



The EUSTACE daily LSAT dataset: A collection of 35000+ unique in-situ daily temperature series with inhomogeneities identified

Yuri Brugnara (1,2), Renate Auchmann (1,2), Stefan Brönnimann (1,2), and Elizabeth Good (3)

(1) Oeschger Centre for Climate Change Research, University of Bern, Bern, Switzerland, (2) Institute of Geography, University of Bern, Bern, Switzerland, (3) Met Office Hadley Centre, Exeter, United Kingdom

In-situ land surface air temperature (LSAT) observations are affected by several problems, in particular by inconsistencies in space and time in the measurement procedures, and by the lack of information concerning these procedures. Moreover, human errors often contaminate the quality of the data in different ways. The absence of an official global repository for meteorological observations means that often data undergo numerous modifications by different users and different versions of the same record can be found even within the same data collection.

Within the EUSTACE project, which aims at a global reconstruction of daily temperature by combining in-situ and satellite observations, we assembled public and restricted collections of in-situ daily maximum and minimum air temperature observations. We removed duplicate data by comparing each year of each record with those of the neighbouring stations, in order to minimise the loss of unique observations. Moreover, we applied an advanced quality control algorithm to all data series, while station coordinates were checked against a Digital Elevation Model. We also carried out a homogeneity assessment for each record, by using four different breakpoint detection algorithms (three relative tests and one absolute test) and taking into account additional relevant information (changes of data source, changes of reporting resolution, data gaps).

The final result is a quality-controlled dataset of ca. 750 million unique observations over the period 1850-2015 that will be assimilated into other EUSTACE products. Each breakpoint is accompanied by a measure of its likelihood, so that each user can customise the results depending on her/his needs. The dataset will be publicly available in NetCDF format on the Centre for Environmental Data Analysis data portal (<http://www.ceda.ac.uk>), with the exception of a small fraction of the data (due to data policy limitations) that can be requested to the original data providers; nevertheless, the information derived from this fraction (e.g., breakpoints) will be public.