Is the storminess in the Twentieth Century Reanalysis really inconsistent with observations? A reply to the comment by Krueger et al. (2013b)

**Xiaolan L. Wang · Y. Feng · G. P. Compo ·
F. W. Zwiers · R. J. Allan · V. R. Swail ·
P. D. Sardeshmukh**

Abstract In a recent study of trends and low frequency variability of extra-tropical cyclone activity in the ensemble of Twentieth Century Reanalyses, we concluded that "For the North Atlantic-European region and southeast Australia, the 20CR cyclone trends are in agreement with trends in geostrophic wind extremes derived from in-situ surface pressure observations". This conclusion has been challenged by Krueger et al. (Clim Dyn, submitted, 2013b), because a recent study (doi:10.1175/JCLI-D-12-00309.1, by the same lead author) comparing annual 95th percentiles (P95) of geostrophic wind speed (geo-wind) derived from surface pressure observations and from the 20CR found that "20CR-geostrophic storminess deviates to a large extent from the observation-based curve" in the period prior to 1950. In this reply, we show that our conclusion is valid; and we clarify that several factors contribute to the reported inconsistencies between the 20CR and observation-based geo-wind extremes. These include the choice of index that is used to represent the temporal variation of extremes (e.g., annual vs. seasonal percentiles), the use of different sampling intervals (6-hourly vs. 3-hourly), and the presence of very large errors in the observations that were not identified, corrected, or excluded in any of the previous studies of observation-based geo-wind extremes. We show that the time series of consecutive seasonal P95 geo-winds derived from the observations and from 20CR are in good agreement back to about 1893, with some deviation earlier when the observations (especially digitized data) remain limited and are more uncertain. We find that the correlation between the 20CR and observation-based geo-wind extremes (P95) time series for the full 134-year record is highly significant statistically, with and without the correction or exclusion of the newly identified erroneous SLP values. The agreement between 20CR and observations is further improved after the correction or exclusion of these erroneous values.

Keywords Geostrophic wind extremes · Cyclone activity · Trend analysis · Low-frequency variability

X. L. Wang (✉)
Climate Research Division, Atmospheric Science
and Technology Directorate, Science and Technology Branch,
Environment Canada, 4905 Dufferin Street, Toronto,
ON M3H 5T4, Canada
e-mail: Xiaolan.Wang@ec.gc.ca

X. L. Wang · Y. Feng · V. R. Swail
Climate Research Division, Science and Technology Branch,
Environment Canada, Toronto, ON, Canada

G. P. Compo · P. D. Sardeshmukh
CIRES, University of Colorado, Boulder, CO, USA

G. P. Compo · P. D. Sardeshmukh
Physical Sciences Division, NOAA Earth System Research
Laboratory, Boulder, CO, USA

F. W. Zwiers
Pacific Climate Impacts Consortium, University of Victoria,
Victoria, BC, Canada

R. J. Allan
Hadley Centre, Met Office, Exeter, UK

1 Introduction

Extremes of extratropical geostrophic wind speed (geo-wind) derived from long-term historical sub-daily surface pressure observations have been used to infer historical storminess conditions and trends over the northeast Atlantic-European region (Wang et al. 2009, 2011, referred to as W09 and W11 hereafter; Alexandersson et al. 1998, 2000; Schmidt et al. 1998; WASA 1998) and over southeast Australia

(Alexander et al. 2011). This is because surface pressure observations are, relatively, more reliable and temporally homogeneous than surface wind observations, and extratropical geo-wind extremes derived from surface pressure observations well approximate surface wind speed extremes (W09).

Recently, Wang et al. (2012), referred to as W12 hereafter, applied an objective cyclone tracking algorithm to analyze trends and low frequency variability of extra-tropical cyclone activity in the ensemble of Twentieth Century Reanalyses (20CR; Compo et al. 2011). On the basis of comparing linear trends estimated from a seasonal cyclone activity index (CAI) time series and seasonal 95th geo-wind percentiles in winter (JFM), they concluded that, for the North Atlantic-European region, the 20CR cyclone trends are in agreement with trends in geo-wind extremes derived from in-situ surface pressure observations. This conclusion has been challenged by Krueger et al. (2013b) because a recent study by the same lead author (Krueger et al. 2013a) comparing *annual* 95th geo-wind percentiles derived from surface pressure observations with those from the 20CR found that "20CR-geostrophic storminess deviates to a large extent from the observation-based curve" in the period prior to 1950.

In this reply, after briefly describing the data and methodology in Sect. 2, we corroborate in Sect. 3 that our conclusion is valid; and we clarify in Sect. 4 that several factors contribute to the apparent inconsistencies between the 20CR and observation-based geo-wind extremes reported by Krueger et al. (2013a, b). Finally, we summarize our conclusions in Sect. 5.

2 Data and methodology

The 10 triangles analyzed in both W09 and W11 are shown in Fig. 1. We also show station Armagh because we use the data from this station to help verify suspect values in some of the other stations (see "Appendix" and Figs. 6, 7). The 20CR gridpoints used to approximate the stations, also shown in Fig. 1 (red diamonds), are exactly the same as in Krueger et al. (2013a, b). In addition, the 50-km EASE (Equal Area SSM/I Earth) gridpoints used to obtain the North Sea regional mean CAI values in W12 are shown in Fig. 1 (thin blue crosses). In W12, CAI values in 5×5 arrays of 50-km EASE-gridpoints were aggregated to represent the CAI at the 250-km EASE grid scale (the center point in the 5×5 array of 50-km EASE-gridpoints; see Fig. 1, thick black crosses), obtaining 11 CAI time series (one for each of 11,250-km EASE-gridpoints). Each of these CAI time series was standardized before being averaged to obtain the regional mean series. There was a minor error in calculating the regional mean CAI values in W12, namely, only 9 CAI time series (cyan dots in Fig. 1),
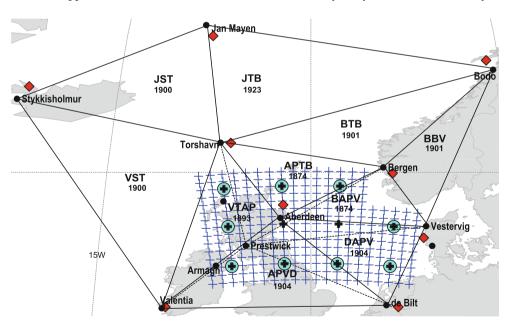

