Earth Syst. Sci. Data Discuss., https://doi.org/10.5194/essd-2018-39
Manuscript under review for journal Earth Syst. Sci. Data
Discussion started: 3 May 2018

A rescued dataset of sub-daily meteorological observations for Europe and the southern Mediterranean region, 1877–2012

Linden Ashcroft[1,2], Joan Ramon Coll[1], Alba Gilabert[1], Peter Domonkos[1], Enric Aguilar[1], Javier Sigro[1], Mercè Castella[1], Per Unden[3], Ian Harris[4], Phil Jones[4,5] and Manola Brunet[1,4]

[1]Centre for Climate Change, Department of Geography, Universitat Rovira i Virgili, Tarragona, Spain
[2]Australian Bureau of Meteorology, Melbourne, Australia
[3]Swedish Meteorological and Hydrological Institute, Folkborgsvägen, Norrköping, Sweden
[4]Climate Research Unit, School of Environmental Sciences, University of East Anglia, Norwich, United Kingdom
[5]Center of Excellence for Climate Change Research, Department of Meteorology, King Abdulaziz University, Jeddah, Saudi Arabia

Correspondence to: Linden Ashcroft (linden.ashcroft@bom.gov.au)

Abstract. Sub-daily meteorological observations are needed for input to and assessment of high-resolution reanalysis products to improve understanding of weather and climate variability. While there are millions such weather observations that have been collected by various organizations, many are yet to be transcribed into a useable format.

Under the auspices of the European Union funded Uncertainties in Ensembles of Regional ReAnalysis (UERRA) project, we describe the compilation and development of a digital dataset of 8.8 million meteorological observations rescued across the European and southern Mediterranean region, many of them Essential Climate Variables (ECVs) as defined by the Global Climate Observing System (GCOS). By presenting the entire chain of data preparation, from the identification of regions lacking in digitized sub-daily data and the locating of original sources, through the digitization of the observations to the quality control procedures applied, we provide a rescued dataset that is as traceable as possible for use by the research community.

Data from 127 stations and of 15 climate variables in the northern Africa and European sectors have been prepared for the period 1877 to 2012. Quality control of the data using a two-step semi-automatic statistical approach identified 3.5 % of observations that required correction or removal, on par with previous data rescue efforts.

In addition to providing a new sub-daily meteorological dataset for the research community, our experience in the development of this UERRA sub-daily dataset gives us an opportunity to share guidance on future data rescue projects.

All data are available on PANGAEA: https://doi.pangaea.de/10.1594/PANGAEA.886511.

Keywords: meteorological data, Europe, Mediterranean, essential climate variables, data rescue

Earth Syst. Sci. Data Discuss., https://doi.org/10.5194/essd-2018-39
Manuscript under review for journal Earth Syst. Sci. Data
Discussion started: 3 May 2018

1. Introduction

Digitizing meteorological observations into a useable modern format is crucial for long-term climate monitoring and assessment. High-quality, long-term observations are needed for almost all aspects of meteorological and climatological research, but many spatial and temporal gaps still exist in data products currently used by the international research community (Brunet and Jones, 2011). For this reason, meteorological data rescue and recovery is becoming increasingly important, particularly in developing countries and for the early instrumental period, as data are often only available in paper format and are at great risk of being permanently lost (Brunet and Jones, 2011; Page et al., 2004; World Meteorological Organization, 2016).

In the last 20 years, many initiatives have been established to recover and digitize land-based meteorological observations at national, regional, and international scales. The Atmospheric Circulation Reconstructions over the Earth initiative (ACRE, Allan *et al.*, 2011) coordinates data rescue across the globe, while other projects such as MEditerranean DAta REscue (MEDARE, www.omm.urv.cat/MEDARE/index.html) and Historical Instrumental Climatological Surface Time Series Of The Greater Alpine Region (HISTALP, www.zamg.ac.at/histalp/) focus on particular regions (Auer et al., 2007; Brunet et al., 2014a, 2014b). Additional initiatives on a national to regional scale, led by meteorological agencies (e.g. Kaspar *et al.*, 2015 in Germany) and research projects (e.g. Ashcroft *et al.*, 2014; Brunet *et al.*, 2006, 2014a) have located and digitized historic observations, and ensured that they are made available to the scientific community.

Many of these projects have focused on the rescue of daily, monthly and or annually-averaged data, as these observations form the basis of long-term climate analysis. Daily maximum temperature, minimum temperature and precipitation totals are often the top priority for digitization, because these variables are used to monitor changes in climate and the incidences of extreme weather events, as well as to identify economic and agriculturally important long-term variations in precipitation (Brunet et al., 2006; Moberg et al., 2006). The development of the 20th Century Reanalysis product — which uses only sub-daily atmospheric pressure observations as input for a global reanalysis — has also benefited from national and regional data rescue activities, resulting in an increase in atmospheric pressure data recovery in recent years (Compo et al., 2011; Cram et al., 2015). Far fewer recovery efforts have been made to uncover sub-daily meteorological observations of other variables, despite the fact that they are the necessary input to global and regional reanalysis products, the output of which can greatly improve the understanding of atmospheric circulation and of high-temporal resolution extreme events (e.g. Cannon *et al.*, 2015; Stickler *et al.*, 2014).

This paper presents the experience and resultant dataset of a two-year digitization effort aimed at recovering sub-daily meteorological data. Our work formed part of Uncertainties in Ensembles of Regional ReAnalysis (UERRA, http://uerra.eu/), a project under the European Union 7th Framework Programme. The goal of UERRA was to produce ensembles of European regional reanalyses at high temporal resolution for several decades, with an estimate of the associated uncertainties in the resulting datasets. A key component of UERRA is the recovery of sub-daily surface meteorological observations to provide input to and assess the quality of the various reanalysis products.

In this paper we describe our complete data rescue process to provide sufficient details, as much as possible, for a fully traceable dataset. In Sect. 2 we explain how we identified target regions for data rescue across Europe and the neighboring southern Mediterranean region to maximise improvement in spatial and temporal coverage of existing data, as well as potential sources of sub-daily data for digitization. We present the methods employed to minimise errors in the digitization process and the steps required to take the data from a disparate set of sources to a unified database.

Earth Syst. Sci. Data Discuss., https://doi.org/10.5194/essd-2018-39
Manuscript under review for journal Earth Syst. Sci. Data
Discussion started: 3 May 2018

In Sect. 3 we provide details on the quality assurance and control procedures used to reduce errors in the dataset, including visual checks, semi-automatic statistical methods and an automatic spatial comparison method. We present the dataset and quality control results in Sect. 4. Finally, we give details about how to access the data, as well as some practical ideas for future data recovery projects, based on our experiences with this particular project.

2. Methods and materials

2.1. Identifying gaps in sub-daily data availability

The primary goal of the data rescue efforts within UERRA has been to improve spatial and temporal coverage of input data for future regional gridded and reanalysis climate products over the European domain. Adding new observations to the available network of datasets from which these products are derived will help to reduce uncertainties, ultimately helping to improve the understanding of European weather and climate. This involves, as a first step, identifying the basic station data used in current reanalysis products available at the European Centre for Medium Weather Forecasts (ECMWF) and other relevant databases that contain digitized observations.

To identify gaps in the available sub-daily climate record, we first conducted an extensive examination of the Meteorological Archival and Retrieval System (MARS) of the ECMWF. MARS is home to the primary data input of the current European reanalysis products available from ECMWF, and so stations that are identified in data sources (see Sect. 2.2) but not present in MARS, or stations with low percentages of sub-daily data are likely candidates for data recovery. We divided our search into pre-1950 and post-1950 data availability, to align with the temporal focus of the proposed UERRA regional reanalysis products and ECMWF historical reanalyses such as ERA-20C (https://www.ecmwf.int/en/research/climate-reanalysis/era-20c). The variables of interest were several Essential Climate Variables (ECVs) as defined by the Global Climate Observing System (GCOS) as well as two other variables, deemed useful for reanalysis by the UERRA research team. The ECVs we focussed on were air temperature (TT), atmospheric pressure (sea level pressure, PP and station level pressure, SP), wind speed (WS) and wind direction (WD), relative humidity (RH), dew point temperature (DP) and daily rainfall (RR). The non-ECVs were snow depth (SD) and fresh snowfall (FS).

Next, we examined other sub-daily data repositories where rescued observations are likely to be stored to further minimise potential duplication of data digitization efforts. We cross-referenced our MARS results with the station list of the International Surface Pressure Databank (ISPD, Cram *et al.*, 2015), the Koninklijk Nederlands Meteorologisch Instituut (KNMI) European Climate Assessment and Dataset (ECA&D: http://eca.knmi.nl/), and the national climate data systems of countries whose data may not yet be in a multi-national repository. In particular, we examined the data available from the National Climate Data Management Systems (CDMS) of the Romanian Meteorological Administration (NMA-RO) and the National Meteorological and Hydrological Services (NMHS) of countries in the Western Balkans, including Albania; Bosnia & Herzegovina; the Republic of Macedonia; Montenegro; and the Republic of Serbia.

With this data availability information, we were able to identify the Mediterranean, Eastern Europe and Scandinavia as three key sub-regions within the European sector where MARS and other data repositories were lacking in sub-daily data for reanalysis development, particularly in the post-1950 period (Fig. 1, Table S1).

The high percentage of stations with data for less than 20 % of the 1950–2010 period (Fig. 1) illustrates the lack of sub-daily observations in these sectors. Gaps are clear in the southern and eastern Mediterranean countries, Sweden and

Earth Syst. Sci. Data Discuss., https://doi.org/10.5194/essd-2018-39
Manuscript under review for journal Earth Syst. Sci. Data
Discussion started: 3 May 2018

Norway for the 1960s and 1970s (Table S1), as well as across the Balkan region. The relatively dense spatial coverage of the stations also suggests there is a high likelihood that observations have been taken at many places in these regions, but have not yet been made available in a standardised format.

2.2. Locating and assessing scans of sub-daily data sources

As well as identifying gaps in the digitized sub-daily record available for Europe, we also needed to locate sources of undigitized sub-daily data. We undertook extensive consultation with NMHS across the three identified regions of poor data coverage, in an attempt to identify and recover paper or scanned data sources suitable for digitization. Priorities were given to data sources already available as scanned images, and stations with variables identified as important for the development and verification of regional reanalyses (see Sect. 2.1). Recovered precipitation observations from NMA-RO were digitized internally, and then provided to us in digitized quality controlled format, using a similar quality control format to ours (see Sect. 3.2). Discussion with the Norway and Swedish NMHS uncovered data for these countries that had been digitized, but were not yet provided to international data repositories. Similarly, the Catalan Meteorological Service (MeteoCat), which has an open data policy, offered us their digitized data for the recent period 1998–2015 to be transferred to relevant global repositories through our effort. Data sharing was organised between these regions and ECMWF without the need for observations to be transcribed from paper format, and will therefore not be discussed in the current study. Political and financial difficulties prevented many countries we contacted, particularly in northern Africa and the Balkans regions, from providing original data sources to us for digitization.

Original data sources were provided in scanned format by Deutscher Wetterdienst (the German Meteorological Office, DWD), the Slovenian Environmental Agency (SEA), and Agencia Estatal de Meteorología (the Spanish Meteorological Service, AEMET), via MeteoCat. Close consultation with these NMHS enabled us to identify valuable and previously undigitized data sources. From these sources, stations with minimal data available in MARS were selected for digitization.

The World Meteorological Organization (WMO) MEDARE initiative and the precursor project to UERRA, the EU-funded EURO4M project (http://www.euro4m.eu/), located key records of data for the Middle Eastern, Balkan and southern Mediterranean regions from the Serbian NMHS online climatological scanned repository (http://www.hidmet.gov.rs/ciril/meteorologija/klimatologija_godisnjaci.php), the United States of America's National Oceanic and Atmospheric Administration/National Climatic Data Center (NOAA/NCDC) Climate Data Modernization Project (CDMP: http://library.noaa.gov/Collections/Digital-Documents/Foreign-Climate-Data-Home), the British Atmospheric Data Centre (BADC, http://badc.nerc.ac.uk/browse/badc/corral/images/metobs) and other national meteorological services (see Brunet *et al.*, 2014a, 2014b for details). Daily maximum and minimum temperature, precipitation, and sub-daily atmospheric air pressure observations from some of these sources were digitized under the auspices of EURO4M and MEDARE, but many other observations were unable to be transcribed due to project constraints. UERRA therefore provides a valuable opportunity to rescue the previously undigitized values from these sources (Brunet et al., 2014b).

Table 1 provides detail of the data sources identified for digitization, while Fig. 2 shows several examples of the data sources used. All of the variables included in each source are listed in Table 1, although not all were digitized under the auspices of UERRA. The majority of data sources from CDMP are secondary, meaning that they are collations or summaries of observations that have been prepared in a central location. Unfortunately, secondary data sources are more prone to transcription errors than original series, as they have been transferred from the original readings. Many were handwritten, although a small subset was typed.

2.3. Digitizing method

Once the data sources had been identified and catalogued, a group of 11 digitizers were employed 15 hours a week over a two-year period to digitize the data. The digitization team was made up of undergraduate and postgraduate geography students from the University Rovira i Virgili, who all had some knowledge of meteorological variables and European climate. The digitizers worked on desktop computers in a computer lab, with large screens and standard keyboards. They were also given the option of working from home on their personal laptops.

The digitizers received initial training sessions, as well as online instructions and monthly in-person meetings to discuss issues and introduce new digitization tasks. Digitization was done using a strict "key as you see" method, meaning that the digitizers typed the values that were provided in the data images, rather than using any coding system. This follows standard best practise outlined by the WMO (2016).

Budget constraints made it unfeasible to employ double-keying, a suggested method of improving digitized data quality (Brönnimann et al., 2006). We tested optical character recognition (OCR) and speech recognition technologies, but the diverse nature of each task and the time and cost associated with training the software for each data source made these options unfeasible. However, the digitizers were trained in self-assessment techniques aimed at reducing data errors. Digitizers were asked to carefully cross check their values with the original source values for the 10th, 20th and 30th day of each month to make sure that no days had been skipped or repeated. Days with missing data were recorded in metadata files, along with any other variations in the data source. Where data sources included monthly totals and summaries, digitizers were instructed to calculate these values from their daily transcribed data, to check accuracy.

The data sources were in a number of different formats (see Fig. 2). The two main formats were one month (or day) to a page for a single station, and one day to a page for a network of stations. Depending on the source structure, each digitizer was in charge of digitizing values from a station (e.g. Egyptian and Moroccan sources, Fig. 2a and b), a time period (e.g. Slovenia, Fig. 2c), or a variable (e.g. Lebanon, Fig. 2d).

In several cases, not all of the data on a sheet were required to be digitized, as they had already been transcribed as part of EURO4M and MEDARE. To help digitizers with the complex layout of the source images, templates were developed in Microsoft Excel for some sources that were as close as possible to the format of the original data source (e.g. Fig. 3). Borders and shading within the files were used to help the digitizer keep track of their work, and date columns were pre-filled with the correct dates to reduce the occurrence of errors associated with leap years. The development of templates was not always possible due to time constraints, although it was employed for all sources with hourly data (see Table 1).

