

Digitizing observations from the Met Office Daily Weather Reports for 1900–1910 using citizen scientist volunteers

Article

Published Version

Creative Commons: Attribution 4.0 (CC-BY)

Open Access

Craig, P. M. ORCID: <https://orcid.org/0000-0001-9213-4599>
and Hawkins, E. ORCID: <https://orcid.org/0000-0001-9477-3677> (2020) Digitizing observations from the Met Office Daily Weather Reports for 1900–1910 using citizen scientist volunteers. *Geoscience Data Journal*, 7 (2). pp. 116-134. ISSN 2049-6060 doi: 10.1002/gdj3.93 Available at <https://centaur.reading.ac.uk/91043/>

It is advisable to refer to the publisher's version if you intend to cite from the work. See [Guidance on citing](#).

To link to this article DOI: <http://dx.doi.org/10.1002/gdj3.93>

Publisher: Wiley

All outputs in CentAUR are protected by Intellectual Property Rights law, including copyright law. Copyright and IPR is retained by the creators or other copyright holders. Terms and conditions for use of this material are defined in the [End User Agreement](#).

www.reading.ac.uk/centaur

CentAUR

Central Archive at the University of Reading

Reading's research outputs online

Digitizing observations from the Met Office Daily Weather Reports for 1900–1910 using citizen scientist volunteers

Philip M. Craig | Ed Hawkins

Department of Meteorology, National Centre for Atmospheric Science, University of Reading, Reading, UK

Correspondence

Philip M. Craig, Department of Meteorology, NCAS, University of Reading, Reading, UK.
Email: p.craig@reading.ac.uk

Funding information

Met Office, Grant/Award Number: P105978; UK National Centre for Atmospheric Science; European Union, Grant/Award Number: 690462; UK Research & Innovation (UKRI) Strategic Priorities Fund UK Climate Resilience programme

Abstract

We describe the digitization of 1.8 million sub-daily and daily weather observations which were recorded in the UK Met Office Daily Weather Reports during the 1900–1910 period. The data were rescued from scanned images of the original documents by 2,148 volunteer citizen scientists using the weatherrescue.org website. The rescued observations include dry and wet bulb temperatures, daily maximum and minimum temperatures, daily rainfall amounts and sub-daily sea-level pressure from 72 different locations across western Europe. These observations will be used to fill gaps in existing pressure, temperature and rainfall records and are one of the largest recoveries of weather data by citizen scientists. The value of these additional observations is highlighted by comparing the pressure observations to the Twentieth Century Reanalysis version 3 ensemble for some specific case studies.

KEYWORDS

citizen science, daily weather report, data rescue, weather observations

1 | INTRODUCTION

In September 1860, the UK Met Office (then known as the Meteorological Department of the Board of Trade), under the direction of Admiral Robert FitzRoy, began publishing Daily Weather Reports (DWRs) which consisted of tabulated weather observations from stations around Great Britain and Ireland (GBI). These stations communicated the observations daily to the Met Office in London by telegraph (Lempfert, 1913; Walker, 2011). Fitzroy used these observations to

provide storm warnings and the first ever public ‘weather forecasts’, starting in August 1861. The observations from the DWRs have yet to be digitized so do not currently exist in any databases, although similar data from a small number of stations have been digitized from other sources.

The DWRs are handwritten which makes them more difficult to digitize with optical character recognition (OCR) than typeset documents (Brönnimann *et al.*, 2006) so must be digitized manually. Any attempts to digitize large quantities of handwritten weather observations are extremely

Dataset Identifier: <https://catalogue.ceda.ac.uk/uuid/235ff4a040854dcd8dfb754bbb898479>

Creator: Philip Craig, Ed Hawkins and 2148 volunteers

Title: Met Office Daily Weather Reports 1900–1910

Publisher: CEDA

Publication year: 2019

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2020 The Authors. *Geoscience Data Journal* published by Royal Meteorological Society and John Wiley & Sons Ltd.

time-consuming and can result in many typographical errors (Le Blancq, 2010). Coll *et al.* (2019) manually digitized 610,000 observations of temperature, rainfall, sunshine and snow depth for stations in the Balkans and Central Europe with a real-time quality control (QC) method to avoid transcription errors such as repeating or skipping values. Here, we use citizen scientist volunteers to digitize the 1900–1910 Met Office DWRs.

Hawkins *et al.* (2019) previously used this citizen science approach to digitize 1.5 million typeset observations of pressure, temperature, rainfall and wind speed from the Ben Nevis Observatory and Fort William in Scotland from 1883 to 1904, and other similar projects have also been successful, for example oldWeather.org and southern-weatherdiscovery.org. Slonosky *et al.* (2018) and Sieber and Slonosky (2019) provide recommendations on how to engage volunteers as citizen scientists, such as providing small amounts of data to digitize rather than a whole page. Ryan *et al.* (2018) discuss how to integrate data rescue activities into university degree courses using daily rainfall for Ireland in the 19th century.

The rescued observations from these projects can then be added to international digital observation archives, such as the International Surface Pressure Databank (ISPD), which contains observations of surface pressure and mean sea-level pressure (mslp) from land and sea locations across the world (Cram *et al.*, 2015). Data coverage is, however, sparse and inconsistent in most regions. For example, the most recent version of ISPD (ISPD v4.7; Compo *et al.*, 2019) has no pressure observations across England and Wales at the beginning of the 20th Century, and the stations in Scotland (four), Ireland (three) and the Channel Islands (one) were all digitized separately rather than as one project. The lack of observations in ISPD and similar datasets means that the reconstructed atmospheric circulation over northern Europe in extended reanalyses such as the 20th Century Reanalysis (20CR; Compo *et al.*, 2011; Slivinski *et al.*, 2019) is very uncertain for this time period and produces inconsistent trends in storminess in 20CR and other centennial reanalyses (Befort *et al.*, 2016; Bloomfield *et al.*, 2018; Rohrer *et al.*, 2019). Pressure data from GBI are useful for improving the quality of reanalyses downstream in central Europe, and the eleven years considered for this dataset cover a cool period before the onset of early 20th-century warming (Hegerl *et al.*, 2018) and include the Santa Maria volcanic eruption in 1902 which suppressed North Atlantic tropical cyclone activity (Yang *et al.*, 2019).

There are also many gaps in the UK Met Office Integrated Data Archive System (MIDAS) dataset (Met Office, 2019) in the early 20th century. Of the many British and Northern Irish stations in the 1900–1910 DWRs, only four have temperature or rainfall data in MIDAS for this time period. Other stations have data starting later, but these are often incomplete records. Data from MIDAS are the basis of the Met

Office gridded dataset for climate variables (HadUK-Grid; Hollis *et al.*, 2019). The addition of data from new locations and insertion of missing data from existing locations will help to improve the quality of the HadUK-Grid dataset. The ISPD will eventually be discontinued and replaced with the Global Land and Marine Observations Dataset (GLAMOD)—a plan for which is outlined in Thorne *et al.* (2017).

In this paper, we describe the dataset of pressure, temperature and rainfall observations recovered from the 1900–1910 DWRs with a citizen science project. In Section 2, we discuss the DWRs in detail and the stations at which observations were taken. The data collection by citizen scientists and QC procedure are outlined in Section 3. Section 4 provides details regarding availability of the dataset and analyses the output to highlight the advantages of these new data and the effectiveness of the volunteers. Section 5 provides a summary of the key findings and gives an outlook for future data rescue efforts.

2 | DATA SOURCES

2.1 | The daily weather reports

The data recovered are from the Met Office DWRs from 1900 to 1910. The DWRs are stored as monthly collections, and the scanned copies are freely available to download from the National Meteorological Library and Archive (NMLA, <https://digital.nmla.metoffice.gov.uk/>). The NMLA has DWRs from September 1860 to December 1980, after which they were replaced by the Daily Weather Summary which ran until December 2002. The 1900–1910 DWRs provide weather observations from stations around GBI, plus several stations from western Europe.

For the 11 years covered in this dataset, the DWRs consisted of four pages each day. The first page is a table of observations split into blocks of stations by region with columns of weather observations starting with the previous evening's observations on the left, followed by the current day's morning and ending on the right with observations covering the previous 24 hr (Figure 1). The 'Yesterday Evening' columns contained observations of mslp, dry bulb temperature (T_{dry}), wind direction/force and weather conditions. The 'This Morning' columns contained observations of mslp, the change in pressure over the last 24 hr, T_{dry} and wet bulb temperature (T_{wet}), wind direction/force, weather conditions and sea conditions. The 'Past 24 Hours' columns reported the maximum and minimum temperatures (T_{max} , T_{min}), weather conditions, sunshine hours and total rainfall. Pressure observations were given in inches of mercury (inHg), temperature in degrees Fahrenheit (°F) and rainfall in inches (in). We have only recovered pressure, temperature and rainfall observations for this dataset.

DAILY WEATHER REPORT

for 8 a.m. on Wednesday, 1st JULY, 1903.

Issued by the METEOROLOGICAL OFFICE, 63, Victoria Street, London. W. N. SHAW, Secretary.