Fig. 1 The pressure triangles that were analyzed in Wang et al. (2009). All triangles with a *dotted line* are supplementary triangles (see Sect. 2 and Table 2 of Wang et al. 2009). The first year of geo-wind data is also shown in each of the 10 triangles. The *red diamonds* represent the 20CR gridpoints that are used to approximate the stations for calculating geo-winds from the 20CR data. The *thin blue crosses* indicate the set of 50-km EASE-gridpoints over which the cyclone activity index (CAI) was averaged to obtain the North Sea

regional mean CAI values for comparison with the regional mean geo-wind extremes. The *thick black crosses* represent the 250-km EASE-gridpoints; CAI values were aggregated for the 5×5 arrays of 50-km EASE-gridpoints centered at each of these 250-km EASE-gridpoints (see W12). The *cyan dots* show the set of 250-km EASE-gridpoints over which the regional mean CAI curve shown in Fig. 12a of W12 was obtained

were used to obtain the regional mean. This error has an effect on the 11-year Gaussian filtered CAI curve reported in W12 but no discernible effect on the trend estimate (see the magenta and blue curves in Fig. 2a; they share the same trend estimate).

Before proceeding to any analysis, we screen the sea level pressure (SLP) observations from stations show in Fig. 1 for large errors (errors greater than 20 hPa) using the procedure detailed in the "Appendix", and correct or exclude the identified erroneous values, obtaining a "new" observational data set (NewObs). As a result of the screening for large errors, we found 108 segments (short periods within a data record) with erroneous SLP values (Figs. 6, 7). The Aberdeen and Torshavn records contain the most identified errors. For the entire collection, almost all of these errors occur in the pre-1948 period and appear to have been introduced primarily during the digitization of paper records or from other post-measurement processing procedures. The errors are usually on the order of tens of hPa (Figs. 6, 7) and, as would be expected, have notable

effects on the observation-based geo-wind extremes. These erroneous values were not identified, nor corrected or excluded in any of the previous studies using the surface pressure data of the WASA (Waves and Storms in the North Atlantic) project (Alexandersson et al. 1998, 2000; Schmidt et al. 1998; WASA 1998; W09; W11; Krueger et al. 2013a; and other pressure tendency studies using the WASA data), although W09 had already identified and excluded 49 random errors in their analysis. These errors, and those identified in W09, are all in the WASA pressure data set (Schmidt et al. 1997; WASA 1998). The WASA data set has been incorporated into the International Surface Pressure Databank (ISPD, Yin et al. 2008), which was assimilated into the 20CR and also used in W09 and W11. Thus, all these errors are apparently present in Krueger et al. (2013a, b) and Alexandersson et al. (1998, 2000); most are also present in W09 and W11 (except the 49 identified in W09). These post-measurement digitization and processing errors are much larger than measurement errors and represent a major source of uncertainty in the
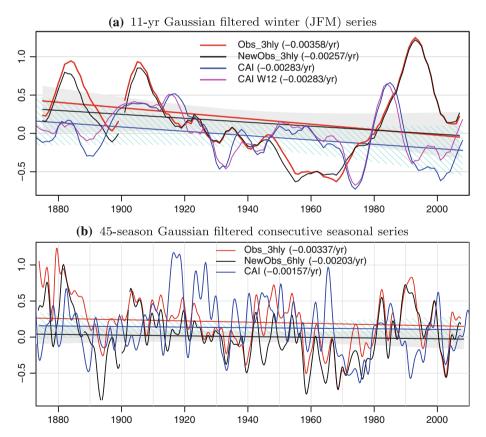

(a) 11-yr Gaussian filtered winter (JFM) series

Obs_3hly (−0.00358/yr)
NewObs_3hly (−0.00257/yr)
CAI (−0.00283/yr)
CAI W12 (−0.00283/yr)

(b) 45-season Gaussian filtered consecutive seasonal series

Obs_3hly (−0.00337/yr)
NewObs_6hly (−0.00203/yr)
CAI (−0.00157/yr)

Fig. 2 The Gaussian filtered series of the North Sea regional averages of standardized cyclone activity index (CAI), and of standardized seasonal P95 geo-winds (*black and red curves*) **a** in winter and **b** in all four seasons consecutively. The *red and black curves* are based on the 3-hourly or 6-hourly (as indicated) geo-winds derived from the Obs and NewObs data (see Sect. 2), respectively. The *blue (magenta) curve* represents averages of CAI over the 11 points shown as *thick black crosses* (the 9 points shown as *cyan dots*) in Fig. 1. The numbers in parentheses are the trend estimates for the corresponding time series. The cyan hatching represents the 95 % confidence interval of the *blue trend line*, and the *grey shading*, of the *black trend line*. The *red (magenta) curve* in **a** is a copy of the *black (dashed) curve* in Fig. 12a of W12. As in W12, the standardization is relative the mean and standard deviation of the period 1961–1990, and the seasons are defined as JFM, AMJ, JAS, and OND

