

Analysis and Forecasting of Extreme Temperature and Precipitation

Across the Complex Terrain of British Columbia

by

Pedro Ivo Odon

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF
THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

in

The Faculty of Graduate Studies

(Atmospheric Sciences)

THE UNIVERSITY OF BRITISH COLUMBIA

(Vancouver)

May 2019

© Pedro Ivo Odon 2019

The following individuals certify that they have read, and recommend to the Faculty of Graduate and Postdoctoral Studies for acceptance, the dissertation entitled:

Analysis and Forecasting of Extreme Temperature and Precipitation

Across the Complex Terrain of British Columbia

submitted by Pedro Ivo Odon in partial fulfillment of the requirements for the degree of Doctor of Philosophy
in Atmospheric Sciences

Examining Committee:

Roland Stull, Atmospheric Sciences
Supervisor

Gregory West, Atmospheric Sciences
Supervisory Committee Member

Ian McKendry, Geography
Supervisory Committee Member

Phillip Austin, Atmospheric Sciences
University Examiner

Brett Eaton, Geography
University Examiner

Abstract

The ability to forecast extreme temperature and precipitation events not only helps satisfy the public's desire to better prepare for such events, but can also provide valuable information about the future risks of such events to emergency managers, regional planners, and policy-makers at all levels of government. Additionally, understanding extreme weather events in the context of climate change can help provide confidence and insight into risk assessment.

This dissertation advances extreme weather forecasting over the complex topography of British Columbia (BC) while accounting for changes in intensity and frequency of extreme events due to nonstationarity.

First the problem of finding a dataset to provide climatological distributions is addressed. Weather station data coverage, quality, and temporal completeness across BC degrade outside of population centres, and as one goes back in time. This data paucity motivates the search for the best reanalysis to serve as a climatological reference dataset. The 2-m temperature and daily accumulated precipitation of the reanalyses are compared with observations from meteorological stations distributed over the complex terrain of British Columbia. The observations are separated into climate regions by Principal Component Analysis (PCA) and K-means clustering, and new verification metrics are introduced to evaluate the best performing reanalysis. Upon thorough evaluation, the Japanese 55-year Reanalysis (JRA-55) was found to be best. Its biases are largely explained by the inability of the coarse-resolution reanalysis to represent terrain characteristics.

The second component of this work combines, downscales and bias corrects the best performing reanalysis using the high-spatial-resolution Parameter-Elevation Regressions on Independent Slopes Model (PRISM) dataset and using surface weather station observations. This results in a high-resolution, long-term gridded dataset that is spatially and temporally complete, yielding a very-high-resolution surface analysis (VHRSA). Biases and mean absolute errors are substantially improved over the JRA-55. The VHRSA not only renders a solution to the paucity of observational data across BC, it also has the potential to be a valuable dataset for research and operational use

in meteorology, climate studies, and hydrology, to name a few.

Next, this dataset is used to create a high-resolution, bias-corrected ensemble forecast using the North American Ensemble Forecast System (NAEFS). The post-processed NAEFS is more skillful than the raw NAEFS forecast out to a forecast lead time of 10 days for both 2-m temperature, and daily accumulated precipitation, according to the verification metrics evaluated.

Statistical temporal stationarity of extreme values of precipitation and temperature are assessed for the 60-year VHRSA period. Statistical nonstationarity is tested to determine if important temporal changes are required to characterize present-day extreme levels. It is determined that nonstationary distributions should be used to represent annual minima values of daily minimum 2-m temperature during summer months and late winter.

Finally, an extreme, or situational awareness index is presented: the Parametric Extreme Index (PEI). It can be used to alert forecasters and other end users of future extreme temperature and precipitation events. The index is shown to be more skillful than the Standardized Anomaly (SA) as the rarity of the event increases, with a higher number of hits and a lower number of misses. The PEI also takes into account the nonstationarity signal of minimum 2-m temperature.

Preface

This dissertation is composed of five chapters which resulted in the publication of three journal articles. These articles provide the contents of Chapters 1, 2, and 3. The details of each article are described in more detail below. Chapter 4 is an article being written as of the time of this writing. The conclusion is unique to this dissertation.

Chapter 1

A portion of Chapter 1 has been published in the Canadian Meteorological and Oceanographic Society (CMOS) Bulletin:

- Odon, P., West, G., and Stull, R. (2017). Vancouver Fall and Winter 2016/17: How Bad Was It? CMOS Bulletin SCMO, 45(4):9-12.

The objective of the paper is to motivate studying the extreme weather that British Columbia (BC) faces throughout fall and winter, and the impacts of such events, in particular in Southwest BC where most of the population resides. P. Odon carried out the research, analyzed the results, and wrote the paper. Professor Stull provided the computational and financial resources for the study. Professor Stull and Dr. West provided background knowledge on meteorology and insights into the impacts on BC Hydro. They also helped editing the paper.

Chapter 2

Chapter 2 has been published in Journal of Applied Meteorology and Climatology:

- Odon, P., West, G., and Stull, R. (2018). Evaluation of reanalyses over British Columbia. Part I: Daily and Extreme 2-m temperature. Journal of Applied Meteorology and Climatology, 57(9):2091-2112.

The paper evaluates how different the reanalysis datasets represent daily and extreme 2-m temperature across the complex terrain of BC. P. Odon carried out the research, analyzed the results, and wrote the manuscript. Professor Stull provided the computational and financial resources for the study and together with Dr. West, edited the contents of the manuscript. Additionally, Dr West provided input into the direction of the research.

Chapter 3

This is the second part of a two-part study. Chapter 2 is *Part I*. Chapter 3 has been published in *Journal of Applied Meteorology and Climatology*:

- Odon, P., West, G., and Stull, R. (2019). Performance of Reanalyses across British Columbia. Part II: Evaluation of Daily and Extreme Precipitation, (in press).

The objective of this paper is to evaluate how the different reanalysis datasets represent daily and extreme precipitation across BC. P. Odon carried out the research, analyzed the results, and wrote the manuscript. Professor Stull provided the computational and financial resources for the study and together with Dr West, edited the contents of the manuscript.

Table of Contents

Abstract	iii
Preface	v
Table of Contents	vii
List of Tables	x
List of Figures	xii
Acknowledgements	xxii
Dedication	xxiii
1 Introduction	1
1.1 Understanding Extreme Weather Events and its Impacts — A Case Study	2
1.1.1 Fall	2
1.1.2 Winter	9
1.1.3 Anatomy of an Arctic Outbreak	14
1.1.4 Final Considerations	14
1.2 Extreme Data Analysis	16
1.3 A Historical Record	17
1.4 A Gridded Climatological Dataset	18
1.5 Forecasting Extreme Weather Events	19
1.6 Dissertation Organization	19
2 Performance of Reanalyses across British Columbia. Part I: Evaluation of Daily and Extreme 2-m Temperature	21
2.1 Introduction	21
2.2 Data and Methodology	22
2.2.1 Weather station data	25
2.2.2 CFSR	27

Table of Contents

2.2.3	ERA-Interim	27
2.2.4	JRA-55	28
2.2.5	MERRA-2	28
2.3	Climate zones	29
2.4	Verification Metrics	32
2.4.1	Daily T2M	32
2.4.2	Extreme T2M	33
2.4.3	Analysis of Variance (ANOVA)	35
2.5	Nonstationarity	35
2.5.1	Daily T2M	36
2.5.2	Extreme T2M	37
2.6	Results and discussion	37
2.6.1	Daily T2M	37
2.6.2	Extreme T2M	48
2.7	Stationarity	57
2.8	Conclusion	62
3	Performance of Reanalyses across British Columbia. Part	
	II: Evaluation of Daily and Extreme Precipitation	64
3.1	Introduction	64
3.2	Data and Methodology	65
3.2.1	Weather station data	68
3.2.2	CFSR	71
3.2.3	ERA-Interim	71
3.2.4	JRA-55	71
3.2.5	MERRA-2	72
3.2.6	PRISM dataset	72
3.3	Climate zones	73
3.4	Verification Metrics	75
3.4.1	Daily PCP	75
3.4.2	Extreme PCP	76
3.4.3	Kruskal-Wallis Analysis	79
3.5	Results and discussion	80
3.5.1	Daily PCP	80
3.5.2	Extreme PCP	93
3.6	Stationarity	99
3.6.1	Daily PCP	99
3.6.2	Extreme PCP	105
3.7	Conclusion	105

Table of Contents

4	Analysis and Forecast of High-resolution Extreme Weather	109
4.1	Introduction	109
4.2	Data	110
4.2.1	Weather station data	110
4.2.2	PRISM	113
4.2.3	NAEFS	113
4.3	The VHRSA: Downscaling and bias-correcting the JRA-55	113
4.3.1	Verification of the VHRSA Daily and Extreme T2M	117
4.3.2	Verification of the VHRSA Daily and Extreme PCP	119
4.4	Bias-corrected NAEFS	123
4.4.1	Verification of the Forecast	129
4.4.2	Verification of Daily PCP	130
4.4.3	Verification of Daily Maximum and Minimum T2M	133
4.5	Stationarity	134
4.5.1	Extreme T2M	137
4.5.2	Extreme PCP	139
4.6	The Parametric Extreme Index	139
4.6.1	Verification of the PEI	141
4.6.2	PEI Performance across T2M and PCP	147
4.7	Conclusion	148
5	Conclusion	152
	Bibliography	156
	Appendices	
A	Surface weather stations	180
B	Train and test stations	184

List of Tables

2.1	Overview of the four reanalysis datasets examined in this study.	24
2.2	Seasonally averaged systematic error (SE in °C) and random error (RE in °C) of mean values, and Kolmogorov-Smirnov statistic (KS) of daily maximum T2M by climate zone	39
2.3	Seasonally averaged systematic error (SE in °C) and random error (RE in °C) of mean values, and Kolmogorov-Smirnov statistic (KS) of daily minimum T2M by climate zone	42
2.4	Seasonally averaged systematic error of 2- and 30-year return levels of daily maximum T2M by climate zone (SE2 and SE30 respectively in °C)	49
2.5	Seasonally averaged systematic error of 2- and 30-year return levels of daily minimum T2M by climate zone (SE2 and SE30 respectively in °C)	52
3.1	Overview of the four reanalysis datasets examined in this study.	67
3.2	Averaged systematic error of: monthly precipitation total (M), two-sample χ^2 -statistic (χ^2), 31-day-window percentage of wet days (W), and 30-year return levels of 1- (1D30), 3- (3D30), 7- (7D30), and 14-day (14D30) accumulated precipitation across wetter climate zones Northwest, Maritime West, Maritime East and Southwest, drier climate zones North, South Central, Central and Southeast, and all climate zones in BC (all systematic errors but two-sample χ^2 -statistic in %)	81
4.1	Averaged systematic error and MAE of monthly precipitation total and monthly mean daily maximum and minimum T2M (MSE and MMAE respectively), of 2-year return levels (SE2 and MAE2 respectively), averaged systematic error of two-sample χ^2 -statistic (χ^2) and KS statistics, across all test stations (all T2M errors but KS statistic in °C; all PCP errors but two-sample χ^2 -statistic in %)	122
4.2	Contingency table.	129

List of Tables

A.1	Surface weather stations description. Station name and abbreviation, longitude and latitude in degrees, elevation in meters, variable as either 2-m temperature (T) or 1-day accumulated precipitation (p), network and whether it is a SYNOP station.	180
B.1	Train surface weather stations description. Station name and abbreviation, longitude and latitude in degrees, elevation in meters, variable as either 2-m temperature (T) or 1-day accumulated precipitation (p), network station.	184
B.2	Test surface weather stations description. Station name and abbreviation, longitude and latitude in degrees, elevation in meters, variable as either 2-m temperature (T) or 1-day accumulated precipitation (p), network station.	186

List of Figures

1.1	A fine-resolution computer model wind forecast for December 12th, 2017 at 1300 UTC. Vector orientation shows direction, colour shows speed. Terrain is grey-shaded. Geographic locations mentioned in text shown with red labels, highways labelled in red.	4
1.2	Fifteen-day accumulated precipitation anomaly centered on November 8th, 2017, relative to 1979-2017 period. Data from the the ECMWF interim reanalysis (ERA-Interim) (Dee et al., 2011).	6
1.3	Cumulative number of rainy days at Vancouver International Airport. Time series for all years between 1937/38 and 2016/17 in grey, 1998/99 in red and 2016/17 in black.	7
1.4	Consecutive rainy days at Vancouver International Airport for the past 5 years in grey, with fall/winter 2016/17 in black.	8
1.5	Fifteen-day averaged minimum 2-m temperature anomaly centered on December 11th, 2016, relative to 1979-2017 period.	10
1.6	Daily minimum 2-m temperature time series for Vancouver, smoothed using a 5-day rolling window for readability. Time series for all years in grey, 1937/38-2016/17 average in blue, 2012/13-2016/17 average in red, 1992/92 in orange and 2016/17 in black. Days with snowfall in 2016/17 indicated with green dashed lines.	11
1.7	(a) Public transit buses paralyzed by road conditions during the December 11th, 2016 arctic outbreak (Fritz, 2016). (b) A Google Maps screen capture, where orange and red colours indicate widespread very slow or stopped traffic during the February 3rd, 2017 arctic outbreak.	13

List of Figures

1.8	(a) 50.0-kPa geopotential height (km) and 2-m temperature (°C), (b) sea level pressure (kPa) and 6-hr precipitation (mm) for December 4th, 2016 at 1200 UTC. (c) and (d), as in (a) and (b), but for December 6th, 2016 at 1200 UTC. Data from the the ECMWF interim reanalysis (ERA-Interim) (Dee et al., 2011).	15
2.1	Location of ECCC (red) and BC Hydro (blue) weather stations and British Columbia population distribution (orange). Upside-down triangles indicate valley stations; upright triangles indicate non-valley stations. Dashed lines delineate the dominant climate zones: North, Central, Maritime West, Southwest and Southeast.	26
2.2	Correlation matrix and five dominant climate zone clusters in BC.	30
2.3	Observed and reanalysis mean, systematic error, and random error magnitude for daily maximum T2M, averaged over stations in the Central climate zone. The vertical dashed lines indicate the change in seasons and the colored dots represent the seasonal average.	40
2.4	Observed and reanalysis mean, systematic error, and random error magnitude for daily minimum T2M, averaged over stations in the North climate zone. The vertical dashed lines indicate the change in seasons and the colored dots represent the seasonal average.	43
2.5	a) Kolmogorov-Smirnov (KS) statistic results for daily maximum T2M, for all four reanalyses by time of year, averaged across all locations in the Southeast climate zone. The vertical dashed lines indicate the change in seasons and the colored dots represent the seasonal average. KS statistic values closer to zero are better. b) As an example, KS-statistic results for daily maximum T2M on March 1 at Cranbrook (YXC) are shown.	45
2.6	Terrain elevation error in meters for ERA-Interim. Other reanalyses' errors are similar.	46

List of Figures

2.7	Mean systematic and mean random error of daily maximum (TMAX) and minimum (TMIN) T2M for each of the 57 stations as a function of terrain reanalysis error. The solid lines show the linear regression fits. The nearly horizontal lines show the random error; the two lines are nearly on top of each other.	47
2.8	30-year return level and systematic error for maximum T2M, for all four reanalyses for Southeast climate zone. The vertical dashed lines indicate the change in seasons and the colored dots represent the seasonal average.	50
2.9	30-year return level and systematic error for minimum T2M, for all four reanalyses for Central climate zone. The vertical dashed lines indicate the change in seasons and the colored dots represent the seasonal average.	53
2.10	Mean systematic error of 2- and 30-year return levels (RL) of extreme maximum (TMAX) and minimum (TMIN) T2M for each of the 57 stations as a function of terrain reanalysis error. The solid lines show the linear regression fits.	55
2.11	MAE ($^{\circ}\text{C}$) of daily maximum and minimum T2M, and of extreme (2- and 30-year return levels) maximum and minimum T2M. MAE is averaged across all 57 stations. Smaller values (closer to the origin) are better.	56
2.12	(a) Mean linear trend of daily minimum T2M; (b) Standard deviation linear trend of daily minimum T2M; (c) Days when $N(\mu(t), \sigma)$ is statistically significant. The vertical dashed lines indicate the change in seasons and the horizontal dashed lines delineate from top to bottom the North, Central, Maritime, Southwest and Southeast climate zones respectively. . .	58
2.13	(a) Mean linear trend of daily maximum T2M; (b) Standard deviation linear trend of daily maximum T2M; (c) Days when $N(\mu(t), \sigma)$ is statistically significant. The vertical dashed lines indicate the change in seasons and the horizontal dashed lines delineate from top to bottom the North, Central, Maritime, Southwest and Southeast climate zones respectively. . .	59

List of Figures

2.14	(a) Location linear trend of extreme maximum T2M; (b) Days when $GEV(\mu(t), \sigma, \kappa)$ for extreme maximum T2M is statistically significant; (c) Location linear trend of extreme minimum T2M; (d) Days when $GEV(\mu(t), \sigma, \kappa)$ for extreme minimum T2M is statistically significant. The vertical dashed lines indicate the change in seasons and the horizontal dashed lines delineate from top to bottom the North, Central, Maritime, Southwest and Southeast climate zones respectively.	61
3.1	Location of ECCC (red) and BC Hydro (blue) weather stations used for precipitation analysis and British Columbia population distribution (orange). Upside-down triangles indicate valley stations; upright triangles indicate non-valley stations. Dashed lines delineate the dominant climate zones North, Northwest, Central, South Central, Maritime West, Maritime East, Southwest and Southeast.	69
3.2	Cramer correlation matrix and eight dominant climate zone clusters in BC. Crosses indicate stations where differences in precipitation distribution are statistically significant.	74
3.3	(a) Diagram illustrating the 31-day centered rolling window for July 9th, performed over the 31-year period; (b) GEV model fitted over the 31-day, 31-year centered window for 1-day precipitation total at Vancouver International Airport (YVR); (c) Quantile plot for the GEV fitted model for the same day, location, and variable; If the GEV is a reasonable model, the points on the quantile plot should lie close to the unit diagonal; (d) Return level, return period plot for the same day, location, and variable; showing the precipitation that corresponds to a given return period.	78
3.4	Observed and reanalysis running centered 31-day precipitation totals and systematic error averaged over stations in the Northwest climate zone. The vertical dashed lines indicate the change in seasons and the colored dots represent the seasonal average.	82
3.5	Observed and reanalysis running centered 31-day precipitation totals and systematic error averaged over stations in the Central climate zone. The vertical dashed lines indicate the change in seasons and the colored dots represent the seasonal average.	84

List of Figures

3.6	Percentage of (a) wet days; (b) two-sample χ^2 -statistic; percentage of (c) dry days, (d) very light, (e) light (f) moderate, (g) heavy and (h) extreme precipitation events for Maritime West climate zone. The vertical dashed lines indicate the change in seasons and the colored dots represent the seasonal average.	86
3.7	Percentage of (a) "wet days"; (b) Two-sample χ^2 -statistic; Percentage of (c) "dry days", (d) "very light", (e) "light" (f) "moderate", (g) "heavy" and (h) "extreme" precipitation events for North climate zone. The vertical dashed lines indicate the change in seasons and the colored dots represent the seasonal average.	88
3.8	a) Winter precipitation totals of (a) PRISM. Winter precipitation totals bilinearly interpolated to PRISM grid of (b) CFSR, (c) ERA-Interim, (d) JRA-55 and (e) MERRA-2. Systematic error of (f) CFSR, (g) ERA-Interim, (h) JRA-55 and (i) MERRA-2. The dots represent weather station location.	90
3.9	JRA-55 mean systematic error of 31-day precipitation totals for each of the 66 stations as a function of PRISM mean systematic error. The solid lines show the linear regression fits.	92
3.10	30-year return level and systematic error of 1-day precipitation total, for all four reanalyses for Southwest climate zone. The vertical dashed lines indicate the change in seasons and the colored dots represent the seasonal average.	94
3.11	30-year return level and systematic error of 1-day precipitation total, for all four reanalyses for North climate zone. The vertical dashed lines indicate the change in seasons and the colored dots represent the seasonal average.	96
3.12	MAE of 31-day precipitation totals, χ^2 -statistic, "wet days", and 2-year and 30-year return levels of 1-, 3-, 7-, and 14-day precipitation totals. MAE is averaged across all 66 stations. Values closer to 0 at the origin of the plot are better.	98

List of Figures

3.13	Seasonal precipitation total at Vancouver International Airport (YVR) for July 9th (grey), 10-year left moving average (black), decadal averages of seasonal precipitation total (1981-1990, 1991-2000 and 2001-2010) (blue) and 30-year seasonal precipitation total averaged over the study period 1981-2010 (green). The red line shows the linear regression fit to the three decadal mean values. Individual 90% confidence intervals (CI; green dotted line) and multiple $(1 - \alpha_{Walker}) \times 100\%$ CI (red dotted line) are drawn; $\alpha_0 = 0.10$. This example indicates that a stationary distribution is appropriate for Jul 9th because the three decadal average values all fall within both CI.	101
3.14	(a) Mean linear trend of season accumulated precipitation; (b) Mean linear trend of days with precipitation. The vertical dashed lines indicate the change in seasons and the horizontal dashed lines delineate from top to bottom the North, Central, Northwest, Maritime West, Maritime East, Southwest, South Central and Southeast climate zones respectively. Units are percent change over the 30-year period.	103
3.15	Mean linear trend of (a) “very light” , (b) “light”, (c) “moderate” and (d) “extreme” precipitation events. The vertical dashed lines indicate the change in seasons and the horizontal dashed lines delineate from top to bottom the North, Central, Northwest, Maritime West, Maritime East, Southwest, South Central and Southeast climate zones respectively. Units are percent change over the 30-year period.	104
4.1	Location of ECCC (blue) and BC Hydro (yellow) T2M weather stations and British Columbia population distribution (red). .	111
4.2	Location of ECCC (blue) and BC Hydro (yellow) PCP weather stations and British Columbia population distribution (red). .	112

