Rescuing historical weather observations improves quantification of severe windstorm risks

Hawkins, E. ORCID: https://orcid.org/0000-0001-9477-3677, Brohan, P., Burgess, S. N., Burt, S. ORCID: https://orcid.org/0000-0002-5125-6546, Compo, G. P. ORCID: https://orcid.org/0000-0001-5199-9633, Gray, S. L. ORCID: https://orcid.org/0000-0001-8658-362X, Haigh, I. D., Hersbach, H. ORCID: https://orcid.org/0000-0001-5330-7071, Kuijjer, K., Martinez-Alvarado, O. ORCID: https://orcid.org/0000-0002-5285-0379, McColl, C., Schurer, A. P. ORCID: https://orcid.org/0000-0002-9176-3622, Slivinski, L. and Williams, J. ORCID: https://orcid.org/0000-0002-8421-4481 (2023) Rescuing historical weather observations improves quantification of severe windstorm risks. Natural Hazards and Earth System Sciences, 23 (4). pp. 1465-1482. ISSN 1684-9981 doi: 10.5194/nhess-23-1465-2023 Available at https://centaur.reading.ac.uk/111770/

Nat. Hazards Earth Syst. Sci., 23, 1465–1482, 2023
https://doi.org/10.5194/nhess-23-1465-2023

Natural Hazards and Earth System Sciences
Open Access
EGU

Rescuing historical weather observations improves quantification of severe windstorm risks

Ed Hawkins[1], Philip Brohan[2], Samantha N. Burgess[3], Stephen Burt[1], Gilbert P. Compo[4,5], Suzanne L. Gray[6], Ivan D. Haigh[7], Hans Hersbach[3], Kiki Kuijjer[7], Oscar Martínez-Alvarado[1,6], Chesley McColl[4,5], Andrew P. Schurer[8], Laura Slivinski[4,5], and Joanne Williams[9]

[1]National Centre for Atmospheric Science, Department of Meteorology, University of Reading, Reading, UK
[2]Met Office Hadley Centre, Exeter, UK
[3]Copernicus Climate Change Service, ECMWF, Reading, UK
[4]Cooperative Institute for Research in Environmental Sciences, University of Colorado at Boulder, Boulder, USA
[5]NOAA Physical Sciences Laboratory, Boulder, USA
[6]Department of Meteorology, University of Reading, Reading, UK
[7]School of Ocean and Earth Science, National Oceanography Centre, University of Southampton, Southampton, UK
[8]School of Geosciences, University of Edinburgh, Edinburgh, UK
[9]National Oceanography Centre, Liverpool, UK

Correspondence: Ed Hawkins (ed.hawkins@ncas.ac.uk)

Received: 16 October 2022 – Discussion started: 27 October 2022
Revised: 9 February 2023 – Accepted: 2 March 2023 – Published: 24 April 2023

Abstract. Billions of historical climatological observations remain unavailable to science as they exist only on paper, stored in numerous archives around the world. The conversion of these data from paper to digital could transform our understanding of historical climate variations, including extreme weather events. Here we demonstrate how the rescue of such paper observations has improved our understanding of a severe windstorm that occurred in February 1903 and its significant impacts. By assimilating newly rescued atmospheric pressure observations, the storm is now credibly represented in an improved reanalysis of the event. In some locations this storm produced stronger winds than any event during the modern period (1950–2015) and it is in the top-4 storms for strongest winds anywhere over land in England and Wales. As a result, estimates of risk from severe storms, based on modern period data, may need to be revised. Examining the atmospheric structure of the storm suggests that it is a classic Shapiro–Keyser-type cyclone with "sting-jet" precursors and associated extreme winds at locations and times of known significant damage. Comparison with both independent observations and qualitative information, such as photographs and written accounts, provides additional evidence of the credibility of the atmospheric reconstruction, including sub-daily rainfall variations. Simulations of the storm surge resulting from this storm show a large coastal surge of around 2.5 m, comparing favourably with newly rescued tide gauge observations and adding to our confidence in the reconstruction. Combining historical rescued weather observations with modern reanalysis techniques has allowed us to plausibly reconstruct a severe windstorm and associated storm surge from more than 100 years ago, establishing an invaluable end-to-end tool to improve assessments of risks from extreme weather.

1 Introducing Storm Ulysses

Extreme wind events are among the costliest natural disasters in Europe. Significant effort is dedicated to understanding the risk of such events, usually using observed storms in the modern era (e.g. Roberts et al., 2014), and synthetic event sets or ensemble seasonal hindcasts designed to sample a wider range of plausible storms (e.g. Sharkey et al., 2020; Walz and Leckebusch, 2019). Severe historical storms that occurred before around 1950 are largely unstudied because atmospheric reanalyses usually only cover the modern

era, and atmospheric reanalyses that do exist for earlier periods may not represent severe storms plausibly due to the sparseness of the observations available to constrain the atmospheric circulation. However, it is likely that some earlier historical windstorms were more extreme and/or followed different tracks from those in the modern era. Expanding the numbers of reconstructed severe historical storms will improve our understanding of the risks from such events today and in the future.

Achieving this goal requires making more historical observations available to be used in reanalyses by (1) improving access to already digitized observations and (2) extracting additional observations from archival material. Here we demonstrate how the digitization of weather observations from paper archives has improved the reconstruction of one particular extreme storm and enabled the creation of a credible reanalysis of the event.

Between 26–27 February 1903 a violent windstorm passed across Ireland and the UK, causing many deaths, several shipwrecks, and considerable damage to infrastructure. For example, the Royal National Lifeboat Institution (RNLI) recorded 10 major rescues of crew from ships in distress, and *The Times* newspaper reported damage across the country, with considerable numbers of injuries and loss of life. In a special report on the event, Shaw (1903) described locations where damage or casualties occurred both on land and at sea. Figure 1 reproduces the summary figure from Shaw (1903), which also indicates the estimated path of the storm.

Figure 2 includes three photographs showing trees uprooted in Dublin (Ireland), damage to a pier in Morecambe, and a train blown over in Cumbria (both in NW England). A written account of the storm experienced in Carlisle (NW England) is also included. The damage in Ireland even inspired a passage in the novel *Ulysses*, written by James Joyce, with the events set the year after the storm in 1904:

> O yes, J. J. O'Molloy said eagerly. Lady Dudley was walking home through the park to see all the trees that were blown down by that cyclone last year and thought she'd buy a view of Dublin.

To pay homage, this windstorm is called Storm Ulysses (Met Eireann, 2017).

2 Reconstructing Storm Ulysses

Modern dynamical reconstructions of historical windstorms rely on reanalyses that assimilate observations of surface pressure that were taken at the time, over both land and ocean, into an atmospheric model, in a similar process to making the initial conditions for a modern weather forecast. In this study we use the NOAA-CIRES-DOE 20th Century Reanalysis version 3 system (20CRv3; Compo et al., 2011; Slivinski et al., 2019a; Slivinski et al., 2021), which has previously produced atmospheric reconstructions for the 1806–2015 period at 0.7° horizontal resolution, generating 3-hourly data, with 80 ensemble members to sample uncertainty in the reconstruction.

We perform novel experiments with this reanalysis system to demonstrate the value of assimilating additional surface pressure observations to better constrain the atmospheric circulation during Storm Ulysses. We evaluate the reanalyses against independent observations and then use the reanalyses to drive a storm surge model and compare against newly rescued tide gauge observations.

Figure 3a shows the synoptic situation according to 20CRv3 (ensemble mean) at 09:00 UTC on 27 February 1903, with a low-pressure cyclone situated over the British and Irish Isles. However, the reanalysis is uncertain about some details of the synoptic situation, with regions of >7 hPa ensemble standard deviation over northern UK (Fig. 3d). The depth of the low in the ensemble mean (967 hPa) is shallower than an estimate made soon after the event (around 960 hPa; Shaw, 1903), but note that the minimum pressure in individual ensemble members is 960 ± 9 hPa (1 standard deviation), highlighting that position and timing uncertainty is making the storm appear shallower in the ensemble mean. The red dots in Fig. 3d represent the locations of available pressure observations, which are assimilated between 06:00 and 12:00 UTC on this day to produce the reanalysis. These observations are relatively sparse, preventing the reanalysis from being able to accurately identify the location of the low pressure and hence represent the severity of the storm. For example, there were no available observations over England or Wales. This is a common feature of such reanalyses when examining extreme events occurring many decades ago (Brönnimann et al., 2013; Meyer et al., 2017) and currently limits the usefulness of such reconstructions for examining individual severe weather events.

However, since the International Surface Pressure Database (Cram et al., 2015) version 4 (Compo et al., 2019) used within 20CRv3 was assembled, two citizen science projects have rescued additional pressure observations for this period and region, which can be used to improve the reanalysis. Thousands of volunteers transcribed millions of meteorological observations from scanned copies of paper records (Hawkins et al., 2019; Craig and Hawkins, 2020), and a few additional records have been digitized specifically for a short period around this event. In total, pressure observations from 89 locations have been added (60 over the British and Irish Isles), with most providing two observations per day (see Appendix A for more details).

The 20CRv3 system has been used to repeat the assimilation process for Storm Ulysses including these new observations. An additional experiment was performed that also included a small improvement to the data assimilation scheme, which ensured that the 20CRv3 ensemble was more representative of the uncertainty (see Appendix B for more details). Figure 3b and c show the synoptic situation in the improved reanalysis experiments; note an additional isobar

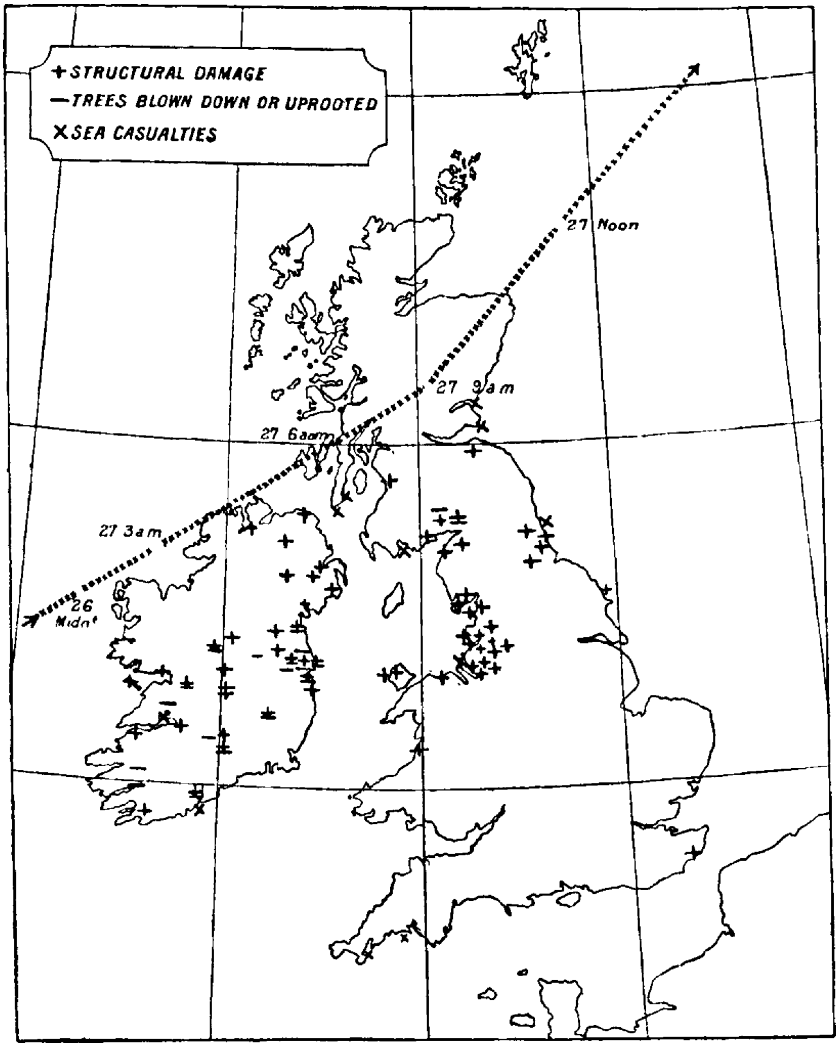

Figure 1. Post-storm estimate for the track of Storm Ulysses and locations of damage caused. Map taken from Shaw (1903).

highlighting a deeper low pressure which is more consistent with the estimate made at the time. Across the ensemble members in these two additional experiments, the minimum low pressure depths are 960 ± 5 and 960 ± 3 hPa, respectively, highlighting the improved confidence in the position and timing of the storm. The isobars are also closer together over the British and Irish Isles, meaning that the highest wind speeds over both land and sea have increased by 15 %–20 %. The increased density of available observations (Fig. 3e; dark blue dots) has reduced the uncertainty in the reconstruction, and the ensemble spread is further reduced when the assimilation scheme is improved (Fig. 3f), becoming more reliable when compared with independent data (see Appendix B).

