'European Climate Assessment & Dataset (ECA&D)'

Algorithm Theoretical Basis Document (ATBD)

Author: Project team ECA&D,
: Royal Netherlands Meteorological Institute KNMI
Date : December 9, 2021
Version: 11

Contents

1 Project description

1.1 Objectives

The European Climate Assessment & Dataset project (ECA&D) started in 2003 as the follow-up to ECA (for which KNMI was responsible member since 1998). Between 2003 and 2008 the project has been partially funded by EUMETNET. From 2009 onwards, KNMI has committed itself to fund ECA&D. ECA&D has now obtained the status of Regional Climate Centre (RCC) for high resolution observation data in WMO Region VI (Europe and the Middle East).

The objective of ECA&D is to analyze the temperature and precipitation climate of WMO region VI, with special focus on trends in climatic extremes observed at meteorological stations. For this purpose, a dataset of 20th-century daily surface air temperature and precipitation series has been compiled (Klein Tank et al. 2002a) and tested for homogeneity (Wijngaard et al. 2003).

To enable periodic assessments of climate change on a European scale, a sustainable system for data gathering, archiving, quality control, analysis and dissemination has been realized. Data gathering refers to long-term daily resolution climatic time series from meteorological stations throughout Europe and the Mediterranean provided by contributing parties (mostly National Meteorological Services (NMSs)) from over 40 countries. Most series cover at least the period 1946–now. Archiving refers to transformation of the series to standardized formats and storage in a centralized relational database system. Quality control uses fixed procedures to check the data and attach quality and homogeneity flags. Analysis refers to the calculation of (extremes) indices according to internationally agreed procedures specified by the CCL/CLIVAR/JCOMM Expert Team on Climate Change Detection and Indices (ETCCDI, `http://www.clivar.org/organization/etccdi/etccdi.php`). Finally, dissemination refers to making available both the daily data (including quality flags) and the indices results to users through a dedicated website.

Recently, the necessary steps have been completed for an improved operational ECA&D system as the first implementation of a Regional Climate Centre (RCC) functionality for high resolution observational data and extremes indices in WMO Region VI. This implies that the system has been made more sustainable/transparent and has been embedded into KNMIs information infrastructure. This ensures ongoing support, guarantees well-performing up-and-running services and documentation, backup- and maintenance procedures.

1.2 Users

Because of its daily resolution, the ECA dataset enables a variety of climate studies, including detailed analyses of changes in the occurrence of extremes in relation to changes in the mean. Web statistics, personal contacts and references in numerous publications, advice reports and applications show that ECA&D serves many users. Also the ECA&D report "Climate of Europe, assessment of observed daily temperature and precipitation extremes" (Klein Tank et al. 2002b) and its successor "Towards an operational system for assessing observed changes in climate extremes" (van Engelen et al. 2008) have received much praise. The project is widely recognized as an example of KNMIs leading European role in the area of climate data exchange and research.

1.3 Requirements

1. Not all countries will be able to submit their contribution in a standardized format at regular time intervals. Therefore, the continuation of individual treatment of each participant is crucial for success. This implies that dedicated solutions should be developed for each data provider, with the level of automation dependent on the technical and manpower possibilities of the respective participants.

2. The data come with different use permissions. We are allowed to redistribute some series to the general public, whereas others are only for index calculation and use in the calculation of the gridded data products. The system should allow for different permission flags.

3. Since there is always a time lag between the most recent data contributed by participants and the present date, the observations from SYNOP messages for the same or nearby stations that are transmitted through the Global Telecommunication System (GTS) should temporarily be used to fill the gap. Once the 'official' series are available from the data providers in participating countries, the temporary SYNOP data should be replaced. Regular updates, using SYNOP data and readily available participant data (see § 2.1) are on a monthly basis. Requesting updates from all data participants is done on a less frequent basis. Each update of the daily data will be followed by a recalculation of quality control scores, indices, climatology, trends and homogeneity. This is followed by a calculations of provisional gridded datafiles for precipitation and daily maximum, minimum and averaged temperature for the past month.

4. The minimum set of metadata for each series, which is required to judge the quality and representativeness of the observations, is described in Aguilar et al. (2003). Metadata information is important

since not all station observations conform closely to the recommendations of instrumentation, exposure and siting which are given in the WMO-CIMO Guide. Moreover, the recommendations have changed over time. The minimum set of metadata should be stored along with the data series. Some of these metadata are used in the blending process.

5. The system should adopt and comply with (inter)nationally agreed standards as much as possible. This refers both to data format and database standards as well as metadata description standards.

6. A subset of the stations with ECA&D series is part of the GCOS Surface Network (GSN). For some of these stations, the daily series are collated and archived also at the WMO World Data Center A in Asheville (U.S.A.). Discrepancies between the series in ECA&D and those in GSN should be carefully monitored. Data series in GSN that are not part of ECA&D will be copied. (Camuffo & Jones (Eds.) 2002).

7. The ECA&D website, as a dissemination tool for data and indices results, should be easily accessible and flexible for many users. Researchers and operational climatologists have very different requirements. The possibility of different interfaces should be explored ranging from bulk download to customizable queries through the data and indices results. Also the output formats on screen and print should be flexible providing reports in different layouts. The daily data should be available to users in different stages of processing. This means that the 'raw' data files (as received from the participants, including explanatory e-mails) as well as the reformatted and quality-controlled data should be stored.

8. The European Environment Agency (EEA) relies on the extremes indices for its European state of the environment reports, which are issued at regular intervals and aim to support sustainable development (EEA-JRC-WHO 2008). Contacts with responsible authors at EEA have learned that they would prefer using up-to-date information also for their annual assessments in particular with respect to index anomaly maps for individual years.

9. The existence of copies of (subsets of the) ECA dataset elsewhere on the Internet in reformatted files should be discouraged. Already, STARDEX (`http://www.cru.uea.ac.uk/projects/stardex/`), GDCN (`http://lwf.ncdc.noaa.gov/oa/climate/research/gdcn/gdcn.html`) and the Climate Explorer (`http://climexp.knmi.nl/`) extracted and published copies of the entire dataset. The problem is that these adhoc copies often stay without regular updates. To improve this situation, specific agreements with responsible persons should be reached

so that the required subsets are delivered straight from the ECA&D source or provided at the ECA&D website.

10. In several WMO working groups, KNMI has indicated its willingness to offer help to other continents, in particular Africa and South America to run similar projects as ECA&D. Part of this capacity building will consist of infrastructure (web) issues. KNMI is also involved in an Indonesia project where ECA&D will provide the infrastructure for analysis and distribution of data. In addition, there is the intention to use the ECA&D system of presenting index results for worldwide indices collected by the ETCCDI. To be prepared for these future requests, the developed system should keep into account such extensions.

11. The developed web interface should run easily on workstations and Internet PCs typically used in participating countries. This means that also lower capacity PCs (e.g. using MS Windows 95 on 386 processor PCs with 8 Mb ram and 800x600 screen resolution with 256 colors and 56k modem) should be able to use the interface without difficulties. All popular web browsers should be supported (MS Internet Explorer, Netscape Navigator, Mozilla, Opera, Lynx). Performance of the system should meet minimum standards. For all parts of the user interface maximum waiting time (assuming optimum Internet speed and advanced PCs or workstations) should at maximum be in the order of 3 to 5 seconds.

12. Operational guarantees for the system outside the KNMI firewall, which has the website and database running, are on a 8/5 basis. Bringing the system up and running again at the next working day is satisfactory, provided that the archived data are in no danger. User access monitoring facilities should be used to count the number of hits and to determine user preferences. This information is to be used primarily for further improvements of the system.

13. The technical solutions should benefit from the routine backup- and maintenance procedures KNMI employs. Optimal use should be made of KNMI information systems and infrastructure to ensure ongoing support and to guarantee up-and-running services from the ECA database and website and to ensure restoring data, with no loss. Regular and reliable backup procedures should be maintained. On the other hand, changes in the KNMI infrastructure should not negatively affect the results of the ECA&D project.

1.4 Infrastructure and software

At the moment, two dedicated ECA&D systems are in use: the develop & test environment and the operational system (outside the firewall). All procedures are run on a developer platform and the results are copied to the operational platform. The operating system is Linux. The web-server is Apache. A MySQL database is used to store the data and corresponding metadata. Most of the software used to update the database are written in Bash, Fortran and C code. More details about the infrastructure can be found in the internal document about the ECA&D infrastructure.

1.5 Data flow

The necessary steps in data processing are:

1. New data import
2. Quality control
3. Blending
4. Indices calculation
5. Climatology calculation
6. Trend calculation
7. Homogeneity analysis
8. E-OBS gridded dataset
9. Website

For each step, the main method is described in the sections below.

2 New data import

2.1 Design rules

2.1.1 NMHSs and data holding institutes

Participant data comes in various file formats. Importing this data into the database tables is entirely done by hand, running relevant scripts to do the conversions. The conversions differ for each data source. Dependent on the permissions granted by the data providers, data series can either be: public or non-public. Non-public data are only used in the calculation of the trends, indices and the gridded datasets, while the public data are published on the web as well. Most station series are updated irregularly, each time after the data providers are contacted.

A few stations are updated on a monthly basis. The stations in Norway, Sweden, Slovenia, Germany, Montenegro, Spain, Wageningen-Haarweg, Portugal (Sogrape S.A. company), Italy-Emilia Romagna and Flanders are updated via the data they provide on their websites. For Dutch stations, a link is made with the Dutch meteorological database. Luxembourg Airport and stations from Bosnia-Herzegovina are updated via monthly emails send by the data provider of those stations. Data for stations in Estonia, the Czech Republic, Switzerland, Ireland, Spain and Finland are provided every month by FTP.

Halfyearly updates are received from the Catalan Meteorological Service, and yearly updates are received from the UK Met Office, MétéoFrance and the Danish meteorological service.

Some NMHSs provide data aggregated over an interval which differs strongly from the 0-0 interval (in either UTC or local time). Examples are the use of the 9-9 UTC interval for temperature of 7-7 UTC for precipitation. Measurements are then often related to the date of the day at which the interval ends. The chances that the daily maximum temperature actually occurred on the date labelled to the measurement are small; such maxima often occur in the afternoon rather than the morning. For precipitation, a similar problem exists, where the chance are high that the largest amount of precipitation actually occurred on the day preceding the the date labelled to the measurement. When the metadata provided by the participant indicates such situation, ECA&D staff adjusts the date of the measurements of daily maximum temperature and daily sums of precipitation by a shift of one day backwards. This shift is documented in the element description.

GCOS Surface Network (GSN)

The GCOS Surface Network (GSN) is established as an outcome of the Second World Climate Conference in 1992 to ensure that the observations and information needed to address climate-related issues are obtained and made available to all potential users.

