EUMETNET/ECSN optional programme:

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

Algorithm Theoretical Basis Document (ATBD)

Projectnumber	:	EPJ029135
	:	
Author	:	Albert Klein Tank,
	:	Royal Netherlands Meteorological Institute KNMI (KS-KA)
	:	
Date	:	28 June 2007
Version	:	4

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). The project is partially funded by EUMETNET.

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 daily dataset of 20th-century surface air temperature and precipitation series has been compiled (Klein Tank et al., 2002a) and tested for homogeneity (Wijngaard et al., 2003).

To enable European climate assessments on a regular basis, a sustainable system for data gathering, archiving, quality control, analysis and dissemination is 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 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). Finally, dissemination refers to making available both the daily data (inclusive quality flags) and the indices results to users through a dedicated website.

Recently, efforts are directed towards 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 means making the system more sustainable/transparent and embedding the system into KNMIs information infrastructure to ensure ongoing support and to guarantee well-performing up-and-running services. It includes establishing general 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 report “Climate of Europe, assessment of observed daily temperature and precipitation extremes” (Klein Tank et al., 2002b) has received much praise. The project

is widely recognized as an example of KNMIs leading European role of climate data exchange and research.

The ECA&D infrastructure is used in several related activities: UNIDART (EUMETNET project) builds a uniform user interface to the ECAD database and other meteorological databases; ENSEMBLES (EU-FP6 project) develops a gridded dataset of daily temperature and precipitation for model evaluation; and MILLENNIUM (EU-FP7 project) uses a subset of long-term climate series for paleo studies.

1.3 Scope

All parts of the system above are described in this ATBD. Several related activities are however explicitly excluded.

- Producing gridded datasets or other derived datasets from the basic station data in ECA&D is not part of the system. Research activities that aim at such products (e.g. in ENSEMBLES) will be supported as much as possible.

- The analysis of indices only includes the indices defined in the former ECA project. For these indices only time series plots, trend maps and anomaly maps are presented. Area average trends for individual countries, Giorgi and Francisco (2000) - regions and/or entire Europe are not included.

1.4 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 man-power 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 or use in ENSEMBLES. 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. Available updates of data series should at least be implemented every 6 month, with newly acquired daily datasets issued and described, as well as indices results updated.

4. The minimum set of metadata for each series, which is required to judge the quality and representativiness 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. Additional public data sources have become available from EU projects, such as the IMPROVE book and CDROM (Camuffo and Jones, 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. Unlike the former ECA website, where data and indices results could only be downloaded in one fixed format, 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, inclusive 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, 2004). 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 (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 publish copies of the entire dataset. The problem is that these ad hoc 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 the help will consist of infrastructure (web) issues. 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 are not very strict. It is allowed that the system is down for several days (e.g. during the weekend), as long as the archived data are in no danger. Bringing the system up and running again at the next working day is satisfactory. 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 general backup- and maintenance procedures KNMI is employing. Optimal use should be made of KNMI information systems and infrastructure to ensure ongoing support and to guarantee up-and-running services from the ECAD 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.

1.5 Infrastructure and software

At the moment, two dedicated ECA&D systems are in use: the developer system bcsecd.knmi.nl (alias ecadev.knmi.nl) and the operational system (outside the firewall), which is a coupled system of two blades bhlbe11a.knmi.nl and bhlbe12a.knmi.nl. Each blade contains a home-directory /webdata/ecad, which is an NFS-mount to a shared disk quotom. The contents of the two blades therefore mirrors each other.

The computer systems are located in the computer room in the cellar of KNMI building A. All procedures are run on a developer platform and the results are copied to the operational platform. The operating system is Linux. The webserver is Apache. For monitoring purposes the open source software Nagios is applied http://www.nagios.org). The database is implemented on MySQL. The MySQL database is ECAD. The operational system contains no MySQL server, but makes use of an external server. For information on other packages that are used see the websites: PHP, Grads (http://grads.iges.org/grads/grads.html), Jakarta (http://tomcat.apache.org/), Open Source package R (http://www.r-project.org), Nagios (http://www.nagios.org). Log files are stored in the $HOME/logs directory. netCDF files are produced and can be accessed via a DODS server or via GrADS.