Figure S1 shows the mean sea level pressure evolution of the storm in 20CRv3 and the two experimental versions of the reanalysis. The experiments with additional observations show minimum pressures around 956 hPa at slightly earlier times than shown in Fig. 3.

3 Reconstructions of wind speeds and the atmospheric circulation

3.1 Surface winds

To examine the severity of this storm in more detail, we first consider the near-surface wind speeds. Figure 4a–c shows the wind footprints of Storm Ulysses for 20CRv3 and the two experiments; these are maps of the ensemble mean maximum 10 m wind speed experienced at each location using instantaneous 3-hourly reanalysis data during the storm. Gust strength data are not available from the reanalysis. The two experiments produce higher wind speeds, particularly in areas of known significant damage such as eastern Ireland and northern England (see Fig. 1). Given that damage related to wind is approximately proportional to the cube of the wind speed (Lamb, 1991; Klawa and Ulbrich, 2003), even a small

Figure 2. Visual descriptions of damage from Storm Ulysses. Top: photographs from Dublin (left) and Morecambe (right). Middle: photograph of a train blown over on Leven viaduct. Bottom: written account of the storm in Carlisle and a map of locations in the photos or named in the text. The Dublin image was supplied by Aida Yared. The Leven photograph was taken by a Mr Alexander, assistant engineer for the Furness railway. The Morecambe image is a scan of a postcard owned by one of the paper authors.

increase in wind strength can result in significantly increased storm damage.

To quantify the relative strength of these simulated winds, it is necessary to compare against other windstorm events in the same reanalysis. We chose to compare with all events during 1950–2015 in 20CRv3, as this represents the period typically available from commonly used reanalyses of the modern period (such as ERA5; Hersbach et al., 2020). If a historical storm is unusual relative to this modern period, then it adds significant information about windstorm risk. Note that 20CRv3 does not yet extend beyond 2015.

Figure 4d–f shows the ranking of the winds experienced during Storm Ulysses compared to all events during the 1950–2015 period. For the original 20CRv3 reanalysis, Storm Ulysses is not particularly unusual, with wind speeds in the top-10 events for some small areas (Fig. 4d). However, when the reanalysis is better constrained by additional observations, Storm Ulysses is in the top-5 strongest wind events for larger areas across the UK, Ireland, and the North Sea (Fig. 4e). Once the assimilation process is also improved, the reanalysis of Storm Ulysses produces the strongest winds of any event for some locations (Fig. 4f), demonstrating the value of having additional observations to constrain the atmospheric circulation to better understand risks.

When looking across the whole of England and Wales, the peak 10 m wind speed over land during Storm Ulysses in the

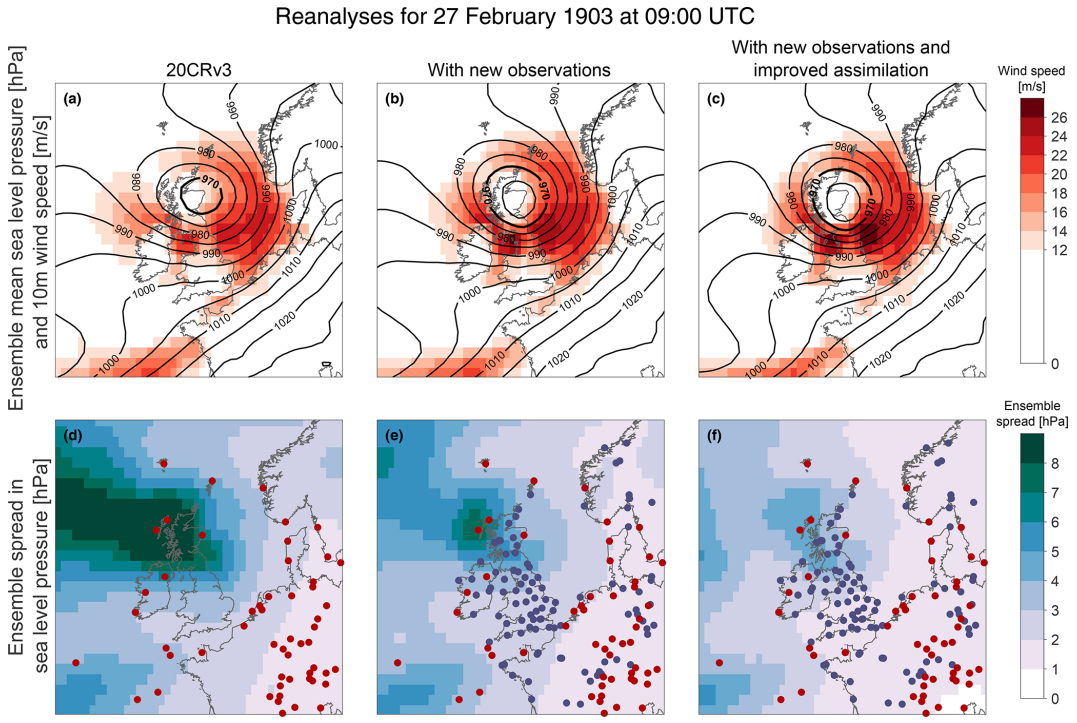

Figure 3. Reconstructing the atmospheric circulation during Storm Ulysses. Synoptic situation at 09:00 UTC on 27 February 1903. Isobars of sea level pressure (hPa, black contours) and wind strength at 10 m (m s^{-1}, red shading) from the ensemble mean of 80 reanalysis fields are shown from 20CRv3 and the two different experiments (top row, **a–c**). Standard deviation of the ensemble of sea level pressure reanalysis fields ("ensemble spread") for the same time (hPa, blue shading, bottom row, **d–f**). Locations with available surface pressure observations in 20CRv3 are shown as red dots, and new added observations in the experiments are shown as dark blue dots. Observations from both land stations and ships are shown, but there are very few available ship observations in this region at this time.

improved reanalysis is similar to the three most severe storms in the modern era, as represented by 20CRv3. Those storms occurred in 1990 (Burns Day Storm), 1997 (Yuma), and 1998 (Fanny), each affecting a slightly different part of the country. Note that this comparison is restricted to the ensemble mean of simulated 10 m winds from instantaneous 3-hourly data and does not account for gusts, so it may miss some of the most extreme winds from any particular storm (e.g. the 1987 Great Storm). If also including both Ireland and Scotland, the Boxing Day Storm of 1998 produced stronger winds than Storm Ulysses in this reanalysis. Regardless of the precise rankings, Storm Ulysses is an extreme windstorm in the context of the modern era, and we can now say that with confidence, even though it occurred over 100 years ago.

This type of historical information is highly relevant to sectors such as insurers, who need to understand the risks of extreme windstorms over the ocean (Buchana and McSharry, 2019) and over the land (Koks and Haer, 2020). Windstorm catalogues (e.g. Roberts et al., 2014) tend to consider the more recent period only, although it is recognized that this may not give a complete picture (Zimmerli and Renggli, 2015). Incorporating detailed information from significant historical storms such as Storm Ulysses is likely to improve estimates of windstorm risk.

3.2 Atmospheric conditions and sting-jet precursors

Although only surface pressure observations are assimilated, they can substantially constrain the lower part of the atmosphere in the reanalyses, and the three-dimensional structure of the storm provides valuable information. We first examine 850 hPa wet-bulb potential temperature (θ_w) during the storm, with a focus on 06:00 UTC on 27 February (Fig. 5). This quantity is commonly used to identify warmer and cooler air masses. High values of relative humidity (RH) at 700 hPa are indicated by stippling to highlight the approximate location of the cloud head. The features visible, such as the hooked cloud head and developing warm seclusion, indicative of frontal fracture, are consistent with a classic Shapiro–Keyser-type cyclone (Shapiro and Keyser, 1990). These features are more pronounced in the improved reanalysis experiments, and Figs. S2 and S3 show their development during the storm.

Figure 5 (bottom row) shows wind speed at 850 hPa and highlights two separate regions of higher wind speeds; this level was chosen to avoid contamination from strong orographic signals. In all the reanalyses there is an extended area of strong winds in the cyclone's warm sector ($\theta_w > 284$ K), but in the experiments, the strongest winds occur in an appar-

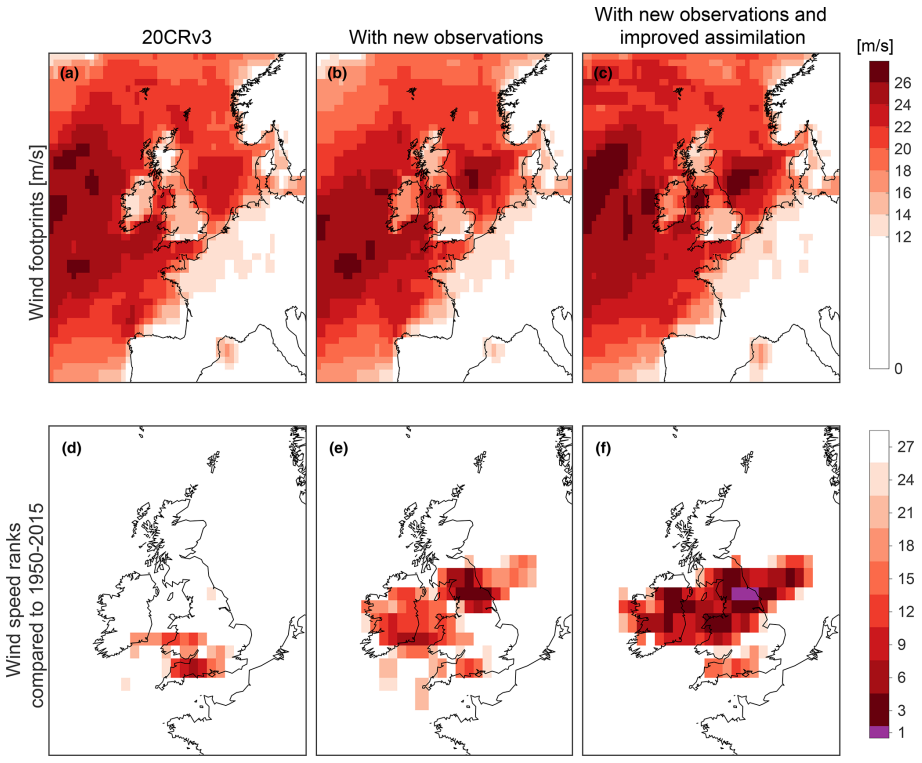

Figure 4. Comparing wind speeds with other events in the modern era. Wind footprints (top row, m s^{-1}, a–c, the maximum 10 m wind speed experienced at each location using instantaneous 3-hourly reanalysis data during the storm) and ranking of 10 m wind speed compared to all events during 1950–2015 (bottom row, d–f) for Storm Ulysses. The columns show 20CRv3 and the two experiments with the reanalysis system. Purple colours indicate locations where Storm Ulysses would have been the strongest observed had it occurred during 1950–2015.

ent frontal fracture zone, just to the south of the low pressure centre at the rear of the cold front and extending rearwards from this as the storm develops (see Figs. S4 and S5). Such strong winds found to the cold side of the bent-back front, which lies along the inner edge of the cloud head, are typically attributable to the cold conveyor belt jet. As this jet extends to the south of the storm, the alignment with the storm's direction of travel yields strong Earth-relative winds. These intense wind jets are typical for this type of storm but are not present in the ensemble mean of 20CRv3. However, a small number of individual ensemble members in 20CRv3 do have a coherent wind jet in this region.