Here we use the GSN to update series in countries that do not regularly (or not at all) contribute to ECA&D. The list with countries for which the GSN stations are updated is:

AG: Algeria	LO : Slovakia
AJ : Azerbaijan	MD: Moldova
BO: Belarus	MT: Malta
CY: Cyprus	PL : Poland
EG: Egypt	RO : Romania
GR: Greece	RS : Russia
HU: Hungary	SY : Syria
IC : Iceland	TS : Tunisia
IT : Italy	TU : Turkey
LG: Latvia	TX : Turkmenistan
UP: Ukraine	:

World Radiation Data Centre (WRDC)

The World Radiation Data Centre (WRDC) is located in St. Petersburg at the Main Geophysical Observatory of the Russian Federal Service for Hydrometeorology and Environmental Monitoring. The WRDC was established in 1964, and since that time it centrally collects, archives and publishes radiometric data for the world, to ensure the availability of these data for research purposes by the international scientific community. The WRDC is one of recognised World Data Centres sponsored by the World Meteorological Organization (WMO) on the basis of Resolution 12 XIV Session of the EC and Resolution 6, XXXVI Session of the EC. Gobal radition data (daily averages) from WRDC are accessed by ECA&D for the European domain to provide data in countries that do not supply global radiation dirctly to ECA&D.

Global Summary Of the Day (GSOD)
The station density of stations providing daily values of windstrength to ECA&D is very inhomogeneous. In order to increase the coverage over Europe for the NMHSs that do not supply windspeed data directly, the NOAA/NCEI repository Global Summary Of the Day (GSOD) is used. Historical data are generally available for 1929 to the present, with data from 1973 to the present being the most complete. In deriving the summary of day data, a minimum of 4 observations for the day must be present (allows for stations which report 4 synoptic observations/day). The data are converted to constant units (e.g, knots) which are not part of the International System of Units (SI-units) which gives slight rounding error from the originally reported values. As the data in ECA&D adhere to the SI-units, new rounding errors may be introduced. The data are reported and summarized based on Greenwich Mean Time (GMT, 0000Z - 2359Z) since the original synoptic/hourly data are reported and based on GMT.

Baseline Surface Radiation Network (BSRN)
The Baseline Surface Radiation Network (BSRN) provides radiation measurements together with collocated surface and upper-air meteorological observations and station metadata in an integrated database. BSRN is a project of the Data and Assessments Panel from the Global Energy and Water Cycle Experiment (GEWEX) under the umbrella of the World Climate Research Programme (WCRP) and as such is aimed at detecting important changes in the Earth's radiation field at the Earth's surface which may be related to climate changes.

Although the BSRN has a modest number of stations in Europe, the quality and maintenance of these stations is exceptional, which is the motivation to include this network in ECA&D.

Global radiation from the BSRN is calcuated using a sum of diffuse and direct radiation, where the latter is weighted with the cosine of the solar senith angle. The solar zenith angle can be calculated for any given time and location and the approximate algorithm of Michalsky (1988) is used here. Note that this algorithm (appended to the end of the article) is based on *single-precision* calculations while the implemented algorithm in ECA&D uses *double-precision*. The advise from the author or the algorithm is that it probably makes sense to use double precision, however, the calculations are only good to 0.01 degrees. In addition, there is further uncertainty caused by refraction, which is unknown because atmospheric profiles are assumed to be those of the US Standard Atmosphere. This, of course, is especially a problem at low sun (Michalsky, personal communication, 04/10/2018). Based on these considerations, the effects of refraction on the daily-average global radiation are ignored.

2.1.2 Synoptical data through the GTS

The data provided by the participants is always received with some delay. It is not possible for all of the participants to deliver (near) real time data, because of validation and verification. To update each series at the time that participant data has not yet arrived, SYNOP messages are used. The source for these synoptical data is the ECMWF MARS-archive (see `http://www.ecmwf.int/services/archive/`). This archive is a complete and consistent representation of SYNOP messages distributed over the GTS. Synoptical data is retrieved from the MARS-archive only for WMO-Region VI and countries in North Africa. For technical reasons, this is translated to be all land stations that fit in the rectangle 90N/40W and 10N/80E. Data retrieval is restricted to the reports of the main hours 00, 06, 12 and 18 UT.

Daily minimum and maximum temperatures are reported at 6 UT and 18 UT respectively, and refer to the 12-hour periods preceding these points in time. These values therefore do not cover a full 24-hour period. The impact of this mismatch and issues related to the quality of the synoptical data is currently under investigation.

Daily values for the following 13 elements are derived from the SYNOP messages:

Daily maximum temperature TX In the synoptical report of 18 UT, the daily maximum temperature is given for that day. This daily maximum temperature is the highest temperature recorded between 06 UT and 18 UT (according to WMO specifications).

Daily minimum temperature TN In the synoptical report of 06 UT, the daily minimum temperature is given for that day. This daily minimum temperature is the lowest temperature recorded between 18 UT (previous day) and 06 UT (according to WMO specifications).

Daily mean temperature TG If the daily maximum temperature (TX) and the daily minimum temperature (TN) is known, mean daily temperature is calculated as TG=(TX+TN)/2. This combines the TX reading of 18 UT and that of TN of 6 UT of the current day.

Daily mean sea level pressure PP Whenever sea level pressure data is available at 00, 06, 12 and/or 18 UT, daily mean sea level pressure is calculated as the average of the available values.

Daily precipitation amount RR Whenever synoptical 12-hourly precipitation data is available at 06 and 18 UT, daily precipitation is calculated as the sum of RR of 18UT of the current day and RR of 6 UT of the next day.

Daily mean snow depth SD Whenever synoptical snow depth data is available at 00, 06, 12 and/or 18 UT, daily mean snow depth is calculated as the average snow depth of the available values.

Daily mean cloud cover CC Whenever synoptical cloud cover data is available at 00, 06, 12 and/or 18 UT, mean daily cloud cover is calculated as the average of the available values. This value in percent is converted to octas by ROUND((cloud cover_in_percents/100)*8).

Sunshine duration SS Whenever synoptical sunshine duration is available (in minutes) at 00, 06, 12 and/or 18 UT, daily sunshine duration is calculated as the summation of the available values.

Daily mean humidity HU Whenever synoptical humidity data is available (in percents) at 00, 06, 12 and 18 UT, daily mean humidity is calculated as the mean of the available values.

Daily mean wind speed FG Whenever synoptical wind speed data is available (in m/s) at 00, 06, 12 and 18 UT, daily mean wind speed is calculated as the mean of the available values.

Daily maximum wind gust FX Whenever synoptical wind gust data is available (in m/s) at 00, 06, 12 and 18 UT, daily maximum wind gust is taken as the maximum of the available values.

Wind direction DD Whenever synoptical wind direction data is available (in degrees) at 12 UT, that value is taken as the wind direction.

Global radiation QQ Whenever Global Radiation sums or averages are available at 00, 06, 12 and 18 UT, daily mean Global radiation (in W/m^2) is calculated as the mean of the four available values.

2.2 Current implementation

Within the ECA&D relational database, various types of tables are distinguished: core tables that hold the unique raw data, working tables that hold temporarily stored data and so-called derived tables that hold derived data calculated according to the rules specified in the remainder of this document. Derived data is updated by running the various processes. It is necessary to store these derived data for better performance of subsequent procedures and/or the website. Data for different elements xx are stored in separate tables. Based on the use permissions that participants have given to their data, two different targets are distinguished. Likewise, tables have extensions for the targets: public and mixed. *Mixed* indicates public data combined with non-public data. The data in the *mixed* tables are used for indices, trends and gridding, while only the data in *public* are available for download on the website. Data in the *public* tables are a subset of those in the *mixed* tables.

The SYNOP messages from the ECMWF MARS-archive are downloaded on a monthly basis. The data archive comes in a BUFR-format, a WMO defined format for irregular spaced point data. To process this BUFR-formatted archive, the ECMWF BUFRDC subroutines are used. These subroutines expand the BUFR-file into ASCII-readable data, which is processed further. The subroutines extract only the data required, i.e. TX, TN, PP, RR, SD, CC, HU and SS, corresponding respectively with BUFR-fields:
12014 (maximum temperature at 2 m, past 12 hours),
12015 (minimum temperature at 2 m, past 12 hours),
10051 (pressure reduced to mean sea level),
13022 (total precipitation past 12 hours),
13013 (total snow depth),
20010 (cloud cover (total)),
13003 (relative humidity),
14031 (total sunshine),
11011 (wind direction),
11012 (wind speed),
11041 (wind gust).
14028 (global radiation).

After extraction into a ASCII-formatted file, every TX, TN, PP, RR, SD, CC, HU, SS, FG, FX, DD and QQ (and the calculated TG) of a synoptical station is stored in a temporary table. When the complete ASCII-file is processed, another process reads this temporary table and determines the daily values. Details about the programs that do this, can be found in an accompanying internal document.

2.3 Conversion from sunshine duration to global radiation

In order to provide a historical perspective for global radiation that is as long as possible, sunshine duration records (which generally extend further back in time) are used to provide a proxy for the corresponding global radiation records. The conversion is documented in detail by van der Schrier et al. (2021). Here a summary is given.

The relationship between daily values of sunshine duration and global radiation is non-linear. A detailed analysis is given by Suehrcke et al. (2013) who derived a relationship which is generalization of the Angstrom-Prescott formula (Allen et al. 1994a)[1]. Albeit the study of Suehrcke et al. (2013) is for monthly values, we use this as the basis for daily amounts. A considerable amount of documentation on the linear relation of the original Angstrom-Prescott formula is available in Duffie & Beckman (1991).

The reason for the non-linearity between sunshine duration and global radiation is explained by Suehrcke et al. (2013). When the cloud cover of the sky increases, clouds also tend to become less transparent for solar radiation; they become optically thicker. This means a reduction in sunshine fraction not only reduces the received clear sky radiation, but simultaneously also reduces the solar radiation transmittance through the clouds. This additional change in cloud transmittance with sunshine fraction causes the non-linearity in the sunshine-radiation relationship. It was also found that the relation changes in space and and with the seasons.

The model proposed by Suehrcke et al. (2013) is:

$$K = \bar{K}_{\text{clear}} \left[\beta + (1 - \beta)\, s^{\gamma} \right] \tag{1}$$

where K is the fraction of radiation at the earth surface (fraction of maximum value), $\bar{K}_{\text{clear}}$ is the daily clearness index for a cloudless day ($\bar{K}_{\text{clear}} \approx 0.7$), β and γ are parameters of the model and S is the average time fraction of bright sunshine ($0 \leq S \leq 1$).

The parameters of the free-parameter model (1) will have a seasonality (Suehrcke et al. 2013), so estimates of the parameters, optimal in a least-squares sense, are made separately in the four seasons. Initially, data from the last five seasons that are present in the records are used to make these estimates. Note that the evaluation of (1) involves calculating the clearness index (the actual amount of global radiation that reaches the earth surface in relation to clear sky global radiation) and the daily sunshine fraction (actual amount of sunshine in relation to the length of day). These parameters are calculated using the yearday and the latitude as input, and the formula used are documented by Allen et al. (1994a).

The estimate of the parameters is made using a method where the sum of squared residuals between the observed clearness index and the calculated

[1] Available online at: http://www.fao.org/3/X0490E/x0490e07.htm#solar radiation

clearness index using eq. 1 is minimized for all observations. The minimization uses Newton's method to search for a minimum in the sum of squared residuals as a function of the three parameters. This method requires a valid starting point, and for all calculations the Angström-Prescott linear relation was used as the starting point (relating to: $K_{textclear} = 0.76$, $\beta = 0.29$ and $\gamma = 1$ when parameters of the Angström-Prescott relation are chosen as $a = 0.22$ and $b = 54$).

