2011.

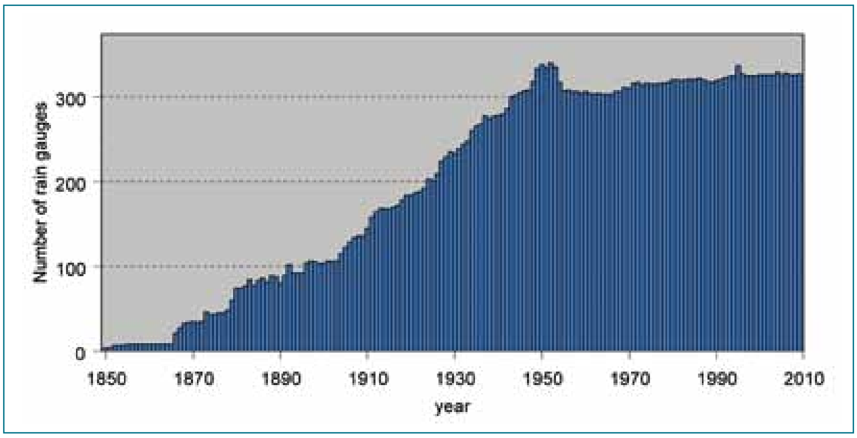


Figure 2.1.
Number of rain gauges in the KNMI manual rain gauge network for the period 1850-2009.

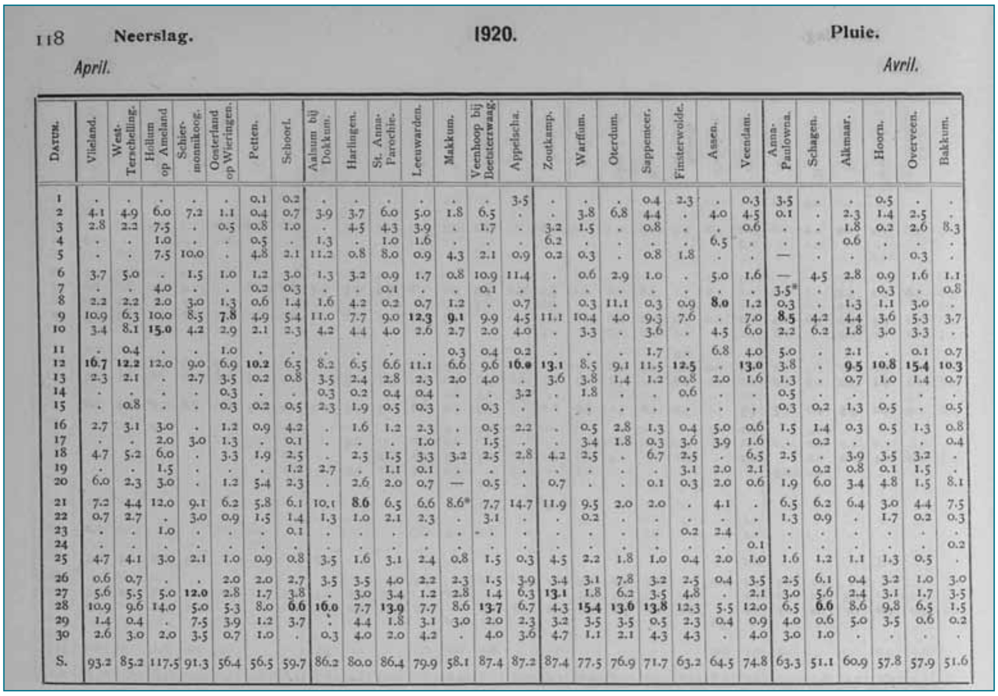
Figure 2.2.
Geographical distribution of the operational rain gauges in the KNMI manual rain gauge network.

Due to an extensive study of Braak (1945) on the wind-induced error of precipitation measurements, KNMI started in 1946 lowering the height of the rain gauges to 0.40 m above ground level. Braak found an average wind-induced error of 5.5% for precipitation measurements at 1.50 m. In January 1947 about 65% of the gauges had been lowered to the height of 0.40 m. The remainder of the gauges was gradually lowered in the next 6 years. Reduction of the measurement height to 0.40 m about halved the wind-induced error. From 1962 onwards, KNMI introduced a new type of rain gauge with a surface area of 200 cm^2 while the height of the rim remained at 0.40 m above ground level. The effect of this change on the measured precipitation amounts is negligible.

Since 1946 the precipitation measurements have been subjected to extensive quality control on a routine basis. Suspect values are traced by comparing the daily measurements with those from neighbouring stations and could often be recovered after consulting the observer. Before 1951 the location of the precipitation stations were inspected infrequently and metadata for that period is scarce. From 1953 onwards stations are regularly visited by stations inspectors and reports of the visits have been stored at KNMI. Further information can be found in Buishand and Velds (1980).

2.2 Data sources

The main hardcopy data source for the digitization of the 1850-1950 daily precipitation amounts is the KNMI yearbooks (see Figure 2.3 for an example). These books contain the meteorological observations of all KNMI stations in the Netherlands. Till 1931 the daily precipitation measurements were also included in these books but from 1932 onwards they were published in separate precipitation yearbooks. The publication of the regular yearbooks ended with the year 1980 and that of precipitation with 1990. The yearbooks are not always exhaustive. Therefore, we sometimes had to go back to more original sources in the KNMI archive. Till 1895 handwritten journals are available and from 1896 onwards also the so-called monthly precipitation charts. Figure 2.4 shows an example of such a chart. The observer noted the daily precipitation amounts on the charts and sent them back to KNMI after the completion of the month. None of the sources is complete, but by combining them we were able to digitize almost the full range of measurements.



118 Neerslag. 1920. Pluie.

April. Avril.

Datum.	Vlieland.	West-Terschelling.	Hollum op Ameland	Schiermonnikoog.	Oosterland op Wieringen.	Petten.	Schoorl.	Aalsum bij Dokkum.	Harlingen.	St. Anna-Parochie.	Leeuwarden.	Makkum.	Veenhoop bij Beetsterzwaag.	Appelscha.
1	.	.	.	.	.	0.1	0.2	.	.	.	.	.	.	3.5
2	4.1	4.9	6.0	7.2	1.1	0.4	0.7	3.9	3.7	6.0	5.0	1.8	6.5	.
3	2.8	2.2	7.5	.	0.5	0.8	1.0	.	4.5	4.3	3.9	.	1.7	.
4	.	.	1.0	.	.	0.5	.	1.3	.	1.0	1.6	.	.	.
5	.	.	7.5	10.0	.	4.8	2.1	11.2	0.8	8.0	0.9	4.3	2.1	0.9
6	3.7	5.0	.	1.5	1.0	1.2	3.0	1.3	3.2	0.9	1.7	0.8	10.9	11.4
7	.	.	4.0	.	.	0.2	0.3	.	.	0.1	.	.	0.1	.
8	2.2	2.2	2.0	3.0	1.3	0.6	1.4	1.6	4.2	0.2	0.7	1.2	.	0.7
9	10.9	6.3	10.0	8.5	**7.8**	4.9	5.4	11.0	7.7	9.0	**12.3**	**9.1**	9.9	4.5
10	3.4	8.1	**15.0**	4.2	2.9	2.1	2.3	4.2	4.4	4.0	2.6	2.7	2.0	4.0
11	.	0.4	.	.	1.0	.	.	.	.	.	.	0.3	0.4	0.2
12	**16.7**	**12.2**	12.0	9.0	6.9	**10.2**	6.5	8.2	6.5	6.6	11.1	6.6	9.6	**16.0**
13	2.3	2.1	.	2.7	3.5	0.2	0.8	3.5	2.4	2.8	2.3	2.0	4.0	.
14	.	.	.	.	0.3	.	.	0.3	0.2	0.4	0.4	.	.	3.2
15	.	0.8	.	.	0.3	0.2	0.5	2.3	1.9	0.5	0.3	.	0.3	.
16	2.7	3.1	3.0	.	1.2	0.9	4.2	.	1.6	1.2	2.3	.	0.5	2.2
17	.	.	2.0	3.0	1.3	.	0.1	.	.	.	1.0	.	1.5	.
18	4.7	5.2	6.0	.	3.3	1.9	2.5	.	2.5	1.5	3.3	3.2	2.5	2.8
19	.	.	1.5	.	.	.	1.2	2.7	.	1.1	0.1	.	.	.
20	6.0	2.3	3.0	.	1.2	5.4	2.3	.	2.6	2.0	0.7	—	0.5	.
21	7.2	4.4	12.0	9.1	6.2	5.8	6.1	10.1	**8.6**	6.5	6.6	8.6*	7.7	14.7
22	0.7	2.7	.	3.0	0.9	1.5	1.4	1.3	1.0	2.1	2.3	.	3.1	.
23	.	.	1.0	.	.	.	0.1	.	.	.	.	.	.	.
24	.	.	.	.	.	.	.	.	.	.	.	.	.	.
25	4.7	4.1	3.0	2.1	1.0	0.9	0.8	3.5	1.6	3.1	2.4	0.8	1.5	0.3
26	0.6	0.7	.	.	2.0	2.0	2.7	3.5	3.5	4.0	2.2	2.3	1.5	3.9
27	5.6	5.5	5.0	**12.0**	2.8	1.7	3.8	.	3.0	3.4	1.2	2.8	1.4	6.3
28	10.9	9.6	14.0	5.0	5.3	8.0	**6.6**	**16.0**	7.7	**13.9**	7.7	8.6	**13.7**	6.7
29	1.4	0.4	.	7.5	3.9	1.2	3.7	:	4.4	1.8	3.1	3.0	2.0	2.3
30	2.6	3.0	2.0	3.5	0.7	1.0	.	0.3	4.0	2.0	4.2	.	4.0	3.6
S.	93.2	85.2	117.5	91.3	56.4	56.5	59.7	86.2	80.0	86.4	79.9	58.1	87.4	87.2

