CrossMark

Reconstructing El Niño Southern Oscillation using data from ships' logbooks, 1815–1854. Part I: methodology and evaluation

Hannah G. Barrett[1] · Julie M. Jones[1] · Grant R. Bigg[1]

Abstract The meteorological information found within ships' logbooks is a unique and fascinating source of data for historical climatology. This study uses wind observations from logbooks covering the period 1815 to 1854 to reconstruct an index of El Niño Southern Oscillation (ENSO) for boreal winter (DJF). Statistically-based reconstructions of the Southern Oscillation Index (SOI) are obtained using two methods: principal component regression (PCR) and composite-plus-scale (CPS). Calibration and validation are carried out over the modern period 1979–2014, assessing the relationship between re-gridded seasonal ERA-Interim reanalysis wind data and the instrumental SOI. The reconstruction skill of both the PCR and CPS methods is found to be high with reduction of error skill scores of 0.80 and 0.75, respectively. The relationships derived during the fitting period are then applied to the logbook wind data to reconstruct the historical SOI. We develop a new method to assess the sensitivity of the reconstructions to using a limited number of observations per season and find that the CPS method performs better than PCR with a limited number of observations. A difference in the distribution of wind force terms used by British and Dutch ships is found, and its impact on the reconstruction assessed. The logbook reconstructions agree well with a previous SOI reconstructed from Jakarta rain day counts, 1830–1850, adding robustness to our reconstructions.

✉ Hannah G. Barrett
hbarrett2@sheffield.ac.uk

[1] Department of Geography, University of Sheffield, Winter Street, Sheffield S10 2TN, UK

Comparisons to additional documentary and proxy data sources are provided in a companion paper.

Keywords Ships' logbooks · ENSO reconstruction · El Niño · Climate change and variability

1 Introduction

The logbooks of ships which historically travelled the world's oceans contain a vast amount of meteorological information useful for studies of historical climate. These logbooks present a unique and exciting record of daily weather observations including wind force and direction, air temperature, pressure, and the state of the weather and the sea from the pre-instrumental era (Garcia-Herrera et al. 2005). Initial studies using this data source focused on individual events such as the reconstruction of daily synoptic conditions during key nautical battles (Wheeler 2001) or of sea ice cover during the summer of 1816 (Catchpole and Faurer 1985). More recently, the scope of these investigations has expanded to reconstructions of larger scale pressure fields (Gallego et al. 2005; Küttel et al. 2010), westerly winds (Barriopedro et al. 2014), temperature (Brohan et al. 2012) and precipitation (Hannaford et al. 2015) over much longer time periods. Jones and Salmon (2005) were the first to attempt a reconstruction of El Niño Southern Oscillation (ENSO) using logbook observations from across the Indian Ocean ($16°$N to $40°$S). However, gaps in the spatial and temporal coverage of the data were found to be a major restriction to their reconstruction (Jones and Salmon 2005). This paper builds on their initial attempt, by using the increased amount of data available since their work, and by focusing only on regions with strong ENSO signals.

El Niño Southern Oscillation is the dominant mode of interannual climate variability in the tropics and is an important part of the natural climate system (Diaz and Markgraf 1992). ENSO is a coupled atmosphere–ocean phenomenon associated with a see-saw pattern of atmospheric pressure and sea surface temperatures (SSTs) across the equatorial Pacific, with periodicities of 2 to 7 years. Its two main phases, El Niño and La Niña, are associated with regional changes in precipitation, sea level pressure and temperature. These extremes result in hazards such as flooding, drought and wildfires which have large socioeconomic consequences. Although ENSO is centred in the equatorial Pacific, its teleconnections have a near global extent making it a hugely important part of the climate system on sub-decadal timescales.

The most reliable record of ENSO behaviour is from instrumental data, which are obtained from in-situ measurements of atmospheric and oceanic variables. The Southern Oscillation Index (SOI), an atmospheric index of the pressure difference between Tahiti and Darwin, extends back to 1876 (Ropelewski and Jones 1987). This was extended further using early pressure data from Jakarta and Tahiti to 1841, and from rain-day counts from Jakarta back to 1829 (Können et al. 1998). The record of Jakarta rain-day counts was used to create an SOI covering June–November, 1829–1850, and provides the most direct record to compare to the logbook reconstructions. The Niño 3.4 index, based on SST, extends back to 1877 (Bunge and Clarke 2009). Understanding ENSO variability prior to these instrumental records helps to place current ENSO variability in a longer term context. This is important as it enables an assessment of whether modern ENSO behaviour is the result of natural climate variability or has been influenced by anthropogenic climate change (Gergis et al. 2006). Current understanding of ENSO behaviour during the early to mid-nineteenth century comes from a range of data sources. Proxy records, such as tree rings, corals and ice cores, are widely used as indicators of past climate and environmental change but have a range of uncertainties which are attached to them. They are less direct measures of the climate than data from ships' logbooks. Recent multi-proxy studies have attempted to bring together a range of reconstructions, however a lack of consensus on ENSO behaviour on an interannual time scale remains (Emile-Geay et al. 2013). Climate models provide an alternative source of information on past ENSO variability, but their ability to produce realistic ENSO simulations is still limited, and conflicts between models are apparent in multi-model comparison projects (Bellenger et al. 2014). The reconstructions using the data from logbooks presented in this paper provide a 'bridge' between modern instrumental records and proxy records. Here, we compare the logbook reconstructions to the early SOI records from the Jakarta observations, as both records are

from direct, well-dated observations (Können et al. 1998). Comparison between the logbook and multi-proxy reconstructions will be discussed further in a companion paper (Barrett et al. 2016).

Created as part of an international project completed in 2004, the Climatological Database for the World's Ocean (CLIWOC) provides access to a digital database of over 280,000 daily weather observations from mostly Dutch, British, Spanish and French national archives (Können and Koek 2005). The CLIWOC database spans 1662 to 1855, however, over 99% of the observations occur in the period 1750 to 1855 (Küttel et al. 2010). They provide a significant extension to modern instrumental records. However, the process of digitization is expensive and time-consuming so, following the CLIWOC project, there were still over 90% of European, mostly British, logbook collections yet to be digitised (Garcia-Herrera et al. 2005). An additional 891 logbooks have since been digitised in the UK from English East India Company (EEIC) records, providing a further 273,000 observations from 1789 to 1834 (Brohan et al. 2012). This paper considers the combined databases in order to utilise the highest number of observations possible. Despite these two large digitisation projects there is still a long way to go in order to fully exploit this resource (Wheeler 2014).

This paper is organised as follows: Sect. 2 describes the data sources used to carry out the historical ENSO reconstructions. Focus on spatial and temporal availability of the logbook data is provided. Sect. 3 explains the data selection procedure and the reconstruction and validation methodologies. It also introduces a new method of assessing the impact of changing the number of observations per season on reconstruction skill, which is particularly important given this characteristic of logbook data. Sect. 4 presents the results of the logbook data analysis with a focus on the wind force terms used. It also presents the results of an evaluation of the reconstruction methods and presents the historical ENSO reconstructions, along with comparison to the early Jakarta rain-day SOI record. The reconstructions and key findings are then discussed in Sect. 5, while Sect. 6 concludes this part of the study. A companion paper (Barrett et al. 2016) compares the results of this paper to previous ENSO reconstructions, documentary and proxy-based, and discusses the implications of these results.

2 Data

Wind direction and strength derived from ships' logbooks were used to reconstruct the Southern Oscillation Index. Due to a lack of overlap between the logbook data and the instrumental SOI, reanalysis data from the modern period (1979–2014) were used to establish statistical relationships

between zonal wind and the SOI. These relationships were then applied to the logbook wind data to reconstruct the historical SOI.

2.1 Reanalysis data

Sub-daily 10 m zonal wind values were obtained from ERA-Interim Reanalysis data (Dee et al. 2011). These data cover the period from 1979 to present, with 2013/2014 being the last DJF season used in this paper. Reanalysis wind data were re-gridded to $7.5° \times 8°$ latitude–longitude, following previous studies using logbook data which found the use of an $8° \times 8°$ latitude–longitude grid provided a good resolution to enable suitable logbook data density, without compromising the climatology (Gallego et al. 2005; Jones and Salmon 2005; Küttel et al. 2010; Hannaford et al. 2015). The re-gridding of the reanalysis data was carried out using climate data operators (CDO, https://code.zmaw.de/projects/cdo). ERA-Interim data are available at four synoptic time steps 00UTC, 06UTC, 12UTC and 18UTC. In order to best match the logbook data, which is taken at local noon, the daily value in each grid box was taken as the one closest to its local noon. These daily 'noon' values were then converted into monthly and seasonal values, based on the four standard meteorological seasons (DJF, MAM, JJA and SON). Only DJF values are used in the paper as this is the season in which ENSO events typically reach their peak intensity (Bellenger et al. 2014), but also in which a suitable amount of logbook data availability was found.

2.2 Instrumental ENSO data

A monthly SOI was obtained from the Australian Bureau of Meteorology (BoM) extending from 1876 to present (Bureau of Meteorology 2015). The Troup SOI method was used, which calculates the standardised anomaly of the mean sea level pressure difference between observations from Tahiti and Darwin. The seasonal SOI was calculated from these monthly values. Other indices of ENSO are available and preliminary investigations included a range of indices, but the SOI was selected as it is an atmospheric measure of ENSO and is therefore closely linked to winds and circulation patterns, so had the strongest correlations with zonal winds (not shown).

2.3 Ships' logbook data

The focus of this paper is on the wind data obtained from ships' logbooks, combining the data from the Climatological Database for the World's Oceans (CLIWOC) and digitised EEIC records. Wind direction was observed on-board ships using a 32-point compass and reported with respect to the direction from which wind was blowing (Wheeler 2005). For wind force, a range of descriptive terminology was used to make observations, as quantitative measures of wind speed had not yet been introduced. The Beaufort Force scale was not introduced into the Royal Navy until 1838, however prior to this there is evidence to suggest a fairly standardized vocabulary for recording wind force was already in place, with terms passed down through generations of sailors (Wheeler 2005). The various wind force terms found within the logbooks of the different countries were recorded in CLIWOC's 'Multilingual Meteorological Dictionary' (Garcia-Herrera et al. 2003). This included British, Dutch, Spanish and French terms, with translations carried out on a country by country basis. This reference guide was also used for transforming the wind force terms found within the additional EEIC logbooks (Brohan et al. 2012). These descriptive wind force terms were then transformed into Beaufort Force equivalents, and finally into modern SI units of ms^{-1} (Können and Koek 2005). The wind force and direction observations were used to calculate zonal winds for use in the reconstructions.

When combined, the CLIWOC and EEIC databases provide over half a million records globally over the period 1750 to 1854. However, 43% of these do not include observations of both wind force and direction. As both of these variables are required to calculate zonal wind, only those records with both wind force and direction are used here. The spatial and temporal distributions of these 283,943 wind observations are shown in Figs. 1 and 2. Figure 1 shows that logbook observations are concentrated along the tracks of key historical shipping routes, which has resulted in their uneven global distribution. The highest concentrations of data points are found in the Atlantic and Indian Oceans, and around the tip of South Africa. Very few observations are found over the Pacific Ocean as there were no regular shipping routes in this area during this period (Gallego et al. 2005).