STATIONS.	YESTERDAY EVENING. (See note on last page.)					THIS MORNING. (See note on last page.)										PAST 24 HOURS.				
	Barom. At 32° and M. S. L.	Temp. Dry Bulb.	Wind Dir.	Force (in M.P.H.)	Weather	Barom. At 32° and M. S. L.	Change since Yester- day.	Temp. Dry Bulb.	Wet Bulb.	Change since Yester- day.	Wind Dir.	Force (in M.P.H.)	Weather	Sea (0 to 5)	Temp.		Weather Sun- shine Hours.	Rainfall. Inch.		
															Max.	Min.				
															Inch.	"			"	"
SCANDINAVIA.																				
Haparanda ...	29.65	53	SW	2	0	29.67	-0.02	51	50	-3	NW	4	0	61	48	—	—	0.04		
Hernösand ...	29.87	57	W	6	0	29.76	-0.11	59	57	-3	NW	4	0	70	52	—	—	—		
Stockholm ...	29.95	64	W	2	0	30.00	+0.05	62	58	-6	N	2	0	75	57	—	—	—		
Wishy ...	30.02	60	2	0	0	30.01	-0.01	57	52	-5	2	0	0	73	54	—	—	—		
Karlstad ...	29.96	63	SW	2	0	30.07	+0.11	59	52	-1	NNE	2	0	71	54	—	—	—		
Färder (St. Anna) ...	29.99	64	WNW	5	0	30.05	+0.06	61	55	-6	WNW	1	0	66	47	—	—	—		
Bodö ...	29.80	48	SW	4	0	29.92	+0.12	49	48	-1	NW	2	0	58	46	—	—	0.20		
Christiansund ...	30.03	50	W	6	0	30.03	-0.00	53	50	-3	W	2	0	52	48	—	—	0.20		
Skudesnaes ...	30.14	54	W	2	0	30.08	-0.06	53	52	-1	SE	4	0	55	52	7.0	—	—		
Sumburgh Head ...	30.03	53	3	0	0	29.93	-0.10	51	51	-2	SW	3	0	57	49	0	—	0.18		
Stornoway ...	29.96	55	SW	4	0	29.85	-0.11	57	55	-2	WNW	5	0	57	49	0	—	0.18		
Malin Head ...	30.10	58	W	6	0	30.04	-0.06	59	55	-4	SW	5	0	59	55	2.0	—	0.04		
Blackpool Pt. ...	30.11	60	SW	5	0	30.05	-0.06	61	58	-3	SW	7	0	62	57	2.0	—	0.01		
Valencia ...	30.26	61	SW	4	0	30.22	-0.04	59	56	-3	SW	7	0	64	56	0	—	0.01		
BRITISH ISLANDS.																				
Roche's Point ...	30.21	60	WNW	3	0	30.27	+0.06	59	57	-2	SW	3	0	63	56	0	—	—		
Parsonstown ...	30.24	63	W	2	0	30.21	-0.03	61	57	-4	W	2	0	65	53	—	—	0.3		
Donaghadee ...	30.21	59	S	1	0	30.18	-0.03	61	58	-3	SW	2	0	64	53	0.0	—	0.07		
Liverpool Obsy. ...	30.20	60	WNW	3	0	30.23	+0.03	59	55	-4	WNW	3	0	67	52	—	—	—		
Holyhead ...	30.25	57	SW	3	0	30.27	+0.02	56	52	-4	SW	4	0	61	53	0.0	—	—		
Pembroke (St. Ann's) ...	30.26	56	SW	4	0	30.30	+0.04	56	52	-4	SW	3	0	59	54	6.0	—	—		
Scilly (St. Mary's) ...	30.27	62	SW	2	0	30.29	+0.02	58	56	-2	SE	2	0	60	56	2.0	—	1.0		
Jersey (St. Aubin's) ...	30.28	62	WNW	4	0	30.27	-0.01	61	57	-4	E	3	0	67	48	2.0	—	0.15		
Portland Bill ...	30.26	62	W	1	0	30.34	+0.08	60	58	-2	SE	2	0	64	53	0	—	—		
Dungeness ...	30.24	62	SW	1	0	30.35	+0.11	62	59	-3	WNW	1	0	68	52	2.0	—	0.14		
Wick ...	30.02	58	SW	3	0	29.90	-0.12	56	55	-1	SW	1	0	62	50	0	—	0.38		
Nairn ...	30.03	56	W	1	0	29.95	-0.08	60	56	-4	WNW	4	0	62	54	—	—	0.15		
Aberdeen ...	30.10	56	S	3	0	29.99	-0.11	60	56	-4	SW	4	0	65	54	0.0	—	0.22		
Leith ...	30.14	63	SW	3	0	30.05	-0.09	62	58	-4	SW	4	0	67	57	0	—	0.01		
North Shields ...	30.22	61	SW	2	0	30.16	-0.06	62	56	-6	SW	2	0	67	55	0	—	0.04		
SPAIN, &c.																				
Spurn Head ...	30.29	65	WNW	3	0	30.28	-0.01	60	58	-2	WNW	3	0	68	57	—	—	—		
Yarmouth ...	30.21	68	SE	2	0	30.29	+0.08	65	63	-2	W	3	0	65	54	0	—	—		
Nottingham ...	30.24	63	WNW	2	0	30.35	+0.11	60	56	-6	W	3	0	71	51	—	—	—		
Oxford ...	30.22	72	WNW	2	0	30.33	+0.11	64	56	-8	W	1	0	75	52	0	—	—		
London (Brixton) ...	30.22	72	WNW	2	0	30.33	+0.11	64	56	-8	W	1	0	75	52	0	—	—		
FRANCE.																				
The Scaw (Skagen) ...	30.07	59	WNW	4	0	30.11	-0.04	59	54	-5	NW	4	0	63	55	—	—	—		
Fanö ...	30.26	56	WNW	3	0	30.25	-0.01	59	57	-2	W	4	0	66	55	—	—	—		
Cuxhaven ...	30.27	65	NW	3	0	30.22	-0.05	61	57	-4	WNW	3	0	66	57	—	—	—		
The Helder ...	30.21	62	NW	1	0	30.29	+0.08	62	57	-1	NW	2	0	63	54	—	—	—		
Brussels ...	30.19	65	NW	4	0	30.27	+0.08	64	56	-2	SE	1	0	73	56	—	—	—		
Berlin ...	30.19	65	NW	4	0	30.28	+0.09	61	57	-6	NW	3	0	73	54	—	—	—		
Frankfurt ...	30.23	65	N	2	0	30.21	-0.02	65	58	-7	N	2	0	74	57	—	—	—		
Munich ...	30.24	65	W	2	0	30.35	+0.11	62	58	-6	SE	1	0	75	57	—	—	—		
SPAIN, &c.																				
C. Gris Nez ...	30.23	63	NE	1	0	30.22	-0.01	59	58	-1	SW	2	0	70	54	—	—	—		
Brest ...	30.29	64	WNW	2	0	30.27	-0.02	57	51	-6	E	3	0	68	52	—	—	0.16		
Lorient ...	30.26	63	W	2	0	30.32	+0.06	57	55	-2	E	1	0	72	54	—	—	0.12		
Rochefort ...	30.27	69	N	3	0	30.25	-0.02	69	60	-9	E	4	0	81	57	0.0	—	0.12		
Biarritz ...	30.31	64	WNW	3	0	30.23	-0.08	63	63	0	E	1	0	71	59	—	—	—		
PARIS.																				
Paris ...	30.21	73	NE	3	0	30.26	+0.05	61	58	-3	ENE	2	0	77	53	—	—	—		
Belfort ...	30.21	62	ENE	3	0	30.23	+0.02	59	57	-2	ENE	3	0	76	55	—	—	0.16		
Lyons ...	30.25	69	WNW	2	0	30.25	+0.00	64	58	-6	WNW	2	0	72	54	—	—	0.12		
Nice ...	29.97	65	W	4	0	30.04	-0.09	64	60	-4	E	1	0	79	59	—	—	0.12		
Perpignan ...	30.14	79	WNW	3	0	30.15	+0.01	74	62	-12	W	3	0	85	68	—	—	—		
SPAN. PEN.																				
Corunna ...	30.25	72	WNW	5	0	30.25	-0.00	64	64	0	SW	3	0	84	68	—	—	—		
Lisbon ...	30.21	67	SW	4	0	30.16	-0.05	65	64	-1	W	2	0	72	64	—	—	0.10		
Azores (P. Delgada) ...	30.21	67	SW	4	0	30.16	-0.05	65	64	-1	W	2	0	72	64	—	—	0.10		

NOTE: The Sunshine for London given in this Table is recorded at the Tower of the Watney Training College, Westminster; the Sunshine entered in Dungeness is recorded at Littlestone-on-Sea.

* Observations for this column not generally received. † Observations received too late for insertion in Charts on next page.

NOTE: The Sunshine for London given in this Table is recorded on the Tower of the Watkin Training College, Westminster; the Sunshine entered in Dungeness is recorded at Littlestone-on-Sea.

* Observations for this column not generally recorded.

† Observations received too late for insertion in Charts on next page.

FIGURE 1 An example of page 1 of DWR from Wednesday 1 July 1903. Stations are listed in blocks by region in the column on the left-hand side of the page, and the observations are presented in columns with 'yesterday evening' first, then 'this morning' and 'past 24 hours'

The content of the second, third and fourth pages changed several times. The second and third pages generally consisted of hand-drawn charts, handwritten notes regarding the previous day's weather and handwritten forecasts. Page 4 contained various additional observations and information useful for this dataset. For example, the previous day's 2 p.m. observations for various stations

in 1900 were digitized but were no longer published in the DWRs after 1900.

Each month's worth of DWRs is appended with pages containing 'Additions and corrections to the reports daily' (Figure 2). The first page of this is the 'Additions' page and takes the same columnar format as page 1 of the DWR (Figure 1), but the rows are ordered by days and each

ADDITIONS AND CORRECTIONS TO THE																			
REPORTS FOR JULY, 1903.																			
1. REPORTS RECEIVED TOO LATE FOR INSERTION DAILY.																			
DATE.	STATIONS.	YESTERDAY EVENING				THIS MORNING										PAST 6 HOURS			
		Barom.	Temp.	Wind.	Weather.	At 6 P.M.	Change since 6 P.M.	At 9 A.M.	Change since 9 A.M.	At 12 M.	Change since 12 M.	At 3 P.M.	Change since 3 P.M.	At 6 P.M.	Change since 6 P.M.	Barom.	Temp.	Wind.	Weather.
1 st	Corwen...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
2 nd	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
3 rd	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
4 th	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
5 th	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
6 th	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
7 th	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
8 th	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
9 th	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
10 th	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
11 th	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
12 th	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
13 th	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
14 th	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
15 th	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
16 th	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
17 th	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
18 th	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
19 th	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
20 th	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
21 st	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
22 nd	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
23 rd	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
24 th	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
25 th	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
26 th	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
27 th	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
28 th	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
29 th	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
30 th	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
31 st	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...

FIGURE 2 An example of the monthly additions and corrections pages for July 1903. Stations are listed by day in the column on the left-hand side, and the observations are presented in the same order as page 1 of the DWRs

station which reported too late on that day. The second page ('Corrections') is of the same format as the additions page (Figure 1) and contains observations which supersede the values reported on page 1 of the DWR.

2.2 | Stations

Data from 72 stations from GBI and other European countries were recovered for this dataset (Figure 3). The stations included on page 1 of the DWRs changed from year to year, expanding as time progressed (summarized by Lempfert (1913) for 1867–1913), but 49 stations were included on page 1 of the DWR for the full 11 years considered here. Some stations stopped reporting during this period and were replaced by another nearby, specifically Loughborough/Nottingham, Wiesbaden/Frankfurt and Blonduos/Isafjord. One peculiarity is Nancy which replaced Belfort on page 1 of the DWR for 5 days in 1909, only for Belfort to reappear. During 1901–1903, Bath and Clacton were in the 'additional observations' on page 4 but were promoted to page 1 for 1904 onwards (we only digitized the 1903 Bath and Clacton observations for this dataset). The most significant development in the coverage of stations for the DWRs was the inclusion of

observations from the Faroe Islands and Iceland in 1907 after a telegraph cable was laid down in summer of 1906 (Lempfert, 1913).

2.2.1 | British and Irish stations

It is important to locate the appropriate metadata for each station. Various documents in the NMLA (summarized in Table 1) provide latitude and longitude coordinates and elevations of the stations. There are inconsistencies between barometer and rain gauge heights between some documents, but we have used the Annual Reports as the definitive source if any inconsistencies were found. The specific NMLA documents and other sources used to locate each station and find their elevations are noted in the final column of Table 2.

The latitude and longitude coordinates were only presented in degrees and arc minutes so the accuracy is poor compared to modern-day standards. In addition, the coordinates available from NMLA documents can sometimes be erroneous with some stations appearing to be in the sea; some of these coordinates are carried over into MIDAS. For example, the same coordinates (59°51'N, 1°17'W) for Stornoway appear in various documents

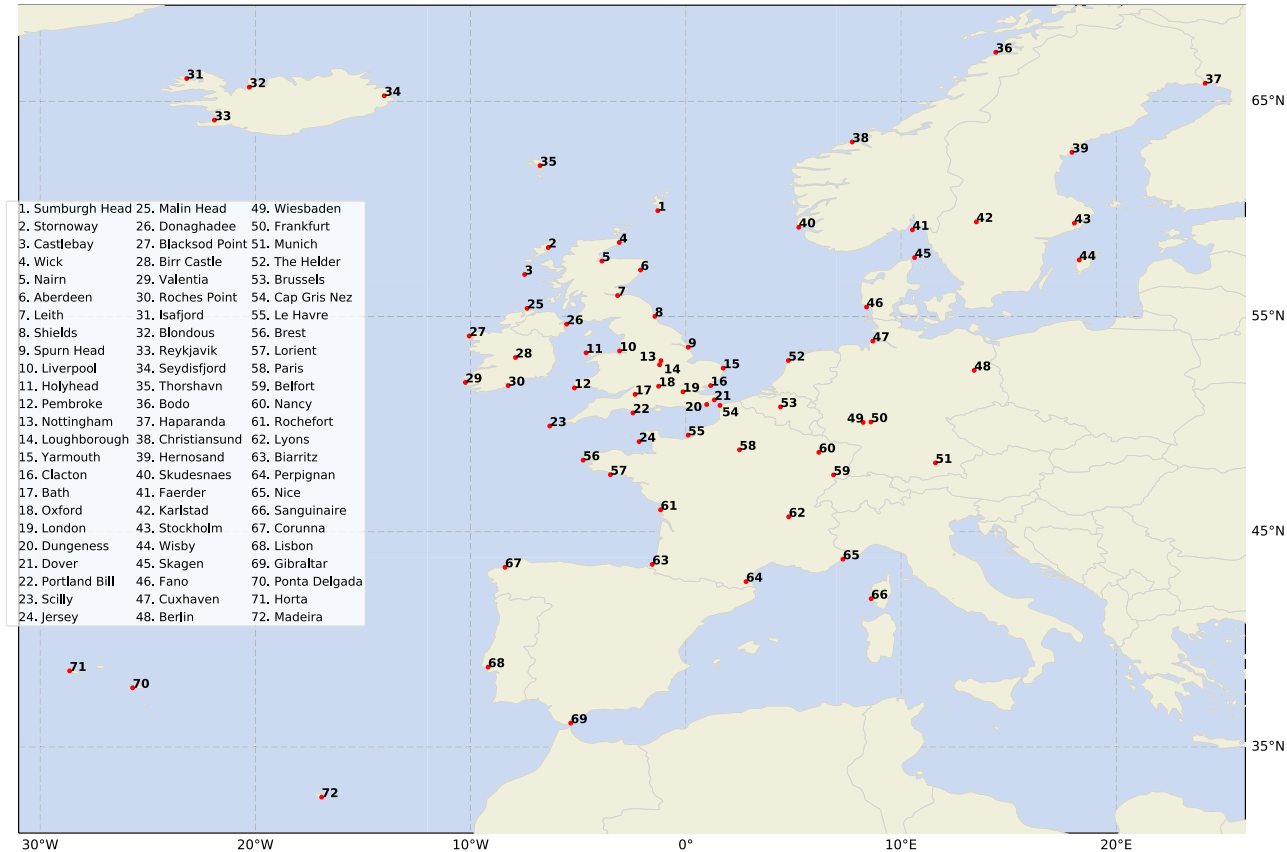


FIGURE 3 Map of all 72 stations included in the dataset from the 1900–1910 DWRs. The country borders shown are the modern day borders

TABLE 1 List of key documents which were used to find the metadata of the British and Irish stations in the 1900–1910 DWRs. Documents labelled with an asterisk are both referred to an Annual Reports in the main text of this paper. After 1920, when control of the Met Office passed from the Board of Trade to the Air Ministry (Walker, 2011), the content of the Annual Reports began to change with less information about the stations

Document type	Available information	Available years	Available from
Reports of the Meteorological Council to the President and Council of the Royal Society*	Exhaustive list of stations which reported to the Met Office Latitude and longitude coordinates Rain gauge elevations above sea level	1857–1905/6	Online at NMLA
Annual Reports of the Meteorological Committee to the Lords Commissioners of His Majesty’s Treasury*	Observer names and/or employers Station inspection reports	1906/7–1919/20	
Monthly Weather Reports (MWRs)	Latitude and longitude coordinates Barometer and rain gauge elevations above sea level	1884–1993	Online at NMLA
British Rainfall	Rain gauge elevation above sea level Information regarding the specific site of the station, for example street or building name	1860–1968	Online at NMLA

but this location is actually off the coast of Lewis on the wrong side of Stornoway Harbour. The observing equipment moved around Stornoway from the beginning of its use as a telegraphic reporting station in 1872 with stints at the castle, a school and various houses (summarized at www.lewisweather.uk).