There are not usually observations of wind speed at 850 hPa anywhere for this historical period, but there is one existing high-frequency record from the period of Storm Ulysses to which we can compare the reanalyses at this height. Meteorologists living at an atmospheric observatory on the summit of Ben Nevis (1345 m above sea level, at 56.8° N, 5.0° W) recorded detailed weather observations manually every hour from 1883–1904, including temperature, rainfall, pressure, wind speed, and wind direction (Hawkins et al., 2019). The hourly pressure observations from this observatory are included as some of the new observations added into the reanalysis. This observatory was usually at a height of roughly 850 hPa, but during Storm Ulysses the observed pressure fell to 810 hPa at 05:00 UTC. The summit observers measured force 10–11 winds from 02:00–03:00 UTC on 27 February, which, on the extended wind scale used, is equivalent to around 45 m s^{-1} (Hawkins et al., 2019). The improved reanalysis shows the highest 850 hPa wind speeds of 28–38 m s^{-1} (5 %–95 % range) at 03:00 UTC on 27 February, whereas 20CRv3 simulates 11–40 m s^{-1} (5 %–95 % range) for the same time. Although this is only a single location, it is an encouraging agreement on the timing of peak winds at this elevation. It is difficult to evaluate the amplitude of the wind speeds given that the reanalysis has a coarse resolution relative to the orography in this region, but the improved reanalysis appears more consistent with the available observations.

The wind speed at 850 hPa is often used as an estimate of maximum surface gust speed (Hart et al., 2017), so it can also be compared to information about known damage near sea level. It is notable that the train on the Leven viaduct (Fig. 2) was blown over at 05:30 UTC (Board of Trade, 1903), consistent with the timing of 850 hPa winds, and therefore potential surface gusts, of above 40 m s^{-1} in that region in the improved reanalyses (Fig. 5). 20CRv3 does not simulate such strong winds at this location.

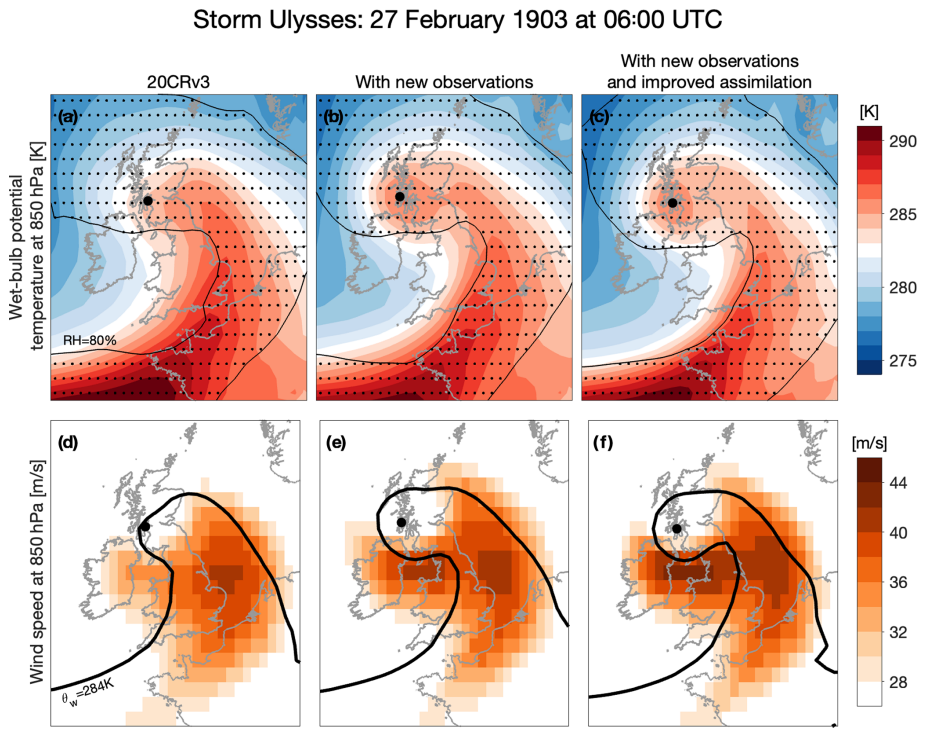

Figure 5. Ensemble mean of wet-bulb potential temperature at 850 hPa (top row, a–c) and wind speed at 850 hPa (bottom row, d–f) in the original 20CRv3 and two experiments with the reanalysis system (columns) at 06:00 UTC on 27 February 1903. The filled black circle represents the position of the mean sea level pressure minimum. The top panels also include stippling and a contour representing the location of the cloud head using relative humidity (RH) with respect to ice of above 80 % at 700 hPa. The 284 K isotherm is indicated with the thick black contour in the bottom panels.

It is often the case that the greatest damage from Shapiro–Keyser windstorms comes from meso- and convective-scale phenomena, and this type of cyclone is known to produce sting jets. A sting jet is "a coherent air flow that descends from mid-levels inside the cloud head into the frontal-fracture region of a Shapiro–Keyser cyclone over a period of a few hours leading to a distinct region of near-surface stronger winds" (Clark and Gray, 2018; after Browning, 2004). These small-scale features cannot be explicitly resolved in the relatively low-resolution model used to generate the available reanalyses, but a metric has been developed for diagnosing precursor conditions suitable for sting-jet formation (Martínez-Alvarado et al., 2012). Mesoscale instability release has been shown to occur in storms with intense sting jets (Gray et al., 2011; Volonté et al., 2018), and the precursor metric assesses the presence in the storm's cloud head of a type of mesoscale convective instability called conditional symmetric instability using a diagnostic called DSCAPE (downdraught slantwise convective available potential energy). The metric was shown to be skilful in identifying storms (from low-resolution model output) in which sting jets developed in corresponding high-resolution simulations capable of resolving mesoscale instability release (Martínez-Alvarado et al., 2012), and it is now applied routinely by the Met Office to provide information relevant to issuing severe wind warnings (Gray et al., 2021).

Figure 6 (top row) shows the track of the storm (defined as the location of minimum interpolated sea level pressure every 3 h) and the number of ensemble members in which DSCAPE is above 200 J kg^{-1} (a typical threshold that identifies a sting-jet precursor, while also requiring RH > 80 % at the level where the DSCAPE threshold is exceeded). Previously a threshold on the number of neighbouring grid points, or grid points within a neighbourhood, in the cloud head with significant DSCAPE has been used as an indicator of the likelihood of a sting jet (e.g. Gray et al., 2021). Instead, we adopt a simpler approach by calculating the fraction of ensemble members with one or more grid points where the DSCAPE threshold is reached to determine the ensemble probability of the presence of mesoscale convective instability. The number of ensemble members with this precursor, and hence the probability of a sting jet, increases as the reanalysis improves. Over half (55 %) of the ensemble members in the improved reanalysis show some precursors during the storm at locations in the cloud head to the north-west of the track of the storm. These precursors appear several hours before the strongest winds are observed to the south of the low pressure, as typically seen in such storms. In 20CRv3, only 30 % of ensemble members show such a precursor. For

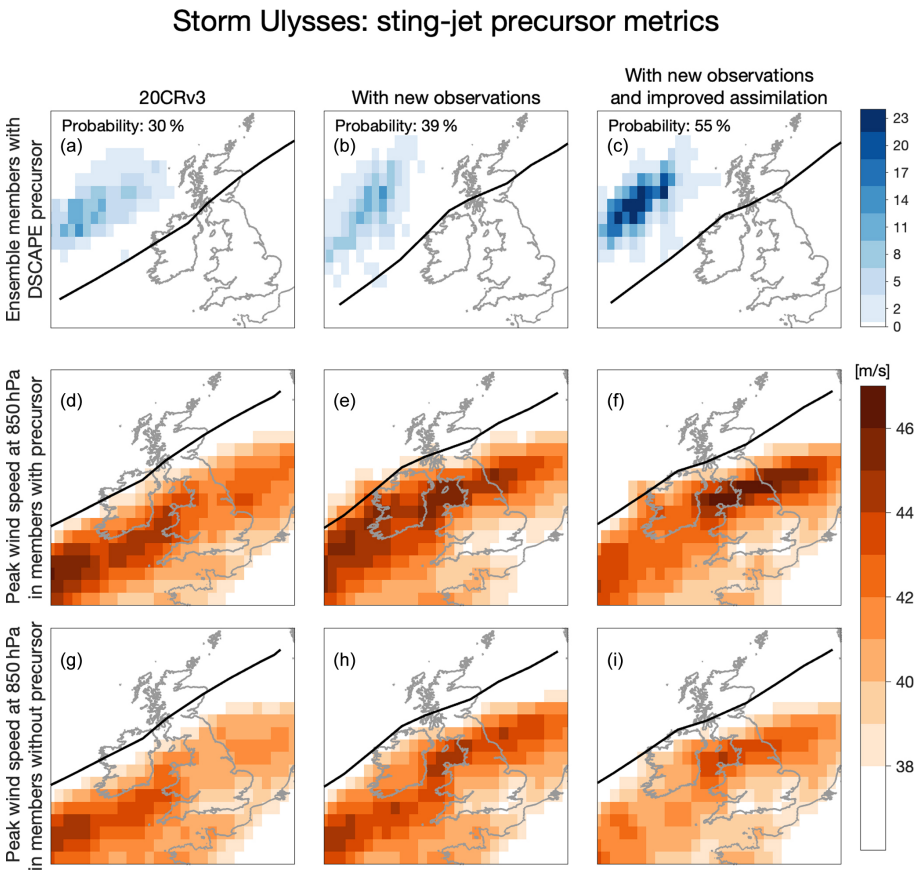

Figure 6. DSCAPE metric during Storm Ulysses, showing the number of ensemble members with a sting-jet precursor at that location at any point during the storm **(a–c)**. The number of members (out of 80) increases as the reanalysis is improved, and the probability values signify the fraction of ensemble members with at least one grid point where the precursor is present at any time during the storm. The bottom two rows **(d–i)** show the maximum wind speed at 850 hPa in two sub-ensembles, using members with **(d–f)** and without **(g–i)** a sting-jet precursor. Peak winds $>45\,\mathrm{m\,s^{-1}}$ (100 mph) are seen in the improved reanalysis at locations where known significant damage occurred over land, but only in the members with a precursor. The track of the minimum sea level pressure is shown by the black line in each panel, which varies slightly between reanalyses.

the DSCAPE precursor likelihood, there is a clear difference between the experiments that only differ due to the assimilation scheme changes (39 % vs 55 %). We suggest that this may be because DSCAPE is a threshold-based binary metric, meaning that the reduction in ensemble spread has a larger effect.

