

**STATISTICAL CHARACTERISATION OF
WIND FIELDS OVER COMPLEX TERRAIN
WITH APPLICATIONS IN BUSHFIRE
MODELLING**

Rachael Elizabeth Quill

A thesis submitted in fulfilment of the requirements of the degree of

Doctor of Philosophy



School of Physical, Environmental and Mathematical Sciences

The University of New South Wales, Canberra

November 2017

ORIGINALITY STATEMENT

I hereby declare that this submission is my own work and to the best of my knowledge it contains no materials previously published or written by another person, or substantial proportions of material which have been accepted for the award of any other degree or diploma at UNSW or any other educational institution, except where due acknowledgement is made in the thesis. Any contribution made to the research by others, with whom I have worked at UNSW or elsewhere, is explicitly acknowledged in the thesis.

I also declare that the intellectual content of this thesis is the product of my own work, except to the extent that assistance from others in the project's design and conception or in style, presentation and linguistic expression is acknowledged.

Rachael Elizabeth Quill

COPYRIGHT STATEMENT

I hereby grant the University of New South Wales or its agents the right to archive and to make available my thesis or dissertation in whole or part in the University libraries in all forms of media, now or here after known, subject to the provisions of the Copyright Act 1968. I retain all proprietary rights, such as patent rights. I also retain the right to use in future works (such as articles of books) all or part of this thesis or dissertation.

I also authorise University Microfilms to use the 350 word abstract of my thesis in Dissertation Abstract International (this is applicable to doctoral theses only).

I have either used no substantial portions of copyright material in my thesis or I have obtained permission to use copyright material; where permission has not been granted I have applied/will apply for a partial restriction of the digital copy of my thesis or dissertation.

Rachael Elizabeth Quill

AUTHENTICITY STATEMENT

I certify that the Library deposit digital copy is a direct equivalent of the final officially approved version of my thesis. No emendation of content has occurred and if there are any minor variations in formatting, they are the result of the conversion to digital format.

Rachael Elizabeth Quill

Abstract

The propagation of bushfire across the landscape is dependent on a variety of environmental factors, but the wind, in particular, has a major effect on both the speed and direction of fire propagation. As such, bushfire spread models, which underpin successful bushfire management, require accurate knowledge of the pattern of winds across the landscape. This can be problematic over complex terrain where winds exhibit considerable spatial variability due to wind-terrain interactions, and where detailed measurements of wind characteristics are comparatively rare.

This thesis contributes two new wind datasets to address the previous lack of data available to develop and validate wind models over complex terrain. It also details analyses that focus on the statistical characterisation of wind as joint wind direction distributions, which represent the directional wind response to changing topography and surface roughness. A novel method for toroidal surface fitting is introduced and implemented to estimate the true continuous response from discrete observed data. This new method, which relies on a conceptually simple adaptation of planar techniques, is compared to the limited range of available toroidal surface estimation techniques and is shown to perform as well as, if not better than, these more sophisticated methods.

Monte Carlo simulations are employed to highlight the sensitivity of statistical comparison tests to alternative distribution structures, and to validate bivariate and circular extensions of the Kolmogorov-Smirnov test. These tests are applied to directional wind response pairs, showing that vegetation regrowth has a significant but varying impact across complex terrain.

Finally, this thesis demonstrates how statistical approaches can be used to complement current physics-based wind modelling methods. The resulting probabilistic representations provide more accurate predictions of wind direction variability, and are better suited to emerging ensemble-based bushfire prediction frameworks. As such, they provide a superior characterisation of uncertainty across the fire modelling process; ultimately enabling fire managers to make more informed decisions.

Acknowledgements

Firstly, to my supervisors, Jason Sharples and Leesa Sidhu. Your guidance has been invaluable. With your doors always open, you have not only guided this work to what you see today, but you have guided me through my academic apprenticeship. You have encouraged me to go beyond my comfort zone at every turn, seek out new ideas and embrace every opportunity. For that, I thank you.

Thank you for the financial support provided by UNSW Canberra, and acknowledgement goes to the Bushfire and Natural Hazards Cooperative Research Centre for financial support and supervision from Graham Thorpe. Particular thanks also goes to the AustMS Women in Mathematics Special Interest Group, the International Society for Non-Parametric Statistics and the CSIRO-ANZIAM Student Support Scheme for financial support enabling me to travel far and wide to communicate this research.

To my collaborators, Natalie Wagenbrenner, Kangmin Moon, Julia Piantadosi and Isaac Towers, thank you for your insights. You have helped to enlivened this work.

Observed data forms the foundation of this research, and the collection of such would not have been possible without the cooperation of NSW National Parks and Wildlife and the National Arboretum Canberra. Thank you to both agencies.

Data collection is never an individual pursuit. To Jason, Ben, Bob, Peter, Nick, Mum, Dad, Barbara, Tim, Katie, Hud, Hannah, Sarah, John and Therese, your efforts over the past three years are greatly appreciated, and I hope the payment in biscuits was sufficient! A huge thank you goes to Colin Symons, who worked tirelessly with me to upgrade the stations and braved the snow to help install them.

My fellow PEMS students; Billie, Hud, Mona, Zhibing, Chad, thank you for the lunch breaks, tea breaks, yoga breaks, Ikea breaks, and ‘its-time-for-a-break’ breaks! I miss you all already. And to the proof-readers; Peter, Mum, Ian and Ben, I will always be grateful.

Thank you to my friends and family for your love and support. No matter where you are in the world, you help me enjoy life. To Billie and to Julia, thank you for always thinking of cake. And to Sarah, thank you for always having the kettle on.

And finally, to Ben, you are my everything. From fiancé to husband (and soon to be dad), you are my rock. You have done the heavy lifting; you have listened; you have laughed; you have consoled; and you have celebrated. I could not have done this without you. Thank you.



The first of many weather station set ups.
UNSW Canberra Field Site, February 2014.

Published Work

Much of the work completed as part of this PhD thesis has been prepared and submitted for peer-review through journal publications, as well as being communicated to the research and professional communities through conference presentations, academic posters and non-academic written work.

The following publications and conference proceedings correspond to research presented in this thesis.

Quill, R., Sharples, J.J., Sidhu, L.A. and Towers, I. (In Prep) Understanding uncertainty in bushfire modelling and characterising uncertain wind inputs.

Quill, R., Sharples, J.J. and Sidhu, L.A. (Under Review) Non-parametric comparison of wind direction to assess the impacts of surface roughness. *Statistical Modelling*, submitted March 2017.

Quill, R., Sharples, J.J. and Sidhu, L.A. (In Prep) Sensitivity analysis of Kolmogorov-Smirnov style statistics for univariate and bivariate data.

Quill, R., Sharples, J.J. and Sidhu, L.A. (In Prep) Estimation of directional wind response using noisy bivariate circular data: a comparison of approaches.

Quill, R., Sharples, J.J., Sidhu, L.A., Wagenbrenner, N. and Forthofer, J. (In Prep) A wind direction evaluation of a diagnostic wind model over complex terrain in the context of ensemble-based fire spread modelling.

Quill, R., Moon, K., Sharples, J.J., Sidhu, L.A., Duff, T. and Tolhurst, K. (2016) Wind speed reduction induced by post-fire vegetation regrowth. Proceedings from the Bushfire and Natural Hazards CRC & AFAC conference, Brisbane, 30 Aug-1 Sep 2016.

Quill, R., Sharples, J.J. and Sidhu, L.A. (2015) Effects of post-fire vegetation regrowth on wind fields over complex terrain. In Weber, T., McPhee, M.J. and Anderssen, R.S. (eds) MODSIM2015, 21st International Congress on Modelling and Simulation. Modelling and Simulation Society of Australia and New Zealand, Dec 2015, pp. 490496. ISBN: 978-0-9872143-5-5. <http://www.mssanz.org.au/modsim2015/A4/quill.pdf>.

The following conference presentations also relate to work conducted for this thesis.

Quill, R., Sharples, J.J. and Sidhu, L.A. (2017) A Monte Carlo evaluation of the sensitivity of bivariate Kolmogorov-Smirnov style tests. ANZIAM 2017, Hahndorf, 5-9 Feb 2017.

Quill, R., Sharples, J.J. and Sidhu, L.A. (2016) The sensitivity of Kolmogorov-Smirnov style tests: a simulation study. The Australian Statistical Conference 2016, Canberra, 5-9 Dec 2016.

Quill, R., Moon, K., Sharples, J.J., Sidhu, L.A., Duff, T., Tolhurst, K. (2016) Impacts of topography and post-fire regrowth on wind speed reduction. Bushfire and Natural Hazards CRC & AFAC 2016, Brisbane, 30 Aug-1 Sep2016.

Quill, R., Sharples, J.J., Sidhu, L.A. and Piantadosi, J. (2016) Statistical comparison of bivariate circular distributions in the characterisation of wind direction response over complex terrain. The 26th Conference of The International Environmetrics Society, Edinburgh, 18-22 Jul 2016.

Quill, R., Sharples, J.J., Sidhu, L.A. and Piantadosi, J. (2016) An evaluation of Kolmogorov-Smirnov style statistics for bivariate circular data. The 3rd International Society for Nonparametric Statistics Conference, Avignon, 11-16 Jun 2016.

Quill, R., Sharples, J.J., Sidhu, L.A., N.S. Wagenbrenner, J.M. Forthofer (2016) Evaluation of operational wind field models over complex terrain. The 5th Fire Behaviour and Fuels Conference, Melbourne, 11-15 Apr 2016.

Quill, R., Sharples, J.J. and Sidhu, L.A. (2016) Analysing the impacts of vegetation and topography on wind fields over complex terrain. The 5th Fire Behaviour and Fuels Conference, Melbourne, 11-15 Apr 2016.

Quill, R., Sharples, J.J., Sidhu, L.A. and Piantadosi, J. (2016) Sensitivity analysis of Kolmogorov-Smirnov style statistics for univariate and bivariate data. ANZIAM 2016, Canberra, 7-11 Feb 2016.

Quill, R., Sharples, J.J. and Sidhu, L.A. (2015) Effects of post-fire vegetation regrowth on wind fields over complex terrain. MODSIM 2015, Gold Coast, 31 Nov-4 Dec 2015.

Quill, R. (2015) Developing a statistical characterisation of wind fields over complex terrain. Annual NSW/ACT ANZIAM Joint Meeting, Sydney, 25-26 Nov 2015.

Quill, R. (2015) Non-parametric comparison of regression surfaces to assess the impacts of vegetation regrowth on wind fields. ANZIAM 2015, Gold Coast, 2-5 Feb 2015.

The following academic posters showcased research from this thesis.

Quill, R. (2016) Statistical characterisation of wind fields over complex terrain for bush-fire modelling applications. AFAC, Brisbane 30 Aug-1 Sep 2016.

Quill, R., Sharples, J.J., Sidhu, L.A. and Thorpe, G. (2015) Evaluation of operational models for wind variability over complex terrain. AFAC, Adelaide 1-3 Sep 2015.

Quill, R., Sharples, J.J., Sidhu, L.A. and Thorpe, G. (2015) Impacts of vegetation regrowth on wind direction over complex terrain. AFAC, Adelaide 1-3 Sep 2015.

Finally, this piece of non-academic written work communicated some of the research highlights from this thesis to a fire industry audience.

Fire Modelling in an Uncertain World, by **Rachael Quill**. *Wildfire Magazine*, Volume 25.2, March/April 2016. <http://wildfiremagazine.org/article/fire-modeling-in-an-uncertain-world/>

Reproduced in *Fire Australia*, Winter 2016. <http://www.bnhcrc.com.au/news/2016/fire-modelling-uncertain-world-0>

Contents

Originality Statement	i
Copyright Statement	ii
Authenticity Statement	iii
Abstract	v
Acknowledgements	vii
Published Work	ix
Chapter 1 Introduction	1
1.1 Background	1
1.2 Research Contributions	2
1.3 Thesis Overview	3
Chapter 2 Literature Review	7
2.1 Fire Modelling in an Uncertain World	7
2.1.1 Forward Rate of Spread	8
2.1.2 Fire Spread Prediction	14
2.1.3 Uncertainty in Bushfire Prediction	16

2.1.4	Modelling Uncertainty	18
2.2	Wind Modelling in the Context of Bushfire	23
2.2.1	Operational Wind Field Models	27
2.2.2	Vertical Wind Profiles	29
2.3	Probabilistic Approaches in Wind Modelling	30
2.3.1	Joint Wind Direction Distributions	33
2.3.2	Parametric Estimation	35
2.3.3	Non-parametric Estimation	36
2.4	“The Literature Gap”	38
Chapter 3	Data Collection	41
3.1	Equipment and Software	42
3.1.1	Data Collection Equipment	42
3.1.2	Data Processing	45
3.2	Case Study I: Flea Creek Valley, NSW	46
3.2.1	Description	46
3.2.2	Data Summary, 2007	50
3.2.3	Data Summary, 2014	52
3.3	Case Study II: National Arboretum Canberra, ACT	57
3.3.1	Description	57
3.3.2	Data Summary, 2015	60
3.4	Concluding Remarks	62
Chapter 4	Probabilistic Representation of Wind Direction	67
4.1	Estimation of Toroidal Surfaces	69

4.1.1	Estimation Techniques	69
4.1.2	Simulation Study Data and Methods	73
4.1.3	Simulation Study Results	75
4.1.4	Analysis of the TPS Spline with Algorithm 1	80
4.1.5	Simulation Study Discussion	88
4.2	Estimated Directional Wind Response	91
4.2.1	Optimising Algorithm 1	91
4.2.2	Prevailing Wind Direction Analysis	93
4.2.3	Flea Creek Valley, 2007	95
4.2.4	Flea Creek Valley, 2014	96
4.2.5	National Arboretum Canberra, 2015	97
4.3	Concluding Remarks	98
Chapter 5 Assessment of the Impacts of Vegetation on Wind Fields over Complex Terrain		101
5.1	Background	103
5.1.1	Literature	103
5.1.2	Wind Data Comparisons	106
5.2	Univariate Wind Direction	113
5.2.1	Simulation Study	113
5.2.2	Conditional Wind Direction Distributions	130
5.3	Bivariate Wind Direction	134
5.3.1	Simulation Study	134
5.3.2	Joint Directional Wind Response Distributions	145
5.4	Concluding Remarks	151

Chapter 6	Evaluation of Wind Field Modelling over Complex Terrain	155
6.1	Predicted Wind Direction using a Diagnostic Wind Model	157
6.1.1	Background	157
6.1.2	Wind Direction Data and Modelling	160
6.1.3	Evaluation Results	166
6.1.4	Discussion	172
6.2	Wind Speed Reduction Beneath the Canopy	173
6.2.1	Background	173
6.2.2	Empirical Wind Speed Reduction Profiles	175
6.2.3	Evaluation Results	176
6.2.4	Discussion	179
6.3	Concluding Remarks	182
Chapter 7	Discussions and Conclusion	185
7.1	Wind Direction: A New Perspective	185
7.1.1	New Datasets	185
7.1.2	Directional Wind Response Distributions	187
7.2	Statistical Techniques for Analysing Wind Direction	188
7.2.1	Comparison Tests	188
7.2.2	Application of Statistical Techniques	190
7.3	Statistical Characterisation of Wind in the Context of Bushfire . . .	191
7.3.1	New Knowledge in the Existing Framework	191
7.3.2	New Modelling Approaches	192
7.4	Conclusion	193
7.5	Further Work	193

Appendix A Wind Data Summaries	195
A.1 Calibration Analysis	195
A.2 Flea Creek Valley 2007	197
A.3 Flea Creek Valley 2014	200
A.4 National Arboretum Canberra 2015	203
Appendix B Directional Wind Response	207
B.1 Flea Creek Valley, 2007	207
B.2 Flea Creek Valley, 2014	207
B.3 National Arboretum Canberra, 2015	207
Appendix C Matlab Code for Toroidal Estimation	213
Appendix D Matlab Code for Surface Comparison Tests	215
D.1 Simulation Procedure	215
D.2 Univariate KS Style Tests	216
D.3 Bivariate KS Style Tests	217
D.4 Mean Squared Difference Tests	219
References	221

List of Tables

3.1	Descriptions of the FCV sites (A1 to A5) in 2007. Adapted from Sharples et al. [2010, Table 1].	48
3.2	Descriptions of the FCV sites (B1 to B11) in 2014.	49
3.3	Summary of wind data collected across FCV between April and June 2007.	51
3.4	Summary of observed wind speeds (ms^{-1}) across FCV between April and June 2007.	51
3.5	Summary wind data collected across FCV between April and December 2014.	54
3.6	Summary of observed wind speeds (ms^{-1}) across FCV between April and December 2014.	54
3.7	Descriptions of the NAC sites (C1 to C11) in 2015.	60
3.8	Summary of wind data collected across the NAC between April and December 2015.	61
3.9	Summary of observed wind speeds (ms^{-1}) across the NAC between April and December 2015.	61
5.1	Sample sizes for FCV pairs.	110
5.2	Location descriptions for comparison pairs at the NAC in 2015. . .	111
5.3	Sample sizes for NAC pairs.	113

5.4	KS test results for univariate conditional wind direction distribution pairs from FCV and the NAC.	131
5.5	Kuiper’s test results for univariate conditional wind direction distributions pairs from FCV and the NAC.	132
5.6	Bivariate KS test results for directional wind response distribution pairs from FCV and the NAC.	145
5.7	Bivariate Kuiper’s test results for directional wind response distribution pairs from FCV and the NAC.	146
5.8	Bivariate squared difference test results (p -values) for directional wind response distributions pairs from FCV and the NAC.	148
6.1	WindNinja with native solver and WindNinja with momentum solver wind direction (WD, $^{\circ}$) and wind speed (WSp, ms^{-1}) predictions for each site across FCV.	165
6.2	Proportion of agreement between predicted wind direction (as compass point) and observed wind direction distribution at each site across FCV, conditional on observing a wind direction of WNW, and a minimum wind speed of T at B1.	166
6.3	Proportion of overlap between predictions and observations at 1-minute time steps. ‘Overlap’ is taken to be a prediction of wind direction in the same compass sector, within one sector (or $\pm 22.5^{\circ}$) or within two sectors (or $\pm 45^{\circ}$).	171
6.4	Average wind speeds for each site across FCV, at minimum wind speed thresholds, T , observed at A1 in 2007 and B1 in 2014.	176
6.5	Average wind speeds across the NAC, at minimum wind speed thresholds, T , observed at C1.	179
A.1	Proportion of winds originating from each compass direction across FCV between April and June 2007.	198

A.2	Proportion of winds originating from each compass direction across FCV between April and December 2014.	203
A.3	Proportion of winds originating from each compass direction across the NAC between April and December 2015.	204

List of Figures

2.1	Observed joint directional wind response distribution represented (a) in planar view and (b) on the torus. Data collected by author. . . .	34
3.1	Davis Vantage Pro 2 Portable Automatic Weather Stations during the calibration phase; (a) set up at the UNSW Canberra Field Site, and (b) sensor array and console box. Photographs taken by author.	43
3.2	Data download using (a) Weatherlink® , and (b) the Raspberry Pi® system. Photographs taken by author.	44
3.3	Photographs of data collection at FCV in (a) 2007 and (b) 2014. Photographs taken by (a) J.J. Sharples and (b) author.	44
3.4	Maps of (a) south-eastern Australia, (b) the Brindabella region and (c) the FCV transect (with 50m contours). Station locations in 2007 (red; A1 to A5) and 2014 (blue; B1 to B11) are indicated.	47
3.5	Observed wind direction roses across FCV at (a-e) A1 to A5 between April and June 2007.	53
3.6	Observed wind direction roses across FCV at (a-k) B1 to B11 between April and December 2014.	55
3.6	(Cont'd) Observed wind direction roses across FCV at (a-k) B1 to B11 between April and December 2014.	56
3.7	Map of (a) south-eastern Australia, (b) the Canberra region, and (c) the NAC. Station locations in 2015 (C1 to C11) are indicated. . . .	58

3.8	Photograph of data collection at the NAC in 2015. Photograph taken by author from Dairy Farmer’s Hill.	59
3.9	Observed wind direction roses across the NAC at (a-k) C1 to C11 between April and December 2015.	63
3.9	(Cont’d) Observed wind direction roses across the NAC at (a-k) C1 to C11 between April and December 2015.	64
4.1	Conceptual diagram of data collection across FCV in 2014, and the observed directional wind response on the (a) east-facing slope at B4, (b) valley floor at B6 and (c) west-facing slope at B10.	68
4.2	Schematic diagram of Algorithm 1 with one iteration ($r = 1$) applied to a 2×2 data grid ($n = 2$).	73
4.3	Continuous surface f_A , with $\mu_A = \tau_A = 90^\circ$	74
4.4	Average time (seconds) taken for individual approximations for each estimation method applied to Y_A . Diamonds represent (orange) Cyclic Cubic and (blue) KDE von Mises. Lines represent (orange) Cubic, (green) TPS and (blue) KDE Normal, in conjunction with Algorithm 1.	76
4.5	Average (solid) Mean Squared Error, and 95% CI (dotted), for each estimation method applied to Y_A , over 250 MC simulations. Diamonds represent (orange) Cyclic Cubic and (blue) KDE von Mises. Lines represent (orange) Cubic, (green) TPS and (blue) KDE Normal, in conjunction with Algorithm 1.	78
4.6	Average (solid) and 95% CI (dotted) for the (left) mean difference in edge values and (right) mean difference in edge derivatives, over 250 MC simulations, for (a,b) Cubic, (c,d) KDE Normal and (e,f) TPS, applied to Y_A . Metrics calculated row-wise (red) and column-wise (blue). Note that the figures contain different scales when compared by column.	79

4.7	Continuous surfaces (a) f_B ($\mu_B = \tau_B = 45^\circ$), and (b) f_C ($\mu_C = \tau_C = 180^\circ$).	81
4.8	Average (solid) and 95% CI (dotted) for the (a) mean difference in edges values, (b) mean difference in edge derivatives and (c) MSE values, over 250 MC simulations, for the application of the TPS spline to Y_B , with $\mu_B = \tau_B = 45^\circ$. Metrics calculated row-wise (red) and column-wise (blue).	82
4.9	Average (solid) and 95% CI (dotted) for the (a) mean difference in edge values, (b) mean difference in edge derivatives, and (c) MSE values, over 250 MC simulations, for the application of the TPS spline to Y_C . Metrics calculated row-wise (red) and column-wise (blue).	83
4.10	Average (solid) and 95% CI (dotted) for the (a) mean difference in edge values, (b) mean difference in edge derivatives and (c) MSE values, over 250 MC simulations, for the application of the TPS spline, with $p = 0.001$, to Y_A . Metrics calculated row-wise (red) and column-wise (blue).	84
4.11	Average (solid) and 95% CI (dotted) for the (a) mean difference in edge values, (b) mean difference in edge derivatives and (c) MSE values, over 250 MC simulations, for the application of the TPS spline, with $p = 0.1$, to Y_A . Metrics calculated row-wise (red) and column-wise (blue).	85
4.12	Average (solid) and 95% CI (dotted) for the (a) mean difference in edge values, (b) mean difference in edge derivatives and (c) MSE values, over 250 Monte Carlo simulations, for the application of the TPS spline to Y_A , with 10×10 discrete grid and 360×360 quasi-continuous grid. Metrics calculated row-wise (red) and column-wise (blue). Note the change of scale in (c).	86

4.13	(a) Mean difference in edge values, (b) mean difference in edge derivatives, and (c) MSE values, for a single simulation, for the application of the TPS spline to Y_A , with 36×36 discrete grid and 360×360 quasi-continuous grid. Metrics calculated row-wise (red) and column-wise (blue).	87
4.14	(a) Mean difference in edge values, (b) mean difference in edge derivatives, and (c) MSE values, for a single simulation, for the application of the TPS spline to Y_A , with 16×16 discrete grid and 720×720 quasi-continuous grid. Metrics calculated row-wise (red) and column-wise (blue).	88
4.15	Mean difference in (top) edge values and (bottom) edge derivatives of the estimated directional wind response distributions across FCV in 2014, using the TPS spline, on the (a,b) east-facing slope at B4, (c,d) valley floor at B6 and (e,f) west-facing slope at B10. Note that the figures contain different scales when compared by column. . . .	92
4.16	Comparison between wind direction observed at Mount Ginini and FCV at (a) A1 in 2007, (b) B2 in 2014 and (c) B1 in 2014, as well as between Canberra Airport and NAC at (d) C1 in 2015. Red lines indicate isolines on the estimated surfaces.	94
4.17	Estimated directional wind response distributions from FCV in 2007, at (a) A2, (b) A3, (c) A4 and (d) A5. Red lines indicate isolines on the estimated surfaces.	95
4.18	Estimated directional wind response distributions from FCV in 2014, at (a) B4, (b) B6, (c) B8 and (d) B10. Red lines indicate isolines on the estimated surfaces.	96
4.19	Estimated directional wind response distributions from the NAC in 2015, at (a) C2, (b) C5, (c) C7 and (d) C9. Red lines indicate isolines on the estimated surfaces.	97
5.1	Map of coincidental station locations across FCV between 2007 (red; A1 to A5) and 2014 (blue; B1, B4, B6, B8 and B10).	107

5.2	Observed conditional wind direction distributions from 2007 at (a) FCV1 on the leeward slope (A2), (b) FCV2 on the valley floor (A3), (c) FCV3 on the lower windward slope (A4), and (d) FCV4 on the upper windward slope (A5), assuming a WNW observed at A1.	108
5.3	Observed conditional wind direction distributions from 2014 at (a) FCV1 on the leeward slope, (B4) (b) FCV2 on the valley floor (B6), (c) FCV3 on the lower windward slope (B8), and (d) FCV4 on the upper windward slope (B10), assuming a WNW observed at B1.	109
5.4	Map of the NAC with station locations from 2015 (yellow; C1, C2, C5, C7 and C9).	111
5.5	Observed conditional wind direction distributions from the NAC at (a) C2 on the clear leeward slope, (b) C5 on the pine leeward slope, (c) C7 on the pine 45° leeward slope and (d) C9 on the pine cross slope.	112
5.6	Power of the univariate KS test to (a) percentage change in mean and (b) relative change in standard deviation for the Normal distribution, at a 1% level.	122
5.7	Power of the univariate Kuiper's test to (a) percentage change in mean and (b) relative change in standard deviation for the Normal distribution, at a 1% level.	124
5.8	Approximations for the asymptotic p -values for the univariate (a) KS test (Equation 5.4) and (b) Kuiper's test (Equation 5.8).	125
5.9	Power of the univariate Kuiper's test to (a) percentage change in mean and (b) relative change in concentration for the von Mises distribution with centre mode, at a 1% level.	127
5.10	Power of the univariate Kuiper's test to (a) percentage change in mean and (b) relative change in concentration for the von Mises distribution with edge mode, at a 1% level.	128

5.11	Power of the bivariate KS test to (a) percentage change in mean and (b) relative change in standard deviation for the bivariate Normal distribution, at a 1% level. Lines represent mean changes in the y direction. Triangles represent mean changes in the x direction. . . .	139
5.12	Power of the bivariate Kuiper's test to (a) percentage changes in mean and (b) relative changes in standard deviation for the bivariate Normal distribution, at a 1% level. Lines represent mean changes in the y direction. Triangles represent mean changes in the x direction.	142
5.13	Power of the bivariate Kuiper's test to (a) percentage changes in mean and (b) relative changes in concentration for the bivariate von Mises distribution, at a 1% level. Lines represent mean changes in the y direction. Triangles represent mean changes in the x direction.	143
6.1	Observed average hourly wind directions at each site across FCV, at (a) 0300h, (b) 0900h, (c) 1500h and (d) 2100h.	161
6.2	WindNinja with native solver prediction over FCV, using domain averaged wind speeds of 1ms^{-1}	164
6.3	WindNinja with momentum solver prediction over FCV, using domain averaged wind speeds of 1ms^{-1}	164
6.4	Observed conditional wind direction distributions at (a) B4 on the east-facing slope, (b) B7 on the valley floor, (c) B8 on the lower west-facing slope and (d) B11 on the upper west-facing slope, assuming a WNW observed at B1. Predictions shown for the native solver (red) and the momentum solver (green).	167
6.5	Observed (blue) and predicted (red) unconditional wind direction distributions at (a) B1 on the western ridge top, (b) B4 on the east-facing slope, (c) B7 on the valley floor, (d) B8 on the lower west-facing slope and (e) B11 on the upper west-facing slope.	170

6.6	Relative wind speed observed between 2007 and 2014 at (a) FCV0 on the ridge top, (b) FCV1 on the east-facing slope, (c) FCV2 on the valley floor, (d) FCV3 on the lower west-facing slope and (e) FCV4 on the upper west-facing slope. Findings of Moon et al. [2013, In Press] are indicated with red lines; RWS = 1.00 (solid), and RWS = 0.82 (dotted).	178
6.7	Relative wind speed observed between cleared east-facing slope and the parallel Radiata pine stand at the National Arboretum Canberra for (a) N0 on the ridge top, (b) N1 on the mid-slope and (c) N2 on the lower slope. Findings of Moon et al. [2013, In Press] are indicated with red lines; RWS = 0.40 (solid), RWS = 0.20 (dashed), and RWS = 0.08 (dotted).	180
A.1	Absolute percentage difference between PAWS and SA _{5m} wind speed observations between 19th and 23rd February 2014. Red line shows SA _{5m} ± 1ms ⁻¹	196
A.2	Wind direction time series for PAWS (blue dots) and SA _{5m} (red line) between 19th and 23rd February 2014.	197
A.3	Data records across FCV between April and December 2007.	197
A.4	Time series of observed wind speeds (ms ⁻¹) across FCV at (a-e) A1 to A5 between April and June 2007.	199
A.5	Data records across FCV between April and December 2014.	200
A.6	Time series of observed wind speed (ms ⁻¹) across FCV at (a-k) B1 to B11 between April and December 2014.	201
A.6	(Cont'd) Time series of observed wind speed (ms ⁻¹) across FCV at (a-k) B1 to B11 between April and December 2014.	202
A.7	Data records at the NAC between April and December 2015.	203
A.8	Time series of observed wind speeds (ms ⁻¹) across the NAC at (a-k) C1 to C11 between April and December 2015.	205

A.8 (Cont'd) Time series of observed wind speeds (ms^{-1}) across the NAC at (a-k) C1 to C11 between April and December 2015.	206
B.1 Observed discrete directional wind response distributions across FCV at (a-l) A2 to A5, with minimum wind speed thresholds of $T = 0\text{ms}^{-1}$, $T = 2\text{ms}^{-1}$ and $T = 4\text{ms}^{-1}$ at A1, between April and June 2007.	208
B.2 Observed discrete directional wind response distributions across FCV at (a-o) B2 to B6, with minimum wind speed thresholds of $T = 0\text{ms}^{-1}$, $T = 2\text{ms}^{-1}$ and $T = 4\text{ms}^{-1}$ at B1, between April and December 2014.	209
B.3 Observed discrete directional wind response distributions across FCV at (a-o) B7 to B11, with minimum wind speed thresholds of $T = 0\text{ms}^{-1}$, $T = 2\text{ms}^{-1}$ and $T = 4\text{ms}^{-1}$ at B1, between April and December 2014.	210
B.4 Observed discrete directional wind response distributions across the NAC at (a-o) C2 to C6, with minimum wind speed thresholds of $T = 0\text{ms}^{-1}$, $T = 2\text{ms}^{-1}$ and $T = 4\text{ms}^{-1}$ at C1, between April and December 2015.	211
B.5 Observed discrete directional wind response distributions across the NAC at (a-o) C7 to C11, with minimum wind speed thresholds of $T = 0\text{ms}^{-1}$, $T = 2\text{ms}^{-1}$ and $T = 4\text{ms}^{-1}$ at C1, between April and December 2015.	212

CHAPTER 1

Introduction

To improve bushfire management and mitigation, and to provide more accurate, timely and comprehensive warnings to the community, a sound basis for modelling the propagation of bushfires is essential. The spread of bushfires is highly sensitive to wind speed and direction; in particular, sudden changes in wind characteristics across the landscape can result in drastic changes in the rate and direction of fire spread. Accurate estimation of wind fields across the landscape and over time is thus a crucial component of bushfire spread modelling. However, estimation of wind fields across rugged terrain is far from trivial. The complex interactions between prevailing winds and varying landscape features are often over-simplified within operational bushfire models due to computational constraints.

The research presented in this thesis aims to improve the understanding of wind flow over complex terrain through developing a statistical characterisation. This improved understanding will complement current approaches to provide a more informed wind modelling framework, including considerations of variability and uncertainty, for bushfire prediction.

1.1 Background

In 2003, the western region of the Australian Capital Territory, and adjoining areas of New South Wales, experienced devastating bushfires. The fires reached into the suburbs of Canberra, destroying 500 homes and claiming 4 lives, and significant areas of vegetation were destroyed across mountain ranges west of the city. In the years since, the Canberra 2003 bushfires have become arguably some of the

most scientifically significant bushfires ever experienced in Australia. Great strides forward have been made in the understanding of extreme and volatile bushfire behaviour. For instance, it was noted that on a number of occasions the fire would spread in a direction perpendicular to the prevailing wind, a phenomenon now known as vorticity-driven lateral spread [Sharples et al., 2010, Simpson et al., 2013].

To better understand some of the fire behaviour phenomena that occurred in 2003, wind characteristics were measured in a number of case studies in the region in 2007. One such case study, Flea Creek Valley, sits within the Brindabella National Park in New South Wales, and is very close to where some of the 2003 fires ignited and began rapidly propagating towards Canberra [Sharples et al., 2010]. After over a decade of vegetation regrowth across the region (with no major fires occurring since 2003), in 2014 wind characteristics were again measured across Flea Creek Valley in order to develop an understanding of how changing vegetation impacts wind flow over complex landscapes.

To develop a statistical characterisation of wind fields across the landscape, wind direction was recast in terms of probability distributions that represent what is referred to as ‘directional wind response’. Directional wind response distributions were constructed as the joint distributions of concurrent prevailing wind directions and those observed at the surface within the valley. These distributions statistically describe the response of the prevailing wind direction to changes in the surface landscape.

1.2 Research Contributions

As described previously, the overarching aim of this thesis is to better understand wind flow over complex terrain from a statistical perspective, in the context of bushfire prediction. This thesis develops four key contributions to research across the fields of statistics, applied mathematics and bushfire modelling.

1. Provide new wind field datasets from regions of undulating and complex terrain at resolutions relevant to surface fire spread and behaviour.

2. Develop a statistical representation of directional wind response in the landscape through understanding approximation of toroidal surfaces using an algorithmic adaptation of planar techniques.
3. Understand the impacts of topography and vegetation on directional wind response through the employment of statistical comparison techniques.
4. Utilise a statistical representation and understanding of wind response across complex terrain to evaluate wind field models used in current operational fire prediction.

This thesis will complement the range of available wind field models by providing datasets for input and validation, as well as comparative methods for improving accuracy, reducing computational requirements and understanding uncertainty. The proposed statistical approaches permit better accommodation of the uncertainties inherent in the structure of near-surface wind fields over complex terrain.

In addition, the probabilistic characterisation of wind fields is more naturally suited to the emerging state-of-the-art fire propagation prediction systems which incorporate ensemble and stochastic approaches by evaluating numerous spread scenarios under a range of different conditions. These forecasting methods allow fire managers to gain an understanding of the scope of possible spread scenarios, and by directly considering the associated uncertainties, managers can have confidence in the use of ‘best guess’ predictions. Moreover, these uncertainties can be used to understand the consequences and impacts of outliers in the scenario set, such as ‘worst case’ scenarios, particularly when these outliers may result in significantly different impacts compared to the single ‘most likely’ prediction.

1.3 Thesis Overview

In addressing the aim of the research, this thesis is structured in accordance with the key scientific contributions detailed in the previous section.

Chapter 2 firstly reviews the current literature relevant to this thesis. This chapter provides a detailed background of the development of bushfire modelling in Australia and across the world. The current suite of wind modelling approaches used in

bushfire prediction are reviewed, before probabilistic approaches to wind modelling available from other fields of research are also considered.

Chapter 3 describes the data collected for analysis throughout this thesis, achieving the first key contribution. Two regions are studied; as previously detailed, Flea Creek Valley is the motivating case study for this research. Data analysed by Sharples et al. [2010] are considered in conjunction with newly collected data from across the valley. The second case study is based at the National Arboretum Canberra, which provides a topographical and vegetative landscape that allows for more controlled statistical analyses of wind fields over varying surface conditions.

Chapter 4 begins to recast wind direction data in probabilistic terms; developing the second key contribution. Directional wind response is defined, and the chapter addresses the question of how to estimate true directional wind response given noisy observed data. The toroidal nature of joint wind direction distributions provides more complex conditions for estimation. The efficacy of a simple algorithmic adaptation to traditional planar techniques in estimating toroidal surfaces is investigated in this chapter. A simulation study is presented to compare this approach to existing toroidal estimation methods.

In developing the third key contribution of this research, Chapter 5 presents a study of statistical techniques that are available for the comparison of directional wind response distributions. A simulation study is used to better understand the power of these techniques against known alternative distribution structures. The results of the simulation study allow for better interpretation of statistical comparisons between observed datasets. Application to the case studies across Flea Creek Valley and the National Arboretum Canberra investigate the impacts of changing vegetation or differing topography on directional wind response over the landscape.

Chapter 6 develops the fourth contribution of this research through an evaluation of current wind modelling techniques that are used within fire prediction frameworks in Australia. This chapter deals with two elements of wind flow across complex terrain that are directly input into bushfire models: wind direction and vegetation-induced wind speed reduction. Firstly, the probabilistic representations of wind direction considered throughout this thesis are used to evaluate the deterministic

wind model currently used in operational fire modelling. The evaluation shows the potential for probabilistic application of such physics-based models to better capture variation and uncertainty in wind direction. Secondly, wind speed reduction profiles developed over flat terrain are tested against data collected for this study over complex and undulating terrain. This evaluation aims to better understand the applicability of such empirical profiles across the broader landscape and highlight areas that require further research.

Finally, Chapter 7 draws together the core findings of this thesis. The discussion is framed in terms of bushfire modelling and prediction, as well as within the broader context of applied statistics and mathematics. Conclusions are drawn in relation to the overarching aim of this thesis and some examples are given of research areas which could be further explored.

CHAPTER 2

Literature Review

To understand the need to better characterise wind fields over complex terrain, it is important to contextualise the issue within the history of bushfire modelling in Australia and across the globe. The models used operationally today have developed throughout the 20th century and into the 21st. Although modelling frameworks have in many ways taken advantage of the technological advancements made in computing during that time, the fundamental scientific underpinning of fire dynamics and how it spreads across the landscape has remained relatively unchanged. The first section of this chapter outlines these developments and clearly highlights the role that wind modelling has in accurately predicting fire spread and behaviour. Current wind models used in fire research and operational prediction are then discussed in Section 2.2.

In emerging literature, the need to deal with uncertainty in fire prediction has become evident, and through statistical and probabilistic approaches, fire modelling frameworks are now beginning to deal with the issue. Rigorous scientific understanding of this uncertainty in relation to fire model outputs, as well as modelling inputs is still, however, in its infancy. Section 2.3 therefore provides some discussion on how probabilistic approaches to wind modelling can help to better characterise uncertainty in fire prediction.

2.1 Fire Modelling in an Uncertain World

Bushfire modelling has been developing across the world since the turn of the 20th century. In the early 1900s, fire services in the USA began to develop systematic

forms of fire management and control, from regulating burning of slashings and penalising camp-fires to providing sufficient fire patrols throughout forest areas [Beals, 1914]. Concurrently, weather researchers began to investigate the idea of predicting forest fire spread. Beals [1914] appears to have been one of the earliest authors to consider the use of wind forecasts in understanding the potential for large fire spread, with examples across the USA used to conceptually verify his ideas. Given the prevalence of bushfires (or wildfires), the most prominent research in the field has originated from Australia and Northern America [Pastor et al., 2003]. Fire affected European countries such as Russia and Portugal have also contributed significantly to the field.

Mathematical fire models began to develop from the mid-20th century; Pastor et al. [2003] gives a comprehensive review of such models developed between 1946 and 2003. The fundamental models for calculating forward rate of spread developed in the USA and Australia were constructed using both physical and empirical approaches [McArthur, 1966, 1967, Rothermel, 1972]. With these as a basis, fire models have developed by incorporating more detailed physical and numerical modelling frameworks. State-of-the-art fire models now sit within complex decision support tools capable of predicting not only forward rate of spread but numerous other variables such as fire perimeter, spotting distances and risk of damage to property [Tolhurst et al., 2008]. Applications of probabilistic fire modelling approaches, such as those outlined by Cruz [2010], have made significant progress in recent years, but discussions of model validation and uncertainty are still in their infancy within both the research and operational bushfire communities.

2.1.1 Forward Rate of Spread

Two core models make up much of the fundamental bases for fire modelling across the globe. The empirical models developed by McArthur [1967] were based on the analysis of small experimental fires conducted in native Australian eucalypt forests, as well as a limited collection of wildfires. Whereas, the semi-physical model developed by Rothermel [1972] was based on the physical principles of fire ignition and propagation, and empirically parameterised using experimental and laboratory fires.

Taken to be the foundation of bushfire modelling in Australia, McArthur [1967] empirically developed the Forest Fire Danger Meter (FFDM; the Mark 5 meter is now in operational use) using data from over 800 small experimental fires burnt for between 15 and 60 minutes through a variety of natural eucalypt forests. A number of wildfires were also analysed, including the 1939 fires which formed the benchmark for the most extreme prediction outcomes. The meter (a circular slide-rule) facilitates the calculation of a Forest Fire Danger Index (FFDI) ranging from 0 to 100, which is translated into nominal categories for communication of potential fire risk to the public. The meter allows for the calculation of the forward rate of spread of a fire as well as fire behaviour characteristics such as flame height and average spotting distance. Additionally, the meter reflects information on likelihood of ignition, difficulty of suppression and potential damage caused by the fire.

A decade later, Noble et al. [1980] empirically fitted equations to the FFDM outputs, enabling its use within automated fire modelling systems. For the Mark 5 FFDM, the FFDI is given by

$$\text{FFDI} = 2 D^{0.987} \exp(-0.45 + 0.0338 T - 0.0345 H + 0.0234 V), \quad (2.1)$$

where V is the wind speed (kmh^{-1}) measured at 10 metres (≈ 33 feet) above clear ground, T is temperature ($^{\circ}\text{C}$), H is relative humidity (%) and D is the Drought Factor. The forward rate of spread of a fire (R , kmh^{-1}) was empirically formulated by Noble et al. [1980] using the FFDI and the fuel weight (W , t ha^{-1}),

$$R = 0.0012 \text{FFDI} \cdot W. \quad (2.2)$$

The approximations suggest that the core factors in fire danger and spread are wind speed and the characteristics of the fuel. However, the meter and equations only convey empirical correlations between environmental variables and rate of spread; they do not provide any insight into the physical processes driving fire propagation.

The exact relationship between fire behaviour and wind velocity is complex. It was shown by McArthur [1967] that rate of spread would increase in proportion to the square of wind velocity, but Noble et al. [1980] incorporated it into the

exponential for FFDI. For meteorological standards, wind speeds were recorded at 10m (≈ 33 ft) above the ground in the open and included as such in Equation 2.1. However, McArthur [1967] recognised that for fire behaviour research it was standard to measure wind speeds at around 1.5m (≈ 5 ft) above ground; winds at this height have a direct influence on the flames of a surface fire [Rothermel, 1972]. McArthur [1967] therefore presented vertical profiles of wind speeds to translate 10m winds to 1.5m winds in three different forest types.

In Equation 2.1, the Drought Factor (D) represents a measure of fuel availability, dependent upon the Keetch-Byram Drought Index [Keetch, 1966], days since last rain and the current rainfall. Temperature and relative humidity are used as proxy for fuel moisture content which represents the proportion of the available fuel that will burn. From the meter, fuel type was also expected to impact on rate of spread such that fires in low quality or open forest areas would be expected to spread faster than estimated since wind speeds would be less impeded, however spotting potential would be lower. Finally, McArthur [1967, page 10] also suggested that as fuel quantity doubled so did rate of spread, recognising that this was perhaps the “only factor over which man can exercise some control”.

For the adjustment of rate of spread under different slope conditions, McArthur [1967] suggested that the rate of spread would increase by a factor of two for every 10° incline in slope (up to 20°). Noble et al. [1980] incorporated such an adjustment through an exponential in terms of the slope in degrees (θ , $^\circ$),

$$R_\theta = R \exp(0.069 \theta). \quad (2.3)$$

It is assumed by Noble et al. [1980] that the same function would also apply down-slope, i.e. for negative θ , but Sullivan et al. [2014] highlighted that this should not in fact be the case. Sullivan et al. [2014] outlined a correction factor for down-slope rates of spread based on the assumption that the rate of spread of a large fire across undulating terrain could be well approximated by the rate of spread across flat terrain. This correction factor could then be calculated for a number of different rate of spread models.

In 2001, the Australian Commonwealth Scientific and Industrial Research Organisation (CSIRO) began Project Vesta [Gould, 2008], which empirically developed look-up tables and equations for fire characteristics as a replacement for the McArthur FFDM as well as developing a more detailed understanding of the relationships between fire behaviour and environmental conditions. These relationships were developed through experimental burns, followed by regression analysis between fire behaviour characteristics and fuel and weather variables. Results showed a significant, and approximately linear, relationship between rate of spread and wind speed [Gould, 2008], rather than the exponential relationship given by Noble et al. [1980].

McArthur [1966] also outlined a Grassland Fire Danger Meter (GFDM), again structured as a circular slide-rule. The GFDM facilitated the calculation of a Grassland Fire Danger Index (GFDI) and the related rate of spread, as well as provided information on the difficulty of suppression, area burnt and flame height for a grassland fire. The GFDM incorporated slightly different input variables to those of the FFDM to capture the different processes governing fire spread through grassland landscapes.

Since grasslands are fully open to the force of the wind, wind velocity was considered the most significant variable in determining the shape of the fire. However, the fuel moisture content of the grass was considered one of the most important factors in determining fire danger, thus rate of spread and likelihood of ignition [McArthur, 1966]. As in the FFDI, fuel moisture was represented using temperature and relative humidity, but fuel availability was considered in the GFDI by incorporating the grass curing process rather than the Drought Index. The assumption made for the Mark 4 GFDM (still in operation) is that fuel weight is approximately $3\text{-}4\text{t ha}^{-1}$ for grassland regions, and so the variable is not explicitly incorporated into the equations derived for the meter [Cruz et al., 2015]. The rate of spread is then represented as a simple linear function of GFDI [Noble et al., 1980],

$$R = 0.13 \text{ GFDI}. \tag{2.4}$$

As for the FFDI, Equation 2.1, the impact of wind on rate of spread of a grassland fire was incorporated through an exponential term. There were no explicit adjust-

ment equations given by Noble et al. [1980] for rate of spread of grassland fire up a slope, but the GFDM again suggests that the rate of spread would increase by a factor of two for each 10° increase in slope, up to 20° [McArthur, 1966].

In the 1990s, Cheney et al. [1993] re-expressed the average rate of spread (ms^{-1}) of a propagating grass fire directly in terms of wind speed and fuel moisture content;

$$R = a U_2^b \exp(c M_{fp}) \quad (2.5)$$

where U_2 is the wind speed measured at 2 metres above open ground (ms^{-1}) and M_{fp} is the predicted fuel moisture content (%). Furthermore, Cheney et al. [1998] suggested that the behaviour of fire spread would change when wind speed passed a critical threshold of 5kmh^{-1} (measured at 10m above the ground in the open; U_{10}). The type of grassland was also classified and taken into account through multiple parameterisations of the renewed grassland fire model.

Across Australia, a number of other fuel types exist, including heathland, mallee forest and pine. Cruz et al. [2015] provides a comprehensive review of rate of spread models developed across these various vegetation types. Consistently, it is shown that fuel characteristics and weather variables are strongly related to the spread of fires through the landscape in Australia. In particular, the relationship to wind speed is commonly defined in terms of a power law or exponential. For example, a number of models for Australian shrublands, including heath, mallee and buttongrass show rate of spread to be proportional to U^a where $a \in [1.0, 1.4]$, with wind speed (U) measured at either 2m or 10m [e.g. Marsden-Smedley and Catchpole, 1995, Catchpole et al., 1998a, McCaw, 1997].

The fire danger rating system in Australia is continuing to develop as new understanding emerges and as fire behaviours continue to adapt to changing environmental conditions, e.g. climate change and an increased peri-urban population. In 2009, after the Black Saturday fires in Victoria, the extreme environmental conditions experienced over that period forced an extension of the FFDI categorisation system to include a ‘Catastrophic’ category representing danger indices greater than 100.

In Australia, the small experimental fires conducted by McArthur, and analysis of a few wildfires, have remained the basis of bushfire modelling for almost half a century. In fact, although explicitly designed for eucalypt forest and south-eastern Australian grasslands, the McArthur meters have been applied in many regions [Noble et al., 1980]. These models have been shown to consistently under-predict forward rate of spread by a factor of two or more, particularly for large fires or under severe conditions [Burrows, 1994]. While for grassland fires, models have been updated to utilise the CSIRO grassland equations [Cheney et al., 1993, 1998], and despite new insights from Project Vesta, the McArthur meters and equations for forest fires are still widely operational today.

In contrast to the empirical approaches taken in Australia, a semi-physical model developed by Rothermel [1972] has been considered the basis for modelling forward rate of spread in the USA for the past half-century. Rothermel's model built on the work completed by Fons [1946] and the equations set out by Frandsen [1971], assuming that a fire propagated as a series of ignitions and defining the rate of fire spread as proportional to the heat provided to a patch of fuel divided by the amount of heat required to ignite that area.

While the effect of wind and slope on rate of spread remains qualitatively similar across the set of fire rate of spread models, the relationships are incorporated using different functional forms. For the McArthur meters and equations, wind speed forms a multiplicative factor in the calculation of FFDI and GFDI, and slope is accounted for through a multiplicative adjustment [Noble et al., 1980]. Rothermel [1972] accounted for wind and slope using additive adjustment factors, ϕ_W and ϕ_S respectively. Essentially, the adjusted rate of spread is equal to the no-wind and flat-ground rate of spread multiplied by $(1 + \phi_W + \phi_S)$.

With the assumption of no-wind and flat-ground (i.e. $\phi_W = \phi_S = 0$), the parameterisation of the physical model was based on empirical data from laboratory experiments [Rothermel, 1972]. The wind coefficient and slope factor were parameterised using wind tunnel experiments and the empirical results from McArthur's investigations in Australia [McArthur, 1967, 1969]. Rothermel [1972] suggested that both wind and slope increase the rate of spread through exposing the potential fuel

ahead of the fire to more radiative and convective heat. The wind coefficient was defined by a power law in terms of wind velocity at ‘mid-flame’ height (rather than at 10m, used by McArthur [1967]), while the slope factor was naturally a function of the slope angle. Both factors were also related to fuel parameters, including fuel particle surface-area-to-volume ratio and the packing ratio of the fuel bed.

Rothermel [1972] applied the model using a number of fuel models representing a range of North American fuel types, from grasslands through to timber forests. The idealised input parameters showed a range of potential spread scenarios and highlighted that the model would be useful within fuel management decision-making processes.

2.1.2 Fire Spread Prediction

The models of both McArthur [1966, 1967] and Rothermel [1972] predict forward rate of spread assuming that the fire will propagate predominantly in a direction defined by a combination of the directional effects of wind and slope [Sharples, 2008]. In addition, both models facilitate the derivation of a number of fire behaviour characteristics, including rate of increase for the fire perimeter. McArthur [1966, 1967] suggested that the perimeter of forest fires would increase at three to four times the forward rate of spread, while the perimeter of a grass fire would increase at a rate two and a half times that of the forward rate of spread. Even prior to the fire danger meters, Mitchell [1937] suggested that as a rule of thumb, the rate of increase of a fire perimeter would be three times the forward rate of spread.

These rules of thumb are still utilised alongside substantial field experience by fire fighters and fire behaviour analysts to manually predict fire spread across the landscape. However, since the turn of the century, and with advancements in computational capacity, forward rate of spread models have been integrated into two-dimensional fire prediction systems which predict the spatial spread of fires and their developing behaviour. These systems are now capable of incorporating complex environmental conditions.

In Australia, the most commonly used state-of-the-art fire prediction tool for research and operations is known as Phoenix Rapidfire [Tolhurst et al., 2008]. The

framework contains a number of fire spread and behaviour models to account for numerous fuel types, including the empirically developed McArthur Mark 5 Meter [McArthur, 1967], the equations developed by Project Vesta [Gould, 2008] and the formulae given by Beck [1995] for forest fuels in Western Australia. Phoenix Rapidfire also incorporates CSIRO grassland fire spread models, including those given by Cheney et al. [1998], as well as fire spread models for heathland vegetation [Catchpole et al., 1998a], buttongrass [Marsden-Smedley, 1993, Marsden-Smedley and Catchpole, 1995], pine [Cruz and Fernandes, 2008, Cruz et al., 2008] and mallee [McCaw, 1997, 1998]. The fire behaviour for Phoenix Rapidfire is therefore tuned to experimental fires or relatively small wildfires using 10m open wind speeds as input, which is modelled using a number of downscaling wind models discussed in Section 2.2.

Within the framework, one-dimensional forward rate of spread is spatialised using Huygens' principle [Richards, 1993], based on a geometric template dependent on wind and slope. Phoenix Rapidfire therefore allows for operational prediction of fire spread over spatial grid sizes between 100m and 200m in resolution, at time intervals of between 1 and 15 minutes [Tolhurst et al., 2008]. Similar frameworks used across the world also incorporate Huygens' principle to spatialise one-dimensional forward rate of spread, and therefore show similar dependence and sensitivities to wind inputs. For example, in the Canadian fire growth simulation system, known as Prometheus [Tymstra et al., 2010], the spatial spread of fire is determined through principles of elliptical fire growth defined by environmental parameters.

In the USA, FARSITE is the key framework used for fire spread modelling both operationally and within research [Finney, 2004]. It is based on the fundamentals from Rothermel [1972] and again uses Huygens' principle to spatialise one-dimensional fire spread models. The Wildland Fire Dynamics Simulator (WFDS) [US Forest Service, 2013, Mell et al., 2009] is also used to model detailed fire dynamics for fire behaviour research. The model was initially developed for building fires and incorporates detailed computational fluid dynamics and combustion models to predict fire behaviour, but it is therefore too computationally intensive for operational use. Within fire behaviour research, WFDS has been utilised to better understand the

interactions between bushfires and the urban environment at what is termed the Wildland-Urban Interface (WUI) [Rehm and Mell, 2009].

2.1.3 Uncertainty in Bushfire Prediction

In an investigation into the uncertainty of fire rate of spread prediction models, Cruz and Alexander [2013] defined an exact prediction as one within a $\pm 5\%$ error band around an observed value. With this definition, it was found that only 3% of rate of fire spread predictions (35 out of 1278) could be considered exact, with most models under-estimating rate of spread. In fact, Cruz and Alexander [2013, page 20] suggested that “one could argue that perhaps the only certainty about wildland fire behaviour prediction is that it is extremely unlikely that a prediction will match the observed fire behaviour characteristics”.

In practice, while McArthur [1977] “felt” that the grassland fire danger meter could predict rate of spread within $\pm 20\%$, Cruz and Alexander [2013] suggested a bound of $\pm 35\%$ as an acceptable benchmark for model performance, after finding that most empirical fire rate of spread models had shown mean absolute percentage errors between 20 and 40%. Albin [1976] indicated, and Alexander and Cruz [2013] reiterated, that the main sources of error in rate of spread predictions were lack of model applicability, internal model inaccuracy and input data errors. Uncertainties were also expected to be higher under operational settings due to logistical and time constraints.

What might be termed the ‘butterfly effect’ describes the potential for error accumulation through modelling processes. Such error propagation is clearly an issue in fire modelling since rate of spread (with an accuracy of, say, $\pm 35\%$) is used to calculate a variety of other fire characteristics, in particular fire perimeter and behaviours such as flame height or spotting potential. One way to deal with such accumulation, suggested by Cruz and Alexander [2013], is to use simpler models. However, with increasing understanding of fire dynamics and increased computational ability, as well as increasing demands by the public and fire management to predict greater detail, it appears increasingly likely that models will become more complex.

As noted by Tolhurst et al. [2008], fire modelling frameworks have become increasingly reliant on quality input data. However, the paradox of fire modelling uncertainty is that modelling over longer temporal or spatial scales is actually more accurate than over finer scales due to small variations in wind, fuel and dynamics that will unlikely ever be fully captured in models [Cruz and Alexander, 2013]. Catchpole et al. [1993, 1998b] suggested that this inherent fine-scale variability in laboratory fires accounted for as much as $\pm 20\%$ of the unexplained variability in observed rates of spread.

In bushfire prediction, uncertainty is currently considered in relation to the final outcome from fire spread and behaviour models. In both operations and research, state-of-the-art prediction methods have been primarily used in a deterministic manner; producing a single prediction of fire rate of spread or final perimeter for a discrete set of input values. Model outcomes are evaluated and interpreted in relation to measured metrics as a fire develops or against final fire perimeters, but these metrics are difficult to observe in the field.

During a fire event, operational focus is naturally on the management and suppression of the fire rather than on data collection. Validation of models using wildfires has therefore relied on limited remote sensing platforms, such as line-scans or aerial imagery collected during the event and post-event data such as final burnt area. While Duff et al. [2012, 2013] and Cui and Perera [2010] considered how fire perimeter prediction can be assessed against final burnt area, evaluations are complicated by factors such as suppression efforts as well as the ability to accurately observe predicted fire behaviour variables while a fire is actively spreading and being fought. There are limited studies or datasets that report multiple rates of spread for a given fire over a period or area, and so it is hard to analyse the uncertainty of fire rate of spread against observations (particularly for wildfires) [Cruz, 2010]. Many researchers have therefore conducted more detailed validation of their models using data from experimental or prescribed fires where conditions can be significantly calmer than those experienced in large wildfire scenarios [Alexander and Cruz, 2013]. Of the few studies that report multiple rates of spread for experimental fires, uncertainties around 30% to 40% were observed [Cruz, 2010].

These issues in model evaluation lead to questions of how to effectively capture uncertainty in the fire prediction process. Understanding model uncertainty in terms of model inputs, model parameters and model structure [Guillaume et al., 2016] is therefore particularly useful. However, there has been limited literature focussed on understanding where in the fire modelling process uncertainty might develop or how such uncertainty might be characterised, or even reduced. Penman et al. [2013] conducted a sensitivity analysis of Phoenix Rapidfire using a simulation study and found that fire behaviour was most strongly influenced by fire weather, rather than suppression and fuel treatments. Cruz [2010] also highlighted the gaps in dealing with uncertainty in the current literature and operations. For instance, the non-existence of confidence intervals in deterministic approaches to quantify error in prediction leaves the estimation of output uncertainty the sole responsibility of the end users.

2.1.4 Modelling Uncertainty

Through consideration of probabilistic approaches, operational bushfire prediction frameworks are beginning to capture and understand uncertainty [e.g. Twomey and Sturges, 2016, Andrews et al., 2007]. Cruz [2010] considered that ensemble modelling would allow identification of prediction limits defining the uncertainty associated with a prediction, provide an ability to make probabilistic predictions, and allow investigation of unlikely events or surprises in the system. J. D. Kepert (pers. comm., 30 Aug. 2016) also suggested that probabilistic prediction would allow for better understanding of likelihood or risk in management, as well as providing seamless and consistent predictions that avoid inconsistent forecasts which in turn improve user confidence. It was noted by Cruz [2010] that ensembles could also be used to extend the interpretation of predicted outcomes, but issues may arise in the interpretation of probabilistic results. Nominal categories of probabilities in terms of low to high risk could be used to better communicate pure probabilities, but this does not entirely alleviate the issues of interpretation.

Bates and Granger [1969] have been credited with pioneering the idea of combining forecasts to reduce the mean error of an individual prediction in ensemble-style modelling. The ideas behind ensemble modelling with Monte Carlo sampling are

based on the Law of Large Numbers in probability theory which states that, with increasingly large sampling sizes, the distribution of predictions asymptotically approaches the true process [e.g. Metropolis and Ulam, 1949, Feller, 1971, Lindgren, 1976]. In application, this results in running a model, or multiple models, a number of times to perturb its inputs, parameters or structure, and produce a distribution of predictions that represents the true range of possible outcomes. The wide ranging applications of ensemble modelling include economics [e.g. Garratt et al., 2011], ecology [e.g. Grenouillet et al., 2011], meteorology [e.g. Gneiting and Raftery, 2005] and climate [e.g. Molteni et al., 1996], as well as fire modelling [e.g. Cruz, 2010].

Kourtz [1972] indicated the potential improvement in reliability by producing probability distributions of fire danger indices rather than single outputs. In the days prior to the use of large computational systems, it was suggested that each weather variable be assigned a distribution by the forecaster rather than a single estimate. The probabilities could be sampled using a Monte Carlo sampling technique to produce a distribution of the predicted Fire Danger Index. In principle, almost half a century ago, Kourtz [1972] was suggesting the use of the type of ensemble modelling that is only recently being developed for operational fire prediction.

A number of authors have studied the ideas of ensemble modelling in application to fire prediction. Salazar and Bradshaw [1986] suggested prediction of distributions might be preferable to deterministic approaches for fire modelling, while Beer [1991] outlined a comparison of deterministic and statistical approaches to bushfire modelling. Anderson et al. [2007] also considered the results of fire modelling under systematic perturbations of climatic input variables, finding that the inclusion of perturbed weather data allowed better prediction of the variation in fire spread, particularly on the flanks and at the back of the fire.

The ensemble method used by Cruz [2010] was set up such that the prediction was given by inputs obeying $x = \bar{x} + x'$, where $x' \sim N(0, \sigma^2)$. Monte Carlo selection was used to sample 1000 ensemble members drawn from the Gaussian probability distribution functions of x , to produce a distribution of fire rate of spread predictions. A key question in this set up is how to generate a valid and realistic ensemble, i.e. how to estimate the spread of the variables, σ^2 . The variables considered by Cruz [2010]

were wind speed, air temperature and relative humidity. The spread of each variable was estimated arbitrarily for this study so as to illustrate the ensemble method. However, it is noted that these will be state and spatially dependent, and will vary with environmental factors such as atmospheric conditions. Wind direction was not accounted for by Cruz [2010], but there was noted to be a lack of influence of wind direction variability for fast-moving grass fires under high wind speed conditions. This has been shown to be very different for forest fires and Moore et al. [1988] noted the greater variance of wind in complex terrain in comparison to flat terrain. Finney et al. [2011] considered one of the most technically thorough ensemble wildland fire simulation methods within the fire literature. Hundreds, if not thousands, of ensemble members were used to produce simulations of fire spread and behaviour over 7-day periods for 91 fires. Results of the simulations were provided in terms of probability contours as well as final fire size, and were analysed in terms of variation between ensemble members as well as in comparison with observed fires. Through time series analysis of historical data, weather sequences were generated to mimic the range of realistic weather conditions for a given case study over the burning period. The simulation was also conducted using both two-dimensional and one-dimensional fire simulators to allow a comparison of the modelling approaches. It was noted by Finney et al. [2011] that added uncertainty in the modelling also originated from the fire analyst utilising the model and the operational choices that were made, for instance the selection of weather stations for input data or editing of fuel data from local experts or crews on the ground.

In addition, Finney et al. [2011] highlighted the sensitivity of the ensemble results to the number of ensemble members used in the prediction. From the generation of weather time series over a period of 7 days, a total of 247^7 different simulated weather conditions were possible, so a large number of ensemble members were required to produce reliable and repeatable results. Increasing the member numbers from 256 to 4,096 saw a marked increase in the variation between simulations, i.e. fire predictions. However, many other ensemble approaches taken for fire modelling (particularly in the operational setting) have analysed significantly fewer ensemble members, for example in the order of 25-30 [Twomey and Sturgess, 2016, French et al., 2014d]. The larger the sample size, or number of simulated fire predictions,

the more extreme events are captured within the ensemble. In some cases, this may not be operationally necessary and so lower numbers of members would be sufficient. However, when considering ‘worst case’ scenarios or trying to understand rare events with major consequences, larger samples sizes are required to capture the extremes.

Unfortunately, it is computationally expensive, and thus time-consuming, to achieve the number of ensemble members studied by Cruz [2010] and Finney et al. [2011]. There are in fact very few operational fire prediction systems that account for uncertainty in fire modelling inputs. The Fire Impact and Risk Evaluation Decision Support Tool (FireDST) developed by a collaborative research group in Australia used multiple deterministic fire spread predictions from Phoenix Rapidfire and combined them to calculate the likelihood of impact [French et al., 2013, 2014d]. Within this calculation, input parameters were sampled from uniform or equally weighted point distributions, rather than the Gaussian distributions used by Finney et al. [2011], and each fire spread scenario was therefore treated as equally likely.

Using an Area Difference Index as measure of performance, French et al. [2014d] compared the application of FireDST using Phoenix Rapidfire to a two-dimensional spatialisation of the CSIRO Grassland Fire and Forest Fire models using Huygens’ principle. The project also considered three case studies to evaluate the application of FireDST and understand the variability in simulating real bushfires [French et al., 2014a,b,c]. Although this analysis considered a thorough range of important fire spread drivers, the results considered only visual comparisons of the predicted fire perimeter to that observed. The model outputs were clearly highly sensitive to changes in wind conditions, including bias correction measures made to modelled wind conditions in the post-event analysis, and simulation studies highlighted the models’ additional sensitivity to slope. Unfortunately, however, the FireDST modelling framework has yet to progress past the proof-of-concept stage.

The only operational ensemble-based fire prediction framework currently used in Australia is SABRE (Simulation Analysis-based Risk Evaluation) developed by the Queensland Fire and Emergency Services [Twomey and Sturgess, 2016]. The framework uses a 25-member ensemble of Phoenix Rapidfire runs with input data selected

from standard distributions fitted around mean values [Queensland Fire and Emergency Services, 2015, Twomey and Sturgess, 2016]. Phoenix Rapidfire currently incorporates some estimates of fire fighting activities, but accurately incorporating their effect on fire perimeters into the model is complex. Verification of outputs from frameworks such as SABRE is therefore difficult where model outputs are almost always expected to be ‘incorrect’.

Other probabilistic approaches to fire modelling include a number of studies which focus on assessing risk of fire spread or fire behaviours [e.g. Penman et al., 2013, Preisler et al., 2004]. McRae and Sharples [2011] outlined a conceptual framework for assessing risk of extreme bushfires where a probabilistic approach was considered to define transitions between stages of fire growth and to calculate the likelihood of fire decay, persistence or escalation. In addition, Mandel et al. [2009] considered a number of statistical techniques for data assimilation in fire modelling. These approaches aim to update fire model parameters as the fire propagates, using new information gained from incoming data streams.

The prediction of uncertainty in bushfire spread and behaviour can also be handled through stochastic modelling techniques. Early in the literature, Dayananda [1977] outlined a simple stochastic model for the occurrence of human-caused forest fires in a particular region, using only the number of visitors to the region and simple environmental indices. Han and Braun [2014, page 434] later suggested that a “realistic fire growth simulator, based on rate of spread, should have a stochastic component”. Stochastic approaches to fire spread prediction have not only been shown to capture wider variation in fire spread than ensemble methods (J. J. Sharples, pers. comm., 2016), but are also less computationally demanding [Han and Braun, 2014]. Moreover, Boychuk et al. [2007, 2009] detailed a forest fire growth model using stochastic techniques where fire growth was represented as a random phenomenon modelled over a Markov chain lattice. These methods allow for the variation of environmental inputs as the fire propagates through time and space. Similar to those obtained through ensemble methods, stochastic models are able to produce probabilistic outputs such as probability contours for fire risk and distributions of area burnt.

Aside from input error, Cruz [2010] highlights the errors in prediction arising from model structure and parameters, i.e. systemic model errors, such as fire observation errors or weather interactions that are not captured within current models. It is suggested that an ensemble of models and parameterisations would shed light on the uncertainty of understanding around the relationships used to develop the different conceptualisations of fire spread [Cruz, 2010, Cruz and Fernandes, 2008].

However, it still remains that ensemble or stochastic approaches, and probabilistic outputs, are relatively new to the industry, particularly the end users of these fire models. It has been suggested that further education and continued research into these approaches are required to improve understanding of their advantages and promote better acceptance of the models [French et al., 2013].

2.2 Wind Modelling in the Context of Bushfire

For effective modelling of bushfire spread across complex landscapes, input variables need to be modelled at all scales relevant to bushfire behaviour; in particular, it was found that wind accounted for much of the variability in fire spread [Alexander and Cruz, 2006]. Finney et al. [2011] also showed that within ensemble fire simulations, wind direction was a key variable in accounting for uncertainty in predictions. Variation in wind direction had a significant impact on the difference between outputs produced by a one-dimensional fire model as compared to a two-dimensional simulator. This highlights the interaction between uncertainties of input variables and the uncertainties in the model structures themselves. These are aspects of fire modelling that still need to be better understood and addressed in research, yet so far in the literature, limited formal sensitivity analyses have been conducted on fire models [e.g. Penman et al., 2013].

While broad-scale weather interactions undoubtedly impact the spread of large bushfires through mechanisms such as synoptic wind changes or the formation of thunderstorms, most fire models use input winds at standardised heights of either 10m or 20ft to capture the near-surface wind dynamics that drive surface fire behaviour. For instance, McArthur [1967] used winds at 10m (\approx 33ft) but also showed profiles from 1.5m (\approx 5ft), and while Cheney et al. [1993] used 2m winds,

Cheney et al. [1998] returned to using 10m winds. Within Phoenix Rapidfire, standardised 10m open winds are therefore used as the input across the fire spread models. Rather than defining a specific height for wind inputs, Rothermel [1983] describes ‘mid-flame’ height of winds as related to the height of the fuel-bed, which therefore requires no prior knowledge of the burning process. To account for the height discrepancy between standardised wind observations at 20ft and the required estimates of ‘mid-flame’ wind speeds for Rothermel’s model, Andrews [2012] investigated modelling the wind speed adjustment factor (WAF).

Interactions between wind and terrain add further complexities to the relationship between wind and fire. Whiteman [2000] and Banta et al. [1990] identified a number of the mechanisms produced by the interaction of wind flow and rugged terrain, such as thermal-driven flows or dynamic channelling. Interactions between the fire itself and winds also form an evolving feedback loop which is extremely difficult to capture within fire modelling systems [e.g. Mandel et al., 2011, Peace et al., 2015]. Motivated by events during the 2003 Canberra bushfires, atypical (unexpected and potentially dangerous) fire spread over complex terrain was investigated by Sharples et al. [2012] and Simpson et al. [2013]. It was shown that atypical lateral spread could only be explained by wind-terrain-fire interactions. In particular, it became clear that flow separation was a necessary condition for the occurrence of such spread. Simpson et al. [2013] suggested that high wind speeds increased the likelihood of the behaviour occurring, with further work identifying relevant wind speed and terrain thresholds its occurrence [Simpson et al., 2016].

The dependence of fire spread on wind characteristics and the complexities of their interactions with the terrain, highlight the need for accurate and timely wind field predictions as input into fire modelling systems used for both ahead-of-time analysis and real-time operations. Crosby and Chandler [1966] highlighted the importance of accurate wind observations in fire modelling as well as management operations, while Noble et al. [1980] recognised the sensitivity of fire to wind, and the requirement of accurate data for model inputs.

For many studies of bushfire rate of spread, the inputs of wind speed and direction have been taken as uniform across either time or space, and sometimes both.

Sullivan and Knight [2001] investigated the variation of wind speeds surrounding a small experimental fire. The study (part of CSIRO's Project Vesta) highlighted not only the variability in wind speed across the area but also the difficulties in measuring (and thus predicting) winds at 'mid-flame' height, which directly drive the direction and speed of the fire front. This fundamental problem has persisted for around 80 years [Mitchell, 1937], despite the improvements in technology and understanding.

Mesoscale Numerical Weather Prediction (NWP) systems are able to provide accurate real-time weather predictions over a range of appropriate spatio-temporal scales. In Australia, the Bureau of Meteorology has developed the Australian Community Climate and Earth-System Simulator (ACCESS) model, which runs operationally at scales between 4km and 25km for local, regional and global weather predictions [Bureau of Meteorology, 2017], but can be used for small-scale high-resolution modelling down to approximately 400m [Wells et al., 2014]. However, in the operational context, these models do not resolve winds at sufficient spatial and temporal scales to capture the detailed topographic effects that influence surface fire behaviour. Wagenbrenner et al. [2016] also highlight the limited ability of traditional broad-scale weather prediction to capture the variability of wind fields across complex terrain, indicating the need for downscaling models to better predict meteorological variables at the finer resolutions relevant to fire spread modelling.

A number of different techniques for modelling two-dimensional wind fields across the landscape have been developed over the past half-century. These techniques have aimed to capture the dynamic variability caused by the interactions between the landscape and air flow above it; in rugged terrain these interactions are complex and sometimes confounding. The modelling techniques range from detailed numerical computations through to empirical adjustment factors.

Computational Fluid Dynamics (CFD) models involve computing numerical solutions to the full array of fluid equations conserving mass, energy and momentum in flow across the landscape. These models include resolution of the Navier-Stokes equations in numerous forms, inclusion of different turbulence schemes and estimation of atmospheric conditions [e.g. Vosper et al., 2002, Burton et al., 2006]. The

physics governing flow through the boundary layer (close to the surface), including interactions with vegetation canopies, are detailed by Ross and Vosper [2005], Finnigan [2000], Belcher et al. [2012], Mathew [2010] and Potter et al. [2002], amongst others. Operationally, the wind modelling software WindWizard, developed in the USA by Butler et al. [2006], uses CFD methods. The UK Meteorological Office has also used similar methodologies within the BLASIUS (Boundary Layer Above Stationary Inhomogeneous Uneven Surfaces) model to investigate flow over hills for applications such as pollution dispersion and wind energy [Allen, 2006, Vosper et al., 2013].

Full CFD models require considerable computational resources in order to solve the equations described above, and so are not necessarily appropriate for operational use in prediction of fire spread and behaviour in real, or faster-than-real, time. Simplifications of these methods are therefore commonly used within operational modelling. Mass-consistent models allow for the resolution of a subset of the equations dealt with in the previous paragraph, namely conserving mass across the landscape without resolving conservation of momentum [e.g. Forthofer et al., 2014b]. These models therefore capture the dominant characteristics of wind flow across the terrain without such significant computational demand.

In addition, look-up tables are often used in operations to give adjustment factors for fire rate of spread given the terrain features, particularly slope and aspect (R. H. D. McRae, pers. comm., 2014). These tables typically suggest an increase in rate of spread up hill and a decrease in speed down hill. The tables have been empirically developed and so inherently capture features such as wind reversals caused by large-scale turbulence over complex terrain. Similarly, wind multipliers also provide adjustment factors for wind speeds across hill slopes, showing a peak of wind speed at the ridge top [e.g. Paterson and Holmes, 1993, Holmes et al., 1997, Glanville and Kwok, 1997]. The method for computation of such wind multipliers across varying terrain was developed for application across Australia, primarily for the assessment of wind hazard or risk to activities including construction and agriculture [Yang et al., 2014]. While such detailed computation of wind multipliers has been used in the mitigation of risk from natural disasters such as tropical cyclones, they have yet to be applied within the bushfire context.

Both the full CFD technique and look-up tables are capable of incorporating finer-scale aspects of turbulence in wind flow across complex terrain, and therefore the components of the flow important for developing fine-scale, potentially atypical, fire behaviour. However, CFD methods are highly computationally demanding, reducing their ability to predict fire spread in real, or even near-real, time, and look-up tables lack the sophistication or computational ability to predict wind behaviours across wide regions.