List of Figures

4.3	(a) 30-year monthly mean of daily maximum T2M PRISM climate on August 28th. (b) 2-year return level systematic error between training stations and reanalysis interpolated to training stations as a function of monthly mean daily maximum T2M bias between JRA-55 and PRISM interpolated to training stations. The solid line shows the linear regression fit. (c) Downscaling difference derived by subtracting (a) from (e). (d) JRA-55 daily maximum T2M on 28 Aug 2017. (e) Monthly mean of daily maximum T2M JRA-55 climate bilinearly interpolated to PRISM grid. (f) VHRSA daily maximum T2M on 28 Aug 2017.	116
4.4	(a) Observed, JRA-55 and VHRSA monthly mean daily maximum T2M averaged over test stations. (b) 2-year return levels of observed, JRA-55 and VHRSA maximum T2M averaged over test stations. (c) Monthly mean bias of JRA-55 and VHRSA daily maximum T2M averaged over test stations. (d) As in (c) but for 2-year return levels. (e) As in (c) but for MAE. (f) As in (e) but for 2-year return levels. (g) KS statistic for daily maximum T2M. The vertical dashed lines indicate the change in seasons and the coloured dots represent the errors seasonal averages. In (a) and (b), values closer to observations (black line) are better. In (c)-(g), values closer to zero are better.	118
4.5	(a) Observed, JRA-55 and VHRSA monthly mean daily minimum T2M averaged over test stations. (b) 2-year return levels of observed, JRA-55 and VHRSA minimum T2M averaged over test stations. (c) Monthly mean bias of JRA-55 and VHRSA daily minimum T2M averaged over test stations. (d) Same as in (c) but for 2-year return levels. (e) Same as in (c) but for MAE. (f) Same as in (e) but for 2-year return levels. (g) KS statistic for daily minimum T2M. The vertical dashed lines indicate the change in seasons and the coloured dots represent the errors seasonal averages. Values of bias, MAE, and KS Statistic closer to zero are better.	120

List of Figures

4.6	(a) Observed, JRA-55 and VHRSA mean of monthly PCP totals averaged over test stations. b) 2-year return levels of observed, JRA-55 and VHRSA PCP averaged over test stations. (c) Monthly mean bias of bias-corrected JRA-55 PCP averaged over test stations. (d) Same as in (c) but for 2-year return levels. (e) Same as in (c) but for MAE. (f) Same as in (e) but for 2-year return levels. g) χ^2 -statistic of PCP. The vertical dashed lines indicate the change in seasons and the coloured dots represent the errors seasonal averages. Values of bias, MAE, and χ^2 -statistic closer to zero are better. . . .	121
4.7	(a) NAEFS raw Canadian control member forecast of daily minimum T2M on previous day 2 Jan, 2017; (b) Same as in (a) but for VHRSA; (c) Ensemble mean raw NAEFS forecast on current day 3 Jan, 2017; (d) Ensemble mean bias-corrected NAEFS forecast on current day 3 Jan, 2017. DMB method done separately for each NAEFS member and then averaged to generate the bias-corrected ensemble mean. (e) VHRSA daily minimum T2M on 3 Jan, 2017 presented for comparison with (c) and (d). The dots represent weather station location.	126
4.8	(a) NAEFS raw Canadian control member forecast of 1-day accumulated PCP on previous day 7 Nov, 2016; (b) Same as in (a) but for VHRSA; (c) Ensemble mean raw NAEFS forecast on current day 8 Nov, 2016; (d) Ensemble mean bias-corrected NAEFS forecast on current day 8 Nov, 2016. DMB method done separately for each NAEFS member and then averaged to generate the bias-corrected ensemble mean. (e) VHRSA 1-day accumulated PCP on 8 Nov, 2016 presented for comparison with (c) and (d). The dots represent weather station location.	127
4.9	(a) NAEFS raw Canadian control member forecast of daily maximum T2M on previous day 27 Aug, 2017; (b) Same as in (a) but for VHRSA; (c) Ensemble mean raw NAEFS forecast on current day 28 Aug, 2017; (d) Ensemble mean bias-corrected NAEFS forecast on current day 28 Aug, 2017. DMB method done separately for each NAEFS member and then averaged to generate the bias-corrected ensemble mean. (e) VHRSA daily maximum T2M on 28 Aug, 2017 presented for comparison with (c) and (d). The dots represent weather station location.	128

List of Figures

4.10 a) TS for events above the 95th percentile during Fall 2016; (b) Same as in a) but for H; (c) Same as in a) but for F ; d) Systematic error of daily PCP; (e) Same as in a) but for b ; (f) Same as in a) but for d ; (g) Same as in a) but for QBS; (h) Same as in a) but for CRPS. Values of TS and H closer to one are better. Values of F and systematic error closer to zero are better. Combined values of $b = 0$ and $d = 1$ are better. Values of QBS and CRPS closer to zero are better.	132
4.11 a) TS for events above the 90th percentile during Summer 2017; (b) Same as in a) but for H; (c) Same as in a) but for F ; d) Systematic error of daily maximum T2M. Values of TS and H closer to one are better. Values of F and systematic error closer to zero are better.	133
4.12 a) Systematic error of daily minimum T2M during Winter 2016/17; (b) Same as in a) but for QBS; (c) Same as in a) but for b ; (d) Same as in a) but for d ; (e) Same as in a) but for CRPS. Values of systematic error, QBS and CRPS closer to zero are better. Combined values of $b = 0$ and $d = 1$ are better.	135
4.13 (a) Trend of extreme minimum T2M in July; (b) Locations where trend is statistically significant; (c) Locations where $GEV(\mu(t), \sigma, \kappa)$ for extreme minimum T2M is statistically significant.	138
4.14 (a) Trend of extreme maximum T2M in August; (b) Locations where trend is statistically significant.	140
4.15 (a) The NCDF (purple) and the GCDF (black) show the probability (y-axis) that the extreme value ($F_c^{-1}(p_c)$; x-axis) will not be exceeded. (b) The NAEFS-derived probability that $F_c^{-1}(p_c)$ will not be exceeded is given by $F_f(F_c^{-1}(p_c))$. The complement of the NCDF, given by $1 - F_f(F_c^{-1}(p_c))$, computes the NAEFS-derived probability that $F_c^{-1}(p_c)$ will be exceeded. The PEI is related to the area above the NCDF bounded by $p_c = 0$ and $p_c = F_c(F_f^{-1}(1))$ ($p_c = 0.12$ in the figure). PEI=0.02 indicating low extreme values have a low probability of exceedance.	142
4.16 Day 1 PEI over BC on 3 Jan 2017.	143
4.17 Day 1 PEI over BC on 8 Nov 2016.	144
4.18 Day 1 PEI over BC on 28 Aug 2017.	145

List of Figures

4.19 EDS as function of 1 - BR for Day 1 PEI and SA across all stations during summer of 2017. Rarer events are to the right. Values of EDS closer to 1 are better.	147
A.1 Elevation of 2-m temperature and 1-day accumulated precipitation stations.	183

Acknowledgements

I thank my research supervisor, Dr. Roland Stull, who provided means and funding for my dissertation, helped writing and editing journal papers, and provided me with a great environment for my studies. I also wish to thank the other members of my supervisory committee, Drs. Alex Cannon, Greg West and Ian McKendry.

I am grateful for the positive and supportive environment in the office, and the help of many of the members of the Weather Forecast Research Team. Without it I'm not sure my transition into Atmospheric Sciences would have been possible. I must specifically thank Maggie Campbell and David Siuta for their patience, always allowing me to stop by their desks and ask questions about Atmospheric Sciences or coding; Dr. Greg West for help writing and editing journal papers, and many conversations which helped spark many of the ideas presented in this dissertation.

Dr. Henryk Modzelewski, Roland Schigas and Tim Chui provided excellent assistance with IT questions, computing resources and coding.

BC Hydro, the Canadian Natural Science and Engineering Research Council (NSERC), and Mitacs are acknowledged for providing funds for this research.

Dedication

To my wife Andreia. This degree would not be possible without your unconditional love and support. This degree is yours as much as it is mine.

Chapter 1

Introduction

The goal of this research is to provide accurate forecasts of extreme daily maximum and minimum 2-m temperature (hereafter T2M) and 1-day accumulated precipitation (hereafter PCP) over the mountainous terrain of British Columbia (BC), Canada. Forecasting extreme events as well as future trends requires accurately identifying such patterns in space and time. During the October-March storm season, southwest BC is affected by extreme rainfall events in low-elevations, typically falling as snow at mid to high elevations (Odon et al., 2017). One type of storm process, known as an atmospheric river, makes landfall on the Coast Mountains of BC, producing heavy to extreme precipitation due to orographic lifting (Ralph et al., 2004, 2005). During winter, arctic air outbreaks can result in record-breaking low temperatures. These and other types of extreme cool season weather events cause floods, property and infrastructure damage, power outages, and business and public-transportation disruption (Odon et al., 2017). During summer, long droughts (Odon et al.) in combination with record-breaking summer temperatures across BC result in high electricity demand, livestock suffering, wildfire fighting costs that exceeded half a billion dollars in 2017 alone, and insured property losses that reach hundreds of millions of dollars (Phillips, 2018).

Complicating matters further, global climate models project an increase in the frequency and intensity of extreme temperature and precipitation events (Kharin and Zwiers, 2000; Zwiers and Kharin, 2005; Zwiers and Wehner, 2013). This will make the impacts of such events, such as increases in winter runoff leading to flooding, and overwhelmed drainage and sewage-system capacity (White et al., 2016; Sun et al., 2018b), more difficult to manage and mitigate.

This necessitates increasing our understanding and ability to forecast extreme weather events. Such information can provide earlier and more accurate detection of such events, which in turn can help community responders, government stakeholders and even the media to take appropriate action to mitigate damage and reduce casualties (Herring et al., 2017).

Although numerical weather prediction has greatly improved over the

past decades, forecasting extreme T2M and PCP events, especially extreme PCP beyond two days, remains challenging (Lalaurette, 2003).

This dissertation has four goals: (1) Employ an appropriate statistical model that specifically addresses extreme T2M and PCP events. Such a model provides accurate estimation of future extreme levels only if sufficient historical data is available. Extreme data modelling, however, leads to a waste of information when implemented in practice, since only the maximum observed values over a specified period are used. This creates a scarcity of data. Even without the scarcity problem, the simple availability of accurate and complete historical data can be an impediment. This leads to the next two goals: (2) evaluate which reanalysis best represents observations to serve as a surrogate climatological dataset with respect to daily and extreme minimum and maximum T2M, and daily and extreme PCP over BC; and (3) based on that evaluation, improve the best reanalysis if possible to create a more accurate gridded climatological dataset to better understand T2M and PCP extreme levels. Finally, (4) combine this gridded dataset with gridded forecasts to create bias-corrected forecasts that provide more accurate information about extreme forecast events.

Motivation as to why this research is important and its potential benefits are provided in the next section, via a case-study overview of the extreme cool season of 2016/2017 in southwest BC. Each of the four goals are introduced in the four following sections, with details presented in subsequent chapters.

1.1 Understanding Extreme Weather Events and its Impacts — A Case Study

The fall and winter seasons of 2016/17 were noteworthy for the cold, wet weather they brought to regions of British Columbia (BC). Although all of BC experienced some impacts, the South Coast region saw particularly large impacts from an exceptionally wet fall that quickly transitioned into a persistently cold, snowy winter.

1.1.1 Fall

October saw significant snow accumulations at mid to high elevations, and record rainfall amounts in many low-elevation areas across the South Coast. Temperatures warmed to record levels ahead of an atmospheric river event on November 8th, 2016. Many daily records were set, with new monthly

1.1. Understanding Extreme Weather Events and its Impacts — A Case Study

records of 19.4°C in Vancouver and 22.4°C in Abbotsford (see Fig. 1.1 for all mentioned locations). The combination of record warmth and rain-on-snow led to large snowmelt contributions to runoff, in addition to the heavy rainfall itself.

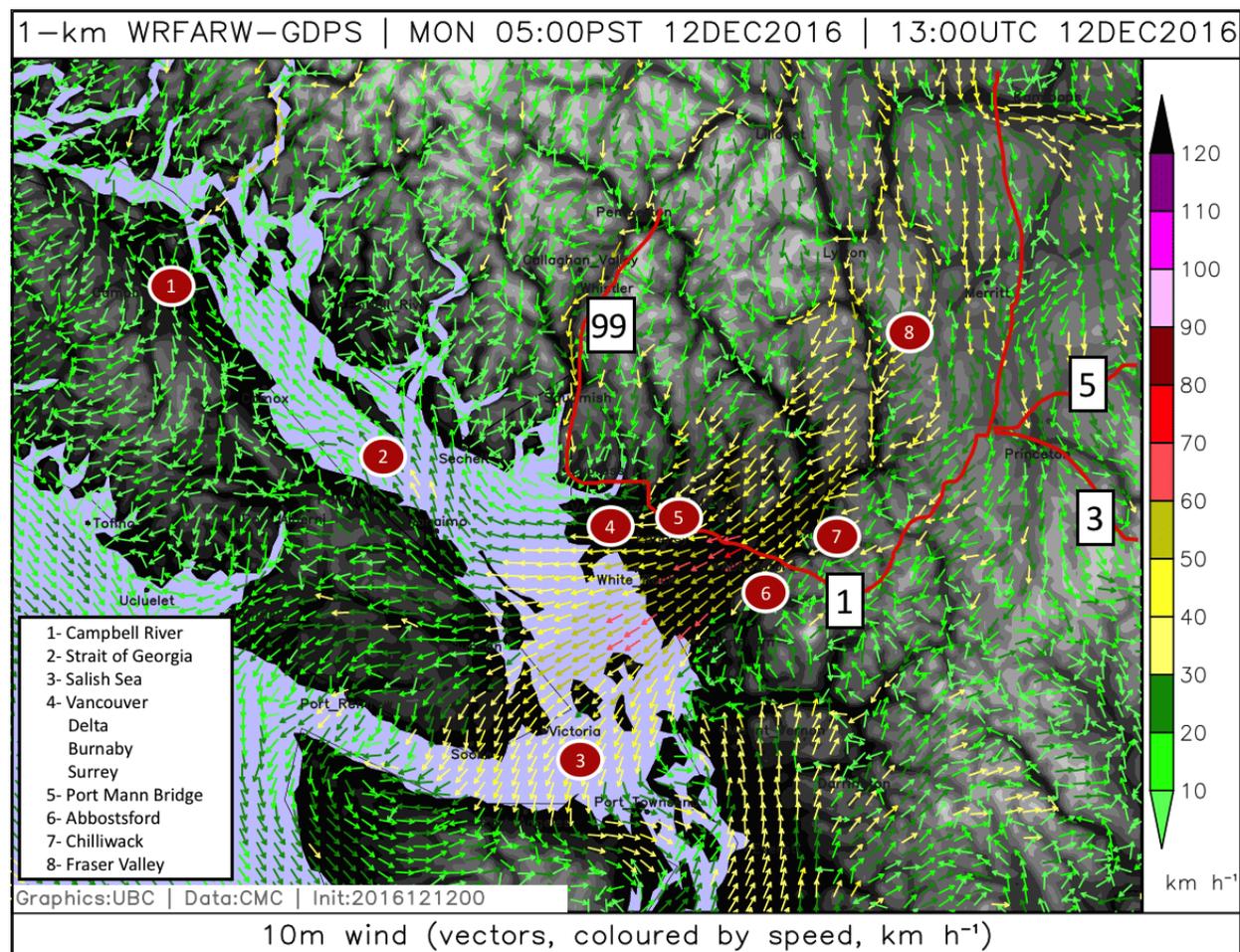


Figure 1.1: A fine-resolution computer model wind forecast for December 12th, 2017 at 1300 UTC. Vector orientation shows direction, colour shows speed. Terrain is grey-shaded. Geographic locations mentioned in text shown with red labels, highways labelled in red.

1.1. *Understanding Extreme Weather Events and its Impacts — A Case Study*

Figure 1.2 illustrates the fifteen-day precipitation anomalies for the first half of November. In several locations of the Coast it precipitated twice more than normal.

For the province’s primary electric utility, BC Hydro, the situation made for challenging reservoir operations, balancing dam safety requirements with minimizing downstream flooding. Although most reservoirs across the South Coast were full or near full, the most concerning was the Campbell River system. The upper part of the watershed received the normal November monthly precipitation amount compressed into the first eight days of the month (BC Hydro, 2016). Furthermore, one-week and two-week reservoir inflows set new records, and were estimated to exceed 1-in-100 year volumes.

In emergency coordination calls and meetings, BC Hydro worked with nearby towns, regional districts, and the province, deciding to increase discharges to a record-tying $600 \text{ m}^3 \cdot \text{s}^{-1}$, enough to fill an olympic-sized swimming pool every four seconds. This was done to mitigate the risk of overtopping the dam in subsequent storms, which would’ve meant passing the full reservoir inflows ($1,100 \text{ m}^3 \cdot \text{s}^{-1}$), flooding communities downstream (BC Hydro, 2016).

At Vancouver International Airport (hereafter, Vancouver) there were only three days without rain in October and only two in November. In fact, the fall and winter 2016/17 period had the second highest frequency of rain since 1937/38, with 121 days of rain during the 182-day (6-month) period (Fig. 1.3). Only the 1998/99 period was higher, with 131 days of rain. Furthermore, this 2016/17 period featured 22 consecutive days of rain (October 12th - November 2nd; Fig. 1.4). This was the longest stretch in the past six years, and the ninth longest since 1937/38.

1.1. *Understanding Extreme Weather Events and its Impacts — A Case Study*

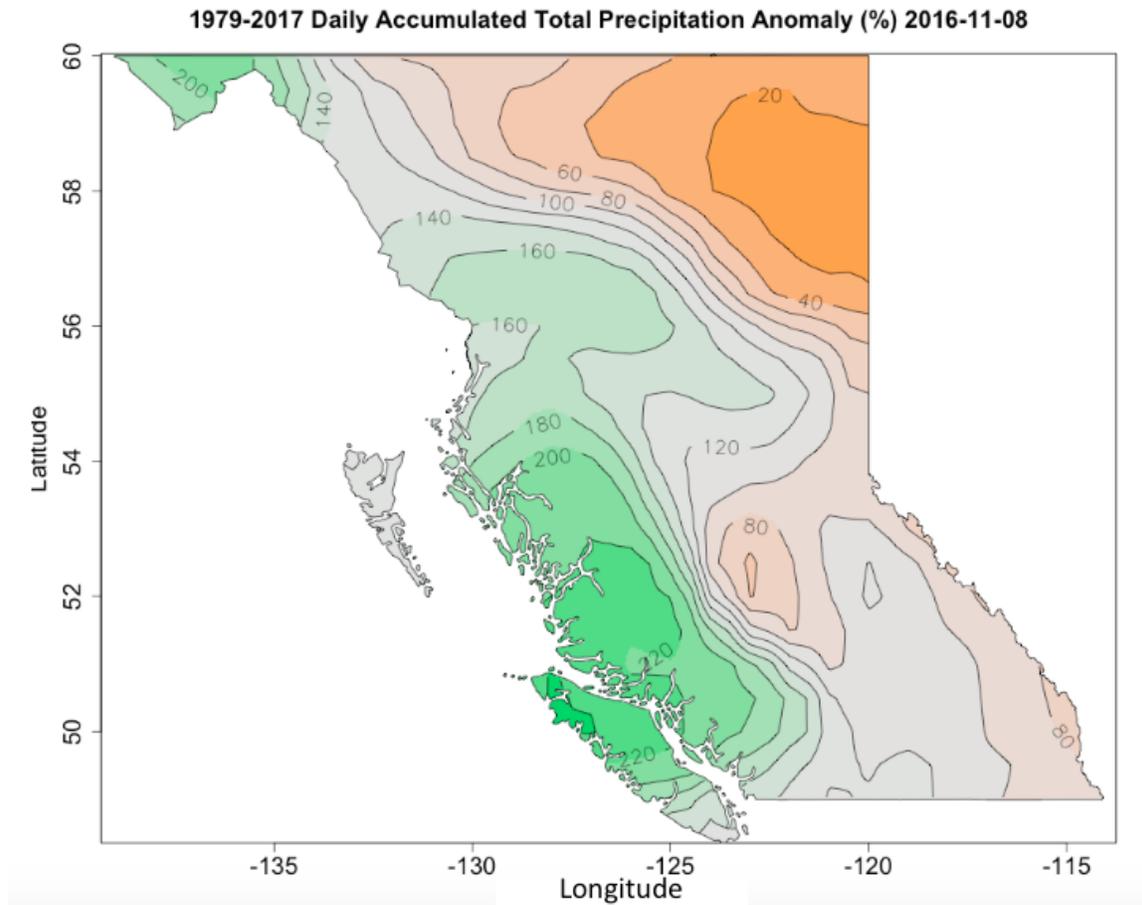


Figure 1.2: Fifteen-day accumulated precipitation anomaly centered on November 8th, 2017, relative to 1979-2017 period. Data from the the ECMWF interim reanalysis (ERA-Interim) (Dee et al., 2011).