The digitizers were required to upload their data to a central server every 15 days, include a count of the number of values digitized, and a copy of the data transcribed so far. This method ensured that the digitizers were making progress, the data were being regularly backed up, and that the digitized observations could be regularly checked (see Sect. 3).

2.4. Conversion to standard units

While all quality control and assessment was applied to the data in their original units, the data were also converted to standard units, to be used in widespread meteorological products and statistical quality control procedures (Table 2). Data sources and available metadata were examined closely to ensure the conversions were as accurate as possible, and any changes to units within the same source were captured. Many atmospheric pressure observations needed to be

Earth Syst. Sci. Data Discuss., https://doi.org/10.5194/essd-2018-39
Manuscript under review for journal Earth Syst. Sci. Data
Discussion started: 3 May 2018

converted from mm of mercury to hectopascals, and station level pressure data reduced to sea level pressure for quality control testing. This step involved a detailed examination of the data sources to identify station height information and any instrument movements that may have occurred.

3. Quality assessment of digitized data

Quality control (QC) procedures are crucial to identify non-systematic errors that could be hidden in time series. These errors can occur as a result of issues with original sources, the method of data collection, transcription, or the digitization process. Ensuring that data are digitized with the utmost consideration for data reliability, and applying a QC procedure to the digitized observations are essential steps in the preparation and analysis of climate data. This is particularly the case for daily or sub-daily data, as these observations are used in the calculation of monthly and annual means.

An ideal QC procedure must be transparent and rigorous to ensure internal data consistency, temporal and spatial coherence, and traceability for future data users. A well-defined and executed QC routine will be able to flag data errors from time series that could compromise the analysis of natural climate variability and anthropogenic climate change, including the study of extreme variables. This is the key to avoid incorrect climate interpretations induced by data errors in a climate change context (Aguilar et al., 2003; Brunet et al., 2006)

An exhaustive QC application was vital for our study, but given the large number of observations, completely manual QC was not a feasible method of correcting the data. However, a completely automated procedure, such as that used for global databases (Dunn et al., 2012) would also be sub-optimal, as the digitized data do not cover a wide geographic area and consistent time period. We therefore decided that a multiple-step process would be the best approach.

Figure 4 outlines the multiple steps of the data quality assurance and control procedures used in the development of the dataset. As outlined in Sect. 2, efforts were made before digitization to minimise the introduction of errors, including a detailed assessment of each data source, the development of templates for many sources, and the selection of qualified digitizers. During and after digitization, the digitized data were then subjected to quality control and assurance testing. The structure of the testing (Fig. 4) can be summarised as a basic visual check, statistical testing at the individual station level, and spatial testing across comparable networks.

Note that homogenization is not included in this procedure. Although the homogenization of data to remove non-climatic features of a long-term instrumental record is crucial for the assessment of climate variability and change (e.g. Peterson *et al.*, 1998), homogeneity assessment of sub-daily data is a highly complex task that is still in development within the research community (Venema et al., 2012).

3.1. Visual cross-checking

Values uploaded by digitizers were systematically compared to the original source images by climatologists familiar with the sources, and occasionally other digitizers. The aims of these initial visual crosschecks was to provide timely feedback to the digitizers if common errors were occurring, identify subtle errors in the order of the data that may not be picked up in statistical procedures, and also make a preliminary assessment of the quality of the data from each particular source (Table 1). Additionally, regular reporting of data completed helped us identify any digitizers who were having trouble with their tasks and needed extra assistance.

For every fourth year of data, two or three days of observations were selected at three monthly intervals for visual cross checking with the original source. Additional ad hoc checks were made if a known issue existed in the data source e.g. if the period covered by the data source contained a leap year, or the source pages were known to be out of order. Although these checks only covered a small percentage of the total digitized data, we felt it was sufficient to identify the general quality of work done by individual digitizers and for each source.

In more than 60 % of stations tested, only a small number (less than 5 %) of the checked values required correction. Visual cross checking of data from stations with a larger number of errors identified the occasional skipped day or duplicated value, which meant that a large percentage of observations needed to be shifted by one time step. The majority of these errors were found in data for Egypt and Algeria, from sources that had already been flagged as difficult to read and containing date order errors. In two cases, digitizers were asked to repeat their work.

3.2. Individual station quality control (SAQC method)

After the basic visual quality checks, the digitized data were subjected to a range of statistical quality control tests. Due to the highly variable nature of the different data sources, and their disparate geographical spread, data from each station were examined individually in this step. Statistical quality control was conducted using a semi-automatic quality control (SAQC) procedure developed by the Centre for Climate Change at URV (http://www.c3.urv.cat/softdata.php). SAQC comprised of three separate programs that can be applied to data in text file format: one examining temperature, wind, relative humidity and dewpoint observations; another assessing sea level pressure data, and a final check on sub-daily rainfall data, daily snow depth and snow fall. The tests were largely adapted from existing automatic quality control procedures developed for sub-daily data at a global scale (e.g. Dunn *et al.*, 2012; Durre *et al.*, 2010), but were adapted for the UERRA dataset to enable more manual examination of the resultant flags. The tests applied within SAQC (Table 3) can be largely grouped into four groups depending on the degree of QC applied (Aguilar et al., 2003):

- *Gross errors tests*: QC tests that detect and flag obviously erroneous values (date order check, date errors, unrealistic values, data repetitions and non-numeric value tests).
- *Tolerance tests*: QC tests that detect and flag those values considered outliers with respect to their own-defined upper/lower limits (climatic outliers, bivariate comparisons, monthly mean of absolute increments, and unusual distribution of values tests).
- *Internal consistency check*: QC tests which detect and flag incoherencies between associated elements within each record (Interval and DP/FS/SD inconsistency test, RH/DP/TT comparison tests, precipitation and snow totals test)
- *Temporal coherency*: QC tests which detect and flag a given value that is not consistent with the amount of change that might be expected in a variable in any time interval according to adjacent values (Flat line test, big jump test, summer snow test and irregular temporal evolution).

Each program was applied at a country level, producing a list of values flagged by each test at each station. The results of each test were then manually cross referenced against the original source data, and corrected or removed by a trained climatologist. The removal or correction of each value was recorded using a flag system, to clearly document the nature of the identified errors and results (Table 4). An example of the air temperature evolution in Port Said (Egypt) taken at 0800 and 1400 for the short period 1939–1940 and resultant QC flags is shown in Fig. 5, highlighting various types of errors, outliers and extreme values over a short time period.

In the initial testing of the SAQC procedure, the tests for duplicate values, monthly mean of absolute increments and unusual distribution of values tests were found to be overly sensitive, resulting in many valid observations being

Earth Syst. Sci. Data Discuss., https://doi.org/10.5194/essd-2018-39
Manuscript under review for journal Earth Syst. Sci. Data
Discussion started: 3 May 2018

flagged for assessment. Many of the legitimate errors identified by these tests were also found by others, so the thresholds on these tests were relaxed to make the task of checking flagged values more manageable.

3.3. Spatial quality assurance (HQC method)

The final QC procedure consisted of subjecting data from neighboring stations to spatial quality control tests. Only data that had been checked by visual and automatic QC were subjected to this procedure. The spatial QC process was conducted using an adapted version of the procedure used in the development of the U.K. Met Office Hadley Centre Global Sub-Daily Station Observations dataset (HadISD v2.0.1.2016p; Dunn et al., 2012, 2016). Adaptation of some tests (Table 5) was required as the UERRA dataset had low spatial resolution and included observations taken at inconsistent times, often converted from units with coarse resolution. Automatically running HQC in its standard form led to a large number of false positive flags being identified, automatically removing a significant number of correct observations being removed from the dataset (Fig. 6).

To reduce the number of false positive flags, the minimum number of neighboring stations required for HQC testing was reduced from ten to five, and the percentage of non-missing observations per month allowed was reduced from 75 % to 66 %. Tests that looked for streaks of identical values, or non-uniform distributions in the frequency of values were also slackened to account for the fact that many observations were converted from different units.

The 127 stations were then split into networks according to their correlation, spatial distance, observing times, overlapping observing periods and variables observed. Six appropriate networks were identified (Table 6), but unfortunately it was not possible to include all stations, periods, variables and observing times. The heterogeneous characteristics of the dataset, the high spatial distance and low spatial resolution of the stations and the inconsistent coverage of the variables included in the dataset meant that only about 4.3 million observations (over 48 % of the total dataset) could be subjected to HQC.

For example, it was not possible to apply HQC to data from Cyprus, Lebanon and Spain due to the low number of stations in each country and the large spatial distance from the neighboring country stations. We were also unable to analyse fresh snow and snow depth, precipitation or relative humidity data, as the HadISD QC does not assess these variables. Moreover, several stations (such as those in Germany and Slovenia, network 6 in Table 6) provided hourly data, but there were not enough neighboring stations with sufficiently high temporal resolution to allow for more than several observing times per day to be checked.

3.4. Final check

After the HQC was applied, a final check was made to ensure that the conversion procedures had been applied correctly and that all flags were realistic i.e. that a flag of 3 was associated with a value that had been removed from the final datasets.

4. Results

4.1. Spatial and temporal data distribution

A total of 8.8 million observations were digitized from 127 stations in 15 countries (Table 7 and Table S2). The majority of located sources provided sub-daily temperature observations, wind speed and wind direction, and atmospheric pressure. A small network of sources from Germany, Lebanon and Slovenia contained hourly data for a number of variables, contributing to the high number of observations for those countries. Additional sources from

Earth Syst. Sci. Data Discuss., https://doi.org/10.5194/essd-2018-39
Manuscript under review for journal Earth Syst. Sci. Data
Discussion started: 3 May 2018

Slovenia contained sub-daily rainfall data, while sources provided by the SEA and DWD included daily snowfall and snow depth data from Slovenia and Germany.

Long records (> 30 years) with many variables were successfully recovered from stations in Egypt, Tunisia and Algeria, although only the Egyptian stations provide observations more than once a day (Fig. 7). Shorter but more widespread observations were rescued across Morocco, Turkey and in the Balkans region, while the snowfall observations in Germany only covered the west of the country.

The largest number of observations (more than 28 %) came from Slovenia (Fig. 8a); even though we only had data for three stations in Slovenia, the observations were hourly, included nine variables, and covered more than 20 years. Around 15 % of the rescued observations came from Egypt, and almost 12 % from Turkey. Both of these countries have a large number of stations in the recovered network, and a variety of variables over a long period of time (Fig. 7).

More than 21 % (1.8 million) of the rescued observations were sub-daily temperature measurements, with wind speed and direction measurements totalling over 17 % (Fig. 8b). There were around 20,000 more wind direction observations than wind speed; this is because very early Tunisian and Egyptian wind speed observations were qualitative (e.g. light, moderate) and were not digitized. Relative humidity data made up around 16 % of the rescued dataset, while sea level pressure and station level pressure contributed a similar amount at just over 15 % (around 1.4 million values). Over 160,000 fresh snow and 160,000 snow depth values (more than 3.5 % combined) were also recovered from Germany and Slovenia from as far back as the 1950s, representing a significant increase in snow observations across the region.

Due to the temporal coverage of the Slovenian data (1950–1978), as well as the dedicated focus of the UERRA project on post-1950 observations, the mid-20th Century was the most well represented period in the rescued dataset (Fig. 8c). Almost 60 % of the dataset covered the 20 years from 1950 to 1969. Observations from Cyprus and northern Africa provided data from the late 19th century, and records from Serbia were recovered up to 2012.

Finally, the most common observing times for the variables rescued were 0700, 1400 and 2100, reflecting standard observing practises over the European region in the 20th century. Tunisian observations were only available for 0700, and for many other countries where observations were only available once a day in the early part of the record, these observations were inevitably in the morning also. Two German stations included a small number of half hourly observations (Fig. 8d).

4.2. Semi-automatic quality control results (SAQC)

All rescued sub-daily data were subjected to quality control routines to identify erroneous values or chains of values in the time series (Sect. 3). A total of 3.2 % of observations, around 268,000, were flagged as suspicious for the whole UERRA dataset using SAQC (Fig. 9).

Removing correct values that have been flagged (false positives) is a common QC issue, and manual examination ensured that these important observations – often of extreme events – are retained for future studies. The majority of the values flagged (1.5 % of the total number of values) were corrected after manual examination, with just over 1 % of the total number of observations deleted due to errors in the source image or issues with the readability of the original values. Over 27, 000, 0.3 % of the total number of observations, were flagged but then found to be correct after examination.

Despite being the country with the smallest number of observations, the largest percentages of flagged values found were for Bosnia and Herzegovina and the Czech Republic (~8 % of the total number of data digitized, Fig. 10a). For Bosnia and Herzegovina a large section of observations from one station needed to be set to missing due to digitizer error, and for the Czech Republic observations a digitization error was able to be corrected by shifting data by one day. The hand-written nature of the Czech data, together with the absence of data templates (only used in Slovenian, Spanish

Earth Syst. Sci. Data Discuss., https://doi.org/10.5194/essd-2018-39
Manuscript under review for journal Earth Syst. Sci. Data
Discussion started: 3 May 2018

and German data sources) may go some way to explaining the large number of flagged values among both countries. The countries with the largest number of observations (Egypt and Slovenia) had about 3 % of their observations corrected or verified, and less than 2 % removed under the SAQC procedure.

A similar amount of flagged values were proportionally found in all rescued observations distributed by variables, except for precipitation (RR, Fig.10b). The high number of precipitation flags is due to two factors. Firstly, several digitizers inadvertently recorded zero rainfall values as missing, or missing rainfall as zero. The format of the Slovenian data sources changed over the period, with some years having hourly rainfall data and others only providing observations three or four times a day. This issue can significantly skew any resulting analysis and so was corrected wherever it was identified.

Secondly, during the latter part of the Slovenian record, some daily rainfall totals were calculated inconsistently, using a midnight to midnight sum occasionally rather than a 0700–0700 total. The six-hourly observations were QCed based on these totals, but the daily rainfall totals calculated in this way were removed from the final version of the dataset to ensure consistency.

SAQC flags distributed by decade show a similar pattern to the distribution of observations, with a peak in the mid-20th century (Fig.10c). The higher number of fl17 flags (observations set to missing as no value could be found in the source image) during the 1940s may reflect data issues during the Second World War, particularly for Egypt and Algeria, where there were issues associated with the ordering of the original source files. Flagged values were relatively evenly distributed across observation times (Fig.10d), although the lower absolute numbers of half hourly observations made for a higher proportion of flagged observations proportionally found in all rescued observations distributed by observation times.

4.3. Spatial quality control results (HQC) Quality control results

Temperature was the variable with the smallest number of flagged values overall by HQC, with the exception of network 2 where data source resolution and the high percentage of temporal gaps lead to extra flags (Fig. 11). The variable with the highest proportion of flagged values in network 2 was sea level pressure.

