**I**ntegrated 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 3.2-b

# Recommended Homogenization Techniques based on Benchmarking Results

José A. Guijarro (AEMET)

Version 2, 28/02/2019

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.

**Summary**

This report summarizes the procedures followed to generate homogeneous databases, their perturbation with realistic inhomogeneities and other quality problems and missing data to be used as benchmarks for the evaluation of available homogenization and quality control packages applied to the nine essential climatic variables used in the INDECIS project. The results of the benchmarking are discussed, deriving recommendations on the applicability of the tested packages.

(A separate deliverable 3.2-a deals with the study of parallel series.)

**Index**



## 1. Introduction

It has long been recognized the need to remove unwanted perturbations form observational series before drawing conclusions about climate variability. Many methods have been developed so far to homogenize the series, and extensive comparisons on their performance were done during the COST Action "HOME" (Venema *et al.*, 2012) and the Spanish project MULTITEST (Guijarro *et al.*, 2017), but they were limited to monthly series of temperature and precipitation. Worth mentioning is the benchmarking exercise on daily mean temperature series undertaken by Killick (2016).

The INDECIS project needs homogenized series of nine different variables at the daily scale, and none or very little experience exists on most of them. Hence the need of the work developed by Work Package 3, devoted to assess the reliability of existing homogenization packages when applied to the homogenization of these varied set of daily climatological variables.

In the first place, a study has been done by means of selected pairs of observed parallel series in order to determine their behavior in the real world. Secondly, benchmark datasets of homogeneous series and different kind of perturbations were generated to test available homogenization packages, compare their results, and study their suitability to homogenize the real series stored at the European Climate Assessment & Dataset (Klein Tank *et al.*, 2002), that will allow the rest of the teams to compute their indices and study their variability.

