

ON WEATHER AND WAVES: APPLICATIONS TO COASTAL ENGINEERING



Justin Pringle

University of KwaZulu Natal

Submitted in fulfillment of the academic requirements for the degree of
*Doctor of Philosophy in Coastal Engineering, College of Agriculture,
Engineering and Science, University of KwaZulu Natal, Durban*

2015

As the candidate's Supervisor I agree/do not agree to the submission of this thesis.

.....
Prof. Derek Stretch

.....
Date

Declaration

I *Justin James Pringle* declare that

1. The research reported in this thesis, except where otherwise indicated, is my original research.
2. This thesis does not contain other persons data, pictures, graphs or other information, unless specifically acknowledged as being sourced from other persons.
3. This thesis does not contain other persons' writing, unless specifically acknowledged as being sourced from other researchers. Where other written sources have been quoted, then:
 - a Their words have been re-written but the general information attributed to them has been referenced.
 - b Where their exact words have been used, then their writing has been placed in italics and inside quotation marks, and referenced.
4. This thesis does not contain text, graphics or tables copied and pasted from the Internet, unless specifically acknowledged, and the source being detailed in the thesis and in the References sections.

Signed

Justin Pringle
2015

Declaration

Details of contribution to publications that form part and/or include research presented in this thesis

Pringle. J, Stretch. D, Bárdossy. A. 2014. Automated classification of atmospheric circulation patterns that drive regional wave climates. *Natural Hazards and Earth System Sciences* 14, 2145–2155.

Pringle - Lead Author.

Stretch - Supervisor.

Bárdossy - Co-Supervisor

Pringle, J. J., Stretch, D. D., Bárdossy, A., 2015. On linking atmospheric circulation patterns to extreme wave events for coastal vulnerability assessments. *Natural Hazards*. <http://dx.doi.org/10.1007/s11069-015-1825-4>

Pringle - Lead Author.

Stretch - Supervisor.

Bárdossy - Co-Supervisor

Corbella, S., Pringle, J., Stretch, D., 2015. Assimilation of ocean wave spectra and atmospheric circulation patterns to improve wave modelling. *Coastal Engineering* 10, 40–53.

Corbella and Pringle - Lead Authors. Corbella contributed to the wave tracking section and Pringle contributed to the Classification. Both authors had equal contribution to the paper.

Stretch - Supervisor.

Pringle, J., Stretch, D., 2015. Stochastic simulation of regional wave climates conditioned on synoptic scale meteorology. part 1: Methodology. Coastal Engineering (Submitted to Coastal Engineering).

Pringle - Lead Author.

Stretch - Supervisor.

Pringle, J., Stretch, D., 2015. Stochastic simulation of regional wave climates conditioned on synoptic scale meteorology. part 2: Applications in coastal vulnerability assessment (Submitted to Coastal Engineering).

Pringle - Lead Author.

Stretch - Supervisor.

Pringle, J., Stretch, D., 2015. Atmospheric classification as a framework for assessing future coastal vulnerability. (Submitted to Coastal Engineering).

Pringle - Lead Author.

Stretch - Supervisor.

Bárdossy, A., Pegram, G., Sinclair, S., Pringle, J., Stretch, D., 2015. Circulation patterns identified through spatial rainfall and ocean wave fields in southern africa. *Frontiers in environmental science*. 3:31. doi: 10.3389/fenvs.2015.00031

Bárdossy - Lead Author.

Pegram & Sinclair - Classification and Precipitation section.

Pringle & Stretch - Classification and Wave section.

Signed

Justin Pringle
2015

Acknowledgements

To Neville, Wilma, Lee and Margeret through your eyes you helped me see and with your thoughts I am.

On my journey I have encountered many inspiring people, each in there own way guiding me along my path. Jade you have shown me only love, kindness and patience even when faced with the onslaught of my thought processes. My parents: Duncan and Eden you have always put my needs first and I am grateful for all you have sacrificed for my personal development. Ceara and James I am grateful for the base you have provided me when I would lose myself along the way.

To my supervisor Derek Stretch I owe so much of my personal development to your relentless quest to push me further and explore the unknown confidently. I am grateful for your support both in the technical and personal spheres of my life.

Prof Bárdossy I am grateful for your help and guidance. Your input to this thesis was invaluable. This study was inspired by your innovation.

Here's to unproductive Wednesdays with Stef and Derrick and endless discussions with Pier and Vuli all within the context of a faulty air conditioner.

To the support staff at Civil Engineering: Ooma Chetty, Sydney Mpungose (*Wena wendlovu!*) and Logan Govender I am grateful for your help, support and willingness to drop everything at a moments notice.

It is within this fantastically intricate universe that this research must now stand, and stand it shall. Therefore to the one God who provides the context, I am grateful.

Abstract

Shoreline erosion in response to extreme wave events can be severe. The reduction in beach width leaves development within the hinterland exposed and vulnerable to future wave attack. Wave climates are a fundamental driver of coastal erosion and changes to wave height, direction and period can severely impact a coastline. These changes are directly linked to changes within the principle drivers of wave climates namely synoptic scale atmospheric circulation. The links are complex and if they can be clarified they can be used to provide insight into wave climates and improve the evaluation of future climate scenarios. The coupling between atmospheric circulation and wave climates provides a tool for risk assessment that is strongly based on fundamental physical processes. This study is focused on exploring this relationship and its effect on coastal vulnerability.

A statistical classification algorithm is utilized to explore the relationship between synoptic scale circulation patterns and regional wave climates. The algorithm is fully automated and discrete atmospheric patterns are derived through an optimization procedure. It is driven to an optimal solution through statistical links between regional wave climates and atmospheric circulation patterns (CPs). The classification is based on the concept of fuzzy sets and differs from standard classification techniques. It employs a "bottom-up" approach as the classes (or CPs) are derived through a procedure that is guided by the wave climate. In contrast existing classification techniques first explore the atmospheric pressure space while links to the variable of interest are only made post classification.

The east coast of South Africa was used as a case study. Wave data off the Durban coastline were utilized to evaluate the drivers of the wave climate. A few dominant patterns are shown to drive extreme wave events. Their persistence and strong high-low coupling drive winds toward the coastline and result in extreme wave events. The sensitivity of the algorithm to key input parameters such as the number of CP classes and temporal resolution of the data was evaluated. The Shannon entropy is introduced to measure the performance of the algorithm. This method benefits from

incorporating the link between atmospheric CPs and the wave climate.

A new stochastic wave simulation technique was developed that is fundamentally based on the CPs. This technique improves the realism of stochastic models while retaining their simplicity and parsimony relative to process-based models. The simplicity of the technique provides the framework to evaluate coastal vulnerability at site specific locations. Furthermore the technique was extended to evaluate changes in wave behaviour due to climate change effects.

To those who have gone before, to those who are present and to those who are still
to come. I dedicate this to you. Live, Love, Learn, Life.

Contents

| | |
|--|-------------|
| Contents | xi |
| List of Figures | xvii |
| List of Tables | xxv |
| Nomenclature | xxvi |
| 1 Introduction | 1 |
| 1.1 Coastal Vulnerability | 1 |
| 1.2 Problem Definition | 3 |
| 1.2.1 Research Question | 3 |
| 1.3 Motivation | 4 |
| 1.3.1 Risk Assessment | 4 |
| 1.3.2 Global Wave Models | 5 |
| 1.4 Aims and Objectives | 6 |
| 1.5 Approach | 7 |
| 1.6 Thesis Structure | 7 |
| 2 Literature Review | 11 |
| 2.1 Coastal Vulnerability | 11 |
| 2.1.1 Statistical Wave models | 12 |
| 2.1.2 Global Wave Hindcast Data | 18 |
| 2.1.3 Shoreline Response Models | 19 |
| 2.2 Atmospheric Classification | 21 |
| 2.2.1 Unsupervised Classification Techniques | 22 |
| 2.2.2 Supervised Classification | 25 |
| 2.2.3 Lagrangian Classification | 28 |

Contents

| | | |
|----------|--|-----------|
| 2.2.4 | An Optimal Classification | 29 |
| 2.3 | Summary of Chapter | 30 |
| 3 | Atmospheric circulation patterns & wave climates | 33 |
| 3.1 | Introduction | 34 |
| 3.2 | Methods | 36 |
| 3.2.1 | Case Study Site | 36 |
| 3.2.2 | Sources of Data | 37 |
| 3.2.3 | Classification Methods | 37 |
| 3.2.4 | Optimization Methods | 40 |
| 3.2.5 | Classification Quality | 42 |
| 3.3 | Results | 43 |
| 3.3.1 | Dominant CP Classes | 43 |
| 3.3.2 | CP Variability | 46 |
| 3.3.3 | CP rules and extreme events | 48 |
| 3.4 | Discussion | 52 |
| 3.5 | Conclusions | 53 |
| 4 | On linking atmospheric circulation patterns to extreme wave events for coastal vulnerability assessments. | 55 |
| 4.1 | Introduction | 56 |
| 4.2 | Method | 58 |
| 4.2.1 | Case Study Site | 58 |
| 4.2.2 | Optimization and Classification | 60 |
| 4.2.3 | Classification Quality Measures | 61 |
| 4.3 | Results | 64 |
| 4.3.1 | Previous Classification Results | 64 |
| 4.3.2 | Classification Quality | 65 |
| 4.3.3 | Sensitivity to Temporal Resolution in the data | 68 |
| 4.3.4 | CP Similarity | 70 |
| 4.4 | Discussion | 72 |
| 4.4.1 | CP links to wave climates | 72 |
| 4.4.2 | Applications to Wave Modelling | 73 |
| 4.5 | Conclusions | 74 |

| | | |
|--------------|--|----------------|
| 5 | Assimilation of ocean wave spectra and atmospheric circulation patterns to improve wave modelling | 77 |
| 5.1 | Introduction | 78 |
| 5.2 | Methods | 80 |
| 5.2.1 | Case study site | 80 |
| 5.2.2 | Partitioning of spectral wave energy | 83 |
| 5.2.3 | Swell tracking | 84 |
| 5.2.4 | Atmospheric classification | 86 |
| 5.2.5 | Linking swell events to circulation patterns | 89 |
| 5.3 | Results | 89 |
| 5.3.1 | Partitioning KwaZulu-Natal wave data | 89 |
| 5.3.2 | Swell origins and associated circulation patterns | 93 |
| 5.3.3 | Average wave directional spectra and associated circulation patterns | 95 |
| 5.4 | Discussion | 99 |
| 5.5 | Conclusions | 102 |
| 6 | Stochastic simulation of regional wave climates conditioned on synoptic scale meteorology. Part1: Methodology | 103 |
| 6.1 | Introduction | 104 |
| 6.2 | Methods | 106 |
| 6.2.1 | Case Study Site | 106 |
| 6.2.2 | Data Sources | 107 |
| 6.2.3 | General Approach | 108 |
| 6.2.4 | Classification of Atmospheric Circulation patterns | 108 |
| 6.2.5 | Wave Climate Simulation | 110 |
| 6.2.6 | Physical Limits | 115 |
| 6.2.7 | Copula Based Simulation | 116 |
| 6.2.8 | Statistical Comparisons | 119 |
| 6.3 | Results | 119 |
| 6.3.1 | CP Simulation Statistics | 119 |
| 6.3.2 | Wave climate dependence structures | 121 |
| 6.3.3 | Univariate Return Periods | 123 |
| 6.3.4 | Wave Climate Distribution | 125 |
| 6.4 | Discussion | 126 |

Contents

| | | |
|----------|--|------------|
| 6.5 | Conclusions | 129 |
| 7 | Stochastic simulation of regional wave climates conditioned on synoptic scale meteorology. Part 2: Applications in coastal vulnerability assessment | 131 |
| 7.1 | Introduction | 132 |
| 7.2 | Method | 134 |
| 7.2.1 | Case Study Site | 134 |
| 7.2.2 | Data Sources | 134 |
| 7.2.3 | Automated Classification | 135 |
| 7.2.4 | Wave Climate Simulation | 136 |
| 7.2.5 | Model Validation | 137 |
| 7.3 | Results | 143 |
| 7.3.1 | Observed and Modelled Wave Directional Statistics | 143 |
| 7.3.2 | Longshore Transport | 146 |
| 7.3.3 | Cross-shore Transport | 147 |
| 7.4 | Discussion and Conclusion | 152 |
| 8 | Atmospheric classification as a framework for assessing future coastal vulnerability. | 155 |
| 8.1 | Introduction | 156 |
| 8.2 | Method | 158 |
| 8.2.1 | Case Study Site | 158 |
| 8.2.2 | Data Sources | 159 |
| 8.2.3 | Classifying Atmospheric Circulation | 160 |
| 8.2.4 | Classification Similarity | 162 |
| 8.2.5 | Wave Simulation | 163 |
| 8.2.6 | Evaluating Changes in Wave climate | 164 |
| 8.3 | Results | 164 |
| 8.3.1 | Measured and reanalysed wave heights | 164 |
| 8.3.2 | Classification Comparison | 168 |
| 8.3.3 | Future Wave Climate | 171 |
| 8.4 | Discussion | 177 |
| 8.5 | Conclusion | 179 |

| | | |
|--|---|------------|
| 9 | Synthesis and Conclusion | 181 |
| 9.1 | Introduction | 181 |
| 9.2 | Improvements to Coastal Vulnerability Assessment | 181 |
| 9.3 | An Alternative to Reanalysed Wave Datasets | 183 |
| 9.4 | Elements of Originality | 184 |
| 9.5 | Conclusion | 185 |
| 9.6 | Suggestions for Future Research | 185 |
| References | | 187 |
| Appendix A Circulation Patterns identified by spatial rainfall and ocean wave fields in Southern Africa | | |
| | | 197 |
| A.1 | Introduction | 198 |
| A.2 | Methodology | 200 |
| A.2.1 | Classification quality | 203 |
| A.2.2 | Non-uniqueness issues | 203 |
| A.2.3 | Data used | 204 |
| A.3 | Classification for Precipitation | 205 |
| A.3.1 | Objective functions | 205 |
| A.3.2 | Spatial extent of classification and non uniqueness of CP set selection | 205 |
| A.4 | Classification for waves | 210 |
| A.4.1 | Case study description | 210 |
| A.4.2 | Dominant CP Classes | 212 |
| A.4.3 | CP Variability | 214 |
| A.4.4 | Variability Within Classes | 214 |
| A.4.5 | CP rules and extreme events | 216 |
| A.5 | Discussion and conclusions | 218 |
| Appendix B Additional Wave Height and Circulation Pattern Statistics | | |
| | | 221 |
| B.1 | Introduction | 221 |
| B.2 | Method | 222 |
| B.3 | Results | 223 |
| B.4 | Discussion | 229 |

Contents

| | |
|---|------------|
| Appendix C Classification Algorithm Pseudocode | 231 |
| C.1 Optimization Algorithm | 231 |

List of Figures

| | | |
|-----|--|----|
| 3.1 | Locations of the wave observation buoys at Durban and Richards Bay, along the KwaZulu Natal coastline. | 38 |
| 3.2 | Average anomaly patterns for all CP classes:1–8. Positive anomaly contours are shown as the dashed line while negative contours are solid. . . | 46 |
| 3.3 | Average anomaly pattern for CP03 (a) with (b) the anomaly with highest DOF, (c) the anomaly with lowest DOF value, while (d) shows the standard deviation for all CP03 anomalies. | 49 |
| 3.4 | (a) Average CP03 with (+) symbols indicating the centers of all negative anomalies (low pressures) contributing to the class. (b) & (c) show actual CP’s for the dates 19/03/2007 and 30/08/2006 respectively, both of which were classified as members of the CP03 class. | 50 |
| 3.5 | CP’s associated with the six largest significant wave heights for the dates (a) 19/3/2007, (b) 5/5/2001, (c) 18/3/2001, (d) 3/4/2001, (e) 23/9/1993 and (f) 19/3/2001. | 51 |
| 4.1 | Locations of wave observations at Durban and Richards Bay on the KwaZulu-Natal coastline of South Africa | 59 |
| 4.2 | The CP anomaly that according to Pringle <i>et al.</i> (2014) drives approximately 40–60% of extreme wave events at the case study site (Fig 4.1). | 65 |
| 4.3 | Classification quality measures (a) Shannon entropy, (b) PCR, (c) EV and (d) WSD for classifications using different numbers of classes showing actual values (points). Solid lines indicate approximate general trends in the quality measures. | 67 |

List of Figures

| | | |
|-----|--|----|
| 4.4 | The outcome of different classification schemes where the number of specified classes varied from 1–16. The results reveal a persistence common class that occurs in each scheme as is shown in sub-plots (a) – (l). The common class is an anomalous low pressure east of the coastline. Solid contours delineate positive (high) pressure anomalies whereas dashed contours represent negative (low) pressure anomalies. | 71 |
| 5.1 | Map of South Africa showing KwaZulu-Natal with locations of the wave recorders. | 81 |
| 5.2 | A map of the Indian Ocean with the dashed lines showing the grouping segments. The small blue dots show a spatial grouping of swell sources and the larger black dots show a grouping by wave characteristics. The grouping by wave characteristics has the following parameters: $1.5\text{m} < H_s < 2.0\text{m}$, $10\text{s} < T_p < 12\text{s}$, $90^\circ < \theta < 135^\circ$. The large black dot indicates the location of the Durban wave recording buoy | 87 |
| 5.3 | Wind roses for Durban (31/05/2002 – 31/07/2013) and Richards Bay (19/08/1993 – 31/07/2013 | 90 |
| 5.4 | An example of a directional energy spectrum (a) and frequency spectrum (d) partitioned into their respective swell (b & e) and wind-wave spectra (c & f) | 92 |
| 5.5 | The minimum energy threshold for the Durban data as defined by $A/(\omega_p^4 + B)$. The dots show the total energy of each swell partition. | 93 |
| 5.6 | A map of the Indian Ocean with the coloured dots showing all the swell sources for Durban. The colours depict the swell periods as defined in the legend. The large black dot indicates the location of the Durban wave recording buoy | 94 |
| 5.7 | The origins of swell events and associated CP anomalies using the spatial grouping method. The class assigned by the classification algorithm is shown in the top panel. The swell origins (dots) and associated average CP anomalies are shown in the centre panel. The associated wave directional spectra are plotted in the bottom panel. The characteristic CPs are classified as (a) CP15, (b) CP10, (c) CP12 (d) CP15 and (e) CP10. | 96 |

| | | |
|------|--|-----|
| 5.8 | The origins of swell events and associated CP anomalies using the wave characteristics grouping method. The class assigned by the classification algorithm is shown in the top panel. The swell origins (dots) and associated average CP anomalies are shown in the centre panel. The associated wave directional spectra are plotted in the bottom panel. The characteristic CPs are classified as (a) CP03, (b) CP03, (c) CP08, (d) CP07 and (e) CP09. | 97 |
| 5.9 | The average wave directional spectra for associated CPs: (a) CP03, (b) CP07, (c) CP08 and (d) CP09. | 98 |
| 5.10 | An example of storm wave events in 2001 with two different driving weather systems: (a) a mid-latitude cyclone occurring on 2001/04/03, and (b) a tropical cyclone occurring on 2001/04/08. Both events may have been grouped as one storm by traditional threshold based methods. A threshold of 3.5m is shown by the dashed line in (c) | 102 |
| 6.1 | Locations of the wave observation buoys at Durban and Richards Bay, along the KwaZulu Natal coastline (Pringle <i>et al.</i> , 2014). | 107 |
| 6.2 | Transition probability matrices showing strong diagonal dominance in CP transitions. Arbitrary CP numbers are assigned to the classes and are shown on the plot as the x and y axes accordingly. | 112 |
| 6.3 | Cumulative frequency distributions for CPs associated with small wave heights (solid line) and a CP associated with large waves (dashed line). Also shown are the exponential tails for each distribution (red) calculated from Eq. 6.5. | 114 |
| 6.4 | (a) A sample scatter plot of log transformed wave heights with 6-hour separation showing strong dependence and (b) the lag one wave height plot of the marginals $F(H_s^t)$ and $F(H_s^{t+1})$ showing the rank correlation. | 115 |
| 6.5 | Observed wave height and directions for a particular CP. | 116 |
| 6.6 | Example of a simulated storm sequence showing the wave height (a), wave period (b) and wave direction (c). | 117 |
| 6.7 | Average anomaly patterns for all CP classes 1 – 16. | 120 |

List of Figures

- 6.8 Observed and simulated lag one wave heights for particular CPs: CPs associated with large wave heights (a&b) and (b) a CP associated with low wave heights. (a) & (c) show CPs associated with large waves and their dependence structure for Autumn, (b) shows a CP associated with small waves for the summer months. The black dots indicate the observed temporal structure and grey dots are simulated from the N12 copula. 122
- 6.9 Observed (black dots) and simulated (grey dots) scatter plots of wave heights and wave periods for a particular CP. The solid line indicates the wave steepness limit using Eq. 6.6. 123
- 6.10 Simulated (red line) wave height return periods for the KwaZulu Natal Coastline. Scatter points show the observed values, the black solid line is the fitted extreme value distribution after Corbella & Stretch (2012d) and the dashed line shows the ERA-Interim reanalysis data. The shaded region indicates the 80th and 20th percentile of wave height extreme obtained from the markov simulation. 124
- 6.11 Wave roses for observed (left) and simulated (right) wave climates with the associated CPs (centre). Surface wind streamlines are also shown in the plot. The line width indicates the relative magnitude of the wind velocity. Significant wave heights for the wave roses are given by the colour legend below the plot. 127
- 7.1 Locations of the wave observation buoy at Durban and the Bluff beach profile used herein along the KwaZulu Natal coastline (Pringle *et al.*, 2014). 135
- 7.2 Equilibrium shoreline response to an increase in water level (S) according to Kriebel & Dean (1993). 139
- 7.3 Equilibrium beach profile envelop according to Jara *et al.* (2015). The black line indicates the EBP envelop that comprises all the individual EBPs (dashed lines). 142
- 7.4 Equilibrium energy function that relates the incoming wave energy to an equilibrium shoreline position (Jara *et al.*, 2015). 143

7.5 Seasonal wave roses for observed wave data (left panel) and modelled (right panel) for the seasons: (a) Summer, (b) Autumn, (c) Winter and (d) Spring. The significant wave heights associated with the wind rose directions are given by the colour legend shown below the plot. 145

7.6 Probability distributions for (a) the observed longshore transport rates and (b) simulated longshore transport rates. The median (red line) and the 95th% confidence limits (shaded) are also shown in the plot 146

7.7 The erosion return periods based on the convolution method after Kriebel & Dean (1993) showing the median (red line) and 80% confidence limits (shaded). Scatter points show erosion volumes and return periods for the observed dataset. Erosion volumes previously estimated by Corbella & Stretch (2012a) are shown by the grey crosses. 148

7.8 The 0 MSL contour shoreline position for the beach profile used herein for the period 1993–2009. The observed shoreline positions are shown as scatter points connected with dashed lines. Modelled shoreline positions are shown with the solid black line. The wave height data for the period is also shown. 149

7.9 Example of simulated 0 MSL contour shoreline positions using the synthetic wave dataset for the period 1990–2090. Wave height data are shown in the top panel 151

7.10 Shoreline position PDF (a) and CDF (b) for the 0 MSL contour as calculated using the simulated waves. The median (red line) and the 5th and 95th% confidence limits (shaded) are also shown in the plot. . . 152

8.1 Locations of the wave observation buoys at Durban and Richards Bay, along the KwaZulu Natal coastline (Pringle *et al.*, 2014). 159

8.2 The relationship between the measured and ERA-reanalysed wave data are shown in (a) the lagged Pearson's correlation coefficients, (b) scatter plot of wave heights and (c) Quantile-Quantile plot 165

8.3 The March 2007 storm event showing (a) the storm track with corresponding CP at the peak of the storm event. (b) The observed (solid line) and ERA modeled (dashed line) wave heights at the Durban wavewider buoy for March 2007. 166

List of Figures

| | | |
|------|--|-----|
| 8.4 | (a) A typical mid-latitude cyclone with associated track and CP two days before the storm peak. (b) The observed (solid line) and modeled (dashed line) wave heights at the Durban waverider buoy for May 2001. | 167 |
| 8.5 | (a) The tracks of a tropical cyclone (black) and a cut-off low (red) and CP at time of the storm peak. (b) The observed (solid line) and ERA modeled (dashed line) wave heights at the Durban waverider buoy for Feb 1997. | 167 |
| 8.6 | Classes derived using observed wave heights (a) and modelled wave heights (b). According to Table 8.2 these classes have the highest degree of association. | 171 |
| 8.7 | (a) The class with the largest ($\sim 54\%$) contribution to extreme wave events according to modelled data (a) and (b) its corresponding class with strongest association according to Table 8.2. | 171 |
| 8.8 | (a) The CP class with the second largest ($\sim 16\%$) contribution to extreme wave events according to modelled data and (b) its corresponding class with strongest association according to Table 8.2. | 172 |
| 8.9 | Return period statistics and associated wave heights for two different future scenarios: (a) RCP2.6 and (b) RCP8.5. Wave heights and return periods for the distant future (2050-2100) are shown in blue whereas those in the near future (2010-2050) are shown in red. Shaded regions indicate the 80% confidence limits. | 174 |
| 8.10 | Wave roses derived from the the HadleyGEM2-ES RCP8.5 scenario for the (a) near (2010-2050) and (b) distant future (2050-2100). | 176 |
| A.1 | The Kruger climate regions of South Africa after Kruger (2004). | 206 |
| A.2 | Three different CPs and the related distributions of daily precipitation in Mpumalanga in Region 5 of Fig. 1. the maps on the left are CPs 8, 2 and 5, appearing as blue (middle), red (dry) and green (wet) lines in the Fig. of frequency distributions. | 207 |
| A.3 | 2 pairs of CPs: top two similar to each other and the bottom two likewise, selected from the CPs chosen on 2 sets of gauges in Region 6. | 208 |
| A.4 | Locations of the wave observation buoys at Durban and Richards Bay, along the KwaZulu Natal coastline. | 211 |

A.5 Average anomaly patterns for 8 CP classes derived from the regional wave climate data. Negative (low) pressure anomaly contours are shown as solid lines while positive pressure contours are dashed. 214

A.6 Average anomaly pattern for CP03 (a) showing (b) the anomaly with highest DOF, (c) the anomaly with lowest DOF value, and (d) the standard deviation for all CP03 anomalies. 216

A.7 (a) Average anomaly pattern CP03 with (+) symbols indicating the centers of all negative anomalies (low pressures) contributing to the class. (b) & (c) show actual CP's for the dates 19/03/2007 and 30/08/2006 respectively, both of which were classified as members of the CP03 class. 217

A.8 CP's associated with the six largest significant wave heights for the dates (a) 19/3/2007, (b) 5/5/2001, (c) 18/3/2001, (d) 3/4/2001, (e) 23/9/1993 and (f) 19/3/2001. 218

B.1 CP correlation coefficients (black) and wave heights (grey) for the year 2007. The correlation coefficients were calculated between CPs of consecutive 6-hourly time steps. 224

B.2 (a) Correlation coefficients (black) and wave heights (grey) for the storm event occurring during the month of February 2007. (b) and (c) are the CP averaged values for the persistence period and storm period respectively.. . . . 226

B.3 (a) Correlation coefficients (black) and wave heights (grey) for the storm event occurring during the month of March 2007. (b) and (c) are the CP averaged values for the persistence period and storm period respectively 227

B.4 (a) Correlation coefficients (black) and wave heights (grey) for the storm event occurring during the month of May 2007. (b) and (c) are the CP averaged values for the persistence period and storm period respectively 228

List of Tables

| | | |
|-----|---|-----|
| 3.1 | The allocation of months to seasons | 39 |
| 3.2 | CP Occurrence frequencies and wave height statistics associated with each CP class. | 47 |
| 3.3 | Six of the most extreme wave events on record and their associated CPs for the period 1992 to 2009. | 52 |
| 4.1 | The allocation of months to seasons | 60 |
| 4.2 | The occurrence statistics and associated wave height statistics for the CP shown in Fig 4.2 after (Pringle <i>et al.</i> , 2014). | 66 |
| 4.3 | Relative differences (%) to key statistics when the CP classification was run for 1-day and 6-hour temporal resolutions and with 16 CP classes. Note that the labels of the CP classes as CP01, CP02, etc is arbitrary, but CP99 denotes an unclassified class. The wave height threshold θ used to define the extreme wave events was $\theta = 3.5$ m for this analysis. | 69 |
| 4.4 | A matrix of cross-correlation coefficients between the different classification schemes showing the persistence of a dominant class contained in all classifications. The classification was not run for 2, 4, 6 and 7 classes. The table is symmetric about the central diagonal. | 70 |
| 5.1 | Historical wave recording instruments at Durban and Richards Bay, their operating periods and water depth (from Corbella & Stretch (2014b)) | 82 |
| 5.2 | The allocation of months to seasons | 83 |
| 5.3 | CP occurrence frequencies and their associated wave event statistics. The wave height threshold θ used to define the wave events was $\theta = 3.5$ m for this data. | 99 |
| 6.1 | The allocation of months to seasons | 106 |

List of Tables

| | | |
|-----|---|-----|
| 6.2 | Seasonal wave climate statistics between the Simulated and Observed data. Standard errors for the sampled statistics (one standard deviation) are shown in brackets. | 119 |
| 7.1 | The allocation of months to seasons | 137 |
| 7.2 | Average longshore transport for observed and simulated wave data. The approximate range between 5 th and 95 th percentiles calculated from the simulated data are also indicated. | 146 |
| 8.1 | The allocation of months to seasons | 159 |
| 8.2 | Contingency table of the two classifications. The rows contain CP classes derived from the observed wave climate. The columns are classes derived from reanalysed wave data. Values in the table are normalised by the number of cases in each column (%) | 169 |
| 8.3 | Key CP statistics using the ERA reanalysis wave data to delineate the classes for the period 1992–2009. Comparisons with classifications using observed wave data to delineate classes are shown in brackets. | 170 |
| 8.4 | Maximum and minimum estimated differences in CP occurrence frequencies between historical and HadGEM2-ES GCM predictions. The table also shows the differences between the near future and distant future as predicted by the HadGEM2-ES GCM for two scenarios. The long term changes are calculated as the difference between CP occurrence frequencies in near future (2010-2050) and distant future (2050-2100). | 173 |
| 8.5 | Average statistics for important wave variables (H_s, T_p, θ) obtained from the stochastic wave simulation for the near and distant future based on two future scenarios. | 175 |
| A.3 | CP Occurrence frequencies and wave height statistics associated with each CP class. | 215 |
| A.4 | Six of the most extreme wave events on record and their associated CPs for the period 1992 to 2009. | 217 |

Chapter 1

Introduction

1.1 Coastal Vulnerability

The coastline is a dynamic environment subject to continual erosion and accretion with small net changes that when accumulated over many years result in significant morphological change (Cowell *et al.*, 2003). While coastal erosion is a natural process damage to infrastructure near the coastline has severe socio-economic, environmental and physical impacts (Prinos & Galiatsatou, 2010). However tourist attractions and ease of living drive high population densities and economic development (Palmer *et al.*, 2011; Prinos & Galiatsatou, 2010; Stive *et al.*, 2010). According to the United Nations approximately 50% of the worlds population live within 60 km of the coast and this figure is set to rise by an amount of 75% by the year 2020 (UN, 1993). The proximity of development to the shoreline intensifies the effect of erosion and leaves it vulnerable to future wave attack (Palmer *et al.*, 2011; Pethick, 2001). Therefore the socio-economic and physical parameters that drive coastal vulnerability create a complex environment from which to quantify the associated risk. This study focusses primarily on regional wave climates as a physical driver of coastal vulnerability.

The general responses of local authorities to coastal vulnerability are to accommodate the changes and retreat, defend (or protect the coast) or do nothing ((IPCC), 1992). Coastal defence is an attractive mitigation strategy that can protect a stretch of beach against future severe erosion events. This safeguards infrastructure and tourist attractions while preventing relocation. Typically defence strategies consist of either hard or soft engineering or a combination of the two. Soft engineering comprises predominantly of nourishment strategies whereas hard engineering includes the design of breakwaters and groynes etc. The success of such strategies depends largely upon

Introduction

the risk of failure during their design lifetime. However in a multivariate environment assessing the risk of failure of a design is difficult. It is well known that the combined effect of the wave climate, sediment supply, inter-arrival times and durations of large wave events as well as sea level rise are fundamental drivers of coastal erosion (Callaghan *et al.*, 2008; Corbella & Stretch, 2012c). Therefore risk assessment and the evaluation of extreme events depends on different combinations of the aforementioned parameters. The process is complicated further when considering that the wave climate consists of the wave height, wave direction and wave period. Therefore it is not necessarily only the 100 year wave height that can cause the most severe shoreline erosion or damage to coastal defence structures. In contrast it is the combination of the above parameters in such a way that they produce the worst case scenario. For example it has been shown that multiple storms following in quick succession can have a higher erosive potential than a single storm with a large wave height (Karunarathna *et al.*, 2013). This is primarily a function of the beach recovery time and antecedent beach conditions. Higher storm frequencies prevent the natural recovery of a beach if the preceding storm has not eroded to the equilibrium beach profile of the following storm (Corbella & Stretch, 2012c). The reduction of the buffer zone between development and the sea allows wave attack to penetrate further than the preceding storm (Forbes *et al.*, 2004). A storm event is defined as the time from which the wave heights exceeds a certain threshold to the time it falls below this threshold. The storm threshold is based on observations of wave heights that cause severe coastal erosion (Corbella & Stretch, 2012c).

This study proposes a new approach towards risk assessment. The technique is based on atmospheric circulation patterns (CPs) that drive regional wave climates. The location, shape and persistence of CPs provide information on characteristics such as the wave height, direction, period and storm duration. Furthermore they provide a physically based mechanism to stochastically simulate regional wave climates. The transitions between CPs define different states within the wave record such as periods of calm conditions or stormy conditions. The advantages of stochastically simulating synthetic wave records in this manner are:

- The technique is based on the physical drivers of regional wave climates.
- It is not computationally demanding to simulate long wave sequences.
- Extreme events are well captured and wave.

- The multivariate complexities required to simulate a wave sequence are simplified by the occurrence of specific CPs.
- Changes in wave behaviour due to climate change effects are easily quantified.
- The technique can be applied to any region improving coastal risk assessments.

1.2 Problem Definition

Quantifying risk within a multivariate environment such as regional wave climates is complex. Physical links between variables are difficult to describe as they involve processes over a range of spatial and temporal scales. However regional wave climates are strongly linked to atmospheric circulation patterns. Therefore if these links can be clarified then they can be used to provide a simplified means to quantify risk.

1.2.1 Research Question

Can statistical links between atmospheric circulation patterns and regional wave climates be utilized to simulate continuous wave sequences and evaluate coastal vulnerability within the context of climate change?

The focus of this study is on the wave climate along the KwaZulu Natal coastline. The research question can be divided into the following sub-questions:

1. Can a suitable classification algorithm be developed that captures the links between atmospheric circulation patterns and regional wave climates?
2. What are the atmospheric circulation patterns that drive extreme wave events along the KwaZulu Natal coastline?
3. Can the performance of the classification algorithm be evaluated so as to identify the optimal configuration of input parameters?
4. Do the CPs describe the dependence structures between wave climate variables well and can they be used to stochastically simulate regional wave climates?
5. Can the algorithm be applied to regions that do not have observed wave data?
6. Can the link between CPs and waves be exploited to evaluate future wave scenarios due to climate change effects?

1.3 Motivation

The KwaZulu Natal coastline on the east coast of South Africa has a high energy wave climate and there are a number of weather systems driving wave development. Cold fronts associated with mid latitude cyclones, cut-off lows and tropical cyclones are cited as the three main physical drivers of the KwaZulu Natal wave climate (Corbella & Stretch, 2012d; Mather & Stretch, 2012; Preston-Whyte & Tyson, 1988; Rossouw *et al.*, 2011; Taljaard, 1967). Each of the aforementioned weather systems are associated with different wave characteristics. For example cold fronts are associated with long period waves from a southerly direction whereas cut-off lows can drive waves towards the coastline from a southeast-east direction (Preston-Whyte & Tyson, 1988; Rossouw *et al.*, 2011). This complexity makes it difficult to quantify coastal vulnerability and assess risk. This is because it is difficult to delineate the source of the incoming wave energy. Furthermore wave observations along the KwaZulu Natal coastline are limited to 18 years (1992–2009) and directional wave data are only available from 2007. This limits return period analysis of storm events.

Recent advances in coastal risk assessment studies have been to incorporate multivariate statistics and the inter-dependence structures between variables (Corbella & Stretch, 2013; De Michele *et al.*, 2007; Li *et al.*, 2013). However these statistical models are not fundamentally based on the drivers of wave climates namely the synoptic scale circulation patterns (CPs). Therefore the inter-dependence structures between wave variables have no direct links to the physical drivers. However if the links between CPs and regional wave climates can be clarified they can be used to simplify coastal vulnerability assessment and simulate synthetic wave records with characteristics similar to the observed data. CPs provide a natural way to define storm events. Their transition between states inherently defines storm duration and their shape and intensity control the strength and direction of the event. It is this combination of wave height, period, direction, storm duration and the storm inter-arrival time that can lead to significant shoreline erosion and subsequent coastal vulnerability concerns.

1.3.1 Risk Assessment

Within the context of climate change and non-stationary statistics it has become evermore important to accurately quantify coastal erosion and its associated risk (Callaghan *et al.*, 2008; Corbella & Stretch, 2012b). Recent advances in risk assessment studies have highlighted the complexity of the problem and provide manners in

which to evaluate coastal risk. Previous methods utilized a benchmark approach in which the largest wave height on record was chosen for design criteria (Callaghan *et al.*, 2008; Isaacson & MacKenzie, 1981). However this practice is flawed and provides little insight into complex environments (Callaghan *et al.*, 2008; Corbella & Stretch, 2013). Coastal erosion depends on a number of factors such as the wave height, direction, storm duration and water level. In this complex multivariate environment the return period of an event forced by two or more variables differs significantly from that forced by a single variate (Callaghan *et al.*, 2008; Corbella & Stretch, 2012b; Hawkes *et al.*, 2002; Li *et al.*, 2013). For example a 50 year wave height will not result in a 50 year beach erosion event (Callaghan *et al.*, 2008). Therefore a simple univariate approach is unable to accurately quantify the effects of a sea storm. However a multivariate approach although not as simple as univariate lends itself to being more attractive. Various multivariate models have been used. For example Corbella & Stretch (2013); De Michele *et al.* (2007); Li *et al.* (2013) utilized copula methods whereas Callaghan *et al.* (2008) incorporated a joint probability to quantify risk. The various models were shown to be effective tools for coastal risk assessment. However it is difficult to quantify the sources of wave energy in complex environments without direct links to the physical drivers. This can lead to inaccurate descriptions of the dependence structures between wave climate variables.

It is well known that the drivers of regional wave climates are atmospheric circulation patterns (CPs). Therefore a technique able to incorporate the links between CPs and regional wave climates will provide a physically meaningful risk assessment study. Furthermore wave data associated with CPs can be used to extend current observations and simulate synthetic wave climates of any given length with characteristics similar to the observed. The CPs also provide the framework to assess future wave climate scenarios due to climate change effects.

1.3.2 Global Wave Models

The use of global wave hindcast datasets from wave modelling centres such as NCEP (National Centre for Environmental Prediction) or ECMWF (European Centre for Medium-Range Weather Forecast) is another approach to evaluate coastal risk. The datasets span approximately 30–40 years and can be used to extend or replace wave observations. However the datasets rely strongly upon data reanalysis and accurate wind field forcing (Caires *et al.*, 2004; Chawla *et al.*, 2013; Tolman *et al.*, 2002). Accurate wave prediction requires near perfect surface wind field input that are generated

by a global climate model (GCM) (Swail & Cox, 1999). The coarse grid of the GCM reduces the models' ability to accurately describe compact weather systems. Therefore the models tend to under predict extreme wave heights nearshore (Chawla *et al.*, 2013; Swail & Cox, 1999; Tolman *et al.*, 2002). Furthermore Chawla *et al.* (2013); Stopa & Cheung (2014) suggest the use of either dataset is application specific. For example the NCEP dataset describes wave variability well but is not suited to long term modelling, however the ERA dataset is better suited for long term modelling. The ability for a model to describe extreme events well is fundamentally important for risk assessment studies. This study proposes that the links between CPs and observed wave data provide a physically based solution to the aforementioned problems.

1.4 Aims and Objectives

Aim:

Evaluate the statistical links between atmospheric circulation patterns and the KwaZulu Natal wave climate so as to improve current risk assessment studies and predict future wave climate scenarios.

Objectives:

1. Develop a statistical model that links atmospheric circulation patterns and regional wave climates.
2. Determine the atmospheric drivers of extreme events along the KwaZulu Natal coastline.
3. Develop a suitable method to evaluate model performance.
4. Stochastically simulate continuous wave sequences conditioned on the CP occurrences.
5. Test the applicability of the model to regions with no observed wave data.
6. Demonstrate the advantage of this simulation technique to assess future wave climates due to climate change effects.

1.5 Approach

The east coast of South Africa is used as a case study. This is because the coastline is frequented with large waves from several different weather sources. Waverider buoys situated off the Durban coastline provide an 18 year wave observation record. This data was used within a classification algorithm to delineate links between CPs and the regions' wave climate.

The atmospheric circulation patterns that drive the KwaZulu Natal coastline were clarified using an objective fuzzy rule based classification algorithm. The algorithm optimizes the shape and orientation of high and low pressure regions of a set of CPs that explain specific aspects of the wave climate. The algorithm explores the direct links between synoptic scale circulation and regional wave climates. Wave climate statistics and the relationships between different variables conditioned on the CPs were used to simulate continuous synthetic wave records.

The coastline also has 37 years of beach profile data that was used to evaluate the performance of the wave climate simulation model. A validation exercise was conducted to evaluate different characteristics and properties of the simulated waves such as the wave height return periods as well as longshore and cross shore beach responses. A new shoreline evolution model was used to demonstrate the feasibility of simulating continuous wave sequences.

1.6 Thesis Structure

This thesis continues with a literature review of the main subject areas. This is followed by 6 chapters that are reproductions of papers. The papers are in original form and are explicitly linked. Key results of all the papers are related in Chapter 9: Synthesis and Conclusions. The appendices A - C contain additional information relevant to the study area. The thesis structure is as follows:

Chapter 2 reviews aspects pertinent to the main study areas in the form of a literature review.

Chapter 3 presents the development of a new wave based atmospheric classification algorithm.

Chapter 4 evaluates the sensitivity of the algorithm to various input parameters specifically the number of classes and temporal resolution of the data.

Introduction

Chapter 5 demonstrates the ability of the classification algorithm to delineate between the circulation pattern types that drive complex wave climates. The work presented in this chapter formed part of a collaborative effort of which the classification algorithm was fundamentally important. The authors' contribution is as follows:

- The CP classification algorithm was used to classify the circulation types associated with the different swell origins.
- The classes derived from the classification algorithm were used to explore the relationship between CPs and wave directional spectra.
- The author was significantly involved in writing the research paper.

Chapter 6 presents a new stochastic wave simulating technique that is from the deduced atmospheric circulation patterns.

Chapter 7 presents a validation exercise for the modelling technique described in chapter 6. Furthermore it demonstrates the advantage of simulation continuous wave sequences for coastal vulnerability assessments.

Chapter 8 highlights how the classification algorithm can be applied to regions with little to no wave observations. The chapter also demonstrates the advantage of the CP-Wave simulation technique in future wave climate estimation due to climate change effects.

Chapter 9 relates the key results from the previous chapters, presents conclusions and discusses future work.

Appendix A presents the effectiveness of the classification algorithm in explicitly describing the atmospheric drivers of specific physical processes. The authors' contribution was as follows:

- The results of the classification method developed in Chapter 3 were used to compare different classification regimes.
- The author wrote the section on linking waves to CPs and the discussion thereof.

Appendix B highlights properties of the circulation pattern regimes that drive wave development for the east coast of South Africa.

Appendix C presents the classification algorithm in pseudo code format.

Chapter 2

Literature Review

2.1 Coastal Vulnerability

Within the context of climate change and non-stationary statistics it has become evermore important to accurately quantify coastal vulnerability and associated risk (Callaghan *et al.*, 2008; Corbella & Stretch, 2012b). Recent advances in risk assessment studies have highlighted the complexity of the problem and provide means by which to evaluate coastal risk. With respect to coastal erosion, previous methods utilized a benchmark approach in which the largest wave height on record was chosen for design criteria (Callaghan *et al.*, 2008; Isaacson & MacKenzie, 1981). However this practice is flawed as it provides little insight into complex environments (Callaghan *et al.*, 2008; Corbella & Stretch, 2013). Coastal erosion depends largely on a number of factors such as the wave height, direction, storm duration and water level. In general the larger/longer the aforementioned values the greater the erosive potential of the storm. Furthermore storms with different characteristics can have the same erosive potential. For example a storm with a large wave height and short period can result in the same amount of erosion as a storm with a small wave height and longer period (Corbella & Stretch, 2012b; De Michele *et al.*, 2007). In this complex multivariate environment the return period of an event forced by two or more variables differs significantly from that forced by a single variate (Callaghan *et al.*, 2008; Corbella & Stretch, 2012b; Hawkes *et al.*, 2002; Li *et al.*, 2013). For example a 50 year wave height will not result in a 50 year beach erosion event (Callaghan *et al.*, 2008). Therefore a simple univariate approach is unable to accurately quantify the effects of a sea storm. However a multivariate approach although not as simple as univariate lends itself to being more attractive. The general approach to evaluate coastal erosion is to simulate

wave climates and water levels as a Markov process and quantify the associated erosion using a structure function that links the wave climate with an associated nearshore process such as cross shore erosion (e.g. Callaghan *et al.*, 2008; Corbella & Stretch, 2012b; De Michele *et al.*, 2007; Guedes Soares & Cunha, 2000).

2.1.1 Statistical Wave models

Long records of observed wave data are fundamental to accurately quantify coastal vulnerability. A long record of wave observations is beneficial because it contains important information on possible wave states, storm inter-arrival times and their temporal grouping. However the availability of wave data is limited to observations from wave buoys, ships and satellites and is focussed primarily in the Northern Hemisphere (Caires *et al.*, 2004). Wave data from buoys are region specific, they only provide information at one location and record lengths for a few locations span 30-40 years (see e.g. Caires *et al.*, 2004; Callaghan *et al.*, 2008; Gilhousen, 1999). With the onset of satellite technology it is possible to obtain global wave fields however it is not possible to monitor the ocean everywhere at the same time (Cotton & Carter, 1994; Monbet *et al.*, 2007). The quality and quantity of wave observations limits their application in risk assessment and evaluation of future wave climate scenarios. Therefore an approach to simulate long wave records to either extend or replace observed data is favourable. Process based approaches are limited in their applicability because they are computationally demanding because they describe processes over a range of temporal and spatial scales. However statistical models provide a means to address this issue.

Statistical techniques to simulate waves have gained significant attention and several techniques have been explored (e.g. Callaghan *et al.*, 2008; Corbella & Stretch, 2012b; Guedes Soares & Ferreira, 1996; Monbet *et al.*, 2007; Solari & Losada, 2011). The techniques include auto-regressive simulation (Guedes Soares & Ferreira, 1996), joint probability (Callaghan *et al.*, 2008) and copulas (Corbella & Stretch, 2012b; De Michele *et al.*, 2007; Li *et al.*, 2013). Synthetic wave records derived from these techniques are modelled as a Monte Carlo process to evaluate all possible wave states. This is advantageous because the models describe the average conditions as well as the expected variability both of which are strategically useful in risk assessment studies (Guedes Soares & Cunha, 2000). There are currently two approaches with regard to simulating wave time series. The first is to simulate a sequence of storms with specific durations and inter-arrival times and the second is to simulate complete temporal se-

quences of waves (Solari & Losada, 2011). Any coastal process such as erosion can easily be incorporated with the use of a suitable structure function (e.g. Callaghan *et al.*, 2008; Corbella & Stretch, 2012b).

Autoregressive Models

A time series involving a random process can be described as a linear combination of the output variable at previous times and an error term such that

$$X_t = k + \sum_{i=1}^p \phi_i X_{t-i} + \epsilon_t \quad (2.1)$$

where X_t is the output variable at time t , ϕ_i is a constant of order i , k is a constant and ϵ is a normally distributed error term. Equation 2.1 is referred to as an autoregressive model and it is a special case of an autoregressive moving average model of order p i.e. ARMA($p, 0$) (Box & Jenkins, 1976). The autoregressive moving average model is defined as

$$X_t = k + \epsilon_t + \sum_{i=1}^p \phi_i X_{t-i} + \sum_{i=1}^q \theta_i \epsilon_{t-i} \quad (2.2)$$

where θ_i are constants of order q . Equation 2.2 describes an autoregressive moving average process of order p and q i.e. ARMA(p, q).

The application of auto-regressive techniques to wave climate simulation provides a simple means to generate synthetic wave records. Guedes Soares & Ferreira (1996) showed that the advantage of this technique is that it is able to infill missing data and extend the current observation record. However the approach of Guedes Soares & Ferreira (1996) was to only simulate wave height. Although this provides little insight into complex wave climates their technique represented the first attempt to statistically simulate long-term wave sequences. To address this Guedes Soares & Cunha (2000) developed a bivariate autoregressive model for simulating wave height and period.

Auto-Regressive (AR) or Auto-Regressive Moving Average (ARMA) techniques require Gaussian and stationary statistics (Box & Cox, 1964). However the statistics associated with wave climates are not stationary nor Gaussian (Guedes Soares & Ferreira, 1996). Intuitively different seasons are associated with different weather systems that drive different wave states. Therefore the Box-Cox transformations must first be applied to the dataset to normalize the variance (Box & Cox, 1964). The

general Box-Cox transformations are

$$f_{\lambda}(X_t) = \begin{cases} \lambda^{-1}(X_t^{\lambda} - 1), & X_t \geq 0, \lambda > 0 \\ \ln X_t, & X_t > 0, \lambda = 0 \end{cases} \quad (2.3)$$

where X_t is the random variable of interest at time t and λ is a power parameter used for the transformation. Guedes Soares & Ferreira (1996) found that this transformation stabilized the variance but the mean was still subject to seasonality. To address this issue they applied a second transformation

$$Y_t = \ln(X_t) - m_t \quad (2.4)$$

where m_t was an estimation of the monthly mean. Additional to the data transformations Guedes Soares & Ferreira (1996) found that gaps in the dataset were difficult to deal with. To address this issue they analysed the data in a piecewise manner with subseries that contained no gaps.

A fundamental weakness of all ARMA models is that they are not well suited to simulate data that exhibit asymmetric properties such as tail dependence (Scotto & Guedes Soares, 2000). Scotto & Guedes Soares (2000) suggest the use of non-linear autoregressive models to handle asymmetric distributions however found little improvement over a standard AR model.

Multivariate Approaches

A multivariate simulation of wave climate variables such as wave height (H), wave period (T) and wave direction (θ) provides significantly more insight into the characteristics of regional wave climates than a univariate approach. Advances in multivariate simulation techniques have seen the development of several different approaches such as bivariate autoregressive models, joint probability methods and most recently the use of copulas (see Callaghan *et al.*, 2008; Corbella & Stretch, 2012b; De Michele *et al.*, 2007; Guedes Soares & Cunha, 2000).

In a bivariate case it is possible to simulate pairs of values based on the bivariate conditional distributions. For example Monbet & Prevosto (2000) transformed the empirical marginal distributions of wave height and period to Gaussian such that

$$Z_i = \phi^{-1}(F_x(X_i)), \forall i \quad (2.5)$$

where Z_i is the Gaussian variate, ϕ is the Gaussian distribution, F_x is the marginal distribution for variable X_i at the i^{th} observation in year j (wave height or period). The wave height and period were then simulated based on the Bayes formula as described in Monbet & Prevosto (2000) as

$$\begin{aligned}\bar{H}_s^{st}(i, j) &= F_{H^{st}}^{(i)-1}(\phi(\bar{Z}_H(i, j))) \\ \bar{T}_p^{st}(i, j) &= F_{T^{st}|H^{st}}^{(i)-1}(\phi(\bar{Z}_T(i, j)|H_s^{st} = \bar{H}_s^{st}(i, j)))\end{aligned}\tag{2.6}$$

where \bar{H}_s^{st} and \bar{T}_p^{st} are the simulated stationary wave height and period, \bar{Z}_H and \bar{Z}_T are the Gaussian variates obtained after removing the mean and variance for the wave height and period respectively. The assumption that wave heights and periods estimated from Equation 2.6 are stationary is not strictly valid. However Monbet & Prevosto (2000) argue that the estimated joint probability distribution of H_s and T_p is stationary over a small time interval of the observations around t . Therefore empirical distributions were based on observations on a small time interval around the reference time. However their method provides little information on the temporal structure of the wave heights, storm durations and their inter-arrival times.

In their study Callaghan *et al.* (2008) utilized the joint probability (JPM) between variates to simulate synthetic wave records so as to evaluate coastal erosion. The method focussed on simulating individual storms as a Markov process with the storm inter-arrival time modelled as a Poisson process. Wave parameters were simulated in a pairwise manner using conditional bivariate distributions of the dependency between wave height and storm duration, wave period and wave height, tidal anomaly and wave height. The wave direction was simulated using the empirical distribution. Extreme value distribution functions were fitted for each of the variables simulated because their technique focussed on simulating storm events. Storm events were defined as the time from which the wave height exceeded 3 m to the time it fell below 3 m. Coastal erosion was estimated with a structural function that linked the wave climate to the cross shore beach response. The time convolution model after Kriebel & Dean (1993) was used to estimate the beach response due to wave forcing. Physical links to storms were not included and independent events had to be manually assessed.

Copulas

A prevalent factor resulting in the success of all the aforementioned statistical wave models is their ability to describe the dependence structures that exist between vari-

Literature Review

ables and the manner in which this is done. All the aforementioned models have only focussed on simulating pairs of values based on the bivariate distribution functions for each pair. However extending this to higher dimensions is difficult. For example Callaghan *et al.* (2008) had to simulate several pairs of variables in order to accurately describe the multivariate environment. Intuitively the variables typically used to characterize regional wave climates such as wave height, direction and period are interdependent. This dependence can be attributed to the meteorological systems that drive wave development. However Callaghan *et al.* (2008) found that the wave direction was weakly related to wave height observed wave data in their study. Copulas are a mathematically tractable approach to handle inter-dependence in multivariate environments.

A copula can be defined as function that couples multivariate distributions to their individual marginal distribution functions (Nelson, 2006). The advantage of exploiting copulas in multivariate modelling is that the dependence structure between variables is completely described by the copula independently of the individual marginal distribution functions. For a set of continuous random variables y_1, \dots, y_n with marginal distribution functions $F(y_1), \dots, F(y_n)$ their joint cumulative multivariate distribution function $H(y_1, \dots, y_n)$ is described by the copula $C\{\cdot\}$ such that (Sklar, 1959)

$$H(y_1, \dots, y_n) = C\{F(y_1), \dots, F(y_n)\}. \quad (2.7)$$

Equation 2.7 is formally referred to as Sklar's Theorem and it has strong implications for multivariate modelling (Nelson, 2006). Sklar's Theorem provides the advantage of selecting the appropriate dependence model (defined by the copula) independently from the choice of the individual marginal distribution functions (Genest & Favre, 2007). Copulas describe the dependence between variables in the following manner as described in Genest & Favre (2007); Nelson (2006):

Considering the 2-dimensional case for simplicity. For a pair of random continuous variables (X, Y) , if X and Y are independent then their joint distribution is defined by the product copula $C = \Pi$, with $\Pi(u, v) = uv \forall u, v \in I$, where $I = [0, 1]$. Furthermore if X and Y are interdependent then the copula defining their joint distribution must lie within the *Fréchet-Hoeffding* bounds:

$$W(u, v) = \max(u + v - 1, 0) \text{ and } M(u, v) = \min(u, v).$$

Thus for every copula C and (u, v) pair in the unit square $I^2 = I \times I$,

$$W(u, v) \leq C(u, v) \leq M(u, v).$$

The application of copulas to multivariate simulation is achieved using the conditional inversion method (e.g. Corbella & Stretch, 2012b; De Michele *et al.*, 2007; Genest & Favre, 2007; Nelson, 2006). It is possible to generate a vector (U_1, \dots, U_n) of uniform values where $U_n \in [0, 1] \forall n$ interdependent variables with copula $C = C(u_1, \dots, u_n)$, where $u_n = F_n(y_n)$ from the conditional distribution

$$\begin{aligned} C_n(u_n|u_1, \dots, u_{n-1}) &= P(U_n \leq u_n | U_1 = u_1, \dots, U_{n-1} = u_{n-1}) \\ &= \frac{\partial_{u_1, \dots, u_{n-1}} C(u_1, \dots, u_n)}{\partial_{u_1, \dots, u_{n-1}} C(u_1, \dots, u_{n-1})} \end{aligned} \quad (2.8)$$

by setting $U_n = C_n^{-1}(q_n|u_1, \dots, u_{n-1})$ where q_n is a random sample from the uniform distribution on the interval $(0,1)$. It is not always possible to explicitly solve Equation 2.8 and therefore numerical methods must be utilized.

With respect to simulating waves, Archimedian copulas have been exploited because of their ease of use (e.g. Corbella & Stretch, 2012b; De Michele *et al.*, 2007). Nelson (2006) describes Archimedian copulas, their construction and application. They are a special class of copulas that are the solution to the functional equation (Nelson, 2006)

$$\phi(C(u, v)) = \phi(u) + \phi(v) \quad (2.9)$$

where ϕ is the generator of the copula. The generator function is a continuous decreasing convex function in the interval I to $[0, \infty]$. The advantage of using Archimedian copulas is that they are defined by a single dependence parameter, θ , that is related to the Kendall's tau statistic (τ_k) through

$$\tau_k = 1 + 4 \int_0^1 \frac{\phi(\theta, t)}{\phi'(\theta, t)} dt \quad (2.10)$$

Risk assessment studies that incorporate copulas to model interdependence between variables have recently become an attractive approach to quantify the effect of storms on coastal risk (Corbella & Stretch, 2012b, 2013; De Michele *et al.*, 2007; Li *et al.*, 2013, 2014). For example Corbella & Stretch (2012b, 2013); De Michele *et al.* (2007) used Archimedian copulas to model the inter-dependencies between the wave height (H), wave period (T), wave direction (A), storm duration (D) and storm inter-arrival time (I). However Li *et al.* (2013, 2014) used a Gaussian copula to model

similar interdependencies of wave climate variables. The gaussian copula suffers the drawbacks of being symmetrical and it has weak tail dependence which is not well suited for modelling extreme events.

Limitations of the Statistical Models

The limitations of the above mentioned statistical models are well documented. For example Corbella & Stretch (2013) highlight the significance that their copula model is not linked to physical meteorological forcing. Therefore identifying independent events and delineating wave sources in complex environments is difficult. Currently independent events are delineated using a time threshold or they are manually assessed (e.g. Callaghan *et al.*, 2008; Corbella & Stretch, 2012b; Li *et al.*, 2013). This can lead to errors when evaluating the dependence between variables. For example Callaghan *et al.* (2008) found no relationship between wave direction and wave height. However weather systems that drive wave development attribute different characteristics to wave climate variables. This implies that wave height, direction and period are interdependent. For example along the east coast of South Africa mid-latitude cyclones associated with cold fronts drive long period waves from a southerly to south westerly direction. In contrast deep cutoff lows drive large waves from a south-easterly to easterly direction. Furthermore the persistence of a particular system in a specific region has significantly different storm characteristics than a system that moves past in a short duration. If links between synoptic scale circulation and regional wave climates can be delineated they can be exploited within risk assessment studies.

2.1.2 Global Wave Hindcast Data

An additional approach to simulating waves is to use existing wave hindcast datasets available from wave modelling centres such as NCEP (National Centre for Environmental Prediction) or ECMWF (European Centre for Medium-Range Weather Forecasting). The hindcast wave data is simulated using global wave models. The NCEP dataset is derived using the WAVEWATCH III (WWIII) model. The WWIII model numerically solves a linear balance equation for the spectral wave action density A given as

$$\frac{DA(k, \theta; \mathbf{x}, t)}{Dt} = S(k, \theta; \mathbf{x}, t) \quad (2.11)$$

where $k = \frac{2\pi}{L}$ is the wave number, L is the wave length, θ is the wave direction, \mathbf{x} is the vector containing all locations and t is the time. The right side of the equation represents the nonconservative source and sink terms pertaining to wave development. Examples of such terms given in Tolman *et al.* (2002) are wind input, shallow water processes and dissipation due to breaking etc. The ECMWF modelling group utilize the WAVE Model referred to as WAM (WAAMDIG, 1988). The model is also a third generation wave model and based on Equation 2.11 (Stopa & Cheung, 2014). Model differences are attributed to differences in numerical solution techniques and differences in approximations of wave physics (Tolman *et al.*, 2002).

Wave hindcast datasets span 30-40 years and are significantly useful in regions with little or no wave data. However the datasets suffer several drawbacks. Firstly they rely strongly upon accurate wind field forcing and data reanalysis (Swail & Cox, 1999). Therefore their use for estimating future wave climate scenarios is compromised. Secondly extreme wave events are not well defined specifically within the nearshore regions. Furthermore wave climate estimation nearshore is limited because of complex wind and bathymetric effects (Caires *et al.*, 2004; Chawla *et al.*, 2013; Stopa & Cheung, 2014; Swail & Cox, 1999; Tolman *et al.*, 2002). When comparing wind and wave data from ECMWF and NCEP Stopa & Cheung (2014) found that NCEP data better describe the variability and upper percentiles of wave observations whereas ECMWF data was more homogeneous through time and therefore suited for long term modelling. Therefore care must be taken when selecting which model to use especially when considering the purpose of model application.

2.1.3 Shoreline Response Models

Fundamental to successful coastal vulnerability assessments are accurate estimations of nearshore processes driven by regional wave climates. These processes include cross shore or longshore beach responses to wave forcing. Another important property of the shoreline response models is that they should be easily incorporated into long records of either simulated or observed wave data.

Longshore Transport

In regards to estimating longshore transport numerous empirical formulae exist. However in their study Schoonees & Theron (1996) found the Kamphuis formula (Kamphuis, 1991) to be the most universally applicable. The advantage of the Kamphuis

formula is that it describes the empirical relationship between wave height, direction, period and longshore transport. It is given as

$$Q = 6.4 \times 10^4 H_{sb}^2 T_p^{1.5} m^{0.75} d^{-0.25} \sin^{0.6}(2\theta_b) \quad (2.12)$$

where Q is the longshore transport (m^3/yr), H_{sb} , T_p and θ_b are the significant wave height, period and direction at breaking respectively, m is the beach slope and d is the water depth. Other longshore transport formulae exist and another common approach is to use the CERC formula (CERC, 1984a)

$$Q = \frac{K_l}{\gamma^{1/16}} \rho g^{1.5} H_{sb}^{2.2} \sin(2\theta_b) \quad (2.13)$$

where γ is the breaker index, ρ is the density of water, g is the gravitational constant and K_l is an empirical constant and ranges between 0.2-0.39 (CERC, 1984a; Schoonees & Theron, 1993).

Cross-shore Beach Response

Currently there are three approaches to estimating cross shore processes within a probabilistic wave environment: (a) analytically, (b) semi-empirically and (c) process based (e.g. Callaghan *et al.*, 2013, 2008; Corbella & Stretch, 2013). These methods are incorporated within the probabilistic framework with the use of a structure function to evaluate the desired process, for example coastal erosion. Furthermore it is expected that a more sophisticated model improves the assessment. However when evaluating cross shore erosion Callaghan *et al.* (2013) found this to not necessarily be the case.

Analytical approaches attempt to solve the shoreline response equation

$$\frac{dy(t)}{dt} = k(y_{eq}(t) - y(t)) \quad (2.14)$$

which governs shoreline response as an exponential process in which the current shoreline position ($y(t)$) approaches equilibrium ($y_{eq}(t)$) under steady-state condition (Miller & Dean, 2004). Kriebel & Dean (1993) solved Equation 2.14 using convolution such that

$$y(t) = k \int_0^t y_{eq}(\tau) e^{-k(t-\tau)} d\tau + y(0). \quad (2.15)$$

However this approach is limited because it requires an idealized forcing function ($y_{eq}(t)$). Therefore applying the beach response model for continuous wave sequences

is difficult. To address this Miller & Dean (2004) numerically solved Equation 2.14. Yates *et al.* (2009) extended the approach of Miller & Dean (2004) by relating the shoreline position to the incoming wave energy so that

$$\frac{dS}{dt} = C^\pm E^{1/2}(E - E_{eq}(S)) \quad (2.16)$$

where C^\pm denotes change rate coefficients for accretion(C^+) and erosion (C^-) and the equilibrium energy ($E_{eq}(S)$) is linearly related to the shoreline position S . In an attempt to simplify Equation 2.16, Jara *et al.* (2015) removed the weighting factor $E^{1/2}$ and suggested expressing $E_{eq}(S)$ as a quadratic function of the shoreline position. In addition to relying on the wave disequilibrium their approach included a static equilibrium model based on a bi-parabolic equilibrium beach profile.

Process based or semi-empirical approaches such as XBEACH (Roelvink *et al.*, 2009) or SBEACH (Larson *et al.*, 1990) are a more sophisticated approach to estimate cross-shore beach response. However their inclusion within a probabilistic framework is computationally demanding. For example applying SBEACH directly to a 1000-year long (repeated 2000 times) probabilistic simulation of storm events took approximately 40 days to run on one processor (Callaghan *et al.*, 2013). The XBEACH model is significantly more numerically intensive than SBEACH and therefore cannot directly be used in a probabilistic framework. To address this Callaghan *et al.* (2013) linearly interpolated erosion values from a pre-run structured grid of beach erosion values using XBEACH.

2.2 Atmospheric Classification

Classification can be broadly defined as the technique in which similar cases are grouped together in such a way that they share common properties while remaining distinctly different between cases of another group (Huth *et al.*, 2008). Such a method provides a simple manner in which to evaluate complex inter-relationships without the need to explicitly derive them. However classification techniques themselves are not simple.

It is well known that synoptic scale atmospheric circulation drives local meteorological variables such as precipitation, temperature and herein regional wave climates (Bárdossy, 2010). These links are complex and difficult to describe in a simplified manner. However classification can be used to gain insight into these relationships.

Classification can be conducted either in a supervised or unsupervised manner however the goals remain the same to group variables that share similar properties together (Bárdossy, 2010). Examples of unsupervised techniques include k-means clustering, self organizing maps (SOMs) and principle component analysis (PCA) (e.g. Hewitson & Crane, 2002; Huth *et al.*, 2008; Kohonen, 1990). Atmospheric classification is usually applied to gridded pressure values at sea level or a specific geopotential height for a specified region.

2.2.1 Unsupervised Classification Techniques

Unsupervised classification utilizes self learning techniques to derive classes that represent different states in the variable domain. The classes are then linked to a variable of interest after they have been derived. For example in atmospheric sciences the classes comprise of gridded pressures and are linked to surface variables such as precipitation or temperature (e.g. Bárdossy *et al.*, 1995; Hewitson & Crane, 2002; Huth *et al.*, 2008). The following sections review common unsupervised classification techniques (Huth *et al.*, 2008).

Self Organizing Maps

Self organizing maps commonly referred to as SOMs are a category within artificial neural network models (Kohonen, 1989, 1990, 1991, 1995). Kohonen (1990) describe it as the process in which a neural network of connected nodes evolve in such a manner as to best describe an input signal. A signal is shown to the network whereby the nodes actively update their positions through a learning technique providing a simplified yet detailed description of the input signal (Hewitson & Crane, 2002; Kohonen, 1990). The learning technique utilizes a Euclidean distance measure to quantify performance (Hewitson & Crane, 2002).

