

Millions of historical monthly rainfall observations taken in the UK and Ireland rescued by citizen scientists

Article

Published Version

Creative Commons: Attribution 4.0 (CC-BY)

Open Access

Hawkins, E. ORCID: <https://orcid.org/0000-0001-9477-3677>,
Burt, S. ORCID: <https://orcid.org/0000-0002-5125-6546>,
McCarthy, M., Murphy, C., Ross, C., Baldock, M., Brazier, J.,
Hersee, G., Huntley, J., Meats, R., O'Grady, J., Scrimgeour, I.
and Silk, T. (2023) Millions of historical monthly rainfall
observations taken in the UK and Ireland rescued by citizen
scientists. *Geoscience Data Journal*, 10 (2). pp. 246-261.
ISSN 2049-6060 doi: <https://doi.org/10.1002/gdj3.157>
Available at <https://centaur.reading.ac.uk/104088/>

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/gdj3.157>

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



DATA ARTICLE

Millions of historical monthly rainfall observations taken in the UK and Ireland rescued by citizen scientists

Ed Hawkins¹ | Stephen Burt¹ | Mark McCarthy² | Conor Murphy³ |
Catherine Ross⁴ | Mike Baldock⁵ | John Brazier⁵ | Gill Hersee⁵ |
Jacqui Huntley^{5,6} | Richard Meats⁵ | John O'Grady⁵ | Ian Scrimgeour⁵ | Tim Silk⁵

¹Department of Meteorology, National Centre for Atmospheric Science, University of Reading, Reading, UK

²National Climate Information Centre, Exeter, UK

³Department of Geography, Irish Climate Analysis and Research UnitS (ICARUS), Maynooth University, Maynooth, Ireland

⁴National Meteorological Archive, Exeter, UK

⁵RainfallRescue.org, Reading, UK

⁶Department of Archaeology, University of Durham, Durham, UK

Correspondence

Ed Hawkins, National Centre for Atmospheric Science, Department of Meteorology, University of Reading, Reading, UK.
Email: e.hawkins@reading.ac.uk

Funding information

Natural Environment Research Council, Grant/Award Number: NE/S015574/1; National Centre for Atmospheric Science

Abstract

Recovering additional historical weather observations from known archival sources will improve the understanding of how the climate is changing and enable detailed examination of unusual events within the historical record. The UK National Meteorological Archive recently scanned more than 66,000 paper sheets containing 5.28 million hand-written monthly rainfall observations taken across the UK and Ireland between 1677 and 1960. Only a small fraction of these observations were previously digitally available for climate scientists to analyse. More than 16,000 volunteer citizen scientists completed the transcription of these sheets of observations during early 2020 using the RainfallRescue.org website, built using the Zooniverse platform. A total of 3.34 million observations from more than 6000 locations have so far been quality controlled and made openly available. This has increased the total number of monthly rainfall observations

Dataset

Identifier: 10.5281/zenodo.5770389

Creator: Ed Hawkins

Dataset correspondence: e.hawkins@reading.ac.uk

Title: Rainfall Rescue v1.1.0

Publisher: Zenodo

Publication year: 2021

Resource type: Dataset

Version: 1.1.0

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.

© 2022 The Authors. *Geoscience Data Journal* published by Royal Meteorological Society and John Wiley & Sons Ltd.

that are available for this time period and region by a factor of six. The newly rescued observations will enable longer and much improved reconstructions of past variations in rainfall across the British and Irish Isles, including for periods of significant flooding and drought. Specifically, this data should allow the official gridded monthly rainfall reconstructions for the UK to be extended back to 1836, and even earlier for some regions.

KEYWORDS

atmospheric science, climate, citizen science, data rescue, rainfall

1 | INTRODUCTION

In 1677, Richard Townley started regular measurements of the amount of rain that fell at Townley Hall, near Burnley in northern England (Folland & Wales-Smith, 1977). His measurements continued until 1704, and other interested observers subsequently took up the challenge of measuring rainfall. The equipment to record rainfall was slowly developed and, by 1820, there were a few dozen rain gauge sites across the UK and Ireland. Many of these historical observations were taken by volunteer observers who published the data in private diaries, academic journals (e.g., Townley, 1694) or other publications.

By 1860, when George Symons, in the form of the British Rainfall Organization (BRO), began coordinating the collection of rainfall data (Glasspoole, 1952; Pedgley, 2010; Walker, 2010), there were several hundred rain gauges being diligently monitored, and this number increased rapidly over the following 30 years. By 1919, there were a few thousand operational rain gauges, many published in the annual volumes of *British Rainfall* (Burt, 2010), and the BRO became part of the UK Met Office. The archives of the BRO were transferred to the Met Office, but the vast majority of the rainfall observations taken before 1961 have never been digitized from the original paper copies. After 1961, the digitized data are far more complete.

Using the rainfall data currently digitally available, it is possible to produce an estimate of average rainfall for England & Wales back to 1766 (Wigley et al., 1984) and to reconstruct the spatial pattern of rainfall for every month back to 1862 (Hollis et al., 2019). However, the available data represents only a small fraction of what could be available, particularly for the period before 1961. For the island of Ireland, there are reconstructions available back to 1711, which include documentary evidence (Murphy et al., 2018), and daily rainfall data continues to be recovered from Irish archives (Ryan et al., 2018, 2021).

This paper describes the digitization of an enormous archive of rainfall observations known as ‘the Ten Year rainfall sheets’. These sheets include monthly rainfall amounts,

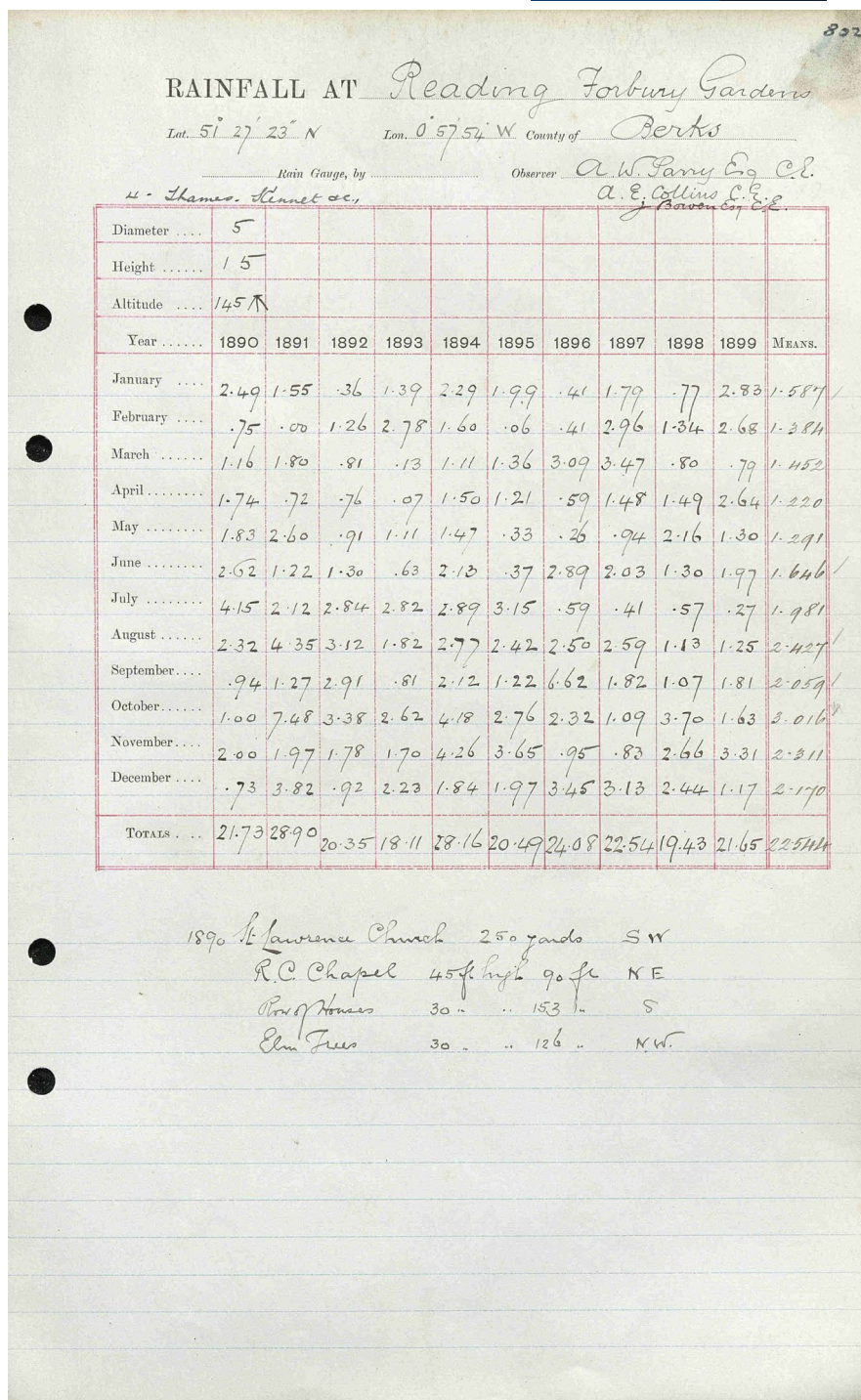
measured across the UK, Ireland and the Channel Islands between 1677 and 1960, although data for locations in the Republic of Ireland are not available after 1940 when responsibility for the country’s meteorological records transferred to Met Éireann. The transcription of these observations from their paper format into digital data will transform our understanding of historical weather variations but would have taken several person-years of effort using traditional manual techniques. The COVID-related lockdown in early 2020 provided a unique opportunity to ask for help in transcribing the observations from volunteer members of the public who had additional spare time and the desire to usefully contribute to science.

