WIND-WAVE EXTREMES IN A CHANGING CLIMATE FROM ATMOSPHERE AND WAVE MODEL ENSEMBLES

by

Alberto Meucci
https://orcid.org/0000-0002-0278-756X

In fulfillment of the requirements for the degree of

Doctor of Philosophy

©January 2020

Department of Infrastructure Engineering
The University of Melbourne

Declaration of Authorship

This is to certify that:

- this thesis comprises only my original work except where indicated,
- due acknowledgment has been made in the text to all other material used,
- this thesis is less than 100 000 words in length, exclusive of tables, bibliographies and appendices.

Alberto Meucci, January 2020

Abstract

Marine offshore structures and operations, as well as coastline defences, rely on accurate estimates of the design sea state, defined as the maximum significant wave height which can be expected over an N year period. To find design sea state estimates, Extreme Value Analysis (EVA) statistical approaches are commonly used. However, due to the paucity of data available for extreme sea states, and the consequent challenges in modelling such phenomena, design sea state estimates are characterized by large confidence limits, which further complicate strategies and policies for resilient design of marine structures and operations. Furthermore, an additional challenge is posed by a changing climate, which introduces further uncertainties in the prediction of future design sea states.

The present work deals with the EVA statistical uncertainties of ocean surface wind speed and significant wave height, developing and testing a novel ensemble approach to EVA. It then applies this novel approach to estimate future changes in extreme significant wave height by the end of the 21st Century. In this way, the present thesis finds, for the first time, statistically significant changes in extreme wave height in distributed regions of the ocean. Such an approach has the potential to be further refined with higher resolution models and new climate model projections, which would significantly improve global estimates both from past and future model ensembles.

In parallel, this work finds an inconsistency in the trends of wind and wave mean climate throughout the 20th Century from climate models and reanalyses, questioning our confidence level as to a possible climate change signal for ocean surface wind speed and wave height. However, the work finds general agreement between datasets and statistical approaches, for a Southern Ocean wind speed and wave height increase over the last part of the 20th Century, that is also projected to further increase by the end of the 21st Century.

Preface

This thesis is mainly composed of three research studies performed during my PhD. Two of these have been published in the Journal of Climate, and appear in Chapter 3 and 5 of the present thesis (©American Meteorological Society. Used with permission.). A third study was published in Science Advances and appears in Chapter 7 of the present work.

The first study (*Meucci et al.*, 2018) (©American Meteorological Society. Used with permission.), Chapter 5, was performed in collaboration with Prof. Øyvind Breivik from the University of Bergen, and the Norwegian Meteorological Institute in Bergen, Norway. This study was funded by the Australian Research Council through Grants DP130100215 and DP160100738, and by the Research Council of Norway through the project ExWaMar (Grant 256466) and supported by the European Research Area for Climate Services through the ERA-4CS project WINDSURFER.

The second study (*Meucci et al.*, 2020a) (©American Meteorological Society. Used with permission.), Chapter 3, was developed during a visiting period in Bergen, Norway, in collaboration with Dr Ole Johan Aarnes and Prof. Øyvind Breivik, with the support of the Integrated Marine Observing System (IMOS) in compiling and archiving the altimeter dataset, the EU programme ERA4CS through the WINDSURFER project, and the Research Council of Norway through the ExWaMar project (Grant 256466).

For the third study (*Meucci et al.*, 2020b), Chapter 7, I conceived the idea under the guidance of my supervisor Kernot Prof. Ian R. Young, and performed the model simulations and statistical analysis, receiving inputs from Dr Mark Hemer, from the CSIRO Oceans and Atmosphere, Hobart, Australia, and Prof. Roshanka Ranasinghe of the Department of Water Science and Engineering, IHE-Delft, the Harbour, Coastal and Offshore Engineering, Deltares, and the Water Engineering and Management, Faculty of Engineering Technology, University of Twente, Enschede, The Netherlands. My colleague, and soon PhD, Ebru Kirezci performed the global coastlines analysis of Fig. 7.18. The study was supported by the Australian Research Council through grant DP160100738, and by the Australian Commonwealth National Environmental Science Program Earth Systems

and Climate Change Hub.

The rest of the thesis chapters (Chapters 1, 2, 4, 6, 8) introduce and describe the theoretical aspects and the motivations behind the present work, and link together the three journal articles.

Acknowledgements

I wish to acknowledge and pay respect to the Wurundjeri people, and their Elders past and present, as the traditional custodians of the land on which the University of Melbourne is located, and where I lived and enjoyed my PhD years. I wish also to acknowledge the Melbourne Research Scholarship through which this PhD was funded.

This work is the product of three fruitful years in Melbourne, where I significantly grew both professionally and personally. This was possible thanks to the interactions I had with so many inspiring people.

First and foremost, I would like to express my sincere gratitude to my advisor Kernot Professor Ian Robert Young for the continuous support, for his patience, motivation, immense knowledge, and all the interesting discussions. His guidance helped me to sail smoothly through my research, and gave me the possibility to experience what world-class research means.

I wish also to thank the rest of my thesis committee, Prof. Alexander Babanin and A/Prof. Alessandro Toffoli, for their insightful comments and encouragement, and for the opportunity they gave me to teach in ocean engineering. This helped me to structure and deepen my knowledge, and to find new ways of communicating my research.

A special thanks goes to Prof. Øyvind Breivik, who kindly hosted me at the Norwegian Meteorological Institute and helped me through an important part of my PhD studies, and to Dr. Ole Johan Aarnes, with whom I had great discussions on statistics and wave heights. I wish also to thank Dr. Mark Hemer for his collaboration, his knowledge and guidance on future climate projections, my colleague Ebru Kirezci for the collaboration and support, and Prof. Roshanka Ranasinghe for contributing to my research with his international experience. Thank you also to my past and present colleagues with whom I had a lot of inspiring discussions that improved my research and sparked my motivation. Without all of this precious professional interactions and support it would not have been possible to conduct this research.

Throughout these years a huge support came from my friends in Australia and back in my home country with whom I walked together and shared joy and pain.

Some believed in me from the beginning and I will never forget the strength that their words gave me.

I dedicate this thesis, and all the publications that stemmed from it, to my family. Their support was invaluable and made this unique experience possible. A special thank goes to my brother. All our phone calls kept me motivated and pushed me to do my best.

Contents

Declaration of Authorship ii

Abstract iv

Preface vi

Acknowledgements viii

List of Tables xiv

List of Figures xvi

1 Introduction 2
- 1.1 Ocean surface wind and wave extremes . . . 2
- 1.2 Research objective . . . 5
- 1.3 Thesis outline . . . 6

2 Wind-wave climate and extremes from model and observational datasets 10
- 2.1 Introduction . . . 10
- 2.2 Towards fully coupled atmosphere-ocean-wave models . . . 12
  - 2.2.1 Coupled atmosphere-wave models . . . 14
- 2.3 Atmosphere-wave model sources of uncertainties . . . 15
  - 2.3.1 Modelling ocean surface wind . . . 15
  - 2.3.2 Modelling the waves . . . 17
  - 2.3.3 Modelling extreme sea states . . . 19
- 2.4 Model assimilation of observations . . . 21
- 2.5 Ocean observation uncertainties . . . 22
- 2.6 The ensemble probabilistic approach . . . 24
- 2.7 Conclusions . . . 25

3 Trends from 20th century models compared to satellite altimeter observations 28
3.1 Introduction/Summary . . . 28
3.2 Background . . . 29
3.3 Datasets and methodology . . . 31
3.3.1 ERA-20CM . . . 32
3.3.2 ERA-20C . . . 33
3.3.3 CERA-20C . . . 33
3.3.4 Observation Feedback Archive (OFA) . . . 34
3.3.5 Satellite altimeters . . . 37
3.3.6 Trend analysis . . . 37
3.4 Century-long trends . . . 38
3.4.1 +24 h forecast lead time trends . . . 39
3.5 Twentieth century climate . . . 42
3.6 Model, reanalysis and satellite trends . . . 43
3.7 Discussion . . . 45
3.8 Conclusions . . . 47

4 Wind and wave Extreme Value Analysis 50
4.1 Introduction . . . 50
4.2 Univariate statistical analysis . . . 52
4.2.1 Basic assumptions . . . 53
4.2.2 Statistical modelling . . . 53
4.2.3 Confidence intervals . . . 54
4.3 Extreme Value Analysis . . . 55
4.3.1 Asymptotic models . . . 56
4.3.2 Threshold models . . . 57
4.4 Wind and wave Extreme Value Analysis (EVA) . . . 58
4.5 Conclusions . . . 60

5 Wind and Wave Extremes from Atmosphere and Wave Model Ensemble forecasts 62
5.1 Introduction . . . 63
5.2 Estimation of wind and wave extremes . . . 66
5.3 Wind and wave extremes from ensemble model predictions . . . 69
5.4 Methodology criteria . . . 71
5.5 Extreme value analysis estimates . . . 85
5.5.1 ENS direct return estimates . . . 86
5.5.2 ENS Initial Distribution Method . . . 90
5.5.3 ERA-Interim POT method . . . 92
5.5.4 In situ measurements—NDBC buoys . . . 94

5.6 Discussion . . . 96
5.7 Conclusions . . . 98

6 Future projections of wind-wave extremes 100
6.1 Introduction . . . 100
6.2 Atmospheric-Ocean Global Climate Models (GCM) . . . 101
6.3 Future wind-wave projections . . . 103
6.3.1 Sources of uncertainty in wind-wave projections . . . 104
6.4 Future wind-wave extreme projections . . . 104
6.5 Conclusion . . . 105

7 Projected 21st Century changes in extreme wind-wave events 108
7.1 Introduction/Summary . . . 108
7.2 Background . . . 108
7.3 Dataset and Methodology . . . 110
7.3.1 EVA technique . . . 111
7.3.2 Preliminary tests on raw model data . . . 114
7.3.3 Robustness of the results . . . 116
7.3.4 Model contributions . . . 121
7.3.5 Statistical significance . . . 121
7.4 Results and findings . . . 125
7.4.1 Model Ensemble . . . 125
7.4.2 Extreme value projections . . . 129
7.4.3 Uncertainty assessment . . . 132
7.4.4 Change in frequency of extremes . . . 133
7.4.5 Tropical Cyclone generated wind-wave extremes . . . 133
7.4.6 Change along global coastlines . . . 137
7.5 Discussion and Conclusions . . . 138

8 Conclusions and Future Research 140
8.1 Future work . . . 141

Acronyms 144

Bibliography 148

List of Tables

3.1 Characteristics of the ECMWF century-long models. 32

3.2 ICOADS 2.5 dataset. Surface pressure and wind speed observation counts classified by report type ID as archived in the Observation Feedback Archive (OFA) of the ECMWF (*Hersbach et al.*, 2015b). . 34

5.1 Ensemble approaches applied throughout the next part of the thesis. The EVA ensemble approach was first tested on a single operational forecast model from European Centre for Medium-Range Weather Forecasts (ECMWF). The method was then applied to an inter-model ensemble constructed by forcing the same WaveWatch-III (WWIII) model with different Global Circulation/Climate Model (GCM) surface wind fields. 62

5.2 H_s Centered anomaly correlation, r_{ACC} as in Eq. (5.1), for the three test locations (Fig. 5.2) for day 9 and day 10 lead times. The comparison shows higher correlation at shorter lead times and tests the two bounds (+216 h and +240 h) of the 5 lead times used in the present work 74

5.3 H_s Actual Correlation (Pearson) coefficient for comparison with the Centered Anomaly Correlation for the same test locations and lead times as in Table 5.2 74

5.4 Comparison of normalized monthly average differences between the ENS forecasts and satellite (SAT) measurements (*Young et al.*, 2017). The maximum normalized differences $\Delta\overline{H}_s/\overline{H}_s^{SAT}$ are shown along with the month in which they occur. 78

5.5 Comparison of normalized monthly average differences between the ENS forecast data set and ERA-Interim. The maximum normalized differences $\Delta\overline{H}_s/\overline{H}_s^{ERAI}$ are shown along with the month in which they occur. 82

5.6 DRE comparison with EVA of buoy data measurements averaged over the synoptic times. EVA of 100-year return values of H_s^{100} are: A - buoys peaks over 90% threshold fitted with Generalized Pareto distribution; B - ENS direct return estimates from 1000 ranked peaks. (B-A)/A - differences normalized on the buoy estimates. ENS DRE 95% bootstrap confidence intervals are reported as reference for the comparison. 95
5.7 DRE comparison with EVA of buoy data measurements averaged over the synoptic times. EVA of 100-year return values of 10 m wind speed (U_{10N}^{100}) are: A - buoys peaks over 90% threshold fitted with Generalized Pareto distribution; B - ENS direct return estimates from 1000 ranked peaks. (B-A)/A - differences normalized by the buoy estimates. ENS DRE 95% bootstrap confidence intervals are reported as reference for the comparison. 96

7.1 Ensemble of Climate Models forcing winds and wave model characteristics. 111
7.2 The percentage change in 100-year significant wave height along coastline by 2100 for both RCP4.5 and 8.5. 138

List of Figures

2.1 Operational forecast coupled atmosphere, ocean, and wave models, as implemented by ECMWF from June 2018 - Integrated Forecasting System (IFS) Cy45r1. 13

3.1 1901–2010 ICOADS 2.5 surface wind and surface pressure observation yearly counts divided by report type, as archived in the ECMWF Observational Feedback Archive. Refer to Table 3.2 for the description of report type ID. 35
3.2 1901–2010 OFA observation counts in base 10 logarithmic scale, binned on a $2^\circ \times 2^\circ$ grid. The three different datasets assimilated in ERA-20C and CERA-20C are: (a) ICOADS2.5 surface wind (b) ICOADS2.5 surface pressure (c) ISPD2.2 surface pressure. 36
3.3 1901–2010 percentage-per-decade trends. Comparison between the three twentieth century ECMWF datasets for H_s (left column) and U_{10} (right column). (a,b) ERA-20CM, (c,d) ERA-20C, (e,f) CERA-20C. Only the 95% significant trends from the SKTT are plotted in the figure. White regions represent areas where the trend is not statistically significant. Sea-ice cover areas that limit the H_s domain are hatched. 39
3.4 Difference in CERA-20C 1901–2010 percentage per decade trends between analysis time ANA and forecast lead time +24 h FC24. ANA - FC24: (a) H_s (b) U_{10}. Sea-ice cover areas that limit the H_s domain are hatched. 40
3.5 Yearly averages of monthly mean values at four different locations indicated in Fig. 3.4b. For ERA-20CM and CERA-20C the envelope of the ensemble values are plotted together with the mean. Column I: ANA H_s. Column II: FC24 H_s. Column III: ANA U_{10}. Column IV: FC24 U_{10}. 41
3.6 Annual averages of monthly mean values. Differences between ERA-20CM and CERA-20C (Eq. 3.2). H_s (left column) and U_{10} (right column). (a,b) 1901–2010, (c,d) 1901–1930, (e,f) 1985–2010. Sea-ice cover areas that limit the H_s domain are hatched. 42

3.7 1985–2010 percentage-per-decade trends. Comparison between altimeter dataset trends ($2° \times 2°$) (*Young and Ribal*, 2019; *Ribal and Young*, 2019), ERA-20CM and CERA-20C at analysis time ($1.5° \times 1.5°$). Left column (a,c,e) H_s trends. Right column (b,d,f) U_{10} trends. Areas of SKTT 95% statistical significance level are dotted. Sea-ice cover areas are hatched. 44

4.1 Target objective of the present thesis work following a scheme adapted from *Field et al.* (2012). 51

5.1 Procedure for advanced lead times ensemble forecasts pooling. Construction of the synthesized dataset to find Direct Return Estimates out of the 1000 highest peaks. (a) Each ensemble forecast diverges with time. (b) For each ensemble member, the maximum value is selected from lead times +216 h to +240 h. (c) The selected maximum values are representative of a 30 h period. Each period is independent, as shown by the vertical dotted lines. (d) The selected maximum values represent independent realizations of extreme wind speed/wave height and are pooled for the EVA. 70

5.2 Locations used for testing the methodology criteria: P40 - *Ekofisk Oil Field* WMO code LF5U (56.50° N, 3.20° E); P35—*Heidrun Oil Field*—WMO code LF3N (65.30° N, 7.30° E); B16—*K5 buoy*, WMO code 64045 (59.10° N, 11.40° W). 73

5.3 Centered anomaly correlation: U_{10N} for ENS member 1 and ENS member 50 at +216 h lead time from the 2010-2016 ENS data set. The highest correlations are found at the Horn of Africa, due to the regularity of the Somali jet and Monsoon events in this region. Tropical areas also show higher values than higher latitudes. 73

5.4 Centered Anomaly Correlation: H_s for ENS member 1 and ENS member 50 at +216 h lead time from the 2010-2016 ENS data set. The values show higher correlation in tropical areas which are more swell dominated. 75

5.5 Quantile-Quantile plots of ensemble member 1 against ensemble member 50 at +216 h lead time for the test locations: (a,b) P40 North Sea, P35 (c,d) and B16 (e,f) in the North Atlantic. The left column (a,c,e) show H_s [m] while the right column (b,d,f) shows U_{10N}. The Q-Q result is shown by the red solid line, which closely approximates the dashed diagonal, indicating a similar PDF. Also shown is a scatter plot of individual predictions (dots). 76

5.6 Test locations over the global dataset domain. 77

5.7 Comparison of $\overline{H}_s$ (left) and $\overline{U}_{10}$ (right) monthly averages between the ENS at +240 h lead time and SATellites for the test locations: (a,b) P40 North Sea, (c,d) P35 North Atlantic, (e,f) B16 North Atlantic. 79

5.8 Quantile-Quantile plots of ensemble member 1 at +216 h lead time and ERA-Interim for the testing locations: (a,b) P40 North Sea, (c,d) P35 North Atlantic, (e,f) B16 North Atlantic. On the left side the comparison for H_{s} (a,c,e) while on the right side for $U_{10\mathrm{N}}$ (b,d,f). 80

5.9 Comparison of $H_{s,90}$ (left) and $U_{10,90}$ (right) monthly 90$^{\mathrm{th}}$ percentiles between the ENS at +240 h lead time and SATellites for the test locations: (a,b) P40 North Sea, (c,d) P35 North Atlantic, (e,f) B16 North Atlantic. 81

5.10 Comparison of $H_{s,90}$ (left) and $U_{10,90}$ (right) monthly 90$^{\mathrm{th}}$ percentiles between the ENS at +240 h lead time and SATellites for the test locations in the Northern Hemisphere (NH): (a,b) NWPO1 North West Pacific Ocean, (c,d) NAO3 North Atlantic Ocean, (e,f) NIO1 North Indian Ocean. (codes refer to test locations in Fig. 5.6). . . . 83

5.11 Comparison of $H_{s,90}$ (left) and $U_{10,90}$ (right) monthly 90$^{\mathrm{th}}$ percentiles between the ENS at +240 h lead time and SATellites for the test locations in the Southern Hemisphere (SH): (a,b) SO1 Southern Ocean, (c,d) SEPO1 South East Pacific Ocean, (e,f) SAO3 South Atlantic Ocean (codes refer to test locations in Fig. 5.6). 84

5.12 ENS 2010-2016 Direct Return Estimates of 100 year Significant Wave Height H_{s}^{100} with Δt = 6 h × 5 = 30 h. The estimates are taken from a reconstructed time equivalent period of 750 years. 87

5.13 ENS 2010-2016 Direct Return Estimates of 100 year wind speed $U_{10\mathrm{N}}^{100}$ with Δt = 6h x 5 = 30h. 750 years time equivalent period. . . 88

5.14 Width of ENS 2010-2016 Direct Return Estimate 95 confidence intervals for 100 year Significant Wave Height H_{s}^{100}. These are computed from 500 bootstraps of the 1000 highest peaks selected from the maximum of 5 different forecast lead times. 89

5.15 Width of ENS 2010-2016 Direct Return Estimate 95 confidence intervals for 100 year wind speed $U_{10\mathrm{N}}^{100}$. These are computed from 500 bootstraps of the 1000 highest peaks selected from the maximum of 5 different forecast lead times. 90

5.16 ENS 2010-2016 100 years return value estimates of H_{s} from the Initial Distribution Method fitted with a Gumbel distribution. Values extracted from ENS#1 at +222h lead time and ENS#50 at +240h. 91

5.17 ENS 2010-2016 100 years return value estimates of U_{10N} from the Initial Distribution Method fitted with a Weibull 2 Parameter distribution. Values extracted from ENS#1 at +222h lead time and ENS#50 at +240h. 92
5.18 ERA-Interim 1979-2017 Peaks Over Threshold 100 years RP Significant Wave Height H_s^{100} - 99.8% threshold, fitted with an exponential distribution. 93
5.19 ERA-I 1979-2017 Peaks Over Threshold 100 years RP wind speed U_{10N}^{100} - 99.8% threshold, fitted with an exponential distribution. . . . 93
5.20 NDBC buoys that meet the requirements for EVA comparison to ENS model forecasts 94

7.1 Mean, $\overline{H}_s$, and 99th percentile, H_s^{99}, significant wave height obtained for the present 1979–2005 ensemble. In the lower section: differences in mean and 99th percentile significant wave height between the historical (1979–2005) and future (2081–2100) time periods for RCP4.5 and 8.5. The grey area are the locations outside of the ensemble domain. 112
7.2 Schematic of the ocean storms pooling method. **a**, extreme wave heights are selected from each CMIP5 model as values over the 90th percentile threshold with a minimum time distance of 48 h (storm independent peaks). The peaks are then ranked and the highest 1000 form the tail of the dataset to which an exponential function is fitted. The stacked histograms show the 1000 peaks, with contributions from all 7 models (Section 7.3). **b, c** show stacked histograms with bias correction and **d, e** with bias correction. The histograms of relative contributions from each model show a more uniform distribution with bias correction. 113
7.3 Scatter and Quantile-Quantile (Q-Q) plots at three selected locations (positions marked on plots) for combinations of different pairs of CMIP5 models (no bias correction). Large scatter indicates the models are independent, similar Q-Q results indicate they are identically distributed. 115
7.4 Test for similarity between distribution of extremes — **a**, historical dataset (1979–2005), **b**, end of 21st Century (2081–2100). Root Mean Square Error (RMSE) analysis of the 90th percentile extreme quantiles for each model compared to the inter-model ensemble extreme quantiles (see Section 7.3). 117
7.5 1979–2005 100-year return period significant wave height, derived from Peaks over 99.6th percentile Threshold of each global wave model run. Results obtained without bias correction. 118

7.6 Changes in the 100-year return period significant wave height derived from 1000 data peaks at the end of the 21st Century — RCP8.5 (see Section 7.3). **a**, Results obtained without bias correction, **b**, Results obtained with bias correction applied at the 99.8^{th} percentile level. 119

7.7 Significant wave height return period analysis. The extreme estimates are derived at different return levels (1 in 10, 20, 50, 100-year return period (RP)) from the ensemble 1000 highest peaks, fitted with an exponential distribution. 120

7.8 The relative contributions over individual models to the ensemble of peaks used for the EVA. Selected locations shown around the world for **a** historical (1979–2005) and **b** future projection RCP8.5 (2081–2100) non-bias-corrected datasets. 122

7.9 Bootstrap estimates of the 95% confidence intervals of the 100-year return period significant wave height for **a**, historical dataset 1979–2005, **b**, end of 21_{st} Century RCP8.5. 123

7.10 Confidence in the historical and future projection estimates of H_{s}^{100} at selected locations, and changes in standardized variable probability distributions for the surrounding 5°x5° regions. **a**, **d** and **g**, compare the confidence intervals of the 7 models ensemble Z standardized variable (see Section 7.3) with each model Peak over Threshold approaches at comparable threshold percentile, and the average of each model EVA. **b**, **e** and **h** compare the histograms of the historical inter-model ensemble of Z standardized variables with the future projections, focusing on the highest peaks tail. **c**, **f** and **i** are Quantile-Quantile (Q-Q) plots comparing the historical inter-model Z ensemble and future projections. The differences between the historical and future projections in the tail of the PDF illustrate projected changes in the frequency and magnitude of extreme wave heights. Note that **a**, **d** and **g**, are analyses at a grid point, whereas, **b**, **c**, **e**, **f**, **h** and **i** are analyses over 5°x5° regions surrounding the selected locations. 124

7.11 Changes in the 100-year return period significant wave height at the end of the 21^{st} Century — RCP8.5, derived from Peaks over 99.6^{th} percentile Threshold of each global wave model run. Results obtained without bias correction. 125

7.12 Magnitude of the 100-year return period significant wave height H_s^{100} [m] resulting from three different dataset EVAs. **a**, 1979–2005 inter-model ensemble of the bias-corrected Z standardised variable highest peaks (see Section 7.3), pooled from global wave model runs forced with 7 GCMs surface winds, **b**, 1985–2018 calibrated altimeter dataset using the Peaks over Threshold approach with exponential distribution fit, **c**, 1979–2005 Peaks over 99.6$^{\text{th}}$ percentile Threshold for the single global wave model run forced with a reference NOAA CFSR wind speed (exponential distribution fit). 127

7.13 100-year return period significant wave height differences between standardized variable corrected inter-model ensemble technique and a single CFSR-forced model analysed using a Peaks over 99.6$^{\text{th}}$ percentile Threshold, Extreme Value Analyses. 128

7.14 95% Confidence intervals of the 100-year return period significant wave height, H_s^{100}, historical time slice. **a**, bootstrap estimates from the inter-model ensemble Extreme Value Analysis, **b**, Peaks over Threshold exponential fit confidence intervals of the CFSR-forced single model extreme value estimates. This figure shows the difference of the confidence interval width between the ensemble technique applied in the main manuscript, and the single model Peak Over Threshold confidence intervals. The possibility to synthetize a dataset of 189 years of data significantly increases the confidence around the H_s^{100} estimates. 130

7.15 Percentage change in the 100-year return period significant wave height by the end of the 21$^{\text{st}}$ Century relative to the 1979–2005 period. Standardized variable analysis approach is used for the intervals 1979–2005 and 2081–2100. **a**, RCP4.5 mid emission scenario. **b**, RCP8.5 high emission scenario. The regions with statistically significant changes at 5% level (see Section 7.3, Eq.7.7) are hatched. 131

7.16 Changes in extreme event frequency. **a**, changes in the number of extreme events per year over the historical 90$^{\text{th}}$ percentile threshold between the historical dataset 1979–2005 and future projection 2081–2100 for RCP8.5. **b**, as for **a** but for the historical 99.6$^{\text{th}}$ percentile threshold. The similar spatial variability demonstrates consistency of the results for different selected thresholds. 134

7.17 Comparison of the bias corrected Z-ensemble wave height with NDBC buoy and satellite altimeter measurements. Quantile-Quantile (Q-Q) plots at four selected locations (positions marked on top map). **a**, **b**, TC locations where the ensemble underestimates extremes. **c**, **d**, extra-TC regions where the ensemble performs well against buoy and satellite measurements, with slightly higher extremes. 135
7.18 Percentage change in 100-year extreme value significant wave height along the global coastline between the historical dataset 1979–2005 and future projection 2081–2100 for RCP8.5. 137

Chapter 1

Introduction

The inherent inability to precisely define how random natural processes unfold, drives human interest in statistics. Data analysis uses statistics to make sense of available information and derive useful insights to apply to every day life. Descriptive and inferential statistics are essential tools for engineers and technicians to understand nature randomness. Some natural processes are more complicated than others, and in these cases an accurate assessment may be unachievable. Mathematical theories explain the evolution of the main driving forces, but the complete understanding of a process may be missing, significantly challenging human reasoning. This is particularly true for extreme events that are characterized by a low level of predictability. These events affect human activities and may cause significant damage and losses of lives. Extremes are usually difficult to measure, and the assessment of past, present and projected future events is complex. A more intricate picture emerges if we consider meteorological-ocean (metocean) extremes. The present thesis focuses on improving the confidence level over the prediction of ocean surface wind speed and wave height extremes. It does so developing innovative ensemble EVA techniques that consider past and future climate changes. The following sections provide an introduction to ocean surface wind and wave extremes, the challenges in their estimation, and the pressing research questions. It leads then to the present research thesis motivation and objective.

1.1 Ocean surface wind and wave extremes

Wind-wave climate

In the present work we are specifically interested in wind-waves, which are the waves generated by the action of the wind over the ocean surface. Wind duration and fetch—the length of the ocean surface over which the wind blows—influence the generation of waves. The ocean surface dynamically interacts with the wind,

affecting wind drag and the wave generation process (*Young*, 1999b).

Wind-waves play a fundamental role in the atmosphere-ocean system. They drive currents, enhance air-sea flux exchange, impact trajectories of floating objects, and interact with human activities and operations at sea (*Holthuijsen*, 2007; *Veritas*, 2010; *Babanin et al.*, 2012). The global ocean wind-wave climate is characterized by seasonality. The largest seasonal variations in mean wave height and wind speed are found at high latitudes in the Northern Hemisphere (*Young*, 1999a). Tropical areas have generally moderate surface wind speed and wave height, except for isolated extreme events such as tropical cyclones, typhoons or hurricanes (*Holthuijsen et al.*, 2012). In contrast to the Northern Hemisphere, the Southern Ocean has a smaller seasonal variation (summer to winter) (*Young and Holland*, 1996; *Young*, 1999a) Here, regularly occurring westerly winds blow approximately every 2-3 days over the water surface with unlimited fetches, continuously generating large wave heights (*Babanin et al.*, 2019).

In the Arctic and Antarctic, wind-waves and swells—waves generated from a distant storm that propagated across the ocean—, affect the marginal ice-zones and the evolution of ice cover, and may further exacerbate climate change impacts on these fragile ecosystems (*Overeem et al.*, 2011; *Thomson and Rogers*, 2014). Wind-waves and swell also affect world coastlines, playing a significant role on beach erosion and inundation levels (*Goodwin et al.*, 2006; *Slott et al.*, 2006; *Marcos et al.*, 2019).

Modelled and observed wind-wave time series enable us, with ever increasing confidence, to robustly estimate the average sea state, comprised of wind waves and swell. The challenge arises if we are to determine a design sea state estimate for unusual or rare extreme events at sea.

Wind-wave extremes

Estimates of ocean extremes are characterized by a low level of confidence (*Coles et al.*, 2001; *Caires and Sterl*, 2005; *Holthuijsen*, 2007; *Goda*, 2010; *Vinoth and Young*, 2011; *Takbash et al.*, 2018). This is due to the paucity of observations available (both length of time series and spatial resolution), and the inability of current models to fully resolve the complexity of extreme sea states. In meteorological and ocean sciences, the statistical analyses of extremes events has become a high priority, as the severity of extreme events may cause major damage with loss of human lives. Consequently, there is a great need to improve the predictability of ocean extreme events, as they are crucial for the safe design, management and development of both coastal and offshore structures.

The statistical investigation of extremes has its origins in the early 20^{th} century. Scientific interest in the topic sparked in the 1950s, introducing the discipline today known as Extreme Value Analysis, EVA (*Coles et al.*, 2001). EVA

is the dominant tool used to determine extreme estimates or the probability of exceedance of specific design values. EVA deals with the randomness of natural extreme events focusing on the extreme deviations from the median of the probability distribution; that is, the tail of the probability distribution of the dataset. Due to the relatively few values generally defining the tail of the distribution (limited observed extremes) and the limited duration of observations, extrapolation of the probability distribution to the desired probability level—one in N years event—results in low confidence in the estimate (*Holthuijsen*, 2007; *Goda*, 2010). Nevertheless, EVA has essential practical value for engineering applications, and relevant social and economic importance.

EVA can be applied to either observational or model datasets. Observations of ocean surface wind speed and wave height come from various sources, these are in-situ and remote sensing measurements. Measuring natural extremes is a challenging task, particularly for ocean wind and waves. Limitations in the observational datasets further complicate the analysis of extremes. Only relatively short time series of observations are typically available to perform an EVA, and the return period sought is in most cases larger than the length of the time series analysed. This means that extrapolation of the probability distribution is almost always required. Models guarantee long and homogeneous time series but have limitations connected to parametrizations and unresolved scales (*Cavaleri et al.*, 2007; *Cavaleri*, 2009). On top of this, models are affected by the limitations of observational datasets used to calibrate them.

A traditional EVA approach, applied to observational or modelled time series, offers just a glimpse into the complicated random processes that characterize extreme sea states. Uncertainties dominate past climate and future prediction of atmosphere and ocean extremes. In a changing climate, the construction and maintenance of marine systems, as well as the management and conservation of world coastlines, need to rely on robust extreme value analyses, capable of supporting resilient and sustainable structural design.

Wind-wave extremes in a changing climate

In the past few years there has been increasing interest in climate change impact on wind-wave. Studies from an extensive dataset of satellites observations found significant changes in the Southern Hemisphere mean surface wind speed over the past 30 years (*Young et al.*, 2011; *Young and Ribal*, 2019). Extreme events may also be affected by climate change, but the changes associated with mean values may not be related to the changes in the extremes; that is, the extremes may vary differently from the average climate (*Wigley*, 2009; *Cooley*, 2009). Deriving physically meaningful wind-wave trends is complex. Current studies show no general consensus on extreme trends over the past century, as well as for future projections

Morim et al. (2018, 2019).

The low confidence in wind-wave EVA estimates impedes understanding of how the magnitude and frequency of wind-wave extremes are changing from historical realizations to future projections. Thus, there is a strong demand for robust analyses of past, present and future wind-wave climate. Innovative approaches will prove to be of paramount importance to support sustainable and resilient design of marine structures, marine operations and coastal defences.

1.2 Research objective

The present work aims to develop a new statistical approach to estimate extreme surface wind speed and wave height return period values in a changing climate. It focuses on the wind speed at 10 m above sea level, and the significant wave height. These are the values commonly used in engineering practice to define the design sea state. Wind speed and significant wave height can be derived from global phase-averaged atmosphere-ocean models, as well as observational datasets, as described in Chapter 2. The objective is to increase confidence in the estimates of the changes in extreme conditions due to climate change.

To achieve this objective, the thesis has been developed in three phases. First, we analyse past ocean surface wind speed and wave heights extending back to the beginning of the 20$^{\text{th}}$ century. We compare a global climate model with two reanalyses, where the numerical model is calibrated with in-situ observations of the atmosphere-ocean system. The objective is to understand if a climate signal—long-term trend linked to climate change—can be detected from the model evolution of the wind-wave field. Particular focus is given to a review of available observational datasets, and how these are assimilated in the two reanalyses. The aim is to identify possible spurious trends caused by the assimilation, over the years, of differing numbers of observations, characterized by increased resolution.

Following this analysis of the past climate, the work develops an innovative EVA technique, introduced by *Breivik et al.* (2013, 2014), to estimate wind speed and wave height return period from medium range operational forecast datasets. The objective is to reduce the uncertainties connected to the estimation of the 100-year design return periods of wind speed and significant wave height, to provide increased confidence in the design of marine systems and coastal defences around the world. This is done using an ensemble of independent forecasts, equivalent to a time series far larger than the ones used for traditional EVA approaches.

The third phase of the work applies this novel ensemble EVA approach to future projections of wave climate, as derived from global wave model runs forced with different GCM surface wind fields. The aim is to understand possible estimates of the changes in extreme significant wave height by the end of the 21$^{\text{st}}$ century. We

also compare the magnitude and frequency of extreme events between the past few decades and the end of the 21st century. This analysis provides a basis for estimating possible changes in wave extremes over the coming century.

The three projects composing the present thesis, have been published or submitted to scientific journals. The aim of these publications is to describe the evolution of wind-wave extremes in a changing climate, and to further reduce the uncertainties connected to their estimates. As a result, the present statistical approaches provide a basis for the design of safe and resilient marine structures and operations.

1.3 Thesis outline

The thesis outlines the development of the three projects described above and developed in collaboration with the Norwegian Meteorological Institute in Bergen (Norway), and the Australian Commonwealth Scientific and Industrial Research Organisation, CSIRO Hobart (Tasmania). The chapter outlines are as follows:

Chapter 2 - Wind-wave climate and extremes from model and observational datasets This chapter is an extended literature review of model and observational datasets of ocean surface wind speed and wave height. The chapter describes in detail the advantages and disadvantages of these products in determining past and present wind-wave climate. It raises questions as to our current ability to reproduce and predict ocean surface wind and waves, with a particular focus on the model datasets used throughout this PhD work.

Chapter 3 - Trends from 20th century models compared to satellite altimeter observations This work was developed as part of a visiting period at the Norwegian Meteorological Institute in Bergen, Norway. It analyses the impact of assimilated observations on global atmopshere-ocean wave models. The study describes past century trends derived from an ECMWF climate model and two reanalyses. The climate model trend results (no assimilated observations) are compared with the results from the two reanalyses, which assimilate observations of ocean surface pressure and wind speed. An additional comparison with an extended dataset of calibrated satellite altimeter ocean observations, discusses the models' trend statistical significance for both wind speed and wave height. Finally, we discuss whether the climate trends found are real or spurious. This chapter is based on the following journal publication:

- Meucci, A., Young, I. R., Aarnes, O. J., and Breivik, Ø (2020a). Comparison of wind speed and wave height trends from twentieth-century models

and satellite altimeters. Journal of Climate, 33(2):611–624. (©American Meteorological Society. Used with permission.)

Chapter 4 - Wind and wave Extreme Value Analysis This chapter introduces extreme statistical analysis. After considering the challenges in observing and estimating metocean extremes, the chapter describes the state-of-the-art in EVA and common traditional approaches used for metocean variables. We show the limitations connected to traditional EVA approaches, and argue for the need to consider ensemble model realizations for enhanced EVA. Also, we suggest that the increasing number of observations, as well as increasing model resolution and performance, will further improve statistical inference of extremes, increasing the confidence in the estimates.

Chapter 5 - Wind and wave extremes from atmosphere and wave model ensemble forecasts This chapter develops an innovative EVA technique, first introduced by *Breivik et al.* (2013, 2014), to estimate global 100-year return period values of ocean surface wind speed and wave height. The analysis is applied to an ensemble of independent and identically distributed advanced lead time forecasts. In this way, we are able to reconstruct a synthetic dataset equivalent to 750 years of data, far longer than commonly available time series. The ensemble approach is demonstrated to be a potential tool to apply to independent model realizations of the atmosphere-ocean system, increasing the confidence level of extreme estimates (*Meucci et al.*, 2018). This chapter is based on the following journal publication:

- Meucci, A., Young, I. R., and Breivik, Ø. (2018). Wind and wave extremes from atmosphere and wave model ensembles. *Journal of Climate*, 31(21):8819–8842. (©American Meteorological Society. Used with permission.)

Chapter 6 - Future projections of wind-wave extremes This part of the thesis discusses current state-of-the-art wind-wave climate extreme projections. It investigates to what extent wind-wave extremes are changing as a result of climate change. We investigate available GCM datasets and comment on their skill in representing the wind-wave climate, as well as extremes. We further discuss the sources of uncertainties connected to climate projections, the challenges to estimate future extremes, and potential new techniques to reduce the effect of these uncertainties. The aim of the chapter is to give a critical analysis of the current level of confidence in future wind-wave climate extreme projection.

Chapter 7 - Projected 21st Century changes in extreme wind-wave events In this chapter we apply the ensemble technique, developed for operational forecast datasets (chapter 5), to an ensemble of global wave model runs, forced with seven different GCM surface wind fields (*Meucci et al.*, 2020b). The inter-model ensemble technique is able to make use of all seven independent GCMs extreme information available to synthesize a large dataset of independent ocean storms, and reduce the confidence intervals associated with estimates of future changes in wave height extremes. The wave height extremes magnitude and frequency changes towards the end of 21st century are discussed for two Representative Concentration Pathway (RCP): RCP4.5 and RCP8.5. This chapter is based on the following journal publication:

- Meucci, A., Young, I. R., Hemer, M., Kirezci, E., and Ranasinghe, R. (2020b). Projected 21st century changes in extreme wind-wave events. Science Advances, 6(24).

Chapter 8 - Conclusions and future research The conclusive chapter discusses the results achieved in the present thesis, and introduces work in progress, with planned future activities to improve the statistical analysis of wind-wave extremes.

Chapter 2

Wind-wave climate and extremes from model and observational datasets

2.1 Introduction

Our current knowledge of sea state and our ability to predict its evolution are based on numerical atmospheric models coupled with phase-averaged third-generation wave models. We use the terminology phase-averaged as the sea state is represented as a linear superposition of waves with random phase (*Holthuijsen*, 2007), and third-generation as they allow the evolution of the wave spectrum through the computation of non-linear wave-wave interactions, as opposed to first and second-generation wave models where spectral forms were imposed parametrically. These models are calibrated with continuously increasing amounts of atmosphere-ocean observations, that consists of large comprehensive datasets of in-situ and remote sensing measurements.

Thanks to the increasing number and quality of modelled and observational time series, our ability to forecast short-term average sea state has been significantly improving (*Cavaleri et al.*, 2018). Specialized research centres and international meteorological agencies focus on resolving the atmosphere-ocean system through these global mathematical models, calibrated with real-time observations. Centres such as the European Centre for Medium-Range Weather Forecasting (ECMWF), the American National Oceanic and Atmospheric Administration (NOAA), the NASA Global Modelling and Assimilation Centre (GMAO), the UK Meteorological Office (Met Office), the Australian Bureau of Meteorology (BoM) and the Japanese Meteorological Agency (JMA), have cutting-edge research experience and technologies to run global and localized models with constantly increas-

ing forecast ability. Today's computational capabilities allow these global models to represent ocean surface wind speed and wave height conditions at an hourly time scale, with less than a 100 km horizontal resolution. Models are calibrated with an increasing amount of observations, and thus are able to accurately represent past and present synoptic sea states across the globe.

These products are a valuable resource to understand, predict and map past, present, and future global ocean climate. However, the interpretation of model results, or observational time series, is not as straightforward as one might think. Due to their intrinsic uncertainties, and unresolved physical processes, models are not able to accurately predict wind and wave surface fields, if not calibrated with observations or tuned at the right parametrized values (*Cavaleri et al.*, 2007). Besides, observational time series are still limited in space and time to perform accurate long-term statistical analyses.

An even more complicated picture emerges at extreme conditions, where the observations are mostly biased (*Large et al.*, 1995; *Bender et al.*, 2010), and the model abilities to reproduce extreme sea states are further reduced (*Powell et al.*, 2003; *Cavaleri*, 2009; *Babanin et al.*, 2019). These limitations affect our ability to estimate ocean wind speed and wave height extreme events, and to quantify possible past and future changes connected to climate change. Here, we refer to climate, as opposed to weather, as the long-term variations in the Earth's atmosphere-ocean system, and long-term statistics as the branch of statistics that relates to extreme estimates. That is, the estimate of the exceedance probability or return period—one in N years event—of extremes.

Wind-wave climate extreme statistics refer to long-term time windows of usually decades or longer. At this scale ocean conditions are not statistically stationary, and specific statistical approaches are needed to investigate ocean wind and wave characteristics (*Holthuijsen*, 2007). These analyses rely on the accuracy and homogeneity of the time series dataset used. Available observational datasets are generally too short to confidently estimate both climate trends and extremes, thus model reanalysis products, that run back many decades, and are calibrated with available observations, represent a potential alternative. These models are able to reproduce long stretches of climate, however, reanalysis time series are non-homogeneous, as they assimilate, over their computational time period, differing amounts of observations with a varying level of accuracy. This is a strong limitation if we want to study wind and wave climate, as it might introduce spurious trends (*Bengtsson et al.*, 2004). Furthermore, these datasets are non-stationary, thus impeding the use of common EVA approaches (*Bengtsson et al.*, 2004; *Aarnes et al.*, 2015).

The present thesis aims to investigate long-term changes in ocean surface wind speed and wave height, with a specific focus on extremes, critical to safe operations

at sea, and for the stability of the world's coastlines. The magnitude and frequency of occurrence of future extreme events is also investigated. This chapter reviews the current state-of-the-art of global observations and models of the atmosphere-ocean system, highlighting their potential and limitations. These subjects are framed within a discussion of the level of confidence in detecting climate trends, and predicting extremes of ocean surface wind speed and wave height. To do this we discuss and elucidate the sources of uncertainties that characterize global modelled and observational datasets. We show why it is difficult to identify physically sound trends in the wind-wave climate, and what are the aspects to be aware of when performing these analyses. Furthermore, we investigate and understand the level of confidence around extreme value statistical estimates of wind speed and wave height, as these estimates are crucial to safe design of coastal and offshore maritime structures and operations. This is done by investigating the sources of uncertainties in coupled atmosphere-wave models, with special attention to issues that limit our understanding of trends, extreme climate, and the connection between these two aspects. The focus then shifts to observational datasets. It is noted that the two parts are often intertwined as models are calibrated with observations.

Once all the sources of uncertainties are explained, we introduce the concept of a probabilistic ensemble approach for Numerical Weather Prediction (NWP) models, showing the potential of using these datasets for a novel approach to estimate statistical extremes of ocean surface wind speed and wave height. At the end of this review chapter, the reader should be familiar with the current level of confidence around ocean wind and wave climate trend estimates and the probability of exceedance of extreme events.

2.2 Towards fully coupled atmosphere-ocean-wave models

The atmosphere-ocean system is extremely complex to understand and to model. In their quest to further improve forecast abilities, international meteorological centres are moving towards fully coupled operational systems. In this, the atmosphere model, the ocean (and ice) model, and the wave model, dynamically interact. At the moment, the ECMWF is implementing a coupled atmosphere, wave and ocean model (Fig. 2.1) with the plan to move to a fully coupled system model in the future with the introduction of ice modelling associated to the ocean model.

Fig. 2.1 shows the three coupling processes that characterize current operational forecasts at ECMWF. The objective is to model the atmosphere-ocean system resolving the dynamical interactions between atmospheric flows and the

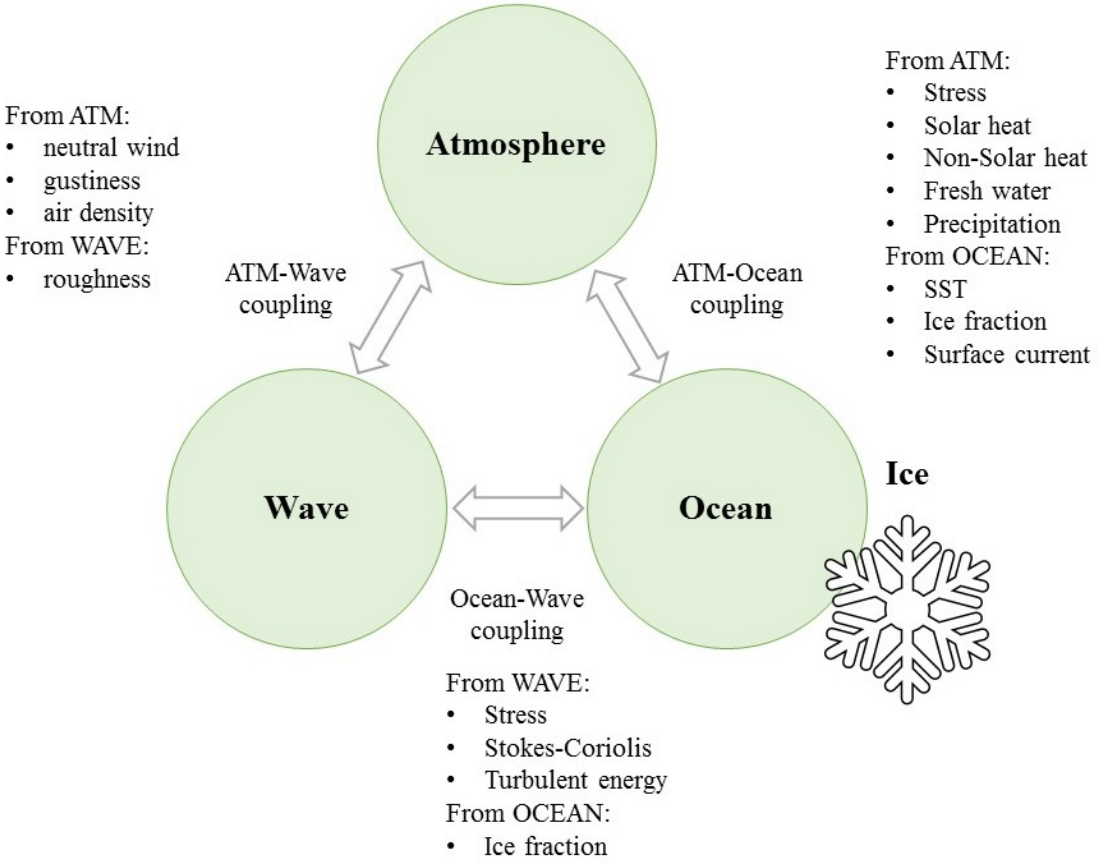


Figure 2.1: Operational forecast coupled atmosphere, ocean, and wave models, as implemented by ECMWF from June 2018 - IFS Cy45r1.

ocean, and in this way improve the forecast skill of the operational datasets. For the purpose of this thesis, we focus on the coupling between the atmospheric and the wave models, hereinafter atmosphere-wave models (upper-left link in Fig. 2.1).

In the following sections, we refer to different types of models. As a reference for the reader, below we present the range of model products for ocean surface wind speed and wave height. Global models are classified as hindcasts, reanalyses, operational, and climate models. A hindcast is a numerical model that runs for an historical period to reproduce the states of the atmosphere and the ocean without assimilating any observations. A reanalysis is a hindcast that includes data assimilation. Operational models are NWP models for the prediction of weather. These models concentrate most of the efforts and capacities of meteorological centres as they run at fine resolutions, and assimilate observations in real-time to produce the best estimate of the initial state of the atmosphere, fundamental to improve models' forecast skills. Climate models simulate the evolution of Earth's climate

through the introduction of drivers such as radiative and convective forces. These models are used to estimate future projections of the atmosphere and ocean and study the effect of climate change. The most common reference are Global Climate Models (GCMs) which model the ocean and atmosphere state on a global grid under different greenhouse gas emissions scenarios.

We study the long-term statistics of ocean wind speed and wave height from a series of different atmosphere-wave and climate models. In each of these studies, we describe the main characteristics and limitations of the model type or version used, focusing on the limitations that challenge the studies for climate trends and extreme values. Coupled atmosphere-wave models are used in Chapter 3 and Chapter 5. For future projections we use an ensemble of different GCM projected wind fields to force a global wave model (Chapter 7), and reduce the confidence intervals around future projections of wave height extremes.

2.2.1 Coupled atmosphere-wave models

Today, it is common for atmosphere models to dynamically interact with third-generation wave models. The neutral wind, gustiness and air density products of the atmospheric model are implemented into the wave model. The wave model feeds back to the atmosphere the modified drag coefficient based on the newly computed ocean surface roughness (*Janssen*, 2004). The mathematical basis of atmosphere-wave models are the Navier-Stokes evolution equations of a two-layer incompressible fluid. Since the density of the air is much smaller than the density of water, ocean waves can be described by the Navier-Stokes equation for a one layer fluid in a gravitational field (*Komen et al.*, 1994). The Laplace solution of these equations describes the linear theory of ocean gravity waves.

Due to the computational impossibility of fully resolving the Navier-Stokes equation at the global scale, and since waves are a stochastic process, the sea surface is resolved by means of a wave spectrum, that evolves through space and time (*Young*, 1999b; *Ochi*, 2005; *Holthuijsen*, 2007). As already mentioned above, these are random-phase/amplitude wave models. The most common third-generation global ocean wave models are the WAve Modeling (WAM) (*WAMDI*, 1988; *Komen et al.*, 1994) developed at the ECMWF, and the WWIII model (*Tolman and Chalikov*, 1996) at the NOAA. In these models, the evolution of the wave spectrum, and thus the physical mechanisms of ocean wave evolution, are defined by the energy balance equation, derived from the above mentioned Navier-Stokes equations. The energy balance equation can be represented in its most simple form in Eq. 2.1 (*Young*, 1999b):

$$\frac{\partial N}{\partial t} + \nabla \cdot (\vec{C}_g N) = S_{tot} \tag{2.1}$$

The left part of the equation derives from linear theory assumptions. N is the wave action, defined as E/ω_0 where: E is the wave energy density spectrum and ω_0 is the relative angular frequency measured in a frame of reference moving with the current (*Young*, 1999b); $\vec{C_g}$ is the group velocity vector. The right part of the equation is a source/sink term that is defined with the assumption of linear superposition of physical processes. This is represented as a summation of independent processes adding or subtracting energy to the waves. The main processes of deep water ocean wave physics are (Eq. 2.2):

$$S_{tot} = S_{in} + S_{nl} + S_{ds} \tag{2.2}$$

S_{in} is the atmospheric, or wind, input, S_{nl} accounts for the non-linear interactions between waves (*Hasselmann*, 1962) and S_{ds} is the dissipation due to white capping, or dissipation in deep water. The present thesis focuses on global statistical analyses of deep water ocean conditions. In the following section, we focus our attention on describing the uncertainties connected to the physical processes happening in deep water conditions. Note that, global models consider intermediate to shallow water source terms as well, but here we limit our review to the uncertainties in modelling deep water ocean areas.

2.3 Atmosphere-wave model sources of uncertainties

The reliability of coupled atmosphere-wave models have greatly improved in the past few years. However, many uncertainties still characterize these products at various scales, from the global meteorological boundary conditions to the unresolved physical phenomena of the interaction between the atmosphere and the ocean, a number of uncertainties affect the correct model representation of surface wind speed and wave height. Furthermore, the observational sources, from which measurements of the atmosphere and the oceans are assimilated into the models, have their intrinsic uncertainties, that translate to challenges in determining the initial conditions from where the model simulations are initialized. Thus, on the one hand we have a limited understanding of the physical processes involved in the wind-wave coupled climate system (*Cavaleri et al.*, 2012), and on the other hand, we face challenges in both in-situ and remote observations of the oceans.

2.3.1 Modelling ocean surface wind

The uncertainties in representing the atmosphere-ocean system challenge accurate model representation of the ocean surface wind speed and wave height. The first

cluster of uncertainties to consider are related to the atmospheric models' ability to correctly represent surface wind speed. Here, we refer, by international convention, to the wind speed at 10 m above mean sea level, U_{10}.