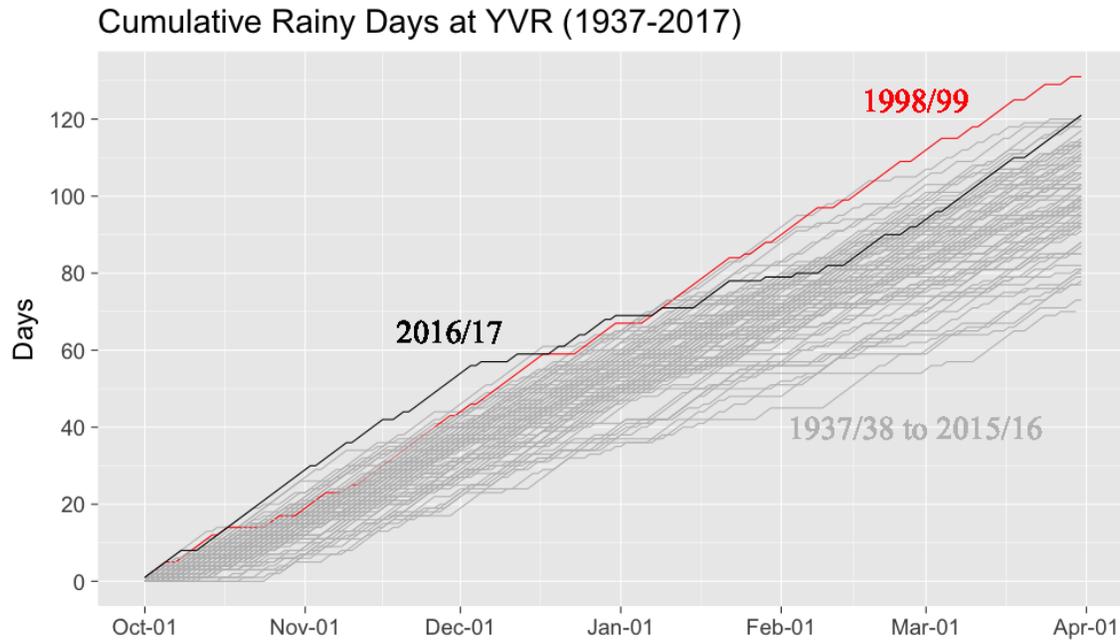


Figure 1.3: Cumulative number of rainy days at Vancouver International Airport. Time series for all years between 1937/38 and 2016/17 in grey, 1998/99 in red and 2016/17 in black.

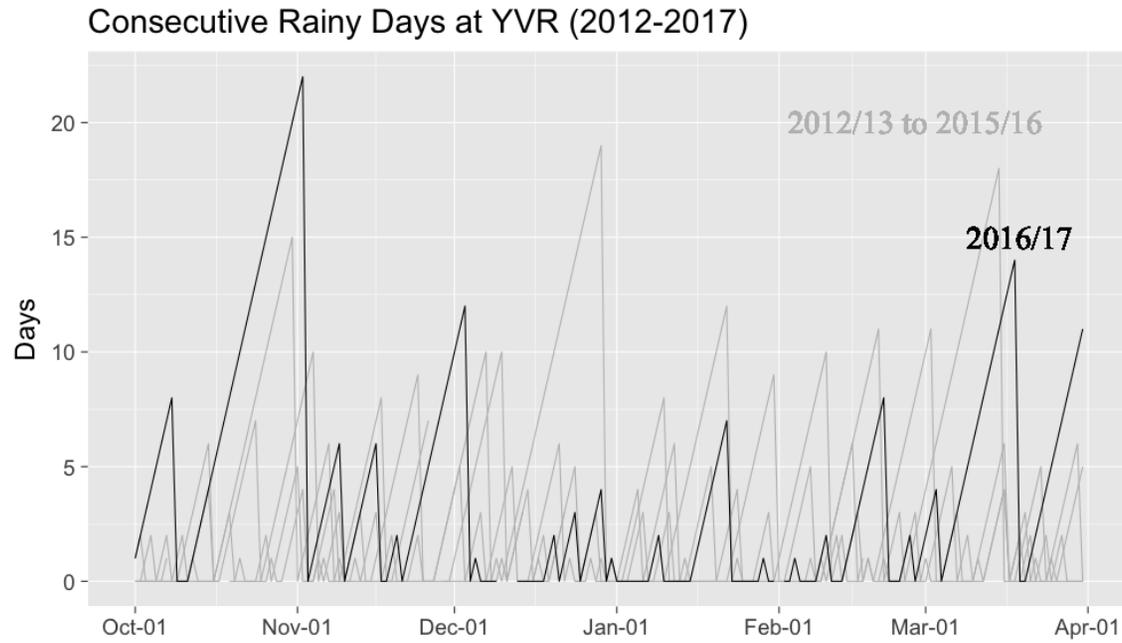


Figure 1.4: Consecutive rainy days at Vancouver International Airport for the past 5 years in grey, with fall/winter 2016/17 in black.

1.1.2 Winter

In December the pattern abruptly became much cold and drier. Five arctic air outbreaks brought anomalous cold temperatures, and snow to British Columbia in December through February.

Figure 1.5 shows the 15-day average minimum 2-m temperature anomaly across British Columbia centered on second arctic outbreak on December 11th, 2016. It shows the difference between what British Columbians would expect for that time of year and what actually happened. Temperatures were much colder than average, with much of BC more than 10°C below normal.

In Vancouver, daily minimum temperatures were below freezing for most of December, January and the first part of February. This period was colder than winters in recent memory (5-year average, red line in Fig. 1.6) and the long term average (80-year average, blue line in Fig. 1.6). There were 54 days of below freezing daily minimum temperature during the entire 6-month period, more than any such period since 1992/93 (orange line in Fig. 1.6).

The five arctic outbreaks can be identified in the smoothed minimum temperature time series by the large departures from the 80-year average, occurring on approximately December 5th and 11th, 2016, January 1st and 10th, 2017 and February 2nd, 2017; most followed by snow days (green dashed line in Fig. 1.6).

These weather events directly impacted the general population as well as government and industry. Garbage was left uncollected for weeks in neighbourhoods of Vancouver, Burnaby, Surrey and Delta (Correia, 2016; McElroy, 2016) due to persistent snow and ice cover in alleys and lanes.

Across Metro Vancouver, many side streets and sidewalks were left un-cleared and unsalted. Residents had difficulties simply getting around the city and questioned the Vancouver Mayor's commitment to the issues (Laanela, 2016). Nearby cities also experienced salt shortages, rationing their supplies. Stores and wholesalers were also having trouble meeting the demand (of Vancouver, 2017).

By the end of December, the city of Vancouver had spent \$2.5 million on snow and ice reduction, triple the amount used in the previous two winter periods combined (of Vancouver, 2017). Even after the unusual cold winter, Vancouver still needed to address the nearly 15,000 potholes (almost double than normal) affecting drivers (Vancouver, 2017).

1.1. *Understanding Extreme Weather Events and its Impacts — A Case Study*

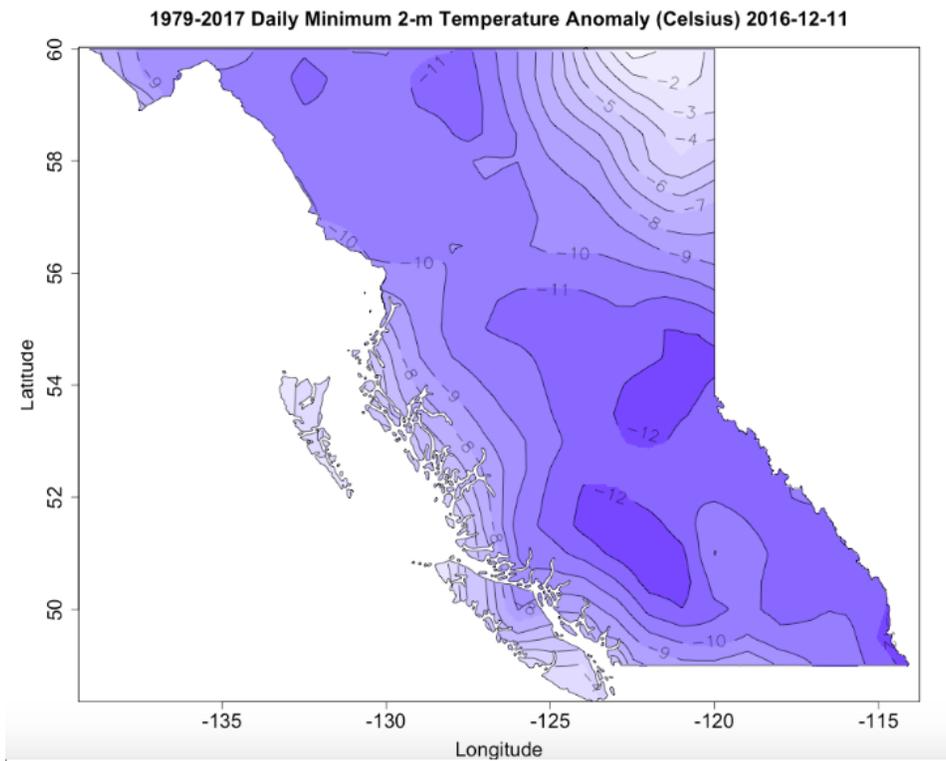


Figure 1.5: Fifteen-day averaged minimum 2-m temperature anomaly centered on December 11th, 2016, relative to 1979-2017 period.

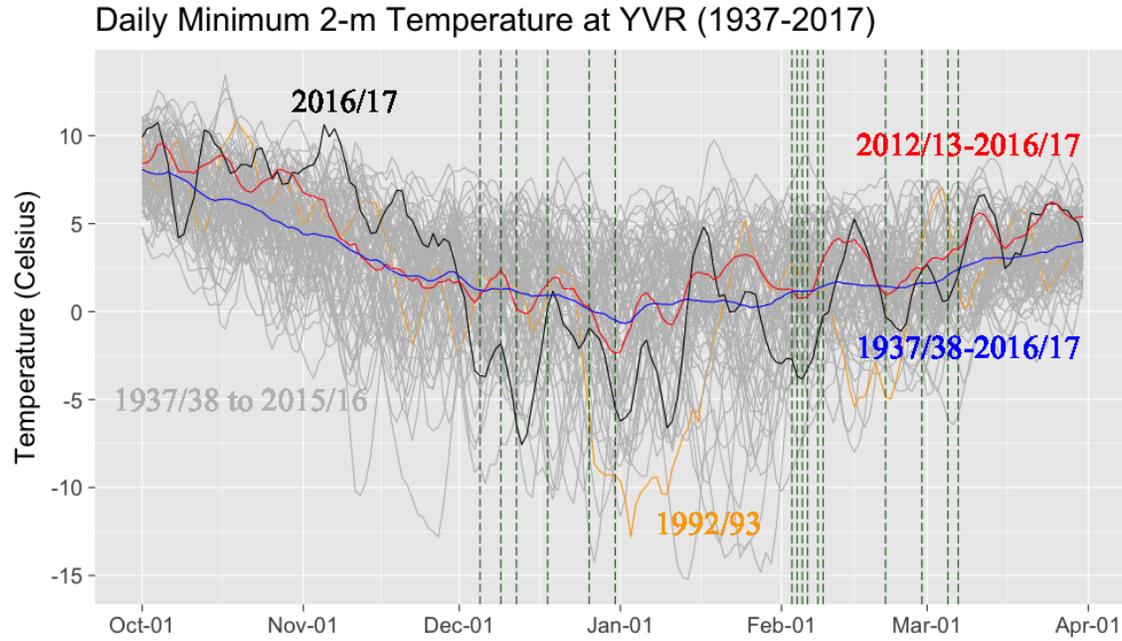


Figure 1.6: Daily minimum 2-m temperature time series for Vancouver, smoothed using a 5-day rolling window for readability. Time series for all years in grey, 1937/38-2016/17 average in blue, 2012/13-2016/17 average in red, 1992/92 in orange and 2016/17 in black. Days with snowfall in 2016/17 indicated with green dashed lines.

1.1. *Understanding Extreme Weather Events and its Impacts — A Case Study*

Long waits, cancellations and delays also left commuters questioning Vancouver’s public transportation system readiness for such cold winter weather (McElroy, 2017a) (Fig. 1.7a). Additionally, strong arctic outflow winds over the Salish Sea during the outbreaks led B.C. Ferries to cancel sailings, impacting ferry commuters (Brend, 2016).

During the arctic outbreaks, approximately \$5 million was spent to operate an ice-clearing cable collar system on the Port Mann Bridge —the province’s primary east-west corridor for both commercial and commuter traffic. By contrast, just \$300,000 was spent to operate it in 2015/16. Lane closure due to crews clearing the bridge and Highway 1 led to major traffic problems during the last arctic outbreak on February 3rd, 2017 (Saltman, 2017) (Fig. 1.7b).

During the third arctic outbreak on January 3rd, 2017, BC Hydro set a new record for power consumption, breaking the old record set on Nov 29th, 2006 (Paetkau, 2017). Finally, in the last outbreak in early February, a storm cycle brought record-breaking snow and freezing rain to the Fraser Valley. Abbotsford observed 57.8 cm of snow and there were reports of up to 80 cm in Chilliwack (MacMahon, 2017). Freezing rain accumulated 2-4 cm (Meuse, 2017). 361,000 BC Hydro customers lost power (Luymes, 2017) and Vancouver was cut off from the rest of BC with highways 1, 3, 5 and 99 all closed (McElroy, 2017b). The large-scale features providing the atmospheric context of these intense arctic outbreaks will be described next.

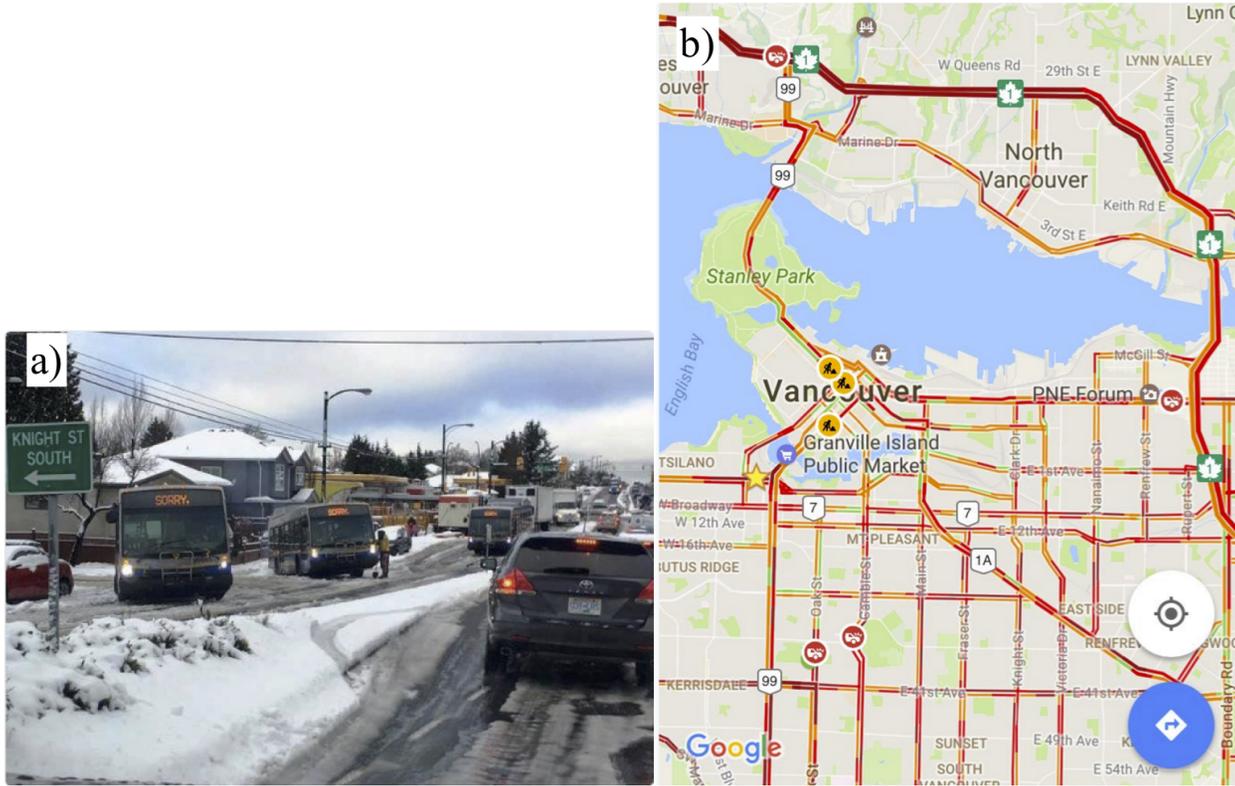


Figure 1.7: (a) Public transit buses paralyzed by road conditions during the December 11th, 2016 arctic outbreak (Fritz, 2016). (b) A Google Maps screen capture, where orange and red colours indicate widespread very slow or stopped traffic during the February 3rd, 2017 arctic outbreak.

1.1.3 Anatomy of an Arctic Outbreak

The five arctic outbreak events were each comprised of a cold continental airmass associated with a strong arctic high pressure system in Alaska and the Yukon, combined with a warmer, moist air mass associated with a Pacific low pressure system. The lows played an important role in producing periods of snow across the South Coast.

In far northern latitudes, during winter the sun angle is low or below the horizon. This means incoming solar radiation is very limited, while outgoing radiation from the earth's surface continues unabated. This creates a negative surface energy budget and the air cools, building very cold airmasses at the surface. Such airmasses are associated with strong high pressure such as the ones seen over Alaska and the Yukon during the first outbreak between December 4th, 2016 and December 6th, 2016 (blue shading in Fig. 1.8a, and red sea level pressure contours in Fig. 1.8b).

The cold, dense, stable airmass flows through valleys, fjords, and straits on its way to Vancouver, partially blocked by the higher mountainous terrain of British Columbia (Figs. 1.8c and d). Where the valleys, fjords and straits widen, the cold air spreads and thins, accelerating into arctic outflow winds (Jackson, 1996). An example is shown in a fine-resolution computer model forecast for the second arctic outbreak on December 12th (Fig. 1.1). These outflow winds have impacts of their own, like the cancelled BC Ferries sailings in the Strait of Georgia mentioned in the previous section.

Often an arctic outbreak event will draw to a close with the approach of a Pacific low pressure system ushering in moist, mild air. In the December 5th outbreak, an upper-level low over the Gulf of Alaska moved southward down the coast (Figs. 1.8a and c). With low-level arctic air in place, the upper-level trough and associated surface low pressure brought 5-15 cm of snow across the South Coast, with 5.4 cm measured in Vancouver (Figs. 1.8c and d). While this is relatively little snow compared to snowfalls in other parts of Canada, it caused major disruptions.

1.1.4 Final Considerations

The fall season was abnormally wet for multiple months, especially in terms of frequency of rain. This culminated in an early November storm cycle that caused headaches for emergency management personnel, and featured greater than 1-in-100-year cumulative flows for parts of Vancouver Island. A series of five arctic outbreaks led to a well-below-normal winter, and an abnormally large number of days with below-freezing minimum temperatures.

1.1. Understanding Extreme Weather Events and its Impacts — A Case Study

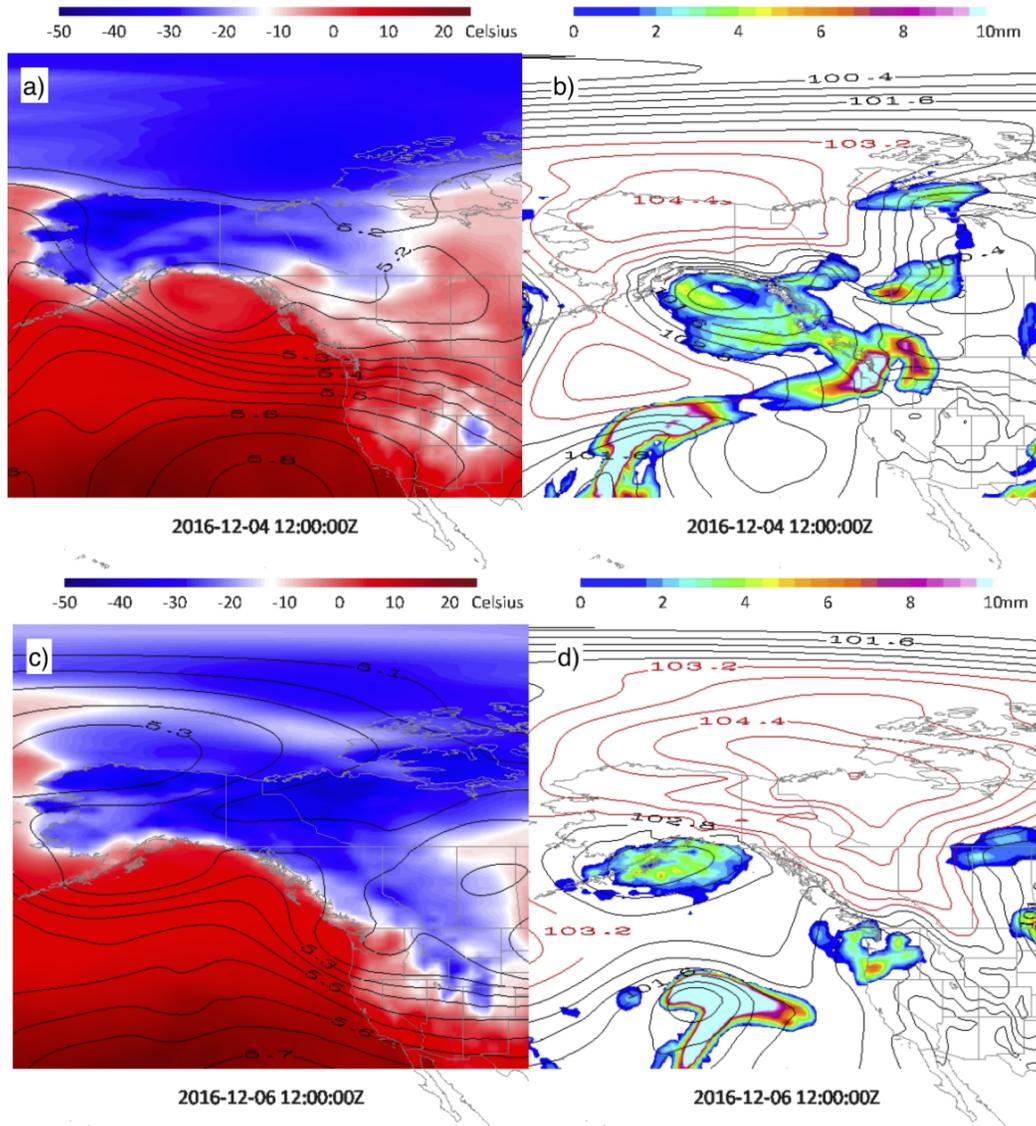


Figure 1.8: (a) 50.0-kPa geopotential height (km) and 2-m temperature ($^{\circ}\text{C}$), (b) sea level pressure (kPa) and 6-hr precipitation (mm) for December 4th, 2016 at 1200 UTC. (c) and (d), as in (a) and (b), but for December 6th, 2016 at 1200 UTC. Data from the the ECMWF interim reanalysis (ERA-Interim) (Dee et al., 2011).