2 | THE TEN YEAR RAINFALL SHEETS

The Ten Year rainfall sheets are held in the UK National Meteorological Archive and were historically maintained by the British Rainfall Organization and then the Rainfall Branch of the Meteorological Office (Craddock, 1976). Each file consists of loose-leaf pages which followed a design laid out by George Symons (see Figure 1). Each pre-printed form gives the monthly and annual rainfall totals for the relevant decadal period and has space for metadata on the type and site of the rain gauge and some location data. This metadata is not always completed. Stations were initially arranged within each file by county and the files themselves were formed into decadal blocks (hence ‘Ten Year Sheets’). From 1910, each station was given a number, which brought stations within each river catchment area together, and this concept was further improved in 1951, but prior to 1910, stations could be found anywhere within the relevant county.

The data falls into two sets comprising data from 1677-1886 and data from 1860-1960. The first series is compiled from a number of different published and unpublished sources including Luke Howard’s ‘Climate of London’, the *Gentleman’s Magazine*, the *Philosophical Transactions of*

FIGURE 1 An example Ten Year rainfall sheet for Reading Forbury Gardens for the 1890s giving the location, county, directions to landmarks (bottom), station number (top right), observer's names, type of rain gauge and altitude. The rainfall amounts are in inches with an annual total (bottom row), and the decadal average of the month (right column). The decadal averages are not always present, and much of the metadata such as latitude and longitude is often not given (or reliable) on early sheets. Later in the series, more metadata are included, with some rainfall inspector notes starting to appear in the early 1900s and site National Grid References from the 1950s



the Royal Society and the *Manchester Memoirs* (Craddock, 1976). The second series consists of data received directly into the British Rainfall Organization to 1919, and the Meteorological Office thereafter, from observers and station authorities, and clerical archives maintained by BRO/Met Office staff until their manual completion ceased after the 1980s, as computerized records were more readily prepared.

Data collected for, and sent directly to, BRO or the Meteorological Office was subject to a level of quality

control and verification, although where records have been amended or added, there is often not a record of when or by whom. For example, a small number of sheets contain adjustments to the values without reasons given. In these cases, we have used the adjusted values. There is a greater risk of inaccuracy in data drawn from materials first published elsewhere, for which the original observations cannot be verified (Craddock, 1976). This accounts for almost all of the pre-1820 data. While references to the original source are often given, Craddock notes that

in many cases metadata on the rain gauge site was not transcribed. To reduce this loss of information, he created an index of early sites giving some information on the source, site or observer, and this index is held in the National Meteorological Archive, although the individual ten-year sheets often contain more information than is in the index. Craddock (1976) discusses several of the challenges in creating homogeneous series from data before and after 1820.

The physical arrangement of the data, resulting in observations from a single station being spread through the entire series of files, made it somewhat complex to exploit this unique data source. Early attempts to identify long-period variations in rainfall included those by Symons (1865), Salter (1921), Nicholas and Glasspoole (1931) and Glasspoole (1941), while in the mid-1970s, Craddock attempted to create a series of homogenous rainfall records representing different districts of Britain.

The potential for analysis after digitization meant the Ten Year rainfall sheets were identified as a priority for scanning by the UK National Meteorological Archive in 2018. Scanning of all data and the Craddock index was completed in 2019 and the records were uploaded to the Met Office Digital Library and Archive (2020).

3 | RAINFALL RESCUE

The volume of data contained within the Ten Year sheets meant that a traditional manual transcription approach would have taken several person-years of effort. The availability of scanned copies of the Ten Year sheets enabled the creation of a citizen science project to ask volunteers to transcribe the observations into digital form more efficiently.

3.1 | Building the web platform

The Zooniverse platform (www.zooniverse.org) offers a set of flexible tools to enable the construction of online citizen science projects, which ask for volunteers to complete numerous simple but highly useful tasks. Several projects have previously used the Zooniverse to successfully digitize historical weather observations (e.g., oldWeather.org, JungleWeather.org; SouthernWeatherDiscovery.org; [Climate History Australia](http://ClimateHistoryAustralia.org)). The WeatherRescue.org project is described in Hawkins et al. (2019) and Craig and Hawkins (2020). Other online citizen science projects have been highly successful, such as the DRAW project (Sieber & Slonosky, 2019). However, the number of observations contained in the Ten Year rainfall sheets is considerably

larger than any of these previous citizen science digitization efforts.

In early 2020, the RainfallRescue.org website was designed and built to enable the transcription of the Ten Year rainfall sheets (see example in Figure 1). The website was launched in March 2020 and the project received widespread media interest (e.g., Amos, 2020; Harvey, 2020). Within just 16 days the entire set of 66,000 sheets had been successfully transcribed, with more than 16,000 volunteers contributing. The Zooniverse reported this as being the most successful project they had ever hosted in terms of the number of volunteers involved and the tasks completed.

3.2 | Website design choices

The 66,000 sheets were of a uniform design, which enabled a single set of questions to be asked of the volunteers. Volunteers were asked to select a particular year of interest and were shown a station sheet which included that year. They were then asked to transcribe the 12 monthly rainfall amounts and the annual total for that year, before being shown another sheet to transcribe. One key decision was on the number of repeat transcriptions per observation that we would require to ensure individual typing errors were largely eliminated. The factors in this decision included: the handwriting was generally readable, there was a cross-check available on the values from adding up the transcribed monthly values to compare with the transcribed annual total, and the ability to increase the number of repeats if required. We chose 4 repeats per value and ensured that if 3 or more volunteers agreed on the transcribed monthly or annual value that it would be provisionally accepted. Around 98% of the values were initially accepted with this consensus choice. The remaining 2% of tasks were added back onto the website to obtain additional transcriptions. These additional transcriptions resulted in around 99.5% consensus, with the remaining values being genuinely hard to read, or from columns where there had been a mistake in the addition on the original hand-written page. In total, around 23 million rainfall amounts were typed by the volunteers.

As well as the data itself, the location information also required transcription. Separate tasks asked the volunteers to type the location name, the grid reference (if available) and the station number (if available). In hindsight, the observer's name should have also been requested as this ended up being a useful way to link different sheets together and identify locations through census information or other historical sources (see below). Overall, we estimated that around 100 million keystrokes were typed by the 16,000 volunteers in 16 days.

After the initial transcription phase, the volunteers were asked to help with additional tasks using online collaborative spreadsheets. Around 100 volunteers contributed to re-examining the individual pages where there had not been a consensus amongst the original transcribers on the location name and resolved these disagreements, usually performed by one volunteer per sheet. Many of the grid references written on the sheets themselves were accurate, but some were not, presumably due to imperfect mapping at the time. Likely grid references were then identified through mapping research by the same subset of volunteers, although this task was only performed for the sheets for 1900-1909 and 1951-1960.

3.3 | Raw data output

A total of 5.28 million unique observations were transcribed, starting with Richard Townley's observations in 1677, and ending with observations from thousands of rain gauges in 1960. This number of raw observations can be compared with the 0.51 million observations for the 1862-1960 period and 3.39 million observations for 1862-2019 period, which are currently used in the Met Office rainfall reconstruction dataset called HadUK-Grid (Hollis et al., 2019). The raw Rainfall Rescue data therefore increases the number of observations available for the pre-1961 period by an order of magnitude and more than doubles the total number of monthly rainfall observations, which are digitally available for the UK (Figure 2). Figure S1 shows the distribution of rainfall sheets by country and by county.