Atmospheric models have generally low biases for moderate wind speed (i.e. <20 m/s). Conversely, large biases are found at high wind speeds (>25 m/s), where the values are underestimated when compared to independent satellite and buoys observations (*Stopa and Cheung*, 2014). *Stopa and Cheung* (2014) investigated biases in the ocean surface wind speed comparing two major reanalyses: the National Centers for Environmental Prediction (NCEP) Climate Forecast System Reanalysis (CFSR) and the ECMWF European ReAnalysis Interim (ERA-I). They found that, if we exclude very low winds, both CFSR and ERA-I underestimate the wind speed when compared to observations, and ERA-I shows low variability. Also, it is clear that both datasets are affected by the changing number of observations assimilated throughout the years. Discontinuities have been observed not only in the CFSR dataset (*Chawla et al.*, 2013; *Rascle and Ardhuin*, 2013), and ERA-I (*Aarnes et al.*, 2015), but also in 20th Century climate model and reanalyses as shown by *Wohland et al.* (2019) and *Meucci et al.* (2020a)

A more recent comparison between modelled wind speed from the ECMWF IFS Operational forecast - Cycle 41 round 1 (Cy41r1) coupled atmosphere-wave model, confirms that the extremes are biased low compared to a comprehensive set of observations in the North Sea (*Pineau-Guillou et al.*, 2018). *Cavaleri* (2009) suggests that a major part of this underestimation is related to the model resolution. Large uncertainties also characterize the observations of high wind speed (more than 30 m/s), as it is challenging to obtain accurate measurements of wind speed at the extremes (*Zeng and Brown*, 1998; *Howden et al.*, 2008; *Babanin et al.*, 2019). Thus, it is difficult to understand where the "truth" stands between the model and observations.

For the purpose of the present work two issues of atmospheric model wind speed are particularly interesting. On one side the inhomogeneity of the past model climate is an issue due to different amounts of observations assimilated into the reanalyses throughout the years, with increasing quality. On the other side, the challenges in reproducing the extreme events, where models and observations usually underestimate the true values is also of importance.

Also, the representation of wind in the atmospheric model has a direct impact on wave model performance, as wave forecasts with lowest errors in the wind speed are also the ones with lowest errors in the wave height (*Bidlot et al.*, 2002). However, the errors in the atmospheric wind values are not sufficient to explain the biases found in modelled wave height (*Cavaleri et al.*, 2007; *Cavaleri*, 2009). Wave model inaccuracies in surface stress formulation and parametrization play a critical role. In the following sections, we describe the wave model sources of

uncertainties in the representation of wave height.

2.3.2 Modelling the waves

Recent comprehensive reviews of the state-of-the-art of wave modelling by *Cavaleri et al.* (2007, 2018) identified the main sources of uncertainties in wave models. As the present thesis is focused on the climate and long-term statistics of deep water wind speed and wave height (deep water is where the ratio of the water depth to the wavelength of more than 0.5 (*Young*, 1999b)), we describe only the sources of uncertainties related to modelling waves in these areas. It is well known that wave models perform better in deep waters than in coastal areas or enclosed basins, as also shown in Fig. 2.5 of the *Cavaleri et al.* (2018) review. Globally, the scatter index is around 10% for deep waters, and more than 20% for coastal areas and enclosed basins when compared to observations (*Cavaleri et al.*, 2018).

However, if we are to perform long-term statistical analyses of ocean surface wind speed and wave height, additional limitations and uncertainties also emerge in deep water conditions. The objective of the following sections is to explore the main causes for these limitations, and how these motivate the approaches followed in the present thesis. It should be noted that, the uncertainties and caveats described do not take anything away from the incredible progress that has been made in wave modelling, making wave models unquestionably critical tools for practical applications, and to improve our understanding of the atmosphere-ocean system.

Physical processes

The main physical processes of wave evolution are represented or parametrized by the three source terms in Eq. 2.2, right side of Eq. 2.1. The first term S_{in} describes the air-sea interaction processes. The *Cavaleri et al.* (2007) review outlines several steps forward which have been made to reduce errors at strong winds over the water surface, especially with the introduction of the gustiness effect. However, the momentum transfer at high wind speed is still not well understood (*Powell et al.*, 2003; *Cavaleri et al.*, 2007), and models are tuned for the bulk of wind data observations following *Miles* (1957) theory, which is probably not accurate under extreme conditions. Phenomena such as sea spray affect the wind drag over the ocean, but it is still not clear to what extent sea spray impacts the action of wind over the sea (*Powell et al.*, 2003; *Makin*, 2005; *Richter and Sullivan*, 2013). It may have quite a large impact on the representation of the sea state during storms such as Tropical Cyclones (*Holthuijsen et al.*, 2012). Further uncertainties are added by sudden changes in wind direction, which affect the waves. This process is still not

well understood and thus not properly represented by the models (*Van Vledder and Holthuijsen*, 1993; *Cavaleri et al.*, 2007).

A second class of uncertainties concerns the representation of the four-wave non-linear interactions, S_{nl}, theorized by *Hasselmann* (1962) and confirmed experimentally by the Joint North Sea Wave Project (JONSWAP) (*Hasselmann et al.*, 1973). According to *Cavaleri et al.* (2007), the four-wave non-linear interaction term is possibly the process that is formally best defined by current wave models. However, solving the energy density integral of the *Hasselmann* (1962) equation is time-consuming. Thus, to facilitate operational models a Discrete Interaction Approximation (DIA) was introduced by *Hasselmann et al.* (1985). The DIA comes with some shortcomings, such as in the case of shallow water conditions, and in its ability to model directional spreading, thus reducing the confidence of the model results in coastal areas. Furthermore, the role of the four-wave interaction mechanism is still not well known for cases such as turning winds, or mixed wind-sea and swell.

The third class of uncertainties concerns the wave energy dissipation in deep water, parametrized by the spectral dissipation term, S_{ds}. This is an area of research that has been extensively explored in the past few years, as it is also the least understood of the ocean wave processes (*Cavaleri et al.*, 2007). Research studies have been recently published from observational experiments that lead to novel approaches to the S_{ds} source term parametrization (*Young and Babanin*, 2006; *Babanin et al.*, 2010; *Ardhuin et al.*, 2010; *Zieger et al.*, 2015). These studies have been incorporated into new versions of global wave models (*Tolman et al.*, 2013). Despite these advances, the prediction of swell height and arrival time are still today characterized by uncertainties, which is believed to be an indication of our lack of understanding of dissipation.

Numerical uncertainties

Global wave models use finite-difference schemes to solve the differential equation of the wave balance (Eq. 2.1) over a grid. The numerical uncertainties connected with such schemes have been studied for decades, and their impact on today's model versions mainly involves the propagation of swell (*Cavaleri et al.*, 2007). Increasing spectral resolution (frequency and direction) reduces the numerical error, but global models have computational constraints due to the need to resolve large domains. For the purpose of this thesis (climate trends and EVA) the sources of uncertainties deriving from the numerical scheme are not significant compared to other sources of error.

Model resolution

The model resolution is an important aspect to consider if we are to analyse long-term climate and extreme statistics from model time series. Firstly, a coarse global model resolution might result in some small island groups not being represented. In this case, the spectral energy that should be dissipated/reflected by these islands is not resolved, as well as the effect on the wind speed. This will potentially result in errors in the representation of the wave height and surface wind speed, such as in the case of the Hawaiian archipelago for horizontal resolutions of 1° or more (*Meucci et al.*, 2018). This is an issue to take into account when comparing the model climate, with in-situ and remote sensing observations. The case of the Hawaiian islands will be shown in Chapter 5, where a comparison of the operational forecast time series, with a National Data Buoys Center (USA) (NDBC) buoy shows a strong normalized difference in the extreme value estimates of the 10-m surface wind speed. Indeed, the model resolution is especially critical for a correct representation of the extremes (*Cavaleri and Bertotti*, 2006; *Cavaleri*, 2009; *Timmermans et al.*, 2017). *Cavaleri and Bertotti* (2006) compared the results from the same meteorological model wind forcing at different resolutions and showed that the maximum values of both wave height and wind speed increase with the model resolution. In addition, the highest peaks are probably also underestimated due to incorrect model physics (*Cavaleri and Bertotti*, 2006). *Timmermans et al.* (2017) demonstrated how the difference in wind forcing resolution impacts global wave model results, especially in Tropical Cyclone areas. In that study, an increase of horizontal resolution from 1.0° to 0.25° significantly improved the level of accuracy of the wave model. Model resolution has also been shown to be critical for future projections of the extremes (*Shimura et al.*, 2015, 2017).

2.3.3 Modelling extreme sea states

In this section, we focus on model and observational sources of uncertainties that affect the estimation of extreme values at sea. The statistical aspects of this topic are described in detail in Chapter 4, which introduces the last two elements of the present thesis on EVA of ocean surface wind speed and wave height.

Cavaleri (2009) treats the main aspects and uncertainties related to modelling extreme ocean waves. Firstly, the physics of the extremes are not well understood. Research has only recently started to consider the effect of sea spray and the separation of flows in extreme sea states, i.e. more than 30 m/s wind speed. The physics of extreme sea states might be significantly different from that found in more benign conditions and implemented in models.

Models tend to tune their input forcing parametrization to the bulk of the data based on *Miles* (1957) mechanism (*Ardhuin et al.*, 2010), which works quite well

for wind speed below 20 m/s, but has never been demonstrated to work at high wind speeds. It is important to note that the wind input source term has a large impact on the wave field (*Young*, 1999b).

Furthermore, models do not resolve currents well in some areas of the ocean. This causes errors in estimating energy concentrated in these areas, potentially impacting the ability to properly model the peaks (*Babanin et al.*, 2017; *Romero et al.*, 2017).

It is helpful to compare model extreme estimates with observational time series, but in many cases, both in-situ and remote sensing observations are questionable at high values of wind speed and wave height, due to challenges connected to the measurement techniques used. This issue is explored further in the introduction of Chapter 5 as a preamble and motivation to the EVA approach developed.

It has been demonstrated by numerous studies (*Bidlot et al.*, 2002; *Caires and Sterl*, 2005; *Cavaleri*, 2009; *Caires*, 2011) that wave models tend to underestimate the peaks, and this is not only due to biases in the wind speed fields as shown by *Pineau-Guillou et al.* (2018), but also due to the other aspects discussed in this section.

A particular issue is predictions for Tropical Cyclones. These are localized extreme events, particularly challenging to measure and model, but that have significant importance for safety at sea and on the coasts. We dedicate the following subsection to this topic.

Tropical Cyclones

Intense Tropical Cyclones (TC), also called hurricanes, or in Asia typhoons, are some of the deadliest natural disasters and their increasing frequency has been argued by many (*Emanuel*, 2005; *Landsea et al.*, 2006; *Elsner et al.*, 2008; *Bender et al.*, 2010). These phenomena have been studied for years and measurements are available from GPS drop sondes (*Powell et al.*, 2003), airplane remote sensing, buoys, platforms, Synthetic Aperture Radar (SAR) and Scanning Radar Altimeter (SRA) data (*Ochi*, 2003). However, TC extremes are still not easily predicted. From a recent review by *Young* (2017) it is clear that even after 30 years of observations, we still lack a complete understanding of the physics of TCs. Many variables affect the evolution and occurrence of these extreme events. To describe TC track and intensity four variables are usually considered: p_0, pressure at the centre of the cyclone, the *track* of the cyclone, V_{fm}, the velocity of the forward movement, and R, the cyclone's radius. A Monte Carlo simulation is commonly used to obtain the initial probability distribution of these variables, to then model the TC evolution in terms of track and intensity. In many engineering applications, the parametric model of *Young* (1988) is used to determine the equivalent fetch of a TC and from it, by means of the JONSWAP relationship, the maximum

wave heights, and furthermore the directional spectra (*Young*, 2006). Recently third-generation wave models have also been tested to simulate TCs, but the main uncertainties are still related to the correct estimation of the wind field forcing conditions (*Liu et al.*, 2017). However, at the current resolution, global models cannot reproduce localized extreme events such as the Tropical Cyclones. The need for increased resolution of TC wind fields that force wave models is emphasized also in the work of *Timmermans et al.* (2017). In addition, *Shimura et al.* (2015) found that the ability of global climate models to accurately simulate TCs is of crucial importance for future projections of extreme ocean waves.

TC formation and evolution is connected to global atmosphere-ocean states. As pointed out by *Breivik et al.* (2014), the operational global model at ECMWF, produces forecasts that usually perform well for extra-tropical areas but do not correctly reproduce TCs. This is due to model resolution, and uncertainties in initial conditions and model error propagation. The impact of the perturbation methods, used for simulating initial conditions and ensemble predictions, on TCs, have been investigated by *Puri et al.* (2001) and *Lang et al.* (2012). However, these studies refer to previous operational forecast models that had slightly different perturbation schemes and a coarser resolution. After the release, in March 2016, of the new ECMWF IFS model cycle, data are now globally available at a horizontal resolution of 9km for the HRES (High RESolution) model. Studies are needed to understand if this resolution is sufficient to correctly reproduce TC wind and wave fields.

2.4 Model assimilation of observations

Data assimilation techniques are used to integrate dynamical models with measurements of the atmosphere-ocean system. A Four-Dimensional Variational scheme (4D-Var) (*Rabier et al.*, 1998) is used at ECMWF to assimilate measurements of the atmosphere-ocean system over an assimilation window. It was introduced more than 20 years ago in operational models, and today it is still considered as the best cost-effective data assimilation technique to estimate initial conditions of the atmosphere-ocean system and initialize operational forecasts and reanalyses. An Ensemble Data Assimilation (EDA) approach was added in 2010 to the IFS at ECMWF (*Isaksen et al.*, 2010), to estimate the uncertainty around the initial conditions on which the global model is initialized. In this approach, ten independent lower-resolution 4D-Var assimilations are run with slightly perturbed observations, to estimate the uncertainties for the initial state of the atmosphere-ocean system. As *Lorenz* (1963) first indicated, slightly perturbed initial conditions might evolve into completely different meteorological systems and preclude a correct analysis of the atmosphere-ocean system. Thus, data assimilation techniques constantly cor-

rect and guide models towards the best estimate of the Earth's weather conditions, as measured by different real-time observations.

Data assimilation is fundamental to reduce the error and scatter index of model results, improving the synoptic estimates of ocean surface wind speed and wave height (*Saha et al.*, 2010; *Dee et al.*, 2011). However, it comes with great computational costs and some drawbacks. As already mentioned, over time (years) reanalyses have assimilated a different number of observations, characterized by an increasing level of accuracy, which translate into inhomogeneities in the datasets that may affect the estimates of long-term climate variations (*Bengtsson et al.*, 2004; *Weisse*, 2010; *Wohland et al.*, 2019). This is a critical aspect to consider if we are to compute long-term climate variations of physical parameters. Chapter 3 shows how contrasting trends are found between climate models and reanalyses. This mainly occurs because reanalyses assimilate observation whereas climate models don't.

Atmosphere-wave models rely on satellite measurements, which are increasingly denser both in space and time. Thus, buoy measurements are important to check the reliability of the models, as these are not assimilated, hence, independent datasets. Data assimilation has little impact at the extremes where the intrinsic challenges of measuring extreme wind and waves, hinder a proper calibration of the models, even with the best ensemble uncertainty estimates of the initial conditions.

2.5 Ocean observation uncertainties

The above discussion indicates that, to accurately represent the atmosphere-ocean system, numerical models need to be combined with real-time observations. This is critical for a detailed knowledge of synoptic systems and their short-term forecasts, as well as long-term variation of the climate with important repercussions for human activities and global economies. This is why international associations such as the Joint technical Commission for Oceanography and Marine Meteorology (JCOMM), under the World Meteorological Organization Intergovernmental Oceanographic Commission (WMO-IOC), coordinate oceanographic and marine meteorological observations with data management and services, to produce observational time series for the validation of operational forecasts. The International Comprehensive Ocean-Atmosphere Data Set (ICOADS) is the largest dataset of collected observations from ships, platforms, and buoys around the globe. This dataset, in its latest release (*Freeman et al.*, 2017), contains measurements that date back to the 17$^{\text{th}}$ Century, and have been incorporated in the latest atmosphere-ocean coupled models, further improving their quality (*Freeman et al.*, 2017). Of particular interest to this project are the derived products of this dataset such as the wind speed and the wave height (*Gulev and Grigorieva*, 2004, 2006), as

this has a direct impact on the long-term analyses performed in the present work, being used in many of the models that we investigate. Specific attention has to be paid to the version of the ICOADS dataset that is incorporated in the models that we analyse, as, for instance, long-term characteristics of wind speed has been modified after the work of *Tokinaga and Xie* (2011).

The Australian Integrated Marine Observing System (IMOS) also collects observations of the ocean surface. It now archives a 33 year dataset of calibrated surface wind speed and wave height observations from 13 satellite altimeter missions (*Zieger et al.*, 2009; *Ribal and Young*, 2019). This unique observational dataset has been used to study the wind-wave climate of the global ocean, showing increasing trends in the wind speed, and varying distribution of the wave height (*Young et al.*, 2011; *Young and Ribal*, 2019).

Significant work has been undertaken to collect an impressive amount of in-situ and remote-sensing observations, which now represent an invaluable resource to enhance the quality of global models. However, even though in the past few decades, there has been a significant improvement in the number and quality of ocean observations, we still need improvement and further investments in observational techniques and coverage (*Ardhuin et al.*, 2019). This would help understand unresolved physical processes and better calibrate models (*Ardhuin et al.*, 2019; *Kent et al.*, 2019). This is particularly true for extremes and poorly observed areas such as the Southern Ocean (*Babanin et al.*, 2019).

There is also a significant bias between buoys and platforms at high wind speeds. Buoy winds are lower than platform winds, by about 3m/s at 25m/s (*Pineau-Guillou et al.*, 2018). This could be due to the transformation of wind measurements to 10 m above sea level. However, there are also questions about whether buoys underestimate high wind speeds as they are affected by the wave motion or by sheltering effects (*Zeng and Brown*, 1998; *Pineau-Guillou et al.*, 2018).

Furthermore, buoy measurements are limited in spatial distribution and thus cover only a small part of the ocean, with the longest time series of ocean wind speed and wave height being measured by NDBC buoys with a maximum time span of 30 years, and mostly located around the North American continent.

In contrast, satellite measurements cover the majority of the global oceans but are limited by their over-pass times, and limited cross-track resolution, and thus may miss storms. This is particularly important if we want to estimate extreme values (*Takbash et al.*, 2018), and might result in underestimated extreme value return periods. Satellite measurements are also an important dataset for climate trend studies, but special care is needed for the poorly observed initial period before 1991.

Thus, to understand the results of a long-term analysis of ocean surface wind speed and wave height we must consider all of the above aspects of the observa-

tional datasets. These can be considered as additional sources of uncertainties, that in the case of assimilation in a global model, interact with the model sources of uncertainties. Hence, it is not trivial to understand how these sources of uncertainties interact in global models. This creates an intricate network of uncertainties that challenges common statistical approaches (*Coles et al.*, 2001).

As a result of these complexities, ensemble probabilistic approaches represent a reasonable way forward (*Saetra and Bidlot*, 2004; *Breivik et al.*, 2014; *Meucci et al.*, 2018). The following section introduces such ensemble techniques, from their theoretical basis to application in operational forecast models, and finally to the potential they have for ocean wind and wave statistical analyses. Ensemble approaches are the linking concept throughout this thesis. This concept is growing in importance as recent reference studies suggest the use of these tools is fundamental to understanding the various sources of uncertainties around estimates of ocean wind speed and wave height (*Morim et al.*, 2019).

2.6 The ensemble probabilistic approach

As stated by *Cavaleri* (2009) "[...] we do not have enough data at our disposal to have a continuous deterministic picture of the oceans. It seems that, at least for certain aspects, we will have to live more and more in a probabilistic world.". This point is particularly interesting for the future development of wave modelling research. The ensemble probabilistic approach is indeed the baseline approach of this PhD thesis. It was first introduced in the field of meteorological forecasting to improve predictability, the guiding light of the many efforts of international forecasters to avoid dangerous and costly errors in forecasts (*Lewis*, 2005).

Predictability means the estimation of uncertainty (*Palmer and Hagedorn*, 2006). Weather and climate, and thus ocean wind and waves, are, by their nature, characterized by uncertainty. Such hydro-dynamically complex systems can develop significantly different states from slightly different initial conditions, as explained by *Lorenz* (1963) in his well-known chaos theory. We deal with uncertainties not only in determining future states of the atmosphere but also in defining the current analysis time states (*Epstein*, 1969), i.e initial conditions of the models. Thus as a response to these intrinsic uncertainties, probabilistic approaches and Ensemble Forecasts are now commonly used in weather predictions, as explained in the historical review of *Lewis* (2005). In addition, these approaches are also used in hindcast (see ERA-20CM (*Hersbach et al.*, 2015a)) and reanalysis models, with the CERA-20C (*Laloyaux et al.*, 2018) and the latest ECMWF global reanalysis ERA5 being typical examples.

In the past few decades numerous approaches have been implemented at the ECMWF for the development, maintenance and constant upgrade of Atmosphere-

Ocean coupled model Ensemble Prediction Systems (Ensemble (ENS)) at the global scale (*Molteni et al.*, 1996). These models face challenges posed by computing capacity; hence many efforts today are focused on models capable of producing forecasts in the most efficient ways. The current ECMWF Integrated Forecasting System (IFS) is based on a 4-D Variational assimilation scheme to estimate the higher resolution forecast members combined with an ensemble data assimilation procedure to find ensemble members initial conditions (*Rabier et al.*, 1998; *Isaksen et al.*, 2010). Once the initial states are found, a Singular Vector scheme propagates the errors of the ensemble members over a 10-15 day medium-range forecast (*Buizza and Palmer*, 1995). A Variable Resolution approach has recently been integrated in the ECMWF Operational datasets to efficiently use computing power and provide more accurate short-range forecasts (*Buizza et al.*, 2007). Even though these models are characterized by initial condition errors, and errors arising from the model representation of the Atmosphere-Ocean system, they have demonstrated the potential benefits for decision making processes with a more reliable and cost-effective strategy than deterministic forecasts (*Saetra and Bidlot*, 2004; *Leutbecher and Palmer*, 2008). In this thesis we exploit the potential of ensemble datasets to develop novel statistical approaches that reduce the uncertainties presented above.

2.7 Conclusions

The present chapter reviewed the main sources of uncertainties in modelled and observational time series that challenge the analysis of long-term changes in ocean surface wind speed and wave height. It serves as an introductory review to guide the reader through the motivations underpinning the three projects developed in this PhD thesis. These projects investigate wind-wave climate long-term variations to understand whether we can demonstrate a connection between the changes in metocean extremes and climate change. We take advantage of an ensemble of model realizations of the atmosphere-ocean system to apply a novel EVA approach to ocean surface wind speed and wave height, first to an historical dataset of operational forecasts, and then to an ensemble of GCM-forced future projections. To achieve this, the present PhD is composed of three different but complementary projects. The results of the ensemble products for each project are discussed in light of their performance and compared with an extensive global dataset of satellite altimeter observations (*Young and Ribal*, 2019; *Ribal and Young*, 2019). The use of ensemble products is motivated by the various sources of uncertainties that characterise global atmosphere-wave models that we described in this chapter. It has been discussed and demonstrated (*Saetra and Bidlot*, 2004; *Cavaleri et al.*, 2007; *Morim et al.*, 2019) that a probabilistic approach with the use of ensembles is

a convenient way forward in the investigation of ocean wind and wave climate. The potential of such datasets is demonstrated and discussed throughout the present thesis. The ensemble products help to improve the confidence around past, present and future wind and wave climate. The next chapter is a preliminary analysis of 20^{th} century mean climate trends in ocean surface 10-m wind speed, U_{10} , and significant wave height, H_{s}. Here, we discuss if true or spurious trends can be derived from different 20^{th} Century model datasets.

Chapter 3

Trends from 20$^{\text{th}}$ century models compared to satellite altimeter observations

3.1 Introduction/Summary

The material presented in this chapter is an expanded version of *Meucci et al.* (2020a) publication (©American Meteorological Society. Used with permission).

The trends in marine 10-m wind speed (U_{10}) and significant wave height (H_{s}) found in two century-long reanalyses are compared against a model-only integration. Reanalyses show spurious trends due to the assimilation of an increasing number of observations over time. The comparisons between model and reanalyses show that the areas where the discrepancies in U_{10} and H_{s} trends are greatest, are also the areas where there is a marked increase in assimilated observations. Large differences in the yearly averages question the quality of the observations assimilated by the reanalyses, resulting in unreliable U_{10} and H_{s} trends before the 1950s. Four main regions of the world oceans are identified where the trends between model and reanalyses deviate strongly. These are the North Atlantic, the North Pacific, the Tasman Sea and the western South Atlantic. The trends at +24 hour lead time are markedly weaker and less correlated with the observation count. A 1985–2010 comparison with an extensive dataset of calibrated satellite altimeters shows contrasting results in H_{s} trends, but similar U_{10} spatial trend distributions, with general agreement between model, reanalyses and satellite altimeters on a broad increase in wind speed over the Southern Hemisphere.

3.2 Background

Long-term changes in ocean surface wind speed and wave height have received increasing attention as climate change impacts have become evident (*Young et al.*, 2011; *Mentaschi et al.*, 2018; *Luijendijk et al.*, 2018; *Young and Ribal*, 2019). Changes in surface wind speed and wave height may dramatically affect coastal communities (*Ranasinghe*, 2016), as well as offshore operations (*Bitner-Gregersen et al.*, 2018), but uncertainties still characterize our current knowledge of long-term variations in these quantities (*Rhein et al.*, 2013).

Trend analyses of global ocean surface wind speed and wave height rely on observational and modelled datasets. Long time series of surface wind speed and wave height observations originate from shipboard measurements that date back to 1854, the year marking the beginning of an internationally organised system for recording shipboard meteorological observations (*Maury*, 1853; *Cardone et al.*, 1990; *Woodruff et al.*, 1998). These datasets are mainly derived from Voluntary Observing Ships (VOS). Studies of VOS marine wind speed have shown that these datasets are affected by changes in measuring techniques. This is especially relevant in the first part of the century, where the absence of universal standards for estimated winds (formalization of sea-state equivalent Beaufort scale took place in 1946 (*Thomas et al.*, 2008)), may have permanently compromised trend studies of ocean surface wind speed over this period (*Peterson and Hasse*, 1987; *Ramage*, 1987). These findings were confirmed by *Cardone et al.* (1990) who voiced general skepticism on the possibility of removing the spurious trends caused by the presence of inhomogeneities before 1950. After 1950, VOS wind observations show an increasing trend in some areas of the oceans, due to a growing number of observations as well as changes in anemometer heights, instrument calibration and ship size (*Cardone et al.*, 1990). After bias correcting VOS wind datasets, an increasing signal still remains in the surface wind speed (*Thomas et al.*, 2008; *Tokinaga and Xie*, 2011), leaving open questions as to whether these residual trends are real or affected by other artificial factors (*Tokinaga and Xie*, 2011).

VOS wave height observations are less affected by changes in measurement approaches. However, the coding system had several changes, with the most significant happening in 1950 (*Kent et al.*, 2019). An extensive analysis of VOS wave height time series (*Gulev and Grigorieva*, 2004) found positive trends in the Atlantic from 1950 to 2002, but no significant trends on the total 1885–2002 period, again questioning the quality of the measurements in the first part of the 20$^{\text{th}}$ century. Upward trends in the Pacific (1950–2002) were found to be considerably weaker compared to the North Atlantic region (*Gulev and Grigorieva*, 2004). However, the analysis was limited to the main ship routes, with limited coverage in the Southern Hemisphere.

Buoy measurements started in the 1970s and cover only a small area of the

oceans. Trends in wind speed and wave height have been investigated from buoy measurements in some areas of the oceans such as the Northeast Pacific (*Allan and Komar*, 2000; *Gower*, 2002; *Ruggiero et al.*, 2010) and the Southern Ocean (*Hemer et al.*, 2010). However, these trend analyses may also be affected by spurious effects and step changes in the records due to changes in instrumentation, buoy hulls and measurement techniques (*Gemmrich et al.*, 2011; *Thomas and Swail*, 2011).

The advent of ocean-observing satellites in the late 1970s represents a fundamental step forward in the understanding of marine wind and wave climate. These datasets guarantee global coverage of the ocean, drastically increasing the amount of observations in previously poorly observed areas such as the Southern Ocean. Today these datasets constitute time series covering a period of over 30 years (*Young et al.*, 2011; *Young and Ribal*, 2019).

However, for trend analysis even longer datasets are desired. This would reduce the effect of natural variability (*Dobrynin et al.*, 2015; *Kumar et al.*, 2016), and help identify potential climate change signals (*Weisse*, 2010).

Reanalyses and hindcasts (without assimilation) have been widely used for both weather and climate applications at both regional and global scales (*Gregow et al.*, 2016) as they represent the best available interpolated datasets in space and time. Trend studies have also taken advantage of reanalyses' long and spatially homogeneous time series that cover the whole globe (*Semedo et al.*, 2011; *Bertin et al.*, 2013; *Wang et al.*, 2012, 2013; *Aarnes et al.*, 2015). The European Centre for Medium-Range Weather Forecasts (ECMWF) and the National Oceanic and Atmospheric Administration (NOAA) have recently extended reanalyses back to the beginning of the twentieth century (*Compo et al.*, 2011; *Poli et al.*, 2016; *Laloyaux et al.*, 2018). These datasets are interesting candidates for investigation of long-term wind and wave climate. However, changes caused by increasing numbers of assimilated observations throughout the years, may produce spurious trends (*Bengtsson et al.*, 2004; *Weisse*, 2010; *Aarnes et al.*, 2015; *Wohland et al.*, 2019). Inconsistent surface wind speed trends have been found between the ECMWF century-long reanalyses (ERA-20C and CERA-20C) and the twentieth century NOAA climate reanalysis (20CR) (*Compo et al.*, 2011). Considering the North Atlantic and North Pacific, *Wohland et al.* (2019) argue that spurious trends might be connected to the growing number of marine wind speed observations assimilated by the ECMWF reanalyses. Furthermore, the ECMWF ERA-20C reanalysis (*Poli et al.*, 2016) and the NOAA 20CR reanalyses were found to be inconsistent in the low frequency variability of the Northern Hemisphere winter seasons (*Befort et al.*, 2016).

To date, no studies have been carried out on long-term wave height trends of these reanalysis products. Waves are a combination of local wind-sea and swell coming from distant storms (*Young*, 1999b). Despite being entirely forced by the

wind field, the long-term trends of wave height may be affected by low-frequency variability, e.g. increasing number of cyclones, in the form of a swell contribution (*Young*, 1999a; *Gulev and Grigorieva*, 2006). This is also found in climate projections, where significant changes in extra-tropical swell can be found toward the end of the 21st century (*Shimura et al.*, 2016; *Breivik et al.*, 2019).

This work aims to investigate the relevance of trend analysis of 10-m surface wind speed, and significant wave height, obtained from the ECMWF ERA-20C (*Poli et al.*, 2016) and CERA20C (*Laloyaux et al.*, 2018) twentieth century reanalyses and compare with the ECMWF ERA-20CM (*Hersbach et al.*, 2015a) twentieth century atmosphere-wave model-only integration (i.e. without assimilation). To further investigate the relevance of U_{10} and H_s trends from twentieth century climate models and reanalyses, we compare the ECMWF century-long datasets, with and without assimilation of in-situ observations, with the trends of an independent dataset of calibrated satellite altimeter observations (*Ribal and Young*, 2019; *Young and Ribal*, 2019). The objective is to investigate twentieth century U_{10} and H_s long-term changes, and to evaluate the relevance of climate studies derived from twentieth century models. Furthermore, data assimilation impacts will be assessed by comparing with trends at +24 h forecast lead time (*Aarnes et al.*, 2015). This analysis provides a mechanism to determine the level of confidence in the climate change signal in ocean wind speed and significant wave height over the twentieth century.

Section 3.3 describes the datasets used for the trend analysis, and explains the approach used to calculate trend magnitude and statistical significance. Section 3.4 describes the century-long trends. Section 3.5 analyses the differences between each model representation of climate averages over the twentieth century. Section 3.6 concentrates on comparing the climate model and reanalysis trends with the satellite altimeters trends. Sections 3.7 and 3.8 discuss the possibility of a real climate signal and our confidence in the estimated trends. Finally, we include our recommendations on how to interpret trends from reanalyses.

3.3 Datasets and methodology

ECMWF (*Dee et al.*, 2014; *Buizza et al.*, 2018) and NOAA (*Compo et al.*, 2011) have recently developed state-of-the-art re-analyses that cover the whole twentieth century. The aim is to obtain a better understanding of the past to improve current weather forecasts and climate studies (*Buizza et al.*, 2018). ECMWF developed three different century-long datasets, (i) ERA-20CM (*Hersbach et al.*, 2015a), (ii) ERA-20C (*Poli et al.*, 2016) and (iii) CERA-20C (*Laloyaux et al.*, 2018). Table 3.1 describes the characteristics of the three models.

ERA-20C and CERA-20C are interesting in the context of trend analysis be-

Table 3.1: Characteristics of the ECMWF century-long models.

	ERA-20CM	ERA-20C	CERA-20C
Ensemble members	10	1	10
Assimilation atmosphere	Model-only integration	Marine wind, surface pressure	Marine wind, surface pressure
Atmosphere/waves/land	Coupled	Coupled	Coupled
Coupled ocean/ice	Prescribed HadISST2.1.0.0	Prescribed HadISST2.1.0.0	Coupled
Assimilation ocean	None	None	Temperature, Salinity
Radiative forcing	CMIP5	CMIP5	CMIP5
Horizontal resolution	atm. 125 km / ocean 166 km	atm. 125 km / ocean 166 km	atm. 125 km / ocean 110 km
Integrated Forecasting System (IFS)	Cy38r1 (2012)	Cy38r1 (2012)	Cy41r1 (2015)
Observation Feedback Archive (OFA)	None	Public	Public

cause in contrast to ERA-20CM, which is a model-only integration, they assimilate in-situ observations of surface pressure and marine surface winds. Neither of the reanalyses assimilates satellite or in-situ ocean wave height measurements. This allows us to compare their trends with independent satellite altimeter wave height measurements. The atmosphere horizontal resolution of the three model integrations is 125 km, whereas the ocean model resolution is 166 km for ERA-20CM and ERA-20C, and 110 km for CERA-20C. The models employ the same CMIP5 radiative forcing scheme. The 10-m surface wind speed, U_{10}, and significant wave height, H_s, are available as monthly averages of daily means, or as monthly averages of daily outputs at specific synoptic times. The latter have been selected for the trend analysis of the period from 1901 to 2010, and to perform a comparison with independent satellite altimeter trends for the overlapping 1985–2010 period. The main characteristics of each dataset used in this study are described in the following paragraphs in the context of trend analysis.

3.3.1 ERA-20CM

ERA-20CM (*Hersbach et al.*, 2015a) is a 10-member ensemble model-only simulation that covers the period 1899–2010. The model version is the Integrated Forecasting System (IFS) cycle 38r1[1], released in 2012. No observations are assimilated in ERA-20CM, thus the model is not able to accurately reproduce actual synoptic states. Despite this, ECMWF argues that the ERA-20CM model still performs well in terms of long-term climate, being consistent with other CMIP5 models. As noted also by *Dee et al.* (2014) the models require boundary conditions that are implemented from other models that in turn depend on observations from multiple sources. ERA-20CM boundary conditions include SST and sea-ice states from the Hadley Centre Global Sea Ice and Sea Surface Temperature data set version 2 (HadISST2) model (*Rayner et al.*, 2006), developed as part of the same ERA-CLIM European project.

[1]ECMWF Cy38r1: https://bit.ly/2OXuHFF

3.3.2 ERA-20C

ERA-20C (*Poli et al.*, 2016) is a single-member reanalysis for the period 1900–2010. It assimilates in-situ observations of surface pressure from the International Surface Pressure Databank (ISPD) version 2 (*Cram et al.*, 2015) and both marine surface pressure and surface wind speed from the International Comprehensive Ocean-Atmosphere Data Set (ICOADS) version 2.5 (*Woodruff et al.*, 2011). The model has the same atmospheric general circulation configuration as the control member of ERA-20CM and employs the same model version. Surface boundary conditions are also provided by HadISST. This allows for a direct comparison of the trend analysis between the two models. In well-sampled regions, the assimilation of observations adds precision at the synoptic scale, however, a negative impact has been observed on trends and low frequency variability (*Poli et al.*, 2016). Two main aspects of the model integrations will be investigated: the observational constraints and the realism of long-term trends.

3.3.3 CERA-20C

CERA-20C is the first atmosphere-ocean coupled climate reanalysis of the twentieth century (*Buizza et al.*, 2018). It is a 10-member ensemble and covers the period from 1901 to 2010 (*Laloyaux et al.*, 2018). It is based on the CERA coupled data assimilation system (*Laloyaux et al.*, 2016), and employs the ECMWF IFS cycle 41r1 of 2015. Like ERA-20C, it assimilates surface pressure and marine wind observations. In addition, it assimilates ocean temperature and salinity profiles. Coupled models have been shown to be crucial for synoptic weather representations. They are especially relevant in the prediction of tropical cyclones (*Mogensen et al.*, 2017), and for a consistent global transport of mass, water, and energy at the relevant time scales (*Dee et al.*, 2014). Despite these models being potentially the best available to reconstruct the earth climate system and thus wind and wave conditions, open questions remain as to the impact of data assimilation on long-term trends and how these compare to ERA-20CM and ERA-20C.

All models described are coupled to the same version of the ECMWF global third generation wave model WAM (*WAMDI*, 1988; *Janssen*, 2004), which resolves the two dimensional wave spectrum, taking into account advection, wind input, bottom friction, nonlinear interactions and dissipation due to white capping. At each time step the wave model interacts with the atmosphere as surface roughness is fed back to the atmospheric boundary layer scheme of IFS.

3.3.4 Observation Feedback Archive (OFA)

ECMWF produces an Observation Feedback Archive (OFA) (*Dee et al.*, 2014) where all observations used in the data assimilation are archived with the corresponding model values interpolated to the observation location (*Hersbach et al.*, 2015b). Here we investigate the feedback archive of ICOADS v2.5 (*Woodruff et al.*, 2011) surface pressure and wind speed and ISPD v2.2 (*Cram et al.*, 2015) surface pressure.

ICOADS v2.5

ICOADS is regarded as the reference long-term marine surface dataset (*Woodruff et al.*, 2011). The ICOADS monthly summary statistics have been produced at a $2° \times 2°$ resolution covering the period 1800–1960, and $1° \times 1°$ since 1960. There are 8 observed variables; sea surface temperature (SST), air temperature, wind speed, wind components, sea level pressure (SLP), total cloudiness and relative humidity. There are also 14 derived variables. No attempt is made to account for observing system changes and measurement biases (*Woodruff et al.*, 2011). Critical metadata such as instrument type and placement are used to improve accuracy. However, effects of changes in measurement techniques and the growing number of observations may still affect estimates of long-term trends in the dataset. In Table 3.2 the surface pressure and wind speed observation counts are listed by measurement technique. These are the observations that have been assimilated in the ERA-20C and CERA-20C reanalyses.

Table 3.2: ICOADS 2.5 dataset. Surface pressure and wind speed observation counts classified by report type ID as archived in the Observation Feedback Archive (OFA) of the ECMWF (*Hersbach et al.*, 2015b).

ID	Description	# pressure obs.	# wind obs.
16005	Drifting and mooring buoys	N/A	845712
16008	Ship	5593209	4845002
16049	Ocean station vessel on station	81830	120308
16050	Ocean station vessel off station	32842	55488
16051	Station or ship on ice	17022	7944
16052	Ocean bottle and low-resolution conductivity temperature depth CTD and XCTD	N/A	21734
16053	Mechanical or digital or micro bathythermograph MBT	N/A	33062
16054	Expandable bathythermograph XBT	N/A	4496
16055	Coastal-marine automated network CMAN	289524	12384
16056	Undulating oceanographic recorder UOR	7	380
16057	Fixed ocean platform or rig	20211	12256
16061	High resolution conductivity temperature depth CTD and XCTD	N/A	812

Fig. 3.1 shows the yearly counts classified by report type over the complete duration of the reanalysis. The vast majority of observations assimilated by ECMWF reanalyses were collected from ships, i.e. report type 16008. As shown in Fig. 3.1,

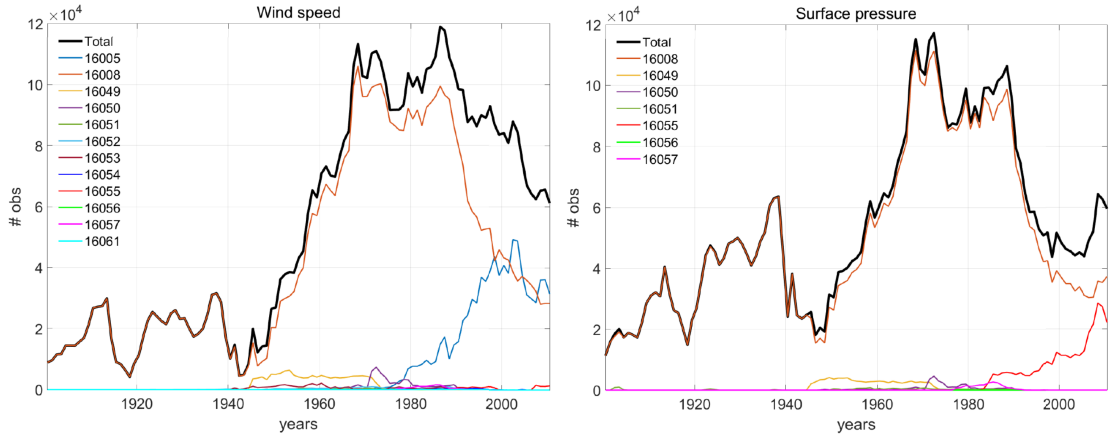


Figure 3.1: 1901–2010 ICOADS 2.5 surface wind and surface pressure observation yearly counts divided by report type, as archived in the ECMWF Observational Feedback Archive. Refer to Table 3.2 for the description of report type ID.

ship measurements started to decrease in the last part of the 20$^{\text{th}}$ century, and observation of the oceans now heavily relies on satellites and drifting buoys (*Kent et al.*, 2006). The observation counts derived from the OFA are crucial to understanding what are the main drivers of trend differences between models and reanalyses (*Wohland et al.*, 2019).

ISPD v2.2

The ISPD is the world's largest surface and sea-level pressure dataset. The dataset consists of observations from land stations, marine observations and tropical cyclone best track reports (*Cram et al.*, 2015). This dataset was first used in the NOAA twentieth century reanalysis, 20CR, and then also assimilated by the ECMWF ERA-20C and CERA-20C reanalyses.

Fig. 3.2 compares global observation counts of the three observational datasets considered here (ICOADS v2.5 surface wind, ICOADS v2.5 surface pressure and ISPD v2.2 surface pressure). The counts are binned on a $2° \times 2°$ grid. The comparison shows how the majority of ocean in-situ observations have been collected in the Northern Hemisphere with some isolated areas in the Southern hemisphere such as the Tasman sea, between Australia and New Zealand, and the South Atlantic ship routes that connect South American and African coasts to the Northern Hemisphere. Fig. 3.2a and Fig. 3.2b clearly show the main ocean ship routes. In the ICOADS wind speed observations plot (Fig. 3.2a) the tropical ocean mooring buoy array can be clearly observed. Fig. 3.2c shows the ISPD surface pressure observation density, mainly composed of land observations with localized areas of

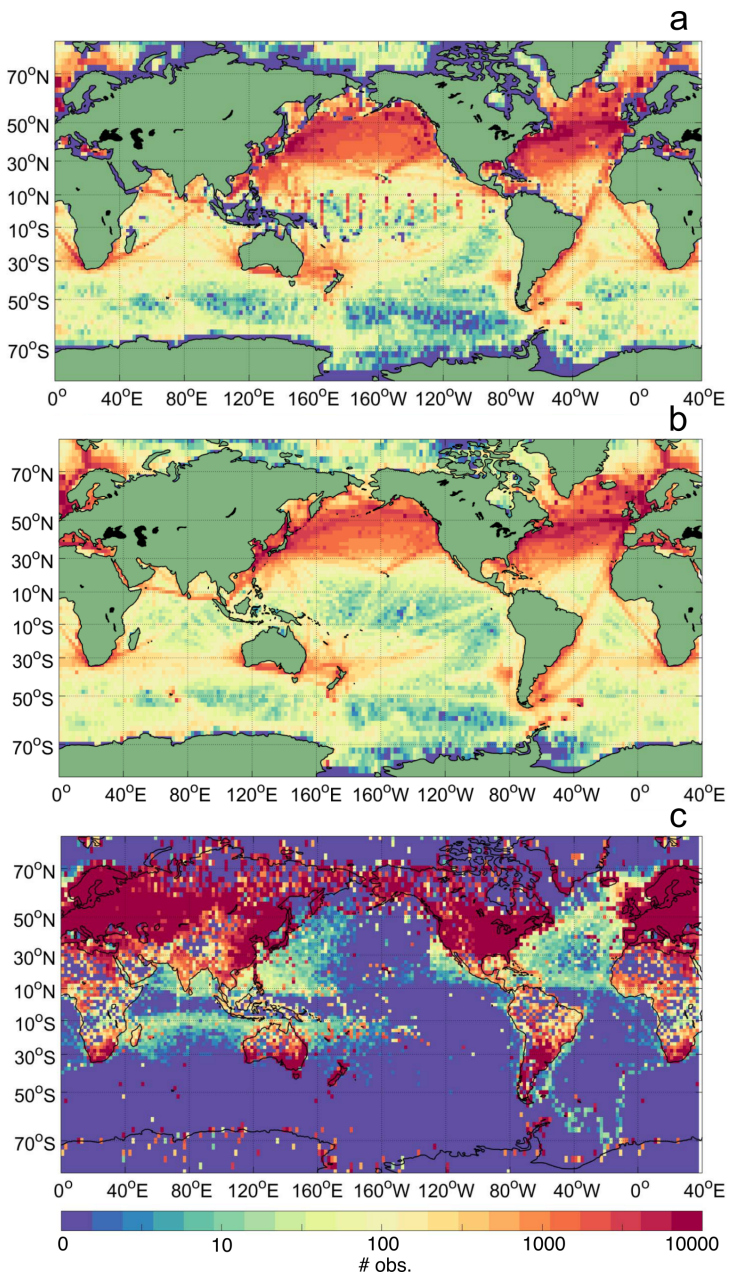


Figure 3.2: 1901–2010 OFA observation counts in base 10 logarithmic scale, binned on a $2° \times 2°$ grid. The three different datasets assimilated in ERA-20C and CERA-20C are: (a) ICOADS2.5 surface wind (b) ICOADS2.5 surface pressure (c) ISPD2.2 surface pressure.

ocean observations. Considering the distribution of the observation density shown in Fig. 3.2, we focus on the ICOADS dataset to investigate the data assimilation

impact on the model climate trends.

3.3.5 Satellite altimeters

To further assess long-term climate performance of the ECMWF reanalyses and model-only integrations, we compare the trends of the last part of the century-long model runs with the trends found from the largest available dataset of satellite measurements of U_{10} and H_s (*Young and Ribal*, 2019). This archive is calibrated and cross-validated against buoys from the United States National Data Buoy Center (NDBC) and independent buoy datasets (*Ribal and Young*, 2019). The calibrated satellite altimeter observations span the period 1985 to 2018. To compare with the ECMWF model integrations, we analyse the overlapping 1985–2010 window, noting that before 1991 the amount of altimeter data is relatively sparse, as only one satellite was operational. The horizontal resolution of this satellite altimeter dataset is $2^\circ \times 2^\circ$. The comparison is particularly interesting given the fact that the ECMWF century-long model integrations do not assimilate satellite observations to avoid impact of sudden changes in the assimilated observations (*Weisse*, 2010; *Sasaki*, 2016).

3.3.6 Trend analysis

To perform the trend analysis we select the monthly mean H_s and U_{10} values from the three ECMWF century-long models and the satellite altimeter observations. We apply a non-parametric trend analysis using the Theil-Sen estimator (*Theil*, 1950; *Sen*, 1968). Compared to a regression analysis, this method is more robust for non-normally distributed data such as in the case of H_s, and has the advantage of reducing the impact of potential outliers. The approach has previously been used to estimate trends of marine wind speed and significant wave height from model reanalyses (*Wang and Swail*, 2001; *Aarnes et al.*, 2015) and from satellite observations (*Young et al.*, 2011; *Young and Ribal*, 2019). The trend magnitude is found by selecting the median of the slopes (change per unit of time) computed for each month, between distinct pairs (i, j) of monthly mean values, as shown in Eq. (3.1),

$$\text{median}\left(\frac{X_j - X_i}{t_j - t_i}\right) \quad 1 \leq i \leq j \leq n \quad \text{with} \quad i \neq j, \tag{3.1}$$

where n is the number of years. X takes respectively the monthly mean values of H_s or U_{10}. The Sen's slope is computed for each month, and $t_j - t_i$ is the difference in time (years). The trend at each location of the globe is the median of the monthly slopes. To evaluate the significance of the trend at each grid point location, we perform a Mann-Kendall test (*Mann*, 1945; *Kendall*, 1948) adapted to account for

seasonality and serial dependence (*Hirsch et al.*, 1982; *Hirsch and Slack*, 1984). To test the serial dependence of monthly mean H_s and U_{10}, we computed the Pearson correlation between consecutive months (monthly lag-1 correlation) for the three ECMWF models considered in this study. The results show that two consecutive monthly mean values are globally correlated with a Pearson coefficient generally higher than 0.4. This is the case for both H_s and U_{10} monthly means from the three ECMWF models. Thus, serial dependence is accounted for in the seasonal Mann-Kendall trend test (hereinafter SKTT), with the covariance term introduced by *Dietz and Killeen* (1981).

3.4 Century-long trends

The H_s and U_{10} 1901–2010 trends from the three ECMWF twentieth century models are shown in Fig. 3.3 as percentage change per decade. Only the trends with 95% significance from the SKTT are shown. The ensemble model trends are found for the 10-member ensemble mean. Fig. 3.3 shows that the ERA-20C and CERA-20C reanalyses trends differ significantly from the ERA-20CM model trends. Two areas stand out with a significant positive trend in the ERA-20CM, the Southern Ocean high latitudes (5–10 cm/s/decade change in U_{10}; 1–2 cm/decade change in H_s) where the different sea-ice cover derived from the model boundary condition may play a role (Fig. 3.3b), and the western tropical Pacific (3–7 cm/s/decade for U_{10}; no significant trend in the H_s), where the strong recent La Niña episode affects the trend of the entire century (*de Boisséson et al.*, 2014). The trend in the Southern Ocean could be correlated with the different surface roughness related to the changes in sea-ice cover found by the SST/sea-ice conditions of HadISST2 (*Rayner et al.*, 2006). The H_s domain is masked at the latitudes of sea-ice cover (hatched regions in Fig. 3.3a). The H_s trends are spatially uniform compared to the U_{10} trends. A slight decrease in H_s ($\approx$-0.5 cm/decade) is depicted at mid-latitudes with the exception of the South Atlantic region. In contrast, the deterministic reanalysis ERA-20C shows mostly increasing trends with magnitudes in the Northern Hemisphere that exceed +1.75% per decade (up to 18 cm/s/decade change in U_{10} and 5–8 cm/decade change in H_s). The CERA-20C spatial trend distribution is very similar to the ERA-20C results, but with reduced magnitude (up to 8 cm/s/decade change in U_{10} and up to 5 cm/decade change in H_s). This could be related to the 10-member ensemble averaging process, or to the dampening effect of the coupled ocean model caused by the assimilation of subsurface ocean measurements (*Wohland et al.*, 2019). A significant increase in both H_s and U_{10} is found, especially in the Northern Hemisphere, for both ERA-20C and CERA-20C. The spatial distribution of the strongest trends found in these reanalyses is remarkably similar to the spatial distribution of the areas

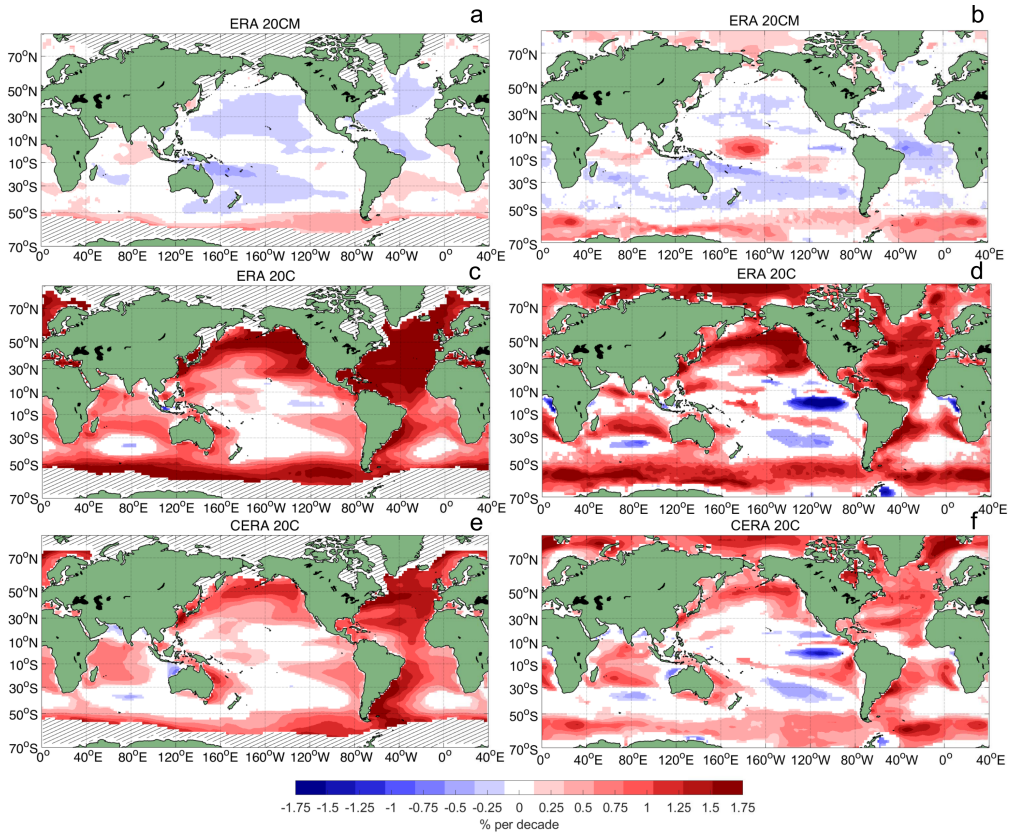


Figure 3.3: 1901–2010 percentage-per-decade trends. Comparison between the three twentieth century ECMWF datasets for H_s (left column) and U_{10} (right column). (a,b) ERA-20CM, (c,d) ERA-20C, (e,f) CERA-20C. Only the 95% significant trends from the SKTT are plotted in the figure. White regions represent areas where the trend is not statistically significant. Sea-ice cover areas that limit the H_s domain are hatched.

with the highest observation counts, as shown in Fig. 3.2. This is particularly true for the ICOADS dataset (Fig. 3.2a,b) and in agreement with the analysis by *Wohland et al.* (2019). Given the similarities between ERA-20C and CERA-20C trends, and assuming that the ensemble approach is more reliable in finding relevant trends, we concentrate our further investigations only on the comparison between the two ensemble models ERA-20CM and CERA-20C.