The result was a 6-month period of weather that, by the metrics discussed herein, was the worst in recent memory, and among the worst in history for the South Coast of BC. Chapter 4 gives more details about how extreme these seasons were, and how well they were forecast.

1.2 Extreme Data Analysis

Over the last few decades, techniques that can provide forecast users with advanced warning of extreme weather events have been developed. To avoid data waste, these techniques work by comparing model forecasts of variables to historical means and climate distributions (Lalauette, 2003; Dutra et al., 2013). For instance, the extreme forecast index (EFI) (Lalauette, 2003; Zsoter, 2006) developed at the European Centre for Medium-Range Weather Forecasts (ECMWF) indicates how far the forecast distribution for an event deviates from the climate distribution. Another example is the situational awareness table developed by National Oceanic and Atmospheric Administration and the National Weather Service (NOAA/NWS) (Grumm and Hart, 2001; Graham et al., 2013). It provides standardized anomalies, percentiles, and return periods of forecasts by comparing them to a reanalysis or to a so-called “model climate”.

Such techniques provide forecasters and users with more information on which to base their decisions by making extreme events easier to identify and placing them in historical context. However, there is room for improvement because they do not utilize extreme data analysis methods.

Extreme levels can be estimated by fitting an appropriate distribution to a sample of observed extreme values recorded over a specified period. According to the extremal types theorem, in the limit as the sample size of observed extreme values increases, the distribution stabilizes and approaches one of the three extreme value distributions (EVD), regardless of the distribution from which the original extreme values were drawn from (Gnedenko, 1943; De Haan, 1976). The three EVD distributions can be simplified to one Generalized Extreme Value (GEV) distribution function (Jenkinson, 1955). This theorem is analogous to the central limit theorem, which states that in the limit as the sample size of mean values becomes large, the shape of the sampling distribution converges to a Gaussian distribution. Both theorems suggest extreme and mean values behave differently and as such should be analyzed separately.

The above explanation might seem unnecessary at first, but to make the point clear, it implies an appropriate statistical method of extremes needs

to be employed, rather than defining an extreme value as being a certain distance away from mean values in order to avoid data waste. Second, the convergence behaviour of the EVD is quite important. It implies that a long historical record is needed for extreme data analysis to be correctly employed and justify data waste. Therefore, to lay the ground work for an improved statistical model of extreme weather events, first the best historical dataset needs to be identified to accurately place extreme T2M and PCP into a historical context. The challenge is a lack of long-term complete historical surface weather station records at many locations across the province (Karl et al., 1993; Odon et al., 2018).

A new methodology to address extreme weather events is introduced in Chapter 2 and used thereafter throughout the entire dissertation.

1.3 A Historical Record

Weather station coverage across BC is limited outside of the southwest BC population centres (Karl et al., 1993; Odon et al., 2018), and often the quality and completeness of station records degrades as well. This data paucity motivates the search for the best model climate from a gridded reanalysis dataset. This would provide a long, continuous record with complete spatial coverage over BC.

A reanalysis is a 3-D gridded dataset created by combining rerun model forecasts with observed weather data via data assimilation. The result is a long-term gridded dataset that is spatially complete and physically consistent. They contain a large variety of atmospheric variables including many that are not directly observed (Dee et al., 2011), potentially rendering a feasible solution to the paucity of observational data over BC.

Some of the most well-known reanalysis datasets are the Climate Forecast System Reanalysis (CFSR) from the National Centers for Environment Prediction (NCEP), the ECMWF interim reanalysis (ERA-Interim), the Japanese 55-year Reanalysis (JRA-55) from the Japanese Meteorological Agency (JMA), and the Modern Era Retrospective-Analysis for Research and Applications (MERRA-2) from the National Aeronautics and Space Administration (NASA). They represent the current generation of reanalyses with improvements over previous generation reanalyses R1 (NCEP/NCAR Reanalysis I) and R2 (NCEP/DOE Reanalysis II), ERA-15 (ECMWF 15-year Re-Analysis) and ERA-40 (ECMWF 40-year Re-Analysis), JRA-25 (Japanese 25-year Reanalysis), and MERRA (Modern Era Retrospective-Analysis for Research and Applications) (Dee et al., 2011; Saha et al., 2010).

Even though the observational data assimilated by the four reanalyses are largely identical, their assimilation methods are not, resulting in significant differences (Mooney et al., 2011; Saha et al., 2010; Koster et al., 2016). Such differences can have an impact on how well extreme weather events are represented.

New techniques to assess the performance of the new generation reanalyses — CFSR, ERA-Interim, JRA-55 and MERRA-2 — with respect to daily and extreme maximum and minimum T2M, and daily and extreme PCP over mountainous BC are described in Chapters 2 and 3 respectively.

1.4 A Gridded Climatological Dataset

Reanalyses are often downscaled to a higher resolution (Cosgrove, 2003; Juang and Kanamitsu, 1994; Rasmussen et al., 2011; Stefanova et al., 2012; Abatzoglou, 2013). High-resolution gridded temperature and precipitation datasets are widely used because of the benefits of spatial completeness and fine-scale features due to topography. Such details are important for modelling applications in fields like hydrology, ecology, and agriculture (Thornton et al., 1997; Mote et al., 2005; Abatzoglou, 2013; Stoklosa et al., 2015).

Across BC there are two types of state-of-the-art datasets suited for long-term analysis: the very high-resolution Parameter-Elevation Regressions on Independent Slopes Model climatology (PRISM; Daly et al. (1994, 1997, 2002)), which provides access to a 30-arc-second (~ 800 m) gridded monthly-mean and annual-mean maximum and minimum temperature, and precipitation values for the 1981-2010 period (PCIC et al., 2014); and the Canadian homogenized stations dataset, which accounts for non-climatic changes in the data such as changes in station siting, instrumentation, time of observation and procedure (Vincent, 1998; Vincent and Gullett, 1999; Vincent et al., 2002; Mekis, 2005; Mekis and Brown, 2010; Mekis and Vincent, 2011). However, PRISM lacks temporal resolution and the stations lack spatial completeness.

The temporal resolution of the reanalysis, the spatial resolution of PRISM, and the homogeneity and ground truth of the stations are therefore combined to create a new, very-high-resolution surface analyses (hereafter referred to as the VHRSA). The methodology is described in Chapter 4.

1.5 Forecasting Extreme Weather Events

Mountain ranges can dramatically impact precipitation distribution due to orographic lifting and blocking, among other processes. Similarly, temperature exhibits large spatial variability in complex terrain. This is in part due to typical changes in temperature with height (i.e., the lapse rate), but also due to mountain-specific processes like cold air pooling, heating and cooling of near-slope air, diabatic processes involved in cross-barrier airmass transformation (Smith et al., 2003), and blocking of temperature advection, among others. All these processes make the spatial distribution of precipitation and temperature quite complicated in the complex terrain of BC (Deng et al., 2005; Astsatryan et al., 2015). Lack of model resolution to resolve terrain features and processes make accurate forecasts in mountainous regions particularly challenging (Junker et al., 1992; Kunz and Kottmeier, 2006; Smith et al., 2010; Haren et al., 2015).

Although numerical weather prediction models are far less frequently verified over mountainous regions, several studies investigated whether high-resolution models performed better than their lower-resolution counterparts across complex terrain. ?, ? and ? suggest that at higher resolutions (< 4 km) there are still major problems associated with precipitation and temperature forecasts. The model microphysics parameterizations lead to overprediction along the steep windward slopes and underprediction in the lee of major barriers. In contrast, Schirmer and Jamieson (2015), Weusthoff et al. (2010), Garvert et al. (2005) and Ikeda et al. (2010) concluded that higher-resolution models were equal or better at simulating orographic influences compared to the low-resolution models. However, all studies agree that higher resolutions models better resolve terrain features.

The forecast methodology described in Chapter 4 both bias corrects and downscales the North American Ensemble Forecast System (NAEFS) using the VHRSA dataset.

1.6 Dissertation Organization

The outline of the dissertation is as follows. Chapters 2 and 3 evaluate reanalyses to identify the one that yields the best long-term and complete T2M and PCP historical dataset over BC, respectively. A new methodology is also introduced in both chapters to address how extremes in T2M and PCP vary across space and time (around the calendar year). Chapter 4 introduces a new statistical model that is used to forecast extreme T2M and

1.6. Dissertation Organization

PCP, that accounts for the issues of model resolution and non-stationarity (changes due to climate change, urbanization, and other effects). The results are summarized in the conclusion.

Chapter 2

Performance of Reanalyses across British Columbia. Part I: Evaluation of Daily and Extreme 2-m Temperature

2.1 Introduction

Performance of the new generation reanalyses — CFSR, ERA-Interim, JRA-55 and MERRA-2 — is assessed with respect to daily and extreme maximum and minimum 2-m temperature (hereafter T2M) over mountainous BC. The purpose of this Chapter is to identify which reanalysis best represents observations as a surrogate climatological dataset. The end goal is for this reanalysis to serve as a tool for the creation of a new extreme forecast index in Chapter 4.

Previous studies have shown differences in T2M between the ERA-Interim and CFSR, and previous generation reanalyses ERA-40 and R1 over the Tibetan Plateau in central Asia (Bao and Zhang, 2013). Betts et al. (2009) showed that the seasonal cycle of mean T2M is higher than observations over the Mississippi River basin for both ERA-40 and ERA-Interim. Berg et al. (2003) found that ERA-40 has substantial biases in T2M and dewpoint temperatures over land in North America. Wang et al. (2011) evaluated the mean T2M performance over the entire globe from a mix of new and previous generation reanalyses (ERA-40, R1, R2 and CFSR). More recently, Lindsay et al. (2014) evaluated the performance of mean T2M over the Arctic for CFSR, MERRA-1, ERA-Interim, and JRA-25. All studies agree that newer reanalyses outperform previous generation reanalyses.

Both daily and extreme maximum and minimum T2M are analyzed, (defined in the Data and Methodology section), and compared with weather

stations in BC for the period 1980-2010. T2M is examined because of its importance and impact on ecosystems (Parmesan et al., 2000), agriculture (Rosenzweig et al., 2001), tourism (Patz et al., 2005; White et al., 2016), urbanization and health (Curriero et al., 2001; White et al., 2016; Odon et al., 2017), and due to observational data availability (Jones et al., 1985; Peterson and Vose, 1997; Jones et al., 1999).

Second, trends in both daily and extreme T2M during the study period are examined in order to determine if significant statistical changes occurred over time. This is to assess whether or not a stationary climatological distribution is appropriate to represent present-day daily and extreme distributions.

The outline of the Chapter is as follows. In section 2.2, a brief description of the different reanalyses and of the weather-station observations is given. In sections 2.3- 2.5 a methodology for dividing BC into climate zones, the various metrics used for evaluating daily and extreme reanalysis T2M, and a method for assessing statistical nonstationarity are described. In section 2.6, the daily and extreme T2M from the reanalyses are evaluated. In section 2.7, the trends of both daily and extreme T2M are examined, and the corresponding changes in return levels and return periods of extremes are discussed. The results are summarized in section 2.8.

2.2 Data and Methodology

Daily maximum/minimum T2M from 57 weather stations from 1 Jan 1980 to 31 Dec 2010 are used in this study for evaluation of the CFSR, ERA-Interim, JRA-55 and MERRA-2 reanalyses. A description of the weather station dataset is given below. The authors chose the 1980-2010 as the overlapping period for comparison because the MERRA-2 began in 1980 (Gelaro, 2015); and because in 2011 the CFSR was extended using NCEP's Climate Forecast System Version 2 (CFSv2) operational model, and differences between the models used to produce CFSR and the operational CFSv2 may affect data evaluation across the 2010/2011 boundary period (Saha et al., 2014).

A description of the different reanalyses and of the weather station dataset is given below. A summary of the reanalyses atmospheric models and configurations is presented in Table 2.1.

T2M from the weather stations used in this study are not assimilated by the CFSR Wang et al. (2011), or the MERRA-2 (Bosilovich et al., 2015). The SYNOP stations (surface synoptic observations) are directly assimilated by the ERA-Interim (Dee et al., 2011) and the JRA-55 (Kobayashi

2.2. *Data and Methodology*

et al., 2015) (see Appendix A for assimilated stations). Therefore evaluating against these observations provides a dependent measure of accuracy. A discussion of how the reanalyses perform across non-assimilated stations is given in section 2.6.

Table 2.1: Overview of the four reanalysis datasets examined in this study.

Institution	Reanalysis Model	Assimilation method	Period	Download grid (<i>lat</i> × <i>lon</i>)	Time interval	Reference	
NCEP/NCAR	CFSR	CFS T382/L64 (global horizontal resolution ~ 38 km)	3D-Var GSI (current CFSv2)	1979-2011 as	$0.5^\circ \times 0.5^\circ$ (~ 50 km)	0000, 0600, 1200 and 1800 UTC	Saha et al. (2010)
ECMWF	ERA- Interim	IFS T255/L60 (global horizontal resolution ~ 79 km)	4D-Var	1979-current	$0.5^\circ \times 0.5^\circ$ (~ 50 km)	0000, 0600, 1200 and 1800 UTC	Dee et al. (2011)
JMA	JRA-55	JMA T319/L60 (global horizontal resolution ~ 55 km)	4D-Var	1958-2012 (current JCDAS)	as $0.5616^\circ \times 0.5616^\circ$ (~ 55 km)	0000, 0600, 1200 and 1800 UTC	Ebita et al. (2009)
NASA	MERRA- 2	GEOS- 5.12.4 AGCM (<i>lat</i> × <i>lon</i>) $0.5^\circ \times 0.625^\circ$ /L72	3D-Var GSI	1980-current	$0.5^\circ \times 0.625^\circ$	0030, 0130,... ...2330 UTC	Gelaro (2015); Gelaro et al. (2017)

2.2.1 Weather station data

The stations were initially selected because of their proximity to population centres and so that they would be geographically dispersed around BC. They represent conditions for a mixture of coastal, intermountain and continental climates. Ideally, variations in climatological time series should be caused only by changes in weather and climate. However, decades-long time series can be affected by inhomogeneities such as missing data, changes in instrumentation, station relocation, changes in the local environment surrounding the weather station such as urbanization or land use cover, and changes in time of observation, to name a few (Peterson et al., 1998; Mekis and Hogg, 1999; Jones et al., 1985). Some changes cause sudden discontinuities while other changes, such as changes in the local environment around the station, cause gradual trends in the data.

Of the 74 stations initially selected for the study, 8 were eliminated due to either station relocation or discontinuation during the 1980 to 2010 study period. Additionally, eight stations with more than 4% missing data were excluded, and one station was eliminated due to poor data quality. Of the remaining 57 stations, 48 are from Environment and Climate Change Canada (ECCC) and nine are from BC Hydro, the primary electric utility company for BC.

Figure 2.1 shows the locations of all 57 stations overlaid with the population distribution across BC, and with the province's political boundaries. Red and blue stations represent ECCC and BC Hydro stations, respectively, and their corresponding three-letter abbreviations that are referenced in the paper. Fifty-two of the stations are located in valleys (indicated by upside-down triangles), and five are in non-valley locations (upright triangles). Of the 4.8 million people living in BC, about half live in the Vancouver metropolitan area in southwest BC.

Temporal data homogeneity for ECCC stations was assessed by identifying non-climatic shifts in the annual and monthly means of the daily maximum and minimum T2M using techniques based on regression models (Vincent, 1998; Wang et al., 2007). The non-climatic shifts were mainly due to station relocation, or changes in observing practices and automation (Vincent and Gullett, 1999). Monthly and daily maximum and minimum temperatures were then adjusted for the shifts identified in the temperature annual series (Vincent et al., 2002).

BC Hydro station data were manually quality controlled based on range limits, spatiotemporal consistency, present weather conditions, and are used to verify reanalysis performance at non-valley stations. Initial stationarity analysis indicated potentially spurious trends in the data, so they are excluded from the analysis of nonstationarity in daily and extreme T2M. Furthermore, ECCC stations with more than 1% missing data are also excluded, leaving 26 ECCC stations for the nonstationarity analysis.

2.2.2 CFSR

In 2010, NCEP introduced the CFSR. Previous NCEP reanalyses have been among the most used NCEP products in history. Many known errors in the assimilation of observational data and execution of previous reanalyses were corrected in the CFSR, resulting in a superior product in most respects (Saha et al., 2010).

CFSR uses NCEP’s global coupled atmosphere-ocean model. It consists of a spectral triangular atmospheric grid (Saha et al., 2006) at a horizontal resolution of T382 (~ 38 km) and a hybrid sigma-pressure system with 64 vertical levels extending from the surface to approximately 0.26 hPa.

CFSR was the first NCEP global reanalysis to directly assimilate satellite radiances, and to use three-dimensional variational data assimilation (3D-Var) in a Gridpoint Statistical Interpolation (GSI) scheme rather than a Spectral Statistical Interpolation (SSI) scheme.

The T2M variable in the CFSR is derived primarily from satellite radiances and radiosonde information. Observations of T2M from land stations are not assimilated (Saha et al., 2010).

2.2.3 ERA-Interim

The ECMWF introduced the ERA-Interim in 2011, in part to replace ERA-40 (Dee et al., 2011). The ERA-Interim configuration uses a spectral T255

horizontal resolution (~ 79 km), and a hybrid sigma-pressure system with 60 vertical levels with the top of the atmosphere located at 0.1 hPa.

Some of the improvements over ERA-15 and ERA-40 include a new humidity analysis; new model physics where the prognostic equations are solved using a semi-Lagrangian scheme, a variational bias correction of satellite radiances (Dee et al., 2011); and improvements on various technical aspects of reanalysis such as data selection, quality control, bias correction, and performance monitoring; each of which can have a major impact on the quality of the reanalysis. Additionally, the use of 4D-Var for the atmospheric analysis in ERA-Interim is a major step forward. The improved 4D-Var data assimilation scheme allows the cost function to be minimized over a 12-h assimilation time interval rather than a single time, as is the case for ERA-40 and CFSR.

The T2M variable is directly assimilated by the ERA-Interim from SYNOP stations (Dee et al., 2011).

2.2.4 JRA-55

In 2011 the JMA produced the Japanese 55-year Reanalysis (JRA-55) (Ebita et al., 2009). The JRA-55 extends 55 years from 1958 to 2012, and will be continued in real time as the JMA Climate Data Assimilation System (JCDAS). JRA-55 uses a spectral model integrated at a TL319 (~ 55 km) horizontal resolution with 60 vertical levels up to 0.1 hPa in hybrid sigma-pressure coordinates.

It employs a 4D-Var scheme which seeks the initial condition that best fits the forecast to the observations within a 6-h assimilation interval rather than a single time. The reanalysis also contains a new radiation scheme, variational bias correction for satellite radiances, an update on dynamical and physical processes such as the prognostic equations being solved in a semi-Lagrangian form rather than Eulerian (Ebita et al., 2011; Takeuchi et al., 2013). These upgrades significantly reduce model biases versus the JMA's previous generation reanalysis, enhance the dynamical consistency of analysis fields, and advance the handling of satellite radiances.

Similarly to ERA-Interim, T2M is directly assimilated from SYNOP stations (Ebita et al., 2011).

2.2.5 MERRA-2

The new Modern Era Retrospective-Analysis for Research and Applications (MERRA-2) was produced by NASA in 2015. The grid used for MERRA-2

is 0.5° latitude 0.625° longitude (~ 55 km) with 72 vertical levels in hybrid sigma-pressure coordinates, from the surface to 0.01 hPa.