There are some small caveats to this comparison of the number of observations available. Some of the sheets are for locations in the Channel Islands and what is now the Republic of Ireland, both of which are not part of the HadUK-Grid dataset. Observations from the Republic of Ireland are not included on the sheets for the 1940s and 1950s as Met Éireann was responsible for coordinating those observations after its formation in 1936. Those data have already been digitized and are openly available. In addition, pre-1940 daily rainfall amounts for the Republic of Ireland have also been digitized (Ryan et al., 2021). Further work will ensure that the disparate datasets from the Republic of Ireland are reunited. Many of the Ten Year rainfall sheets are for locations with just a handful of observations, and often less than a decade. These stations are less useful for creating local time series of rainfall variations but can still be helpful in larger scale reconstructions (Simpson & McCarthy, 2019). However, the effort required to identify precise coordinates for each site means there has been a focus on the longer series.

3.4 | Location concatenation and quality control

The raw data is not usable for long-term climate analysis as it consists of ten year segments from individual stations, many of which have no location information. A few dedicated volunteers (who are also co-authors on this paper) collated the rainfall time series for more than 6000 locations across the UK and Ireland by combining the data from different decades.

This process required identifying those sheets that were linked together by being observations from the same location. Often the location names written on the sheets would change slightly from decade to decade and so were not a perfectly reliable method of linking sheets. The observer's name was often a helpful clue to which sheet was a continuation of a previous decade.

Once the linked sheets were identified, the transcribed data from each sheet had to be combined into a single spreadsheet format. In the published version of the dataset, the best estimate of the location of each rain gauge is given with a grid reference, and the corresponding latitude, longitude and approximate elevation. To identify the grid reference, many of the locations required additional research into the history of the named site or searching ancestry and census archives to identify observers and their homes. Old maps (e.g. Ordnance Survey 1900s series, see later) contain some marked rain gauges that also helped identify locations more precisely. Additional clues on many of the Ten Year sheets included distance and direction to local landmarks such as churches and railway stations (see Figure 1). Such details would have been vital in the early years, because site inspectors would require them to find the sites when visiting. Many sites were run by interested volunteers, located at vicarages, schools, mental hospitals, reservoirs, water works, lighthouses, parks, hills, manor houses, on canals, at railway stations and even royal palaces. Industrial sites were also used occasionally, such as collieries, iron and steelworks, flax mills and a chocolate factory.

Once the decades were joined together, additional quality control measures were applied, namely: (1) removing values that were described on the sheets as estimated, (2) ensuring that the sum of monthly values equalled the annual total, (3) identifying rain gauge moves, (4) removing values that were copied or duplicated from other locations and (5) resolving the remaining values for which the original transcriptions had disagreed. In addition, there were occasions where the written measurement is an accumulation over two or more months. These are indicated by curly brackets on the original Ten Year sheets and are given as -999 in the final dataset.

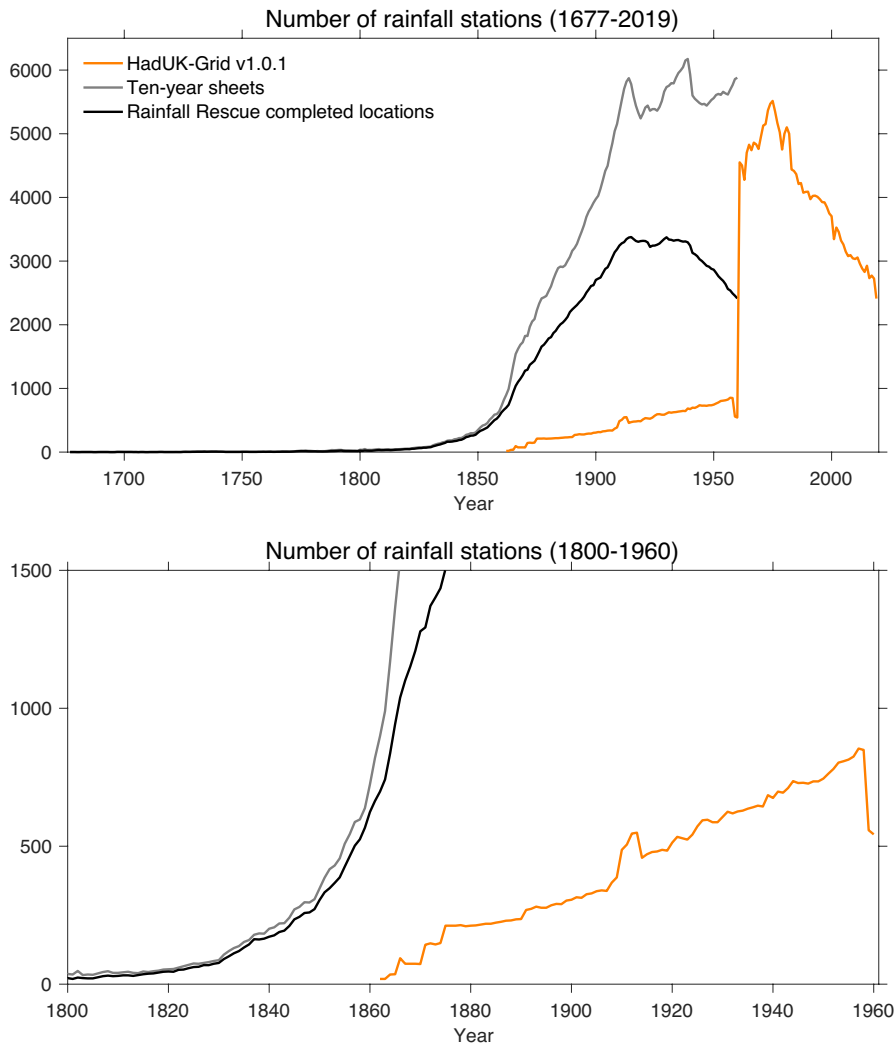


FIGURE 2 The number of locations with rainfall observations available in HadUK-Grid v1.0.1 (orange), the raw Rainfall Rescue observations (grey), and v1.1.0 of the Rainfall Rescue dataset (black) between 1677 and 2019 (top) and 1800–1960 (bottom)

Overall the quality of the records appears to be remarkably good and a testament to the thoroughness of the original observers and the BRO. Errors and duplications were confined to a relatively small number of records and were mostly declared on the records, e.g., readings were copied from other locations to complete a year or decade for averaging purposes. Further quality control tests will be performed when this data is integrated with other already available digital observations.

In this data release (v1.1.0), 6093 locations have been identified and the sheets linked together. In many cases, there was more than one rain gauge present, and sometimes a rain gauge was moved a reasonable distance, so effectively starting a new site. In total, 8549 time series have been produced, containing around 3.34 million observations. In some cases the records cover more than 100 years at a single location, but most are a few decades in length, and many have gaps (see Figure S2). Occasionally the time series cover only a few years, especially pre-1860 and in regions where there were very few rain gauges at that time, e.g. Wales.

Figure 2 summarizes how many stations are available for each year in the raw data and v1.1.0 dataset, compared

to the number in HadUK-Grid. Some of the Rainfall Rescue time series will be duplicates of those already available in HadUK-Grid, but the vast majority will be new. Figure 3 shows the locations available for selected years.

During 2020, the initial data from 95 new locations were provided to the UK Met Office for use in their annual ‘State of the UK climate’ report (Kendon et al., 2020) and the complete dataset will be used in subsequent reports. Currently, the spatially gridded rainfall dataset for the whole of the UK is available for each month since 1862, but the Rainfall Rescue data will: (i) add considerable additional spatial detail, and (ii) allow this dataset to be extended backwards to at least 1836 for the whole UK, and even earlier for some regions.

The focus for v1.1.0 of the Rainfall Rescue dataset has been on ensuring the earlier observations are combined as a priority, which has resulted in a slight decline in data availability from around 1920 to 1950 (Figure 2). There are many locations available in the raw data to fill this temporal gap, and this will be the focus of future combination efforts. Note that v1.1.0 of the dataset already includes data from more rain gauges for each year between around 1900–1960