observations. The error model of Krueger et al. ([2013a](#)) considers only a 1 hPa standard deviation measurement error, and thus does not fully represent the uncertainty in the observations.
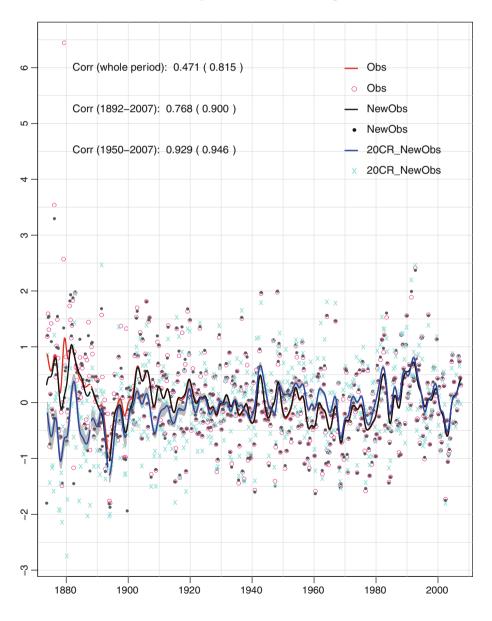
The observed sub-daily SLP time series (typically with two or three values daily in the early decades, and 3-hourly or hourly in the recent decades) were interpolated to a 3-hourly data series using the same procedure as in W09 and W11 (natural spline interpolation; as explained in W09, interpolation is necessary because the hours of observations vary from station to station, and also over time). Since the available 20CR data are 6-hourly, we sample the resulting 3-hourly observations at the same 6-hourly time steps (0000, 0600, 1200, 1800 hours) as in 20CR, and we exclude (set to missing) the 20CR time steps where the observations in the NewObs data set are missing, obtaining the 20CR_NewObs data set. Thus, both the 20CR

and observations (20CR_NewObs and NewObs) have exactly the same number of non-missing geo-wind data points, at the same sequence of 6-hourly time steps. Also, exactly the same method was used to calculate geo-winds from the observed and 20CR 6-hourly SLP data, and the same methods were used to derive the annual and seasonal 95th percentiles (P95). For comparison purposes, we also consider annual and seasonal P95 values based on geo-winds from 3-hourly observational SLP data. These are indicated with "_3hly" hereafter.

3 Observed and 20CR trends

First, we show that our conclusion that "For the North Atlantic-European region and southeast Australia, the 20CR cyclone trends are in agreement with trends in

Fig. 3 North Atlantic regional averages of standardized consecutive seasonal P95 geo-winds (*symbols*) and the corresponding 45-season Gaussian filtered series (*curves*). The seasonal P95 values are obtained using the method of W11 to diminish aliasing effects. The NewObs and Obs indicate that the geo-winds are calculated from SLP data with and without correction (or exclusion in some cases) of the newly identified errors, respectively. The 20CR_NewObs is the 20CR in which time steps corresponding to missing values in the NewObs data set have been masked off (set to missing). The *grey shading* represents the 20CR_NewObs ensemble spread. Discontinuities in the curves represent periods of missing geo-winds. The correlations between the *black* and *blue* curves (filtered series) are reported without parentheses on the graph, and those between the *dots* and *crosses* (unfiltered series) are reported in parentheses
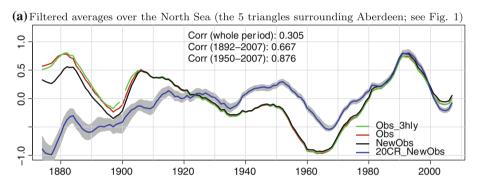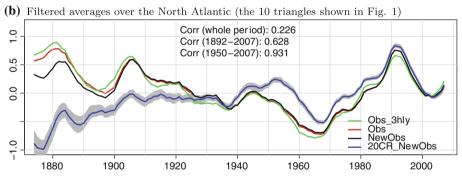
geostrophic wind extremes derived from in-situ surface pressure observations" is valid. To this end, we obtain annual P95 values conventionally (i.e., as the 95th percentile considering all sub-daily geo-winds in each calendar year), and we also obtain consecutive seasonal P95 values using the method of W11. Specifically, we first calculate the seasonal P95 of all sub-daily geo-winds in moving 91-day windows, obtaining a daily time series of moving season P95 values. Then, a 91-day moving average procedure is applied to the daily series of moving season P95 values, obtaining a daily time series of 91-day moving averaged values of moving season P95. This latter daily series is sampled seasonally, at four mid-season days of each year, to obtain the consecutive seasonal P95 values analyzed in this study. The power spectrum of such a
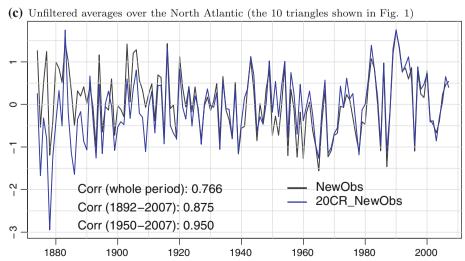
consecutive seasonal P95 series is presented by the green curve in Fig. 1 of W11, indicating that this series contains little aliasing effect (W11).

For the North Sea region (the area of the 5 triangles: APTB, BAPV, DAPV, APVD, and VTAP; see Fig. 1), as shown in Fig. 2, the agreement between the linear trend estimates for the 20CR CAI time series and for seasonal P95 geo-wind time series is even better than reported in W12 after the correction or exclusion of the newly identified data errors. The linear trends are estimated using the method detailed in Wang and Swail (2001), which is based on the Kendall's tau (Sen 1968; Kendall 1955; Mann 1945) and also accounts for lag-1 autocorrelation. The 95 % confidence interval for the trend is estimated using the variance of the corresponding residual series (von Storch and Zwiers 1999).

