

# A NEW APPROACH TO HOMOGENIZE DAILY PEAK WIND GUSTS: AN APPLICATION TO THE AUSTRALIAN SERIES

Cesar **AZORIN-MOLINA**<sup>1</sup>, Jose A. **GUIJARRO**<sup>2</sup>, Tim R. **McVICAR**<sup>3,4</sup>, Blair **TREWIN**<sup>5</sup>, Andrew **FROST**<sup>6</sup>, Deliang **CHEN**<sup>1</sup>

- (1) **University of Gothenburg**, Department of Earth Sciences - Regional Climate Group, Gothenburg, Sweden;
- (2) **State Meteorological Agency (AEMET)**, Balearic Islands Office, Palma de Mallorca, Spain;
- (3) **CSIRO Land and Water**, Canberra, ACT, Australia;
- (4) **Australian Research Council Centre of Excellence for Climate System Science**, Sydney, Australia;
- (5) **National Climate Centre, Australian Bureau of Meteorology (BoM)**, Melbourne, Australia;
- (6) **Australian Bureau of Meteorology**, Sydney, Australia



\*Corresponding author: Cesar Azorin-Molina, Regional Climate Group (<http://rcg.gvc.gu.se>), Department of Earth Sciences, University of Gothenburg, Box 460, 405 30 Gothenburg, Sweden.  
E-mail: [cesar.azorin-molina@gu.se](mailto:cesar.azorin-molina@gu.se)

**ABSTRACT.** Daily Peak Wind Gusts (DPWG) time-series are valuable data for the evaluation of wind related hazard risks to different socioeconomic and environmental sectors. Yet wind time-series analyses can be impacted by several artefacts, both temporally and spatially, that may introduce inhomogeneities that mislead the studies of their decadal variability and trends. The aim of this study is to present a new strategy in the homogenization of a challenging climate extreme such as the DPWG using 548 time-series across Australia for 1941-2016. This automatic homogenization of DPWG is implemented in the recently developed Version 3.0 of the R package Climatol. The new approach is an advance in homogenization of climate records as identifies 353 breakpoints based on monthly data, splits the daily series into homogeneous sub-periods, and homogenizes them without needing the monthly corrections. The major advantages of this homogenization strategy are its ability to: (i) automatically homogenize a large number of DPWG series, including short-term ones and without needing site metadata (e.g., the change in observational equipment in 2010/2011 was correctly identified); (ii) use the closest reference series even not sharing a common period with candidate series or presenting missing data; and (iii) supply homogenized series, correcting anomalous data (quality control by spatial coherence), and filling in all the missing data. The NCEP/NCAR reanalysis wind speed data was also trialled in aiding homogenization given the station density was very low during the early decades of the record; however, reanalysis data did not improve the homogenization. Application of the new approach found a reduced range of DPWG trends based on site data, and an increased negative regional trend of this climate extreme, compared to raw data and homogenized data using NCEP/NCAR. The analysis produced the first homogenized DPWG dataset to assess and attribute long-term variability of extreme winds across Australia.

