

Numerical and Statistical Modelling of Australian Severe Weather

Joshua Hartigan
Bachelor of Mathematics (Honours)

Thesis submitted in fulfilment of the requirements for the degree of
Doctor of Philosophy
Under the supervision of: James Brown, Lance M. Leslie, and Shev MacNamara

School of Mathematical and Physical Sciences,
Faculty of Science,
University of Technology Sydney

Submitted October 2022

CERTIFICATE OF ORIGINAL AUTHORSHIP

I, Joshua Hartigan declare that this thesis, is submitted in fulfilment of the requirements for the award of Doctor of Philosophy, in the School of Mathematical and Physical Sciences at the University of Technology Sydney.

This thesis is wholly my own work unless otherwise referenced or acknowledged. In addition, I certify that all information sources and literature used are indicated in the thesis.

This document has not been submitted for qualifications at any other academic institution.

This research is supported by the Australian Government Research Training Program.

Production Note:

Signature: Signature removed prior to publication.

Date: 30/05/2022

Acknowledgements

This research represents a challenging but rewarding four years where I have had the privilege to apply my background in mathematics to better understand various forms of severe weather, including thunderstorms, for which I have been passionate about for the majority of my life. There are numerous people who have helped make this dream a reality and provided valuable support along the way.

Firstly, I would like to thank my first summer research supervisor, Dr Judy-anne Osborn, who encouraged me to pursue my passion in severe weather and helped set up my honours research project in modelling how severe convection interacts with the sea-breeze air mass. I would also like to thank my honours supervisors, Joshua Soderholm and Robert Warren for taking me on despite the long distance, and supporting me throughout that project. I thank my piano teacher, Luba Totoeva, for her connection with Em/Prof. Alex Novikov, which led me to my PhD supervisors Prof. Lance Leslie and Dr Shev MacNamara. I also thank Luba for the countless hours teaching me piano, providing me a much needed break and emotional outlet throughout my candidature.

To my supervisors: thank you for all of the support you have provided throughout my candidature, and for providing the room necessary for me to grow as a researcher, and explore many of my interests in severe weather. Lance, you provided valuable intellectual stimulation and so much deep knowledge from your career in atmospheric and climate science. Shev, you were a constant source of positivity and interesting insights from a mathematician's point of view. I also thank Prof. James Brown for coming aboard my supervisory panel and providing much needed administrative support.

I would also like to thank my colleagues both at UTS and outside of UTS. The School of Mathematical and Physical Sciences has provided valuable financial and administrative support throughout my candidature. My appreciation goes out especially to the school manager, Elizabeth Gurung Tamang, for always making the time for me to talk and sort out different forms, and providing me with opportunities to work with other areas of the school, the Faculty of Science and the Finance team at UTS. Alex Bishop and Mohammad Mahdi Ahmadian, who completed their PhDs alongside me, provided fun and refreshing conversations every Wednesday that I was on campus, prior to the COVID-19 outbreak. Some of my PhD work required significant amounts of computing power (particularly for the numerical simulations and development of the SCT climatology), to which I am grateful to Mike Lake, the UTS scheme manager of the NCI supercomputing facility, for providing sufficient computation time on the supercomputer and always being available to help with any problems I had running code on it. Further thanks go to Peter Chan and Bruce Buckley at IAG for arranging insurance data for my research project, which has helped with choosing case studies and provided a useful form of verification of my numerical simulations. I would like to thank Ron Holle, John Cramer and Ryan Said of Vaisala Inc. for assisting in my request for GLD360 lightning data, and Joshua Soderholm and Alain Protat for providing weather radar data — both data sets were vital in the development of the severe convective thunderstorm hazards climatology. The weather radar data provided an additional form of verification of my numerical simulations. I was also fortunate to begin collaborating with Dr Milton Speer during my candidature, who has provided a number of additional research avenues to understand Australian severe weather, and whose expertise and meteorological background I truly admire.

I want to thank my friends for supporting me and understanding when I would miss social events so I could work on my project. I would like to acknowledge and thank my parents, Mer-

ren and Peter, for providing love and support, and always encouraging me to follow my dreams. Without both of you, getting to where I am now would have been unachievable. Thanks also goes to my sister, Brea, for always being there for me. Finally I want to thank my partner, Matthew, for being a constant source of love and support, and for keeping me grounded. I hope to provide the same level of support and encouragement as you complete your PhD.

Abstract

Various forms of severe weather frequently impact communities around Australia, including drought, floods, bushfires, severe convective thunderstorms and tropical cyclones. Such events have varied temporal scales and can have a significant effect on society, especially when multiple events occur shortly after each other or at the same time. As the global warming trend continues, the threat of severe weather impacting regions likely will change. Along with using global circulation models and regional climate models to understand how future warming will affect the risk of severe weather, it is important to study severe weather historically, and trends in severe weather and related variables, as this highlights the performance of modelling efforts, and what is currently occurring. Additionally, statistical modelling can be used to assess the attributes of severe weather, at the time scales important to forecasting and planning for these events. However, non-linear statistical methods including machine learning have seldom been used in the atmospheric and climate sciences to help understand severe weather, especially in Australia. Meanwhile, numerical modelling can be applied to past events, such as severe convection, to help understand how the event developed and improve future forecasting of similar events. Yet, there are few numerical modelling case studies of severe convective storms in Australia, particularly for tornadic storms, with much of the knowledge Australian forecasters rely on coming from studies and understandings developed overseas. This thesis applies trend analysis and statistical modelling to improve understanding of drought across southeast Australia, and severe convective storm hazard environments across Australia. Numerical modelling also is applied to a case study of a tornadic storm in Australia's most populous city, Sydney.