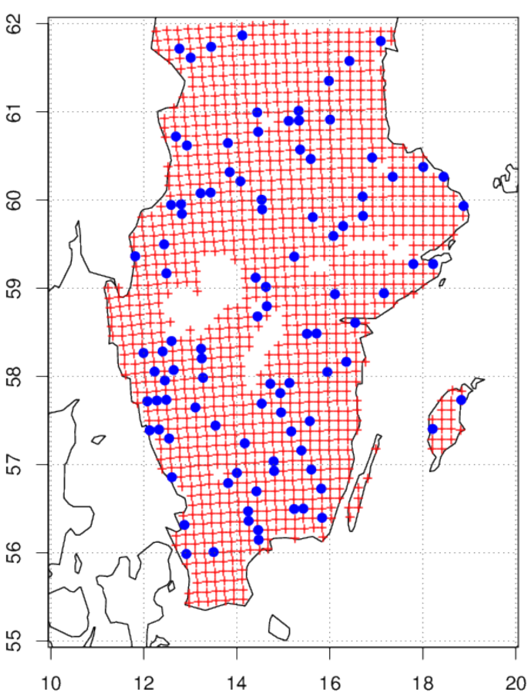
## 2. Generation of Benchmarks

### *2.1. Homogeneous databases*

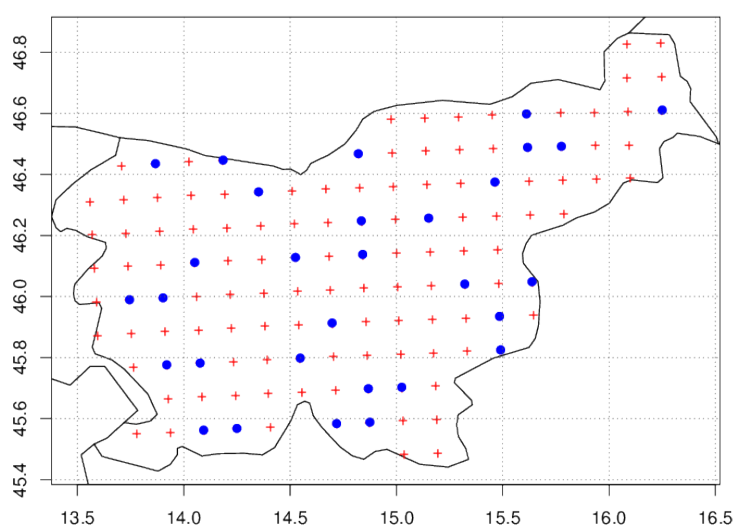
After considering other alternatives to obtain suitable datasets with granted homogeneity, it was decided that the best way would be to use the outputs of a Regional Climate Model. The Royal Netherlands Meteorological Institute (KNMI) Regional Atmospheric Climate Model (RACMO) version 2, driven by Hadley Global Environment Model 2 - Earth System (MOHC-HadGEM2-ES) (Collins et al., 2008), was chosen because it has a high spacial resolution (0.11 ◦ ) and provides the studied variables at daily time steps.

The series were obtained from the historical experiment (1950-2005) for two European areas with different climates: southern Sweden and Slovenia. A random selection of 100 and 30 points for South Sweden and Slovenia respectively was used to build the homogeneous benchmarks:

**Southern Sweden grid points and a 100 sample**




**Slovenian grid points and a 30 sample**



These RCM series use a 360 days/year calendar. As this can pose problems to homogenization packages, normally devised to deal with a realistic calendar, the 5 or 6 lacking days were evenly distributed along the year and their data were estimated by a linear interpolation.

The nine variables drawn from the RCM simulation were (units as in ECA&D):

1. CC : Cloud Cover (oktas)
2. FG : Wind Speed (0.1 m $s^{-1}$)
3. HU : Relative Humidity (1 %)
4. PP : Sea Level Pressure (0.1 hPa)
5. RR : Precipitation Amount (0.1 mm)
6. SD : Snow Depth (cm)
7. SS : Sunshine (0.1 hours)
8. TN : Minimum Temperature (0.1 °C)
9. TX : Maximum Temperature (0.1 °C)

### *2.2. Production of the inhomogeneous benchmarks*

The introduction of biases in the homogeneous series was done by simulating relocations. First, closest pairs of RCM series were used to build a database of differences and ratios between nearby locations. Then, for every random sub-period to perturb in the homogeneous series, a difference (or ratio, depending on the variable) series was randomly chosen and applied, modified by a random factor, to bias the sub-period.

The mean break frequency was set to 4 in 100 years (as found in previous studies on European series), and network-wide trends were introduced in some variables (0.4 °C/100 years in temperature, -5%/100 years in precipitation and wind, and -2%/100 years in relative humidity).

Finally, missing values and other quality problems (different from biases) were added to generate four different flavors of the perturbed benchmarks out of the initial clean (homogeneous) RCM dataset. They were archived into files named as **cc_fF-ver.tgz**, where:

**cc**=country code (se=Sweden; si=Slovenia)

**F**=Flavor (B=Breaks, C=Clean, M=Missing data, Q=Quality issues/not breaks):

C (AKA ”Baboon”)

B (AKA ”Weird Baboon”)

BM (AKA ”Angry Baboon”)

QM (AKA ”Tall Baboon”)

QMB (AKA ”Ballistic Baboon”)

**ver**=version (csv=CSV; ecad=ECAD)

These archives can be downloaded from http://www.indecis.eu/benchmarking.php. Most are offered in CSV version only, but conversion to/from ECAD format is straightforward by means of an R script.

Warning: Biases, missing data and quality issues were applied **independently** to each variable. Therefore, no consistency can be expected between, e.g., maximum and minimum temperatures. (Thanks to Yizhak Yosef for noting this.)