Datum.	Zoutkamp.	Warffum.	Oterdum.	Sappemeer.	Finsterwolde.	Assen.	Veendam.	Anna-Paulowna.	Schagen.	Alkmaar.	Hoorn.	Overveen.	Bakkum.
1	.	.	.	0.4	2.3	.	0.3	3.5	.	.	0.5	.	.
2	.	3.8	6.8	4.4	.	4.0	4.5	0.1	.	2.3	1.4	2.5	.
3	3.2	1.5	.	0.8	.	.	0.6	.	.	1.8	0.2	2.6	8.3
4	6.2	.	.	.	.	6.5	.	.	.	0.6	.	.	.
5	0.2	0.3	.	0.8	1.8	.	.	—	.	.	.	0.3	.
6	.	0.6	2.9	1.0	.	5.0	1.6	—	4.5	2.8	0.9	1.6	1.1
7	.	.	.	.	.	.	.	3.5*	.	.	0.3	.	0.8
8	.	0.3	11.1	0.3	0.9	**8.0**	1.2	0.3	.	1.3	1.1	3.0	.
9	11.1	10.4	4.0	9.3	7.6	.	7.0	**8.5**	4.2	4.4	3.6	5.3	3.7
10	.	3.3	.	3.6	.	4.5	6.0	2.2	6.2	1.8	3.0	3.3	.
11	.	.	.	1.7	.	6.8	4.0	5.0	.	2.1	.	0.1	0.7
12	**13.1**	8.5	9.1	11.5	**12.5**	.	**13.0**	3.8	.	**9.5**	**10.8**	**15.4**	**10.3**
13	3.6	3.8	1.4	1.2	0.8	2.0	1.6	1.3	.	0.7	1.0	1.4	0.7
14	.	1.8	.	.	0.6	.	.	0.5	.	.	.	.	.
15	.	.	.	.	.	.	.	0.3	0.2	1.3	0.5	.	0.5
16	.	0.5	2.8	1.3	0.4	5.0	0.6	1.5	1.4	0.3	0.5	1.3	0.8
17	.	3.4	1.8	0.3	3.6	3.9	1.6	.	0.2	.	.	.	0.4
18	4.2	2.5	.	6.7	2.5	.	6.5	2.5	.	3.9	3.5	3.2	.
19	.	.	.	.	3.1	2.0	2.1	.	0.2	0.8	0.1	1.5	.
20	0.7	.	.	0.1	0.3	2.0	0.6	1.9	6.0	3.4	4.8	1.5	8.1
21	11.9	9.5	2.0	2.0	.	4.1	.	6.5	6.2	6.4	3.0	4.4	7.5
22	.	0.2	.	.	.	.	.	1.3	0.9	.	1.7	0.2	0.3
23	.	.	.	.	0.2	2.4	.	.	.	.	.	.	.
24	.	.	.	.	.	.	0.1	.	.	.	.	.	0.2
25	4.5	2.2	1.8	1.0	0.4	2.0	1.0	1.6	1.2	1.1	1.3	0.5	.
26	3.4	3.1	7.8	3.2	2.5	0.4	3.5	2.5	6.1	0.4	3.2	1.0	3.0
27	**13.1**	1.8	6.2	3.5	4.8	.	2.1	3.0	5.6	2.4	3.1	1.7	3.5
28	4.3	**15.4**	**13.6**	**13.8**	12.3	5.5	12.0	6.5	**6.6**	8.6	9.8	6.5	1.5
29	3.2	3.5	3.5	0.5	2.3	0.4	0.9	4.0	0.6	5.0	3.5	0.6	0.2
30	4.7	1.1	2.1	4.3	4.3	.	4.0	3.0	1.0	.	.	.	.
S.	87.4	77.5	76.9	71.7	63.2	64.5	74.8	63.3	51.1	60.9	57.8	57.9	51.6

Figure 2.3.
Example a table with rainfall data from the KNMI yearbook of the year 1920.

Figure 2.4.
Example of a handwritten monthly precipitation chart.

2.3 Digitization, quality control and homogenization

2.3.1 Digitization

All data were manually keyed into spreadsheets. This labour intensive work was mainly performed by students working part-time on the project. We have considered using Optical Character Recognition (OCR), but found manual keying just as efficient. To facilitate the digitization process, the spreadsheets were prepared in such a way that they reflected the page layout of the yearbooks. Monthly totals were automatically calculated in the spreadsheets to facilitate the comparison with the monthly totals in the hardcopy source material. Discrepancies between the two generally indicated typing errors, which could then be corrected on the spot. Special markups were used for days with missing data and multi-day precipitation amounts.

2.3.2 Quality control

Quality control consisted of two steps. The first step concerned the correction for actual typing errors. Monthly totals of the digitized data were compared with those in the source material as described above. Almost all typing errors could be corrected using this procedure. Incidentally, placement of the digitized daily data within a month was shifted one or more days. In cases where there was a suspicion that this occurred, corrections were made. The resulting data files are, as far as possible, free of typing errors and contain the data as original as possible. These data files will be retained at KNMI and will be known as the 'uncorrected' datasets'.

The second step concerns the distribution of multi-day precipitation amounts and filling in of missing data. Multi-day amounts are sometimes present when the (voluntary) observer was not able to measure one or more days. We developed a procedure to estimate daily precipitation amounts from neighbouring stations using universal kriging. This procedure can be used both for filling in missing days and for the distribution of multi-day amounts. In the latter case, the daily estimates are used to proportionally distribute the multi-day amount over the days concerned. The procedure is described in detail in Jilderda (2011).

Before performing the second quality control step, all stations were put together in a one table of the KNMI database system. A metadata file was created consisting of station numbers, names, positions and start and stop dates. Sometimes there was confusion about stations names and numbers but most problems could be solved by comparing the several hardcopy data sources of the measurements.

2.3.3 Homogenization

For this data source we planned to use the complete daily precipitation dataset (1850-2009) for studying long-term trends in extreme rainfall. Because inhomogeneities may affect these trends, we decided to homogenize[2] the data to get rid of inhomogeneities for instance caused by station relocations. Homogenization of datasets with a large number of time series is, however, time consuming and is generally only feasible when some kind of automatic method is available. We therefore started a cooperation with the NOAA/National Climatic Data Center in the US. The NOAA group developed software for objectively homogenizing a large number of temperature time series by the method of pairwise comparisons (Menne and Williams, 2009). We implemented the software at KNMI and adapted it for use with the rainfall data. This is described in detail in Buishand et al. (2011).

For the study of inhomogeneities it is important to obtain knowledge about station relocations, changes in the nearby environment of the rain gauges, etc. This kind of information is available in reports made by station inspectors, though the information is scarce before 1951. Since 1 January 2004 KNMI stores these reports in a digital archive but before that time only hardcopy report were available. In the course of the CS8 project, we made digital images of all these hardcopy reports (see e.g. Brandsma, 2008) to facilitate their access. KNMI is now in the process of entering the most relevant data into a meta-database.

2.4 Deliverables and examples of results

2.4.1 Deliverables

The main deliverables of this part of the project are:

1. A dataset with all daily precipitation time series in the Netherlands for the 1850-1950 period (12300 station years, 4.5 million measurements). This quality controlled data is freely available via the KNMI website, along with the post-1950 data which was disclosed in January 2009.
2. A set of digital images of all KNMI (precipitation) yearbooks. The images are available from the KNMI website and enable the user of the data to go back to the main hardcopy source of precipitation data.
3. A set of digital images of all reports made by station inspectors. These images are available upon request for each station. The information in the reports most relevant for the quality of a stations series, is now be entered into a meta-database. For each station, this information will be made available via the internet in the coming years.
4. Publications on the: (a) homogenization and analysis of the data (Buishand et al, 2011), (b) estimation of missing values and distribution of multi-day precipitation amounts (Jilderda, 2011), and (c) digitization and scanning of historical climate data (Brandsma, 2008).

2 A numerical series representing the variations of a climatological element is called homogeneous if the variations are caused only by fluctuations in weather and climate (Conrad and Polak, 1962).

2.4.2 Examples of results

Figure 2.5 shows the mean annual precipitation amounts in the 1850-2009 period using all available stations (without homogeneity corrections). A clear upward trend is visible.

After homogenization among others the following indices were calculated from the data: (a) annual precipitation; (b) precipitation in the summer half-year; (c) annual number of days with precipitation greater than 20 mm; (d) annual number of days with precipitation greater than 30 mm. Figure 2.6 shows the spatial distribution of the trends of these indices for the 1910-2009 period. Homogenization of pre-1910 data was not feasible with the current method because of the small number of stations and the decreasing quality of the measurements. The figure is based on the records from 102 stations, which had no more than 5% missing daily precipitation over the whole 100-year period.

Increases of 200-250 mm (30-35%) in mean annual precipitation are found in regions along the western and northern coast, whereas the increase ranges between 70 and 160 mm (10-20%) along the eastern border and the southeast of the country (Figure 2.6a). Parts of the coastal region show increases in mean summer precipitation between 90 and 120 mm (25-30%, Figure 2.6b). The trend is much smaller in the eastern part of the country but positive everywhere. The trend in the number of exceedances of the 20 mm threshold is also always positive (Figure 2.6c). For a large part of the country increases of 2 to 3 days (70-105%) are found. The area around The Hague and Rotterdam in the west of the country shows the largest increase in the number of exceedances of the 30 mm threshold (up to about 1.5 days (Figure 2.6d) or about a threefold increase). Most of the stations exhibiting abnormal trends in the annual precipitation amounts also have exceptional trends in the exceedance frequencies of the 20 mm and 30 mm thresholds.

Figure 2.7 compares the linear fit with the loess smoother (Cleveland, 1979) for the number of exceedances of the 30 mm threshold shown. The loess smooth for this index has a relative minimum around 1980 which is outside the 95%-confidence band. There is a rather strong increase in the number of days after 1980 according to the loess smooth. This increase amounts to 78% of the total change over the 1910-2009 period. The figure shows that in the year 1998 there are on average 2.8 days with precipitation larger than the 30 mm threshold. Omission of that point leads to only a marginal change in the shape of the loess.

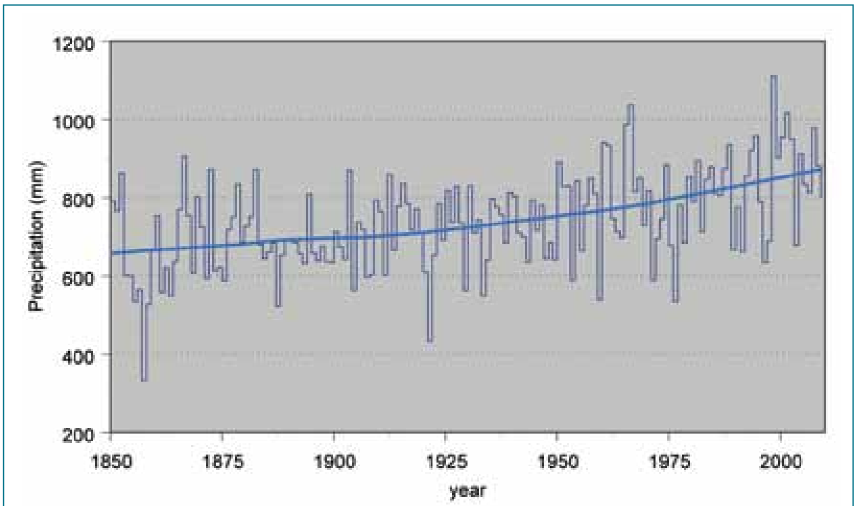


Figure 2.5.
Mean annual precipitation amounts in the Netherlands for the period 1850-2009. The smooth line is a loess smoother with span 0.45.