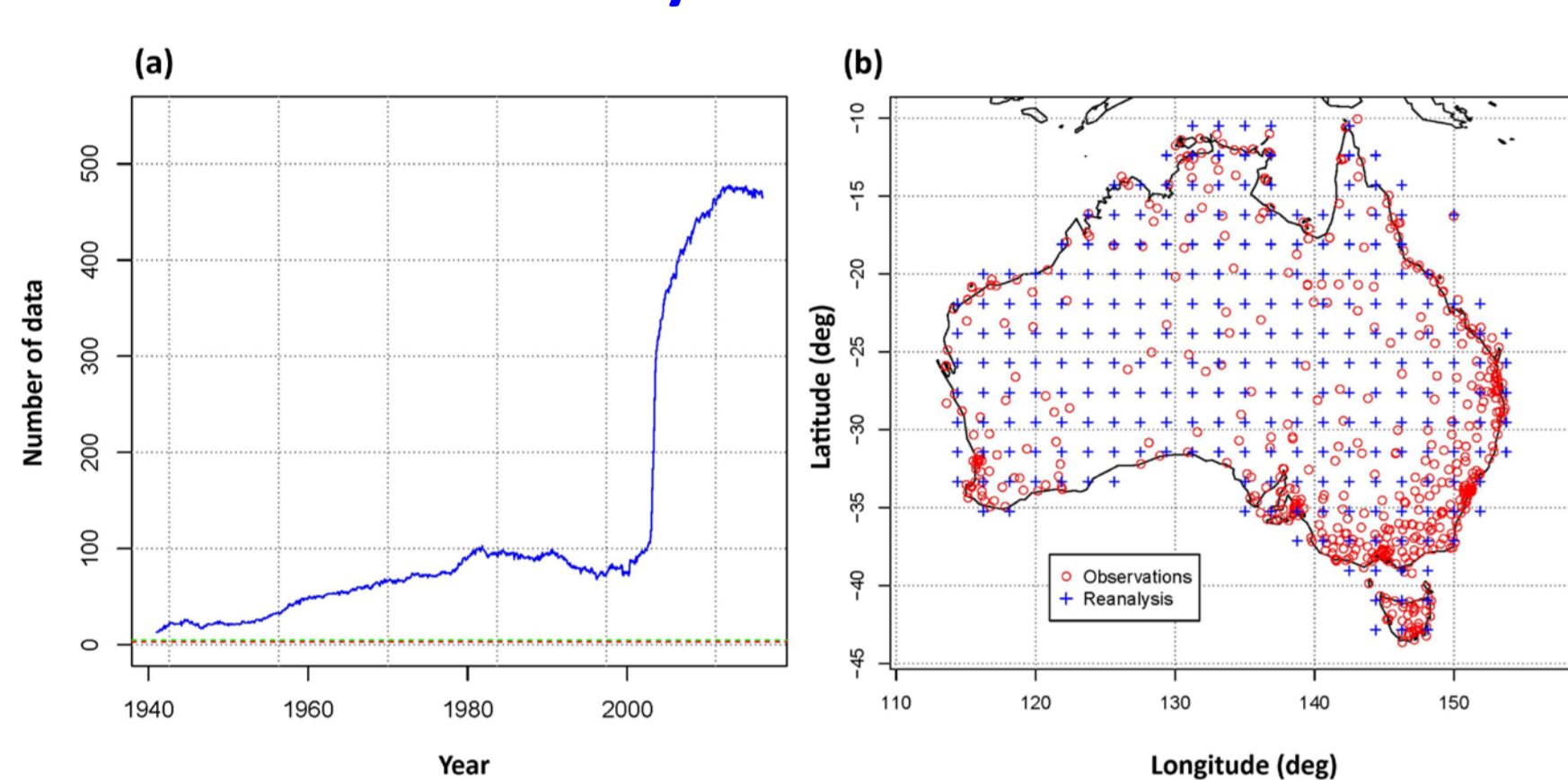
## 1. STATE-OF-THE-ART and OBJECTIVE

- (i) Most of previous homogenization studies have focused on air temperature and precipitation, with very few efforts on winds.
- (ii) Only very few approaches has attempted to homogenize wind speed series, mainly at monthly basis.
- (iii) The aim of this study is to present a new statistical approach to homogenize daily peak wind gusts (DPWG) at daily basis using long-term data series.