3.4.1 +24 h forecast lead time trends

To assess to what extent the differences in the CERA-20C trends are related to the assimilated observations, and how this is sustained through the model forecast

time, we performed a trend analysis at the +24 h forecast lead time (FC24). Note that, the +24 h lead time is the only forecast dataset available online for these models. *Aarnes et al.* (2015) used the +48 h forecast lead time to lessen the impact of data assimilation on the ERA-Interim trends. The percentage-per-decade trend

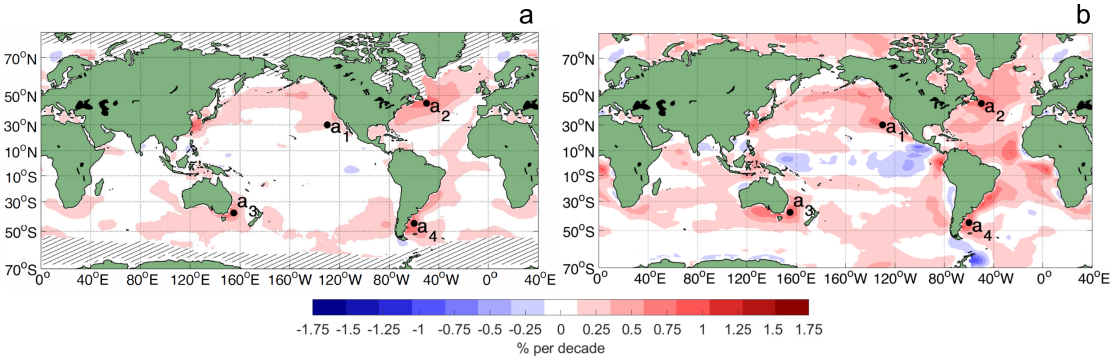


Figure 3.4: Difference in CERA-20C 1901–2010 percentage per decade trends between analysis time ANA and forecast lead time +24 h FC24. ANA - FC24: (a) H_s (b) U_{10}. Sea-ice cover areas that limit the H_s domain are hatched.

difference between CERA-20C at analysis time (ANA) and +24 h forecast lead time (FC24) is shown in Fig. 3.4 (ANA-FC24). The trend in the analysis is larger then the forecast (positive values in Fig. 3.4), especially in the areas where the main differences in the number of observations are found, i.e. in the Northern Hemisphere and in localized areas of the Southern Hemisphere (Fig. 3.2a,b). The differences are larger for U_{10} (Fig. 3.4b), because the wind speed is a parameter directly assimilated from the ICOADS2.5 dataset, whereas H_s is a derived variable (Fig. 3.4a). Again, the H_s spatial variability is more uniform compared to the wind speed due to the different nature of these two variables. Differences between analysis (ANA) and +24 h forecast (FC24) H_s are less pronounced, showing a smaller impact of data assimilation on H_s climate, but remarkably similar to the ICOADS pattern in Fig 3.2a.

To further investigate the areas where the observational datasets significantly impact the century-long trends, we select four locations with the largest differences between ANA and FC24 trends. The locations are shown in Fig. 3.4b. At these locations we compute the yearly averages for all three models. Fig. 3.5 shows the yearly average values of U_{10} and H_s for the reanalyses ERA-20C and CERA-20C (green and red lines) and the model-only integration ERA-20CM (blue lines) at ANA and FC24. Large differences are found in the first part of the century. At location a_1 the reanalyses show a change in the trend direction across the century, with an unusual decrease in the first part of the 20th century. Again, the comparison between ANA and FC24 confirms that the data assimilation effect



Figure 3.5: Yearly averages of monthly mean values at four different locations indicated in Fig. 3.4b. For ERA-20CM and CERA-20C the envelope of the ensemble values are plotted together with the mean. Column I: ANA H_s. Column II: FC24 H_s. Column III: ANA U_{10}. Column IV: FC24 U_{10}.

is reduced in the forecast case of U_{10} but not significantly for H_s. The FC24 ensemble spread is larger than the ANA spread as expected. Location a_3 in the Tasman sea shows a significant reduction in the average difference between the two reanalyses and ERA-20CM at FC24 compared to ANA. Fig. 3.5 generally reveals large differences in the first part of the 20th century. This difference is also present for H_s at FC24. However, the U_{10} ANA yearly average difference between models is significantly reduced at FC24.

3.5 Twentieth century climate

We compare the H_s and U_{10} yearly averages and investigate the differences between ERA-20CM model and CERA-20C reanalyses climate at the global scale. The normalized difference in the yearly averages is found as in Eq. (3.2),

$$d = \frac{\overline{x}^{\mathrm{ERA20CM}} - \overline{x}^{\mathrm{CERA20C}}}{\overline{x}^{\mathrm{ERA20CM}}}, \tag{3.2}$$

where $\overline{x}$ is the yearly average respectively of H_s or U_{10} monthly means. The differences are normalized over the ERA-20CM yearly averages to obtain a percentage difference. The results are shown in Fig. 3.6 for three different time periods:

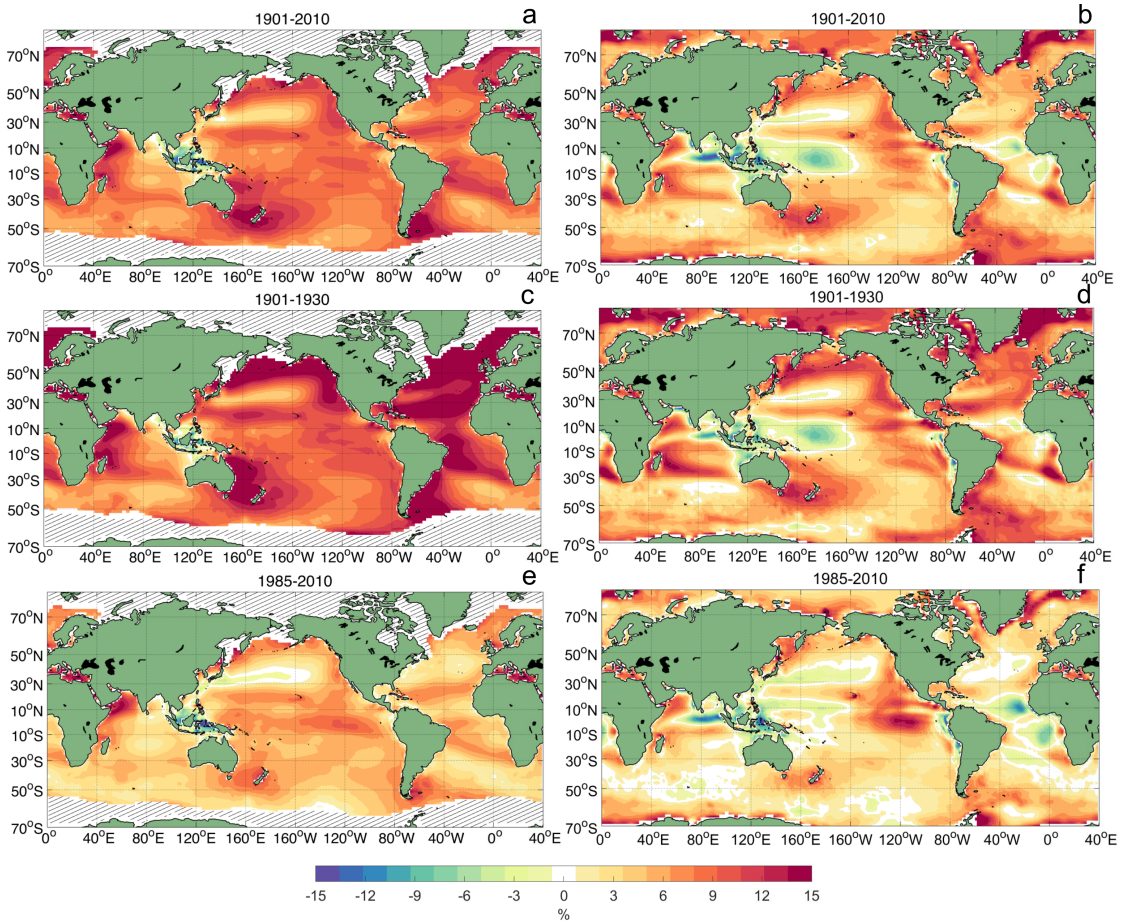


Figure 3.6: Annual averages of monthly mean values. Differences between ERA-20CM and CERA-20C (Eq. 3.2). H_s (left column) and U_{10} (right column). (a,b) 1901–2010, (c,d) 1901–1930, (e,f) 1985–2010. Sea-ice cover areas that limit the H_s domain are hatched.

1901–2010 (Fig. 3.6a,b), 1901–1930 (Fig. 3.6c,d), 1985–2010 (Fig. 3.6e,f). It is interesting to note that the yearly average H_s differences between the two models are generally higher than the U_{10} differences, showing that the impact of data assimilation of U_{10} and SLP have a marked effect on wave climate. In the first part of the century, 1901–1930 (Fig. 3.6c,d) the climate differences are larger, whereas the last part of the century, 1985–2010 (Fig. 3.6e,f), show a better agreement between the models. This is consistent with the location analysis performed in Fig. 3.5. The ERA-20CM climate is consistently higher for both H_s and U_{10} yearly averages throughout the century. Some areas show closer agreement, such as the 30° to 40° N latitude band in the Western North Pacific, where we find significantly smaller differences between the models than other regions. Initially, this singular pattern of similarity might be attributed to the average ship routes that follows higher latitudes on the great circle between Asia and North America. However, a more detailed analysis of observation counts shows that also at these latitudes the number of observations increases throughout the dataset time period. Given that ERA-20CM and CERA-20C both have SST boundary conditions dictated by the HadISST2 model, these similarities could be related to Sea Surface Temperature (SST) induced winds. This area is indeed impacted by the Kuroshio Extension phenomenon, characterized by strong seasonal variations of SST temperature that affect surface winds (*Nonaka and Xie*, 2003). These findings, if confirmed, could further demonstrate the impact of SST boundary conditions on modelled ocean surface climate. In general, H_s and U_{10} yearly average differences show similar spatial distribution. For the complete 1901–2010 dataset, the absolute H_s differences are between -0.15 m and 0.5 m, whereas, the U_{10} differences are between -1.8 m/s and 2 m/s. The H_s and U_{10} maximum absolute differences are found at the high latitudes. Larger discrepancies between ERA-20CM and CERA-20C are again found in the areas with most assimilated observation from the ICOADS dataset (Fig. 3.2a,b).

3.6 Model, reanalysis and satellite trends

Considering the large difference in H_s and U_{10} climate representation in the first part of the century, and to further investigate the ECMWF dataset performance in long-term climate analyses, we here compare the 1985–2010 model and reanalysis trends with trends calculated from calibrated satellite altimeter dataset (*Young and Ribal*, 2019; *Ribal and Young*, 2019). The trends are shown in Fig. 3.7 in terms of percentage change per decade. The dotted areas are regions where the trends are significant at the SKTT 95% level. Few locations show a statistically significant trend in H_s for both the satellite altimeter (Fig. 3.7a) and ERA-20CM (Fig. 3.7c), demonstrating general agreement between the datasets. In contrast, the

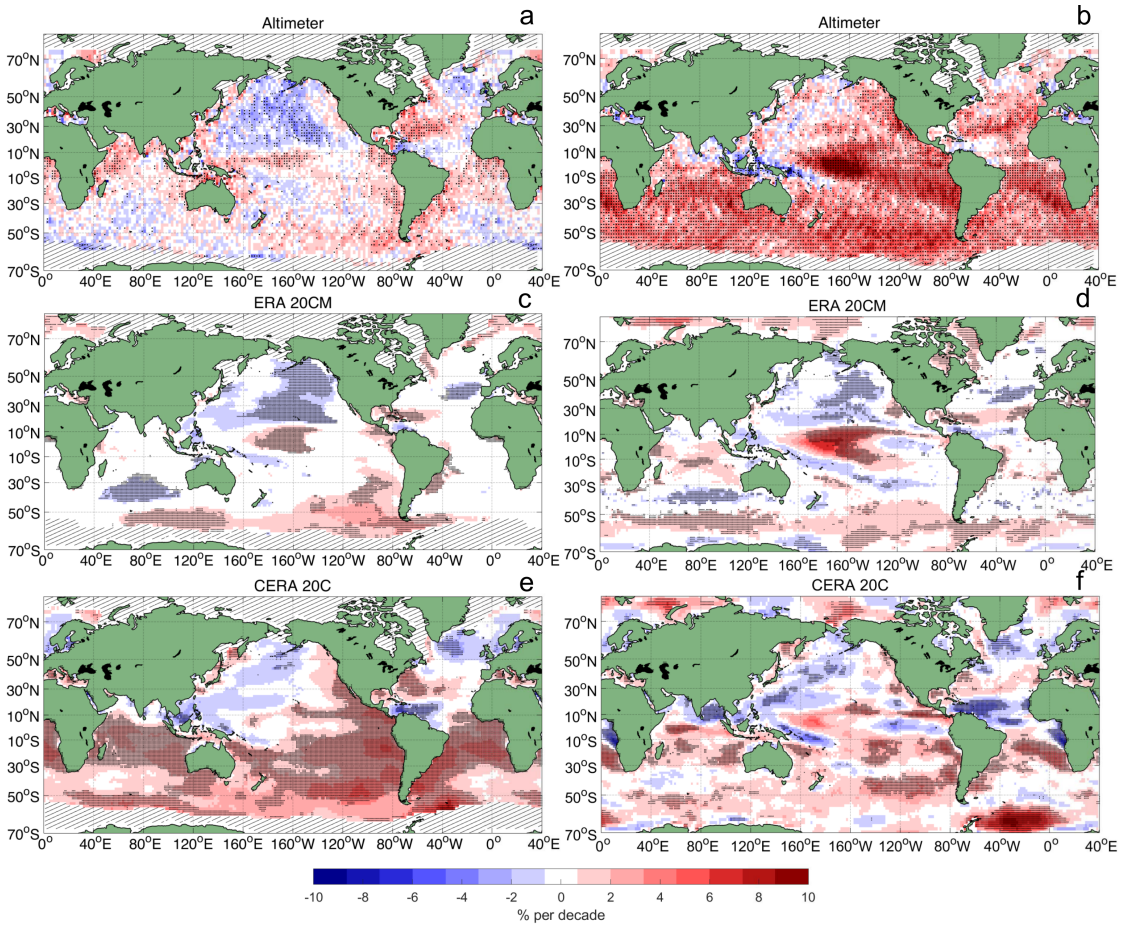


Figure 3.7: 1985–2010 percentage-per-decade trends. Comparison between altimeter dataset trends ($2^\circ \times 2^\circ$) (*Young and Ribal*, 2019; *Ribal and Young*, 2019), ERA-20CM and CERA-20C at analysis time ($1.5^\circ \times 1.5^\circ$). Left column (a,c,e) H_s trends. Right column (b,d,f) U_{10} trends. Areas of SKTT 95% statistical significance level are dotted. Sea-ice cover areas are hatched.

altimeter U_{10} trends (Fig. 3.7b) are largely significant in the Southern Hemisphere, whereas the ERA-20CM U_{10} trends (Fig. 3.7d), show only a few statistically significant areas of the oceans with lower trend magnitudes. Different U_{10} trends are also found in the North East Pacific. Here, the altimeters depict an increasing trend that is in contrast to the decreasing values found in ERA-20CM. However, the results do show similar spatial distribution for both H_s and U_{10} between the satellite altimeter and ERA-20CM results. For instance, the U_{10} altimeter and ERA-20CM trends (Fig. 3.7b,d) both show: a significant increase in the Central Pacific (*de Boisséson et al.*, 2014), a positive band across the Tropical Atlantic, and a general increase across the Southern Ocean. Also, both figures show an

increase in the central Indian Ocean and a reduction in the Southern Indian Ocean (agreement less clear). Also, the H_s altimeter and ERA-20CM trends (Fig. 3.7a,c), show: a small increase in the Southern Ocean west of South America, the increase in the central Pacific already noted for U_{10}, a decrease in the North Pacific and a slight increase in the Atlantic east of the Gulf of Mexico.

A greater spatial difference in distribution is found between the altimeter trends (Fig. 3.7a,b) and the CERA-20C trends at analysis time (Fig. 3.7e,f). Compared to both satellite and ERA-20CM results, the CERA-20C H_s trends (Fig. 3.7e) show larger differences than the U_{10} trends. CERA-20C results show a generally increasing U_{10} trend in the Southern Hemisphere (Fig. 3.7f) that is further amplified in the H_s trends (Fig. 3.7e). The reason for the large trends in the H_s coupled CERA-20C reanalysis in the Southern Hemisphere is not clear. We speculate that the large positive trends in southern hemisphere mid latitudes (Fig. 3.7e) may be associated with an increase of swell propagating from the Southern Ocean. However, this possibility requires further analysis of, for instance, changes in wave period.

Although there are differences in magnitude, all three datasets compared here, show increasing U_{10} in the Southern Hemisphere, with similar spatial distribution. The analysis of the CERA-20C FC24 trends (not shown here) does not differ significantly from the results obtained at ANA shown in Fig. 3.7e,f. This means that data assimilation for the period 1985–2010 is not affecting U_{10} trends as strongly as is the case when the whole twentieth century is considered.

3.7 Discussion

The present work found generally inconsistent twentieth century H_s and U_{10} trends for an ECMWF model-only integration (ERA-20CM) and reanalyses (ERA-20C and CERA-20C). This is also in agreement with previous studies that compared the ECMWF reanalysis datasets with the NOAA twentieth century reanalysis (20CR) (*Befort et al.*, 2016; *Wohland et al.*, 2019). NOAA 20CR assimilates ISPD surface pressure measurements but does not show any inconsistencies in trends compared with ECMWF's model-only integration, ERA-20CM (*Compo et al.*, 2006, 2011; *Befort et al.*, 2016; *Wohland et al.*, 2019). This suggests that the main impact on trends might be caused by the ICOADS dataset assimilation in the ECMWF reanalyses. It is not clear if the assimilated ISPD surface pressure measurements (Fig. 3.2c) also have a significant impact on reanalysed U_{10} and H_s trends.

It seems at first counter-intuitive that the difference between the free-running ERA-20CM and the two reanalyses, ERA-20C and CERA-20C, is greatest in the early period where observations are scarce. The reason is, as *Laloyaux et al.* (2018) explain (and first noted by *Poli et al.* 2015), that the observation error for pressure

observations in ERA-20C were too small at the start of the century, putting too much weight to the observations located in the subtropical high pressure belt. This generated large positive surface pressure increments over the unobserved Antarctic region. The same effect was found over the Arctic, but not as strongly as in the southern hemisphere. To alleviate this, the observation errors were increased in CERA-20C for the early 20th century and then allowed to decrease with for the more recent, data-rich period (J. Nicolas, pers. comm., 2019, and *Laloyaux et al.* 2018, Fig. 11). We note here, however, that there remains a large discrepancy between ERA-20CM and CERA-20C in all locations investigated, both northern and southern hemisphere, and the improvement over ERA-20C is only marginal in terms of surface wind speed (see Fig. 3.5, third column).

This work found large positive trends for ERA-20C and CERA-20C reanalyses in areas that correspond to the highest number of observation counts of ICOADS2.5 assimilated data (Fig. 3.2). Further analysis at four locations around the globe (Fig. 3.5), shows that the quality of the observations may be as important as the increasing number of observations in affecting potential spurious trends, especially in the first part of the 20$^{\text{th}}$ Century, where the U_{10} and H_{s} monthly averages significantly differ between reanalyses and the climate model. This is particularly true for the first half of the century (Fig. 3.5), where the main average differences are found between the ERA-20CM and the two reanalyses ERA-20C and CERA-20C. Since the observations in this part of the century originate only from shipboard observations (Fig. 3.1), we believe that the quality of the measurements, already questioned in reference studies (*Ramage*, 1987; *Peterson and Hasse*, 1987; *Cardone et al.*, 1990), have a major impact. As a result, it may not be possible to extract reliable trend estimates from reanalysis that extend to the period before 1950.

ERA-20CM seems to perform better than the reanalyses in resolving the spatial distribution of H_{s} and U_{10} long-term trends. As pointed out by *Hersbach et al.* (2015a), ERA-20CM outperforms reanalyses such as ERA-20C at large time scales (more than a year) thus being a potential reference model to analyse H_{s} and U_{10} trends over the last century. Although the ERA-20CM spatial trend distribution proved to be similar to altimeter significant wave height trends, it should be noted that the ERA-20CM realizations, especially in variation in temperature, are closely related to the spread in the HadISST2 ensemble (*Hersbach et al.*, 2015a). That is, the sea surface temperature boundary condition may be a significant contributor to this observed trend. Indeed, we found similarities between ERA-20CM model-only integration and CERA-20C coupled model reanalysis in the yearly average climate for regions such as the Kuroshio Extension area (Fig. 3.6), where SST-induced winds might explain the effect of the SST boundary conditions on H_{s} and U_{10} in the reanalyses. The North Pacific, the North Atlantic, the Tasman sea and the South Atlantic coast of Africa and South America, are regions characterized by

large discrepancies between model and reanalyses trends (Fig. 3.3), suggesting a significant impact of assimilated ICOADS observations (Fig. 3.2a,b) on long-term climate analysis of these areas.

The 1985–2010 comparison with satellite altimeter trends further investigated the ECMWF datasets ability in representing long-term climate variations in H_s and U_{10}. The comparison confirmed the superior quality of the ERA-20CM model-only integration in describing the spatial distribution of H_s trends compared to the CERA-20C reanalysis. The CERA-20C H_s trends found in the Southern Hemisphere mid-latitudes are somewhat confusing. Model, reanalyses and satellite altimeter observations agree on a general increase of U_{10} over the Southern Hemisphere, in the last part of the twentieth century. However, only the satellite trend results are consistently statistically significant, whereas ERA-20CM and CERA-20C are only sparsely statistically significant at the SKTT 95% level.

Both model and satellite wind and wave time series have limitations. However, these are invaluable tools to understand the global wind-wave climate, and it is fundamental to continue to compare and analyse independent observational and model datasets. It is not possible to identify a general best performing dataset to estimate wind-wave climate trends. However, we might argue that the absence of data assimilation in ERA-20CM, and the larger time span, if compared to the satellite observations, makes this model the more suitable for the analysis of century-long trends. Furthermore, knowledge of the performance and structure of these datasets, may provide insight as to the most appropriate model results for particular areas of the oceans, or for a specific time period. Also, these tools may benefit practical applications if their limitations with respect to trend estimates are understood.

The contrasting H_s and U_{10} trend results found in this study, show that the connection between the wind and wave climate trends is not as straightforward as one might think. The model wave height trends may significantly differ from the wind speed trend estimates. Future wind and wave climate trend studies should consider this aspect of the global models.

3.8 Conclusions

Twentieth century reanalyses show spurious trends in 10-m surface wind speed, U_{10}, and significant wave height, H_s throughout the whole 1901–2010 period. The increasing number of observations, as well as changes in the quality of the data ingested in these reanalyses, mainly affect four areas of the oceans where H_s and U_{10} trend analyses show the largest discrepancies between the model-only integration and the reanalyses. These are the North Atlantic, the North Pacific, the Tasman Sea and the South Atlantic region east of South America. The com-

parison with satellite altimeter trends showed agreement on a general increase of Southern Hemisphere U_{10}. Considerable care must be exercised when considering the assimilation of surface wind measurements before the 1950s. A comparison with the newly available ERA-5 reanalysis (from 1950 until today), may further help detecting data assimilation impacts, and investigate the nature of H_{s} and U_{10} climate trend estimates from state-of-the-art reanalyses.

Chapter 4

Wind and wave Extreme Value Analysis

4.1 Introduction

The first part of this thesis dealt with model and observational time series of the ocean surface wind speed and wave height. In particular, we focused on the constraints of these datasets that limit an accurate representation of the long-term average and extreme climate. In chapter 3 we described a published work (*Meucci et al.*, 2020a) on ocean surface wind speed and wave height past Century climate trends, that shows what are the challenges in estimating physically meaningful climate trends of metocean variables.

This thesis focus now shifts to wind and wave Extreme Value Analysis (EVA). Two published works are presented in chapter 5 (*Meucci et al.*, 2018) and chapter 7 (*Meucci et al.*, 2020b) on a novel approach to the estimation of wind speed and wave height extremes, that responds to increasing interest by climate scientists and engineers in extreme events. These are the events that most affect human life and activities, resulting in large social and economic losses. The IPCC Special Report on Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaptation (SREX) (*Field et al.*, 2012) describes these events as disasters. The report introduces the concept of risk, as a measure of the severity of climate extreme impacts. The objective is to classify and define the areas of interest for disaster risk management strategies and policies. Disaster risk management depends on three different interconnected factors as shown in Fig. 4.1, adapted from the IPCC SREX report (*Field et al.*, 2012).

These are weather and climate events (with a particular focus on extremes), exposure, and vulnerability. Exposure means the presence of people, infrastructure, economic, or cultural assets. Vulnerability is the susceptibility of being affected

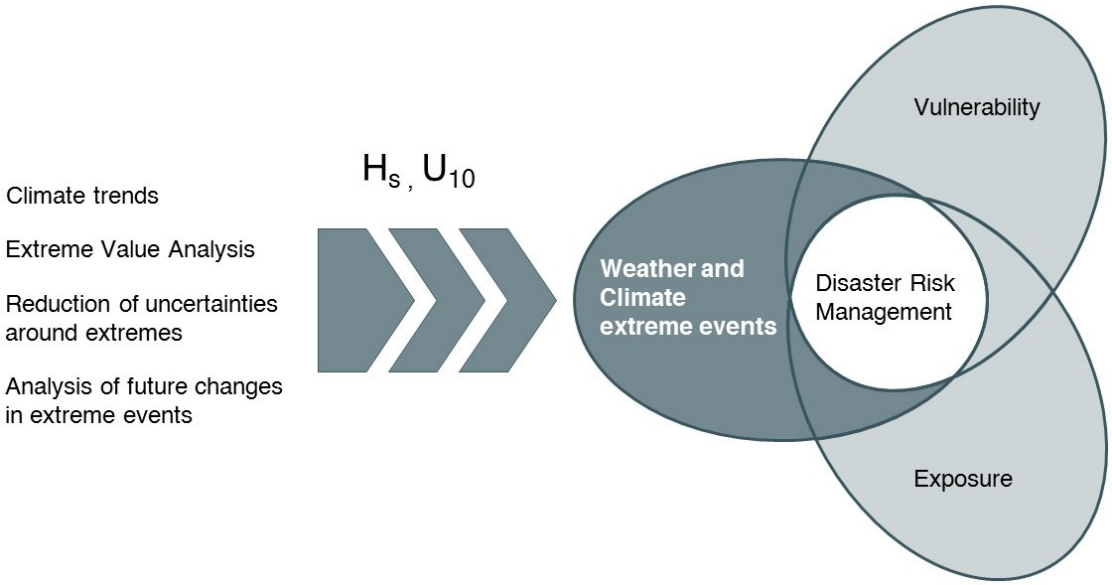


Figure 4.1: Target objective of the present thesis work following a scheme adapted from *Field et al.* (2012).

by extreme events. As an example, communities that live in coastal or delta areas are exposed to the effect of sea-level rise or extreme storm surges. Furthermore, they are even more vulnerable if they live in economically underdeveloped regions such as the very populous delta areas of countries like Indonesia or Bangladesh.

The objective of scientists and engineers in the metocean field, and other areas of research and applications, is to develop sustainable and resilient strategies for disaster risk management policies. These measures must respond to society's needs and be resilient to possible climate change effects. Adaptation strategies are needed to answer the pressing questions and challenges posed by a changing climate. This is particularly crucial for extreme events, the main cause of disasters.

To achieve this we must act on the three factors shown in Fig. 4.1. The present study contributes to this, producing robust EVA studies to estimate extreme weather and climate events. As in chapter 3 of the work, the focus is on the H_s and U_{10} parameters (Fig. 4.1). However, producing robust and reliable extreme estimates is challenging, as extreme events are rare, and few data are available to make assessments regarding changes in their frequency or intensity (*Field et al.*, 2012). Thus, we need to infer the behaviour of parameters of a population, based on limited random samples of the same population. This is done with statistical modelling techniques that are capable of robustly predicting the variability around the mean of a population. However, this might result in large uncertainties in inferring the probability levels of the extremes. This is because

the available datasets to perform the analyses are much shorter than the return period—once in N years events— sought, for common engineering practice (50, 100 years or more). To cope with these limitations, Extreme Value Theory (EVT) has developed and evolved significantly in the past few decades.

This chapter provides an introduction to univariate statistical analysis tools for EVT (or EVA). Univariate means that a single variable is considered in the analysis, as opposed to multivariate techniques that consider a set of different variables that interplay and affect each other. The chapter is propaedeutic to the next two research studies developed in the course of this PhD to estimate ocean surface wind speed and wave height extremes, and their changes due to climate change. After a brief discussion of the statistical basis, we direct attention to metocean extremes and a novel ensemble EVA approach implemented throughout the rest of the thesis.

4.2 Univariate statistical analysis

Statistical analyses infer the behaviour of a specific population based on a limited sample, or random variable, X. To each event of a random variable we can associate a probability, thus constructing a probability distribution $F(x) = Pr\{X < x\}$, where x is each measured value in the sample space Ω of the random variable X. In EVT we assume that the random variable is continuous; this means that a probability density function exists. The probability density function is the derivative in space of the probability distribution, $f(x) = dF/dx$.

Most common statistical approaches are based on a number of preliminary assumptions that simplify the analyses. These are, the variables have to be independent and identically distributed (i.i.d.), and the dataset time series stationary. This is never exactly the case for weather, climate or ocean variables. Nonetheless, these are commonly accepted assumptions if preliminary tests are performed in their support.

In the following sections, we describe univariate statistical analyses and the connected assumptions used throughout the next part of the thesis. We acknowledge that this chapter is not exhaustive of the currently known theory around extreme value estimates, but we believe this to be sufficient for this work. The aim is to describe a novel approach using independent realizations of the same population drawn from model ensemble products. We describe this novel approach further in details in chapter 5 methodology section 5.4.

4.2.1 Basic assumptions

Independent and Identically Distributed continuous Random variables

Two random variables are said to be i.i.d. when the outcome of one does not affect the probability distribution of the other (*Coles et al.*, 2001). This means that their joint density function factorizes as in Eq. 4.1.

$$f_{X_1,X_2}(x_1, x_2) = f_{X_1}(x_1) f_{X_2}(x_2) \tag{4.1}$$

This can be extended to a set of variables defined as mutually independent if their joint density function factorizes as shown in Eq. 4.2.

$$f_{X_1,...,X_k}(x_1, ..., x_k) = \prod_{i=1}^{k} f_{X_i}(x_i) \tag{4.2}$$

The correlation of independent random variables is zero. However, as mentioned this is seldom the case for meteorological or ocean variables, where low correlation can exist, but never zero. Throughout this thesis we test at what level the variables are i.i.d. comparing variables with scatter and quantile-quantile plots. These are commonly used model diagnostics that justify the use of a specific statistical model.

Stationary processes

To define stationarity we need to introduce the concept of a random process, that is a sequence of random variables $X_1, ..., X_k$. A random process is stationary if the joint distribution of a subset of a random variable viewed n points later in time, remains unchanged (*Coles et al.*, 2001). This may be a controversial assumption in EVA, as scientists agree that climate is not stationary. Thus, in each of the studies performed we explain the stationarity assumption, and how this is justified for the results we are aiming for.

4.2.2 Statistical modelling

Statistical modelling techniques are used to infer the characteristics of a population, based on random samples of the same population. Statistical models can be parametric or non-parametric. Here we focus on parametric statistical modelling and use them to find a family of functions consistent with the random variable realizations $x_1, ..., x_k$. The function identified as the most consistent with the data is called the *estimator*, and the particular value obtained from this function is called the *estimate*. This is never a precise procedure as the estimate is never the exact

value of the population. This is an important source of uncertainty in determining weather and climate extremes, that adds to both modelled and observational uncertainties discussed in chapter 2.

To understand how the estimator is biased against the population, the bias or the mean-square error are common reference values, as well as the standard error to measure the accuracy around the estimate. The bias can be identified and corrected. The standard error is more of a true measure of how precise our statistical model is and therefore it must be kept low.

Following is a brief description of the available methods of unknown parameter estimation, followed by the most common ways of computing confidence intervals. This result will be useful when considering the next chapters methodology and results.

Estimation methods

There are a number of available methods to estimate parameters. The most commonly used are: the method of moments, the maximum likelihood method, the least square method, the probability-weighted moments or L-moments, and graphical methods (*Galambos*, 1994). In this thesis, we apply the most general and robust method, as well as the most commonly applied: the maximum likelihood method (*Coles et al.*, 2001). The objective is to find the unknown true value, θ_0, of a parameter θ that defines the density function. We call a likelihood function the probability of the observed data as a function of θ. The method maximizes the likelihood function to estimate the true parameter θ_0 of a distribution of data.

The maximum likelihood estimates are found with numerical methods. Model diagnostics graphical methods such as quantile-quantile plots are used throughout this thesis to evaluate the accuracy of the parameter estimates. Also, the maximum likelihood estimates are compared to novel ensemble approach results from direct in-sample estimates.

4.2.3 Confidence intervals

To quantify the precision of an estimator we can compute a confidence interval, that is a range of values within which we have a specified level of confidence that our parameter value would fall. Confidence intervals can be computed by previously assuming a pivot function ϕ of θ_0 whose distribution does not depend on θ_0, so that for any value of $0 < \alpha < 1$ we have (*Coles et al.*, 2001):

$$\Pr\{\phi_l \leq \phi \leq \phi_u\} = 1 - \alpha \tag{4.3}$$

or rewritten as a function of θ:

$$\Pr\{\theta_l \leq \theta \leq \theta_u\} = 1 - \alpha \tag{4.4}$$

The range varies according to the level of confidence, 1 - α, we are seeking. Many practical applications aim at a 95% level of confidence ($\alpha = 0.05$) which means that if we re-run our statistical model a hundred times we will have an estimate that will, on average, fall 95 times out of 100 in the interval computed. This is the level of confidence commonly used for metocean extreme analyses (*Caires and Sterl*, 2005; *Caires*, 2011).

Confidence intervals can be computed also with a bootstrap method. This is a non-parametric technique that re-samples multiple times, usually between 500 to 1000 times is sufficient, a set of data (sample), and obtains the upper and lower bound of a range of values around the extreme estimate (*Breivik and Aarnes*, 2017). In this thesis, we use this methodology as it has the advantage of no underlying function assumption.

4.3 Extreme Value Analysis

Extreme Value Theory (EVT), or Extreme Value Analysis (EVA), studies the statistical behaviour of extremes. This branch of statistics has received increasing interest in the past few decades as it has numerous applications in many scientific and technical fields (*Galambos*, 1994; *Castillo*, 2012).

EVA is used to find the probability of exceedance of extreme events. This is useful not only in Ocean Engineering, the topic of the present thesis, but in many other disciplines such as structural engineering, where return level extreme loads are needed to design structures, or hydraulic engineering, where the EVA focus is on estimating the one in N years flood event, to safely plan cities along rivers or in delta areas. In meteorology EVA is applied to estimate rainfall extremes that might significantly affect large populated areas or wind speed extremes that might result in dangerous loads on structures. Other EVA approaches are found in material and fatigue strength studies, or applied to vehicle traffic data to understand how often we experience peak numbers of cars in a specific area, or also to internet traffic data. In the pharmaceutical industry, and for pollution extreme estimates, EVA is used to understand dangerous rare events.

In this chapter we focus on the EVA approaches commonly used for metocean applications. In particular, we describe the two main approaches to estimate extremes, as presented by *Coles et al.* (2001). These are asymptotic models and threshold models. Also, we dedicate a small section to the Initial Distribution Method (IDM), which is not an EVA, but is still used in some practical ocean engineering applications, where data from extremes are mostly missing. We analyse the IDM and discuss its many limitations when used to estimate extremes.

For a complete theoretical explanation of the EVA the reader should refer to *Coles et al.* (2001), *Castillo* (2012) or *Dey and Yan* (2016). The works of *Galambos* (1994) and *Castillo* (2012) also provide a detailed list of the main domains of application of EVA.

4.3.1 Asymptotic models

Asymptotic models study the statistical behaviour of the maxima (or minima) of random variables (Eq. 4.5).

$$M_n = max\{X_1, ..., X_k\} \tag{4.5}$$

As we approximate sample means by a normal distribution following the central limit theorem, analogously at the extremes we use the extremal types theorem to approximate the extreme distribution to a family of functions. The extremal types theorem affirms that re-scaled i.i.d. sample maxima $M_n^* = (M_n - b_n)/a_n$, based on a sequence of constants $\{a_n > 0\}$ and $\{b_n\}$ converge to three distribution families: the Gumbel, the Frechet and the Weibull. These can be combined into the Generalized Extreme Value (GEV) distribution function family (Eq. 4.6).

$$G(z) = \exp\left\{ -\left[1 + \xi\left(\frac{z-\mu}{\sigma}\right)\right]^{-1/\xi}\right\}, \tag{4.6}$$

where $\mu \in \mathbb{R}$ is the location parameter, $\sigma > 0$ is the scale parameter, and $\xi \in \mathbb{R}$ is the shape parameter. The parameters can be estimated by one of the various statistical models mentioned in section 4.2.2. To estimate the extremes, quantile data are blocked into sequences. This methodology is called Block Maxima. Usually, in weather and climate extremes statistical analysis, a common way to sequence data is to select the Annual Maxima. This is, one value is selected for each year, Annual Maxima Method (AMM).

By inverting Eq. 4.6 we can estimate the return level z_p, associated with the return period $1/p$ as in Eq. 4.7.

$$z_p = \begin{cases} \mu - \frac{\sigma}{\xi}\left[1 - \{-\log(1-p)\}^{-\xi}\right], & \text{if } \xi \neq 0 \\ \mu - \sigma\log(-\log(1-p)\}, & \text{if } \xi = 0 \end{cases} \tag{4.7}$$

These are the values of interest in designing structures or planning operations to mitigate risk. For a desired return period, for instance, 50 or 100 years, we find the probability level, and from the fitted function, we obtain the design extreme value estimate, such as a design sea state. These are reference values for engineering practice. Probabilities of occurrence are determined according to the time interval. For instance, the probability of occurrence of the 100-year event is 1% for any given

year. Conversely, the probability of exceedance of the 100-year return value in a 100 years' time is 63% and not a 100% (Table 1.1 in *Nott* (2006)).

Although asymptotic models are attractive because they are easy to implement, the sequencing technique imposes a trade-off between large biases in the estimates, if blocks are too small, or too large variances, if blocks are too big. Furthermore, the sequencing technique wastes a lot of useful information that could otherwise be used to increase the accuracy of our estimates. This is the reason why the popularity of threshold models significantly increased after the pioneering work of *Pickands* (1975). The following section describes the threshold model approach to EVA.

4.3.2 Threshold models

Threshold models select i.i.d. extreme event realizations of a random variable over a chosen threshold u. This way, for a sufficiently large value of u, the distribution of $(X - u)$, conditional on $X > u$, approximately follows the Generalized Pareto (GP) family, Eq. 4.8 (*Coles et al.*, 2001):

$$H(y) = 1 - \left(1 + \frac{\xi y}{(\sigma + \xi(u - \mu))}\right) \tag{4.8}$$

A modelling technique, such as the maximum likelihood method, can be used to estimate the values of the GP ($H(y)$) function parameters. As was done for the asymptotic models, the return levels can be computed as the probability of exceedance of a value x, this time conditional on $X > u$ (Eq. 4.9).

$$\Pr\{X > x \mid X > u\} = \left[1 + \xi\left(\frac{z - \mu}{\sigma}\right)\right]^{-1/\xi}, \tag{4.9}$$

thus factorizing $\Pr\{X > u\} = \zeta_u$:

$$\Pr\{X > x\} = \zeta_u \left[1 + \xi\left(\frac{z - \mu}{\sigma}\right)\right]^{-1/\xi} \tag{4.10}$$

We can now compute the level x_m that is exceeded on average once every m observations for $\xi \neq 0$ (Eq. 4.11).

$$x_m = u + \frac{\sigma}{\xi}\left[(m\zeta_u)^{\xi} - 1\right], \tag{4.11}$$

This expression is commonly applied to find the N years return level (Eq. 4.12), where $m = N \times n_y$ where n_y is the number of observations per years.

$$z_N = \begin{cases} u + \frac{\sigma}{\xi}\left[(N n_y \zeta_u)^{\xi} - 1\right] & \text{if } \xi \neq 0 \\ u + \sigma \log(N n_y \zeta_u) & \text{if } \xi = 0 \end{cases} \tag{4.12}$$

To find the return level (or return period) value we need to substitute the three values of the scale, shape and location parameters obtained from the best function fit to the extreme data.

The recommended practice for extreme wave analysis (*Mathiesen et al.*, 1994), also described in *Holthuijsen* (2007) and used in *Vinoth and Young* (2011), *Meucci et al.* (2018), and *Takbash et al.* (2018), finds the return level as in Eq. 4.13.

$$\Pr\{x < x_{RP}\} = 1 - \frac{N_y}{RP \times N_{POT}}, \tag{4.13}$$

where N_y is the number of years in the dataset, N_{POT} is the number of peaks over the selected threshold and RP is the return period, which for this thesis has been chosen as 100 years. The extreme estimate is found by substituting the probability level into the fitted function.

Threshold approaches have their limitations. A challenging aspect of such models is the selection of an appropriate threshold. The choice of a threshold is a trade-off between bias and variance as it is for the choice of block size in the block maxima asymptotic models. Some techniques have been proposed to select the value of u, such as the mean residual life plot (or mean excess plot) (*Coles et al.*, 2001), but a best strategy still does not exist. A common approach is to keep the threshold u as low as possible to provide a model that better resembles the data. However, special care needs to be taken in this procedure as too low a threshold may violate the asymptotic assumptions (*Dey and Yan*, 2016). We can test how stable the threshold choice is by fitting a GP function to different time series constructed with varying thresholds. This way, a sufficiently stable threshold level might be found.

4.4 Wind and wave EVA

The design and operation of marine systems rely on accurate statistical analyses of historical datasets that provide the practitioner with extreme value return period estimates, that refer to the value exceeded once in N years. In non-polar regions wind and wave loads define the main stresses for structures and ships, and an accurate estimation of their extreme value estimates associated with given return periods is crucial for a cost-effective and safe design of structures and ship operations. Wind and wave extremes also affect coastal areas. Common values in the design of coastal defences are the 50-year or 100-year return period. A robust

estimate of these return period values is important as in 2010, 290 million people worldwide lived below the 100-year flood level and US$9600 billion of assets were exposed to inundation (*Hinkel et al.*, 2014). In some locations of the world, such as the Netherlands, practitioners and engineers refer to an even larger return period, 10000-year, to safely design coastal defence structures. To note that, these estimates are based on historical datasets. Thus, the values are extrapolated from the present sea state conditions, not considering the possible effects of climate change, as it is done with return period estimates based on future projections (Chapter 7).

The reference parameters investigated in the present thesis are: the Significant Wave Height (SWH) or H_s, defined as the mean of the highest one-third of the waves in a wave record, and the wind speed at 10 m above the mean sea level (U_{10}). Today a large amount of wind speed and significant wave height data is available from global models, as well as satellite observations. This forms a solid base of time series on which we can apply long-term statistic approaches. The best performing statistical approach to estimate metocean extremes is the Peaks Over Threshold (POT), due to the possibility of selecting a larger number of extremes than the asymptotic models (*Coles et al.*, 2001; *Holthuijsen*, 2007). The topic of wind and wave extreme analysis is well treated in many practical reports such as the reference World Meteorological Organization (WMO) report (*Lopatoukhin et al.*, 2000). To perform sufficiently accurate extreme analyses, ocean storm peaks should be selected from time series sampled at an interval less than 6 h, otherwise, at a coarser time resolution ($>$ 6 h) some storms may be missed. This issue has been discussed in the case of satellite observations, where the detection of extremes is limited to the satellite passing time over a same region. Furthermore, as independence is a basic assumption for EVAs, we must select peaks from independent storms. This is commonly done by considering only peaks separated by more than 48 h, guaranteeing that each peak is selected from a different storm (*Lopatoukhin et al.*, 2000). If these criteria are met, we can apply the EVA.

As mentioned above, in some practical cases the Initial Distribution Method (IDM) is applied. In the following section, we briefly discuss the case of the IDM. This is a method whose use is no longer recommended due to the inappropriate manner it uses to extrapolate extreme estimates, but it is still often applied. We also apply an IDM in chapter 5 to compare the performance with common EVA.

Initial Distribution Method (IDM)

Ocean engineers are often confronted with a paucity of extreme sea state observations, and with poor resolution of wave models. Thus, to estimate the return period of extreme wave height, or wind speed, for offshore and coastal structures and operations, engineers relied for many years on the Initial Distribution Method

(IDM). This approach considers all available data at a specific location of interest. This means that not only extreme events but also the bulk of the data is used to fit a density function to the probability distribution and extrapolate return levels of extreme wind speed and wave height events. The function choice is facilitated by past experience as, for instance, it is known that wave height long-term analyses are well described by a *Gumbel*, a *Log-normal*, a *Fisher-Tippet I* or a *Weibull* distribution (*Tucker and Pitt*, 2001; *Alves and Young*, 2003; *Holthuijsen*, 2007). Once the function is identified the parameters are computed by the same statistical models discussed in section 4.2.2, such as the least-squares method or the maximum-likelihood method. The return level to extrapolate the value of the extreme estimate for a chosen return period is then found similarly to a POT approach as in Eq. 4.14.

$$\Pr\{x < x_N\} = 1 - \frac{d}{RP \times 24 \times 365}, \tag{4.14}$$

where, d is the decorrelation time scale expressed in hours. d is usually set to 3 h after the reference works of (*Tucker and Pitt*, 2001; *Cooper and Forristall*, 1997).

Noting the limited data that were available in many locations of the ocean, this approach is a pragmatic solution to the problem, and it is still used in some applications. The advantage is clear, this method uses all available data. However, the method is strongly affected by the bulk of the dataset hindering an accurate extrapolation of the extreme estimates. That is, the results are mostly affected by small values and less so by the few extremes. Nowadays, EVA approaches such as POT are commonly used in preference to the IDM as they provide better results (*Meucci et al.*, 2018; *Takbash et al.*, 2018)

4.5 Conclusions

The probability of exceedance of extreme events has to date been poorly explored due to the numerous sources of uncertainties around extreme estimates (*Morim et al.*, 2018, 2019). The objective of the second part of this thesis is to enhance estimates of confidence around ocean surface wind speed and wave height extreme estimates. We do so by developing a novel approach to estimate metocean extremes using an ensemble of i.i.d. model realizations of the atmosphere-ocean system. The technique is first applied to the ECMWF operational forecast model (chapter 5), then, extended to an inter-model ensemble of WWIII runs forced with seven different GCM surface wind fields.

Chapter 5

Wind and Wave Extremes from Atmosphere and Wave Model Ensemble forecasts

In the next part of the thesis we present an in-depth review of different methodologies and approaches to estimate H_s and U_{10} extremes. The novel ensemble EVA approach presented in this chapter shows great potential for robust long-term extreme analyses. This approach is later applied to future extreme projections (chapter 7). Table 5.1 shows the ensemble statistical approaches applied in the next part of the thesis with the main characteristics of the datasets analysed.

Table 5.1: Ensemble approaches applied throughout the next part of the thesis. The EVA ensemble approach was first tested on a single operational forecast model from ECMWF. The method was then applied to an inter-model ensemble constructed by forcing the same WWIII model with different GCM surface wind fields.

	Single-model Ensemble EVA ECMWF Operational forecasts (chapter 5)	Inter-model Ensemble EVA WWIII forced with 7 GCMs (chapter 7)	
Variables	H_s; U_{10}	H_s	
Time window	2010–2016	1979–2005	2081–2100
Time resolution	6h	6h	6h
Horizontal resolution	1°× 1°	1°× 1°	1°× 1°
Dataset years	6	27	20
Equivalent time period [years]	750	189	140

The present chapter is taken from the published work of *Meucci et al.* (2018) (©American Meteorological Society. Used with permission.). This work develops

an innovative approach to wind speed and significant wave height extreme value analysis. The approach is based on global atmosphere-wave model ensembles, the members of which are propagated in time from the best estimate of the initial state, with slight perturbations to the initial conditions, to estimate the uncertainties connected to model representations of reality. The low correlation of individual ensemble member forecasts at advanced lead times means they can be assumed to be independent and allows us to perform inference statistics. The advantage of ensemble probabilistic forecasts is that it is possible to synthesize an equivalent dataset of duration far longer than the simulation period. This allows the use of direct inference statistics to obtain extreme value estimates. A short time series of six years (from 2010 to 2016) of ensemble forecasts is selected to avoid major changes to the model physics and resolution and thus ensure stationarity. This time series is used to undertake extreme value analysis. The study estimates global wind speed and wave height return periods by selecting peaks from ensemble forecasts from +216 to +240 hours' lead time from the operational ensemble forecast data set of the European Centre for Medium-Range Weather Forecasts (ECMWF). The results are compared with extreme value analyses performed on a commonly used reanalysis data set, ERA-Interim, and buoy data. The comparison with traditional methods demonstrates the potential of this novel approach for statistical analysis of significant wave height and wind speed ocean extremes at the global scale.

5.1 Introduction

Uncertainties in the prediction of wind and wave extremes challenge the design and construction of marine systems. Design and construction of these systems rely on accurate statistical analyses of historical data sets, which provide extreme value return period estimates, i.e. values exceeded on average once every N years, where N is the defined return period. Inference statistics approaches are based on fitting a specified probability distribution to the available data series of observations and then, with preliminary hypotheses, estimating the extremes corresponding to the return period. In disciplines such as atmospheric and ocean sciences or hydrology, extreme value analysis (EVA) has become common practice (*Ferreira and Soares*, 1998; *Coles et al.*, 2001; *Caires*, 2011, 2016). Instead of considering the whole data set, EVA concentrates on the tail of the probability distribution function. For such analyses, it is desirable to have long continuous observational records. In practice, however, such records are seldom available. This hinders a reliable evaluation of the probability of occurrence of the extreme values (EV) (*Young*, 1999a; *Alves and Young*, 2003; *Holthuijsen*, 2007; *Zieger et al.*, 2009; *Aarnes et al.*, 2012; *Breivik et al.*, 2013, 2014). Satellite measurements are a potentially promising data source with which to estimate wind and wave extremes. They provide global coverage,

and their temporal duration, across multiple platforms, is now approximately 30 years (*Zieger et al.*, 2009; *Young et al.*, 2017). However, the spatial density of such systems is such that storm events may be missed or underestimated (*Young et al.*, 2011). In addition, there are questions about the accuracy of satellite measurements under extreme conditions (*Young et al.*, 2017; *Young and Donelan*, 2018). The different instrument biases may also affect the EV estimates. In contrast, buoys, ships and platform measurements are limited in space but may have long time series. Although buoy and platform measurements are often regarded as "ground truth" and are used to calibrate and validate both models (*Bidlot et al.*, 2002) and satellites (*Zieger et al.*, 2009; *Young et al.*, 2017), questions remain about their ability to measure extreme wind speeds and wave heights (*Large et al.*, 1995; *Zeng and Brown*, 1998; *Taylor and Yelland*, 2001; *Howden et al.*, 2008; *Bender et al.*, 2010; *Jensen et al.*, 2015). A solution to these constraints is to use numerical reanalyses. These provide long term continuous time series on which to perform EVA, but are affected by analysis states and model biases (*Caires et al.*, 2004; *Sterl*, 2004; *Stopa and Cheung*, 2014; *Aarnes et al.*, 2015).

In order to address the stochastic variability in global weather systems, meteorological agencies have adopted numerical weather prediction (NWP) models of atmosphere-ocean dynamics which include a probabilistic element. These models adopt an ensemble of operational forecasts that allows the estimation of uncertainties in forecasts, and thus of the atmosphere-ocean system predictability (*Palmer and Hagedorn*, 2006). Weather and climate are, by their nature, characterized by uncertainty. In such hydrodynamically complex systems significantly diverging states arise from slightly different initial conditions, as shown by the pioneering work by *Lorenz* (1963).

The ensemble approach deals with uncertainties, not only in determining future states of the atmosphere, but also by calculating an *analysis* (*Epstein*, 1969), i.e. the initial state of the model. Probabilistic methods and ensemble forecasts are now commonly used in weather prediction (*Lewis*, 2005). In the past few decades the atmosphere-ocean coupled global ensemble prediction systems (ENS) at the the European Centre for Medium-Range Weather Forecasts (ECMWF) has been continually developed and upgraded, in terms of model physics, assimilation methods, ingestion of observations and numerical resolution (*Molteni et al.*, 1996).

The current ECMWF Integrated Forecasting System (IFS) is based on a four-dimensional (time and the three spatial dimensions) variational assimilation scheme (4DVAR) to estimate the initial state (*Rabier et al.*, 1998). This is combined in an ensemble data assimilation (EDA) procedure where an ensemble of 4DVAR analyses is created which reflects the uncertainty of the analysis (*Isaksen et al.*, 2010). The EDA in turn is blended with singular vector estimates (*Buizza and Palmer*, 1995) to generate an ensemble of initial conditions that contain the fastest

growing modes to ensure a realistic ensemble spread with lead time. A variable resolution approach is employed to allocate more resources to the first 10 days where forecasts are more accurate, and less (typically half the resolution) for the remainder, to day 15 (*Buizza et al.*, 2007) and in some cases even further.