Fig. 4 a, b The 11-year Gaussian filtered series, and **c** the unfiltered series, of regional averages of standardized annual P95 geo-winds over the indicated regions. The annual P95 values are determined conventionally by using all 6-hourly geo-winds in each calendar year, except for the Obs_3hly curve, which is derived from all 3-hourly geo-winds (it is a copy of the annual P95 curve shown in Fig. 2 of W09). The NewObs and Obs indicate that the geo-winds are calculated from SLP data with and without correction/ exclusion of the newly identified errors, respectively. The 20CR_NewObs is the 20CR in which time steps corresponding to missing values in the NewObs data set have been masked off (set to missing). The *grey shading* represents the 20CR_NewObs ensemble spread (from the minimum to the maximum values among the 56 members). Discontinuities in the curves represent periods of missing geo-winds. The correlations (Corr) reported on each panel are those between the *black* and *blue curves* for the indicated periods

(a) Filtered averages over the North Sea (the 5 triangles surrounding Aberdeen; see Fig. 1)

Corr (whole period): 0.305
Corr (1892–2007): 0.667
Corr (1950–2007): 0.876

Obs_3hly
Obs
NewObs
20CR_NewObs

(b) Filtered averages over the North Atlantic (the 10 triangles shown in Fig. 1)

Corr (whole period): 0.226
Corr (1892–2007): 0.628
Corr (1950–2007): 0.931

Obs_3hly
Obs
NewObs
20CR_NewObs

(c) Unfiltered averages over the North Atlantic (the 10 triangles shown in Fig. 1)

Corr (whole period): 0.766
Corr (1892–2007): 0.875
Corr (1950–2007): 0.950

NewObs
20CR_NewObs

In winter (JFM), the linear trend estimate for the 20CR CAI time series is closer to that for the NewObs_3hly P95 geo-wind time series than to that for the Obs_3hly P95 geo-wind time series (see the numbers in parentheses in Fig. 2a). In other words, the agreement between the blue and black trend lines in Fig. 2a is better than that between the blue and red trend lines [the latter is what was shown in Fig. 12a of W12]. This is also true for the consecutive seasonal time series, as shown in Fig. 2b. Also, the 95 % confidence interval of the consecutive seasonal CAI trend estimate (cyan hatching) overlaps substantially with that of the NewObs consecutive seasonal P95 geo-wind trend estimate (grey shading in Fig. 2b). Note that the Obs_3hly and Ne-wObs_3hly seasonal P95 geo-winds in Fig. 2, and in Fig. 12a of W12, are derived conventionally from all 3-hourly geo-winds in each season of each year (i.e. same as in W09). But the NewObs (also labeled as "NewObs_6hly" in Fig. 2b) seasonal P95 geo-winds are derived using the method of W11 and thus contains little aliasing effect. Both the CAI values and the NewObs seasonal P95 geo-winds in Fig. 2b are from 6-hourly data and thus are more comparable.

The difference between the pair of blue and black trend lines in Fig. 2a or b is statistically insignificant, because the linear trend estimated for the time series of the differences between the 20CR CAI time series and the NewObs_3hly or NewObs P95 geo-wind time series is insignificant. The 95 % confidence interval for the trend estimated from the difference time series is $(-0.00482, 0.00692)$ for Fig. 2a, and $(-0.00832, 0.00071)$ for 2b, both indicating that the trend in the differences is insignificant.

Note that we discussed both linear trends and low-frequency variability in W12, which is clearly specified even in the title of the study, and that we concluded only that the linear trend estimates are consistent with each other. We computed linear trends, because it is a common summary of one aspect of a time series and linear trends are of broad interest to users, and because low-frequency variability can exist with or without a long-term linear trend.

4 Factors contributing to the reported inconsistencies

Next, we show that the inconsistencies between the 20CR and observation-based geo-wind extremes reported by Krueger et al. (2013a, b) are, to a large extent, an artifact of using annual percentiles to represent extremes. The curves in Fig. 3 represent 45-season (11.25-year) Gaussian filtered series of the consecutive standardized seasonal P95 values, derived from the uncorrected observations (red), from the newly corrected observations (black), and from 20CR data with missing values in the newly corrected observations being excluded (blue). The symbols in Fig. 3 indicate the unfiltered consecutive seasonal P95 values. Note that in

Fig. 3 the seasons are defined as DJF, MAM, JJA, and SON, to be consistent with W09 and W11.

In general, the 20CR and observation-based consecutive seasonal P95 series are in good agreement, especially over the period since 1893 (Fig. 3, blue and black curves). The correlation between the unfiltered series (dots and crosses in Fig. 3) is 0.815, 0.900, and 0.946 for the whole period (1874–2007), the period from 1892 to 2007, and the period from 1950 to 2007, respectively (Fig. 3, the numbers in parentheses). All correlations are highly significant (the 99.99 % critical value for sample correlations is 0.160 for sample size $N = 134 \times 4 = 536$). The slightly lower correlation for the whole period is due to the deviation in the pre-1893 period. Note that this deviation is substantially smaller than that of the annual P95 time series shown in Fig. 4 and in Krueger et al. (2013a, b). The correlation between the filtered series (black and blue curves in Fig. 3) is lower than that between the unfiltered series (dots and crosses), particularly for the whole period due to the pre-1893 discrepancy between the observational P95 values and those from 20CR.