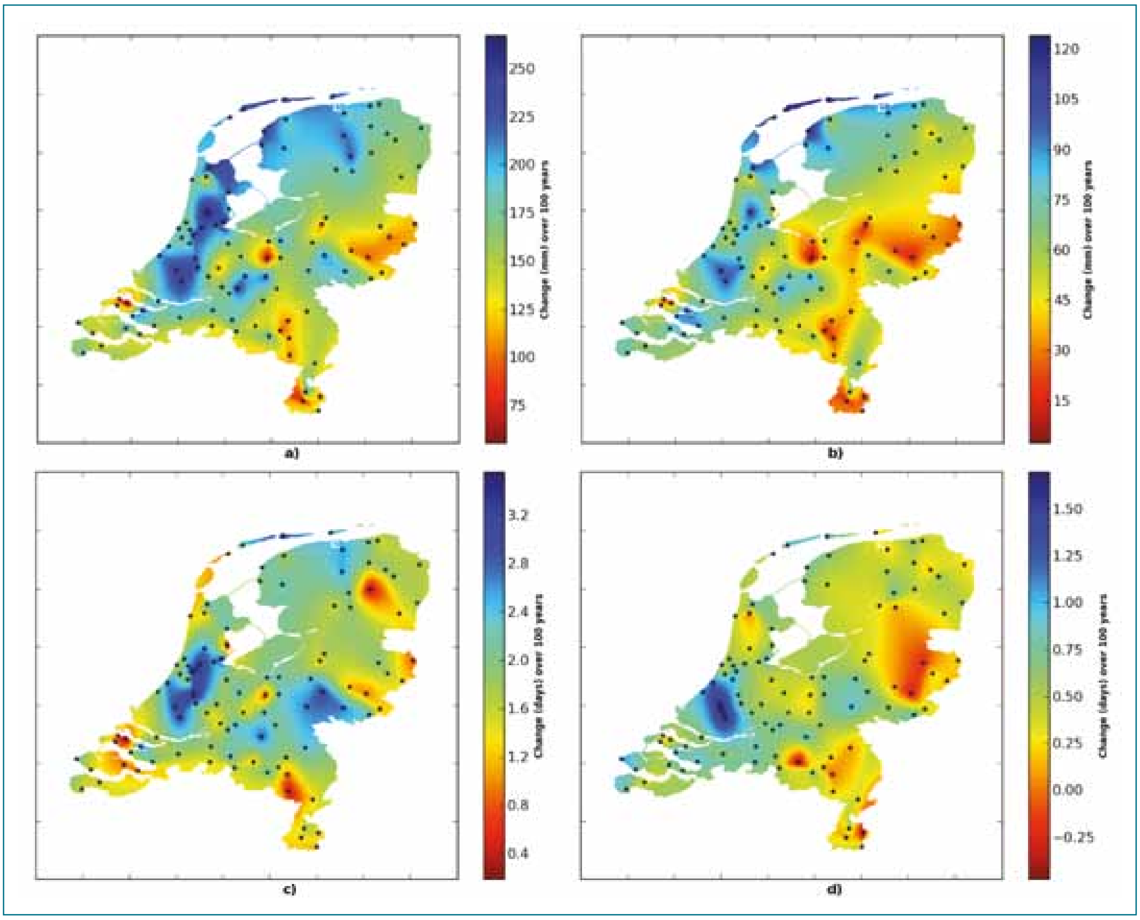


Figure 2.6.
Changes in the mean of four precipitation indices over the period 1910–2009: (a) annual precipitation (mm); (b) precipitation in the summer half-year (mm); (c) annual number of days with precipitation greater than 20 mm; (d) annual number of days with precipitation greater than 30 mm. Black dots indicate the station positions.

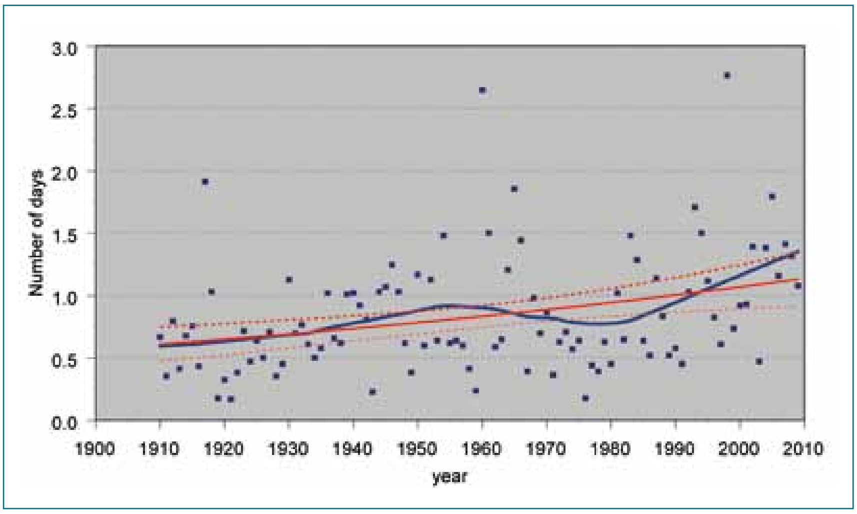


Figure 2.7.
Annual averages of the number of days with precipitation greater than 30 mm for the period 1910–2009. The blue curve is based on a loess smoother with a span of 0.45. The red dashed lines mark a pointwise 95% - confidence band for the linear trend (red line).

The increase in annual mean precipitation is statistically significant for almost all stations (97%). Increasing trends in the Dutch precipitation were also mentioned in the KNMI climate reports. The increase found here is much larger than the 85 mm found by Van Boxel and Cammeraat (1999) for the precipitation in the Netherlands for the 20th century. These authors use, however, only five precipitation series over the 1904-1998 period, where part of the data, from 1980 onwards, were derived from automatic rain gauges. The automatic rain gauges underestimate precipitation to about 5%. This partly explains the lower values of Van Boxel and Cammeraat. In addition, they use only a slightly different period (1904-1998 instead of 1910-2009).

The increase in the summer half-year precipitation is statistically significant along the coast. Closer inspection reveals that the trend can mainly be attributed to strong precipitation increases since 1980. This trend may be related the increase of about 1°C in the North Sea temperature since 1951 (Lenderink et al., 2009).

The winter precipitation (not presented here) also shows statistically significant increasing trends. Increasing trend in winter precipitation have also been found in Belgium (Schmith, 2001) the western part of Germany (Rapp and Schönwiese, 1995) and other parts of Europe (Moberg and Jones, 2005).

2.5 Possible applications of the data

Together with the post-1950 data, the daily precipitation dataset produced here has many potential applications, among others:

1. Calculation of design standards for dikes, canals and drainage systems by Ministries, waterboards, municipalities and consultancies. In general the estimates of design levels improve with increasing length of the time series.
2. Assessment of the historical adaptive capacity of the water systems to the observed precipitation changes, which are larger than the projected changes due to climate change.
3. Use as base-line climate in climate impact studies.
4. Validation of (regional) climate models. It is of interest to know how well these models are able to simulate e.g. the observed trends in extreme indices in the past century.
5. Assessment of precipitation trends and variability and their areal distribution.

The data has not been used in the other KvR projects, mainly due to time constraints. The data is, however, available for use in the new research program KvK (the successor of KvR).

3. High resolution precipitation from rainfall strip charts

This chapter describes the creation of the dataset with 321 station years of 5-minute precipitation from 20th century rainfall strip charts and paper rolls. Such high-resolution precipitation values are important for the design of fast responding systems like sewerages and tunnels. Because of the labor-intensive extraction work, we developed an automatic curve extraction (ACE) algorithm for processing bulk quantities of rainfall strip charts and rolls. The framework largely automates the processing of rainfall curves on the strip charts and paper rolls and can be applied to other elements as well.

3.1 Continuous precipitation recording in the Netherland

Self-recording rain gauges (see Figure 3.1 for an example) have been applied for continuous rainfall measurements at a selected set of KNMI stations since the end of the 19th century. At first, rainfall was recorded on daily and sometimes weekly rainfall strip charts. Thereafter, from about 1980 through 1993, paper rolls were used containing the rainfall registration for about 10-20 days. Finally, from 1994 onwards rainfall measurements are transferred electronically and operationally stored at 10-minutes resolution (for some selected stations at 1-minute resolution). Until now, the strip charts and paper rolls have been used mostly for extracting hourly values[3]. In infrastructural design (e.g. sewer systems, tunnel drainage) there is, however, a need for long high-resolution rainfall series with resolutions much higher than 1 hour. Fortunately, the charts and rolls can be used to extract this kind of information with a time resolution of about 5 to 10 minutes.

Figure 3.1.
Example of a self-recording rain gauge (without lid).

3.2 Data sources

In total 321 station years were digitized, consisting of the following 6 stations: (a) De Bilt (1897–1993), (b) Eelde (1954–1993), (c) Den Helder/De Kooy (1954–1993), (d) Vlissingen (1954–1993), (e) Beek (1954–1993) and (f) Amsterdam (1920–1983).
The total number of daily rainfall strip charts in our archive is about 110,000 and the number of rolls about 5,000. Figure 3.2 shows a an example of a daily strip chart for the station De Bilt. The typical size of strip charts is 40 by 8 cm. In contrast to daily strip charts, rolls contains on average 15 days and can be up to 40 m in length with heights varying between 14 and 29 cm. The length of a single day on a roll is about 1.5 m. Figure 3.3 gives an example of piece of a roll for De Bilt.

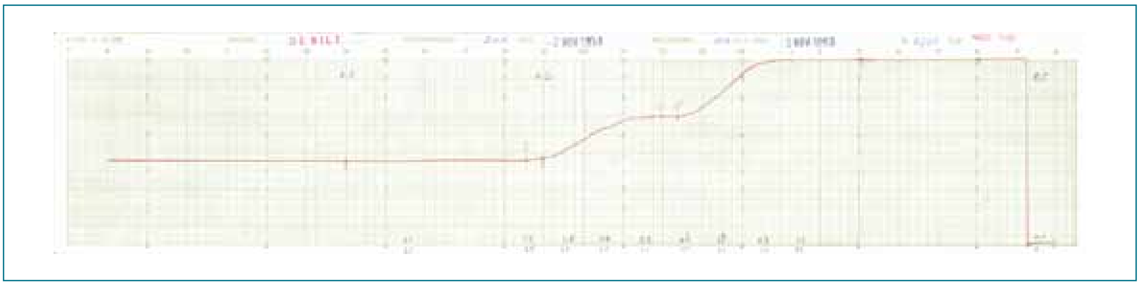


Figure 3.2.
An example of a daily rainfall strip chart for De Bilt.

3 An exception is the dataset with 15-minute precipitation amounts for Eelde, De Bilt and Beek for the period 1955-1979 and a 5-minute time series for De Bilt with a length of 12 years (1928, 1933, 1951-1960).

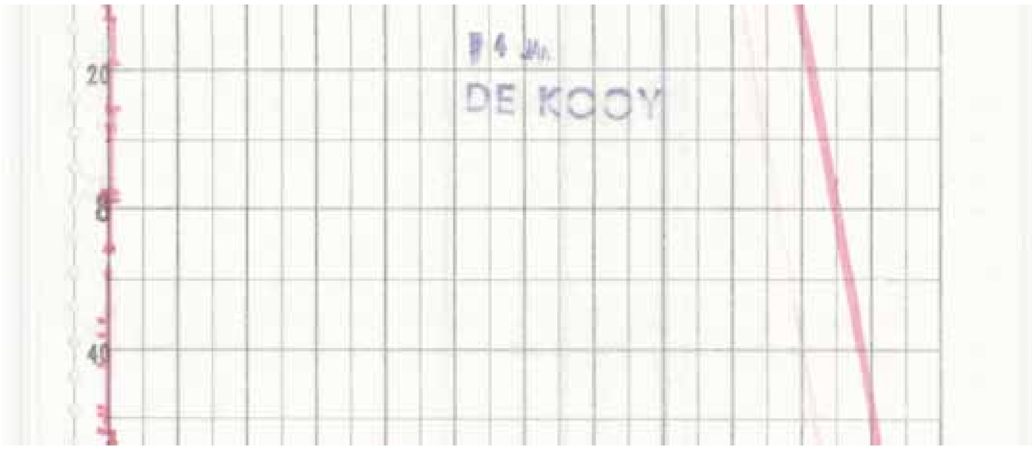


Figure 3.3.
Example part of a roll for De Kooy. The complete rol is almost 250 times as large in vertical direction.