It uses an upgraded 3D-Var assimilation scheme based on the GSI with a 6-h update cycle. Some other improvements over NASA’s previous generation reanalysis include an updated physics model, aerosol assimilation, corrections in precipitation for land surface and imbalances in water and energy cycles, to name a few (Rienecker et al., 2011; Gelaro, 2015; Gelaro et al., 2017).

Similarly to CFSR, the T2M is derived primarily from satellite radiances, and radiosonde, aircraft and wind-profiler information. Observations of T2M from land stations are not assimilated (Gelaro et al., 2017).

2.3 Climate zones

The location of BC immediately east of the Pacific Ocean, and its complex topography, produce distinctive climate zones that vary with elevation, location relative to ocean and mountains, and latitude (Peel et al., 2007). Climate zones vary from the wettest in Canada to the hottest and driest, with the formidable barrier of the Coast Mountains serving as the dividing-line between the two.

Due to temperature ranges and variability that are unique to each climate zone (Moore et al., 2008), reanalysis performance is evaluated separately for each zone.

To determine the different climate zones, Principal Component Analysis (PCA) is conducted on the daily maximum and minimum T2M correlation matrix for each station. The corresponding station correlation is calculated from the daily temperature values for the entire study period. The high correlation values (Fig. 2.2) are owed to annual seasonal effects, which in section 2.5 becomes an important factor as the authors want to evaluate the degree to which the reanalyses capture the annual cycle across the different climate zones.

The Principal Components are ordered by the size of their respective eigenvalues — their rank corresponds to their relative importance in describing temperature variations. 93% of the variability of both maximum and minimum temperature can be explained by the first four components, which are retained, leading to a significant reduction of data while retaining most of the variance.

A K-means clustering analysis is then performed on the four components to find a natural grouping of the data, where each component belongs to

2.3. Climate zones

the cluster with the nearest centroid. One to 10 clusters were tested and a five-cluster solution was chosen (Fig. 2.2). After five clusters, the additional decrease in the within-cluster dissimilarity is not substantial, and the number of stations per cluster becomes too low. The five climate zones (encircled by dashed lines in Fig. 2.1) are subsequently matched to those identified by Chilton (1981) and Moore et al. (2008):

1. *Islands and Coast Mountains (hereafter Maritime):*

Heavily influenced by the Pacific Ocean, this zone includes island stations on the immediate coast, windward of the Coast Mountains. It has mild winters and cool summers. Fall, winter, and spring feature frequent landfalling low pressure systems, while summers are fairly dry. High elevations accumulate very deep snowpacks in the winter and are extensively glaciated.

2. *Coast Mountains (hereafter Southwest):*

The climate of the Coast Mountains is similar to the Maritime climate, with a slightly less pronounced maritime influence. This more protected region has slightly more sunshine and greater temperature variability relative to the Maritime climate zone.

3. *Interior Plateau (hereafter Central):*

On the lee side of the Coast Mountains, the Interior Plateau is a broad, elevated region broken occasionally by narrow valleys. In this drier and more continental climate, seasonal and diurnal differences in temperature are much greater than at the coast. Summers tend to be hot and dry; winters cooler and less moist. The Okanagan Valley is the southernmost, hottest, and driest part of the province and in all of Canada.

4. *Columbia Mountains and Southern Rockies (hereafter Southeast):*

Farther east, westerly winds again ascend the Columbia and Rocky Mountains. These ranges also restrict easterly flow of cold, continental arctic air into the region. High elevations are wet and cool. Valley bottoms are semi-arid with hot summers, and frequent, cold temperatures below temperature inversions in the cool season.

5. *The Interior Plains, Northern and Central Plateaus and Mountains (hereafter North):*

In the Northern and Central Plateaus and Mountains, winters are colder and drier than in the South, due to frequent influxes of continental arctic air, and fewer Pacific storms. Summers are short, warm, and wetter than southern BC due to the northward migration of the storm track.

To the east of the northern Rocky Mountains lies an extension of the Great Plains. It sees limited precipitation due to the many topographic barriers between it and the Pacific Ocean, and very frequent influxes of continental arctic air due to proximity and a lack of topographic barriers to the north and east. This area experiences a continental climate with long, cold, dry winters and short, warm summers.

2.4 Verification Metrics

The statistical behaviors of daily and extreme T2M are compared between observed weather station data, and their corresponding location in the reanalyses. In order to evaluate the agreement between observations at station locations and reanalyses at grid points, the methods Nearest Neighbor, Inverse Distance Weighting (IDW), Bilinear and Bicubic interpolation are investigated for comparison with the observed values. No method is superior across all locations and IDW is adopted. A more detailed description of each interpolation method can be found in Mooney et al. (2011) and Stahl et al. (2006).

2.4.1 Daily T2M

To obtain a smooth climatology of daily maximum and minimum T2M averaged over 31 years for each calendar day, a 31-day centered rolling window was applied to both reanalysis and observed temperatures. The rolling average is the unweighted mean of the 961 sample values centered around a calendar day. This window length was chosen because it smooths out noise related to extreme weather events while still capturing monthly variations in daily T2M. For the reanalyses, the daily maximum (minimum) T2M is defined as the highest (lowest) value of the six-hourly T2M outputs for CFSR, ERA-Interim and JRA-55; and as the highest (lowest) value of the hourly T2M outputs for MERRA-2 in a calendar day. The Canadian meteorological convention is followed, where it defines a calendar day to be from 0601 UTC of the current day to 0600 UTC of the following day (Meteorological Service of Canada, 2015) .

The bias (or systematic error), the random error, and the two-sample Kolmogorov-Smirnov (KS) statistic are then computed to estimate how accurately each reanalysis captures T2M.

Bias is the average difference between the reanalysis and the observation. The random error, measured by centered root mean squared error (CRMSE), expresses how concentrated the errors are from the mean bias on a given calendar day. The two-sample KS statistic is used to determine the largest absolute difference between the reanalysis empirical cumulative distribution function (ECDF) and the observation ECDF. It detects the general difference between the two distributions.

2.4.2 Extreme T2M

In addition to societal impacts, extreme temperatures are a concern for power utilities primarily because of their effect on load. That is, during periods of extreme heat, electrical load (demand) is high due to air conditioning. The same goes for extreme cold and heating. Other impacts include transmission line thermal ratings — wherein power transmission can be limited during periods of hot weather, limiting a utility’s ability to supply power for consumption or market trading. It is important for utilities to identify and anticipate such events, so that they can plan optimal system operations in light of abnormal load considerations.

A return level is a value of T2M that is expected to occur, on average once every return period. For extreme T2M, the 2- and 30-yr return levels are examined, which represent significant and extreme departures from normal, respectively (2- and 30-year return levels are also known as 2- and 30-year recurrence intervals, or 0.5 and $0.0\bar{3}$ Annual Exceedance Probabilities (AEP)).

To do this, a 31-day centered rolling window is used over the 31-year period to obtain 31 values of annual maximum (minimum) T2M for each calendar day, one value for each of the 31 years. In other words, the 31 annual maximum (minimum) T2M in a given 31-day centered rolling window are the highest (lowest) values in each of the 31 years, which then comprise 31 values of extreme T2M for each calendar day. As with daily T2M, the window length struck a balance between having a large enough sample size while ensuring that each day within the window would have a similar climatological distribution.

Return levels can be estimated by fitting an appropriate distribution to a sample of extreme values. According to the theorem of extremal types, in the limit as the sample size of extreme values becomes large, the distribution sta-

bilizes and approaches one of the three extreme value distributions (EVD), regardless of the distribution from which the original T2M was drawn (Gnedenko, 1943; De Haan, 1976). This theorem for extreme values is analogous to the central limit theorem for mean values, which states that in the limit as the sample size of mean values becomes large, the sampling distribution converges to a Gaussian distribution. Both theorems suggest extreme and mean values behave differently and as such should be analyzed differently. The three EVD distributions can be simplified to one Generalized Extreme Value (GEV) distribution function (Jenkinson, 1955).

A GEV dresses these 31 sample values for each calendar day by the method of L moments. The method of L moments is chosen because it produces more efficient estimates of return levels for small sample sizes than the method of maximum likelihood (Hosking et al., 1985; Hosking, 1990). The small sample size of 31 extremes values is a concern because: 1) the three EVD distributions converge at different rates (Davis, 1982), and 2) the daily T2M from which the extreme values are sampled are normally distributed and serially correlated, resulting in a slower convergence rate to either of the three distributions (Leadbetter et al., 1983). Thus, a goodness-of-fit test is conducted to evaluate the closeness of the fitted GEV distributions to the data.

A Lilliefors test compares the largest absolute difference between the fitted GEV cumulative distribution function (CDF) and the observation ECDF. The null hypothesis is that the observed data is drawn from a GEV distribution. A sufficiently large critical difference between the fitted GEV CDF and the observation ECDF results in the null hypothesis being rejected.

A parametric bootstrap procedure determines this critical difference. Namely, 100 samples of size 31 are generated from the fitted GEV distribution for each calendar day at each station. Then, 100 critical differences are derived from the comparison of each generated sample ECDF and the fitted GEV CDF. The 90th percentile of the resulting collections of critical differences is used as the critical difference for the rejection of the null hypothesis that the sample originates from a GEV distribution, which corresponds to accepting or rejecting the null hypothesis at the $\alpha_0 = 0.10$ significance level.

At this 10% significance level, 10% of the tests are expected to exceed a critical difference, rejecting the null hypothesis. However, there are 57 weather stations and 365 calendar days totaling $N_0 = 20805$ independent Lilliefors tests. To reduce the probability of incorrectly rejecting one or more of the N_0 true null hypotheses, Walker's criterion with $\alpha_{Walker} = 1 - (1 - \alpha_0)^{1/N_0} = 5.06 \times 10^{-6}$ is regarded as significant. Namely, the null hypothesis is rejected when the critical difference between the fitted GEV

CDF and the observations ECDF exceeds the $(1 - \alpha_{Walker})^{\text{th}}$ percentile of the resulting collections of critical differences between the fitted GEV CDF and the generated ECDF of the samples (Wilks, 2016).

Less than 4% (6%) of the locations and calendar days are rejected during the 1980-2010 study period, suggesting maximum (minimum) extreme T2M behaviour can be described by a GEV distribution.

The 2- and 30-year return levels are then estimated from the GEV distribution for each calendar day. The bias of the 2- and 30-year return levels are calculated to estimate how well each reanalysis captures extreme T2M.

The 2- and 30-yr return levels of maximum and minimum extreme T2M are chosen because less than 1% of their 90% respective confidence intervals overlap, indicating the difference between the two return levels is statistically significant. The sampling uncertainty of the estimates of the 2- and 30-yr return levels are determined by a parametric bootstrap procedure. 100 samples of size 31 are generated from the fitted GEV distribution for each calendar day at each station. Then, the 2- and 30-year return levels are estimated from each generated sample. The 5th and 95th percentiles of the resulting collection of 2- and 30-year return levels is used as lower and upper bounds of the 90% confidence intervals for the true 2- and 30-year return levels.

2.4.3 Analysis of Variance (ANOVA)

The mean systematic error of daily maximum and minimum T2M, and of 2- and 30-year return levels are calculated for each location from all calendar day systematic errors. Comparisons between mean systematic errors of the reanalyses (CFSR, ERA-Interim, JRA-55 and MERRA-2) and of the observed levels of T2M (daily maximum, daily minimum, extreme maximum and extreme minimum) are made using a two-way ANOVA. Tukey's honest significant difference (HSD) is applied following statistical significance in the ANOVA to identify differences in pairwise comparisons of reanalyses mean systematic errors.

2.5 Nonstationarity

Statistical nonstationarity is tested to determine if there are important temporal changes in daily and extreme T2M. Specifically, is nonstationarity substantial enough to require a more complex characterization of extreme levels that takes into account temporal changes, or is a comparatively simpler stationary model accurate enough to represent daily and extreme T2M?

2.5.1 Daily T2M

For daily T2M, a 91-day centered rolling window is used to obtain 2821 values of maximum and minimum T2M for each calendar day over the 31 years. The rolling window is extended to 91 days to increase the sample size while still accounting for seasonal variations. A Gaussian distribution dresses the 2821 sample values, where the parameters of the Gaussian distribution are estimated by the method of maximum likelihood.

An advantage of maximum likelihood over L moments for parameter estimation is its adaptability to changes in model. The maximum likelihood estimation (MLE) of nested models leads to a simple likelihood ratio test (LRT) procedure of any one nonstationary model where the mean and standard deviation vary with time, against a stationary model where the mean and standard deviation are constant. Because the method of L moments are defined for only identically distributed random samples, the use of time-dependent parameters in a Gaussian distribution prevent us from using the method of L moments.

First, a Shapiro-Wilks test tests the null hypothesis that the observed data is drawn from a Gaussian distribution at the $\alpha_{FDR} = 0.05$ level of significance. The False Discovery Rate (FDR) procedure is similar to Walker's criteria, but less strict. The collection of p -values from the $N_0 = 20805$ Shapiro-Wilks tests are arranged in ascending order. The null hypotheses are rejected if their respective p -values are lower than a threshold level p_{FDR} that depends on the distribution of ordered p -values (Wilks, 2016).

For the locations and calendar days where the Shapiro-Wilks test is tested during the 1980-2010 study period, more than 90% of the results suggest both maximum and minimum daily T2M can be described by a Gaussian distribution.

Second, the null hypothesis that all 31 years of daily maximum/minimum T2M are drawn from the same Gaussian distribution are tested against the alternate hypothesis that each of the 31 years of the daily temperature data is drawn from a different Gaussian distribution due to time dependencies in the parameters. This alternate hypothesis is tested against the null hypothesis using a LRT.

To perform the LRT, it is necessary to fit 31 Gaussian distributions separately to the each year of data, and compare these 31 distributions with the single Gaussian distribution fit using the full data set for each calendar day at each station. In principle, time dependence can be assumed on both parameters. However, if the changes in either of the parameters is negligible, it is advantageous to keep them constant.

The Gaussian distribution parameters are assumed to vary with time as $\mu(t) = \mu_0 + \mu_1(t - t_0)$, and $\sigma(t) = \sigma_0 + \sigma_1(t - t_0)$. The slope coefficients μ_1 and σ_1 represent the annual rate of change in mean and standard deviation of daily T2M respectively. The null hypotheses $\mu_1 = 0$ and $\sigma_1 = 0$ are tested against the alternate hypotheses that $\mu_1 \neq 0$ and $\sigma_1 \neq 0$.

With models $M_0 = N(\mu, \sigma)$ and $M_1 = N(\mu(t), \sigma(t))$, the LRT compares the difference between the maximized log-likelihoods under models M_0 and M_1 respectively. A sufficiently large difference indicates that the nonstationary model explains substantially more of the variation in the data than the stationary model. Small differences suggest that the increase in model size does not bring worthwhile improvements in the capacity of the model to explain the data. This critical difference is determined by the χ^2 distribution. Model M_0 is rejected at the $\alpha_{FDR} = 0.05$ level of significance.

2.5.2 Extreme T2M

For extreme maximum and minimum T2M, the rolling 31-day centered rolling window is maintained for the 31 year period. A GEV dresses these 31 sample values by the method of maximum likelihood. A nonstationary GEV distribution is compared, where only the location parameter is allowed to exhibit trend, with a stationary GEV distribution with constant location, scale and shape parameters.

The GEV distribution location parameter is assumed to vary with time as $\mu(t) = \mu_0 + \mu_1(t - t_0)$, where the slope coefficient μ_1 represent the annual rate of change in location values of extreme T2M.

With models $M_1 = GEV(\mu(t), \sigma, \kappa)$ and $M_0 = GEV(\mu, \sigma, \kappa)$, the alternate hypothesis is tested against the null hypothesis that the extremes of temperature are drawn from the same GEV distribution using a LRT at the $\alpha_{FDR} = 0.05$ level of significance. Namely, the alternate hypothesis is that a nonstationary model explains substantially more of the variation in the data, and consequently changes in return levels should be accounted for.

2.6 Results and discussion

2.6.1 Daily T2M

The Central climate zone is representative of the overall performance of the different reanalyses for daily maximum T2M across BC (Figure 2.3). The seasonal cycle of temperature is well captured by all reanalyses (Fig. 2.3). All four reanalyses show a negative (cold) systematic error (Fig. 2.3; Table 2.2)

throughout the entire calendar year. ERA-Interim and JRA-55 outperform CFSR and, to a lesser degree, MERRA-2. Additionally, MERRA-2 has a less consistency in performance, and usually has a higher systematic error during spring. MERRA-2 and CFSR also tend to have a higher (worse) random error during spring (Fig. 2.3). The Maritime climate zone has the lowest systematic across all reanalyses all year long, and the lowest season-to-season temperature variation (Figure not shown; Table 2.2). The North climate zone has the highest random error across all reanalyses during winter time, and the highest season-to-season temperature variation. This might be expected to some extent since higher observed variability typically makes for poorer model forecast random error, and the reanalyses are based on model forecasts.

Table 2.2: Seasonally averaged systematic error (SE in °C) and random error (RE in °C) of mean values, and Kolmogorov-Smirnov statistic (KS) of daily maximum T2M by climate zone

	Winter			Spring			Summer			Fall		
	SE	RE	KS	SE	RE	KS	SE	RE	KS	SE	RE	KS
North												
CFSR	-2.05	1.44	0.29	-5.59	1.12	0.53	-4.22	0.91	0.40	-3.86	0.86	0.43
ERA-Interim	-1.88	1.33	0.23	-3.47	0.69	0.37	-3.71	0.54	0.38	-3.21	0.79	0.36
JRA-55	-1.67	1.56	0.28	-3.80	0.78	0.39	-3.20	0.53	0.34	-3.01	0.77	0.35
MERRA-2	-3.32	1.52	0.30	-6.11	1.12	0.57	-4.24	0.75	0.42	-3.76	0.85	0.40
Central												
CFSR	-2.15	1.11	0.30	-4.80	0.99	0.50	-4.36	0.81	0.40	-3.68	0.73	0.40
ERA-Interim	-1.55	0.97	0.18	-2.67	0.56	0.30	-3.15	0.58	0.62	-2.67	0.64	0.30
JRA-55	-2.20	0.96	0.25	-2.75	0.52	0.30	-2.98	0.54	0.29	-2.88	0.64	0.31
MERRA-2	-3.61	1.12	0.37	-4.98	1.01	0.57	-2.84	0.64	0.42	-2.89	0.80	0.40
Maritime												
CFSR	-0.81	0.76	0.25	-1.40	0.79	0.39	-1.35	0.76	0.41	-1.02	0.65	0.35
ERA-Interim	-0.05	0.72	0.25	-0.05	0.74	0.32	-0.29	0.81	0.30	-0.27	0.67	0.31
JRA-55	-0.37	0.80	0.29	-0.38	0.88	0.33	-1.03	0.98	0.36	-0.62	0.79	0.34
MERRA-2	-0.56	0.73	0.24	-0.66	0.77	0.32	-0.45	0.81	0.30	-0.40	0.66	0.28
Southwest												
CFSR	-2.84	0.75	0.39	-3.23	0.90	0.39	-2.57	0.97	0.28	-2.74	0.79	0.34
ERA-Interim	-2.57	0.72	0.35	-2.28	0.67	0.31	-2.16	0.74	0.27	-2.54	0.68	0.32
JRA-55	-2.44	0.73	0.33	-2.13	0.76	0.29	-2.49	0.77	0.28	-2.53	0.77	0.32
MERRA-2	-3.15	0.78	0.44	-3.15	0.98	0.39	-1.94	0.75	0.24	-1.98	0.76	0.26
Southeast												
CFSR	-4.87	0.98	0.47	-7.83	1.19	0.67	-7.02	1.01	0.52	-5.91	0.86	0.51
ERA-Interim	-4.19	1.03	0.39	5.99	0.72	0.53	-6.28	0.71	0.47	-5.25	0.83	0.46
JRA-55	-4.63	0.94	0.43	-5.52	0.73	0.51	-5.54	0.67	0.43	-4.95	0.84	0.44
MERRA-2	-5.61	1.06	0.14	-7.51	1.10	0.65	-5.64	0.76	0.46	-4.58	1.04	0.42

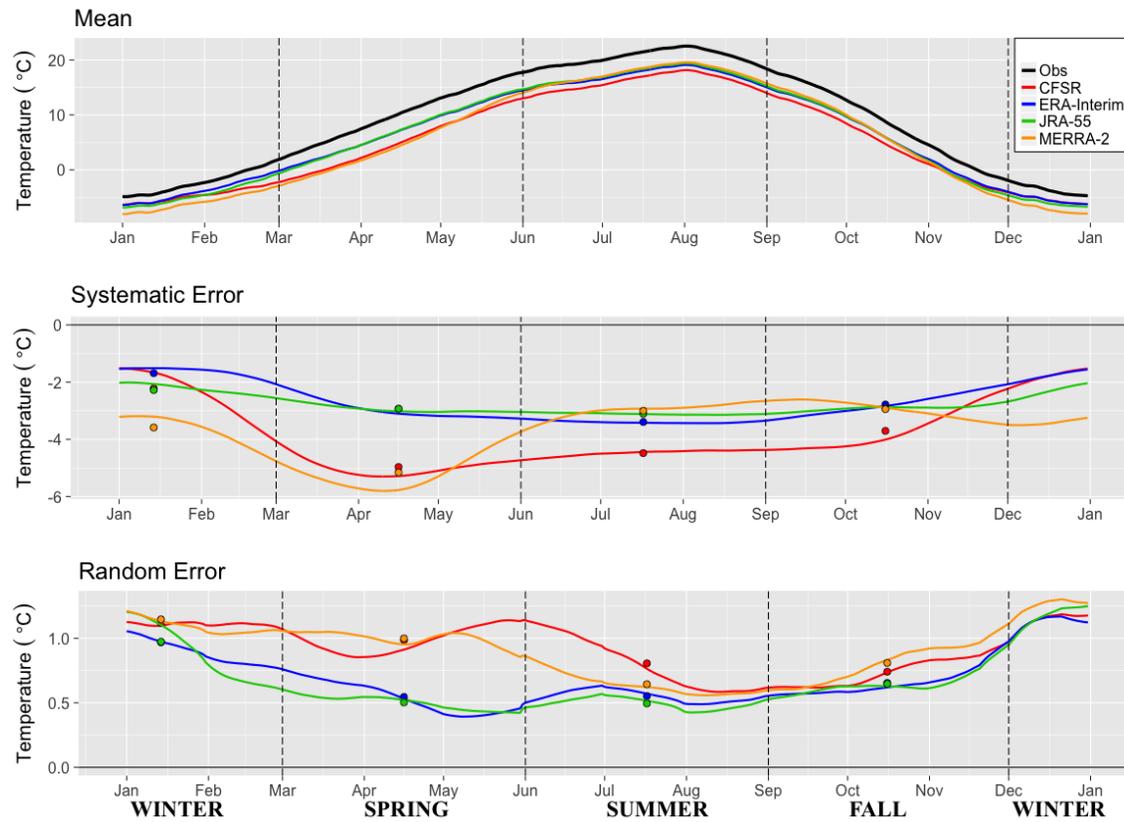


Figure 2.3: Observed and reanalysis mean, systematic error, and random error magnitude for daily maximum T2M, averaged over stations in the Central climate zone. The vertical dashed lines indicate the change in seasons and the colored dots represent the seasonal average.