## 3. Benchmarking results

### *3.1. Statistical metrics to evaluate the results*

Homogenized series resulting from methodologies applied by the participants are expected in the same CSV format as that of the inhomogeneous benchmark (keeping unaltered the order of the columns: no station ID checking will be performed). Evaluation statistical metrics are then calculated by:

a) Reading the Clean benchmark (the solution)

b) Reading the homogenized series provided by each method.

c) Calculating Root Mean Squared Errors (RMSE) and differences in means, standard deviations, maximum values, minimum values and trends (6 metrics) on:

c.1) The daily series

c.2) Monthly series of means, standard deviations, maximum values, minimum values, percentile 10 and percentile 90 (6 monthly series).

c.3) Monthly series of special indices for three selected variables:

RR: no. of days reaching or exceeding 1, 10 and 30 mm (3 monthly series).

TN: no. of days with TN<=0 °C and TN>=20 °C (2 monthly series).

TX: no. of days with TX>=25 °C (1 monthly series).

### *3.2. Results of the applied methods*

As of the time of writing this document (February 28th, 2019), complete results for all benchmark flavors are available only from ACMANT v.4 (AC4) and CLIMATOL v.3 (C3m), provided by their developers (Peter Domonkos and José A. Guijarro respectively). Additionally, Petr Štěpánek sent results from his Distribution Adjustment by Percentiles method (DAP) for benchmark flavor B and Southern Sweden.

The big number of metrics described in the previous section has been graphically summarized by means of box-plots, which are available in a separate annex named INDECIS-WP3_boxplots.pdf. In order to avoid an excessive number of pages, only boxplots of RMSE and trend errors have been included in this document, that will be timely updated as new results are provided by the participants.