Given the automatic nature of the HQC tests, all values flagged by this step were removed and given a flag of 36. Values that were subjected to HQC were therefore marked with an additional flag (a prefix of 3), to clearly identify the level of testing applied to each individual observation (see Table 5 and Fig. 12). This means that observations which were corrected or verified in the SAQC round of testing (and given a flag of 2 or 4) but passed the spatial QC procedure had a final flag of fl32 or fl34, ensuring that information from both rounds of QC were retained to maintain the traceability of the QC procedure.

In total about 64,000 values were flagged and subsequently removed by HQC, around 0.7 % of the total dataset (Fig. 12). While the HQC tests were unable to be applied to all of the observations, these results are similar to the findings of other large-scale QC efforts (Dunn et al., 2012). Around 3.9 %, or about 330,000 observations were flagged by both QC procedures (Fig. 12). A total of 2.1 % of the data were removed as a result of SAQC and HQC testing, with 1.5 % corrected during the SAQC process. Only 0.3 % were flagged but later verified during SAQC, although this includes many legitimate extreme events that are crucial for calibrating and verifying the tail end of atmospheric behaviour that can have the largest societal impact. These results are generally on par with the percentage of keying errors identified in similar digitization efforts (Brönnimann et al., 2006).

4.4. Additional digitization quality assurance checks

In the final data check, a small conversion problem was detected with the sea level pressure at two Slovenian stations (around 318,000 values). The vast majority of these observations passed both SAQC and HQC, with large errors identified and flagged appropriately. However, these observations were also marked with an additional prefix of 4 (Table 4) to signify that additional QC may be required by future users.

Incidental errors throughout the digitization process, namely digitizers keying the same data twice, gave us an additional opportunity to examine the quality of several data sources. In particular, these opportunistic situations allowed us to identify the likely percentage of errors that would be identified using a double keying technique.

Zagazig, Egypt, 1932

The 0800 WD, WS and RH data for Zagazig, Egypt in 1932 were digitized twice by different digitizers: once using a template where every station on a page was digitized together, and once without a template but extracting only data from Zagazig from each source page. A total of 70 disagreements were found out of 1098 values, just over 6 % of the overlapping data. Interestingly, all but one disagreement was found to be due to errors in the data digitized using the template. A total of eight values were entered into an incorrect row, six values were misread by the digitizer as they were hard to read, and 55 errors were as a result of skipped days i.e. entire pages of data were skipped. All of the skipped days errors occurred in relative humidity, indicating that the digitizer worked through the source by digitizing one complete column at a time, rather than reading across each row. The one error in the non-templated data was due to an incorrect row being read.

Egypt 1931

Two digitizers inadvertently digitized 0800 SLP, TT, WS, WD and RH data for 11 stations in Egypt in 1931, both using the same template. A total of 308 differences were found between the two versions, 1.6 % of the 19800 values digitized. Checking the differences with the original source images revealed that 79 % were errors from one digitizer, and 21 % from the second digitizer. The most common error type was an incorrect row or column being read (54 % of errors), or the misreading of a value that was hard to decipher (43 %). Only 4 % of the errors identified were put down to gross typographical errors (e.g. 999 instead of 99).

These two Egyptian examples highlight a number of key issues with data digitization. The first is that the reliability of digitized data depends to a large extent on the reliability of the person digitizing those data. In both cases there was a clear separation between the two digitizers, even though (in the case of Egypt 1931), both digitizers used the same method. The second is that templates created without input from digitizers may not always achieve the best result. Indeed, follow up surveys with the digitizers suggested that several of the digitizers did not enjoy using templates, and preferred to work on spreadsheets they designed themselves.

Finally, these opportunistic analyses show that many of the errors made in the digitization process are small, particularly with daily or sub-daily data. Reading the value from a nearby station that is given in the row below the station of interest, or accidentally shifting the data by one day is very difficult to identify using automatic or semiautomatic quality control procedures. Triple, or even five-time keying of data is the best way to overcome these issues. However this is simply not feasible for many digitation projects due to limited funds, or a prioritisation of quantity over quality. While we cannot say that the dataset is free from errors, the methods we have used have removed or flagged the majority of suspect values.

5. Discussion

The quality of this UERRA dataset is due in large part to the extensive and multi layered quality control procedures employed to minimise errors from the data source or the transcription process. Meteorological data come in a wide range of formats, and preparing these data to be ingested into a national database, or shared among the research community, is not a trivial task. It can be time consuming, expensive and often difficult (Brönnimann et al., 2006). In particular, the transcription of the original observations (hereafter referred to as digitization) requires a lot of work-hours and resources. The preparation of data sources for keying, digitizing the values themselves and then assessing their quality are equally important components of the process. Without a reliable method of digitization and a standard method to assess the quality of sources, the accuracy of the final dataset can be jeopardised, potentially wasting the entire project.

There are some overarching guidelines currently available to assist organizations and communities who are conducting their own data recovery project. However, they are generally brief when it comes to specifics of the digitization method. Original WMO guidelines on climate data rescue (Tan et al., 2004), for example, include minimum information on the best method of data digitization, but instead focus on locating original data sources and data management.

In their guide for digitizing manuscript climate data, Brönnimann et al. (2006) describe the use of speech recognition, optical character recognition (OCR) and manual key entry. On balance, they found key entry to be the most efficient method of digitizing data, in terms of speed, error rate and amount of post-processing required. The WMO updated data rescue guidelines (World Meteorological Organization, 2016) support this finding, suggesting that OCR techniques are expensive and only appropriate for certain sources, while the human eye is still better when translating hand written observations.

The currently accepted best practise for manual data digitization is to double, or sometimes triple-key data using a "key what you see" method that employs templates which match the data source (Healy et al., 2004; World Meteorological Organization, 2016). Citizen science efforts, largely focussed on observations taken at sea, in fact require a value to be keyed five times (Eveleigh et al., 2013). Coupled with an automatic quality control procedure, these features of the digitization process are important for providing the best possible opportunity for data accuracy.

However, in reality this method is prohibitively expensive and not feasible for many small data recovery projects. Single data entry with visual checking is often the most cost effective and successful way of recovering valuable climate data for analysis, even though there are known issues around the resultant data quality. Based on our experience, we provide five key recommendations for other data rescue initiatives that might lack the resources to employ double or triple keying techniques:

- **Conduct a complete assessment of each data source before digitization**

It is vital to understand the limitations and issues of original data images and sources before the digitization process begins (Brönnimann et al., 2006). This is particularly the case if the data are provided in pre-scanned format. Checking every page of the original data source before providing it for digitization will save time and effort in the long term. Identify any mistakes in the page order, missing pages, images that are too dark or light to be read, or any changes in format or data units, to make an assessment of the data source quality.

Earth Syst. Sci. Data Discuss., https://doi.org/10.5194/essd-2018-39
Manuscript under review for journal Earth Syst. Sci. Data
Discussion started: 3 May 2018

- **Develop user friendly templates**

As suggested in the WMO's data rescue best practises, we still believe that the use of templates acts to reduce the number of digitization errors. Although templates will not remove issues associated with the original source, it will give digitizer the best chance to replicate what they see on the page. Templates that include automatic visualisation of the observations, highlight outliers, or enforce regular breaks would help to improve the quality of the resultant data.

We make this suggestion despite our case study finding that data digitized using a template included more errors than data digitized without (see Sect. 4.4.1). In this case study, one digitizer was asked to key data for more than 20 stations into a template, while the other digitized observations from only one station (one row per page of data source) without a template. Clearly there is a balance between the repetitive nature of keying in multiple rows of data, and the high chance of error associated with picking out one row of data in a complex table. Perhaps a compromise can be found here, where digitizers are charged with the slightly less repetitive task of finding one station in a page of observations, but still entering these data into a pre-formatted spreadsheet. Another suggestion could be to develop the templates in collaboration with the digitization team.

- **Involve digitizers in quality control procedures**

One potentially time saving method that can be employed to reduce digitization errors is to involve the digitizers in the quality assurance and quality control of the data. It is true that unreliable digitizers may also make unreliable quality control assessors, but by asking digitizers to run QC on data keys by others, they will become more aware of common errors they may make in their own work. This step can also help to identify errors within the data source, since other problems may occur due to poor observational practices leading to erroneous instrument readings or to mistakes when transcribing the data into secondary sources, among many other things (Brönnimann et al., 2006; Hunziker et al., 2017).

Another aspect of engaging digitizers in QC is to provide near-real-time feedback on the quality of the digitized data. It is no use finding that one digitizer is making a consistent error in their work if that task is already complete. Conducting QC as soon as data become available means you can advise the digitizers about their errors and hopefully make them more conscientious in the future.

- **Do not underestimate the value of manual checking quality control results**

Most of the currently available QC tests are based on statistical tests and are intended to identify individual errors or a chain of erroneous values. An alternative is visual QC checks, which, although existing, are not totally well developed nor employed and, therefore, data quality issues that may appear systematically can remain inadvertently in the data series (Hunziker et al., 2017).

Although manually checking the results of any QC procedure is very time consuming and tedious, our work suggests that for data rescue projects – particularly for critical spatial or temporal gaps – it is a necessary step to minimise the number of observations removed. Completely automated QC procedures used for global products run the risk of removing large swaths of data that can be corrected by a close examination of the reasons behind the flag. For example, if data from a station is out by one day due to a digitization error, it will likely be removed in any spatial analysis with neighbors. Automatic quality control procedures can also remove real extreme events or other observations that are correct but trigger flags as they have been converted from a coarser unit to those used in modern observations.

The value of manually assessing QC results means that it is also necessary to use an appropriate QC procedure. A QC tool that produces a large number of false quality flags will cause the project to lose a lot of time validating observations. For that reason it may be appropriate to tailor the QC procedure for different sources, providing that any variations of the procedure applied is recorded.

- **Provide all versions of the final dataset to enable traceability**

Finally, as with all dataset development, it is crucial to retain all versions of the data, from the original images to the raw keyed data, through all of the quality control iterations and any conversions applied. Manual checking of values may mean that it is not possible to create a truly reproducible product, but accompanying each data value with a quality flag and keeping every version of the data can create, as much as possible, a dataset that is traceable.

6. Data availability

The digitized dataset is available through the World Data Center PANGAEA (https://doi.pangaea.de/10.1594/PANGAEA.886511). The digitized dataset has also been provided to international data repositories including the International Surface Pressure Databank, the International Surface Temperature Initiative, and the C3S Lot 2 project through the British Science and Technology Facilities Council (STFC)/Centre for Environmental Data Analysis (CEDA), ECMWF's MARS Archive, the Global Precipitation Climatology Centre Dataset, the ECA&D and HadISD. In each case, the original, quality controlled and converted versions of the data are provided along with details of each source and quality control flag, to make the dataset as traceable as possible. The original data scans are available through each data repository (Table S2) and through the Universitat Rovira i Virgili Centre for Climate Change (ftp://130.206.36.123, user: C3_UERRA, password: c3uerra17).

7. Conclusions

This study describes our process of identifying, digitizing and quality controlling an extensive set of sub-daily meteorological observations for use by the wider research community. These multiple, complex steps are often overlooked when data are used for research, and yet without them, there would be no data.

The data we have rescued as part of the UERRA project totals 8.8 million observations from 15 countries, spanning 1879 to 2012. The observations cover the Mediterranean region, as well as eastern and central Europe, addressing data scarcity in these regions as identified in currently existing weather and climate data repositories.

Observations of temperature, atmospheric pressure, wind, humidity, and precipitation have been recovered from a wide range of original sources, from field books to daily weather registers kept for an entire country. Some sources were typed while others were hand written; some were provided in standard meteorological units, while others needed extensive conversion to be comparable with modern standards.

These observations have also been subjected to extensive quality control, making them useful for the development and verification of regional reanalysis, as well as potential studies of high-resolution weather at a station level. We developed a suite of semi-automatic QC programs to assess the quality of sub-daily data. These programs, which enabled us to manually check flagged values, were complemented by an automatic spatial QC procedure adapted from the global HadISD method. The QC procedure flagged 3.9 % of the total number of observations digitized, with 2.1 % of the total number removed, 1.5 % corrected, and 0.3 % retained as correct observations. Using an unadapted version of the HadISD procedure, or not checking the flagged values where possible, would have resulted in a large number of correct observations being removed, affecting the ability of future data users to examine extreme events and other high resolution atmospheric behaviour in a part of the world with limited observational coverage. These QC results are on par with other data rescue activities, and show the value of manually checking observations that have been digitized from paper. It is our hope that these observations support and improve the next generation of international and European weather and climate services.

8. Author contributions

PU was the coordinator of the UERRA project, and MB and PJ coordinated the data rescue component of WP1. LA, JRC, AG and MC managed the digitization and visual cross-check procedure. PD, EA, LA and JS developed the SAQC procedure. AG developed and ran the HQC procedure. IH provided information and analysis on the MARS data archive availability. The initial draft of the paper was written by LA with contributions from JRC for Sect. 3.2, 4.2 and 5, AG for Sect. 3.3, 4.3 and 5, and MB for Sect. 5. All authors contributed to the revision of the text.

9. Competing interests

The authors declare that they have no conflict of interest.

10. Acknowledgements

This work was undertaken as part of the Uncertainties in Ensembles of Regional ReAnalysis project (UERRA, SPAce grant number 607193). Many thanks to the UERRA digitization team: I. López, D. Azuara, J. Paradinas, A. Domènech, E. Pla, E. Romero, A. Robert, J. Tarragó, C. Bonfill, G. Vandellòs, R. Guerra, A. Àvila, A. Balart, M. Martin, E. Cebrián, P. Sabaté and A. Boqué. We are also very grateful to Robert Dunn for his help with the use of the HadISD quality control code, and to Lynda Chambers and Blair Trewin for their feedback on earlier versions of the manuscript.

11. References

Aguilar, E., Auer, I., Brunet, M., Peterson, T. C. and Wieringa, J.: Guidelines on Climate Metadata and Homogenization, World Meteorol. Organ., (1186), 55, 2003.

Allan, R., Brohan, P., Compo, G. P., Stone, R., Luterbacher, J., Brönnimann, S., Allan, R., Brohan, P., Compo, G. P., Stone, R., Luterbacher, J. and Brönnimann, S.: The International Atmospheric Circulation Reconstructions over the Earth (ACRE) Initiative, Bull. Am. Meteorol. Soc., 92(11), 1421–1425, doi:10.1175/2011BAMS3218.1, 2011.

Ashcroft, L., Gergis, J. and Karoly, D. J.: A historical climate dataset for southeastern Australia, 1788-1859, Geosci. Data J., 1(2), 158–178, doi:10.1002/gdj3.19, 2014.

