Reconstruction of the historical climate of the Southern Ocean from whaling ships’ logbooks

Praveen Rao Teleti

Scott Polar Research Institute, Department of Geography
University of Cambridge

This dissertation is submitted for the degree of

Doctor of Philosophy

Christ’s College

February 2020
Declaration

This dissertation is the result of my own work and includes nothing which is the outcome of work done in collaboration except as declared in the Acknowledgements or where specifically indicated in the text. It is not substantially the same as any dissertation that has already been submitted or is being concurrently submitted for a degree, diploma or other qualification at the University of Cambridge or at any other University or similar institution except as declared in the Acknowledgements or where specifically indicated in the text. This dissertation does not exceed the length limits prescribed by the Degree Committee for the Faculty of Earth Sciences and Geography.

Praveen Rao Teleti
February 2020
Reconstruction of the historical climate of the Southern Ocean from whaling ships’ logbooks

Praveen Rao Teleti

The estimated century-long trends in the Southern Ocean climate are uncertain, as the data used to estimate these changes are relatively short (~60 years) compared to possible long-term climate variability found in the region. Due to the lack of longer land-based meteorological records, logbooks of Christian Salvesen Whaling Company’s whaling ships operating in the Southern Ocean in the 1930s and 1950s are investigated in this thesis. A historical climate dataset is produced from the meteorological observations from the whaling logbooks.

The mean sea-level pressure (MSLP) changes across the Southern Ocean diverge: for example, in the northern reaches (55°S latitudinal band) MSLP in the historical period (the 1930s and 1950s) is found to be lower than modern climatology (1981-2010). It is in contrast to the southern reaches (65°S latitudinal band) where MSLP is found to be decreasing over the same period. A historical Southern Annular Mode (SAM) index from 1930-1960 is generated from the whaling dataset, and a significant positive trend is found. Subsequently, historical cyclonic frequency is estimated using a semi-supervised cyclone identification algorithm. The average number of cyclones per year for the historical period is statistically lower than the average for the 1999-2008 period over a control area in the Weddell Sea.

Finally, the whaling dataset is assimilated into two (CERA-20C and 20CRv2c) current-generation reanalyses using an offline data assimilation method. The uncertainty in MSLP fields over the assimilation window decreased in both reanalyses by ~40% in the area of observations post-assimilation.

Overall it is shown that meteorological observations from the whaling logbooks can be utilised to reconstruct historical climate in terms of MSLP variability, climate modes (e.g. SAM) and identification of individual cyclones, and to improve the representation of past climate in the reanalyses. In summary, it has been demonstrated that the Southern Ocean climate near the Antarctic coast in the early/mid-20th Century was characterised by higher MSLP and less cyclonic activity compared to the modern period.
Except the LORD build the house, they labour in vain that build it: except the LORD keep
the city, the watchman waketh but in vain.

Psalms 127:1 (KJV)
Acknowledgements

First, I wish to begin by thanking my two supervisors, Dr G. Rees and Prof. J. Dowdeswell, for diligently guiding this research to completion. I could not have hoped for better individuals to add their strengths to supervise my PhD and to hold a genuine interest in my success and welfare. In some respects this thesis would not have been possible if not for Dr C. Wilkinson, University of East Anglia, along with sharing historical whaling logbook images to be used in this thesis, he also shared expertise and passion for all things historical and marine. Discussions with Dr G. Marshall, British Antarctic Survey, are appreciated and helped to improve many chapters in this thesis. I thank Nehru Trust for Cambridge University and Christ’s College for their generosity to fund this PhD studentship.
Table of contents

List of figures xvii
List of tables xix
Nomenclature xxi

1 Introduction 1
  1.1 Importance of Southern Ocean climate . . . . . . . . . . . . . . . . . . . . 2
  1.2 Prelude to modern Southern Ocean and Antarctic climate . . . . . . . . . . 4
  1.3 Atmospheric circulation changes in the last 60 years . . . . . . . . . . . . . 6
  1.4 State of Reanalyses over the 20th century . . . . . . . . . . . . . . . . . . . 9
  1.5 Data-rescue efforts . . . . . . . . . . . . . . . . . . . . . . . . . . . . .10
  1.6 Major global gridded datasets . . . . . . . . . . . . . . . . . . . . . . . .14
  1.7 Lack of meteorological observations in the Southern Ocean . . . . . . . . . 21
  1.8 Significance of whaling logbooks . . . . . . . . . . . . . . . . . . . . . . .22
  1.9 Thesis aims and structure . . . . . . . . . . . . . . . . . . . . . . . . . . .23

2 Climate dataset from whaling ships’ logbooks 27
  2.1 Introduction . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .27
  2.2 Data and Methods . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .30
    2.2.1 Description of logbooks . . . . . . . . . . . . . . . . . . . . . . . . .30
## Table of contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.2.2</td>
<td>Instructions and error detection/correction</td>
<td>34</td>
</tr>
<tr>
<td>2.2.3</td>
<td>Standardisation</td>
<td>41</td>
</tr>
<tr>
<td>2.3</td>
<td>Results and future steps</td>
<td>42</td>
</tr>
<tr>
<td>2.3.1</td>
<td>Spatial and Temporal spread of dataset</td>
<td>42</td>
</tr>
<tr>
<td>2.3.2</td>
<td>Planned use of the historical dataset</td>
<td>45</td>
</tr>
<tr>
<td>3</td>
<td>MSLP variability in the Southern Ocean and historical SAM index</td>
<td>47</td>
</tr>
<tr>
<td>3.1</td>
<td>Introduction</td>
<td>47</td>
</tr>
<tr>
<td>3.2</td>
<td>Data and Methods</td>
<td>51</td>
</tr>
<tr>
<td>3.2.1</td>
<td>Christian Salvesen whaling dataset</td>
<td>52</td>
</tr>
<tr>
<td>3.2.2</td>
<td>Observational datasets</td>
<td>53</td>
</tr>
<tr>
<td>3.2.3</td>
<td>Spatial aggregation</td>
<td>54</td>
</tr>
<tr>
<td>3.2.4</td>
<td>Generation of SAM index</td>
<td>54</td>
</tr>
<tr>
<td>3.2.5</td>
<td>Bootstrapping with Monte Carlo realisations</td>
<td>57</td>
</tr>
<tr>
<td>3.2.6</td>
<td>Subsets of MSLP observations</td>
<td>58</td>
</tr>
<tr>
<td>3.2.7</td>
<td>SAM indices from MSLP subsets</td>
<td>58</td>
</tr>
<tr>
<td>3.2.8</td>
<td>Statistical Appraisal</td>
<td>59</td>
</tr>
<tr>
<td>3.2.9</td>
<td>Randomised ANOVA test</td>
<td>60</td>
</tr>
<tr>
<td>3.3</td>
<td>Results</td>
<td>60</td>
</tr>
<tr>
<td>3.3.1</td>
<td>Temporal spread of whaling seasons</td>
<td>60</td>
</tr>
<tr>
<td>3.3.2</td>
<td>Comparison and validation of the whaling dataset</td>
<td>61</td>
</tr>
<tr>
<td>3.3.3</td>
<td>Seasonal variability of MSLP across the Southern Ocean</td>
<td>66</td>
</tr>
<tr>
<td>3.3.4</td>
<td>Changes in the Semi-Annual cycle</td>
<td>70</td>
</tr>
<tr>
<td>3.3.5</td>
<td>Seasonal changes across decades</td>
<td>72</td>
</tr>
<tr>
<td>3.3.6</td>
<td>New marine observations-based SAM indices</td>
<td>74</td>
</tr>
<tr>
<td>3.3.7</td>
<td>Trends in the historical SAM indices</td>
<td>77</td>
</tr>
<tr>
<td>3.4</td>
<td>Discussion</td>
<td>77</td>
</tr>
</tbody>
</table>
# Table of contents

3.4.1 Spatio-temporal distribution of whaling dataset .......................... 77
3.4.2 Validation of whaling dataset ............................................. 79
3.4.3 Decadal changes over the Southern Ocean ......................... 80
3.4.4 Monthly Variability and Semi-Annual cycle changes .......... 81
3.4.5 Choice of high-latitude band ........................................... 83
3.4.6 Marine observations-based historical SAM and its trend ....... 84
3.5 Conclusions ............................................................................. 87

4 A novel approach for historical cyclone detection .............. 89
4.1 Introduction ............................................................................. 89
4.2 Data ......................................................................................... 93
4.2.1 Christian Salvesen whaling dataset (historical) ................. 93
4.2.2 ICOADS meteorological observations (modern) .............. 93
4.2.3 Cyclone track datasets ..................................................... 94
4.3 Methodology ........................................................................... 95
4.3.1 Method One ................................................................. 95
4.3.2 Method Two ................................................................. 96
4.3.3 Randomised ANOVA test .............................................. 103
4.4 Results ..................................................................................... 104
4.4.1 Number of cyclone encounters using Method One .......... 104
4.4.2 Identification and tracking of cyclones using Method Two .. 105
4.4.3 Semi-supervised cyclone detection algorithm ................. 113
4.4.4 Comparison of cyclonic frequency ................................. 115
4.5 Discussion and Summary ..................................................... 118
4.5.1 Method One ................................................................. 118
4.5.2 Method Two ................................................................. 119
4.5.3 Limitation of methods .................................................. 122
## Table of contents

4.5.4 Comparison: Historical vs Modern ........................................ 123
4.6 Conclusions ............................................................................... 124

5 Offline data assimilation of historical whaling ships’ observations .................................................. 127

5.1 Introduction ............................................................................. 127
5.2 Data Sources and Methods ....................................................... 131
  5.2.1 Datasets ........................................................................... 131
  5.2.2 Methods .......................................................................... 133
5.3 Results ..................................................................................... 135
  5.3.1 Decadal means and uncertainty ......................................... 135
  5.3.2 Seasonal variability ............................................................ 136
  5.3.3 Effect of data assimilation ............................................... 139
  5.3.4 Quantification of the sensitivity ....................................... 140
  5.3.5 Reduction of Error (RE) .................................................... 145
5.4 Discussion ............................................................................... 146
  5.4.1 Measurement of ensemble uncertainties ......................... 146
  5.4.2 Differences in reanalyses across seasons and decades ......... 148
  5.4.3 Data assimilation methodology ....................................... 151
  5.4.4 Post-assimilation changes and their sensitivities .............. 151
  5.4.5 RE scores ........................................................................ 152
5.5 Conclusions ............................................................................ 154

6 Conclusions .................................................................................. 157

6.1 Overview .................................................................................. 157
6.2 Summary of the thesis .............................................................. 158
6.3 Limitations, future work and perspectives .............................. 161

References ..................................................................................... 165
Appendix A  Supplementary Information  187
# List of figures

1.1 A polar stereographic projection of the Southern Ocean .................. 3
1.2 Antarctic station MSLP trends for 1971–2000 ............................. 7
1.3 The focus areas of the ACRE’s regional chapters ......................... 11
1.4 Leith Harbour whaling station on South Georgia .......................... 22

2.1 A typical catch book used in the 1930s ..................................... 31
2.2 A typical Chief Officer’s logbook issued in the 1950s ...................... 32
2.3 Ship tracks from the Christian Salvesen Whaling Co. logbooks .......... 36
2.4 The whaling ships positions grouped by whaling season ................. 43
2.5 Spatial density plot of the whaling dataset ................................. 44

3.1 Longitudinal distribution of MSLP data in the whaling dataset ........ 55
3.2 Latitudinal distribution of MSLP data within each Antarctic sector .... 55
3.3 Spatial distribution of whaling dataset ................................. 56
3.4 Temporal spread of whaling seasons ....................................... 61
3.5 Comparison between ICOADS and whaling dataset ....................... 62
3.6 Whaling dataset correlation with Antarctic stations ...................... 65
3.7 Seasonal variability across the Southern Ocean ............................ 67
3.8 Changes in the Semi-Annual cycle ....................................... 71
3.9 New marine observations-based SAM indices ............................. 76
3.10 Trends in the historical SAM indices ................................... 78
<table>
<thead>
<tr>
<th>Figure</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.11</td>
<td>Spatio-temporal changes in the circumpolar trough movement</td>
<td>82</td>
</tr>
<tr>
<td>4.1</td>
<td>Cyclone Jo during January-February 2000</td>
<td>98</td>
</tr>
<tr>
<td>4.2</td>
<td>Ideal case and pdf plot for Method Two</td>
<td>102</td>
</tr>
<tr>
<td>4.3</td>
<td>Cyclone encounters from Method One</td>
<td>105</td>
</tr>
<tr>
<td>4.4</td>
<td>Whaling season 1935-36 ship tracks</td>
<td>106</td>
</tr>
<tr>
<td>4.5</td>
<td>Whaling season 1935-36 meteorological snapshot</td>
<td>107</td>
</tr>
<tr>
<td>4.6</td>
<td>Whaling season 1935-36 week 11 and 12 cyclones</td>
<td>109</td>
</tr>
<tr>
<td>4.7</td>
<td>Whaling season 1956-57 ship tracks</td>
<td>110</td>
</tr>
<tr>
<td>4.8</td>
<td>Whaling season 1956-57 meteorological snapshot</td>
<td>111</td>
</tr>
<tr>
<td>4.9</td>
<td>Whaling season 1956-57 week 7 and 8 cyclones</td>
<td>112</td>
</tr>
<tr>
<td>4.10</td>
<td>Cyclonic frequency comparison: Historical vs Modern</td>
<td>116</td>
</tr>
<tr>
<td>5.1</td>
<td>Decadal means and uncertainty 20CRv2c</td>
<td>136</td>
</tr>
<tr>
<td>5.2</td>
<td>Decadal means and uncertainty CERA-20C</td>
<td>137</td>
</tr>
<tr>
<td>5.3</td>
<td>Seasonal variability CERA-20C and 20CRv2c</td>
<td>138</td>
</tr>
<tr>
<td>5.4</td>
<td>Effect of data assimilation 20CRv2c</td>
<td>141</td>
</tr>
<tr>
<td>5.5</td>
<td>Effect of data assimilation CERA-20C</td>
<td>142</td>
</tr>
<tr>
<td>5.6</td>
<td>Pre- minus post-assimilation for 20CRv2c</td>
<td>144</td>
</tr>
<tr>
<td>5.7</td>
<td>Pre- minus post-assimilation for CERA-20C</td>
<td>144</td>
</tr>
<tr>
<td>5.8</td>
<td>The RE values for CERA-20C and 20CRv2c</td>
<td>147</td>
</tr>
<tr>
<td>A.1</td>
<td>Outliers detected from the Southern Harvester ship for the 1956-57 whaling season</td>
<td>187</td>
</tr>
<tr>
<td>A.2</td>
<td>Conceptualisation of SAM, SAO along with stations used to calculate SAM index</td>
<td>188</td>
</tr>
<tr>
<td>A.3</td>
<td>Effect of data assimilation 20CRv2c in the 1950s</td>
<td>189</td>
</tr>
<tr>
<td>A.4</td>
<td>Effect of data assimilation CERA-20C in the 1950s</td>
<td>190</td>
</tr>
<tr>
<td>A.5</td>
<td>A sample composite of cyclones in the 1935-36 whaling season</td>
<td>191</td>
</tr>
</tbody>
</table>
List of tables

2.1 List of vessels from the Christian Salvesen whaling logbooks ............. 33
2.2 Beaufort wind scale with description ........................................... 38
2.3 Sea State and Swell Scale ............................................................... 39

3.1 Correlations between sub-and Antarctic meteorological stations .......... 64
3.2 Seasonal variability across the months ........................................... 68
3.3 Seasonal variability across the zonal bands .................................... 69
3.4 Seasonal variability across the time periods .................................... 69
3.5 Seasonal changes across decades .................................................. 73

4.1 Method Two model parameters ...................................................... 100
4.2 Method One cyclone encounters .................................................... 104
4.3 Cyclones detected in Case 1 ......................................................... 108
4.4 Cyclones detected in Case 2 ......................................................... 111
4.5 Total number of cyclones estimated from Method Two ....................... 114
4.6 Number of cyclones detected passing through control area ............... 117
4.7 Number of cyclones detected in the modern period ......................... 117

5.1 Decadal means and uncertainty CERA-20C and 20CRv2c .................. 137
5.2 Seasonal variability across zonal bands ....................................... 139
**Nomenclature**

20CRv2c  Twentieth Century Reanalysis version 2c

ABS    Amundsen-Bellinghausen Sea

ACC    Antarctic Circumpolar Current

ACRE   Atmospheric Circulation Reconstructions over the Earth

AMOC   Atlantic Meridional Overturning Circulation

ASL    Amundsen Sea Low

BP     Before Present

C3S    Copernicus Climate Change Service

CERA-20C Coupled European Reanalysis - 20th Century

CLIWOC Climatological Database for the World’s Oceans

CMIP   Coupled Model Intercomparison Project

CRU    Climate Research Unit

DA     Data Assimilation

ECMWF  European Centre for Medium-Range Weather Forecasts

EEIC   English East India Company

ENSO   El Niño-Southern Oscillation

ERA    European Reanalysis
Nomenclature

FFT  Fast Fourier Transform
GCM  Global Climate Model
GCOS Global Climate Observing System
GHCN Global Historical Climate Network
GMST  global mean surface temperature
HadCRUT  Hadley Centre-Climate Reseach Unit Temperature dataset
HadSLP  Hadley Centre Sea Level Pressure dataset
HadSST  Hadley Centre Sea Surface Temperature dataset
HBC  Hudson Bay Company
IBTrACS  International Best Track Archive for Climate Stewardship
ICOADS  International Comprehensive Ocean-Atmosphere Data Set
IMILAST  Intercomparison of Mid-Latitude Storm Diagnostics
IMMA  International Maritime Meteorological Archive
IPCC  Intergovernmental Panel on Climate Change
ISPD  International Surface Pressure Databank
MSLP  Mean Sea-level Pressure
NIWA  National Institute of Water and Atmospheric Research
NMS  National Meteorological Service
NOAA  National Oceanic and Atmospheric Administration
NWP  Numerical Weather Prediction
OCR  Optical Character Recognition
PSA  Pacific-South American pattern
RCP  Representative Concentration Pathway
Nomenclature

RECLAIM Reclaim Marine Observations
SAM Southern Annular Mode
SAO Semi-Annual Oscillation
SCAR Scientific Committee on Antarctic Research
SG South Georgia
SMRU Sea Mammal Research Unit
SST Sea Surface Temperature
UNFCCC United Nations Framework Convention on Climate Change
WMO World Meteorological Organisation
Chapter 1

Introduction

The global climate has witnessed great changes, many unprecedented, in the last 200 years: the three hottest years since temperature records began have been observed in the last ten years (MetOffice 2020), the shrinking of many glaciers around the globe has been observed (Zemp et al. 2019), and the drastic reduction in the Arctic sea ice and disintegration of northern permafrost has taken place (Parkinson 2019, Turetsky et al. 2019). Knowledge and understanding of these changes have only been possible due to long-term meteorological observations. These long-term observations are the result of many nations starting the near-modern land-based network of meteorological stations in Europe and their many colonial administrative regions and, by the late nineteenth century, station coverage expanded to include all inhabited continents and major shipping routes. An inventory of the earliest pre-1850 land-based observations is presented in Brönnimann et al. (2019). Today, we can observe global climate change in many important climate parameters, with some such as temperature (Hadley Centre-Climate Research Unit Temperature version 4 (HadCRUT4); Morice et al. 2012), going back to the mid-to-late nineteenth century.

These meteorological observations are the bedrock of our understanding of the climate system and are useful as a baseline against which to assess the state of the current climate with respect to past climate. The differences in the global temperature (current minus pre-industrial baseline) are the measure used in the United Nations Framework Convention on Climate Change (UNFCCC) and resulting Paris Agreement of 2015 that aims to ’prevent dangerous anthropogenic interference with the climate system’. The Paris Agreement defines an ambition to limit global temperature change to below 1.5°C or 2°C above pre-industrial
levels (the period 1850-1900 as defined by the Intergovernmental Panel on Climate Change (IPCC); Intergovernmental Panel on Climate Change 2014).

However, global mean surface temperature (GMST) is highly uncertain before 1950 (Hawkins et al. 2017, Schurer et al. 2014), as it relies on proxy data, sparse instrumental data for many areas and climate models with uncertain radiative forcing. In addition, different GMST datasets disagree on the amount of warming since 1850 by more than 10%. Whether we have 0.62°C (according to HadCRUT4) or 0.48°C (Berkeley Earth; Rohde et al. 2013) remaining before reaching 1.5°C above this baseline is critically important as this disagreement implies more than 20% uncertainty in the allowed carbon budget to meet the Paris Agreement goals solely due to the choice of GMST dataset. The differences result from combining scarce historical observations with proxy and interpolated data, especially over the oceans and in the Southern Hemisphere in the historical period.

On the other hand, a suite of coupled climate models (Fifth phase of the Coupled Model Intercomparison Project (CMIP5); Taylor et al. 2011) has consistently underestimated Antarctic sea-ice: many of the models show drastic reduction whereas the observations show a significant increase in Antarctic sea-ice until the last few years (de Lavergne et al. 2014; Parkinson 2019). This apparent discrepancy has been attributed primarily to inadequate representation of the processes driving these changes in the numerical models, due to lack of long observational datasets for accurate parametrisation (Hobbs et al. 2016). A large component of uncertainty in the GMST changes and inaccurate modelling of Antarctic sea ice arises due to the lack of long-term observations in regions such as the Southern Ocean, which limits our understanding of changing global climate. Southern Ocean climate, in turn, plays a major role in the global climate, and the understanding of the Southern Ocean climate is therefore essential for understanding global climate.

1.1 Importance of Southern Ocean climate

The Southern Ocean (defined as the circumpolar mass of water polewards of 60°S) covers approximately 14% of world’s surface and plays a vital role in global climate due to its size and volume (Fig. 1.1). It affects net global solar radiation by changing planetary albedo via its ability to reflect solar radiation from sea ice or to absorb it through open waters. In addition, in the process of forming seasonal sea ice, the production of cold, dense water takes place, which is a significant driver of the global ocean circulation. The strong westerly winds that blow over the Southern Ocean drive the world’s largest and strongest current system,
1.1 Importance of Southern Ocean climate

The Antarctic Circumpolar Current (ACC), and are recognized to be the dominant driving force for the Southern Ocean’s meridional overturning circulation, acting as a medium for propagation of climate signals (teleconnections) across the globe. Furthermore, the Southern Ocean is the world’s most biologically productive ocean and acts as a significant sink for CO$_2$, playing a vital role in modulating global CO$_2$ concentration levels.

Fig. 1.1 A polar stereographic projection (south of 30°S) showing the Southern Ocean and its division into five sectors. The sectors are divided as defined by Parkinson and Cavalieri (2012). ABS is Amundsen-Bellinghausen Sea.

The CO$_2$ and energy fluxes in the Southern Ocean form an important link between the regional atmosphere and cryosphere and the deep global ocean. Furthermore, Antarctica
holds 90% of the world’s freshwater as ice sheets, which reach elevations of over 4000 m. The changes in the intensity and direction of winds and cyclonic frequency play a significant role in the ice-sheet mass balance, thus affecting modern and future global sea-level. Before discussing the current state of the Southern Ocean and Antarctic climate, their key processes and cycles, it is useful to briefly discuss the prelude to the modern climate since the start of the Holocene some 10,000 years ago.

1.2 Prelude to modern Southern Ocean and Antarctic climate

The current climate of Antarctica and the Southern Ocean did not evolve in isolation but in response to global climatic changes. Many studies using a variety of proxies (e.g., pollen records, tree-rings, ice-cores, lake sediment cores, glacial grounding lines among others) have pieced together distinct changes in the global climate throughout its history. I will briefly highlight the abrupt changes in the Antarctic climate in the last 10,000 years or so (the Holocene), since the last glacial maximum about 20,000 years ago (e.g. RAISED Consortium Bentley et al. 2014). It is generally believed that, at the start of the most recent inter-glacial cycle, sudden destabilisation of the southern Laurentide Ice Sheet due to early-Holocene optimum warm temperatures caused a large amount of western Canadian freshwaters to discharge into the North Atlantic (Gregoire et al. 2012). This catastrophic freshening disturbed the thermohaline gradient between the colder North Atlantic and warmer South Atlantic. It led to near-total shut down of the Atlantic Meridional Overturing Circulation (AMOC), providing barriers to heat transport from equatorial regions to Northern high latitudes. This period of shut-down of the AMOC is thought to have caused the Younger Dryas (YD) period, which cooled the Northern Hemisphere to near-glacial conditions between 12,900–11,500 years Before Present (BP) (Carlson 2013). Proxy evidence for this climate cooling can be found in various sediments, glacier-margin positions and in the isotopic record in Greenland ice cores (Severinghaus et al. 1998).

The shut down of AMOC had profound global impact as it caused a breakdown in the inter-hemispheric heat transport, resulting in a very cold Northern Hemisphere but warmer Southern Hemisphere. Nearly all long-term sediments and ice-core proxies point to this climatic optimum near the start of the Holocene (11500-9000 BP) in the Antarctic (Masson et al. 2000). It was followed by the retreat of the marine parts of the West Antarctic to
their current position between 7000-9000 BP (Conway et al. 1999). A dramatic drop in temperatures around 8200 BP is visible in the East Antarctic ice-core isotope record (Masson et al. 2000). A change in solar variability was detected alongside a long-term decrease in greenhouse gases. The period between 7500-5000 BP saw an intensification of westerlies coincident with the cooling of West Antarctica as indicated by the ice-cores from Siple Dome in West Antarctica (Browne et al. 2017). After 5000 BP, a milder climate appeared with milder temperatures in both West and East Antarctica. The westerlies also weakened during this period lasting until 2500 BP.

The evidence for another temperature drop is found in Siple Dome around 6400-6200 BP, followed by an increase in temperature over the East Antarctic Ice Sheet starting at 6000 BP that continued until 1200 BP. Several inland lakes in the McMurdo Dry Valleys reached high water levels, while coastal lakes dried up during the same period (Hall et al. 2010). Around 1500 BP the Southern Hemisphere westerlies begin to intensify accompanied by deepening of the Amundsen Sea Low (ASL), a quasi-permanent region of low pressure located in the Amundsen Sea between 60°S-70°S (see Fig. A.2; Hosking et al. 2013, Turner et al. 2013). It was followed by cooling during 1200-1000 BP (Mayewski and Maasch 2006).

Overall, these changes appear to be the product of fluctuations in insolation, the large-scale production of aerosols from diverse sources including volcanic eruptions, and changes in greenhouse gases. These signals are then superimposed on the internal variability of the Antarctic ice sheets and climate, leading to the observed changes. The understanding of these long-term changes is vital to demonstrate the effect of external and internal forcing on the Antarctic climate system and how it modulates our modern global climate.

Furthermore, it is observed that circulation and temperature changes occur in succession; for example, reconstructions of Northern and Southern Hemisphere temperature and circulation changes found that large-scale circulation changes always preceded temperature changes (Mayewski et al. 2009). However, the current global warming observed since 1850 was not preceded by large-scale circulation changes. It has been suggested that these post-1850 temperature changes are forced by non-natural factors, e.g. anthropogenic effects (Mayewski et al. 2009). However, we are not able to confirm this hypothesis due to the lack of long-term instrumental meteorological observations, especially in the Southern Ocean. Some circulation changes have been detected from available observations, and a better understanding of these changes can shed more light into the proposed anthropogenic warming in the Southern Ocean in the modern period.
1.3 Atmospheric circulation changes and their effect over the last 60 years

One of the principal modes of atmospheric circulation in the Southern Hemisphere is the Southern Annular Mode (SAM; see Fig. A.2), which is essentially a dipole between southern mid-latitude and high-latitude mean sea-level pressure (MSLP) (Gong and Wang 1999, Marshall 2003). The SAM pattern is found in many atmospheric fields in the Southern Hemisphere as a leading mode of variability, explaining as much as 35% of observed variability (Thompson and Wallace 2000). A positive trend is observed in the SAM index in the austral summer and autumn, as pressures have decreased around the coast of the Antarctic and increased in mid-latitudes (Thompson et al. 2011). This is attributed primarily to ozone depletion since the 1970s, and greenhouse gases; however, the evidence for the impact of greenhouse gases is conflicting (Treguier et al. 2010). The exact mechanism by which the signal from the loss of stratospheric ozone descends to the surface level is under active investigation, and probable pathways and the effects of other forcing factors on SAM variability are examined in Fogt et al. (2009). The positive trend in SAM computed from mid-latitude and coastal Antarctic stations has been observed since 1957 (Marshall 2003).

The annual and seasonal analysis of MSLP of various Antarctic coastal stations reveals a more precise picture (Turner et al. 2005). All stations show a negative annual trend for the 1971-2000 period, except for Orcadas (60.7°S 44.7°W) on the South Orkney Islands which has small insignificant positive trend (Fig. 1.2). Stations in East Antarctica show larger negative trends than their western counterparts both on annual and seasonal time-scales. The significance of these trends also varies spatially. The stations located on the East Antarctic coast show higher statistical significance than all other stations compared. The greatest negative trend is found in the summer, followed by winter, for most of the stations analysed. The seasonal variability of MSLP of Antarctic coastal stations could be explained in relation to another pressure gradient that exists between 50°S in the open water and 65°S at the Antarctic coast due to uneven lagging of heating and cooling in these regions (see Fig. A.2). It leads to biannual expansion and contraction of the circumpolar trough, a belt of low-pressure known as the Semi-Annual Oscillation (SAO; Broeke 2000).

The expansion phase of the SAO occurs between September to December and March to June, with contraction usually between December to March and June to September. The SAO is observed to have been weakening since the 1980s and is believed to be partly responsible for coastal MSLP deepening (Broeke 1998, Broeke 2000). Evidence for this is provided by
1.3 Atmospheric circulation changes in the last 60 years

Fig. 1.2 Antarctic station MSLP trends for 1971–2000, adapted from Turner et al. (2005).
the movement of the circumpolar trough to its southernmost position during summer (Broeke 2000). The observed weakening of the SAO is manifested in a decrease in its strength or restricted movement or both (Meehl et al. 1998). Meehl et al. also found that the circumpolar trough is not apparent in the MSLP fields around the 50°S latitude band in summer and winter (its climatological position during summer and winter), suggesting that since the 1980s the circumpolar trough has been mainly confined to its more southerly position throughout the year. Importantly, the position of the circumpolar trough is known to modulate storm tracks in the Southern Ocean (Thompson and Solomon 2002).

In addition, previous studies have found that large-scale climate modes, for example, the SAM and El Niño-Southern Oscillation (ENSO), have impacted on the intensity and frequency of Southern Hemisphere cyclones (Pezza et al. 2008). As an example, when tracking cyclones in the Southern Ocean south of 60°S, Grieger et al. (2018) recognised a positive trend in cyclone frequency in the Austral summer over the 1979-2008 period, which is possibly linked to the positive summer trend of SAM during the same interval (Diamond and Renwick 2015). These observations are supported by many climate modelling studies which have suggested that increased cyclonic activity accompanies the predicted stronger SAM over the Southern Ocean (e.g. Lynch et al. 2006). Furthermore, the observed increase in the number of cyclones around the Antarctic Peninsula in recent years (Lubin et al. 2008) is suggested to be one of the reasons for the rapid warming of the Antarctic Peninsula (Marshall 2007) and a reduction in the sea-ice extent in the region to its west (Parkinson 2019).

Hence, climatic modes (e.g. SAM, SAO) linked to cyclones are modulating Southern Ocean climate through changes in the sea ice. For example, cyclones play a significant role in the heat exchange between sea and air by exerting dynamic and thermodynamic forces on sea ice (Simmonds and Wu 1993). A recent study investigating satellite-tracked sea-ice motion vectors found large and statistically significant trends that are strongly correlated to local cyclonic winds (Holland and Kwok 2012). Raphael et al. (2019) also found that increased cyclonic activity in the Ross Sea has contributed to freshening and increasing sea ice in that region (Simmonds 2015). Holland and Kwok suggested that underestimation of cyclonic winds in Global Climate Models (GCMs) is a major reason behind their consistent failure to simulate increasing Antarctic sea ice, even in the modern satellite-observation period (Zhang 2007, Eisenman et al. 2011, Zhang 2013).

Relating to another major climatic feature in the Southern Ocean domain, the Amundsen Sea Low (ASL), it has been observed that cyclonic flow around this low-pressure centre drives warm mid-latitude winds into the Antarctic Peninsula/Amundsen-Bellinghausen Sea
1.4 State of Reanalyses over the 20th century

(ABS) region, and a cold continental wind over the Ross Sea (Lefebvre et al. 2004, Holland and Kwok 2012, Hosking et al. 2013, Turner et al. 2015). The observed dipole of large-scale expansion and reduction in the Ross Sea and the ABS, respectively, is linked to a combination of strengthening of SAM and deepening ASL (Holland and Kwok 2012, Hosking et al. 2013).

The influence of cyclones is not just restricted to sea ice. The cyclonic systems transport moisture from mid-latitudes to the Antarctic continent, thus playing a significant role in the mass balance of the Antarctic ice sheets (Grieger et al. 2018). Noting the influence of these synoptic systems on the weather and its extremes over the Southern Ocean and in the coastal and continental regions of Antarctica, it is important to detect and quantify cyclonic activity over the Southern Ocean. However, cyclone tracks for the Southern Hemisphere have been relatively under-examined compared with their northern counterparts, for a number of reasons that include both data sparsity and difficult logistics.

Overall, climatic indices and cyclonic intensity and frequency have a major impact on Southern Ocean climate. Therefore, reconstructing historical Southern Ocean climate, in terms of climatic indices and cyclonic activities are important to understand long-term changes in the Southern Ocean climate. Moreover, the large-scale interaction between Antarctic climate and these climatic modes and cyclonic activity is complicated, since each region responds to different atmospheric modes of variability at different time-scales (Lefebvre and Goosse 2008, Raphael and Hobbs 2014, Teleti and Luis 2016). The changing Southern Ocean climate compounds the difficulty in the understanding of the relationship between these modes and sea ice. For example, the effect of ENSO events on sea ice has been found to be significant when SAM is weak or in phase with the driving ENSO event. However, the strength of such time-varying relationships is only based on a short (40 year) length of observations (Stammerjohn et al. 2008, Fogt et al. 2009). Besides, the available observations do not depict conditions in the wider region of the Southern Ocean; to do so, we need global climate datasets.

1.4 State of Reanalyses over the 20th century

Whereas single-location land-based meteorological observations have enhanced the understanding of climate at that place or region, a broader understanding of global or hemispheric climate can be extracted if simultaneous observations from wide regions are analysed systematically. In this regard, long-term multi-ensemble reanalyses are being developed (e.g. 20CRv2c (Compo et al. 2011), CERA-20C (Laloyaux et al. 2018), ERA-20C (Poli et al.
The reanalyses objectively combine observations and the state of a numerical weather prediction model to generate an estimate of the global climate. However, due to the small number of observations available in the Southern Ocean and Antarctica concerning historical climate records, global reanalyses are extremely poor in quality over that region in the pre-satellite period (Schneider and Fogt 2018). Many studies have found reanalyses are not of a quality that is useful climatologically; for example, the two century-long reanalyses CERA20C and 20CRv2c do not attain useful climate-predictive quality until 1979 in CERA20C and 2000 in 20CRv2c (Schneider and Fogt 2018). In addition, many multi-model ensembles generally disagree with observed changes (e.g. in sea ice) in the Southern Ocean for the recent period (since the 1980s), due to unresolved relationships between climate variables which result from the lack of long-term observations in this region (Zunz et al. 2013, Shu et al. 2015, Jones et al. 2016).

In this context of minimal constraining meteorological data for the Southern Ocean, I will describe several notable data-rescue projects and initiatives undertaken to improve reanalyses. These projects identify, digitalise and extract meteorological data from historical data sources, are presented in the next section.

1.5 Data-rescue efforts

To further the development of reanalyses and enhance the data available to them, the international Atmospheric Circulation Reconstructions over the Earth (ACRE) initiative, led by various institutions in UK, USA, Germany and Switzerland, coordinates between the reanalyses and data-rescue communities. The main task of the initiative is to support numerous national and international meteorological organisations and the climate data community to recover historical instrumental surface-terrestrial and marine global weather observations for climate applications. The ACRE initiative has been instrumental in the digitisation and extraction of millions of historical meteorological observations (Fig. 1.3), in collaboration with the International Surface Pressure Databank (ISPD, Cram et al. 2015), the International Comprehensive Ocean–Atmosphere Data Set (ICOADS; http://icoads.noaa.gov), the International Surface Temperature Initiative (ISTI; Rennie et al. 2014) and NOAA's NCDC Climate Database Modernization Program (CDMP). ACRE's contribution to ICOADS dataset could be found in DCK no. 246 and 247 amounting to more than 146,000 observations (Freeman et al. 2017). In the past, ACRE has facilitated data-rescue projects working in partnership with, for example, the British Library, National Archives and Royal Navy, to
image and extract observations from the English East India Company (EEIC) ship logbooks (the 1780s–1830s) and Royal Navy’s Arctic ship logbooks (1818-1825).

Fig. 1.3 The focus areas of ACRE’s regional chapters across the world in conjunction with C3S data-rescue services, along with dates and places of annual international ACRE workshops conducted. Adapted from Brönnimann et al. (2018).

Other data-rescue projects, such as the CLIWOC (Climatological Database for the World’s Oceans; García-Herrera et al. 2005) project, aim to explore the meteorological information contained in pre-1854 ships’ logbooks from around the world. The project digitised the daily data from logbooks in the 1750–1850 period from the voyages of English, Spanish, Dutch and French ships. Although the CLIWOC database was in principle designed to serve climate studies, other onboard events such as sea-ice presence were also fed into the database to make it useful for multidisciplinary studies as well. The main variables extracted for the CLIWOC database consist of the date, geographical position, wind direction, wind force, present weather, sea state, sea-ice reports and, starting from the turn of the 19th century, temperature and air pressure.

Even though wind direction and wind force estimates were qualitative rather than quantitative, such observations when converted to quantitative equivalents are very accurate when compared with modern instruments (Ayre et al. 2015). Other global climate datasets such as ICOADS (discussed below) have standardised and used these historical ship-based observations to estimate wind conditions. Following the CLIWOC project, the RECLAIM project
was established to continue building on the knowledge gained during CLIWOC. Now the RECLAIM project is a contributory project to ICOADS (https://icoads.noaa.gov/reclaim/). There has been considerable interest and encouragement from research organisations in recent times for systematic data-rescue operations. Major organisations, including WMO, NOAA, ECMWF and others, have issued guidelines and initiated major data-rescue projects around the world (WMO 2017, Brönnimann et al. 2018). Notably, the Copernicus Climate Change Service (C3S), implemented by ECMWF on behalf of the European Union, has established climate data rescue services in conjunction with the ACRE initiative (Fig. 1.3).

These data-rescue projects involve experts and specialist researchers working on ships’ logbooks to extract meteorological observations. The method produces high-quality data but is expensive due to the constant engagement of researchers and requires a relatively long time due to manual data extraction. One of the narrowest bottlenecks of historical data extraction has been a lack of reliable and efficient automated processes to deal with hundreds of thousands of weather journals and ship logbooks which are written by hand. Many new archives have been located, catalogued and photographed by the data-rescue initiatives. However, there are at least as many data to be rescued as are currently available in digital archives for the period prior to 1950 (Allan et al. 2011).

As an example, largely forgotten sources of observations in the form of hard copies and images of whaling logbooks from the Southern Ocean contain more than 32,000 images that are waiting to be extracted from the Vestfold Archive, Sandefjord, alone (Wilkinson 2016b). This is partly because traditional Optical Character Recognition (OCR) software performs poorly when applied to tabular hand-written weather reports and logbooks. Newer tools based on Artificial Intelligence are being developed; however, they are still in the prototype stage, and manual extraction remains the only feasible method to extract historical meteorological observations.