The differences between the P95 values in the filtered Obs and NewObs (red and black curves, respectively), and between the unfiltered Obs and NewObs (pink circles and black dots, respectively), shown in Fig. 3 are purely due to the effect of the newly identified observational errors that were included in the Obs data set, but are either corrected or excluded in the NewObs data set. These errors, especially the long run of very large errors (greater than 30 hPa) in the Aberdeen record (Fig. 6, top panel), result in a few very large outliers (up to about 6.5 standard deviations in 1879) in the observation-based geo-wind extremes (see Fig. 3). Their effects are particularly notable in the first decade (compare Fig. 3, red and black curves). This is because only two of the 10 triangles (APTB and BAPV) have any geo-wind data for the pre-1893 period. Both of these triangles include the erroneous Aberdeen record, and one also includes the erroneous Torshavn record (Fig. 1). Since there were only two geo-winds triangles available in the pre-1893 period, the uncertainty in the observationally estimated curve is expected to be substantially larger in this early period.

It is important to note that 20CR assimilated marine observations and other station data in the early period, in addition to the few stations that were used in the geo-wind calculation. For each pressure triangle, the geo-winds derived from the 20CR data indirectly involve observations in the vicinity of the triangle and farther afield. That is, several types of observational information were used in the 20CR, which enables more comprehensive quality control (QC) of the observational data, potentially resulting in geo-wind estimates that are less affected by observational errors than the geo-winds derived from observational pressure triangles. For example, 143 out of the 146 newly identified erroneous values in the Aberdeen record for the period 1871–1921 were

rejected by the 20CR QC system (see Compo et al. 2011 "Appendix B" for a detailed description of the 20CR QC system). For the Aberdeen record for year 1879, the 20CR rejected 98 values, including the long run of large errors (38 erroneous values) in October 1879 that were identified in this study. Some of the errors identified by the 20CR are probably smaller than 20 hPa, so that the procedure of screening for large errors conducted in this study (see "Appendix") cannot detect them. This may be an additional reason for the deviation between the 20CR and observed geo-winds in the pre-1893 period (Fig. 3, blue and black curves, respectively). Further in-depth analysis of the marine and other station data collectively is necessary to find the causes behind the remaining deviation; we plan to undertake this time consuming task in the near future. We believe that the uncertainty also requires further investigation in this early period, both for the observations and 20CR. More in-depth quality assurance of the pressure data and digitization of more observed data in the early period, such as being coordinated by the Atmospheric Circulation Reconstructions over the Earth initiative (Allan et al. 2011), will help reduce uncertainty.

Next, despite the known aliasing issues, we reconsider the conventional annual P95 geo-winds as in previous studies (e.g., Alexandersson et al. 1998, 2000; Krueger et al. 2013a). Figures 4a, b show the 11-year Gaussian filtered regional averages of the standardized annual P95 geo-winds for both the North Sea and the North Atlantic regions (see Fig. 1). For the NewObs and 20CR_NewObs data sets, the unfiltered series of standardized annual P95 geo-winds are also shown in Fig. 4c. The annual P95 values are derived conventionally, as in W09, Alexandersson et al. (1998, 2000), and Krueger et al. (2013a). The Obs_3hly (green) curve is the annual P95 curve shown in Fig. 2 of W09 (also in Fig. 1 of Krueger et al. 2013b), which was based on 3-hourly geo-winds derived from the SLP data without the correction or exclusion of the errors shown in Figs. 6, 7.

The difference between the green and red (Obs_3hly and Obs) curves in Fig. 4a, b arises solely from the sampling interval, i.e., 6-hourly (Obs) versus 3-hourly (Obs_3hly) geo-winds. For the North Atlantic averages, sampling from 3-hourly geo-winds gives higher annual P95 values in the early decades (and lower values in the 1960s) than sampling from 6-hourly geo-winds (Fig. 4b, green and red curves). For the North Sea area, the differences are smaller but still noticeable in the early decades (Fig. 4a, green and red curves). In Krueger et al. (2013a, b), the observed geo-winds were derived from 3-hourly data, but the 20CR geo-winds were 6-hourly (because the available 20CR data are 6-hourly). This contributes modestly to the deviation between the 20CR and observed low-pass filtered annual P95 time series, particularly for the full domain (Fig. 4b; compare blue and green curves vs. blue and red curves).

The difference between the red and black (Obs and NewObs) curves in Fig. 4a, b is purely due to the correction or exclusion of the newly identified erroneous SLP values shown in Figs. 6, 7. The effect of this is particularly notable in the pre-1900 period. Since there are only two geo-wind triangles available in the pre-1893 period and both triangles are included in the North Sea area, the differences in the regional averages between the North Sea and the North Atlantic regions (compare the same colour curves in Fig. 4a, b) in this early period are small; they are purely due to the standardization that is based on the mean and variance of the whole period for each curve.

As can be seen from the blue and black curves in Figs. 3 and 4b, the deviations between the 20CR and observation-based consecutive standardized seasonal P95 series are much smaller than those between the corresponding annual P95 series. The inconsistencies between 20CR and observation-based geo-wind extremes reported by Krueger et al. (2013a) are mainly in the annual P95 time series and are, to a large extent, due to the annual sampling (other contributors are described above). The annual sampling convolves the very different storminess regimes in different seasons. The resulting annual P95 time series suffers from aliasing between the effects of low-frequency variability and the annual cycle. For example, differences in the annual cycle between the 20CR and station-data based P95 geo-winds (see Fig. 5) would be aliased and shown as differences in the low-frequency variability. On the contrary, the annual cycle in both the mean and variance of geo-wind extremes is effectively diminished from our consecutive standardized seasonal P95 time series. This is because our seasonal P95 values are derived from 4 distributions (one for each season) for each year and are then standardized in each of the four seasons of year, separately (i.e., the standardization is with respect to the mean and variance in each season).

Differences could exist in the annual cycle of both the mean and variance, as shown in Fig. 5, because of (1) the small differences between the 20CR and station-based triangles, (2) the fact that 20CR used more observational information (marine and other station data) than the station-data based geo-winds, and (3) limitations of the 20CR model resolution. In particular, 20CR shows lower variance of P95 geo-winds over triangles DAPV, VTAP, and APVD, but higher mean P95 values with higher variability over triangle BBV, than the station-data based counterpart (Fig. 5).