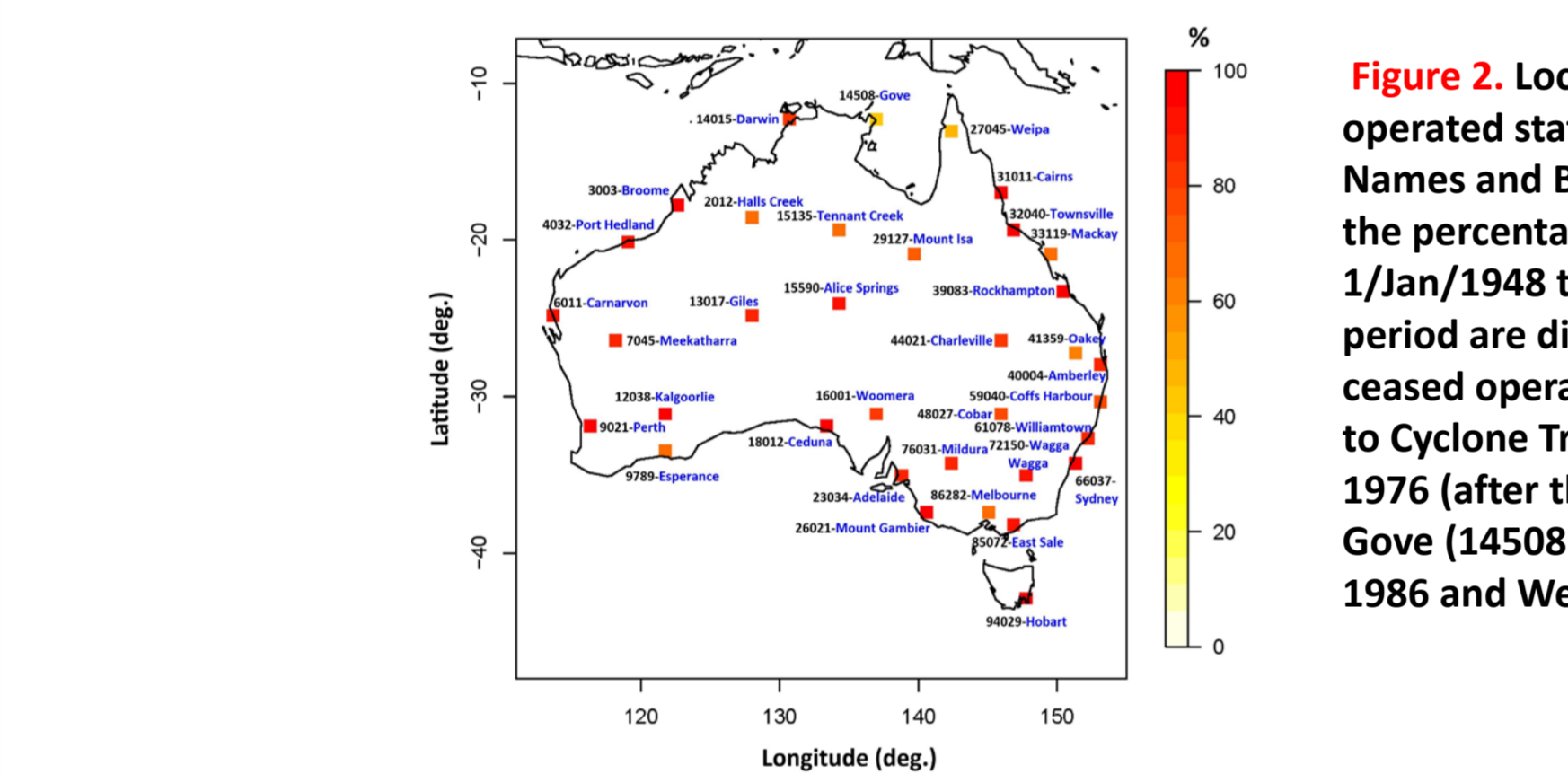


## 2. DATA AND METHODS

### Observed and reanalysis DPWG datasets

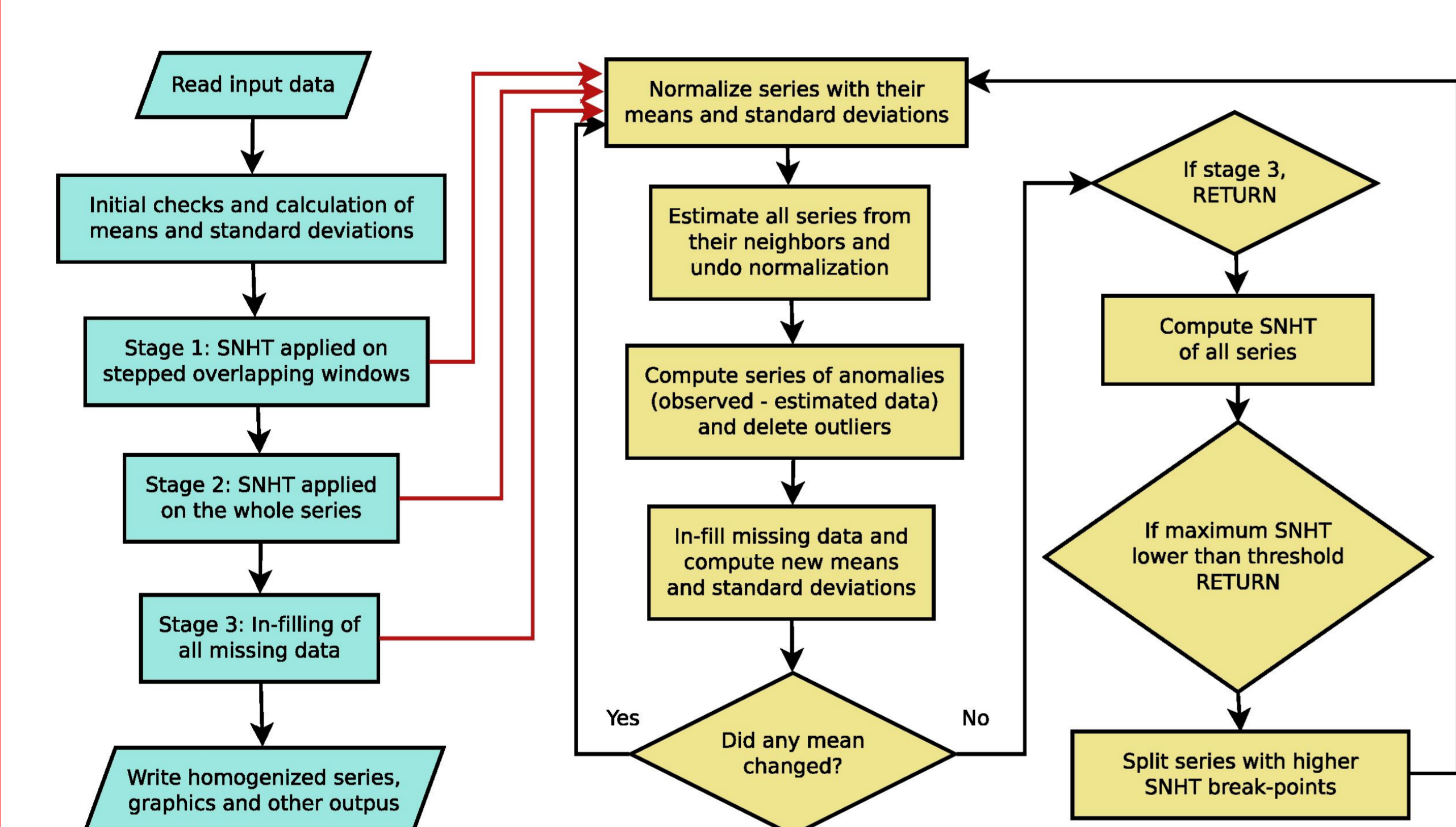


**Figure 1.** Data characteristics: (a) tracks the BoM observed data availability over 1/Jan/1941 to 31/Dec/2016 for the 548 stations. Dashed green and red lines show desirable and minimum availability thresholds for a reliable homogenization and quality control of the series; (b) shows the location of the 548 stations and center-points of the 245 NCEP/NCAR reanalysis grid-cells (791 series in total) used here.



**Figure 2.** Location of the 35 BoM operated stations used in the analysis. Names and BoM station ID along with the percentage of observations in their 1/Jan/1948 to 31/Dec/2016 (69 years) period are displayed. Darwin (14015) ceased operations on 24/Dec/1974 (due to Cyclone Tracy) and recommenced in 1976 (after the station was re-built), Gove (14508) commenced records in 1986 and Weipa (27045) in 1994.

## 3. THE CLIMATOL PACKAGE



**Figure 3.** Overview of the homogenization procedure in the Climatol package with the operations performed within each stage.

The main advantages of Climatol V3 compared to other methods are its ability:

- (i) To automatically homogenize a large number of series, including short-term ones;
- (ii) To use the closest reference series even though they do not have a common observation period with the candidate series or present missing data.
- (iii) To supply homogenized series, correcting anomalous data (quality control by spatial coherence) and filling in all the missing data.