Auer, I., Böhm, R., Jurkovic, A., Lipa, W., Orlik, A., Potzmann, R., Schöner, W., Ungersböck, M., Matulla, C., Briffa, K., Jones, P., Efthymiadis, D., Brunetti, M., Nanni, T., Maugeri, M., Mercalli, L., Mestre, O., Moisselin, J.-M., Begert, M., Müller-Westermeier, G., Kveton, V., Bochnicek, O., Stastny, P., Lapin, M., Szalai, S., Szentimrey, T., Cegnar, T., Dolinar, M., Gajic-Capka, M., Zaninovic, K., Majstorovic, Z. and Nieplova, E.: HISTALP—historical instrumental climatological surface time series of the Greater Alpine Region, Int. J. Climatol., 27(1), 17–46, doi:10.1002/joc.1377, 2007.

Brönnimann, S., Annis, J., Dann, W., Ewen, T., Grant, A. N., Griesser, T., Krähenmann, S., Mohr, C., Scherer, M. and Vogler, C.: A guide for digitising manuscript climate data, Clim. Past, 2(3), 191–207, doi:10.5194/cpd-2-191-2006, 2006.

Brunet, M. and Jones, P.: Data rescue initiatives: bringing historical climate data into the 21st century, Clim. Res., 47(1), 29–40, doi:10.3354/cr00960, 2011.

Brunet, M., Saladié, O., Jones, P., Sigró, J., Aguilar, E., Moberg, A., Lister, D., Walther, A., Lopez, D. and Almarza, C.: The development of a new dataset of Spanish Daily Adjusted Temperature Series (SDATS) (1850–2003), Int. J. Climatol., 26(13), 1777–1802, doi:10.1002/joc.1338, 2006.

Brunet, M., Gilabert, A., Jones, P. and Efthymiadis, D.: A historical surface climate dataset from station observations in Mediterranean North Africa and Middle East areas, Geosci. Data J., 1(2), 121–128, doi:10.1002/gdj3.12, 2014a.

Brunet, M., Jones, P. D., Jourdain, S., Efthymiadis, D., Kerrouche, M. and Boroneant, C.: Data sources for rescuing the rich heritage of Mediterranean historical surface climate data, Geosci. Data J., 1(1), 61–73, doi:10.1002/gdj3.4, 2014b.

Cannon, D. J., Brayshaw, D. J., Methven, J., Coker, P. J. and Lenaghan, D.: Using reanalysis data to quantify extreme wind power generation statistics: A 33 year case study in Great Britain, Renew. Energy, 75, 767–778, doi:10.1016/j.renene.2014.10.024, 2015.

Compo, G. P., Whitaker, J. S., Sardeshmukh, P. D., Matsui, N., Allan, R. J., Yin, X., Gleason, B. E., Vose, R. S., Rutledge, G., Bessemoulin, P., Brönnimann, S., Brunet, M., Crouthamel, R. I., Grant, A. N., Groisman, P. Y., Jones, P. D., Kruk, M. C., Kruger, A. C., Marshall, G. J., Maugeri, M., Mok, H. Y., Nordli, Ø., Ross, T. F., Trigo, R. M., Wang, X. L., Woodruff, S. D. and Worley, S. J.: The Twentieth Century Reanalysis Project, Q. J. R. Meteorol. Soc., 137(654), 1–28, doi:10.1002/qj.776, 2011.

Cram, T. A., Compo, G. P., Yin, X., Allan, R. J., McColl, C., Vose, R. S., Whitaker, J. S., Matsui, N., Ashcroft, L., Auchmann, R., Bessemoulin, P., Brandsma, T., Brohan, P., Brunet, M., Comeaux, J., Crouthamel, R., Gleason, B. E., Groisman, P. Y., Hersbach, H., Jones, P. D., Jónsson, T., Jourdain, S., Kelly, G., Knapp, K. R., Kruger, A., Kubota, H., Lentini, G., Lorrey, A., Lott, N., Lubker, S. J., Luterbacher, J., Marshall, G. J., Maugeri, M., Mock, C. J., Mok, H. Y., Nordli, Ø., Rodwell, M. J., Ross, T. F., Schuster, D., Srnec, L., Valente, M. A., Vizi, Z., Wang, X. L., Westcott, N., Woollen, J. S. and Worley, S. J.: The International Surface Pressure Databank version 2, Geosci. Data J., 2(1), 31–46, doi:10.1002/gdj3.25, 2015.

Dunn, R. J. H., Willett, K. M., Thorne, P. W., Woolley, E. V., Durre, I., Dai, A., Parker, D. E. and Vose, R. S.: HadISD: A quality-controlled global synoptic report database for selected variables at long-term stations from 1973-2011, Clim. Past, 8(5), 1649–1679, doi:10.5194/cp-8-1649-2012, 2012.

Dunn, R. J. H., Willett, K. M., Parker, D. E. and Mitchell, L.: Expanding HadISD: quality-controlled, sub-daily station data from 1931, Geosci. Instrumentation, Methods Data Syst., 5(2), 473–491, doi:10.5194/gi-5-473-2016, 2016.

Durre, I., Menne, M. J., Gleason, B. E., Houston, T. G., Vose, R. S., Durre, I., Menne, M. J., Gleason, B. E., Houston, T. G. and Vose, R. S.: Comprehensive Automated Quality Assurance of Daily Surface Observations, J. Appl. Meteorol. Climatol., 49(8), 1615–1633, doi:10.1175/2010JAMC2375.1, 2010.

Eveleigh, A., Jennett, C., Lynn, S. and Cox, A. L.: “I want to be a captain! I want to be a captain!”: gamification in the old weather citizen science project, Proc. First Int. Conf. Gameful Des. Res. Appl. - Gamification ’13, 79–82, doi:10.1145/2583008.2583019, 2013.

Healy, A. F., Kole, J. A., Buck-Gengler, C. J. and Bourne, L. E.: Effects of Prolonged Work on Data Entry Speed and Accuracy., J. Exp. Psychol. Appl., 10(3), 188–199, doi:10.1037/1076-898X.10.3.188, 2004.

Hunziker, S., Gubler, S., Calle, J., Moreno, I., Andrade, M., Velarde, F., Ticona, L., Carrasco, G., Castellón, Y., Oria, C., Croci-Maspoli, M., Konzelmann, T., Rohrer, M. and Brönnimann, S.: Identifying, attributing, and overcoming common data quality issues of manned station observations, Int. J. Climatol., 37(11), 4131–4145, doi:10.1002/joc.5037, 2017.

Kaspar, F., Tinz, B., Mächel, H. and Gates, L.: Data rescue of national and international meteorological observations at Deutscher Wetterdienst, Adv. Sci. Res., 12, 57–61, doi:10.5194/asr-12-57-2015, 2015.

Moberg, A., Jones, P. D., Lister, D., Walther, A., Brunet, M., Jacobeit, J., Alexander, L. V., Della-Marta, P. M., Luterbacher, J., Yiou, P., Chen, D., Klein Tank, A. M. G., Saladié, O., Sigró, J., Aguilar, E., Alexandersson, H., Almarza, C., Auer, I., Barriendos, M., Begert, M., Bergström, H., Böhm, R., Butler, C. J., Caesar, J., Drebs, A., Founda, D., Gerstengarbe, F.-W., Micela, G., Maugeri, M., Österle, H., Pandzic, K., Petrakis, M., Srnec, L., Tolasz, R., Tuomenvirta, H., Werner, P. C., Linderholm, H., Philipp, A., Wanner, H. and Xoplaki, E.: Indices for daily temperature and precipitation extremes in Europe analyzed for the period 1901–2000, J. Geophys. Res., 111(D22), D22106, doi:10.1029/2006JD007103, 2006.

Page, C. ., Nicholls, N., Plummer, N., Trewin, B., Manton, M., Alexander, L., Chambers, L., Choi, L. E., Collins, D. A., Gosai, A., Della-Marta, P., Haylock, M. ., Inape, K., Laurent, V., Maitrepierre, L., Makmur, E. P., Nagamigawa, H., Ouprasitwond, N., McGree, S., Pahalad, J., Salinger, M. ., Tibig, L., Tran, T. D., Vediapan, K. and Zhai, P.: Data Rescue in the Southeast Asia and South Pacific Region: Challenges and Opportunities, Bull. Am. Meteorol. Soc., 85(10), 1483–1489, doi:10.1175/BAMS-85-10-xxxx, 2004.

Peterson, T. C., Easterling, D. R., Karl, T. R., Groisman, P., Nicholls, N., Plummer, N., Torok, S., Auer, I., Boehm, R., Gullett, D., Vincent, L., Heino, R., Tuomenvirta, H., Mestre, O., Szentimrey, T., Salinger, J., Førland, E. J., Hanssen-Bauer, I., Alexandersson, H., Jones, P. and Parker, D.: Homogeneity adjustments of in situ atmospheric climate data: a review, Int. J. Climatol., 18(13), 1493–1517.

Stickler, A., Brönnimann, S., Valente, M. A., Bethke, J., Sterin, A., Jourdain, S., Roucaute, E., Vasquez, M. V., Reyes, D. A., Allan, R. and Dee, D.: ERA-CLIM: Historical Surface and Upper-Air Data for Future Reanalyses, Bull. Am. Meteorol. Soc., 95(9), 1419–1430, doi:10.1175/BAMS-D-13-00147.1, 2014.

Tan, L. S., Burton, S., Crouthamel, R., van Engelen, A., Hutchinson, R., Nicodemus, L., Peterson, T. C. and Rahimzadeh, F.: Guidelines on Climate Data Rescue WMO/TD No. 1210. [online] Available from: http://www.wmo.int/pages/prog/wcp/wcdmp/documents/WCDMP-55.pdf, 2004.

Venema, V. K. C., Mestre, O., Aguilar, E., Auer, I., Guijarro, J. A., Domonkos, P., Vertacnik, G., Szentimrey, T., Stepanek, P., Zahradnicek, P., Viarre, J., Müller-Westermeier, G., Lakatos, M., Williams, C. N., Menne, M. J., Lindau, R., Rasol, D., Rustemeier, E., Kolokythas, K., Marinova, T., Andresen, L., Acquaotta, F., Fratianni, S., Cheval, S., Klancar, M., Brunetti, M., Gruber, C., Prohom Duran, M., Likso, T., Esteban, P. and Brandsma, T.: Benchmarking homogenization algorithms for monthly data, Clim. Past, 8(1), 89–115, doi:10.5194/cp-8-89-2012, 2012.

World Meteorological Organization: Guidelines on Best Practices for Climate Data Rescue 2016., 2016.

Earth Syst. Sci. Data Discuss., https://doi.org/10.5194/essd-2018-39
Manuscript under review for journal Earth Syst. Sci. Data
Discussion started: 3 May 2018

Tables

Table 1: An overview of the data sources used in this project. More information on the precise temporal coverage of each location and units are provided with the dataset, available on Pangaea. Variables given in bold in the variable column have been digitized as part of this project: not all available variables and time periods were digitized in this project due to time and funding constraints. Each source can be found at ftp://130.206.36.123, u: C3_UERRA, p: c3uerra17, folder: C3_UERRA_datasources_images, where each source is listed under their source code. In the source location column, NOAA-CDMP represents the National Oceanic and Atmospheric Administration Climate Data Modernisation Project. The variables are represented by acronyms similar to those used in the main text: temperature (TT), relative humidity (RH), dew point temperature (DP), mean sea level pressure (PP), and station level pressure (SP), wind direction (WD), wind speed (WS), wet bulb temperature (WB), precipitation (RR), snow depth (SD) and fresh snow (FS).

Table 2: List of conversions applied to digitized data, where x represents the original unit and y is the converted value. Full details of the conversion applied to data from each station is given in Table S2.

Table 3: Descriptions of the SAQC tests applied for each climate variable. Variable acronyms are as those described in Table 1. The programs used to apply each test are available at http://www.c3.urv.cat/softdata.php.

Table 4: Description of quality control flags applied to data during the SAQC and HQC procedures.

Table 5: List of HQC tests applied to data, specifying which test ran for which variable in their original state (R), and which tests were adapted (A) to be applied to the UERRA data set. Full adaptation details are provided in the text. For more details on individual tests, see Dunn et al. (2012).

Table 6: HQC networks with the countries, periods, observing times and some comments.

Table 7: Summary of stations digitized as part of the UERRA project. The variables are temperature (TT), relative humidity (RH), dew point temperature (DP), wind speed (WS), wind direction (WD), air pressure (PP, including sea level pressure and station level pressure), wet bulb temperature (WB), total snow depth (SD), fresh snow (FS) and precipitation (RR). The digitized dataset is available through the World Data Center PANGAEA (https://doi.pangaea.de/10.1594/PANGAEA.886511), in the format of one file for each variable and country

FIGURES

Figure 1: Stations with monthly mean sea level pressure data in MARS across the three identified regions of interest: a) the Mediterranean b) Eastern Europe and c) Scandinavia. The colors indicate the percentage of data available for 1950–2010.

Figure 2: Examples of the different data source formats found for digitization: a) Egypt, 1939, where each row is data for a different station on one day; b) Morocco 1968, where each row is a data for a different station on one day; c) Kredarica, Slovenia 1970, where each row is data of a different variable for one station on one day; d) Ksara, Lebanon 1939, where each row is atmospheric pressure data for one day at one station. Data images that have been permitted to be shared by the data source owner permission are available online through the Universitat Rovira i Virgili's Centre for Climate Change.

Figure 3: Examples of the templates used in data digitization. Shaded rows and columns in the templates represent data that are not to be digitized a) The template for the Slovenian data sources picks out the rows that require digitizing (acronyms used in this sheet: wind direction WD, wind speed WS, atmospheric pressure SLP, temperature T, relative humidity RH, precipitation P, snow depth SD and fresh snow FS). Note that rows for the daily values are formatted to match the location of the data in the original source. b) the template for temperature data from Spanish data sources with the columns labelled with variables and hours (dry bulb temperature TD, relative humidity HU and dew point temperature PR).

Figure 4: A schematic of the digitization quality assurance and quality control procedures used in the development of the dataset.

Figure 5: Air temperature evolution (in ºC) in Port Said station (Egypt) taken at 0800 (in black) and 1400 (in grey) for the period 1939–1940. Different errors flagged by SAQC are marked with solid colored squares: an outlier (pink), outlier and intervariable (IV) error (yellow), IV error (orange), and big jump, IV error and outlier (red). The decision made by manual checking is shown by rectangular outlines: flagged values identified as transcription errors are outlined by a red border, flags due to a data duplication error are outlined in blue, and flags that were found to be valid extremes are outlined in green.

Figure 6: Percentage of flagged values using the standard QC tests developed for the U.K. Met Office Hadley Centre Global Sub-Daily Station Observations (HadISD), and the percentage of values flagged using HadISD tests specifically adapted for the UERRA dataset. The variable acronyms are the same as those given in the text: temperature (TT), dew point temperature (DP), mean sea level pressure (PP), wind direction (WD), and wind speed (WS).

Figure 7: Spatial coverage of 8.8 million observations digitized showing the station locations. The approximate length of the record at each station is indicated by the size of the pie symbol; the number of observations per day is represented by the color of the pie pieces; and the different variables available at each station are indicated by which wedges are shaded based on the legend in the top right corner. The variables are represented by acronyms Variable acronyms are as those described in the caption to Table 1, apart from SLP, which represents station and sea level pressure.