In Figure 3.2 several features of a strip chart can be distinguished: on the paper-colored background a green colored grid is printed with a red colored curve on top of it representing the cumulative rainfall amount. On top of the chart, the name of the station (De Bilt) and two dates are stamped. The first date represents the date the strip was positioned into the pluviograph, the second corresponds to the date the strip was taken off. The exact times are usually handwritten alongside the stamps. In general, the times are fixed at 8:00 UTC, but variations of some minutes sometimes occur. Note that the start and end time of recording are not necessarily equal to the grid start and end time. Also visible are some markings written down by an observer or analyst after the recording took place. On the far right of the strip, a so-called tip-over moment is visible as a nearly vertical line drawn in the red curve color. This line corresponds to the time when the pluviograph automatically emptied its reservoir (here when the reservoir contains 10 mm of rainfall). In rare cases, the emptying was done by the observer. The tip-over moments are a common feature and need to be detected in order to successfully extract the rainfall data.

In contrast to charts, rolls (Figure 3.3) have no fixed times for placement and removal. Similar stamps or annotations of measurement location are mostly visible only at the beginning or end of the roll together with a written date and the recording times. The basic information on rolls is identical to that on charts. There is large variation in the layout of both strips and rolls. The most notable variation is in the colors used for the grid and the curve on the strip charts. Grid color is most often green, but in early years red and black were also applied. Red and blue are the most commonly used curve colors. In rare cases (0.2%), the grid and curve have almost the same red color and are therefore hardly distinguishable. This complicates the detection process. The background (paper) color is almost always close to white. However, on very old charts this has darkened to a yellow-brownish color.

With respect to the grid and the curve, the rolls have a similar appearance to the charts, with the exception that the grid is oriented vertically instead of horizontally. The color scheme only slightly varies. The grid is mostly dark gray colored on a light gray background with a red colored rainfall curve on top of it. On the roll images, the day transitions correspond exactly to the 24 hour markings visible on the graph. In some cases, time markings are visible that are manually drawn by an observer. These time markings can be of any time and can be located anywhere on the roll independent of the underlying grid. Furthermore, in some cases, a roll contains two separate grids, e.g. for recording another meteorological variable. Therefore, the upper and lower boundary of the correct grid must be known as well. Since for rolls the exact positions of the grid boundaries and the type and position of markers are not known in advance, manual pre-processing is necessary.

A variation that appears for strip charts only is the presence of curved vertical axes. This occurs for the so-called balance type pluviographs that were occasionally used in the first part of the 20th century. The graph start and end times may also vary through the years. A meta-database was constructed defining for each measurement location the periods with identical features. The start and end times of the graph and the presence of a curved axis are derived from this database.

3.3 Automatic curve extraction (ACE)

3.3.1 Introduction

At the beginning of the project, we studied the feasibility of using existing curve extraction software for digitizing the strip charts and paper rolls. Several commercial and free software packages are available for that purpose. However, these packages are meant to trace one line at a time and require manual navigation, such as indicating the position of the axes, indicating the beginning and the end of the curve to be traced and specifying the color of that curve. The packages also have difficulties handling background grids and curved axes, which are common for meteorological strip charts. Consequently, they were not suitable for processing large amounts of strip charts.

Alternatively we developed a framework for automatic curve extraction (ACE).
Figure 3.4 shows the five basic steps of the ACE framework: (1) image scanning, where the strip charts or paper rolls are scanned to digital images, (2) pre-processing, consisting of metadata extraction and defining day transitions on the roll images, (3) applying the automatic curve extraction (ACE) method to determine the coordinates of the lines on the images (all steps within the dashed box), (4) result post-processing with a graphical user interface to visually inspect the results of the curve extraction method and to correct possible errors resulting from step (3), and (5) result aggregation which translates the coordinates of cumulative rainfall curves to time series with the desired time resolution.

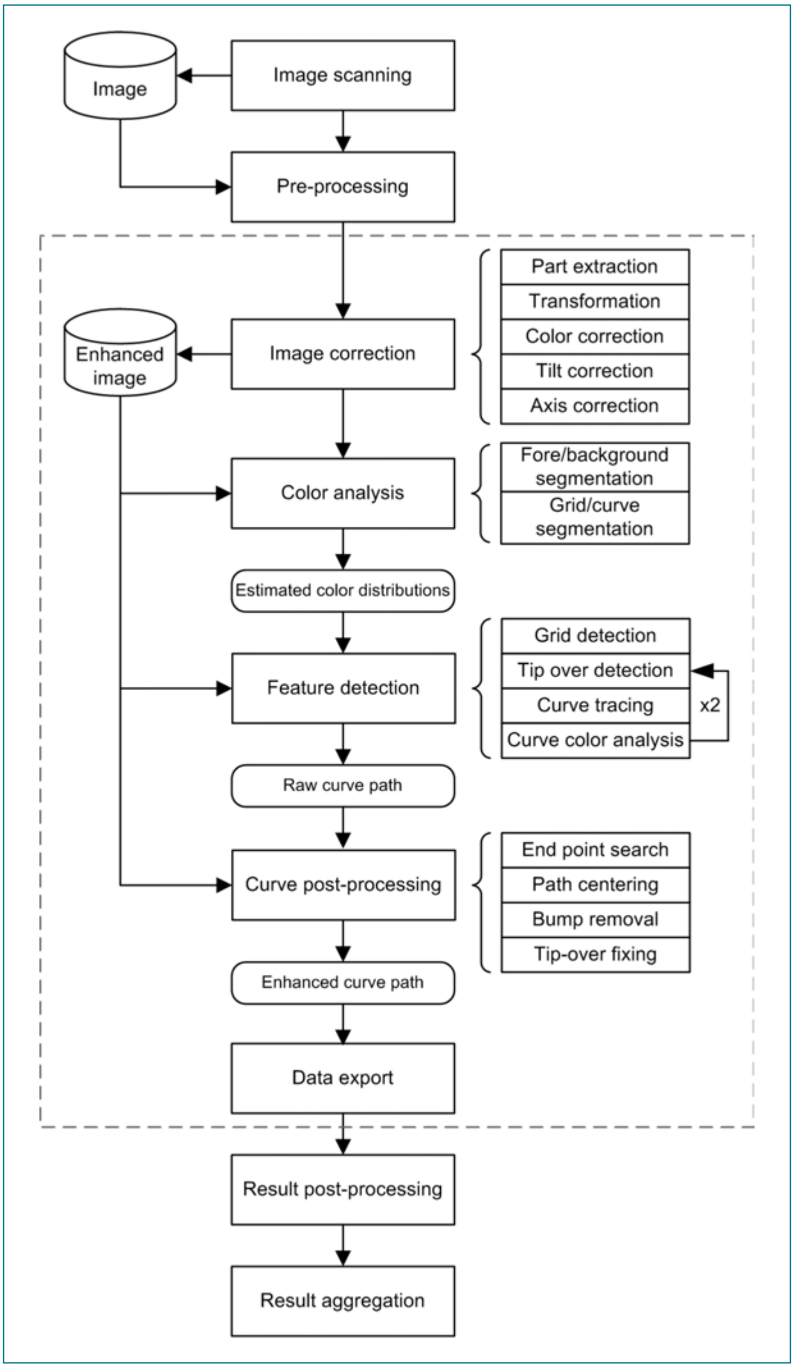


Figure 3.4.
Schematic overview of the automatic curve extraction framework. The part within the dashed box is fully automatic.

Below we briefly describe the steps of the ACE framenwork. A detailed description is given in Van Piggelen et al. (2010, 2011).

3.3.2 Image scanning and preprocessing

For the scanning of the strip charts and paper rolls we purchased two scanners: a Canon DR5010C transit scanner and a Contex Chameleon G600 scanner (see Figure 3.5). The Canon scanner was used for scanning the strip charts at a speed of 20 charts per minute at a maximum resolution of 600 dots per inch (dpi). The Contex scanner was used to scan the long paper rolls. This scanner is designed for scanning large image sizes and scans rolls with a speed of 150 cm/min at 300 dpi resolution. Using these resolutions, a number of 10 to 15 pixels per curve and grid line thicknesses is achieved, which is sufficient for ACE.

Figure 3.5.
Canon DR5010C transit scanner used for scanning the strip charts (left) and Contex Chameleon G600 scanner used for scanning the long paper rolls (right).

For each measurement location, periods with identical features (e.g. the start and end times of the graphs and the presence of a curved vertical axis) have been identified. This information was stored in a meta-database and used later on during the processing of the images. A property of the rolls is that they contain multiple day recordings in a sequence. For a correct detection of the rainfall values on the rolls, the exact start and end date and time of the roll recording are needed, as well as the exact locations of the day transitions and the upper and lower grid boundaries. Sometimes, only arbitrary time markings made by the observer were available. Before the application of ACE, the roll images were therefore digitally cut to make sure that each day on these images was processed correctly. To avoid the error-prone automatic detection of these lines and markings, the day transitions and grid boundary positions were identified manually. To facilitate this work, we developed a special application, called Roller, capable of handling the large bitmap images and providing the necessary tools to accomplish the marking of the marker and grid position locations.

Besides the actual rainfall information, the strip charts and rolls also contain (written) metadata relevant for further use such as station name, date, problems with the rain gauge, etc. For each strip chart and paper roll, we extracted this type of information by visual inspection and importing it into a metadatabase.

3.3.3 Application of automatic curve extraction (ACE)

We developed the automatic curve extraction (ACE) method to determine the coordinates of the cumulative precipitation curves on the digital images (all steps within the dashed box of Figure 3.4). ACE uses a color detection procedure that automatically separates the background of the charts and rolls from the grid and subsequently the precipitation curve. The precipitation curve is then detected by the minimization of a cost function.

ACE is suitable for the detection of only a single curve on a strip chart image. In rare cases, multiple curves were present and manual post-processing was needed. Furthermore, ACE assumes that the strip charts are positioned correctly onto the pluviograph during recording. Without this assumption there is no frame of reference and precise correction becomes problematic. For the charts and rolls processed here, there is no indication that strip charts were incorrectly positioned. For charts with curved vertical axes, we assumed that the curvature corresponds to the length of the arm of the pen. Correction of this curvature will then correctly reposition the rainfall curve line onto the now rectilinear grid. ACE is described in detail in Van Piggelen et al. (2011).

3.3.4 Post processing and results aggregation

The results of ACE are examined using a software program called PostACE. We developed this tool to visually inspect the results of the automatic extraction and to make adjustments where needed. Figure 3.6 shows and example screen of PostACE.

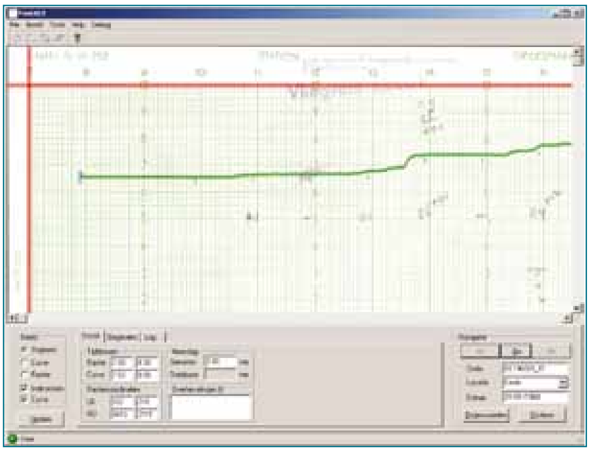


Figure 3.6.
Example screen of the post processing tool PostACE, used for visually inspecting the results of ACE and for making corrections (if needed).

