Digitisation of Historical Observations for ERA-CLIM

Alexander Stickler[1,2], Stefan Brönnimann[1,2], Maria Antónia Valente[3], Alexander Sterin[4], Sylvie Jourdain[5], Jorge Guzman[6], Rob Allan[7], Dick Dee[8]

[1]Oeschger Centre for Climate Change Research, Bern, Switzerland [2]Inst. of Geography, Univ. Bern, Switzerland [3]Instituto Dom Luiz, Univ. Lisbon, Portugal [4]RIHMI-WDC, Obninsk, Russia
[5]Météo-France, Toulouse, France [6]Univ. del Pacífico, Santiago de Chile, Chile [7]ACRE Program Manager, MetOffice Hadley Centre, Exeter, UK [8]ECMWF, Reading, UK
(alexander.stickler@giub.unibe.ch, Phone: +41 31 631 88 43)

Introduction

Aside from improvements in the assimilation and numerical modelling schemes, new reanalyses can significantly profit from the recovery, imaging and digitisation of historical observations (see also Poster Z215). Here, we present the status and selected examples of the digitisation of historical surface, aerological and radiation measurements in the framework of ERA-CLIM (www.era-clim.eu), an EU FP7 project designed to prepare the data necessary for a new reanalysis going back to the beginning of the 20th century. One peculiarity of this reanalysis is the fact that it will, for the first time, assimilate upper-air data from the time before 1948.

Digitisation and Metadatabase

After digital high resolution imaging (Fig. 4), digitisation was done either by Optical Character Recognition (OCR, if possible) or by keying. A metadatabase holding the station inventories as well as additional metadata describing the sources, observational methods etc. has been created and will be made publicly available.

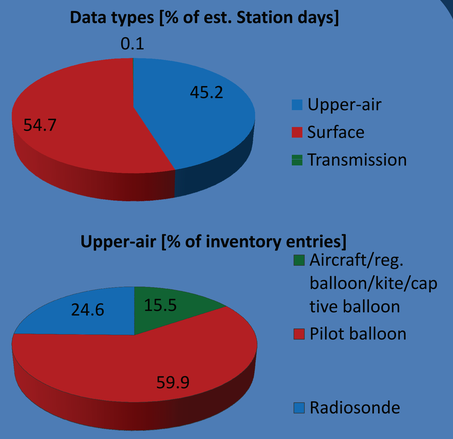


Fig. 3: Upper panel: Distribution of the estimated number of station days inventorised over the different data types. Lower panel: Distribution of the inventorised upper-air data over different measurement platforms.

Fig. 4: KAISER photo equipment, consisting of a table with tripod and a lighting with high frequency fluorescent tubes used for digital high resolution imaging.

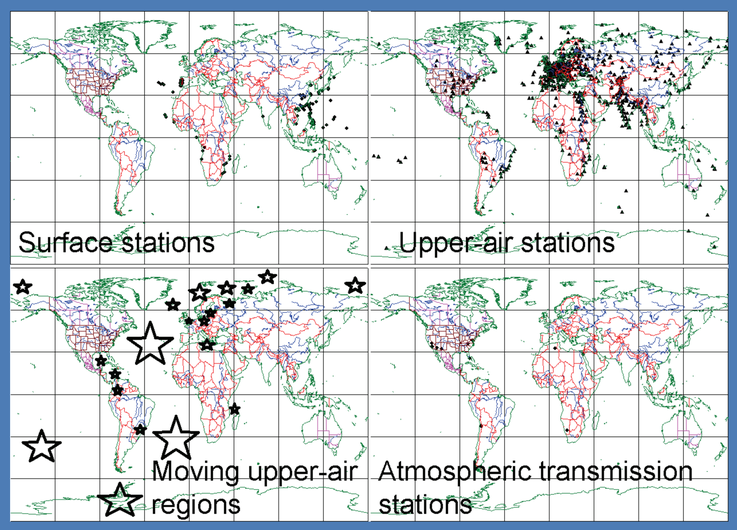


Fig. 1: Maps showing the global distribution of the measurement stations that have been inventorised and are being digitised in the framework of ERA-CLIM. The different observational platforms are separated in the different panels for clarity. For moving upper-air, stars denote regions for which data sources have been inventorised.