In many ways SOMs are similar to traditional cluster analysis techniques in that the SOM will place nodes in such a way that is representative of the distribution of data points (Hewitson & Crane, 2002). For example nodes are densely spaced where data points are densely spaced and visa versa. However Hewitson & Crane (2002) argue that they differ in two fundamental ways. The first being the way in which the groups are defined and secondly when the SOM is viewed collectively it describes the structure of the data set. For example if two states commonly occur together in the dataset then their associated nodes will be located near to each other. The initial aim

2.2 Atmospheric Classification

of the SOM technique is not to identify individual groups or clusters within the data set. Rather the aim is to best describe the data set and if groups or clusters exist within the data set then they are reflected within the SOM (Hewitson & Crane, 2002). For example the SOM technique applied to atmospheric classification is described as follows (Hewitson & Crane, 2002; Skivic & Francis, 2012):

For a data set that consists a time series of pressure on a 5x10 grid the SOM will have n numbers of nodes each associated with a reference vector that consists of 50 coefficients. The size of the SOM (number of nodes) is user defined. It is expected that a larger SOM will represent the data to a higher degree in comparison to a smaller more general SOM. For each time realization the pressure data is presented to the SOM. The similarity between the reference vector for each node and the pressure data is calculated. The reference vector for the best match node is then updated by a user defined learning rate. The advantage of the SOM technique compared to cluster analysis is the nodes surrounding the best match node are also adjusted. The updating scheme is given as (Kohonen, 1990)

$$m_i(t + 1) = m_i(t) + h_{ci} \cdot [x(t) - m_i(t)] \quad (2.17)$$

where $m_i(t)$ is the reference vector for node i at time t . The data set at time t is given as $x(t)$ and h_{ci} is a neighborhood function relating to the best match node c applied to the i th node. It is given as

$$h_{ci}(t) = \alpha(t) \cdot \exp\left(-\frac{\|r_c - r_i\|^2}{2\sigma^2(t)}\right) \quad (2.18)$$

where $\alpha(t)$ is a user defined learning rate that decreases with time, $\|r_c - r_i\|$ is the distance between the best matching node c and node i and $\sigma(t)$ is referred to as the radius of training and defines which nodes to update.

The use of SOMs in atmospheric science has gained significant attention (Hewitson & Crane, 2002; Hong *et al.*, 2005; Skivic & Francis, 2012). The advantage of SOMs in synoptic climatology is the data is treated as a continuum (Hewitson & Crane, 2002). This provides a physically meaningful framework from which to evaluate the complex environment. Atmospheric circulation is not a set of clearly defined states but rather forms part of a continuum in which systems transition between states smoothly (Huth *et al.*, 2008). However a drawback in the method is that links to surface variables are only made post classification. In other words the SOM technique is a top down

approach as atmospheric states are derived first after which they are best matched to the occurrence of a particular surface variable.

Cluster Analysis

Cluster analysis methods are commonly used in classification and are arguably the most natural approach (Huth *et al.*, 2008). The k means clustering algorithm is popular approach although the use of simulated annealing as an optimizing tool in clustering analysis has been shown to be an improvement over the latter (Huth *et al.*, 2008; Phillip *et al.*, 2007). A drawback of the k means technique is that it is sensitive to the preliminary selection of node centroids therefore making it unstable (Huth *et al.*, 2008). The aim of the k means method is to maximise the similarity between data that belong to a particular cluster (Huth *et al.*, 2008). There are currently two methods by which this is done. The first is the algorithm developed in Lloyd (1982) and the second is that developed in Hartigan & Wong (1979). In the Lloyd (1982) algorithm the data are randomly divided into k clusters and the locally optimal solution is found by alternating between two steps (Slonim *et al.*, 2013). The first step assigns data to the nearest cluster centroid which is followed by updating the clusters' centroid (Lloyd, 1982; Slonim *et al.*, 2013). The algorithm developed in Hartigan & Wong (1979) attempts to search for k partitions that have locally optimal within-type sum of squares. This is achieved by moving points between clusters and evaluating the new within-type sum of squares and accepting changes that result in lower values.

Principle Component Analysis

Principle component analysis (PCA) is an eigentechnique that has two uses in classification (Huth *et al.*, 2008; Richman, 1986). Firstly it can be used as a pre-processing tool before classification to reduce colinearity between variables and as a data compression tool. Secondly the technique can be exploited in itself as a classification method (Richman, 1986). In order for PCA to be used as a classification tool the data matrix must be organized so that the gridded values are in rows and the time realizations are in columns (referred to as T-mode) (Compagnucci & Richmond, 2008; Huth *et al.*, 2008). For example in atmospheric classification the gridded pressures are the rows while the time realizations of the CP are the columns. If the data matrix is configured so that the gridded values are the columns and the time realizations are the rows (referred to as S-mode), the PCA algorithm only detects the modes of variance

that are not representative of individual circulation patterns (e.g. Huth *et al.*, 2008). A description of the T-mode and S-mode PCA is given in Compagnucci & Richmond (2008). For T-mode PCA the data matrix \mathbf{X}_t is of order $(t \times n)$ where t is the length of the record and n is the size of the grid. The PCA algorithm is the solution to the formulation $\mathbf{X}_t = \mathbf{F}\mathbf{A}^T$ where \mathbf{F} is a matrix of the principle components and \mathbf{A} relates the components of \mathbf{F} to the input variables in \mathbf{X}_t , for example \mathbf{A} could be the correlation matrix.

2.2.2 Supervised Classification

A supervised classification scheme is one in which classes (or groups) are derived by optimizing their fulfilment in meeting a specific objective. In terms of atmospheric classification the object is based on a variable of interest. The variable of interest can be rainfall, temperature or the wave climate. In other words the classes which comprise of atmospheric pressure data have strong links to the variable of interest. The classification consists of two parts. Firstly an *optimization procedure* is used to derive a set of classes. Secondly a *classification method* is used to assign CP realizations to their respective classes.

Fuzzy Logic in Atmospheric Classification

Bárdossy *et al.* (1995) were the first to utilize the concept of fuzzy logic within the framework of atmospheric classification science. The approach is based on the concept of fuzzy sets after Zadeh (1965). The fuzzy sets are used to mathematically quantify imprecise statements such as ‘high’ or ‘low’ pressure (Bárdossy *et al.*, 1995; Roffel & Chin, 1993; Zadeh, 1965). The concept of a fuzzy set after Zadeh (1965) is that objects in reality do not precisely belong to a particular class. For example at what temperature is something defined as being hot or cold? Therefore there exists some degree of vagueness attached with the objects membership to a particular class (Roffel & Chin, 1993). Fuzzy logic through the use of fuzzy sets provides a manner in which to quantify this membership grade. A fuzzy set is formally defined as follows (Zadeh, 1965):

for a set of points Y where an element of Y is denoted by y then the fuzzy set of X in Y is defined through a membership function $f_X(y)$. The membership function assigns to each element in Y a membership grade that is a real number in the

interval $[0,1]$. The value unity implies that y does belong to X whereas 0 implies that y does not belong to X .

Therefore this approach can be used within pattern recognition analysis by incorporating a set of fuzzy rules. Bárdossy *et al.* (1995) define a fuzzy rule as a simple ‘if A then B ’ statement. The object A is a fuzzy condition formulated with fuzzy sets and B is the consequence which is also a fuzzy set. The fuzzy rule attaches a degree of truth to the membership of an object to a class thus replacing binary logic. In terms of atmospheric classification the fuzzy rules represent the different CP classes. The rules contain spatial arrangements of fuzzy numbers that relate to the spatial arrangement of high and low pressures within the CP class.

The argument for the use of fuzzy logic as a tool in atmospheric classification is simple. Since CP realizations form part of a continuum it follows that each realization belongs to some extent to all classes (Bárdossy *et al.*, 1995; Huth *et al.*, 2008). Therefore there is a degree of ambiguity when assigning CPs to classes and fuzzy logic is an attempt to quantify this ambiguity. CPs are assigned to classes based on the degree to which they belong to a class and a CP is assigned to a class with which it shares the highest degree of fit (DOF). However Huth *et al.* (2008) argue that the manner in which CPs are assigned to classes can lead to erroneous classification.

Fuzzy Classification

In Bárdossy *et al.* (1995) CP classes were defined by spatially distributed fuzzy numbers. The fuzzy numbers attempt to represent different pressure states such that

1. very low values $(-\infty, -3, 0)_T$,
2. medium low values $(-4, -0.85, 4)_T$,
3. medium high values $(0.25, 0.85, 4)_T$,
4. very high values $(0, 3, +\infty)_T$ and
5. any value $(-\infty, 0, +\infty)_T$.

The fuzzy numbers ($n = 1 \dots 5$) represent different triangular membership functions. The functions (denoted by subscript T) assign a membership grade (value between $[0,1]$) to pressure values at each grid point based on their relative strengths of association to the fuzzy numbers. The triangular membership functions are defined as

(Bárdossy *et al.*, 2002)

$$\mu(x, y, z)_T(P) = \begin{cases} \frac{P-x}{y-x} & \text{if } x \leq P \leq y \\ \frac{P-z}{y-z} & \text{if } y < P \leq z \\ 0 & \text{else} \end{cases} \quad (2.19)$$

where P is the pressure value at a location. For example a very high pressure value that coincides with the location of fuzzy number 4 will be assigned a value of 1 whereas if it were a very low value it will be assigned 0. The fuzzy number $n = 5$ is used to describe values that have no effect on the CP and ensures spatially compact features. Therefore each CP class c can be defined by the vector $\mathbf{V}_c(i), i \in I$, where i are the locations in grid I and the integers $\mathbf{V}_c(i)$ are the fuzzy numbers described above.

The classification method can be summarised as following (Bárdossy *et al.*, 2015):

1. At each time realization t the membership grades $\mu_{V_c(i)}(P(i, t))$ for the pressure values $P(i, t)$ at each location i are calculated.
2. The membership grades are then combined in an overall DOF value such that

$$DOF(c, t) = \prod_{n=1}^4 \left[\frac{1}{N(c, i \forall V_c(i) = n)} \sum_{i=1}^{N(c, i \forall V_c(i) = n)} \mu_{V_c(i)}(P(i, t))^{Q_n} \right]^{\frac{1}{Q_n}} \quad (2.20)$$

where $N(c, i \forall V_c(i) = n)$ are the amounts for each fuzzy number $n = 1, \dots, 4$ present in the CP class c . The power Q_n reflects the relative importance for each of the fuzzy numbers.

The CP classes are derived through an optimization technique that is based on a variable of interest. The variable of interest is incorporated within a set of objective functions that relate to particular properties of the variable the user is interested in. For example the following objective functions were used in Bárdossy *et al.* (2002) to link CPs with rainfall:

$$O_1(\theta) = \sum_{i=1}^S \sqrt{\frac{1}{T} \sum_{t=1}^T (p(CP(t))_i - \bar{p}_i)^2}, \quad (2.21)$$

and

$$O_2 = \sum_{i=1}^S \frac{1}{T} \sum_{t=1}^T T \left| \ln \left(\frac{z(CP(t))_i}{\bar{z}_i} \right) \right| \quad (2.22)$$

where S is the number of rainfall stations, T is the total time period used for the classification, $p(CP(t))_i$ is the probability of the precipitation exceeding a threshold θ for a given CP at station i , and \bar{p}_i is the probability of the precipitation exceeding threshold θ without classification, $z(CP(t))_i$ is the mean precipitation for the given CP at time t and station i and \bar{z}_i is the mean precipitation without classification. Different thresholds θ were incorporated to delineate between CPs driving high and low rainfall events. The objective functions are linearly combined into an overall objective function $O = \alpha O_1 + \dots + \alpha O_m$ where the different weights $(\alpha_1, \dots, \alpha_m)$ reflect the relative importance of each objective function (Bárdossy, 2010).

The aforementioned objective functions are specific to precipitation. However it is possible to incorporate any variable of interest but the objective remain the same, to derive a set of CP classes with statistics of the variable of interest that differ significantly from those in the unclassified case (Bárdossy, 2010). Given the nature of the classification scheme the set of all possible CP combinations is significantly large (Bárdossy, 2010). There are 5 fuzzy numbers for each CP class that can be combined in any manner within the specified region I for a specified number of classes. Therefore a simulated annealing technique after Aarts & Korst (1989) is utilized as an optimization procedure.

2.2.3 Lagrangian Classification

The previous sections have dealt largely with Euclidean classification techniques. That is to say that the classification is applied to a temporal sequence of gridded atmospheric pressures. This attempts to divide the dataset into separate, clearly defined states that occur frequently. However it is not necessarily correct to evaluate atmospheric processes in this manner since atmospheric circulation is more of a continuum than a set of well defined states (Huth *et al.*, 2008). Therefore an alternative approach is to analyse air trajectories approaching a region with the aim to classify different circulation types (Ramos *et al.*, 2014; Stohl & Scheifinger, 1994). This method outlined in Ramos *et al.* (2014) incorporates a two stage clustering technique in which a set of 5 distinct air streams represents the flow for a given day. Once the flow for each day is characterized by the air streams a classification technique is applied to the dataset. The classification is based on the temporal evolution of latitude, longitude, height specific humidity and the distance to the region of interest of the representative flows. Details of the technique can be found in Ramos *et al.* (2014).

2.2.4 An Optimal Classification

What is a good classification? It is important to remember that the aim of a classification is to delineate distinct classes that provide insight into the high dimensional environment. Therefore utilizing the correct number of classes that describe important features of the environment without loss of generality is fundamental to a successful classification (Huth *et al.*, 2008). It is possible to quantify the classification quality through a set of measurements described in Huth *et al.* (2008). With respect to gridded atmospheric pressures these are: (a) the explained variance (EV), (b) the pattern correlation ratio and (c) the within class standard deviation.

The explained variance is defined as,

$$EV = 1 - \frac{ss_i}{ss_t} \quad (2.23)$$

with

$$ss_i = \frac{\sum_{k=1}^K \sum_{j=1, j \neq i}^{N_k} \sum_{i=1}^{N_k} \sum_{l=1}^L \sum_{m=1}^M (x_{lmik} - x_{lmjk})^2}{\sum_{k=1}^K N_k(N_k-1)}, \quad (2.24)$$

$$ss_t = \frac{\sum_{j=1, j \neq i}^N \sum_{i=1}^N \sum_{l=1}^L \sum_{m=1}^M (x_{lmi} - x_{lmj})^2}{N(N-1)}, \quad (2.25)$$

where ss_i is the mean total sum of squares within a class, ss_t is the mean total sum of squares without classification, and x_{lmik} the anomaly value at the (l, m) grid point for the i th index belonging to k th class, with K the total number of classes.

The pattern correlation ratio is defined as,

$$PCR = \frac{Pc_b}{Pc_i} \quad (2.26)$$

with

$$Pc_b = \frac{\sum_{l=1, l \neq k}^K \sum_{k=1}^K \sum_{j=1}^{N_l} \sum_{i=1}^{N_k} r(\mathbf{x}_{ik}, \mathbf{x}_{jl})}{\sum_{l=1, l \neq k}^K \sum_{k=1}^K N_k N_l}, \quad (2.27)$$

$$Pc_i = \frac{\sum_{k=1}^K \sum_{j=1, j \neq i}^{N_k} \sum_{i=1}^{N_k} r(\mathbf{x}_{ik}, \mathbf{x}_{jk})}{\sum_{k=1}^K N_k(N_k-1)}, \quad (2.28)$$

where Pc_b is the mean correlation coefficient between fields from different classes, Pc_i is the mean correlation coefficient between fields within a given class, $r(\mathbf{x}_{ik}, \mathbf{x}_{jl})$ is the correlation coefficient between the i th index for the k th class and the j th index for the l th class, \mathbf{x} a vector containing all anomaly values, and N_k is the number of anomalies belonging to the k th class.

The within class standard deviation is defined as,

$$WSD = \frac{1}{K} \sum_{i=1}^K sd_i \quad (2.29)$$

where sd_i is the standard deviation within the i th class. These three methods provide useful insight into the performance of the classification with regards to deriving a set of dissimilar classes with reduced within class variability. However these methods do not describe the ability of the classification to delineate the drivers of a surface variable of interest.

2.3 Summary of Chapter

Coastal vulnerability is a global concern and attempts to quantify risk relies strongly upon the accurate description of regional wave climates. Wave observations offer the ideal approach to understand and characterize regional wave climates. However good quality data is limited to a few regions that have access to wave buoy data at which the record lengths span 30–40 years. Wave buoys are limited in that they describe wave conditions at a single location. Satellite data attempts to address this and provides a means to describe global wave fields. However relative to the size of the ocean all observed wave data are sparse and it is not possible to monitor the entire ocean at the same time. To address this significant attention has been focussed on the development of wave models to supply accurate estimates of waves nearshore and on a global scale.

With regard to characterising regional wave climates using wave models, two approaches exist. The first is to develop a stochastic model that statistically describes the relationships between wave variables, or utilize wave data from existing wave hindcast datasets available from wave modelling groups such as NCEP or ECMWF. The latter approach requires an additional step to transform waves from offshore to nearshore. Statistical models are advantageous because they are not computationally demanding and can be used to simulate any given record length. However accurate wave predictions require descriptions of the complex, non-linear dependence structures between wave variables. Copulas are a mathematically tractable solution to this. Global wave hindcast data suffers two drawbacks. Firstly accurate descriptions of waves relies strongly upon data reanalysis and accurate wind field forcing and extreme values are not well captured. Furthermore nearshore processes are not well described because of complex non-linear interaction. Grid refinement is required to handle this and

therefore waves must be transformed from offshore to nearshore. Therefore statistical models are well suited to model nearshore wave processes. With regard to wave time series modelling there are two approaches. The first is to simulate individual storm events with specific durations and inter-arrival times. The second is to simulate a continuous wave sequence. Current statistical methods are not linked to the physical drivers (atmospheric circulation patterns) of waves and therefore independent events and dependence structures are difficult to describe. Furthermore they require good records of wave observations so as to describe the dependence structures between wave variables. However if the links between regional wave climates and synoptic scale circulation patterns can be delineated they can be exploited to simulate synthetic wave records in a physically based manner. Atmospheric classification provides a means to achieve this.

Atmospheric classification is a means to simplify complex scenarios of high dimensionality into a set of clearly defined states that occur often enough to generalise. In meteorological sciences the classes contain gridded atmospheric pressures or geopotentials over a particular regions. The classes are then linked to a variable of interest on the surface such as precipitation or waves herein. Several methods exist to delineate CP classes such as k means clustering, principle component analysis (PCA) and self-organizing maps (SOMS). Although these methods have been shown to accurately delineate between states, links to a surface variables are only made post classification. This is referred to as a ‘top down’ approach. Since the goal is to find CPs that drive a variable of interest an approach which incorporates this variable is favoured. A fuzzy rule based classification algorithm has been developed that incorporates the desired variable to drive the model to an optimal solution. The algorithm has only been applied to precipitation but it has been shown to provide significant insight into the synoptic scale circulation patterns driving rainfall events. Further developments of the algorithm are required to link it to waves. A suitable method is also required to evaluate the performance of the algorithm with regard to the optimal number of classes required for a successful classification.

This review of literature suggests a method of quantifying coastal risk by stochastically simulating long sequences of regional wave climates based on their atmospheric drivers. Atmospheric circulation patterns strongly associated with waves can be identified using a fuzzy classification algorithm that is based on the statistical properties of the waves. The interdependencies between wave climate variables associated with each atmospheric circulation pattern can be completely described using copulas. This CP-

Literature Review

Wave simulation technique improves the realism of stochastic models while retaining their simplicity and parsimony relative to process-based models.

Chapter 3

Atmospheric circulation patterns & wave climates

Abstract

The wave climate is a fundamental driver of coastal vulnerability and changing trends in wave heights, periods and directions can severely impact a coastline. In a diverse storm environment, the changes in these parameters are difficult to detect and quantify. Since wave climates are linked to atmospheric circulation patterns an automated and objective classification scheme was developed to explore links between synoptic scale circulation patterns and wave climate variables, specifically wave heights. The algorithm uses a set of objective functions based on wave heights to guide the classification and find atmospheric classes with strong links to wave behaviour. Spatially distributed fuzzy numbers define the classes and are used to detect locally high and low pressure anomalies. Classes are derived through a process of simulated annealing. The optimized classification focuses on extreme wave events. The east coast of South Africa was used as a case study. The results show that three dominant patterns drive extreme wave events. The circulation patterns exhibit some seasonality with one pattern present throughout the year. Some 50 - 80% of the extreme wave events are explained by these three patterns. It is evident that strong low pressure anomalies east of the country drive a wind towards the KwaZulu-Natal coastline which results in extreme wave conditions. We conclude that the methodology can be used to link circulation patterns to wave heights within a diverse storm environment. The circulation patterns agree with qualitative observations of wave climate drivers. There are applications to the assessment of coastal vulnerability and the management of coastlines worldwide.

3.1 Introduction

Wave climates are strongly linked to atmospheric circulation. The link is complex and its direct functional description can be difficult to derive. However atmospheric circulation can be classified into discrete patterns (CPs), which represent different links

to wave behavior. If these links can be clarified they can be used to assess changes in the wave climate.

Understanding the wave climate at a given region is of fundamental importance to coastal planners, managers and engineers. Coastal erosion depends strongly on the extreme wave events, which in turn are driven by atmospheric circulation. Therefore changes in circulation patterns can change wave climate parameters (wave height, direction etc). This has a direct impact on the location and severity of beach erosion. Wave climates along coastlines dominated by a single storm system are easiest to define and changes in circulation patterns are reflected in wave observations. For example, Komar *et al.* (2010) were able to evaluate changes in the wave climate of the North Pacific and North Atlantic oceans. The shorelines along these coasts are typically dominated by one or two main storm systems with inherent seasonality. However in a diverse storm environment the trends can be more difficult to identify and quantify.

The aim of this study is to utilize existing statistical methods to identify the atmospheric circulation patterns driving the wave climate at a given location, with particular reference in extreme wave events. These events are defined here as periods during which significant wave heights exceed a threshold of 3.5 m. Physical links between atmospheric features and wave heights are complex and nonlinear. Therefore attempts to model wave characteristics derived from circulation patterns can be difficult and time consuming. Statistical knowledge gained from the observations of wave climates and pressure fields allow insight into this complex relationship without the need for explicit physical coupling. This can be a useful tool for risk analysis since it provides insight into the source of extreme events. If we can identify the circulation patterns that drive extreme events, then their occurrence (or the occurrence of similar patterns) can have a degree of risk attached to it. For example the risk could be the likelihood of an extreme wave event, of severe erosion, of extended storm durations, or a combination of all three.

Circulation patterns are herein described in terms of pressure anomalies on the 700 hPa geopotential. The types/classes or groups of anomalies can be specified by two approaches: 1) those specified prior to classification; 2) those that are derived and evolve during the classification process (Huth *et al.*, 2008). In the past, anomaly patterns were identified by experts in the field: examples are the Hess-Brezowski catalogue or the Lamb classification (Hess & Brezowsky, 1952; Huth *et al.*, 2008; Lamb, 1972). However the power of modern computers provides a means to generate numerical solutions to complicated algorithms that can automate the process. It is important

to note that atmospheric circulation patterns are not a set of separated, well-defined states. CPs change smoothly between states that form part of a continuous sequence of events (Huth *et al.*, 2008). Therefore the classes (or types) merely represent simplified climatic events responsible for specific variables of interest. While automated derivations of classification types utilize objective reasoning, according to Huth *et al.* (2008) the procedure as a whole cannot be considered fully objective. A number of subjective decisions are still employed. For example the number of CPs to use and the method of differentiating classes. Existing objective based classification algorithms such as self-organizing maps (SOMS), principle component analysis (PCA) and cluster analysis provide effective ways to visualize the complex distribution of synoptic states (Hewitson & Crane, 2002; Huth *et al.*, 2008). These approaches are fundamentally based on only the predictor variables (atmospheric pressure anomalies in our case). Links to surface weather variables are only made once the classification technique has been carried out. This can therefore lead to non-optimal links to the variable of interest (Bárdossy, 2010).

The classification method used for this study is a fuzzy rule based algorithm developed by Bárdossy *et al.* (1995). The classification technique aims to find strong links between atmospheric CPs and a variable of interest, for this study the wave height. The algorithm was originally used to link atmospheric CPs with rainfall events (Bárdossy *et al.*, 2002). In the present study the method has been adapted to use wave heights to guide the classification procedure. The main aim of this paper is to investigate the feasibility of using this method to identify CPs that are the main drivers of regional wave climates for application to coastal vulnerability assessments.

3.2 Methods

3.2.1 Case Study Site

The KwaZulu-Natal (KZN) coastline (Fig. 3.1) has a high energy wave climate. Tropical cyclones, mid-latitude (extra-tropical) cyclones and cut-off lows have been cited as important drivers of the local wave climate (Corbella & Stretch, 2012d; Mather & Stretch, 2012; Rossouw *et al.*, 2011). Tropical cyclones that become stationary to the south east of Madagascar can occasionally drive large wave events that cause severe beach erosion in KZN (Corbella & Stretch, 2012d; Mather & Stretch, 2012). Cut-off lows are deep low pressure systems that are displaced from the normal path of

west-east moving mid-latitude cyclones (Preston-Whyte & Tyson, 1988). Instabilities within the westerly zonal flow, due to the high wind shear, create vortices (low and high pressure), which can become cut-off and move equatorward (Preston-Whyte & Tyson, 1988). This diverse storm environment leads to seasonality within the wave climate. On average autumn and winter are associated with the largest wave energy, while summer has the smallest (Corbella & Stretch, 2012d). Seasons are defined according to Table 3.1.

3.2.2 Sources of Data

Wave data was obtained from wave measurement buoys at two locations along the KwaZulu-Natal coastline (Fig. 3.1) for the period 1992–2009. A comparison of the wave data from the Durban and Richards Bay measurement locations by Corbella & Stretch (2012d) showed a strong correlation. Therefore, where necessary, the two data sets were used to fill in missing data to provide a continuous wave record for the KwaZulu-Natal coastline. The data comprised significant wave heights, maximum wave heights, wave periods, and wave directions at 3-hourly intervals. However only daily mean values of significant wave heights were used for the analysis reported here.

The CP classification procedure was applied to daily normalized pressure anomalies that describe atmospheric circulation patterns. The anomalies were derived from the 700 hpa geopotential height with a grid resolution of 2.5° ($10^\circ S$ $0^\circ E$ – $50^\circ S$ $50^\circ E$). This region is of significant size to include CPs responsible for swell waves (Corbella *et al.*, 2015). Geopotential heights were obtained from the ERA-Interim data set for the period 1979–2009 (<http://apps.ecmwf.int/datasets/>). Let $k(i, t)$ be the geopotential height at location i and time t , then the anomaly at location i and time t is defined as

$$h(i, t) = \frac{k(i, t) - \overline{k(i)}}{\sigma(i)} \quad (3.1)$$

where $\overline{k(i)}$ and $\sigma(i)$ are the average and standard deviation of the geopotential at location i .

3.2.3 Classification Methods

The classification used herein comprises two parts: (1) An optimization procedure in which a set of classes defining atmospheric states are derived using an optimization process as described in §3.2.4, and (2) A classification method that involves a process

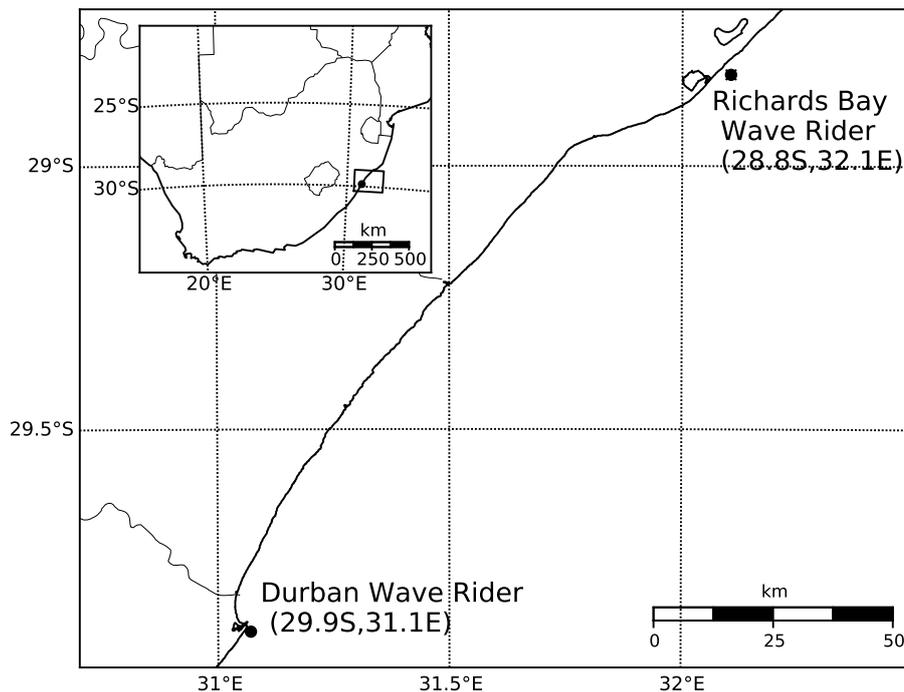


Fig. 3.1 Locations of the wave observation buoys at Durban and Richards Bay, along the KwaZulu Natal coastline.

of assigning CPs to the classes.

The aim is to identify a classification in which the set of classes defining atmospheric pressure fields can explain the occurrence of wave events at a specified location. There are many ways in which classification algorithms can be constructed. Classifications can be subjective, objective or a mixture of both (Bárdossy, 2010; Huth *et al.*, 2008). Objective classification algorithms employ a self-learning technique whereby atmospheric classes are derived through an optimization procedure (for examples see: Bárdossy, 2010; Bárdossy *et al.*, 2002; Hewitson & Crane, 2002; Huth *et al.*, 2008). Since the goal is to gain insight into the drivers of a regional wave climate it follows that the wave climate should be included within the optimization procedure (see § 3.2.4). Then the set of CP classes that are derived have strong links to the regional wave climate. Furthermore, classes linked with these variables explain, as best possible, their occurrences. This is a useful tool in guiding the algorithm to an optimal solution. Classifying CPs linked with extreme wave events is the focus of this study. The classifying procedure uses wave heights as the dependent variable to find classes of the independent variable, atmospheric pressure anomalies. The method of classifi-

Table 3.1 The allocation of months to seasons

| Season | Months |
|--------|--------------------|
| Summer | January – March |
| Autumn | April – June |
| Winter | July – September |
| Spring | October – December |

cation is described in detail by Bárdossy (2010); Bárdossy *et al.* (1995, 2002). Only a brief overview is given here.

The classification method used herein is fuzzy-rule based which incorporates the use of fuzzy-sets (Zadeh, 1965). This allows the algorithm to handle imprecise statements such as 'strong high pressure' or 'low pressure' (Bárdossy *et al.*, 1995). The CPs at each time realization are assigned to a certain CP class or group. Each CP class is defined by a rule which comprises a number of fuzzy set membership functions. The n^{th} CP class is described by the fuzzy rule n as a vector $\mathbf{c}_n = [V(1, n), \dots, V(i, n), \dots, V(K, n)]$, for all available grid locations $(1, \dots, K)$, where V is the matrix containing all CP rules $(1, \dots, n)$, the index $V(i, n)$ is the fuzzy set number corresponding to the location i for rule n . The rules consist of the following fuzzy sets:

Fuzzy set number 0 - any type of anomaly,

Fuzzy set number 1 - strong positive anomaly,

Fuzzy set number 2 - weak positive anomaly,

Fuzzy set number 3 - weak negative anomaly and

Fuzzy set number 4 - strong negative anomaly.

The fuzzy set numbers $(1, \dots, 4)$ describe the locations of different pressure types, however the fuzzy set number 0 is irrelevant for the CP classification. In general most of the grid locations belong to this fuzzy set number. The algorithm only considers patterns with structures corresponding to the arrangement of the fuzzy set numbers $1, \dots, 4$.

From the fuzzy set numbers described above, a membership grade μ at location i can be assigned for each daily anomaly pattern as

$$\mu_n(i, t) = g(\mathbf{c}_n(i), t) \quad (3.2)$$

where n is the fuzzy rule, $g(\mathbf{c}_n(i), t)$ is the membership function for the fuzzy set number j at location i at time t (Bárdossy *et al.*, 2002). The membership grade μ at each location ranges between 0–1 based on the membership function for the location specific fuzzy number. A value of 0 implies that the anomaly value has no association with the fuzzy number and a value of 1 implies the anomaly is strongly associated with the fuzzy number. It follows that a combination of the membership grades provide insight into the performance of each CP rule in relation to the daily anomaly patterns. A degree of fit (DOF) is computed for each CP rule and the rule with the highest DOF value is assigned to the circulation pattern class for that day. The degree of fit is defined as follows (Bárdossy *et al.*, 1995, 2002):

$$DOF(n, t) = \prod_{j=1}^4 \left[\frac{1}{N(n, j)} \sum_{i=1}^{N(n, j)} \mu_n(i, t)^{P_j} \right]^{\frac{1}{P_j}} \quad (3.3)$$

where t is the day, $N(n, j)$ is the number of grid points corresponding to the fuzzy set number j for fuzzy rule n , the term $\sum_{i=1}^{N(n, j)} \mu_n(i, t)$ sums all the membership grades at various locations corresponding to the fuzzy set number j for rule n , and the exponent P_j is a parameter that allows us to emphasize the influence of selected rules on the *DOF*.

The CP rules were obtained via an optimization procedure following Bárdossy *et al.* (2002) and is described in § 3.2.4.

3.2.4 Optimization Methods

The goal of the optimization is to derive a set of CP classes or rules defining dominant circulation patterns in a particular region. The rules are strongly linked to a variable of interest. The optimization procedure should maximize dissimilarity between the CP types while minimizing the variability within the classes. The significant wave height (H_s) was selected as the ‘guiding’ variable for this study. The algorithm considers both the daily average significant wave height and the daily maximum significant wave height. The optimization procedure was carried out for the period containing all wave data (1992–2009). A simulated annealing algorithm following Aarts & Korst (1989) is used in the optimization procedure. Details of the process are given in Bárdossy *et al.* (2002). The algorithm may be briefly outlined as follows:

- (1) Randomly assigned CP rules are initialized and their performance is evaluated through an objective function O .

-
- (2) The initial ‘annealing temperature’ is set to q_0 .
 - (3) A rule n is selected randomly.
 - (4) A location i is selected randomly.
 - (5) A fuzzy number c^* is selected randomly.
 - (6) If $c_n(i) = c^*$ return to step 2.
 - (7) Set $c_n(i) = c^*$ and run the classification.
 - (8) Calculate the new performance O^* for the new rules.
 - (9) If $O^* > O$, accept the change.
 - (10) If $O^* \leq O$, accept the change with probability $\exp\left(\frac{O-O^*}{q_j}\right)$.
 - (11) If the change has been accepted, replace O by O^* .
 - (12) Repeat steps 2 – 10 a specified number of iterations.
 - (13) Decrease the ‘annealing temperature’ such that $q_{new} < q_{old}$.
 - (14) Repeat the steps 2 – 12, until the number of accepted changes becomes less than a predefined limit.

The optimization process relies strongly upon a set of objective functions. The objective functions are based on the extreme wave events, wave heights and storm duration as discussed in § 3.2.4.

Objective Functions

A good classification contains classes with corresponding wave statistics which differ from the statistics calculated without classification. The goal of the classification is to obtain a set of CP rules which correspond to the occurrence of extreme waves. Extreme waves events are defined where $H_s \geq 3.5\text{m}$. Therefore the objective functions used within the algorithm are designed to optimize the CP occurrences which coincide with extreme wave events. These are relatively rare events. A random classification leads to the same probability of occurrence as the mean for each rule, which is undesirable (Bárdossy, 2010). A good classification should lead to rules that differ from the climatological mean for the selected variable, in this case the wave height.

The intention of this classification is to find CPs that drive extreme wave events. Three objective functions were used as the performance measures. The first objective function relates to the conditional probability of an event based on the occurrence of a CP class. It is given as:

$$O_1(\theta) = \sqrt{\sum_{t=1}^T h_{CP(t)} (p(H_s \geq \theta | CP(t)) - \bar{p})^2} \quad (3.4)$$

where θ is a predefined threshold, T is the total number of days, $h_{CP(t)}$ is the frequency of the CP class, $p(CP(t))$ is the frequency that the threshold is exceeded for a given CP on a day t , \bar{p} is the unclassified probability of exceedance for all days in period T . The advantage of incorporating a predefined threshold θ is to allow the algorithm to evaluate different scenarios. For this study two different thresholds were considered. The first relating to extreme wave events whereby $\theta_1 = 3.5$ m and the second whereby $\theta_2 = 2.5$ m. The second threshold allows the algorithm to explore a larger dataset from which the classes can be derived.

Another useful measure of the performance relates to the mean significant wave heights. The ratio between the CP class averaged wave heights to the unclassified mean provides information on the separability of the classes from the mean. Therefore the second objective function incorporating average significant wave heights is defined as

$$O_2 = \sum_{t=1}^T h_{CP(t)} \left| \frac{H_s(CP(t))}{\overline{H_s}} - 1 \right| \quad (3.5)$$

where $H_s(CP(t))$ is the mean significant wave height on a day with the given CP(t) class and $\overline{H_s}$ is the mean daily wave height without classification.

A weighted linear combination of Eq. 3.4, Eq. 3.5 was used to optimize the solution to the classification algorithm. The weights were chosen to emphasize the importance of certain objective functions to others and to correct for the range in values between the different objective functions.

3.2.5 Classification Quality

Classification quality refers to the ability of the algorithm to maximize dissimilarity between a set of CP classes, while minimizing variability within daily CP realizations. This study focuses on classifying CPs driving extreme wave events, therefore there are two criteria for measuring the classification quality. The first is the ability of

the classification to explain extreme wave events. The second is the variability of the classifications within each CP class. There exists an optimal number of CP rules which successfully explain extreme events and daily CP realizations. Too few rules implies that the resulting CPs do not allow a proper distinction of the causal mechanisms and would lead to classes which have statistics similar to the unclassified case. Too many classes increases the computational effort and captures features that are not general and do not correspond to the wave generating mechanisms. Bárdossy (2010) suggests utilizing the objective functions as a measure of the classification quality. Huth *et al.* (2008) list a number of different quality measures that explain the separability between and variability within CP classes. For this study the variability within the classes as well as the degree of fit are used as measures of the classification quality. This provides insight into the performance of the classes with respect to their ability to explain average CPs.

The variability of extreme events is defined as the position of the lowest anomaly relative to the average pattern. This was assumed as the storm centre. Wave events are driven by storms associated with low pressures (i.e. negative anomalies). The performance of the CP classes in explaining extreme wave events can be measured by their relative contribution to extreme events, namely $p(CP|H_s \geq \theta)$, where θ is a predefined threshold (for this study 3.5 m). A classification strongly linked to the wave climate should define classes whose frequency of occurrence correspond to the average and extreme wave events. This implies that CPs driving extreme events should occur infrequently, whereas CPs driving the average wave climate should occur more frequently.

3.3 Results

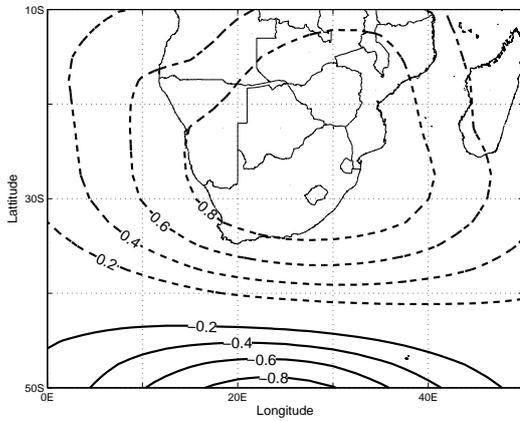
3.3.1 Dominant CP Classes

The objective functions (Eqs. 3.4 & 3.5) were used to derive a set of CP classes which explain extreme wave events. Fig. 3.2 shows the average anomaly patterns for all the CP classes. CP99 refers to an unclassified class. Useful statistical parameters relative to this study for a given CP class are: (a) Frequency of occurrence, (b) Percentage contribution of extreme events, (c) average and maximum significant wave heights (\overline{H}_s). These parameters are obtained from the classification and are shown in Table 3.2.

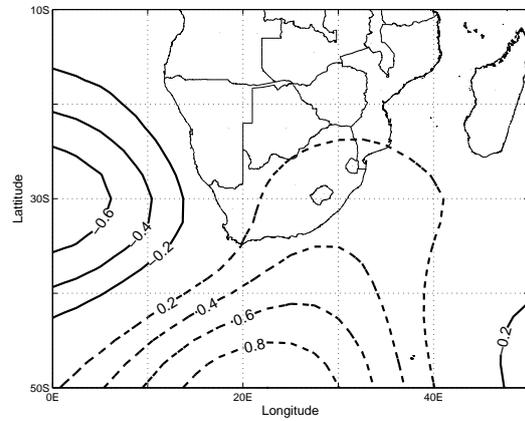
The results show two trends in CPs that drive wave development. Firstly CP01

and CP02 (Figs. 3.2(a) & (b)) according to Table 3.2 occur most frequently ($\sim 17\%$ of the time). CP01 resembles that of mid-latitude cyclones which frequently travel in a west to east direction south of the country, while CP02 resembles the high pressure systems that follow the mid-latitude cyclones. Secondly, Table 3.2 shows that CP03 is associated with 30 – 60% of all extreme wave events. The large contribution by this class to extreme events is present all year round with the highest contribution in winter ($\sim 65\%$). CP03 (Fig. 3.2(c)) occurs infrequently (7 – 9% of the time), however it is associated with large wave heights. For example average and maximum significant wave heights associated with CP03 range from 2.4 – 3.0 m and 5.0 – 8.5 m respectively. CP05 and CP06 (Figs. 3.2(e),(f)) according to the classification are responsible for about 30% of extreme events in spring and summer respectively. CP06 represents low pressure anomalies southeast of Madagascar. This appears to resemble the strong low pressure systems that are associated with tropical cyclones. According to Mather & Stretch (2012) low pressure systems southeast of Madagascar can cause large swells. CP05 resembles low pressure systems over the interior and which extend southwards.

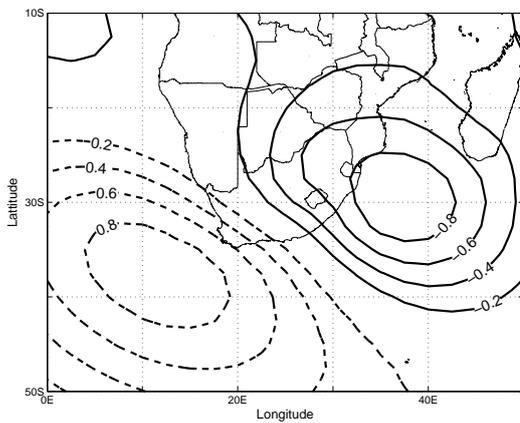
No time lag was considered when deriving the CP classes. This constrains the algorithm to only consider CPs occurring on the day of a wave event and assumes that extreme events are driven by relatively stationary CPs.



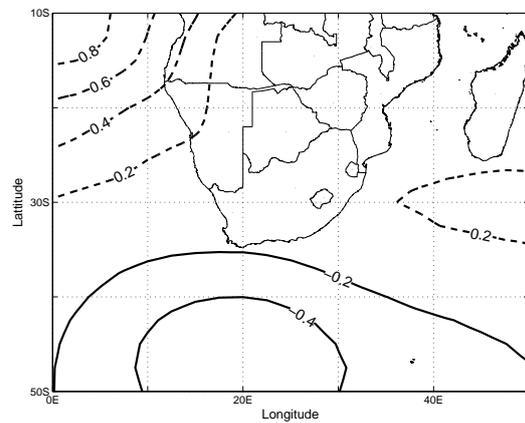
(a) CP01



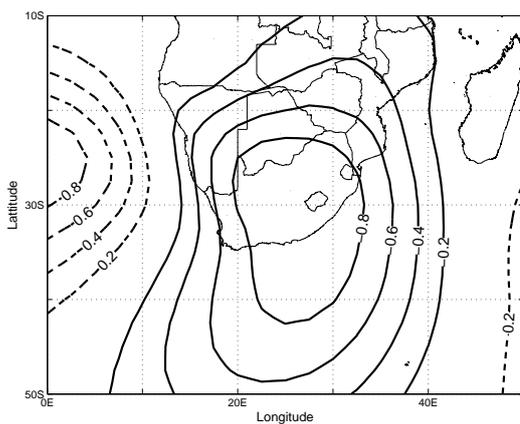
(b) CP02



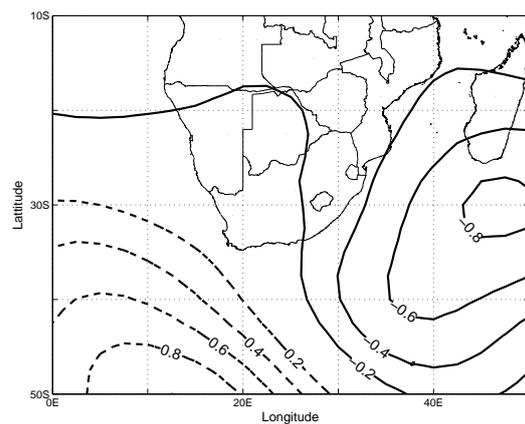
(c) CP03



(d) CP04



(e) CP05



(f) CP06

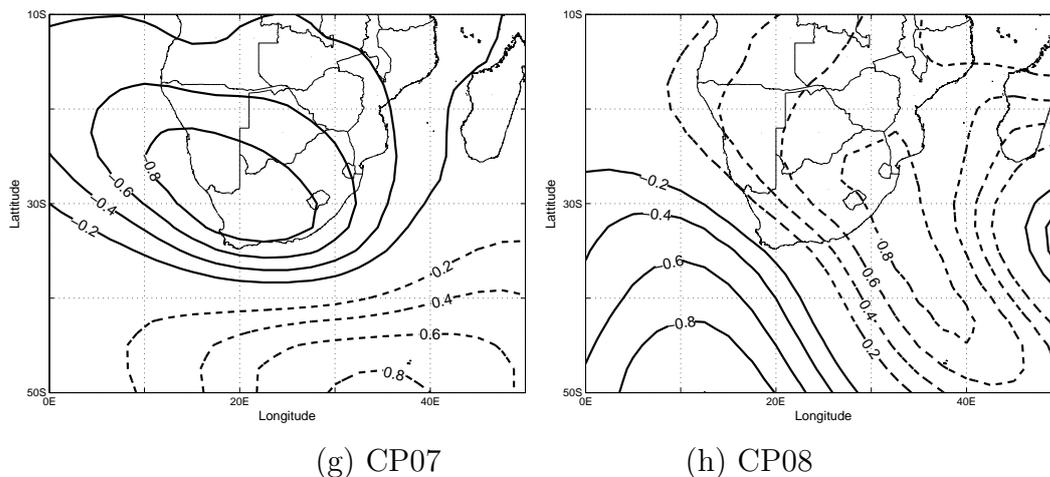


Fig. 3.2 Average anomaly patterns for all CP classes:1–8. Positive anomaly contours are shown as the dashed line while negative contours are solid.

3.3.2 CP Variability

Degree of Fit (DOF)

The degree of fit relates to how well the CP for each day is classified as a given class relative to the rule file. The larger the degree of fit the stronger the relation between the CP and the CP class. Fig. 3.3 shows the average anomaly pattern for CP03 together with the CPs associated with both the highest and lowest degree of fit value for that class. CP03 is associated with cut-off lows to the east/south-east of South Africa. The pattern also shows a strong high pressure region to the southwest. The combination of strong cut off lows occurring in conjunction with high pressure regions are important features for channeling waves towards the eastern coastline. Fig. 3.3(c) is the CP with the lowest degree of fit for the class CP03 and it shows only a weak anomaly pattern.

Variability Within Classes

It is expected that in the vicinity of the regions defining rule types (high or low pressures) the standard deviation should be low. This is because the classification is based on the location of these rules in comparison to the anomaly patterns for specific days. Whereas the locations of “any anomaly” rule types (fuzzy number 0) are expected to have significant variability, the variability in the vicinity of negative anomalies (low pressures) can be attributed to the movement of the low pressure

3.3 Results

Table 3.2 CP Occurrence frequencies and wave height statistics associated with each CP class.

| STATISTICS | CP01 | CP02 | CP03 | CP04 | CP05 | CP06 | CP07 | CP08 | CP99 ^a |
|---|------|------|------|------|------|------|------|------|-------------------|
| Occurrence frequency ($p(CP)\%$) | | | | | | | | | |
| Summer | 18 | 18 | 8.0 | 13 | 7.5 | 8.1 | 5.6 | 15 | 8.3 |
| Autumn | 18 | 19 | 8.0 | 11 | 10 | 7.2 | 5.1 | 13 | 8.8 |
| Winter | 16 | 17 | 8.1 | 12 | 11 | 8.4 | 4.5 | 14 | 8.9 |
| Spring | 17 | 16 | 9.1 | 12 | 9.4 | 7.7 | 5.0 | 15 | 9.2 |
| All Seasons | 17 | 17 | 8.3 | 12 | 9.6 | 7.8 | 5.1 | 14 | 8.8 |
| Threshold exceedance for a given CP ($p(H_s \geq \theta CP)\%$) | | | | | | | | | |
| Summer | – | 0.4 | 8.0 | 0.3 | – | 3.5 | 2.6 | 0.7 | – |
| Autumn | 1.2 | 1.5 | 12 | 1.6 | 2.1 | 2.0 | 5.6 | 0.6 | 5.0 |
| Winter | 0.9 | 0.8 | 14 | – | 0.9 | 0.4 | 2.3 | 0.8 | 0.4 |
| Spring | 0.3 | – | 4.6 | 0.6 | 2.2 | – | 0.7 | – | – |
| All Seasons | 0.6 | 0.7 | 9.6 | 0.6 | 1.4 | 1.5 | 2.8 | 0.5 | 1.4 |
| Exceedance contribution ($p(CP H_s \geq \theta)\%$) | | | | | | | | | |
| Summer | – | 5.6 | 50 | 2.8 | – | 22 | 11 | 8.3 | – |
| Autumn | 7.7 | 10 | 33 | 6.4 | 7.7 | 5.1 | 10 | 2.6 | 17 |
| Winter | 7.5 | 7.5 | 64 | – | 5.7 | 1.9 | 5.7 | 5.7 | 1.9 |
| Spring | 4.5 | – | 55 | 9.1 | 27 | – | 4.5 | – | – |
| All Seasons | 5.8 | 7.4 | 48 | 4.2 | 7.9 | 6.9 | 8.5 | 4.2 | 7.4 |
| Average H_s (m) for each CP | | | | | | | | | |
| Summer | 1.8 | 1.9 | 2.5 | 1.8 | 1.8 | 2.2 | 2.2 | 1.9 | 1.9 |
| Autumn | 1.8 | 1.9 | 2.7 | 1.9 | 2.0 | 2.0 | 2.1 | 1.9 | 2.1 |
| Winter | 2.0 | 2.0 | 2.9 | 1.9 | 2.1 | 2.0 | 2.2 | 2.0 | 1.9 |
| Spring | 2.0 | 1.9 | 2.4 | 1.9 | 2.2 | 2.0 | 2.2 | 2.0 | 2.0 |
| All Seasons | 1.9 | 1.9 | 2.6 | 1.9 | 2.1 | 2.1 | 2.2 | 2.0 | 2.0 |
| Standard deviation of H_s (m) for each CP | | | | | | | | | |
| Summer | 0.48 | 0.49 | 1.1 | 0.49 | 0.53 | 0.76 | 0.74 | 0.61 | 0.49 |
| Autumn | 0.58 | 0.66 | 1.0 | 0.70 | 0.76 | 0.66 | 0.90 | 0.55 | 1.0 |
| Winter | 0.58 | 0.61 | 0.94 | 0.55 | 0.66 | 0.58 | 0.66 | 0.55 | 0.67 |
| Spring | 0.51 | 0.49 | 0.84 | 0.52 | 0.71 | 0.50 | 0.61 | 0.50 | 0.49 |
| All Seasons | 0.54 | 0.57 | 1.0 | 0.56 | 0.69 | 0.63 | 0.74 | 0.56 | 0.70 |
| Max H_s (m) for each CP | | | | | | | | | |
| Summer | 3.4 | 4.0 | 8.5 | 3.7 | 3.4 | 5.0 | 5.2 | 5.6 | 3.3 |
| Autumn | 4.0 | 5.5 | 5.7 | 5.5 | 6.3 | 4.3 | 5.1 | 4.0 | 5.4 |
| Winter | 4.2 | 3.8 | 5.6 | 3.4 | 3.8 | 3.5 | 4.3 | 4.8 | 3.6 |
| Spring | 3.9 | 3.3 | 5.3 | 4.5 | 5.4 | 3.4 | 3.7 | 3.5 | 3.3 |
| All Seasons | 4.2 | 5.5 | 8.5 | 5.5 | 6.3 | 5.0 | 5.2 | 5.6 | 5.4 |

^aCP99 is the unclassified class. Blank entries imply zero occurrences in the data set

systems. It is also expected that high pressure systems are more stable with lower variability in their positions. For example Fig. 3.3(d) shows lower variability in the vicinity of the high pressure region while high variability (standard deviation of 1) in the low pressure region. This can be attributed to the movement of low pressure systems around the negative anomaly. Fig. 3.3(d) shows high standard deviation values in comparison to the mean negative anomaly. This could also indicate that CP anomalies driving extreme events (cut-off lows) are associated with strong negative anomalies.

3.3.3 CP rules and extreme events

Daily CP's classified as in a certain class for extreme wave events ($H_s \geq 3.5m$) were compared to the average patterns for that class. Fig. 3.4 shows the average pattern for CP03 together with selected extreme events corresponding to CP03. The centres of the CP's are shown as "+" symbols in the plot. A centre is defined as the location of the peak negative anomaly. The variability within the class is apparent. However the majority of CPs classified as CP03 resemble strong cut-off lows to the east/southeast of the country. It is apparent that these strong cut-off lows drive extreme wave events. Fig. 3.5 and Table 3.3 show CP's associated with the six largest significant wave height events. Four out the six events have been classified as CP03, the class contributing to the majority of extreme events. The concept of CPs belonging, to some degree, to all the classes is evident in Fig. 3.5(f). This shows a similar pattern to CP04 and CP08, which both represent low pressures southeast of Madagascar. However according to the classification this CP belongs to class CP08 and not CP04. From visual inspection it appears to resemble class CP04 better than CP08. Fig. 3.5(a) is the CP associated with the March 2007 storm which caused severe coastal erosion along the KwaZulu-Natal coastline (Corbella & Stretch, 2012c; Mather & Stretch, 2012) with significant wave heights reaching 8.5m.

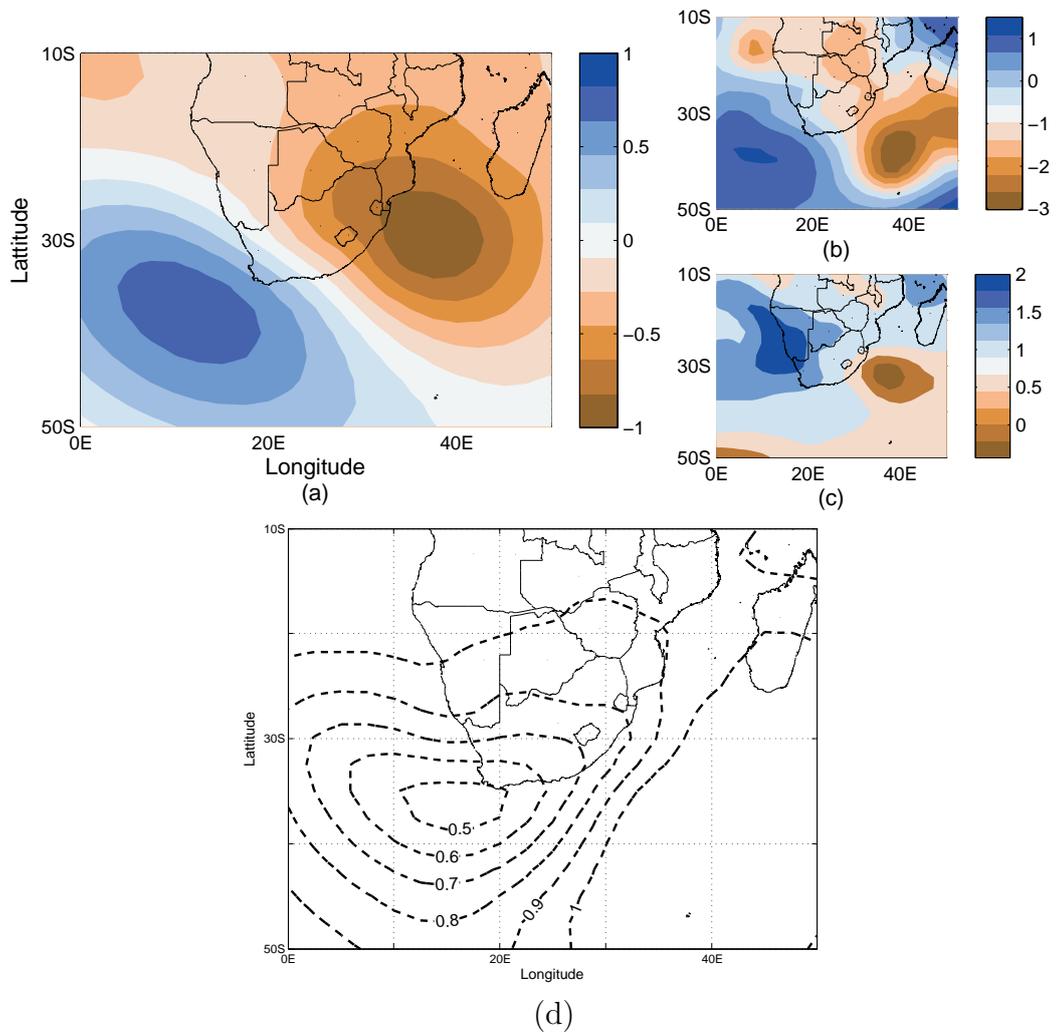


Fig. 3.3 Average anomaly pattern for CP03 (a) with (b) the anomaly with highest DOF, (c) the anomaly with lowest DOF value, while (d) shows the standard deviation for all CP03 anomalies.

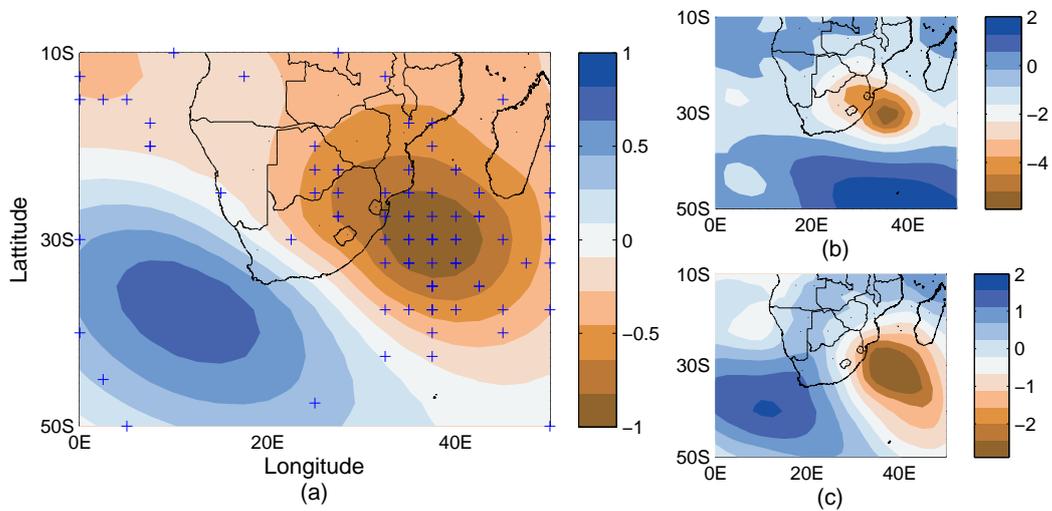


Fig. 3.4 (a) Average CP03 with (+) symbols indicating the centers of all negative anomalies (low pressures) contributing to the class. (b) & (c) show actual CP's for the dates 19/03/2007 and 30/08/2006 respectively, both of which were classified as members of the CP03 class.

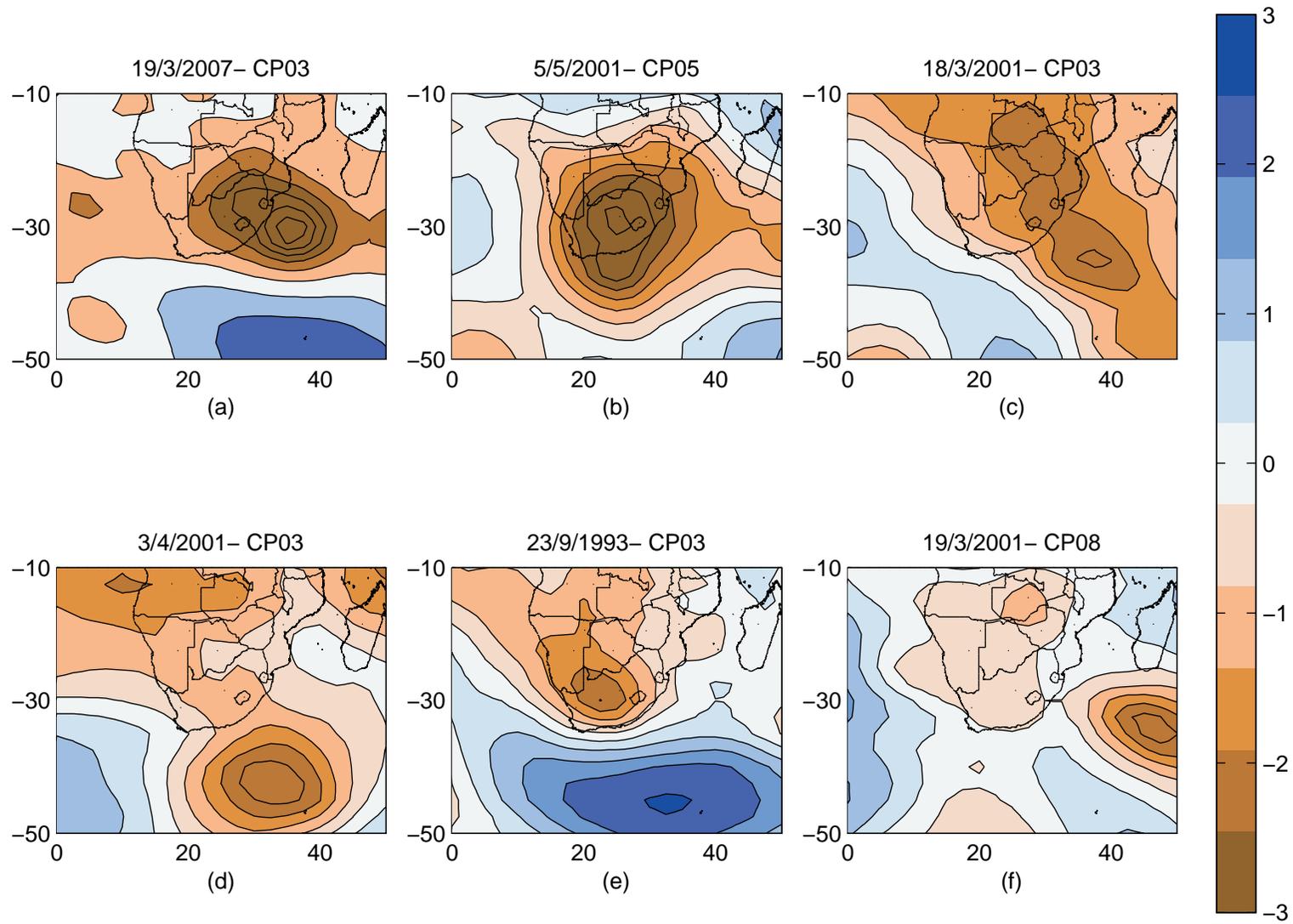


Fig. 3.5 CP's associated with the six largest significant wave heights for the dates (a) 19/3/2007, (b) 5/5/2001, (c) 18/3/2001, (d) 3/4/2001, (e) 23/9/1993 and (f) 19/3/2001.

Table 3.3 Six of the most extreme wave events on record and their associated CPs for the period 1992 to 2009.

| Fig. 3.5 | Date | CP | H_s (m) |
|----------|------------|------|-----------|
| (a) | 19/03/2007 | CP03 | 8.50 |
| (b) | 05/05/2001 | CP05 | 6.30 |
| (c) | 18/03/2001 | CP03 | 5.92 |
| (d) | 03/04/2001 | CP03 | 5.66 |
| (e) | 23/09/1993 | CP03 | 5.64 |
| (f) | 19/03/2001 | CP08 | 5.63 |

3.4 Discussion

Classifying circulation patterns is a useful tool for investigating the occurrence of certain patterns over a given region. There are many different techniques used for classifying CPs, each of which has its benefits and drawbacks (Huth *et al.*, 2008). Classification can be subjective or objective (to a degree). However the goal is always to group similar patterns into individual classes. A useful application for engineering purposes is utilizing a variable of interest to ‘guide’ the algorithm to find CPs linked to its occurrence. Bárdossy *et al.* (2002) successfully implemented this to classify CPs that explained wet and dry events in Europe.

The emphasis of the present study has been on the statistical link between atmospheric circulation patterns and extreme wave events. This is the first time the method described here has been used in this context and it has the potential to improve current methods of risk analysis. The benefit of fuzzy logic as a classification tool is that each daily CP belongs, to some degree, to all the CP classes. This is characteristic of atmospheric circulation where daily CPs form part of a continuum rather than a set of individual states as suggested by the derived CP classes (Huth *et al.*, 2008). However a potential drawback of the method is the manner in which the CPs on each day are assigned to a class (Huth *et al.*, 2008). The degree of fit (§ 3.3.2) used in this study incorporates the connectivity to a given class through AND/OR combinations of HIGH/LOW and NOT HIGH/NOT LOW anomalies as described in Bárdossy *et al.* (1995). However this technique has been successful in associating CPs with rainfall events (e.g. Bárdossy, 2010; Bárdossy *et al.*, 2002)

In the context of our case study site on the east coast of South Africa the most

frequent CPs are low and high pressure anomalies located south of the country. This can be attributed to the west–east progression of mid-latitude cyclones which frequent this area. They are major contributors to the wave climate along the South African coastline (Rossouw *et al.*, 2011). The low pressure systems can become isolated after being displaced towards the equator and can become stationary (Preston-Whyte & Tyson, 1988). These stationary cut-off lows can drive the development of extreme wave events. Table 3.2 indicates that the dominant CP that drives extreme events along the KwaZulu-Natal coastline is CP03, which is associated with abnormally low pressure to the east-southeast. CPs classified as CP03 resemble cut-off lows and the pattern agrees with speculations by Corbella & Stretch (2012d); Mather & Stretch (2012); Rossouw *et al.* (2011) concerning drivers of extreme waves. Low pressure anomalies linked to storms east to south-east of South Africa drive wind fields that direct the wave attack toward the coastline.

Callaghan *et al.* (2008); Corbella & Stretch (2013) highlight the importance of identifying independent storms for risk analysis of extreme wave events. One limitation with the methods described herein is that it is difficult to evaluate the independence of the different CPs. A particular storm in various stages of development may belong to a number of CP classes rather than a single class according to the classification scheme. Examples are cut-off lows that become detached from extra-tropical cyclones traveling west to east in the region south of the country. The process of storm development drives wave development. This suggests that it may be better to locate a specific type of CP at any location rather than a specified type of CP at a fixed location.

3.5 Conclusions

A fuzzy rule based classification method has been adapted to identifying the atmospheric circulation patterns that drive regional wave climates. The east coast of South Africa was used as a case study. The method is based on normalized anomalies in daily 700 hPa geo-potential heights. The CP classes are derived from an optimization procedure which is guided by a variable of interest, in this case wave heights. The classification shows a strong anomaly pattern east-southeast of South Africa which explains 30 – 60% of extreme wave events. This CP type explains extreme events in all seasons. However it occurs infrequently ($\sim 8\%$ of the time) and is associated with large wave heights ranging from 5.0 – 8.5 m. Frequently occurring CP classes have similar structure to mid-latitude cyclones or translational low pressure systems

Atmospheric circulation patterns & wave climates

(followed by a zone of high pressure) that occur south of South Africa (Taljaard, 1967).

The methodology discussed here appears to be new in the context of wave climate analysis and has potential for application to risk assessment studies in coastal management and engineering.

Chapter 4

On linking atmospheric circulation patterns to extreme wave events for coastal vulnerability assessments.

Abstract

Atmospheric circulation patterns (CPs) are fundamental drivers of regional wave climates. A fuzzy rule based classification algorithm has recently been developed to identify these atmospheric features. The algorithm is guided by wave heights and optimises the location, shape and strength of a set of CP classes in order to find features that drive extreme waves. This paper focuses on a method for evaluating the performance of CP classification algorithms and reducing the subjectivity in the selection of classification parameters. We suggest a method based on entropy to quantify the classification quality and provide a means to objectively define an optimal number of CP classes. We also explore the sensitivity to the temporal resolution of the data. For our case study site the entropy measure indicates that a good quality classification requires 15–20 CP classes. However regardless of the number of classes used there is a persistent, common class that explains a large proportion of extreme wave events. The methods described here contribute to developing a new framework for improved statistical wave modelling in coastal vulnerability risk assessments.