High DSCAPE values have previously been found to be an indicator for strong surface winds in such datasets (Hart et al., 2017; Clark and Gray, 2018). This can be examined for Storm Ulysses by splitting the reanalysis ensembles into two. The maximum wind speeds at 850 hPa in the ensemble members with a sting-jet precursor are clearly larger than in those members without a sting-jet precursor in each of the reanalyses (Fig. 6, bottom two rows). Ensemble mean wind speeds of $>45\,\mathrm{m\,s^{-1}}$ (100 mph) are simulated across members of the improved reanalysis with a precursor and occur at locations where known significant damage occurred (Fig. 1). The members without a precursor have significantly lower wind speeds, and the ensemble mean does not reach $45\,\mathrm{m\,s^{-1}}$.

Overall, the improved reanalysis appears to be a credible representation of the storm. Simple comparisons have highlighted the value of both quantitative observations and qualitative information to evaluate the plausibility of the reanalyses. The photographic and written evidence is notable for enabling an evaluation of the reconstructions for aspects of the storm for which detailed instrumental measurements are not available.

4 Assessing rainfall variations during Storm Ulysses

An important test of the credibility of the Storm Ulysses reconstructions is to compare with additional independent data that are not assimilated into the reanalyses. North-west Europe, and the UK and Ireland in particular, have detailed instrumental observations that can be used for such an evalua-

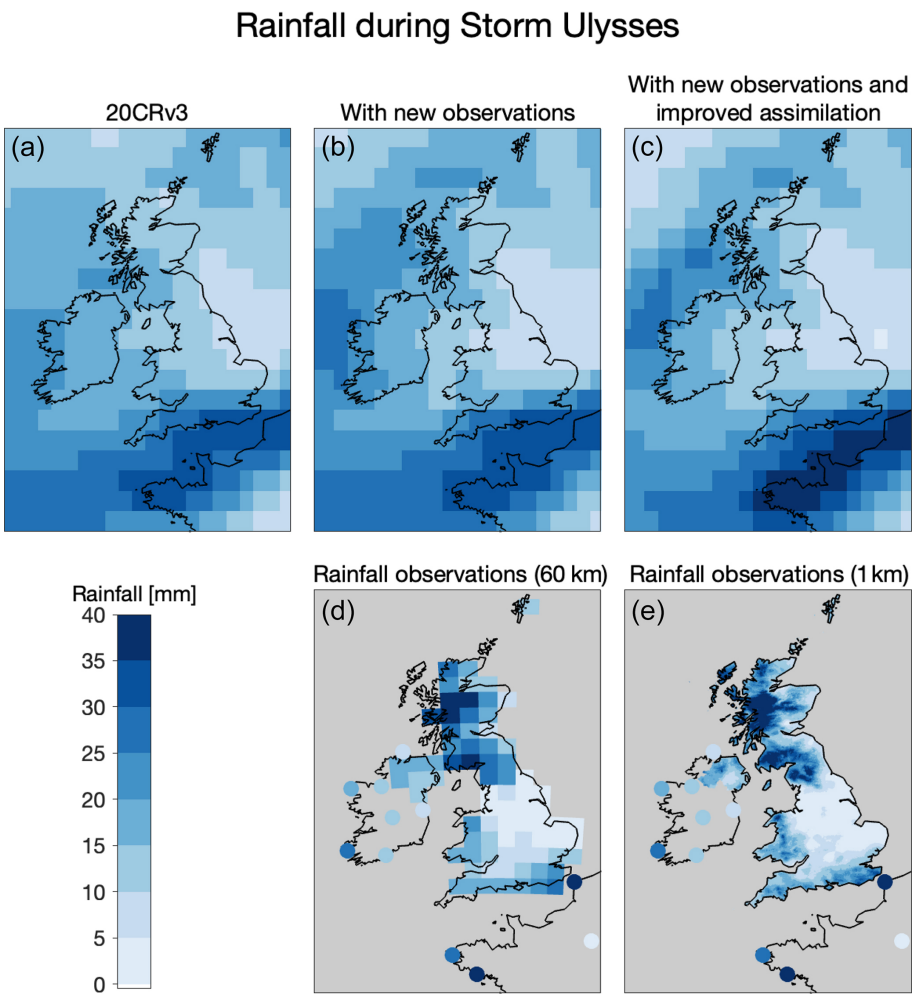

Figure 7. Assessing rainfall variations during the storm. Rainfall in millimetres for the 48 h period between 09:00 UTC on 26 February and 09:00 UTC on 28 February 1903 in the three versions of the reanalysis **(a–c)**, compared with gridded rainfall reconstructions for the UK, interpolated from in situ observations (HadUK-Grid, on two different spatial scales; **d, e**). The 60 km dataset roughly matches the spatial resolution of 20CRv3. Other available individual station rainfall observations for Ireland and France are shown with filled circles (see Appendix A). None of the rainfall data are assimilated in the reanalyses, and so the observational data and reanalysis output are independent.

tion. In this case, both daily and even sub-daily rainfall observations can be independently compared with rainfall estimates generated within the reanalyses (see Appendix A for details). Figure S6 shows how precipitation varies during the storm in all the original 20CRv3 and two experimental reanalyses.

Figure 7 compares the rainfall totals derived from interpolated in situ observations (Hollis et al., 2019) and from the reanalysis for the 2 days of the storm and shows good agreement in the broad spatial patterns between 20CRv3 and the observations. None of the reanalysis experiments capture the large rainfall over the mountainous regions of the UK, presumably due to the coarse resolution of the reanalysis compared to the small spatial scales of the steep orography. However, the reanalysis with additional observations and improved assimilation has drier conditions over the central UK and wetter conditions over northern France than the original version, in even better agreement with the independent rainfall observations. The average ensemble spread in rainfall totals is reduced by 25 % across the domain in this improved reanalysis compared to 20CRv3 (not shown).

It is also possible to compare higher frequency rainfall data. During this period the UK and Ireland had five meteorological observatories that were taking hourly observations of rainfall, which can also be compared with the reanalysis. The blue bars in Fig. 8 show these observations, integrated over the same 3 h periods as the reanalysis. The grey and red lines show 20CRv3 and the improved reanalysis, highlighting that even on a sub-daily timescale there is reasonable agreement with the observations, both for the timing of the rainfall and the amounts. The exception is Fort William, which is in the most mountainous region of the UK, where the reanalysis is

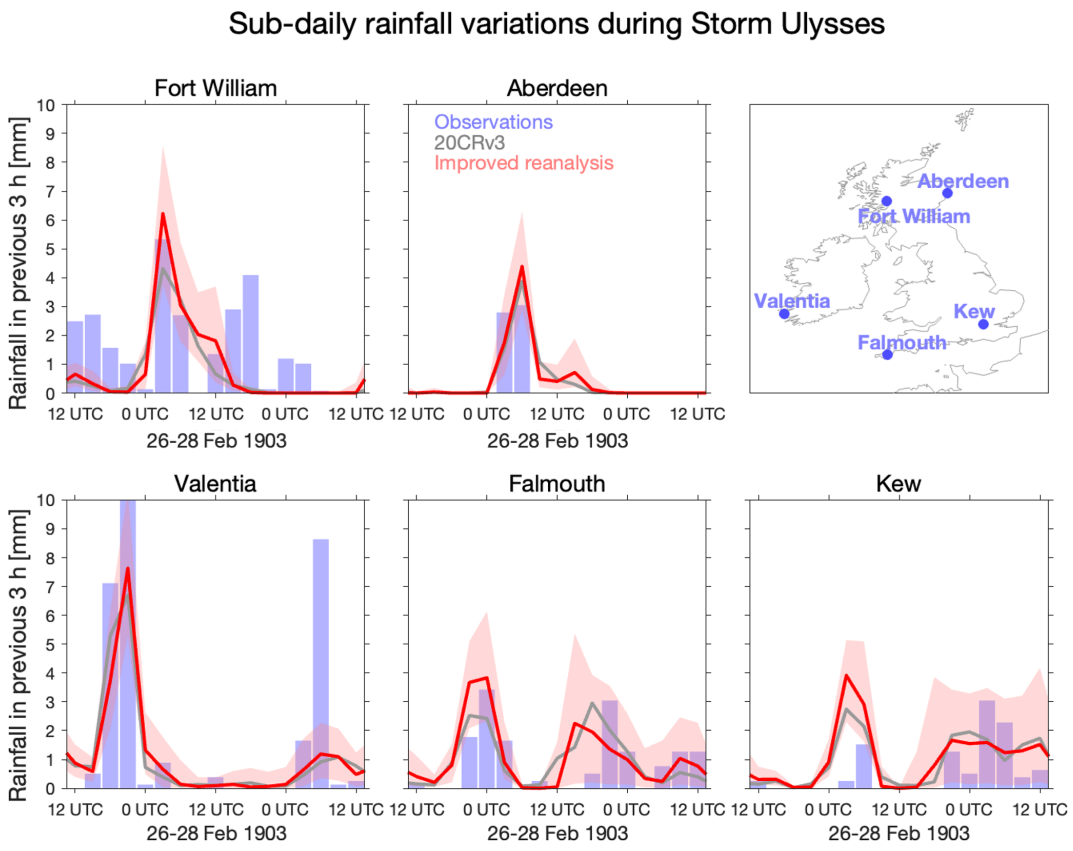

Figure 8. Assessing high-frequency rainfall variations. Rainfall every 3 h during Storm Ulysses from 20CRv3 (ensemble median in grey) and the reanalysis with new observations and improved assimilation (ensemble median and 16 %–84 % range in red) between 12:00 UTC on 26 and 12:00 UTC on 28 February 1903 for five locations across the British and Irish Isles. The 16 %–84 % range is roughly equivalent to showing the ensemble standard deviation but is more appropriate for a non-normally distributed variable, such as high-frequency rainfall amounts. These locations have hourly rainfall observations available which have been integrated over the same 3 h periods (blue bars) as produced by the reanalysis. None of the rainfall data are assimilated in the reanalyses, and so the observational data and reanalysis output are independent.

too low-resolution to represent the variable orography. There is more rain in the observations than the reanalysis for this location (and also for Valentia to a lesser extent); however, the timing of peak rainfall amounts is well represented. Although the ensemble mean rainfall does not change notably between 20CRv3 and the improved reanalysis, the average ensemble spread across each location and 3 h period is reduced by 24 % in the improved version (not shown).

Slivinski et al (2021) demonstrated that interannual variability in rainfall is well represented in 20CRv3. This study extends those comparisons to an individual event. It provides evidence that 20CRv3 produces plausible estimates of rainfall during this extreme storm over most parts of the UK and Ireland and that this representation is further improved in the experiments performed. Further spatial downscaling would likely be required to reliably represent rainfall variations in mountainous regions.

5 Assessing the coastal storm surge

Windstorms also produce coastal storm surges. We use the reanalysis winds and atmospheric pressure fields to drive the UK Continental Shelf 3 (CS3) model, which is a hydrodynamic numerical ocean model of the entire north-west European continental shelf with a resolution of approximately 12 km. These simulations produce estimates of storm surge heights around the British and Irish Isles (see Appendix C for more details). This type of approach has previously been adopted globally (e.g. Muis et al., 2016; Tadesse and Wahl, 2021) and regionally (e.g. Haigh et al., 2014) using different reanalyses, including for specific extreme events (Choi et al., 2018; Meyer et al., 2022) with mixed success.