The two main challenges, first, a lack of automated tools for data extraction and, secondly, the time-consuming process of manual data extraction, are being addressed by using a citizen-science based approach. An example is the OldWeather project (hosted by Zooniverse, https://www.oldweather.org/), where members of the public are invited to look at the logbook pages of historical ships and transcribe the meteorological observations recorded on them. The OldWeather project focuses on US-based ships and logbooks operating in the Arctic. By contrast, Southern Weather Discovery (also hosted by Zooniverse, https://www.zooniverse.org/projects/drewdeepsouth/southern-weather-discovery) focuses on extracting data from logbooks of ships plying the Southern Ocean. It is overseen by ACRE
Antarctica, C3S Data Rescue Service and NIWA New Zealand. Both these projects and similar other projects (Weather Detective, Weather Rescue; https://www.zooniverse.org/projects/edh/weather-rescue/) have extracted and constructed datasets of high-quality meteorological observations of the Southern Ocean from historical ship logbooks comprising long series of daily-resolution observations.

The extracted observations have proved invaluable for numerical climate reanalyses, leading to newer reanalyses such as the 20CRv3 (Slivinski et al. 2019), and ERA-5 (Hersbach et al. 2019). Many of these reanalyses now span the whole 20th Century, some even providing the state of climate as far back as 1806 (20CRv3). ACRE’s biggest strength is in undertaking project-focused digitisation efforts using regional affiliates, which now span every region and continent, by bringing in experts from different fields (e.g. archivists, historians and climate scientists) to answer a specific research question of historical climate (See Fig. 1.3). The data generated from these activities incrementally improve our understanding of past climate, and more such efforts are required to keep pace with advances in the data assimilations to produce accurate representations of historical climate.

However, before the ACRE and similar reanalyses-focused data-rescue initiatives, many studies attempted to create globally gridded climate datasets (e.g. air-/sea-surface temperature, pressure) which are the basis for many international climate-change detection and adaptation policies from international organisations (e.g. UNFCCC, IPCC). These are now discussed briefly.

Climate analysis and reconstruction using historical ship logbooks have been going on for some decades; however, investigations of polar high-latitudes, and especially the Southern Ocean and Antarctica have only been undertaken quite recently. As early as 1970, Oliver and Kingston (1970) were the first to use historical ships’ logbooks to examine synoptic weather in modern times. They concluded that plotted information from ships’ logbooks is acceptable as an additional source of meteorological evidence, along with land-based weather stations, in the reconstruction of synoptic charts. Soon after that, Moodie and Catchpole (1975) analysed historical weather records of the Hudson Bay Company (HBC) outposts in estuaries of Hudson Bay. Examining the dates of ice freezing and thawing a reasonable estimate of the severity of winter and prevailing temperatures was made.

Such studies have been reviewed by Jones and Bradley (1992) and by BráZdíl et al. (2005). In a similar vein, Lamb (1982) while re-constructing climate throughout human history, forecast that historical ships’ logbooks would provide essential climatic observations for the remotest regions of Earth. Many attempts have been made to use historical climate
observations recorded in diverse formats in scientific studies. The earliest studies (e.g. Moodie and Catchpole 1975) used weather records from a single station to study past climatic changes; however, climate information from point sources was insufficient to infer climatic changes over large remote regions. To understand climate change at the regional-global scale, a systematic aggregation of large numbers of observations collected over many years is required from a network of fairly widely spread observing stations.

There was also a need to systematically organise the collected climatic data into a readily usable global grid format. One of the pioneering studies was Alexander and Mobley (1976), who provided SSTs (Sea Surface Temperatures) and sea ice climatologies for both the Arctic and Antarctic on a $1^\circ \times 1^\circ$ grid. They used the U.S. Navy Fleet Numerical Weather Center’s (FNWC) monthly ice charts and the Navy Hydrographic Office’s monthly mean ice pack limits for Arctic and Antarctic sea ice climatologies, respectively. SSTs were first interpolated into a $1^\circ \times 1^\circ$ grid from the 125×125 Navy grid of FNWC (approx. 1.44°×2.88° lat-long global grid) for the Northern Hemisphere and from the 2.5° grid of Washington and Theil (1970) for the Southern Hemisphere. The resulting gridded data points were merged to form a global SST dataset. There have been many attempts to measure historical temperature changes on a global scale. Relatively more attention has been given to Northern Hemisphere temperature reconstruction; studies such as Jones et al. (1986a) and Jones et al. (1986b), made huge strides in pulling together large amounts of data on a hemispheric scale. However, undertaking the same exercise for the Southern Hemisphere was not only difficult because a large portion of it was covered by the ocean but also because meteorological stations were fewer and often separated by long distances.

In the following section, I will briefly summarise current major spatio-temporal global gridded climate datasets which were generated using the network of land-based weather stations, floating ocean buoys and historical ships’ logbooks, while discussing various methodologies employed to extract, to standardise and to render them into gridded datasets.

1.6 Major global gridded datasets

CRUTEM4 dataset

A major effort to combine land-based climate observations from around the world produced a global dataset, the Climate Research Centre Temperature (CRUTEM4) dataset version 4 (Jones et al. 1986a, Jones et al. 1986b, Jones et al. 2012). In the latest update,
the dataset spans from 1850. The input to the dataset comprises 5583 station records, and the dataset is generated as anomalies of temperature (that is, the difference between observations for a given time and mean temperatures for the reference period, in this case, 1961–1990). The station data are collected from a variety of sources, including several World Meteorological Organization (WMO) and Global Climatological Observation System (GCOS) initiatives, as well as several national and other initiatives (coordinated by National Meteorological Services, NMSs) and scientific publications.

However, the CRUTEM4 dataset does not use the NMSs daily and hourly Synoptic Reports (SYNOP) because SYNOP data are operational in nature, and so are not always extensively quality controlled by NMSs. Secondly, their coverage tends to be denser in regions where the CRUTEM4 dataset already has many sources. Another aspect is data treatment; raw data can only be used to its fullest potential if it is homogenised. Data is said to be homogenised if the variations exhibited by the data are solely the result of the vagaries of the weather and climate (Jones et al. 2012). However, real data is plagued with many inhomogeneities that include: changes in instrumentation, exposure, and measurement technique; changes in station location (both position and elevation); changes in observation times and the methods used to calculate monthly averages; and changes in the environment of the station, particularly with reference to urbanisation that affects the representativeness of the temperature records.

After homogenisation (as described in Brohan et al. 2006, Trewin 2010), the data are aggregated into equally spaced spatial grids. In the CRUTEM4 dataset, the Climate Anomaly Method is used (CAM; Jones 1994), where each station is reduced to anomalies from monthly means calculated for a reference period (1961-1990). This method requires each station to have at least 20 years of values for the reference period, and estimates mean values where possible using nearby stations that have long records. Station temperature anomaly values are then averaged together in a non-weighted fashion for all stations within each $5^\circ \times 5^\circ$ grid box. The resulting grid box time series differ in the number of stations between grid boxes and through time. Large-scale hemispheric and global series is computed by a weighted average of all the grid boxes with data; the weights are the cosine of the central latitude of the grid box to account for the large number of stations in low-/mid-latitudes relative to high-/polar-latitudes.

**ICOADS**

Datasets like the CRUTEM4 enabled a better understanding of climatic changes in the short- and long-term over land. However, the need remained to broaden the understanding
of climate over the ocean, which covers more than 70\% of the globe. For many hundreds of years, sailors observed and later systematically measured weather on the high seas. The Brussels Maritime conference in 1853 marked the start of international efforts to standardise the system of meteorological observations at sea. Even though the standard set of observations was adopted at different times by different countries, meteorological measurements at sea slowly became more uniform.

To make use of these standardised maritime logbooks, a global climate dataset was inaugurated by Slutz et al. (1985), who launched the Comprehensive Ocean-Atmosphere Data Set (COADS), later updated by Woodruff et al. (1987). Now called International-COADS (ICOADS), this brought a large number of Ocean-Atmosphere observations from varied sources together, and the dataset was widely utilised for generating SST climatologies (e.g. Kaplan et al. 1998, Kennedy et al. 2011b, Kennedy et al. 2011a). The observational data were supplied from a variety of national and international sources, including ships’ logbooks from 1854 onwards. Data collected by ships of opportunity (volunteer commercial and research vessels) during 1854-1979, later updated to 2012, were edited and summarised into monthly and annual climatologies.

ICOADS data are in-situ marine meteorological observations, mainly from ships and buoys, and from many different national and international data sources. A wide variety of measured and visually estimated variables are utilised, including all those necessary to estimate ocean surface fluxes (e.g. Josey et al. 1999, Grist and Josey 2003). This long-term global ocean record has supported research in many other scientific domains; recent examples include global atmospheric reanalyses (Kistler et al. 2001, Uppala et al. 2005, Compo et al. 2011, Laloyaux et al. 2018, Slivinski et al. 2019, Hersbach et al. 2019), satellite and in situ blended analyses (Rayner 2003), ground-truthing for remotely sensed or pre-instrumental proxy data (Dunbar et al. 1994), and assessment of global anthropogenic emissions from ships (Corbett et al. 1999, Corbett and Koehler 2003).

ICOADS is currently at release 3 (Freeman et al. 2017), with individual observations from 1662 to 2014, with monthly summaries for 2°×2° grids covering 1800–2014, and for 1°×1° boxes since 1960. The dataset comprises more than 40 core variables including many meteorological parameters, positional and ship-related meta-data. The dataset has been updated in two modes; ‘delay’ mode when all observations are subjected to full quality control steps, and ‘real-time’ mode when not including observations from all sources, allowing for quicker processing.
As expected, combining a very large number of logbooks recorded over many decades involved a major effort. The first release of the COADS dataset combined the Atlas dataset produced by NCDC (National Climatic Data Center) and data from the Historical Sea Surface Temperature (HSST) Data Project. The 1854-1969 period of the Atlas was extended through 1979 using NCDC’s 1970s Decade dataset, and other additions to later years such as buoy and bathy-thermograph data. Other data were included because of their high quality (Ocean Station Vessels) or remote location (South African Whaling). Since 1979, the Global Climate Observing System (GCOS) observations have been collected and utilised.

The major output from enhanced trimming (that is eliminating records which fall outside 4.5 times the standard deviations from the smoothed median) was a record for each 2°-month-year box (with at least one acceptable observation) containing the complete matrix of 19 variables \( \times \) 14 statistics. Afterwards, these basic records were sorted into time order for temporal analysis of the dataset. Instrumental, observational and coding methods, entry instructions and ship construction all have undergone significant changes throughout the period covered by the ICOADS dataset. Methods of observation or instruments in use are not explicitly recorded in many logbooks, leading to random observational data jumps. These inhomogeneities are compounded by variations in data density and by errors that may occur at every stage of observation, digitisation and processing.

Projects such as ICOADS and CLIWOC have painstakingly collected, extracted and arranged vast collections of marine observations spanning over two centuries. However, due to lack of meta-data, inherent systematic bias in observations, and a range of different errors make these datasets unsuitable for studying climate change. Hence, secondary datasets (e.g. HadSST) address the inconsistencies and discrepancies found in the primary datasets and offer a more consistent and more precise picture of global changes.

**HadSST**

An SST dataset, Hadley Centre Sea Surface Temperature dataset version 3 (HadSST3; Kennedy et al. 2011b, Kennedy et al. 2011a), which spans more than 160 years from 1850 to present-month, has global coverage and offers a unique opportunity to study and understand SST changes at regional and global scales. The dataset was generated by combining data provided by updated ICOADS version 2 (Worley et al. 2005) along with data from several additional sources; for example, the World Ocean database (Levitus et al. 2000) and the Marine Data Bank – MOHSST (Met Office Historical SST; Parker et al. 1995). The HadSST3 dataset comprises anomalies which were generated using mean values for a reference period (1961-1990).
Thorough quality-control procedures were followed to remove systematic biases from both SST and marine air temperature (MAT) data and to combine in-situ and satellite SSTs in a consistent manner. The SST data are taken from ICOADS, from 1850 to 1994, and MOHSST observations. The HadSST3 (Kennedy et al. 2011b, Kennedy et al. 2011a) is produced by taking in-situ measurements of SSTs from ships and buoys, rejecting measurements that fail quality checks, converting the measurements to anomalies by subtracting climatological values from the measurements, and calculating a robust average of the resulting anomalies on a $5^\circ \times 5^\circ$ monthly grid. After gridding the anomalies, bias adjustments are applied to reduce the effects of spurious trends caused by changes in SST measuring practices. The uncertainties due to under-sampling and measurement error have been calculated for the gridded monthly data, as have the uncertainties on the bias adjustments following the procedures described in Kennedy et al. (2011b) and Kennedy et al. (2011a).

**HadISST**

The Hadley Centre Global Sea Ice and Sea Surface Temperature dataset (HadISST; Rayner 2003) is a combination of monthly globally complete fields of SST and sea ice concentration starting from 1871. It is generated by using many known sources of ships’ logbooks, making it the most comprehensive sea ice dataset to date. HadISST uses a reduced-space optimal interpolation applied to SSTs from the Marine Data Bank (mainly ship tracks) and ICOADS through 1981, and a blend of in-situ and adjusted satellite-derived SSTs for 1982-onwards. The sea ice data are taken from a variety of sources including digitised sea ice charts and passive microwave retrievals. HadISST is primarily intended to be used as boundary conditions for atmospheric models. The Met Office HadISST 2.1 project (Titchner and Rayner 2014) aims to fill in spatial and temporal gaps left by the previous generation HadISST 1.0 dataset. The released dataset has global (Arctic and Antarctic) sea ice concentrations from 1850 to 2007 on a $1^\circ$ grid. The dataset contains only sea ice concentration with no SST observations.

**HadCRUT4**

A land-ocean combined dataset is the Hadley Centre-CRU Temperature dataset (HadCRUT4). It is based on CRUTEM4 and HadSST3, which are discussed above. CRU has developed land dataset (CRUTEM4) in conjunction with the Hadley Centre, and the SST dataset (HadSST3) was developed solely by the Hadley Centre. The combined dataset is available on a $5^\circ \times 5^\circ$ grid from 1850 to present (Morice et al. 2012).

**HadSLP2**
HadSLP2 global sea-level pressure dataset has been painstakingly produced using a large number of monthly land and marine pressure observations on a 5° latitude-longitude grid from 1850 to 2004 (Allan and Ansell 2006). The primary data sources include the Met Office Library and Archives, annual and monthly reports from various Ministries and Departments from many countries, and WMO meteorological statistics handbooks. In addition, observations from the Global Historical Climate Network (GHCN), versions 1 and 2 (Vose et al. 1992), and the Global Climate Observing System (GCOS) Surface Network (GSN), along with observations from many smaller data-rescue projects, were combined to create a global dataset of pressure observations. More than 2000 stations have been included, out of which 615 stations have more than 100 years of records. The project had a focus on improving the coverage over conventionally sparse regions such as Africa, Asia and much of the Southern Hemisphere, especially Antarctica. The marine observations were sourced from ICOADS (Worley et al. 2005). The quality checked marine observations were gridded and combined with land-based observations.

 Antarct i c MSLP datasets

 Jones and Wigley (1988) constructed a gridded MSLP dataset across the Antarctic and Southern Ocean. They utilised MSLP data from the World Meteorological Centre in Melbourne from 1972 to 1985 and data published along with the NOTOS system for 1957-1962. Antarctic station data were used in conjunction to reconstruct atmospheric variability in the high southern latitude. Jones (1990) extended the MSLP dataset of Jones and Wigley (1988) back to 1951 and then 1911 using different sets of stations. In addition, Jones and Lister (2007) produced a comprehensive review of four historically reconstructed MSLP datasets for the Southern Hemisphere and found Jones (1991)’s dataset to be in general agreement with the new datasets. In addition, Fogt et al. (2018) reconstructed Antarctic coastal station MSLP data back to 1905 using statistical reconstruction methods.

 ISTI

 The International Surface Temperature Initiative (ISTI; Thorne et al. 2011) produced world’s most comprehensive and provenance-tracked set of land-based meteorological database called, Global Land Surface Meteorological Databank version 1 (Rennie et al. 2014). It holds monthly timescale mean, maximum, and minimum temperature for approximately 40,000 stations globally.

 This is the global repository for all monthly timescale individual land surface observations from the 1800s to present and uses data deriving from sub-daily, daily, and monthly
observations. It brings together data from more than 45 sources to create a single merged dataset. It is used in the creation of various integrated global temperature resources, most notably Global Historical Climatology Network Monthly (GHCN-M) v4 (Menne et al. 2018).

**ISPD**

The International Surface Pressure Databank (ISPD version 2) is the world’s most extensive collection of surface and sea-level pressure observations (Cram et al. 2015). It merges data from primarily three data sources: observations from land stations, marine observing systems, and tropical cyclone best-track pressure reports. The land-based observations are derived mainly from the Integrated Surface Database (Smith et al. 2011), which consists of national and international hourly and synoptic surface pressure observations collected from many sources. The observations are quality checked, and duplicates are eliminated prior to inclusion into ISPD. The marine component is derived from ICOADS release 2.5, while the cyclone core pressures are derived from the International Best Track Archive for Climate Stewardship (IBTrACS; Knapp et al. 2010).

The IBTrACS dataset consists of post-season global tropical cyclone best-track position and intensity observations, and reports collected from each of the World Meteorological Organization (WMO) Regional Specialized Meteorological Centers, Tropical Cyclone Warning Centers, and other national agencies. These observations are based on satellite images, meteorological stations and tracking methods used by each reporting agency, which are aggregated into a composite cyclone database. The ISPD dataset is a collection of raw observations and, unlike other global gridded datasets, it consists of observations from 1768 to present.

More generally, a large number of climate datasets are available offering differing resolution, data handling and interpolation methods; for example, Reynolds’s OISST, NASA’s GISTEMP, and Kaplan SST (Kaplan et al. 1998). Despite the availability of large numbers of datasets, the Southern Ocean and Antarctica remain the regions that have the least climatological data (Freeman et al. 2017). Even the most comprehensive instrumental datasets, for example, HadCRUT4 (Morce et al. 2012) and Kaplan SST (Kaplan et al. 1998), have null data or significant gaps in the regions southward of 40° S. Of all the global climate datasets discussed above, none of the datasets has anything approaching sufficient data coverage in the Southern Ocean.
1.7 Lack of meteorological observations in the Southern Ocean

To overcome the dearth of data in the Southern Ocean, studies like Jones (1990), Parkinson (1990) and more recently Edinburgh and Day (2016) have utilised observations taken during Antarctic expeditions. Jones (1990) has analysed air temperature records from 26 expeditions to Antarctica that have wintered for at least nine months, between 1898 and 1958. A map of temperatures is generated, dividing the whole region into a $5\degree \times 5\degree$ lat-long grid extrapolating the temperatures using a cubic spline smoothing function. Interestingly, Parkinson (1990) explored ships’ logs maintained by pioneering explorers (e.g. Cook, Bellingshausen, Wilkes, and Ross) in the 18th-19th centuries. She observed that ice-edge records agree well with satellite records from 1973-76, implying no large changes have occurred in the meantime. A recent article (Edinburgh and Day 2016) has used historical Antarctic ships’ logbooks from the period 1897-1917, comparing ship records to the more recent satellite sea-ice observations. They argue that the sea ice extent has not changed substantially in the last 100 years or so; however, such conclusions rely on a small set of expeditions (total of 11). Nevertheless, this study provides evidence that historic Antarctic ships’ logbooks are very valuable to study long-term environmental changes in the Southern Ocean.

Generally, historical temperature, circulation and sea ice reconstructions in the Southern Ocean and Antarctica differ from the observations in the modern instrumental period. Fan et al. (2014) suggested that there were colder conditions and the possibility of greater sea ice extent before the start of the modern satellite era in 1979. This is supported by recent reconstructions of recovered early satellite imagery, which indicate that in the early 1960s, Antarctic sea ice could have been more extensive than at present (Gallaher et al. 2014).

Some studies employed meteorological observations from Antarctic stations to examine MSLP variability. King and Harangozo (1998) examined meteorological records for the Antarctic Peninsula and suggested that there has been an increase in the northerly component of the atmospheric circulation over the Peninsula since the late 1950s. Also, sea-ice extent data available indicate that the ice edge to the west of the Peninsula lay to the north of recently observed extremes during the very cold conditions prevailing in the late 1950s. General conclusions from the numerous studies presented above are that climate in the Southern Ocean and Antarctica was different in the first half of 20th century to that of the later period; sea-ice extent was larger before the 1970s, and temperatures were cooler during the 1950s. Reported long-term changes in the Southern Ocean climate have been debated due to lack
of reliable data in the global historical datasets in these regions. The nature of long-term changes is also ambiguous as conflicting trends are reported, and the problem is compounded further due to the relatively short study periods.

1.8 Significance of whaling logbooks

There is a source of ancillary observations in the form of whaling ships’ logbooks in the Southern Ocean, which contains large numbers of meteorological observations and can assist in reconstructing historical Southern Ocean climate. As discussed earlier (Sections 1.5 and 1.6) the understanding of long-term variability of the Southern Ocean climate is hampered by the dearth of continuous meteorological observations. Even though the earliest observations date to the 18th Century from exploratory voyages such as that of Capt. James Cook’s *Endeavour* expeditions, such observations are few (ICOADS3.0 DCK 246, Freeman et al. 2017). All of these data are temporally and spatially sparse, and serve primarily to demonstrate the high spatial variability of Southern Ocean climate over a variety of time scales. Whaling ships were the only ships regularly plying the Southern Ocean in the first half of the 20th Century (Tønnessen and Johnsen 1982).

![Leith Harbour whaling station](image-url)  
Fig. 1.4 Leith Harbour whaling station on South Georgia with whaling ships in the port laid up for the winter and "Discovery 2" at jetty in the harbour in the 1930s. (© South Georgia Govt., SG 2020)
1.9 Thesis aims and structure

Commercial whaling in the Southern Ocean was carried out for the better part of the 20th Century until the international moratorium on whaling was put in place in 1982. The whaling ships were the most common ships operating in the Southern Ocean, and due to the nature of whaling, ships returned to the same regions year after year, creating the sustained collection of weather observations of the region and producing stationarity amid the largely transitory nature of marine observations. One of the largest whaling companies to operate in the Southern Ocean was the Christian Salvesen Whaling Company, based at Leith, Scotland (Vamplew 1975). The company engaged in whaling operations around the world, operating shore-based stations at Grytviken and Leith Harbour in South Georgia (Fig. 1.4), and numerous factory ships in the Southern Ocean.

For political-economic reasons, the 1930s and 1950s were the most prolific periods of whaling in the Southern Ocean (Tønnessen and Johnsen 1982). Therefore, ship-based observations from the Christian Salvesen Whaling Company’s logbooks of whaling ships operating in the Southern Ocean in the historical period (1930s and 1950s) are investigated in this thesis, as they offer a valuable window into the historical climate of the region and fill in knowledge gaps left by previous data-rescue efforts (e.g. ICOADS). In addition, according to the RECLAIM project (Wilkinson et al. 2011), a contributing project to ICOADS, whaling ships’ logbooks have not been prioritised in the current data-rescue efforts, nonetheless contain a large number of meteorological observations especially in the Polar regions.

1.9 Need for this work and thesis aims and structure

The need for more observations in the Southern Ocean is clear, and a lack of long-term meteorological observations in the region hinders the investigation of climate variability and change at high southern latitudes (Jones et al. 2016). In this context, data from whaling logbooks should be fully utilised to fill in the data gaps in the Southern Ocean. The extracted observations can then be used to reconstruct long-term circulation changes and historical cyclonic frequency in the Southern Ocean.

This thesis brings under-utilised whaling logbooks to the forefront and demonstrates that the quality of the meteorological observations contained in them is on a par with other observational datasets. In addition, the usefulness of historical whaling observations is further demonstrated by assimilating the historical observations into multi-ensemble numerical reanalyses of past climate to improve the state of reanalyses in the historical period over the Southern Ocean. This thesis will demonstrate that whaling logbooks contain
high-quality meteorological observations that can be used to reconstruct historical climate and also, despite the relatively small number of observations extracted, these observations can have a significant impact on the understanding of Southern Ocean climate.

The thesis has two main aims: first, to create a historical climate dataset from the logbooks of Christian Salvesen whaling ships traversing the Southern Ocean in the 1930s and 1950s; and, secondly, using this climate dataset to reconstruct variability in the Southern Ocean climate during the historical period (1930s and 1950s) and comparing these decades with modern observations. I begin by describing the process of data acquisition from whaling logbooks and the subsequent error-correction and standardisation required to produce a historical climate dataset. After that, I validate resultant meteorological dataset against existing station-based observations. In addition, I reconstruct MSLP variability in the Southern Ocean and a SAM index from the whaling dataset. I estimate cyclonic frequency during the period of investigation, using a semi-supervised cyclone identification and tracking algorithm, which is then compared with modern cyclonic frequency. Finally, I assimilate historical observations from the whaling dataset into 20th Century numerical reanalyses to improve the representation of MSLP fields over the Southern Ocean during the study period.

The thesis is structured as follows:

Chapter 2 describes the process of data extraction from Christian Salvesen whaling logbooks. Due to significant failures of Optical Character Recognition (OCR) software systems to extract observations from tabular logbook pages, a manual data extraction approach is adopted. The extracted climate variables include air temperature, air pressure, wind conditions and others, along with positional and time information. The extracted data are standardised, error-corrected and formatted in an internationally accepted data structure in order to be useful for further analysis. This Chapter has been published as Teleti et al. (2019), and the dataset created is made publicly available in international repositories.

Like any historical data source, the resultant meteorological dataset needs to be validated against existing observations. In this case, early Antarctic and sub-Antarctic station-based meteorological observations are used to validate the whaling dataset. In addition, MSLP variability in the Southern Ocean during the months of observation is analysed and compared with modern climatology. To capture hemispheric-scale atmospheric variability, a longer marine observations-based Southern Annular Mode (SAM) climate index is constructed in Chapter 3.
1.9 Thesis aims and structure

Having demonstrated the quality of the whaling dataset, an attempt is made to recognise fine-scale weather features from the whaling dataset in Chapter 4. Synoptic scale weather systems (extra-tropical cyclones) are identified and tracked through space and time using historical observations from the logbook-derived climate dataset. A semi-supervised cyclone identification and tracking algorithm is developed to estimate cyclonic frequency using the whaling dataset. The estimated number of cyclones for each whaling season, historical cyclonic frequency, is then compared to modern cyclonic frequency.

The penultimate chapter (Chapter 5) is focused on improving the representation of Southern Ocean MSLP fields by assimilating individual observations in the whaling dataset into numerical reanalysis models (CERA-20C and 20CRv2c). A computationally inexpensive ‘offline’ data assimilation method is developed and used to reduce uncertainty in the Southern Ocean. The improvement in the reanalyses post-assimilation is presented to assess the impact of the historical data.

The final chapter of the thesis (Chapter 6) draws out the main conclusions of the work and makes suggestions for further work.
Chapter 2

Climate dataset from whaling ships’ logbooks

In this Chapter, I demonstrate how to extract historical meteorological data from whaling logbooks to create a historical climate dataset.

2.1 Introduction

The Southern Ocean is the least documented climatic region of the globe (Jones et al. 2016). Much of the climate data over the Southern Ocean collected in the last 40 years or so is from polar orbiting satellites. Before the advent of satellites, meteorological information was drawn largely from exploratory expeditions to Antarctica and the Southern Ocean. A small number of scientific stations were built on the Antarctic Peninsula and sub-Antarctic islands starting from the early twentieth century and, with the impetus of the International Geophysical Year (IGY) 1957-58, many more scientific stations on the Antarctic coast were established. To understand long-term climatic changes in Antarctic climate, many previous studies have turned to meteorological measurements taken at these few early stations (e.g. Turner et al. 2005; Chapman and Walsh 2007; Steig et al. 2009; Nicolas and Bromwich 2014; Fogt et al. 2018). Due to near-continuous measurements taken at these stations since the 1950s, a clearer picture of climatic processes over Antarctica has emerged. However, understanding of climatic patterns and processes over the Southern Ocean remains less clear. Many multi-model ensembles generally disagree with observed changes (e.g. in sea ice) in the Southern Ocean for the common period (since the 1980s), due to the unresolved
relationship between climate variables which results from the lack of long-term observations in this area (Zunz et al. 2013; Shu et al. 2015; Jones et al. 2016).

To address the issue of a lack of meteorological observations over the oceans in general, many attempts have been made to systematically assemble marine observations combining different data sources taken on-board scientific, commercial and cruise vessels and from drifting and fixed ocean buoys. One of the largest such efforts is ICOADS Release 3.0 (International Comprehensive Ocean-Atmosphere Data Set, Freeman et al. 2017), which is the most comprehensive dataset that combines data from in-situ marine meteorological observations, mainly from ships and buoys, and from many different national and international data sources. It is described in detail in Chapter 1.

A dataset covering both land and ocean regions, the Hadley Centre-CRU Temperature dataset (HadCRUT4, Morice et al. 2012), has been developed by the Climate Research Unit (CRU) at the University of East Anglia in conjunction with the Hadley Centre (UK Meteorological Office), in which the marine component of the dataset is derived from ICOADS. A further dataset focusing on surface and sea-level pressure observations, the International Surface Pressure Databank (ISPD, version 2, Cram et al. 2015), contains observations from land stations, marine observing systems, and tropical cyclone best-track pressure reports. Similarly, the marine component of this dataset is derived from ICOADS.

Despite tremendous advances in the extraction and assimilation of data to produce global climate datasets, large regions of the Southern Ocean remain poorly represented, with a heavy reliance on observations from early Antarctic and Southern Ocean expeditions as the primary sources for these datasets. Logbooks from these voyages do provide the first-ever weather observations in these regions but are not sustained over time. Other sources of logbooks from commercial, fishing and whaling vessels traversing the Southern Ocean have been largely overlooked. The North American and European whaling industry focused on the Southern Ocean soon after over-fishing led to the collapse of suitable stocks in the northern high-latitudes (Tønnessen and Johnsen 1982). The first whaling fleet, the Dundee whaling expedition, is known to have visited the Falkland Islands in 1892-93 (Headland 2009). Since then until the 1980s, except for the years during two world wars, whaling ships hunted and caught whales almost every year in the Southern Ocean. A key advantage of using whaling logbooks as a source of meteorological data is that vessels usually visited the same whaling grounds year after year, providing sustained temporal coverage in these regions.

The reasons for under-utilisation of this source may be because, as many of these logbooks are housed in different countries, private and commercial whaling logbooks are a relatively
low priority compared to the more accessible national Antarctic expeditions in digitalisation efforts, combined with wide usage of non-English languages in the logbooks. Also, concerns raised regarding the quality of whaling data made the digitalisation of whaling logbooks a low priority. For this study, I have chosen to extract and assimilate a large number of observations from commercial whaling ships to create a historical meteorological dataset. Although the period of interest was from the early 20th Century to the International Whaling Commission’s whaling moratorium in 1986, the spatial and temporal span of the data recovered was ultimately dictated by the data sources located and accessed. To locate such logbooks, I have consulted a report published by the RECLAIM project (https://icoads.noaa.gov/reclaim/), a contributory project to ICOADS, which listed all the identified archives of ships’ logbooks in the Southern Ocean. It was found that the Centre for Research Collections, University of Edinburgh, contains a limited number of logbooks of the Christian Salvesen Whaling Company, a British whaling interest that operated a number of harbour and ship-based production facilities in the Southern Ocean from 1908 to 1963 (Vamplew 1975).

Fortunately, the logbooks have been scanned into 2,700 digital images by the RECLAIM project and were made available for this study (Wilkinson 2016a). These logbooks are from whaling expeditions undertaken during the 1930s and 1950s, and I use these logbook images to extract meteorological observations to construct a climate dataset of the Southern Ocean. The time period also reflects the two most prolific whaling decades in the 20th Century, shaped by political and economic conditions (Jackson 1978, Tønnessen and Johnsen 1982). The newly extracted historical data are stored and made available in an internationally accepted format to streamline assimilation with existing datasets, preserving and extending existing international climate datasets. To bring historical observations to the same standard as modern ones, historical observations are standardised and homogenised into modern units. In the following sections, I show the detailed methods of error detection-correction and standardisation for each variable or group of variables in the dataset, along with suitable storage formats. More details about observations and procedures used to make the meteorological measurements, and also efforts to convert those observations into modern units, are discussed in Section 2.2. After passing observational data through stringent quality-control checks and standardisation processes, the resulting dataset is presented in Section 2.3. I then offer conclusions and future research tasks as a result of the newly created dataset.
2.2 Data and Methods

2.2.1 Description of logbooks

The British Admiralty, with the help of the Meteorological Office’s Marine Observers Handbook (MOH) (HMSO 1930, HMSO 1950), issued meticulously detailed instructions on all aspects of weather observation-taking and record-keeping, including lists of instruments and observable parameters, and the methods and frequency with which these weather observations were to be taken. The observations were usually made by ships’ navigating officers, first officers or other experienced scientific personnel on board. The essential set of observed parameters include air temperature, wind conditions, sea-level pressure, sea state and a general description of the weather, with time, date and position.

Logbooks in the current collection contain observations from 18 years or whaling seasons from the two decades of the 1930s and 1950s. Hand-written Chief Officers’ (commonly known as the deck) logbooks are the most common type of logbook in the current collection (Table 2.1). Deck logbooks were the principal source of information on navigation and weather observations, and were duly retained for legal and insurance purposes (Fig. 2.2). A small number of catch books (Fig. 2.1) and H1-9 reports typed from original logbooks, issued by the British Ministry of Transport and Civil Aviation and US Hydrographic Office, respectively, are also present. The catch books recorded daily whale-catch numbers, the amount of blubber processed, and corresponding oil produced; they also include records of weather and positional information.

All these documents record observations at noon except for deck logbooks, which record 4-hourly observations throughout the season. The logbooks also contain reports of floating ice in the form of icebergs and sea ice. Sightings of sea ice, and more often icebergs, are found in both types of logbooks, more so in the earlier Norwegian-language logbooks which describe different types of sea ice and icebergs in Norwegian terms which can only be approximately translated into English. Even though there were a moderate number of observations concerning sea ice and icebergs and their distance and direction relative to the ship, consistency in reporting is poor. That is unsurprising as sea-ice types were not standardised at that time, and type-ambiguity can produce misleading results. Hence, the sea-ice and iceberg information was not included in the preparation of this dataset; however, such information can be obtained from the authors on request.
2.2 Data and Methods

Fig. 2.1 A typical catch book used in the 1930s, recorded in Norwegian. Day of the week, Date, Midday position, Weather description including Barometer and Temperature observations from columns 1 to 4. In addition, Number of whales and their species, Amounts of blubber harvested and the oil extracted for each day, are recorded from columns 4 to 6. The top of logbook records Name of the vessel and start and end of week recorded on this page. Column 7 records replenishments and column 8 records detailed description of whales caught e.g. species, length, sex, whether pregnant or lactating etc. (Wilkinson 2016a)
Fig. 2.2 A typical Chief Officer’s whaling logbook issued in the 1950s, presented logbook page is from Southern Venturer during 1952-53 whaling season which, describes Time, Speed of the Vessel, Compass courses (magnetic variation wrt to True North), Wind direction and force, Sea-state, Barometer, Temperature observations and general-weather remarks, from Left to Right columns. The top of the logbook records Name of the Vessel and Date, hence this logbook records four sub-daily observations. The middle of the page records the position of the ship at Noon. The bottom of logbook records names of Ship’s Master recording the logbook and counter-signed by the Chief Officer. (Wilkinson 2016a)
Table 2.1 List of vessels, whaling seasons, start-end dates, length of logbook observations and type of documents used in this study of Southern Ocean whaling logbooks from the Christian Salvesen Whaling Company.

<table>
<thead>
<tr>
<th>Name of the vessel</th>
<th>Whaling season</th>
<th>Start date</th>
<th>End date</th>
<th>Length of the season(days)</th>
<th>Type of document</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1939-40</td>
<td>17-Dec-1939</td>
<td>17-Mar-1940</td>
<td>91</td>
<td>Catch Book</td>
</tr>
<tr>
<td><strong>New Sevilla</strong></td>
<td>1932-33</td>
<td>27-Oct-1932</td>
<td>03-Apr-1933</td>
<td>158</td>
<td>H1-9</td>
</tr>
<tr>
<td></td>
<td>1933-34</td>
<td>29-Oct-1933</td>
<td>09-Mar-1934</td>
<td>131</td>
<td>H1-9</td>
</tr>
<tr>
<td></td>
<td>1934-35</td>
<td>23-Nov-1934</td>
<td>22-Apr-1935</td>
<td>150</td>
<td>H1-9</td>
</tr>
<tr>
<td></td>
<td>1939-40</td>
<td>09-Dec-1939</td>
<td>10-Mar-1940</td>
<td>92</td>
<td>H1-9</td>
</tr>
<tr>
<td><strong>Salvestria</strong></td>
<td>1931-32</td>
<td>10-Jan-1932</td>
<td>25-Mar-1932</td>
<td>75</td>
<td>Catch Book</td>
</tr>
<tr>
<td></td>
<td>1933-34</td>
<td>29-Oct-1933</td>
<td>09-Mar-1934</td>
<td>131</td>
<td>Catch Book</td>
</tr>
<tr>
<td></td>
<td>1934-35</td>
<td>24-Nov-1934</td>
<td>03-Apr-1935</td>
<td>130</td>
<td>Catch Book</td>
</tr>
<tr>
<td></td>
<td>1935-36</td>
<td>18-Nov-1935</td>
<td>17-Feb-1936</td>
<td>91</td>
<td>Catch Book</td>
</tr>
<tr>
<td><strong>Sourabaya</strong></td>
<td>1932-33</td>
<td>21-Sep-1932</td>
<td>06-Mar-1933</td>
<td>166</td>
<td>Catch Book</td>
</tr>
<tr>
<td></td>
<td>1933-34</td>
<td>10-Sep-1933</td>
<td>06-Mar-1934</td>
<td>177</td>
<td>Catch Book</td>
</tr>
<tr>
<td></td>
<td>1934-35</td>
<td>01-Dec-1934</td>
<td>31-Mar-1935</td>
<td>120</td>
<td>Catch Book</td>
</tr>
<tr>
<td><strong>Southern Harvester</strong></td>
<td>1950-51</td>
<td>30-Nov-1950</td>
<td>09-Mar-1951</td>
<td>98</td>
<td>Deck logbook</td>
</tr>
<tr>
<td></td>
<td>1951-52</td>
<td>22-Nov-1951</td>
<td>14-Feb-1952</td>
<td>84</td>
<td>Deck logbook</td>
</tr>
<tr>
<td></td>
<td>1953-54</td>
<td>17-Dec-1953</td>
<td>22-Mar-1954</td>
<td>95</td>
<td>Deck logbook</td>
</tr>
<tr>
<td></td>
<td>1956-57</td>
<td>18-Dec-1956</td>
<td>19-Mar-1957</td>
<td>91</td>
<td>Deck logbook</td>
</tr>
<tr>
<td></td>
<td>1959-60</td>
<td>11-Dec-1959</td>
<td>07-Apr-1960</td>
<td>118</td>
<td>Deck logbook</td>
</tr>
<tr>
<td><strong>Southern Venturer</strong></td>
<td>1950-51</td>
<td>09-Dec-1950</td>
<td>12-Mar-1951</td>
<td>92</td>
<td>Deck logbook</td>
</tr>
<tr>
<td></td>
<td>1953-54</td>
<td>07-Dec-1953</td>
<td>19-Mar-1954</td>
<td>102</td>
<td>Deck logbook</td>
</tr>
<tr>
<td></td>
<td>1955-56</td>
<td>01-Feb-1956</td>
<td>30-Apr-1956</td>
<td>89</td>
<td>Deck logbook</td>
</tr>
<tr>
<td></td>
<td>1956-57</td>
<td>09-Dec-1956</td>
<td>07-May-1957</td>
<td>149</td>
<td>Deck logbook</td>
</tr>
<tr>
<td></td>
<td>1957-58</td>
<td>06-Feb-1958</td>
<td>09-May-1958</td>
<td>92</td>
<td>Deck logbook</td>
</tr>
<tr>
<td><strong>Svend Foyn</strong></td>
<td>1940-41</td>
<td>30-Jan-1941</td>
<td>14-May-1941</td>
<td>104</td>
<td>Catch Book</td>
</tr>
</tbody>
</table>
The opportunity to use automated Optical Character Recognition (OCR) software to extract data from these records is severely restricted due to the frequent use of cursive writing, the presence of irregular abbreviations and the tabular structure of the logbook pages themselves. Trial OCR runs produced very high misread rates, even with training data, in a format that was very time consuming to edit manually. In view of the quality of data acquired, and the time required to produce it using OCR methods, manual extraction was the preferred method for obtaining data from the logbooks. I have extracted close to 12,000 unique data records. Each record contains a number of positional, meteorological and meta-data fields, including latitude/longitude, air temperature, wind conditions, sea-level pressure, vessel name, port of registry, among others. All extracted raw data were added to a relational database for simple and systematic access.