4.1 Introduction

Atmospheric features drive regional wave climates. Their occurrence and persistence control wave development including extreme wave events that can cause severe coastal erosion. Therefore atmospheric circulation patterns (CPs) are fundamental drivers of coastal vulnerability.

The physical links between ocean waves and atmospheric circulation is complex since it involves processes over a range of spatial and temporal scales. Current global wave models rely heavily on reanalysis with data assimilation to produce reasonable hind-cast wave data, but remain poor at reproducing extreme events nearshore (Caires *et al.*, 2004; Chawla *et al.*, 2013; Stopa & Cheung, 2014; Swail & Cox, 1999; Tolman *et al.*, 2002). Furthermore they are limited in their ability to predict future climate

scenarios. However event timing is generally well captured by models, so they can be used to identify the drivers of regional wave climates. Recently, the classification of synoptic scale atmospheric circulation with links to regional wave climates has provided insights that suggest that CPs can be used to statistically model regional wave climates (Corbella *et al.*, 2015; Espejo *et al.*, 2014; Pringle *et al.*, 2014). A more detailed discussion of this approach is presented in § 4.4.2.

There are several reasons why classifying atmospheric drivers and their links to waves should have an important role in coastal vulnerability and climate change assessments. It provides a natural way to identify independent storm events for risk assessment. Corbella & Stretch (2012b) identified independent events based on auto-correlation, while Li *et al.* (2013) separated independent events by requiring a minimum inter-event time, and Callaghan *et al.* (2008) manually assessed independent events based on the meteorological features associated with them. The transition of CPs between classes seems to be the most physically meaningful method for delineating independent events. Atmospheric CPs also contain important information relevant to the distributions of wave height, direction and period. For example when a particular CP type occurs the associated wave height, direction and period can be (statistically) predicted. CP classification can also be used to extend current data sets and to infill missing data (Hewitson & Crane, 2002). Finally the prediction and evaluation of climate change impacts on coastal vulnerability would be more robust if they are linked to changes in the atmospheric circulation patterns that are the basic drivers of extreme wave events.

Objective atmospheric classification algorithms have been used extensively in the past to cluster similar circulation patterns into classes representing different atmospheric states Bárdossy (2010); Bárdossy *et al.* (1995, 2002); Hess & Brezowsky (1952); Huth *et al.* (2008); Lamb (1972). Automated classification techniques that have been used to great effect include clustering analysis, principle component analysis (PCA) and self organizing maps (SOMS) (Hewitson & Crane, 2002; Huth *et al.*, 2008). While these methods have been shown to provide insight into the CPs that are typical of particular regions, links to surface weather phenomena (such as precipitation, temperature or the wave climate) are only found after the classes are derived and are not part of the class derivation itself. This study aims to elucidate the links between atmospheric CPs and coastal vulnerability events. In such cases coupling a variable of interest associated with those events to CPs after the classes have been derived may not be the most effective approach. Alternatively, if the variable of interest is used

On linking atmospheric circulation patterns to extreme wave events for coastal vulnerability assessments.

to guide the CP classification to an optimal solution we can gain more direct insights into the types of CPs that drive those events. In fact Bárdossy *et al.* (2015) have shown how two guiding variables (wave height and precipitation) used by the same classification algorithm produce significantly different results.

Pringle *et al.* (2014) were the first to apply statistical methods to finding a set of atmospheric states (or circulation patterns) with strong links to wave behaviour. This new approach to wave climate analysis may have important implications for coastal vulnerability and risk assessment. It uses a fuzzy rule based classification algorithm to identify CPs that drive wave development and builds on the approach previously introduced by Bárdossy *et al.* (1995) for linking CPs to rainfall. The classification uses objective reasoning and is guided by the wave climate data to give an optimal solution. However Huth *et al.* (2008) argue that so-called objective classifications are still subjective to some extent, such as in the selection of certain key parameters. For example in their classification study Pringle *et al.* (2014) used eight CP classes whose patterns were derived from daily data. The decision to use eight classes was subjective and the effect on the model performance was not evaluated. Nor were the effects of the selected temporal resolution of the data. The sensitivity of the algorithm to such parameters can be important. Therefore this paper addresses these issues and suggests an objective means to select the best combination of model parameters. Applications of CP based wave modelling techniques, which are currently under development for coastal vulnerability assessments, are then discussed.

4.2 Method

4.2.1 Case Study Site

The wave climate along the KwaZulu Natal coastline of South Africa is driven by a number of circulation pattern (CP) types. Those driving wave development have been cited as mid-latitude cyclones (cold fronts), tropical cyclones, and cut off lows (Corbella & Stretch, 2012d; Mather & Stretch, 2012; Rossouw *et al.*, 2011). Austral autumn and winter months (Table 4.1) experience the largest wave heights on average (Corbella & Stretch, 2012d). During these months the tracks of mid-latitude cyclones move equator-ward by a few degrees, a result of high pressure cell concentrations shifting northward (Preston-Whyte & Tyson, 1988).

Wave heights at 3 hourly intervals were obtained from wave observation buoys at

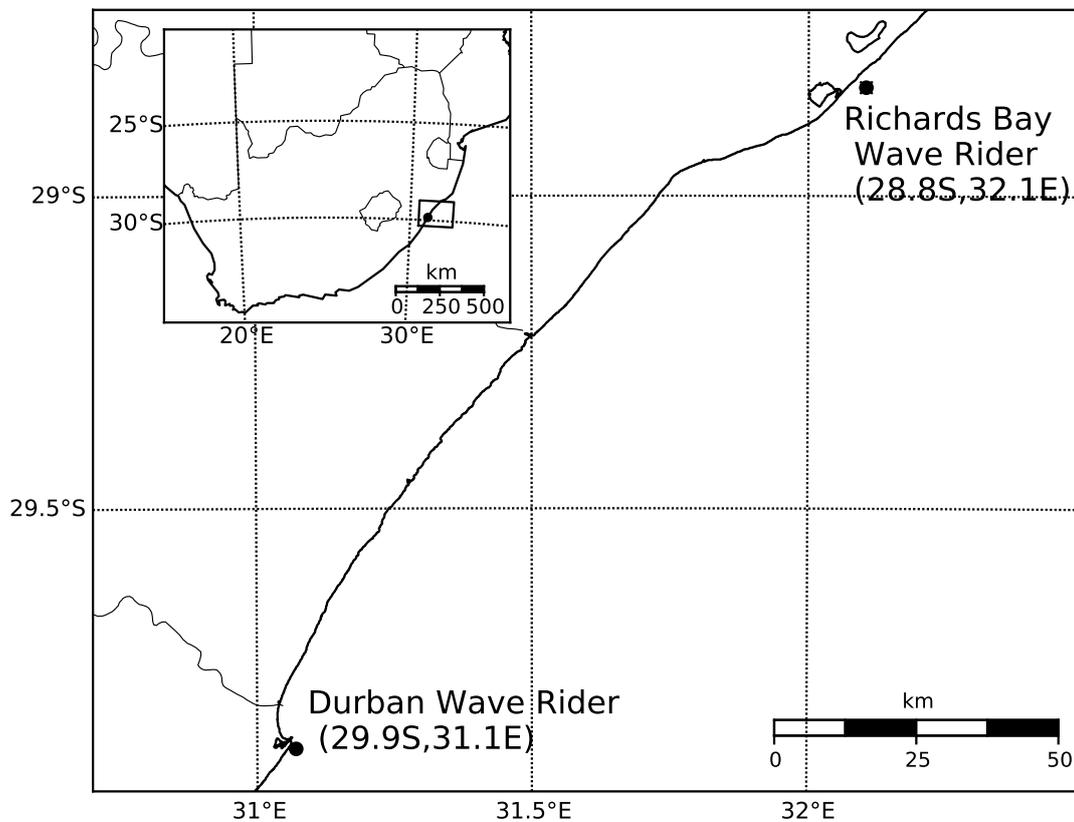


Fig. 4.1 Locations of wave observations at Durban and Richards Bay on the KwaZulu-Natal coastline of South Africa

Durban and Richards Bay (refer Fig. 4.1) for the period (1992–2009). Corbella & Stretch (2012d) found a strong correlation between wave heights from the two wave buoys. Therefore the Richards Bay Wave Rider was used to infill missing data within the Durban Wave Rider record to derive one continuous wave record.

The identification and classification of atmospheric circulation patterns used height anomalies of the 700 hPa geo-potential with a $2.5^\circ \times 2.5^\circ$ grid resolution. Pressure anomalies are defined as

$$h(i, t) = \frac{k(i, t) - \mu(i, j(t))}{\sigma(i, j(t))} \quad (4.1)$$

where $j(t)$ represents the natural annual cycle of the mean and standard deviation at location i and time t . The pressure data was from the ERA-Interim dataset (<http://apps.ecmwf.int/datasets/>) for the period 1979 – 2012. It is noted that the ERA-

On linking atmospheric circulation patterns to extreme wave events for coastal vulnerability assessments.

Interim dataset also provides reanalyzed wave data for the same period. However only the observed wave height is used herein.

Table 4.1 The allocation of months to seasons

| Season | Months |
|--------|----------------------|
| Summer | December – February |
| Autumn | March – May |
| Winter | June – August |
| Spring | September – November |

4.2.2 Optimization and Classification

Atmospheric classification that is guided by a variable of interest provides insight into the physical mechanisms driving its occurrence. For this study the significant wave height was used as the variable of interest and the focus was on extreme wave events. An extreme wave event was defined as the time from which the significant wave height exceeds 3.5 m to the time it reduces below 3.5 m. The threshold of 3.5 m was chosen because observations suggest wave heights exceeding this value are responsible for significant coastal erosion (Corbella & Stretch, 2013).

In this study atmospheric classification refers to both an *optimisation procedure*, in which a set of classes are derived, and a *classification method*, by which CPs are assigned to the classes.

The classification method herein is based on fuzzy logic and detailed descriptions of the algorithm are given by Bárdossy (2010); Bárdossy *et al.* (1995, 2002); Pringle *et al.* (2014). The use of fuzzy logic allows the algorithm to define areas of relatively ‘high’ or ‘low’ pressures.

At each time interval a degree of fit (DOF) is computed for all classes. The DOF evaluates the spatial similarity of the anomaly at a certain time to each of the classes. Subsequently the class corresponding to the highest DOF is assigned as the CP for that time.

The classes themselves are derived from an optimisation procedure that maximises an objective function linked to the significant wave height. The objective function serves as a ‘guide’ for the optimisation algorithm in order to obtain a set of clearly defined classes with statistics significantly different from the mean. The objective function is itself a function that depends on wave height (H_s) and key statistical

parameters thereof such that

$$O = f(g_1(H_s), g_2(p(H_s \geq \theta | CP(t), p(\theta))); w_1, w_2) \quad (4.2)$$

where θ is a wave height threshold, $CP(t)$ is the class assigned to day t , $p(H_s \geq \theta | CP(t))$ is the probability of an event for a given wave height threshold θ conditioned on each CP class, and $p(\theta)$ is the unconditional probability of an event for a given threshold. The values w_1 and w_2 are weights applied to the functions g_1 and g_2 . The weights reflect the relative importance and account for the range of values between g_1 and g_2 . A simulated annealing algorithm is used for the optimisation.

4.2.3 Classification Quality Measures

It is important to quantify the sensitivity of the CP classification algorithm to changes in the parameters that define the methodology. For example, what are the effects of changing the number of classes used for classification or of using different temporal resolutions in the data? A good classification contains dissimilar classes while reducing the variability within each class. Furthermore for this study it should be able to balance the degree of certainty (or confidence) with which a particular subset of CP classes explains the extreme wave events while reducing the total number of classes required.

Changing the number of CP classes has direct implications on the classification quality. If the number of classes is too many the classification is unable to capture general climatic features that describe extreme wave events. Alternatively, too few classes result in CPs that are unable to describe climatic mechanisms in sufficient detail to distinguish between them (Pringle *et al.*, 2014).

Huth *et al.* (2008) list a number of quality measures such as the explained variance (EV), pattern correlation ratio (PCR) and within type standard deviation (WSD). These measures all relate to the model's ability to maximise dissimilarity between classes while reducing the variability within classes. The explained variance incorporates a type of Euclidean distance measure using the sum of squares. It is given by

$$EV = 1 - \frac{SS_i}{SS_t} \quad (4.3)$$

On linking atmospheric circulation patterns to extreme wave events for coastal vulnerability assessments.

with

$$ss_i = \frac{\sum_{k=1}^K \sum_{j=1, j \neq i}^{N_k} \sum_{i=1}^{N_k} \sum_{l=1}^L \sum_{m=1}^M (x_{lmik} - x_{lmjk})^2}{\sum_{k=1}^K N_k(N_k - 1)}, \quad (4.4)$$

$$ss_t = \frac{\sum_{j=1, j \neq i}^N \sum_{i=1}^N \sum_{l=1}^L \sum_{m=1}^M (x_{lmi} - x_{lmj})^2}{N(N-1)}, \quad (4.5)$$

where ss_i is the mean total sum of squares within a class, ss_t is the mean total sum of squares without classification, and x_{lmik} the anomaly value at the (l, m) grid point for the i th index belonging to k th class, with K the total number of classes.

Another useful measure of the classification performance is the pattern correlation ratio. This is defined as the ratio between the mean correlation coefficient between classes and the mean correlation coefficient within the classes (Huth, 1996). It is given by

$$PCR = \frac{Pc_b}{Pc_i} \quad (4.6)$$

with

$$Pc_b = \frac{\sum_{l=1, l \neq k}^K \sum_{k=1}^K \sum_{j=1}^{N_l} \sum_{i=1}^{N_k} r(\mathbf{x}_{ik}, \mathbf{x}_{jl})}{\sum_{l=1, l \neq k}^K \sum_{k=1}^K N_k N_l}, \quad (4.7)$$

$$Pc_i = \frac{\sum_{k=1}^K \sum_{j=1, j \neq i}^{N_k} \sum_{i=1}^{N_k} r(\mathbf{x}_{ik}, \mathbf{x}_{jk})}{\sum_{k=1}^K N_k(N_k - 1)}, \quad (4.8)$$

where Pc_b is the mean correlation coefficient between fields from different classes, Pc_i is the mean correlation coefficient between fields within a given class, $r(\mathbf{x}_{ik}, \mathbf{x}_{jl})$ is the correlation coefficient between the i th index for the k th class and the j th index for the l th class, \mathbf{x} a vector containing all anomaly values, and N_k is the number of anomalies belonging to the k th class.

The within class standard deviation also provides insight into the ability of the classification to minimize variability within a class and it is given by

$$WSD = \frac{1}{K} \sum_{i=1}^K sd_i \quad (4.9)$$

where K is the total number of classes and sd_i is the standard deviation within the i th class.

The aforementioned performance measures are only able to quantify the quality of the classification with respect to one variable, i.e. pressure anomalies. However for this study it is also important to evaluate how well the classification performs in deriving CPs that are associated with extreme wave events. For this we propose the

Shannon Entropy.

Shannon Entropy as a measure of classification quality

A good classification should balance the degree to which a particular subset of CP classes explains extreme wave events while reducing the total number of classes required. Furthermore this performance measure should be based on the wave climate and be strongly coupled with the classification but used in an independent manner.

Since our objective is to successfully identify atmospheric circulation patterns that drive extreme wave events, a method that quantifies the model's performance based on this objective is required. Simply measuring the variability within and between classes offers little knowledge of how well the classification explains extreme events. Bárdossy (2010) used the objective functions themselves to identify an optimal number of CP classes. However the objective functions are not an independent measure of the classification quality since they are used in the optimization process to derive the classes.

The Shannon entropy (Shannon, 1948) provides useful information on the certainty of an outcome or the expected information quantity and is defined as

$$H = - \sum_{n=1}^N (p_n \log p_n) \quad (4.10)$$

where p_n is the probability of occurrence of event n and N is the total number of events.

Shannon entropy has found wide application. Originally used to gage the certainty of an outcome it is extensively used in information theory (Shannon, 1948). It has also been shown to be a useful tool in statistical modeling. For examples see Petrov *et al.* (2013), Gotovac & Gotovac (2009). Equation 4.10 has the following important properties particular to the present study (Shannon, 1948):

- (a) A smaller number of classes contains more information i.e. more CPs are assigned to each class. This reduces the certainty with which the classification explains extreme wave events therefore increasing the entropy,
- (b) H is always positive unless all but one p_n has the value unity, and
- (c) the entropy of two independent events is the sum of their individual entropies.

On linking atmospheric circulation patterns to extreme wave events for coastal vulnerability assessments.

The contribution of CPs to extreme events provides useful insight into the drivers of those events. This information is also unbiased towards the frequency of occurrence since it is conditioned on the occurrence of a CP given an event. Therefore it is possible to identify a class that drives extreme events and occurs frequently, as an example. However it is better to find classes that occur infrequently but also contribute to the event space since extreme wave events tend to occur infrequently. Hence it is possible to use the contribution of CPs to events defined as $P_n = p(CP|H_s \geq 3.5)$ to calculate the entropy as follows:

$$H = -\frac{1}{K} \sum_{n=1}^K P_n \ln P_n \quad (4.11)$$

where P_n is the contribution of class n to extreme events and K is the total number of classes. Equation 4.11 is a measure of the average entropy per CP class. Increasing the number of classes improves the ability of the classification algorithm to find classes driving extreme wave events. Therefore P_n for the classes driving the events should increase whereas P_n for all other classes should reduce. This increases the certainty of the CP classes driving the events thus reducing the average entropy value. However, increasing the number of classes eventually leads to diminishing returns. A larger set of classes improves the classification's ability to capture specific climatic features while simultaneously reducing its ability to capture general features.

4.3 Results

4.3.1 Previous Classification Results

Pringle *et al.* (2014) have previously applied the fuzzy rule based classification algorithm to link atmospheric CPs to regional wave climates. The KwaZulu-Natal coastline (Fig. 4.1) was used as a case study site. Their results showed that 40 – 60% of extreme wave events were associated with one type of CP. Fig. 4.2 shows the anomaly pattern for that CP and the statistics associated with the CP class are given in Table 4.2. It was noted that the high-low coupling in the pressure field drives strong winds towards the coastline that can in turn lead to large wave events.

It is important to verify whether this anomaly pattern is distinct or whether it comprises several CP types because of the low number of classes used by Pringle *et al.* (2014). In other words if the number of classes is increased above the 8 classes used by Pringle *et al.* (2014)), what are the associated spatial and statistical changes to this

CP class?

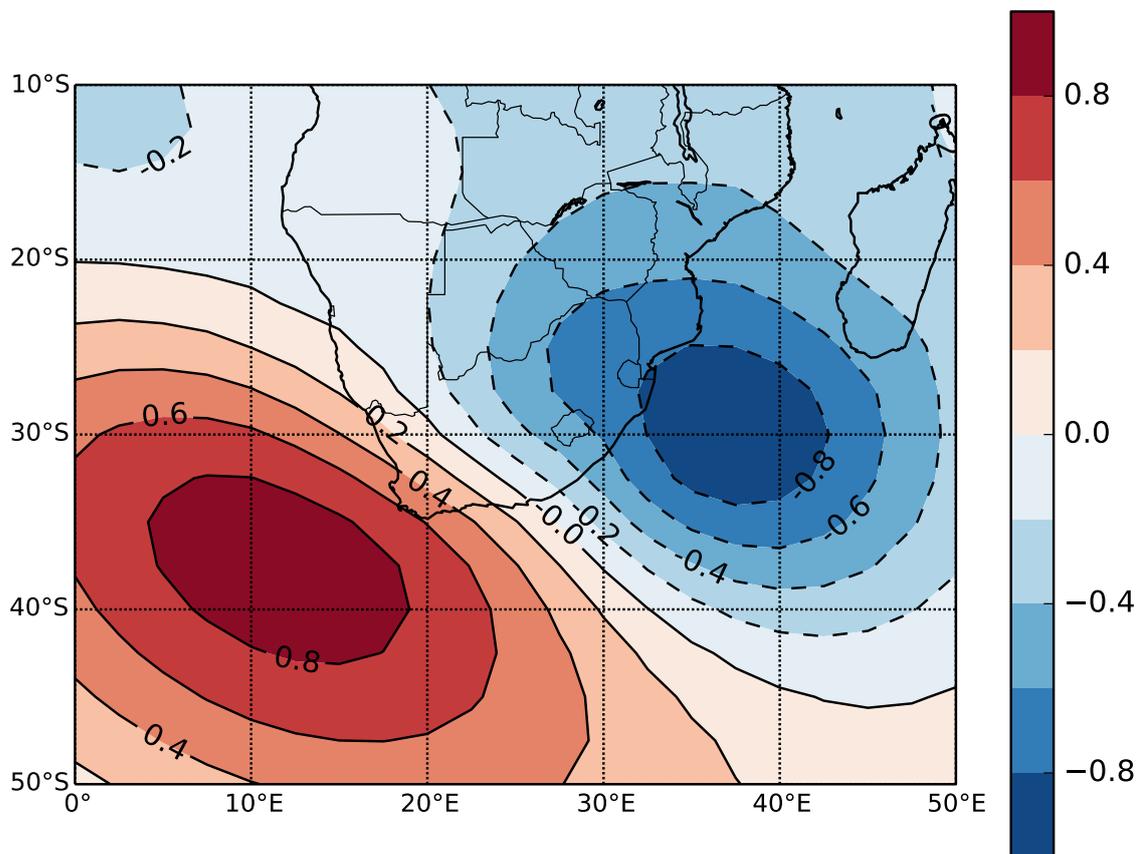


Fig. 4.2 The CP anomaly that according to Pringle *et al.* (2014) drives approximately 40–60% of extreme wave events at the case study site (Fig 4.1).

4.3.2 Classification Quality

The Shannon entropy (Eq. 4.11), PCR (Eq. 4.6), WSD (Eq. 4.9) and EV (Eq. 4.3) were used to evaluate the classification quality as discussed in § 4.2.3. The optimized CP classes were derived for the time period 2001 – 2009 and then the classification (assigning CPs to classes) was performed for the period 1990 – 2000. Fig. 4.3(a) and (b) show that the entropy and PCR provide the most insight into the performance of the classification whereas the EV and WSD provide little insight. In general the EV is shown to increase with increasing number of classes while WSD decreases. This is expected since more classes provide the algorithm with the ability to explain a larger number of atmospheric states thereby reducing the variance within each class.

On linking atmospheric circulation patterns to extreme wave events for coastal vulnerability assessments.

Table 4.2 The occurrence statistics and associated wave height statistics for the CP shown in Fig 4.2 after (Pringle *et al.*, 2014).

| Statistics | |
|---|-----|
| <i>Frequency of occurrence: $h(CP)$ %</i> | |
| Summer | 8.0 |
| Autumn | 8.0 |
| Winter | 8.1 |
| Spring | 9.1 |
| Annual | 8.3 |
| <i>Contribution to extreme events: $p(CP H_s \geq 3.5 \text{ m})$ %</i> | |
| Summer | 50 |
| Autumn | 33 |
| Winter | 64 |
| Spring | 55 |
| Annual | 48 |
| <i>Probability of an event given a CP Class : $p(H_s \geq 3.5 \text{ m} CP)$ %</i> | |
| Summer | 8.0 |
| Autumn | 12 |
| Winter | 14 |
| Spring | 4.6 |
| Annual | 9.6 |

However neither are related to the wave climate. Therefore they are unable to describe the loss of generality of the classification with the increase in the number of classes in comparison to the entropy and PCR. Fig. 4.3(d) shows that the WSD approaches a value of 1 with increasing number of classes. The classes are based on normalised anomalies within the 700 hPa geopotential therefore on average the classes should have a mean of 0 and standard deviation of 1.

In the context of previous classification techniques and their relative performance Huth *et al.* (2008) evaluated the performance of a number of different classification methods. The techniques evaluated ranged from k -means clustering, T-mode principle component analysis (PCA) and correlation based, to mention a few. Their classification was based on sea level pressures. Results showed EV values ranging from 10 – 40% and PCR values ranging from 40 – 60%. These results are comparable with those obtained for this study.

The average entropy for each class and the PCR (Fig. 4.3(a) and (b)) provide

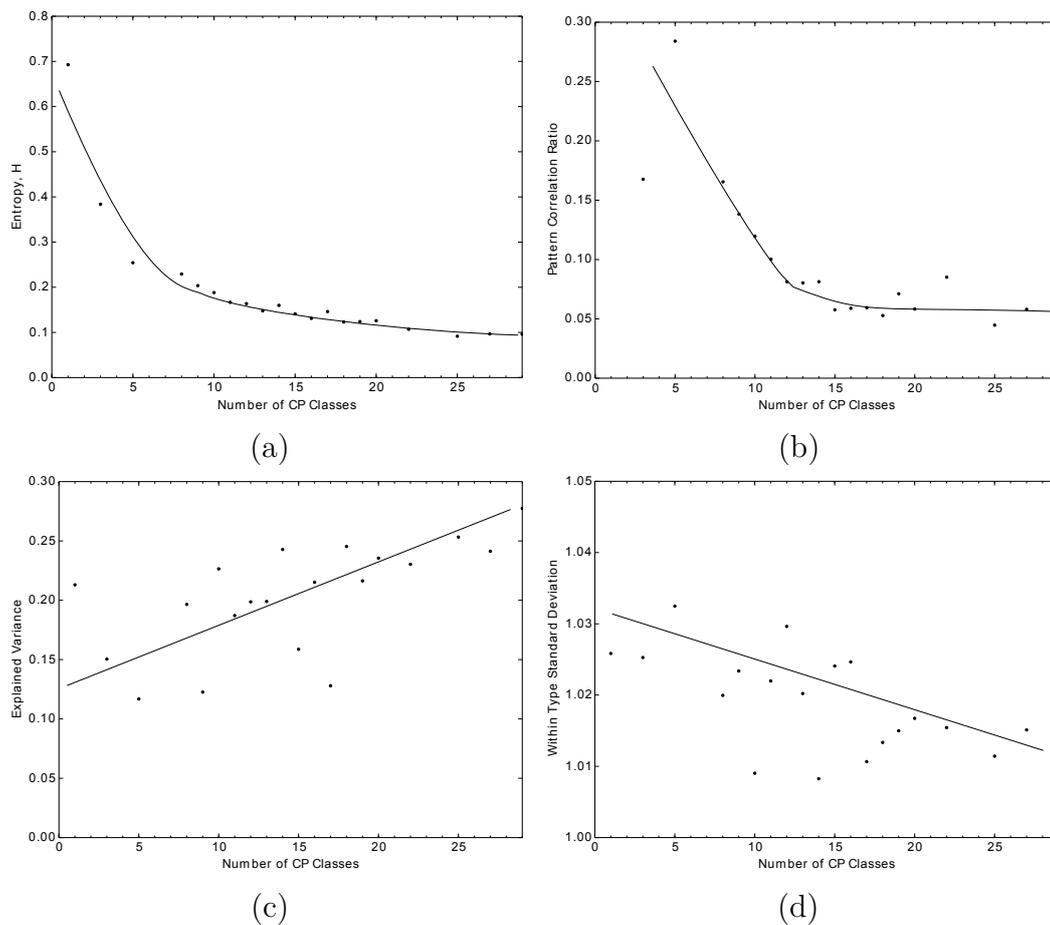


Fig. 4.3 Classification quality measures (a) Shannon entropy, (b) PCR, (c) EV and (d) WSD for classifications using different numbers of classes showing actual values (points). Solid lines indicate approximate general trends in the quality measures.

interesting insight into the model performance. Both measures show a rapid decline followed by only small changes with increasing number of CP classes. It is expected that both the entropy and PCR will decrease with increasing number of classes. The increase in classes improves the ability of the algorithm to locate a set of discrete well separated classes. Improving the correlation within classes, while decreasing the correlations between classes, results in a better classification and decreases the PCR value. It is clear that the PCR shows little change for more than about 15 classes. Similarly the entropy shows only slow changes beyond about 10 classes. This is attributed to the improvement in the model's ability to locate CPs driving extreme events. However increasing the number of classes reduces the classifications ability to capture general climate features. Therefore there is a point of diminishing returns which is evident

On linking atmospheric circulation patterns to extreme wave events for coastal vulnerability assessments.

in the small amount of information gained by increasing the number of CP classes beyond 10 to 15.

4.3.3 Sensitivity to Temporal Resolution in the data

The time interval of the pressure data was reduced from daily to 6-hourly. The relative differences in the classification statistics were evaluated with the results shown in Table 4.3. The statistics evaluated were: (a) the frequency of a particular class ($p(CP)$), and (b) the contribution of a CP class to the total number of events ($p(CP|H_s \geq 3.5)$). A negative change implies that for the higher resolution pressure data the class contributes less towards the statistic of interest. The number of classes used in the classification was 16. This number was chosen in accordance with Fig. 4.3 which shows that it produces a good quality classification. There is little improvement in classification quality for a greater number of classes. The daily and 6-hourly classifications were run using the classes derived from daily data.

The classification captures the occurrence of the CP classes well at both temporal resolutions, with less than 5 % change for all classes. Table 4.3 shows that $p(CP|H_s \geq 3.5)$ is more sensitive to changes in the temporal resolution of the pressure data. This reflects the duration of the storm events and the detail with which they are captured. Storms typically last a few days therefore a classification using 6-hourly data is able to capture significantly more detail than one using daily data.

Table 4.3 Relative differences (%) to key statistics when the CP classification was run for 1-day and 6-hour temporal resolutions and with 16 CP classes. Note that the labels of the CP classes as CP01, CP02, etc is arbitrary, but CP99 denotes an unclassified class. The wave height threshold θ used to define the extreme wave events was $\theta = 3.5$ m for this analysis.

| Statistics | CP01 | CP02 | CP03 | CP04 | CP05 | CP06 | CP07 | CP08 | CP09 | CP10 | CP11 | CP12 | CP13 | CP14 | CP15 | CP16 | CP99 |
|---|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|
| <i>Difference in occurrence frequency $p(CP)$ %</i> | | | | | | | | | | | | | | | | | |
| Summer | 1 | -1 | 0 | 0 | 0 | -1 | 0 | 2 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 |
| Autumn | 0 | -2 | 1 | 0 | 1 | 0 | 0 | 1 | 1 | 0 | 0 | 1 | -1 | 0 | 1 | 0 | 0 |
| Winter | 0 | 1 | 1 | 0 | 0 | -1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | -1 | 0 | 0 |
| Spring | 0 | 0 | 0 | 0 | 0 | 0 | -1 | 2 | 0 | 0 | -1 | -1 | -1 | 0 | 0 | 0 | 0 |
| Annual | 0 | 0 | 0 | 0 | 0 | -1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| <i>Difference in conditional probability of an event given a CP $p(H_s \geq 3.5 CP)$ %</i> | | | | | | | | | | | | | | | | | |
| Summer | 0 | -2 | 0 | 2 | -2 | -3 | -4 | 0 | -6 | 0 | 0 | -1 | -2 | 0 | 0 | 0 | 0 |
| Autumn | 0 | 1 | 0 | -5 | -11 | -16 | -6 | -3 | -15 | -6 | -3 | -2 | -25 | -4 | -4 | 1 | 0 |
| Winter | 0 | -1 | -1 | -6 | -7 | -11 | -8 | -4 | -3 | 0 | 2 | 0 | -3 | 0 | 0 | 0 | 0 |
| Spring | -1 | -3 | -7 | -2 | 0 | 0 | -8 | -5 | -8 | 0 | 0 | -2 | 1 | -1 | -2 | -2 | 0 |
| Annual | -0 | -1 | -2 | -3 | -5 | -7 | -6 | -3 | -8 | -2 | -0 | -2 | -8 | -1 | -1 | -1 | 0 |
| <i>Difference in contribution to extreme events $p(CP H_s \geq 3.5)$ %</i> | | | | | | | | | | | | | | | | | |
| Summer | 0 | -10 | 0 | 31 | -4 | -4 | -5 | 0 | -20 | 6 | 0 | 3 | -4 | 0 | 6 | 0 | 0 |
| Autumn | 4 | 1 | 6 | -2 | -3 | -6 | 2 | -1 | 0 | -2 | 2 | 4 | -5 | 0 | 0 | 1 | 0 |
| Winter | 0 | -1 | 5 | 2 | -5 | -20 | -1 | -3 | 9 | 0 | 8 | 0 | 3 | 0 | 2 | 2 | 0 |
| Spring | 1 | -8 | -7 | 10 | 0 | 0 | 14 | -4 | -5 | 0 | 3 | -6 | 5 | 4 | -2 | -4 | 0 |
| Annual | 2 | -2 | 3 | 3 | -3 | -7 | 1 | -2 | 0 | -1 | 4 | 1 | -1 | 1 | 2 | 0 | 0 |

^aCP99 is the unclassified class.

On linking atmospheric circulation patterns to extreme wave events for coastal vulnerability assessments.

4.3.4 CP Similarity

Classification schemes were carried out where the total number of CP classes was varied in the range 1 – 16. Since the lowest entropy was found for greater than 15 classes (Fig. 4.3) it was not necessary to calculate the entropy for classifications using less than 8 CP classes, and those with 2, 4, 6, and 7 classes were omitted.

The results from all the classification schemes reveal an interesting feature. Fig. 4.4 shows a CP class that is similar for all the classifications undertaken. This class is an anomalous low pressure region east of the KwaZulu-Natal coastline that drives a strong wind towards the coastline resulting in large wave conditions. Table 4.4 shows the correlation coefficients between the class described above for a particular classification scheme and all other schemes.

The average CP is very similar to that found by Pringle *et al.* (2014). Marginal differences can be noted in the strength, location and high-low coupling. According to Pringle *et al.* (2014) this class explains 40 – 60% of extreme wave events. It is important to note how this pattern becomes more refined and stronger as the number of classes increases. This suggests that there may be more than one type of CP driving the extreme wave events. However this specific class is associated with the majority of extreme events.

Table 4.4 A matrix of cross-correlation coefficients between the different classification schemes showing the persistence of a dominant class contained in all classifications. The classification was not run for 2, 4, 6 and 7 classes. The table is symmetric about the central diagonal.

| CP Classes : | 1 | 3 | 5 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 |
|--------------|------|------|------|------|------|------|------|------|------|------|------|----|
| 1 | 1 | - | - | - | - | - | - | - | - | - | - | - |
| 3 | 0.98 | 1 | - | - | - | - | - | - | - | - | - | - |
| 5 | 0.91 | 0.93 | 1 | - | - | - | - | - | - | - | - | - |
| 8 | 0.94 | 0.96 | 0.82 | 1 | - | - | - | - | - | - | - | - |
| 9 | 0.95 | 0.97 | 0.87 | 0.96 | 1 | - | - | - | - | - | - | - |
| 10 | 0.97 | 0.98 | 0.85 | 0.97 | 0.97 | 1 | - | - | - | - | - | - |
| 11 | 0.97 | 0.97 | 0.93 | 0.92 | 0.97 | 0.97 | 1 | - | - | - | - | - |
| 12 | 0.96 | 0.95 | 0.87 | 0.87 | 0.91 | 0.94 | 0.96 | 1 | - | - | - | - |
| 13 | 0.98 | 0.96 | 0.89 | 0.90 | 0.95 | 0.96 | 0.96 | 0.95 | 1 | - | - | - |
| 14 | 0.90 | 0.89 | 0.79 | 0.79 | 0.84 | 0.87 | 0.93 | 0.95 | 0.86 | 1 | - | - |
| 15 | 0.91 | 0.87 | 0.85 | 0.82 | 0.83 | 0.90 | 0.86 | 0.87 | 0.92 | 0.83 | 1 | - |
| 16 | 0.91 | 0.93 | 0.96 | 0.85 | 0.88 | 0.90 | 0.93 | 0.92 | 0.90 | 0.93 | 0.87 | 1 |

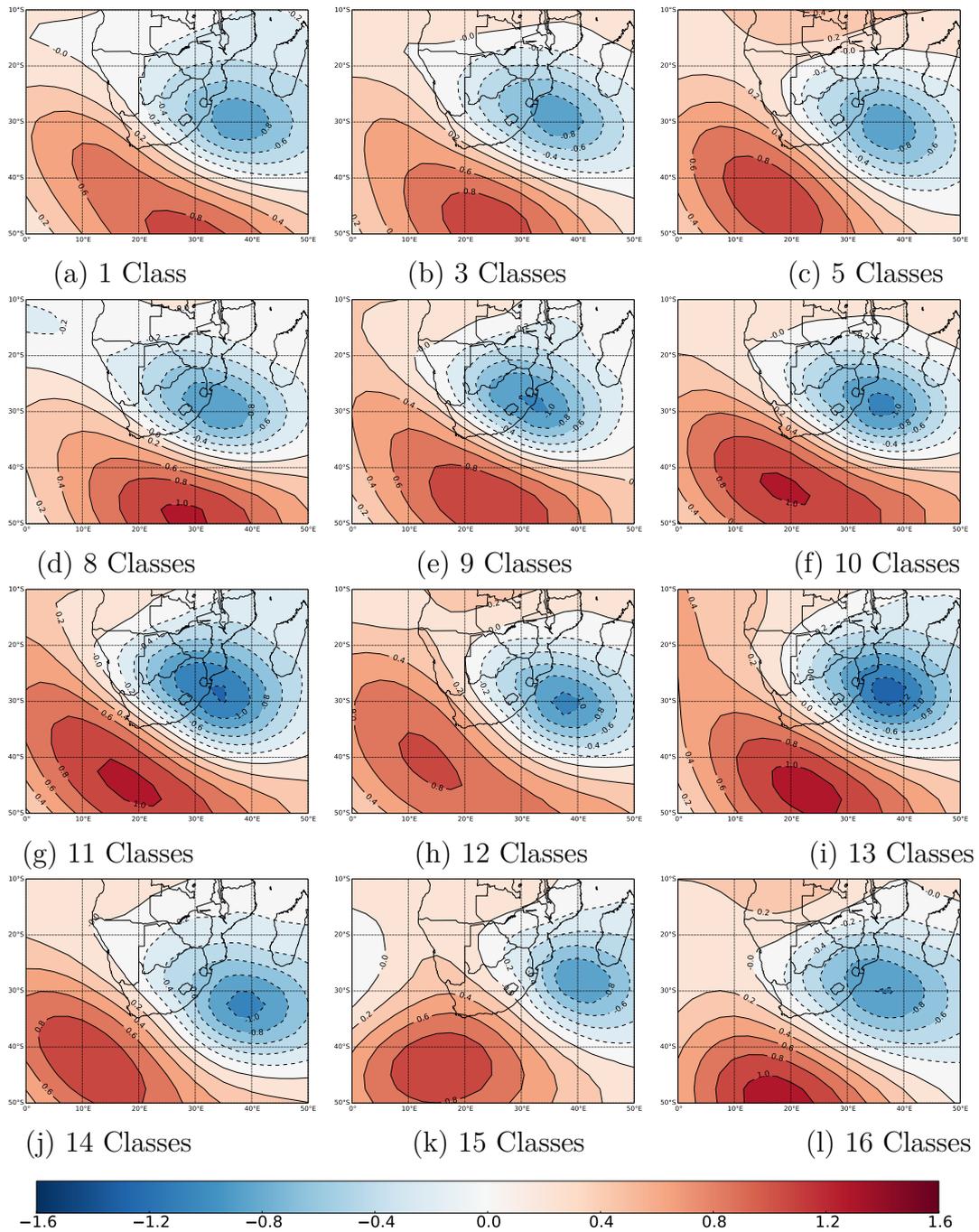


Fig. 4.4 The outcome of different classification schemes where the number of specified classes varied from 1–16. The results reveal a persistence common class that occurs in each scheme as is shown in sub-plots (a) – (l). The common class is an anomalous low pressure east of the coastline. Solid contours delineate positive (high) pressure anomalies whereas dashed contours represent negative (low) pressure anomalies.

4.4 Discussion

The aim of this study was to describe refinements in a methodology for linking wave climates to atmospheric CPs, including an objective way to select key model parameters and to quantify the performance of CP classification algorithms. This is part of an effort to develop new methods to statistically simulate regional wave climates with links to CPs for applications to coastal vulnerability assessment - this is discussed further in § 4.4.2.

4.4.1 CP links to wave climates

In diverse storm environments it can be difficult to distinguish between the atmospheric circulation patterns that drive wave development. However it has been shown that statistical links between atmospheric CPs and regional wave climates can provide strong insights into this physical coupling. Pringle *et al.* (2014) introduced an objective classification algorithm that is guided by the wave height in order to find atmospheric features with strong links to extreme wave events. However the selection of model parameters was to some extent subjective. Therefore it is important to establish the best combination of parameters that produce a realistic classification. In the present work we have proposed and tested an objective method based on the Shannon entropy to evaluate the performance of the classification.

A good objective classification is one in which the causal mechanisms of the variable of interest are captured well with the least number of CP classes required. It was found that the Shannon entropy is a good measure of classification quality. This is because a classification that realistically explains extreme wave events requires a good degree of certainty with which it is able to explain those events. For the case study site it is evident that a realistic classification requires 10 – 20 CP classes. The entropy is also shown to agree with other performance measures such as the pattern correlation ratio as described by Huth (1996). Both quality measures initially change rapidly with increasing number of classes, followed by more gradual changes. The advantage of using the Shannon entropy over other performance measures is that it couples the CP classes with their ability to explain extreme wave events, whereas the performance measures described in Huth *et al.* (2008) are based solely on the pressure values.

The duration of extreme wave events is typically in the order of hours. Therefore increasing the temporal resolution of the data improves the classification. The statistic that showed the most sensitivity to temporal resolution at the case study site was the

contribution of a CP class to the number of recorded events. This is due to the improved detail with which the classification is able to capture the extreme wave events. There were only small changes in the frequencies of the classes. This indicates that the ability of the classification to successfully identify atmospheric states is good.

A persistent CP class that drives wave events along the KwaZulu-Natal coastline is shown in Fig. B. Regardless of the number of CP classes used in the classification, an anomalous low pressure region east of the coastline causes a large percentage of extreme wave events. This pattern is similar to that found by Pringle *et al.* (2014). There are minor differences in the strength, location and in the high-low coupling. This has implications for coastal vulnerability and risk analysis for the region. The CP class shows that low pressure regions at a specific location are likely to produce extreme wave events. However, one aspect that requires further research is that the classification is currently not able to provide details of how the low pressure anomalies move towards this location. For example a cut-off low and a mid-latitude cyclone that move into this region will be classified as in the same class. This could be a source of error when using this methodology for identifying independent events.

4.4.2 Applications to Wave Modelling

The use of CPs to model waves is a new and interesting approach to coastal vulnerability assessment. It differs from purely statistical models by preserving links to the physical drivers of waves while still not needing to explicitly model complex physical processes over wide spatial and temporal scales.

Current global wave models suffer from the following drawbacks (Caires *et al.*, 2004; Swail & Cox, 1999):

1. Accurately estimating extreme wave heights is limited by the accuracy (or lack thereof) of wind field inputs, which in turn is a function of model grid resolution.
2. Without the benefit of hindsight through re-analysis and data assimilation, global wave models do not yet provide a means to accurately assess future wave climate scenarios.
3. Site specific wave climate estimation generally requires an additional downscaling process.

Statistical methods with links to atmospheric CPs can provide solutions to these problems (Pringle & Stretch, 2015).

On linking atmospheric circulation patterns to extreme wave events for coastal vulnerability assessments.

1. They are based on observed wave data and have been shown to provide insight into complex wave climates by elucidating the various types of weather systems that are associated with specific wave event characteristics (Corbella *et al.*, 2015; Espejo *et al.*, 2014; Pringle *et al.*, 2014).
2. CP sequences can be modelled as a Markov process to form the basis for synthetic wave climate simulation. This is similar to the methodology proposed by Bárdossy & Plate (1991) but in their study it was applied to precipitation modelling.
3. CPs have been demonstrated to be useful for downscaling from regional climate models to site specific locations (e.g. Bárdossy & Pegram, 2011; Camus *et al.*, 2014).

Developing CP-based wave climate simulation requires accurate descriptions of the dependence structure between wave height, period and direction for each CP class (see e.g. Corbella *et al.*, 2015). Thence, given a sequence of simulated CPs, it is possible to simulate a synthetic wave climate that can be used for coastal vulnerability assessment or design applications. This methodology is currently under development.

4.5 Conclusions

This paper focuses on a method for evaluating the performance of CP classification algorithms and reducing the subjectivity in the selection of classification parameters. The algorithm discussed here was first introduced by Pringle *et al.* (2014) for use in wave climate analysis and simulation and was therefore strongly linked to wave statistics. The Shannon entropy has been proposed as an objective measure for selecting the optimal number of classes and other parameters because it links both the CPs and their ability to explain extreme wave events.

For the case study site in KwaZulu-Natal, South Africa, the classification quality is highest for 15 – 20 CP classes based on the Shannon entropy.

The sensitivity of the classification to temporal resolution is relatively low. However, the occurrence of a specific CP given an extreme wave event can be sensitive to the increases in temporal resolution of the pressure data. This can be attributed to an improvement in the classification method's ability to capture the details of extreme wave events.

Classifications with different numbers of CP classes have all shown the presence of a similar anomaly pattern that drives a large percentage of extreme wave events, namely a region of low pressure east-southeast of the KwaZulu-Natal coastline.

The methods described here for linking atmospheric circulation patterns to extreme wave events provide a potentially useful new framework for coastal vulnerability assessments that should be generally applicable to any geographical region. Relating extreme wave events directly to their physical drivers should improve the robustness of statistical wave models and thereby facilitate their use for design and vulnerability assessments, and the assessment of future climate change effects.

Chapter 5

Assimilation of ocean wave spectra
and atmospheric circulation
patterns to improve wave modelling

Abstract

The modelling of waves associated with extreme events is fundamental to coastal engineering design and coastal vulnerability assessments. The storm systems that drive extreme wave events are associated with specific atmospheric circulation patterns (CPs). In this paper the link between these circulation patterns and wave spectra is explored as a means to improve wave modelling in engineering applications. The methodology involves partitioning wave spectra into low frequency swell and locally generated wind waves. The origin(s) of the swell waves can then be estimated in order to link them to the atmospheric circulation pattern(s) that generated them. A method based on fuzzy logic and fuzzy sets is used to identify and classify the atmospheric circulation patterns. Finally the spectral characteristics associated with specific circulation patterns can be obtained. The methodology is tested using a case study on the east coast of South Africa. The atmospheric circulation patterns driving low frequency swell events resemble those previously identified as the dominant drivers of significant wave events in the region. The link between wave spectra and CPs can be used to study the impacts of specific CPs on the coastline. For example the spectra associated with swell produced by tropical cyclones can be used to model their potential impacts. This new methodology may improve the inputs to spectral wave models and aid studies of climate change impacts. It may also help in identifying statistically independent storm events and improve multivariate statistical models of such events.

5.1 Introduction

The relationship between atmospheric features and ocean wave energy spectra is complex. However, understanding this relationship may improve the application of spectral wave models in coastal vulnerability assessments and in the analysis and design of coastal infrastructure.

Atmospheric states can be classified into distinct circulation patterns (CPs) that provide insights into regional wave climates (Camus *et al.*, 2014; Pringle *et al.*, 2014). The insights can be used to improve statistical models of extreme ocean waves and even to develop a means to simulate long term wave records (Espejo *et al.*, 2014). Direct measurements of waves are often constrained by budgets that allow only short or intermittent data sets. In contrast atmospheric pressure data is globally available over the past century. This paper focuses on a new method of developing links that can help to improve the inputs required for spectral wave models such as SWAN (Booij *et al.*, 1999). Spectral wave models use the wave action balance to perform wave transformations. This process requires the specification of the wave spectrum as a boundary condition. The spectral shapes defined by the default parameter values in spectral wave models may be inappropriate for waves driven by a particular forcing. The aim of this paper is to provide a method of estimating typical spectral characteristics that are applicable to specific forcing mechanisms. For example if the effects of a cyclone are to be modelled the proposed method can identify which spectral characteristics are appropriate for waves produced by that forcing mechanism.

In addition the method provides a means of identifying statistically independent events associated with synoptic scale features. For statistical analyses of extrema a set of independent events generally needs to be identified. Traditionally, independent storm events are defined from significant wave height data using a wave height threshold (e.g. Callaghan *et al.*, 2008; Corbella & Stretch, 2013; Mendez *et al.*, 2008). Storm events are assumed to begin when the wave height threshold is exceeded and to end when the wave height falls below and remains below the threshold for a specified time period. The minimum inter-event time period is a subjective choice but may be estimated using autocorrelation (e.g. Corbella & Stretch, 2013). This classical method has numerous flaws, the most significant of which is that it does not distinguish between wave events driven by different meteorological systems, each of which contains its own set of statistics. For example if a tropical cyclone occurs within the user defined inter-event time after a cut-off low both independent events will be recorded as a single event. The classical method would therefore produce an incorrect statistical distribution by containing two independent events within one. By partitioning the wave spectra and cross assigning the wave partitions we can define actual independent events without requiring a subjective inter-event time period.

In this paper we discuss a method of estimating the spectral characteristics of waves caused by various forcing mechanisms using a partitioning and swell tracking algorithm

and then linking them to atmospheric CPs. We initially describe the proposed methods of linking wave spectrum characteristics to the CPs (§ 5.2). Results from our case study site at Durban, South Africa are then presented (§ 5.3). Finally we discuss the results and the limitations of our approach (§ 5.4).

5.2 Methods

In order to link wave spectral characteristics to CPs we need to estimate the origin(s) of the waves. The waves produced near the coast are typically forced by local winds that do not have a long enough fetch to develop a significant amount of wave energy. For this reason we are interested in swell waves that develop far offshore and are driven by large scale circulation patterns. These swell waves generally have higher energies and are of particular interest for coastal engineering applications.

To estimate the origin of swell waves we first have to ensure that we are only considering the swell component of the total wave energy. The energy spectra are therefore partitioned into unique swell and wind waves by means of a partitioning algorithm (§ 5.2.2). The spectral partitions are then cross assigned into a collection of independent swell events (§ 5.2.3). Using the linear wave theory deep water wave dispersion relationship and spherical geometry the swell origins can be estimated (§ 5.2.3). The origins of the events are then grouped by location or spectral characteristics (§ 5.2.3). The grouped origins are then used to search the global geo-potential height data during the times when the swell is estimated to have been produced (§ 5.2.5). The circulation patterns within the origin groups are then averaged into a characteristic circulation pattern for those swell events.

5.2.1 Case study site

The province of KwaZulu-Natal on the east coast of South Africa has two relatively long records of wave data measured at Durban and Richards Bay (Fig. 5.1). Descriptions of the wave characteristics for this region, including directional energy spectra, are given by Corbella & Stretch (2012d, 2014b). The spectral wave data are derived from wave recording buoys that are listed in Table 5.1. The wave recording buoys sample at a rate of 1.28 Hz and the data sets have spectral data available at 3-hour, 1-hour and 0.5-hour intervals.

Data for geo-potential heights that are required to investigate atmospheric circu-

lation patterns were obtained from the ERA-Interim data set (<http://apps.ecmwf.int/datasets/>) for the period 1979 – 2013 at 6 hourly intervals. Circulation patterns are defined in this study as normalized anomalies derived from the 700 hPa geopotential with a grid resolution of 2.5° ($10^\circ\text{S } 0^\circ\text{E} - 50^\circ\text{S } 50^\circ\text{E}$). The 700 hPa geopotential anomalies indicate regions of relatively high and low pressures and can be used to infer wind fields without accounting for boundary layer effects (Bárdossy *et al.*, 2015). Furthermore they are less noisy than those based on surface pressures which facilitates the automated CP classification procedure.

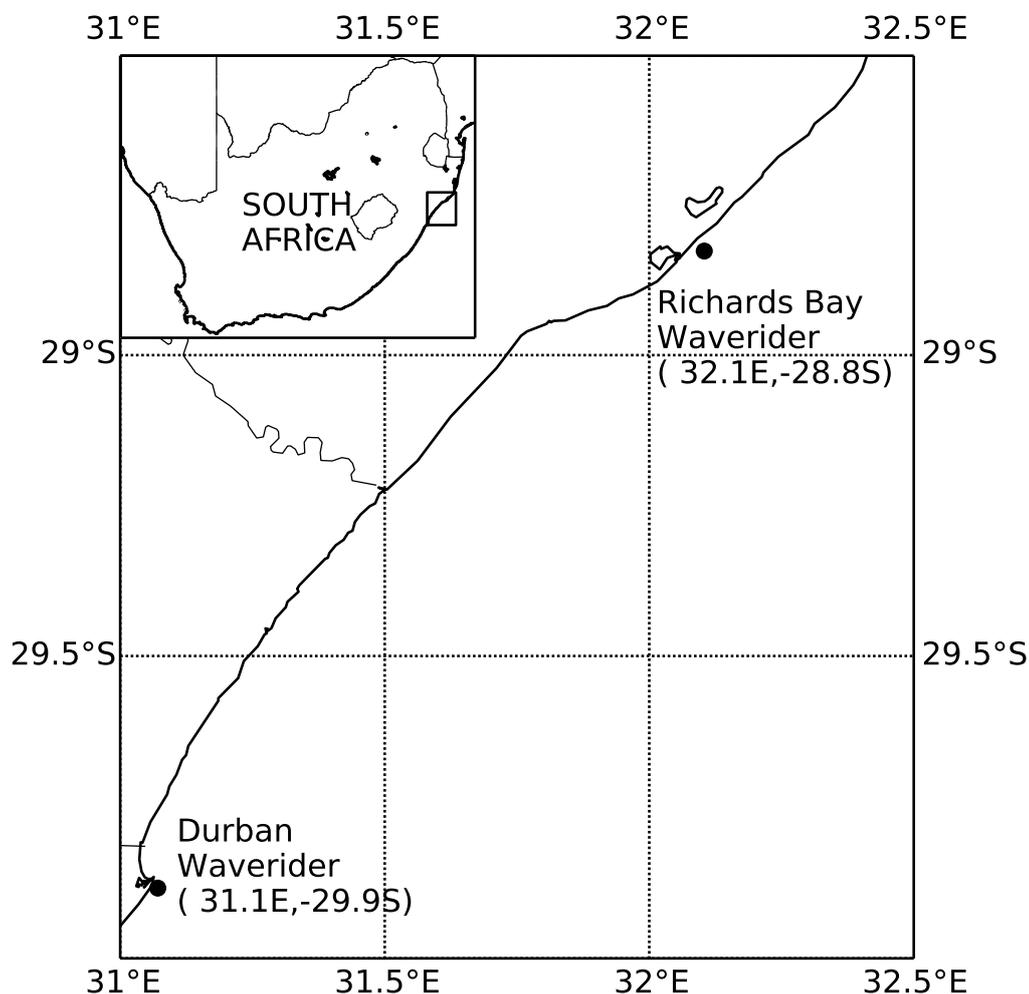


Fig. 5.1 Map of South Africa showing KwaZulu-Natal with locations of the wave recorders.

Wind data is required to identify the locally generated wind wave components of the wave energy spectra. Durban and Richards Bay both have land based wind

Assimilation of ocean wave spectra and atmospheric circulation patterns to improve wave modelling

recording instruments for their port control. Durban and Richards Bay wind data was available from 31/05/2002 to 31/07/2013 and 19/08/1993 to 31/07/2013 respectively. It should be noted that coastal wind gradients may produce dramatically different wind velocities between offshore and onshore locations. In this case the buoys are within 2 km of the wind recording instruments and are a representative measurement. Since the data period is restricted by the availability of the wave data (Table 5.1) Durban was only analysed for approximately 6 years and Richards Bay was analysed for approximately 16 years.

Table 5.1 Historical wave recording instruments at Durban and Richards Bay, their operating periods and water depth (from Corbella & Stretch (2014b))

| Instrument | Date | Depth (m) |
|-----------------------|-------------------------|-----------|
| Durban | | |
| Directional Waverider | 23/08/2007 – 30/04/2013 | 30 |
| Richards Bay | | |
| 3D Directional Buoy | 11/06/1997 – 14/10/2002 | 22 |
| Richards Bay | | |
| Directional Waverider | 08/11/2002 – 30/04/2013 | 22 |

Mid-latitude cyclones (with associated cold fronts), coastal lows, cut-off lows and tropical cyclones have been cited as the main swell producing mechanisms along the KwaZulu-Natal coast (Corbella & Stretch, 2012d; Mather & Stretch, 2012; Rossouw *et al.*, 2011). The reader is referred to Hunter (1987); Preston-Whyte & Tyson (1988) and Taljaard (1995) for a detailed description of South African weather systems. Coastal lows form closer to the coast than cut-off lows, mid-latitude and tropical cyclones. Therefore they are typically associated with smaller wave heights and shorter periods. Mid-latitude cyclones that traverse west to east in the region south of the country, drive long period southwesterly – southeasterly swell waves and occur during the austral winter months. Cut-off lows are deep, low pressure systems that become cut-off from the west–east moving mid-latitude cyclones (Preston-Whyte & Tyson, 1988). Their persistence can produce large swell waves with long periods. Tropical cyclones produce long period waves typically from the north-east or east-north-east and occur during late austral summer/early autumn months. However Mather & Stretch (2012) argue that tropical cyclones that become stationary south of Madagascar can drive severe wave conditions resulting in extensive coastal erosion. Seasons are defined in Table 5.2.

The wave data from Durban and Richards Bay will be used to demonstrate a

method of linking spectral wave data to the above-mentioned circulation patterns.

Table 5.2 The allocation of months to seasons

| Season | Months |
|--------|----------------------|
| Summer | December – February |
| Autumn | March – May |
| Winter | June – August |
| Spring | September – November |

5.2.2 Partitioning of spectral wave energy

The partitioning of spectral wave data is the separation of spectral energy into swell energy and wind-wave or sea energy. The swell energy is usually characterised by waves produced large distances away from the coast and have longer periods and higher energy than local wind-waves. The proposed method is focussed on the origin of the more energetic swell waves. In order to estimate the origin of specific swell waves one must first remove the locally formed wind wave energy and partition the swell waves that are produced from different forcing locations. Numerous authors have taken an interest in developing partitioning algorithms (e.g. Gerling (1992); Hanson & Phillips (2001); Portilla *et al.* (2009); Voorrips *et al.* (1997)) and Aarnes & Krogstad (2001) provide a general review of these algorithms. The algorithm implemented in this paper is a mixture of those proposed in Hanson & Phillips (2001) and Voorrips *et al.* (1997).

Direct measurements of directional wave spectra $S(\omega, \theta)$, where ω is frequency and θ is direction, are often not archived because they are either considered unnecessary or impractically large to store. Most available spectral data is therefore limited to 1-dimensional frequency spectra $S(\omega)$. Applying the partitioning algorithm to two-dimensional spectra is similar to applications with one-dimension spectra. The reader is referred to Hanson & Phillips (2001) for a detailed description of the partitioning algorithm. Here we provide only a brief summary of this partitioning algorithm.

Initially the spectral peaks are isolated and separated using a steepest ascent algorithm. A number of these isolated peaks will contain wind-wave peaks and these are identified by a wave age criterion. All the peaks that satisfy the criterion are considered to be local wind-waves and are removed from the partitioned data. The remaining peaks are all swell energy but may belong to the same swell system. Mutual swell peaks are combined if they satisfy a peak separation criterion. The peak

Assimilation of ocean wave spectra and atmospheric circulation patterns to improve wave modelling

separation criterion is satisfied if the peaks are within a threshold distance of each other. The threshold contains a calibration factor κ which may be unique to particular data sets. Finally any partitions whose energy falls below a specified energy threshold ($A/(\omega_p^4 + B)$) are removed, where A and B are calibration factors.

Partitioning statistics

Partitioning statistics are required for the tracking algorithm (§ 5.2.3). The relevant partitioning statistics are a partitioning identification number, the observation time (t), the total energy (E), the significant wave height (H_s), the mean (ω_m) and peak frequency (ω_p) and the mean direction (θ_m).

The total spectral energy is defined as

$$E = \int_{\omega} \int_{\theta} S(\omega, \theta) d\theta d\omega, \quad (5.1)$$

The significant wave height is defined as

$$H_s = 4\sqrt{E}, \quad (5.2)$$

and the mean frequency and mean direction are defined as

$$\omega_m = \frac{E}{\int_{\omega} \int_{\theta} S(\omega, \theta) \omega^{-1} d\theta d\omega}, \quad (5.3)$$

$$\theta_m = \arctan \left(\frac{\int_{\omega} \int_{\theta} S(\omega, \theta) \sin \theta d\theta d\omega}{\int_{\omega} \int_{\theta} S(\omega, \theta) \cos \theta d\theta d\omega} \right), \quad (5.4)$$

respectively.

5.2.3 Swell tracking

Cross Assignment of Wave Partitions

It cannot be assumed that all the partitioned spectra are uncorrelated and many swell partitions may represent only one wave system. These dependent events need to be collected by cross assignment criteria. Although both Hanson & Phillips (2001) and Voorrips *et al.* (1997) provide cross assignment criteria neither have any theoretical reason to be preferred over the other. The cross assignment criteria, as with other partitioning criteria, are used mainly as a means to calibrate the cross assignments.

For these reasons we create our own criteria, the bounds of which are determined iteratively.

Two partitions, A and B , are cross assigned if they satisfy the following criteria:

1. $\frac{|E_A - E_B|}{E_A} < 0.2$
2. $|\theta_{mB} - \theta_{mA}| < 20^\circ$
3. $\frac{|\omega_{mA} - \omega_{mB}|}{\omega_{mA}} < 0.015$

Each partition is compared to the 6 consecutive partitions. The comparison of the 6 consecutive partitions is based on the assumption that it is unusual for more than 3 swell sources to be identified in any one recording period. The latter has never occurred in the available case study data. This is however a function of the specific data set and the comparison length may need to be changed for other locations.

Swell origin identification

The swell source may be estimated from linear wave theory and the partitioning statistics. The group velocity, $C_g = g/4\pi d$, can be rewritten in terms of frequency and travel time to find the distance d to the wave origin, namely

$$d = \frac{g}{4\pi \left(\frac{d\omega}{dt}\right)}. \quad (5.5)$$

Assuming that the frequency at the observation point is increasing over time and that the wave origin time, t_0 , can be calculated at $\omega = 0$ we can calculate the wave origin time as

$$t_0 = -b \left(\frac{d\omega}{dt}\right)^{-1}, \quad (5.6)$$

where b is the ω intercept at $t = 0$. The rate of change of the frequency ($\frac{d\omega}{dt}$) is calculated from the best fit line between the observation time (t) and mean frequency (ω_m) of the cross assignment groups.

The distance to the swell origin can then be used with spherical geometry to estimate the coordinates of the swell source. The source latitude and longitude are given by Bartsch (1974) as:

$$\begin{aligned} \phi &= \sin^{-1}(\sin \phi_0 \cos \theta_d + \cos \phi_0 \sin \theta_d \cos \vartheta), \\ \varphi &= \varphi_0 + \sin^{-1} \left(\frac{\sin \theta_d \sin \vartheta}{\cos \phi} \right), \end{aligned} \quad (5.7)$$

Assimilation of ocean wave spectra and atmospheric circulation patterns to improve wave modelling

where ϕ_0 is the observation latitude, φ_0 is the observation longitude, ϑ is the mean direction of the cross assignment group, $\theta_d = d/R_E$ is the angular distance from the swell origin and R_E is the radius of the earth.

Grouping of independent swell events

Before the characteristic circulation patterns can be assigned to swell origins the cross assignment groups have to be collated into swell that have similar points of origin. This is done on the assumption that waves formed at similar locations will have similar forcing systems. To investigate the plausibility of this assumption we also carried out an additional alternative grouping based on the wave characteristics.

The spacial groupings are defined by segments of circles extending from the wave recording instrument into the ocean. The segment size may be defined at a user's discretion. Here we use segments of 22.5° divided into intervals of 5° . Fig. 5.2 illustrates a segment of independent swell events.

The grouping by wave characteristics is defined in terms of significant wave height, peak period and directions. All swell events that fall within a similar range of these characteristics are grouped together. If the assumption that similar wave events are formed by similar CPs is correct, then the two grouping methods should have a strong correlation. Fig. 5.2 shows an example comparing the two methods. The small blue dots are the swell sources grouped by location and the larger black dots are the swell sources grouped by wave characteristics. The two methods of grouping show similar results. The wave characteristics grouping contains less points as it is limited to a specific range of wave heights, whereas the location grouping contains all the swell events within a segment irrespective of wave height.

5.2.4 Atmospheric classification

Automated classification provides significant insight into high dimensional problems. Different states are grouped together through a clustering technique such as k mean, principle component analysis or artificial neural networks (Hewitson & Crane, 2002; Huth *et al.*, 2008). However in the aforementioned techniques, links to surface weather phenomena are only made after the classes (or groups) are found (e.g. Camus *et al.*, 2014). The classification algorithm used in this study was first developed by Bárdossy *et al.* (1995) and adapted by Pringle *et al.* (2014) to incorporate wave climate analysis. The algorithm consists of two parts, firstly a set of CP classes are derived using an

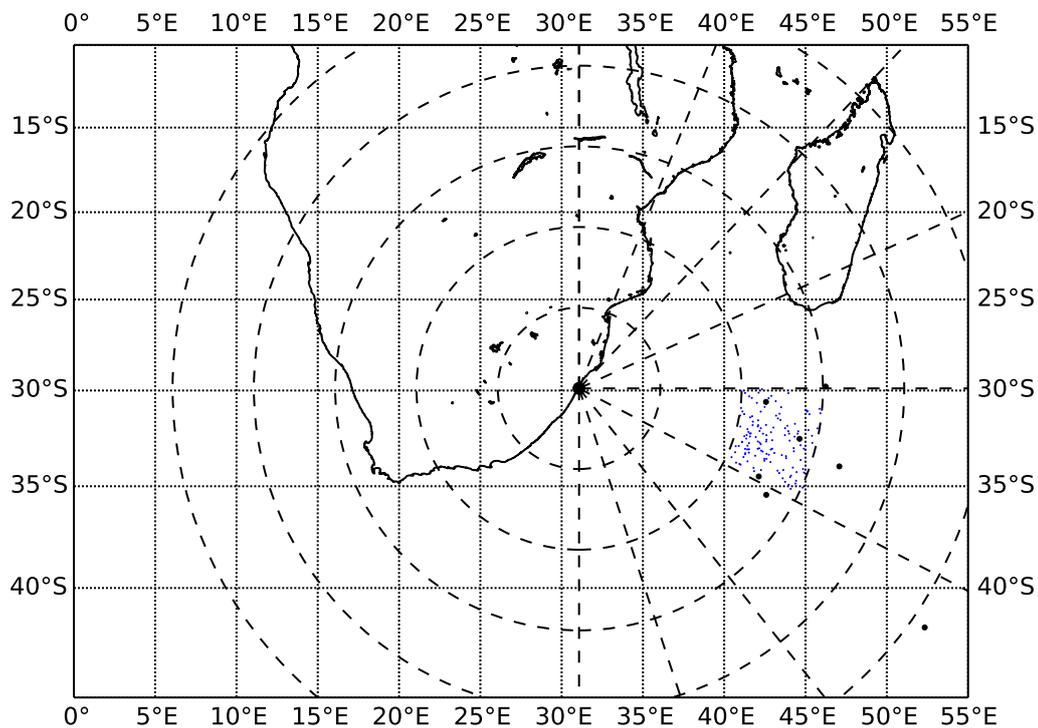


Fig. 5.2 A map of the Indian Ocean with the dashed lines showing the grouping segments. The small blue dots show a spatial grouping of swell sources and the larger black dots show a grouping by wave characteristics. The grouping by wave characteristics has the following parameters: $1.5\text{m} < H_s < 2.0\text{m}$, $10\text{s} < T_p < 12\text{s}$, $90^\circ < \theta < 135^\circ$. The large black dot indicates the location of the Durban wave recording buoy

optimization procedure and secondly CP anomalies at each time are assigned to a particular class. The optimization procedure is guided by the significant wave height (H_s) to an optimal solution using a set of objective functions and simulated annealing. It follows that the classes derived in this supervised manner will have strong links to wave behaviour and should provide insight into the causal mechanisms of regional wave climates (Pringle *et al.*, 2014).