Figure 9a, b show maps of the height of the simulated maximum storm surge during Storm Ulysses using 20CRv3 and the reanalysis with additional observations and improved assimilation. More precisely, the non-tidal residual is shown, which indicates the difference between the water height

driven by the meteorological forcing after the astronomical tidal component has been removed. The improved reanalysis fields drive a larger storm surge on the north-west coast of England and around the Irish coasts and a smaller storm surge in other locations (Fig. 9c), with considerably reduced uncertainty. This is consistent with the pattern of stronger winds shown in Fig. 3.

Figure 9d, e also compare the simulated storm surge with the same metric derived from high-frequency tide gauge observations for two sites near Liverpool (within 15 km of each other). The tide gauge data have recently been digitized from paper archives in another citizen science project (see Appendix A). These two sites are close to the region of strongest winds during Storm Ulysses and the peak simulated storm surge. The improved reanalysis fields drive a larger surge (by 0.35 m) than the original reanalysis, in better agreement with the observations. However, the observed storm surge (around 2.5 m) is still slightly larger than the simulations, hinting that the reanalysis might still be underestimating the wind strength, and could be further improved through, for example, improved resolution or addition of more pressure observations to better constrain the wind fields. Alternatively, the coastal surge model could be slightly underestimating the local response to the winds, and increased spatial resolution may help resolve this. There is also more variability in the observations than the reanalysis, but this is not unexpected due to complex local tidal features in this region. This verification against independent data is encouraging for the credibility of the reanalysis and for using the reanalysis to estimate storm surges at other locations and time periods where tide gauge data are not available.

There are no reports of flooding in Liverpool during this storm because the maximum surge occurred during neap tides and not at high tide. As a result, the skew surge (peak observed height minus peak predicted height during the tidal cycle) was around 1.2 m. Overall, the storm surge is one of the 10 largest observed events between 1857–1903 (the period of data recently rescued) and is larger than any observed event in the available modern Liverpool tidal records (1991–2021). This suggests that this storm surge was a roughly once-per-decade event and would likely have caused flooding if the peak surge had occurred during high spring tides. Improved knowledge of such events will inform risk estimates of coastal flooding, especially as sea levels in this region have increased by around 0.2 m since this storm.

6 Benefits of rescuing observations for improving reanalyses and estimating risk

The recovery of historical weather observations from paper archives is informing our knowledge and understanding of the risks from extreme weather. Any approach to estimating risk by identifying plausible worst-case outcomes (Thompson et al., 2017) or developing storylines of severe weather events (Shepherd et al., 2018) would benefit from longer sampling of real-world behaviour and improved historical knowledge (Woo and Johnson, 2018; Pinto et al., 2019).

As an example, during the modern period (1950–2015) the maximum instantaneous wind speed (based on 3-hourly data) experienced at a grid point in northern England (54.4° N, 2.0° W) was 21.2 m s^{-1} in 20CRv3. During Storm Ulysses, this location was in the region of peak winds over land and experienced a wind speed of 22.0 m s^{-1}. The modern period data suggest that the unprecedented winds experienced during Storm Ulysses would be rarer than once in 100 years for that location. Having a credible reconstruction for such a rare event provides valuable information on plausible risks and potential damage. Note that these quoted wind speeds will be substantially less than sustained wind speeds or gusts, motivating future downscaling of this storm to better quantify the extreme nature of the winds.

This end-to-end case study demonstrates how combining modern weather forecasting and data assimilation techniques with measurements of surface pressure taken over 100 years ago can credibly reconstruct details of one of the most severe European windstorms in the instrumental period, providing support for the capability of this reanalysis approach to reconstruct extreme events in general. This study is also a clear example of how the addition of newly rescued meteorological and related climatological observations can directly improve reanalyses of such extreme events and provide independent validation of the reconstruction.

Comprehensive rescue of existing weather observations, stored on paper in various archives, would allow much more precise and accurate reconstructions of many other similar events across Europe (and any region with enough data available), including other extreme weather events such as heatwaves and floods (Brönnimann et al., 2018). Better knowledge and understanding of these historical extreme events would allow observed trends in extreme events to be put into a longer-term context and help identify where present-day and future risks have been underestimated because such extreme events may not have yet been observed during the modern period.

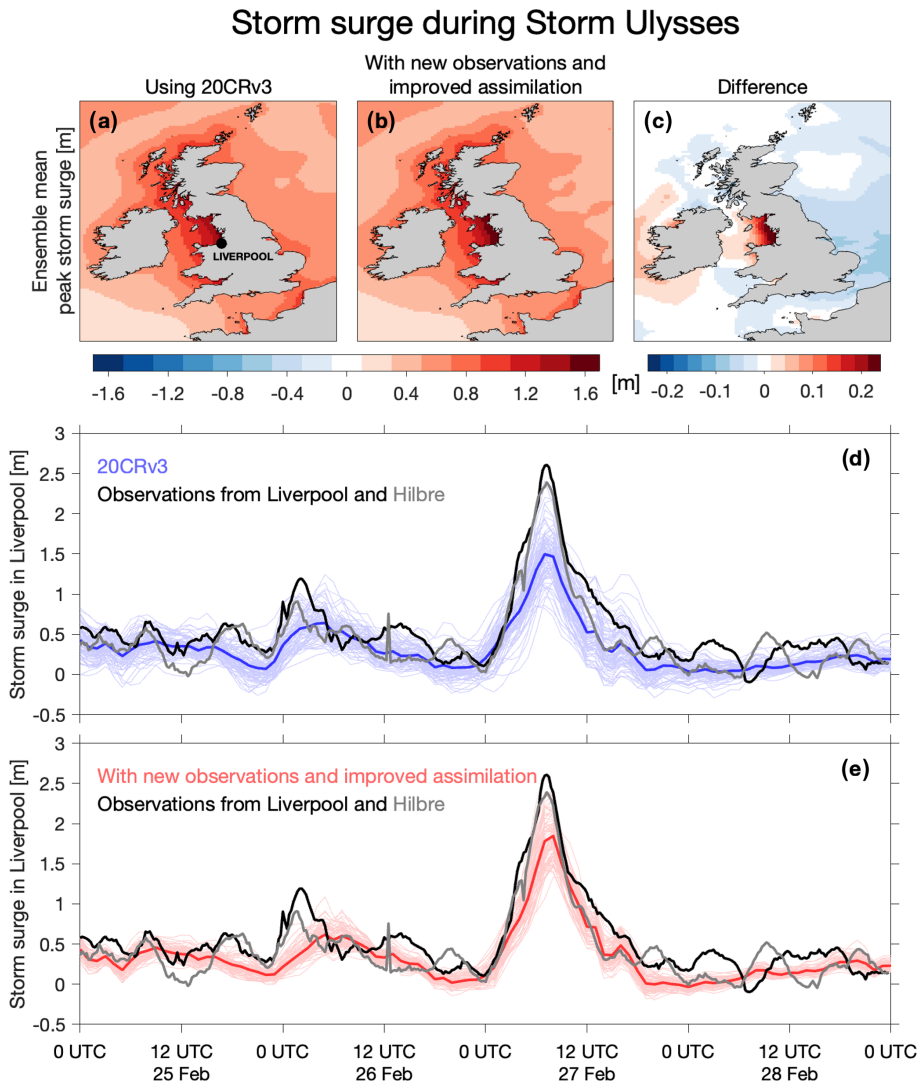

Figure 9. Storm surge simulations for Storm Ulysses. Top: maps of the ensemble mean of the maximum storm surge during Storm Ulysses, for two versions of the reanalysis **(a, b)**, and the difference **(c)**. Bottom: simulated storm surge height at Liverpool for each ensemble member (thin lines) and ensemble mean (thick lines) using two versions of the reanalysis **(d, e)**, compared with the newly rescued tide gauge observations for Liverpool Docks and Hilbre Island, for 25–28 February 1903.

Appendix A: Additional observations

The additional pressure and rainfall observations used come from a range of sources. The largest component comes from the Weather Rescue citizen science project (Craig and Hawkins, 2020), which digitized 11 years of the UK Met Office *Daily Weather Report*s (1900–1910; e.g. Fig. A1). These reports include twice-daily surface pressure observations and daily rainfall amounts from 57 locations across the UK, Ireland, and north-west Europe. The new pressure data also include hourly observations taken on the summit of Ben Nevis in Scotland and in the nearby town of Fort William, which were also transcribed by volunteers (Hawkins et al., 2019). The final source of data is 19 *Stations of the Second Order* and 11 locations with *Climatological Returns* in the UK Met Office digital archives (NMLA, 2023), which were additionally transcribed for the period around Storm Ulysses, with two pressure observations per day. Note that in 1903, when Storm Ulysses occurred, Ireland was not an independent country and was part of the UK. This can be seen in Fig. A1, which describes observations from locations in present-day Ireland as being part of the "British Islands". We have used present-day national boundaries when describing locations in the text and use "British and Irish Isles" as a more inclusive term where appropriate. Pressure observations from tens of more locations are potentially available for the British and Irish Isles (and other countries) for this event, including hourly data from several sites, but these have not yet been digitized from the paper archives.

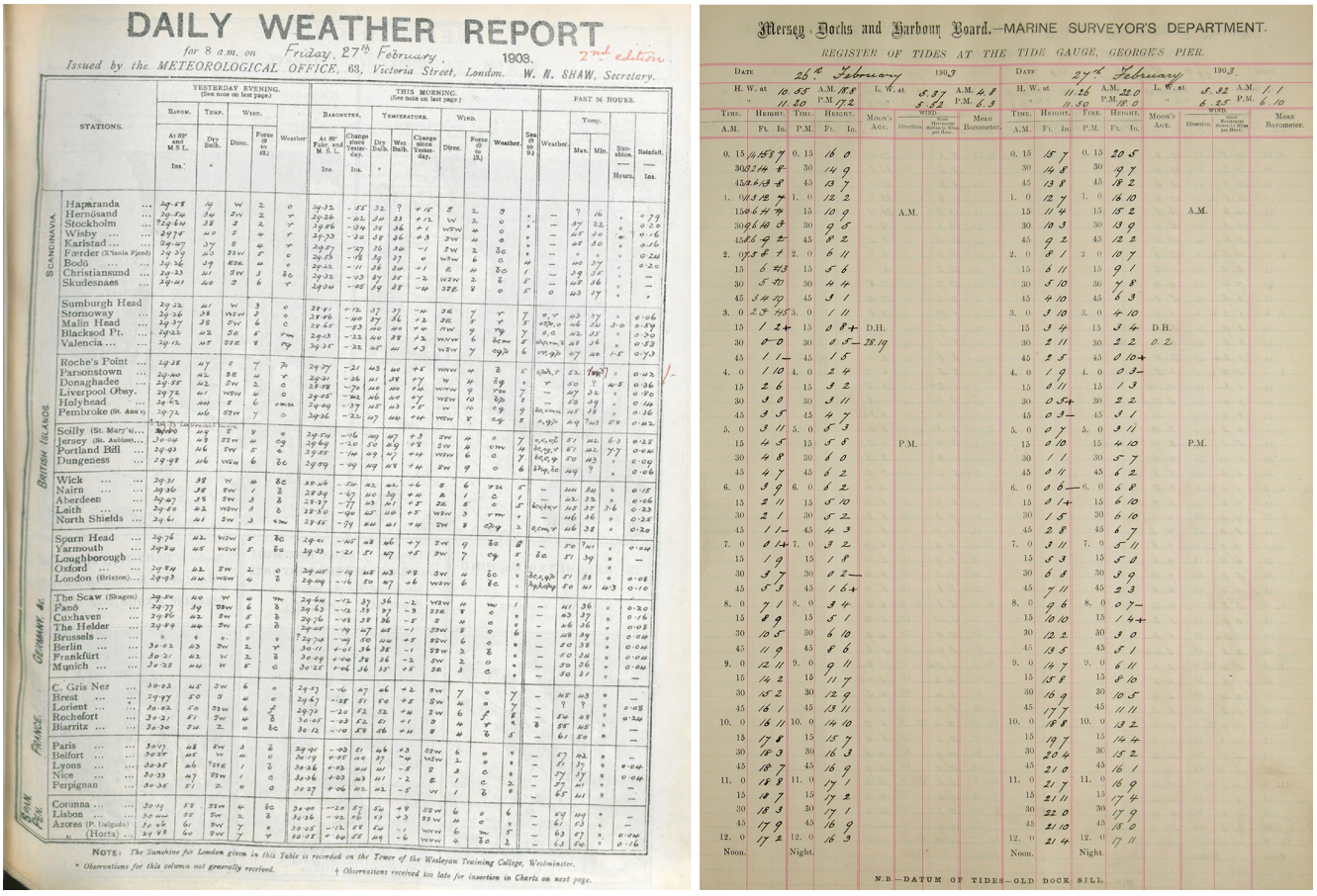

Figure A1. The UK *Daily Weather Report* observations page for 27 February 1903 (left; from NMLA, 2023) and tide gauge measurements from Liverpool Docks for 26–27 February 1903 (right; supplied by authors).