Metadata for the 30 GBI stations are summarized in Table 2. The latitude and longitude coordinates are presented in decimal form and rounded to two decimal places. The Annual Reports provide the name and/or organization of the observer which helped to locate the specific site where the observations were taken. The specific locations of the

TABLE 2 Metadata for all 30 British and Irish stations included in the dataset. The longitude/latitude co-ordinates are presented in decimal form and rounded to two decimal places (but stored with greater accuracy in the SEF files). The elevations refer to the height of the rain gauge above sea level in metres

Station	lat/lon	Elevation (m)	Reporting period	Observer(s)	Site	Notes
Sumburgh Head	59.92°N, 1.30°W	34.1 ^{''}	11 years	Rev. William Brand	Church of Scotland, Dunrossness	Annual Reports 1871, 1902/3
Stornoway	58.21°N, 6.39°W	8.8	11 years	J. MacKenzie	Kenneth St Scotland St. from Spring 1902	Annual Reports 1902/3, 1911/12 www.lewisweather.uk
		15.5		W. Grant (July 1907–)	Lewis St.	
Castlebay	56.96°N, 7.49°W	11.3	09/07/1907–31/12/1910	James Smith	Old Schoolhouse	Annual Report 1911/12 Canmore
Wick	58.44°N, 3.09°W	24.4	11 years	James Sinclair Jessie Sinclair	Market Place	Annual Reports 1893/4, 1902/3, 1911/12
Nairn	57.58°N, 3.90°W	25.0	11 years	Miss Penny	Delnies School (now demolished)	Annual Report 1902/3 British Rainfall 1900–10
Aberdeen	57.16°N, 2.10°W	14.0	11 years	G.A. Clarke	Kings College, University of Aberdeen	Annual Report 1902/3 Geddes (1955)
Leith	55.97°N, 3.17°W	5.8	11 years	T. Richardson A.J. Bottrill	Post Office, Constitution Street	Annual Report 1902/3 Historic Environment Scotland
North Shields	55.01°N, 1.44°W	30.2	11 years	W.B. Clark, Post Office	Dockwray Square	Annual Reports 1874, 1875, 1902/3 British Rainfall 1903–10
Spurn Head	53.58°N, 0.11°E	7.9	11 years	A.S. Badcock, Lightkeeper	Lighthouse	Annual Reports 1902/3
Liverpool	53.40°N, 3.07°W	57.3	11 years	W.E. Plummer	Bidston Observatory, The Wirral	Annual Report 1902/3 Reynolds (1954)
Holyhead	53.32°N, 4.63°W	14.6	11 years	T. Chope	Sailor's Home, Prince of Wales Road (now Sea Cadets Hall)	Annual Report 1893/4, 1902/3
Pembroke	51.68°N, 5.18°W	45.7	11 years	G.H. Dunsford, Lightkeeper	Lighthouse, St. Ann's Head	Annual Report 1902/3
Nottingham	52.95°N, 1.15°W	58.5	26/07/1903–31/12/1910	A. Brown & P. Boobyer	Nottingham Castle (1903)	Annual Reports 1902/3
		25.0			Trent Lane (from 1904)	British Rainfall 1903–10
Loughborough	52.76°N, 1.22°W	44.5	01/01/1900–16/1/1903	W. Berridge	Forest Road	Annual Report 1902/3 Met. Magazine, March 1892

(Continues)

TABLE 2 (Continued)

Station	lat/lon	Elevation (m)	Reporting period	Observer(s)	Site	Notes
Yarmouth	52.60°N, 1.74°E	3.0	11 years	G.T. Watson	Sailor's Home, 23 Marine Parade	Annual Report 1902/3
Clacton	51.79°N, 1.15°E	16.5	01/01/1903–31/12/1910	A.W. Shadick (Town Council)	Town Hall	Annual Report 1902/3
Bath	51.39°N, 2.36°W	20.1	01/01/1903–31/12/1910	W. H. Symons, Medical Officer of Health	Henrietta Park	Annual Report 1902/3 Marriott (1901)
Oxford	51.76°N, 1.26°W	63.4	11 years	W. Wickham, Radcliffe Observer	Radcliffe Observatory	Annual Report 1902/3 Smith (1968)
London	51.46°N, 0.13°W	23.5	01/01/1900–29/12/1904	F. Gaster	Acre Lane, Brixton	Annual Report 1902/3 Bench (1981)
	51.50°N, 0.23°W	8.2	30/12/1904–31/12/1910	HM Office of Works	St. James' Park	Walker (2011)
Dungeness	50.91°N, 0.97°E	7.9	11 years	J.G. Williams, Lightkeeper	Old Lighthouse	Annual Report 1902/3
Dover	51.13°N, 1.32°E	70.4	04/08/1907–31/12/1910	W.C. Hawke, Borough Surveyor	Elevation suggests the castle	Annual Report 1911/12 British Rainfall 1908–10
Portland Bill	50.51°N, 2.46°W	53.9	11 years	W.J. Batton, Lightkeeper	Old Higher Lighthouse (until May 1906)	Annual Reports 1902/3, 1911/12
		5.8			New Lighthouse (from May 1906)	1906 MWR
Scilly	49.91°N, 6.32°W	19.8	11 years	A. Hicks	Signal Station, The Garrison	Annual Report 1902/3
Jersey	49.20°N, 2.13°W	7.6	11 years	John Fisher	Rue du Croquet, St. Aubin	Annual Report 1902/3 Le Blancq (2010)
Malin Head	55.38°N, 7.37°W	70.1	11 years	A.C. Hailstone	Signal Station	Annual Report 1902/3
Donaghadee	54.64°N, 5.53°W	12.2	11 years	W. Keown	Old Coastguard Station (2–10 Warren Road)	Annual Report 1902/3 National Museums NI
Blacksod Point	54.10°N, 10.06°W	11.3	11 years	A. Marshall, Coastguard	Lighthouse	Annual Report 1902/3
Birr Castle/ Parsonstown	53.10°N, 7.91°W	53.3	11 years	J.L. Roe	Birr Castle	Annual Report 1902/3
Valentia	51.94°N, 10.24°W	9.1	11 years	J.E. Cullum	Observatory	Annual Report 1902/3 Murphy (1990)
Roches Point	51.79°N, 8.25°W	12.8	11 years	B. Kelleher/J. Mountjoy, Post Office	Lighthouse	Annual Report 1902/3 Dublin City Council

stations in Aberdeen (Geddes, 1955), Liverpool (Reynolds, 1954), Oxford (Smith, 1968) and Valentia (Murphy, 1990) were already known, but further investigation was required to locate other stations (Table 1).

Some stations are not in the location that their name in the DWR may suggest. For example, the *Liverpool* observations were taken at the Bidston Observatory (Reynolds, 1954) on The Wirral across the River Mersey from Liverpool, and the *Pembroke* observations were taken at St Ann's Head Lighthouse, across the Milford Haven Waterway from the town of Pembroke. The London observations moved from Brixton to St James' Park in December 1904 (Table 2). The Brixton observations were taken by Frederic Gaster at his house (Bench, 1981), but he retired from his role at the Met Office in late 1903. The St. James' Park site came into operation on 8 November 1904 but did not replace Brixton in the DWR until 30 December of that year. For other stations where there were various potential locations for the DWR observations (e.g., Sumburgh Head, Dover, Nottingham), we compared the annual rainfall totals from British Rainfall to the DWR observations to determine which stations took observations for the DWRs.

2.2.2 | Stations outside of Great Britain and Ireland

Finding the specific locations of stations outside GBI was a more difficult task as the documents currently available from the NMLA give little or no information regarding foreign stations. However, we were able to find specific locations of 20 stations which are discussed in this section. Meteorological observations were taken at the Old Astronomical Observatory in Stockholm (Moberg *et al.*, 2002; Yaskell, 2008), and thrice-daily mslp data are already included in ISPD. Also included in ISPD are Nordby (Fanø), Tørshavn, Wisby, Haparanda, Härnösand and Bodø which were rescued as part of the North Atlantic Climatological Dataset (NACD, Frich *et al.*, 1996) and Waves and Storms in the North Atlantic (WASA, Carretero *et al.*, 1998) projects. Coordinates for Haparanda, Härnösand, Wisby and Bodø are presented in Schmith *et al.* (1997), but the Bodø coordinates have not been used in our dataset since they point to a location just off the coast in the Saltfjorden. Brandt (1994a) provides details about the various locations of the Fanø/Nordby weather station over the years. The Tørshavn station was also at various locations (Brandt, 1994b) and was at the telegraph station during the period which overlaps with the DWRs used for this dataset (1907–1910).

The 1909/10 Annual Report gives some information regarding three French stations in the appendices. The Brest observations were taken at Pointe St. Mathieu, the Rochefort observations were taken on Ile d'Aix, and the Lorient

observations were taken on Ile de Groix. The 1912 DWRs are more specific regarding the Lorient station, highlighting that the observations were taken at the Bec Melen semaphore station on the island. The Bulletin International also provided some information regarding other stations in the DWRs. The observations from Paris, Lyon, Perpignan and Nice all took place at various observatories. Other documents from Meteo-France revealed that the Sanguinaire station was a lighthouse and the Biarritz station was a semaphore station.

Documents from the Infante D. Luiz Observatory in Lisbon (<http://sign.fc.ul.pt/anais.html>; also the site of the DWR observations) provided the barometer, rain gauge and dry bulb thermometer heights for Lisbon, Ponta Delgada and Madeira. Chazarra *et al.* (2012) describe the history of meteorological observations in the Azores and Madeira, revealing that the Ponta Delgada station was at the former Convent of Grace, Horta observations were taken at the civil government building, and the Madeira observations were recorded at Forte Pico in Funchal.

Wheeler (2011), after Hurst (1959), describes the history of meteorological observations in Gibraltar. In this dataset, there is only five months of observations for Gibraltar as it did not appear in the DWRs until August 1910. At the time, the station was located at South Bastion on the west of the peninsula close to the harbour. The 1909/10 Annual Report contains the height of the rain gauge in a list of stations from the colonies and dependencies, and the 1912 MWR included Gibraltar in its annual summary which provided the barometer height.

2.3 | Observing times

The observing times at the DWR stations were mostly steady throughout the time period considered for this dataset. Until December 1906, information regarding the observing times of each country was presented in tabular form on page 4 of the DWRs (Figure 4). No further information was published regarding observing times until November 1908 when some brief and unsubstantial notes were published on page 4. The most significant change in observing times occurred in July 1908 when the morning observing time at the GBI stations changed from 8 a.m. to 7 a.m. in order to bring these stations in line with the rest of Europe. All the observing times stored in this dataset are in UTC.

No observing times were published in the DWRs for the Icelandic and Faroese stations due to the omission of this information after December 1906. For Tørshavn, we have used the same observing times as Fanø (7 a.m. and 7 p.m. UTC) since the Faroe Islands were (and remain so) a Danish territory and therefore its meteorological observations were under the jurisdiction of the Danish Meteorological Institute (DMI). The Annual Reports confirm that the Icelandic stations were

TABLE SHOWING THE LOCAL STANDARD TIMES AND THE CORRESPONDING GREENWICH TIMES OF THE OBSERVATIONS RECORDED ON PAGE 1.				
COUNTRIES, &c.	Hours of Morning Observations by Local Standard Time.	Corresponding Greenwich Times.	Hours of Evening Observations by Local Standard.	Corresponding Greenwich Times.
British Isles	8 a.m. (West European)	8 a.m.	6 p.m.* (W.E.T.)	6 p.m.*
Holland	7 a.m. (" ")	7 a.m.	7 p.m. (W.E.T.)	7 p.m.
Germany, Denmark, Norway, Sweden }	8 a.m. (Mid-European)	7 a.m.	8 p.m. (M.E.T.)	7 p.m.
Belgium	7 a.m. (West European)	7 a.m.	Not published	—
France	7 a.m. (Paris) ...	6.51 a.m.	6 p.m. (Paris)	5.51 p.m.
Spain	8 a.m. (West European)	8 a.m.	6 p.m. (W.E.T.)	6 p.m.
Portugal	8 a.m. (Lisbon) ...	8.36 a.m.	8.23 p.m. (Lisbon)	4 p.m.
Azores { Ponta Delgada ...	6 a.m. (Local) ...	7.43 a.m.	9 p.m. (Local)	10.43 p.m.
{ Horta	6 a.m. (Local) ...	7.55 a.m.	9 p.m. (Local)	10.55 p.m.