The application of model (1) to all stations where we have an overlap in daily sunshine and global radiation records gives a set of over 220 estimates of the paramaters. For each of the four seasons, these estimates are interpolated over the European domain using a simple spline.

In the conversion of a sunshine duration record to a global record, the interpolated dataset is sampled for the corresponding grid box and the corresponding season, and a proxy for global radiation is made using eq. 1. These converted sunshine duration data ar stored in the ECA&D database with element description 'Global radiation amount 0 - 0 UTC derived from sunshine duration' .

3 Quality control

3.1 Design rules

Quality control (QC) procedures flag each individual observation in a series. Separate QC procedures are performed for the station series (non-blended) and the blended series. Three QC flags are currently implemented:

- Flag=0: 'valid'
- Flag=1: 'suspect'
- Flag=9: 'missing'

The following conditions apply for each element.

daily precipitation amount RR:
... must be equal or exceed 0 mm
... must be less than 300.0 mm
... must not be repetitive (i.e. exactly the same amount) for 10 days in a row if amount larger than 1.0 mm
... must not be repetitive (i.e. exactly the same amount) for 5 days in a row if amount larger than 5.0 mm
... dry periods receive flag = 1 (suspect), if the amount of dry days lies outside a 14·bivariate standard deviation

daily mean surface air pressure PP:
... must exceed 900.0 hPa

. . . must be less than 1080.0 hPa
. . . must not be repetitive (i.e. exactly the same) for 5 days in a row

daily maximum temperature TX:
. . . must exceed -90.0 °C
. . . must be less than 60.0 °C
. . . must exceed or equal daily minimum temperature (if exists)
. . . must exceed or equal daily mean temperature (if exists)
. . . must not be repetitive (i.e. exactly the same) for 5 days in a row
. . . must be less than the long term average daily maximum temperature for that calendar day + 5 times standard deviation (calculated for a 5 day window centered on each calendar day over the whole period)
. . . must exceed the long term average daily maximum temperature for that calendar day - 5 times standard deviation (calculated for a 5 day window centered on each calendar day over the whole period)

Daily minimum temperature TN:
. . . must exceed -90.0 °C
. . . must be less than 60.0 °C
. . . must be less or equal to daily maximum temperature (if exists)
. . . must be less or equal to daily mean temperature (if exists)
. . . must not be repetitive (i.e. exactly the same) for 5 days in a row
. . . must be less than the long term average daily minimum temperature for that calendar day + 5 times standard deviation (calculated for a 5 day window centered on each calendar day over the whole period)
. . . must exceed the long term average daily minimum temperature for that calendar day - 5 times standard deviation (calculated for a 5 day window centered on each calendar day over the whole period)

Daily mean temperature TG:
. . . must exceed -90.0 °C
. . . must be less than 60.0 °C
. . . must exceed or equal daily minimum temperature (if exists)
. . . must be less or equal to daily maximum temperature (if exists)
. . . must not be repetitive (i.e. exactly the same) for 5 days in a row
. . . must be less than the long term average daily mean temperature for that calendar day + 5 times standard deviation (calculated for a 5 day window centered on each calendar day over the whole period)
. . . must exceed the long term average daily mean temperature for that calendar day - 5 times standard deviation (calculated for a 5 day window centered on each calendar day over the whole period)

Daily snow depth SD:
. . . must exceed or equal 0.0 cm
. . . must be less than 300.0 cm if station elevation is less or equal to 400 m
. . . must be less than 800.0 cm if station elevation is between 400 m and 2000

m
. . . must be less than 1500.0 cm if station elevation is equal to or more than 2000 m

Daily cloud cover CC:
. . . must exceed or equal 0
. . . must be less than or equal 8

Daily humidity HU:
. . . must exceed or equal 5.0%
. . . must be less than or equal to 100.0%

Daily sunshine duration SS:
. . . must exceed or equal 0.0 h
. . . must be less than the maximum daylength plus a 5% margin, valid for the corresponsing yearday and latitude of the station (calculation rules below).

Daily global radiation QQ:
. . . must exceed or be equal to 3% of the Top Of Atmosphere (TOA) radiation
. . . must be less than the maximum amount of global radiation received at the earth surface on a clear day plus a 5% margin, valid for the corresponsing yearday, latitude and elevation of the station (calculation rules below).

Daily mean wind speed FG:
. . . must exceed or equal 0.0 m/s
. . . must be less than or equal to 46 m/s
. . . must not be repetitive (i.e. exactly the same value) for 6 days in a row if value larger or equal to 2.0 m/s

Daily maximum wind gust FX:
. . . must exceed or equal 0.0 m/s
. . . must be less than or equal to 76 m/s
. . . must not be repetitive (i.e. exactly the same value) for 5 days in a row if value larger or equal to 4.0 m/s

Daily mean wind direction DD:
. . . must exceed or equal 0.0 degrees
. . . must be less than or equal to 360 degrees
. . . must not be repetitive (i.e. exactly the same value) for 6 days in a row if value larger than 0.5 degrees

The default QC flag is 0 ('valid'). If one of the conditions above is not met: a QC flag of 1 ('suspect') is assigned. If data is missing: QC=9 ('missing'). The conditions are tested in an automated procedure, but a manual intervention is possible for non-blended series and the manual QC flag will be propagated to the blended series. For instance, precipitation extremes flagged 'suspect' are overruled if supplementary evidence exists (e.g. from radar images or weather charts) that the particular extreme is plausible.

If for a calendar day 10 or more samples exist, then the long-term average or standard deviation is calculated for that day. In order to adjust the day-to-day variability associated with the sampling, the long-term averages are smoothed. The (smoothed) long-term average is only calculated if the total number of days present is 25 or more. If a calendar day does not meet these requirements (e.g. for a leap day), the quality checks associated with long term averages are not performed for that day.

In the comparative quality control tests (i.e. daily maximum temperature $\geq$ daily averaged temperature $\geq$ daily minimum temperature) a caveat exists. As described in Section 2.1, some daily maximum temperature series may have been date-shifted by ECA&D staff with one day backwards. When this is the case, the comparative test is based on a comparison of the daily minimum and daily averaged temperatures of today, combined with the daily maximum temperature of yesterday.

The calculation rules for the TOA, the maximum daylength and the amount of global rdiation received at the surface of the earth on a clear day are sourced from Allen et al. (1994a) and reproduced below.

The extraterrestrial radiation, R_a, for each day of the year and for different latitudes can be estimated from the solar constant, the solar declination and the time of the year by:

$$R_a = \frac{24}{\pi} G_{sc} d_r \left[\omega_s \sin(\phi)\sin(\delta) + \cos(\phi)\cos(\delta)\sin(\omega_s)\right] \tag{2}$$

where

R_a	extraterrestrial radiation	[MJ m^{-2} day^{-1}]
G_{sc}	solar constant	108.08 MJ m^{-2} day^{-1}
d_r	inverse relative distance Earth-Sun	
w_s	sunset hour angle	[rad]
j	latitude	[rad]
d	solar decimation	[rad]

R_a is expressed in equation 2 in MJ m^{-2} day^{-1}. The latitude, j, expressed in radians is positive for the northern hemisphere and negative for the southern hemisphere.

The inverse relative distance Earth-Sun, d_r, and the solar declination, δ, are given by:

$$d_r = 1 + 0.033 \cos\left(\frac{2\pi}{365} j\right) \tag{3}$$

$$\delta = 0.409 \sin\left(\frac{2\pi}{365} j - 1.39\right) \tag{4}$$

where j is the number of the year between 1 (Jan 1) and 365 or 366 (Dec 31).

The sunset hour angle, w_s, is given by:

$$w_s = \arccos\left[-\tan(j)\tan(\delta)\right] \tag{5}$$

For the winter months in latitudes greater than 55° (N or S), the equations for Ra have limited validity.

The maximum possible duration of sunshine, N, is given by Equation 6:

$$N = \frac{24}{\pi}\omega_s. \tag{6}$$

For the quality control of global radiation, the maximum amount of radiation that is received at the earth surface is uses. This relation uses a linearization of Beer's radiation law and documented by Allen et al. (1994a):

$$R_a = (0.75 + 0.00002z))\, R_0 \tag{7}$$

Here R_a is the global radiation received on the surface of the earth on a cloudless day and R_0 is the top of the atmosphere radiation. The altitude of the station is denoted by z. One of the assumptions underlying this equation is that the average solar angle is about 50°.

4 Blending

4.1 Design rules

The procedure to calculate the optimal combination of ECA station and nearby station (which can be an ECA station or a synoptical station) has the following steps (applying spherical trigonometry):

1. Convert LAT and LON into decimal degrees. E.g. for station De Bilt this yields

Latitude:	52:06N	LAT_{ECA}	= 52+6/60 = 52.10
Longitude:	05:11E	LON_{ECA}	= 5+11/60 = 5.18

2. For every other station, also convert LAT and LON into decimal degrees

Latitude:	HH_{LA}:MM_{LA}	LAT_{OTHER}	= $HH_{LA}+MM_{LA}/60$
Longitude:	HH_{LO}:MM_{LO}	LON_{OTHER}	= $HH_{LO}+MM_{LO}/60$

If Latitude on southern hemisphere: $LAT_{OTHER} = -1 \cdot LAT_{OTHER}$
If Longitude on western hemisphere: $LON_{OTHER} = -1 \cdot LON_{OTHER}$

3. Find a combination ECA-OTHER station by minimizing the distance (here in km):

   distance = radius_earth $\times$ ARCCOS(SIN(atan$\cdot$LAT$_{\mathrm{ECA}}$) $\times$ SIN(atan $\cdot$ LAT$_{\mathrm{OTHER}}$) + COS(atan $\cdot$ LAT$_{\mathrm{ECA}}$) $\times$ COS(atan $\cdot$ LAT$_{\mathrm{OTHER}}$) $\times$ COS(atan $\cdot$ (LON$_{\mathrm{OTHER}}$ - LON$_{\mathrm{ECA}}$)))
   where: radius_earth = 6378.137 kilometers, and atan = ARCTAN(1)/45

   Substituting for De Bilt, with LAT/LON from WMO synoptical or ECA-stations yields:

   distance = radius_earth $\times$ ARCCOS(SIN(atan $\cdot$ 52.10) $\times$ SIN(atan $\cdot$ LAT$_{\mathrm{OTHER}}$) + COS(atan $\cdot$ 52.10) $\times$ COS(atan $\cdot$ LAT$_{\mathrm{OTHER}}$) $\times$ COS(atan $\cdot$ (LON$_{\mathrm{OTHER}}$ - 5.18)))

   Repeat distance for every OTHER station, keeping LAT$_{\mathrm{ECA}}$ and LON$_{\mathrm{ECA}}$ fixed (in the example above, for De Bilt). The OTHER station with lowest distance is the station that is nearest to De Bilt (in this example). Only data from stations that are no more than 12.5 km away from the original ECA station, is used.

4. As a last step, the difference in elevation of the ECA station and OTHER station is considered. Only data from stations located within 25 m height difference is taken into account.

Next, the blended series are constructed. Suppose we have a station series from 1900 until 2005, with missing data between 1930 and 1935 and also after 2005. Now that we know what other stations are nearby we are considering the data from these stations to 'infill' the gaps or data values that are flagged as suspect during QC (as illustrated in the figure below; see also § 3).