About 50% of the code in the home directory is in use. The other half is test code from various authors. Each directory with code that is relevant for the operational processes has a readme.txt file describing the main purpose of the code and the subroutines called. However, due to the redundancy and suboptimal organization of the code, it is difficult to find ones way.

1.6 Data flow

The necessary steps in data processing are:

1. New data import
2. Blending
3. Quality control
4. Indices calculation
5. Homogeneity analysis
6. Output generation
7. Website

For each step, the design rules, current implementation and necessary changes are described in the sections below.

2 New data import

2.1 Design rules

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, for indices only, or for ENSEMBLES. Public data are published on the web in addition to the indices results.

The data provided by the participants is always received with some delay. It is not possible for 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 values for the following 9 elements are derived from the SYNOP messages:

1. 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).

2. 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).

3. 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.

4. Daily mean sea level pressure PP

Whenever sea level pressure data is available at 00, 06, 12 and 18 UT, daily mean sea level pressure is calculated as $\sum$ PP / 4.

5. Daily precipitation amount RR

Whenever synoptical 12-hourly precipitation data is available at 06 and 18 UT, daily precipitation is calculated as ∑ RR.

6. Daily mean snow depth SD

Whenever synoptical snow depth data is available at 00, 06, 12 and 18 UT, daily mean snow depth is calculated as ∑ SD / 4.

7. Daily mean cloud cover CC

Whenever synoptical cloud cover data is available at 00, 06, 12 and 18 UT, mean daily cloud cover is calculated as ∑ CC / 4. This value in percent is converted to octa's by ROUND((cloudcover_in_percents/100)*8).

8. Sunshine duration SS

Whenever synoptical sunshine duration is available (in minutes) at 00, 06, 12 and 18 UT, daily sunshine duration is calculated as ∑ SS / 4.

9. Daily mean humidity HU

Whenever synoptical humidity data is available (in percents) at 06, 12 and 18 UT, daily mean humidity is calculated as ∑ HU / 3.

2.2 Current implementation

Within the ECAD 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, four different targets are distinguished. Likewise, tables have extensions for the targets: *daily*, *indices*, *grid*, *all*.

The core tables of the ECAD database are:

country:
country codes (derived from ISO 3166) and the associated country name. For example: country code 'nl' and country name 'The Netherlands'. See also:
http://www.iso.ch/iso/en/prods-services/iso3166ma/02iso-3166-code-lists/index.html

directors:
directors of the participating institutes (linked with participants-table), with their affiliation, etc.

elements:
defines all elements, and the way they are measured and/or determined.

homogeneity:
derived table with homogeneity test results for blended series

indices:
indices together with units, how they are calculated, what elements are required for index calculation, etc.

month_*xx*_blended_*target*:
derived table that contains monthly series derived from the daily series following WMO-rules.

nearby_stations:
derived table that links participant stations to nearby WMO synoptical stations (and other stations in the stations table) used for updating in the blending process.

participants:
participants with their affiliation, etc.

permissions:
types of permission flags that are allowed.

seasons:
definition of 3 months (DJF, MAM, JJA, SON) and 6 months (AMJJAS, ONDJFM) seasons

series:
links series to element, station and participant and sets permissions

series_blended_*target*:
specifies what series are used as the sources for each blended series

series_blended_*target*_derived:
calculated start and stop dates of each blended series

series_derived:
calculated start and stop dates of series

series_indices:
calculated index values and smoothed average for each location

series_trends:
calculated trends for locations, indices and seasons

series_*xx*:
daily values for element xx (see table element) for each station

series_*xx*_blended_*target*:
blended daily series for element xx for each location for each target

series_*xx*_blended_*target*_nosynop:
blended daily series for element xx for each location and target without using SYNOP updates

stations:
stations for which series are provided

synops:
all data from SYNOP messages (multiple observations per day)

synops_derived:
calculated start and stop date for each SYNOP series

synops_*xx*:
daily data derived from SYNOP messages

wmostations:
all WMO synoptical stations, which might be used to extend the data series received from participants

For retrieving the SYNOP messages from the ECMWF MARS-archive the script run.sh is used at the ECMWF machine. It is executed on a monthly basis from the crontab by "*sh run.sh –3 daily*". In this way, a new data archive is created at the ECMWF MARS-computer every month and send to KNMI. 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, 12015, 10051, 10004, 13013, 20010, 13003 and 14031.

