EMS Annual Meeting Abstracts
Vol. 10, EMS2013-160, 2013
13th EMS / 11th ECAM

Preparing Historical UpperAir Data Sets Based on Hardcopy Documents: RIHMI Experience

A. Sterin and D Nikolaev
RIHMI-WDC, Obninsk, Russian Federation (sterin@meteo.ru)

Digitizing hardcopy data, depending hardcopy document features and specific, requires case-by-case decisions on choice of digitizing manner (manual or OCR), on variable transformations, data management operations, quality check and, finally, on output formats. All decisions taken together, create technologies, though separate elements of these technologies may be of different origin and may be just different instruments.
The paper describes solutions on digitizing hardcopy data at RIHMI-WDC (Obninsk, Russia). The elements of technologies include use of OCR (for printed documents, if OCR is acceptable), or use of manual filling MS ACCESS tables, if OCR is not acceptable. For OCR technologies, collecting data into MS EXCEL and their primary check based on conditional formatting option of MS EXCEL is used.
Further steps include import of EXCEL or ACCESS tables into SAS System software, numerous data transformation and processing. For further quality check, calculation of statistics and data visualization are applied. Each suspected value discovered at this stage, is re-checked from the primary hardcopy documents including corrigendum lists of each publication. Again, the processing is done after correcting the values.
Finally, the output into formats most acceptable for ERA CLIM is provided. The paper contains the detailed description of each step of the technologies. Very simple solutions such as those based on OCR (ABBYY FineReader), MS EXCEL and MS ACCESS, are shown to be effective as elements of constructed technologies.