Evaluating meteorological reanalysis for identifying fair weather conditions in historical atmospheric electricity data

Article

Published Version

Mkrtchyan, H. ORCID: https://orcid.org/0000-0002-2921-1384, Nicoll, K. A. ORCID: https://orcid.org/0000-0001-5580-6325 and Harrison, R. G. ORCID: https://orcid.org/0000-0003-0693-347X (2025) Evaluating meteorological reanalysis for identifying fair weather conditions in historical atmospheric electricity data. Quarterly Journal of the Royal Meteorological Society. e5066. ISSN 1477-870X doi: 10.1002/qj.5066 Available at https://centaur.reading.ac.uk/123638/

To link to this article DOI: http://dx.doi.org/10.1002/qj.5066

Publisher: Royal Meteorological Society

Received: 26 November 2024 | Revised: 17 June 2025 | Accepted: 8 July 2025

DOI: 10.1002/qj.5066

Quarterly Journal of the Royal Meteorological Society RMetS

RESEARCH ARTICLE

Evaluating meteorological reanalysis for identifying fair-weather conditions in historical atmospheric electricity data

H. Mkrtchyan | **K. A. Nicoll** | **R. G. Harrison**

Department of Meteorology, University of Reading, Reading, United Kingdom

Correspondence
K.A. Nicoll, Department of Meteorology, University of Reading, Reading, United Kingdom.
Email: k.a.nicoll@reading.ac.uk

Funding information
UK Research and Innovation, Grant/Award Number: EP/X024547/1

Abstract

Surface atmospheric electricity measurements, particularly of the vertical electric field (or potential gradient, PG) began in the 1770s, becoming more widespread in the United Kingdom during the nineteenth century. In the twentieth century, PG measurements were systematically obtained by the Met Office, at their observatory sites of Kew, Eskdalemuir and Lerwick. These records' importance is now increasingly recognised, due to the inherent global atmospheric electric circuit (GEC) information, which is an embedded part of the climate system. Earlier data have the limitation that, until about 1960, preferred daily PG data values were selected using the geomagnetically-informed approach of identifying quiet and disturbed days. It is now known that classification by local weather conditions, identifying the data recorded during 'Fair Weather' (FW) or dry, 'Non-Hydrometeors' (NH) circumstances, is superior for obtaining GEC signals. However, the necessary weather information is only available at a subset of PG measurement sites globally. For other sites, meteorological reanalysis – and many such data products are available spanning different times and scales – offers a new approach for retrospective weather-based classification of PG data. This work investigates applying ERA5 reanalysis to selecting PG data obtained during FW and NH conditions by using the Lerwick site as a testbed, comparing with direct weather observations (including wind speed, precipitation, cloud base height, and pressure) which were originally used for PG data selection. ERA5-based PG data selection is shown to yield quality improvements, especially for data downgraded by the geomagnetic approach, for example our method reclassifies 20% of disturbed weather data as fair weather. This offers a general route to improving long-term atmospheric electricity data obtained at non-meteorological sites globally.

KEYWORDS

data recovery, global electric circuit, historical data, potential gradient, reanalysis

1 | INTRODUCTION

The fair-weather potential gradient (PG), in the absence of local effects, is related to the global distribution of thunderstorms and electrified rain clouds. Good quality PG measurements classified in this way are relatively uncommon, but have occasionally been recorded at certain locations worldwide, including at Lerwick observatory in Shetland, United Kingdom, where measurements were made from 1925 to 1984 (Harrison & Riddick, 2022). Similar long time series measurements of PG were also made at the other UK Met Office sites at Eskdalemuir and Kew, London. These long records potentially provide a wealth of information about variations in the global electric circuit (GEC), which links current flow in the disturbed regions of the planet (from thunderstorms and shower clouds) to the fair-weather regions of the atmosphere (Wilson, 1921). The GEC is essentially an embedded aspect of the climate system, and responds to internal variability, such as El Niño Southern Oscillation (ENSO) (Harrison *et al.*, 2011, 2022; Slyunyaev *et al.*, 2021). By analysing historical PG data, valuable insights into long-term patterns, trends, and changes in the atmospheric system are expected, to improve understanding of processes related to climate and perhaps refine climate models. Additionally, comparing older data with more recent observations allows assessment of how the global circuit may have evolved over time, to investigate potential links to climate change.

The global representativity of PG measurements is significantly influenced by local weather conditions, making it essential to use PG from a site where these conditions are well-monitored and documented. Lerwick provides comprehensive associated data that allows for the effective removal of local weather-related effects, thereby enhancing the likelihood of retrieving GEC signals. This makes Lerwick a highly suitable testbed for refining fair-weather criteria, with the object of reclassifying historical PG measurements.

There are direct indications that the PG data from Lerwick can have global representativity. Beyond the established link with ENSO, the Lerwick PG data show the diurnal variation typical of the GEC, that is, similar in shape to the variation in global thunderstorm area of a minimum at 0300 UTC and maximum at 1900 UTC. This is conventionally known as the *Carnegie curve* (Harrison, 2013; Whipple & Scrase, 1936). In the earlier period of PG measurements, the data at Lerwick were selected on the appearance of the daily chart traces, following the geomagnetic practice of 'quiet' and 'disturbed' days. Since the 1960s, atmospheric electricity data at the UK sites and more generally have been classified, at least, into hours without rainfall (termed 'Non-Hydrometeors') and, ideally 'Fair Weather' (OYB, 1965). The criteria used to determine 'Fair Weather' seek to minimise local factors which may otherwise adversely influence the PG measurements (Harrison & Nicoll, 2018).

Only a subset of atmospheric electricity (AE) measurement sites globally have had co-located meteorological measurements, and hence classification and selection of those sites' PG data using the local weather conditions has hitherto not been possible. In this study, we aim to explore how ERA5 meteorological reanalysis data can be used to apply weather-based classification. This potentially offers retrospective classification of early PG data at non-meteorological sites, and therefore the possibility of recovering a large amount of historical AE data in which there may be GEC signals.

The approach is evaluated using data from the Lerwick site, where direct meteorological data is also available which can be compared with the weather information from ERA5. To explore how suitable conditions can be recognised in earlier records (pre-1960s), relevant weather parameters have been extracted from ERA5 and compared with surface meteorological observations from Lerwick. With digitized hourly PG data for Decembers (other months are yet to be digitised) now available from 1927 to 1954 (Harrison *et al.*, 2023; Mkrtchyan *et al.*, 2024) and ERA5 data accessible from 1940 onward, this research focuses on reclassifying the historical PG measurements at Lerwick in the overlap from 1940 to 1954.

This paper is structured in the following way. Section 2 describes the PG data available from the Lerwick site as well as the ERA5 reanalysis data. Section 3 provides a comparison of the ERA5 data with the directly observed local weather data (so-called Met Office Integrated Data Archive System [MIDAS]) to identify fair-weather criteria. In Section 4, we apply the fair-weather criteria derived from ERA5 data to the historical PG measurements, enabling a focused analysis of PG behaviour during fair-weather periods. Section 5 discusses the effects and limitations of the data selection on PG, and conclusions on the usefulness and applicability of the method are provided in Section 6.

2 | DATA AND METHODOLOGY

2.1 | Observatory and historical data

Atmospheric electricity measurements were undertaken almost continuously at Lerwick observatory, Shetland, United Kingdom between 1925 and 1984, providing a long series of hourly PG measurements (Harrison *et al.*, 2025; Harrison & Riddick, 2022). The earlier data were classified

TABLE 1 Potential gradient (PG) data from Lerwick observatory.

Period	Type	Classification system	Digitisation status	Citation
1927–1956	Handwritten	Daily electric character	Hourly values in progress (Zooniverse AtmosElec Project: https://www.zooniverse.org/projects/hripsi-19/atmoselec-atmospheric-electricity-for-climate)	
			December hourly values digitised (1940–1950)	Harrison (2022)
			Monthly mean values digitised	Harrison *et al.* (2023)
1957–1963	Printed	Daily electric character	Hourly values not digitised	
			Monthly mean values digitised	Harrison *et al.* (2023)
1964–1984	Printed	Meteorological	Hourly averaged PG for each month (1964–1980)	Mkrtchyan *et al.* (2024)
			Monthly mean values digitised	Harrison *et al.* (2023)

using a daily character method, based on approaches derived from geomagnetic measurements. Only the later (post-1965) values were classified using the weather criteria method. A summary of when the different classification systems were applied is given in Table 1, and the details of the daily character method and weather criteria methods are described in Section 2.4.

The PG data were originally recorded by hand on a variety of standard paper reporting forms, from which printed summary sheets were generated. These have all been scanned for preservation. These summary sheets indicate which days were classified as a disturbed, no hydrometeor or fair-weather day. Despite the challenges posed by variable formats and a range of styles of handwriting, our digitisation efforts (which are currently ongoing through manual classification, and the AtmosEleC citizen science Zooniverse project [https://www.zooniverse.org/projects/hripsi-19/atmoselec-atmospheric-electricity-for-climate]) have so far successfully converted the hourly December PG data from 1927 to 1954 (Mkrtchyan *et al.*, 2024), and the data summary sheets from 1957 to 1983 (Harrison *et al.*, 2023) into a usable digital format. From 1940 to 1964 these sheets did not include indications of fair-weather days and hours, whereas during the years 1964 to 1983, much more detailed information is included. It is therefore for these earlier years with poorer metadata that we investigate applying the additional insights provided by meteorological reanalysis.

2.2 | ERA5 reanalysis dataset

Meteorological reanalysis data combine such observations as are available with numerical modelling techniques to provide a consistent and detailed record of past weather and climate conditions. ERA5 is the fifth generation of atmospheric reanalysis produced by European Centre for Medium-Range Weather Forecasts (ECMWF) under the Copernicus Climate Change Service (Dutra *et al.*, 2020), with a spatial resolution of 31 × 31 km. The ERA5 data were obtained from Hersbach *et al.* (2023) and the quantities used for this study are:

- Surface pressure: a measure of the weight per unit area of the air column vertically above a point on the Earth's surface. In ERA5 the units of this parameter are hectopascals (hPa).
- Total precipitation (TP): this represents the combined liquid and frozen water, including rain and snow, falling to the Earth's surface. It consists of large-scale and convective precipitation, excluding fog, dew, or atmospheric evaporation. We have converted hourly TP from ERA5 into a 12-hour accumulation (for 0900 and 2100 UTC) to make it consistent with the rainfall observations from MIDAS. Units have also been converted from m to mm.
- Wind speed (WS): this is derived (in $m{\cdot}s^{-1}$) from u and v wind components, representing eastward and northward air movement at 10 m above the Earth's surface. Here we use the mean WS, which is calculated from the vector mean of u and v using the following equation:

$$WS = \sqrt{u^2 + v^2} \tag{1}$$

- Cloud base height (CBH): this is calculated by searching from the second-lowest model level upwards, to the level where cloud fraction becomes greater than 1% and condensate content greater than 1×10^{-6} $kg{\cdot}kg^{-1}$. Fog (i.e., cloud in the lowest model layer) is not considered when deriving CBH.
- Total cloud cover (TCC) is the proportion of a grid box covered by cloud. TCC is a single-level field calculated from the cloud occurring at different model levels through the atmosphere. Assumptions are made about the degree of overlap/randomness between clouds at different heights. Cloud fractions vary from 0 to 1. For

this study we have converted the units into oktas (i.e., eighths of sky coverage).