The classification method herein is based on fuzzy logic and fuzzy-sets after Zadeh (1965). The use of fuzzy logic allows the algorithm to quantify the meaning of statements such as ‘strong low pressure’ (Bárdossy *et al.*, 1995). Details of the method and its performance can be found in (Bárdossy, 2010; Bárdossy *et al.*, 1995, 2002; Pringle *et al.*, 2014; Stehlik & Bárdossy, 2003). Only a brief overview is given here.

The CP classes are defined as a fuzzy rule which consists of spatially distributed fuzzy numbers that correspond to fuzzy-set membership functions. The j^{th} class is

Assimilation of ocean wave spectra and atmospheric circulation patterns to improve wave modelling

described by the vector $\mathbf{v}_j = [\mathbf{H}(\mathbf{1}, \mathbf{j}), \dots, \mathbf{H}(\mathbf{K}, \mathbf{j})]$ for all grid points $(1, \dots, k, \dots, K)$. The matrix \mathbf{H} contains all the rules $(1, \dots, j)$. To each grid point k one of the following fuzzy-set membership functions can be assigned based on the anomaly value (Pringle *et al.*, 2014):

- Fuzzy set number 0 - any type of anomaly,
- Fuzzy set number 1 - strong positive anomaly,
- Fuzzy set number 2 - weak positive anomaly,
- Fuzzy set number 3 - weak negative anomaly and
- Fuzzy set number 4 - strong negative anomaly.

A degree of fit (DOF) is then assigned to each class for the CP anomaly at time t . This means that the anomaly on any given day can belong to a degree to all classes. However the class with the highest DOF value is assigned as the CP for that particular day.

The classes are derived through an optimization procedure. The optimization is based on a set of objective functions. It is through the objective functions that the wave height is incorporated (Pringle *et al.*, 2014). The first objective function corresponds to extreme wave events. An event is defined as the time from when $H_s \geq 3.5\text{m}$ to the time $H_s < 3.5\text{m}$.

$$O_1(\theta) = \sqrt{\sum_{t=1}^T (p(H_s \geq \theta | CP(t)) - \bar{p})^2} \quad (5.8)$$

where θ is a predefined threshold, T is the total number of days, $p(H_s \geq \theta | CP(t))$ is the conditional probability of an extreme event for a given CP on a day t , \bar{p} is the unconditional probability of exceedance for all days in period T . Two different thresholds were considered : $\theta_1 = 3.5 \text{ m}$ and $\theta_2 = 2.5 \text{ m}$). The two thresholds force the algorithm to consider CPs driving both extreme wave events and medium size wave events.

The second objective function corresponds to the average significant wave heights and is defined as

$$O_2 = \sum_{t=1}^T \left| \ln \left(\frac{H_s(CP(t))}{\overline{H_s}} \right) \right| \quad (5.9)$$

where $H_s(CP(t))$ is the mean significant wave height at time t with a given CP class and $\overline{H_s}$ is the mean daily wave height without classification.

The optimization procedure maximizes a linear combination of Equations 5.8 and 5.9. This implies that the optimization aims to define a set of distinct classes based on the wave height that have statistics significantly different from the mean (or unclassified case).

5.2.5 Linking swell events to circulation patterns

The circulation patterns associated with storm swell energy are obtained by ensemble averaging the anomalies that occurred at the estimated times of the swell origins for each grouping of independent swell events. This average (or characteristic) CP can then be plotted and compared with the location of the inferred grouped swell origins. § 5.3 presents the results for all the swell groupings.

The characteristic CPs are then classified according to § 5.2.4 into the classes that best describe their appearance. The choice of the number of classes used in the classification is generally subjective but Pringle *et al.* (2015) have provided an objective method for estimating the optimal number of classes. They suggested a range of 15 – 20 and therefore 16 CP classes were used for the present study. The 16 CP classes to which the characteristic CPs are assigned were delineated according to the optimization procedure discussed in § 5.2.4 using the significant wave height ERA-Interim pressure at a 6-hourly temporal resolution.

5.3 Results

The results obtained from analysis of the Durban and Richards Bay data are similar so we focus on presenting results from the Durban data.

5.3.1 Partitioning KwaZulu-Natal wave data

Wind data is crucial in the separation of wind waves from swell waves. Fig. 5.3 shows the wind roses of Durban and Richards Bay.

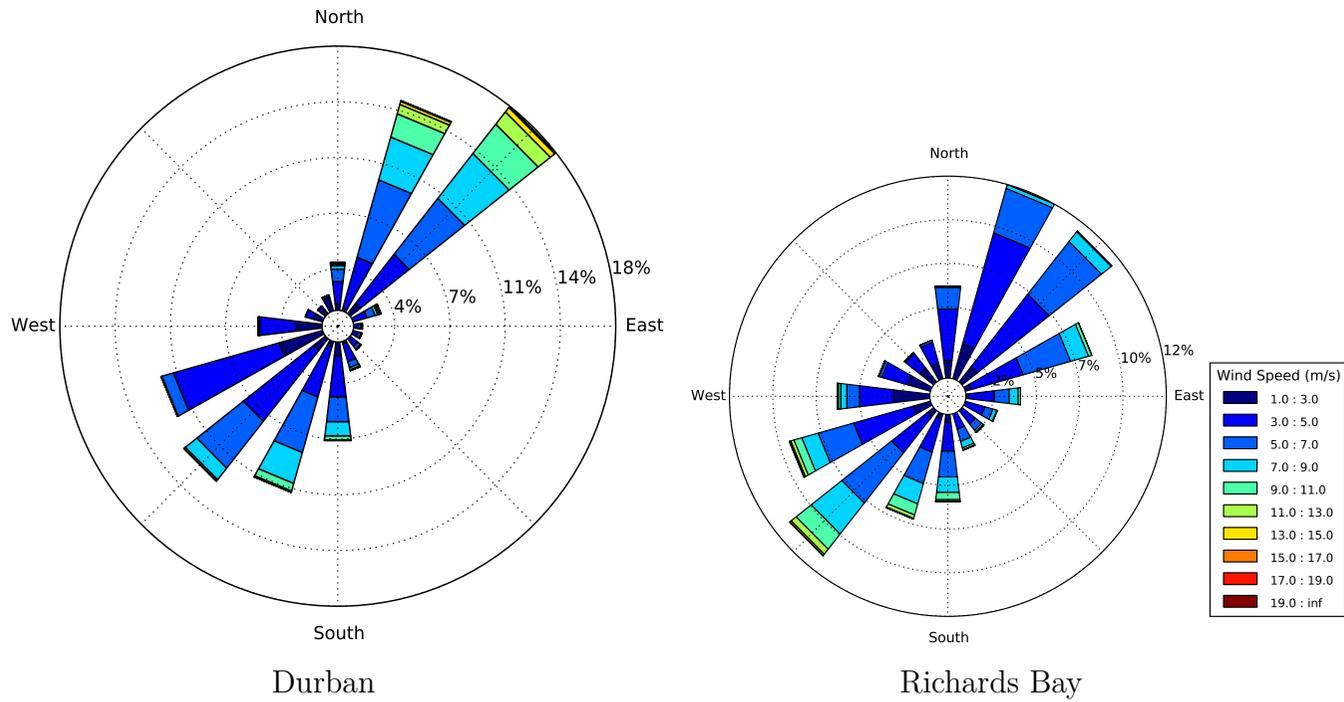


Fig. 5.3 Wind roses for Durban (31/05/2002 – 31/07/2013) and Richards Bay (19/08/1993 – 31/07/2013)

An example of partitioned data, both in terms of the associated frequency spectra and the directional energy spectra, is shown in Fig. 5.4. Fig. 5.4a shows the unpartitioned spectral data and the recorded wind vector, Fig. 5.4b shows the partitioned swell component in the data and Fig. 5.4c shows the partitioned wind-wave component. The plots show that the partitioning algorithm has effectively separated the locally generated wind-waves from the swell waves. The 2-dimensional plots illustrate how the wind waves are significantly different from the swell waves in both energy, direction and period.

The calibration factor κ for the peak separation criterion was calculated iteratively and a value of 0.8 was adopted. The function $A/(\omega_p^4 + B)$ was used to define the peak energy threshold following Hanson & Phillips (2001). The parameters A and B were chosen as $A = 1.0 \times 10^{-5}$ and $B = 1.0 \times 10^{-3}$ respectively. The intention of the threshold is to eliminate noise and Fig. 5.5 shows the spectral energy for the Durban data relative to the specified energy threshold.

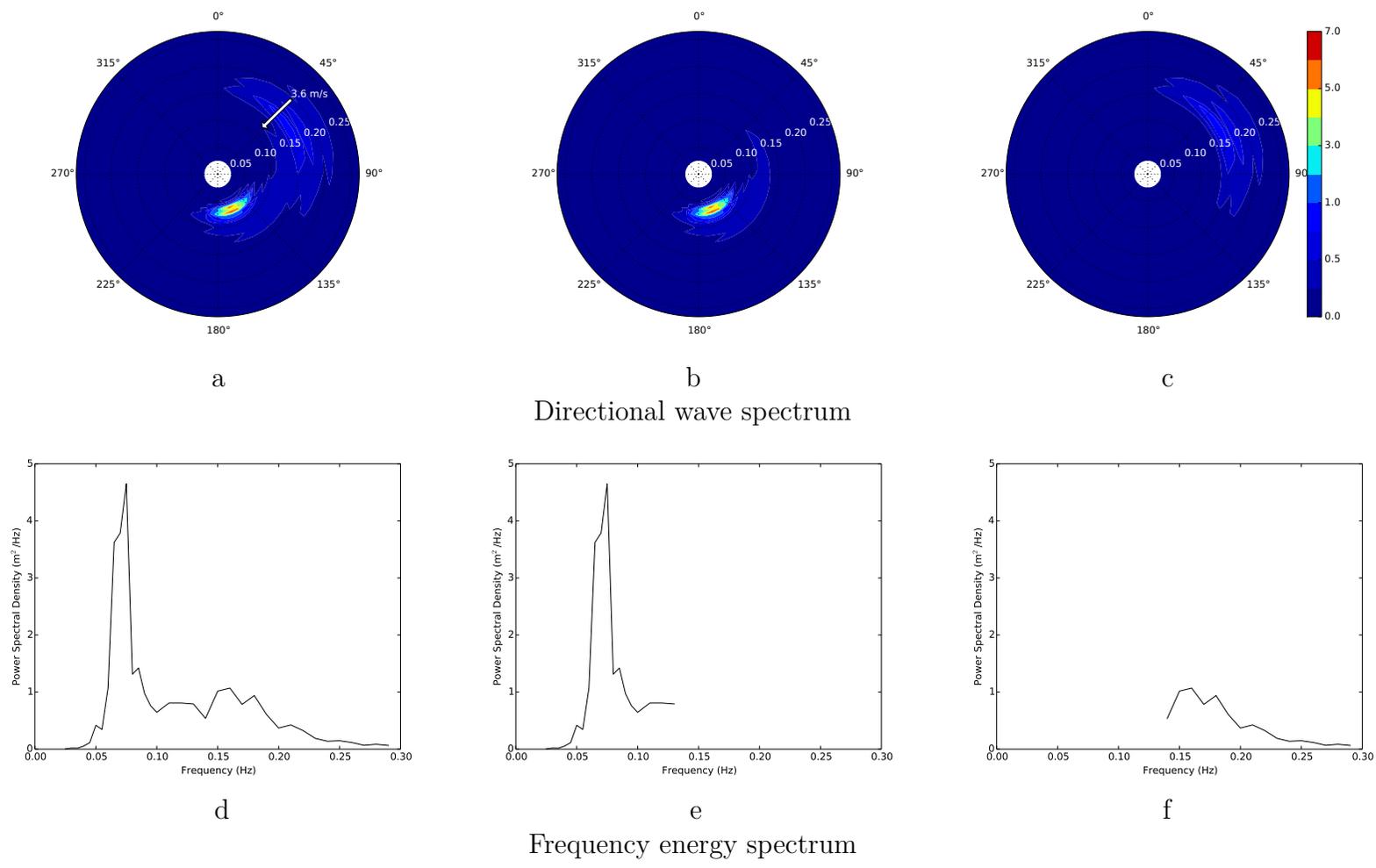


Fig. 5.4 An example of a directional energy spectrum (a) and frequency spectrum (d) partitioned into their respective swell (b & e) and wind-wave spectra (c & f)

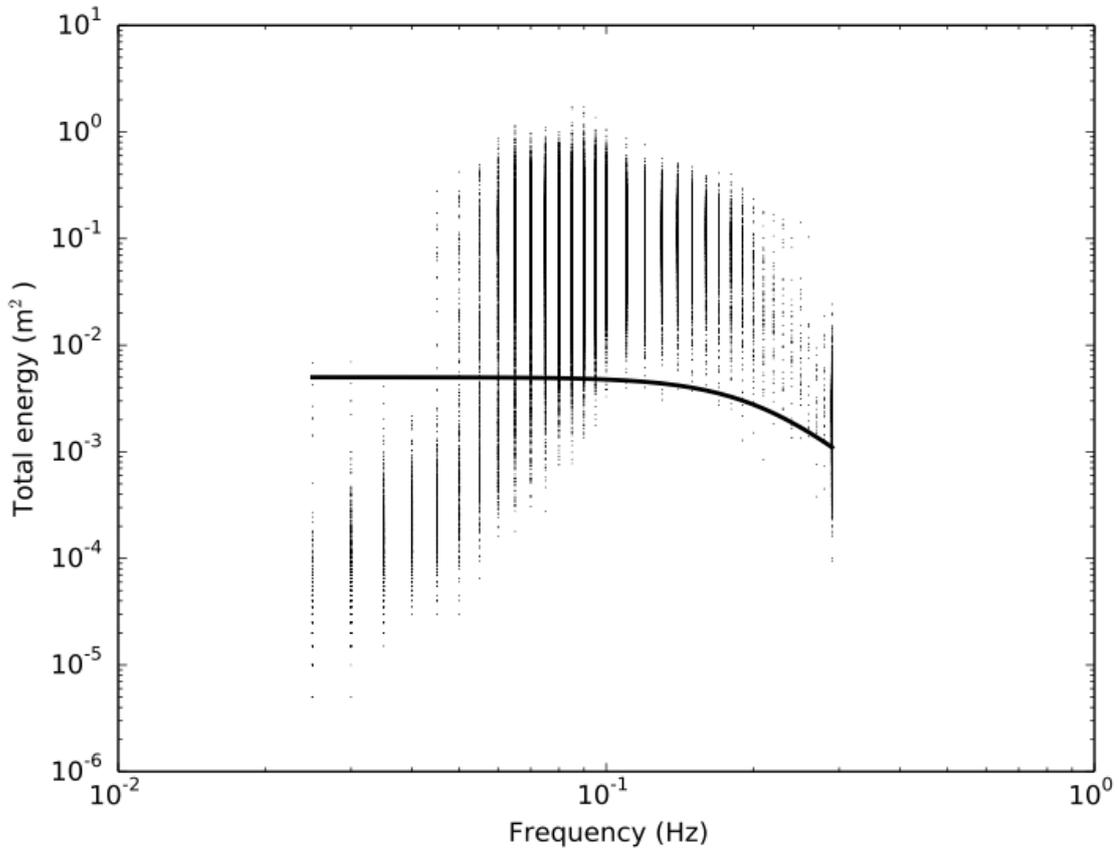


Fig. 5.5 The minimum energy threshold for the Durban data as defined by $A/(\omega_p^4 + B)$. The dots show the total energy of each swell partition.

5.3.2 Swell origins and associated circulation patterns

The derived spatial origins of the cross-assigned swell are shown in Fig. 5.6. The dots indicate the origin of the swell and the colour of the dots indicates the associated wave periods. As expected the shorter wave periods have origins closer to the coast than the longer wave periods. The majority of the swell wave events originate from the south east and have wave periods above 10 seconds.

Fig 5.7 shows five characteristic CP anomalies, the classes to which the classification algorithm assigns the anomalies, and the associated wave directional spectra. The swell origins shown as dots in Fig. 5.7 were grouped according to the spatial grouping method described in § 5.2.3. Visually, the characteristic CPs only weakly resemble the CP classes to which they are assigned. Furthermore the associated CP anomalies show a high degree of variability. While their shapes are similar to the identified classes

Assimilation of ocean wave spectra and atmospheric circulation patterns to improve wave modelling

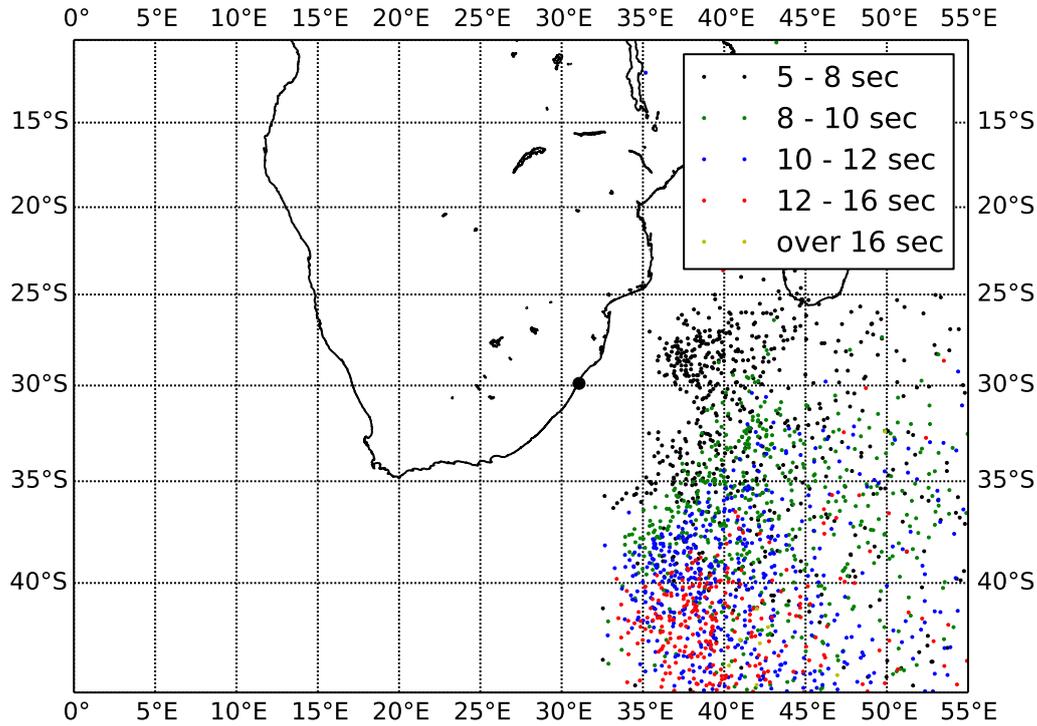


Fig. 5.6 A map of the Indian Ocean with the coloured dots showing all the swell sources for Durban. The colours depict the swell periods as defined in the legend. The large black dot indicates the location of the Durban wave recording buoy

their strengths are significantly weaker (values ranging from about -0.5 to +0.5). This may be attributed to the movement of low pressure cells within the vicinity of the swell origins or inaccuracy in the location and grouping of the swell origins. According to this method of swell grouping, the characteristic CPs are similar to west–east moving mid-latitude cyclones that frequent the region south of the country. The large area of high pressure shown in Fig. 5.7 (c) is typical of the pattern associated with the passage of mid-latitude cyclones.

Grouping the swell origins using the associated wave characteristics (Fig. 5.8) yields results with stronger associated CPs and therefore improved fidelity in associating CPs with the swell origins. The associated CPs were classified as (a) & (b) CP03, (c) CP08, (d) CP07 and (e) CP09. Their statistics are shown in Table 5.3. The wave parameters used for grouping the swell origins were as follows:

(a) $H_s = 1.5 - 2.0 \text{ m}$, $\omega = 0.1 - 0.088 \text{ Hz}$, $\theta = 90 - 135^\circ$

(b) $H_s = 1.5 - 2.0 \text{ m}$, $\omega = 0.0875 - 0.075 \text{ Hz}$, $\theta = 90 - 135^\circ$

-
- (c) $H_s = 1.5 - 2.0 \text{ m}$, $\omega = 0.0875 - 0.075 \text{ Hz}$, $\theta = 135 - 180^\circ$
 - (d) $H_s = 2.0 - 2.5 \text{ m}$, $\omega = 0.1 - 0.088 \text{ Hz}$, $\theta = 90 - 135^\circ$
 - (e) $H_s = 2.0 - 2.5 \text{ m}$, $\omega = 0.088 - 0.075 \text{ Hz}$, $\theta = 135 - 180^\circ$

Wave heights above 3.0 m were not considered since there were too few events on record during the time period considered using directional wave data (see Table 5.1). According to Table 5.3 CP07 is associated with the majority of extreme wave events, while CP09 contributes the largest percentage of extreme wave events during winter. All CP classes occur infrequently (4% – 6%). Figs. 5.7 and 5.8 show that the main driver of wave events is the strong coupling of low and high pressures to the east of the country. The location and orientation of this coupling defines the characteristics of the wave events such as wave height, direction and period.

5.3.3 Average wave directional spectra and associated circulation patterns

Atmospheric circulation patterns contain important information on wave height (or energy), direction and frequency which is of particular importance to coastal vulnerability assessment. This is shown specifically in Fig. 5.9 which contains average wave directional spectra for each CP. It is also clear that the classification algorithm is able to capture CP classes with significantly different wave characteristics. CP07 and CP09 Fig. 5.9(b) and (d) are associated with the largest wave energies which is also reflected in Table 5.3 as these two CPs contribute the most towards extreme wave events. The direction of wave attack is strongly linked to the orientation of the high-low pressure coupling. For example CP09 has a high-low pressure coupling such that it drives a strong southerly wind and hence wave energy is from the south. CP03 and CP08 Fig 5.9 are associated with low to medium wave energies. Fig 5.9(a) suggest that the higher frequency easterly component of the wave energy is driven by the low/high coupling to the east of the country. However the atmospheric driver of the lower frequency southerly component is unclear. Typically this pattern (CP03) resembles the high pressures that follows the west-east propagating mid-latitude cyclones which drives southerly to south-easterly waves which explains the southerly component of the wave energy. This is also seen in Fig 5.9(c) and (d) in which CP08 and CP09 resemble mid-latitude cyclones with a strong component of southerly wave energy.

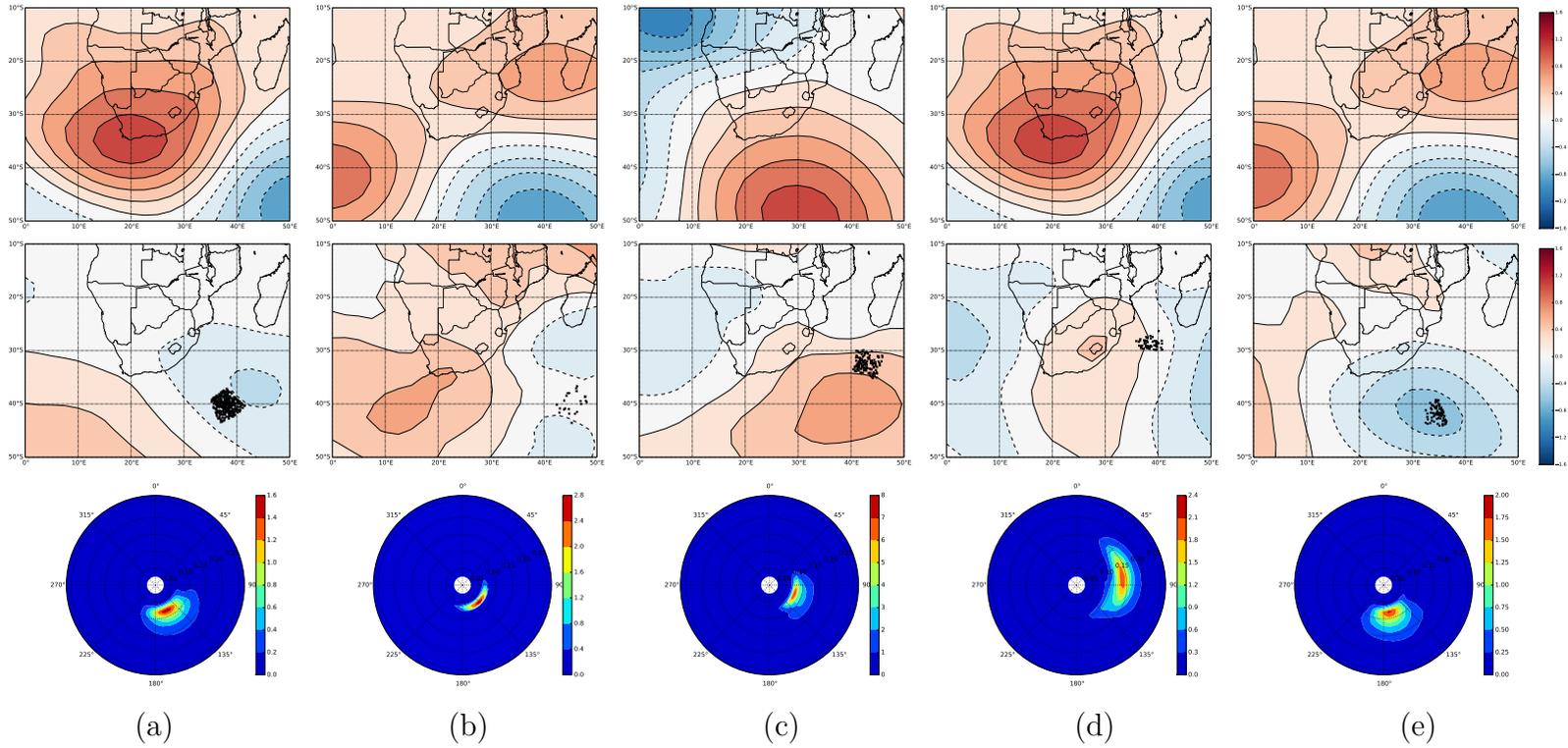


Fig. 5.7 The origins of swell events and associated CP anomalies using the spatial grouping method. The class assigned by the classification algorithm is shown in the top panel. The swell origins (dots) and associated average CP anomalies are shown in the centre panel. The associated wave directional spectra are plotted in the bottom panel. The characteristic CPs are classified as (a) CP15, (b) CP10, (c) CP12 (d) CP15 and (e) CP10.

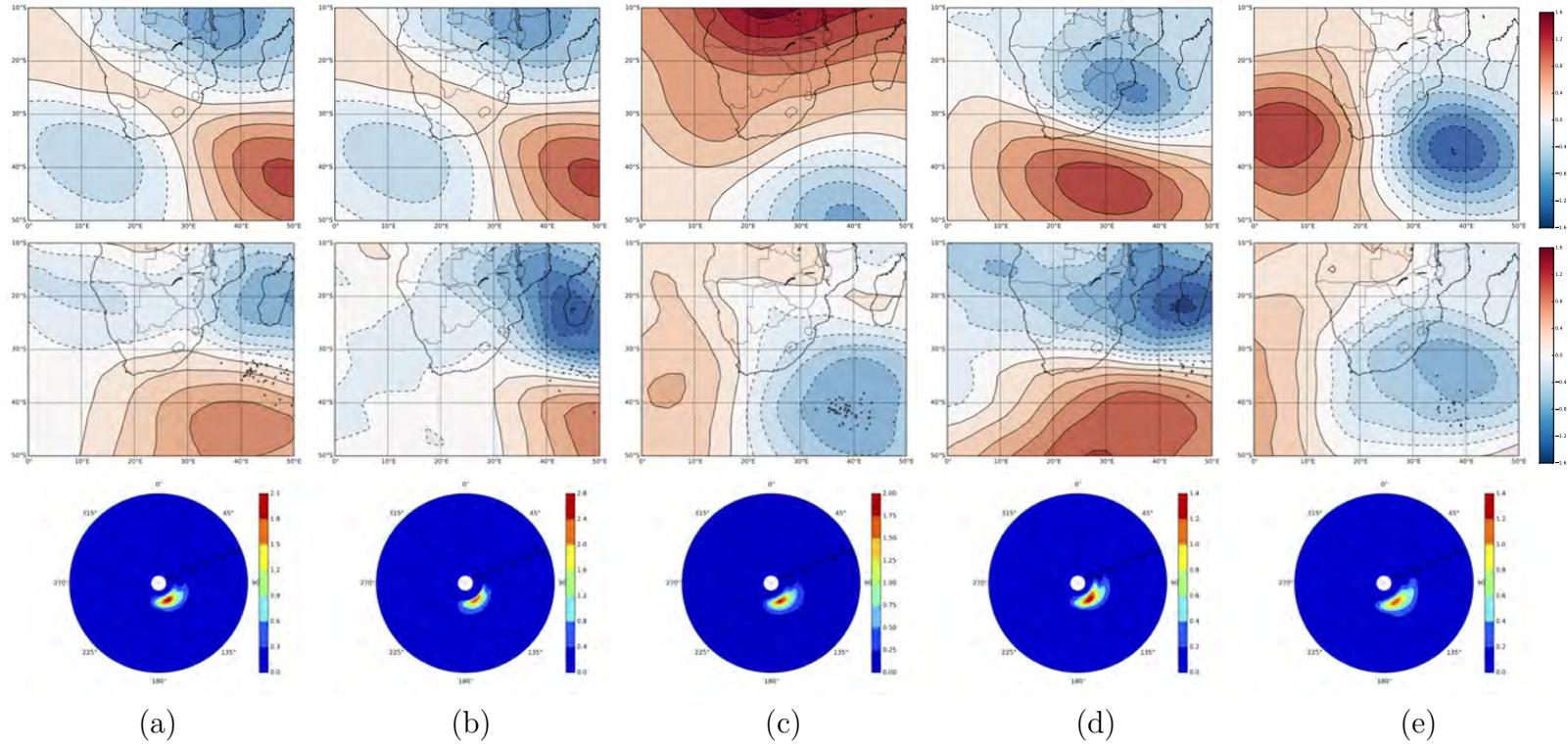


Fig. 5.8 The origins of swell events and associated CP anomalies using the wave characteristics grouping method. The class assigned by the classification algorithm is shown in the top panel. The swell origins (dots) and associated average CP anomalies are shown in the centre panel. The associated wave directional spectra are plotted in the bottom panel. The characteristic CPs are classified as (a) CP03, (b) CP03, (c) CP08, (d) CP07 and (e) CP09.

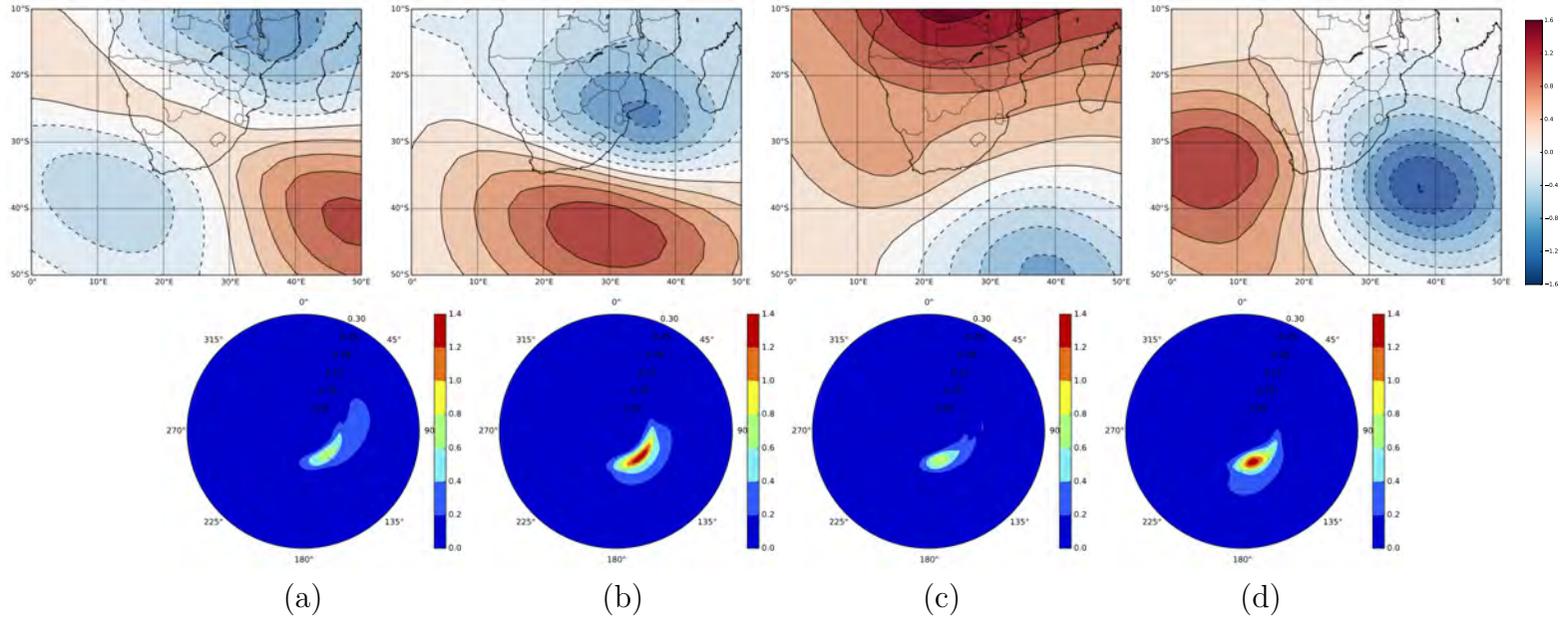


Fig. 5.9 The average wave directional spectra for associated CPs: (a) CP03, (b) CP07, (c) CP08 and (d) CP09.

Table 5.3 CP occurrence frequencies and their associated wave event statistics. The wave height threshold θ used to define the wave events was $\theta = 3.5$ m for this data.

| CP | CP03 | CP07 | CP08 | CP09 |
|---|------|------|------|------|
| <i>Occurrence frequency ($p(CP)$) %</i> | | | | |
| Summer | 6.54 | 6.49 | 5.85 | 4.15 |
| Autumn | 5.41 | 5.55 | 4.83 | 5.27 |
| Winter | 6.07 | 4.66 | 4.30 | 5.31 |
| Spring | 5.53 | 5.70 | 5.58 | 4.74 |
| All Seasons | 5.89 | 5.60 | 5.13 | 5.13 |
| <i>Contribution to extreme events ($p(CP \theta)$) %</i> | | | | |
| Summer | 0.00 | 26.6 | 0.00 | 0.00 |
| Autumn | 7.89 | 10.5 | 0.66 | 10.5 |
| Winter | 9.52 | 11.1 | 4.76 | 17.5 |
| Spring | 5.26 | 44.7 | 0.00 | 7.89 |
| All Seasons | 7.38 | 16.2 | 1.48 | 11.1 |

5.4 Discussion

The main proposition underpinning this study was that linking the occurrence of storm waves at a location of interest to atmospheric circulation patterns can improve analysis and modelling in coastal vulnerability assessments or coastal engineering design applications. We have demonstrated that spectral partitioning into non-local swell components and local wind-wave components can be used to track the origins of the swell waves and thus link them to their associated atmospheric circulation patterns. The successful establishment of these links is complex and our experimentation has highlighted some of the difficulties in achieving this.

The partitioning procedure contains various calibration factors that affect the outcomes e.g. for spectral peak separation and defining the energy threshold. The reader is referred to Hanson & Phillips (2001) for an explanation of the calibration factors. The cross assignment conditions in the swell tracking algorithm (§ 5.2.3) may also be site specific. The dependence on user defined calibrations makes the procedure less robust and detracts from its generality.

Perhaps the largest uncertainty is contained within Eqs. 5.5 and 5.6. Due to the dispersive nature of surface waves we expect waves with the largest wave periods to

Assimilation of ocean wave spectra and atmospheric circulation patterns to improve wave modelling

have traveled from the greatest distances. However, there are occasions when $d\omega/dt$ for wave events with long wave periods are similar to those with small wave periods. Such events may therefore be grouped unrealistically close to one another, perhaps due to neglecting dissipation effects. Incorrect spatial and temporal placement can create errors when linking CPs with the swell origins. Therefore outlier values need to be removed from the spatial and/or wave characteristic groupings prior to analysis of their associated CPs. Here we eliminated frequency values that were more than one standard deviation from the mean.

Fig. 5.6 shows all the unique swell events for the Durban data and illustrates some of the above-mentioned issues with the procedure. The swell origins are the small dots and they are coloured according to the magnitude of their wave periods. The swell origins largely behave as expected with the shorter period waves forming closer to shore than the longer period waves. However there are examples when this is not the case due to the reasons noted above. Some events are also estimated to have formed over Madagascar which is obviously not possible.

The linking of CPs to wave origins based on the spatial grouping method (Fig. 5.7) showed high variability that suggests erroneous coupling of storm origins with CP anomalies. However the method of grouping based on wave characteristics (Fig. 5.8) performed much better in this regard. Pringle *et al.* (2014) noted that it is possible for a specific storm to belong to a number of classes during its development because the set of classes represent different atmospheric states with no attached temporal information. The relative orientation of high and low pressure anomalies is also not directly incorporated into the CP classification procedure. This may be important for discriminating between waves from different directions because the locations of high and low pressures influences the direction of the surface wind field that drives wave formation. In general the strongest winds occur in the zone between high and low pressures where the pressure gradient is high. However the results in Fig. 5.9 indicate that the CP classification algorithm can successfully discriminate between wave spectra of varying strength and direction. Therefore we propose that the CPs are a physically meaningful way to describe complex wave climates.

Although the algorithm has limitations it can have significant benefits in some applications. The linking of spectral wave energy to CPs is valuable in spectral wave models. As shown in § 5.3 different circulation patterns produce different spectral shapes. It would therefore be erroneous to assume the same spectral shape for all wave conditions or to use the default spectral characteristics provided by some spectral wave

models (e.g. SWAN). The link between the spectral data and the CPs can be used to fit a parametric model to the spectral data to study the impacts of specific CPs on the coastline. For example the specific spectral shape for swell produced by tropical cyclones can be used to model their potential impacts. This may also be useful for the study of climate change phenomena where one could study the changes in forcing mechanisms as opposed to directly investigating the wave data.

Another important application of the algorithm may be found in the statistical analysis of extreme wave events and wave climate modelling. Firstly it provides a physically meaningful way to identify independent events. Furthermore it provides a method to differentiate and model wave directional spectra in complex wave climates. The classical method of defining independent events uses wave height thresholding and cannot reliably define statistically independent events. The proposed algorithm avoids the mixing of statistically independent events by linking them to specific meteorological systems in terms of their CPs. An example is shown in Fig 5.10 of a recorded wave event on the east coast of South Africa with two different causal weather systems. These two events may have been grouped as one using the two week time threshold proposed by Corbella & Stretch (2012d). Li *et al.* (2014) used the classical method of defining storm independence when they proposed a method of modelling multivariate storm parameters. In their method they identified that their model failed to capture the physical limits of the storm driving forces. Camus *et al.* (2011) utilised classification algorithms to group wave events with similar characteristics within a complex wave climate for modelling purposes. The more recent work by Camus *et al.* (2014); Espejo *et al.* (2014) also developed the links between atmospheric CPs and the aforementioned wave climate. However these links were found after CP classification, whereas in the method presented here the derivation of the CP classes is based *a priori* on their link to the wave climate. This distinction and its benefits are further discussed by Bárdossy *et al.* (2015). The results shown in § 5.3.2, 5.3.3 and Figs 5.7, 5.8 indicate that grouping events with similar characteristics on the assumption that they are driven by similar CPs is valid at our case study site. Exploiting these links can significantly improve statistical models.

Assimilation of ocean wave spectra and atmospheric circulation patterns to improve wave modelling

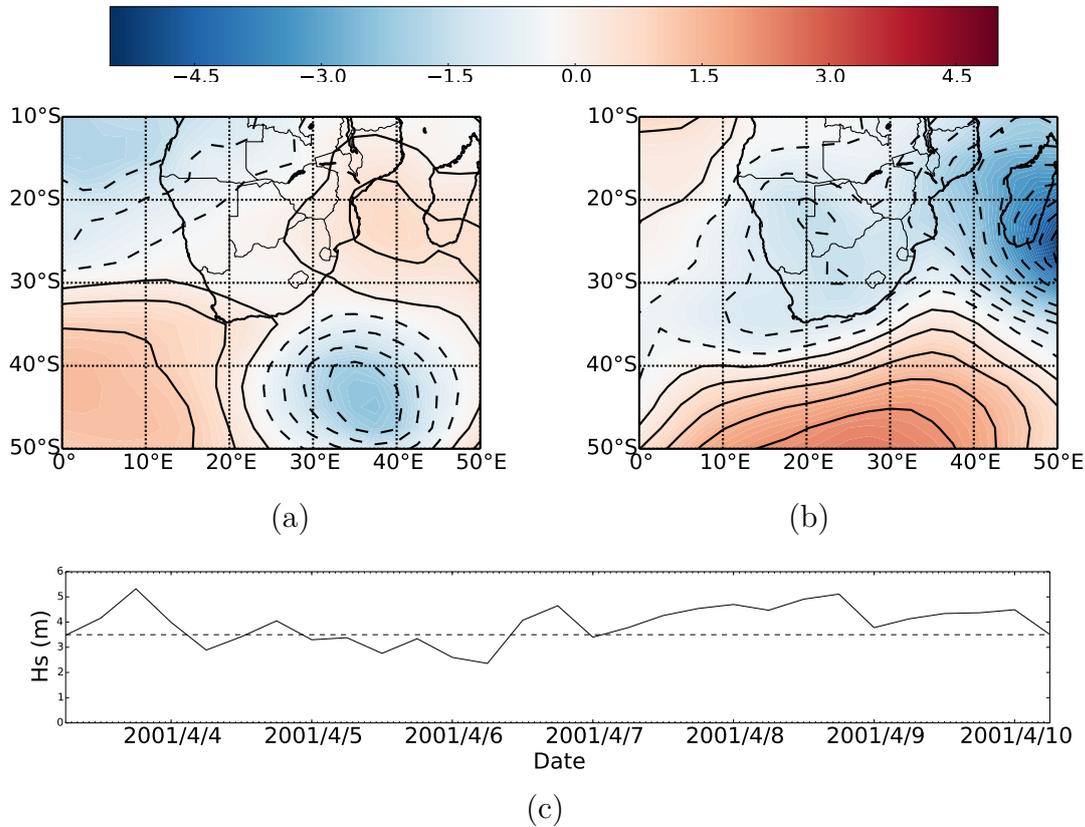


Fig. 5.10 An example of storm wave events in 2001 with two different driving weather systems: (a) a mid-latitude cyclone occurring on 2001/04/03, and (b) a tropical cyclone occurring on 2001/04/08. Both events may have been grouped as one storm by traditional threshold based methods. A threshold of 3.5m is shown by the dashed line in (c)

5.5 Conclusions

In this paper we have proposed a new method for linking atmospheric circulation patterns to wave spectral characteristics. This seems to be the first attempt at exploring this link for regional wave data in the south west Indian Ocean. The wave spectral energy distributions have been shown to be related to the relative positions and orientation of low and high pressure anomalies. The new methodology may lead to improved inputs to spectral wave models, aid in the study of climate change impacts, and improve the identification of statistically independent wave events and multivariate statistical models.

Chapter 6

Stochastic simulation of regional
wave climates conditioned on
synoptic scale meteorology. Part1:
Methodology

Abstract

Statistical modelling of wave climates is an important tool in coastal/ocean engineering design and vulnerability assessments. Modern techniques of multivariate modelling that exploit copulas are now being developed and used for risk assessment applications in diverse fields ranging from finance to hydrology and coastal engineering. Many such statistical models do not directly exploit the physical links between events of interest, such as floods or extreme storm waves, and their fundamental drivers. On the other hand, process-based models that attempt to include those links, are subject to modelling errors due to our limited understanding of the processes and/or limitations in the available computational resources to adequately resolve those processes. In this paper we introduce a new mixed approach to the stochastic simulation of wave climates that is conditioned on synoptic scale meteorological circulation patterns (CPs) as the key drivers of waves. Copulas are used for the multivariate dependence structure in the model while the CP occurrences are treated as a Markov chain. Simulated wave time series are shown to reproduce observed wave statistics from a case study site, including extremum statistics. The new techniques presented here should improve statistical modelling while retaining their simplicity and parsimony relative to full process-based models.

6.1 Introduction

It is important for coastal engineers, planners and managers to have detailed characterisations of regional wave climates in order to carry out design and/or coastal vulnerability assessments. Long term wave observations are generally needed to quantify risks and specify design criteria. However when wave observations are limited to relatively short periods, risk assessment and future projections are more uncertain especially within the context of climate change. The application of global wave models seems an attractive solution. However the accuracy of wave data derived from

these models depends strongly on wind field inputs that are themselves predicted by a model that relies strongly on reanalysis using data assimilation. As a consequence, they can be poor at predicting extreme events, and are unable to accurately predict future wave climate scenarios (Caires *et al.*, 2004; Chawla *et al.*, 2013; Stopa & Cheung, 2014; Swail & Cox, 1999; Tolman *et al.*, 2002). Multivariate statistical models have recently been used to address this issue (e.g. Callaghan *et al.*, 2008; Corbella & Stretch, 2013). However, a drawback of purely statistical models is that they are not directly linked to the physical drivers of the wave climate and therefore can potentially produce unrealistic results.

Synoptic scale atmospheric circulation patterns (CPs) that can be identified through classification techniques have recently been shown to have strong links to regional wave climates (Camus *et al.*, 2014; Corbella *et al.*, 2015; Espejo *et al.*, 2014; Pringle *et al.*, 2014). They provide a physically meaningful way to describe changes in wave state and to identify independent storm events. These characteristics can be exploited to model wave climates statistically. Bárdossy & Pegram (2011); Camus *et al.* (2014) have shown that CPs can be used to successfully downscale output from regional climate models. Such models are limited in their applicability because the downscaling technique can only correct statistical properties of the estimated wave climate. CPs have not yet been used to directly model regional wave climates based on the associated intra-dependence structure between variables such as wave height (H_s), direction (D) and period (T_p).

The simulation of wave climates can be approached in two different ways (Solari & Losada, 2011). The first is to simulate entire wave records and the second is to simulate individual storm events with different durations and inter-arrival times (see e.g. Callaghan *et al.*, 2008; Corbella & Stretch, 2013; Solari & Losada, 2011). Simulating storms typically requires the subjective selection of a wave height threshold to define the start and end of a storm (Corbella & Stretch, 2013; Li *et al.*, 2014). In contrast simulating entire wave records reproduces the natural changes in wave states between the extreme events.

In the past AR (auto-regressive) or ARMA (auto-regressive moving average) models have been used to statistically model regional wave climates. Examples of these techniques are given by Cunha & Guedes Soares (1999); Guedes Soares & Ferreira (1996); Solari & Losada (2011). However these models are purely statistical with no links to the physical mechanisms driving wave development. Furthermore the use of AR or ARMA models requires stationary statistics. Wave climates are typically

Stochastic simulation of regional wave climates conditioned on synoptic scale meteorology. Part1: Methodology

non-stationary and not normal. Therefore the general Box-Cox transformations are required to normalise the data after which the data must be transformed further to ensure it is stationary (Box & Cox, 1964). However CPs provide a physically meaningful framework to model non-stationary statistics. For example the transitions between CPs and their seasonal or longer term variations can naturally delineate changes in wave states. Links to CPs can therefore be exploited to simplify non-stationary statistical wave modeling. Such an approach is also well suited to the task of simulating long wave records for investigating climate change effects.

In this paper we present a mixed approach to the stochastic simulation of wave climates that is conditioned on the synoptic scale circulation patterns that drive them. This new modelling approach is first described and then some simulation results are presented that illustrate the model's efficacy in reproducing the statistics of an observed wave climate at a case study site.

6.2 Methods

6.2.1 Case Study Site

The regional wave climate along the east coast of South Africa (Fig. 6.1) is commonly referred to as highly energetic (Mather & Stretch, 2012; Rossouw *et al.*, 2011; Taljaard, 1967). A number of weather patterns are associated with the generation of this wave climate. Mid-latitude cyclones, cut-off lows and tropical cyclones have been linked to major wave events (Corbella *et al.*, 2015; Mather & Stretch, 2012; Pringle *et al.*, 2014). However the role of tropical cyclones is debatable and according to Mather & Stretch (2012) they tend to drive large wave events only when they become stationary south of Madagascar and north-east of Durban.

Table 6.1 The allocation of months to seasons

| Season | Months |
|--------|----------------------|
| Summer | December – February |
| Autumn | March – May |
| Winter | June – August |
| Spring | September – November |

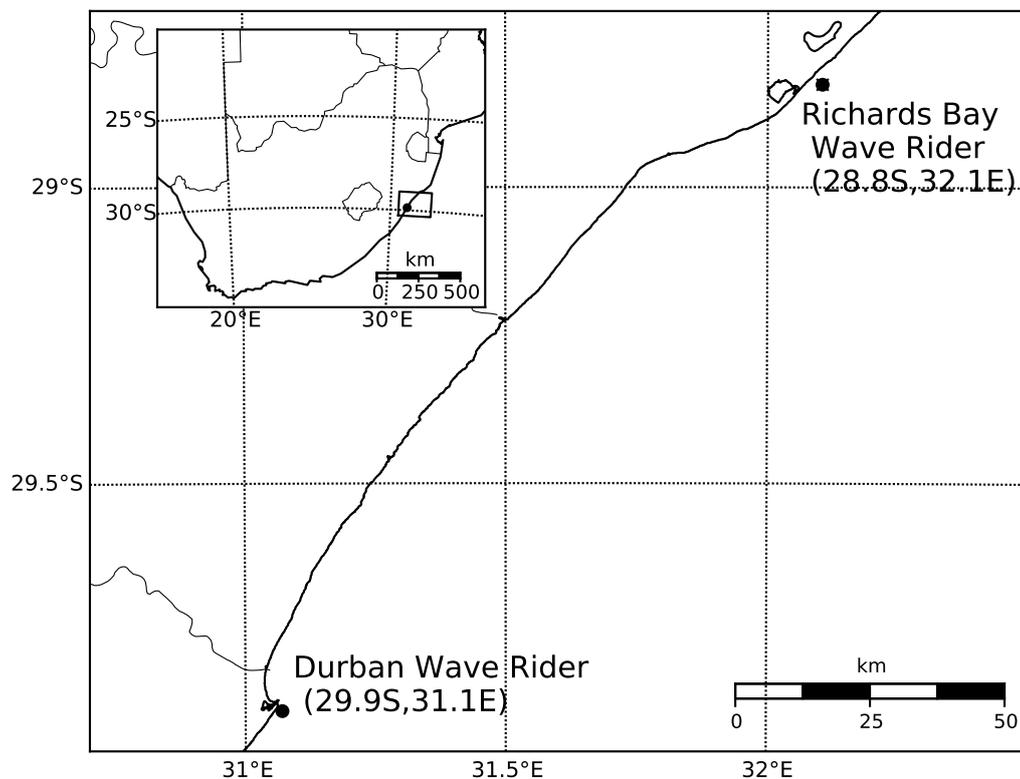


Fig. 6.1 Locations of the wave observation buoys at Durban and Richards Bay, along the KwaZulu Natal coastline (Pringle *et al.*, 2014).

6.2.2 Data Sources

The atmospheric classification algorithm used herein was applied to pressure anomalies on the 700 hPa geopotential height with a grid resolution of 2.5° ($10^\circ\text{S } 0^\circ\text{E} - 50^\circ\text{S } 50^\circ\text{E}$) at a 6-hourly temporal resolution. This region is of sufficient size to include weather patterns responsible for swell waves (Corbella *et al.*, 2015). Positive and negative anomalies indicate regions of relatively high and low surface pressures respectively. Anomalies are defined relative to the mean and normalised by the standard deviation for each Julian day $j(t)$. The anomaly h at location i and time t is thus given by

$$h(i, t) = \frac{k(i, t) - \mu(i, j(t))}{\sigma(i, j(t))} \quad (6.1)$$

where $k(i, t)$ are the geopotential heights and $\mu(i, j(t))$ and $\sigma(i, j(t))$ are the mean and standard deviation values at location i and for Julian day $j(t)$. The Julian day function provides a smooth transition between CP states on different days. The pressure data was obtained from the ERA-Interim dataset (<http://apps.ecmwf.int/datasets/>) for the

Stochastic simulation of regional wave climates conditioned on synoptic scale meteorology. Part1: Methodology

period 1979–2012 at 6-hour intervals.

The 700 hPa geopotential height was selected for the atmospheric classification because it has been widely used previously in similar weather pattern classifications (see e.g. Bárdossy *et al.*, 2015) and this prior knowledge and experience can be used for comparison purposes. Tests using sea level pressures showed only small differences in the outcomes. The lower level pressures are also typically more "noisy" (i.e. have more small scale variability due to boundary layer effects), which can adversely affect the optimisation process used to derive the CPs.

Wave data used to guide the classification algorithm were obtained from the Durban and Richards Bay waverider buoys (Fig. 6.1). A previous study has shown a strong correlation between wave height data from the two buoys (Corbella & Stretch, 2012d). Therefore data from the Richards Bay waverider were used to supplement missing data within the Durban dataset. The data comprises significant wave height, direction and period at 3-hour intervals. Wave height and period have been recorded for approximately 18 years from 1992–2009 whereas wave direction measurements have only been recorded from 2007.

6.2.3 General Approach

This section briefly outlines the new CP-wave simulation technique after which a more detailed description is provided in the sections that follow.

Firstly a fuzzy rule based classification algorithm is applied to gridded anomalies within the 700 hPa geopotential height at a 6-hourly temporal resolution to delineate CPs that are associated with wave development. A transition matrix is calculated to describe the movement between classes. Secondly the distributions of wave climate variables for each CP are calculated and their dependence structures are evaluated using Archimedean copulas. A pseudo-random CP sequence is simulated based on the transition matrix for N years at a 6-hourly interval. From the simulated CP sequence wave climate variables are simulated based on the associated CP based dependence structures.

6.2.4 Classification of Atmospheric Circulation patterns

Pattern recognition algorithms aim to provide insight into multi-dimensional processes (Huth *et al.*, 2008). They involve *identifying* common states (or classes) which exhibit similar properties and then *grouping* similar states together using a classification tech-

nique (Huth *et al.*, 2008). Examples include k -means clustering, principle component analysis (PCA) and the use of artificial neural networks (ANN) (Hewitson & Crane, 2002; Huth *et al.*, 2008). In the case of climate studies the links between different classes and a surface variable of interest are typically only made after the classes have been derived. However the technique used herein is distinct in that it has intrinsic links to wave characteristics because wave heights are incorporated into an optimisation procedure that identifies an optimal set of classes using objective functions. The algorithm is based on fuzzy logic and was developed by Bárdossy *et al.* (1995) and adapted by Pringle *et al.* (2014) to assess the drivers of regional wave climates. Bárdossy *et al.* (2015) have highlighted an advantage of including the variable of interest directly into the classification method by showing that the deduced CP classes can differ for different guiding variables.

The first type of objective function used to derive the optimal set of CP classes is designed to select classes associated with strong wave events, i.e. those classes within which the frequency of exceeding specified wave height thresholds is significantly higher than the unconditional exceedance frequency. Whence

$$O_1(\theta) = \sqrt{\sum_{t=1}^T (p(H_s \geq \theta | CP_t) - \bar{p})^2} \quad (6.2)$$

where θ is the prescribed wave height threshold applied over the time period T , CP_t is the class assigned at time t and \bar{p} is the unclassified relative frequency of an event. Two wave height thresholds were used (Pringle *et al.*, 2014): (a) $\theta = 3.5$ m, since wave heights greater than this are associated with severe coastal erosion on the KwaZulu-Natal coastline (Corbella & Stretch, 2013); (b) $\theta = 2.5$ m, a threshold associated with midrange wave heights. The addition of the midrange wave height threshold reduces the ‘noise’ or ‘spiky’ behaviour of the objective function and improves the optimization.

A second type of objective function was used to select CP classes that are associated with wave heights significantly different from the unconditional mean. Whence

$$O_2 = \sum_{t=1}^T \left| \frac{H_s(CP_t)}{(\overline{H_s})} - 1 \right| \quad (6.3)$$

where $(\overline{H_s})$ is the unconditional average significant wave height and $H_s(CP_t)$ is the average significant wave height for the given CP class assigned at time t .

Stochastic simulation of regional wave climates conditioned on synoptic scale meteorology. Part1: Methodology

A linear combination of the objective functions (Eq. 6.2 and 6.3) was used to locate an optimal set of CP classes using a simulated annealing optimisation algorithm. Further details of our CP classification technique used herein are given in Bárdossy (2010); Bárdossy *et al.* (1995, 2015); Corbella *et al.* (2015); Pringle & Stretch (2015); Pringle *et al.* (2014). The outcome of the CP classification process is shown in Fig 6.7 where the average anomalies associated with the full set of derived CPs are presented. The number of CPs used in the classification (sixteen in this case) was based on the entropy criterion suggested and tested by Bárdossy *et al.* (2015); Pringle & Stretch (2015).

6.2.5 Wave Climate Simulation

A complex, challenging and arguably the most important part of wave climate simulation is the modelling of their multivariate dependency structure. The dependence structure between variables such as wave height (H_s), direction (D) and period (T) is typically modelled using copulas or a joint probability method (JPM) (Callaghan *et al.*, 2008; Corbella & Stretch, 2013; De Michele *et al.*, 2007; Li *et al.*, 2013). In high dimensional environments the relationships between variables can be unclear and difficult to derive. However a link to CPs can provide insight into these relationships and can simplify their modelling. For example the temporal sequence of wave climate variables (H_s , D and T) depend strongly on the occurrence of different CPs. Therefore the CPs drive changes in the state of the wave climate. For example some CPs are associated with calm periods while others are associated with extreme events. The natural transitions of CPs between different states can therefore provide a physically based temporal sequence of wave climate variables.

A summary of the steps used for the wave climate simulation is as follows:

1. Classify the observed pressure data as discussed in § 6.2.4.
2. Calculate the occurrence statistics for each CP and derive a CP transition matrix (e.g. Table 6.2).
3. Calculate the distributions of wave climate variables for each CP and evaluate their dependence structure (e.g. Fig. 6.4,6.5,6.8 & 6.9).
4. Simulate a pseudo-random CP sequence based on the transition matrix for N years and 6-hour intervals.

5. Simulate wave climate variables for the CP sequence using the CP based dependence structure (e.g. Fig 6.6).
6. Repeat steps 4-5 M times.

Simulating a CP sequence

A Markov Chain was used to model the transitions between CPs. A transition probability matrix was constructed to describe the temporal sequencing of the CPs. The CP sequence can be expressed in terms of conditional probabilities $P(CP_t | CP_{t-1}) = p_{ij}^{(s)}$ where $p_{ij}^{(s)}$ $i, j = 1 \dots n$ is the probability of a specified CP occurring at time t given the CP at the preceding time step ($t - 1$) during season s , where n is the total number of CP classes ($n=16$ in this case). The transition probabilities were estimated for each season to account for the seasonal features of the wave climate: seasons are defined in Table 6.1. The conditional probabilities $p_{ij}^{(s)}$ made up the transition matrices for each season. A preliminary analysis of the transition probability matrix shows strong diagonal dominance due to the high CP persistence over the 6-hour time intervals. This is shown in Fig 6.2. In other words the probability of a CP remaining the same for consecutive time steps is high. This pattern stability or persistence drives large waves towards the coast since the CP associated wind field has time to develop the sea state. The transition probability matrix also shows that changes in CP regimes favour transitions to particular successor CPs. For example in Fig 6.7 the CP in row 1 column 1 is most likely to move to the CP in row 1 column 2 since that has the dominant transition probability for the transition to another CP.

The simulated CP sequence was 100 years long with a 6-hour temporal resolution. This process was repeated 101 times to obtain a suitable sample size and account for a number of different scenarios. From the simulated CP sequence it was then possible to simulate a set of 101 wave climates each spanning 100 years.

Marginal Distributions

In order to simulate wave climates conditioned on the occurrence of different CPs it is important to accurately describe the relationships of H_s , T and D within each CP. The marginal distributions for H_s , T were modeled using a non-parameteric kernel density estimation (KDE) technique (Rosenblatt, 1956), whereas wave directions (D) were sampled from the empirical distribution. The fixed width kernel density estimator is

Stochastic simulation of regional wave climates conditioned on synoptic scale meteorology. Part1: Methodology

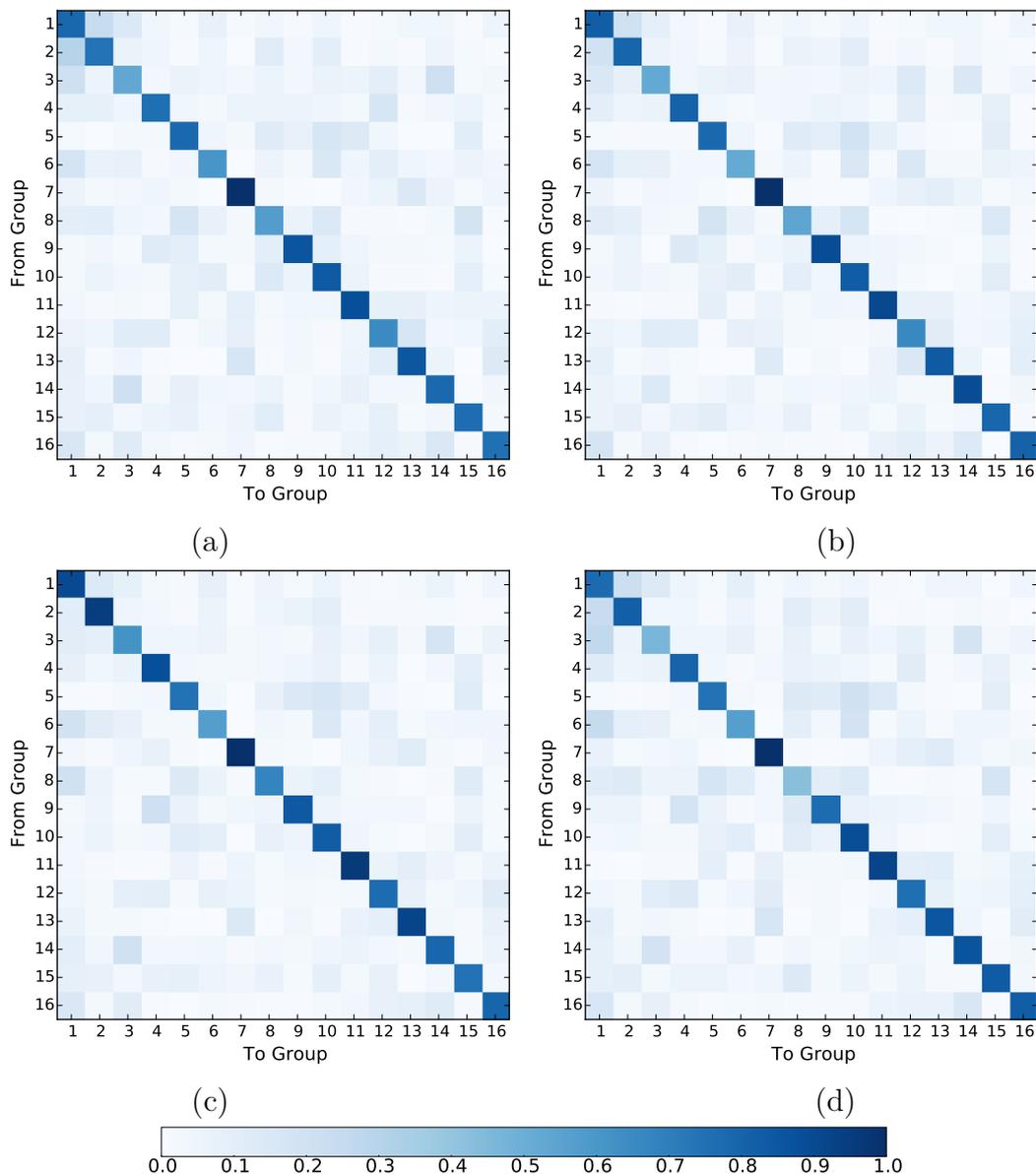


Fig. 6.2 Transition probability matrices showing strong diagonal dominance in CP transitions. Arbitrary CP numbers are assigned to the classes and are shown on the plot as the x and y axes accordingly.

defined as

$$\hat{f}(x) = \frac{1}{dN} \sum_{i=1}^N K\left(\frac{x - x_i}{d}\right) \quad (6.4)$$

where d is the kernel width (referred to as the bandwidth), N is the total number of sample points and K is a kernel function describing the shape of the individual kernels. The kernel function $K(\cdot)$ has a mean of zero and density of one: in this case the kernels were assumed to be Gaussian $N(0, 1)$. The KDE technique has the advantage that it does not rely on fitting a particular distribution to the dataset. However the method does not describe extreme values well. This is because it is based on observations that are limited by the maximum value on record. To account for this limitation exponential tails were fitted to the two largest observed values on record. This can be justified by an assumption that the tails of distributions often approach an exponential distribution (Gyasi-Agyei & Pegram, 2014; Maidment, 1993). An exponential tail is given by

$$F(z) = 1 - (1 - p_e) \exp[-(z - z_e)/L] \quad (6.5)$$

where the pair (z_e, p_e) corresponds to the second largest value z_e and its associated cumulative frequency p_e . The parameter L is derived by substituting the largest observed value and its associated cumulative frequency. Examples of this technique are shown in Fig 6.3 for CPs associated with both small and large waves. This method of estimating extremes was only applied to (H_s) values.

Simulating Waves

The aim of this study was to develop a simple method to simulate regional wave climates driven by CP occurrences. However there are fundamental properties of wave mechanics that must be incorporated. An example is the existence of temporal dependence in wave heights e.g. large waves tend to follow large waves and similarly for small waves. Other examples include the relationships between wave heights, wave periods, and wave directions.

In order to correctly model the temporal sequence of wave heights their lag one (6-hour) correlation was investigated. Temporal correlations at larger time lags are more difficult to interpret because there is limited data at larger lag times due to our approach of conditional sampling on CP occurrences. The Kendall's tau statistic showed strong correlation between wave heights at a 6-hour temporal shift. The log transform was applied to wave height values to normalise and reduce sampling bias.

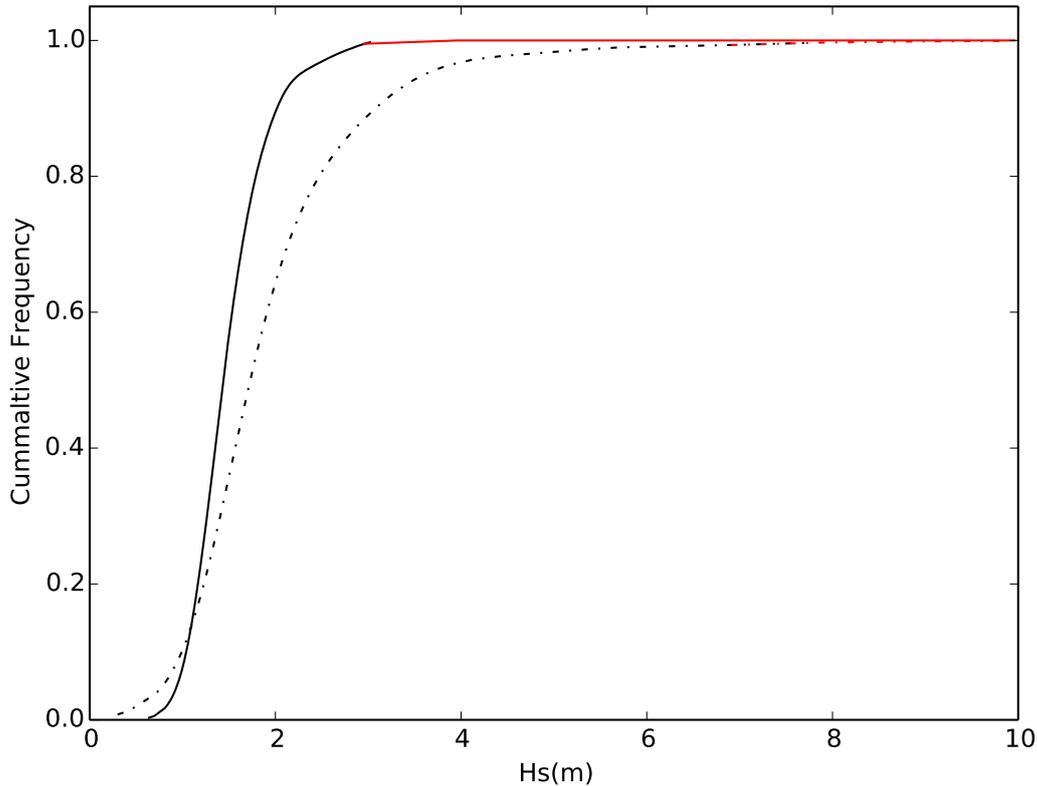


Fig. 6.3 Cumulative frequency distributions for CPs associated with small wave heights (solid line) and a CP associated with large waves (dashed line). Also shown are the exponential tails for each distribution (red) calculated from Eq. 6.5.