Earth Syst. Sci. Data Discuss., https://doi.org/10.5194/essd-2018-39
Manuscript under review for journal Earth Syst. Sci. Data
Discussion started: 3 May 2018

Figure 8: Distribution of the digitized observations by a) country, b) variable, c) decade and d) hour of observation. The length of each bar shows the number of observations digitized (in millions), with orange indicating any observations removed during quality control. Variable acronyms are as those described in Table 1. Country codes are as those listed in Table 7

Figure 9: Percentages of flagged/not flagged values derived from SAQC application to the UERRA dataset. The green wedge represents data that passed SAQC; orange wedges represent values that have been removed (1.5 % of the total); blue wedges represent data that were corrected (1.3 % of the total) and purple wedges show the data that were verified (0.3 % of the total). Flag codes given are explained in Table 4.

Figure 10: Total counts (in percentage) of error flags by countries (a), variables (b), observation times (c) and decades (d) derived from SAQC application to the UERRA dataset. Purple indicates values that were flagged but verified; blue indicates values that were flagged and corrected; and red and orange indicate values that were flagged and removed as errors. Variable acronyms are as those described in Table 1. Flag descriptions are given in Table 4.

Figure 11: The percentage of values flagged within each network (see Table 6) tested using the HQC automatic procedure. Variable acronyms are as explained in the caption for Table 1, noting that not all variables were included in each network.

Figure 12: The percentage distribution of quality control flags in the UERRA dataset. Values that have passed QC are represented in green (QC flags fl10, fl40 and fl30); values that were flagged but verified as correct are shown in purple (fl14, fl44 and fl34); values that were flagged but corrected are shown in blue (fl12 fl42, fl32); and values that were flagged and removed are shown in orange (fl11, fl13, fl15, fl17 and fl36). The darkness of the colors indicates the level of QC applied for each flag. Lighter colors represent values that were only subjected to semi-automatic quality control (SAQC, fl codes that begin with 1 and 4), while darker colors indicate values subjected to both SAQC and spatial HQC procedures (fl codes that begin with 3). See Table 4 for additional flag details.

Earth Syst. Sci. Data Discuss., https://doi.org/10.5194/essd-2018-39
Manuscript under review for journal Earth Syst. Sci. Data
Discussion started: 3 May 2018

Table 1: An overview of the data sources used in this project. More information on the precise temporal coverage of each location and units are provided with the dataset, available on Pangaea. Variables given in bold in the variable column have been digitized as part of this project: not all available variables and time periods were digitized in this project due to time and funding constraints. Each source can be found at ftp://130.206.36.123, u: C3_UERRA, p: c3uerra17, folder: C3_UERRA_datasources_images, where each source is listed under their source code. In the source location column, NOAA-CDMP represents the National Oceanic and Atmospheric Administration Climate Data Modernisation Project. The variables are represented by acronyms similar to those used in the main text: temperature (TT), relative humidity (RH), dew point temperature (DP), mean sea level pressure (PP), and station level pressure (SP), wind direction (WD), wind speed (WS), wet bulb temperature (WB), precipitation (RR), snow depth (SD) and fresh snow (FS).

Source code and name	Country (or countries) covered	Time period covered (continuous)	Source provider	Primary or secondary	Typed or hand written	Variables	Details
So04: Bulletin Météorologique de l'Algérie	Algeria	1877–1938	NOAA-CDMP	Secondary	Hand-written	**TT, PP, WS, WD**, Cloud, weather condtions, RR, Tmax, Tmin	Sub-daily observations, multiple variables and stations per day on each page, average reliability and some issues with chronological order of pages. One file per year.
So12: Annales de l'Observatoire de Ksara	Lebanon	1921–1971	NOAA-CDMP	Secondary	Hand-written	**TT, PP, RH**, summaries of wind, sunshine, evaporation, rainfall, clouds, weather	Hourly observations, one variable per month per station on each page, good readability and source in good chronological order. One file per year.
So06: Bulletin Meteorologique du Maroc	Morocco, Algeria	1953–1968, 1977–1978	NOAA-CDMP	Secondary	Typed	**TT, PP, ST, DP, WS, WD**, Cloud, RR, Tmax, Tmin	Sub-daily observations, multiple variables and stations per day on each page, good readability and source in good chronological order. One file per year.
So63: Cyprus Meteorological Returns	Cyprus	1881–1922	UK Met Office	Primary	Hand-written	**TT**, ST, WS, WD, Cloud, RR	Sub-daily observations, multiple variables for one station on each data, good readability and in good chronological order. One file per month.
So62: German station observing books	Germany	1958–1978	German Meteorological Service DWD	Primary	Hand-written	**TT, PP, ST, DP, RH, WB, WD, WS, RR, FS, SD**	Daily snowfall data for some stations, one station per month on each page, hourly or half-hourly observations provided for two locations (Dresden and Brocken), multiple variables for one station per day on each page. One file per year. Good readability and in good chronological order
So16: Egypt daily weather report	Egypt	1907–1957	NOAA-CDMP	Secondary	Hand-written	**TT, PP, RH, DP, WS, WD**, RR, Cloud, Visibility, Weather	Sub-daily observations from multiple stations and variables for one day on each page, Average readability and some issues with chronological order. One file per year.

So54: Instituto Nacional de Meteorología Banco de Datos	Spain	1954–1984	Provided by MeteoCat, but containing data from the Spanish Meteorological Agency (AEMET)	Secondary	Hand-written and typed	**TT, PP, ST, RH, DP, WD, WS,** Cloud, RR	Sub-daily data one station per month on each page, good readability and in good chronological order. One file per month.
So56: Meteoroloski godisnjak 1 - klimatoloski podaci	Bosnia and Herzegovina, Croatia, Republic of Serbia	1949–2012	Provided by the Republic Hydrometeorological Institute of Serbia	Secondary	Typed	**TT, ST, RH, WS, WD,** Vapour pressure, RR, SD, Cloud, Visibility, Weather	Sub-daily data one station per month on each page, good readability and in good chronological order. One file per year.
So64: Rocenka-annuaire	Czech Republic, Slovak Republic	1940–1968	NOAA-CDMP	Secondary	Hand-written	**TT, PP, ST, RH, WS, WD,** Visibility, Cloud, RR, Weather	Sub-daily data one station per month on each page, good readability and in good chronological order. One file per year.
So65: Slovenian meteorological observing books	Slovenia	1950–1978	Provided by the Slovenian Environmental Agency	Primary	Hand-written	**TT, PP, ST, RH, DP, SD, FS, RR, WS, WD,** WB, Cloud, Visibility, Weather	Hourly data, one station per day on each page, in good chronological order but difficult to read at times. One file per day.
So61: Yillik Meteorolojoi Bülteni	Turkey	1962–1971	NOAA-CDMP	Secondary	Typed	**TT, ST, RH, WS, WD,** Tmax, Tmin, Cloud, Evaporation, RR, Weather	Sub-daily data one station per month on each page, good readability and in good chronological order. One file per year.

Earth Syst. Sci. Data Discuss., https://doi.org/10.5194/essd-2018-39
Manuscript under review for journal Earth Syst. Sci. Data
Discussion started: 3 May 2018

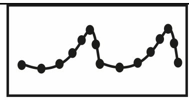
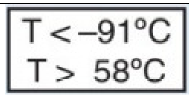
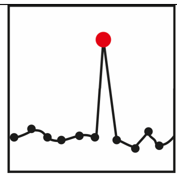
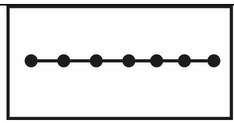
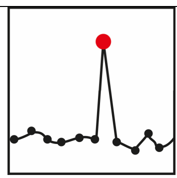
Table 2: List of conversions applied to digitized data, where x represents the original unit and y is the converted value. Full details of the conversion applied to data from each station is given in Table S2.

	Original units	**final units**	**Details**
Wind speed conversions	Beaufort scale	m/s	replacement of x with y using the following map: 0=0 1=1 2=2.6 3=4.6 4=6.7 5=9.3 6=12.3 7=15.4 8=19 9=22.6 10=26.8 11=30.9 12=35, from WMO Code 1100
	Turkish 17-point power scale	m/s	replacement of x with y using the following map: 0=0 1=0.9 2=2.4 3=4.4 4=6.7 5=9.3 6=12.3 7=15.5 8=18.9 9=22.6 10=26.4 11=30.5 12=34.8 13=39.2 14=43.8 15=48.6 16=53.5 17=58.6 taken from data source (average of wind range used)
	9 point power scale	m/s	replacement of x with y using the following map: 0=0 1=1 2=2.6 3=4.6 4=6.7 5=9.3 6=12.3 7=15.4 8=19 9=28.8, from 1931 French instruction book http://bibliotheque.meteo.fr/exl-php/vue-consult/mf_-_recherche_avancee/ISO00008704
	km/h	m/s	y=x/3.6 rounded to 1 decimal place
	knots	m/s	y=x*0.514444 rounded to 1 decimal place
Wind direction conversions	16 point compass scale	degrees	replacement of x with y using the following map: C=0 NNE=22.5 NE=45 ENE=67.5 E=90 ESE=112.5 SE=135 SSE=157.5 S=180 SSW=202.5 SW=225 WSW=247.5 W=270 WNW=292.5 NW=315 NNW=337.5 N=360
	32 point direction scale	degrees	y=x*11.25
	degrees/10	degrees	y=x/10.0
Pressure conversions	mmhg	hPa	y=x*1.33224 rounded to 1 decimal place
	hpa*10	hPa	y=x/10
Temperature conversions	degF	degC	y=(x-32)*(5/9) rounded to 1 decimal place

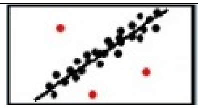
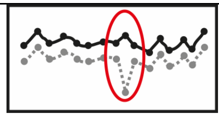
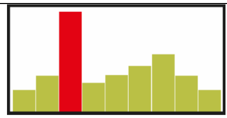
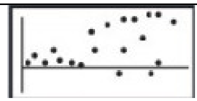
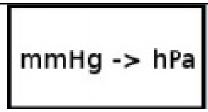
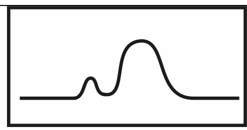
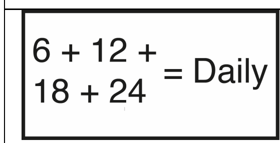
Earth Syst. Sci. Data Discuss., https://doi.org/10.5194/essd-2018-39
Manuscript under review for journal Earth Syst. Sci. Data
Discussion started: 3 May 2018

Open Access Earth System Science Data Discussions

Table 3: Descriptions of the SAQC tests applied for each climate variable. Variable acronyms are as those described in Table 1. The programs used to apply each test are available at http://www.c3.urv.cat/softdata.php.

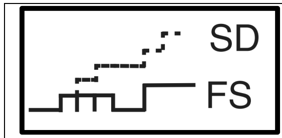
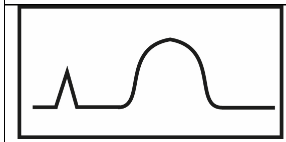
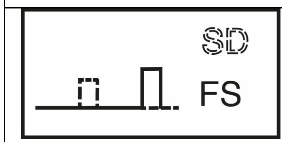
SAQC test		Brief description	Variables
	Date order check	Detect erroneous calendar date order	TT/DP/RH/WD/WS/RR/PP
	Data repetitions	Flag repeated entire months	TT/DP/RH/WD/WS/PP
	Unrealistic values	Flag values outside world record limits and physically impossible values	TT/DP/RH/WD/WS/PP
	Climatic outliers	Flag values out of an established threshold	TT/DP/RH/WS/PP/RR/FS/SD
	Duplicate values	Detect at least 10 identical consecutive values (later relaxed to 30 identical values)	TT/DP/RH/WD/WS
	Big jumps and sharp spikes	Large differences between adjacent values	TT/DP

Earth Syst. Sci. Data Discuss., https://doi.org/10.5194/essd-2018-39
Manuscript under review for journal Earth Syst. Sci. Data
Discussion started: 3 May 2018

	Bivariate outliers	Differences between adjacent values that are larger than the bivariate distribution	PP
	Intervariable inconsistency	Flag internal inconsistencies among variables	TT/RH/WS
	DP inconsistency	Differences between observed and calculated DP	TT/RH/DP
	Monthly mean of absolute increments	Mean monthly increments below/above of the climatic normal increment	TT/RH/WD/WS
	Irregular temporal evolution	Unexpected temporal evolution	TT/RH/WS
	Unit changes	Unit changes (mmHg-hPa)	PP
	Unusual distribution of values	Distribution of values in each month includes a secondary peak	TT,DP, VV
	Precipitation totals	Sum of sub-daily RR data = daily RR total	RR

Earth Syst. Sci. Data Discuss., https://doi.org/10.5194/essd-2018-39
Manuscript under review for journal Earth Syst. Sci. Data
Discussion started: 3 May 2018

Open Access Earth System Science Data Discussions

	Snow totals	Sum of fresh snow <= total snow depth	FS/SD
	Summer snow	Snowfall between May and October	FS/SD
	Non-numeric values	Removal of non-numeric values	PP/RR/FS/SD
	FS/SD inconsistency	Flag total SD that increase without a FS fall and inverse. FS is not accompanied by SD.	FS/SD

Earth Syst. Sci. Data Discuss., https://doi.org/10.5194/essd-2018-39
Manuscript under review for journal Earth Syst. Sci. Data
Discussion started: 3 May 2018

Table 4: Description of quality control flags applied to data during the SAQC and HQC procedures.

Flag	**Flag description**	**Expert decision**
fl10	Passed statistical quality control	Retained
fl11	Identified as suspect and removed due to gross digitizer error	Removed
fl12	Identified as suspect, found to be a digitization error, corrected	Corrected
fl13	Identified as suspect, found to be a digitization error, removed	Removed
fl14	Identified as suspect but retained as correct after expert examination	Retained
fl15	Identified as suspect, found to be a source error and removed	Removed
fl17	Identified as suspect, no observation found in source, removed	Removed
fl30	Passed SAQC and HQC	Retained
fl32	Corrected in SAQC, passed HQC	Retained
fl34	Retained as correct in SAQC, passed HQC	Retained
fl36	Identified as suspect in HQC, removed	Removed
fl40	Passed statistical quality control but updated to correct units after location of accurate metadata	Retained
fl42	Identified as suspect, found to be a digitization error and corrected, then updated to correct units after location of accurate metadata	Corrected
fl44	Identified as suspect but retained as correct after expert examination, then updated to correct units after location of accurate metadata	Retained

Earth Syst. Sci. Data Discuss., https://doi.org/10.5194/essd-2018-39
Manuscript under review for journal Earth Syst. Sci. Data
Discussion started: 3 May 2018

Table 5: List of HQC tests applied to data, specifying which test ran for which variable in their original state (R), and which tests were adapted (A) to be applied to the UERRA data set. Full adaptation details are provided in the text. For more details on individual tests, see Dunn et al. (2012).