2.2.1 Operational Wind Field Models

Mass-consistent diagnostic models are generally preferred for operations due to the computational constraints on producing useful, fine-scale wind model outputs in real, or near-real, time. A number of downscaling models are available in the literature including WindStation [Lopes, 2003] which incorporates a number of sub-models and solvers. However, WindNinja [Forthofer et al., 2014b] is currently the most widely used model across the USA in operational fire management, and is also incorporated within Phoenix Rapidfire in Australia (Tolhurst et al., 2008).

Despite its prolific use within operational fire spread models, it is noted throughout the literature [e.g. Butler et al., 2015] that there is a lack of wind data observed across complex terrain for effective evaluation of such wind modelling approaches. Commonly used datasets, such as Askervein Hill [Taylor and Teunissen, 1987], Tighvein Hill [e.g. Vosper et al., 2002, Burton et al., 2006] or Bolund Hill [Berg et al., 2011], represent landscapes that, although real, are selected due to their simplistic topographic or vegetation features, e.g. flat plains surrounding a single hill with limited vegetation. Wind data collected from more complex terrain, i.e. within mountain ranges, have been used to identify landscape-scale patterns or boundary layer dynamics such as thermal flows or dynamic channelling through valleys (Whiteman and Doran, 1993, Gross and Wipperman, 1987, Weber and Kaufmann, 1998). Surface wind data collected at finer scales, relevant to fire behaviour and spread across mountainous landscapes, such as those collected by Sharples et al. [2010], are rare.

In the evaluation of the development of WindNinja, Forthofer et al. [2014b] conducted a comparison of mass and mass-and-momentum conserving wind field models

with observed wind data from Askervein Hill and Waterworks Hill. Although Waterworks Hill exemplified slightly more complex topography, both regions exhibited limited vegetation cover. Both the mass and mass-and-momentum conserving wind models accurately predicted wind speeds on windward slopes. However, both models showed the greatest limitations in modelling wind speed and direction on leeward slopes. A comparison of fire spread simulations driven with a full CFD wind model versus a mass-consistent wind model also showed stark differences on the leeward side of Askervein Hill [Forthofer et al., 2014a]. By accounting for conservation of momentum in the CFD model, the fire spread was predicted to significantly reduce in speed on the leeward slope due to lee-side recirculation, while the mass-consistent model showed little reduction in rate of spread as the fire propagated over the hill.

Butler et al. [2015] introduced a high-resolution dataset collected over yet more complex terrain, with data collected over a tall, isolated mountain sitting on a plain at the foot of a mountain-valley range. The mountain, which rose 800m above the valley floor and had a horizontal scale of 4km, was considerably larger than Askervein Hill and provided more complex features for analysis of wind flow. A secondary canyon site was also investigated, but again the vegetation across both study sites was limited to grasses and low-level brush.

Wagenbrenner et al. [2016] compared the application of a previous version of WindNinja (version 2.5.2), a mass conserving model, against the observations made by Butler et al. [2015]. The study highlighted the improvements that downscaling models make on broad-scale NWP outputs, particularly in relation to the variability of wind speeds experienced across complex terrain. As in previous studies, the biggest improvements shown by the mass-consistent solver in WindNinja were found on windward slopes and ridge tops under high-wind conditions. However, the study reiterated the limitations of the mass-consistent solver in predicting wind speed and direction on leeward slopes.

Forthofer et al. [2014b] warned that users should be aware of model limitations and interpret results cautiously where appropriate. It was also noted by Wagenbrenner et al. [2016] that the limitations on leeward slopes were to be expected from WindNinja 2.5.2. The scheme was designed to account for only mass conservation across

the landscape and did not take into consideration momentum conservation, which would cause features such as recirculation on leeward slopes. The consequences of not capturing such wind features are significant in the fire modelling context with considerable impacts on fire propagation across the landscape, and characteristics such as flow separation and recirculation on leeward slopes being linked to extreme fire behaviours [Simpson et al., 2016]. Recent development of WindNinja has incorporated a momentum solver which allows the model to capture more complex flow patterns across the landscape while maintaining operational computation times (N. S. Wagenbrenner, pers. comm., 2015).

2.2.2 Vertical Wind Profiles

Vertical wind profiles within the boundary layer, but above the canopy, are often assumed to follow a logarithmic profile [Touma, 1977]. This profile becomes disturbed near the surface due to roughness of varying lengths, from topographical scales down to the scale of the vegetation. There is a significant body of work on wind behaviour over topographical features such as hills, escarpments and wind breaks [e.g. Holmes et al., 1997, Glanville and Kwok, 1997, Cleugh, 2002, Mobbs et al., 2005]. Equally, there is a wealth of literature regarding the physics of wind disturbances within the canopy [e.g. Finnigan, 2000, Belcher et al., 2012, Ross, 2008, Byronscott, 1990, Wood, 2000]. However, few authors have considered the combined effects of topography and vegetation on vertical wind profiles [e.g. Allen, 2006].

Particularly in the context of bushfire, the landscape is very rarely smooth or homogeneous, and in relation to surface fires, it is the wind within the canopy that is of most interest. Cruz and Alexander [2013] noted that aside from topographical features, the principal drivers behind the behaviour of spreading fires are fuel moisture and wind speed. Van Wagner [1989] recognised that the prediction of surface fires may be more difficult than that of crown fires due to the complexity of understory fuels. Alongside this recognition, it can also be asserted that the variation of wind fields within the vegetation layer adds further complications to the modelling of surface fires spreading beneath the canopy.

As with the wind modelling techniques described in the previous section, full CFD methods to resolve the physical equations describing vertical wind profiles through vegetation and over complex terrain are computationally expensive, so simplifications are required for operational prediction time frames. In the current suite of fire spread prediction models, wind speeds measured in the open environment (with no vegetation or above the vegetation layer) are translated to predict wind speeds within forests or vegetation at ‘mid-flame’ height using adjustment factors. The ‘wind reduction factor’ (WRF) [Cionco, 1972, Rothermel, 1972] and ‘wind adjustment factor’ (WAF) [Andrews, 2012] are defined empirically for a number of structural vegetation features including crown ratios and vegetation age. However, both the WRF and WAF assume the wind reduction profile to be constant throughout the vegetation layer despite the physics-based literature suggesting otherwise [e.g. Finnigan, 2000, Belcher et al., 2012]. Moon et al. [2013] presented empirical wind speed profiles for different forest types, showing that wind speed profiles within the canopy were in fact non-constant.

It is in the interest of the fire research industry to better understand, and subsequently model, the variability of wind fields within vegetation layers. Recent work [Moon et al., In Press] has shown that wind reduction profiles within the canopy depend on open wind speeds and height within the layer. An empirical Generalised Additive Modelling System (GAMS) is under development to model wind reduction profiles under various conditions (K. Moon, pers. comm., 2015). However, this new model has been developed using data collected in flat terrain areas, where the impacts of topography were intentionally minimised, so does not represent wind speed reduction beneath the canopy in areas of complex or mountainous terrain.

2.3 Probabilistic Approaches in Wind Modelling

As outlined in Section 2.1.3, probabilistic approaches have only recently been applied to operational bushfire modelling despite their advantages being highlighted a number of times over the past 50 years. Although more recent ensemble approaches produce probabilistic fire spread outputs, they still rely heavily on deterministic variable inputs. Twomey and Sturgess [2016] use standard probability distribution functions, such as the Normal or Uniform, around average input values given by

deterministic models, such as WindNinja. Wind direction, wind speed and wind speed reduction factor are all sampled using a Normal distribution centred around the predicted mean value, with advice from meteorologists determining relationships and distribution parameters (B. Twomey, pers. comm., 1 Sept 2016). One benefit of such statistical approaches to fire modelling inputs is the reduced computational demand in characterising dynamic wind behaviours that are physically complex, thus allowing for uncertainty characterisation without validation through full CFD models.

In the context of fire spread prediction, wind fields have been considered in probabilistic terms in only a limited number of studies [e.g. Sharples et al., 2010]. Work in the wind energy sector has contributed analyses of a number of individual sites and the fitting of probability distributions to wind speed and wind direction datasets [e.g. Erdem and Shi, 2011, Carta et al., 2008]. These studies consistently show that the variability of wind direction at a given point is not well represented by a Uniform or Normal distribution.

The most well-known univariate and unimodal distribution used to estimate wind speed is the Weibull distribution [e.g. Lackner and Elkinton, 2007]. However, modelling wind direction distributions is more complex due to the circular nature of the variable $\phi \in [0, 2\pi)$, which is defined such that

$$\phi = \phi + 2c\pi, \quad c \in \mathbb{Z}. \quad (2.6)$$

There are numerous other natural circumstances where circular data arise [Maradia and Jupp, 2000]. For example, Goodyear [1970] considered the orientation of fish species in relation to sun exposure, while Fisher and Powell [1989] investigated the flow direction of an ancient river by considering the orientation of fossils. Fisher [1993] outlined some of the theory behind circular statistics and, in particular, highlighted the use of wrapped or circular distributions such as the von Mises distribution.

Within the wind energy literature there have been numerous investigations into joint wind speed and wind direction distributions since the 1970s [e.g. Smith, 1971, McWilliams et al., 1979, McWilliams and Sprevak, 1980, Garcia-Rojo, 2004]. Many

of these studies have been aimed at optimising wind farm layouts for the most efficient production of energy, and therefore have taken the next step to characterise the joint distribution for modelling purposes. The literature has progressed through considering the univariate distributions of wind speed and wind direction individually, through to developing estimations of multivariate, cylindrical distributions. A number of different models have been used to estimate cylindrical distributions by combining linear and circular marginal distributions. For instance, Carta et al. [2008] used a truncated Normal-Weibull distribution to represent the marginal distribution of wind speed, while applying a von Mises distribution for the marginal distribution of wind direction. Similarly, Erdem and Shi [2011] considered the Angular-Linear approach introduced by Johnson and Wehrly [1978] to model the joint probability distribution of wind speed and direction.

Zhang et al. [2011, 2013] outlined a multivariate multimodal wind distribution (MMWD) model for cylindrical wind speed-wind direction data, which used non-parametric kernel density estimation (or the Parzen-Rosenblatt Window method). The model was shown to be effective across univariate (wind speed), bivariate (wind speed and direction) and multivariate (wind speed, wind direction and air density) distributions, using both visual and quantitative assessments of performance. Through a comparison with other bivariate unimodal models, the study highlighted the need for a model capable of capturing the multimodal nature of the joint distributions. Such multimodal distributions were estimated by García-Portugués et al. [2013a,b] using circular-linear kernel density estimation in application to wind direction and air pollution data. Additionally, the authors note that application of the introduced algorithm could easily be extended to circular-circular densities.

Both temporal and spatial uncertainty across the fire prediction process must be considered. Although a number of recent studies consider statistical approaches to spatial modelling of circular data with application to wave directions on the Adriatic coast of Italy [Alégria et al., 2016, Lagona and Picone, 2016], there are limited studies dealing with similar spatial statistical modelling of wind direction over complex terrain.

2.3.1 Joint Wind Direction Distributions

Over the last two decades of the 20th century, a collection of authors [Gross and Wippermann, 1987, Whiteman and Doran, 1993, Weber and Kaufmann, 1998] began to investigate the relationship between synoptic winds and winds experienced across complex terrain, particularly within valley structures. The investigations saw the development of joint wind direction distributions showing the relationship between wind direction at a location on a ridge top or recorded at the synoptic level, and wind direction observed at a point within a valley. These bivariate distributions were analysed to diagnose the presence of wind flow mechanisms such as forced channelling, pressure-driven channelling, thermal driving and downward momentum transport [Whiteman and Doran, 1993]. The studies focussed on these idealised mechanisms using empirical distributions, but the investigations were limited to physical interpretation with no efforts made to characterise the distributions or incorporate them into modelling approaches.

Sharples et al. [2010] used similar joint wind direction distributions to identify the surface wind behaviours across complex terrain in south-eastern Australia. The effect of wind speed was investigated by comparing changes in distribution structure as minimum wind speed thresholds (measured at the local ridge top) increased. The multimodal distributions indicated that wind reversals were prevalent on leeward slopes, and became more so with higher wind speeds. As discussed previously, these reversals were shown to provide some of the conditions necessary for atypical fire spread [Simpson et al., 2013, 2016].

In considering the bivariate joint distribution of wind direction, the two circular variables $\phi, \psi \in [0, 2\pi)$, i.e. the prevailing wind direction observed above a valley or on the ridge top (ϕ), and the concurrent wind direction observed at the surface within the valley (ψ), are defined according to Equation 2.6. The bivariate circular frequency distribution, f , of two such variables is then defined over the torus, $(\phi, \psi) \in S^1 \times S^1 = \mathbb{T}^2$, such that

$$f : \mathbb{T}^2 \rightarrow \mathbb{R}, \quad \int_{\mathbb{T}^2} f d\phi d\psi = 1. \quad (2.7)$$

Sharples et al. [2010] collected wind direction data in 22.5° bins relating to the 16 points of the compass, i.e. N, NNE, NE, and so on. Such observed discrete data, in n bins, are recorded at the direction pairs $X_{ij} = \{(\phi_i, \psi_j) : i, j = 1, \dots, n\}$. The observed bivariate discrete dataset, $Y(X_{ij})$, is considered to be a noisy realisation of the underlying true continuous process [Sharples et al., 2010], and the following model is proposed;

$$Y(X_{ij}) = f(X_{ij}) + \epsilon(X_{ij}), \quad (2.8)$$

where $\epsilon \sim N(0, \sigma^2)$ is a random noise component. The observed discrete data, $Y(X_{ij})$, are represented either on the plane or over the torus as shown in Figure 2.1, while $f(X_{ij})$ is the underlying continuous distribution function to be estimated.

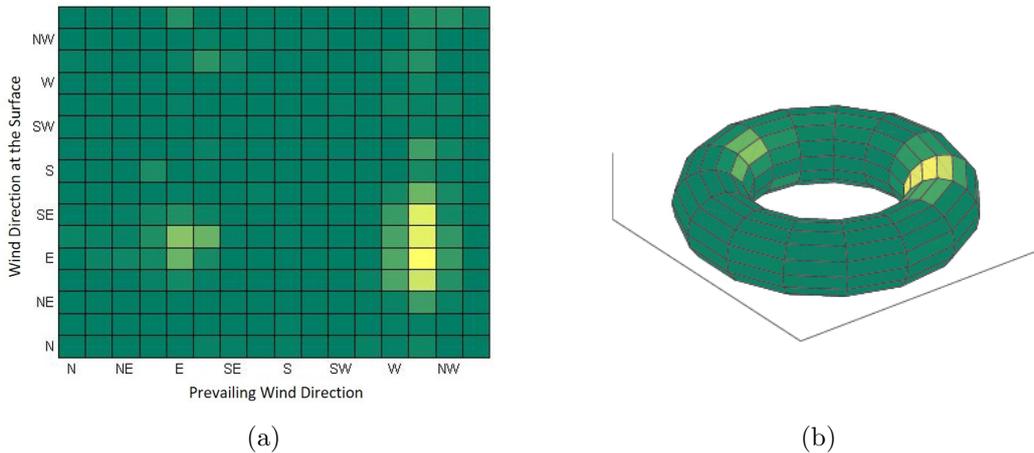


Figure 2.1: Observed joint directional wind response distribution represented (a) in planar view and (b) on the torus. Data collected by author.

Parametric estimation is commonly used for modelling data with known probability distributions, but there are complexities in modelling multimodal, multivariate toroidal distributions, and particularly in estimating and interpreting the parameter space [e.g. Coles, 1998, Kent et al., 2008]. Non-parametric estimation can therefore be used for exploratory analysis and inference regarding surface shapes and structures where known distributions are less suitable.

When considering a multimodal multivariate wind distribution model, Zhang et al. [2011, 2013] suggested that quantitatively characterising uncertainty should be a future area of research. It should be noted here that any smooth estimator of a surface is inherently biased due to the trade-off made between bias and variance. This problem prevents such estimators being used to construct confidence intervals

for the consideration of uncertainty. Alternative tools such as the SiZer (Significant Zero crossing of the derivatives) map have been developed for curves by Chaudhuri and Marron [1999] and adapted for circles by Oliveira et al. [2014b]. A spherical adaptation is currently under development [Vuollo and Holmström, 2017] and a toroidal version would be possible, however such techniques rely on simple visual interpretation which can be complicated to present in higher dimensions.

2.3.2 Parametric Estimation

Univariate wind direction distributions can be estimated and analysed using circular distributions such as the von Mises or wrapped Normal distributions [Fisher, 1993]. In bioinformatics, bivariate circular data arise in the identification of proteins, since pairs of angles between amino acids along the backbone of different proteins form unique bivariate relationships [e.g. Singh et al., 2002, Mardia, 2013]. The bivariate von Mises distribution has been used to estimate the joint distribution of angles between two amino acids along the backbone of known proteins. The distribution structures, i.e. modal locations and shapes, of an unknown protein are then compared to known protein characteristics in order to identify it [e.g. Mardia et al., 2007, Taylor et al., 2012].

The full bivariate von Mises distribution has eight parameters, interpretation of which becomes complex [Kent et al., 2008]. Therefore, a number of models have been proposed as simplifications of the bivariate von Mises distribution: the Cosine model and the Sine model [Singh et al., 2002, Kent et al., 2008]. The Sine model, for instance, can be given by

$$f(\phi, \psi) \propto \exp \{ \kappa_1 \cos(\phi - \mu_1) + \kappa_2 \cos(\psi - \mu_2) + \delta \sin(\phi - \mu_1) \sin(\psi - \mu_2) \}, \quad (2.9)$$

where μ_1 and μ_2 are the mean values of ϕ and ψ , and κ_1 and κ_2 are the respective concentration values. The fifth parameter, δ , describes the covariance between the two variables. In the literature, a number of formal parameterisation techniques exist and have been utilised to fit such a distribution to data, including maximum likelihood estimation, method of moments and the least squares method [e.g. Kent and Tyler, 1988, Mardia et al., 2007, 2009]. In the context of bushfire modelling,

physical interpretation of these parameters is plausible in terms of wind direction measurements that might be observable in the field.

However, these models are still under development within an active area of research, and when considering their application to joint wind direction distributions over complex terrain, it is important to understand their key assumptions. One of the most important assumptions made amongst the wind energy literature is that wind speed and wind direction are independent variables [Zhang et al., 2011, 2013]. This makes it possible to construct joint distributions using products of univariate distributions. With an aim to model joint wind directions using the bivariate von Mises distribution, the μ and κ parameters of the Sine Model, Equation 2.9, can be simply interpreted as the mean and variance of the prevailing and surface wind directions. The dependence between the observations at the two points is dealt with in the covariance term, δ [Kent et al., 2008]. However, further work is necessary to understand both spatial and temporal correlations in wind direction across the landscape.

2.3.3 Non-parametric Estimation

Non-parametric estimation is very useful for accurate approximation and understanding of density shapes, which can then be applied in exploratory analysis and inference. It is possible to quantify modal structures and probabilities without the restrictions of known distribution forms. The estimation of bivariate circular, or toroidal, surfaces is also an active area of research [e.g. Panzera and Taylor, 2012, Di Marzio et al., 2013], but it is again plausible that it is an area of research that can be exploited for probabilistic wind and fire modelling.

Non-parametric estimation techniques can essentially be thought of as belonging to two classes: interpolation and smoothing. Interpolation techniques, such as the cubic spline, are efficient and straightforward to run by fitting the estimated surface to the set (or a subset) of the exact data points. Bivariate surface approximations are often dealt with using tensor products of two univariate splines with the appropriate boundary conditions, such as periodicity to handle data circularity [R Core Team, 2016, MATLAB®, 2016]. However, exact interpolation between data points leaves these methods susceptible to exaggeration of noise in the observed dataset.

Under the assumption that wind data are affected by noise [Sharples et al., 2010], exact interpolation is therefore unlikely to represent the underlying bivariate wind direction distributions.

In contrast, smoothing techniques are designed to account for inherent noise by optimising the trade-off between bias and variance. For instance, the thin plate smoothing spline is defined to allow a trade-off between the fidelity of the spline to the observed data points and the smoothness of the resulting surface [Wahba, 1990]. Many smoothing spline techniques have been applied to univariate and bivariate linear data, with much literature researching their properties and applications [e.g. Meinguet, 1979, Hutchinson and Gessler, 1994, Hutchinson, 1995, Hancock and Hutchinson, 2006].

For bivariate data, smoothing splines have also been used to estimate curves over cylindrical and spherical spaces [e.g. Fan and Martin, 2013, Egerstedt and Martin, 2010]. For surface estimation, Wahba and Wendelberger [1980] and Wahba [1981] defined a spherical roughness penalty to construct a thin sphere smoothing spline. Although a number of authors have considered the construction of splines to estimate curves over the torus [Egebrand et al., 2010, Goodman et al., 1989, Lee and Tang, 1989], there is limited literature on the application or construction of splines to estimate a surface defined over the torus.

Kernel density estimation is also commonly used to estimate curves and surfaces for exploratory analysis and statistical inference. With the definition of a toroidal kernel, Di Marzio et al. [2011] constructed a kernel density estimator over the d -dimensional torus. Taylor et al. [2012] used toroidal kernel density estimates (based on products of the von Mises distribution) to validate a number of protein structures using bivariate circular data, but it was noted that the results of the study relied heavily on the selection of the bandwidth which controlled the smoothness of the estimated surface.

An alternative simple strategy used in a number of practical applications for handling circular (particularly univariate) data is to wrap, or repeat, the dataset and conduct the estimation on the expanded set, which explicitly exhibits its circular

nature [Silverman, 1986]. Despite the potential for this method to increase computational demand for the estimation process, it is readily available (and interpretable) to practitioners from a wide range of fields. In order to estimate the underlying continuous structure of joint wind direction distributions, Sharples et al. [2010] used this process to adapt thin plate smoothing spline methods and identify opposing edges of the planar joint wind direction distributions. By identifying these opposing edges, Sharples et al. [2010] were able to capture some of the toroidal nature of the bivariate wind direction distributions and provide realistic estimates of the underlying continuous process.

2.4 “The Literature Gap”

Bushfire modelling techniques used across the world have been empirically developed for over half a century, predominantly on the basis of small-scale experimental fires. Although state-of-the-art prediction systems are today capable of modelling more complex dynamics of fire behaviour, there are still significant flaws in the modelling process. The reliance on deterministic modelling has long been challenged by the suggestion of probabilistic approaches, which may not categorically improve on model outputs but will provide the user with complementary information to facilitate more informed discussions of accuracy and uncertainty [Cruz, 2010]. In recent research, these ideas have developed significantly, but there is still much work to be done to incorporate probabilistic input variables and fully quantify uncertainty throughout the fire modelling process.

Accurate prediction of wind field characteristics is key to the validity of fire predictions, but lack of data for development and evaluation remains a limitation of wind models currently used within operational bushfire modelling. As outlined in Chapter 1, the first aim of this research is to collect wind data at scales relevant to surface fire behaviour. These data are presented in Chapter 3, and discussed in a probabilistic context throughout this thesis.

Incorporating probabilistic approaches to wind modelling allows for a computationally efficient way to bridge the gap between currently operational diagnostic models and the full set of dynamics acting on wind fields in the environment. Probabilistic

approaches also allow for quantification of uncertainty around the inputs of bushfire models, and production of input distributions that are well suited to emerging fire modelling frameworks. However, techniques currently used within ensemble-based fire modelling frameworks do not well represent the true variability of wind speed and direction across complex terrain. This research will utilise the newly collected data to evaluate current wind modelling techniques with an aim to better represent this variability in wind characteristics.

Joint wind direction distributions were empirically developed in the late 20th century for physical interpretation of broad-scale wind flow over complex terrain. Sharples et al. [2010] used similar joint distributions to analyse surface wind flow in order to understand the processes behind atypical fire spread. Within this research, these distributions are further analysed to provide a platform for better characterisation and modelling of uncertainty in wind direction across the landscape.

Parametric approximation of joint wind direction distributions would ideally provide wind direction models parameterised with physical features that could be observed in the field. However, the development of parametric bivariate circular distributions is still an active area of research, with parameter interpretation a core issue. Non-parametric estimation and comparison of joint wind direction distributions alternatively allows for quantification of distribution features without the restraints of known distribution functions. Such analysis will provide probabilistic information that can complement current physics-based wind modelling techniques that are used operationally. However, toroidal non-parametric techniques are not readily available to practitioners, and this research also aims to understand how best to handle bivariate circularity in datasets.

CHAPTER 3

Data Collection

As discussed in the previous chapter, there is a lack of wind datasets available for evaluation of wind models, such as those employed in bushfire modelling, particularly at fine resolutions. In numerous contributions to fire behaviour research, it has been highlighted that surface wind behaviours have considerable impacts on surface fire spread [e.g. Rothermel, 1972, Andrews, 2012]. Variations in surface wind flow can create conditions for the development of extreme fire behaviour that lead to significant changes in the spread of a bushfire across the landscape [e.g. Sharples et al., 2012, Cheney et al., 2001, Wildland Fire Associates, 2013].

The first aim of this research is to contribute a new wind dataset at the spatial and temporal resolutions relevant to surface bushfire behaviour. Wind data were collected at such scales across a number of case studies around the Australian Capital Territory (ACT) and New South Wales (NSW) in 2007. Analysis of these data were presented by Sharples et al. [2010], where certain aspects of mountain meteorology were linked to extreme fire behaviours. This research is utilised as the motivating example for the studies detailed throughout this thesis.

This chapter provides a comprehensive account of the data collection process that formed the basis for much of the analysis presented in this thesis. Firstly, Section 3.1 details the equipment and software used to collect and process the data, before Sections 3.2 and 3.3 describe the study areas and summarise the observed wind characteristics. This research firstly reanalyses the data collected in 2007 and analysed by Sharples et al. [2010] from Flea Creek Valley (FCV) in the Brindabella

National Park, NSW. As an extension of this research, wind data were again collected across the same transect of FCV in 2014. In 2015, a second case study region was identified at the National Arboretum Canberra (NAC) in the ACT, where data were collected across two contrasting stands of vegetation. Section 3.4 closes this chapter with some concluding remarks.

3.1 Equipment and Software

3.1.1 Data Collection Equipment

Eleven Davis¹ Vantage Pro 2TM Portable Automatic Weather Stations (PAWS) were used to collect meteorological data for this research (Figure 3.1). The stations collected data on numerous meteorological variables including wind speed, wind direction, temperature, relative humidity, rainfall and solar radiation. This research focusses on the analysis of observed wind characteristics. Wind speed was observed using cup anemometers mounted at 5 metres above ground level (Figure 3.1(a)), with a resolution of 0.4ms^{-1} and accuracy of $\pm 1\text{ms}^{-1}$ (within a range of 0.5ms^{-1} and 89ms^{-1}). Wind direction was observed using a wind vane, also mounted at 5 metres above ground level, and was recorded in 22.5° bins according to the 16 points of the compass.

The stations were capable of measuring data at 1-minute, 5-minute, 10-minute, 30-minute and 1-hour intervals. Within these intervals, instantaneous observations were taken every 3 seconds. For wind speed, the recorded value for a specific interval was calculated as the average of the instantaneous observations, while recorded wind directions were calculated as the most dominant wind direction observed over the time interval. All other meteorological variables were observed using a sensor array at approximately 2 metres above ground level, and data were recorded via the console and data logger held within a weatherproof box at approximately 1 metre above ground (Figure 3.1(b)).

To calibrate the eleven weather stations prior to deployment, comparisons were made with wind observations recorded by a permanent automatic weather station at the UNSW Canberra Field Site (Figure 3.1 (a)). Comparisons were also conducted

¹www.davisnet.com

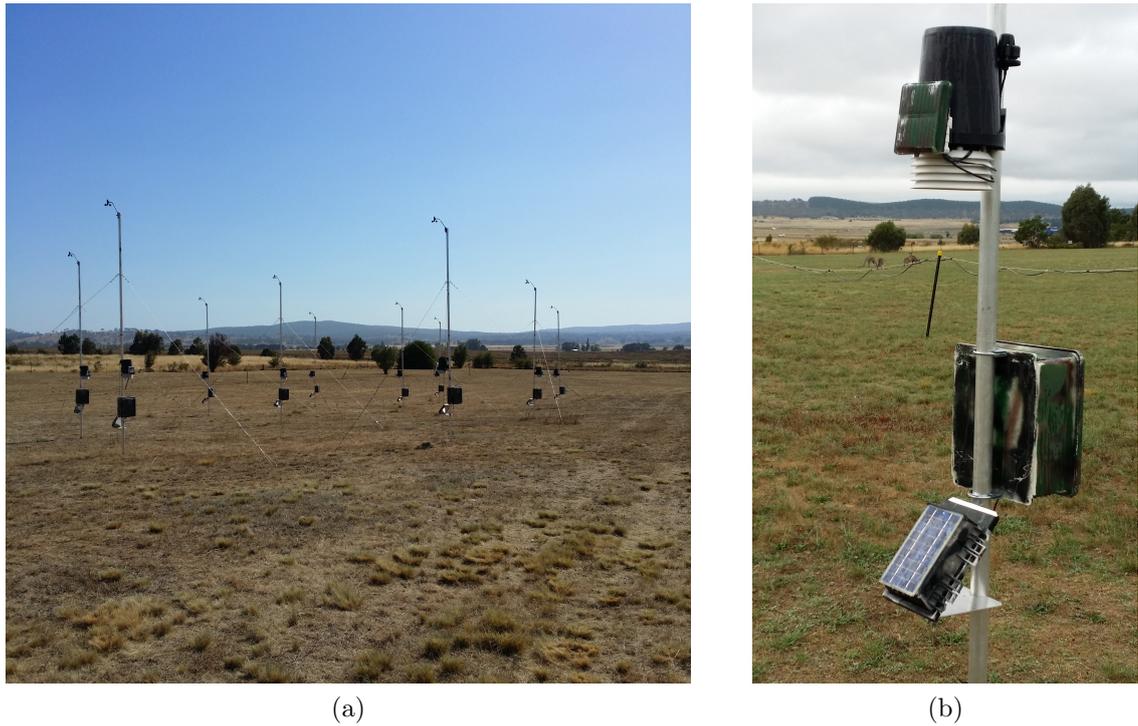


Figure 3.1: Davis Vantage Pro 2 Portable Automatic Weather Stations during the calibration phase; (a) set up at the UNSW Canberra Field Site, and (b) sensor array and console box. Photographs taken by author.

between a reference PAWS and a weather station at the Woodlawn Bioreactor in New South Wales. Analysis of these comparisons showed that the PAWS were accurate and consistent in the measurement of both wind speed and wind direction (Appendix A).

Initially, 30-minute data observation intervals were used due to the minimal data storage available in the in-built Davis data loggers (128kB). If left uncollected, the loggers would simply overwrite themselves so, in the field, data were collected every 4-6 weeks by downloading the files to a laptop, using the Weatherlink® software provided with the stations (Figure 3.2(a)). Batteries were set up to recharge using a solar panel fixed to the console boxes, but some difficulties were encountered in the winter months due to low solar exposure, particularly at sites with heavy vegetation or south facing aspects (Figure 3.3(b)). Sharples et al. [2010] experienced similar issues in the winter months of 2007 due to a lack of sunlight, but vegetation cover was much less of a problem since it consisted of only young regrowth across Flea Creek Valley after the 2003 Canberra bushfires (Figure 3.3(a)).

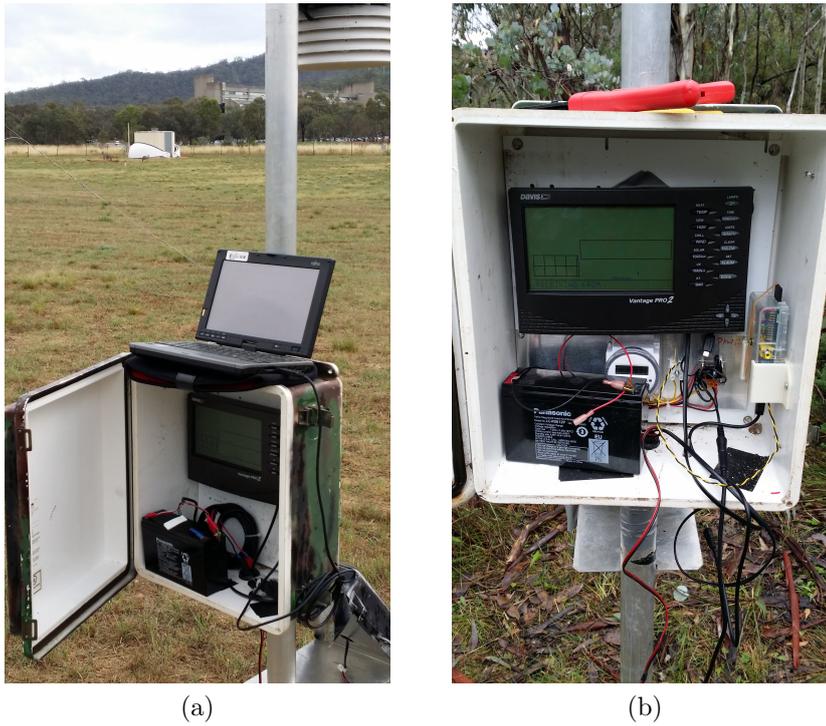


Figure 3.2: Data download using (a) Weatherlink® , and (b) the Raspberry Pi® system. Photographs taken by author.

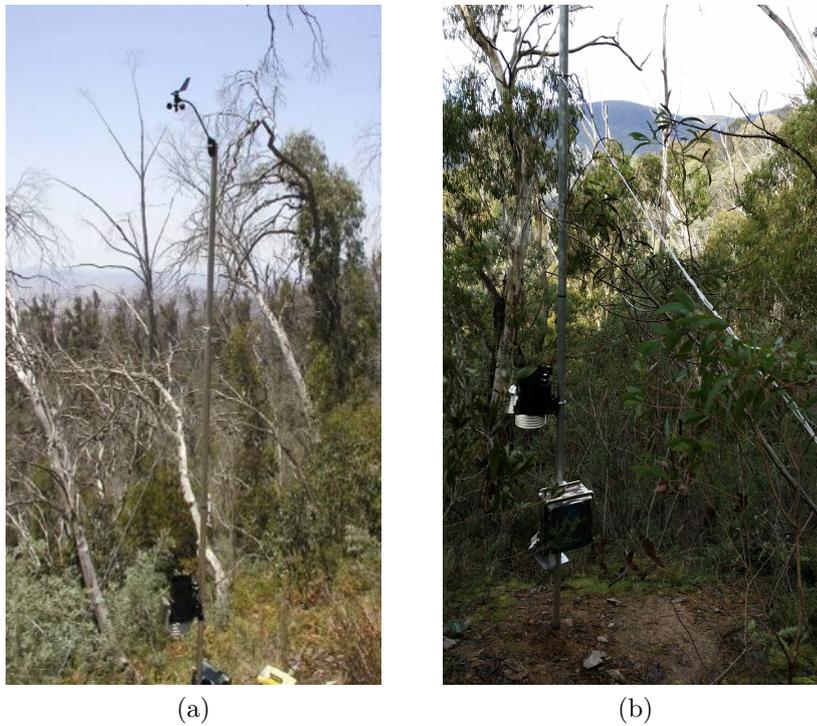


Figure 3.3: Photographs of data collection at FCV in (a) 2007 and (b) 2014. Photographs taken by (a) J.J. Sharples and (b) author.

In the winter of 2014, the stations were modified by the author to improve data storage capacity. This involved installation of Raspberry Pi® (RPi) systems (Figure 3.2(b)), which also allowed for more frequent observations, that is, down to 1-minute intervals. Data files were automatically downloaded from the PAWS data logger to an external SD card every 24 hours, with downloads taking approximately 4-5 minutes during which time new observations could not be recorded. Due to the extra power usage of the RPi systems (even when running for only a few minutes a day), battery issues continued with a number of stations, again particularly those with limited solar exposure. In these areas, extra batteries were added to stations and data collection, with maintenance checks, continued to be conducted every 4-6 weeks. Figures A.3, A.5 and A.7 in Appendix A show the data records for each dataset considered within this research. The gaps in the records indicate times of no data collection which were predominantly due to insufficient battery power.

3.1.2 Data Processing

The software packages used for data collection, processing and analysis included Weatherlink®, MATLAB® [2012, 2016] and R [R Core Team, 2016]. Data presented by Sharples et al. [2010] were analysed using code converted from Fortran® to MATLAB® [2012]. In 2014, prior to the installation of the RPi system, data were downloaded from the PAWS using the Weatherlink® software provided with the stations. This software produced Comma-Separated-Value (CSV) files that were consequently analysed in MATLAB® [2012]. With the installation of the RPi system, raw data were collected directly from the stations in CSV format via the external SD card. The data in these files were coded according to the Davis documentation [Davis Instruments Corp., 2009] in the hexadecimal system. These raw data were converted to the decimal system and the standard units for each variable in MATLAB® [2012] prior to analysis.

Within MATLAB® [2012], data were cleaned to account for the battery outages discussed in the previous subsection as well as data anomalies. Aside from battery and data storage challenges, a number of other data collection issues were experienced in the field due to interference from wildlife. These interferences took many forms and included cables chewed by small mammals, ant infestations, guy ropes

pulled out by larger animals (most likely kangaroos), rain buckets filling with forest debris (including bird faeces) and the potential for bird interferences with wind vanes.

Due to excessive rain bucket debris across both case studies, rain data were discarded from all datasets. With maintenance checks conducted every 4-6 weeks, major interferences such as chewed cables or pulled out guy ropes could be identified so that the appropriate data were also discarded. Other minor, or short-term, interferences such as birds or debris hitting the weather vane were more difficult to identify. The analysis presented in this thesis focusses on understanding the distributions of wind data across long time periods, and the uncertainty surrounding individual data points is intrinsically handled by the statistical approaches employed. However, data anomalies caused by the above mentioned interferences may be an important consideration in future analysis using these datasets, particularly for finer scale applications such as the analysis of forest micro-meteorology.

Information from all the observed meteorological variables was used to diagnose the above anomalies and gaps. For instance, assessment of the observed solar radiation at each station provided a proxy for diagnosis of battery recharge issues. Once clean, daily data files were concatenated to produce a complete record for each station. Data could then be manipulated to compile information for each variable across all sites. Individual variables were then summarised using summary statistics, time series plots and histograms such as those shown in the following sections and in Appendix A. Appendix A also details the storage and access of the data used for analysis throughout this thesis.

3.2 Case Study I: Flea Creek Valley, NSW

3.2.1 Description

Flea Creek Valley (FCV) lies within the Brindabella National Park, 70km west of Canberra (Figure 3.4(a)). With elevations ranging from approximately 500m up to 1420m at Mount Coree (Figure 3.4(b)), the area is considered rugged terrain according to the surface roughness classifications defined by Weber et al. [2008]

and McRae and Sharples [2013]². The temperate high-country climate leads to warm summers and cool winters (with snow fall recorded most years). The vegetation across the region is broadly native eucalypt forest, with some patches of pine plantation.

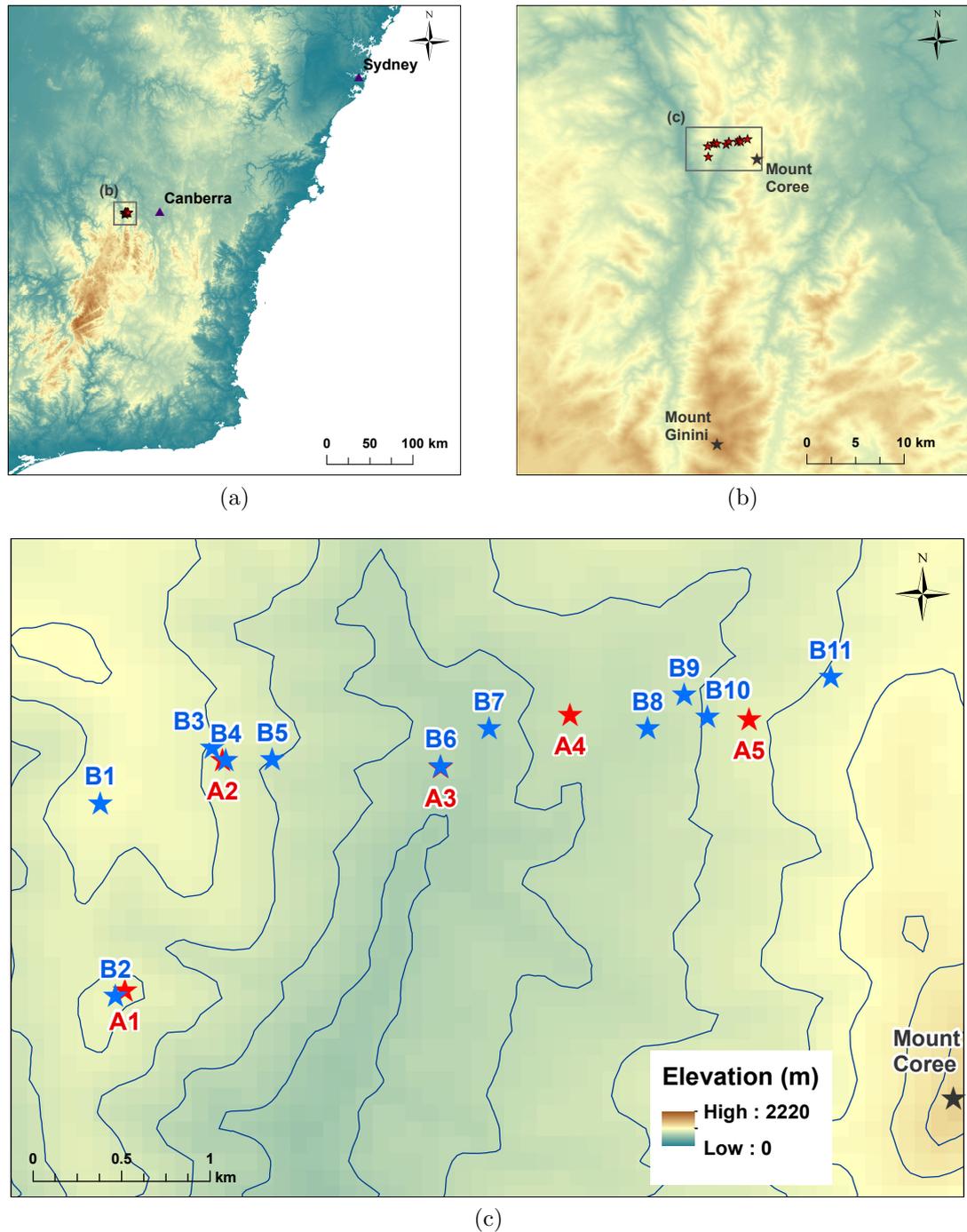


Figure 3.4: Maps of (a) south-eastern Australia, (b) the Brindabella region and (c) the FCV transect (with 50m contours). Station locations in 2007 (red; A1 to A5) and 2014 (blue; B1 to B11) are indicated.

²www.highfirerisk.com.au/maps

As noted by Sharples et al. [2010], the synoptic patterns across the region are dominated by high-pressure weather systems which produce west-northwesterly (WNW) winds during the summer, and westerly winds throughout the winter. Flea Creek Valley is aligned along an approximate north-south axis, thus approximately perpendicular to the prevailing winds experienced across the region. Significant topographical features surrounding the valley include Mount Coree approximately 3km to the east of the valley floor, and Webbs Ridge to the west, on which stations A1, B1 and B2 were located. Flea Creek joins the Goodradigbee River around 6km downstream of the study site.

Wind data were collected across a 3-4km transect of the valley between January and October 2007, using five PAWS [Sharples et al., 2010]. Between April and December 2014, wind data were collected across the same transect of the valley, using eleven PAWS³. Figure 3.4(c) indicates the locations the PAWS in both years, and the details of each station site are given in Tables 3.1 and 3.2. Details of the 2007 sites are adapted from Sharples et al. [2010, Table 1], and topography details for the 2014 sites were obtained through ArcGIS [ESRI, 2011] analysis of the SRTM⁴ 90m Digital Elevation Map (DEM).

Table 3.1: Descriptions of the FCV sites (A1 to A5) in 2007. Adapted from Sharples et al. [2010, Table 1].

	Lat.	Long.	Elev.	Slope	Asp.	Description
A1	148.76619	-35.30373	1026m	0°	NA	Ridge-top site. Relatively dense acacia seedling and other regrowth outside 3m. Larger eucalypts with cambial growth approx. 4-5m away. Very little canopy immediately overhead but relatively dense canopy over station surrounds.
A2	148.77113	-35.29196	955m	20-25°	105°	E-facing slope on western side of FCV. Sparse eucalypt regrowth approx. 3-4m from station. Some taller trees, some with cambial growth, approx. 4-5m from station. Partially intact canopy overhead.
A3	148.78222	-35.29232	787m	5-10°	150°	Near bottom of FCV on small knoll. Some burnt trees approx. 3-4m away. Scattered canopy overhead.
A4	148.78878	-35.28964	850m	10-15°	300°	W-facing slope near bottom of eastern sidewall of FCV. Very dense (dead) bracken up to 1m high. Some larger trees, some with cambial regrowth approx. 5-6m away. Sparse canopy overhead.
A5	148.79787	-35.28990	1000m	20-25°	315°	NW-facing slope on eastern sidewall of FCV. Very sparse acacia regrowth and bracken within a radius of 5-6m. Some larger eucalypts 7-8m from station. Relatively sparse canopy overhead.

³Scientific Research Licence #SL101338, NSW National Parks and Wildlife Service.

⁴<http://www.chiar-csi.org.au/data/srtm-90m-digital-elevation-database-v4-1>

Table 3.2: Descriptions of the FCV sites (B1 to B11) in 2014.

Description	Lat.	Long.	Elev.	Slope	Asp.	Vegetation (April)	Vegetation (Dec)
B1 Western ridge top	148.76508	-35.29411	1095m	Flat	NA	Light scrub up to 1m. Individual trees to N and W (foliage at 5-15m). Medium to thick coverage up to 10m, approx. 3m away to the S.	Scrub up to 2m, approx. 5m away. Sparse canopy up to 15m, approx. 3m away.
B2 Western ridge top (1km south)	148.79787	-35.2899	1000m	Gentle	South	Scrub up to 1m, intermediate foliage 2-3m away. Several surrounding trees with heavy foliage at 5-7m. Large 10-15m tree, 3m to the East.	Surrounding thick brush/bracken up to 5m, and dense canopy up to 15m.
B3 E-facing slope under powerline	148.77061	-35.29131	993m	Low-Medium	E	Surrounding scrub and trees up to 2-4m. Open above, under powerlines	Brush up to 5m surrounding station within 2m. No higher canopy. Cleared for powerlines.
B4 E-facing slope	148.77131	-35.29197	972m	Steep	SE	No scrub. Intermediate foliage at 1-5m. Large individual trees up to 10m to N and SE (up-slope and down-slope). Downslope foliage at same alt. as anemometer.	Dense scrub to 2m, and canopy up to 15m surrounding station, approx. 3m away. Clear downslope of station.
B5 SE-facing slope	148.77383	-35.292	900m	Very Steep	SE	Steep slope clearing. Clear to S-SSE, and ENE. Some 1-2m light scrub.	Light scrub above and below station. Dense trees surrounding station approx. 10-15m away from clearing.
B6 Valley floor, on knoll	148.78222	-35.29225	791m	nearly Flat	NE	Surrounding medium density canopy up to 5-10m.	Sparse canopy but some directly over station. Dense low bush ground cover to 1m.
B7 Valley floor	148.78467	-35.29033	771m	Medium	SSE	Clear to N & S. Some scrub up to 1.5m, and surrounding foliage at 5m. Trees to E & W to 15m.	Medium density ground cover to 1m. Saplings to 2m surrounding station. Sparse canopy directly above. Canopy cover from 4m away from station.
B8 W-facing slope, near gully	148.79725	-35.29064	840m	Medium	WNW	Little to no scrub. Open to W. Trees & foliage at 2-15m to the E. Trees up to 10m to the S & N.	Medium density ground cover to 1m. Saplings to 2m surrounding station. Sparse canopy directly above. Canopy cover from 4m away.
B9 W-facing slope	148.76458	-35.28844	889m	Low-Medium	W	Scrub to 1.5m. Tree up to 3m to E. Surrounding medium density foliage up to 5-10m, approx. 5m away.	Medium density groundcover. Low shrubs to 1m. Medium density canopy cover.
B10 W-facing slope	148.79575	-35.66667	912m	Medium	W	Open to NW. Rest enclosed at 4-15m. Scrub to 2m. Trees to N from 2-6m away with med foliage 4-10m.	Low density scrub cover approx. 2m away from station. Surrounding medium to low density canopy at 5m.
B11 W-facing slope, eastern ridge	148.80203	-35.28789	1032m	Medium	W	Dense scrub up to 1.5m. Trees up to 7-15m to E, approx. 5m away. Surrounding medium density foliage up to 7-15m.	Medium density canopy up to 10-12m. Dense scrub cover and low grass to 1-1.5m height. Fallen logs.

The Flea Creek area was heavily impacted by the 2003 Canberra bushfires, in particular the McIntyres Hut fire which ignited a few kilometres north-west of the study area and burnt through the region and into the ACT. These fires claimed 4 lives, and destroyed over 500 homes. The vegetation across FCV and the surrounding regions was also devastated. Table 3.1 describes the station sites across FCV from the 2007 data collection, only 4 years after the fires [Sharples et al., 2010]. Table 3.2 then describes the sites across the valley in 2014 (eleven years since the fires). There are clear differences in language; where Sharples et al. [2010] describes “burnt trees” and “sparse canopy”, the current vegetation is described as “medium density” with “large eucalypts” surrounding the stations. Figure 3.3 exemplifies these differences in vegetation between the two years.

3.2.2 Data Summary, 2007

Throughout the analysis presented in this thesis, surface winds are considered in relation to the prevailing wind conditions to represent the response of wind flow to changing surface conditions. For each case study, and over each time period, the prevailing wind conditions are indicated by data collected at the local ridge top. Although these sites cannot be assumed to accurately represent the prevailing wind conditions, analyses later in Chapter 4 (Section 4.2.2) show that the wind directions observed at these sites are at least indicative of the prevailing winds. In 2007 at FCV, the ridge top station was located at A1 (Figure 3.4(c)). However, in 2014, winds observed at this site (B2) were no longer indicative of the prevailing conditions (likely due to the change in vegetation structure), and the station at B1 was used as the ridge top station.

From January to July 2007, data were collected at 30-minute intervals, from July onwards data were collected at 1-hour intervals due to battery issues experienced throughout the winter resulting from a lack of solar recharge [Sharples et al., 2010]. In 2014, 30-minute data were also collected from April through to July. In July 2014, the RPi system was installed and 1-minute data were collected through to December. Sharples et al. [2010] detailed the data collected and the initial analysis of wind direction response over Flea Creek Valley in 2007. The 30-minute data

collected by Sharples et al. [2010] were analysed in this thesis for comparison with the 30-minute data collected in 2014.

Table 3.3 summarises the data collected between 15th April and 21st June 2007. Over 3,000 observations were taken at each site across the valley, with very few invalid data points (*NaN*). This research focusses on wind direction observations, particularly in relation to three prevailing wind speed thresholds; 0ms^{-1} , 2ms^{-1} and 4ms^{-1} , observed at the ridge top stations. The second half of Table 3.3 shows the number of wind direction observations recorded when the wind speeds observed at the ridge top station A1 were above these three thresholds. The values clearly indicate that wind speed conditions were relatively slow since less than 1% of the wind direction data were observed at ridge top wind speeds greater than 4ms^{-1} .

Table 3.3: Summary of wind data collected across FCV between April and June 2007.

	A1	A2	A3	A4	A5
Total Observations	3,210				
<i>NaN</i>	<i>2</i>	<i>0</i>	<i>2</i>	<i>3</i>	<i>4</i>
Effective Observations	3,208	3,210	3,208	3,207	3,206
Wind Direction Observations					
A1 $\geq 0\text{ms}^{-1}$	3,191	2,809	2,823	2,964	2,161
A1 $\geq 2\text{ms}^{-1}$	396	396	396	396	333
A1 $\geq 4\text{ms}^{-1}$	23	23	23	23	16

Table 3.4 summarises the wind speeds experienced across FCV in 2007, with Figure A.4 in Appendix A showing the time series of wind speeds recorded across the collection period. Again, it is clear that this case study deals with a relatively low wind speed environment. Wind speeds within the valley are particularly low, at less than 60% of those experienced at the ridge top, with modal values of zero.

Table 3.4: Summary of observed wind speeds (ms^{-1}) across FCV between April and June 2007.

	A1	A2	A3	A4	A5
Mean	0.96	0.44	0.48	0.43	0.33
% of A1	—	45	56	45	35
Std	0.87	0.74	0.65	0.56	0.51
Mode	0.4	0	0	0	0
Min	0	0	0	0	0
Median	0.9	0	0.4	0.4	0
75th	1.3	0.4	0.9	0.9	0.4
95th	2.7	2.2	1.8	1.3	1.3
Max	4.9	4.9	4.5	3.6	3.1

Figure 3.5 shows the wind roses observed across the valley from April to July 2007, with the percentage values given in Table A.1 in Appendix A. The wind direction behaviours across FCV in 2007 are discussed in greater detail by Sharples et al. [2010] as well as later in this thesis. Figure 3.5(a) shows the bimodal structure of the prevailing wind directions across the region. A strong westerly mode represents the most dominant prevailing westerlies, while a smaller easterly mode indicates the presence of easterly prevailing winds, perhaps linked to a supposed ‘sea breeze’ effect documented for the Canberra region [Sharples et al., 2010]. On the east-facing slope, easterly winds dominate the flow pattern (Figure 3.5(b)), while on the valley floor, a bimodal north-south flow pattern exists that aligns with the axis of the valley (Figure 3.5(c)). On the west-facing slope, westerlies aligning with the dominant prevailing winds are most often experienced (Figures 3.5(d) and (e)).

3.2.3 Data Summary, 2014

Between 14th April and 10th July 2014, 30-minute data were collected across Flea Creek Valley. The installation of the RPi systems in July allowed for collection of 1-minute data until 15th December 2014. Table 3.5 outlines the data collected throughout the entire collection period from April to December 2014. With the increase in observation frequency, the volumes of data were far greater than collected in 2007. However, due to the data collection issues discussed previously, the numbers of invalid data points (*NaN*) were also far greater across the entire valley.

The number of observations with increasing ridge top wind speeds observed at B1 indicate a relatively low wind speed environment, as seen in 2007. Between April and July, average wind speeds were extremely low; even at the ridge top, average wind speeds were less than 1ms^{-1} , with a maximum wind speed of only 6.70ms^{-1} (see also Figure A.6 in Appendix A). Between July and December, the wind speeds remained low, despite stronger winds typically experienced in Spring and Summer; the average wind speed at B1 remained below 1ms^{-1} , with a maximum of only 5.81ms^{-1} . The reduction of wind speeds from the ridge top to the valley stations was enhanced in 2014 from 2007, with wind speeds across the valley only reaching a maximum of around 40% of those experienced at B1, except for high on the west-facing slope at B11 near the eastern ridge top (Table 3.6).

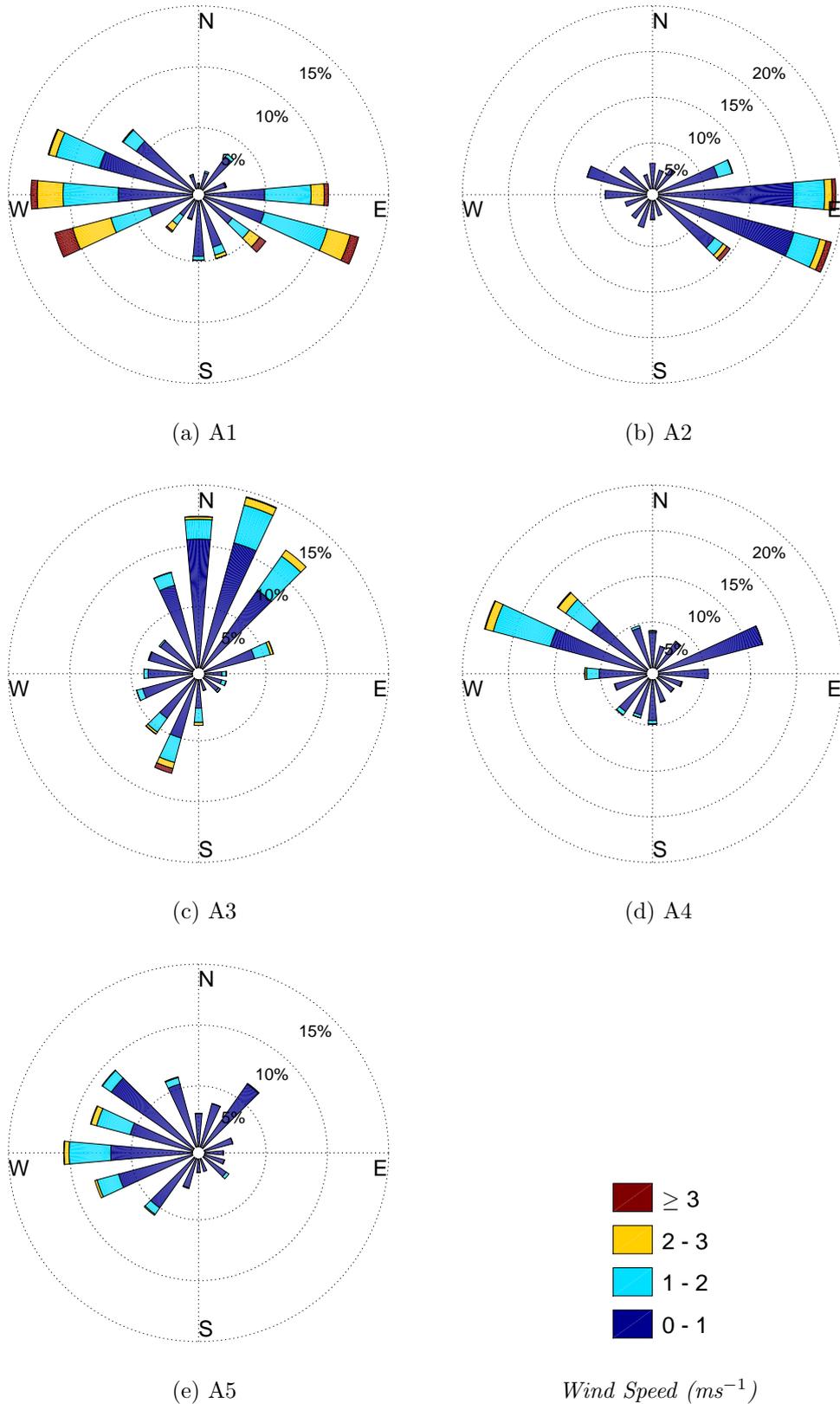


Figure 3.5: Observed wind direction roses across FCV at (a-e) A1 to A5 between April and June 2007.

Table 3.5: Summary wind data collected across FCV between April and December 2014.

	B1	B2	B3	B4	B5	B6
Total Observations	237,577					
<i>NaN</i>	65,896	156,077	153,980	36,800	44,160	24,428
Effective Observations	171,681	81,500	83,597	200,777	193,417	213,149
Wind Direction Observations						
$B1 \geq 0\text{ms}^{-1}$	141,844	47,554	39,890	100,267	55,562	88,705
$B1 \geq 2\text{ms}^{-1}$	13,784	7,465	6,239	10,384	6,426	12,338
$B1 \geq 4\text{ms}^{-1}$	605	438	405	378	129	588
	B7	B8	B9	B10	B11	
Total Observations						
<i>NaN</i>	47,812	127,861	166,313	63,453	38,308	
Effective Observations	189,765	109,716	71,264	174,124	199,269	
Wind Direction Observations						
$B1 \geq 0\text{ms}^{-1}$	44,326	42,863	26,527	65,152	96,419	
$B1 \geq 2\text{ms}^{-1}$	7,522	6,002	3,945	10,992	11,543	
$B1 \geq 4\text{ms}^{-1}$	283	166	105	537	567	

Table 3.6: Summary of observed wind speeds (ms^{-1}) across FCV between April and December 2014.

	B1	B2	B3	B4	B5	B6	B7	B8	B9	B10	B11
Mean	0.92	0.39	0.18	0.22	0.16	0.4	0.13	0.15	0.36	0.24	0.62
% of <i>B1</i>	—	42	20	24	17	43	14	16	39	26	67
Std	0.77	0.47	0.38	0.38	0.37	0.62	0.35	0.33	0.58	0.44	0.76
Mode	0.89	0	0	0	0	0	0	0	0	0	0
Min	0	0	0	0	0	0	0	0	0	0	0
Median	0.89	0.45	0	0	0	0	0	0	0	0	0.45
75th	1.34	0.89	0.45	0.45	0	0.89	0	0	0.45	0.45	0.89
95th	2.24	1.34	0.89	0.89	0.89	1.79	0.89	0.89	1.79	1.34	2.24
Max	6.70	4.02	4.02	4.47	5.36	6.71	5.81	3.58	4.47	4.47	6.71

The wind roses shown in Figure 3.6 (and percentages given in Table A.2 in Appendix A) show patterns of wind flow consistent with those observed in 2007. At the ridge top stations, both B1 and B2, the dominant wind direction was again westerly to north-westerly (Figures 3.6(a) and (b)). On the east-facing slopes at B3 and B4, dominant easterly modes were again observed (Figures 3.6(c) and (d)), and on the valley floor at B7 a bimodal north-south wind behaviour was again evident (Figure 3.6(g)). On the west-facing slope at B9 and B11, strong westerly winds were again observed, in line with the dominant westerly prevailing wind direction (Figures 3.6(i) and (j)).

At the remaining sites across the valley, B5, B6, B8 and B10, greater variation in wind direction is observed, and patterns are less consistent with those experienced

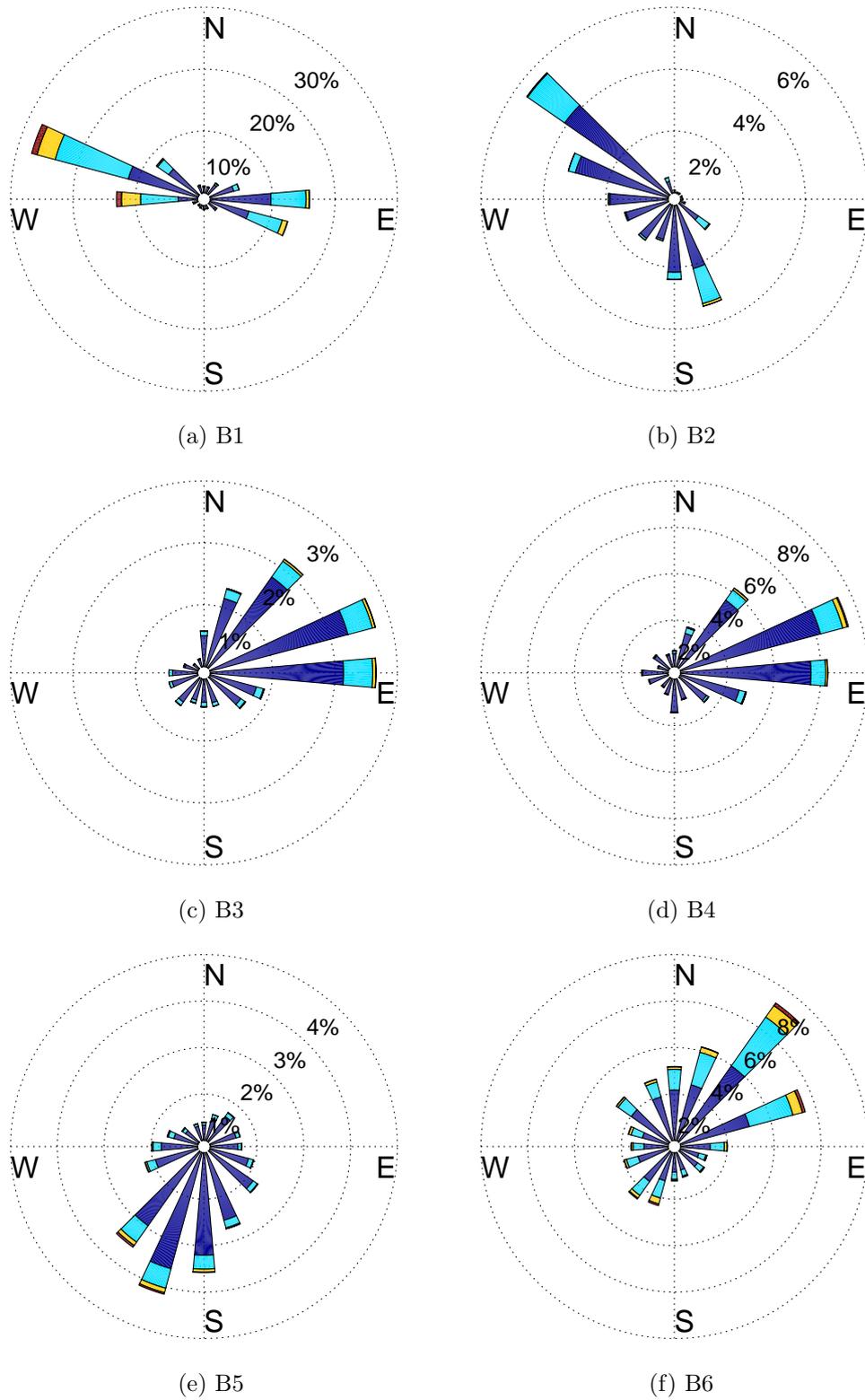


Figure 3.6: Observed wind direction roses across FCV at (a-k) B1 to B11 between April and December 2014.

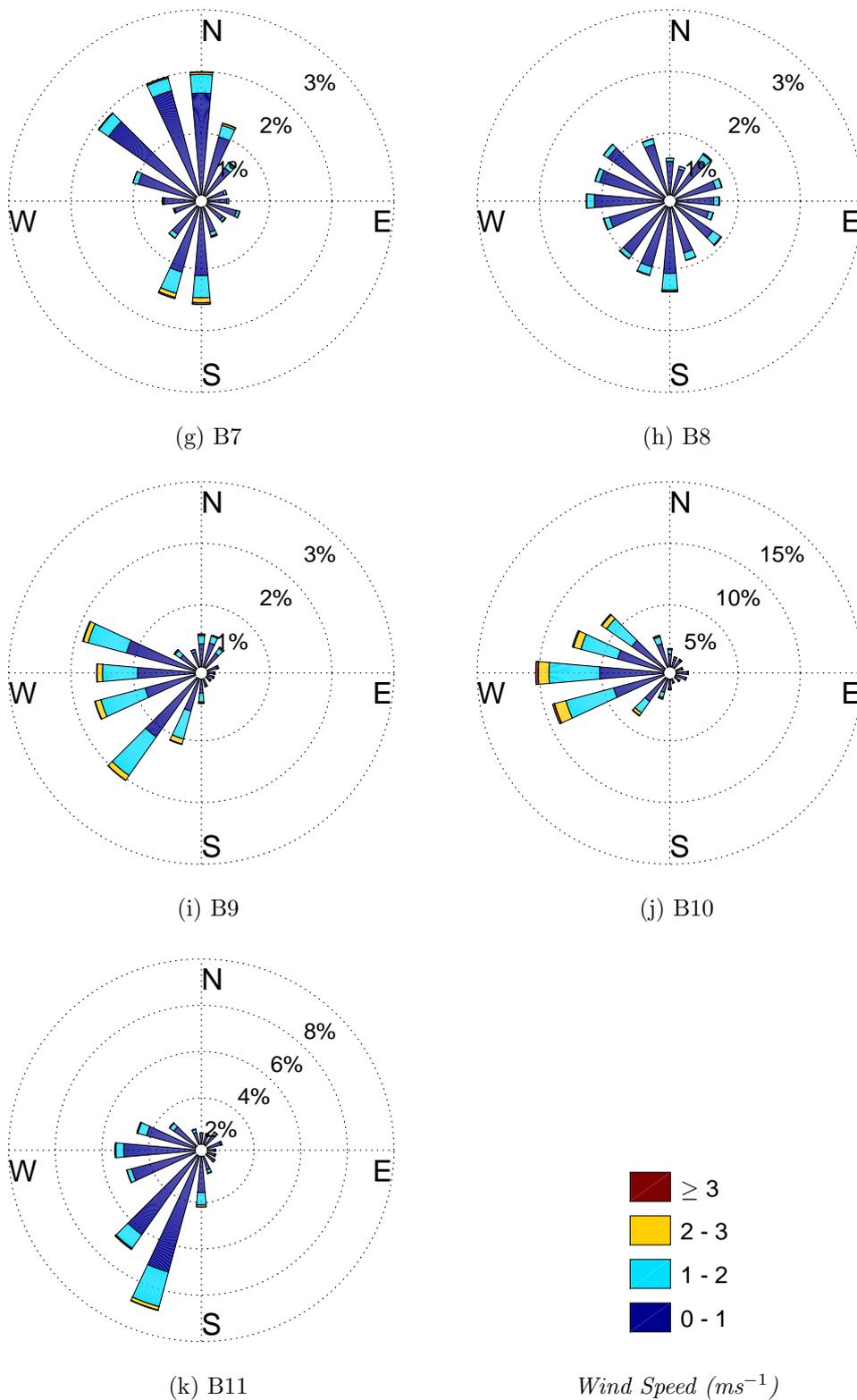


Figure 3.6: (Cont'd) Observed wind direction roses across FCV at (a-k) B1 to B11 between April and December 2014.

in 2007 (Figures 3.6(e), (f), (h) and (k), respectively). At each of these sites, it is likely that effects from local topography have had a significant impact on the wind direction behaviours. For instance, B6 is situated on the top of a knoll at the bottom of the valley which may result in increased variation in the observed wind direction. On the eastern slope of the valley, B5 in fact sits on a steep localised south-facing slope which would have significant impacts on local flows. Similarly, B8 is located close to a small gully on the valley wall which may again impact local wind flow. In addition, the increased vegetation resulting from the uninterrupted regrowth since 2007 may have impacted on the variation of wind directions observed at each site across Flea Creek Valley.

3.3 Case Study II: National Arboretum Canberra, ACT

3.3.1 Description

The National Arboretum Canberra (NAC) is located on the western shore of Lake Burley Griffin above Scrivener Dam in the ACT, overlooking Canberra from the west. To the west of the NAC, there are views out across the Murrumbidgee River to Uriarra State Forest and the eastern foothills of the Brindabella Ranges. Figures 3.7(a) and (b) show a regional maps indicating the location of the study area, and Figure 3.8(a) shows a photograph of the study region and its surroundings.

The region can be classified as undulating terrain [Weber et al., 2008, McRae and Sharples, 2013]⁵, with elevations ranging between 500m and 700m. The highest points in the local area are Black Mountain and Mount Ainslie, both reaching approximately 780m (Figure 3.7(b)). Although at a lower altitude, the temperate climate across the Canberra region is broadly similar to that of the Brindabella region with warm dry summers and cool winters, though snowfall is more rare. The synoptic wind patterns are again dominated by high-pressure systems producing WNW prevailing winds in the summer, and westerly winds in the winter.

The National Arboretum Canberra was developed after the Canberra 2001 and 2003 fires, and is made up of nearly 100 different forest stands of rare, endangered and symbolic trees. Theses stands are now at various levels of growth, with many

⁵www.highfirerisk.com.au/maps

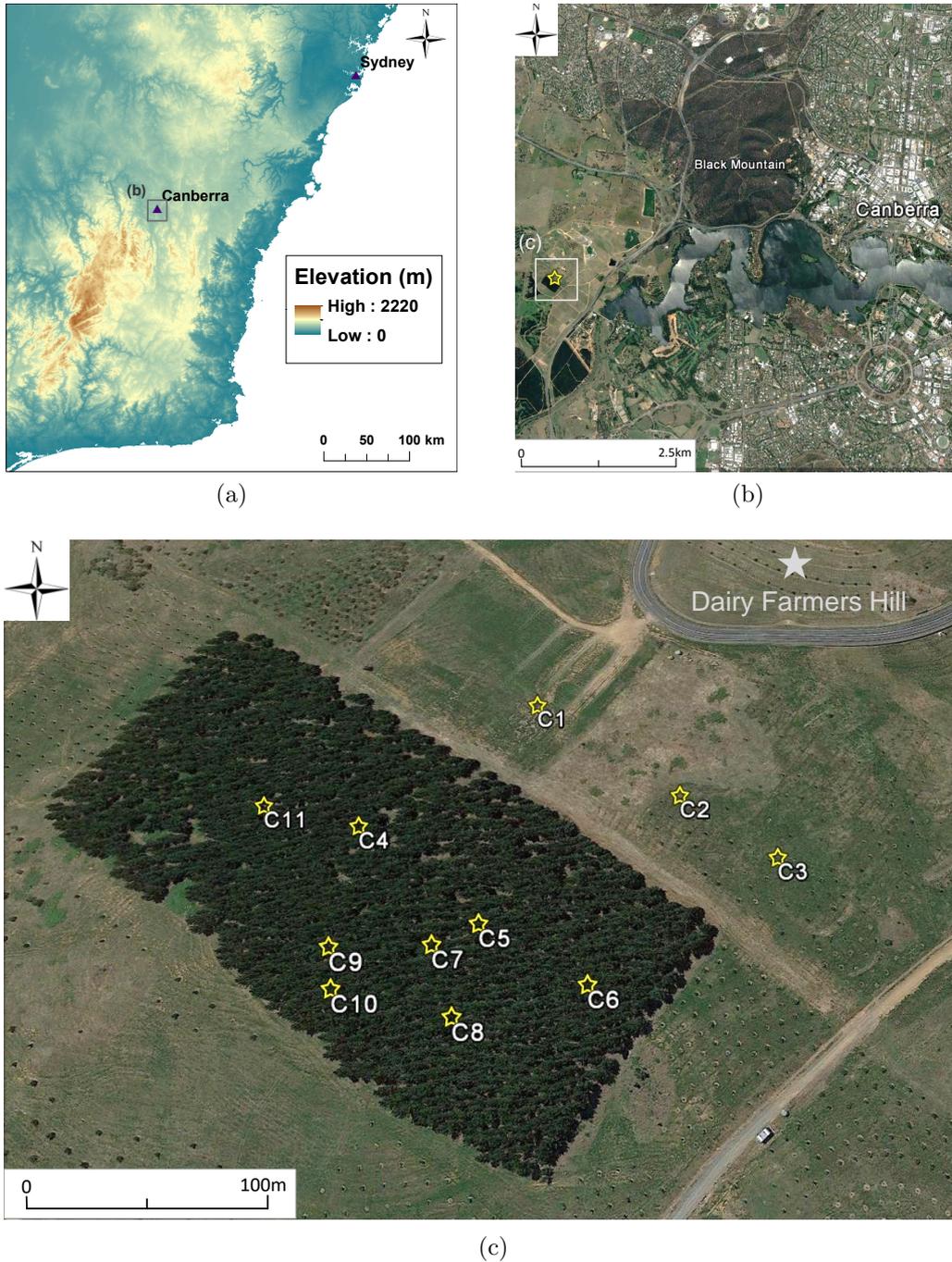


Figure 3.7: Map of (a) south-eastern Australia, (b) the Canberra region, and (c) the NAC. Station locations in 2015 (C1 to C11) are indicated.



Figure 3.8: Photograph of data collection at the NAC in 2015. Photograph taken by author from Dairy Farmer’s Hill.

still very young, but two forest stands are nearly 100 years old⁶. The NAC was selected as the second study area for this project since it would be possible to take advantage of these uniform vegetation stands for analysis of the wind fields across the area. Permission was granted by the NAC board to analyse the impacts of either contrasting vegetation cover or topography on wind fields, while controlling the other variable as much as possible.