Figure 2 shows the interannual variation in the number of wind observations globally from 1750 to 1854. Although the number of records increases in the period of EEIC data coverage, 1789 to 1834, the number of records with wind observations does not increase to a similar extent. This is a result of CLIWOC focusing on obtaining wind observations whereas, digitisation of EEIC logbooks prioritised temperature and pressure measurements (Brohan et al. 2012).

Although the coverage of observations appears highly concentrated in Fig. 1, the data density on a given day is extremely sparse and is a key limitation to studies using this data source at present. To overcome this data sparsity, daily observations were averaged and aggregated into seasonal values, and seasonal reconstructions were carried out. Wind observations were averaged and
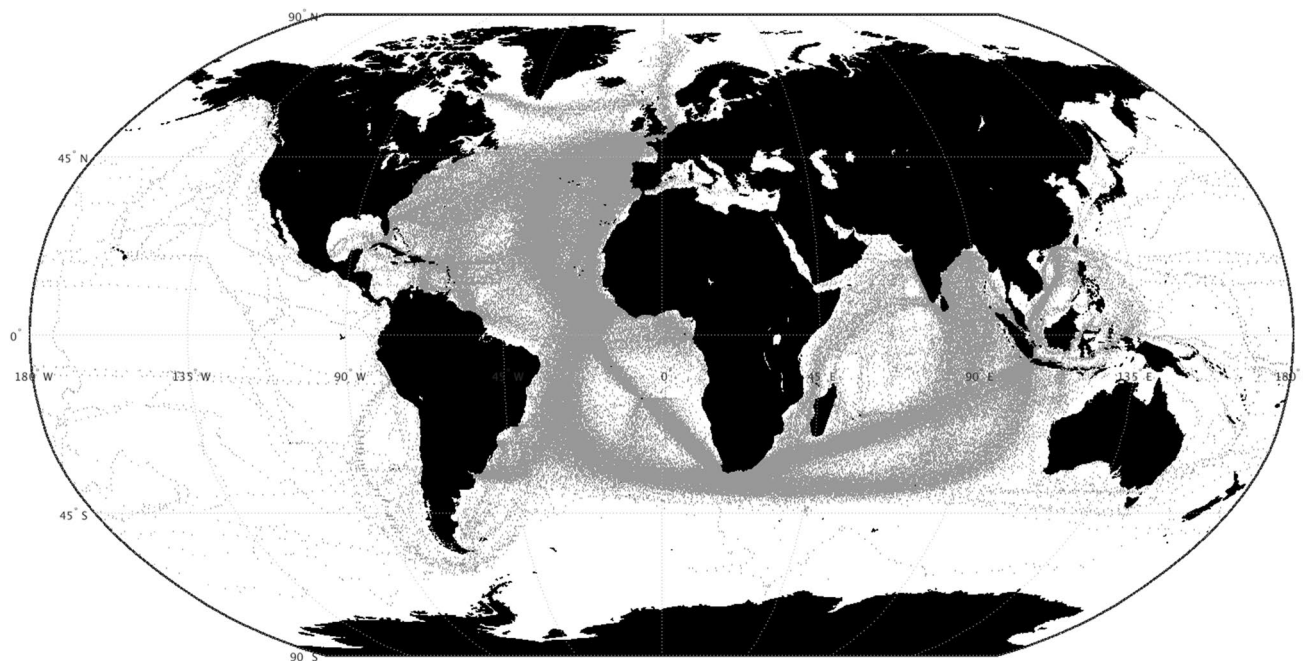
Fig. 1 Location of data entries from ships' logbooks with both wind force and wind direction observations from the digitised archives of CLI-WOC and EEIC, 1750–1854
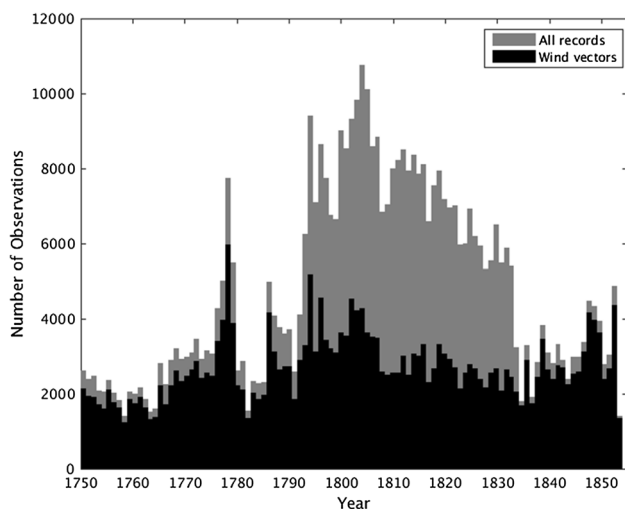

Fig. 2 Yearly distribution of all logbook data entries (*grey*) and those with wind vector observations (*black*), 1750–1854

aggregated into a 7.5° by 8° latitude–longitude grid, matching the resolution of the re-gridded ERA-Interim zonal winds. As ships often travelled in convoy, observations taken on the same day from within a given grid box were averaged into a daily value and counted as one observation for that day. Given that the logbook observations were taken at local noon, the effect of averaging observations from the same day is not influenced by the effects of the diurnal cycle.

3 Method

3.1 Selection of predictor grid boxes

Ideally, logbook data from the equatorial Pacific would be used as this is the centre of action for ENSO. However, the lack of logbook data in this core region results in dependence on regions with strong ENSO teleconnections. To identify these regions, Pearson's correlation coefficients were calculated between the re-gridded, seasonal 'noon' zonal winds from ERA-Interim and the seasonal Southern Oscillation Index, 1979/80 to 2013/14. Both data sets were detrended prior to calculation of correlations. Grid boxes whose correlations were significant at the 90% significance level were considered as potential grid boxes suitable for use in the reconstructions. The correlation coefficients during the DJF season are shown in Fig. 3. Strong positive correlations are found during this season across much of the central and Eastern Indian Ocean, and the Maritime continent, a region with strong ENSO teleconnections (Aldrian and Susanto 2003). During El Niño events high pressure anomalies are typical over the Western Pacific and negative zonal wind anomalies are typical over the Maritime Continent (Wang and Fiedler 2006). The opposite is the case for La Niña events, with positive zonal wind anomalies typically in this region. These patterns lead to the large area of positive correlation between the SOI and zonal wind over the Maritime Continent and into the Indian Ocean. Logbook data from this region of strong teleconnections is used
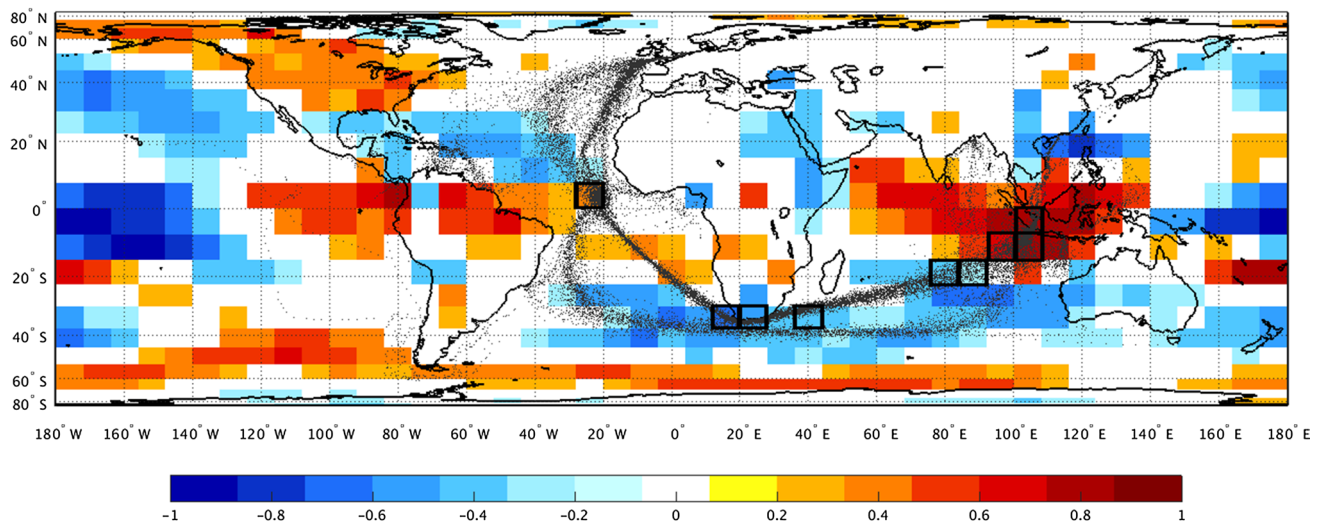
Fig. 3 Pearson's correlation coefficient between de-trended mean local noon DJF zonal wind and DJF Southern Oscillation Index, 1979/80–2013/14. Only those significant at 90% are shown. Grid boxes used in the reconstructions are outlined and are numbered 1 to 9 from West to East. *Black dots* indicate location of DJF logbook wind observations 1815/16–1853/54

for the reconstruction. A significant ENSO signal is also found within the zonal winds in other regions of the globe, which means that logbook data from more remote regions can also be exploited.

After analysis of regions with significant correlations, the main factor limiting use of individual grid boxes for reconstruction is the availability of wind data from logbooks, also indicated in Fig. 3. Following extensive analysis of the availability of DJF observations per grid box per season, the nine grid boxes indicated in Fig. 3 were selected for use in the reconstruction. These span the eastern and central Indian Ocean (grid boxes 5–9), around the coast of South Africa (grid boxes 2–4) and one from the mid-Atlantic (grid box 1). There are significant correlations between zonal winds and the SOI during DJF seasons in these grid boxes, with coefficients ranging from +0.86 to −0.60. The use of observations from different regions follows recent ENSO reconstructions which have used proxy data from multiple teleconnection regions to provide a more robust ENSO signal (Braganza et al. 2009).

The logbook data availability from these nine grid boxes during DJF seasons 1815/16 to 1853/54 are shown in Fig. 4, and back to 1750 in Supplementary Fig. 1 (S1). There is an average of 13 observations per grid box per season, however there is large interannual variability in the availability of data. In a given year, the amount of data across the nine grid boxes is also non-uniform. Prior to 1815 data availability in these grid boxes is reduced and missing data hinders reconstruction. There are only four seasons between 1750 and 1815 where data are available in all nine grid boxes (1788, 1791, 1802 and 1811), therefore the years after 1815 were the focus of the ENSO reconstructions, with SOI

reconstructions for these four additional years available in Supplementary Table 1 (Table S1).