Ensemble prediction systems were not developed with the intention of extreme value estimation, but *Breivik et al.* (2013) demonstrated a practical application to ocean wind and wave extremes estimation. Their work was focused on the North-East Atlantic Ocean and the North Sea. Using a data set of operational ensemble forecasts from 1999 to 2009 they introduced four criteria with which to validate an EVA and then found the 100-year return period values of 10 m neutral wind speed ($U_{10\mathrm{N}}$) and significant wave height (H_s). We will consistently use the 10 m neutral wind speed as calculated for the wave model component of the IFS. This may at high wind speeds differ from the ordinary 10 m wind speed and was shown by *Aarnes et al.* (2015) to be more appropriate than the latter. Comparing the ensemble EV with the ERA-40 (*Uppala et al.*, 2005), and ERA-Interim (ERA-I, see *Dee et al.* 2011) re-analyses and the NORA10 hindcast (*Reistad et al.*, 2011; *Aarnes et al.*, 2012; *Furevik and Haakenstad*, 2012), they found that the predicted extreme values were comparable with NORA10, but higher than ERA-40 and ERA-Interim. ERA-40 and ERA-Interim are indeed known to underestimate the upper percentiles of H_s (*Stopa and Cheung*, 2014). The upper percentiles in enclosed seas, here the North Sea, are biased low due to the coarse resolution of ensemble forecasts.

Breivik et al. (2014) extended the approach to the global scale and applied EVA to find the 100-year return values over the world's oceans. Results were again compared with ERA-Interim and observational data showing comparable results but with much reduced confidence intervals compared to traditional EVA.

The objective of this work is to further develop this approach for wind and wave extreme value analysis from ensemble forecasts. The present methodology considers 5 different forecast lead times for a larger synthesized data set, and thus tightens the estimated confidence levels compared to those found by *Breivik et al.* (2014). This is achieved by selecting peaks from each of the ensemble forecasts to ensure independence. The synthesized data set has an equivalent length of 750 years (3 times longer than *Breivik et al.* (2014)). Thus, EVA can be achieved without the need to fit and extrapolate a probability distribution function to the data—for a 100 year return period, the estimated wave height/wind speed is "in-sample". The approach employs operational forecasts from a relatively short period (six years) to ensure that the data set is stationary and that the model resolution remains unchanged. Finally, the resulting extreme value estimates of wind speed and wave height are compared with traditional approaches (Peaks Over Threshold) for buoy data at specific sites and global values from satellite altimeter data and long du-

ration model reanalysis (ERA-Interim) data bases.

The chapter is structured as follows: in Section 5.2, a summary of traditional methods for EVA shows the potential for model ensemble EVA. Section 5.3 and Section 5.4 describe the methodology, the criteria followed and the data sets used. Section 5.5 presents the results of the EVA with a comparison with different data sets, traditional long-term statistical methods and EVA of buoy measurements. The discussion in Section 5.6 leads to the conclusions (Section 5.7), outlining the strengths and limitations in estimating ocean wind and wave extremes from model ensembles.

5.2 Estimation of wind and wave extremes

A large body of literature and research has been devoted to estimating U_{10} and H_s extremes. Records of these quantities started with ship visual observations, and for many years relied on this approach (*WMO*, 1998; *Gulev and Grigorieva*, 2004; *Holthuijsen*, 2007). In more recent times estimates from buoys and fixed platforms have become more common. Whilst recent studies demonstrated these estimates may be biased at extreme values (*Large et al.*, 1995; *Powell et al.*, 2003; *Bender et al.*, 2010; *Durrant et al.*, 2013), they are still the best estimates of the "ground truth" and a valuable resource to validate and compare numerical model results (*Bidlot et al.*, 2002). However, these data are limited in their time and space coverage and remote sensing techniques, such as satellite measurements, are becoming more common (*Zieger et al.*, 2009; *Young et al.*, 2011, 2017). Satellite records are better distributed in space but sparse in time (*Vinoth and Young*, 2011).

Based on three years of satellite altimeter data, a first evaluation of wind speed U_{10} and H_s for the development of an ocean atlas was performed by *Young and Holland* (1996). The database was subsequently updated to a 10-year satellite-model combined series (*Young*, 1999a). The records showed the potential for EVA on satellite measurements (*Young*, 1999a; *Alves and Young*, 2003; *Vinoth and Young*, 2011). More recently *Zieger et al.* (2009) collected and calibrated data available from various satellite missions to produce a reliable data set of 23 years, subsequently extended to 30 years by *Young et al.* (2017). These observations from altimeters, scatterometers and radiometers are a valuable resource to understand ocean extremes. Such a continuous data set helps evaluate the global wind and wave climatology and their trends (*Young et al.*, 2011). These satellite observations have now been used to estimate the rate of increase of extreme values of global wind speed and wave height (*Young et al.*, 2012; *Zieger et al.*, 2014).

Similarly, EVA of wind speed and significant wave height can be based on model data obtained from hindcasts or reanalyses (*Aarnes et al.*, 2012, 2015; *Caires and*

Sterl, 2005). The goal of a reanalysis is to produce a data set which is statistically homogeneous in space over a long time period which is not affected by model changes and thus statistically stationary. These reanalyses rely on data assimilation from in situ measurements and satellite records. An example of the impact is the improvement of results for the Southern Ocean due to satellite measurement assimilation (*Caires et al.*, 2004). In an inter-comparison of the different wind and wave reanalysis data set, *Caires et al.* (2004) showed that the accuracy of such models has improved, particularly those produced by ECMWF.

ECMWF has generated a series of increasingly sophisticated reanalyses, starting with ECMWF European ReAnalysis 15 (ERA-15) (*Gibson et al.*, 1997) which covered the period 1979-1993. This was followed by ECMWF European ReAnalysis 40 (ERA-40) (*Uppala et al.*, 2005) which covered the period 1957-2002. Several global extreme value analyses of significant wave height have been based on the ERA-40 data set. The Royal Meteorological Institute of the Netherlands (KNMI) Atlas was based on ERA-40 model results, and produced results which generally underestimate wind speed and wave height extremes (*Sterl and Caires*, 2005). Due to changes in data assimilation over time (*Sterl and Caires*, 2005), the consistency and hence the reliability of the results from the model was limited. After ERA-40, ERA-I (*Dee et al.*, 2011) became the new standard to apply EVA. ERA-I covers the period from January 1979 until the end of August 2019, and it has been used also to evaluate ocean extremes and potential trends of wind and waves (*Aarnes et al.*, 2012, 2015). However, ERA-I still underestimates wind speed and wave height, and according to *Stopa and Cheung* (2014) particular attention must be paid to the analysis of the upper percentiles of the data, which may not be well represented by the model. The latest ECMWF European ReAnalysis 5 (ERA5), has become the new standard, and a significant amount of research has already been performed on this dataset. However, there are still no comprehensive evaluation of the performance of this new reanalysis for EVA. In general, all re-analysis data sets have limitations due to inhomogeneity in data assimilation (*Sterl*, 2004). For this reason, ECMWF produces an hindcast version of the ERA5 called ERA5-h, where no data is assimilated into the model. This dataset could prove to be of significant interest in evaluating the impact of data assimilation.

EVAs of data sets are also affected by the methodology used to estimate return periods. The Initial Distribution Method (IDM) is a common approach that fits a probability density function (PDF) to the whole body of data available, and extrapolates the chosen PDF to the desired return period. The IDM is commonly used when the size of the available data set is relatively small. As the IDM uses all the data, it can return stable estimates in such situations. However, there are a number of limitations to this approach. First, the extrapolation of the tail of the PDF to the desired probability level (return period) is done with a fit dominated

by the bulk of the data in the body of the PDF. Hence, the accuracy with which the tail is defined is questionable. Second, there is no theoretical basis for the selection of the appropriate PDF form to fit to the data.

Extreme value analysis based on independent maxima (or minima), as opposed to IDM, which considers the entire data set falls in two categories, (i) asymptotic models and (ii) threshold models (*Coles et al.*, 2001). The extrema must be independent and identically distributed (i.i.d). Independent, so that observations at any particular time are not correlated with other data points close in space or time, and identically distributed, to apply common inference statistics theorems. For asymptotic models the extreme value theorem indicates that a Generalized Extreme Value (GEV) distribution can be fitted to the data set, and the distribution of asymptotic block maxima (or minima) be used to find the desired return levels. The time series is typically constructed by extracting the annual maxima (AM). The lack of long measurement time series, and the fact that only one value is retained per year, often prevents robust estimates of the desired return periods with AM.

The small data sets resulting from AM is partly overcome by threshold models. A Peaks-over-threshold (POT) approach (*Pickands*, 1975) fits peaks above a threshold u, to a Generalized Pareto (GP) distribution (*Coles et al.*, 2001). This method allows the selection of a larger number of extreme values compared to the AM method, improving statistical inferences. There is, however, no theoretically robust approach to selecting the threshold u. Once the peaks are selected, attention must be paid to ensure each of these values is independent (i.e. not selected from the same storm). A common practice is to consider 48 h storm independence, i.e. the peaks must be at least 48 hours apart (*Lopatoukhin et al.*, 2000; *Caires and Sterl*, 2005). The POT method is often strongly dependent on the threshold selection, and this in turn depends on the geographical location, time period and data set itself (*Caires and Sterl*, 2005; *Holthuijsen*, 2007). The ultimate choice tends to be a trade-off between bias and variance (*Caires*, 2016). That is, a low threshold may lead to bias but will result in a larger data set and reduced variance in the fit to the PDF.

In each of the approaches described above (IDM, AM, POT), the measured time series is shorter than the desired return period. As a result, the PDF needs to be extrapolated to the desired probability level (return period). If the time series were longer than the desired return period, extrapolation would not be necessary and the EV could be directly determined, in-sample, from the tail of the empirical PDF of the recorded data or from a theoretical PDF fitted to the data.

5.3 Wind and wave extremes from ensemble model predictions

Breivik et al. (2013, 2014) showed that at advanced lead times the individual ensemble members of the Ensemble Prediction System (ENS) correlate only weakly, and hence represented independent realizations of the wind or wave field. As such, it is possible to synthesize a data set the equivalent of hundreds of years. The desired return period can then be determined as above without extrapolation of the PDF. That is, using a Direct Return Estimate (DRE) from an in-sample measure of the empirical distribution. *Breivik et al.* (2013, 2014) calculated 100-year return estimates of 10 m wind speed and significant wave height ($U_{10\mathrm{N}}^{100}$, H_{s}^{100}), from a 9-year data set of ensemble forecasts at 10 days' (+240 h) lead time. The essence of the approach is that the ensemble predictions at long lead times can be pooled to create the equivalent of a very long time series. *Breivik et al.* (2013) proposed four criteria (see below) that must be met. However, the method still comes with some caveats. First, the possible influence of model updates on the biases could trigger a spurious trend. Second, as mentioned above, the work of *Breivik et al.* (2013, 2014) was performed only on day 10 (+240 h) lead time forecasts. The potential to include shorter lead times in the evaluation of the return periods, to further increase the data set, was considered but not pursued beyond a cursory EVA based on pooling +228 h and +240 h forecasts. Finally it was assumed that each forecast was representative of a six-hour period. There is little theoretical justification for this assumption, leaving doubts as to the possible bias introduced in the final estimates.

To answer these caveats, the methodology adopted here is based on the ECMWF ENS data set (that consists of forecasts issued twice daily at 00 and 12 UTC) over a six-year period, from 3 March 2010 to 29 February 2016. In this period the spatial resolution was unchanged at approximately 32 km for the atmosphere and about 54 km for the wave field. Every forecast is composed of 50 ensemble members derived from slightly perturbed initial conditions, plus one control member modeled with the best estimate of the initial conditions, and one HRES (Higher RESolution) member that is the most accurate forecast run at higher spatial and temporal resolution. We chose to focus on the 50 ENS dataset that we have available. The present analysis could be extended to add the control member that has the same resolution as the ensemble, thus adding an additional member. The HRES member is not considered as it is from a model run at twice the resolution of the ensemble.

In this study we focus on the ensemble of 50 perturbed members interpolated on a $1.0^\circ \times 1.0^\circ$ global grid, excluding the ice-infested areas (south of 70° S and north of 80° N). ENS forecasts are taken at advanced lead times to assure i.i.d.

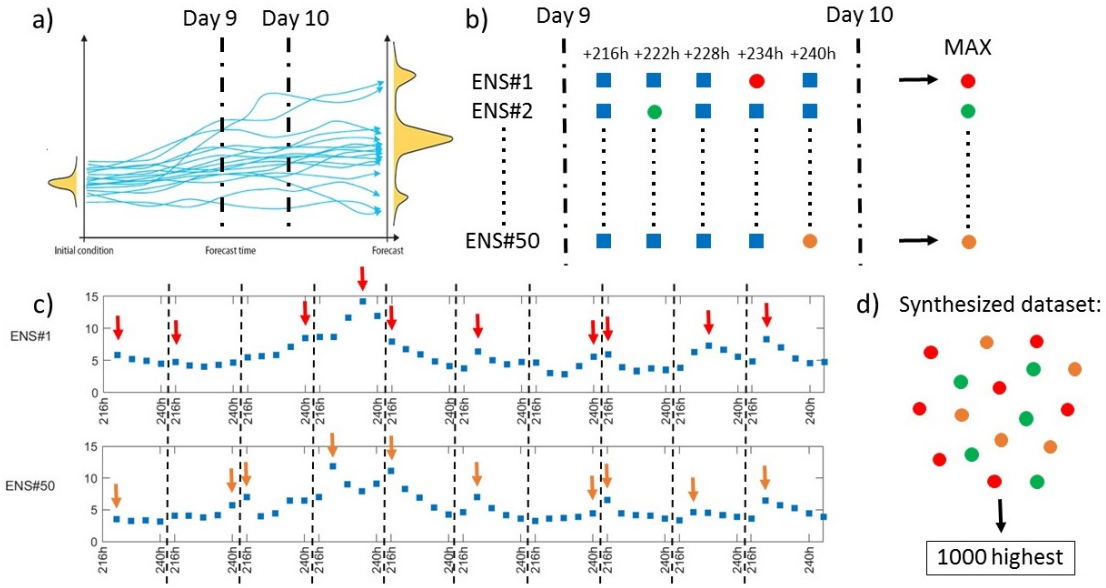


Figure 5.1: Procedure for advanced lead times ensemble forecasts pooling. Construction of the synthesized dataset to find Direct Return Estimates out of the 1000 highest peaks. (a) Each ensemble forecast diverges with time. (b) For each ensemble member, the maximum value is selected from lead times +216 h to +240 h. (c) The selected maximum values are representative of a 30 h period. Each period is independent, as shown by the vertical dotted lines. (d) The selected maximum values represent independent realizations of extreme wind speed/wave height and are pooled for the EVA.

data, as detailed in Section 5.4. In contrast to *Breivik et al.* (2013, 2014), who selected values at a fixed lead time of 10 days, we select the maximum value of the five lead times from day 9 to day 10 (+216 h, +222 h, +228 h, +234 h and +240 h). The objective is to estimate wind and wave extremes for a 100-year return period ($U_{10\mathrm{N}}^{100}$, H_{s}^{100}).

The operation of the ECMWF ensemble prediction system and the EVA methodology adopted are shown diagrammatically in Fig. 5.1. A new set of ensemble forecasts are issued twice daily (00 and 12 UTC). As noted above, 50 perturbed ensemble members are generated, each with slightly different initial conditions. As shown in Fig. 5.1a, these forecasts will diverge in time. If these ensemble forecasts remain representative of the true wind/wave climate (both means and extremes) and they diverge to the extent that they are not significantly correlated, they will each represent a possible realization of the wind/wave field (in other words, a random draw from climatology). If this is the case, they could all be used in an EVA. *Breivik et al.* (2014) selected values at a lead time of 10 days (+240 h). In the

analysis here, we consider lead times from Day 9 to Day 10 (+216h to +240h). Over this 30-hour period we select the maximum wind speed/wave height (Fig. 5.1b) from each member.

Thus, the 30-h maximum is kept for each of the 50 ensemble members. It is, however, important to note that the values obtained for two consecutive forecasts, which commenced 12 hours apart for say ENS#1, are not related as they are started from initial states 12 hours apart and then diverge as the forecast evolves (out to Days 9 and 10). Indeed, the perturbations added to each member at analysis time are random (see discontinuities in Fig. 5.1c). This is quite different to traditional EVA (e.g. POT, AM, IDM) where values are selected from a continuous time series. As will be shown below, it is thus necessary to assign some representative time interval to each of these independent values to arrive at an equivalent length of our data set.

Such an analysis is valid, and thus provides relevant results, only if the sample of data has specific characteristics (*Coles et al.*, 2001). These characteristics can be represented by the following methodological criteria.

5.4 Methodology criteria

Breivik et al. (2013, 2014) proposed four criteria required for statistical validity of the ensemble data when applied to EVA:

1. No significant correlation between ensemble members at advanced lead times, i.e. i.i.d. ensemble members
2. Model climatology comparable to the observed climatology distribution
3. No spurious trend in the data set due to model updates
4. Forecasts are representative for a specific time interval

The testing procedure for each of these criteria is described in the following sections.

Criterion 1: Independent and Identically Distributed (i.i.d) ensemble members

To perform an EVA the data sample (in this case the ensemble forecasts) must be i.i.d. (*Coles et al.*, 2001). This translates into the requirement of independence of the upper percentiles of the sampled ensemble members (the tails). The ENS forecast skill drops around day 9/10 (*Buizza*, 1997), and as a result one could expect also a lower level of correlation for the upper percentiles of the individual

members, the part of the data set of interest for the EVA. The objective is thus to first test for independence and identical distribution of the ensemble members at advanced lead times. We wish to assess the validity of independence at 9 days' lead time (+216 h), and then to select the peak from several lead times from each forecast (i.e. +216 h, 222 h, 228 h, 234 h, 240 h). To test the independence of the ensemble members a centered anomaly correlation (ACC, see *Wilks* 2011; *von Storch and Zwiers* 1999) is computed for two randomly chosen ensemble members and evaluated at +216 h and +240 h lead time. The ACC is calculated as follows.

$$r_{\mathrm{ACC}} = \frac{\sum_{t=1}^{T} \left(x_i^{'}(t) - \overline{x}_i^{'}\right)\left(x_j^{'}(t) - \overline{x}_j^{'}\right)}{\sqrt{\sum_{t=1}^{T} \left(x_i^{'}(t) - \overline{x}_i^{'}\right)^2 \sum_{t=1}^{T} \left(x_j^{'}(t) - \overline{x}_j^{'}\right)^2}} \tag{5.1}$$

where

- r_{ACC} is the anomaly correlation coefficient between x_i and x_j (x represents H_{s} and $U_{10\mathrm{N}}$ in the present application)
- x_i and x_j are two randomly chosen ensemble members (in this case ENS1 and ENS50)
- t is time level, from 1 to T ($T = 2192\,\mathrm{days} \times 2 = 4384$ values); total of the 2010-2016 six years data set at 2 forecasts per day
- $x_i^{'}$ and $x_j^{'}$ are the anomalies computed by subtracting the monthly mean for each year, and each month of the 2010-2016 ENS data set

Testing was performed initially at the same three reference grid points for the North Atlantic Ocean and the North Sea as used by *Breivik et al.* (2013) (Fig. 5.2).

- P40: *Ekofisk Oil Field* WMO code LF5U (56.50° N, 3.20° E)
- P35: *Heidrun Oil Field* WMO code LF3N (65.30° N, 7.30° E)
- B16: *K5 buoy* WMO code 64045 (59.10° N, 11.40° W)

The ACC (5.1) for H_{s}, computed for lead times of +216 h and +240 h over the six-year ENS data set, are shown in Table 5.2, and in Table 5.3 in terms of linear correlation (i.e. Pearson product-moment). This is computed as in Eq. (5.1), without previously subtracting monthly means (i.e. not considering anomalies).

As expected, the correlation is higher for shorter lead times and is higher for the open ocean sites, compared to enclosed seas (P40). The reason why lower anomaly correlations are found in the North Sea compared to the North Atlantic is the

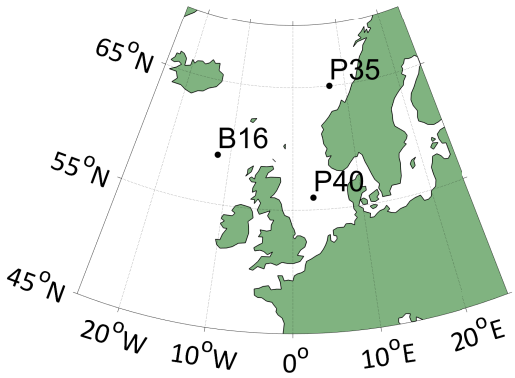


Figure 5.2: Locations used for testing the methodology criteria: P40 - *Ekofisk Oil Field* WMO code LF5U (56.50° N, 3.20° E); P35—*Heidrun Oil Field*—WMO code LF3N (65.30° N, 7.30° E); B16—*K5 buoy*, WMO code 64045 (59.10° N, 11.40° W).

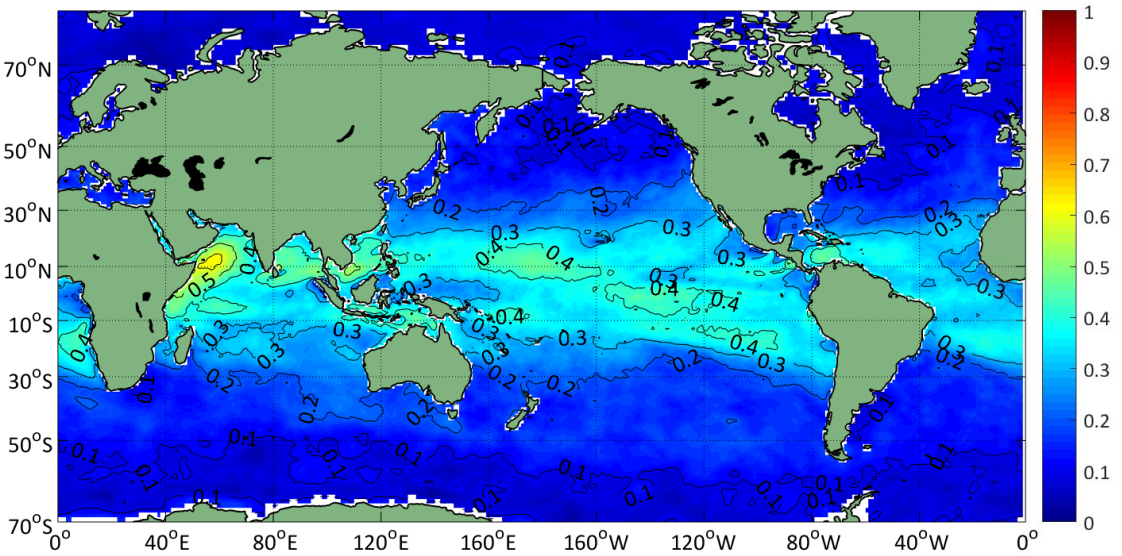


Figure 5.3: Centered anomaly correlation: U_{10N} for ENS member 1 and ENS member 50 at +216 h lead time from the 2010-2016 ENS data set. The highest correlations are found at the Horn of Africa, due to the regularity of the Somali jet and Monsoon events in this region. Tropical areas also show higher values than higher latitudes.

Table 5.2: H_s Centered anomaly correlation, r_{ACC} as in Eq. (5.1), for the three test locations (Fig. 5.2) for day 9 and day 10 lead times. The comparison shows higher correlation at shorter lead times and tests the two bounds (+216 h and +240 h) of the 5 lead times used in the present work

Lead time	**+216 h**	**+240 h**
P40 (North Sea)	0.1588	0.1499
P35 (Open Ocean)	0.2447	0.1841
B16 (Open Ocean)	0.2507	0.1838
Average	0.2181	0.1726

Table 5.3: H_s Actual Correlation (Pearson) coefficient for comparison with the Centered Anomaly Correlation for the same test locations and lead times as in Table 5.2

Lead time	**+216 h**	**+240 h**
P40 (North Sea)	0.3750	0.3651
P35 (Open Ocean)	0.5133	0.4757
B16 (Open Ocean)	0.5577	0.5167
Average for Open Ocean	0.5352	0.4962

reduced swell propagation to this more sheltered region. As such, locally generated wind sea will largely define the wave climate. Wind sea has lower predictability at long lead times than swell (see Fig. 5.3).

The results of Table 5.2 are comparable to the values found by *Breivik et al.* (2013). The small differences found may be due to the different approach used in the subtraction of seasonality in the data, and to the different time spans of the data sets. The present data set spans the period 2010 to 2016, while *Breivik et al.* (2013) analysed data from 1999 to 2009.

The ACC (Eq. 5.1) was computed for both H_s and U_{10N}. Fig. 5.4 shows the global results for H_s at lead time +216 h. Tropical regions show a higher correlation due to the dominance of swell. With the exception of equatorial regions where $0.6 < r_{ACC} < 0.7$, the remainder of the globe shows $r_{ACC} < 0.5$. For EVA, the requirement is for small values of r_{ACC}. This criterion appears to be generally met at lead times greater than +216 h for H_s, with perhaps some questions about the equatorial region. The same test was performed for U_{10N} in Fig. 5.3. The wind speed generally exhibits low correlation over the whole globe, with the exception of the North Indian Ocean near the Horn of Africa. This is most probably due to the regularity of the Somali jet and of the Monsoon events in this area (*Ranjha*

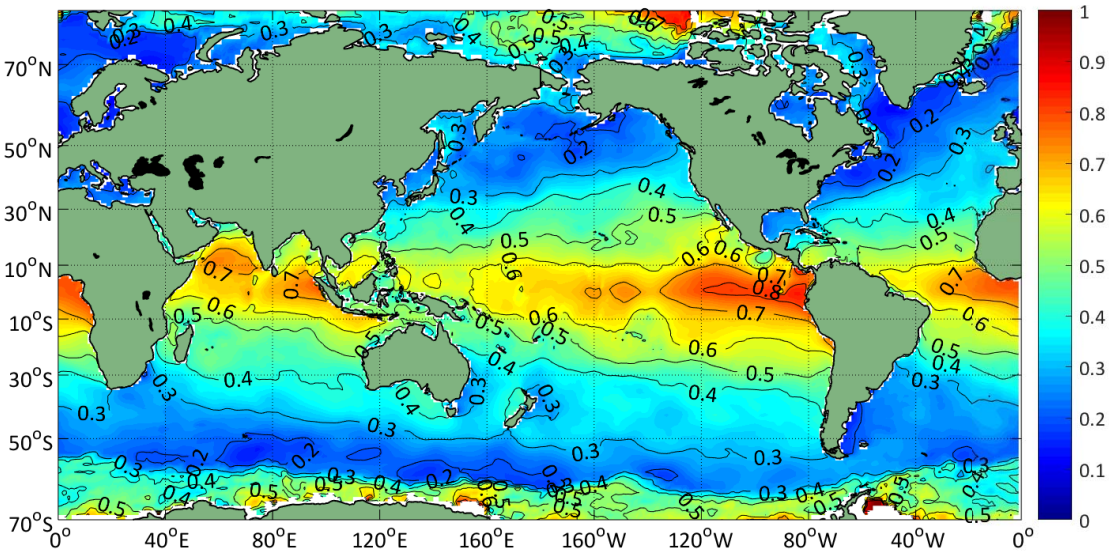


Figure 5.4: Centered Anomaly Correlation: H_s for ENS member 1 and ENS member 50 at +216 h lead time from the 2010-2016 ENS data set. The values show higher correlation in tropical areas which are more swell dominated.

et al., 2015). The reason why we find lower correlation in wind speed than in wave height is because, once generated, waves propagate over large distances (swell), thus increasing the forecast skill at advanced lead times.

In order to test whether the ensemble members were identically distributed, a quantile-quantile (Q-Q) comparison was performed for the three test sites of Fig. 5.2. Q-Q and scatter plots are shown in Fig. 5.5. Again, results are shown for the two ensemble members ENS1 and ENS50. The large amount of scatter confirms that r_{ACC} is low for these locations. Despite this, the Q-Q plots indicate that the PDFs are very similar for the ensemble members shown, supporting the assumption that they are identically distributed.

For comparison with the findings by *Breivik et al.* (2013), Fig. 5.5 presents the three North Atlantic sites shown in Fig. 5.2. The same analysis was carried out at a total of 33 locations around the globe (Fig. 5.6) yielding results consistent with those in Fig. 5.5.

Criterion 2: Modelled and observed climatology

Wind and wave model data (ENS operational forecasts) must represent the observed climatology to be used for an EV estimation. Hence, a second check of the data is performed involving comparisons of:

- Monthly ENS H_s and U_{10N} means and 90th percentiles compared to calibrated

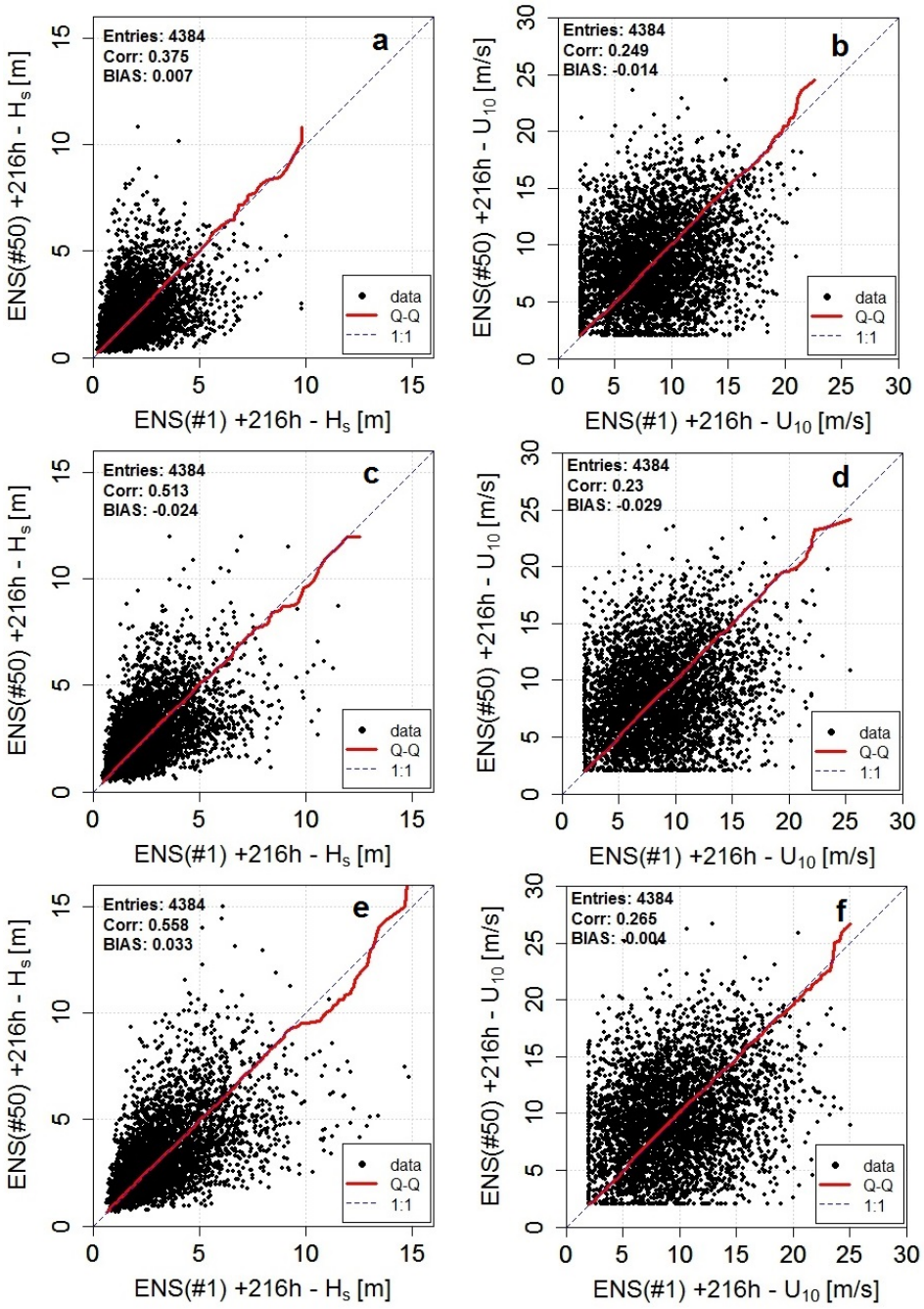


Figure 5.5: Quantile-Quantile plots of ensemble member 1 against ensemble member 50 at +216 h lead time for the test locations: (a,b) P40 North Sea, P35 (c,d) and B16 (e,f) in the North Atlantic. The left column (a,c,e) show H_s [m] while the right column (b,d,f) shows U_{10N}. The Q-Q result is shown by the red solid line, which closely approximates the dashed diagonal, indicating a similar PDF. Also shown is a scatter plot of individual predictions (dots).

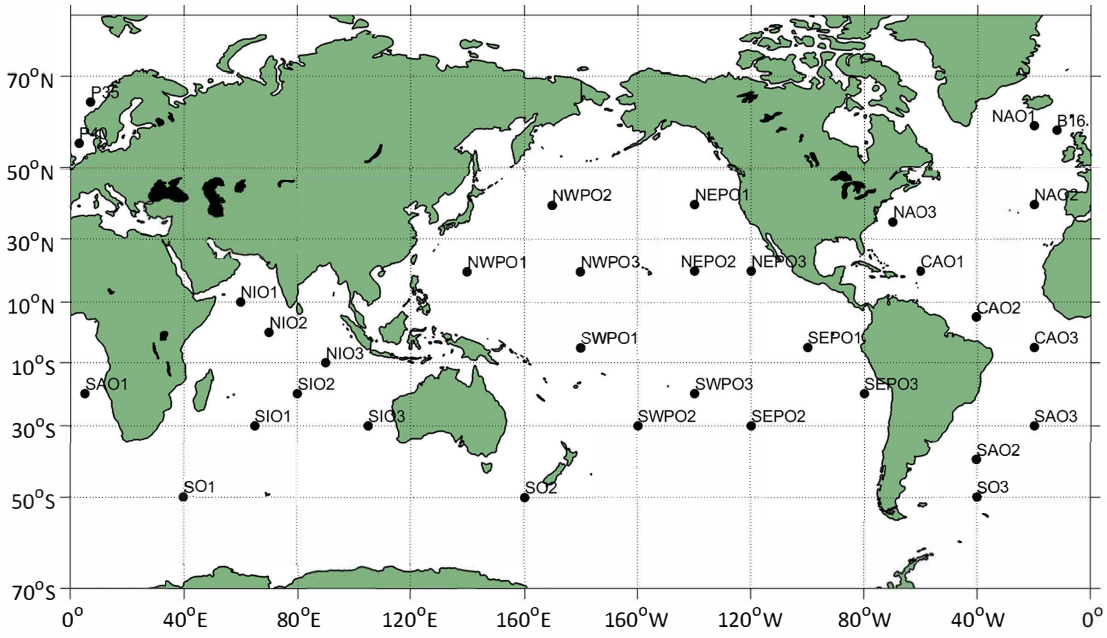


Figure 5.6: Test locations over the global dataset domain.

and validated satellite data

- Monthly ENS H_{s} and $U_{10\mathrm{N}}$ means compared to ERA-Interim

To test the model climatology we first investigate the satellite data set recently calibrated, validated and extended until 1 April 2015 by *Young et al.* (2017). The objective of the comparison is to evaluate the monthly mean and 90$^{\mathrm{th}}$ percentile values of H_{s} and $U_{10\mathrm{N}}$ over the six-year period of the ENS forecasts analysed in this study, i.e. from 2010 to 2016, and compare these to the 28 year average values from the altimeter database (*Young et al.*, 2017). This comparison tests both that the model can reasonably reproduce the climatological mean and extremes and that the 6-year period is representative of the longer-term climate. We could have chosen a higher percentile than 90, but it should be remembered that the sampling pattern of satellite altimeters means that peaks will be missed (*Young et al.*, 2011, 2017). As a result, it is expected that the altimeter will underestimate higher percentiles, the underestimation increasing as the value of the percentile increases.

The comparison is performed for the ensemble ($1^{\circ} \times 1^{\circ}$ resolution) at the three locations mentioned in Section 5.4 (Fig. 5.2). We selected $2^{\circ} \times 2^{\circ}$ satellite regions that surround the model locations. The maximum normalized difference values of $\Delta\overline{H}_{\mathrm{s}}/\overline{H}_{\mathrm{s}}^{SAT}$ are shown in Table 5.4 for the three test locations, with the corresponding occurrence month where $\Delta\overline{H}_{\mathrm{s}} = \overline{H}_{\mathrm{s}}^{\mathrm{ENS}} - \overline{H}_{\mathrm{s}}^{\mathrm{SAT}}$. Here, $\overline{H}_{\mathrm{s}}^{ENS}$ refers to monthly means computed by first averaging the 50 ensemble members (every

12 h), and then determining the ensemble means over the month. Similarly, $\overline{H}_s^{\mathrm{SAT}}$ is the monthly mean of the satellite measurements for the $2^\circ \times 2^\circ$ region.

Fig. 5.7 shows the normalized H_s and $U_{10\mathrm{N}}$ differences and the mean monthly values for both satellite and ENS +240 h, as a function of month for the three locations in Fig. 5.2. The results show that the mean monthly values of H_s for both the satellite (altimeter) and the model at lead time +240 h are in good agreement. Most mean monthly differences $\Delta\overline{H}_\mathrm{s}/\overline{H}_\mathrm{s}^{\mathrm{SAT}}$ are less than 10%. The largest monthly difference is 16.5%, found at the North Sea location P40 (Fig. 5.7a). This most likely represents limitations of model resolution in this more protected near-shore site as well as potential errors caused by selecting satellite data in a location where there are likely to be greater spatial gradients of H_s than in the open ocean. The comparison for the wind speed $U_{10\mathrm{N}}$ shows normalized differences below 10%. In this case the highest difference is found in the open ocean location P35, 9.2% (Fig. 5.7d).

Table 5.4: Comparison of normalized monthly average differences between the ENS forecasts and satellite (SAT) measurements (*Young et al.*, 2017). The maximum normalized differences $\Delta\overline{H}_\mathrm{s}/\overline{H}_\mathrm{s}^{\mathrm{SAT}}$ are shown along with the month in which they occur.

Lead time		**+216 h**	**+240 h**
P40 (North Sea)	Dec	0.143	0.165
P35 (Open Ocean)	Mar	0.125	0.129
B16 (Open Ocean)	May	0.132	0.152

A comparison of mean monthly values was also performed using ERA-Interim for the same period as the ENS forecasts (3 March 2010 to 29 February 2016). The ERA-Interim reanalysis provides data at the same spatial resolution (on a 1.0° grid) as the ENS, and the temporal resolution is every 6 h. Table 5.5 shows the maximum normalized difference values $\Delta\overline{H}_\mathrm{s}/\overline{H}_\mathrm{s}^{\mathrm{ERAI}}$. This table shows maximum differences comparable to the satellite analysis.

The ERA-Interim/ENS results are compared in the Q-Q plots shown in Fig. 5.8. They show a tendency for the ENS forecast to give slightly higher H_s extremes compared to the ERA-Interim for all three reference grid points in the North Sea and North Atlantic basins (P40, P35, B16). The ENS wind speed exhibits upper percentiles quite comparable to ERA-Interim. The ENS data used here is from IFS model cycles *Cy36r1* (released 26 Jan 2010) through *Cy41r1* (replaced 8 March 2016). ERA-Interim is built using IFS model cycle *Cy31r2* released 12 Dec 2006.

A comparison with monthly 90$^{\mathrm{th}}$ percentiles was undertaken to evaluate the climatology of the extremes. Fig. 5.9 shows the normalized H_s and $U_{10\mathrm{N}}$ differences and the monthly 90$^{\mathrm{th}}$ percentile values for both satellite and ENS +240 h, as a

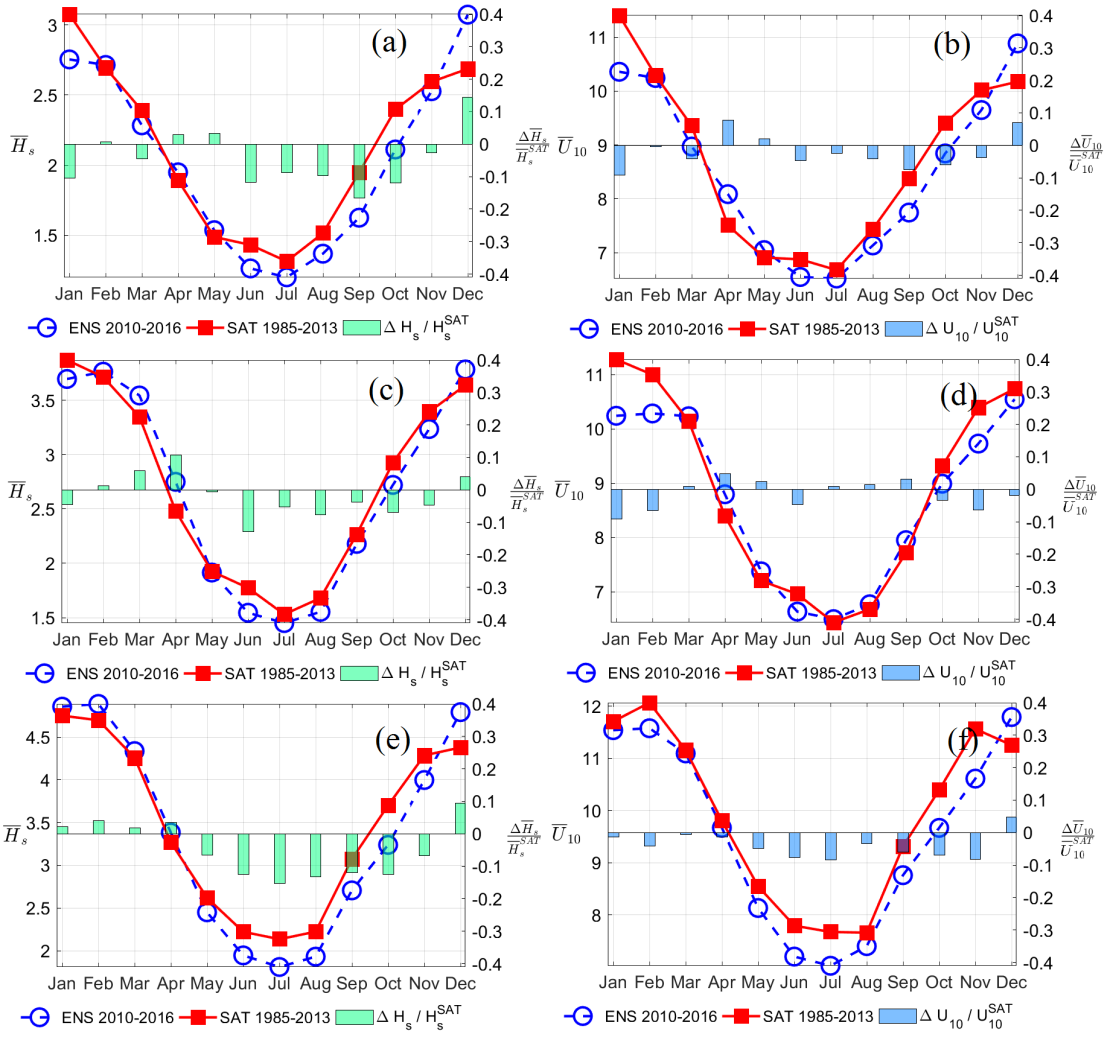


Figure 5.7: Comparison of $\overline{H}_s$ (left) and $\overline{U}_{10}$ (right) monthly averages between the ENS at +240 h lead time and SATellites for the test locations: (a,b) P40 North Sea, (c,d) P35 North Atlantic, (e,f) B16 North Atlantic.

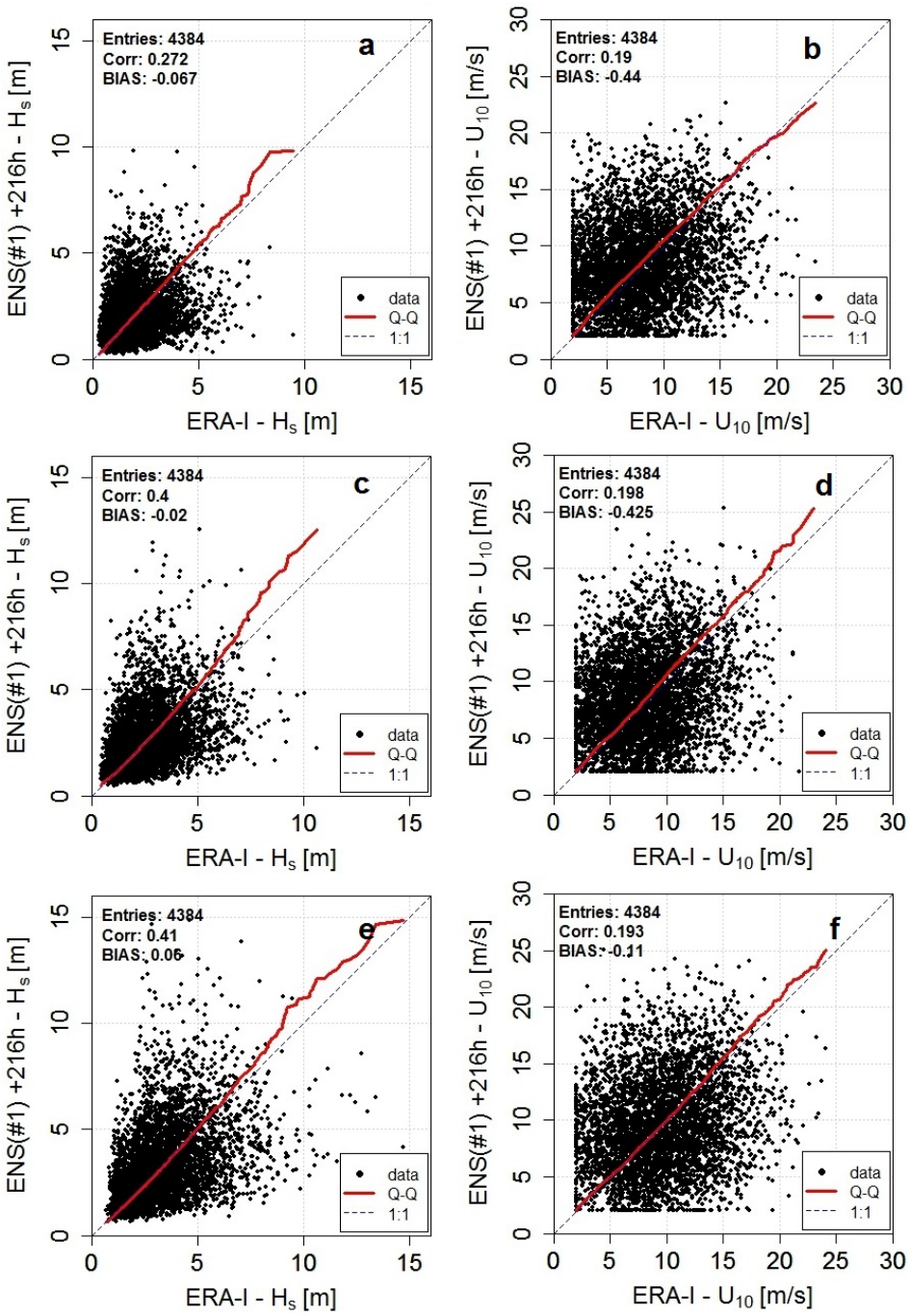


Figure 5.8: Quantile-Quantile plots of ensemble member 1 at +216 h lead time and ERA-Interim for the testing locations: (a,b) P40 North Sea, (c,d) P35 North Atlantic, (e,f) B16 North Atlantic. On the left side the comparison for H_s (a,c,e) while on the right side for U_{10N} (b,d,f).

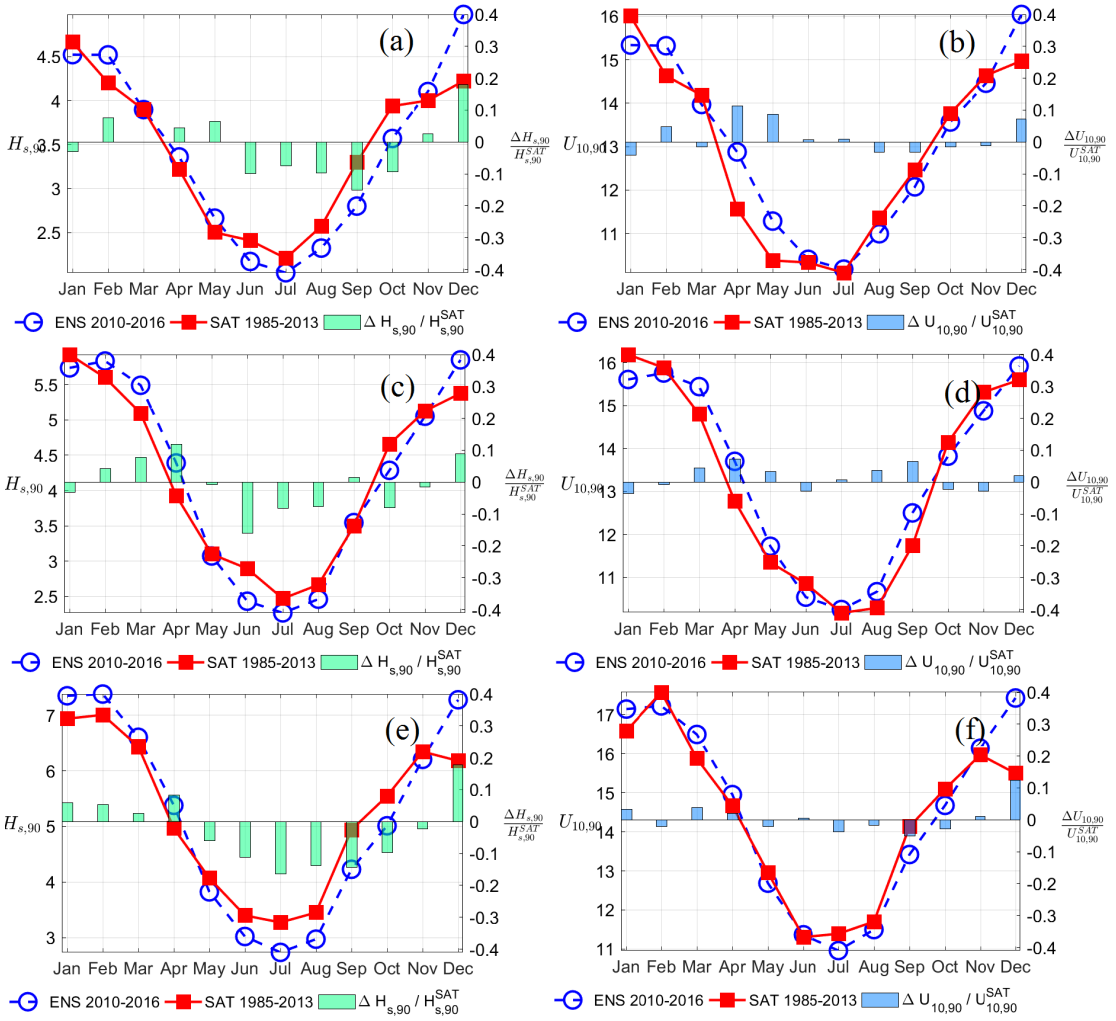


Figure 5.9: Comparison of $H_{s,90}$ (left) and $U_{10,90}$ (right) monthly 90$^{\text{th}}$ percentiles between the ENS at +240 h lead time and SATellites for the test locations: (a,b) P40 North Sea, (c,d) P35 North Atlantic, (e,f) B16 North Atlantic.

Table 5.5: Comparison of normalized monthly average differences between the ENS forecast data set and ERA-Interim. The maximum normalized differences $\Delta\overline{H}_{\mathrm{s}}/\overline{H}_{\mathrm{s}}^{\mathrm{ERAI}}$ are shown along with the month in which they occur.

Lead time		**+216 h**	**+240 h**
P40 (North Sea)	Nov	0.1531	0.1408
P35 (Open Ocean)	Aug	0.1522	0.1543
B16 (Open Ocean)	Jul	0.1498	0.1458

function of month for the three output locations in Fig. 5.2. The results show that the monthly 90$^{\mathrm{th}}$ percentile values of H_{s} for both the satellite (altimeter) and the model at lead time +240 h are also in good agreement. Most monthly differences $\Delta H_{\mathrm{s},90}/H_{\mathrm{s},90}^{\mathrm{SAT}}$ are also less than 10% also for the case of these high percentile values. The largest monthly difference is 17.9%, found again at the North Sea location P40 (Fig. 5.9a) for the reasons already explained for the monthly means. The comparison for the 90$^{\mathrm{th}}$ percentile wind speed $U_{10\mathrm{N}}$ shows normalized differences generally below 10%. In this case the highest difference is found in the open ocean location B16, 12.4% (Fig. 5.9f).

The comparison for an additional six locations, chosen from the total tested (Fig. 5.6), is shown for three Northern Hemisphere points in Fig. 5.10 and for three Southern Hemisphere points in Fig. 5.11. Differences between the model and satellite climatologies are shown at location NWPO1 (Fig. 5.10a and Fig. 5.10b). This is an area visited by tropical cyclones. Due to the coarse resolution (32×32 km for the atmospheric component of IFS), the model is unable to reproduce the intensity of these extreme events and we see significant differences in the monthly 90$^{\mathrm{th}}$ percentiles of significant wave height. It is also interesting to note the differences between the model and the satellite monthly 90$^{\mathrm{th}}$ percentiles at location NIO1 (Fig.5.10e and Fig. 5.10f) during the monsoon months. This is globally the location where the largest differences were found.

The results for the Southern Hemisphere (Fig. 5.11), show that the satellites and model monthly 90$^{\mathrm{th}}$ percentiles are in good agreement, with normalized differences generally below 10%.