If necessary, the extraction with ACE can be repeated within PostACE using user-defined image features, such as altered tip-over moment positions. In case the trace still does not match the exact position of the curve found on the image, correction lines can be drawn to force the vertical position of the curve points across a specified interval. In addition, polygons can be drawn to exclude the enclosed pixels during the curve tracing phase by setting the corresponding probability values to zero (background).

The final step is the combination of the extraction results of the individual images into a precipitation intensity table with arbitrary time resolution and time range. For this task we wrote the program ParseTracks. Here, the time resolution was set to 5 minutes, corresponding to a width of roughly 30 pixels for the strip charts and 60 pixels for rolls. Using the start and end points of each curve with corresponding times, the exact pixel positions of the 5 minute intervals were determined so that each interval exactly matches one of the 12 intervals in an hour. The actual difference in pixel height within each interval corresponds to the precipitation intensity during that period. In addition, the resulting hourly and daily precipitation sums, as calculated from values 5-minute values, were compared with the already known values to search for suspect values. No attempt has been made to transform the 5-minute precipitation such that their sums match the known values of the hourly and daily precipitation amounts. The daily values from the manual gauges have, however, been used to check the plausibility of the daily sums of the 5-minute precipitation.

3.3.5 Homogeneity

The 5-minute precipitation series may contain inhomogeneities due to e.g. changes of instruments, changes in measurements practices or slow or abrupt changes in the surroundings. We did not correct for these kind of inhomogeneities. At present there are no techniques available to make the corrections. The same holds for lower resolution series like hourly or daily series. The difficulty is that the magnitude of the inhomogeneites depends on the weather conditions at the moment of the measurement and that extensive parallel measurements are needed to find transfer functions that relate the old situation to the new. These parallel series are seldom available.

The 5-minute values are also not corrected such that their hourly and daily sums correspond to the already available hourly and monthly sums. It is known that precipitation from pluviographs systematically underestimates the precipitation measured with the manual rain gauges (see Section 2). For instance, the pluviograph automatically siphons when it contains 10 mm of precipitation. In the past, this could take about 15 sec during which no precipitation was measured (Muller and Van Londen, 1983). Depending on the rainfall intensity at the moment of siphoning, this could lead to significant losses of precipitation of about 5% or larger during an extreme event.

Till 1982 KNMI corrected the hourly precipitation sums from strip charts in such a way that their daily totals corresponded to the daily sums of the nearby manual rain gauge. In practice this meant that for each day a correction factor was calculated that was then applied to al hourly totals of that day. However, from 1982 onwards corrections have no longer been applied. The main reason that we refrained from corrections here is, as stated above, the weather dependence of the correction. A constant correction factor does not account for that. Another reason is that today the automatic pluviographs are not situated nearby the manual rain gauges. For instance, in De Bilt the manual rain gauge moved in 1995 from the KNMI measurement field to the yard of a farm 450 m southwards. Due to the large variations of rainfall in space and time, the measurement of the automatic pluviograph and the manual rain gauge are no longer comparable.

3.4 Deliverables and examples of results

3.4.1 Deliverables

The main deliverables of this part of the project are:

1. A dataset with 321 station years of 5-minute precipitation time series in the Netherlands for the stations: (a) De Bilt (1897–1993), (b) Eelde (1954–1993), (c) Den Helder/De Kooy (1954–1993), (d) Vlissingen (1954–1993), (e) Beek (1954–1993) and (f) Amsterdam (1920–1983) the 1850-1950 period. This quality controlled data is freely available via the KNMI website.
2. A set of high-resolution digital images of the rainfall charts and paper rolls (1.4 Tb in total). These images are available upon request for each station.
3. The software framework inclusive documentation, consisting of: Roller, ACE and PostACE. All programs use common code for the various overlapping functionalities. The C++ code of the programs can be adjusted to suit the specific needs of each user. The code is available upon request.
4. Publications on the: (a) automatic curve extraction for digitizing rainfall strip charts (Van Piggelen et al., 2011), (b) automatic curve extraction (ACE) framework documentation (Van Piggelen et al. 2010), and (c) digitization and scanning of historical climate data (Brandsma, 2008).

3.4.2 Examples of results

The overall detection quality of the ACE algorithm is good. In 70% of the total number of days processed the process completed without erroneous results. If detection errors were noted during post-processing, merely an adjustment of one or more of the found grid boundary locations was necessary in most cases. In special cases, usually during days with high-precipitation intensity, relocation of the found curve path was sometimes necessary. Generally, the processing of a single day strip took 30-40 seconds on a Pentium 4 3.0Ghz machine. Depending on the presence and severity of the errors, post-processing usually takes no longer than 1 minute and on average can be done within a fraction of that.

After post processing the data, the results have been aggregated to 5-minute rainfall intensity values. In the case of Beek this yields a file with 36 years of 5-minute data. These data can be used, e.g., to perform extreme value analysis for different durations. As an example, Figure 3.7 shows the Gumbel plots of the 10-min and 60-min rainfall amounts of Beek. For each rainfall amount, the figure shows the corresponding return period *Tr*.

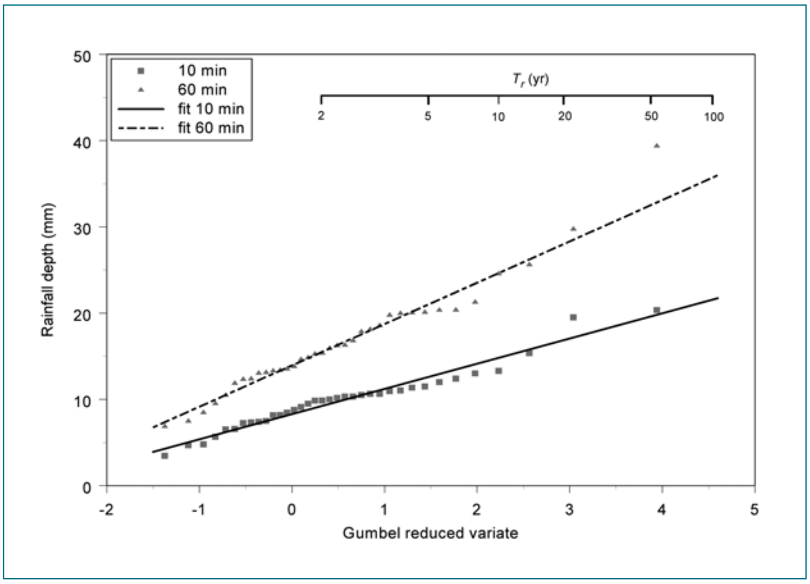


Figure 3.7.
Gumbel plots for the 10-min and 60-min annual maximum rainfall amounts for Beek (1955-1990). The solid and dashes lines are maximum likelihood fits to the observed data. The annual maxima are plotted using the median plotting position.

3.5 Possible applications of the data

Together with the post-1993 high-resolution data, the 5-minute precipitation dataset produced here has many potential applications, among others:

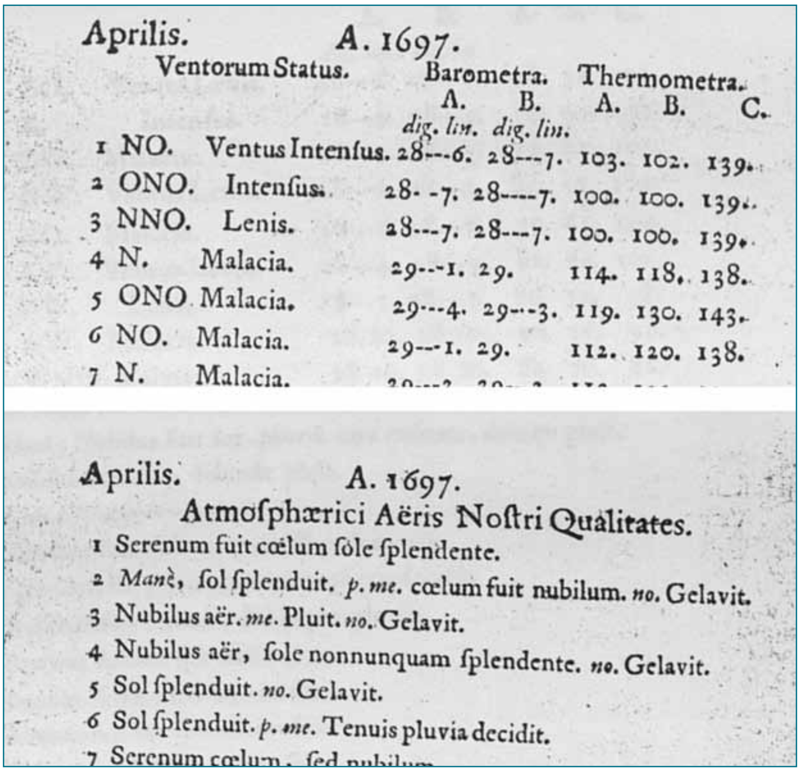
1. Calculation of rainfall depth-duration-frequency curves and design-storm estimation in hydrological analysis and engineering design. These type of analyses mainly use the annual maximum amounts of the rainfall time series for various durations. A peak-over-threshold approach is also possible.
2. Simulation of e.g. urban drainage. Complete high-resolution time series are needed for that purpose.
3. Validation of climate models and calibration of stochastic rainfall models.
4. Use as base-line climate in climate impact studies.
5. Assessment of trends and variability of high resolution precipitation.

4. Daily pre-1850 meteorological observations

This chapter describes the digitization of part of the old pre-1850 meteorological observations in the Netherlands, dating back to about the year 1700. For the reconstruction of the pre-1850 climate, it is important to have as many as possible parallel time series available. In previous projects a large part of these data has already been digitized and published on the KNMI website. The CS8-project adds a signicant number of stations to these data. The digitization of this kind of data is labor-intensive because of the mostly hand-written base material and the non standard way of observing like the use of old units and standards, irregular observation times and (sometimes) the use of Latin as writing language.

4.1 Pre-1850 meteorological observations in the Netherlands

Meteorological measurements in the Netherlands started around the year 1700. Figure 4.1 shows an example of one of the earliest daily observations. The figure shows two snapshots of observations in April 1697 from the meteorological diary of Wolferd Senguerd (1646–1724). In contrast to most pre-1850 observations, these observations were printed instead of hand-written. Note that the daily description of wind speed and weather conditions are in Latin. The figure shows air pressure measurements for two barometers reported in old-style Rhineland inches divided into 12 lines. The figure also shows air temperatures of three thermometers reported in integer values of an unspecified unit (the units differ among the thermometers).


Aprilis. A. 1697.