In order to provide as continuous a historical reconstruction as possible, seasons in which individual grid boxes had less than three observations per season were infilled via longitudinal interpolation. Following methods of previous studies (Küttel et al. 2010; Hannaford et al. 2015), where available, wind observations from adjacent grid boxes which exceeded the selected number threshold were used. Interpolation was carried out longitudinally as the boxes to the east/west had better data availability than those to the north/south due to the historic shipping routes in the regions used. For example, for a given grid box if there were less than the minimum threshold of observations for the DJF season in a given year, then the DJF value from that year in the adjacent grid box which exceeds the threshold would be used, and the difference between the ERA-Interim reanalysis climatological means of the two grid boxes in the reanalysis applied, to gain a value for the grid box with low observations. If not enough observations were available, then the year was deemed missing.

3.2 Reconstruction and validation methods

Two common approaches to past climate reconstruction are composite-plus-scale (CPS) and climate field reconstruction (CFR) (Mann et al. 2005). Principal component regression (PCR) is a well-established method of CFR and is used in previous studies containing logbook data (Jones and Salmon 2005; Küttel et al. 2010; Hannaford et al. 2015). CPS and CFR are complementary methods, with CFR providing reconstruction of spatial patterns based
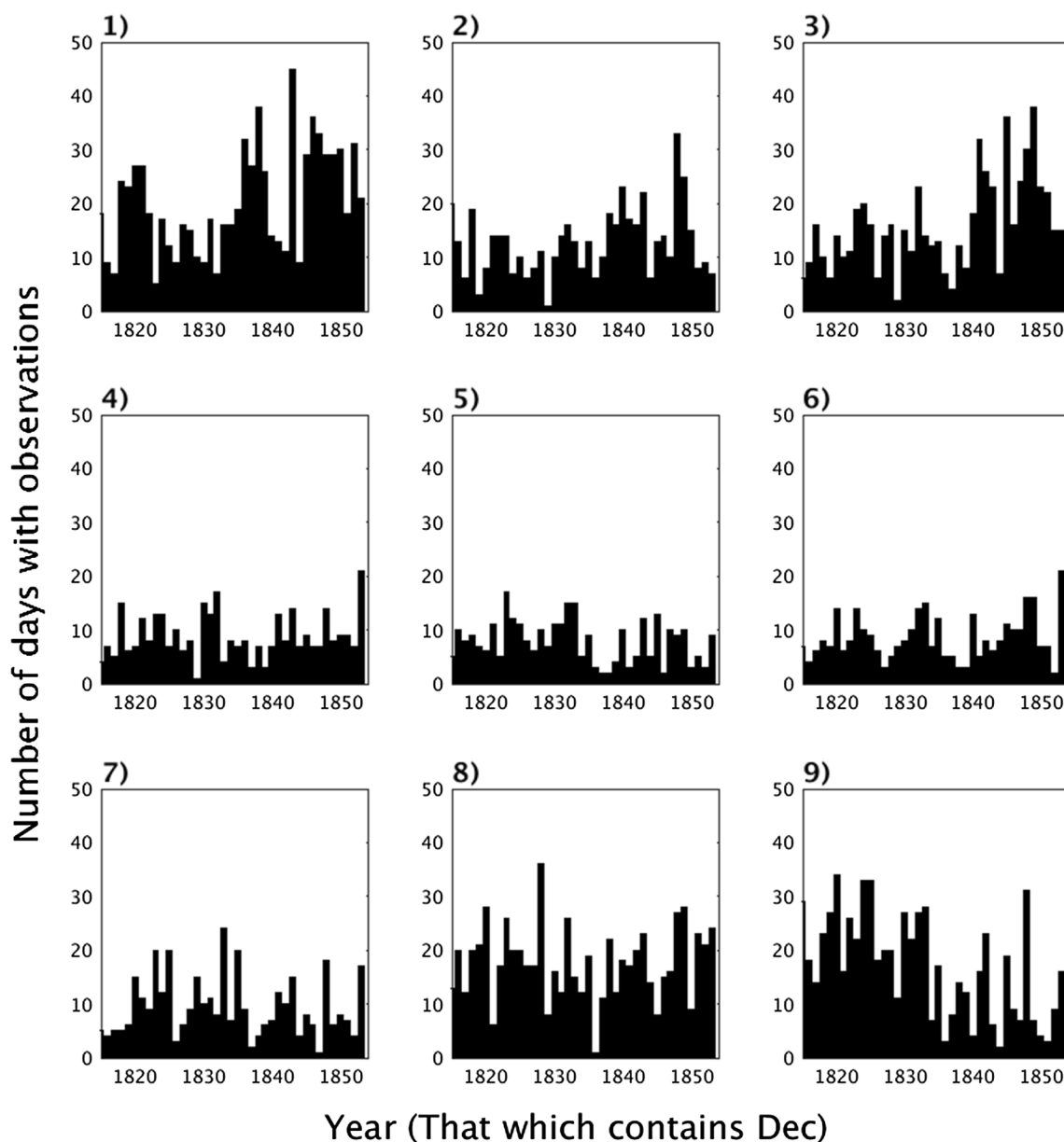
Fig. 4 Number of days with observations per DJF season in the nine predictor grid boxes, 1815/16–1853/54

on assumptions of stationarity between the predictor and large scale patterns, and CPS involving a simpler procedure based on local statistical relationships (Mann et al. 2008). Recent ENSO reconstructions have used both composite-plus-scale and climate field reconstruction methods (Wilson et al. 2010; Emile-Geay et al. 2013). Here we carry out reconstructions using both PCR and CPS methodologies (see Supplementary info for a detailed methodology).

For both methods, seasonal zonal wind anomalies were calculated using detrended, re-gridded ERA-Interim zonal wind, over the period 1979/80–2013/14. ERA-Interim zonal winds and DJF Southern Oscillation Index are detrended prior to model fitting in order to avoid an

influence from modern trends in the reconstruction method. However, after fitting and validation, the non-detrended data are used to produce the reconstruction. Uncertainties were calculated using ± 1.96 standard deviations of the residuals from the fitting as the 95% confidence intervals (PAGES 2 K Consortium 2013).

3.3 Assessing reconstruction skill

A number of statistics were used to assess the quality of the reconstruction models following the example of previous ENSO-reconstructions studies (Emile-Geay et al. 2013). Pearson's correlation coefficients and coefficient of

determination were calculated between the fitting reconstruction and the instrumental SOI. Cross validation was also carried out using the same period as the model fitting (1979–2013) to test the regression relationships (Wilks 2005). This method has been used in previous PCR-based reconstructions where the period over which regressions are derived is too short for other validation methods (Jones and Widmann 2003; Hannaford et al. 2015). Reduction of error (RE) skill scores were also calculated as an objective measure to assessing the performance of the statistical reconstruction models (Cook et al. 1994). RE scores range from −∞ to +1, where one suggests complete agreement with the predictand, zero indicates that the reconstruction is of same value as the climatology, and negative scores suggest no meaningful information can be gained from the reconstruction.

3.4 Threshold tests

Due to the limited availability of logbook data, seasonal averages must be calculated using only a restricted number of days of observations. Previous logbook studies have used different thresholds for defining the minimum number of observations that should be used in the calculation of seasonal averages. Küttel et al. (2010) and Neukom et al. (2013) only used seasons with observations for at least 3 days in each $8° \times 8°$ grid box, whereas Hannaford et al. (2015) used a threshold of at least one-tenth of the number of days in the season. The key aim is to find a balance between high spatial resolution and a good signal-to-noise ratio from the observations (Küttel et al. 2010). Here, we present a new methodology for quantifying the uncertainty of individual years within a historical reconstruction based upon the number of logbook observations available per season.

A minimum observation threshold test was carried out to assess the impact of using only a few observations per season from modern reanalysis data to calculate the seasonal means. These thresholds ranged from 1 day of observations per season per grid box up to 10 days of observations. This range was chosen as it covers both the threshold used by Küttel et al. (2010) and Neukom et al. (2013), and that used by Hannaford et al. (2015). At each threshold, the PCR and CPS methods were carried out 100 times over the fitting period using seasonal datasets composed of randomly selected daily mean values from ERA-Interim reanalysis data. The eigenvectors for PCR, and the correlation weights for CPS, were those calculated with the full seasonal data set, as these are what are used in the historical reconstructions. The seasonal mean was calculated using the dataset obtained from the selected threshold number of observations in each grid box. At each threshold (1 to 10), the mean correlation coefficients between the reconstructed SOI and the instrumental SOI, and the RE skill scores, were calculated.

The minimum data thresholds tests are based on a 'worst-case scenario', where each grid box contains only the selected minimum threshold number of observations. In reality the number of logbook observations per grid box is non-uniform (Fig. 4), with thresholds being exceeded in most of the grid boxes. Therefore, an additional test was carried out to directly replicate the data availability in each of the years of the logbook reconstruction. For each year of the logbook reconstruction, fitting and validation was carried out using the data availability for that given year. For example, the data availability in 1815/16 DJF season in the predictor grid boxes was replicated for all seasons throughout the fitting period (1979–2013), and the RE score calculated. Therefore, each year of the historical reconstruction has its own RE score assigned to it based on a fitting which replicates its data availability throughout a 35-year period. The RE scores obtained from this analysis were then used to assign a reconstruction skill to each of the years of the reconstruction, flagging up the years that need to be treated with more caution than others. This is a new method that can be applied to logbook reconstructions so as to provide an uncertainty value for each year within a historical reconstruction.

The initial threshold tests, which investigated the skill of the reconstruction methods using observations ranging from one per grid box per season up to 10 observations (see Supplementary info and Table S2 for detailed results), found that CPS performs better then PCR when a limited number of observations are available. In reality, the number of observations in each grid box is not the same in a given year and varies throughout the time period. Therefore, when replicating the actual data availability from the logbook years more directly applicable RE scores for the historical reconstructions are obtained. The RE value for the individual year relates to the skill of a reconstruction that could be carried out assuming the data availability during that given year is replicated over the entire reconstruction period, and the results are shown in Fig. 5.