2.2.2 Instructions and error detection/correction

In the following sections, I consider the instructions issued by the Admiralty/MOH to understand better the procedures followed in acquiring these shipboard meteorological observations. The use of these instructions is two-fold; first, instructions can point to sources of error and, secondly, they preserve the meta-data of observations, which could be used in future for comparing observations taken using different methods and adjusting them as necessary. Manually extracted data are not without their own challenges. The most common type of error is the gross or observational error; that is, misreading the numbers on instrument scales, together with faulty recording in logbooks which leads to typographical and transcriptional errors. To ensure internal consistency and temporal coherency, statistical tests were employed to identify cases of zero-variance (‘consecutive identical values’) or high variance (‘outliers’) for intra-day and consecutive inter-day observations. Once such cases were identified, they were either made missing or replaced by suitable value (usually the mean of temporally neighbouring values).

Positional Information

Each meteorological record must contain a valid position to be useful. The ship’s position (latitude and longitude) was observed and recorded once a day at noon as a usual practice; hence, noon position is assigned to all the observations taken during the preceding and following 12 hours. The Admiralty Manual (Chapter I, Admiralty 1938) instructs that positional observation were to be taken with reference to true north rather than magnetic
north. The latitudinal and longitudinal position was recorded in degrees and minutes of four cardinal directions, which has been now converted into the degree and decimal system to facilitate further processing. A spatial plot of all raw data points (Fig. 2.3a) shows that some locations have unrealistically large latitudinal and longitudinal differences from one day to the next, suggesting erroneous locations.

Distance travelled between two positions (taken at mid-day of consecutive days) is considered a good indicator of spurious jumps in positions. To this end, all observations belonging to a ‘Ship ID’, which is a unique combination of vessel name and season, were queried from the database, and distance between neighbouring positions was calculated. The distance was measured in nautical miles (nm) following ‘rhumb lines’, as it was common practice to travel along rhumb lines at a constant compass bearing, rather than following arcs of a great circle. Assuming distance data follow a normal distribution, most of the values would be within three times the standard deviation ($\sigma$) from the mean ($\bar{x}$); any values outside of this bracket were treated as suspicious.

According to the central-limit theorem, the higher the difference between an individual value and the data mean, the more likely the value is to be an artefact of gross and/or transcriptional errors. I used the Generalized ESD (extreme studentised deviate) test (Rosner 1983) to flag outliers for each unique Ship ID group. The test removes the observations that maximise $R_i = \max_i \left| \frac{x_i - \bar{x}}{\sigma} \right|$, which is the spread of individual values away from the mean value. The test then re-computes a number of R-values depending on estimates of the maximum number of outliers in the dataset. The exact number of outliers and their positions within the dataset are determined by finding the largest $i$ such that $R_i > \lambda_i$, $\lambda_i$ is the critical t-value of R.

When outliers were plotted (not shown), most clustered together except for a few with very large deviations. Upon closer inspection, it was found that un-clustered outliers were very likely to be erroneous values; hence, they were replaced by the average of temporally neighbouring values. However, clustered groups of outliers showed an interesting pattern; almost all were from the days either at the start or end of the logbook entries. This could be explained by the fact that ships covered large distances to and from whaling grounds and their resupply base on South Georgia at the beginning and close of the whaling season (Supplementary Fig. A.1). For the rest of the season, ships were usually drifting within whaling grounds and covering smaller amounts of daily mileage. I re-performed the statistical test on the corrected data, and no outliers were flagged barring values at the start and end of whaling seasons which confirms that distance data were a heterogeneous mixture of large and small distances. Corrected positions are shown in Fig. 2.3b.
Time information

Time information is also a part of the set of variables that must be present for each data point. Time-keeping instruments were supplied to all British whaling ships for navigational purposes from the Royal Observatory, Greenwich, or from the nearest chronometer depot (Chapter XI, Admiralty 1938). Each ship was allowed three chronometers and one deck watch to be used for day-to-day record keeping and navigation purposes. Navigating officers on board carefully installed and maintained the chronometers in a designated room. Great effort was made to keep correct GMT time as an essential aid to navigation. Each of the three chronometers was compared with the other two, and readings from two chronometers showing near-identical time were taken as the correct GMT time and used to set the deck watch. If possible, other methods of time-keeping (e.g. wireless time signals, telegraphic time signals and astronomical observations) were also used to correct the on-board chronometers.

The longitudinal position was computed by following Admiralty Navigation Manual instructions (Misc. Chart 86, Admiralty 1938). The globe was divided into 24 time zones, each spanning 15° of longitude. Each ship’s local noon-time was set as the time when the sun reached its highest point in the sky. The difference between deck (GMT) time and local
time was measured, and, if it was positive, then longitude was calculated to be 15° or its fraction East for each hour of difference and vice-versa if it was negative. Conversely, GMT time can be determined if the longitude of the ship’s position and local time are known.

Whaling ships operated in two different time frames: deck time and factory or kitchen time, for operational reasons. The fact that ships remained in a time-zone for many days or weeks on end while catching whales required the whaling operations to align to the particular local time-zone. Crew shifts, meal times and other on-board activities followed local time; however, all ships were recommended to use four principal hours 0000, 0600, 1200 and 1800 GMT (deck time) for the observations and recording of meteorological parameters (Chapter XIV, Admiralty 1938). Hence, it is assumed that the times recorded in deck logbook are GMT whereas catch book entries followed local time, which changed when the ship passed from one time-zone to another. Each recorded time entry in the catch book was placed in its respective time-zone according to the ship’s longitudinal position. The local time was then converted into GMT using the procedure outlined earlier.

**Wind Conditions**

Before the widespread use of instrumental anemometers, the direction and force of the wind were visually estimated. Wind direction was specified as a point of the true compass from which wind blows, and was observed to the nearest true compass point (Chapter XIV, Admiralty 1938). The wind force was expressed by means of the Beaufort wind scale (Simpson 1906), a 13-point numerical scale devised in 1808 and named after Admiral Sir Francis Beaufort. The Beaufort scale was used to record the wind force in the deck logbooks and catch books. A companion table was printed in the preface of each logbook, which supplied conversion scales and visual criteria to aid observers on a ship’s deck (Table 2.2).

In our present collection, some early logbooks use wind terms to describe wind strength. All of the encountered wind terms were resolved to one of the levels of the Beaufort scale. Once wind strength is established in terms of the Beaufort scale, it can be expressed easily in knots or m/s. Thereafter, wind force and direction data were separately tested for outliers using the Generalised ESD test. A small fraction (less than 0.5%) of the values for each Ship ID were found to be outliers. Erroneous values were replaced by the average of neighbouring values, if available, or otherwise made null in the database.
Table 2.2 Beaufort wind scale with description and visual criteria (Chapter XIV, British Admiralty 1938)

<table>
<thead>
<tr>
<th>Beaufort Number</th>
<th>Wind speed (NM per hour or knots)</th>
<th>Wind description</th>
<th>Visual criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>Less than 1</td>
<td>Calm</td>
<td>Sea mirror-smooth</td>
</tr>
<tr>
<td>1</td>
<td>1-3</td>
<td>Light air</td>
<td>Small wavelets like scales; no foam crests</td>
</tr>
<tr>
<td>2</td>
<td>4-6</td>
<td>Light breeze</td>
<td>Waves short and more pronounced; foam</td>
</tr>
<tr>
<td>3</td>
<td>7-10</td>
<td>Gentle breeze</td>
<td>has glassy appearance</td>
</tr>
<tr>
<td>4</td>
<td>11-16</td>
<td>Moderate breeze</td>
<td>Waves are longer and white</td>
</tr>
<tr>
<td>5</td>
<td>17-21</td>
<td>Fresh breeze</td>
<td>Waves more pronounced and long; white</td>
</tr>
<tr>
<td>6</td>
<td>22-27</td>
<td>Strong breeze</td>
<td>Larger waves form</td>
</tr>
<tr>
<td>7</td>
<td>28-33</td>
<td>Strong wind</td>
<td>Sea heaps up; wind blow foam in streaks</td>
</tr>
<tr>
<td>8</td>
<td>34-40</td>
<td>Fresh gale</td>
<td>Height of waves increases visibly; foam</td>
</tr>
<tr>
<td>9</td>
<td>41-47</td>
<td>Strong gale</td>
<td>is blown in dense streaks</td>
</tr>
<tr>
<td>10</td>
<td>48-55</td>
<td>Whole gale</td>
<td>High waves with long over hanging crests</td>
</tr>
<tr>
<td>11</td>
<td>56-65</td>
<td>Storm</td>
<td>Waves so high that ships within sight are hidden in the troughs; air filled with spray</td>
</tr>
</tbody>
</table>

**Sea state and swell**

The sea state and swell are closely related to prevalent wind conditions. The difference is that sea state is defined as those waves caused by ambient wind conditions, whereas swell is produced by waves formed by past wind action, or by wind blowing at a distance. Careful observations of sea state and swell were vital to the detection of weather systems, e.g. cyclones forming and passing by at a distance from the ship (Table 2.3). A short swell means a swell where the length or distance between each successive wave crest is relatively small. A long swell means a swell in which the length or distance is large. A low swell means a swell where the height between the lowest and highest part of the swell is small. A heavy swell means a swell where height is great.

Both sea state and swell were recorded in the deck log by means of the adjusted Douglas Sea State and Swell scale (WMO Code table 3700, Manual on Codes, No. 306, part A) (Table 2.3). The direction of the swell was specified in a similar way to that of wind direction; that is, the point of the 16-point compass from which the swell travels. Sea state observations were converted into the numerical height of wave fields using Table 2.3. Wave height data were passed through the Generalized ESD test to flag outliers, and flagged observations were replaced by the mean of temporally neighbouring values, if available, or otherwise made null. Both sea-state and swell wave heights were standardised to metres.
Table 2.3 Ten-point Sea State (left) and Swell Scale (right), adapted from Douglas Sea and Swell Scale (WMO Code table 3700).

<table>
<thead>
<tr>
<th>Sea State</th>
<th>Swell Scale</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scale No.</td>
<td>Description</td>
</tr>
<tr>
<td>0</td>
<td>Calm(Glassy)</td>
</tr>
<tr>
<td>1</td>
<td>Calm(Rippled)</td>
</tr>
<tr>
<td>2</td>
<td>Smooth</td>
</tr>
<tr>
<td>3</td>
<td>Slight</td>
</tr>
<tr>
<td>4</td>
<td>Moderate</td>
</tr>
<tr>
<td>5</td>
<td>Rough</td>
</tr>
<tr>
<td>6</td>
<td>Very rough</td>
</tr>
<tr>
<td>7</td>
<td>High</td>
</tr>
<tr>
<td>8</td>
<td>Very high</td>
</tr>
<tr>
<td>9</td>
<td>Precipitous</td>
</tr>
</tbody>
</table>

Air and sea-surface temperature

One of the main meteorological parameters of interest is temperature, and all whaling ships were equipped with mercury thermometers to measure it. Most of the logbooks record only air temperature, which was measured using the dry bulb of a psychrometer. The psychrometer structure was placed about 5 feet above the upper deck in the open air, as free as possible from the sun’s radiation or warm air from galleys, engine and boiler rooms (Chapter XIII, Admiralty 1938). To maximise the exposure to the ambient weather, the psychrometer was usually hung towards the windward side of the ship. At least 15 minutes were allowed to pass after placing it in position for temperatures to settle before making observations.

Ship-design changes from sail to diesel had more impact on the methodology of sea-surface temperature (SST) measurements than air temperatures. Originally a wooden or a canvas bucket was lowered to sample and draw up sea water before thermometers enclosed in metal cases were put into the bucket to measure SSTs. The water was drawn clear of any discharge from the ship, and the thermometer was kept inserted in the water for several minutes to obtain stable readings. If a canvas bag was used, it was not placed in a draught as evaporation could lower the measured temperature. When ships became diesel-engine driven, SSTs were measured directly from seawater that was let in to cool the engines in the engine
room. Measurements were taken with a mercury thermometer placed as close as possible to the water inlet (Kent et al. 2010).

As it was more usual to measure air temperatures than SSTs, if logbooks were ambiguous about the source of measurement, it was considered to be air temperatures. Where available, both temperatures were passed through a Generalized ESD test to flag outliers, separated by Ship ID groups and parameters. The flagged outliers were replaced with reference to neighbouring values. Almost all observations were recorded in degrees Fahrenheit, which were converted to degrees Celsius in the database.

**Sea-level air pressure**

Along with wind and swell observations, atmospheric pressure measurements were vital in keeping ships safe from hazardous storms in the high seas. Vessels in the path of powerful storms could capsize or be severely damaged by the waves and sea swell. Hence, great care was taken to observe and record changes in atmospheric pressure as wireless (radio) weather forecasts were not available for the Southern Ocean in the 1930s and 1950s. All British whaling ships were supplied with Kew-pattern marine mercurial barometers (Chapter XIII, Admiralty 1938). Each barometer was fitted with a Gold slide which could offset the pressure-column reading for latitude, height of instrument above sea-level and temperature. It was recommended to note the Gold slide offset before reading the barometer, as heat from the body of the observer could affect the Gold slide more quickly than the barometer.

The above corrections were necessary to reduce the barometer observations to sea-level and latitude of 45º (the standard unit of atmospheric pressure; Chapter XIII, Admiralty 1938). When a ship moved violently, the mercury jerked up and down along the scale, and the mean of the highest and lowest heights was taken as the value of the barometric pressure. Logbooks in the earlier years of present collection recorded pressure in inches of the mercury column. By the 1950s this had changed to millibars. Once again, I performed a Generalised ESD test to determine extraction and typographical errors. I found no outliers in the data; indeed, air pressure readings appeared to have been very carefully and meticulously observed and recorded. To bring uniformity to the observations in the dataset, a standard formula was used to convert mercury-column inches into hecto-Pascals (hPa or millibars).
2.2 Data and Methods

2.2.3 Standardisation

The aim of producing a readily accessible dataset cannot be achieved without adhering to some generally accepted format. The International Maritime Meteorological Archive (IMMA) format (Woodruff 2007) comprises a comprehensive set of marine climate variables, including the most commonly reported meteorological variables with the time, location, and ship-related meta-data, among others. IMMA1 (IMMA version 1, the latest version; Smith et al. 2016), adopted in the ICOADS 3 dataset, is an internationally accepted format to integrate historical weather information from diverse platforms and national/international sources. A number of earlier formats exist, ranging from 1853 Maritime Conference conventions to the WMO’s early International Maritime Meteorological (IMM) punched-card format (Yoshida 2004, WMO 1952), to Global Telecommunication System (GTS) codes used to transmit weather reports to land-based stations from ships; and many other reporting practises. With the myriad of conventions and formats used to record, transfer and archive historical and near-contemporary weather data, it was necessary to devise a common format that would make different data sources compatible with each other. Hence, the IMMA format was produced by retaining the best features and concepts of previously used formats but providing a new format that is better aligned with modern electronic data services and storage.

The IMMA format is designed to be flexible in terms of the number of meteorological and meta-data fields to the level of individual records in the datasets. The IMMA format record consists of an essential set of parameters called the Core, followed by a number of different attachments (attms). The Core is divided into locational and meteorological sections, incorporating many of the most commonly used parameters in a standardised form (fields listed in IMMA, Supp. D, more information can be found in the IMMA report; ICOADS (Freeman et al. 2017)). I am using Ship metadata (Meta-vos) attm (C7) to store meta-data from the logbooks, including recruiting country and the country of registration of the ship, the types of thermometers, barometers and other details. I have constructed the dataset following the recommendations regarding format, conventional codes (e.g. country codes, variable indicators etc.) and precision for each variable in the Core (C0) and Meta-vos attm (C7) fields.
2.3 Results and future steps

I have created a readily accessible, standardised and quality checked IMMA-compliant dataset of meteorological observations from the Southern Ocean. This dataset is a result of the first-ever study to extract meteorological observations solely from whaling logbooks in the Southern Ocean. Each record contains a number of positional, meteorological and meta-data parameters found in the Christian Salvesen Co. whaling logbooks of the 1930s and 1950s. Each parameter was manually extracted from logbooks and stored in a relational database. All data points were passed through statistical tests to flag and correct erroneous values. To make the dataset accessible and inter-portable with other marine datasets, data were homogenised and standardised. Each record in our dataset was produced according to recommendations in the IMMA format, bringing them to a level similar to that of existing international datasets relating to historical meteorological records such as ICOADS.

In total, the assembled dataset contains close to 12,000 observations recorded during 4604 observation-days spanning two decades. It contains 71 variables in total, including 48 and 23 variables for Core and Ship meta-data sections, respectively. All data records have positional and time fields due to our quality systems’ insistence on non-null data for these fields. Wind conditions, air pressure and temperature fields are more populous than other climatic fields. The meta-data fields are collected from various sources and are provided alongside the meteorological observations. For example, the dimensions of the observing platform (e.g. ship) are not recorded in the logbooks, but by searching the UK Shipping Registry against the name of the vessel mentioned in the logbook, the Official Number (ON), type of vessel, number of engines, dimensions and other ancillary information are obtained. In the following section, I inspect the spatial and temporal characteristics of the dataset.

2.3.1 Spatial and Temporal spread of dataset

A preliminary exploration of the new dataset was undertaken. All observations were divided into separate seasons and plotted (Fig. 2.4). I have collated data from four and nine seasons in the 1930s and 1950s, respectively. These two decades (1930s and 1950s) were the only time period represented in the current collection of Salvesen logbooks. Interestingly, this period could contain far more observations than any other period in the whaling history in the Southern Ocean, due to the increased number of whaling expeditions at this time (Jackson 1978, Tønnessen and Johnsen 1982). The collected data from the 1930s show that the whaling
Fig. 2.4 The whaling ships positions contained in the dataset grouped by whaling season (defined as period between 1-Sept of the current year to 31-Aug of next year). The number of observations in each whaling season is shown in the parentheses.
activity was much more confined to South Georgia and the northern and western edges of the Weddell Sea in the pre-WWII period (Fig. 2.4). The highest number of observations for any season in the 1930s decade comes from the 1933/34 season, yet the longest season was the 1932/33 season (Table 2.1; Fig. 2.4). The whaling season would typically start between mid-September and early-November and continue until March of the next year, making the length of the season in this part of the Southern Ocean between about 150 and 170 days (Table 2.1).

Fig. 2.5 Map of positional data from Christian Salvesen Whaling Co. ships operating in the Southern Ocean in the 1930s and 1950s. All data points are binned in 1° grid boxes and all grid boxes shown contain at least one observation from the whole period.

On the other hand, whaling positions in the 1950s are much more spread out, covering almost the whole circumpolar Southern Ocean (Fig. 2.4). This could be partly due to the scarceness of whaling stocks around South Georgia and the larger number of competitors in the 1950s than in the 1930s. The whaling seasons started much later in the 1950s as compared with the 1930s but the length of the season is longer (typically about 180-200 days). The start and length of the whaling season were dictated by weather conditions and international regulations (Jackson 1978, Tønnessen and Johnsen, 1982). Even though the length of seasons in the 1930s and 1950s are fairly comparable (Table 2.1), 70% of the meteorological data in the dataset comes from 1950s. This is due to the fact that, on average,
there were six observations per day in the 1950s as compared to 1.46 observations in the 1930s.

Next, to better understand the spatial density of the dataset, all positional data were binned in a 1° latitude-longitude grid box (Fig. 2.5). The average number of observations in a grid box is weighted by the cosine of the central latitude of the grid box. As shown in the spatial density plot (Fig. 2.5), the Weddell Sea sector is the most densely populated region, followed by the Amundsen and Bellingshausen seas (ABS) sector and the Indian Ocean sector. The data density reflects the whaling practices of that time; for example, the marginal sea-ice zones around the Weddell Sea and much of the Antarctic coastline were popular whaling grounds due to abundance of whale pods and shelter provided by the ice pack, which dampens waves, during bad weather (McLaughlin 1976, Tønnessen and Johnsen 1982).

2.3.2 Planned use of the historical dataset and data rescue efforts

Observational meteorological datasets such as ICOADS3 and ISPD contain the World’s largest collection of global marine observations; however, significant gaps exist in the Southern Ocean. These datasets are routinely used to constrain NWP model runs to produce long-term climate datasets derived from observational data, also known as climate reanalysis (e.g. the Twentieth Century Reanalysis version 2 (20CRv2); Compo et al. 2011). The meteorological observations derived from historical logs of whaling vessels in this study complement these existing observational datasets by filling a considerable data void in the Southern Ocean. The observations can also be fed into the growing number of numerical reanalysis climate models used to generate increasingly accurate pre-modern climate datasets. Furthermore, the relatively dense set of observations comprising the whaling dataset can be used independently to study and to reconstruct circumpolar or regional climate by using appropriate data analysis. Such an analysis can enhance our understanding of the background climatology of the Southern Ocean from the pre-satellite era, with which it is important to compare more recent observations and trends. For example, the MSLP observations can be used to reconstruct the Southern Annular Mode (SAM; Marshall 2003). Similarly, storminess and frequency of depressions tracking through the Southern Ocean is another example of evidence concerning synoptic climatology that can be derived from these historical datasets.

In addition, the RECLAIM project has identified a large cache of largely unknown Christian Salvesen Whaling Co. logbooks at the Sea Mammal Research Unit (SMRU),
Climate dataset from whaling ships’ logbooks

University of St. Andrew’s, Scotland (Wilkinson and Wilkinson 2018). In total 14,900 new images have been captured, with most logbooks from British-flagged whale factory ships (together with one Norwegian and two Japanese vessels). These logbooks contain the usual positional, weather and sea-conditions and ice-condition records. Time and resource permitting, data extracted from the logbooks found at St. Andrew’s, together with the dataset presented in this study, can close substantial gaps in our knowledge of Southern Ocean climate in the pre-satellite period. A further collection of whaling logbooks at the Vestfold Archive in Sandefjord, Norway, has been identified and photographed, producing more than 30,000 digital images of logbooks, catch books and day reports (Wilkinson 2016b). This study has shown that whaling logbooks can be used to extract valuable meteorological observations from a little known area of the World and points to their large untapped potential for data-rescue efforts and the construction of historical meteorological datasets.

A review of data-rescue priorities and practices undertaken by Brönnimann et al. (2018), states that to make these data-rescue efforts more effective, they should be continuous and coordinated with a long-term goal. One of the largest international data-rescue initiatives, Atmospheric Circulation Reconstructions over the Earth (ACRE; www.met-acre.org/) coordinates over 50 organisations in various countries. The ACRE initiative both undertakes and facilitates the retrieval of historical instrumental surface, terrestrial and marine weather observations. A few examples of ACRE supported and/or inter-linked projects are ISPD, the U.K. Colonial Registers and Royal Navy Logbooks project (www.corral.org.uk), ICOADS3, RECLAIM, the International Environmental Data Rescue Organisation (IEDRO; www.iedro.org), and NOAA’s NCDC Climate Database Modernization Program (CDMP; www.ncdc.noaa.gov/oa/climate/cdmp/cdmp.html). Other similar initiatives are the Euro-Climhist database (www.euroclimhist.unibe.ch), Tambora (www.tambora.org), and Old Weather (oldweather.org). In the Southern Ocean region, the Southern Weather Discovery project, part of ACRE Antarctica (www.zooniverse.org/projects/drewdeepsouth/southern-weather-discovery), is at the forefront of data-rescue endeavours by using the collective efforts of many citizen scientists to decipher, translate and extract meteorological information from station observations, weather diaries, and ship logbooks.

The current dataset is made available in csv (comma separated values) format file and an accompanying description file. The dataset is held in the Apollo digital database, University of Cambridge Data Repository (doi.org/10.17863/CAM.31530), with free access. The data would be submitted to be included into ICOADS future releases.
Chapter 3

MSLP variability in the Southern Ocean and historical SAM index

In this Chapter, seasonal and decadal MSLP variability over the Southern Ocean are analysed and an historical marine observations-based SAM index is generated from the whaling dataset.

3.1 Introduction

The Southern Hemisphere westerlies are a very important factor in the Southern Ocean and global climate through their role in driving the Antarctic Circumpolar Current, upwelling and the out-gassing of marine $CO_2$ from the Southern Ocean as well as hydroclimatic variability on the southern continents including Antarctica (Russell et al. 2006; Ho et al. 2012; Moreno et al. 2018). Indices describing the Southern Annular Mode (SAM), a zonally symmetric dipole structure in sea-level pressure between high and middle Southern latitudes (Gong and Wang 1999) that varies with time, can be defined in many ways. For example, some studies have used differences between MSLP observations in the Southern Hemisphere mid-latitudes and high-latitudes to generate SAM indices using station-based observations (Marshall 2003; Visbeck 2009; see Fig. A.2). Whereas others have used the first principal component of empirical orthogonal function (EOF) analysis of geopotential height anomalies poleward of 20°S from reanalyses to generate SAM indices (e.g. Thompson and Wallace 2000; Fogt et al. 2012).
A SAM index is considered to be of positive polarity when pressure levels at mid-latitudes are anomalously high, and those at high-latitudes are low, and vice versa. Regardless of the methodologies used to generate different SAM indices, the SAM has a positive relationship with the Southern Hemisphere westerlies (Thompson and Wallace 2000, Swart and Fyfe 2012). During the periods of positive SAM, zonal winds at the southern end of the core westerlies belt (50-55°S; Moreno et al. 2018) tend to be stronger and move closer to Antarctica. By contrast, at the northern edge, zonal winds become weaker and move further away to the north. During negative SAM periods the westerlies behave in the opposite way (Thompson et al. 2000; Thompson and Wallace 2000).

This spatio-temporal variability of SAM and, in turn, the westerlies, has been observed to have a significant impact on regional climate in the Southern Hemisphere (Kidston et al. 2009; Ho et al. 2012; Moreno et al. 2018). Strother et al. (2015), for example, found long-distance pollen from the South American continent in the lake sediments of South Georgia (∼54°S 36°W) located in the core westerlies belt, indicating that either intensification of westerlies or positional changes in the westerly-wind belt, or both, have occurred several times in this 7,000 year-long pollen record. On the other hand, SAM’s impact on more recent (last ∼50 yr) instrumental temperature and pressure records from Antarctica and sub-Antarctic islands is well documented by Fogt et al. (2012), Marshall et al. (2011), and Turner et al. (2005), among others.

On intra-year seasonal time scales, the cycles of circumpolar trough around Antarctica are observed to dominate temperature, precipitation records and the long-term variability of Antarctic sea ice (Hurrell and van Loon 1994; Broeke 2000; Bracegirdle et al. 2008). These cycles of circumpolar trough exist in response to the movement of air-masses from continental Antarctica to its periphery driven by uneven heating and cooling of the troposphere above Antarctica and the Southern Ocean (Bracegirdle et al. 2008).

The clearer understanding of the observed relationship between SAM index (or circumpolar troughs) and climate variables (e.g. the circumpolar westerlies, MSLP, temperature) is constrained by the brevity of observations in the regions of their occurrence. The problem is further compounded by the existence of different versions of a single climate index such as SAM. A recent comparative study by Barrucand et al. (2018) brought this problem to the forefront. They found that the characteristics of different SAM indices varied significantly depending on the source of data and how those data were handled computationally. For example, they found that SAM indices, even though some were generated from the same
dataset but drawn from different regions, showed quite different spectral and climatological properties (their Figure 2).

This point was also made in earlier studies, where it was noted that certain SAM indices were more useful in explaining variability in some regions than others (Ho et al. 2012; Jones et al. 2009). This was due to SAM’s interaction with other climate indices (e.g. El Niño-Southern Oscillation (ENSO)) which led to the observed variability in the southern extra-tropical regions (Yuan and Martinson 2001; Fogt et al. 2011). For example, the relative phase of SAM and the PSA pattern (Pacific–South American pattern, a stationary Rossby-wave train emanating from the tropical central Pacific in response to ENSO events) (Mo and Ghil 1987), dictated the overall impact of ENSO on Antarctic atmospheric fields (Mo 2000; Fogt and Bromwich 2006). Another study found evidence of the cumulative impact of these climate indices on the long-term variability of Antarctic sea ice (Teleti and Luis 2016 and references therein).

In such circumstances, it is important to investigate whether the definition and data used to generate a particular SAM index are able to explain observed trends and variability in the climatic parameters at a broad scale. Ideally, the SAM index should be generated from observations close to its action centres (regions where the physical variables manifest the climatological index pattern clearly and consistently; for the SAM index these are at \(\sim 40^\circ S\) and \(\sim 65^\circ S\); Gong and Wang 1999). However, due to the lack of longer time series of observations from the more southerly of these latitudinal bands, SAM indices in the past have been reconstructed using statistical regression techniques based on tree-ring proxies or station MSLP records from the northern centre of action (Jones et al. 2009).

These issues could be addressed by using marine observations from the SAM action centres. The variability contained in the meteorological observations from these regions would be more representative of the actual changes than their proxies like tree-rings or station MSLP records. Thus, the SAM index generated from these observations would be a ‘true’ SAM index representing the hemispheric changes in a Hemisphere that has large marine component. This has been not possible until recently because of the lack of sufficient numbers of observations in the southern high-latitudes; however, the release of ICOADS v3 (Freeman et al. 2017) has meant that a large collection of historical raw meteorological observations has been made available. The primary sources of these observations in the Southern Ocean are logbooks from whaling ships which contain many meteorological observations at the daily or better frequency and additional information about the number and species of whales caught and oil produced, among other ancillary details.
The historical data from whaling logbooks were not completely ignored before ICOADS. A pioneering study by de la Mare (1997), for example, used the locations logged by the whaling fleets in the Southern Ocean as a proxy for the summer sea-ice edge. His study, and that of de la Mare (2009), showed that the sea-ice edge was further northwards in the 1950s than at present by more than 2° of latitude. However, several studies, most notably Vaughan (2000) and Ackley et al. (2003), have raised concerns about the methods used in de la Mare’s study, but also expressed doubts regarding the trustworthiness of the whaling data itself taken from the International Whaling Commission database. For example, these studies found that large scale false information from Soviet Whaling from 1947 to 1972 were reported to the International Whaling Commission database. As data-rescue efforts are often constrained by the resources available, different projects have to compete for limited resources.

The doubts expressed in the previous studies regarding the quality of the whaling data have brought the priority of the data-rescue of the whaling logbooks down in comparison to other marine sources (Wilkinson 2019, personal communication). Not only that, but the whaling logbooks were once the property of commercial entities; hence, the data-rescue efforts have been directed more towards national/international scientific expeditions to the Antarctic. This is one of the reasons that the representation of the Southern Ocean in terms of meteorological records has remained poor despite the growth in observations from other regions in international observational datasets (e.g. ICOADS).

I produced a meteorological dataset from whaling logbooks of Christian Salvesen ships operating in the Southern Ocean over the historical period (1930s and 1950s) in Chapter 2 (published in Teleti et al. 2019). This work was done to complement existing observations in the ICOADS dataset, but it also serves to demonstrate that meteorological observations from whaling logbooks are of similarly high quality as observations from any expedition or mail ship’s logbooks. Hence, this chapter aims to, first, validate the whaling dataset from Chapter 2 by comparing this dataset to existing observational dataset (ICOADS) to demonstrate the quality and veracity of the whaling records. Secondly, by combining observations from the whaling dataset and ICOADS, a larger dataset is created which is used to understand MSLP variability and to detect the presence and movement of the circumpolar trough around Antarctica in the historical period. The observed conditions are then compared to the modern climatological period (1981-2010), to estimate the climate changes in the Southern Ocean since the historical period.

Lastly, a historical marine observations-based SAM index is computed using a Whaling-ICOADS combined dataset (now containing meteorological observations close to two action
3.2 Data and Methods

The primary data source for this study, a Christian Salvesen Whaling dataset, is briefly described here. One of the largest sources of marine observations in the Southern Ocean is the International Comprehensive Ocean-Atmosphere Data Set (ICOADS; Freeman et al. 2017), and land-based observations for the same region are contained in the Global Historical Climatology Network version 1 (GHCNv1; Vose et al. 1992) and the Scientific Committee on Antarctic Research’s Reference Antarctic Data for Environmental Research (SCAR-
MSLP variability in the Southern Ocean and historical SAM index

READER; Turner et al. 2004), which are discussed below. In this chapter, I use these land- and marine-based observational datasets to compare with and extend the Christian Salvesen whaling dataset.

3.2.1 Christian Salvesen whaling dataset

The Christian Salvesen whaling dataset (Teleti et al. 2019, Chapter 2) contains meteorological observations extracted from logbooks of whaling ships operating in the Southern Ocean during the 1930s and 1950s. Extracted observations include a number of weather parameters, air and sea-surface temperatures, MSLP, wind force (speed) and direction. All data were stringently quality checked and error corrected (Teleti et al. 2019). The observations were also standardised into modern units and scales to match with contemporary data. The dataset is formatted according to the internationally accepted IMMA format (Smith et al. 2016).

In this chapter, I will focus on MSLP observations from the Christian Salvesen dataset for a number of reasons. First, MSLP observations were very uniform with a high degree of internal consistency, requiring almost no correction (Chapter 2). High-quality raw data ensures that any derived information represents changes in the physical parameter rather than an artefact of erroneous measurements. As supported by many previous studies using historical sources, MSLP observations were found to be of better quality compared to other historical climate parameters, e.g. temperature and rainfall observations (Ashcroft et al. 2014, Brunet et al. 2014). Secondly, many earlier studies (e.g. Allan and Ansell 2006, Ansell et al. 2006, Brohan et al. 2010, Küttel et al. 2010), have used historical land- and marine-based observations to reconstruct long-term compilations of monthly and daily MSLP datasets.

All MSLP observations in the Christian Salvesen dataset are either daily or sub-daily. I focus on fluctuations that are longer than daily cycles by averaging sub-daily observations into a single daily observation and assigning it the lat/long position of the noon position of the given whaling vessel. This ‘binning’ of sub-daily observations into a single daily observation with an accompanying noon position is reasonable given that a ship would usually have moved less than approximately 30 nautical miles in the course of a day during whaling and noon position is, therefore, a good approximation of a ship’s position throughout that day (Chapter 2). Daily observations are aggregated into monthly means if at least 21 days in a calendar month have valid daily observations (WMO Report No. 100; WMO 2011).
3.2 Data and Methods

3.2.2 Observational datasets

ICOADS

The International Comprehensive Ocean-Atmosphere Data Set (ICOADS; Freeman et al. 2017) is a comprehensive collection of in-situ marine meteorological observations, mainly from ships and buoys, derived from many different national and international data sources. It is described in great detail in Chapter 1. A 30-year MSLP modern climatology is generated using 1° spatial and monthly resolution data from 1981 to 2010 covering the Southern Ocean.

Station data sources

I use the first version of Global Historical Climatology Network (GHCNv1), a collection of monthly observations from approximately 6,000 temperature, 7,500 precipitation and 2,000 pressure stations from 1880 to 1988 (Vose et al. 1992). The current version of the dataset is 4, but this does not contain MSLP observations; hence, version 1 is used here to acquire MSLP data for the sub-tropical land stations in the Southern Hemisphere for the period of investigation. For the Antarctic and sub-Antarctic stations, I use the Databank of Antarctic Surface Temperature and Pressure Data (NDP-032), collected and published by Jones and Reid (2001), available at https://cdiac.essdive.lbl.gov/epubs/ndp/ndp032/ndp032.html. As some of the Antarctic stations were established many (~70) years ago, stations have often been moved to different locations and undergone instrument changes since their establishment but station records do not explicitly record or adjust for these changes. It is possible that some station records contain observations made at multiple locations but amalgamated under one station name.

The Scientific Committee on Antarctic Research’s REference Antarctic Data for Environmental Research (SCAR-READER) project (Turner et al. 2004) was specifically set up to reprocess available meteorological observations from different stations and avoid, or at least mark, factors affecting data quality such as instrument and location changes. The resulting meteorological records from the SCAR-READER dataset are more reliable but at the cost of shortening their length relative to the NDP-032 dataset. Six stations (Grytviken, Orcadas, Faraday/Vernadsky, Rothera Point, Bellingshausen and Signy are shown in 3.3) in the Antarctic and sub-Antarctic domain have long enough MSLP records to be compared with the whaling data. Out of these six stations, Grytviken, Orcadas and Signy station records have the same start year in both NDP-032 and SCAR-READER datasets - I use data records from SCAR-READER for these stations.
For the remaining three stations, I use the NDP-032 dataset as these stations have longer record length in the NDP-032 dataset than the SCAR dataset. As Turner et al. (2004) report, these stations were only moved short distances from their original locations, and it is anticipated that climatological conditions would remain largely same and these changes would not affect the quality of the records too adversely. These station records from NDP-032 have been successfully used to generate station-based SAM index by Visbeck (2009).

### 3.2.3 Spatial aggregation

To understand the longitudinal distribution of MSLP data in the Christian Salvesen whaling dataset, I separated data positions by longitude into five Antarctic sectors (Weddell Sea, Indian Ocean, Pacific Ocean, Ross Sea and Amundsen-Bellingshausen seas (ABS) following Parkinson and Cavalieri 2012). The longitudinal distribution of data for each Antarctic sector is shown in Figure 3.1. The greatest number of observations are located in the Weddell Sea sector, followed by the ABS sector and the Indian Ocean sector. It is observed that the number of observations in the Weddell Sea sector is about three times higher than observations in all other sectors put together. The latitudinal distribution of the data grouped by Antarctic sectors is given in Fig. 3.2. It can be seen that most of the observations are concentrated within the 55º-60º and 60º-65º S latitude bands, again in the Weddell Sea sector.

In order to generate meaningful statistics, I have binned the data into five Antarctic longitudinal sectors and five latitudinal zonal bands of 5° each (Fig. 3.3). The size of latitudinal bands is chosen to capture typical weather systems found in the Southern Ocean, whose average radius size is 2º latitude (Lim and Simmonds 2007). This combination of latitudinal bands and longitudinal sectors retains information density well by not over-stretching limited data over large regions and provides well-defined data-rich regions for further analysis. In addition, circumpolar zonal means are generated from sectoral means; this ensures that sparsely populated sectoral-bands are given equal weighting compared to the denser ones.