This revealed strong dependence in both the upper and lower tails of the distribution - an example is shown in Fig. 6.4(a) &(b).

The strong tail dependence suggests that a simple AR model would not capture such detail correctly and therefore a copula based approach was considered. Details of the copula approach are discussed in § 6.2.7

Observed sample pairs of wave period (T) and wave height (H_s) indicate that the largest values of each variable were correlated whereas no clear correlation is evident for low wave heights. An example of this relationship is shown in Fig 6.9. Therefore the wave period was simulated using a copula approach conditioned on the wave height for the given CP at each time step. The copula approach is well suited to describe complex relationships between variables. Marginal distribution functions for H_s and T were estimated according to § 6.2.5.

The data indicated no clear relationship between wave height and average direction for small to average wave heights. However Fig 6.5 shows that larger waves ($H_s \geq 3m$)

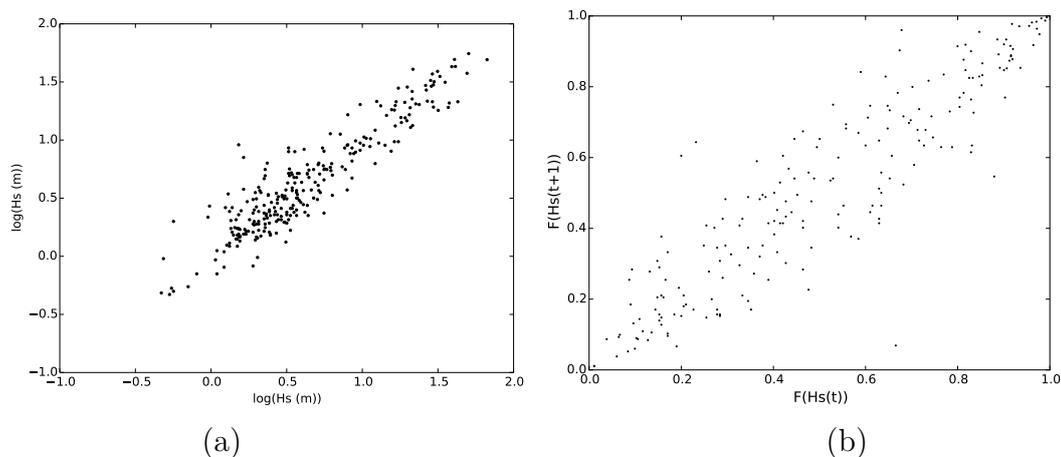


Fig. 6.4 (a) A sample scatter plot of log transformed wave heights with 6-hour separation showing strong dependence and (b) the lag one wave height plot of the marginals $F(H_s^t)$ and $F(H_s^{t+1})$ showing the rank correlation.

are associated with wave directions between $135^\circ - 180^\circ$. Therefore the wave directions were simulated conditioned on the wave height from their empirical distributions for each CP. The reason for this choice in simulating wave directions was that the available data for wave directions at the case study site were limited to approximately 5 years. Therefore fitting conditional marginal distributions to the dataset is difficult. Furthermore the empirical distributions limit the wave directions to physically realistic values. This prevents the simulation of waves with directions that are unrealistic. An alternative approach where the wave simulation is conditioned on the wave direction (rather than CP occurrence) is discussed in § 6.4.

6.2.6 Physical Limits

Naturally occurring physical limits on wave climate parameters prevent the occurrence of unrealistic values. However in some instances it is possible to simulate such values using statistical models. Examples of the physical limitations include wave breaking due to steepness or water depth and the direction of the approaching waves relative to the coastline. Waves can only approach a coastline from its seaward side which is a function of the coastal orientation. Wave steepness is defined as the ratio of wave height (H) to wave length (L) and for deep water is given as (e.g. Goda, 2008)

$$\frac{H}{L} = \frac{2\pi H}{gT^2} \quad (6.6)$$

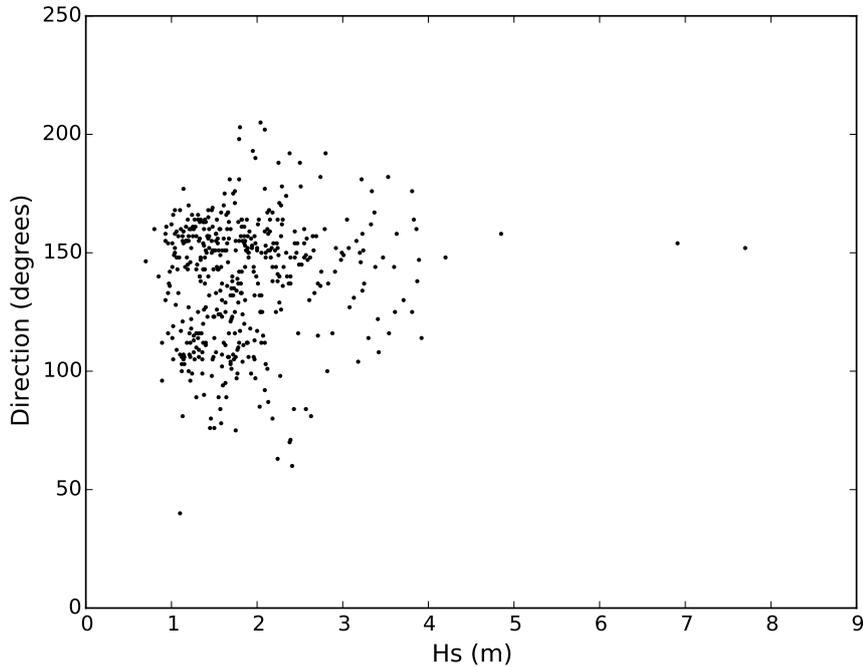


Fig. 6.5 Observed wave height and directions for a particular CP.

where g is the gravitational constant and T is the wave period. The maximum wave steepness was assumed as $\frac{1}{7}$ (Michelle, 1893).

While the link with CPs constrain the wave climate parameters to a certain range on average, it is still possible to simulate unrealistic combinations of values. For example simulating a small period with a large wave height or a wave direction that cannot occur relative to the coastal orientation. Therefore to account for this the wave period T was limited by Eq 6.6 given the simulated wave height. Significant attention has been focussed on large wave heights or limiting the wave height due to steepness and water depth restrictions. However it is also possible to simulate waves that are unrealistically small. To address this the simulated waves were restricted by a lower bound that was the smallest observed wave height. Wave direction was simulated according to § 6.2.5.

6.2.7 Copula Based Simulation

Copulas have been widely used for describing and modelling the dependence structure in multivariate statistics. Recently they have been applied to risk assessment within the coastal engineering context including coastal erosion prediction based on multivariate wave statistics (Corbella & Stretch, 2012a, 2013; Li *et al.*, 2013, 2014).

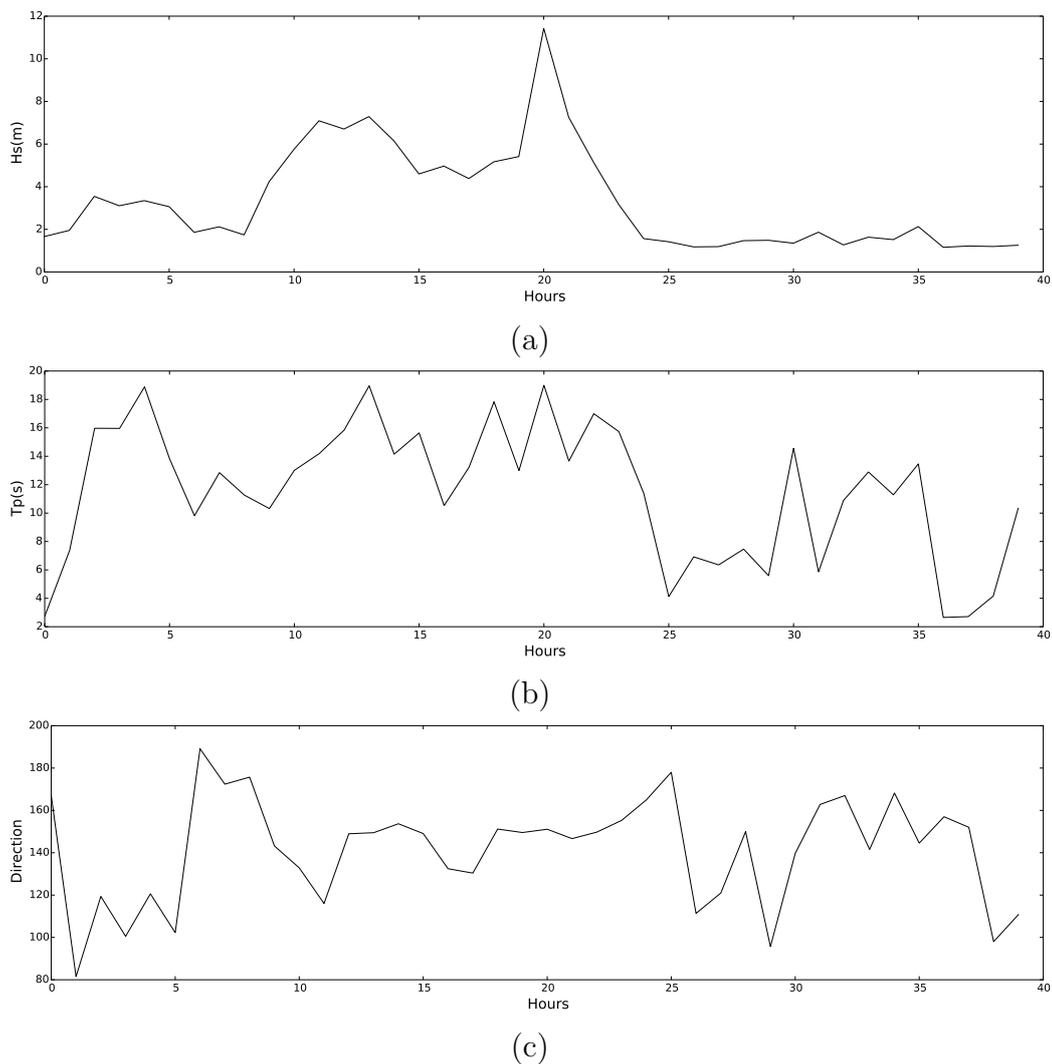


Fig. 6.6 Example of a simulated storm sequence showing the wave height (a), wave period (b) and wave direction (c).

An advantage of copulas is that the dependence structure between random variables is completely described by the copula function independently of the individual marginal distribution functions. Sklar (1959) showed that a copula can describe the joint cumulative multivariate distribution function $H(x_1, \dots, x_n)$ for continuous random variables x_1, \dots, x_n such that

$$H(x_1, \dots, x_n) = C\{F(x_1), \dots, F(x_n)\} \quad (6.7)$$

where $C\{\cdot\}$ is the copula function and $F(x_1), \dots, F(x_n)$ are the individual marginal distributions functions for x_1, \dots, x_n continuous random variables.

Stochastic simulation of regional wave climates conditioned on synoptic scale meteorology. Part1: Methodology

Archimedean copulas provide a relatively simple framework for constructing a copula from an observed set of random variables. They can be described by a single parameter based on the sampled Kendall's tau statistic. A detailed description of Archimedean copulas is given by Nelson (2006). In the bivariate case an Archimedean copula $C(u, v)$ can be formally defined as

$$C(u, v) = \phi^{-1}(\phi(u) + \phi(v)) \quad (6.8)$$

where ϕ is a generator function that is continuous and strictly decreasing from $[0,1]$ to $[0,\infty]$, $u = F(X)$ and $v = F(Y)$ are the marginal distribution functions for the random variables X, Y.

Two different copulas were used to simulate the pairs H_s - H_s and H_s - T values for a given CP. Firstly the temporal sequence of wave heights was simulated using the N12 copula. This copula is well suited to describe the relationship between variables that exhibit strong dependence especially in the extremes as shown in Fig 6.4. The N12 copula is described in Nelson (2006) with generator function

$$\phi(q) = \left(\frac{1}{q} - 1\right)^\theta \quad (6.9)$$

where q is a value between (0,1) and $\theta = 2/(3(1 - \tau))$ with τ being the Kendalls tau statistic.

The temporal sequence of wave heights were simulated as follows:

1. Randomly select initial $H_s^{(0)}$ value given CP at $t = 0$.
2. Set $h_s^{(0)} = \ln(H_s^{(0)})$.
3. Calculate the percentile $u = F(h_s^{(0)})$ given CP_{t+1} and marginal distribution function F .
4. Calculate the percentile v conditioned on u as $P(V \leq v|U = u) = \frac{\partial C(u,v)}{\partial u}$.
5. Return the transformed wave height value at the given quantile level v as $H_s^{(t+1)} = \exp(F^{-1}(v))$, where F is the marginal distribution function.
6. Repeat $N - 1$ times, where N is the length of the simulation time.

Given the sequence of wave heights it is possible to simulate the conditional sequence of wave periods T . The Gumbel-Hougaard copula was used to simulate the

wave period conditioned on wave height. This copula is defined by the generator function

$$\phi(q) = (-\ln(q))^\theta \quad (6.10)$$

where q is a random variable on the interval (0,1) and $\theta = 1/(1 - \tau)$. The Gumbel-Hougaard copula has no lower tail dependence but strong upper tail dependence. This dependence structure is expected for wave heights and periods since large waves are generally associated with large periods, which is demonstrated in Fig 6.9.

6.2.8 Statistical Comparisons

A basic requirement for statistical models is that they can generate simulations that reproduce the statistical properties of the observed data that is being modelled. Comparisons used for this study include seasonal average wave heights, directions and periods as well as return period statistics and event probabilities. An event probability is defined as $P(H_s \geq \theta)$ where θ is a predefined wave height threshold (3.5m herein). The univariate wave height return period is used herein as it is a simple measure and easily compared to previous studies. The return period is defined as (Goda, 2008; Salvadori, 2004)

$$RP = \frac{\mu_T}{1 - p} \quad (6.11)$$

where μ_T is the average inter-arrival of storms and p is a probability level.

6.3 Results

6.3.1 CP Simulation Statistics

Table 6.2 Seasonal wave climate statistics between the Simulated and Observed data. Standard errors for the sampled statistics (one standard deviation) are shown in brackets.

| Season | Summer | | Autumn | | Winter | | Spring | |
|------------------------|-------------|------|-------------|------|-------------|------|-------------|------|
| | Sim | Obs | Sim | Obs | Sim | Obs | Sim | Obs |
| $H_s(m)$ | 1.6 (0.01) | 1.6 | 1.7 (0.01) | 1.7 | 1.7 (0.01) | 1.6 | 1.8 (0.01) | 1.6 |
| T (s) | 10.5 (0.03) | 9.4 | 10.2 (0.05) | 10.3 | 10.3 (0.04) | 10.8 | 10.5 (0.03) | 9.5 |
| D (°) | 128 (0.3) | 130 | 135 (0.3) | 136 | 142 (0.3) | 143 | 131 (0.3) | 132 |
| $P(H_s \geq \theta)$ % | 0.39 (0.08) | 0.32 | 2.68 (0.22) | 2.89 | 1.53 (0.17) | 1.20 | 0.96 (0.11) | 0.70 |

Table 6.2 shows statistics for mean wave height, mean direction, mean period and

Stochastic simulation of regional wave climates conditioned on synoptic scale meteorology. Part1: Methodology

$P(H_s \geq \theta)$ obtained from all the simulated sequences as well as the observed data. Values from the simulated sequences are shown with their standard deviations. The seasonal mean values show little variability. Only small differences are apparent between observed and simulated values for wave periods and directions. The simulated wave series accurately match the observed exceedance probability statistics of the extreme wave events. It is therefore evident that the CP based simulation technique presented herein is able to simulate wave climates that accurately reproduce the observed wave statistics.

Fig 6.7 shows the average anomaly patterns for all CP classes. Large wave events are strongly associated with the CPs in row 2 column 3 and row 3 column 1. The orientation of the high-low pressures associated with these CPs drives strong winds and subsequently large waves towards the east coast of South Africa.

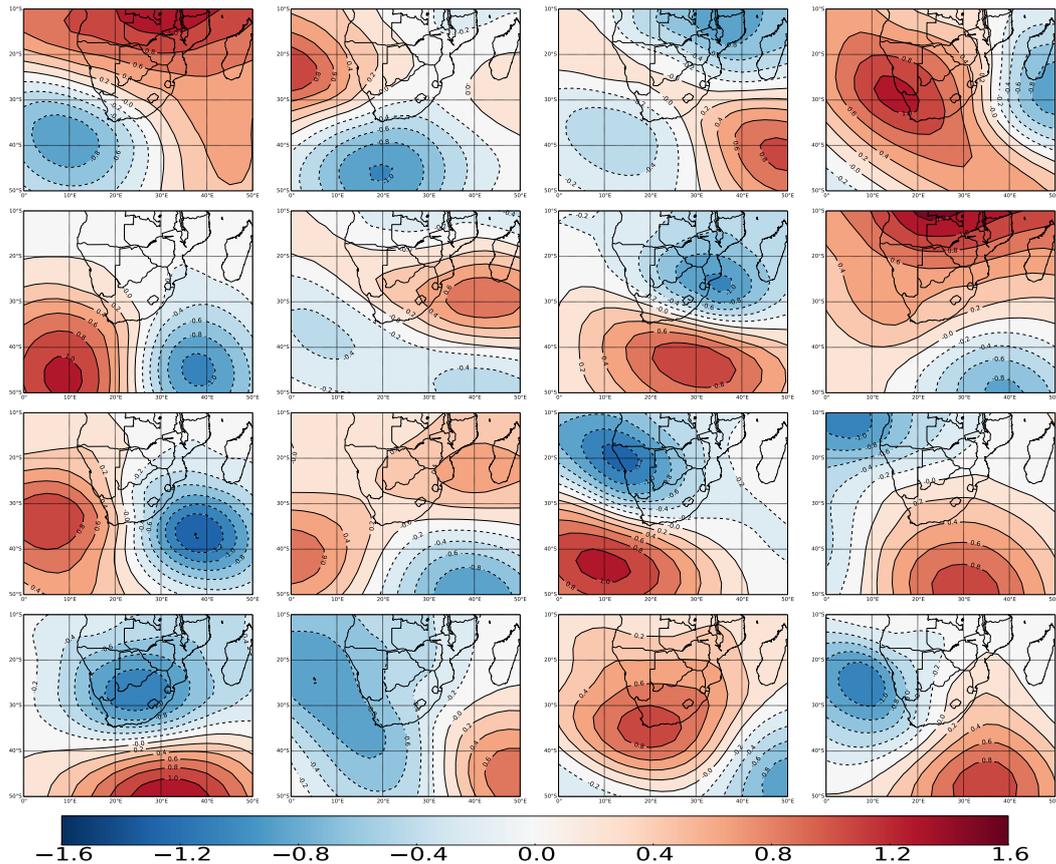


Fig. 6.7 Average anomaly patterns for all CP classes 1 – 16.

6.3.2 Wave climate dependence structures

Fig. 6.8 shows scatter plots (both observed and simulated data) that indicate the correlation between the wave height at times t_i and t_{i+1} , referred to herein as the lag one dependence structure for significant wave heights. The data from three CPs are included in Fig 6.8 as examples of those associated with large and smaller wave heights for summer and autumn. In general autumn months are associated with the largest waves and summer with the smallest (Corbella & Stretch, 2012d). It is apparent that the N12 Archimedean copula models the lag one wave height dependence structure well. This copula exhibits positive dependence in both the upper and lower tails of the distribution. It is expected that the temporal structure of waves follows this pattern i.e. large waves follow large waves and the same goes for small waves. This is reflected in the value of the Kendall's tau statistics, which were in the range 0.5 – 0.8 for all CPs and for all seasons.

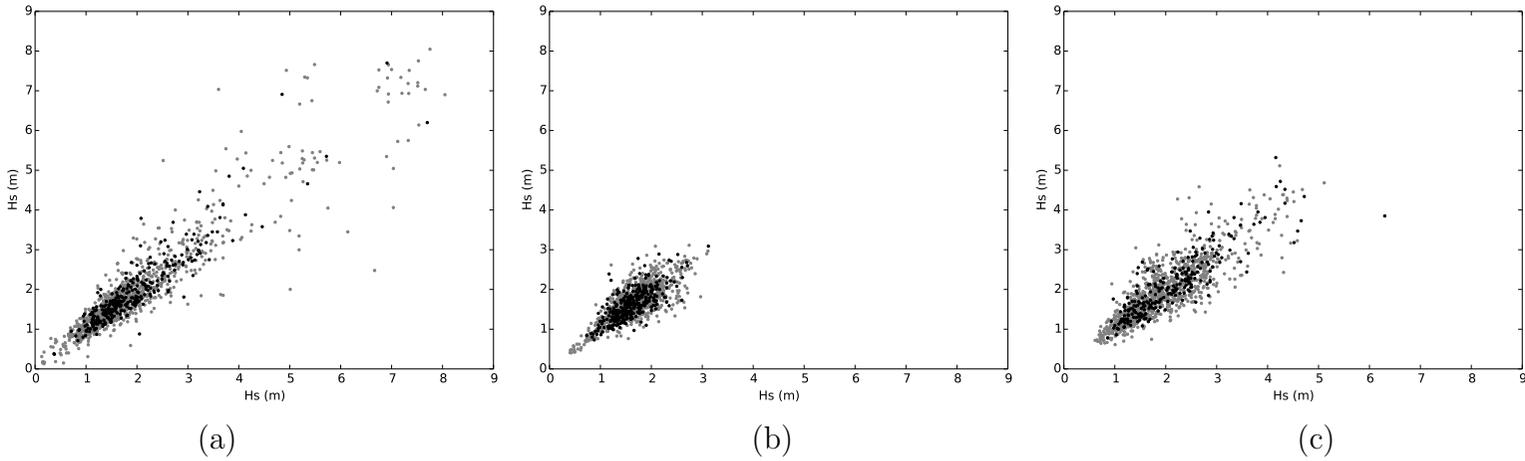


Fig. 6.8 Observed and simulated lag one wave heights for particular CPs: CPs associated with large wave heights (a&b) and (b) a CP associated with low wave heights. (a) & (c) show CPs associated with large waves and their dependence structure for Autumn, (b) shows a CP associated with small waves for the summer months. The black dots indicate the observed temporal structure and grey dots are simulated from the N12 copula.

The relationship between wave height and wave period is more complex and Fig 6.9 reveals some interesting features. There is little correlation between the two variables for small to medium wave heights. However Fig 6.9 shows a stronger correlation at larger values. The clustering of simulated periods at small wave heights can be attributed to the minimum wave height threshold imposed on the simulation and discussed in § 6.2.6. It is interesting to note that the observations of wave height and period appear to cluster in streaks of constant periods for varying wave heights. This can be attributed to the quantisation of the wave periods in the processing of the wave buoy data.

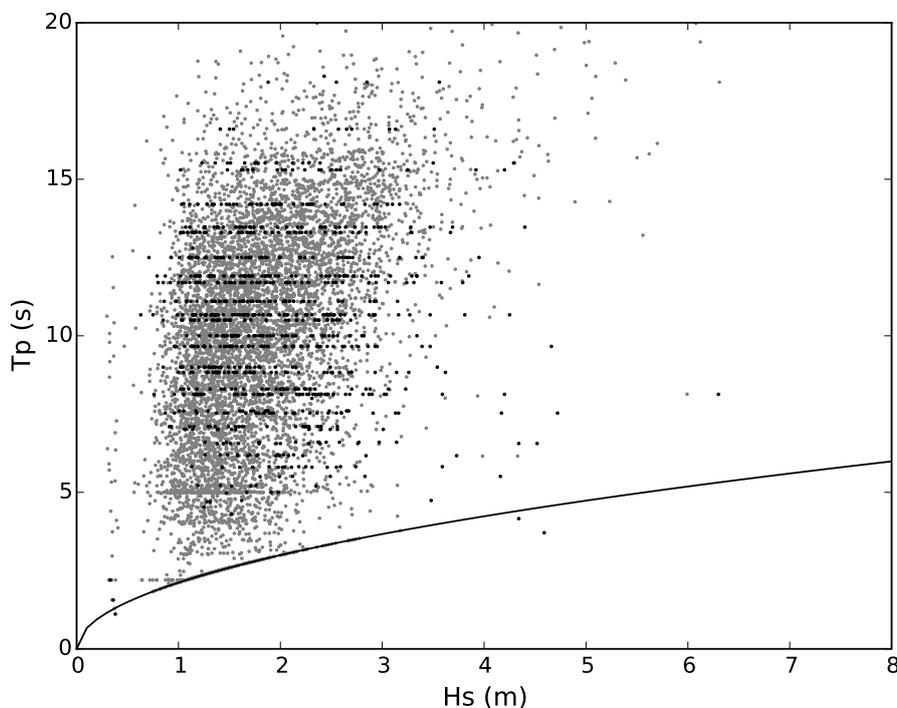


Fig. 6.9 Observed (black dots) and simulated (grey dots) scatter plots of wave heights and wave periods for a particular CP. The solid line indicates the wave steepness limit using Eq. 6.6.

6.3.3 Univariate Return Periods

Comparison of the return periods from simulated and observed significant wave height series suggest they are in good agreement. Fig 6.10 shows the univariate wave height extrema analysis and associated return periods for the simulated (red line), and ob-

Stochastic simulation of regional wave climates conditioned on synoptic scale meteorology. Part1: Methodology

served data (scatter points). Results using the ERA-Interim reanalysis data are also shown (dashed line). To account for nearshore effects the ERA-Interim reanalysis data was transformed from deep water to the location of the wave rider (see Fig. 6.1) using the spectral wave model SWAN (Simulating WAVes Nearshore) (Booij *et al.*, 1999). The 80th and 20th percentiles of the wave height extrema were obtained from the simulated wave sequences and are included in the plot. Estimated significant wave height return periods from the study of Corbella & Stretch (2012d) are shown as the solid black line. The return periods were calculated using the peak over threshold (POT) technique used in Corbella & Stretch (2012d). In their study a monthly window was used to delineate between independent events, the threshold was defined as $H_s \geq 3.5m$.

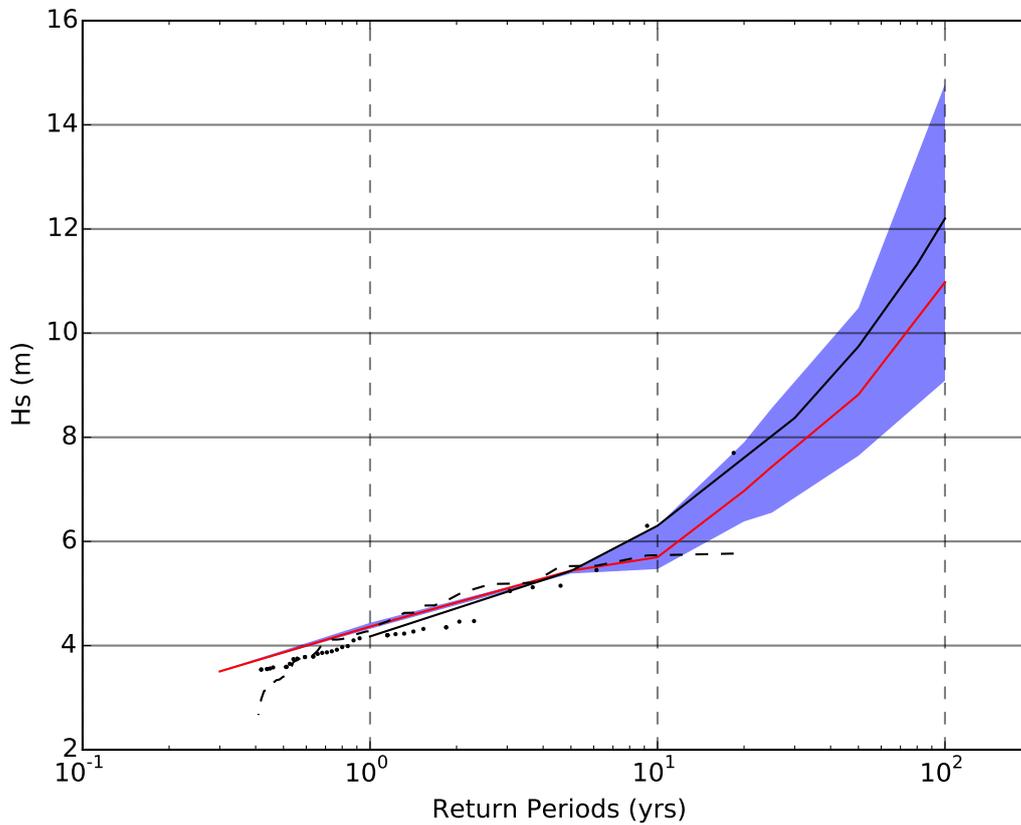


Fig. 6.10 Simulated (red line) wave height return periods for the KwaZulu Natal Coastline. Scatter points show the observed values, the black solid line is the fitted extreme value distribution after Corbella & Stretch (2012d) and the dashed line shows the ERA-Interim reanalysis data. The shaded region indicates the 80th and 20th percentile of wave height extreme obtained from the markov simulation.

In the case of the simulated wave heights greater uncertainty is expected at very

large wave heights because they occur seldom within the 100 year sequence. The simulated wave heights are larger than the observed values at low return periods. However the model captures the general trend in wave height extrema well. For example the return period associated with the largest observed wave height on record ($H_s = 8.5m$) is approximately 30 yrs which agrees with the study by Corbella & Stretch (2012d). The reanalysed wave data from the ERA-Interim dataset shows two interesting features. Firstly at small return periods (< 2 yrs) the wave heights match the simulated waves well. However the returns periods for wave heights above 5 m do not agree well. This suggests that the reanalysed wave data does not reproduce event peaks well, a feature discussed in detail by Caires *et al.* (2004); Swail & Cox (1999). Furthermore this indicates that caution should be exercised when using reanalysed wave data from global models for the analysis of wave height extrema.

6.3.4 Wave Climate Distribution

An important model validation exercise is to demonstrate that it is capable of reproducing the correct wave directional statistics. Fig 6.11 shows wave roses of observed and simulated wave climates for given CPs. The Fig. also shows where the majority of waves come from and their associated wave heights (energy). It is evident from Fig 6.11 that the simulated waves reproduce the correct wave directional distribution for a given CP. Furthermore the spatial orientation of the high and low pressures within the associated CPs can be linked to the shape of the wave roses. This is expected because the CPs are the fundamental drivers of the wave climate. The top two panels show pressure anomalies consistent with wave directional statistics. However the relationship between CP orientation and wave direction is not as clear in the bottom panel. The orientation of the pressure gradient in the bottom panel suggests a strong east-southeasterly component in the wave directional statistics but the wave rose shows a strong south-southeasterly component. Since the anomalies are derived from the 700 hpa pressure level, wind fields at the surface can differ in direction from that suggested by the orientation of the pressure gradients in the CP. This is due to frictional effects near the surface and is demonstrated in Fig 6.11 where the CP averaged surface wind fields are shown. The surface winds reveal interesting features pertaining to wave directional statistics. For example winds in the top panel along the east coast of South Africa have a southwest component, however observed wave statistics suggest the waves are arriving from the south-southeast. Similarly winds in the middle panel suggest waves will have a strong easterly component but observations

show a strong east-southeast component. Therefore the differences between wave and wind directions is attributed to refraction effects that are a function of the coastline orientation and the location of the wave-rider buoy (see Fig 6.1). Corbella & Stretch (2014a) explored the effects of refraction on wave data from the wave-rider buoy in Durban and indicated that deep water waves could be refracted by between 6° and 16° .

6.4 Discussion

Recent developments in statistical wave climate modelling have given new insights into complex wave environments, particularly with respect to the dependence structure of wave characteristics. Furthermore these models are computationally efficient and therefore provide a simplified means to simulate many different possible scenarios. A disadvantage of these techniques is they are typically not directly linked to the physical drivers of regional wave climates. This study attempts to bridge that gap through the development of a new approach that uses synoptic scale atmospheric circulation patterns as a framework within which to simulate wave climates.

Simulated time series of wave climate variables using the techniques described have been shown to reproduce the statistics of observed values. A key feature of the simulated waves is that they include the correct number of extreme events. This is important for risk assessment studies. Furthermore the temporal dependence structure is well modelled thus ensuring that the ordering of the simulated waves is correct. The wave height return periods associated with the simulated waves is in good agreement with those previously deduced by Corbella & Stretch (2012d) from observed wave heights. It is also important to note the limitations in analyses based on the ERA-Interim wave data. The ERA-Interim wave data seems to reproduce observed statistics for wave heights below 5m but not for wave heights above this threshold. This point has been previously discussed by Caires *et al.* (2004); Swail & Cox (1999) and illustrates the continuing need for accurately measured data and statistical models for the analysis of extreme events.

The decision to model wave periods conditioned on wave heights is based on the observation that large wave heights typically occur simultaneously with large wave periods (Fig 6.9). This trend is reproduced in the simulated data. However due to the dispersion relationship for water waves it is possible that given a sequence of wave heights those with large periods can arrive before those with large wave heights and

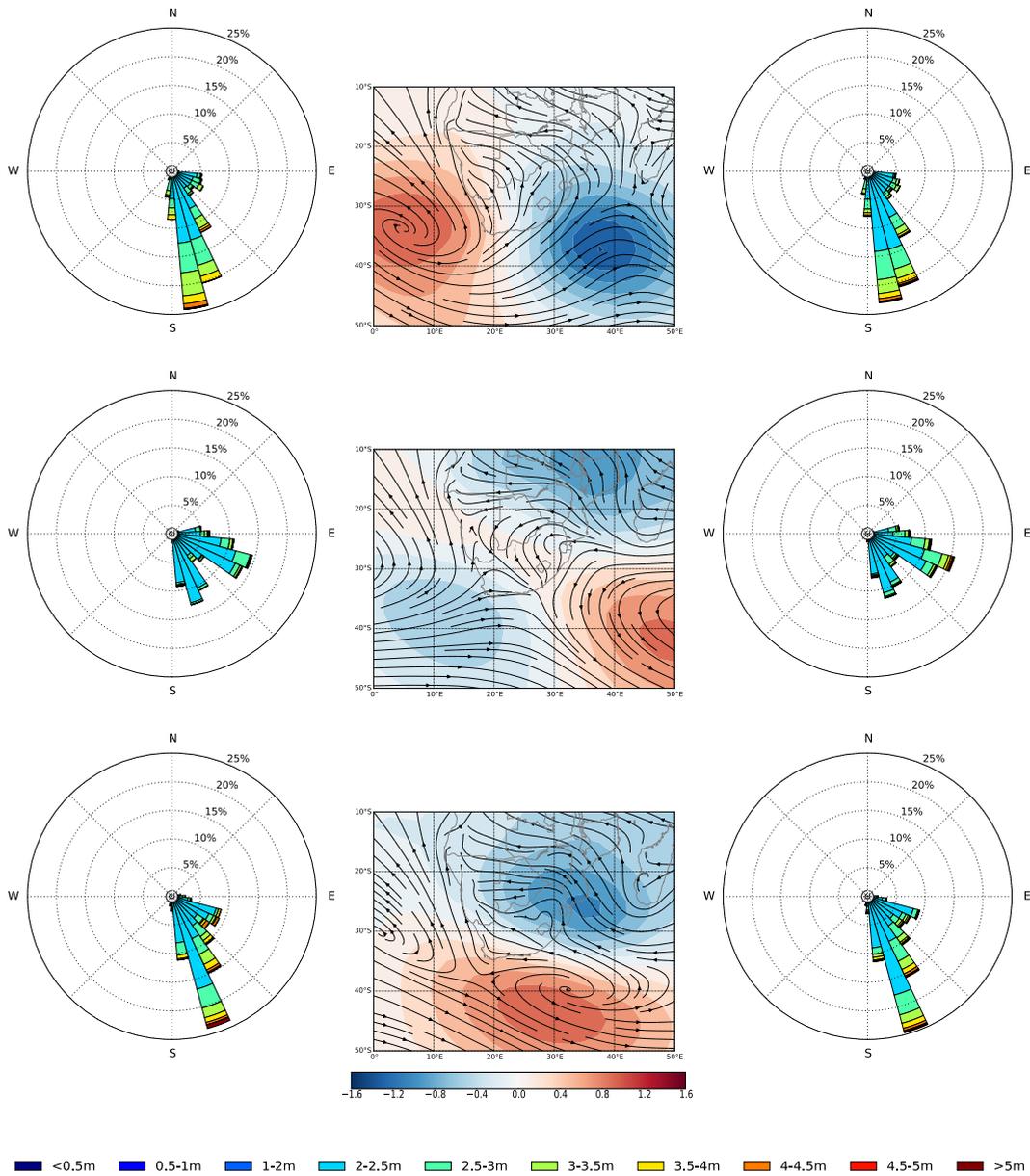


Fig. 6.11 Wave roses for observed (left) and simulated (right) wave climates with the associated CPs (centre). Surface wind streamlines are also shown in the plot. The line width indicates the relative magnitude of the wind velocity. Significant wave heights for the wave roses are given by the colour legend below the plot.

Stochastic simulation of regional wave climates conditioned on synoptic scale meteorology. Part1: Methodology

shorter periods.

Wave directions were sampled from the empirical distribution conditioned on the wave height for each CP. The simulated wave climate exhibits similar characteristics to the observed values in terms of their relative frequencies. Including the wave direction in the simulation has strong implications for future climate prediction due to climate change since changes in CP statistics may have an effect on the directional distribution of the wave energy.

An alternative approach to a CP-based simulation method as presented in this study is to model the wave climate conditioned on direction which in turn is defined by the CP occurrences. For example consider the following steps:

1. At a given time t sample the CP based on CP_{t-1} and the transition matrix.
2. The CP_t is associated with a specific wave direction distribution from which a direction can be sampled.
3. Given this direction it is possible to conditionally sample the associated wave height and period.

This alternative approach assumes that different wave heights and periods come from different directional sectors for any given CP, a property demonstrated in Fig 6.11. It was not used in the present study because of the relatively short record of directional wave data available at the case study site (approximately 5 yrs).

We note that some of the choices made for the modelling presented here are likely to be case study specific. For example the wave height thresholds used in the objective functions discussed in §6.2 for the CP optimisation process. In this case our choices reflect a focus on coastal vulnerability issues, but the algorithm allows for different specifications of the objective functions if the focus is different. The outcome of the CP selection and classification process and associated statistics will then be different. We have not yet explored the effects of this on the wave climate simulations in detail, but they should be investigated in future work.

The approach to simulating waves described herein provides a framework for assessing future wave climate scenarios associated with climate change effects. This is strategically important to coastal planners and engineers when assessing future coastal vulnerability concerns. Pressure data obtained from global climate models (GCMs) for future climate change scenarios can be used to drive this wave simulation technique. Changes in occurrence frequencies for different CPs will directly affect the

wave simulation. This is advantageous because the simulated waves do not require data assimilation which is not an option when simulating future wave scenarios. Furthermore the synthetic wave sequences provide a continuous record from which it is possible to assess nearshore processes and beach response.

6.5 Conclusions

This study presents a new approach for stochastically simulating wave climates based on the occurrence of synoptic scale meteorological circulation patterns. The results demonstrate that this methodology is a viable alternative to current stochastic approaches for simulating regional wave climates. Atmospheric circulation patterns provide an ideal framework for reconstructing the natural dependence structure within wave climate variables such as wave height, direction and period, while copulas provide the required flexibility to achieve this. Our new modelling approach has an advantage of over purely statistical methods since the CPs used for conditioning the simulations are the major drivers of the wave climate and have direct physical links to wave characteristics. This should add reliability to predictions based on such models.

There is a need for additional work to test the methodology at different case study sites and to explore coastal engineering applications of this modelling approach.

Chapter 7

Stochastic simulation of regional wave climates conditioned on synoptic scale meteorology. Part 2: Applications in coastal vulnerability assessment

Abstract

Stochastic models have a long history in the simulation of synthetic data sequences for application to risk assessment e.g. in stochastic hydrology. Their use in coastal engineering has recently grown with the application of multivariate models to assess the risks from sea level rise due to warming of the global climate. In this study we illustrate some typical coastal vulnerability applications of a recently developed stochastic wave climate model that has links to synoptic scale atmospheric circulation patterns (CPs). The modelling approach used allows for the simulation of a long continuous time series of waves that are linked to their associated CPs. The time series can be used for various types of coastal vulnerability assessments. Several examples are illustrated using the east coast of South Africa as a case study. The results demonstrate the robustness of the modelling technique and the potential applications in coastal vulnerability assessments. The link to CPs facilitates the application of these methods to the assessment of future climate changes on coastal vulnerability.

7.1 Introduction

Recent advances in coastal vulnerability and risk assessment have seen the use of probabilistic models to stochastically simulate regional wave climates (e.g. Callaghan *et al.*, 2008; Corbella & Stretch, 2013; Pringle & Stretch, 2015). An advantage of these approaches compared to process-based wave models is their cost effectiveness for simulating long records. This is beneficial where wave observations are limited and simulations can be used to extend wave datasets for risk assessment studies. Although process based models aim to resolve the physical drivers of wave development they are computationally expensive for long term simulations and are therefore limited to shorter time durations.

Another approach is to use global wave hindcast datasets that are available from wave modelling centres such as NCEP (National Centre for Environmental Predic-

tion) or ECMWF (European Centre for Medium-Range Weather Forecasting). These datasets typically span 30 - 40 years. However the accuracy of global wave models relies strongly upon reanalysis and accurate wind field forcing (Caires *et al.*, 2004; Chawla *et al.*, 2013; Tolman *et al.*, 2002). Chawla *et al.* (2013); Stopa & Cheung (2014) found that in general the NCEP and ERA datasets under-predict extreme events nearshore. However the NCEP model better describes wave height variability, especially in the upper percentiles, whereas the ERA model is more homogeneous through time and therefore better suited to long term modelling (Stopa & Cheung, 2014). Therefore the best choice of model is application specific and both may be inaccurate in characterising extreme events which are of fundamental importance in risk assessment studies.

Recently Pringle & Stretch (2015) developed a new stochastic simulation technique that is linked to the physical drivers of regional wave climates, namely synoptic scale circulation patterns (CPs). Their modelling approach was to simulate a continuous wave sequence rather than simulating independent storm events with various durations and inter-arrival times as was previously done by Callaghan *et al.* (2008); Corbella & Stretch (2013) amongst others.

Espejo *et al.* (2014); Pringle *et al.* (2014) were the first to exploit links between CPs and regional wave climates in coastal vulnerability studies. Although their approaches differed they were both based on the statistical classification of CPs into classes or states that exhibit similar properties. The classification technique used in Pringle *et al.* (2014) and in the present work, is based on fuzzy logic and wave data is explicitly used to guide the algorithm to a suitable solution. Both studies have presented a strong basis from which it is possible to evaluate coastal risk in a simplistic manner. For example Camus *et al.* (2014) used the CPs to downscale multivariate wave climate from global wave hindcast datasets and Pringle & Stretch (2015) have utilized the CPs to simulate regional wave climates.

Atmospheric CPs present a natural way to describe transitions between wave states. For example some CPs are associated with extreme wave events while others are associated with calmer periods and the transition between states depends naturally on the development and movement of CPs. Furthermore the CPs contain information on the dependence structure between wave climate variables such as wave height (H_s), wave direction (θ) and wave period (T). Therefore at each time step it is possible to estimate the wave height, direction and period conditional on a particular CP occurrence.

Pringle & Stretch (2015) used a case study to demonstrate their model's efficacy in reproducing key statistics from the observed wave climate. In this follow-on paper we focus on demonstrating specific applications of the new model as a tool for coastal vulnerability analyses including: (a) the seasonal distribution of coastal erosion and/or accretion due to cross-shore processes; (b) longshore sediment transport processes; and (c) long-term shoreline evolution using shoreline models, in this case the recent model of Jara *et al.* (2015). Observations from a case study site are used to validate the modelling results where possible and illustrate the utility of the new wave modelling approach.

7.2 Method

7.2.1 Case Study Site

The wave climate along the east coast of South Africa is driven by a number of weather systems that result in an energetic environment (Preston-Whyte & Tyson, 1988; Rossouw *et al.*, 2011; Taljaard, 1967). The three most common weather systems that drive large wave events are mid-latitude cyclones, cold fronts and tropical cyclones (Corbella & Stretch, 2012c; Mather & Stretch, 2012; Schumann, 1988; Taljaard, 1967). Although the contribution of tropical cyclones is questionable, those that become stationary south of Madagascar can drive large waves towards the east coast of Southern Africa (Mather & Stretch, 2012).

Durban is a coastal city located on the east coast of South Africa. Given its highly energetic coastline and the location of its port, the city has constantly battled against coastal erosion. Subsequently long term beach monitoring and a sand by-pass scheme have been employed as mitigation strategies (Barnett, 1999). Since 1973 the local authorities have recorded beach profiles every 3 months along a 100km length of coastline extending both south and north from the city. Fig 7.1 shows the locations of the Durban wave rider buoy and recorded beach profile used herein.

7.2.2 Data Sources

Wave height data used for the CP classification and wave simulation model were obtained from two wave rider buoys off the coast of KwaZulu Natal at Durban and Richards Bay. The data was for the period 1992 – 2009 at 3-hour intervals. However wave directional data have only been recorded from 2007. The Richards Bay wave

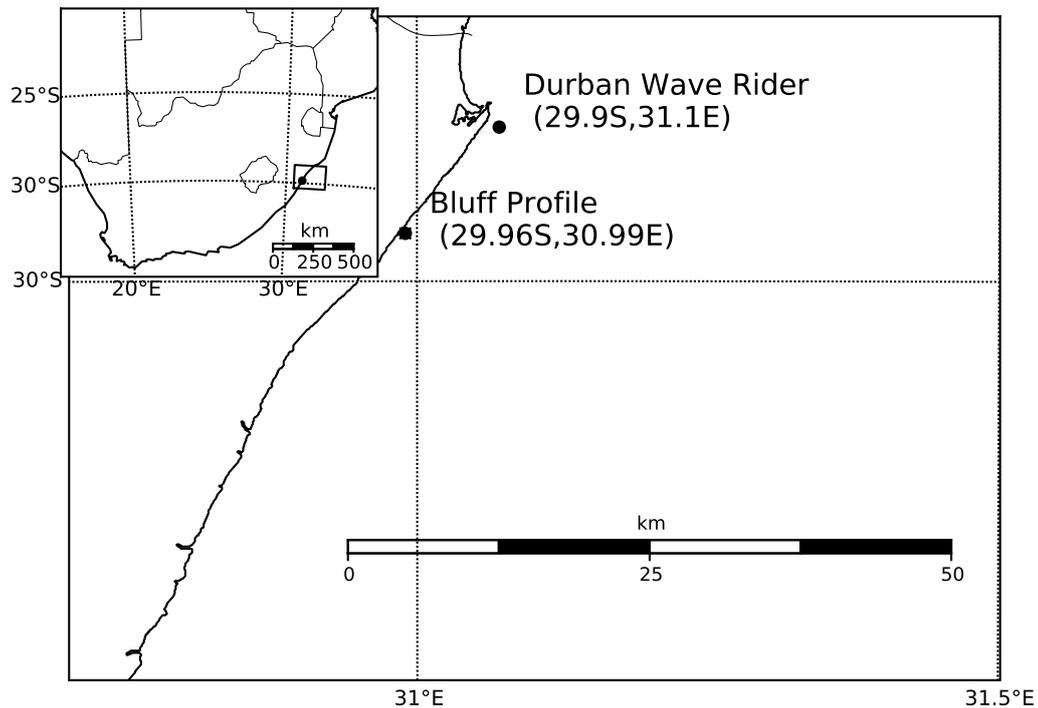


Fig. 7.1 Locations of the wave observation buoy at Durban and the Bluff beach profile used herein along the KwaZulu Natal coastline (Pringle *et al.*, 2014).

rider data was used to supplement missing data within the Durban record. Corbella & Stretch (2012c) showed a strong correlation between the two datasets.

Atmospheric pressure data used in the classification algorithm was obtained from the ERA-Interim dataset (<http://apps.ecmwf.int/datasets/>) for the period 1979 – 2012 at a grid resolution of $2.5^\circ(10^\circ\text{S } 0^\circ\text{E} - 50^\circ\text{S } 50^\circ\text{E})$ at 6-hour intervals.

7.2.3 Automated Classification

The objective based classification was applied to anomalies within the 700 hPa geopotential. Detailed descriptions of the algorithm are given in Bárdossy *et al.* (1995, 2002); Pringle *et al.* (2014).

The algorithm comprises of two components. Firstly a *classification method* that is based on fuzzy logic is applied to the anomalies and groups them into a set of distinct classes. The fuzzy logic is used in a manner that quantifies subjective statements such as ‘high pressure’ or ‘low pressure’. Furthermore the fuzzy logic is exploited to locate and group patterns with similar spatial orientations and locations of high/low pressure regions. Secondly an *optimization procedure* is then used to maximise an

Stochastic simulation of regional wave climates conditioned on synoptic scale meteorology. Part 2: Applications in coastal vulnerability assessment

object function based primarily on the wave height (H_s) and properties thereof. The objective function is described as

$$O = f(g_1(\overline{H_s}), g_2(p(H_s \geq \theta | CP(t), p(\theta))); w_1, w_2) \quad (7.1)$$

where g_1 and g_2 are functions based on the following statistical properties of the wave height: the average wave height $\overline{H_s}$ and $p(H_s \geq \theta | CP(t))$, the probability of a wave heights being above a certain threshold θ conditioned on the CP occurrence at time t . The weights w_1 and w_2 express the relative importance of the two functions g_1 and g_2 . The objective function is used in such a manner as to derive CP classes with statistics different from the mean.

7.2.4 Wave Climate Simulation

A brief overview of the wave climate simulation model is given here with full details provided in Pringle & Stretch (2015).

As mentioned earlier atmospheric CPs classified into distinct classes contain important information about wave climate variables and the dependence structure that links them. Furthermore the transitions between CP classes can be described as a Markov process. Therefore given a set of wave climate variables (H_s, T_p, θ) and a CP at some time t_i it is possible to predict the CP and subsequently the wave climate variables at time t_{i+1} . Hence the model depends on two important parameters, namely the occurrence of a particular CP at time t_i and the dependence structure between wave climate variables given the CP at time t_{i+1} . The technical steps required for the wave climate simulation are as follows (Pringle & Stretch, 2015):

1. Calculate relative statistical properties for each CP and derive the CP transition matrix.
2. Calculate the distributions of wave climate variables for each CP and evaluate dependence structures.
3. Simulate a pseudo-random CP sequence based on the transition matrix for N years at a 6-hour interval.
4. Simulate wave climate variables based on the CP sequence and associated CP based dependence structures.
5. Repeat steps 3 – 4 M times.

Pringle & Stretch (2015) used bivariate copulas of the Archimedean family to model the dependence structure between wave climate variables. The dependence structure modelled included the lag one temporal wave height structure ($H_s(t_i)|H_s(t_{i-1})$) and the wave period (T_p) conditioned on H_s . Wave directions for each CP were modelled from their empirical distribution conditioned on the wave height (H_s). This decision was based on the limited amount of directional wave data. Furthermore directional wave data for each CP showed a strong association with the CPs able to delineate the directional spread of wave energy into clearly defined sectors.

7.2.5 Model Validation

The model was applied to several validation exercises in order to test its ability to accurately describe regional wave climates and the nearshore processes associated therewith. The east coast of South Africa was used as a case study. One important aim of the validation exercises was to demonstrate the usefulness of simulating an entire wave record. The wave model was run at a 6-hour temporal resolution for 100 years and this was repeated 101 times to obtain a representative sample of all possible wave climate scenarios. Given the long temporal simulations of synthetic waves it is important to select validation exercises that are not computationally demanding.

Seasonal Wave Energy Directional Statistics

Statistical properties of the directional wave energy for each season are presented using wave roses. The wave roses provide a means to independently compare directional wave energy sources and their occurrence probabilities for the simulated and observed wave climates. Wave energy E is directly related to the significant wave height (H_s) squared i.e. $E \propto H_s^2$. Occurrence probabilities for wave heights from different directional sectors were calculated for each season defined in Table 7.1.

Table 7.1 The allocation of months to seasons

| Season | Months |
|--------|----------------------|
| Summer | December – February |
| Autumn | March – May |
| Winter | June – August |
| Spring | September – November |

Longshore Sediment Transport

Longshore sediment transport is important in that it is one of the factors that influences long term shoreline change, design of breakwaters and navigation channels etc (Schoonees & Theron, 1993). Therefore it is a nearshore process that is fundamental to accurate coastal vulnerability assessments. There are numerous empirical longshore transport formulas available, for example the CERC formula (CERC, 1984b), the Van Der Meer equation (Van der Meer, 1990) and the Kamphuis equation (Kamphuis, 1991): they have been reviewed by Schoonees (2000); Wang *et al.* (1998). The Kamphuis equation was used here, where the longshore transport Q (m^3/year) is given by

$$Q = 6.4 \times 10^4 H_{sb}^2 T_p^{1.5} m^{0.75} d^{-0.25} \sin^{0.6}(2\theta_b) \quad (7.2)$$

where H_{sb} is the significant wave height at breaking (in metres), T_p is the peak wave period (in seconds), m is the beach slope, d is the water depth (in metres) and θ_b is the wave angle at breaking. One reason this model was selected is because it includes both wave period and direction as drivers of longshore transport which illustrates the utility of a wave climate simulation technique that includes wave height, direction and period.

Cross-shore Sediment Transport

In order to calculate the cross-shore profile response given the long temporal simulation of the synthetic wave climate, a simple and reliable erosion structural function is required. The function should not be computationally demanding and should be easily incorporated into a Markov chain process. The time convolution shoreline response model based on the concept of equilibrium profiles and proposed by Kriebel & Dean (1993) is a suitable tool given the above criteria. Fig 7.2 shows an example of the equilibrium shoreline profile concept proposed by Kriebel & Dean (1993). A similar approach was recently adopted by Callaghan *et al.* (2008); Corbella & Stretch (2012b) to estimate erosion from simulated wave storm sequences. An alternative approach would be to use a process-based or semi-empirical model such as XBEACH or SBEACH (Larson *et al.*, 1990; Roelvink *et al.*, 2009). Although they are computationally demanding it is possible to use these models to estimate cross-shore responses for synthetic wave sequences (e.g. Corbella & Stretch, 2012b; Ranasinghe *et al.*, 2013). However Ranasinghe *et al.* (2013) argue that the increase in model sophistication was not justifiable even though XBEACH was shown to most accurately predict the beach

profile.

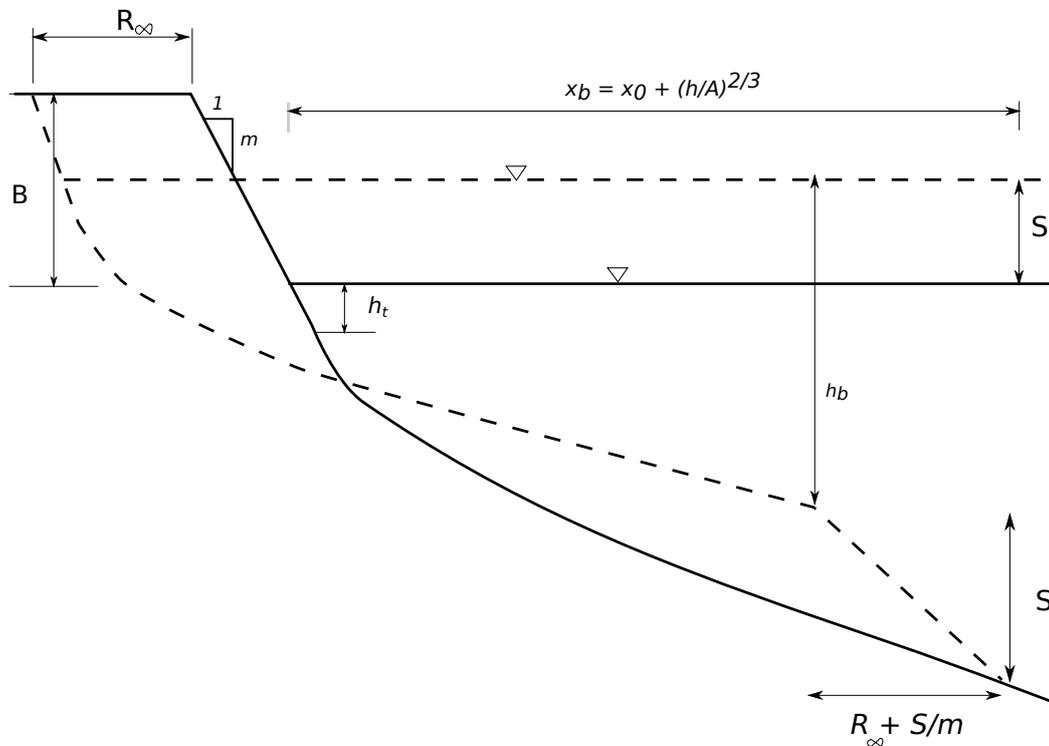


Fig. 7.2 Equilibrium shoreline response to an increase in water level (S) according to Kriebel & Dean (1993).

According to Kriebel & Dean (1993) equilibrium beach profiles can be described by

$$h = \begin{cases} -B & x \leq -\frac{B}{m} \\ mx & -\frac{B}{m} \leq x \leq \frac{4A^3}{9m^3} \\ A(x - x_0)^{\frac{2}{3}} & x \geq \frac{4A^3}{9m^3} \end{cases} \quad (7.3)$$

where m is the beach-face slope estimated from beach profile data, B is the height of the berm and A is an empirical constant based upon the particle size and settling velocity. The beach response is calculated using the convolution method such that

$$R(t) = \alpha R_{\infty} \int_0^t f(t) e^{-\alpha(t-\tau)} dt \quad (7.4)$$

where $R(t)$ is the berm retreat at time t , α reflects the erosion time scale, τ is a time lag and $f(t) = \sin^2\left(\pi \frac{t}{D}\right)$, D is the total storm duration, and R_{∞} defines the maximum

Stochastic simulation of regional wave climates conditioned on synoptic scale meteorology. Part 2: Applications in coastal vulnerability assessment

retreat at $t = \infty$ under steady forcing conditions. In this case

$$R_\infty = \frac{S \left(x_b - \frac{h_b}{m} \right)}{\left(B + h_b - \frac{S}{2} \right)} \quad (7.5)$$

where S is the total surge (sum of the storm surge and wave setup). The volume eroded is then calculated so that

$$\frac{V(t)}{V_\infty} = \frac{R(t)}{R_\infty} \quad (7.6)$$

which is based on the assumption that the elevation contours erode at the same rate. The maximum eroded volume above mean sea level is given by

$$V_{m\infty} = R_\infty B + \frac{S^2}{2m} - \frac{2S^{5/2}}{5A^{3/2}} \quad (7.7)$$

The convolution model is used herein as a validation exercise. The concept of eroded volume return periods is used to compare the simulated waves to the observed waves. However it is noted that since the storm surge was not simulated, the total surge S was based only on the maximum wave setup $\bar{\eta}_{max}$ given by

$$\bar{\eta}_{max} = \frac{40 - 3\gamma_b^2}{128} \gamma_b H_b \quad (7.8)$$

where $\gamma_b = 0.78$ and H_b is the breaking wave height. Eq. 7.8 is assumed for saturated wave conditions (Dean & Dalrymple, 1991).

A disadvantage of the Kriebel & Dean (1993) model is that the solution to Eq. 7.4 requires an idealized forcing function $f(t)$. This is a problem when evaluating continuous wave records as wave forcing cannot be easily idealized with a simple function (Miller & Dean, 2004). Therefore the Kriebel & Dean (1993) model (KD93) used herein was used to estimate beach erosion volumes due to storm events ($H_s \geq 3.5m$) only. The wave height threshold was chosen because observations suggest that wave heights greater than this cause severe beach erosion (Corbella & Stretch, 2013). Furthermore the observed shoreline response to waves greater than the threshold is shown in Fig. 7.8. The robustness of the simulated waves was tested against the observed wave record by evaluating the erosion return periods. The return periods were calculated using the peak over threshold method. Independent storm events were identified using a minimum two week inter-arrival time based on autocorrelation (Corbella &

Stretch, 2012b) .

Shoreline Evolution Model

To demonstrate the utility of having a method to simulate continuous wave records a shoreline evolution model was applied to the synthetic dataset. Such models are useful tools that can be relatively simple and easy to incorporate into a Markov Chain sequence. Furthermore it is possible to predict shoreline positions both short and long term.

The model selected for the present work was developed by Jara *et al.* (2015) as an improvement to the model proposed by Yates *et al.* (2009). For ease of reference it will be referred to herein as the dynamic equilibrium model (DEM). Only a brief description of the model is given here – a detailed description is given in the cited references.

The DEM is a model for shoreline position $S(t)$ at time t given by

$$\frac{dS(t)}{dt} = -C^\pm [S(t) - S_\infty] \quad (7.9)$$

where the equilibrium shoreline position S_∞ is assumed to depend on the incoming wave energy E , C^+ is a rate parameter associated with accretion, and C^- a rate parameter associated with erosion. Under steady conditions and given enough time, the shoreline will converge to the equilibrium position for a given wave energy.

The solution to Eq. 7.9 requires a function to relate S_∞ to E . Yates *et al.* (2009) first defined the equilibrium wave energy as a linear function of the shoreline position. However Jara *et al.* (2015) suggested the use of a quadratic function to better describe the relationship between the two variables. The equilibrium energy function (EEF) leads to a set of possible equilibrium beach profiles (EBPs) given the incoming wave energy. Whence Jara *et al.* (2015) define an EBP envelope that encompasses all

Stochastic simulation of regional wave climates conditioned on synoptic scale meteorology. Part 2: Applications in coastal vulnerability assessment

possible EBPs as (refer Figs. 7.3 & 7.4)

$$\left\{ \begin{array}{l} \frac{h_t - h_{bmax}}{\tan\phi} + \frac{(h_{bmax}^{3/2} - h_t^{3/2}) \left(x_t(h_t + B) - V_s - \frac{3}{5}h_{bmax}^{5/2} + Bh_{bmax}^{3/2} \right)}{\frac{3}{5}(h_t^{5/2} - h_{bmax}^{5/2}) + B(h_t^{3/2} - h_{bmax}^{3/2})} = 0, \\ S_{min} = x_t - \left(\frac{h_{bmax}}{A} \right)^{3/2} - \frac{h_t - h_{bmax}}{\tan\phi}, \\ S_{max} = x_t - \frac{x_t(h_t + B) - V_s}{\frac{3}{5}h_t + B} \end{array} \right. \quad (7.10)$$

where S_{min} and S_{max} are the boundary shoreline positions for the EBP envelope, h_{bmax} is the maximum breaker depth, B is the berm height, V_s is the volume of sediment contained in the EBP, x_t is the distance to a stable toe at a water depth of h_t , ϕ is the sediment critical angle of repose and A is a constant based on the sediment settling velocity.

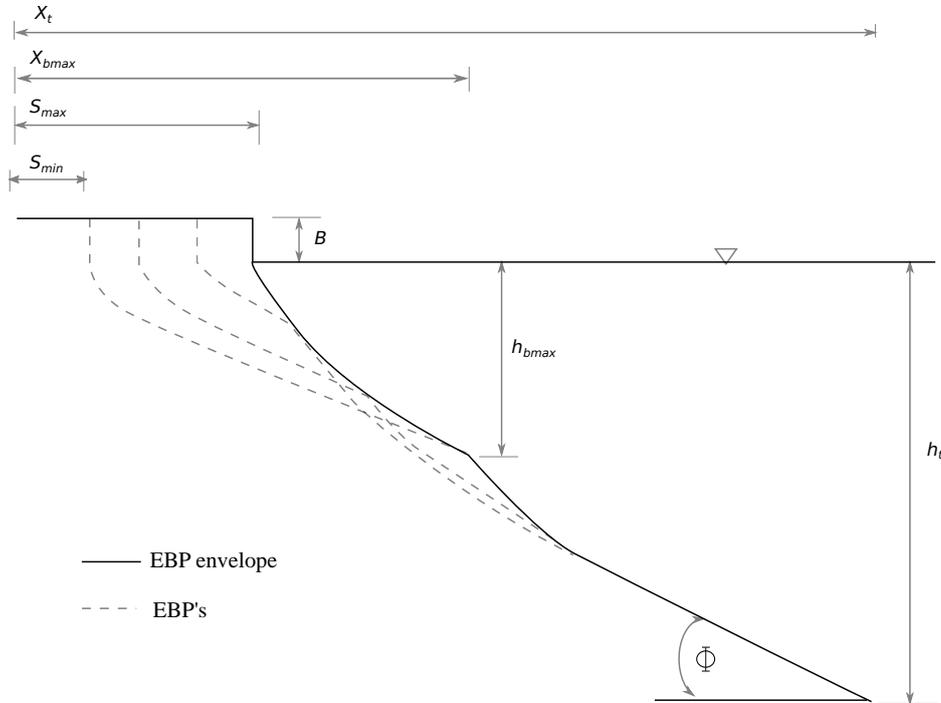


Fig. 7.3 Equilibrium beach profile envelop according to Jara *et al.* (2015). The black line indicates the EBP envelop that comprises all the individual EBPs (dashed lines).

The EEF (Fig. 7.4) is used to determine the state of the current shoreline position relative to the equilibrium shoreline position S_∞ given the incoming wave energy E . If the current shoreline position is larger than S_∞ the beach will erode and visa versa for $S < S_\infty$. The incoming wave energy E is calculated as

$$E = \left(\frac{\gamma_b}{4}\right)^2 h_b^2 \quad (7.11)$$

where $\gamma_b = 0.78$ and h_b is the water depth at breaking. To obtain the maximum energy E_{max} the term h_b is substituted with the value h_{bmax} obtained by Eq. 7.10.

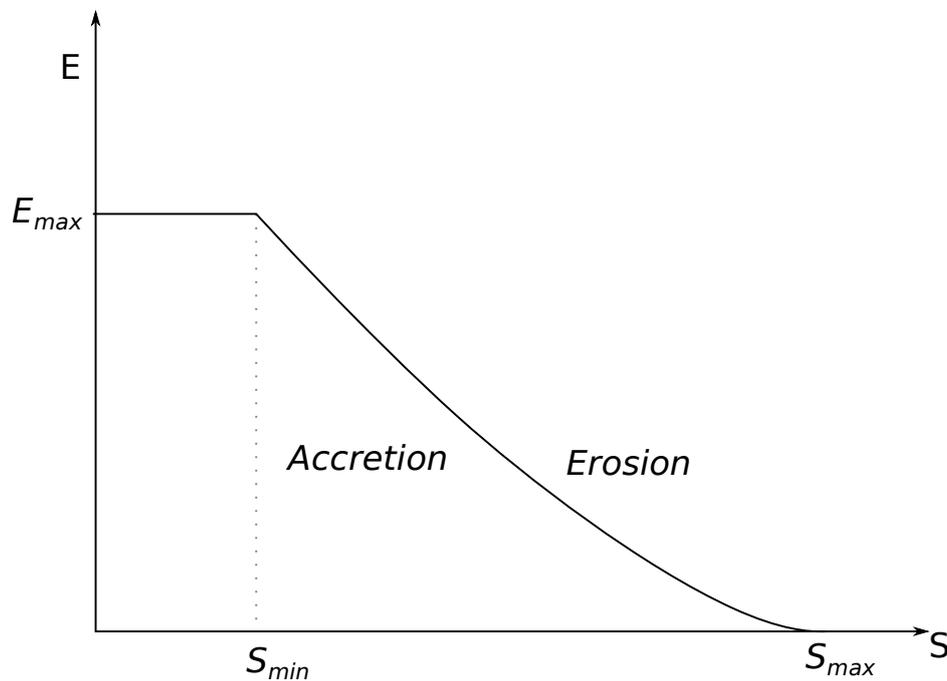


Fig. 7.4 Equilibrium energy function that relates the incoming wave energy to an equilibrium shoreline position (Jara *et al.*, 2015).