The logic that is applied when constructing the blended series is as follows. First, valid data from nearby ECA stations is taken to 'infill' the gaps, i.e. days with qc=1 or missing data. If no valid data from nearby ECA stations is available, valid data from nearby synoptical stations is taken to 'infill' the gaps. If there is less than 10 years difference between the year of the last date of the series and the current date, the series are extended with synop data from nearby synoptical stations as well. More details about the blending process can be found in the accompanying internal document.

The extension of validated series with (unvalidated) synop data has some consequences for the quality of the resulting blended series. This is the principle motivation to limit the length of the synop data series which is added to existing validated series. These issues are the subject of van den Besselaar et al. (2012).

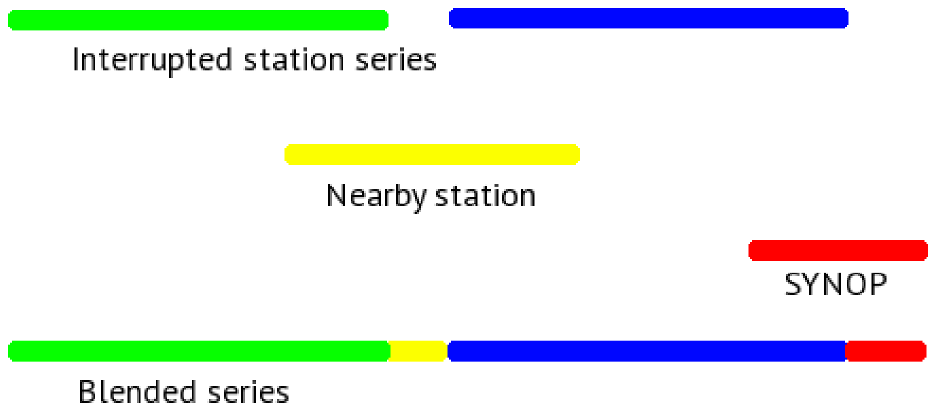

Figure 1: Blending figure

5 Indices calculation

5.1 Design rules

Indices are calculated for the *mixed* blended series only and over a time span which is as long as the record allows. For an index to be calculated for a particular year, at least 350 days with valid daily data must exist. For an index to be calculated for a half-year period, at least 175 days with valid daily data must exist. For an index to be calculated for a seasonal period, at least 85 days with valid daily data must exist. For an index to be calculated for a monthly period, at least 25 days with valid daily data must exist. Indices results are stored in the database only if a series contains at least 10 years of valid data.

A total of 72 indices are calculated on the basis of the blended daily series for the categories Cold, Drought, Heat, Pressure, Rain, Snow, Sunshine, Temperature, Wind and Compound. The acronyms are: TG, TN, TX, DTR*, ETR, GD4, GSL*, vDTR, CFD, FD*, HD17, ID*, CSDI*, TG10p, TN10p*, TX10p*, SU*, TR*, WSDI*, TG90p, TN90p*, TX90p*, RR*, RR1, SDII*, CDD, CWD*, R10mm*, R20mm*, RX1day*, RX5day*, R75p, R75pTOT, R95p, R95pTOT*, R99p, R99pTOT*, PP, SPI3, SPI6, SD, SS, TXx*, TNx*, TXn*, TNn*, SSp, PET, SD1, SD5cm, SD50cm, CD, CW, WD, WW, CSU, RH, CC, CC2, CC6, PRCPTOT, UTCI, TCI, TCI60, TCI80, HI, BEDD, FXx, FG6Bft, FGcalm, FG, DDnorth, DDsouth, DDwest, DDeast. Those with * are part of the ETCCDI list of 27 worldwide indices available from `http://cccma.seos.uvic.ca/ETCCDI/indices.shtml`.

For the CDD and CWD spell duration indicators, a spell can continue into the next year and is counted against the year in which the spell ends. For example, a dry spell (CDD) beginning in December 2000 and ending in

January 2001 is contributes to the dry spells of 2001.

The exact definition of each index is given in the next sections. Each index is calculated as annual , winter half-year (ONDJFM), summer half-year (AMJJAS), winter (DJF), spring (MAM), summer (JJA), autumn (SON) and monthly values.

5.2 Calculation of percentiles

Zhang et al. (2005) brought to the attention that percentiles, calculated on the basis of data from a 'base'-period of the record, and subsequently applied to data from the 'out-of-base' period, will introduce inhomogeneities in the resulting exceedance series. The inhomogeneities are strongest for high percentiles and for data with strong auto correlation.

In their article, they offer an alternative way to calculate percentiles when they are applied to the base-period. This method of calculating percentiles is adopted by ECA&D. This procedure is: (Zhang et al. 2005, §4)

1. The 30-yr base period is divided into one 'out of base' year, the year for which exceedance is to be estimated, and a 'base period' consisting of the remaining 29 yr from which the thresholds would be estimated.

2. A 30-yr block of data is constructed by using the 29-yr base period dataset and adding an additional year of data from the base period (replicating one year in the base period). This constructed 30-yr block is used to estimate thresholds.

3. The out-of-base year is then compared with these thresholds, and the exceedance rate for the out-of-base year is obtained.

4. Steps 2 and 3 are repeated an additional 28 times, by repeating each of the remaining 28 in-base years in turn to construct the 30-yr block.

5. The final index for the out-of-base year is obtained by averaging the 29 estimates obtained from steps 2, 3 and 4.

5.2.1 Empirical quantile estimation

The quantile of a distribution is defined as

$$Q(p) = F^{-1}(p) = \inf\{x : F(x) \geq p\}, 1 < p < 1$$

where $F(x)$ is the distribution function. Let $\{X_{(a)}, \ldots, X_{(n)}$ denote the order statistics of (i.e. sorted values of X), and let $\hat{Q}_i(p)$ denote the ith sample quantile definition. The sample quantiles can be generally written as

$$\hat{Q}_i(p) = (1-\gamma)X_{(j)} + \gamma X_{(j+1)}.$$

Hyndman & Fan (1996) suggest a formula to obtain medium un-biased estimate of the quantile by letting $j = \text{int}(p * n + (1 + p)/3))$ and letting $\gamma = p * n + (1 + p)/3 - j$, where $\text{int}(u)$ is the largest integer not greater than u. The empirical quantile is set to the smallest or largest value in the sample when $j < 1$ or $j > n$ respectively. That is, quantile estimates corresponding to $p < 1/(n + 1)$ are set to the smallest value in the sample, and those corresponding to $p > n/(n + 1)$ are set to the largest value in the sample.

5.3 Smoothing of indices

Next to the actual index values, smoothed index values are provided based on the application of a LOWESS (locally weighted scatterplot smoothing) smoother. This smoother fits simple models to localized subsets of the data to build up a function that describes the deterministic part of the variation in the data, point by point.

The code is based on routines provided by W. S. Cleveland (Bell Laboratories, Murray Hill NJ).

The smoother span f gives the proportion of points in the plot which influence the smooth at each value. The value of f is set to:

$$f = \frac{30}{\text{length of record in years}}.$$

This gives higher values for f when the length of the series is short, giving more smoothness.

The number of 'robustifying' iterations which should be performed is set to 3.

The parameter δ is used to speed up computation: instead of computing the local polynomial fit at each data point it is not computed for points within δ of the last computed point, and linear interpolation is used to fill in the fitted values for the skipped points. This parameter is set to 1/100th of the range of the input data, which is generally regarded as a standard value.

5.3.1 Cloudiness indices

CC

- *Mean of daily cloud cover (oktas)*

Let CC_{ij} be the daily cloud cover at day i of period j. Then mean values in period j are given by:

$$CC_j = \frac{\sum_{i=1}^{I} CC_{ij}}{I}$$

CC2

- *Mostly sunny days (cloud cover $\leq$ 2 oktas) (days)*

Let CC_{ij} be the daily cloud cover at day i of period j. Then counted is the number of days where:

$$CC_{ij} \leq 2\, oktas$$

CC6

- *Mostly cloudy days (cloud cover $\geq$ 6 oktas) (days)*

Let CC_{ij} be the daily cloud cover at day i of period j. Then counted is the number of days where:

$$CC_{ij} \geq 6\, oktas$$

5.3.2 Cold indices

GD4

- *Growing degree days (sum of TG > 4 °C) (°C)*

Let TG_{ij} be the daily mean temperature at day i of period j. Then the growing degree days are:

$$GD4_j = \sum_{i=1}^{I} (TG_{ij} - 4 \mid TG_{ij} > 4\,^{\circ}C)$$

GSL

- *Growing season length (days)*

Let TG_{ij} be the mean temperature at day i of period j. Then counted is the number of days between the first occurrence of at least 6 consecutive days with:

$$TG_{ij} > 5\,^{\circ}C$$

and the first occurrence after 1 July of at least 6 consecutive days with:

$$TG_{ij} < 5\,^{\circ}C$$

CFD

- *Maximum number of consecutive frost days (TN < 0°C) (days)*

Let TN_{ij} be the daily minimum temperature at day i of period j. Then counted is the largest number of consecutive days where:

$$TN_{ij} < 0\,^{\circ}C$$

FD

- *Frost days (TN < 0° C) (days)*

Let TN_{ij} be the daily minimum temperature at day i of period j. Then counted is the number of days where:

$$TN_{ij} < 0\,^{\circ}C$$

HD17

- *Heating degree days (sum of* 17 $^{\circ}C$ *- TG) (° C)*

Let TG_{ij} be the daily mean temperature at day i of period j. Then the heating degree days are:

$$HD17_j = \sum_{i=1}^{I}(17\,^{\circ}C - TG_{ij})$$

ID

- *Ice days (TX < 0° C) (days)*

Let TX_{ij} be the daily maximum temperature at day i of period j. Then counted is the number of days where:

$$TX_{ij} < 0\,^{\circ}C$$

CSDI

- *Cold-spell duration index (days)*

Let TN_{ij} be the daily minimum temperature at day i of period j and let $TN_{in}10$ be the calendar day 10th percentile calculated for a 5-day window centred on each calendar day in the 1961–1990 period. Then counted is the number of days per period where, in intervals of at least 6 consecutive days:

$$TN_{ij} < TN_{in}10$$

TG10p

- *Days with TG < 10th percentile of daily mean temperature (cold days) (days)*

Let TG_{ij} be the daily mean temperature at day i of period j and let $TG_{in}10$ be the calendar day 10th percentile calculated for a 5-day window centred on each calendar day in the 1961–1990 period. Then counted is the number of days where:

$$TG_{ij} < TG_{in}10$$

TN10p

- *Days with TN < 10th percentile of daily minimum temperature (cold nights) (days)*

Let TN_{ij} be the daily minimum temperature at day i of period j and let $TN_{in}10$ be the calendar day 10th percentile calculated for a 5-day window centred on each calendar day in the 1961–1990 period. Then counted is the number of days where:

$$TN_{ij} < TN_{in}10$$

TX10p

- *Days with TX < 10th percentile of daily maximum temperature (cold day-times) (days)*

Let TX_{ij} be the daily maximum temperature at day i of period j and let $TX_{in}10$ be the calendar day 10th percentile calculated for a 5-day window centred on each calendar day in the 1961–1990 period. Then counted is the number of days where:

$$TX_{ij} < TX_{in}10$$

TXn

- *Minimum value of daily maximum temperature (° C)*

Let TX_{ij} be the daily maximum temperature on day i of period j. Then the minimum daily maximum temperature for period j is:

$$TXn_j = min(TX_{ij})$$

TNn

- *Minimum value of daily minimum temperature (° C)*

Let TN_{ij} be the daily minimum temperature on day i of period j. Then the minimum daily minimum temperature for period j is:

$$TNn_j = min(TN_{ij})$$

5.3.3 Compound indices

The indices CD, CW, WD, and WW are based on Beniston (2009).