Test	**TT**	**TD**	**SLP**	**FF**	**WD**
Duplicate months check	R	R	R	R	R
Odd cluster check	R	R	R	R	R
Frequent values check	A	A	A		
Diurnal cycle check	R	R	R	R	R
Distributional gap check	A	A	A		
Known record check	A	A	A	A	A
Repeated streaks/unusual spell frequency check	A	A	A	A	A
Climatological outliers check	R	R			
Spike check	R	R	R		
T and Td cross-check		R			
Unusual variance checks	R	R	R	R	R
Wind checks				A	A
Nearest-neighbor data check	A	A	A		
Station clean-up	A	A	A	A	A

Table 6: HQC networks with the countries, periods, observing times and some comments.

Network	Countries	Number of stations	Period	Variables	Observing times	% of data checked
1	Morocco	8	1953-1968	PP, TT, WD, FF DP	6 or 7h/18h (2)	94.5 %
2	Algeria, Tunisia	25	1886-1938	PP, TT, WD, FF	7h (1)	91.5 %
3	Egypt	21	1907-1957	PP, TT, WD, FF DP	6h or 8h/12h or 14h/18h or 20h (4)	69.6 %
4	Turkey	25	1962-1971	PP, TT, WD, FF	7h/14h/21h (3)	100 %
5	Slovenia, Croatia, Bosnia and Herzegovina, Serbia	10	1950-2012	PP, TT, WD, FF	7h/14h/21h (3)	72.7 %
6	Slovenia, Czech Republic, Germany (2 stations)	11	1948-1968	PP, TT, WD, FF	7h/14h/21h (3)	46.99 %

Earth Syst. Sci. Data Discuss., https://doi.org/10.5194/essd-2018-39
Manuscript under review for journal Earth Syst. Sci. Data
Discussion started: 3 May 2018

Table 7: Summary of stations digitized as part of the UERRA project. The variables are temperature (TT), relative humidity (RH), dew point temperature (DP), wind speed (WS), wind direction (WD), air pressure (PP, including sea level pressure and station level pressure), wet bulb temperature (WB), total snow depth (SD), fresh snow (FS) and precipitation (RR). The digitized dataset is available through the World Data Center PANGAEA (https://doi.pangaea.de/10.1594/PANGAEA.886511), in the format of one file for each variable and country.

Country	Country code	Number of stations	Period covered	Variables	Number of observations per day	Total digitized	Total after QC	Percentage of data removed in SAQC
Algeria	ALG	21	1877–1968	DP, WS, WD, SLP, TT	4	684114	665369	2.74
Bosnia and Herzegovina	BOH	2	1953–1984	WS, WD, PP, RH, TT	3	125831	115894	7.90
Croatia	CRO	2	1949–1984	WS, WD, PP, RH, TT	3	391789	390141	0.42
Cyprus	CYP	2	1881–1922	TT	2	45070	45068	0.00
Czech Republic	CZE	7	1948–1968	WS, WD, PP, RH, TT	6	379582	377843	0.46
Egypt	EGY	18	1907–1957	DP, WS, WD, PP, RH, TT	6	1371436	1336281	2.56
Germany	GER	23	1958–1978	WS, WD, FS, PP, RH, RR, SD, TT, WB, DP	Up to 24	697308	692750	0.65
Lebanon	LBN	1	1930–1939	PP, RH, TT	24	262944	254044	3.38
Morocco	MAR	8	1910–1968	DP, WS, WD, PP, TT	4	340563	336170	1.29
Serbia	SER	3	1949–2012	WS, WD, PP, RH, TT	3	358898	356058	0.79
Slovak Republic	SLO	2	1940–1967	WS, WD, PP, RH, TT	6	248751	247541	0.49
Slovenia	SLV	3	1950–1978	DP, WS, WD, FS, PP, RH, RR, SD, TT	Up to 24	2507878	2437163	2.82
Spain	ESP	5	1954–1984	WS, WD, PP, RH, TT, DP	5	194274	192670	0.83
Tunisia	TUN	5	1886–1938	WS, WD, PP, TT	1	174900	170480	2.53
Turkey	TUR	25	1962–1971	WS, WD, PP, RH, TT	3	1028898	1017871	1.07
Total		*127*				*8812236*	*8627338*	*2.00*

Earth Syst. Sci. Data Discuss., https://doi.org/10.5194/essd-2018-39
Manuscript under review for journal Earth Syst. Sci. Data
Discussion started: 3 May 2018

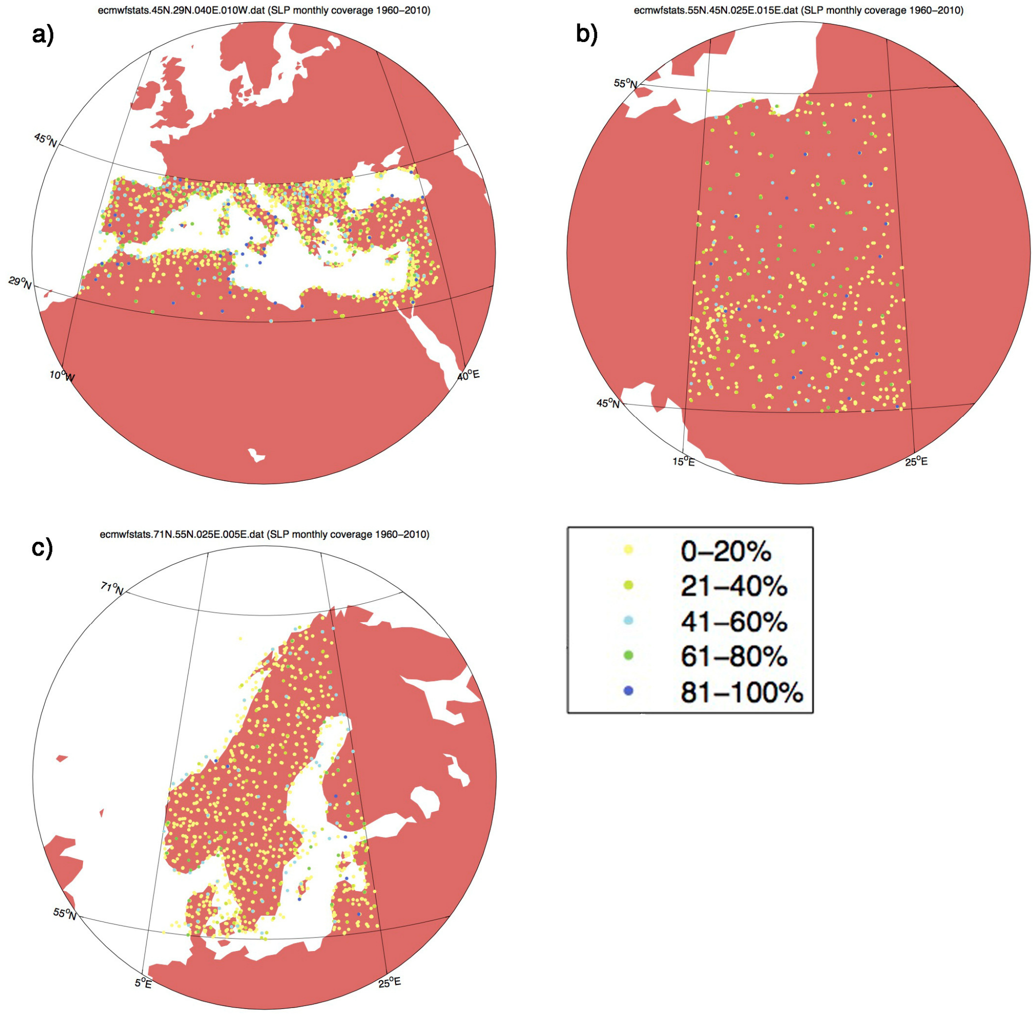


Figure 1: Stations with monthly mean sea level pressure data in MARS across the three identified regions of interest: a) the Mediterranean b) Eastern Europe and c) Scandinavia. The colors indicate the percentage of data available for 1950–2010.

Earth Syst. Sci. Data Discuss., https://doi.org/10.5194/essd-2018-39
Manuscript under review for journal Earth Syst. Sci. Data
Discussion started: 3 May 2018

Open Access Earth System Science Data Discussions

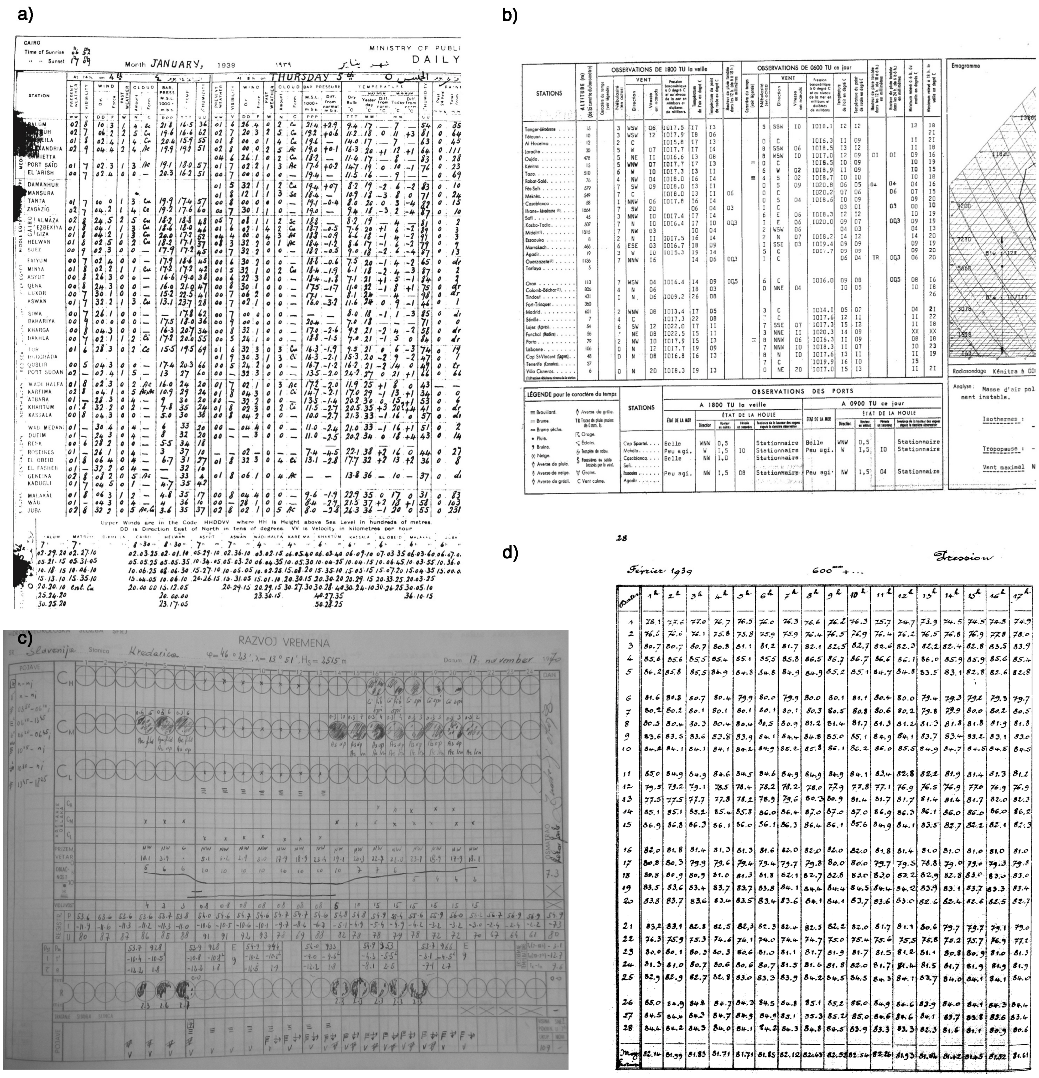
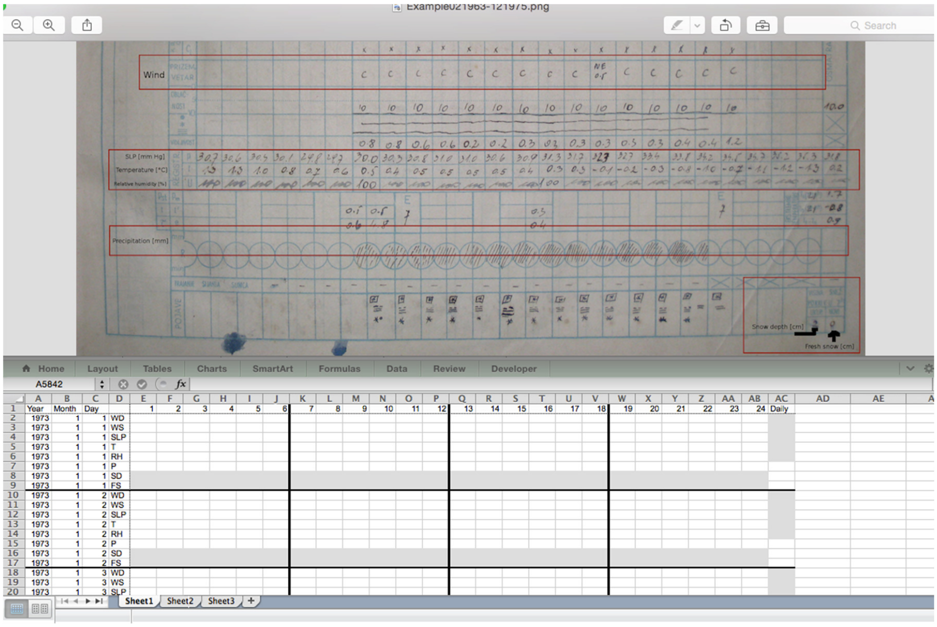


Figure 2: Examples of the different data source formats found for digitization: a) Egypt, 1939, where each row is data for a different station on one day; b) Morocco 1968, where each row is a data for a different station on one day; c) Kredarica, Slovenia 1970, where each row is data of a different variable for one station on one day; d) Ksara, Lebanon 1939, where each row is atmospheric pressure data for one day at one station. Data images that have been permitted to be shared by the data source owner permission are available online the Universitat Rovira i Virgili's Centre for Climate Change.