Daily minimum T2M and their errors are illustrated in Figure 2.4 for the North climate zone, which is representative of other zones. Similar to daily maximum T2M, all four reanalyses capture season-to-season temperature variations regardless of the temperature differences between the different climate zones, but generally exhibit a cold bias (Figure not shown; Table 2.3). In contrast, the systematic error is much smaller for daily minimum than for daily maximum T2M across all zones. Again, ERA-Interim and JRA-55 outperform CFSR and MERRA-2 for systematic and random errors, both of which are higher during winter months. For the North, systematic error is positive (warm) during winter months for the ERA-Interim and JRA-55, and in the Maritime it is positive all year across all reanalyses (Figure not shown; Table 2.3).

Table 2.3: Seasonally averaged systematic error (SE in °C) and random error (RE in °C) of mean values, and Kolmogorov-Smirnov statistic (KS) of daily minimum T2M by climate zone

	Winter			Spring			Summer			Fall		
	SE	RE	KS	SE	RE	KS	SE	RE	KS	SE	RE	KS
North												
CFSR	0.48	1.78	0.20	-1.42	1.17	0.24	-1.51	0.66	0.28	-0.62	1.01	0.25
ERA-Interim	1.13	1.37	0.19	-0.32	0.88	0.17	-0.48	0.53	0.17	-0.58	0.86	0.21
JRA-55	1.93	1.71	0.25	-0.51	0.82	0.17	-1.02	0.53	0.26	-0.67	0.86	0.26
MERRA-2	-3.03	1.56	0.23	-3.89	1.01	0.22	-2.24	0.66	0.37	-3.35	1.09	0.35
Central												
CFSR	0.84	1.56	0.14	-0.77	0.99	0.17	-1.60	0.68	0.40	-0.18	0.90	0.15
ERA-Interim	0.98	1.24	0.14	-0.01	0.75	0.18	-0.17	0.61	0.62	-0.09	0.83	0.15
JRA-55	1.16	1.34	0.16	-0.24	0.74	0.16	-0.64	0.63	0.29	-0.37	0.83	0.15
MERRA-2	-3.54	1.49	0.24	-2.86	0.89	0.33	-1.59	0.71	0.29	-2.39	1.08	0.27
Maritime												
CFSR	2.40	0.87	0.35	2.10	0.66	0.37	1.11	0.63	0.29	2.08	0.80	0.37
ERA-Interim	2.71	0.88	0.33	2.74	0.68	0.40	1.96	0.31	0.38	2.52	0.83	0.37
JRA-55	2.79	0.93	0.39	2.89	0.70	0.47	1.72	0.67	0.37	2.67	0.83	0.43
MERRA-2	1.57	0.86	0.29	1.58	0.63	0.31	0.90	0.63	0.26	1.58	0.80	0.32
Southwest												
CFSR	-0.65	0.79	0.19	-0.94	0.68	0.26	-1.79	0.69	0.27	-0.82	0.70	0.22
ERA-Interim	-1.08	0.84	0.21	-0.40	0.59	0.22	-0.19	0.59	0.20	-0.66	0.65	0.21
JRA-55	-0.14	0.74	0.22	0.23	0.59	0.26	-0.23	0.62	0.19	-0.05	0.69	0.26
MERRA-2	-3.08	0.91	0.22	-1.94	0.69	0.34	-1.00	0.61	0.29	-1.58	0.75	0.28
Southeast												
CFSR	-3.43	1.41	0.24	-3.89	1.16	0.41	-4.22	0.80	0.54	-3.03	1.08	0.33
ERA-Interim	-2.94	1.19	0.24	-2.62	0.83	0.32	-2.41	0.74	0.37	-2.57	1.05	0.29
JRA-55	-2.22	1.10	0.25	-2.52	0.81	0.33	-2.32	0.81	0.35	-2.18	0.97	0.28
MERRA-2	-7.83	1.30	0.55	-7.83	1.13	0.60	-3.84	0.79	0.51	-4.69	1.36	0.46

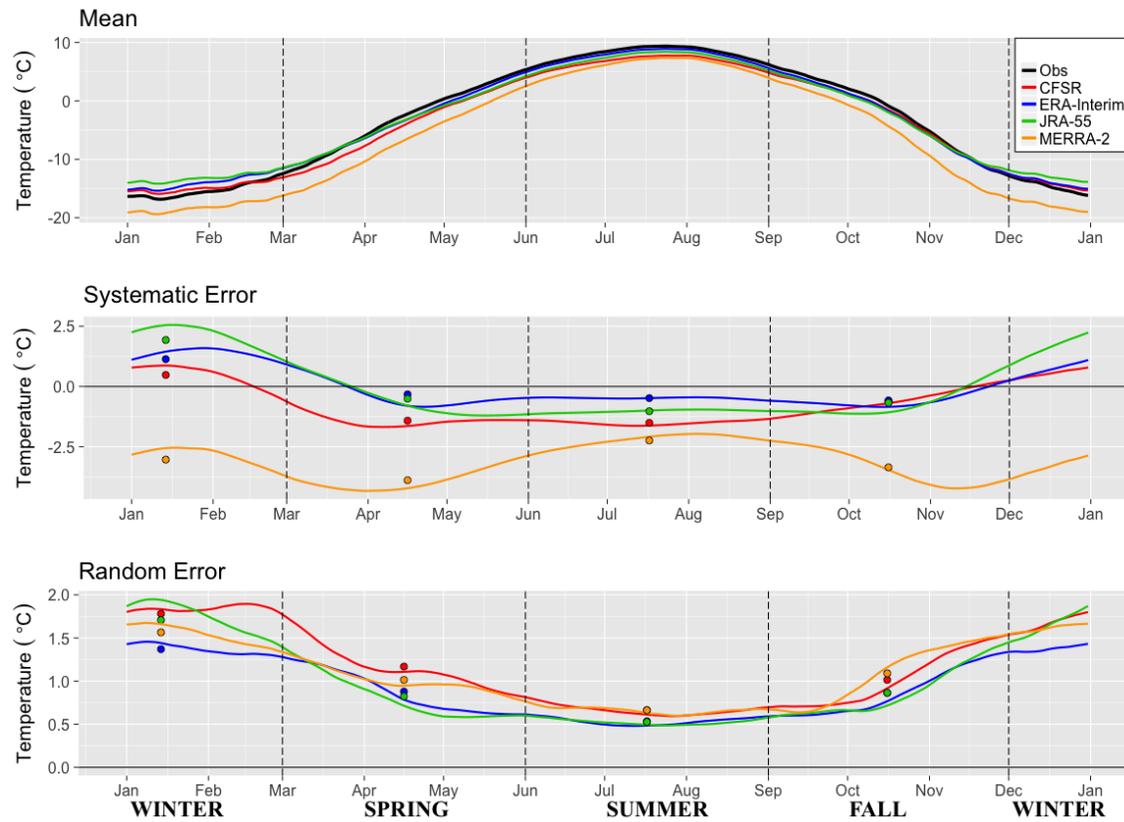


Figure 2.4: Observed and reanalysis mean, systematic error, and random error magnitude for daily minimum T2M, averaged over stations in the North climate zone. The vertical dashed lines indicate the change in seasons and the colored dots represent the seasonal average.

Finally, a KS statistic is used to determine how well the reanalyses captures the ECDF of observed daily maximum T2M across the Southeast climate zone (Fig. 2.5). Values closer to zero are better, which would indicate no difference between the observation ECDF and that of the reanalysis. ERA-Interim and JRA-55 generally outperform CFSR and MERRA-2. MERRA-2 once again exhibits high variability in performance throughout the seasons, followed closely by CFSR (Table 2.2). MERRA-2 performs best in the Maritime climate zone. For daily minimum T2M the disparity between reanalyses is more pronounced, with Era-Interim and JRA-55 consistently outperforming CFSR and MERRA-2 (Table 2.3). Furthermore, daily minimum T2M is consistently better captured by the reanalyses than maximum T2M.

A reality of the relatively sparse surface station network in the complex terrain of BC is that most stations are located in valleys, and higher-elevation stations are lacking (Fig. 2.6). This is especially true for stations with quality, long records. With this caveat in mind, the effects of reanalysis terrain error (reanalysis terrain elevation minus real-world station elevation) are examined. Due to the relatively coarse resolution of the reanalyses, they tend to reduce terrain amplitude, exhibiting negative terrain error over mountain ridges, and positive errors over valleys. Reanalysis daily T2M systematic error is strongly related to reanalysis terrain error, but not climate zone (Fig. 2.7). Namely, larger positive terrain errors are associated with larger negative temperature biases. That is, typically valleys exhibit a cold bias in reanalyses, and a warm bias over mountain ridges.

The regression line is steeper for CFSR and MERRA-2, indicating that ERA-Interim and JRA-55 suffer from terrain-related biases to a lesser extent. Additionally, minimum T2M have less terrain-related systematic error than maximum T2M. The magnitudes of the random errors are smaller and uncorrelated with terrain error across all reanalyses.

2.6. Results and discussion

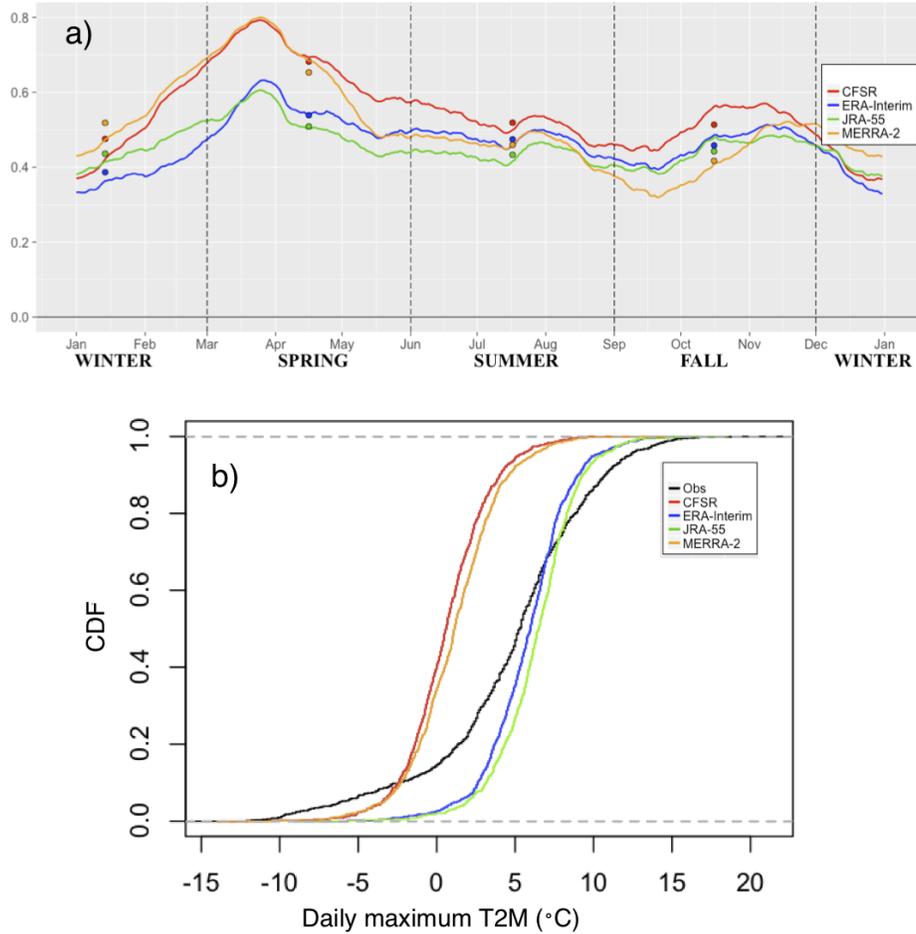


Figure 2.5: a) Kolmogorov-Smirnov (KS) statistic results for daily maximum T2M, for all four reanalyses by time of year, averaged across all locations in the Southeast climate zone. The vertical dashed lines indicate the change in seasons and the colored dots represent the seasonal average. KS statistic values closer to zero are better. b) As an example, KS-statistic results for daily maximum T2M on March 1 at Cranbrook (YXC) are shown.

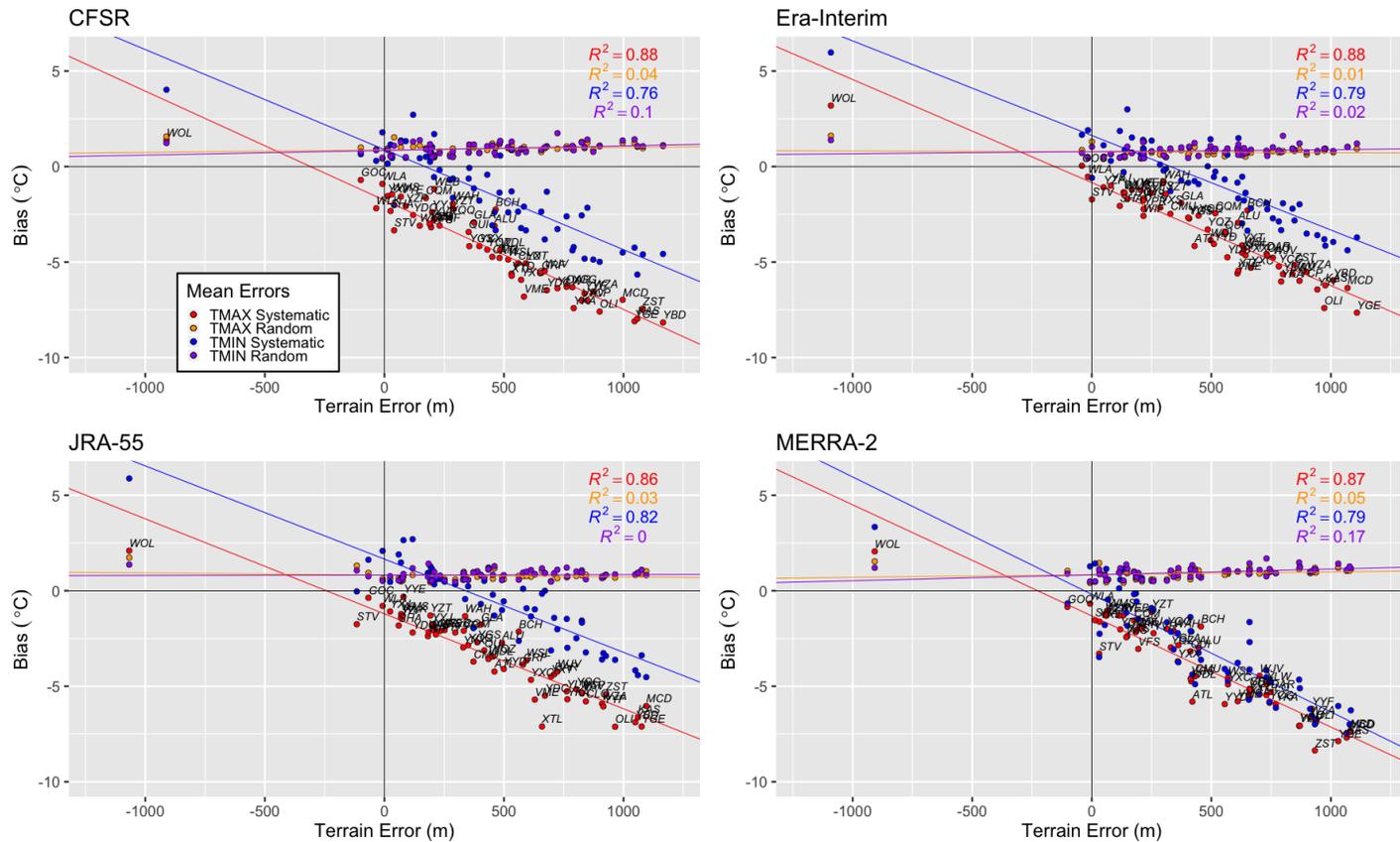


Figure 2.7: Mean systematic and mean random error of daily maximum (TMAX) and minimum (TMIN) T2M for each of the 57 stations as a function of terrain reanalysis error. The solid lines show the linear regression fits. The nearly horizontal lines show the random error; the two lines are nearly on top of each other.

2.6.2 Extreme T2M

The Southeast climate zone has the highest maximum T2M 30-year return levels, and the performance of the reanalyses here is representative of the overall performance across BC (Fig. 2.8; Table 2.4). All four reanalyses show negative (cold) systematic error (Fig. 2.8) throughout the entire calendar year. ERA-Interim and JRA-55 have a consistent systematic error and outperform MERRA-2 and CFSR. MERRA-2 once again has a larger variation in performance. CFSR also generally performs worse but has a more consistent bias. The Maritime climate zone has the lowest 30-year return levels, with values below 30°C during summer (Figure not shown).