Both the altimeter and ERA-Interim comparisons were performed for all the locations in Fig. 5.6 around the globe. Although there are some regional variations (e.g. both models under-estimate wind speeds and wave heights in the Somali Jet), generally the IFS accurately reproduces the climatology (mean and 90$^{\mathrm{th}}$ percentile) well. There remains a question about how well any model reproduces the extreme above the 90$^{\mathrm{th}}$ percentile. However, a rigorous global comparison is not possible as altimeter data is also questionable at such extremes (e.g. under-sampling). The ultimate test will be the magnitudes of the resulting extreme value estimates (see

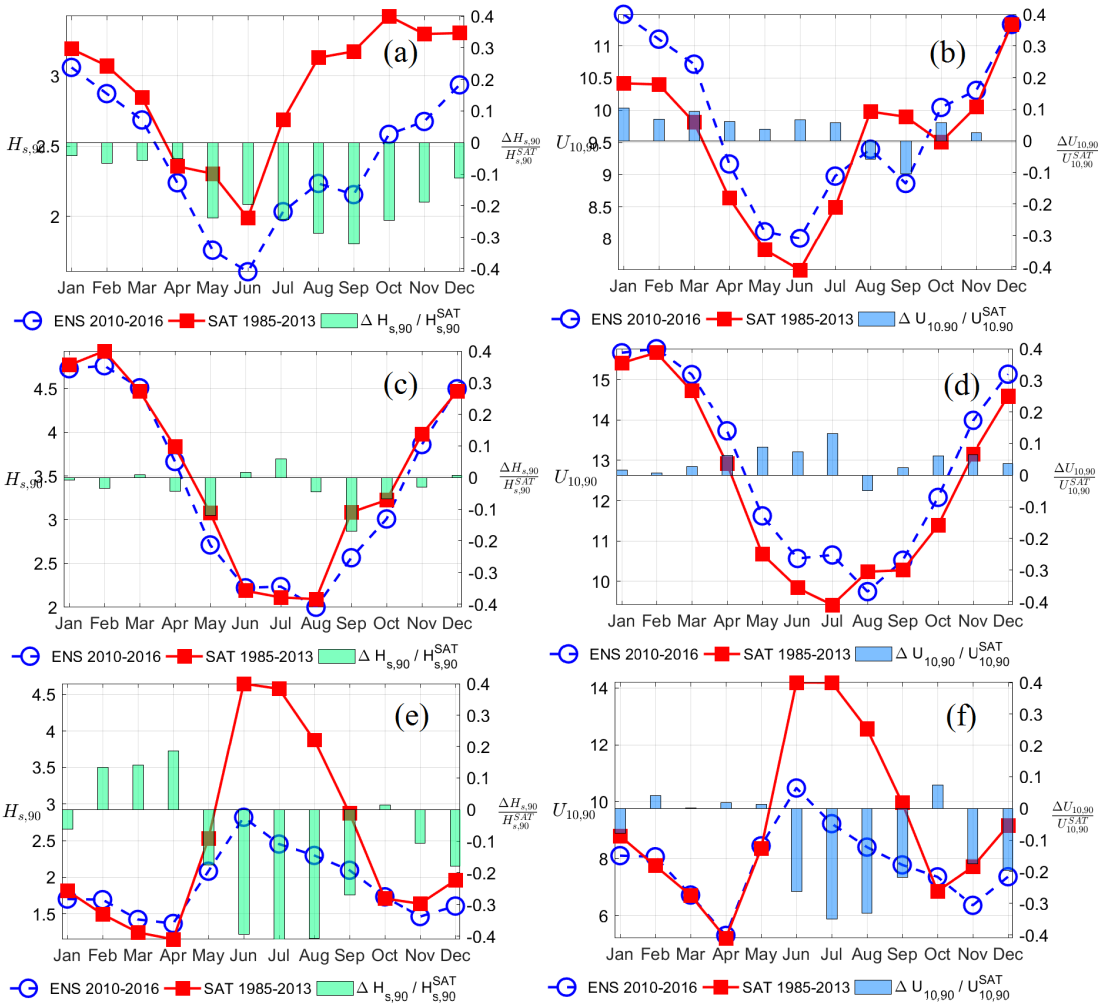


Figure 5.10: Comparison of $H_{s,90}$ (left) and $U_{10,90}$ (right) monthly 90$^{\text{th}}$ percentiles between the ENS at +240 h lead time and SATellites for the test locations in the Northern Hemisphere (NH): (a,b) NWPO1 North West Pacific Ocean, (c,d) NAO3 North Atlantic Ocean, (e,f) NIO1 North Indian Ocean. (codes refer to test locations in Fig. 5.6).

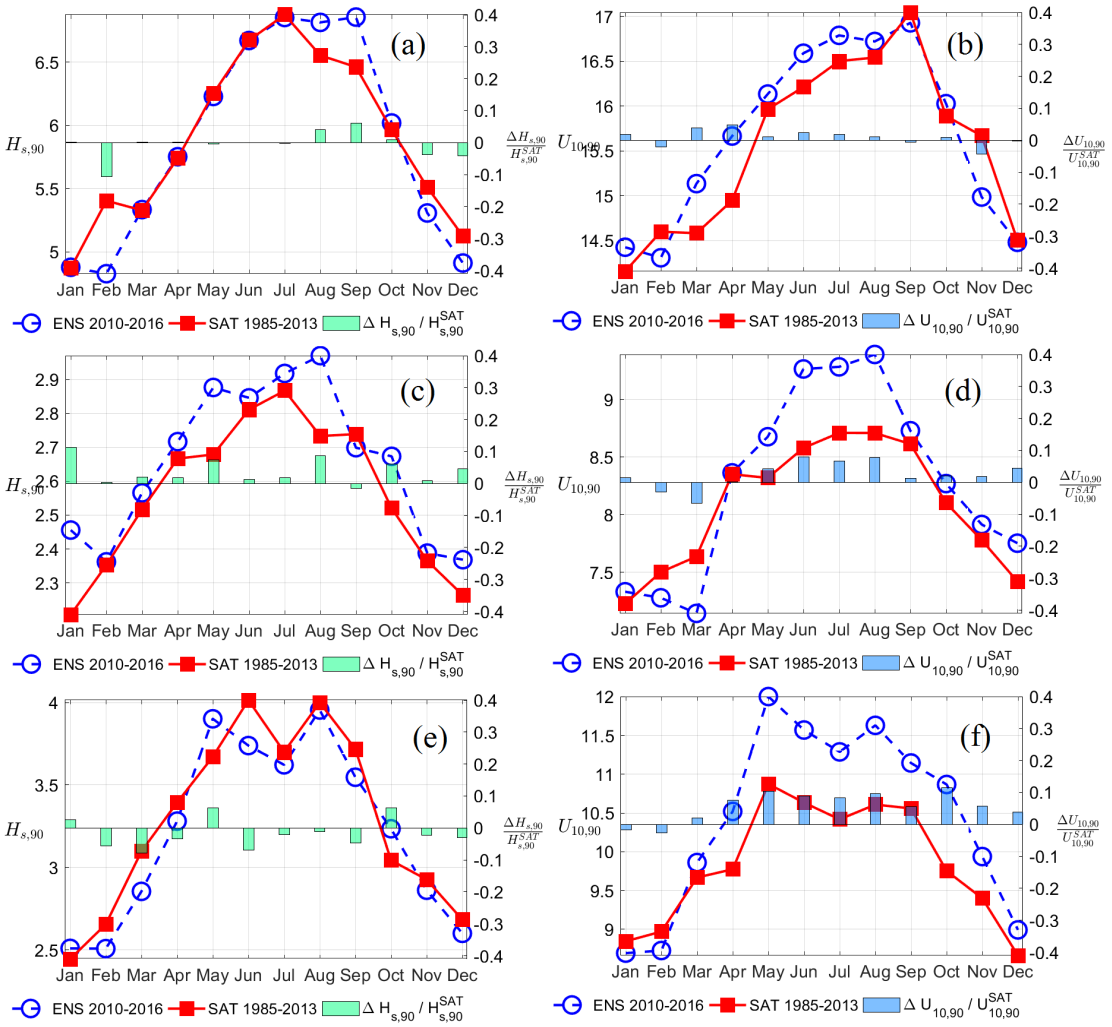


Figure 5.11: Comparison of $H_{s,90}$ (left) and $U_{10,90}$ (right) monthly 90$^{\text{th}}$ percentiles between the ENS at +240 h lead time and SATellites for the test locations in the Southern Hemisphere (SH): (a,b) SO1 Southern Ocean, (c,d) SEPO1 South East Pacific Ocean, (e,f) SAO3 South Atlantic Ocean (codes refer to test locations in Fig. 5.6).

Section 5.5).

Criterion 3: Spurious trend due to model updates

The ECMWF IFS was significantly upgraded in February 2010 and March 2016 with major changes to the horizontal resolution[1]. To minimize the impact from model updates, we selected only forecasts between March 2010 and February 2016 as the horizontal resolution remained unchanged in this period.

Criterion 4: Representative time interval

The last criterion regards the representative time interval of each forecast: Δt. This parameter is of crucial importance for the EV estimates. The value determines the equivalent time period for the ensemble data set as defined by *Breivik et al.* (2013),

$$T_{\text{eq}} = M \times N \times \Delta t, \tag{5.2}$$

where M is the number of forecasts, i.e. twice a day for six years (4384 values), N is the number of ensemble members (50). *Breivik et al.* (2013, 2014) assumed a representative value of $\Delta t = 6$ hours. Although this seemed to yield reasonable results, there is no strong theoretical justification for the choice. Noting that r_{ACC} was relatively small at a lead time of +216 h, there is justification for using lead times longer than or equal to this value. Therefore we consider the period from +216 h to +240 h. That is, +216 h, +222 h, +228 h, +234 h and +240 h. For each ensemble member, the maximum H_{s} or $U_{10\text{N}}$ is then selected from among these lead times. This maximum is thus the representative maximum for this 30 h interval and $\Delta t = 5 \times 6\,\text{h} = 30\,\text{h}$. This approach provides a sounder justification for the estimation of the time interval than that assumed by *Breivik et al.* (2014).

As noted in Section 3 and in Figure 5.1, this representative interval should not be confused with the 12h interval between forecasts in the IFS. Although forecasts are initiated every 12 hours, we take values at long lead times (+216 h to +240 h). Therefore, values from successive forecasts are independent. If forecasts were issued every 24 h (for example), it would not change the representative interval $\Delta t = 30$ h. It would, however, mean that the total number of data points would be halved in the analysis. See Section 5.5 for further details.

5.5 Extreme value analysis estimates

Here we compare global return period estimates obtained using the ensemble analysis described in Section 3 with the Initial Distribution Method applied to the

[1]See the list of major model updates maintained by ECMWF: `https://goo.gl/2G5A7x`.

same short-duration model data and POT estimates obtained with longer duration ERA-Interim and buoy data sets.

5.5.1 ENS direct return estimates

In order to undertake EVA using direct return estimates (DRE) of 100-year return values for significant wave height, H_{s}^{100}, and 10 m neutral wind speed, $U_{10\mathrm{N}}^{100}$, the highest 1000 peaks were extracted from the 2010-2016 ENS data set for each point in the interpolated grid (see Fig. 5.1d). The selection of just the 1000 highest peaks has a negligible probability of contamination for this large dataset, and it saves a large amount of computational time (*Breivik and Aarnes*, 2017).

At each location, the maximum value of the five lead times was selected and the values were sorted to rank the 1000 highest peaks. From these peaks the return period is then found directly from an interpolation based on the equivalent time period. This is computed as in Eq. (5.2) (Fig. 5.1),

$$T_{\mathrm{eq}} = 6\,\mathrm{years}\times 365.25\times 2\,\mathrm{forecasts\ a\ day}\times 50\,\mathrm{ens\ members}\times 6\,\mathrm{h}\times 5\,\mathrm{lead\ times} = 750\,\mathrm{years}. \tag{5.3}$$

Thus from a 6 year ensemble archive we obtain a 750-year equivalent time period. The 100-year return period value can then be directly estimated from the (in sample) ranked data set. The ratio $\tau = 750/100 = 7.5$ shows that the 7.5^{th} peak of the 1000 highest peaks is the estimate for the 100-year return value. This value can then be interpolated directly from the ranked data without the need to assume a PDF or extrapolate to a desired probability level.

Fig. 5.12 shows global values of H_{s}^{100} using this approach. Despite the fact that the Southern Ocean is well known for large wave conditions, the largest extremes occur in the Northern hemisphere (both the Atlantic and Pacific Oceans). This result is consistent with previous model estimates of extreme wave heights (*Caires et al.*, 2004; *Sterl and Caires*, 2005; *Breivik et al.*, 2014), as well as satellite estimates (*Vinoth and Young*, 2011).

Tropical cyclone paths in both the Atlantic and Pacific tend to track east to west and then turn north along the eastern shorelines of these basins. The paths of these low pressure systems appear to be reproduced in Fig. 5.12 with the extreme waves in both the Atlantic and Pacific extending from high latitudes towards the Equator along the eastern boundaries of each basin. Also, baroclinic instability could contribute to the variability of the extremes in these areas due to the contrast of cold air over land with warm air over the ocean. Globally, the values are comparable to the estimates for the 2003-2012 period of *Breivik et al.* (2014) but slightly lower. There are two possible causes for this difference. First, in the present analysis, we select the highest value of the 5 lead times compared

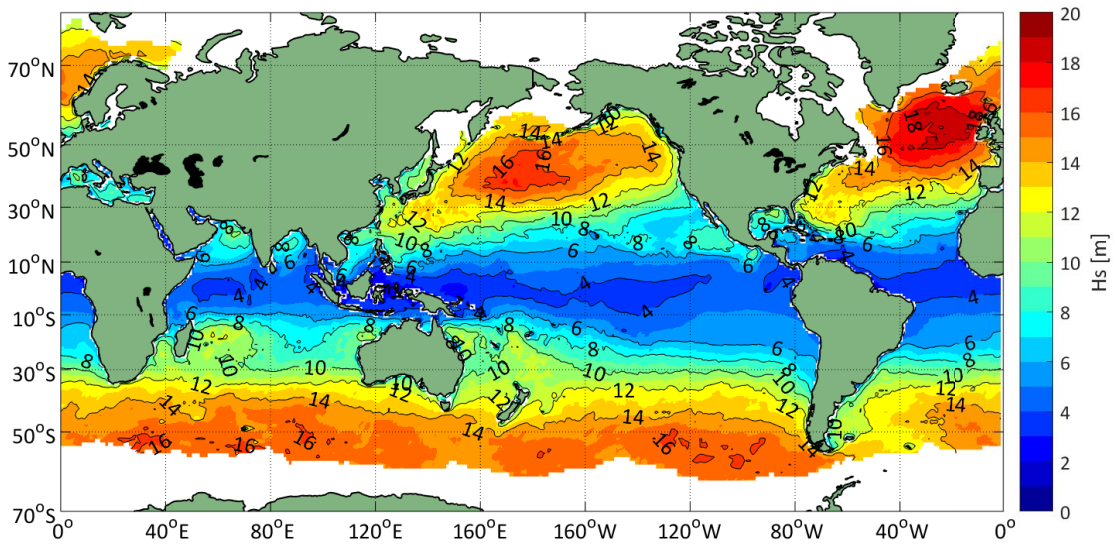


Figure 5.12: ENS 2010-2016 Direct Return Estimates of 100 year Significant Wave Height H_{s}^{100} with $\Delta t = 6$ h $\times$ 5 $= 30$ h. The estimates are taken from a reconstructed time equivalent period of 750 years.

to *Breivik et al.* (2014) who selected only the value at one lead time. However, in these two cases the probability levels at the 100-year event are also different [7.5/1000 for the present analysis compared to 2.29/1000 for *Breivik et al.* (2014)]. Secondly, the actual simulation periods are different [2010–2016 for the present analysis and 2003–2012 for *Breivik et al.* (2014)].

Similarly, we derived 100-year return values of neutral 10 m wind speed, U_{10N}^{100}, also using direct return estimates from the 7.5^{th} ranked highest peaks (Fig. 5.13). The global estimates of extreme values of U_{10N}^{100} compare well with *Breivik et al.* (2014), except for the Northern hemisphere mid-latitudes of the Western Pacific and Western Atlantic oceans where the return period estimates found in the present work are smaller than those found by *Breivik et al.* (2014). The spatial distribution of U_{10N}^{100} is similar to H_{s}^{100}, as one might expect. However, there are notable differences. The U_{10N}^{100} values exhibit a greater degree of spatial noise compared to H_{s}^{100}. This reflects the greater spatial variability of wind speed, U_{10N}. Significant wave height, H_{s}, responds to wind speed over an extended fetch, meaning that it is a spatially and temporally integrated quantity and the extremes are thus smoother.

There is also a suggestion of a band of higher U_{10N}^{100} values across the Pacific at a latitude of approximately 10° N. This is most likely a signature of the tropical cyclone belt. Generally, however, the spatial resolution of the atmospheric ensemble (32×32 km) is insufficient to fully resolve tropical cyclone winds.

The 95% confidence intervals of the direct return estimates were computed

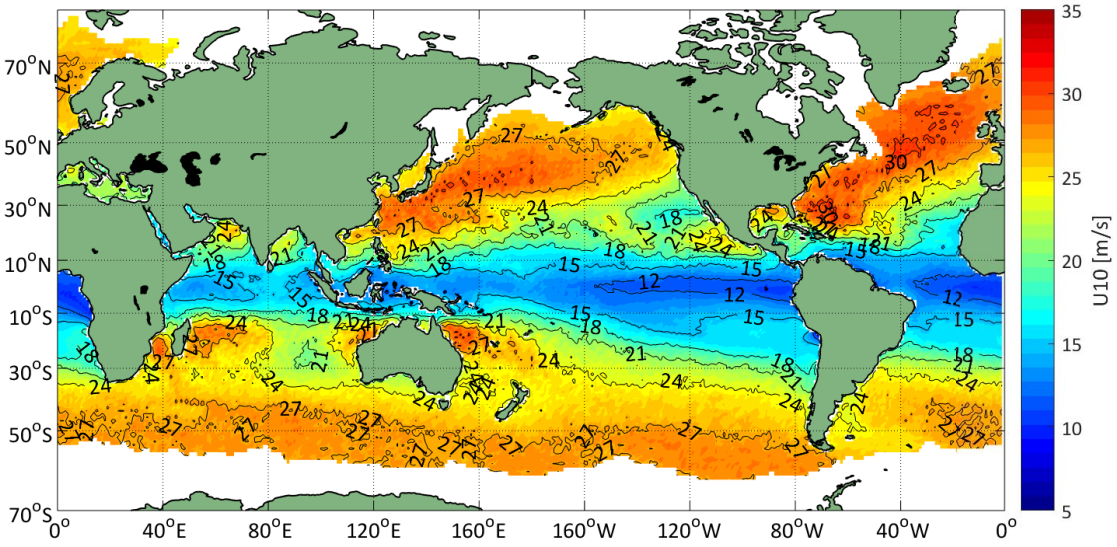


Figure 5.13: ENS 2010-2016 Direct Return Estimates of 100 year wind speed $U_{10\mathrm{N}}^{100}$ with Δt = 6h x 5 = 30h. 750 years time equivalent period.

from 500 bootstraps of the 1000 highest peaks considered in the EVA (*Breivik et al.*, 2014; *Breivik and Aarnes*, 2017). As outlined by *Breivik and Aarnes* (2017), tail estimation where only the k highest values in a bootstrap sequence are required, can be greatly economized by keeping only a subset of the original data set consisting of the K_0 highest values. Obviously, $K_0 > k$ as a bootstrap must be formed by at least the k highest values. However, such an estimate based on k bootstrapped values drawn from the K_0 highest has a probability of being biased. This can be called the probability of contamination, p_c (see Eq 2 by *Breivik and Aarnes* 2017). The bias appears because the bootstrapped values are sometimes slightly too high as they should have been drawn from the discarded bulk of the data set (values smaller than $X_{N,N-K_0}$), where $X_{N,i}$ is the i-th lowest of N entries in a data set, cf *Breivik and Aarnes* (2017). The bias is normally extremely small, and the probability of contamination can be made arbitrarily small by making K_0 sufficiently large. In our case we keep the $K_0 = 1000$ peaks and should therefore bootstrap a slightly smaller number ($k \approx 925$) to keep the probability of contamination below 99%. The bias introduced by bootstrapping 1000 values can be shown to be negligible and for practical reasons we therefore set $k = K_0 = 1000$. The results for H_s^{100} are shown in Fig. 5.14. The values shown in the figure are the total confidence limit span (i.e. $\mathrm{CI}_{0.975} - \mathrm{CI}_{0.025}$). The results show that the span of the confidence limits increases from the Equator to higher latitudes, largely reflecting the magnitude of the H_s^{100} values. Values of the span are less than 0.5 m at the Equator and increase to approximately 1.5 m at higher

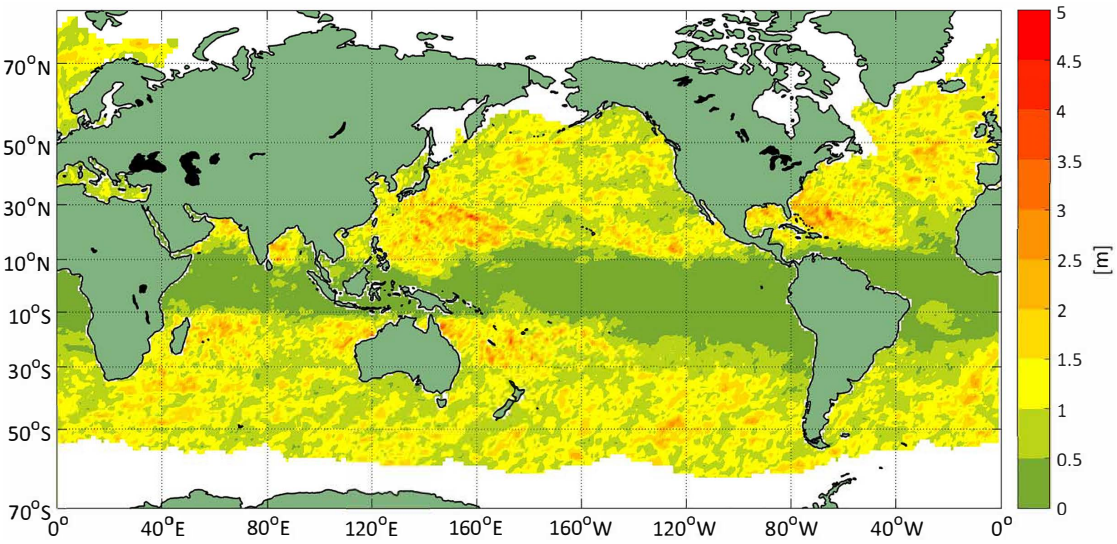


Figure 5.14: Width of ENS 2010-2016 Direct Return Estimate 95 confidence intervals for 100 year Significant Wave Height H_s^{100}. These are computed from 500 bootstraps of the 1000 highest peaks selected from the maximum of 5 different forecast lead times.

latitudes. There is some noise in the results, reflecting the variability which comes about from interpolating between values in the in-sample PDF. This variability can be reduced by smoothing the PDF by fitting a predetermined function form for the PDF. Following *Breivik et al.* (2014), an exponential distribution was fitted to the 500 bootstrap samples of the 1000 highest peaks. This results in smaller values for the confidence limit span and a smoother distribution. The span was smaller than reported by *Breivik et al.* (2014), reflecting the 750-year synthetic data set in the present analysis compared to 229 years in *Breivik et al.* (2014). This occurs despite the fact that the period considered in the present analysis is 6 years compared to the 10 years for *Breivik et al.* (2014). Compared to traditional EVA techniques (e.g. POT applied to a 30-year period) the present analysis results in quite small values of confidence limit span, demonstrating the advantage of the in-sample estimation technique.

The confidence limits for U_{10N}^{100} are shown in Fig. 5.15. The distribution is similar to H_s^{100}. Values range from approximately 1 m s^{-1} at the Equator to 3 m s^{-1} at higher latitudes. Both Fig. 5.14 and Fig. 5.15 show area of higher confidence limits for regions impacted by tropical cyclones, reflecting the variability of relatively rare events.

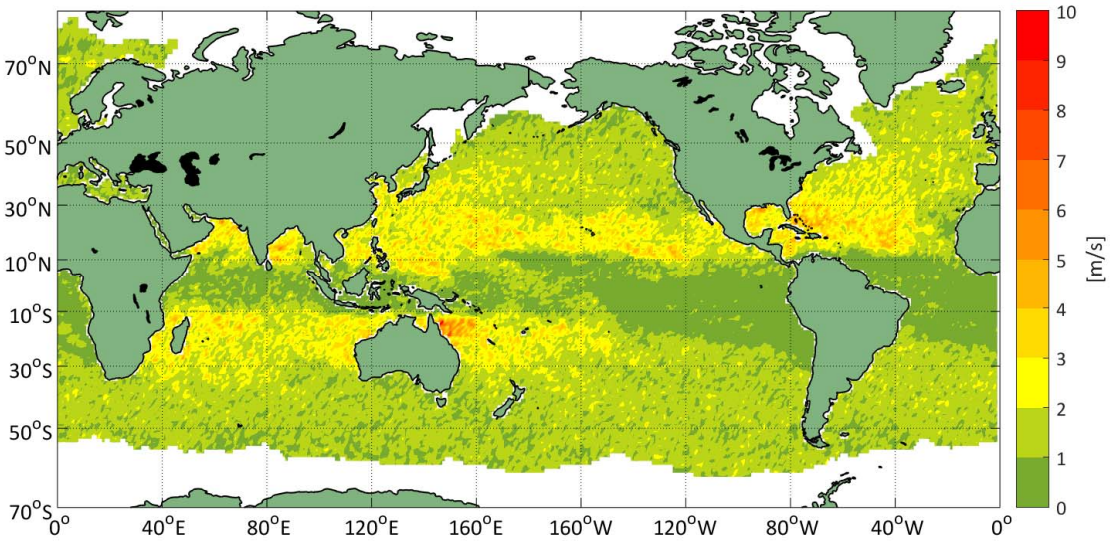


Figure 5.15: Width of ENS 2010-2016 Direct Return Estimate 95 confidence intervals for 100 year wind speed U_{10N}^{100}. These are computed from 500 bootstraps of the 1000 highest peaks selected from the maximum of 5 different forecast lead times.

5.5.2 ENS Initial Distribution Method

As noted above, the present analysis uses an archive of six years' worth of ECMWF ensemble forecasts issued twice daily with 50 members. In order to make an assessment of how the extreme value analysis of the ensemble compares against a more traditional approach *of a six-year time series* we have performed an IDM analysis where we have selected two ensemble members and one lead time from each forecast (+222 h for ENS#1 and +240 h for ENS#50). This represents the same amount of data as a six-year time series with six-hourly resolution (which is the temporal resolution of for example ERA-Interim). A six-year time series would be too short for traditional EVA. A POT analysis is not feasible with such a short record, let alone AM (which would consist of only six values). As noted in Section 5.2, the IDM approach has a number of theoretical limitations and hence we do not advocate its use. Nevertheless, for such a short time series, it would be the only alternative and it is included here as a comparison to the DRE drawn from the same climatology. A decorrelation value of 3 hours was chosen for the IDM analysis (*Vinoth and Young*, 2011). As there is no agreed theoretical choice for the IDM PDF, a goodness of fit test was performed for each of the three testing points P40, P35 and B16 (Fig. 5.2) and to the additional locations of Fig. 5.6. The best fits were achieved with a Gumbel distribution for H_s and a Weibull 2 Parameters distribution for the wind speed. The IDM results applied to the ENS data are

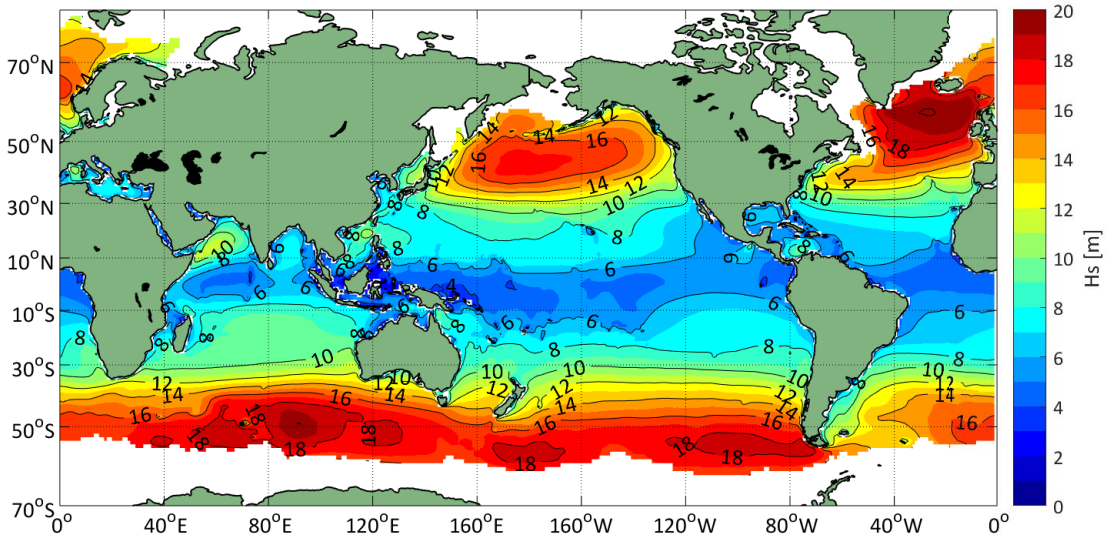


Figure 5.16: ENS 2010-2016 100 years return value estimates of H_s from the Initial Distribution Method fitted with a Gumbel distribution. Values extracted from ENS#1 at +222h lead time and ENS#50 at +240h.

shown in Fig. 5.16 for H_s^{100}, and Fig. 5.17 for U_{10N}^{100}. The spatial distributions are generally similar to the DRE estimates (Fig. 5.12 and Fig. 5.13). As noted above, the IDM effectively fits a pre-determined PDF to the core of the PDF and relies on the assumed shape of the theoretical distribution to model the tail of the PDF. As previously noted by *Vinoth and Young* (2011), this results in smooth spatial distributions, as is evident in Figs. 5.16 and 5.17. Compared to the DRE, the IDM generally yields higher extreme value estimates. As noted, this is a function of the shape of the tail of the assumed PDF and the assumed decorrelation scale (3 hrs). Therefore, this would change if different theoretical PDFs were used for the IDM fit.

The other significant difference compared to the DRE is the fact that the IDM shows the maximum values in both hemispheres of comparable magnitude. This is not surprising, as previous studies [e.g. (*Young*, 1994, 1999a)] have shown that the mean monthly values of H_s and U_{10N} are comparable in both hemispheres. Therefore, as the fit of the IDM PDF to the data is largely determined by these mean values, the resulting extremes must also be similar. In contrast, the DRE results show that, although the mean values may be similar, the PDFs for the Northern Hemisphere case have a "heavier tail". That is, although the mean conditions may be similar, the high latitudes for the Northern Hemisphere are stormier than at the same latitudes in the Southern Hemisphere. This highlights a significant shortcoming of the IDM approach.

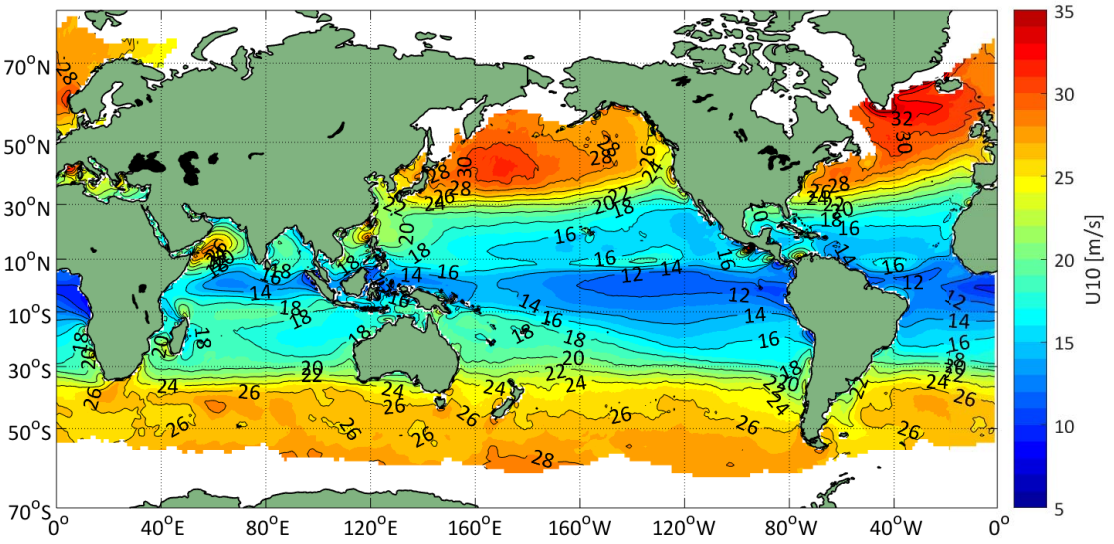


Figure 5.17: ENS 2010-2016 100 years return value estimates of U_{10N} from the Initial Distribution Method fitted with a Weibull 2 Parameter distribution. Values extracted from ENS#1 at +222h lead time and ENS#50 at +240h.

5.5.3 ERA-Interim POT method

It is common practice to determine extremes (design sea states) using reanalysis data sets such as ERA-Interim. As we have a relatively long reanalysis data set (38 years), it is possible to undertake a POT analysis. As this has a far sounder theoretical basis, it represents a more compelling comparison than the IDM approach.

As the ERA-Interim data set is long, this analysis also tests whether adopting the relatively short 6-year data set for the DRE introduces any significant bias associated with a time period which is not representative of the general climate. The POT EVA was applied to the full ERA-Interim data set (1979-2017) with a threshold at each location chosen at the local 99.8$^{\text{th}}$ percentile. An exponential distribution was then fitted to the data, the results appearing in Figs 5.18 and 5.19 for H_{s} and U_{10N}, respectively.

The spatial distributions are remarkably similar to the DRE results (Figs. 5.12 and 5.13). As for the DRE, the maximum values are located in the Northern Hemisphere and the maxima tend to be displaced towards the western sides of the North Atlantic and North Pacific, consistent with the locations of the major storm tracks.

The maximum values of H_{s}^{100} and U_{10N}^{100} are both slightly smaller than the DRE analysis (H_{s}^{100} - 14m versus 16m, U_{10N}^{100} - 27m/s versus 28m/s). These differences are consistent with previous observations that ERA-Interim tends to underestimate

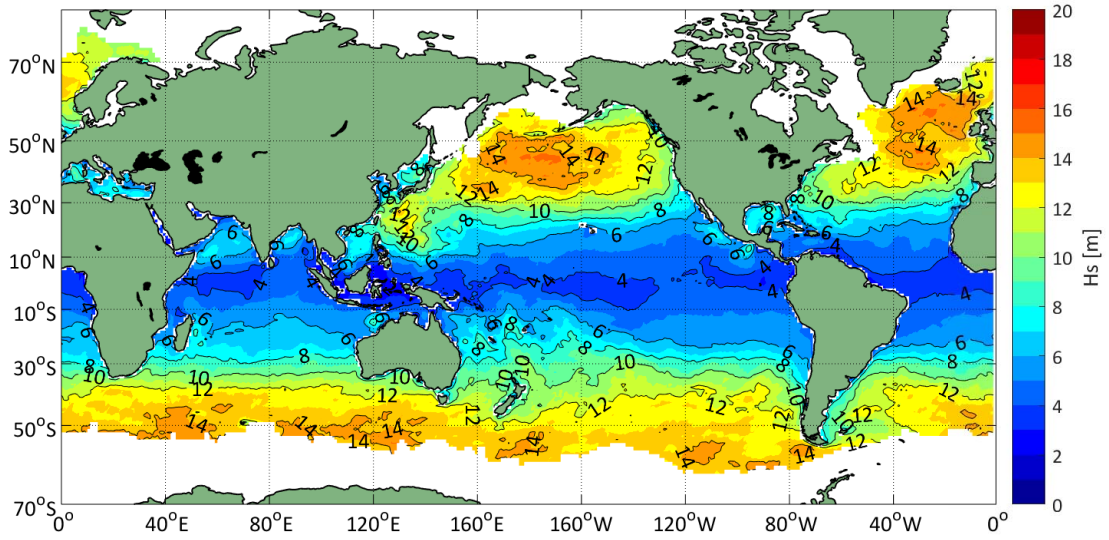


Figure 5.18: ERA-Interim 1979-2017 Peaks Over Threshold 100 years RP Significant Wave Height H_s^{100} - 99.8% threshold, fitted with an exponential distribution.

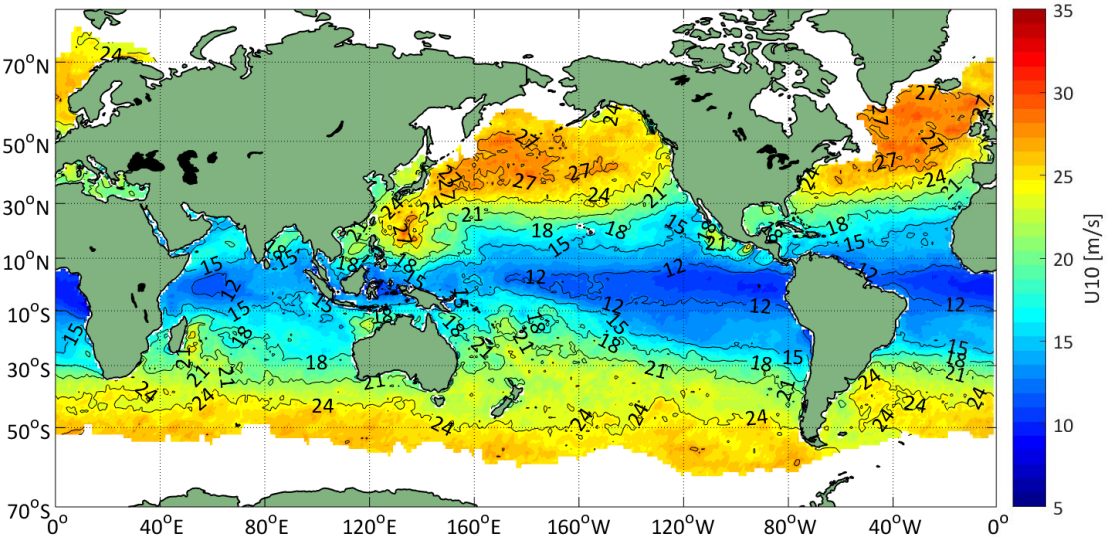


Figure 5.19: ERA-I 1979-2017 Peaks Over Threshold 100 years RP wind speed U_{10N}^{100} - 99.8% threshold, fitted with an exponential distribution.

the means (*Stopa and Cheung*, 2014) and it seems that this behavior is present also for the extremes. Nevertheless, these differences are comparable to the DRE confidence limits and smaller than those which would be applicable to the POT analysis. As such, the results are in agreement and provide strong evidence that the DRE yields results comparable to more commonly used EVA approaches.

5.5.4 In situ measurements—NDBC buoys

The ENS return values are here compared with buoy data. In this case, we use selected buoys from the American National Data Buoy Center (NDBC). Buoys were selected on the criterion that the mooring water depth is more than 300 meters, that the buoy is more than 100 km offshore (to ensure the model data are not influenced by land, considering the relative coarse native resolution of the model of about 54 km for the wave field and 32 km for the atmosphere, here interpolated to a $1.0° \times 1.0° \approx 110$ km), and that the time series is longer than 20 years. The buoys that meet these requirements are shown in Fig. 5.20.

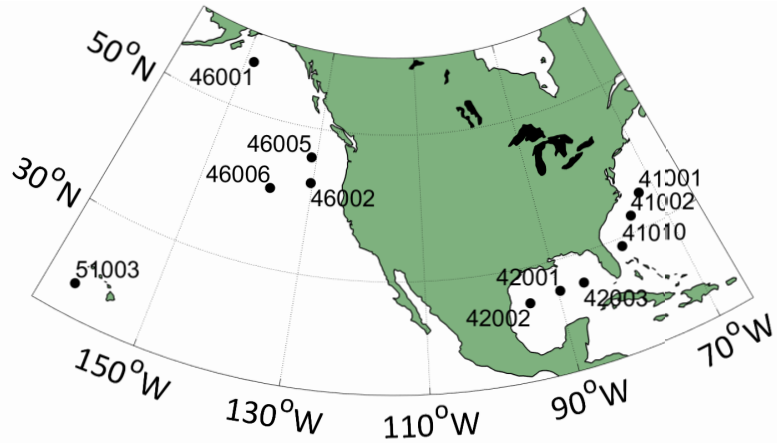


Figure 5.20: NDBC buoys that meet the requirements for EVA comparison to ENS model forecasts

NDBC buoys record wave height and wind speed every hour. In order to compare these measurements to the 6 h forecasts of the ENS data set used in the EVA, we average the buoy data at a ± 2-hour interval, centered at synoptic times, 00–06 and 12–18 UTC (*Bidlot et al.*, 2002). We then select peaks above the 90$^{\text{th}}$ percentile, with a separation time greater than 48 h. The time separation ensures storm independence following *Lopatoukhin et al.* (2000). A Peaks Over Threshold (POT) analysis was then applied to the data fitted to a Generalized Pareto (GP) distribution. The results are shown in Table 5.6 for H_{s}^{100} and Table 5.7 for $U_{10\text{N}}^{100}$. For comparison the ENS DRE at the closest grid point to the buoy location were

selected. In the tables the 100 years ENS DRE return periods are reported with the corresponding 95% confidence intervals, computed as described in Section 5.5.1. The comparison shows generally good agreement with the DRE results for H_s^{100}, except for the buoys in the Gulf of Mexico (WMO 42001, 42002 and 42003) which give approximately 10% higher return value estimates than the ENS DRE analysis. This is arguably due to the model being incapable of accurately reproducing (at the resolution of ≈110 km) the physics of this enclosed sea and the presence of unresolved tropical cyclones.

Regarding the winds, the major difference is found for Hawaii (WMO51003). Some of the islands are smaller than the model resolution. Thus, the model cannot correctly represent the topography of the area, which may provide shelter for in-situ wind measurements. In general there is good agreement between DRE results from model ensembles, and EVA results from buoys.

Table 5.6: DRE comparison with EVA of buoy data measurements averaged over the synoptic times. EVA of 100-year return values of H_s^{100} are: A - buoys peaks over 90% threshold fitted with Generalized Pareto distribution; B - ENS direct return estimates from 1000 ranked peaks. (B-A)/A - differences normalized on the buoy estimates. ENS DRE 95% bootstrap confidence intervals are reported as reference for the comparison.

	A	B		(B-A)/A
WMO	POT GP 90% [m]	ENS H_s^{100} DRE [m]	CI95 H_s^{100} DRE	Normalized diff %
41001	11.6	11.61	[10.51, 12.16]	0.09
41002	12.4	12.45	[11.46, 13.40]	0.40
41010	10.32	10.83	[10.12, 11.40]	4.94
42001	8.73	9.92	[8.85, 10.33]	13.63
42002	7.77	8.56	[7.41, 9.06]	10.17
42003	9.61	10.58	[9.85, 11.59]	10.09
46001	12.05	11.98	[11.54, 12.45]	-0.58
46002	13.79	13.39	[13.09, 14.09]	-2.90
46005	12.55	13.66	[12.98, 14.02]	8.84
46006	14.34	13.43	[13.22, 14.34]	-6.35
51003	6.48	7.09	[6.76, 7.23]	9.41

Table 5.7: DRE comparison with EVA of buoy data measurements averaged over the synoptic times. EVA of 100-year return values of 10 m wind speed (U_{10N}^{100}) are: A - buoys peaks over 90% threshold fitted with Generalized Pareto distribution; B - ENS direct return estimates from 1000 ranked peaks. (B-A)/A - differences normalized by the buoy estimates. ENS DRE 95% bootstrap confidence intervals are reported as reference for the comparison.

	A	B		(B-A)/A
WMO	buoys U_{10N}^{100} POT GP 90% [m/s]	ENS U_{10N}^{100} DRE [m/s]	CI95 U_{10N}^{100} DRE	Normalized diff %
41001	26.91	27.36	[26.34, 28.32]	1.67
41002	25.15	28.09	[27.3, 29.11]	11.69
41010	23.97	27.8	[26.18, 29.27]	15.98
42001	24.69	26.12	[24.62, 27.23]	5.79
42002	24.45	24.87	[23.26, 26.41]	1.72
42003	26.94	26.48	[25.47, 27.09]	-1.71
46001	24.39	24.99	[24.54, 25.55]	2.46
46002	25.42	26.31	[25.49, 26.93]	3.50
46005	24.55	26.65	[25.6, 27.39]	8.55
46006	24.45	26.01	[24.95, 26.36]	6.38
51003	16.35	20.3	[19.02, 21.49]	24.16

5.6 Discussion

The present work extends the method by *Breivik et al.* (2013, 2014) to estimate wind and wave extremes from operational forecast model ensembles. We focused on the significant wave height and wind speed from a 6-year archive (2010-2016) of the ECMWF operational forecast model. We derived direct return estimates (DRE) of H_s^{100} and U_{10N}^{100} from the 1000 highest peaks, extracted from the maximum value of 5 different lead times (+216 h, +222 h, +228 h, +234 h and +240 h). This methodology strengthens the approach proposed by *Breivik et al.* (2013) for the construction of synthesized dataset with a more theoretically justified forecast representative interval Δt. Also, because the synthesized dataset is longer, there is a reduction in the confidence intervals of the return period estimates, compared to *Breivik et al.* (2014).

The global results for both wind speed and wave height DRE estimates are in good agreement with a POT analysis of the 38-year long ERA-Interim data set. Similarly, results are in good agreement with a POT analysis of a range of buoy locations. Both of these comparisons provide strong validation for the approach.

The H_s^{100} estimates are found to be similar to the results found by *Breivik et al.* (2014) from the 2003-2012 archive, obtained from a single lead time +240 h (day 10). Values found in this work are slightly lower due to the different method used to identify peaks and assign the 100-year probability levels. The present wind speed estimates are generally smaller, especially in the western regions of the North Atlantic and North Pacific basins, and for the Southern Ocean. Differences are caused by the different probability levels, as mentioned for wave heights, and by the different model cycles considered. It is also possible that the weather of individual years may have influenced the estimates slightly. It is also worth noting that there are uncertainties in high wind speed measurements from buoys, platforms, ships or satellites, against which these models are calibrated (*Powell et al.*, 2003). As a consequence, forecast models may be biased in the representation of strong winds (*Pineau-Guillou et al.*, 2018), thus the evaluation of wind speed return period values obtained from these models must be carefully considered.

Furthermore, the higher correlations between ensemble members found in the tropical areas suggests paying particular attention to the evaluation of the H_s^{100} estimates in these areas. However, the correlation of the extremes from the tail of the PDF have been demonstrated to be lower than the anomaly correlation found for the whole body of the data (*Breivik et al.*, 2013).

As mentioned, in the DRE of H_s^{100} (Fig. 5.12) and U_{10N}^{100} (Fig. 5.13), there is a suggestion of what is most likely a tropical cyclone (TC) belt in the Central Pacific, at around 10° N. In support of this, ongoing research on EVA from satellite measurements show a similar shape in the North Central Pacific. The present DRE analysis shows possible TC tracks also along the Asian east coast, typically known for its typhoons, the east coast of Madagascar, the North West and North East of Australia, and along the east coast of North America. This demonstrates the strength of the DRE methodology compared to the IDM estimates in detecting the possible TC influence on wind and wave extremes. That is, the IDM which extrapolates the extremes from a distribution function fitted to the entire body of the data smooths the spatial distribution of extremes, masking possible TC influences. Even though the DRE showed this potential, increased model resolution is still needed to accurately estimate extremes in TC areas. The confidence limits were also larger in TC areas, denoting higher uncertainties for the direct return estimates of the 100 year return periods. Today, ECMWF produces a HRES forecast at 9 km resolution, and 50 ensemble members at 19 km resolution. The radius to maximum winds of a TC is typically 10 to 60 km. These increased resolution model ensembles could provide useful return value estimates of TC extremes.

The present work also compared the ENS IDM estimates, with an IDM approach for the ERA-Interim reanalysis. The ERA-Interim H_s^{100} results were gen-

erally lower compared to the ENS data set. This was consistent with the Q-Q analysis for the testing locations (Fig. 5.8). The new ECMWF ERA-5 climate reanalysis data set (*Hersbach and Dee*, 2016), based on the model cycle 41r2, could be of interest to compare with model ensemble output analysis conducted in this study. Finally, the comparison with NDBC buoy EVA showed good agreement with the ENS DRE for both H_{s}^{100} and $U_{10\mathrm{N}}^{100}$. This demonstrated that the present methodology can be applied to estimate extremes of ocean wind and waves.

5.7 Conclusions

This chapter has extended the approach to wind and wave extreme estimates using atmosphere-wave model ensembles from the ECMWF operational forecasts by *Breivik et al.* (2013, 2014). The main advantage of the ensemble approach presented here is the ability to synthesize a long duration equivalent dataset for EVA. Considering ensemble members at advanced lead times as independent realizations of atmospheric and ocean surface wave states allows the evaluation of extreme values with classical extreme value methods (*Coles et al.*, 2001). This paper demonstrated the potential of this innovative approach in evaluating wind and wave extremes over the global oceans. With the expected increase in NWP model resolution over the next decade, further improvement is to be expected in model estimates of the ocean and atmosphere (*Bauer et al.*, 2015) that will further strengthen the present approach. Focus on tropical cyclone winds and wave height in ensemble forecasts is the natural next step in order to understand whether increased resolution could correctly reproduce these extreme events, and hence provide reliable EVAs.

A low correlation between ensemble members is accepted in EVT. However, the level of correlation should always be tested as it may affect the analysis. It is also important to note that, the 100-year significant wave height and wind speed results are representative of a specific period, in this case the February 2010 – March 2016 time window. This means that the 750 years synthesized dataset represents only the climate conditions present in those years, thus, excluding long-term climate oscillations. Further research should be focus on applying this EVA approach to different historical time slices and compare the 100-year return period results.

Stationarity has been assumed in this chapter work, however this may not always be the case for return value estimates of atmospheric and oceanic variables, as shown by (*Young et al.*, 2011; *Young and Ribal*, 2019). Future EVA should consider possible increases or decreases of significant wave height and wind speed extremes for the different ocean basins (*Hemer et al.*, 2013; *Wang et al.*, 2014; *Aarnes et al.*, 2015), and integrate the present approach with climate projections. This approach is implemented in the next step of the present thesis (Chapter 7),

where an ensemble of future projections is used to determine the changes in the 100-year significant wave height by the end of the 21st Century.

Chapter 6

Future projections of wind-wave extremes

6.1 Introduction

The first five chapters of the present thesis investigated wind and wave climate trends from past century models and reanalysis, and introduced a novel approach to estimate extreme ocean surface wind speed and wave height. The novel EVA approach used an ensemble of independent and identically distributed operational forecasts, to reduce the confidence intervals of the 100-year return period estimates (chapter 5). To meet the stationarity requirement, the analysis was done on a six years dataset, from 2010 to 2016. This means that the 100-year wind and wave extreme estimates were derived from a short time window assumed to be representative of the long-term wind and wave extreme climate. This is a common assumption in reference EVA studies. However, the analysis of mean climate trends (chapter 3), and recent reference studies (*Young et al.*, 2011; *Young and Ribal*, 2019), showed that the ocean wind speed and wave height climate is not stationary. This is a crucial aspect to consider if we are to estimate future metocean extremes, and design climate resilient coastal and offshore structures and operations.

The present level of knowledge on future wind-wave extremes is limited to a few studies (*Wang et al.*, 2014; *Shimura et al.*, 2015; *Appendini et al.*, 2017) with no reliable future extreme value analysis. The next part of the thesis focuses on estimating future wind-wave extremes using Global Circulation (or Climate) Model (GCM) projections. This is done applying the novel ensemble EVA approach, introduced in chapter 5, to an ensemble of global wave model runs forced with seven independent GCM surface wind speed fields. This extensive dataset is used to pool ocean storms and estimate long-term extremes. In this way, the future changes in extreme events are found by comparing historical and future projection

datasets. The ensemble approach results show, for the first time, statistically significant changes in extreme wind-wave events by the end of the 21st Century.

The present chapter describes the theory behind future projections of the Earth's climate. Then, it focuses on future projections of wind-wave events, discussing the major sources of uncertainty connected to this field of research (*Wang et al.*, 2015; *Morim et al.*, 2018, 2019). Overall, the objective of this chapter is to provide an introduction to the motivations behind the novel ensemble EVA approach to estimate future wind-wave extreme projections, presented in chapter 7.

6.2 Atmospheric-Ocean Global Climate Models (GCM)

The Earth's climate varies in response to internal and external factors. The Earth's internal variability is characterized by large scale decadal, or inter-decadal oscillations that may complicate climate patterns. To distinguish a climate *signal* from the background *noise* of the Earth's internal variability is not straightforward. Available observational time series are often too short, and studies in several science fields have developed tentative indicators, such as the signal-to-noise ratio (*Hawkins and Sutton*, 2012), to understand if a signal can be discerned. The inhomogeneity of the observational time series and model reanalyses also poses numerous challenges, as demonstrated for the case of metocean variables, chapter 3 (*Meucci et al.*, 2020a). In addition to internal factors, external anthropogenic (greenhouse gas emissions) or natural (solar or volcanic activity) factors drive the Earth's climate changes.

The objective of climate scientists is to understand how the Earth's climate is changing from past observational and model time series, and additionally to project the climate into the future. Atmospheric and ocean Global Circulation Models (GCM) are the most advanced tools available today to simulate the response of the global climate system to natural or anthropogenic factors. GCMs are numerical models that mathematically solve the Navier-Stokes equations on a three-dimensional grid, and project future climate states under different scenarios. Due to the limitations imposed by the high computational costs of running such models, the GCMs resolution are typically three hours, on a temporal scale, and 200 km, on a spatial scale. Thus, chaotic small scale systems are not well resolved by GCMs. Nonetheless, these models are capable of representing daily weather and long-term climate characteristics with reasonable accuracy.

The main variables considered in GCMs to estimate the Earth's climate changes are: solar radiance, volcanic eruptions, solar radiance absorption and reflection,

energy and momentum transfer, greenhouse gases and aerosol, sea ice, clouds, water vapour, and land surface use. As outlined by the recent Intergovernmental Panel on Climate Change (IPCC) Assessment Reports (AR4, AR5), greenhouse gas (GHG) emissions are warming the Earth. Following the preliminary Special Report on Emission Scenarios (SRES) evaluated in the IPCC AR4, which considered no GHG emission mitigation strategies, Representative Concentration Pathways RCP, that cover the period 1850-2100 (*Van Vuuren et al.*, 2011), were developed to understand the effects of different GHG emission scenario on the Earth's climate, including different mitigation strategies (IPCC AR5). These pathways follow different human development trajectories that affect greenhouse gas emissions and concentration. They are termed *representative* as they describe a set of different climate scenarios, all published in the literature. Four different RCPs were selected by a panel of IPCC experts: RCP2.6, RCP4.5, RCP6, RCP8.5 (*Moss et al.*, 2008). The numbers that label each pathway are based on the different radiative forcing target levels (W/m^2) found from reference research studies. The RCP2.6 is the very low forcing level, RCP4.5 and RCP6 are the mid-level emissions scenarios, and RCP8.5 is the very high emission scenario.