First, trends in drought are understood through analysis of change to percentiles, bootstrap resampling of periods within the time series, and permutation testing to understand statistical significance. This has been applied to key areas in southeast Australia, including the Hunter Valley Region, Sydney Catchment Area (SCA), Canberra city, and the Northern Murray-Darling Basin. It was found that, overall, there are concerning changes to precipitation mean and variance particularly over the cool season, suggesting a reduction in streamflow across the region. Coupled with increasing mean temperatures for most seasons across southeast Australia causing an increase in potential evapotranspiration, there appears to be a trend towards more frequent and severe drought. Wavelet analysis was applied to assess potential climate drivers within the time series, and how the relationships with these climate drivers have changed over time. Finally, statistical modelling was applied to further understand the key climate drivers associated with precipitation in the region, and assess how well different modelling techniques perform on prediction over the climate-scale. These models generally performed better than the climatology, and highlighted the influence of global warming on precipitation with Tasman Sea Surface Temperature Anomalies a frequently selected attribute. This is particularly apparent in the SCA, where models performed particularly well on predicting annual precipitation, despite trends in the time series.

On 16 December 2015, a severe thunderstorm and associated tornado affected Sydney, causing widespread damage and insured losses of \$206 million. This storm severely impacted the suburb of Kurnell, with significant damage sustained to Sydney's desalination plant, which supplies up to 15% of Sydney water during drought. High resolution numerical simulations of this storm were conducted using the Weather Research and Forecasting (WRF) model on a double nested domain using ECMWF's ERA5 reanalysis data for initial conditions. The results from simulations with the Morrison microphysics scheme are compared to those from the National Severe Storms Laboratory double-moment 4-ice microphysics scheme. Both simulations produced severe convective storms that followed similar paths to the observed storm, but the Morrison scheme did not display the same morphology. On the other hand, the storm simulated with

the NSSL scheme displayed cyclical low- and mid-level mesocyclone development, which was observed in the Kurnell storm. This highlights how supportive the atmosphere was for development of severe rotating thunderstorms with the potential to produce tornadoes. The work presented in this thesis displays the applicability of the WRF model to studying the causes of high-impact Australian thunderstorms.

In the final part of this thesis, statistical modelling is applied to better understand the atmospheric variables most closely associated with severe convective storm hazards (i.e., hail over 2 cm, wind gusts above 90 km/hr or tornadoes). The best performing models were then applied to the Bureau of Meteorology's BARRA reanalysis data set. This highlighted differing regions of Australia that are most likely affected by the individual hazards. There was a high frequency of hail environments over eastern Australia, particularly from the central coast of NSW into southeast Queensland, and the far southeast of Australia. There is a similar peak in the frequency of tornadic environments, with an additional high frequency of environments across much of southern Australia, likely due to cool-season tornadoes. Meanwhile, for environments supportive of severe convective wind, there was a high frequency of environments over part of eastern Australia. There was a very high frequency of environments for all three hazards over much of northern Australia. This is suggested to be an overestimation, except possibly for wind environments, as the majority of data that the statistical models were trained on, was obtained from the southeast of Australia. For wind environments, there were more reports in northern Australia, so the model is likely better trained to assess environment frequency across both regions. These results highlight the importance of developing a spatially balanced data set, along with categorically balanced data, or otherwise breaking a large area into smaller regions, prior to training statistical models. Spatial changes in the frequency of severe thunderstorm hazard environments were assessed between two periods in the BARRA data set. It was found that hail and tornadic environments have increased in frequency over eastern Australia during spring, tornadic environments have also increased in frequency over southeast and northwest Australia during summer, while they have decreased over southwest Australia during autumn and winter. There also is an increase in the frequency of environments supportive of severe wind thunderstorms over Western Australia and eastern Australia during spring, and parts of southwest and southeast Australia during summer.

Supporting Publications

Hartigan, J., MacNamara, S., and Leslie, L.M. Comparing precipitation and temperature trends between inland and coastal locations. ANZIAM J, 60:C109–C126, 2018. DOI: 10.21914/anziamj.v60i0.13967

Hartigan, J., MacNamara, S., and Leslie, L.M. Trends in precipitation and temperature in Canberra. In Elsworth, S. (ed.) MODSIM2019, 23rd International Conference on Modelling and Simulation. Modelling and Simulation Society of Australia and New Zealand, December 2019, pp. 1181–1187, 2019. DOI: 10.36334/modsim.2019.K7.hartigan

Hartigan, J., MacNamara, S., and Leslie, L.M. Application of Machine Learning to Attribution and Prediction of Seasonal Precipitation and Temperature Trends in Canberra, Australia. Climate, 8(6):76, 2020. DOI: 10.3390/cli8060076

Hartigan, J., MacNamara, S., Leslie, L.M., and Speer, M. Attribution and Prediction of Precipitation and Temperature Trends within the Sydney Catchment Using Machine Learning. Climate, 8(10):120, 2020. DOI: 10.3390/cli8100120

Hartigan, J., MacNamara, S., Leslie, L.M., and Speer, M. High resolution simulations of a tornadic storm affecting Sydney. ANZIAM J, 62:C1–C15, 2020. DOI: 10.21914/anziamj.v62.16113