### 3.2.4 Generation of SAM index

To generate a SAM index, as constructed in previous studies (e.g. Gong and Wang 1999, Marshall 2003), mid-latitude (∼40°S) and high-latitude (∼65°S) zonal MSLP data are required. The Christian Salvesen whaling (ChriSal) dataset does not contain observations
3.2 Data and Methods

Fig. 3.1 Longitudinal distribution of MSLP data in the whaling dataset for five Antarctic sectors: a) Amundsen-Bellinghausen Sea (ABS), b) Weddell Sea (WDL), c) Indian Ocean (IO), d) Pacific Ocean (PO) and Ross Sea (RS) sectors. Each sectoral sub-plot shows the number of observations binned in 1° longitude steps and the numbers in parenthesis denote the total number of observations in each sector.

Fig. 3.2 Latitudinal distribution of MSLP data within each Antarctic sector: Amundsen-Bellinghausen Sea (ABS), Weddell Sea (WDL), Indian Ocean (IO), Pacific Ocean (PO) and Ross Sea (RS) in the whaling dataset.
Fig. 3.3 Spatial distribution of all MSLP data points binned into 1° box (contour levels on the right), overlaid with latitudinal bands (B1, B2, B3, B4, B5) and Antarctic longitudinal sectors. Locations of land-based meteorological stations used to validate Christian Salvesen whaling (ChriSal) MSLP data are denoted by triangles (on the left).
from mid-latitudes; however, there exist many meteorological stations at those latitudes which were used previously by Marshall (2003) to construct the SAM index. I use the same six mid-latitude stations used by Marshall (2003) (shown in Suppl. Fig. A.2) to generate 40°S normalised zonal monthly anomalies. Not all of the six stations began observations prior to 1930s; only Hobart and Puerto Montt did. I begin the time series of data with these two stations and add others into the zonal mean as they come online.

To generate high-latitude normalised zonal monthly anomalies, I use the B3 zonal band, which is the closest to 65°S latitude band (see Section 3.4.5). Both zonal anomalies are normalised over the 30-year period from 1930 to 1959. The difference between the normalised average MSLP of the mid-latitude (40°S) and high-latitude (B3) zonal bands is used to compute ChriSal SAM:

\[
SAM_{ChriSal} = P_{z40}^* - P_{zB3}^*
\]

(3.1)

where, \( P_{z40}^* \) is the normalised average MSLP of 40°S zonal band and \( P_{zB3}^* \) is the normalised average MSLP of B3 zonal band.

To generate SAM indices from ChriSal-ICOADS merged dataset, observations within a 5-degree latitude for the mid-latitude band (40°-45°S) and high-latitude (60°-65°S, B3 band) are used to generate SAM index using Eq 3.1.

### 3.2.5 Bootstrapping with Monte Carlo realisations

When two datasets of unequal sizes are to be compared, a direct comparison made between these datasets will be biased towards the larger dataset. To rectify this, the larger dataset is sub-sampled such that a subset drawn is similar in size to the smaller dataset and then compared to the smaller dataset. In other words, multiple subsets of fixed size (size of the smaller dataset) are drawn from, the larger dataset in random order and the number of subsets drawn is kept high to eliminate the influence of chance on the derived statistics. The number of subsets needed to be drawn depends on the confidence limits one is willing to accept of the desired statistic at a certain p-value. For example, we can denote 95% confidence interval (CI) of a statistic as \( C.I. = 2 \times 1.96 \sqrt{\frac{(1-P)P}{n}} \) where \( P \) is p-value and \( n \) is number of iterations (subsets). The larger the number of iterations, the narrower the CI and, hence, the more precise the statistic. The number of iterations is inversely proportional to the square
of CI required. For CI of 0.1, 73 iterations and CI of 0.01, 7300 iterations are sufficient at 0.05 p-value. For this study, I have chosen CI of 0.02, which requires 1000 iterations and this number of iterations falls within the range of the number of iterations recommended to determine true population statistics from Monte Carlo realisations.

3.2.6 Subsets of MSLP observations

One of the primary aims of this study is to assess the quality of ChriSal dataset in comparison to the much larger ICOADS dataset. As a measure of quality, I correlate the ChriSal dataset to the station observations, then comparing that to the correlation obtained from ICOADS dataset. To compute an unbiased statistic, for example, the correlation between station observations and marine observations from ChriSal and ICOADS datasets, I generate 1000 ChriSal-sized subsets of ICOADS dataset, following a process described in Section 3.2.5. Each subset of ICOADS dataset is then correlated to station records. The resulting correlation coefficients are compared to the correlation coefficient obtained from ChriSal and station records. This method provides an effective way to allow a like-to-like comparison of ChriSal and ICOADS datasets.

3.2.7 SAM indices from MSLP subsets

Now, to assess the usefulness of ChriSal dataset in comparison to much larger ICOADS dataset, the historical SAM index is generated from ChriSal dataset and compared to that generated from ICOADS dataset. As will be shown in Sections 3.3.2 and 3.3.3, both ChriSal and ICOADS datasets have similar data characteristics and quality; hence, they can be combined into one dataset that is larger than ICOADS. A number of SAM indices are generated by an approach similar to the one taken in the earlier section (Section 3.2.6), but in this instance, different sized subsets are generated.

I draw three categories of subsets comprising of 5,000-, 10,000- and 20,000-pairs of observations from ChriSal-ICOADS merged dataset. The number of pairs is chosen such that the first category, the 5000-pairs subset, is approximately equivalent to the actual number of observations present in the ChriSal dataset in the high-latitude band used to generate SAM index. The second category, 10,000-pairs subset, represents a dataset double the size of ChriSal and the third, 20,000-pairs subset is a dataset four times the size of ChriSal dataset. Again I draw 1000 subsets of 5,000-, 10,000- and 20,000-pairs of randomly-drawn
observations each from ChriSal-ICOADS merged dataset. Further, by treating each of these subsets as an independent dataset, a SAM index is generated by the method outlined in Section 3.2.4. The resulting three categories of SAM indices (1000-each), along with SAM indices generated from ChriSal dataset and full ChriSal-ICOADS merged dataset, are compared to existing historical SAM indices available in the literature.

### 3.2.8 Population correlation and p-value

The correlation coefficients and p-values computed by comparing new SAM indices (generated in Section 3.2.7) to existing SAM indices represent only the sample correlation of a subset of the population (ChriSal-ICOADS merged dataset). Therefore, to compute the composite population correlation coefficient and p-value, I employ meta-analysis.

The most widely used meta-analysis method, Fisher’s transform (Fisher 1948) combines p-values from independent tests (with same hypothesis) performed on subsets of a population (a sample) to generate a composite p-value for the population using the formula,

$$
\chi^2_{2k} = -2 \sum_{i=1}^{k} \ln(p_i),
$$

where $p_i$ is the p-value for $i^{th}$ hypothesis test and $k$ is number of tests combined.

To obtain the composite p-value, I calculate the area under the chi-squared curve to the right of summed chi-square with $2k$ degrees of freedom.

Similarly, the composite correlation coefficient of a population can be determined by $z$-transforming sample correlation coefficients. Fisher’s $z$ transformation of the sample correlation $r$ is

$$
z = \frac{1}{2} \log \left( \frac{1+r}{1-r} \right),
$$

which is approximately normally distributed with mean $\log \left( \frac{1+\rho}{1-\rho} \right)$ and variance $\frac{1}{\sqrt{n-3}}$, where $\rho$ is the population correlation coefficient and $n$ is the number of observations in each test.

Once a $z$ value is computed from each sample correlation coefficient, all $z$ values are averaged to obtain mean $\bar{z}$ and lower and upper limits of 95% confidence interval of mean $\bar{z}$, $\bar{z}_{0.025}$ and $\bar{z}_{0.975}$. An estimate of the composite correlation coefficient ($r$) with its corresponding 95% confidence limits is then generated from the following inverse transformation:

$$
r = \tan(z) = \frac{2e^z-1}{2e^z+1} \text{ for } z = \bar{z}, \bar{z}_{0.025} \text{ and } \bar{z}_{0.975}.
$$
3.2.9 Randomised ANOVA test

When the sizes of groups to be compared are small (with unknown distribution), a randomised ANOVA test is performed to test the null hypothesis that different groups have the same mean. To do so, first, an F-statistic ($F_{ORG}$) is calculated from an ANOVA test on the original data groups. Second, all observations irrespective of groups are mixed and drawn at random order to fill the groups and then the ANOVA test is performed on new groups to calculate a new F-statistic ($F_{NEW}$). The fraction of $F_{NEW} > F_{ORG}$ is computed, which gives $p$, the probability of obtaining an F-statistic as large as obtained with observed data if the null hypothesis were true.

The second step should be repeated for all possible permutations of the ordering the data in groups, which, in a practical sense, is near impossible. Iterations of more than 1000 are suggested to be sufficient to approximate a $p$-value. In this study, I use 5000 iterations to randomise the ANOVA test to examine if the means of the MSLP in different bands and decades are equal.

3.3 Results

Three main analyses are performed. First, the ChriSal dataset is validated against the larger ICOADS dataset. Secondly, seasonal and long-term MSLP changes over the Southern Ocean are identified. Finally, a marine observations-based SAM index and its trend are generated, in order to be compared with existing SAM indices. However, before going into band-sectoral MSLP data analysis, I first examine the temporal characteristics of the whaling data as the temporal spread of data dictates the availability of data in various sectoral-bands.

3.3.1 Temporal spread of whaling seasons

The ChriSal whaling dataset is divided into a number of seasons (a whaling season is defined as the period from 1 September of the current year to 31 August of the following year) for each year where data were available in the 1930s and 1950s. The whaling seasons in the dataset began between early September and October for the years in the 1930s, whereas for the years in the 1950s the seasons started in early December (Fig. 3.4). Typically ending in late February/early March and late April/early May, for the 1930s and 1950s, respectively. Each whaling season is on average 160-180 days long, and seasons are consistently longer
in the 1950s than in the 1930s. From a meteorological point of view, these longer seasons provide more data from spring, summer and well into late autumn in the 1950s.

Fig. 3.4 Start of the whaling season (in days from 1st September of that year, blue) and length of the season (in days, red) for the whole whaling dataset. The year indicates the first part of the whaling season (e.g. 1932 represents the 1932-33 whaling season).

3.3.2 Comparison and validation of the whaling dataset

To gain a better understanding of the spatio-temporal properties of the ChriSal whaling dataset, I compare it with the similar but larger ICOADS observational dataset in order to assess the quality of the whaling dataset.

Comparison between ICOADS and ChriSal

The first comparison of the Christian Salvesen whaling dataset is made with the ICOADS dataset. I chose the B3-WDL sectoral band (Weddell Sea sector in 60°-65°S band, Fig. 3.3) because of the abundance of data points in that band. Figure 3.5a shows the MSLP daily data from the Christian Salvesen and ICOADS datasets averaged over the same sectoral band over the 1930 to 1960 period. It is observed that both ChriSal and ICOADS datasets have observations only for a portion of the year (summer), as seen in the temporal analysis of the whaling season in Section 3.3.1. The Twentieth Century reanalysis (20CRv2c, Compo
et al. 2011) dataset averaged over the same sectoral band is also shown to contrast with observational datasets (Fig. 3.5a).

The ChriSal and ICOADS datasets have similar lower mean daily MSLP values and much larger standard deviations (mean=983.53 s.d.=11.63 hPa and mean=985.76 s.d.=7.42 hPa, respectively) compared with the 20CRv2c data (mean=990.26 s.d.=4.3 hPa), calculated over the period covered by the ChriSal dataset (Fig. 3.5a). The standard deviations in the ChriSal and ICOADS datasets are similar and almost 2-3 times the standard deviations found in the 20CRv2c data (Fig. 3.5b). This shows that the current reanalyses still have poor data quality in the historical period but also demonstrates that the ChriSal and ICOADS datasets have similar characteristics.

Both ChriSal and ICOADS datasets show large fluctuations, implying that the observational datasets are able to capture high variability in the daily MSLP observations in the Southern Ocean (Fig. 3.5b). As the B3-WDL sectoral band is known to contain frequent cyclonic activity (Simmonds and Wu 1993), large fluctuations are expected but are not present in the 20CRv2c data. It could be inferred that such reanalysis products fail to reproduce the rapid pressure fluctuations observed in the observational whaling dataset over the period of
investigation. It should also be noted that the ChriSal dataset has observations during many periods (e.g. 1940, 1941) where ICOADS observations are absent.

Correlation with sub-Antarctic and Antarctic stations

Several sub-Antarctic and Antarctic Peninsula meteorological stations were established in the early 20\textsuperscript{th} Century and have provided continuous weather observations since then. I have selected six such stations with longest records (Table 3.1). All of the earliest-established stations are located close to coasts; hence, it is possible that weather observed at these stations is influenced by maritime weather of the surrounding seas (Turner et al. 2005) and can be used to cross-validate marine observations. The stations with the longest records (Grytviken and Orcadas; for the location of the stations see Fig. 3.3) overlap with both decades present in the ChriSal dataset, and while other stations cover only the 1950s. I compare the ChriSal observations with meteorological observations from sub-Antarctic and Antarctic stations with long observational records to cross-validate the whaling dataset. The correlation obtained with the ChriSal dataset is compared with correlations obtained from the ChriSal-sized subsets of ICOADS dataset (described in Section 3.2.5).
Table 3.1 Correlations between sub-Antarctic and Antarctic meteorological station MSLP observations and various sectoral-band MSLP observations in the Christian Salvesen whaling dataset and ChriSal-sized ICOADS subsets (** indicate significance at 5% level and the locations of both stations and sectoral bands are shown in Fig. 3.3).

<table>
<thead>
<tr>
<th>Station Name</th>
<th>Position</th>
<th>Observing since</th>
<th>ChriSal Corr.</th>
<th>ICOADS Com. Corr.</th>
<th>95% CI of Composite Corr.</th>
<th>Sectoral Band</th>
</tr>
</thead>
<tbody>
<tr>
<td>Grytviken</td>
<td>54° 3' S 36 5' W</td>
<td>1905</td>
<td>0.31</td>
<td>0.71**</td>
<td>0.25</td>
<td>B2-WDL</td>
</tr>
<tr>
<td>Orcadas</td>
<td>60° 7' S 44° 7' W</td>
<td>1903</td>
<td>0.68**</td>
<td>0.79**</td>
<td>0.30</td>
<td>B3-WDL</td>
</tr>
<tr>
<td>Faraday/Vernadsky</td>
<td>65°4' S 64° 4' W</td>
<td>1944</td>
<td>0.84**</td>
<td>0.92**</td>
<td>0.15</td>
<td>B4-ABS</td>
</tr>
<tr>
<td>Rothera Point</td>
<td>67° 3' S 68° 8' W</td>
<td>1946</td>
<td>0.82**</td>
<td>0.93**</td>
<td>0.20</td>
<td>B4-ABS</td>
</tr>
<tr>
<td>Bellingshausen</td>
<td>62° 2' S 58° 9' W</td>
<td>1944</td>
<td>0.84**</td>
<td>0.90**</td>
<td>0.28</td>
<td>B4-ABS</td>
</tr>
<tr>
<td>Signy</td>
<td>60° 7' S 45° 6' W</td>
<td>1947</td>
<td>0.84**</td>
<td>0.73**</td>
<td>0.45</td>
<td>B3-WDL</td>
</tr>
</tbody>
</table>
Data from shore stations (listed in Table 3.1) are compared with the ChriSal and ICOADS subsets from the latitudinal-sectoral bands they lie in. Figure 3.6 shows the histogram of correlation coefficients obtained from comparing a thousand ICOADS subsets with each station. The composite correlation coefficient for the population alongside the 95% Confidence Interval (CI) of the composite correlation is generated as outlined in Section 3.2.6. The correlation coefficients (Table 3.1) obtained from the ChriSal dataset are also shown as thick red lines in Figure 3.6.

The comparison of ChriSal’s B2-WDL sectoral band with Grytviken shows that ICOADS subsets much more closely covary than the ChriSal dataset and ChriSal correlation coefficient is found to be insignificant as well (Fig. 3.6a, Table 3.1). However, the comparison of the ChrisSal’s B3-WDL sectoral band with Orcadas shows that the ICOADS subsets still more closely covary than with the ChriSal dataset. But the ChriSal correlation coefficient is closer to the ICOADS composite correlation coefficient and is significant at the 5% level. In the third comparison, B3-WDL sectoral band with Signy shows that the ChriSal dataset outperforms ICOADS subsets. The ChriSal correlation coefficient is not only larger than...
the ICOADS composite correlation coefficient but also outside 95% CI of the composite correlation coefficient. The quartiles of the distribution of the correlation coefficient between the B3-WDL sectoral band and Signy show a large range; even, in this case, the ChriSal correlation coefficient is better (Fig. 3.6c).

Similarly, comparisons of the ChrisSal B4-ABS sectoral band with Faraday, Rothera Point and and Bellinghausen show that ICOADS subsets still covary more closely than with the ChriSal dataset, but the ChriSal correlation coefficient is within the 95% CI of ICOADS composite correlation coefficient in Fig. 3.6f (Bellinghausen station). All of the compared stations and corresponding bands of ChriSal dataset are strongly correlated except between Grytviken and the B2-WDL band of ChriSal dataset at 5% significance level.

3.3.3 Seasonal variability of MSLP across the Southern Ocean

The ChriSal dataset is shown to be of equally good quality, if not better, compared to the ICOADS dataset and contributes in periods and sectors where ICOADS does not have data (Section 3.3.2). As also seen in the earlier sections (Section 3.2 and 3.3.1), ICOADS and ChriSal datasets strongly agree with and complement each other. To fully understand MSLP variability across the Southern Ocean, I have merged the ChriSal and ICOADS datasets to create a dataset larger than ICOADS. I investigate the ChriSal-ICOADS merged dataset to examine MSLP seasonal variability across the Southern Ocean in the historical period (1930s and 1950s).

The mean monthly MSLP values from various data sources in different latitudinal zonal bands are shown in Fig. 3.7. The monthly MSLP data from ChriSal and ChriSal-ICOADS merged datasets are shown in Fig. 3.7a-b, which are compared to monthly mean MSLP observed in the modern climatological period (1981-2010) in Fig. 3.7c. I have used monthly mean MSLP from ERA-Interim (Dee et al. 2011) and ICOADS datasets over 1981-2010 period to construct data for the modern climatological period. It is observed that, in general, the mean MSLP over the modern climatological period is higher and standard errors are smaller as compared to the historical period. As Fig. 3.7a and b show, the ChriSal dataset closely follows the ChriSal-ICOADS merged dataset, but with higher standard errors; hence, the more certain ChriSal-ICOADS merged dataset is considered for the historical period. In addition, the modern period data is taken from ICOADS dataset because as Fig. 3.7c shows that ERA-Interim overestimates monthly mean MSLP for all zonal bands by up to 7hPa for all months.
3.3 Results

Fig. 3.7 The zonal mean monthly MSLP (hPa) from ChriSal and ChriSal-ICOADS merged datasets (solid and dashed lines, respectively) with standard errors (error bars in hPa) for a) 1930s and b) 1950s decades. Zonal mean monthly MSLP (hPa) from ERA-Interim and ICOADS datasets (solid and dashed lines, respectively) with standard errors (error bars in hPa) for modern climatology (1981-2010) are shown in c).

To compare the historical and modern period, I performed randomised ANOVA test (as outlined in Section 3.2.9) on the ChriSal-ICOADS merged (for historical period) and ICOADS (for modern period) datasets. The p-values thus obtained from randomised ANOVA tests are used to reject or accept the $H_0$ (the means are equal). The analysis is restricted to early-Summer to late-Autumn months; as seen in the earlier Section 3.3.2, both datasets (ChriSal and ICOADS) do not contain any observations in the months outside the limits of the whaling season for the period under investigation over the Southern Ocean. The monthly mean MSLP shows a similar pattern in both decades (1930s and 1950s); however, data from the 1930s are more uncertain (large error bars) than in the 1950s.

In general, it is observed that for spring, before the start of summer in December, MSLP in all bands shows a decreasing tendency. However, from December monthly MSLP in both datasets rises, reaching a relative peak in January before declining in February, then again rising in March and the months after that (Fig. 3.7a, b). The quantum of changes are more certain in the 1950s and in the ChriSal-ICOADS merged dataset than in the 1930s. The variability in individual bands reveals a more accurate picture; for example, the B2 band in the 1930s shows a decreasing trend in the months of October and November and reaches the
lowest MSLP in December. It then starts to recover before dropping again in February, from where it continues to rise after February. Similarly, B3 band (absent before December) rises from December to January before dipping in February to rise again in March and succeeding months. However, the B4 band in the same decade rises from December to January but has a sustained fall through February and March.

The first comparison is made between monthly mean MSLP contained in the various months (November, December, January, February and March) for each zonal band and time period. This test examines whether the mean MSLP within a particular zonal band varies significantly over the months investigated (Table 3.2). In the 1930s, the largest changes (the difference between highest and lowest monthly MSLP within a band) occurred in B4 band followed by B2 and B3. None of the changes observed are statistically significant at 5% level. The changes in the 1950s are varied, 2.74 hPa in B3 to 7.06 hPa in B4 band. However, changes in the B4 sector are statistically significant at 5% level (Fig. 3.7b). In the modern climatology period, the largest change is recorded in B4 with 5.44 hPa, followed by B3 (2.26 hPa) and B2 (1.08 hPa) bands across months. The changes in B4 and B3 are significant, while B2 changes are insignificant (Table 3.2). Overall, in the 1950s B2, B3 and B4 bands are at their lowest in December, December (February in 1930s) and March, respectively. In contrast to the modern period where B2, B3 and B4 have their lowest in January, March and March, respectively.

Interestingly, the bands seem to have their lowest MSLP month in sequence; for example, in the 1930s, B2 and B3 in December, B4 in March, whereas in the 1950s, B2 is lowest in December, B3 in February and B4 in March. By contrast, in the modern climatology period, B2 is lowest in January, B3 and B4 in March.
Table 3.3 The difference (hPa) between the highest and lowest in the mean monthly MSLP at each month (November, December, January, February and March) across the zonal bands (B2, B3 and B4) for each time period. The p-values obtained from computing t-test on the groups of means are shown in the parenthesis and the numbers in the square bracket denote band with lowest MSLP for that decade and month.

<table>
<thead>
<tr>
<th>Decades</th>
<th>November</th>
<th>December</th>
<th>January</th>
<th>February</th>
<th>March</th>
</tr>
</thead>
<tbody>
<tr>
<td>1930s</td>
<td>NA</td>
<td>0.17 (p=0.99) [B4]</td>
<td>1.21 (p=0.71) [B3]</td>
<td>0.28 (p=0.99) [B2]</td>
<td>5.78 (p=0.15) [B4]</td>
</tr>
<tr>
<td>1950s</td>
<td>2.01 (p=0.15) [B3]</td>
<td>4.14 (p=0.04) [B2]</td>
<td>1.86 (p=0.29) [B3]</td>
<td>4.62 (p=0) [B4]</td>
<td>5.37 (p=0) [B4]</td>
</tr>
<tr>
<td>1981-2010 climatology</td>
<td>11.21 (p=0) [B4]</td>
<td>6.60 (p=0) [B3]</td>
<td>5.73 (p=0) [B4]</td>
<td>8.36 (p=0) [B4]</td>
<td>11.11 (p=0) [B4]</td>
</tr>
</tbody>
</table>

Table 3.4 The difference (hPa) between modern and 1950s in the mean monthly MSLP in the zonal bands (B2, B3 and B4) across each November, December, January, February and March months. The p-values obtained from computing t-test on the groups of means are shown in the parenthesis.

<table>
<thead>
<tr>
<th>Decades</th>
<th>November</th>
<th>December</th>
<th>January</th>
<th>February</th>
<th>March</th>
</tr>
</thead>
<tbody>
<tr>
<td>B2</td>
<td>4.04 (p=0)</td>
<td>7.29 (p=0)</td>
<td>5.68 (p=0)</td>
<td>7.08 (p=0)</td>
<td>5.60 (p=0)</td>
</tr>
<tr>
<td>B3</td>
<td>0.51 (p=0.90)</td>
<td>-1.07 (p=0.76)</td>
<td>-2.25 (p=0.10)</td>
<td>-2.32 (p=0.08)</td>
<td>-3.05 (p=0.11)</td>
</tr>
<tr>
<td>B4</td>
<td>-4.17 (p=0.22)</td>
<td>-1.17 (p=0.22)</td>
<td>-3.02 (p=0.06)</td>
<td>-0.97 (p=0.75)</td>
<td>-1.05 (p=0.77)</td>
</tr>
</tbody>
</table>
This pattern of lower monthly MSLP in each band at a specific month could be due to transitioning of the circumpolar trough from one band to another. If that is true, then MSLP of all bands at each month under investigation should also differ, as mean MSLP of the band under the influence of the low pressure circumpolar trough will differ from other bands. Table 3.3 shows that, in the 1950s, the band with lowest MSLP is B3 for November, B2 for December, B3 again for January and B4 for February and March. The largest three differences between the bands for each month have occurred in December, February and March, and each is statistically significant.

The pattern is unclear in the 1930s; the lowest MSLP is recorded in B4 in December, B3 in January, B3 in February and B4 again in March. The differences between the bands are not significant (Table 3.3). In the modern climatology period, the lowest MSLP band for November is in B4, B3 in December, B4 in January, February and March. The differences between the bands for each month are significant, the largest difference is observed in November. Tables 3.2 and 3.3 reveal that certain months have lower monthly MSLP than other months within a band and certain bands have lower monthly MSLP than other bands at each month. The distribution of the low pressure circumpolar trough seems to progress southwards from the beginning of summer (December), starting from B2 band and progressing to B3 and B4 as the autumn (March) approaches.

The third comparison is carried out to assess if the monthly mean MSLP for each band across the months has changed from historical to modern periods (Table 3.4). The zonal band B2 MSLP has significantly increased overall months between the 1950s and modern climatology. In contrast with MSLP in B3 band where it has decreased, especially in February and March. The changes in MSLP of the B4 band is similar to the B3 band, but the largest decrease occurs in November and January over the same period of time(Table 3.4). It is suspected that the semi-annual cycle of the circumpolar trough is altered in the modern period. To test this hypothesis, I generate semi-annual cycles of the circumpolar trough in the next section.

### 3.3.4 Changes in the Semi-Annual cycle

The movement of the circumpolar trough is more strongly manifests in the semi-annual cycle, analysing the semi-annual cycle can reveal a clearer picture of observed changes in the strength and positional variability of the circumpolar trough as seen in the previous section (Section 3.3.3). The semi-annual cycle is strongly observed in the monthly MSLP
3.3 Results

differences between stations located around ∼50°S and ∼65°S latitude bands (van Loon et al. 1993, Meehl et al. 1998). To examine if the semi-annual cycle in the MSLP differences has changed from historical period to the modern period, MSLP differences between Grytviken and Orcadas (See Fig. 3.3 for locations) are generated from each time period (1930s, 1950s and modern) and shown in Fig. 3.8a. The differences are spectrally decomposed by using fast fourier transform (FFT) to reveal various cycles and associated phase angles contained in them (Fig. 3.8c,e). The differences show large month-to-month variations in the 1930s, followed by 1950s, but the differences in the modern period are flattest of all three (Fig. 3.8a). Overall, two peaks at September-October and February-March are observed when the difference Grytviken minus Orcadas is highest, indicating circumpolar trough is at its southernmost position over Orcadas (∼65°S latitude band). While, two troughs are observed, at June-July and December-January, indicating circumpolar trough is at its northernmost position over Grytviken (∼50°S latitude band).

Fig. 3.8 The MSLP differences between, spectral decomposition of differences and phase of each cycle in the spectral decomposition for Grytviken minus Orcadas (a,c,d) and Macquarie minus Dumont Durville (b,d,f), respectively. Differences are available for each decade (1930s, 1950s and modern) except for Macquarie minus Dumont Durville, where only 1950s and modern period are available.

The spectral decomposition of these differences also suggests that semi-annual cycle is strongest in the 1930s, followed by 1950s, where the amplitudes of the six-monthly cycle are larger than any other cycle (Fig. 3.8c). However, in the modern period, the semi-annual
cycle is almost negligible as compared to other cycles. The phase delay in the semi-annual cycles for three time periods is approximately equal.

The same is generated from Macquarie (54.5°S 158.9°E) and Dumont Durville (66.7°S 140.0°E) stations, shown in Fig. 3.8b,d and f. As Dumont Durville was established in 1956, no differences could be generated for the 1930s, and the 1950s period comprises years from 1956-1966. The monthly differences are within the range of 3-4 hPa throughout the year except in the winter when the differences change from 14 hPa at the start of winter to 7hPa in the middle of winter. Similar to differences between Grytviken and Orcadas, 1950s differences show pronounced peaks and troughs than the modern ones, generated from Macquarie and Dumont Durville stations. The 1950s semi-annual cycle is larger than the modern ones, in addition, the six-monthly cycle in the modern period is not the largest signal by amplitude, in contrast to the 1950s (Fig. 3.8d).

Furthermore, the absolute phase delay for the 1950s is 130.79°, while, for the modern period 141.73° (shown in 3.8f), meaning, peaks and troughs in the modern period are delayed by approx. 21 days, which can be observed in the MSLP monthly differences showing a delay of one month in the modern period. In general, the semi-annual cycle of the circumpolar trough has weakened and slightly delayed in the modern period as compared to the historical period.

### 3.3.5 Seasonal changes across decades

The monthly MSLP changes and semi-annual circumpolar cycle for each time period and changes from the historical to the modern period are shown in the previous sections (Sections 3.3.3 and 3.3.4). I will now investigate long-term seasonal changes in the Southern Ocean through the ChriSal-ICOADS merged and ICOADS datasets. As all bands have data mainly for summer months (December, January and February; DJF) therefore the analysis will be restricted to the DJF season. The decadal DJF seasonal means for all zonal bands covering the Southern Ocean are computed for each decade under investigation (1930s and 1950s). To estimate the modern climate of the Southern Ocean, a 30-year climatology (1981-2010) is calculated for each zonal band using monthly 1° gridded ICOADS dataset. The station records (listed in Table 3.1) are used to compare their MSLP in the modern period to the historical period. The rationale being that stations on the coast have strong maritime influence, their records can substitute in the absence of marine observations in those locations and periods. This takes into account cross-validated changes observed in the marine data.
Table 3.5 Differences in the MSLP between the historical (1930s and 1950s) and modern (1981-2010) period across different bands covering the Southern Ocean, computed from the ChriSal-ICOADS merged dataset and station records.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>B2</td>
<td>ChriSal-merged, Grytviken Station</td>
<td>0.3±1.06 (p=0.78)</td>
<td>6.30±0.71 (p=0)</td>
<td>6.60±0.86 (p=0)</td>
</tr>
<tr>
<td></td>
<td>ChriSal-merged, Grytviken Station</td>
<td>-0.14±0.96 (p=0.88)</td>
<td>1.88±0.89 (p=0.04)</td>
<td>1.73±0.66 (p=0.03)</td>
</tr>
<tr>
<td>B3</td>
<td>ChriSal-merged, Orcadas Station</td>
<td>-0.87±1.11 (p=0.44)</td>
<td>1.26±0.78 (p=0.06)</td>
<td>-0.39±0.89 (p=0.57)</td>
</tr>
<tr>
<td></td>
<td>ChriSal-merged, Orcadas Station</td>
<td>-2.33±1.07 (p=0.04)</td>
<td>1.64±1.23 (p=0.16)</td>
<td>-0.14±0.98 (p=0.89)</td>
</tr>
<tr>
<td>B4</td>
<td>ChriSal-merged, Faraday/Vernadsky Station</td>
<td>-1.11±1.2 (p=0.37)</td>
<td>0.27±0.97 (p=0.75)</td>
<td>-0.84±0.92 (p=0.33)</td>
</tr>
<tr>
<td></td>
<td>ChriSal-merged, Rothera Point</td>
<td>NA</td>
<td>0.99±1.18 (p=0.41)</td>
<td>NA</td>
</tr>
<tr>
<td></td>
<td>ChriSal-merged, Bellinghausen Station</td>
<td>NA</td>
<td>-2.89±1.29 (p=0.05)</td>
<td>NA</td>
</tr>
<tr>
<td></td>
<td>ChriSal-merged, Signy Station</td>
<td>NA</td>
<td>-1.69±1.06 (p=0.12)</td>
<td>NA</td>
</tr>
</tbody>
</table>

Decadal means for Grytviken and Orcadas are computed for each decade and for the 30-year (1981-2010) climatological mean. For the rest of the stations only means from the 1950s and modern climatological period are generated, except for Signy station where observations were stopped in 1995, rendering it insufficient for generating modern climatological mean. The differences between each decade for all available zonal bands using different data sources (ChriSal-ICOADS merged datasets and station data) are arranged by the zonal bands and differences are colour-coded by the significance of the difference at 1% and 5% significance level (Table 3.5).

For the B2 zonal band, the difference between the 1950s and 1930s is negative for Grytviken station, but positive for the ChriSal-ICOADS merged dataset (Table 3.5). However, both differences are insignificant at the p<0.05 level. In contrast with the difference found between the modern climatological mean and the 1950s, all data sources report higher MSLP in the modern climatology than 1950s. The differences computed from ChriSal-ICOADS (6.30 hPa) and Grytviken station records (1.88 hPa) are significant at a 5% level. Similarly, all data sources show higher MSLP in the modern climatology than in the 1930s, with the differences significant at a 5% level (Table 3.5).

For the B3 zonal band, the difference between the 1950s and 1930s is negative in all data sources, but the differences are significant at the 5% level only from Orcadas records. The differences between the modern climatological mean and 1950s show higher MSLP in the
modern period, but the differences are insignificant at 5% level (Table 3.5). The difference between the modern climatological mean and the 1930s is negative for both data sources but insignificant at a 5% level.

Many stations in the B4 zonal band do not have observations from the 1930s; hence, no differences between the 1930s and 1950s and modern climatology are computed (Table 3.5). The difference between the 1950s and 1930s in the B4 zonal band is negative from the ChriSal-ICOADS merged dataset though insignificant. The differences between modern climatological mean and 1950s are negative except ChriSal-ICOADS dataset. This difference (modern minus 1950s) is insignificant except in Rothera station. In addition, the difference between modern and the 1930s are all negative (Table 3.5).

So, overall the MSLP changes in the B2 band of Southern Ocean is observed to be increasing from the 1930s to modern climatological period, though the changes between the 1950s and the modern climatological period are more profound. The bands B3 and B4 report a continuous decreasing trend from the 1930s to the modern period; however, the observed changes are insignificant at 5% level. To investigate the effects of these changes on the circulation pattern, historical SAM index is generated in the next section.

### 3.3.6 New marine observations-based SAM indices

In this section, a number of SAM indices are generated from ChriSal and ChriSal-ICOADS merged datasets to investigate the feasibility of constructing a marine-observations based historical SAM index and how this compares with existing SAM indices. Three categories of SAM indices, ChriSal SAM and full ChriSal-ICOADS SAM (generated in Section 3.2.5) are compared with three historical SAM indices (e.g. Visbeck SAM (Visbeck 2009); JW2003 SAM (Jones and Widmann 2003); Fogt2012 SAM (Fogt et al. 2012), shown in Figure 3.9.

The comparison is restricted to the period under investigation (1930 to 1960) and summer (DJF) months. This is because previous studies have shown that historical SAM indices are more certain in the summer months (Fogt et al. 2012). Figure 3.9 also shows composite correlation coefficients (denoted with triangles) with p-values for each category of SAM indices when compared with existing historical SAM indices.

The correlation between three categories of SAM indices and Visbeck SAM shows (Fig. 3.9a) that correlation coefficients for the set of SAM indices generated from 5000-pairs are between zero to 0.7, and many are insignificant at the 5% level (inside the red arc). The composite correlation coefficients for the set of SAM indices generated from 5000-pairs is
approximately 0.5 and marginally significant. By contrast, for 10000- and 20000-pairs SAM indices all correlation coefficients are significant at a 5% level, with composite correlation coefficients of 0.47 each. The correlation coefficients for ChriSal and all-observations ChriSal-ICOADS SAM index are 0.54 and 0.48, respectively. It is interesting to note that ChriSal SAM correlation is higher than all correlations despite being up to 10-times smaller than other datasets (e.g. ChriSal-ICOADS).

Similarly, the correlation between three sets of SAM indices, ChriSal-ICOADS SAM index and JW2003, is shown in Figure 3.9b. The range of correlation coefficients for the 5000-pairs SAM indices is large from negative (not shown) to almost one, due to the small size of the dataset for 5000-pairs SAM indices correlation coefficients. By contrast, none of the 10000- and 20000-pair SAM indices correlation coefficients are negative, and range from 0.5 to 0.9 and 0.6 to 0.8, respectively. It is observed that the smaller dataset (5000-pairs) SAM indices correlation coefficients have a larger range than the correlation coefficients produced from the larger dataset (20000-pairs) SAM indices. The 5000-pairs SAM composite correlation coefficient is 0.72, but insignificant at 5% level. The 10000- and 20000-pairs SAM indices composite correlations are 0.79 and 0.73, respectively, significant at 5% level. The correlation coefficients for ChriSal and all-observations ChriSal-ICOADS SAM index are approximately equal to 0.6 and significant at 5% level. Again, the ChriSal SAM correlation with the JW2004 SAM index is better than similar-sized 5000-pairs SAM indices composite correlation coefficient.

The comparison with the Fogt2012 SAM index (Fig. 3.9c) reveals that 5000-pairs SAM indices composite correlation coefficient is higher than 10000- and 20000-pairs SAM indices composite correlation coefficients, but 10000- and 20000-pairs SAM indices composite correlation coefficients have smaller p-value than 5000-pair SAM indices composite correlation coefficient (Fig. 3.9c). The correlation coefficient for ChriSal SAM index is less than both 5000-pairs and 10000-pairs SAM indices composite correlation coefficients, but higher than 20000-pairs SAM indices composite correlation coefficient. The all-observations ChriSal-ICOADS SAM index is smaller than all other correlation coefficients.

The comparison between ChriSal SAM index and three sets of SAM indices (All, 5000-, 10000- and 20000-pairs) reveals that the ChriSal SAM index covariates strongly with all three sets of SAM indices (Fig. 3.9d). The range of correlation coefficients for three sets of SAM indices is very small as compared to other comparisons (Fig. 3.9a-c). It is found that 5000-, 10000- and 20000-pairs SAM indices composite correlation coefficients vary by the size of the dataset from which they are computed, and they are observed to be in the ascending
Fig. 3.9 An adapted Taylor diagram showing correlation coefficients (the azimuthal angle) with p-values (the radial distance from the origin in inverse order) for a) Visbeck SAM, b) JW2004 SAM, c) Fogt2012 SAM, and d) ChriSal SAM indices, when correlated to ChriSal SAM (except in d), full ChriSal-ICOADS SAM, 5,000-pairs ChriSal-ICOADS SAM, 10,000-pairs ChriSal-ICOADS SAM and 20,000-pairs ChriSal-ICOADS SAM indices. The triangles denote composite correlation coefficients and p-values for the respective sample populations. The region inside the red arc is insignificant at 5% level.
order. The largest correlation and smallest p-value are found when the all-observations ChriSal-ICOADS SAM index is compared with the ChriSal SAM index, with a value of 0.82 (p-value $\approx 0$).