After extraction into a ASCII-formatted file, every TX, TN, PP, RR, SD, CC, HU and SS (and the calculated TG) of a synoptical station is stored in the temporary table **synops**. When the complete ASCII-file is processed, another process reads this temporary table, determining the daily values, which are then stored into the corresponding table

synops_*xx*. Once the complete process is finished, all data from the temporary table is deleted, and the **synops_*xx*** tables hold exactly one record per day per station.

The SYNOP related processes are all combined in the shell script:

```
$HOME/apps/scripts/meta_dailyscripts.sh
```

The shell script:

```
$HOME/apps_new/administration/series_derived.sh
```

calculate the administrative rules required for blending. Each of these scripts calls R-scripts and Java-scripts to perform the actual calculations. The results are written into the ECAD database.

2.3 Necessary changes

The naming conventions are not strictly imposed. There are many derived tables (e.g. those obtained by blending series) that do not have the _derived extension. Moreover, a separation needs to be made for location information (associated with the blended series) and station information (associated with the data sources). Now these two are combined in the station table, which leads to unwanted results on the website. In many tables variables need to be renamed. At present, sta_id is often used where sta_grp or location_id is meant. Moreover, additional tables with metadata information need to be included to comply with metadata standards. This is already implemented for the ECA database in the UNIDART project.

The number of targets based on the data permissions can be reduced from the present 4 to a total of 2. Only *daily* (for public data) and *indices* (for index calculation or gridding) need to be distinguished. This means that all data that are used for gridding can also be subject to indices calculation. This will simplify the processing and improve the performance.

3 Blending

3.1 Design rules

The procedure to calculate the optimal combination of ECA station and nearby SYNOP station has 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 SYNOP 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: $LAT_{OTHER} = LAT_{OTHER} * -1$

If Longitude on western hemisphere: $LON_{OTHER} = LON_{OTHER} * -1$

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

distance = radius_earth * ARCCOS(SIN(atan*LAT_{ECA}) * SIN(atan*LAT_{OTHER}) + COS(atan*LAT_{ECA}) * COS(atan*LAT_{OTHER}) * COS(atan*(LON_{OTHER} – LON_{ECA})))

where: radius_earth = 6366.198 kilometers, and atan = ARCTAN(1)/45

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

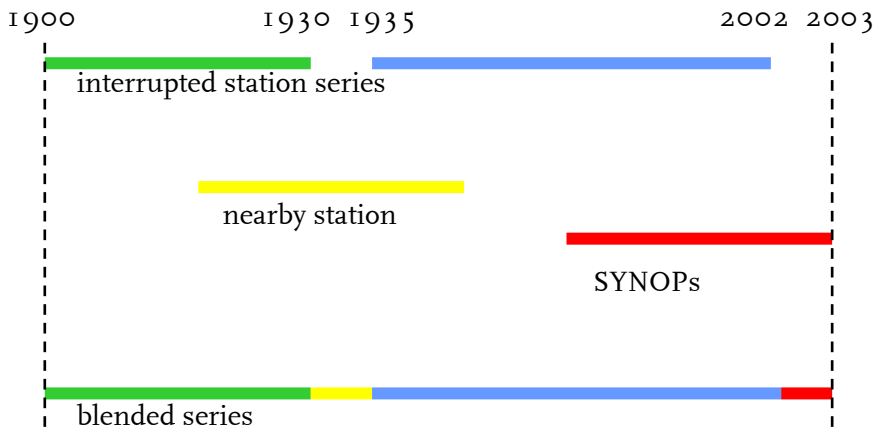
distance = radius_earth * ARCCOS(SIN(atan*52.10) * SIN(atan*LAT_{OTHER}) + COS(atan*52.10) * COS(atan*LAT_{OTHER}) * COS(atan*(LON_{OTHER} – 5.18)))

Repeat *distance* for every SYNOP station, keeping LAT_{ECA} and LON_{ECA} fixed (in the example above, for De Bilt). The SYNOP 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 25 km away from the original ECA-station, is used.