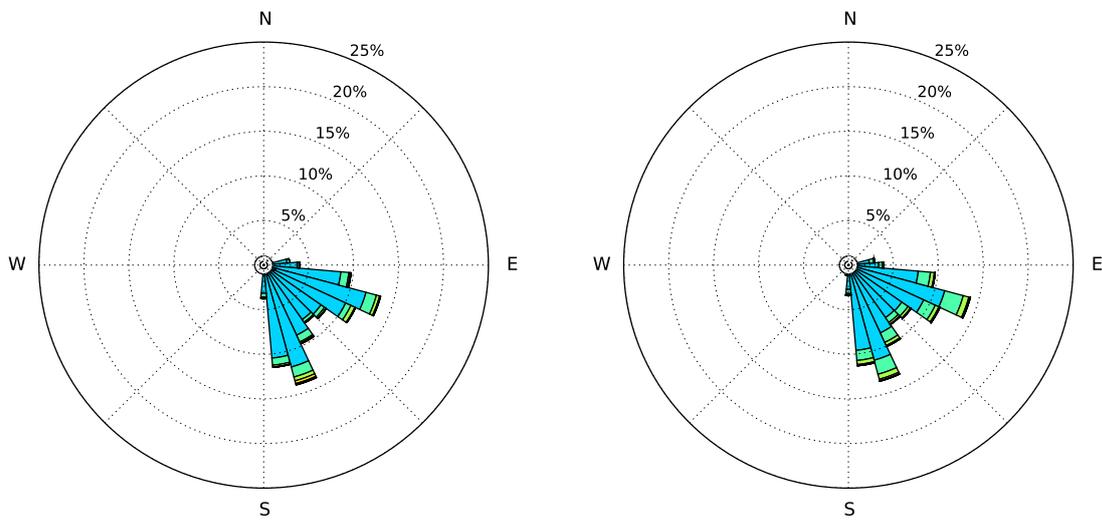
7.3 Results

7.3.1 Observed and Modelled Wave Directional Statistics

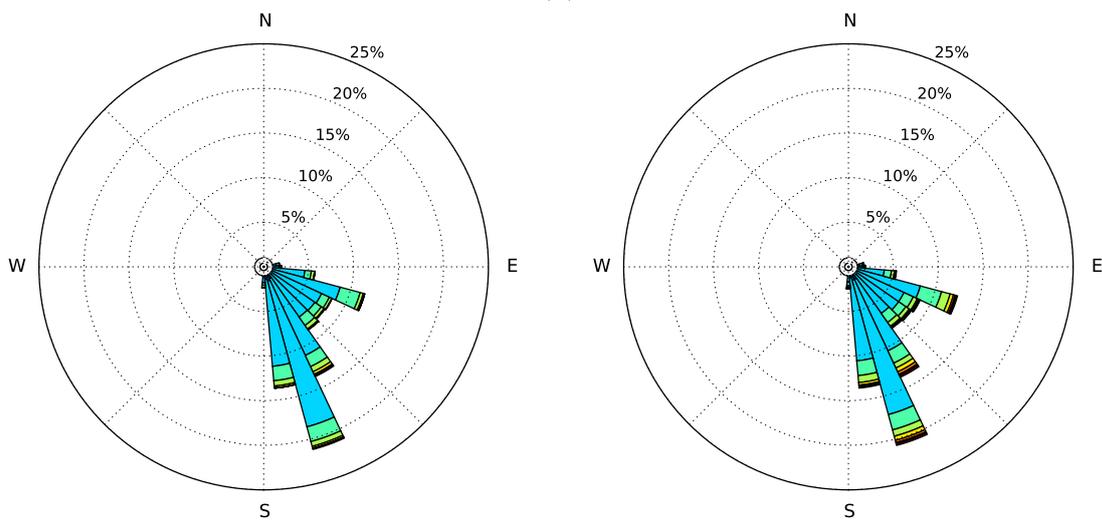
The seasonal wave roses for the observed and simulated wave data are shown in Fig 7.5. In general the modelled data appears similar to the observed. The shape and origin of wave energy for the modelled data is similar to the observed, even in cases where there the directional distribution is bimodal. Furthermore the occurrence frequencies

Stochastic simulation of regional wave climates conditioned on synoptic scale meteorology. Part 2: Applications in coastal vulnerability assessment

of different wave heights for the modeled data compare well with the observed data.



(a)



(b)

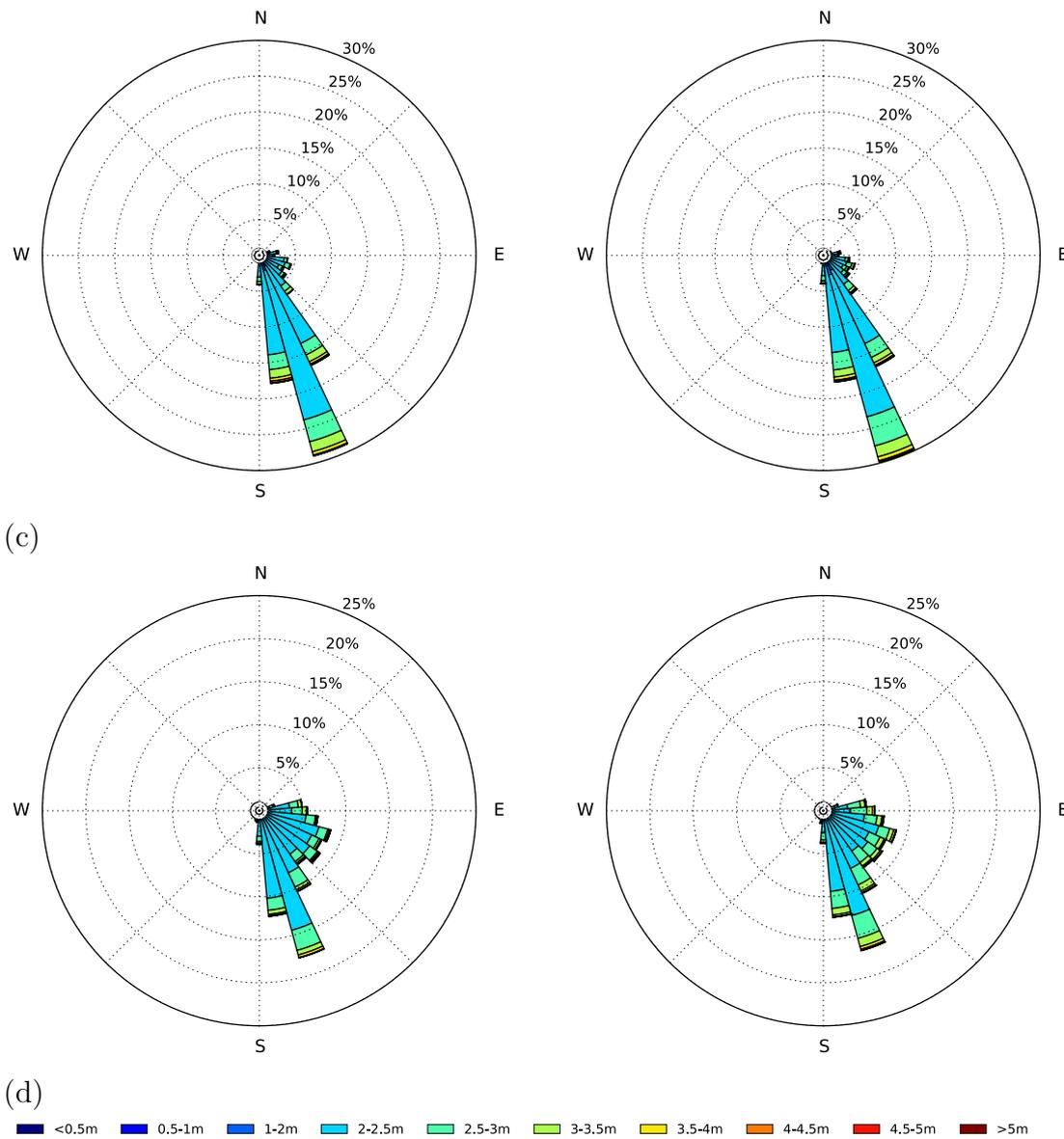


Fig. 7.5 Seasonal wave roses for observed wave data (left panel) and modelled (right panel) for the seasons: (a) Summer, (b) Autumn, (c) Winter and (d) Spring. The significant wave heights associated with the wind rose directions are given by the colour legend shown below the plot.

7.3.2 Longshore Transport

The annual longshore sediment transport rates were calculated according to Eq. 7.2 for both the observed and simulated wave data. Since wave directional measurements were only available from the year 2007, longshore transport statistics calculated from the observed waves are subject to sampling errors due to the short record. Therefore for each 100 year simulated wave climate randomly selected periods spanning the same length as the observed record were sampled.

Fig. 7.6 and Table 7.2 shows that the mean and variance of longshore transport rate calculated from the simulated data is in good agreement with the observed data. However the simulated data is unable to reproduce the bimodal characteristic of the observed transport rates. Furthermore the reasonable closeness between the 5th and 95th percentile confidence limits implies there is relatively little variation of longshore transport within the simulated data.

Table 7.2 Average longshore transport for observed and simulated wave data. The approximate range between 5th and 95th percentiles calculated from the simulated data are also indicated.

| Observed (m^3/year) | Simulated (m^3/year) |
|--------------------------------|----------------------------------|
| 1 293 484 | 1 143 172 (1 050 000, 1 300 000) |

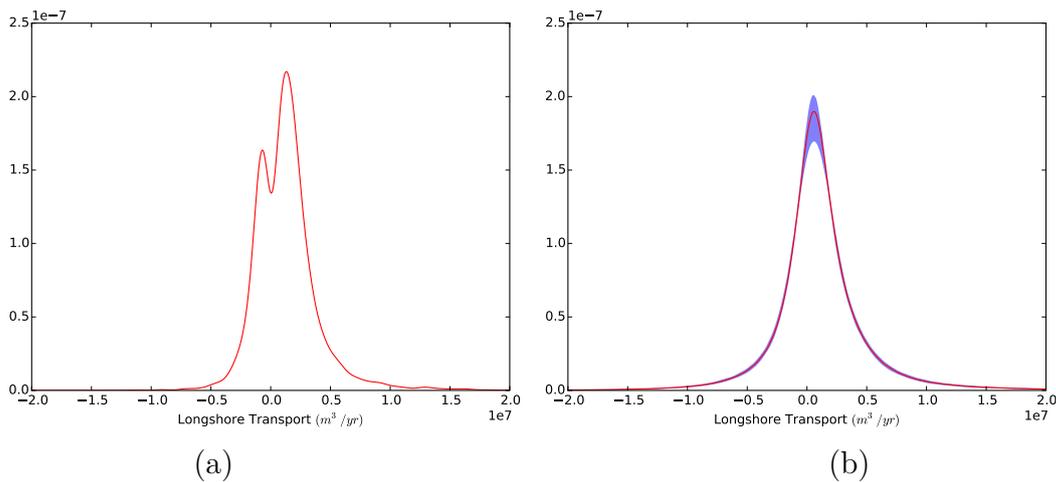


Fig. 7.6 Probability distributions for (a) the observed longshore transport rates and (b) simulated longshore transport rates. The median (red line) and the 95th% confidence limits (shaded) are also shown in the plot

7.3.3 Cross-shore Transport

Both cross-shore beach response models described in § 7.2.5 were applied to the synthetic wave dataset. The time convolution approach was used to estimate erosion return periods, whereas the shoreline evolution model was used to evaluate short and long term shoreline positions and the statistics thereof.

Erosion Return Periods

Fig. 7.7 shows the erosion return periods associated with both simulated and observed wave records. Independent events were delineated with a two week window as used by Corbella & Stretch (2012b). The erosion volumes estimated using the simulated waves are similar to the observed. This demonstrates two key results of the simulation technique. Firstly the simulated storms have similar durations to those of the observed sequence. Furthermore the simulated storms have comparable wave energy to the observed storms. The eroded volumes for return periods less than 40 years are significantly different from those estimated in Corbella & Stretch (2012a). However erosion volumes associated with return periods greater than 40 years are both similar in value and in the slope of the plots. The erosion volumes for associated return periods in Corbella & Stretch (2012a) were estimated using a process based model.

Fig. 7.7 does not show an upper limit as proposed in literature and has a similar shape to the results found in (Callaghan *et al.*, 2008; Ranasinghe *et al.*, 2013). The eroded volumes for different return periods are similar in magnitude to those estimated in Callaghan *et al.* (2008).

The results are tightly distributed for volumes associated with return period < 20 years. However the larger spread of values for return periods > 20 years is expected as the synthetic wave records are only 100 years long.

Shoreline Positions

The DEM was calibrated using beach profile data from a site situated south of Durban (refer Fig. 7.1). The site was chosen because data was not affected by the sand bypass scheme at the port entrance. Furthermore the site is situated on a relatively straight section of coastline. Therefore any divergence in longshore transport rates is negligible. The EBP parameters were estimated using Eq. 7.10 with the following assumptions:

- The critical angle of repose (ϕ) was assumed as 15° .

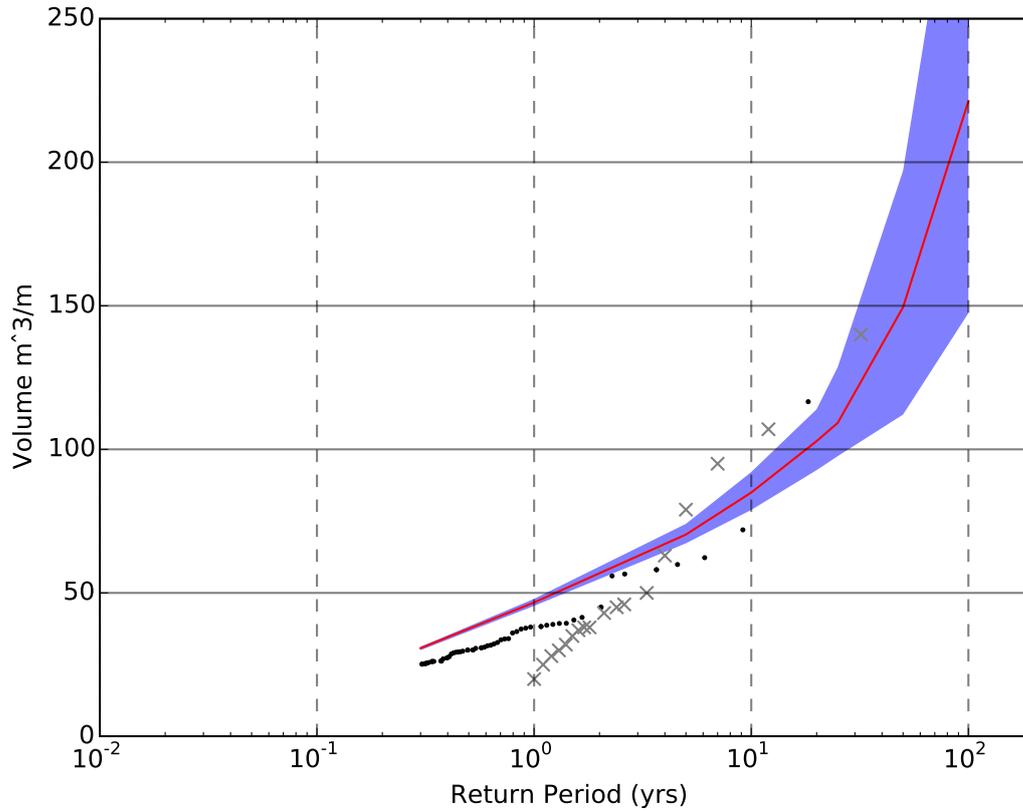


Fig. 7.7 The erosion return periods based on the convolution method after Kriebel & Dean (1993) showing the median (red line) and 80% confidence limits (shaded). Scatter points show erosion volumes and return periods for the observed dataset. Erosion volumes previously estimated by Corbella & Stretch (2012a) are shown by the grey crosses.

- The median grain size D_{50} was estimated as 0.25 mm.
- h_t was assumed equal to the closure depth of 15 m (Mather & Stretch, 2012). According to beach profile data $h_t = 15$ m occurs at a distance $x_t \approx 1200$ m from the reference datum.
- The shoreline position was chosen relative to the 0 MSL contour.
- The maximum shoreline position on record was assumed to be $S_{max} = 75$ m from which V_s was estimated.

The accretion/erosion rate parameters C^\pm were estimated using a simple least squares technique. Although Jara *et al.* (2015) set $C^+ = C^-$, for the present study we estimated $C^+ = 5 \times 10^{-7} \text{ s}^{-1}$ and $C^- = 5 \times 10^{-4} \text{ s}^{-1}$. The calibration suggests that beach

erosion occurs significantly faster than accretion i.e. erosion occurs at a time scale of approximately 3 hours whereas accretion takes approximately 4 months. Fig. 7.8 shows that for the observation period the DEM is able to capture the general trends of erosion and accretion, but does not accurately describe the extreme shoreline positions. This is attributed both to model limitations as well as missing data. Between 1993 – 2002 there are large periods of missing wave and shoreline data increasing the modelling difficulty. However the model performs significantly better between 2002 – 2009 and appears to capture the beach response to the largest wave event on record (March 2007) well.

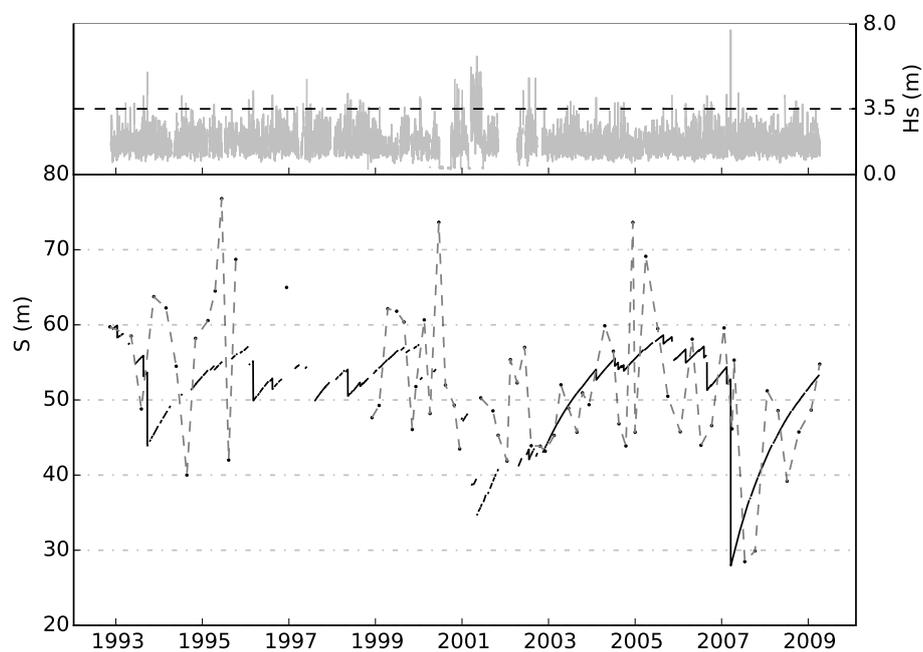


Fig. 7.8 The 0 MSL contour shoreline position for the beach profile used herein for the period 1993–2009. The observed shoreline positions are shown as scatter points connected with dashed lines. Modelled shoreline positions are shown with the solid black line. The wave height data for the period is also shown.

The DEM was applied to the synthetic wave dataset. Fig. 7.9 shows an example of a 100–year shoreline position simulation using the simulated waves. It is interesting to note the effect of storm clustering on shoreline position as well as the effect of single large storm events. For example there are a number of storm events that appear to cluster during the period 2070 – 2079 which result in similar erosion quantities as single events in 2016 and 2029.

Figs. 7.10(a) & (b) show the probability distribution function (PDF) and cumula-

Stochastic simulation of regional wave climates conditioned on synoptic scale meteorology. Part 2: Applications in coastal vulnerability assessment

tive distribution function (CDF) for the simulated shoreline positions with the 5th and 95th% confidence limits. The mean shoreline position for the 0 MSL contour is approximately 50 m from the reference datum and the distributions reveal a heavy lower tail. This is expected since the erosion rate parameter obtained from the calibration is larger than the accretion rate parameter. Coupled with the high energy wave climate this implies that the shoreline often does not fully recover to its equilibrium position between storm events.

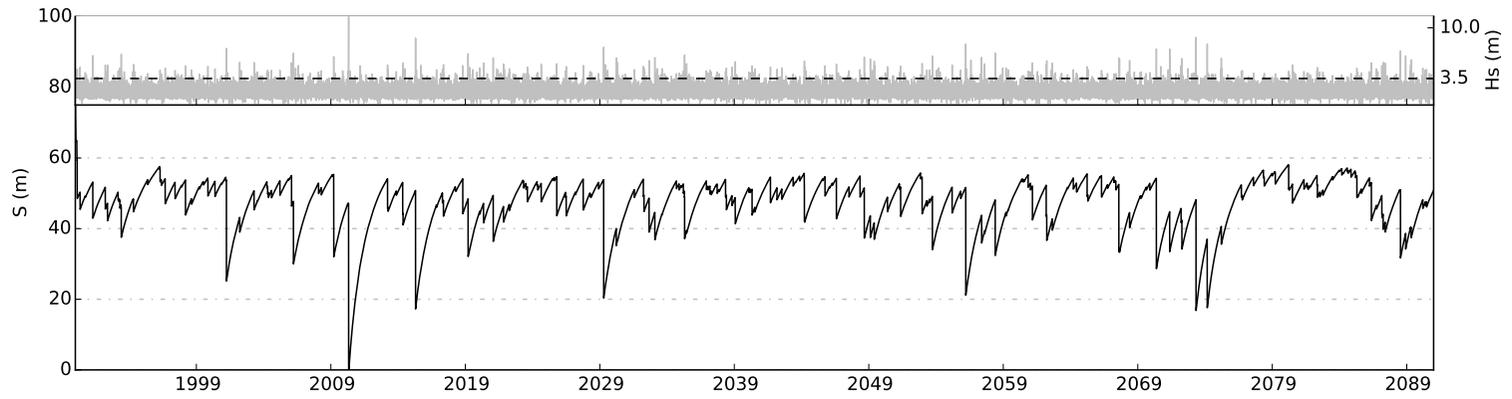


Fig. 7.9 Example of simulated 0 MSL contour shoreline positions using the synthetic wave dataset for the period 1990–2090. Wave height data are shown in the top panel

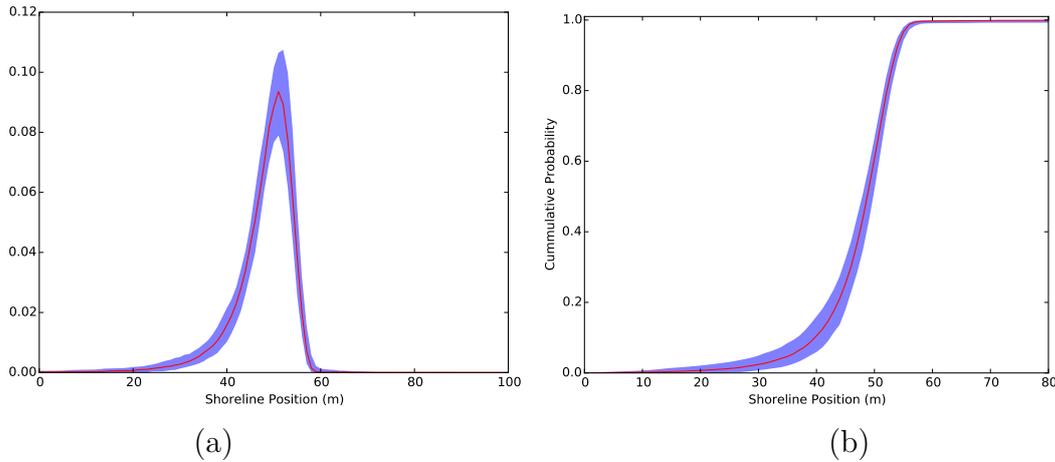


Fig. 7.10 Shoreline position PDF (a) and CDF (b) for the 0 MSL contour as calculated using the simulated waves. The median (red line) and the 5th and 95th% confidence limits (shaded) are also shown in the plot.

7.4 Discussion and Conclusion

This study attempts to demonstrate the performance of a new wave climate simulation technique for evaluating aspects of coastal vulnerability. The stochastic simulation technique was described by Pringle & Stretch (2015) and is linked to the physical drivers of regional wave climates. The synthetic wave dataset spans 100 years at a 6-hour temporal resolution. In general the model performs well and is reasonably able to accurately mimic results from observed wave data in the analysis of nearshore processes such as longshore and cross-shore sediment transport. Furthermore the advantage of simulating continuous wave sequences was demonstrated using the new shoreline evolution model developed by Jara *et al.* (2015).

The simulated wave time series provide a means by which to analyse both long and short term coastal risk assessment associated with longshore and cross-shore sediment transport. Return periods associated with simulated erosion volumes using the synthetic wave data agree well with those from observed wave data. The predicted change in erosion volumes with return period is qualitatively similar to results reported by Callaghan *et al.* (2008). The erosion volumes increase without limitation for increasing return periods which differs from the results of Corbella & Stretch (2012a) who suggested an upper limit. This appears to be a limitation of the Kriebel & Dean (1993) model for large erosion events since realistically an upper limit is expected. Recently

Callaghan *et al.* (2013) showed that semi-empirical and process based approaches can replicate this upper limit. However using those models with long-term synthetic wave time series may be challenging.

Seasonal statistics suggest that the wave model accurately describes the origin and direction of wave energy associated with different wave heights. The use of CPs to drive the model is beneficial since they have been shown to accurately delineate directional wave energy sources. Furthermore the CPs are capable of describing bimodal wave energy sources which is beneficial for modelling purposes.

The new dynamic shoreline evolution model developed by Jara *et al.* (2015) was used to demonstrate the advantage of simulating long continuous wave sequences. Shoreline positions can be simulated for long periods with relatively low computational effort, which is advantageous for coastal risk assessments. In contrast shoreline positions estimated in Yates *et al.* (2009) were based on wave data simulated using a spectral wave model, which is not well suited to simulating long sequences of shoreline changes because it is computationally demanding. A simple least squares technique was used to calibrate the shoreline model using observed wave data. The calibrated shoreline model successfully captured the general trends of observed shoreline positions.

A significant advantage of the stochastic simulation technique presented here is that it is well suited to studying future climate scenarios that include climate change effects. The links to atmospheric CPs should provide a robust framework for evaluating the effects of changes in synoptic scale meteorology obtained from global climate models (GCMs). Such changes have direct effects on surface variables of interest. For example CPs have been used to assess changes in the statistical properties of precipitation (e.g. Abiodun *et al.*, 2015; Bárdossy & Pegram, 2011) and recently on the European wave climate (e.g. Perez *et al.*, 2015). Therefore the stochastic simulation of waves based on GCM outputs for future climate scenarios provides a promising new methodology for quantifying coastal vulnerability within the context of climate change. This is because the effect of the changing climate on nearshore processes such as longshore and cross shore beach response can be easily quantified through the statistical links between CPs and waves.

Chapter 8

Atmospheric classification as a framework for assessing future coastal vulnerability.

Abstract

Recent developments in coastal vulnerability assessment have highlighted the benefits of identifying the synoptic scale atmospheric circulation patterns (CPs) that drive regional wave climates. A useful application of these CPs is to provide a framework to stochastically simulate waves that are conditioned on the weather systems that drive them. This has recently been exploited in the development of a hybrid CP-linked stochastic wave climate simulation technique. The present study is focussed on demonstrating two important properties of the CP classification and wave simulation techniques. Firstly we demonstrate that the outcome of our CP classification algorithm is robust to whether the wave data used is derived from models (with re-analysis) or from direct measurements. Secondly, we show how the new CP-linked wave simulation techniques can be applied to elucidate future wave climates associated with climate change scenarios. To illustrate these methods the future wave climate on the east coast of South Africa was simulated using CPs from HadGEM2-ES GCM simulations for the period 2010 – 2100. Some long-term changes in the occurrence statistics of extreme wave events are found for the case study location. More generally the results suggest broad applicability of the methodology for both current and future coastal vulnerability assessments.

8.1 Introduction

Within the context of climate change the ability to quantify risk in non-stationary environments is of fundamental importance. In coastal environments, coastal engineers and planners require reliable models that can accurately describe regional wave climates and associated risk factors. The application of multivariate statistical models is one method of addressing this issue, but in general they are not directly linked to the meteorological forcing of waves and can sometimes produce unrealistic results (Corbella & Stretch, 2013). Furthermore, non-stationary statistics can be complex and

difficult to incorporate into such models. Alternatively, process-based global wave models have become increasingly attractive due to improvements in their performance (Caires *et al.*, 2004; Chawla *et al.*, 2013; Mínguez *et al.*, 2011; Swail & Cox, 1999; Tolman *et al.*, 2002). The accuracy of these models depends largely upon the wind field inputs that drive them and they currently require re-analysis with data assimilation to produce accurate results. They are therefore unable to accurately predict future wave climate scenarios. Furthermore the ability of global wave models to accurately simulate extreme events depends strongly on the grid resolution. This is particularly important nearshore, an area that is typically poorly resolved by global wave models (Mínguez *et al.*, 2011). The prediction of climate change effects can therefore benefit from hybrid statistical models that retain links to the physical mechanisms that drive regional wave climates.

The use of meteorological features to drive coastal risk assessment models has recently attracted increasing attention in research. For example Zou *et al.* (2013) incorporated an ensemble modelling framework linking meteorological features to coastal flood risk through regional downscaling. Camus *et al.* (2014) proposed the use of weather pattern types as a statistical downscaling framework for regional wave climates. Espejo *et al.* (2014); Pringle *et al.* (2014) have recently shown how atmospheric classification can be a useful tool in coastal vulnerability assessment since it can provide a link between statistical risk assessment models and process-based global wave models. The transition between different atmospheric states or classes is a physically meaningful way to describe wave behaviour. This link was exploited in a recent stochastic wave simulation technique proposed in Pringle & Stretch (2015). Furthermore the use of atmospheric classification provides a useful framework within which to assess climate change effects (e.g. Perez *et al.*, 2015).

Following Bárdossy *et al.* (1995) the atmospheric classification algorithm used herein is based on fuzzy rules with intrinsic links to surface variables, in this case wave characteristics. Pringle *et al.* (2014) were the first to apply this atmospheric classification technique to elucidate the drivers of regional wave climates. The classes are automatically derived in a supervised manner based on their ability to explain wave characteristics such as wave heights, directions, periods and storm durations. Pringle *et al.* (2014) focussed on using the wave height as the variable of interest to classify circulation patterns (CPs).

Wave climate estimation using the latest generation of spectral wave models has become increasingly accurate. While it may never replace the reliability of direct

Atmospheric classification as a framework for assessing future coastal vulnerability.

wave observations it is particularly useful where direct measurements are not available. Furthermore it provides an alternate dataset to use in an atmospheric classification algorithm. Therefore CP classification using modelled wave data has the potential to describe the drivers of wave climates in regions where no wave observations are available. This approach is also well suited to analysing and quantifying future coastal vulnerability concerns.

The aim of this study is to investigate the robustness of our automated fuzzy rule based classification in locating the drivers of regional wave climates using only modelled wave data. If the algorithm can locate the same CPs as those derived from direct wave measurements then this robustness can be exploited in coastal vulnerability at locations without wave measurements. We also demonstrate an application of a CP-linked stochastically wave simulation model for evaluating future wave characteristics under climate change scenarios. This is based on a supposition that changes in regional CP occurrence frequencies predicted by global climate models (GCMs) will in turn drive significant changes in the wave climates in those regions.

8.2 Method

8.2.1 Case Study Site

The KwaZulu-Natal wave climate on the east coast of South Africa, shown in Fig 8.1, has been studied in some detail (Corbella & Stretch, 2012a,b,c,d; Mather & Stretch, 2012). However, it is only recently that Pringle *et al.* (2014) have focussed on the physical drivers of the wave climate, namely the atmospheric circulation patterns associated with major wave events. They noted that stationery low pressure systems south or east of the country drive large waves towards the coastline. Specific weather systems associated with extreme events have been cited as tropical cyclones (TC's), mid-latitude cyclones and cut-off lows (Corbella & Stretch, 2012d; Mather & Stretch, 2012; Pringle *et al.*, 2014; Rossouw *et al.*, 2011; Schumann, 1988). Deep low pressure systems east of the country coupled with a ridging high pressure region behind a frontal low, can cause strong winds to blow towards the coastline with associated extreme wave conditions.

Corbella & Stretch (2013) defined extreme wave events as those where the significant wave height $H_s \geq 3.5\text{m}$ since the 3.5 m threshold was found to delineate events that cause significant coastal erosion. The austral autumn and winter months

are typically associated with the occurrence of the largest wave heights in the region (Corbella & Stretch, 2012d). Seasons are defined in Table 8.1.

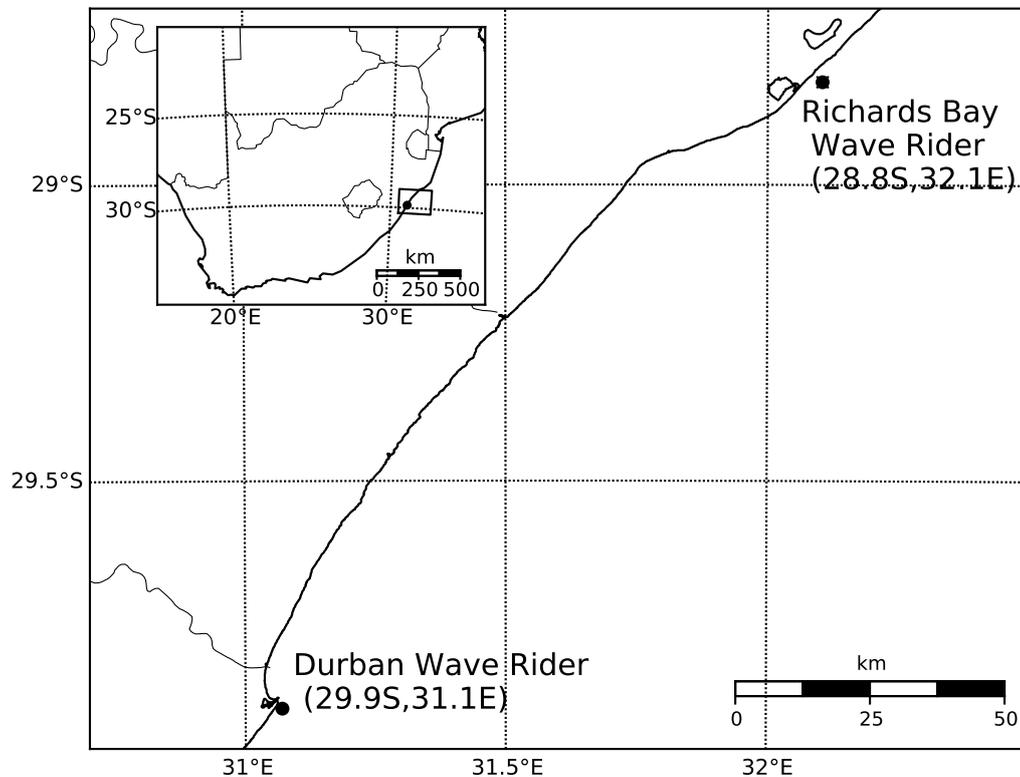


Fig. 8.1 Locations of the wave observation buoys at Durban and Richards Bay, along the KwaZulu Natal coastline (Pringle *et al.*, 2014).

Table 8.1 The allocation of months to seasons

| Season | Months |
|--------|----------------------|
| Summer | December – February |
| Autumn | March – May |
| Winter | June – August |
| Spring | September – November |

8.2.2 Data Sources

Observed and Modeled Wave Data

Observed wave data was obtained from the Durban wave-rider buoy (refer Fig. 8.1) for the period 1992 – 2009. Where necessary missing data was supplemented by

Atmospheric classification as a framework for assessing future coastal vulnerability.

data from the Richards Bay wave-rider buoy (Fig 8.1) to provide a continuous wave record. Corbella & Stretch (2012b) showed a strong relationship between wave heights measured by the two buoys.

Re-analysis swell and significant wave height data were obtained from the ERA-Interim dataset¹. The ERA wave data is based on the third generation WAM model ((Dee *et al.*, 2011; Komen *et al.*, 1994) and uses data assimilation to improve the results. Wave data were extracted for the location 30°S - 31°E (Fig 8.1).

Atmospheric Pressure Data

The classification of atmospheric states is based on pressure anomalies on the 700 hPa geo-potential. Re-analysed atmospheric pressure data were obtained from the ERA-Interim dataset for the area 10°S, 0°E – 50°S, 50°E with a grid resolution of 2.5° and for the period 1972 – 2009.

Pressure fields on the 700 hPa geo-potential for the same region were also obtained from the HadGEM2-ES implementation of the CMIP5 centennial simulations for the period 2010 – 2100 (Collins *et al.*, 2011; Jones *et al.*, 2011). Analysis of the pressure data included two future scenarios or “representative climate pathways” (RCPs): a low emission scenario (RCP2.6) and a high emission scenario (RCP8.5) (Moss *et al.*, 2010).

The CP anomalies are defined as perturbations on a specific pressure level, namely

$$a(x, t) = \frac{p(x, t) - \bar{p}(x, j(t))}{\sigma(x, j(t))} \quad (8.1)$$

where $a(x, t)$ is the anomaly value at location x for time t , $\bar{p}(x, j(t))$ and $\sigma(x, j(t))$ are the mean and standard deviation pressure level at location x for Julian day $j(t)$. The inclusion of the function $j(t)$ is to provide a smooth transition between CP states for different days.

8.2.3 Classifying Atmospheric Circulation

Classification of atmospheric circulation refers here to a computer assisted process whereby (a) states (or classes) with similar properties are *identified*, and (b) *grouped* together. The classification technique used herein differs from most others used in climate studies because it is guided by a surface variable of interest. Wave heights

¹<http://apps.ecmwf.int/datasets/>

are incorporated into an optimisation procedure to identify classes that are the main sources of specific wave characteristics. An outcome of this is that the classes have intrinsic links to wave behaviour. The methodology is based on fuzzy logic and was developed by Bárdossy (2010); Bárdossy *et al.* (1995, 2002). Pringle *et al.* (2014) adapted the algorithm to delineate sources of wave behaviour. Bárdossy *et al.* (2015) demonstrate the advantage of including a variable of interest in the classification by showing that the educed set of CP classes associated with different variables of interest can be distinct.

The goal of the classification is to optimally delineate the physical drivers of regional wave climates. Therefore an optimisation procedure is used that is based on the wave height with the assumption that similar CPs drive similar wave events. Two objective functions are used. The first objective function is designed to identify CP classes associated with large wave events. These CP classes are associated with higher frequencies of exceeding specified wave thresholds than the unconditional exceedance frequency. Whence

$$O_1(\theta) = \sqrt{\sum_{t=1}^T (h(H_s \geq \theta \mid CP(t)) - \bar{h})^2} \quad (8.2)$$

where θ is a prescribed wave height threshold for the time period T , $CP(t)$ is the class assigned to time t and \bar{h} is the unclassified mean frequency of an event. The algorithm considered two wave height thresholds after Pringle *et al.* (2014): (a) $\theta \geq 3.5$ m, wave heights greater than this are associated with severe coastal erosion (Corbella & Stretch, 2013); (b) $\theta \geq 2.5$ m, a threshold associated with midrange wave heights.

The second objective function aims to identify CPs with mean wave height statistics that are different from the unclassified mean. The ratio between the mean wave heights for CP classes and the unclassified mean wave height is incorporated as follows:

$$O_2 = \sum_{t=1}^T \left| \frac{H_s(CP(t))}{(\overline{H_s})} - 1 \right| \quad (8.3)$$

where $(\overline{H_s})$ is the unclassified average significant wave height and $H_s(CP(t))$ is the average significant wave height for the given CP class assigned for time t . A weighted linear combination of Equations 8.2 and 8.3 are used to find an optimal solution to the classification. The weights reflect their relative importance and account for the different average magnitudes of the objective functions. A simulated annealing

Atmospheric classification as a framework for assessing future coastal vulnerability.

algorithm was used in the optimisation process.

8.2.4 Classification Similarity

The classification technique outlined in § 8.2.3 was applied to the ERA-Interim pressures (at the 700 hPa geopotential) using two different wave datasets: (a) direct wave measurements and (b) re-analysed modelled waves. That is, from the same pressure dataset two different sets of CP classes were derived using the measured and reanalysed wave data respectively. The relationship between the two classifications was evaluated using a contingency table. The classifications have a non-numerical basis therefore methods such as regression analysis cannot be used to quantify their similarity (Stehlik & Bárdossy, 2003). The contingency table is a useful measure of the association between the two classifications. For example two classifications that are exactly similar possess classes with the same spatial configurations. Therefore each class in the first classification will occur with a matching class in the second classification. Since the classification technique is stochastic in nature, the class labels assigned to each set of CPs are arbitrary. More importantly the spatial orientation of pressures within the classes derived using reanalysed waves may differ from those obtained from wave measurements. The differences can arise from the optimisation process itself or from differences between the two wave datasets. The contingency table provides a simple means to identify similar classes from each classification.

Two useful statistics can be calculated from the contingency table. They are based on a χ^2 statistic defined as (Hartung *et al.*, 1999)

$$\chi^2 = \sum_{i=1}^I \sum_{j=1}^J \frac{\left(n_{ij} - \frac{n_i \cdot n_j}{n}\right)^2}{\frac{n_i \cdot n_j}{n}} \quad (8.4)$$

where I and J are the number of classes in the first and second classification respectively. I and J also reflect the number of rows and column within the contingency table. n_{ij} is the number of simultaneous occurrences of class (row) i with class (column) j , n_i is the total number of simultaneous occurrences of class (row) i with all the j classes (columns). Similarly n_j is the total number of simultaneous occurrences of class (column) j with all i classes (rows) and n is the total sum of all the occurrences within the contingency table. The term $\frac{n_i \cdot n_j}{n}$ represents the expected number of simultaneous occurrences of class (row) i with class (column) j .

The dependence between the two classifications can be measured by the modified

Pearsons coefficient C and the Cramer coefficient V . Both measures are bounded by 0 and 1 and the stronger the association between classifications the larger the values. The modified Pearsons coefficient is defined as

$$C = \sqrt{\frac{\min(I, J)}{\min(I, J) - 1} \sqrt{\frac{\chi^2}{\chi^2 + n}}} \quad (8.5)$$

whereas the Cramer coefficient is defined as

$$V = \sqrt{\frac{\chi^2}{n(\min(I, J) - 1)}} \quad (8.6)$$

It is expected that for our case study the two classifications should be strongly associated since the CPs that drive the wave events are present in both.

8.2.5 Wave Simulation

Only a brief overview of our CP-based method for the stochastic simulation of waves is presented here since a detailed description is given by Pringle & Stretch (2015). The wave simulation technique exploits the strong links between CPs and wave characteristics. These links are also used for the classification method described in § 8.2.3.

The occurrence of different CPs and their transitions between states provide a physically meaningful way to simulate waves because each CP is associated with wave climate variables that have specific interdependence structures. Pringle & Stretch (2015) showed that the CP sequences can be described as a Markov process. Therefore given a CP at some time t_i and wave variables (H_s, T_p, θ) it is possible to simulate the CP and associated wave variables at time t_{i+1} . Bivariate copulas conditioned on the CPs were used to model the dependence structure between significant wave heights and periods $H_s|T_p$ and between significant wave heights at different times $H_s(t)|H_s(t-1)$. Wave directions were simulated from their empirical distributions conditioned on the wave heights and CP occurrences.

The classification method was applied to future pressure data from HadGEM2-ES GCM simulations for both high and low emission scenarios. Changes in CP occurrences were estimated using the classes derived from the ERA-Interim pressure dataset. The new CP occurrence statistics for the future scenarios were then used to simulate future wave climates based on the CP – wave dependence structure derived from directly measured wave data.

8.2.6 Evaluating Changes in Wave climate

The stochastic wave simulation technique was applied using the CP statistics obtained for the near (2010 – 2050) and distant (2050 – 2100) futures in the RCP2.6 and RCP8.5 climate scenarios. The statistical properties of the simulated waves were used to evaluate changes in wave behaviour for the two future scenarios. This is important because changes in wave behaviour have direct implications for coastal vulnerability assessments. The wave statistics analysed for changes included the seasonally averaged wave heights, directions and periods, wave height return periods, and seasonal wave roses. The return period is defined as (Goda, 2008; Salvadori, 2004)

$$T_R = \frac{\mu_t}{1 - p} \quad (8.7)$$

where μ_t is the average storm inter-arrival time and p is the probability level.

8.3 Results

8.3.1 Measured and reanalysed wave heights

The measured and reanalysed wave data are well correlated for average wave heights. Both the Pearson and Spearman's cross-correlation coefficients are about 0.7 at zero time lag (Fig 8.2(a)). In contrast Fig 8.2(b),(c) and Fig 8.3, 8.4 and 8.5 show the inability of the ERA wave model to accurately reproduce extreme wave events. The model is able to capture general trends such as correct timing of the event and shape but underestimates the peak wave heights. The peak correlation coefficient is at a zero time lag (see Fig 8.2(a)), which suggests the timing is correct. Swail & Cox (1999) found similar trends when comparing ERA and NCEP (National Centre for Atmospheric Research) wave height data. They attributed such differences to the model's inability to sufficiently resolve compact wind fields such as in tropical cyclones and mid-latitude cyclones.

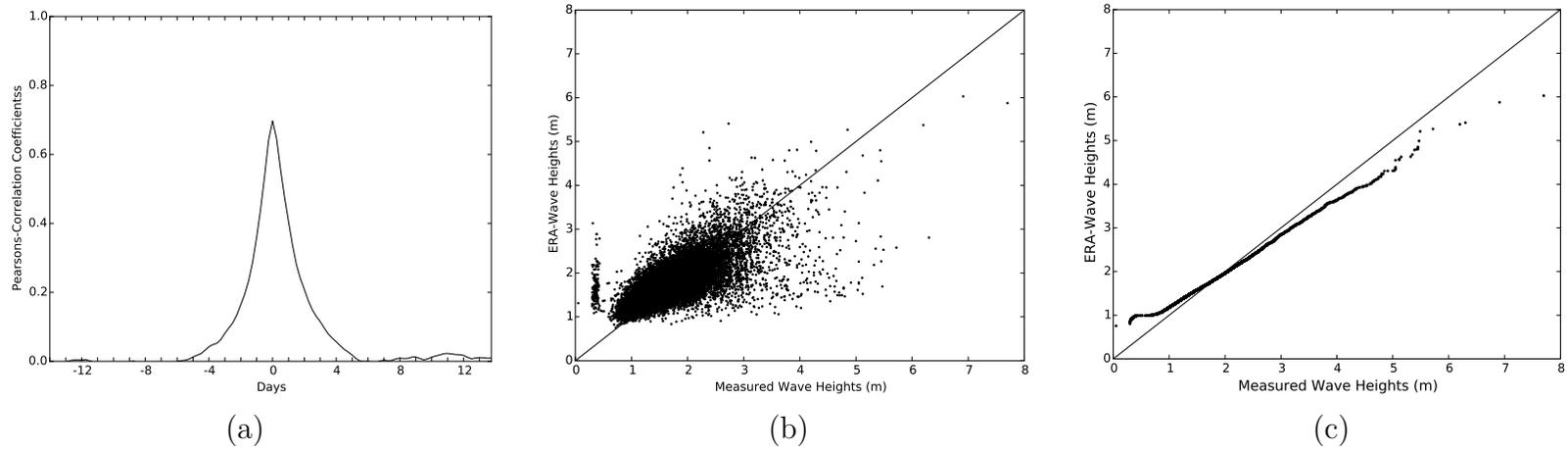


Fig. 8.2 The relationship between the measured and ERA-reanalysed wave data are shown in (a) the lagged Pearson's correlation coefficients, (b) scatter plot of wave heights and (c) Quantile-Quantile plot .

Atmospheric classification as a framework for assessing future coastal vulnerability.

Figs. 8.3, 8.4 and 8.5 show time sequences of three extreme wave events off the KwaZulu-Natal coastline. The sequences are driven by different meteorological features. Fig. 8.3 was an event driven by a strong cut-off low system that became stationary east of the coastline. It is the largest wave event on record for the KwaZulu-Natal coastline and caused severe coastal erosion. Mid-latitude cyclones as shown in Fig. 8.4 are common weather systems for the South African climate. This particular mid-latitude cyclone was responsible for driving an extreme wave event along the KwaZulu-Natal coastline. Fig. 8.4 shows the CP two days before the occurrence of the peak wave height of the extreme event. This is approximately the time it takes for deep water waves to reach the coastline from such a distance. This is based on a estimated group velocity for waves with a period of 10 seconds. Tropical cyclones are sometimes cited as major drivers of extreme wave events along the KwaZulu-Natal coastline (Corbella & Stretch, 2012d; Mather & Stretch, 2012). However they are only capable of driving large waves towards the coastline when they occur at specific locations, for example within the Mozambican channel or south of Madagascar. An example of this is shown in Fig. 8.5. Both the occurrence of a cut off low (south east of the coastline) and a tropical cyclone are shown to drive large waves towards the coastline. Their tracks are shown as red and black respectively.

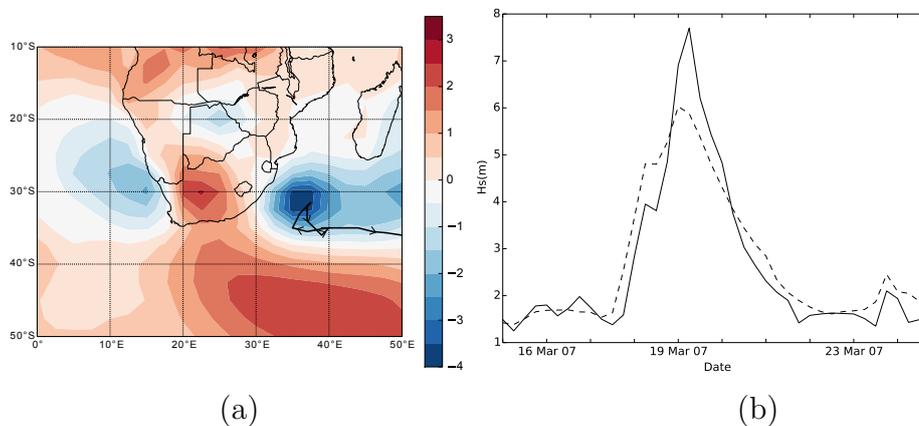


Fig. 8.3 The March 2007 storm event showing (a) the storm track with corresponding CP at the peak of the storm event. (b) The observed (solid line) and ERA modeled (dashed line) wave heights at the Durban waverider buoy for March 2007.

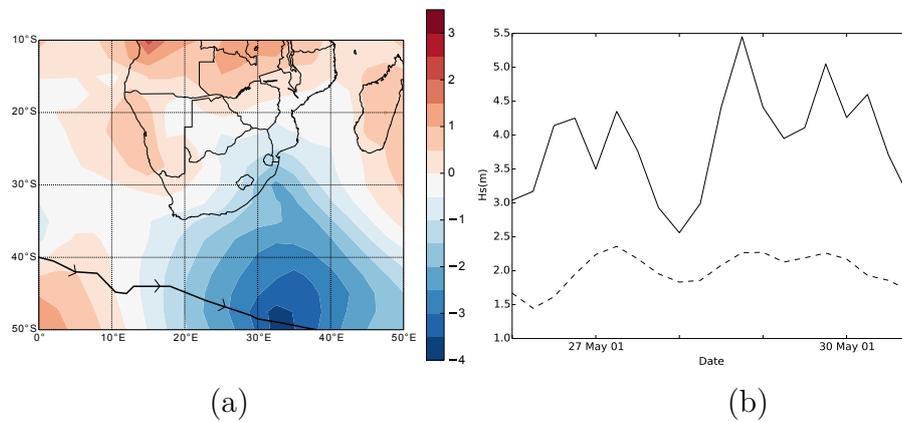


Fig. 8.4 (a) A typical mid-latitude cyclone with associated track and CP two days before the storm peak. (b) The observed (solid line) and modeled (dashed line) wave heights at the Durban waverider buoy for May 2001.

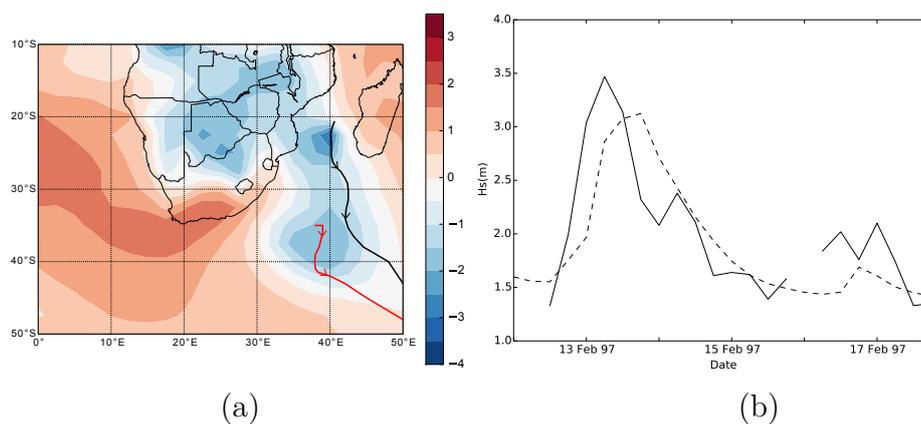


Fig. 8.5 (a) The tracks of a tropical cyclone (black) and a cut-off low (red) and CP at time of the storm peak. (b) The observed (solid line) and ERA modeled (dashed line) wave heights at the Durban waverider buoy for Feb 1997.

8.3.2 Classification Comparison

Two classification procedures were applied to the CP anomalies on the 700 hPa geopotential. Classes were first derived using measured wave data to form the objective functions that guided the classification (§8.2.3). The classification was then re-run using reanalysis wave data. Sixteen CP classes were specified in both cases, which according to Pringle & Stretch (2015) is optimal (in the sense of minimum entropy) and gives a classification outcome that provides strong insight into the physical drivers of the wave climate.

Contingency

The contingency table (Table 8.2) provides a means to statistically compare the two classification outcomes. The similarities between the classifications is evident in the table. For each class derived using reanalysis wave data there is a corresponding class derived from the measured wave dataset that occurs in conjunction with it a large percentage of the time. That is, for each row in Table 8.2 there is a matching column in which the occurrence frequencies are high. The modified Pearson's correlation coefficient and Cramer's coefficient are 0.77 and 0.28 respectively. Although the Cramer's coefficient value seems low, these values are similar to what Stehlik & Bárdossy (2003) found when comparing European classification schemes.

Table 8.2 Contingency table of the two classifications. The rows contain CP classes derived from the observed wave climate. The columns are classes derived from reanalysed wave data. Values in the table are normalised by the number of cases in each column (%)

| CP : | CP01 | CP02 | CP03 | CP04 | CP05 | CP06 | CP07 | CP08 | CP09 | CP10 | CP11 | CP12 | CP13 | CP14 | CP15 | CP16 | CP99 |
|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|
| CP01 | 11 | 38 | 20 | 0 | 36 | 7 | 12 | 0 | 6 | 6 | 13 | 1 | 9 | 4 | 8 | 1 | 6 |
| CP02 | 9 | 11 | 1 | 0 | 25 | 4 | 26 | 1 | 27 | 7 | 9 | 2 | 0 | 4 | 4 | 2 | 7 |
| CP03 | 1 | 4 | 4 | 1 | 12 | 9 | 14 | 1 | 4 | 3 | 8 | 10 | 3 | 6 | 6 | 1 | 11 |
| CP04 | 2 | 5 | 11 | 3 | 7 | 0 | 6 | 6 | 1 | 2 | 8 | 7 | 1 | 3 | 5 | 29 | 1 |
| CP05 | 2 | 2 | 0 | 48 | 1 | 1 | 2 | 15 | 3 | 9 | 0 | 5 | 7 | 4 | 3 | 1 | 0 |
| CP06 | 5 | 4 | 3 | 1 | 5 | 5 | 7 | 0 | 1 | 10 | 4 | 5 | 5 | 0 | 8 | 1 | 12 |
| CP07 | 1 | 2 | 12 | 4 | 2 | 6 | 5 | 14 | 10 | 4 | 9 | 18 | 6 | 3 | 1 | 13 | 2 |
| CP08 | 35 | 12 | 0 | 4 | 3 | 1 | 2 | 4 | 2 | 5 | 7 | 1 | 1 | 3 | 3 | 0 | 1 |
| CP09 | 5 | 1 | 0 | 2 | 2 | 1 | 3 | 34 | 8 | 4 | 0 | 2 | 0 | 14 | 3 | 8 | 0 |
| CP10 | 20 | 7 | 2 | 13 | 3 | 6 | 8 | 4 | 4 | 21 | 4 | 3 | 3 | 30 | 12 | 3 | 2 |
| CP11 | 1 | 0 | 1 | 6 | 0 | 15 | 0 | 9 | 11 | 7 | 1 | 12 | 15 | 9 | 5 | 11 | 0 |
| CP12 | 0 | 0 | 17 | 0 | 0 | 2 | 1 | 1 | 3 | 4 | 4 | 11 | 16 | 1 | 18 | 7 | 1 |
| CP13 | 0 | 1 | 9 | 1 | 2 | 10 | 6 | 6 | 14 | 3 | 1 | 7 | 17 | 8 | 0 | 6 | 7 |
| CP14 | 0 | 5 | 2 | 1 | 1 | 20 | 1 | 1 | 3 | 1 | 5 | 4 | 4 | 8 | 9 | 1 | 28 |
| CP15 | 6 | 3 | 11 | 16 | 1 | 0 | 4 | 4 | 0 | 9 | 21 | 4 | 0 | 0 | 6 | 11 | 0 |
| CP16 | 1 | 2 | 9 | 1 | 0 | 10 | 3 | 0 | 2 | 3 | 5 | 6 | 12 | 2 | 8 | 5 | 11 |
| CP99 | 0 | 0 | 0 | 0 | 0 | 2 | 0 | 0 | 1 | 2 | 0 | 2 | 1 | 0 | 0 | 0 | 11 |

^aCP99 is the unclassified class.

Atmospheric classification as a framework for assessing future coastal vulnerability.

Spatial Similarity

Figs. 8.6, 8.7 and 8.8 show CP class comparisons for the two classifications. Fig. 8.6 refers to the pair of classes that according to Table 8.2 have the strongest association. Both classes have a low pressure region to the south of the country that is bounded by a northerly located high pressure region. This pattern is similar to a mid-latitude cyclone in an early stage of its development. According to the classifications the classes occur approximately 12% and 10% of the time in the measured and ERA-modeled wave scenarios respectively.

Extreme wave events predicted by the ERA wave model are driven predominantly by CPs shown in Figs. 8.7(a) and 8.8(a). According to Table 8.3 these two classes contribute about 54% and 16% to extreme events respectively. The classes appear very similar in appearance to their counterpart classes derived using observed wave data. However according to the wave statistics in Table 8.3 the classes contribute significantly less to extreme wave events. This may be attributed to the fact that classifications based on observed wave data are able to capture significantly more detail than those based on simulated (with reanalysis) wave data.

Table 8.3 Key CP statistics using the ERA reanalysis wave data to delineate the classes for the period 1992–2009. Comparisons with classifications using observed wave data to delineate classes are shown in brackets.

| CLASS | Frequency | Threshold exceedance given CP | Exceedance contribution |
|-------------|-----------|-------------------------------|-------------------------|
| | $p(CP)$ % | $p(\theta CP)$ % | $p(CP \theta)$ % |
| CP05 (CP01) | 10 (12) | – (0.48) | 0.91 (5) |
| CP08 (CP09) | 5.4 (4.8) | 8.6 (3.2) | 54 (11) |
| CP16 (CP04) | 4.8 (5.6) | 3.0 (2.1) | 16 (9.2) |

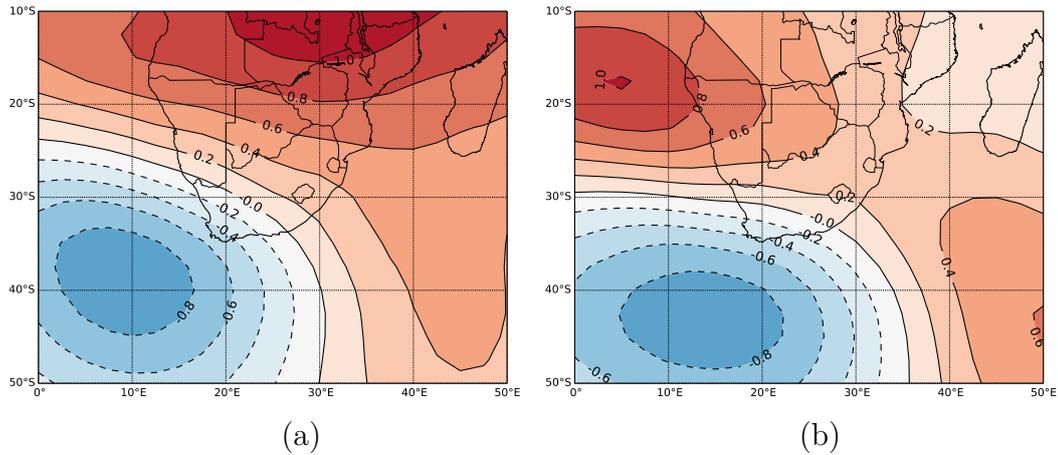


Fig. 8.6 Classes derived using observed wave heights (a) and modelled wave heights (b). According to Table 8.2 these classes have the highest degree of association.

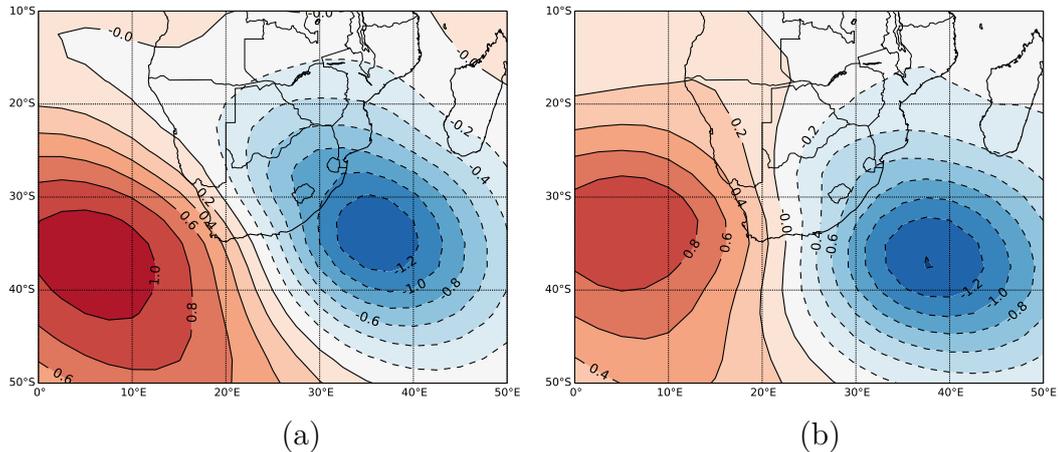


Fig. 8.7 (a) The class with the largest ($\sim 54\%$) contribution to extreme wave events according to modelled data (a) and (b) its corresponding class with strongest association according to Table 8.2.

8.3.3 Future Wave Climate

The CP classification methodology was applied to at the 700 hPa geopotentials obtained from the HadGEM2-ES GCM for the period 2010-2100. This was done to demonstrate the possibility of utilising CPs to evaluate future wave climate scenarios. The sixteen CP classes used for this application were still those derived from the pressure anomalies obtained from the ERA-Interim dataset in conjunction with direct wave measurements (see § 8.3.2). However, CP statistics for the near future (2010-2050) and distant future (2050-2100) GCM simulations were used to simulate waves and evaluate changes for the future wave climate.

Atmospheric classification as a framework for assessing future coastal vulnerability.

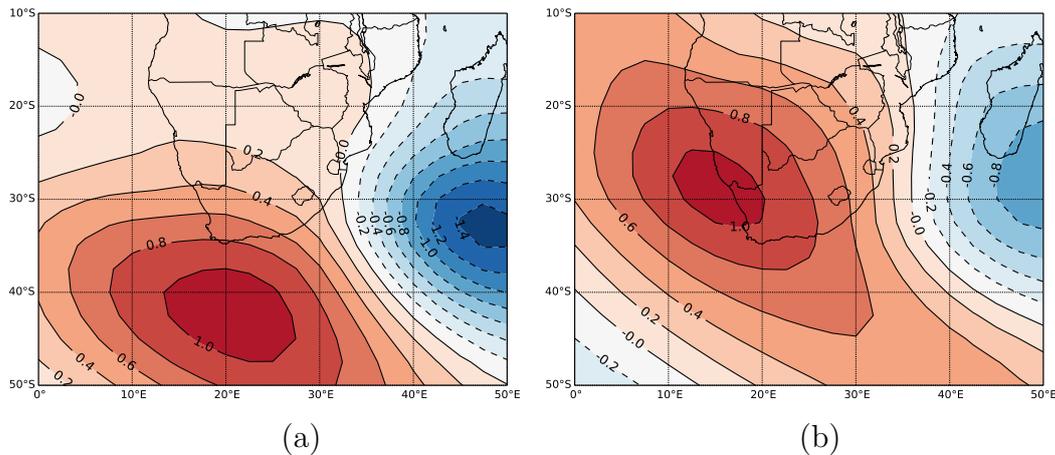


Fig. 8.8 (a) The CP class with the second largest ($\sim 16\%$) contribution to extreme wave events according to modelled data and (b) its corresponding class with strongest association according to Table 8.2.

Future CP Statistics

The maximum and minimum differences in CP occurrence frequencies for two future climate scenarios (RCP8.5 and RCP2.6) between the near and distant future are shown in Table 8.4. The long term changes are calculated as the differences in statistics between the distant future (2050-2100) and near future (2010-2050) as estimated by the HadGEM2-ES GCM. Also shown is the comparison between historical pressures predicted by the HadGEM2-ES GCM and reanalysed pressures for the period 1979-2009. The results in Table 8.4 indicate that there are only small changes in CP occurrences in the distant future for both the high and low emission scenarios. Changes in CP statistics for future climate scenarios will directly affect future wave behaviour. However since Table 8.4 indicates small changes in CP occurrences, it is expected that there will be small changes to wave behaviour. However in terms of longshore sediment transport rates Q even a small change in wave height can have a significant effect since $Q \propto H_s^m$ with $m > 1$ (CERC, 1984b; Kamphuis, 1991). It is also important to evaluate changes in wave direction because a previously sheltered stretch of coast may become vulnerable to wave attack.

Changes in Wave Behaviour

Synthetic wave datasets were simulated based on CP statistics for the near future (2010 – 2050) and distant future (2050 – 2100) and for the two climate scenarios RCP2.6 and RCP8.5. The two periods were chosen to evaluate changes in wave be-

Table 8.4 Maximum and minimum estimated differences in CP occurrence frequencies between historical and HadGEM2-ES GCM predictions. The table also shows the differences between the near future and distant future as predicted by the HadGEM2-ES GCM for two scenarios. The long term changes are calculated as the difference between CP occurrence frequencies in near future (2010-2050) and distant future (2050-2100).

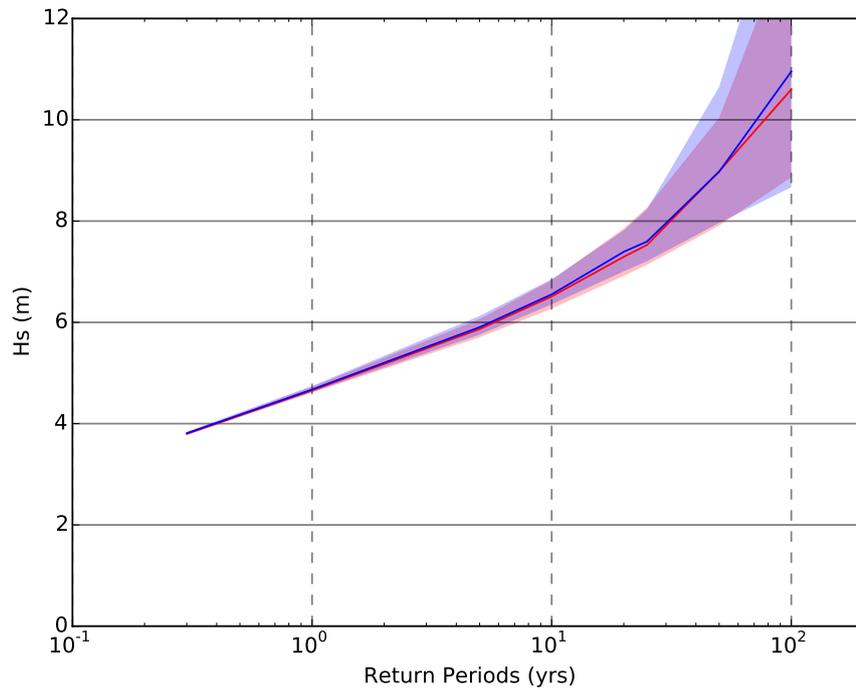
| Statistics: | Min | Max |
|--|-----|-----|
| <i>Historical Comparison in CP Frequency %</i> | | |
| Summer | -2 | 4 |
| Autumn | -2 | 1 |
| Winter | -2 | 2 |
| Spring | -2 | 4 |
| <i>Long Term Changes RCP8.5 %</i> | | |
| Summer | -3 | 3 |
| Autumn | -2 | 4 |
| Winter | -2 | 2 |
| Spring | -3 | 2 |
| <i>Long Term Changes RCP2.6 %</i> | | |
| Summer | -2 | 2 |
| Autumn | -1 | 1 |
| Winter | -2 | 2 |
| Spring | -2 | 2 |

haviour between the near and distant future. The synthetic waves were simulated for 100 years using CP statistics for each period. The length of the datasets were chosen so as to evaluate wave heights associated with the 100 year event. This process was repeated 101 times to obtain a suitable sample size.