Speer, M., Leslie, L.M., Hartigan, J., and MacNamara, S. Changes in Frequency and Location of East Coast Low Pressure Systems Affecting Southeast Australia. Climate, 9(3):44, 2021. DOI: 10.3390/cli9030044

Speer, M.S., Leslie, L.M., MacNamara, S., and Hartigan, J. From the 1990s climate change has decreased cool season catchment precipitation reducing river heights in Australia’s southern Murray-Darling Basin. Sci. Rep., 11(1):1–16, 2021. DOI: 10.1038/s41598-021-95531-4

Speer, M.S., Leslie, L.M., and Hartigan, J. Jet Stream Changes over Southeast Australia during the Early Cool Season in Response to Accelerated Global Warming. Climate, 10(6):84, 2022. DOI: 10.3390/cli10060084

Speer, M.S., Hartigan, J., and Leslie, L.M. Machine Learning Assessment of the Impact of Global Warming on the Climate Drivers of Water Supply to Australia’s Northern Murray-Darling Basin. Water, 14(19):3073, 2022. DOI: 10.3390/w14193073

Contents

| | | |
|----------|--|-----------|
| 1 | Introduction | 22 |
| 1.1 | Motivation | 22 |
| 1.2 | Drought | 23 |
| 1.3 | Convective Storms | 26 |
| 1.3.1 | Buoyancy | 26 |
| 1.3.2 | Deep Moist Convection | 27 |
| 1.3.3 | Vertical Wind Shear and Severe Thunderstorms | 29 |
| 1.4 | Severe Thunderstorm Climatologies | 31 |
| 1.4.1 | What is a Climatology? | 31 |
| 1.4.2 | Severe Convective Storms in Australia | 32 |
| 1.5 | Thesis Aims | 34 |
| 2 | Data and Methodologies | 45 |
| 2.1 | Precipitation, Temperature and Climate Drivers | 45 |
| 2.2 | Trend Analysis | 47 |
| 2.3 | Statistical Attribution and Prediction | 49 |
| 2.3.1 | Linear and Logistic Regression | 49 |
| 2.3.2 | Support Vector Machines | 50 |
| 2.3.3 | Random Forests | 52 |
| 2.4 | The Weather Research and Forecasting Model | 53 |
| 2.4.1 | Microphysics | 60 |
| 2.5 | Other Data | 62 |
| 2.5.1 | Severe Thunderstorm Archive | 62 |
| 2.5.2 | Radar and Lightning Data | 62 |
| 2.5.3 | Reanalysis Data Sets | 63 |
| 3 | Comparing Precipitation and Temperature Trends Between Inland and Coastal Locations | 70 |
| 3.1 | Overview | 70 |
| 3.2 | Introduction | 71 |
| 3.3 | Data and methodology | 71 |
| 3.4 | Results | 73 |
| 3.4.1 | Evolution of precipitation | 73 |
| 3.4.2 | Evolution of temperature | 75 |
| 3.4.3 | Wavelet analysis of precipitation | 75 |
| 3.4.4 | Wavelet analysis of temperature | 77 |
| 3.5 | Discussion and Conclusions | 77 |
| 4 | Application of Machine Learning to Attribution and Prediction of Seasonal Precipitation and Temperature Trends in Canberra, Australia | 81 |
| 4.1 | Overview | 81 |

| | | |
|----------|---|------------|
| 4.2 | Introduction | 82 |
| 4.3 | Materials and Methods | 83 |
| 4.3.1 | Data | 83 |
| 4.3.2 | Statistical Analysis | 83 |
| 4.3.3 | Attribute Selection | 83 |
| 4.3.4 | Training and Prediction | 85 |
| 4.4 | Results | 87 |
| 4.4.1 | Evolution of Precipitation | 87 |
| 4.4.2 | Evolution of Temperature | 90 |
| 4.4.3 | Wavelet Analysis of Precipitation | 92 |
| 4.4.4 | Wavelet Analysis of Temperature | 94 |
| 4.4.5 | Training and Prediction of Precipitation | 96 |
| 4.4.6 | Training and Prediction of TMax | 97 |
| 4.5 | Discussion | 97 |
| 4.5.1 | Drought Vulnerability | 97 |
| 4.5.2 | Attribution and Prediction | 98 |
| 4.6 | Conclusions | 99 |
| 5 | Attribution and Prediction of Precipitation and Temperature Trends within the Sydney Catchment Using Machine Learning | 104 |
| 5.1 | Overview | 104 |
| 5.2 | Introduction | 105 |
| 5.3 | Data and Methodology | 106 |
| 5.3.1 | Data | 106 |
| 5.3.2 | Statistical Analysis | 107 |
| 5.3.3 | Attribute Selection and Prediction | 107 |
| 5.4 | Results | 111 |
| 5.4.1 | Evolution of Precipitation | 111 |
| 5.4.2 | Evolution of Temperature | 114 |
| 5.4.3 | Wavelet Analysis of Precipitation | 117 |
| 5.4.4 | Wavelet Analysis of Temperature | 117 |
| 5.4.5 | Attribution and Prediction of Annual Precipitation | 121 |
| 5.4.6 | Attribution and Prediction of Autumn Precipitation | 122 |
| 5.4.7 | Attribution and Prediction of Winter Precipitation | 123 |
| 5.5 | Discussion | 124 |
| 5.5.1 | Drought Vulnerability | 124 |
| 5.5.2 | Attribution and Prediction | 124 |
| 5.6 | Conclusions | 127 |
| 6 | Impact of accelerated global warming on rainfall and temperature trends in Australia's northern Murray-Darling Basin using statistical analysis and machine learning | 132 |
| 6.1 | Overview | 132 |
| 6.2 | Introduction | 133 |
| 6.2.1 | Australia's Murray-Darling Basin | 133 |
| 6.2.2 | The NMDB | 134 |
| 6.3 | Data and Methodology | 137 |
| 6.3.1 | Data | 137 |
| 6.3.2 | Statistical analysis | 138 |
| 6.3.3 | Attribute selection | 138 |