Earth Syst. Sci. Data Discuss., https://doi.org/10.5194/essd-2018-39
Manuscript under review for journal Earth Syst. Sci. Data
Discussion started: 3 May 2018

a) Slovenia



b) Spain

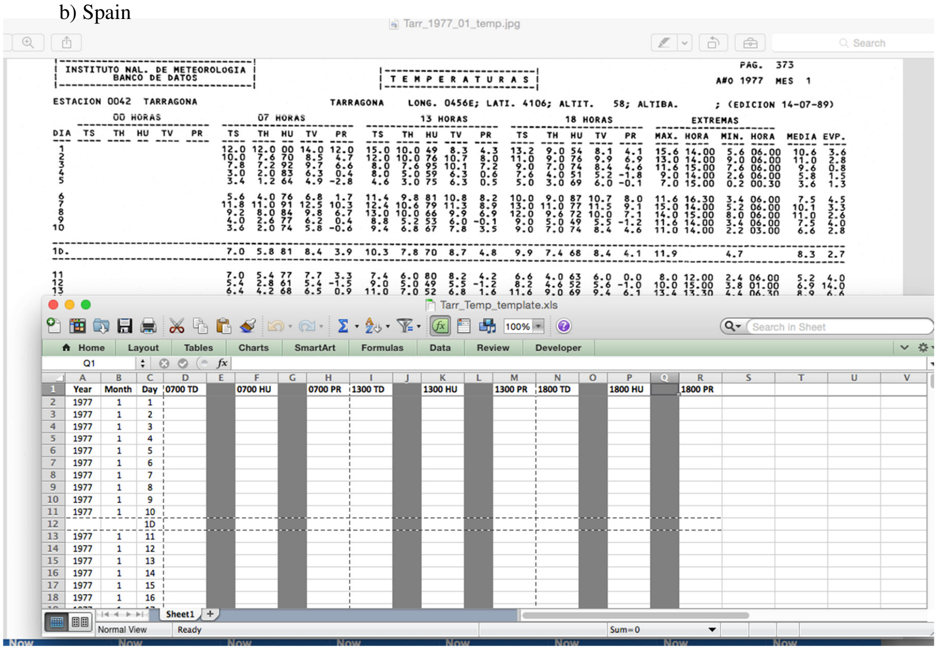


Figure 3: Examples of the templates used in data digitization. Shaded rows and columns in the templates represent data that are not to be digitized a) The template for the Slovenian data sources picks out the rows that require digitizing (acronyms used in this sheet: wind direction WD, wind speed WS, atmospheric pressure SLP, temperature T, relative humidity RH, precipitation P, snow depth SD and fresh snow FS). Note that rows for the daily values are formatted to match the location of the data in the original source. b) the template for temperature data from Spanish data sources with the columns labelled with variables and hours (dry bulb temperature TD, relative humidity HU and dew point temperature PR).

Earth Syst. Sci. Data Discuss., https://doi.org/10.5194/essd-2018-39
Manuscript under review for journal Earth Syst. Sci. Data
Discussion started: 3 May 2018

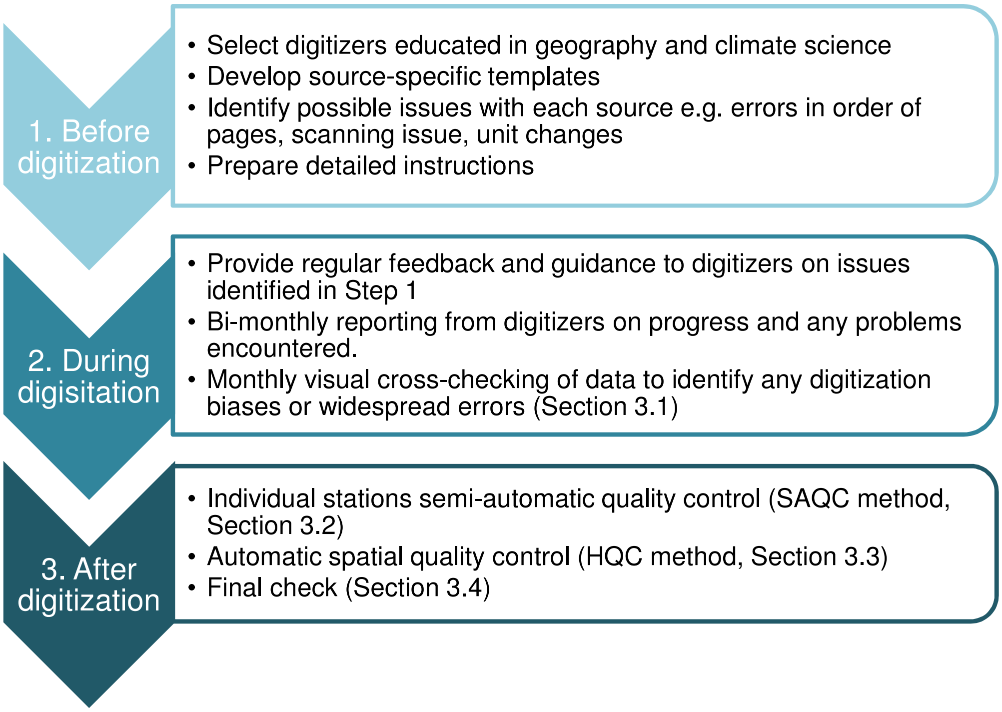


Figure 4: A schematic of the digitization quality assurance and quality control procedures used in the development of the dataset.

Earth Syst. Sci. Data Discuss., https://doi.org/10.5194/essd-2018-39
Manuscript under review for journal Earth Syst. Sci. Data
Discussion started: 3 May 2018

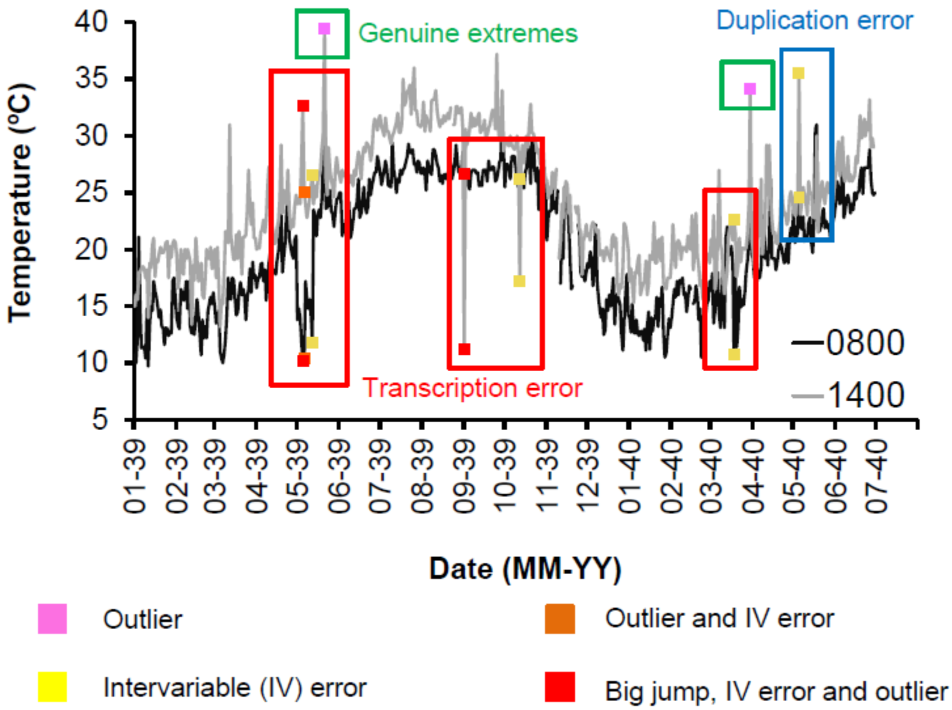


Figure 5: Air temperature evolution (in ºC) in Port Said station (Egypt) taken at 0800 (in black) and 1400 (in grey) for the period 1939–1940. Different errors flagged by SAQC are marked with solid colored squares: an outlier (pink), outlier and intervariable (IV) error (yellow), IV error (orange), and big jump, IV error and outlier (red). The decision made by manual checking is shown by rectangular outlines: flagged values identified as transcription errors are outlined by a red border, flags due to a data duplication error are outlined in blue, and flags that were found to be valid extremes are outlined in green.

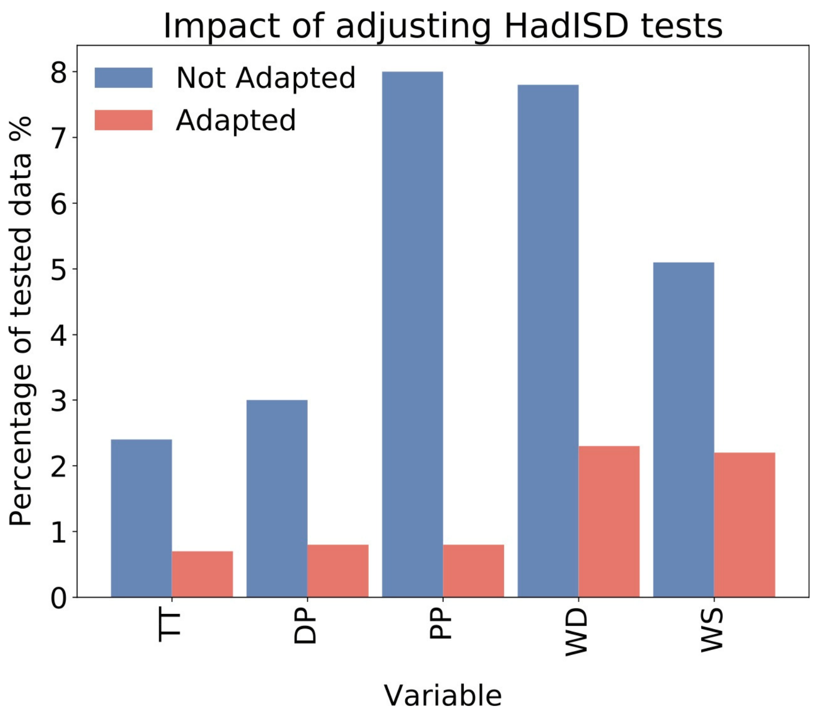


Figure 6: Percentage of flagged values using the standard QC tests developed for the U.K. Met Office Hadley Centre Global Sub-Daily Station Observations (HadISD), and the percentage of values flagged using HadISD tests specifically adapted for the UERRA dataset. The variable acronyms are the same as those given in the text: temperature (TT), dew point temperature (DP), mean sea level pressure (PP), wind direction (WD), and wind speed (WS).

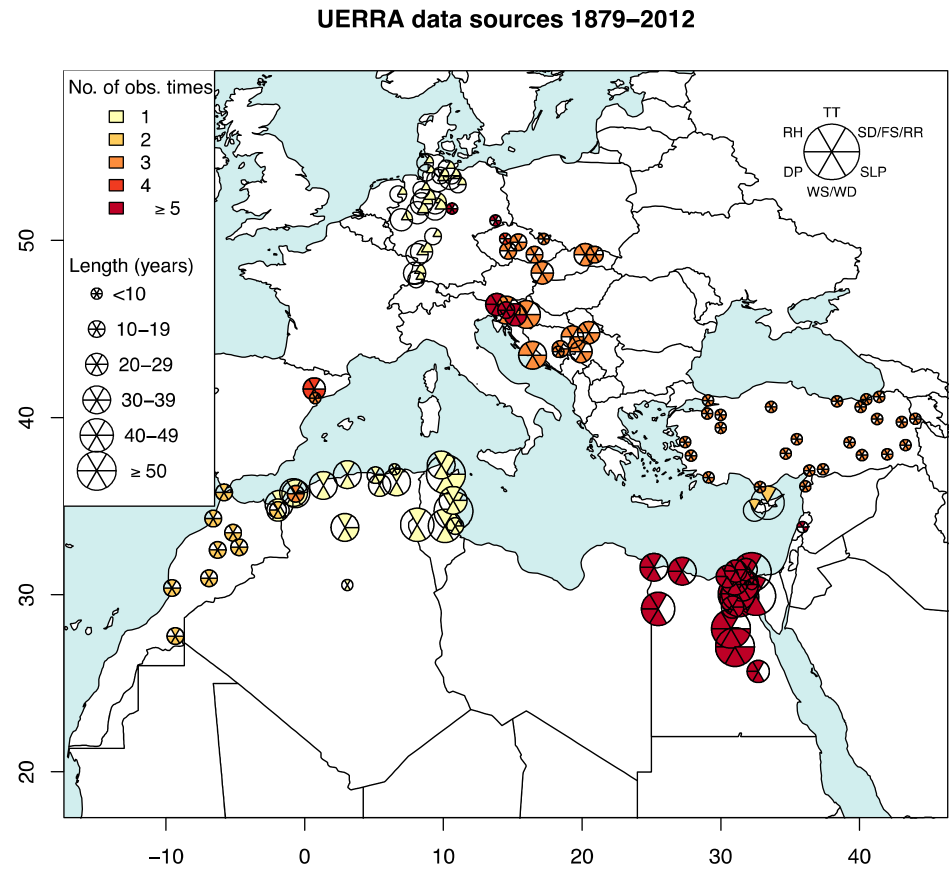
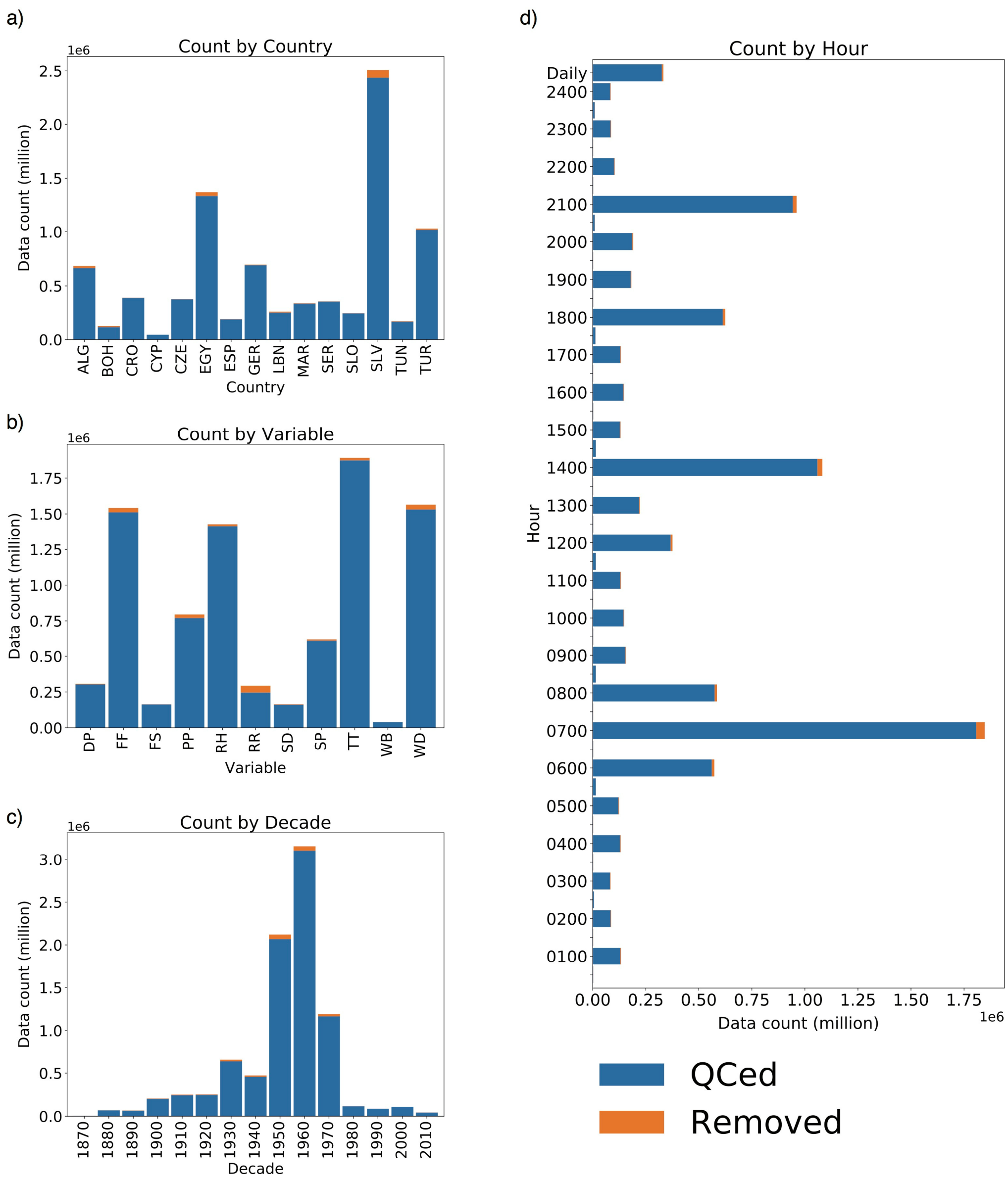


Figure 7: Spatial coverage of 8.8 million observations digitized showing the station locations. The approximate length of the record at each station is indicated by the size of the pie symbol; the number of observations per day is represented by the color of the pie pieces; and the different variables available at each station are indicated by which wedges are shaded based on the legend in the top right corner. The variables are represented by acronyms Variable acronyms are as those described in the caption to Table 1, apart from SLP, which represents station and sea level pressure.