Nevertheless, the correlations between the annual P95 time series of the NewObs and 20CR_NewObs data sets (see the numbers in Fig. 4c) are highly significant statistically. Even the lowest value (0.766), which is obtained for the whole 134-year period (1874–2007) for the North Atlantic region (Fig. 4c), is highly significant (for sample size $N = 134$, the 99.99 % critical value for sample correlations is 0.316). The correlations between the filtered
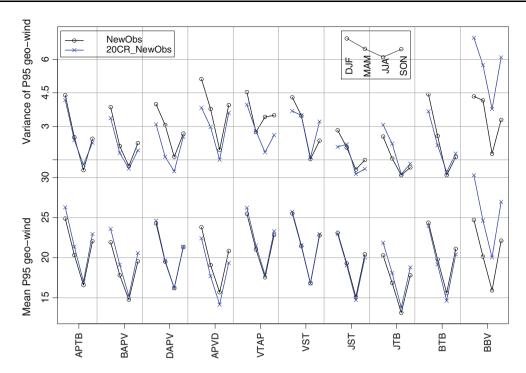
Fig. 5 The annual cycle in the mean and variance of the 20CR and station-data based P95 geo-winds for each of the 10 triangles (see the horizontal axis and Fig. 1)

annual P95 series (black and blue curves in Fig. 4a, b) are much lower, which may be partly due to the "the discrepancy-spreading effect" of the Gaussian filter.

5 Conclusions

In this reply, we have provided further evidence to show that the conclusion comparing linear trends in 20CR storminess and observation-based geo-wind extremes in W12 is valid. We have also clarified that several factors contribute to the apparent inconsistencies between the 20CR and observation-based geo-wind extremes reported by Krueger et al. (2013a, b). These include the choice of index that is used to represent time variation in extremes (e.g., annual vs. seasonal percentiles), the use of different sampling intervals (6-hourly vs. 3-hourly), and the presence of very large errors in the observations (i.e., the WASA pressure data set; Schmidt et al. 1997) that were not identified, nor corrected or excluded in any of the previous studies of observation-based geo-wind extremes (Alexandersson et al. 1998, 2000; Schmidt et al. 1998; WASA, 1998; W09; W11; Krueger et al. 2013a, b).

We have shown that the time series of consecutive seasonal P95 geo-winds derived from the observations and from 20CR are in good agreement starting in 1893, with some deviation in the pre-1893 period for which the observations (especially digitized data) remain limited and are more uncertain. The correlation between the 20CR and

observation-based geo-wind extremes (P95) time series for the full 134-year record is highly significant statistically, with and without the correction or exclusion of the newly identified erroneous SLP values. The agreement between 20CR and observations is further improved after the correction or exclusion of the newly identified erroneous SLP values.

Appendix: The newly identified erroneous SLP data
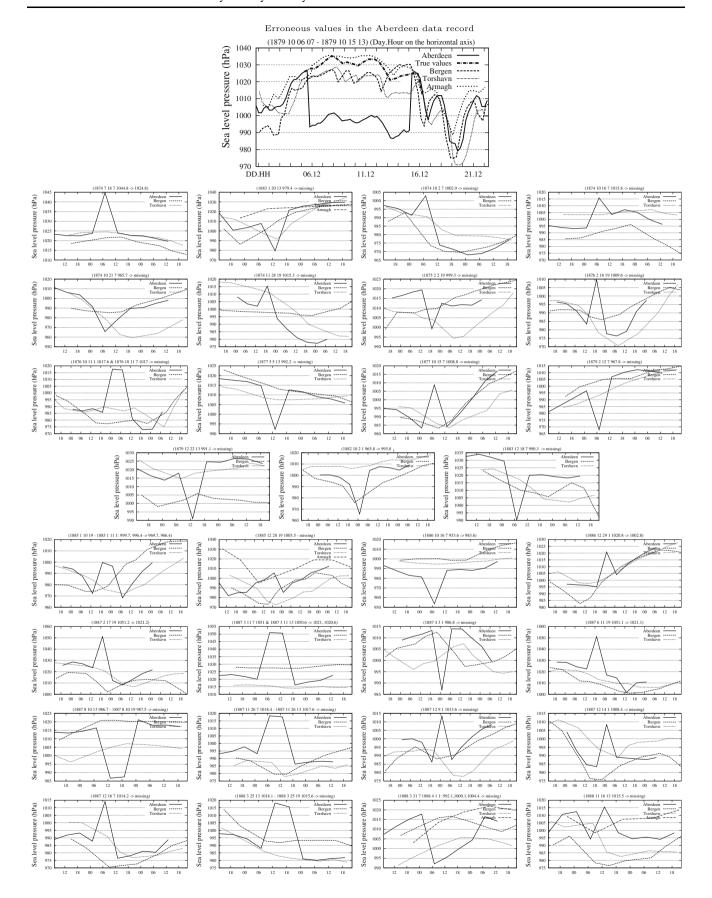
In W09, for any SLP series $\{P_i\}$ ($i = 1, 2, ..., N$), the i-th observation P_i is considered a suspicious value if $|P_{i-1} -$

