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]**, Manola Brunet**[1,3]**,
Enric Aguilar**[1]**, Mercè Castella**[1]**, Javier Sigro**[1]**, Ian Harris**[3]**, Per Unden**[4]**, and Phil Jones**[3,5]

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

Correspondence: 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 of such weather observations that have been collected by various organisations, many are yet to be transcribed into a useable format.

Under the auspices of the Uncertainties in Ensembles of Regional ReAnalyses (UERRA) project, we describe the compilation and development of a digital dataset of 8.8 million meteorological observations of essential climate variables (ECVs) rescued across the European and southern Mediterranean region. By presenting the entire chain of data preparation, from the identification of regions lacking in digitised sub-daily data and the location of original sources, through the digitisation 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 African 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 sub-daily dataset gives us an opportunity to share some suggestions for future data rescue projects.

All versions of the dataset, from the raw digitised data to data that have been quality controlled and converted to standard units, are available on PANGAEA: https://doi.org/10.1594/PANGAEA.886511 (Ashcroft et al., 2018).

1 Introduction

Digitising meteorological observations into a useable modern format is crucial for long-term climate monitoring and meteorological service development. High-quality 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 digitise land-based meteorological observations at national, regional and international scales. The Atmospheric Circulation Reconstructions over the Earth initiative (ACRE, Allan et al., 2011) coordinates climate data rescue across the globe, while other projects such as MEditerranean DAta REscue (MEDARE, www.omm.urv. cat/MEDARE/index.html, last access: 4 February 2018) and Historical Instrumental Climatological Surface Time Series Of The Greater Alpine Region (HISTALP, wwww.zamg.ac. at/histalp, last access: 6 May 2018) focus on particular regions (Auer et al., 2007; Brunet et al., 2014a, b). Additional initiatives on a national to regional scale, led by meteorological agencies (e.g. Kaspar et al., 2015) and research projects (e.g. Ashcroft et al., 2014; Brunet et al., 2006, 2014a), have also located and digitised meteorological 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 digitisation, because these variables are used to monitor changes in climate and the incidences of extreme weather events, both of which are important for the economic and agricultural sectors (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. We define sub-daily variables here as variables observed at least once a day, up to every half an hour. These data, rather than daily values or monthly averages, are necessary input for global and regional reanalysis products, which can greatly improve 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 2-year digitisation effort aimed at recovering sub-daily meteorological data across the European region. Our work formed part of Uncertainties in Ensembles of Regional ReAnalyses (UERRA, http://uerra.eu/, last access: 27 August 2018), 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 was the recovery of sub-daily surface meteorological observations to provide input to and assess the quality of future regional 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. We present the methods used to minimise errors in the digitisation process and the steps required to take the data from a disparate set of sources to a unified database. In Sect. 2 we explain how we identified target regions and likely sources for data rescue across Europe and the neighbouring southern Mediterranean region to maximise improvement in spatial and temporal coverage of existing data. 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 (QC) results in Sect. 4. Finally, we give some practical ideas for future data recovery projects based on our experiences with this particular project, as well as details about how to access the data.

2 Methods and materials

2.1 Identifying gaps in sub-daily data availability

The primary goal of the data rescue efforts within UERRA was to improve spatial coverage of input data for future regional gridded and reanalysis climate products over the European domain. Our aim was not to develop single, long-term data series for particular stations, but rather improve the availability of sub-daily observations anywhere that may be underrepresented in the current observational data used for European reanalysis products. This involved, 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 digitised observations.

To identify gaps in the available sub-daily climate record, we first conducted a visual examination of the data holdings of the International Surface Pressure Databank (ISPD, Cram et al., 2015) and the Koninklijk Nederlands Meteorologisch Instituut (KNMI) European Climate Assessment and Dataset (ECA&D: http://eca.knmi.nl/, last access: 12 Jan-
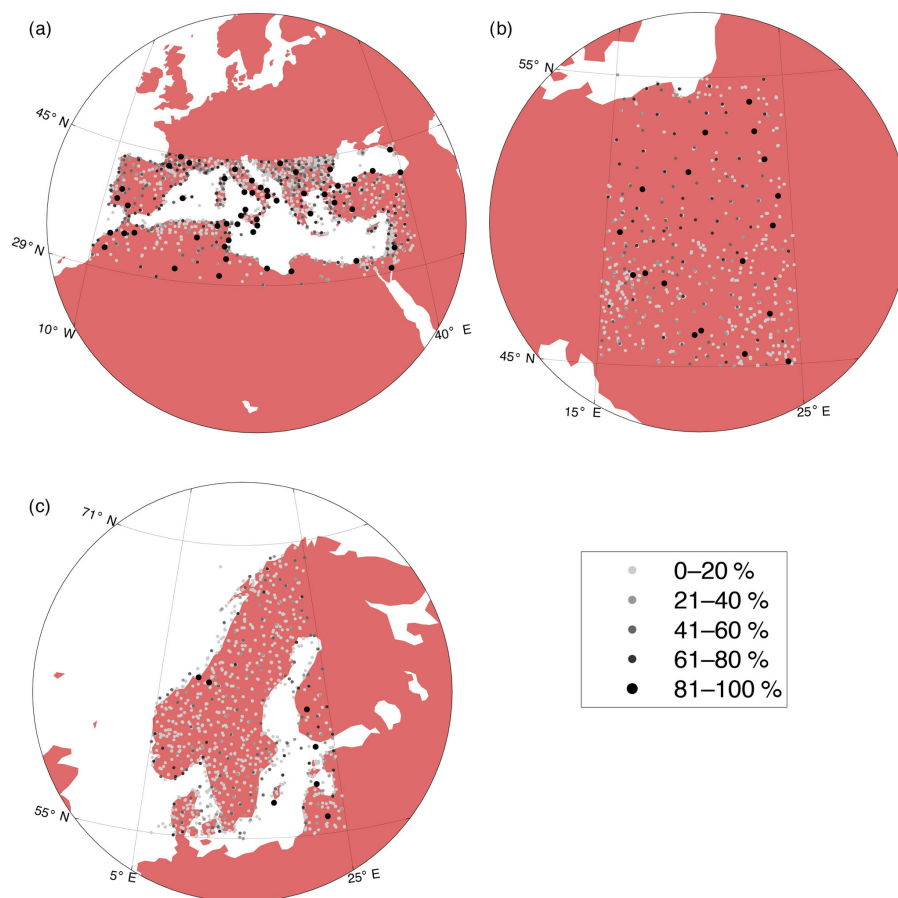
Figure 1. Stations with monthly mean sea level pressure data in MARS across the three identified regions of interest: **(a)** the Mediterranean, **(b)** eastern Europe and **(c)** Scandinavia. The shade and size of the symbols indicate the percentage of data available for 1957–2010.

uary 2018). These databanks provide station lists, regularly updated datasets and online visualisation tools, making it relatively straightforward to identify the regions lacking in sub-daily data. We also examined the holdings of the national climate data systems of countries whose data may not yet be in a multi-national repository. In particular, we checked the data available from the national climate data management systems of countries that had not been included in previous regional data rescue projects, namely the Romanian Meteorological Administration (NMA-RO) and the national meteorological and hydrological services (NMHSs) 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 identified the Mediterranean, eastern Europe and Scandinavia as three key sub-regions within the European sector as lacking in sub-daily data.

We then conducted an extensive examination of the data available for these regions within the Meteorological Archival and Retrieval System (MARS) at ECMWF. MARS is home to the primary data input for the current European reanalysis products available from ECMWF (e.g. Dahlgren et

al., 2016), 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. Interrogating the MARS holdings is not as straightforward as ISPD or ECA&D due to the extremely large number of data sources stored in the system and the registrations required, which is why we conducted our search in this order.

We focussed our search on the three data-sparse sub-regions in the post-1957 period, 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, last access: 22 July 2018). The variables of interest were several atmospheric and terrestrial essential climate variables (ECVs) as defined by the Global Climate Observing System (GCOS, World Meteorological Organization, 2015) that were identified as important for the development and verification of regional reanalyses: air temperature (TT), atmospheric pressure (sea level pressure, PP, and station level pressure, SP), wind speed (WS), wind direction (WD), relative humidity (RH), dew point temperature (DP), daily rainfall (RR), fresh snowfall (FS) and snow depth (SD).