### 3.3.7 Trends in the historical SAM indices

The trends in the different SAM indices are computed over 20$^{th}$ Century, dividing the whole period into three 30-year segments, one of which corresponds to the period of investigation of this thesis (1930-1960) (Fig. 3.10). I have focused on summer SAM indices as numerous past studies (e.g. Visbeck 2009) have indicated that historical SAM indices derived in the seasons other than summer are highly uncertain. The trends are tested using the non-parametric Mann-Kendall trend test and reported at 5% and 10% significance levels. In the first segment (1900-1930), Visbeck SAM, JW2003 SAM and Fogt2012 SAM indices show a very weak negative trend; however, all are insignificant at 10% level. By contrast, in the third segment (1960-1990), Visbeck SAM and JW2003 SAM show positive trends but are insignificant, differing from Fogt2012 SAM which shows an insignificant negative trend over the same period. Only the Marshall SAM shows a significant trend of 0.048±0.019/year over the same period (1960-1996) at 5% significance level.

Finally, in the middle segment (1930-1960), the period of our investigation, Visbeck SAM is the only SAM index which shows a negative trend, although JW2004 SAM and Fogt2012 SAM show a similar magnitude of a positive trend but are insignificant. The ChriSal-ICOADS SAM index shows a positive trend (0.022±0.022/yr), significant at 10% level. Out of all trends computed over different time segments, only ChriSal-ICOADS and Marshall SAM indices show significant positive trends.

### 3.4 Discussion

#### 3.4.1 Spatio-temporal distribution of ChriSal dataset

In this chapter, I have analysed the spatio-temporal distribution of the ChriSal dataset consisting of meteorological observations derived from the logbooks of whaling ships in the Southern Ocean in the 1930s and 1950s (Teleti et al. 2019). I found, in agreement with previous studies (Tønnessen and Johnsen 1982), a change to a later start to the whaling season in the 1950s compared to the 1930s (Fig. 3.4). The whaling season extended from early
September to late February/early March and early December to late April/early May, for the 1930s and 1950s, respectively. This could have been due to stricter regulations, introduced to make whaling more sustainable (Tønnessen and Johnsen 1982). However, it is observed from the season-length data that to make up for the later starting dates in the 1950s, dwindling whale stocks and increased competition from other whaling companies; the seasons were stretched to catch as many whales as possible to fulfil whaling quotas until the stormy and cold winter weather made it difficult and dangerous to continue (Fig. 3.4).

Although such operations proved very detrimental to the whaling stocks, they resulted in longer-span ships’ logbooks. The length of the whaling season and general whaling practices have a bearing on our ability to record weather observations in different latitudinal bands; for example, in a given whaling season the whaling ships first entered whaling grounds in B2 band then moved into B3 and B4 bands (Fig. 3.3) as the season progressed, which is reflected in the timings of the observations dictated by the latitudinal position of that band (Fig. 3.7). Further, the greater concentration of observations in the Weddell Sea sector, followed by the ABS sector and the Indian Ocean sector, can be attributed to a number of possible reasons: first, as almost all of the whaling vessels operated from and had a supply base at South Georgia, the whalers preferred regions that were not too distant, in case ship
malfunction forced them back to South Georgia; second, the marginal sea-ice zones around the Weddell Sea and much of the Antarctic coastline were popular whaling grounds due to the abundance of whale pods and shelter provided by the ice pack, which reduces waves, during bad weather (McLaughlin 1976, Tønnessen and Johnsen 1982).

This relative lack of observations outside the Weddell Sea and ABS sectors can be closed substantially, as observations from many logbooks are yet to be extracted (Wilkinson and Wilkinson 2018). However, the gaps from May to October in the temporal coverage of historical meteorological data in the Southern Ocean are likely to remain even after extraction of additional data sources in the future, because whaling and other ships were not active during the austral winter.

### 3.4.2 Validation of ChriSal dataset

The comparison between daily MSLP observations from ChriSal and ICOADS datasets suggests that their mean values and standard deviations are quite similar (Section 3.3.2). The ICOADS observations have been checked and error-corrected using sophisticated quality assurance processes. The fact that ChriSal observations match the ICOADS data characteristics is a demonstration that the quality of the ChriSal dataset is equal to that of ICOADS. In addition, regardless of the newer models being developed and more data being assimilated, the reanalyses (20CRv2c) have failed to reproduce the MSLP variability recorded in the observations (Fig. 3.5a). The most common cause for this is attributed to unresolved model physics and parametrizations, along with the lack of comprehensive observational datasets in the region (Jones et al. 2016). Then it is imperative that reanalyses should be assimilated with newly extracted historical observations as performed in Chapter 5.

The histograms of ICOADS subsets’ correlation with individual stations (Fig. 3.6) suggest that although composite correlation coefficients are high for most stations, they are not robust as the range of 95% CI of the composite correlation coefficients is too wide (~0.1). That means that factors such as frequency of observations and matching dates in both datasets, yield a large influence on the ICOADS subsets correlations. This is particularly true with Grytviken, Signy and Bellingshausen stations. However, this does not explain why ChriSal-sized ICOADS subsets perform better than the ChriSal dataset. This is probably because the ICOADS subsets are more densely packed with observations than the ChriSal dataset. In other words, a densely packed dataset will have a stronger correlation than a sparse dataset, even when both have equal sizes.
The ChriSal correlation coefficient is usually less than the trailing $20^{th}$ percentile of ICOADS subsets correlation coefficient distribution for many stations (Fig. 3.6). Notable exceptions are Signy, where the ChriSal dataset outperforms the ICOADS subsets to be in the leading 99 percentile range of Signy’s correlation coefficient distribution. This is due to more observations in the ChriSal dataset in B3-WDL sectoral band than the ICOADS dataset in the same band and over the same period of time.

This comparison shows that even a huge dataset such as ICOADS has substantial gaps in the Southern Ocean during the historical period and that the ChriSal dataset is an important addition to the ICOADS dataset. Further, it is demonstrated that ChriSal and ICOADS have similar data qualities and adding the ChriSal dataset will enhance the ICOADS dataset. I have merged the ChriSal and ICOADS datasets to create a larger ChriSal-ICOADS dataset. The ChriSal-ICOADS dataset is used to understand long-term changes over the Southern Ocean in section 3.3.

### 3.4.3 Decadal changes over the Southern Ocean

Decadal summer changes over the Southern Ocean (inclusive of all bands) show that MSLP in the 1930s and 1950s (1950s minus 1930s) are not statistically different except in differences computed from Orcadas station records (Table 3.5). This is expected because the temporal gap between the two decades is probably not wide enough to detect large atmospheric circulation changes between them. However, it is observed that MSLP in the 1930s is higher than 1950s across all analysed bands and stations (Orcadas station records even show significant negative difference), with the only exception being a computed insignificant positive difference from the ChriSal-ICOADS merged dataset in B2. The contrasting signs of changes (positive in B2 and negative in B3 and B4) in the modern minus historical differences, hints that the different changes started to occur in the B2 and rest of the bands sometime in the historical period.

This disparity can be observed as early as the 1950s, as changes between the 1950s and the modern period (modern minus 1950s), present a distinct and diverging variation for the bands considered in this study. The B2 band as evidenced from the ChriSal-ICOADS merged dataset and Grytviken Station, experienced a significant MSLP increase from the 1950s to the modern period; however, the difference is larger in the ChriSal-ICOADS merged dataset than Grytviken Station records (Section 3.3.5, Table 3.5).
While in the B3 and B4 bands changes are mixed, an insignificant increase is observed in the B3 band, and an insignificant decrease in B4 except in the ChriSal-ICOADS merged dataset.

However, if a longer time period is considered, differences (modern minus 1930s) are negative for both B3 and B4 bands, although insignificant. The B2 band shows a significant increase over the same period of time. Hence, there is strong evidence to suggest that the B2 band MSLP have been increasing since the 1930s, while MSLP in B3 and B4 bands were decreasing over the same period. The changes are significant in the B2 band but insignificant in B3 and B4 bands, but the nature of these changes is unambiguous.

The probable reasons for this divergence could be found in the monthly variability and changes in the circumpolar trough cycles in the next section.

### 3.4.4 Monthly Variability and Semi-Annual cycle changes

The individual plotting in Fig. 3.7 of ChriSal and ChriSal-ICOADS merged datasets demonstrates the ability of the ChriSal dataset to record and exhibit MSLP variability similar to that contained in the larger ChriSal-ICOADS merged dataset. However, the ChriSal dataset has a higher standard error due to smaller numbers of observations than the ChriSal-ICOADS merged dataset; hence, the ChriSal-ICOADS merged dataset is used to compute differences. To understand the presence of low MSLP fields found in the monthly MSLP data (Figs. 3.7a, b), we need to re-examine the semi-annual movement of the circumpolar trough as highlighted by Broeke (2000). The cause of expansion (contraction) phases of the circumpolar trough twice a year is attributed to a number of factors, but mainly weakening (strengthening) phases of the mid-tropospheric meridional temperature gradient in the sub-Antarctic region (van Loon and Rogers 1984; Bracegirdle et al. 2008). The difference in the temperature gradient over the Antarctic periphery (~65°S), and further north over the open ocean (~50°S), is due to the different thermal memories of these regions.

The large thermal capacity of the ocean causes temperature changes in the air masses around 50°S to lag behind air masses over Antarctica when cooling occurs at the start of autumn. On the other hand, the negligible thermal memory of the Antarctic ice sheets helps to heat air masses over Antarctica much faster than oceanic regions at the start of spring as solar radiation levels begin to increase. These lead-lag differences in the cooling and warming of air masses over oceanic and Antarctic regions drive the twice-yearly cycles of the circumpolar trough (Meehl 1991). Hence, the expansion phase occurs from March to June.
MSLP variability in the Southern Ocean and historical SAM index

Fig. 3.11 The positions of the circumpolar trough across the months November, December, January, February and March in the a) historical period (1950s) and b) modern period as estimated from Table 3.3.

and September to December, whereas the contraction phase occurs from June to September and December to March, for the first and second cycles of the year, respectively (Meehl et al. 1998; Holland and Raphael 2006).

As the circumpolar trough transits from its northernmost position (~50ºS) to its southernmost position (~65ºS) over the summer season, the zonal bands’ (B2, B3 and B4) MSLP fields capture this movement (Table 3.3). In other words, lower MSLP values in a band relative to other bands, point to the presence of the circumpolar trough in that zonal band during that particular month (Table 3.2). The monthly analysis of the ChriSal-ICOADS merged datasets captures the contracting phase of the first cycle of the circumpolar trough as it begins its southerly movement through the bands as the summer season progresses (B2 to B3 to B4; December to March; Fig. 3.7 and Table 3.3). The pattern of monthly MSLP in various bands shows an agreement with the work of Hurrell and van Loon (1994), to the monthly pattern (shown in their Figure 7) for the months from November to March. However, this pattern is only visible in the historical period but not in the modern period, especially in the B2 band.

When Table 3.3 is inspected alongside Table 3.4, it is observed that the B2 sector has experienced a significant increase; and bands B3 and B4 decrease in MSLP over the modern period. Hence, it is hypothesised that the circumpolar trough has a restricted range of movement in the modern period. This is visually confirmed as well in Fig. 3.11a, where circumpolar trough’s positions are plotted by months according to Table 3.3 in the 1950s. It shows circumpolar trough is present over B3 band in November, B2 in December, then B3 again in January and over B4 during February and March. This pattern is altered in the modern period, circumpolar trough is present over B4 in November, B3 in December, and over B4 during January, February and March (Fig. 3.11b). It is observed that the circumpolar
trough is never present over B2 band in the modern period, as supported by a significant increase in it compared to 1950s (Table 3.4). The alteration of the semi-annual cycle is detected from the differences of MSLP from various stations in Fig. 3.8 as well.

Some previous studies, e.g. van Loon et al. (1993) have supported the idea that the circumpolar trough had stopped moving northward at the start of summer. If the circumpolar trough is not moving northwards, then it must be spending a larger proportion of summer in B3 and B4 bands as seen in Fig. 3.11b. Except in the December, circumpolar trough occupies B4 band for all other months. Interestingly, the months (November, December and January) where these modifications have occurred are, also the same months with largest changes (modern minus 1950s) in the B3 and B4 bands (Table 3.4). Therefore, it is proposed that circumpolar trough occupies B4 band (closer to the Antarctic coast) most of the summer season. In turn, this prolonged presence of circumpolar trough close to the coast for much of the summer is partially responsible for the observed MSLP decrease in the summer in the coastal Antarctic stations (Turner et al. 2005) as found in Table 3.5.

And if the circumpolar trough is occupying latitude bands closer to the coast during summer, there should be increased cyclonic frequency closer to Antarctica, consistent with the fact that circumpolar trough is known to enhance cyclonic activity (Simmonds and Keay 2000). Hence this proposal is supported by many studies which found a significant increasing trend in the number of cyclones around the Antarctic coast in summer (Grieger et al. 2018). The combined deepening of Antarctic station MSLP and increased cyclonic activity will have profound implications on the long-term circulation changes, e.g. SAM index and transport of moisture inland, thus affecting mass-balance of the ice-sheets.

3.4.5 Choice of high-latitude band

The latitudinal bands used in previous studies were not fixed but instead dependent on the availability of meteorological data in the latitude bands close to SAM’s action centres (Jones et al. 2009). However, in this chapter, we are free to choose any zonal band as long as it captures full SAM variability in the period of investigation. To select the high-latitude band, I correlated various zonal bands in the ChriSal-ICOADS merged dataset to find which of these bands had the strongest relationship with the 40ºS zonal band. As SAM is essentially a dipole of mid-latitude and high-latitude pressure fields, the stronger the (negative) relationship between the 40ºS zonal band and one of the ChriSal-ICOADS merged dataset zonal bands (high-latitude band), the stronger the possibility of the action centre being located in that
The strongest correlation was found between the 40ºS zonal band and the B3 (60º-65ºS; Fig. 3.3) zonal band with a correlation coefficient r = -0.37 (significant at 5% level).

### 3.4.6 Marine observations-based historical SAM and its trend

Many new marine observations-based SAM indices are generated using diverse data sources. The SAM indices are restricted to summer season: this is when the SAM is most annular in form, whereas in winter it is most asymmetrical (Fogt et al. 2012). In addition, the summer season is when the observations from which SAM indices are generated are most reliable. Previous studies have also observed that SAM reconstructions are most reliable in the summer due to annular SAM structure and availability of a greater number of stations for reconstruction (Jones et al. 2009). I have compared the season in which SAM indices are considered most robust; hence, the relationship derived from them can also be considered robust.

The first SAM index generated in this chapter is the ChriSal SAM index, made from using normalised zonal means of B3 zonal band from the ChriSal dataset and six mid-latitude station records. The ChriSal SAM index is not a completely marine observations-based SAM index as station records are used to compute the mid-latitude band (Section 3.2.3). This SAM index was generated to examine if a small ChriSal-sized dataset can be used in combination with mid-latitude station records, to generate an observations-based SAM index.

The ChriSal-ICOADS SAM index is a genuinely marine observations-based SAM, as both high-latitude and mid-latitude zonal bands are generated using zonal means of the B3 zonal band and 40º S zonal band from the ChriSal-ICOADS merged dataset, respectively. It is the longest of the SAM indexes generated in this study, as other SAM indices are generated using the randomly-drawn subsets of the ChriSal-ICOADS merged dataset. The number of observations in each of these subsets is chosen to compare the impact of dataset size on the SAM index generated to represent the full variability of the MSLP fields.

Three categories of subsets are generated in this study, containing 5000-, 10000- and 20000-pairs of random-ordered observations from the ChriSal-ICOADS merged dataset. As the primary purpose of these subsets is to generate a diverse range of SAM indices, it is hoped that each pair of observation will contain one mid-latitude and one high-latitude observation, but because I draw observations in random order, it may or may not be true for each pair of observations or even for the whole subset. However, as I draw a large number of subsets (1000), this collection of subsets envelopes all possibilities (from no observations in
3.4 Discussion

either zonal bands to almost all observations in one of the zonal bands). It is evident that the larger the subset, the higher the chances of generating a densely packed SAM index.

Hence, a complete range of SAM indices is generated (from near-empty SAM index to near full SAM index time series) when they are generated from this collection of subsets. Due to limits on the maximum number of observations in each collection of subsets (e.g. 5000-, 10000- and 20000-pairs), a limit is placed on the SAM indices generated from these subsets to represent full MSLP variability over the Southern Ocean. The 5000-pairs subsets are very close to the ChriSal dataset in size, and the SAM indices generated from them work as a benchmark to assess the ability of ChriSal SAM index to explain the variability observed in the MSLP fields in the form of SAM index and also its closeness to existing historical SAM indices.

The shorter ChriSal SAM index outperforms much larger and denser SAM indices when compared to Visbeck SAM index (Fig. 3.9a). The ChriSal SAM index can explain more than 25% of the variability (more than others) in the Visbeck SAM during the period of investigation. This could be due to the fact that Visbeck SAM and ChriSal SAM indices share the same set of mid-latitude station records. The full ChriSal-ICOADS SAM index also correlates well with Visbeck SAM, but the correlation coefficient is smaller than the ChriSal SAM index, which suggests that all other factors being equal, the shared mid-latitude stations have a favourable impact on the ChriSal SAM index.

The comparison with JW2003 shows that both ChriSal and full ChriSal-ICOADS SAM indices correlation coefficients are smaller than SAM indices generated from 5000-, 10000- and 20000-pairs of observations from the ChriSal-ICOADS merged dataset. This is due to the process of JW2003 SAM index generation, and as Jones and Widmann (2003) note, the JW2003 is generated by fitting relationships derived from ERA-40/station MSLP data to SAM index generated through Principle Component Analysis from the ERA-40 data between 1958-2001. Many studies have also noted that reanalyses such as ERA-40 are unreliable in the Southern Ocean before the start of assimilation of satellite-derived data in 1979. It is possible that the SAM index derived from the ERA-40 data is unreliable; hence, larger the marine dataset from which the SAM indices are generated provides smaller the agreement between JW2003 SAM and that SAM index (Fig. 3.9b).

This pattern is repeated in Fig. 3.9c, where the SAM indices are compared to Fogt2012 SAM, the highest correlation is found with SAM indices generated from 5000-pairs of observations. However, the correlation coefficient for the full ChriSal-ICOADS SAM index when compared to Fogt2012 SAM is marginally better than the one found between it and
the JW2003 SAM index. This is possibly due to a slight difference in the generation of JW2003 SAM and Fogt2012 SAM indices. Even though both are generated from station dataset, Fogt2012 SAM was fitted to an observational SAM index, Marshall SAM. However, it does not explain the consistently higher correlation coefficients found for SAM indices generated from 10000- and 20000-pairs of observations from the ChriSal-ICOADS merged dataset. Even though they are artificially generated, it seems that 10000- and 20000-pairs sized datasets are optimal for measuring the relationship between marine observations-based SAM indices and existing SAM indices.

In other words, the size and composition of the dataset used to generate a SAM index are found to have an impact on the correlation between that SAM index and other SAM indices. For example, the SAM index generated from a much larger marine dataset (full ChriSal-ICOADS dataset), is less similar to JW2004 and Fogt2012 SAM indices than the ones generated from a smaller dataset (5000-pairs observations ChriSal-ICOADS dataset). On the contrary, the Visbeck SAM index comparison is more favourable to the SAM indices generated from larger marine datasets, for example, composite correlation coefficients for all three categories of SAM indices are almost same, but the p-values vary from 0.02 to $10^{-5}$, in the order of the sizes. However, there is one exception; that is ChriSal SAM index’s correlation coefficient (0.54) which is higher than the one obtained from full ChriSal-ICOADS dataset (0.48). Finally, the correlation between ChriSal SAM index and other SAM indices generated in this study reveal that the higher the size of the dataset, the higher the correlation. In other words, smaller ChriSal SAM index can explain more than 64% variability in the SAM index generated from full ChriSal-ICOADS dataset.

The trends analysed in Fig. 3.9 reveal that SAM indices have undergone different changes over different periods of time. In the first 30-yr period (1900-1929), all three SAM indices (Visbeck, Fogt2012 and JW2004) show a slightly negative trend although the uncertainty associated with each index is large; hence, none of the trends are significant at a 5% level. In the next 30-yr period, which is also the period of investigation in this study, it is shown that (except Visbeck SAM) all SAM indices are positive. The ChriSal-ICOADS SAM is the only one with a significant trend (0.022±0.022/yr) over this period at 10% significance level.

In the next time period, from 1960 to 2005, the Marshall SAM shows a highly significant positive trend of 0.048±0.019/yr at 5% significance level. Even though other indices (Visbeck, Fogt2012, JW2003) largely follows the Marshall SAM, they fail to show any trend significant or otherwise (Fig. 3.9). These trends have been reported in previous studies as well, using 30-yr periods (Fogt et al. 2012). It is interesting to note that only Marshall SAM and ChriSal-
ICOADS SAM show a significant positive trend; it is proposed that a positive trend in the SAM index started as early as the 1930s. These findings are in line with our knowledge of MSLP changes over mid-latitude and high-latitude zonal bands. As this study has found a decreasing MSLP trend in the high-latitude zonal bands, a positive SAM index can be expected. In the same vein, if MSLP in high-latitude zonal bands has been decreasing since the 1930s, before the start of the Marshall SAM in 1957, a positive trend SAM is to expected as well.

3.5 Conclusions

In conclusion, the whaling dataset (Chapter 2) has been found to be reliable and similar to other Southern Ocean observational datasets in quality for climatological studies. The seasonal atmospheric circulation over the Southern Ocean was found to undergo cyclic changes in the form of expansion and contraction of the circumpolar trough, confirming previously identified circumpolar trough cycles. However, significant alterations to the circumpolar trough cycles have been found in the modern period; it occupies B4 band (closer to the Antarctic coast) most of the summer. Longer time series of observational data will be required to confirm these changes and the causes behind them.

The MSLP increases in the northern reaches are attributed to the general increase in MSLP in the mid-latitudes (Cai et al. 2003). While MSLP decreasing polewards of 60°S, is attributed to ozone depletion over Antarctica. The exact mechanism by which signal from the loss of stratospheric ozone descends to the surface level is under active investigation, probable pathways and effect of other forcing factors on the MSLP variability are examined in Fogt et al. (2009). However, this work proposes that prolonged presence of circumpolar trough near Antarctic coast during summer could be partly responsible for decreasing MSLP trend found in the Antarctic coastal stations during the austral summer.

The new marine observations-based SAM index generated in this chapter correlates with existing SAM indices to varying degrees during the period of investigation (1930-1960). The differences occur between the SAM indices due to the data and methods used to generate them. The trend analysis of the SAM indices reveals that, out of all SAM indices considered, only the SAM index generated in this study (ChriSal-ICOADS SAM index) and the Marshall SAM index show significant trends over any time period investigated. Considering MSLP changes found in the region polewards of 60°S, it is proposed that a positive trend in the SAM index started in the historical period, and is continuing in the modern period. Hence,
future studies should be directed to investigating the long-term relationship between regional Southern Hemisphere climate and the historical SAM index.

An additional line of research could be to compare the new SAM index to historical cyclones recorded in the whaling or similar datasets, as many previous studies have found that large-scale climate modes (e.g. SAM, ENSO) have an impact on Southern Ocean cyclones’ intensity and frequency (Pezza et al. 2008, Grieger et al. 2018). In addition, as it is shown in this chapter that the circumpolar trough is restricted close to the Antarctic coast during summer, a comparison should to made to provide quantification of the relationship between the circumpolar trough and cyclones observed in the historical period vis-a-vis the modern period. In addition, the present Christian Salvesen whaling dataset could be expanded to include many recently digitised logbooks from the same company to make the dataset more robust in terms of time series and locations (Wilkinson and Wilkinson 2018). The extended dataset and, in turn, a new SAM will be beneficial for assessing long-term changes in the Southern Ocean climate.
Chapter 4

A novel approach for historical cyclone detection

In this Chapter, I demonstrate how to utilise historical whaling meteorological data with innovative analyses to identify historical cyclones in the Southern Ocean.

4.1 Introduction

The Southern Ocean’s intense cyclonic activity and a large number of cyclo-genesis events throughout the year are attributed to strong baroclinicity (Simmonds and Lim 2009) and surface fluxes of latent and sensible heat (Uotila et al. 2011). Unlike the Northern Hemisphere, the Southern Hemisphere stationary waves do not contribute significantly to poleward energy transfer at any time of the year, and hence, transient eddies play a major role in balancing radiative energy disparity between Southern Hemisphere mid-/tropical latitudes and high-latitudes (Peixoto and Oort 1992, Grieger et al. 2018). In doing so, eddies create fertile conditions in the form of latent and sensible heat exchanges that lead to observed cyclonic activity.

This cyclonic activity itself has been subject to changes over the recent decades. As previous studies have found, large-scale climate modes, for example, the Southern Annular Mode (SAM) and El Niño-Southern Oscillation (ENSO), have impacted on the intensity and frequency of Southern Hemisphere cyclones (Pezza et al. 2008). As an example, when tracking cyclones in the Southern Ocean south of 60°S, Grieger et al. (2018) found a positive trend in the cyclone frequency in the Austral summer over the 1979-2008 period, which is
possibly linked to the positive summer trend of SAM during the same period (Simmonds 2015). These observations are supported by many climate modelling studies which have suggested that increased cyclonic activity accompanies the predicted stronger SAM over the Southern Ocean (e.g. Lynch et al. 2006). Furthermore, the observed increase in the number of cyclones around the Antarctic Peninsula in recent years (Lubin et al. 2008) is suggested to be one of reasons for the rapid warming of the Antarctic Peninsula (Marshall 2007) and a reduction in the sea-ice extent in the region to its west (Parkinson 2019).

Cyclones play a major role in the heat exchange between sea and air by exerting dynamic and thermodynamic forces on sea ice (Simmonds and Wu 1993). One of the earliest studies to link cyclonic variability to Antarctic sea-ice was Howarth (1983), who found that cyclones have a major impact on Antarctic sea-ice at a regional scale. A more recent study investigating satellite-tracked sea-ice motion vectors found large and statistically significant trends that are strongly correlated to local cyclonic winds (Holland and Kwok 2012). Raphael et al. (2019) also found that increased cyclonic activity in the Ross Sea has contributed to freshening and increasing sea-ice in that region (Simmonds 2015). Holland and Kwok suggested that underestimation of cyclonic winds in Global Climate Models (GCMs) is a major reason behind their consistent failure to simulate increasing Antarctic sea ice, even in the modern satellite observations period (Zhang 2007, Eisenman et al. 2011, Zhang 2013).

The influence of cyclones is not just restricted to sea ice. The cyclonic systems transport moisture from mid-latitudes to the Antarctic continent, thus playing a significant role in the mass balance of the Antarctic ice sheets (Grieger et al. 2018). Noting the influence of these synoptic systems on the weather and its extremes over the Southern Ocean and in the coastal and continental regions of Antarctica, it is important to detect and quantify cyclonic activity over the Southern Ocean. Cyclone tracks for the Southern Hemisphere have been relatively under-examined compared with their northern counterparts, for a number of reasons that include both data sparsity and difficult logistics.

The first major initiative to improve regional synoptic forecasts and to better understand the behaviour of Southern Ocean cyclones, the international Antarctic First Regional Observing Study of the Troposphere (FROST) project, was organized by the Scientific Committee on Antarctic Research (SCAR) (Turner et al. 1996). Since then, many studies have proposed a number of cyclone tracking algorithms (e.g. Murray and Simmonds 1991; Sinclair 1994; Hodges 1995; Serreze 1995; Blender et al. 1997; Wernli and Schwierz 2006; Uotila et al. 2011). Each algorithm makes decisions based on known specific thresholds of a physical parameter or combination of parameters to find and track cyclones using a variety of datasets.
Hence, results from individual studies may differ even when investigating the same region and over the same time period. To facilitate comparison, the multi-institutional Intercomparison of Mid LATitude STorm diagnostics (IMILAST) project was set up to provide a uniform platform to compare algorithms for the identification and tracking of extra-tropical cyclones (Neu et al. 2012). The IMILAST project produced global cyclone tracks and data starting from 1979 until 2011, using a common input dataset from the ERA-Interim reanalysis (Dee et al. 2011). These algorithms and studies have been valuable in identifying the key climatological features (including the variability and trends) of Southern Ocean extra-tropical cyclones.

However, records of cyclone frequency and characteristics in the pre-satellite era are relatively few. Historically, the Southern Hemisphere suffers from a lack of sustained climate observations in general, and cyclone observations are even poorer given that large numbers of cyclones do not make landfall and even those who do were not always observed due to sparse human settlements and meteorological stations (WMO 2019, Harper et al. 2008). To address this, a comprehensive global tropical cyclone dataset, the International Best Track Archive for Climate Stewardship (IBTrACS; Knapp et al. 2010) and a number of cyclone track datasets focusing on Southern Hemisphere oceans have been developed and published, including the Southwest Pacific Enhanced Archive of Tropical Cyclones (SPEArTC; Diamond et al. 2011), and the Australian Bureau of Meteorology National Climate Centre (BOM NCC; Trewin 2010). These efforts have indeed led to some insights into the background frequency of cyclones but are wholly inadequate to investigate long-term changes. For example, the period after the 1960s shows a sudden jump in the number of cyclones detected, likely due to the underestimation of cyclones in the open seas prior to satellite observations (Trewin 2010, Magee et al. 2016).

In such a scenario, it is important to make the best use of commonly available meteorological data (that is, ship observations) to detect open-sea cyclones in the absence of satellite observations to fill in the gaps in the Southern Ocean data record. The combination of ship and land observations has been used to detect and forecast cyclone tracks, intensity and timing of landfall since the early 20th Century (Landsea 2007). The majority of literature on historical cyclones is focused on the North Atlantic hurricanes. Those studies offer useful insights, but the methods developed for the North Atlantic Ocean cannot be applied to the Southern Ocean due to climatological differences. The historical hurricanes in the National Hurricane Center’s (NHC) North Atlantic Hurricane Database (HURDAT2; Landsea and Franklin 2013) were identified and added as long as an individual storm system
was reported by at least two separate observations of sustained tropical storm force winds (> 34 kt) or the equivalent in sea-level pressure (roughly 1005 mb or lower; Landsea et al. 2008). This provides the simplest and most straightforward method to detect historical cyclones; that is, to inspect ships’ logbooks for low air pressure or high wind observations, as has been undertaken by numerous studies investigating historical Atlantic hurricanes (e.g. Fernández-Partagás and Diaz 1996, Landsea et al. 2004).

The above method of using ship observations directly could lead to over- or underestimation depending on the circumstances. For example, the number of cyclones would be over-estimated in the Southern Ocean due to the presence of quasi-permanent moving polar lows. On the other hand, if direct ship observations are used to detect cyclones, the ship has to be relatively close to the storm system; however, ships usually steered away from cyclones for safety (Admiralty 1938). This could lead to gross underestimation, especially if many cyclones formed and moved at a distance from ships (Landsea et al. 2008). Even after the presence of a cyclone has been identified, it is difficult, without additional information, to estimate an accurate location and intensity (Neumann and Hammer 1999).

In this chapter, I aim to address the drawbacks highlighted above by creating new and adapting existing methodologies to identify and track individual cyclones in the Southern Ocean using historical whaling logbook data (Teleti et al. 2019 and Chapter 2 in this thesis). Two separate methods are employed to estimate the number of cyclones in the immediate vicinity and at a distance from observing ships. In the first method, to detect cyclones in the close vicinity of the observing vessel, I use a metric proposed by Lim and Simmonds (2002) that measures pressure tendency (rapid fall and rise of pressure observations) while adjusting for climatological conditions at the latitude of observations. To detect cyclones forming and moving at a distance (around 500 km) from the ships, the second method uses a statistical model which utilises all useful variables recorded in the whaling dataset to identify a cyclone and estimate its position and intensity. Further, to automate the detection of cyclones through the latter method, I have produced a semi-supervised cyclone detecting and tracking algorithm that can be used in the analysis of similar datasets in the future.

This chapter is divided into several sections. Section 4.2 briefly describes the different datasets used in this study, such as the Christian Salvesen whaling dataset, and I expand on the methods used to count the number of ship encounters with cyclones and to generate a linear model to detect and track cyclones. The results of applying these methods to historical whaling dataset are presented in Section 4.3, showing the number of cyclones detected by different methods and the sensitivity of the cyclone detecting algorithm. In
Section 4.4, I discuss the results in a broader context and expand on the effectiveness and uncertainties of using historical observations to detect historical cyclones and comparison of cyclonic frequency in the historical and modern period. Finally, in Section 4.5, I conclude by summarising the issues addressed in this study and offering suggestions on the further use of the cyclone detection algorithm with historical meteorological datasets.

4.2 Data

I primarily use data from two distinct time periods: historical Christian Salvesen whaling logs and modern ICOADS meteorological observations. The modern data are used to produce a linear regression model based on the relationship between simultaneous cyclone and weather observations. Subsequently, the model is applied to the historical data to detect and track historical cyclones.

4.2.1 Christian Salvesen whaling dataset (historical)

The Christian Salvesen whaling dataset (Teleti et al. 2019, Chapter 2) contains meteorological observations extracted from logbooks of whaling ships operating in the Southern Ocean during the 1930s and 1950s. It is described in detail in Chapter 2. In this Chapter, I focus on mean sea-level pressure (MSLP) and wind observations from the Christian Salvesen dataset to detect and track historical cyclones.

4.2.2 ICOADS meteorological observations (modern)

The International Comprehensive Ocean-Atmosphere Data Set 3.0 (ICOADS; Freeman et al. 2017) is a comprehensive collection of in-situ marine meteorological observations, mainly from ships and buoys, derived from many different national and international data sources. It contains raw meteorological observations from 1662 to 2014. It is described in detail in Chapter 2. In this Chapter, I use observations recorded between 1999-2008 in the months of November, December, January, February and March and located south of 30°S latitude. This selection reflects the seasonal and spatial spread of the whaling dataset.
4.2.3 Cyclone track datasets

A number of national/international weather agencies and studies have produced cyclone tracks and data, based on different algorithms and procedures. However, a clear distinction can be made based on the type of data used. The IMILAST project uses ERA-Interim reanalysis dataset whereas IBTrACS uses ship, surface, and satellite observations to produce such tracks.

**IMILAST**

A total of 15 cyclone tracking algorithms contributed to the IMILAST project (a multi-institutional study generated different cyclone tracks, Neu et al. 2012). A common input dataset, the 6-hourly ERA-Interim reanalysis (Dee et al. 2011), was used to detect cyclones globally. The global cyclone tracking data were generated from 1 January 1979 to 31 March 2009, with each cyclone tracked at 6-hour intervals. For this study, I have chosen tracking algorithms denoted by M02 and M18 from the IMILAST database (https://naturalsciences.ch/organisations/proclim/activities/project_imilast/data_download), to reflect the different parameters used to identify cyclones (e.g. minimum MSLP in M02 and vorticity at 850 hPa in M18). More details regarding individual cyclone identification schemes can be found in Neu et al. (2012).

**IBTrACS**

The IBTrACS (Knapp et al. 2010) dataset collates best-track position and intensity observations from many World Meteorological Organization (WMO) Regional Specialized Meteorological Centres (RSMCs) and Tropical Cyclone Warning Centres (TCWCs), as well as other archived sources. The IBTrACS dataset consists mainly of post-season reanalysis of cyclone positions and intensity from all available data sources (e.g. ship, surface, and satellite observations by the contributing agencies worldwide). The cyclones are usually tracked at 6-hour intervals by most agencies, but some do so at 3-hour intervals. The core characteristics of storms included in the dataset are position (latitude and longitude), wind (mean, minimum, maximum), cyclone centre/core pressure (mean, minimum, maximum) and certain additional details (e.g. ocean basin, season).
4.3 Methodology

To identify and track historical cyclones, I employ two methods. 1. Using relative MSLP tendency (rapid fall and rise of pressure observations) to count occurrences of whaling ships’ encounters with cyclones. 2. The historical whaling data is passed through a linear regression model developed from modern meteorological data. Each method is described in detail in the sections below.

4.3.1 Method One

The ships’ MSLP observations alone cannot reveal the existence of cyclones in the wider sectoral area around the ship (100s of km distant); however, they can provide an estimate of the number of cyclones in the immediate vicinity of ships by identifying the number of cyclone encounters made by the ships. As supported by many whaling ship logbook entries containing many mentions of storm in the written weather descriptions.

Before this is done, however, what constitutes a cyclone needs to be defined. Cyclonic systems have a broad definition but Pezza and Ambrizzi (2003) and subsequent studies defined cyclones as systems with core pressure below 980 hPa in the Southern Ocean. I use the same definition of cyclones to identify cyclones from the historical data. To further assist cyclone identification, I also use the core pressure tendency. Here I follow the relative central pressure normalized deepening rate ($NDR_r$) proposed by Lim and Simmonds (2002) and Lim and Simmonds (2007), which is defined as:

$$NDR_r = \frac{P_r}{24hPa/day} \times \frac{\sin 60}{\sin \Theta}$$

where, $P_r = P_c(\lambda, \Theta) - P_{mean}(\lambda, \Theta, t)$ and $\Delta P_r$ is the $P_r$ change over 24 hours.

$P_r$ is the difference between $P_c$, the core cyclone pressure at the location $(\lambda, \Theta)$ and $P_{mean}$, the climatological mean MSLP at the longitude $\lambda$, latitude $\Theta$ and time $t$. The climatological mean is used to account for the presence of quasi-permanent polar lows in the Southern Ocean. Although the $NDR_r$ has been used to identify explosive cyclonic systems by Lim and Simmonds (2002), it can also be used to detect cyclone encounters made by ships as, from a seafarer’s point of view, when a vessel is in a cyclone’s path or in close proximity, MSLP will fall rapidly. The $NDR_r$ generated from ship MSLP observations can help identify these rapid deepening events. In the same vein, when cyclones and ships move away from
each other, MSLP rises quickly again. This rapid fall and rise in the MSLP is counted as one encounter with a cyclone.

To count the number of cyclone encounters, all ShipIDs (each of which is a unique combination of vessel name and season) in the whaling dataset that contained MSLP observations less than 980 hPa were selected. $NDR_r$ was then calculated using Eq. 4.1 for each ShipID by replacing $P_c$ with ship MSLP observations and 30-year mean (1981-2010) monthly ERA-Interim MSLP data was used as background climatology. One cyclone encounter was counted when the difference between a pair (to account for fall and rise of pressure) of adjacent $NDR_r$ values exceeded 0.6 units, the choice of this threshold value is discussed in Section 4.5.1

4.3.2 Method Two

This method takes a different approach to locate approximate cyclone positions using ships’ MSLP and wind observations. The positional information about a cyclone can be divided into direction and distance components with respect to the ship from which observations were taken. The direction of cyclones with respect to a ship can be determined by azimuth angle ($A_z$), $A_z = D - 90^\circ - I$, where D is direction of the wind, I is Inflow angle of the wind (the angle between the wind direction and a tangential direction or, defined as the local trajectory of mass transport to the storm centre) for the Southern Hemisphere (Landsea et al. 2004). Boose et al. (2001) and Boose (2004) successfully determined the impact of historical hurricanes in North America using an inflow angle of $20^\circ$. Similarly, Landsea et al. (2004) used an inflow angle of $20^\circ$ to triangulate the centre of cyclones from multiple wind observations (their Fig. 2). However, in the light of a more recent empirical study (Zhang and Uhlhorn 2012) to measure inflow angle and to account for the approximation of wind directions in the ship logbooks by manual observations (Teleti et al. 2019), an inflow angle of $25^\circ$ is used here. Even so, the results of this study are not affected significantly by using either value.