When fully replicating the data availability of the period 1815/16–1853/54 during the fitting period the mean RE score obtained is 0.32 for the PCR method and 0.60 for CPS. The PCR RE scores are lower than the CPS RE scores for the same data availability, but all are positive except for 1819/1820. Any positive RE score is considered successful, making the difference between success and failure very small in some cases (Cook et al. 1994). Positive values closer to zero have less skill than those closer to one, however all are classed as 'skilful'. In order to account for the reduced skill of positive values in close proximity to zero, a threshold was selected, below which additional caution in the reconstruction should be taken. From the results
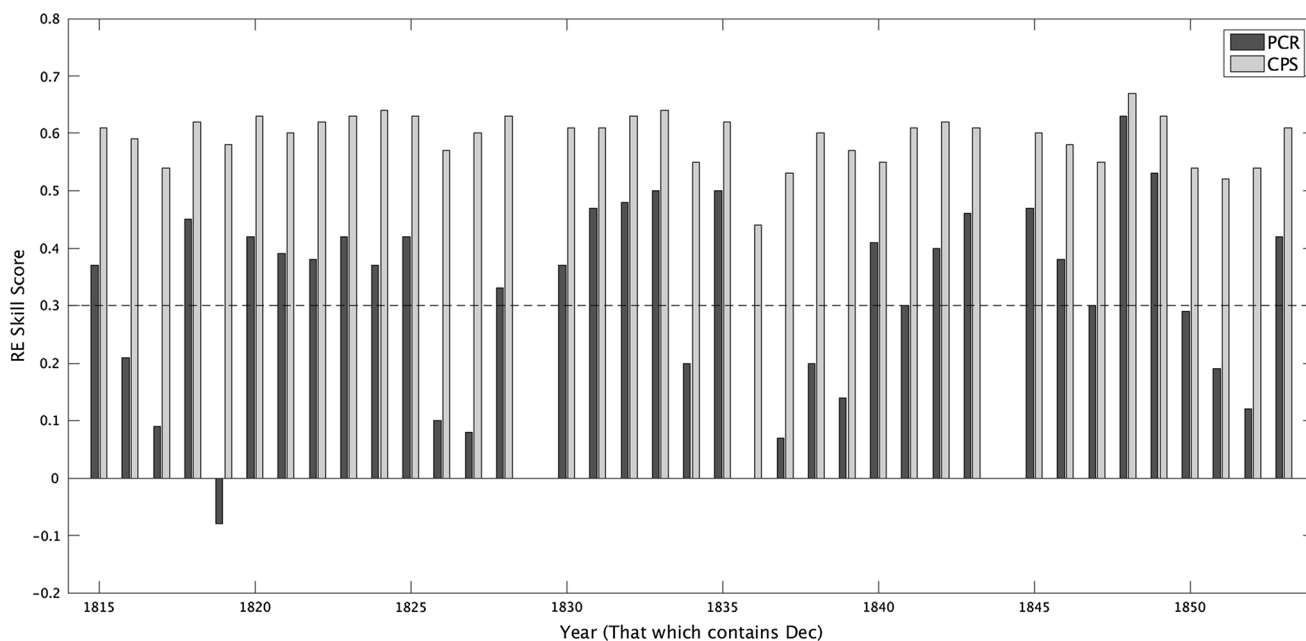
Fig. 5 RE skill scores obtained from PCR (*dark grey*) and CPS (*light grey*) reconstructions carried out over the fitting period, replicating the logbook data availability from each of the DJF seasons, 1815– 1854. 1829/30 and 1844/45 are missing data years. *Dashed line* indicates RE = 0.30

in Fig. 5, an RE score of less than 0.30 was selected as the value at which a year within the historical reconstruction which should be treated with caution. All of the CPS scores are higher than this threshold, whereas 13 of the PCR scores are lower than 0.3 and therefore are highlighted as years with additional uncertainty in the historical PCR reconstructions.

3.5 Data origin and wind force terms

Prior to carrying out the reconstructions, the wind data from the logbooks were analysed. Previous analysis of the wind force conversion terms highlighted that translations were carried out on a country by country basis, and that improvements could still be made on the calibration of wind force terms between countries (Koek and Können 2005). As a result of this, the country of origin of the observations used in this study has been analysed. The temporal coverage of observations from different countries varies throughout the CLIWOC period. Dutch ships dominate the later portion of the CLIWOC period and contributions from UK logbooks to CLIWOC end in 1829. The additional digitised EEIC observations only span from 1787 to 1834. The observations from the grid boxes used in the reconstructions have been divided into the countries from which the ships originate. Figure 6 shows the Dutch and EEIC contributions, which together make up 96% of the observations used during the period 1815–1854. An additional 3% of
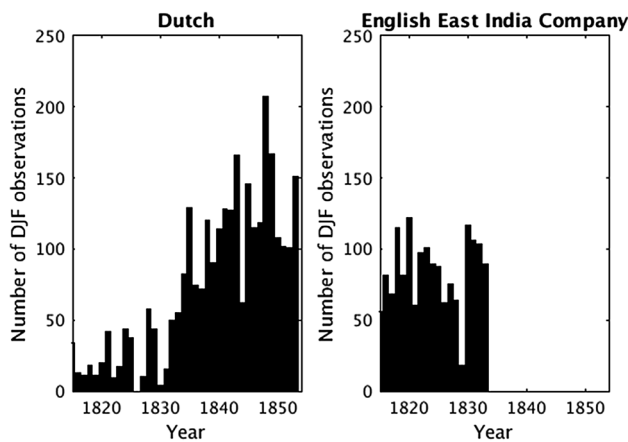

Fig. 6 The number of DJF wind observations from the regions used in the SOI reconstruction which come from Dutch and English East India Company during DJF seasons 1815/16–1853/54

observations are taken from UK CLIWOC ships and <1% from Spanish ships, which are not included in this section of analysis. The period 1815–1833 is dominated by EEIC observations (68%), whereas after 1833 all of the observations are from Dutch ships. This distinct shift in origins of the data source could be important when interpreting the resulting seasonal zonal winds and the reconstructed SOI.

To examine the differences in wind force terms used per country, the frequency of Beaufort Force equivalent wind terms used by the Dutch and EEIC ships during the period

of data overlap, 1815–1833, is assessed in Fig. 7. The data from the nine grid boxes were grouped together, based on similar climatology, into three sub-regions. These sub-regions consist of tropical grid boxes (1, 7, 8 and 9), those from the trade wind region (5 and 6) and grid boxes from around the southern tip of South Africa (2, 3 and 4). The Beaufort Scale ranges from 0 to 12, with higher numbers relating to stronger winds. Looking at both the Dutch and EEIC observations, grid boxes located in the tropics have a higher frequency of low Beaufort Force values, and some of the most extreme values are found in the grid boxes around southern Africa.

Although the Dutch observations come from a smaller sample size, the general distributions from the two countries can be compared. A main feature of the distributions is the almost complete lack of Beaufort Force 3 recorded in the EEIC data. It was used less than 0.01% in the EEIC observations, whereas it is the most common Beaufort Force found in the Dutch data, being used in 32% of the records. The EEIC observations are generally of higher wind forces than the Dutch data. A two sampled Kolmogorov–Smirnov test (Wilks 2011) was used in each of the three sub-regions to test the null hypothesis that the wind force data from the Dutch and EEIC ships come from the same distribution. In all three regions, the null hypothesis can be rejected at 1% significance level, indicating a statistically significant difference in the
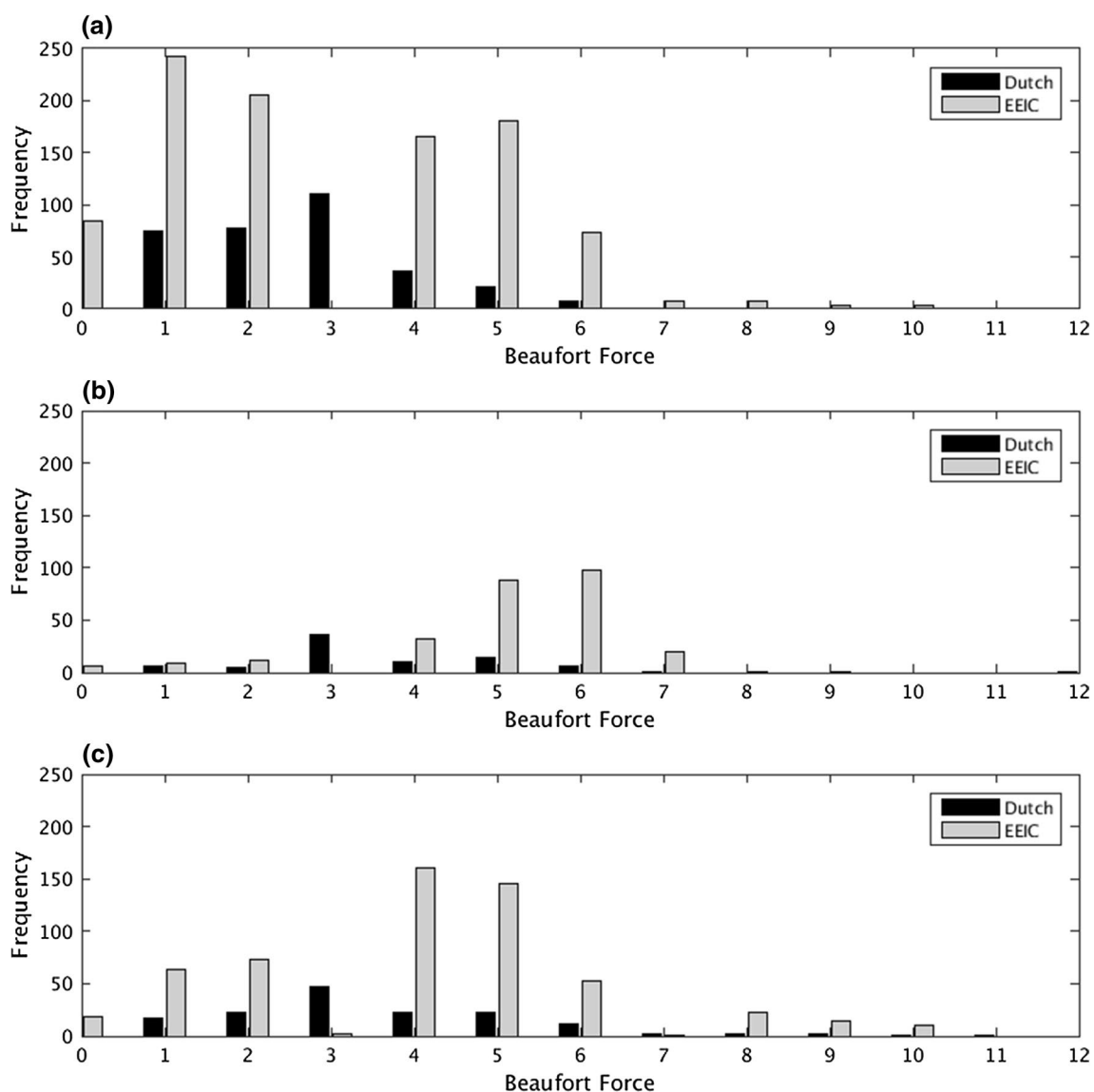

Fig. 7 Frequency of Beaufort-equivalent wind force terms used by the Dutch and EEIC in the **a** Tropics (grid boxes 1 and 7–9) **b** trade wind region (5–6) and **c** southern African grid boxes (2–4)

distributions of the wind force terms used by the Dutch and EEIC from within the same regions.

These preliminary investigations highlighted the difference in wind force terms used by the different countries, and the dominance of Dutch data in the later part of the record. As a result, this was taken into account during the ENSO reconstructions. Firstly, reconstructions were carried out using a climatological mean calculated over the entire period, 1815–1854 (known as variant 'A' later). Following this an alternative method was employed which used two separate means, one for the period dominated by British records, 1815–1833, and one for the later period, dominated by Dutch records, 1834–1854 (variant 'B'). This second method aims to address the impact of using data from countries with differences in the terminology for wind force observations.