| | | |
|----------|--|------------|
| 6.4 | Results and Discussion | 139 |
| 6.4.1 | Precipitation and Temperature Time Series in the Northern Murray-Darling Basin | 139 |
| 6.4.2 | P-values and box-whisker plots for precipitation, TMax and TMin | 139 |
| 6.4.3 | Wavelet analysis of temperature and precipitation 1911–2018 | 142 |
| 6.4.4 | Attribute Selection | 144 |
| 6.5 | Conclusions | 145 |
| 7 | High resolution simulations of a tornadic storm affecting Sydney | 148 |
| 7.1 | Overview | 148 |
| 7.2 | Introduction | 149 |
| 7.3 | Data and methodology | 149 |
| 7.4 | Event overview | 150 |
| 7.5 | Simulation results | 152 |
| 7.6 | Conclusions and future work | 154 |
| 8 | An Australian Severe Convective Thunderstorm Hazard Climatology | 157 |
| 8.1 | Overview | 157 |
| 8.2 | Introduction | 158 |
| 8.3 | Data and Methodology | 160 |
| 8.3.1 | Data | 160 |
| 8.3.2 | Methodology | 162 |
| 8.3.3 | Classification Model Selection | 163 |
| 8.4 | Results | 165 |
| 8.4.1 | Comparison of Pseudo-proximity Soundings to Observed Soundings . . . | 165 |
| 8.4.2 | STA Climatology | 169 |
| 8.4.3 | Model Fitting | 173 |
| 8.4.4 | BARRA Climatology | 186 |
| 8.4.5 | Unbalanced Data Case Study | 193 |
| 8.4.6 | Trends in the BARRA Climatology | 197 |
| 8.5 | Discussion and Conclusions | 205 |
| 9 | Conclusions and Future Work | 216 |

List of Figures

| | | |
|-------|---|----|
| 2.5.1 | Map displaying the three regions where Vaisala Inc. GLD360 lightning network data was obtained from 2014–2018 (black rectangles), and BoM weather radars (red stars) within these regions to determine if detected lightning was associated with severe hail, as determined by calculated MESH ≥ 2 cm. . . . | 64 |
| 3.2.1 | Drought in NSW over the past 24 months, taken from the Bureau of Meteorology (2018). Locations of Newcastle and Scone are included. | 72 |
| 3.4.1 | Time series of total wet season precipitation (top); and box plots of the bootstrapped mean wet season precipitation over 20-year periods (bottom) for Newcastle and Scone. Dashed lines indicate the 5th and 95th (bottom and top red), 10th and 90th (bottom and top green) 15th and 85th (bottom and top dark blue), 20th and 80th (bottom and top pink), and 25th and 75th percentiles (bottom and top light blue). | 74 |
| 3.4.2 | As in Figure 3.4.1 but for TMax during each location’s respective wet season. | 76 |
| 3.4.3 | As in Figure 3.4.1 but for TMin during each location’s respective wet season. | 76 |
| 3.4.4 | Wavelet analysis for total wet season precipitation for Newcastle and Scone. Low values (blue) in the wavelet power spectrum (left) indicate low variability while high values (red) indicate high variability. Peaks in the global power spectrum (right) indicate high variability. The dashed/solid red line indicates the 95% confidence level, while the dashed/solid black line indicates the 90% confidence level. | 77 |
| 3.4.5 | As in Figure 3.4.4, but for TMax (top two rows) and TMin (bottom two rows). | 78 |
| 4.3.1 | Scatter plot of annual precipitation in Canberra against each climate driver that serve as possible attributes, showing non-linear relationships between precipitation and the climate drivers. The thin black line is a linear fit against the data using least squares regression. Correlations of each predictor against precipitation are provided in the box to the lower right of each sub-panel, low correlations ($ r < 0.3$) indicate little linear relationship between a predictor and precipitation. | 84 |
| 4.3.2 | As in Figure 4.3.1 except for TMax. Predictors with $ r > 0.4$ indicate moderate or strong relationships with TMax likely because they are attributes of global warming, or have been influenced by global warming. | 85 |
| 4.4.1 | (a) Annual precipitation time series; and (b) box plots of the bootstrapped mean precipitation over 20-year periods for Canberra, both annually and for all four seasons. Dashed lines indicate the 5th and 95th (bottom and top red), 10th and 90th (bottom and top green) 15th and 85th (bottom and top dark blue), 20th and 80th (bottom and top pink), and 25th and 75th percentiles (bottom and top light blue). | 89 |
| 4.4.2 | (a) Time series and (b) box plots as in Figure 4.4.1 except for TMax. | 90 |
| 4.4.3 | (a) Time series and (b) box plots as in Figure 4.4.1 except for TMin. | 91 |