Earth Syst. Sci. Data Discuss., https://doi.org/10.5194/essd-2018-39
Manuscript under review for journal Earth Syst. Sci. Data
Discussion started: 3 May 2018



104

105 **Figure 8: Distribution of the digitized observations by a) country, b) variable, c) decade and d) hour of observation. The**
106 **length of each bar shows the number of observations digitized (in millions), with orange indicating any observations removed**
107 **during quality control. Variable acronyms are as those described in Table 1. Country codes are as those listed in Table 7.**

108

Earth Syst. Sci. Data Discuss., https://doi.org/10.5194/essd-2018-39
Manuscript under review for journal Earth Syst. Sci. Data
Discussion started: 3 May 2018

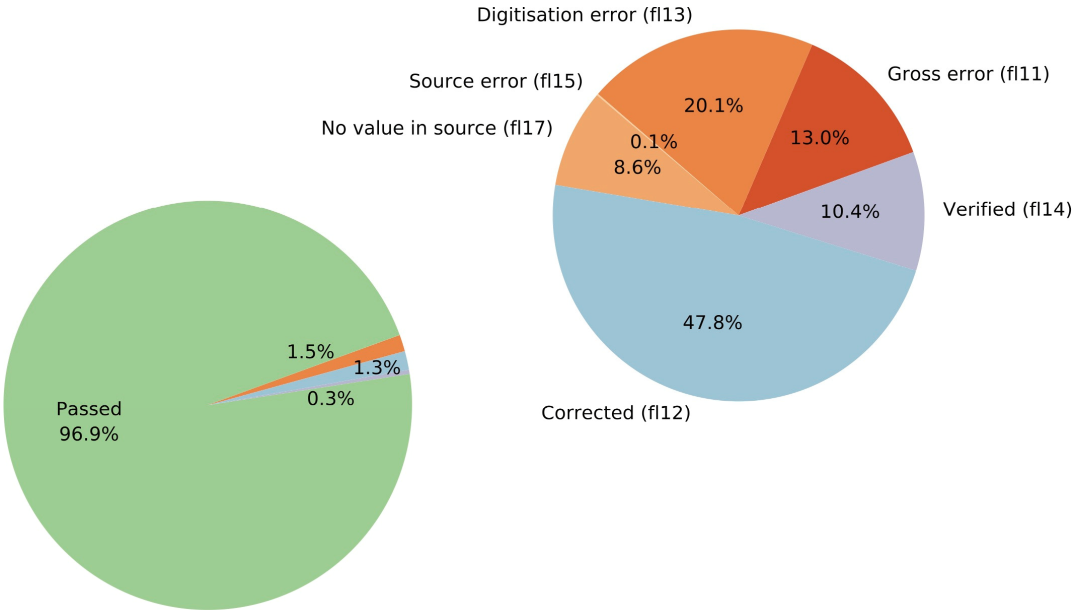


Figure 9: Percentages of flagged/not flagged values derived from SAQC application to the UERRA dataset. The green wedge represents data that passed SAQC; orange wedges represent values that have been removed (1.5 % of the total); blue wedges represent data that were corrected (1.3 % of the total) and purple wedges show the data that were verified (0.3 % of the total). Flag codes given are explained in Table 4.

Earth Syst. Sci. Data Discuss., https://doi.org/10.5194/essd-2018-39
Manuscript under review for journal Earth Syst. Sci. Data
Discussion started: 3 May 2018

Open Access Earth System Science Data Discussions

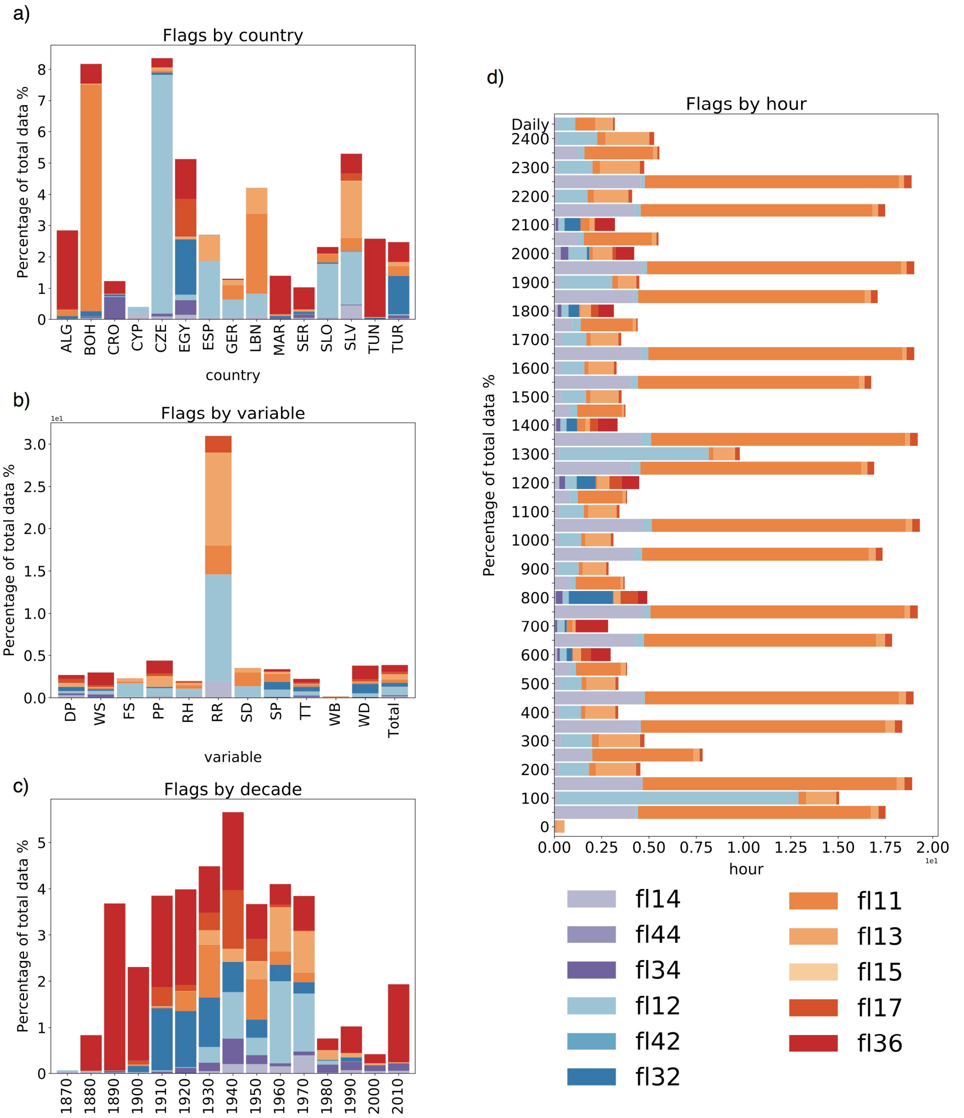


Figure 10: Total counts (in percentage) of error flags by countries (a), variables (b), observation times (c) and decades (d) derived from SAQC application to the UERRA dataset. Purple indicates values that were flagged but verified; blue indicates values that were flagged and corrected; and red and orange indicate values that were flagged and removed as errors. Variable acronyms are as those described in Table 1. Flag descriptions are given in Table 4.

Earth Syst. Sci. Data Discuss., https://doi.org/10.5194/essd-2018-39
Manuscript under review for journal Earth Syst. Sci. Data
Discussion started: 3 May 2018

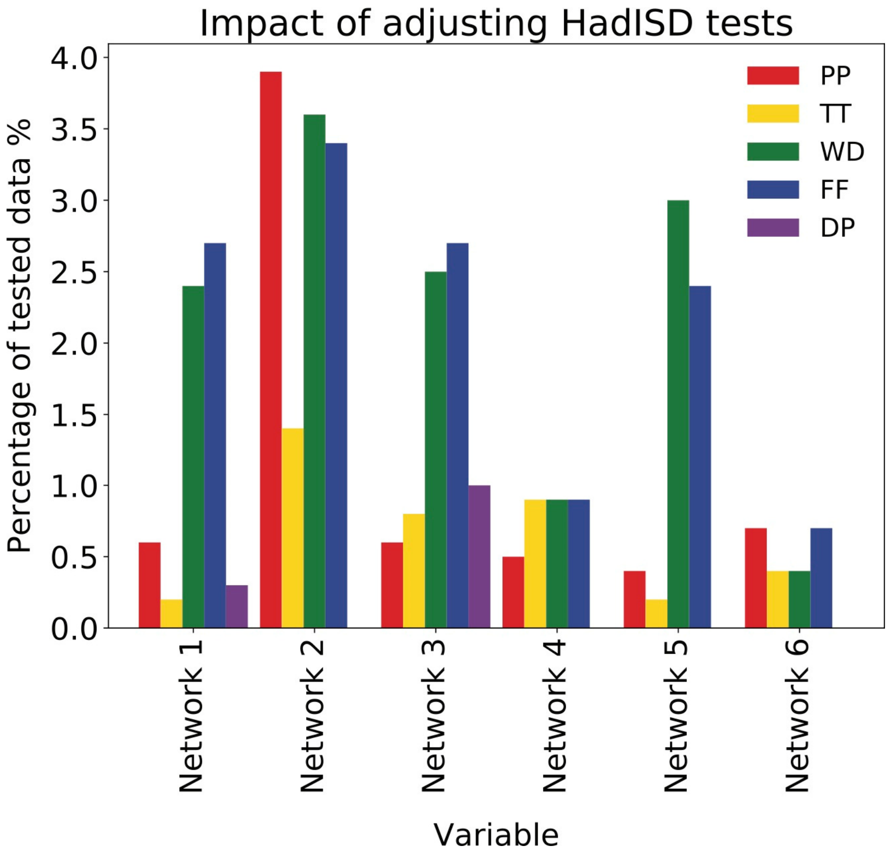


Figure 11: The percentage of values flagged within each network (see Table 6) tested using the HQC automatic procedure. Variable acronyms are as explained in the caption for Table 1, noting that not all variables were included in each network.

Earth Syst. Sci. Data Discuss., https://doi.org/10.5194/essd-2018-39
Manuscript under review for journal Earth Syst. Sci. Data
Discussion started: 3 May 2018

Open Access Earth System Science Data Discussions

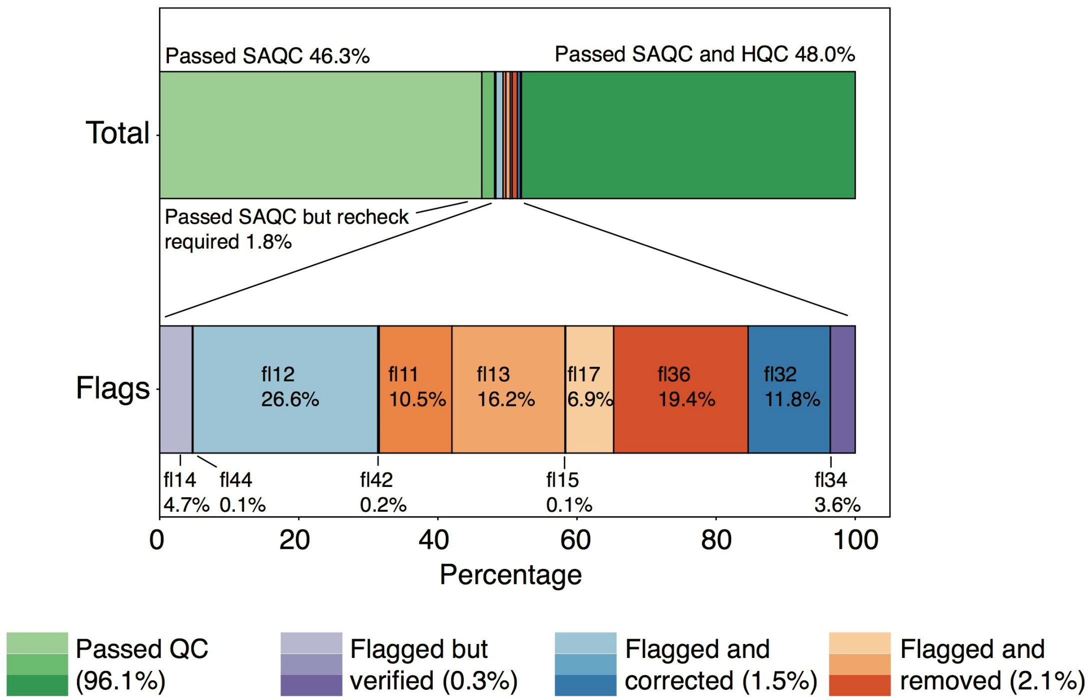


Figure 12: The percentage distribution of quality control flags in the UERRA dataset. Values that have passed QC are represented in green (QC flags fl10, fl40 and fl30); values that were flagged but verified as correct are shown in purple (fl14, fl44 and fl34); values that were flagged but corrected are shown in blue (fl12 fl42, fl32); and values that were flagged and removed are shown in orange (fl11, fl13, fl15, fl17 and fl36). The darkness of the colors indicates the level of QC applied for each flag. Lighter colors represent values that were only subjected to semi-automatic quality control (SAQC, fl codes that begin with 1 and 4), while darker colors indicate values subjected to both SAQC and spatial HQC procedures (fl codes that begin with 3). See Table 4 for additional flag details.