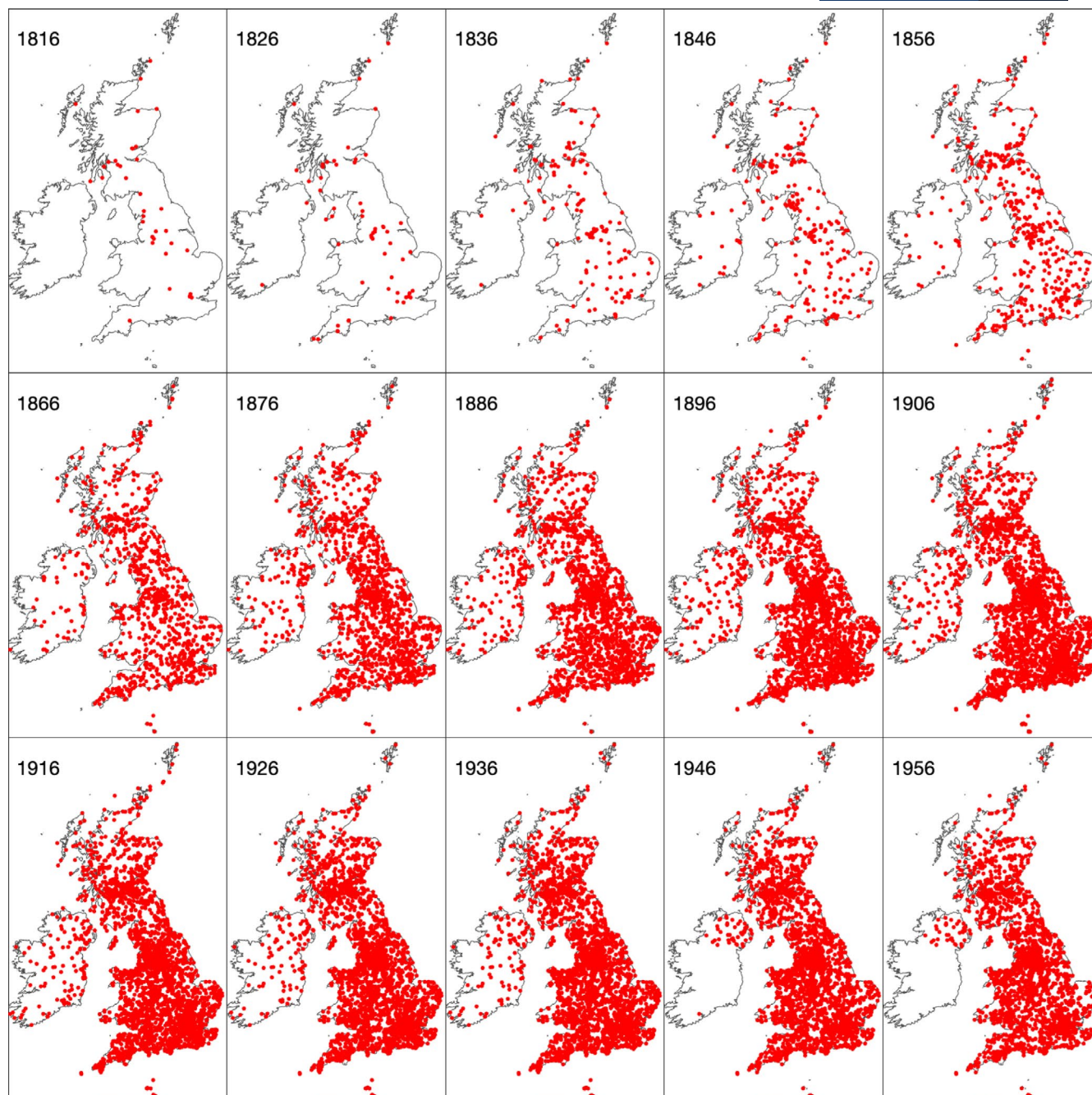


FIGURE 3 The locations of all rain gauges in v1.1.0 of the Rainfall Rescue dataset for specific years between 1816 to 1956. Maps for each year are available as part of the dataset

than are currently available for 2020, due to the decline in the rain gauge observing network since the 1970s.

3.5 | Potential data issues

3.5.1 | Refining locations

Identifying the precise location of the rain gauges often involved research using other sources, as noted above. Many of the early sheets were simply recorded as being

in a town or village and often include no other information. For example, the Dalton (Lancashire) records for 1806–1812 are provided in the dataset without a location as there were two places called Dalton in the county at that time – one in Furness and the other near Skelmersdale – and it has not been possible to provide even a broad location for what is a very early dataset. Added to this, Dr John Dalton was a well-known recorder of weather in Cumberland (now Cumbria) and Lancashire at this time. However, on balance the Dalton location is felt more likely to be a geographical location.

There are several other examples where the town can be identified but the precise location within the town is impossible to determine. Several records have low confidence locations for similar reasons.

More frequently, directions/distances in relation to local churches and railway stations are presented; later versions of the Ten Year sheets (1920s) clearly say that these directions are “from gauge” but, even here, that cannot be assumed as many are, in fact, direction *to* the gauge, and looking at all of the evidence available is essential. This is especially important in more remote areas, where stations/churches can be tens of miles away. Distances typically seem to be ‘as the crow flies’ and are usually given in yards or miles but furlongs are occasionally used e.g. “All Saints, Hereford 1 furlong 6 chains SE”. However, there was at least one occurrence of “church 5 minutes N” leading to discussions about cartographical minutes or time. Looking at family history sites and ascertaining the address of Miss Woodhouse, the observer, it was agreed that she most likely gave the time to walk to the church!

Some records may give the town and a street name, such as “Ross, Broad St”. In this instance, online web searches identified the address of the observers (Purchas & Son, Wine and Spirit Merchants), at numbers 12–13 Broad Street. These searches also located an image of the property in 1953, which was compared with a modern view from StreetMap to confirm the building and identify a precise location (Figure S3). There are many other similar examples where this was possible.

Hand-written notes on the records can also be useful for describing the location, although many refer to obstacles, such as 6 foot high rhubarb or other vegetables, encroaching sweet peas, and a lot of tall gooseberry bushes or raspberry canes! Distances from the house or glass-houses were useful especially for the large country estates where only the house name is given as the location. Glasshouse and kitchen gardens are often marked on older OS maps. Other notes are less enlightening however. For example, Castle Douglas, Slogarie is described as “(gauge) in a glen among the hills”, but subsequent records allowed the location to be refined. On the other hand, the example of extensive notes by Mr Oliver from Langraw, Roxburghshire eventually indicated simply that the house was 30 feet away from the gauge. Occasionally the records would include a diagram showing the precise location of the gauge, such as one on the 1920s record for Finsbury New River Head (Figure S4).

For other locations, there were often several shorter periods of observations in a small area (often less than 3 km in radius). Frequently, a gauge would be handed over from one observer to another in a small town or village, or within family generations, allowing a multi-decade time series to be created which may not otherwise have been

included individually due to their short length. These locations tend to have “-MIX” in the dataset location name, and the different segments have been given slightly different latitude and longitude coordinates as appropriate.

However, many locations are harder to determine precisely, and these uncertain coordinates are usually identified in comments in the dataset combined files.

3.5.2 | Issues with the data

The quality control efforts undertaken in the dataset production will have removed many errors but cannot account for all possible issues. For example, the original observations may have been taken incorrectly, the rain gauge readings were incorrect due to a faulty exposure, or the wrong numbers sent to those collating the Ten Year sheets. There are also some duplicated measurements, often when observations were copied from a nearby rain gauge to avoid a gap in the data. In many cases these have been identified and removed, but there are at least two instances where the same data appears in rain gauges a very long distance apart. For example, some data nominally for ‘Dublin’ is identical to the measurements from two different gauges in Cumbria for the same years. It is believed the Dublin location is incorrect. In another case, data on sheets labelled for Hungerford and sheets for Aylesbury (separated by over 50 km) are identical, but it has proved impossible to identify which location is correct.

The exposures of some rain gauges were noted as not being ideal, which may have caused undercatch or, less frequently, overcatch. Snowfall and objects sheltering the rain gauge are the most common causes. Moves of the rain gauges are not always clear from the Ten Year sheets, and some may not have been noted at all. One rare issue is that when a ‘trace’ of rain was recorded on the sheets, this appears in the dataset as 0.01 inches (entering a ‘trace’ of rainfall—less than 0.005 inches or 0.05 mm—only became official policy in the 1920s). A handful of Ten Year sheets appear to be missing, as there are occasionally 10-year gaps in the combined data corresponding to a complete sheet. These gaps may be possible to fill if the daily rainfall sheets are available in the Archives, but these have not yet been scanned. A few sheets also appear to be filed in the incorrect decade, and these have been identified in the dataset.

Some of the locations have years that are identified in notes on the sheet as being questionable, but unless these are obviously wrong they have usually been retained in the dataset. An exception is some very remote or mountainous locations where two combined data files are provided, one with all the data as transcribed, and one with unusual sequences of years deleted, for example, Covesea Skerries.

Some locations that initially appeared promising are not retained in the final dataset as the data are fragmented or poor. However, the data files and images are still provided in the POOR-LOCATIONS folder of the dataset.

4 | NOTABLE ORGANIZATIONS AND LOCATIONS

Many of the rain gauges were maintained by organizations often linked to the water supply, such as at reservoirs, pumping stations, and water and sewage works. Local councils often used parks as suitable sites for the local authority rain gauge or other climatological instruments. Many other gauges were kept at their homes by volunteer observers, who were often members of the clergy or wealthy landowners, although many were in the care of gardening or estate staff. A large number of the gauge records ended with the observer's death. However, observing was not solely the task of the retired, as many of these men and women had been devoted to recording rainfall for decades. Some of the more widespread organizations who contributed for lengthy periods are briefly highlighted and gratefully acknowledged below. The notable individual observers will be discussed in an additional paper.