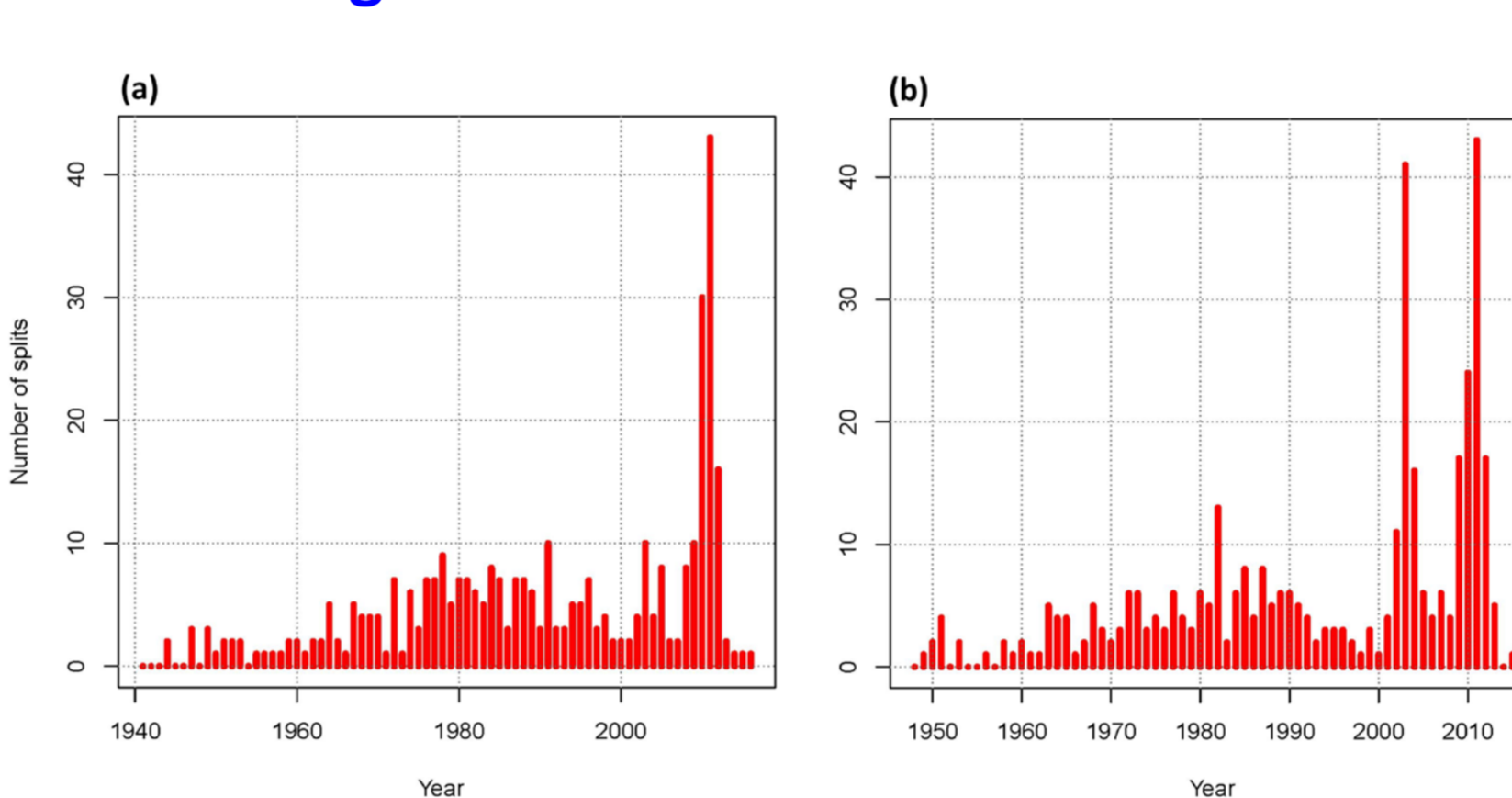
## 5. CONCLUSION

The main conclusion of this study can be summarized as follows:

- (i) The strong decrease of correlations between the observed DPWG series as distance increase cannot be minimized with the use of the NCEP/NCAR reanalysis series as references. This is because reanalyzes series represent mean wind speed (both spatially and temporally) instead of gust winds observed at a station.
- (ii) In spite of the low intersite correlations, the homogenization procedure implemented here using Climatol V3 detected 353 significant break-points in the 548 Australian series with DPWG data for 1941-2016.
- (iii) Only a small proportion of the detected break-points are supported by metadata because not all metadata necessarily produce a shift in the mean, and not all relevant metadata are reported (especially in the firsts decades).
- (iv) This study has produced a quality-controlled and homogenized DPWG dataset that will serve for assessing long-term variability and trends of this extreme weather-related hazard across Australia for 1941-2016.