The high percentage of stations with data for less than 60 % of the 1957–2010 period in MARS (Fig. 1) illustrates the lack of sub-daily observations in these sectors. Gaps are clear in the southern and eastern Mediterranean countries, Sweden, and Norway for the 1960s and 1970s (Table S1 in the Supplement), as well as across the Balkan region. The relatively dense spatial coverage of the stations with less than 60 % data coverage also suggests that sub-daily observations may 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 digitised sub-daily record available for Europe, we also needed to locate sources of undigitised sub-daily data. We undertook extensive consultation with NMHSs across the three identified regions of poor data coverage, in an attempt to identify and recover paper or scanned data sources suitable for digitisation. Priorities were given to data sources already available as scanned images, stations with data from the post-1957 period, and stations where the selected ECVs were recorded (see Sect. 2.1). Recovered precipitation observations from NMA-RO were digitised internally, and then provided to us in digitised quality-controlled format, using a similar quality control format to that used in this study (see Sect. 3.2). Discussion with the Norwegian and Swedish NMHSs uncovered data for these countries that had been digitised, but were not yet provided to international data repositories. Similarly, the Catalan Meteorological Service (MeteoCat), which has an open data policy, allowed their digitised data for the recent 1998–2015 period 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 further here. Political and financial difficulties prevented many other countries we contacted, particularly in northern Africa and the Balkans regions, from providing original data sources to us for digitisation.

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 NMHSs enabled us to identify valuable and previously undigitised data sources. From these sources, stations with minimal data available in MARS were selected for digitisation.

The World Meteorological Organization (WMO) Mediterranean data rescue initiative MEDARE and the precursor project to UERRA, the European Reanalysis and Observations for Monitoring project (EURO4M, http://www.euro4m.eu/, last access: 12 January 2018), located key records of data for the Middle Eastern, Balkan and south-ern Mediterranean regions from the Serbian NMHS online climatological scanned repository (http://www.hidmet.gov.rs/ciril/meteorologija/klimatologija_godisnjaci.php, last access: 4 June 2018), 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, last access: 8 August 2018), the British Atmospheric Data Centre (BADC, http://badc.nerc.ac.uk/browse/badc/corral/images/metobs, last access: 8 August 2018), and other national meteorological services (see Brunet et al., 2014a, b for details). Daily maximum and minimum temperature, precipitation, and sub-daily atmospheric air pressure observations from some of these sources were digitised 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 undigitised values from these sources (Brunet et al., 2014b).

Table 1 provides detail of the data sources identified for digitisation, 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 digitised 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 Digitising method

Once data sources had been identified and catalogued, a group of 11 digitisers were employed for 15 h a week over a 2-year period to digitise the data. The digitisation team was made up of undergraduate and postgraduate geography students from the Universitat Rovira i Virgili (URV), who all had some knowledge of meteorological variables and European climate. The digitisers 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 digitisers received initial training sessions, online instructions and monthly in-person meetings to discuss issues and introduce new digitisation tasks. Digitisation was done using a "key as you see" method, meaning that the digitisers typed the values they could read in the data images, rather than using any coding system. This follows standard best practice outlined by the WMO (2016). Clear, unambiguous errors in the data sources were generally retained by the digitisers and recorded in station metadata files, which were later used when quality controlling the data (see Sect. 3). If a digitiser could not read a value due to poor handwriting or scan-

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 (Ashcroft et al., 2018). Variables given in bold in the variable column have been digitised as part of this project: not all available variables and time periods were digitised in this project due to time and funding constraints. Each source can be found at ftp://130.206.36.123 (last access: 4 June 2018), u: C3_UERRA, p: c3uerra17, folder: C3_UERRA_datasources_images, where the sources are 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), maximum temperature (Tmax) and minimum temperature (Tmin).

| Source code and name	Country (or countries) covered	Time period covered (continuous)	Source provider	Primary or secondary	Typed or hand written	Surface variables	Details
So04: Bulletin Météorologique de l'Algérie	Algeria	1877–1938	NOAA-CDMP	Secondary	Hand-written	**TT, PP, WS, WD,** Cloud, weather conditions, 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 1 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-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.
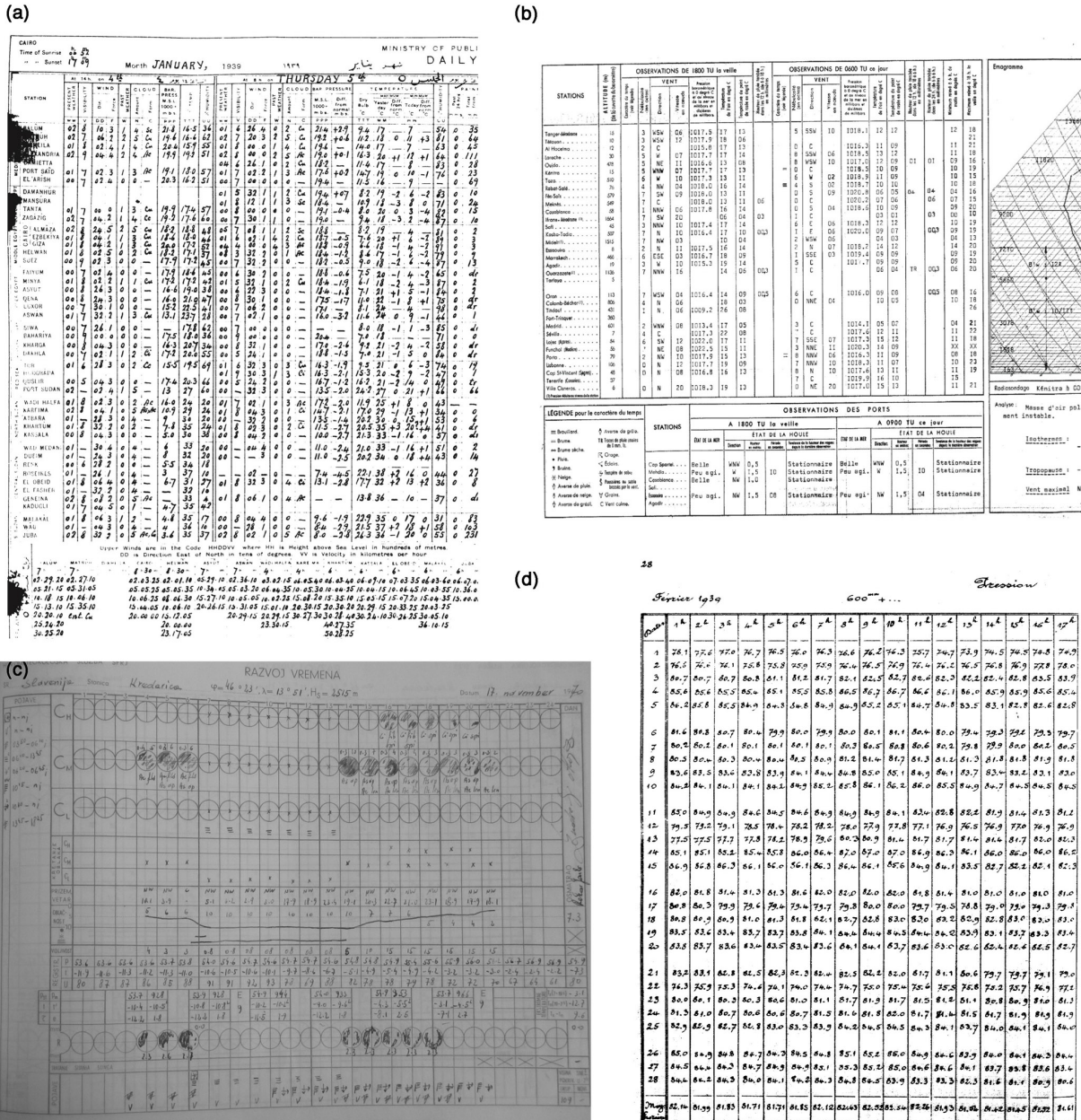
Figure 2. Examples of the different data source formats found for digitisation: **(a)** Egypt, 1939, where each row is observations from a different station on 1 day; **(b)** Morocco, 1968, where each row is observations from a different station on 1 day; **(c)** Kredarica, Slovenia, 1970, where each row is observations of a different variable for one station on 1 day; **(d)** Ksara, Lebanon, 1939, where each row is atmospheric pressure data for 1 day at one station. Data images are available online the Universitat Rovira i Virgili's Centre for Climate Change (see Sect. 6).

ning issues, they represented it by a value of -88.8, while missing values were set to -99.9.