| | | |
|-------|--|-----|
| 4.4.4 | Wavelet analysis for total wet season precipitation for Canberra. Low values (blue) in the wavelet power spectrum (a) indicate low periodicity while high values (red) indicate high periodicity. Peaks in the global power spectrum (b) on the right panel indicate high periodicity. The dashed/solid red line indicates the 95% confidence level, while the dashed/solid black line indicates the 90% confidence level in each plot. The cone of influence is depicted in the wavelet power spectrum by the solid black cone. | 93 |
| 4.4.5 | (a) Wavelet power spectra and (b) global power spectra for TMax, with low and high values in the power spectra indicating low and high periodicity, respectively, the cone of influence and statistical significance bands as already explained in Figure 4.4.4. | 94 |
| 4.4.6 | (a) Wavelet power spectra and (b) global power spectra for TMin, with low and high values in the power spectra indicating low and high periodicity, respectively, the cone of influence and statistical significance bands as already explained in Figure 4.4.4. | 95 |
| 5.2.1 | Location map of the Sydney Catchment Area (outline in dark grey): the six sites where data were obtained and the Sydney central business district (black dots). | 106 |
| 5.3.1 | Scatter plot of annual precipitation across the Sydney Catchment Area (SCA) against each climate driver that serves as a possible attribute. The thin black line shows a linear fit against the data using least squares regression. Correlations of each predictor against precipitation are provided in the box to the lower right of each sub-panel; low correlations ($ r < 0.3$) indicate a weak linear relationship between a predictor and precipitation. | 109 |
| 5.4.1 | (left panels, a–i) Annual precipitation time series summed across all six sites. Dashed lines indicate the 5th and 95th (bottom and top red), 10th and 90th (bottom and top orange), 15th and 85th (bottom and top green), 20th and 80th (bottom and top brown) and 25th and 75th percentiles (bottom and top dark blue); and (right panels, b–j) box plots of the bootstrapped mean annual precipitation over 31-year periods for the SCA. | 112 |
| 5.4.2 | (left panels, a–i) Annual mean TMax time series across the SCA; and (right panels, b–j) box plots of the bootstrapped mean TMax over 31-year periods for the SCA, as in Figure 5.4.1 but for mean TMax. | 115 |
| 5.4.3 | (left panels, a–i) Annual mean TMin time series across the SCA; and (right panels, b–j) box plots of the bootstrapped mean TMin over 31-year periods for the SCA, as in Figure 5.4.1 but for mean TMin. | 116 |
| 5.4.4 | Wavelet analysis for total precipitation across the SCA both annually (a–b) and over the four seasons (c–j). Low values (blue) in the wavelet power spectrum (left panels, a–i) indicate low periodicity, while high values (orange) indicate high periodicity. Horizontal peaks in the global power spectrum (right panel, b–j) indicate high periodicity over the time series. The dashed/solid red (black) line indicates the 95% (90%) confidence level in each plot. The cone of influence is depicted in the wavelet power spectrum by the solid black cone. | 118 |
| 5.4.5 | Wavelet power spectra (left panel, a–i) and global power spectra (right panel, b–j) for TMax, with low and high values in the power spectra indicating low and high periodicity, respectively; the cone of influence and statistical significance bands as already explained in Figure 5.4.4. | 119 |

| | | |
|-------|--|-----|
| 5.4.6 | Wavelet power spectra (left panel, a–i) and global power spectra (right panel, b–j) for TMin, with low and high values in the power spectra indicating low and high periodicity, respectively; the cone of influence and statistical significance bands as already explained in Figure 5.4.4. | 120 |
| 6.2.1 | Map of northern and southern Murray-Darling Basin in southeast Australia The MDB lies within subtropical latitudes (25°S – 38°S) of the Australian continent. Observation stations used for precipitation, TMax and TMin averaging that represent the NMDB are marked and indicated in a legend. (Source: Murray-Darling Basin Authority, G.P.O. Box 1801, Canberra City, ACT 2601 Australia. https://www.mdba.gov.au/sites/default/files/pubs/Murray-Darling_Basin_Boundary.pdf) Reproduced with some place name deletions and insertions via license: Creative Commons Attribution-Non Commercial-NoDerivatives4.0 International Public License (CC BY-NC-ND 4.0) | 134 |
| 6.2.2 | Murray-Darling Basin rainfall deciles Rainfall deciles for the 48 months January 2017 to December 2020 in southeast Australia focusing on the MDB defined by the area within the solid black line. Note the lowest on record in the north of the basin and the very much below or below average rainfall in the rest of the basin. (Reproduced with permission under Creative Commons Attribution Licence 3.0 from the Australian Bureau of Meteorology). Available at : http://www.bom.gov.au/climate/maps/rainfall/?variable=rainfall&map=decile&period=48month&region=md&year=2020&month=12&day=31 | 136 |
| 6.2.3 | Annual deciles of actual evapotranspiration and soil moisture 2018–2019 Map of southeast Australia showing for the MDB region deciles during the 2018–2019 year for, (a) annual area-averaged actual evapotranspiration. Note the decile area of lowest on record in the NMDB, and (b) annual area-averaged soil moisture. Note the decile area of the lowest on record in the NMDB. (Reproduced with permission under Creative Commons Attribution Licence 3.0 from the Australian Bureau of Meteorology. Available at: http://www.bom.gov.au/water/nwa/2019/mdb/climateandwater/climateandwater.shtml). | 137 |
| 6.4.1 | Precipitation, TMax & TMin time series in the NMDB Precipitation time series in the NMDB for (a) Annual, (b) April-May. Dashed lines indicate percentiles 5th and 95th (red); 10th and 90th (orange); 15th and 85th (light green); 20th and 80th (brown); and, 25th and 75th (dark blue). Note the apparent decrease in mean and reduction in values greater than the 75th percentile for April-May since the 1990s and the decrease from 2011 to well below the 5th percentile in 2019 for the annual time series; the time series of TMax in the NMDB for (c) Annual, (d) October-March. Note the steep increase approaching the 100th percentile in both October-March and annual time series, which mirrors the steep decrease in 2019 annual precipitation (Figure 6.4.1b); the time series of TMin in the NMDB for (e) Annual, (f) April-May. Note the almost linear increase in the annual time series since the 1960s and the steep increase with much less variation since 2014, while there is no clear trend in April-May apart from one steep increase after the 1950s and another from 2014. | 140 |