Once I have derived the direction of the cyclone/depression centre relative to the ship, distance can be determined by applying a linear multi-regression model using the Ordinary Least Squares (OLS) method; input parameters included ship MSLP observations, distance between the ship and detected cyclone centre, zonal and meridional (UV) wind components as independent variables, and observed cyclone-centre core pressure as the dependent variable (Eq. 4.2). I have used this particular set of predictors because these variables are readily
available in both the historical and modern datasets used in this study, and pressure and winds are known to respond strongly to the changes in the atmosphere, thus facilitating the construction of a physically robust model. The UV wind components are used instead of wind speed because UV components account for the direction of the wind as well as strength. As shown later in the section, wind direction information is crucial to identify and track cyclones.

\[ P_{\text{cyclone}} = \alpha + \beta_1 P_{\text{ship}} + \beta_2 D_{\text{ship}} + \beta_3 U_{\text{ship}} + \beta_4 V_{\text{ship}} + \varepsilon \] (4.2)

To build the linear model, I have used modern cyclone tracks and meteorological data recorded between 1999-2008. The time-period was chosen to take advantage of the peak number of observations recorded both in-situ and from satellites before observations tailed off post-2008 (Freeman et al. 2017). The 10-year period provides sufficient data points to build a reliable model (for MSLP observations; WMO 2007). First, in order to choose which set of cyclone tracks to use to build the model, I compared a selection of cyclone tracks identified by M02 and M18 (denoted as M02, M18 by Neu et al. 2012) in the IMILAST dataset.

As an example, Figure 1.4a shows ‘Cyclone Jo’ forming near the northern part of Vanuatu (∼17°S 166°E) and travelling south-eastwards over the succeeding two weeks. Geographically, the M02, M18 methodologies closely match IBTrACS’s cyclone track. The IBTrACS dataset recognised Cyclone Jo as a system on 23-Jan-2000 at 00:00:00 whereas M02 and M18 detect it at 25-Jan-2000 06:00:00 and 22-Jan-2000 12:00:00, respectively. The core cyclone pressures through time reported by the three data sources are found to be in general agreement (Fig. 1.4b). This comparison and ten others (not shown) point to the consistent depiction of cyclones in both observations-based (IBTrACS) and reanalysis-based (M02, M18) cyclone datasets, and indicates that information from either dataset can be used to build the linear regression model.

Thus, the data and tracks generated by each methodology were used to model relationships between simultaneous ICOADS weather observations and cyclone data. Table 4.1 shows three models constructed using M02 (Linear Model M02; LM02), M18 (Linear Model M18; LM18) and IBTrACS (Linear Model IBTRACS; LIBTRACS) cyclone tracks with ICOADS observations between 1999-2008. To match the spatial and seasonal characteristics of whaling data, all cyclone tracks and ICOADS observations were filtered to retain only data south of 30°S and for the months of November, December, January, February and
Fig. 4.1 a) Cyclone Jo tracked by IBTrACS, M02 and M18 methodologies in January-February 2000. b) Core pressure of the cyclone from each methodology (hPa). The origin of the cyclone near Vanuatu is encircled in black.
March. Only observations taken within 3 hours (same time interval used by the National Hurricane Centre and Brown et al. (2006) to derive pressure-wind relationship) and 500 km (this restriction is discussed in a later paragraph) of the cyclone time-step and location were retained.

Clearly, each model offers some advantages and disadvantages in the detection of cyclones. Hence, to choose the model to use for detecting historical cyclones, I rely on the metrics generated from these models. The number of suitable data points shows a large variation, with the model built around M18 having the highest number of data points, followed by M02 and IBTrACS (Table 4.1). The metrics chosen to compare are Akaike Information Criteria (AIC), Proportion of Variance Explained (PVE) and F-value. I start by using Ship MSLP observations as the only predictor to model cyclone core pressure, with PVE reaching 0.87 for the M02 and M18 models but remaining low at 0.34 for the IBTrACS model. However, the AIC and F-values remain high, which suggests sub-optimum fit, in which the addition of other predictors could benefit model fit. Subsequently, adding distance between cyclone centre and ICOADS observations as a predictor improves the fit marginally for all models.

Finally, a full model using ship MSLP, distance and ship UV wind components (Eq. 4.2) is shown to improve the fit even further, especially when F-value is considered. Models LM02 and LM18 have the highest PVE among the models for full-set predictors. However, the LM02 model has the lowest AIC and F-values, which suggests that the LM02 model is well-fitted and best-suited among the compared models to represent the relationship between cyclone core pressure, ship observations and positions. Low \( p \)-values (< 0.05) indicates that the null hypothesis: coefficients have no effect, can be rejected for all cases.
Table 4.1 Model statistics using different combinations of variables (rows) for different cyclone core pressure observations (columns) to fit OLS linear model to ICOADS raw observations (All coefficients are significant at the 95% level except $\beta_3$ and $\beta_4$ coefficients of the IBTRACS model). AIC is Akaike information criteria (AIC), PVE is Proportion of Variance Explained (PVE) or unadjusted r-squared and p is p-values.

<table>
<thead>
<tr>
<th>Predictors</th>
<th>LM02 (n=19050)</th>
<th>LM18 (n=27116)</th>
<th>LIBTRACS (n=80)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>AIC PVE F (p)</td>
<td>AIC PVE F (p)</td>
<td>AIC PVE F (p)</td>
</tr>
<tr>
<td>Ship MSLP</td>
<td>117170 0.87 1790400 (0)</td>
<td>163510 0.87 192270 (0)</td>
<td>512.24 0.34 39.50 (0)</td>
</tr>
<tr>
<td>Ship MSLP + Distance</td>
<td>115340 0.88 1068500 (0)</td>
<td>161700 0.88 103750 (0)</td>
<td>495.98 0.47 34.37 (0)</td>
</tr>
<tr>
<td>Ship MSLP + Distance + UU + VV</td>
<td>115200 0.88 34933 (0)</td>
<td>161520 0.88 52261 (0)</td>
<td>499.87 0.47 16.78 (0)</td>
</tr>
<tr>
<td>Coefficients (full model)</td>
<td>53.86 0.946 -0.013 -0.046 -0.036</td>
<td>44.48 0.957 -0.010 -0.049 -0.015</td>
<td>147.21 0.852 -0.029 0 -0.020</td>
</tr>
</tbody>
</table>
4.3 Methodology

Now using the LM02 model, I can predict the core pressure of hypothetical cyclones in the vicinity of our ships by inputting whaling ship data from the 1930s and 1950s to the regression model. However, the pressure at the centre of the cyclone and the size of the system are inter-dependent. The size of a synoptic weather system changes throughout its life cycle as Simmonds and Keay (2000) states, there is no single best definition of size or radius of such systems. Rudeva and Gulev (2007) found that the effective radius varies from 700 to 800 km for cyclones with core pressure deeper than 980 hPa, from 500 to 600 km for the moderate cyclones (980–1000 hPa), and from 350 to 500 km for the shallow cyclones (>1000 hPa) for both winter and summer. Carrasco et al. (2014), Chavas et al. (2016) and other studies have also found that radii for moderate cyclones were about 500 km.

As a trade-off between capturing an adequate number of possible cyclones and detecting genuine cyclones, I consider weather systems as cyclonic if their core pressure is 980 hPa or less and systems with core pressure over 980 hPa are considered depressions. I intend to identify hypothetical moderate cyclones that are assumed to have radii of 500 km; hence, I estimate the pressure at 100, 200, 300, 400, 500 km away from the ship in the direction derived from wind directions, by passing whaling ship observations into linear model LM02. Subsequently, I compute the probability of finding a moderate cyclone as a function of the MSLP observations at the ship (hPa), and the distance between the ship and the centre of the hypothetical cyclone (km). The probability distribution is smoothed by passing a 2-point (2 × 2) binomial filter over the data, as shown in Fig. 4.2a. As expected, moving away from the ship increases the possibility of finding a cyclone; there is more than a 50% probability if a ship’s MSLP observations record 985 hPa or lower. Hence, now the presence of a cyclone can be expressed as a probability % at 100, 200, 300, 400, 500 km from the ship.

Fig. 4.2b shows an illustration of the methodology described above, with ship positions overlaid on the MSLP background field, ship’s wind observations denoted by wind barbs, and grey-shaded lines representing direction-lines protruding from the ship towards the cyclone. Darker to lighter shades represent 100, 200, 300, 400, 500 km from a ship, with the probability of finding a moderate cyclone at that distance shown in parenthesis. To distinguish direction-lines with high probability, ‘×’ is marked at the end of direction-lines when probability exceeds 50% at any distance from the ship (Fig. 4.2b). Frequently, a single cyclone would encounter multiple ships close to its path during its lifetime. The cumulative probability of detection for a hypothetical cyclone (as shown in Fig. 4.2b) is based on the number of observations and is computed as follows:
Fig. 4.2 (a) Probability (in %) for the predicted pressure to be 980 hPa or lower, as a function of the MSLP observations at ship (hPa) and the distance of the ship from the centre of the hypothetical cyclone (km). (b) An illustration of distance-MSLP ship observations relationship with a probable cyclone present. The ships’ locations are denoted by blue dots with wind barb showing wind direction and strength. The grey-shaded lines represent direction-lines protruding from the ship towards the probable cyclone (red ©), darker to lighter shades represent 100, 200, 300, 400, 500 km from the ship. The probability (%) for the predicted pressure to be 980 hPa or lower is shown along the grey-shaded lines with distance in parentheses. The ‘×’ mark is placed at the tip of the direction-line if probability exceeds 50% at any distance from the ship.
4.3 Methodology

\[ P_{Detect} = 1 - \prod_{1}^{n} (1 - p_n) \]  

(4.3)

where \( p_1, p_2, p_3 \ldots p_n \) represent probabilities derived from individual ships and \( P_{Detect} \) is the cumulative probability of existence of cyclone.

As in the illustration (Fig. 4.2b), two direction lines are pointing to a small area on the map, with the combined probability of finding a cyclone at the extended intersection between the direction-lines more than 80% (using Eq. 4.3), which is larger than the individual (54% and 56%) probabilities. Hence, the more direction-lines, especially from different ships pointing to the same area, the higher the chances of finding a cyclone at the extended intersection between the direction-lines. Each estimated cyclone position is denoted by © in red (Fig. 4.2b), and individual direction-lines without a ‘×’ are considered to be pointing to depressions rather than cyclones.

### 4.3.3 Randomised ANOVA test

When the sizes of groups to be compared are small (with unknown distribution), randomised ANOVA tests could be performed to test the null hypothesis that different groups have the same mean. To do so, first, the F-statistic \( F_{ORG} \) is calculated from ANOVA tests on the original data groups. Second, all observations irrespective of groups are mixed and drawn at random order to fill the groups and then, an ANOVA test is performed on new groups to calculate a new F-statistic \( F_{NEW} \). The fraction of \( F_{NEW} > F_{ORG} \) is computed which gives \( p \), the probability of obtaining an F-statistic as large as we obtained with observed data if the null hypothesis were true.

The second step should be repeated for all possible permutations of the ordering the data in groups, which in a practical sense is near impossible. Iterations of more than 1000 are suggested to be sufficient to approximate a \( p \)-value. In this study, I use 5000 iterations to randomise the ANOVA test to test the means of the number of cyclones estimated in different decades.
Table 4.2 Number of cyclone encounters for the 1930s and 1950s, by the month of observation. The number of Ship-days are given in the parentheses.

<table>
<thead>
<tr>
<th>Month</th>
<th>November</th>
<th>December</th>
<th>January</th>
<th>February</th>
<th>March</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>1930s</td>
<td>14 (2024)</td>
<td>23 (6465)</td>
<td>11 (6975)</td>
<td>5 (5730)</td>
<td>8 (3374)</td>
<td>61 (24568)</td>
</tr>
<tr>
<td>1950s</td>
<td>2 (946)</td>
<td>13 (16172)</td>
<td>21 (21735)</td>
<td>21 (20580)</td>
<td>11 (12258)</td>
<td>68 (71691)</td>
</tr>
<tr>
<td>Total</td>
<td>16 (2970)</td>
<td>36 (22617)</td>
<td>32 (28710)</td>
<td>26 (26310)</td>
<td>19 (15634)</td>
<td>129 (96259)</td>
</tr>
</tbody>
</table>

4.4 Results

4.4.1 Number of cyclone encounters using Method One

The instances of cyclone encounter detected using Method One (using $NDR_r$) are shown by the months and positions of ships at the time of cyclone encounters for the 1930s and 1950s (Fig. 4.3a). The distribution of cyclonic encounters reflects the spatial spread of the whaling dataset itself. The encounter positions in the 1930s are concentrated in the north of the Weddell Sea whereas in the 1950s positions are much more diffuse and extend further east to the East Antarctic coast (Fig. 4.3b). Interestingly, cyclone encounters are more organised spatially in the 1930s; for example, cyclone encounters in the late spring (November) are more northerly than those the start of summer (December) and so on. At the start of autumn (March) the cyclone encounters are the southernmost in the 1930s decade. By contrast, for the 1950s, encounters in the month of March are the southernmost observed. The encounters in the earlier months are mixed in the 1950s and do not have the strict order seen the 1930s.

The number of encounters recorded for each month in the 1930s and 1950s is shown in Table 4.2. It should be noted that due to the varying number of ships present and length of their operations in each season, not all months cumulatively cover equal amounts of time. To provide the context for the number of cyclonic encounters for each month of the investigated decades, Ship-days (which is a product of number of the ships and number of days observed by those ships for that month) added for the whole decade are shown in parentheses in Table 4.2. The number of encounters reach a maximum in the month of December and January-February for the 1930s and 1950s, respectively closely following the number of Ship-days for each month. The total number of encounters for each decade are similar even though the number of Ship-days is hugely different, with a grand total of 129 encounters for two decades.
4.4 Results

4.4.2 Identification and tracking of cyclones using Method Two

In this section, the probable positions of cyclones are identified from ship observations using linear regression model LM02. The whaling dataset is divided into a number of seasons (a whaling season is defined as the period from 1st September of the current year to 31st August of the following year) for each year in the 1930s and 1950s. Each whaling season is then divided up into week-long time slices to individually detect and track the movement of cyclones in space and time over the season. Two week-long sections from a whaling season in each decade are presented as the case studies.

Case 1: 1935-36 Whaling season

Number of Vessels operating: 4

Season’s time period: 21-Nov-1935 to 25-Mar-1936

The ship tracks for Hektoria, New Sevilla, Sourabaya and Salvestria are overlain by wind barbs for the whaling season 1935-36 are shown in Figure 4.4. The probable direction of cyclones derived from each ship’s wind observations is shown in Figure 4.5a. The Hektoria’s cyclone directions are completely different from the cyclone directions inferred from the other three ships. This is due to the fact that the Hektoria (Fig. 4.4) is located at a large distance (around 3000 km) from three other three ships (not shown in Fig. 4.5b, c). Distances between the other three ships (New Sevilla, Sourabaya and Salvestria) over time are shown in

Fig. 4.3 The geographical locations of cyclone encounters by month for (a) 1930s and (b) 1950s.
A novel approach for historical cyclone detection

Fig. 4.4 The ship tracks for Hektoria, New Sevilla, Sourabaya and Salvestria overlain by wind barbs for the whaling season 1935-36.

Figure 4.5b. Sourabaya and Salvestria remain within 500 km of each other, and all three ships (excluding Hektoria) are always within 1000 km. As seen in Section 4.3.2, the moderate cyclones have 500 km radius; hence, it is probable that these three ships were under the influence of a single weather system much of the time.

Further evidence is found in Figure 4.5c, with the MSLP observations from New Sevilla, Sourabaya and Salvestria (Hektoria did not record MSLP in its log so is not shown) following each other closely most of the time, even though the ships were often hundreds of kilometres apart. This gives support to the suggestion that three ships (New Sevilla, Sourabaya and Salvestria) were under the same weather system throughout the whaling season whereas Hektoria was not.

I have chosen the period when all ships (except Hektoria) are closest, that is from 30-Jan-1936 to 13-Feb-1936, when they are no more than 500 km apart (Fig. 4.5b); MSLP observations and cyclone directions are also in general agreement. A week-long time-frame is then selected from that period to identify cyclones spatio-temporally. Figure 4.6a shows Week 11 (30-Jan-1936 to 06-Feb-1936) from the first logbook entry of the 1935-36 whaling season. There is only one direction-line with more than 50% probability protruding from New Sevilla. This is supported by the direction-lines from Sourabaya. Hence, a cyclone at ~58°S 15°W can be confidently identified (marked by a red © symbol in Fig. 4.6a). There
Fig. 4.5 a) Cyclone directions derived from wind direction observations for the whaling season 1935-36. b) Distance between the ships New Sevilla, Sourabaya and Salvestria as they operated in the Southern Ocean. As Hektoria was more than 2000 km away from rest of the ships, its distance to other ships is not shown. c) MSLP observations from New Sevilla, Sourabaya and Salvestria. The logbooks of Hektoria did not record MSLP observations.
Table 4.3 The number of cyclones, depressions and new cyclones detected in the week-long time frame.

<table>
<thead>
<tr>
<th>Week</th>
<th>Dates</th>
<th>Cyclones Detected</th>
<th>Depressions Detected</th>
<th>Number of new cyclones</th>
</tr>
</thead>
<tbody>
<tr>
<td>11</td>
<td>30-Jan-1936 ~ 06-Feb-1936</td>
<td>1</td>
<td>5</td>
<td>1</td>
</tr>
<tr>
<td>12</td>
<td>06-Feb-1936 ~ 13-Feb-1936</td>
<td>3</td>
<td>2</td>
<td>2</td>
</tr>
</tbody>
</table>

are also many direction-lines from the different ships pointing to a region around 30°W and 20°W at 64°S, but none have probabilities greater than 50%, so it is assumed that they are likely to be pointing at depressions yet to develop into cyclones.

The next time frame, Week 12 (06-Feb-1936 to 13-Feb-1936), is shown in Fig 4.6b. The previously identified cyclone ~ 58°S 15°W appears near ~ 59°S 20°W and this repeated cyclone is identified by ©. The depressions identified in the previous week (direction-lines without × marks) have developed into cyclones. The direction-lines protruding from all three ships point to a cyclone (located near 64°S 30°W meridian line at the bottom of Fig. 4.6b). There is high confidence that a cyclone was present at this location. Three more cyclone positions have been identified and marked by a red © symbol. Table 4.3 shows that five depressions and one new cyclone were identified for Week 11 and two depressions and four cyclones were identified for Week 12.

**Case 2**: 1956-57 Whaling season

Number of Vessels operating: 3

Season’s time period: 09-Dec-1956 to 22-Mar-1957

In the whaling season of 1956-57, the ships Southern Harvester and Southern Venturer made meteorological observations at the sub-daily resolution, unlike Balaena; hence, wind barbs and cyclone directions are more closely packed (Fig. 4.7 and Fig. 4.8a). Also, in this whaling season ships are more spatially spread out than in the 1935-season (Fig. 4.5). As a result, only Balaena and Southern Harvester are within 500 km mark of one another (Fig. 4.8b); the period they are closest is 20-Jan-1957 to 03-Feb-1957. In addition, both ship’s MSLP and cyclone directions are more similar during this period than in any other (Fig. 4.8b and c).

In order to identify individual cyclones, the period when Balaena and Southern Harvester are closest is divided into one week-long time-frame each. Figure 4.9a shows Week 7 (20- Jan-1957 to 27- Jan-1957) from the first logbook entry of 1956-57 whaling season,
Fig. 4.6 Estimated cyclone © and ship positions for weeks: (a) 30-Jan-1936 to 06-Feb-1936 and (b) 06-Feb-1936 to 13-Feb-1936. The colour of ‘×’ mark denotes the parent ship in the legend and red arrows show approximate track of the cyclone identified by ©R.
direction-lines from both Balaena and Southern Harvester pointing to a cyclone located at ∼68°S 100°W and another cyclone further south as pointed out by Southern Harvester’s direction-lines. A cyclone identified in the previous week (not shown) is denoted by © and is located at approximately ∼68°S 75°W. As the data are sub-daily, they offer an opportunity to track the movements of cyclones in greater detail than before. The direction-lines from Southern Venturer imply the presence of a cyclone at ∼61°S 60°W and direction-lines track the cyclone as it moves eastward to reach ∼60°S 34°W as Week 7 progresses (red arrows in Fig. 4.9a show approximate tracks of the cyclones identified). A typical extra-tropical cyclone moves at an average about 30 km/h (Kossin 2018, Dorst 2019) and the distances between subsequent positions are within this range. Another previously identified cyclone transits from ∼62°S 45°W to ∼66°S 34°W, a distance of 375 km at 31 km/h.

Data from Week 8 (Fig. 4.9b; 27-Jan-1957 to 03-Feb-1957) show the previously tracked cyclone (©) last seen at ∼60°S 34°W moving east, just south of South Georgia and then heading south and west after crossing the prime meridian. One new and two previously identified cyclones are also identified west of the Antarctic Peninsula. A count of the total number of cyclones, new cyclones and depressions detected in each weekly time frame is given in Table 4.4.
4.4 Results

Fig. 4.8 a) Cyclone directions derived from wind direction observations for the whaling season 1956-57. b) Distance between the ships Southern Venturer, Balaena and Southern Harvester as they operated in the Southern Ocean. c) MSLP observations for the same ships.

Table 4.4 The number of cyclones, depressions and new cyclones detected in the week-long time frames from the 1956-1957 whaling season.

<table>
<thead>
<tr>
<th>Week</th>
<th>Dates</th>
<th>Cyclones Detected</th>
<th>Depressions Detected</th>
<th>Number of new cyclones</th>
</tr>
</thead>
<tbody>
<tr>
<td>7</td>
<td>20-Jan-1957 – 27-Jan-1957</td>
<td>4</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>8</td>
<td>27-Jan-1957 – 03-Feb-1957</td>
<td>4</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>
Fig. 4.9 Estimated new cyclones ©, previously identified cyclones ©R and ship positions for weeks: (a) 20-Jan-1957 to 27-Jan-1957 and (b) 27-Jan-1957 to 03-Feb-1957 of the 1956-57 whaling season. The colour of ‘×’ mark denotes the parent ship in the legend and red arrows show approximate tracks of the cyclones identified. The Antarctic Peninsula and South Georgia are denoted by AP and SG respectively.
4.4 Results

4.4.3 Semi-supervised cyclone detection algorithm

The manual analysis of the spatio-temporal ships’ logbook data described above can be transformed into a semi-supervised identification and tracking algorithm, which can ingest a large amount of historical whaling data and count new and persistent cyclones from one time-frame to the next. The algorithm is designed to take data from each whaling season and divide it into week-long segments. The identification and tracking criteria for the proposed semi-supervised algorithm are as follows:

1. If two or more direction-lines, one of which should be more than 50% probability (e.g. ‘×’ marked), are pointing in the same or similar directions (within 30 degrees) and their end points are less than 100 km apart, then they are considered to be pointing at a single cyclonic system (©).

2. If temporally subsequent direction-lines are within 200 km for 4-hourly sub-daily and within 700 km for daily observations, the system is considered to be in motion.

3. If a cyclone identified in the previous time-frame reappears in the current time-frame within 500 km of the last position, the cyclone is assumed to be a repeat or a persistent cyclone (©R).

4. The number of cyclones and depressions are counted along with new cyclones identified for each week.

5. The process is continued until the end of the data time-series or until only one ship is left without any more than 50% probability ‘×’ marked direction-lines.

This algorithm was used to process observations from all whaling seasons in the Christian Salvesen whaling dataset, including the two case studies presented above. The estimated total number of cyclones identified for the 1930s and 1950s, grouped by months, is shown in Table 4.5. As mentioned earlier in Section 4.1, due to an uneven number of ships and observation days in a month for each season, Ship-days are shown in parentheses to place the number of cyclones in perspective. A sample composite of whaling season 1935-36 is shown in Suppl. Fig. A.5.
Table 4.5 Total number of cyclones inferred from ships’ log MSLP data for whaling seasons in the 1930s and 1950s. Ship-days are shown in parentheses.

<table>
<thead>
<tr>
<th></th>
<th>November</th>
<th>December</th>
<th>January</th>
<th>February</th>
<th>March</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>1930s</td>
<td>21 (2024)</td>
<td>39 (6465)</td>
<td>33 (6975)</td>
<td>27 (5730)</td>
<td>18 (3374)</td>
<td>138 (24568)</td>
</tr>
<tr>
<td>1950s</td>
<td>42 (946 )</td>
<td>65 (16172)</td>
<td>66 (21735)</td>
<td>73 (20580)</td>
<td>69 (12258)</td>
<td>315 (71691)</td>
</tr>
<tr>
<td>Total</td>
<td>63 (2970)</td>
<td>104 (22617)</td>
<td>99 (28710)</td>
<td>100 (26310)</td>
<td>87 (15634)</td>
<td>453 (96259)</td>
</tr>
</tbody>
</table>
4.4 Results

4.4.4 Comparison of cyclonic frequency

The number of cyclones detected for each decade and month, as shown in Table 4.5 even with the number of Ship-days, cannot be compared directly with each other. The frequency of observations, number of observations, number of ships, distance/area covered, duration of whaling season and so on, are different for each vessel and season, which introduces a non-standardised view of the number of cyclones detected in each decade. To generate like-for-like comparison, a common time-period and an area with sufficient observations for each whaling season are required. The number of cyclones detected within such areas could then be used to compare cyclone frequency with that of the modern period as well. To do so, first of all, it is necessary to identify a month with most observations common to both decades in the whaling dataset.

The monthly histogram of whaling data (not shown) reveals that the month of January is the most common month for observations in the dataset. The spatial binning of January observations revealed an area south-east of South Georgia to be highly populated. Hence, an area of radius 500 km centred at 60°S 26°W was selected (Solid circle, Fig. 4.10a) to compare the number of cyclones in the month of January for each whaling season in the whaling dataset and the modern period. The radius of 500 km was selected as a reasonable control area where an adequate number of cyclones could be detected both from the whaling dataset and the modern cyclone dataset. As seen earlier (Section 4.2), observations as far as 500 km distant can be used to detect the presence of cyclones, and observations in an extended circle of 1000 km radius centred at 60°S 26°W (Dashed circle, Fig. 4.10a) were also included in the detection of cyclones in the solid circle.

A total of 10 whaling seasons were found to contain at least 21 days of observations in January within the dashed circle (five seasons in each decade) following the WMO guidelines for calculating a monthly average of a meteorological station (WMO Report No. 100; WMO 2011). Interestingly, the number of days of January covered by the whaling seasons in the 1930s is slightly greater than by the whaling seasons in the 1950s within the dashed circle (Fig. 4.10a). This produces a similar number of Ship-days in the 1930s to the 1950s in January, due to the balancing off of a lesser number of ships in the former and a lesser number of observing days in the latter. Thus providing near equal number of Ship-days for objective comparison of the number of cyclones in both decades.

The number of cyclones detected in the 1930s and 1950s that passed through the solid circle (Fig. 4.10a) is shown in Table 4.6. The highest number of cyclones is found in the
Fig. 4.10 a) The area with the number of observations for the month of January for all whaling seasons is enclosed by a solid circle. The number of days of January observed within the area encircled by a dashed line for each whaling season is shown in parentheses. b) The tracks of all cyclones passed through the circle in the month of January for the 1999-2008 period are shown, on average 8.6 cyclones per year passed through the circle area in the month of January during the same period. SG denotes South Georgia.
Table 4.6 Number of cyclones detected passing through the solid circle depicted in Figure 4.10a during the month of January in each available whaling seasons in the 1930s and 1950s in the whaling dataset with the five-year total and average for each decade.

<table>
<thead>
<tr>
<th>Whaling Seasons 1930s</th>
<th>Number of Cyclones</th>
<th>Whaling Seasons 1950s</th>
<th>Number of Cyclones</th>
</tr>
</thead>
<tbody>
<tr>
<td>1932-1933</td>
<td>4</td>
<td>1950-1951</td>
<td>6</td>
</tr>
<tr>
<td>1933-1934</td>
<td>5</td>
<td>1951-1952</td>
<td>2</td>
</tr>
<tr>
<td>1934-1935</td>
<td>2</td>
<td>1953-1954</td>
<td>5</td>
</tr>
<tr>
<td>1935-1936</td>
<td>5</td>
<td>1956-1957</td>
<td>4</td>
</tr>
<tr>
<td>1939-1940</td>
<td>3</td>
<td>1958-1959</td>
<td>4</td>
</tr>
<tr>
<td>Five-year Total (Average)</td>
<td>19 (3.8)</td>
<td>Five-year Total (Average)</td>
<td>21 (4.2)</td>
</tr>
</tbody>
</table>

Table 4.7 Number of cyclones detected passing through solid circle depicted in Figure 4.10b during 1999-2008 period with average and standard deviations for the same period.

<table>
<thead>
<tr>
<th></th>
<th>1999</th>
<th>2000</th>
<th>2001</th>
<th>2002</th>
<th>2003</th>
<th>2004</th>
<th>2005</th>
<th>2006</th>
<th>2007</th>
<th>2008</th>
<th>Average</th>
<th>Std Dev</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>6</td>
<td>18</td>
<td>5</td>
<td>11</td>
<td>6</td>
<td>12</td>
<td>7</td>
<td>9</td>
<td>4</td>
<td>8</td>
<td>8.60</td>
<td>4.17</td>
</tr>
</tbody>
</table>

1935-1936 whaling season in the 1930s and in 1950-1951 in the 1950s. The total number of cyclones in five years of each decade are similar, 19 for 1930s and 21 for 1950s. Figure 4.10b shows the tracks of all the cyclones from the M02 cyclone data (Neu et al. 2012) that passed through the solid circle (shown in Fig. 4.10a) in the month of January during 1999-2008. A total of 86 cyclones were detected in this 10-year period, with an average of 8.6 cyclones per year and the standard deviation of 4.17 cyclones (Table 4.7). The number of cyclones shows large variability, from 18 in 2000 to 4 in 2007.

To assess the differences in the means of these three groups, I employed a randomised ANOVA test (Section 4.3.3). After 5000 iterations of the randomised ANOVA test, the fraction of tests producing F-statistics at least as extreme as the original group is 0.014. This p-value or the probability suggests that there is more than a 95% (98.6%; (1 - p-value) × 100%) probability that the null hypothesis (no difference between the samples) can be rejected. In other words, the differences in the means of the number of cyclones detected in January in the 1930s, 1950s and modern period are statistically significant at 5% significance level.
4.5 Discussion and Summary

4.5.1 Method One

By following the first method in this study, using $NDR_r$ to count the number of cyclone encounters (Section 4.3.1), we are clearly restricted to identifying cyclones that crossed paths with or moved close to a ship. Such occurrences can be assumed to be relatively infrequent as, unsurprisingly, the whaling ships avoided being near cyclones. The Admiralty Navigation Manual (1938) clearly instructed ships to move away as far as possible to the right of the path of a cyclone in the Southern Hemisphere (Reid’s Law of Storms). Hence, the number of encounters detected represents only a small subset of possible cyclones. Nevertheless, this method provides an approximate lower estimate of the number and locations of cyclones and can be usefully applied to large numbers of historical ships’ logbooks which only record MSLP observations. The identified encounter locations (Fig. 4.3) lie between previously reported regions of cyclogenesis and cyclolysis. Near 60°W longitude there are three genesis regions: in the lee of the Andes over South America; near 47° and 32°S, and over the Antarctic Peninsula; and in embayments of the Antarctic coast, specifically in the Weddell Sea as identified by Jones and Simmonds (1993) and Lim and Simmonds (2002). Hence, most of the cyclone encounters are found in the areas where cyclones transit from cyclogenesis and cyclolysis regions. As expected, these areas also correspond to the densest regions of identified cyclone tracks as shown in previous studies (e.g. Wei and Qin 2016).

In addition, the number of cyclone encounters detected is sensitive to the $NDR_r$ threshold used. The threshold proposed by Lim and Simmonds (2002), which detects explosive cyclones with a deepening rate of more than 24 hPa/day (1 hPa/hr) at 60° latitude. However, a deepening rate of 1 hPa/hr or higher is very unlikely to be detected in the ships’ logbooks as strong baroclinicity and weak static stability are associated with explosive cyclone development and when those conditions start to occur ships move away for safety. To reach a reasonable threshold value, therefore, $NDR_r$ values were calculated for typical cases in this whaling dataset; that is, a moderate cyclone with core pressure ~980 hPa briefly crossing the path of a ship (located in the 50°-65°S latitude band) that has an MSLP observation equal to the climatological mean pressure for that location. The average $NDR_r$ value was found to be 0.5 (using Eq. 4.1). By using a higher threshold of 0.8 the method detected too few cyclonic encounters in the 1950s as compared to 1930s. On the other hand, the possibility of encountering a cyclone in the 1950s should be higher due to the greater number of observations in
4.5 Discussion and Summary

the 1950s than in the 1930s, assuming similar conditions existed throughout the two decadal periods.

After testing with different values, a threshold of 0.6 was found to be suitable and was used to count the number of cyclone encounters. Interestingly, the positions of encounters for each month in the 1930s migrate southwards as the whaling season progresses from spring to autumn (Fig. 4.3). These findings are supported by many previous studies which identified a shift in the cyclone tracks southward towards the autumn (Jones and Simmonds 1993; Wei and Qin 2016). The number of encounters detected changes through the months, but on average, summer months are higher than in spring or autumn months. Interestingly, Simmonds and Keay (2000), Grieger et al. (2018) and others also found an increasing trend in the number of cyclones in the summer when analysing data over 1958–97 in the 60°S latitude band. Even so, more data are needed to attribute the greater number of encounters in the summer found in this study to the positive trend detected by previous studies as this could be possibly due to more summer observations in the whaling dataset.

Another notable observation is that the total number of cyclone encounters is not much higher in the 1950s than the 1930s (Table 4.2), even though the 1950s contained high-resolution sub-daily observations compared to daily observations in the 1930s. This is suggestive that the results extracted from Method One do not become biased when high-frequency sub-daily observations are used as compared to daily observations. This feature is reassuring and useful in future work where it can usefully be employed as numerous historical ship logbooks only contain daily-resolution observations to count cyclone encounters using this method.

4.5.2 Method Two

Method Two builds on the meteorological influence of synoptic systems such as cyclones on the weather observed in the surrounding areas. The design matrix used here is collected from 1999-2008, a period with an abundance of in-situ observations including those from research vessels, commercial vessels, drifting buoys and satellite-based observations, offering the most robust and detailed account of the influence of cyclones on weather observations in the Southern Ocean (Freeman et al. 2017). Prior to building the regression models, independent variables (observations) were tested for multi-collinearity and were found to be independent of each other. The subsequent models built were accessed by their statistical properties. As the AIC is a measure of relative goodness-of-fit that penalises for over-fitting it is suitable for
the inter-comparison of models derived from the same data. The AIC, F-values and PVE were ranked in that order to select a model. The model built around M02 tracks (LM02) was found to be the most satisfactory (Table 4.2) in comparison to models built around other datasets (M18, IBTrACS).

This is surprising, given that the model built on IBTrACS tracks, which are derived from post-season analysis of meteorological observations and satellite images, has a poor fit relative to the similar meteorological observations obtained from ICOADS (Table 4.2). A number of studies (e.g. Hodges et al. 2017, Schreck et al. 2014) have found that IBTrACS contains too few cyclones originating or travelling south of 20°S. We can confirm this as M02, M18, IBTrACS have 300, 400 and 10 cyclones, respectively, for the regions south of 30°S between 1999-2008. The probable reason, as pointed out by Hodges et al. (2017), is that as the Southern Hemisphere is sparsely populated especially in regions such as the South Pacific and South Atlantic basin, many cyclones were not detected by the National Weather Centres unless they were likely to make landfall. Hence, too few cyclones south of 30°S are present in the IBTrACS, and the model build on IBTrACS is therefore not suitable for this study.

On the other hand, the high PVE obtained from both LM02 and LM18 (Table 4.2) is a testament to the in situ observations assimilation process in the generation of the ERA-Interim dataset. Unsurprisingly, LM02 (low AIC and F-value) performs better than LM18 as the former uses surface MSLP observations and the latter uses 850hPa geopotential heights to detect track cyclones. As Hoskins and Hodges (2005) and Lakkis et al. (2019) have found, upper tropospheric cyclones do not usually match with surface level cyclones; hence, cyclones detected from the upper levels of the atmosphere (LM18) may not reflect meteorological conditions observed on the surface.

The sign and magnitude of the regression coefficients (of LM02) are of interest, and the physical consistency of the relationships can be checked as follows. The MSLP would decrease moving towards a cyclone centre from the ship location, the distance coefficient ($\beta_2$) is expected to be negative and, in addition, negative $\beta_3$ and $\beta_4$ coefficients are expected as strong wind signals a deeper cyclone. The ship MSLP ($\beta_1$) coefficient suggests, ceteris paribus, that a 1 hPa change in ship MSLP would lead to 0.946 hPa change in the cyclone core pressure; in other words, the lower the ship MSLP observations, the deeper the cyclone in the vicinity and vice-versa, which seems reasonable. Additional evidence can be presented from Pepler et al. (2018), who found the pressure gradient from the centre of a cyclone to the edge of a weather system to be $\sim$1 hPa/100 km. In our case, the regression coefficient $\beta_2$ would
lead to a similar pressure gradient for every 100km, further supporting the methodology of this study.

Method Two not only detects the presence of cyclones but also tracks them in space and time. The tracking of cyclones is dependent on using a reasonable assumption for the translational speed of cyclones. A recent study (Kossin 2018) found a decreasing trend in the translational speed of cyclones since the 1950s; however, many subsequent studies have shown the trend to be an artefact of the changing observational density of cyclones (Lanzante 2019, Moon et al. 2019). While not considering the trend, the cyclones are known to change speeds when travelling across latitude bands (Lanzante 2019). The higher the latitude, the faster the cyclone travels. The mean speed of an extra-tropical cyclone at ∼60ºS is 30 km/hr or 15.1 kts (Kossin 2018, Dorst 2019). Hence, on average, an extra-tropical cyclone would travel more than 600 km in a day.

The weather conditions and their evolution in the case studies presented in Section 4.4.2 have fine-tuned many smaller empirical details in the model leading to a semi-supervised cyclone detection and tracking algorithm (Section 4.4.3). Even the most sophisticated cyclone identification and tracking algorithms used at the National Weather Centres across the world require some manual corrections and adjustments from experienced meteorologists. This is, in part, because the many identifying features of cyclones/hurricanes and their evolution are often hidden or not present at all in the routinely used data sources (e.g. satellite images).

Likewise, the semi-supervised algorithm constructed here requires manual intervention as a number of factors that are assumed in order to construct this algorithm may not hold for many possible scenarios contained in this historical whaling dataset and other historical meteorological datasets at large. For example, a central assumption is that any synoptic system whose core pressure is less than 980 hPa is a cyclone and between 980-1000 hPa is a depression. The Southern Ocean contains a large number of polar lows at any time which are frequently deeper than 980 hPa, but the pressure gradient structure and sustained winds might not be present; hence, from an operational point of view, they would not be termed as cyclones.

In addition, frequent gale force winds experienced by ships in the Southern Ocean could be mis-identified as cyclonic winds by both the linear model developed in this study and also from in situ observations; however, due to changing seasons and, in turn, changing pressure gradients between the Antarctic continent and surrounding open waters and the high elevation of the Antarctic ice sheets, katabatic gravity winds are observed in the Southern Ocean especially near the coast.
Nonetheless, this algorithm is better than many previous statistical models (e.g. CLIPER (Climatology and Persistence); Neumann (1972), SHIFOR (Statistical Hurricane Intensity Forecast); Jarvinen and Neumann (1979), SHIP (Statistical Hurricane Intensity Prediction); DeMaria et al. (2005)). For example, the SHIP scheme uses a set of 15 climatological, persistence and synoptic predictors derived from the initial conditions of the NCEP Medium Range Forecast model to build linear multiple regression models and predicts the intensity of hurricanes in terms of maximum sustained winds. By contrast, our algorithm attempts to estimate both position and intensity of cyclones from the location of observations using in situ meteorological observations. The models built on in situ observations are likely to be more accurate than those from the forecast model (e.g. SHIP); including a diverse range of predictors adds to the robustness of the model.