In 20CRv3, the land stations are all assigned an uncertainty of 1.2 hPa (surface pressure) or 1.6 hPa (sea level pressure), and ship observations are assigned an uncertainty of 2.0 hPa. This is unchanged in our experiments. The added observations mainly come from locations that were regularly inspected by the Met Office, suggesting the data are of high quality and perhaps the uncertainty assigned is too large.

The gridded daily rainfall data in Fig. 7 are from the HadUK-Grid dataset (Hollis et al., 2019), summed over the two days of Storm Ulysses. The individual station observations in Fig. 7 are taken from the *Daily Weather Report*s and the *Stations of the Second Order* books. The hourly rainfall observations in Fig. 8 were digitized manually just for this event from the *Hourly Books* for the four Met Office Observatories (Falmouth, Kew, Aberdeen, and Valentia), stored in the Met Office Archives. Hawkins et al. (2019) contains the hourly rainfall data for Fort William.

Two tide gauge series with 15 minute resolution covering the Ulysses storm were recently recovered from scanned paper records by the UK Tides citizen science project (Matthews et al., 2023). The two nearby sites are Liverpool Docks (53.4052° N, 2.9985° W) and Hilbre Island (53.3851° N, 3.2293° W). The precision of the tide gauge data is 1 inch, and the data for the period from 25–28 February 1903 are available (see Fig. A1 for an example of the scanned pages) and have been screened for quality control. The errors are hard to constrain without contemporary levelling information, but the high coherence of the Liverpool and Hilbre data suggests they are within a few tens of centimetres. The tidal data for the whole period rescued (1857–1903) have not yet been fully quality controlled, so we cannot yet be precise about how extreme the Ulysses storm surge was within the full context of the longer period, but it is certainly in the top-10 events. There are no other digitized high-frequency tide gauge records for the UK for this period, although several do exist on paper.

Appendix B: Improvements to data assimilation scheme

We used the openly available data of the 20CRv3 reanalysis for this storm (Slivinski et al., 2019b) and performed two additional experiments with the same reanalysis system. The first experiment was identical to the original except

for assimilating thousands of additional newly rescued surface pressure observations. The second experiment repeated the first experiment with a change to the data assimilation scheme.

In the 20th Century Reanalysis assimilation system, the background (prior) fields are provided from the underlying numerical weather prediction model with prescribed pentad sea surface temperatures (interpolated to daily), monthly sea ice concentration, and monthly radiative forcing. Surface pressure observations are assimilated with an ensemble Kalman filter (Whitaker and Hamill, 2002) to generate the reanalysis. One common issue with ensemble filters is so-called "ensemble collapse", in which the ensemble spread can collapse to 0 in sequential assimilation cycles (Anderson and Anderson, 1999; Whitaker and Hamill, 2002). Generally, ad hoc inflation methods are needed to increase the ensemble spread. In the 20CRv3 system, the inflation method is relaxation-to-prior-spread (RTPS; Whitaker and Hamill, 2012; Slivinski et al., 2019a), where the analysis ensemble spread is "relaxed" back to the prior ensemble spread by a temporally and spatially varying parameter λ. This parameter depends on the observation network density at that time and location, as well as a hyperparameter p_{relax}:

$$\lambda_{\text{inf}(x,y,t)} = \frac{p_{\text{relax}}\left(\sigma_{\text{b}(x,y,t)} - \sigma_{\text{a}(x,y,t)}\right)}{\sigma_{\text{a}(x,y,t)}} + 1, \tag{B1}$$

where $\left(\sigma_{\text{b}(x,y,t)}\right)$ is the standard deviation of the background ensemble and $\left(\sigma_{\text{a}(x,y,t)}\right)$ is the standard deviation of the analysis ensemble before inflation. The hyperparameter p_{relax} can vary from 0 to 1 and determines the sensitivity of λ to the observation density: the higher p_{relax} is, the more the analysis ensemble can be relaxed back to the prior ensemble. Essentially, the inflation is increased with high observation density and decreased with low observation density (see Fig. 3 of Slivinski et al., 2019a). However, p_{relax} itself needs to be tuned; due to the computational cost, effort, and number of parameters that need to be tuned, only a few initial tests were completed, resulting in $p_{\text{relax}} = 0.9$ in the Northern Hemisphere, 0.7 in the Southern Hemisphere, and a linear transition between the two values in the tropics. However, results from these experiments (see below) suggest that further tuning could be beneficial, since the addition of many new observations did not have as large of an impact on the analysis mean or spread as expected. Therefore, the subsequent experiment with "improved assimilation" was run with the new observations, as well as decreasing p_{relax} to 0.5 everywhere. This ultimately prevented the analysis ensemble spread from being relaxed back to the prior ensemble spread as much, effectively allowing the observations to have a stronger impact, as shown in Fig. 3.

We consider whether the changes to RTPS improve the reanalysis by comparing with independent observations. Ideally the ensemble spread of the reanalysis should be "reliable"; i.e. it appropriately represents the uncertainty given the available observations. This reliability can be tested by comparing with pressure observations that are not assimilated. For the period of Storm Ulysses, we have pressure observations from five additional locations that were withheld from the reanalysis (see right hand panel of Fig. B1). These locations were chosen to cover a wide range of locations across the UK and Ireland. The data were digitized manually for this event, except for Durham which recently became available for a longer period (Burt, 2023).

We have extracted the reanalysis ensemble mean and ensemble spread from the locations and times of these independent observations and performed a mean bias correction. A bias correction is also included within the reanalysis assimilation cycle, so this approach roughly mimics the reanalysis approach. This process allows a root mean square error (RMSE) between the observations and reanalysis and a mean ensemble spread for the times of the observations to be calculated for each location for February 1903 (Fig. B1). A reliable ensemble would show similar values for RMSE and ensemble spread.

For the original reanalysis, the ensemble spread is much larger than the RMSE for each location (blue symbols are to the right of the 1 : 1 dashed line), suggesting that the ensemble is under-confident (or over-dispersive) in the atmospheric circulation patterns. In the experiments with additional observations (red symbols), the ensemble spread and RMSE have both been reduced as expected, but the ensemble remains under-confident. In the experiment where the RTPS parameter is reduced (yellow symbols), the RMSE and ensemble spread now lie closer to the 1 : 1 line, indicating a more reliable ensemble.

Although these results are encouraging that a smaller RTPS parameter is producing a more reliable ensemble, if also accounting for observational uncertainty, the ensemble spread should be slightly smaller than the RMSE; i.e. the plotted points should fall to the left of the 1 : 1 line. It is therefore likely that the ensemble is still slightly under-confident even with the reduced RTPS.

Further experiments for a much longer period would be required to rigorously assess the ensemble, likely including examining other parameters such as the assumed uncertainty in each observation.

We also consider the modern period with a similar set of tests. There is a possibility that the RTPS parameter in the original 20CRv3 might also need to be reduced for the modern period, which would make the comparison of wind ranks in Fig. 4 potentially unfair. However, Fig. B2 highlights that for 2 example years (1953 and 2003, i.e. 50 and 100 years after Storm Ulysses) the original 20CRv3 is roughly reliable when compared to unassimilated pressure data from one location (Reading, UK, 51.5° N, 1.0° W), which is now available for each year in the comparison. The Reading data were assimilated in the two reanalysis experiments, so they are not included in Fig. B1.

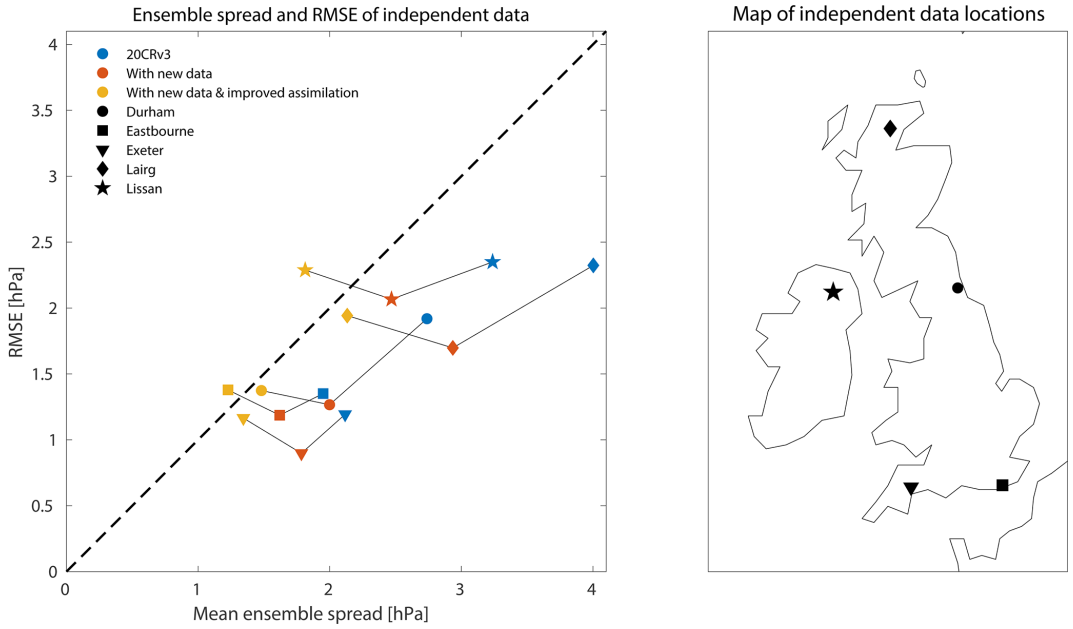

Figure B1. Ensemble reliability. Mean ensemble spread of the reanalysis and RMSE of the reanalysis compared to unassimilated independent observations for five locations (shapes) during February 1903. Different versions of the reanalysis are shown with different colours. None of these observations are assimilated in any version of the reanalysis. Adding the additional observations reduces both the ensemble spread and the RMSE (moving from blue to red symbols), and the improved assimilation has made the reanalysis more reliable (moving from red to yellow symbols, RTPS reduced from 0.9 to 0.5), with a small increase in RMSE. The dashed line represents "perfect" ensemble reliability, when RMSE and ensemble spread are equal.