Table 2.4: Seasonally averaged systematic error of 2- and 30-year return levels of daily maximum T2M by climate zone (SE2 and SE30 respectively in °C)

	Winter		Spring		Summer		Fall	
	SE2	SE30	SE2	SE30	SE2	SE30	SE2	SE30
North								
CFSR	-4.48	-7.66	-6.76	-6.28	-3.66	-3.30	-5.31	-5.51
ERA-Interim	-2.89	-4.38	-3.94	-3.68	-3.65	-3.60	-3.99	-4.46
JRA-55	-3.43	-4.83	-4.10	-3.69	-3.03	-2.72	-3.60	-3.98
MERRA-2	-4.90	-8.35	-8.25	-8.66	-5.05	-4.68	-4.93	-5.72
Central								
CFSR	-5.21	-6.20	-5.82	-5.08	-4.01	-3.84	-4.49	-5.11
ERA-Interim	-2.76	-3.64	-2.97	-2.61	-2.94	-3.22	-2.90	-3.66
JRA-55	-3.68	-4.65	-3.29	-3.27	-2.93	-3.23	-3.37	-4.00
MERRA-2	-6.16	-8.20	-6.60	-6.40	-3.29	-3.15	-3.58	-3.78
Maritime								
CFSR	-2.30	4.72	-2.21	-3.72	-1.85	-4.06	-2.27	-4.01
ERA-Interim	-1.39	-3.83	-1.18	-1.95	-1.38	-2.66	-1.74	-3.16
JRA-55	-1.90	-4.42	-1.71	-2.53	-2.19	-4.00	-2.32	-4.02
MERRA-2	-1.95	-3.95	-1.69	-3.05	-1.28	-3.01	-1.52	-3.07
Southwest								
CFSR	-3.78	-4.69	-3.59	-3.79	-2.69	-3.52	-3.05	-3.71
ERA-Interim	-3.31	-4.12	-2.62	-2.34	-2.12	-2.21	-2.75	-2.98
JRA-55	-3.24	-4.55	-2.55	-2.73	-2.92	-3.53	-3.35	-3.84
MERRA-2	-4.60	-4.84	-4.48	-4.09	-2.33	-2.17	-1.90	-1.79
Southeast								
CFSR	-6.61	-8.73	-8.35	-6.84	-6.81	-6.81	-6.09	-5.76
ERA-Interim	-4.60	-5.36	-5.72	-5.05	-6.08	-5.98	-4.99	-4.74
JRA-55	-5.49	-6.40	-5.80	-5.48	-5.65	-5.81	-5.35	-5.38
MERRA-2	-7.18	-8.98	-8.57	-7.17	-6.22	-5.90	-4.43	-3.72

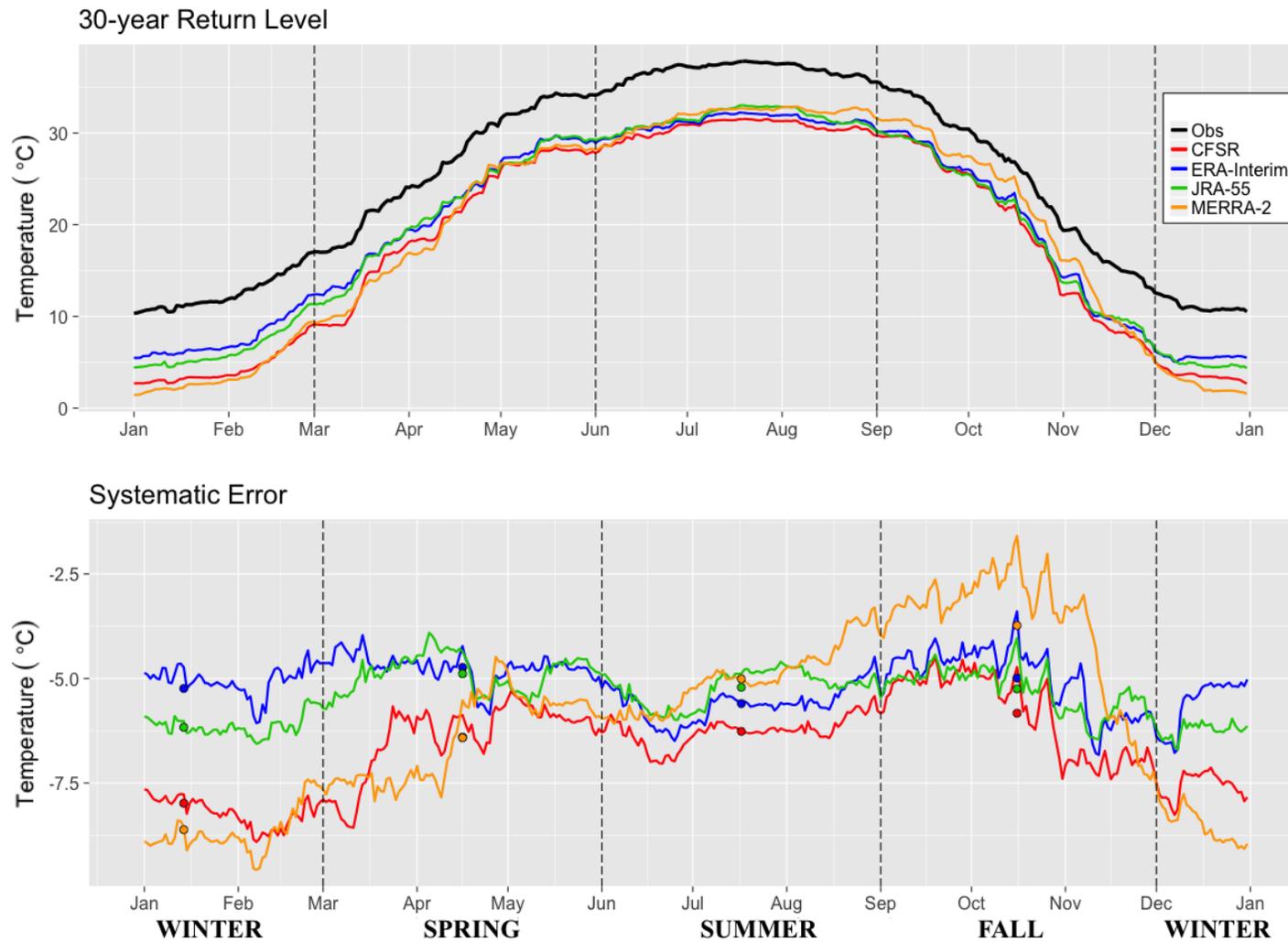


Figure 2.8: 30-year return level and systematic error for maximum T2M, for all four reanalyses for Southeast climate zone. The vertical dashed lines indicate the change in seasons and the colored dots represent the seasonal average.

2.6. Results and discussion

For minimum T2M 30-year return levels, results are mixed. The JRA-55 and ERA-Interim show generally positive systematic errors (larger during the winter) in the North and Central zones, while MERRA-2 and CFSR show mostly negative systematic errors (Fig. 2.9; Table 2.5). This indicates a warm bias in extreme minimum T2M for the JRA-55 and ERA-Interim, and a cold bias for the MERRA-2 and CFSR. In contrast, the Southeast and Southwest climate zones generally show negative systematic error in extreme minimum T2M in all four reanalyses, indicating they are typically too cold (Figure not shown; Table 2.5). The Maritime climate zone has the lowest season-to-season extreme T2M variations (Figure not shown).

Table 2.5: Seasonally averaged systematic error of 2- and 30-year return levels of daily minimum T2M by climate zone (SE2 and SE30 respectively in °C)

	Winter		Spring		Summer		Fall	
	SE2	SE30	SE2	SE30	SE2	SE30	SE2	SE30
North								
CFSR	-1.02	-3.88	-1.75	-4.42	-1.14	-1.80	-0.16	-2.69
ERA-Interim	2.20	1.73	0.95	1.76	0.11	0.04	-0.04	0.11
JRA-55	3.93	4.95	1.05	1.49	-0.59	-0.63	0.11	0.35
MERRA-2	-3.72	-4.74	-3.25	-3.97	-1.11	-1.26	-3.41	-4.50
Central								
CFSR	0.24	-3.04	-1.19	-2.76	-1.34	-1.66	0.57	-0.13
ERA-Interim	2.56	3.74	0.73	1.30	0.73	1.17	0.63	1.37
JRA-55	4.16	5.99	0.98	2.18	0.17	0.35	0.67	1.23
MERRA-2	-3.02	-2.83	-2.06	-2.54	0.02	0.31	-1.51	-1.24
Maritime								
CFSR	3.09	1.90	3.44	3.12	2.38	3.07	3.55	3.41
ERA-Interim	3.14	2.31	3.83	3.94	3.33	3.90	3.60	3.83
JRA-55	3.83	2.74	4.51	4.51	3.43	3.99	4.46	4.17
MERRA-2	1.91	0.87	3.15	3.08	2.64	3.39	3.19	2.99
Southwest								
CFSR	-1.25	-2.86	-1.32	-2.71	-2.32	-2.58	-0.82	-1.89
ERA-Interim	-2.55	-3.84	-0.98	-2.36	-0.08	0.12	-1.09	-2.42
JRA-55	0.15	-0.67	0.74	0.64	0.31	0.69	0.73	-0.03
MERRA-2	-5.32	-6.64	-2.10	-3.30	-0.38	0.01	-1.66	-3.12
Southeast								
CFSR	-6.91	-8.54	-5.11	-8.83	-4.37	-4.85	-3.41	-6.00
ERA-Interim	-3.69	-2.83	-2.77	-3.28	-2.35	-2.16	-2.89	-3.53
JRA-55	-1.16	-0.27	-2.09	-2.32	-2.00	-1.82	-1.88	-2.40
MERRA-2	-8.91	-10.19	-5.78	-7.64	-3.18	-2.84	-4.49	-5.81

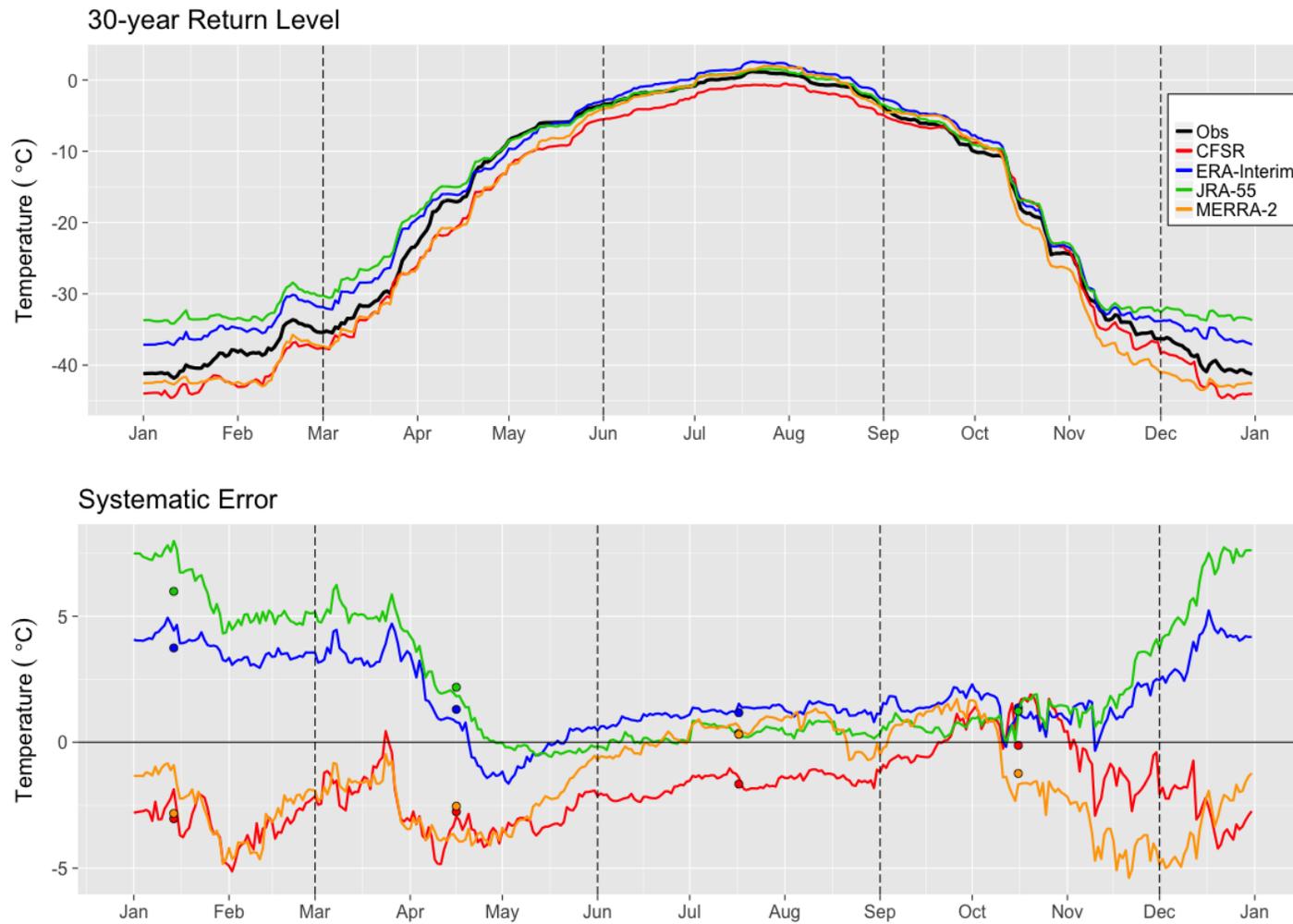


Figure 2.9: 30-year return level and systematic error for minimum T2M, for all four reanalyses for Central climate zone. The vertical dashed lines indicate the change in seasons and the colored dots represent the seasonal average.

ERA-Interim and JRA-55 extreme minimum T2M systematic errors are smaller than CFSR and MERRA-2, which have larger errors during winter months. MERRA-2 once again exhibits the largest variation in performance throughout the calendar year and across different climate zones.

Similar to daily T2M, the systematic errors of 30- and 2-yr return level T2M are related to reanalysis terrain elevation error (Fig. 2.10). Namely, the larger the positive reanalysis terrain elevation error relative to the station's real world elevation, the larger the negative systematic error of the reanalysis temperature. Furthermore, the regression lines are vertically shifted in CFSR and MERRA-2, indicating larger errors across all elevations despite similar horizontal grid spacing (which should give them similar abilities to resolve terrain).

An ANOVA indicates differences in mean systematic error of daily and extreme T2M between the reanalyses at the 1% significance level. After multiple comparisons by Tukey's HSD, daily and extreme minimum T2M are shown to be significantly better captured than daily and extreme maximum T2M. ERA-Interim and JRA-55 consistently significantly outperform CFSR and MERRA-2 for both daily and extreme T2M. If the four reanalyses are separated into two groups — the more accurate JRA-55 and ERA-Interim, and the less accurate CFSR and MERRA-2 — the difference in mean systematic error between groups is statistically significant. Within groups, the JRA-55 (CFSR) mean systematic error is smaller than the ERA-Interim (MERRA-2) mean systematic error, but this difference is not statistically significant.

Finally, Figure 2.11 shows the mean absolute error (MAE) of the reanalyses daily and extreme (2- and 30-year return levels) T2M averaged over the entire study period and all stations. The smaller the enclosed area for a given reanalysis, the better its performance. Daily minimum and maximum T2M have smaller MAE than extreme minimum and maximum T2M. Daily and extreme minimum T2M are better captured by the reanalyses than daily and extreme maximum T2M. Additionally, the errors of 2-year and 30-year return levels are of similar magnitude. Results are similar for non-assimilated stations. ERA-Interim and JRA-55 outperform CFSR and to a lesser degree MERRA-2 (not shown).

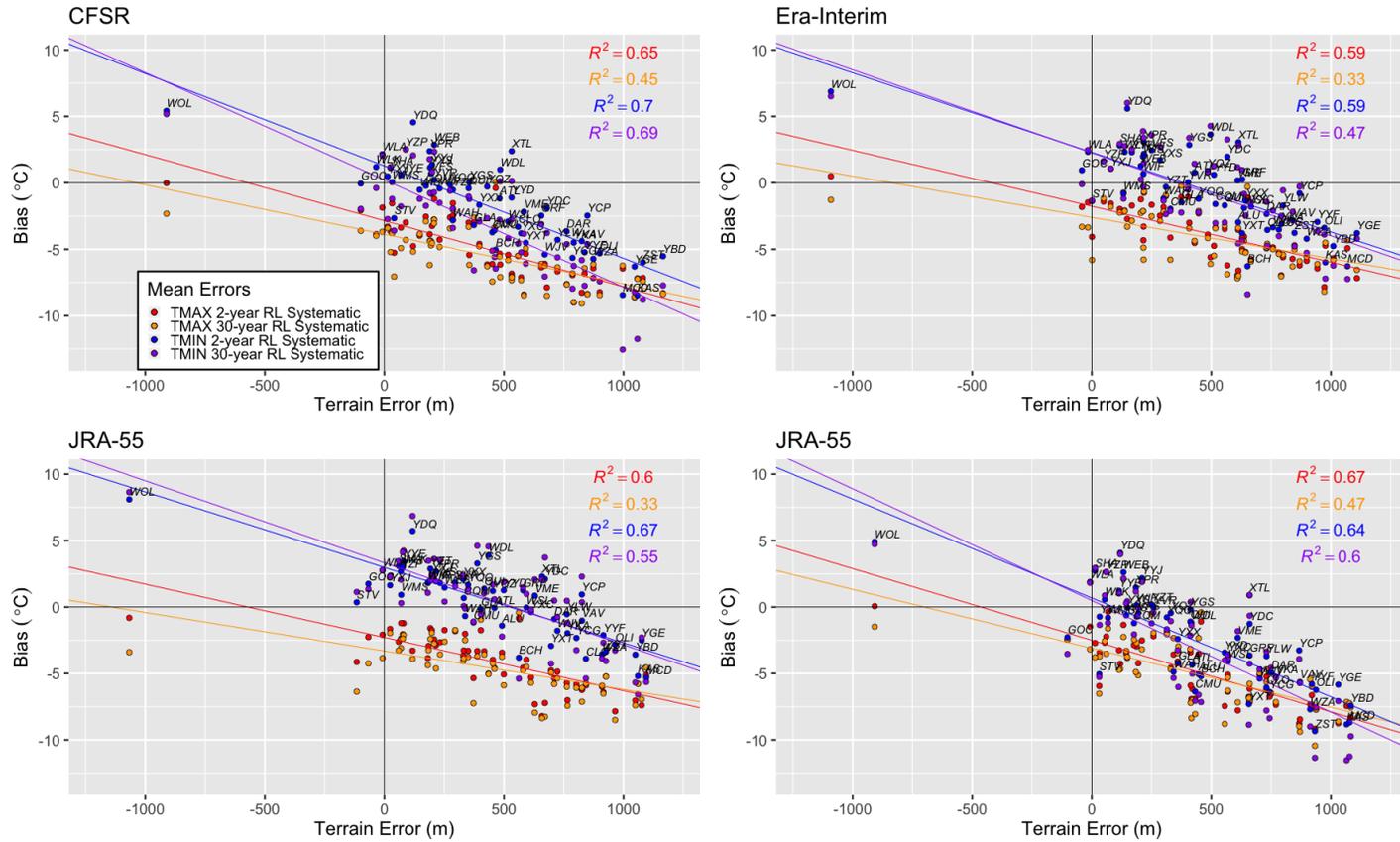


Figure 2.10: Mean systematic error of 2- and 30-year return levels (RL) of extreme maximum (TMAX) and minimum (TMIN) T2M for each of the 57 stations as a function of terrain reanalysis error. The solid lines show the linear regression fits.

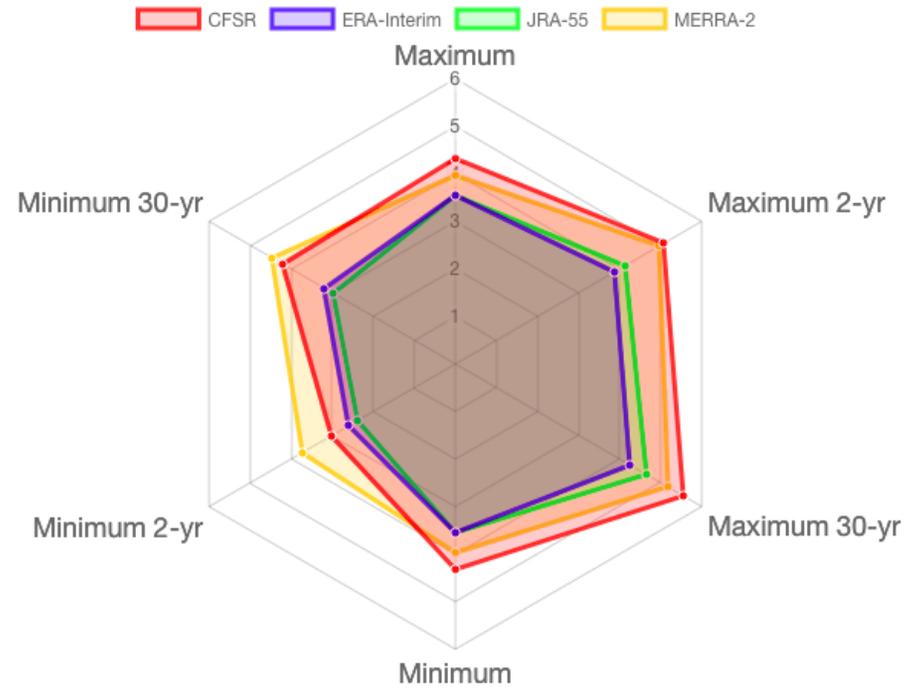


Figure 2.11: MAE ($^{\circ}\text{C}$) of daily maximum and minimum T2M, and of extreme (2- and 30-year return levels) maximum and minimum T2M. MAE is averaged across all 57 stations. Smaller values (closer to the origin) are better.

2.7 Stationarity

In selecting a nonstationary model for the Gaussian distribution, a decision must be made about whether both the mean and standard deviation, or only one of the parameters should be treated as varying in time.

Variations through time of mean and standard deviation values for each calendar day at each station are modelled as a linear trend for daily maximum and minimum T2M. The null hypothesis that there is no trend, which indicates the parameters do not vary with time, is tested against the alternate hypotheses that the parameters are time-dependent at the $\alpha_{FDR} = 0.05$ significance level.

Figure 2.12a indicates a warming trend of mean values of daily minimum T2M across BC during summer, fall and winter, and a weak cooling trend during spring across North and Central climate zones. Furthermore, Figure 2.12b suggests a weak decreasing trend in standard deviation of daily minimum T2M for fall and winter season. However, none of trends are statistically significant.

Under models $M_0 = N(\mu, \sigma)$ and $M_1 = N(\mu(t), \sigma)$, however, the critical difference for comparing these two models is statistically significant at the $\alpha_{FDR} = 0.05$ significance level, mostly near urban areas in the Southwest during summer (Fig . 2.12c). This implies that a linear trend component explains a substantial amount of the variation in the data, and is likely to be a genuine effect in the warming trend process rather than a chance feature in the observed data.

Figure 2.13a suggests a warming trend of mean values of daily maximum T2M across the Southeast and Southwest climate zones during summer, fall and winter. Figure 2.13b shows a decreasing trend in standard deviation of daily maximum T2M during fall and winter across BC, and increasing trend in the spring.

None of the trends in the daily maximum T2M parameters are statistically significant at the $\alpha_{FDR} = 0.05$ significance level, and a stationary Gaussian distribution with model $M_0 = N(\mu, \sigma)$ is accurate enough to represent daily maximum T2M during the study period with few exceptions at the Southeast climate zone (Fig. 2.13c).