The ERA5 reanalysis dataset is widely used by operational weather centres, and often as a state-of-the-art-climatology with which to evaluate forecast models and assess model developments. The hourly resolution of ERA5 output also allows the study of short-term weather events, such as investigating the detailed evolution of weather systems such as hurricanes (Hersbach 2019) or break-up of the stratospheric polar vortex (Hersbach *et al.*, 2020). Overall, in terms of precipitation, previous studies have found ERA5 to be a valuable resource for rainfall studies, but it has potential biases in estimating diverse types of rainfall. For example, it has been observed to have a wet bias for light cumulus rainfall and a dry bias for heavier (convective) precipitation (Chen *et al.*, 2023). These biases are attributed to the convective precipitation scheme and the large-scale rainfall scheme used in ERA5 (Chen *et al.*, 2023). The accuracy of ERA5 WS output has also been investigated, for example Molina *et al.* (2021) compared hourly 10-m windspeed from ERA5 with WS observations from the HadISD network of 245 stations across Europe and found that most of the stations exhibited a high correlation (typically 0.8–0.9) between the observations and ERA5 WS on an hourly basis, with the correlation further improving for longer time frequencies (e.g., 6–24 hours). Similarly, Santos *et al.* (2019) compared ERA5 WSs with observations at a wind farm located in Brazil, demonstrating a correlation of 0.65 for hourly means, and 0.95 for monthly means. In terms of ERA5 temperature performance, Zou *et al.* (2022) investigated the differences between ERA5 surface temperatures and observations from 1080 automatic weather stations in southeast coastal China and concluded that ERA5 can capture daily and monthly temperature variations well. Biases of up to 2 K were, however, found by Zhao and He (2022) and Xu *et al.* (2022) between ERA5 temperatures and observations on the Tibetan Plateau, and attributed to elevation differences between ERA grid points and meteorological sites. The question of how representative ERA5 data are for the island location of Lerwick (located in the Shetland Isles, on an island 80 km long and 25 km wide) therefore arises, as many of the previous studies comparing ERA5 with observational datasets use inland locations far from coastal regions. Section 3 of this paper investigates directly how representative ERA5 data are for Lerwick.

2.3 | MIDAS

MIDAS is the open-data version of the Met Office Integrated Data Archive System (MIDAS) for land surface station data (1853–present) (Met Office, 2019). MIDAS Open consists of hourly and daily weather measurements of parameters such as temperature, rainfall, sunshine, radiation, wind and weather observations such as present weather codes, cloud cover, snow, etc. For this study we have explored WS (measured at 10 m), cloud cover, CBH (measured with an optical cloud base recorder) and pressure at Lerwick (all as hourly values). Precipitation data are also analysed, but this is only available for the 12-hour total rainfall accumulation, with values for 0900 and 2100 UTC.

2.4 | Fair-weather selection

At Lerwick observatory, the classification system for atmospheric electric PG measurements evolved over time. Initially, the approach was to categorise the level of daily disturbance in the measured values, following practice in geomagnetism measurements. Days were simply labelled as either ‘Quiet’ or ‘Disturbed’, depending on the variability observed (Harrison & Nicoll, 2018). The UK Met Office classification system recorded a day with only positive PG values as type 0, while days with negative PG records were categorised as 1 or 2 according to the duration of negative PG. A letter (a, b, or c) was appended after those numbers to signify the range and frequency of extreme PG values encountered (Harrison & Riddick, 2022). However, from 1957, a different approach was introduced and the daily mean PG values were calculated from only those measurements obtained during hours without precipitation. By 1964, a more detailed hour-by-hour classification method was introduced, applying the simultaneous meteorological conditions to classify the data into hours with ‘No Hydrometeors’ (NH) or ‘fair weather’ (FW).

To classify data as having been obtained in ‘fair weather’, four criteria were applied (taken from Harrison & Nicoll, 2018):

1. No hydrometeors (i.e., no rain, hail, or snow).
2. No low stratus clouds (cloud base required to be above 300 m).
3. Limited cumuliform clouds (up to three-eighths if no effect on PG or one-eighth if PG affected).
4. Mean hourly surface WS (at 10 m) below 8 m·s^{-1}.

These criteria aimed to avoid local disturbances that could affect PG measurements, such as charged precipitation, low clouds or convective clouds which might be charged, and high WS generating charged dusts and airborne soil particles. Applying these classifications has successfully allowed identification of global influences on PG data, such as the Carnegie curve and a relationship with sea surface temperatures modulated by El Niño (Harrison *et al.*, 2011, 2022).

3 | COMPARISON OF ERA5 AND MIDAS FOR IDENTIFYING FAIR-WEATHER CONDITIONS

In this section, with the motivation of identifying how representative ERA5 data are for Lerwick, as well as to help define specific fair-weather criteria for ERA5, we compare the available weather information in the ERA5 and MIDAS datasets for the Lerwick site. This comparison primarily concerns the analysis of meteorological conditions and so is independent of the availability of digitised PG data, which varies during the time period of the meteorological data.

In this section we will compare ERA5 reanalysis data with MIDAS ground-based measurements to evaluate the applicability of ERA5 in studies covering the period from

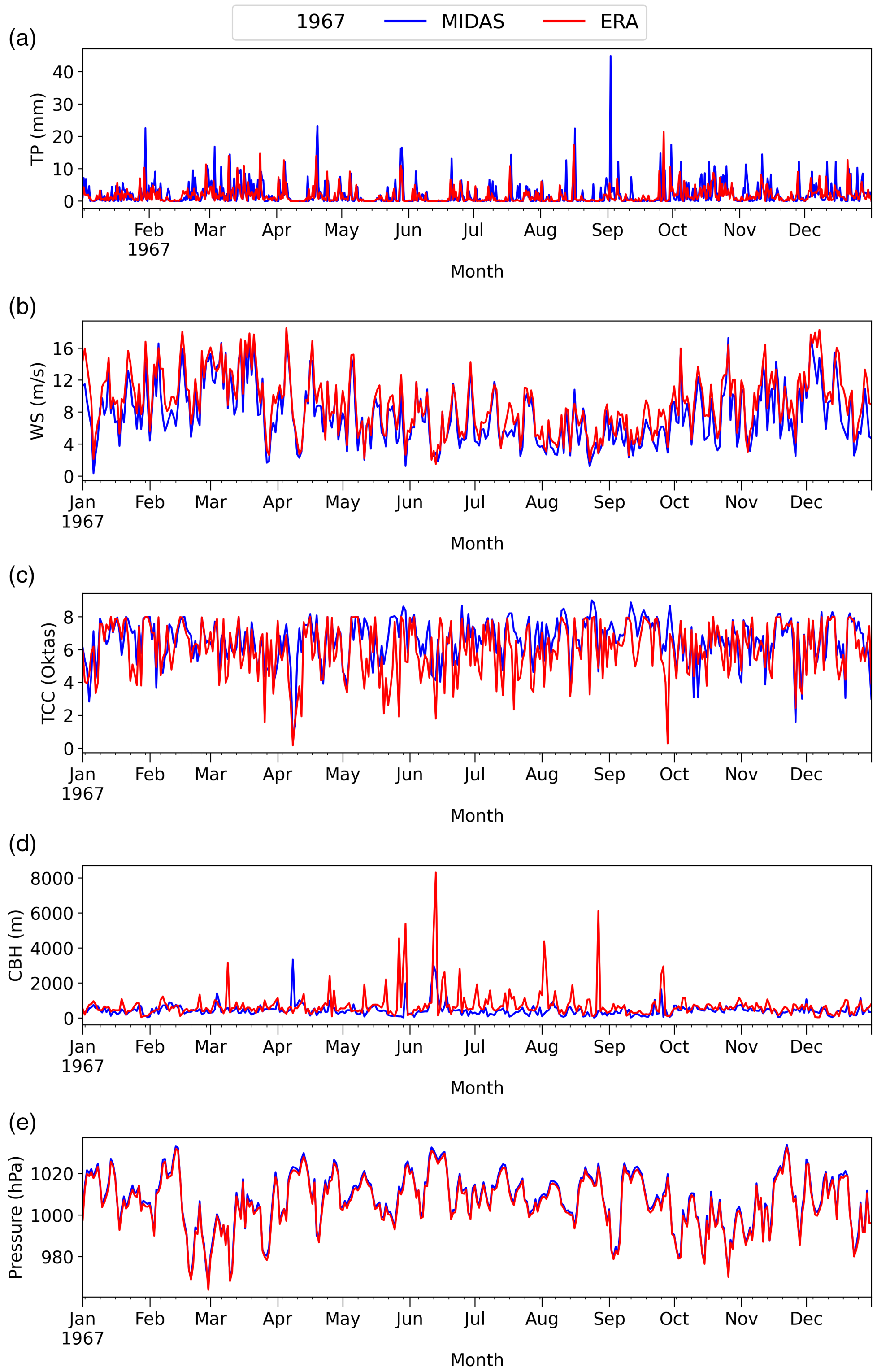


FIGURE 1 Time series of daily means of ERA5 (red) and MIDAS (blue) parameters for (a) total precipitation (TP) (12-hour accumulation), (b) wind speed (WS), (c) total cloud cover (TCC), (d) cloud base height (CBH) and (e) pressure during 1967. Daily means are calculated from the hourly values. [Colour figure can be viewed at wileyonlinelibrary.com]

1964 to 1980 for meteorological parameters which are relevant to defining fair-weather criteria for atmospheric electricity studies. Although hourly ERA5 data are available from 1940, hourly MIDAS records are not available from 1940, so we focus on the later period from 1964 onwards, in order to use the hourly data. It is preferable to select the PG data from 1967 onwards, to minimise the effects of nuclear testing on the PG which are apparent in PG records from Lerwick from the late 1950s until the mid-1960s (Harrison & Riddick, 2022). The data analysis is undertaken, firstly, by comparison of the daily mean time series of precipitation (12-hour accumulation), WS and cloud cover over a year, and secondly by examining specific daily case studies. Figure 1 provides a comparison between daily mean (averaged from the hourly values) observations from MIDAS (in blue) and daily mean data from ERA5 (in red) for different meteorological parameters recorded in 1967. Figure 1a shows TP (12-hour accumulation), demonstrating broad similarity throughout the year between MIDAS and ERA5, but with some major differences in magnitude during certain periods (e.g., September in particular). Precipitation occurs during all seasons, but particularly during the winter months, where there are several instances of larger values of precipitation in the MIDAS data than ERA5 (especially during November and December). Figure 1b shows WS, demonstrating a good general agreement between MIDAS and ERA5, and certainly much better than precipitation. Total cloud cover (TCC) is shown in Figure 1c, with no clear seasonal cycle, and again reasonable general agreement between MIDAS and ERA5. It is evident that there is better agreement during the winter months. Cloud properties are further investigated in Figure 1d, which examines CBH and demonstrates some major discrepancies between MIDAS and ERA5 during the summer months, where ERA5 seems to consistently estimate higher CBHs than MIDAS. This is consistent with the better agreement between ERA5 and MIDAS total cloud cover (TCC) in winter, which will be more fully investigated in Figure 4. Finally, Figure 1e demonstrates good agreement between MIDAS and ERA5 pressure values, which track closely throughout the year, indicating that ERA5 derived pressure values are likely to be very representative of observations at Lerwick. Further details of the distributions and correlations between ERA5 and MIDAS datasets will be analysed in Figures 3 and 4.

To investigate the similarities between ERA5 and MIDAS datasets over a shorter timescale (and to see how hourly values compare, rather than daily averages), Figure 2 displays hourly values of data over a single day for various meteorological parameters from selected days in 1967 (the same year as in Figure 1). One day from winter (21 February 1967) and one day from summer (30 August 1967) are presented for comparison. TP and pressure are not shown here, as the MIDAS database does not include hourly measurements for these variables in the 1960s, only one single daily average for pressure, and 12-hourly accumulation for precipitation. Figure 2a(i),(ii) shows comparisons of WS for the two days. Both days show a diurnal cycle in WS, which is captured by both ERA5 and MIDAS and in general good agreement. Figure 2b(i),(ii) shows more discrepancy in the TCC between ERA5 and MIDAS, as was found in Figure 1, with ERA5 generally producing more cloud cover than the MIDAS observations on these particular days (the largest discrepancy is a difference of 6 oktas at 0200 UTC in Figure 2b(i)). Finally, Figure 2c(i),(ii) also demonstrates large differences in the diurnal cycles of CBH, as well as the magnitudes. Although not shown in Figure 2, it is also instructive to compare the daily precipitation accumulations between MIDAS and ERA5 for these two days. For 21 February 1967 the daily precipitation accumulation was 10 mm for MIDAS and 4.5 mm for ERA5; whilst for 30 August 1967 it was 3.6 mm for MIDAS and 2.6 mm for ERA5, demonstrating reasonable differences on both days. From the initial analysis presented in Figures 1 and 2 it is clear that there are often large discrepancies in the cloud parameters (i.e., TCC and CBH) between ERA5 and MIDAS, which will now be more fully investigated in Figures 3 and 4.