Data were collected using the eleven PAWS between April and December 2015, across the Radiata pine stand (*Pinus radiata*; planted in 2004) and the clear stand adjacent to it (yet to be planted), near Dairy Farmers Hill (Figures 3.7(c) and 3.8). These stands lie along the highest ridges on the NAC site which is approximately perpendicular to the dominant WNW prevailing winds. The Radiata pine stand covers the spur at the southern end of the ridge. Figure 3.7(c) shows the locations of the stations throughout the Radiata pine and adjacent stands, and highlights the experimental design used to analyse wind fields under contrasting vegetation but similar topography, or uniform vegetation but varying topography. Table 3.7 provides the details of each of the sites used for data collection across the NAC. Elevations were estimated using Google Earth⁷, since analysis within ArcGIS, using a 90m DEM, led to a number of stations lying within the same pixels.

⁶www.nationalarboretum.act.gov.au

⁷Google Earth version 7.1.8.3036

Table 3.7: Descriptions of the NAC sites (C1 to C11) in 2015.

	Vegetation	Location	Lat.	Long.	Elev.	Slope
C1	Clear	Ridge Top	149.06275	-35.29381	658m	Flat
C2	Clear	SE-facing slope	149.06350	-35.29422	648m	Medium
C3	Clear	SE-facing slope	149.06440	-35.29447	637m	Medium
C4	Radiata pine	Ridge Top	149.06186	-35.29439	650m	Flat
C5	Radiata pine	SE-facing slope	149.06250	-35.29483	647m	Medium
C6	Radiata pine	SE-facing slope	149.06303	-35.29506	637m	Medium
C7	Radiata pine	S-facing slope	149.06228	-35.29492	645m	Medium
C8	Radiata pine	S-facing slope	149.06239	-35.29519	637m	Medium
C9	Radiata pine	SW-facing slope	149.06178	-35.29492	643m	Medium
C10	Radiata pine	SW-facing slope	149.06181	-35.29508	639m	Medium
C11	Radiata pine	NW-facing slope	149.06133	-35.29425	644m	Medium

Throughout the study period, data were collected at 1-minute intervals across the NAC. As at FCV, there were some difficulties due to the power usage of the RPi systems and the lack of solar recharge, particularly through the winter months and in areas of south-facing slopes. Figure A.7 in Appendix A shows the data record from the NAC through 2015; gaps are due to battery outages as well as difficulties in stabilising the stations in the soft soils present at the NAC; some stations fell over or were knocked down by wildlife (e.g. kangaroos).

3.3.2 Data Summary, 2015

At the NAC, the ridge top station was taken at C1 in the clear stand (Figure 3.7(c)). Analysis against observations taken by the Australian Bureau of Meteorology at Canberra Airport (presented later in Chapter 4, Section 4.2.2), show that observations from this station were indicative of the prevailing wind direction conditions throughout the study period.

Table 3.8 outlines the data collected at the NAC between 28th April and 31st December 2015. As in the second half of 2014 at FCV, the 1-minute observation frequency resulted in large volumes of data collected across the NAC. However, due to the previously discussed data collection issues, there were also a large quantity of invalid data points (*NaN*) across the entire study period (data record gaps are highlighted in Figure A.7 in Appendix A). Despite this, the volume of data collected at the NAC in 2015 far outweighed that collected at FCV in both 2007 and 2014.

Table 3.8: Summary of wind data collected across the NAC between April and December 2015.

	C1	C2	C3	C4	C5	C6
Total Observations	356,171					
<i>NaN</i>	83,658	102,254	3,173	31,659	235,539	114,542
Effective Observations	272,513	253,917	352,998	324,512	120,632	241,629
Wind Direction Observations						
C1 \geq 0ms ⁻¹	259,224	176,031	239,766	67,785	32,583	75,247
C1 \geq 2ms ⁻¹	169,797	118,867	165,852	65,350	30,977	64,192
C1 \geq 4ms ⁻¹	110,586	76,613	107,887	57,888	27,732	51,936
		C7	C8	C9	C10	C11
Total Observations						
<i>NaN</i>		247,419	62,259	117,894	80,831	154,562
Effective Observations		108,752	293,912	238,277	275,340	201,609
Wind Direction Observations						
C1 \geq 0ms ⁻¹		28,479	85,858	74,482	31,801	77,883
C1 \geq 2ms ⁻¹		27,750	79,328	68,251	29,018	71,093
C1 \geq 4ms ⁻¹		25,190	66,786	55,848	24,760	59,977

Table 3.9 summarises the wind speed (ms⁻¹) experienced across the NAC in 2015. On the clear slope, wind speeds were generally higher than those observed at the ridge top at FCV. With less dense vegetation across the region and less complex undulating terrain, these higher winds are likely explained by a lack of surface roughness surrounding the study site. However, considerably stronger winds were observed on the clear slope (C1, C2 and C3) than at all other stations located within the pine stand. The reduction in wind speeds caused by the pine plantation was far greater than that shown across FCV, with wind speeds observed in the pine stand at less than 10% of those observed on the clear ridge top. The time series of wind speeds across the entire study period, shown in Figure A.8 (Appendix A), also highlight this dramatic wind speed reduction.

Table 3.9: Summary of observed wind speeds (ms⁻¹) across the NAC between April and December 2015.

	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11
Mean	3.64	2.30	1.87	0.04	0.13	0.09	0.06	0.22	0.17	0.06	0.31
% of C1	—	63	51	1	3	3	2	6	5	2	9
Std	2.86	2.06	1.69	0.17	0.34	0.26	0.18	0.45	0.39	0.24	0.58
Mode	1.34	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Min	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Median	2.68	1.79	1.34	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
75th	5.36	3.58	2.68	0.00	0.00	0.00	0.00	0.45	0.00	0.00	0.45
95th	8.94	6.26	5.36	0.45	0.89	0.45	0.45	1.34	0.89	0.45	1.34
Max	20.12	16.99	15.20	3.58	3.58	4.02	2.68	4.47	4.47	4.92	5.36

Figure 3.9 shows the wind roses for each site across the NAC. On the ridge top at C1 the strong westerly prevailing winds experienced across the ACT region are again evident, as well as the less dominant easterly prevailing winds (Figure 3.9(a)). On the clear east-facing slope at C2 and C3, similar bimodal west-east modes are also observed (Figures 3.9(b) and (c)), suggesting that wind directions experienced at these locations align with the prevailing winds. Within the pines, the wind directions observed at C4, highlight the increased variability on the ridge top caused by the introduction of pine trees (Figure 3.9(d)).

Within the pine stand, the wind direction behaviours on the east-facing slope, parallel to C2 and C3 are very different. Figures 3.9(e) and (f) show that the dominant wind directions observed at C5 and C6 are easterlies, almost 180° different to those on the parallel clear slope. These easterly modes are also evident at C7 and C8 on the south-east facing slope (Figures 3.9(g) and (h)). On the south-facing slope at the end of the spur, on the cross-slope to the dominant westerly winds, C9 and C10 record easterly and southerly winds (Figures 3.9(i) and (j)). Finally, on the west-facing slope at C11, westerly winds are most commonly observed (Figure 3.9(k)), aligning with the dominant prevailing wind direction and in opposition to the less dominant easterly prevailing winds.

3.4 Concluding Remarks

As discussed in Chapter 2, there is a lack of wind data available at resolutions most relevant to modelling surface fire behaviour, in particular datasets collected over complex terrain are rare. For this reason, a number of gaps exist in the development and validation of wind models for application to bushfire prediction. The first contribution of this thesis has been to introduce two new datasets that fulfil such requirements.

This chapter has outlined two case study regions, and summarised their respective observed wind datasets. The first case study at Flea Creek Valley, while building on that presented by Sharples et al. [2010], provides an opportunity to analyse the impacts of post-fire regrowth on surface winds observed across an area of complex terrain. The second case study at the National Arboretum Canberra complements

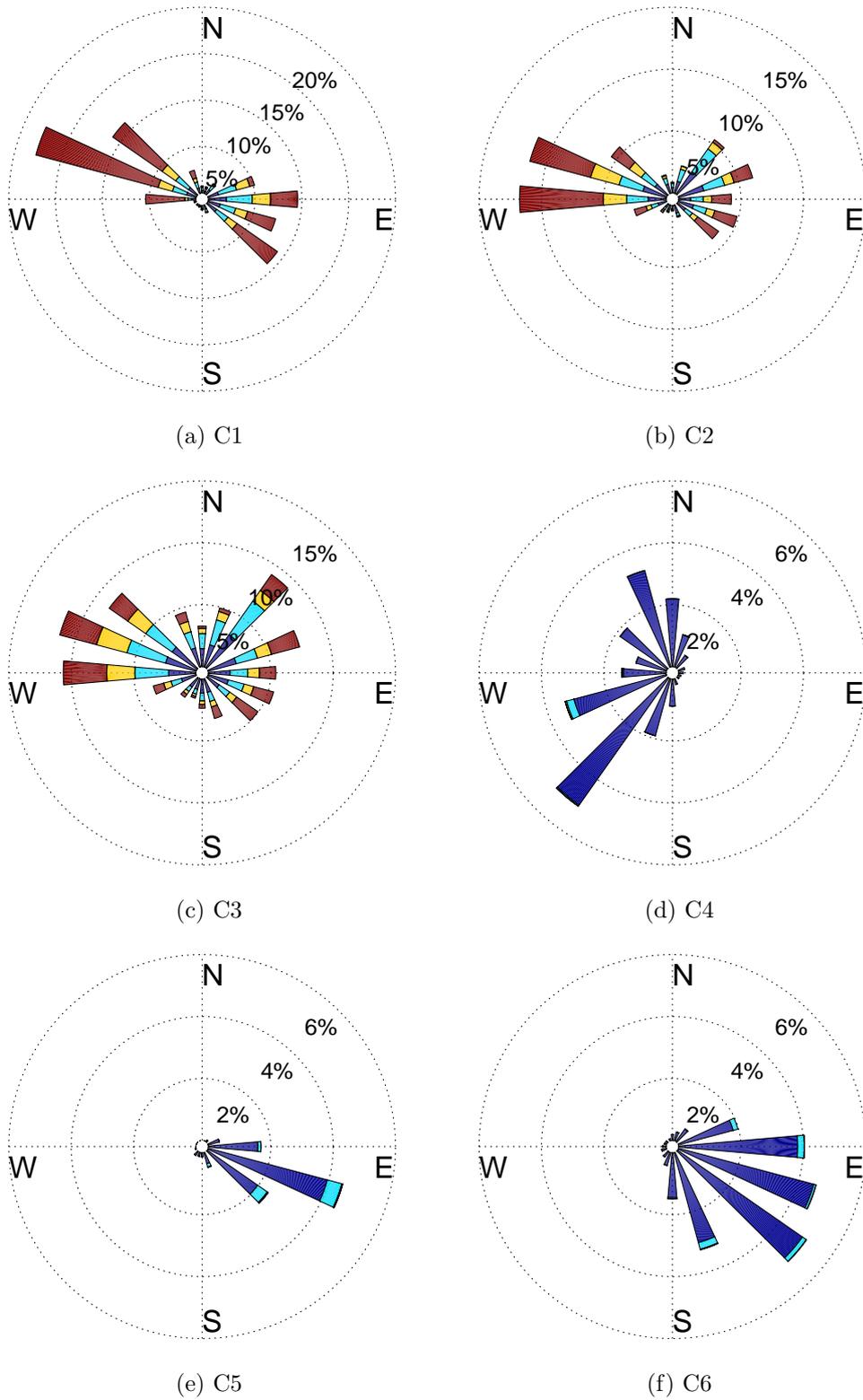


Figure 3.9: Observed wind direction roses across the NAC at (a-k) C1 to C11 between April and December 2015.

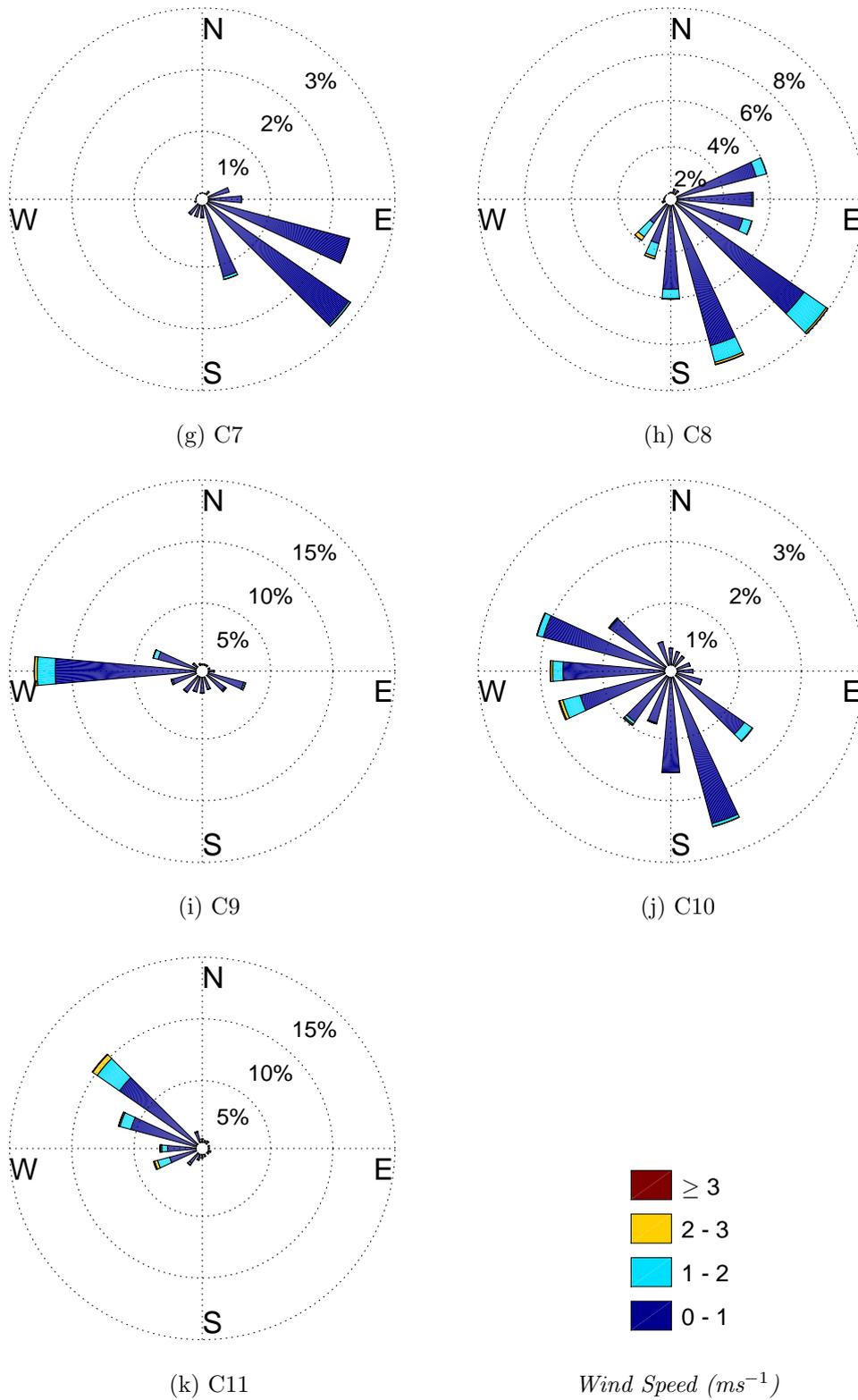


Figure 3.9: (Cont'd) Observed wind direction roses across the NAC at (a-k) C1 to C11 between April and December 2015.

the first by considering a more uniform vegetation environment over which wind characteristics can be compared and contrasted. The availability of these datasets is outlined in Appendix A.

Within this thesis, the data presented in this chapter are assumed to be independent and identically distributed since more detailed study of spatial and temporal correlations are beyond the scope of this research. It is expected that these correlations would be resolution dependent. With regards to temporal correlation, it is reasonable to assume 30-minute data to be uncorrelated and, with wind gusts measured over 3-second intervals, even 1-minute data may only be weakly correlated. Spatial correlations would also depend on temporal resolution as well as vegetation cover. At smaller temporal resolutions, spatial correlation might be time-lagged as prevailing winds move across the valley. However, with significant vegetation cover, it is reasonable to assume minimal spatial correlation between sites.

Through subsequent analysis of these datasets, this thesis seeks to fulfil the three remaining research contributions outlined in Chapter 1, Section 1.2. The data presented in this chapter, could however be utilised in numerous other ways for the advancement of bushfire modelling as well as within other research areas. As a single example, the secondary variables of temperature, relative humidity or solar radiation could be combined with the wind data to provide further insights into micro-meteorology within different topographical and vegetation conditions.

CHAPTER 4

Probabilistic Representation of Wind Direction

A fundamental aim of this research is to recast wind fields in probabilistic terms. To do this, the response of prevailing wind fields to changes in the landscape beneath them are represented as joint distributions of the prevailing wind direction and the wind direction experienced at the surface, as introduced in Chapter 2 (see Section 2.3.1). Throughout this thesis, these joint distributions will be referred to as directional wind response distributions.

Figure 4.1 shows a conceptual diagram of how directional wind response was observed across Flea Creek Valley in 2014 (see Chapter 3, Section 3.2 for more details on data collection). Figures 4.1 (a-c) show the discrete grids of observed joint wind direction distributions at sites on the east-facing slope (B4), the valley floor (B6) and the west-facing slope (B10). The yellow modes of these distributions represent the most frequently observed wind direction pairs for each site, which can be used to identify wind behaviours. For instance, in Figure 4.1(a), the off-diagonal mode evident at B4 indicates the prevalence of wind reversals when the east-facing site is on the leeward slope to westerly prevailing winds. Figure 4.1(b) shows that two modes are apparent on the valley floor at B6 under westerly prevailing winds, one of which crosses the horizontal boundary of the planar data view. This bimodal structure is indicative of broad-scale channelling of wind flow through the valley. Finally, Figure 4.1(c) shows the alignment between westerly prevailing winds and westerly winds experienced on the west-facing slope at B10.

These directional wind response distributions can incorporate the impacts of wind speed through using a wind speed threshold to constrain the data used to construct

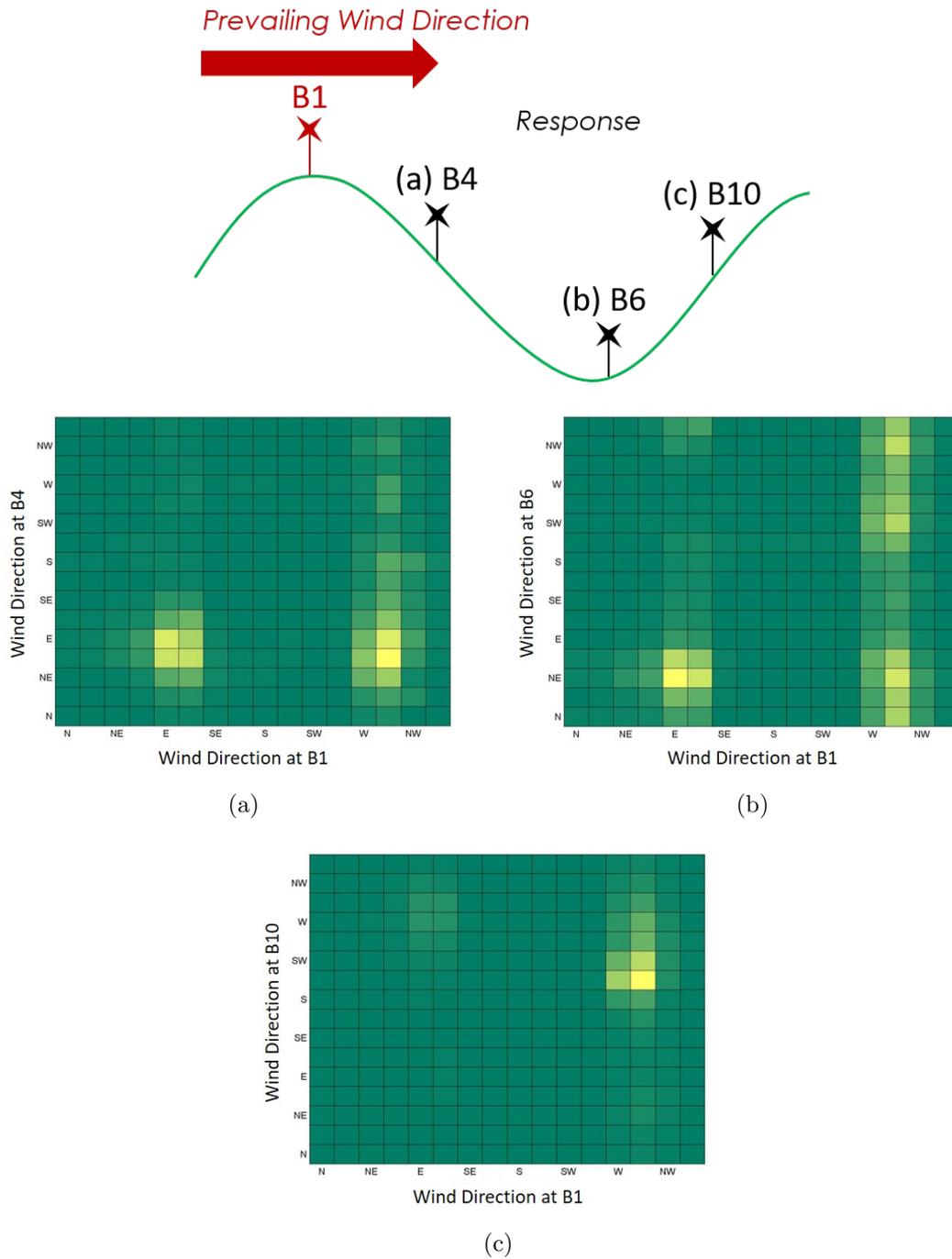


Figure 4.1: Conceptual diagram of data collection across FCV in 2014, and the observed directional wind response on the (a) east-facing slope at B4, (b) valley floor at B6 and (c) west-facing slope at B10.

the distribution. In Appendix B, minimum wind speed thresholds for the prevailing winds (0ms^{-1} , 2ms^{-1} and 4ms^{-1}) are used to show how the directional wind response changes as the prevailing wind speeds increase, similar to those shown by Sharples et al. [2010].

As discussed in Chapter 2, all real datasets contain an unknown level of environmental and instrumentation noise. The example data shown in Figure 4.1 are therefore taken to be noisy realisations of the true directional wind response at each point across the valley. In order to better understand wind behaviour across the landscape, and therefore contribute to a better understanding of fire behaviour in complex terrain, a number of estimation methods can be applied to observed bivariate wind direction data to estimate the true underlying directional wind response.

Due to the circular nature of wind direction, i.e. an observation of northerly winds can be represented at 0° or 360° , analyses of bivariate (thus toroidal) wind direction distributions require specialised statistical techniques. In this chapter, a selection of estimation methods, introduced in Chapter 2, are presented and compared in terms of their overall accuracy and their ability to capture the toroidal nature of the data. The most appropriate estimation technique is then applied to observed wind direction data to estimate the underlying directional wind response across Flea Creek Valley in 2007 and 2014, and the National Arboretum Canberra in 2015.

This research has been presented in the following journal article;

Quill, R., Sharples, J.J. and Sidhu, L.A. (In Prep) Estimation of directional wind response using noisy bivariate circular data: a comparison of approaches.

4.1 Estimation of Toroidal Surfaces

4.1.1 Estimation Techniques

Each of the techniques described in the following subsections were implemented and analysed within MATLAB® [2016]. Further details regarding the code used for this investigation are provided in Appendix C.

4.1.1.1 Cubic spline with and without periodic boundary conditions

The cubic spline estimates a continuous surface by exactly interpolating between the set (or a subset) of known data points. Two forms of the cubic spline were implemented using the `csape` function in MATLAB® [2016], which uses a tensor product of univariate cubic splines to estimate bivariate surfaces. The first form of the spline (herein referred to as ‘Cubic’) had no specified boundary conditions, thus not inherently allowing for periodicity of any given dataset. The second form of the cubic spline (herein referred to as ‘Cyclic Cubic’) was applied using periodic boundary conditions which ensured that the first and second derivatives of the first and last data point in each dimension match. This condition is designed to account for circularity in the dataset.

4.1.1.2 Thin plate smoothing spline

In contrast to exact interpolation, the thin plate smoothing (TPS) spline estimates the unique smooth function, f , that minimises

$$p \sum_j \sum_i |Y(X_{ij}) - f(X_{ij})|^2 + (1 - p) \int_{\mathbb{R}^2} (|\partial_\phi^2 f|^2 + 2|\partial_\phi \partial_\psi f|^2 + |\partial_\psi^2 f|^2) d\phi d\psi, \quad (4.1)$$

with smoothing parameter $p \in (0, 1)$. The smoothing parameter controls the trade-off between data fidelity (considered in the first term) and the smoothness of the resulting surface (accounted for by the second term, which represents a roughness penalty).

Optimal selection of the smoothing parameter ensures that the approximation does not over-fit noise in the data. There are numerous ways in which the smoothing parameter can be estimated or optimised, including generalised cross-validation [Hancock and Hutchinson, 2006] or ordinary cross-validation [Wahba, 1990]. The smoothing parameter can also be fixed [Sharples et al., 2010], and for the sake of methodological comparison in the first section of this study the smoothing parameter in the TPS minimisation function (4.1) is set at $p = 0.01$. This value is selected through visual comparison of the estimated surfaces with those produced by

Sharples et al. [2010], as well as consideration of the accuracy of the cup anemometers used to collect the raw data. This smoothing parameter is fixed throughout the application of Algorithm 1 (introduced in Section 4.1.1.4) to ensure the resulting estimated surfaces are comparable. In more detailed analysis of the TPS spline in Section 4.1.4, the smoothing parameter is varied.

The TPS spline was implemented using the `tpaps` function in MATLAB® [2016]. This technique does not inherently account for the circularity in toroidal data, but as discussed in Chapter 2, there is limited literature on the application of toroidal smoothing splines. As with the non-circular cubic spline and the kernel density method (outlined below), in this study the planar TPS spline technique is adapted using Algorithm 1 (defined later in Section 4.1.1.4), to account for circularity.

4.1.1.3 Kernel Density Estimation with Normal and von Mises kernels

The final form of estimation considered is kernel density estimation (KDE). This method relies on the specification of a kernel distribution to be used as a basis for the approximation. Firstly, it is common for the Normal distribution to be used as a kernel (herein referred to as ‘KDE Normal’), but this does not account for circularity. Therefore, the circular von Mises distribution is also used as a kernel (herein referred to as ‘KDE von Mises’). As with the cubic splines, a product kernel is used for both the KDE methods to approximate the bivariate density. The `ksdensity` function is used in MATLAB® [2016] and code for KDE von Mises was adapted from the `NPCirc` package developed for R [Oliveira et al., 2014a, R Core Team, 2016].

Similar to the TPS spline, the trade off between data fidelity (or bias) and smoothness (or variance) is controlled by a parameter, which for KDE methods is referred to as the bandwidth. This bandwidth can again be selected using numerous methods including various forms of cross-validation, as well as a number of rules constructed for standard kernels such as Silverman’s rule of thumb for Normal kernels [Silverman, 1986] or the plug-in formula devised for toroidal kernels [Oliveira et al., 2012]. In the `ksdensity` package, Silverman’s rule of thumb is implemented to select the optimal bandwidth for the Normal kernel. The optimal bandwidth is first selected

for approximation without application of Algorithm 1, and then fixed to that bandwidth for further iterations. The optimal bandwidth for the von Mises kernel is approximated using likelihood cross-validation (LCV) in this study, with implementation conducted via the `bt.cv` function in R, provided in the `NPCirc` package [Oliveira et al., 2013, 2014a]. Within this package a number of selection methods are available, including the rule of thumb devised by Taylor [2008] (the analogous selection method for the toroidal kernel to that of Silverman’s rule of thumb for the Normal kernel), many of which selected the same or similar bandwidth values to that of the LCV method.

4.1.1.4 Data Wrapping Algorithm

As an extension of the technique discussed by Silverman [1986], to capture circularity in bivariate data, discrete data can be repeated on a row-by-row and column-by-column basis [e.g. Sharples et al., 2010]. While previous studies have considered or applied the technique in limited ways, here a more comprehensive appraisal of data wrapping in the bivariate context is presented.

Algorithm 1 outlines the generalised procedure by which this method can be conducted, and Figure 4.2 gives an example schematic over a single iteration ($r = 1$) with $n = 2$. It is important to note that a single iteration of Algorithm 1 applied to a discrete $n \times n$ data grid, $G = [g_{ij}]_{n \times n}$, produces a discrete $(n + 2) \times (n + 2)$ grid, $G_1 = [g_{ij}]_{(n+2) \times (n+2)}$.

The repeated grid G_r (after r iterations of Algorithm 1) begins to unveil the periodicity of the data in both dimensions, prior to the application of planar estimation methods. The resulting smooth surface is estimated across the entire repeated grid, and then restricted to the extent of G in order to be representative of the torus.

Algorithm 1 Compute r iterations around $n \times n$ data grid $G = [g_{ij}]_{n \times n}$

- 1: **procedure** ITERATION(G, r)
 - 2: $G = [g_{ij}]_{n \times n}$
 - 3: **for** r iterations around G **do**
 - 4: $G_r[r + 1 : r + n, r + 1 : r + n] = G[1 : n, 1 : n]$
 - 5: $G_r[n + r + 1 : n + r + r, :] = G_r[r + 1 : r + r, :]$
 - 6: $G_r[:, n + r + 1 : n + r + r] = G_r[:, r + 1 : r + r]$
 - 7: $G_r[1 : r, :] = G_r[n + 1 : n + r, :]$
 - 8: $G_r[:, 1 : r] = G_r[:, n + 1 : n + r]$
 - 9: **end for**
 - 10: Return $G_r = [g_{ij}]_{(n+2r) \times (n+2r)}$
 - 11: **end procedure**
-

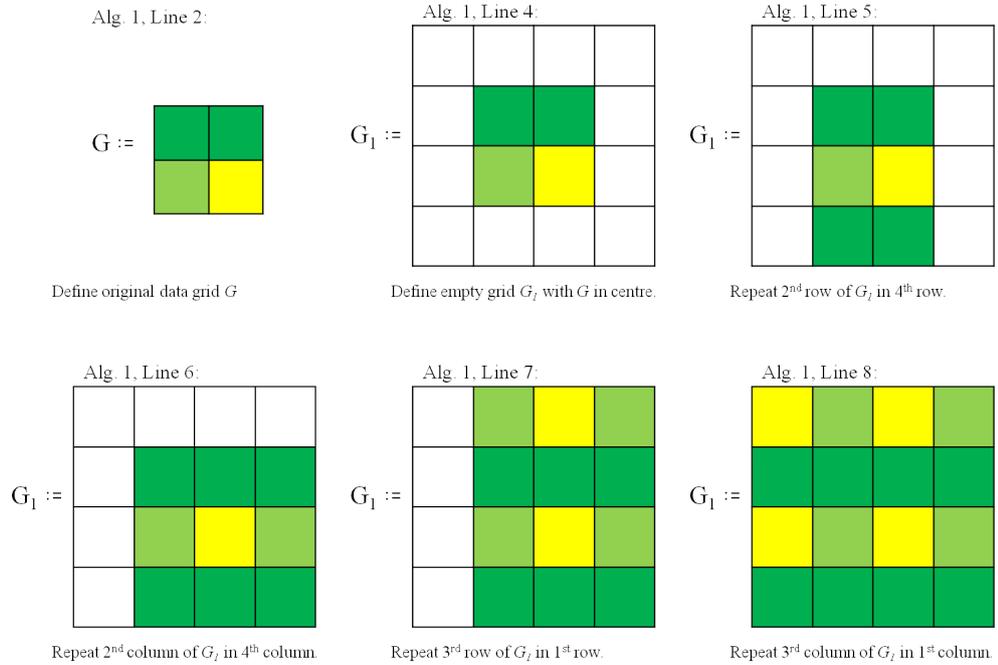


Figure 4.2: Schematic diagram of Algorithm 1 with one iteration ($r = 1$) applied to a 2×2 data grid ($n = 2$).

4.1.2 Simulation Study Data and Methods

4.1.2.1 Toroidal Data

A synthetic toroidal surface, f_k , was simulated such that $f_k = \max(f_k^*, 0)$ with $k = A, B, C$. For instance

$$f_A^*(\phi, \psi) = \sin(\phi - \mu_A) \cos(\psi - \tau_A), \quad (4.2)$$

where $\phi, \psi \in [0^\circ, 360^\circ)$, and the parameters μ_A and τ_A are the mean values of ϕ and ψ respectively. Negative values were truncated to zero to emulate the observed wind direction surfaces. Figure 4.3 shows the continuous surface of f_A with $\mu_A = \tau_A = 90^\circ$. The cases where $k = B, C$ are discussed later in this chapter.

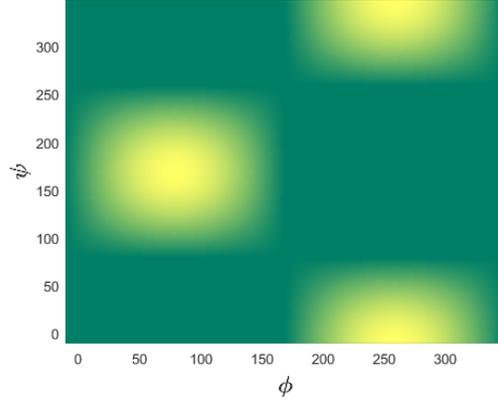


Figure 4.3: Continuous surface f_A , with $\mu_A = \tau_A = 90^\circ$.

Discrete sample data, Y_A , were simulated with noise according to Equation 2.8 (Chapter 2, Section 2.3.1), where $\epsilon_A \sim N(0, \sigma_A^2)$ are independent and identically distributed random variables with $\sigma_A = 0.25$.

All estimation methods described in the previous section were then applied to Y_k to estimate the continuous surface, f_k , from the discrete sample. The planar estimation methods, Cubic, TPS and KDE Normal, were applied to Y_k in conjunction with Algorithm 1. The toroidal estimation methods, Cyclic Cubic and KDE von Mises, were applied to Y_k with no repetition of the data to provide benchmark results for inherently toroidal estimation techniques.

4.1.2.2 Measures of Performance

The estimated smooth surface, \hat{f}_k , was realised on a fine 1° resolution discrete grid, $X_{st} = \{(\phi_s, \psi_t) : s, t = 1, \dots, 360\}$. The performance of the surface estimation methods was assessed using the mean squared error (MSE), as well as two difference metrics calculated separately along the imposed edges of each dimension, ϕ and ψ . To understand the average performance of the estimation techniques, 250 Monte Carlo simulations were conducted, and the mean and 95% confidence intervals (CI) of the following performance metrics were plotted against r , the number of iterations of Algorithm 1.

Mean squared error (MSE) was defined as

$$\frac{1}{360^2} \sum_{s=1}^{360} \sum_{t=1}^{360} \left\{ \hat{f}_k(\phi_s, \psi_t) - f_k(\phi_s, \psi_t) \right\}^2, \quad (4.3)$$

where f_k is the true surface. Although this statistic is named the mean squared difference, it is technically an average since the smoothing methods produce bias in the surface estimates and so the mean squared error should involve a bias term. In this study, however, the term ‘mean’ will be used for this statistic in order to clarify the discussion of its average value across a number of simulations.

The mean difference in edge values, row-wise (along ϕ , with fixed ψ_0 and ψ_{360}), was defined as

$$\frac{1}{360} \sum_{s=1}^{360} \left\{ \hat{f}_k(\phi_s, \psi_0) - \hat{f}_k(\phi_s, \psi_{360}) \right\}, \quad (4.4)$$

and similarly for column-wise (along ψ , with fixed ϕ_0 and ϕ_{360}).

Finally, the mean difference in first derivative at the edges, row-wise (along ϕ , with fixed $\psi_0, \psi_1, \psi_{360}$ and ψ_{361}), was defined as

$$\begin{aligned} \frac{1}{360} \sum_{s=1}^{360} \left\{ \delta \hat{f}_{k,0} - \delta \hat{f}_{k,360} \right\} = \\ \frac{1}{360} \sum_{s=1}^{360} \left\{ \left[\hat{f}_k(\phi_s, \psi_1) - \hat{f}_k(\phi_s, \psi_0) \right] - \left[\hat{f}_k(\phi_s, \psi_{361}) - \hat{f}_k(\phi_s, \psi_{360}) \right] \right\}, \end{aligned} \quad (4.5)$$

and similarly column-wise (along ψ , with fixed $\phi_0, \phi_1, \phi_{360}$ and ϕ_{361}).

Appendix C provides the MATLAB® [2016] code used to calculate these measures.

4.1.3 Simulation Study Results

4.1.3.1 Estimation Time

Figure 4.4 shows the average time taken (over 250 Monte Carlo simulations) to perform an individual surface approximation for each estimation method under increasing iterations of Algorithm 1.

Under zero iterations, it is clear that, compared to either KDE method, the TPS spline and both Cubic splines took very little time to conduct each approximation.

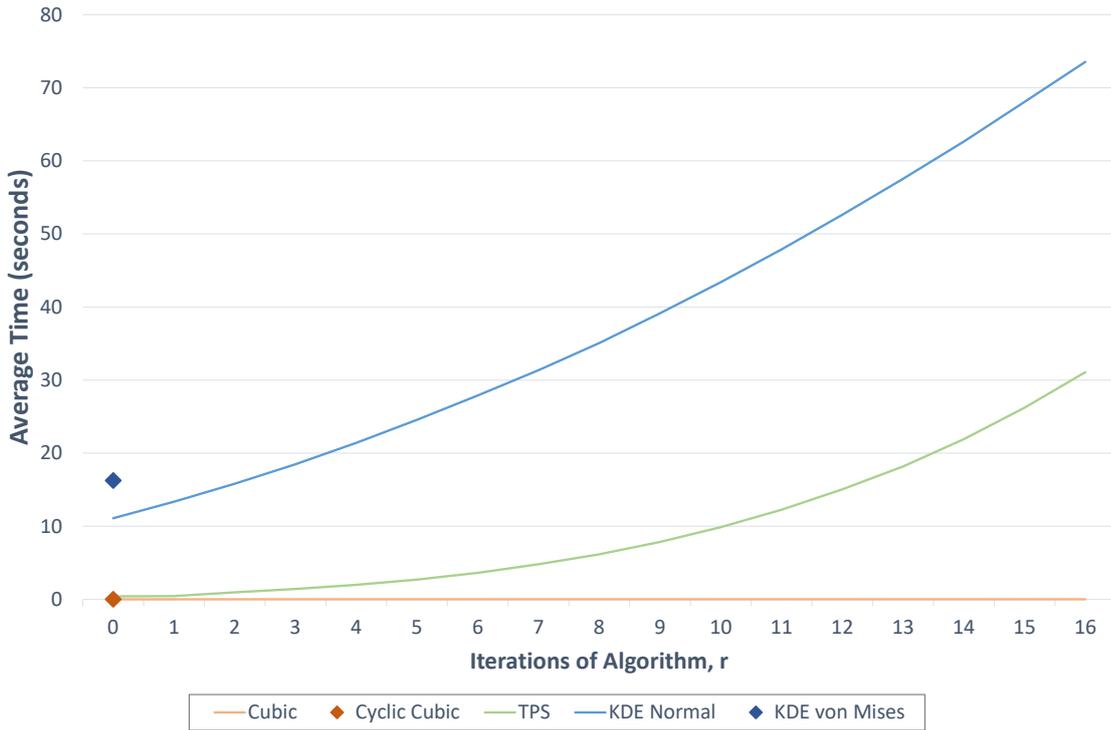


Figure 4.4: Average time (seconds) taken for individual approximations for each estimation method applied to Y_A . Diamonds represent (orange) Cyclic Cubic and (blue) KDE von Mises. Lines represent (orange) Cubic, (green) TPS and (blue) KDE Normal, in conjunction with Algorithm 1.

The KDE von Mises estimation techniques took around 17 seconds to approximate a single surface, while the KDE Normal method took approximately 11 seconds per approximation. Comparatively, the TPS spline with zero iterations of Algorithm 1 took on average 0.41 seconds per approximation, and both Cubic splines took less than 0.002 seconds per approximation.

With increasing iterations of Algorithm 1, the Cubic spline (as an exact interpolation method) took negligible time to conduct the approximations, while the TPS spline and KDE Normal method increased in computational demand. The TPS spline showed an exponential increase in time taken, while the KDE Normal method increased approximately linearly in time taken against number of iterations. The average time taken to conduct an approximation with complete repetition of the discrete observed grid, i.e. $r = 16$, for the TPS spline was approximately 31 seconds, while the KDE Normal method took almost 74 seconds per surface approximation.

4.1.3.2 Overall Surface Estimation Accuracy

Figure 4.5 shows the average mean squared error (MSE), and its associated 95% confidence interval (CI), over 250 Monte Carlo simulations, for each of the planar estimation methods with increasing iterations of Algorithm 1. The average MSE resulting from the toroidal estimation methods without the algorithm are also shown for zero iterations of Algorithm 1 (diamonds).

The Cyclic Cubic method matches the edge derivatives of the surface and is therefore equivalent to performing the Cubic under only one iteration of Algorithm 1. This is evident in Figure 4.5 since the Cubic results converge to a constant MSE, approximately equal to the Cyclic Cubic estimation MSE, after a single iteration.

Surprisingly, the TPS spline shows the lowest MSE results, even without any iterations of Algorithm 1 when the method does not account for circularity. As the toroidal nature of the surface is captured with increasing iterations of the algorithm, the MSE for the TPS spline consistently decreases. After approximately 6-8 iterations of Algorithm 1, the MSE levels out to an approximately constant value.

The largest MSE values are given by the KDE approximation techniques, but with the smallest variation. The KDE von Mises, inherently designed to account for toroidality in the data, performs least well with a very high MSE value. In contrast, the KDE Normal approximation gives MSE values only slightly higher than, but still in the order of, those given by Cubic and TPS methods. As with the Cubic and TPS methods, the KDE Normal results converge to a relatively constant MSE value. This occurs after only one iteration of Algorithm 1.

4.1.3.3 Estimation of Toroidal Characteristics

To indicate how well the planar estimation methods capture the toroidal nature of the datasets, Figure 4.6 shows the average difference in edge values and edge derivatives for the three planar methods under increasing iterations of Algorithm 1. Since the Cubic spline interpolates the data, it matches the edge values and derivatives of the surface after only 4 or 5 iterations of Algorithm 1 (Figures 4.6(a) and (b)). Figures 4.6(c) and (d) also suggest that the KDE Normal approximates the toroidal nature of the surface very well after only a few iterations of Algorithm 1. For the

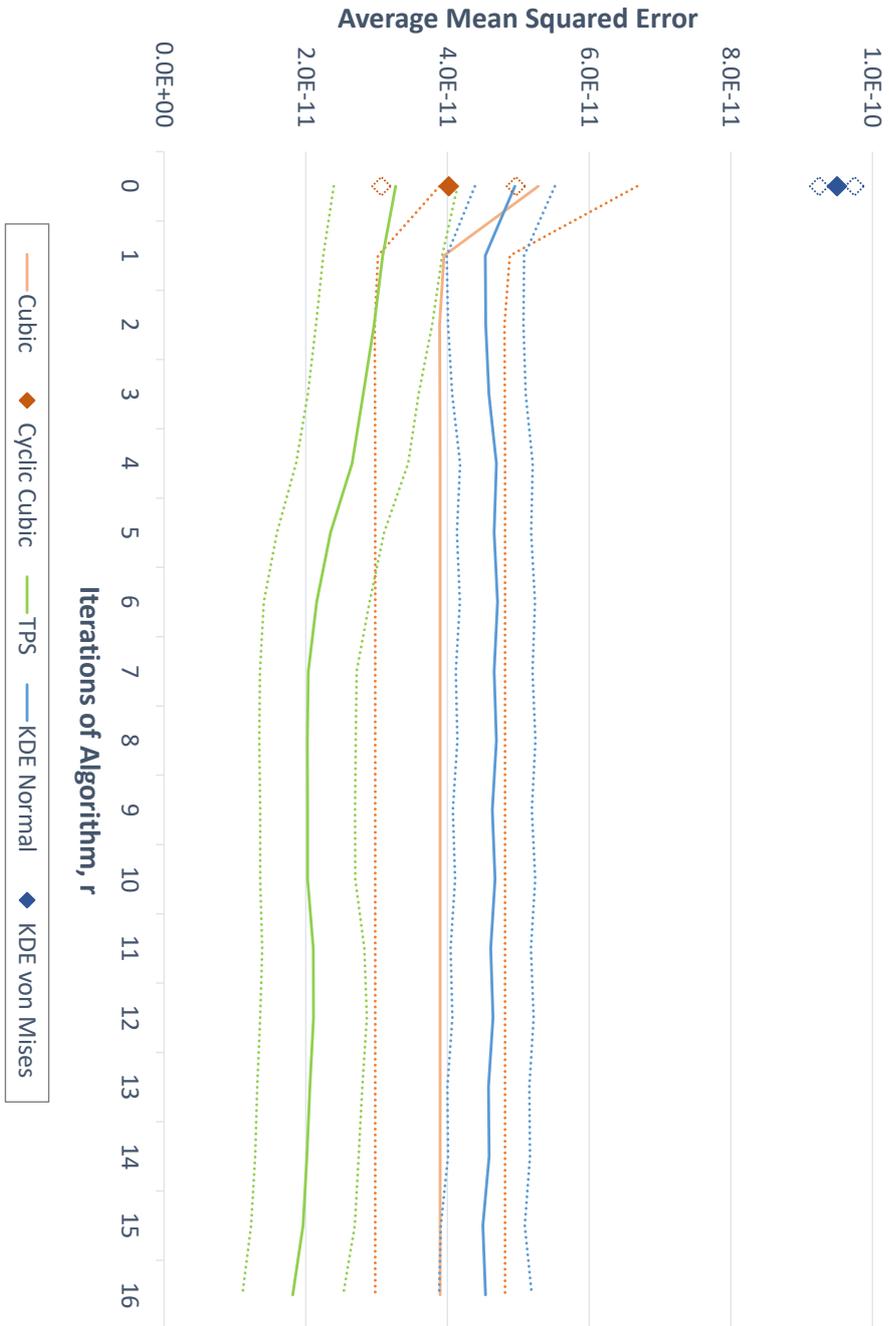


Figure 4.5: Average (solid) Mean Squared Error, and 95% CI (dotted), for each estimation method applied to Y_A , over 250 MC simulations. Diamonds represent (orange) Cyclic Cubic and (blue) KDE von Mises. Lines represent (orange) Cubic, (green) TPS and (blue) KDE Normal, in conjunction with Algorithm 1.

TPS spline, Figure 4.6(f) shows this consistent behaviour with decreasing difference in derivative values as the iterations of Algorithm 1 are increased, although with far greater variation than shown for the Cubic or KDE Normal methods.

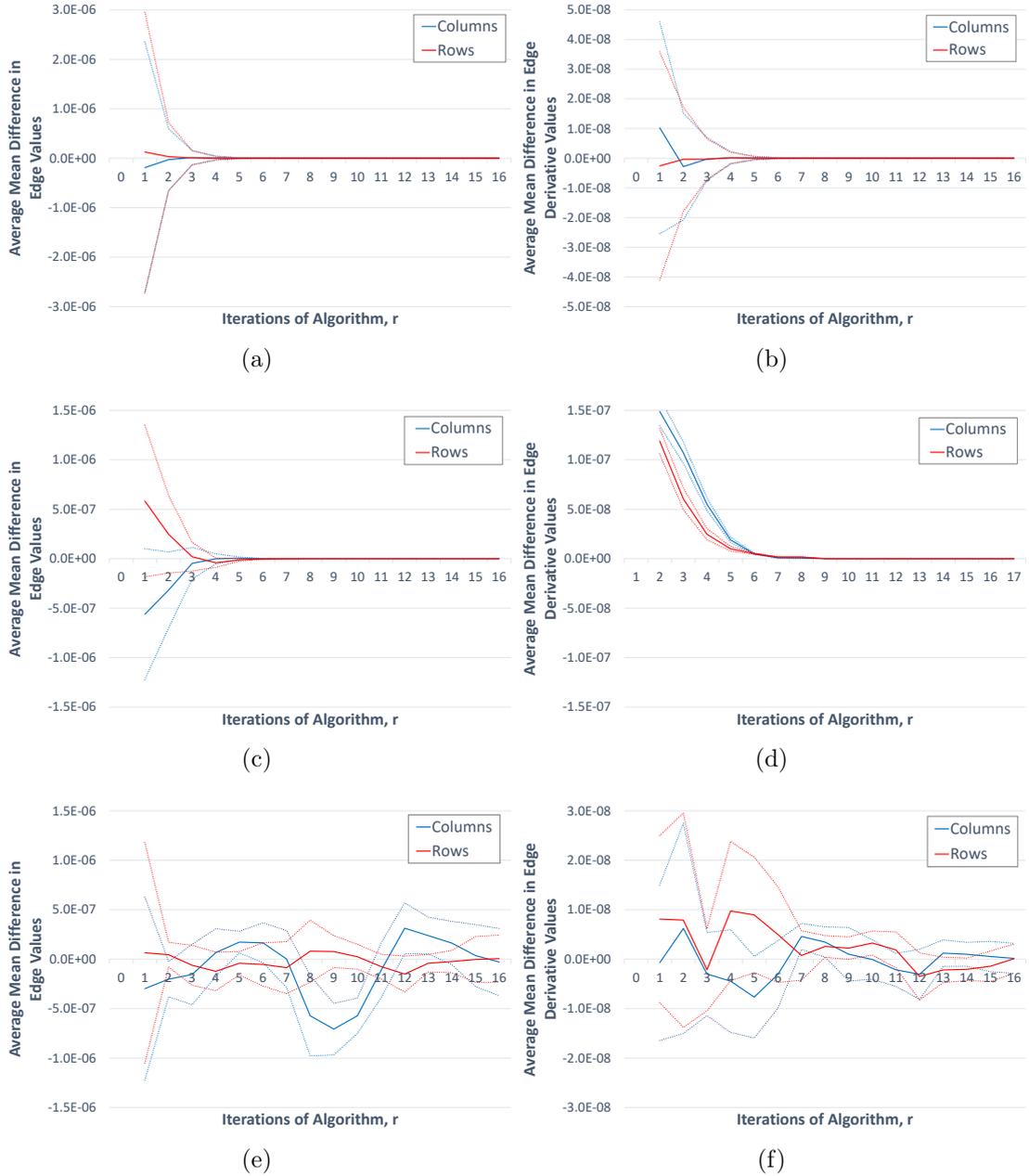


Figure 4.6: Average (solid) and 95% CI (dotted) for the (left) mean difference in edge values and (right) mean difference in edge derivatives, over 250 MC simulations, for (a,b) Cubic, (c,d) KDE Normal and (e,f) TPS, applied to Y_A . Metrics calculated row-wise (red) and column-wise (blue). Note that the figures contain different scales when compared by column.

The difference in edge values for the TPS spline, shown in Figure 4.6(e), appears to vary according to surface structure. When considering the discrete surface Y_A

row-wise, one of the modes shown in Figure 4.3 sits on the horizontal boundary of the planar surface, resulting in limited variation in the edge value differences around zero (shown in red in Figure 4.6(e)). Consider Y_A column-wise, and the transitions between both modes shown in Figure 4.3 sit along the vertical surface edges, resulting in periodic edge value differences (shown in blue in Figure 4.6(e)). The pattern shows a period of approximately 8, which matches that of the periodicity of the synthetic modes constructed across Y_A . This suggests that modal positioning in relation to the planar surface extracted from the torus, i.e. where the torus is ‘cut’ to be represented over the plane, impacts the toroidal accuracy of the surface approximation using the TPS spline.

4.1.4 Analysis of the TPS Spline with Algorithm 1

To investigate further the impacts of surface structure and other parameters on estimation accuracy, additional analyses were conducted applying the TPS spline in conjunction with Algorithm 1. These analyses were conducted using f_A and two additional synthetic toroidal surfaces defined by $f_k = \max(f_k^*, 0)$ with $k = B, C$, where

$$f_B^*(\phi, \psi) = \sin(\phi - \mu_B) \cos(\psi - \tau_B), \quad (4.6)$$

with $\mu_B = \tau_B = 45^\circ$ (Figure 4.7(a)). The sampled discrete surface Y_B was simulated according to Equation 2.8 where $\epsilon_B \sim N(0, \sigma_B^2)$ and $\sigma_B = 0.25$.

Secondly, a bivariate von Mises distribution was simulated using the Sine model [Singh et al., 2002] such that

$$f_C^*(\phi, \psi) = \exp \{ \cos(\phi - \mu_C) + \cos(\psi - \tau_C) + \sin(\phi - \mu_C) \sin(\psi - \tau_C) \}, \quad (4.7)$$

with $\mu_C = \tau_C = 180^\circ$ (Figure 4.7(b)), and Y_C was sampled according to Equation 2.8, where $\epsilon_C \sim N(0, \sigma_C^2)$ and $\sigma_C = 0.5$.

Surfaces were sampled over a discrete 16×16 grid (i.e. 22.5° bins) and estimated over a 360×360 grid (i.e. 1° resolution). To examine the effects of varying the smoothing parameter for the TPS spline estimation (p , Equation 4.1), the surface f_A was estimated assuming three separate values; $p = 0.001$, $p = 0.01$ and $p = 0.1$. The

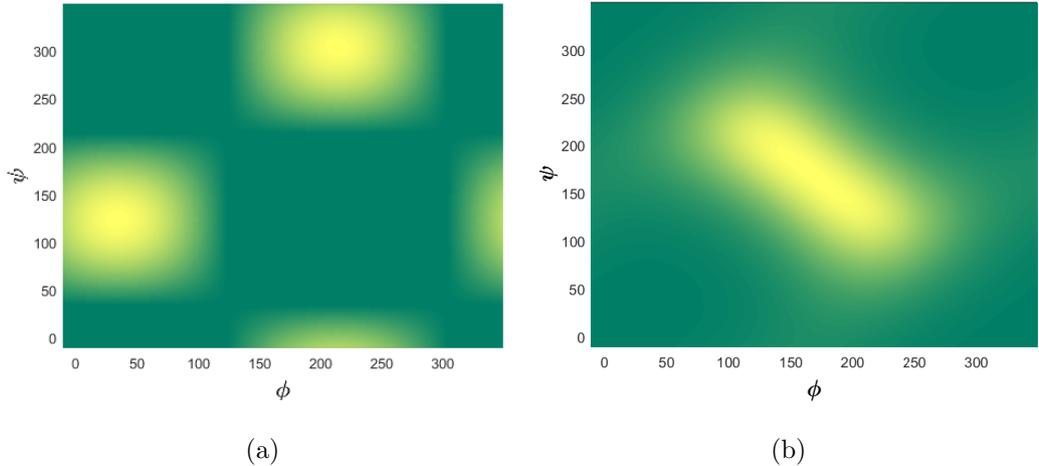


Figure 4.7: Continuous surfaces (a) f_B ($\mu_B = \tau_B = 45^\circ$), and (b) f_C ($\mu_C = \tau_C = 180^\circ$).

results for $p = 0.01$ have been discussed in the previous section and are considered the benchmark for the following analysis.

Finally, the impacts of changing grid sizes on the accuracy of the toroidal estimation were also considered. A number of $n \times n$ discrete grids, Y_A , with $n = 10, 16, 36, (72)$, were sampled, before evaluating the estimated smooth surface, f_A . Results for $n = 16$ are discussed in the previous section and considered the benchmark for this further analysis. In addition, for a 16×16 discrete grid, the estimated smooth surface was also evaluated on a 720×720 quasi-continuous grid (i.e. 0.5° resolution).

4.1.4.1 Shifted Distribution Modes

Figure 4.8 shows the values of the difference metrics when the TPS spline was applied, in conjunction with Algorithm 1, to the sample Y_B . It is clear that the differences in edge values are impacted by the location of the mode on the surface. In Figure 4.8(a), both row and column differences show periodicity that matches that of the modes in Figure 4.7(a). The periodicity is however less pronounced than shown in Figure 4.6(e) since the modes cross both the horizontal and vertical boundaries of the plane.

As shown for Y_A in Figure 4.6(f), the difference in derivatives (Figure 4.8(b)) reduces to near zero after around 6 iterations. The decreasing trend of the mean squared error with increasing iterations of Algorithm 1 is also shown under the shifted surface structure, Y_B , as it was for Y_A in Figure 4.5.

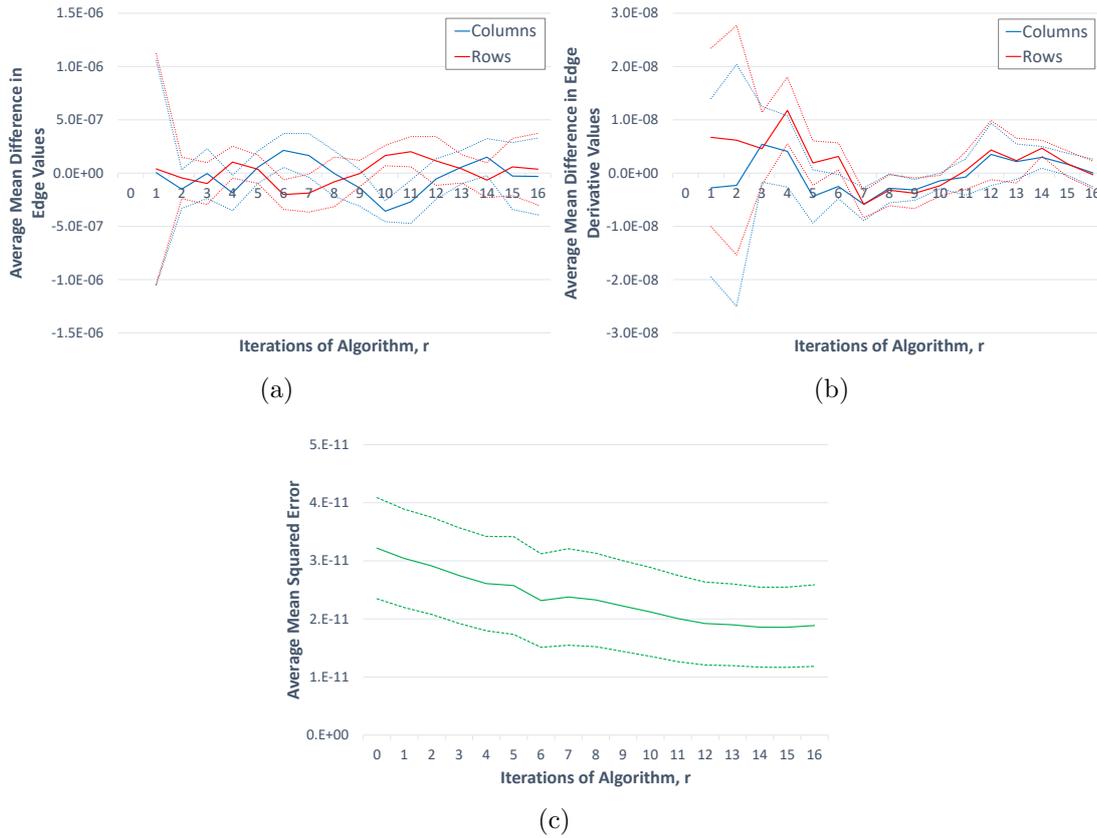


Figure 4.8: Average (solid) and 95% CI (dotted) for the (a) mean difference in edges values, (b) mean difference in edge derivatives and (c) MSE values, over 250 MC simulations, for the application of the TPS spline to Y_B , with $\mu_B = \tau_B = 45^\circ$. Metrics calculated row-wise (red) and column-wise (blue).

4.1.4.2 Bivariate von Mises Surface

For the bivariate von Mises surface, Y_C , the difference metrics along the rows and columns (Figure 4.9(a) and (b)) show much less structure in their pattern than those shown previously for Y_A or Y_B . Since the elongated asymmetric mode shown in Figure 4.7(b) sits in the centre of the surface, it has similar effects on both the row and column difference metrics. The pattern of the differences is however harder to determine in relation to the modal structure. What is evident is that at approximately $r = 10$ the variation in the difference in edge and derivative values is affected, which may relate to some periodicity of the single mode.

Figure 4.8(c) shows much smaller MSE values than those shown for Y_A (Figure 4.5) and Y_B (Figure 4.8(c)). The decreasing trend in MSE remains consistent for this surface, but the variation in MSE values is also considerably smaller than that shown for Y_A and Y_B . This suggests that the TPS spline produces a far more

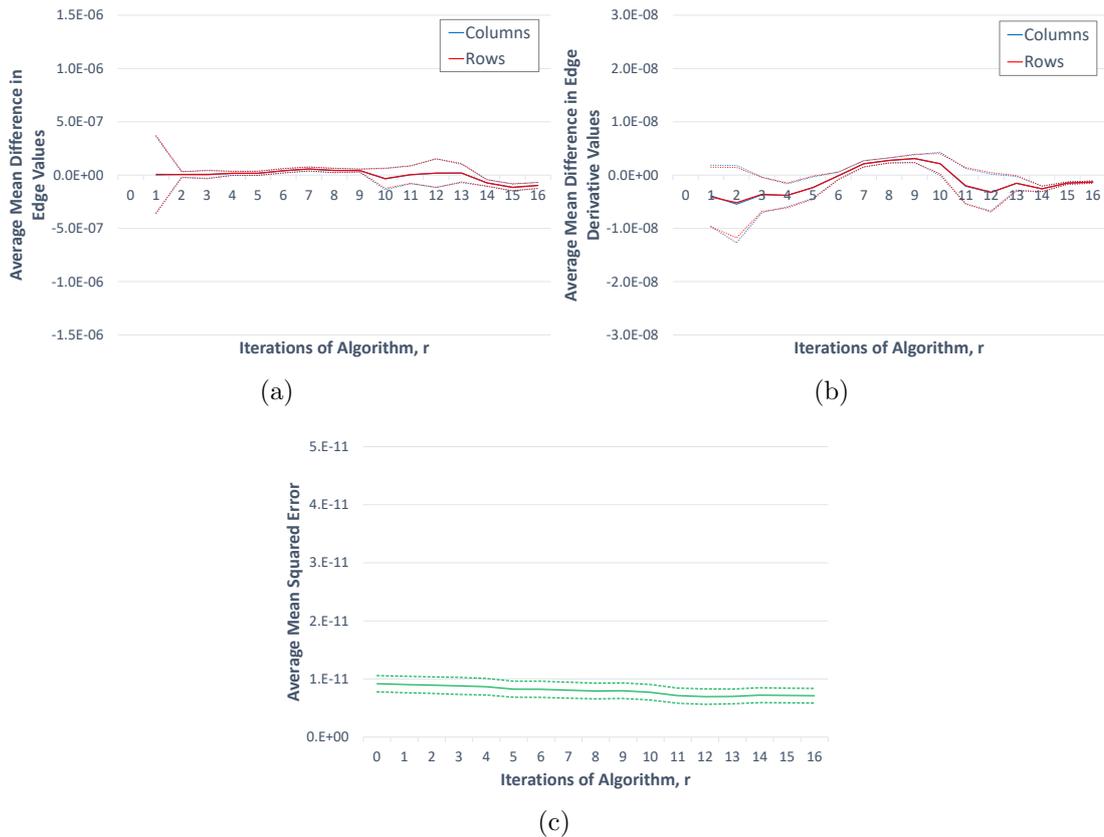


Figure 4.9: Average (solid) and 95% CI (dotted) for the (a) mean difference in edge values, (b) mean difference in edge derivatives, and (c) MSE values, over 250 MC simulations, for the application of the TPS spline to Y_C . Metrics calculated row-wise (red) and column-wise (blue).

accurate overall surface estimation, and estimation of the toroidal characteristics, of the bivariate von Mises surface than it does for the bimodal surface generated using a Sine and Cosine product.

4.1.4.3 Impacts of the Smoothing Parameter

The TPS spline was applied in conjunction with Algorithm 1 to Y_A with $\mu_A = \tau_A = 90^\circ$, with the smoothing parameter in the construction of the TPS spline varied from $p = 0.01$ to $p = 0.001$ and $p = 0.1$. These variations alter how closely the surface is fitted to the observed data points, and thus suggest that the data is expected to contain either more or less noise. By decreasing the parameter by a factor of 10, the estimated surface, \hat{f}_A , becomes increasingly smooth, with the smoothness penalty (second term) in Equation 4.1 given a greater weighting. In contrast, increasing the value of the parameter produces a less smooth estimated surface, with fidelity to the data (first term) given a larger weighting according to

Equation 4.1. Figures 4.10 and 4.11 shows the effects of respectively decreasing and increasing the smoothing parameter on the estimation performance.

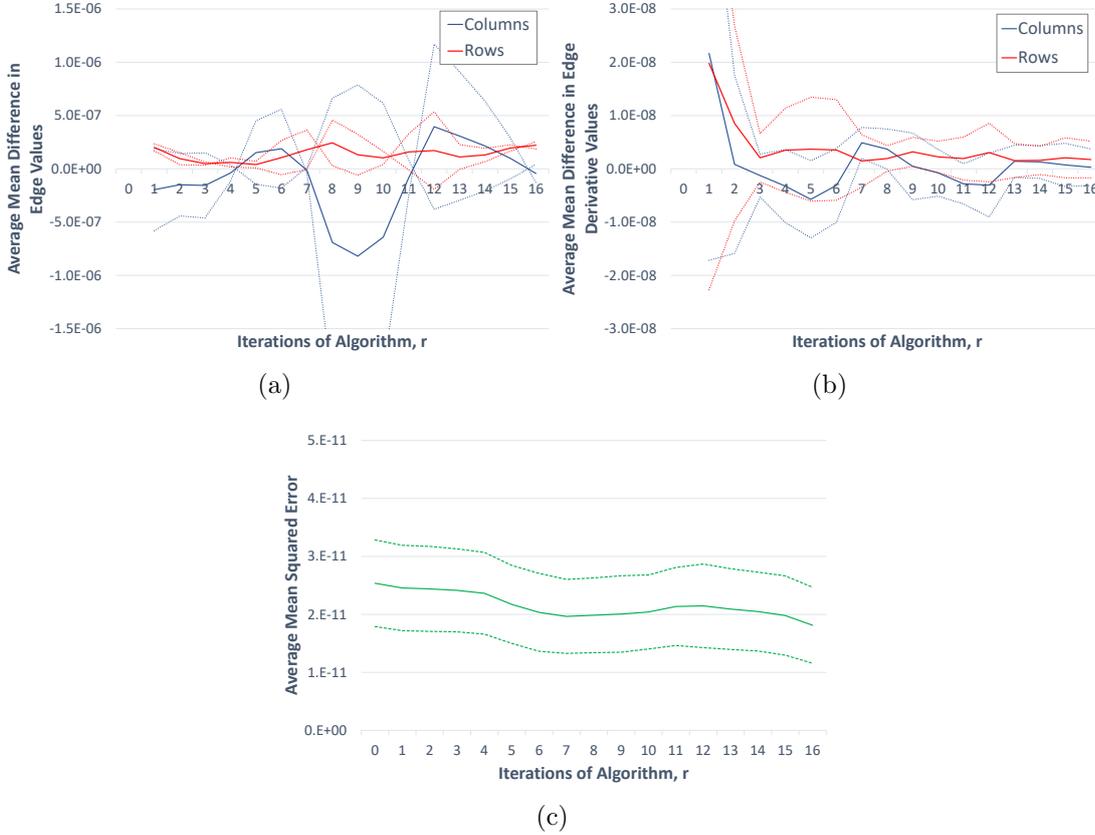


Figure 4.10: Average (solid) and 95% CI (dotted) for the (a) mean difference in edge values, (b) mean difference in edge derivatives and (c) MSE values, over 250 MC simulations, for the application of the TPS spline, with $p = 0.001$, to Y_A . Metrics calculated row-wise (red) and column-wise (blue).

With a decreased smoothing parameter, thus lesser fidelity to the noisy observed data, the variation in difference values shown in Figure 4.10(a) is considerably greater than that shown in Figures 4.9(e) and 4.11(a) for smoothing parameters $p = 0.01$ and $p = 0.1$, respectively. However, the periodicity of the column differences (blue lines) remains evident as the smoothing parameter is both decreased and increased. The difference in row edge values (red lines) also remains relatively consistent across all three smoothing parameters. Additionally, the difference in derivatives (Figures 4.10(b) and 4.11(b)) continues to reduce to approximately zero after 6-8 iterations of Algorithm 1, as seen for the benchmark smoothing parameter, $p = 0.01$, in Figure 4.9(f).

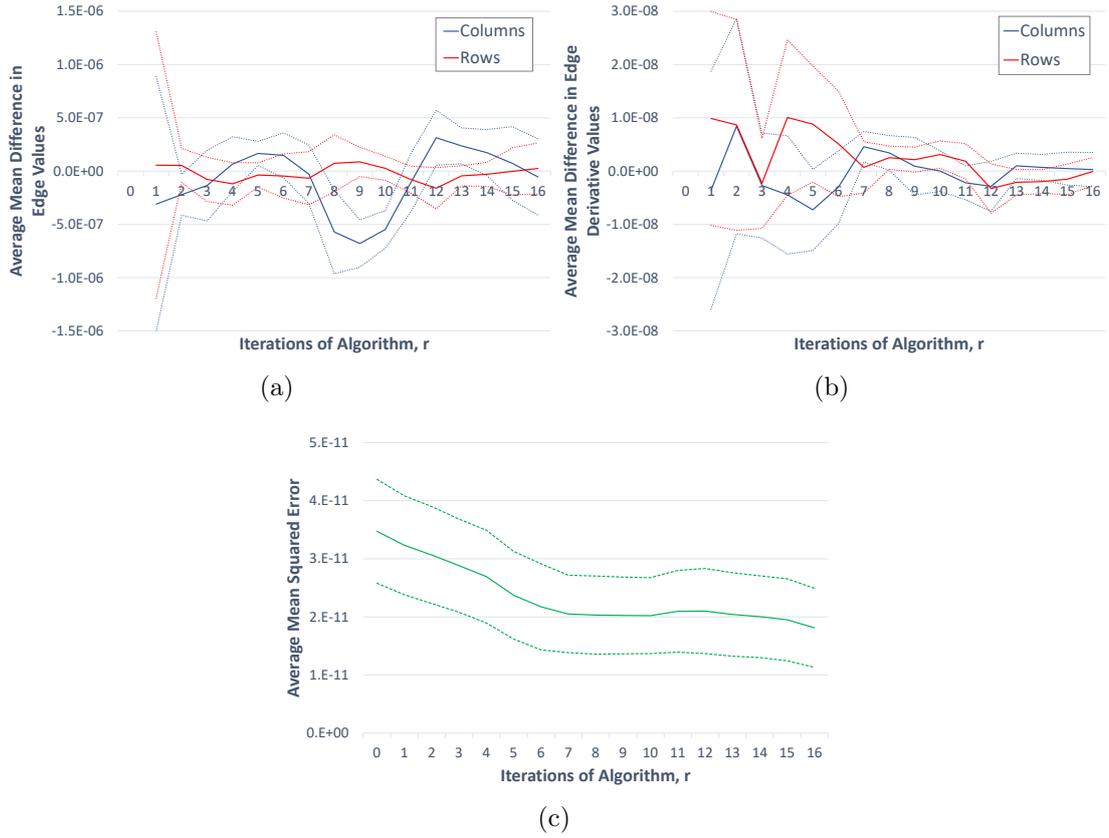


Figure 4.11: Average (solid) and 95% CI (dotted) for the (a) mean difference in edge values, (b) mean difference in edge derivatives and (c) MSE values, over 250 MC simulations, for the application of the TPS spline, with $p = 0.1$, to Y_A . Metrics calculated row-wise (red) and column-wise (blue).

The mean squared errors in estimated surface values (Figures 4.10(c) and 4.11(c)) again show a decreasing trend as the iterations increase, suggesting that revealing more information regarding the periodicity of the data leads to an overall improvement in surface estimation accuracy. This magnitude of the MSE and the decreasing trend remains consistent as the smoothing parameter used in the TPS spline is both decreased and increased. This suggests that the smoothing parameter has little impact on the overall accuracy of the surface estimation, but it can impact the estimation of toroidal characteristics.

4.1.4.4 Differing Discrete and Continuous Grid Sizes

Figure 4.12 and 4.13 illustrate the performance of Algorithm 1 and the TPS spline at estimating f_A (with $\mu_A = \tau_A = 90^\circ$) on a 360×360 quasi-continuous grid from differing discrete observed grid sizes; $n = 10, 36$. Given the increases in computational time with the increases in discrete data up to a 36×36 grid, results from a

single example simulation are presented. Simulations using a 72×72 discrete grid are not presented here but show similar results.

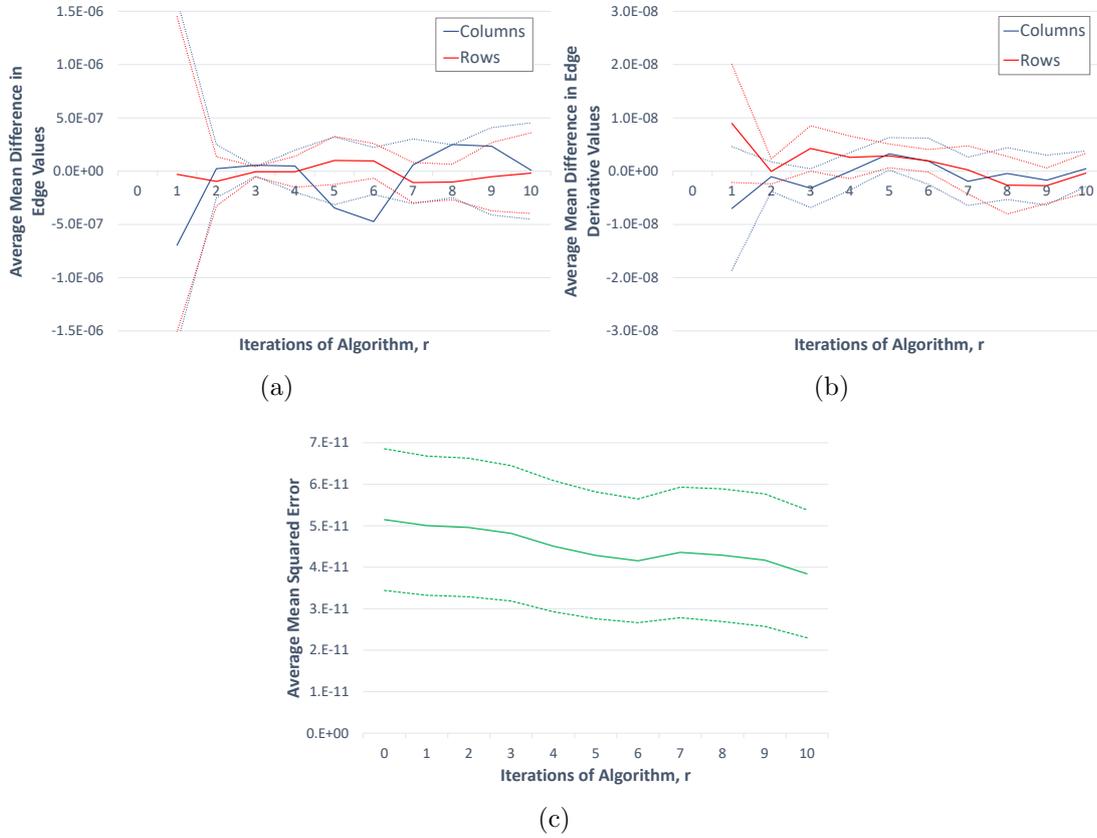


Figure 4.12: Average (solid) and 95% CI (dotted) for the (a) mean difference in edge values, (b) mean difference in edge derivatives and (c) MSE values, over 250 Monte Carlo simulations, for the application of the TPS spline to Y_A , with 10×10 discrete grid and 360×360 quasi-continuous grid. Metrics calculated row-wise (red) and column-wise (blue). Note the change of scale in (c).