FIGURE 4 An example of the table from page 4 of the DWRs showing the local and Greenwich observing times of the DWR stations from July 1906. This information was no longer provided in tabular form after December 1906, but from 8 November 1908, some information regarding observing times was provided on page 4 in the section titled ‘Notes referring to information given on page 1’. The asterisk besides the British Isles evening observing times was used to indicate that the evening observations at Oxford were taken at 8 p.m.

also under the jurisdiction of the DMI at the time but we have used the observing times from 1906 to 1919 data for Reykjavik provided by Trausti Jónsson (pers. comm.). The Icelandic observing times changed during the period considered for this dataset and were converted to UTC using the time zone changes documented at timeanddate.com.

Information regarding the observing times for Gibraltar and Madeira, both of which first appeared in the DWRs in 1910, was also difficult to find. From 1 August 1910, a handwritten note is included on page 4 of the DWRs stating ‘At Gibraltar the evening observations are taken at 9 p.m.’. However, no information is provided regarding the morning observations at Gibraltar so we have assumed that they were taken at 7 a.m. UTC as per the rest of the stations. We could not find any published information for the Madeira observing times, so we have assumed that observations were taken at 6 p.m. and 9 p.m. local time as per the Azores stations (Figure 4). These were telegraphed directly to London by the Meteorological Service of the Azores, and it is likely that a similar arrangement was made regarding the Madeira observations. The times for Madeira were also converted to UTC using the time zone information from timeanddate.com.

3 | DATA COLLECTION AND QUALITY CONTROL

3.1 | Method of data collection

The observations were digitized by 2,148 volunteers using the weatherrescue.org website between December 2017 and July 2018, following on from a previous successful digitization project using the same approach (Hawkins *et al.*, 2019) (H19). Each year from 1900–1910 was completed in turn so that the dataset would be as complete as possible for a set

period, depending on the success of the project. It was decided to finish the data collection when 1910 had been completed due to a drop in the number of active volunteers.

Each of the main DWR observation pages was split into 18 smaller pieces to allow the volunteers to transcribe a subset of the observations in each task (Sieber and Slonosky, 2019). The list of stations was separated into 6 similar-sized groups, and the observations were separated into three sets—morning, evening and daily. This image processing step to produce 18 images per page was a time-consuming process due to the scanned images not always being positioned or scaled in the same way on each page. A semi-automated process was built which assumed a fixed size and position of each group, but required the image position of the top left corner of the tabulated data to be manually identified. Further manual adjustments were required to ensure all the observations in each of the sub-images were visible as the page images were often scaled slightly differently.

Five separate volunteers were asked to transcribe the data shown in each smaller image by column. In H19, three volunteers per image were used, but more were thought to be necessary for the DWRs because the observations were handwritten rather than typeset. If four or more volunteers agreed on the value, then it was accepted, and if three agreed but the other two disagreed with each other, then the value was also accepted. In all other cases, the values were flagged as ‘#N/A’ and checked manually (see below). Fewer transcriptions would have meant faster transcription, but significantly more manual QC was necessary. This is a balance which will vary, depending on specific project circumstances.

As discussed above, some of the observations arrived too late to be included in the original documents, and some were later corrected. For the ‘additions and corrections’ pages, a separate volunteer approach was used. A spreadsheet template was made available to a set of volunteers who entered each of the pages manually and sent the completed template

to the project team. This provided an additional 53,000 observations and corrected errors in the original DWRs, of which there were approximately 30,000 over the 11-year period.

The data collected included two observations from the previous evening (pressure [p_{eve}] and $T_{\text{dry;eve}}$), three from the morning (pressure [p_{mor}], $T_{\text{dry;mor}}$ and T_{wet}) and three daily observations (T_{max} , T_{min} and total rainfall for the last 24 hr). Additional observations such as the wind strength and direction, sunshine and weather types were not rescued as they were deemed a lower priority. All pressure observations had been reduced to 32°F and sea level before transmission to the Met Office for collation.

3.2 | Quality control

After processing the volunteer transcriptions, a spreadsheet in CSV (comma separated values) format was produced for each day, representing a digital equivalent of the scanned document (Figure 1). Each spreadsheet was manually checked for flagged errors, and the corresponding entries in the DWRs were checked for legibility. If the corresponding DWR entries were clear, then the spreadsheet entry would be changed accordingly, but if the DWR entry was unclear (too difficult to read), then the spreadsheet entry was deleted from the final dataset. However, tendency values for the past 24 hr for p_{mor} and $T_{\text{dry;mor}}$ were published in the DWRs and could often be used to correct errors when the values of p_{mor} and $T_{\text{dry;mor}}$ themselves were illegible. After the errors for each year's spreadsheets had been removed, the additions and corrections were accounted for. The additions and corrections were stored as monthly spreadsheets similar to their original format (Figure 2) and could be easily inserted into the daily spreadsheets.

Once the additions and corrections had been accounted for, the following checks were made to the data:

1. If $T_{\text{wet}} > T_{\text{dry;mor}}$, then both values were deleted if they were clear in DWR; if one was unclear, then that value was deleted. A similar process was followed for (i) $T_{\text{min}} \geq T_{\text{max}}$, (ii) $T_{\text{dry;eve}} > T_{\text{dry;max}}$ and (iii) $T_{\text{dry;eve}} < T_{\text{min}}$. We did not make checks similar to (ii) and (iii) for $T_{\text{dry;mor}}$ since it is unclear if $T_{\text{dry;mor}}$ is included in T_{max} ; T_{min} for the previous or subsequent 24 hr.
2. Hyphens in the additions pages for rainfall were left blank in the spreadsheets as it was unclear if these referred to zero rainfall or no data.
3. Any data in the corrections pages with a question mark were deleted in the spreadsheet for the corresponding day.
4. Pressure observations were also compared to 20CRv3 (Slivinski *et al.*, 2019) to find any very large discrepancies (e.g. 1 inHg), and some values were corrected.

There are many occasions where $T_{\text{dry;eve}} > T_{\text{max}}$ and $T_{\text{dry;eve}} < T_{\text{min}}$ but where the values are clearly written—these have been retained instead of removing large quantities of data. Such instances of this were mostly from European stations; we speculate that this may be a result of errors when converting the units from °C to °F (Lempfert, 1954).

In the future, we will flag any suspicious values rather than delete them from the final dataset (Brönnimann *et al.*, 2019). This will also show which values the volunteers were unable to agree on. The raw version of this dataset has not been published, but it can be made available on request.

4 | THE DATASET

The data collection (Section 3.1) and QC (Section 3.2) procedures resulted in the recovery of 1,832,926 observations which consist of twice-daily mslp and T_{dry} , T_{wet} (morning only), T_{max} , T_{min} and rainfall for 72 stations across Europe (Figure 3). In 1900 only, fourteen GBI stations, along with Skudesnaes and Rochefort, sent 2 p.m. observations of sea-level pressure, T_{dry} and T_{wet} to the Met Office (Lempfert, 1954), and these have also been rescued.

4.1 | Pressure observations

4.1.1 | Comparing datasets

We have recovered 464,880 mslp observations, most of which are twice daily. Before 1 July 1908, the pressure observations at GBI stations were not corrected for the difference in gravity between the station location and 45°N. This was first noted in the DWR on 25 February 1906, and approximate gravity corrections were published in the DWR until 1 July 1908 from when it was noted that the gravity correction had now been applied to the GBI observations. Prior to 25 February 1906, no comment is made on this matter but we assume that the gravity correction had not been applied. We have not made any corrections to the observations and publish the data as it was reported in the DWRs.

The spatial coverage of pressure observations has increased substantially compared to ISPD for this time period (Figure 5). The most notable change in coverage is across England and Wales where there were previously no data in ISPD but the data from the DWRs have added 15 locations with twice-daily pressure observations. The DWRs have also provided data from thirteen new French stations and four Icelandic stations and partially filled a gap in the Azores ISPD record between 1888 and 1905.

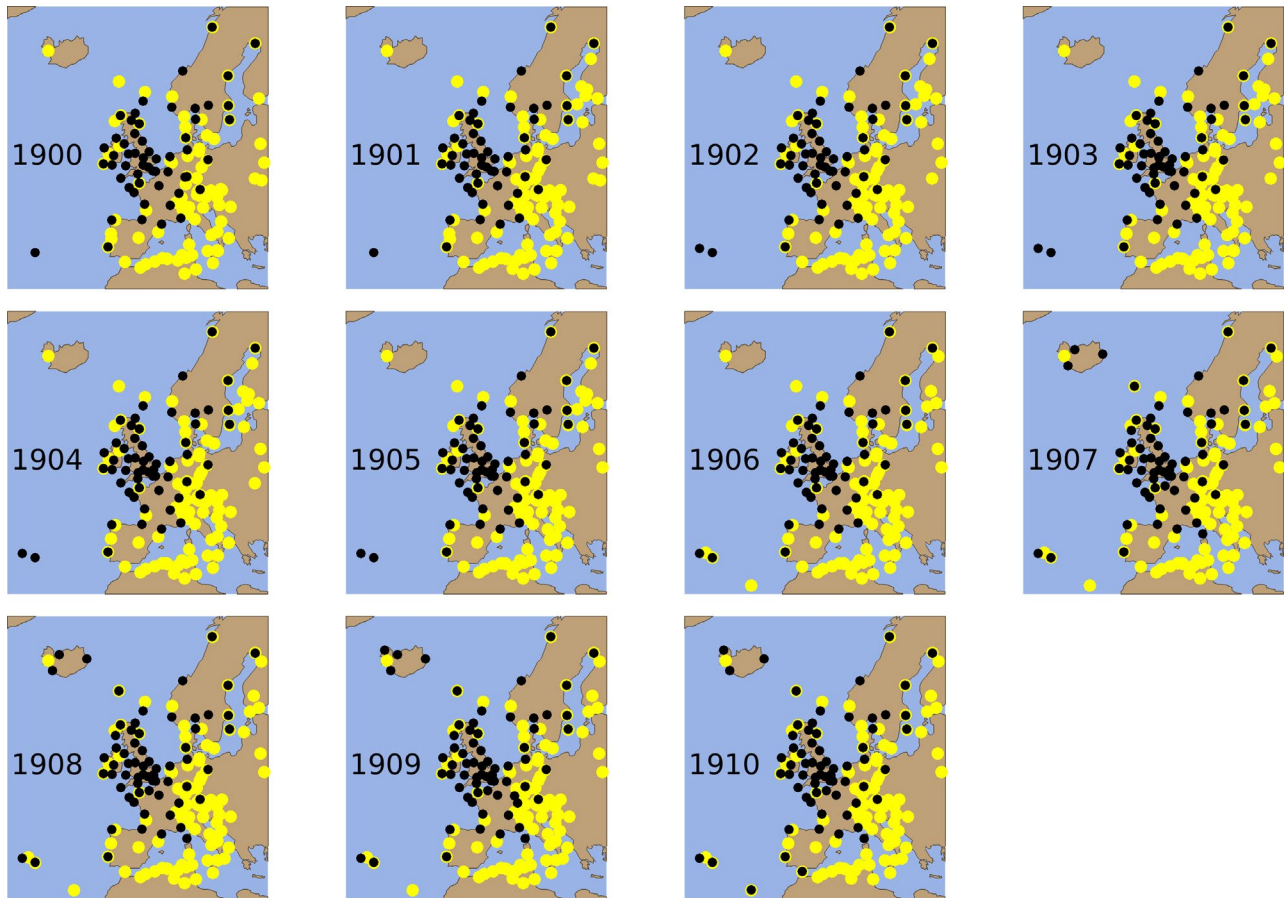


FIGURE 5 Land stations currently in ISPD (yellow) across Europe and the stations from the DWRs (black) for each year in 1900–1910. Some of the DWR stations are the same locations as data already in ISPD but are often at different times of day

The DWRs have helped to fill gaps in the ISPD record, such as at Aberdeen (Figure 6) where ISPD has no 1800 UTC observations between 1 January and 1 July 1900. The DWR observations, taken at the same time, agree well with the ISPD data for the rest of the year with a low RMSE of 0.86 hPa (less than 1.6 hPa which is considered an estimate of observational uncertainty in the 20CRv3 data assimilation system; Slivinski *et al.*, 2019) so can be used to fill this gap at Aberdeen. However, in 1909 at Aberdeen, the DWR data in May and November are offset by one day from the ISPD data—errors which are corrected at the end of both these months. We found no problems with the DWRs and suspect that this issue arose when the 1800 Aberdeen data were digitized prior to submission to ISPD—a day was missed either in the logbook or when the data were entered, and compensated for by repeating the final value of the month. Further errors in the Aberdeen ISPD data were identified by Wang *et al.* (2014) in response to Krueger *et al.* (2014) who claimed that pre-1950 20CR estimates of storminess from Wang *et al.* (2013) are erroneous. However, when Wang *et al.* (2014) corrected erroneous Aberdeen pressure values using the DWRs, they found that 20CR estimates of storminess do indeed agree well with observation-based proxies.