4. As a last step, the difference in elevation of the ECA station and SYNOP station is considered. Only data from SYNOP stations located within 50 m height difference is taken into account.

Next, the blended series are constructed. Suppose we have a station series from 1900 until 2002, with missing data between 1930 and 1935 and also after 2002. 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 Section 4).



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. If no valid data from nearby ECA stations is available, valid data from nearby synoptical stations is taken to "infill" the gaps. Blending is done both using SYNOP data and without using SYNOP data.

3.2 Current implementation

The shell script:

```
$HOME/apps_new/administration/nearby_stations.sh
```

calculates the optimal combinations of ECA stations and SYNOP stations. Note that, for some reason, LAT and LON are stored in the ECAD database in decimal degrees * 3600.

The shell scripts:

```
$HOME/apps/blending/do_all_blend.sh
$HOME/apps/scripts/series_blended_all_derived.sh
$HOME/apps/scripts/series_blended_all.sh
```

perform the actual blending. Each of these scripts calls R-scripts and Java-scripts to perform the actual calculations. The results are written into the ECAD database. Once the blended series are created for each individual element, new blended series are created without SYNOPS as well.

3.3 Necessary changes

The blending procedure requires that groups of stations are manually identified (which should in future result in a separate locations table). This needs to be done already upon importing new data. The present implementation is inconsistent because sta_id is used for individual stations as well as station groups. For many locations stations still have to be grouped.

At present, the preference for finding valid data from nearby ECA stations cannot be specified. In practice, the nearby station with the lowest station number that meets the criteria is taken. Since there is a known ranking in the quality of station series, a selection possibility needs to be created.

It is not clear why the blending process is repeated without using SYNOP data. These blended series without synop data are not used. A simplification here will improve performance.

4 Quality control

4.1 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 positive or zero

...must be less than 299.9 millimeters

...must not be repetitive for 10 days when amount larger than 5.0 mm

...must not be repetitive for 5 days when amount larger than 1.0 mm

...dry periods receive flag = 1 (suspect), when for the specific location the amount of dry days lies outside a 14*bivariate standard deviation

daily mean surface air pressure PP:

...must exceed 600.1 hPa

...must be less than 1080.1 hPa

...must not be repetitive for 5 days

...elevation of the recording station must not exceed 1000 meters

daily maximum temperature TX:

...must exceed –89.9 °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 for 5 days

...must be less than the long term average daily maximum temperature for that calendar day + 5 times standard deviation

...must exceed the long term average daily maximum temperature for that calendar day - 5 times standard deviation

Daily minimum temperature TN:

...must exceed –89.9 °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 for 5 days

...must be less than the long term average daily minimum temperature for that calendar day + 5 times standard deviation

...must exceed the long term average daily minimum temperature for that calendar day - 5 times standard deviation

Daily mean temperature TG:

...must exceed –89.9 °C

...must be less than 60.0 °C

...must exceed or equal daily minimum temperature (if exists)

...must be less or equals to daily maximum temperature (if exists)

...must not be repetitive for 5 days

...must be less than the long term average daily mean temperature for that calendar day + 5 times standard deviation

...must exceed the long term average daily mean temperature for that calendar day - 5 times standard deviation

Daily snow depth SD:

...must exceed or be equal to 0

...must be less than 15 meters

Daily cloud cover CC:

...must exceed or be equal to 0

...must be less than or equal to 8

Daily humidity HU:

...must exceed or be equal to 0

...must be less than or equal to 100

Daily sunshine duration SS:

...must exceed or be equal to 0

...must be less than 24

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 instance, precipitation extremes flagged "suspect" can be overruled if supplementary evidence exists (e.g. from radar images) that that particular extreme is "valid".