The difference in row values remain approximately zero for both discrete grid sizes (red lines in Figures 4.12(a) and 4.13(a)), as shown for the benchmark 16×16 discrete grid in Figure 4.6(e). In contrast, the difference in column edge values (blue lines) appear to exhibit scaled periodicity with the revelation of repeated modes. For example, while a trough in column difference occurs at approximately 9 iterations in Figure 4.6(e), a similar trough occurs at 6 iterations for a 10×10 grid, and 20 iterations for a 36×36 grid. These each relate to when between approximately 55% and 60% of the surface has been repeated.

Equally, the differences in edges derivatives show variation up to approximately $r = 4$ in Figure 4.12(b) and $r = 14$ in Figure 4.13(e), then reducing to around zero with the further iterations. For the benchmark 16×16 discrete grid, this threshold

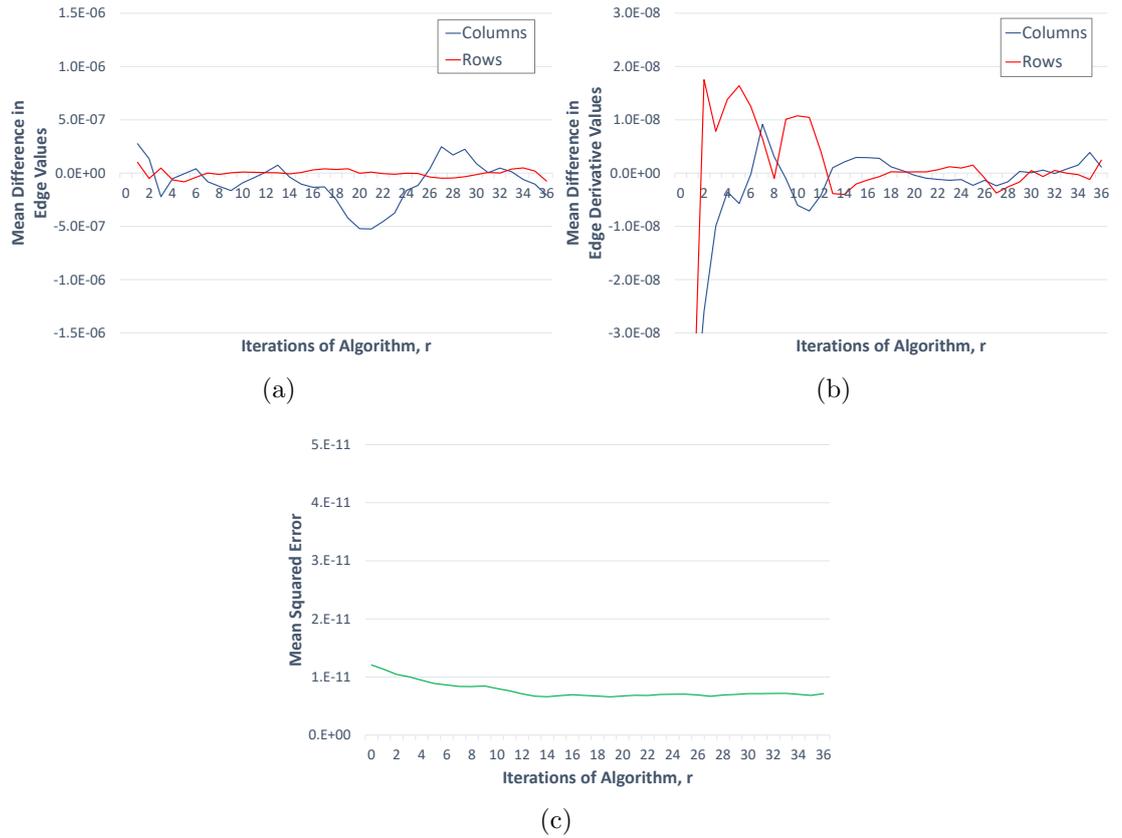


Figure 4.13: (a) Mean difference in edge values, (b) mean difference in edge derivatives, and (c) MSE values, for a single simulation, for the application of the TPS spline to Y_A , with 36×36 discrete grid and 360×360 quasi-continuous grid. Metrics calculated row-wise (red) and column-wise (blue).

was approximately $r = 7$. Each of these thresholds relate to when approximately 40% of the surface has been repeated, thus again showing the scaling in toroidal estimation behaviours for the varying discrete grid sizes.

Figure 4.14 shows the estimation from a 16×16 discrete grid to a 720×720 quasi-continuous grid. Both differences in edge values and derivatives are considerably smaller than those shown for the 360×360 quasi-continuous grid for any discrete grid size. This suggests that with evaluation of the continuous estimated surface over a finer scale grid, the toroidal characteristics of the surface can be more accurately captured. However, a trade-off is made with regards to the computational time taken to evaluate the estimation at this finer resolution.

Figures 4.12(c), 4.13(c) and Figure 4.14(c) show that the mean squared error in estimated surface values reduces as the iterations increase, across all input and output grid sizes tested. This increase in accuracy is very similar to that shown for

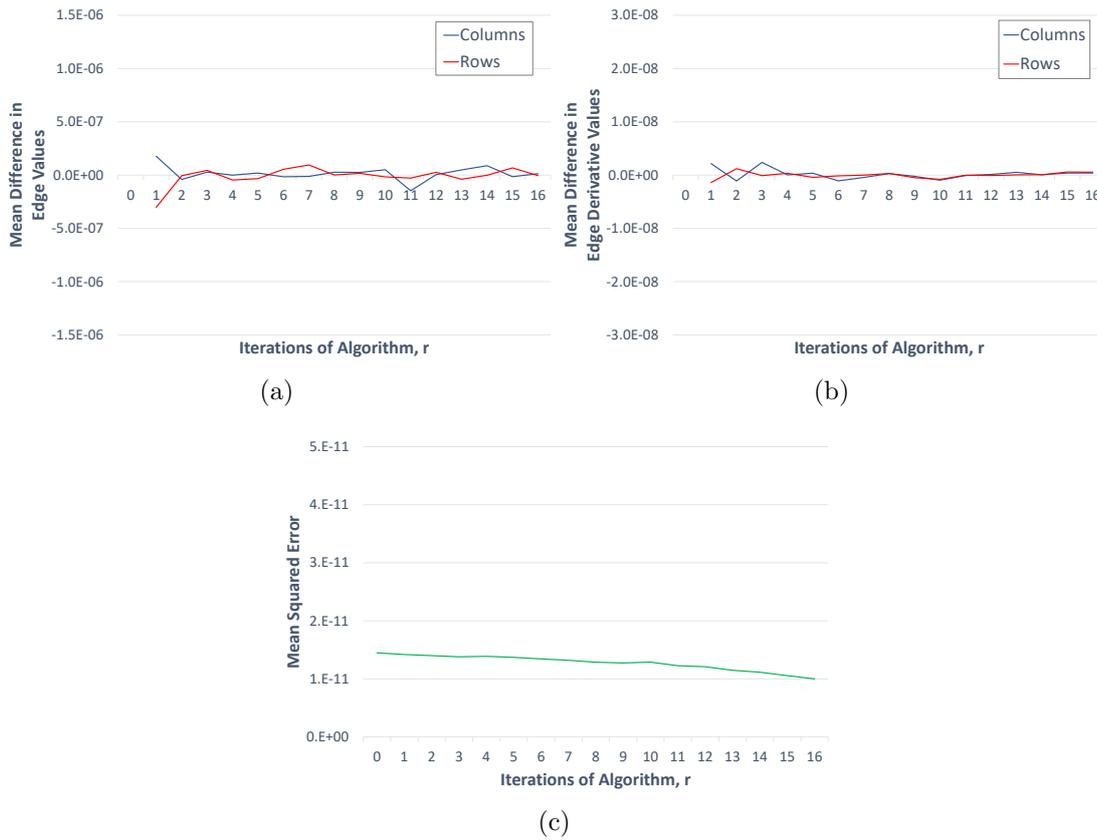


Figure 4.14: (a) Mean difference in edge values, (b) mean difference in edge derivatives, and (c) MSE values, for a single simulation, for the application of the TPS spline to Y_A , with 16×16 discrete grid and 720×720 quasi-continuous grid. Metrics calculated row-wise (red) and column-wise (blue).

the benchmark case in Figure 4.5. The magnitude of the MSE is, however, clearly decreased with increased discrete and quasi-continuous grid sizes; suggesting that the more detailed information provided by using finer input and output resolutions provides better approximations to the underlying surface. However, as suggested above, there are important trade-offs to consider with relation to the practical computation time of surfaces at these finer resolutions.

4.1.5 Simulation Study Discussion

In general, the algorithmic adaptation of planar estimation techniques shows as good, if not better, surface estimation results when compared to the intrinsically toroidal estimation methods. However, unlike the Cubic spline or KDE Normal method, the TPS spline varies in its ability to capture the toroidal nature of the surface at the imposed edges.

More specifically, the overall accuracy of the planar Cubic spline reduces to that of the Cyclic Cubic spline after allowing for edge effects, and then remains constant with increasing iterations of Algorithm 1. As expected, the magnitude of this constant MSE remains within that of the errors constructed over the simulated surface. The difference in edge values and derivatives shows similar behaviour, reducing to zero after around 4 iterations. The times taken for both Cubic splines to run are also negligible compared to the remaining methods, altogether suggesting that this method is most preferable for surface estimation. However, the metrics presented here assess data fidelity without evaluating surface smoothness. In many applications, including wind modelling, observed data are noisy and so an exact interpolation method is far from appropriate in estimating the true underlying process; a smoothing method that accounts for noise is likely more appropriate.

The trade-off in smoothness that comes with the use of smoothing splines or kernel density estimation allows the surface approximation to account for noise in observed data and avoid over-fitting. However, a trade-off in terms of computation time needs to be addressed. While all Cubic run times were negligible, if say 6-8 iterations of Algorithm 1 were applied with the TPS spline to gain an accurate approximation of the toroidal surface, each estimation would take approximately 4-6 seconds. Even without application of the algorithm, the KDE Normal estimation took around 11 seconds while the KDE von Mises approximation took around 16 seconds. Although operational constraints are commonly focussed on computation time and resources, it is important to consider the benefits of each estimation method in relation to a particular application.

While the KDE von Mises intrinsically accounts for toroidality of the surface, the overall estimation accuracy showed the highest MSE value across all methods. Alongside a very slow computation time, it would appear that this technique is yet to provide an appropriate estimation method for practice. The KDE Normal used in conjunction with Algorithm 1 was intended to account for circularity in the dataset, while producing slightly faster computation times. The overall accuracy of the KDE Normal in conjunction with Algorithm 1 showed more promising results, at least within the order of those given by the other methods analysed. The KDE

Normal with Algorithm 1 was also highly accurate in capturing the toroidal characteristics of the surface after only a few iterations, but with computation times up to around 30 seconds per approximation in order to achieve this.

The overall surface accuracy of the TPS spline in conjunction with Algorithm 1 also improved as iterations increased. As with the previous planar estimation techniques, as the algorithm revealed more data, the TPS spline incorporated more information about the shape and toroidality of the surface, thus providing an improved fit. However, the differences in edge values appear to vary in accordance with surface structure, with rows and columns affected differently depending on the position and shape of modes across the surface. In application of such a spline, this must be considered in the pre-processing of toroidal data to form a planar discrete surface. Understanding the periodicity of the discrete dataset also allows for optimal choice of the number of iterations of Algorithm 1 required for accurate surface estimation.

Although changing the smoothing parameter did not impact the overall accuracy of the surface estimation or the general periodicity of the difference metrics, it did impact the variability of the difference in edge values. In particular, when decreasing the smoothing parameter, i.e. estimating a smoother surface, the variation around the difference in edge values increased. In practice, this suggests that the periodicity in the difference metric is more variable and so harder to rely on for the optimisation of the number of iterations of Algorithm 1 required for accurate surface estimation.

Across all grid sizes and smoothing parameter values, the pattern of differences in edge values aligned with the periodicity of the modes in each dimension, suggesting that the unveiling of new modes introduces some instability into the toroidal approximation. However, the difference in edge derivatives tended to zero after a scaled number of iterations, equivalent to revealing approximately 40% of the surface, across all simulations using Y_A . With finer resolution input and output grids, overall accuracy of the surface estimation was considerably improved, but the trade-off in computational demand results in high estimation times which may be impractical for operational use.

The alignment between the surface structure and the periodicity of the difference metrics was reaffirmed when the modal positions of the simulated dataset were

shifted in Y_B . This modal shift resulted in a relative shift in the periodic pattern of the differences in edge values. In addition, the change in distribution structure given by the bivariate von Mises distribution also resulted in a related change in the pattern of the difference metric. However, with an elongated diagonal mode, the periodicity of the metric in each dimension became less obvious.

4.2 Estimated Directional Wind Response

Based on the previous discussion, therefore due to its accuracy, relative speed and simple conceptual underpinning, the thin plate smoothing spline in conjunction with Algorithm 1 is used to estimate continuous directional wind response. The ability to identify clear modes and their related periodicity, within the wind direction data also allows optimisation of the number of iterations of Algorithm 1 that are required for accurate toroidal surface estimation.

Directional wind response is constructed using observed bivariate wind direction distributions as described at the beginning of this chapter, and is estimated across each of the case studies detailed in Chapter 3. The full sets of observed discrete bivariate distributions for each site across Flea Creek Valley in 2007 and 2014, and the National Arboretum Canberra in 2015 are given in Appendix B. For a map of the sites analysed in this section see Figures 3.4 and 3.7 in Chapter 3.

4.2.1 *Optimising Algorithm 1*

Using the distributions observed across Flea Creek Valley in 2014 shown at the beginning of this chapter (Figure 4.1), the suitability of the TPS spline estimation method for application to directional wind response can be assessed. It is also possible to empirically select the optimal iterations of Algorithm 1 for accurate toroidal estimation. Figure 4.15 shows the difference in edge values and derivatives for each site when the TPS spline is applied to these discrete wind direction data.

The differences in first derivatives (Figure 4.15(b,d,f)) reduce to around zero after 6 iterations of the algorithm. In contrast, but as expected from the simulation study, the differences in edge values (Figure 4.15(a,c,e)) show paths that relate to the surface structure of the three datasets. In Figure 4.1 it can be seen that, after

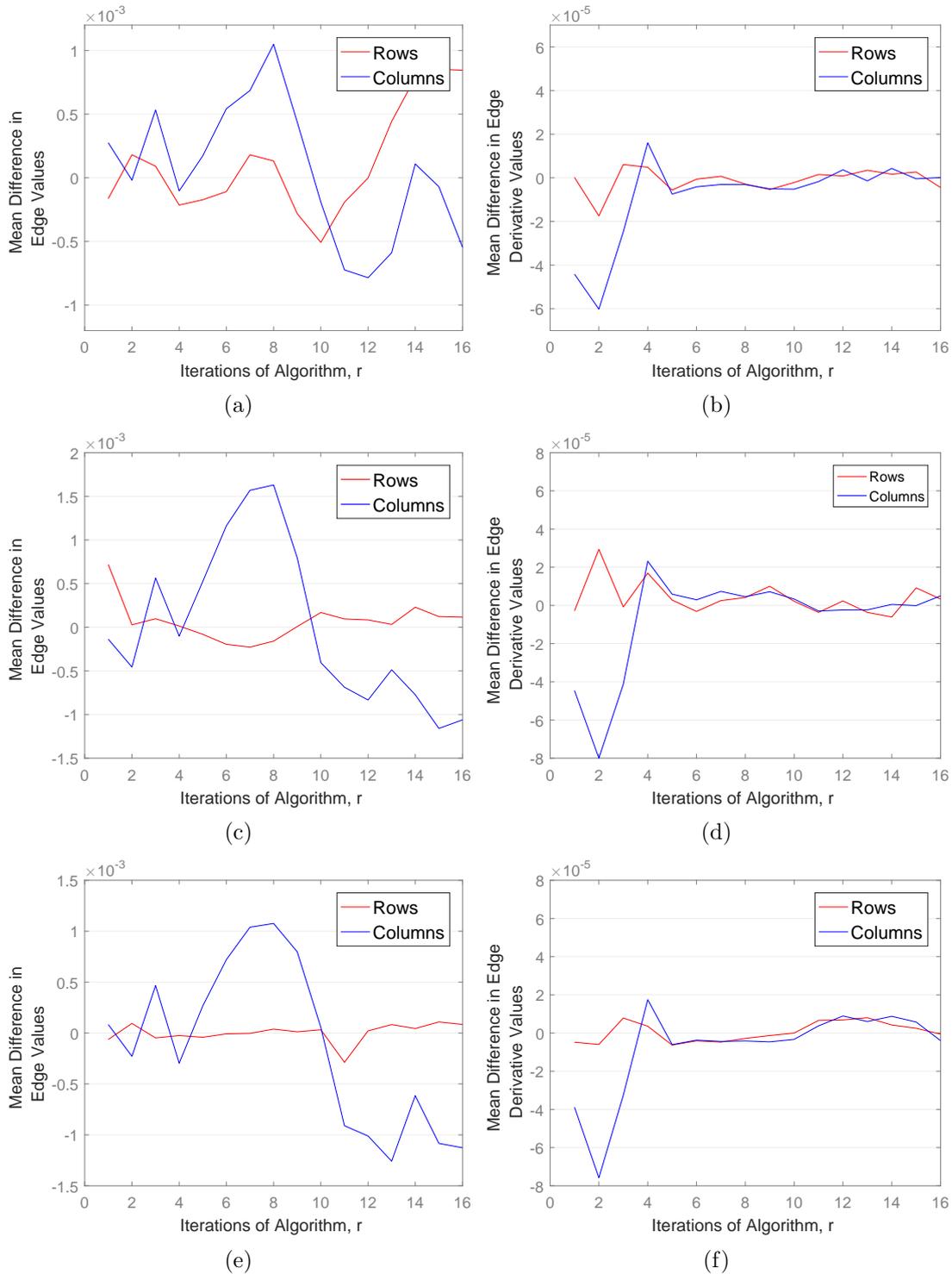


Figure 4.15: Mean difference in (top) edge values and (bottom) edge derivatives of the estimated directional wind response distributions across FCV in 2014, using the TPS spline, on the (a,b) east-facing slope at B4, (c,d) valley floor at B6 and (e,f) west-facing slope at B10. Note that the figures contain different scales when compared by column.

approximately 6 iterations of Algorithm 1 applied to all three surfaces, each of the modes will have been entirely repeated column-wise. Then after 10 iterations the modes would begin to appear again. These patterns relate directly to the column-wise differences in edge values (blue lines) which remain close to zero up to about 4-6 iterations, before indicating some periodicity which peaks around 8 iterations and troughs towards 16 iterations.

Considering the discrete surfaces row-wise, Figures 4.1(b) and (c) exhibit a dominant elongated mode. These modes have a more constant impact on the approximation, and result in differences in edge values close to zero (Figures 4.15(b) and (c), red lines). In contrast, on the east-facing slope at B4 (Figure 4.1(a)) there are distinct modes in the row-wise direction which would account for the periodicity shown in the row differences in Figure 4.15(a).

In the following section, the TPS spline in conjunction with 6 iterations of Algorithm 1, was used to estimate the true continuous directional wind response across both case studies described in Chapter 3. Figures 4.16, 4.17, 4.18 and 4.19 show that after the application of the TPS spline with 6 iterations of Algorithm 1, the contours align well across the planar boundaries, suggesting a good match in edge values. Additionally, the surface estimation shows reasonable smoothness by the TPS spline for practical interpretation of wind direction behaviour, and the average computation time to estimate each surface was approximately 3.6 seconds.

4.2.2 *Prevailing Wind Direction Analysis*

In order to construct directional wind response distributions across both case studies, concurrent prevailing and surface (or within terrain) wind observations were required. The observed ridge top wind data described in Chapter 3 cannot be considered directly equivalent to geostrophic wind flow above the boundary layer across each region. However, through comparison with data collected at nearby automatic weather stations by the Australian Bureau of Meteorology (BoM), it was possible to determine that the sites were indicative of the prevailing winds in the region.

Figure 4.16 shows the joint wind direction distributions between the local BoM data and ridge top stations at FCV in 2007 and 2014, and at the NAC in 2015. Strong

modes along the diagonal indicate a strong alignment between wind directions experienced at the ridge tops and those recorded by BoM.

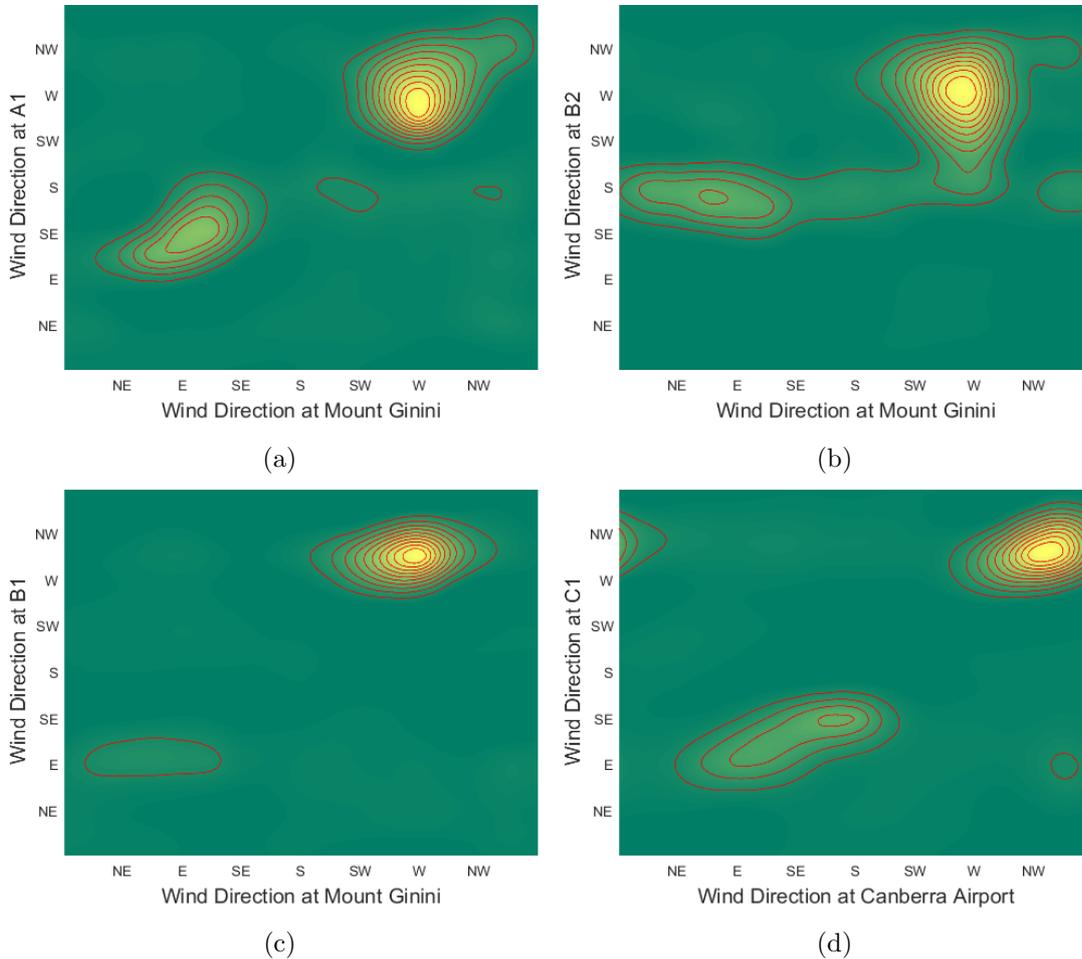


Figure 4.16: Comparison between wind direction observed at Mount Ginini and FCV at (a) A1 in 2007, (b) B2 in 2014 and (c) B1 in 2014, as well as between Canberra Airport and NAC at (d) C1 in 2015. Red lines indicate isolines on the estimated surfaces.

Through analysis of the joint distributions between BoM wind direction data observed at Mount Ginini (Figure 3.4(b), Station No. #070349) and data observed at FCV, Sharples et al. [2010] showed that A1 was indicative of the prevailing winds in the region in 2007 (Figure 4.16(a)). However, in 2014, analysis of the wind response at the same site (B2; Figure 4.16(b)) showed that the ridge top wind directions were no longer a good representation of the prevailing wind direction. The ridge top station in 2014 at FCV was therefore taken to be B1 (Figure 4.16(c)).

BoM wind direction data observed at Canberra Airport (Station No. #070351) were compared to those observed at the ridge top at the NAC in 2015. Figure 4.16(d) shows that the wind directions observed at the ridge top over the NAC aligned with

the wind directions observed at Canberra Airport, so C1 was taken to indicate the prevailing winds for this case study.

4.2.3 Flea Creek Valley, 2007

Figure 4.17 shows the estimated directional wind response distributions for the four stations across FCV from west to east (A2 to A5, Figure 3.4). On the western wall, at A2, the weather station is most often on the leeward slope to prevailing west-northwesterly winds across the valley, and the off-diagonal mode suggests that westerly prevailing winds are experienced as easterlies at the site (Figure 4.17(a)).

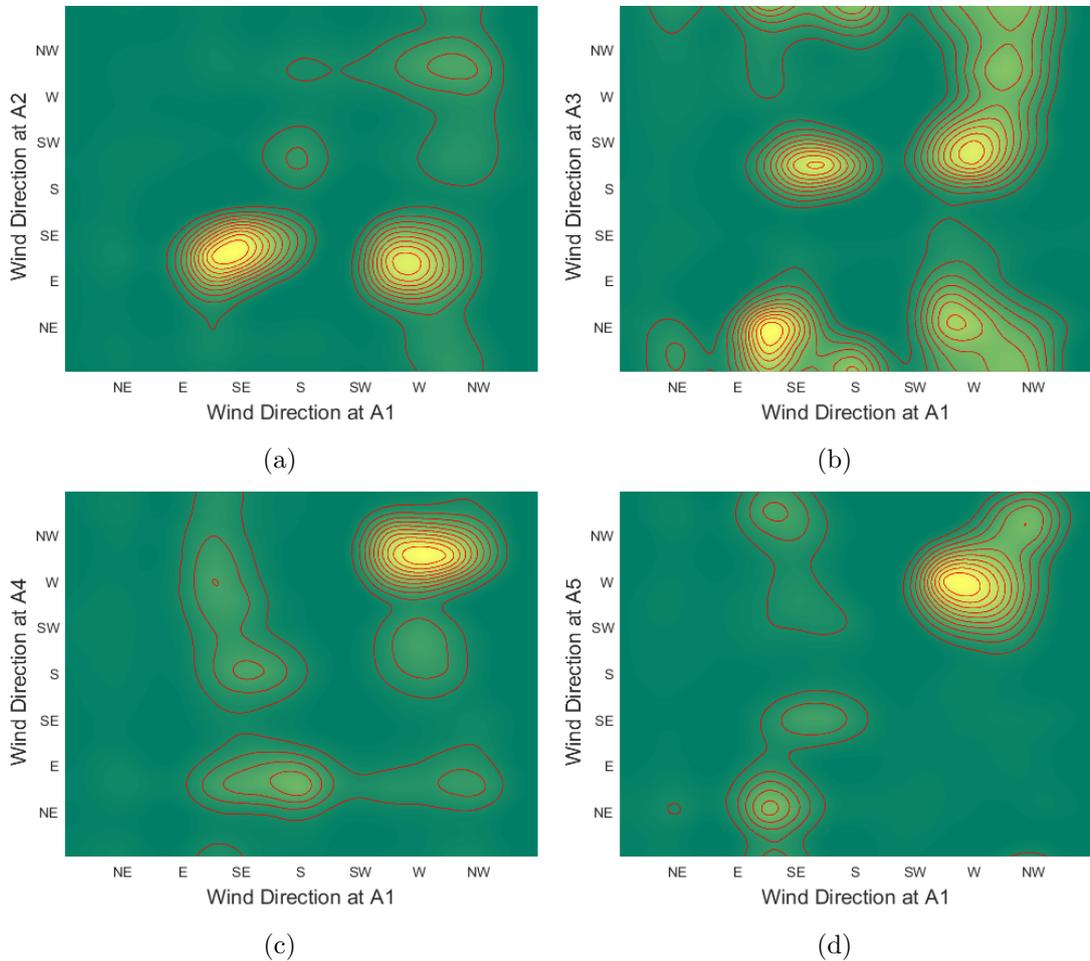


Figure 4.17: Estimated directional wind response distributions from FCV in 2007, at (a) A2, (b) A3, (c) A4 and (d) A5. Red lines indicate isolines on the estimated surfaces.

In contrast, on the eastern wall at A4 and A5, the dominant modes of the distribution lie along the diagonal and so wind directions observed at the surface here align with observed ridge top wind directions (Figures 4.17(c) and (d), respectively). On the valley floor, at A3, four modes are evident at low wind speeds (Figures 4.17(b)),

suggesting that winds at the site are either north-easterly or south-westerly, whether prevailing winds are easterly or northerly. This suggests that channelling through the valley may be the key driving force of winds experienced at this site. With increasing wind speeds, all of these modal patterns become more distinct which suggest mechanical processes rather than thermal forcing.

4.2.4 Flea Creek Valley, 2014

The estimated smooth directional wind response distributions for stations B4, B6, B8 and B10 are shown in Figure 4.18. These stations form a representation of winds across the valley transect from east to west.

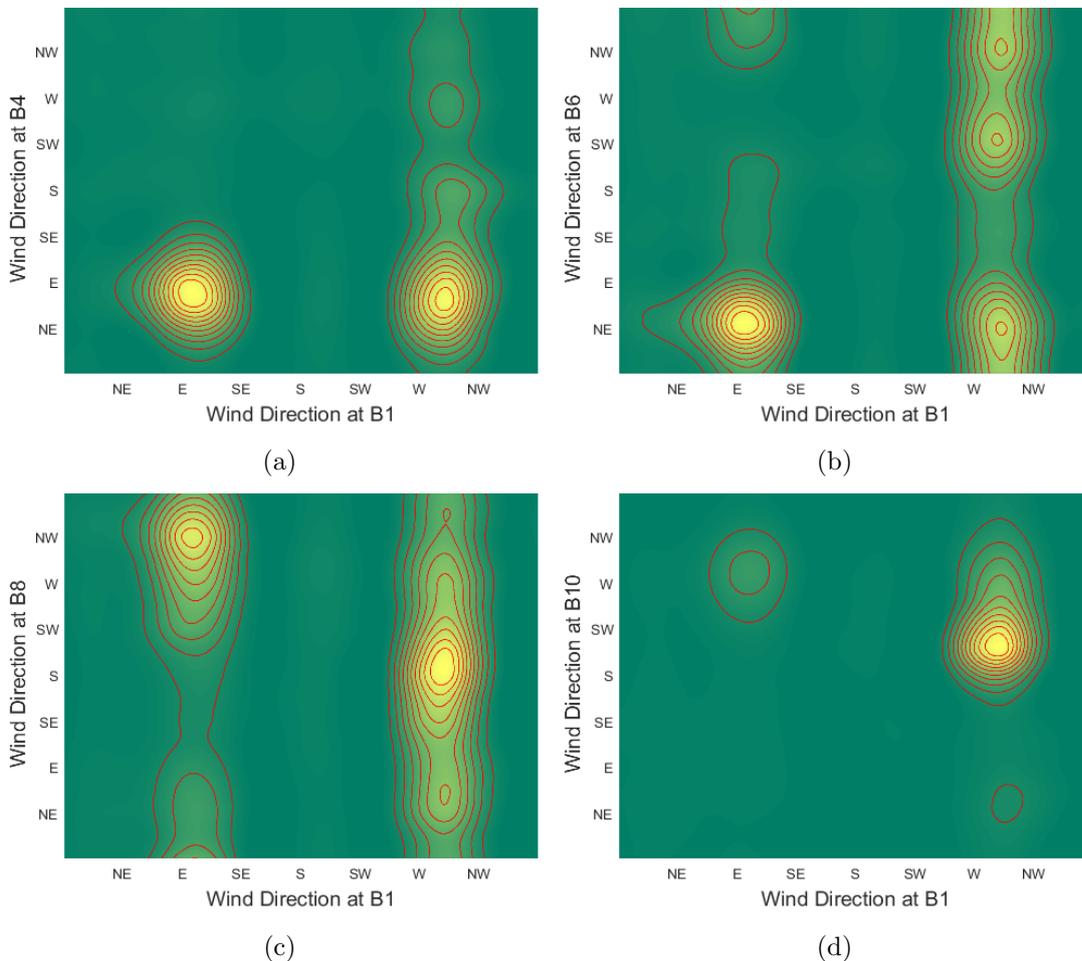


Figure 4.18: Estimated directional wind response distributions from FCV in 2014, at (a) B4, (b) B6, (c) B8 and (d) B10. Red lines indicate isolines on the estimated surfaces.

As described at the beginning of this chapter, the off-diagonal mode present at B4 indicates a region of wind reversal when the site is leeward to the prevailing westerly winds (Figure 4.18(a)). A reflected off-diagonal mode is present at B8

and B10 (Figures 4.18(c) and (d), respectively), again indicating regions of wind reversals when these sites are leeward to the easterly prevailing winds. As previously discussed, the variation shown at B6 and B8 (Figures 4.18(b) and (c), respectively) is likely to result from local topographical or vegetation features, but there is also evidence of channelling through the north-south aligned valley at both these sites.

4.2.5 *National Arboretum Canberra, 2015*

The estimated smooth directional wind response distributions for the stations C2, C5, C7 and C9 at the NAC are shown in Figure 4.19. These directional wind response distributions show considerably less variation than those observed at Flea Creek Valley. This is likely a result of the uniform vegetation type and undulating topography at the NAC which ensures more consistent wind direction behaviour.

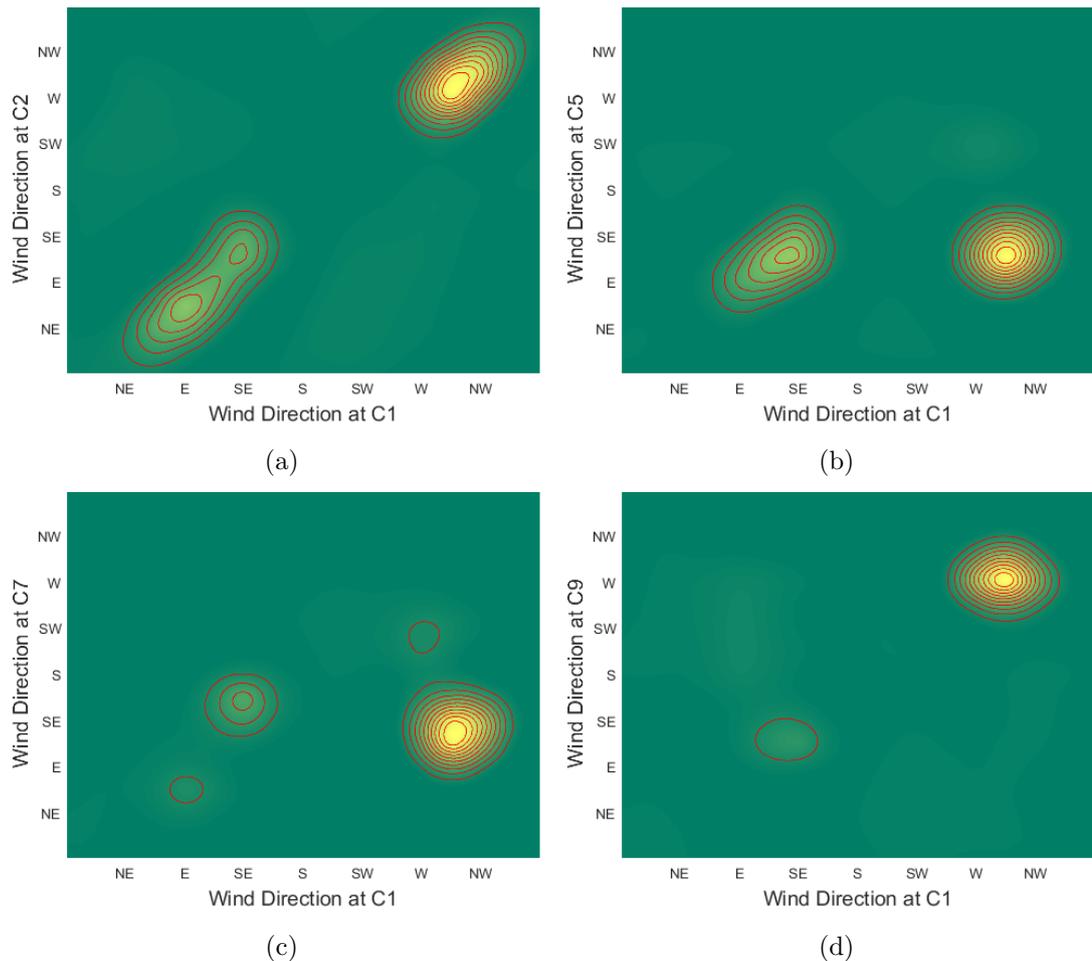


Figure 4.19: Estimated directional wind response distributions from the NAC in 2015, at (a) C2, (b) C5, (c) C7 and (d) C9. Red lines indicate isolines on the estimated surfaces.

At C2, on the cleared east-facing slope of the study area, there are two clear modes along the diagonal which indicate alignment of the observed wind directions with the westerly and easterly prevailing winds (Figure 4.19(a)). Within the pine stand, the westerly prevailing winds appear to produce a wind reversal on the leeward (C5) and 45° leeward (C7) slope, indicated by the off-diagonal modes in Figures 4.19(b) and (c). On the cross slope at C9, on the end of the spur within the pine stand, the mode relating to the westerly prevailing winds moves back to the diagonal (Figure 4.19(d)). This suggests that the wind directions on the cross-slope align with the prevailing winds.

4.3 Concluding Remarks

This chapter has proposed a probabilistic representation of wind direction in terms of directional wind response. In using observed discrete joint wind direction distributions, the true continuous wind response process requires estimation. To this end, the analysis in this chapter compares a number of toroidal surface estimation techniques.

Smoothing and kernel density estimation methods were found to be more computationally expensive than exact interpolation, but the techniques are designed to avoid over-fitting of noise within an observed dataset; an issue very relevant when dealing with wind data. A trade-off between resource constraints and required surface smoothness is therefore unavoidable in the application of such estimation techniques. Introduced in Section 4.1, Algorithm 1 was not only able to improve planar interpolation and smoothing techniques for toroidal data, but also provided better surface estimation results than toroidal kernel density estimation. In practice therefore, it is reassuring to note that the simple intuitive strategy analysed here achieves its goals of capturing bivariate circularity. However, the choice of planar estimation technique, as well as other variables such as grid size and smoothing parameter, remains an important consideration when conducting surface estimation.

The ability of the thin plate smoothing spline, applied in conjunction with Algorithm 1, to capture the toroidal characteristics of datasets at the identified edges was dependent on the structure of the surface. The relationship shown between

toroidal surface estimation accuracy and the surface structure for the TPS spline could be exploited to select an optimal number of iterations required for accurate surface estimation. In practice, this link could be utilised to pre-process toroidal data and ensure modes are positioned appropriately on the planar surface, e.g. without modes on the edge of the planar data view.

In application to directional wind response, the toroidal nature of the surfaces was optimally estimated after approximately 6 iterations of Algorithm 1. The examples showed that the toroidal accuracy of the application of the thin plate smoothing spline, after a limited number of iterations, has the potential to improve computation time for subsequent analyses of wind direction distributions. It was then possible to interpret the resulting accurate and realistic estimates of true wind direction response across complex terrain in relation to broad-scale mountain meteorology as well as fine-scale wind behaviours in amongst the landscape.

The following chapter asks the question of whether or not changes in surface characteristics, such as varying topography or vegetation, have a significant impact on this directional wind response.

CHAPTER 5

Assessment of the Impacts of Vegetation on Wind Fields over Complex Terrain

A quantitative comparison of wind direction distributions under differing physical and environmental conditions will help develop an understanding of the impacts of vegetation and topography on wind flow behaviour across complex terrain features. Such comparison in this chapter aims to identify whether changes in vegetation or topography significantly alter the relationship between prevailing winds and those experienced within the landscape. Through subsequent analysis, it may be possible to quantify these changing relationships by parameterising joint wind response distribution functions, such as that done within the wind energy industry [e.g. Carta et al., 2008, Erdem and Shi, 2011, Zhang et al., 2013]. Such statistical modelling of these distributions using physical parameters such as vegetation cover would feed into emerging probabilistic approaches to bushfire prediction [Finney et al., 2011, French et al., 2014d, Twomey and Sturges, 2016].

Building on the mathematical foundations of the previous chapter, the following hypotheses are proposed to test for equality between two directional wind response distributions:

$$\begin{aligned} H_0 : f_1 &= f_2 \\ H_A : f_1 &\neq f_2, \end{aligned} \tag{5.1}$$

where f_1 and f_2 represent continuous underlying wind response processes which, when estimated as $\hat{f}_k(X_{st})$, are realised at 1° resolution such that $X_{st} = \{(\phi_s, \psi_t) :$

$s, t = 1, \dots, 360\}$. According to Equation 2.8, where $Y = f + \epsilon$, $Y(X_{ij})$ represent discrete wind distributions observed over the wind direction pairs corresponding to the 16 points of the compass (i.e. N, NNE, NE and so on), such that $X_{ij} = \{(\phi_i, \psi_j) : i, j = 1, \dots, 16\}$.

In order to consider the statistical tests which can be used to understand the impacts of vegetation and topography on directional wind response, the problem is first reduced to the unconditional univariate wind distribution case, i.e. $Y(X_i)$. Through non-parametric combination of univariate tests, the analysis is then extended to the bivariate case, $Y(X_{ij})$, to consider the impacts of vegetation and topography on directional wind response distributions.

This chapter first considers the sensitivity and power of univariate and bivariate Kolmogorov-Smirnov (KS) style non-parametric statistical tests to changes in distributional structure. Monte Carlo methods are used to simulate comparisons of known distributions, and the resulting power curves are analysed to determine thresholds for significance in terms of distributional changes in modal location or shape. These findings are then used to better interpret results of the tests when applied to wind direction distributions.

Section 5.1 first reviews the relevant literature on non-parametric statistical tests, before reintroducing the appropriate wind direction data analysed in this chapter. In Section 5.2, the power of the univariate KS and Kuiper's tests are analysed before being applied to univariate conditional wind direction distributions. Section 5.3 then considers the case of bivariate data. The sensitivity of an extended KS test to changes in bivariate distribution structure is first analysed, and evaluated through a comparison with the power behaviour of the univariate test. An extension of Kuiper's test for the comparison of bivariate circular distributions is then proposed. The power of such an extension is analysed in relation to the univariate Kuiper's test and the bivariate KS test to establish whether it is a reasonable proposition. On application to the bivariate wind direction distributions discussed in the previous chapter, these sensitivity results and relationships are considered. Finally, mean squared difference based non-parametric surface comparison tests are applied to bivariate wind direction distributions. Section 5.4 concludes this chapter.

The research presented in this chapter has been summarised in two journal articles;

Quill, R., Sharples, J.J. and Sidhu, L.A. (In Prep) Sensitivity analysis of Kolmogorov-Smirnov style statistics for univariate and bivariate data.

Quill, R., Sharples, J.J. and Sidhu, L.A. (Under Review) Non-parametric comparison of wind direction to assess the impacts of surface roughness. *Statistical Modelling, submitted March 2017.*

5.1 Background

5.1.1 Literature

Non-parametric statistical tests allow for the comparison of observed datasets without the constraints of parametric estimation, enabling comparison of more complex multimodal, multivariate distributions. However, interpretation of non-parametric results in terms of changes in the underlying physical process can be difficult. One way of developing an interpretation of such results is to analyse the power of the applied statistical test.

The power of a given statistical test can be defined as its ability to detect that the null hypothesis of equality is false when comparing two distributions that are in fact different [Lindgren, 1976]. In the literature, Monte Carlo studies are commonly used to evaluate the power of a test against alternative distributions with known structures [e.g Peacock, 1983]. Goodness-of-fit statistical tests can be used to compare a data sample with a known distribution, such as testing for uniformity or normality using the t -test or χ^2 test, and the power of such tests is well studied. For instance, Bergin [1991] compared four tests for uniformity on the circle, and found that all tests were powerful against a unimodal alternative. However, bimodal or direction-avoidance alternatives caused problems. Distribution-free statistical comparison tests are designed to be powerful against any alternative distribution, and are often used to compare two data samples of unknown underlying distribution structure. Such two-sample tests can include omnibus tests, Cramér-von Mises tests and Kolmogorov-Smirnov (KS) style tests. There are many books that cover

the details of these tests, and more, for both linear and circular datasets, including Conover [1980], Neave and Worthington [1988], Jammalamadaka and SenGupta [2001] and Mardia and Jupp [2000].

KS style tests are a widely used and well studied form of non-parametric statistical test. When the parametric forms of the distribution structures of two data samples are unknown, it is natural to consider a comparison of the two empirical distribution functions. This is the basis on which two-sample KS style tests were developed by Kolmogorov [1933] and Smirnov [1939]. The univariate test is not only distribution-free, but also highly efficient and invariant to the ordering of the data sample [Peacock, 1983]. Numerous early studies of these tests provided approximations to the asymptotic behaviour of the statistics as well as tabulated percentage points and critical regions [e.g. Feller, 1948, Stephens, 1965, Maag and Stephens, 1968, Stephens, 1969, 1970, Maag and Dicaire, 1971, Maag, 1973, Koziol, 1980]. Early studies also indicated the power of the KS test against a variety of distributions, showing the test to be more powerful than alternatives such as the χ^2 test [Massey, 1951, Pearson, 1963].

Kuiper [1960] proposed an adaptation of the KS test to provide a distribution-free comparison test for circular datasets that has since been used in numerous studies, particularly for small sample sizes [e.g. Upton and Fingleton, 1989, Mardia and Jupp, 2000]. In particular, Batschelet [1981] highlights its use within biological studies, and Freedman [1979] and Verdoux et al. [1997] exemplify the use of Kuiper's test to understand seasonality within epidemiological research. In comparing the KS and Kuiper's tests, Pearson [1963] and Louter and Koerts [1970] independently found Kuiper's test to be more powerful. However, Louter and Koerts [1970] also concluded (Srinivasan [1971] later concurred) that the power of both tests was very sensitive to the alternative distribution, despite their distribution-free properties.

In many applications, it is necessary to compare bivariate distributions, for instance in astronomy [Babu and Feigelson, 2005] or textural pattern recognition [Ferryanto, 1995]. Peacock [1983] and Justel et al. [1997] outlined a bivariate extension of the KS test which utilised an extended definition of the empirical distribution function

for bivariate data. Conceptually this appears simple, but in practice the extension of the KS test to the multivariate case is not trivial, with issues including the increased computational demand involved with the selection of comparison points [Fasano and Franceschini, 1987]. Peacock [1983, page 621] showed that although the bivariate KS test could not be considered distribution-free, the test was sufficiently powerful against any “reasonable” distribution structure (i.e. one that was not strongly correlated), and therefore could be confidently used in practice. Further study of the extended KS test has continued within the field of astronomy, with a generalised multivariate test analysed by Koen and Siluyele [2007]. Through application of the extended KS test, Koen and Siluyele [2007] found it to be more powerful than a number of alternative tests, as well as easy and fast to run.

Pewsey and Kato [2016] considered a bootstrap goodness-of-fit testing approach which provided insight into the power of a number of toroidal uniformity tests applied to bivariate circular samples [Wellner, 1979, Jupp, 2009]. The test however depends on a parametric approach, requiring knowledge of the marginal distributions and binding densities used to construct the bivariate circular distributions. In this study non-parametric methods are considered, and Pewsey and Kato [2016] note that further work is currently underway to develop a non-parametric version of their approach. There are few non-parametric testing procedures available for toroidal surfaces, and a bivariate extension of the two-sample Kuiper’s test has not yet been tackled in the literature.

Throughout the literature there are a number of other methods presented to test the equality of multivariate distributions, in particular Chiu and Liu [2009] defined an extension of the popular Cramér-von Mises goodness-of-fit test. Dhar et al. [2014] also proposed a bivariate extension of the Q-Q plot and associated statistical tests which were found to be more powerful than the bivariate KS and Cramér-von Mises tests under specified alternative conditions.

In comparing estimated continuous directional wind response distributions, non-parametric surface comparison tests are also considered. For example, Chiang and Puri [1984] proposed a rank order test for parallelism between multi-dimensional regression surfaces, equivalent to testing for equality between standardised surfaces.

However, the rank order test considered surfaces produced through linear regression, where the constants associated with each covariate were known, or had been estimated. Delgado [1993] presented a statistic for testing the equality of non-parametric curves which did not depend on a smoothing term. The level and power of the test was assessed using simulated curves, and the results were found to be promising over a number of models and error distributions.

Non-parametric tests based on the squared difference between surfaces are also popular in the applied literature. Wang and Ye [2010] used medical imagery as motivation to determine a generalised multi-dimensional form of the test statistic derived through non-parametric analysis of covariance by Dette and Neumeyer [2001]. The work went on to establish the generalised form of asymptotic normality for this test statistic. Bowman [2006] considered multi-dimensional air pollution data in the development of a similar test statistic based on the work of Young and Bowman [1995], which drew analogies with one-way analysis of variance. As opposed to the statistic developed by Wang and Ye [2010], which was calculated using the difference between each individual surface, the statistic developed by Bowman [2006] was calculated using the difference between the individual surfaces and the surface generated by the combined dataset.

Wang and Ye [2010] and Bowman [2006] both worked on fixed design problems, where observed data were taken at pre-determined and fixed points on the regression surfaces. Srihera and Stute [2010] added to the field by providing a test which was shift and scale invariant, and considerations were taken for surfaces made up of observations taken at different locations across the surface, i.e. random designs.

5.1.2 Wind Data Comparisons

5.1.2.1 Flea Creek Valley, NSW

Non-parametric comparison tests were performed on the pairs of wind direction distributions observed at approximately coincidental locations across Flea Creek Valley between 2007 and 2014. Figure 5.1 shows a map of the coincidental weather station pairs; FCV1 (A2-B4), FCV2 (A3-B6), FCV3 (A4-B8) and FCV4 (A4-B10).

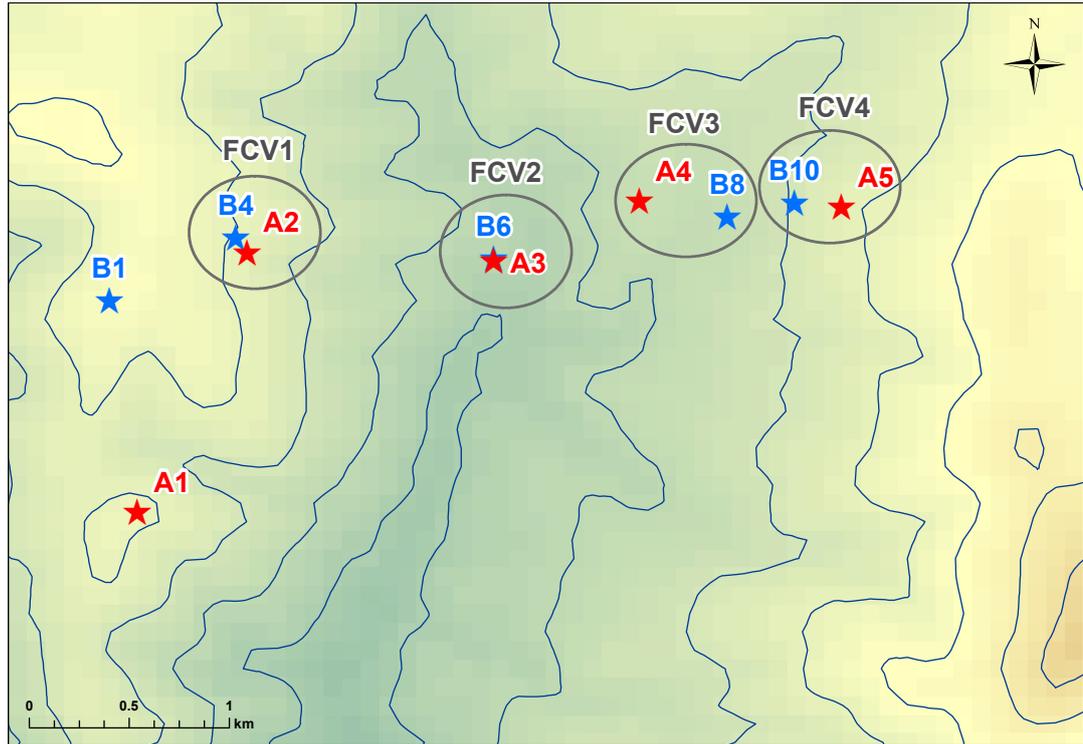


Figure 5.1: Map of coincidental station locations across FCV between 2007 (red; A1 to A5) and 2014 (blue; B1, B4, B6, B8 and B10).

Table 3.1 describes the vegetation present at each site in 2007 [Sharples et al., 2010], while Table 3.2 describes the vegetation in 2014. There are clear differences in language: where Sharples et al. [2010] described “burnt trees” and “sparse canopy”, the 2014 vegetation was described as “medium density” with “large eucalypts” surrounding the stations. This regrowth of vegetation around the station sites would have likely impacted upon the air movement around the anemometers, and thus provides a unique opportunity to better understand the impact of vegetation regrowth on wind direction distributions in complex terrain.

Despite seven years of regrowth, the wind response distributions from the same locations in both years can look visually similar. Figures 5.2 and 5.3 shows the univariate wind direction distributions observed across Flea Creek Valley, conditional on observing a WNW prevailing wind direction at the ridge top. These distributions are linearised from the wind roses shown in Chapter 3, and are discussed in more detail as pairs in relation to the differences between the two sample years which are more visibly evident when represented on the line.

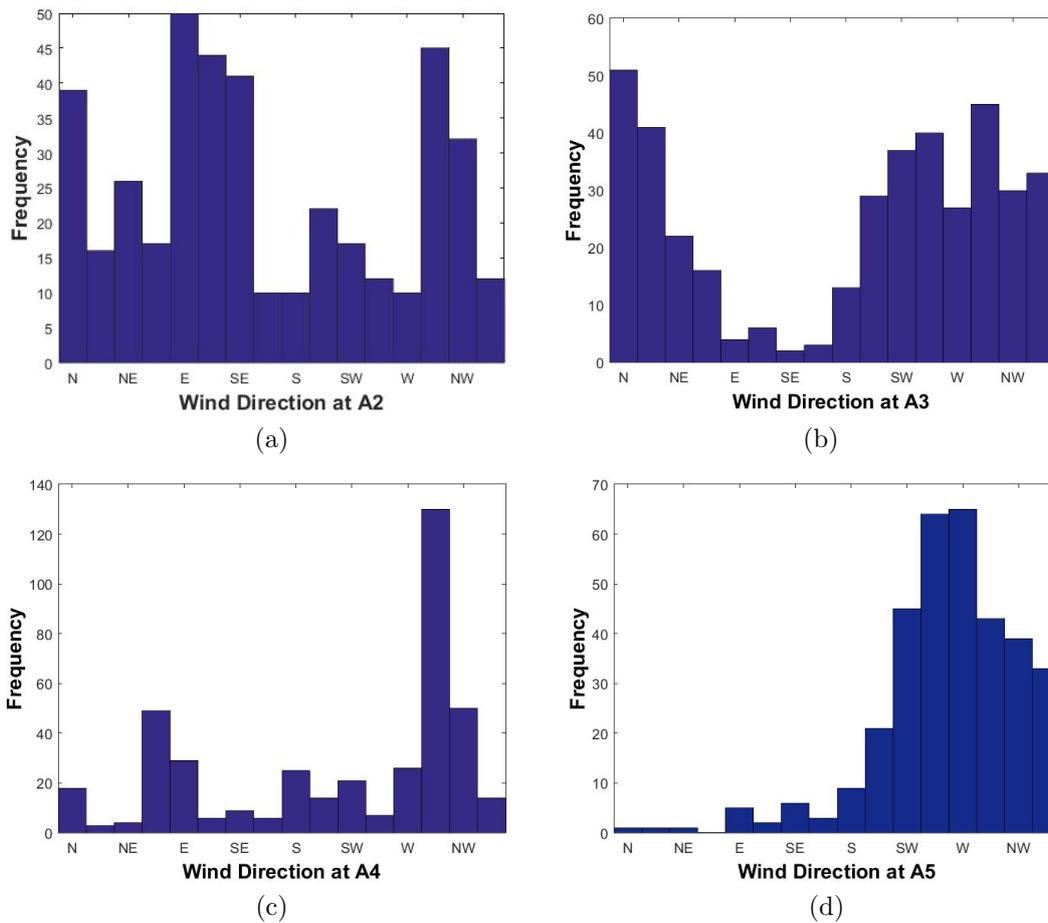


Figure 5.2: Observed conditional wind direction distributions from 2007 at (a) FCV1 on the leeward slope (A2), (b) FCV2 on the valley floor (A3), (c) FCV3 on the lower windward slope (A4), and (d) FCV4 on the upper windward slope (A5), assuming a WNW observed at A1.

FCV1 shows a strong easterly mode in both 2007 and 2014 (Figures 5.2(a) and 5.3(a)), while FCV2 and FCV4 show broad westerly or north-westerly modes in both years (Figures 5.2(b) and 5.3(b), and Figures 5.2(d) and 5.3(d)). FCV3 shows clearer differences between the two years. Figure 5.3(c) shows a wide mode centred around a southerly direction, while Figure 5.2(c) shows a mode around the north-west to northerly direction.

The estimated continuous wind response distributions using the thin plate smoothing spline technique developed and analysed in Chapter 4, for the four points across Flea Creek Valley in both 2007 and 2014 are shown in Figures 4.17 and 4.18 in the previous chapter. Each pair shares a similar number of modes between the two years, with modes generally in similar positions. It would appear that the variability of wind direction has increased in 2014, perhaps due to the increased vegetation

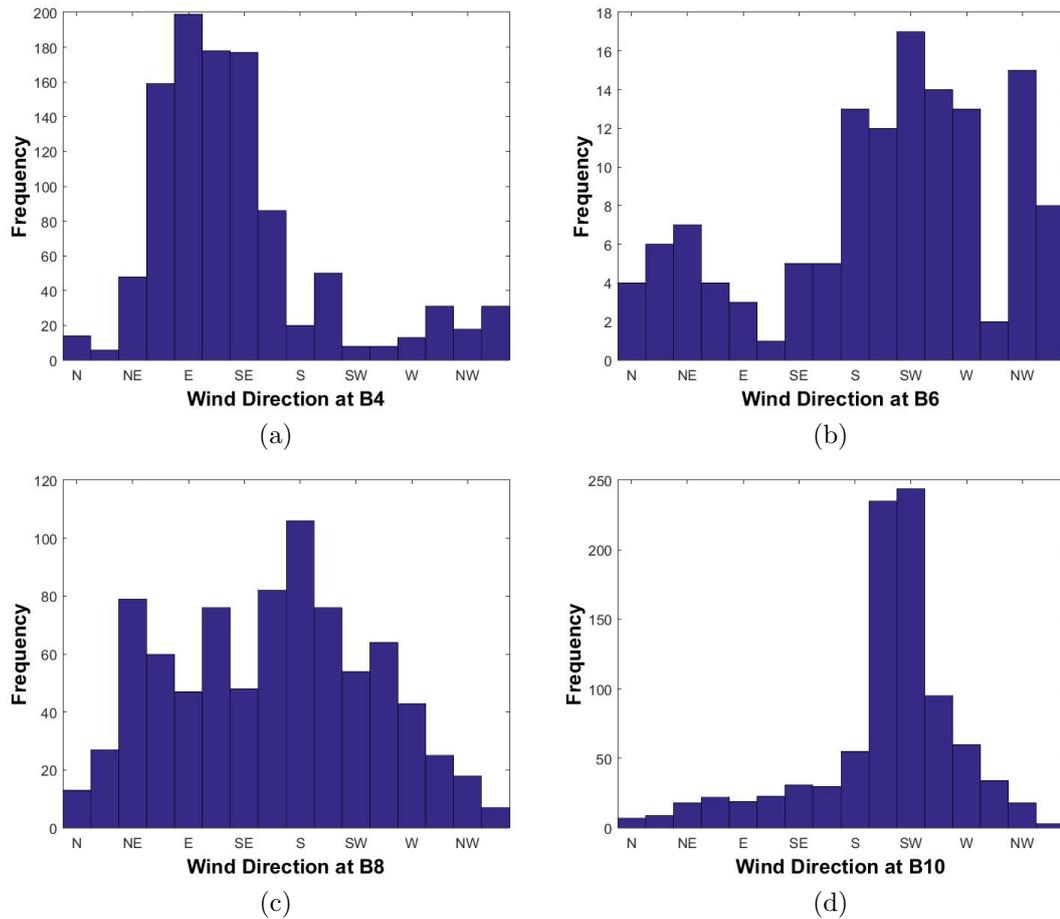


Figure 5.3: Observed conditional wind direction distributions from 2014 at (a) FCV1 on the leeward slope, (B4) (b) FCV2 on the valley floor (B6), (c) FCV3 on the lower windward slope (B8), and (d) FCV4 on the upper windward slope (B10), assuming a WNW observed at B1.

and thus turbulence produced around the weather stations [e.g. Finnigan, 2000, Belcher et al., 2012].

Pairs FCV1 (Figures 4.17(a) and 4.18(a)) and FCV2 (Figures 4.17(b) and 4.18(b)) show the most similarities between the two sample years. For FCV2, four modes are evident in both years suggesting that the wind direction is relatively variable throughout the valley floor and this remains consistent with the increased vegetation. Meanwhile, for FCV1 two dominant modes are evident at approximately (ESE, ESE) and (WNW, ESE), with a third around (WNW, WNW). The off-diagonal mode shown at (WNW, ESE) aligns with the distributions shown in Figures 5.2(b) and 5.3(b), which suggest that under westerly prevailing winds, a wind reversal region develops on this east-facing slope (leeward to the westerly prevailing) so that easterly winds are experienced at the surface.

On the west-facing slope (Figures 4.17(d) and 4.18(d)), FCV4 shows a faint off-diagonal mode at (ESE, WNW) in both years. This again suggests a wind reversal region when the west-facing slope is leeward to the easterly winds. An additional mode is present in 2007 at (ESE, ESE) which represents an alignment of easterly prevailing winds with easterly surface winds on the leeward slope. This is likely to have developed due to a greater proportion of easterly prevailing winds observed in 2007 (approximately 29%) as opposed to 2014 (approximately 15%).

Finally, for FCV3, the stations were not placed in identical positions across the two sample years due to logistical constraints, and so it is not surprising that the distributions show the most discrepancies (as also shown in Figures 5.2(c) and 5.3(c)). While Figure 4.18(c) suggests that westerly prevailing winds produce predominantly southerly surface winds with wide variation, Figure 4.17(c) shows a clear mode at (WNW, WNW) which indicates that surface winds align with the prevailing winds. Both years, however, also show considerable variation across the bivariate distributions.

Table 5.1 shows the data sample sizes for the construction of the conditional and joint distributions between 2007 and 2014. The table highlights the variation in sample size across both case study periods, and also shows that the sample sizes dealt with in the application of these comparison tests are far greater than those handled in the existing literature. For example, Stephens [1965] tabulated percentage points for Kuiper’s statistic up to sample sizes of 100, and Peacock [1983] considered the behaviour of the bivariate KS test up to finite sample sizes of 50.

Table 5.1: Sample sizes for FCV pairs.

	Conditional		Joint	
	2014	2007	2014	2007
FCV1	1,046	403	2,537	2,809
FCV2	129	399	346	2,823
FCV3	825	411	1,676	2,964
FCV4	903	338	1,864	2,161

5.1.2.2 *National Arboretum Canberra, ACT*

Figure 5.4 shows the weather station sites across the NAC used for analysis in this chapter. Table 5.2 details the comparisons made to understand the impacts of

varying vegetation under similar topography, and varying topography under similar vegetation on wind direction across the area.

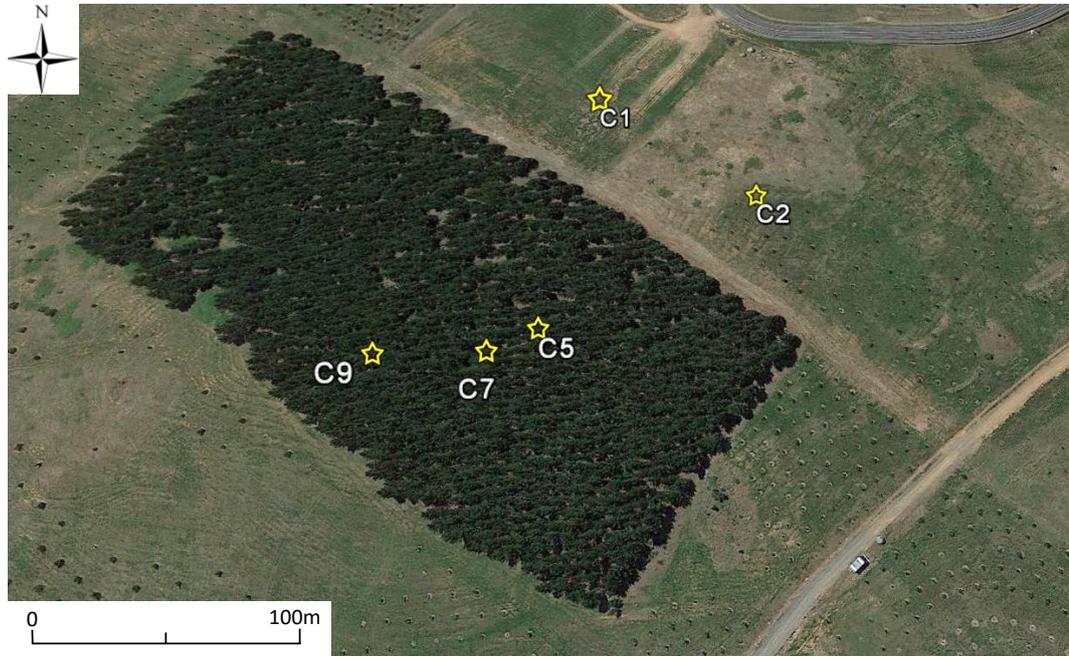


Figure 5.4: Map of the NAC with station locations from 2015 (yellow; C1, C2, C5, C7 and C9).

Table 5.2: Location descriptions for comparison pairs at the NAC in 2015.

NAC1	(C2) Clear Leeward Slope	(C5) Pine Leeward Slope
NAC2	(C5) Pine Leeward Slope	(C7) Pine 45° Leeward Slope
NAC3	(C5) Pine Leeward Slope	(C9) Pine Cross Slope
NAC4	(C7) Pine 45° Leeward Slope	(C9) Pine Cross Slope

Figure 5.5 shows the univariate conditional wind direction distributions observed across the NAC under west-northwesterly (WNW) prevailing wind conditions. Again, these distributions are linearised from the wind roses shown in Chapter 3, and are discussed in more detail here in relation to the differences seen between the pairs detailed in Table 5.2.

Each site experiences a dominant wind direction under the WNW prevailing winds. On the clear leeward slope (C2, Figure 5.5(a)), and on the cross slope within the pine plantation (C9, Figure 5.5(d)), westerly winds are experienced. In contrast, on the leeward slope at C5 and 45° leeward slope at C7 within the pine plantation (Figures

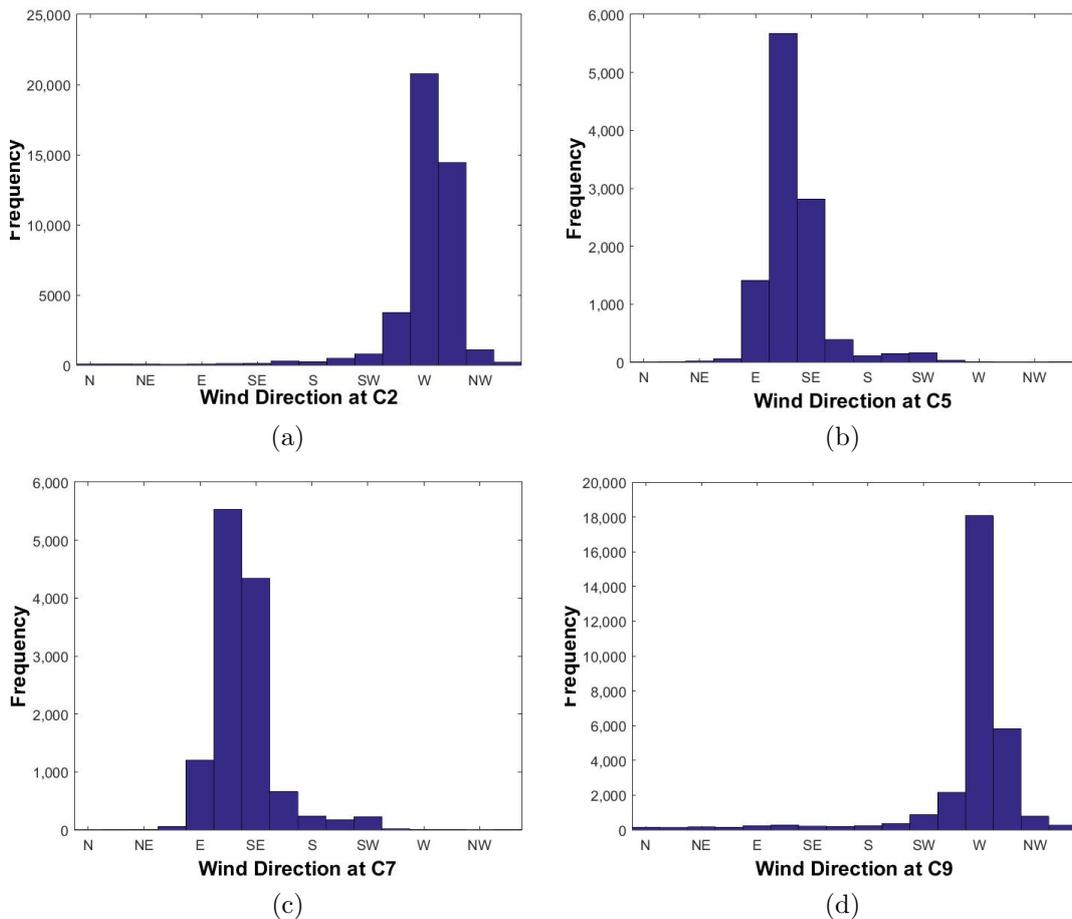


Figure 5.5: Observed conditional wind direction distributions from the NAC at (a) C2 on the clear leeward slope, (b) C5 on the pine leeward slope, (c) C7 on the pine 45° leeward slope and (d) C9 on the pine cross slope.

5.5(b) and (c)), easterly to south-easterly winds are experienced. Perhaps due to far greater sample sizes observed at the NAC (Table 5.3), due to the increased frequency of observations over a longer study period, there is less variation across the four wind direction distributions than observed across FCV. This lack of variability may also be due to the undulating surrounding terrain and the uniformity the vegetation across both the clear and forested stands of the study site.

The estimated continuous directional wind response distributions for the four NAC sites for comparison are shown in Figure 4.19 in Chapter 4. Again the lack of variability observed in Figure 5.5 is evident. All four sites exhibit two modes; one is consistent at approximately (ESE, ESE) suggesting that when the less dominant easterly prevailing winds blow across the region, each of the sites experiences easterly winds, no matter the vegetation or topographical differences between them.

When westerly prevailing winds are experienced, however, the response of wind flow varies across the four sites. As shown in Figure 5.5, the surface winds align with the prevailing westerlies on the cleared leeward slope (C2) and the cross slope within the pines (C9), showing a (WNW, WNW) mode in Figures 4.19(a) and (d). However, the leeward and 45° leeward slopes within the pines (C5 and C7) experience winds of the opposite direction, resulting in a (WNW, ESE) mode in Figures 4.19(b) and (c). This suggests the presence of wind reversal regions within the vegetation at these sites, caused by not just the vegetation but also the relative aspect of the terrain with respect to the prevailing wind direction.

Table 5.3 shows the data sample sizes used to construct the conditional and joint wind direction distributions. This table again highlights the variability in sample size across the study. The sample sizes for the NAC are considerably larger than those from FCV due to the increased length of the study period and the increased sample frequency from 30 minutes to 1 minute. These sample sizes are now many orders of magnitude greater than those used to develop and analyse the statistical comparison tests discussed in the previous section.

Table 5.3: Sample sizes for NAC pairs.

	Conditional		Joint	
	<i>m</i>	<i>n</i>	<i>m</i>	<i>n</i>
NAC1	42,756	10,844	176,031	32,583
NAC2	10,844	12,484	32,583	28,479
NAC3	12,484	30,191	32,583	74,428
NAC4	10,844	30,191	28,479	74,428

5.2 Univariate Wind Direction

5.2.1 Simulation Study

5.2.1.1 Univariate Kolmogorov-Smirnov Test

The Kolmogorov-Smirnov (KS) test, devised by Kolmogorov [1933] and later adapted for two samples by Smirnov [1939], compares the probability distributions of two datasets (as a measure of goodness-of-fit to a known distribution or as a comparison of two samples) by considering the maximum difference between the empirical

distribution functions (EDFs) of each set. There are three key features of the univariate test: it is highly efficient, it is distribution-free and the results of the test do not depend on the ordering of the data [Peacock, 1983].

In the literature, it is shown that the asymptotic behaviour of the two-sample test tends towards the asymptotic behaviour of the one-sample test, i.e. comparing a dataset to a known distributions [Abrahamson, 1967]. For a comparison of two samples with m and n data points, respectively, the quantity $N = \frac{mn}{m+n}$ is defined and the univariate KS test statistic is given by

$$D_N^{(1)} = \sup_x |F_m(x) - G_n(x)|, \quad (5.2)$$

where F_m and G_n are the empirical distribution functions, i.e. $F_m(x) = P(X \leq x)$ [Kolmogorov, 1933, Smirnov, 1939].

Since the statistic $D_N^{(1)}$ is known to be proportional to the sample size [e.g. Kendall and Stuart, 1979], critical values given by Massey [1951] are $d_{0.01} = 1.63/\sqrt{N}$ at the 1% significance level, $d_{0.05} = 1.36/\sqrt{N}$ at the 5%, and $d_{0.10} = 1.22/\sqrt{N}$ at the 10% level. An alternative standardised statistic can also be defined [Kolmogorov, 1933, Smirnov, 1939],

$$Z_N^{(1)} = \sqrt{N}D_N^{(1)}. \quad (5.3)$$

The asymptotic behaviour of this statistic was given by Kolmogorov [1933],

$$P(Z_\infty^{(1)} > z) \simeq 2 \exp(-2z^2), \quad (5.4)$$

accurate to 3 decimal places for $z > 1.08$, i.e. $\alpha < 0.194$, with $Z_\infty^{(1)}$ defined by the transformation

$$1 - Z_N^{(1)}/Z_\infty^{(1)} = 0.20N^{-0.6}. \quad (5.5)$$

Stephens [1970] also introduced a modified Kolmogorov-Smirnov statistic $D_N^* = D_N^{(1)}(\sqrt{N} + 0.12 + 0.11\sqrt{N})$, for which the same asymptotic behaviour as in Equation 5.4 was approximated, since for large N this modified statistic reduces to $Z_N^{(1)}$.