Figure 3 compares the distributions of the daily means of meteorological parameters from MIDAS and ERA5, during 1964 to 1980. Each subplot corresponds to a different meteorological parameter and a summary of the medians and interquartile ranges (IQR) of the distributions is shown in Figure 3f. Figure 3a shows the distribution of WS, where it is seen that, although the IQRs of both distributions are similar, the ERA5 distribution is shifted to higher values (median of 8.3 $m \cdot s^{-1}$) than MIDAS (median of 6.5 $m \cdot s^{-1}$). For the TCC data presented in Figure 3b, MIDAS and ERA5 match well for moderate cloud cover (around 4–6 oktas), and they both show a significantly higher frequency of total overcast conditions (8 oktas), but there is less good agreement between 6 and 8 oktas. Figure 3c shows a comparison of CBHs, which shows reasonable agreement, but again with a shift in the ERA5 distribution to higher values. Atmospheric surface pressure, however, shows very good agreement between MIDAS and ERA5 (Figure 3d), with very similar medians and IQRs. Finally, total precipitation (12-hour accumulation) is compared in Figure 3e, where it is seen that the TP data in both datasets show that low precipitation values are common, and the distributions of ERA5 and MIDAS are broadly similar.

To more fully investigate the correlations between the meteorological parameters from ERA5 and MIDAS, Figure 4 provides scatter plots to display the relationships, along with their correlation coefficients (highlighted in red

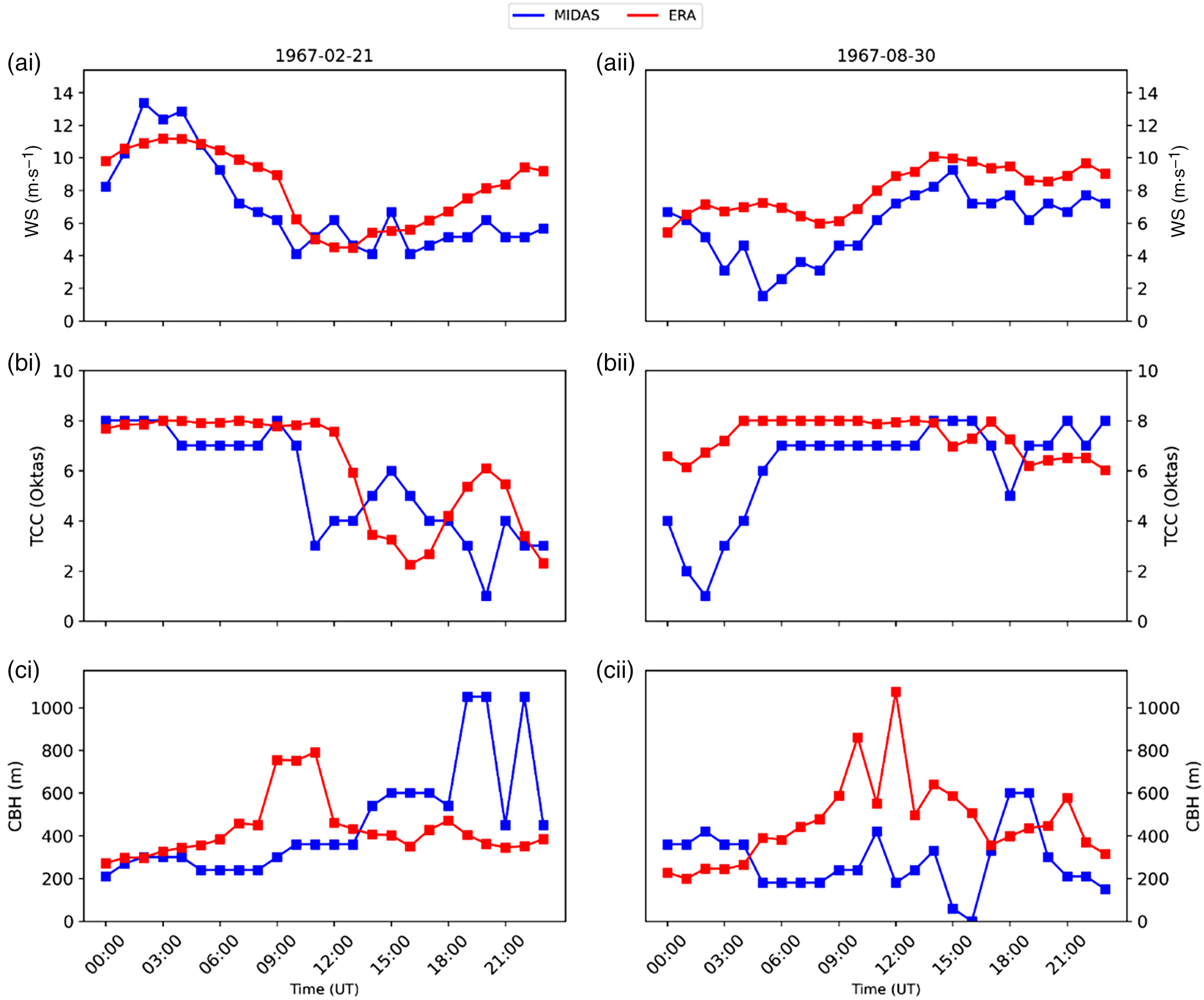
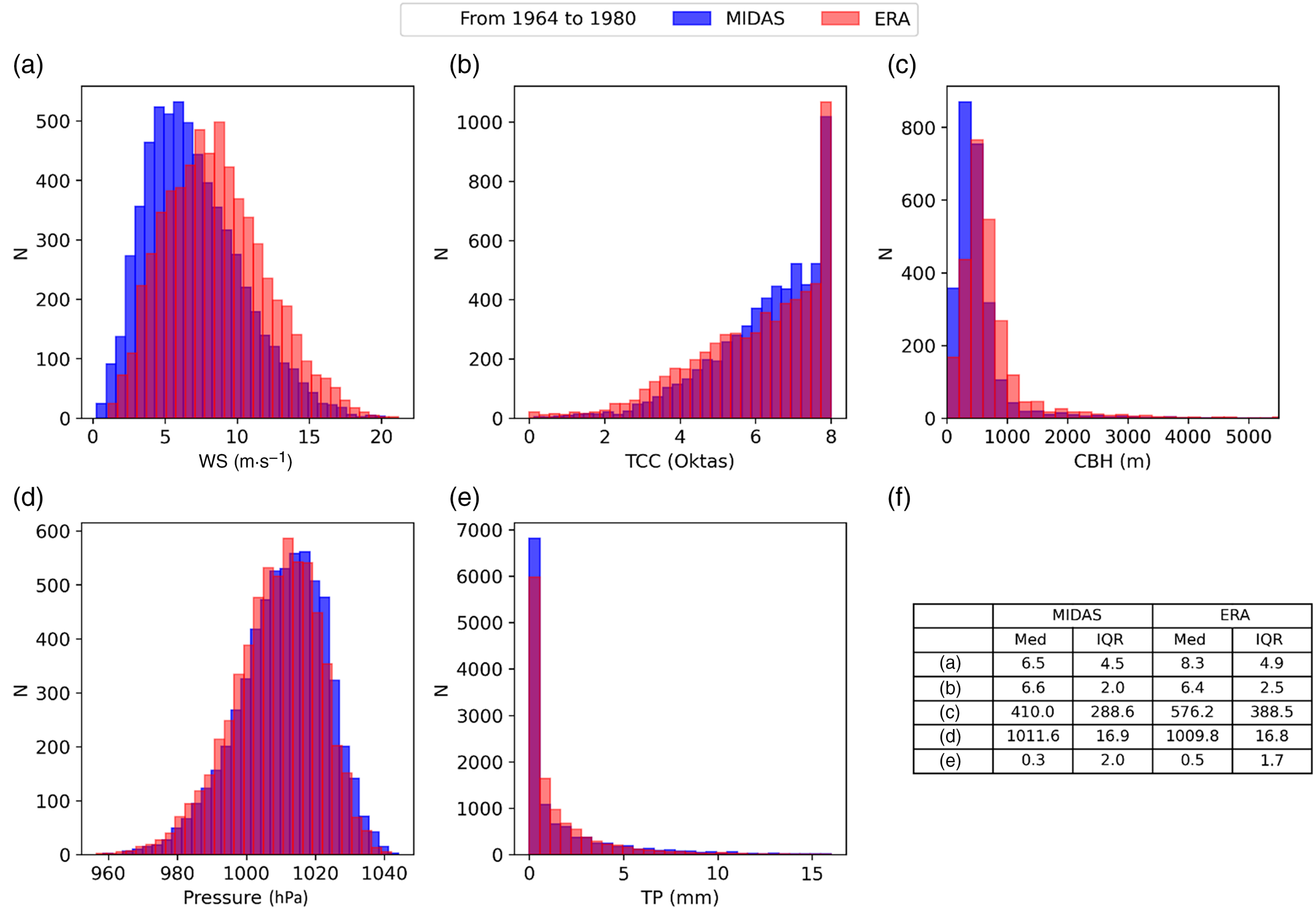


FIGURE 2 Case study illustrations for selected individual days. (i) 21 February 1967; (ii) 30 August 1967. Daily variation of hourly meteorological parameters from ERA5 and MIDAS for (a) wind speed (WS), (b) total cloud cover (TCC), (c) cloud base height (CBH). [Colour figure can be viewed at wileyonlinelibrary.com]

boxes) for data from 1964 to 1980. The data have been plotted for all seasons (black points) and for only December values (in red) to investigate whether there is a seasonal dependence for some parameters (e.g., cloud parameters as suggested by Figure 1). The month of December specifically is chosen as the PG analysis in Section 4 uses only December data (as only the December values have been fully digitised for 1940–1950). A table of the correlation coefficients for the relationship between ERA5 and MIDAS for each of the meteorological parameters by season is also provided in the Supporting Information (Table S1). Table S1 shows that the correlations for all parameters except precipitation are highest during winter months (Dec, Jan, Feb). WS is shown in Figure 4a and demonstrates a strong positive correlation of 0.92 (which is similarly good for December values), indicating a good match between the two datasets, with most points closely following a linear trend. TCC in Figure 4b has a moderate correlation of 0.69, showing some scatter, but this improves to 0.77 when only December values are analysed. CBH in Figure 4c is less well correlated (correlation of 0.39), with a large spread of data points, particularly at higher cloud amounts, indicating significant discrepancies between the datasets. Table S1 demonstrates that the correlation for CBH is best during the winter months, presumably as CBHs are typically lower and there is less spread in the CBH range. The correlation is best (0.55) for December, and the clustering of the red points towards the bottom left of the plot illustrates that the CBHs during December are typically much lower (<1500 m) than during other months, as might be expected. Atmospheric pressure shown in Figure 4d exhibits a near-perfect correlation



	MIDAS		ERA	
	Med	IQR	Med	IQR
(a)	6.5	4.5	8.3	4.9
(b)	6.6	2.0	6.4	2.5
(c)	410.0	288.6	576.2	388.5
(d)	1011.6	16.9	1009.8	16.8
(e)	0.3	2.0	0.5	1.7