Fig. 6 Segments of erroneous SLP values (i.e., the "outliers" in the *solid curves*) in the Aberdeen (ID: 03091) record, in comparison with the corresponding segments from nearest stations available. The *x*-axis is the hour of observation. The date, or date range, of the erroneous values are given on the top of each panel, where the word "missing" or the number(s) following "→" are the replacement for the erroneous values (here "missing" means "set to missing"). The "True values" (correct original observed values) in the *top panel* are extracted from the UK Daily Weather Report books for that period
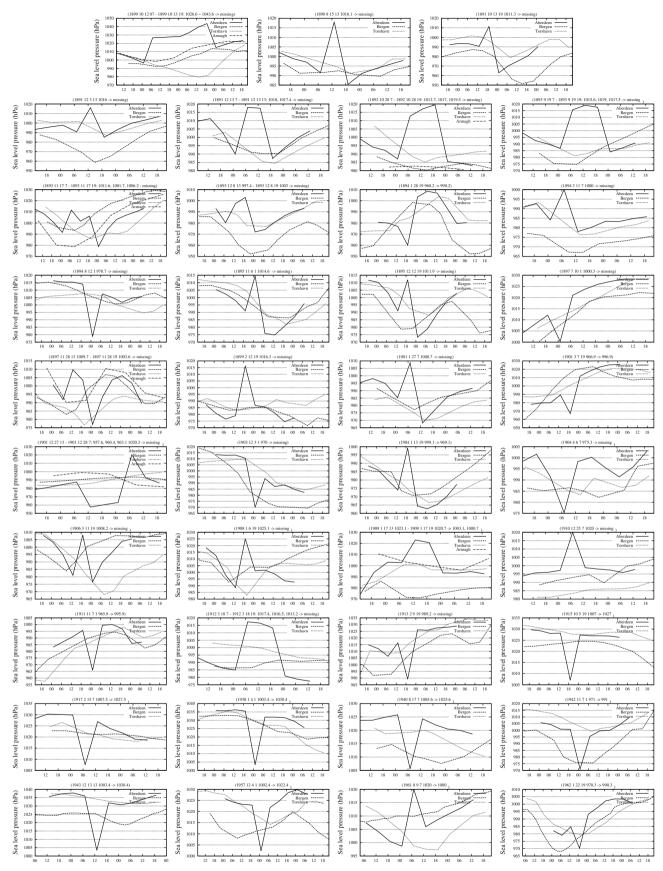
Erroneous values in the Aberdeen data record
(1879 10 06 07 - 1879 10 15 13) (Day.Hour on the horizontal axis)
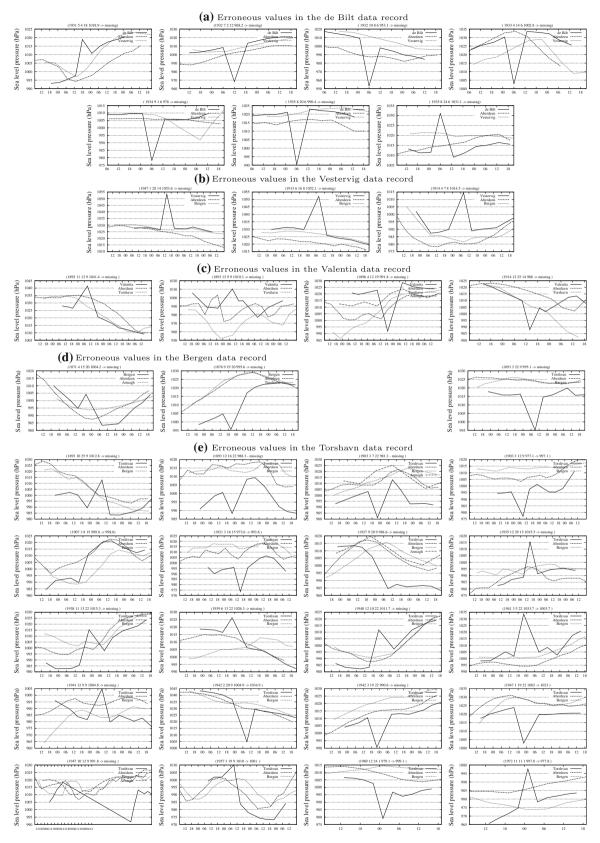
Fig. 6 continued

Fig. 7 Same as in Fig. 6 but for errors in the de Bilt (ID:06260), Vestervig (ID: 21100), Valentia (ID: 03953), Bergen (ID: 01311), and Torshavn (ID: 06011) data records

$P_i| > 20$ hPa **and** $|P_{i+1} - P_i| > 20$ hPa. This screening procedure is good for identifying random, isolated errors; but it would fail to identify erroneous values that occur consecutively. We have modified the procedure, namely, we first identify all P_i such that $|P_{i-1} - P_i| > 20$ hPa **or** $|P_{i+1} - P_i| > 20$ hPa; then, we visually investigate the plot of each data segment that contains such a P_i along the corresponding segments from two or three of the nearest stations if available (see Fig. 6), to determine whether or not the suspect value P_i is erroneous. The second step is essential, because the new screening procedure includes many real cases of large SLP tendencies (including real low pressure systems). For example, about 270 segments of suspect values were identified in the Torshavn record, but only 21 segments were determined to contain error(s); the other suspect values are consistent with values recorded at the neighbouring stations. This step is time consuming but necessary to ensure data quality. In the future, it may be possible to at least partially automate this second step. The longest run of very large errors is found in the Aberdeen record, from 6 October 1879 to 15 October 1879, as shown in the top panel of Fig. 6, where the "True values" (correct original observed values) are extracted from the UK Daily Weather Report books for that period. Except for this long run of errors, the number(s) used to replace the erroneous values (see Figs. 6, 7) are estimated using all values in the segment of error(s) along with those in the corresponding segments of the nearest stations. Such corrections make the data in that short period consistent among all stations in comparison. Best corrections need to be found from the original paper copy of the data records, which are currently not available to us. However, effort is been made to locate these original data. The ACRE (Atmospheric Circulation Reconstructions Over the Earth) initiative (Allan et al. 2011) has located sub-daily or hourly observations for many UK stations (including Aberdeen) and is aiming to have them scanned and digitized for future ISPD updates and historical reanalyses.