4.2 Current implementation

The shell scripts:

```
$HOME/apps/r_index/all_qcserie.sh
$HOME/apps/scripts/process_qca.sh
```

calculate the quality flags for the nonblended series. The shell script:

```
$HOME/apps/r_index/all_qcblend.sh
```

performs the same check for the blended series. Each of these scripts calls R-scripts and Java-scripts to perform the actual calculations. The results are written into the ECAD database.

The series_*xx* tables in the ECAD database contain three QC columns: QC, QCA and QCM. The automated procedure results in a QC flag which is stored under QCA. These flags are copied into QC, unless a manual intervention takes place and the manually specified QCM overrules the flagging based on the automated procedure.

4.3 Necessary changes

Manual interventions to assign QC flags that overrule the automated procedures have been performed in particular for precipitation extremes. For other elements this still has to be done. A check on the treatment of manual QC flags in the blending process is needed. Cases with QCM=1 should not be replaced with also erroneous SYNOP data.

5 Indices calculation

5.1 Design rules

Indices are calculated for blended series only. Indices are calculated for the period 1700 – 2020, to achieve that equal periods are compared for all indices and trends. For an index to be calculated for a particular year, at least 362 days with valid daily data must exist. For an index to be calculated for a half-year period, at least 181 days with valid daily data must exist. For an index to be calculated for a seasonal period, at least 86 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. For a trend to be calculated, at least 80% of the considered period must contain valid data. For example, when calculating a trend for the period 1901-2006, at least 80% of this period (i.e. 85 years) must contain a valid value of the index. To calculate the significance of the trends of indices, the *lm* function in R (fitting linear model) is applied. To calculate the smoothing running mean value, the *lowess* function in the R is applied. Running means are only calculated for series with at least 25 data points.

A total of 41 indices are calculated on the basis of the blended daily series. The acronyms are: RR, RR1, SDII, CDD, CWD, R10mm, R20mm, RX1day, RX5day, R75p, R75pTOT, R95p, R95pTOT, R99p, R99pTOT, SPI6, TG, TG10p, TG90p, GD4, GSL, HD17, CSFI, WSFI, TN, TN10p, TN90p, DTR, vDTR, ETR, FD, CFD, CSDI, TR, TX, TX10p, TX90p, ID, SU, WSDI, and PP.

The exact definition of each index is given in the diagrams below:

RR

- *Precipitation sum (mm)*

Let RR_{ij} be the daily precipitation amount for day *i* of period *j*. Then sum values are given 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 no 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 at wet days is given by:

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

CDD

- *Maximum no 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 no of consecutive days where:

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

CWD

- *Maximum no 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 no 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 no 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 no 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:

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

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:

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

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 no 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:

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

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:

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

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:

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

TG

- *Mean of daily mean temperature (°C)*

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

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

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 no of days where:

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

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 no of days where:

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

SPI6

- *September 6-month Standard Precipitation Index*

Derived by fitting a Gamma distribution to the 6-month precipitation totals between 1 April and 30 September.

See for details: B. Lloyd-Hughes and M.A. Saunders, 2006. A drought climatology for Europe. International Journal of Climatology, **22**, 1571-1592.

TG10p

- *Days with TG < 10th percentile of daily mean temp (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 no of days where:

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

TG90p

- *Days with TG > 90th percentile of daily mean temp (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 no of days where:

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

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:

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

GSL

- *Growing season length (days)*

Let TG_{ij} be the mean temperature at day *i* of period *j*. Then counted is the no of days between the first occurrence of at least 6 consecutive days with:

$$TG_{ij} > 5°C$$

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

$$TG_{ij} < 5°C$$

HD17

- *Heating degree days (sum of 17°C-TG) (°C)*

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

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

CSFI

- *Cold-spell 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 no of days per period where, in intervals of at least 6 consecutive days:

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

WSFI

- *Warm-spell 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 no of days per period where, in intervals of at least 6 consecutive days:

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

TN

- *Mean of daily minimum temperature (°C)*

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

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

TN10p

- *Days with TN < 10th percentile of daily min temp (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 no of days where:

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

TN90p

- *Days with TN > 90th percentile of daily min temp (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 no of days where:

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

DTR

- *Mean of diurnal temperature range (°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:

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

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 difference in period j:

$$vDTR_j = \sum_{i=1}^{I} \left| (TX_{ij} - TN_{ij}) - (TX_{i-1,j} - TN_{i-1,j}) \right| \Big/ I$$

ETR

- *Intra-period extreme temperature range (°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:

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

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 no of days where:

$$TN_{ij} < 0°C$$

CFD

- *Maximum no 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 no of consecutive days where:

$$TN_{ij} < 0°C$$

CSDI

- *Cold-spell duration index (days)*

Let TN_{ij} be the daily minimum temperature at day i of period j and let TN_{inorm} be the calendar day mean calculated for a 5 day window centred on each calendar day in the 1961-1990 period. Then counted is the no of days per period where, in intervals of at least 6 consecutive days:

$$TN_{ij} < TN_{inorm} - 5$$

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 no of days where:

$$TN_{ij} > 20°C$$

TX

- *Mean of daily maximum temperature (°C)*

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

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

TX10p

- *Days with TX < 10th percentile of daily max temp (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 no of days where:

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

TX90p

- *Days with TX > 90th percentile of daily max temp (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 no of days where:

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

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 no of days where:

$$TX_{ij} < 0°C$$

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 no of days where:

$$TX_{ij} > 25°C$$

WSDI

- *Warm-spell duration index (days)*

Let TX_{ij} be the daily maximum temperature at day *i* of period *j* and let TX_{inorm} be the calendar day mean calculated for a 5 day window centred on each calendar day in the 1961-1990 period. Then counted is the no of days per period where, in intervals of at least 6 consecutive days:

$$TX_{ij} > TX_{inorm} + 5$$

PP

- *Mean of daily surface air pressure (hPa)*

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

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

In ECA&D, the trends in indices are calculated for the following periods:

1. 1901 – last year
2. 1946 – last year
3. 1961 – last year
4. 1976 – last year
5. 1979 – last year

Of all years considered in a period, at least 80% of them must contain valid index data (i.e., not missing).

5.2 Current implementation

The shell scripts:

```
$HOME/apps/final_r_index/batch_all.sh
$HOME/apps/r_index/all_trend.sh
```

calculate the indices and trends for the year, half year and 3-months seasons. Each of these scripts calls R-scripts and Java-scripts to perform the actual calculations. The results are written into the ECAD database.

5.3 Necessary changes

For the percentile indices, we still use the old definitions. The new definitions, which deal with the jumps at the beginning and end of the reference period, are computationally too time consuming. The problems with the performance of R are likely due to inefficient programming. R-software from Environment Canada (which includes the more advanced percentile indices) calculates the indices much more efficient, and tests with other software also have shown that calculation can be more efficient. The workaround we chose for the time being is to perform the calculations for groups of stations.

The fixed time periods for which trends are calculated are now specified in the code at several places but should be flexible.

Indices for extremes in other elements than temperature and precipitation need to be developed and implemented.

6 Homogeneity analysis

6.1 Design rules

In any long time series, changes in routine observation practices may have introduced inhomogeneities of nonclimatic origin that severely affect the extremes. Wijngaard et al. (2003) statistically tested the daily ECA series (1901–99) 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 and (3) the annual wet day count RR1 (threshold 1 mm). The use of derived annual variables avoids autocorrelation 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. Petitt 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, $\overline{Y}$ is the mean and s the standard deviation.

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 $n-k$ years:

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

where

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

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

$$T_0 = \max_{1 \le k < n} T(k)$$

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 III.

Buishand range test

In this test, the adjusted partial sums are defined as

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

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 $(S_k^*/s)/\sqrt{n}$ is depicted in the graphs representing the results of this test. 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:

$$R = (\max_{0 \le k \le n} S_k^* - \min_{0 \le k \le n} S_k^*)/s$$

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

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:

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

The X_k is depicted in the graphs representing the results of this test.

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

$$X_E = \max_{1 \le k \le n} |X_k|$$

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

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):

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

Table III. 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 and Moberg, 1997)

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

Table IV. 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

Table V. 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

Table VI. 1% and 5% critical values for N of the Von Neumann ratio test as a function of n. For $n \le 50$ 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

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 and Nelson, 1981). This test gives no information about the location of the shift. Table VI gives critical values for N.