Some large differences can be seen between the two datasets. For example, at Stornoway (Figure 7) on 27 February 1903 there is a difference of about 35 hPa (~1 inHg) between the 0800 DWR and 0900 ISPD observations during a severe storm (Shaw, 1903; Hawkins *et al.*, 2019). The DWR observation (967.15 hPa) is more realistic than the ISPD value (1,002.13 hPa) in such a storm. The error in ISPD is likely to come from a typing error where the value was entered as 29 inHg rather than 28 inHg (Le Blancq, 2010). By using multiple volunteers for each value, the data collection method (Section 3.1) and QC (Section 3.2) prevent this sort of error. Smaller errors may be a result of an observer reading the barometer incorrectly or mistakes made when entering data such as mistaking a number 3 for an 8. The large error at Aberdeen on 3 December 1909 (Figure 6), where ISPD pressure decreases below 970 hPa and DWR pressure increases to 1,000 hPa, was caused by an error in the DWRs where the pressure was corrected to a value 1 inHg too high. This value was therefore deleted from the dataset.

At Stornoway, the ISPD data are systematically greater than the DWR data by about 1.1 hPa throughout 1903 and 1909 (Figure 7; true for all years in 1900–1910, not shown). This cannot be explained by the lack of gravity correction

before July 1908 (+0.03 inHg or ~1.02 hPa for Stornoway) as the difference persists after the gravity correction was added to the DWR pressure observations. We compared the DWR mslp from Stornoway to the original data submitted to ISPD (provided by Alistair Dawson, pers. comm.) which were recovered from documents of the Scottish Meteorological Society (SMS) for the years 1867–1921 and taken at 9 a.m. and 9 p.m. The offsets in Figure 7 are due to a gravity correction applied after submission to ISPD. After July 1908, this was unnecessary as the ISPD data from July 1894 exactly match the DWR observations. The ISPD data are also at the wrong times and location as the SMS appeared to stop taking observations at Stornoway in 1894.

4.1.2 | Comparing case studies

Figure 8 shows comparison of mslp from the 20CRv3 ensemble mean and DWR observations from two significant weather events: the February 1903 storm (Shaw, 1903) and an anticyclone in January 1902 which caused the highest mslp to ever be recorded in GBI (Burt, 2007). The z -scores,

$$z = \frac{\bar{p}_R - p_0}{\sigma}$$

where p_R is the 20CRv3 ensemble mean mslp, p_0 is the observed DWR value, and σ is the ensemble spread (standard deviation), show how many standard deviations separate the observation from the ensemble mean.

For the anticyclone (Figure 8a,c), the 20CRv3 ensemble mean underestimates the DWR observations by a small amount and the ensemble spread is low, presumably because there is little of any dynamic or thermodynamic significance for the model to simulate. The z -scores are quite large despite the proximity of the ensemble mean to the observations, a result of the low ensemble spread. In the case of the February 1903 storm (Figure 8b,d), the ensemble mean overestimates mslp by about 10 hPa near the centre of the cyclone with a huge ensemble spread trailing west from Scotland. Across Europe, where there are more ISPD data already available (Figure 5), the ensemble mean is much closer to observations and the ensemble spread is much lower. The erroneous observation in ISPD at Stornoway (Figure 7) during this storm was not rejected by the 20CRv3 data assimilation system since the ensemble spread is so large. An exploratory rerun of 20CRv3 for 1902 and 1903 with a pre-quality controlled version of this dataset (Laura Slivinski, pers. comm.) gives an improved representation of this storm with a deeper low in the ensemble mean and a reduced ensemble spread. The erroneous

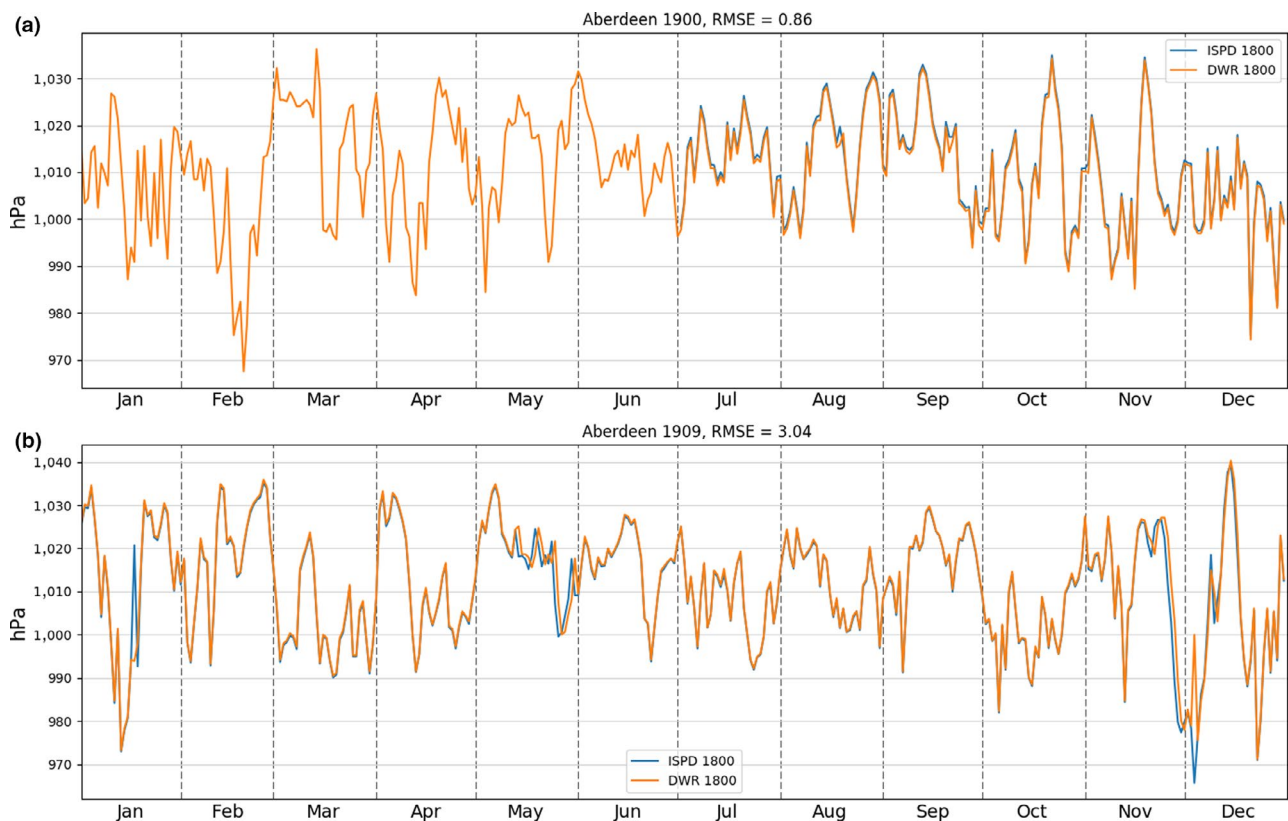


FIGURE 6 Daily evening DWR (orange) and ISPD (blue) mslp (in hPa) for Aberdeen in (a) 1900 and (b) 1909, with the root-mean-square error between the two datasets

Stornoway observation on 27 February was rejected in the exploratory 20CRv3 run, presumably because the added observations had better constrained the ensemble.

4.2 | Rainfall observations

A total of 222,292 rainfall observations were recovered from 67 stations in the DWRs—the Icelandic and Faroese stations did not report rainfall to the Met Office. Skudesnaes did not consistently report rainfall to the Met Office until 18 June 1905, although a small number of observations were published in the DWRs prior to this date. We decided to remove these few observations from the dataset as it is unclear whether or not these were genuine observations.

To highlight the usefulness of the additional rainfall observations, we examine the HadUK-Grid daily precipitation dataset (Hollis *et al.*, 2019) during October 1903 (Figure 9), which is the wettest month in the England and Wales precipitation record (https://www.metoffice.gov.uk/hadobs/hadukp/data/ranked_monthly/HadEWP_ranked_mly.txt; Wigley *et al.*, 1984). The spatial pattern of total rainfall across the UK for October 1903 is reasonably well represented by the 20CRv3 ensemble mean with South-East England receiving the least rainfall and the mountainous regions to the west being the wettest. However, 20CRv3 substantially underestimates the total

rainfall in mountainous regions which is likely a consequence of the model's horizontal resolution and a lack of pressure observations resulting in poor representation of weather systems. However, this sort of underestimation in mid-latitudes is to be expected in global reanalysis datasets (de Leeuw *et al.*, 2015).

Some of the DWR rainfall data are not yet included in HadUK-Grid or MIDAS. For example, the DWR observations for Stornoway are very different from the gridded data throughout 1903 (Figure 10) and MIDAS has no daily or hourly rain observations for the Outer Hebrides for this time period. The new Stornoway rainfall observations therefore represent a big improvement in coverage for the HadUK-Grid. The DWR observations for Oxford, however, are in very good agreement with HadUK-Grid (Figure 11) since MIDAS has similar daily observations from the same site (the Radcliffe Observatory; Smith, 1968; Burt and Burt, 2019). The DWR observations are often slightly less than the gridded data which may be an artefact of the interpolation used to create the gridded dataset or slightly different observing times for the MIDAS data.

4.3 | Temperature observations

We recovered a total of 1,145,802 temperature observations: 469,639 T_{dry} , 229,962 T_{wet} , 222,415 T_{max} and 220,744 T_{min} . Stations in Iceland and the Faroe Islands did not report T_{wet} .

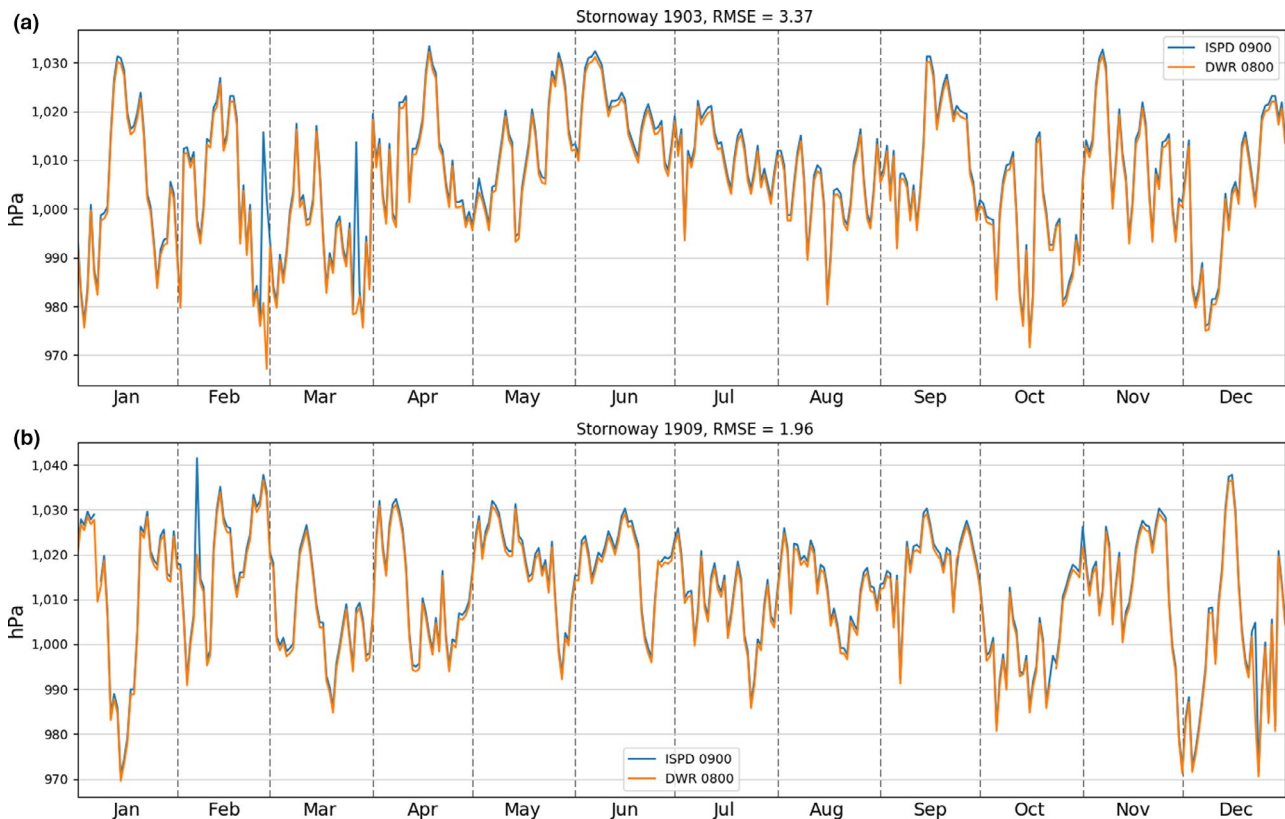


FIGURE 7 Daily morning DWR (orange) and ISPD (blue) mslp (in hPa) for Stornoway in (a) 1903 and (b) 1909, with the root-mean-square error between the two datasets