5.2.1.2 Univariate Kuiper's Test

In many applications, including considering wind direction, data can be inherently circular. In order to apply the KS test as described in the previous section, the circle would need to be cut to reshape the data on the line. Unfortunately, the KS test is not invariant to the location at which the circle is cut, and so Kuiper [1960] introduced an adaptation of the KS statistic to handle such a situation. As with the KS test, Kuiper's test is distribution-free and is commonly used in the univariate form to compare relatively small samples, i.e. less than 100 [e.g. Batschelet, 1981, Freedman, 1979]. However, in contrast to the KS test, there is little consensus in the literature regarding the approximation of the asymptotic behaviour of Kuiper's test statistic.

To account for circularity, Kuiper's test statistic is a combination of the maximum difference and the minimum difference between empirical distribution functions. The one-dimensional, one-sample statistic is defined as

$$V_n^{(1)} = D_n^+ + D_n^-, \quad (5.6)$$

where $D_n^+ = \sup_x \{F_n(x) - g(x)\}$ and $D_n^- = |\inf_x \{F_n(x) - g(x)\}|$, with $F_n(x) = P(X \leq x)$ defined to be the EDF of a sample of size n , and $g(x)$ is some known distribution.

As with the KS test, this statistic is proportional to the sample size, so Kuiper [1960] and Stephens [1970] approximated the asymptotic distribution of $\sqrt{n}V_n^{(1)}$, for $v > 6/5$, as

$$P(\sqrt{n}V_n \geq v) \simeq (8v^2 - 2) \exp(-2v^2) - \frac{8v}{3\sqrt{n}} (4v^2 - 3) \exp(-2v^2), \quad (5.7)$$

which is based on continuous distributions, and is therefore conservative when testing discrete distributions. For large n , the second term tends to zero, thus the approximation tends to

$$P(\sqrt{n}V_n \geq v) \simeq (8v^2 - 2) \exp(-2v^2). \quad (5.8)$$

This approximation was applied by Durbin [1973] and Arsham [1988] for large sample sizes, i.e. $n \rightarrow \infty$.

Stephens [1970] proposed an alternative statistic to account for sample size,

$$V_n^* = V_n (\sqrt{n} + 0.155 + 0.24/\sqrt{n}), \quad (5.9)$$

and used the approximate formula for the asymptotic behaviour of $\sqrt{n}V_n$ given in Equation 5.8, accurate to 3 decimal places for $v > 1.26$, i.e. $\alpha < 0.447$ (or accurate to 2 decimal places for $v > 1.06$ or $\alpha < 0.74$). As for the equivalent KS statistic, for large n this alternative tends to $\sqrt{n}V_n$, therefore the approximate asymptotic behaviour is given by Equation 5.7, for $v > 6/5$.

Stephens [1970] gave the upper quantiles of V_n^* as equivalent to those for $\sqrt{n}V_n$ given by Stephens [1965], i.e. $v_{0.01}^* = 2.001$, $v_{0.05}^* = 1.747$ and $v_{0.10}^* = 1.620$. When substituted for v in Equation 5.8, these critical values give; $P(V_n^* > 2.001) = 0.0106$, $P(V_n^* > 1.747) = 0.0545$ and $P(V_n^* > 1.62) = 0.110$ which are all very close to the prescribed significance levels. Maag and Dicaire [1971] also showed that Stephens' approximation to the distribution of Kuiper's statistic is accurate across a range of percentage points (0.5% to 10%).

As with the two-sample KS test, the quantity $N = \frac{mn}{m+n}$ is defined so that the two-sample Kuiper test statistic is given by

$$V_N^{(1)} = D_N^+ + D_N^-, \quad (5.10)$$

where $D_N^+ = \sup_x \{F_m(x) - G_n(x)\}$ and $D_N^- = |\inf_x \{F_m(x) - G_n(x)\}|$. The EDFs of the two sample datasets of sizes m and n are given by F_m and G_n , respectively. It is shown and assumed throughout much of the literature [e.g. Kuiper, 1960, Abrahamson, 1967, Mardia and Jupp, 2000] that the distribution of V_N (two-sample statistic) tends to the distribution of V_n (one-sample statistic).

For small sample sizes, i.e. when $n \neq m$, and $n + m \leq 28$, or when $n = m$ and $3 \leq n \leq 100$, Maag and Stephens [1968] gave exact percentage points for V_N as defined in Equation 5.10. For larger, unequal sample sizes, the test outlined is based on the assumption that the distribution of $\sqrt{N}V_N$ (defined as in the two-sample

KS test) tends to that of $\sqrt{n}V_n$ as $n, m \rightarrow \infty$. It was conjectured by Maag and Stephens [1968] that the asymptotic behaviour of this two-sample statistic might be approximated by a function of the form

$$P\left(\sqrt{N}V_N \geq v\right) \simeq \sum_{l=1}^{\infty} (8l^2v^2 - 2) \exp(-2l^2v^2) + 2.828 v \sqrt{\frac{nm}{n+m}} \sum_{l=1}^{\infty} (3l^2 - 4l^2v^2) \exp(-2l^2v^2). \quad (5.11)$$

Once expanded, the first term of this approximation gives the asymptotic distribution of V_n and V_n^* given in Equation 5.8. However, it was remarked by Maag and Stephens [1968] that Hodges [1957] found erratic behaviour of the distribution of the KS statistic for large unequal sample sizes, and this behaviour is likely to extend to Kuiper's statistic.

In this study a parallel formation of Kuiper's test is also evaluated. Batschelet [1981] defined the test statistic; $k = mnV_N$. To determine significance, the one-sample critical value table from Stephens [1965] is used alongside a translation to the two-sample test; $L(\alpha) = [mn(m+n)]^{1/2} K(\alpha)$, where $K(\alpha)$ is the one-sample critical value for $n < m$. If $k > L(\alpha)$, the null hypothesis is rejected at level α .

A number of alternative formulations of the asymptotic behaviour of Kuiper's test also exist in the literature [e.g. Abrahamson, 1967, Raghavachari, 1973, Steck, 1969, Upton and Fingleton, 1989], but drawing parallels with the KS test results is more complicated. The aim of this study is to improve understanding of the sensitivity of the KS and Kuiper's tests for better interpretation in application to wind direction data. With this in mind, the analogous formulations of asymptotic behaviour for both the KS test and Kuiper's test are analysed, as well as the statistic given by Batschelet [1981] which is commonly applied in practice.

5.2.1.3 Simulation Procedure

Distributions were constructed to emulate the properties of the discrete binned wind direction distributions observed in this research. These simulated baseline distributions were then compared to similar distributions with shifted structures, such as changes in mean or standard deviation. All simulations were coded in

MATLAB® [2016] using the `kstest` function, with further details of the code provided in Appendix D. The following procedure describes how the univariate KS test was performed. This procedure was conducted for each of the subsequent tests analysed in this study.

1. Simulate a discrete probability density function (PDF) based on a Normal distribution; $N(8, 1)$, with random error $\sigma^2 \sim N(0, 0.01)$. This discrete PDF consists of 16 bins of unit width in order to emulate the structure of the wind direction data presented in Chapter 3.
2. Generate a data sample from this PDF by randomly generating a sample size (N) between 100 and 1000, and then calculating the proportions of the sample allocated to each discrete bin.
3. Repeat steps (1) and (2) to simulate a second sample for comparison. This second sample is constructed either under changes in the mean of the Normal PDF while the standard deviation is held constant, or under relative changes in the standard deviation while the mean is held constant.
4. The KS test is run to compare the two simulated distributions; generating the observed test statistics, $D_N^{(1)}$ and $Z_N^{(1)}$ defined by Equations 5.2 and 5.3, respectively.
5. Significance of the test statistics is calculated using the asymptotic p -value formula for $Z_N^{(1)}$ given in Equation 5.4 (P_Z), and the tabulated critical values for $D_N^{(1)}$ at a 1% significance level, defined by Massey [1951]; $d_{0.01} = 1.63/\sqrt{N}$.

In some cases, the asymptotic distributions of the test statistics are available in the literature, and so the p -values can be approximated. Alternatively, the p -values can be calculated using Monte Carlo (MC) simulation schemes. These types of tests include permutation tests where all possible combinations of the data are permuted, while maintaining the dataset structures. The test statistics observed from each permutation form the distribution under the null hypothesis of equality between the sample distributions [Berger, 2009]. The exact p -value can then be calculated as the percentage of the distribution that is more extreme than each observed statistic.

Bootstrap tests calculate the test statistic for all possible variations of the data set, without restricting the data design. The exact calculation of the p -value was traditionally not conducted under the guise of computational resource restrictions, but with modern computing power, this is generally no longer the case. However, with large datasets, the calculation of bootstrap style tests can still become cumbersome and subsets of all the possible bootstrap simulations (at least 1000) are used to estimate the exact distribution of the test statistics, and thus the p -value. This estimate is therefore a conservative approximation of the p -value [Neuhauser, 2012].

In the case of the KS test, it is known that the asymptotic behaviour and corresponding critical values are conservative in their estimate of significance, meaning that they are more likely to suggest that a pair of distributions are significantly different since it requires more evidence to support the null hypothesis of equality [e.g. Massey, 1951, Slakter, 1965]. MC estimates of the p -values for $D_N^{(1)}$ or $Z_N^{(1)}$ are expected to be less conservative than the approximations to the asymptotic behaviour, but still more conservative than the exact p -value.

In addition to the univariate Normal distributions, simulations of the univariate von Mises distribution were also used to understand the sensitivity of Kuiper's test to changes in modal structure. The von Mises (vM) distribution (considered the circular analogue of the Normal distribution) requires two parameters; mean and concentration. The concentration parameter, κ , can be thought of as the inverse of the standard deviation for the Normal distribution, i.e. a high κ value indicates high concentration and low variance. The baseline von Mises distribution used for this study had mean π and concentration 5, $vM(\pi, 5)$, giving a similar structure to the $N(8, 1)$ baseline distribution. To assess the ability of Kuiper's test to handle circularity, a $vM(0, 5)$ distribution was used as a third baseline.

The MATLAB® [2016] function, `circ_kuipertest`, from the CircStat package [Berens, 2009], uses Batschelet's statistic [Batschelet, 1981]. The code calculates the observed statistic $k = mnV_N$, as well as providing the critical value K and related significance level α of the observed statistic. The significance level can be thought of as equivalent to the p -value of the test statistic and is reported as

such, p_k , but only takes the discrete α values 0.01, 0.02, 0.05, 0.10 etc. due to the constraints of using the tabulated critical values of Stephens [1965]. The code was adapted to also calculate $V_N^{(1)*} = \sqrt{N}V_N$, along with the asymptotic formula for its p -value given by Equation 5.8. To summarise, the variables calculated using the adapted code were;

- $V_N^{(1)} = D_N^+ + D_N^-$,
- $k = mnV_N^{(1)}$ (as in `circ_kuipertest`),
- p_k : p -value from `circ_kuipertest`,
- K : critical value from `circ_kuipertest`,
- $V_N^{(1)*} = \sqrt{N}V_N^{(1)}$, and
- $p_{V^*} = \left(8 \left(V_N^{(1)*}\right)^2 - 2\right) \exp\left(-2 \left(V_N^{(1)*}\right)^2\right)$.

Monte Carlo simulations were also used to form the null distribution of $V_N^{(1)}$, k and $V_N^{(1)*}$, and calculate the approximate p -values for the observed statistics. Further details of this code are given in Appendix D.

To measure the power of the tests at a given significance level α (i.e. 1%, 5% or 10%), the test statistics and p -values were calculated over 1000 MC simulations, each time randomly generating the error distribution. The power was calculated as the percentage of the p -values or critical values that fell within the significance region, i.e. the percentage of p -values less than the α level, or the percentage of test statistics greater than the critical value for the α level.

For incremental changes in mean and standard deviation (or concentration), plots are presented showing the power of the tests using a 1% significance level. Analysis of 5% and 10% significance levels were also conducted but are not presented here. It is expected that for shifts close to zero the power of the test would approximately match the significance level. For a test to be considered powerful, a steep rise in the curve is then expected as the distributions shift further from the baseline. A test might be considered powerful against known alternatives if it shows a rejection rate of say 70% or above. By determining the power of the KS and Kuiper's tests against the manipulated alternatives, it is possible to determine thresholds of modal changes at which the test is sufficiently powerful to reliably detect the difference between two distributions.

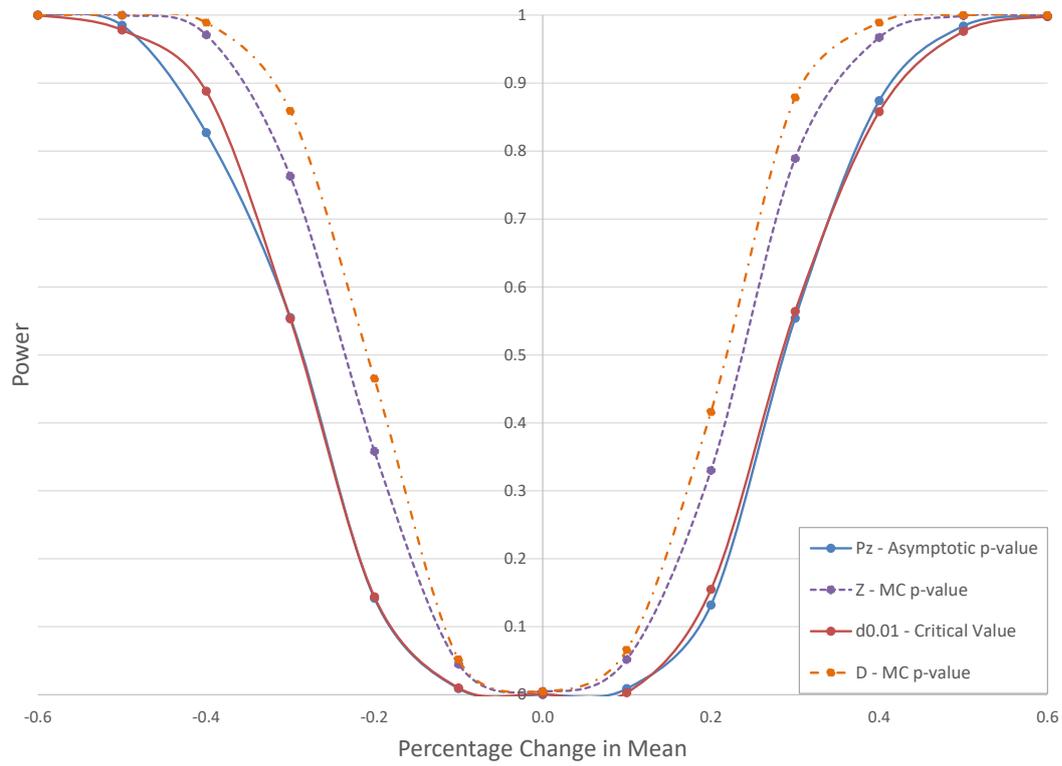
For changes in standard deviation or concentration, it is expected that the power of the test is best analysed with respect to the relative change rather than the absolute, since decreases in standard deviation are constrained and finite, while increases can be infinite. Reflecting this, analysis not reported here shows that the power of the KS and Kuiper's tests against absolute changes in standard deviation or concentration is asymmetric. Relative changes in standard deviation or concentration are therefore considered such that, for instance, halving the standard deviation of the Normal distributions (i.e. reducing $N(0, 1)$ to $N(0, 0.5)$, given by a relative change of -0.5) is equivalent to doubling it (i.e. increasing $N(0, 1)$ to $N(0, 2)$, given by a relative change of $+0.5$).

5.2.1.4 Results

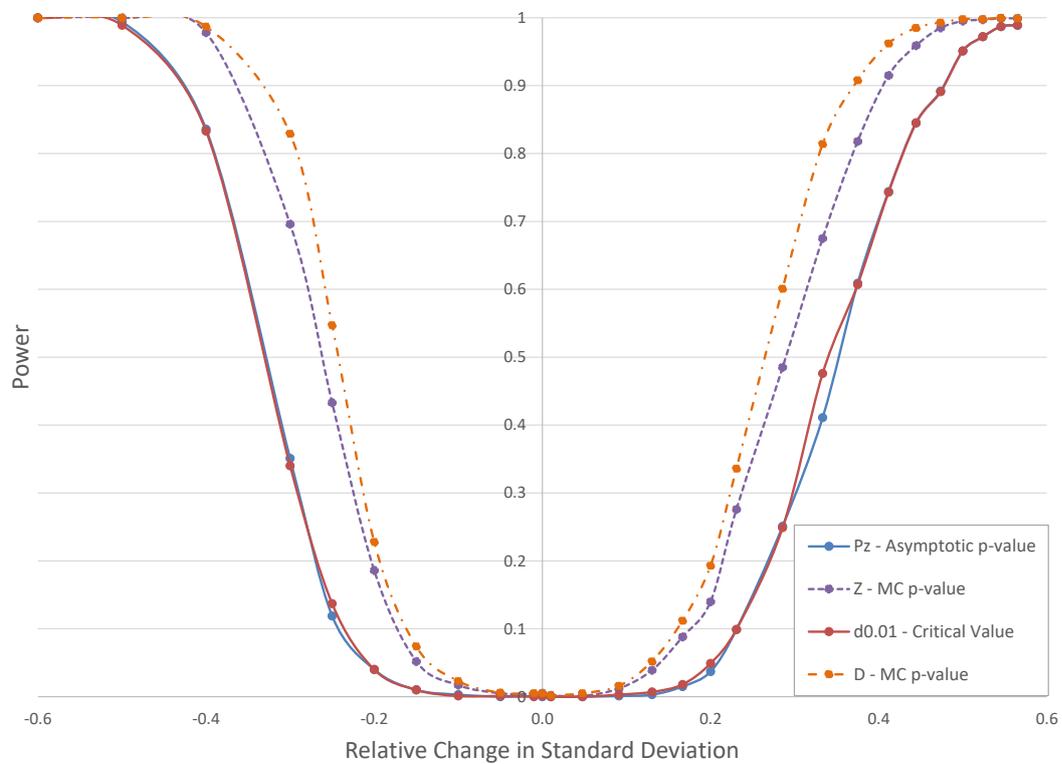
Figure 5.6(a) shows the power of the univariate KS test against shifts in the mean of the Normal distribution, with standard deviation held constant. As expected, the power is symmetric, showing that increases and decreases in the mean value have almost identical impacts on the power of the test. The plot shows that the test is successful when comparing almost identical distributions, with the power level matching the 1% significance level for changes in the mean of less than 10% of the discrete bin size. The power of the test then quickly rises to above 70% as the changes in mean become greater than $\pm 30\%$ or $\pm 40\%$ (depending on the formulation of the p -value). This steepness in the power curve indicates only a narrow region of marginal power at small changes to the mean value (between $\pm 10\%$ and $\pm 40\%$ of the bin size) for all formulations of the p -value.

Figure 5.6(b) shows that the KS test is again successful when comparing similar distributions, with power levels matching the significance levels for relative changes in standard deviation less than ± 0.2 . The test becomes quickly powerful (above 70%) when standard deviation is changed by a factor of ± 0.3 or ± 0.4 (depending on the formulation), again showing a relatively narrow region of marginal power.

Both plots in Figure 5.6 indicate agreement with the literature by showing the conservative nature of the asymptotic p -value estimate and the critical value (solid lines). The MC estimates (dotted lines) show consistently higher power curves, suggesting that they are more sensitive to changes in the mean. The power curves



(a)



(b)

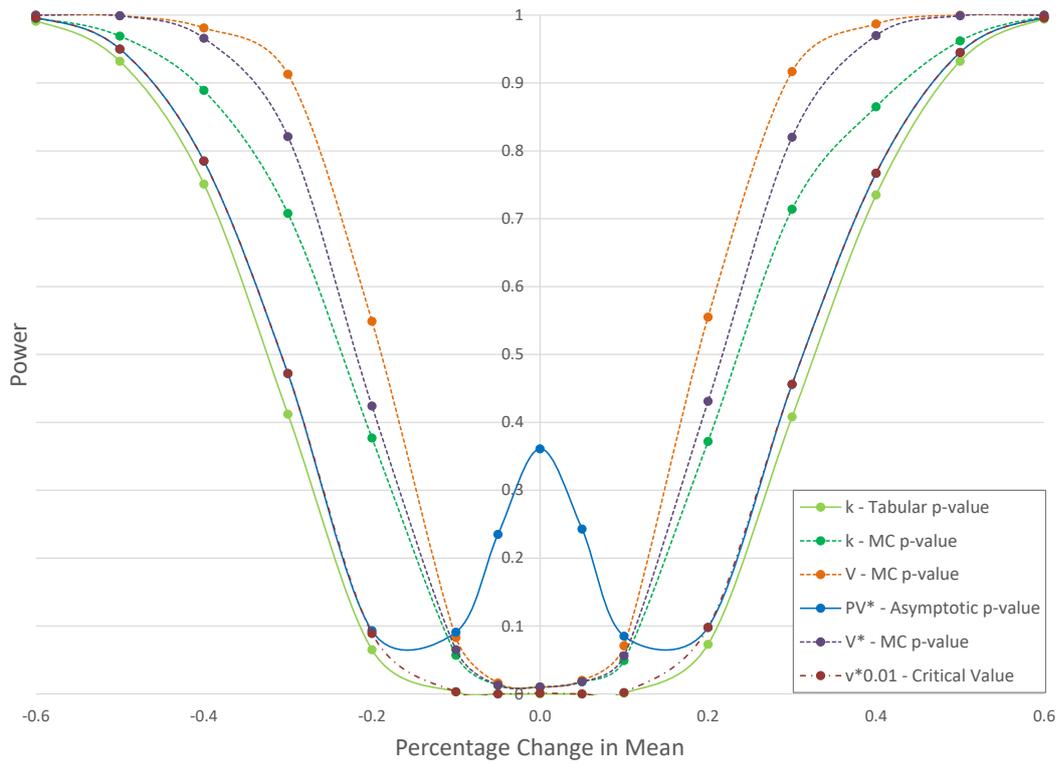
Figure 5.6: Power of the univariate KS test to (a) percentage change in mean and (b) relative change in standard deviation for the Normal distribution, at a 1% level.

for the critical value and asymptotic p -value are almost identical since they are derived from the same formula (Equation 5.4). However, the MC power curves indicate that the $D_N^{(1)}$ statistic is slightly more conservative than the $Z_N^{(1)}$ statistic which is scaled by the sample size.

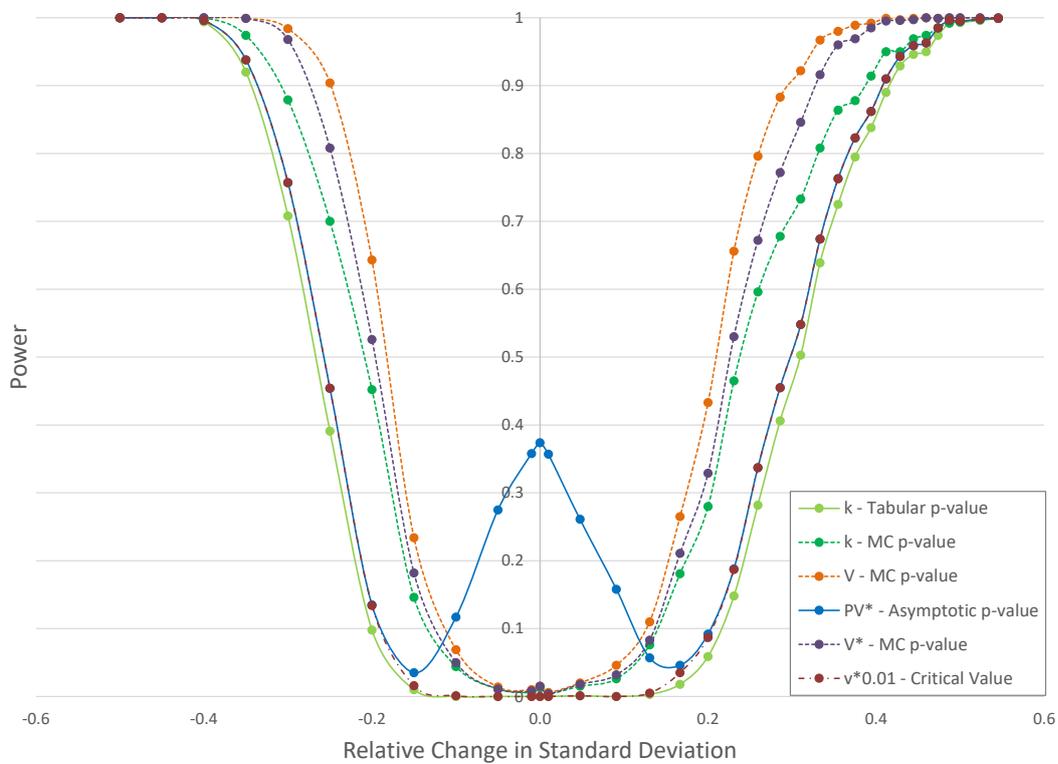
Figure 5.7 shows the sensitivity of Kuiper's test to changes in mean and standard deviation for the Normal distribution. As for the KS test results shown in Figure 5.6, Kuiper's test is symmetrically sensitive to changes in mean and relative changes in standard deviation. The MC p -values (dotted lines) are also consistently more sensitive to the changes than the conservative critical values and asymptotic approximations (solid lines), again concurring with the literature. The magnitude of sensitivity is also similar for Kuiper's test as it was for the KS test, with similar changes in mean (less than $\pm 10\%$ of a discrete bin size) and standard deviation (less than a relative change of ± 0.2) showing power levels equivalent to the 1% significance level. Depending on the test's form, changes in mean greater than approximately $\pm 30\%$ to $\pm 40\%$ are required to give a powerful Kuiper's test (above 70%), again showing a relatively narrow region of marginal power. As noted by Louter and Koerts [1970], Kuiper's test appears to be more powerful against shifts in standard deviation than the KS test, only requiring changes of approximately ± 0.2 or ± 0.3 to reach powers over 70%.

The asymptotic p -value and $v_{0.01}^*$ critical value again exhibit similar power curves, while that for the k statistic appears more conservative. For the MC results, a hierarchy emerges showing V_N to be the most sensitive statistic, followed by the scaled V_N^* statistic (as shown for the KS test). Finally, the k statistic is the most conservative, as was also found with the asymptotic results.

The key difference between the results shown in Figure 5.6 and those shown in Figure 5.7 is given by the approximation to the asymptotic p -values. For changes in mean of less than $\pm 20\%$ and standard deviation of less than ± 0.15 , the power of this p -value increases rather than tending towards the significance level. The power curves show that the approximate asymptotic p -value gives almost 40% false positives when the two distributions being tested are identical except for error. Interestingly, the $v_{0.01}^*$ critical value does not show such a peak around zero changes.



(a)



(b)

Figure 5.7: Power of the univariate Kuiper's test to (a) percentage change in mean and (b) relative change in standard deviation for the Normal distribution, at a 1% level.

For both the KS and Kuiper’s tests, the approximations to the asymptotic p -values have limited validity. Stephens [1970] stated that the approximate asymptotic p -value formula for Kuiper’s test was accurate to 3 decimal places for $v > 1.26$ (or $\alpha > 0.447$), while Kolmogorov [1933] gave the approximate asymptotic p -value for the KS test, accurate for $z > 1.08$. From the simulation results for Kuiper’s test on identical or very similar distributions, the v values were very low (as low as 0.42) meaning that the assumptions underpinning the asymptotic p -value formula were no longer valid. The difference between the two tests can be clearly explained by plotting the approximations of the asymptotic p -value for each case; Equations 5.4 and 5.7 (Figure 5.8).

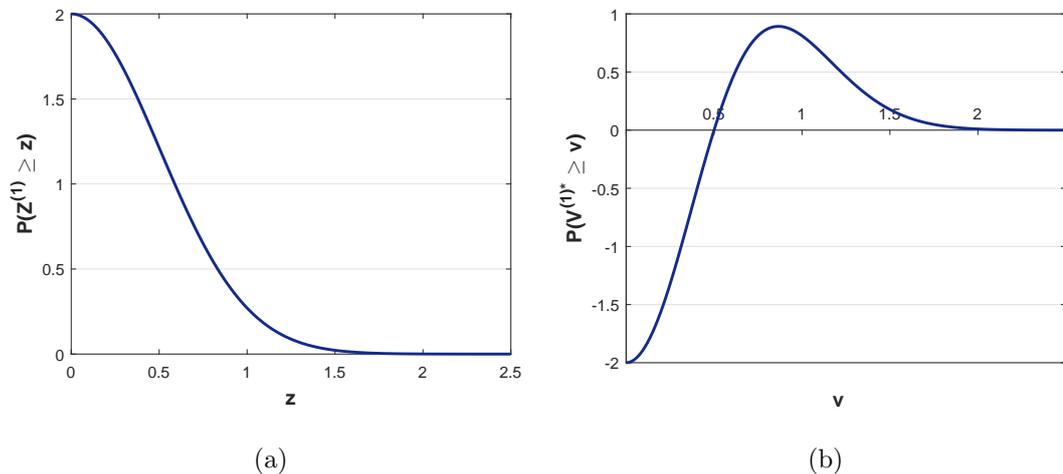


Figure 5.8: Approximations for the asymptotic p -values for the univariate (a) KS test (Equation 5.4) and (b) Kuiper’s test (Equation 5.8).

In the case of the KS test (Equation 5.4), the formula takes a Gaussian form, so peaks at $z = 0$ and monotonically decreases with increasing z values which result from increasingly different distributions. In the approximation of the asymptotic behaviour of the standardised Kuiper’s statistic (Equation 5.8), the negative exponential is multiplied by a term of order v^2 . Therefore, the approximation is no longer monotonic and peaks at approximately $v = 0.866$. With smaller p -values, the quadratic term dominates and therefore, as v approaches zero, the approximation decreases towards -2. This results in small p -value approximations for small values of v generated by similar or identical distributions (or even negative p -values for $v \leq 0.5$, which are truncated to zero). This is confirmed by the power curves since empirical results show that the peak in power occurs as consistent simulations

result in $V_N^{(1)*}$ statistic values below 0.866. Therefore, in contrast to the KS test and with false positive rates reaching 95%, there is clearly a need for caution in the use of the approximation to the asymptotic behaviour of Kuiper's test beyond its range of validity.

Figures 5.9 and 5.10 show the sensitivity of Kuiper's test to changes in the mean and concentration of the von Mises distributions. To assess how well Kuiper's test accounts for circularity, the sensitivity is shown for a mode simulated in the 'centre' of the distribution when represented on the line, $vM(\pi, 5)$ (Figure 5.9), as well as for a mode simulated on the 'edge' of the distribution, $vM(0, 5)$ (Figure 5.10). As with the Normal distribution, the power is consistently symmetric to changes in mean and concentration, and the MC p -value estimates are consistently less conservative than the asymptotic approximations. No matter whether the mode is simulated at the 'centre' or 'edge' of the distribution, Kuiper's test shows very similar sensitivity to changes in the mean and concentration to that of the KS test and Kuiper's test applied to the Normal distribution. This reinforces the concepts behind using Kuiper's test as the circular adaptation of the KS test.

As indicated by Louter and Koerts [1970] and Srinivasan [1971], the sensitivity of Kuiper's test appears to vary according to the distribution; Kuiper's test is more powerful against the von Mises distribution than when applied to the Normal distribution. The power curves shown in both Figures 5.9 and 5.10 are steeper than those in Figures 5.6 and 5.7, resulting in narrower regions of marginal power; between $\pm 20\%$ and $\pm 40\%$ for changes in mean, and between ± 0.2 and ± 0.3 (or ± 0.3 and ± 0.4 , depending on the tests' form) for relative changes in concentration. Despite this increase in power, the relative behaviour of the forms of the test remains consistent. The k statistic continues to be the least sensitive to changes in distribution structure (both in asymptotic form and from the MC results), and the MC power curves indicate that $V_N^{(1)}$ is again less conservative than the scaled $V_N^{(1)*}$.

There is again a clear exception to the similarity between power results from the KS test and Kuiper's test. As with the application of Kuiper's test to the simulated Normal distributions, a peak in power is observed for the asymptotic p -value for changes close to zero. It appears that the behaviour has been exacerbated by

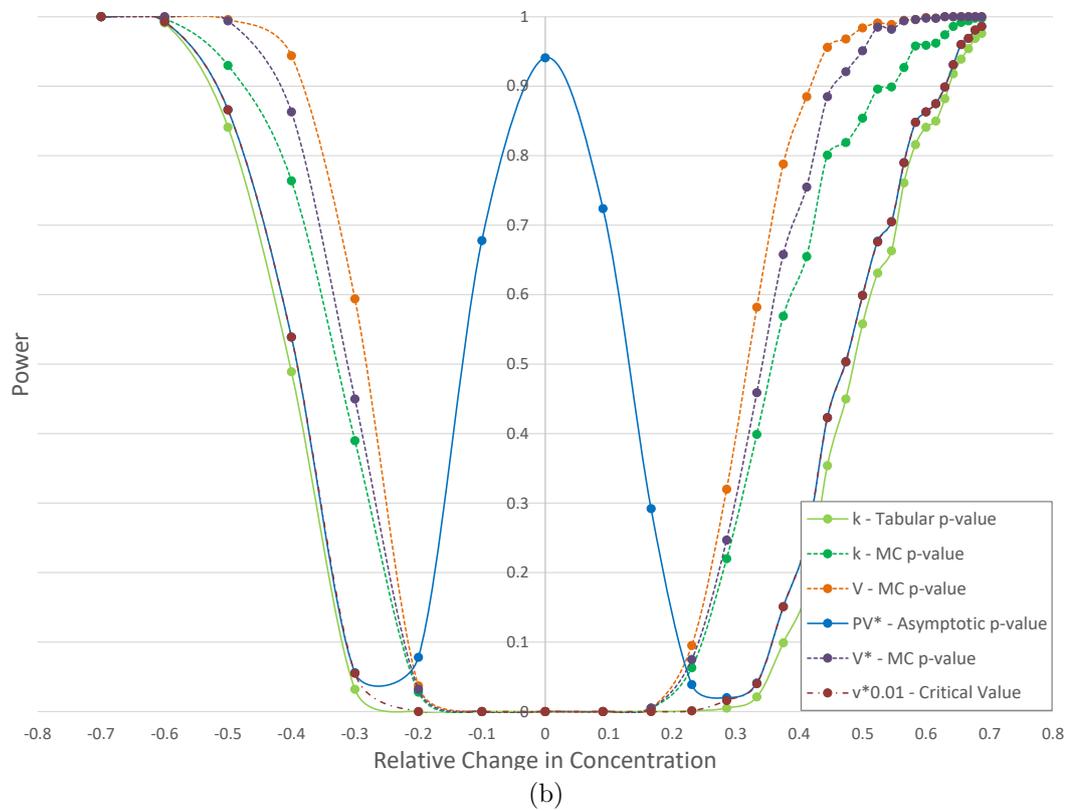
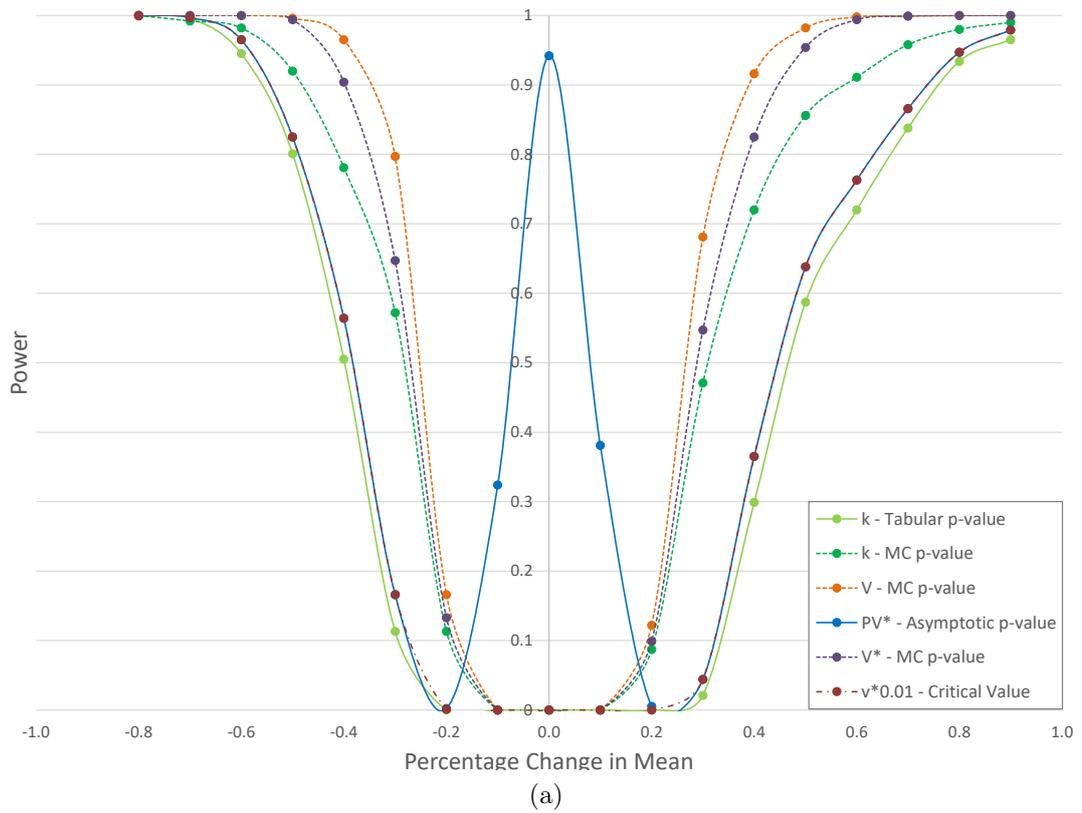


Figure 5.9: Power of the univariate Kuiper's test to (a) percentage change in mean and (b) relative change in concentration for the von Mises distribution with centre mode, at a 1% level.

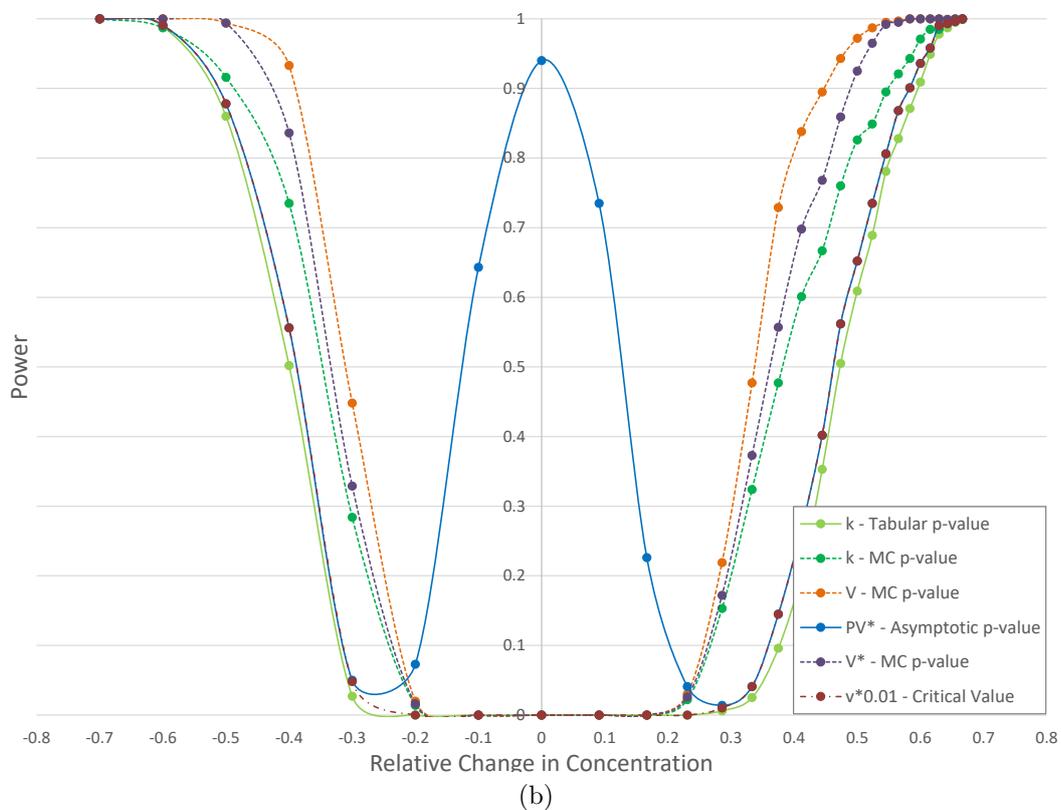
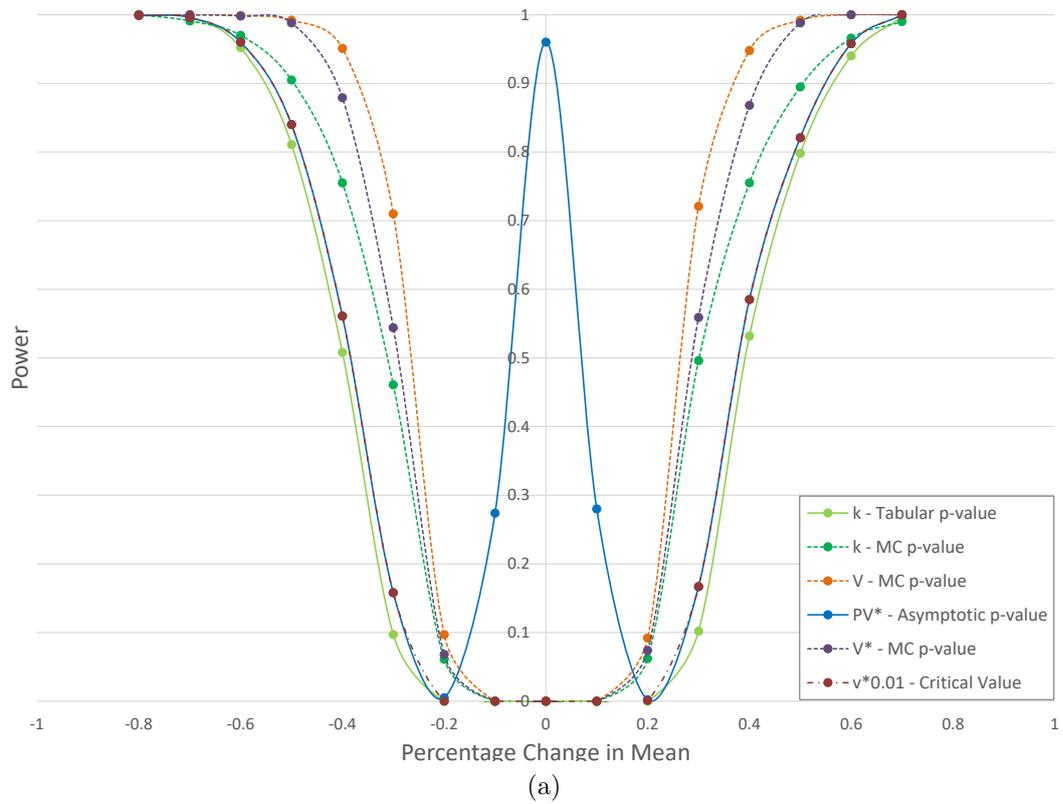


Figure 5.10: Power of the univariate Kuiper's test to (a) percentage change in mean and (b) relative change in concentration for the von Mises distribution with edge mode, at a 1% level.

application to the circular distribution. Power levels reach around 95% when there is no simulated change in mean or concentration, indicating a 95% false positive rate. The width of the peak is approximately the same as found for the Normal distribution for the mean, but has almost doubled in width for shifts in concentration (comparing Figure 5.7(b) to Figures 5.9(b) and 5.10(b)). Looking at the results in more detail, the v values for these simulations become as low as 0.31, considerably below the validity threshold of 1.26 for the use of the approximation.

5.2.1.5 Discussion

Across all results presented for the KS test, it is seen that the MC p -values exhibit higher power across all types of change, suggesting that the MC estimates are more likely to determine a significant difference between distributions given a smaller change in distributional structure. This is to be expected since it is reported numerous times in the literature that the asymptotic approximation to the p -value of the observed $D_N^{(1)}$ statistic is conservative (particularly for discrete distributions), meaning that the test is more likely to indicate that two distributions are significantly different when experiencing only small changes.

The power of the KS test (in all forms) is symmetrical to positive and negative shifts in the mean value. The test is shown to be powerful, giving high power values at relatively small changes to the mean; the power reaches almost 100% after a mean shift equivalent to just $\pm 50\%$ of the discrete bin size. When relative changes in standard deviation are considered, the KS test is clearly symmetric in its power. The test is shown to be relatively powerful again, with only a narrow region of potentially inconclusive test results (between a factor of ± 0.2 and ± 0.4).

Although in this analysis a power threshold of 70% has been used, the question of how powerful a test must be to reliably detect a difference between two distributions is likely to be dependent upon the application. In the case of discrete observed data, only changes significant after a one-bin mean shift may be considered important, in which case the KS test is very powerful. However, the form of the test being used and the desired application have a clear impact on this threshold.

In general, the results for Kuiper's test are very similar to those given by the KS test for the Normal distribution. Both are powerful and symmetric around changes to the mean and standard deviation. In addition, application to both von Mises distributions showed very similar power characteristics to those given for the Normal distribution and the KS test, and results were clearly invariant to the location of the mode on the von Mises distribution. This all provides empirical evidence to support the consistency of Kuiper's test as the circular adaptation of the KS test.

The single variation in behaviour between the KS test and Kuiper's test was the peak in power shown for the asymptotic p -value approximation for identical, or very similar, simulated distributions. The peak in power appeared to worsen with the application to the von Mises distribution, and empirical results indicate that the simulated comparisons resulted in consistently lower values of $V_N^{(1)*}$. The behaviour of this asymptotic approximation can be easily explained with more detailed inspection of the derived formulae for each test, but the exacerbation with the application to the von Mises distribution suggests that the form of the distributions being compared may in fact impact these results, despite the distribution-free nature of the test.

5.2.2 Conditional Wind Direction Distributions

Each of the observed wind direction distributions shown in Figures 5.2, 5.3 and 5.5 is conditional on a WNW prevailing wind direction observed at the reference stations on the ridge top for each case study. The combined sample size, $N = \frac{nm}{n+m}$, for each univariate wind direction dataset is shown in Table 5.4 along with the results of the univariate Kolmogorov-Smirnov test applied to each pair.

Across the leeward and windward slopes of Flea Creek Valley, at pairs FCV1, FCV3 and FCV4, the $D_N^{(1)}$ statistic is significantly greater than the critical values given by $d_{0.01}$ and $d_{0.05}$ at the 1% and 5% significance levels. Similarly, the asymptotic p -values, $P_{Z_N^{(1)}}$, are very small. In addition, the Monte Carlo p -values, $P_{D_N^{(1)}}^{MC}$ and $P_{Z_N^{(1)}}^{MC}$, are zero for these three pairs. These results indicate that there is significant evidence against the null hypothesis of equality in these cases, and therefore suggest that the wind direction distributions from 2007 and 2014 are significantly different.

Table 5.4: KS test results for univariate conditional wind direction distribution pairs from FCV and the NAC.

	N	$D_N^{(1)}$	$d_{0.01}$	$d_{0.05}$	$Z_N^{(1)}$	$P_{Z_N^{(1)}}$	$P_{D_N^{(1)}}^{MC}$	$P_{Z_N^{(1)}}^{MC}$
Flea Creek Valley								
FCV1	291	0.2259	0.0956	0.0797	3.8529	2.55×10^{-33}	0	0
FCV2	97	0.1630	0.1651	0.1377	1.6096	0.0112	0.0008	0.0054
FCV3	274	0.4226	0.0984	0.0821	6.9987	5.70×10^{-43}	0	0
FCV4	246	0.4893	0.1039	0.0867	7.6740	1.41×10^{-51}	0	0
National Arboretum Canberra								
NAC1	8,650	0.9432	0.0175	0.0146	87.7213	0	0	0
NAC2	5,803	0.1165	0.0214	0.0179	8.8713	8.77×10^{-69}	0	0
NAC3	7,978	0.9086	0.0182	0.0152	81.1578	0	0	0
NAC4	8,832	0.9062	0.0173	0.0145	85.1599	0	0	0

On the valley floor, for the pair FCV2, the situation is somewhat different. The value for the $D_N^{(1)}$ statistic is greater than the critical value given at the 5% level, but still smaller than that given for the 1% level, suggesting that for this pair there is not sufficient evidence against the null hypothesis at the 1% significance level (but there is at the 5% level). Similarly, the asymptotic p -value, $P_{Z_N^{(1)}}$, is 0.0112, and therefore the null hypothesis cannot be rejected at the 1% significance level. These results suggest that the wind direction distributions on the valley floor may not be significantly different between 2007 and 2014. The Monte Carlo p -values remain small and suggest that there is a significant difference between the two distributions, but they are relatively close to 0.01 compared to the other three pairs.

Across the National Arboretum Canberra, all pairs are significantly different according to the results shown in Table 5.4. When comparing the leeward slope to the 45° leeward slope within the pines (NAC2), the test statistics $D_N^{(1)}$ and $Z_N^{(1)}$ show the smallest values by far, resulting in the only asymptotic p -value to be non-zero. This suggests that this pair may be more similar than the others, which agrees with a visual assessment of the univariate distributions shown in Figure 5.5. The critical values across all pairs at the NAC are very small (a tenth of those for FCV) due to the large sample sizes resulting from data collected over a longer period and at higher frequency than at FCV.

Table 5.5 shows the results of applying Kuiper’s test to the pairs of univariate conditional wind direction distributions across FCV and the NAC. Similar patterns to

those found using the KS test are evident. At FCV, all four pairs indicate a significant difference in wind direction distributions between 2007 and 2014 according to the statistical results. However, FCV2 shows the largest $P_{V_N^{(1)*}}$ value and shows the only non-zero MC p -value for the $V_N^{(1)}$ statistic. This again indicates that this pair shows the least statistical evidence of significant difference between the two years.

Table 5.5: Kuiper’s test results for univariate conditional wind direction distributions pairs from FCV and the NAC.

	$V_N^{(1)}$	$V_N^{(1)*}$	$P_{V_N^{(1)*}}$	$k^{(1)}$	$L_{0.01}$	$P_{V_N^{(1)}}^{MC}$	$P_{V_N^{(1)*}}^{MC}$	$P_{k^{(1)}}^{MC}$
Flea Creek Valley								
FCV1	0.3619	6.1723	2.46×10^{-31}	152,545	49,454	0	0	0
FCV2	0.2399	2.3686	0.0006	12,348	10,431	0.0005	0	0
FCV3	0.4506	7.4631	1.85×10^{-46}	152,784	40,964	0	0	0
FCV4	0.4893	7.6740	3.31×10^{-49}	149,350	38,944	0	0	0
National Arboretum Canberra								
NAC1	0.9472	88.0977	0	439.2×10^6	10.0×10^6	0	0	0
NAC2	0.1173	8.9326	3.14×10^{-67}	15.9×10^6	3.6×10^6	0	0	0
NAC3	0.9228	82.4233	0	302.1×10^6	7.3×10^6	0	0	0
NAC4	0.9224	86.6820	0	347.7×10^6	8.0×10^6	0	0	0

Across the NAC, the statistical results show that all four wind direction distribution pairs are significantly different as vegetation and aspect changes. All p -values, whether asymptotic or MC, are zero except one. For NAC2, the observed statistic values of $V_N^{(1)}$, $V_N^{(1)*}$ and $k^{(1)}$ are all considerably smaller than for the remaining pairs, and the $P_{V_N^{(1)*}}$ value is the only non-zero p -value but is still extremely small.

5.2.2.1 Discussion

Visually, changes in wind direction are difficult to interpret across Flea Creek Valley due to higher noise levels in the wind distributions (Figures 5.2 and 5.3), but across the National Arboretum Canberra there are clear shifts in the observed wind direction modes as the vegetation and aspect conditions change (Figure 5.5). From the findings of the simulation study, it is expected that the KS style tests would be very powerful against the definitive shifts in mean shown across the NAC where the spread of the modes does not appear to vary. It was shown in the simulation studies that both the KS and Kuiper’s tests were powerful against small changes in mean and standard deviation (or concentration) relative to the size of the discrete bins observed here. This gives confidence in the application of the tests across both

case studies, and the consistent results across all forms of the tests provide strong evidence to support the conclusions.

In the simulation study, the Monte Carlo p -value estimates were shown to be less conservative than the asymptotic approximations. This is evident in the results shown in Tables 5.4 and 5.5 where most MC p -values are zero, indicating that all but one pairings are significantly different. Using the raw test statistics and the asymptotic results, Tables 5.4 and 5.5 show that only the wind direction pairs of FCV2 and NAC2 consistently yielded the least evidence in support of the rejection of equality. This suggests that, except at these two sites, the univariate wind direction distributions changed significantly with the vegetation regrowth experienced at FCV and the changing vegetation and topographical conditions across the NAC.

At FCV2, the observed wind direction distributions show some of the greatest variation, meaning that the KS and Kuiper's tests were inhibited in developing evidence against the null hypothesis and thus determining the distributions to be significantly different. This high variation may be caused by a number of localised features creating increased turbulence around the station, including within-canopy effects and topographical features such as the knoll on which the station was situated within the valley floor [e.g. Finnigan, 2000, Belcher et al., 2012].

The pair NAC2 compares the forested leeward slope at the National Arboretum Canberra to the forested 45° leeward facing slope to the WNW prevailing winds. Figure 5.5 indicates that the observed conditional wind direction distributions were very similar between these two sites. An easterly peak shown in these distributions indicate a wind reversal region beneath the vegetation under WNW prevailing winds. This suggests the existence of vegetation and aspect thresholds for the generation of wind reversals, consistent with the computational findings of Simpson et al. [2013] where such thresholds indicated the sufficient conditions for atypical fire spread.

5.3 Bivariate Wind Direction

5.3.1 Simulation Study

5.3.1.1 Bivariate Kolmogorov-Smirnov Test

Peacock [1983] and Justel et al. [1997] proposed an extension of the univariate KS test to the bivariate case, while maintaining its attractive features (high efficiency, distribution-free and invariant to ordering). The test was extended through considering the EDF of a bivariate sample, $\{x, y\}$, in terms of the four probabilities; $P1 = P(x \leq X, y \leq Y)$, $P2 = P(x \geq X, y \leq Y)$, $P3 = P(x \leq X, y \geq Y)$ and $P4 = P(x \geq X, y \geq Y)$. The bivariate KS test then considers the maximum of the maximum differences between each of these probabilities. After applying the test to a variety of surface structures, Peacock [1983, page 621] concluded that the test was sufficiently distribution-free under all “reasonable” surfaces, but not under those that were highly correlated.

For a comparison of two samples, with m and n data points such that $N = \frac{mn}{m+n}$, the bivariate KS test statistic is defined as

$$D_N^{(2)} = \max(D_N^{P1}, D_N^{P2}, D_N^{P3}, D_N^{P4}), \quad (5.12)$$

where D_N^P defines the maximum absolute differences between the empirical distribution functions, F_m and G_n , constructed over the four probabilities for example

$$D_N^{P1} = \sup_{(x,y)} |F_m^{P1}(x, y) - G_n^{P1}(x, y)|, \quad (5.13)$$

where $F_m^{P1}(x, y) = P(x \leq X, y \leq Y)$ [Gosset, 1987]. Again, as in the univariate case, $D_N^{(2)}$ is expected to be proportional to the sample size [Peacock, 1983, Gosset, 1987], so the standardised statistic is also defined as

$$Z_N^{(2)} = \sqrt{N} D_N^{(2)}. \quad (5.14)$$

In this study, the search for the supremum of the difference between EDFs over all four quadrants of the bivariate space was conducted across only the data points

observed in the dataset, thus employing the methodology suggested by Fasano and Franceschini [1987] to reduce the computational demand of the test.

The asymptotic behaviour of the $Z_N^{(2)}$ statistic was given by Peacock [1983] and Gosset [1987], for the area of interest $P(Z_N^{(2)} > z) \lesssim 0.2$, as

$$P(Z_\infty^{(2)} > z) \simeq 2 \exp(-2(z - 0.5)^2), \quad (5.15)$$

where $Z_\infty^{(2)}$ is derived from the relationship

$$1 - Z_N^{(2)}/Z_\infty^{(2)} = 0.53N^{-0.9}, \quad (5.16)$$

with $m, n \gtrsim 10$.

Peacock [1983] gave a table of critical values for $Z_N^{(2)}$, but with only small sample sizes up to 50; $z_{0.01} = 2.06$, $z_{0.05} = 1.83$ and $z_{0.10} = 1.70$. This is orders of magnitude smaller than the sample sizes used in this study.

5.3.1.2 Bivariate Kuiper's Test

Similar in format to the bivariate extension of the two sample KS test, an extension of Kuiper's test is proposed. For a comparison of two samples, with m and n data points such that $N = \frac{mn}{m+n}$, the bivariate Kuiper's test statistic is defined as

$$V_N^{(2)} = \max(V_N^{P1}, V_N^{P2}, V_N^{P3}, V_N^{P4}), \quad (5.17)$$

where V_N^P defines the univariate Kuiper's statistic between the empirical distribution functions, F_m and G_n , constructed over each of the four probabilities defined in Section 5.3.1.1. For example,

$$V_N^{P1} = (D_N^+)^{P1} + (D_N^-)^{P1}, \quad (5.18)$$

where $(D_N^+)^{P1} = \sup_{(x,y)} \{F_m^{P1}(x, y) - G_n^{P1}(x, y)\}$, with $F_m^{P1}(x, y) = P(x \leq X, y \leq Y)$, and likewise $(D_N^-)^{P1} = |\inf_{(x,y)} \{F_m^{P1}(x, y) - G_n^{P1}(x, y)\}|$.

Using $V_N^{(2)}$ as a basis, two further test statistics could be defined to standardise the value with respect to the sample size. As for the previous tests, the statistic is

standardised with respect to the sample size to give

$$V_N^{(2)*} = \sqrt{N}V_N^{(2)}. \quad (5.19)$$

The third test statistic is drawn in parallel with the univariate statistic defined by Batschelet [1981],

$$k^{(2)} = mnV_N^{(2)}. \quad (5.20)$$

Approximating the asymptotic behaviour of these proposed test statistics lies outside the scope of this study, but it is considered that issues in the approximation might include, for example, the need to show each test statistic to be at least sufficiently distribution-free as in the bivariate extension of the KS test. The statistics' invariance to data ordering, or the location at which the torus is cut to represent the plane, should also be rigorously established. The proposed test statistic may be calculated over each possible toroidal permutation, with non-parametric combination techniques then applied to form an invariant statistic.

It might also be expected that the asymptotic behaviour of an extended Kuiper's statistic takes a similar exponential form to that of the univariate KS and Kuiper's tests and the bivariate KS test. However, the issues shown for this approximation with the univariate Kuiper's test may also extend to the bivariate case. Without approximations to the asymptotic behaviour of the proposed test statistics readily available, this study focusses on the Monte Carlo estimates of the p -values.

5.3.1.3 Simulation Procedure

For the bivariate KS test, a symmetric bivariate Normal (BN) distribution was simulated and tested using the procedure defined in Section 5.2.1.3. The baseline distribution was given by $BN \left(\begin{bmatrix} 8 \\ 8 \end{bmatrix}, \begin{bmatrix} 3 & 0 \\ 0 & 3 \end{bmatrix} \right)$, with an error distribution given by $N(0, 0.001)$. Here, zero cross-correlation is assumed in order to simplify the simulated distributions. In addition, the analysis and modelling of spatial and temporal correlations between wind direction distributions is beyond the scope of this research. However, it is noted that these correlations may not be insignificant,

and analysis of such would make up a vital component of further study leading from this research.

To run the simulations of the bivariate KS test, code was adapted from the function `kstest2`, available in MATLAB® [2016]. Further details of this code are given in Appendix D. From the above discussion, and in parallel to the univariate KS test simulations, the following statistics were calculated using this code;

- $D_N^{(2)} = \max(D_N^{P1}, D_N^{P2}, D_N^{P3}, D_N^{P4})$,
- $Z_N^{(2)} = \sqrt{N}D_N^{(2)}$, and
- $P_Z \simeq 2 \exp(-2(Z_N^{(2)} - 0.5)^2)$.

Monte Carlo methods were also used to construct a null distribution of the test statistics $D_N^{(2)}$ and $Z_N^{(2)}$ over 1000 simulations, and the p -values for the observed statistics were then derived.

As for the bivariate KS test, a bivariate Normal distribution is simulated to understand the sensitivity of these proposed bivariate Kuiper's test statistics. A bivariate von Mises (*BvM*) distribution is also simulated to investigate the application of the statistics to bivariate circular distributions. To reduce the number of parameters in the bivariate von Mises distribution, the Sine model given by Singh et al. [2002] is used. By setting the covariance parameter, λ , to zero, this approximation reduces to the product of two univariate von Mises distributions [Mardia, 2013], and also produces a single symmetric mode similar to that simulated for the bivariate Normal distribution.

A bivariate von Mises distribution (Sine model) is simulated to extend the analysis of the univariate Kuiper's test; $BvM \left(\begin{bmatrix} \pi \\ \pi \end{bmatrix}, \begin{bmatrix} 5 & 0 \\ 0 & 5 \end{bmatrix} \right)$, with errors simulated as for the bivariate KS test. Analysis of the univariate Kuiper's test showed that the formulation of the test statistic handles circularity in the data well, so only one bivariate von Mises distribution is simulated.

To run the simulations of the proposed bivariate test, code was adapted from the functions `circ_kuipertest` and `kstest2`, written in MATLAB® [2016]. Again, further details of this code is given in Appendix D. From the above discussion, and

in parallel to the univariate Kuiper's test simulations, the following statistics were calculated;

- $V_N^{(2)} = \max(V_N^{P1}, V_N^{P2}, V_N^{P3}, V_N^{P4})$,
- $k^{(2)} = mnV_N^{(2)}$, and
- $V_N^{(2)*} = \sqrt{N}V_N^{(2)}$.

Again, Monte Carlo methods were used to form a null distribution of each of the test statistics over 1000 simulations, and the p -values for the observed statistics were then derived.

As with each of the univariate tests, the power of the p -values against changes in mean and standard deviation of the bivariate Normal distribution, and mean and concentration of the bivariate von Mises distribution were calculated over 1000 Monte Carlo simulations. In the bivariate case, the mean was shifted independently in the x and y directions. This allows for comparison of power results with those given by the univariate KS test. The changes in standard deviation and concentration are again considered in relative terms, and were made symmetrically around the single mode of the simulated bivariate Normal and von Mises distributions.

Brief analysis (not presented here) of the sensitivity of the bivariate KS and Kuiper's tests to shifts in mean along the diagonals, i.e. combined symmetric x and y shifts, show very similar behaviours to those discussed in the following section.

5.3.1.4 Results

Figure 5.11(a) shows the sensitivity of the bivariate KS test to shifts in the bivariate mean. The lines on the plots indicate the power of the test as the mean of the bivariate Normal distribution was shifted through the y direction, while triangular points indicate that similar power behaviour was observed for shifts in the x direction. The sensitivity of the bivariate KS test remains symmetric to shifts in the mean, as in the univariate case. When considering the asymptotic p -value derived for the $Z_N^{(2)}$ statistic, the sensitivity behaviour is very similar to that shown in Figure 5.6(a). However, the power curve is less steep in the bivariate case, resulting in a region of marginal power between changes in mean of $\pm 10\%$ up to $\pm 80\%$ of the bin size.

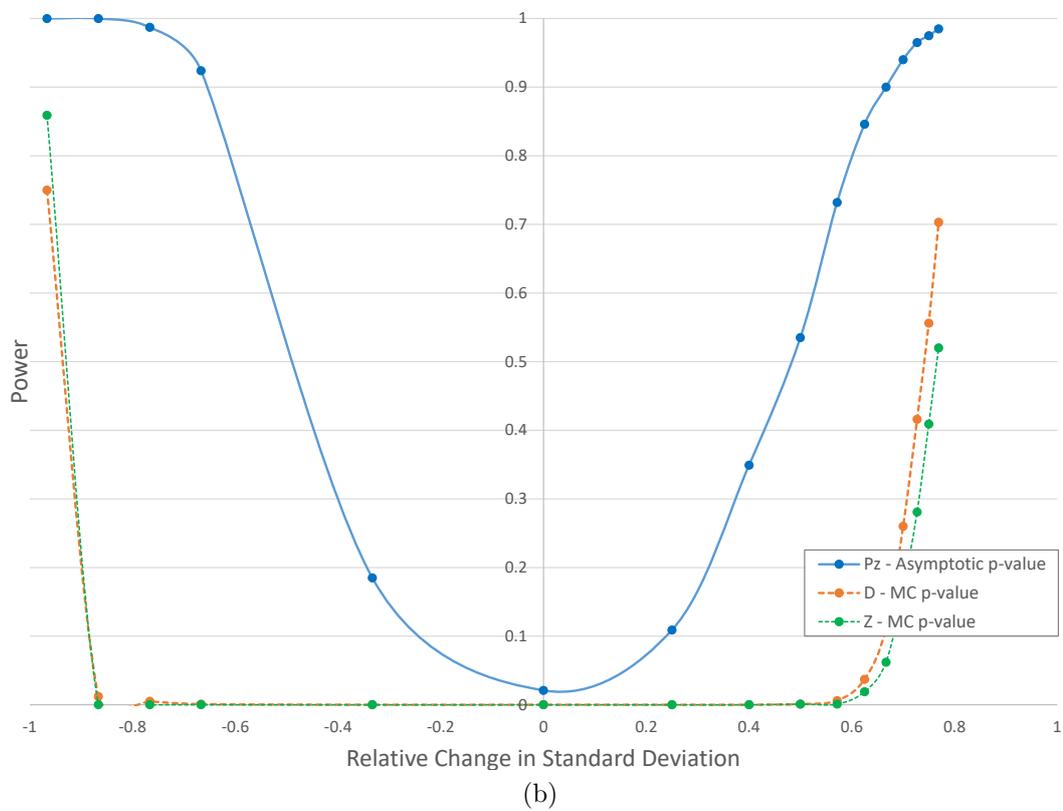
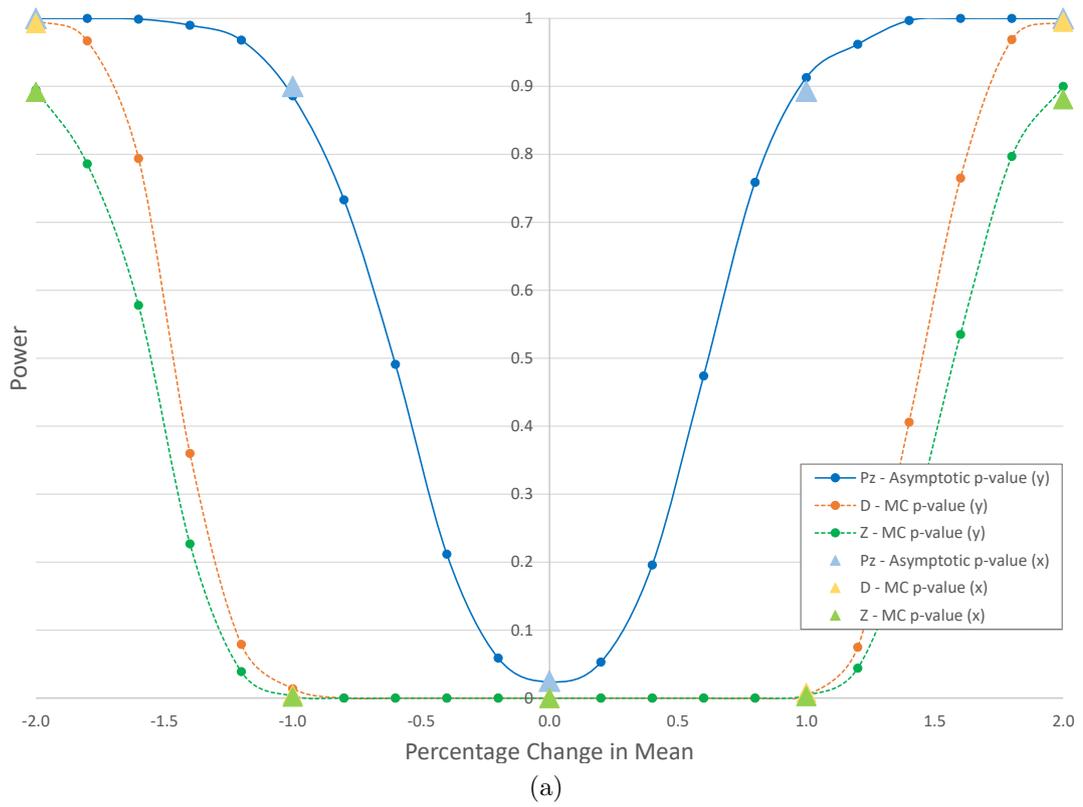


Figure 5.11: Power of the bivariate KS test to (a) percentage change in mean and (b) relative change in standard deviation for the bivariate Normal distribution, at a 1% level. Lines represent mean changes in the y direction. Triangles represent mean changes in the x direction.

Figure 5.11(b) shows the sensitivity of the bivariate KS test to relative changes in the standard deviation of the bivariate Normal distribution. As in Figure 5.6(b), the power curve is symmetric for the asymptotic approximation of the p -value for $Z_N^{(2)}$, but it is again considerably wider. This increased width results in a region of marginal power between a relative change in standard deviation of ± 0.1 and ± 0.6 .

The MC results again indicate that the $D_N^{(2)}$ statistic is less conservative than the scaled $Z_N^{(2)}$ statistic. However, as opposed to those shown for the univariate case in Figure 5.6, the MC estimates in the bivariate case are far more conservative than the asymptotic approximation, i.e. less sensitive to small changes in mean or standard deviation. In this case, the MC estimates show power levels close to (if not equal to) zero for shifts in mean values up to $\pm 100\%$ of the discrete bin size. The power curves for the MC estimates are then very steep, resulting in a powerful test (above 70%) for changes in mean of around $\pm 140\%$ to $\pm 150\%$.

The MC power curves for the relative changes in standard deviation are slightly asymmetric in the bivariate case, but the curves remain very steep. Regions of marginal power are reduced to relative changes of $+0.6$ to $+0.8$ and -0.8 to -1.0 . The asymmetry shown in these curves may develop through the structural constraints of changing the standard deviation (i.e. reductions can only be finite, but increases infinite), and these become more obvious in the bivariate case as perhaps more MC simulations are required to estimate the null distribution of either statistic.

Figure 5.12 shows the sensitivity of the proposed bivariate Kuiper's tests to changes in mean and standard deviation for the bivariate Normal distribution. Only MC estimations for p -values are calculated since no asymptotic approximations exist, but these estimations show very similar behaviour to the MC estimates for the bivariate KS test. The $k^{(2)}$ statistic is again shown to be the most conservative of the three statistics, and $V_N^{(2)*}$ is again shown to be more conservative than the unscaled $V_N^{(2)}$ statistic.

In addition, sensitivity is again symmetric around changes in the mean of the bivariate Normal distribution, with similar behaviours shown for shifts in mean along the x and y directions. All three statistics show very low powers (approximately

zero) up to shifts of $\pm 100\%$ in the mean. The power curves are then very steep, getting up to 70% power under shifts of $\pm 120\%$ or $\pm 140\%$, resulting in very small regions of potentially inconclusive test results. Slight asymmetry is again observed in the power of the tests against relative changes in standard deviation. However, the power curves are almost vertical, giving thresholds for significance at around +0.8 and -0.9 .

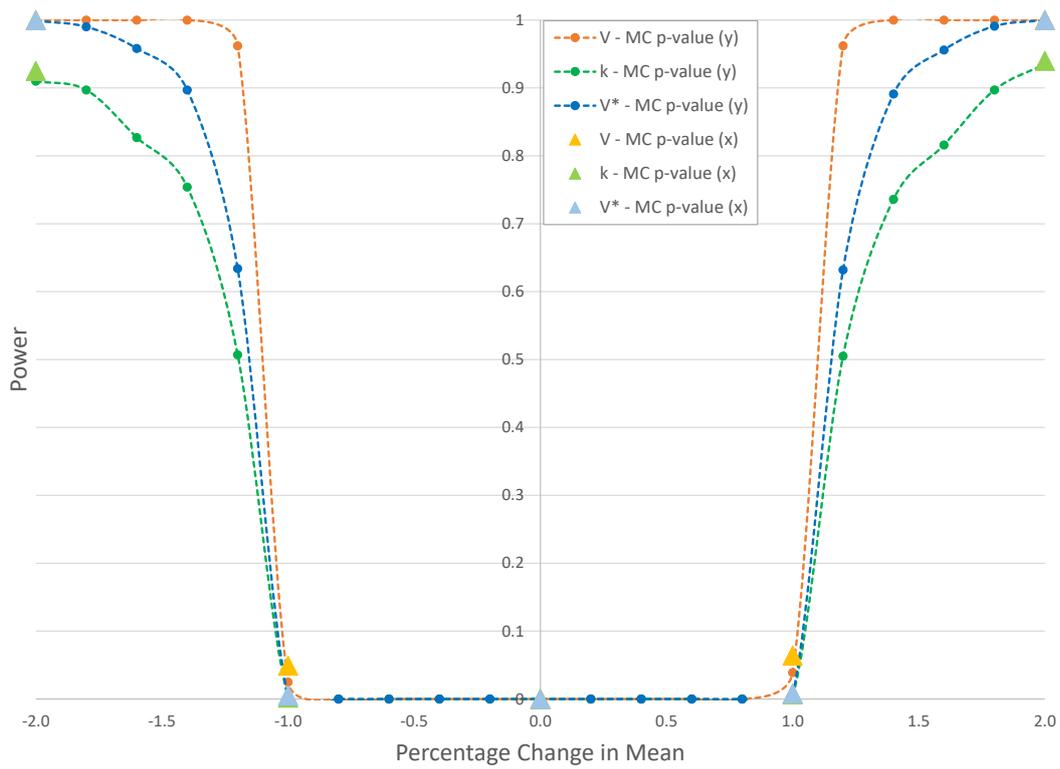
Figure 5.13 shows the sensitivity of the proposed bivariate Kuiper's test to changes in mean and concentration of the bivariate von Mises distribution. Although the behaviours are similar to those found by application to the bivariate Normal distribution, the power curves for the three tests are much closer to each other.

The asymmetry seen in Figure 5.12 is reversed in Figure 5.13 due to the inverse relationship between standard deviation for the Normal and concentration for the von Mises distributions. This provides further evidence that the asymmetry is an artefact of the parameter limits. Interestingly, further increases in concentration beyond approximately +0.9 appear to make no further change to sensitivity. This is due to the discretisation of the surface, meaning that further increases in concentration result in the same single grid cell of maximum probability. The variation shown in the results between increases of +0.9 and +1.0 suggest that further Monte Carlo simulations might be necessary in this bivariate case in order to better approximate the asymptotic behaviour of the test.

