



The high-quality UERRA rescued dataset; QC procedures and results

Joan Ramon Coll (1), Linden Ashcroft (1), Alba Gilabert (1), Enric Aguilar (1), Javier Sigró (1), Phillip Jones (2), Per Uden (3), Manola Brunet (1,2)

(1) Centre for Climate Change, Universitat Rovira i Virgili, Tarragona, Spain, (2) Climatic Research Unit, School of Environmental Sciences, University of East Anglia, Norwich, United Kingdom, (3) Sveriges Meteorologiska och Hydrologiska Institut (SMHI), Sweden

In the framework of the EU-FP7-funded collaborative project entitled Uncertainties in Ensembles of Regional Reanalyses (UERRA, 2014-2017), about 9M station synoptic values were rescued over European data-sparse sub-regions (the Mediterranean, Central Europe and the Balkan regions) for the main climate variables and mostly for the post-1950 period. The newly UERRA dataset derived from data rescue (DARE) effort will be used to produce ensembles of regional reanalysis for several decades and estimate the associated uncertainties.

Data quality control (QC) procedures were required to detect and correct/remove non-systematic errors derived from original sources, digitisation process, data manipulation or transmission that could compromise the final quality of the dataset. For this purpose, a battery of automated quality control (AQC) routines were specifically designed to assess temporal and spatial data consistency from air pressure, air temperature, dew-point, relative humidity, wind speed, wind direction and precipitation (snow-depth and fresh snow) at the sub-daily (daily) scale. After a preliminary visual data inspection to detect and correct skewed values introduced from an erroneous calendar day, AQC tests were automatically applied for the whole UERRA dataset. The results revealed that 9.5% of sub-daily data were flagged as suspicious. Flagged observations were manually cross-checked by using the original sources to ensure the recovery of all possible data. Applying this methodology, 5.2% of flagged data were corrected, 2.2% were validated and 2.1% were set to missing as a new value could not be offered. Once temporal data consistency was assessed, spatial homogeneity checking was also carried out to ensure spatial data consistency among stations where possible. The methodology developed by Dunn et al., 2012 was adapted to the UERRA dataset for this purpose. Intra- and inter-station consistency were tested, flagging 1.7% of the sub-daily data as suspicious. Flagged observations were finally set to missing to avoid that climate time-series are negatively affected by erroneous values.

The application of AQC tests and spatial consistency checking ensured temporal and spatial data coherency to the new high-quality UERRA rescued dataset. The incorporation of about 9M of high-quality station synoptic values to European regional reanalysis will surely help to produce accurate ensembles and estimate the associated uncertainties.

Dunn RJH, Willett KM, Thorne PW, Woolley EV, Durre I, Dai A, Parker DE, Vose RS. 2012. HadISD: a quality-controlled global synoptic report database for selected variables at long-term stations from 1973–2011. *Clim. Past* 8: 1649–1679.