FIGURE 3 Histograms of daily mean values of (a) wind speed (WS), (b) total cloud cover (TCC), (c) cloud base height (CBH), (d) pressure, and (e) total precipitation (TP) (12-hour accumulation), for both ERA5 (red) and MIDAS (blue) for 1964–1980. (f) The median and interquartile ranges (IQR) of the distributions for the WS (a), TCC (b), CBH (c), pressure (d), and TP (e). [Colour figure can be viewed at wileyonlinelibrary.com]

(correlation of 1.00), with the points lying on the 1:1 line. This is likely to be due to the pressure being assimilated in ERA-5 (as suggested by ERA-5 documentation[i] which states that surface pressure, temperature and relative humidity [SYNOP] data from land stations are typically assimilated). Studies in the literature find excellent agreement between measured and predicted surface pressure values (up to 0.98 R^2) for sites at which data has been assimilated (e.g., for the Arctic as discussed by Pernov *et al.*, 2024), provided the mean altitude of the ERA-5 grid cell and the site altitude are similar, which is likely to be the case as the Lerwick site altitude is only 82 m. Finally, TP (Figure 4e) has a correlation of 0.77, showing a positive relationship, though with some noticeable spread, particularly at higher precipitation values, suggesting moderate agreement between the datasets. Unlike the other parameters which show higher correlations during December, precipitation shows a lower correlation. Further analysis finds that out of a total of 11,787 precipitation values, 5880 (50%) of these were classified as Non-Hydrometeor (NH) cases by MIDAS (using a 12-hour total of <0.2 mm as NH), and 3816 (32%) classified as NH by ERA. Only 29% were classified as NH by both ERA and MIDAS, demonstrating that ERA5 overestimates precipitation on dry days in 21% of the cases. For the values classified as NH by MIDAS and not ERA, the difference in precipitation was generally small (mean of <0.4 mm), but a few cases differed by up to 13 mm. This disagreement in precipitation between ERA5 and point observations from MIDAS may be expected due to the highly localised nature of rainfall, and the relatively large size of ERA5 grid squares (31 × 31 km), demonstrating that precipitation derivation from ERA5 for a single site must be interpreted with care. In summary, there are strong correlations between ERA5 and MIDAS data for WS and pressure, but weaker correlations for precipitation, TCC and CBH. The improved correlations between the cloud parameters during winter months (i.e., December) may be related to the likelihood

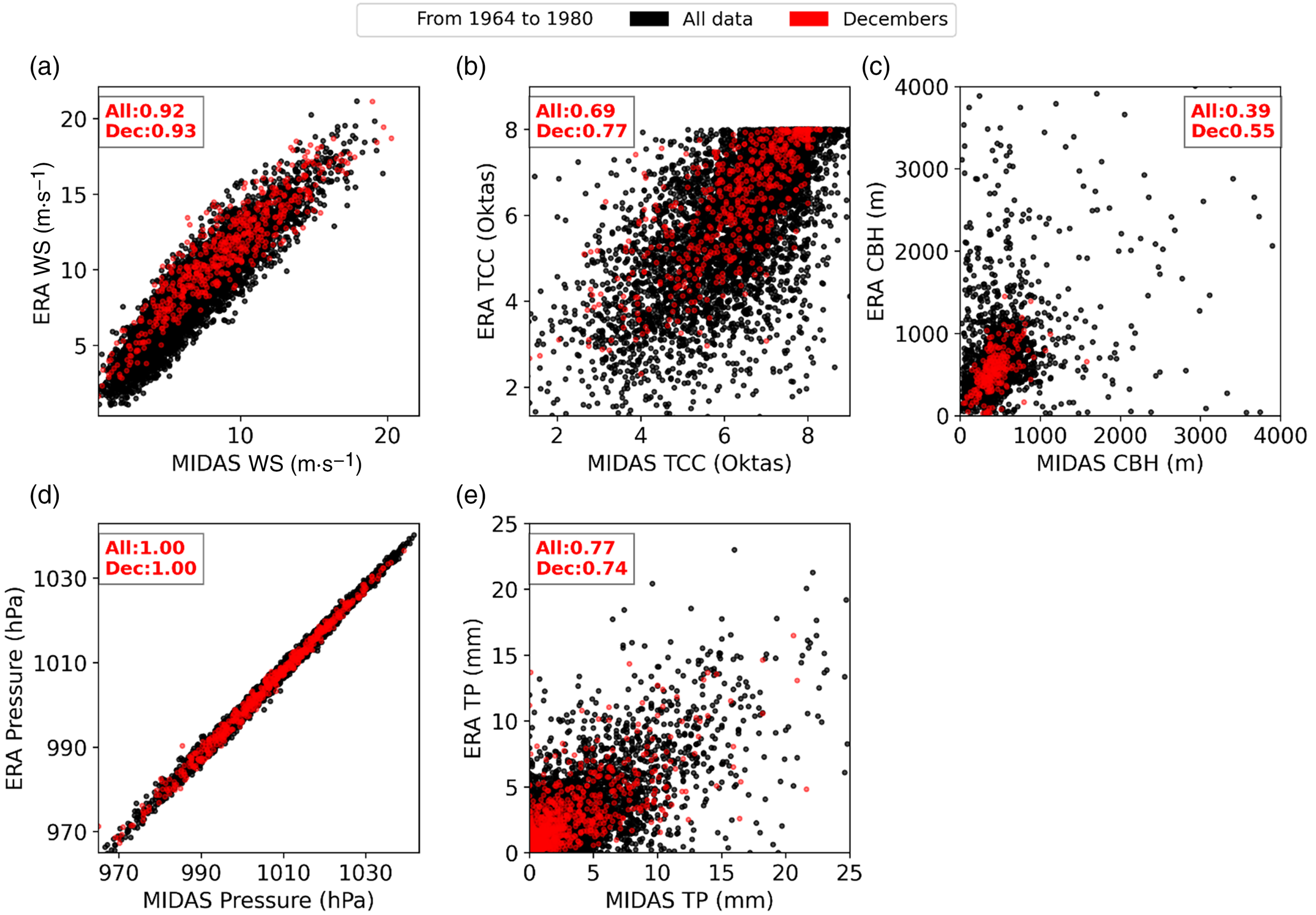


FIGURE 4 Scatter plot of daily means of ERA5 vs MIDAS data for (a) wind speed (WS), (b) total cloud cover (TCC), (c) cloud base height (CBH), (d) pressure, (e) total precipitation (TP) (12-hour accumulation) for 1964–1980. Black dots indicate data for all seasons and the red ones only December values. The red numbers in the top left of the figures denote the correlation coefficient (R) between each of the MIDAS and ERA5 variables. [Colour figure can be viewed at wileyonlinelibrary.com]

that the clouds at Lerwick are more stratiform and extensive during the winter, and therefore more likely to be captured by the relatively coarse horizontal resolution of ERA5 grid squares (31 × 31 km). This is in contrast to summer, when more localised convective clouds occur, which are unlikely to be accurately represented by ERA5. This is potentially more important for small island sites such as Lerwick, where the dimensions of the island (80 × 25 km) only cover a few ERA5 grid squares.

The aim of the analysis in this section has been to assess how representative ERA5 reanalysis data are of actual meteorological observations from Lerwick (represented by the MIDAS dataset), so that a set of criteria for fair weather can be determined using the ERA5 dataset. Section 2.4 summarises the meteorological criteria that have historically been used (e.g., by the UK Met Office) to classify fair-weather conditions for atmospheric electricity studies, and these are related to WS, cloud and precipitation. Figures 1–4 all demonstrate that WS and pressure are the quantities best estimated by ERA5. The histogram in Figure 3a confirms that selecting WS values between 2 and 8 m·s^{-1} (which represent the 0.8th and 46.5th percentiles of the ERA5 WS distribution respectively) to represent fair-weather values (as in Section 2.4) is sound, therefore we will implement this in our initial definition of ERA5 fair-weather criteria. Although there has generally been no inclusion of pressure in the typical definition of fair weather, we choose to use this for our criteria due to the ability of ERA5 to accurately represent pressure observations at Lerwick. Since fair-weather conditions tend to occur during periods of relatively high pressure, we will select conditions where the pressure is greater than 1000 hPa (23rd percentile of the ERA5 pressure distribution). Such conditions are likely to minimise periods of precipitation, which is known to have a major influence on PG. Figure S1 in the Supporting Information examines the relationship between pressure and precipitation from MIDAS more fully and demonstrates that although selection of pressure above 1000 hPa does not remove all of the precipitation, the amount of precipitation decreases for pressures above 1000 hPa. For example, if a precipitation event is defined as >0.2 mm (the minimum resolution of a tipping-bucket rain gauge), 16.1% of the total number of data points in Figure S1 would be classed as precipitation. Of these, approximately half (8.3%) are removed when pressure higher than 1000 hPa is selected. Although selecting a higher value for the pressure (e.g., >1020 hPa) would reduce the number of points with precipitation further, the number of data points left after applying such criteria is much reduced (e.g., for the analysis performed in Figure 6, the number of points is reduced by half when using >1020 hPa). This precludes any meaningful analysis to be performed, therefore 1000 hPa is retained. ERA5's determination of precipitation at Lerwick is reasonable (particularly for winter months) but far from perfect, but since it is not the absolute value of precipitation that is important for fair weather – only the criterion that there is no precipitation – we also implement a criterion based on TP < 0.05 mm·hour^{-1} for fair weather. The choice of 0.05 mm·hour^{-1} is derived from analysis performed in Figures 10 and S2 in the Supporting Information and discussed more fully in Section 4.2. Although the cloud parameters of TCC and CBH do not show particularly high correlations between ERA5 and MIDAS observations, it is known that the influence of charge in low-level stratiform cloud on PG is an important one and should be included if possible (Harrison & Nicoll, 2018). We therefore implement criteria based on CBH, selecting only values above 500 m. This is stricter than the original Met Office fair-weather criteria (discussed in Section 2.4) which employed a minimum CBH of 300 m, but it is informed by our findings at the University of Reading site, also in the United Kingdom (Harrison *et al.*, 2017).

4 | APPLICATION OF ERA5 FAIR-WEATHER CRITERIA TO PG MEASUREMENTS

4.1 | Period 1940–1950

The following section applies the fair-weather criteria for ERA5 meteorological parameters determined in Section 3 (i.e. WS between 2 and 8 m·s^{-1}, pressure above 1000 hPa, precipitation below 0.05 mm·hour^{-1}, and CBH above 500 m) to the Lerwick PG data to assess its effectiveness for selecting fair-weather PG measurements. Figure 5 investigates this by presenting three different daily case studies from 1940 to 1950 (from the database of Harrison, 2022), which show hourly PG and ERA5 meteorological data (WS, pressure, total precipitation and CBH). The red-shaded region in each plot shows time periods which are identified by fair weather using the ERA5