CD

- *Days with TG < 25th percentile of daily mean temperature and RR < 25th percentile of daily precipitation sum (cold/dry days)*

Let TG_{ij} be the daily mean temperature at day i of period j and let $TG_{in}25$ be the calendar day 25th percentile calculated for a 5-day window centred on each calendar day in the 1961–1990 period. Let RR_{wj} be the daily precipitation amount at wet day w ($RR \geq 1.0$ mm) of period j and let $RR_{wn}25$ be the 25th percentile of precipitation at wet days in the 1961–1990 period. Then counted is the number of days where:

$$TG_{ij} < TG_{in}25 \qquad \text{and} \qquad RR_{wj} < RR_{wn}25$$

CW

- *Days with TG < 25th percentile of daily mean temperature and RR > 75th percentile of daily precipitation sum (cold/wet days)*

Let TG_{ij} be the daily mean temperature at day i of period j and let $TG_{in}25$ be the calendar day 25th percentile calculated for a 5-day window centred on each calendar day in the 1961–1990 period. Let RR_{wj} be the daily precipitation amount at wet day w ($RR \geq 1.0$ mm) of period j and let $RR_{wn}75$ be the 75th percentile of precipitation at wet days in the 1961–1990 period. Then counted is the number of days where:

$$TG_{ij} < TG_{in}25 \qquad \text{and} \qquad RR_{wj} > RR_{wn}75$$

WD

- *Days with TG > 75th percentile of daily mean temperature and RR < 25th percentile of daily precipitation sum (warm/dry days)*

Let TG_{ij} be the daily mean temperature at day i of period j and let $TG_{in}75$ be the calendar day 75th percentile calculated for a 5-day window centred on each calendar day in the 1961–1990 period. Let RR_{wj} be the daily precipitation amount at wet day w ($RR \geq 1.0$ mm) of period j and let $RR_{wn}25$ be the 25th percentile of precipitation at wet days in the 1961–1990 period. Then counted is the number of days where:

$$TG_{ij} > TG_{in}75 \qquad \text{and} \qquad RR_{wj} < RR_{wn}25$$

WW

- *Days with TG > 75th percentile of daily mean temperature and RR > 75th percentile of daily precipitation sum (warm/wet days)*

Let TG_{ij} be the daily mean temperature at day i of period j and let $TG_{in}75$ be the calendar day 75th percentile calculated for a 5-day window centred on each calendar day in the 1961–1990 period. Let RR_{wj} be the daily precipitation amount at wet day w ($RR \geq 1.0$ mm) of period j and let $RR_{wn}75$ be the 75th percentile of precipitation at wet days in the 1961–1990 period. Then counted is the number of days where:

$$TG_{ij} > TG_{in}75 \qquad \text{and} \qquad RR_{wj} > RR_{wn}75$$

UTCI

- *Mean of the Universal Thermal Climate Index*

The assessment of the thermophysiological effects of the atmospheric environment is one of the key issues in human biometeorology. To quantify these effect, the Universal Thermal Climate Index (UTCI) is developed in COST action 730.

Input to the UTCI are mean radiant temperature, air temperature, water vapour pressure and windspeed. The algorithm for the mean radiant temperature is the one given by Fanger (1970). Documentation for the mean radiant temperature and the UTCI is available via http://www.utci.org.

In ECA&D, data for daily averaged windspeed, relative humidity and sunshine duration are available which we will use as input for calculating the mean radiant temperature and the UTCI. For this purpose, sunshine duration is related to an estimate of direct solar radiation (which is required as input to the mean radiant temperature) and relative humidity is related to water vapour pressure using values of the daily maximum and minimum temperatures. The algorithms used for these conversions are documented elsewhere Allen et al. (1994b,a); Duffie & Beckman (1991).

Since the UTCI relates to human activity, we use the daily *maximum* temperature as the air temperature since we can expect outdoor human activity to occur during the day time.

In the mean radiant temperature, the projected area factor f_g is included which is related to the angle of the sun. Again since outdoor activity is in the day time, we take the average of f_g over the period 10:00h to 20:00h.

TCI

- *Mean of the Tourism Climatic Index*

The Tourism Climatic Index (TCI) represents a quantitative evaluation of world climate for the purposes of tourism and is a composit measure of the climatic well-being of tourists. The TCI is aimed at the tourists involved in site-seeing or light outdoors activities. The TCI is originally defined by Mieczkowski (1985) as a weighted sum of several factors:

$$\mathrm{TCI} = 2\,(4\mathrm{CI}_d + \mathrm{CI}_a + 2\mathrm{R} + 2\mathrm{S} + W)$$

with CI_d the daytime Comfort Index, CI_a the daily Comfort Index, R the monthly precipitation sum, S the average daily sunshine duration and W the average windspeed.

The value of the TCI vary between 100 ('Ideal') to <10 ('Impossible'). The rating categories of the Tourism Climatic Index are shown in Table 1.

Mieczkowski (1985) defines tables which relate the various inputs to the TCI (eq. 8) to climatic parameters. These discrete relations are interpolated using a chebyshev approximation to yield a continuous relation between daily climate data and the TCI.

Mieczkowski (1985) uses a thermal comfort rating based on the effective temperature as calculated by the American Society of Heating, Refrigerating and Air Conditioning Engineers (ASHRAE 1972), and Mieczkowski (1985) has translated the effective temperature in a rather ad-hoc fashion to the thermal comfort rating. The relation between temperature and humidity leading to the thermal comfort rating is captured in a simple look-up table.

The relation between precipitation, daily sunshine duration and wind are given by Mieczkowski (1985). Following Perch-Nielsen et al. (2010), the latter input has been modified. The 'wind chill index' used orginally appeared to be seriously flawed and is replaced by the wind chill equivalent temperatures defined by Osczevski & Bleustein (2005).

TCI	description
90-100	ideal
80-89	excellent
70-79	very good
60-69	good
50-59	acceptable
40-49	marginal
30-39	unfavourable
20-29	very unfavourable
10-19	extremely unfavourable
<10	impossible

Table 1: A classification scheme for the Tourism Climatic Index.

TCI60

- *Days where the Tourism Climatic Index $\geq$ 60*

The Tourism Climatic Index (TCI) represents a quantitative evaluation of world climate for the purposes of tourism and is a composit measure of the climatic well-being of tourists.

Let TCI_{ij} be the daily value of the Tourism Climatic Index at day i of period j. Then counted is the number of days where:

$$\text{TCI}_{ij} \geq 60$$

The value TCI=60 represents the lowest level where the climatic conditions for sight-seeing or light outdoors activities following the classification of the Tourism Climatic Index is ‘good’.

TCI80

- *Days where the Tourism Climatic Index ≥ 80*

The Tourism Climatic Index (TCI) represents a quantitative evaluation of world climate for the purposes of tourism and is a composit measure of the climatic well-being of tourists.

Let TCI_{ij} be the daily value of the Tourism Climatic Index at day i of period j. Then counted is the number of days where:

$$\text{TCI}_{ij} \geq 80$$

The value TCI=80 represents the lowest level where the climatic conditions for sight-seeing or light outdoors activities following the classification of the Tourism Climatic Index is ‘excellent’.

5.3.4 Drought indices

CDD

- *Maximum number of consecutive dry days (RR < 1 mm) (days)*

Let RR_{ij} be the daily precipitation amount for day i of period j. Then counted is the largest number of consecutive days where:

$$RR_{ij} < 1\ mm$$

SPI6

- *6-Month Standardized Precipitation Index*

SPI is a probability index based on precipitation. It is designed to be a spatially invariant indicator of drought. SPI6 refers to precipiation in the previous 6-month period (+ indicates wet; - indices dry).

See for details and the algorithm: Guttman (1999).

SPI3

- *3-Month Standardized Precipitation Index*

SPI is a probability index based on precipitation. It is designed to be a spatially invariant indicator of drought. SPI3 refers to precipiation in the previous 3-month period (+ indicates wet; - indices dry).

See for details and the algorithm: Guttman (1999).

PET

- *Potential EvapoTranspiration*

PET is an index which gives the FAO-endorsed potential evapotranspiration as calculated by the Penman-Monteith parametrization. Here reference crop evapotranspiration is a measure for potential evapotranspiration. Reference crop evaporation is defined as the rate of evaporation from an idealized grass reference crop with a fixed crop height of 0.12 m, an albedo of 0.23, and a surface resistance of 70 s m^{-1}. In terms of of its evaporation rate, such a crop closely resembles the reference crop of an extensive surface of short green grass cover of uniform height, actively growing, completely shading the ground, and not short of water.

The equation used for estimating the reference crop evaporation is based on the Penman-Monteith approach;

$$ET_0 = \frac{0.408\Delta(R_n - G) + \gamma\frac{900}{T+273}U_2(ea - e_d)}{\Delta + \gamma(1 + 0.34U_2)}$$

where
ET_0 : reference crop evapotranspiration
R_n : net radiation at crop surface (using ECA&D elements: sunshine duration and cloud cover)
G : soil heat flux (using ECA&D element: daily averaged temperature)
T : daily averaged temperature
U_2 : daily averaged windspeed at 2 m height (using ECA&D element: daily averaged wind speed at 10 m)
$(e_a - e_d)$: vapour pressure deficit (using ECA&D elements: relative humidity and daily averaged temperature)
Δ : slope vapour pressure (using ECA&D element: daily averaged temperature)
γ : psychrometric constant (using ECA&D element: daily averaged sea-level pressure)
900 : coefficient for the reference crop
0.34 : wind coefficient for the reference crop
This equation is referred to as the FAO Penman-Monteith equation. See Allen et al. (1994b) and Allen et al. (1994a) for details.

HI

- *Huglin Index*

The Huglin Index is an index specifically aimed at grap growth (Huglin 1978) and defined using daily averaged temperature TG_i and the daily maximum temperature TX_i for day i in the period 1 April to 30 September:

$$HI = \sum_{01/04}^{30/09} \frac{(TG_i - 10) + (TX_i - 10)}{2} K$$

where K is a daylength coefficient. The daylight coefficient is a function of the latitude of the station but a clear definition is absent. The value of K is determined using table 2.

latitude	daylight coefficient K
$\leq$ 40°N	1.00
40°N-42°N	1.02
42°N-44°N	1.03
44°N-46°N	1.04
46°N-48°N	1.05
48°N-50°N	1.06

Table 2: The daylight coefficient as used in the Huglin index as a function of latitude.