4.1 | Northern lighthouse board

The Lighthouses constructed and maintained by the Northern Lighthouse Board (NLB) have provided an important source of rainfall data in some of the most remote parts of Scotland and also the Isle of Man. Symons (1865) identified 48 lighthouses or shore signal stations in Scotland and the Isle of Man. Four of these (Pentland Skerries, Kinnaird Head, Inchkeith and Mull of Kintyre) have data spanning over 147 years from 1813 to 1960 (not always contiguous), but these were all terminated sometime during the 1960s. A further 35 lighthouses were identified providing records starting after 1865. In total 83, NLB lighthouses provided about 72,000 monthly readings in this dataset.

However, there are some issues with the data from lighthouses. Their exposed aspect sometimes led to over-recording of rainfall (e.g., water from waves entering the gauge), but more often under-recording of rainfall (e.g., location on cliff faces resulted in rain being carried over the gauge by updrafts). Other occasional issues noted on the sheets included puffins burrowing, causing the gauge to topple over! The data from lighthouses requires further checking and comparisons with other nearby gauges where possible before use.

We note that many of the original English lighthouse records (pre-1939) were unfortunately lost in the bombing of Trinity House in London during World War 2. However, the Rainfall Rescue dataset does include records from some English lighthouses, such as Dungeness and Hartlepool Heugh, copies of whose records were also in the Met Office archives.

4.2 | Reservoirs and water companies

There was a vast increase in demand for water from the mid-19th century onwards, with a further increase in the early 20th century, due to a combination of factors such as industrialization, urbanization and transport. Many water companies were formed during these periods, often from amalgamation or buy out of smaller private companies, in order to satisfy this demand. Each will have had their own Civil Engineers who were tasked with understanding local hydrology as well as geology. Many set up large numbers of rain gauges in the catchment being investigated, usually in respect of construction of a reservoir. Some of these gauges continued over decades but many were in operation for only a few years. A specific example of the Catcleugh Reservoir in Northumberland is described in the Supplementary Information.

4.3 | The Manchester, Sheffield and Lincolnshire Railway

The Manchester, Sheffield and Lincolnshire Railway (MS&LR) operated rain gauges for about 70 years, from around 1850 to 1922. There were gauges in about 50 different locations, spread across MS&LR's canal and railway networks in Derbyshire, Cheshire, Lancashire, the West Riding of Yorkshire and Lincolnshire. MS&LR was renamed the Great Central Railway in 1897 and merged with several other railways to form the London and North Eastern Railway (LNER) at the start of 1923, at which point the rain gauge records abruptly stopped. A few rain gauges in Derbyshire and Cheshire have sporadic observations in the 1840s. These were referenced in an unsuccessful proposal by MS&LR to sell clean rainwater falling on its reservoir catchment moorlands to Manchester Corporation in 1848 (Homersham, 1848).

MS&LR rain gauges made up 35 of the ~500 stations listed in the first *British Rainfall* volume covering 1860–1861. 45 gauges were recorded in the 1922 volume of *British Rainfall*, the final listing before the gauges were decommissioned, with 30 of gauges being in use for the full 1860–1922 period. In 1923, *British Rainfall* gave the

MS&LR gauges a farewell paragraph, with a warning that the gauges might have under-recorded the amount of rainfall due to their shallow funnel design (many of them could also have been over 70 years old by the end).

MS&LR owned and operated a number of canals and waterways in addition to its railway networks. The MS&LR rainfall sheets sometimes provide only vague location information on them, just a town name (e.g., Barnsley, Macclesfield, Retford) and an elevation, not making it clear whether this was a canal-based location or a railway-based location, and with no ‘nearest church’ type directions provided either. However, eleven of the MS&LR rain gauges are marked on old Ordnance Survey maps, making their exact location clear. For example, New Holland Station, Lincolnshire (which connected with MS&LR’s Humber ferry to Hull) is shown in Figure 4. More details on the MS&LR rain gauges are given in the Supplementary Information.

4.4 | The Stye

The Stye, altitude around 1,100 ft (330 m), is located on Seathwaite Fell at the southern end of Borrowdale in the Lake District, about 2 km south of Seathwaite village. The Stye’s rainfall received particular attention from the British Rainfall Organization, as the rain gauges there

often recorded larger annual totals than any other location in Britain, sometimes over 200 inches [5,000 mm] in a year or 40 inches [1,000 mm] in a month (Figure 5 and Supplementary Information). There are 28 ten-year rainfall sheets for The Stye between 1850 and 1960, for nine different rain gauges, up to four of them running in parallel at times. Two further sheets seem to be missing, both from the 1910s.

5 | INTERESTING EXAMPLE COMMENTS ON THE SHEETS

The Ten Year sheets often include small comments describing events or issues during that particular decade. A number of interesting or humorous comments are listed below, but there are many others.

Woolwich Eltham High Street, 1944: “Gauge destroyed by enemy action”.

West Ayton (North Riding) readings stopped in September 1949: “too old to bother now”.

Harter Fell, Middleton-in-Teesdale, November 1876: “no readings as gauge stolen”.

The Hall, Sunderland, 1866: “Rev Iliff (the observer) thinks his observations hopelessly wrong”, followed by a comment in 1869: “Rev Iliff had his right arm broken in June so was prevented from taking his observations

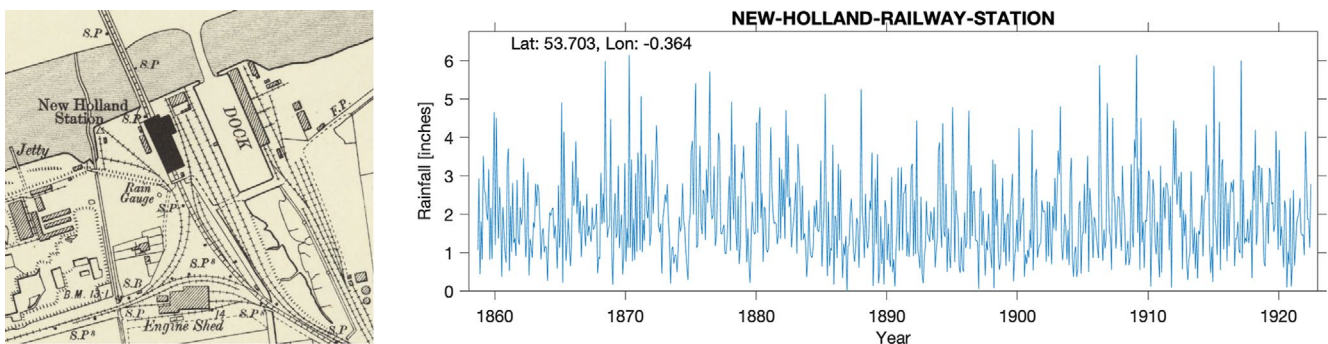


FIGURE 4 (left) New Holland Railway Station gauge identified on a 1908 Ordnance Survey 25-inch map (image from National Libraries of Scotland, maps.nls.uk), and (right) the rainfall time series from 1859–1922

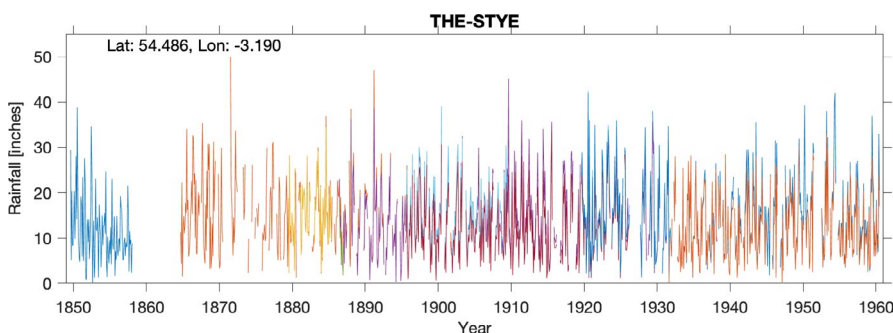


FIGURE 5 The monthly rainfall time series for the rain gauges at The Stye, with different colours representing the different gauges

regularly and a few weeks afterwards a road was made through his garden and his instrument meddled with”.

Saffron Walden Audley End (Essex) by J. Bryan (the observer), 1876: “I am afraid, there is not much dependence on this gauge... I find the funnel often unlouvered (by) curious persons taking it off to see the inside”.

Sevenoaks Chevening Gardens, September 1892: “Gauge emptied by child”.

Stourmouth Rectory Kent, 1863: “Gauge found choked with a bird’s nest”.

Walden Head near Aysgarth, 1875: “Gauge destroyed by tourists”.

