Digitization of Historical Observations for ERA-CLIM

Alexander Stickler

Oeschger Centre for Climate Change Research and Inst. of Geography, Univ. Bern, Switzerland

Stefan Brönnimann

Oeschger Centre for Climate Change Research and Inst. of Geography, Univ. Bern, Switzerland

Maria Antónia Valente Instituto Dom Luiz, Univ. Lisbon, Portugal

> Alexander Sterin RIHMI-WDC, Obninsk, Russia

Sylvie Jourdain Météo-France, Toulouse, France

Jorge Guzman, Univ. del Pacífico, Santiago de Chile, Chile

Rob Allan ACRE Program Manager, MetOffice Hadley Centre, Exeter, UK

Dick Dee

ECMWF, Reading, UK

Aside from improvements in the assimilation and numerical modeling schemes, new reanalyses can significantly profit from the recovery, imaging and digitization of historical observations. This is especially true for data-sparse regions and times. Here, we present the status and selected examples of the digitization of historical surface, aerological and radiation measurements in the framework of ERA-CLIM, an EU FP7 project designed to prepare the data necessary for a new reanalysis going back to the beginning of the 20th century. Project goals of ERA-CLIM are, amongst others 1) the improvement of the available observational record for the early 20th century, 2) the preparation of datasets needed for a planned global reanalysis, and 3) the development of an Observation Feedback Archive facility for users that will enable better quality characterization and possibly error correction of the primary observational data sources. Several project partners take part in the digitisation: Univ. of Bern (UBERN, lead), the Russian Inst. for Hydrological and Meteorological Information (RIHMI), Météo-France, Univ. of Lisbon, the international Atmospheric Circulation Reconstructions over the Earth (ACRE) initiative via the MetOffice Hadley Centre and Univ. del Pacífico (Chile). Manual processing (keying) as well as software-based optical character recognition (OCR) have been used for digitisation. Because data rescue is a work intensive activity, a significant fraction of the whole project budget has been devoted to this work package.

The data rescue activities of ERA-CLIM specifically focus on the data-sparse regions, such as the Tropics, the polar regions and the world's oceans before 1957. From the time before 1957, large amounts of surface data from former colonies and from overseas territories of European countries (e.g. Portugal and France) are being digitized. These surface data make up ca. 55% of the estimated total station days that have been inventorized. Another 45% of the inventorized data consist of upperair (aerological) observations. A relatively tiny fraction (< 1%) are atmospheric transmission measurements from 13 stations worldwide (1902-50). In case of the very early upper-air observations before the 1930s, even Europe and North America still hold an important quantity of data to be recovered in digital form. E.g., although monthly means of US kite data from the 1920s to the 1940s had already been digitized recently, the single ascent data has only now been added. Major sources of ascent data for Europe and North Africa, from the German Daily Weather Report, published by Deutsche Seewarte, the KNMI aerological reports, the Finnish Meteorological Yearbooks, and the daily climate reports (1919-1939) from Météo-France, have been found to hold much more pre-WW II data than what has previously been available in digital format. Other early global sources that have been imaged are some publications by the Commission for the Exploration of the Free Atmosphere (1900s to 1920s, "International Days"). Furthermore, many expeditions to remote places like Greenland, Antarctica or tropical regions in the first half of the 20th century collected aerological observations, and many were even organized for this very purpose.

An administrative web interface as well as a fully searchable metadatabase have been developed at UBERN to facilitate image, data and digitisation job handling and exchange. The inventorization of all identified data sources and their digital imaging have almost been completed. The digitization is still ongoing and will partly be continued for some time. However, due to the vast amount of records identified and the large amount of data, especially in the case of upper-air (1638 inventory entries, 1.3 mio. station days) and surface data (1.55 mio. station days), a prioritization in agreement with the goals and deliverables of the project turned out to be necessary. At UBERN, the largest sources of upper-air data found were the NOAA Central Library Foreign Data section (27%, data from many countries) and Indian upper-air and weather bulletins (27%), followed by many smaller sources, each making up less than 5% of the total amount of data. For the whole WP1, 16% of the upper-air data inventory entries stem from aircraft, registering balloon, kite or captive balloon measurements, 25% from radiosondes, and the largest part, 60%, are pilot balloon wind observations.

In case of the surface data, additional records (to the ones mentioned in the project proposal) have been identified in East Asia and around the Philippine Sea. Data from two additional atmospheric transmission stations were found, one in Algeria, one in the SW US. For aerological data, many more pre-1940 European station records have been imaged from the above-mentioned sources. Additionally, WW II data from Korea and China and 1930s data from several Chinese stations have been rescued. Also, more data from the French West Indies. New Caledonia, Polynesia, Antarctica and from the Southern Ocean (1940s and 50s), early US kite data, and data from French weather ships have been recovered. Finally, additional large aerological sources for the oceans and polar regions (German merchant and naval ships (eg. S.M.S. Planet 1906-07) and "Meteor" Expedition (1920s), Norwegian "Maud" Expedition (1918-25) and US Operation "Highjump" (Antarctic, 1940s)) were found. Quality control and reformatting of the data have started and will soon be finished so that the first version of the data can be delivered on time in June.

Corresponding Author:

Name:	Alexander Stickler
Organization:	Oeschger Centre for Climate Change Research and Inst. of Geography
Address:	Univ. Bern
	Inst. of Geography
	Hallerstrasse 12
	CH-3012 Bern
	Switzerland