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

1. 1901 – last year
2. 1946 – last year
3. 1961 – last year
4. 1976 – last year
5. 1979 – last year

Of all years considered in a period, at least 80% of them must contain valid data (i.e., not missing). Only temperature series and precipitation series are tested on homogeneity. Other elements, like surface air 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 highest of the two category values (hence the least favourable) 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.

At the website the trends in the climate change indices are only presented for series that are classified as "useful" in the considered period.

6.2 Current implementation

The shell scripts:

```
$HOME/apps/r_index/all_homogeneity.sh
$HOME/apps/r_index/r_update_db.sh
$HOME/apps/scripts/indices_seasons_derived.sh
```

perform homogeneity analysis and move the results of indices calculation and homogeneity analysis to the final tables and update the administration.

6.3 Necessary changes

It is undesirable that the indices results obtained in an earlier section are moved to the final destination tables in the database only after homogeneity analysis is finished. This requires that indices calculation and homogeneity analysis are always run in combination. Results should be saved in the final tables after each step of data processing.

The end year (last year) for the fixed time periods for which homogeneity analysis is performed is now specified in the code at several places but should be flexible.

Homogeneity tests for other elements than temperature and precipitation need to be developed and implemented.

7 Output generation

7.1 Design rules

Extractions of the ECAD database are prepared for use on the website (bulk download of daily data), as well as use in ENSEMBLES.

For ensembles both daily data and derived monthly data are extracted for each blended series. In addition, the homogeneous sub-periods between 1961 and now are identified, using the three homogeneity tests from Section 6 that indicate the break year and iteratively apply these tests to homogeneous subsections of each blended series.

7.2 Current implementation

The shell scripts:

```
$HOME/apps/makesets/makeallsets.sh
$HOME/ensembles/make_ensembles_wmo.sh
$HOME/ensembles/make_homoperiods.sh
```

make the derived files for the data download section of the ECA&D website, the datafiles for the ENSEMBLES project and the homogeneous periods for the ENSEMBLES project.

7.3 Necessary changes

None.

8 Website

8.1 Design rules

The look and feel of the website is modeled after the former website (http://www.knmi.nl/samenw/eca) so that visitors recognize the new site as a logical evolution of the old one.

The main categories of the website are:

1. Home: homepage that introduces the project and provides news items
2. FAQ
3. Daily data: download of bulk and customized datasets based on interactive queries of the ECAD 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. Publications
6. Links: 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 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 ECAD database it is always up-to-date.

8.2 Current implementation

Most web pages are dynamically generated using scripts and queries that are embedded in php pages. In addition, a map server is active to display maps. The implemented map server is the open source development environment for building spatially-enabled internet applications: MapServer (see http://mapserver.gis.umn.edu). All documents related to the web interface rest in the htdocs directory. All functionality and interactivity is made possible without the use of high-tech utilities, like Java and Flash. In stead, a minimal configured PC with a standard browser and telephone Internet-connectivity is

considered as the main target to develop for (note: large downloads require a broadband Internet connection, though).

8.3 Necessary changes

In the original design and in many of the php pages there is an additional category "Participant login", which was intended for restricted user access (with authentication by username and password). The intention was to use this section for participants to update and upload their own data and metadata from here. However, this option has never been implemented.

A number of new functionalities are requested, which all operate along the lines of the existing functions. This means that on the fly connections to the database are made. Among these new functions are:

9 Maintainance

9.1 Design rules

Frequent backups are required of all ECA&D data, key source code and web pages. At regular intervals, all (new) information from the developer system needs to be copied to the operational system.

9.2 Current implementation

The shell scripts:

```
$HOME/apps/scripts/cleanup.sh
$HOME/scripts/backup_ECAD_database.sh
$HOME/scripts/backup_website_in_htdocs.sh
$HOME/scripts/backup_processes_in_home.sh
$HOME/scripts/backup_participantdata_in_rawdata.sh
$HOME/scripts/copy_to_external.sh
```

clean up the temporary ECAD database tables, backup the ECAD database tables and the website, processes in $home dir and participant data in $home/rawdata dir. The destination is the ecad/backup dir. Finally, the data and web pages are copied from the developer system to the operational system.