Dartmoor Chagford White Ridge, 1928: “Gauge disturbed by ponies”.

Perth (The Academy), 1936: “G. somewhat out of shape having been struck by lawn-mower”.

Banstead Mental Hospital, Dec 1951: “Gauge hidden by inmates.” (The record did not resume until July 1954.)

Leeds, Allerton Hall “site unsatisfactory. Obs refuses to consider new site. Blacklisted”.

Shotley Bridge, Durham “1894 Mr Coulson died in September & the record for the remainder of the year is unsatisfactory”.

Ingbirchworth, Brown’s Gauge, 1870s “From the record kept by the observer at the Reservoir. The observations are carelessly entered & the arithmetic is very faulty. In a few cases there is doubt whether small quantities were left in the gauge at the end of the month. Nevertheless the record is substantially correct”.

6 | LONG-TERM RAINFALL VARIATIONS AND EXTREME EVENTS

This dataset provides snapshots of rainfall variations at individual locations. The longest running station is at the Oxford Radcliffe Observatory, where data exists for every month from 1767 to today (Burt & Burt, 2020), but the majority of the stations included in v1.1.0 of the dataset are for 20–60 years, though not always continuously (Figure S2).

Many applications of this rainfall data will require the integration of data from all locations together to produce longer time series and/or spatial averages by piecing together the individual series. This is not a trivial task. For example, there are recent reconstructions for Chatsworth House (Harvey-Fishenden et al., 2019) and Carlisle (Todd et al., 2015). Hollis et al. (2019) describe the detailed statistical approach to produce HadUK-Grid, which is an estimate of rainfall on a 1 km grid across the UK every month since 1862.

Here, we apply a simplified version of the Hollis et al. (2019) approach¹ to produce estimates of rainfall across the UK for each month back to 1836. Figure 6 shows the UK average rainfall for each season from 1836–2019, with the Rainfall Rescue data shown for 1836–1960 and the current HadUK-Grid series for 1862–2019. The two series agree well for the overlapping 1862–1960 period, with the Rainfall Rescue series being very slightly drier overall. There are notably dry winters in the 1840s and 1850s, which will be interesting to examine in further work.

For the England and Wales average (Figure 7), the comparison can be extended further back using the EWP series (Alexander & Jones, 2001). Again, the Rainfall Rescue series is slightly drier than EWP, especially before 1860 in spring and summer. Murphy et al. (2020) discuss the potential limitations of EWP due to under-representation of snowfall in winter and also suggest that the summer values are slightly too large in the early EWP record. The Rainfall Rescue dataset will allow a more detailed investigation of these issues.

As further examples, we show maps of rainfall in the four wettest and driest months during the 1836–1960 period for the UK as a whole, just using the Rainfall Rescue v1.1.0 dataset (Figures 8 and 9). These illustrate the variations in rainfall that can occur around the UK in both time and space, and the spatial detail that it is possible to produce. The vastly increased number of stations now available will increase confidence and reduce uncertainty in these types of map and time series. October 1903 remains the wettest UK month on record (~219 mm for the UK average) and February 1932 is the driest on record (~10 mm for the UK average) in this dataset, although there are often large differences between rainfall anomalies across the country and different regions will often have other months for their wettest and driest on record.

7 | SUMMARY AND LESSONS LEARNT

The Rainfall Rescue project has digitized the ‘Ten Year rainfall sheets’, which are 66,000 pages containing 5.28 million hand-written monthly rainfall observations taken all over the UK and Ireland between 1677 and 1960. The digitization was achieved during March and April 2020 by

¹We estimate rainfall anomalies for each station for each month (calculated using the Hollis et al. (2019) monthly varying climatology), and interpolate between observations after applying a Box-Cox transformation. The interpolated anomalies are recombined with the gridded climatology to produce an estimate of rainfall across the complete 1km grid which can be spatially averaged.

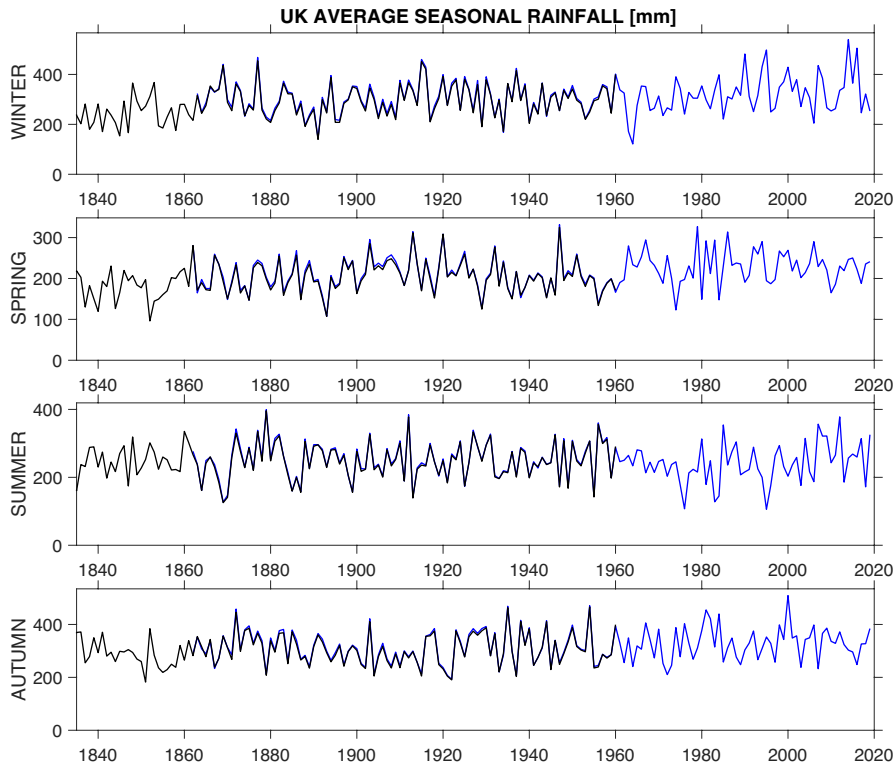


FIGURE 6 UK average seasonal rainfall since 1836 using Rainfall Rescue data (black) and the HadUK-Grid series (blue)

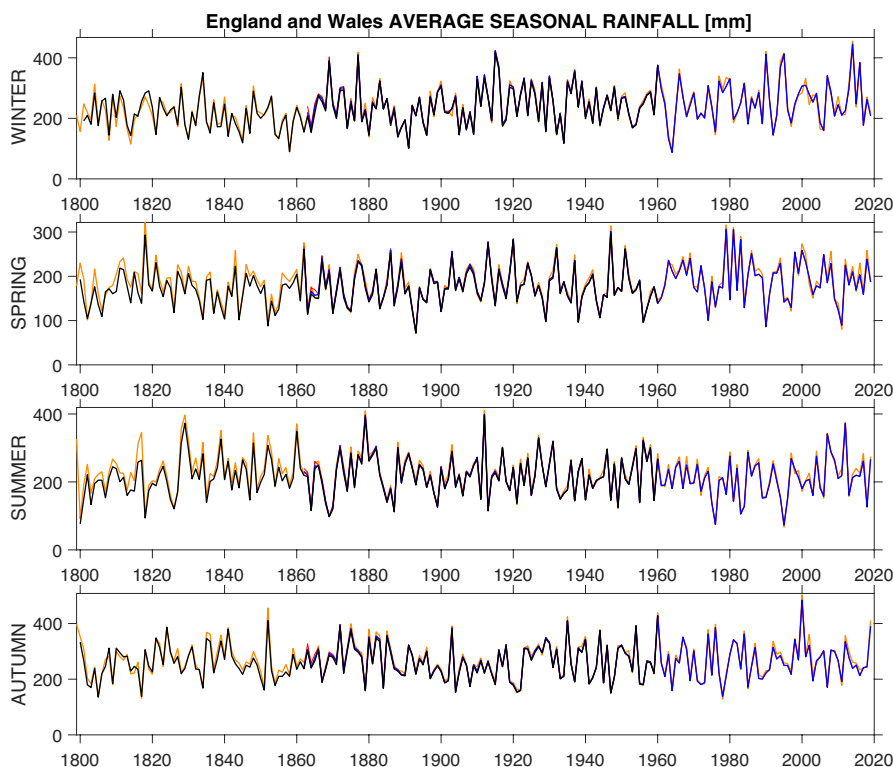


FIGURE 7 England and Wales average seasonal rainfall since 1800 using Rainfall Rescue data (black), the HadUK-Grid series (blue) and the EWP series (orange)

more than 16,000 volunteers using the RainfallRescue.org website, built using the Zooniverse web platform. This is believed to be the largest climate-related citizen science data rescue project ever completed.