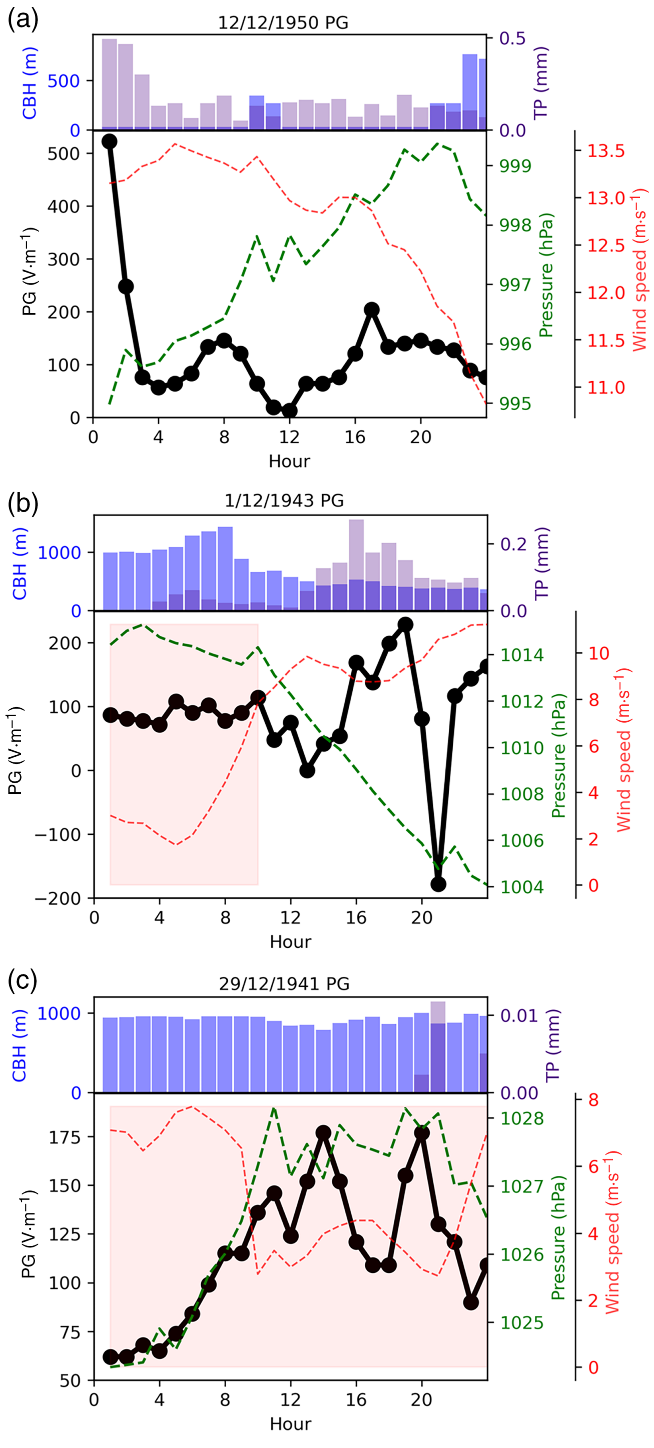


FIGURE 5 Three case studies of individual days demonstrating the daily variation of ERA5 meteorological parameters and PG during (a) non-fair weather, (b) half-fair weather, (c) fair weather. The shaded red region indicates when the fair-weather criteria are fulfilled according to wind speed (WS) between 2 and 8 m·s^{-1}, pressure above 1000 hPa, cloud base height (CBH) above 500 m, and TP below 0.05 mm·hour^{-1}. The top panel displays the CBH (blue bars) and total precipitation (TP) (purple bars), illustrating changes throughout the day. The bottom panel presents PG (dotted black), pressure (green) and WS (red dashed lines). [Colour figure can be viewed at wileyonlinelibrary.com]

criteria. Figure 5a shows an example of what the ERA5 classification would define as 'non-fair weather', with WS above 11 $m \cdot s^{-1}$, pressure between 995 and 999 hPa, and continuous cloud with a base height of 30 m for most of the day, rising to 1000 m during the late evening. Precipitation is also present throughout the day. Although the PG is generally less than 200 $V \cdot m^{-1}$, there is a large increase to 500 $V \cdot m^{-1}$ at 0100 UTC, which may be due to snowfall (as precipitation in the hours between 2100 UTC on the day before and 0900 UTC was recorded in the observations, and the air temperature was around 2°C). This day was originally classified as '1a' day type (using the daily electric character method based on the PG), which has been used in the past to examine GEC behaviour (Harrison & Nicoll, 2018). Our analysis criteria suggest that this day should not be interpreted as globally representative as it had far from fair-weather conditions.

The second case study in Figure 5b features a period of fair weather until 0900 UTC, indicated by the red-shaded area in the figure. However, as the day progresses, conditions shift to non-fair weather, with an increase in WS from 2 to 10 $m \cdot s^{-1}$, a decrease in pressure from 1014 to 1004 hPa, a decrease in CBH to 500 m and the onset of precipitation at 1200 UTC. The PG is relatively steady at around 100 $V \cdot m^{-1}$ until 1000 UTC when it becomes much more variable and goes negative at 2000 UTC. This transition provides a clear example of varying weather conditions within a single day and their impact on the PG measurements. The electrical-character day type was 1a/2a which has previously been interpreted as non-fair weather, but as is evident here, at least 10 hours of this day were likely to have been fair weather. Finally, for the third case study (Figure 5c) fair-weather conditions are present throughout the entire day (WS 3–8 $m \cdot s^{-1}$, pressure 1024 to 1028 hPa, CBH 900 m), as indicated by the continuous, red-shaded area. Under these conditions, the PG data follows a shape resembling the Carnegie curve (until light rain occurs at the end of the day), demonstrating globally representative conditions. This is generally only observable when local meteorological influences on the PG are negligible. These three case studies help to illustrate that the PG behaviour during these days is generally consistent with that expected from the meteorological conditions identified by ERA5. Further, the fair-weather criteria derived in Section 3 appear to be able to select PG data which is also consistent with typical PG behaviour during fair-weather conditions (e.g., $0 < PG < 200\ V \cdot m^{-1}$). It is also evident that this method helps to select individual fair-weather hours, rather than just a whole day, which has likely led to a much reduced fair-weather dataset being available for GEC studies historically.

Focusing more on the average expected diurnal variation in PG during fair weather (i.e., the Carnegie curve), Figure 6 investigates the effect of using different ERA5 criteria to select fair-weather conditions with the hourly PG values selected accordingly. For this, each of the meteorological variables included in our fair-weather criteria (i.e., pressure, TP, CBH and WS) will be analysed individually (Figure 6b–e), and then together (Figure 6f), as described in Table 2. The first and third columns of Figure 6 show all the available 'raw' hourly PG values which satisfy the different fair-weather criteria, and the second and fourth columns show the corresponding relative mean of the hourly PG values (in blue) and median PG (green). N.B. not all hourly values are available due to some missing data, presumably from instrument issues. The hourly PG values from the Carnegie curve (from Harrison, 2013, for November, December, January) are also shown in red. The top right legend of each figure indicates the number of hours of data retained in each case. Figure 6a(i) shows the hourly PG values during all meteorological conditions (using December data from 1940 to 1950), with PG values ranging from −200 to 600 $V \cdot m^{-1}$, with a lot of variability. When all of the 1687 PG values are averaged together by hour, Figure 6a(ii) demonstrates a diurnal variation very similar to the Carnegie shape, showing how effective averaging over many points is in reducing variability within a dataset.

The subsequent plots in Figure 6 show that the number of hourly PG datapoints is reduced as the various fair-weather criteria are applied (decreasing from 1687 to 557), with the WS criterion most dramatically reducing the number of points (Figure 6e), versus the others. When all of the meteorological criteria are applied (Figure 6f), the number of negative values and large ($>300\ V \cdot m^{-1}$) PG values decreases, demonstrating the removal of likely non-fair-weather PG values. Although there is good agreement between the Carnegie curve and the averaged PG data when no criterion is applied (shown in Figure 6a(ii)), this is typically a result of averaging many values together, as the raw data in Figure 6a(i) demonstrate large variability in the individual PG values. As the meteorological criteria become more strict, only the key data points carrying the GEC signal remain, and the Carnegie curve signal is identifiable, even without averaging. Thus application of the fair-weather meteorological criteria acts to identify those PG data points which carry the GEC signal most strongly.

Figure 6 analyses PG data for all electrical-character days (i.e. 0a, 1a, 2a) but analysis of the individual day types has also been performed, with the results for the 1a/2a days shown in Figure S2 in the Supporting Information. Traditionally, 1a/2a days have been regarded as non-fair-weather days and not used for GEC studies; however, application of the ERA5 fair-weather criteria demonstrates good agreement with the Carnegie curve and that 20% of the 1a/2a hourly values are actually

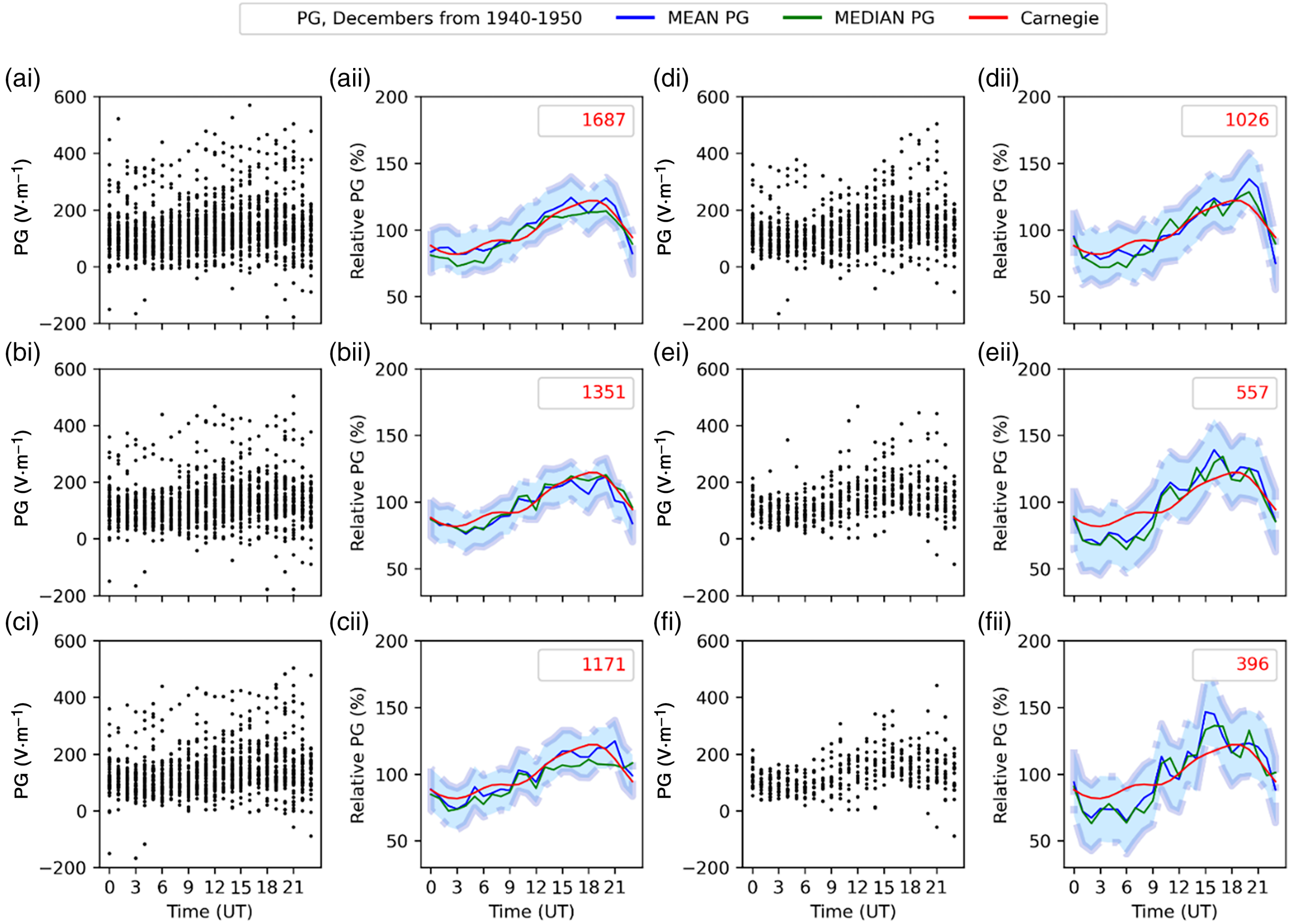


FIGURE 6 Diurnal variation in potential gradient (PG) at Lerwick (using available hourly PG from December in 1940–1950, taken from Harrison, 2022) under different fair-weather meteorological criteria for ERA5 specified in Table 2, that is, (a) no criteria, (b) pressure, (c) total precipitation (TP), (d) cloud base height (CBH), (e) wind speed (WS), (f) pressure, TP, CBH and WS. Columns 1 and 3 show the hourly PG values under these criteria. Columns 2 and 4 show the corresponding mean hourly PG (blue), median hourly PG (green) and Carnegie PG values (red) for each set of criteria (plotted as relative PG value which is a proportion of the mean of the remaining values). The blue envelope shows the standard deviation in the mean of the relative PG and the red number in the top right corner gives the number of PG values retained in each case. [Colour figure can be viewed at wileyonlinelibrary.com]

TABLE 2 Fair-weather meteorological criteria for ERA5 which have been used in Figures 6 and S2. The coloured blue boxes highlight which meteorological criteria are selected for Figures (a) to (f) in Figures 6 and S2.