	Ventorum Status.		Barometra. A. dig. lin.	Barometra. B. dig. lin.	Thermometra. A.	Thermometra. B.	Thermometra. C.
1	NO.	Ventus Intenſus.	28---6.	28---7.	103.	102.	139.
2	ONO.	Intenſus.	28--7.	28----7.	100.	100.	139.
3	NNO.	Lenis.	28---7.	28---7.	100.	100.	139.
4	N.	Malacia.	29---1.	29.	114.	118.	138.
5	ONO	Malacia.	29---4.	29---3.	119.	130.	143.
6	NO.	Malacia.	29---1.	29.	112.	120.	138.
7	N.	Malacia.					

Aprilis. A. 1697.

Atmoſphærici Aëris Noſtri Qualitates.

1 Serenum fuit cœlum ſole ſplendente.
2 *Manè*, ſol ſplenduit. *p. me.* cœlum fuit nubilum. *no.* Gelavit.
3 Nubilus aër. *me.* Pluit. *no.* Gelavit.
4 Nubilus aër, ſole nonnunquam ſplendente. *no.* Gelavit.
5 Sol ſplenduit. *no.* Gelavit.
6 Sol ſplenduit. *p. me.* Tenuis pluvia decidit.

Figure 4.1.
Two snapshots from the dairy of Senguerd who preformed daily meteorological observation in Leiden at the end of the 17th century.

Senguerd was professor in natural philosophy at Leiden University and performed his observations and measurements in a house in Leiden whose exact location is unknown. His measurements over the 1697-1698 period (Senguerdius, 1699) have been preserved and have recently been digitized at KNMI. Können and Brandsma (2005) reviewed the daily pressure readings of Senguerd, converted them to modern units, and reduced them to 0°C. The 2-year series runs parallel with the Paris 1665–1713 and London 1697–1708 pressure series. Although the series covers a time span of 23 months only, it can be regarded as a useful addition to the very few pressure series that extend back into the 17th century.

After Senguerd many more people, often professors, doctors and vicars, started carrying out meteorological observations and measurements. This situation continued till after the founding of KNMI by Buys-Ballot in 1854. In spite of the absence of a meteorological institute, the pre-1850 observations were to some extent organized. From the start of the 18th century, a kind of measuring network was established by Petrus van Musschenbroek (1692–1761). He coordinated the measurements of the network and followed the guidelines of the English Royal Society.

In the first half of the 19th century Wenckebach brought together many of the early observations and compared them. About a century later Labrijn (1945) published a PhD-thesis on the climate of the Netherlands during the last two and a half centuries. In his thesis he presented an overview of all of the then known ancient meteorological observations and measurements and reconstructed long time series of several meteorological variables. In the eighties of the 20th century, Van Engelen and Geurts further extended the inventory of Labrijn with additional stations and presented for each station an extensive overview of the available metadata. For 12 stations they digitized the (sub)-daily meteorological measurements. Their work is published in five reports named 'Historische weerkundige waarnemingen' (KNMI-publication 165)[4]. The data itself is freely available from the KNMI website.

In the year 2000 KNMI started the program Hisklim (Historical Climate), defined in Brandsma et al. (2000). Hisklim aims at digitization, database building and homogenization of historical records in the Netherlands. The HISKLIM program-description defines all Dutch data sources that still had to be digitized (status in the year 2000). Many of these sources have been digitized since then, among others the 19th century KNMI yearbooks, a unique long hourly climate series of Amsterdam in the period 1784-1963 (Wallbrink and Brandsma, 2009) and 17th and 18th century maritime logbooks within the scope of the CLIWOC (CLImatological database for the World's OCeans 1750-1850) project (Können and Koek, 2005)[5].

Hisklim defined the remainder of the pre-1850 data sources of land stations in the Netherlands as one of the sources that could only be digitized and published on the Internet with the help of external funds. This remainder of this chapter describes how we undertook this task within the scope of the CS8 project.

4 Within the scope of BSIK-CS8 these reports have been digitized and are available as pdf-files from the KNMI-website at http://www.knmi.nl/klimatologie/daggegevens/antieke_wrn/zwanenburg_literatuur.html

5 See also: http://www.knmi.nl/cliwoc.

4.2 Data sources

All potential data sources (stations) were defined in the Hisklim program description (Brandsma et al., 2000, p.30-31). For some stations no hardcopies were available at KNMI. We contacted several archives in the Netherlands to check the availability of these sources.

Figure 4.2 shows 13 stations that were finally selected for digitization in the CS8 project. It concerns stations with (sub)daily observations of for instance air temperature, air pressure, wind direction and speed. Of all 13 stations (mostly) hand-written hardcopy base material was available.

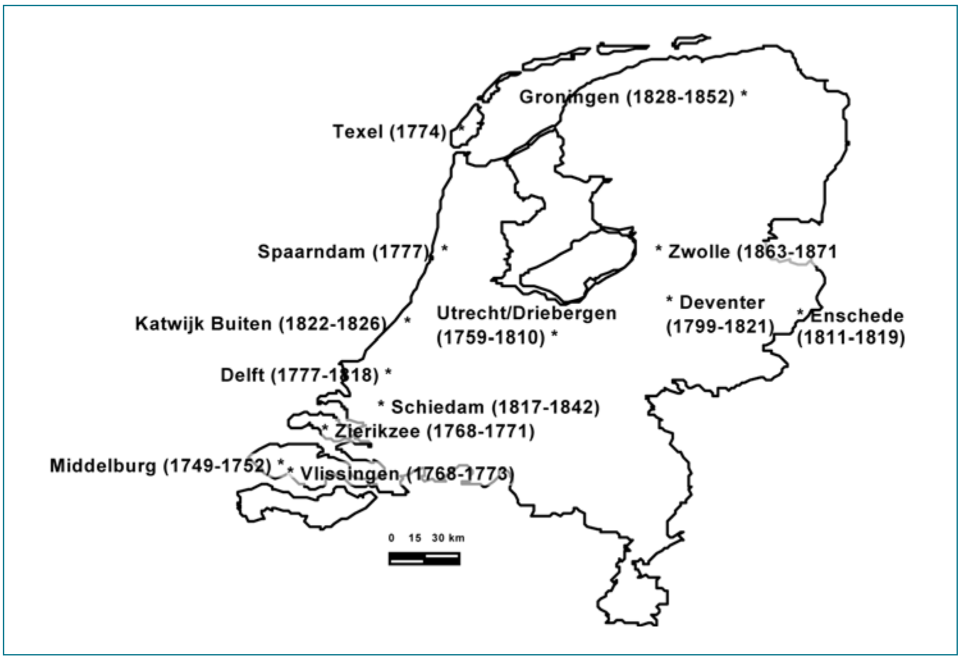


Figure 4.2.
Position of the 13 stations for which the data has been digitized in CS8 including the period for which observations were available.

4.3 Digitization and quality control

4.3.1 Digitization

Before keying the data into spreadsheets, we transformed all hardcopy source material into digital images. Figure 4.3 shows the so-called CopiBook bookscanner that was purchased for this purpose. The scanner can be used for scanning books and old documents containing climate observations and metadata. The scanner has a book cradle to prevent damage of books and scans two color or grayscale pages with a resolution of 300 dpi in 7 or $2^1/_2$ seconds, respectively. It provides, among others, automatic location and cropping of the documents.

Figure 4.3.
CopiBook bookscanner.

The use of digital images has two important advantages compared to the use of hardcopy data. First, the images can be used by more persons at the same time and at any location. Second, together with the digitized data also the images of the original data can easily be provided (e.g. via the Internet). An advantage of the latter is that the user has the possibility to go back to the original documents. Especially for the older handwritten data this may be advantageous. The digital images may serve as an extra backup of the hardcopy data in case of calamities. A disadvantage may be the sustainability of the files. File formats like jpeg of gif will likely change in the future to other formats. This may require file conversions of the original files. In that process errors may easily be made.

For the actual keying of the data we used several working places with two screens next to each other, one with an image and one with a spreadsheet. All data were manually keyed into spreadsheets. As for the other parts of CS8, most of this labour intensive work was performed by students working part-time on the project. The spreadsheets were prepared in such a way that they reflected the page layout of the source material. The work is often complicated by frequent changes in page layouts, number of variables and measuring frequencies. Also the use of Latin delays the digitization.

4.3.2 Quality control

Quality control for this data source was complicated because monthly means or totals were often not available. Typing errors were mainly corrected by visual inspection of the data. The data delivered by the project consist of the data as it is in the original source material. Interpretation and homogenization of the data requires additional research which is outside the scope of the CS8 project. An example of this type work can be found in Können and Brandsma (2005) as mentioned in the beginning in this chapter.

Several factors may influence the homogeneity of the stations series, e.g. station relocations, changes in instrument type or position or changes in the environment. This information is mostly scarce. For the data presented here, Geurts and Van Engelen (1992) present the known potential sources of inhomogeneity.

4.4 Deliverables and examples of results

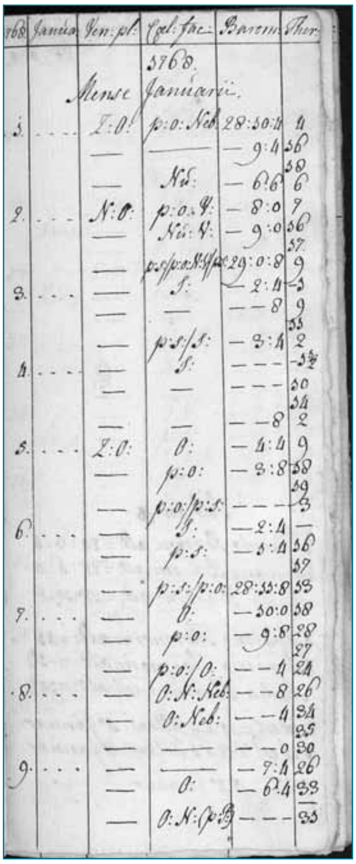
4.4.1 Deliverables

The main deliverables of this part of the project are:

1. Data files with (sub)daily meteorological measurements/observations of 13 stations in the Netherlands in the pre-1850 period. In total it concerns about 150 station years and 110.000 measurements. The data (corrected for typing errors) is freely available via the KNMI website, along with pre-1850 data digitized in previous projects.
2. A set of digital images of the hardcopy source material (available at KNMI) of the pre-1850 stations in the Netherlands. The images are available from the KNMI website and enable the user of the data to go back to the original data source.
3. Searchable pdf files of the reports of Van Engelen and Geurts containing, among others, most of the metadata of the pre-1850 stations. These are freely available from the KNMI website.
4. Publication on the digitization and scanning of historical climate data (Brandsma, 2008).

4.4.2 Example of results

Figure 4.3 gives an example page of the weather diary of Van de Muelen together with the digitized data in the accompanying spreadsheet. His weather observations (partly in Latin) and measurements were taken partly in Utrecht and partly in Driebergen and cover the period 1759–1810.The digitization of this long series constituted a significant part of the digitization of the pre-1850 stations.