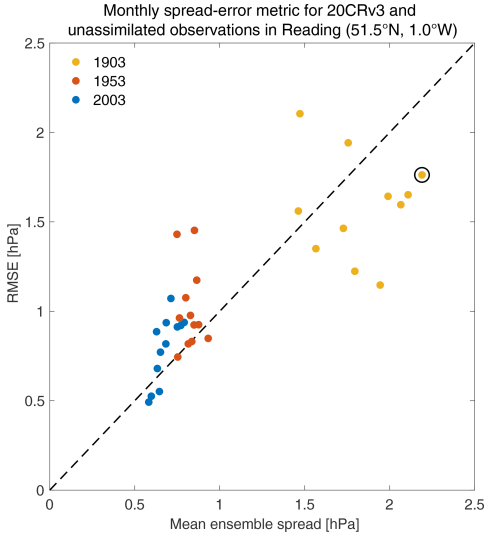

Figure B2. Ensemble reliability across years. Mean ensemble spread of the reanalysis and RMSE of the reanalysis (both in hPa) compared to unassimilated independent observations from one location where data are available for an extended period (one dot per month from three different years). For the modern period (using 1953 and 2003 examples), the ensemble is roughly reliable, so there is no evidence that the RTPS parameter needs to be altered for that time period. The open circle indicates February 1903.

Appendix C: Storm surge model

To model the storm surge and tide we used the UK Continental Shelf 3 (CS3) model. This model was developed at the National Oceanography Centre in the UK and is based on a finite-difference discretization of the fully nonlinear, depth-averaged Navier–Stokes equations (Proctor and Flather, 1983; Flather et al., 1991). CS3 was extensively used for operational forecasting by the Met Office from 1991 to 2006 and is one of the most validated operational storm surge forecasting models in the world. The model covers the entire north-west European continental shelf on a $1/9°$ latitude by $1/6°$ longitude grid, giving a resolution of approximately 12 km. We applied tidal forcing at the open lateral boundaries using the 15 largest constituents derived from a harmonic analysis of a larger-area ocean model. Wind stress was calculated using a quadratic stress law, where the drag coefficient is derived from observations using the parameterization of Smith and Banke (1975). We ran the hydrodynamic model 160 times from the start of 25 February to the end of 28 February 1903, simulating total water level (e.g. tides plus storm surges) using wind (u and v components) and atmospheric pressure fields from each of the original 80 reanalysis ensemble members and then for the 80 improved ensemble members. The reanalysis produces 3-hourly winds and pressure fields, which were interpolated to hourly for the simulations. We also ran an additional tide-only simulation. We sub-

tract the predicted astronomical tidal heights from each of the total water level simulations to estimate the storm surge components. We save model results for each model grid cell every 10 min and calculate maps of the maximum storm surge over the event for each original and improved ensemble member. We also extracted the storm surge time series at the nearest point to the Liverpool tide gauges.

Data availability. The complete 20th Century Reanalysis dataset is openly available (https://portal.nersc.gov/project/20C_Reanalysis/, last access: 19 April 2023). For the short period around Storm Ulysses, the data from 20CRv3 and the reanalysis experiments performed are available here: https://doi.org/10.5281/zenodo.7838019 (Hawkins, 2023a). The additional observations used are available here: https://doi.org/10.5281/Zenodo.7765124 (Hawkins, 2023b).

Supplement. The supplement related to this article is available online at: https://doi.org/10.5194/nhess-23-1465-2023-supplement.

Author contributions. EH conceived and led the project and analysis, with contributions from all co-authors. PB, GPC, HH, CM, and LS assisted with the design and running of the reanalysis experiments. KK and IDH carried out the storm surge experiments. SNB and SB assisted with data recovery, and JW provided the tidal data. SLG and OMA provided guidance on the sting-jet analysis. EH prepared the paper with contributions from all co-authors.

Competing interests. The contact author has declared that none of the authors has any competing interests.

Disclaimer. Publisher's note: Copernicus Publications remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Acknowledgements. We thank the observers who took the original weather observations over a century ago, those who collated the data so carefully at the time, and the archivists who have preserved the paper material ever since. We also gratefully acknowledge the thousands of citizen scientist volunteers who gave their spare time to help digitize and recover the weather and tide gauge observations used, and Zooniverse for providing the citizen science platform. We also thank Andy Matthews and Elizabeth Bradshaw for the retrieval of Liverpool and Hilbre tide gauge data. Support for the 20th Century Reanalysis Project version 3 dataset is provided by the U.S. Department of Energy, Office of Science Biological and Environmental Research (BER), by the National Oceanic and Atmospheric Administration Climate Program Office, and by the NOAA Physical Sciences Laboratory. This research used resources of the National Energy Research Scientific Computing Center (NERSC), a U.S. Department of Energy Office of Science User Facility located at Lawrence Berkeley National Laboratory, operated under Contract No. DE-AC02-05CH11231. Ed Hawkins was supported by the National Centre for Atmospheric Science. Ed Hawkins and Andrew P. Schurer were supported by the NERC GloSAT project. Philip Brohan was supported by the Met Office Hadley Centre Climate Programme funded by BEIS. Gilbert P. Compo, Laura Slivinski, and Chesley McColl were supported in part by the NOAA cooperative agreement NA22OAR4320151, by the NOAA Climate Program Office and NOAA Physical Sciences Laboratory.

Financial support. This research has been supported by the Natural Environment Research Council (grant no. NE/S015574/1).

Review statement. This paper was edited by Joaquim G. Pinto and reviewed by three anonymous referees.

References

Anderson, J. L. and Anderson, S. L.: A Monte Carlo implementation of the nonlinear filtering problem to produce ensemble assimilations and forecasts, Mon. Weather Rev., 127, 2741–2758, 1999.

Board of Trade: Furness Railway, https://www.railwaysarchive.co.uk/documents/BoT_LevenViaduct1903.pdf (last access: 20 September 2022), 1903.

Brönnimann, S., Martius, O., Franke, J., Stickler, A., and Auchmann, R.: Historical weather extremes in the "Twentieth Century Reanalysis", edited by: Brönnimann, S. and Martius, O., Weather extremes during the past 140 years, Reihe G Grundlagenforschung: Vol. G89 (pp. 7–17), Bern: Geographica Bernensia, https://doi.org/10.4480/GB2013.G89.01, 2013.

Brönnimann, S., Brugnara, Y., Allan, R. J., Brunet, M., Compo, G. P., Crouthamel, R. I., Jones, P. D., Jourdain, S., Luterbacher, J., Siegmund, P., Valente, M. A., and Wilkinson, C. W.: A roadmap to climate data rescue services, Geosci. Data J., 5, 28, https://doi.org/10.1002/gdj3.56, 2018.

Browning, K. A.: The sting at the end of the tail: damaging winds associated with extratropical cyclones, Q. J. R. Meteorol. Soc., 130, 375–399, https://doi.org/10.1002/wea.3889, 2004.

Buchana, P. and McSharry, P. E.: Windstorm risk assessment for offshore wind farms in the North Sea, Wind Energy, 22, 1219, https://doi.org/10.1002/we.2351, 2019.

Burt, S. D.: A twice-daily barometric pressure record from Durham Observatory in north-east England, 1843–1960, Geosci. Data J., 10, 3–17, https://doi.org/10.1002/gdj3.135, 2023.

Choi, B. H., Kim, K. O., Yuk, J. H., and Lee, H. S.: Simulation of the 1953 storm surge in the North Sea, Ocean Dynam., 68, 1759, https://doi.org/10.1007/s10236-018-1223-z, 2018.

Clark, P. A. and Gray, S. L.: Sting jets in extratropical cyclones: a review, Q. J. R. Meteorol Soc, 144, 943–969, https://doi.org/10.1002/qj.3267, 2018.

Whitaker, J. S., Sardeshmukh, P. D., Matsui, N., Allan, R. J., Yin, X., Gleason, B. E., Vose, R. S., Rutledge, G., Bessemoulin, P., Brönnimann, S., Brunet, M., Crouthamel, R. I., Grant, A. N., Groisman, P. Y., Jones, P. D., Kruk, M. C., Kruger, A. C., Marshall, G. J., Maugeri, M., Mok, H. Y., Nordli, Ø., Ross, T. F., Trigo, R. M., Wang, X. L., Woodruff, S. D., and Worley, S. J.:

The Twentieth Century Reanalysis Project. Q. J. R. Meteorol. Soc., 137, 1–28, https://doi.org/10.1002/qj.776, 2011.

Compo, G. P., Slivinski, L. C., Whitaker, J. S., Sardeshmukh, P. D., McColl, C., Brohan, P., Allan, R., Yin, X., Vose, R., Spencer, L. J., Ashcroft, L., Bronnimann, S., Brunet, M., Camuffo, D., Cornes, R., Cram, T. A., Crouthamel, R., Dominguez-Castro, F., Freeman, J. E., Gergis, J., Giese, B. S., Hawkins, E., Jones, P. D., Jourdain, S., Kaplan, A., Kennedy, J., Kubota, H., Blancq, F. L., Lee, T., Lorrey, A., Luterbacher, J., Maugeri, M., Mock, C. J., Moore, K., Przybylak, R., Pudmenzky, C., Reason, C., Slonosky, V. C., Tinz, B., Titchner, H., Trewin, B., Valente, M. A., Wang, X. L., Wilkinson, C., Wood, K., and Wyszynski, P.: The International Surface Pressure Databank version 4, Research Data Archive at the National Center for Atmospheric Research, Computational and Information Systems Laboratory, https://doi.org/10.5065/9EYR-TY90, 2019.

Craig, P. and Hawkins, E.: Digitizing observations from the Met Office Daily Weather Reports for 1900–1910 using citizen scientist volunteers, Geosci. Data J., 7, 116, https://doi.org/10.1002/gdj3.93, 2020.

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

Flather, R., Proctor, R., and Wolf, J.: Oceanographic forecast models, pp. 15–30, in: Computer modelling in the environmental sciences, edited by: Farmer, D. G. and Rycroft, M. J. (Based on the proceedings of a conference organised by the Natural Environment Research Council in association with the Environmental Mathematics Group of the Institute of Mathematics and its Applications, British Geological Survey, Keyworth, April 1990), Oxford: Clarendon Press, 379 pp., ISBN 0198533942, 1991.

Gray, S. L., Martínez-Alvarado, O., Baker, L. H., and Clark, P. A.: Conditional symmetric instability in sting jet storms, Q. J. R. Meteorol. Soc., 137, 1482–1500, https://doi.org/10.1002/qj.859, 2011.

Gray, S. L., Martínez-Alvarado, O., Ackerley, D., and Suri, D.: Development of a prototype real-time sting jet precursor tool for forecasters, Weather, 76, 369–373, https://doi.org/10.1002/wea.3889, 2021.

Haigh, I. D., Wijeratne, E. M. S., MacPherson, L. R., Pattiaratchi, C. B., Mason, M. S., Crompton, R. P., and George, S.: Estimating present day extreme total water level exceedance probabilities around the coastline of Australia: tides, extra-tropical storm surges and mean sea level, Clim. Dynam., 42, 121, https://doi.org/10.1007/s00382-012-1652-1, 2014.

Hart, N. C. G., Gray, S. L., and Clark, P. A.: Sting jet windstorms over the North Atlantic: climatology and contribution to extreme wind risk, J. Clim., 30, 5455–5471, https://doi.org/10.1175/JCLI-D-16-0791.1, 2017.

Hawkins, E.: Ulysses Storm Data v1.0, Zenodo [data set], https://doi.org/10.5281/zenodo.7838019, 2023a.

Hawkins, E.: Weather Rescue Data v2.0 (v2.0), Zenodo [data set], https://doi.org/10.5281/zenodo.7765124, 2023b.