## 4. RESULTS

### 4.1. Homogenization of DPWG

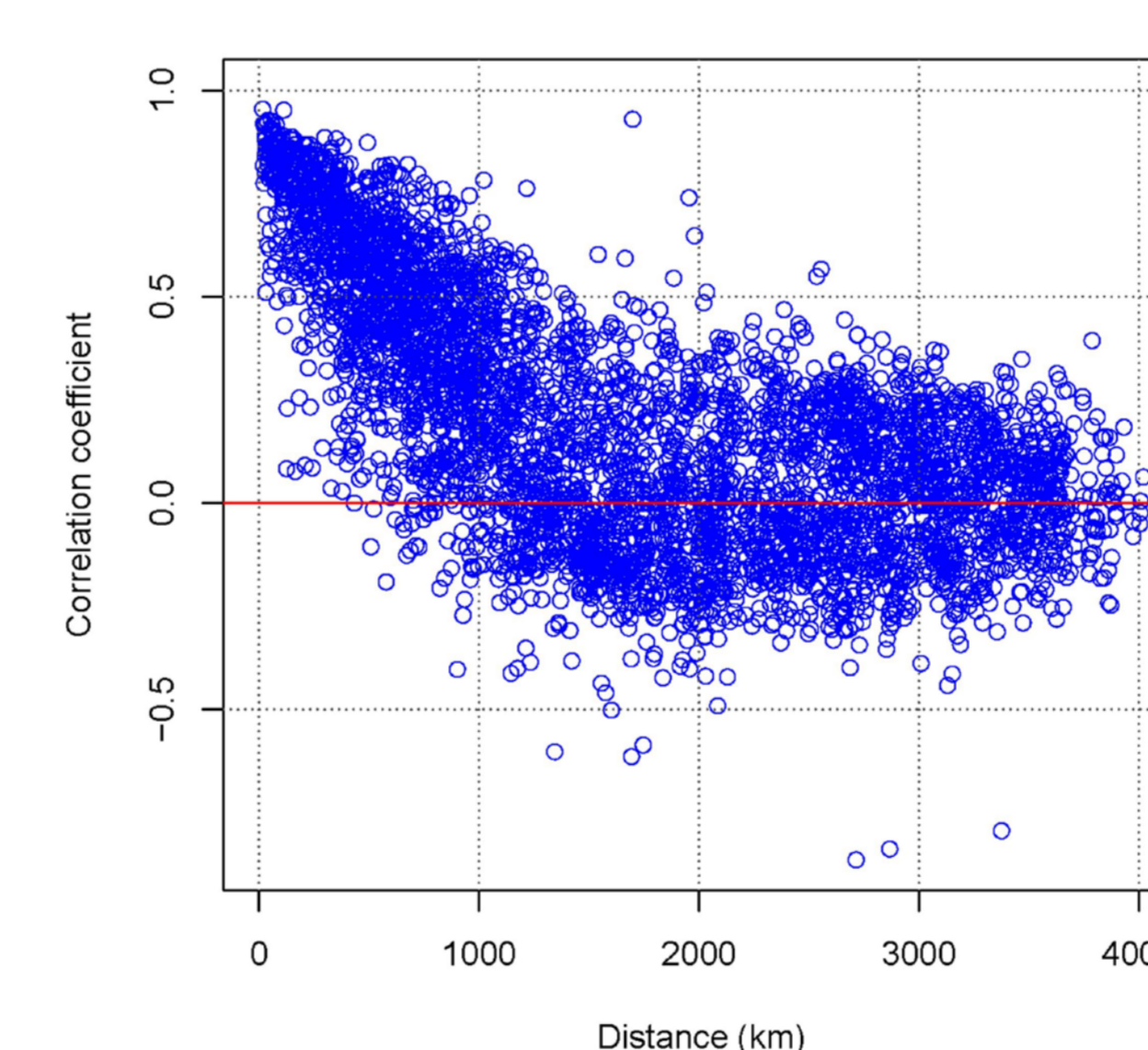


**Figure 4.** Number of splits per year applied in the homogenization of the DPWG series (a) without (1941-2016) and (b) with the support of NCEP/NCAR reanalysis series (1948-2016). For comparison purposes, note that the start year for both graphs differs.

Dataset	Breakpoints	Parameter	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
1941-2016	353	SNHT	0.90	4.70	8.00	9.69	13.25	42.20
Without NCEP/NCAR		RMSE	0.21	0.46	0.66	0.74	0.93	3.23
1948-2016	384	SNHT	0.90	4.50	7.70	9.47	12.70	48.30
Without NCEP/NCAR		RMSE	0.21	0.46	0.66	0.74	0.91	3.23
1948-2016	401	SNHT	1.00	4.30	6.90	8.59	11.00	49.10
With NCEP/NCAR		RMSE	0.19	0.53	0.75	0.83	1.06	3.22