Recently, in parallel with the RCPs, a group of climate scientists have developed and published Shared Socio-Economic Pathway(s) (SSP). These are pathways of global development trend that, in parallel with the RCPs, are being investigated with Coupled Model Intercomparison Project (CMIP) phase 6 climate models, that will support the preparation of the IPCC sixth assessment report (AR6).

A large number of GCMs have been developed internationally to investigate the uncertainties around future climate. All of these models show different biases and variability. This results in a series of inter-model uncertainties particularly important in the assessment of a potential climate signal for specific variables.

It is common practice to assess the significance of a detected climate signal by reviewing different research studies at the global and local scales. The change is consider significant if it is agreed by most of the scientists that studied it. Low to high confidence labels are assigned to each of the Earth's climate change phenomena studied, comparing the different analyses and approaches published in the literature. A different problem is, whether to attribute a signal of change to anthropogenic factors.

In the following sections we focus on the analysis of future wind-wave projections, and further extend the discussion to wind-wave extremes, as an introduction to the analysis in chapter 7.

6.3 Future wind-wave projections

Atmospheric and ocean GCMs do not include wind-wave parametrizations, thus no direct wind-wave output is currently available, and the understanding of changes in wind-wave climate is limited to the relationship with other geophysical variables such as mean sea level pressure or sea surface temperature (*Cavaleri et al.*, 2012; *Hemer et al.*, 2013; *Field et al.*, 2014). To downscale the results of atmospheric GCMs, and generate wind-wave climate projections, independent studies used either statistical or dynamical approaches. A statistical approach exploits the relationship with other well known variables, such as mean sea level pressure, to project the wind-wave climate (*Wang and Swail*, 2006). A dynamical downscaling approach consists instead of running high-resolution regional climate models forced with GCM simulated climate. The surface wind speed created in this way, is used to force a global wave model and project wind-wave conditions for the future (*Hemer et al.*, 2013). To collect and coordinate future wind-wave climate projections, and answer questions around poorly analysed and understood wind-wave climate changes (*Stocker*, 2014), the Coordinated Ocean Wave Climate Project (COWCLIP) group (*Hemer et al.*, 2012) was formed, under the support of the World Climate Research Programme and the Joint Technical Commission for Oceanography and Marine Meteorology of the World Meteorological Organization (WMO) and Intergovernmental Oceanographic Commission (IOC) of UNESCO. The objective is to produce a series of wind-wave climate projections and assess the uncertainties connected with the assessment of future wind-wave climate.

As part of this community driven effort, a review of 91 global and regional 21st century wind-wave climate projection studies was performed by *Morim et al.* (2018). Across these models a consensus on an increase of mean significant wave height across the Southern Ocean was found. *Wang et al.* (2015) suggest this to be connected with the dipole pattern developed in the Southern Ocean where contrasting effects of the increasing greenhouse gases emissions and the ozone depletion, are causing a poleward shift of the westerly wind belt (*Sigmond et al.*, 2011). Together with the Southern Ocean region, the most significant projected increase in SWH is found in the Arctic region (*Dobrynin et al.*, 2012), with already observed changes in the wave height due to the melting ice sheet which generates new ice-free waters (*Waseda et al.*, 2018).

Moving to other regions, there is general consensus on an increase in mean significant wave height also in the Eastern Tropical Pacific (*Wang et al.*, 2015; *Morim et al.*, 2018), where the strengthening of the trade winds plays a role (*de Boisséson et al.*, 2014). In contrast, the North Atlantic shows a general decrease in SWH from low to mid latitude, with some discussion on a increase at the high latitudes, due to a possible poleward shift of extra-tropical cyclone tracks (*Hemer et al.*, 2013; *Wang et al.*, 2015; *Aarnes et al.*, 2017). An increase is found across all seasons and

different scenarios in the Baltic Sea (*Morim et al.*, 2018). All the other regions of the global oceans lack consensus on future changes in SWH (*Morim et al.*, 2018).

A more comprehensive analysis of such projected changes in wave climate is contained in *Morim et al.* (2018) and *Morim et al.* (2019).

6.3.1 Sources of uncertainty in wind-wave projections

In the *Morim et al.* (2018) and *Morim et al.* (2019) wind-wave climate projection studies, different sources of uncertainty were identified and classified. In particular, three dominant sources of uncertainties emerged. These are the RCP, the inter-model uncertainties due to different biases and variability of the GCMs, and finally the wave model limitations in representing the wave climate, also discussed in this thesis in chapter 2. In the same way *Wang and Swail* (2006) previously defined three main sources of uncertainty. These are, the scenario, the model, and the method uncertainties.

In general, climate model (or inter-model) uncertainties are the most dominant of the three main sources of uncertainty (*Wang et al.*, 2015; *Morim et al.*, 2019). In particular they account for 40% of the total variance around the global oceans, except for the Eastern Tropical Pacific, dominated by the regularly blowing trade winds, where the inter-model uncertainty account for only 5 to 10% of the total variance *Wang et al.* (2015). As this is smaller than the climate signal, we could argue that a significant change is to be expected in the Eastern Tropical Pacific wind-wave climate. An additional source of uncertainty is added by the downscaling techniques.

6.4 Future wind-wave extreme projections

Climate change may affect extreme events in unexpected ways, that could be significantly different from the changes found in the mean values (*Cooley*, 2009). Therefore, it is crucial to study the impacts of climate change on the extremes, as opposed to the impacts on the mean climate. However, as extremes are particularly rare events, it is challenging to find robust statistical analysis of their future changes, due to uncertainties in both historical and future projections of extreme values.

Research studies on future projections of extreme wind-wave events lack consensus everywhere around the globe (*Morim et al.*, 2018). There are only a few exceptions found for the higher percentiles in the Southern Ocean and North Atlantic, where there is a larger consensus on the possible changes at the extremes (*Morim et al.*, 2018).

Future extremes projections are characterized by large confidence limits, as many physical processes are still not well understood (*Field et al.*, 2012). This significantly reduces the level of confidence on the reliability of model simulations, and ultimately the extreme estimates, let alone their projections into the future.

Two different atmospheric systems dominate wind-wave extreme events: extra-tropical cyclones and the Tropical Cyclones (TC). A medium level of consensus has been reached around extra-tropical cyclones events. For instance, there is medium confidence in the scientific community that these events will decrease in frequency in each hemisphere (*Field et al.*, 2012), thus reducing also the number of wind-wave extremes.

The picture is more complicated for Tropical Cyclones as these are localized events with projections for reduced frequency, but increased intensity, and for these reasons it is very challenging to predict tropical cyclone extremes. Research studies show some consensus on an increase of the average intensity of tropical cyclones (*Field et al.*, 2012; *Walsh et al.*, 2016; *Knutson et al.*, 2019), this means more frequent category 4 and 5 TCs. It is still not clear if this will happen in all the ocean regions impacted by TCs. The frequency is expected to either decrease or remain unchanged, according to the climate science community (*Field et al.*, 2012; *Walsh et al.*, 2016; *Knutson et al.*, 2019).

The changes of these extreme atmospheric events directly affect wind-wave extremes, and there are many unresolved questions which result in uncertainties for future projections. This is particularly evident when we analyse the short observational time series available, and consider the limited physics of current GCMs. Thus, it is important to consider the different sources of uncertainties connected with future extreme estimates, and this thesis focuses on reducing statistical and inter-model sources of uncertainty.

The next chapter presents an ensemble approach to estimate future extreme projections, taking advantage of a series of global wave model runs forced by independent GCMs wind fields. This allows the pooling of ocean storms to represent an equivalent time period of more than 100 years. Such an approach is an advantage when we estimate 100-year return period events, and how these values are changing between historical and future periods.

6.5 Conclusion

The present chapter described the current level of knowledge on wind-wave climate and future projections of extremes. The general view presented suggests ways to critically analyse the next chapter EVA results. It should be noted that, climate change studies are constantly improving, thus further increasing the level of understanding around wind-wave event changes. Whether the changes found in these

studies, and in the present thesis, can be attributed to anthropogenic influence, or to internal or external natural factors, is still an open question. In general, particular care is needed for the interpretation of both past climate trends (chapter 3) and future projections (chapter 7) of wind-wave events. The next chapter proposes a novel approach that reduces the uncertainties around future projections of wind-wave extremes, and it finds, for the first time, statistically significant changes in extreme wind-wave events.

Chapter 7

Projected 21st Century changes in extreme wind-wave events

7.1 Introduction/Summary

The present chapter is an extended edited version of a work published in Science Advances entitled "Projected 21st Century changes in extreme wind-wave events" (*Meucci et al.*, 2020b).

An innovative approach is used to estimate the magnitude of global changes in extreme wave conditions by 2100, as a result of projected climate change. This approach generates a synthetic dataset from an ensemble of wave models forced by independent climate simulation winds, enhancing statistical confidence associated with projected changes in extreme wave conditions. Under two IPCC representative greenhouse gas emission scenarios (RCP4.5 and RCP8.5), we find that the magnitude of the 1 in 100 year significant wave height event increases by 5 to 15% over the Southern Ocean by the end of the 21st century, compared to the 1979–2005 period. The North Atlantic shows a decrease at low to mid latitudes ($\approx$ 5 to 15%) and an increase at high latitudes ($\approx$ 10%). The extreme significant wave height in the North Pacific increases at the high latitudes by 5 to 10%. The ensemble extreme value analysis approach used, means that these projected changes are, for the first time, shown to be statistically significant.

7.2 Background

Human activities in both coastal and offshore regions are strongly affected by local wind generated wave (wind-wave), events. There has been increasing interest in changes in wind-wave climate both historically and for future projections (*Hemer et al.*, 2013; *Young and Ribal*, 2019; *Morim et al.*, 2019). Long-term satellite mea-

surements have shown that wind-wave climate is changing over the global oceans (*Young and Ribal*, 2019). Coordinated international efforts (*Hemer et al.*, 2010) have been focused on collecting and analyzing international datasets and models to understand mean and higher percentile global ocean wind-wave climate. Although understanding how mean conditions are projected to change is important, it is the extreme wave conditions which are critical for both coastal and offshore areas. The design sea state for offshore and coastal structures is typically defined as the maximum significant wave height which can be expected over an N year period (*Veritas*, 2010). Offshore structures, such as oil rigs, or wind farms, and coastal protection structures (rock breakwaters, sea walls), usually consider the 1 in 100 year significant wave height as the design sea state.

Wind-wave extremes are also crucial for the determination of coastal sea levels and coastal erosion, and changes in the climate may further exacerbate the already predicted strong societal and economic impacts of wind-waves on the worlds coasts (*Hinkel et al.*, 2014; *Luijendijk et al.*, 2018). It has been estimated (*Hinkel et al.*, 2014) that in 2010, 290 million people worldwide lived below the 100-year flood level and US$9600 billion of assets were exposed to inundation. Therefore, it is of paramount importance to understand the implications of climate change on extreme wind-wave conditions. In many scientific fields, the one in N year values, (i.e. the return values) of extreme events is estimated using Extreme Value Analysis (EVA) (*Coles et al.*, 2001).

Long and homogeneous time series are needed to perform a robust analysis and confidently estimate such extreme values. However, spatial and temporal sampling (*Gulev and Grigorieva*, 2004; *Holthuijsen*, 2007), along with observational bias issues (*Large et al.*, 1995; *Powell et al.*, 2003; *Bender et al.*, 2010; *Durrant et al.*, 2013; *Young et al.*, 2017; *Young and Donelan*, 2018), limit estimates of extreme significant wave heights. Thus, extreme value estimates are typically characterized by large statistical uncertainty.

Dynamical and statistical analysis approaches have been applied to investigate possible changes in the wind-wave climate from projections of the mean and percentile values (*Dobrynin et al.*, 2012; *Hemer et al.*, 2013; *Morim et al.*, 2019), yet, aside from a few studies (*Wang et al.*, 2014; *Shimura et al.*, 2015; *Appendini et al.*, 2017; *Timmermans et al.*, 2017), a reliable analysis of the projected changes in wind-wave extremes is lacking. This results in a lack of consensus on whether the frequency or magnitude of these extreme events are impacted by climate change.

The present analysis aims to quantify how the magnitude and frequency of wind-wave extreme events may change under future projected climate scenarios. As such, we need to determine the difference between historical estimates of extremes and future projections. Determining such differences when both the historical estimates and future projections of extremes contain large statistical

uncertainty is problematic.

Therefore, here we use a novel approach which can reduce uncertainties attributable to standard EVA methods (i.e. enhance confidence in the estimated extreme values) as well as quantifying inter-model and inter-scenario uncertainties (*Wang and Swail*, 2006; *Wang et al.*, 2015; *Morim et al.*, 2018, 2019). In this approach, we analyze an ensemble of Global Climate Models (GCMs) used to force a wave model, as independent realizations of ocean wave conditions, thus enabling the pooling of the highest extreme wave conditions (after applying bias correction to reduce the effect of different GCM forcing, see Section 7.3 and Fig. 7.2). This approach forms a large dataset which increases the confidence of the statistical estimates thereof.

This approach has been shown to work well when applied to single model ensemble forecasts of historical extreme sea states (*Sterl et al.*, 2009; *de Winter et al.*, 2012; *Breivik et al.*, 2014; *Meucci et al.*, 2018), but has not previously been applied to extreme values extracted from different GCMs, and to future projections of global extremes.

Although at the present resolution, GCMs are not able to resolve localized extreme events such as the tropical cyclones (*Shimura et al.*, 2015; *Timmermans et al.*, 2017), the strength of this approach is that it produces more robust results compared to previous studies that have attempted to address the intrinsically uncertain nature of the ocean-atmosphere state (*Mentaschi et al.*, 2017).

7.3 Dataset and Methodology

An ensemble of global wave model runs of WAVEWATCH-III (*Tolman et al.*, 2009) (WWIII v. 3.14), forced with surface winds simulated by different Global Climate Models (GCMs), was used to develop a dataset of storm wave conditions. These GCMs are part of the Coupled Model Intercomparison Project Phase 5 (CMIP5), feeding into the Fifth Assessment Report (AR5) of the Intergovernmental Panel on Climate Change (IPCC). We selected the 7 WWIII model runs that performed best in terms of wave climate prediction (*Hemer and Trenham*, 2016) and applied an Extreme Value Analysis (*Coles et al.*, 2001) to model output. The GCMs used to generate the surface winds which forced the WWIII wave model (CMIP5/WWIII) were: ACCESS1.0, BCC-CSM1.1, GFDL-CM3, HadGEM2-ES, INMCM4, MIROC5, MRI-CGCM3 (*Hemer and Trenham*, 2016). The characteristics of each model are described in Table 7.1.

The data are available on the CSIRO data servers (CAWCR dataset at https://doi.org/10.4225/08/55C991CC3F0E8).

As noted above, this ensemble is a subset of the larger group used by *Morim et al.* (2019) to investigate changes in mean and 99$^{\text{th}}$ percentile significant wave

Table 7.1: Ensemble of Climate Models forcing winds and wave model characteristics.

	CMIP5	Atmosphere	Wave model		
	phase	Res. Lon. × Lat.	Res. Lon. × Lat.	Interval	Period
ACCESS1.0	5	1.88 × 1.25	1.0 × 1.0	6 h	1979–2005, 2081–2100
BCC-CSM1.1	5	2.8 × 2.8	1.0 × 1.0	6 h	1979–2005, 2081–2100
GFDL-CM3	5	2.5 × 2.0	1.0 × 1.0	6 h	1979–2005, 2081–2100
HadGEM2-ES	5	1.88 × 1.25	1.0 × 1.0	6 h	1979–2005, 2081–2100
INMCM4	5	2.0 × 1.25	1.0 × 1.0	6 h	1979–2005, 2081–2100
MIROC5	5	1.4 × 1.4	1.0 × 1.0	6 h	1979–2005, 2081–2100
MRI-CGCM3	5	1.1 × 1.1	1.0 × 1.0	6 h	1979–2005, 2081–2100

height by 2100. The smaller group was used here as the present analysis of extremes required 6-hourly data rather than monthly statistics. Such data were archived for these 7 models. In addition, the hourly data means an increase in the amount of data by a factor of approximately 700 per model. Hence, the smaller ensemble was also required to render the problem computationally tractable. The present subset of models is drawn from Clusters 2 and 4 of *Morim et al.* (2019), which were broadly representative of the larger ensemble of models. However, to confirm that this is the case, the climatology for the sub-set was determined as the mean and 99^{th} percentile significant wave height, and for the historical time period (1979–2005). These results are shown in Fig. 7.1a, b, and are comparable to Figure 1 of *Morim et al.* (2019).

Similarly the percentage changes in and between the historical (1979–2005) and future (2081–2100) time periods for both RCP4.5 and 8.5 are shown in Fig. 7.1c, d, e, f. Again, this is directly comparable with Figure 1 ($\Delta\overline{H}_{\text{s}}$) and Figure 2 ($\Delta H_{\text{s}}^{99}$) of *Morim et al.* (2019). These comparisons provide confidence that the present sub-set of models is representative of the larger ensemble consisting of 83 models.

7.3.1 EVA technique

Our aim here was to pool the data from the ensemble of the 7 bias-corrected models. To this end, first, a threshold was set at the 90^{th} percentile for each of the 7 models. Storm-independent peaks were then selected above this threshold (Fig. 7.2a), with independence guaranteed by ensuring peaks are separated by a minimum of 48 h (*Lopatoukhin et al.*, 2000). The peak values from the pooled dataset were then ranked and the highest 1000 were selected (*Breivik et al.*, 2013; *Meucci et al.*, 2018). This process of selecting storm peaks from each of the models in the ensemble and pooling them to form extreme value histograms is shown diagrammatically in Fig. 7.2. The procedure was done for both historical (Fig. 7.2b,d)

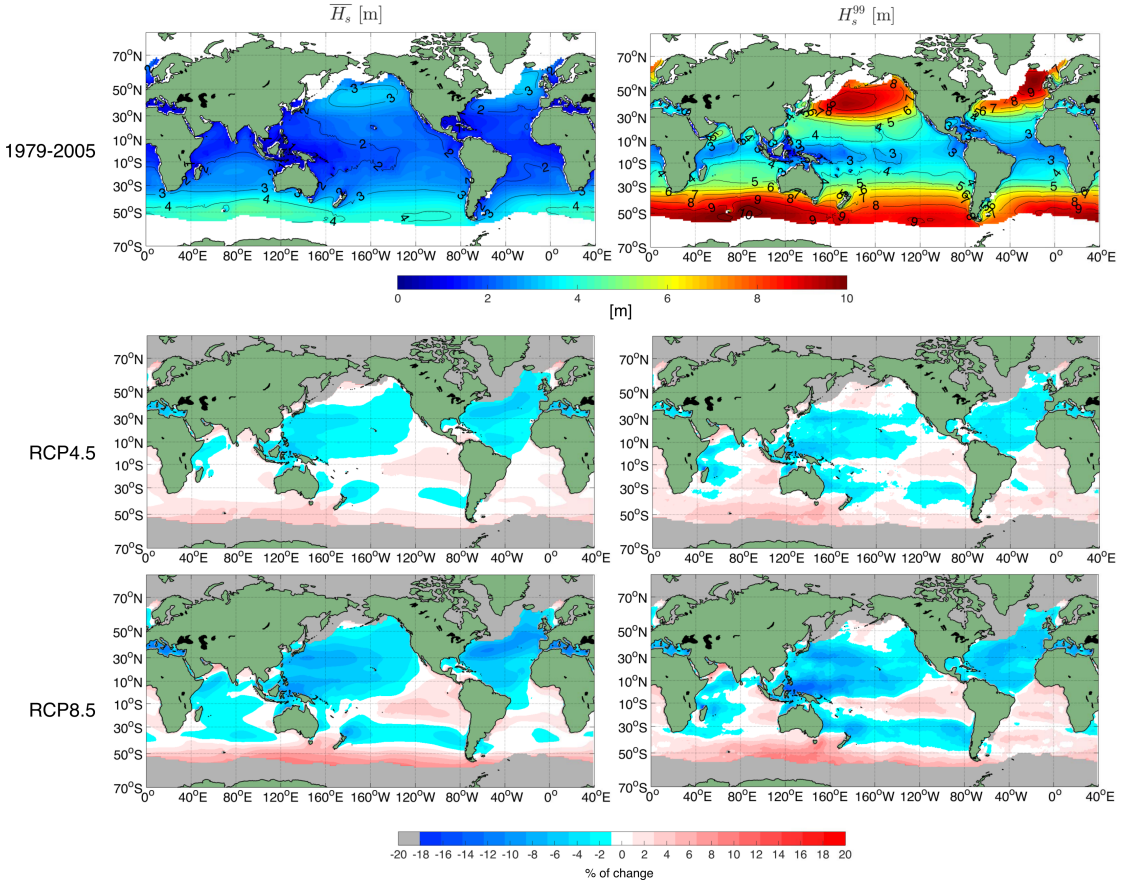


Figure 7.1: Mean, $\overline{H}_{\mathrm{s}}$, and 99$^{\mathrm{th}}$ percentile, H_{s}^{99}, significant wave height obtained for the present 1979–2005 ensemble. In the lower section: differences in mean and 99th percentile significant wave height between the historical (1979–2005) and future (2081–2100) time periods for RCP4.5 and 8.5. The grey area are the locations outside of the ensemble domain.

and future projections 6-hourly datasets (Fig. 7.2c,e), with and without bias correction (see paragraph 7.3.3). As for a Peaks over Threshold (PoT) approach, we fitted an exponential distribution to the 1000 highest peaks and found the 100-year return period value at the desired probability level from the relationship (*Vinoth and Young*, 2011; *Takbash et al.*, 2018):

$$\Pr\{x < x^{100}\} = 1 - \frac{N_Y}{100 \cdot N_{POT}} \tag{7.1}$$

where N_{POT} is the number of peaks (1000), and N_Y is the equivalent duration in years of the pooled dataset from which the peaks were extracted.

In order to perform the EVA, it is necessary to define a representative time

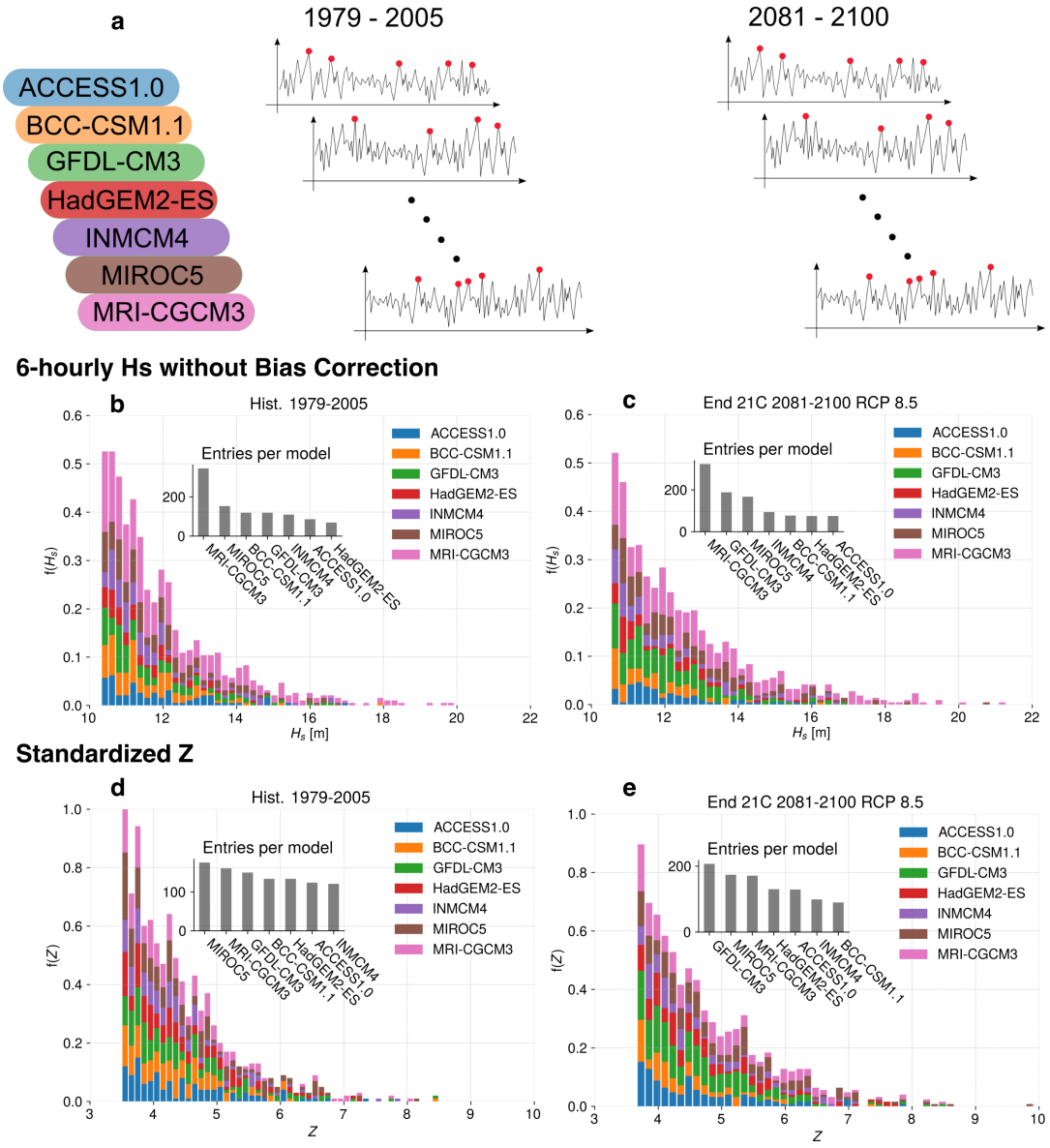


Figure 7.2: Schematic of the ocean storms pooling method. **a**, extreme wave heights are selected from each CMIP5 model as values over the 90$^{\text{th}}$ percentile threshold with a minimum time distance of 48 h (storm independent peaks). The peaks are then ranked and the highest 1000 form the tail of the dataset to which an exponential function is fitted. The stacked histograms show the 1000 peaks, with contributions from all 7 models (Section 7.3). **b, c** show stacked histograms with bias correction and **d, e** with bias correction. The histograms of relative contributions from each model show a more uniform distribution with bias correction.

period of the pooled dataset (i.e. define N_{POT}). Once the extremes from each of the models are pooled, we no longer have a single time series. Rather we have a pooled set of extremes from multiple models, of which we verified the level of independence and identical distribution (Fig. 7.3).

Therefore, to compute the equivalent time period for this pooled dataset N_Y, we assumed that each of the model 6-hourly outputs are representative of a 6h hour time window, this value has been tested in previous literature (*Breivik et al.*, 2013; *Meucci et al.*, 2018). Considering leap years for the historical dataset (1979–2005), the equivalent time is given by Eq. 7.2.

$$T_{\text{eq}}^{\text{hist}} = 27\,\text{years} \times 365.25 \times 4\,\text{hindcasts a day} \times 6\,\text{h} \times 7\,\text{models} = 189\,\text{years} \quad (7.2)$$

and for the future projection (2081–2100), the equivalent time is given by Eq. 7.3.

$$T_{\text{eq}}^{\text{proj}} = 20\,\text{years} \times 365.25 \times 4\,\text{projections a day} \times 6\,\text{h} \times 7\,\text{models} = 140\,\text{years} \quad (7.3)$$

As discussed above and in the following sections, the synthesized time period guarantees smaller return period confidence intervals, reducing the uncertainties. Eq. 7.2 and Eq. 7.3 define the values of N_Y in Eq. 7.1. Note that as N_Y is larger than the desired return period (100 years), we are in sample and no extrapolation of the empirical probability distribution function (epdf) is required (hereafter, for clarity, we refer to the probability distribution function of the synthesized data as epdf, and to the probability density function fitted to these data as fpdf). The exponential distribution fitted to the epdf of the data is hence used only to smooth the tail of the epdf and aid interpolation of the extreme significant wave height at the desired probability level (i.e. one in 100-year event) (see Fig. 7.2). A similar approach to that applied here has been used to estimate projected changes in extreme significant wave height at selected North Sea locations (*Sterl et al.*, 2009; *de Winter et al.*, 2012). In contrast to the present approach, however, this previous analysis used only annual maxima from a single model.

7.3.2 Preliminary tests on raw model data

Data from the ensemble of models can be pooled to undertake EVA only if the data complies with specific criteria (*Breivik et al.*, 2013; *Meucci et al.*, 2018). That is, the data must be independent and identically distributed (i.i.d.) (*Coles et al.*, 2001). To investigate if the data are i.i.d. we selected several locations around the globe and constructed scatter plots of significant wave height values for the historical model datasets. The scatter plots are for values at the specific locations

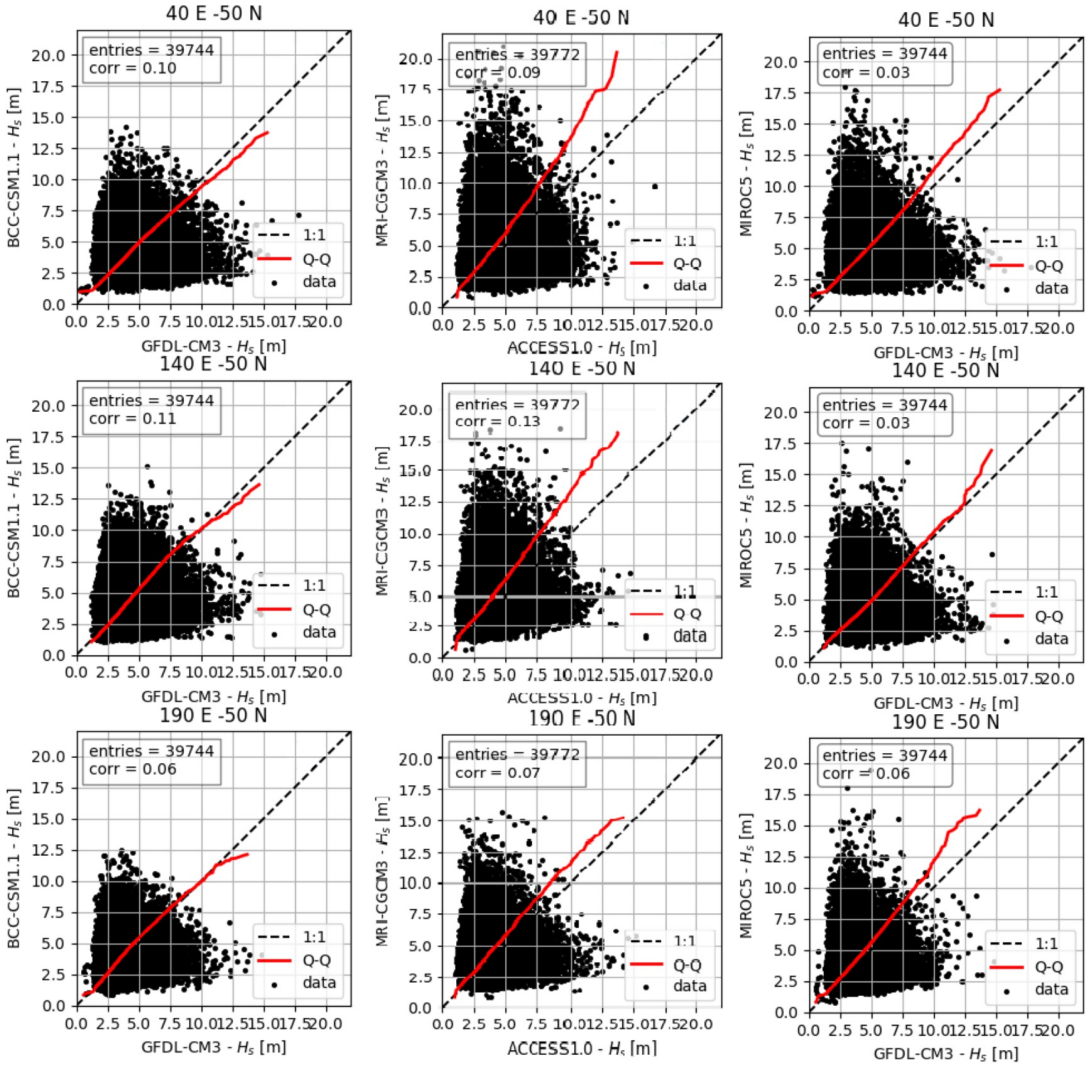


Figure 7.3: Scatter and Quantile-Quantile (Q-Q) plots at three selected locations (positions marked on plots) for combinations of different pairs of CMIP5 models (no bias correction). Large scatter indicates the models are independent, similar Q-Q results indicate they are identically distributed.

and the same times. Plots were performed between each combination of two models (Fig. 7.3). As the models are forced with different wind fields, the scatter plots show low correlation (correlation coef. ≤ 0.13) and it is assumed that the models are sufficiently independent to perform the analysis (*Breivik et al.*, 2013; *Meucci et al.*, 2018). In addition, Quantile-Quantile (Q-Q) plots (Fig. 7.3) show that generally the models are identically distributed, i.e. Q-Q line is close to the 1:1 line indicating similar epdfs. This is particularly true for the bulk of the data, with greater divergence at the extremes. To further test the relationship between the models at the extreme values, we extracted the peaks over the 90$^{\text{th}}$ percentile for each m model and compared the quantiles with the quantiles of the inter-model ensemble of extremes taken above the same percentile. We then computed the extreme quantiles RMSE (Root Mean Square Error) to define a performance parameter e_{RMSE} (Eq. 7.4).

$$e_{\text{RMSE}} = 1 - \frac{|\frac{\sum \text{RMSE}_{\text{m}}}{n} - H_s^{100}|}{H_s^{100}} \tag{7.4}$$

The performance parameter e_{RMSE} (Eq. 7.4) describes the similarity (or discordance) between model distributions of the extremes. The smaller the e_{RMSE}, the more similar the distribution of extremes between models, and thus the more robust is the ensemble EVA approach. RMSE_{m} is the root mean square error of the m model 90$^{\text{th}}$ percentile extreme quantiles in relation to the inter-model ensemble 90$^{\text{th}}$ percentile extreme quantiles, where n is the number of models (in this case 7), and H_s^{100} is the 100-year return period value estimated with the inter-model EVA approach (i.e. respectively for historical and future projection datasets). The e_{RMSE} results for both historical and future projections datasets (Fig. 7.4), show reasonable agreement between model extremes ($e_{\text{RMSE}} \leq 0.05$), except for the tropical areas, affected by tropical cyclones, that are still not adequately represented by the models, due to coarse resolution and missing physics. These results show that some GCMs seem to capture tropical cyclone characteristics whilst others still miss them. This is clearly shown by the EVA performed on each single global wave model run over the 1979–2005 historical time slice (Fig. 7.5).

7.3.3 Robustness of the results

To account for different biases and variances between the various models, we tested two bias correction approaches and a bias uncorrected approach. As outlined above, we describe the results obtained from a standardized variable correction, which we believe works quite well. This approach allows a comparison between models with different biases and variances (*Aarnes et al.*, 2017) by standardizing the significant wave height values to each model's historical mean μ_m^{hist} and stan-

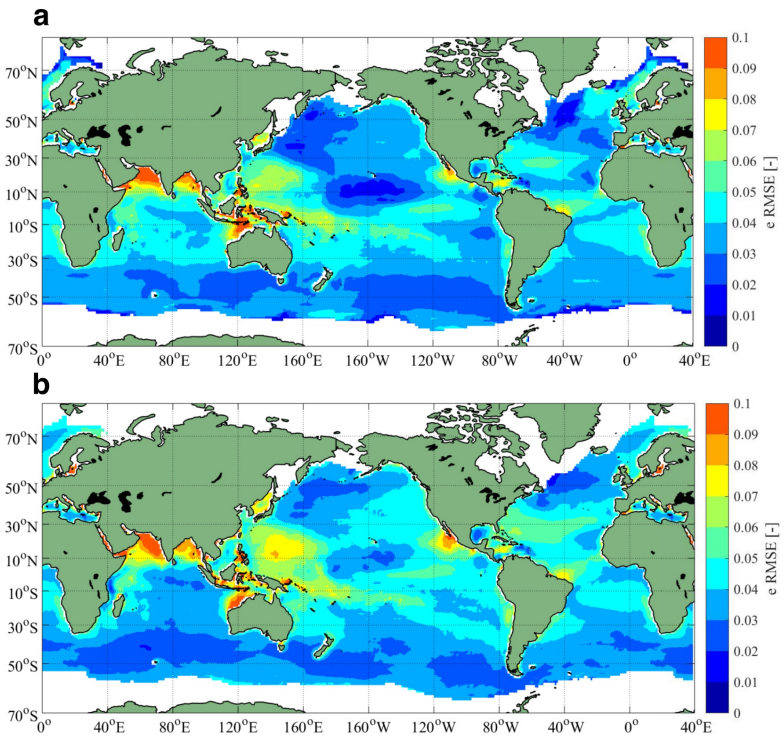


Figure 7.4: Test for similarity between distribution of extremes — **a**, historical dataset (1979–2005), **b**, end of 21st Century (2081–2100). Root Mean Square Error (RMSE) analysis of the 90th percentile extreme quantiles for each model compared to the inter-model ensemble extreme quantiles (see Section 7.3).

dard deviation σ_m^{hist} (here $m = 1, 2, ..., 7$ for the 7 models). This is done for both historical (Eq. 7.5) and future projection (Eq. 7.6) significant wave height values $H_{\text{s,m}}$.

$$Z_m^{\text{hist}} = \frac{H_{\text{s,m}}^{\text{hist}} - \mu_m^{\text{hist}}}{\sigma_m^{\text{hist}}} \tag{7.5}$$

$$Z_m^{\text{proj}} = \frac{H_{\text{s,m}}^{\text{proj}} - \mu_m^{\text{hist}}}{\sigma_m^{\text{hist}}} \tag{7.6}$$

The values of $H_{\text{s,m}}$ were standardized for each of the seven models individually, then, the storm data (extreme values) were pooled. The EVA was then conducted on the 1000 highest Z values extracted from the pooled Z_m variables. We then found the 100-year Z^{100} return period and from this computed H_{s}^{100}, inverting Eq. 7.5 and 7.6. The inversion was done using the 7-model ensemble historical mean, μ_{ens}^{hist}, and standard deviation, $\sigma_{ens}^{\text{hist}}$. The changes $\Delta H_{\text{s}}^{100}$ between historical and future projection datasets are shown in Fig. 7.15. A second analysis directly

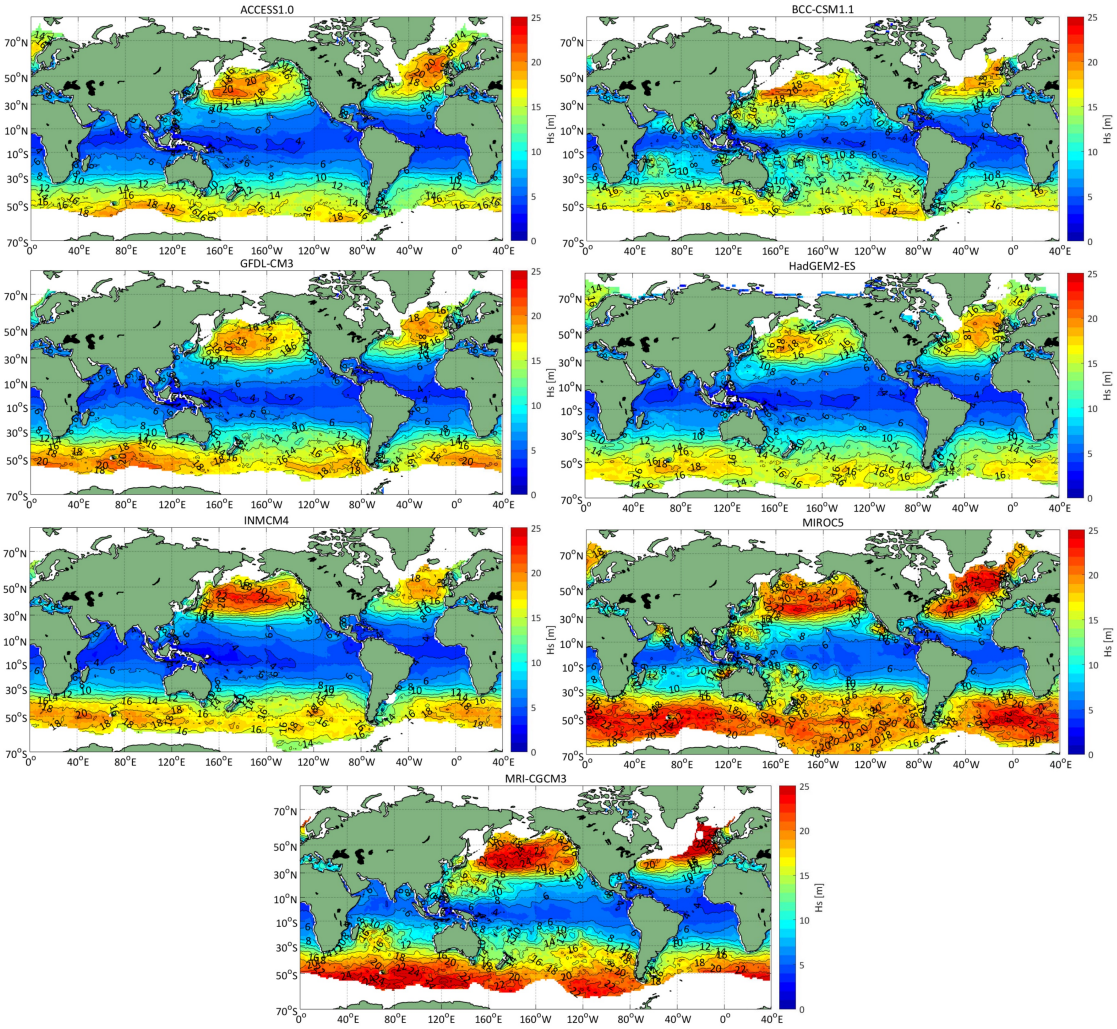


Figure 7.5: 1979–2005 100-year return period significant wave height, derived from Peaks over 99.6$^{\text{th}}$ percentile Threshold of each global wave model run. Results obtained without bias correction.

compared the extreme significant wave height values from the inter-model ensemble of the historical dataset with the extreme significant wave heights of the inter-model ensemble of future projections. No bias correction was used in this analysis. The PoT analysis was performed on the 1000 highest significant wave height peaks selected as described above. The 100-year return period changes at the end of the 21st Century (Fig. 7.6a) show a similar spatial distribution to the Z-corrected results (Fig. 7.15), especially at the mid and high latitudes of both hemispheres.

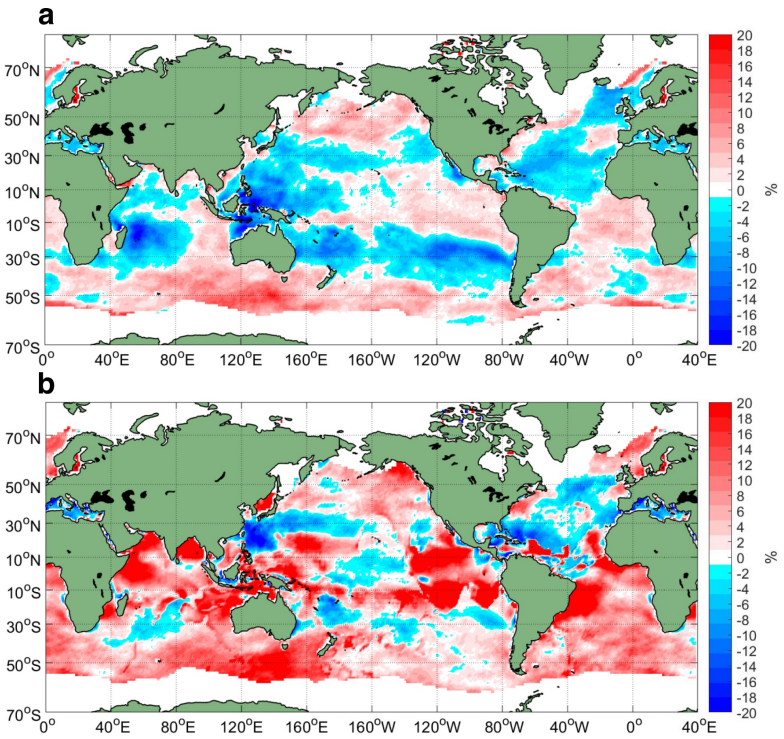


Figure 7.6: Changes in the 100-year return period significant wave height derived from 1000 data peaks at the end of the 21st Century — RCP8.5 (see Section 7.3). **a**, Results obtained without bias correction, **b**, Results obtained with bias correction applied at the 99.8th percentile level.

The main differences are found in the tropical areas where, as already mentioned above, the models have limitations in the representation of extremes (Tropical Cyclones areas in Fig. 7.5). In a third approach, we extracted the 1000 highest peaks from bias corrected historical and future datasets. We bias corrected each of the CMIP5/WWIII outputs at the 99.8th percentile level, based on the 99.8th quantile of a reference climate model run. That is the WWIII wave model forced with surface winds from the Climate Forecast System Reanalysis (CFSR) of the American National Centers for Environmental Predictions (CFSR/WWIII run).

The delta found from the 99.8$^{\text{th}}$ percentile quantile of the reference runs, was then added or subtracted to the cumulative distribution function of each of the CMIP5/WWIII runs, shifting the tail of the extremes to match the chosen quantile level. This process is similar to a correction with a quantile-based mapping method (*Li et al.*, 2010) but limited to a single quantile in order to preserve the variance of each model tail of the extremes. The changes in H_{s}^{100} (Fig. 7.6b) using this approach show greater variability than the first two approaches, but the general distributions are similar, especially at the mid and high latitudes of both hemispheres, demonstrating the robustness of the results in these regions of the global oceans.

Changes in the tropical areas remain an open question, and this bias correction approach introduces larger deviations in the results for these areas. In addition to this, we test the consistency of H_{s} future changes, performing an EVA at different return levels. These correspond to the 1 in 10, 20, 50, and 100-year return periods. The EVA extreme estimates for each return level are shown in Fig. 7.7, with the percentage of change of the 2081–2100 RCP8.5 projections relative to the 1979–2005 period. The results show the different magnitude of the estimates for both

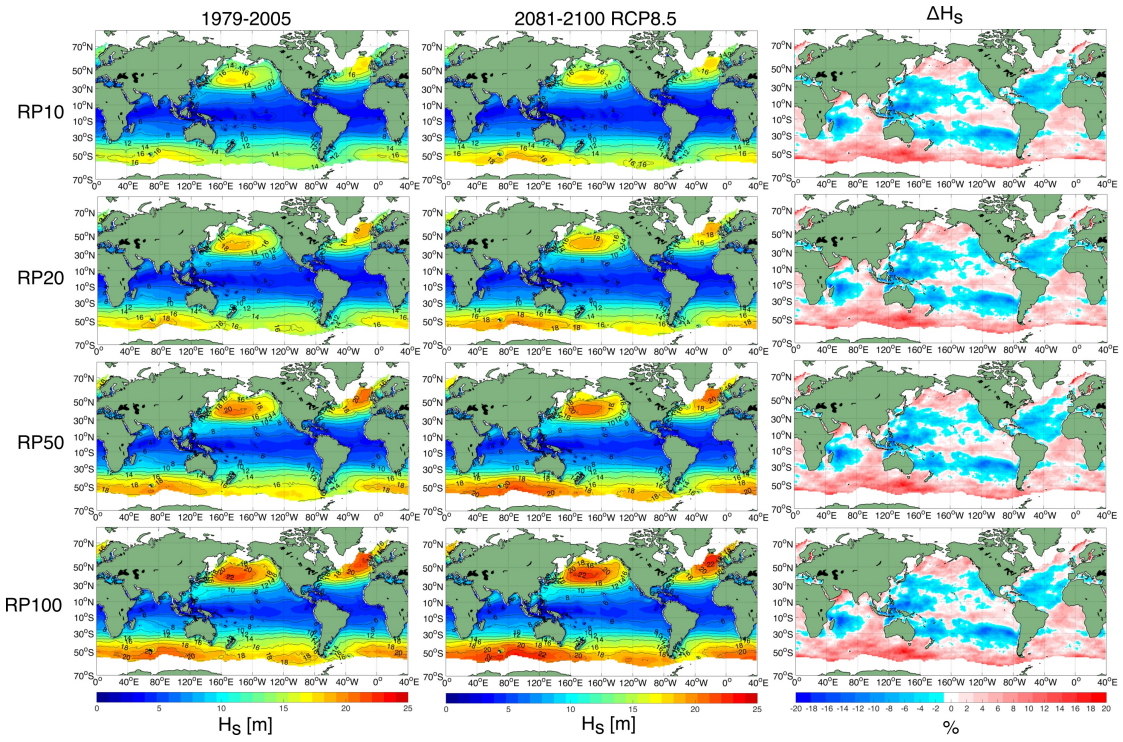


Figure 7.7: Significant wave height return period analysis. The extreme estimates are derived at different return levels (1 in 10, 20, 50, 100-year return period (RP)) from the ensemble 1000 highest peaks, fitted with an exponential distribution.

historical and future projection datasets, from the smallest (10-year return period) to the largest (100-year return period). The percentage of change in significant wave height by the end of the 21st Century, relative to the 1979–2005 period, is consistent through the different return period analyzed.

7.3.4 Model contributions

An important element of the pooled model approach is that all the models contribute to the EVA estimates. That is, the peaks composing the tail of the pooled extremes epdf come from all CMIP5/WWIII runs, rather than a small number of models dominating the selected extreme values. To test this, the model contributing each of the peaks in the ensemble EVA was tracked. The number of peaks from each model making up the highest 1000 Z values in the pooled extreme dataset was determined at selected test locations. The model contribution to the 1000 peaks for non-bias-corrected data is shown for these locations as pie charts (Fig. 7.8). This figure shows that even without any bias correction, all 7 models contribute to the fitted probability density function (fpdf) and thus the return period estimation. The bias correction process further ensures an even more uniform contribution to the peaks across the individual models of the ensemble, as shown in Fig. 7.2.

7.3.5 Statistical significance

Once the 1000 peaks are extracted from the inter-model ensemble, we can bootstrap the confidence intervals at the 95% level (*Breivik and Aarnes*, 2017). The bootstrap technique is performed by resampling 500 times the 1000 pooled peaks. For each grid point we compute the 100-year return value fitting the exponential distribution to each permutation cycle (500) sample of the 1000 values. The 95% confidence level is given by the 0.025 and 0.975 percentiles. The confidence intervals for both historical and future projection H_s^{100} (Fig. 7.9) are significantly smaller than results found with common EVA techniques (*Breivik et al.*, 2014; *Meucci et al.*, 2018) (see Fig. 7.10a, d, g). Taking advantage of the reduced statistical uncertainty we here introduce a test to evaluate the statistical significance of the extreme significant wave height changes at the end of the 21st Century. The significant wave height changes at the end of the 21st Century are regarded as statistically significant if the 95% confidence limits of the historical and future extremes do not overlap, that is:

$$|\Delta H_s^{100}| \geq 0.5 \times (\mathrm{CI}_{95,hist} - \mathrm{CI}_{95,proj}) \tag{7.7}$$

where ΔH_s^{100} is the difference between historical and future projections of H_s^{100} (Fig. 7.15), CI_{95} are the confidence intervals (i.e. the range of values in which we

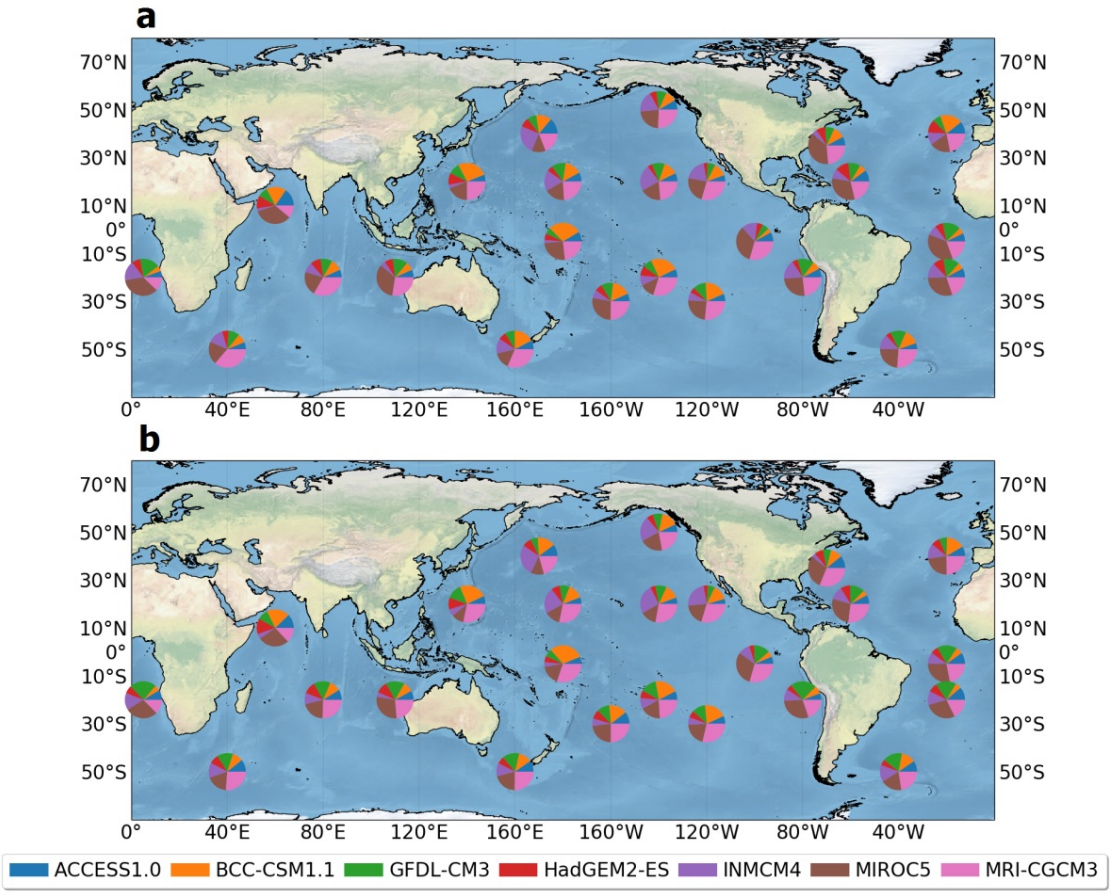


Figure 7.8: The relative contributions over individual models to the ensemble of peaks used for the EVA. Selected locations shown around the world for **a** historical (1979–2005) and **b** future projection RCP8.5 (2081–2100) non-bias-corrected datasets.

are 95% confident to find H_s^{100} for the historical $CI_{95,hist}$ and the future projection $CI_{95,proj}$ inter-model ensembles).