| | | |
|-------|---|-----|
| 6.4.2 | Box and whisker plots of NMDB precipitation, TMax & TMin Box and whisker plots of NMDB for (a) mean April-May precipitation, and (b) variance; (c) mean annual TMax; (d) mean annual TMin; (e) mean April-May TMin. | 142 |
| 6.4.3 | Wavelets for precipitation Wavelets representing NMDB precipitation for the periods (a) annual, (b) April-May, (c) JJAS, and (d) October-March. The dashed black and red lines are the 90th and 95th confidence percentiles, respectively. | 143 |
| 7.4.1 | Atmospheric conditions from ERA5 reanalysis data, at the time of development of the Kurnell tornadic storm. Variables shown are (a) mean sea level pressure (hPa); (b) 500 hPa geopotential height (m) represented as thin black contours, and temperature ($^{\circ}\text{C}$) represented by filled contours; (c) convective available potential energy (CAPE; J kg^{-1}); and (d) total totals index (TT; K, Miller, 1972). | 151 |
| 7.4.2 | Filled contours display radar reflectivity (dBZ), thin black contours are vertical velocity (every 10 m s^{-1} starting at 5 m s^{-1}), and vectors show wind speed and direction, from the Kurnell radar interpolated to 2.5 km above sea level, at (a) 2325 UTC and (b) 2331 UTC. The location of Kurnell is provided by a solid green triangle. White regions are where no reflectivity is observed by the radar. The white circle in the figure is due to the storm propagating over the Kurnell radar, where sampling does not take place. However, vectors still exist in this circle as the 3D wind field is retrieved using multiple radars that sample this region. | 151 |
| 7.5.1 | Variables shown are lowest model level reflectivity (filled contours; dBZ), wind speed and direction (vectors), and cold pool outline (thick grey contour representing $\theta' < -1 \text{ K}$), and mid-level vertical velocity (eta-level = 17, approximately 5 km above sea level, thin black contours; every 10 m s^{-1} beginning at 5 m s^{-1}), for the NSSL simulation. The location of Sydney is provided by a solid green circle, and the location of Kurnell is provided by a solid green triangle. | 153 |
| 7.5.2 | Filled contours display moist potential temperature perturbations (K) at the lowest model level, while vectors display lowest model level wind speed (m s^{-1}) and direction. All figures are from the NSSL simulation. This figure is focused on the region of rotation, so the coastline is not included but is immediately to the left of the domain shown. | 153 |
| 7.5.3 | Simulated (a) maximum surface hail diameter (cm) and (b) maximum 10 m wind speed (m s^{-1}), between 2210 and 2320 UTC for the NSSL simulation. The location of Sydney is provided by a solid green circle, and the location of Kurnell is provided by a solid green triangle. | 154 |
| 8.4.1 | Scatterplots of bulk shear computed over different levels (left column) and CAPE computed for different parcel types (right column), comparing vertical profiles from the BARRA data set against observed rawinsonde profiles. The thin black line represents the line $y = x$, where points ideally should lie. . . . | 167 |
| 8.4.2 | As in Figure 8.4.1, but for the coastal rawinsonde sites only. | 168 |
| 8.4.3 | As in Figure 8.4.1, but for the inland rawinsonde sites only. | 169 |
| 8.4.4 | Binned density of total reports in the STA from 1990–2018 after manual filtering of reports, for (a) hail, (b) wind, and (c) tornadoes. | 170 |
| 8.4.5 | Binned density of all hail reports in the STA for each month from 1990–2018. | 172 |
| 8.4.6 | As in Figure 8.4.5 but for wind reports. | 172 |