**Table 1.** Number of detected breakpoints (splits applied) and statistical summaries of the SNHT and RMSE metrics (in  $m s^{-1}$ ) after the monthly homogenization of the DPWG series for 1941-2016, and for 1948-2016 without and with the NCEP/NCAR reanalysis reference series. Statistics refer to Min. (minimum value), Q1 (first quartile), median, mean, Q3 (third quartile) and Max. (maximum value) for the detected breakpoints.

### 4.2. Assessment of the added value of reanalysis



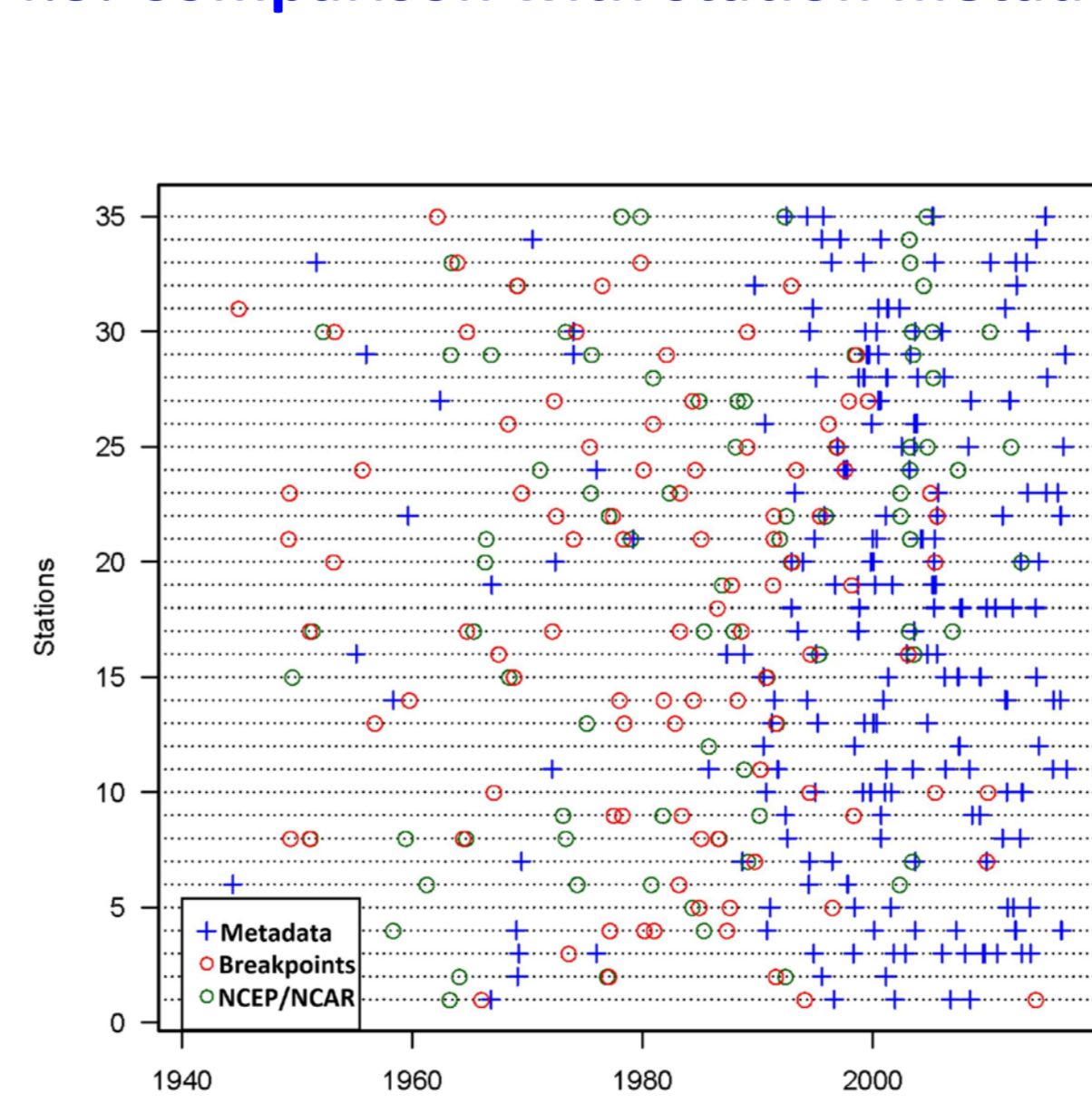
**Figure 6.** Scatter plot (correlogram) of the monthly correlation coefficients versus distance (up to ~4000 kms) of a 100 sample of DPWG series. Correlation coefficients were calculated between the first differences of the DPWG series to minimize the influence of inhomogeneities.