Budget constraints made it unfeasible to use double-keying, a suggested method of improving digitised data quality where the same data are transcribed twice (Brönnimann et al., 2006; World Meteorological Organization, 2016). 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 to each data source made these options unfeasible. However, the digitisers were trained in self-assessment techniques aimed at reducing data errors. Digitisers 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
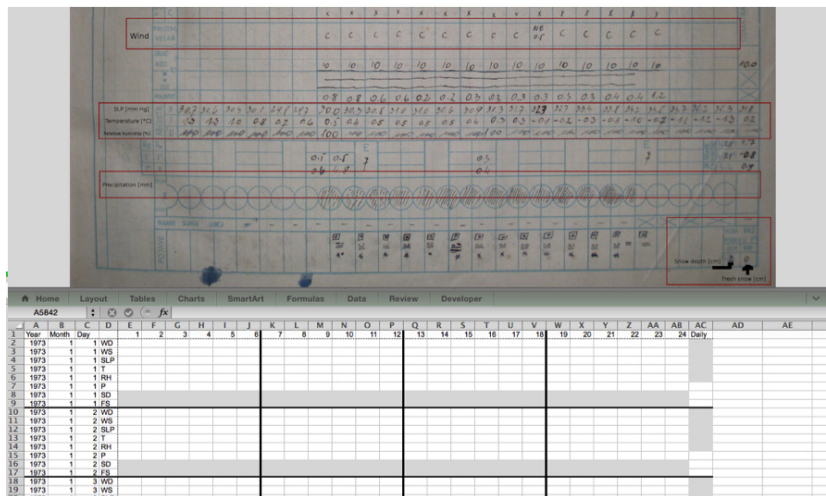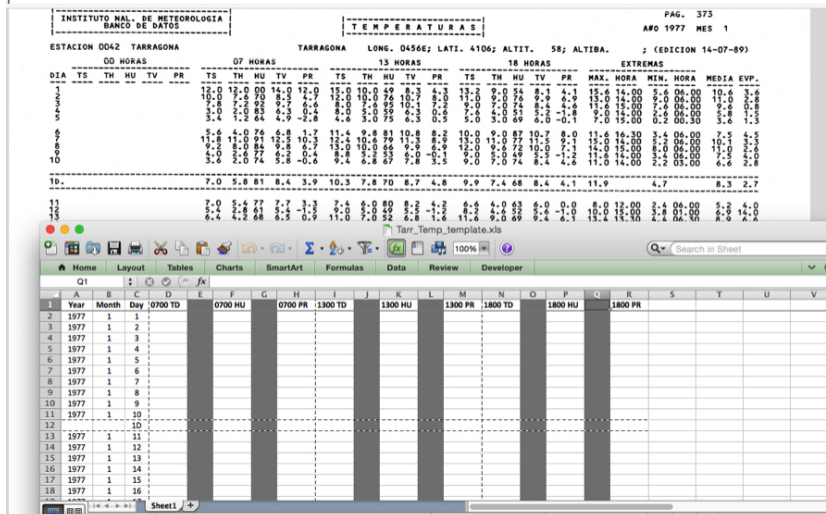
Figure 3. Examples of the templates used in data digitisation. Shaded rows and columns in the templates represent data that are not to be digitised. **(a)** The template for the Slovenian data sources picks out the rows that require digitising: wind direction (WD), wind speed (WS), atmospheric pressure (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).

other variations in the data source, such as repeated pages in the scanned file or temporary changes in the table structure. Where data sources included monthly totals and summaries, digitisers were also instructed to calculate these values from their daily transcribed data, to check accuracy.

The data sources were in a number of different formats. The two main formats were 1 month (or day) to a page for a single station, and 1 day to a page for a network of stations. Depending on the source structure, each digitiser was in charge of digitising 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). English and Catalan translations of the relevant column and row headings were provided to the digitisers for each source, as well as the various wind strength scales (see Sect. 2.4).

In several cases, not all of the data on a sheet were required to be digitised, as they had already been transcribed as part of EURO4M and MEDARE. To help digitisers with the complex layout of the source images, templates were developed in Microsoft Excel for some sources that were as close as

Table 2. List of conversions applied to digitised 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 S3.

	Original units	Final units	Details
Wind speed conversions	Beaufort scale	m s^{-1}	Replacement of x with y using the following map: $0 = 0.0$, $1 = 1.0$, $2 = 2.6$, $3 = 4.6$, $4 = 6.7$, $5 = 9.3$, $6 = 12.3$, $7 = 15.4$, $8 = 19$, $9 = 22.6$, $10 = 26.8$, $11 = 30.9$, $12 = 35$, from WMO Code 1100 (Da Silva et al., 1995)
	Turkish 17-point power scale	m s^{-1}	Replacement of x with y using the following map: $0 = 0.0$, $1 = 0.9$, $2 = 2.4$, $3 = 4.4$, $4 = 6.7$, $5 = 9.3$, $6 = 12.3$, $7 = 15.5$, $8 = 18.9$, $9 = 22.6$, $10 = 26.4$, $11 = 30.5$, $12 = 34.8$, $13 = 39.2$, $14 = 43.8$, $15 = 48.6$, $16 = 53.5$, $17 = 58.6$ taken from data source (average of wind range used)
	9-point power scale	m s^{-1}	Replacement of x with y using the following map: $0 = 0.0$, $1 = 1.0$, $2 = 2.6$, $3 = 4.6$, $4 = 6.7$, $5 = 9.3$, $6 = 12.3$, $7 = 15.4$, $8 = 19$, $9 = 28.8$, from 1931 French instruction book (Angot, 1931)
	km h^{-1}	m s^{-1}	$y = x/3.6$ rounded to 1 decimal place
	knots	m s^{-1}	$y = x \times 0.514444$ rounded to 1 decimal place
Wind direction conversions	16-point compass scale	degrees	Replacement of x with y using the following map: $C = 361$, $NNE = 22.5$, $NE = 45$, $ENE = 67.5$, $E = 90$, $ESE = 112.5$, $SE = 135$, $SSE = 157.5$, $S = 180$, $SSW = 202.5$, $SW = 225$, $WSW = 247.5$, $W = 270$, $WNW = 292.5$, $NW = 315$, $NNW = 337.5$, $N = 360$
	32-point direction scale	degrees	$y = x \times 11.25$
	degrees / 10	degrees	$y = x/10$
Pressure conversions	mmhg	hPa	$y = x \times 1.33224$ rounded to 1 decimal place
	hpa × 10	hPa	$y = x/10$
Temperature conversions	°F	°C	$y = (x - 32) \times (5/9)$ rounded to 1 decimal place

possible to the format of the original data source (see Fig. 3 for several examples). Borders and shading within the files were used to help the digitiser 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. While the development of templates was not always possible due to time constraints, templates were used for all sources with very high-resolution data (e.g. observations every hour, see Table 1).

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

2.4 Conversion to standard units

While visual quality control and assessment were 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 in particular needed to be converted from millimetres 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. In most cases, only the station height information could be located, but any changes identified were recorded in the coordinates accompanying the final dataset.
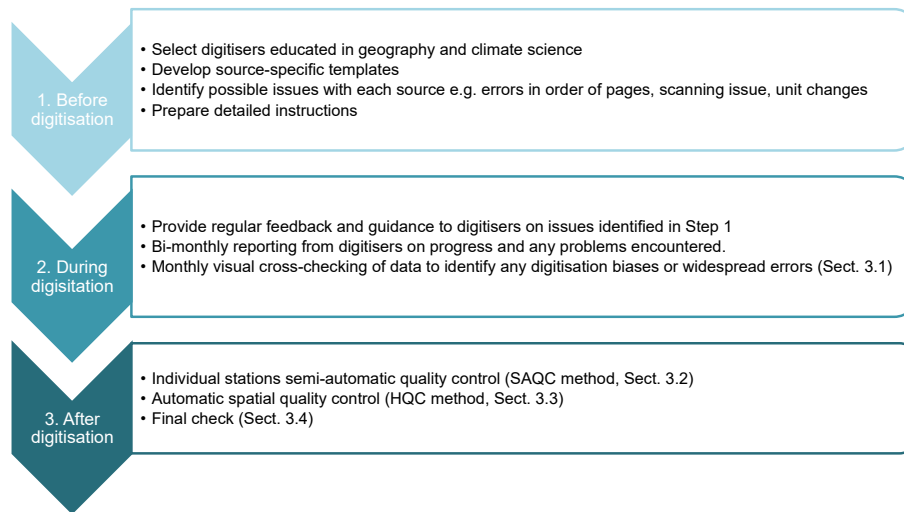
Figure 4. A schematic of the quality assurance and quality control procedures used in the development of the dataset.

3   Quality assessment of digitised data

Quality control procedures are crucial to identify non-systematic errors or shed light on systematic biases in a time series. This is particularly the case for daily or sub-daily data, as these observations are used in the calculation of monthly and annual means. Errors can occur as a result of issues with original sources, the method of data collection, transcription in the original source or the digitisation process.