9.3 Necessary changes

It turns out that backups are made on the same system that holds the original database and code. This is unwanted and should be changed. For the time being, a workaround is implemented, which copies the relevant backup files to the workstation bhw045.knmi.nl.

References

Aguilar, E., I. Auer, M. Brunet, T.C. Peterson and J. Wieringa, 2003. Guidelines on climate metadata and homogenization. WMO/TD No. 1186.

Alexandersson H. 1986. A homogeneity test applied to precipitation data. J. Climatol. 6: 661-675.

Buishand TA. 1981. The analysis of homogeneity of long-term rainfall records in the Netherlands. KNMI Scientific Report WR 81-7: De Bilt, The Netherlands.

Buishand TA. 1982. Some methods for testing the homogeneity of rainfall records. J. Hydrol. 58: 11-27.

Camuffo, D. and P.D. Jones (Eds.), 2002. Improved Understanding of Past Climate Variability from Early Daily European Instrumental Sources. Climatic Change, Vol. 53, no. 1-3.

Giorgi, F. and R. Francisco, 2000. Uncertainties in regional climate change predictions. A regional analysis of ensemble simulations with the HADCM2 GCM. Clim. Dyn., 16, 169-182.

Jarušková D. 1994. Change-point detection in meteorological measurement. Mon. Wea. Rev. 124: 1535-1543.

Klein Tank, A.M.G. and Coauthors, 2002a. Daily dataset of 20th-century surface air temperature and precipitation series for the European Climate Assessment. Int. J. of Climatol., 22, 1441-1453.

Klein Tank, Albert, Janet Wijngaard and Aryan van Engelen, 2002b. Climate of Europe; Assessment of observed daily temperature and precipitation extremes. KNMI, De Bilt, the Netherlands, 36pp.

Lloyd-Hughes, B. and M.A. Saunders, 2006. A drought climatology for Europe. Int. J. of Climatol., 22, 1571-1592.

Pettitt AN. 1979. A non-parametric approach to the change-point detection. Appl. Statist. 28: 126-135.

Von Neumann J. 1941. Distribution of the ratio of the mean square successive difference to the variance. Ann. Math. Statist. 13: 367-395.

Wijngaard, J.B., A.M.G. Klein Tank and G.P. Konnen, 2003. Homogeneity of 20th century European daily temperature and precipitation series. Int.J. of Climatol., 23, 679-692.

Appendix

Sequence of scripts

The logical order for the full cycle of scripts to perform the entire analysis after new participant data have been added to the database or after a new SYNOPS file has become available is:

```
$HOME/apps/scripts/meta_dailyscripts.sh
$HOME/apps_new/administration/series_derived.sh
$HOME/apps_new/administration/nearby_stations.sh
$HOME/apps/r_index/all_qcserie.sh
$HOME/apps/scripts/process_qca.sh
$HOME/apps/blending/do_all_blend.sh
$HOME/apps/scripts/series_blended_all_derived.sh
$HOME/apps/scripts/series_blended_all.sh
$HOME/apps/r_index/all_qcblend.sh
$HOME/apps/final_r_index/batch_all.sh 50
$HOME/apps/r_index/all_trend.sh
$HOME/apps/r_index/all_homogeneity.sh
$HOME/apps/r_index/r_update_db.sh
$HOME/apps/scripts/indices_seasons_derived.sh
$HOME/apps/makesets/makeallsets.sh
$HOME/ensembles/make_ensembles_wmo.sh
$HOME/ensembles/make_homoperiods.sh
$HOME/apps/scripts/cleanup.sh
$HOME/scripts/backup_ECAD_database.sh
$HOME/scripts/backup_website_in_htdocs.sh
$HOME/scripts/backup_processes_in_home.sh
$HOME/scripts/backup_participantdata_in_rawdata.sh
$HOME/scripts/copy_to_external.sh
```