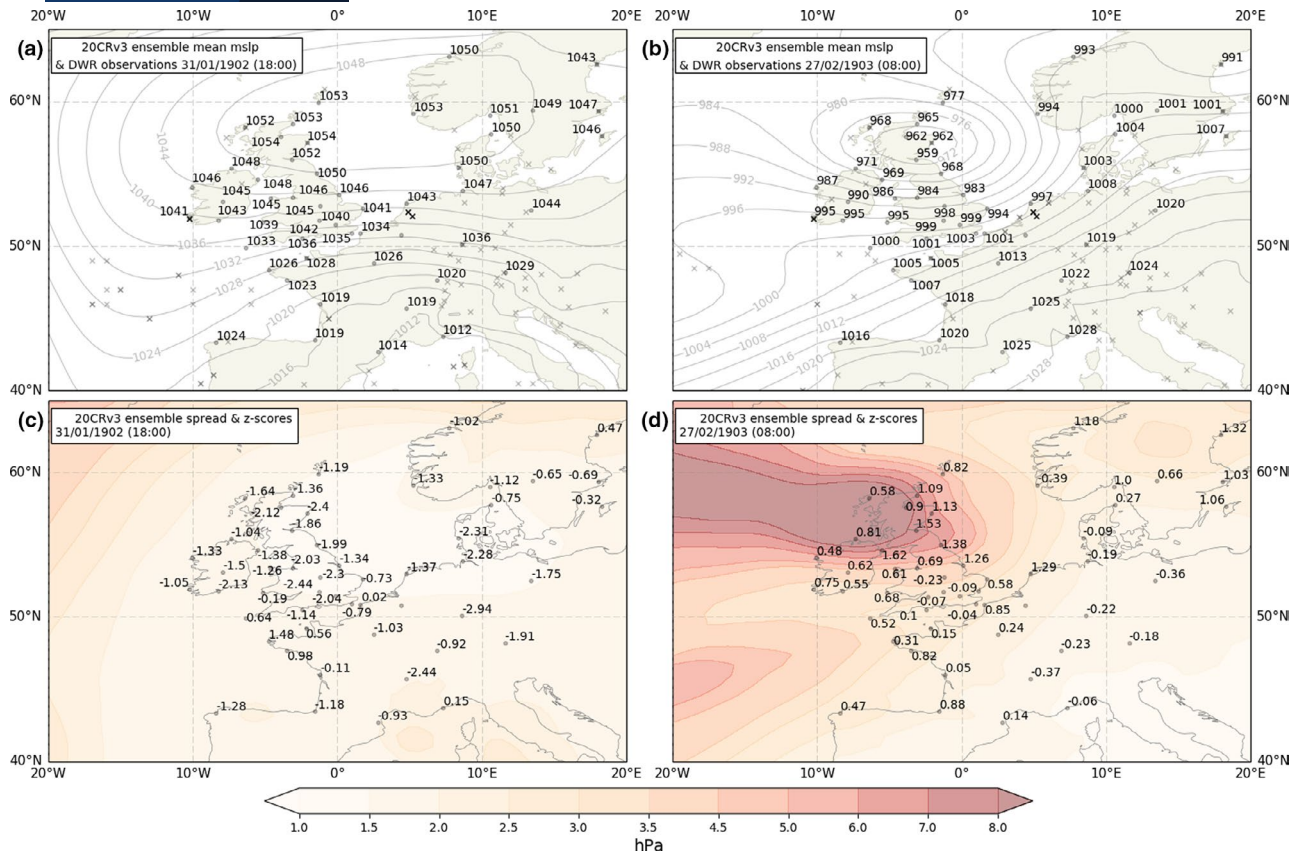


FIGURE 8 Ensemble mean 20CRv3 mslp (contours) and DWR mslp observations (grey circles and black numbers) with approximate gravity corrections applied (a,b) and 20CRv3 ensemble spread (contours) and z-scores (grey circles and black numbers) for 31 January 1902 anticyclone (a,c) and 27 February 1903 storm (b,d). The time stated in each panel is the time of the British and Irish DWR observations, although observations from Europe within 3 hr of the DWR observations are also included. The grey crosses in panels (a) and (b) indicate the locations of land and sea pressure observations in ISPD within three hours of the DWR observations. Darker crosses indicate multiple observations at these locations within three hours of the DWR observations. Units are hPa

$T_{\max}$ or $T_{\min}$ to the Met Office, and there is no $T_{\max}$ and $T_{\min}$ data for Karlstad. From 15 November 1908, the German stations replaced T_{wet} with relative humidity. We recovered a total of 3,042 relative humidity observations from the German stations.

No gridded dataset for daily UK temperatures is currently available for 1900–1910 (HadUK-Grid starts in 1960; Hollis *et al.*, 2019), but daily $T_{\max}$ and $T_{\min}$ for many stations are available in MIDAS. Only three DWR stations (Oxford, Bath and Liverpool) have temperature data in MIDAS during 1900–1910 with the record at Bath starting in 1904. As an example comparison, the DWR $T_{\max}$ at these stations agrees well with MIDAS during a heatwave in September 1906 with only some small discrepancies (Figure 12). These differences in $T_{\max}$ may have been caused by problems with transmission to the Met Office or errors when the data were transcribed for submission to MIDAS.

The 20CRv3 ensemble mean of $T_{\max}$ during the September 1906 heatwave was generally underestimated compared to the DWR observations (Figure 13). This is

particularly notable at the Scottish stations which fall inside contours 2–6°C above the observed temperature. At inland stations in the south of England (Bath, Oxford and London), where the heatwave peaks in GBI, $T_{\max}$ is underestimated less substantially than it is further north. Underestimation of $T_{\max}$ appears to occur across GBI during the fortnight centred around the heatwave. The maximum temperatures at Nairn and Nottingham are consistently greater than the ensemble mean and are even outside the ensemble range (Figure 14a,b). However, at coastal stations such as Nairn the 20CRv3 values in Figure 14 are calculated using bilinear interpolation, and at the horizontal resolution of 20CRv3 (~70 km), this will include using values from above the sea or more mountainous inland regions. Altitude may play a role in the biases at Nairn and Nottingham, but we cannot determine how significant a factor; this is because we do not know the thermometer heights for these stations and we do not currently have access to the 20CRv3 orography.

Not all locations show the same underestimation by 20CRv3. At Clacton (not shown), for example, DWR $T_{\max}$ is much closer to the ensemble mean, and during the peak of the

FIGURE 9 Total rainfall (in mm) across the United Kingdom for October 1903 in HadUK-Grid (left) and the 20CRv3 ensemble mean (right)

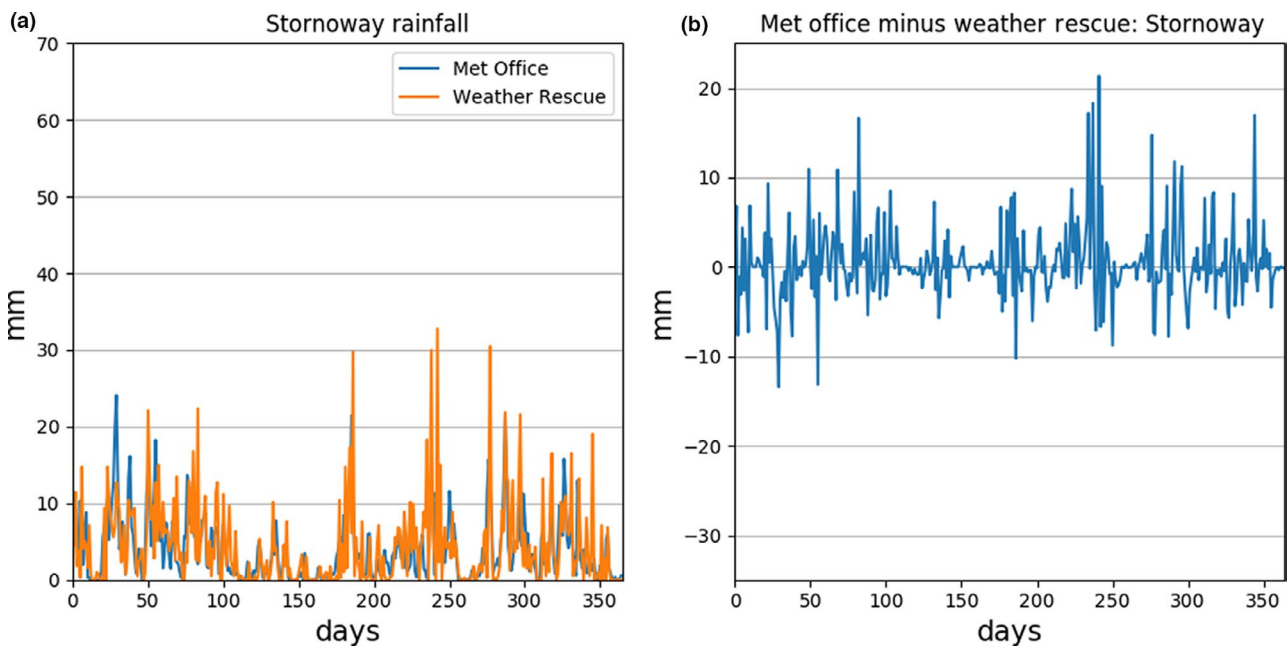
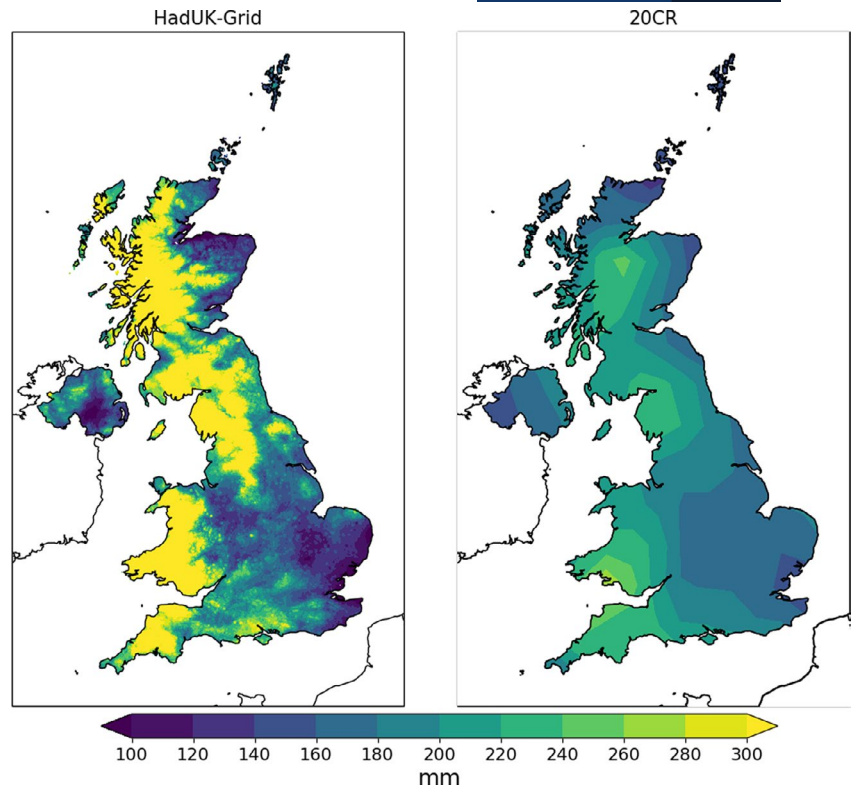


FIGURE 10 Daily rainfall (in mm) in 1903 at Stornoway with (a) comparison between HadUK-Grid (blue) and DWRs (orange) and (b) the difference between the two datasets as HadUK-Grid minus DWR

heatwave, the ensemble mean overestimates $T_{\max}$ by 4–5°C. At two stations on mainland Europe, the ensemble mean performs much better with remarkably good agreement at Paris and Berlin which is almost perfect at times and consistently within the ensemble spread (Figure 14c,d). Paris is not yet included in ISPD and has no other locations nearby in ISPD (Figure 5) which makes the performance of the 20CRv3 ensemble mean all the more remarkable.