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 well-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 events (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 by cross-checking all observations against the original source was not feasible. However, a completely automated procedure that tests data against that of neighbouring values, such as that used for global databases (Dunn et al., 2012), would also be sub-optimal, as the digitised 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. A different version of the dataset was produced after each step, enabling users to ultimately access the original data, as well as data that had undergone one or two rounds of quality testing.

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 digitisation 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 digitisers. During and after digitisation, the digitised 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 homogenisation is not included in this QC procedure. Although the homogenisation 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 under development within the research community (Venema et al., 2012).

3.1   Visual cross-checking

A selection of values uploaded by digitisers were systematically compared to the original source images by postgraduate researchers and other digitisers at the Centre for Climate Change at URV familiar with the sources. The aims of these initial visual cross-checks was to provide timely feedback to the digitisers if common digitisation 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 digitisers who were having trouble with their tasks and needed extra assistance.

For every fourth year of data, 2 or 3 days of observations were selected at three monthly intervals for visual cross-checking with the original source. This was completed for data from all sources. 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 digitised data, we felt it was sufficient to identify the general quality of work done by individual digitisers 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, digitisers were asked to repeat their work.

3.2　Individual station quality control (SAQC method)

After the basic visual cross-checks, the digitised 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. Data were also examined in their original temporal resolution, and not converted to daily averages, as averaging the sub-daily values would make it difficult to identify the erroneous value. Statistical quality control was conducted using a semi-automatic quality control (SAQC) procedure (Universitat Rovira i Virgili, 2014). The SAQC method was 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 was modified for our dataset to enable more manual examination of the resultant flags. Full details of the procedure, the relevant software and instructions for use are available from A.Q.C. Software menu at http://www.c3.urv.cat/softdata.php (last access: 8 August 2018).

SAQC comprised of three separate programs that can be applied to the data at their original time resolution 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 data. 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*. These are 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*. These are QC tests that detect and flag those values considered outliers with respect to their own defined upper and lower limits (climatic outliers, bivariate comparisons, monthly mean of absolute increments, and unusual distribution of values tests).

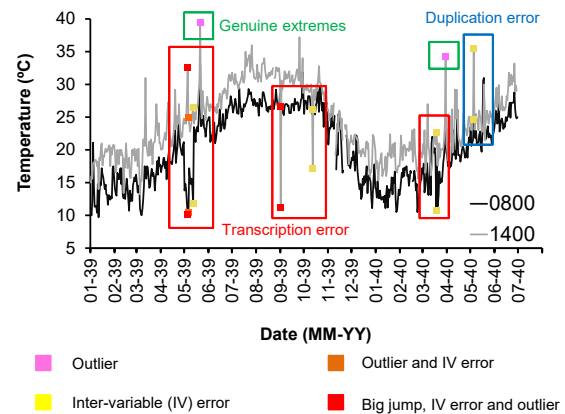- *Inter-variable check*. These are QC tests which detect and flag inconsistencies between associated elements

Figure 5. Air temperature evolution (in °C) in Port Said station (Egypt) taken at 08:00 (in black) and 14:00 (in grey) for the period 1939–1940. Different errors flagged by SAQC are marked with solid coloured squares: an outlier (pink); outlier and inter-variable (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: values identified as transcription errors are outlined by a red border, values flagged due to a data duplication error are outlined in blue, and values that were found to be valid extremes are outlined in green. Values found to be errors were corrected and given a flag of fl12 in the quality-controlled version of the datasets, and values found to be correct were retained and given a flag of fl14.

within each record (interval and DP/FS/SD inconsistency test, RH/DP/TT comparison tests, precipitation and snow totals test).

- *Temporal coherency*. These are 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 produced a list of values flagged by each test at each station. The combined key results were then manually cross-referenced against the original source data, and corrected or removed from the quality-controlled version of the dataset. 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 08:00 and 14:00 local time 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 flagged for assessment. Many of the legitimate errors identified by

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 (last access: 8 August 2018).
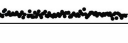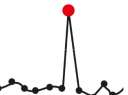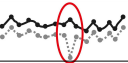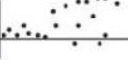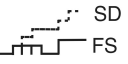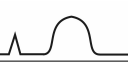
	SAQC test	Brief description	Variables
APRIL 31	Date order check	Detect erroneous calendar date order	TT/DP/RH/WD/WS/RR/PP
	Data repetitions	Flag repeated entire months	TT/DP/RH/WD/WS/PP
T < −91°C T > 58°C	Unrealistic values	Flag values outside world record limits and physically impossible values	TT/DP/RH/WD/WS/PP
	Climatic outliers	Flag values out of an established threshold	TT/DP/RH/WS/PP/RR/FS/SD
	Duplicate values	Detect at least 30 identical consecutive values	TT/DP/RH/WD/WS
	Big jumps and sharp spikes	Large differences between adjacent values	TT/DP
	Bivariate outliers	Differences between adjacent values that are larger than the bivariate distribution	PP
ΔT≈ΔW ≈ RH?	Inter-variable inconsistency	Flag internal inconsistencies among variables	TT/RH/WS
	DP inconsistency	Flag differences between observed and calculated DP	TT/RH/DP
	Monthly mean of absolute increments	Flag all values when mean monthly increments below/above of the climatic normal increment	TT/RH/WD/WS
	Irregular temporal evolution	Flag values that show unexpected temporal evolution	TT/RH/WS
mmHg -> hPa	Unit changes	Automatic unit changes from millimetres of mercury (mmHg) to hectopascals (hPa)	PP
	Unusual distribution of values	Flag values where the distribution in each month includes a secondary peak	TT, DP, VV
6 + 12 + 18 + 24 = Daily	Precipitation totals	Flag values when sum of sub-daily RR data does not equal daily RR total	RR
SD FS	Snow totals	Flag values when sum of fresh snow <= total snow depth	FS/SD
	Summer snow	Flag snowfall between May and October	FS/SD
34.!7 °C	Non-numeric values	Flag non-numeric values	PP/RR/FS/SD
SD FS	FS/SD inconsistency	Flag total SD that increases without a FS falland or decreases with a FS value/ flag FS that is not accompanied by SD	FS/SD
34.!7 °C	Non-numeric values	Flag non-numeric values	PP/RR/FS/SD

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

Flag	Flag description	Expert decision
fl10	Passed all SAQC tests	Retained
fl11	Identified as suspect and removed due to gross digitiser error	Removed
fl12	Identified as suspect, found to be a digitisation error, corrected	Corrected
fl13	Identified as suspect, found to be a digitisation 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 digitisation 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

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 and automatic quality assurance (HQC method)

The final QC procedure consisted of subjecting data from neighbouring stations to spatial quality control tests, as well as rerunning several individual station checks in a fully automated way as a second-round check for gross errors that may have slipped through SAQC. Only data that had been checked by visual means and SAQC were subjected to this procedure and as with SAQC, the data were examined in their original temporal format to avoid removing valid data. This QC process (Hadley quality control, or HQC) was conducted using an adapted version of the procedure used in the development of the UK Met Office Hadley Centre Global Sub-Daily Station Observations dataset (HadISD v2.0.1.2016p; Dunn et al., 2012, 2016). Due to time constraints, only data digitised as part of this project were used in the spatial quality assessment, although future work could make use of the existing HadISD dataset as a reference network.

Automatically running HQC with the standard thresholds used in the development of the global HadISD dataset led to a large number of false positive flags being identified (Fig. 6), as the rescued dataset had low spatial coverage and included observations taken at inconsistent times, often converted from units with coarse resolution. To reduce the number of false positive flags and increase the number of stations that could be checked, some of the HadISD tests were adapted (Table S2). The minimum number of neighbouring stations required for HQC testing was reduced from 10 to 5, and the percentage of non-missing observations per month
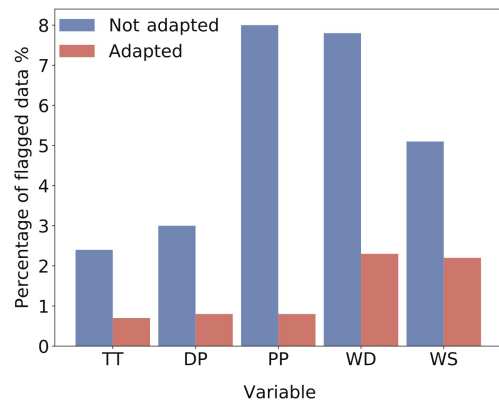

Figure 6. Percentage of flagged values using the standard QC tests developed for the UK Met Office Hadley Centre Global Sub-Daily Station Observations (HadISD), and the percentage of values flagged using HadISD tests specifically adapted for this dataset (HQC). The variable acronyms are the same as those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

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.