4.5.3 Limitation of methods

The number of cyclones detected by the algorithm needs to be carefully compared within the whaling dataset and with modern cyclone dataset in view of their limitations. Like many studies using historical data, the methods developed in this study have certain limitations. In general, the accuracy of predictions made by either of the methods is limited by the accuracy of the whaling dataset itself. For example, the MSLP observations were carried out using mercurial marine barometers fitted with correctional scales, and some were recorded in inches of Mercury while others were in millibars (Teleti et al. 2019 and Chapter 2). The meteorological dataset published by the same study standardised the observations by converting all observations to hPa; hence, invariably errors are unavoidable due to, e.g. precision of measurement scale, rounding-off errors. Additional uncertainty comes from the positions recorded by the ships themselves. The latitude and longitude are estimated to the nearest 0.1° (leading to 6 nm or 11 km margin of error) in the historical period due to imprecise instruments (Landsea and Franklin 2013).

The algorithm and the model rely on statistical relationships similar to the hurricane wind-pressure equations derived by Brown et al. (2006), but based on only a small number of predictors (four) which may not be able to capture changes accurately for all scenarios. Unfortunately, historical datasets do not contain many meteorological variables to draw upon, and this limits the accuracy of the intensity and position of cyclones detected. Naturally, as the relationships derived in this study are empirical, they only hold for regions (south of 30ºS) and seasons (November-December-January-February-March) from which they are derived. A
number of studies, however, have successfully utilized ship observations to produce statistical models that detect the presence of cyclones not in the immediate vicinity of a given vessel (e.g. Chang and Guo 2007) and also predict future track positions (e.g. DeMaria et al. 2005).

4.5.4 Comparison: Historical vs Modern

Lastly, the distribution of the number of cyclones in space and time estimated by the semi-supervised algorithm reveals an interesting pattern (Table 4.5). The number of cyclones in the 1930s increases monotonically from November to December, peaks in December and then decreases from January to March. It should be noted that the highest number of observations are present in December, but it is uncertain whether the peak number of cyclones in December is naturally occurring or if it is an artefact of higher Ship-days. While the pattern in the 1950s is less clear, the number of cyclones increases from November to December then decreases until February before attaining a peak in March. The number of cyclones detected in the 1930s and 1950s cannot be compared even with the knowledge of Ship-days because clustering of ships and distance travelled through a season cannot be satisfactorily reconciled for all seasons recorded in the whaling dataset.

In addition, such standardisation would be meaningless for the cyclones detected in the modern period, making historical-modern cyclonic frequency comparison impossible. The comparison presented in Section 4.4.4, provides a level-field for making such comparisons. Overall, the number of cyclones in the 1950s is slightly more than in the 1930s, but this could be due to temporally denser and spatially more spread-out observations in the former than in the latter. However, the average number of cyclones in both decades is lower than the average calculated over the modern period (1999-2008). It is confirmed by the randomised ANOVA test performed on the number of cyclones in the historical period (1930s and 1950s) and modern period, and the latter is found to be significantly higher than the former.

In addition, an indirect comparison of cyclonic frequency can be made by comparing extrapolated historical cyclone numbers with modern ones. For example, the average number of cyclones in the Southern Ocean per summer season (DJF) was found to be $\sim$190 from 1979-2013 (Wei and Qin 2016). Comparing these modern data, first, with the 1930s, a total of 99 cyclones in the DJF season were derived from observations from six whaling seasons in the 1930s, which is equivalent to $\sim$17 cyclones per DJF season. However, the spatial spread of the observations in a typical whaling season was no more than 50 degrees of longitude; hence, to offset the restricted coverage of observations multiplication by a factor 6 or 7 is
needed to reproduce circumpolar coverage. The adjusted circumpolar number of cyclones per DJF season is 119, and the similarly adjusted number of cyclones per DJF season is 143 derived from all ten seasons in the 1950s. These estimated numbers of cyclones are less than the average found by Wei and Qin (2016) but provide a good approximation of cyclones over the largely unobserved region and time-period.

4.6 Conclusions

This is the first ever work to use whaling data to identify and track historical cyclones in the Southern Ocean. Also, the first to estimate and compare the historical cyclonic frequency with modern cyclonic frequency in the Southern Ocean to the author’s knowledge. Two methods have been proposed to capture different aspects of the influence of cyclones on the local weather recorded through meteorological observations. The first method used the relative deepening rate of pressure observations to detect the approach of cyclones towards whaling ships or vice-versa. The number of cyclonic encounters detected is sensitive to the deepening rate threshold applied, and this threshold has been framed taking the climatological conditions of the region under investigation into account. The monthly number of encounters in the 1930s and 1950s decades could not be compared directly with each other, due to a number of factors such as different Ship-days recorded and distance/area covered for that month.

The estimation of cyclones using a linear multi-regression model developed in this study identifies and tracks individual cyclones in space and time. The probability of identifying a cyclone at a given distance from pressure and wind observations on-board a whaling ship is enhanced when multiple ships are close by. On average in each week, two new cyclones and one previously identified cyclone are identified from the whaling dataset. However, the identification and tracking algorithm developed and used in this study is sensitive to the definition of cyclones and the translational speed assumed; this is, in turn, dependent on the model and the observations used to calibrate the model. Overall, cyclonic frequency over a control area (in the Weddell Sea) appears to have increased significantly from the start of whaling dataset (1930s) to the modern period (2000s); however, more data from the historical period would be required to strengthen this claim.

A number of innovations could be included in future studies to enhance the accuracy and robustness of the algorithm developed here. Additional predictors could be incorporated into the model, given that several studies (e.g. SHIP) have demonstrated a clear link between
sea-surface temperatures (SSTs) in the path of a cyclone and the observed intensity of the cyclone. This could be achieved by combining different datasets (e.g. ICOADS) with whaling datasets to estimate SST fields in the path of the cyclone, and even the climatological SSTs of the region can improve model skill. As cyclones are strongly non-linear processes in terms of pressure-gradients, wind speeds employing a neural network to model cyclone-meteorological observations can address non-linearity issues faced by the linear models. The neural network based cyclone intensity and tracks forecast have been shown to out-perform linear regression when using climatological and persistence variables (Baik and Hwang 1998). This is because more flow-dependent information is extracted from the data in the standard back-propagation neural network as compared to linear regression. Hence, further study should be conducted using neural networks to model cyclone intensity based on SST data in addition to the variables used in this study.
Chapter 5

Offline data assimilation of historical whaling ships’ observations

In this Chapter, I demonstrate how to assimilate historical whaling meteorological data into modern reanalyses to improve the representation of MSLP fields in the historical period.

5.1 Introduction

The global climate change is manifested in increasing global temperatures (Intergovernmental Panel on Climate Change 2014), a marked decrease in Arctic and, more recently, Antarctic sea ice (Simmonds 2015; Parkinson 2019), combined with the thinning of mountain glaciers and ice caps around the world (Zemp et al. 2019). Meteorological observations provide a fundamental basis to understand the changing climate system. The longer the observational records, the more fully we can assess the past and predicted-future changes in the climate. Many European nations started the near-modern land-based network of meteorological stations in their own countries and colonial administrative regions and, by the late 19th century, coverage expanded to include all inhabited continents and many parts of the oceans. Today, we can estimate global climate change from many important climate records, with some such as temperature (HadCRUT4; Morice et al. 2012) going back to the mid-to-late 19th century. Naturally, observations from such a long period are compromised by known and unknown measurement bias resulting from, for example, changing techniques (observing at a particular time of the day, recording only a specific statistic, etc.), instruments, changes in the observational spatio-temporal densities and so on. The disparity in the observational
density of stations has led to gaps in knowledge of climate and its variability at different times and in different regions.

To resolve these shortcomings, numerous efforts have been directed at developing climate reanalysis models which are dynamically constrained by known observations, to expand the understanding of the evolving global climate. A reanalysis system consists broadly of a combination of numerical weather prediction (NWP) model and a data assimilation (DA) scheme to produce best estimates of past states of the climate (e.g. temperature, air pressure, wind) (Fujiwara et al. 2017). The NWP models propagate information from the previous states of the climate to the next while maintaining physical and dynamical laws, starting with model initial conditions. The DA system then attempts to combine optimally past observations and model states, both of which are incomplete and/or inaccurate. These reanalyses produce forecasts for the state of Earth’s climate, with reduced errors and associated uncertainties compared to both observations and states of the model (Carrassi et al. 2018). Hence, an atmospheric reanalysis produces the best estimate of the state of Earth’s atmosphere by assimilating observational data into NWP model forecasts, including the observations and variables from a first-guess background model state.

The reanalyses have been categorised into two groups by the type of assimilated observations: NWP-like reanalyses and climate reanalyses (Zhou et al. 2018). The former continues to assimilate all available data (e.g. surface stations, ships, buoys, radiosondes, aeroplanes and satellites), while the latter assimilates only selected observations (e.g. surface-level observations). The NWP-like reanalyses include ERA-Interim, JRA-55, MERRA, MERRA-2. Climate reanalyses include 20CRv2c, ERA-20C, CERA-20C. The NWP-like reanalyses are designed to take full advantage of modern data sources such as short-to-medium term weather forecasts, whereas climate reanalyses are suitable for ingesting fewer, but longer, historical observations. For more detailed inter-comparison of reanalyses approaches, refer to the S-RIP project (http://s-rip.ees.hokudai.ac.jp) (Fujiwara et al. 2017). Here, I briefly highlight important features and differences in the available reanalyses relevant to this study of historical climate. I focus on three climate reanalyses (ERA-20C, CERA-20C and 20CRv2c) that span whole the 20th Century.

The two reanalyses from ECMWF, ERA-20C and CERA-20C reanalyses, run coupled models combining atmosphere–land–ocean–waves (IFS version Cy38r1) and atmosphere–ocean–land–waves–sea ice (IFS version Cy41r2) components of the Earth system, respectively. In addition, both ERA-20C and CERA-20C assimilate only selected conventional observations of surface pressure and marine wind; the greenhouse gases forcing are the same as CMIP5
(Taylor et al. 2011) and atmospheric lower boundary conditions for sea-surface temperature (SST) are taken from HadISST2.1.0.0 (Titchner and Rayner 2014). A particular difference is that ERA-20C assimilates sea-ice conditions from HadISST2.1.0.0, whereas CERA-20C uses a coupled model to simulate sea-ice fields and assimilates upper-air observations (Laloyaux et al. 2018). The ERA-20C uses a deterministic four-dimensional variational (4D-Var) system based on a 10 member model ensemble producing 3-hourly state of the atmosphere and spatio-temporally varying background errors (Poli et al. 2013). CERA-20C simultaneously assimilates atmospheric and ocean observations in a coupled Earth system model based on a 4D-Var method (using a longer analysis window to include more past and future observations) with a common 24 hr assimilation window shared by the atmospheric and ocean components (Laloyaux et al. 2016). It produces a 6-hourly 10 member ensemble of state of the atmosphere and spatio-temporally varying background errors.

Of the ERA-20C and CERA-20C reanalyses, CERA-20C is chosen for this study due to its superior performance over ERA-20C for the Antarctic region (Schneider and Fogt 2018). The remaining reanalysis, 20CRv2c, is the only one of the three that uses only surface and sea-level pressure observations, unlike CERA-20C, which also uses marine wind observations. The 20CRv2c uses SSTs from HadISST1.1 (Rayner 2003) and sea-ice fields from COBE SST (Ishii et al. 2005) as boundary conditions. The greenhouse gas forcings are derived from monthly 15° gridded estimates of CO2 from WMO observations. The 20CRv2c uses an ensemble Kalman filter technique to assimilate observations to produce 56 equally probable members of climate state ensemble.

At this point, it is pertinent to note differences between Variational (e.g. 4D-Var) and Stochastic (e.g. ensemble Kalman Filter) assimilation techniques used in these reanalyses. The 4D-Var method minimises a cost function that penalises differences between observations and model background state (Courtier et al. 1994; Talagrand 2010). It iteratively optimises the fit between assimilated observations and the time-varying forecast trajectory within the full assimilation window (Park and Županski 2003). The convergence of the assimilation algorithm depends on finding the global minimum for the cost function. Of all of the reanalyses that use 4D-Var technique, only a simplified incremental 4D-Var (in two nested loops and with reduced control space and physics) is implemented because a full implementation of 4D-Var is computationally intensive, requiring constant updating of adjoint models; this remains impractical at present (Fujiwara et al. 2017).
The method is implemented by deriving adjoint models; this is carried out by differ- entiation of the direct model and transposition of a linear tangent model (Carrassi et al. 2018).

By contrast, a more probabilistic approach is taken in ensemble Kalman Filter (EnKF) assimilation techniques (Evensen 1994). An ensemble of forecasts is used to define a probability distribution function of the background states (the priory) which is then combined with observations and associated uncertainties to derive a probability distribution of analysis states (the posterior). The optimal analysis state is determined by applying a Monte Carlo approximation to the Kalman filter (Kalman 1960) to this posterior distribution. The algorithm updates every ensemble member to a different set of observations perturbed by random noise, making each ensemble equally likely and independent (Brönnimann 2015). The mean state of the atmosphere and attendant uncertainties are derived from the ensemble mean and uncertainties.

The implementation of EnKF is less intensive because unlike the 4D-Var technique, it does not require perturbation forecasts and adjoint models. Even so, both methods are based on updating initial conditions and covariance matrices which, although powerful, remain computationally intensive and are usually implemented using national or international supercomputing facilities. For example, as Laloyaux et al. (2018) note, the full-mode running of reanalyses CERA-20C requires 20,000 cores solving 500,000 variational problems at a pace of one every 30s, the generation of CERA-20C took seven months using up to 5% of the total computing resources at ECMWF, Reading, UK.

Hence, the motivation for this study is to investigate computationally inexpensive methods to assimilate new historical meteorological observations into reanalyses. My approach uses a third class of techniques known as nudging data assimilation (NDA). This is a process where the model state is moved towards observations (reducing the distance between model state and observations), without running the full model (Blum et al. 2009). Hence, NDA is also referred to as ‘offline’ assimilation as it is applied to the analysis state without updating initial conditions (Brönnimann 2015). The NDA is performed by adding a forcing (nudging) term to the model dynamics that relaxes the model state towards the observations over the assimilation window (Blum et al. 2009). The nudging term is defined as a product of the difference between observations and model state and a nudging coefficient. Nudging coefficients can be determined by both heuristic and physical considerations, and these techniques have been successfully applied to single, deterministic forecasts by Auroux and Blum (2008), Widmann et al. (2010) and Pazó et al. (2016).
In this Chapter, I apply NDA ‘offline’ assimilation methods to multi-ensemble reanalyses (20CRv2c and CERA-20C) to nudge the model state towards the new historical meteorological observations extracted from logbooks of whaling ships in the Southern Ocean (Teleti et al. 2019 and Chapter 2). Through this process, I attempt to estimate the inherent uncertainty in the reanalyses, their sensitivity and the uncertainties to the data assimilation. The chapter is structured as follows: Section 5.2 briefly describes the different reanalyses datasets used in this study (e.g., 20CRv2c, CERA-20C) and the Christian Salvesen whaling dataset with a description of offline assimilation method. Section 5.3 shows the results of comparisons between reanalyses, effects of offline assimilation on both reanalyses and their uncertainties and the relative sensitivity of the reanalyses to new assimilated observations. In Section 5.4, the results are discussed in a broader context linked to the quality of climate reanalyses in the first-half of 20th Century and the effectiveness of such offline methods; and finally, in Section 5.5, I conclude by offering suggestions for the expansion of the offline methods to other reanalyses and further improvements to the offline assimilation methods.

5.2 Data Sources and Methods

The whaling dataset (Teleti et al. 2019) produced in Chapter 2 is the main source of historical meteorological observations to be assimilated into MSLP fields of two multi-member reanalyses (e.g. 20CRv2c and CERA-20C). As introduced above, 20CRv2c and CERA-20C reanalyses datasets with 6-hourly 56- and 10-member ensembles are used.

5.2.1 Datasets

Christian Salvesen Whaling dataset

The Christian Salvesen whaling dataset (Teleti et al. 2019, Chapter 2 in this thesis) contains meteorological observations extracted from logbooks of whaling ships operating in the Southern Ocean during the 1930s and 1950s. It is described in detail in Chapter 2. In this Chapter, I primarily assimilate MSLP observations into reanalyses because, as noted in previous studies, the quality of reanalyses is dependent on the quality of input observations (Laloyaux et al. 2018). The MSLP observations in the Christian Salvesen whaling dataset were found to be very uniform with a high degree of internal consistency, requiring almost no correction (Teleti et al. 2019).
Twentieth-Century Reanalysis, Version 2c (20CRv2c)

The Twentieth-Century Reanalysis, Version 2 (20CRv2; Compo et al. 2011), generated by the Physical Sciences Division (PSD) of the Earth System Research Laboratory (ESRL) of the National Oceanic and Atmospheric Administration (NOAA) and the Cooperative Institute for Research in Environmental Sciences (CIRES) at the University of Colorado, is a historical comprehensive global atmospheric circulation gridded dataset (that is, a product of NWP model) that covers the time period from 1871–2012. An Ensemble Kalman Filter data assimilation method was used, and the National Center for Environmental Prediction’s (NCEP) global numerical weather prediction land/atmosphere model provided background fields as ‘first guess’. This NCEP model has a Gaussian grid with horizontal resolution of $\sim 2^\circ$ latitude $\times 1.875^\circ$ longitude (model resolution T62 with 94$\times$192 grid points), and has 28 model atmosphere levels with a top at 10 hPa.

An upgraded version (20CRv2c; Compo et al. 2015) was generated using the same model but with an improved lower boundary condition that comprises monthly fields of COBE-SST2 sea ice concentrations (Hirahara et al. 2014) and SSTs from an ensemble of Simple Ocean Data Assimilation with Sparse Input version 2 (Giese et al. 2016) corrected to COBE-SST2 for latitudes $>60^\circ$N. Furthermore, additional surface pressure observations, contained in the International Surface Pressure Databank version 3 (Cram et al. 2015), were taken into account. It is the only major reanalysis that assimilates only surface observations and extends from 1850 to 2014. The 20CRv2c dataset provides both analysis (e.g., MSLP) and forecast (e.g., V10m, T2m, SH2m, TCLC) fields. In this study, I use MSLP analysis fields based on the 56-member ensemble, which are available on a regular $2^\circ \times 2^\circ$ global grid from 90$^\circ$N to 90$^\circ$S latitude and from 0$^\circ$E to 360$^\circ$E longitude (91$\times$180 grid points). By contrast, the forecast fields are provided on a Gaussian grid covering 88.542$^\circ$N–88.542$^\circ$S and 0$^\circ$E–358.125$^\circ$E.

CERA-20C

The ECMWF Coupled Ocean–Atmosphere Reanalysis of the Twentieth Century (CERA-20C, Laloyaux et al. 2018) assimilates surface pressure observations from the International Surface Pressure Databank (ISPD; Cram et al. 2015), and additionally assimilates sea surface winds from the International Comprehensive Ocean–Atmosphere Data Set (ICOADS; Freeman et al. 2017). The CERA-20C also assimilates ocean temperature and salinity from the Hadley Centre’s EN4 dataset (Good et al. 2013).
5.2 Data Sources and Methods

5.2.2 Methods

Reanalyses Uncertainty

To compare uncertainties in the ensemble-based reanalyses, previous studies (e.g. Fyfe et al. 2013) have used the 2.5–97.5% ranges of the ensemble estimates (ensemble spread) and the 95% uncertainty ranges of the ensemble-mean (ensemble-mean uncertainty). These are derived by dividing the 2.5–97.5% ranges by the square root of the number of ensembles; 56 for 20CRv2c and 10 for CERA-20C. The ensemble-mean uncertainty normalises for different ensemble sizes, termed standard error, was similarly computed by Schneider and Fogt (2018). The ensemble means are compared using a two-tailed standard t-test and differences are reported at 5% significance level.

Offline data assimilation

The offline method adjusts the MSLP ensembles produced by the reanalyses (20CRv2c and CERA-20C) by nudging the MSLP ensembles closer to the new historical MSLP observations being assimilated. The conceptual cost function of the NDA algorithm can be defined (Blum et al. 2009) as:

$$J(G) = \int_0^T \left\langle W(X - X^0), X - X^0 \right\rangle dt + \left\langle K(G - \hat{G}), (G - \hat{G}) \right\rangle$$  \hspace{1cm} (5.1)

where X is a model state matrix, $X^0$ is an observational matrix, $\hat{G}$ and G represent nudging coefficients and the estimated optimal nudging coefficients, respectively; and W and K are specified weighting matrices. The state of the atmosphere is defined by

$$\frac{\partial X}{\partial t} = F(X) + G(X^0 - X)$$

with initial conditions, $X(0) = V$.

For this study, I have defined the cost function (J) using the least absolute shrinkage and selection operator (LASSO) regression analysis method in Eq. 5.2. The LASSO is one of the shrinkage methods which are used to retain a subset of the predictors while discarding the rest. The predictor subset produces a model that is interpretable and has possibly lower prediction error than the full model. The model is typically, trained on data $(x_1, y_1), \ldots, (x_N, y_N)$ to estimate the regression coefficients, $\beta_j$, for p predictors. Each $x_i = (x_{i1}, x_{i2}, \ldots, x_{ip})^T$ is a vector of feature measurements for the $i^{th}$ case. The minimisation of the cost function is carried out by reducing the RHS of the Eq. 5.2. The first term of RHS of Eq. 5.2 is the sum of residual sum of square and the second term is a sum of absolute values of the regression coefficients multiplied by the shrinkage/nudging coefficient ($\lambda$).
where, $y_i$ is predicted model state, $\beta_0$ is the linear intercept and $\beta_j$ is regression coefficient for each predictor $j$.

This offline assimilation method was developed by Brohan (2018) to integrate newly extracted station MSLP observations from Daily Weather Reports from the British Isles from 1856 to 1960 into reanalyses. The details of implementation are available at http://brohan.org/offline_assimilation/.
The model code developed for that project is stored on Github, a digital repository, which is made available under GNU Lesser General Public (open source) License as published by the Free Software Foundation.

I have used a similar cost function to constrain reanalyses output using our new historical whaling observations. The assimilation window is 8 hours long, centred on the analysis time of each reanalysis similar to many current reanalyses (Fujiwara et al. 2017). For each analysis time, the relationship between MSLP ensemble values at one of the global grid points and MSLP ensemble values at the locations of the observations falling within the assimilation window are modelled.

The last term of the cost function can be thought as a summation of the modulus of coefficients which is less than or equal to $s$, e.g. $|\beta_1| + |\beta_2| \cdots \leq s$. Here, $s$ is a constant that exists for each value of the nudging coefficient $\lambda$, and the equations are referred to as constraint functions. Due to the linear combination of coefficients, the region of constraints forms a rectangle in the feature space and has corners at each of the axes. The values of coefficients are determined by finding the first point of intersection between the constraint region and RSS, and will often occur at an axis. When this is so, one of the coefficients will be equal to zero, and at higher dimensions, many of the coefficients estimates may equal zero simultaneously. Hence, LASSO reduces over-fitting by restricting values of coefficients but also performs feature or predictor selection (Hastie et al. 2009). The model with the best fit will generally include only a subset of the predictors that are deemed truly informative.

**Reduction of Error**

To evaluate the effects of assimilation of new observations on reanalyses with respect to a reference dataset, I examined the reduction of error (RE) skill score (Cook et al. 1994).
I use the Hadley Centre Sea Level Pressure dataset (HadSLP2; Allan and Ansell 2006) as the reference in the assessment process. The RE is calculated between the absolute values of the ensemble mean of the reanalysis at a time-step and the reference monthly value for that time-step at each grid point. Hence, the RE compares, in this case, the closeness of the background state of a reanalysis to the reference with respect to the post-assimilation state and provides a spatial map of the impact of assimilation of new observations expressed as deviations from the reference. The RE can be defined as follows:

\[
RE = 1 - \frac{\sum (x^a_i - x^{ref}_i)^2}{\sum (x^f_i - x^{ref}_i)^2}
\]

(5.3)

where \(x^a\) is the ensemble mean of the reanalysis post-assimilation, \(x^f\) is the ensemble mean of the reanalysis pre-assimilation (model background state), \(x^{ref}\) is the reference dataset, and \(i\) refers to the time step.

**Computational environment and practical implementation**

The analyses are performed using Ubuntu 16.04 LTS operated workstation configured with Intel® Core™ i7 CPU X 990 with clock frequency 3.47GHz and GeForce 210/PCIe/SSE2 graphics card. The data manipulation is carried out using standard and custom code written in MATLAB® R2017b. The NDA is performed using python3 programming language code primarily written by Brohan (2019), customised and adapted for whaling dataset using Anaconda repository management tool.

**5.3 Results**

I investigate uncertainty in both reanalyses (20CRv2c and CERA-20C) over the periods covered by our whaling dataset, the 1930s and 1950s. I focus on zonal uncertainty across decades and seasons. Subsequently, I assimilate new observations into reanalyses and highlight changes in their ensemble-mean and ensemble-mean uncertainty structure.

**5.3.1 Decadal means and uncertainty**

The quality of reanalyses data in the historical period (1930s and 1950s) is assessed by comparing zonal means and uncertainties polewards of 60ºS. I plot decadal MSLP zonal
means and uncertainties for the 1930s (a) and 1950s (b) in the 20CRv2c and CERA-20C reanalyses in Figures 5.1 and 5.2, respectively. The ensemble spread (grey shading) is higher in the 1930s than in the 1950s and is also greater in the high latitudes than in the mid-latitudes. The ensemble-mean uncertainty (green shading) varies as a function of latitude (Table 5.1) and, for the same latitude band, uncertainty is higher in the CERA-20C than in the 20CRv2c. Even within each reanalysis, the range of uncertainties is quite large; the ensemble-mean uncertainty from the least uncertain 60ºS latitude band in the 1950s is less than 28% of the uncertainty from the most uncertain 90ºS zonal band in the 1930s in 20CRv2c. For CERA-20C, the range is even higher; the highest ensemble-mean uncertainty is 3.6 times of the lowest uncertainty.

All zonal latitude ensemble-means in the 1930s and 1950s were tested using a two-tailed t-test and were found to be significantly different from each other at each zonal latitude band in 20CRv2c (Table 5.1). Whereas for the CERA-20C, zonal means for the 1930s and 1950s were significantly different only between 58ºS to 64ºS, comparing zonal ensemble-means in the 1930s and 1950s for the CERA-20C, all ensemble-means at all latitude bands were found to be significantly different at a 95% level.

### 5.3.2 Seasonal variability

To compare seasonal uncertainty in the reanalyses, I investigate the seasonal ensemble-mean MSLP values and uncertainty in the CERA-20C and 20CRv2c reanalyses generated by combining the 1930s and 1950s decades. The seasons are divided into the austral summer
5.3 Results

Fig. 5.2 Ensemble mean MSLP values and uncertainty in the CERA-20C reanalysis for the 1930s (a) and 1950s (b) decades. The grey shading indicates the 2.5–97.5% ranges of the ensemble estimates and green shading shows the 95% uncertainty ranges of the ensemble means, derived by dividing the 2.5–97.5% ranges by the square root of the number of ensembles.

Table 5.1 The decadal zonal MSLP ensemble means and 95% uncertainty ranges of the ensemble means for 90°S, 80°S, 70°S and 60°S latitude bands, in hPa.

<table>
<thead>
<tr>
<th>Latitude band</th>
<th>Reanalyses</th>
<th>1930s</th>
<th>1950s</th>
</tr>
</thead>
<tbody>
<tr>
<td>90°S</td>
<td>CERA-20C</td>
<td>1006.83 ± 0.33</td>
<td>1006.94 ± 0.17</td>
</tr>
<tr>
<td></td>
<td>20CRv2c</td>
<td>1013.88 ± 0.14</td>
<td>1010.55 ± 0.09</td>
</tr>
<tr>
<td>80°S</td>
<td>CERA-20C</td>
<td>998.18 ± 0.27</td>
<td>997.75 ± 0.14</td>
</tr>
<tr>
<td></td>
<td>20CRv2c</td>
<td>1005.95 ± 0.11</td>
<td>1003.23 ± 0.06</td>
</tr>
<tr>
<td>70°S</td>
<td>CERA-20C</td>
<td>990.16 ± 0.27</td>
<td>990.29 ± 0.12</td>
</tr>
<tr>
<td></td>
<td>20CRv2c</td>
<td>995.18 ± 0.12</td>
<td>992.61 ± 0.06</td>
</tr>
<tr>
<td>60°S</td>
<td>CERA-20C</td>
<td>989.21 ± 0.14</td>
<td>989.55 ± 0.09</td>
</tr>
<tr>
<td></td>
<td>20CRv2c</td>
<td>993.80 ± 0.07</td>
<td>991.66 ± 0.04</td>
</tr>
</tbody>
</table>
(December-January-February or ‘DJF’), autumn (MAM), winter (JJA) and spring (SON). The zonal DJF seasonal ensemble-means and 95% uncertainty ranges of the ensemble means, together with its spread, are shown for CERA-20C (Fig. 5.3a) and 20CRv2c (Fig. 5.3b) reanalyses.

Fig. 5.3 The zonal MSLP DJF seasonal ensemble mean (black lines). Grey shading indicates the 2.5–97.5% ranges of the ensemble estimates, green shading shows the 95% uncertainty ranges of the ensemble means, derived by dividing the 2.5–97.5% ranges by the square root of the number of ensemble members for the CERA-20C (a) and 20CRv2c (b), and spatial differences between DJF seasonal means of CERA-20C and 20CRv2c are overlaid by cumulative uncertainty (black lines; in hPa) in (c).

The ensemble spreads (grey shading) are generally higher for 20CRv2c than CERA-20C not only for the DJF season (Fig. 5.3a,b) but for other seasons as well (Table 5.2). However, the ensemble-mean uncertainty ranges (green shading) for the 20CRv2c are smaller than CERA-20C. For example, for the least uncertain season and zonal band, DJF at 60°S, the uncertainty of CERA-20C is 33% higher than 20CRv2c (Table 5.2). For the most uncertain season, JJA, the difference between the reanalyses is more pronounced; CERA-20C is more than 50% higher than 20CRv2c in the 90°S zonal band. In general, uncertainty increases polewards for all seasons (e.g. the DJF season reaching a maximum of 0.52 hPa and 0.45 hPa at 90°S compared to 0.33 hPa and 0.22 hPa at 60°S for CERA-20C and 20CRv2c, respectively). Seasonally, the JJA season has the maximum uncertainty in all zonal bands; 0.90 hPa and 0.43 hPa at 90°S compared to 0.46 hPa and 0.26 hPa at 60°S for CERA-20C and 20CRv2c, respectively.

The spatial difference between the DJF seasonal means of CERA-20C and 20CRv2c are overlaid by cumulative uncertainty is shown in Fig. 5.3c. The differences between reanalyses are relatively small for most regions in the range of 100 hPa to -100 hPa. However, both reanalyses point to the region of high uncertainty off the Ross Sea known to contain Amundsen Sea Low (ASL; Turner et al. 2013).
Table 5.2 The zonal MSLP seasonal ensemble means and 95% uncertainty ranges of the ensemble means for 90°S, 80°S, 70°S and 60°S latitude bands, in hPa.

<table>
<thead>
<tr>
<th>Latitude band</th>
<th>Reanalyses</th>
<th>DJF</th>
<th>MAM</th>
<th>JJA</th>
<th>SON</th>
</tr>
</thead>
<tbody>
<tr>
<td>90°S</td>
<td>CERA-20C</td>
<td>1004.36±0.52</td>
<td>1009.61±0.46</td>
<td>1009.60±0.90</td>
<td>1003.96±0.36</td>
</tr>
<tr>
<td></td>
<td>20CRv2c</td>
<td>1003.80±0.45</td>
<td>1015.26±0.42</td>
<td>1019.74±0.43</td>
<td>1010.07±0.36</td>
</tr>
<tr>
<td>80°S</td>
<td>CERA-20C</td>
<td>998.69±0.52</td>
<td>999.62±0.39</td>
<td>998.50±0.99</td>
<td>993.67±0.31</td>
</tr>
<tr>
<td></td>
<td>20CRv2c</td>
<td>999.85±0.35</td>
<td>1007.02±0.34</td>
<td>1009.64±0.34</td>
<td>1001.86±0.32</td>
</tr>
<tr>
<td>70°S</td>
<td>CERA-20C</td>
<td>993.26±0.52</td>
<td>991.21±0.36</td>
<td>990.45±0.87</td>
<td>985.99±0.35</td>
</tr>
<tr>
<td></td>
<td>20CRv2c</td>
<td>992.59±0.29</td>
<td>994.15±0.32</td>
<td>997.58±0.36</td>
<td>991.25±0.30</td>
</tr>
<tr>
<td>60°S</td>
<td>CERA-20C</td>
<td>990.08±0.33</td>
<td>989.50±0.25</td>
<td>990.38±0.46</td>
<td>987.55±0.32</td>
</tr>
<tr>
<td></td>
<td>20CRv2c</td>
<td>990.99±0.22</td>
<td>992.36±0.25</td>
<td>995.21±0.26</td>
<td>992.36±0.17</td>
</tr>
</tbody>
</table>

5.3.3 Effect of data assimilation

The MSLP contours from the 56-ensemble members and the ensemble mean at 10-Dec-1932 12:00:00 in 20CRv2c reanalysis, before (a) and after (b) assimilating whaling ships’ observations, are shown in Figure 5.4. The red dots in the panel (b) indicate the locations of the new historical observations being assimilated. The huge spread of individual ensemble MSLP contour lines (thin blue) indicate low agreement among ensemble members giving rise to high uncertainty in almost all regions showed in Figure 5.4 and regions where the ensemble spread (standard deviations) is less than 3 hPa are indicated by thick black lines. The new assimilated observations (red dots) constrain each ensemble member to reflect new information in the MSLP fields. A new pattern starts to emerge in the form of newly identified weather systems with more certain mean MSLP fields.

The uncertainty before assimilation (Fig. 5.4a) is higher than after (Fig. 5.4b), with more thick black mean contours especially around the new observations (red dots, Fig. 5.4b). For example, a cyclonic system that was misplaced away from the locations of the new observations in Figure 5.4a, is now seen to be deeper and moved north-east after the assimilation (Fig. 5.4b). Conversely, the narrow and deep cyclonic system present near 60°S 100°E (Fig. 5.4a) is made less deep and spread out. The ASL is present in the area and with the almost same magnitude in both panels. The high-pressure regions are also modified by bringing tongues of the mid-latitude high-pressure belt towards the tip of South America in Figure 5.4b. The new observations have also produced a previously unidentified cyclonic system in the reanalyses (Fig. 5.4). The existing pattern of MSLP fields is either reinforced
or diluted according to the relationship between grids at those observation points and the rest of the grid points.

The effect of assimilation on CERA-20C is shown in Figure 5.5 for the same analysis day and time. Almost all regions are highly uncertain, as with 20CRv2c (Fig. 5.4). The MSLP fields in the CERA-20C reanalysis (Fig. 5.5) are broadly similar to those of 20CRv2c. CERA-20C also shows a misplaced cyclonic system in the same way as in the 20CRv2c reanalysis, but with much greater uncertainty. After assimilation, the cyclonic system deepens and is relocated to the north-east (Fig. 5.5a). Another cyclonic system near 60ºS 100ºE (Fig. 5.4a) is also deepened after the assimilation (Fig. 5.5b). A new depression has developed from a low tongue-structure in Figure 5.5a and is moved near 60ºS 180ºE in Figure 5.5b.

Some weather features not present before assimilation are visible afterwards. Interestingly, those features can also be identified in both reanalyses. For example, a low-pressure tongue formed in the Ross Sea after assimilation in 20CRv2c (Fig. 5.4b), is similar to a depression identified near 60ºS 180ºE in CERA-20C (Fig. 5.5b). Even though no information is shared between the reanalyses except the new assimilated historical ship observations, the existence of similar features independently confirms the validity of the assimilation process and the utility of information added from new observations. In addition, the uncertainty, in general, is reduced after the assimilation, in that better constrained MSLP values appear not just limited to the observation sites.

Another similar comparison is shown in Supplementary Figs. A.3 and A.4 for 13-Dec-1957 12:00:00 for both reanalyses. The effect on the reanalyses is similar to that in the 1930s except that the reanalyses are more certain, hence, the effect of new observations is less profound.

### 5.3.4 Quantification of the sensitivity of ensemble means and uncertainties

Here, I quantify the changes in the ensemble means and uncertainties post-assimilation for both 20CRv2c and CERA-20C reanalyses.

As shown earlier (Section 5.3.3), the new whaling observations reduce uncertainties; Figure 5.6 shows ensemble mean contours of 20CRv2c at 10-Dec-1933 12:00:00 before (a) and after (b) assimilation overlaid by the 95% uncertainty ranges of the ensemble means. Here, ensemble-means uncertainty ranges are shown instead of standard deviations, to
Fig. 5.4 The MSLP contours from 56-ensemble members (thin blue lines) and ensemble mean (thick black lines) at 10-Dec-1932 12:00:00 in 20CRv2c reanalysis, before (a) and after (b) assimilating whaling ships’ observations. The thin blue lines indicate ensemble contours and the contours where ensemble spread (standard deviations) is less than 3 hPa are indicated by thick black lines. Yellow dots represent observations present in the dataset and red dots represent newly added whaling dataset observations.
Fig. 5.5 The MSLP contours from 10-ensemble members (thin blue lines) and ensemble mean (thick black lines) at 10-Dec-1932 12:00:00 in CERA-20C reanalysis, before (a) and after (b) assimilating whaling ships’ observations. The thin blue lines indicate ensemble contours and the contours where ensemble spread (standard deviations) is less than 3 hPa are indicated by thick black lines. The existing observations (yellow dots) used in the reanalysis were not provided hence not shown here. Red dots represent newly added whaling dataset observations.
normalise for the different number of ensemble members in each reanalysis. Before the assimilation, the area around the new observations is one of the most uncertain regions (Fig. 5.6a). After assimilation, it becomes the region of least uncertainty (the area around red dots; Fig. 5.6b). The new observations bring increasing certainty to the surrounding area in each ensemble, hence reducing the spread within the ensemble members. The reduction in the uncertainty is about 80% in the 3,000 km\(^2\) area in the immediate vicinity of the observations and 60% in the 7,000 km\(^2\) area in the extended region.
Fig. 5.6 The 20CRv2c ensemble mean MSLP contours at 10-Dec-1933 12:00:00 before (a) and after (b) assimilation overlaid by the 95% uncertainty ranges of the ensemble means (in hPa) and, the difference between (b) minus (a) is calculated only where 95% uncertainty in ensemble means in after assimilation (b) is less than 1.2 hPa, shown in (c), contours at -4 (red) and 4 (blue) hPa.

Fig. 5.7 Same as Fig. 5.6 but for CERA-20C reanalysis data and, the difference between (b) minus (a) is calculated only where 95% uncertainty in ensemble means in after assimilation (b) is less than 2.5 hPa, shown in (c), contours at -7 (red) and 7 (blue) hPa.
The magnitude of changes in the MSLP fields are shown in Fig. 5.6c but are limited to regions with less than 1.2 hPa ensemble-mean uncertainty (equivalent to 9 hPa standard deviations). The extended area around the observations and the Ross Sea region show a deepening of up to 4 hPa, but the MSLP fields in the Weddell Sea and Southern Ocean near 60°S 100°E and South of New Zealand increase simultaneously by the same amount.