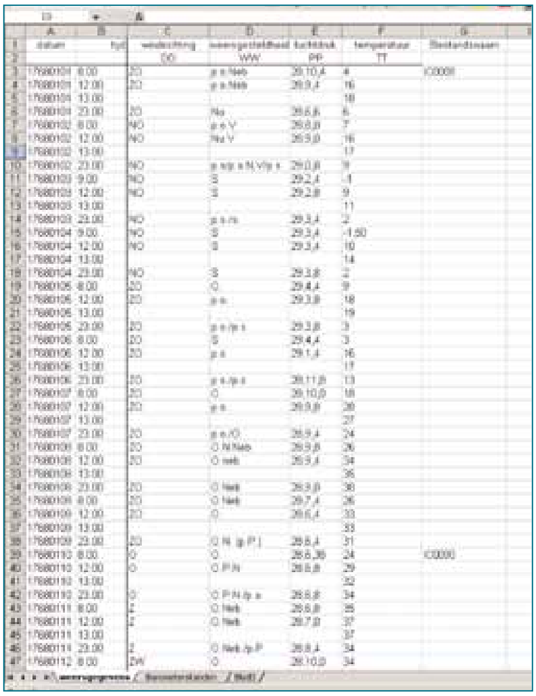


Figure 4.3.
Example page from the weather diary of Jan Carel van der Muelen (left) and the digitized data in the spreadsheet (right). Note that the code in far right column of the spreadsheet refers to the file name of the accompanying image.

As Figure 4.3 (left image) shows, Van der Muelen used abbreviations to express the weather conditions (fourth column from left). These abbreviations pointed to Latin terms. Table 4.1 present the abbreviations together with the accompanying Latin terms and our translation of the Latin.

Table 4.1.
Translation of the Latin terms and their abrevations as used in the Van der Muelen diary.

Abbreviation	Latin term	Translation
T	*Lignificat Tonitra*	Thunder and lightning
P	*Pluviam*	Rain
V	*Ventrum*	Wind
Pr	*Procellam*	Storm wind
C:S	*Coelum serenum*	Clear sky
C:p:S	*Coelum pene serenum*	About clear sky
C:o	*Coelum obducti*	Completely overcast sky
C:p:o	*Coelum pene obductum*	Partly overcast sky
A:B	*Aurora borealis*	Northerly light
C:Nu	*Coelum nubilum*	Cloudy
G	*Grandinum*	Hail
N	*Nixum*	Snow
Neb	*Nebulam*	Mist/fog
P:P	*Parum pluvia*	Little rain
P:c	*Pluvio copiosa*	Abundant rain
P:N	*Parum nivis*	Little snow
N:c	*Nivem copiosa*	Abundant snow
V:f	*Ventum forte*	Strong wind

4.5 Possible applications of the data

The main application of old pre-1850 meteorological observations is in climate reconstruction. For climate reconstruction, it is important to have as many as possible parallel time series available. This allows for the construction of homogeneous time series of weather variables. The parallel time series them selves do not need to be homogeneous for that purpose.

The observations may also be used to study the impact of weather on society, like agricultural production, diseases, floods, etc. An example of this kind of work is the books of Buisman about weather, wind and water in the low countries[6]. The books of Buisman show the impact of weather on the lives of people in Netherlands in the past thousand years.

6 Jan Buisman. Duizend jaar weer, wind en water in de lage landen (part 1-5)

5. Hourly observations of Batavia/Jakarta (1866-1980)

This chapter describes the digitization of the hourly temperature measurements of Batavia/Jakarta in Indonesia. Indonesia is situated in a data-sparse area where long time series are scarce. These series are, however, needed to validate climate models for that area, to construct a baseline climate and to develop climate scenarios. The KNMI archives contain a large amount of hardcopy meteorological observations and measurements for Indonesia for the 1850-1980 period. The remainder of the CS8 budget has been used to digitize a part of this unique set of data. The larger part of the data is being digitized within in a cooperation between KNMI and the Indonesian meteorological office BMKG (see www.didah.org).

5.1 Historical meteorological observations in Indonesia

The earliest meteorological measurements in Indonesia made by the Dutch date from about 1758. Till about 1850 the measurements are fragmentary and their quality is questionable (Braak, 1921). From 1866 onwards hourly measurements were performed at the Batavia observatory (see Figure 5.1). These consisted of, among others: air temperature, air pressure, moisture, wind speed and direction, cloud cover and rainfall. In 1942 Batavia was renamed to Jakarta.

Figure 5.1.
Batavia observatory in the year 1900.

From 1879 onwards a network of daily precipitation stations was set up comparable to the network of daily precipitation stations in the Netherlands (see Chapter 2).

Besides the Batavia observatory and the precipitation network several other meteorological stations were set up and operated, yielding (sub)daily readings of various meteorological measurements and observations.

5.2 Data sources

Table 5.1 gives an overview of the hardcopy Indonesian data available in the KNMI archives. Of these data the hourly temperature data of Batavia/Jakarta were selected for digitization in the CS8 project. The Batavia/Jakarta data were published in yearbooks. Figure 5.2 shows an example page from such a yearbook. In the post-war period 1946-1958 almost no measurements were made and no year books have been published, with the exception of a combined yearbook for the 1948-1950 period (with incomplete measurements).

Table 5.1.
Inventory of the Indonesian climate data available in the KNMI archives.

period	data	amount
1866-1980	hourly observations of air pressure, temperature, moisture, etc. of Batavia/Jakarta	~ 115 years of hourly observations (on average 7 elements/obs)
1879-1916	daily precipitation sums of various stations	on average 300 precipitation stations
1917-1965	monthly statistics derived from daily values. Per month: precipitation sums, number of wet/dry days, extremes, etc)	On average 2000 stations
until 1889 notably 1848-1858	subdaily observations of various stations	Already digitized
1910-1973	monthly and yearly observations of various stations	On average 200 stations

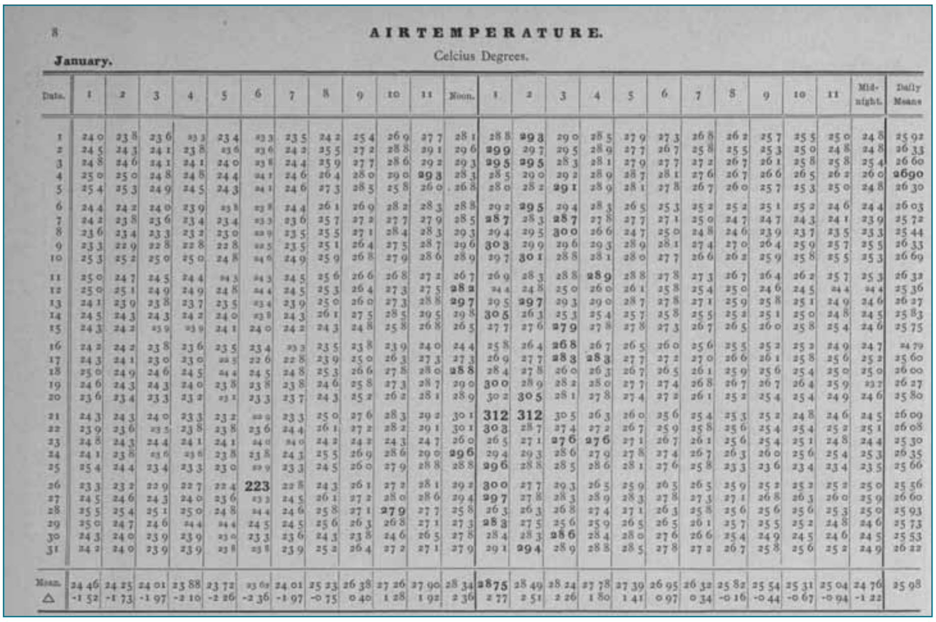


Figure 5.2.
Example page of hourly temperature data of January 1910 in the Batavia yearbook of 1910.

5.3 Digitization and quality control

5.3.1 Digitization

Before digitizing the data into spreadsheets, we transformed all hardcopy source material of Indonesia into digital images with the CopiBook bookscanner (see Chapter 4). Because of the enormous amount of data and the relatively good quality of the images, we decided to study the feasibility of using optical character recognition (OCR) software. We found that ABBYY FineReader (v.10.0) gives acceptable results. Finereader was used to transform the digital images to actual data in spreadsheets, which were later combined into one data file.

5.3.2 Quality control

Quality control consisted mainly of the correction for actual typing errors. Monthly and daily means/totals of the digitized data were compared with those in the source material. Almost all typing errors could be corrected using this procedure. Some errors were also detected by visual inspection of the time series.

The hourly series may contain inhomogeneities due to e.g. changes of instruments, changes in measurements practices or slow or abrupt changes in the surrounding. We did not correct for these for the same reasons as mentioned at the end of Section 3.3.4 about the high-resolution precipitation series.

5.4 Deliverables and examples of results

5.4.1 Deliverables

The main deliverables of this part of the project are:

1. A dataset with the hourly air temperature measurements of Batavia/Jakarta for the 1866-1980 period (about 102 station years, 894.120 measurements). This quality controlled data is freely available via the KNMI website.
2. A set of digital images of all hardcopy Indonesian data available in the KNMI archives. The images are available from the KNMI website and enable the user of the data to go back to the main hardcopy source of the data.
3. Publication: digitization and scanning of historical climate data (Brandsma, 2008).

5.4.2 Example of results

Figure 5.3 presents the annual mean temperature in Batavia/Jakarta as calculated from the hourly measurements. The figure shows a warming trend of about 1.5°C in the 1866-1980 period, which is much larger than the global warming trend in that period. This may partly be caused by the enormous growth of the city of Batavia/Jakarta in that period.

Figure 5.4 shows the mean diurnal temperature cycle of four adjacent periods between 1866 and 1980. The figure shows that the increase in the annual temperature is mainly visible in the daytime temperatures while the nighttime temperatures remained relatively constant.

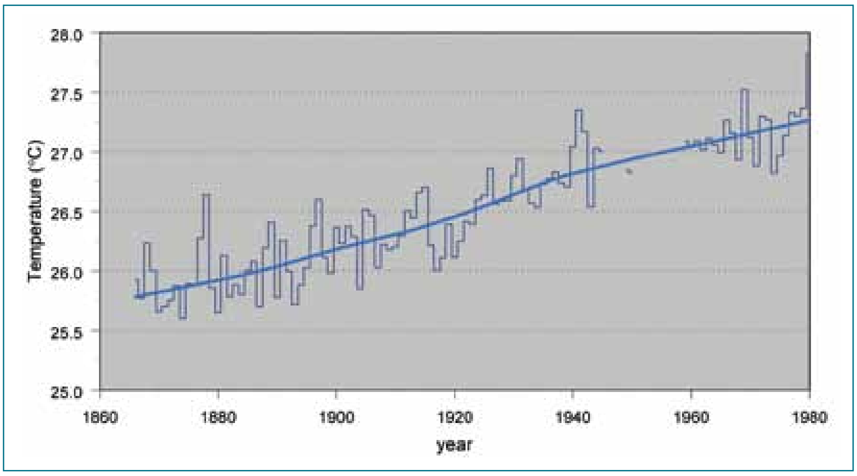


Figure 5.3.
Annual mean temperature of Batavia/Jakarta in the period 1866-1980. The thick smooth line is based on a loess smoother with automatic span selection.