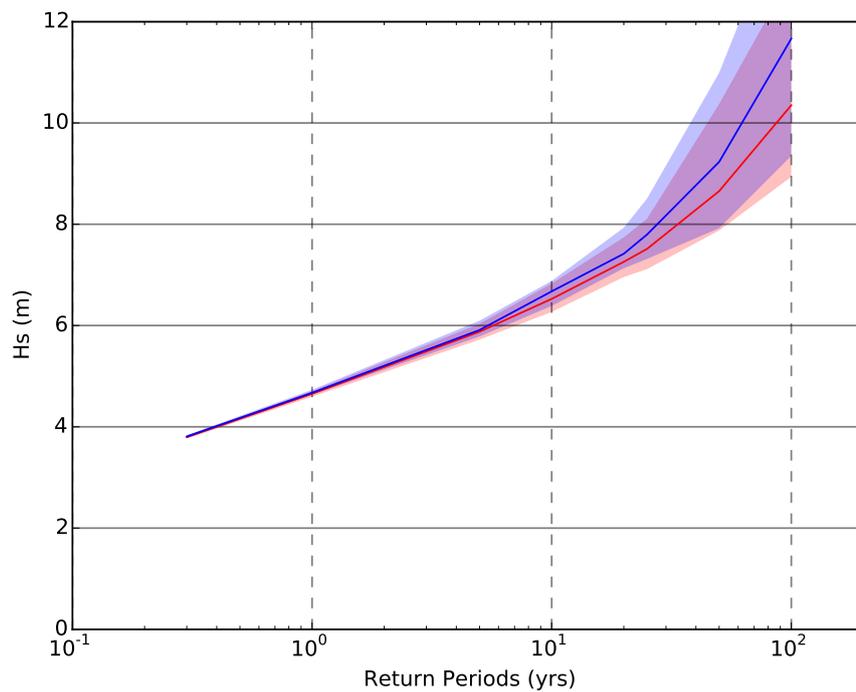
Fig. 8.9 shows the return periods and associated wave heights for the two future scenarios and between the two periods. According to the HadGEM2-ES simulation there are no significant changes in wave height extrema between the near and distant future for the RCP2.6 emission scenario. However for the RCP8.5 emission scenario, wave heights associated with return periods greater than 5 years will be larger in the distant future. This has direct implications for coastal vulnerability.

In general according to the simulations based on the output from the HadGEM2-ES GCM there will be no significant changes in wave direction along the east coast of South Africa. Fig. 8.10 shows the wave roses derived from the simulated wave data for the RCP8.5 emission scenario in the near and distant future. They suggest that there will be a decrease in wave energy from the east-southeast sector while wave energy from the south-southeast sector will increase. However these changes are small (in the

Atmospheric classification as a framework for assessing future coastal vulnerability.



(a)



(b)

Fig. 8.9 Return period statistics and associated wave heights for two different future scenarios: (a) RCP2.6 and (b) RCP8.5. Wave heights and return periods for the distant future (2050-2100) are shown in blue whereas those in the near future (2010-2050) are shown in red. Shaded regions indicate the 80% confidence limits.

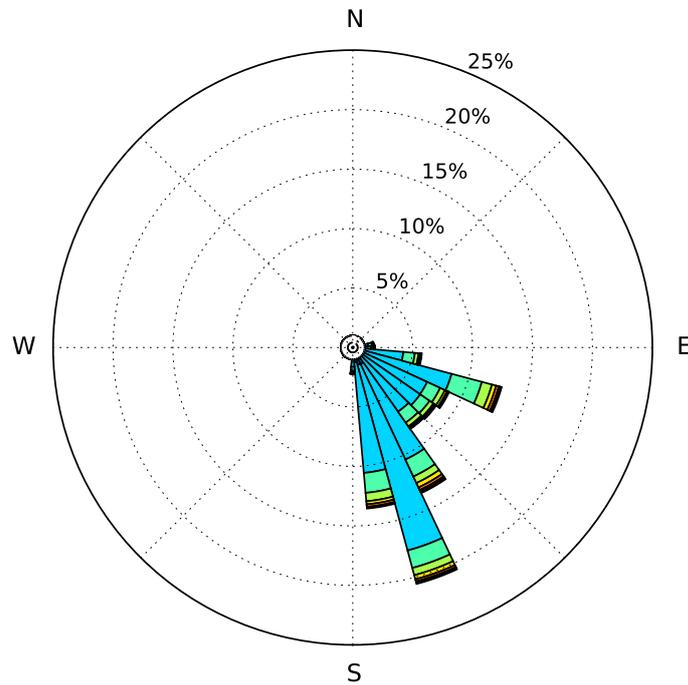
order of 1 or 2%).

Table 8.5 shows no significant change in the seasonal statistics of wave climate variables for both future climate scenarios between the near and distant future.

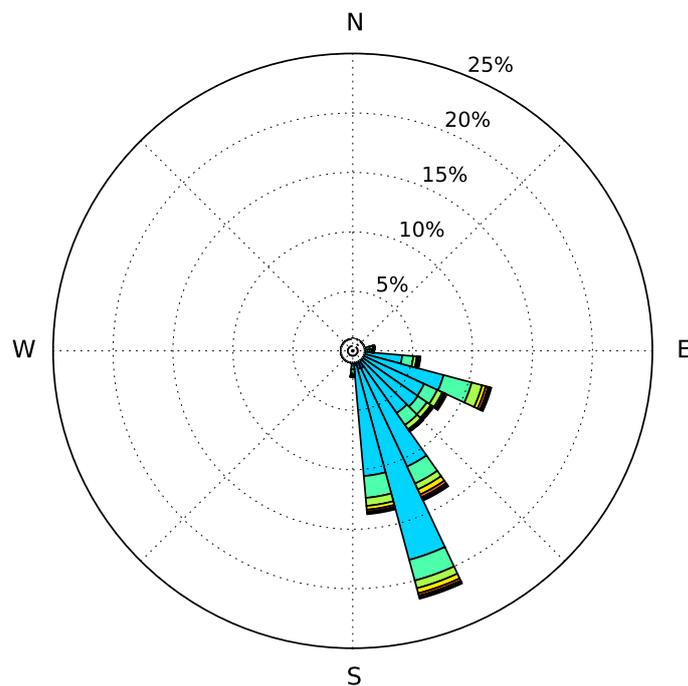
Table 8.5 Average statistics for important wave variables (H_s, T_p, θ) obtained from the stochastic wave simulation for the near and distant future based on two future scenarios.

| Statistics: | Summer | Autumn | Winter | Spring |
|--|--------|--------|--------|--------|
| <i>RCP2.6 Near Future (2010-2050)</i> | | | | |
| H_s (m) | 1.6 | 1.7 | 1.7 | 1.8 |
| T_p (s) | 10.5 | 10.2 | 10.4 | 10.5 |
| θ (degrees) | 130 | 136 | 143 | 131 |
| <i>RCP2.6 Distant Future (2050-2100)</i> | | | | |
| H_s (m) | 1.6 | 1.7 | 1.7 | 1.8 |
| T_p (s) | 10.5 | 10.2 | 10.4 | 10.5 |
| θ (degrees) | 128 | 135 | 143 | 131 |
| <i>RCP8.5 Near Future (2010-2050)</i> | | | | |
| H_s (m) | 1.6 | 1.7 | 1.7 | 1.8 |
| T_p (s) | 10.5 | 10.3 | 10.4 | 10.5 |
| θ (degrees) | 129 | 135 | 143 | 131 |
| <i>RCP8.5 Distant Future (2050-2100)</i> | | | | |
| H_s (m) | 1.6 | 1.7 | 1.7 | 1.8 |
| T_p (s) | 10.5 | 10.3 | 10.4 | 10.5 |
| θ (degrees) | 128 | 136 | 143 | 141 |

Atmospheric classification as a framework for assessing future coastal vulnerability.



(a)



(b)



Fig. 8.10 Wave roses derived from the the HadleyGEM2-ES RCP8.5 scenario for the (a) near (2010-2050) and (b) distant future (2050-2100).

8.4 Discussion

Classification of synoptic scale circulation patterns into discrete states with strong links to ocean wave behaviour has been proposed as a new framework for coastal risk assessment (Pringle *et al.*, 2014). The separation of atmospheric features into classes relies strongly upon a selected variable of interest such as the wave height. Previous studies have only applied such classification techniques to locations such as the KwaZulu-Natal coastline where there are reliable wave observation records (Pringle *et al.*, 2014). However with improvements in simulated reanalysed wave data, it is important to clarify the ability of the classification scheme to delineate CPs based only on such simulated datasets. If the methodology can be shown to be insensitive to the source of wave data then the model can be used for future risk assessment studies on a global scale.

Wave climates predicted by global models have improved due to increased resolution, improved model formulations, and the application of reanalysis with data assimilation. However the ability of such models to predict extreme wave events for coastal engineering design and vulnerability applications still requires a clear understanding of the fidelity of these simulated data sets. In general a strong relationship exists between observed and modelled (with reanalysis) wave data. However extreme events can be relatively poorly reproduced in terms of significant wave heights although the timing of the events appears approximately correct. For our case study site, observed and modelled wave data were shown for different CPs such as mid-latitude cyclones, cut-off lows and tropical cyclones. As noted in §8.2.1 these CPs are fundamental drivers of the regional wave climate generally, and extreme events in particular.

To test the robustness of the CP classification technique reanalysis wave data from the ERA-Interim dataset was used as the guiding variable. The classes derived from this wave data were compared to those derived from observed wave data. There is a strong inter-dependence between the two classification outcomes. Furthermore there is good spatial resemblance between classes and the occurrence frequencies for similar classes are in broad agreement. This implies that the technique is robust with respect to small differences in the wave data used.

The implications of our results for coastal risk assessment are important. For example the use of reanalysis data allows one to apply the technique to regions where direct wave measurements are not available. Furthermore the results demonstrate an advantage of using CPs as a framework for assessing the implications of future climate

Atmospheric classification as a framework for assessing future coastal vulnerability.

changes on waves. The recently developed stochastic wave simulation technique based on the CPs was applied to output from a GCM for different future scenarios. This technique differs from typical CP-based downscaling techniques for assessing climate change because continuous sequences of waves are simulated based on the CP-Wave links (e.g. Abiodun *et al.*, 2015; Bárdossy & Pegram, 2011; Perez *et al.*, 2015). In general CP-based downscaling only uses the statistical properties of a variable of interest associated with individual CPs to correct GCM predictions of the variable of interest. Recently (Perez *et al.*, 2015) showed how the relationship between a variable of interest and CPs can be used to predict monthly or seasonal statistics thereof .

The CPs can also be used to develop quantile-quantile transforms between observed and GCM-predicted values (e.g. Bárdossy & Pegram, 2011). The Q-Q transforms can then be applied to correct the statistical properties of variables predicted by the GCM. However such techniques are limited to statistical measures and cannot be used to explicitly describe complex processes, for example beach response due to wave forcing. Therefore simulating continuous wave sequences are advantageous when assessing nearshore processes such as cross-shore and longshore beach responses (e.g. Pringle *et al.*, 2015).

According to the GCM used herein significant changes in the regional wave climate for our case study location are only identifiable for the high emission scenario. The changes show up in the wave height return period statistics and suggest that wave heights with return periods $T_R \geq 5$ years will increase in the distant future. An additional advantage of incorporating CPs into future wave climate assessment studies is that they incorporate important information on wave directional characteristics. For example if GCMs predict a future change in CP occurrence frequencies this may imply an associated change in wave directional characteristics. This could in turn have implications for coastal vulnerability since a change in wave direction may lead to coastlines becoming more exposed to wave attack. For our case study location only small changes in wave directional statistics are predicted for the simulation periods used. However small changes in wave statistics may have significant effects on processes such as sediment transport rates and associated beach erosion. In their study Corbella & Stretch (2012b) estimated that beach erosion will increase at a rate of 0.14%/year/storm and 0.2%/year/storm for 0.0057 m/yr increases in H_s along the east coast of South Africa. Furthermore Perez *et al.* (2015) estimated that on average wave heights and periods in the North Atlantic and mediterranean will decrease in the distant future.

8.5 Conclusion

Synoptic scale atmospheric classification has been shown to be a robust framework for coastal risk assessment by comparing classes derived using reanalysis wave data and those derived from directly measured wave data. In general the results suggest that there is a strong inter-dependence between the classifications. This has important implications for coastal risk assessment including those associated with future climate change scenarios. This has been demonstrated using a recent stochastic wave simulation technique that is linked to circulation patterns that are the main drivers of waves. The technique can be applied to any location including those where wave data are limited to those obtained from global wave models.

Chapter 9

Synthesis and Conclusion

9.1 Introduction

This chapter presents a summary of the findings of the research papers (Chapters 3-8) described herein regarding the statistical links between atmospheric circulation patterns and regional wave climates. The conclusions are presented followed by recommendations for future research.

9.2 Improvements to Coastal Vulnerability Assessment

The primary focus of this research was to present a new framework to statistically simulate regional wave climates so as to improve current coastal vulnerability assessment techniques. Statistical methods have an advantage over process based approaches because they are less computationally demanding. Therefore it is possible to easily generate large synthetic wave datasets with statistics that are similar to wave observations. The datasets can then be incorporated within vulnerability assessment studies. However current statistical simulation techniques are not directly linked to the meteorological features that drive wave climates. This makes it difficult to delineate wave characteristics from various sources. Furthermore purely statistical models are not restricted by the physical limitations of wave development such as the wave direction relative to the coastal orientation. This research has addressed these issues by identifying the statistical links between regional wave climates and atmospheric circulation patterns. The links are then used to stochastically simulate synthetic wave datasets.

Synthesis and Conclusion

Classification is a useful tool when identifying trends in complex environments because it attempts to group similar states together thus reducing data dimensionality. Atmospheric classification was used herein to delineate the links between wave development and synoptic scale circulation (see Chapter 3). The classification method used herein was based on fuzzy logic and statistical links to wave behaviour were used to optimize the shape of the different CP classes. Therefore the CP classes are explicitly linked to wave behaviour. This approach differs from other classification techniques that derive the classes first and then link them to a variable of interest. This has direct implications when linking a surface variable to the occurrence of specific meteorological features. It is shown in Appendix A that the drivers of different surface variables such as waves and precipitation can differ significantly. The fuzzy rule based classification described in Chapter 3 was able to successfully delineate between the meteorological drivers of the KwaZulu Natal wave climate. Furthermore extreme wave events were shown to be associated with specific meteorological features. This has direct implications for simulating waves because the occurrences of these meteorological features are strongly associated with large waves.

The optimal number of CP classes required to explain wave behaviour was found to be about 16 (Chapter 4). This was based on the ability of the classes to explain wave behaviour, particularly the occurrence of extreme events. A method based on the Shannon entropy was used to quantify classification quality. This approach differs from other classification quality measures in that it directly links the CP classes to waves. Furthermore increasing the temporal resolution of the pressure data improved the detail with which the classification captured extreme wave events. A low pressure region to the east of the KwaZulu Natal coastline is strongly associated with large waves. The high/low coupling drives strong winds towards the coastline.

The classification was applied together with a swell tracking algorithm to evaluate sources of wave energy along the KwaZulu Natal coastline (Chapter 5). This research demonstrated the ability of the classification to identify the CPs associated with different wave energy directional spectra. The orientation of the high/low pressure regions are directly linked with the shape of the spectra. This is a beneficial property of the classification that can be exploited for modelling purposes.

The CP-wave links provide a means to improve stochastic simulations of regional wave climates while retaining a physical basis. Chapter 6 presents a wave simulation technique that is based on the occurrences of different CPs. The ability of the classification to delineate wave behaviour was exploited to simulate the waves. The CPs

9.3 An Alternative to Reanalysed Wave Datasets

naturally define changes in the wave state between periods associated with large waves and those associated with smaller waves. Archimedean Copulas were used to describe the dependence structure between wave climate variables conditioned on the CPs. The synthetic wave dataset exhibited similar characteristics to the observed dataset. An extensive validation exercise described in Chapter 7 demonstrated two important properties of the simulated waves. Firstly they exhibit statistics that are similar to the observed dataset. Secondly they can be easily incorporated into vulnerability assessments involving complex nearshore processes. Of particular interest is their use in simulating continuous sequences of a process of interest. For example the waves were applied to a new semi-empirical cross shore beach response model. The model relates shoreline positions to the incoming wave energy and the shoreline can either accrete or erode depending on the steady state equilibrium shoreline position.

The wave simulation model is strategically important for evaluating future wave climate scenarios due to climate change effects. Changes in CP frequencies predicted by global climate models for different future climate scenarios will have a direct impact on wave behaviour. This can be easily incorporated into the wave simulation technique based on the assumption that the wave statistics for each CP remain unchanged. This is demonstrated in Chapter 8 where changes in CP frequencies predicted by the Hadley-GEM2-ES model were used to evaluate future wave climates scenarios. Significant changes to wave behaviour were only identified for a high emission scenario because these changes were only noticeable in the wave height return period statistics. In general the wave heights at given return periods are predicted to increase in the distant future for the east coast of South Africa. Furthermore Chapter 8 demonstrated that the classification algorithm can be applied to regions with little to no observed wave data. This is of fundamental importance because the classification algorithm relies strongly upon wave data. However this research has demonstrated that reanalysed wave data can be used in the classification algorithm as a supplement to wave observations.

9.3 An Alternative to Reanalysed Wave Datasets

The use of global wave datasets available from wave modelling centres such as NCEP and ECMWF has become more attractive in coastal vulnerability studies due to improvements in modelling techniques. Furthermore the datasets span approximately 30-40 years. However the reanalysed data still suffer significant drawbacks. For ex-

ample extreme events are not well captured because they rely strongly upon accurate wind field forcing and reanalysis with data assimilation. The accurate descriptions of extreme events are fundamental to coastal vulnerability assessments. Furthermore the modelling techniques are not well suited to assess future wave climate scenarios due to climate change effects.

The CP-wave simulation technique discussed herein is an attempt to address these issues. The simulation technique is not computational demanding lending itself to generating large synthetic wave datasets. Extreme events are well described and the CP-wave simulation technique is well suited to assess future wave behaviour.

9.4 Elements of Originality

This study has presented a new framework to assess coastal vulnerability by focussing on the meteorological drivers of regional wave climates. An existing classification algorithm, originally developed to evaluate the drivers of rainfall events, was modified to explore the links between waves and synoptic scale circulation patterns. The application of this classification procedure to aspects of coastal engineering is completely new. Furthermore the circulation patterns derived herein have strong links to wave behaviour because the classification algorithm is driven by the waves to an optimal solution.

The relationship between the waves and the CPs were then used to stochastically simulate waves. This stochastic simulation technique presents the first mixed approach to wave simulation. The simplicity of statistical models is exploited whilst retaining the links to the physical drivers of waves. The simulation technique is well suited to assess future wave behaviour because it is based on synoptic scale meteorology. Atmospheric circulation is generally well predicted by GCMs because it involves processes that range large spatial scales. Furthermore the continuous sequences of waves are well suited to evaluate nearshore processes required for current and future coastal risk assessments.

This study has shown that the classification algorithm can be applied to regions in which only reanalysed wave data is available. Therefore the CP-wave simulation technique can be applied to these regions, improving current and future coastal vulnerability assessments. It is likely that wave simulations based on reanalysed datasets will under estimate extreme wave events. However future improvements in global wave modelling techniques will address these issues and improve wave climate estimation.

9.5 Conclusion

The statistical links between atmospheric circulation patterns and regional wave climates were exploited to delineate wave behaviour along the east coast of South Africa. A fuzzy rule based classification algorithm was developed that optimizes the shape of a set of CPs that drive wave development. The algorithm was applied to anomalies in the 700 hPa geopotential heights. This study has shown that only a few types of CPs are associated with large wave events along the east coast of South Africa. Strong high/low pressure regions to the east of the South Africa can drive large waves towards the coastline. Therefore the classification provides significant insight into the drivers of complex wave climates. A sensitivity study was applied to the algorithm to quantify the optimal arrangement of the classes and reduce subjectivity in selection of the number of CP classes. The Shannon entropy indicated that 16 CP classes provided significant insight into wave behaviour particularly in regard to extreme events.

Analysis of wave directional spectra indicated that the shape of the CPs are explicitly linked with the shape of the spectra. For example the orientation of the high and low pressures are directly related to the distribution of wave energy. This characteristic was then exploited to stochastically simulate synthetic wave sequences with properties similar to the observed wave data.

The research presented herein has provided the framework from which to stochastically simulate regional wave climates with direct links to meteorological features that drive wave development. The advantage of simulating continuous sequences of waves is that they can be easily incorporated within beach response models. Furthermore the links between CPs and waves were shown to provide the framework to assess future wave behaviour due to climate change effects.

9.6 Suggestions for Future Research

The use of CPs to provide insight into complex wave climates is an attempt to simplify the high dimensionality of the environment. However a drawback of the classification technique presented herein is that it treats atmospheric circulation as a set of separate distinct states. Atmospheric circulation forms more of a continuum with different CPs varying smoothly between states. Therefore a Lagrangian approach to classifying different air trajectories may address this issue.

Currently the classification algorithm used herein optimizes the CP classes based

Synthesis and Conclusion

on objective functions formulated on the wave height. These objective functions can be modified further to include other variables such as wave direction and wave period. The inclusion of such variables may provide additional insights into the links between synoptic scale circulation and wave behaviour.

An alternative approach to the CP-wave simulation technique proposed herein is to simulate the wave climate conditioned on wave direction. This approach can be justified because it has been shown that the CPs delineate the associated wave energy directional distributions. Therefore the CPs can be used to define the wave direction from which the wave period and wave height can be inferred. This alternative approach has been discussed in Chapter 6.

References

- Aarnes, J. E., Krogstad, H. E., 2001. Partitioning sequences for the dissection of directional ocean wave spectra: A review. Technical Report Work Package 4 (WP4) of the EnviWave (ENG-2001-00017) Research Program, EU Energy, Environment and Sustainable Development Program, 23.
- Aarts, E., Korst, J., 1989. Simulated annealing and boltzmann machines: a stochastic approach to combinatorial optimization and neural computing. John Wiley & Sons, Chichester.
- Abiodun, B., Omar, S., Lennard, C., Jack, C., 2015. Using regional climate models to simulate extreme rainfall events in the Western Cape, South Africa. *International Journal of Climatology*.
- Bárdossy, A., 2010. Atmospheric circulation pattern classification for South-West Germany using hydrological variables. *Physics and Chemistry of the Earth* 35, 498–506.
- Bárdossy, A., Duckstein, L., Bogardi, I., 1995. Fuzzy rule-based classification of atmospheric circulation patterns. *International Journal of Climatology* 15, 1087–1097.
- Bárdossy, A., Pegram, G., 2011. Downscaling precipitation using regional climate models and circulation patterns towards hydrology. *Water Resources Research* 47, 1–18.
- Bárdossy, A., Pegram, G., Sinclair, S., Pringle, J., Stretch, D., 2015. Circulation patterns identified through spatial rainfall and ocean wave fields in southern africa. *Frontiers in environmental science* 3.
- Bárdossy, A., Plate, E., 1991. Modeling daily rainfall using a semi-markov representation of circulation pattern occurrence. *Journal of Hydrology* 122, 33–47.
- Bárdossy, A., Stehlik, J., Caspary, H. J., 2002. Automated objective classification of daily circulation patterns for precipitation and temperature downscaling based on optimized fuzzy rules. *Climate Research* 23, 11–22.
- Barnett, K., 19–25 April 1999. The management of durban’s beaches: A historical perspective. In: *Fifth International Conference of Coastal and Port Engineering in Developing Countries*. COPEDEC V, Cape Town.
- Bartsch, H., 1974. *Handbook of Mathematical Formulas*. Academic Press, New York.

References

- Booij, N., Ris, R., Holthuijsen, L. H., 1999. A third-generation wave model for coastal regions 1. model description and validation. *Journal of Geophysical Research* 104, 7649–7666.
- Box, G., Cox, D., 1964. An analysis of transformations. *Journal of the royal statistical society: series B* 26, 211–252.
- Box, G., Jenkins, G., 1976. *Time series analysis, forecasting and control*, 2nd Edition. Holden-Day, San Francisco.
- Caires, S., Steel, A., Bidlut, J., Graham, N., Swail, V., 2004. Intercomparison of different wind-wave reanalysis. *Journal of Climate* 17 (10), 1893–1913.
- Callaghan, D., Ranasinghe, R., Roelvink, D., 2013. Probabilistic estimation of storm erosion using analytical, semi-empirical and process based storm erosion models. *Coastal Engineering* 82, 64–75.
- Callaghan, D. P., Nielson, P., Short, A., Ranasinghe, R., 2008. Statistical simulation of wave climate and extreme beach erosion. *Coastal Engineering* 55, 375–390.
- Camus, P., Mendez, F., Medina, R., no, A. C., 2011. Analysis of clustering and selection algorithms for the study of multivariate wave climate. *Coastal Engineering* 58, 453–462.
- Camus, P., Menéndez, M., Méndez, F., Izaguirre, C., Espejo, A., Cánovas, V., Pérez, J., Rueda, A., Losada, I., Medina, R., 2014. A weather-type statistical downscaling framework for ocean wave climate. *Journal of Geophysical Research: Oceans* 119, 1–17.
- CERC, 1984a. *Shore protection manual*. Vol. 1. Vicksburg, Miss. :Dept. of the Army, Waterways Experiment Station, Corps of Engineers, Coastal Engineering Research Center.
- CERC, 1984b. *Shore protection Manual*. U.S. army corps of engineers, Coastal engineering research centre, Washington, D.C.: U.S. government printing office.
- Chawla, A., Spindler, D., Tolman, H., 2013. Validation of a thirty year wave hindcast using the climate forecast system reanalysis winds. *Ocean Modelling* 70, 189–206.
- Collins, W., Bellouin, N., Doutriaux-Boucher, M., Gedney, N., Halloran, P., Hinton, T., Hughes, J., Jones, C., Joshi, M., Liddicoat, S., Martin, G., O’Connor, F., Rae, J., Senior, C., Sitch, S., Totterdell, I., Wiltshire, A., Woodward, S., 2011. Development and evaluation of an Earth-system model HadGEM2. *Geosci. Model Dev. Discuss.* 4, 997–1062.
- Compagnucci, R., Richmond, M., 2008. Can principal component analysis provide atmospheric circulation or teleconnection patterns? *International Journal of Climatology* 28, 703–726.
- Corbella, S., Pringle, J., Stretch, D., 2015. Assimilation of ocean wave spectra and atmospheric circulation patterns to improve wave modelling. *Coastal Engineering* 10, 40–53.

- Corbella, S., Stretch, D., 2012a. Multivariate return periods of sea storms for coastal erosion risk assessment. *Natural Hazards and Earth System Sciences* 12, 2699–2708.
- Corbella, S., Stretch, D., 2014a. Directional wave spectra on the east coast of south africa. *Journal of the South African Institution of Civil Engineering* 56, 53–64.
- Corbella, S., Stretch, D. D., 2012b. Predicting coastal erosion trends using non-stationary statistics and process-based models. *Coastal Engineering* 70, 40–49.
- Corbella, S., Stretch, D. D., 2012c. Shoreline recovery from storms on the east coast of South Africa. *Natural Hazards and Earth System Sciences* 12, 11–22.
- Corbella, S., Stretch, D. D., 2012d. The wave climate on the the kwazulu natal coast. *Journal of the South African Institute of Civil Engineering* 54 (2), 45–54.
- Corbella, S., Stretch, D. D., 2013. Simulating a multivariate sea storm using Archimedean copulas. *Coastal Engineering* 76, 68–78.
- Corbella, S., Stretch, D. D., 2014b. Wave spectra on the east coast of south africa. *South African Institution of Civil Engineering* 56 (3), 53–64.
- Cotton, P., Carter, D., 1994. Cross-calibration of TOPEX, ERS-1, and GEOSAT wave heights. *Journal of Geophysical Research* 99 (C12), 25025–25033.
- Cowell, P., Stive, M. J. F., Niedoroda, A. W., Swift, D. J. P., de Vriend, H. J., Buijsman, M. C., Nicholls, R. J., Roy, P. S., Kaminsky, G. M., Cleveringa, J., Reed, C. W., de Boer, P. L., 2003. The coastal-tract (part 2): Applications of aggregated modeling of lower-order coastal change. *Journal of Coastal Research* 19 (4), 828–848.
- Cunha, C., Guedes Soares, C., 1999. On the choice of data transformation for modelling time series of significant wave height. *Ocean Engineering* 26, 489–506.
- De Michele, C., Salvadori, G., Passoni, G., Vezzoli, R., 2007. A multivariate model of sea storms using copulas. *Coastal Engineering* 54, 734–751.
- Dean, R., Dalrymple, R., 1991. *Water Wave Mechanics for Engineers and Scientists*. Vol. 2. World Scientific, Singapore.
- Dee, D., Uppala, S., Simmons, A., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M., Balsamo, G., Bauer, P., Bechtold, P., Beljaars, A., van de Berg, L., Bidlot, J., Bormann, N., Delsol, C., Dragani, R., Fuentes, M., Geer, A., Haimberger, L., Healy, S., Hersbach, H., Hólm, E., Isaksen, J., Kállberg, P., Köhler, M., Matricardi, M., McNally, A., Monge-Sanz, B., Morcette, J., Park, B., Peubey, C., de Rosnay, P., Tavolato, C., Thépaut, J., Vitart, F., 2011. The era-interim reanalysis: configuration and performance of the data assimilation system. *Quarterly journal of the royal meteorological society* 137, 553–597.
- Espejo, A., Camus, P., Losada, I., Méndez, F., 2014. Spectral ocean wave climate variability based on atmospheric circulation patterns. *Journal of Physical Oceanography* 44, 2139–2152.

References

- Forbes, D., Parker, G., Manson, G., Ketch, L., 2004. Storms and shoreline retreat in the southern gulf of st. lawrence. *Marine Geology* 210, 169–204.
- Genest, C., Favre, A., 2007. Everything you always wanted to know about copula modelling but were afraid to ask. *Journal of Hydrologic Engineering* 12 (4), 347–368.
- Gerling, T. W., 1992. Partitioning sequences and arrays of directional ocean wave spectra into component wave systems. *Journal Atmospheric and Oceanic Technology* 9, 444–458.
- Gilhousen, D., 1999. Improvement in national buoy center measurements. In: Swail, V. (Ed.), *Achievements in Marine Climatology*. Environment-Canada, pp. 79–89.
- Goda, Y., 2008. *Random Seas and Design of Maritime Structures*, 2nd Edition. World Scientific Publishing Co. Pte. Ltd., Singapore.
- Gotovac, H., Gotovac, B., 2009. Maximum entropy algorithm with inexact upper entropy bound based on Fup basis functions with compact support. *Journal of Computational Physics* 228, 9079–9091.
- Guedes Soares, C., Cunha, C., 2000. Bivariate autoregressive models for the time series of significant wave height and mean period. *Coastal Engineering* 40, 297–311.
- Guedes Soares, C., Ferreira, A., 1996. Representation of non-stationary time series of significant wave height with autoregressive models. *Probabilistic engineering mechanics* 11, 139–148.
- Gyasi-Agyei, Y., Pegram, G., 2014. Interpolation of daily rainfall networks using simulated radar fields for realistic hydrological modelling of spatial rains field ensembles. *Journal of Hydrology* 519, 777–791.
- Hanson, J. L., Phillips, O. M. ., 2001. Automated analysis of ocean surface directional wave spectra. *Journal of Atmospheric and Oceanic Technology* 18, 277–293.
- Hartigan, J., Wong, M., 1979. A k-means clustering algorithm. *Journal of the Royal Statistical Society. Series C (Applied Statistics)* 28 (1), 100–108.
- Hartung, J., Bärbel, E., Klösener, K., 1999. *Statistik*. R Oldenbourg, München.
- Hawkes, P., Gouldby, B., Tawn, J., Owen, M., 2002. The joint probability of waves and water levels in coastal engineering design. *Journal of Hydraulic Research* 40 (3), 241–251.
- Hess, P., Brezowsky, H., 1952. *Katalog der Großwetterlagen Europas*. Ber. DT. Wetterd. in der US-Zone 33, Bad Kissingen, Germany.
- Hewitson, B., Crane, R., 2002. Self-organizing maps: applications to synoptic climatology. *Climate Research* 22, 13–26.
- Hong, Y., Hsu, K., Sorooshian, S., Gao, X., 2005. Self-organizing nonlinear output (sono): A neural network suitable for cloud patch-based rainfall estimation at small scales. *Water Resources Research* 41 (3).

- Hunter, I., 1987. The Weather of the Agulhas Bank and the Cape Town South Coast. Thesis for the Department of Oceanography. University of Cape Town, Cape Town.
- Huth, R., 1996. An intercomparison of computer-assisted circulation classification methods. *International Journal of Climatology* 16, 893–922.
- Huth, R., Beck, C., Philipp, A., Demuzere, M., Ustrnul, Z., Cahynová, M., Kyselý, J., Tveito, O. E., 2008. Classifications of atmospheric circulation patterns: recent advances and application. *Annals of the New York academy of sciences* 1146, 105–152.
- (IPCC) Intergovernmental Panel on Climate Change, 1992. Global climate change and the rising challenge of the sea. Tech. rep., Coastal Zone Management Subgroup, Ministry of Transport, Public Works and Water Management – Tidal Waters Division, The Hague.
- Isaacson, M., MacKenzie, N., 1981. Long-term distributions of ocean waves: a review. *Journal of Waterway, Port, Coastal and Ocean Engineering* 107 (WW2), 93–109.
- Jacobeit, J., 2010. Classifications in climate research. *Physics and Chemistry of the Earth, Parts A/B/C* 35 (9–12), 411 – 421.
- Jara, M., González, M., Medina, R., 2015. Shoreline evolution model from a dynamic equilibrium beach profile. *Coastal Engineering* 99, 1–14.
- Jones, C., Hughes, J., Bellouin, N., Jones, G., Knight, J., Liddicoat, S., O’Connor, F., Andres, R., Bell, C., Boo, K., Bozzo, A., Butchart, N., Cadule, P., Corbin, K., Doutriaux-Boucher, M., Friedlingstein, P., Gornall, J., Gray, L., Halloran, P., Hurtt, G., Ingram, W., Lamarque, J., Law, R., Meinhausen, M., Osprey, S., Palin, E., Parsons Chini, L., Raddatz, T., Sanderson, M., Stellar, A., Schurer, A., Valdes, P., Wood, N., Woodward, S., Yoshioka, M., Zerroukat, M., 2011. The HadGEM2-ES implementation of CMIP5 centennial simulations. *Geosci. Model Dev.* 4, 543–570.
- Kamphuis, J., 1991. Alongshore sediment transport rate. *Journal of Waterways, Port, Coastal and Ocean Engineering, ASCE* 117 (6), 624–641.
- Karunaratna, H., Pender, D., Ranasinghe, R., Short, A., Reeve, D., 2013. The effects of storm clustering on beach profile variability. *Marine Geology* 348, 103–112.
- Kohonen, T., 1989. Self-organization and associative memory, 3rd Edition. Springer-Verlag, Berlin.
- Kohonen, T., 1990. The self-organizing map. *Proc IEEE* 78 (9), 1464–1480.
- Kohonen, T., 1991. Self-organizing maps: Optimization approaches. In: *Proceedings of the International Conference on Artificial Neural Networks*. Espoo, Finland, pp. 981–990.
- Kohonen, T., 1995. Self-organizing maps. Springer-Verlag, Heidelberg.
- Komar, P. D., Allan, J. C., Ruggiero, P., 2010. Ocean wave climates: Trends variations due to earth’s changing climate. In: Kim, Y. (Ed.), *Handbook of coastal and ocean engineering*. World Scientific Publishing Co., Hackensack, USA, Ch. 35, pp. 972–995.

References

- Komen, G., Cavaleri, L., Donelan, M., Hasselmann, K., Janssen, P., 1994. Dynamics and modelling of ocean waves. Cambridge University Press, Cambridge, UK.
- Kriebel, D., Dean, R., 1993. Convolution method for time dependent beach-profile response. *Journal of Waterway Port. C. Ocean Eng.* 119 (2), 204–226.
- Kruger, A. C., 2004. Climate of South Africa: Climate regions. Report WS45, South African Weather Service, Pretoria.
- Lamb, H. H., 1972. British Isles weather types and a register of the daily sequence of circulation patterns, 1861–1971. Meteorological Office, Geophysical Memoir No. 116, London: HMSO.
- Larson, M., Kraus, N., Byrnes, M., 1990. Sbeach: Numerical model for simulating storm-induced beach change, Report 2, Numerical Formulation and Model Tests, Technical Report CERC -89-9. Tech. rep., US Army Engineer Waterways Experiment Station, Vicksburg, MS.
- Li, F., van Gelder, P., Callaghan, D., Jongejan, R., van Heler, C., Ranasinghe, R., 2013. Probabilistic modeling of wave climate and predicting dune erosion. *Jornal of Coastal research* 65, 760–765.
- Li, F., van Gelder, P., Vrijling, J., Callaghan, D., Jongejan, R., Ranasinghe, R., 2014. Probabilistic estimation of coastal dune erosion and recession by statistical simulation of storm events. *Applied ocean research* 47, 53–62.
- Lloyd, S., 1982. Least squares quantization in pcm. *IEEE Transactions on Information Theory* 28 (2), 129–137.
- Maidment, D., 1993. *Handbook of Hydrology*.
- Mather, A. A., Stretch, D. D., 2012. A perspective on sea level rise and coastal storm surge from southern and eastern Africa: a case study near Durban, South Africa. *Water* 4, 237–259.
- Mendez, F. J., Menendez, M., Luceno, A., Medina, R., Graham, N. E., 2008. Seasonality and duration in extreme value distributions of significant wave height. *Ocean Engineering* 35, 131–138.
- Michelle, J., 1893. On the highest wave in water. *Philosophy Magazine* 36, 430–435.
- Miller, J., Dean, R., 2004. A simple new shoreline change model. *Coastal Engineering* 51, 531–556.
- Mínguez, R., Espejo, A., Tomás, A., Méndez, F., Losada, I., 2011. Directional calibration of wave reanalysis databases using instrumental data. *Journal of Atmospheric and Oceanic Technology* 28, 1466–1485.
- Monbet, V., Ailliot, P., Prevosto, M., 2007. Survey of stochastic models for wind and sea state time series. *Probabilistic engineering mechanics* 22, 113–126.

- Monbet, V., Prevosto, M., 2000. Bivariate simulation of non-stationary and non-gaussian observed processes: Application to sea state parameters. In: Proceedings of the Tenth (2000) International Offshore and Polar Engineering Conference. Seattle, USA.
- Moss, R., Edmonds, J., Hibbard, K., Manning, M., Rose, S., van Vuuren, D., Carter, T., Emori, S., Kainuma, M., Kram, T., Meehl, G., Mitchell, J., Nakicenovic, N., Riahi, K., Smith, S., Stouffer, R., Thompson, A., Weyant, J., Wilbanks, T., 2010. The next generation of scenarios for climate change research and assessment. *Nature* 463, 747–756.
- Nelson, R., 2006. An Introduction to Copulas, 2nd Edition. Springer, New York, USA.
- Palmer, B. J., R. Van der Elst, Mackay, F., Mather, A. A., Smith, A. M., Bundy, S. C., Thackery, Z., Leuci, R., Parak, O., 2011. Preliminary coastal vulnerability assessment for KwaZulu-Natal, South Africa. *Journal of Coastal Research* 64, 1390–1395.
- Perez, J., Menedez, M., Camus, P., Mendez, F., Losada, I., 2015. Statistical multi-model climate projections of surface ocean waves in Europe. *Ocean Modelling*. URL <http://dx.doi.org/10.1016/j.ocemod.2015.06.001>
- Pethick, J., 2001. Coastal management and sea-level rise. *Catena* 42, 307–322.
- Petrov, V., Guedes Soares, C., Gotovac, H., 2013. Predicting extreme significant wave heights using maximum entropy. *Coastal engineering* 74, 1–10.
- Philipp, A., Beck, C., Huth, R., Jacobeit, J., 2014. Development and comparison of circulation type classifications using the cost 733 dataset and software. *International Journal of Climatology* Article in Press.
- Phillip, A., Della-Marta, P., Jacobeit, J., Fereday, D., Jones, P., Moberg, A., Wanner, H., 2007. Long term variability of daily North Atlantic European pressure patterns since 1850 classified by simulated annealing clustering. *Journal of Climate Change* 20, 231–253.
- Portilla, J., Ocamp-Torres, F. J., Monbaliu, J., 2009. Spectral partitioning and identification of wind sea and swell waves. *Journal of Atmospheric and Oceanic Technology* 26, 107–122.
- Preston-Whyte, R. A., Tyson, P. D., 1988. The atmosphere and weather of Southern Africa. Oxford University Press, Cape Town.
- Pringle, J., Stretch, D., 2015. Stochastic simulation of regional wave climates conditioned on synoptic scale meteorology. part 1: Methodology. *Coastal Engineering* (Submitted).
- Pringle, J. J., Stretch, D. D., Bárdossy, A., 2014. Automated classification of the atmospheric circulation patterns that drive regional wave climates. *Natural Hazards and Earth System Sciences* 14, 2145–2155.

References

- Pringle, J. J., Stretch, D. D., Bárdossy, A., 2015. On linking atmospheric circulation patterns to extreme wave events for coastal vulnerability assessments. *Natural Hazards*, 1–15.
URL <http://dx.doi.org/10.1007/s11069-015-1825-4>
- Prinos, P., Galiatsatou, P., 2010. Coastal flooding: Analysis and assessment of risk. In: Kim, Y. (Ed.), *Handbook of coastal and ocean engineering*. World Scientific Publishing Co., Hackensack, USA, Ch. 38, pp. 1039–1071.
- Ramos, A., Sprenger, M., Wernli, H., Durán-Quesada, A., Lorenzo, M., Gimeno, L., 2014. A new circulation type classification based upon lagrangian air trajectories. *Frontiers in Earth Science* 2, 1–18.
- Ranasinghe, R., Callaghan, D., Roelvink, D., June 2013. Does a more sophisticated storm erosion model improve probabilistic erosion estimates? In: *7th international conference on Coastal Dynamics*. Coastal Dynamics, France, pp. 1277–1286.
- Richman, M., 1986. Rotation of principle components. *Journal of Climatology* 6, 293–335.
- Roelvink, D., Reniers, A., van Dongeren, A., van Thiel de Vries, J., McCall, R., Lescinski, J., 2009. Modelling storm impacts on beaches, dunes and barrier islands. *Coastal Engineering* 56, 1133–1152.
- Roffel, B., Chin, R., 1993. Application of fuzzy logic in the control of polymerization reactors. *Control engineering practice* 1 (2), 267–272.
- Rosenblatt, M., 1956. Remarks on some nonparametric estimates of a density function.
- Rossouw, J., Coetzee, L., Visser, C., 2011. A south african wave climate study. *Coastal Engineering Proceedings* 1 (18).
URL <https://icce-ojs-tamu.tdl.org/icce/index.php/icce/article/view/3618>
- Salvadori, G., 2004. Bivariate return periods via 2-copulas. *Statistical Methodology* 1, 129–144.
- Schoonees, J., 2000. Annual variation in the net longshore sediment transport rate. *Coastal Engineering* 40, 141–160.
- Schoonees, J., Theron, A., 1993. Review of the field-data base for longshore sediment transport. *Coastal Engineering* 19, 1–25.
- Schoonees, J., Theron, A., 1996. Improvement of the most accurate longshore transport formula. In: *Coastal Engineering Proceedings*. pp. 3652–3665.
- Schumann, E., 1988. Climate and weather off natal. In: Schumann, E. (Ed.), *Lecture Notes on Coastal and Estuarine Studies*. Springer-Verlag, New York, Ch. 4, pp. 81–100.
- Scotto, M., Guedes Soares, C., 2000. Modelling the long-term time series of significant wave height with non-linear threshold models. *Coastal Engineering* 40, 313–327.

- Shannon, C., 1948. A mathematical theory of communication. *Bell System Tech. J.* 27, 379–423.
- Skivic, N., Francis, J., 2012. Self-organizing maps: A powerful tool for the atmospheric sciences. In: Johnsson, M. (Ed.), *Applications of Self-Organizing Maps*. InTech, Ch. 13, pp. 251–268.
URL <http://www.intechopen.com/books/applications-of-self-organizing-maps/self-organizing-maps-a-powerful-tool-for-the-atmospheric-sciences>
- Sklar, A., 1959. Fonctions de répartition à n dimensions et leurs marges. *Publ. Inst. Stat. Univ. Paris* 8, 229–231.
- Slonim, N., Aharoni, E., Cramer, K., 2013. Hartigan's K-Means versus Lloyd's K-Means-is it time for a change? In: *23rd International Joint Conference on Artificial Intelligence*. Beijing, China, pp. 1677–1684.
- Solari, S., Losada, M., 2011. Non-stationary wave height climate modelling and simulation. *Journal of geophysical research* (C09032).
- Stehlik, J., Bárdossy, A., 2003. Statistical comparison of european circulation patterns and development of a continental scale classification. *Theory of Applied Climatology* 76, 31–46.
- Stive, M., Ranasinghe, R., Cowell, P., 2010. Sea level rise and coastal erosion. In: Kim, Y. (Ed.), *Handbook of coastal and ocean engineering*. World Scientific Publishing Co., Hackensack, USA, Ch. 37, pp. 1023–1038.
- Stohl, A., Scheifinger, H., 1994. A weather pattern classification by trajectory clustering. *Meteorol. Z.* 3, 333–336.
- Stopa, J., Cheung, K., 2014. Intercomparison of wind and wave data from the ECMWF reanalysis interim and the NCEP climate forecast system reanalysis. *Ocean Modelling* 75, 65–83.
- Swail, V., Cox, A., 1999. On the use of ncep-ncar reanalysis surface marine wind fields for long-term north atlantic wave hindcast. *Journal of Atmospheric and Oceanic Technology* 17, 532–545.
- Taljaard, J., 1967. Development, distribution and movement of cyclones and anticyclones in the south hemisphere during the IGY. *Journal of Applied Meteorology* 6 (6), 973–987.
- Taljaard, J., 1995. *Atmospheric Circulation Systems, Synoptic Climatology and Weather Phenomena of South Africa, Part 2: Atmospheric Circulation Systems in the South African region*. South African Weather Bureau, Pretoria.
- Tolman, H., Balasubramanian, B., Burroughs, L., Chalikov, D., Chao, Y., Chen, H., Gerald, V., 2002. Development and implementation of wind-generated ocean surface wave models at NCEP. *Weather Forecasts*, 311–333.

References

- UN, 1993. Report of the united nations conference on environment and development. Vol. 1. Rio de Janeiro, pp. 3–14, resolutions adopted by the conference, New York, United Nations.
- Van der Meer, J., 1990. Static and dynamic stability of loose materials. In: Pilarczyk, K. (Ed.), Coastal Protection. Balkema, Rotterdam.
- Voorrips, A. C., Makin, V. K., Hasselmann, S., 1997. Assimilation of wave spectra from pitch and roll buoys in a north sea wave model. *Journal of Geophysical Research* 102, 5829–5849.
- WAAMDIG, 1988. The wam model - a third generation ocean wave prediction model. *Journal of Physical Oceanography* 18 (12), 1775–1810.
- Wang, P., Kraus, N., R.A Davis, Jr., 1998. Total longshore sediment transport rate in the surf zone: field measurements and empirical predictions. *Journal of Coastal Research* 14 (1), 269–282.
- Yates, M., Guza, R., O'Reilly, W., 2009. Equilibrium shoreline response: observation and modelling. *Journal of Geophysical Research* 114, C09014.
- Zadeh, L., 1965. Fuzzy sets. *information control* 8 (3), 338–353.
- Zou, Q., Chen, Y., Cluckie, I., Hewitson, R., Pan, S., Peng, Z., Reeve, D., 2013. Ensemble prediction of coastal flood risk arising from overtopping by linking meteorological, ocean, coastal and surf zone models. *Quarterly journal of the royal meteorological society* 139, 298–313.

Appendix A

Circulation Patterns identified by spatial rainfall and ocean wave fields in Southern Africa

Abstract

This paper presents the application of Fuzzy Rule Based Circulation Patterns (CPs) classification in the description and modelling of two different physical processes: rainfall regimes and ocean waves. Large ocean waves are typically generated over fetches of the order of thousands of kilometres far off shore, whereas rainfall is generated by local atmospheric variables including temperature, humidity, wind speed and radiation over the area of concern. The spatial distribution of these variables is strongly dependent on regional pressure patterns, which are similar for associated weather and wind behaviour on a given day. The choice of the CP groupings is made by searching for those CPs which generate (i) different daily rainfall patterns over mesoscale regions and (ii) wave heights from different directions at chosen shoreline locations. The method used to choose the groupings of CPs is a bottom-up methodology using simulated annealing, ensuring that the causative CPs are responsible for the character of the results. This approach is in marked distinction to top-down approaches such as k-means clustering or Self Organising Maps (SOMS) to identify several classes of CPs and then analysing the effects of those CPs on the variables of choice on given historical days. The CP groups we define are often different for the two phenomena (rainfall and waves) simply because different details of the pressure fields are responsible for wind and for precipitation. The region chosen for the application is the province of KwaZulu-Natal in South Africa, using the same set of raw geopotential heights to represent the pressure patterns, but selecting from the set those typical patterns affecting ocean waves on the one hand and regional rainfall on the other.

A.1 Introduction

Local weather (precipitation, temperature wind) and related phenomena such as floods, storm and waves are strongly dependent on atmospheric processes. These processes are very complex and highly non linear. On the other hand due to the continuity

of the atmospheric conditions these local phenomena are imbedded into large scale features. It is both of theoretical and practical importance to understand these links. Among the atmospheric variables, pressure is the driver for flow and transport. Air pressure at the land surface is a variable which can be measured simply and with good accuracy. Selected observations and meteorological models provide information on high altitude pressure conditions. Air pressure and geopotential heights are among the best modeled quantities. Therefore it is reasonable to use them as a basis to describe atmospheric circulation.

The relationship between local variables and atmospheric circulation described with the help of pressure conditions is a complicated and highly non linear one. On the other hand it reasonable to assume that similar circulation conditions cause similar local meteorological conditions. Therefore an appropriate classification can help to quantify the relationship between local variables, such as amount of rain on a day in a region or wave height and period on a shoreline, and circulation. The intention is to define classes on the basis of the circulation and the local variables via conditional distributions. The behavior of the local variable described for example using the conditional distribution should be different for the different classes from the unconditional distribution of the variable. A comprehensive summary of classification methods can be found in Jacobeit (2010).

There are different ways one can define groups of circulation patterns:

1. to define CPs using atmospheric variables only
2. to define CPs using atmospheric variables by taking the local variable into account for setting up the patterns
3. to define CPs using a combination of atmospheric variables and local variables

The first method intends to find typical distinct patterns of the atmospheric variables. The patterns differ in their defining space. Their link to local variables is determined after the classification and might for some variables yield a good distinction in the behavior while for others not.

The second method acknowledges the fact that relatively small differences of the atmospheric variables might lead to very different behavior of the local variables. This means that the classification should not intend to distinguish the patterns by producing very different CPs but to group CPs which to some extent are similar but explain the target variable as well as possible. On the other hand it is important that

Circulation Patterns identified by spatial rainfall and ocean wave fields in Southern Africa

the classification is done on the basis of the atmospheric variables only so that the classification can be used for time periods lacking the observation of the local variables.

The goals of various classification schemes can differ widely. While the classification of pressure fields on a purely statistical basis might reveal specific features of atmospheric dynamics, it might not provide the best basis for the explanation of the behavior of surface variables such as wind, temperature or precipitation. Under certain circumstances relatively small differences in the atmospheric conditions can lead to very different behavior of the surface variables. If one is interested in explaining the relationship between surface variables and atmospheric conditions then the purpose is to obtain classes of circulation patterns with distinct conditional probabilities or distributions of the selected variables.

The first classifications were developed for regions in Europe on a subjective basis (Lamb (1972) for Great Britain and Hess & Brezowsky (1952) for Germany). Automated classifications were developed for different regions in Europe, North America and China. In this paper fuzzy rule based classifications are developed for precipitation and waves in South Africa.

A.2 Methodology

The basic methodology of fuzzy rule-based classification was described for explaining precipitation behavior in Bárdossy *et al.* (1995). The main ideas are summarized here.

The classification is performed on anomalies of daily air pressure based variables: sea level pressures or geopotential heights g . These data can be obtained from reanalysis products on a regular grid. The anomalies are calculated under consideration of the annual cycle of both the mean and the standard deviation of the observed pressure. The anomaly at gridpoint x and time t is calculated as:

$$d(x, t) = \frac{g(x, t) - \bar{g}(x, J(t))}{s_g(x, J(t))} \quad (\text{A.1})$$

where $J(t)$ is the Julian date corresponding to day t (which is a value between 1 and 366), $\bar{g}(x, J(t))$ is the mean and $s_g(x, J(t))$ is the standard deviation of g at location x and Julian day $J(t)$. These anomalies d have zero mean and unit standard deviation for each grid location.

The classification is based on the location of certain anomalies. For each gridpoint x of the pressure grid G five different possibilities of anomalies are considered to define

a classification class. A triangular fuzzy number is assigned to each of these possible anomalies:

1. positive anomaly $(0, 3, +\infty)_T$
2. negative anomaly $(-\infty, -3, 0)_T$
3. non positive anomaly $(-4, -0.85, 0.25)_T$
4. non negative anomaly $(0.25, 0.85, 4)_T$
5. non representative anomaly $(-\infty, 0, +\infty)_T$

where the subscript T denotes a triangular fuzzy number with the membership function:

$$\mu_{(a,b,c)_T}(d) \begin{cases} \frac{d-a}{b-a} & \text{if } a \leq d \leq b \\ \frac{d-c}{b-c} & \text{if } b < d \leq c \\ 0 & \text{else} \end{cases} \quad (\text{A.2})$$

This means that for the classification of each gridpoint one of the above possible classes is assigned. Thus a class k can be defined as a vector $j_k(x)x \in G$ where each $j_k(x)$ is an integer such that $1 \leq j_k(x) \leq 5$.

For a given day t and a given set of rules $j_k(x) x \in G$ and $k = 1, \dots, K$ the classification is performed as follows:

1. For each rule k and each location x the membership value of $d(x, t)$ in the fuzzy set corresponding to $j_k(x)$ is calculated. These values are $\mu_{j_k(x)}(d(x, t))$
2. These individual membership values are combined to an overall degree of fulfillment of the rule k by calculating:

$$\text{DOF}(k, t) = \prod_{m=1}^4 \left[\frac{1}{\#\{x \mid j_k(x) = m\}} \sum_{x \mid j_k(x) = m} \mu_{j_k(x)}(d(x, t))^{q_m} \right]^{\frac{1}{q_m}} \quad (\text{A.3})$$

3. The k_0 index corresponding to the maximal $\text{DOF}(k, t)$ is selected as class for day t

For further details the readers are referred to Bárdossy *et al.* (1995).

The fuzzy rule based classification is performed to explain the behavior of one or more selected variables. These are denoted with the variable V . Two types of objective functions are considered. The first one relates the CPs to the exceedence of certain

Circulation Patterns identified by spatial rainfall and ocean wave fields in Southern Africa

thresholds of V . The objective is to define classes with frequencies of exceedances which differ from the unconditional frequency of occurrences. The first is O_1 :

$$O_1(\theta) = \left(\sum_{t=1}^T h_{CP(t)}(V > \theta | CP(t)) - h(V > \theta) \right)^{\frac{1}{2}} \quad (\text{A.4})$$

where h stands for the frequency of an event, V is the variable under interest and θ is a prescribed threshold. This objective expresses that a classification should help to decide if a threshold is exceeded or not.

The second objective type is related to the mean magnitude of the variable V with respect to the circulation patterns. It is formulated as:

$$O_2 = \frac{1}{T} \sum_{t=1}^T \left| \frac{(V | \bar{CP}(t))}{\bar{V}} - 1 \right| \quad (\text{A.5})$$

The second objective function evaluates the ability of the algorithm to derive classes with average values of V that are different from the unclassified average. In other words this objective function measures the separability of the classes from the mean.

These objective functions can be used in combination - for example for different thresholds θ and even for different variables. A weighted combination of them is used as the classification objective function.

The value of the objective functions depends on the classification, which itself depends on the selected rules $\{j_k(x) | x \in G, k = 1, \dots, K\}$, where K is the number of rules. Once objective functions for the classification are defined, different classifications can be compared with respect to their ability to explain the variability of the surface variable investigated. The higher the objective function values the better the classification is. Thus one can use an optimization procedure to find rules $j_k(x)$ which maximize the objective functions. The number of possible rule systems for a given number of patterns is given by the combinatorial:

$$N_K = \binom{5^{|G|}}{K}$$

This is usually an extremely large number (for example for a small pressure grid 10×10 with $|G| = 100$ nodes and $K = 10$ rules the number of possible rule systems is $> 10^{340}$) thus the best rules cannot be found by trying all possibilities. Instead an optimization method has to be applied. Simulated annealing provides a reasonable alternative as described in Bárdossy *et al.* (2002).

A.2.1 Classification quality

The quality of the classification can be measured by the above defined objective functions. Besides that other measures can be defined, which can be used for comparison, for example to find the optimal number of classes. For this one can use thresholds for the variable V . The entropy of the conditional distributions can be used to measure the binary quality of a classification:

$$H(K) = \sum_{k=1}^K h_k (p_k \log_2 p_k + (1 - p_k) \log_2 (1 - p_k)) \quad (\text{A.6})$$

where p_k is the probability of the exceedance of the threshold on a day with the k -th CP, and h_k is the frequency of the k -th CP. The quantity $H(0)$ is the entropy in the case no classification was performed:

$$H(0) = p_0 \log_2 p_0 + (1 - p_0) \log_2 (1 - p_0) \quad (\text{A.7})$$

A classification provides information for the variable V if $H(K) < H(0)$. In general a classification using K classes is better than another with L classes if:

$$H(K) < H(L)$$

A.2.2 Non-uniqueness issues

Different CP classifications can be obtained for the same region. In Philipp *et al.* (2014) the authors investigated 27 automatic circulation pattern classification approaches and found that different classification methods lead to very different classifications. The Fuzzy Rule Based classification method can lead to different classifications as the stochastic optimization obtained using simulated annealing does not necessarily lead to the same classification if another sequence of random numbers is selected. A different choice of stations or a different choice of thresholds or weights for the objective functions can lead to different classifications too. Different classifications can be compared from the viewpoint of their performance or from the viewpoint of the similarity of their classes. While the calculation of the objective functions is straightforward, the similarity of the classifications requires further measures.

Two statistics can be calculated based on the contingency tables of the pairs of classifications. For two different classifications n_{ij} is the number of days in class i for the first and in class j in the second classification. The χ^2 statistics to compare the

Circulation Patterns identified by spatial rainfall and ocean wave fields in Southern Africa

classifications is calculated as:

$$\chi^2 = \sum_{i=1}^r \sum_{j=1}^s \frac{(n_{ij}n - n_i.n.j)^2}{n_i.n.jn} \quad (\text{A.8})$$

where r is the number of classes in the first classification, s is the number of classes in the second classification, n_i is the number of days in class i for the first classification, $n.j$ is the number of days in class j for the second classification and n is the total number of days. If the classifications are independent then the χ^2 value is small, the bigger it is the more the classifications resemble each other. Based on the χ^2 values two measures of dependence were calculated, the modified Pearson coefficient C :

$$C = \sqrt{\frac{\min(r, s)}{\min(r, s) - 1} \frac{\chi^2}{\chi^2 + n}} \quad (\text{A.9})$$

and the Cramer coefficient (Hartung *et al.* (1999)) V :

$$V = \sqrt{\frac{\chi^2}{n(\min(r, s) - 1)}} \quad (\text{A.10})$$

Both coefficients are bounded by 1, and the higher they are the stronger the association is. The associations of CP classifications for the same geographical regions using different objective functions or different stochastic optimization settings are usually strong, which is a consequence of the fact that the same variables (geopotential heights) are classified.

A.2.3 Data used

The classifications were performed using ECMWF ERA reanalysis data sets. The classification is based on daily normalized anomalies, derived from the 700 hPa geopotential height with a grid resolution of 2.5° ($10^\circ - 50^\circ$ S; $0^\circ - 50^\circ$ W). Geopotential heights were obtained from the ERA-Interim data set of Dee *et al.* (2011) for the period 1979 to 2009 (<http://apps.ecmwf.int/datasets/>).

A.3 Classification for Precipitation

A.3.1 Objective functions

Precipitation has a highly skewed distribution with above 80 % probability of a dry day in many parts of South Africa. Individual very high values of precipitation can lead to "random" optima which are not robust. Instead a new variable related to the average wetness of a selected number of stations was considered:

$$W(t) = \frac{\#\{x_i; Z(x_i, t) > 0\}}{\#\{x_i; Z(x_i, t) \geq 0\}} \quad (\text{A.11})$$

by definition $0 \leq W(t) \leq 1$. This variable is taken as V for the objective functions in Equations (A.4) and (A.5).

The 24 Climate Regions defined by Kruger (2004), slightly modified by concatenating some of the very small regions (mostly in dry areas) with larger ones, were used for the classification. The map with the locations of the regions is shown in Fig. A.1. A set of representative stations was selected for each region and the CPs were classified using the above defined objective function. The classification obtained using this objective function works well for precipitation in all selected regions both for the calibration and the validation time periods. For variety, we will show some CPs and resulting rainfall in Fig. A.2 and specific CPs selected for region 6 in Fig. A.3.

A.3.2 Spatial extent of classification and non uniqueness of CP set selection

We explore three sets of classifications driven by precipitation wetness. The first CP set is conditioned on wetness in region 5 and we compare the differences in rainfall distributions from three distinct sets of CPs. The second set involves determining the robustness of the method when different sets of gauges in a region are used for CP classification. For this purpose we use data from region 6 and randomly sample 2 sets for comparison. The third experiment treats CPs defined over three regions in Fig. A.1. These are region 6 which experiences subtropical summer rainfall and the occasional hurricane, 10 which is typically savannah, experiencing summer rainfall dominated by convective systems, and 22 which has a Mediterranean climate and experiences mostly frontal system winter rainfall.

Due to the stochastic optimization no unique best classification can be achieved.

Circulation Patterns identified by spatial rainfall and ocean wave fields in Southern Africa

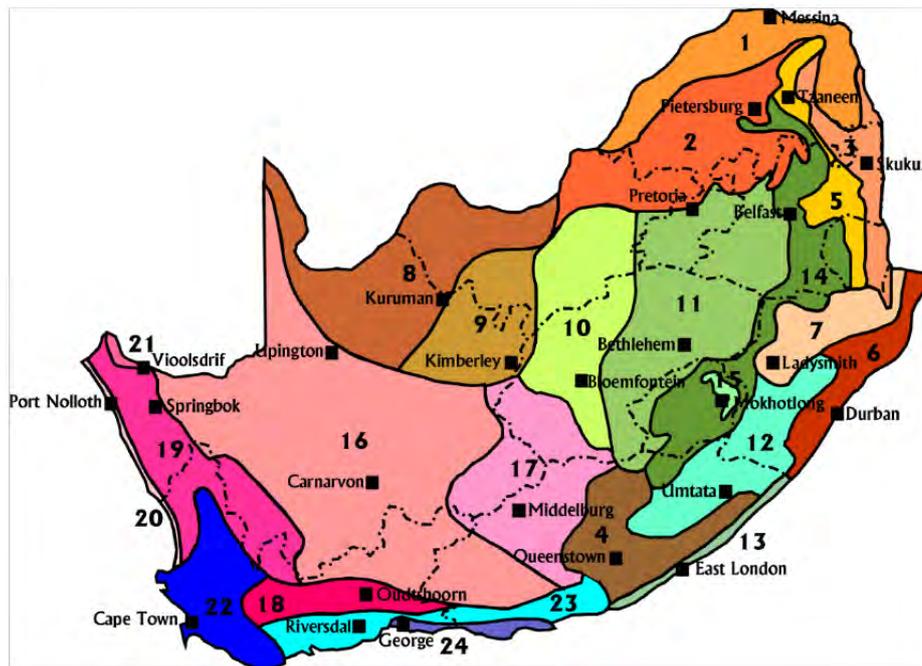


Fig. A.1 The Kruger climate regions of South Africa after Kruger (2004).

It is of interest to see how

- the selection of the stations for the objective function
- the randomness due to stochastic optimizations

influences the results. For this purpose the performance measures were calculated for different classification results.

To introduce the first experiment, which is to demonstrate the link between rainfall regimes and CPs, we offer Fig. A.2, which shows the correspondence between (i) three CPs chosen from a set of 8 in Mpumalanga and (ii) the rainfall distribution at a particular rain-gauge in the region. Fig. A.2 shows the 700 hPa anomalies for three selected CPs and the corresponding distributions of daily precipitation. CP2 is the driest with the highest probability of a dry day. CP5 is the wettest with a much lower probability of a dry day. CP8 is medium wet with statistics between CP5 and CP2. The distributions are significantly different from each other indicating that classification provides useful information on precipitation behavior.

To illustrate the methodology applied in the second experiment, we selected region 6, then we selected some gauges within the region, in different configurations, to

A.3 Classification for Precipitation

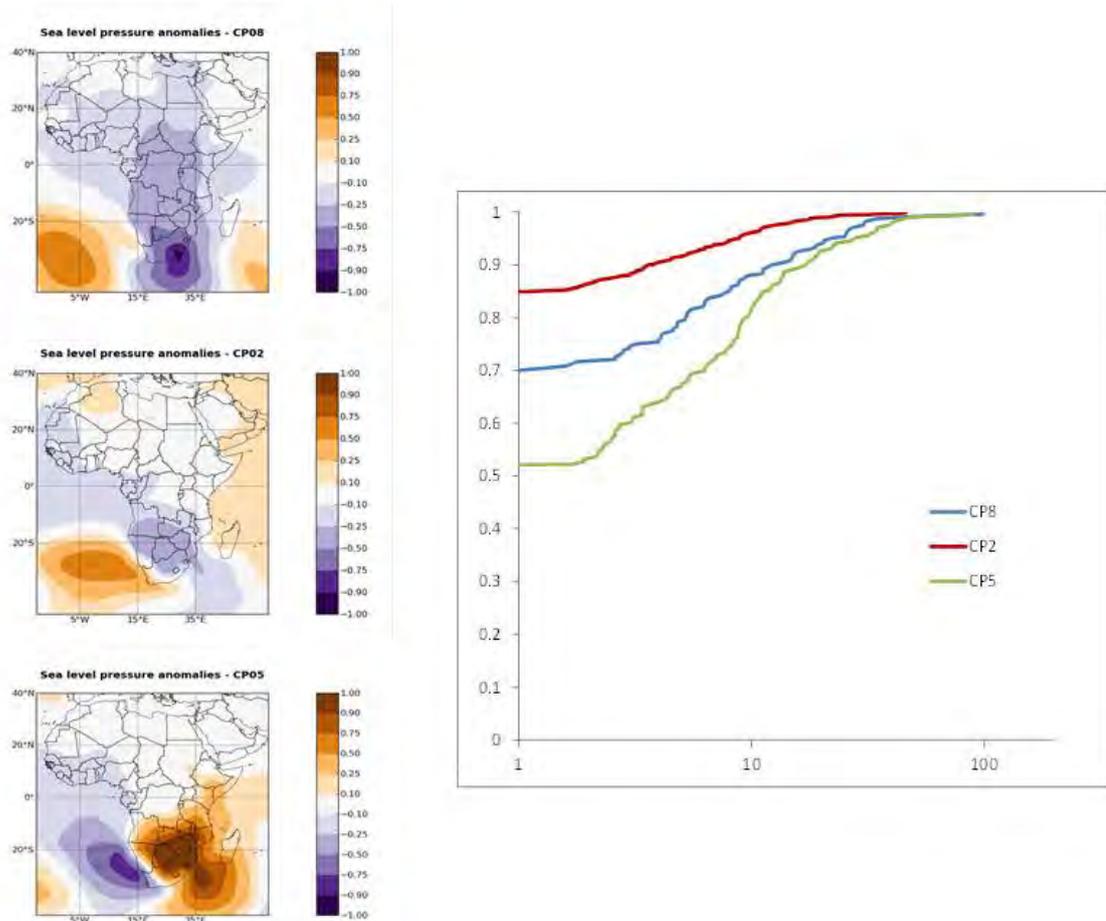


Fig. A.2 Three different CPs and the related distributions of daily precipitation in Mpumalanga in Region 5 of Fig. 1. the maps on the left are CPs 8, 2 and 5, appearing as blue (middle), red (dry) and green (wet) lines in the Fig. of frequency distributions.