5 | SUMMARY AND CONCLUSIONS

In this paper, we have described a dataset of 1,832,926 observations of mslp, temperature and rainfall at 72 locations around Europe which were recovered from the 1900–1910 Met Office DWRs by volunteers in the citizen science project Weather Rescue. Many of these observations are

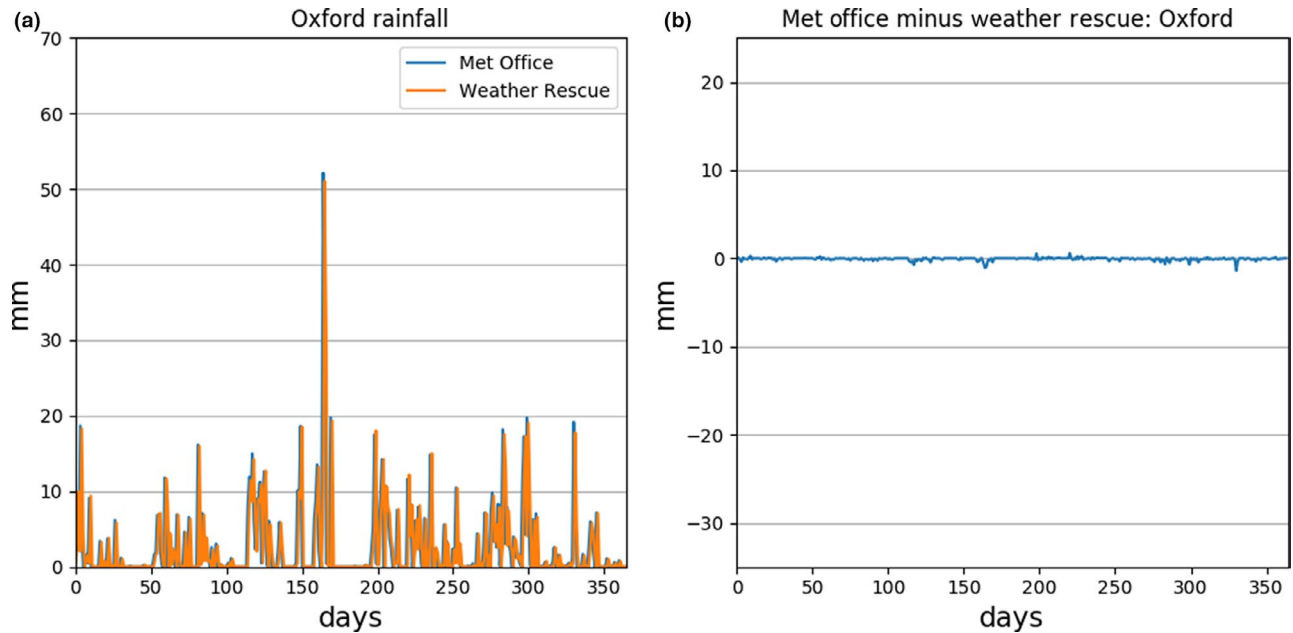


FIGURE 11 Daily rainfall (in mm) in 1903 at Oxford with (a) comparison between HadUK-Grid (blue) and DWRs (orange) and (b) the difference between the two datasets as HadUK-Grid minus DWR

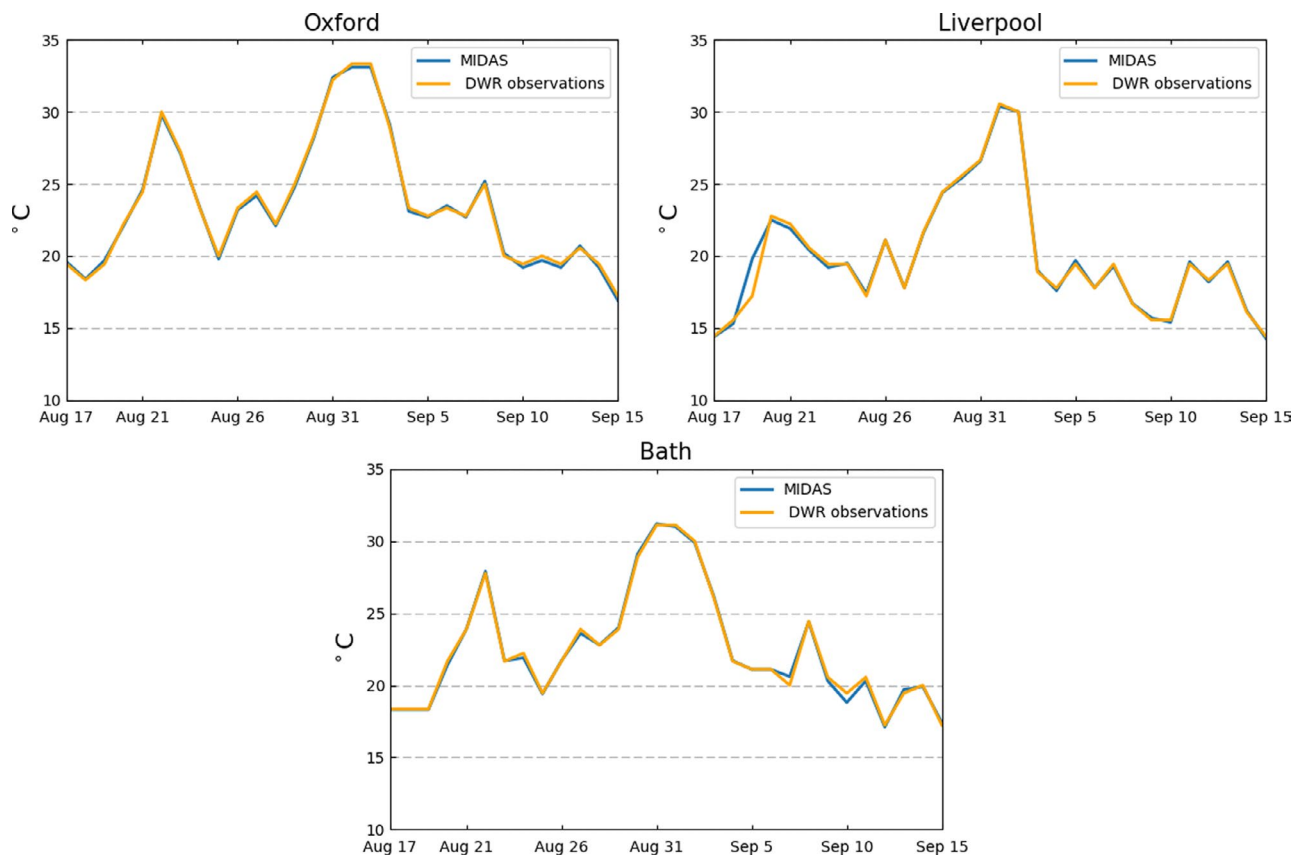


FIGURE 12 MIDAS (blue) and DWR (orange) T_{max} (in °C) during the August/September 1906 heatwave at (a) Oxford, (b) Liverpool and (c) Bath

not presently included in ISPD or MIDAS, and the addition of these observations will represent a significant improvement in spatial coverage for these datasets. For

example, in ISPD there are currently no observations of pressure anywhere in England and Wales in the years 1900–1910.

FIGURE 13 Ensemble mean 20CRv3 2m $T_{\max}$ (contours and grey numbers) and DWR observed $T_{\max}$ (black numbers and markers) on 02/09/1906. Units are $^{\circ}\text{C}$

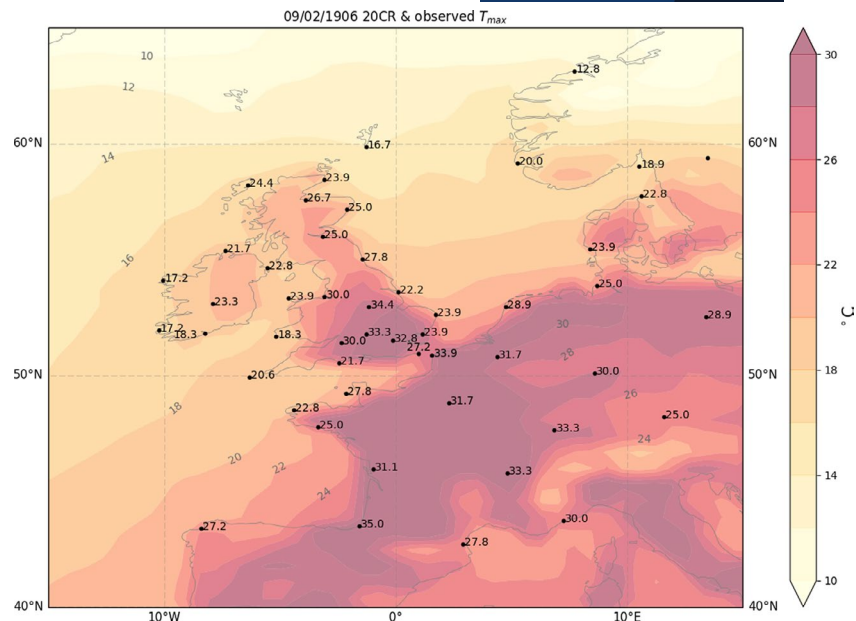
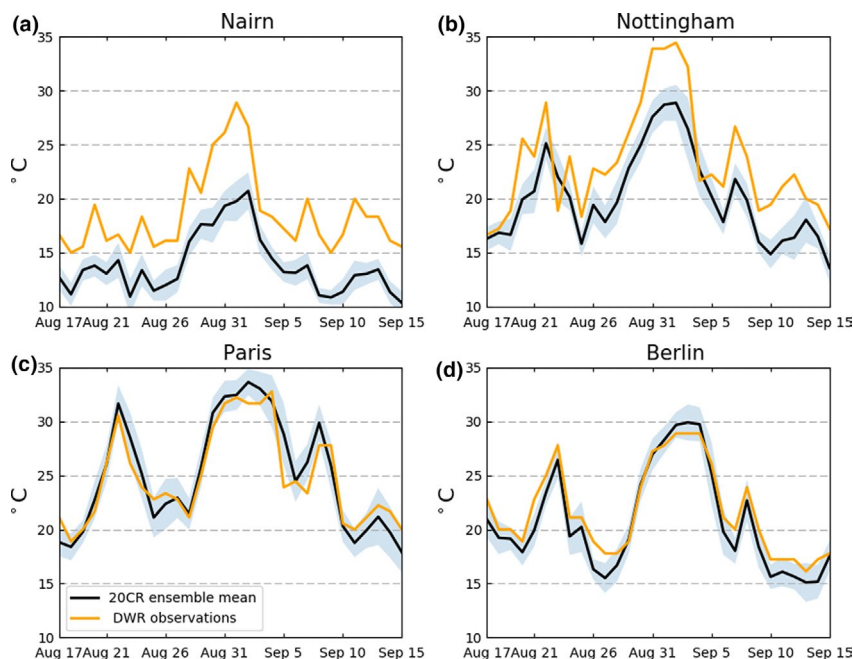


FIGURE 14 Observed DWR $T_{\max}$ (orange line) and ensemble mean 20CRv3 2m $T_{\max}$ (black line) with the ensemble spread (blue shading) at (a) Nairn, (b) Nottingham, (c) Paris and (d) Berlin. Units are $^{\circ}\text{C}$



We have also shown how effective the volunteers were at digitizing the DWRs. Various errors in ISPD at Aberdeen and Stornoway were found when comparing to the observations recovered for this dataset. Common errors that arise when digitizing handwritten records are to confuse digits (e.g., 3 and 8), repeat the previous observation or to type the wrong number (Le Blancq, 2010). By asking multiple volunteers to digitize the same data, we have minimized these errors and where the volunteers were unable to agree the values were flagged as errors and rechecked during QC. The citizen science project has therefore been a more efficient and accurate way to digitize large quantities of

handwritten weather observations than simply having one person doing this alone.

Various case studies have highlighted the importance of these new observations. The 20CRv3 ensemble mean mslp overestimates the pressure at the centre of a particularly significant storm in February 1903 (Shaw, 1903) by about 10 hPa and produces a large ensemble spread of more than 6 hPa (Figure 8). The inclusion of the new pressure observations from this dataset, and the correction of errors at Stornoway and Aberdeen, will help to reduce the ensemble spread in 20CRv3 and should give an ensemble mean closer to the observed values.

In addition to the 1900–1910 DWRs, the 1861–1874 DWRs have also been digitized by volunteers on weathrescue.org and pressure observations (3-hourly) from the 1919–1960 DWRs have also been digitized (Lisa Alexander, pers. comm.). Both these datasets are undergoing QC and will be published separately.