Criteria\figure labels	(a)	(b)	(c)	(d)	(e)	(f)
No criteria	■					
Pressure > 1000 hPa		■				■
Total precipitation (TP) < 0.05 mm·hour^{-1}			■			■
Cloud base height (CBH) > 500 m				■		■
Wind speed (WS) between 2 and 8 m·s^{-1}					■	■

fair weather. This suggests that the 1a/2a classifications, despite variable meteorological conditions, are nevertheless relatively useful in identifying times from which fair-weather electrical characteristics can be extracted.

To investigate more fully the effect of applying the various fair-weather meteorological criteria to the Lerwick PG data, Figure 7 examines the distributions of hourly PG measurements from Figure 6 (i.e., using the data from the plots in columns 1 and 3 of Figure 6). Each subplot (b–f) in Figure 7 compares the distribution of all PG data (Figure 7a) (blue) with PG data selected according to the meteorological criteria (red) from Table 2. The

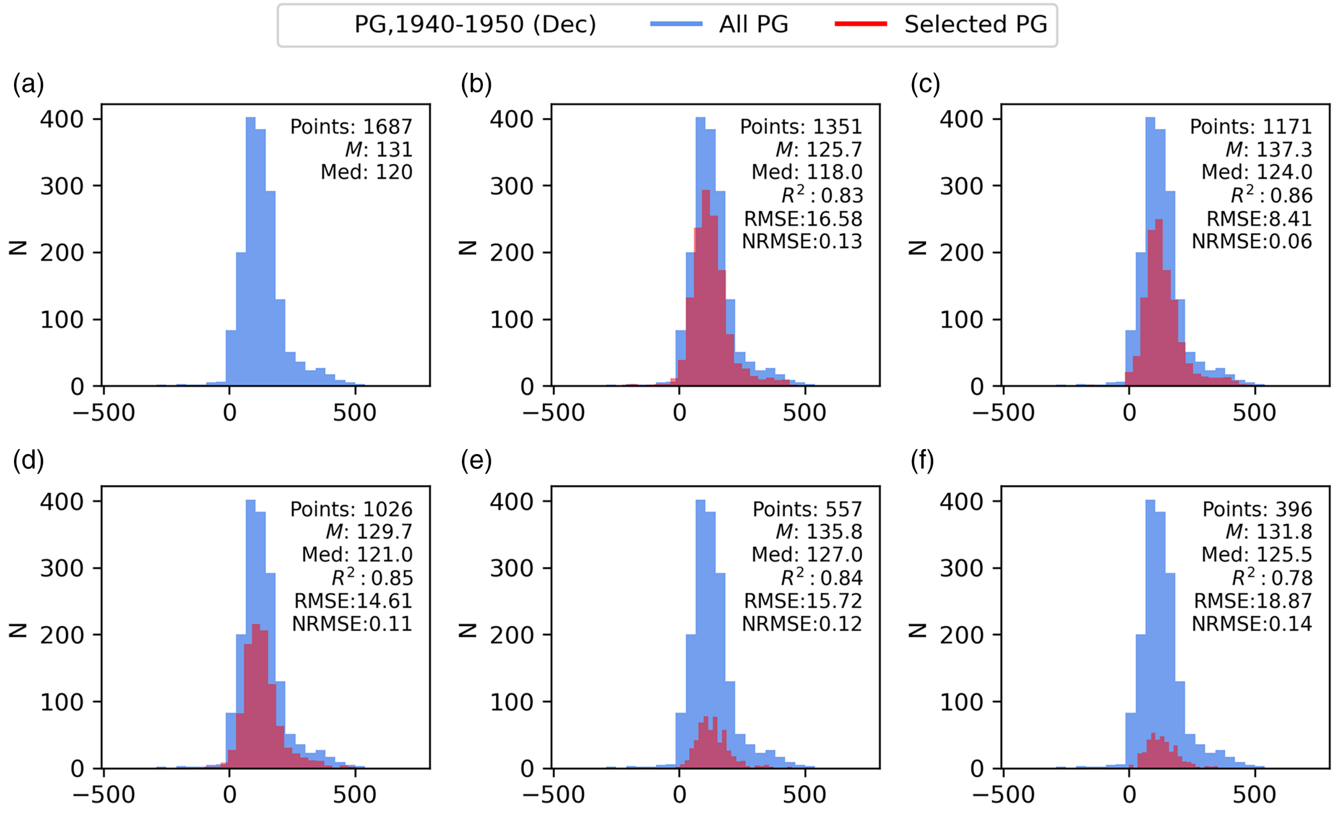


FIGURE 7 Distribution of hourly PG values (from Decembers in 1940–1950). The blue bars are for PG during all meteorological conditions (with 1687 values) and red bars are PG selected according to the different fair-weather criteria (given in Table 2). The number on the top line at the right-hand side in each plot is the total sample size before and after applying the fair-weather selection criteria. Statistical metrics such as *M* (mean), Med (median), R^2, RMSE (root-mean-squared error), and NRMSE (normalised RMSE) are also given. [Colour figure can be viewed at wileyonlinelibrary.com]

plots indicate a narrowing of the PG distribution when fair-weather criteria are applied, with a noticeable shift in the mean (*M*) and median (Med) values. The numbers on the top line in the right-hand side in each plot represent the total sample size before and after applying the fair-weather selection criteria. Statistical metrics such as R^2, RMSE (root-mean-squared error), and NRMSE (normalised RMSE) are also included in each subplot, demonstrating the correlation and variability reduction between the total and selected datasets.

The analysis in this section confirms that the initial criteria suggested for fair-weather selection of PG data using ERA5 in Section 3 are sensible. These criteria will now be applied to the later PG data from Lerwick to investigate how well the ERA criteria perform against the meteorological classification originally applied to the PG data at Lerwick observatory.

4.2 | Period 1967–1980

The latter part of the PG data record at Lerwick (from 1967 to 1980) has so far only digitised the monthly mean PG values by hour of day, as well as the number of hours of Fair Weather (FW) and NH per month (as described in Table 1, taken from Harrison *et al.* (2023)). Figure 8 shows the daily mean PG time series for this time period for (a) FW hours, and (b) NH hours as originally classified by the Met Office. Perhaps unexpectedly, there is a wider range of PG values during FW than for the NH cases, but the histograms of the PG distributions in Figure 8 show that the difference between the mean values is small (158 V·m^{-1} for FW and 160 V·m^{-1} for NH).

For the final analysis, Figures 9 and 10 investigate the ability of the ERA5 fair-weather criteria to accurately select FW and NH hours respectively. For this, our fair-weather criteria of WS between 2 and 8 m·s^{-1}, pressure above 1000 mb, TP below 0.05 mm·hour^{-1} and cloud base height (CBH) above 500 m are applied to all hourly ERA data from 1967 to 1980. The number of hours of fair weather per month is derived from ERA5 and shown in Figure 9, alongside the number of hours of fair weather per month as classified by Lerwick observatory (data taken from Mkrtchyan *et al.* (2024)).

Figure 9a,b shows the distribution of the number of FW hours according to ERA5 and Lerwick respectively, where

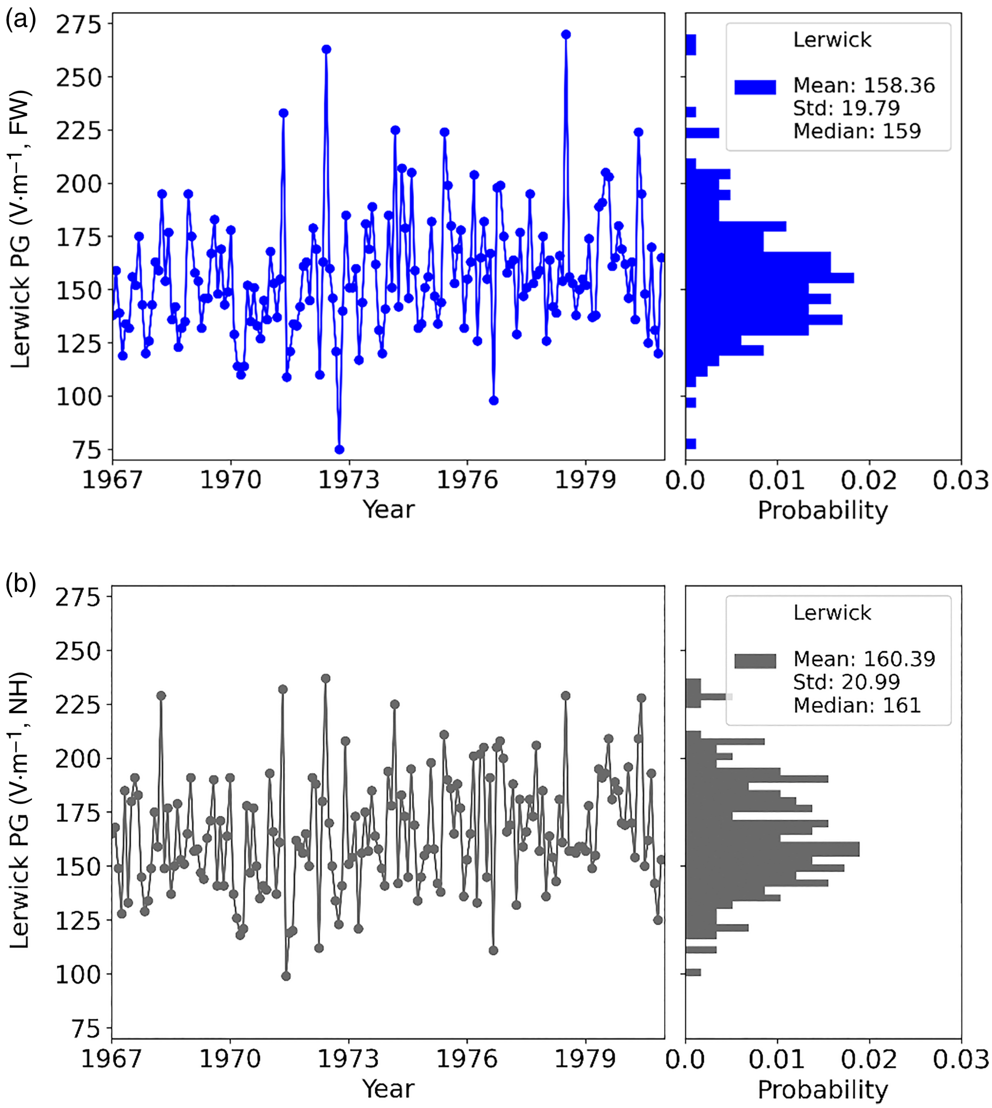


FIGURE 8 Left hand plots: time series of daily mean PG values for (a) Fair Weather (FW) and (b) Non-Hydrometeors (NH) from 1967 to 1980 (as explained in Section 2.2). Right-hand plots: histogram of distribution of daily mean PG values for (a) FW hours and (b) NH hours. [Colour figure can be viewed at wileyonlinelibrary.com]

it is seen that the means of the distributions are quite different (151 hours for ERA5 and 217 hours for Lerwick). There are also discrepancies in the tails of the distributions, with ERA5 deriving many more small numbers of FW hours per month (i.e. <100 hours) than Lerwick, but also Lerwick classifying many more months with large numbers of FW hours (i.e. >300 hours). The correlation between the two datasets is 0.59 as shown in Figure 9c. Despite these discrepancies, the time series in Figure 9d shows reasonable agreement between the two datasets, with many of the features showing similar behaviour, particularly from 1974 onwards.