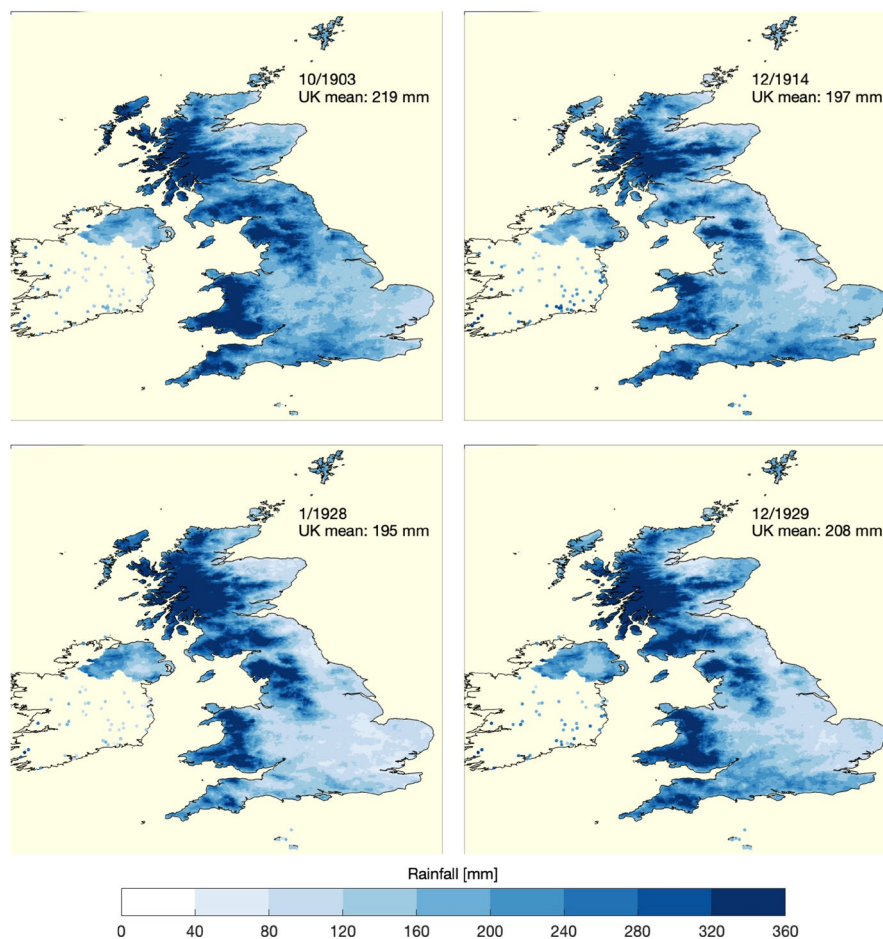
Subsequent detailed work by a small team of dedicated volunteers (who are co-authors on this paper) has enabled the release of 3.34 million observations, which have been

collated, combined and initially quality controlled into time series for 6093 locations (Hawkins, 2021). These observations will enable the reconstruction of monthly rainfall variations across the UK back to 1836, and even earlier for some regions.

A number of lessons have been learnt about running such citizen science projects, which may be useful

FIGURE 8 Example maps of gridded rainfall estimate for the four wettest UK months in the 1836–1960 period, just using the Rainfall Rescue v1.1.0 dataset. The values for stations outside the UK are shown as dots

UK rainfall patterns for example months



for similar projects. The media interest was critical for recruiting volunteers, but the timing of the project was slightly fortuitous with the scanned copies of the sheets becoming available just before the COVID pandemic. The number of volunteers was higher than expected, and this was at least partly because the pandemic-related lockdowns gave some people more spare time and a desire to do useful projects, but also meant others were looking for a welcome distraction.

Also important was the continued engagement and conversation with the volunteers and providing provisional results in a timely manner. The project ‘Talk Forum’ allowed the project team and volunteers to interact easily. For example, the monthly rainfall observations for a particular year were initially requested to be transcribed in two groups, but many volunteers commented that this was awkward, and the project website was changed to ask for all twelve monthly values and annual total in the same task. We also chose to show the whole image of the Ten Year sheet in each task, rather than undertake image cropping for the particular part of the page being transcribed. This had the advantage of being easier for the project team and allowed the volunteers to see potentially

valuable context within the sheet, but may have slowed down the typing process. Given the number of volunteers who joined Rainfall Rescue, the advantages of this choice clearly outweighed the disadvantages, but this may not always be the case with similar transcription projects.

As noted above, the transcription of the observer’s name (and potentially other parts of the metadata) would have been a huge benefit in joining locations together. The project would also have benefitted from a clearer set of instructions for unusual situations on the sheets and a single place where volunteers could see the lessons being learnt during the transcription process. Such transcription projects would also benefit from a more flexible set of tools within the Zooniverse infrastructure, and this is planned development.

Most importantly, the engagement of a group of keen volunteers in the location identification, concatenation of decadal sheets and in the quality control process has been essential and immensely valuable. Placing our trust in the volunteers has been a highly rewarding step as they have enabled the data release to be completed far faster and in larger volume than could have been achieved by the project team alone.

UK rainfall patterns for example months

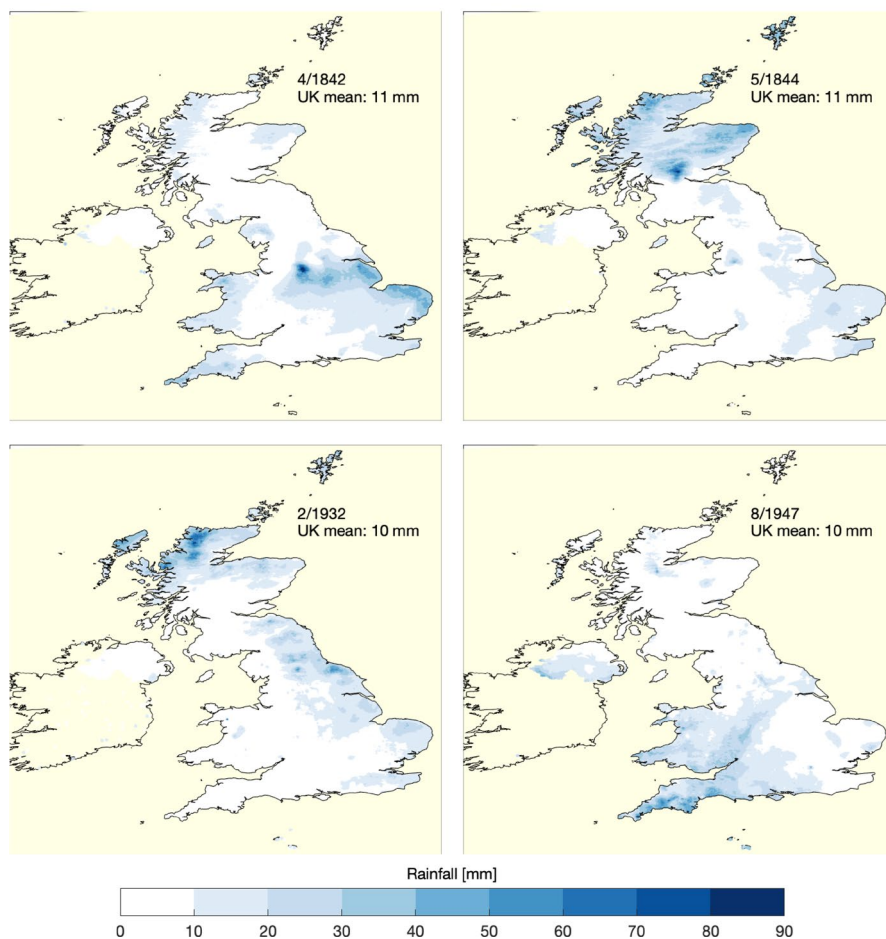


FIGURE 9 Example maps of gridded rainfall estimate for the four driest UK months in the 1836–1960 period, just using the Rainfall Rescue v1.1.0 dataset. Note the different colour scale from Figure 7

ACKNOWLEDGEMENTS

We thank the 16,000 volunteers who contributed so much of their spare time for the transcription of the observations, and the tens of volunteers who continued to provide invaluable input during the construction of the final dataset. We also gratefully acknowledge the efforts of the thousands of volunteer and paid observers who diligently measured the rainfall, often continuously for several decades, and for keeping such detailed records from which we can all benefit today. The British Rainfall Organization deserves much credit for coordinating the rainfall measurement activities since 1859. EH was funded by the National Centre for Atmospheric Science and the NERC GloSAT project.