Inventory	# entries	# unique station names	% entries imaged	% entries digitised or in process of digitisation	# station days	% station days digitised or in process of digitisation
Upper-air	1,638	909	81	37	1,266,623	43
Surface	121	110	100	7	1,541,586	8
Moving upper-air	87	-	100	53	8,870	58
Transmis-sion	17	13	100	100	3,729	100

Table 1: Key statistics and status of the ERA-CLIM inventory.

Project Status

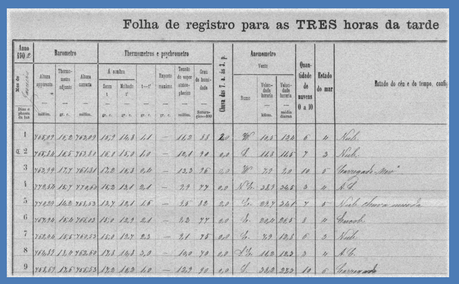
The inventory of all identified data sources and their digital imaging have almost been completed. The digitisation is still ongoing and will partly continue for some time. Some key statistics are presented in Table 1. However, due to the vast amount of records identified and the large amount of data, especially in the case of upper-air and surface data, a prioritisation in agreement with the goals and deliverables of the project turned out to be necessary. The largest single 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. These sources include e.g. the German Daily Weather Report, the KNMI aerological reports, the Finnish Meteorological Yearbooks, the daily climate reports (1919-1939) from Météo-France, and reports from polar and oceanic expeditions. For the whole project, 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. (Fig. 3 lower panel) 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. The upper-air data will be included in CHUAN (Stickler et al., 2010).

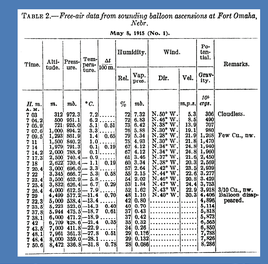
Conclusions

The digitised data are expected to be very valuable for the new ERA-CLIM reanalysis as well as for other reanalysis and research projects. The prolongation of the currently available observational series into the past is of crucial importance for our understanding of the climate system. All data digitised in the framework of ERA-CLIM will be made freely available, acknowledging a more than ever publicly demanded open access policy. If you are interested in the project visit www.era-clim.eu. If you want to contribute to the important task of digitising old weather data, please have a look at Poster Z215 and www.data-rescue-at-home.org.

Data Rescue Activities

The data rescue activities of ERA-CLIM specifically try to focus on the data-sparse regions, such as the Tropics, the polar regions and the world's oceans before 1957 (Fig. 1). 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, Fig. 2 upper panel) are being digitised. These surface data make up ca. 55% of the estimated total station days that have been inventorised. Another 45% of the inventorised data consist of upper-air (aerological) observations (Fig. 2 middle and lower panel). A relatively tiny fraction (< 1%) are atmospheric transmission measurements from 13 stations worldwide (1902-50) (Fig. 3 upper panel). 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. Fig. 2, middle panel).





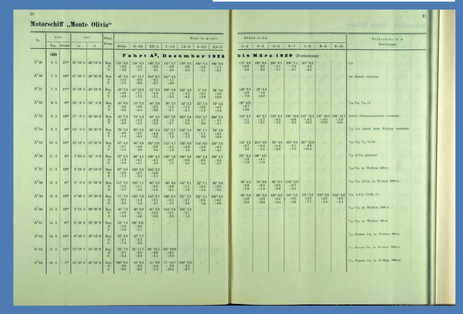


Fig. 2: *Upper panel: Surface station data from Angra (Azores) recorded in 1902, one of the oldest Portuguese stations in operation. Middle panel: US kite data reported in the "Monthly Weather Review Supplements" 1915-18. Lower panel: Pilot balloon data from the German Merchant Marine 1928-31, here: MS "Monte Olivia" on the Hamburg-South America Line.*

References

Stickler A et al. (2010): The Comprehensive Historical Upper-Air Network. *Bull. Amer. Meteor. Soc.* 91: 741-751. DOI: 10.1175/2009BAMS2852.1.