This analysis has been repeated in Figure 10 for the NH hours derived from ERA5 using TP less than 0.05 mm·hour^{-1}. Panels (a) and (b) demonstrate that the shape of distributions of the number of NH hours per month are very different between ERA5 and Lerwick observatory, with the Lerwick values being much more normally distributed than ERA5. The ERA5 data have a wider range compared to the original Lerwick observatory classification, and Figure 10c shows that in general there are more NH hours detected by ERA5 than Lerwick for the same data points (with a correlation of 0.7). There are, however, broad similarities in the time series of NH hours from both datasets, with Figure 10d showing that the fluctuations in the data are similar. The seasonal variation in NH hours is much more evident in the ERA5 values, however, presumably due to climatological assumptions about precipitation in the ERA5 reanalysis model. The exact choice of which value of TP to classify as NHs is investigated in Figure S3 in the Supporting Information. This demonstrates the effect of using a range of values for TP (from 0.001 to 0.1 mm·hour^{-1}) and shows that the best agreement between ERA5-derived and Lerwick-classified NH hours is for total precipitation below 0.05 mm·hour^{-1}. For values less than 0.05 mm·hour^{-1} the number of NH hours is considerably underestimated, and above 0.05 mm·hour^{-1} they are overestimated.

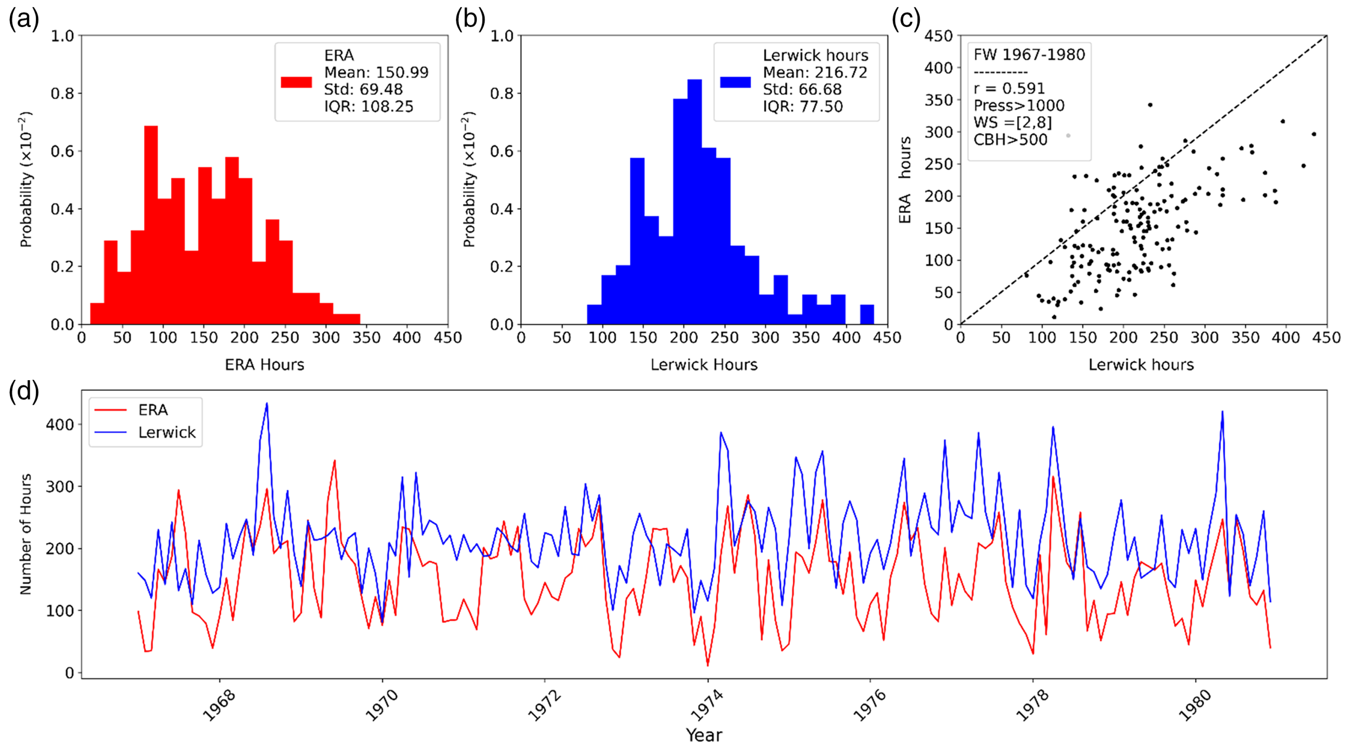


FIGURE 9 Comparison of Fair Weather (FW) hours per month derived from ERA5 using meteorological FW criteria described in Section 3, and those classified directly by Lerwick observatory from 1967 to 1980. Histograms show the distribution of FW hours derived from (a) ERA5 and (b) Lerwick Observatory (with the mean, standard deviation (Std) and interquartile range (IQR) shown in top right). (c) The number of FW hours from ERA5 plotted against FW hours from Lerwick Observatory (with the FW criteria for ERA5 specified in the top left of the plot). The dashed line is the 1:1 line; (d) the time series of ERA5 FW hours (red) and Lerwick Observatory FW hours (blue). [Colour figure can be viewed at wileyonlinelibrary.com]

5 | DISCUSSION

This work has investigated whether meteorological reanalysis data from the ERA5 dataset can be used to define FW conditions for atmospheric electricity measurements. The analysis in Section 2 investigates how effective ERA5 is at deriving certain meteorological parameters important for fair weather, which includes WS, cloud parameters (TCC and CBH) and precipitation. Although not historically used for FW determination, we have also included surface pressure in our analysis. It is well known that direct comparisons between model parameters, such as ERA5 (which represent an average over a model grid box) and site-specific observations have significant limitations (Hersbach *et al.*, 2020). The good correlation between ERA5 and the MIDAS observations for pressure and WS at a specific site is encouraging and expected as these dynamically related variables tend to be better simulated by models. Some of the observations (most likely, pressure) may have been assimilated and have therefore influenced the reanalysis to some extent as well; however, precipitation and cloud within ERA5 are associated with parameterisation and higher uncertainty, which is reflected in the poorer correlations with MIDAS. It is known that the uncertainty is more pronounced in regions that are poorly sampled since ERA5 integrates observations from surface and radiosondes, and only includes satellite data from 1979. Studies, including those by Sheridan *et al.* (2020), have indicated that ERA5 shows poorer performance in coastal areas and the Arctic, where observations are scarce, highlighting potential limitations in these regions. Further, the few observations associated with a remote island location are likely to lead to large uncertainties for localised meteorological phenomena, such as convective cloud and rainfall. Hence, for this study we limited the ERA5 data to the immediate grid square for Lerwick, which may be a more demanding application of the ERA5 data than for representing sites situated in extensive homogeneous terrain.

The final choice of suitable fair-weather criteria for ERA5 was based on balancing the choice of parameters which are accurately derived by ERA5 with those which can significantly influence atmospheric electrical variables. Precipitation has a large influence on PG; however, the reduced reliability of ERA5 for precipitation (particularly in the case of convective rainfall events) means

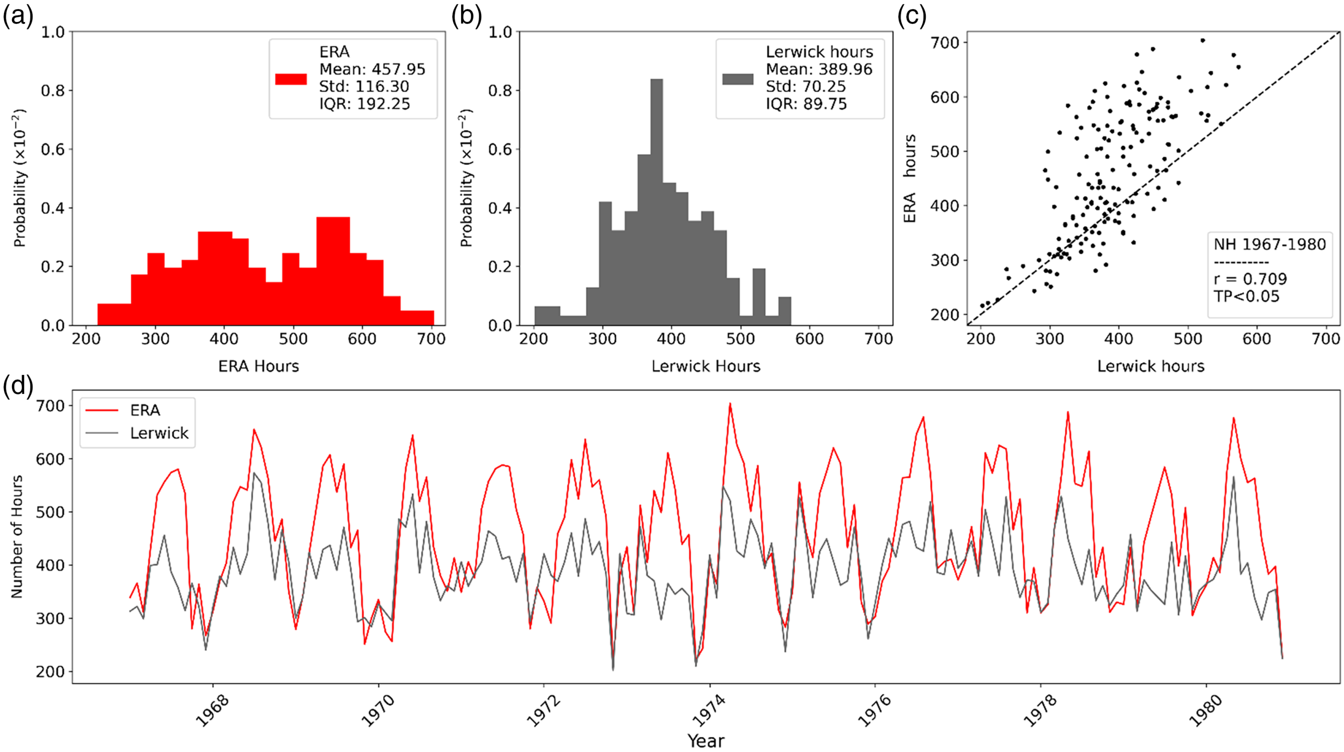


FIGURE 10 Comparison of Non-Hydrometeor (NH) hours per month derived from ERA5 using meteorological criteria (i.e., ERA5 total precipitation (TP) <0.05 mm·hour^{-1}), and those found classified directly at Lerwick observatory from 1967 to 1980. Details are as per the caption for Figure 9. (d) The time series of ERA5 NH hours (red) and Lerwick Observatory NH hours (grey). [Colour figure can be viewed at wileyonlinelibrary.com]

that it is less well suited in identifying FW and NH circumstances. To minimise these effects, criteria for surface pressure were also implemented, as a rough proxy for rainfall, given the greater confidence in the ERA5 surface pressure data. By only selecting surface pressure values greater than 1000 hPa, disturbed weather conditions (and likely those with precipitation) were less likely, but not excluded entirely. Although relatively large uncertainty in the ERA5 cloud parameters was also demonstrated in Section 2, the height of the cloud base is known to be an important factor in controlling the variability in PG, and more so than TCC (Harrison & Nicoll, 2018). Hence, CBH was included in the final fair-weather criteria. Our choice of 500 m CBH is more conservative than the original Met Office requirements of 300 m, but less stringent than requiring a CBH of 1000 m as suggested by analysis of cloud base effects at Reading (Harrison *et al.*, 2017). The analysis in Figure 4 demonstrates that ERA5 CBH data correlate better with the MIDAS observations during Decembers compared to annual values. We speculate that this is due to the greater frequency of extensive stratus cloud/frontal systems during Lerwick winters. Their large horizontal extent likely means that ERA5 is more able to accurately simulate such cloud decks, in comparison to much smaller, more localised convective cloud systems, which may only be one or two grid squares in spatial extent. Hence, applying reanalysis data to other sites may need further local refinement in the quantities considered.