Circulation Patterns identified by spatial rainfall and ocean wave fields in Southern Africa

classify the Circulation Pattern anomalies [CPs] which are the cause of different types of rainfall over this region.

We find we get similar CPs, enough similarity of shape to pick a set, [note that the labelling within each set is random, so we match by correlation of shape, not label]. Fig. A.3 shows the 700 hPa anomalies for two different classifications.

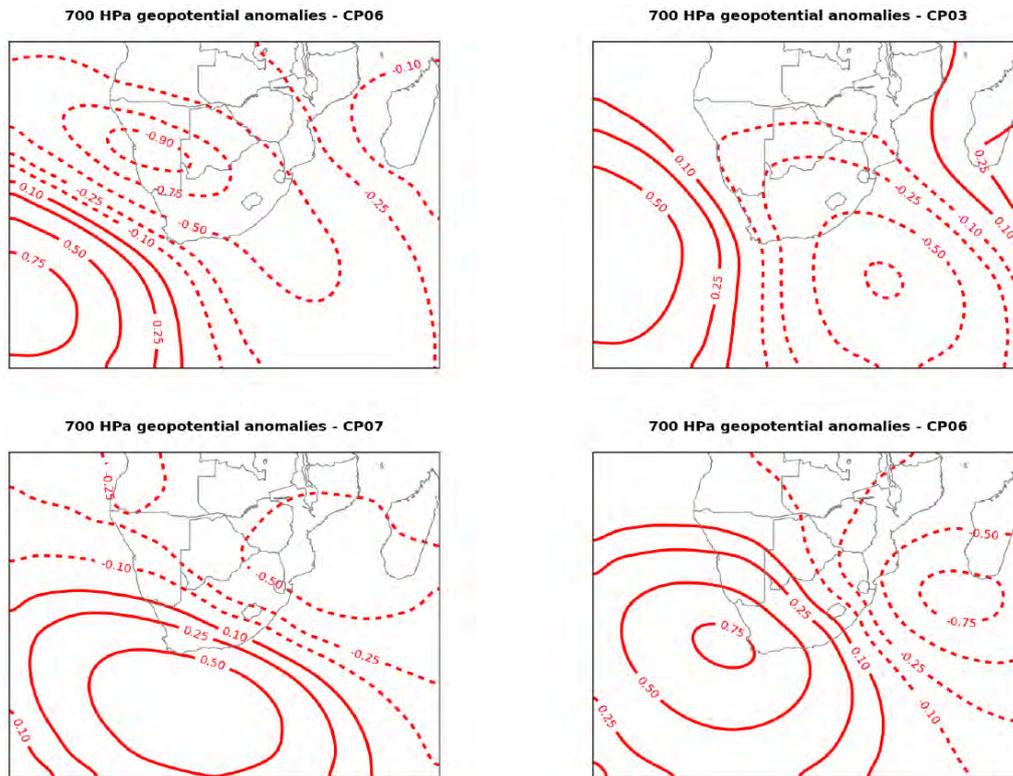


Fig. A.3 2 pairs of CPs: top two similar to each other and the bottom two likewise, selected from the CPs chosen on 2 sets of gauges in Region 6.

This similarity allows us to settle on one set of CPs per region (and season) because we have a robust method. In the next section we turn to the Infilling/repair problem – an extension of which we will use later in spatial interpolation between gauges.

The third experiment was to make comparisons between CP sets within and between 3 regions. This started by selecting 5 random subsets of 12 gauges from the historical datasets in each of the regions. CPs were conditioned on each of the 5 subsets in each region independently. We then set out to determine whether the similarity of the sets within each region was materially greater than between regions. Because these are categorical data it does not make sense to compute correlation coefficients

A.3 Classification for Precipitation

Table 1. CP similarities for different classifications using the Pearson coefficient in the regions numbered 6,10 and 22 in Fig. 1

| Region | 6 | 10 | 22 |
|--------|-------|-------|-------|
| 6 | 0.699 | 0.617 | 0.640 |
| 10 | 0.617 | 0.674 | 0.592 |
| 22 | 0.640 | 0.592 | 0.698 |

Table 2. CP similarities for different classifications using the Cramer coefficient in the regions numbered 6, 10 and 22 in Fig. 1

| Region | 6 | 10 | 22 |
|--------|-------|-------|-------|
| 6 | 0.263 | 0.213 | 0.226 |
| 10 | 0.213 | 0.247 | 0.201 |
| 22 | 0.226 | 0.201 | 0.262 |

between sets, so Pearson and Cramer coefficients (equations A.9 and A.10) based on the χ^2 were the statistics used.

We chose three climate regions within South Africa to compare sets of CPs. The regions chosen and shown in Fig. 1 are: region 6 (KwaZulu-Natal), region 10 (Free State) and region 22 (Western Cape). We used 5 different random groupings of gauges for classifications in each region, thus for each region we derived 5 independent sets of CPs, making 15 in all.

The purpose of this calculation was to determine if the CPs derived for a given region had a higher inter-association than between regions. These coefficients were calculated for all pairs of classifications for the same regions. Then the averages of the association measures were calculated for classifications corresponding to the same (excluding the comparison of a classification with itself), and corresponding to different regions leading to 2 sets of 3 by 3 matrices.

The average of the 20 within-region statistics are compared with the averages of the 25 between-region statistics and appear in Tables A.1 and A.2 (the difference in the numbers in each group is because we omitted the diagonal elements of comparing a set with itself). The result is that there is a significant difference between the within against the between coefficients, supporting our exploitation of this CP selection procedure based on daily wetness.

All classifications show significant dependence, which of course is reasonable as the atmospheric conditions on the same region are classified in all cases. On the other

hand the within block average statistics are larger than the between block counterpart, inferring that there is a stronger within relationship than between blocks. The conclusion is that the choosing of CPs dependent on regions is valuable and worth the effort.

A.4 Classification for waves

Atmospheric circulations drive regional wave climates through atmosphere-ocean interactions. In particular they control the generation of the extreme wave events that cause severe coastal erosion. They are therefore also fundamental drivers of coastal vulnerability. The link between the wave climates and atmospheric circulation is complex. However, statistical models that link synoptic scale atmospheric circulation to regional wave characteristics have recently been shown to give significant insights (Pringle *et al.*, 2014). We propose that the classification of atmospheric drivers can improve coastal vulnerability assessments and the prediction of climate change effects. For example it provides a natural way to identify and isolate the effects of independent storm events, which is required for extremum analysis. For example Corbella & Stretch (2012b) identified independent events based on the autocorrelation or by using a 6-hour inter-arrival window. The transition of CPs between classes is a physically more meaningful method for defining independent events. Atmospheric CPs also contain important information regarding the distribution of wave height, direction and period, because when a particular CP type occurs the associated wave height, direction and period can be (statistically) predicted. Linking wave events to CPs can also be used to extend current data sets and to infill missing data (Hewitson & Crane, 2002). Finally the prediction and evaluation of climate change impacts on coastal vulnerability would be more robust if linked to changes in the atmospheric circulation patterns that are the basic drivers of wave climates and extreme wave events.

A.4.1 Case study description

The KwaZulu-Natal (KZN) coastline (Fig. A.4) is associated with a high energy wave climate. A number of weather types have been cited as the drivers of this wave climate. For example tropical cyclones, mid-latitude (extra-tropical) cyclones and cut off lows (Mather & Stretch (2012); Corbella & Stretch (2012d); Rossouw *et al.* (2011)). The location and persistence of tropical cyclones (TC's) are believed to drive large wave

A.4 Classification for waves

events that cause severe beach erosion in KZN (Mather & Stretch (2012); Corbella & Stretch (2012d)). Cut-off lows are deep low pressure systems that are displaced from the normal path of west-east moving mid-latitude cyclones (Preston-Whyte & Tyson (1988)). They are caused by instabilities within the westerly zonal flow due to the high wind shear. Vortices can become cut-off and move equator-ward (Preston-Whyte & Tyson (1988)). These features lead to seasonality in the wave climate and the occurrence of storms in particular. On average autumn and winter (April to September) are associated with the largest wave energy, while summer (January to March) has the smallest (Corbella & Stretch (2012d)). The significant wave height (H_s) is the key variable of interest for coastal vulnerability applications. Our algorithm considers both the daily average significant wave height and the daily maximum significant wave height. Wave data for the period 1992 to 2009 were obtained from wave buoys located near Durban and Richards Bay on the KwaZulu-Natal coastline (Fig. A.4).

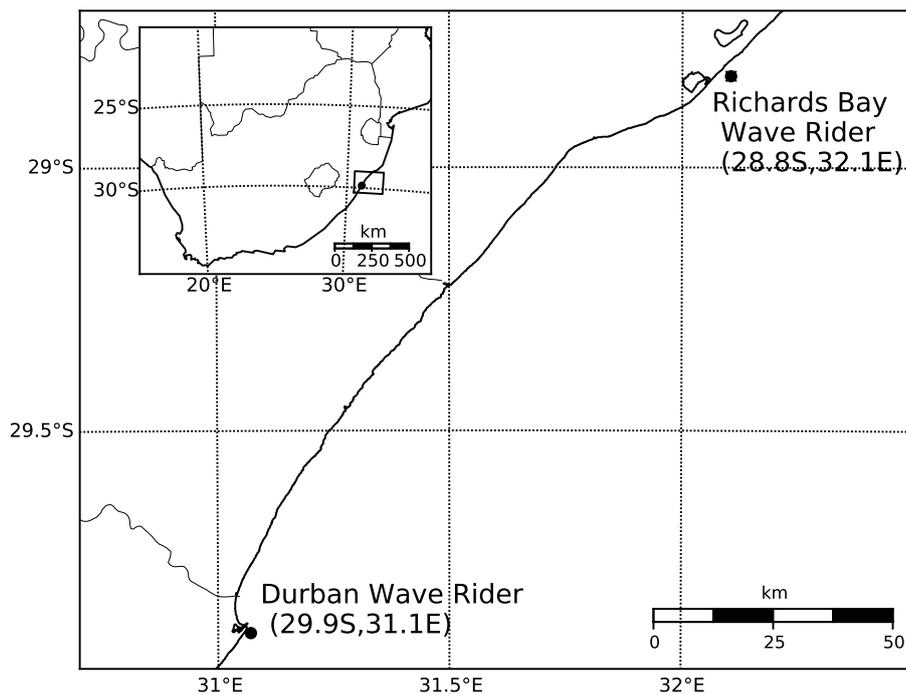


Fig. A.4 Locations of the wave observation buoys at Durban and Richards Bay, along the KwaZulu Natal coastline.

In this application the goal of the classification is to obtain a set of CP classes which explain extreme wave events. Wave heights larger than 3.5m have been shown to cause significant erosion along the KwaZulu-Natal coastline (Corbella & Stretch

Circulation Patterns identified by spatial rainfall and ocean wave fields in Southern Africa

(2013)). The objective functions (Equations A.4 and A.5) were used as performance measures.

The wave height threshold θ can be exploited to incorporate various scenarios. Two different thresholds were used. The first relates to the occurrence of extreme events ($\theta_1 = 3.5\text{m}$), while the second relates to midrange wave heights ($\theta_2 = 2.5\text{m}$). The second objective function type is also used to provide a more detailed classification.

To account for the persistence of CPs during extreme events Eq. A.4 was modified to include storm durations, defined as the duration of wave height excursions above 3.5m.

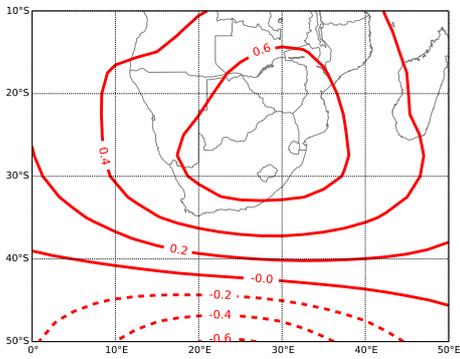
A.4.2 Dominant CP Classes

Fig. A.5 shows the average anomaly patterns for all the CP classes. CP99 refers to an unclassified class. Useful statistical parameters for each CP class are their frequency of occurrence, their contribution to extreme events, and the average and maximum significant wave heights (H_s) associated with them. These parameters are shown in Table A.3.

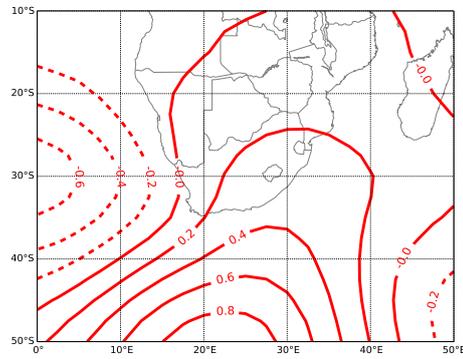
The results reflect a number of trends in CP-wave generation. However only the two most significant trends are discussed herein. Firstly CP01 and CP02 (Fig. A.5a) occur most frequently (about 17% of the time). Both CP classes resemble a mid-latitude cyclone in its different stages of development. The CP01 resembles the central low pressure region of a mid-latitude cyclone as it moves from west to east south of the country, while CP02 resembles the high-pressure region that follows. Secondly, Table A.4 shows a class (CP03) that comprises 60% of all extreme wave events. The CP03 is shown to contribute significantly to extreme events all year round with highest contribution in winter (65%). However CP03 (Fig. A.5c) occurs infrequently (9% of the time). Its occurrence is associated with large average and maximum significant wave heights ranging from 2.4 to 3.0 m and 5.0 to 8.5 m respectively. CP05 and CP06 (Fig. A.5e & f) contribute approximately to 30% of extreme events in spring and summer respectively. The CP06 resembles a pattern similar to tropical cyclones south of Madagascar. Tropical cyclones located within this region have been cited to drive large swells toward the KwaZulu-Natal Coast (Mather & Stretch, 2012). The CP05 resembles low-pressure systems over the interior.

The algorithm only considered CPs associated with extreme events at the time the event was recorded. In other words no time lags were considered when deriving the CP classes. Therefore the algorithm assumes that extreme events are driven by

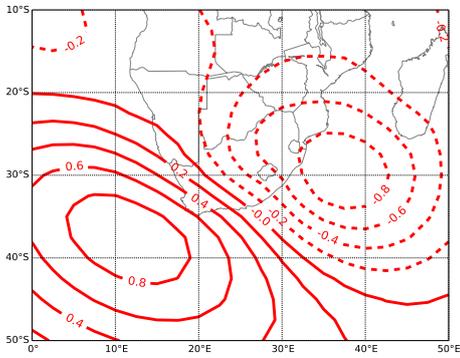
relatively stationary CPs.



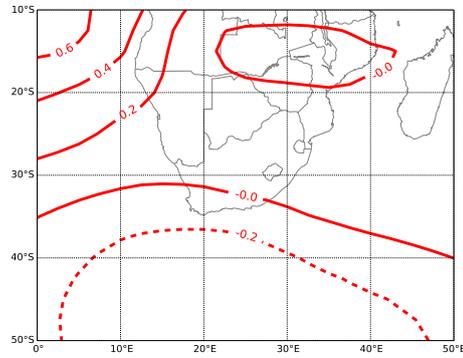
(a) CP01



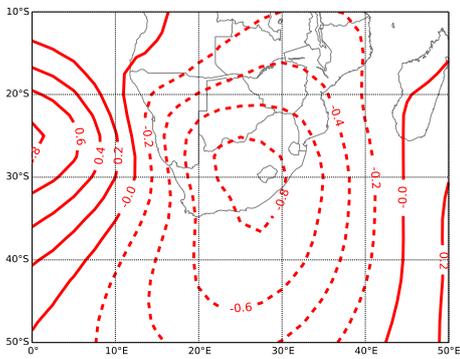
(b) CP02



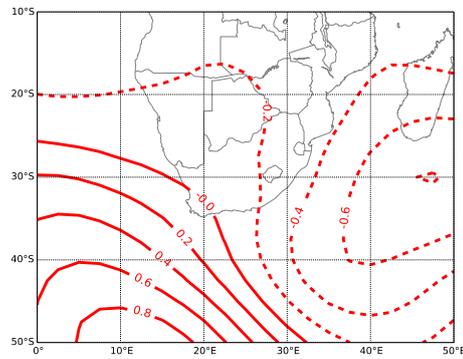
(c) CP03



(d) CP04



(e) CP05



(f) CP06

Circulation Patterns identified by spatial rainfall and ocean wave fields in Southern Africa

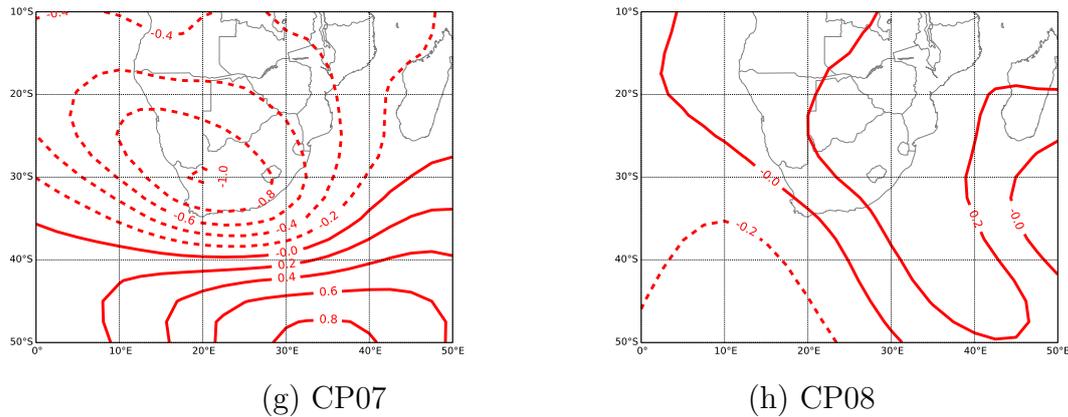


Fig. A.5 Average anomaly patterns for 8 CP classes derived from the regional wave climate data. Negative (low) pressure anomaly contours are shown as solid lines while positive pressure contours are dashed.

A.4.3 CP Variability

The degree of fit (DOF) describes the membership of a CP for given day to a particular class. The larger the degree of fit the stronger the belief that the CP belongs to a particular class. Fig. A.6 shows the average anomaly pattern for CP03 and the CPs associated with both the strongest and weakest membership for that class. CP03 is associated with a strong low pressure region east/south-east of South Africa. The pattern also shows a strong high-pressure region to the southwest. The coupling between the strong low and high pressure drive strong winds and subsequently large waves towards the coastline. The CP with the weakest membership to CP03 is shown to be a weak anomaly pattern (refer to Fig. A.6(c)).

A.4.4 Variability Within Classes

It is expected that in regions of high pressure pattern variation should be low. This is attributed to the stability of high pressures in their positions. In contrast pattern variation in regions of low pressure should be higher primarily due to the movement of low pressure systems. This is reflected in Fig. A.6(d), which shows significantly larger variation (standard deviation of 1) in the vicinity of the low-pressure region. The magnitude of the variation also indicates that CPs driving extreme events are associated with strong low pressures.

A.4 Classification for waves

Table A.3 CP Occurrence frequencies and wave height statistics associated with each CP class.

| STATISTICS | CP01 | CP02 | CP03 | CP04 | CP05 | CP06 | CP07 | CP08 | CP99 ^a |
|---|------|------|------|------|------|------|------|------|-------------------|
| Occurrence frequency ($p(CP)\%$) | | | | | | | | | |
| Summer | 18 | 18 | 8.0 | 13 | 7.5 | 8.1 | 5.6 | 15 | 8.3 |
| Autumn | 18 | 19 | 8.0 | 11 | 10 | 7.2 | 5.1 | 13 | 8.8 |
| Winter | 16 | 17 | 8.1 | 12 | 11 | 8.4 | 4.5 | 14 | 8.9 |
| Spring | 17 | 16 | 9.1 | 12 | 9.4 | 7.7 | 5.0 | 15 | 9.2 |
| All Seasons | 17 | 17 | 8.3 | 12 | 9.6 | 7.8 | 5.1 | 14 | 8.8 |
| Threshold exceedance for a given CP ($p(H_s \geq \theta CP)\%$) | | | | | | | | | |
| Summer | – | 0.4 | 8.0 | 0.3 | – | 3.5 | 2.6 | 0.7 | – |
| Autumn | 1.2 | 1.5 | 12 | 1.6 | 2.1 | 2.0 | 5.6 | 0.6 | 5.0 |
| Winter | 0.9 | 0.8 | 14 | – | 0.9 | 0.4 | 2.3 | 0.8 | 0.4 |
| Spring | 0.3 | – | 4.6 | 0.6 | 2.2 | – | 0.7 | – | – |
| All Seasons | 0.6 | 0.7 | 9.6 | 0.6 | 1.4 | 1.5 | 2.8 | 0.5 | 1.4 |
| Exceedance contribution ($p(CP H_s \geq \theta)\%$) | | | | | | | | | |
| Summer | – | 5.6 | 50 | 2.8 | – | 22 | 11 | 8.3 | – |
| Autumn | 7.7 | 10 | 33 | 6.4 | 7.7 | 5.1 | 10 | 2.6 | 17 |
| Winter | 7.5 | 7.5 | 64 | – | 5.7 | 1.9 | 5.7 | 5.7 | 1.9 |
| Spring | 4.5 | – | 55 | 9.1 | 27 | – | 4.5 | – | – |
| All Seasons | 5.8 | 7.4 | 48 | 4.2 | 7.9 | 6.9 | 8.5 | 4.2 | 7.4 |
| Average H_s (m) for each CP | | | | | | | | | |
| Summer | 1.8 | 1.9 | 2.5 | 1.8 | 1.8 | 2.2 | 2.2 | 1.9 | 1.9 |
| Autumn | 1.8 | 1.9 | 2.7 | 1.9 | 2.0 | 2.0 | 2.1 | 1.9 | 2.1 |
| Winter | 2.0 | 2.0 | 2.9 | 1.9 | 2.1 | 2.0 | 2.2 | 2.0 | 1.9 |
| Spring | 2.0 | 1.9 | 2.4 | 1.9 | 2.2 | 2.0 | 2.2 | 2.0 | 2.0 |
| All Seasons | 1.9 | 1.9 | 2.6 | 1.9 | 2.1 | 2.1 | 2.2 | 2.0 | 2.0 |
| Standard deviation of H_s (m) for each CP | | | | | | | | | |
| Summer | 0.48 | 0.49 | 1.1 | 0.49 | 0.53 | 0.76 | 0.74 | 0.61 | 0.49 |
| Autumn | 0.58 | 0.66 | 1.0 | 0.70 | 0.76 | 0.66 | 0.90 | 0.55 | 1.0 |
| Winter | 0.58 | 0.61 | 0.94 | 0.55 | 0.66 | 0.58 | 0.66 | 0.55 | 0.67 |
| Spring | 0.51 | 0.49 | 0.84 | 0.52 | 0.71 | 0.50 | 0.61 | 0.50 | 0.49 |
| All Seasons | 0.54 | 0.57 | 1.0 | 0.56 | 0.69 | 0.63 | 0.74 | 0.56 | 0.70 |
| Max H_s (m) for each CP | | | | | | | | | |
| Summer | 3.4 | 4.0 | 8.5 | 3.7 | 3.4 | 5.0 | 5.2 | 5.6 | 3.3 |
| Autumn | 4.0 | 5.5 | 5.7 | 5.5 | 6.3 | 4.3 | 5.1 | 4.0 | 5.4 |
| Winter | 4.2 | 3.8 | 5.6 | 3.4 | 3.8 | 3.5 | 4.3 | 4.8 | 3.6 |
| Spring | 3.9 | 3.3 | 5.3 | 4.5 | 5.4 | 3.4 | 3.7 | 3.5 | 3.3 |
| All Seasons | 4.2 | 5.5 | 8.5 | 5.5 | 6.3 | 5.0 | 5.2 | 5.6 | 5.4 |

^aCP99 is the unclassified class. Blank entries imply zero occurrences in the data set

Circulation Patterns identified by spatial rainfall and ocean wave fields in Southern Africa

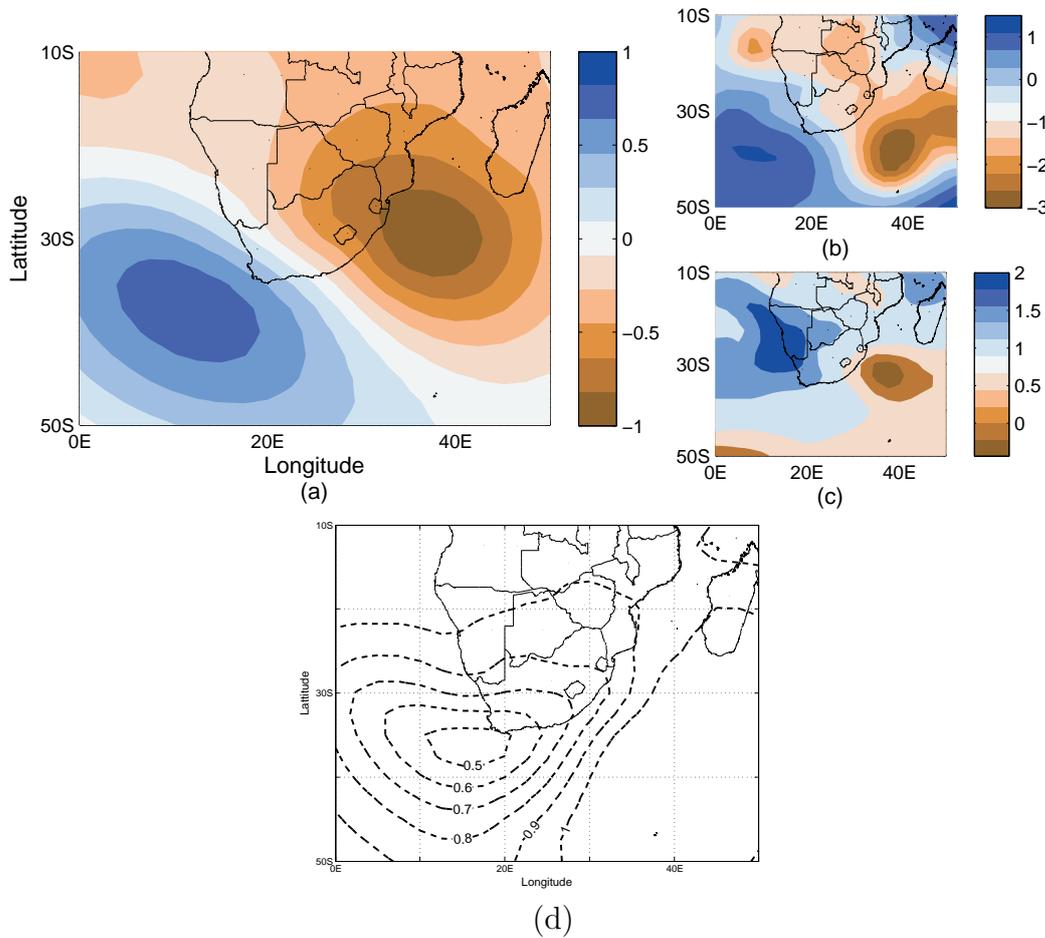


Fig. A.6 Average anomaly pattern for CP03 (a) showing (b) the anomaly with highest DOF, (c) the anomaly with lowest DOF value, and (d) the standard deviation for all CP03 anomalies.

A.4.5 CP rules and extreme events

Daily CP realisations associated with extreme wave events ($H_s \geq 3.5\text{m}$) were compared to the average class pattern to which they belong. Fig A.7 shows the average pattern for CP03 together with selected extreme events corresponding to CP03. The locations of the peak negative anomalies are shown in the plot. Significant pattern variability within the class is apparent. Fig. A.8 and Table A.4 show the CPs associated with six of the largest significant wave height events. The majority of the six events have been classified as belonging to CP03. Fig. A.8(f) from visual inspection shows a pattern similar to CP04 and CP08. Both classes are associated with low-pressure regions southeast of Madagascar. However the CP has been classified as belonging to class

Table A.4 Six of the most extreme wave events on record and their associated CPs for the period 1992 to 2009.

| Fig | Date | CP | H_S (m) |
|-----|------------|------|-----------|
| (a) | 19/03/2007 | CP03 | 8.50 |
| (b) | 05/05/2001 | CP05 | 6.30 |
| (c) | 18/03/2001 | CP03 | 5.92 |
| (d) | 03/04/2001 | CP03 | 5.66 |
| (e) | 23/09/1993 | CP03 | 5.64 |
| (f) | 19/03/2001 | CP08 | 5.63 |

CP08 and not CP04. From a visual account it appears to better resemble class CP04. Fig. A.8(a) & (c) are the CPs associated with the March 2007 storm which caused severe coastal erosion along the KwaZulu-Natal coastline (Mather & Stretch (2012); Corbella & Stretch (2012c)) with significant wave heights reaching 8.5m.

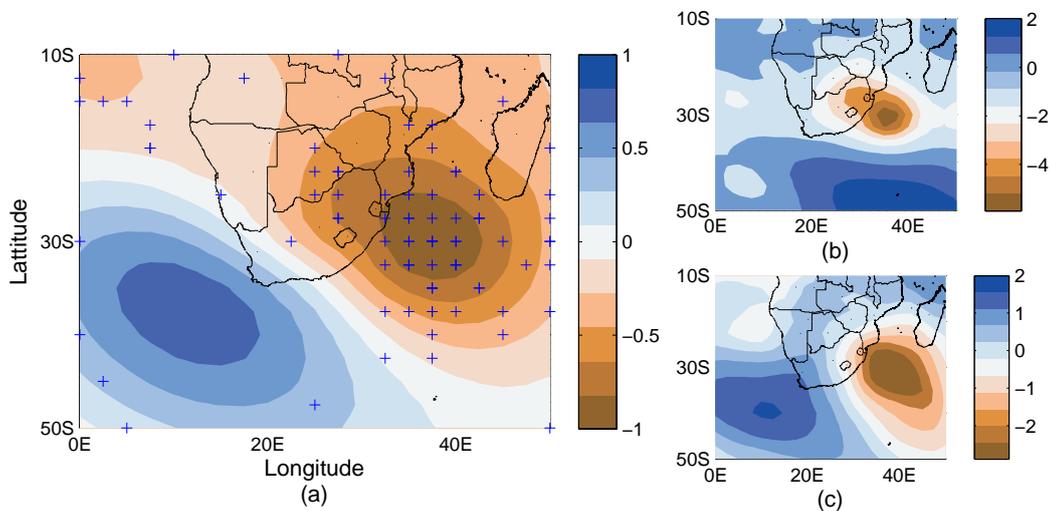


Fig. A.7 (a) Average anomaly pattern CP03 with (+) symbols indicating the centers of all negative anomalies (low pressures) contributing to the class. (b) & (c) show actual CP's for the dates 19/03/2007 and 30/08/2006 respectively, both of which were classified as members of the CP03 class.

Circulation Patterns identified by spatial rainfall and ocean wave fields in Southern Africa

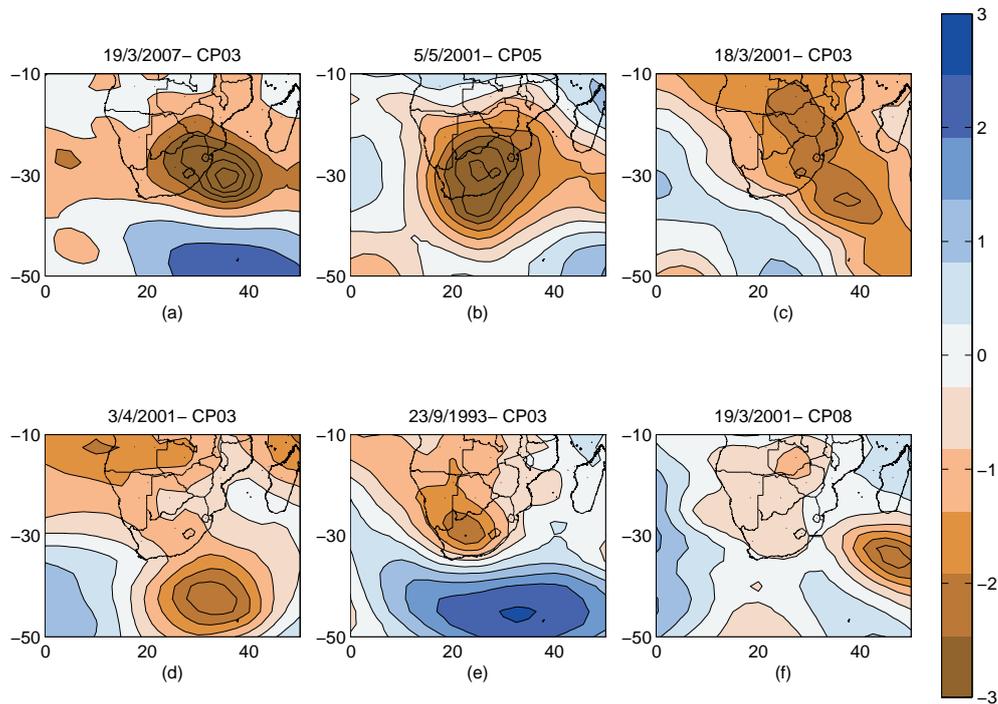


Fig. A.8 CP's associated with the six largest significant wave heights for the dates (a) 19/3/2007, (b) 5/5/2001, (c) 18/3/2001, (d) 3/4/2001, (e) 23/9/1993 and (f) 19/3/2001.

A.5 Discussion and conclusions

In this paper different problems related to the fuzzy rule based classification of circulation patterns were discussed. It was shown that reasonable classifications can be developed for precipitation and waves in South Africa. We did this by selecting long sequences of Circulation Patterns in the region surrounding Southern Africa and conditioning the selection of the CPs on selected variables germane to the task; daily wetness for rainfall and wave height frequencies of ocean waves. The classifications yield classes with significantly different behavior depending on the classification goal. It was shown in the precipitation part of the study, that using different stations of a climatic region for the classification objective leads to different classifications, but the performance of these classifications is nevertheless very similar. Different classifications for the same region are more similar than classifications corresponding to different regions. This justifies the use of different classifications for the different regions and variables, because they add discriminatory power to the modelling procedure.

In the case of the wave study, the emphasis was on the statistical link between

atmospheric CPs and extreme wave events. The results show that the classification algorithm is able to identify CPs that drive extreme wave events along the KwaZulu-Natal coastline. The most frequent CPs associated with wave generation are revealed as low and high pressure anomalies south of the country. These reflect easterly traversing mid-latitude cyclones and their associated high pressures anomalies. The CP class that drives the majority of extreme wave events (labeled CP03) is associated with strong low pressure anomalies east of the country. A high/low pressure coupling drives strong winds towards the coastline and generates large waves. It is not clear what weather regimes are associated with these events since both mid-latitude cyclones and cut-off lows may belong to the CP03 class depending on their location. The classification algorithm does not currently identify CP anomalies as part of an overall structure developing in space and time: each class could be a snapshot in the temporal evolution of (perhaps distinct) CPs. Therefore a particular weather system may be associated with a number of classes during its development, which makes for added complexity and requires further research.

There are similarities in the CP classes driven by the two surface variables because in some instances they are linked to similar synoptic weather systems. However there are also significant differences in the details of the pressure fields. These differences are expected because large ocean waves are typically generated over fetches of the order of thousands of kilometres far off shore, whereas some rainfall is generated by local atmospheric variables including temperature, humidity, wind speed and radiation over the area of concern.

Appendix B

Additional Wave Height and Circulation Pattern Statistics

B.1 Introduction

Of particular interest to this study is the effect of CP characteristics on wave development. These characteristics include CP persistence (or lack thereof) in a certain region, the movement of the CP before, during and after the storm event and the orientation of high and low pressure regions. This study has focussed predominantly on the latter as it is well suited for classification. However the translation of a CP across southern Africa is also of significant importance. This is not explicitly described with the classification technique used within this thesis.

The movement of CPs can be measured in a number of different ways. We propose two different approaches. Firstly the Pearson's correlation coefficient as a measure of persistence (or stability) and secondly the temporally averaged CP as a way to describe the CP movement. The persistence of a CP in a particular region can simply be defined as the time during which the Pearson's correlation coefficient does not vary significantly. However a rapid change in the Pearson's correlation coefficient implies a change in CP regime. The correlation coefficient also provides insight into changes within the circulation regime. If a CP is persistent in a region that is favourable for wave development relative to the east coast it is expected that it will drive large waves towards the coastline. This is because the CP has had sufficient time to develop the sea state through wind-wave interaction culminating in a large wave event. Furthermore the time it takes for the storm peak to reach the coastline in relation to the time the CP has remained in a certain region is also important. This is a function of wave

physics (e.g. wave dispersion) but primarily it is due to the CP characteristics.

This appendix describes CP persistence and movement characteristics. Wave height and pressure data for the year 2007 were used as an example to demonstrate the CP characteristics associated with wave development. Three extreme wave events occurred during the year with the largest event on record occurring towards the end of March 2007.

B.2 Method

The Pearson's correlation coefficient was used to measure change in circulation regime as well as CP persistence. The CPs are defined on a region with resolution 2.5° ($10^\circ\text{S } 0^\circ\text{E} - 50^\circ\text{S } 50^\circ\text{E}$). The CP at time t_i was compared to the CP at time t_{i+1} such that

$$\rho = \frac{\overline{\mathbf{x} \cdot \mathbf{y}} - \mu_x \mu_y}{\sigma_x \sigma_y} \quad (\text{B.1})$$

where μ_x, σ_x and μ_y, σ_y are the mean and standard deviations for the CPs at time t_i and t_{i+1} respectively, calculated as

$$\mu = \frac{1}{G} \sum_{i=1}^G a_i, \quad (\text{B.2})$$

$$\sigma = \sqrt{\frac{1}{G} \sum_{i=1}^G (a_i - \mu)^2} \quad (\text{B.3})$$

where G is the total number of grid points and a_i is the anomaly value at the i^{th} grid point. The term $\overline{\mathbf{x} \cdot \mathbf{y}}$ is the mean of the dot product between the anomaly values for CPs at t_i and t_{i+1} . Therefore using Equation B.1 it is possible to describe persistent periods as well as changes to the circulation regime. Persistent periods are defined as periods with high and consistent ρ values.

Temporally averaged CPs provide insight into the movement of CPs. The shape of the average CP describes its translation across southern Africa. Temporal averages were calculated during storm events as well as during periods of persistence.

B.3 Results

Pearson's correlation coefficient values between CPs at consecutive time steps were calculated for the year 2007 and are shown in Fig. B.1. The Fig. reveals two

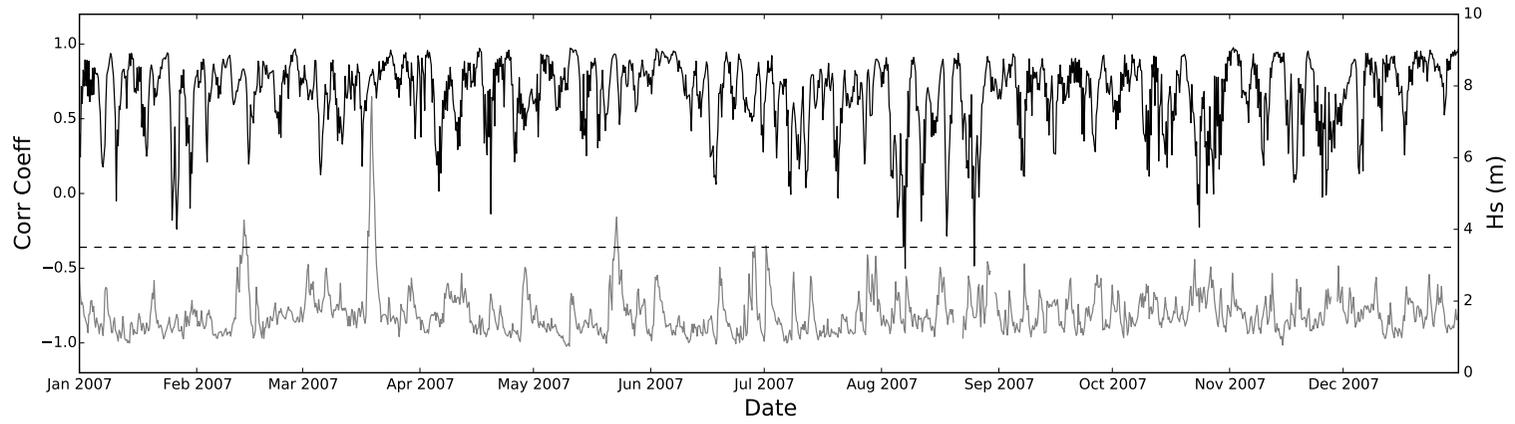


Fig. B.1 CP correlation coefficients (black) and wave heights (grey) for the year 2007. The correlation coefficients were calculated between CPs of consecutive 6-hourly time steps.

interesting features. Firstly there are "stable" periods where the CPs are well correlated followed by rapid changes. The rapid changes indicate a change in CP state. The change can be attributed to a break down of the circulation pattern structure (reducing values) or the rapid translation of a low pressure structure across southern Africa. Increasing correlation values indicate the transition of a CP into a stable state.

The periods associated with CP stability are also associated with large wave events. Figs. B.2, B.3 and B.4 show the relationship between the correlation coefficients and wave heights as well as the movement of the CPs within these periods. The CPs were temporally averaged for both the stable period and during the storm event ($H_s \geq 3.5m$). The average CP during the February and March storm events show a persistent low pressure region east of the coastline which suggests this storm event is associated with a CP that remains stationary over the region. However the storm event occurring in May (Fig B.4) is associated with a CP that translates in a south-easterly direction. This is indicated in the elliptical shape of the low pressure region. It is interesting to note the location of the storm peak (maximum wave height) within the stable period. The peak can occur near the end of the period as in Figs. B.2 and B.4 or in the middle such as Fig. B.3.

Additional Wave Height and Circulation Pattern Statistics

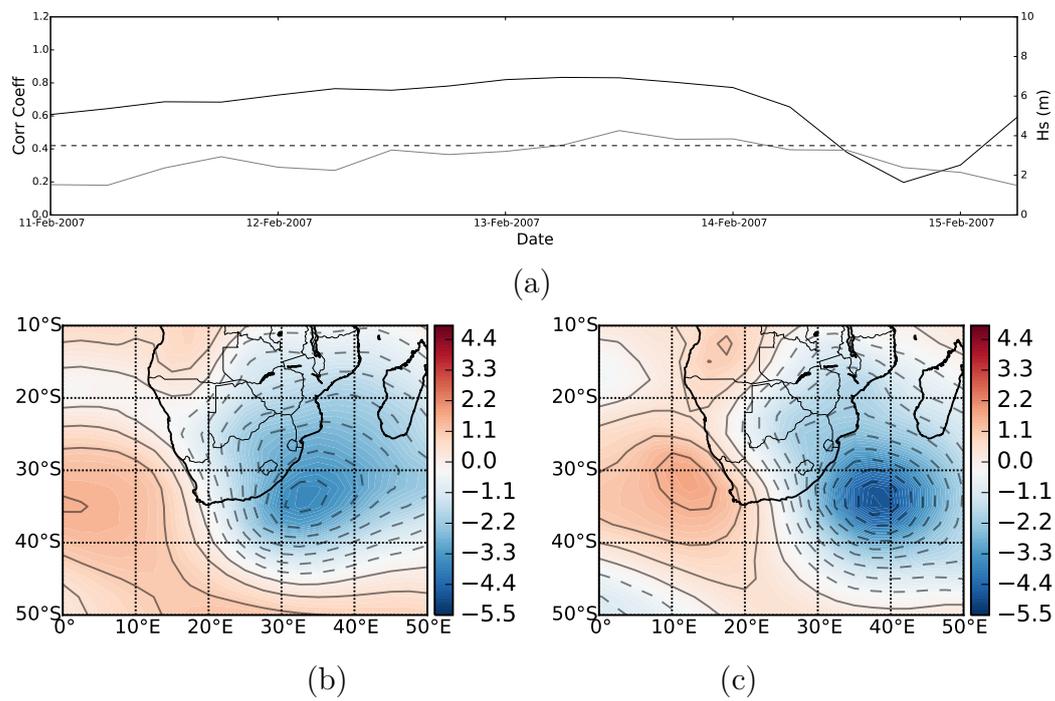


Fig. B.2 (a) Correlation coefficients (black) and wave heights (grey) for the storm event occurring during the month of February 2007. (b) and (c) are the CP averaged values for the persistence period and storm period respectively..

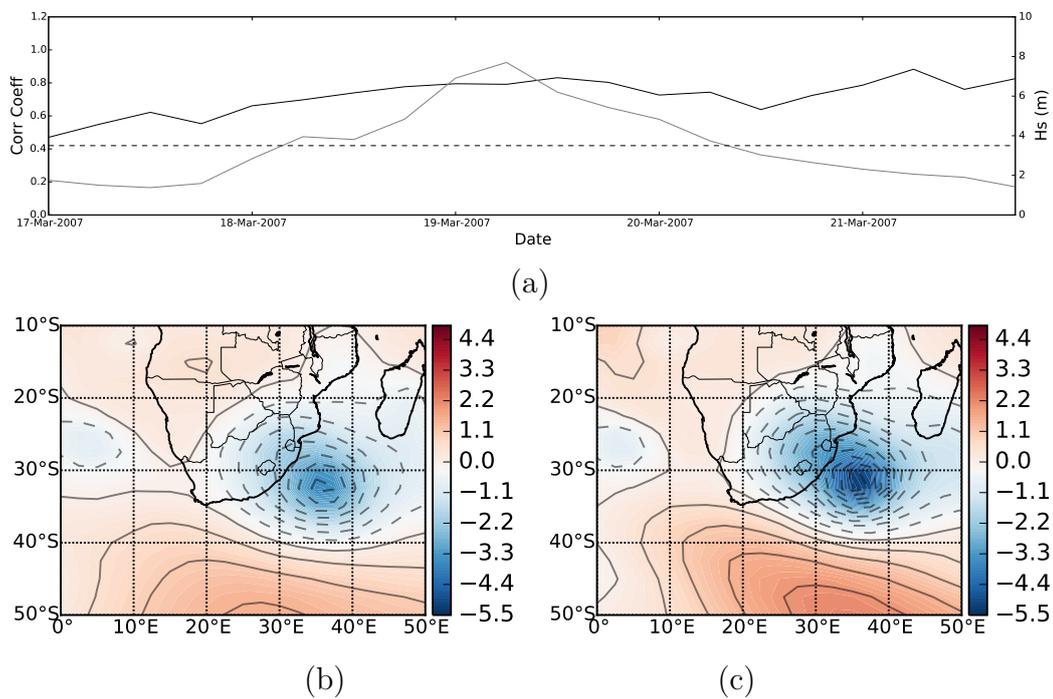
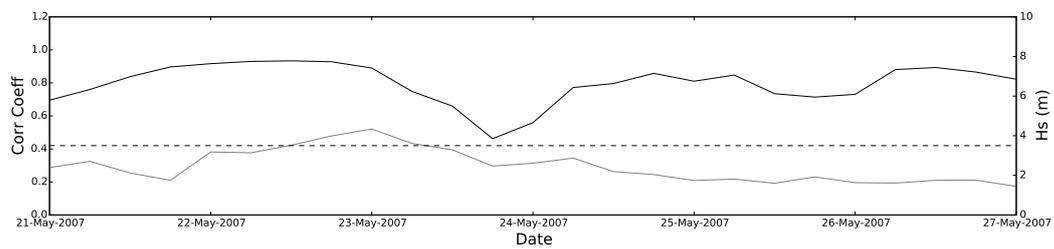
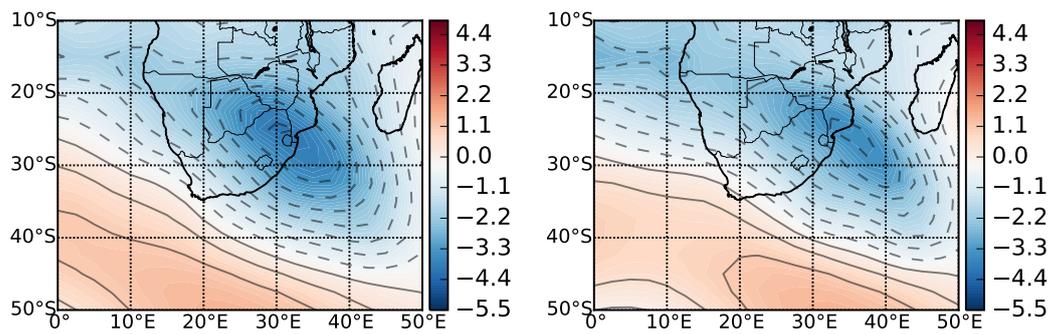


Fig. B.3 (a) Correlation coefficients (black) and wave heights (grey) for the storm event occurring during the month of March 2007. (b) and (c) are the CP averaged values for the persistence period and storm period respectively

Additional Wave Height and Circulation Pattern Statistics



(a)



(b)

(c)

Fig. B.4 (a) Correlation coefficients (black) and wave heights (grey) for the storm event occurring during the month of May 2007. (b) and (c) are the CP averaged values for the persistence period and storm period respectively

B.4 Discussion

The movement of CPs has a strong influence on wave development especially large wave events. The Pearson's correlation coefficient is a simple tool that provides useful insight into these processes. Extreme wave events are strongly associated with CP persistence and storm peaks can either occur towards the middle of the stable period or towards the end. The arrival of the storm peak relative to time the CP has remained persistent in a certain region is a function of the distance of the CP relative to the coastline. Furthermore the more persistent the CP the larger the associated wave heights. This is attributed to the duration the CP has had to drive a strong wind towards the coastline. Furthermore the longer the distance that the wind has blown over the sea (or fetch), the larger the waves.

The correlation coefficients between consecutive CPs reveals useful insight into the movement and transitions of CPs between states. The temporal sequence of correlation coefficients for the year 2007 reveal a cyclic fluctuation between stable/correlated periods and uncorrelated periods. The uncorrelated periods indicate changes in the CP state indicative of the movement of low pressure regions across southern Africa. However the correlation coefficients do not provide information on the shape and direction a particular CP moves. Therefore they are only useful as a primary evaluation tool. The temporal averaged CPs provide more insight into the direction the CP is moving and has the potential to improve current classification techniques.

The use of both the correlation coefficients and the temporally averaged CP can be used to improve classification as follows: The correlation coefficients define the periods in which to temporally average CPs. Ideally this is done during the more correlated periods which are defined as the periods between the very low values (or troughs) in the temporal sequence of correlation coefficients. Classification can then be applied to the set of averaged CPs.

Appendix C

Classification Algorithm Pseudocode

This appendix presents the pseudocode for both the classification and optimization algorithms defined in this study.

The optimization algorithm includes the classification algorithm. Therefore to simplify this section only the optimization algorithm will be explained in detail.

C.1 Optimization Algorithm

The optimization algorithm consists of three main subroutines: (a) an *Initialization* function that reads in the pressure and wave data, (b) *Classification* in which CPs are assigned to classes and (c) *Optimization* in which the classes are derived. Procedures of strategic importance are described as follows:

Classification Algorithm Pseudocode

Procedure 1 Initialization

Reads in pressure data, wave data and initial random definitions of CP classes.
Converts pressure values to anomalies.

Input: Pressure data file,

wave data file,

CP class file,

Files containing average and standard deviations of pressure values,

number of grid points in the y-direction \mathbf{Y} ,

number of grid points in the x-direction \mathbf{X} ,

number of CP classes (\mathbf{n}),

length of record \mathbf{T} .

Output: Anomaly data matrix ($\mathbf{P} \in \mathbb{R}^{\mathbf{T} \times \mathbf{Y} \times \mathbf{X}}$),

Wave height vector ($\mathbf{W} \in \mathbb{R}^{\mathbf{T}}$),

CP Class matrix ($\mathbf{C} \in \mathbb{R}^{\mathbf{n} \times \mathbf{Y} \times \mathbf{X}}$).

Variables:

$\mathbf{C} \in \mathbb{R}^{\mathbf{n} \times \mathbf{Y} \times \mathbf{X}}$

$\mathbf{P} \in \mathbb{R}^{\mathbf{T} \times \mathbf{Y} \times \mathbf{X}}$

$\mathbf{W} \in \mathbb{R}^{\mathbf{T}}$

$\mathbf{AVG} \in \mathbb{R}^{J \times \mathbf{Y} \times \mathbf{X}}$

$\mathbf{STD} \in \mathbb{R}^{J \times \mathbf{Y} \times \mathbf{X}}$

► J is the length of the year

$\mathbf{BA} \in \mathbb{R}^{\mathbf{Y} \times \mathbf{X}}$

$\mathbf{BS} \in \mathbb{R}^{\mathbf{Y} \times \mathbf{X}}$

► Temp vectors of length $\mathbf{Y} * \mathbf{X}$ to store values.

OPEN(File : average pressure data file as \mathbf{A})

OPEN(File : standard deviation pressure data file as \mathbf{S})

for $i = 1 \rightarrow J$ **do** ► Read and store average and standard deviations of pressure data

dummy \leftarrow **READ**(\mathbf{A})

► Dummy variable of time step in file

dummy \leftarrow **READ**(\mathbf{S})

$(\mathbf{BA}(k), k = 1 \rightarrow \mathbf{Y} * \mathbf{X}) \leftarrow$ **READ**(\mathbf{A} , *float*)

$(\mathbf{BS}(k), k = 1 \rightarrow \mathbf{Y} * \mathbf{X}) \leftarrow$ **READ**(\mathbf{S} , *float*)

$LL \leftarrow 1$

for $k = 1 \rightarrow \mathbf{Y}$ **do**

for $l = 1 \rightarrow \mathbf{X}$ **do**

$\mathbf{AVG}(k, l) \leftarrow \mathbf{BA}(LL)$

$\mathbf{STD}(k, l) \leftarrow \mathbf{BS}(LL)$

$LL \leftarrow LL + 1$

end for

end for

end for

Procedure 1 Initialization (continued)

OPEN(File : CP classes file as **CP**)
for $i = 1 \rightarrow n$ **do** ▶ Read and store CP classes
 dummy \leftarrow **READ**(**CP**) ▶ Dummy variable of class number in file
 (**BA**(k), $k = 1 \rightarrow Y * X$) \leftarrow **READ**(**CP**, *int*)
 LL \leftarrow 1
 for $k = 1 \rightarrow Y$ **do**
 for $l = 1 \rightarrow X$ **do**
 C(k, l) \leftarrow **BA**(*LL*)
 LL \leftarrow *LL* + 1
 end for
 end for
end for
OPEN(File : Pressure data file as **Pr**)
OPEN(File : Wave data file as **Wa**)
for $i = 1 \rightarrow T$ **do** ▶ Read and store pressure and wave data
 dummy \leftarrow **READ**(**Pr**) ▶ Dummy variable of time step in file
 (**BA**(k), $k = 1 \rightarrow Y * X$) \leftarrow **READ**(**Pr**, *float*)
 W \leftarrow **READ**(**Wa**, *float*)
 LL \leftarrow 1
 for $k = 1 \rightarrow Y$ **do**
 for $l = 1 \rightarrow X$ **do**
 P(k, l) \leftarrow **BA**(*LL*)
 LL \leftarrow *LL* + 1
 end for
 end for
end for
Reading data complete now convert pressure data to anomalies
for $i = 1 \rightarrow T$ **do**
 J \leftarrow **JULIANDATE**(i) ▶ Convert loop variable i to a Julian date
 for $k = 1 \rightarrow Y$ **do**
 for $l = 1 \rightarrow X$ **do**
 P(i, k, l) \leftarrow $\frac{\mathbf{P}(i, k, l) - \mathbf{AVG}(J, k, l)}{\mathbf{STD}(J, k, l)}$
 end for
 end for
end for

Classification Algorithm Pseudocode

Procedure 2 Classification

Input: number of CP classes (\mathbf{n}),

Anomaly data matrix ($\mathbf{P} \in \mathbb{R}^{\mathbf{T} \times \mathbf{Y} \times \mathbf{X}}$),

CP Class matrix ($\mathbf{C} \in \mathbb{R}^{\mathbf{n} \times \mathbf{Y} \times \mathbf{X}}$),

► \mathbf{C} contains the definitions of the CP classes derived in the optimization

number of grid points in the y-direction \mathbf{Y} ,

number of grid points in the x-direction \mathbf{X} ,

length of record \mathbf{T}

Output: A vector $\mathbf{c} \in \mathbb{R}^{\mathbf{T}}$ containing CP class labels for all time steps in \mathbf{T} .

Variables:

$\mathbf{pex} \in \mathbb{R}^4$

► user defined exponent values for fuzzy numbers $1, \dots, 4$

$\mathbf{N} \in \mathbb{R}^4$

► vector to store counts of different fuzzy numbers $1, \dots, 4$

$\mu \in \mathbb{R}^4$

► vector to store membership grades

Set all values of \mathbf{N} and μ to 0

$\mathbf{DOF} = 1$

$\mathbf{Class} = 1$

$\mathbf{Max} = 0$

for $i=1 \rightarrow \mathbf{T}$ do

 for $j=1 \rightarrow \mathbf{n}$ do

 for $k = 1 \rightarrow \mathbf{Y}$ do

 for $l = 1 \rightarrow \mathbf{X}$ do

 Fuzzy Number $\leftarrow \mathbf{C}(j, k, l)$

 Anomaly $\leftarrow \mathbf{P}(i, k, l)$

 grade $\leftarrow \text{MEMBERSHIP}(\text{FuzzyNumber}, \text{Anomaly})$

 ► Calculate the membership grade for anomaly value based on the fuzzy number

$\mu(\text{FuzzyNumber}) \leftarrow \mu(\text{FuzzyNumber}) + \text{grade}^{\mathbf{pex}(\text{FuzzyNumber})}$

$\mathbf{N}(\text{FuzzyNumber}) \leftarrow \mathbf{N}(\text{FuzzyNumber}) + 1$

 end for

 end for

 for $m = 1: 4$ do

$\mathbf{DOF} \leftarrow \mathbf{DOF} \times \left(\frac{\mu(m)}{\mathbf{N}(m)} \right)^{1/\mathbf{pex}(m)}$

 end for

 if $\mathbf{DOF} > \mathbf{Max}$ then

 Class $\leftarrow i$

 Max $\leftarrow \mathbf{DOF}$

 end if

end for

$\mathbf{c}(\mathbf{T}) \leftarrow \mathbf{Class}$

Set all values of \mathbf{N} , μ and \mathbf{Max} to 0.

Set \mathbf{DOF} and $\mathbf{Class} = 1$

end for

Procedure 3 Membership

This procedure assigns a membership grade $\in [0, 1]$ to the anomaly at a grid point based on its value and the fuzzy number at the same grid point. It is a function that is called from the CLASSIFICATION function.

Input: Fuzzy Number (**f**) and the Anomaly value (**a**)

Output: Membership grade μ

function TRIANGULAR(**a**,**x**,**y**,**z**)

 ► Calculate the membership grade using a triangular function

$$\mu = \begin{cases} \frac{a-x}{y-x} & \text{if } x \leq a \leq y \\ \frac{a-z}{y-z} & \text{if } y < a \leq z \\ 0 & \text{else} \end{cases}$$

return μ

end function

if **f** == 1 **then**

$\mu \leftarrow \text{TRIANGULAR}(\mathbf{a}, -\infty, -3, 0)$

else if **f** == 2 **then**

$\mu \leftarrow \text{TRIANGULAR}(\mathbf{a}, -4, -0.85, 4)$

else if **f** == 3 **then**

$\mu \leftarrow \text{TRIANGULAR}(\mathbf{a}, 0.25, 0.85, 4)$

else if **f** == 4 **then**

$\mu \leftarrow \text{TRIANGULAR}(\mathbf{a}, 0, 3, +\infty)$

end if

Classification Algorithm Pseudocode

Procedure 4 Optimization

This optimization procedure uses simulated annealing to optimize the shape of the CP classes.

Input: Number of CPs \mathbf{n} ,

Wave vector $\mathbf{W} \in \mathbb{R}^{\mathbf{T}}$,

Anomaly matrix $\mathbf{P} \in \mathbb{R}^{\mathbf{T} \times \mathbf{Y} \times \mathbf{X}}$,

CP classes $\mathbf{C} \in \mathbb{R}^{\mathbf{n} \times \mathbf{Y} \times \mathbf{X}}$,

grid points in the y-direction \mathbf{Y} ,

grid points in the x-direction \mathbf{X} , Time period \mathbf{T} .

Output: An optimized CP class matrix $\mathbf{C} \in \mathbb{R}^{\mathbf{n} \times \mathbf{Y} \times \mathbf{X}}$.

$\mathbf{c} \leftarrow \text{CLASSIFICATION}(\mathbf{n}, \mathbf{P}, \mathbf{C}, \mathbf{Y}, \mathbf{X}, \mathbf{T})$ ▶ Perform initial classification

$\mathbf{Ob1}, \mathbf{Ob2}, \mathbf{Ob3} \leftarrow \text{CALCOBJECTIVEFUNCS}(\mathbf{c}, \mathbf{W})$

▶ Calculate initial objective functions

$\mathbf{Temp} \leftarrow 7.85 * \mathbf{n}$

▶ Randomly set initial annealing temperature

for $kt = 1 \rightarrow 150$ **do**

▶ Start annealing

for $i = 1 \rightarrow 5000$ **do**

$\mathbf{R} \leftarrow \text{RANDOM}(\mathbf{n})$

▶ Randomly select a CP class

$\mathbf{IX} \leftarrow \text{RANDOM}(\mathbf{X})$

▶ Randomly select x-location

$\mathbf{IY} \leftarrow \text{RANDOM}(\mathbf{Y})$

▶ Randomly select y-location

$\mathbf{C}(R, IY, IX) \leftarrow f \in [1, \dots, 4], f \neq \mathbf{C}(R, IY, IX)$

▶ Change fuzzy number at location (IX, IY) for class R

$\mathbf{c2} \leftarrow \text{CLASSIFICATION}(\mathbf{n}, \mathbf{P}, \mathbf{C}, \mathbf{Y}, \mathbf{X}, \mathbf{T})$

▶ Perform new classification with changed CP class

$\mathbf{Ub1}, \mathbf{Ub2}, \mathbf{Ub3} \leftarrow \text{CALCOBJECTIVEFUNCS}(\mathbf{c2}, \mathbf{W})$

$\mathbf{O} \leftarrow \alpha \mathbf{Ob1} + \beta \mathbf{Ob2} + \gamma \mathbf{Ob3}$

$\mathbf{U} \leftarrow \alpha \mathbf{Ub1} + \beta \mathbf{Ub2} + \gamma \mathbf{Ub3}$

▶ Calculate old and new objective functions. Weights are user defined.

$\mathbf{O} \leftarrow w - \mathbf{O}$

$\mathbf{U} \leftarrow w - \mathbf{U}$

▶ Change maximization scheme to minimizing

if $\mathbf{U} < \mathbf{O}$ **then**

Accept change

else

Accept change with probability $\exp\left(\frac{\mathbf{O} - \mathbf{U}}{\mathbf{Temp}}\right)$

end if

if Change i accepted **then**

$\mathbf{O1}, \mathbf{O2}, \mathbf{O3} \leftarrow \mathbf{U1}, \mathbf{U2}, \mathbf{U3}$

end if

Reduce annealing temperature.

end for

end for

Procedure 5 Calculate Objective Functions

This procedure calculates the objective functions required for the optimization. It is called in the optimization procedure.

Input: Classification $\mathbf{c} \in \mathbb{R}^{\mathbf{T}}$,
 Wave data $\mathbf{W} \in \mathbb{R}^{\mathbf{T}}$,
 Period \mathbf{T} .

Output: Objective functions **O1**, **O2** and **O3**.

Variables:

$\mathbf{h}_{\text{CP}} \in \mathbb{R}^{\mathbf{n}} \leftarrow 0$ ▶ Frequency of CP class
 $\mathbf{hs}_{\text{CP}} \in \mathbb{R}^{\mathbf{n}} \leftarrow 0$ ▶ Average wave height of CP class
 $\mathbf{hs} \leftarrow 0$ ▶ Average wave height
 $\mathbf{pr}_{3.5} \in \mathbb{R}^{\mathbf{n}} \leftarrow 0$ ▶ Probability of wave height exceeding 3.5 m for a given CP
 $\mathbf{pe}_{3.5} \leftarrow 0$ ▶ Probability of wave height exceeding 3.5 m
 $\mathbf{pr}_{2.5} \in \mathbb{R}^{\mathbf{n}} \leftarrow 0$ ▶ Probability of wave height exceeding 2.5 m for a given CP
 $\mathbf{pe}_{2.5} \leftarrow 0$ ▶ Probability of wave height exceeding 2.5 m
for $i = 1 \rightarrow \mathbf{T}$ **do**
 $cp \leftarrow \mathbf{c}(i)$
 $\mathbf{h}_{\text{CP}}(cp) \leftarrow \mathbf{h}_{\text{CP}}(cp) + 1$
 $\mathbf{hs}_{\text{CP}}(cp) \leftarrow \mathbf{hs}_{\text{CP}}(cp) + \mathbf{W}(i)$
 $\mathbf{hs} \leftarrow \mathbf{hs} + \mathbf{W}(i)$
 if $\mathbf{W}(i) \geq 2.5$ **then**
 $\mathbf{pr}_{2.5}(cp) \leftarrow \mathbf{pr}_{2.5}(cp) + 1$
 $\mathbf{pe}_{2.5} \leftarrow \mathbf{pe}_{2.5} + 1$
 end if
 if $\mathbf{W}(i) \geq 3.5$ **then**
 $\mathbf{pr}_{3.5}(cp) \leftarrow \mathbf{pr}_{3.5}(cp) + 1$
 $\mathbf{pe}_{3.5} \leftarrow \mathbf{pe}_{3.5} + 1$
 end if
end for
 $\mathbf{hs} \leftarrow \mathbf{hs}/\mathbf{T}$
 $\mathbf{pe}_{2.5} \leftarrow \mathbf{pe}_{2.5}/\mathbf{T}$
 $\mathbf{pe}_{3.5} \leftarrow \mathbf{pe}_{3.5}/\mathbf{T}$
for $i = 1 \rightarrow \mathbf{n}$ **do**
 $\mathbf{hs}_{\text{CP}}(i) \leftarrow \mathbf{hs}_{\text{CP}}(i)/\mathbf{h}_{\text{CP}}(i)$
 $\mathbf{O1} \leftarrow \mathbf{O1} + \mathbf{hs}_{\text{CP}}(i)/\mathbf{hs}$
 $\mathbf{pr}_{2.5}(i) \leftarrow \mathbf{pr}_{2.5}(i)/\mathbf{h}_{\text{CP}}(i)$
 $\mathbf{O2} \leftarrow \mathbf{O2} + (\mathbf{pr}_{2.5}(i) - \mathbf{pe}_{2.5})^2$
 $\mathbf{pr}_{3.5}(i) \leftarrow \mathbf{pr}_{3.5}(i)\mathbf{h}_{\text{CP}}(i)$
 $\mathbf{O3} \leftarrow \mathbf{O3} + (\mathbf{pr}_{3.5}(i) - \mathbf{pe}_{3.5})^2$
end for