We find that many of the identified errors were introduced during the digitization of paper records, such as misreading of the data in inches of Hg (e.g. 29 instead of 30 in. or vice versa), or mis-typing the decimal values, such as .0 as .9, or .2 as .9 in., or vice versa, or swapping a pair of decimal values, or other typographical errors (see Figs. 6, 7). Mis-recorded observations in inches of Hg in the early period are also possible. Such mistakes create large errors, because the early observations are recorded in inches of Hg and an error of .9 in. Hg is an error of about 30 hPa. Errors can also be introduced during the conversion of the values in inches of Hg to hPa.

Note that the error-screening procedure (buddy check) here aims at detecting large errors (greater than 20 hPa), so that smaller errors would go undetected. Smaller errors are always harder to detect. This might be the case for the quality control (QC) system used in the 20CR. Some of the errors identified here might not have been identified by the 20CR QC system and may have been assimilated in the 20CR; no doubt some were identified and excluded in 20CR but not here. For example, 143 out of the 146 newly identified erroneous values in the Aberdeen record for the period 1871–1921 were rejected by the 20CR. For the Aberdeen record for year 1879, the 20CR rejected 98 values, including the 38 erroneous values in October 1879 and two errors in February and December 1879 (see Fig. 6). Some of the errors identified by the 20CR are probably smaller than 20 hPa, so that the above procedure of screening for large errors cannot detect them. These are completely different QC systems; neither of them is perfect. But it is arguable that the 20CR QC system (see Compo et al. 2011 "Appendix B" for a complete description) is more comprehensive since it uses all available data (including marine data) in the area.

References

Allan R, Brohan P, Compo GP, Stone R, Luterbacher J, Brönnimann S (2011) The international atmospheric circulation reconstructions over the Earth (ACRE) initiative. Bull Am Meteorol Soc 92:1421–1425. doi:10.1175/2011BAMS3218.1

Alexandersson H, Tuomenvirta H, Schmith T, Iden K (2000) Trends of storms in NW Europe derived from an updated pressure data set. Clim Res 14:71–73

Alexandersson H, Schmith T, Iden K, Tuomenvirta H (1998) Long-term variations of the storm climate over NW Europe. Glob Atmosphere Ocean Syst 6:97–120

Compo GP, Whitaker JS, Sardeshmukh PD, Matsui N, Allan RJ, Yin X, Gleason BE Jr, Vose RS, Rutledge G, Bessemoulin P, Brönnimann S, Brunet M, Crouthamel RI, Grant AN, Groisman PY, Jones PD, Kruk MC, Kruger AC, Marshall GJ, Maugeri M, Mok HY, Nordli , Ross TF, Trigo RM, Wang XL, Woodruff SD, Worley SJ (2011) The Twentieth Century Reanalysis project. Q J R Meteorol Soc 137:1–28. doi:10.1002/qj.776

Kendall MG (1955) Rank correlation methods. Charles Griffin, London, p 196

Krueger O, Schenk F, Feser F, Weisse R (2013a) Inconsistencies between long-term trends in storminess derived from the 20CR reanalysis and observations. J Clim 26(3):868–874. doi:10.1175/JCLI-D-12-00309.1

Krueger O, Feser F, Baerring L, Kaas E, Schmith T, Tuomenvirta, von Storch H (2013b) Comment on Trends and low frequency variability of extra-tropical cyclone activity in the ensemble of Twentieth Century Reanalysis. Clim Dyn, submitted

Mann HB (1945) Non-parametric tests against trend. Econometrica 13:245–259

Schmidt T, Alexandersson H, Iden K, Tuomenvirta H (1997) North Atlantic-European pressure observations 1868–1995 (WASA dataset 1.0; CD-ROM included). Technical report 97–3. Danish Meteorological Institute, Copenhagen, Denmark

Schmidt T, Kaas E, Li T-S (1998) Northeast Atlantic storminess 1875–1995 re-analyzed. Clim Dyn 14:529–536

Sen PK (1968) Estimates of the regression coefficient based on Kendall's tau. J Am Stat Assoc 63:1379–1389

von Storch H, Zwiers FW (1999) Statistical analysis in climate research. Cambridge University Press, Cambridge, p 484

Wang XL, Feng Y, Compo GP, Swail VR, Zwiers FW, Allan RJ, Sardeshmukh PD (2012) Trends and low frequency variability of extra-tropical cyclone activity in the ensemble of Twentieth Century Reanalysis. Clim Dyn, in press, doi:10.1007/s00382-012-1450-9

Wang XLL, Swail VR (2001) Changes of extreme wave heights in Northern Hemisphere oceans and related atmospheric circulation regimes. J Clim 14:2204–2221

Wang XLL, Wan H, Zwiers FW, Swail VR, Compo GP, Allan RJ, Vose RS, Jourdain S, Yin X (2011) Trends and low-frequency variability of storminess over western Europe, 1878–2007. Clim Dyn 37:2355–2371. doi:10.1007/s00382-011-1107-0

Wang XLL, Zwiers FW, Swail VR, Feng Y (2009) Trends and variability of storminess in the Northeast Atlantic region 1874–2007. Clim Dyn 33:1179–1195. doi:10.1007/s00382-008-0504-5

WASA (1998) Changing waves and storms in the Northeast Atlantic? Bull Am Meterol Soc 79:741–760

Yin X, Gleason BE, Vose RS, Compo GP, Matsui N (2008) The International surface Pressure Databank (ISPD) Version 2.2. National Climatic Data Center, Asheville, pp 1–12. Accessible at ftp://ncdc.noaa.gov/pub/data/ispd/doc/ISPD2_2.pdf