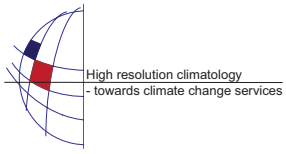
EMS Annual Meeting Abstracts
Vol. 7, EMS2010-554, 2010
10th EMS / 8th ECAC



Aspects of quality insurance in digitizing historical climate data in Germany

H. Mächel (1), J. Behrends (1), and A. Kapala (2)

(1) Deutscher Wetterdienst, Klima und Umwelt, Offenbach, Germany (hermann.maechel@dwd.de), (2) Meteorological Institute, University of Bonn, Bonn, Germany

This contribution presents some of the problems and offers solutions regarding the digitization of historical meteorological data, and explains the need for verification and quality control. For the assessment of changes in climate extremes, long-term and complete observational records with a high temporal resolution are needed. However, in most countries, including Germany, such climate data are rare. Therefore, in 2005, the German Weather Service launched a project to inventory and digitize historical daily climatic records in cooperation with the Meteorological Institute of the University of Bonn.

Experience with Optical Character Recognition (OCR) show that it is only of very limited use, as even printed tables (e.g. yearbooks) are not sufficiently recognized (10-20% error). In hand-written records, the recognition rate is about 50%. By comparing daily and monthly values, it is possible to auto-detect errors, but they can not be automatically corrected, since there is often more than one error per month. These erroneous data must then be controlled manually on an individual basis, which is significantly more error-prone than direct manual input.

Therefore, both precipitation and climate station data are digitized manually. The time required to digitize one year of precipitation data (including the recording of daily precipitation amount and type, snow amount and type, and weather events such as thunder storms, fog, etc.) is equivalent to about five hours for one year of data. This involves manually typing, reformatting and quality control of the digitized data, as well as creating a digital photograph. For climate stations with three observations per day, the working time is 30-50 hours for one year of data, depending on the number of parameters and the condition of the documents.

Several other problems occur when creating the digital records from historical observational data, some of which are listed below.

Older records often used varying units and different conventions. For example, a value of 100 was added to the observed temperatures to avoid negative values. Furthermore, because standardization of the observations was very low when measurements began up to 200 years ago, the data often reflect a greater part of non-climatic influences. Varying daily observation times make it difficult to calculate a representative daily value. Even unconventional completed tables cost labor and requires experienced and trained staff.

Data homogenization as well as both manual and automatic quality control may address some of these problems.