Indecis

Sectorial Climate Services

Integrated approach for the **d**evelopment across **E**urope of user oriented **c**limate **i**ndicators for GFCS high-priority **s**ectors: Agriculture, disaster risk reduction, energy, health, water and tourism

Work Package 3

Deliverable 3.4

Release of INDECIS-QCHDS

(Station data and gridded versions)

E. Aguilar[1], J. Gyijarro[2], L.Y.A. Randriamarolaza[1], O. Skrynyk[1]

[1]Centre for Climate Change (C3), Rovira i Virgili University (URV), Vila-seca – Spain
[5]State Meteorological Agency (AEMET), Balearic Islands Office, Spain

This report arises from the Project INDECIS which is part of ERA4CS, an ERA-NET initiated by JPI Climate, and funded by FORMAS (SE), DLR (DE), BMWFW (AT), IFD (DK), MINECO (ES), ANR (FR), with co-funding by the European Union's Horizon 2020 research and innovation programme

TABLE OF CONTENTS

1. Introduction

The Integrated approach for the development across Europe of user-oriented climate indicators for GFCS high-priority sectors: agriculture, disaster risk reduction, energy, health, water and tourism (INDECIS, http://www.indecis.eu/index.php) was developed to work on the entire Climate Services process, i.e., "from data to service". Quality assurance check and homogenization procedures applied to empirical time series of the essential climate variables (ECV) are the key milestones of the Climate Service creating (Aguilar et al., 2003). To this end, INDECIS WP3, Data Quality and Homogeneity, has produced quality controlled and homogenized datasets of the target ECVs at the daily time resolution using automated procedures, integrating existing approaches and developing new ones.

2. INDECIS Quality Control software INQC

The project INDECIS included the creation of a new software suite for quality control of daily climatological data. This **IN**DECIS **Q**uality **C**ontrol Software, **INQC**, is available as a standard R-CRAN package (https://CRAN.R-project.org/package=INQC). It can Quality Control (QC) daily data of air temperature (maximum, mean, minimum), atmospheric precipitation, sea level pressure, relative humidity, cloud cover, sunshine duration, snow depth and wind speed.

In the framework of the project, the INQC software was intended to work with the entire European Climate Assessment and Dataset (ECA&D) dataset, so it provides a reasonably fast solution to perform QC of large datasets of daily climatological time series and automatically flag errors (e.g, physically impossible values, tx > tn), suspect values (e.g., statistical outliers) and values which represent a collective error (i.e., flat strikes). The software includes more than 20 fully parametrizable quality assurance tests, which can be run by issuing a single wrapper command. This command creates flagged series and statistics, which a software user (a climatologist) can use to either refine the QC by manual

checks or to filter out the flagged values. In addition, if the user needs a more personalized experience, each test can be run as a stand-alone function.

As it was mentioned above, originally INQC was created in order to QC the ECA&D station data. However, INQC's functions can be used to deal with quality control problems in any other climatological data sets with the daily time resolution. To facilitate its application, additional functions, which can transform input data files from the HOME Cost and RClimDex formats to the ECA&D one were developed.

3. Homogenization software Climatol

The R package Climatol (https://CRAN.R-project.org/package=climatol) is a homogenization software that has been selected as the main homogenization tool in the INDECIS project. The effectiveness of the software has been evaluated in several benchmark tests where it demonstrated good results, which are comparable in terms of accuracy to other well established and tested homogenization algorithms. One of Climatol's key feature characteristics is that it can be used automatically, which significantly increases its applicability to large data sets such as ECA&D (Klein Tank et al., 2002). Several versions of the software have been released since its creation. In the scope of the INDESIS project, Climatol 3.1.1. was used.

The Climatol detection method (Guijarro, 2018) is based on the standard normal homogeneity test (SNHT) (Alexandersson, 1986). For any candidate time series, Climatol uses data from neighboring stations to create a single composite reference series as their optionally weighted average. This composite series is used further to create time series of anomalies (in order to detect breaks) and to estimate all missing data and all sub-periods/segments after break point detected. From the statistical point of view, the approach employed in the estimation process is equivalent to applying a type II linear regression model.

4. Release of INDECIS quality controlled and homogenized data sets

Quality controlled and homogenized versions of the ECA&D station time series were published in the Data Portal of the INDECIS project (http://www.indecis.eu/data.php). Only PUBLIC series listed in the corresponding files 'permissions.txt' are intended for downloading and usage.

For instance, Table 1 contains information regarding available quality controlled files for each climate variable. In addition to the file permissions.txt, three folders can be found in the Data Portal (http://www.indecis.eu/ecad_data_series.php):

QCConsolidated: files with identical names to those at the ECA&D website and with identical format (check https://www.ecad.eu for details). The QC field includes additional codes:

0: the observations passed QC
1: the observation is erroneous
2: the observations is strongly suspect
3: the obsercation is suspect
4: the observations failed a collective QC test (e.g., data strikes)

QC: full report of all the tests passed by each observation, presented in a spreadsheet-like format. Rows are observations, columns are QC tests applied. A "1" in any column indicates that this observation did not pass this particular test.
QCSummary: statistical summary of the QC process.

After applying INQC, the quality controlled time series were homogenized with the Climatol software. The INDECIS homogenized ECA&D station data can be downloaded from (http://www.indecis.eu/ecad_data_homog.php).

Table 1. Summary of the QCed station ECA&D data files available in the INDECIS Data Portal

Variable	permission.txt		
	Total	Public	Non-public
CC	2562	2347	215
FG	2132	1852	280
HU	3302	2831	371
PP	1951	1748	203
RR	18500	14169	4331
SS	1290	1083	207
TN	7646	6007	1639
TX	7547	5939	1608

5. Conclusions

Obtaining in the scope of the INDECIS project the ECA&D time series, quality controlled with INQC and homogenized with Climatol, is an important milestone in the climate service production. In addition, the release of these data on the project Data Portal can facilitate and promote their further usage in different aspect of climate analysis.

References

Aguilar, E., Auer, I., Brunet, M., Peterson, T.C. and Wieringa J. (2003) WMO Guidelines on climate metadata and homogenization. WCDMP No.53, WMO-TD No. 1186, WMO, Geneva, Switzerland.

Alexandersson, H. (1986) A homogeneity test applied to precipitation data. *Journal of Climatology*, 6(6), 661-675. https://doi.org/10.1002/joc.3370060607.

Guijarro, J.A. (2018) Homogenization of climatic series with Climatol. Version 3.1.1. Guide

Klein Tank, A.M.G., Wijngaard, J.B., Können, G.P. *et al.* (2002) Daily dataset of 20th-century surface air temperature and precipitation series for the European Climate Assessment. *International Journal of Climatology*, 22(12), 1441-1453. https://doi.org/10.1002/joc.773.