| | | |
|--------|---|-----|
| 8.4.7 | As in Figure 8.4.5 but for tornado reports. | 173 |
| 8.4.8 | Receiver Operating Characteristic (ROC) curves for each statistical model for hail SCTs described in Table 8.8, with those developed through forward selection on the left and those through backward selection on the right. . . . | 179 |
| 8.4.9 | Precision-Recall (PR) curves (solid blue lines) for each statistical model for hail SCTs described in Table 8.8, with those developed through forward selection on the left and those through backward selection on the right. Solid, dark grey lines depict lines of constant CSI, while dashed light grey lines depict lines of constant bias. | 180 |
| 8.4.10 | ROC curves for each statistical model for tornadic SCTs described in Table 8.10, as outlined in Figure 8.4.8. | 181 |
| 8.4.11 | PR curves for each statistical model for tornadic SCTs described in Table 8.10, as outlined in Figure 8.4.9. | 182 |
| 8.4.12 | ROC curves for each statistical model for wind SCTs described in Table 8.12, as outlined in Figure 8.4.8. | 183 |
| 8.4.13 | PR curves for each statistical model for wind SCTs described in Table 8.12, as outlined in Figure 8.4.9. | 184 |
| 8.4.14 | Average frequency of environments supportive of SCT hazards (in days per year) for (a) hail, (b) wind, (c) tornadoes as diagnosed by the SVR Poly F model, and (d) tornadoes as diagnosed by the LogR B model. | 187 |
| 8.4.15 | 1990–2018 average frequency of environments supportive of hail SCTs per year in the BARRA reanalysis data set. | 188 |
| 8.4.16 | As in Figure 8.4.15 but for tornadic SCTs using the SVM Poly F model. . . . | 191 |
| 8.4.17 | As in Figure 8.4.15 but for tornadic SCTs using the LogR B model. | 192 |
| 8.4.18 | As in Figure 8.4.15 but for wind SCTs. | 194 |
| 8.4.19 | Average frequency of environments supportive of SCT hazards for (a) hail, and (b) wind with from the best performing models developed using the unbalanced data set. | 196 |
| 8.4.20 | Filled contours display average change in frequency of environments supportive of hail SCTs between 1990–2003 and 2004–2018 using bootstrap resampling. Change is represented as a monthly average number of days (a) annually, and over (b) autumn, (c) winter, (d) spring, and (e) summer. The thick black contour displays statistically significant changes from permutation testing where the p-value < 0.1. | 199 |
| 8.4.21 | As in Figure 8.4.20 but for environments supportive of tornadic SCTs using the SVM Poly F model. | 201 |
| 8.4.22 | As in Figure 8.4.20 but for environments supportive of tornadic SCTs using the LogR B model. | 202 |
| 8.4.23 | As in Figure 8.4.20 but for environments supportive of wind SCTs. | 204 |

List of Tables

| | | |
|-----|---|-----|
| 2.1 | List of locations where weather station data was obtained to assess trends in precipitation and temperature over regions of southeast Australia. | 46 |
| 2.2 | Number of SCT reports by hazard type, for the raw STA data set and the filtered data set. | 63 |
| 3.1 | P-values for the permutation test on the difference in the mean wet season precipitation between two 20-year periods. Tests conducted used 5000 resamples. | 74 |
| 3.2 | As in Table 3.1 but for the difference in the mean maximum temperature (TMax), and for the mean minimum temperature (TMin) during the wet season. | 75 |
| 4.1 | Percentage of folds selecting attributes for precipitation using LR, and SVR with the radial basis (RBF) and polynomial (poly) kernel functions. | 86 |
| 4.2 | Same as Table 4.1 except for TMax. | 87 |
| 4.3 | p-values for the permutation test on the difference in the mean wet season precipitation, mean maximum temperature or mean minimum temperature between 1939–1958 and 1999–2018; and 1979–1998 and 1999–2018. Tests conducted used 5000 resamples. Text in bold face highlight statistical significance at the 95% confidence level, text in italics highlight statistical significance at the 90% confidence level. | 88 |
| 4.4 | Performance of each model on the testing data set (2006–2017). The best performing model for predicting annual precipitation and TMax is highlighted, based on low RMSE, and both high skill and correlation. | 96 |
| 5.1 | Percentage of folds selecting attributes for precipitation in linear regression (LR); support vector regression (SVR) with the radial basis (RBF) and polynomial (Poly) kernel functions; and random forests (RF). | 111 |
| 5.2 | p-values for the permutation test on the difference between the mean and variance of precipitation, TMax or TMin between 1958–1988 and 1989–2019, annually and across all four seasons. Tests conducted used 5000 resamples. Text in bold face highlights statistical significance at the 95% confidence level; text in italics highlights statistical significance at the 90% confidence level. | 113 |
| 5.3 | Performance of each of the annual precipitation models on the testing data set (2011–2019). The best performing model for predicting annual precipitation is highlighted, determined by low RMSE, and both high correlation and skill. | 121 |
| 5.4 | Performance of each of the autumn precipitation models on the testing data set (2011–2019), similar to Table 5.3. | 123 |
| 5.5 | Performance of each of the winter precipitation models on the testing data set (2011–2019), similar to Table 5.3. | 123 |