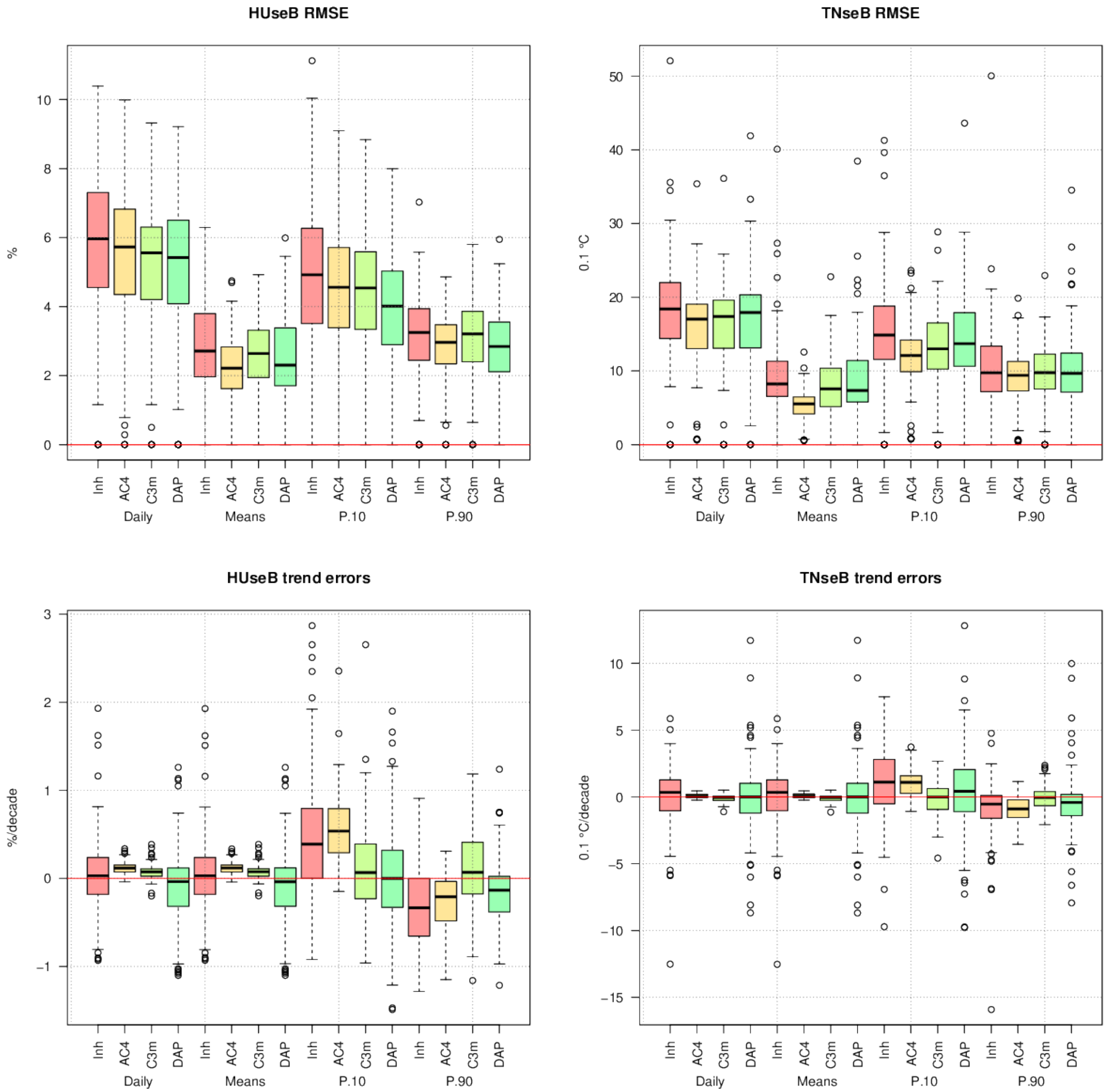



Fig. 1: Examples of evaluation metrics displayed as box-plots. Root Mean Squared Errors (top row) and trend errors (bottom row) of Southern Sweden relative humidity (left column) and minimum temperature (right column) in Weird Baboon (breaks only, no missing data) for inhomogeneous (problem) series (Inh) and results by ACMANT (AC4), CLIMATOL (C3m) and DAP. Both metrics are calculated on Daily series and on monthly Means and Percentiles 10 and 90 (P10 and P90).

*3.2.1. ”Weird Baboon” (flavor B: Biases only)*

Cloud Cover (CC) is a variable with a limited number of values (0 to 8, in integers) and no breaks seem to have been detected on these series in Sweden. Also very small reductions of trend errors appear in Slovenian series.

Wind Speed (FG): Boxplots show a clear reduction of trend errors in both areas, although RMSE are improved only in Sweden.

Relative Humidity (HU): RMSE do not show a dramatic improvement, but again it is in trend errors where the benefits of homogenization are more noticeable.

Surface Pressure (PP): This variable seems difficult to homogenize. No improvements are seen in RMSE, and the reduction in trend errors is clear only in Slovenia.

Total Rainfall (RR): The spatial and temporal variability of this element makes it quite difficult to homogenize. Therefore, very little improvement (if any) is achieved in both RMSE and trend errors.

Snow Depth (SD): This variable appeared as the most difficult to homogenize, to the point that no results are offered for their series, whose homogenization is discouraged.

Sunshine Duration (SS): No reduction of RMSE is achieved after homogenization, and only a reduction of trend errors can be appreciated.

Minimum Temperature (TN): The reduction in RMSE is not dramatic, but trend errors are very much reduced after homogenization in both areas.

Maximum Temperature (TX): Only in Slovenia can be seen a clear reduction in RMSE, but trend errors are dramatically improved in both areas.

*3.2.2. ”Angry Baboon” (flavor BM: Biases plus missing data)*

Currently here are less results for this benchmark. Results are similar to those of Weird Baboon, although the presence of missing data adds more difficulty and reduces a bit the benefits of the homogenization.

*3.2.3. ”Tall Baboon” (flavor QM: Quality issues plus missing data)*

Tall Baboon (and Ballistic Baboon) are not intended to be homogenized directly, since they are provided to test Quality Control software on them. However, homogenization methods can be applied to test their robustness to the (sometimes big) quality problems introduced in these benchmarks.