Data from each country 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 5), but it was not possible to include all stations, periods, variables and observing times. The heterogeneous characteristics of the dataset, the high spatial separation and irregular distribution of the stations, and the inconsistent coverage of the variables in-

Table 5. The networks used in the spatial and automatic quality control analysis (HQC), including the period, variables and observing times examined. Note that not all observing times were examined in HQC due to neighbouring data availability.

Network	Countries	Number of stations used	Period checked by HQC	Variables checked by HQC	Observing times examined	% of complete data checked
Morocco	Morocco	8	1953–1968	PP, TT, WD, FF DP	6 or 7 h/18 h (2)	94.5 %
North Africa	Algeria, Tunisia	25	1886–1938	PP, TT, WD, FF	7 h (1)	91.5 %
Egypt	Egypt	21	1907–1957	PP, TT, WD, FF DP	6 h or 8 h/12 h or 14 h/18 h or 20 h (4)	69.6 %
Turkey	Turkey	25	1962–1971	PP, TT, WD, FF	7 h/14 h/21 h (3)	100 %
Central Balkans	Slovakia, Croatia, Bosnia-Herzegovina, Serbia	10	1950–2012	PP, TT, WD, FF	7 h/14 h/21 h (3)	72.7 %
Central Europe	Slovenia, Czech Republic, Germany (2 stations)	11	1948–1968	PP, TT, WD, FF	7 h/14 h/21 h (3)	46.99 %
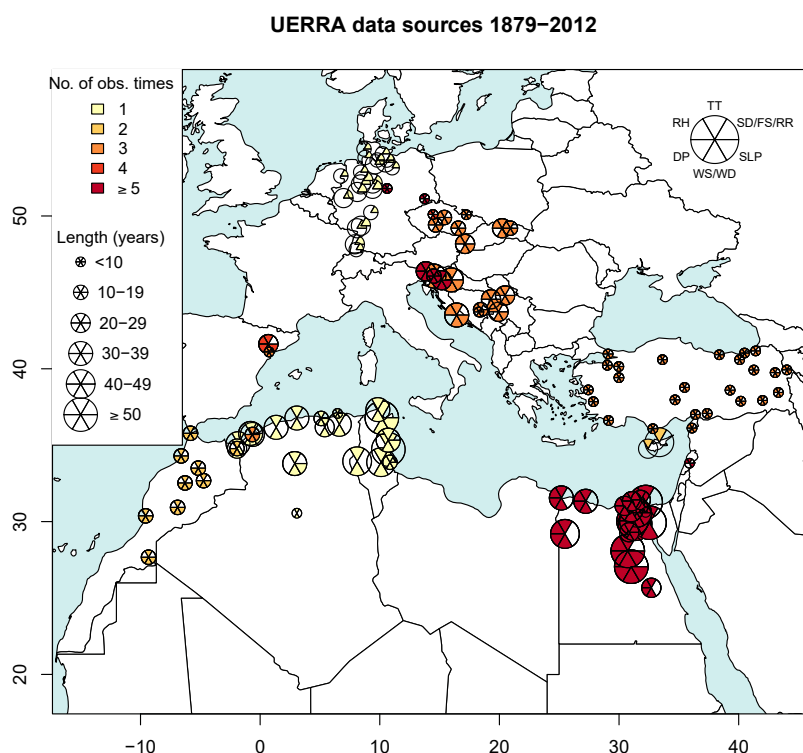
Figure 7. Spatial coverage of 8.8 million observations digitised, showing the station locations. The approximate length of the record at each station is indicated by the size of the pie symbol; the number of observations per day is represented by the colour of the pie pieces; and the different variables available at each station are indicated by which wedges are shaded based on the legend in the top right corner. Variable acronyms are as those described in the caption to Table 1, apart from SLP, which represents station and sea level pressure.

cluded 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 distance between the stations of neighbouring countries. We were also unable to

automatically analyse fresh snow and snow depth, precipitation, or relative humidity data, as the HadISD QC does not assess these variables as raw input. Moreover, several stations (such as those in Germany and Slovenia, the central Europe network in Table 5) provided hourly data, but there were not enough neighbouring stations with sufficiently high

Table 6. Summary of stations digitised as part of this 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 digitised 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 digitised	Total after QC	Percentage of data removed in SAQC
Algeria	ALG	21	1877–1968	DP, WS, WD, SLP, TT	4	684 114	665 369	2.74
Bosnia-Herzegovina	BOH	2	1953–1984	WS, WD, PP, RH, TT	3	125 831	115 894	7.90
Croatia	CRO	2	1949–1984	WS, WD, PP, RH, TT	3	391 789	390 141	0.42
Cyprus	CYP	2	1881–1922	TT	2	45 070	45 068	0.00
Czech Republic	CZE	7	1948–1968	WS, WD, PP, RH, TT	6	379 582	377 843	0.46
Egypt	EGY	18	1907–1957	DP, WS, WD, PP, RH, TT	6	1 371 436	1 336 281	2.56
Germany	GER	23	1958–1978	WS, WD, FS, PP, RH, RR, SD, TT, WB, DP	Up to 24	697 308	692 750	0.65
Lebanon	LBN	1	1930–1939	PP, RH, TT	24	262 944	254 044	3.38
Morocco	MAR	8	1910–1968	DP, WS, WD, PP, TT	4	340 563	336 170	1.29
Serbia	SER	3	1949–2012	WS, WD, PP, RH, TT	3	358 898	356 058	0.79
Slovak Republic	SLO	2	1940–1967	WS, WD, PP, RH, TT	6	248 751	247 541	0.49
Slovenia	SLV	3	1950–1978	DP, WS, WD, FS, PP, RH, RR, SD, TT	Up to 24	2 507 878	2 437 163	2.82
Spain	ESP	5	1954–1984	WS, WD, PP, RH, TT, DP	5	194 274	192 670	0.83
Tunisia	TUN	5	1886–1938	WS, WD, PP, TT	1	174 900	170 480	2.53
Turkey	TUR	25	1962–1971	WS, WD, PP, RH, TT	3	1 028 898	1 017 871	1.07
Total		127				8 812 236	8 627 338	2.00

temporal resolution to allow for more than a subset of observing times per day to be checked.

4 Results

4.1 Spatial and temporal data distribution

A total of 8.8 million observations were digitised from 127 stations in 15 countries (Tables 6 and S3). Long records (> 30 years) of many variables were successfully recovered from stations in Egypt, Tunisia and Algeria, although only the Egyptian stations provided observations more than once a day (Fig. 7). Shorter but more widespread observations were rescued across Morocco, Turkey and 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 observations; this is because very early Tunisian and Egyptian wind speed observations were qualitative (e.g. light, moderate) and were not digitised. 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 % of the full dataset combined) were also recovered from Germany and Slovenia from as early as the 1950s, representing a significant increase in snow observations across the region.
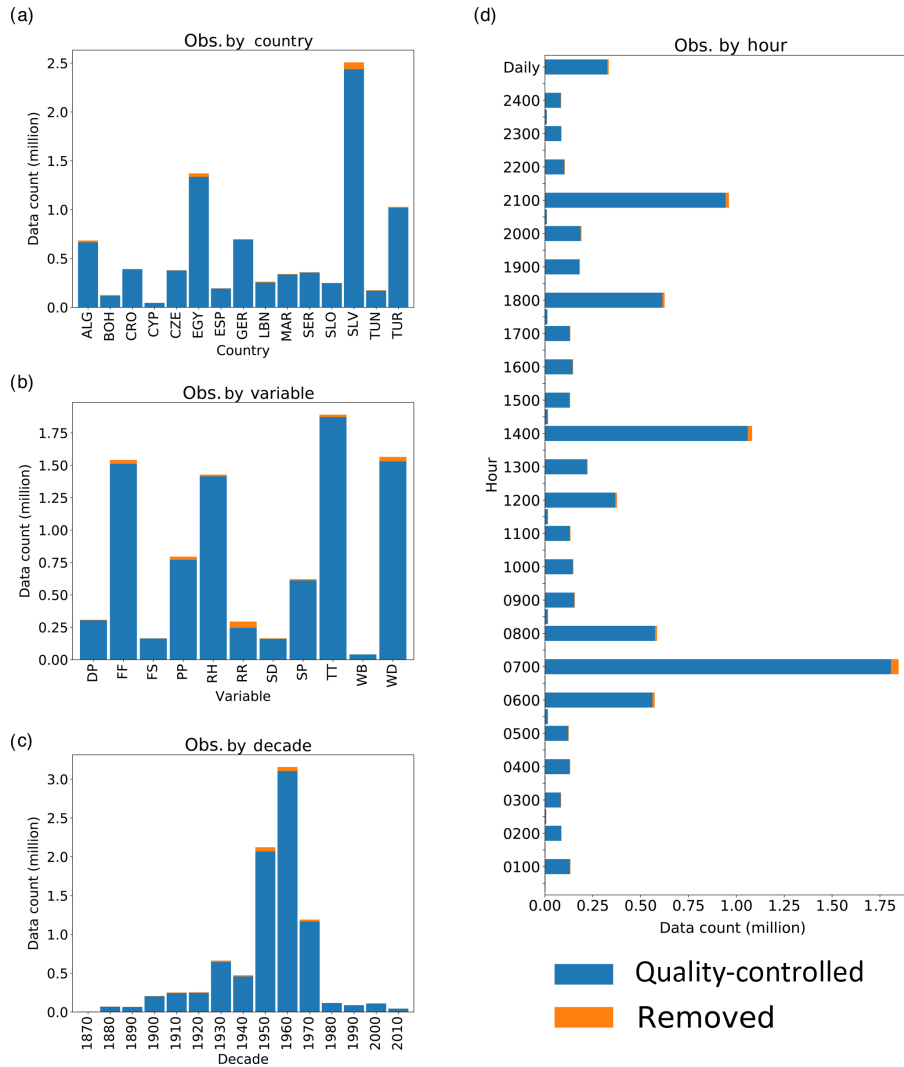
Figure 8. Distribution of the digitised observations by (a) country, (b) variable, (c) decade and (d) hour of observation. The length of each bar shows the number of observations digitised (in millions), with orange indicating any observations flagged and removed during SAQC. Variable acronyms are as those described in Table 1. Country codes are as those listed in Table 6.