| | | |
|------|---|-----|
| 6.1 | P-values from permutation testing differences in interval means and variances P-values from permutation testing differences in interval means and variances for April–May, JJAS, October–March and annual precipitation, TMax and TMin, based on area averages of observing stations in the northeast part of the NMDB. Marginally significant values (p-value ≤ 0.10) are in bold italics. Note that the p-value for each variance test is calculated after one sample has had bias correction in the mean. Key points to note are the significant and highly significant p-values (p-value < 0.05 and p-value < 0.01 , respectively) for the April–May mean and variance precipitation decreases from 1965–1991 to 1992–2018; and the highly significant increases in mean TMin, and in mean TMax for most of the periods. | 141 |
| 6.2 | Major precipitation attributes identified for each time period The five major precipitation attributes identified, for each time period. They had the highest percentages of appearances in the 10-fold, cross-validation of the machine learning schemes, applied to the 1965–2018 observed precipitation data set. | 144 |
| 8.1 | Example 2×2 contingency table outlining the possible outcomes from a simple yes/no classification model. Here, <i>a</i> represents a true positive, <i>b</i> a false positive, <i>c</i> a false negative and <i>d</i> a true negative. | 165 |
| 8.2 | Information for each rawinsonde site in Australia that was used to compare against pseudo-proximity soundings derived from the BARRA reanalysis data set. | 166 |
| 8.3 | Percentage of folds that indices calculated using the surface-based (SB), most-unstable (MU) and mixed-layer (ML) parcel formulations appeared in. | 173 |
| 8.4 | Percentage of folds selecting attributes for hail SCTs using LogR, SVM with the radial basis (RBF) and polynomial (Poly) kernel functions, and random forests (RF) using either forward (F) or backward (B) selection techniques to search through the space of attributes. | 174 |
| 8.5 | As in Table 8.4 but for tornadic SCTs. | 174 |
| 8.6 | As in Table 8.4 but for wind SCTs. | 175 |
| 8.7 | Contingency statistics on the testing data set for the best performing models on hail SCTs. The best performance among the models for each contingency statistic is highlighted in bold face. | 176 |
| 8.8 | Features and tuning parameters selected for the best performing models presented in Table 8.7. | 176 |
| 8.9 | As in Table 8.7 but for tornadic SCTs. | 176 |
| 8.10 | Features and tuning parameters selected for the best performing models presented in Table 8.9. | 176 |
| 8.11 | As in Table 8.7 but for wind SCTs. | 177 |
| 8.12 | Features and tuning parameters selected for the best performing models presented in Table 8.11. | 177 |
| 8.13 | Contingency statistics from application of the discriminant relationship described in (Allen et al., 2011, , their Equation 7), applied to each of the test data sets used for hail, tornado and wind SCTs. | 186 |
| 8.14 | Percentage of folds that indices calculated using the surface-based (SB), most-unstable (MU) and mixed-layer (ML) parcel formulations appeared in, when developing the statistical models on an unbalanced data set. | 193 |
| 8.15 | As in Table 8.7 but for hail SCTs developed with unbalanced data. | 195 |

| | | |
|------|---|-----|
| 8.16 | Features and tuning parameters selected for the best performing models presented in Table 8.15. | 195 |
| 8.17 | As in Table 8.7 but for wind SCTs developed with unbalanced data. | 195 |
| 8.18 | Features and tuning parameters selected for the best performing models presented in Table 8.17. | 195 |

Abbreviations

- AMO: Atlantic Multidecadal Oscillation
- ARW: Advanced Research WRF
- BARRA: Bureau of Meteorology Atmospheric Reanalysis
- BoM: Bureau of Meteorology
- CAPE: Convective Available Potential Energy
- CIN: Convective Inhibition
- DMC: Deep Moist Convection
- DMI: Dipole Mode Index
- ECMWF: European Centre for Medium-Range Weather Forecasts
- EHI: Energy Helicity Index
- EL: Equilibrium Level
- ENSO: El-Niño Southern Oscillation
- GLD360: Global Lightning Dataset 360
- GlobalSSTA: Global Sea Surface Temperature Anomalies
- GlobalT: Global Temperature Anomalies
- GRIB: General Regularly Distributed Information in Binary form
- IOD: Indian Ocean Dipole
- IQR: Interquartile Range
- IPO: Interdecadal Pacific Oscillation
- JJAS: June, July, August, September (the southern Australian cool season)
- LapR: Lapse Rate
- LCL: Lifted Condensation Level
- LFC: Level of Free Convection
- LI: Lifted Index
- LR: Linear Regression
- LogR: Logistic Regression
- LSM: Land-surface model
- MATLAB: MATrix LABoratory
- MCS: Mesoscale Convective System

- MDB: Murray-Darling Basin
- MESH: Maximum Estimated Size of Hail
- ML: Mixed-Layer
- MSE: Mean Square Error
- MU: Most-Unstable
- NMDB: Northern Murray-Darling Basin
- NSSL: National Severe Storms Laboratory
- NSW: New South Wales
- OT: Overshooting Cloud-Top
- PBL: Planetary Boundary Layer
- PDO: Pacific Decadal Oscillation
- Poly: Polynomial kernel
- PW: Precipitable Water
- QLCS: Quasi-Linear Convective System
- RBF: Radial Basis Function kernel
- RF: Random Forest
- RK3: Runge-Kutta third-order integration scheme
- RMSE: Root Mean Square Error
- RSS: Residual Sum of Squares
- S06: 0–6 km bulk shear
- SA: South Australia
- SB: Surface-Based
- SAM: Southern Annular Mode
- SCA: Sydney Catchment Area
- SCP: Supercell Composite Parameter
- SCT: Severe Convective Thunderstorm
- SD: Standard Deviation
- SHERB: Severe Hazards in Environments with Reduced Buoyancy Index
- SHIP: Significant Hail Parameter
- SOI: Southern Oscillation Index

- SRH: Storm-Relative Helicity
- STA: Severe Thunderstorm Archive
- STP: Significant Tornado Parameter
- SVM: Support Vector Machine
- SVR: Support Vector Regression
- SWEAT: Severe Weather Threat Index
- TMax: Mean maximum temperature
- TMin: Mean minimum temperature
- TPI: Tripole Index for the Interdecadal Pacific Oscillation
- TSSST: Tasman Sea Sea Surface Temperature Anomalies
- TT: Total Totals
- UK: United Kingdom
- USA: United States of America
- UTC: Universal Coordinated Time
- VWS: Vertical Wind Shear
- WA: Western Australia
- WRF: Weather Research and Forecasting