4 Results

4.1 Reconstruction skill

The reconstruction skill of both the PCR and CPS methods were assessed over the fitting period, 1979–2013, using the re-gridded ERA-Interim data and instrumental Southern Oscillation Index for the DJF season. Figure 8 shows the SOI reconstructions along with the instrumental SOI for this period. Due to differences in the methodologies, the CPS method produces a normalised reconstruction, whereas the PCR method produces a reconstruction

with variability similar to the non-normalised instrumental record.

For the PCR method, Principal Components 1, 2 and 6 were retained due to their high correlation with the instrumental SOI. The coefficient of multiple determination during the fitting period is 0.89 (significant at the 1% level) (Fig. 8), thus 80% of the variance of the SOI is explained by the PCR model. The correlation between the PCR-reconstructed SOI and the detrended instrumental SOI from cross validation is 0.86, with an RE score of 0.73. For the CPS reconstruction, a correlation of 0.87 (significant at the 1% level) was found during the fitting period, thus 75% of the variance is explained by the CPS method (Fig. 7). For both methods, when comparing the non-detrended SOI with the reconstructed SOI, high RE values are obtained, with a RE score of 0.80 for the PCR method and 0.75 for the CPS method. Overall, the PCR method has slightly higher reconstruction skill than the CPS, however both perform well and are therefore both used to carry out historical reconstructions.

4.2 Modern wind patterns during strong El Niño and La Niña events

To test these modern reconstructions further and to see if the selection of predictor grid boxes is physically sensible, the wind patterns from some of the strongest ENSO events in this record were analysed. The DJF zonal wind anomalies in two recent strong El Niño events, 1982/83 and 1997/98, and two strong La Niña events, 1998/99
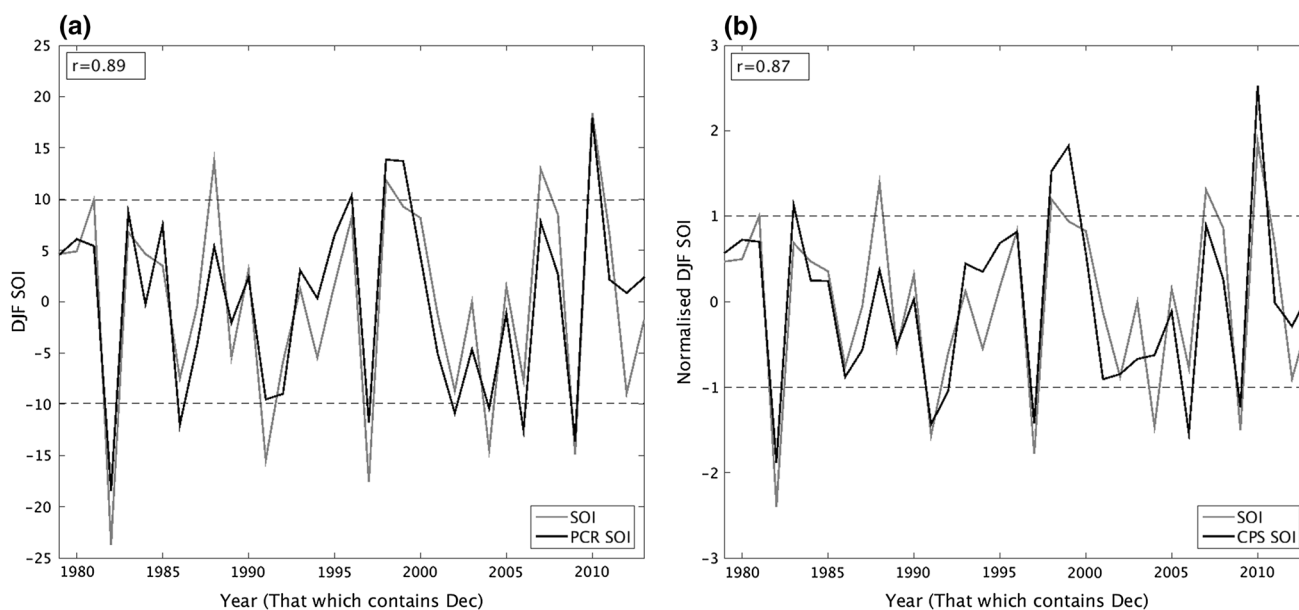

Fig. 8 Instrumental DJF Southern Oscillation Index and reconstructed DJF SOI using **a** principal component regression and **b** composite-plus-scale (normalised values), 1979/80–2013/14. *Dashed lines* represent ±1 standard deviation, the threshold used to define the main ENSO events

and 2010/11 are shown in Figs. 9 and 10, respectively. Although the 1982/83 and 1997/98 El Niño events were both strong events, the difference in the strength of the wind anomalies in predictor grid boxes is clear from Fig. 9. The negative anomalies in the Indian Ocean are stronger in 1997/98, whereas there is a much larger region of positive anomalies around the tip of Southern Africa in the 1982/83 event. Therefore, using grid boxes from both these key regions in the historical reconstructions increased the chance of identifying key anomalous years in the past. The same is true for La Niña events (Fig. 10), with much stronger positive anomalies in the
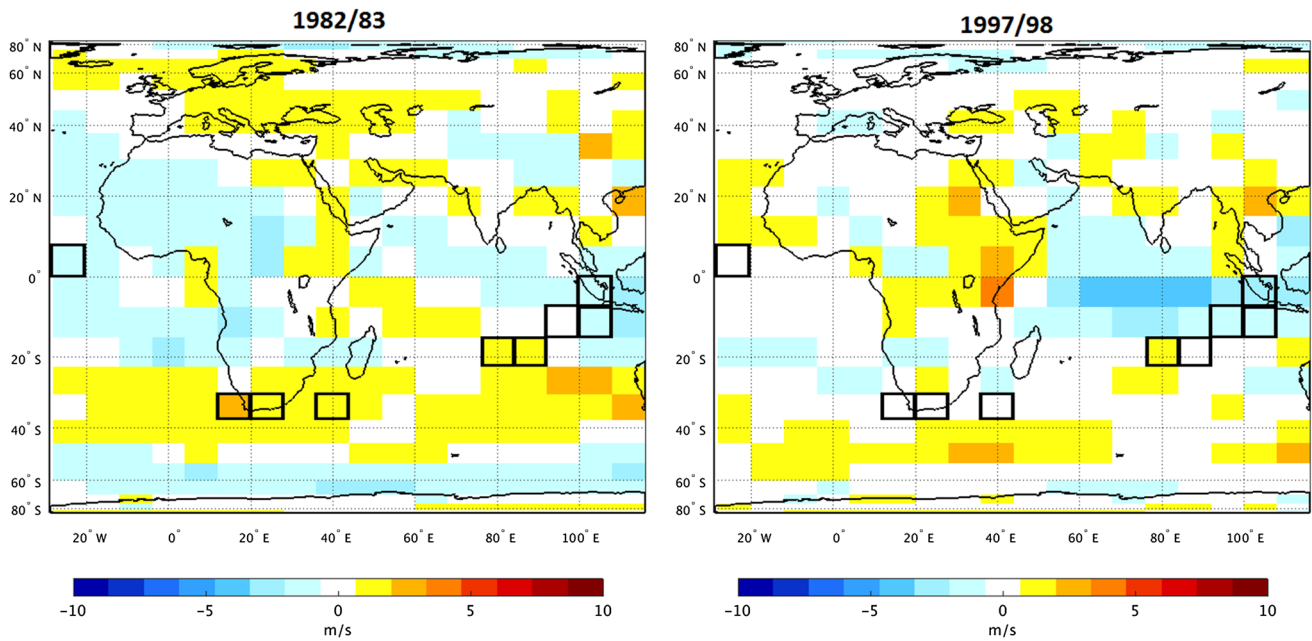

Fig. 9 DJF zonal wind anomalies (ms^{-1}) during El Niño events in 1982/83 (*left*) and 1997/98 (*right*), relative to 1979–2013 climatology. Predictor *grid boxes* used in the reconstruction are outlined in *black*
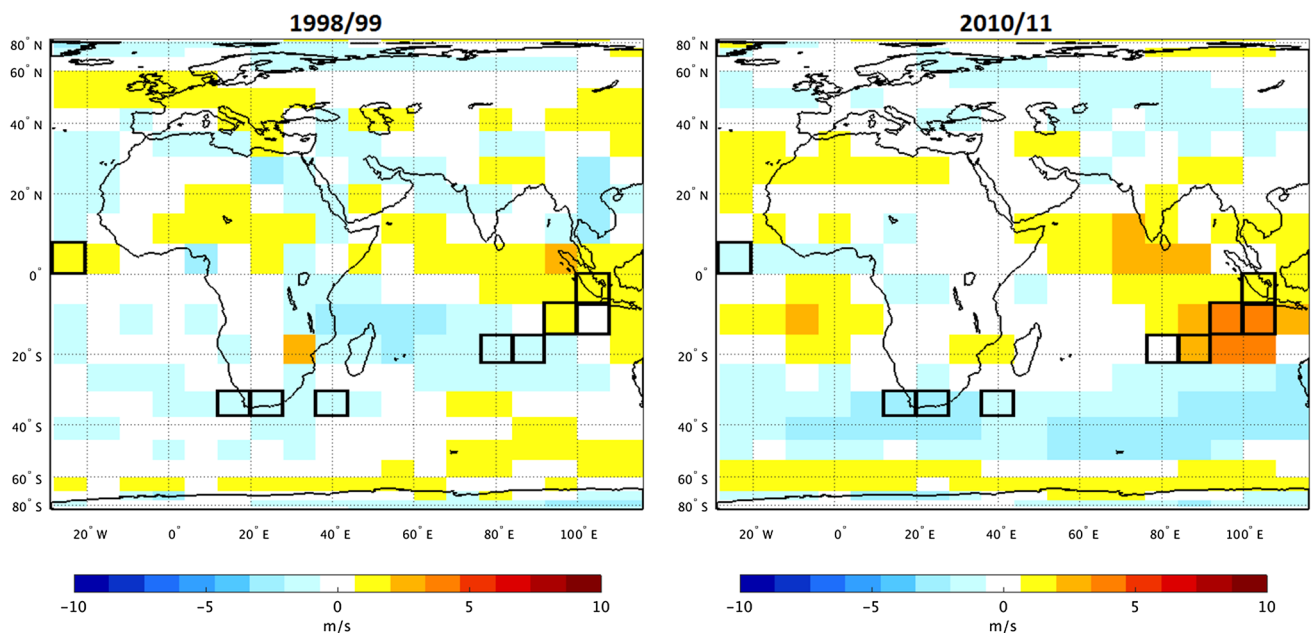

Fig. 10 DJF zonal wind anomalies (ms^{-1}) during La Niña events in 1998/99 (*left*) and 2010/11 (*right*), relative to 1979–2013 climatology. Predictor *grid boxes* used in the reconstruction are outlined in *black*

eastern Indian Ocean in the 2010/11 event, compared to the 1998/99 event.

Even though the expected anomalies are not found in all of the grid boxes used in the four events, the reconstructed SOI for these years show values very close to the true SOI (Fig. 8). In most cases, the sign of the wind anomalies in the predictor grid box fits with what is expected from the correlations indicated in Fig. 3. Therefore, the use of anomalies from different regions is beneficial for the reconstruction approach and helps provide a more robust ENSO signal.