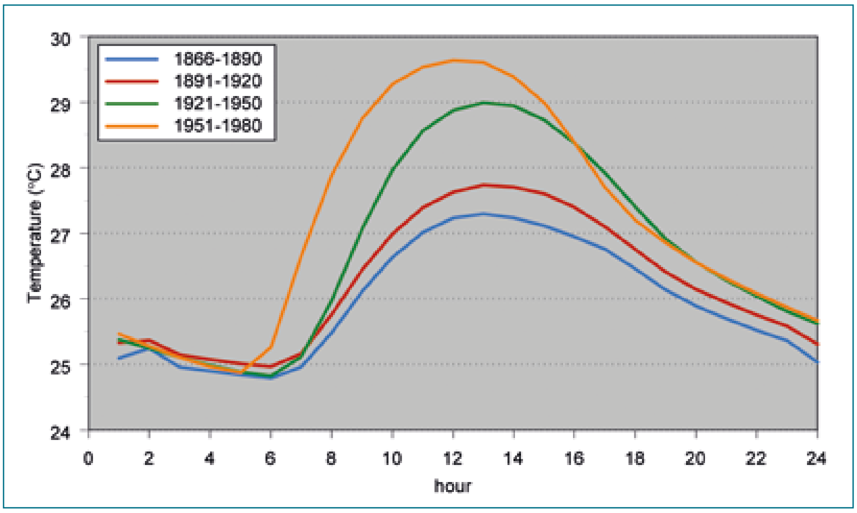


Figure 5.4.
Mean diurnal temperature cycle of four adjacent periods between 1866 and 1980.

5.5 Possible applications of the data

The main application of the hourly temperatures of Batavia/Jakarta are:

1. Validation of climate models for this data sparse area.
2. Construction of baseline climate and development of climate scenarios.
3. Study of long term temperature trends and variability
4. Study of the impact of the urban heat island on the temperature of Batavia/Jakarta.

The value of the dataset will increase when the other variables become available (see www.didah.org). The daily values of the digitized data for Indonesia and neighboring countries are also being collected in the Southeast Asian Climate Assessment & Dataset (SACA&D)[7], which enables the analysis of the extreme events in the region.

7 See at: http://saca-bmkg.knmi.nl

6. Discussion and conclusions

The CS8 project digitized and disclosed a large amount of historical climate data for research and the general pubic. Climate data rescue, digitization and disclosure receives an increasing interest in the World. Programs like Hisklim (Brandsma, 2000) have been emerging in other parts of the World too. A well know project is The international Atmospheric Circulation Reconstructions over the Earth (ACRE) initiative (www.met-acre.org) that undertakes and facilitates the recovery of historical instrumental surface terrestrial and marine global weather observations. Another example is the The MEditerranean climate DAta Rescue initiative (www.omm.urv.cat/MEDARE).

In the CS8 project we explored the use of various digitization techniques. In some cases manual keying the data into spreadsheets is still the best option. We used this for the daily rainfall data (which contains a lot of zero's) and the hand-written pre-1850 data. For the rainfall strip charts and paper rolls manual typing was not an option. We therefore developed an automatic curve extraction (ACE) algorithm for processing bulk quantities of rainfall strip charts and rolls. ACE can be applied to other variables as well. Today the archives of meteorological and hydrological institutes still contain huge amounts of strip charts that contain information not yet available in digital form. The method presented here has the potential to be used in those cases as well. Finally, we used OCR to digitize the yearbooks of Batavia/Jakarta. OCR may speed up the digitization when good quality digital images are available of the (printed) hardcopy measurements.

New digitization techniques are being developed. An example is a technique where that makes use of crowd digitization (see e.g. www.data-rescue-at-home.org). The idea is that the digital images with (mostly) handwritten data are made available via Internet together with a tool for keying in the data. A crowd of volunteers can then key in climate data at home and thus contribute to the creation of new datasets.

For most of the digitization activities it is important to first scan the hardcopy source material to high-quality digital images. This is not only important for data rescue and preservation, but it also enables the users of climate data to go back to the source material. The latter is increasingly being recommended for good quality climate data and may even become a requirement in the future. Accordingly, in CS8 we made digital images of almost all the source material and presented this freely on the Internet along with the data files of the digitized data.

Finally, digitization and disclosure of historical climate data has a somewhat old-fashioned and dusty image. This is undeserved! The old climate data represent the reference that the World needs to assess the present and future climate. Numerous scientific journal papers and books are published each year making use of these data and many infrastructural works can only reliably be designed because of the existence of long-term climate time series. Although the actual digitization work is often labour-intensive, new techniques are becoming available to scan, digitize and disclose the climate data via the Internet in a more efficient way than was possible previously. In summary, there is plenty of reason to invest time and money in this type of work.

7. References

Braak C., [1921]. Het klimaat van Nederlandsch Indië, Deel I: Algemene hoofdstukken, Javasche Boekhandel & Drukkerij (529pp, in Dutch).

Brandsma, T., F. Koek, H Wallbrink & G.P. Konnen, 2000. Het KNMI-programma Hisklim (HIStorisch KLIMaat) KNMI publication: PUBL-191 (in Dutch).

Brandsma, T., 2008. Data rescue and digitization: tips and tricks resulting from the Dutch experience International Workshop on Rescue and Digitization of Climate Records in the Mediterranean Basin, 28/11/2007-30/11/2007, Manola Brunet and Franz Kuglitsch (Ed), 2008, Geneva, Switzerland, World Meteorological Organization, WMO, WCDMP No. 67, WMO-TD No.1432, 47-54.

Buishand, T.A., & C.A. Velds, 1980. Klimaat van Nederland 1: Neerslag en verdamping. KNMI (26 pp., in Dutch).

Buishand T.A., G. De Martino, J.N. Spreeuw & T. Brandsma, 2011. Homogeneity of precipitation series in the Netherlands and their trends in the past century. Submitted to International Journal of Climatology.

Cleveland W.S., 1979. Robust locally weighted regression and smoothing scatterplots. Journal of the American Statistical Association 74: 829-836.

Conrad, V. & L.D. Pollak, 1962. Methods in Climatology. Harvard Univ. Press, Cambridge, Mass., 459 pp.

Geurts H.A.M. & A.F.V. van Engelen, 1992. Historische Weerkundige waarnemingen, vol. 5, Beschrijving Historische meetreeksen. KNMI publication 165-V (in Dutch).

Jilderda, R., 2011. Inschatting van ontbrekende waarden in de dagelijkse afappingen van de neerslag verricht vóór 1 januari 1951. KNMI Intern rapport-XX (in Dutch; in preparation).

Klein Tank A.M.G., J.B. Wijngaard, G.P. Können, et al., 2002. Daily dataset of 20th-century surface air temperature and precipitation series for the European Climate Assessment. International Journal of Climatology, 22, 1441-1453.

Können, G.P. & F.B. Koek, 2005. Description of the CLIWOC database. Climatic Change 73, 1-2, 117-130.

Können G.P. & T. Brandma, 2005. Instrumental pressure observations from the end of the 17th century: Leiden (The Netherlands). Int. J. Climatology, 2005, 25, 8, 1139-1145, doi:10.1002/joc.1192.

Labrijn, A., 1945. Het klimaat van Nederland gedurende de laatste twee en een halve eeuw, Med. en Verh. 49, KNMI, De Bilt, 114 pp. (in Dutch, with extended summary in English).

Lenderink G., E. van Meijgaard & F. Selten, 2009. Intense coastal precipitation in the Netherlands in response to high sea surface temperatures: analysis of the event of August 2006 from the perspective of a changing climate. Climate Dynamics (32), nr. 1, 19-33.

Menne, M.J. & C.N. Williams (Jr), 2009. Homogenization of temperature series via pairwise comparisons. Journal of Climate 22(7): 1700-1717. DOI: 10.1175/2008JCLI2263.1.

Moberg A. & P.D. Jones, 2005. Trends in indices for extremes in daily temperature and precipitation in central and western Europe, 1901-99. International Journal of Climatology (25), 1149-1171.

Muller S.H. & A. van Londen, 1983. Het beoordelen van regenmeters, met als voorbeelden de Thies-regenmeter en de elektrische KNMI-regenmeter, KNMI scientific report WR-83-16, KNMI, De Bilt, 48 pp (in Dutch).

Parmet B.W.A.H., T.A. Buishand, T. Brandsma & R. Mulders, 1999. Design discharge of the large rivers in The Netherlands - towards a new methodology.

Hydrological Extremes: Understanding, Predicting, Mitigating, Gottschalk et al. (Ed), 1999, Birmingham, International Union of Geodesy and Geophysics (IUGG), IAHS, ISBN 1901502856, 269-272.

Piggelen, H.E. van, T. Brandsma, H. Manders & J.F. Lichtenauer, 2011. Automatic curve extraction for digitizing rainfall strip charts. J. Atm. Oceanic Technol., 2011, 28, 7, 891-906, DOI: 10.1175/2011JTECHA1505.1.

Piggelen, H. van, J. Lichtenauer & T. Brandsma, 2010. Automatic Curve Extraction (ACE) Framework. KNMI-publication: PUBL-227, 27/12/2010, pp50.

Rapp J. & C. Schönwiese, 1995. Atlas der Niederschlags- und Temperaturtrends in Deutschland 1891-1990. Frankfurter geowissenschaftliche Arbeiten/B, Band 5. Johann Wolfgang Goethe-Universität, Frankfurt am Main.

Schmith T., 2001. Global warming signature in observed winter precipitation in Northwestern Europe. Climate Research (17), no. 3, 263-274.

Senguerdius W., 1699. Atmospheraerici Aeris. Lugduni Batavorum, Apud Cournerlium Boutesteyn: Leiden. (Available from library KNMI).

Van Boxel J. & E. Cammeraat, 1999. Wordt Nederland steeds natter? Een analyse van de neerslag in deze eeuw. Meteorologica (8), nr. 1, pag. 12-15 (in Dutch).

Van Engelen A.F.V. & H.A.M. Geurts, 1985. Historische Weerkundige waarnemingen, Vol 4, Nicolaus Cruquius (1678–1754) and his meteorological observations. KNMI publication 165-IV.

Wallbrink, H. & T. Brandsma, 2008. DIGISTAD: Disclosure of the hourly meteorological observations of the Amsterdam City Water Office 1784-1963. KNMI publication: PUBL-220, 4/4/2008, pp26.

Climate changes Spatial Planning

Climate change is one of the major environmental issues of this century. The Netherlands are expected to face climate change impacts on all land- and water related sectors. Therefore water management and spatial planning have to take climate change into account. The research programme 'Climate changes Spatial Planning', that ran from 2004 to 2011, aimed to create applied knowledge to support society to take the right decisions and measures to reduce the adverse impacts of climate change. It focused on enhancing joint learning between scientists and practitioners in the fields of spatial planning, nature, agriculture, and water- and flood risk management. Under the programme five themes were developed: climate scenarios; mitigation; adaptation; integration and communication. Of all scientific research projects synthesis reports were produced. This report is part of the Climate scenarios series.

Climate scenarios

The projects in this field are designed to obtain high quality climate information and scenarios relevant for the Netherlands. The projects both focus on an improved monitoring and modelling of regional climate variability, and at the construction of tailored climate change scenarios suitable for exploring spatial adaptation options, such as flood retention areas or coastal defense. In all fields special attention is devoted to extreme climate conditions. The climate scenarios are designed and developed jointly with a number of key stakeholders.

Programme Office Climate changes Spatial Planning

c/o VU University Amsterdam, FALW
De Boelelaan 1085
1081 HV Amsterdam
The Netherlands
T +31 20 598 7318
office@klimaatvoorruimte.nl

c/o Alterra, Wageningen UR
P.O. Box 47
6700 AA Wageningen
The Netherlands
T +31 317 48 6540
info@klimaatvoorruimte.nl

www.climatechangesspatialplanning.nl