BEDD

- *Biologically Effective Degree Days*

The Biologically Effective Degree Days index has been specifically targeted to describe grape growth (Gladstones 1992). The BEDD index is based on a growing degree days measure. Let TX_i and TN_i be the daily maximum and daily minimum temperature for day i. Then BEDD is calculated by

$$BEDD = \sum_{01/04}^{30/09} \min \left[\max \left[\left(\frac{TX_i + TN_i}{2} \right) - b, 0 \right], 9 \right],$$

where $b = 10$ is an appropriate value for grape growth.

5.3.5 Heat indices

SU

- *Summer days ($TX > 25$ °C) (days)*

Let TX_{ij} be the daily maximum temperature at day i of period j. Then counted is the number of days where:

$$TX_{ij} > 25\,^{\circ}C$$

TR

- *Tropical nights (TN > 20 °C) (days)*

Let TN_{ij} be the daily minimum temperature at day i of period j. Then counted is the number of days where:

$$TN_{ij} > 20\,^{\circ}C$$

WSDI

- *Warm-spell duration index (days)*

Let TX_{ij} be the daily maximum temperature at day i of period j and let $TX_{in}90$ be the calendar day 90th percentile calculated for a 5-day window centred on each calendar day in the 1961–1990 period. Then counted is the number of days per period where, in intervals of at least 6 consecutive days:

$$TX_{ij} > TX_{in}90$$

TG90p

- *Days with TG > 90th percentile of daily mean temperature (warm days) (days)*

Let TG_{ij} be the daily mean temperature at day i of period j and let $TG_{in}90$ be the calendar day 90th percentile calculated for a 5-day window centred on each calendar day in the 1961–1990 period. Then counted is the number of days where:

$$TG_{ij} > TG_{in}90$$

TN90p

- *Days with TN > 90th percentile of daily minimum temperature (warm nights) (days)*

Let TN_{ij} be the daily minimum temperature at day i of period j and let $TN_{in}90$ be the calendar day 90th percentile calculated for a 5-day window centred on each calendar day in the 1961–1990 period. Then counted is the number of days where:

$$TN_{ij} > TN_{in}90$$

TX90p

- *Days with TX > 90th percentile of daily maximum temperature (warm day-times) (days)*

Let TX_{ij} be the daily maximum temperature at day i of period j and let $TX_{in}90$ be the calendar day 90th percentile calculated for a 5-day window centred on each calendar day in the 1961–1990 period. Then counted is the number of days where:

$$TX_{ij} > TX_{in}90$$

TXx

- *Maximum value of daily maximum temperature (°C)*

Let TX_{ij} be the daily maximum temperature on day i of period j. Then the maximum daily maximum temperature for period j is:

$$TXx_j = max(TX_{ij})$$

TNx

- *Maximum value of daily minimum temperature (°C)*

Let TN_{ij} be the daily minimum temperature on day i of period j. Then the maximum daily minimum temperature for period j is:

$$TNx_j = max(TN_{ij})$$

CSU

- *Maximum number of consecutive summer days (TX > 25°C) (days)*

Let TX_{ij} be the daily maximum temperature for day i of period j. Then counted is the largest number of consecutive days where:

$$TX_{ij} > 25°C$$

5.3.6 Humidity index

RH

- *Mean of daily relative humidity (%)*

Let HU_{ij} be the daily relative humidity at day i of period j. Then mean values in period j are given by:

$$RH_j = \frac{\sum_{i=1}^{I} HU_{ij}}{I}$$

5.3.7 Pressure index

PP

- *Mean of daily sea level pressure (hPa)*

Let PP_{ij} be the daily sea level pressure at day i of period j. Then mean values in period j are given by:

$$PP_j = \frac{\sum_{i=1}^{I} PP_{ij}}{I}$$

5.3.8 Rain indices

RR

- *Precipitation sum (mm)*

Let RR_{ij} be the daily precipitation amount for day i of period j. Then sum values are give by:

$$RR_j = \sum_{i=1}^{I} RR_{ij}$$

RR1

- *Wet days (RR ≥ 1 mm) (days)*

Let RR_{ij} be the daily precipitation amount for day i of period j. Then counted is the number of days where:

$$RR_{ij} \geq 1\ mm$$

SDII

- *Simple daily intensity index (mm/wet day)*

Let RR_{wj} be the daily precipitation amount for wet day w ($RR \geq 1.0$mm) of period j. Then the mean precipitation amount of wet days is given by:

$$SDII_j = \frac{\sum_{w=1}^{W} RR_{wj}}{W}$$

CWD

- *Maximum number of consecutive wet days (RR ≥ 1 mm) (days)*

Let RR_{ij} be the daily precipitation amount for day i of period j. Then counted is the largest number of consecutive days where:

$$RR_{ij} \geq 1\ mm$$

R10mm

- *Heavy precipitation days (precipitation ≥ 10 mm) (days)*

Let RR_{ij} be the daily precipitation amount for day i of period j. Then counted is the number of days where:

$$RR_{ij} \geq 10\ mm$$

R20mm

- *Very heavy precipitation days (precipitation ≥ 20 mm) (days)*

Let RR_{ij} be the daily precipitation amount for day i of period j. Then counted is the number of days where:

$$RR_{ij} \geq 20\ mm$$

RX1day

- *Highest 1-day precipitation amount (mm)*

Let RR_{ij} be the daily precipitation amount for day i of period j. Then maximum 1-day values for period j are:

$$RX1day_j = \max\left(RR_{ij}\right)$$

RX5day

- *Highest 5-day precipitation amount (mm)*

Let RR_{kj} be the precipitation amount for the five-day interval k of period j, where k is defined by the last day. Then maximum 5-day values for period j are:

$$RX5day_j = \max\left(RR_{kj}\right)$$

R75p

- *Days with RR > 75th percentile of daily amounts (moderate wet days) (days)*

Let RR_{wj} be the daily precipitation amount at wet day w ($RR \geq 1.0$ mm) of period j and let $RR_{wn}75$ be the 75th percentile of precipitation at wet days in the 1961–1990 period. Then counted is the number of days where:

$$RR_{wj} > RR_{wn}75$$

R75pTOT

- *Precipitation fraction due to moderate wet days (> 75th percentile) (%)*

Let RR_j be the sum of daily precipitation amount for period j and let RR_{wj} be the daily precipitation amount at wet day w ($RR \geq 1.0$ mm) of period j and $RR_{wn}75$ the 75th percentile of precipitation at wet days in the 1961–1990 period. Then $R75pTOT_j$ is determined as:

$$R75pTOT_j = 100 \times \frac{\sum_{w=1}^{W} RR_{wj},\ where\ RR_{wj} > RR_{wn}75}{RR_j}$$

R95p

- *Days with RR > 95th percentile of daily amounts (very wet days) (days)*

Let RR_{wj} be the daily precipitation amount at wet day w ($RR \geq 1.0$ mm) of period j and let $RR_{wn}95$ be the 95th percentile of precipitation at wet days in the 1961–1990 period. Then counted is the number of days where:

$$RR_{wj} > RR_{wn}95$$

R95pTOT

- *Precipitation fraction due to very wet days (> 95th percentile) (%)*

Let RR_j be the sum of daily precipitation amount for period j and let RR_{wj} be the daily precipitation amount at wet day w ($RR \geq 1.0$ mm) of period j and $RR_{wn}95$ the 95th percentile of precipitation at wet days in the 1961–1990 period. Then $R95pTOT_j$ is determined as:

$$R95pTOT_j = 100 \times \frac{\sum_{w=1}^{W} RR_{wj},\ where\ RR_{wj} > RR_{wn}95}{RR_j}$$

R99p

- *Days with RR > 99th percentile of daily amounts (extremely wet days) (days)*

Let RR_{wj} be the daily precipitation amount at wet day w ($RR \geq 1.0$ mm) of period j and let $RR_{wn}99$ be the 99th percentile of precipitation at wet days in the 1961–1990 period. Then counted is the number of days where:

$$RR_{wj} > RR_{wn}99$$

R99pTOT

- *Precipitation fraction due to extremely wet days (> 99th percentile) (%)*

Let RR_j be the sum of daily precipitation amount for period j and let RR_{wj} be the daily precipitation amount at wet day w ($RR \geq 1.0$ mm) of period j and $RR_{wn}99$ the 99th percentile of precipitation at wet days in the 1961–1990 period. Then $R99pTOT_j$ is determined as:

$$R99pTOT_j = 100 \times \frac{\sum_{w=1}^{W} RR_{wj},\ where\ RR_{wj} > RR_{wn}99}{RR_j}$$

5.3.9 Snow indices

SD

- *Mean of daily snow depth (cm)*

Let SD_{ij} be the daily snow depth at day i of period j. Then mean value of period j is given by:

$$SD_j = \frac{\sum_{i=1}^{I} SD_{ij}}{I}$$

SD1

- *Snow days (SD ≥ 1 cm) (days)*

Let SD_{ij} be the daily snow depth for day i of period j. Then counted is the number of days where:

$$SD_{ij} \geq 1\ cm$$

SD5cm

- *Number of days with SD ≥ 5 cm (days)*

Let SD_{ij} be the daily snow depth for day i of period j. Then counted is the number of days where:

$$SD_{ij} \geq 5\ cm$$

SD50cm

- *Number of days with SD ≥ 50 cm (days)*

Let SD_{ij} be the daily snow depth for day i of period j. Then counted is the number of days where:

$$SD_{ij} \geq 50\ cm$$

5.3.10 Sunshine indices

SS

- *Sunshine duration (hours)*

Let SS_{ij} be the daily sunshine duration for day i of period j. Then sum values are given by:

$$SS_j = \sum_{i=1}^{I} SS_{ij}$$

SSp

- *Sunshine duration fraction with respect to daylength (%)*

Let SS_{ij} be the daily sunshine duration amount for day i of period j and $SS_{ij}^{\max}$ the maximum daylight hours for day i of period j. Sum values in period j are given by:

$$SS_j = \sum_{i=1}^{I} SS_{ij} \text{ and } SS_j^{\max} = \sum_{i=1}^{I} SS_{ij}^{\max}.$$

The index is then given by

$$Sp_j = \frac{SS_j}{SS_j^{\max}} \times 100\%$$

The maximum daylight hours are calculated based on theory given in Allen et al. (1994a). The yearday j for month M and day D can be determined by

$$j = int\left(275\frac{M}{9} - 30 + D\right) - 2 \tag{8}$$

which is from (Allen et al. 1994a, eq. 1.26), provided that: if $M < 3$, then $j = j + 2$ and if leap year and $M > 2$, then $j = j + 1$.

Given the yearday, the maximum daylight hours N [h] can be calculated using (Allen et al. 1994a, eq. 1.34)

$$N = \frac{24}{\pi}\omega_s \tag{9}$$

where ω_s is the sunset hour angle [rad]. This can be calculated by (Allen et al. 1994a, eq. 1.23)

$$\omega_s = \arccos\left(-\tan\phi\tan\delta\right), \tag{10}$$

where δ is the solar declination [rad] (Allen et al. 1994a, eq. 1.25)

$$\delta = 0.409\sin\left(\frac{2\pi}{365}j - 1.39\right) \tag{11}$$

and ϕ the latitude [rad] of the station (negative for southern hemisphere).