The analysis presented in Figures 6 and S1 demonstrates the usefulness of applying ERA5 data to historical PG measurements which have been classified using the electrical-character method. Previous studies have tended to use only the 0a days for GEC studies, as these were thought to be least influenced by local phenomena and therefore be more globally representative. However, there are indications that this could be site-dependent: unlike Lerwick, at Eskdalemuir, the 0a classification is found to yield different median values to the other classifications (see Figure 8 of Harrison & Riddick, 2024). For Lerwick, our analysis demonstrates that around 20% of the 1a/2a data also meet the FW criteria, suggesting that data from the site, which have historically been considered as poor quality, may well contain useful information and should therefore not be disregarded.

Although this study has focused solely on the Lerwick location, it is likely that the technique of using reanalysis data to identify FW criteria will also be useful at other sites measuring atmospheric electricity. For this, careful attention should be paid to the representability of reanalysis data at individual locations, particularly in the case of coastal, high-altitude or high-latitude sites, as well as consideration of what constitutes appropriate fair-weather criteria for specific sites, which may well differ from those used at UK sites such as Lerwick.

6 | CONCLUSION

The usefulness of several reanalysis variables in ERA5 for atmospheric electricity data selection has been explored by comparing the WS, pressure, TCC, precipitation and CBH in ERA5 with direct observations at the Lerwick observatory site. Following these comparisons, selected ERA5 data have been applied to classifying past atmospheric electricity data, specifically that of the hourly PG which was obtained at Lerwick observatory from 1925 to 1984 (Harrison & Riddick, 2022). This study has identified the following criteria from ERA5 as useful for classifying hourly fair-weather conditions at Lerwick: WS between 2 and 8 $m \cdot s^{-1}$, pressure above 1000 hPa, precipitation less than 0.05 $m \cdot hour^{-1}$, and CBH above 500 m.

In the early period, from 1940 to 1950, the original classification of the PG data was solely based on the 'Electrical-Character Method', which summarized the overall appearance of each day's electrogram trace. This led to categorising PG records as 0a, 1a, or 2a, in increasing amounts of disturbance, and in the most disturbed cases with added letters b or c. Our analysis shows that applying selected meteorological parameters from the ERA5 dataset to some of the existing historical classifications can successfully identify fair-weather conditions within days which were previously regarded as of poor quality and with an assumed unlikely prospect of global representativity. Using the reanalysis data reclassifies 20% of hourly data within 1a or 2a days as actually fair weather, which suggests that these days are nevertheless useful for extracting atmospheric electricity data without strong local influences. This indicates an additional possible source of data for GEC studies.

In conclusion, this work demonstrates that meteorological reanalysis data, and specifically the ERA5 dataset, can provide the relevant local parameters with which fair-weather atmospheric electricity conditions at a site can be identified. The same principles can be readily applied to PG measuring sites which lack their own co-located meteorological measurements, for improving the quality and global representativity of their PG data.

ACKNOWLEDGEMENTS

The authors acknowledge the work of many observers at Lerwick observatory who have collected the data, and the many supporters, including through Zooniverse, who have keyed PG data from the Lerwick site. Hripsime Mkrtchyan acknowledges funding from UKRI for Postdoc Guarantee Fellowship Grant EP/X024547/1 (AtmosEleC).

DATA AVAILABILITY STATEMENT

PG data from Lerwick are archived at the University of Reading Research Data archive at https://doi.org/10.17864/1947.001367 (Mkrtchyan *et al.*, 2024; Harrison, 2022; Harrison *et al.*, 2023). Hourly ERA5 meteorological data were obtained from Hersbach *et al.* (2023), downloaded from the Copernicus Climate Change Service (2023). MIDAS meteorological observation data were downloaded from Met Office (2019). Data used in preparing the figures in this paper can be found at https://figshare.com/s/8999d393e024debd7aca. The Carnegie data were obtained from Harrison (2013).

ENDNOTE

[i]https://confluence.ecmwf.int/display/CKB/ERA5%3A+data+documentation#ERA5:datadocumentation-Table15.

ORCID

H. Mkrtchyan https://orcid.org/0000-0002-2921-1384
K. A. Nicoll https://orcid.org/0000-0001-5580-6325
R. G. Harrison https://orcid.org/0000-0003-0693-347X

REFERENCES

Chen, X., Cao, D., Liu, Y., Xu, X. & Ma, Y. (2023) An observational view of rainfall characteristics and evaluation of ERA5 diurnal cycle in the Yarlung Tsangbo grand canyon, China. *Quarterly Journal of the Royal Meteorological Society*, 149(753), 1459–1472. Available from: https://doi.org/10.1002/qj.4468

Dutra, E., Muñoz-Sabater, J., Boussetta, S., Komori, T., Hirahara, S. & Balsamo, G. (2020) Environmental lapse rate for high-resolution land surface downscaling: an application to ERA5. *Earth and Space Science*, 7(5), e2019EA000984. Available from: https://doi.org/10.1029/2019ea000984

Harrison, G. (2022) *Atmospheric electricity data for El Niño-southern oscillation studies*. University of Reading. Dataset. Available from: https://doi.org/10.17864/1947.000409

Harrison, R.G. (2013) The Carnegie curve. *Surveys in Geophysics*, 34(2), 209–232. Available from: https://doi.org/10.1007/s10712-012-9210-2

Harrison, R.G., Joshi, M. & Pascoe, K. (2011) Inferring convective responses to El Niño with atmospheric electricity measurements at Shetland. *Environmental Research Letters*, 6, 044028. Available from: https://doi.org/10.1088/1748-9326/6/4/044028

Harrison, R.G., Mkrtchyan, H. & Nicoll, K.A. (2025) Atmospheric electricity data from Lerwick during 1964 to 1984. *Geoscience Data Journal*, 12(3), e70009.

Harrison, R.G. & Nicoll, K.A. (2018) Fair weather criteria for atmospheric electricity measurements. *Journal of Atmospheric and Solar-Terrestrial Physics*, 179, 239–250. Available from: https://doi.org/10.1016/j.jastp.2018.07.008

Harrison, R.G., Nicoll, K.A. & Aplin, K.L. (2017) Evaluating stratiform cloud base charge remotely. *Geophysical Research*

Letters, 44, 6407–6412. Available from: https://doi.org/10.1002/2017GL073128

Harrison, R.G., Nicoll, K.A., Joshi, M. & Hawkins, E. (2022) Empirical evidence for multidecadal scale global atmospheric electric circuit modulation by the El Niño-southern oscillation. *Environmental Research Letters*, 17(12), 124048. Available from: https://doi.org/10.1088/1748-9326/aca68c

Harrison, R.G., Nicoll, K.A. & Mkrtchyan, H. (2023) *Lerwick observatory monthly mean potential gradient 1925–1984*. University of Reading. Dataset. Available from: https://doi.org/10.17864/1947.000505

Harrison, R.G. & Riddick, J.C. (2022) Atmospheric electricity observations at Lerwick geophysical observatory. *History of Geo- and Space Sciences*, 13, 133–146. Available from: https://doi.org/10.5194/hgss-13-133-2022

Harrison, R.G. & Riddick, J.C. (2024) Atmospheric electricity observations at Eskdalemuir geophysical observatory. *History of Geo- and Space Science*, 15, 5–16. Available from: https://doi.org/10.5194/hgss-15-5-2024

Hersbach, H. (2019) ECMWF's ERA5 reanalysis extends back to 1979. *ECMWF Newsletter*, 158, 1.

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J. et al. (2020) The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, 146, 1999–2049. Available from: https://doi.org/10.1002/qj.3803

Hersbach, H., Bell, B., Berrisford, P., Biavati, G., Horányi, A. & Muñoz Sabater, J. (2023) ERA5 hourly data on single levels from 1940 to present. Copernicus Climate Change Service (C3S) Climate Data Store (CDS). Available from: https://doi.org/10.24381/cds.adbb2d47 (Accessed on 25/08/2023).

Met Office. (2019) *Met Office MIDAS open: UK land surface stations data (1853-current)*. Centre for Environmental Data Analysis. Available from: http://catalogue.ceda.ac.uk/uuid/dbd451271eb04662beade68da43546e1/ (Accessed September 2023)

Mkrtchyan, H., Harrison, G. & Nicoll, K. (2024) *Lerwick observatory monthly mean potential gradient by hour of day 1964-1984*. University of Reading. Dataset. Available from: https://doi.org/10.17864/1947.001367

Molina, M.O., Gutiérrez, C. & Sánchez, E. (2021) Comparison of ERA5 surface wind speed climatologies over Europe with observations from the HadISD dataset. *International Journal of Climatology*, 41(10), 4864–4878.

OYB. (1965) *The observatories' year book, annual volumes for 1965*. London, UK: Meteorological Office, HMSO, p. 21. Available from: https://geomag.bgs.ac.uk/data_service/data/yearbooks/YBpdf/YB_ESKLER_1965.pdf

Pernov, J.B., Gros-Daillon, J. & Schmale, J. (2024) Comparison of selected surface level ERA5 variables against in-situ observations in the continental Arctic. *Quarterly Journal of the Royal Meteorological Society*, 150(761), 2123–2146.

Santos, J., Sakagami, Y., Haas, R., Passos, J., Machuca, M.N., Radünz, W.C. et al. (2019) Wind speed evaluation of MERRA-2, ERA-interim, and ERA5 reanalysis data at a wind farm located in Brazil. In: *Proceedings of the ISES Solar World Congress*, pp. 1–10.

Sheridan, S.C., Lee, C.C. & Smith, E.T. (2020) A comparison between station observations and reanalysis data in the identification of extreme temperature events. *Geophysical Research Letters*, 47, e2020GL088120. Available from: https://doi.org/10.1029/2020GL088120

Slyunyaev, N.N., Ilin, N.V., Mareev, E.A. & Price, C.G. (2021) A new link between El Niño-southern oscillation and atmospheric electricity. *Environmental Research Letters*, 16, 044025. Available from: https://doi.org/10.1088/1748-9326/abe908

Whipple, F.J.W. & Scrase, F.J. (1936) Point discharge in the electric field of the earth. In: *Geophysical memoirs of the meteorological office, 68*. London: His Majesty's Stationery Office. Available from: https://digital.nmla.metoffice.gov.uk/download/file/IO_efb7fd62-a542-4266-aac6-cbdeae3311b6

Wilson, C.T.R. (1921) III. Investigations on lighting discharges and on the electric field of thunderstorms. *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character*, 221(582–593), 73–115.

Xu, W., Lei, X., Chen, S., Yu, T., Hu, Z., Zhang, M. et al. (2022) How well does the ERA5 reanalysis capture the extreme climate events over China? Part II: extreme temperature. *Frontiers in Environmental Science*, 10, 921659.

Zhao, P. & He, Z. (2022) A first evaluation of ERA5-land reanalysis temperature product over the Chinese Qilian Mountains. *Frontiers in Earth Science*, 10, 907730.

Zou, J., Lu, N., Jiang, H., Qin, J., Yao, L., Xin, Y. et al. (2022) Performance of air temperature from ERA5-land reanalysis in coastal urban agglomeration of Southeast China. *Science of the Total Environment*, 828, 154459.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Mkrtchyan, H., Nicoll, K.A. & Harrison, R.G. (2025) Evaluating meteorological reanalysis for identifying fair-weather conditions in historical atmospheric electricity data. *Quarterly Journal of the Royal Meteorological Society*, e5066. Available from: https://doi.org/10.1002/qj.5066