Statistically significant values of changes in the H_s^{100} are shown with hatching in Fig. 7.15. As can be seen in this figure, most areas where the changes in H_s^{100} are greater than 5% in magnitude (positive or negative) are statistically significant. One of the major advantages of the pooled ensemble approach used here is that the reduction in confidence limits of the extreme value estimates means that differences in extreme values can be determined with greater statistical confidence.

A further demonstration of the increase is statistical confidence due to the pooled ensemble approach is demonstrated Fig. 7.11 which shows the changes in H_s^{100} by the end of the 21st Century for RCP8.5 calculated using a Peaks over 99.6th

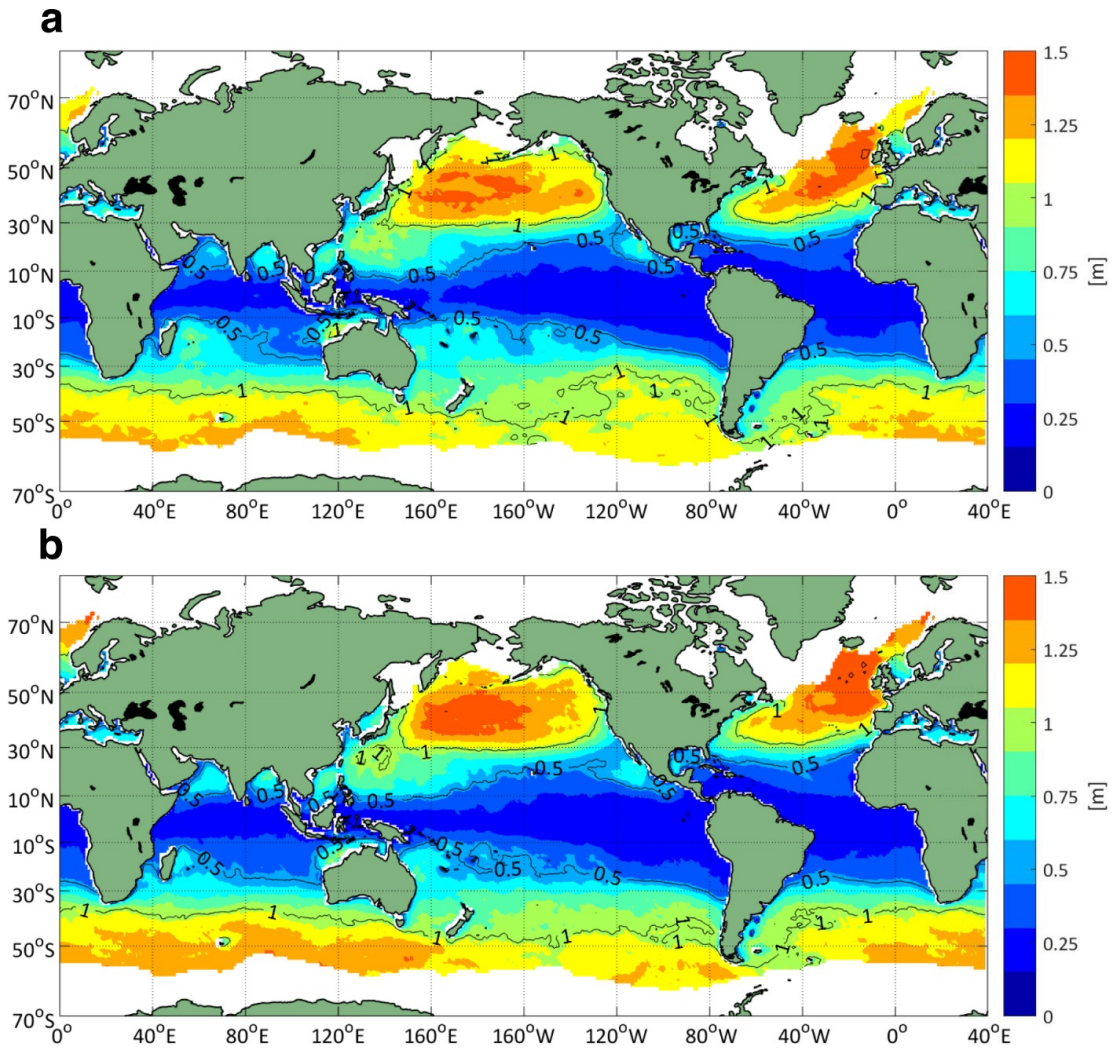


Figure 7.9: Bootstrap estimates of the 95% confidence intervals of the 100-year return period significant wave height for **a**, historical dataset 1979–2005, **b**, end of 21_{st} Century RCP8.5.

percentile Threshold (PoT) for each of the individual models in the ensemble. Fig. 7.11 can be directly compared with the ensemble result Fig. 7.15b. Although some of the spatial variations evident in Fig. 7.15b are evident when considering the models individually, the increase in statistical variability is evident. Applying the same approach as above to determine statistical confidence, we find almost none of changes in H_{s}^{100} are statistically significant at the 95^{th} percentile level when using individual models.

A further comparative test was undertaken by evaluating the average of the

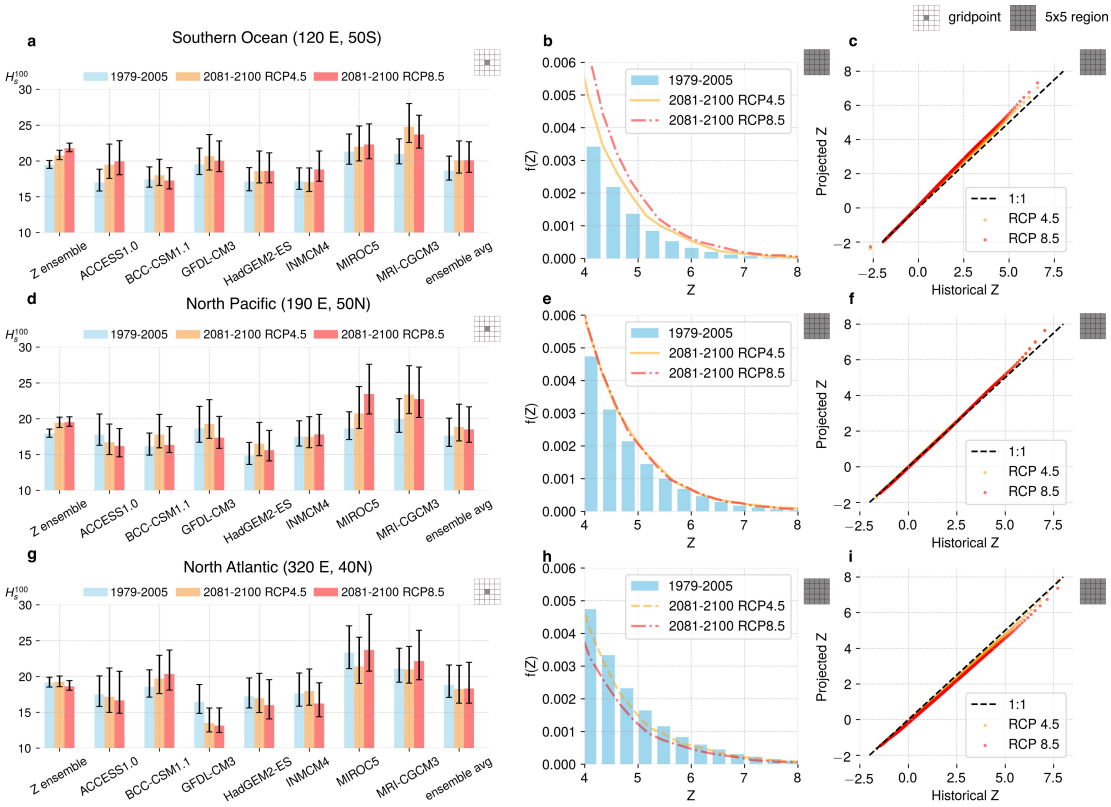


Figure 7.10: Confidence in the historical and future projection estimates of H_s^{100} at selected locations, and changes in standardized variable probability distributions for the surrounding 5°x5° regions. **a**, **d** and **g**, compare the confidence intervals of the 7 models ensemble Z standardized variable (see Section 7.3) with each model Peak over Threshold approaches at comparable threshold percentile, and the average of each model EVA. **b**, **e** and **h** compare the histograms of the historical inter-model ensemble of Z standardized variables with the future projections, focusing on the highest peaks tail. **c**, **f** and **i** are Quantile-Quantile (Q-Q) plots comparing the historical inter-model Z ensemble and future projections. The differences between the historical and future projections in the tail of the PDF illustrate projected changes in the frequency and magnitude of extreme wave heights. Note that **a**, **d** and **g**, are analyses at a grid point, whereas, **b**, **c**, **e**, **f**, **h** and **i** are analyses over 5°x5° regions surrounding the selected locations.

changes in H_s^{100} from the individual models in Fig. 7.9. This yielded a result comparable to the ensemble result, Fig. 7.15b., augmenting confidence in the robustness of the ensemble approach. However, as shown in Fig. 7.10a, d, g, the confidence limits using such an averaging approach are still much larger than for

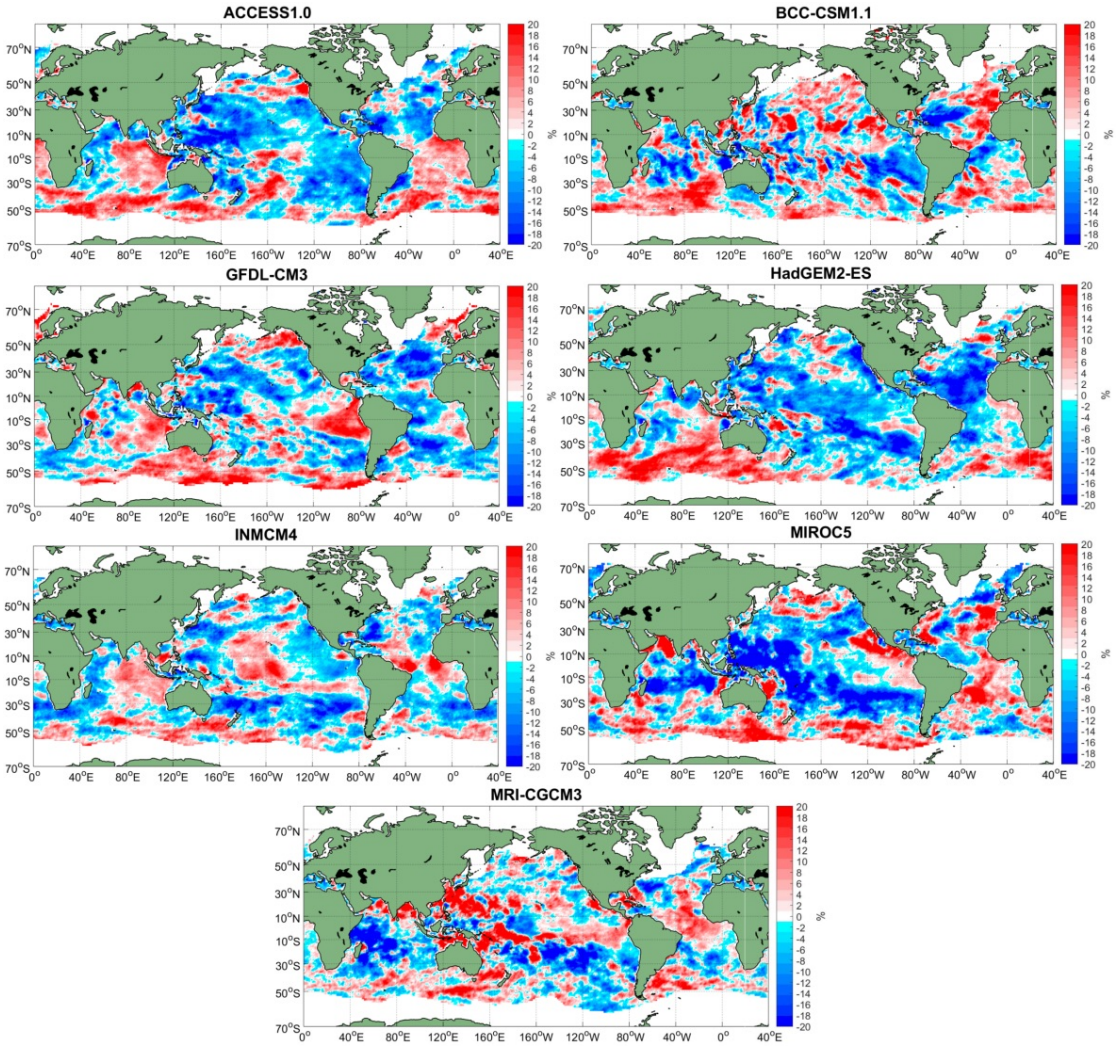


Figure 7.11: Changes in the 100-year return period significant wave height at the end of the 21st Century — RCP8.5, derived from Peaks over 99.6th percentile Threshold of each global wave model run. Results obtained without bias correction.

the ensemble approach.

7.4 Results and findings

7.4.1 Model Ensemble

We investigated changes in global 100-year return period significant wave heights, H_s^{100}, over the 21st Century, from an inter-model ensemble of 7 global wave model runs

(using the same WAVEWATCH III v3.14 (*Tolman et al.*, 2009) wave model) forced with independent GCM surface winds, part of the international Coupled Model Intercomparison Project phase 5, CMIP5 (*Taylor et al.*, 2012).

As explained in Section 7.3, the present ensemble is part of a larger grouping of models (*Morim et al.*, 2019) which has been used to examine projected changes in mean and percentile wave conditions. The sub-set of 7 models was chosen as the estimation of extreme conditions requires at least 6-hourly data over the full duration of the model runs, in contrast to the estimation of mean and percentile projections. Such hourly data was available for this sub-set which was previously tested for its wind-wave simulation skill (*Hemer and Trenham*, 2016). The smaller number of ensemble members renders the present analysis computationally feasible.

The sub-set was further tested to ensure it was representative of the larger ensemble, reproducing comparable mean and 99^{th} percentile climatology, as well as projected changes in both quantities by 2100 (see Fig. 7.1). The ocean wave extremes were pooled from the ensemble of global wave model runs, which collectively represent an equivalent time period much larger than commonly adopted EVA approaches (see Section 7.3). To be able to pool the model results in this manner, the data must be independent and identically distributed (*Coles et al.*, 2001).

In the present case, we made the pragmatic assumption that the model results are independent, as each wave model run is forced with a different GCM, independently initiated. To be identically distributed, the models must generate comparable extreme values (*Meucci et al.*, 2018) (see Section 7.3). These assumptions are verified by pair-wise model comparisons as shown in Fig. 7.3. The tail of the empirical probability distribution function (epdf) of the synthesized dataset was constructed as in Fig. 7.2, which demonstrates that there are contributions from each of the 7 models (identically distributed, see Section 7.3).

Although all models contribute as shown in Fig. 7.2, the contributions are not uniform, as shown by the nested grey bar plots (Fig. 7.2b, c). This occurs because the stacked colored bars represent the 1000 peak ensemble of non-bias corrected H_{s} model results. A bias correction (see Section 7.3), performed using the Z standardized variable, largely reduces the uneven contribution of GCMs such as the MRI-CGCM3 (Fig. 7.2d, e). After the bias correction, we fitted a probability density function (fpdf) to the pooled data and interpolated the 100-year return period, as is commonly done in the Peaks over Threshold (PoT) EVA approach (see Section 7.3). Note that with the ensemble dataset the 100-year value can be determined by interpolation, rather than extrapolation, which would be necessary in a conventional EVA. In the case of the ensemble analysis, the required probability level (100-year value) is within the sampled probabilities.

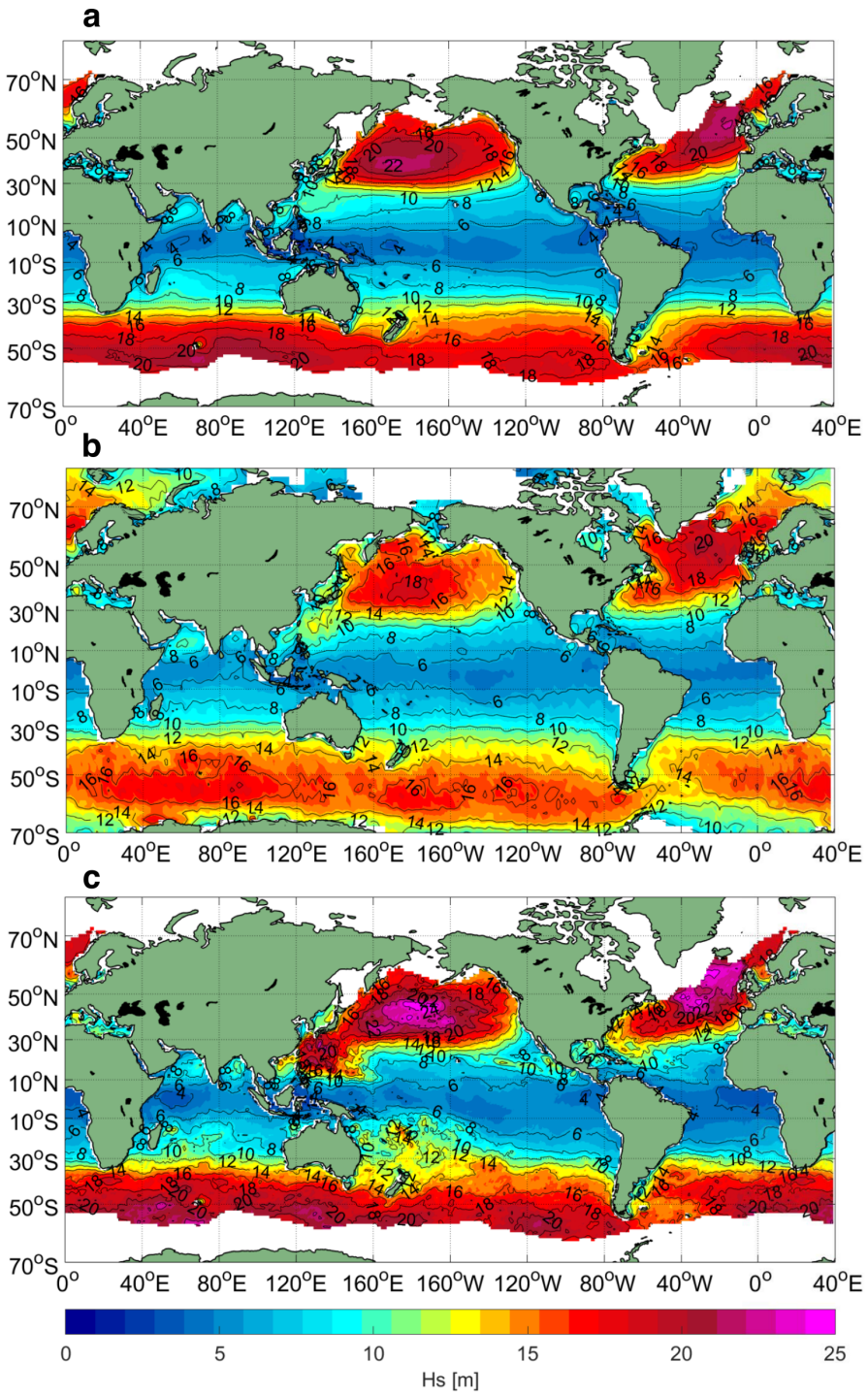


Figure 7.12: Magnitude of the 100-year return period significant wave height H_s^{100} [m] resulting from three different dataset EVAs. **a**, 1979–2005 inter-model ensemble of the bias-corrected Z standardised variable highest peaks (see Section 7.3), pooled from global wave model runs forced with 7 GCMs surface winds, **b**, 1985–2018 calibrated altimeter dataset using the Peaks over Threshold approach with exponential distribution fit, **c**, 1979–2005 Peaks over 99.6$^{\text{th}}$ percentile Threshold for the single global wave model run forced with a reference NOAA CFSR wind speed (exponential distribution fit).

The performance of the inter-model ensemble approach is demonstrated in Fig. 7.12. This figure shows the H_s^{100} estimates obtained with the model ensemble, compared with conventional PoT approaches applied to both altimeter data (34) and a single wave model forced with NOAA CFSR winds (*Saha et al.*, 2010). The comparison of the H_s^{100} values obtained from both the CFSR-forced wave model run, and the satellite dataset, shows the level of uncertainties around the extreme estimates (Fig. 7.12). The wave model forced with the CFSR high quality wind field, which has been calibrated against in-situ and remote ocean observations (*Saha et al.*, 2010), acts as a model benchmark to compare with the climate simulations. These climate simulations run just with prescribed boundary conditions, without any in-situ or remote sensing observation assimilated throughout the years. Note that this CFSR-forced reanalysis simulation is independent of the ensemble used (i.e. not a member of the ensemble). Fig. 7.13 shows the differences between the

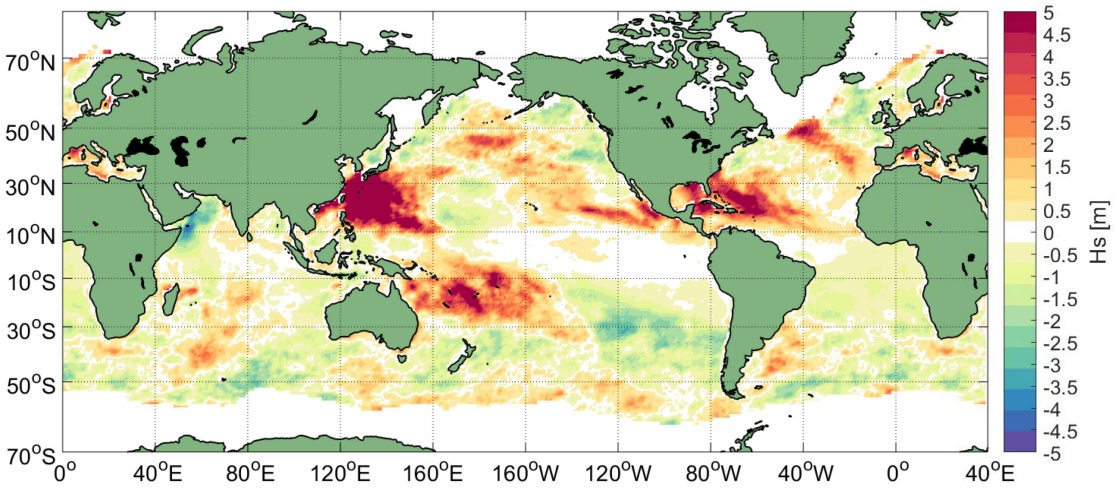


Figure 7.13: 100-year return period significant wave height differences between standardized variable corrected inter-model ensemble technique and a single CFSR-forced model analysed using a Peaks over 99.6th percentile Threshold, Extreme Value Analyses.

model ensemble and the CFSR-forced global distributions of H_s^{100}. Although there are differences of magnitude between the various approaches, the spatial distributions obtained with the inter-model ensemble of pooled extremes (Fig. 7.12a), is remarkably similar to the other datasets processed using traditional EVA approaches (Fig. 7.12b, c). The main differences are found in the tropical cyclones areas, where the GCMs are unable to adequately model extremes due to their coarse resolution (*Timmermans et al.*, 2017) and altimeter data due to their sampling density (*Takbash et al.*, 2018). The spatial distribution and magnitude of the extreme values estimated from the inter-model ensemble are also remarkably

similar to previous global significant wave height EVAs (*Caires and Sterl*, 2005; *Breivik et al.*, 2014; *Wang et al.*, 2015; *Meucci et al.*, 2018), providing further support for the multi-model approach. Confidence limits for the magnitude of the H_s^{100} values vary due to different dataset characteristics and statistical approaches. The advantage of pooling extreme wave data from an inter-model ensemble of wave models, is the reduction of the intrinsic uncertainties connected to EVA estimates (*Morim et al.*, 2018). The pooled dataset, in effect, extends the length of the available model time series, being representative of a time window longer than the return period sought (1 in 100 years in this case). This increases the level of confidence in the extreme value estimates. Fig. 7.14 shows a comparison of 95% confidence limits for H_s^{100} obtained for both the inter-model ensemble and the single CFSR-forced model, indicating a reduction in the magnitude by a factor of approximately 3 (see Section 7.3). As a result, confidence in the determination of projected changes in significant wave height extremes by the end of the 21st Century using the inter-model ensemble is also improved.

7.4.2 Extreme value projections

In this study an historical (1979–2005) inter-model ensemble was compared with an inter-model ensemble of future projections at the end of the 21st Century (2081–2100), for two different Representative Concentration Pathways: RCP4.5 intermediate emissions scenario, and RCP8.5, high emissions scenario (*Van Vuuren et al.*, 2011). We determined the global H_s^{100} differences between these two periods for both emission scenarios (Fig. 7.15), with model extreme values bias corrected using a standardized variable correction technique, which largely reduces the effect of GCM differences, caused by differences in both physics and model resolution (see Section 7.3).

Results show that the Southern Ocean is characterized by an overall increase in the H_s^{100} at the end of 21st Century for both emission RCPs (Fig. 7.15). In this region the inter-model ensemble estimates show increases of approximately 5% (RCP4.5) to 15% (RCP8.5) for H_s^{100}. The South Pacific shows decreasing wave extremes around 30° S, and a slight increase in the significant wave extremes around 10S. This indicates a possible strengthening of the trade winds by the end of the 21st Century, in agreement with recent studies (*England et al.*, 2014). Increases between 5% and 10% (both RCPs) are indicated in the Indian Ocean west of Australia and in large areas of the South Atlantic. The changes highlighted above are all statistically significant, as indicated by the hatching in Fig. 7.15 (see Section 7.3). The Northern Hemisphere shows a more irregular pattern. The models predict increases in the H_s^{100} at the high latitudes of both the North Pacific and North Atlantic. However, tests of statistical significance (see Section 7.3 and Fig. 7.15 hatching), indicate the North Atlantic increases are not statistically

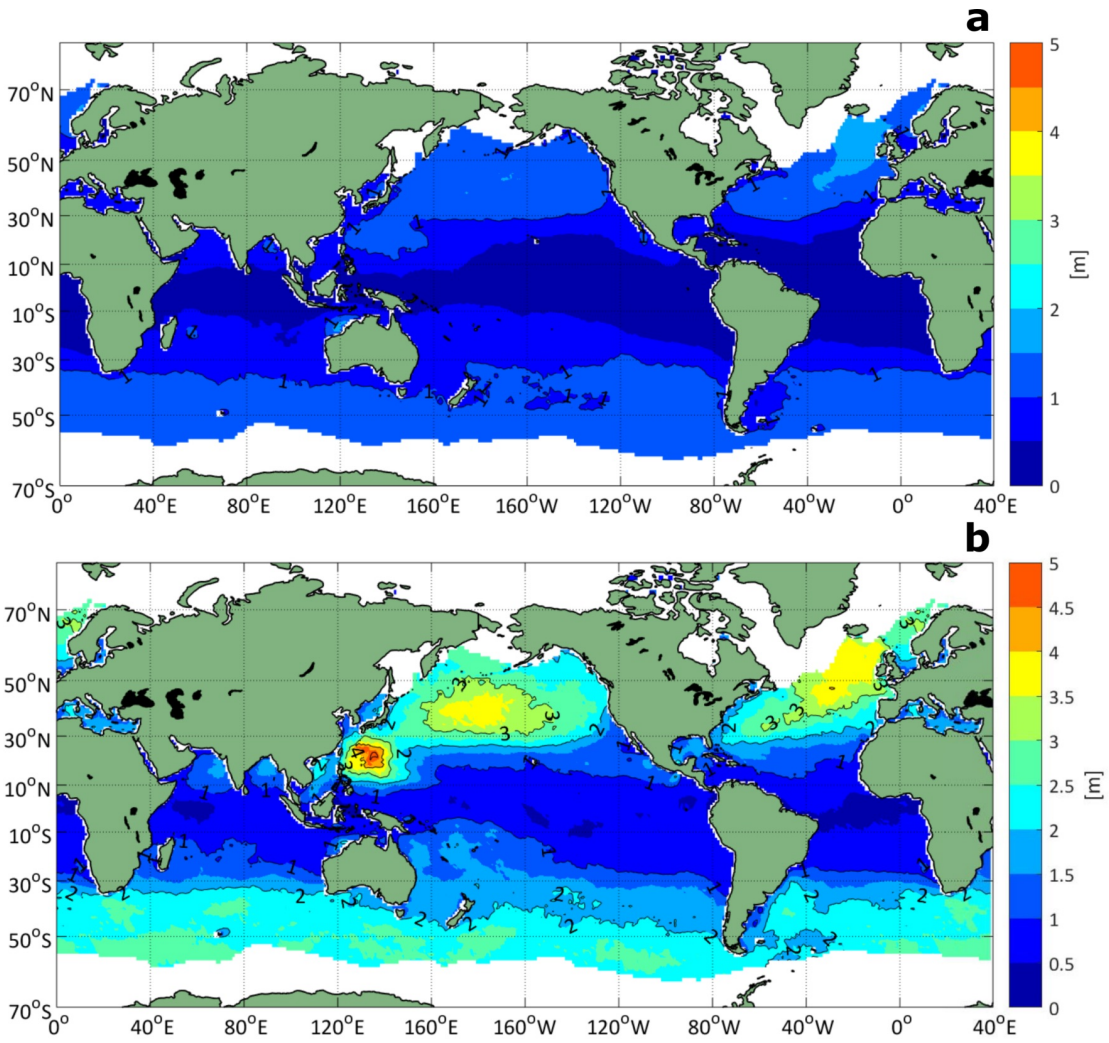


Figure 7.14: 95% Confidence intervals of the 100-year return period significant wave height, H_s^{100}, historical time slice. **a**, bootstrap estimates from the inter-model ensemble Extreme Value Analysis, **b**, Peaks over Threshold exponential fit confidence intervals of the CFSR-forced single model extreme value estimates. This figure shows the difference of the confidence interval width between the ensemble technique applied in the main manuscript, and the single model Peak Over Threshold confidence intervals. The possibility to synthetize a dataset of 189 years of data significantly increases the confidence around the H_s^{100} estimates.

significant. The projected decrease in extreme significant wave heights of -5% to -15% in the mid and low latitudes of the North Atlantic basin is, however, statistically significant.

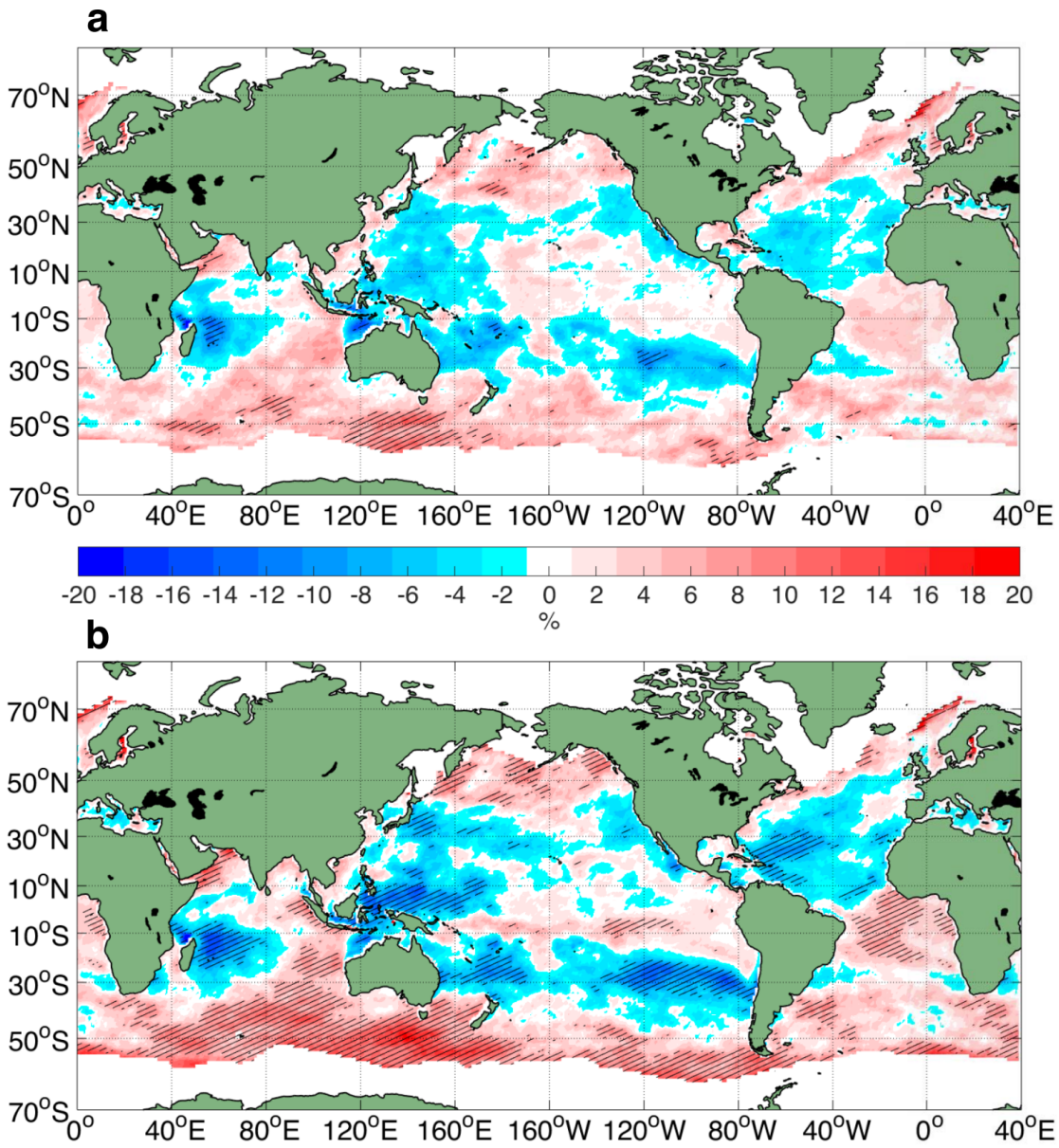


Figure 7.15: Percentage change in the 100-year return period significant wave height by the end of the 21st Century relative to the 1979–2005 period. Standardized variable analysis approach is used for the intervals 1979–2005 and 2081–2100. **a**, RCP4.5 mid emission scenario. **b**, RCP8.5 high emission scenario. The regions with statistically significant changes at 5% level (see Section 7.3, Eq.7.7) are hatched.

7.4.3 Uncertainty assessment

To account for differences in the wave climatology across the models in the ensemble, a bias correction approach using a Z standardized variable (Section 7.3) was used. To investigate the projected changes in extreme significant wave heights in more detail, we examined the empirical probability distribution function (epdf) of this Z standardized variable (Section 7.3) at specific locations (Fig. 7.10). This Z ensemble strongly reduces the uncertainties associated with the H_s^{100} estimates both from the historical 1979–2005 dataset and the two future projection scenarios. Fig. 7.10a, d, g show that the 95% confidence intervals of the H_s^{100} estimates at three selected locations from: the ensemble of models, each individual model and the average across the models. The confidence limits represented by the error bars in the graphs, are considerably smaller for the ensemble approach than for any of the single models when analyzed using a common PoT analysis. Fig. 7.10 b, c; e, f; h, i show epdfs from 5°x5° regions around these three specific locations, the averaging over the region being used to further reduce statistical variability in the epdfs. The Southern Ocean region around location (120° E, 50° S) is an area where the extreme significant wave height is projected to increase. The epdfs (Fig. 7.10b, c) show that this occurs due to an increase in both the frequency and the magnitude of extreme events for both RCP4.5 and RCP8.5. The changes in magnitude are depicted by a shift towards the right of the future scenarios epdfs if compared to the histogram of the historical dataset. The changes in the Southern Ocean event frequency are shown by a translation of the future scenario epdfs along the Y-axis (Fig. 7.10b). The Q-Q plot (Fig. 7.10c) analyzes the details of these changes. Here, the changes in frequency are shown by the departure of the future scenarios red and orange quantiles from the 1:1 line in the region of plot represented by the bulk of the dataset.

The error bars show that the projected changes by the end of 21st Century are statistically significant for this location (i.e. the error bars do not overlap) (Fig. 7.10a). The North Pacific region around location (190° E, 50° N) is also an area where the extreme significant wave height is projected to increase. In this case the comparison with the historical dataset (Fig. 7.10e, f) shows that the epdfs of the two projected emission scenarios are similar in terms of frequency, but there is an increase in the magnitude of extreme waves by the end of the century. The confidence limits (Fig. 7.10d) indicate that this change is statistically significant for both future scenarios. The North Atlantic location (320° E, 40° N) is representative of an area where the extreme significant wave height is projected to decrease. The 5°x5° region epdfs (Fig. 7.10h, i) show that this decrease is a result of a decrease in both the frequency and magnitude of the extremes, for both RCP4.5 and RCP8.5 emission scenarios. The confidence intervals, however, indicate the projected changes are not statistically significant (Fig. 7.10g).

7.4.4 Change in frequency of extremes

An additional investigation of changes in the frequency of extreme events was undertaken by determining, for each single wave model, the number of extreme events above defined thresholds for both the historical (1979–2005) and future projection (2081–2100) datasets. For each model the threshold was set as a percentile of the historical dataset. The number of extremes above this value, for both the historical and future projection datasets of each model, were obtained, and then the 7-model difference in the number of extremes (count/years) between historical and future projection was computed (see Section 7.3). Initially we tested a 99.6^{th} percentile threshold, that approximately corresponds to the percentile, over the total number of model outputs, of the 1000 storm events pooled for the EVA performed at each location of the globe. To confirm the consistency of the results we also tested a 90th percentile threshold.

Fig. 7.16a shows the change in the number of events above the 90^{th} percentile level and Fig. 7.16b shows the case for the 99.6^{th} percentile threshold. For the Southern Ocean the increase in the H_{s}^{100} is associated with an increase in the number of extreme storms. This is a consistent result for both thresholds (Fig. 7.16). Thus, the change in frequency of these events impacts the return period estimates. In the North Pacific we found an increase in significant wave height at the end of the 21^{st} Century (Fig. 7.15), but there is no significant increase in the frequency of occurrence of these events (Fig. 7.16). Here, it is an increase in the magnitude of the events which is responsible for the changes in the 100-year return period significant wave height. The projected decrease in values in the North Atlantic is associated with a decrease in the frequency of the extremes (approx. 1 event less per year for the 99.6^{th} percentile threshold).

7.4.5 Tropical Cyclone generated wind-wave extremes

Tropical Cyclone (TC) generated wind-wave extremes deserve separate discussion as the coarse resolution of the GCMs and the physics schemes used may limit the representation of TC extreme winds, and consequently the modelled wind-wave extremes. The present 7-model ensemble dataset performance in reproducing TC wind-waves was analyzed and validated against observations by *Shimura et al.* (2017) in the Western North Pacific region. The analysis shows that 3 out of 7 models can represent at least 80% of the TCs frequency, as well as winds over 30 m/s. These models are the BCC-CSM1.1, the MIROC5, and MRI-CGCM3. The ability of these models to depict TC wind-wave extremes is also shown in Fig. 7.17 where, for selected NDBC buoy locations, we compare the bias corrected 7-model ensemble wave heights with buoy and satellite altimeter wave height measurements. Additionally, in Fig. 7.5 an EVA is performed for each of the wave

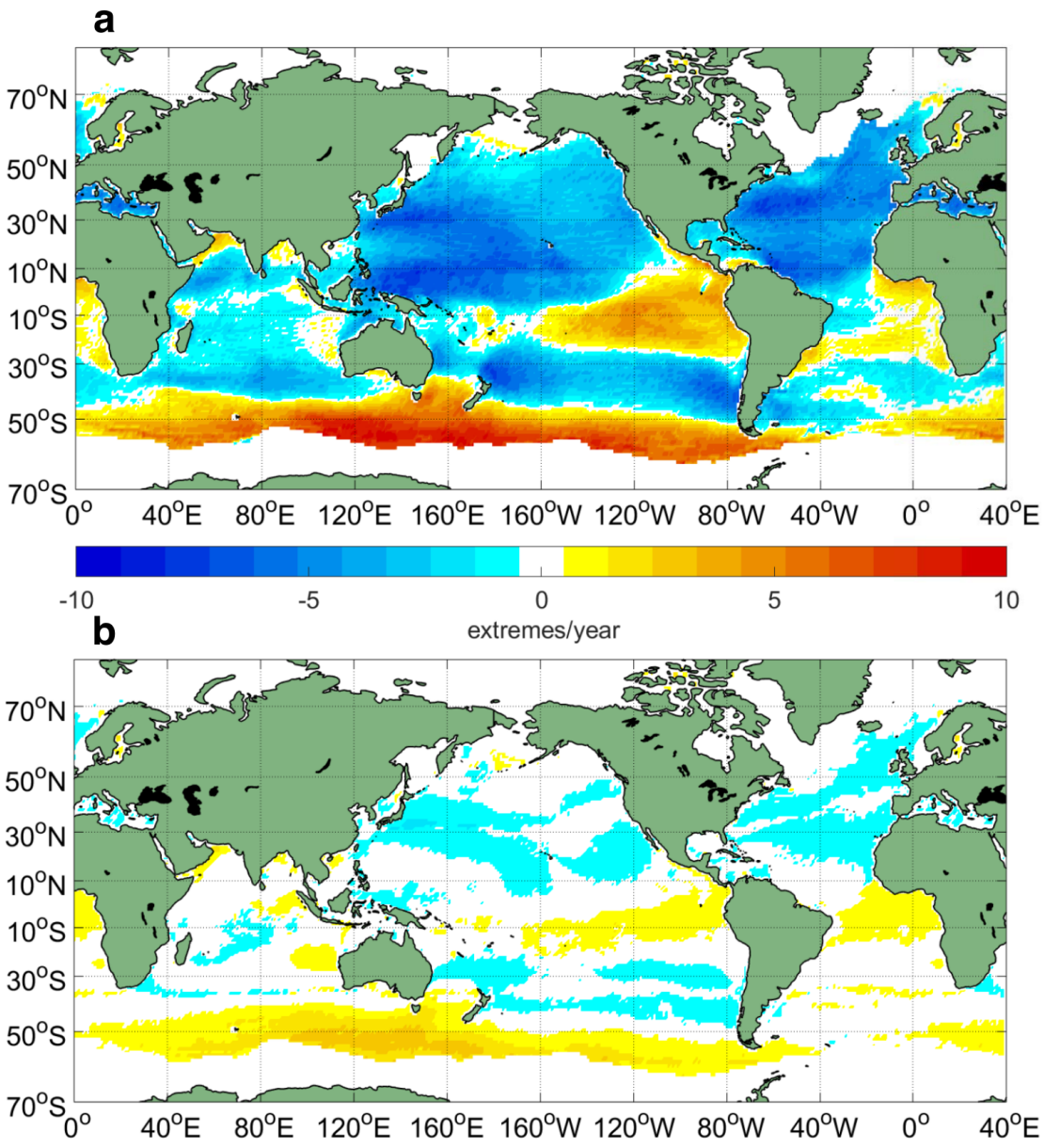


Figure 7.16: Changes in extreme event frequency. **a**, changes in the number of extreme events per year over the historical 90th percentile threshold between the historical dataset 1979–2005 and future projection 2081–2100 for RCP8.5. **b**, as for **a** but for the historical 99.6th percentile threshold. The similar spatial variability demonstrates consistency of the results for different selected thresholds.

model runs for the 1979-2005 historical period to show the differences between each model EVA performance in TC regions. Thus, the 7-model ensemble has

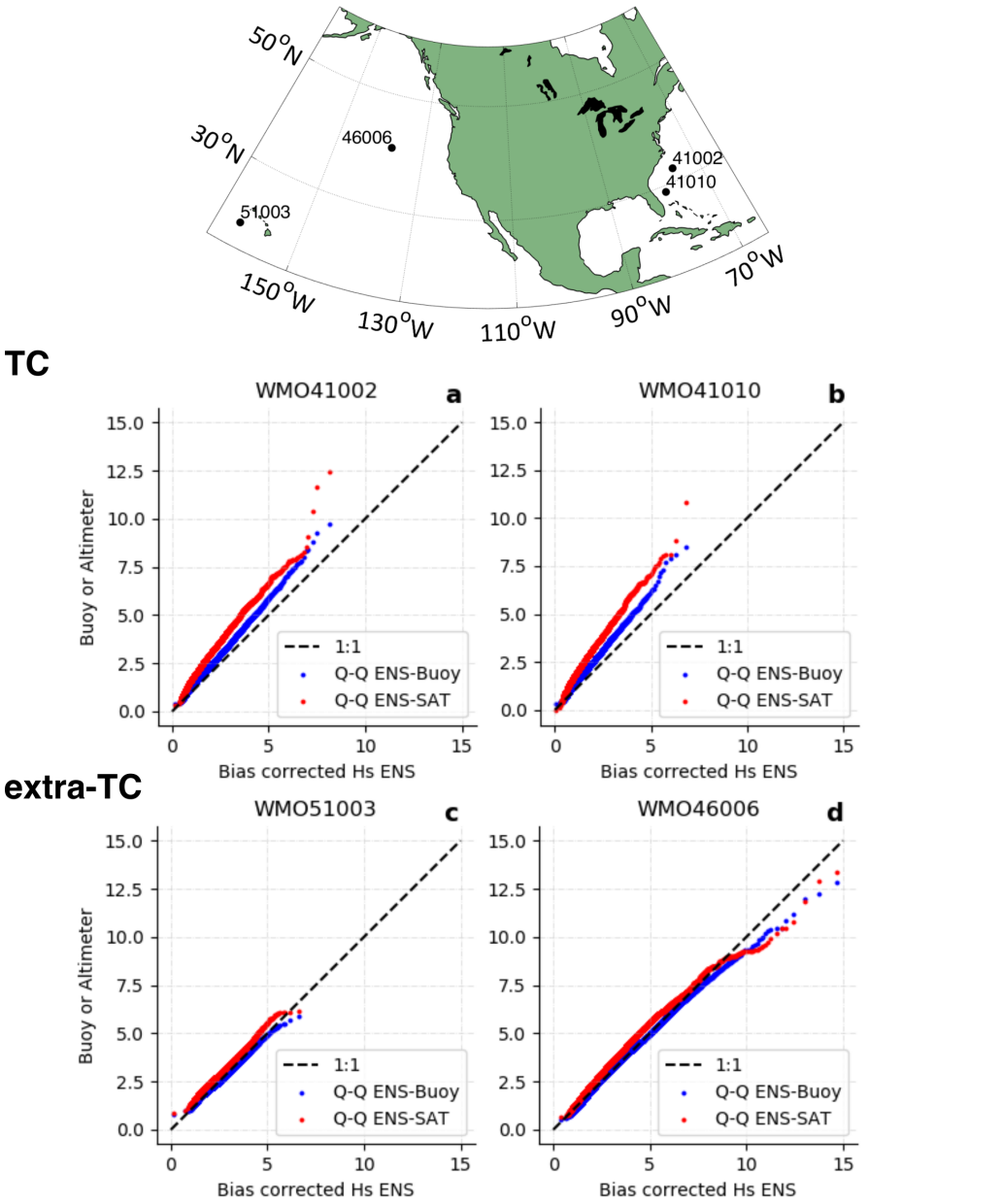


Figure 7.17: Comparison of the bias corrected Z-ensemble wave height with NDBC buoy and satellite altimeter measurements. Quantile-Quantile (Q-Q) plots at four selected locations (positions marked on top map). **a**, **b**, TC locations where the ensemble underestimates extremes. **c**, **d**, extra-TC regions where the ensemble performs well against buoy and satellite measurements, with slightly higher extremes.

only limited accuracy in representing such events, and, as can be seen in Fig ..., underestimates extremes in tropical cyclone regions. However, it is useful to investigate the magnitude and frequency changes (Fig. 7.15 and Fig. 7.16) in major TC regions (the Western and Eastern North Pacific, the North Indian, South Indian, and South Pacific, the Caribbean/Gulf of Mexico, and the open North Atlantic - https://www.ncei.noaa.gov/news/inventory-tropical-cyclone-tracks), and how the GCM projections perform in these regions. Fig. 7.15 shows statistically significant decreases in the magnitude of H_s^{100} in TC regions, except for the North West Pacific. Fig. 7.16 shows that the frequency of TC extreme wind-wave events decreases globally by 2100. According to the recently published IPCC Special Report on the Ocean and Cryosphere in a Changing Climate (*Collins et al.*, 2019), and reference reviews on global future projections in TC events (*Walsh et al.*, 2016; *Knutson et al.*, 2019), there is medium confidence on a general future projected increase in Category 4 and 5 TC events. If this is the case, H_s^{100} future projected values should also increase as a result of the shift in magnitude of TC events. This differs from the results obtained with the present 7-model ensemble EVA (Fig. 7.15). However, there is also emerging evidence (low confidence) on a reduction in the frequency of TCs (*Knutson et al.*, 2019; *Collins et al.*, 2019), which is in agreement with Fig. 7.16 results. It should be noted that both the intensity and frequency of the 1000 pooled peaks will impact the EVA. Thus, the reduction in frequency would partially counteract the increase in TC intensity, thus reducing the H_s^{100} future projected estimates.

As noted above, global wind-wave extreme projections in TC regions are affected by CMIP5 GCM resolution, and model physics that limit the confidence around the estimates of TC wind-wave events (*Walsh et al.*, 2016). However, the present EVA approach can infer some information on TCs, derived from a limited number of models contributing to the ensemble. In the future, this approach could take advantage of CMIP6 higher resolution GCMs, with improved physics, to force a high resolution global wave model (horizontal resolution of at least $0.25°$(*Timmermans et al.*, 2017)), to estimate TC generated wind-wave extreme future changes. In this way, the ensemble approach performance would be significantly improved, as a much larger amount of information on TC wind-wave events would be embedded in the pooled highest peaks. In that case, we could estimate changes in TC generated wind-wave extremes, as opposed to non-TC generated waves, arguably still dominant in the present ensemble dataset (*Shimura et al.*, 2017). At the present level of accuracy, special care must be taken in interpreting the changes in H_s^{100} in TC regions.

7.4.6 Change along global coastlines

A potential consequence of the projected changes in extreme wave conditions will be enhanced shoreline erosion and possibly coastal flooding in some regions.

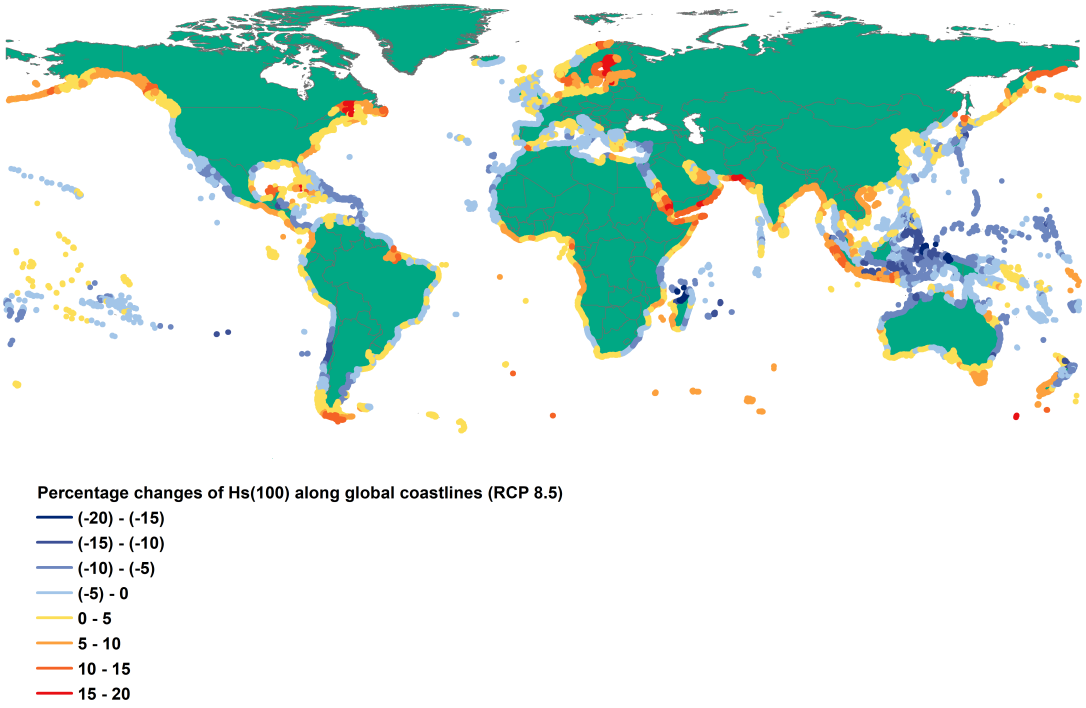


Figure 7.18: Percentage change in 100-year extreme value significant wave height along the global coastline between the historical dataset 1979–2005 and future projection 2081–2100 for RCP8.5.

Fig. 7.18 shows the projected changes in H_s^{100} along the worlds coastline in 2100 for RCP8.5. This was determined by assigning the change in H_s^{100} at the closest model point to the coastal segment. Clearly, this assessment ignores potential nearshore refraction and shoaling at this global scale.

Table 7.2 shows the lengths and percentage of global coastline changes in H_s^{100} in 5% increments. A total of 59% of the worlds coastlines are projected to experience an increase in extreme wave conditions for the RCP8.5 (Table 7.2). Consistent with Fig. 7.15b, high latitude regions of both hemispheres are projected to have an increase in extreme significant wave height, whilst lower latitudes generally see a decrease. It should be noted, however, that due to model resolution, tropical cyclone generated waves are poorly resolved. Therefore, there is less confidence in the projected changes in H_s^{100} in these tropical regions (30° S to 10° S, 10° N to 30° N) (see Section 7.3). Although Fig. 7.15b shows increases in H_s^{100} in the Southern

Table 7.2: The percentage change in 100-year significant wave height along coastline by 2100 for both RCP4.5 and 8.5.