5.3.11 Temperature indices

TG

- *Mean of daily mean temperature ($^\circ C$)*

Let TG_{ij} be the mean temperature at day i of period j. Then mean values in period j are given by:

$$TG_j = \frac{\sum_{i=1}^{I} TG_{ij}}{I}$$

TN

- *Mean of daily minimum temperature ($^\circ C$)*

Let TN_{ij} be the minimum temperature at day i of period j. Then mean values in period j are given by:

$$TN_j = \frac{\sum_{i=1}^{I} TN_{ij}}{I}$$

TX

- *Mean of daily maximum temperature ($^\circ C$)*

Let TX_{ij} be the maximum temperature at day i of period j. Then mean values in period j are given by:

$$TX_j = \frac{\sum_{i=1}^{I} TX_{ij}}{I}$$

DTR

- *Mean of diurnal temperature range ($^\circ C$)*

Let TX_{ij} and TN_{ij} be the daily maximum and minimum temperature at day i of period j. Then the mean diurnal temperature range in period j is:

$$DTR_j = \frac{\sum_{i=1}^{I} (TX_{ij} - TN_{ij})}{I}$$

ETR

- *Intra-period extreme temperature range ($^\circ C$)*

Let TX_{ij} and TN_{ij} be the daily maximum and minimum temperature at day i of period j. Then the extreme temperature range in period j is:

$$ETR_j = \max(TX_{ij}) - \min(TN_{ij})$$

vDTR

- *Mean absolute day-to-day difference in DTR (°C)*

Let TX_{ij} and TN_{ij} be the daily maximum and minimum temperature at day i of period j. Then calculated is the absolute day-to-day differences in period j:

$$vDTR_j = \frac{\sum_{i=2}^{I} |(TX_{ij} - TN_{ij}) - (TX_{i-1,j} - TN_{i-1,j})|}{I}$$

5.3.12 Wind indices

FXx

- *Maximum value of daily maximum wind gust (m s^{-1})*

Let FX_{ij} be the daily maximum wind gust on day i of period j. Then the maximum daily maximum wind gust for period j is:

$$FXn_j = max(FX_{ij})$$

FG6Bft

- *Days with daily averaged wind ≥ 6 Bft (10.8 m s^{-1}) (days)*

Let FG_{ij} be the daily averaged wind strength at day i of period j. Then counted is the number of days where:

$$FG_{ij} \geq 10.8\, ms^{-1}$$

FGcalm

- *Calm days (FG ≤ 2 m s^{-1}) (days)*

Let FG_{ij} be the daily averaged wind strength at day i of period j. Then counted is the number of days where:

$$FG_{ij} \leq 2\, ms^{-1}$$

FG

- *Mean of daily mean wind strength (m s^{-1})*

Let FG_{ij} be the mean wind strength at day i of period j. Then mean values in period j are given by:

$$FG_j = \frac{\sum_{i=1}^{I} FG_{ij}}{I}$$

DDnorth

- *Days with northerly winds (-45° < DD ≤ 45°) (days)*

Let DD_{ij} be the daily value of the wind direction at day i of period j. Then counted is the number of days where:

$$-45\,^\circ < DD_{ij} \leq 45\,^\circ$$

DDsouth

- *Days with southerly winds (135° < DD ≤ 225°) (days)*

Let DD_{ij} be the daily value of the wind direction at day i of period j. Then counted is the number of days where:

$$225\,^\circ < DD_{ij} \leq 315\,^\circ$$

DDeast

- *Days with easterly winds (45° < DD ≤ 135°) (days)*

Let DD_{ij} be the daily value of the wind direction at day i of period j. Then counted is the number of days where:

$$45\,^\circ < DD_{ij} \leq 135\,^\circ$$

DDwest

- *Days with westerly winds (225° < DD ≤ 315°) (days)*

Let DD_{ij} be the daily value of the wind direction at day i of period j. Then counted is the number of days where:

$$225\,^\circ < DD_{ij} \leq 315\,^\circ$$

6 Climatology calculations

6.1 Design rules

Climatologies for all indices described in Sect. 5.1 are calculated. Normal periods used in ECA&D are 1951–1980, 1961–1990, 1971–2000 and 1981–2010. A climatological value for a particular index and a particular station is calculated if at least 70% of the data are available.

These climatologies are used in the 'indices of extremes' webpages. Both anomalies of an index, for a particular year and season, can be plotted with respect to the 1961-1990 climatology, and maps of the 1951–1980, 1961–1990, 1971–2000 and 1981–2010 climatologies can be plotted.

7 Trend calculation

7.1 Design rules

A trend is calculated for each of the indices and for each of the aggregation periods for which the indices are calculated. Of all values considered in a period, at least 70% of them must contain valid index data (i.e., not missing) for the trend to be calculated.

Calculation of the trend value is done by a least squares estimate of a simple linear regression. The regression is performed by routine e02adf Numerical Algorithms Group (NAG, `http://www.nag.co.uk/`), where all points have equal weight. Data points with 'missing' values are not part of the inputdata for this routine. The routine calculates a least-squares polynomial approximation of degree 0 and 1, using Chebyshev polynomials as the basis. Subsequent evaluation of the Chebyshev-series representation of the polynomial approximation are carried out using NAG's e02aef routine. These routines give a value for the intercept a_0 and a value of the slope a_1:

$$\mathbf{y}_i = a_0 + a_1\mathbf{x}_i + \mathbf{e}_i,$$

with $\mathbf{e}_i$ a residual.

This follows (von Storch & Zwiers 1999, §8.3.8). To test the null hypothesis that the slope a_1 has a value of 0 against the hypothesis that the slope is distinguishable from 0, we calculated

$$t = \frac{a}{\left(\sigma_E/\sqrt{S_{XX}}\right)}.$$

This value is then compared against critical values from the t-distribution with $n-2$ degrees of freedom. Here

$$\sigma_E^2 = \frac{1}{n-2}\sum_{i=1}^{N}(\mathbf{y}_i - a_0 - a_1\mathbf{x}_i)^2$$

is the squared sum of errors of the fit and

$$S_{XX} = \sum_{i=1}^{N}(\mathbf{x}_i - \overline{\mathbf{x}})^2\,.$$

Because we have fitted a linear model that depends upon only one factor, the t and F tests are equivalent. In fact: $F = t^2$, and the square of a t random variable with $n-2$ degrees of freedom is distributed as $F(1, n-2)$. We will use the F-statistic here, which is identical to a two-sided t-test. The F-statistic is calculated by

$$F = \frac{SSR}{\sigma_E^2},$$

where

$$SSR = \sum_{i=1}^{N} (a_0 + a_1 \mathbf{x}_i - \overline{\mathbf{y}})^2 \,.$$

The t-test is not robust against departures from the independence assumption. In general, time series in climatology will be auto correlated. Under these circumstances, the t-test becomes too liberal and rejects the null-hypothesis too often. Having some auto correlation in a series actually decreases the number of degrees of freedom. To account for this, an estimate of the equivalent sample size is made (von Storch & Zwiers 1999, §6.6.8). The equivalent sample size is then:

$$n'_x = \frac{n_x}{1 + 2\sum_{k=1}^{n_\mathrm{x}-1} \left(1 - \frac{k}{n_\mathrm{x}}\right) \rho_x(k)}$$

where $\rho_x(k)$ is the auto correlation function and n_x the number of degrees of freedom. Note the factor 2 in the denominator; it is missing in von Storch & Zwiers (1999, eq. 6.26) but should be there.

Given the number of degree of freedom and the t-value, a significance level can be calculated. This calculation makes use of the Numerical Recipes function `BETAI` Press et al. (1989), for the calculation of the incomplete beta function.

For each of the indices described in § 5.1 the trend is calculated over the following periods:

1. 1851 – last year
2. 1901 – last year
3. 1951 – last year
4. 1901 – 1950
5. 1951 – 1978
6. 1979 – last year

8 Homogeneity analysis

8.1 Design rules

In any long time series, changes in routine observation practices may have introduced inhomogeneities of non-climatic origin that severely affect the extremes. Wijngaard et al. (2003) statistically tested the daily ECA series (1901–1999) of surface air temperature and precipitation with respect to homogeneity. Their methodology has been implemented in ECA&D. A two-step approach is followed. First, four homogeneity tests are applied to evaluate the daily series using the testing variables: (1) the annual mean of the diurnal temperature range DTR (= maximum temperature - minimum temperature), (2) the annual mean of the absolute day-to-day differences of the diurnal temperature range vDTR, (3) the annual wet day count RR1 (threshold 1 mm), (4) the annual number of snow days (SD >= 1cm) SD1, (5) the annual mean of sea-level pressure PP, (6) the annual sum of sunshine duration SS, (7) the annual mean of relative humidity RH and (8) the annual mean of cloud cover CC. The use of derived annual variables avoids auto correlation problems with testing daily series. Second, the test results are condensed for each series into three classes: 'useful–doubtful–suspect'.

The four homogeneity tests are:

1. Standard Normal Homogeneity Test (SNH, Alexandersson (1986))
2. Buishand Range test (BHR, Buishand (1982))
3. Pettitt test (PET, Pettitt (1979))
4. Von Neumann Ratio test (VON, von Neumann (1941))

All four tests suppose under the null hypothesis that in the series of a testing variable, the values are independent with the same distribution. Under the alternative hypothesis the SNH, BHR and PET test assume that a step-wise shift in the mean (a break) is present. These three tests are capable to locate the year where a break is likely. The fourth test (VON) assumes under the alternative hypothesis that the series is not randomly distributed. This test does not give information on the year of the break. The calculus of each test is described below (from Wijngaard et al. 2003).

Y_i (i is the year from 1 to n) is the annual series to be tested, $\bar{Y}$ is the mean and s the standard deviation.

8.1.1 Standard normal homogeneity test

Alexandersson (1986) describes a statistic $T(k)$ to compare the mean of the first k years of the record with that of the last $n-1$ years:

$$T(k) = k\bar{z}_1^2 + (n-k)\bar{z}_2^2 \qquad k = 1,\ldots,n$$

where

$$\bar{z}_1 = \frac{1}{k}\frac{\sum_{i=1}^{k}(Y_i - \bar{Y})}{s} \quad \text{and} \quad \bar{z}_2 = \frac{1}{n-k}\frac{\sum_{i=k+1}^{n}(Y_i - \bar{Y})}{s}$$

If a break is located at the year K, then $T(k)$ reaches a maximum near the year $k = K$. The test statistic T_0 is defined as:

$$T_0 = \max\left(T(k)\right) \qquad \text{for } 1 \le k < n$$

The test has further been studied by Jarušková (1994). The relationship between her test statistic $T(n)$ and T_0 is:

$$T_0 = \frac{n(T(n))^2}{n - 2 + (T(n))^2}$$

The null hypothesis will be rejected if T_0 is above a certain level, which is dependent on the sample size. Critical values are given in Table 3.