Due to the temporal coverage of the Slovenian data (1950–1978), as well as the dedicated focus of the UERRA project on post-1957 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 07:00, 14:00 and 21:00, reflecting standard observing practices over the European region in the 20th century. Tunisian observations were only available for 07:00, and for many other countries where observations were only available once a day in the early part of the record, these observations were also inevitably in the morning. Two German

stations included a small number of half-hourly observations (Fig. 8d).

4.2 Semi-automatic quality control (SAQC) results

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 dataset using SAQC (Fig. 9).

Flagging correct values (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
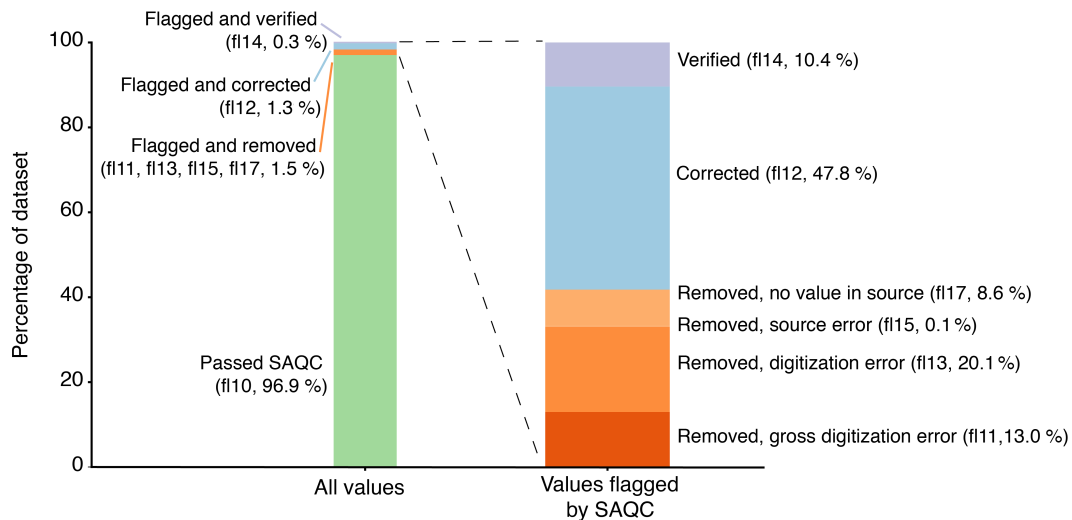
Figure 9. Percentages of flagged and not flagged values derived from SAQC application to this dataset. Panel (**a**) shows all datasets, while (**b**) breaks down data that were flagged as possible errors by SAQC. Flag codes given are explained in Table 4.

removed from the quality-controlled version of the dataset due to errors in the source image or issues with the readability of the original values. This includes observations recorded as -88.8 by digitisers (hard to read, see Sect. 2.3). Over 27 000, or 0.3 % of the total number of observations, were flagged but then found to be correct after examination.

Despite being among the countries with the smallest number of observations, the largest percentages of flagged values found were for Bosnia-Herzegovina and the Czech Republic ($\sim$ 8 % of the total number of data digitised, Fig. 10a). For Bosnia-Herzegovina a large section of observations from one station was given a flag of fl11 and removed due to an extensive digitiser error that could not be reconciled. A digitisation error in the Czech Republic observations was able to be corrected by shifting data by 1 day, resulting in a large number of fl12 flags (corrected based on original source). The handwritten nature of the Czech data, together with the absence of data templates (only used in Slovenian, Spanish 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), which was only available for Slovenian stations. The high number of precipitation flags is due to two factors. Firstly, several digitisers 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 3 or 4 times a day. Reporting no rainfall as missing data could significantly affect any future analysis of rainfall frequency using

these data, and so these values were corrected, resulting in a number of fl12 (corrected based on original source) flags. 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 07:00–07:00 total. The 6-hourly observations from the same stations were quality-controlled 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, and given a flag of fl15 (removed due to source error).

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 some original source files were ordered incorrectly. This resulted in a number of values being ascribed to the wrong date. 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 during these times (compare Fig. 8d and 10d).

4.3 Spatial quality control results (HQC)

In total about 64 000 values were flagged and subsequently removed by HQC, around 0.7 % of the total dataset. Temperature was the variable with the smallest number of flagged values overall by HQC, with the exception of the northern African network, where data source resolution and the high number of missing values caused HQC to flag and remove extra values (Fig. 11). The variable with the highest propor-
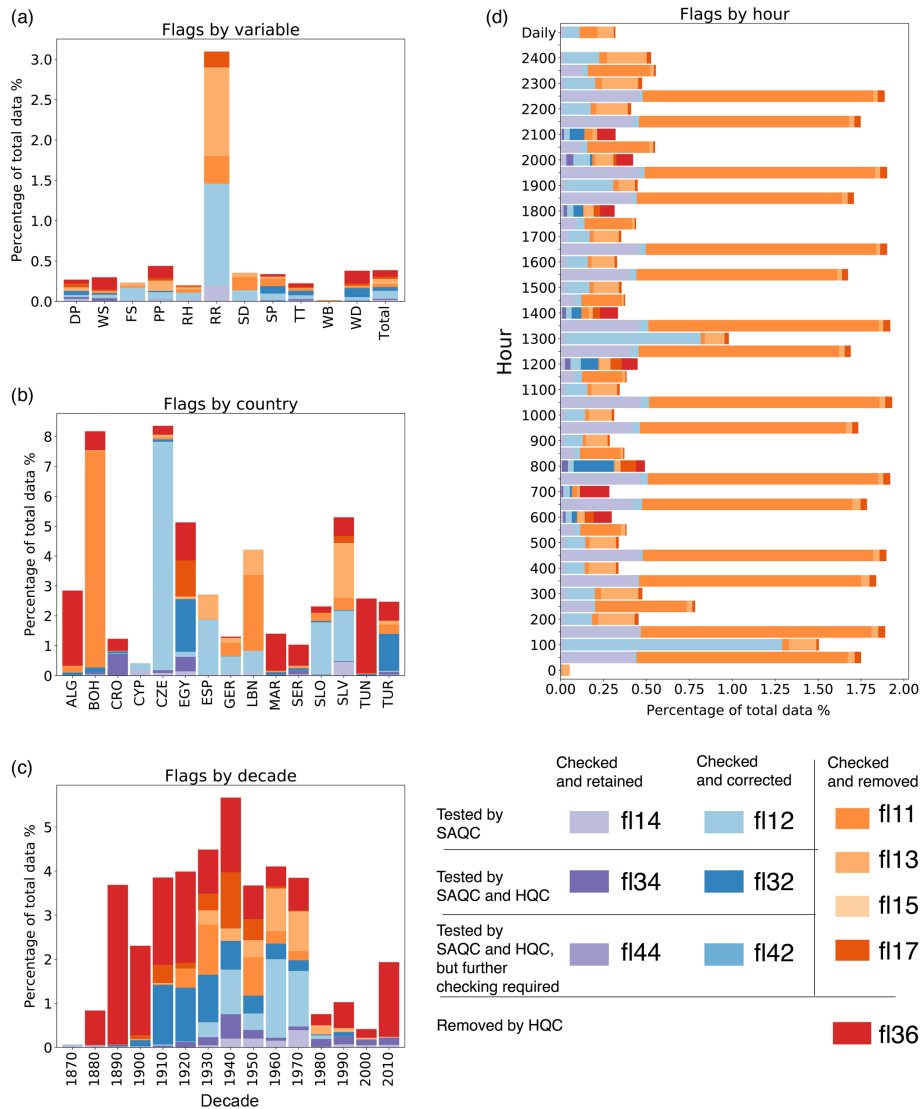
Figure 10. Total counts (in percentage) of error flags by countries **(a)**, variables **(b)**, observation times **(c)** and decades **(d)** derived from SAQC application to the dataset. Purple indicates values that were flagged but verified; blue indicates values that were flagged and corrected; and 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.

tion of flagged values in the northern African network was sea level pressure.