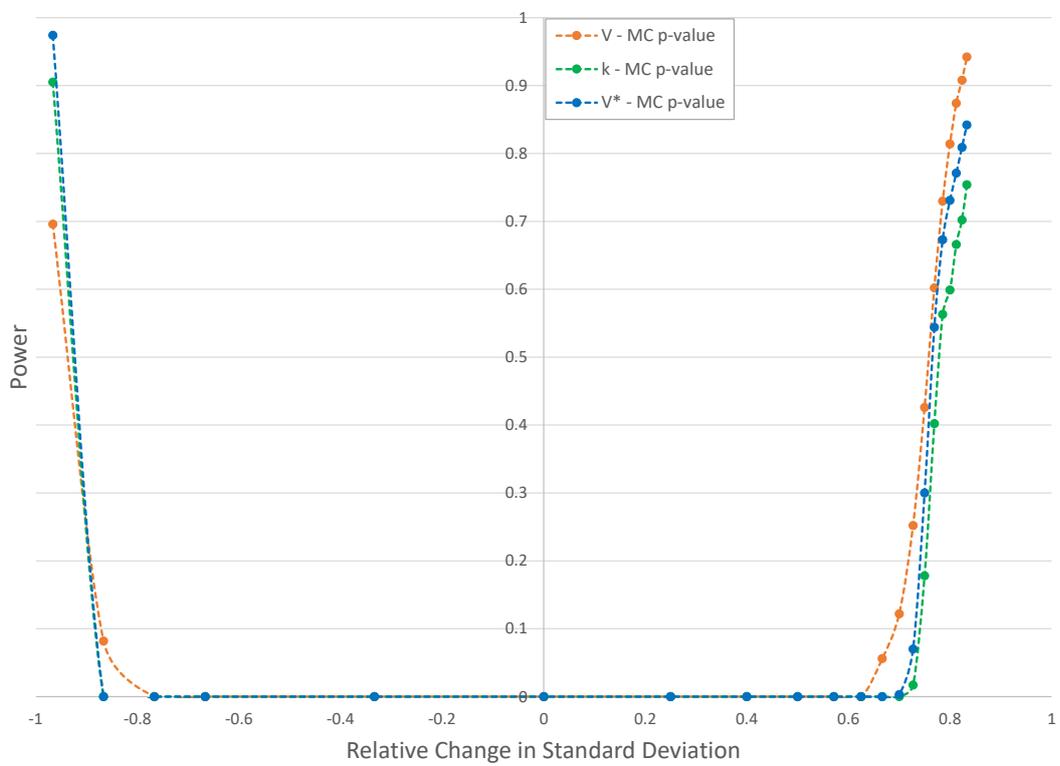
5.3.1.5 Discussion

The approximation to the asymptotic p -value given for the bivariate KS test behaves in a very similar way to that for the univariate KS test. This suggests that the extension of the KS test from the univariate version is indeed consistent, and both tests are comparable in terms of power for application. This begs the question of whether a similar extension can be applied to Kuiper's test.

For the proposed bivariate Kuiper's test, Figures 5.12 and 5.13 show very different behaviours to those shown for the univariate Kuiper's test in Figures 5.7, 5.9 and 5.10. However, the MC results are similar to those shown by the MC estimates for the bivariate KS test in Figure 5.11. For both sets of bivariate test statistics, the

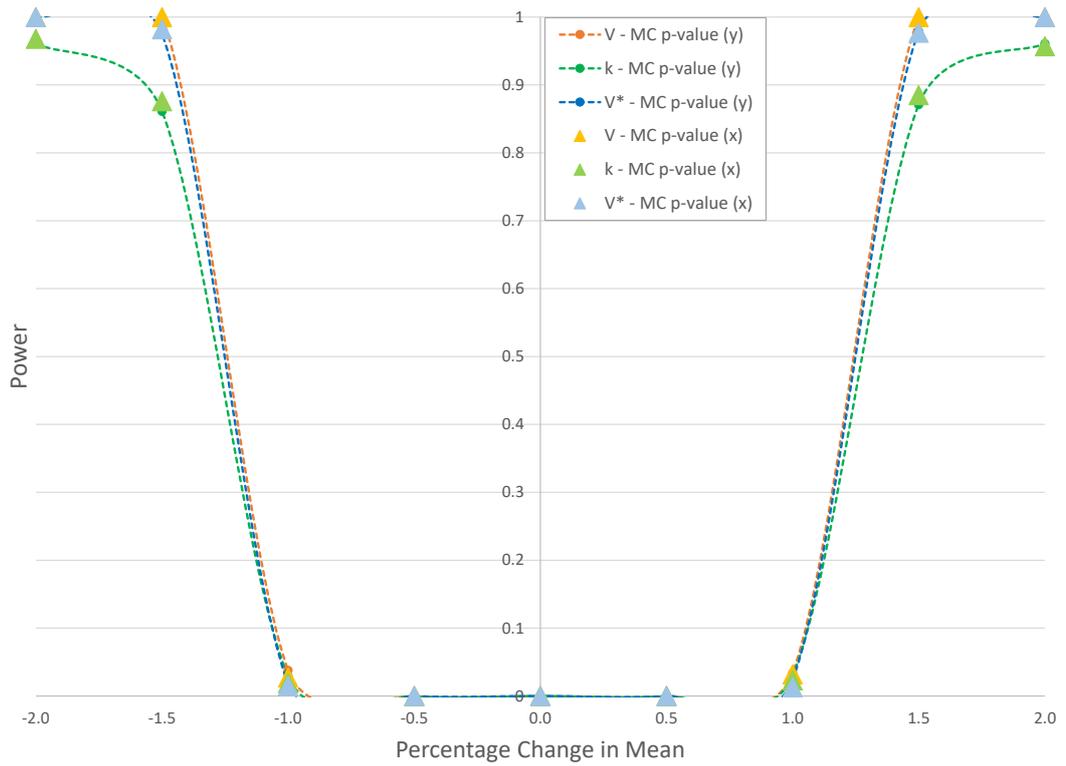


(a)

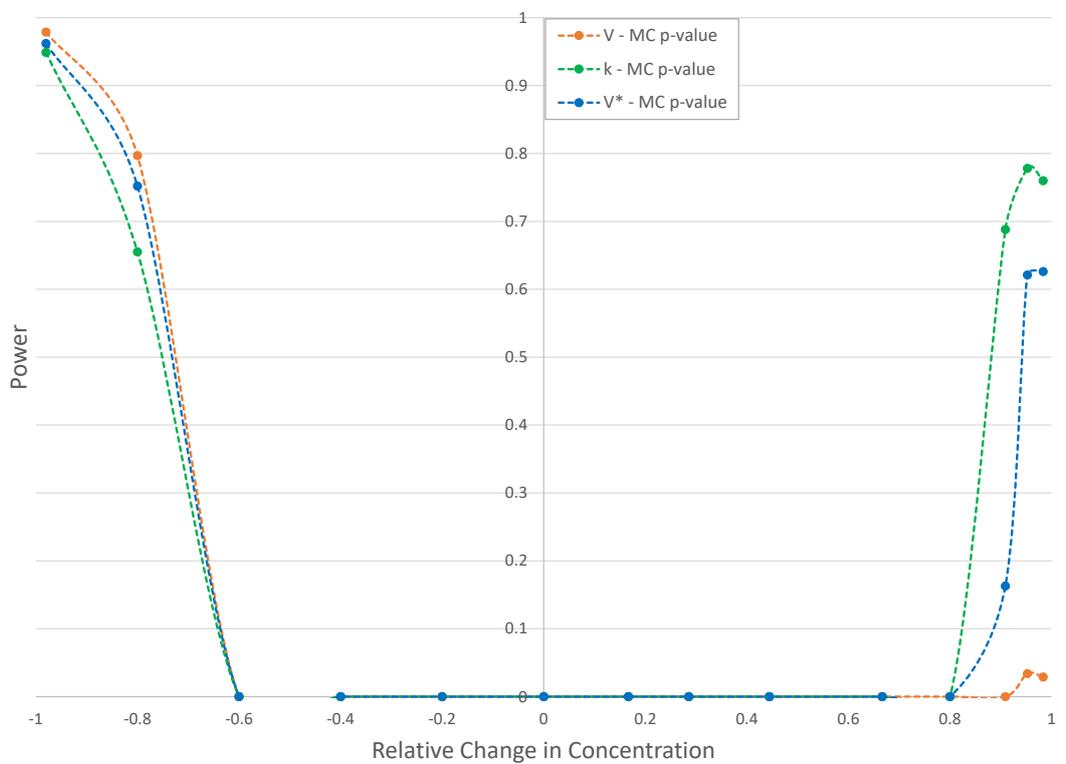


(b)

Figure 5.12: Power of the bivariate Kuiper's test to (a) percentage changes in mean and (b) relative changes in standard deviation for the bivariate Normal distribution, at a 1% level. Lines represent mean changes in the y direction. Triangles represent mean changes in the x direction.



(a)



(b)

Figure 5.13: Power of the bivariate Kuiper's test to (a) percentage changes in mean and (b) relative changes in concentration for the bivariate von Mises distribution, at a 1% level. Lines represent mean changes in the y direction. Triangles represent mean changes in the x direction.

MC p -value estimates are very conservative compared to the asymptotic p -value approximations. However, the power curves are very steep, resulting in narrow regions of inconclusive test results and definitive thresholds for significance.

The wide MC power curves observed for both bivariate tests may be explained in a number of ways. For instance, the increased dimensional complexity of the bivariate data is not explicitly taken into account in the bivariate extension of the tests and may result in the greater conservatism. It is conjectured that larger changes in mean and standard deviation (or concentration) are required to develop statistical evidence of a significant difference between two bivariate distributions. Equally, by considering the maximum statistics over the four EDF probabilities, the tests inherently account for larger differences between the distributions, making it more difficult to develop statistical evidence against the null hypothesis when distributions are similar.

As suggested in the previous section, the results may also not reflect the true behaviour of the bivariate test statistics, and simply more MC simulations may be required in the bivariate case to capture the added complexity. A brief investigation into increasing the MC runs ten-fold (to 10,000) appeared to have limited impacts on the sensitivity results. However, since the bivariate extension at least squares the complexity of the univariate problem, perhaps at least 100,000 MC simulations would be required to adequately capture the true behaviour of the test statistics.

The similarities in MC p -value sensitivity shown between the bivariate tests suggest that the proposed bivariate extension of Kuiper's test is a reasonable circular extension of the bivariate KS test. Unfortunately, however, more insight is required to understand whether the extension from the univariate Kuiper's test is valid. Further research is required to develop an approximation to the asymptotic behaviour of the proposed test statistics. It may be possible that an approximation based on the consistent exponential form derived for the KS tests would yield similar relationships between the univariate and bivariate cases, which would become an important consideration in application.

5.3.2 Joint Directional Wind Response Distributions

5.3.2.1 KS Style Tests

Table 5.6 shows the results of the bivariate KS test comparing the discrete observed directional wind response distributions from both case studies. Across Flea Creek Valley, all of the wind response pairs can be considered statistically different, with very small asymptotic p -values. As discussed previously, the MC p -value estimates are less conservative than the bivariate asymptotic approximations, and this is evident in Table 5.6.

As in the univariate case, the lowest $D_N^{(2)}$ and $Z_N^{(2)}$ values and largest asymptotic p -value are given by FCV2, which visually shows the most similarities between the two years (Figures 4.17 and 4.18). The MC p -value for the $Z_N^{(2)}$ statistic is also reasonably large, suggesting less evidence of significant difference between the directional wind responses observed in 2007 and 2014.

Table 5.6: Bivariate KS test results for directional wind response distribution pairs from FCV and the NAC.

	N	$D_N^{(2)}$	$Z_N^{(2)}$	$P_{Z_N^{(2)}}$	$P_{D_N^{(2)}}^{MC}$	$P_{Z_N^{(2)}}^{MC}$
Flea Creek Valley						
FCV1	1,333	0.3309	11.84	4.45×10^{-112}	0.0002	0.0002
FCV2	308	0.2931	5.02	3.79×10^{-18}	0.0001	0.1166
FCV3	1,071	0.4574	14.94	1.79×10^{-181}	0	0
FCV4	1,001	0.4617	14.57	2.10×10^{-172}	0	0
National Arboretum Canberra						
NAC1	27,494	0.6173	102.18	0	0	0.0001
NAC2	15,197	0.2936	36.19	0	0.3067	0.2923
NAC3	22,662	0.7461	112.24	0	0	0
NAC4	20,598	0.7148	102.53	0	0	0

For the comparison of directional wind response at the NAC, similar patterns are again shown between the bivariate and the univariate test results. All asymptotic p -values for the $Z_N^{(2)}$ statistic are found to be zero in this case, but the less conservative MC p -value estimates show very high values for NAC2. This strongly indicates that this pair are not significantly different, suggesting that the directional wind response observed at the NAC is the same on the leeward slope as on the 45° leeward slope

within the pines. Visual inspection of the distributions in Figure 4.19 confirms the similarities between the distributions.

Table 5.7 shows the results of the bivariate Kuiper’s test between directional wind response distributions, which are again consistent with the previous bivariate and univariate KS and Kuiper’s test results. Across Flea Creek Valley, the valley floor pair (FCV2) shows the smallest statistic values, and gives a very large MC p -value for the $k^{(2)}$ statistic, suggesting that the pair of distributions are not significantly different. Similarly, a large p -value is found for NAC2 using $k^{(2)}$, alongside small values for all three test statistics, again suggesting that this is the only pair of the four to not show evidence of significant difference. As consistent with Table 5.6, non-zero p -values for the $k^{(2)}$ statistic are also shown for FCV1 and NAC1.

Despite the conclusive results shown for the $k^{(2)}$ statistic, the p -values for both $V_N^{(2)}$ and $V_N^{(2)*}$ are all zero. In analysing the sensitivity of these newly proposed statistics, it was found in Section 5.3.1 that $k^{(2)}$ was the most conservative and so least likely to determine two distributions as significantly different, thus concurring with the results shown in Table 5.7.

Table 5.7: Bivariate Kuiper’s test results for directional wind response distribution pairs from FCV and the NAC.

	$V_N^{(2)}$	$V_N^{(2)*}$	$k^{(2)}$	$P_{V_N^{(2)}}^{MC}$	$P_{V_N^{(2)*}}^{MC}$	$P_{k^{(2)}}^{MC}$
Flea Creek Valley						
FCV1	0.5823	2.7525	1171	0	0	0.001
FCV2	0.5204	1.1767	147	0	0	0.6423
FCV3	0.7333	3.1637	1100	0	0	0
FCV4	0.6728	2.8048	818	0	0	0
National Arboretum Canberra						
NAC1	1.0959	23.9658	1.88×10^6	0	0	0.0324
NAC2	0.5042	8.2086	1.42×10^5	0	0	0.1399
NAC3	1.4740	29.2914	1.08×10^6	0	0	0
NAC4	1.4081	26.6852	9.07×10^5	0	0	0

5.3.2.2 Non-parametric Surface Comparison Tests

To compare the estimated continuous distributions, this study also considers two non-parametric statistics based on the squared difference between the two estimated

surfaces, $\hat{f}_1(X_{st})$ and $\hat{f}_2(X_{st})$, where $X_{st} = \{(\phi_s, \psi_t) : s, t = 1, \dots, 360\}$. In the estimation of the quasi-continuous surface, each point on the realised 360×360 grid had a calculated proportion value such that the entire surface sums to unity. The grid points were therefore taken to be the data points in these comparison tests. In order to compare two estimated surfaces, the biases in each estimate should be controlled for. In this research, this is done through fixing the smoothing parameter of the thin-plate-smoothing spline used to estimate the quasi-continuous surfaces $\hat{f}_1(X_{st})$ and $\hat{f}_2(X_{st})$.

The first test statistic is defined by Wang and Ye [2010], as the mean squared difference between the two surfaces;

$$T_W = \frac{1}{c^2} \sum_{s,t}^c \left\{ \hat{f}_1(X_{st}) - \hat{f}_2(X_{st}) \right\}^2, \quad (5.21)$$

where $c = 360$, the number of quasi-continuous data points in each dimension.

Under the null hypothesis, H_0 , it is assumed the two surfaces, $\hat{f}_1(X_{st})$ and $\hat{f}_2(X_{st})$, are drawn from the same distribution. Under this assumption, a third smooth surface, $\hat{f}(X_{st})$, can be defined as the surface estimated using both datasets combined. In practice, this is constructed by combining the proportions of data at each point on the discrete 16×16 grid, then estimating the continuous combined surface, $f(X_{st})$. The second statistic, developed by Bowman [2006], is analogous to analysis of variance by forming a function of the squared difference between each individual smooth surface and this combined null surface,

$$T_B = \frac{1}{\nu^2} \sum_k^c \sum_{s,t}^c \left\{ \hat{f}_k(X_{st}) - \hat{f}(X_{st}) \right\}^2, \quad (5.22)$$

with $\nu^2 = \frac{1}{d^2} \sum_k \sum_{i,j}^d \left\{ Y_k(X_{ij}) - \hat{f}(X_{ij}) \right\}^2$, where $d = 16$ is the number of observed data points in each dimension, thus scaling out the effect of error variance.

Establishing theoretical limit distributions of these statistics under the null hypothesis has been shown to be difficult and slow [Wang and Ye, 2010], and no asymptotic results are available. Therefore an empirical approach is taken by using a bootstrap algorithm to produce a Monte Carlo style scheme similar to those used for the KS

and Kuiper’s tests. The bootstrap procedure is given in Appendix D. In order to form a distribution of the test statistics T_W and T_B under the null hypothesis, the two observed datasets are combined under the assumption that both samples are drawn from the same population. The bootstrap is performed on the residuals taken between each of the two individual samples and the estimated quasi-continuous surface from the combined dataset. These pairs of bootstrapped residuals are added onto this combined surface to generate pairs of bootstrapped samples. The sample pairs are then used to calculate the two test statistics under the null hypothesis.

For this study, the application of the above bivariate statistical comparison tests was conducted in MATLAB® [2016] with code developed using, amongst others, the `bootstrp` function. Further details can be found in Appendix D.

Table 5.8 shows the MC p -value results of the T_W and T_B statistical comparison tests applied to the estimated quasi-continuous directional wind response surfaces for FCV and NAC. Table 5.8 shows a third estimation method labelled ‘Truncated TPS’. By defining a cut-off across the TPS surfaces such that very small values were assumed zero, potential bias in the test statistics could be mitigated against. The cut-off across the TPS surfaces was taken to be 0.00005; any value smaller than this is forced to zero. This cut-off was chosen through consideration of the average surface value, and using the visual surfaces as a guide.

Table 5.8: Bivariate squared difference test results (p -values) for directional wind response distributions pairs from FCV and the NAC.

	Cubic		TPS		Truncated TPS	
	T_W	T_B	T_W	T_B	T_W	T_B
Flea Creek Valley						
FCV1	0.946	0.825	0.235	0.521	0.013	0.165
FCV2	0.996	0.946	0.865	0.920	0.272	0.465
FCV3	0.540	0.576	0.005	0.148	0	0.010
FCV4	0.567	0.612	0	0.119	0	0.037
National Arboretum Canberra						
NAC1	0.415	0.542	0	0.044	0	0.044
NAC2	1	1	1	0.982	1	0.992
NAC3	0.201	0.422	0	0.050	0	0.065
NAC4	1	1	0.999	0.988	0	0.035

Of note is that, across all estimation methods and both tests, the pairs FCV2 and NAC2 give the largest p -values. This suggests that despite the statistical techniques,

the least evidence against the null hypothesis is developed for these pairs, and so the distributions are not considered significantly different. This is consistent with the results shown for the bivariate KS and Kuiper's test analysed in the previous section. However, aside from this, Table 5.8 suggests that the applied statistical techniques can have considerable impacts on the results.

Under cubic estimation, all comparison pairs across both case studies are found to be not significantly different for both the T_W and T_B test statistics. Cubic approximation exactly interpolates the data points without allowing for noise, and so does not produce a smooth surface. When comparing two surfaces, it is difficult for both statistical tests based on the squared difference between the estimated surfaces to build statistical evidence against equality. This is because all distributions constructed under the null hypothesis will exhibit high levels of variation across the entire surface, and so distinct modal patterns are difficult to identify.

When the TPS spline is applied to the directional wind response distributions, it is possible to differentiate between the different comparisons and infer more complex relationships. At Flea Creek Valley, FCV1 and FCV2 continue to show very high p -values for both tests, thus continuing to suggest no significant difference between the estimated directional wind response between the two years. For FCV3 and FCV4, the p -values for the T_B test remain relatively high and so again suggest no significant difference between the surfaces, but the T_W statistic gives small p -values which suggest that the pairs are significantly different between the sample years.

This contradictory result highlights the differences between the construction of the test statistics. By taking the difference between each surface and the combined surface, the T_B statistic appears to be less sensitive to changes in distributions as peaks are averaged out in the calculation of the combined surface. This makes it harder for the test to develop statistical evidence against the null hypothesis. This is also evident in the NAC results, where p -values for T_B are larger than those for T_W for NAC1 and NAC3, where the T_W results suggest that the surfaces are significantly different.

When the TPS spline approximation is truncated, differences between the surfaces become more obvious so both T_B and T_W are able to identify shifts in individual

modes across the distributions. In this form, the results for both bivariate tests agree with those shown using the bivariate and univariate KS style tests on the discrete data. In this case, FCV2 and NAC2 stand out as the pairs with the highest p -values suggesting no significant difference between the distributions. However, it is still evident that T_B is less sensitive than T_W . At the 1% significance level, all other pairs are still shown to be significantly different using T_B , while for T_W FCV1 gives a p -value slightly above the threshold.

5.3.2.3 Discussion

Across Tables 5.6 and 5.7, only the wind direction pairs of FCV2 and NAC2 were consistently shown not to have sufficient evidence to support the rejection of equality. As in the univariate case, this suggests that directional response distributions across most of FCV and the NAC have changed significantly due to the vegetation regrowth and changing vegetation and topographical conditions. The test results from the NAC, as well as visual inspection of the directional wind response distributions in Figure 4.19, again reinforce the existence of vegetation and aspect thresholds for the generation of wind reversal regions. This research indicates that these wind reversals persist beneath the canopy where in fact more turbulent behaviours would be expected to dominate wind flow [Finnigan, 2000, Belcher et al., 2012].

In contrast to the univariate case, it was shown in Section 5.3.1, that the MC p -value estimates for the bivariate KS style tests are considerably more conservative. This inverse feature of the bivariate tests is evident in the results shown in Tables 5.6 and 5.7 where the asymptotic p -values are zero or very small compared to those given by the MC estimates. Table 5.7 also highlights that the $k^{(2)}$ statistic is expected to be more conservative than the other proposed bivariate Kuiper statistics, again concurring with the findings in Section 5.3.1.

For the bivariate mean squared difference based statistical tests applied to the estimated quasi-continuous directional wind response distributions, the surface approximation method had a clear effect on the test results. Under cubic approximation, all distributions could not be considered significantly different due to the high level

of variation across the interpolated surfaces. Although very computationally efficient, it is unlikely that this method would be utilised to estimate directional wind response due to its susceptibility to noise, highly likely to be present in wind data.

The TPS spline provided much smoother and more reasonable approximations to the true underlying directional wind response shown in Figures 4.17, 4.18 and 4.19, and allowed a better understanding of the impacts of using different test statistics in the comparisons shown in Table 5.8. The T_B statistic was perhaps not sensitive enough to identify shifts in wind direction modes since they were averaged out in the comparison of each surface with the combined dataset. However, using the TPS suggested that FCV1 and FCV2, as well as NAC2 and NAC4 were not significantly different. At FCV1, wind reversal modes were dominant in both years which suggests that the increased vegetation in the area has not changed the generation of wind reversals on the leeward slope. For NAC4, the test result is less convincing against the visual representation of the distributions (Figure 4.19) since the off-diagonal mode present on the 45° leeward slope appears to move back to the diagonal on the cross slope within the pines.

Finally, when the TPS is truncated to avoid bias toward small estimation values, the T_W results agree with those given by the KS style tests, while the T_B p -values remain considerably larger. Using the truncation allows the tests to distinguish between individual modes and identify when they have shifted. As discussed earlier, the construction of the T_W statistic allows this test to be more sensitive to such changes.

5.4 Concluding Remarks

To better interpret the results of non-parametric statistical comparisons, it is important to understand the power of the available tests against a variety of alternatives. By considering the power of the KS and Kuiper's tests against changes in mean and variance, or concentration, of both the Normal and von Mises distributions, it was shown that both tests exhibit steep power curves under all conditions. However, thresholds for significance determined from these power curves were dependent on

the form of the test statistic and the method of determining significance. In practice, this allows researchers to ensure the appropriate application of either the KS or Kuiper's test. Regions of marginal power, resulting in inconclusive test results, are defined in terms of percentage changes in mean and relative changes in spread. Although it may be difficult to determine when these conditions occur, practitioners should avoid applying the tests under such circumstances.

For bivariate data, the extended KS test is also shown to be powerful against changes in mean and standard deviation above defined thresholds. Although the test becomes more conservative using Monte Carlo methods to approximate the p -value of the test statistic, it does provide a much more conclusive threshold for significance than the asymptotic approximation. Similar behaviour is observed for a proposed bivariate extension of Kuiper's test, suggesting that the proposition is a reasonable extension of the bivariate KS test to account for circularity.

Further research is required to develop an understanding of the asymptotic behaviour of the proposed bivariate Kuiper's test, but a caution is provided in the analysis of the univariate version. The power curve of the univariate Kuiper's test revealed extremely high false positive rates when comparing identical distributions using an asymptotic approximation of the test statistics' behaviour. Although this behaviour occurred outside the specified region of validity for this approximation, similar approximations in the bivariate case may also suffer from the same issue.

Application of these statistical comparison tests suggests that both conditional univariate wind direction distributions and bivariate directional wind response distributions were clearly impacted by vegetation and topography across both case studies. Vegetation regrowth had a significant impact at FCV, but visually similar behaviours could still be identified. On the valley floor, where there was considerable variation in wind direction, perhaps caused by persistent within-canopy turbulence or the impacts of local topography, the vegetation regrowth appeared to have limited impact on the distribution of wind direction. At the NAC, the introduced vegetation in the pine stand had a clear and structural impact on wind direction, while changing topography within the pine stand also impacted the distribution of

wind direction. The shift in the wind direction mode experienced under WNW prevailing winds clearly indicated that vegetation and aspect were important factors in the generation of recirculation or wind reversal regions within the vegetation.

Given these findings there is a clear need to better characterise the spatial variation in wind direction distributions. The empirical identification of conditions for generation of different wind mechanisms will play a significant role in progressing towards a parametric model for wind direction distributions that can be used to better incorporate the spatial variability of the inputs in fire spread models, particularly across complex landscapes.

The aim of this chapter is to compare non-parametric statistical tests, and so the results of each test discussed in the chapter are considered independently. However, the tests may be applied simultaneously to each set of wind direction data (univariate discrete, bivariate discrete and bivariate quasi-continuous). When applying more than one test to the same set of data it is easier to build evidence to reject the null hypothesis, and so a statistical correction must be applied to the significance level. One such technique is the Bonferroni correction which divides the desired significance level by the number of tests applied [Bonferroni, 1936, Dunn, 1959, 1961]. For instance, for the univariate and bivariate discrete datasets, both the KS and Kuiper tests are applied so an adjusted significance level of $0.01/2 = 0.0050$ can be used. For the quasi-continuous estimated surfaces, two non-parametric tests are applied, each under 3 different estimation methods, and so the adjusted significance level of $0.01/6 = 0.0017$ may be used. In all three cases, the application of a Bonferroni correction does not alter the discussions arising from the results in this chapter. In fact, for some results, the correction strengthens the conclusions detailed above.

There are many areas that warrant further investigation, including analysis of the statistical methods used to estimate and analyse wind direction distributions, and consideration of which methods are most practically useful in the context of wind and fire modelling. Further data collection would also be valuable in quantifying features such as surface roughness across these case studies (and new ones), leading towards parametrisation of wind models using physical and observable parameters.

The results and conclusions of these simulation studies are also limited to understanding the power of KS style tests under changes to unimodal distributions. Brief investigations using idealised bimodal distributions were considered by varying the location and spread of individual modes. Resulting power curves showed very similar behaviours to those discussed in this chapter. Further development of this research would consider the sensitivity of KS style tests to changes in more complicated distribution structures, such as simultaneous changes to multiple modes or simultaneous changes to mean and variance, or concentration.

CHAPTER 6

Evaluation of Wind Field Modelling over Complex Terrain

For effective modelling of fire spread, input variables need to be accurately represented across a range of relevant spatial scales. This is particularly true for wind. Although it is known that broad-scale weather interactions have considerable impacts on the spread of large bushfires through mechanisms such as synoptic wind changes or the formation of thunderstorms, there are numerous studies on fire behaviours driven by phenomena at scales of only 10s to 100s of metres within the landscape and close to the surface [e.g. Sharples et al., 2012, Simpson et al., 2013]. With emerging research into the dynamics of extreme fire behaviour and the variability of fire spread, it is increasingly important for wind models used in operational fire prediction to capture areas of complex flow across rugged terrain. Typically, wind for bushfire modelling is realised at a height of 10m or 20ft. Within the context of bushfire modelling uncertainty discussed in Chapter 2, an evaluation of weather predictions used for current operational models is vital to understanding the uncertainty of model outputs used within the fire management process. However, meteorological datasets observed at spatial resolutions relevant to extreme fire behaviour, particularly near-surface winds, are rare.

This chapter first considers the evaluation of a diagnostic wind model that is currently used operationally within state-of-the-art fire modelling frameworks across Australia and the USA. In contrast to previous evaluation studies, this research considers how the traditional deterministic approach to wind modelling might be improved through ensemble-style application, providing outputs that can be represented in probabilistic terms. The probabilistic understanding of wind fields

across complex terrain developed throughout this thesis can complement the current physics-based deterministic models. In particular, probabilistic representations of wind model outputs allow for a discussion of uncertainty in the context of ensemble-based fire prediction frameworks, leading to a better understanding of fire prediction uncertainty.

As wind direction modelling is analysed across complex landscapes, wind speed as a key driving force for surface fire spread cannot be ignored. In the current suite of operational fire spread models, wind speeds measured in the open environment (above the vegetation layer) are modified to represent wind speeds at ‘mid-flame’ height using adjustment factors. In general, these adjustment factors assume constant vertical wind speed profiles throughout the vegetation layer. However, empirical studies have shown that wind speeds beneath canopies can vary significantly with height above ground as well as with forest type and prevailing wind speed.

The second section of this chapter considers the variability of wind speed within vegetation across complex terrain. This study uses the data collected for this research over complex terrain to evaluate empirical wind reduction profiles developed for a number of forest types in areas of flat terrain. Through better understanding of wind speed reduction beneath the canopy across landscapes from mountainous ranges through to flat plains, wind speed reduction models for bushfire spread prediction can be adapted to incorporate the variation observed in vertical wind speed profiles within the vegetation layer.

The research presented in this chapter has been summarised in the following journal article and conference proceedings;

Quill, R., Sharples, J.J., Sidhu, L.A., Wagenbrenner, N. and Forthofer, J. (In Prep) A wind direction evaluation of a diagnostic wind model over complex terrain in the context of ensemble-based fire spread modelling.

Quill, R., Moon, K., Sharples, J.J., Sidhu, L.A., Duff, T. and Tolhurst, K. (2016) Wind speed reduction induced by post-fire vegetation regrowth. Proceedings from the Bushfire and Natural Hazards CRC & AFAC conference, Brisbane, 30 Aug - 1 Sep 2016.

6.1 Predicted Wind Direction using a Diagnostic Wind Model

6.1.1 *Background*

As discussed in Chapter 2, mesoscale Numerical Weather Prediction (NWP) is accurate and operational for real-time weather prediction, but the horizontal resolutions (from 4km up to 25km) are not sufficient to resolve underlying terrain for surface fire behaviour in an operational context. In addition, Butler et al. [2015] highlight the limited nature of traditional broad-scale weather prediction in understanding the variability of wind fields across complex terrain; indicating the need for downscaling models to better predict meteorological variables at finer spatial resolutions. As an example of one such downscaling model, WindNinja is currently used operationally in Australia within the state-of-the-art fire modelling framework Phoenix Rapidfire [Forthofer et al., 2014b, Tolhurst et al., 2008]. As a diagnostic model, WindNinja is generally preferred for operations (as opposed to prognostic approaches of full computational fluid dynamics models) due to the computational constraints on producing useful, fine-scale wind model outputs in real or near-real time [Forthofer et al., 2014b].

In order to understand the uncertainty of these downscaling models, evaluation of their performance must be completed in comparison to observed wind dynamics. However, it is often noted throughout the literature [e.g. Butler et al., 2015] that there is a lack of wind data observed across complex terrain for effective evaluation of wind modelling approaches used in operational fire modelling. Key wind data sources have previously included Askervein Hill [e.g. Forthofer and Butler, 2007, Lopes, 2003], Tighvein Hill [e.g. Vosper et al., 2002, Burton et al., 2006] or Bolund Hill [e.g. Berg et al., 2011]. These data represent landscapes that, although real, are simplistic in their topographical features, e.g. flat plains surrounding a single hill with limited vegetation. While wind data have been collected from more complex terrain, i.e. within mountain ranges, to identify important landscape-scale wind flow features such as mechanical channelling through valleys [e.g. Gross and Wippermann, 1987, Whiteman and Doran, 1993, Weber and Kaufmann, 1998], data collected at finer scales relevant to fire behaviour and spread across mountainous landscapes are rare.

A visual comparison of predicted fire spread using output from a full computational fluid dynamics wind model with that from a mass-consistent wind field model showed the stark differences between the two models on the leeward side of Askervein Hill [Forthofer et al., 2014a]. By accounting for conservation of momentum in the wind field, the fire spread was predicted to significantly reduce in speed on the leeward slope, while using the mass-consistent wind model showed little reduction in rate of spread.

Further to this study, the authors conducted a comparison of mass and mass-and-momentum conserving wind field models with observed wind data from Askervein Hill and Waterworks Hill, the latter of which exemplified “slightly more complex topography” [Forthofer et al., 2014b, page 973]. Both cases exhibit limited vegetation cover across the study regions. The results of this comparison showed that although both the mass and mass-and-momentum conserving wind models accurately predicted wind speeds on windward slopes, both models showed the greatest limitations in the modelling of wind speed and direction on leeward slopes. In fact, Forthofer et al. [2014b] warned that users of such models should be aware of these limitations and interpret results cautiously where appropriate.

Butler et al. [2015] introduced a high-resolution dataset collected over yet more complex terrain but still limited to grasses and low-level brush vegetation. Wagenbrenner et al. [2016] compared the application of a previous version of WindNinja (version 2.5.2), a mass-conservation model, against these observations. The improvements downscaling models, such as WindNinja, make on broad-scale numerical weather prediction were highlighted, particularly in relation to the variability of wind speeds experienced across complex terrain [Wagenbrenner et al., 2016].

However, the study reiterated the limitations of the WindNinja 2.5.2 model in predicting wind speed and direction on leeward slopes. Since this scheme was designed to account for only mass-conservation across the landscape and not momentum conservation, which is the cause of features such as recirculation on leeward slopes, Wagenbrenner et al. [2016] noted that these limitations were to be expected. With characteristics such as flow separation on leeward slopes being linked to extreme fire behaviours [Sharples et al., 2012, Simpson et al., 2013] as well as being shown to

have significant impacts on fire propagation across the landscape, the consequences of not capturing such wind features are significant in the fire modelling context.

As discussed in Chapter 2, operational wind and fire models are often deterministic in their nature, while wind fields and fire spread are driven by a range of processes, many of which are beyond current physical modelling capabilities. To deal with the gaps between reality and model outputs, we can consider the underlying processes in terms of probabilistic approaches. While statistical approaches have limitations and rely upon previous system behaviours (which can include outliers), they are able to capture the inherent variability of wind and fire across the landscape and are better suited to emerging ensemble-based fire prediction frameworks.

Specifically, SABRE [Twomey and Sturgess, 2016] and FireDST [French et al., 2014d], incorporate probabilistic approaches to fire modelling through variation of input variables across distributional structures. However, these frameworks account for variability in input parameters in relatively simplistic terms, assuming standard distributional behaviour, such as Normal, Uniform or point processes, around a deterministically predicted mean value. Wind fields, as input to fire spread prediction, have been recast in a probabilistic light in a limited number of studies [e.g. Sharples et al., 2010]. Throughout these studies, and within the research presented in this thesis, the variability of wind direction at a given point is consistently shown to not be well represented by a Normal or Uniform distribution. Moreover, the structure of wind direction distributions can vary considerably across the landscape. However, there have been very few (if any) studies into the characterisation of this spatial wind direction variability and the parameterisation of such distributions using physical variables.

The aim of this study is to evaluate the diagnostic wind model WindNinja in the context of bushfire prediction over areas of complex terrain and vegetation using the data collected over Flea Creek Valley in 2014. By considering wind fields in probabilistic terms, this study also aims to advance discussions of uncertainty in wind and fire modelling, and progress towards spatial characterisation of wind direction distributions for input into ensemble-based fire prediction frameworks.

6.1.2 Wind Direction Data and Modelling

6.1.2.1 Winds over Flea Creek Valley in 2014

In this study, the 1-minute wind direction data collected between July and December 2014 across Flea Creek Valley are used to evaluate the diagnostic wind model, WindNinja. To account for the accuracy of the instrumentation, data points with observed wind speeds less than 0.4ms^{-1} were excluded from the analysis.

In Chapter 3, Figure 3.6 indicates the distribution of wind directions observed at the eleven weather stations across the valley. In summary, wind data collected from the ridge top stations on both the western and eastern sides of the valley (B1, B2 and B11) indicate the most frequently observed WNW prevailing wind direction. Less frequent easterly prevailing winds were also observed at all three sites. On the east-facing slopes of the valley (B3, B4 and B5), observed dominant easterly wind directions indicate the prevalence of wind reversals when these slopes were leeward to the WNW prevailing winds.

On the valley floor (B6 and B7), observed wind directions were more variable than at other sites but northerly and southerly modes suggest that channelling or thermal flows may have dominated wind movement through this region. Finally, on the west-facing slope (B8, B9 and B10), north-westerly winds indicate a northerly bias in the winds when this slope was windward to the WNW prevailing winds. Channelling indicated on the valley floor may have induced this effect on the eastern slopes, but local topographical features such as nearby gullies could have also caused this deviation from the prevailing wind direction.

Figure 6.1 shows the average hourly wind speed and direction for 0300h, 0900h, 1500h and 2100h at each site across the valley. As expected, wind speeds increased during the afternoon, with wind directions at the north-western ridge top station (B1) rotating from NE overnight (0300h) through to NW in the afternoon (1500h). This is echoed by a similar regional wind direction change observed at Canberra Airport. This change of direction was coupled with the increase in wind speed across the valley. However, the higher wind speeds were only felt along the valley

floor and on the west-facing (or windward) slope of the valley. The east-facing leeward slope winds remained relatively low during the afternoon period.

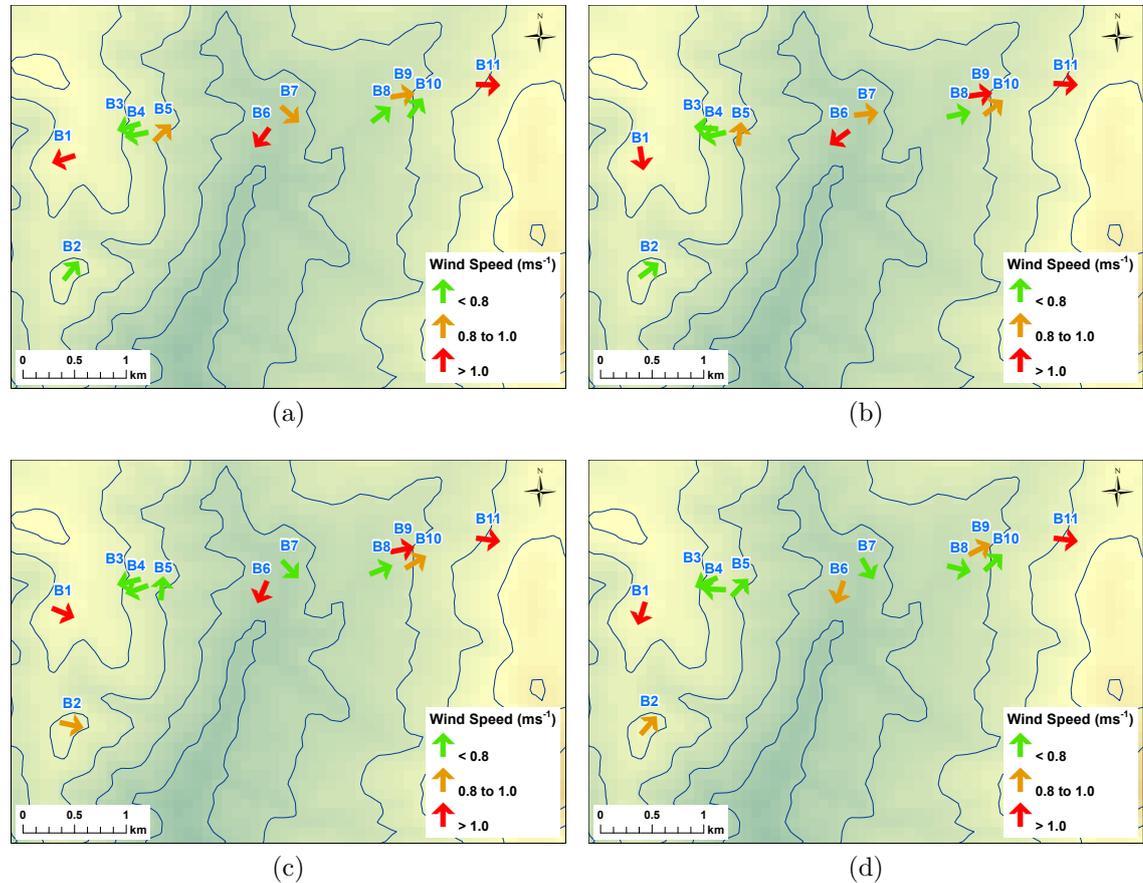


Figure 6.1: Observed average hourly wind directions at each site across FCV, at (a) 0300h, (b) 0900h, (c) 1500h and (d) 2100h.

Despite the almost 180° change in prevailing wind direction measured at B1, the remaining stations showed relatively stable wind directions. Each other station experienced consistent wind directions throughout the night and day, suggesting that diurnal effects had little impact on wind flow beneath the canopy across the valley. Most prominently on the valley floor (B6 and B7), northerly average flows agreed with the dominant northerly modes shown in Figure 3.6, with average northerlies experienced throughout the day and night. However, no strong average southerly wind directions were shown across the 24 hours. It appears that the southerly mode shown at B7 in Figure 3.6 manifested as more westerly wind directions around 0900h (Figure 6.1(b)). This suggests that channelling through the valley (forming the north and south modes shown in the wind rose for B7) was mechanically forced due

to frictional gradients occurring at short time scales, rather than thermal gradients induced as part of a diurnal cycle.

It is also clear that local topography had a significant impact on wind direction at specific sites. For instance, the south-western ridge top station (B2) consistently experienced SW winds, even when B1 (representative of the prevailing wind) recorded NE winds at 0300h. Also, the southerly winds experienced at B5 on the east-facing valley slope were likely due to the localised steep south-facing slope on which the station was situated. In this case, very localised mechanical flows may have been evident as the winds shifted to flow eastwards overnight.

On the east-facing slope of the valley, B3 and B4 showed consistent easterly wind directions up the slope of the valley wall. The easterlies experienced at 1500h, when the prevailing winds were westerly, indicate the existence of a recirculation region within the canopy, concurring with Figure 4.16 as well as analyses conducted by Sharples et al. [2010] which showed the prevalence of leeward slope eddies.

Finally, on the west-facing slopes (B8, B9, B10 and B11), consistent average westerly winds were observed throughout the day and night. This again concurs with the wind roses shown in Figure 4.16, but also shows that even when easterlies were experienced at B1 on the western ridge top, westerlies were recorded on the eastern slope. This identifies further potential for a recirculation region on the leeward west-facing slope under easterly prevailing winds, i.e. the converse to those shown at B3 and B4 on the east-facing slope under westerly winds at 1500h.

6.1.2.2 Traditional Deterministic Wind Modelling

The diagnostic wind model WindNinja was used to predict the wind field across Flea Creek Valley. Two solver options were used within the model; a mass conserving solver (packaged within WindNinja 2.5.2, referred to herein as the ‘native solver’), and a beta version of a mass-and-momentum conserving solver (later modified and released within WindNinja 3.0.0, referred to herein as the ‘momentum solver’). Input parameters for the model runs included a surface roughness set to ‘Trees’, input and output winds at 5 metres above the surface and 1ms^{-1} average domain wind speeds. Modelled wind directions were adjusted to align with observations

of WNW (298°) winds at B1 on the western ridge top. Using a domain averaged wind speed of 1ms^{-1} , modelled wind speeds at B1 were approximately 2ms^{-1} for WindNinja with both solvers.

The wind field prediction is defined at 5 metres above the clear ground or any vegetation layer, whereas wind observations were taken at 5 metres above the ground within the approximately 15-metre-high vegetation layer. To account for this, it is common to adjust wind speeds using wind reduction factors which are discussed later in this chapter [e.g. Andrews, 2012, Moon et al., In Press]. However, for fire modelling applications wind directions are not transformed beneath the canopy. Therefore, the wind directions predicted at 5 metres above the canopy are taken to indicate the wind directions predicted beneath the canopy.

Outputs from both models were analysed in ArcGIS [ESRI, 2011] to generate the 5-metre predicted wind fields shown in Figures 6.2 and 6.3. Table 6.1 shows the wind speed and wind direction outputs given by WindNinja with native solver and with momentum solver for each of the station sites. Using WindNinja with native solver (Figure 6.2), the predicted wind field was relatively smooth across the valley, maintaining a dominant WNW direction across both the leeward and windward slopes, as highlighted in the predictions given in Table 6.1. Wind speeds were highest across the western and eastern ridge tops, with very low speeds recorded on the valley floor.

Using the momentum solver (Figure 6.3), the predicted wind field showed some subtle variations. These variations were most prominent on the leeward slope where small lateral circulations were shown around the gully features near B2 and B5. Larger-scale channelling down the valley (shown by the northerly bias in predicted wind directions in Table 6.1) was strongly enhanced with the additional momentum conservation conditions. Wind speeds were again highest across the ridge tops, with the addition of some variations around small topographical features. In addition to the fastest winds observed at B1 and B2 on the western ridge top, wind speeds were shown to be much faster across the eastern windward slope around B10 and B11 than on the western leeward slope.

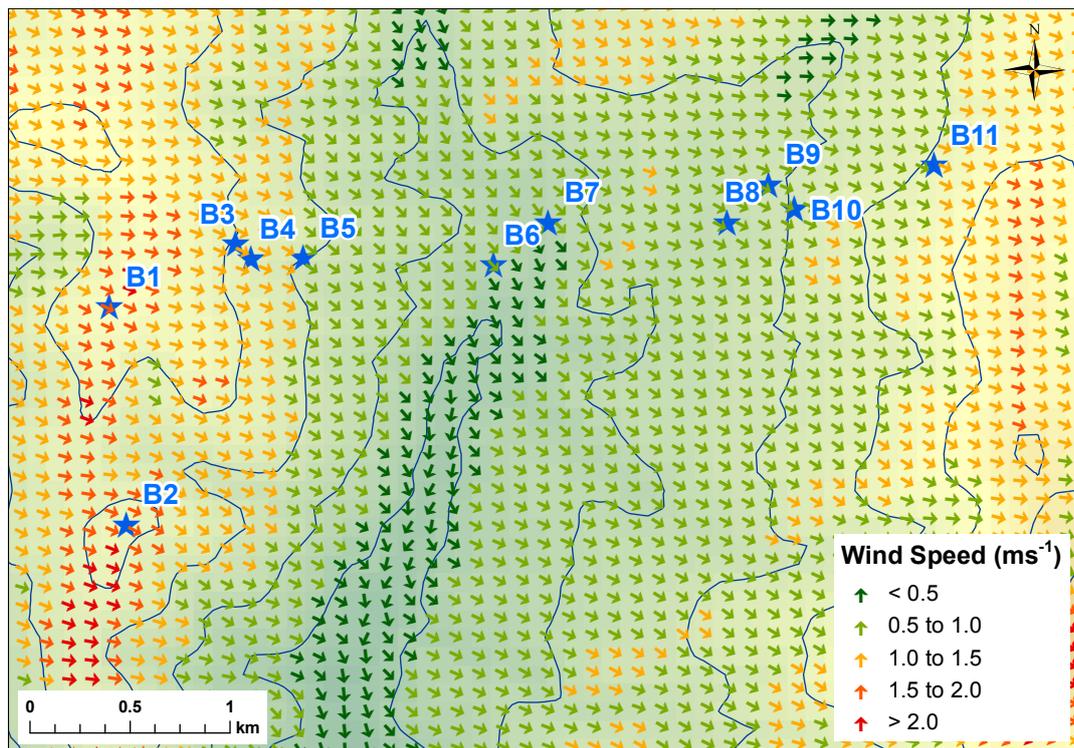


Figure 6.2: WindNinja with native solver prediction over FCV, using domain averaged wind speeds of 1ms^{-1} .

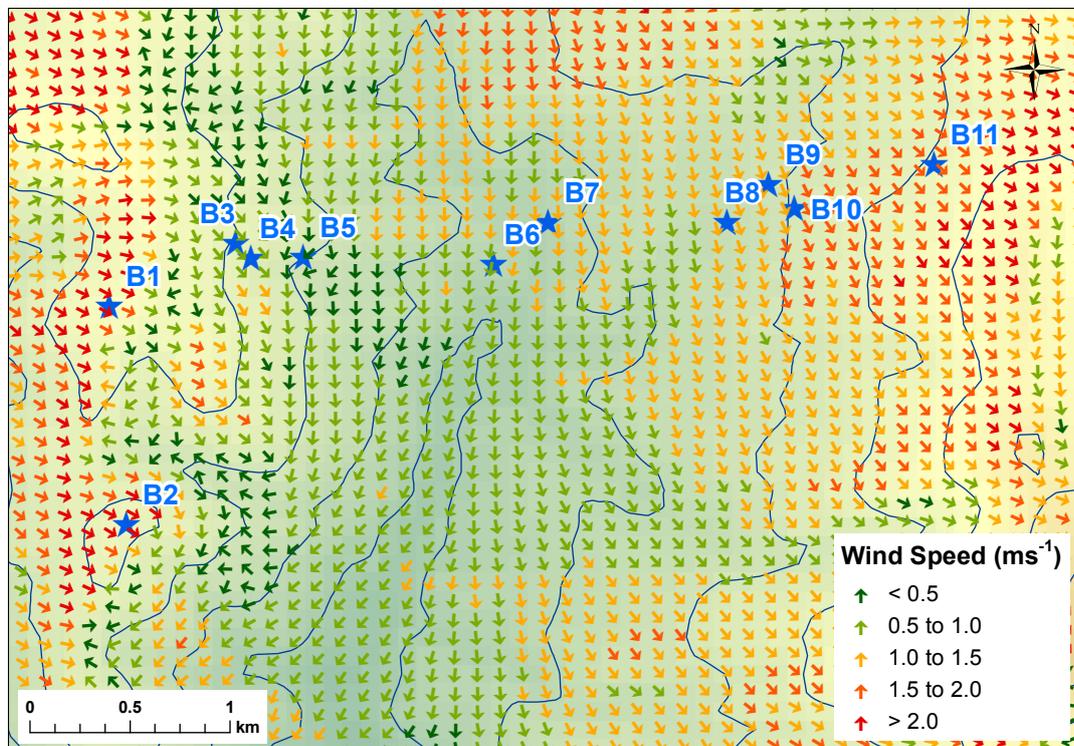


Figure 6.3: WindNinja with momentum solver prediction over FCV, using domain averaged wind speeds of 1ms^{-1} .

Table 6.1: WindNinja with native solver and WindNinja with momentum solver wind direction (WD, °) and wind speed (WSp, ms⁻¹) predictions for each site across FCV.

	Native Solver			Momentum Solver		
		WD	WSp	WD	WSp	
B1	298	WNW	1.99	298	WNW	2.28
B2	299	WNW	1.94	303	WNW	2.08
B3	298	WNW	1.25	316	NW	0.86
B4	298	WNW	1.25	317	NW	0.96
B5	312	NW	0.93	48	NE	0.07
B6	322	NW	0.60	9	N	0.97
B7	331	NNW	0.55	359	N	1.13
B8	309	WNW	0.54	344	NNW	1.16
B9	300	WNW	0.58	341	NW	1.28
B10	300	WNW	0.60	335	NNW	1.48
B11	297	WNW	1.03	313	NW	1.91

6.1.2.3 Ensemble Approach to Wind Modelling

For a comparison between observed wind direction distributions and a probabilistic representation of WindNinja, predicted unconditional wind direction distributions were constructed by running WindNinja with the momentum solver in an ensemble-style framework. Since the momentum solver forms the next generation of wind models used operationally with fire prediction, the native solver was not analysed within this framework.

Due to the deterministic nature of the model, it was run once for each of the wind directions relating to the 16 points of the compass to generate a look-up table for the modelled wind directions at each of the eleven sites across the valley given the wind direction at the ridge top station, B1. This results in a 16-by-11 look-up table, or wind library. Using the 1-minute data from B1 throughout the study period i.e. July to December 2014, this look-up table was cross referenced at each time point in order to construct expected frequency distributions of wind directions for each site across Flea Creek Valley. These predicted wind direction distributions were then compared to the observed unconditional wind direction distributions at each station site.

6.1.3 Evaluation Results

6.1.3.1 Deterministic Prediction

Figure 6.4 shows the observed conditional wind direction distributions at B4 on the east-facing slope, B7 on the valley floor, and B8 and B11 on the west-facing slope, for a prevailing wind speed threshold (T) of 0ms^{-1} and WNW wind direction measured at B1. Predictions from Table 6.1 are shown in red (native solver) and green (momentum solver). For all stations, Table 6.2 gives the proportion of the observed distributions that agree with each model prediction for increasing wind speed thresholds (T), which is taken to be the number of observations in the predicted segment divided by the total observations for the time period. In general, the percentage agreements are low due to the deterministic nature of the predictions; the single model outputs are not capable of capturing the variability of the distributions observed in Figure 6.4.

Table 6.2: Proportion of agreement between predicted wind direction (as compass point) and observed wind direction distribution at each site across FCV, conditional on observing a wind direction of WNW, and a minimum wind speed of T at B1.

T		Native Solver				Momentum Solver		
		0.4ms^{-1}	2ms^{-1}	4ms^{-1}		0.4ms^{-1}	2ms^{-1}	4ms^{-1}
B1	WNW	–	–	–	WNW	–	–	–
B2	WNW	0.1892	0.2070	0.1561	WNW	0.1892	0.2070	0.1561
B3	WNW	0.0321	0.0177	0.0210	NW	0.0229	0.0143	0.0126
B4	WNW	0.0173	0.0134	0.0095	NW	0.0370	0.0165	0.0143
B5	WNW	0.0228	0.0123	0.0000	NE	0.0718	0.0498	0.0141
B6	NW	0.0932	0.0566	0.0462	N	0.0850	0.0657	0.0838
B7	NW	0.1348	0.1161	0.0898	N	0.1802	0.1550	0.0898
B8	NNW	0.0504	0.0447	0.0120	NNW	0.0394	0.0358	0.0000
B9	WNW	0.2007	0.1830	0.1475	NW	0.0343	0.0140	0.0164
B10	WNW	0.0633	0.0792	0.0561	NNW	0.0333	0.0298	0.0374
B11	WNW	0.1561	0.1747	0.1754	NW	0.0253	0.0283	0.0994

The highest agreements, for model predictions with either solver, are found on the ridge tops (B2 and B11) and valley floor (B6 and B7) where the models are able to capture the dominant wind direction modes of the broader scale wind field. On the western ridge top (B2), there is no difference between the predictions from both models, whereas on the eastern ridge top (B11) and valley floor (B6 and B7), the momentum solver prediction shows a bias towards northerly winds. This bias has little impact on the percentage agreements observed on the valley floor, but the

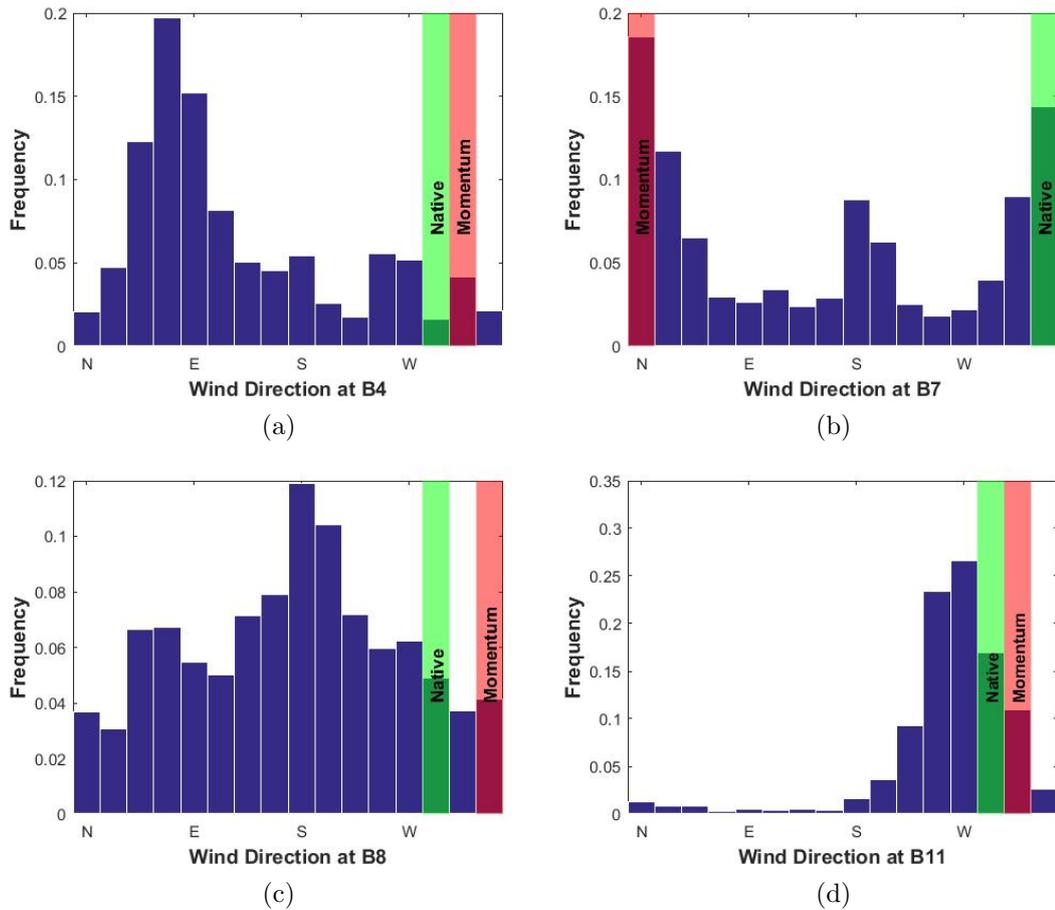


Figure 6.4: Observed conditional wind direction distributions at (a) B4 on the east-facing slope, (b) B7 on the valley floor, (c) B8 on the lower west-facing slope and (d) B11 on the upper west-facing slope, assuming a WNW observed at B1. Predictions shown for the native solver (red) and the momentum solver (green).

agreement values between observation and prediction at B11 are much lower for the momentum solver than for the native solver at all wind speed thresholds, i.e. 0.0253 at $T = 0\text{ms}^{-1}$ as opposed to 0.1561 for the native solver. The highest individual agreement values are found at B2 with $T = 2\text{ms}^{-1}$ where the percentage agreement reaches above 20%.

For many of the sites (e.g. from B2 to B9), across both the native and momentum solver, the agreement reduces as the wind speed threshold increases. This suggests that both versions of the model are less capable of predicting wind direction behaviour at higher prevailing wind speeds. The observed wind direction distributions (shown in Appendix A) indicate that the variance of dominant modes decreases as wind speed thresholds increase, resulting in a lower percentage agreement if the model does not accurately predict the key mode. However, due to the

relatively low wind speeds experienced throughout the study period the highest wind speed thresholds also have smaller sample sizes to construct the distributions for comparison, somewhat reducing the reliability of subsequent conclusions. Further model runs with higher domain-averaged wind speed might also indicate different behaviours when the momentum solver is used (it was found that increased wind speeds under the native solver had little impact on predicted wind direction).

On the western wall of the valley, leeward to the WNW prevailing winds, neither model is able to capture the easterly winds observed. As seen in Figures 6.2 and 6.3, the model with either solver predicts predominantly westerly flows across the entire valley when the prevailing winds are from the WNW. The observations at B3 and B4 clearly show dominant easterly modes at these stations (e.g. Figure 6.4(a)), suggesting the existence of recirculation within the vegetation on the leeward slope. The discrepancies between predictions and observations result in extremely low agreement percentages of 3.2% or less for the native solver, and 3.7% or less for the momentum solver. With the concentration of modes (or reduction in variance) as prevailing wind speeds increase, the percentage agreements drop to below 1% and 1.5% for the native and momentum solvers, respectively.

The persistence of leeward slope easterly modes under higher wind speed conditions suggests that the recirculation is mechanically driven. Analysis of the timing of similar easterly modes experienced in the same location across the valley by Sharples et al. [2010] also showed limited diurnal patterns, suggesting that this recirculation region is in fact an area of persistent leeward slope eddies within the vegetation layer, rather than thermally driven up-slope winds. While WindNinja with either solver is not intended to predict within-canopy flows, with no mechanism for wind direction adjustment these eddies are not captured within fire modelling frameworks.

Finally, on the eastern slope (B8, B9 and B10) agreement percentages are somewhat larger than for the western slope. The momentum solver predicts a considerable northerly bias to the flow through the valley, and this appears to have the greatest impact on the eastern slope. Due to this, the native solver performs better than the momentum solver at all three sites for all wind speed thresholds. In particular, at B9 the native solver predicts a WNW direction which captures the edge of the

mode, while the momentum solver misses the mode by predicting a NW direction, resulting in agreement values of less than 3% as opposed to values up to 20% given by the native solver. This dramatic difference is also in part due to the discretisation of wind direction, i.e. the binning of observations, which results in a considerable difference in observations between two adjacent bins.

6.1.3.2 Ensemble Prediction

Figure 6.5 shows the discrete observed unconditional wind direction distributions for B1, B4, B7, B8 and B11 across Flea Creek Valley (with $T = 0\text{ms}^{-1}$), as well as the predicted wind direction distributions produced using an ensemble-style application of WindNinja with the momentum solver. On the western and eastern ridge tops (B1 and B11, respectively), the predictions capture the dominant modal structures observed at the stations. In particular, at B1 the model captures the dominant WNW prevailing wind directions and the secondary easterly prevailing wind direction (Figure 6.5(a)). At B11, although a bimodal distribution is predicted, the modes are very concentrated, covering only single wind direction bins (Figure 6.5(e)). This lack of variation shown in the predicted wind direction distributions is evident across the valley.

Under ensemble-style prediction, the model generally predicts at least one mode coincident with the observed dominant wind direction modes at each site. At B3 and B4, in contrast to the deterministic predictions, the predicted distributions pick up the wind reversal modes, e.g. the dominant easterly mode shown in Figure 6.5(b), while at B9 and B10, the dominant westerly modes are predicted. Through the valley floor (B6 and B7), the model indicates strong bimodal structures to the wind direction distributions which are somewhat evident in the observations, but dampened by considerable variation (e.g. Figure 6.5(c)).

The high variation in the wind directions observed at the valley floor sites reduces the efficacy of the prediction, but the variation itself may be induced by local features affecting the wind field which are at higher resolutions than the model can resolve. For instance, B6 is located on top of a knoll at the bottom of the valley, which may induce localised flows or eddies which cannot be represented at the resolution used to predict the wind field (i.e. 90m). This is again evident

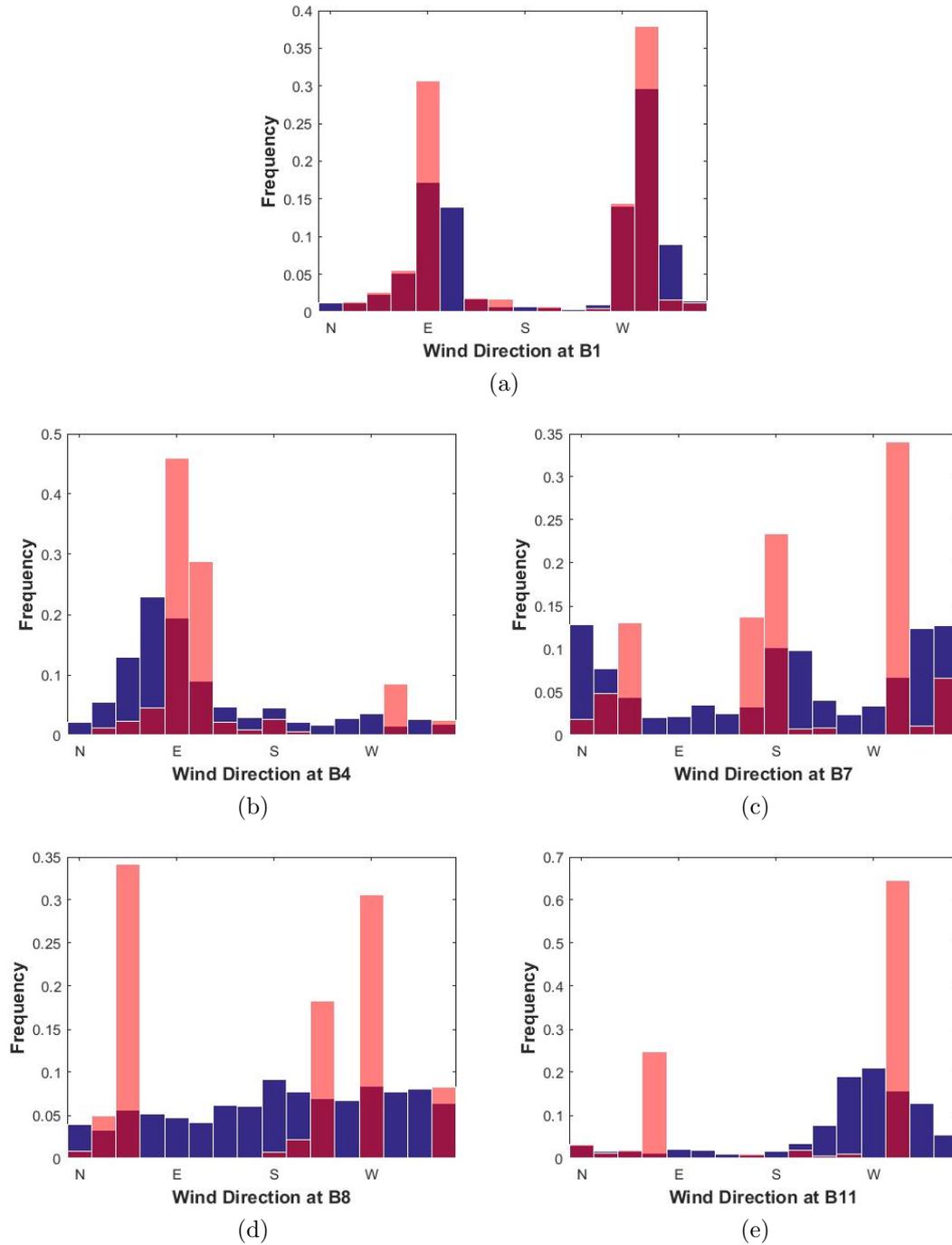


Figure 6.5: Observed (blue) and predicted (red) unconditional wind direction distributions at (a) B1 on the western ridge top, (b) B4 on the east-facing slope, (c) B7 on the valley floor, (d) B8 on the lower west-facing slope and (e) B11 on the upper west-facing slope.

at B8 on the eastern slope (Figure 6.5(d)); observed wind directions are almost uniform around the compass, whereas the model predicts an approximately bimodal distribution representative of winds along the valley axis.

Table 6.3 shows the proportion of time that WindNinja with the momentum solver predicted the same wind direction as observed, or within one or two compass sectors (i.e. $\pm 22.5^\circ$ or $\pm 45^\circ$). At the western ridge top (B1) the model is very accurate, consistently predicting wind directions within one sector of the observations. At the remaining ridge top stations (B2 and B11), as well as on the western slope (B3 and B4) and on the eastern slope (B9), the model predicts wind directions within the same compass quadrant as those observed (i.e. within $\pm 45^\circ$) over 50% of the time. These sites show the greatest similarity between the observed and predicted wind direction distributions.

Table 6.3: Proportion of overlap between predictions and observations at 1-minute time steps. ‘Overlap’ is taken to be a prediction of wind direction in the same compass sector, within one sector (or $\pm 22.5^\circ$) or within two sectors (or $\pm 45^\circ$).

	Same Sector	\pm One Sector	\pm Two Sectors
B1	0.7352	1.0000	1.0000
B2	0.1210	0.3797	0.5396
B3	0.1577	0.4049	0.5812
B4	0.1485	0.4072	0.5843
B5	0.0345	0.1056	0.1706
B6	0.1214	0.3374	0.4674
B7	0.0646	0.1886	0.3184
B8	0.0558	0.1652	0.3051
B9	0.1469	0.3973	0.5768
B10	0.0617	0.2055	0.3699
B11	0.1140	0.3760	0.5843

The proportions of overlap between observed and predicted wind directions are lower at locations across the valley where greater variation was shown in the observed distributions (i.e. B5, B7, B8 and B10). Table 6.3 shows an overlap within $\pm 45^\circ$ for B6 of 47%. At this site, a strong bimodal distribution is predicted, while a secondary wind direction mode was not strongly observed. The lowest overlap proportions are shown at B5, with only 17% overlap within two compass sectors. On inspection of the predicted distribution (not shown here), it appears that the model has not captured the structure of the observed wind direction distribution, i.e. a dominant southerly mode is observed while the model predicts a bimodal distribution.

6.1.4 Discussion

In the pursuit of accurate fire spread and behaviour prediction, the accuracy and uncertainty of model inputs must be considered. Traditional deterministic physics-based models can be complemented with probabilistic information formed by empirical data. This study shows a stark comparison between the application of a diagnostic wind model using the traditional deterministic approach and an ensemble-style approach incorporating empirical information. The differing results have considerable consequences when considered within fire modelling frameworks.

In the deterministic application of WindNinja with both native and momentum solvers, the model outputs were able to capture broad-scale dominant wind direction features, i.e. dominant westerly winds on the ridge tops and eastern (windward) slopes and the northerly flows through the valley floor. With the addition of the momentum solver, WindNinja was able to better capture some topographic impacts on wind flow across the valley, including recirculation through gullies and larger-scale channelling. However, large-scale northerly channelling through the valley predicted by the model was not strongly observed in the data. This over-prediction of channelling by the momentum solver led to lower performance on the windward slope; reducing percentage agreements from 20% down to only 3%. In the context of fire, this significant difference between predicted and observed wind direction may cause considerable difference between predicted and observed fire spread.

Both model solvers were extremely limited by their deterministic nature, leading to very small agreement percentages between single predictions and observed distributions. As previously noted in the literature [Forthofer et al., 2014b,a, Butler et al., 2015, Wagenbrenner et al., 2016], the models performed least well on the leeward slope. On this slope, dominant modes representative of leeward slope recirculation or eddies were not captured by the model with either solver. It has now been well researched and shown that such leeward slope eddies can create the necessary conditions for dangerous and extreme fire behaviour [Sharples et al., 2012, Simpson et al., 2013]. Through an inability to model important wind behaviours there is potential to misrepresent the spread and behaviour of a wildfire in the landscape.

Ensemble-style application of the WindNinja model with momentum solver clearly gave results in probabilistic terms that were more informative than the individual deterministic predictions. The predicted wind direction distributions were able to capture many of the key structures of the wind distributions observed across the valley. Bimodal distributions were predicted at a number of sites where the deterministic application of the model was only able to predict a single outcome. The ensemble predictions were also able to capture the recirculation modes observed on the western slope of the valley, leading the way for better modelling of the potential for associated extreme fire behaviours in such areas of the terrain.

Although the ensemble predictions were able to capture key wind direction modes at each site, they showed a lack of variability compared to the observations. In areas where wind directions were highly variable, overlap percentages between predictions and observations (to within an entire compass quadrant) could be as low as 17%. There are always potential improvements to be made to better capture underlying physical processes, but dynamic downscaling models can still be limited by resolution. Mechanisms existing at finer scales will continue to contribute uncertainty to model outputs. From this study, it is clear that the ensemble-style application of WindNinja shows differing levels of accuracy across the landscape. There is potential to identify expected levels of model uncertainty depending on physical features such as the terrain or vegetation. Using developing ensemble-based approaches [French et al., 2014d, Twomey and Sturges, 2016], this quantification of input uncertainty can be fed through the fire modelling framework to allow assessment of uncertainty around operational fire spread and behaviour predictions.

6.2 Wind Speed Reduction Beneath the Canopy

6.2.1 Background

Vertical wind profiles within the boundary layer are most often described using a logarithmic profile [Touma, 1977]. This profile becomes disturbed near the surface due to roughness of varying length scales, from topographical scales down to vegetation. These disturbances close to the ground, or at ‘mid-flame’ height are the wind patterns that drive surface bushfires beneath the canopy. In the current suite of

fire spread prediction models, wind speeds measured in the open environment (with no vegetation or above the vegetation layer) are translated to predict wind speeds within forests or vegetation using adjustment factors. The ‘wind reduction factor’ (WRF) [Cionco, 1972, Rothermel, 1972] is used within operational fire spread modelling in Australia, and is essentially defined as the ratio of open wind speeds and those observed within the vegetation. In the USA, the reciprocal of the WRF is utilised and known as the ‘wind adjustment factor’ (WAF) [Andrews, 2012], taking sheltered wind speeds as the numerator and open wind speeds as the denominator.

Both the WRF and WAF have been empirically derived for a number of structural vegetation features including crown ratios and vegetation age, and assume the wind reduction profile to be constant throughout the vegetation layer. However, Moon et al. [2013] presented empirical wind speed reduction profiles for different forest types, showing that profiles within the canopy were in fact non-constant. These wind profiles varied considerably with prevailing wind speeds as well as height above ground within the canopy. Cruz and Alexander [2013] noted that aside from topographical features, the principal drivers behind the behaviour of spreading fires are fuel moisture and wind speed. Alongside the recognition by Van Wagner [1989] that the prediction of surface fires may well be more difficult than that of crown fires due to the complexity of understorey fuels, the variation of wind fields within the vegetation layer adds further complications to the modelling of surface fires spreading beneath the canopy.

Fire spread prediction models are not immune to the effects of error accumulation, and it has been noted that the main sources of errors in fire model predictions include input data error [Cruz and Alexander, 2013]. It is within the interest of the fire research industry to better understand, and therefore model, the variability of wind fields within vegetation layers. A Generalised Algebraic Modelling System (GAMS) is being used to develop an empirical model for wind speed reduction profiles under various conditions (K. Moon, pers. comm., 2015). This new model is based upon data collected in flat terrain areas, where the impacts of topography were intentionally minimised. Although there is a significant body of work on wind behaviour over topographical features [e.g. Holmes et al., 1997, Glanville and Kwok, 1997, Cleugh, 2002] and within canopies [e.g. Finnigan, 2000, Belcher et al., 2012],

there are only a limited number of studies that consider the combined impacts of vegetation over complex terrain [e.g. Allen, 2006].

The data collected as part of this research provide an ideal opportunity to extend the work of Moon et al. [2013, In Press] to understand the impacts of vegetation regrowth on wind speeds experienced within the canopy over complex terrain. Wind data collected across FCV and the NAC are used to evaluate the applicability of the empirical wind speed reduction profiles described by Moon et al. [2013, In Press] to rugged landscapes, particularly valley structures within mountain ranges, and ridge lines or spurs across undulating hills.

6.2.2 Empirical Wind Speed Reduction Profiles

Moon et al. [2013, In Press] describe the collection and analysis of wind data from seven different vegetation types in Victoria, Australia. Data were collected at heights of 1, 2, 5, 10 and 15m using guyed-masts with horizontal cup anemometers. At each site, average 30-minute wind speed measurements were taken between four closely located weather stations, with a fifth station located at a nearby ‘open environment’ site. Data were collected over approximate 1-month periods.