Period	Years	No. of stations	NCEP/NCAR	4 references		10 references	
				Mean correl.	Mean dist.	Mean correl.	Mean dist.
1948-2016	69	139	No	0.635	227	0.577	355
			Yes	0.575	108	0.546	180
1948-2002	55	86	No	0.580	291	0.517	463
			Yes	0.554	116	0.512	204
1948-1982	35	58	No	0.592	356	0.519	542
			Yes	0.547	122	0.513	215
1948-1965	15	7	No	0.501	626	0.532	1244
			Yes	0.558	131	0.512	220

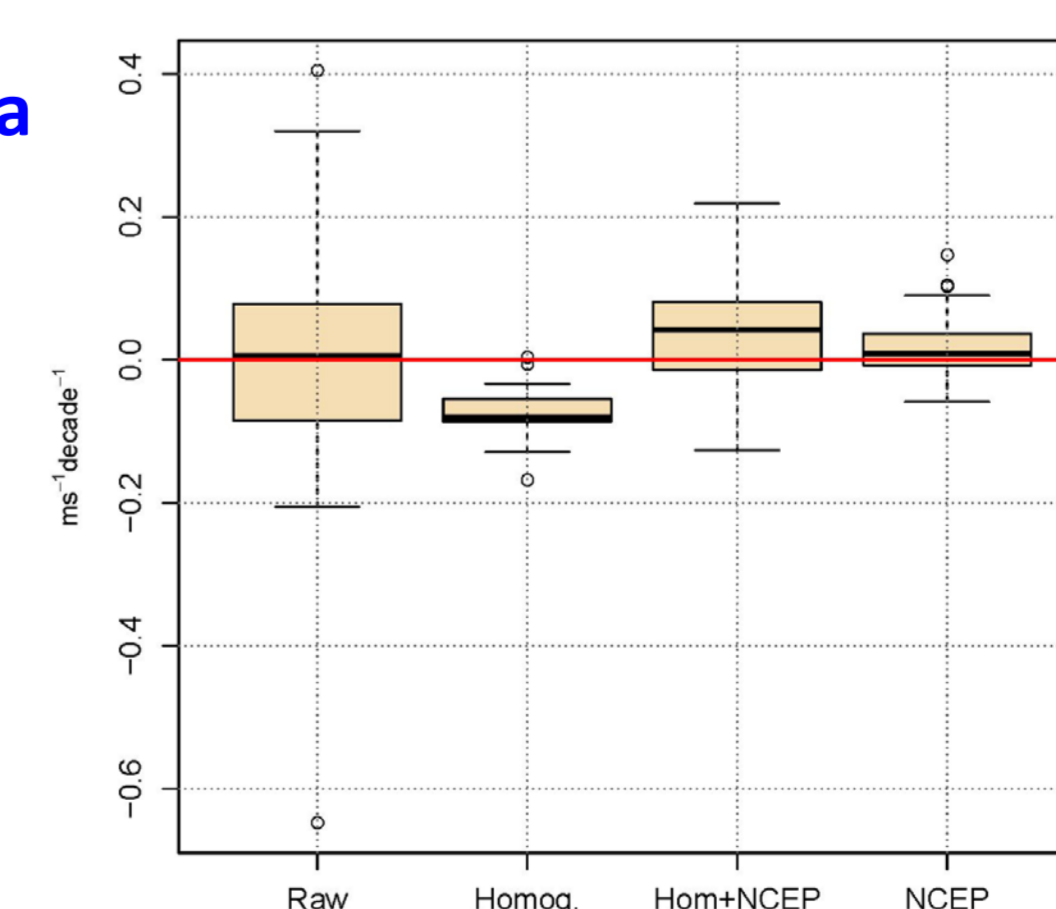
**Table 2.** Mean correlation coefficients ( $r$ ) and distances (in km) from the closest 4 and 10 references stations with a minimum of 10 years of common observations in different periods, with and without reference series from the NCEP/NCAR reanalysis.