Given the automatic nature of the HQC tests, all values flagged by this step were removed from the final version of the dataset and given a flag of fl36. 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). For example, observations which were corrected or verified in the SAQC round of testing and given an initial flag of fl12 or fl14 but passed the HQC procedure had a final flag of fl32 or fl34, ensuring that information from both rounds of QC were retained.

While the HQC tests were unable to be applied to all of the observations, these results are similar to the findings of the HadISD spatial QC analyses (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 tails of atmospheric behaviour which can have the largest societal impact. These percentages of flagged values are similar to those identified by Brönnimann et al. (2006), who found transcription error rates of

0.2 % to 3 % for hourly temperature and upper air observations.

4.4   Additional digitisation quality assurance checks

In the final data check, a small conversion problem was detected with the atmospheric 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 marked with a prefix of "4" rather than "1" (subjected to SAQC) or "3" (subjected to SAQC and HQC) in the final dataset, to signify that additional QC may be required by future users.

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

4.4.1   Zagazig, Egypt, 1932

The 08:00 WD, WS and RH data for Zagazig, Egypt, in 1932 were digitised twice by different digitisers: once using a template where every station on a page was digitised together, and once without a template but extracting only data from Zagazig from each source page (see Fig. 2a). A total of 70 disagreements were found out of 1098 values, just over 6 % of the overlapping data. Interestingly, all but one disagreement was due to errors in the data digitised using the template. A total of eight values were entered into an incorrect row, six values were misread by the digitiser 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 digitiser worked through the source by digitising 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.

4.4.2   Egypt 1931

Two digitisers inadvertently digitised 08:00 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 19 800 values digitised. Checking the differences with the original source images revealed that 79 % were errors from one digitiser, and 21 % from the second digitiser. 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).
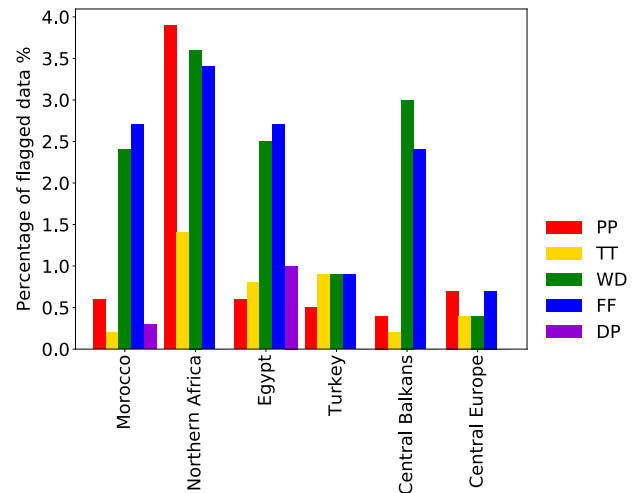

Figure 11. The percentage of values flagged within each network (see Table 5) tested using the HQC automatic procedure. Variable acronyms are as explained in the caption for Table 1, noting that not all variables were included in each network.

These two Egyptian examples highlight a number of key issues with data digitisation. The first is that the reliability of digitised data depends to a large extent on the reliability of the person digitising those data. In both cases there was a clear separation between the two digitisers, even though (in the case of Egypt 1931) both digitisers used the same method. The second is that templates created without input from digitisers may not always achieve the best result. Indeed, follow-up surveys with the digitisers suggested that several of the digitisers 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 digitisation process are small. Reading the value from a nearby station that is given in the row below the station of interest, or accidentally shifting the data by 1 day is very difficult to identify using automatic or semi-automatic quality control procedures. Double-keying, which is considered standard practice for many data entry activities (Barchard and Pace, 2011), would be the best way to overcome these issues, or even triple-keying, which is the method used by a number of citizen science activities (e.g. Old Weather, www.oldweather.org, last access: 8 August 2018). However, this was simply not feasible for this digitation project due to limited resources. While we cannot say that the final version of the dataset from this study is free from errors, the methods we have used have removed or flagged the majority of suspect values.

5   Discussion

Procedures used in this study to identify, digitise and quality control data are an example of the effort required to prepare
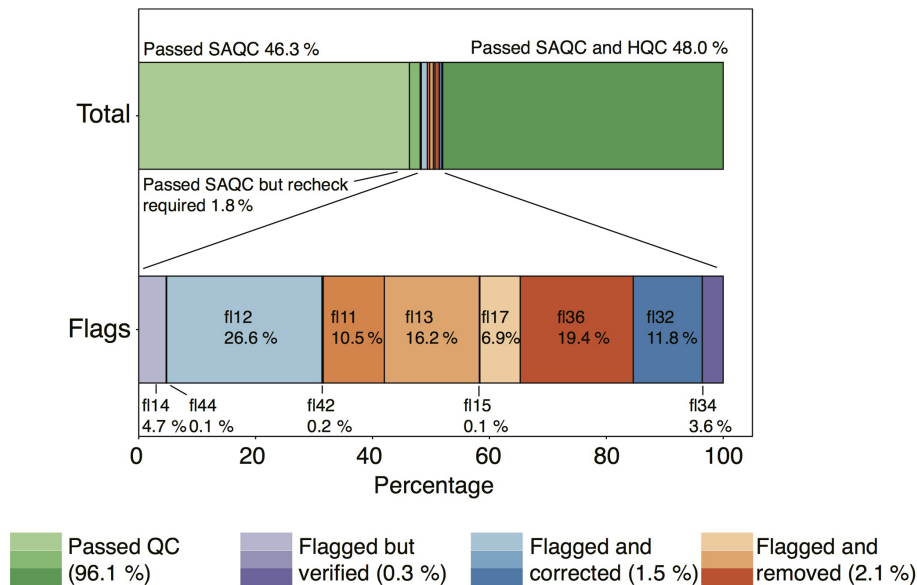
Figure 12. The percentage distribution of quality control flags in the dataset. Values that have passed QC are represented in green (QC flags fl10, fl40 and fl30); values that were flagged but verified as correct are shown in purple (fl14, fl44 and fl34); and values that were flagged but corrected are shown in blue (fl12, fl42 and fl32); and values that were flagged and removed are shown in orange (fl11, fl13, fl15, fl17 and fl36). The darkness of the colours indicates the level of QC applied for each flag. Lighter colours represent values that were only subjected to semi-automatic quality control (SAQC, fl codes that begin with 1), darker colours indicate values subjected to both SAQC and spatial HQC procedures (fl codes that begin with 3), and the colours in the middle represent the small number of values that may need to be rechecked (fl codes that begin with 4). See Table 4 for additional flag details.

an observational dataset for analysis. 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 difficult (Brönnimann et al., 2006). In particular, the transcription of the original observations (referred to here as digitisation) requires a lot of work hours and resources. Without a reliable method of digitisation and a standard method to assess the quality of sources, the accuracy and usability of the final dataset can be jeopardised.

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

In their guide for digitising manuscript climate data, Brönnimann et al. (2006) describe the use of speech recognition, optical character recognition and manual key entry. On balance, they found key entry to be the most efficient method of digitising data, in terms of speed, error rate and the 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 hu-

man eye is still better when translating hand-written observations.

The currently accepted best practice for manual data digitisation 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; Ryan et al., 2018; World Meteorological Organization, 2016). Citizen science efforts that make use of large numbers of volunteers in fact require a value to be keyed at least 3, and up to 5, times (Eveleigh et al., 2013). Coupled with an automatic quality control procedure, these features of the digitisation process are important for providing the best possible opportunity for data accuracy.

However, in reality it is prohibitively expensive and not feasible for many small data recovery projects to use all of these features. Single data entry with visual checking is often the most cost-effective 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 digitisation.*

It is vital to understand the limitations and issues of original data images and sources before the digitisation process begins (Brönnimann et al., 2006), particularly if

the data are provided in pre-scanned format. Checking every page of the original data source before providing it for digitisation 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. With this information it then becomes possible to provide improved instructions to digitisers, develop better templates and tools for digitisation, or even re-scan data sources if possible.