AUTHOR CONTRIBUTION

Ed Hawkins: Conceptualization (lead); Data curation (lead); Formal analysis (lead); Funding acquisition (lead); Investigation (lead); Methodology (lead); Project administration (lead). **Stephen Burt:** Project

administration (equal); Writing – review & editing (equal). **Mark Paul McCarthy:** Conceptualization (equal); Resources (equal); Writing – review & editing (equal). **Conor Murphy:** Investigation (equal); Writing – review & editing (equal). **Catherine Ross:** Resources (equal); Writing – review & editing (equal). **Mike Baldock:** Investigation (equal); Validation (equal); Writing – review & editing (equal). **John Brazier:** Investigation (equal); Validation (equal); Writing – review & editing (equal). **Gill Hersee:** Investigation (equal); Validation (equal); Writing – review & editing (equal). **Jacqui Huntley:** Investigation (equal); Validation (equal); Writing – review & editing (equal). **Richard Meats:** Investigation (equal); Validation (equal); Writing – review & editing (equal). **John O’Grady:** Investigation (equal); Validation (equal); Writing – review & editing (equal). **Ian Scrimgeour:** Investigation (equal); Validation (equal); Writing – review & editing (equal). **Tim Silk:** Investigation (equal); Validation (equal); Writing – review & editing (equal).

OPEN PRACTICES



This article has earned an Open Data badge for making publicly available the digitally shareable data necessary to reproduce the reported results. The data is available at <https://doi.org/10.5281/zenodo.5770389>. Learn more about the Open Practices badges from the Center for Open Science: <https://osf.io/tvyxz/wiki>.

ORCID

Ed Hawkins  <https://orcid.org/0000-0001-9477-3677>
 Stephen Burt  <https://orcid.org/0000-0002-5125-6546>
 Mark McCarthy  <https://orcid.org/0000-0003-2692-1417>
 Conor Murphy  <https://orcid.org/0000-0003-4891-2650>

REFERENCES

- Alexander, L. & Jones, P. (2001) Updated precipitation series for the UK and discussion of recent extremes. *Atmospheric Science Letters*, 1, 142. <https://doi.org/10.1006/asle.2001.0025>
- Amos, J. (2020). *Help needed to rescue UK's old rainfall records*. <https://www.bbc.co.uk/news/science-environment-52040822> Accessed 25 May 2021.
- Burt, S. (2010) British Rainfall 1860–1993. *Weather*, 65, 121. <https://doi.org/10.1002/wea.603>
- Burt, S. & Burt, T. (2020) *Oxford Weather and Climate since 1767*. Oxford University Press.
- Craddock, J. (1976) Annual rainfall in England since 1725. *Quarterly Journal of the Royal Meteorological Society*, 102, 823. <https://doi.org/10.1002/qj.49710243412>
- Craig, P. & Hawkins, E. (2020) Digitizing observations from the Met Office Daily Weather Reports for 1900–1910 using citizen scientist volunteers. *Geoscience Data Journal*, 7, 116. <https://doi.org/10.1002/gdj3.93>
- Folland, C. & Wales-Smith, B. (1977) Richard Towneley and 300 years of regular rainfall measurement. *Weather*, 32, 438. <https://doi.org/10.1002/j.1477-8696.1977.tb04501.x>
- Glasspoole, J. (1941) Variations in annual, seasonal and monthly rainfall over the British Isles, 1870–1939. *QJRMS*, 67, 5. <https://doi.org/10.1002/qj.49706728803>
- Glasspoole, J. (1952) The British Rainfall Organisation. *Weather*, 7, 85. <https://doi.org/10.1002/j.1477-8696.1952.tb01449.x>
- Harvey, F. (2020). *Call for isolated Britons to help digitise historical rainfall data*. <https://www.theguardian.com/science/2020/mar/26/call-isolated-britons-help-digitise-historical-rainfall-data> Accessed 1 June 2021
- Harvey-Fishenden, A., Macdonald, N. & Bowen, J. (2019) The early dry weather fears of Britain's water-transport network. *Regional Environmental Change*, 19, 2325–2337. <https://doi.org/10.1007/s10113-019-01524-5>
- Hawkins, E. (2021). *Rainfall Rescue dataset v1.1.0*. <https://doi.org/10.5281/zenodo.5770389>
- Hawkins, E.D., Burt, S., Brohan, P., Lockwood, M., Richardson, H., Roy, M. et al. (2019) Hourly weather observations from the Scottish Highlands (1883–1904) rescued by volunteer citizen scientists. *Geoscience Data Journal*, 6, 160. <https://doi.org/10.1002/gdj3.79>
- Hollis, D., McCarthy, M., Kendon, M., Legg, T. & Simpson, I. (2019) HadUK-Grid—A new UK dataset of gridded climate observations. *Geoscience Data Journal*, 6, 151. <https://doi.org/10.1002/gdj3.78>
- Homersham, S. (1848) *Report to the directors of the Manchester, Sheffield, and Lincolnshire railway company comparing the quantity, quality and price of the water that can be supplied to the inhabitants of Manchester and Salford*. Available at <https://books.google.co.uk/books?id=2RUHAAAQAAJ> Accessed 25 Jun 2021.
- Kendon, M., McCarthy, M., Jevrejeva, S., Matthews, A., Sparks, T. & Garforth, J. (2020) State of the UK climate 2019. *International Journal of Climatology*, 40, 1–69. <https://doi.org/10.1002/joc.6726>
- Met Office Digital Library and Archive (2020) *Ten Year rainfall sheets*. Available at https://digital.nmla.metoffice.gov.uk/SO_d383374a-91c3-4a7b-ba96-41b81cfb9d67/ Accessed 24 November 2020.
- Murphy, C., Burt, T.P., Broderick, C., Duffy, C., Macdonald, N., Matthews, T. et al. (2018) A 305-year continuous monthly rainfall series for the island of Ireland (1711–2016). *Climate of the Past*, 14, 413. <https://doi.org/10.5194/cp-14-413-2018>
- Murphy, C., Wilby, R.L., Matthews, T., Horvath, C., Crampsie, A., Ludlow, F. et al. (2020) The forgotten drought of 1765–1768: Reconstructing and re-evaluating historical droughts in the British and Irish Isles. *International Journal of Climatology*, 40, 5329. <https://doi.org/10.1002/joc.6521>
- Nicholas, F. & Glasspoole, J. (1931) General monthly rainfall over England and Wales 1727–1931. *British Rainfall*, 1931, 299.
- Pedgeley, D. (2010) The British Rainfall Organization, 1859–1919. *Weather*, 65, 115. <https://doi.org/10.1002/wea.579>
- Ryan, C., Duffy, C., Broderick, C., Thorne, P.W., Curley, M., Walsh, S. et al. (2018) Integrating data rescue into the classroom. *BAMS*, 99, 1757. <https://doi.org/10.1175/BAMS-D-17-0147.1>
- Ryan, C., Murphy, C., McGovern, R., Curley, M. & Walsh, S. (2021) Ireland's pre-1940 daily rainfall records. *Geoscience Data Journal*, 8, 11. <https://doi.org/10.1002/gdj3.103>
- Salter, M. (1921) *The Rainfall of the British Isles*. London University Press.
- Sieber, R. & Solonsky, V. (2019) Developing a flexible platform for crowdsourcing historical weather records. *Historical Methods*, 52, 164. <https://doi.org/10.1080/01615440.2018.1558138>
- Simpson, I.R. & McCarthy, M.P. (2019) Structural and sampling uncertainty in observed UK daily precipitation extremes derived from an intercomparison of gridded data sets. *International Journal of Climatology*, 39, 128. <https://doi.org/10.1002/joc.5789>
- Symons, G. (1865). *On the Rainfall of the British Isles*. Report to the British Association for the Advancement of Science, page 192; online copy available at <https://www.biodiversitylibrary.org/page/29381656#page/264/mode/1up> Accessed 25 June 2021.
- Todd, B., Macdonald, N. & Chiverrell, R.C. (2015) Revision and extension of the composite Carlisle rainfall record, NW England: 1757–2012. *International Journal of Climatology*, 35, 3593–3607. <https://doi.org/10.1002/joc.4233>
- Townley, R. (1694) VI. A letter from Richard Townley, of Townley in Lancashire, Esq; containing observations on the quantity of rain falling monthly, for several Years successively: communicated to the Royal Society. *Philosophical Transactions of the Royal Society of London*, 18, <https://doi.org/10.1098/rstl.1694.0014>

- Walker, M. (2010) The man behind the British Rainfall Organization. *Weather*, 65, 117. <https://doi.org/10.1002/wea.595>
- Wigley, T.M.L., Lough, J.M. & 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. <https://doi.org/10.1002/joc.3370040102>

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

How to cite this article: Hawkins, E., Burt, S., McCarthy, M., Murphy, C., Ross, C., Baldock, M., et al (2022) Millions of historical monthly rainfall observations taken in the UK and Ireland rescued by citizen scientists. *Geoscience Data Journal*, 00, 1–16. Available from: <https://doi.org/10.1002/gdj3.157>