4.3 Historical ENSO reconstructions

Following, fitting and validation of both reconstruction methods, historical reconstructions were carried out, using the model fitted over the whole period. Figure 11 shows the DJF SOI reconstructions using PCR and CPS over 1815–1854. Variant A reconstructions, (PCR A and CPS A) were calculated using a climatological mean from the entire period, 1815–1853, whereas, variant B (PCR B and CPS B) were calculated using two climatological values, one covering 1815 to 1833 and one for the later period 1834 to 1854. This takes into account the change in data source from mostly EEIC in the earlier period to all Dutch records after 1833. Years with RE scores less

than 0.3 are indicated with the historical PCR reconstructions (light grey bars), taking into account the results of the minimum data availability threshold testing.

Table 1 shows the correlation coefficients between the four historical reconstructions. The four reconstructions are all significantly correlated and show a high degree of similarity to each other. The strongest correlations are found between the two CPS reconstructions, $r = 0.88$, followed by the two PCR reconstructions, $r = 0.86$. However, correlations between the reconstructions using the different methods remain high, suggesting good agreement between the two reconstruction methodologies. Within the PCR reconstructions, some of the more

Table 1 Pearson's correlation coefficients between historical reconstructions of DJF SOI 1815/16–1853/54 using principal component regression (PCR) and composite-plus-scale (CPS) (normalised), with (a) the same mean throughout and (b) using 1815–1833 and 1834–1853 means

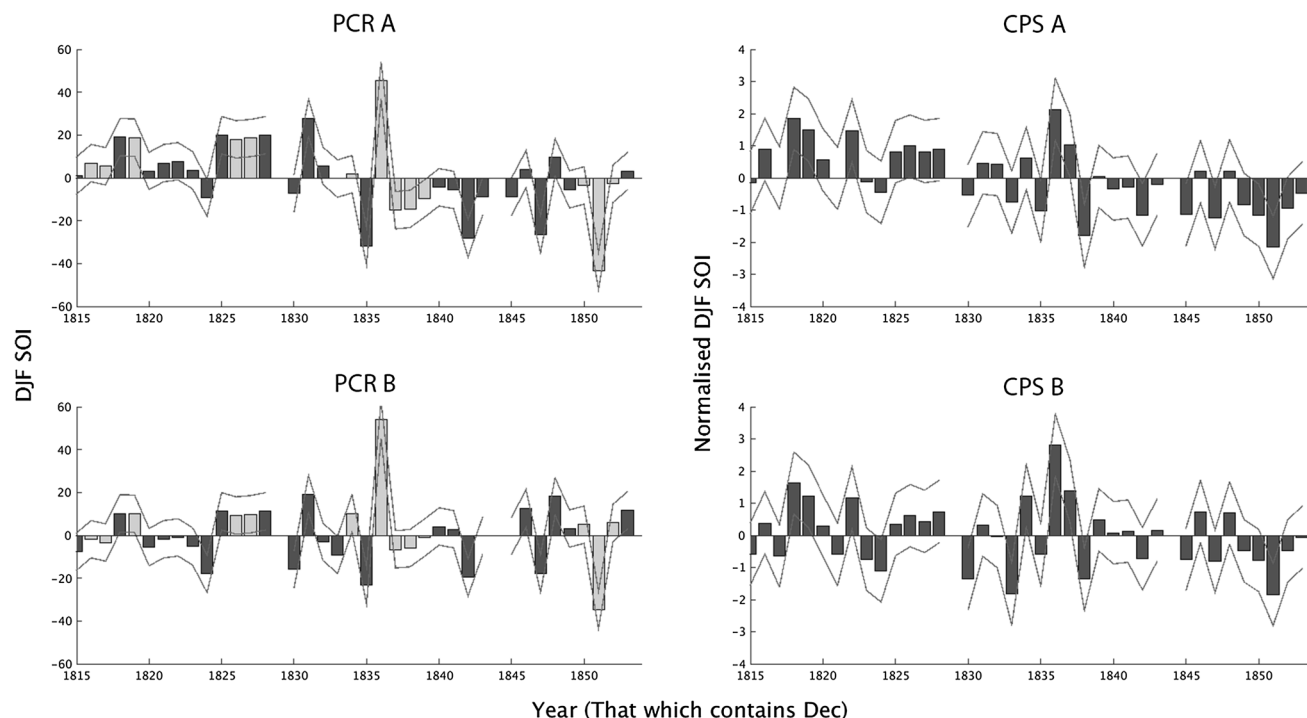|	PCR A	PCR B	CPS A	CPS B
PCR A	–			
PCR B	0.86	–		
CPS A	0.81	0.65	–	
CPS B	0.67	0.75	0.88	–

All significant at 1% level

Fig. 11 Reconstructions of DJF SOI 1815/16–1853/54 using principal component regression (PCR) and composite-plus-scale (CPS) (normalised), with **a** the same mean throughout and **b** using 1815–1833 and 1834–1853 means. Thin lines show the 95% confidence intervals. *Light grey bars* indicate years with RE scores less than 0.3. No data for 1829/30 and 1844/45

extreme events are those with additional uncertainty (i.e. RE < 0.3). However, these events are still classified as main ENSO events in the CPS reconstructions, where reconstruction skill is higher. Therefore, similar signals are being provided by the two methods.

In order to assess the differences between the reconstructions, additional statistical analysis was performed. Least squares fitting was carried out to identify any statistically significant trends in the four reconstructions. A negative trend in SOI over the reconstruction period is suggested from Fig. 11, with the SOI generally decreasing during the period 1815/16 to 1853/54. Results from linear regression for the four reconstructions found that a statistically significant ($p < 0.05$) negative trend is found when only one climatology is used throughout the entire period, for both the PCR and CPS methods (Table S3). When variant B is used, to take into account the change in dominance from EEIC to Dutch observations, this significant linear trend is no longer present in either of the statistical reconstructions.

Analysis of the mean and two-sampled t tests was carried out to assess the difference in SOI before and after 1833/34 from each of the four reconstructions (Table S4). It was found that the reconstructions without the adjustments in the mean (Variant A) have significantly different SOI in the earlier period compared to the later period. In contrast, for those reconstructions where adjustments to the mean were made (Variant B), this statistically significant difference was not found. Therefore, the impact of using logbook data from two periods dominated by countries which use different wind force terminology is significant and should be addressed.

To identify the main ENSO events from these reconstructions, El Niño events were classified as seasons with a negative SOI value more than one standard deviation from the mean, and La Niña as a positive SOI more than one standard deviation from the mean (Stahle et al. 1998). Figure 12 shows DJF seasons classified as El Niño and La Niña in the four reconstructions. Two events are found in all four of the reconstructions and have the most extreme

SOI values, an El Niño event in 1851/52, and a La Niña event in 1836/37. These are named as high confidence events. Attention is then given to those ENSO events found in at least three of the four logbook reconstructions, which are seen as medium confidence events. Between 1815/16 and 1853/54, five additional events fall into this category; two La Niña events 1818/19 and 1819/20, and three El Niño events 1835/36, 1842/43 and 1847/48. Taking into account the significantly more positive SOI values found in the reconstructions of variant A prior to 1834, the criteria for medium confidence events is relaxed to only the two reconstructions using variant B needed to classify El Niño events prior to 1834. This results in two additional El Niño events, 1824/25 and 1830/31. Grouping together these high and medium confidence events, there are six main El Niño events and three main La Niña events during the logbook reconstruction period.

The zonal wind anomalies in the predictor grid boxes in the two high confidence events are shown in Fig. 13. There is a clear difference in the signal between El Niño, 1851/52 and La Niña, 1836/37. In 1851/52, negative zonal wind anomalies are found in the three grid boxes closest to Indonesia. This negative wind anomaly was also apparent in the more recent El Niño 1997/98 and in a similar region in 1982/83 (Fig. 9). In contrast, strong positive zonal wind anomalies are found in this region during the historical La Niña 1836/37, as well as modern events in 2010/11 and 1998/99 (Fig. 10).

The El Niño signal from the grid boxes around the southern tip of Africa is of positive wind anomalies, which is seen more strongly in the 1982/83, modern event than the 1997/98 event, however such a signal fits with that suggested from the modern correlation coefficients between zonal wind and SOI in this region (Fig. 3). The signal from these grid boxes is less clear, in the 1836/37 La Niña event, as negative anomalies would be expected. However, this raises the issue of inter-event diversity of ENSO, which is returned to in Part II (Barrett et al. 2016). The spatial pattern of wind anomalies during

Fig. 12 Frequency of El Niño (negative SOI events) and La Niña (positive SOI events) in logbook reconstructions, 1815/16–1853/54
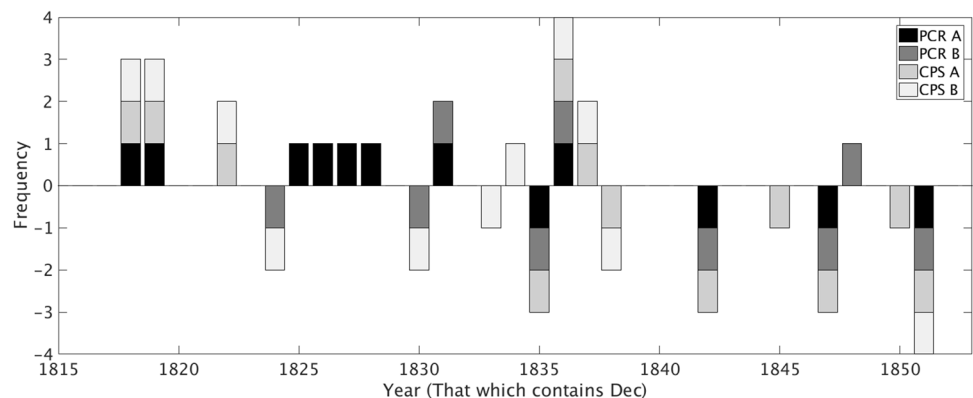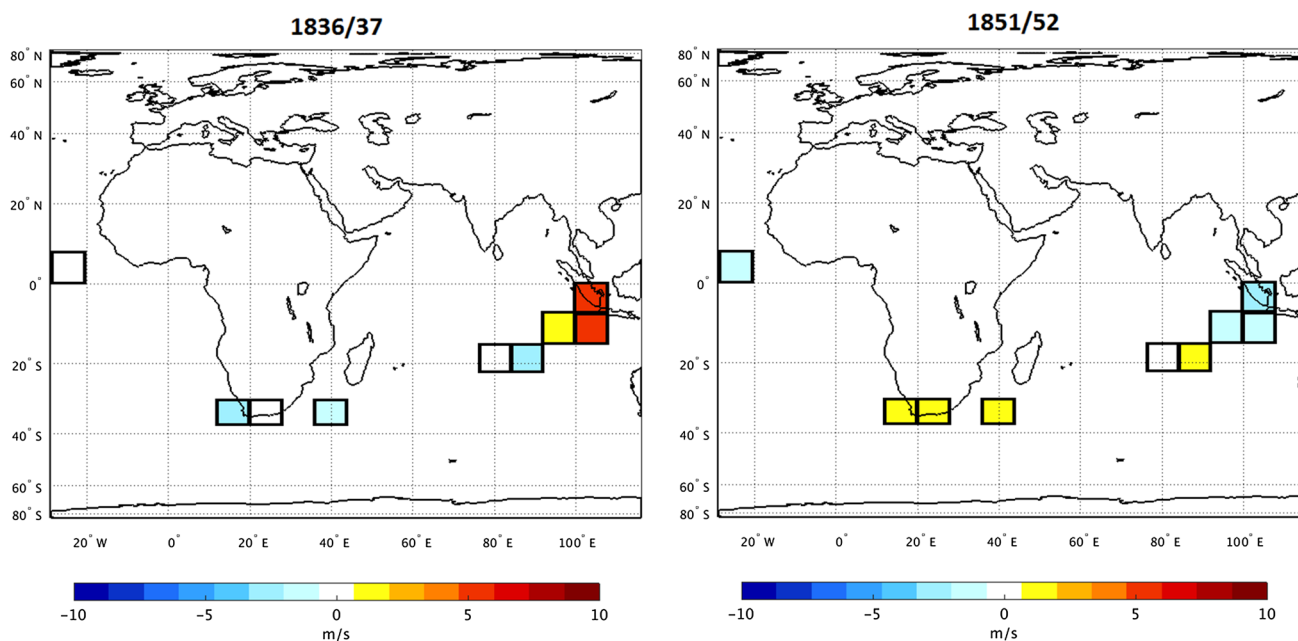
Fig. 13 La Niña, 1836/37, (*left*) and El Niño, 1851/52, (*right*) zonal wind anomalies (ms^{-1}), relative to 1834/35–1853/54 DJF climatology