To avoid the effects of topography, the stations were located in flat areas, and all stations were at least 20 times the vegetation height away from the vegetation boundary to avoid edge effects. To account for the accuracy of the cup anemometers, wind speeds below 1kmh^{-1} ($\approx 0.278\text{ms}^{-1}$) were excluded from the analysis presented in Moon et al. [In Press]. For the case studies analysed in this thesis, wind speeds below 0.4ms^{-1} ($\approx 1.4\text{kmh}^{-1}$) were excluded, again to account for the accuracy of the instrumentation.

Akin to the WRF, Moon et al. [2013, In Press] calculated the reduction of wind speed induced by each forest type, termed Relative Wind Speed (RWS), as the wind speed measured within the vegetation, U_V , divided by the wind speed measured at the nearby ‘open’ site, U_O ;

$$\text{RWS} = \frac{U_V}{U_O}. \quad (6.1)$$

In this study, relative wind speeds were calculated under increasing minimum prevailing wind speed thresholds (observed at the ridge top sites) to understand the changes in *RWS* as prevailing wind speeds increased; $T = 0.4\text{ms}^{-1}$ ($\approx 1.4\text{kmh}^{-1}$), $T = 2\text{ms}^{-1}$ ($\approx 7.2\text{kmh}^{-1}$) and $T = 4\text{ms}^{-1}$ ($\approx 14.4\text{kmh}^{-1}$). Relative wind speed results for each case study were compared to expected *RWS* values given in Moon et al. [2013, Fig. 3] for 10kmh^{-1} to 20kmh^{-1} winds, as well as the wind speed profiles and results shown in Moon et al. [In Press, Figs. 2 and 3].

6.2.3 Evaluation Results

6.2.3.1 Relative Wind Speed at Flea Creek Valley

In this study, wind speeds are compared at the ridge top stations and at the four co-located stations between the two sample years (using the 30-minute data observed between April and July). These sites are shown in Figure 5.1, and a summary of average wind speeds observed across Flea Creek Valley in 2007 and 2014 are given in Table 6.4. The ridge top pair (A1-B1) will be referred to as ‘FCV0’, while the four co-located pairs across the valley will be referred to as FCV1, FCV2, FCV3 and FCV4, as defined in Chapter 5 (Section 5.1.2.1).

Table 6.4: Average wind speeds for each site across FCV, at minimum wind speed thresholds, T , observed at A1 in 2007 and B1 in 2014.

T	2007			2014		
	0.4ms^{-1}	2ms^{-1}	4ms^{-1}	0.4ms^{-1}	2ms^{-1}	4ms^{-1}
FCV0	1.2583	2.6655	4.2722	1.2194	2.7704	4.6071
FCV1	0.8484	1.3801	2.5013	0.4798	0.4753	0.6250
FCV2	0.9670	1.3848	1.8590	0.7735	0.9516	1.3974
FCV3	0.8699	1.3225	1.5974	0.5550	0.7253	1.1122
FCV4	0.8365	1.2587	1.6217	0.6507	0.9182	1.3379

At Flea Creek Valley, 2014 wind speeds were considered relative to 2007 wind speeds where vegetation already existed. Therefore, the expected relative wind speed between the two years was calculated by taking the ratio of *RWS* for the regrowth open forest and *RWS* for the mature open forest, as defined by Moon et al. [2013]. For the normalised height of 0.3 (i.e. at 5m in 15m high vegetation), the relative wind speeds shown in Moon et al. [2013, Fig. 3] were 0.11 for both regrowth and mature open forest, for $10\text{-}20\text{kmh}^{-1}$ open winds. If height was read directly at 5m, the *RWS* values were 0.11 and 0.09 for regrowth and mature open forest,

respectively. Thus, using this direct 5m vegetation height, the RWS from regrowth to mature open forest was approximately 0.82, while considering the normalised height the RWS was 1.00. In addition, when considering the results shown in Moon et al. [In Press, Fig. 2], the RWS for both regrowth and mature open forests at a normalised height of 0.3 was approximately 0.15, giving a RWS value between the two forest types of 1.00. Moon et al. [In Press, Fig. 3] also shows that the RWS between regrowth and mature forest is approximately 1.00.

Figure 6.6 shows the relative wind speed results from Flea Creek Valley. For the western ridge top pair (FCV0), all RWS values at this site correspond well with the findings of Moon et al. [2013, In Press]. After accounting for the relative change in vegetation, RWS values appear around 1.00, indicating that the increased vegetation between the two years has had very little impact on wind speeds experienced on the ridge top.

On the valley floor (FCV2) and eastern ridge top (FCV4), RWS values are at the lower end of the expected RWS range, indicating that the increased vegetation at these sites has induced higher levels of wind speed reduction than those observed by Moon et al. [2013, In Press]. On the walls of the valley, relative wind speeds appear considerably lower than those observed on flat terrain. On the predominantly windward slope (the eastern valley wall, FCV3), RWS values are around 0.6, with increased wind speeds having limited impact on the RWS values.

In contrast, on the predominantly leeward slope (or western valley wall, FCV1), RWS values are approximately 0.7 for the lowest wind speed threshold and values reduce consistently as the threshold increase. With open wind speeds above 4ms^{-1} , wind speeds experienced beneath the mature canopy in 2014 are only a third of the speeds experienced under the sparse canopy of 2007. These results suggest that for increasing open wind speeds, greater reduction of wind speed beneath the canopy on the leeward slope is experienced. This concurs with the results shown in Moon et al. [In Press, Fig. 3], but goes against the profiles shown in Moon et al. [2013].

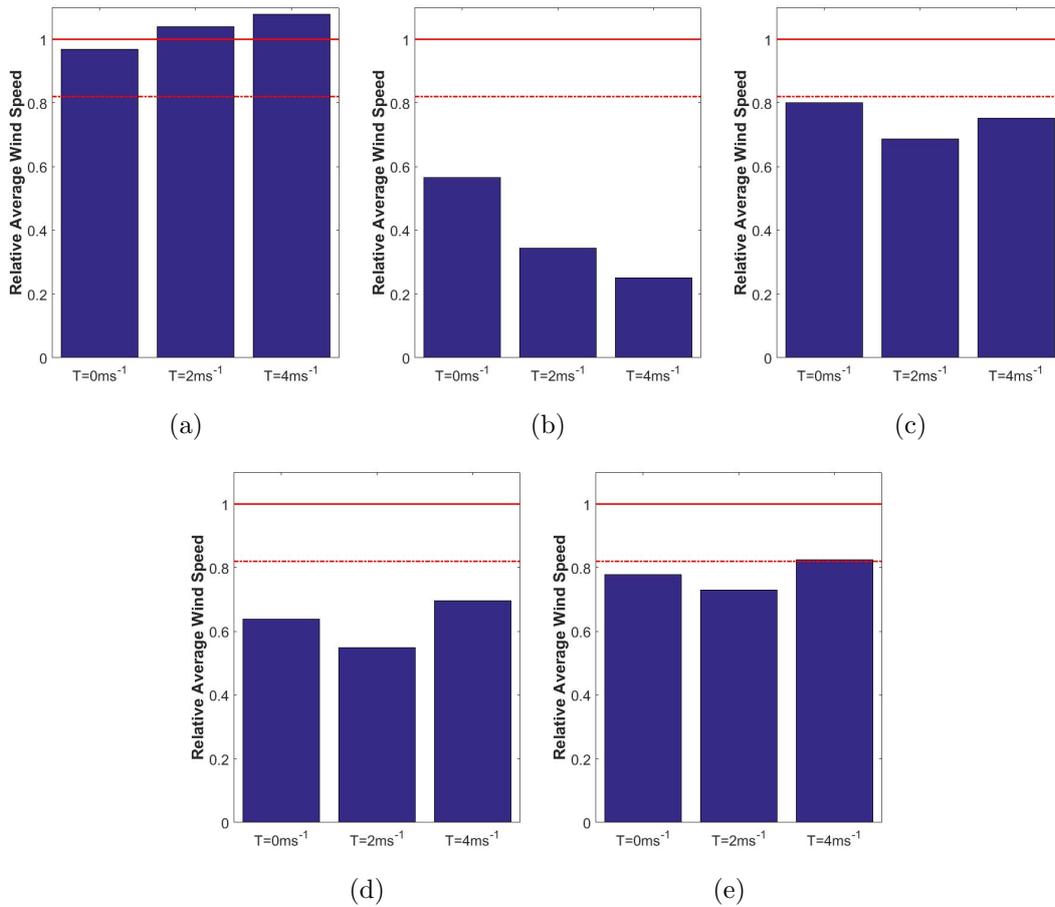


Figure 6.6: Relative wind speed observed between 2007 and 2014 at (a) FCV0 on the ridge top, (b) FCV1 on the east-facing slope, (c) FCV2 on the valley floor, (d) FCV3 on the lower west-facing slope and (e) FCV4 on the upper west-facing slope. Findings of Moon et al. [2013, In Press] are indicated with red lines; $RWS = 1.00$ (solid), and $RWS = 0.82$ (dotted).

6.2.3.2 Relative Wind Speed at the National Arboretum Canberra

In this study, the ‘open’ wind speed was taken to be that recorded on the clear slope at C1. The RWS was calculated between each weather station down the cleared east-facing slope and the parallel station on the east-facing slope within the pine plantation, therefore maintaining similar topographical features between stations. The pairs compared in this chapter will be referred to as ‘N0’ (C1-C4) on the ridge top, ‘N1’ (C2-C5) on the mid-slope and ‘N2’ (C3-C6) on the lower slope. These sites are shown in Figure 3.7, and a summary of wind speeds observed across the National Arboretum Canberra in 2015 is given in Table 6.5.

Results from the NAC were directly comparable to results from the ‘Pine plantation’ class in Moon et al. [2013, In Press]. From Moon et al. [2013, Fig. 3], relative wind

Table 6.5: Average wind speeds across the NAC, at minimum wind speed thresholds, T , observed at C1.

T	Clear Slope			Pine Stand		
	0.4ms^{-1}	2ms^{-1}	4ms^{-1}	0.4ms^{-1}	2ms^{-1}	4ms^{-1}
N0	3.8754	5.2196	6.5378	0.5913	0.5921	0.5968
N1	2.7591	3.4209	4.2462	0.8171	0.8215	0.8342
N2	2.2523	2.6587	3.1843	0.6547	0.6545	0.6529

speeds for mature pine forest at a normalised height of 0.3 (or a direct height of 5m) and a wind speed between 10kmh^{-1} and 20kmh^{-1} , were around 0.035. Moon et al. [In Press, Fig. 3] indicated a higher RWS value of approximately 0.1 for a normalised height of 0.3. Furthermore, Moon et al. [In Press, Fig. 3] showed that for wind speeds over approximately 4ms^{-1} , the RWS stabilised at approximately 0.08. For lower wind speeds, however, RWS values increase to approximately 0.2 for 2ms^{-1} wind speeds, and up to 0.4 for wind speeds as low as 0.4ms^{-1} .

Figure 6.7 shows the relative wind speeds between the parallel transects of the clear slope and the pine plantation slope at the NAC. At the ridge-top stations (N0), average wind speeds of 3.9ms^{-1} , 5.2ms^{-1} and 6.5ms^{-1} were recorded for the three increasing open wind speed thresholds, respectively. At these wind speeds, it is expected from Moon et al. [In Press] that the RWS values have reached stabilisation at 0.08. It is clear from Figure 6.7 that the observed RWS values approach this value as the wind speed threshold increases.

Down the predominantly lee-slope of the transect (N1, N2), RWS values are much higher, but average open wind speeds are lower; between 2.8ms^{-1} and 4.2ms^{-1} for N1 and between 2.3ms^{-1} and 3.2ms^{-1} for N2. At these lower open wind speeds, it is expected from Moon et al. [In Press, Fig. 3] that RWS values should be between 0.2 and 0.4, and results shown in Figure 6.7 agree with this expectation.

6.2.4 Discussion

Figures 6.6 and 6.7 show that results across sites representative of broader-scale terrain or undulating landscapes (i.e. FCV ridge top and valley floor, or the NAC transect) exhibit good agreement with the findings of Moon et al. [2013, In Press] for the relevant forest types and wind speed thresholds (allowing for the relative change in forest type at FCV). On the leeward and windward slopes of FCV, wind

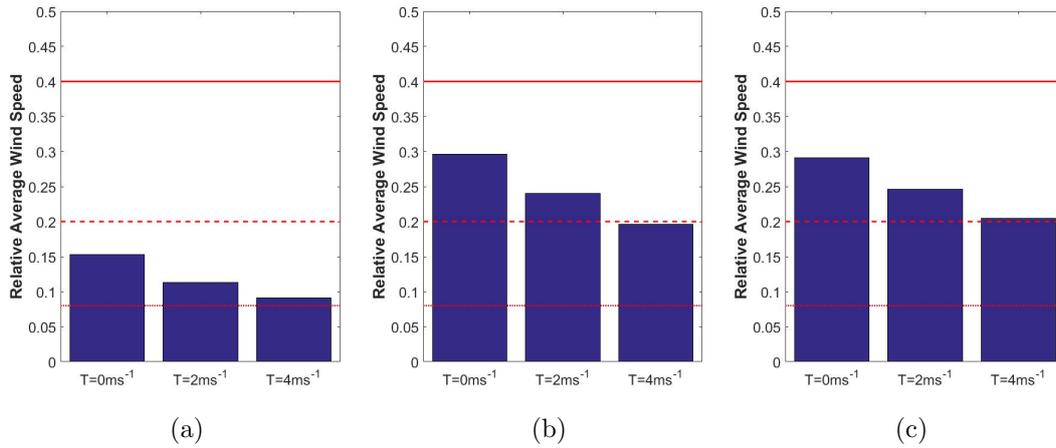


Figure 6.7: Relative wind speed observed between cleared east-facing slope and the parallel Radiata pine stand at the National Arboretum Canberra for (a) N0 on the ridge top, (b) N1 on the mid-slope and (c) N2 on the lower slope. Findings of Moon et al. [2013, In Press] are indicated with red lines; RWS = 0.40 (solid), RWS = 0.20 (dashed), and RWS = 0.08 (dotted).

speed reduction is much more significant between 2007 and 2014 than suggested by Moon et al. [2013, In Press]. These results suggest that the terrain features have a compounding role to play in wind speed reduction, and there is still further analysis required to better understand wind speed reduction beneath the canopy over complex terrain.

In further research, the compounding effects of topography may be characterised through the consideration of drag coefficients and streamlining [as noted by Moon et al., In Press]. Consideration of vegetation structure and penetrability, as well as three-dimensional turbulence will also be relevant to this future discussion. As an immediate extension of this study, RWS could also be calculated between each NAC station within the vegetation and the station on the clear ridge top. Differences between such results and those shown in Figure 6.7 would indicate any compounding effects of topography on RWS.

In this study, the average height of vegetation at both FCV and the NAC was significantly less than heights reported by Moon et al. [2013, In Press]. In addition, the vegetation structure was not quantified for both case studies, and it is important to consider where in the strata the 5m observations would sit. At the NAC, within the pine stand, the 5m wind observations were within the dense pine canopy, but at FCV, within the open forest, it is possible that the observations were made within a

secondary maxima in the wind speed profile. The comparison of normalised height results goes a considerable way to account for this with good agreement between results, yet further quantification of vegetation structure would be necessary to progress this research and characterise the impacts of vegetation and topography on wind speed reduction.

It is noted by Moon et al. [In Press] that the vegetation structure is itself dynamic, and varies over time. This variation is contemplated over long periods of time, with vegetation growth and interference due to human or natural causes. With shorter periods of data collection, i.e. less than one year, the seasonality of plant density, particularly through the open eucalypt forest may have a significant impact on relative wind speeds. In this study, although data collection periods spanned considerably different time scales (nine months for the case studies compared to one month collected by Moon et al. [2013, In Press]), there was good agreement between results at the broader landscape scale. Further investigation into the impacts of seasonality on RWS is possible with this dataset and it would be expected to further advance the discussion of drag effects and penetrability as areas for characterisation of the impacts of vegetation and topography on wind speed reduction. This could have significant implications for the application of wind speed reduction factors or models in fire spread prediction.

This study is limited by the caveats of wind data collection in the field. Data were collected using low-cost Davis cup anemometers which have been reported to show a bias towards lower wind speeds [Moon et al., In Press]. This form of data collection records horizontal wind speeds; limiting analysis to the horizontal, while vertical wind flow is unaccounted for and may have significant impacts on fire spread below the canopy. As noted by Moon et al. [In Press], more accurate data collection and more detailed analysis would be possible with three-dimensional sonic anemometers.

Possible edge effects at the NAC, observed on both the clear slope and within the Radiata pine stand, also need to be considered as a limiting factor in this study. Stations were at distances in the order of only a few times the height of the vegetation from the boundary of the pine stand, rather than 20 times the height of the vegetation as used by Moon et al. [2013, In Press]. In light of this, it might be

expected that the wind speed reduction observed at the stations would be less than that observed further away from the edge of the vegetation. However, in the results shown in Figure 6.7, this does not appear to be the case. Furthermore, analysis of wind direction across the NAC has shown that edge effects have minimal impacts on the wind fields experienced at the study sites.

Finally, it should be noted that reported wind speeds in this study were generally relatively low, but in application to fire spread prediction, high wind speeds are of most relevance in the context of extreme bushfire behaviour. Despite this, the focus here is on conditions driving surface fires beneath the canopy where conditions may be less extreme. Indeed, as noted by Moon et al. [In Press], in cases of prescribed burning conditions are ideally mild, so understanding low wind speed interactions within the vegetation layer is important in the development of accurate bushfire spread models.

6.3 Concluding Remarks

From the studies presented in this chapter, it is clear that observations of wind characteristics over areas of broad-scale topography or undulating terrain align well with deterministic wind direction predictions, as well as empirical wind speed profiles developed for flat terrain. However, within more complex landscape features, i.e. valley walls and leeward slopes, modelled wind fields can experience significant departures from observed wind behaviours. As expected from the discussions in Chapter 5, these results enhance the understanding of the significant impact that complex terrain has on wind flow and its variability.

Adaptation and further development of existing modelling techniques is therefore required to improve prediction of wind and fire (as well as their associated uncertainties) across the landscape. Probabilistic approaches used to complement current approaches are capable of better capturing variability in wind characteristics where traditional methods are limited. The approach highlighted in the first half of the chapter is not only better suited to the emerging probabilistic fire prediction frameworks, but also allows quantification of uncertainties to provide more detailed information to fire managers within operational time-scales.

The collection and analysis of wind data within complex terrain has a number of caveats, which have perhaps led to the lack of such datasets being available for model validation. As such, the evaluations in this chapter are constrained by the limits of the datasets themselves. For instance, conditional wind direction distributions were considered only under WNW prevailing winds across Flea Creek Valley, while significant easterly prevailing winds were also observed. Aggregation of the 1-minute observation intervals to longer time frames, such as 10-minute or 1-hour intervals, might also reveal different characteristics of the wind at varying resolutions. Analysis of these different scenarios may result in different conclusions to the evaluation of existing methods. However, the results presented in this chapter strongly highlight that the evaluation of models employed within the fire spread modelling process is a necessary step to reducing errors and characterising uncertainties in fire prediction.

CHAPTER 7

Discussions and Conclusion

This chapter will discuss and conclude the findings of this thesis in reference to the aim set out in Chapter 1: to improve understanding of wind flow over complex terrain from a statistical perspective, in the context of bushfire prediction.

The findings of the research are discussed in three sections. Firstly, the implications of the new wind direction datasets introduced in Chapter 3 and the probabilistic representation of wind direction shown in Chapter 4 are considered. Secondly, the methodological implications of the sensitivity analysis detailed in Chapter 5 are discussed. Thirdly, an in-depth discussion of the new knowledge and approaches that this thesis can contribute to modelling frameworks leads from Chapter 6.

Following these discussions, the thesis is concluded before some future directions for research are outlined.

7.1 Wind Direction: A New Perspective

7.1.1 New Datasets

An issue identified across the literature regarding wind and bushfire prediction, is the lack of data available to verify developing models and analysis. The validation of current wind models used for operational fire prediction has therefore been limited to a few case studies where terrain and vegetation have been restricted to create idealised scenarios. For instance, vegetation may be minimised to low-level grasses to remove its impacts on wind flow [e.g. Wagenbrenner et al., 2016, Butler et al.,

2015, Forthofer et al., 2014b], or flat terrain areas are chosen to avoid topographical influences on wind speed [e.g. Moon et al., In Press].

The datasets collected and analysed in this research provide a unique set of wind measurements across a variety of vegetation types in both complex and undulating landscapes in south-eastern Australia. The complicated environments in which the data for this research were collected are in many ways their asset; however they also contribute to the limitations of the datasets. The complex topography and vegetation of the Flea Creek Valley case study limits the generalisation of results. However, through analysis of the same location over two time periods it was shown that the regrowth of mixed Eucalyptus forest (akin to much of the forests across south-eastern Australia) had a significant impact on wind direction behaviours. In particular, it was clear that what was relatively uniform regrowth across the region had different impacts depending on the aspect of the terrain in relation to the dominant prevailing winds.

The National Arboretum Canberra case study attempted to reduce some of the complexities of the natural environment surrounding the data collection. Through taking advantage of discrete stands of vegetation types, it was possible to compare wind fields over similar topography with contrasting vegetation, and likewise through similar vegetation with changing topography. This study helped to reinforce the findings from Flea Creek Valley that vegetation and aspect in relation to the prevailing winds play a significant role in the occurrence of wind reversal regions, or persistent leeward slope eddies, within the canopy.

Despite some of the limitations of the data collection, such as the use of cup anemometers resulting in the collection of only horizontal winds with an accuracy of 0.4ms^{-1} , very large quantities of data were collected throughout this research. These data were at temporal resolutions down to one minute and spatial resolutions captured in only a handful of similar case studies across the globe. The research presented in this thesis did not allow for analysis of the detailed behaviours of the wind direction and speed at fine temporal scales, or spatial analysis such as correlations between sites, but the data remain available for further study (Appendix A). These datasets can therefore continue to contribute to the further development

and validation of current wind and fire modelling frameworks in realistic landscapes where fine-scale (temporal and spatial) fire behaviours may develop.

7.1.2 Directional Wind Response Distributions

It is clear from the literature discussed in Section 2.1.3 that dealing with uncertainty in fire modelling is a key issue in the prediction of fire spread and behaviour. Recent developments consider the use of ensemble-based or stochastic modelling frameworks to predict a range of possible fire outcomes which can be characterised with likelihoods and probabilities [e.g. French et al., 2013, 2014d, Twomey and Sturges, 2016]. A probabilistic representation of input variables, such as wind fields, is ideally suited to these emerging frameworks and can help to provide accurate characterisation of uncertainty throughout the modelling process. At present, simplified probabilistic representations of wind characteristics, such as direction and speed, are formed with Uniform or Normal distributions [Twomey and Sturges, 2016]. However, the literature discussed in Section 2.3, and the data detailed in Chapter 3, suggest that wind direction distributions are often not well described by these standard functions, and in fact can vary significantly across the landscape.

Joint wind direction distributions, referred to in this research as directional wind response distributions, have been used to identify broad-scale meteorology as well as smaller-scale wind mechanisms relevant to bushfire behaviour in mountainous terrain [e.g. Whiteman and Doran, 1993, Sharples et al., 2010]. Using the discrete data collected in this research, the underlying continuous directional wind response process can be estimated. Unfortunately, however, there are limited estimation methods designed for bivariate circular (or toroidal) datasets that are readily available to practitioners.

It was shown in Chapter 4 that a conceptually simple adaptation to planar estimation methods can be used to accurately estimate toroidal surfaces. In some cases, the adaptation performed better than existing toroidal estimation techniques. However, there are a number of elements that were not handled in great detail in this research that may have significant impacts on the results presented, including the selection of smoothing parameters and bandwidth for the estimation methods. There are also clear considerations that must be made when applying the simple adaptation,

including the impacts of surface structure on the toroidality of the estimation. With these considerations in mind, it is possible to pre-process a toroidal dataset such that planar estimation methods can be optimally applied to provide an accurate toroidal surface approximation.

The resulting estimated continuous directional wind response distributions can be used to identify key wind behaviours and quantify their likelihoods. Consideration of the timings of distribution modes can also identify the mechanisms driving wind behaviours. For example, the timing of apparent channelling in Flea Creek Valley showed that thermal flows did not play a key role in wind direction behaviour through the valley (Section 6.1). Having estimated wind direction distributions and identified these key wind behaviours, it becomes of interest to understand how physical features such as vegetation and topography impact upon them. The following section deals with how statisticians can look to resolve such questions.

7.2 Statistical Techniques for Analysing Wind Direction

7.2.1 Comparison Tests

To understand the impacts of vegetation or topography on wind direction distributions, a statistical hypothesis testing framework was developed in Chapter 5. Given two observed distributions (either univariate or bivariate), the null hypothesis assumed that both datasets originated from the same underlying distribution, while the alternative hypothesis stated that the two originated from different processes. To test such a hypothesis, a number of statistics are available; within this thesis, two approaches were taken. The first considered Kolmogorov-Smirnov (KS) style tests based on the difference between empirical distribution functions. Sensitivity analysis in Chapter 5 highlighted the power of these tests, and the means by which statistical significance is established, against structured alternative distributions.

The two-sample univariate KS test for the comparison of linear datasets was extended by Kuiper [1960] to account for circularity, such as that found in wind direction data. The sensitivity of these tests to changes in Normal and von Mises distributions showed that both tests exhibited similarly powerful behaviour to changes

in mean and standard deviation, or concentration. This equivalent behaviour indicates that Kuiper's test is indeed a valid extension of the KS test. However, results for the approximation of the asymptotic behaviour of Kuiper's statistic for distributions that were very similar suggested that, unlike for the KS test, extreme caution should be taken in applying the approximation outside its valid range.

The extension of the KS test to compare bivariate distributions has received less analysis in the literature but is well used in certain fields, e.g. astronomy [Peacock, 1983]. Comparison of the sensitivity of the bivariate KS test to that of the univariate test revealed very similar behaviours for the approximation to the asymptotic behaviour of the test statistic. This again suggests a robust extension of the univariate KS test. However, the use of Monte Carlo simulations to discern significance of an observed test statistic was far less powerful under the bivariate extension.

It would appear that no bivariate extension currently exists for Kuiper's test to compare bivariate circular datasets. Using the basis of the bivariate KS test, an extended Kuiper's test statistic was proposed. An approximation to the asymptotic behaviour of such a statistic was outside the scope of this thesis, and so Monte Carlo simulations were relied upon to determine the significance of an observed test statistic. As in the case of the bivariate KS test, the Monte Carlo p -values were shown to be not very powerful for small changes in distribution structure, but showed a strong threshold for determining significance that could be utilised in application of the test. Similarities between the behaviour of the bivariate KS test and that of the proposed bivariate Kuiper's test suggest it to be a viable bivariate extension of Kuiper's test. However, more theoretical work is required to understand and analyse its asymptotic properties, as well as to determine whether it might exhibit similar behaviour to the asymptotic properties of the univariate Kuiper's test.

The second statistical approach taken in Chapter 5 considered the mean squared difference between the estimated bivariate directional wind response distributions. Under this approach, the distributions were treated as surfaces or images to be compared, thus circularity was not inherently accounted for. It was found that these statistics were sensitive to the surface estimation methods, particularly to

small surface values approximated away from the key distribution modes, and so truncating the surfaces above a constant (but small) value allowed for more useful results. In practice, this may only be possible when the practitioner can be assured that these small values are insignificant to the comparison. In the case of directional wind response, it was often clear where the modes were centred and there were only ever a few. In more complex images, it would be important to consider a truncation that minimises the loss of information while maximising the usefulness of the test results.

The sensitivity studies in Chapter 5 were able to shed light on the sensitivity of univariate and bivariate two-sample tests to changes in standard unimodal distributions. A number of lessons were then available in the application of such tests throughout the chapter. However, the simulated distributions were highly idealised, and changes in mean and standard deviation, or concentration, were dealt with independently. In reality, a significant test result may arise from a variety of differences between two distributions. It may well be possible to discern these visibly, but further parametric estimation and comparison of distributions may be required to fully interpret distribution changes.

7.2.2 Application of Statistical Techniques

In the application of the above statistical tests to both univariate and bivariate wind direction distributions, in general it was found that vegetation regrowth at Flea Creek Valley and the changes in vegetation and aspect at the National Arboretum Canberra had significant impacts on observed directional wind response. Although results varied somewhat between different tests, there were two clear exceptions to this finding.

At Flea Creek Valley, the exception was a site where high levels of variation were observed. At this site, the high variation persisted through the regrowth of vegetation and so the difference between the distributions was not found to be statistically significant. At the National Arboretum Canberra, no significant difference was suggested between the wind direction distributions observed on the forested leeward and 45° leeward slopes. Significant differences shown to exist between the 45° leeward slope and the cross slope indicated a threshold for aspect in relation to

the prevailing wind for the existence of wind reversal regions beneath the canopy. Equally, significant differences between the cleared and forested slopes also indicated the existence of a threshold for surface roughness in order to produce wind reversal regions on the leeward slope.

The most important finding from these comparisons was that the impacts of vegetation varied considerably across the landscape. Therefore as fire research approaches the challenge of statistically modelling wind flow, it must consider the relationship between wind flow and its environment as dynamic.

7.3 Statistical Characterisation of Wind in the Context of Bushfire

7.3.1 New Knowledge in the Existing Framework

The knowledge gained from new datasets and analysis shown throughout this thesis can be used to enhance and complement current wind and fire modelling approaches. The empirical information drawn from analysis of wind direction distributions builds on previous understanding of wind mechanisms that can form conditions susceptible to extreme or dynamic fire behaviour. For instance, the requirement of sufficiently steep terrain, increased surface roughness and a leeward aspect for the formation of wind reversals within the canopy concurs with that found in computational research into atypical lateral spread of bushfires [Simpson et al., 2016].

Although the studies detailed in this research do not yet provide comprehensive quantification, there is clear potential for probabilistic information derived from directional wind response distributions to be applied within the fire modelling process. Likelihood information regarding wind behaviours, such as channelling or reversals, could be overlaid as simple filters within existing frameworks, or enhance existing terrain filters such as that detailed by Sharples et al. [2012]. Inevitably, the limitations of the datasets and analysis within this research would be paramount when considering the generalisation of such information across broader regions, and more validation would be required with the potential for further data collection.

Aside from a more detailed understanding of wind direction developed through empirical analysis, an investigation into vegetation induced wind speed reduction

across both case studies was presented in Chapter 6. The new data allowed for verification of existing approaches to wind speed reduction modelling. Although empirical wind speed reduction profiles developed by Moon et al. [2013, In Press] were accurate across broad-scale terrain, it was shown that more complex topographical features could have a significant compounding impact on wind speed reduction. As a key variable in the driving of surface fires beneath the canopy, this work helps to shed light on the variation of such a parameter across the landscape and indicates where new research is required.

7.3.2 New Modelling Approaches

Probabilistic representation of parameters such as wind direction or wind speed allows for quantification of uncertainty for current and emerging modelling frameworks. A comparison between current wind model outputs and distributions of observed wind directions detailed in Chapter 6 highlighted the limitations of deterministic modelling approaches, and provided numerical values for varying levels of uncertainty across Flea Creek Valley. The accuracy levels of a single deterministic model output compared to a full distribution of observed wind directions were extremely low in some cases, and distributions most often did not fit the standard probability distribution functions utilised in emerging ensemble-based fire prediction frameworks.

With an ensemble-style application of the operational downscaling wind model WindNinja, it was possible to more accurately replicate the full distributions of wind direction observed across Flea Creek Valley. This positive result shows that it is possible to easily adapt current physics-based models in conjunction with new empirical probabilistic information to better capture the variation in wind direction that is experienced over complex terrain. However, further work is required in order to improve the predictions of the full wind direction distributions. Future directions in ensemble-based fire modelling are looking towards incorporation of the stochastic behaviour expected by wind flow to better characterise variation in wind direction and speed. Through a more detailed understanding and better characterisation of the spatial variability of inputs to fire models, it is possible to better capture uncertainty of the modelled outputs. This increased information provides decision

makers with better tools and understanding when faced with the management of a bushfire.

7.4 Conclusion

In the process of modelling bushfire as it spreads and changes behaviour across the landscape, the accurate modelling of its driving forces is fundamental. Wind speed and direction at the surface have a direct influence over how and where a fire propagates, but wind flow over complex terrain is driven by complicated physical mechanisms that cannot all be captured by real-time physics-based models. A statistical characterisation of wind flow in complex terrain can capture the uncertainty related to wind modelling, and allow analysis of scenarios and their likelihoods.

The research presented in this thesis details an analysis of wind direction in probabilistic terms. Through the collation of new datasets and a fresh perspective on the representation of wind direction, this research has developed a unique understanding of wind behaviours across complex and undulating terrain. Key insights include, but are not limited to, the distinct variation in directional wind response according to changing vegetation and topography across the landscape. Through investigation into the techniques available to handle wind direction data, better interpretation of statistical analyses of directional wind response is now available. Finally, this research has highlighted how new approaches to wind modelling can draw from and complement current approaches within fire prediction. This opens a pathway for developing more accurate frameworks capable of handling uncertainty and providing more detailed information to decision makers and fire managers.

7.5 Further Work

As is the case for many projects, this work has posed a number of new research questions that are yet to be answered. Here, some of these possible further questions are introduced.

- What are the likelihoods of certain wind behaviours within complex terrain?

Using probabilistic representations of directional wind response, it is possible to extend the work in this thesis to analyse the likelihood of given wind mechanisms such as leeward slope eddies or channelling through valleys. Generalisation of such likelihoods across the landscape would depend upon the constraints of this dataset and new data used for validation, but applicable probabilistic information could then be directly fed into wind and fire modelling frameworks.

- How can directional wind response be statistically modelled?

New developments with mixtures of bivariate von Mises distributions would allow for a model of multimodal bivariate circular distributions. The use of bivariate copulas may be employed to deal with the spatial correlations between prevailing and surface winds. A key issue within such parametric modelling would be the interpretation of parameters. For fire modelling applications, it would be required that these parameters be defined in terms of observable physical variables such as vegetation, slope or aspect.

- Is the extension of Kuiper's test reasonable, robust and consistent?

A theoretical basis for the extension of Kuiper's test to the bivariate case needs to be developed to ensure the statistic is robust and consistent. A key issue in this area is to determine whether the test statistic is distribution-free. It may also be preferable to develop an approximation to the asymptotic behaviour of the extended statistic in the same form as that of the univariate statistic, and equivalent KS statistics.

- How sensitive are fire modelling frameworks to variation in their inputs?

Probabilistic representations of wind direction should now be tested within ensemble-based fire models to understand the repercussions of their application. Detailed sensitivity analyses of fire modelling frameworks to variation in such input values have yet to be completed in detail. Such analyses would allow researchers and end users to identify the most significant inputs, parameters and structures in the modelling process and identify areas where further research is required.

APPENDIX A

Wind Data Summaries

The data used for analysis throughout this thesis are stored in Comma-Separated-Value format. Individual data files for each station over each case study (or time period) exist, as well as data files compiled for each variable; wind speed, wind direction, temperature, relative humidity and solar radiation. These data files may be accessed on request through the author, or the School of Physical, Environmental and Mathematical Sciences at UNSW Canberra.

A.1 Calibration Analysis

After the 2007 data collection by Sharples et al. [2010], the portable automatic weather stations (PAWS) were placed in storage for approximately 6 years, before being recommissioned in 2014 for the research presented in this thesis. To reaffirm the accuracy of the PAWS ($\pm 5\%$ or $\pm 1\text{ms}^{-1}$, whichever greater), the stations were set up at the UNSW Canberra Field Site, where observations from a sonic anemometer (SA) were available for comparison.

Using the SA, wind speed and wind direction observations were recorded at a height of 10 metres, logging at 10Hz intervals. The SA wind data were converted to 1-hour time intervals and compared with that recorded by the PAWS between 19th and 23rd February 2014. To account for the height discrepancy between the 10m SA and the 5m PAWS observations, the 1/7th power law was used [Touma, 1977];

$$SA_{5m} = (5/10)^{0.143} SA_{10m}.$$

Figure A.1 shows the absolute percentage difference between the mean wind speed data recorded across the eleven PAWS and the wind speed recorded by the SA, with the red line indicating $\pm 1\text{ms}^{-1}$ on the SA_{5m} observations. This figure, as well as individual station analysis not presented here, shows that the wind speed observations recorded by all eleven PAWS were generally within this accuracy region.

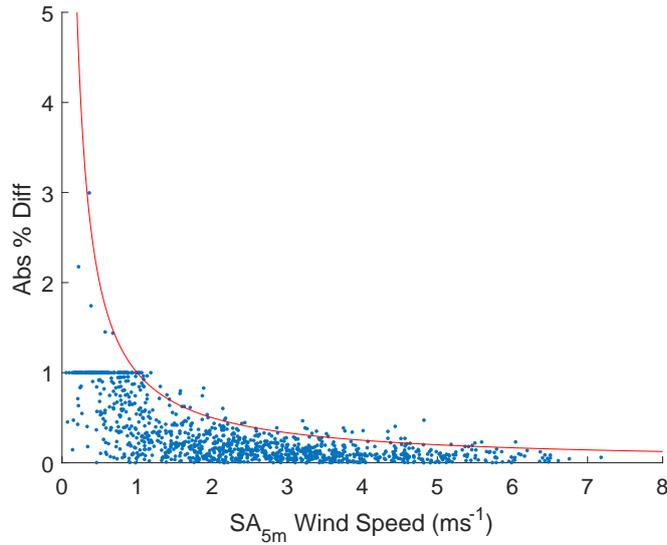


Figure A.1: Absolute percentage difference between PAWS and SA_{5m} wind speed observations between 19th and 23rd February 2014. Red line shows $\text{SA}_{5m} \pm 1\text{ms}^{-1}$.

Figure A.2 shows the time series of wind direction recorded by the eleven PAWS (blue dots) and those recorded by the SA (red line). The figure again indicates that the PAWS wind direction observations were consistent with the SA_{5m} observations throughout the calibration time period. Therefore, the PAWS were taken to be recording consistent and accurate wind speed and direction observations.

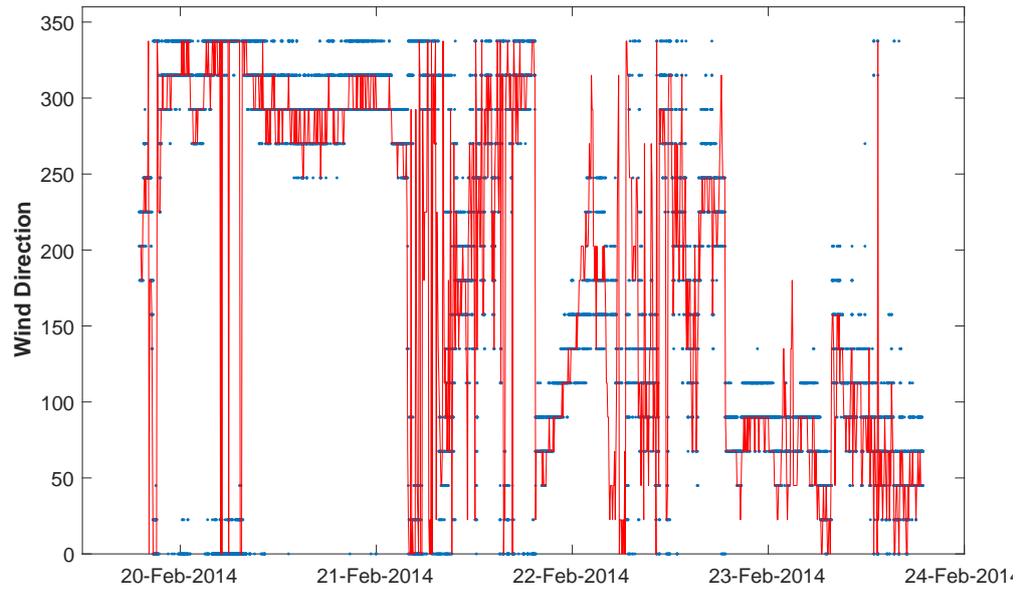


Figure A.2: Wind direction time series for PAWS (blue dots) and SA_{5m} (red line) between 19th and 23rd February 2014.

A.2 Flea Creek Valley 2007

Figure A.3 shows the data record from the five weather stations across Flea Creek Valley in 2007. The gap in the data record was due to battery outages over the winter of 2007, and lack of access to the valley due to snow [Sharples et al., 2010]. After this gap, the weather stations were set to observe data at 1-hour intervals, rather than the 30-minute intervals that had been used up until July.

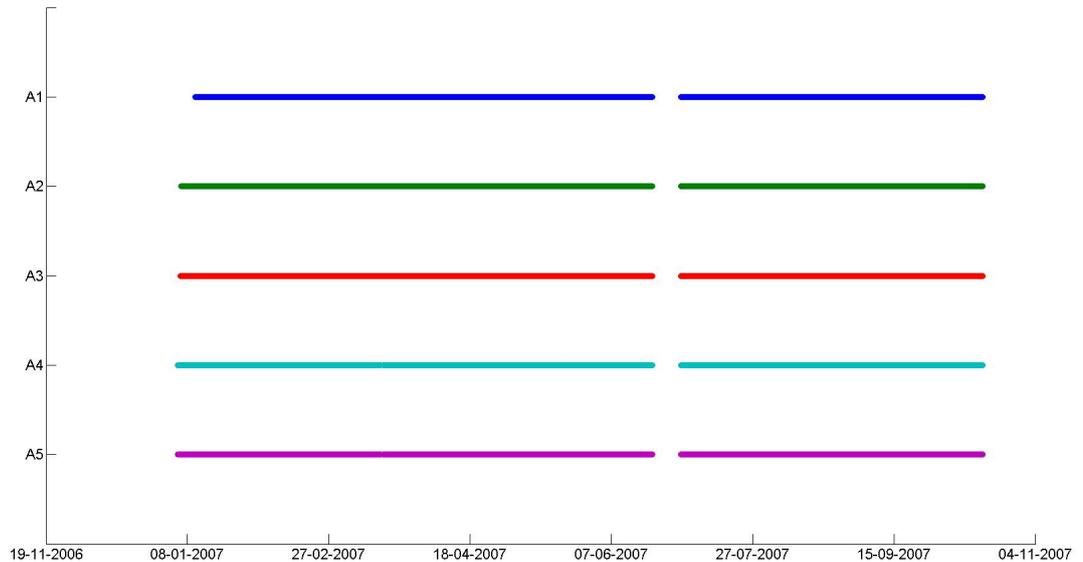


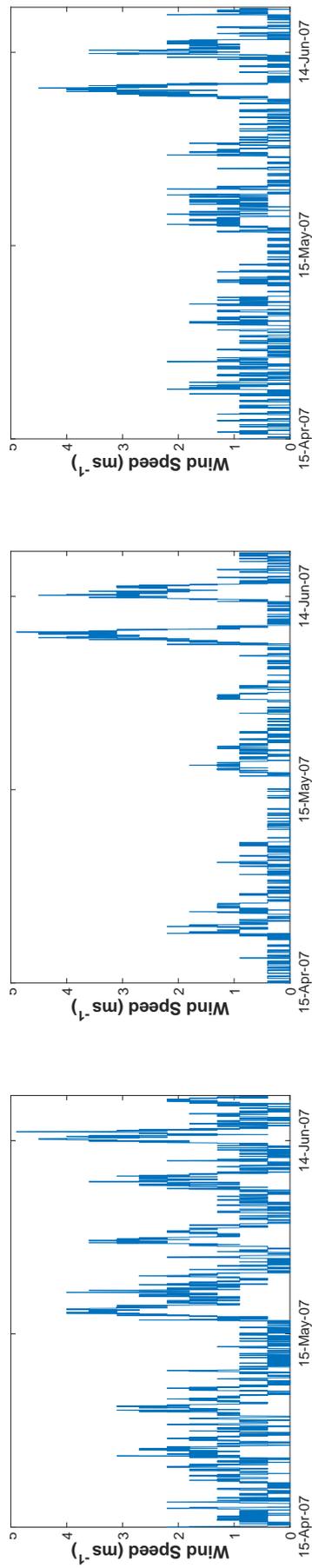
Figure A.3: Data records across FCV between April and December 2007.

Figure A.4 shows each of the wind speed time series for the data collected across Flea Creek Valley between April and June 2007.

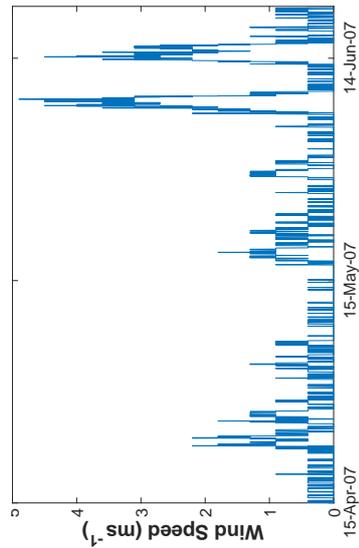
Table A.1 summarises the proportions of wind direction data observed within each compass sector, for each station across Flea Creek Valley between April and June, 2007. The empty data points were discarded from analysis.

Table A.1: Proportion of winds originating from each compass direction across FCV between April and June 2007.

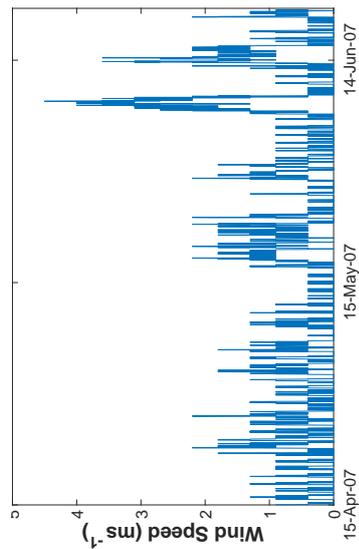
	A1	A2	A3	A4	A5
<i>Empty</i>	0.03	0.01	0.01	0.00	0.00
Calm	0.01	0.07	0.07	0.06	0.26
N	0.00	0.03	0.12	0.04	0.03
NNE	0.02	0.02	0.15	0.02	0.04
NE	0.04	0.03	0.12	0.04	0.06
ENE	0.02	0.08	0.06	0.12	0.02
E	0.10	0.19	0.02	0.05	0.02
ESE	0.13	0.19	0.02	0.03	0.02
SE	0.06	0.10	0.02	0.02	0.03
SSE	0.05	0.02	0.01	0.03	0.01
S	0.05	0.02	0.04	0.05	0.01
SSW	0.02	0.03	0.08	0.04	0.03
SW	0.03	0.03	0.06	0.05	0.06
WSW	0.12	0.02	0.05	0.04	0.08
W	0.13	0.04	0.04	0.07	0.10
WNW	0.12	0.07	0.04	0.18	0.09
NW	0.07	0.04	0.03	0.12	0.09
NNW	0.01	0.02	0.08	0.05	0.06



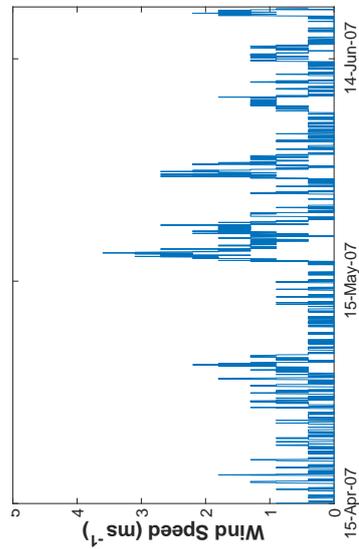
(a) A1



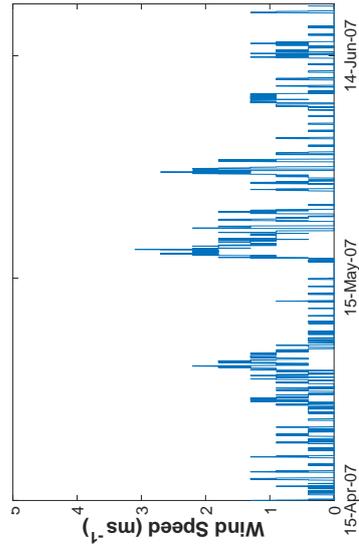
(b) A2



(c) A3



(d) A4



(e) A5

Figure A.4: Time series of observed wind speeds (ms^{-1}) across FCV at (a-e) A1 to A5 between April and June 2007.

A.3 Flea Creek Valley 2014

Figure A.5 shows the record for the data collected across Flea Creek Valley between April and December 2014. The gaps in the record are most often caused by battery outages due to a lack of solar radiation at the sites, as well as a number of other challenges discussed in Chapter 3. The Raspberry Pi systems were installed in July 2014 and the observation frequency was increased to 1-minute rather than 30-minute intervals, but continued battery issues were experienced. After completed testing, B3 was installed in September 2014.

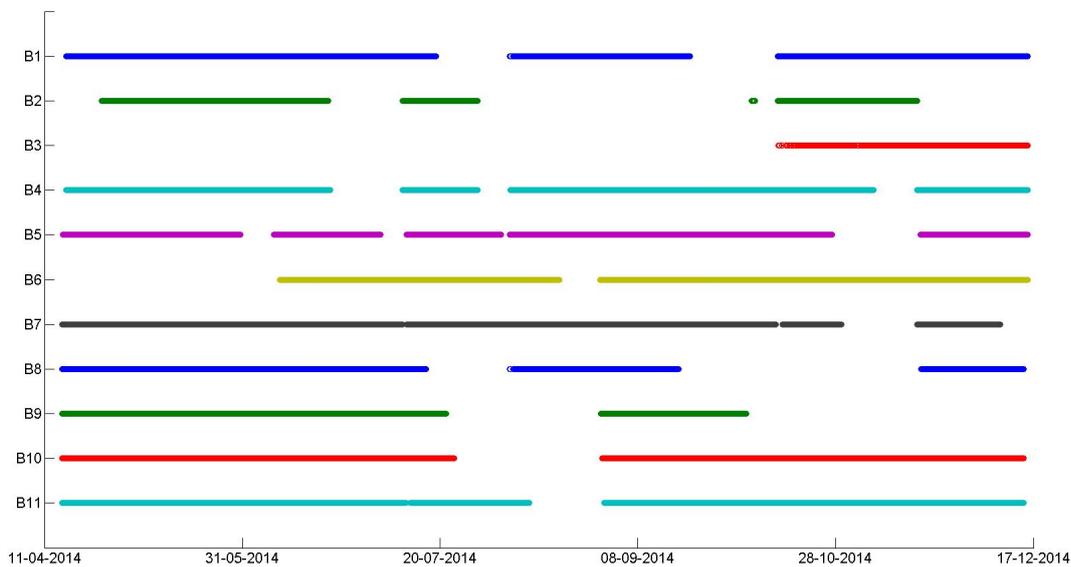


Figure A.5: Data records across FCV between April and December 2014.

Figure A.6 shows the time series of wind speed data collected at each station across Flea Creek Valley in 2014. The gaps in the data highlighted in Figure A.5 are also evident in these time series.

Table A.2 summarises the proportion of observations within each wind direction sector from the data collected across Flea Creek Valley in 2014. A greater proportion of ‘Empty’ data points were recorded due to battery issues experienced throughout data collection. A greater proportion of ‘Calm’ observations were also recorded as compared to the 2007 Flea Creek Valley data shown in Table A.1.

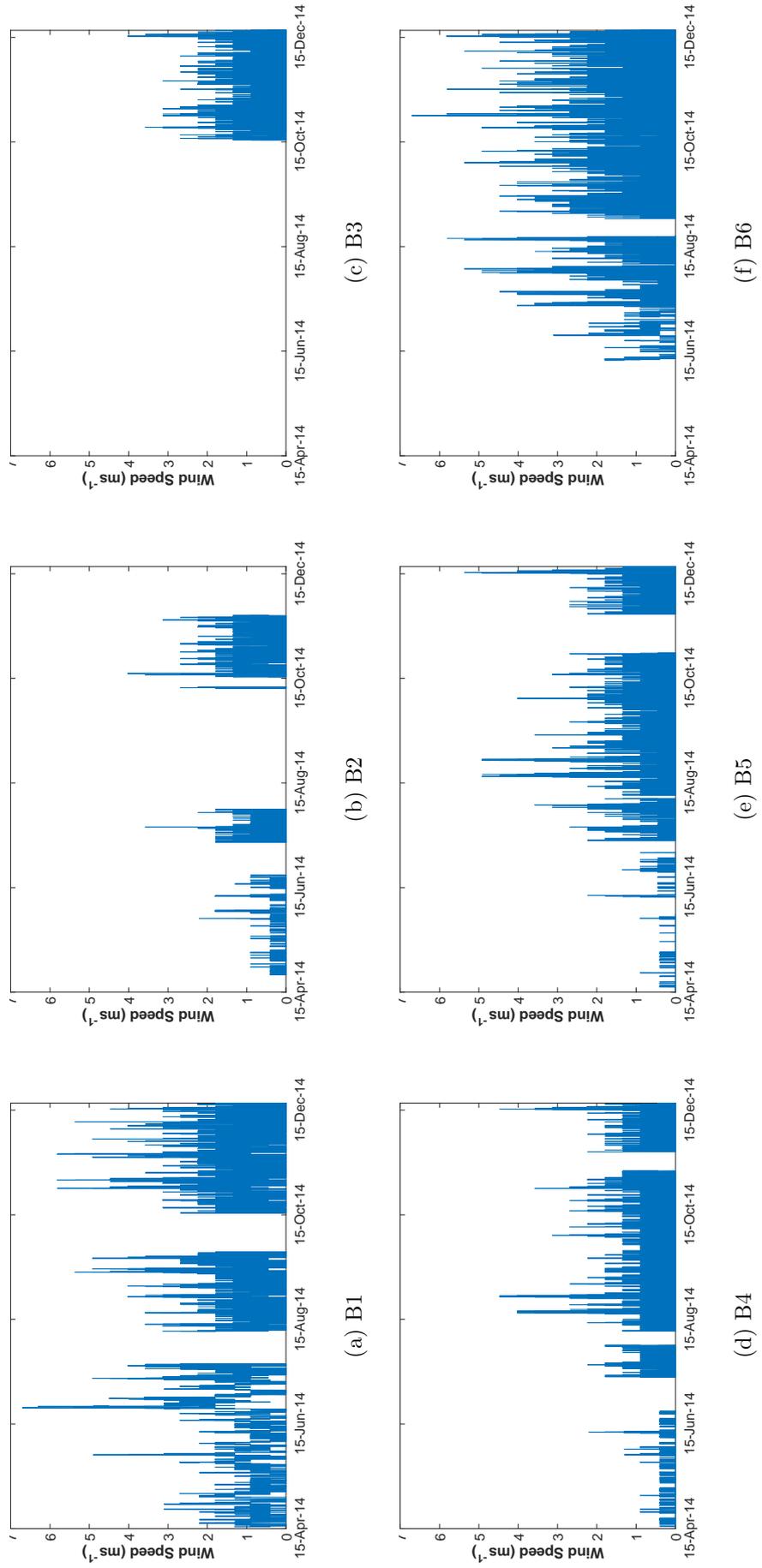


Figure A.6: Time series of observed wind speed (ms^{-1}) across FCV at (a-k) B1 to B11 between April and December 2014.

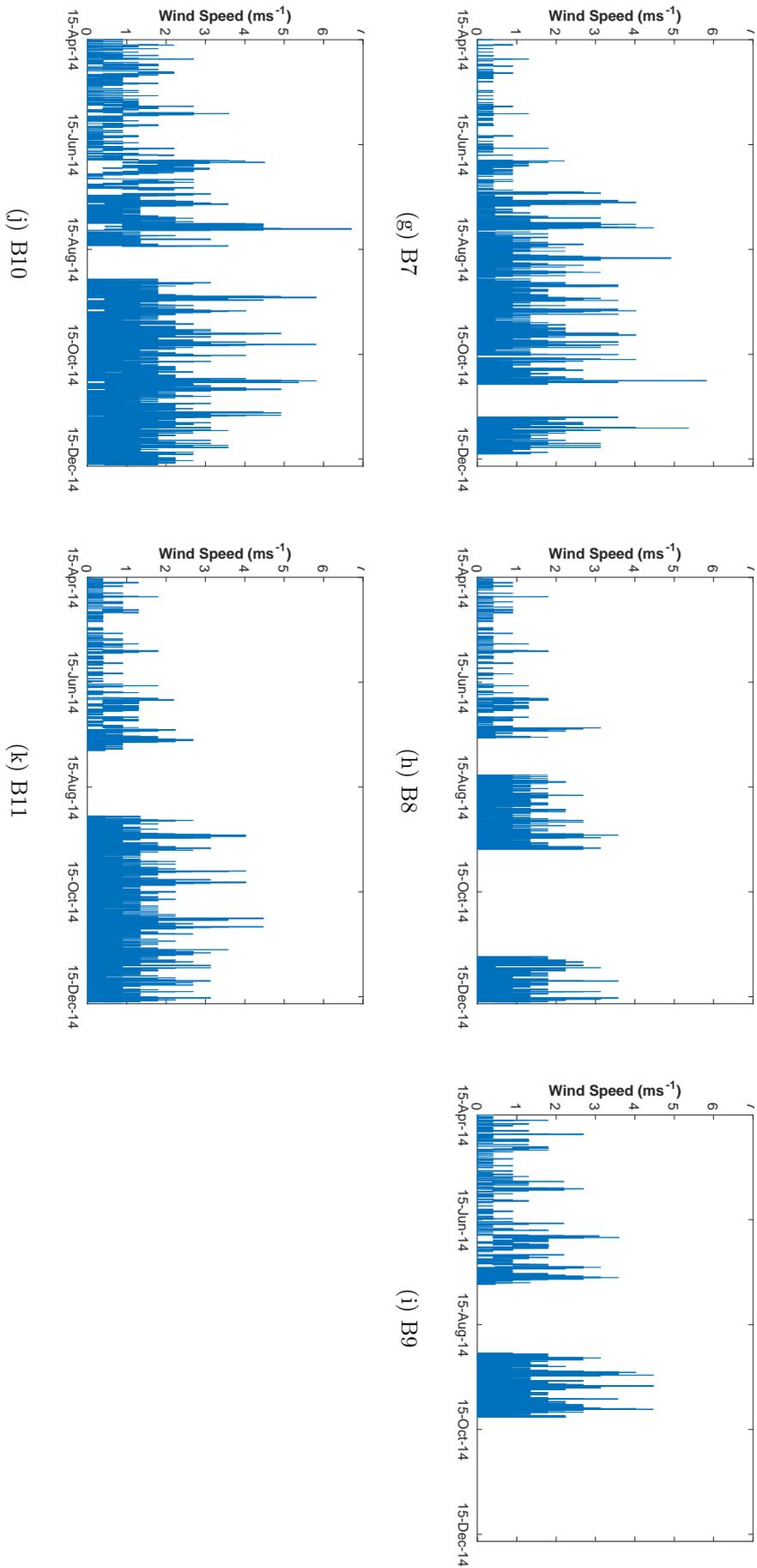


Figure A.6: (Cont'd) Time series of observed wind speed (ms^{-1}) across FCV at (a-k) B1 to B11 between April and December 2014.

Table A.2: Proportion of winds originating from each compass direction across FCV between April and December 2014.

	B1	B2	B3	B4	B5	B6	B7	B8	B9	B10	B11
<i>Empty</i>	0.28	0.66	0.65	0.15	0.19	0.10	0.20	0.54	0.70	0.27	0.16
Calm	0.13	0.11	0.18	0.31	0.49	0.41	0.54	0.28	0.14	0.40	0.28
N	0.01	0.00	0.01	0.01	0.01	0.04	0.03	0.01	0.01	0.01	0.02
NNE	0.01	0.00	0.01	0.03	0.01	0.04	0.02	0.01	0.01	0.01	0.01
NE	0.02	0.00	0.02	0.05	0.02	0.07	0.01	0.01	0.01	0.01	0.01
ENE	0.03	0.00	0.03	0.10	0.01	0.06	0.01	0.01	0.01	0.01	0.01
E	0.10	0.00	0.03	0.10	0.01	0.02	0.01	0.01	0.00	0.01	0.01
ESE	0.08	0.00	0.01	0.05	0.02	0.01	0.01	0.01	0.00	0.01	0.01
SE	0.01	0.01	0.01	0.03	0.02	0.01	0.01	0.01	0.00	0.01	0.01
SSE	0.00	0.03	0.01	0.02	0.03	0.01	0.01	0.01	0.00	0.01	0.01
S	0.01	0.03	0.01	0.04	0.04	0.01	0.02	0.01	0.00	0.02	0.01
SSW	0.00	0.02	0.01	0.01	0.05	0.02	0.02	0.01	0.01	0.07	0.02
SW	0.00	0.02	0.01	0.01	0.04	0.03	0.01	0.01	0.02	0.06	0.05
WSW	0.01	0.02	0.01	0.02	0.02	0.03	0.01	0.01	0.02	0.04	0.10
W	0.08	0.02	0.01	0.03	0.02	0.02	0.01	0.02	0.02	0.04	0.12
WNW	0.18	0.04	0.00	0.01	0.01	0.02	0.02	0.02	0.02	0.03	0.08
NW	0.05	0.05	0.00	0.01	0.01	0.04	0.03	0.02	0.01	0.02	0.06
NNW	0.01	0.00	0.00	0.01	0.01	0.03	0.03	0.01	0.01	0.01	0.03

A.4 National Arboretum Canberra 2015

Figure A.7 shows the record of data collected across the National Arboretum Canberra in 2015. Again, regular gaps in the data record exist due to interruptions in data collection, mostly resulting from battery outages. Data were collected at 1-minute observation intervals throughout the data collection period.

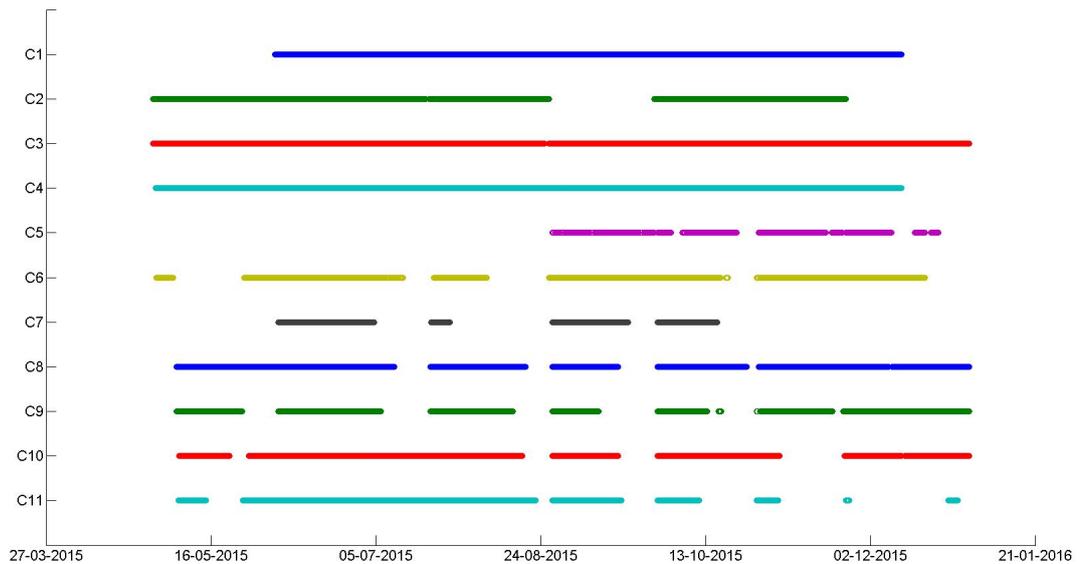


Figure A.7: Data records at the NAC between April and December 2015.

Figure A.8 shows the wind speed time series for each weather station across the National Arboretum Canberra in 2015. The clear slope time series (C1, C2 and C3) clearly show far higher observed wind speeds than the remaining stations within the Radiata pine stand.

Table A.3 summarises the proportions of wind directions observed within each compass sector for each site across the National Arboretum Canberra. Again, due to battery issues, there is a relatively high proportion of ‘Empty’ data points. The percentage of ‘Calm’ observations is considerably higher within the Radiata pine stand than on the cleared slope (up to 79% compared to only 11%).

Table A.3: Proportion of winds originating from each compass direction across the NAC between April and December 2015.

	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11
<i>Empty</i>	0.23	0.29	0.01	0.09	0.66	0.32	0.69	0.17	0.33	0.23	0.43
Calm	0.04	0.08	0.10	0.68	0.24	0.45	0.23	0.48	0.37	0.61	0.29
N	0.01	0.01	0.03	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.00
NNE	0.01	0.02	0.05	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00
NE	0.01	0.05	0.09	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00
ENE	0.05	0.06	0.08	0.00	0.00	0.02	0.00	0.04	0.00	0.00	0.00
E	0.10	0.04	0.05	0.00	0.02	0.04	0.01	0.03	0.00	0.00	0.00
ESE	0.07	0.05	0.05	0.00	0.04	0.05	0.02	0.03	0.03	0.00	0.00
SE	0.09	0.04	0.05	0.00	0.02	0.05	0.03	0.08	0.02	0.02	0.00
SSE	0.01	0.01	0.03	0.00	0.00	0.03	0.01	0.07	0.01	0.03	0.00
S	0.00	0.00	0.02	0.01	0.00	0.01	0.00	0.04	0.01	0.02	0.00
SSW	0.00	0.01	0.02	0.02	0.00	0.00	0.00	0.02	0.01	0.01	0.00
SW	0.00	0.01	0.02	0.05	0.00	0.00	0.00	0.02	0.02	0.01	0.01
WSW	0.00	0.03	0.04	0.03	0.00	0.00	0.00	0.00	0.02	0.02	0.04
W	0.05	0.12	0.11	0.01	0.00	0.00	0.00	0.00	0.13	0.02	0.03
WNW	0.18	0.11	0.11	0.01	0.00	0.00	0.00	0.00	0.04	0.02	0.06
NW	0.11	0.05	0.09	0.02	0.00	0.00	0.00	0.00	0.00	0.01	0.10
NNW	0.03	0.02	0.05	0.03	0.00	0.00	0.00	0.00	0.00	0.00	0.01

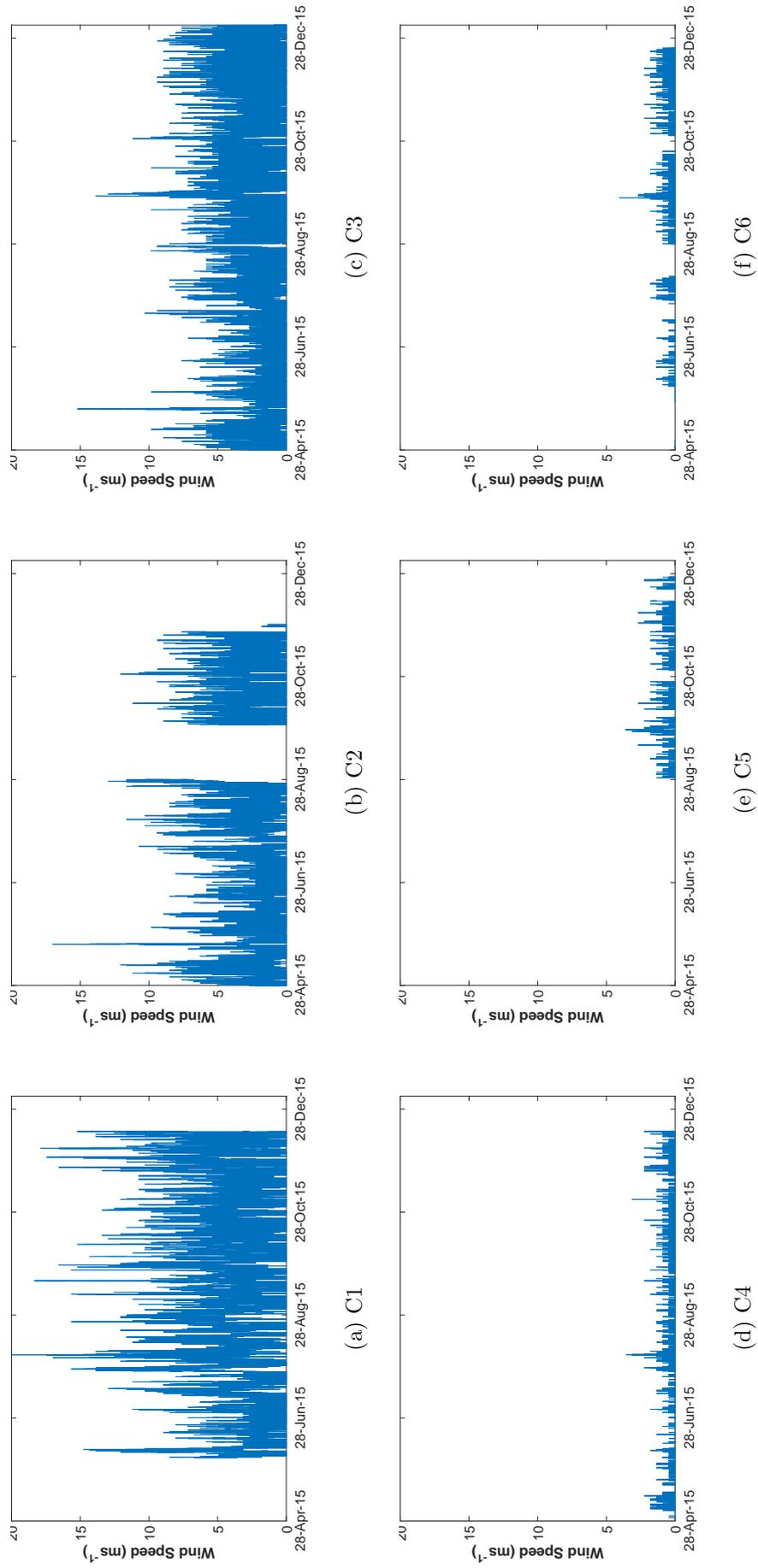


Figure A.8: Time series of observed wind speeds (ms^{-1}) across the NAC at (a-k) C1 to C11 between April and December 2015.

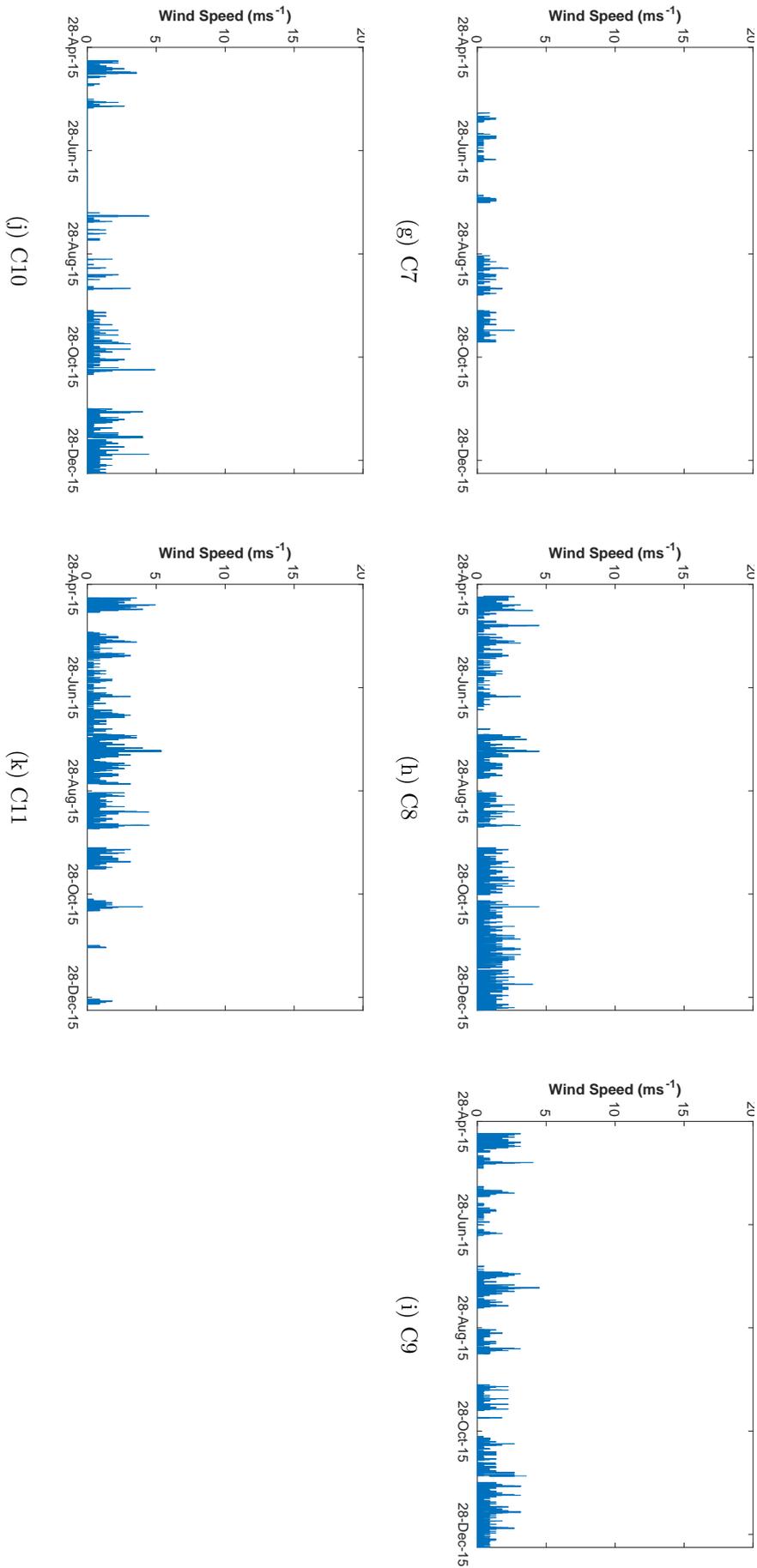


Figure A.8: (Cont'd) Time series of observed wind speeds (ms^{-1}) across the NAC at (a-k) C1 to C11 between April and December 2015.

APPENDIX B

Directional Wind Response

B.1 Flea Creek Valley, 2007

Figure B.1 shows the discrete observed directional wind response distributions for the valley sites (A2 to A5) across Flea Creek Valley between April and June 2007. At each site, the distributions are shown for three minimum ridge top wind speed thresholds, observed at A1; $T = 0\text{ms}^{-1}$, $T = 2\text{ms}^{-1}$ and $T = 4\text{ms}^{-1}$.

B.2 Flea Creek Valley, 2014

Figures B.2 and B.3 show the discrete observed directional wind response distributions for the valley sites (B2 to B11) across Flea Creek Valley between April and December 2014. At each site, the distributions are shown for three minimum ridge top wind speed thresholds, observed at B1; $T = 0\text{ms}^{-1}$, $T = 2\text{ms}^{-1}$ and $T = 4\text{ms}^{-1}$.

B.3 National Arboretum Canberra, 2015

Figures B.4 and B.5 show the discrete observed directional wind response distributions for the sites (C2 to C11) across the National Arboretum Canberra in 2014. At each site, the distributions are shown for three minimum ridge top wind speed thresholds, observed at C1; $T = 0\text{ms}^{-1}$, $T = 2\text{ms}^{-1}$ and $T = 4\text{ms}^{-1}$.