### 4.4. Trends

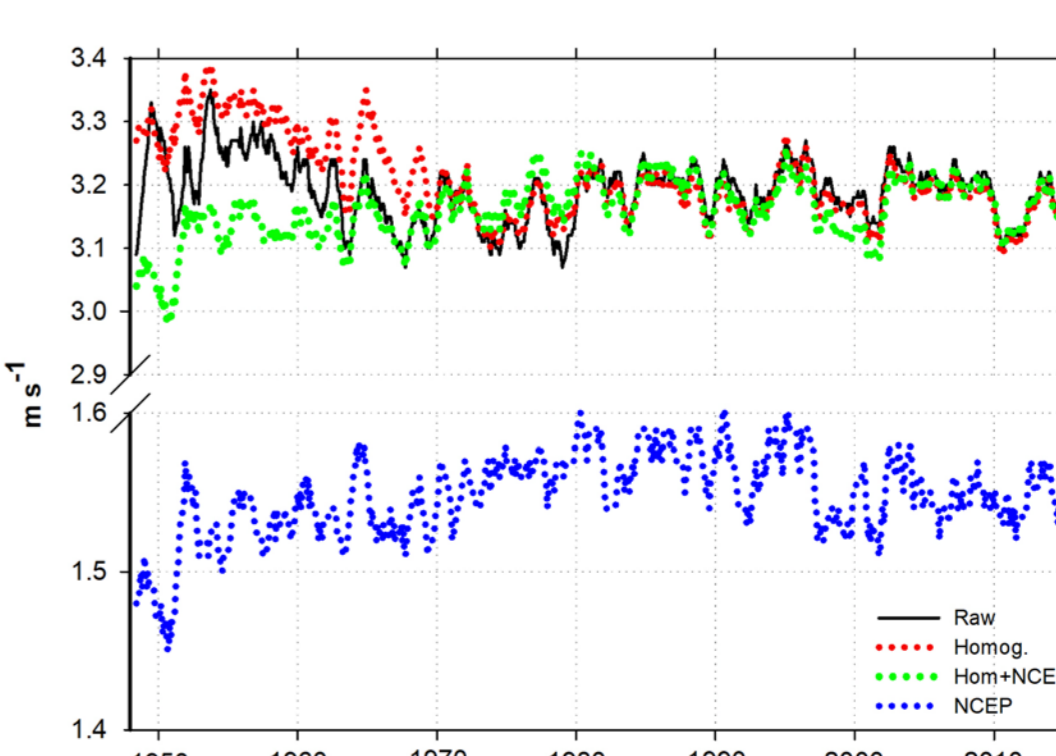
### 4.3. Comparison with station metadata



**Figure 7.** Dates of the breakpoints detected in the homogenizations of the 35 selected stations with (green circles) and without (red circles) the NCEP/NCAR references, plotted on reported metadata dates related to wind instrumentation changes (blue crosses).



**Figure 8.** Box-and-whisker plots of DPWG trends (in  $m s^{-1} dec^{-1}$ ) for all 35 selected stations computed from the original (Raw) series, the homogenized series without (Homog.) and with (Homog.+NCEP/NCAR) reanalysis references, plus trends of the 35 NCEP/NCAR reference series drawn from the closest grid cells to the selected series. The median (black line), the 25th and 75th percentile range (boxes); whiskers extend to the most extreme data points which are no more than 1.5 times the interquartile range from the box; data outside these limits (outliers) are plotted as small circles.



**Figure 9.** Running annual means of DPWG monthly averages for raw, homogenized with and without the NCEP/NCAR reanalysis across Australia for 1948-2016. The same for the NCEP/NCAR reanalysis but using mean wind speed (i.e., with lower values compared to DPWG)