	RCP4.5		RCP8.5	
H_s^{100}	Coastline		Coastline	
(% change)	(Km)	(%)	(Km)	(%)
-20% to -15%	9,643	0.89	7,399	0.69
-15% to -10%	13,130	1.22	25,281	2.34
-10% to -5%	69,208	6.42	120,625	11.18
-5% to 0%	277,810	25.76	285,227	26.45
0% to 5%	499,537	46.32	365,741	33.91
5% to 10%	168,420	15.62	18,2163	16.89
10% to 15%	33,053	3.06	68,087	6.31
15% to 20%	7,737	0.72	24,015	2.23

Ocean up to 20%, these large increases are generally at latitudes further south than the majority of continental coastlines. The southern tip of South America is projected to experience an increase of approximately 20%, with the west coast of New Zealand and Tasmania experiencing an increase of 10 to 15%. Increases of 10 to 15% are also projected for the coasts of the North Pacific (Canada and Kamchatka Peninsula). As shown in Fig. 7.15 (see Section 7.3) changes in the magnitude of H_s^{100} less than 5% are generally not statistically significant.

7.5 Discussion and Conclusions

The ensemble approach adopted here has the major advantage that the resulting values of H_s^{100} have considerably smaller confidence intervals than if estimates from a single model were used. This is an advantage when investigating changes in values of H_s^{100}. The reduction in the magnitude of the confidence intervals can be seen at specific locations in (Fig. 7.10a, d, g), and globally in Fig. 7.14 and 7.9. This allows the determination of projected changes in extremes with much greater statistical confidence. As a result, the present analysis has been able to estimate statistically significant changes in extreme significant wave height on a global basis. Although projected changes in extreme events appear robust, intra-model uncertainties, as well as uncertainties that arise from atmospheric downscaling and wind-wave modelling are not addressed by this analysis (*Morim et al.*, 2019). Furthermore, some regions of the global oceans affected by local climate variability, such as tropical cyclones, present higher uncertainty due to model limitations in representing these natural phenomena at the current model resolutions (*Shimura et al.*, 2015; *Walsh et al.*, 2016; *Appendini et al.*, 2017; *Timmermans et al.*, 2017;

Shimura et al., 2017; *Collins et al.*, 2019; *Knutson et al.*, 2019). Whether tropical cyclones are increasing in frequency or magnitude remains an open question (*Landsea et al.*, 2006; *Walsh et al.*, 2016; *Patricola and Wehner*, 2018; *Collins et al.*, 2019; *Knutson et al.*, 2019). However, increasing resolution and improved physics of next generation models will further address remaining sources of uncertainties, with increasingly accurate estimates of the changes in future projected extreme significant wave heights. Furthermore, we advocate for a continuous archiving of data from GCMs, to ensure long climate record. These GCMs are already run over centennial time-scales, but high temporal data is archived only for limited time slices.

Chapter 8

Conclusions and Future Research

Marine structures and coastal defences are impacted by the action of waves. Extreme waves are particularly dangerous as they pose the highest risk to human lives and assets. Throughout this thesis, we discussed the importance of accurately estimating the design sea state, that is the maximum significant wave height which can be expected over a N year period. We particularly focused on the 100-year return period, as this is the most common design sea state sought for many engineering applications. We investigated the challenges related to reliable estimation of the 100-year return period, and analyzed a series of works on the topic.

The thesis then developed, and tested, a novel Extreme Value Analysis (EVA) approach which improves the confidence around 100-year return period estimates of ocean surface wind speed and wave height. This approach proved to be fundamental in reducing the statistical uncertainties around design sea state estimates. Comparison with other datasets and previous EVA showed the validity of such an approach.

The advantage of this approach is that we were able to synthesise an equivalent time series far longer than the ones commonly available from observations and model output. Furthermore, comparing our results with different models we were able to identify the best available model datasets to estimate long-term extremes.

This initial work considered the wave climate as stationary. However, we know that this is not the case as shown by studies on satellite measurements (*Young et al.*, 2011; *Young and Ribal*, 2019), in-situ observations (*Allan and Komar*, 2000; *Gower*, 2002; *Gulev and Grigorieva*, 2004; *Ruggiero et al.*, 2010; *Hemer et al.*, 2010), and model hindcasts and reanalyses (*Aarnes et al.*, 2015; *Wohland et al.*, 2019).

Past climate trends are still under discussion due to many limitations and inconsistency found in the trends (*Bengtsson et al.*, 2004; *Weisse*, 2010; *Gower*, 2002; *Hemer et al.*, 2010). This led to investigation of such changes and we found inconsistency in the trends of wind and wave mean climate as determined from 20^{th}

Century climate models and reanalyses. However, we identified agreement between datasets on a possible signal in Southern Hemisphere wind speed changes over the last part of the 20th Century that could trigger changes in the wave climate.

As the mean wave climate is changing, we applied the ensemble approach methodology, developed in the first part of the thesis, to an inter-model ensemble of global WWIII wave model runs, forced with seven different GCM surface wind fields. The comparison between the 1979–2005 and 2081–2100 periods, found, for the first time, statistically significant projected changes in wind-wave extremes by the end of the 21st Century, for the high emission scenario RCP8.5.

This is an important finding of the thesis, as for the first time we were able to reduce the confidence interval around future extreme estimates, and provide insight into how the extreme wave climate is projected to change.

The main limitation of the three projects presented is the interpretation of Tropical Cyclone generated waves. We acknowledged, throughout the thesis, that particular care is needed in the interpretation of the 100-year return period estimates in TC regions. This is due to the coarse resolution of the global models, that limit a correct representation of these localized events.

However, this work has strong potential for the estimation of TC extreme wind-waves. Models with resolution higher than 0.25° can be used (*Timmermans et al.*, 2017), to robustly estimate TC wind-wave 100-year return period values.

In conclusion, the present thesis first developed the ensemble EVA approach (*Breivik et al.*, 2013, 2014; *Meucci et al.*, 2018) for a stationary case, and then applied such an approach to find future wind-wave extreme estimates. The aim is to contribute to the IPCC sixth Assessment Report (AR6), and improve the understanding of extreme wind-wave events in a changing climate. This will support resilient strategies for the design of marine offshore structures and coastal defences.

8.1 Future work

As pointed out by *Morim et al.* (2019), the natural way forward in the understanding of the wind-wave climate, and thus its extremes, is the use of ensemble products, as the information that we can extrapolate from such products may reduce, or explain, some of the numerous sources of uncertainty in the determination of wind-wave climate (*Morim et al.*, 2018). Thus, following are a number of possible extensions to the present thesis, that we believe could significantly contribute to research in this field.

Interpretation and improvement of ensemble reanalysis products

An interesting study would be to elaborate on aspects involved in the sources of uncertainty connected to the predictability of extreme wind-wave climate, using ensemble spread products, now available from novel reanalyses such as the ECMWF ERA5. This product considers the uncertainties connected to the atmospheric flow. No ocean wave ensemble spread is produced by ECMWF, but it is of scientific value to investigate how the wave model RMSE compares to the ensemble spread of the atmospheric product. A possible outcome of the study could be to suggest the production of an ocean ensemble spread as is done for the atmospheric model.

Ensemble of TC generated wind-waves

To improve the estimation of the 100-year wind-wave return period in TC regions, the present thesis ensemble EVA approach could be applied to a series of high-resolution model runs of virtually simulated TCs. Such information, with the available high-resolution, would significantly improve the extreme value estimates, giving insights into these crucial phenomena. This would then help practitioners in mapping the extreme climate of TC regions.

The same approach can then be applied for future extremes, using different RCPs from a series of high resolution GCMs (*Shimura et al.*, 2015).

CMIP6 global climate models

A logical extension to the future extremes work presented in chapter 7 would be to use the recently developed CMIP6 models. To do this we would need to run a global wave model with the new surface wind products, and evaluate the performance for mean wave climate. In this way, we could extrapolate information on the extremes, and apply the inter-model ensemble approach as in chapter 7.

Acronyms

4D-Var Four-Dimensional Variational scheme

AMM Annual Maxima Method

BoM Australian Bureau of Meteorology

CFSR Climate Forecast System Reanalysis

CMIP Coupled Model Intercomparison Project

COWCLIP Coordinated Ocean Wave Climate Project

Cy41r1 Operational forecast - Cycle 41 round 1

DIA Discrete Interaction Approximation

ECMWF European Centre for Medium-Range Weather Forecasts

EDA Ensemble Data Assimilation

ENS Ensemble

ERA-15 ECMWF European ReAnalysis 15

ERA-40 ECMWF European ReAnalysis 40

ERA-I ECMWF European ReAnalysis Interim

ERA5 ECMWF European ReAnalysis 5

EVA Extreme Value Analysis

EVT Extreme Value Theory

GCM Global Circulation/Climate Model

GEV Generalized Extreme Value

GMAO NASA Global Modelling and Assimilation Centre

GP Generalized Pareto

i.i.d. independent and identically distributed

ICOADS International Comprehensive Ocean-Atmosphere Data Set

IDM Initial Distribution Method

IFS Integrated Forecasting System

IMOS Integrated Marine Observing System

IPCC Intergovernmental Panel on Climate Change

JCOMM Joint technical Commission for Oceanography and Marine Meteorology

JMA Japanese Meteorological Agency

JONSWAP Joint North Sea Wave Project

NCEP National Centers for Environmental Prediction

NDBC National Data Buoys Center (USA)

NOAA American National Oceanic and Atmospheric Administration

NWP Numerical Weather Prediction

POT Peaks Over Threshold

RCP Representative Concentration Pathway

SAR Synthetic Aperture Radar

SRA Scanning Radar Altimeter

SRES Special Report on Emission Scenarios

SREX Special Report on Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaptation

SSP Shared Socio-Economic Pathway(s)

SWH Significant Wave Height

VOS Voluntary Observing Ships

WAM WAve Modeling

WMO World Meteorological Organization

WMO-IOC World Meteorological Organization Intergovernmental Oceanographic Commission

WWIII WaveWatch-III

Bibliography

Aarnes, O. J., Abdalla, S., Bidlot, J.-R., and Breivik, Ø. (2015). Marine wind and wave height trends at different ERA-Interim forecast ranges. *Journal of Climate*, 28(2):819–837.

Aarnes, O. J., Breivik, Ø., and Reistad, M. (2012). Wave extremes in the northeast Atlantic. *Journal of Climate*, 25(5):1529–1543.

Aarnes, O. J., Reistad, M., Breivik, Ø., Bitner-Gregersen, E., Ingolf Eide, L., Gramstad, O., Magnusson, A. K., Natvig, B., and Vanem, E. (2017). Projected changes in significant wave height toward the end of the 21st century: Northeast Atlantic. *Journal of Geophysical Research: Oceans*, 122(4):3394–3403.

Allan, J. and Komar, P. (2000). Are ocean wave heights increasing in the eastern north pacific? *Eos, Transactions American Geophysical Union*, 81(47):561–567.

Alves, J. H. G. and Young, I. R. (2003). On estimating extreme wave heights using combined Geosat, Topex/Poseidon and ERS-1 altimeter data. *Applied Ocean Research*, 25(4):167–186.

Appendini, C. M., Pedrozo-Acuña, A., Meza-Padilla, R., Torres-Freyermuth, A., Cerezo-Mota, R., López-González, J., and Ruiz-Salcines, P. (2017). On the role of climate change on wind waves generated by tropical cyclones in the gulf of mexico. *Coastal Engineering Journal*, 59(02):1740001.

Ardhuin, F., Rogers, E., Babanin, A. V., Filipot, J.-F., Magne, R., Roland, A., Van Der Westhuysen, A., Queffeulou, P., Lefevre, J.-M., Aouf, L., et al. (2010). Semiempirical dissipation source functions for ocean waves. part i: Definition, calibration, and validation. *Journal of Physical Oceanography*, 40(9):1917–1941.

Ardhuin, F., Stopa, J. E., Chapron, B., Collard, F., Husson, R., Jensen, R. E., Johannessen, J., Mouche, A., Passaro, M., Quartly, G. D., and et al. (2019). Observing sea states. *Frontiers in Marine Science*, 6.

Babanin, A. V., Onorato, M., and Qiao, F. (2012). Surface waves and wave-coupled effects in lower atmosphere and upper ocean. *Journal of Geophysical Research: Oceans*, 117(C11).

Babanin, A. V., Rogers, W. E., Camargo, R. D., Doble, M., Durrant, T., Filchuk, K., Ewans, K., Hemer, M., Janssen, T., Kelly-Gerreyn, B., and et al. (2019). Waves and swells in high wind and extreme fetches, measurements in the southern ocean. *Frontiers in Marine Science*, 6.

Babanin, A. V., Tsagareli, K. N., Young, I., and Walker, D. J. (2010). Numerical investigation of spectral evolution of wind waves. part ii: Dissipation term and evolution tests. *Journal of Physical Oceanography*, 40(4):667–683.

Babanin, A. V., van der Weshuijsen, A., Chalikov, D., and Rogers, W. E. (2017). Advanced wave modeling, including wave-current interaction. *Journal of Marine Research*, 75(3):239–262.

Bauer, P., Thorpe, A., and Brunet, G. (2015). The quiet revolution of numerical weather prediction. *Nature*, 525(7567):47–55.

Befort, D. J., Wild, S., Kruschke, T., Ulbrich, U., and Leckebusch, G. C. (2016). Different long-term trends of extra-tropical cyclones and windstorms in ERA-20C and NOAA-20CR reanalyses. *Atmospheric Science Letters*, 17(11):586–595.

Bender, L., Guinasso Jr, N., Walpert, J., and Howden, S. D. (2010). A comparison of methods for determining significant wave heights applied to a 3-m discus buoy during hurricane Katrina. *Journal of atmospheric and oceanic technology*, 27(6):1012–1028.

Bengtsson, L., Hagemann, S., and Hodges, K. I. (2004). Can climate trends be calculated from reanalysis data? *Journal of Geophysical Research: Atmospheres*, 109(D11).

Bertin, X., Prouteau, E., and Letetrel, C. (2013). A significant increase in wave height in the North Atlantic Ocean over the 20th century. *Global and Planetary Change*, 106:77–83.

Bidlot, J.-R., Holmes, D. J., Wittmann, P. A., Lalbeharry, R., and Chen, H. S. (2002). Intercomparison of the performance of operational ocean wave forecasting systems with buoy data. *Weather and Forecasting*, 17(2):287–310.

Bitner-Gregersen, E. M., Vanem, E., Gramstad, O., Hørte, T., Aarnes, O. J., Reistad, M., Breivik, Ø., Magnusson, A. K., and Natvig, B. (2018). Climate change and safe design of ship structures. *Ocean Engineering*, 149:226–237.

Breivik, Ø. and Aarnes, O. J. (2017). Efficient bootstrap estimates for tail statistics. *Natural Hazards and Earth System Sciences*, 17(3):357.

Breivik, Ø., Aarnes, O. J., Abdalla, S., Bidlot, J.-R., and Janssen, P. A. (2014). Wind and wave extremes over the world oceans from very large ensembles. *Geophysical Research Letters*, 41(14):5122–5131.

Breivik, Ø., Aarnes, O. J., Bidlot, J.-R., Carrasco, A., and Saetra, Ø. (2013). Wave extremes in the northeast Atlantic from ensemble forecasts. *Journal of Climate*, 26(19):7525–7540.

Breivik, Ø., Carrasco, A., Staneva, J., Behrens, A., Semedo, A., Bidlot, J.-R., and Aarnes, O. J. (2019). Global Stokes Drift Climate under the RCP8.5 Scenario. *J Climate*, 32(6):1677–1691.

Buizza, R. (1997). Potential forecast skill of ensemble prediction and spread and skill distributions of the ecmwf ensemble prediction system. *Monthly Weather Review*, 125(1):99–119.

Buizza, R., Bidlot, J.-R., Wedi, N., Fuentes, M., Hamrud, M., Holt, G., and Vitart, F. (2007). The new ECMWF VAREPS (variable resolution ensemble prediction system). *Quarterly Journal of the Royal Meteorological Society*, 133(624):681–695.

Buizza, R., Brönnimann, S., Haimberger, L., Laloyaux, P., Martin, M. J., Fuentes, M., Alonso-Balmaseda, M., Becker, A., Blaschek, M., Dahlgren, P., et al. (2018). The EU-FP7 ERA-CLIM2 project contribution to advancing science and production of Earth system climate reanalyses. *Bulletin of the American Meteorological Society*, 99(5):1003–1014.

Buizza, R. and Palmer, T. (1995). The singular-vector structure of the atmospheric global circulation. *Journal of the Atmospheric Sciences*, 52(9):1434–1456.

Caires, S. (2011). Extreme value analysis: wave data. *Joint WMO/IOC Technical Commission for Oceanography and Marine Meteorology (JCOMM) Technical Report*, (57).

Caires, S. (2016). A comparative simulation study of the annual maxima and the peaks-over-threshold methods. *Journal of Offshore Mechanics and Arctic Engineering*, 138.

Caires, S. and Sterl, A. (2005). 100-year return value estimates for ocean wind speed and significant wave height from the ERA-40 data. *Journal of Climate*, 18(7):1032–1048.

Caires, S., Sterl, A., Bidlot, J., Graham, N., and Swail, V. (2004). Intercomparison of different wind–wave reanalyses. *Journal of Climate*, 17(10):1893–1913.

Cardone, V. J., Greenwood, J. G., and Cane, M. A. (1990). On trends in historical marine wind data. *Journal of Climate*, 3(1):113–127.

Castillo, E. (2012). *Extreme value theory in engineering*. Elsevier.

Cavaleri, L. (2009). Wave modeling - Missing the peaks. *Journal of Physical Oceanography*, 39(11):2757–2778.

Cavaleri, L., Abdalla, S., Benetazzo, A., Bertotti, L., Bidlot, J.-R., Breivik, Ø., Carniel, S., Jensen, R., Portilla-Yandun, J., Rogers, W., et al. (2018). Wave modelling in coastal and inner seas. *Progress in oceanography*, 167:164–233.

Cavaleri, L., Alves, J.-H., Ardhuin, F., Babanin, A., Banner, M., Belibassakis, K., Benoit, M., Donelan, M., Groeneweg, J., Herbers, T., et al. (2007). Wave modelling–the state of the art. *Progress in oceanography*, 75(4):603–674.

Cavaleri, L. and Bertotti, L. (2006). The improvement of modelled wind and wave fields with increasing resolution. *Ocean engineering*, 33(5-6):553–565.

Cavaleri, L., Fox-Kemper, B., and Hemer, M. (2012). Wind waves in the coupled climate system. *Bulletin of the American Meteorological Society*, 93(11):1651–1661.

Chawla, A., Spindler, D. M., and Tolman, H. L. (2013). Validation of a thirty year wave hindcast using the climate forecast system reanalysis winds. *Ocean Modelling*, 70:189–206.

Coles, S., Bawa, J., Trenner, L., and Dorazio, P. (2001). *An Introduction to Statistical Modelling of Extreme Value Theory*, volume 208. Springer, London.

Collins, M., Sutherland, M., Bouwer, L., Cheong, S.-M., Frlicher, T., Jacot Des Combes, H., Koll Roxy, M., Losada, I., Dwi Susanto, R., McInnes, K., Ratter, B., Rivera-Arriaga, E., Susanto, R., Swingedouw, D., and Tibig, L. (In press 2019). *Extremes, Abrupt Changes, and Managing Risk - Chap. 6 in IPCC Special Report on the Ocean and Cryosphere in a Changing Climate*.

Compo, G. P., Whitaker, J. S., and Sardeshmukh, P. D. (2006). Feasibility of a 100-year reanalysis using only surface pressure data. *Bulletin of the American Meteorological Society*, 87(2):175–190.

Compo, G. P., Whitaker, J. S., Sardeshmukh, P. D., Matsui, N., Allan, R. J., Yin, X., Gleason, B. E., Vose, R. S., Rutledge, G., Bessemoulin, P., et al. (2011). The twentieth century reanalysis project. *Quarterly Journal of the Royal Meteorological Society*, 137(654):1–28.

Cooley, D. (2009). Extreme value analysis and the study of climate change. *Climatic change*, 97(1-2):77.

Cooper, C. K. and Forristall, G. Z. (1997). The use of satellite altimeter data to estimate the extreme wave climate. *Journal of Atmospheric and Oceanic Technology*, 14(2):254–266.

Cram, T. A., Compo, G. P., Yin, X., Allan, R. J., McColl, C., Vose, R. S., Whitaker, J. S., Matsui, N., Ashcroft, L., Auchmann, R., et al. (2015). The international surface pressure databank version 2. *Geoscience Data Journal*, 2(1):31–46.

de Boisséson, E., Balmaseda, M. A., Abdalla, S., Källén, E., and Janssen, P. A. E. M. (2014). How robust is the recent strengthening of the Tropical Pacific trade winds? *Geophys Res Lett*, pages 4398–4405.

de Winter, R. C., Sterl, A., de Vries, J. W., Weber, S. L., and Ruessink, G. (2012). The effect of climate change on extreme waves in front of the dutch coast. *Ocean dynamics*, 62(8):1139–1152.

Dee, D., Balmaseda, M., Balsamo, G., Engelen, R., Simmons, A., and Thépaut, J.-N. (2014). Toward a consistent reanalysis of the climate system. *Bulletin of the American Meteorological Society*, 95(8):1235–1248.

Dee, D., Uppala, S., Simmons, A., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M., Balsamo, G., Bauer, P., et al. (2011). The ERA-Interim reanalysis: Configuration and performance of the data assimilation system. *Quarterly Journal of the royal meteorological society*, 137(656):553–597.

Dey, D. K. and Yan, J. (2016). *Extreme value modeling and risk analysis: methods and applications*. CRC Press.

Dietz, E. J. and Killeen, T. J. (1981). A nonparametric multivariate test for monotone trend with pharmaceutical applications. *Journal of the American Statistical Association*, 76(373):169–174.

Dobrynin, M., Murawski, J., Baehr, J., and Ilyina, T. (2015). Detection and attribution of climate change signal in ocean wind waves. *Journal of Climate*, 28(4):1578–1591.

Dobrynin, M., Murawsky, J., and Yang, S. (2012). Evolution of the global wind wave climate in CMIP5 experiments. *Geophysical Research Letters*, 39(18).

Durrant, T. H., Greenslade, D. J., and Simmonds, I. (2013). The effect of statistical wind corrections on global wave forecasts. *Ocean Modelling*, 70:116–131.

Elsner, J. B., Kossin, J. P., and Jagger, T. H. (2008). The increasing intensity of the strongest tropical cyclones. *Nature*, 455(7209):92–95.

Emanuel, K. (2005). Increasing destructiveness of tropical cyclones over the past 30 years. *Nature*, 436(7051):686–688.

England, M. H., McGregor, S., Spence, P., Meehl, G. A., Timmermann, A., Cai, W., Gupta, A. S., McPhaden, M. J., Purich, A., and Santoso, A. (2014). Recent intensification of wind-driven circulation in the pacific and the ongoing warming hiatus. *Nature climate change*, 4(3):222.

Epstein, E. S. (1969). Stochastic dynamic prediction. *Tellus*, 21(6):739–759.

Ferreira, J. and Soares, C. G. (1998). An application of the peaks over threshold method to predict extremes of significant wave height. *Journal of Offshore Mechanics and Arctic Engineering*, 120(3):165–176.

Field, C. B., Barros, V., Stocker, T. F., and Dahe, Q. (2012). *Managing the risks of extreme events and disasters to advance climate change adaptation: special report of the intergovernmental panel on climate change*. Cambridge University Press.

Field, C. B., Barros, V. R., Dokken, D., Mach, K., MAS-TRANDREA, M., Bilir, T., Chatterjee, M., Ebi, K., ES-TRADA, Y., Genova, R., et al. (2014). Ipcc 2014: Summary for policymakers in climate change 2014: Impacts, adaptation, and vulnerability. part a: Global and sectoral aspects. contribution of working group ii to the fifth assessment report of the intergovernmental panel on climate change. *Contribution of Working Group II to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*, pages 1–32.

Freeman, E., Woodruff, S. D., Worley, S. J., Lubker, S. J., Kent, E. C., Angel, W. E., Berry, D. I., Brohan, P., Eastman, R., Gates, L., et al. (2017). Icoads release 3.0: a major update to the historical marine climate record. *International Journal of Climatology*, 37(5):2211–2232.

Furevik, B. R. and Haakenstad, H. (2012). Near-surface marine wind profiles from rawinsonde and NORA10 hindcast. *J Geophys Res*, 117(D23):14.

Galambos, J. (1994). Extreme value theory for applications. In *Extreme value theory and applications*, pages 1–14. Springer.

Gemmrich, J., Thomas, B., and Bouchard, R. (2011). Observational changes and trends in northeast pacific wave records. *Geophysical Research Letters*, 38(22).

Gibson, J., Kållberg, P., Uppala, S., Hernandez, A., Nomura, A., and Serrano, E. (1997). ERA description. Shinfield Park, Reading.

Goda, Y. (2010). *Random seas and design of maritime structures*. World scientific.

Goodwin, I. D., Stables, M. A., and Olley, J. M. (2006). Wave climate, sand budget and shoreline alignment evolution of the iluka–woody bay sand barrier, northern new south wales, australia, since 3000 yr bp. *Marine Geology*, 226(1-2):127–144.

Gower, J. (2002). Temperature, wind and wave climatologies, and trends from marine meteorological buoys in the northeast pacific. *Journal of Climate*, 15(24):3709–3718.

Gregow, H., Jylhä, K., Mäkelä, H., Aalto, J., Manninen, T., Karlsson, P., Kaiser-Weiss, A., Kaspar, F., Poli, P., Tan, D., et al. (2016). Worldwide survey of awareness and needs concerning reanalyses and respondents views on climate services. *Bulletin of the American Meteorological Society*, 97(8):1461–1473.

Gulev, S. K. and Grigorieva, V. (2004). Last century changes in ocean wind wave height from global visual wave data. *Geophysical Research Letters*, 31(24).

Gulev, S. K. and Grigorieva, V. (2006). Variability of the winter wind waves and swell in the north atlantic and north pacific as revealed by the voluntary observing ship data. *Journal of Climate*, 19(21):5667–5685.

Hasselmann, K. (1962). On the non-linear energy transfer in a gravity-wave spectrum part 1. general theory. *Journal of Fluid Mechanics*, 12(4):481–500.

Hasselmann, K., Barnett, T., Bouws, E., Carlson, H., Cartwright, D., Enke, K., Ewing, J., Gienapp, H., Hasselmann, D., Kruseman, P., et al. (1973). Measurements of wind-wave growth and swell decay during the joint north sea wave project (jonswap). *Ergänzungsheft 8-12*.

Hasselmann, S., Hasselmann, K., Allender, J., and Barnett, T. (1985). Computations and parameterizations of the nonlinear energy transfer in a gravity-wave spectrum. part ii: Parameterizations of the nonlinear energy transfer for application in wave models. *Journal of Physical Oceanography*, 15(11):1378–1391.

Hawkins, E. and Sutton, R. (2012). Time of emergence of climate signals. *Geophysical Research Letters*, 39(1).

Hemer, M., Wang, X., Church, J., and Swail, V. (2010). Modeling proposal: Coordinating global ocean wave climate projections. *Bulletin of the American Meteorological Society*, 91(4):451–454.

Hemer, M. A., Fan, Y., Mori, N., Semedo, A., and Wang, X. L. (2013). Projected changes in wave climate from a multi-model ensemble. *Nature climate change*, 3(5):471–476.

Hemer, M. A. and Trenham, C. E. (2016). Evaluation of a CMIP5 derived dynamical global wind wave climate model ensemble. *Ocean Modelling*, 103:190–203.

Hemer, M. A., Wang, X. L., Weisse, R., and Swail, V. R. (2012). Advancing wind-waves climate science: The cowclip project. *Bulletin of the American Meteorological Society*, 93(6):791–796.

Hersbach, H. and Dee, D. (2016). ERA5 reanalysis is in production. *ECMWF newsletter*, 147:7.

Hersbach, H., Peubey, C., Simmons, A., Berrisford, P., Poli, P., and Dee, D. (2015a). ERA-20CM: a twentieth-century atmospheric model ensemble. *Quarterly Journal of the Royal Meteorological Society*, 141(691):2350–2375.

Hersbach, H., Poli, P., and Dee, D. (2015b). The observation feedback archive for the ICOADS and ISPD data sets. (18):31.

Hinkel, J., Lincke, D., Vafeidis, A. T., Perrette, M., Nicholls, R. J., Tol, R. S., Marzeion, B., Fettweis, X., Ionescu, C., and Levermann, A. (2014). Coastal flood damage and adaptation costs under 21st century sea-level rise. *Proceedings of the National Academy of Sciences*, 111(9):3292–3297.

Hirsch, R. M. and Slack, J. R. (1984). A nonparametric trend test for seasonal data with serial dependence. *Water Resources Research*, 20(6):727–732.

Hirsch, R. M., Slack, J. R., and Smith, R. A. (1982). Techniques of trend analysis for monthly water quality data. *Water resources research*, 18(1):107–121.

Holthuijsen, L. H. (2007). *Waves in oceanic and coastal waters*. Cambridge University Press.

Holthuijsen, L. H., Powell, M. D., and Pietrzak, J. D. (2012). Wind and waves in extreme hurricanes. *Journal of Geophysical Research: Oceans*, 117(C9).

Howden, S., Gilhousen, D., Guinasso, N., Walpert, J., Sturgeon, M., and Bender, L. (2008). Hurricane Katrina winds measured by a buoy-mounted sonic anemometer. *Journal of Atmospheric and Oceanic Technology*, 25(4):607–616.

Isaksen, L., Haseler, J., Buizza, R., and Leutbecher, M. (2010). The new ensemble of data assimilations. *ECMWF Newsletter*, 123:17–21.

Janssen, P. (2004). *The interaction of ocean waves and wind*. Cambridge University Press.

Jensen, R. E., Swail, V. R., Bouchard, R. H., Riley, R. E., Hesser, T. J., Blaseckie, M., and MacIsaac, C. (2015). Field laboratory for ocean sea state investigation and experimentation: FLOSSIE: Intra-measurement evaluation of 6N wave buoy systems. In *14th Int. Workshop on Wave Hindcasting and Forecasting and Fifth Coastal Hazard Symposium*, volume A1, Key West, FL. WMO/IOC JCOMM.

Kendall, M. G. (1948). Rank correlation methods.

Kent, E. C., Berry, D. I., Woodruff, S. D., and Taylor, P. K. (2006). Voluntary observing ships: A vital marine observing system in decline. *CLIVAR Exchanges*, 38:20–21.

Kent, E. C., Rayner, N. A., Berry, D. I., Eastman, R., Grigorieva, V. G., Huang, B., Kennedy, J. J., Smith, S. R., and Willett, K. M. (2019). Observing requirements for long-term climate records at the ocean surface. *Frontiers in Marine Science*, 6.

Knutson, T., Camargo, S. J., Chan, J. C., Emanuel, K., Ho, C.-H., Kossin, J., Mohapatra, M., Satoh, M., Sugi, M., Walsh, K., et al. (2019). Tropical cyclones and climate change assessment: Part i. detection and attribution. *Bulletin of the American Meteorological Society*, (2019).

Komen, G. J., Cavaleri, L., and Donelan, M. (1994). *Dynamics and modelling of ocean waves*. Cambridge university press.

Kumar, P., Min, S.-K., Weller, E., Lee, H., and Wang, X. L. (2016). Influence of climate variability on extreme ocean surface wave heights assessed from ERA-Interim and ERA-20C. *Journal of Climate*, 29(11):4031–4046.

Laloyaux, P., Balmaseda, M., Dee, D., Mogensen, K., and Janssen, P. (2016). A coupled data assimilation system for climate reanalysis. *Quarterly Journal of the Royal Meteorological Society*, 142(694):65–78.

Laloyaux, P., de Boisseson, E., Balmaseda, M., Bidlot, J.-R., Broennimann, S., Buizza, R., Dalhgren, P., Dee, D., Haimberger, L., Hersbach, H., et al. (2018). CERA-20C: A coupled reanalysis of the Twentieth Century. *Journal of Advances in Modeling Earth Systems*, 10(5):1172–1195.

Landsea, C. W., Harper, B. A., Hoarau, K., and Knaff, J. A. (2006). Can we detect trends in extreme tropical cyclones? *Science*, 313(5786):452–454.

Lang, S., Leutbecher, M., and Jones, S. (2012). Impact of perturbation methods in the ECMWF ensemble prediction system on tropical cyclone forecasts. *Quarterly Journal of the Royal Meteorological Society*, 138(669):2030–2046.

Large, W., Morzel, J., and Crawford, G. (1995). Accounting for surface wave distortion of the marine wind profile in low-level ocean storms wind measurements. *Journal of Physical Oceanography*, 25(11):2959–2971.

Leutbecher, M. and Palmer, T. N. (2008). Ensemble forecasting. *Journal of Computational Physics*, 227(7):3515–3539.

Lewis, J. M. (2005). Roots of ensemble forecasting. *Monthly weather review*, 133(7):1865–1885.

Li, H., Sheffield, J., and Wood, E. F. (2010). Bias correction of monthly precipitation and temperature fields from Intergovernmental Panel on Climate Change AR4 models using equidistant quantile matching. *Journal of Geophysical Research: Atmospheres*, 115(D10).

Liu, Q., Babanin, A., Fan, Y., Zieger, S., Guan, C., and Moon, I.-J. (2017). Numerical simulations of ocean surface waves under hurricane conditions: Assessment of existing model performance. *Ocean Modelling*, 118:73–93.

Lopatoukhin, L., Rozhkov, V., Ryabinin, V., Swail, V., Boukhanovsky, A., and Degtyarev, A. (2000). Estimation of extreme wind wave heights. Technical report, WMO.

Lorenz, E. N. (1963). Deterministic nonperiodic flow. *Journal of the atmospheric sciences*, 20(2):130–141.

Luijendijk, A., Hagenaars, G., Ranasinghe, R., Baart, F., Donchyts, G., and Aarninkhof, S. (2018). The state of the world's beaches. *Scientific Reports*, 8(1).

Makin, V. K. (2005). A note on the drag of the sea surface at hurricane winds. *Boundary-Layer Meteorology*, 115(1):169–176.

Mann, H. B. (1945). Nonparametric tests against trend. *Econometrica: Journal of the Econometric Society*, pages 245–259.

Marcos, M., Rohmer, J., Vousdoukas, M. I., Mentaschi, L., Le Cozannet, G., and Amores, A. (2019). Increased extreme coastal water levels due to the combined action of storm surges and wind waves. *Geophysical Research Letters*, 46(8):4356–4364.

Mathiesen, M., Goda, Y., Hawkes, P. J., Mansard, E., Martín, M. J., Peltier, E., Thompson, E. F., and Van Vledder, G. (1994). Recommended practice for extreme wave analysis. *Journal of hydraulic Research*, 32(6):803–814.

Maury, M. (1853). Maritime conference held at brussels, for devising a uniform system of meteorological observations at sea. *August and September*.

Mentaschi, L., Vousdoukas, M. I., Pekel, J.-F., Voukouvalas, E., and Feyen, L. (2018). Global long-term observations of coastal erosion and accretion. *Scientific reports*, 8(1):12876.

Mentaschi, L., Vousdoukas, M. I., Voukouvalas, E., Dosio, A., and Feyen, L. (2017). Global changes of extreme coastal wave energy fluxes triggered by intensified teleconnection patterns. *Geophysical Research Letters*, 44(5):2416–2426.

Meucci, A., Young, I. R., Aarnes, O. J., and Breivik, Ø. (2020a). Comparison of wind speed and wave height trends from twentieth-century models and satellite altimeters. *Journal of Climate*, 33(2):611–624.

Meucci, A., Young, I. R., and Breivik, Ø. (2018). Wind and wave extremes from atmosphere and wave model ensembles. *Journal of Climate*, 31(21):8819–8842.

Meucci, A., Young, I. R., Hemer, M., Kirezci, E., and Ranasinghe, R. (2020b). Projected 21st century changes in extreme wind-wave events. *Science Advances*, 6(24).

Miles, J. W. (1957). On the generation of surface waves by shear flows. *Journal of Fluid Mechanics*, 3(2):185–204.

Mogensen, K. S., Magnusson, L., and Bidlot, J.-R. (2017). Tropical cyclone sensitivity to ocean coupling in the ECMWF coupled model. *J Geophys Res: Oceans*, 122(5):4392–4412.

Molteni, F., Buizza, R., Palmer, T. N., and Petroliagis, T. (1996). The ECMWF ensemble prediction system: Methodology and validation. *Quarterly journal of the royal meteorological society*, 122(529):73–119.

Morim, J., Hemer, M., Cartwright, N., Strauss, D., and Andutta, F. (2018). On the concordance of 21st century wind-wave climate projections. *Global and planetary change*.

Morim, J., Hemer, M., Wang, X. L., Cartwright, N., Trenham, C., Semedo, A., Young, I., Bricheno, L., Camus, P., Casas-Prat, M., et al. (2019). Robustness and uncertainties in global multivariate wind-wave climate projections. *Nature Climate Change*, 9(9):711–718.

Moss, R., Babiker, W., Brinkman, S., Calvo, E., Carter, T., Edmonds, J., Elgizouli, I., Emori, S., Erda, L., Hibbard, K., et al. (2008). Towards new scenarios for the analysis of emissions: Climate change, impacts and response strategies.

Nonaka, M. and Xie, S.-P. (2003). Covariations of sea surface temperature and wind over the kuroshio and its extension: Evidence for ocean-to-atmosphere feedback. *Journal of Climate*, 16(9):1404–1413.

Nott, J. (2006). *Extreme events: a physical reconstruction and risk assessment*. Cambridge University Press.

Ochi, M. (2003). *Hurricane generated seas*, volume 8. Elsevier.

Ochi, M. K. (2005). *Ocean waves: the stochastic approach*, volume 6. Cambridge University Press.

Overeem, I., Anderson, R. S., Wobus, C. W., Clow, G. D., Urban, F. E., and Matell, N. (2011). Sea ice loss enhances wave action at the arctic coast. *Geophysical Research Letters*, 38(17).

Palmer, T. and Hagedorn, R. (2006). *Predictability of weather and climate*. Cambridge University Press.

Patricola, C. M. and Wehner, M. F. (2018). Anthropogenic influences on major tropical cyclone events. *Nature*, 563(7731):339.

Peterson, E. and Hasse, L. (1987). Did the beaufort scale or the wind climate change? *Journal of physical oceanography*, 17(7):1071–1074.

Pickands, J. I. (1975). Statistical inference using extreme order statistics. *The Annals of Statistics*, pages 119–131.

Pineau-Guillou, L., Ardhuin, F., Bouin, M.-N., Redelsperger, J.-L., Chapron, B., Bidlot, J.-R., and Quilfen, Y. (2018). Strong winds in a coupled wave–atmosphere model during a North Atlantic storm event: evaluation against observations. *Quarterly Journal of the Royal Meteorological Society*, 144(711):317–332.

Poli, P., Hersbach, H., Dee, D., Berrisford, P., Simmons, A., and Laloyaux, P. (2015). ERA-20C deterministic. ERA Report Series 20, European Centre for Medium-Range Weather Forecasts. http://www.ecmwf.int/publications.

Poli, P., Hersbach, H., Dee, D. P., Berrisford, P., Simmons, A. J., Vitart, F., Laloyaux, P., Tan, D. G., Peubey, C., Thépaut, J.-N., et al. (2016). ERA-20C: An atmospheric reanalysis of the twentieth century. *Journal of Climate*, 29(11):4083–4097.

Powell, M. D., Vickery, P. J., and Reinhold, T. A. (2003). Reduced drag coefficient for high wind speeds in tropical cyclones. *Nature*, 422(6929):279–283.

Puri, K., Barkmeijer, J., and Palmer, T. (2001). Ensemble prediction of tropical cyclones using targeted diabatic singular vectors. *Quarterly Journal of the Royal Meteorological Society*, 127(572):709–731.

Rabier, F., Thépaut, J.-N., and Courtier, P. (1998). Extended assimilation and forecast experiments with a four-dimensional variational assimilation system. *Quarterly Journal of the Royal Meteorological Society*, 124(550):1861–1887.

Ramage, C. (1987). Secular change in reported surface wind speeds over the ocean. *Journal of climate and applied meteorology*, 26(4):525–528.

Ranasinghe, R. (2016). Assessing climate change impacts on open sandy coasts: A review. *Earth-science reviews*, 160:320–332.

Ranjha, R., Tjernström, M., Semedo, A., Svensson, G., and Cardoso, R. M. (2015). Structure and variability of the oman coastal low-level jet. *Tellus A: Dynamic Meteorology and Oceanography*, 67(1):25285.

Rascle, N. and Ardhuin, F. (2013). A global wave parameter database for geophysical applications. part 2: Model validation with improved source term parameterization. *Ocean Modelling*, 70:174–188.

Rayner, N., Brohan, P., Parker, D., Folland, C., Kennedy, J., Vanicek, M., Ansell, T., and Tett, S. (2006). Improved analyses of changes and uncertainties in sea surface temperature measured in situ since the mid-nineteenth century: The HadSST2 dataset. *Journal of Climate*, 19(3):446–469.

Reistad, M., Breivik, Ø., Haakenstad, H., Aarnes, O. J., Furevik, B. R., and Bidlot, J.-R. (2011). A high-resolution hindcast of wind and waves for the north sea, the norwegian sea, and the barents sea. *Journal of Geophysical Research: Oceans*, 116(C5).

Rhein, M., Rintoul, S., Aoki, S., Campos, E., Chambers, D., Feely, R., Gulev, S., Johnson, G., Josey, S., Kostianoy, A., et al. (2013). Observations: ocean climate change 2013: The physical science basis. contribution of working group i to the fifth assessment report of the intergovernmental panel on climate change ed tf stocker et al. *Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA*.

Ribal, A. and Young, I. R. (2019). 33 years of globally calibrated wave height and wind speed data based on altimeter observations. *Scientific Data*, 6(77).

Richter, D. H. and Sullivan, P. P. (2013). Sea surface drag and the role of spray. *Geophysical Research Letters*, 40(3):656–660.

Romero, L., Lenain, L., and Melville, W. K. (2017). Observations of surface wave–current interaction. *Journal of Physical Oceanography*, 47(3):615–632.

Ruggiero, P., Komar, P. D., and Allan, J. C. (2010). Increasing wave heights and extreme value projections: The wave climate of the us pacific northwest. *Coastal Engineering*, 57(5):539–552.

Saetra, Ø. and Bidlot, J.-R. (2004). Potential benefits of using probabilistic forecasts for waves and marine winds based on the ECMWF Ensemble Prediction System. *Weather and forecasting*, 19(4):673–689.

Saha, S., Moorthi, S., Pan, H.-L., Wu, X., Wang, J., Nadiga, S., Tripp, P., Kistler, R., Woollen, J., Behringer, D., et al. (2010). The ncep climate forecast system reanalysis. *Bulletin of the American Meteorological Society*, 91(8):1015–1058.

Sasaki, W. (2016). Impact of satellite data assimilation in atmospheric reanalysis on the marine wind and wave climate. *Journal of Climate*, 29(17):6351–6361.

Semedo, A., Sušelj, K., Rutgersson, A., and Sterl, A. (2011). A global view on the wind sea and swell climate and variability from ERA-40. *Journal of Climate*, 24(5):1461–1479.

Sen, P. K. (1968). Estimates of the regression coefficient based on Kendall's tau. *Journal of the American statistical association*, 63(324):1379–1389.

Shimura, T., Mori, N., and Hemer, M. A. (2016). Variability and future decreases in winter wave heights in the Western North Pacific. *Geophys Res Lett*, 43(6):2716–2722.

Shimura, T., Mori, N., and Hemer, M. A. (2017). Projection of tropical cyclone-generated extreme wave climate based on cmip5 multi-model ensemble in the western north pacific. *Climate Dynamics*, 49(4):1449–1462.

Shimura, T., Mori, N., and Mase, H. (2015). Future projections of extreme ocean wave climates and the relation to tropical cyclones: Ensemble experiments of mri-agcm3. 2h. *Journal of Climate*, 28(24):9838–9856.

Sigmond, M., Reader, M., Fyfe, J., and Gillett, N. (2011). Drivers of past and future southern ocean change: Stratospheric ozone versus greenhouse gas impacts. *Geophysical Research Letters*, 38(12).

Slott, J. M., Murray, A. B., Ashton, A. D., and Crowley, T. J. (2006). Coastline responses to changing storm patterns. *Geophysical Research Letters*, 33(18).

Sterl, A. (2004). On the (in) homogeneity of reanalysis products. *Journal of Climate*, 17(19):3866–3873.

Sterl, A., Brink, H. v. d., Vries, H. d., Haarsma, R., and Meijgaard, E. v. (2009). An ensemble study of extreme storm surge related water levels in the north sea in a changing climate. *Ocean Science*, 5(3):369–378.

Sterl, A. and Caires, S. (2005). Climatology, variability and extrema of ocean waves: The web-based KNMI/ERA-40 wave atlas. *International Journal of Climatology*, 25(7):963–977.

Stocker, T. (2014). *Climate change 2013: the physical science basis: Working Group I contribution to the Fifth assessment report of the Intergovernmental Panel on Climate Change*. Cambridge University Press.

Stopa, J. E. and Cheung, K. F. (2014). Intercomparison of wind and wave data from the ecmwf reanalysis interim and the ncep climate forecast system reanalysis. *Ocean Modelling*, 75:65–83.

Takbash, A., Young, I. R., and Breivik, O. (2018). Global wind speed and wave height extremes derived from long-duration satellite records. *Journal of Climate*, (2018).

Taylor, K. E., Stouffer, R. J., and Meehl, G. A. (2012). An overview of CMIP5 and the experiment design. *Bulletin of the American Meteorological Society*, 93(4):485–498.

Taylor, P. K. and Yelland, M. J. (2001). Comments on "On the effect of ocean waves on the kinetic energy balance and consequences for the inertial dissipation technique". *Journal of Physical Oceanography*, 31(8):2532–2536.

Theil, H. (1950). A rank-invariant method of linear and polynominal regression analysis (parts 1-3). In *Ned. Akad. Wetensch. Proc. Ser. A*, volume 53, pages 386–392.

Thomas, B. and Swail, V. (2011). Buoy wind inhomogeneities related to averaging method and anemometer type: application to long time series. *International Journal of Climatology*, 31(7):1040–1055.

Thomas, B. R., Kent, E. C., Swail, V. R., and Berry, D. I. (2008). Trends in ship wind speeds adjusted for observation method and height. *International Journal of Climatology: A Journal of the Royal Meteorological Society*, 28(6):747–763.

Thomson, J. and Rogers, W. E. (2014). Swell and sea in the emerging arctic ocean. *Geophysical Research Letters*, 41(9):3136–3140.

Timmermans, B., Stone, D., Wehner, M., and Krishnan, H. (2017). Impact of tropical cyclones on modeled extreme wind-wave climate. *Geophysical Research Letters*, 44(3):1393–1401.

Tokinaga, H. and Xie, S.-P. (2011). Wave-and anemometer-based sea surface wind (WASWind) for climate change analysis. *Journal of Climate*, 24(1):267–285.

Tolman, H. L., Banner, M. L., and Kaihatu, J. M. (2013). The nopp operational wave model improvement project. *Ocean Modelling*, 70:2–10.

Tolman, H. L. and Chalikov, D. (1996). Source terms in a third-generation wind wave model. *Journal of Physical Oceanography*, 26(11):2497–2518.

Tolman, H. L. et al. (2009). User manual and system documentation of WAVEWATCH III TM version 3.14. *Technical note, MMAB Contribution*, 276:220.

Tucker, M. J. and Pitt, E. G. (2001). *Waves in ocean engineering*. Number Volume 5.

Uppala, S. M., Kållberg, P., Simmons, A., Andrae, U., Bechtold, V., Fiorino, M., Gibson, J., Haseler, J., Hernandez, A., Kelly, G., et al. (2005). The ERA-40 re-analysis. *Quarterly Journal of the royal meteorological society*, 131(612):2961–3012.

Van Vledder, G. P. and Holthuijsen, L. (1993). The directional response of ocean waves to turning winds. *Journal of Physical Oceanography*, 23(2):177–192.

Van Vuuren, D. P., Edmonds, J., Kainuma, M., Riahi, K., Thomson, A., Hibbard, K., Hurtt, G. C., Kram, T., Krey, V., Lamarque, J.-F., et al. (2011). The representative concentration pathways: an overview. *Climatic change*, 109(1-2):5.

Veritas, D. N. (2010). Environmental conditions and environmental loads, dnv-rp-c205. *Recommended Practice, October*.

Vinoth, J. and Young, I. (2011). Global estimates of extreme wind speed and wave height. *Journal of Climate*, 24(6):1647–1665.

von Storch, H. and Zwiers, F. (1999). *Statistical Analysis in Climate Research*. Cambridge University Press.

Walsh, K. J., McBride, J. L., Klotzbach, P. J., Balachandran, S., Camargo, S. J., Holland, G., Knutson, T. R., Kossin, J. P., Lee, T.-c., Sobel, A., et al. (2016). Tropical cyclones and climate change. *Wiley Interdisciplinary Reviews: Climate Change*, 7(1):65–89.

WAMDI, G. (1988). The WAM model - A third generation ocean wave prediction model. *Journal of Physical Oceanography*, 18(12):1775–1810.

Wang, X. L., Feng, Y., Compo, G., Swail, V., Zwiers, F., Allan, R., and Sardeshmukh, P. (2013). Trends and low frequency variability of extra-tropical cyclone activity in the ensemble of twentieth century reanalysis. *Climate Dynamics*, 40(11-12):2775–2800.

Wang, X. L., Feng, Y., and Swail, V. (2012). North Atlantic wave height trends as reconstructed from the 20th century reanalysis. *Geophysical Research Letters*, 39(18).

Wang, X. L., Feng, Y., and Swail, V. R. (2014). Changes in global ocean wave heights as projected using multimodel CMIP5 simulations. *Geophysical Research Letters*, 41(3):1026–1034.

Wang, X. L., Feng, Y., and Swail, V. R. (2015). Climate change signal and uncertainty in cmip5-based projections of global ocean surface wave heights. *Journal of Geophysical Research: Oceans*, 120(5):3859–3871.

Wang, X. L. and Swail, V. R. (2001). Changes of extreme wave heights in northern hemisphere oceans and related atmospheric circulation regimes. *Journal of Climate*, 14(10):2204–2221.

Wang, X. L. and Swail, V. R. (2006). Climate change signal and uncertainty in projections of ocean wave heights. *Climate Dynamics*, 26(2-3):109–126.

Waseda, T., Webb, A., Sato, K., Inoue, J., Kohout, A., Penrose, B., and Penrose, S. (2018). Correlated increase of high ocean waves and winds in the ice-free waters of the arctic ocean. *Scientific reports*, 8(1):4489.

Weisse, R. (2010). *Marine climate and climate change: storms, wind waves and storm surges*. Springer Science & Business Media.

Wigley, T. M. (2009). The effect of changing climate on the frequency of absolute extreme events. *Climatic Change*, 97(1-2):67.

Wilks, D. S. (2011). *Statistical methods in the atmospheric sciences*, volume 100. Academic press.

WMO (1998). Guide to wave analysis and forecasting. Technical report, WMO.

Wohland, J., Omrani, N.-E., Witthaut, D., and Keenlyside, N. S. (2019). Inconsistent wind speed trends in current twentieth century reanalyses. *Journal of Geophysical Research: Atmospheres*, 124(4):1931–1940.

Woodruff, S., Diaz, H., Elms, J., and Worley, S. (1998). COADS Release 2 data and metadata enhancements for improvements of marine surface flux fields. *Physics and Chemistry of the Earth*, 23(5-6):517–526.

Woodruff, S. D., Worley, S. J., Lubker, S. J., Ji, Z., Eric Freeman, J., Berry, D. I., Brohan, P., Kent, E. C., Reynolds, R. W., Smith, S. R., et al. (2011). ICOADS Release 2.5: extensions and enhancements to the surface marine meteorological archive. *International journal of climatology*, 31(7):951–967.

Young, I. (1994). Global ocean wave statistics obtained from satellite observations. *Applied Ocean Research*, 16(4):235–248.

Young, I. and Donelan, M. (2018). On the determination of global ocean wind and wave climate from satellite observations. *Remote Sensing of Environment*, 215:228–241.

Young, I. R. (1988). Parametric hurricane wave prediction model. *Journal of Waterway, Port, Coastal, and Ocean Engineering*, 114(5):637–652.

Young, I. R. (1999a). Seasonal variability of the global ocean wind and wave climate. *International Journal of Climatology*, 19(9):931–950.

Young, I. R. (1999b). *Wind generated ocean waves*, volume 2. Elsevier.

Young, I. R. (2006). Directional spectra of hurricane wind waves. *Journal of Geophysical Research: Oceans*, 111(C8).

Young, I. R. (2017). A review of parametric descriptions of tropical cyclone wind-wave generation. *Atmosphere*, 8(10):194.

Young, I. R. and Babanin, A. V. (2006). Spectral distribution of energy dissipation of wind-generated waves due to dominant wave breaking. *Journal of Physical Oceanography*, 36(3):376–394.

Young, I. R. and Holland, G. (1996). *Atlas of the Oceans: Wind and Wave Climate*. Pergamon.

Young, I. R. and Ribal, A. (2019). Multiplatform evaluation of global trends in wind speed and wave height. *Science*, 364(6440):548–552.

Young, I. R., Sanina, E., and Babanin, A. (2017). Calibration and cross-validation of a global wind and wave database of altimeter, radiometer and scatterometer measurements. *Journal of Atmospheric and Oceanic Technology*, (2017).

Young, I. R., Vinoth, J., Zieger, S., and Babanin, A. V. (2012). Investigation of trends in extreme value wave height and wind speed. *Journal of Geophysical Research: Oceans*, 117(C11).

Young, I. R., Zieger, S., and Babanin, A. V. (2011). Global trends in wind speed and wave height. *Science*, 332(6028):451–455.

Zeng, L. and Brown, R. A. (1998). Scatterometer observations at high wind speeds. *Journal of Applied Meteorology*, 37(11):1412–1420.

Zieger, S., Babanin, A. V., Rogers, W. E., and Young, I. R. (2015). Observation-based source terms in the third-generation wave model wavewatch. *Ocean Modelling*, 96:2–25.

Zieger, S., Babanin, A. V., and Young, I. R. (2014). Changes in ocean surface wind with a focus on trends in regional and monthly mean values. *Deep Sea Research Part I: Oceanographic Research Papers*, 86:56–67.

Zieger, S., Vinoth, J., and Young, I. R. (2009). Joint calibration of multiplatform altimeter measurements of wind speed and wave height over the past 20 years. *Journal of Atmospheric and Oceanic Technology*, 26(12):2549–2564.