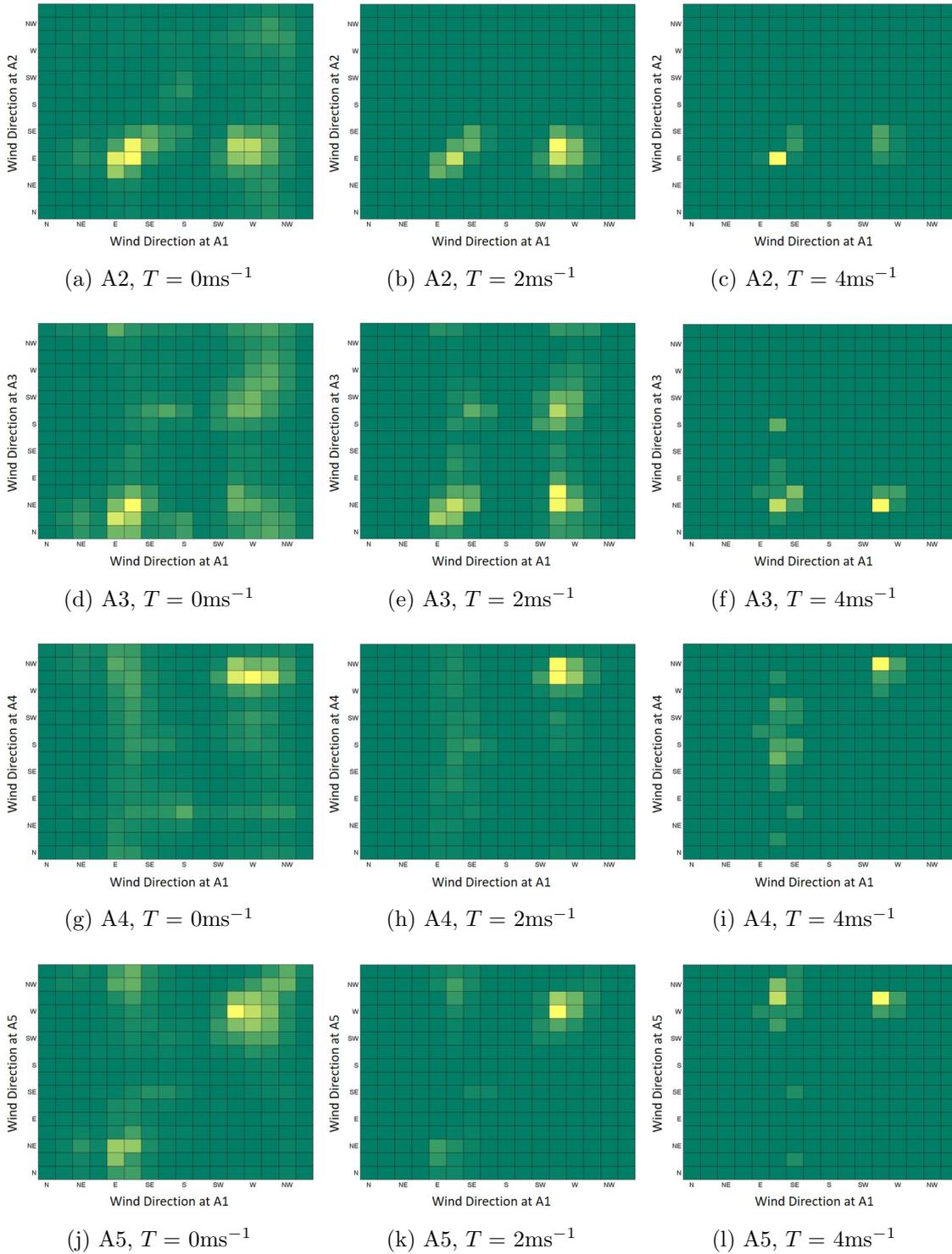


Figure B.1: Observed discrete directional wind response distributions across FCV at (a-l) A2 to A5, with minimum wind speed thresholds of $T = 0\text{ms}^{-1}$, $T = 2\text{ms}^{-1}$ and $T = 4\text{ms}^{-1}$ at A1, between April and June 2007.

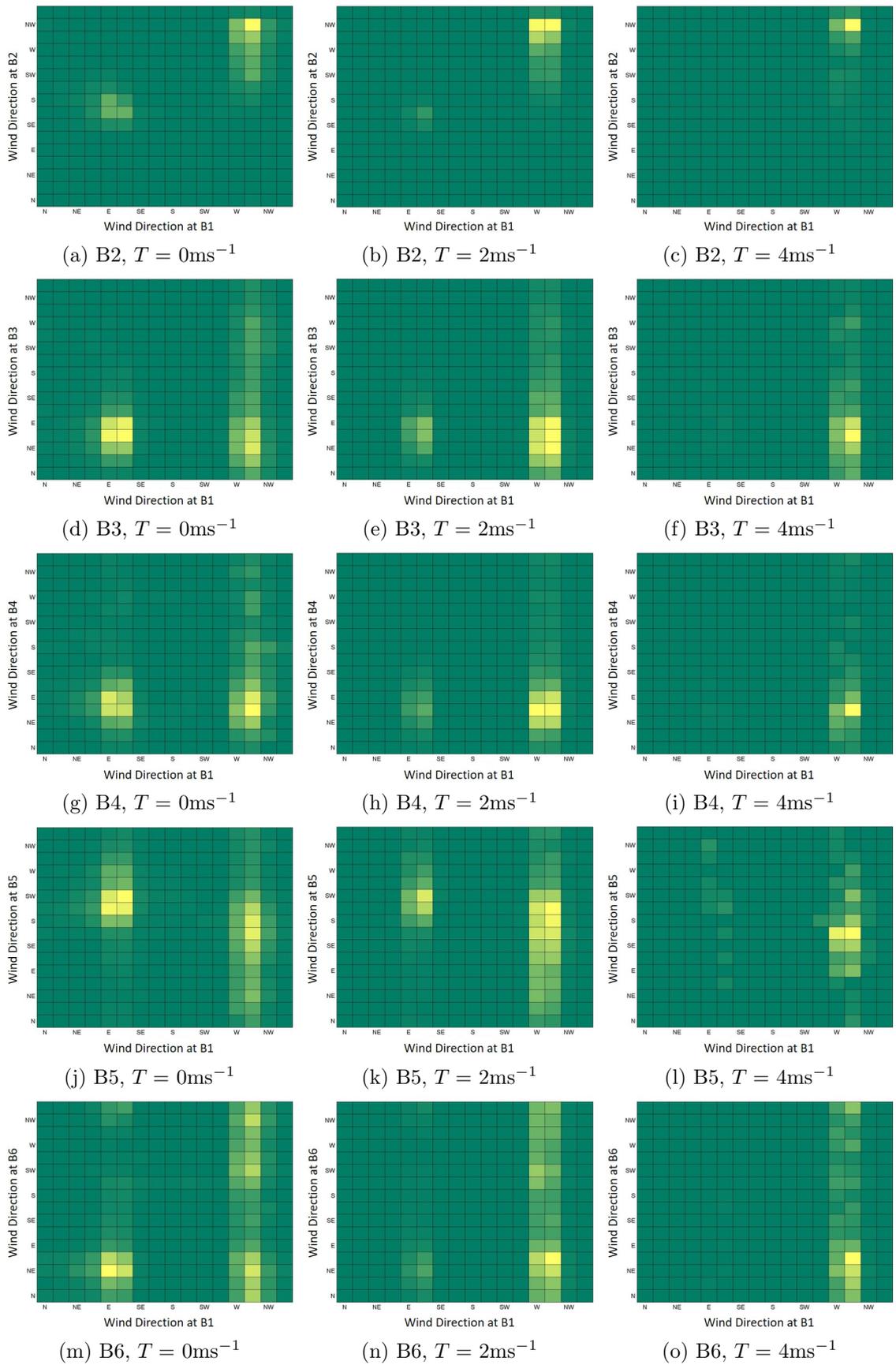


Figure B.2: Observed discrete directional wind response distributions across FCV at (a-o) B2 to B6, with minimum wind speed thresholds of $T = 0\text{ms}^{-1}$, $T = 2\text{ms}^{-1}$ and $T = 4\text{ms}^{-1}$ at B1, between April and December 2014.

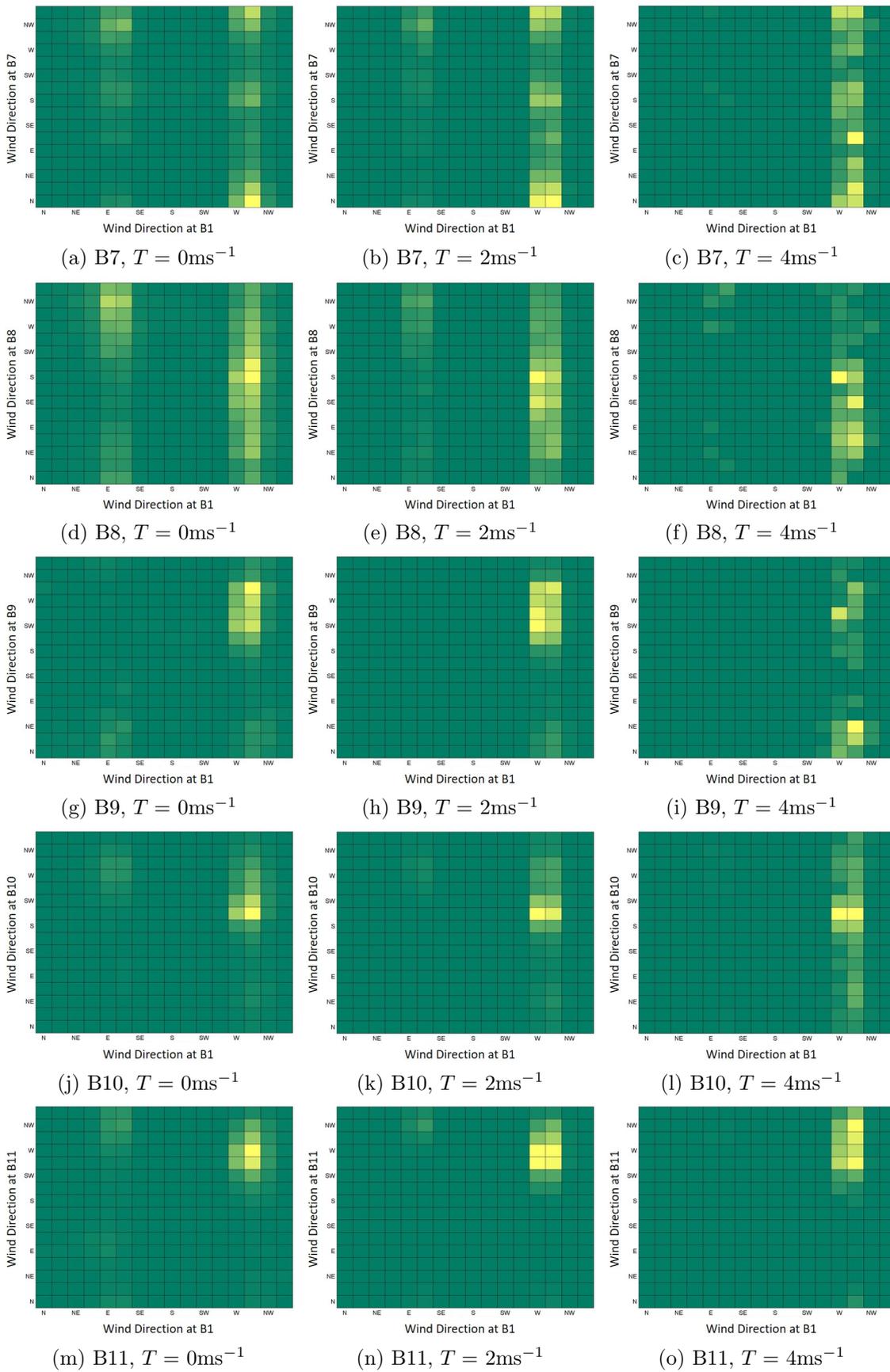


Figure B.3: Observed discrete directional wind response distributions across FCV at (a-o) B7 to B11, with minimum wind speed thresholds of $T = 0\text{ms}^{-1}$, $T = 2\text{ms}^{-1}$ and $T = 4\text{ms}^{-1}$ at B1, between April and December 2014.

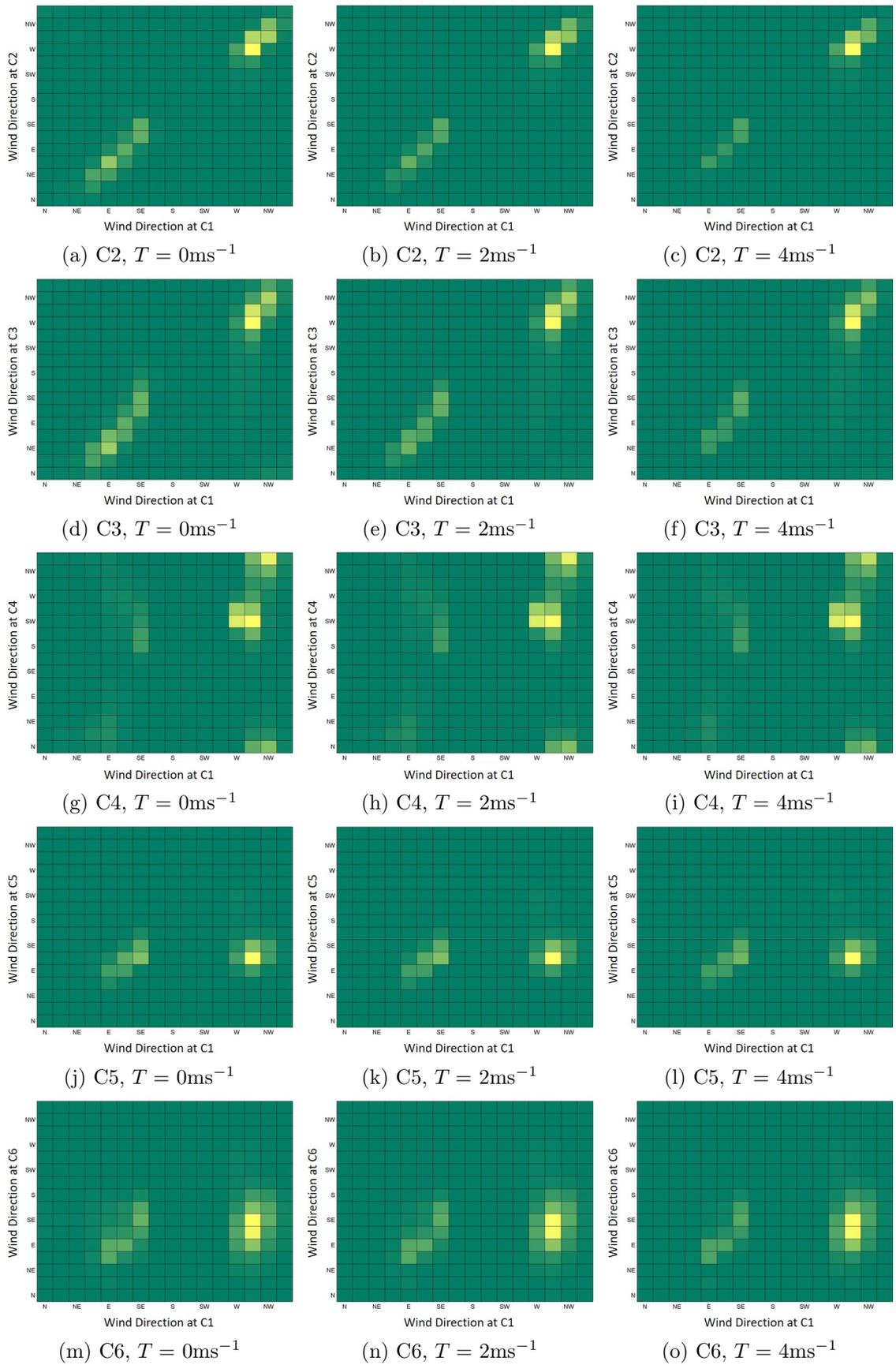


Figure B.4: Observed discrete directional wind response distributions across the NAC at (a-o) C2 to C6, with minimum wind speed thresholds of $T = 0\text{ms}^{-1}$, $T = 2\text{ms}^{-1}$ and $T = 4\text{ms}^{-1}$ at C1, between April and December 2015.

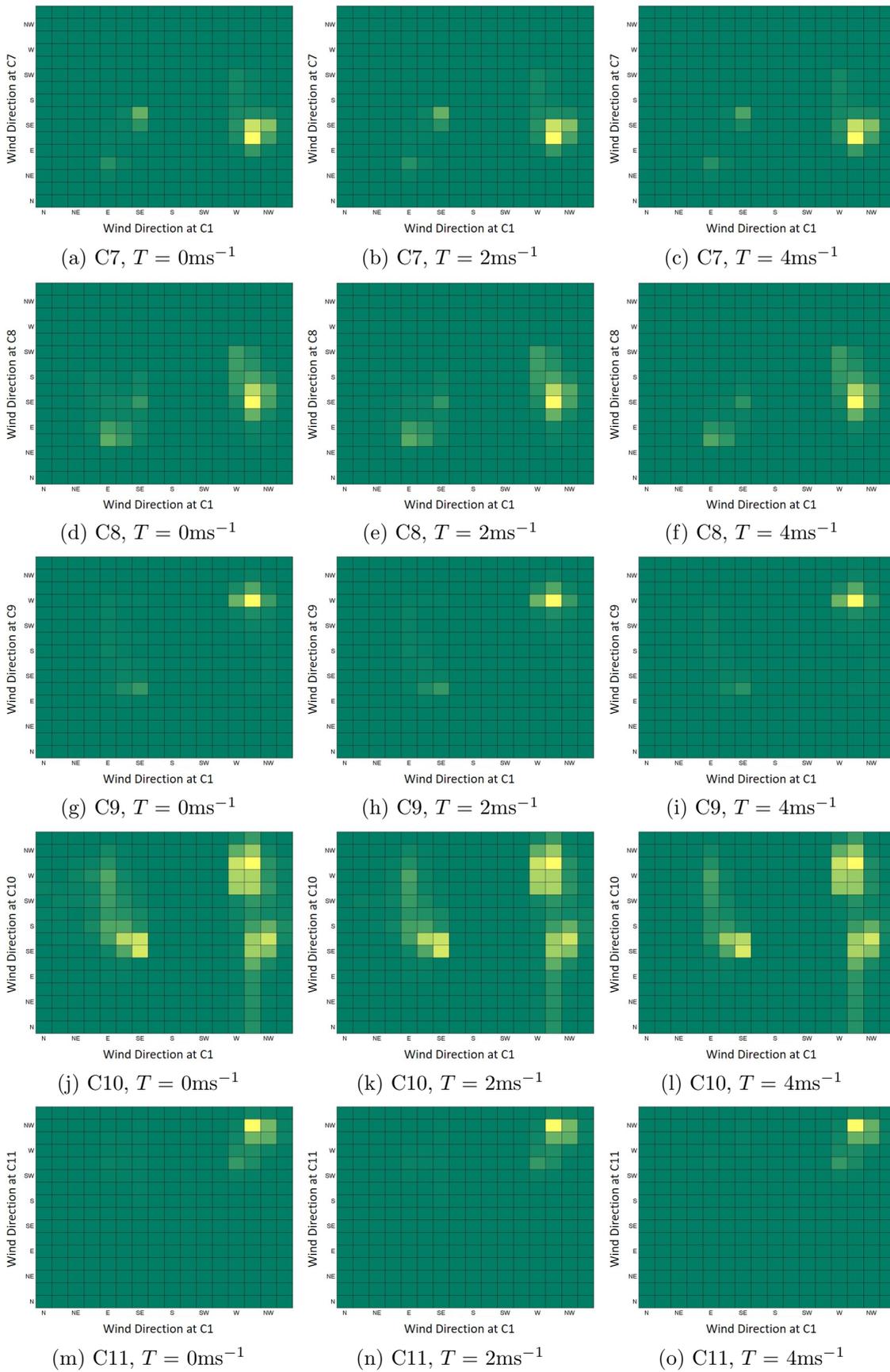


Figure B.5: Observed discrete directional wind response distributions across the NAC at (a-o) C7 to C11, with minimum wind speed thresholds of $T = 0\text{ms}^{-1}$, $T = 2\text{ms}^{-1}$ and $T = 4\text{ms}^{-1}$ at C1, between April and December 2015.

APPENDIX C

Matlab Code for Toroidal Estimation

This appendix outlines the MATLAB® [2012, 2016] code used to analyse and apply the toroidal estimation techniques discussed in Chapter 4.

Initially, the input parameters are defined:

```
s=16; % size of discrete grid
c=360; % size of quasi-continuous grid
k=16; % number of iterations of Algorithm 1
a=250; % number of Monte Carlo runs
```

The ‘true’ continuous distribution is then simulated over a $c \times c$ grid, with no noise and standardised to sum to unity. Within a Monte Carlo scheme, i.e. $\mathbf{b}=1:\mathbf{a}$, the toroidal estimation is conducted using the cubic spline, thin plate smoothing spline or kernel density estimation methods.

First, the discrete ‘observed’ distribution, \mathbf{y} , is simulated with random noise over an $\mathbf{s} \times \mathbf{s}$ grid. Then, for each iteration, i.e. $\mathbf{n}=0:\mathbf{k}$, Algorithm 1 is implemented:

```
% Algorithm 1
yS=zeros(s+(2*n),s+(2*n));
yS(n+1:n+s,n+1:n+s)=y;
yS(s+n+1:s+n+n,:)=yS(n+1:n+n,:);
yS(:,s+n+1:s+n+n)=yS(:,n+1:n+n);
yS(1:n,:)=yS(s+1:s+n,:);
yS(:,1:n)=yS(:,s+1:s+n);
```

Each estimation method is performed on this repeated grid, yS , which is individually reformatted for input into each function. The resulting surface, m , is evaluated over the quasi-continuous $c \times c$ grid and standardised to sum to unity. The following functions are used for the estimation techniques:

```
% Cubic spline with default boundary conditions
dough0 = csape({x1_1,x2_1},yS);
% Cubic spline with periodic boundary conditions
dough0 = csape({x1_1,x2_1},yS,'periodic');
% TPS spline, adapted to give iterative solutions for all data sizes
t=tpaps2(dirxy,dirdat,0.001);
% KDE von Mises, using LCV for k=0 and then fixed
[f,xi,bw]=ksdensity2(kx,kxi,'Kernel','circ_vmpdf',...
                    'bandwidth',[49.9947 49.9947]);
% KDE Normal, using Scott's Rule for k=0 and then fixed
[f,xi,bw]=ksdensity2(kx,kxi,'bandwidth',[42.2090 42.2090]);
```

To calculate the bandwidth for the kernel density estimation, the function `bw.CV` is called from the `NPCirc` package in R [Oliveira et al., 2013, 2014a].

Finally, the difference metrics are calculated for each Monte Carlo run, b :

```
% Mean Squared Error
mse(b,n+1)=mean(mean((m(1:c,1:c)-y(1:c,1:c)).^2));
% Difference in Edge Values
dif_t(b,n+1)=sum(m(1,:)-m(c+1,:))/size(m,1); % rows
dif_l(b,n+1)=sum(m(:,1)-m(:,c+1))/size(m,2); % cols
% Difference in Edge Derivatives
der_t(b,n+1)=sum(diff(m(1:2,:))-diff(m(c+1:c+2,:)))/size(m,1); % rows
der_l(b,n+1)=sum(diff(m(:,1:2),1,2)-...
                diff(m(:,c+1:c+2),1,2))/size(m,2); % cols
```

The mean and standard deviation of these difference metrics for each iteration are calculated and used to plot the figures discussed in Chapter 4.

APPENDIX D

Matlab Code for Surface Comparison Tests

This appendix outlines the MATLAB® [2012, 2016] code used to conduct the simulation studies discussed in Chapter 5.

D.1 Simulation Procedure

Firstly, the following pseudo-code provides the general format of the simulation procedure, as well as the Monte Carlo schemes used to estimate the p -values and power of the test statistics.

```
MC=1000; % number of Monte Carlo runs
% Pre-define matrices for test statistics and p-values.
stat_value=zeros(1,MC); stat_pvalue=zeros(1,MC);
for s=1:MC
    % Simulate standardised distributions with error, e.g. uni Normal.
    y1; y2;
    % Generate two samples based on the distributions, with random
    % sample sizes between 100 and 1000.
    S1; S2
    % Perform the statistical test.
    [stat, p]=test(S1,S2);
    % Allocate statistics and p-values to matrices.
    stat_value(s)=stat; stat_pvalue(s)=p;
    % Calculate p-values using Monte Carlo permutations
```

```
mcp=1000;
stat_mcpvalue=zeros(length(mcp),1);
for sm=1:mcp;
    % Combine data sets and randomly split to calculate KS stat
    % under null of equality.
    A1; A2;
    [stat_m, p_m]=test(A1,A2);
    stat_mc(sm)=stat_m;
end
% Calculate MC p-values for statistic.
stat_pmc=length(stat_mc(stat_mc>stat))/length(stat_mc);
stat_mcpvalue(s)=stat_pmc;
end
% Calculate level of each p-value (asymptotic and MC), e.g. stat_pvalue
level1=length(stat_pvalue(stat_pvalue<0.01))/length(stat_pvalue);
level5=length(stat_pvalue(stat_pvalue<0.05))/length(stat_pvalue);
level10=length(stat_pvalue(stat_pvalue<0.1))/length(stat_pvalue);
```

D.2 Univariate KS Style Tests

The `kstest2` function, available in MATLAB® [2016], was used for the analysis and application of the univariate KS test.

The code used for the analysis and application of the univariate Kuiper's tests was adapted from the `circ_kuipertest` function (available in the `CircStat` package for MATLAB® [2016]). The adaptation added the capacity to calculate the $V_N^{(1)*}$ statistic and p -value.

```
function [d1, d2, v, k, p_k, K, v1, p_v1] = circ_kuipertest2(alpha1,
    alpha2, res, vis_on)
...
% Calculate v1-statistic
v1 = sqrt(n*m/(m+n)) * v;
```

```

% Calculate p-value for v1
pv1 = (8*v1^2 - 2) * exp(-2*v1^2);
if pv1>=0.447; p_v1=1;
elseif pv1<=0; p_v1=0;
else p_v1=pv1;
end

```

D.3 Bivariate KS Style Tests

The code used for the analysis and application of the bivariate KS test was adapted from the function, `kstest_2s_2d`¹, available for MATLAB® [2016], and redefined as `kstest_2s_biv`.

```

function [pValue, KSStatistic, Zn] = kstest_2s_biv(x1, x2, alpha)
...
% A function handle to perform comparisons in all possible directions
fhCounts = @(x, edge)([
sum(x((x(:, 1) >= edge(1)) & (x(:, 2) >= edge(2)),3))...
sum(x((x(:, 1) <= edge(1)) & (x(:, 2) >= edge(2)),3))...
sum(x((x(:, 1) <= edge(1)) & (x(:, 2) <= edge(2)),3))...
sum(x((x(:, 1) >= edge(1)) & (x(:, 2) <= edge(2)),3))]);
KSStatistic = -inf;
for iX = 1:size(x1,1) %(n1+n2)
    % Choose a starting point
    edge = x1(iX,1:2);
    % Estimate the CDFs for both distributions around this point
    vfCDF1 = fhCounts(x1, edge)./sum(x1(:,3));
    vfCDF2 = fhCounts(x2, edge)./sum(x2(:,3));
    % Two-tailed test statistic
    vfThisKSTS = abs(vfCDF1 - vfCDF2);
    fKSTS = max(vfThisKSTS);

```

¹by Qiuyan Peng @ ECE/HKUST, and adapted by Dylan Muir (13th October, 2012). Accessed 18th February 2015, via <http://au.mathworks.com/matlabcentral/fileexchange/38617-two-dimensional-2d-paired-kolmogorov-smirnov-test>

```
% Final test statistic is the max abs diff in CDFs
if (fKSTS > KSstatistic)
    KSstatistic = fKSTS;
end
end
% Gosset Z calculation and P estimation
n = n1 * n2 / (n1 + n2);
Zn = sqrt(n) * KSstatistic;
pValue = 2 * exp(-2 * (Zn - 0.5).^2);
```

The code used for the analysis and application of the bivariate Kuiper's test was adapted from the `kstest_2s_biv` function to calculate the proposed statistics.

```
function [VStat, kStat, VStar] = kuiper_biv(x1, x2, alpha)
...
VStat = -inf;    kStat = -inf;    VStar = -inf;
for iX = 1:size(x1,1) %(n1+n2)
    % Choose a starting point
    edge = x1(iX,1:2);
    % Estimate the CDFs for both distributions around this point
    vfCDF1 = fhCounts(x1, edge)./sum(x1(:,3));
    vfCDF2 = fhCounts(x2, edge)./sum(x2(:,3));
    % Two-tailed test statistic
    dplus = max([0 vfCDF1-vfCDF2]);
    dminus = max([0 vfCDF2-vfCDF1]);
    % Calculate v-statistic
    vfThisV = dplus + dminus;
    ThisV=max(vfThisV);
    % Final test statistic is the maximum stat in CDFs
    if (ThisV > VStat)
        VStat = ThisV;
    end
end
```

```

% Calculate k-statistic
kStat = n1 * n2 * VStat;
% Calculate v1-statistic
VStar = sqrt(n1*n2/(n1+n2)) * VStat;
end

```

D.4 Mean Squared Difference Tests

Finally, the following procedure and MATLAB® [2016] code were used to perform the mean squared difference based tests conducted in Chapter 5.

1. Calculate the observed discrete bivariate histograms, Y_k . Each set is given by $k = 1, 2$, and $k = 3$ represents the equally weighted combined set.
2. Estimate the continuous surface, m_k , for each dataset, using the selected estimation method; Cubic, TPS or Truncated TPS.
3. Calculate the observed test statistics, T_W and T_B , using the following code;

```

% Calculate T_W (Wang and Ye, 2010)
w=360*360;
T_W=(1/w)*sum(sum((m(:, :, 1) - m(:, :, 2)).^2));

% Calculate T_B (Bowman, 2006)
e=zeros(16,16,2);
for k=1:2
    for n=1:16
        for p=1:16
            % Calculate the combined residuals (yS_i - m_3)
            e(n,p,k)=yS(n+1,p+1,k)-m(round(n*22.5),round(p*22.5),3);
        end
    end
end
sigma2=(1/256)*(sum(sum((e(:, :, 1)).^2))+sum(sum((e(:, :, 2)).^2)));
T_B=(1/sigma2)*(sum(sum((m(:, :, 1)-m(:, :, 3)).^2))...
                +sum(sum((m(:, :, 2)-m(:, :, 3)).^2)));

```

4. Perform the bootstrap procedure to construct the estimated distribution of each test statistic: T_W and T_B .

(a) Draw a bootstrap residual, e_k^* , using the following code;

```
% Bootstrap Residual
re=reshape(e(:,:,k),1,16*16);
[bstat,bsam]=bootstrp(1,[],re);
for boot=1:length(re)
    reb(boot)=re(bsam(boot));
end
eb(:,:,k)=reshape(reb,16,16);
```

(b) Construct a bootstrap sample under the null, $Y_k^* = m_3 + e_k^*$.

(c) Estimate the continuous bootstrap surface, m_k^* , using the selected estimation method from (2).

(d) Calculate T_W^* and T_B^* , using the code outlined in (3).

(e) Repeat (a)-(d) 1000 times to construct distributions of T_W^* and T_B^* under the null hypothesis.

(f) Calculate the p -values for each observed statistic; T_W and T_B , according to these distributions.

References

- I. G. Abrahamson. Exact Bahadur efficiencies for the Kolmogorov-Smirnov and Kuiper one- and two-sample statistics. *The Annals of Mathematical Statistics*, 38(5):1475–1490, 1967.
- F. A. Albini. Estimating wildfire behavior and effects. General Technical Report INT-30. Technical report, USDA Forest Service, Intermountain Forest and Range Experiment Station, Ogden, UT, 1976.
- A. Alégria, M. Bevilacqua, and E. Porcu. Likelihood-based inference for multivariate space-time wrapped-Gaussian fields. *Journal of Statistical Computation and Simulation*, 86(13):2583–2597, 2016.
- M. E. Alexander and M. G. Cruz. Evaluating a model for predicting active crown fire rate of spread using wildfire observations. *Canadian Journal of Forest Research*, 36(11):3015–3028, 2006.
- M. E. Alexander and M. G. Cruz. Limitations on the accuracy of model predictions of wildland fire behaviour: A state-of-the-knowledge overview. *The Forestry Chronicle*, 89(03):372–383, 2013.
- T. Allen. Flow over hills with variable roughness. *Boundary-Layer Meteorology*, 121(3):475–490, 2006.
- K. Anderson, G. Reuter, and M. D. Flannigan. Fire-growth modelling using meteorological data with random and systematic perturbations. *International Journal of Wildland Fire*, 16(2):174–182, 2007.
- P. L. Andrews. *Modeling wind adjustment factor and midflame wind speed for Rothermel’s surface fire spread model*. United States Department of Agriculture/Forest Service, Rocky Mountain Research Station, 2012.
- P. L. Andrews, M. Finney, and M. Fischetti. Predicting wildfires. *Scientific American*, (August):47–55, 2007.

- H. Arsham. Kuiper's p -value as a measuring tool and decision procedure for the goodness-of-fit test. *Journal of Applied Statistics*, 15(2):131–135, 1988.
- G. J. Babu and E. D. Feigelson. Astrostatistics: Goodness-of-fit and all that! In *Astronomical Society of the Pacific*, volume 351, page 127, 2-5 October 2005.
- R. M. Banta, G. Berri, W. Blumen, G. A. Carruthers, G. A. Dalu, D. R. Durran, J. Egger, J. R. Garratt, S. R. Hanna, J. C. R. Hunt, R. N. Meroney, W. Miller, W. D. Neff, M. Nicolini, J. Peagle, R. A. Piekle, R. B. Smith, D. G. Strimaitis, T. Vukicevic, and C. D. Whiteman. *Atmospheric Processes Over Complex Terrain*. Boston, Mass.: American Meteorological Society, Boston, Mass., 1990.
- J. M. Bates and C. W. J. Granger. The combination of forecasts. *Journal of the Operational Research Society*, 20(4):451–468, 1969.
- E. Batschelet. *Circular Statistics in Biology*. Mathematics in Biology. Academic Press Inc. (London) Ltd, 1981.
- E. A. Beals. The value of weather forecasts in the problem of protecting forests from fire. *Monthly Weather Review*, 42(2):111–119, 1914.
- J. A. Beck. Equations for the forest fire behaviour tables for Western Australia. *CALM Science*, 1(3):325–348, 1995.
- T. Beer. Bushfire-control decision support systems. *Environment International*, 17(2):101–110, 1991.
- S. E. Belcher, I. N. Harman, and J. J. Finnigan. The wind in the willows: Flows in forest canopies in complex terrain. *Annual Review of Fluid Mechanics*, 44(1):479–504, 2012.
- P. Berens. CircStat: A MATLAB toolbox for circular statistics. *Journal of Statistical Software*, 31(10):1–21, 2009.
- J. Berg, J. Mann, A. Bechmann, M. S. Courtney, and H. E. Jorgensen. The Bolund Experiment, Part I: Flow over a steep, three-dimensional hill. *Boundary-Layer Meteorology*, 141(2):219, 2011.
- V. Berger. A socratic dialogue. *Journal of Modern Applied Statistical Methods*, 8(1):316–321, 2009.
- T. M. Bergin. A comparison of goodness-of-fit tests for analysis of nest orientation in Western Kingbirds (*Tyrannus verticalis*). *The Condor*, 93(1):164–171, 1991.
- C. E. Bonferroni. *Teoria statistica delle classi e calcolo delle probabilita*. Libreria internazionale Seeber, 1936.

- A. W. Bowman. Comparing nonparametric surfaces. *Statistical Modelling*, 6(4): 279–299, 2006.
- D. Boychuk, W. J. Braun, R. J. Kulperger, Z. L. Krougly, and D. A. Stanford. A stochastic model for forest fire growth. *INFOR*, 45(1):9–16, 2007.
- D. Boychuk, W. J. Braun, R. J. Kulperger, Z. L. Krougly, and D. A. Stanford. A stochastic forest fire growth model. *Environmental & Ecological Statistics*, 16(2): 133–151, 2009.
- Bureau of Meteorology. About the ACCESS model. http://www.bom.gov.au/australia/charts/about/about_access.shtml, 2017. Accessed: March 2017.
- N. D. Burrows. *Experimental development of fire management model for Jarrah (Eucalyptus marginata) Forest*. PhD thesis, Australian National University, 1994.
- R. R. Burton, S. B. Vosper, and S. D. Mobbs. Underlying structures in orographic flow: An analysis of measurements made on the Isle of Arran. *Quarterly Journal of the Royal Meteorological Society*, 132(618):1457–1466, 2006.
- B. W. Butler, M. Finney, L. Bradshaw, J. Forthofer, C. McHugh, R. Stratton, and D. Jimenez. WindWizard: A new tool for fire management decision support. In *Proceedings RMRS-P-41*, Fuels management-how to measure success, pages 787–796, Portland, OR, 2006.
- B. W. Butler, N. S. Wagenbrenner, J. M. Forthofer, B. K. Lamb, K. S. Shannon, D. Finn, R. M. Eckman, K. Clawson, L. Bradshaw, P. Sopko, S. Beard, D. Jimenez, C. Wold, and M. Vosburgh. High-resolution observations of the near-surface wind fields over an isolated mountain and in a steep river canyon. *Atmospheric Chemistry and Physics*, 15:3785–3801, 2015.
- R. A. D. Byrnescott. The effects of ridge-top and lee-slope fires upon rotor motions in the lee of a steep ridge. *Mathematical and Computer Modelling*, 13(12):103–112, 1990.
- J. A. Carta, P. Ramirez, and C. Bueno. A joint probability density function of wind speed and direction for wind energy analysis. *Energy Conversion and Management*, 49(6):1309–1320, 2008.
- E. A. Catchpole, W. R. Catchpole, and R. C. Rothermel. Fire behavior experiments in mixed fuel complexes. *International Journal of Wildland Fire*, 3(1):45–57, 1993.

- W. R. Catchpole, R. A. Bradstock, J. Choate, L. G. Fogarty, N. Gellie, G. McCarthy, W. L. McCaw, J. B. Marsden-Smedley, and G. Pearce. Cooperative development of equations for heathland fire behaviour. In *Proceedings of 3rd International Conference on Forest Fire Research and 14th Conference on Fire and Forest Meteorology*, volume 2, pages 16–20, 1998a.
- W. R. Catchpole, E. A. Catchpole, B. W. Butler, R. C. Rothermel, G. A. Morris, and D. J. Latham. Rate of spread of free-burning fires in woody fuels in a wind tunnel. *Combustion Science and Technology*, 131(1-6):1–37, 1998b.
- P. Chaudhuri and J. S. Marron. SiZer for exploration of structures in curves. *Journal of the American Statistical Association*, 94(447):807–823, 1999.
- N. P. Cheney, J. S. Gould, and W. R. Catchpole. The influence of fuel, weather and fire shape variables on fire-spread in grasslands. *International Journal of Wildland Fire*, 3(1):31–44, 1993.
- N. P. Cheney, J. S. Gould, and W. R. Catchpole. Prediction of fire spread in grasslands. *International Journal of Wildland Fire*, 8(1):1–13, 1998.
- N. P. Cheney, J. S. Gould, and L. McCaw. The Dead-Man Zone - a neglected area of firefighter safety. *Australian Forestry*, 64(1):45–50, 2001.
- C.-Y. Chiang and M. L. Puri. Rank-order tests for the parallelism of several regression surfaces. *Journal of Statistical Planning and Inference*, 10(1):43–57, 1984.
- S. N. Chiu and K. I. Liu. Generalized Cramér-von Mises goodness-of-fit tests for multivariate distributions. *Computational Statistics and Data Analysis*, 53(11):3817–3834, 2009.
- R. M. Cionco. A wind-profile index for canopy flow. *Boundary-Layer Meteorology*, 3(2):255–263, 1972.
- H. A. Cleugh. Field measurements of windbreak effects on airflow, turbulent exchanges and microclimates. *Australian Journal of Experimental Agriculture*, 42(6):665–677, 2002.
- S. Coles. Inference for circular distributions and processes. *Statistics and Computing*, 8(2):105–113, 1998.
- W. J. Conover. *Practical Nonparametric Statistics*. John Wiley and Sons, 2nd edition, 1980.
- J. S. Crosby and C. C. Chandler. Get the most from your wind speed observation. *Fire Control Notes*, 27(4):12–13, 1966.

- M. G. Cruz. Monte Carlo-based ensemble method for prediction of grassland fire spread. *International Journal of Wildland Fire*, 19(4):521–530, 2010.
- M. G. Cruz and M. E. Alexander. Uncertainty associated with model predictions of surface and crown fire rates of spread. *Environmental Modelling & Software*, 47(0):16–28, 2013.
- M. G. Cruz and P. M. Fernandes. Development of fuel models for fire behaviour prediction in maritime pine (*Pinus pinaster Ait.*) stands. *International Journal of Wildland Fire*, 17(2):194–204, 2008.
- M. G. Cruz, M. E. Alexander, and P. A. M. Fernandes. Development of a model system to predict wildfire behaviour in pine plantations. *Australian Forestry*, 71(2):113–121, 2008.
- M. G. Cruz, J. Gould, M. E. Alexander, A. Sullivan, L. McCaw, and S. Matthews. *A guide to rate of fire spread models for Australian vegetation*. CSIRO Land and Water Flagship and AFAC, Canberra ACT and Melbourne VIC, 2015.
- W. Cui and A. H. Perera. Quantifying spatio-temporal errors in forest fire spread modelling explicitly. *Journal of Environmental Informatics*, 16(1):19–26, 2010.
- Davis Instruments Corp. Vantage Pro and Vantage Pro2 Serial Communication Reference Manual. Technical Report Rev 2.3 (02/09/2009), 2009.
- P. W. A. Dayananda. Stochastic models for forest fires. *Ecological Modelling*, 3(4):309–313, 1977.
- M. A. Delgado. Testing the equality of nonparametric regression curves. *Statistics and Probability Letters*, 17(3):199–204, 1993.
- H. Dette and N. Neumeier. Nonparametric analysis of covariance. *The Annals of Statistics*, 29(5):1361–1400, 2001.
- S. S. Dhar, B. Chakraborty, and P. Chaudhuri. Comparison of multivariate distributions using quantile-quantile plots and related tests. *Bernoulli*, 20(3):1484–1506, 2014.
- M. Di Marzio, A. Panzera, and C. C. Taylor. Kernel density estimation on the torus. *Journal of Statistical Planning and Inference*, 141(6):2156–2173, 2011.
- M. Di Marzio, A. Panzera, and C. C. Taylor. Non-parametric regression for circular responses. *Scandinavian Journal of Statistics*, 40(2):238–255, 2013.
- T. J. Duff, D. M. Chong, P. Taylor, and K. G. Tolhurst. Procrustes based metrics for spatial validation and calibration of two-dimensional perimeter spread models: A

- case study considering fire. *Agricultural and Forest Meteorology*, 160(0):110–117, 2012.
- T. J. Duff, D. Chong, and K. G. Tolhurst. Quantifying spatio-temporal differences between fire shapes: Estimating fire travel paths for the improvement of dynamic spread models. *Environmental Modelling & Software*, 46(0):33–43, 2013.
- O. J. Dunn. Confidence intervals for the means of dependent, Normally distributed variables. *Journal of the American Statistical Association*, 54(287):613–621, 1959.
- O. J. Dunn. Multiple comparisons among means. *Journal of the American Statistical Association*, 56(293):52–64, 1961.
- J. Durbin. Distribution theory for tests based on the sample distribution function. In *Distribution Theory for Tests Based on the Sample Distribution Function*, pages 1–64. SIAM, Philadelphia, 1973.
- F. Egebrand, M. B. Egerstedt, and C. F. Martin. Smoothing splines on the torus. In *Mathematical Theory of Networks and Systems*, Budapest, Hungary, 2010.
- M. B. Egerstedt and C. F. Martin. *Control Theoretic Splines: Optimal Control, Statistics, and Path Planning*. Princeton Series in Applied Mathematics. Princeton University Press, 2010.
- E. Erdem and J. Shi. Comparison of bivariate distribution construction approaches for analysing wind speed and direction data. *Wind Energy*, 14(1):27–41, 2011.
- ESRI. *ArcGIS Desktop: Release 10*. Environmental Systems Research Institute, Redlands, CA, 2011.
- W. Fan and C. F. Martin. Smoothing splines on the manifold of lines in \mathbb{R}^2 . In *59th World Statistics Congress*, Hong Kong, 2013. International Statistical Institute.
- G. Fasano and A. Franceschini. A multidimensional version of the Kolmogorov-Smirnov test. *Monthly Notices of the Royal Astronomical Society*, 225(1):155–170, 1987.
- W. Feller. On the Kolmogorov-Smirnov limit theorems for empirical distributions. *The Annals of Mathematical Statistics*, 19(2):177–189, 1948.
- W. Feller. Law of large numbers for identically distributed variables. In *An introduction to probability theory and its applications*, volume 2, pages 231–234. 1971.
- S. Ferryanto. A Kolmogorov-Smirnov type statistic for detecting structural changes of textures. *Pattern Recognition Letters*, 16(3):247–256, 1995.

-
- M. Finney. FARSITE: Fire Area Simulator – Model Development and Evaluation. Technical Report Res Pap RMRS-RP-4 Revised, USDA Forest Service, Rocky Mountain Research Station, 2004.
- M. A. Finney, I. C. Grenfell, C. W. McHugh, R. C. Seli, D. Trethewey, R. D. Stratton, and S. Brittain. A method for ensemble wildland fire simulation. *Environmental Modeling and Assessment*, 16(2):153–167, 2011.
- J. Finnigan. Turbulence in plant canopies. *Annual Review of Fluid Mechanics*, 32(1):519–571, 2000.
- N. I. Fisher. *Statistical analysis of circular data*. University Press, Cambridge, first edition, 1993.
- N. I. Fisher and C. M. Powell. Statistical analysis of two-dimensional palaeocurrent data: Methods and examples. *Australian Journal of Earth Sciences*, 36(1):91–107, 1989.
- W. L. Fons. Analysis of fire spread in light forest fuels. *Journal of Agricultural Research*, 72(3):93–121, 1946.
- J. Forthofer and B. Butler. Difference in simulated fire spread over Askervein Hill using two advanced wind models and a traditional uniform wind field. In *Proceeding RMRS-P-46CD*, The fire environment - innovations, management and policy, pages 123–127, Destin, FL, 2007.
- J. M. Forthofer, B. W. Butler, C. W. McHugh, M. A. Finney, L. S. Bradshaw, R. D. Stratton, K. S. Shannon, and N. S. Wagenbrenner. A comparison of three approaches for simulating fine-scale surface winds in support of wildland fire management. Part II. An exploratory study of the effect of simulated winds on fire growth simulations. *International Journal of Wildland Fire*, 23(7):982–994, 2014a.
- J. M. Forthofer, B. W. Butler, and N. S. Wagenbrenner. A comparison of three approaches for simulating fine-scale surface winds in support of wildland fire management. Part I. Model formulation and comparison against measurements. *International Journal of Wildland Fire*, 23(7):969–981, 2014b.
- W. H. Frandsen. Fire spread through porous fuels from the conservation of energy. *Combustion and Flame*, 16(1):9–16, 1971.

- L. S. Freedman. The use of a Kolmogorov-Smirnov type statistic in testing hypotheses about seasonal variation. *Journal of Epidemiology and Community Health (1979-)*, 33(3):223–228, 1979.
- I. French, B. Cechet, T. Yang, and A. Sanabria. FireDST: Fire Impact and Risk Evaluation Decision Support Tool - model description. In J. Piantadosi, R. S. Aderssen, and J. Boland, editors, *MODSIM2013, 20th International Congress on Modelling and Simulation*, Adelaide, Australia, 2013. Modelling and Simulation Society of Australia and New Zealand Inc.
- I. French, M. Woolf, B. Cechet, T. Yang, and A. Sanabria. Acknowledging and understanding variability in simulating bushfires. Part 1 - Evaluation of FireDST against the Kilmore fire of 7 February 2009. Technical report, Geoscience Australia, 2014a.
- I. French, M. Woolf, B. Cechet, T. Yang, and A. Sanabria. Acknowledging and understanding variability in simulating bushfires. Part 2 - Evaluation of FireDST against the Wangary fire of 10 January 2005. Technical report, Geoscience Australia, 2014b.
- I. French, M. Woolf, B. Cechet, T. Yang, and A. Sanabria. Acknowledging and understanding variability in simulating bushfires. Part 3 - Evaluation of FireDST against the Mt Hall fire of 24 December 2001. Technical report, Geoscience Australia, 2014c.
- I. A. French, T. J. Duff, R. P. Cechet, K. G. Tolhurst, J. D. Kepert, and M. Meyer. FireDST: a simulation system for short-term ensemble modelling of bushfire spread and exposure. In *Advances in Forest Fire Research*. Imprensa da Universidade de Coimbra, Coimbra, 2014d.
- E. García-Portugués, R. Crujeiras, and W. González-Manteiga. Exploring wind direction and SO₂ concentration by circular-linear density estimation. *Stochastic Environmental Research & Risk Assessment*, 27(5):1055–1067, 2013a.
- E. García-Portugués, R. M. Crujeiras, and W. González-Manteiga. Kernel density estimation for directional-linear data. *Journal of Multivariate Analysis*, 121:152–175, 2013b.
- R. Garcia-Rojo. Algorithm for the estimation of the long-term wind climate at a meteorological mast using a joint probabilistic approach. *Wind Engineering*, 28(2):213–223, 2004.

- A. Garratt, J. A. Mitchell, S. P. Vahey, and E. C. Wakerly. Real-time inflation forecast densities from ensemble Phillips curves. *The North American Journal of Economics and Finance*, 22(1):77–87, 2011.
- M. J. Glanville and K. C. S. Kwok. Measurements of topographic multipliers and flow separation from a steep escarpment. Part II. Model-scale measurements. *Journal of Wind Engineering and Industrial Aerodynamics*, 6971(0):893–902, 1997.
- T. Gneiting and A. E. Raftery. Weather forecasting with ensemble methods. *Science*, 310(5746):248–249, 2005.
- T. N. T. Goodman, S. L. Lee, and A. Sharma. Approximation and interpolation by complex splines on the torus. *Proceedings of the Edinburgh Mathematical Society (Series 2)*, 32(02):197–212, 1989.
- C. P. Goodyear. Terrestrial and aquatic orientation in the Starhead Topminnow, *Fundulus Notti*. *Science*, 168(3931):603–605, 1970.
- E. Gosset. A three-dimensional extended Kolmogorov-Smirnov test as a useful tool in astronomy. *Astronomy and Astrophysics*, 188(1):258–264, 1987.
- J. S. Gould. *Project Vesta Fuel Structure, Fuel Dynamics and Fire Behaviour*. Fire in Dry Eucalypt Forest. Melbourne: CSIRO Publishing, Melbourne, 2008.
- G. Grenouillet, L. Buisson, N. Casajus, and S. Lek. Ensemble modelling of species distribution: The effects of geographical and environmental ranges. *Ecography*, 34(1):9–17, 2011.
- G. Gross and F. Wippermann. Channeling and countercurrent in the Upper Rhine valley: Numerical simulations. *Journal of Climate and Applied Meteorology*, 26(10):1293–1304, 1987.
- J. H. A. Guillaume, R. J. Hunt, A. Comunian, R. S. Blakers, and B. Fu. Methods for exploring uncertainty in groundwater management predictions. In A. J. Jakeman, O. Barreteau, R. J. Hunt, J.-D. Rinaudo, and A. Ross, editors, *Integrated Groundwater Management: Concepts, Approaches and Challenges*, pages 711–737. Springer Nature, SpringerLink.com, 2016.
- L. Han and J. W. Braun. Dionysus: a stochastic fire growth scenario generator. *Environmetrics*, 25(6):431–442, 2014.

- P. A. Hancock and M. F. Hutchinson. Spatial interpolation of large climate data sets using bivariate thin plate smoothing splines. *Environmental Modelling & Software*, 21(12):1684–1694, 2006.
- J. L. Hodges. The significance probability of the Smirnov two-sample test. *Arkiv for Matematik*, 3(5):469–486, 1957.
- J. D. Holmes, R. W. Banks, and P. Paevere. Measurements of topographic multipliers and flow separation from a steep escarpment. Part I. Full scale measurements. *Journal of Wind Engineering and Industrial Aerodynamics*, 6971(0):885–892, 1997.
- M. F. Hutchinson. Interpolating mean rainfall using thin plate smoothing splines. *International Journal of Geographical Information Systems*, 9(4):385–403, 1995.
- M. F. Hutchinson and P. E. Gessler. Splines-more than just a smooth interpolator. *Geoderma*, 62(1):45–67, 1994.
- S. R. Jammalamadaka and A. SenGupta. *Topics in Circular Statistics-Vol 5*. World Scientific Publishing Company, River Edge, United States, 2001.
- R. A. Johnson and T. E. Wehrly. Some angular-linear distributions and related regression models. *Journal of the American Statistical Association*, 73(363):602–606, 1978.
- P. E. Jupp. Data-driven tests of uniformity on product manifolds. *Journal of Statistical Planning and Inference*, 139(11):3820–3829, 2009.
- A. Justel, D. Pea, and R. Zamar. A multivariate Kolmogorov-Smirnov test of goodness of fit. *Statistics and Probability Letters*, 35(3):251–259, 1997.
- J. J. Keetch. Quality control in fire danger rating. *Fire Control Notes*, 27(4):11–16, 1966.
- M. Kendall and A. Stuart. *The Advanced Theory of Statistics*, volume 2: Inference and Relationship. Charles Griffin and Company Ltd, London and High Wycombe, fourth edition, 1979.
- J. T. Kent and D. E. Tyler. Maximum likelihood estimation for the wrapped Cauchy distribution. *Journal of Applied Statistics*, 15(2):247–254, 1988.
- J. T. Kent, K. V. Mardia, and C. C. Taylor. Modelling strategies for bivariate circular data. In *Proceedings of the Leeds Annual Statistical Research Conference, The Art and Science of Statistical Bioinformatics*, Leeds University Press, Leeds, pages 70–73, 2008.

- C. Koen and I. Siluyele. Multivariate comparisons of the period-light-curve shape distributions of Cepheids in five galaxies. *Monthly Notices of the Royal Astronomical Society*, 377(3):1281–1286, 2007.
- A. N. Kolmogorov. Sulla determinazione empirica di una legge di distribuzione. *Giornale dell' Istituto Italiano degli Attuari*, 4:1–11, 1933.
- P. Kourtz. Probability makes fire danger index more reliable. *Fire Control Notes*, 33(4):11–12, 1972.
- J. A. Koziol. Percentage points of the asymptotic distributions of one and two sample Kuiper statistics for truncated or censored data. *Technometrics*, 22(3):437–442, 1980.
- N. H. Kuiper. Tests concerning random points on a circle. *Indagationes Mathematicae (Proceedings)*, 63:38–47, 1960.
- M. Lackner and C. Elkinton. An analytical framework for offshore wind farm layout optimization. *Wind Engineering*, 31(1):17–31, 2007.
- F. Lagona and M. Picone. Model-based segmentation of spatial cylindrical data. *Journal of Statistical Computation and Simulation*, 86(13):2598–2610, 2016.
- S. L. Lee and W. S. Tang. Complex homogeneous splines on the torus. *Approximation Theory and its Applications*, 5(4):31–42, 1989.
- B. W. Lindgren. *Statistical Theory*. Macmillan Publishing Co., Inc, New York, third edition, 1976.
- A. M. G. Lopes. WindStation: a software for the simulation of atmospheric flows over complex topography. *Environmental Modelling & Software*, 18(1):81–96, 2003.
- A. S. Louter and J. Koerts. On the Kuiper test for normality with mean and variance unknown. *Statistica Neerlandica*, 24(2):83–87, 1970.
- U. R. Maag. The asymptotic expansion of the distribution of the goodness of fit statistic V_{NN} . *The Annals of Statistics*, 1(6):1185–1188, 1973.
- U. R. Maag and G. Dicaire. On Kolmogorov-Smirnov type one-sample statistics. *Biometrika*, 58(3):653–656, 1971.
- U. R. Maag and M. A. Stephens. The V_{NM} two-sample test. *The Annals of Mathematical Statistics*, 39(3):923–935, 1968.
- J. Mandel, J. D. Beezley, J. L. Coen, and M. Kim. Data assimilation for wildland fires. *IEEE Control Systems*, 29(3):47–65, 2009.

- J. Mandel, J. D. Beezley, and A. K. Kochanski. Coupled atmosphere-wildland fire modeling with WRF 3.3 and SFIRE 2011. *Geoscientific Model Development*, 4(3):591–610, 2011.
- K. V. Mardia. Statistical approaches to three key challenges in protein structural bioinformatics. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 62(3):487–514, 2013.
- K. V. Mardia and P. E. Jupp. *Directional Statistics*. Wiley Series in Probability and Statistics. John Wiley and Sons, Ltd, Chichester, New York, Weinheim, Brisbane, Singapore, Toronto, 2000.
- K. V. Mardia, C. C. Taylor, and G. K. Subramaniam. Protein bioinformatics and mixtures of bivariate von Mises distributions for angular data. *Biometrics*, 63(2):505–512, 2007.
- K. V. Mardia, J. T. Kent, G. Hughes, and C. C. Taylor. Maximum likelihood estimation using composite likelihoods for closed exponential families. *Biometrika*, 2009.
- J. B. Marsden-Smedley. *Fuel characteristics and fire behavior in Tasmanian Buttongrass Moorlands*. Department of Environment & Land Management, Parks & Wildlife Service, 1993.
- J. B. Marsden-Smedley and W. R. Catchpole. Fire behaviour modelling in Tasmanian buttongrass moorlands. II. Fire behaviour. *International Journal of Wildland Fire*, 5(4):215–228, 1995.
- J. Massey, F. J. The Kolmogorov-Smirnov test for goodness of fit. *Journal of the American Statistical Association*, 46(253):68–78, 1951.
- J. Mathew. Large eddy simulation. *Defence Science Journal*, 60(6):598–605, 2010.
- MATLAB®. *2012b*. The MathWorks, Inc., Natick, Massachusetts, United States., 2012.
- MATLAB®. *2016b*. The MathWorks, Inc., Natick, Massachusetts, United States., 2016.
- A. G. McArthur. Weather and grassland fire behaviour. Technical Report Leaflet No. 100, Forestry and Timber Bureau, 1966.
- A. G. McArthur. Fire behaviour in eucalypt forests. Technical Report Leaflet No.107, Forestry and Timber Bureau, 1967.

- A. G. McArthur. The Tasmanian bushfires on 7th February, 1967, and associated fire behaviour characteristics. In *Technical Co-operation Programme. Mass Fire Symposium*, volume 1, page 23. Maribyrnong, Victoria: Defence Standards Laboratories, 1969.
- A. G. McArthur. Fire Danger Rating Systems. Technical Report FAO Document FO: FFM/77/3e01, Food and Agriculture Organization of United Nations, Rome, Italy, 1977.
- L. McCaw. Research as a basis for fire management in mallee-heath shrublands of south-western Australia. In *Proceedings of the 3rd International Conference on Forest Fire Research and 14th Fire and Forest Meteorology Conference.* (Ed. DX Viegas), pages 2335–2348, 1998.
- W. L. McCaw. *Predicting fire spread in Western Australian mallee-heath scrubland.* PhD thesis, University of New South Wales, Australian Defence Force Academy, 1997.
- R. H. D. McRae and J. J. Sharples. A conceptual framework for assessing the risk posed by extreme bushfires. *Australian Journal of Emergency Management*, 26(2):47–53, 2011.
- R. H. D. McRae and J. J. Sharples. A process model for forecasting conditions conducive to blow-up fire events. In J. Piantadosi, R. S. Anderssen, and J. Boland, editors, *MODSIM2013, 20th International Congress on Modelling and Simulations*, pages pp. 2506–2512. Modelling and Simulation Society of Australia and New Zealand, 2013.
- B. McWilliams and D. Sprevak. The estimation of the parameters of the distribution of wind speed and direction. *Wind Engineering*, 4:227–238, 1980.
- B. McWilliams, M. M. Newmann, and D. Sprevak. The probability distribution of wind velocity and direction. *Wind Engineering*, 3(4):269–273, 1979.
- J. Meinguet. Multivariate interpolation at arbitrary points made simple. *Zeitschrift für angewandte Mathematik und Physik ZAMP*, 30(2):292–304, 1979.
- W. Mell, A. Maranghides, R. McDermott, and S. L. Manzella. Numerical simulation and experiments of burning douglas fir trees. *Combustion and Flame*, 156:2023–2041, 2009.
- N. Metropolis and S. Ulam. The Monte Carlo method. *Journal of the American Statistical Association*, 44(247):335–341, 1949.

- J. A. Mitchell. Rule of thumb for determining rate of spread. *Fire Control Notes*, 1(7):395–396, 1937.
- S. D. Mobbs, S. B. Vosper, P. F. Sheridan, R. Cardoso, R. R. Burton, S. J. Arnold, M. K. Hill, V. Horlacher, and A. M. Gadian. Observations of downslope winds and rotors in the Falkland Islands. *Quarterly Journal of the Royal Meteorological Society*, 131(605):329–351, 2005.
- F. Molteni, R. Buizza, T. N. Palmer, and T. Petroliaigis. The ECMWF ensemble prediction system: Methodology and validation. *Quarterly Journal of the Royal Meteorological Society*, 122(529):73–119, 1996.
- K. Moon, T. J. Duff, and K. G. Tolhurst. Characterising forest wind profiles for utilisation in fire spread models. In *20th International Congress on Modelling and Simulation*, Adelaide, Australia, 2013.
- K. Moon, T. J. Duff, and K. G. Tolhurst. Sub-canopy forest winds: understanding wind profiles for fire behaviour simulation. *Fire Safety Journal*, In Press. doi: <http://dx.doi.org/10.1016/j.firesaf.2016.02.005>.
- G. E. Moore, L. B. Milich, and M. K. Liu. Plume behaviors observed using lidar and SF6 tracer at a flat and hilly site. *Atmospheric Environment (1967)*, 22(8): 1673–1688, 1988.
- H. Neave and P. Worthington. *Distribution-free Tests*. Unwin Hyman Ltd, London, UK, 1988.
- M. Neuhauser. *Nonparametric Statistical Tests: A Computational Approach*. CRC Press, Taylor & Francis Group, FL, USA, 2012.
- I. R. Noble, A. M. Gill, and G. A. V. Bary. McArthur’s fire-danger meters expressed as equations. *Australian Journal of Ecology*, 5(201-203), 1980.
- M. Oliveira, R. M. Crujeiras, and A. Rodríguez-Casal. A plug-in rule for bandwidth selection in circular density estimation. *Computational Statistics and Data Analysis*, 56(12):3898–3908, 2012.
- M. Oliveira, R. M. Crujeiras, and A. Rodríguez-Casal. Nonparametric circular methods for exploring environmental data. *Environmental and Ecological Statistics*, 20(1):1–17, 2013.
- M. Oliveira, R. M. Crujeiras, and A. Rodríguez-Casal. NPCirc: An R package for nonparametric circular methods. *Journal of Statistical Software*, 61(9):26, 2014a.

- M. Oliveira, R. M. Crujeiras, and A. Rodríguez-Casal. CircSiZer: an exploratory tool for circular data. *Environmental and Ecological Statistics*, 21(1):143–159, 2014b.
- A. Panzera and C. C. Taylor. Nonparametric smoothing of circular data. In *46th Scientific Meeting of the Italian Statistical Society*, 2012.
- E. Pastor, L. Zarate, E. Planas, and J. Arnaldos. Mathematical models and calculation systems for the study of wildland fire behaviour. *Progress in Energy and Combustion Science*, 29(2):139–153, 2003.
- D. A. Paterson and J. D. Holmes. Computation of wind flow over topography. *Journal of Wind Engineering and Industrial Aerodynamics*, 4647(0):471–476, 1993.
- M. Peace, T. Mattner, G. A. Mills, J. D. Kepert, and L. McCaw. Fire-modified meteorology in a coupled fire-atmosphere model. *Journal of Applied Meteorology and Climatology*, 54(3):704–720, 2015.
- J. A. Peacock. Two-dimensional goodness-of-fit testing in astronomy. *Monthly Notices of the Royal Astronomical Society*, 202(Feb 1983):615–627, 1983.
- E. S. Pearson. Comparison of tests for randomness of points on a line. *Biometrika*, 50(3/4):315–325, 1963.
- T. D. Penman, L. Collins, O. F. Price, R. A. Bradstock, S. Metcalf, and D. M. O. Chong. Examining the relative effects of fire weather, suppression and fuel treatment on fire behaviour - A simulation study. *Journal of Environmental Management*, 131(0):325–333, 2013.
- A. Pewsey and S. Kato. Parametric bootstrap goodness-of-fit testing for Wehrly-Johnson bivariate circular distributions. *Statistics and Computing*, 26(6):1307–1317, 2016.
- M. C. Potter, D. C. Wiggert, M. Hondzo, and T. I.-P. Shih. *Mechanics of Fluids*. Brooks/Cole Cengage Learning, third edition, 2002.
- H. K. Preisler, D. R. Brillinger, R. E. Burgan, and J. W. Benoit. Probability based models for estimation of wildfire risk. *International Journal of Wildland Fire*, 13(2):133–142, 2004.
- Queensland Fire and Emergency Services. 2014-2015 Annual Report. Technical report, The State of Queensland, Queensland Fire and Emergency Services, 2015.
- M. Raghuvaran. Limiting distributions of Kolmogorov-Smirnov type statistics under the alternative. *The Annals of Statistics*, 1(1):67–73, 1973.

- R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2016.
- R. G. Rehm and W. Mell. A simple model for wind effects of burning structures and topography on wildland-urban interface surface-fire propagation. *International Journal of Wildland Fire*, 18(3):290–301, 2009.
- G. D. Richards. The properties of elliptical wildfire growth for time dependent fuel and meteorological conditions. *Combustion Science and Technology*, 95(1-6):357–383, 1993.
- A. N. Ross. Large-eddy simulations of flow over forested ridges. *Boundary-Layer Meteorology*, 128(1):59–76, 2008.
- A. N. Ross and S. B. Vosper. Neutral turbulent flow over forested hills. *Quarterly Journal of the Royal Meteorological Society*, 131(609):1841–1862, 2005.
- R. C. Rothermel. A mathematical model for predicting fire spread in wildland fuels. Research Paper INT-115. Technical report, US Department of Agriculture, Forest Service, Intermountain Forest and Range Experiment Station, Ogden, UT., 1972.
- R. C. Rothermel. How to predict the spread and intensity of forest and range fires. General Technical Report INT-143. Technical report, US Department of Agriculture, Forest Service, Intermountain Forest and Range Experiment Station, Ogden, UT., 1983.
- L. A. Salazar and L. S. Bradshaw. Display and interpretation of fire behavior probabilities for long-term planning. *Environmental Management*, 10(3):393–402, 1986.
- J. J. Sharples. Review of formal methodologies for wind–slope correction of wildfire rate of spread. *International Journal of Wildland Fire*, 17:179–193, 2008.
- J. J. Sharples, R. H. D. McRae, and R. O. Weber. Wind characteristics over complex terrain with implications for bushfire risk management. *Environmental Modelling & Software*, 25(10):1099–1120, 2010.
- J. J. Sharples, R. H. D. McRae, and S. R. Wilkes. Wind-terrain effects on the propagation of wildfires in rugged terrain: fire channelling. *International Journal of Wildland Fire*, 21(3):282–296, 2012.
- B. W. Silverman. *Density Estimation for Statistics and Data Analysis*. Chapman and Hall Ltd, 1986.

-
- C. C. Simpson, J. J. Sharples, J. P. Evans, and M. F. McCabe. Large eddy simulation of atypical wildland fire spread on leeward slopes. *International Journal of Wildland Fire*, 22(5):599–614, 2013.
- C. C. Simpson, J. J. Sharples, and J. P. Evans. Sensitivity of atypical lateral fire spread to wind and slope. *Geophysical Research Letters*, 43(4):1744–1751, 2016.
- H. Singh, V. Hnizdo, and E. Demchuk. Probabilistic model for two dependent circular variables. *Biometrika*, 89(3):719–723, 2002.
- M. J. Slakter. A comparison of the Pearson chi-square and Kolmogorov goodness-of-fit tests with respect to validity. *Journal of the American Statistical Association*, 60(311):854–858, 1965.
- N. V. Smirnov. On the estimation of the discrepancy between empirical curves of distribution for two independent samples. *Moscow University Mathematics Bulletin*, 2(2), 1939.
- O. E. Smith. Application of distributions derived from bivariate Normal density function. In *Bulletin of the American Meteorological Society*, volume 52, page 222. American Meteorological Society, 45 Beacon St, Boston, MA 02108-3693, 1971.
- R. Srihera and W. Stute. Nonparametric comparison of regression functions. *Journal of Multivariate Analysis*, 101(9):2039–2059, 2010.
- R. Srinivasan. On the Kuiper test for normality with mean and variance unknown. *Statistica Neerlandica*, 25(3):153–157, 1971.
- G. P. Steck. The Smirnov two sample tests as rank tests. *The Annals of Mathematical Statistics*, 40(4):1449–1466, 1969.
- M. A. Stephens. The goodness-of-fit statistic V_n : Distribution and significance points. *Biometrika*, 52(3/4):309–321, 1965.
- M. A. Stephens. Results from the relation between two statistics of the Kolmogorov-Smirnov type. *The Annals of Mathematical Statistics*, 40(5):1833–1837, 1969.
- M. A. Stephens. Use of the Kolmogorov-Smirnov, Cramér-von Mises and related statistics without extensive tables. *Journal of the Royal Statistical Society. Series B (Methodological)*, 32(1):115–122, 1970.
- A. L. Sullivan and I. K. Knight. Estimating error in wind speed measurements for experimental fires. *Canadian Journal of Forest Research*, 31(3):401–409, 2001.

- A. L. Sullivan, J. J. Sharples, S. Matthews, and M. P. Plucinski. A downslope fire spread correction factor based on landscape-scale fire behaviour. *Environmental Modelling & Software*, 62:153–163, 2014.
- C. C. Taylor. Automatic bandwidth selection for circular density estimation. *Computational Statistics & Data Analysis*, 52(7):3493–3500, 2008.
- C. C. Taylor, K. V. Mardia, M. Di Marzio, and A. Panzera. Validating protein structure using kernel density estimates. *Journal of Applied Statistics*, 39(11):2379–2388, 2012.
- P. A. Taylor and H. W. Teunissen. The Askervein Hill project: Overview and background data. *Boundary-Layer Meteorology*, 39(1):15–39, 1987.
- K. G. Tolhurst, B. Shields, and D. Chong. Phoenix: development and application of a bushfire risk management tool. *Australian Journal of Emergency Management*, 23(4):47–54, 2008.
- J. S. Touma. Dependence of the wind profile power law on stability for various locations. *Journal of the Air Pollution Control Association*, 27(9):863–866, 1977.
- B. Twomey and A. Sturgess. Simulation Analysis-based Risk Evaluation (SABRE) fire: operational stochastic fire spread decision support capability in the Queensland Fire and Emergency Service. In *AFAC 2016*, Brisbane, 2016.
- C. Tymstra, R. W. Bryce, B. M. Wotton, S. W. Taylor, and O. B. Armitage. Development and structure of Prometheus: the Canadian wildland fire growth simulation model. *Natural Resources Canada, Canadian Forest Service, Northern Forestry Centre, Information Report NOR-X-417*. (Edmonton, AB), 2010.
- G. J. G. Upton and B. Fingleton. *Spatial Data Analysis by Example*. Wiley Series in Probability and Mathematical Statistics. John Wiley & Sons, 1989.
- US Forest Service. Wildland-Urban Fire Models: Overview of the Wildland-Urban Fire Dynamics Simulator (WFDS). <https://www.fs.fed.us/pnw/fera/wfds/index.shtml>, 2013. Accessed: March 2017.
- C. E. Van Wagner. Prediction of crown fire behaviour in conifer stands. In *10th Conference on Fire and Forest Meteorology*, pages 207–213, Ottawa, Canada, 1989.
- H. Verdoux, N. Takei, R. Cassou de Saint-Mathurin, and M. Bourgeois. Analysis of the seasonal variation of schizophrenic births using a Kolmogorov-Smirnov type statistic. *European Psychiatry*, 12(3):111–116, 1997.

-
- S. Vosper, E. Carter, H. Lean, A. Lock, P. Clark, and S. Webster. High resolution modelling of valley cold pools. *Atmospheric Science Letters*, 14(3):193–199, 2013.
- S. B. Vosper, S. D. Mobbs, and B. A. Gardiner. Measurements of the near-surface flow over a hill. *Quarterly Journal of the Royal Meteorological Society*, 128(585):2257–2280, 2002.
- V. Vuollo and L. Holmström. A scale space approach for spherical density. In *The Second Advances in Directional Statistics (ADISTA) International Workshop*, Rome, Italy, 2017.
- N. S. Wagenbrenner, J. M. Forthofer, B. K. Lamb, K. S. Shannon, and B. W. Butler. Downscaling surface wind predictions from numerical weather prediction models in complex terrain with WindNinja. *Atmospheric Chemistry and Physics*, 16:5229–5241, 2016.
- G. Wahba. Spline interpolation and smoothing on the sphere. *SIAM Journal on Scientific and Statistical Computing*, 2(1):5–16, 1981.
- G. Wahba. *Spline Models for Observational Data*. CBMS-NSF Regional Conference Series in Applied Mathematics. Society for Industrial and Applied Mathematics, Philadelphia, 1990.
- G. Wahba and J. Wendelberger. Some new mathematical methods for variational objective analysis using splines and cross validation. *Monthly Weather Review*, 108(8):1122–1143, 1980.
- X.-F. Wang and D. Ye. On nonparametric comparison of images and regression surfaces. *Journal of Statistical Planning and Inference*, 140(10):2875–2884, 2010.
- R. O. Weber and P. Kaufmann. Relationship of synoptic winds and complex terrain flows during the MISTRAL field experiment. *Journal of Applied Meteorology*, 37(11):1486–1496, 1998.
- R. O. Weber, R. H. D. McRae, J. J. Sharples, and G. A. Mills. Highfire Risk Project Stakeholder’s Research Report: Evidence-based policy for fire risk management in and around Australia’s high country. Technical report, Bushfire Cooperative Research Centre, 2008.
- J. A. Wellner. Permutation tests for directional data. *The Annals of Statistics*, 7(5):929–943, 1979.

- T. Wells, C. S. Yeo, and R. J. B. Fawcett. Meteorological and fire behaviour lessons learned from the Aberfeldy fire, Victoria, 17th January 2013. In *AFAC 2014*, Wellington, New Zealand, 2014.
- C. D. Whiteman. *Mountain meteorology: fundamentals and applications*. Oxford University Press, New York, 2000.
- C. D. Whiteman and J. C. Doran. The relationship between overlying synoptic-scale flows and winds within a valley. *Journal of Applied Meteorology*, 32(11):1669–1682, 1993.
- Wildland Fire Associates. Granite Mountains IHC Entrapment and Burnover Investigation: Yarnell Hill Fire - June 30, 2013. Technical report, Arizona Division of Occupational Safety and Health, November 2013.
- N. Wood. Wind flow over complex terrain: A historical perspective and the prospect for large-eddy modelling. *Boundary-Layer Meteorology*, 96(1/2):11–32, 2000.
- T. Yang, K. Nadimpali, and B. Cechet. Local wind assessment in Australia. computation methodology for wind multipliers. Technical Report Record 33, Geoscience Australia, 2014.
- S. G. Young and A. W. Bowman. Non-parametric analysis of covariance. *Biometrics*, 51(3):920–931, 1995.
- J. Zhang, S. Chowdhury, A. Messac, and L. Castillo. Multivariate and multimodal wind distribution model based on kernel density estimation. In *ASME 2011 5th International Conference on Energy Sustainability*, Washington DC, USA, 2011.
- J. Zhang, S. Chowdhury, A. Messac, and L. Castillo. A multivariate and multimodal wind distribution model. *Renewable Energy*, 51(0):436–447, 2013.