Overall, metrics do not change much (as expected from a benchmark without breaks), but bigger outliers are usually removed by CLIMATOL.

*3.2.4. ”Ballistic Baboon” (flavor QMB: Quality issues, missing data and biases)*

In this case the correction of breaks improve RMSE a bit, and trend errors are much reduced in many variables. The exceptions are total rainfall (RR) and sunshine hours (SS).

## 4. Discussion and recommendations

Results analyzed so far show Snow Depth (SD) as the most difficult variable to homogenize, due to its high proportion of zeros and strong seasonality. Right now the recommendation is to use their raw series, without attempting any homogenization.

The benefits of homogenization of Cloud Cover (CC), Surface Pressure (PP), Total Rainfall (RR) and partially in Sunshine Duration (SS) seem quite limited in their reduction of errors. Anyway, as no worsening is generally introduced in these series, their homogenization can contribute to eliminate outliers and infill any missing data, facilitating the calculation of climate indices from them.

Both ACMANT and CLIMATOL produced clear improvements in the other variables. ACMANT showed a better reduction of RMSE and less biased trend corrections in wind speed (FG), while CLIMATOL is best in relative humidity (HU), differences being less noticeable in other variables. Therefore, both methods could be applied to the homogenization of ECAD daily series. ACMANT has in favor its simplicity and a shorter computing time, while CLIMATOL provides lots of diagnostic graphics and is more robust to outliers and missing data, returning always complete series with all gaps infilled.

## 5. Conclusions and future work

Benchmarking results available so far indicate that the homogenization of the ECAD dataset can be initiated without delay, in order to provide homogenized series to the other INDECIS teams in a timely manner.

That does not preclude the analysis of new homogenization results provided by participants in a near future, which will greatly improve the knowledge about the performance of different methods on the little explored world of the homogenization of daily climate series.

## 6. Acknowledgements

Project INDECIS 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 (Grant 690462).

ECAD and EURO-CORDEX are also acknowledged, since their data are been crucial for this project.

Especial thanks are given to Núria Pérez Zanón for having extracted the RCM series and studied their suitability as a realistic homogeneous dataset, and to Antonello Squintu and Victor Venema for their invaluable contributions to a discussion meeting in Vila-seca where most strategic procedures were decided.

## 7. References

Collins, W.J., Bellouin N., Doutriaux-Boucher M., Gedney N., Hinton T., Jones C. D., Liddicoat S., Martin G., O'Connor F., Rae J., Senior C., Totterdell I., Woodward S., Reichler T., Kim J. (2008): Evaluation of the HadGEM2 model. MetOffice Hadley Centre Technical Note 74, 47 pp.

Guijarro, J.A., López, J.A., Aguilar, E., Domonkos, P., Venema V.K.C., Sigró J., Brunet, M. (2017): Comparison of homogenization packages applied to monthly series of temperature and precipitation: The MULTITEST project. 9 th Seminar for homogenization and quality control in climatological databases and 4 th conference on spatial interpolation techniques in climatology and meteorology (Budapest, April 3-7), WMO WCDMP-No.85, pp. 46-62.

Killick, R.E. (2016): Benchmarking the Performance of Homogenisation Algorithms on Daily Temperature Data. PhD Thesis, University of Exeter, 249 pp.

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

Venema, V., Mestre, O., Aguilar, E., Auer, I., Guijarro, J.A., Domonkos, P., Vertacnik, G., Szentimrey, T., Stepanek, P., Zahradnicek, P., Viarre J., Müller-Westermeier, G., Lakatos, M., Williams, C.N., Menne, M., Lindau, R., Rasol, D., Rustemeier, E., Kolokythas, K., Marinova, T., Andresen, L., Acquaotta, F., Fratianni, S., Cheval, S., Klancar, M., Brunetti, M., Gruber, C., Prohom Duran, M., Likso, T., Esteban, P. And Brandsma, T. (2012): Benchmarking homogenization algorithms for monthly data. Clim. Past, 8, 89-115.