Similarly, for the CERA-20C reanalysis at the same date and time, the area around the new meteorological observations is highly uncertain up to 4 hPa before assimilation (Fig. 5.7a). After assimilation, the uncertainty drops to 0.7 hPa in that area and many parts of the domain under investigation (Fig. 5.7b). This is equivalent to an 83% reduction in uncertainty around the observations and up to 65% in many regions in the domain. The differences in the MSLP fields (After minus before) are shown in Figure 5.7c, although limited to regions with uncertainty less than 2.5 hPa (equivalent to 8 hPa standard deviations). The area around the observations, off the coast of Southern America and the area around a newly identified cyclonic system at the northern limits of the Ross Sea, is deepened by up to 7 hPa whereas values are increasing in the Weddell Sea, the Bellinghausen Sea and off the Wilkes Land coast by the same amount.

### 5.3.5 Reduction of Error (RE)

The RE values are computed based on the ratio of pre- and post-assimilation anomalies with respect to the monthly field of the same time-step of the HadSLP2 dataset. The two separate time-steps used previously are considered in this analysis: 10-Dec-1932 12:00:00 and 13-Dec-1957 12:00:00. The maximum value of one for RE is obtained when the post-assimilation state is equal or very close to the reference dataset at that grid location. Negative RE values indicate that the pre-assimilation background state is closer to the reference dataset than the post-assimilation state. RE values close to zero indicate that pre- and post-assimilation states are almost equally distinct from the reference dataset; in other words, changes made in the post-assimilation state are minimal with respect to the reference dataset vis-a-vis the pre-assimilation state.

At the 10-Dec-1932 12:00:00 time-step, RE scores for CERA-20C (Fig. 5.8a) show large areas where the pre-assimilated reanalysis state is much closer to the HadSLP2 dataset than the post-assimilated reanalysis state (areas shaded with blue). These blue regions are often observed over oceans where both background state of CERA-20C and the HadSLP2 dataset are generated from the same primary observations, hence closer. In contrast with
regions (areas shaded red) where new observations (green dots) are present, move CERA-20C post-assimilation state closer to reference dataset than background state. For the 20CRv2c (at the same time-step), the spatial map shows a similar pattern but with a larger proportion of white regions, where departures from the reference dataset are minimal for pre- and post-assimilation states (Fig. 5.8c).

For the second time-step (13-Dec-1957 12:00:00) the pre-assimilation state is closer to the reference dataset than the post-assimilation state in fewer regions than in the earlier time-step for CERA-20C reanalysis (Fig. 5.8b). Further, I find larger regions where the changes from pre- to post-assimilation states are minimal from the reference dataset. In addition, for the 20CRv2c reanalysis, a very large proportion of regions show no perceptible changes in the post-assimilation state (Fig. 5.8d). In general, all four plots show the movement of post-assimilation states towards the observational reference dataset, with an exception in the Weddell Sea at the first time-step (Fig. 5.8a).

5.4 Discussion

5.4.1 Measurement of ensemble uncertainties

Although CERA20C and 20CRv2c are both multi-member ensembles, their sampling of the sources of uncertainty are quite different (Schneider and Fogt 2018). The individual ensemble members in CERA-20C are generated by perturbations accounting for errors in the observational records and forecast model. During the generation of ensemble members, a flow-dependent background error matrix is computed which becomes the basis for combining forecast and observational data, and thus of estimates of ensemble spread (Laloyaux et al. 2016). By contrast, in 20CRv2c, at each analysis time the ensemble mean is the previous analysis, and the ensemble spread is constrained by the analysis error, which is the difference between the ‘first guess’ forecast and the observations (Tippett et al. 2003, Compo et al. 2011). Overall, CERA20C and 20CRv2c do not estimate the same sources of uncertainty; hence, comparing ensemble-mean uncertainty may not be a like-for-like comparison (Schneider and Fogt 2018). Nonetheless, both reanalyses produce uncertainty statistics based on ensemble standard deviations, standard error or a similar metric to assess the quality of reanalysis at specific places and times (Compo et al. 2011, Laloyaux et al. 2018). Thus, the ensemble-mean uncertainty used in this study provides a valid comparison of reanalyses.
Fig. 5.8 Spatial map of the impact of data assimilation on reanalyses presented by RE values, at 10-Dec-1932 12:00:00 for CERA-20C (a) and 20CRv2c (c) and 13-Dec-1957 12:00:00 for CERA-20C (b) and 20CRv2c (d). The green dots indicate the positions of observations assimilated at each time-step. The Weddell Sea is denoted by WS in (a).
5.4.2 Differences in reanalyses across seasons and decades

The comparison between the 1930s and the 1950s emphasises the beneficial effect of inclusion and assimilation of a greater number of observations, as the mean uncertainty in the former is almost double that in the latter decade (Table 5.1). The uncertainty reduction from the 1930s to 1950s in both reanalyses has been attributed to the increased number of available observations for reanalyses (Compo et al. 2011, Laloyaux et al. 2018). The ensemble-mean uncertainty in the CERA-20C is found to be consistently higher than 20CRv2c between the decades (Table 5.1). These findings are in agreement with Schneider and Fogt (2018), who found that for the time period under investigation, CERA-20C has lower quality than 20CRv2c in the region polewards of 60ºS. It is only after the 1960s that CERA-20C’s standard error drops lower than that of 20CRv2c, and this directly follows the numbers of observations assimilated into reanalyses (Schneider and Fogt 2018). Furthermore, the lower ensemble mean MSLP in the 1950s than in the 1930s, and in CERA-20C than in the 20CRv2c, can be attributed to increased numbers of observations assimilated; hence, a more realistic and tightly coupled model, respectively.

The contrast between CERA-20C and 20CRv2c is even starker in the seasonal analysis. Although the ensemble spread is much higher in 20CRv2c than CERA-20C, mean uncertainty is higher in the latter than the former (Fig. 5.3). The large ensemble spread in 20CRv2c is due to ensemble members being drawn from a large posterior distribution. However, mean uncertainty becomes smaller when normalised for the number of ensemble members (Fig. 5.3b). On the other hand, the smaller ensemble spread in CERA-20C is achieved by the coupled model’s feedback loops which help to evolve all components of the Earth system together tightly and is also aided by a longer assimilation window. However, this process spreads too much information in the form of large-scale increments in the ensemble space during poorly observed periods (Laloyaux et al. 2018). Hence, greater mean uncertainty is found in the CERA-20C reanalysis.

The average magnitude of mean uncertainty increases from seasons DJF to MAM to SON to JJA for both reanalyses (Table 5.2). The higher uncertainty in JJA compared to DJF is likely due to a combination of the unresolved and unconstrained asymmetric circulation in models, greater inter-annual variability, and fewer observations in the Austral winter period of JJA (Schneider and Fogt 2018). A similar uncertainty is also shown by the Laloyaux et al. (2018), in their Figure 11, for the SON season in the region polewards of 60ºS. In the same vein, lower uncertainty among the ensemble members in the mid-latitudes and progressively higher uncertainty towards the South Pole in both reanalyses (decadal and seasonally)
can be attributed to limited observations in the interior of the Antarctic continent. Due to a lack of observations polewards of 60ºS, the models adjust the atmospheric circulation to regions where it is least constrained and this gives rise to increased uncertainty as compared to mid-latitudes (Laloyaux et al. 2018).

The spatial uncertainty of the individual reanalyses is similar to uncertainty in the difference between CERA-20C and 20CRv2c, as uncertainties are additive; the mean difference and associated uncertainty for the DJF season are shown in Fig. 5.3c. One of the highly variable regions of the Southern Ocean is the region of the ASL in the box bounded by 50ºW, 180ºW, 60ºS and 75ºS (Turner et al. 2013). In previous studies (e.g. Hosking et al. 2013, Fogt et al. 2012), it was found that the ASL shows maximum variability in the Austral summer, a finding confirmed in this study. The austral summer has particular relevance to our study because the whaling dataset to be assimilated contains mostly summer observations and, as demonstrated above (Figs. 5.6, 5.7), the assimilation process can help alleviate some uncertainty.

Although CERA-20C comprises a coupled ocean-atmosphere-sea ice model and 4D-Var assimilation system, the decadal and seasonal ensemble-mean uncertainties are higher (almost twice) than 20CRv2c in the 1930s and 1950s (Table 5.1, Figs. 5.1 and 5.2). This discrepancy should be viewed in the light of a number of differences; for example, assimilation methods – observations assimilated and the estimation of true covariance in each reanalysis. The variational assimilation method (incremental 4-D Var) used in CERA-20C has been found to be better than EnKF methods for reanalyses that assimilate only pressure observations (Whitaker et al. 2008). Both reanalyses assimilate surface pressure observations, but a crucial distinction is that CERA-20C additionally assimilates marine wind and subsurface temperature and salinity profiles for the ocean component of the coupled model. However, assimilation of additional observations cannot cause increases in observed uncertainty. In addition, the longer assimilation window of 24 hr in CERA-20C offers a chance to include many more observations in the analysis cycle than the 6 hr window of 20CRv2c, thereby improving the estimate of background-error covariances and lowering uncertainty.

Furthermore, CERA-20C assimilates only a subset of available quality-corrected observations; hence it is possible that essential information contained in the omitted ones is being lost. This contrasts with 20CRv2c which uses all available quality-corrected observations. In the example presented by Laloyaux et al. (2018), the tropical storms and hurricanes are poorly identified in the CERA-20C as compared to 20CRv2c, due to rejection of cyclone ‘Best’ track data in the former. This situation stems from the quality assurance system of
CERA-20C, which rejects observations too different from the background, which is often the case for the periods and regions poorly observed (e.g. the pre-satellite era in the Southern Hemisphere). This is a cause for concern.

Lastly, the finite ensemble size can lead to an underestimation of the true covariance. Many methods for inflating the covariance have been developed to address this issue (e.g., Anderson 2009, Whitaker et al. 2008, Whitaker and Hamill 2012). Kalnay et al. (2007) state that an ensemble size of 80-100 was necessary to estimate the background error covariance accurately. To offset for sub-optimal ensemble-estimated background-error covariances derived from fewer numbers of ensemble members (56) than the recommendation in the 20CRv2c, two methods were used to correct these errors; covariance inflation and distance-dependent covariance localization. Due to the sparse observational density for much of the reanalysis, simple multiplicative covariance inflation was used as described in Whitaker et al. (2004). The ensemble covariance was forced to zero to avoid filter divergence arising from spurious long-distance correlations. To this effect, a spatial filter with a horizontal localization distance of 4000 km was used with the same localization function as Whitaker et al. (2004).

In contrast with 20CRv2c, CERA-20C with 10 ensemble members estimates and updates the atmospheric background-error covariance matrix by proportionally weighing the climatological forecast error estimates computed over 2014 using the full observing system and from 10 daily forecasts produced for each assimilation cycle. The climatological error estimates are given 85% while daily error estimates are given 15% weight as giving too much emphasis to a small number of daily forecasts error estimates will lead to inconsistency in the background error correlation calculation (Laloyaux et al. 2018). I believe that, by using this distribution of weights method, the error covariance is optimally estimated as a large proportion is drawn from comparatively certain periods, and only a small proportion arises from highly uncertain daily forecasts from the pre-satellite era under investigation. Therefore, it is unlikely that the particular method of error covariance estimation in CERA-20C is the cause of higher uncertainty than in 20CRv2c.

After considering all three major potential factors for higher uncertainty, using a subset of available quality-corrected observations (white-listed observations) in CERA-20C is most likely the cause for the higher uncertainty than 20CRv2c. Hence, adopting less stringent requirements for Southern Ocean observations can improve the uncertainty in the CERA-20C.
5.4.3 Data assimilation methodology

The core feature of the LASSO nudging data assimilation process is the generation of lean (sparse) models. The physical interpretation of lean models with limited predictors is convenient as only the most important predictors are retained, and others are discarded. For example, observations recorded at distant points from the point of interest will have a negligible climatological effect but using methods such as Ordinary Least Square (OLS) regression would result in negligible but non-zero coefficients. This situation can be avoided by using LASSO, which makes coefficients of unimportant predictors exactly zero.

The points of intersection or values of regression coefficients depend on nudging coefficient $\lambda$. The range for $\lambda$ is from 0 to $\infty$. A zero value is equivalent to the OLS regression method (where all variables are included in the model), and large $\lambda$ values penalise the large coefficients to such an extent that all predictors are dropped from the model. Any increase in $\lambda$ from zero is beneficial as it reduces the value and number of coefficients, thus reducing the variance (hence avoiding over-fitting), without losing any important features. Even so, after a certain threshold, the model starts losing important properties, giving rise to bias in the model and thus under-fitting. To arrive at the optimal value of the nudging coefficient, I tested different values for $\lambda$ from 0.2, 0.5, 1, 10 to minimise Root Mean Square Error (RMSE) between model and predicted values. It was found that, in the majority of cases in our analysis, when the number of observations within an assimilation window was fewer than 4, feature reduction is not a major concern; hence, I have used $\lambda = 1$ as a compromise between bias reduction and retaining only important predictors.

An optimal nudging (ON) method has been proposed by Vidard et al. (2003) to compute coefficients and has been found similar in performance to Kalman filtering. However, the proposed ON technique requires the computation of the adjoint state of the model equations, resulting in major computational requirements, which is contrary to the purpose of this study.

5.4.4 Post-assimilation changes and their sensitivities

The state of the atmosphere in the 20CRv2c reanalysis at 10-Dec-1932 12:00:00 is nudged towards the observations (Fig. 5.6). There are a few notable observations before I consider the changes in the post-assimilation reanalysis state. Four polar lows are identified circling the Antarctic continent. The whole region south of 40°S is encapsulated within 0.5 hPa ensemble-mean uncertainty, which is twice the uncertainty found in the sub-tropics. In
addition, the surrounding region of would-be assimilated observations (60ºS 0ºE) is highly uncertain (1.2 hPa) as it is close to a depression. The effect of assimilating whaling ships’ observations are quite remarkable (Fig. 5.6b). The uncertainty structure of the analysis state remains largely the same except for the region where new observations are located. The LASSO method’s sparse model is more dependent on the strength of coefficients than the distance between locations of interest and locations of observations. It essentially acts as an improved spatial filter similar to the one implemented in 20CRv2c reanalysis.

The objective limits to the influence of new observations provide assurance that the nudging method used does not result in changes that are physically and numerically inconsistent. Now the uncertainty around new observations is comparable to that found in the sub-tropics and has a clear effect on identifying individual depressions with greater confidence (Fig. 5.6b). Furthermore, the ensemble-means are also changed, deepening and shifting the depression found south-west of the observations. Some depressions have deepened, and some are made shallower; however, a new depression has formed in the Ross Sea as seen in both Fig. 5.6 b and c.

For the same analysis time, CERA-20C reanalysis shows that a region south of 40ºS is highly variable in terms of uncertainty (Fig. 5.7a). Some depressions are highly uncertain (4 hPa) (e.g. in the Weddell Sea and Pacific Ocean sector of the Southern Ocean), whereas other depressions are anomalously certain (four times more certain, 0.7 hPa) (e.g. in the Indian Ocean sector and area of the AS Low). This again emphasises the deficiency of CERA-20C to estimate error covariance accurately. Post-assimilation uncertainty structure is substantially altered, the contours are segmented, and regions distant from the observations undergo large changes (Fig. 5.7b). Possibly due to tightly-coupled CERA-20C model, however, most uncertain regions remain so except the area around the new observations. The differences plot (Fig. 5.7c) presents a number of regions deepened and shallowed by almost twice the amount of that of 20CRv2c (Fig. 5.6c). The alternating structure in both uncertainty and ensemble-mean in the post-assimilation analysis state (Fig. 5.7b,c) suggests an underlying zonal wave-number 3 propagation similar to the one found in the MSLP fields by previous studies (Cerrone et al. 2017).

5.4.5 RE scores

As the reference dataset HadSLP2 is derived from ICOADS observations, RE scores computed with respect to HadSLP2 offer us ways to investigate the difference between pre-
assimilation reanalyses states and observational reference. I can also assess the response of reanalyses to the changes made by the DA method in the post-assimilation states. At the first time-step, the pre-assimilation background state in many regions is closer to the reference dataset than post-assimilation states, especially over oceans (Fig. 5.8a,c). Although HadSLP2 was generated using a previous version of ICOADS with fewer observations than the current version. The fact that pre-assimilation states match HadSLP2 indicates that not many new observations have been assimilated into the reanalyses over the Southern Ocean since the generation of HadSLP2.

The changes made in the reanalyses post-assimilation display an interesting pattern (Fig. 5.8b, d); much of Antarctica, together with regions close to land-areas and assimilated observations, are closer to the observational reference. This could be due to well-represented atmospheric processes over Antarctica and assimilation of observations drawn from these regions, respectively. However, for the second time-step, the biggest change has been the regions where the reanalysis state has changed minutely in the post-assimilation as compared to pre-assimilation states (white areas, Fig. 5.8b,d). This demonstrates better quality of reanalyses, as pre- and post-assimilation states do not show a large deviation from the reference dataset, except at the locations of the newly assimilated observations and regions climatologically coupled with those locations (Fig. 5.8b,d).

Ideally, the spatial map of RE score should only display regions where improvements are brought by the assimilation of new observations, and the rest of the domain should experience negligible changes; this scenario is almost reached in the 20CRv2c reanalysis at the second time-step (Fig. 5.8d).

Comparing both reanalyses and time-steps, it is observed that, at the first time-step, the reanalyses and reference dataset are largely the same in the pre-assimilation states. However, in post-assimilation states, more regions are tightly coupled, consequently adding new observations leads to widespread changes in the wider domain which may not be physically consistent. By contrast, in the second time-step, when reanalyses have better quality, most of the changes are negligible except in locations around new observations. Hence, RE scores are useful to assess the relative impact of new observations on the reanalyses when measured as deviations from the observational reference.
5.5 Conclusions

In this Chapter I have successfully assimilated new historical whaling-log meteorological observations (Teleti et al. 2019, Chapter 2 in this thesis) into two century-long reanalyses (20CRv2c and CERA-20C). I demonstrate that the offline methods used in this study can provide a simpler and computationally inexpensive way to assimilate historical meteorological data and contribute to the broad understanding of historical climate, in our case in the very data-sparse Southern Ocean. Such processes serve a two-fold purpose. The first is providing the reanalyses community with a system to contextually evaluate suitability, alongside statistical testing, of would-be assimilated observations for the inclusion into reanalyses without running costly assimilation methods. The second is that the data-rescue community can utilise these tools to take stock of their efforts, to help them focus current and future data-rescue efforts on specific regions and times for maximum effectiveness without the need to wait for the next release of completely new and computationally costly reanalyses.

The quality of both reanalyses (20CRv2c and CERA-20C) in the historical period is found to be poor over the Southern Ocean domain. In general, 20CRv2c is found to be better than CERA-20C, and the 1950s better than the 1930s, Summer months better than Winter and the 60° zonal band better than the 90° zonal band. The differences in the reanalyses are at their highest in autumn, followed by winter, summer and spring. But all seasons show large uncertainty in the MSLP fields over the ASL region in the Ross Sea. The results of assimilation (sensitivity) of new observations on reanalyses depend on the background state and inherent coupling in each reanalysis. For example, CERA-20C retains a coupled wave-like impact structure albeit confined, 20CRv2c mostly reinforces existing weather features post-assimilation. Further, it is found that, as being relatively more uncertain CERA-20C benefits from data assimilation than 20CRv2c, and shows the strong influence of new observations on its MSLP fields.

In addition, the effects of data assimilation in the 1950s are more subtle because the increased number of observations already assimilated in the reanalyses compared to the 1930s, which means that, climate fields are more certain in the 1950s. The reference dataset such as global gridded HadSLP2 dataset is very useful to assess modern reanalyses, which revealed that background states of CERA-20C are closer to reference dataset than the post-assimilation states, with more improvements in the 1950s. While, 20CRv2c in the 1930s and 1950s show large areas where post-assimilation states are minimally affected with respect to pre-assimilation states, except where new observations are assimilated. I have demonstrated
that nudging method can be used realistically to assimilate new historical observations into 
multi-ensemble reanalyses in the Antarctic domain. The pressure fields have been shown to 
improve in terms of certainty post-assimilation.

The current assimilation method could be improved by weighting would-be assimilated 
observations with respect to neighbouring observations and prior estimates (Tavolato and 
Isaksen 2015). Two new but computationally inexpensive data assimilation methods could 
be investigated in the future. The first method is Backward-Forward Nudging (BFN; Auroux 
and Blum 2008) process consists of forward nudging step similar to the current method used 
in this Chapter, but with an additional backward nudging step where final state derived from 
forward step is nudged towards initial state. In other words, the final state derived from the 
forward step becomes the initial condition for the backward step and process is repeated until 
cost function converges. By nudging the model state both forward and backward in time, 
eliminates numerical inconsistency, e.g. assimilation shock to the model state, providing a 
smoother flow of climate information across time-steps.

The second method is the sequential implementation of the Ensemble Kalman Filter called 
Ensemble Kalman Fitting (EKF; Franke et al. 2017). This method updates the ensemble 
mean and member anomalies individually without explicitly updating the covariance matrices. 
Similar to current method used in this study, to reduce the influence of spurious correlations 
in the background error covariance matrix, an ensemble square root filter is applied to restrict 
the influence of distant observations on climate fields. But they use much longer assimilation 
window of 6 months to assimilate palaeoclimatological records, hence leading to coarser 
updates and reanalysis resolution.

The extension of current work would be to constrain the multi-ensemble reanalyses with 
BFN process and compare those results with those obtained from the EKF method with much 
shorter assimilation window, to understand how these processes combine uncertainties in 
the observational data with reanalyses. The comparison would point to a data assimilation 
methodology that can better utilise historical observations to improve the representation of 
historical atmospheric circulation in the reanalyses.
Chapter 6

Conclusions

6.1 Overview

Due to a lack of long land-based meteorological records around the Southern Ocean, historical ship-based observations are investigated in this thesis. The Christian Salvesen Whaling Company’s logbooks of whaling ships operating in the Southern Ocean in the 1930s and 1950s offer an invaluable window into the historical climate of the region, especially that of the Weddell Sea sector. An historical climate dataset is produced from the meteorological observations extracted from eight and ten whaling seasons/years from the 1930s and 1950s decades, respectively, from the logbooks.

The MSLP changes across the Southern Ocean diverge; for example, in the northern reaches (55°S latitudinal band) MSLP in the historical period (1930s and 1950s) is found to be lower than the modern period (1981-2010) climatology, suggesting a positive trend in the MSLP at least since the 1950s. In contrast, in the southern reaches (65°S latitudinal band) MSLP is found to be decreasing over the same period. In addition, significant changes are detected in the semi-annual cycle of the Southern Hemisphere’s circumpolar trough as it is found to be occupying latitude bands closer to the Antarctic coast for a larger portion of the summer in the modern period than historically. This view is supported by increased cyclonic frequency in the summer during the modern period (Grieger et al. 2018), pointing to the expected influence of the circumpolar trough in the Antarctic coastal zone (Thompson and Wallace 2000).

Furthermore, this thesis proposes that the prolonged presence of the circumpolar trough close to the Antarctic coast is partly responsible for the decreasing MSLP found around the
Antarctic coast during summer in the modern period (Turner et al. 2005). The historical SAM index generated from 1930-1960 in this thesis has been found to be increasing during this period, suggesting that the positive trend found in the SAM index in the modern period (Marshall 2003) began as far back as the 1930s.

Subsequently, Southern Ocean historical cyclonic frequency is estimated using a semi-supervised cyclone identification algorithm. The total number of cyclones detected for the 1930s and 1950s is similar despite differences in the number and spatial distribution of observations in both decades. The average number of cyclones per year for the historical period is statistically lower than the average for the 1999-2008 period over a control area (in the Weddell Sea), confirming previous research which found an increasing trend in cyclones in the Southern Ocean. Finally, the whaling dataset is assimilated into two (CERA-20C and 20CRv2c) current-generation numerical climate reanalyses using an offline data assimilation method. The uncertainty in the MSLP fields decreased in both reanalyses by \( \sim 40\% \) in the area of newly added whaling ship observations post-assimilation.

Overall, it has been shown that meteorological observations from whaling logbooks can be utilised to reconstruct historical climate in terms of MSLP variability, climate modes (e.g. SAM), the identification of individual cyclones and to improve the representation of past climate in numerical reanalyses. The almost indistinguishable changes observed (historical compared to modern) in this thesis could be attributed to the brevity of the observations restricting our ability to detect possible anthropogenic effects.

6.2 Summary of the thesis

A readily accessible, standardised and quality checked IMMA-compliant dataset of historical meteorological observations from the Southern Ocean is created in Chapter 2. The dataset is the result of the first-ever study to extract meteorological observations solely from whaling logbooks in the Southern Ocean. Each record contains a number of positional, meteorological and meta-data parameters found in the Christian Salvesen Co. whaling logbooks of the 1930s and 1950s. Each parameter is manually extracted from logbooks and stored in a relational database. All data points are passed through statistical tests to detect and correct erroneous values. To make the dataset accessible and inter-portable with other marine datasets, data are homogenised and standardised. Each record in this dataset is produced according to recommendations in the IMMA format, bringing them to a level similar to that of existing international datasets relating to historical meteorological records such as ICOADS.
In total, the assembled dataset contains more than 12,000 observations recorded during 4604 observation-days spanning two decades. It contains 71 variables in total, including 48 and 23 variables for the Core variable and Ship meta-data sections, respectively. It should be noted that not all variables are populated in the dataset. Wind conditions, air pressure and temperature fields are more populous than other climatic fields. Meta-data fields such as the Official Number (ON), type of vessel, number of engines and dimensions are collected from various sources and are provided alongside the meteorological observations.

The whaling dataset (Teleti et al. 2019, Chapter 2 in this thesis) is found to be reliable in Chapter 3, refuting some of the earlier studies which raised concerns regarding whaling data. The whaling dataset is found to be similar to other marine observational datasets (e.g. ICOADS) in terms of its quality for climatological studies. The seasonal MSLP analysis of the whaling dataset confirmed previously identified semi-annual expansion and contraction cycles of the circumpolar trough over the Southern Ocean. However, a significant alteration to the circumpolar trough cycles, in the form of a prolonged presence of the circumpolar trough close to the Antarctic coast for a larger proportion of the summer, is found in the modern period.

In addition, MSLP in the northern reaches of the Southern Ocean (∼55°S) is found to be increasing significantly at least since the 1930s, while in the southern reaches (∼65°S latitudinal band) MSLP is weakly decreasing over the same period. The MSLP increases in the northern reaches are attributed to a general increase in MSLP in the mid-latitudes. By contrast, the decreasing MSLP polewards of 60°S is attributed in part at least to ozone depletion over Antarctica. However, this thesis proposes that the prolonged presence of the circumpolar trough near the Antarctic coast could be partly responsible for the decreasing MSLP trend in the austral summer.

The new marine observations-based SAM index generated in Chapter 3 correlates with existing SAM indices to varying degrees during the period of investigation (1930-1960). The differences occur between these SAM indices due largely to the data and methods used to generate them. The trend analysis reveals that, out of all the SAM indices considered, the SAM index generated using the full ChriSal-ICOADS merged dataset has a significant positive trend over the 1930-1960 period. However, the trend is not as strong as the positive trend found in the Marshall SAM index in the post-1957 period but nonetheless lends credence to the proposal that the SAM has been positive at least since the 1930s. Even so, more data are required to strengthen this claim.
In Chapter 4, historical meteorological observations from whaling ships’ logbooks (Teleti et al. 2019, Chapter Two in this thesis) are used to identify cyclones during the 1930s and 1950s. Two methods are proposed to capture different aspects of the influence of cyclones on the local weather recorded through meteorological observations. The first method uses the relative deepening rate of pressure observations to detect the relative approach of cyclones to whaling ships. The monthly number of encounters in the 1930s and 1950s decades could not be compared directly with each other, due to a number of factors such as the different numbers of ship-days recorded and the distance/area covered for that month by whaling vessels.

The second method uses a linear multi-regression model to identify and track individual cyclones in space and time across the Southern Ocean. The probability of identifying a cyclone at a given distance from pressure and wind observations on-board a whaling ship is enhanced when multiple ships are close by. On average, each week, two new cyclones and one previously identified cyclone are identified. Overall, the average number of cyclones identified in the 1930s and 1950s is statistically lower than the average calculated using modern period (1999-2008) data over a control area (in the Weddell Sea). Based on the available data, it can be concluded that the modern cyclonic frequency is higher than the historical frequency. This assessment is supported by previous workers, who identified a positive trend in the cyclone frequency during summer (Grieger et al. 2018).

The historical meteorological whaling log observations (Teleti et al. 2019, Chapter 2 in this thesis) are successfully assimilated into two multi-ensemble century-long numerical climate reanalyses (20CRv2c and CERA-20C) in Chapter 5. The results of assimilation (sensitivity) of new whaling observations into reanalyses depend on the background state and inherent coupling in each reanalysis. For example, CERA-20C retains a coupled wave-like impact structure, albeit confined, whereas 20CRv2c mostly reinforces existing weather features post-assimilation. Further, it has been found that, because CERA-20C is relatively more uncertain, CERA-20C benefits more from new-data assimilation and shows a stronger influence of new observations on its MSLP fields than 20CRv2c. In addition, the effects of data assimilation in the 1950s are more subtle because the climate fields are better known in the 1950s as compared to the 1930s.

The computationally inexpensive offline Data Assimilation methods serve a two-fold purpose. The first is in providing the reanalyses community with a system to contextually evaluate historical data suitability, alongside statistical testing of would-be assimilated observations for inclusion into reanalyses without running computationally intensive assimilation
6.3 Limitations, future work and perspectives

The evaluation of MSLP changes over the Southern Ocean would have been enhanced by observations in the non-summer months. In addition, the whaling dataset created from the Christian Salvesen logs is biased towards the Weddell Sea and ABS sectors of the Southern Ocean. These spatio-temporal constraints are the consequence of weather conditions and the whalers’ preference for certain whaling grounds. Observational gaps in other sectors in the Southern Ocean could be filled with further whaling data extraction, but gaps in the non-summer months are likely to remain in future datasets because whaling did not take place during the Antarctic winter. As the RECLAIM project has identified, a large cache of largely unknown Christian Salvesen Whaling Co. logbooks is present at the Sea Mammal Research Unit (SMRU), University of St. Andrew’s, Scotland (Wilkinson and Wilkinson 2018), which would be an obvious target for future work. Data extracted from these logbooks kept at St. Andrew’s, if added to the dataset presented in this thesis, can close substantial gaps in our knowledge of the Southern Ocean climate in the pre-satellite period. A further collection of whaling logbooks at the Vestfold Archives in Sandefjord, Norway, has also been identified and photographed, producing more than 30,000 digital images of logbooks, catch books and day reports (Wilkinson 2016b).

Furthermore, to make the data extraction process more efficient, a citizen-science based approach (e.g. as used in the OldWeather, Southern Weather Discovery, Weather Rescue initiatives) could be taken. The process could start by making high-quality photographs of logbooks and relevant historical and contextual information available on the web to members of the public, allowing them to transcribe the observations visually. The process should be interesting and intuitive to keep the volunteer members of the public engaged, leading to a higher throughput of extracted data. In addition, an attempt could be made to automate the
data extraction process, and newer OCR tools based on Artificial Intelligence (AI) should be developed in collaboration with computer-science researchers.

The historical SAM indices generated in this thesis are based on data from less than 20 whaling seasons over a 30 yr period. So too, therefore, are the detected MSLP, and circumpolar troughs semi-annual cycle changes over the Southern Ocean. Stronger arguments could be made on changing climatic patterns if more data were made available; however, the data extracted in this thesis, combined with the existing ICOADS dataset, make up the largest collection of available data for Southern Ocean region and period. Hence, future studies should be directed to investigate the long-term relationship between regional Southern Hemisphere climate and the new SAM index to highlight the role of SAM indexes and tropical indices in understanding the observed climate variability.

An additional line of research could be to compare the new SAM index to historical cyclones recorded in the whaling or similar datasets, given that many previous studies have found that large-scale climate modes (e.g. SAM, ENSO) have an impact on Southern Ocean cyclone intensity and frequency (Pezza et al. 2008, Grieger et al. 2018). In addition, as the circumpolar trough is found to be confined close to the Antarctic coast for longer periods of the year (Sections 3.3.4 and 3.3.5), a comparison should to made to provide quantification of the relationship between the circumpolar trough and cyclones observed in the historical period vis-a-vis the modern period.

It could be postulated that the semi-supervised cyclone detection algorithm developed in Chapter 4 underestimates the number of historical cyclones as the model is built on fewer than five climate parameters. To make the model more robust, a number of innovations could be included in a future study to enhance the accuracy and performance of the semi-supervised cyclone detection algorithm developed here. Additional predictors could be incorporated into the model, given that several studies (e.g. SHIP) have demonstrated a clear link between sea-surface temperatures (SSTs) in the path of a cyclone and the observed intensity of the cyclone. This could be achieved by combining different datasets (e.g. ICOADS) with whaling datasets to estimate SST fields in the path of the cyclone, and even the climatological SSTs of the region may improve model skill.

Furthermore, as cyclones are strongly non-linear processes in terms of pressure-gradients and wind speeds, employing a neural network to model cyclone-meteorological observations could address non-linearity issues faced by the linear models. A neural network based cyclone intensity and track forecast have been shown to out-perform linear regression when using climatological and persistence variables (Baik and Hwang 1998). This is because
more flow-dependent information is extracted from the data in the standard back-propagation neural network as compared to linear regression. Hence, further study should be conducted using neural networks to model cyclone intensity based on SST data in addition to the variables used in this study.

The assimilation method used in Chapter 5 of this thesis, linearly forces the model state towards the historical observations; however, in this process, probabilistic information of the multi-ensemble member states is lost. In addition, at the beginning of each assimilation window, model states fluctuate widely as new information is forced upon it. These challenges could be addressed by two methods. In the first, a Backward-Forward Nudging (BFN; Auroux and Blum 2008) process consisting of a forward nudging step similar to the current method used in Chapter Five could be used, but with an additional backward nudging step where the final state derived from forward step is nudged towards the initial state.

The second method could be the sequential implementation of the Ensemble Kalman Filter called Ensemble Kalman Fitting (EKF; Franke et al. 2017). This method updates the ensemble mean and member anomalies individually without explicitly updating the covariance matrices. Similarly to the method used in Chapter 5, this would reduce the impact of spurious correlations in the background error covariance matrix, applying an ensemble square root filter to restrict the influence of distant observations on climate fields. The extension of current work would be to constrain the multi-ensemble numerical climate reanalyses with the BFN process and compare those results with those obtained from the EKF method with much shorter assimilation window.

This would help the understanding of how these processes combine uncertainties in the observational data with reanalyses. The comparison would point to a data assimilation methodology that can better utilise historical observations to improve the representation of historical atmospheric circulation in climate reanalyses. In addition, the latest version of 20CR reanalysis is v3 (Slivinski et al. 2019), however, at the time of this work it wasn’t publicly released. Hence, similar analysis and assimilation techniques could be performed on the current version, 20CRv3 reanalysis in the future.

In conclusion, this thesis has shown that whaling logbooks can be used to extract valuable meteorological observations from little known areas of the World and points to their large untapped potential for data-rescue efforts and the construction of historical meteorological datasets. The thesis also shows the evolution of Antarctic and Southern Ocean climate since the 1930s to the modern period, but how this climate will evolve in the future is highly dependent on the greenhouse emission path taken from now on. Under the IPCC’s
Representative Concentration Pathway (RCP) 8.5 scenario, by 2070, it is predicted that the global mean surface air temperatures over land would be more than 3.5°C higher than observed in the late 19th century (Intergovernmental Panel on Climate Change 2014). The air temperatures over Antarctica could be warmed by about 3°C, well above the range of centennial natural variations (Intergovernmental Panel on Climate Change 2014, Pol et al. 2014). As the ozone levels recover by mid-21st Century, it will no longer have an effect on the westerlies; thus the observed poleward shift in the westerlies in the summer now will occur throughout the year (Swart et al. 2015).

The scenario also proposes stronger and positive SAM index, enhancement of westerly winds over the Southern Ocean is expected to further isolate Antarctica thereby, preventing the penetration of warm maritime air masses in the interior; so lesser warming in the interior of Antarctica is predicted. An increased cyclone activity and stronger zonal winds due to enhanced positive SAM index during the 21st century in Antarctica (Intergovernmental Panel on Climate Change 2014). With the warmer atmosphere projected for the 21st century, an increase in Antarctic precipitation is predicted that averages 0.42 ± 0.01 mm/yr (Uotila et al. 2007). The Southern Ocean is expected to continue warming, freshening, and acidification (Rintoul et al. 2018, Swart et al. 2018). Mass loss from the West and East Antarctic Ice Sheets will accelerate, contributing to rising sea levels (Rintoul et al. 2018).

The anthropogenic effects were thought to be difficult to detect in the Southern Ocean due to brevity of climate records (Jones et al. 2016) but, the investigations like Parkinson (2019), Rintoul et al. (2018) and Swart et al. (2018) have started to recognise them. Studies like this thesis are also shedding light on the long-term changes in the Southern Ocean; hence, the global nature of the anthropogenic effects is clear. Rapid and substantial steps should be taken to reduce or at least stabilise the greenhouse emissions before the climate tipping points are reached.
References


References


References


References


References


References


References


Appendix A

Supplementary Information

Supplementary Figures

Fig. A.1 Outliers detected from the Southern Harvester ship for the 1956-57 whaling season. The green band denotes first and last 20% of the season time, as it shows most of the outliers are either in the first or last 20% of the season time.
Fig. A.2 A polar stereographic projection (south of 30°S) showing the Southern Ocean, thick red lines indicate 40°S, 50°S and 65°S latitude bands. The SAM index is calculated as normalised MSLP difference between 40°S and 65°S latitude bands; Antarctic and mid-latitude stations used by Marshall (2003) to generate SAM index are shown as squares and circles, respectively. The 50°S and 65°S latitude bands denote extremes in the movement of circumpolar trough, the SAO cycle. Gray patch in the vicinity of 65°S 150°W shows approximate position of Amundsen Sea Low (ASL; Turner et al. 2013).
Fig. A.3 The MSLP contours from 56-ensemble members (thin blue lines) and ensemble mean (thick black lines) at 13-Dec-1957 12:00:00 in 20CRv2c reanalysis, before (a) and after (b) assimilating whaling ships’ observations. The thin blues lines indicate ensemble contours and the contours where ensemble spread (standard deviations) is less than 3 hPa are indicated by thick black lines. Yellow dots represent observations present in the dataset and red dots represent newly added whaling dataset observations.
Fig. A.4 The MSLP contours from 10-ensemble members (thin blue lines) and ensemble mean (thick black lines) at 13-Dec-1957 12:00:00 in CERA-20C reanalysis, before (a) and after (b) assimilating whaling ships’ observations. The thin blues lines indicate ensemble contours and the contours where ensemble spread (standard deviations) is less than 3 hPa are indicated by thick black lines. The existing observations (yellow dots) used in the reanalysis were not provided hence not shown here. Red dots represent newly added whaling dataset observations.
Fig. A.5 A sample composite of whaling season 1935-36 showing two weeks of data in each sub-plot. © and ©R shows a new cyclone and a repeated cyclone respectively.