different events are not always the same, therefore using data from various regions helps to identify different signals. However, overall similarities are found in the typical spatial wind patterns from both historical and modern events.

Additional caution is needed when analysing those years with PCR RE scores less than 0.3. This additional level of uncertainty will be discussed further in Barrett et al. (2016). The limited logbook data availability from these 2 years resulted in RE scores less than 0.3 when using the PCR method, and are therefore flagged as more uncertain years within the reconstruction. However, the CPS reconstructions in these years also indicate strong ENSO events, which supports the classification of these events within the historical record.

5 Discussion

In the companion paper (Barrett et al. 2016) we examine how our new logbook-based SOI reconstructions compare to previous proxy and documentary ENSO reconstructions, and discuss the implications of this comparison for the nature of ENSO reconstruction methods. However, here we discuss the limitations and strengths of our methodology, and also compare our reconstruction with a previous SOI reconstruction which is also based on direct meteorological measurements (Können et al. 1998).

5.1 Difference in wind force terminology

An important finding of this paper is the statistically significant difference between the distributions of wind force terms used by the Dutch and EEIC ships. A shift from using observations from mostly EEIC to entirely Dutch records during the period of reconstruction is found, because of the use of both EEIC and CLIWOC data, but in part influenced by the spatial domain. The majority of previous logbook studies focus on the North Atlantic and only use CLIWOC data (Gallego et al. 2005; Küttel et al. 2010). Hannaford et al. (2015) used both CLIWOC and EEIC data from grid boxes around southern African region, but no clear shift was identified in their study.

The major difference in the distributions of wind force terms between the Dutch and EEIC observations is the very limited appearance of Beaufort Force 3 (BF3) in the British records but the common use of it in the Dutch records. Previously, Koek and Können (2005) highlighted a difference in the frequency of Beaufort force values used by different CLIWOC countries, especially in the range of Force 3 to 5. Beaufort Force 3 is described as a 'Gentle Breeze' under current British classification, and the CLIWOC multi-lingual dictionary found the following historic terms to be representative of equivalent wind force: 'gentle breeze', 'gentle gale', 'gentle trade', feint gale', 'light gale' and 'easy gale', with the latter two uncommon after 1700 (Garcia-Herrera et al. 2003). Of the 22,000 British wind force terms analysed by CLIWOC, less than 0.5% were BF3 equivalent

terms (Wheeler and Wilkinson 2005). Of the observations used in the SOI reconstruction, BF3 was only recorded twice in the British wind force records. The reason for this limited use of BF3 by the British ships is unknown. As the Beaufort terms are conversions of descriptive terms by CLIWOC, it may either be because BF3 equivalents were under recorded in the original logbook observations, or that BF3 equivalents were under-converted by CLIWOC. Determination of which of these is true, is outside of the scope of this paper. In contrast to the British records, Koek and Können (2005) found that Beaufort Force 3 is over-represented in the Dutch records, and this is supported in the results of this paper. The Beaufort Force 3 equivalent terms in Dutch all include the term 'bramzeilskoelte' which refers to the topgallant sail, the upper-most sail that could be used at this wind speed (Koek and Können 2005).

It was also found that in general, the Beaufort Force terms used by the EEIC were higher than those used by the Dutch within the same geographical regions over the same time period. One reason for this is that the method of describing wind force used by the Dutch was distinctly different than that used by EEIC and other countries from within the CLIWOC database. The Dutch method of wind force observations relied heavily on the wind's influence on the sails used, which provides a more explicit means of determining the order of these terms, making them more refined than is possible in the other languages (Garcia-Herrera et al. 2003). In contrast the British relied more on the state of the sea to describe wind force (Wheeler and Wilkinson 2005). This different approach to wind determination is suggested as the likely reason for the differences in the distribution of wind force terms used.

Another reason for this difference could be that some of the Dutch CLIWOC data post 1826 are taken from 'Extract Logbooks', meteorological summaries of original logbook data. These Extract Logs were not made until the 1860s, after the adoption of the Beaufort scale in the Netherlands (Koek and Können 2005). The conversions from wind force terms to Beaufort scales were carried out at the time the Extract Logs were created, and these translations were found to be of better quality than those done by the CLIWOC team (Koek and Können 2005). Therefore, the wind force terms taken from these logbooks could be more reliable.

The implications of higher wind force terms from EEIC than Dutch records is a bias towards stronger winds in the period dominated by EEIC and weaker winds in the period of Dutch observations. A tendency for stronger winds in the first half of the record could promote an increased frequency of La Niña events in the first part of the reconstruction, as during La Niña events stronger westerly winds are typical in the predictor grid boxes closest to the Maritime continent. This was found in all of the historical reconstructions carried out, but more so in those in which the difference in wind force observations was not taken into account. Seven of the eight La Niña events found in the PCR A reconstruction were prior to 1831, thus within the period dominated by stronger EEIC observations and all four of the El Niño events in this reconstruction were post 1835.

The impact of relying solely on Dutch logbooks post 1834 has not been investigated in previous papers which carried out climate reconstructions using logbooks over this period. Wheeler (2005) carried out an assessment of the coherency of wind observations between vessels sailing in convoy and found good agreement. However, a major limitation in this analysis was the tendency for ships in convoy to be from the same country rather than from a mix of nationalities and therefore a mix of recording approaches. The difference between wind observations from different countries has been highlighted here as a potential source of inconsistency and an area that requires further investigation. Less than 5% of British logbooks were exploited by CLIWOC and none after 1828, therefore a vast amount of records are available for future digitisation, as outlined in a recent survey of logbooks in UK archives (Wilkinson 2009). Digitisation of these British records would result in the ability to carry out reconstructions using data from solely British sources. An analysis of wind force terms and reconstructions using only British observations would enable a fuller assessment of the impact of using a split of Dutch and British records.

5.2 Difference in reconstruction methods: PCR and CPS

Reconstructions using both PCR and CPS were carried out to assess the impact of using different methodologies on the resulting historical reconstructions. Recent multi-proxy ENSO studies employed both techniques and found no major differences between the ENSO reconstructions found using CPS and PCR (Wilson et al. 2010). The two methods can be seen as complimentary approaches (Mann et al. 2008). Both methods assume stationarity in the statistical relationships over time, a common limitation in studies of the past (Gergis et al. 2006). Results from this paper suggest good agreement between the two approaches and both provided good reconstruction skill over the fitting period. Nevertheless, it is worth pointing out that the PCR method has slightly higher skill during the fitting period. Although correlations are strong between reconstructions, the ENSO event chronologies differ significantly between reconstructions. Only two ENSO events of high confidence were found: La Niña 1836/37 and El Niño 1851/52. An additional, seven medium confidence events were found: five El Niño (1824/25, 1830/31, 1835/36, 1842/43 and 1847/48)

and two La Niña events (1818/19 and 1819/20). A companion paper (Barrett et al. 2016) will analyse the discrepancies in these events-based chronologies further.

5.3 Threshold tests on number of observations per season

A new method of threshold testing was carried out to assess the reconstruction skill using a limited number of observations. Results showed that, as would be expected, the higher the number of observations used per season, the better the reconstruction skill. Nevertheless, these tests demonstrated that even with very few observations reasonable reconstruction skill can be obtained, with the CPS method performing better than PCR with very limited data availability. Additional analysis which replicated the logbook data available in the predictor grid boxes used in the historical reconstruction, enabled the reconstruction skill in each year to be evaluated, and identified years within the PCR reconstructions where reconstruction skill is lower than $RE = 0.3$ (Fig. 5). It is noteworthy, however, that the CPR reconstruction for all years had an RE significantly greater than 0.3 (Fig. 5). Overall, the threshold results showed that valuable reconstructions can be obtained even when using limited data sources. However, the benefit of further digitisation to increase data density of logbook wind observations is clear, as reconstruction skill can be increased if the number of observations per season can be increased.

5.4 Comparison to Jakarta based record of the SOI

While Part II, focuses on a comprehensive comparison with previous ENSO reconstructions, here we present a comparison of one of the logbook SOI reconstructions (CPS B) with the Jakarta rain-day SOI (Können et al. 1998; and single site pressure SOI, Fig S2; Table S5), as both are from direct observations and can be used together to obtain a more robust ENSO signal from this period. The Jakarta-rain day SOI spans the Indonesian dry season, June–November and the SOI was calculated from a continuous road maintenance record of dry days in Jakarta, 1830–1850. CPS B is focused on for comparison, as the CPS methodology has the higher skill when using sparse data availability (Fig. 5) and variant B accounts for the difference in wind force terms between the earlier and later part of the reconstruction period. Figure 14 shows the logbook CPS B DJF reconstruction along with the Jakarta rain-day SOI, which are significantly correlated (95% significance level). Over the period 1830 to 1850 the strongest El Niño event in CPS B is during 1833/34. The Jakarta rain-day SOI also indicates 1833 to be the most extreme negative SOI (El Niño) during this period. The strongest La Niña in the CPS B reconstruction during this period is in
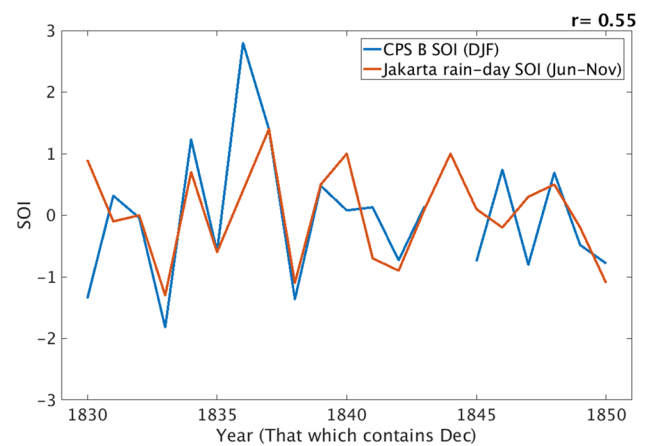

Fig. 14 CPS B logbook DJF SOI (1830/31–1850/51) and Jakarta rain-day June–Nov SOI (1830–1850). Pearson's correlation coefficient (r) shown in top right

1836/37 (and extends into 1837/38). In 1837, the strongest positive SOI (La Niña) is found in the Jakarta rain-day SOI. Therefore, there is a good match between the two most extreme ENSO events during this period. The strong agreement between these records suggests that there is a robust signal from these early instrumental based reconstructions of SOI, giving additional confidence to our logbook reconstructions.