Hawkins, E., Burt, S., Brohan, P., Lockwood, M., Richardson, H., Roy, M., and Thomas, S.: Hourly weather observations from the Scottish Highlands (1883–1904) rescued by volunteer citizen scientists, Geosci. Data J., 6, 160, https://doi.org/10.1002/gdj3.79, 2019.

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., Simmons, A., Soci, C., Abdalla, S., Abellan, X., Balsamo, G., Bechtold, P., Biavati, G., Bidlot, J., Bonavita, M., De Chiara, G., Dahlgren, P., Dee, D., Diamantakis, M., Dragani, R., Flemming, J., Forbes, R., Fuentes, M., Geer, A., Haimberger, L., Healy, S., Hogan, R. J., Hólm, E., Janisková, M., Keeley, S., Laloyaux, P., Lopez, P., Lupu, C., Radnoti, G., de Rosnay, P., Rozum, I., Vamborg, F., Villaume, S., and Thépaut, J.-N.: The ERA5 global reanalysis, Q. J. R. Meteorol. Soc., 146, 1999, https://doi.org/10.1002/qj.3803, 2020.

Hollis, D., McCarthy, M. P., Kendon, M., Legg, T., and Simpson, I.: HadUK-Grid – A new UK dataset of gridded climate observations, Geosci. Data J., 6, 151, https://doi.org/10.1002/gdj3.78, 2019.

Klawa, M. and Ulbrich, U.: A model for the estimation of storm losses and the identification of severe winter storms in Germany, Nat. Hazards Earth Syst. Sci., 3, 725–732, https://doi.org/10.5194/nhess-3-725-2003, 2003.

Koks, E. E. and Haer, T.: A high-resolution wind damage model for Europe, Sci. Rep., 10, 6866, https://doi.org/10.1038/s41598-020-63580-w, 2020.

Lamb, H. H.: Historic Storms of the North Sea, British Isles and Northwest Europe, Cambridge University Press, ISBN 978-0521619318, 1991.

Martínez-Alvarado, O., Gray, S. L., Catto, J. L., and Clark, P. A.: Sting jets in intense winter North-Atlantic windstorms, Environ. Res. Lett., 7, 024014, https://doi.org/10.1088/1748-9326/7/2/024014, 2012.

Matthews, A., Bradshaw, E., and Williams, J.: Using citizen science to digitise 3 million hand-written tide-gauge data entries for Liverpool and Hilbre Island, in preparation, 2023.

Met Eireann: Exceptional weather events, https://www.met.ie/cms/assets/uploads/2017/08/Feb1903_storm.pdf (last access: 14 December 2022), 2017.

Meyer, E. M. I., Weisse, R., Grabemann, I., Tinz, B., and Scholz, R.: Reconstruction of wind and surge of the 1906 storm tide at the German North Sea coast, Nat. Hazards Earth Syst. Sci., 22, 2419–2432, https://doi.org/10.5194/nhess-22-2419-2022, 2022.

Meyer, L., Hunziker, R., Weber, J., and Zürcher, A.: An Analysis of the "Great Gale of October 1881" using the Twentieth Century Reanalysis, in: Historical Weather Extremes in Reanalyses, edited by: Brönnimann, S., Geographica Bernensia, G92, 91–100, https://doi.org/10.4480/GB2017.G92.08, 2017.

Muis, S., Verlaan, M., Winsemius, H. C., Aerts, J. C. J. H. J. H., and Ward, P. J.: A global reanalysis of storm surges and extreme sea levels, Nat. Commun., 7, 11969, https://doi.org/10.1038/ncomms11969, 2016.

NMLA: Land observations, https://digital.nmla.metoffice.gov.uk/SO_7c59f237-7add-4d78-9c99-4e4210a926e1/ (last access: 6 February 2023), 2023.

Pinto, J. G., Pantillon, F., Ludwig, P., Déroche, M., Leoncini, G., Raible, C. C., Shaffrey, L. C., and Stephenson, D. B.: From Atmospheric Dynamics to Insurance Losses: An Interdisciplinary Workshop on European Storms, B. Am. Meteorol. Soc., 100, ES175–ES178, https://doi.org/10.1175/BAMS-D-19-0026.1, 2019.

Proctor, R. and Flather, R. A.: Routine storm surge forecasting using numerical models: procedures and computer programs for use on the CDC Cyber 205E at the British Meteorological Office, Institute of Oceanographic Sciences Report, No. 167. 171 pp., 1983.

Roberts, J. F., Champion, A. J., Dawkins, L. C., Hodges, K. I., Shaffrey, L. C., Stephenson, D. B., Stringer, M. A., Thornton, H. E., and Youngman, B. D.: The XWS open access catalogue of extreme European windstorms from 1979 to 2012, Nat. Hazards Earth Syst. Sci., 14, 2487–2501, https://doi.org/10.5194/nhess-14-2487-2014, 2014.

Shapiro, M. A. and Keyser, D.: Fronts, jet streams and the tropopause, in: Extratropical Cyclones, Newton CW, edited by: Holopainen, E. O., American Meteorological Society: Boston, MA, 167–191, https://doi.org/10.1007/978-1-944970-33-8_10, 1990.

Sharkey, P., Tawn, J. A., and Brown, S. J.: Modelling the spatial extent and severity of extreme European windstorms, J. R. Stat. Soc. C, 69, 223–250, https://doi.org/10.1111/rssc.12391, 2020.

Shaw, W. N.: The meteorological aspects of the storm of February 26–27, 1903, Q. J. R. Meteorol. Soc., 29, 233, https://doi.org/10.1002/qj.49702912801, 1903.

Shepherd, T. G., Boyd, E., Calel, R. A., Chapman, S. C., Dessai, S., Dima-West, I. M., Fowler, H. J., James, R., Maraun, D., Martius, O., Senior, C. A., Sobel, A. H., Stainforth, D. A., Tett, S. F. B., Trenberth, K. E., van den Hurk, B. J. J. M., Watkins, N. W., Wilby, R. L., and Zenghelis, D. A.: Storylines: an alternative approach to representing uncertainty in physical aspects of climate change, Clim. Change, 151, 555–571, https://doi.org/10.1007/s10584-018-2317-9, 2018.

Slivinski, L. C., Compo, G. P., Whitaker, J. S., Sardeshmukh, P. D., Giese, B. S., McColl, C., Allan, R., Yin, X., Vose, R., Titchner, H., Kennedy, J., Spencer, L. J., Ashcroft, L., Brönnimann, S., Brunet, M., Camuffo, D., Cornes, R., Cram, T. A., Crouthamel, R., Domínguez-Castro, F., Freeman, J. E., Gergis, J., Hawkins, E., Jones, P. D., Jourdain, S., Kaplan, A., Kubota, H., Le Blancq, F., Lee, T.-C., Lorrey, A., Luterbacher, J., Maugeri, M., Mock, C. J., Kent Moore, G. W., Przybylak, R., Pudmenzky, C., Reason, C., Slonosky, V. C., Smith, C. A., Tinz, B., Trewin, B., Valente, M. A., Wang, X. L., Wilkinson, C., Wood, K., and Wyszyński, P.: Towards a more reliable historical reanalysis: Improvements for version 3 of the Twentieth Century Reanalysis system, Q. J. R. Meteor. Soc., 145, 2876, https://doi.org/10.1002/qj.3598, 2019a.

Slivinski, L. C., Compo, G. P., Whitaker, J. S., Sardeshmukh, P. D., Giese, B. S., McColl, C., Brohan, P., Allan, R., Yin, X., Vose, R., Titchner, H., Kennedy, J., Spencer, L. J., Ashcroft, L., Bronnimann, S., Brunet, M., Camuffo, D., Cornes, R., Cram, T. A., Crouthamel, R., Dominguez-Castro, F., Freeman, J. E., Gergis, J., Hawkins, E., Jones, P. D., Jourdain, S., Kaplan, A., Kubota, H., Blancq, F. L., Lee, T., Lorrey, A., Luterbacher, J., Maugeri, M., Mock, C. J., Moore, K., Przybylak, R., Pudmenzky, C., Reason, C., Slonosky, V. C., Smith, C., Tinz, B., Trewin, B., Valente, M. A., Wang, X. L., Wilkinson, C., Wood, K., and Wyszynski, P.: NOAA-CIRES-DOE Twentieth Century Reanalysis Version 3, Research Data Archive at the National Center for Atmospheric Research, Computational and Information Systems Laboratory, https://doi.org/10.5065/H93G-WS83, 2019b.

Slivinski, L. C., Compo, G. P., Sardeshmukh, P. D., Whitaker, J. S., McColl, C., Allan, R. J., Brohan, P., Yin, X., Smith, C. A., Spencer, L. J., Vose, R. S., Rohrer, M., Conroy, R. P., Schuster, D. C., Kennedy, J. J., Ashcroft, L., Brönnimann, S., Brunet, M., Camuffo, D., Cornes, R., Cram, T. A., Domínguez-Castro, F., Freeman, J. E., Gergis, J., Hawkins, E., Jones, P. D., Kubota, H., Lee, T. C., Lorrey, A. M., Luterbacher, J., Mock, C. J., Przybylak, R. K., Pudmenzky, C., Slonosky, V. C., Tinz, B., Trewin, B., Wang, X. L., Wilkinson, C., Wood, K., and Wyszyński, P.: An Evaluation of the Performance of the Twentieth Century Reanalysis Version 3, J. Climate, 34, 1417, https://doi.org/10.1175/JCLI-D-20-0505.1, 2021.

Smith, S. D. and Banke, E.: Variation of the sea surface drag coefficient with wind speed, Q. J. R. Meteorol. Soc., 101, 665–673, 1975.

Tadesse, M. G. and Wahl, T.: A database of global storm surge reconstructions, Scientific Data, 8, 125, https://doi.org/10.1038/s41597-021-00906-x, 2021.

Thompson, V., Dunstone, N. J., Scaife, A. A., Smith, D. M., Slingo, J. M., Brown, S., and Belcher, S. E.: High risk of unprecedented UK rainfall in the current climate, Nat. Commun., 8, 107, https://doi.org/10.1038/s41467-017-00275-3, 2017.

Volonté, A., Clark, P. A., and Gray, S. L.: The role of mesoscale instabilities in the sting-jet dynamics of windstorm Tini, Q. J. R. Meteorol. Soc., 144, 877–899, https://doi.org/10.1002/qj.3264, 2018.

Walz, M. A. and Leckebusch, G. C.: Loss potentials based on an ensemble forecast: How likely are winter windstorm losses similar to 1990?, Atmos. Sci. Lett., 20, e891, https://doi.org/10.1002/asl.891, 2019.

Whitaker, J. S. and Hamill, T. M.: Ensemble data assimilation without perturbed observations, Mon. Weather Rev., 130, 1913–1924, https://doi.org/10.1175/1520-0493(2002)130<1913:EDAWPO>2.0, 2002.

Whitaker, J. S. and Hamill, T. M.: Evaluating methods to account for system errors in ensemble data assimilation, Mon. Weather Rev., 140, 3078–3089, 2012.

Woo, G. and Johnson N. F.: Stochastic modelling of possible pasts to illuminate future risk, in: Oxford handbook of complex disaster risks, edited by: Shultz, J., Rechkemmer, A., and Johnson, N. F., Oxford University Press, Oxford, ISBN 9780197554982, 2018.

Zimmerli, P. and Renggli, D.: Winter storms in Europe: messages from forgotten catastrophes, Swiss Re, https://www.swissre.com/risk-knowledge/mitigating-climate-risk/winter-storms-in-europe/winter-storms-in-europe-messages.html (last access: 17 April 2023), 2015.