Table 3: 1% critical values for the statistic T_0 of the single shift SNHT as a function of n (calculated from the simulations carried out by Jarušková (1994)) and the 5% critical value (Alexandersson 1986).

n	20	30	40	50	70	100
1%	9.56	10.45	11.01	11.38	11.89	12.32
5%	6.95	7.65	8.10	8.45	8.80	9.15

8.1.2 Buishand range test

In this test, the adjusted partial sums are defined as

$$S_0^* = 0 \quad \text{and} \quad S_k^* = \sum_{i=1}^{k} (Y_i - \bar{Y}) \quad k = 1, \ldots, n$$

When a series is homogeneous the values of S_k^* will fluctuate around zero, because no systematic deviations of the Y_i values with respect to their mean will appear. If a break is present in year K, then S_k^* reaches a maximum (negative shift) or minimum (positive shift) near the year $k = K$. The significance of the shift can be tested with the 'rescaled adjusted range' R, which is the difference between the maximum and the minimum of the S_k^* values scaled by the sample standard deviation:

$$R = (\max S_k^* - \min S_k^*)/s \qquad 0 \le k \le n \quad \text{for max and min separately}$$

Buishand (1982) gives critical values for $R/\sqrt{n}$ (see Table 4).

Table 4: 1% and 5% critical values for $R/\sqrt{n}$ of the Buishand range test as a function of n (Buishand 1982); the value of $n = 70$ is simulated.

n	20	30	40	50	70	100
1%	1.60	1.70	1.74	1.78	1.81	1.86
5%	1.43	1.50	1.53	1.55	1.59	1.62

8.1.3 Pettitt test

This test is a non-parametric rank test. The ranks $r_1, \ldots, r_n$ of the $Y_1, \ldots, Y_n$ are used to calculate the statistics:

$$X_k = 2\sum_{i=1}^{k} r_i - k(n+1) \quad k = 1, \ldots, n$$

If a break occurs in year E, then the statistic is maximal or minimal near the year $k = E$:

$$X_E = \max |X_k| \quad \text{for } 1 \leq k \leq n$$

The significance level is given by Pettitt (1979). Critical values for X_E are given in Table 5.

Table 5: 1% and 5% critical values for X_E of the Pettitt test as a function of n; values are based on simulation.

n	20	30	40	50	70	100
1%	71	133	208	293	488	841
5%	57	107	167	235	393	677

8.1.4 Von Neumann ratio

The von Neumann ratio N is defined as the ratio of the mean square successive (year to year) difference to the variance (von Neumann 1941):

$$N = \sum_{i=1}^{n-1}(Y_i - Y_{i+1})^2 / \sum_{i=1}^{n}(Y_i - \bar{Y})^2$$

When the sample is homogeneous the expected value is $N = 2$. If the sample contains a break, then the value of N tends to be lower than this expected value (Buishand 1981). If the sample has rapid variations in the mean, then values of N may rise above two (Bingham & Nelson 1981). This test gives no information about the location of the shift. Table 6 gives critical values for N.

In ECA&D, test results are calculated for the following periods (identical to the trend periods):

1. 1851 – last year
2. 1901 – last year
3. 1951 – last year
4. 1901 – 1950

Table 6: 1% and 5% critical values for N of the von Neumann ratio test as a function of n. For $n \leq 50$ these values are taken from Owen (1962); for $n = 70$ and $n = 100$ the critical values are based on the asymptotic normal distribution of N (Buishand 1981).

n	20	30	40	50	70	100
1%	1.04	1.20	1.29	1.36	1.45	1.54
5%	1.30	1.42	1.49	1.54	1.61	1.67

5. 1951 – 1978
6. 1979 – last year

Of all years considered in a period, at least 70% of them must contain valid data (i.e., not missing). Only temperature series and precipitation series are tested on homogeneity. Other elements, like sea level pressure are not tested. The test results are condensed into a single flag for each series according to:

- Class 1: 'useful' – 1 or 0 tests reject the null hypothesis at the 1% level
- Class 2: 'doubtful' – 2 tests reject the null hypothesis at the 1% level
- Class 3: 'suspect' – 3 or 4 tests reject the null hypothesis at the 1% level

For temperature, where two variables are tested, the two categories are calculated separately for each variable. If the results are different, the least favourable category is assigned to the temperature series of the station. If not all 4 individual tests can be calculated the flag is 'missing'. This means the homogeneity of the series in the considered period could not be determined.

On the website the trends in the climate change indices are only presented for series that are classified as 'useful' or 'doubtful' in the considered period.

For snow cover the index SD1 (number of snow days) was used for the homogeneity tests, and for relative humidity the index RH (mean of daily relative humidity), for sea level pressure the index PP (mean of daily sea level pressure), for cloud cover the index CC (mean of daily cloud cover), and for sunshine the index SS (mean of daily sunshine). For the indices CW, CD, WW, WD and PET the homogeneity results of the temperature series are used.

9 E-OBS gridded dataset

9.1 Design rules

The E-OBS dataset is the gridded version of the ECA&D station data for precipitation, sea-level pressure, minimum, mean and maximum temperature using all the *mixed* blended series. Only the quality control flags 'valid' are taken into account, but no check is made to include the homogeneity results. E-OBS has 2 different grid resolutions. Data is made available on a 0.1 and 0.25 degree regular lat-lon grid They cover the area: 25N-75N x 40W-75E. Daily uncertainties and elevation files are made available as well, where the uncertainty is based on an ensemble of realizations. For further information on the generation of the ensemble, we refer to the Cornes et al. (2018) paper.

The E-OBS dataset targets users of regional climate models and climate change analysis. Twice a year there will be a full update covering the period 1950 to either the end of last year of until the end of the previous June. Additional, data from the current year will be made available through monthly updates.

The dataset is available as NetCDF files not only for the whole period, but also for 15 year chunks.

Users will need to register at least their e-mail address for a mailing list before they receive the location of website where to download the data.

For more information about the E-OBS gridded datasets we refer the reader to Cornes et al. (2018), Haylock et al. (2008), Hofstra et al. (2008) and van den Besselaar et al. (2011).

10 Website

10.1 Design rules

The main categories of the website are:

1. Home: homepage that introduces the project and provides news items
2. FAQ: answers to several question about the project
3. Daily data: download of bulk and customized datasets based on interactive queries of the ECA database; the results of these queries range from PDF-documents of station metadata to zipped downloadable datasets
4. Indices of extremes: visualization of indices results through diagrams and maps using similar interactive selections as for daily data
5. Project info: project information, publications based on ECA data and links to relevant external websites and related projects

The interactive web interface uses (pull down) menus that together build a query, including time period selection, station/country selection and element/index selection. Based on this query selections of daily data can be retrieved or indices/trends/anomaly plots or maps can be shown. The content of each pull down menu is linked to the choice made in another pull down menu. For instance if country selection is 'The Netherlands' only stations for that country are shown in the menu item station selection. There are no restrictions to the order of the selections. Because the website information is directly (on the fly) retrieved from the ECA database it is always up-to-date.

References

Aguilar, E., Auer, I., Brunet, M., Peterson, T., & Wieringa, J. 2003, WMO/TD No. 1186

Alexandersson, H. 1986, J. Climatol., 6, 661

Allen, R. G., Smith, M., Perrier, A., & Pereira, L. S. 1994a, ICID Bulletin, 43, 35

Allen, R. G., Smith, M., Perrier, A., & Pereira, L. S. 1994b, ICID Bulletin, 43, 1

ASHRAE. 1972, Handbook of Fundamentals (New York, USA: American Society of Heating, Refrigerating and Air-Conditioning Engineers, Inc.)

Beniston, M. 2009, GRL, 36, L07707, DOI:10.1029/2008GL037119

Bingham, C. & Nelson, L. 1981, Technometrics, 23, 285

Buishand, T. 1981, KNMI Scientific Report, de Bilt, The Netherlands

Buishand, T. 1982, J. Hydrol., 58, 11

Camuffo, D. & Jones, P. (Eds.) 2002, Climatic Change, 53

Cornes, R. C., van der Schrier, G., van den Besselaar, E. J. M., & Jones, P. D. 2018, J. Geophys. Res. (Atmospheres), 123, 9391–9409, doi:10.1029/2017JD028200

Duffie, J. A. & Beckman, W. A. 1991, Solar Engineering of Thermal Processes (New York: John Wiley & Sons, inc.)

EEA-JRC-WHO. 2008, EEA Report No 4/2008, JRC Reference Report No JRC47756, DOI:10.2800/48117

Fanger, P. O. 1970, Thermal comfort (Copenhagen: Danish Technical Press)

Gladstones, J. 1992, Viticulture and environment (Adelaide: Winetitles)

Guttman, N. B. 1999, J. Amer. Water Resources Assoc., 35, 311

Haylock, M., Hofstra, N., Klein Tank, A., et al. 2008, J. Geophys. Res., 113, D20119, DOI:10.1029/2008JD10201

Hofstra, N., Haylock, M., New, M., Jones, P., & Frei, C. 2008, J. Geophys. Res., 113, D21110, DOI:10.1029/2008JD010100

Huglin, P. 1978, Comptes Rendus de l'Académie d'Agriculture, 1117

Hyndman, R. J. & Fan, Y. 1996, The American Statistician, 50, 361

Jarušková, D. 1994, Mon. Wea. Rev., 124, 1535

Klein Tank, A., Wijngaard, J., Können, G., et al. 2002a, Int. J. of Climatol., 22, 1441

Klein Tank, A., Wijngaard, J., & van Engelen, A. 2002b, KNMI, De Bilt, the Netherlands, 36p., available from `http://eca.knmi.nl/publications`

Michalsky, J. J. 1988, Solar Energy, 40, 227

Mieczkowski, Z. 1985, The Canadian Geographer/Le Géographe canadien, 29, 220

Osczevski, R. & Bleustein, M. 2005, Bull. Amer. Meteor. Soc., 86, 1453, doi:10.1175/BAMS-86-10-1453

Owen, D. 1962, Handbook of Statistical Tables (Addison Wesley: Reading, UK)

Perch-Nielsen, S. L., Amelung, B., & Knutti, R. 2010, Climatic Change, 103, 363, doi:10.1007/s10584-009-9772-2

Pettitt, A. 1979, Appl. Statist., 28, 126

Press, W. H., Flannery, B. P., Teukolsky, S. A., & Vetterling, W. T. 1989, Numerical Recipes. The Art of Scientific Computing (FORTRAN version) (Cambridge: Cambridge University Press)

Suehrcke, H., Bowden, R. S., & Hollands, K. G. T. 2013, Solar Energy, 92, 160, doi:10.1016/j.solener.2013.02.026

van den Besselaar, E., Haylock, M., van der Schrier, G., & Klein Tank, A. 2011, J. Geophys. Res. (Atmospheres), 116, D11110, doi:10.1029/2010JD015468

van den Besselaar, E. J. M., Klein Tank, A. M. G., van der Schrier, G., & Jones, P. D. 2012, J. Geophys. Res. (Atmospheres), 117, D07101, doi:10.1029/2011JD016687

van der Schrier, G., Knap, W., Dirksen, M., et al. 2021, Environmental Science Letters, to be submitted, , -

van Engelen, A., Klein Tank, A., van der Schrier, G., & Klok, L. 2008, KNMI Publication 216, available from `http://eca.knmi.nl/publications`

von Neumann, J. 1941, Ann. Math. Statist., 13, 367

von Storch, H. & Zwiers, F. W. 1999, Statistical Analysis in Climate Research (Cambridge: Cambridge University Press)

Wijngaard, J., Klein Tank, A., & Können, G. 2003, Int.J. of Climatol., 23, 679

Zhang, X., Hegerl, G., Zwiers, F. W., & Kenyon, J. 2005, J. Climate, 18, 1641