

The EUSTACE daily LSAT dataset: A collection of 36000+ unique in-situ daily temperature series with inhomogeneities identified

Yuri Brugnara (1,2) and Stefan Brönnimann (1,2)

(1) Oeschger Centre for Climate Change Research, University of Bern, Bern, Switzerland, (2) Institute of Geography, University Of Bern, Bern, Switzerland

EUSTACE is a EU Horizon 2020 project whose goal is to produce daily estimates of surface air temperature since 1850 across the globe for the first time, by combining surface and satellite data using novel statistical techniques. As part of this, tens of thousands of in-situ data series were collected from various providers.

Surface in-situ 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 series can be found even within the same collection, with usually no information on the modifications applied.

Within EUSTACE public collections (mainly GHCN-D, ISTI, and ECA&D) of daily maximum and minimum temperature observations, together with some additional dataset with restricted data policy, were merged. Duplicates were removed by comparing the observations in each individual year of each series with the data of the stations located within a 200 km radius. Moreover, an automatic quality control algorithm was applied to all series. We also carried out an homogeneity assessment, by using four different breakpoint detection algorithms (three relative tests and one absolute test), and provide breakpoint locations for each series.

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. The dataset will be publicly available in NetCDF format, with the exception of a small fraction of the data (due to the policy of data providers); nonetheless the information derived from this fraction (e.g., for the homogeneity tests) will be public.