6 Conclusions

Here we present the development of ENSO reconstructions using wind observations from ships' logbooks. Temperature and pressure data from ships' logbooks provide an additional source of data, but are also subject to data availability issues, and additional sources of measurement uncertainty. Therefore, they are not used in the reconstruction approach presented here. Thus, we have built upon Jones and Salmon's (2005) preliminary logbook SOI reconstruction by using a larger amount of data, a focus on regions with strong ENSO teleconnections and an analysis of multiple methods of reconstructions. Two different reconstruction methodologies were used, PCR and CPS and both were found to be suitable methods for reconstruction ENSO using data from ships' logbooks.

We developed methodologies to assess the impact of data availability on reconstruction quality. We found that good reconstruction skill can be obtained with limited data availability. As expected, the reconstruction skill increases with the number of observations available. However, it was found that statistically significant skill can be obtained when using surprisingly few observations per season. We expect that these results will be region specific, reflecting

the ability of few observations to capture seasonal climate in a given region, and we recommend that this analysis is repeated by researchers developing reconstructions using logbook data from other geographical regions.

We identified a statistically significant difference in the mean and distribution of the basic Dutch and EEIC wind force observations, which leads to a potentially spurious trend in the reconstructions. We address this by splitting the study period into two, an earlier period dominated by British records (1815/16–1833–34) and a later period from which only Dutch observations are available (1834/35–1853/54). For these two periods, different climatological values were calculated and applied to the reconstruction methods, thus producing two reconstructions for each statistical reconstruction method. The resulting historical SOI time series had strong correlations with each other, however some differences were found when converting this into an events-based chronology. The most robust CPS B logbook reconstruction compares well to the Jakarta rain-day SOI, (Können et al. 1998) which spans much of the reconstruction period. The strong agreement between these two records indicate a common signal from direct observations, which together can be used to obtain a more robust understanding of ENSO during this period. In Barrett et al. (2016) we compare both sets of reconstructions against existing reconstructions, to determine which method produces a reconstruction with a higher agreement to existing reconstructions.

Digitisation of additional logbook records would enable a higher number of observations to be used. The SOI reconstruction skill largely depends on the data availability during a given year, as it varies between grid boxes and seasons. Therefore, the number of observations needed to obtain good reconstruction skill may be different in other seasons and/or geographical regions with different climatological variability, and should therefore be assessed on an individual basis. Fig. S1 shows, for example, that the current SOI reconstruction could be greatly extended in temporal length, in principal back to 1750, with targeted additional digitisation of existing logbooks. There thus remains a potentially rich source of marine records for studies of past climate variability.

Compliance with ethical standards

Conflict of interest　The authors declare that they have no conflict of interest.

References

Aldrian E, Susanto DR (2003) Identification of three dominant rainfall regions within Indonesia and their relationship to sea surface temperature. Int J Climatol 23:1435–1452

Barrett HG, Jones JM, Bigg GR (2016) Reconstructing El Niño Southern Oscillation using data from ships' logbooks, 1815–1854, Part II: comparisons with existing ENSO reconstructions and implications. Clim Dyn, submitted

Barriopedro D, Gallego D, Alvarez-Castro MC, García-Herrera R, Wheeler D, Peña-Ortiz C, Barbosa, SM (2014) Witnessing North Atlantic westerlies variability from ships logbooks (1685–2008). Clim Dynam 43:939–955

Bellenger H, Guilyardi É, Leloup J, Lengaigne M, Vialard J (2014) ENSO representation in climate models: from CMIP3 to CMIP5. Clim Dynam 42(7–8):1999–2018

Braganza K, Gergis JL, Power SB, Risbey JS, Fowler AM (2009) A multiproxy index of the El Niño-Southern Oscillation, A.D. 1525–1982. J Geophys Res 114:D05106

Brohan P, Allan R, Freeman E, Wheeler D, Wilkinson C, Williamson F (2012) Constraining the temperature history of the past millennium using early instrumental observations. Clim Past 8:1551–1563

Bunge L, Clarke AJ (2009) A verified estimation of the El Niño Index Niño-3.4 since 1877. J Clim 22(14):3979–3992

Bureau of Meteorology (2015) Southern Oscillation Index (SOI) since 1876. http://www.bom.gov.au/climate/current/soihtm1.shtml. Accessed 21 Oct 2015

Catchpole AJW, Faurer MA (1985) Ships' log-books, sea ice and the cold summer of 1816 in Hudson Bay and Its approaches. Arctic 38(2):121–128

Climate Data Operators (CDO) https://code.zmaw.de/projects/cdo

Cook ER, Briffa KR, Jones PD (1994) Spatial regression methods in dendroclimatology—a review and comparison of two techniques. Int J Climatol 14(4):379–402

Dee DP et al (2011) The ERA-Interim reanalysis: configuration and performance of the data assimilation system. Q J Roy Meteor Soc 137:553–597

Diaz HF, Markgraf V (1992) El Niño: historical and paleoclimatic aspects of the Southern Oscillation. Cambridge University Press, Cambridge

Emile-Geay J, Cobb KM, Mann ME, Wittenberg AT (2013) Estimating central equatorial pacific SST variability over the past millennium. Part I: methodology and validation. J Clim 26:2302–2328

Gallego D, Garcia-Herrera R, Ribera P, Jones PD (2005) Seasonal mean pressure reconstruction for the North Atlantic (1750–1850) based on early marine data. Clim Past 1:19–33

García-Herrera R, Wheeler D, Können GP, Koek FB, Prieto MR (2003) CLIWOC multilingual meteorological dictionary, an English-Spanish-Dutch-French Dictionary of wind terms used by Mariners from 1750 to 1850, KNMI publication 205

García-Herrera R, Wilkinson C, Koek FB, Prieto MR, Calvo N, Hernandez E (2005) Description and general background to ships' logbooks as a source of climatic data. Clim Change 73:13–36

Gergis J, Braganza K, Fowler A, Mooney S, Risbey J (2006) Reconstructing El Niño- Southern Oscillation (ENSO) from high-resolution palaeoarchives. J Quat Sci 21(7):707–722

Hannaford M, Jones JM, Bigg GR (2015) Early-nineteenth-century southern African precipitation reconstructions from ships' log-books. Holocene 25(2):379–390

Jones PD, Salmon M (2005) Preliminary reconstructions of North Atlantic Oscillation and Southern Oscillation Index measures of wind strength and directions taken during the CLIWOC period. Clim Change 73:131–154

Jones JM, Widmann M (2003) Instrument- and tree-ring-based estimated of the Antarctic Oscillation. J Clim 16:3511–3523

Koek FB, Können GP (2005) Determination of wind force and present weather terms: the Dutch Case. Clim Change 73:79–95

Können GP, Koek FB (2005) Description of the CLIWOC database. Clim Change 73(1–2):117–130

Können GP, Jones PD, Kaltofen MP, Allan RH (1998) Pre-1866 extensions of the Southern Oscillation Index using early Indonesian and Tahitian meteorological readings. J Clim 11:2325–2339

Küttel M, Xoplaki E, Gallego D, Luterbacher J, García-Herrera R, Allan R, Barriendos M, Jones P, Wheeler D, Wanner H (2010) The importance of ship log data: reconstructing North Atlantic, European and Mediterranean sea level pressure fields back to 1750. Clim Dynam 34:1115–1128

Mann ME, Rutherford S, Wahl E, Ammann C (2005) Testing the fidelity of methods used in proxy-based reconstructions of past climate. J Clim 18:4097–4107

Mann ME, Zhang Z, Hughes MK, Bradley RS, Miller SK, Rutherford S, Ni F (2008) Proxy-based reconstructions of hemispheric and global surface temperature variations over the past two millennia. PNAS 105(36):13252–13257

Neukom R, Nash DJ, Endfield GH, Grad SW, Grove CA, Kelso C, Vogel CH, Zinke J (2013) Multi-proxy summer and winter precipitation reconstructions for southern Africa over the last 200 years. Clim Dynam 42(9–10):2713–2726

PAGES 2 K consortium (2013) Continental-scale temperature variability during the past two millennia. Nature Geosci 6:339–346

Ropelewski CF, Jones PD (1987) An extension of the Tahiti–Darwin Southern Oscillation Index. Mon Weather Rev 115:2161–2165

Stahle DW, D'Arrigo RD, Krusic PJ, Cleaveland MK, Cook ER, Allan RJ, Cole JE, Dunbar RB, Therrell MD, Gay DA, Moore MD, Stokes MA, Burns BT, Villanueva-Diaz J, Thompson LG (1998) Experimental dendroclimatic reconstruction of the Southern Oscillation. Bull Am Meteorol Soc 79:2137–2152

Wang C, Fiedler PC (2006) ENSO variability and the eastern tropical Pacific: a review. Progr Oceanogr 69:239–266

Wheeler D (2001) The weather of the European Atlantic seaboard during October 1805: an exercise in climatology. Clim Change 48:316–385

Wheeler D (2005) An examination of the accuracy and consistency of ships' logbook weather observations and records. Clim Change 73:97–116

Wheeler D (2014) Hubert Lamb's treasure trove – ships logbooks in climate research. Weather 69:133–139

Wheeler D, Wilkinson C (2005) The determination of logbook wind force and weather terms: the English case. Clim Change 73:57–77

Wilkinson C (2009) British Logbooks in UK Archives, 17th–19th Centuries—a survey of the range, selection and suitability of British logbooks and related documents for climatic research. CRU RP12

Wilks DS (2005) Statistical methods in the atmospheric sciences. Elsevier, London

Wilks DS (2011) Statistical methods in atmospheric sciences: an introduction. 3rd edn. Academic Press, San Diego

Wilson R, Cook E, D'Arrigo R, Riedwyl N, Evans MN, Tudhope A, Allan R (2010) Reconstructing ENSO: the influence of method, proxy data, climate forcing and teleconnections. J Quat Sci 25(1):62–78