It is likely that a more bespoke transcription website would make the volunteer efforts slightly more efficient. We also chose five repeats per image which potentially slowed transcription but probably reduced the number of errors and the amount of manual QC required. This is a tricky balance, and the optimal choice will likely vary across projects. Automatic techniques to split images of large pages of data into different segments and read the text are ideally required to further speed up digitization of historical weather observations. Ensuring coordination across countries and projects to avoid duplication of effort is also necessary although there are benefits to two independent transcriptions of the same data from different sources.

ACKNOWLEDGEMENTS

We thank the 2148 volunteers, without whom the digitization of this dataset would not have been possible. We also thank Philip Brohan and the UK National Meteorological Archive for scanning and providing access to the DWRs. Gil Compo and Laura Slivinski provided early access to the 20CRv3 data. EH was funded by the UK National Centre for Atmospheric Science and by the ERA4CS INDECIS project with funding from the European Union Grant 690462. This work was also supported by the UK Research & Innovation (UKRI) Strategic Priorities Fund UK Climate Resilience programme. The programme is co-delivered by Met Office and NERC on behalf of UKRI partners AHRC, EPSRC and ESRC. Support for the Twentieth Century Reanalysis Project version 3 dataset is provided by the U.S. Department of Energy, Office of Science Biological and Environmental Research (BER), by the National Oceanic and Atmospheric Administration Climate Program Office and by the NOAA Earth System Research Laboratory Physical Sciences Division. This publication uses data generated via the Zooniverse.org platform, development of which is funded by generous support, including a Global Impact Award from Google, and by a grant from the Alfred P. Sloan Foundation.

DATA AVAILABILITY STATEMENT

The dataset is available from the Centre for Environmental Data Analytics (<https://catalogue.ceda.ac.uk/uuid/235ff4a040854dcd8dfb754bbb898479>) in Station Exchange Format (SEF) and CSV format. SEF files provide data for one variable at a particular location along with the station metadata (latitude/longitude coordinates and elevation), and this format is to become the international standard for storing historical

weather data. For storing in SEF format, the units of mslp, temperature and rainfall have been converted to hectopascals (hPa), degrees Celsius (°C) and millimetres (mm), respectively. More details regarding the structure and content of SEF files are available at the C3S Data Rescue GitHub repository (<https://github.com/C3S-Data-Rescue-Lot1-WP3/SEF/wiki>). The SEF files will also be submitted to the C3S Data Rescue Service (<https://data-rescue.copernicus-climate.eu/>).

ORCID

Philip M. Craig  <https://orcid.org/0000-0001-9213-4599>

Ed Hawkins  <https://orcid.org/0000-0001-9477-3677>

REFERENCES

- Befort, D.J., Wild, S., Kruschke, T., Ulbrich, U. and Leckebusch, G. (2016) Different long term trends of extra-tropical cyclones and windstorms in ERA-20C and NOAA-20CR reanalyses. *Atmospheric Science Letters*, 17, 586–595.
- Bench, A.T. (1981) Reminiscences of the Meteorological Office, 1898–1910. *The Meteorological Magazine*, 110, 323–329.
- Bloomfield, H.C., Shaffrey, L.C., Hodges, K.I. and Vidale, P.L. (2018) A critical assessment of the long-term changes in the wintertime surface Arctic Oscillation and Northern Hemisphere storminess in the ERA20C reanalysis. *Environmental Research Letters*, 13, 094004.
- Brandt, M.L. (1994a) *The North Atlantic Climatological Dataset: Dokumenteret stationhistorie for 25140 Nordby, Fanø*. Technical report, Danish Meteorological Institute. (in Danish).
- Brandt, M.L. (1994b) *The North Atlantic Climatological Dataset: Dokumenteret stationhistorie for klima- og synopstationer i Torshavn og Mykines, Færøerne 1872–1994*. Technical report, Danish Meteorological Institute. (in Danish).
- Brönnimann, S., Annis, J., Dann, W., Ewen, T., Grant, A.N., Griesser, T. *et al.* (2006) A guide for digitising manuscript climate data. *Climate of the Past*, 2, 137–144.
- Brönnimann, S., Allan, R., Ashcroft, L., Baer, S., Barriendos, M., Brázdil, R. *et al.* (2019) *Bulletin of the American Meteorological Society*, 100, ES389–ES413.
- Burt, S. (2007) The Highest of the Highs: Extremes of barometric pressure in the British Isles, Part 2 – the most intense anticyclones. *Weather*, 62, 31–41.
- Burt, S. and Burt, T. (2019) *Oxford weather and climate since 1767*. New York, NY: Oxford University Press.
- Carretero, J.C., Gomez, M., Lozano, I., de Elvira, A.R., Serrano, O., Iden, K. *et al.* (1998) Changing waves and storms in the northeast Atlantic? *Bulletin of the American Meteorological Society*, 79, 741–760.
- Chazarra, A., Barceló, A.M., Pires, V., Cunha, S., Silva, A., Marques, J. *et al.* (2012) *Climate Atlas of the Archipelagos of the Canary Islands, Madeira and the Azores: Air Temperature and Precipitation (1971–2000)*. Agencia Estatal de Meteorología de España and Instituto de Meteorologia, Portugal.
- Coll, J.R., van der Schrier, G., Aguilar, E., Rasol, D., Coscarelli, R. and Bishop, A. (2019) Data rescue of daily climate station-based observations across Europe. *Earth System Science Data Discussions*, 1–30.
- Compo, G.P., Whitaker, J.S., Sardeshmukh, P.D., Matsui, N., Allan, R.J., Yin, X. *et al.* (2011) The twentieth century reanalysis project. *Quarterly Journal of the Royal Meteorological Society*, 137, 1–28.

- Compo, G.P., Slivinski, L.C., Whitaker, J.S., Sardeshmukh, P.D., McColl, C., Brohan, P. *et al.* (2019) *The International Surface Pressure Databank version 4*. Research Data Archive at the National Center for Atmospheric Research, Computational and Information Systems Laboratory, Boulder, CO.
- Cram, T.A., Compo, G.P., Yin, X., Allan, R.J., McColl, C., Vose, R.S. *et al.* (2015) The International Surface Pressure Databank version 2. *Geoscience Data Journal*, 2, 31–46.
- Frich, P., Alexandersson, H., Ashcroft, J., Dahlström, B., Demarée, G.R., Drebs, A. *et al.* (1996) *North Atlantic climatological dataset (NACD version 1) - final report*. Technical report, Danish Meteorological Institute.
- Geddes, A.E.M. (1955) The development of the study and practice of meteorology at Aberdeen. *Weather*, 10, 385–389.
- Hawkins, E., Burt, S., Brohan, P., Lockwood, M., Richardson, H., Roy, M. and *et al.* (2019) Hourly weather observations from the Scottish Highlands (1883–1904) rescued by volunteer citizen scientists. *Geoscience Data Journal*, 6, 160–173.
- Hegerl, G., Brönnimann, S., Schurer, A. and Cowan, T. (2018) The early 20th century warming: Anomalies, causes, and consequences. *Wiley Interdisciplinary Reviews: Climate Change*, 9, e522.
- Hollis, D., McCarthy, M., Kendon, M., Legg, T. and Simpson, I. (2019) HadUKGrid: a new UK dataset of gridded climate observations. *Geoscience Data Journal*, 6, 151–159.
- Hurst, G.W. (1959) A brief history of meteorology in gibraltar. *Weather*, 14, 41–46.
- Krueger, O., Feser, F., Bärning, L., Kaas, E., Schmith, T., Tuomenvirta, H. *et al.* (2014) Comment on Trends and low frequency variability of extratropical cyclone activity in the ensemble of twentieth century reanalysis by Xiaolan L. Wang, Y. Feng, G. P. Compo, V. R. Swail, F. W. Zwiers, R. J. Allan, and P. D. Sardeshmukh, *Climate Dynamics*, 2012. *Climate Dynamics*, 42, 1127–1128.
- Le Blancq, F. (2010) Rescuing old meteorological data. *Weather*, 65, 277–280.
- de Leeuw, J., Methven, J. and Blackburn, M. (2015) Evaluation of ERA-Interim reanalysis precipitation products using England and Wales observations. *Quarterly Journal of the Royal Meteorological Society*, 141, 798–806.
- Lempfert, R.G.K. (1913) British weather forecasts: Past and present. *Quarterly Journal of the Royal Meteorological Society*, 39, 173–184.
- Lempfert, R.G.K. (1954) Some Reminiscences of the Meteorological Office of the Year 1902. *The Meteorological Magazine*, 83, 161–166.
- Marriott, W. (1901) Inspection of Stations 1901. *Quarterly Journal of the Royal Meteorological Society*, 27, 206–212.
- Met Office (2019) *Met Office MIDAS open: UK land surface stations data (1853-current)*. <https://catalogue.ceda.ac.uk/uuid/dbd451271eb04662beade68da43546e1>. [Accessed 25th February 2019].
- Moberg, A., Bergström, H., Kringsman, J.R. and Svanred, O. (2002) Daily air temperature and pressure series for Stockholm (1756–1998). *Climatic Change*, 53, 171–212.
- Murphy, E.J. (1990) A brief history of Valentia Observatory. *Weather*, 45, 266–268.
- Reynolds, G. (1954) Weather records at the Liverpool Observatory. *Weather*, 9, 233–241.
- Rohrer, M., Brönnimann, S., Martius, O., Raible, C.S. and Wild, M. (2019) Decadal variations of blocking and storm tracks in centennial reanalyses. *Tellus A*, 71, 1–21.
- Ryan, C., Duffy, C., Broderick, C., Thorne, P.W., Curley, M., Walsh, S. *et al.* (2018) Integrating data rescue into the classroom. *Bulletin of the American Meteorological Society*, 99, 17571764.
- Schmith, T., Alexandersson, H., Iden, K. and Tuomenvirta, H. (1997) *North Atlantic-European pressure observations 1868–1995 (WASA dataset version 1.0)*. Technical report, Danish Meteorological Institute.
- Shaw, W.N. (1903) The meteorological aspects of the storm of february 26–27, 1903. *Quarterly Journal of the Royal Meteorological Society*, 29, 233–262.
- Sieber, R. and Slonosky, V. (2019) Developing a flexible platform for crowdsourcing historical weather records. *Historical Methods: A Journal of Quantitative and Interdisciplinary History*, 52, 164–177.
- Slivinski, L.C., Compo, G.C., Whitaker, J.S., Sardeshmukh, P.D., Giese, B.S., McColl, C. *et al.* (2019) Towards a more reliable historical reanalysis: Improvements for version 3 of the Twentieth century Reanalysis system. *Quarterly Journal of the Royal Meteorological Society*, 145, 2876–2908.
- Slonosky, V., Sieber, R., Burr, G., Podolsky, L., Smith, R., Bartlett, M. *et al.* (2018) From books to bytes: A new data rescue tool. *Geoscience Data Journal*, 6, 58–73.
- Smith, C.G. (1968) The Radcliffe Meteorological Station. *Weather*, 23, 362–367.
- Thorne, P.W., Allan, R.J., Ashcroft, L., Brohan, P., Dunn, R.J.H., Menne, M.J. *et al.* (2017) Toward an integrated set of surface meteorological observations for climate science and applications. *Bulletin of the American Meteorological Society*, 98, 2689–2702.
- Walker, M. (2011) *History of the Meteorological Office*. New York, NY: Cambridge University Press.
- Wang, X.L., Feng, Y., Compo, G.P., Swail, V.R., Zwiers, F.W., Allan, R.J. *et al.* (2013) Trends and low frequency variability of extra-tropical cyclone activity in the ensemble of twentieth century reanalysis. *Climate Dynamics*, 40, 27752800.
- Wang, X.L., Feng, Y., Compo, G.P., Zwiers, F.W., Allan, R.J., Swail, V.R. (2014) Is the storminess in the Twentieth Century Reanalysis really inconsistent with observations? A reply to the comment by Krueger *et al.* (2013b). *Climate Dynamics*, 42, 1113–1125.
- Wheeler, D. (2011) The Gibraltar climatic record Part 3. *Temperature. Weather*, 66, 259–265.
- Wigley, T.M.L., Lough, J.M. and Jones, P.D. (1984) Spatial patterns of precipitation in England and Wales and a revised, homogeneous England and Wales precipitation series. *International Journal of Climatology*, 4, 1–25.
- Yang, W., Vecchi, G.A., Fueglistaler, S., Horowitz, L.W., Luet, D.J., Muñoz, Á.G. *et al.* (2019) Climate impacts from large volcanic eruptions in a high resolution climate model: The importance of forcing structure. *Geophysical Research Letters*, 46, 7690–7699.
- Yaskell, S.H. (2008) From research institution to astronomical museum: A history of the Stockholm Observatory. *Journal of Astronomical History and Heritage*, 11, 146–155.

How to cite this article: Craig PM, Hawkins E. Digitizing observations from the Met Office Daily Weather Reports for 1900–1910 using citizen scientist volunteers. *Geosci Data J.* 2020;00:1–19. <https://doi.org/10.1002/gdj3.93>