– *Develop user-friendly templates.*

Our examination of duplicated data for Zagazig (Sect. 4.4) does not align with the recommendations made by WMO (2016) about the use of templates. In this case study, one digitiser was asked to key data for more than 20 stations into a template, while the other digitised observations from only 1 station (1 row per page of data source) without a template. More errors were made using a template than not using a template, although it must be noted that the template style was unfamiliar to the digitisers, and different digitisers completed the tasks. 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.

Despite this finding, we still believe that the use of templates acts to reduce the number of digitisation errors. Although templates do not remove issues associated with the original source, they do give the digitiser 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. Another suggestion could be to develop the templates in collaboration with the digitisation team.

– *Involve digitisers in quality control procedures.*

One potentially time-saving method that can be employed to reduce digitisation errors is to involve the digitisers in the quality assurance and quality control of the data. It is true that unreliable digitisers may also make unreliable quality control assessors, but by asking digitisers to run QC on data keyed 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, as poor observational practices may lead to erroneous instrument readings or other mistakes when transcribing the data if the data are secondary sources (Brönnimann et al., 2006; Hunziker et al., 2017).

– *Do not underestimate the value of manually checking quality control results.*

Most QC procedures 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 existent, are neither well developed nor employed and, therefore, data quality issues that may appear systematically can remain inadvertently in the data series (Hunziker et al., 2017).

While 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 incorrectly removed as errors. Completely automated QC procedures used for global products run the risk of removing large swathes 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 1 day due to a digitisation error, it will likely be removed in any automatic spatial analysis with neighbours. Flagging and manually examining these errors allows all of the affected observations to be retained by one correction. 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 may cause a 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 are 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 and decisions based on expert knowledge 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

All versions of the digitised dataset are available through the World Data Center PANGAEA (https://doi.org/10.1594/PANGAEA.886511, Ashcroft et al., 2018). Version 1 contains the raw digitised data, which in the original format includes typographical errors and other issues subsequently identified in the quality control procedure. We have retained this information to ensure

transparency of the process, in case it is useful for future users of the dataset. Version 2 contains the data with SAQC applied. Version 3 contains the data with statistical and spatial automated quality control applied, while Version 4 (labelled "convertedvalue") contains the Version 3 data converted to SI units. Full details of the quality control flags, data sources and station information are also provided.

These files have also been provided to international data repositories, including the International Surface Pressure Databank, the International Surface Temperature Initiative, the C3S 311a Lot 2 Global Land and Marine Observations Database service through the British Science and Technology Facilities Council (STFC)/Centre for Environmental Data Analysis (CEDA), ECMWF's MARS Catalogue, the Global Precipitation Climatology Centre Dataset, the ECA&D, and HadISD. Through these repositories and their connections to ECMWF's MARS holdings, future users should be able to develop long-term composite time series of these and other observations from the European sector. The original data scans are available through each data repository (Table S3) 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, digitising and quality controlling an extensive set of sub-daily meteorological observations across Europe and the southern Mediterranean for use by the wider research community. The multiple, complex steps associated with dataset development are often overlooked when data are used for research, and yet without them, there would be no data to analyse. 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 several ECVs, including 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 data.

These observations have also been subjected to extensive semi-automatic and automatic 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. The QC procedure flagged 3.9 % of the total number of observations digitised, with 2.1 % of the total number removed, 1.5 % corrected and 0.3 % retained as

correct observations. These QC results are on par with other data rescue activities. It is our hope that these observations support and improve the next generation of international and European weather and climate services.

The Supplement related to this article is available online at https://doi.org/10.5194/essd-10-1613-2018-supplement.

References

Aguilar, E., Auer, I., Brunet, M., Peterson, T. C., and Wieringa, J.: Guidelines on Climate Metadata and Homogenization, World Meteorological Organization, WMO/TD No. 1186, 55 pp., 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, B. Am. Meteorol. Soc., 92, 1421–1425, https://doi.org/10.1175/2011BAMS3218.1, 2011.

Angot, A.: Instructions météorologiques, Meteofrance, Paris, France, available at: http://bibliotheque.meteo.fr/exl-php/

vue-consult/mf_-_recherche_avancee/ISO00008704 (last access: 4 June 2018), 1931.

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

Ashcroft, L., Coll, J.R., Gilabert, A., Domonkos, P., Aguilar, E., Sigro, J., Castella, M., Unden, P., Harris, I., Jones, P., and Brunet, M.: Meteorological observations for Europe and the southern Mediterranean region, 1877–2012, PANGAEA, https://doi.org/10.1594/PANGAEA.886511, 2018.

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, 17–46, https://doi.org/10.1002/joc.1377, 2007.

Barchard, K. A. and Pace, L. A.: Preventing human error: The impact of data entry methods on data accuracy and statistical results, Comput. Human Behav., 27(5), 1834–1839, https://doi.org/10.1016/J.CHB.2011.04.004, 2011.

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, 137–144, https://doi.org/10.5194/cp-2-137-2006, 2006.

Brunet, M. and Jones, P.: Data rescue initiatives: bringing historical climate data into the 21st century, Clim. Res., 47, 29–40, https://doi.org/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, 1777–1802, https://doi.org/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, 121–128, https://doi.org/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, 61–73, https://doi.org/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, https://doi.org/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. Roy. Meteor. Soc., 137, 1–28, https://doi.org/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., Auch-

mann, 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, 31–46, https://doi.org/10.1002/gdj3.25, 2015.

Dahlgren, P., Landelius, T., Kållberg, P., and Gollvik, S.: A high-resolution regional reanalysis for Europe, Part 1: Three-dimensional reanalysis with the regional HIgh-Resolution Limited-Area Model (HIRLAM), Q. J. Roy. Meteor. Soc., 142, 2119–2131, https://doi.org/10.1002/qj.2807, 2016.

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, 1649–1679, https://doi.org/10.5194/cp-8-1649-2012, 2012.

Da Silva, A. M., Young, C. C., and Levitus, S.: Toward a revised Beaufort equivalent scale, in Proc. Int. COADS Winds Workshop, NOAA-ELR, Kiel, 270–286, 1995.

Dunn, R. J. H., Willett, K. M., Parker, D. E., and Mitchell, L.: Expanding HadISD: quality-controlled, sub-daily station data from 1931, Geosci. Instrum. Method. Data Syst., 5, 473–491, https://doi.org/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, 1615–1633, https://doi.org/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, https://doi.org/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, 188–199, https://doi.org/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, 4131–4145, https://doi.org/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, https://doi.org/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., Lin-

derholm, 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, D22106, https://doi.org/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, J., Tibig, L., Tran, T. D., Vediapan, K., and Zhai, P.: Data Rescue in the Southeast Asia and South Pacific Region: Challenges and Opportunities, B. Am. Meteorol. Soc., 85, 1483–1489, https://doi.org/10.1175/BAMS-85-10-1483, 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 ofin situ atmospheric climate data: a review, Int. J. Climatol., 18, 1493–1517, https://doi.org/10.1002/(SICI)1097-0088(19981115)18:13<1493::AID-JOC329>3.0.CO;2-T, 1998.

Ryan, C., Duffy, C., Broderick, C., Thorne, P. W., Curley, M., Walsh, S., Daly, C., Treanor, M., Murphy, C., Ryan, C., Duffy, C., Broderick, C., Thorne, P. W., Curley, M., Walsh, S., Daly, C., Treanor, M., and Murphy, C.: Integrating data rescue into the classroom, B. Am. Meteorol. Soc., https://doi.org/10.1175/BAMS-D-17-0147.1, online first, 2018.

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, B. Am. Meteorol. Soc., 95, 1419–1430, https://doi.org/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, available at: http://www.wmo.int/pages/prog/wcp/wcdmp/documents/WCDMP-55.pdf (last access: 4 June 2018), 2004.

Universitat Rovira i Virgili: Original scans of data sources used in the development of the UERRA sub-daily dataset, available at: ftp://130.206.36.123 (last access: 8 August 2018), 2018.

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, 89–115, https://doi.org/10.5194/cp-8-89-2012, 2012.

World Meteorological Organization: GSOC-195: Status of the Global Observing System for Climate, available at: https://unfccc.int/sites/default/files/541.pdf (last access: 4 June 2018), 2015.

World Meteorological Organization: Guidelines on Best Practices for Climate Data Rescue 2016, WMO-No. 1182, available at: https://public.wmo.int/en/resources/library/guidelines-best-practices-climate-data-rescue (last access: 8 August 2018, 2016.