Analogously, Fig. 2.14a shows strong warming of extreme maximum T2M during summer, and strong warming of extreme minimum T2M during winter (Fig. 2.14c). The LRT deviance statistic suggests the evidence supporting such a trend is weak, implying no significant improvement over the stationary model at the $\alpha_{FDR} = 0.05$ significance level (Fig. 2.14b,d). During the 31 years of the study period, changes in return levels of extreme

2.7. Stationarity

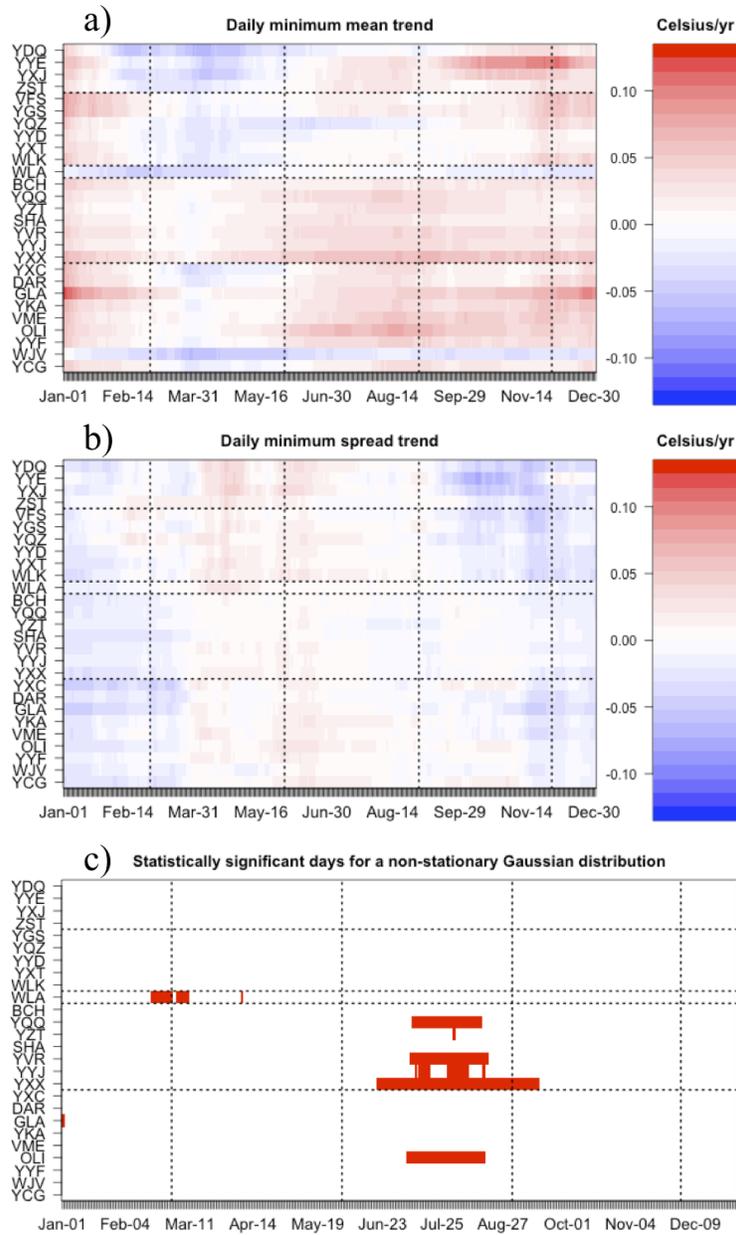


Figure 2.12: (a) Mean linear trend of daily minimum T2M; (b) Standard deviation linear trend of daily minimum T2M; (c) Days when $N(\mu(t), \sigma)$ is statistically significant. The vertical dashed lines indicate the change in seasons and the horizontal dashed lines delineate from top to bottom the North, Central, Maritime, Southwest and Southeast climate zones respectively.

2.7. Stationarity

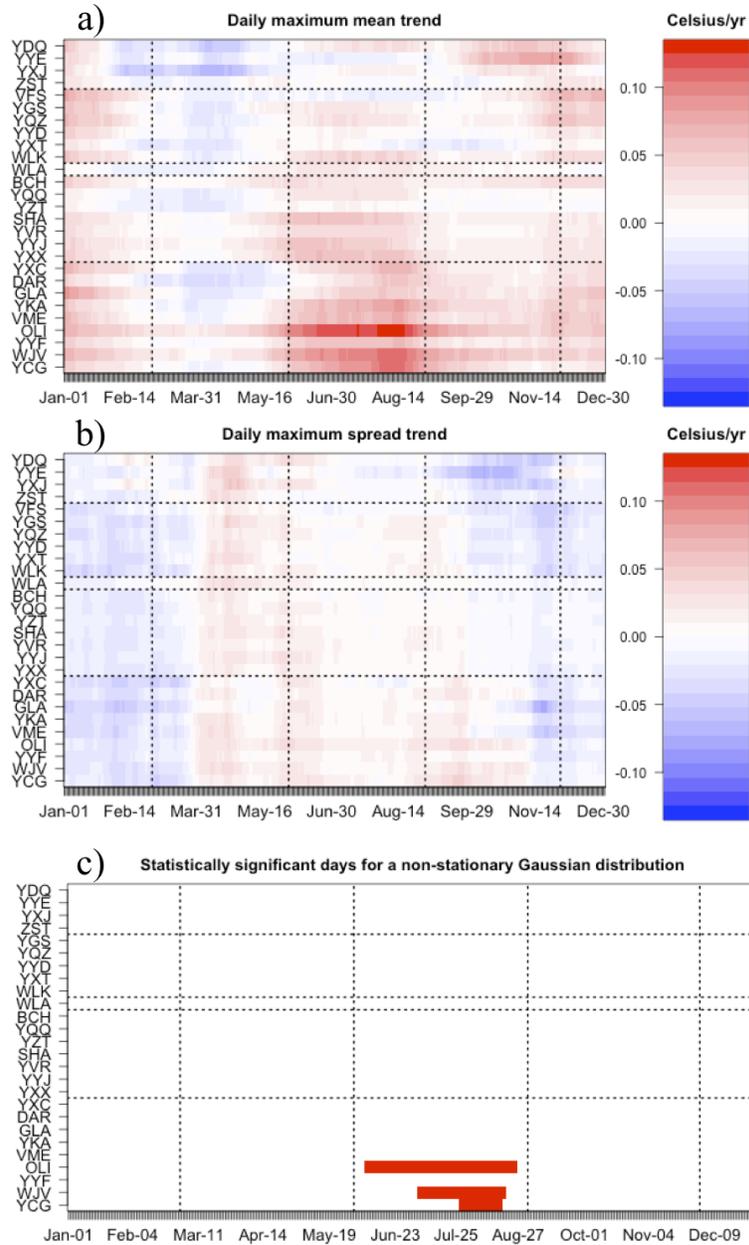


Figure 2.13: (a) Mean linear trend of daily maximum T2M; (b) Standard deviation linear trend of daily maximum T2M; (c) Days when $N(\mu(t), \sigma)$ is statistically significant. The vertical dashed lines indicate the change in seasons and the horizontal dashed lines delineate from top to bottom the North, Central, Maritime, Southwest and Southeast climate zones respectively.

2.7. Stationarity

T2M are generally too weak for nonstationarity to be required in order to characterize extreme levels.

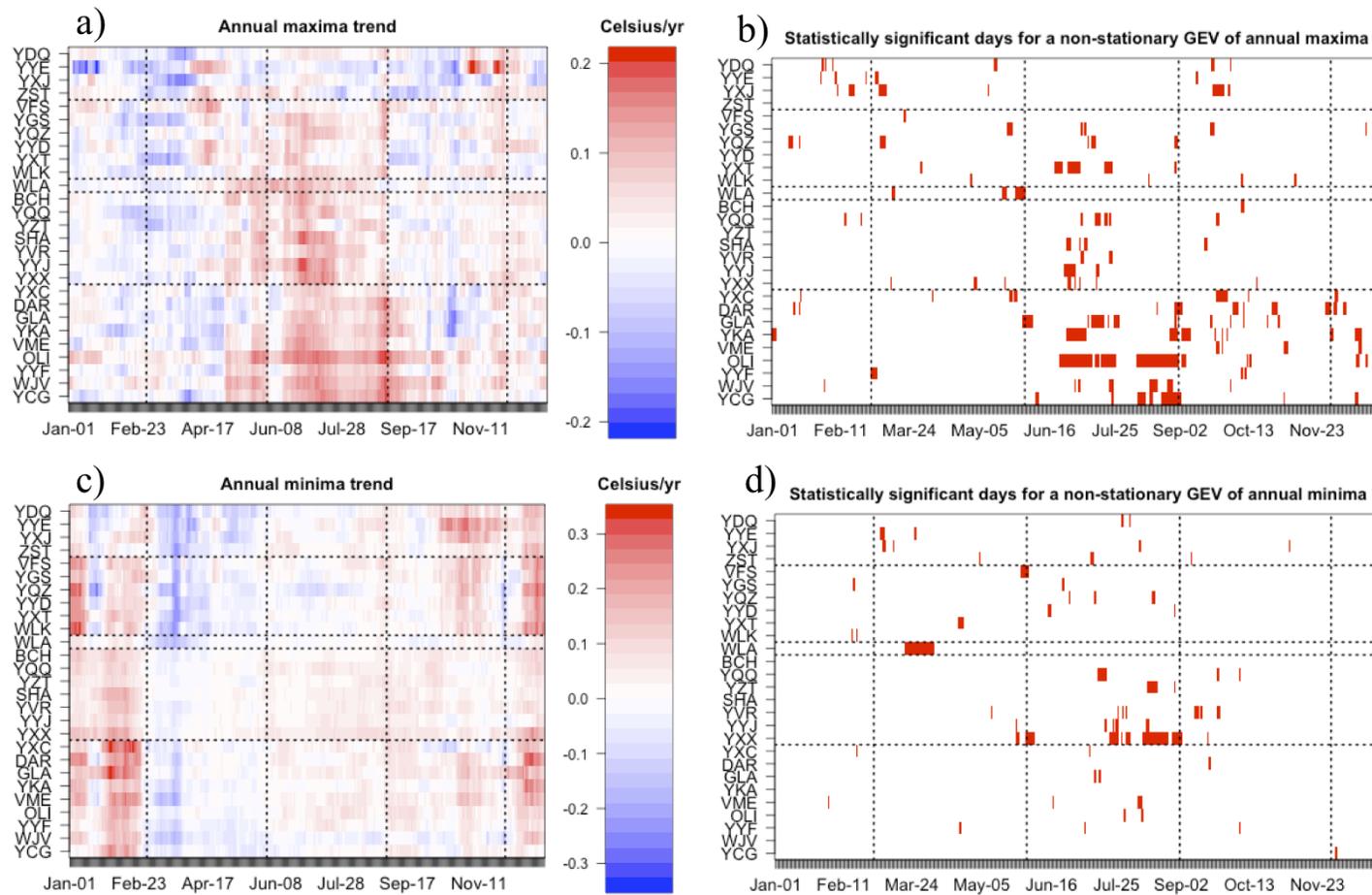


Figure 2.14: (a) Location linear trend of extreme maximum T2M; (b) Days when $GEV(\mu(t), \sigma, \kappa)$ for extreme maximum T2M is statistically significant; (c) Location linear trend of extreme minimum T2M; (d) Days when $GEV(\mu(t), \sigma, \kappa)$ for extreme minimum T2M is statistically significant. The vertical dashed lines indicate the change in seasons and the horizontal dashed lines delineate from top to bottom the North, Central, Maritime, Southwest and Southeast climate zones respectively.

2.8 Conclusion

The skill of reanalysis daily and extreme T2M were investigated across British Columbia (BC) during the 1980-2010 study period. To identify the best reanalyses among CFSR, ERA-Interim, JRA-55 and MERRA-2, the systematic error, random error and KS statistic were used to compare daily maximum and minimum T2M. To evaluate extreme maximum and minimum T2M, systematic error of 2- and 30-year return levels were compared.

ERA-Interim and JRA-55 had the lowest systematic and random error, and more consistent lower errors throughout the year and across all climate zones in BC. One possible explanation besides T2M being assimilated by these reanalyses is differences in the data assimilation methods. In regions where data is sparse, a reanalysis solution is more dependent on the model structure than on observations (Lindsay et al., 2014). ERA-Interim and JRA-55 use 4D-Var while CFSR and MERRA-2 use 3D-Var. 4D-Var better interprets physical information implicit in the meteorological model equations where observations are sparse, and extracts and transfers information into data-void areas in a more consistent way (Thépaut, 2006; Whitaker et al., 2009; Dee et al., 2011). Both issues are relevant for BC as weather station coverage is sparse outside Southwestern BC, and the Pacific Ocean to the west is a well-known data void (Hacker et al., 2003; Spagnol, 2005). Furthermore, the previous generation JRA-25, which has a 3D-Var assimilation method, shows a larger variation in performance in T2M throughout the calendar year (Lindsay et al., 2014) when compared to the results of JRA-55 shown here.

Minimum temperatures are consistently better captured by the reanalyses than maximum temperatures, whether it is daily or extreme T2M. This consistent difference in skill between maximum and minimum temperatures suggest they should be treated separately, rather than assessing hot or cold extreme weather events by looking at different tails of a common distribution of mean T2M (Graham et al., 2013; Hart and Grumm, 2001).

Furthermore, reanalysis daily and extreme maximum T2M generally have a cold bias, as do daily and extreme minimum T2M. Previous studies show a cold bias in T2M in CFSR and ERA-Interim at different pressure levels (Bao and Zhang, 2013), and in MERRA-1, JRA-55 and CFSR over Antarctica (Jones et al., 2016).

ERA-Interim extends as far back as 1979, whereas JRA-55 extends back to 1958, giving the latter a longer record and a larger sample size. A longer historical record is desirable, as a natural difficulty of extreme weather statistics is the limited amount of extreme data. Additionally, the standard errors

2.8. Conclusion

of the estimated parameters and of the return levels are expected to decrease as the sample size increases, whether they were obtained by MLE or method of L moments (Hosking et al., 1985; Hosking, 1990; Cai and Hames, 2010).

There is a noticeable warming trend in the Southwest and Southeast climate zones of BC during summer months for daily maximum T2M; and during summer months for extreme maximum T2M; during summer, fall and winter months for daily minimum T2M; and during winter for extreme minimum T2M. Vincent and Mekis (2006) showed that the number of cold events has significantly decreased while the number of warm events has significantly increased over BC. This warming trend could in part be caused by urbanization surrounding the weather stations (Gunn, 2010; Jones et al., 1985). The population of BC has grown faster than the Canadian national rate, with Southwest BC leading the growth rate (Demography Division, 2016). Despite some evidence of warming and cooling trends, most were not statistically significant. Thus, for this 31-year study period, temporally varying distributions are not needed to represent T2M over BC. It is possible that, were a longer period of record used, more widespread significant trends might be found.

While random error and KS statistic cannot be bias corrected or easily improved otherwise, this study found that systematic errors in daily and extreme T2M are largely explained by the reanalysis terrain elevation error, and thus could be largely corrected/eliminated. Systematic T2M errors are a larger component of the total error than random errors, thus a bias-corrected reanalysis T2M would be a substantially improved dataset, particularly over complex terrain. The methodology presented here should be able to be used in other parts of the world.

Such a correction will be discussed in Chapter 4. First, Chapter 3 will evaluate daily and extreme accumulated precipitation produced by these reanalyses over BC.

Chapter 3

Performance of Reanalyses across British Columbia. Part II: Evaluation of Daily and Extreme Precipitation

3.1 Introduction

In Chapter 2 performance of latest-generation reanalyses with respect to daily and extreme maximum and minimum 2-m temperature (T2M) over mountainous BC is discussed. In this chapter performance with respect to daily and extreme precipitation is assessed (hereafter daily and extreme PCP; defined in section 3.2), and trends in both daily and extreme PCP during the study period are examined to determine if significant statistical changes occurred over the timespan of the dataset.

An added difficulty is that verification of precipitation extremes is more challenging than that of extreme temperatures (Bhend and Whetton, 2013; van Oldenborgh et al., 2013). Because of the small spatial and time scales of precipitation, in general, numerical models do not simulate precipitation as well as they do temperature (Kendon et al., 2014; Ravishankar et al., 2016), and have limited ability to faithfully represent extreme precipitation events (Zhu et al., 2014). For instance, differences in the reanalyses parameterizations of convection and other physical processes can impact how well extreme precipitation events are represented (Dee et al., 2011; Lindsay et al., 2014).

Previous studies have evaluated reanalysis performance with respect to precipitation and moisture fields. Flux tower observations over the Northern Hemisphere of temperature, wind speed, precipitation, downward shortwave radiation, net surface radiation, and latent and sensible heat fluxes were used to evaluate the performance of CFSR, ERA-Interim, ERA-40, and MERRA, where ERA-40 was found to have the lowest precipitation bias and ERA-

3.2. Data and Methodology

Interim best captured precipitation variability (Decker et al., 2012). Hodges et al. (2011) explored how well JRA-25, ERA-Interim, MERRA, and CFSR identify extratropical cyclones over the Northern and Southern Hemispheres, and found that the latest-generation reanalyses better represent cyclones, especially in the Southern Hemisphere. Berg et al. (2003) found that ERA-40 has positive biases in precipitation over land in North America, and Ruiz-Barradas and Nigam (2005) found that ERA-40 also has positive biases in evapotranspiration during the warm season over the U.S. Great Plains. Finally, Bosilovich et al. (2015) found that MERRA-2 has a more consistent global precipitation average than MERRA, and a lower global precipitation bias than JRA-55 and CFSR when compared to the Global Precipitation Climatology Project (GPCP; Adler et al. (2003)).

Precipitation is examined here because of its various financial, societal, and environmental impacts such as hydroelectric power generation (Odon et al., 2017), flooding and water management (White et al., 2016; Odon et al., 2017; Sun et al., 2018b), agriculture (Rosenzweig et al., 2001; Sun et al., 2018b), tourism (Patz et al., 2005; White et al., 2016), health (Curriero et al., 2001; Patz et al., 2005), and flora and fauna (Parmesan et al., 2000).

Furthermore, several studies have noted increases in the frequency and intensity of extreme precipitation events in various parts of the world (Mann and Emanuel, 2006; Krishnamurthy et al., 2009; Donat et al., 2013; Westra et al., 2013; Ravishankar et al., 2016). An increase in extreme precipitation may lead to other impacts such as increase in winter runoff, which in turn may lead to flooding, and challenges in drainage and sewage systems capacities (White et al., 2016; Sun et al., 2018b).

In section 3.2, a brief description of the different reanalyses and of the weather station observations is given. In sections 3.3 and 3.4, we describe the methodology for dividing BC into climate zones, and the various metrics used for evaluating daily and extreme reanalysis PCP. In section 3.5, daily and extreme PCP from the reanalyses are evaluated. In section 3.6, the methods for assessing statistical nonstationarity are introduced, and trends of both daily and extreme PCP are examined. Results are summarized in the conclusion.

3.2 Data and Methodology

Daily accumulated precipitation from 66 weather stations from 1 Jan 1980 to 31 Dec 2010 are used in this study to evaluate the CFSR, ERA-Interim, JRA-55 and MERRA-2 reanalyses. The 1980-2010 study period is chosen

3.2. *Data and Methodology*

because it is the longest overlap between the four reanalyses. MERRA-2 began in 1980 (Gelaro, 2015; Gelaro et al., 2017) and CFSR ended in 31 Dec 2010. From January 1, 2011 forward, the CFSR was extended using NCEP Climate Forecast System Version 2 (CFSv2) operational model. Differences between the model used to produce the CFSR and the operational CFSv2 may affect data evaluation past the extension date (Saha et al., 2014).

A description of the different reanalyses and of the weather station dataset is given below. A summary of the reanalyses atmospheric models and configurations are presented in Table 3.1. A broader description and comparison of the latest and previous generation reanalyses can be found in Chapter 2.

Table 3.1: Overview of the four reanalysis datasets examined in this study.

Institution	Reanalysis Model		Assimilation method	Period	Download grid ($lat \times lon$)	Time interval	Reference
NCEP/NCAR	CFSR	CFS T382/L64 (global horizontal resolution ~ 38 km)	3D-Var GSI	1979-2011 (current CFSv2)	as $0.5^\circ \times 0.5^\circ$ (~ 50 km)	6-h accumu- lation at 0000, 0600, 1200 and 1800 UTC	Saha et al. (2010)
ECMWF	ERA- Interim	IFS T255/L60 (global horizontal resolution ~ 79 km)	4D-Var	1979-current	$0.5^\circ \times 0.5^\circ$ (~ 50 km)	6-h and 12-h accumu- lation at 0000 and 1200 UTC	Dee et al. (2011)
JMA	JRA-55	JMA T319/L60 (global horizontal resolution ~ 55 km)	4D-Var	1958-2012 (current JCDAS)	as $0.5616^\circ \times$ 0.5616° (~ 55 km)	3-h accumu- lation at 0000, 0300, ..., and 2100 UTC	Ebita et al. (2009)
NASA	MERRA- 2	GEOS- 5.12.4 AGCM ($lat \times lon$) $0.5^\circ \times$ 0.625° /L72	3D-Var GSI	1980-current	$0.5^\circ \times$ 0.625°	1-h accumu- lation at 0030, 0130, ..., and 2330 UTC	Gelaro (2015); Gelaro et al. (2017)

Precipitation from the weather stations used in this study are not assimilated by the ERA-Interim (Dee et al., 2011), the MERRA-2 (Bosilovich et al., 2015) or the JRA-55 (Kobayashi et al., 2015); but are indirectly assimilated by the CFSR (see subsection 2b; Wang et al. (2011)). Therefore evaluating against these observations provides a reasonably independent measure of accuracy.

3.2.1 Weather station data

Due to the higher spatial variability of precipitation compared to temperature, more precipitation stations are included in this study. Of the 111 geographically-dispersed precipitation stations initially selected, 45 stations with more than 4% missing data were excluded. Of the remaining 66 stations, 57 are from Environment and Climate Change Canada (ECCC) and 9 are from BC Hydro (Table A.1).

Figure 3.1 shows the locations of all 66 stations overlaid with population distribution across BC. Fifty-seven stations are located in valleys (indicated by upside-down triangles), and nine are in non-valley locations (upright triangles).

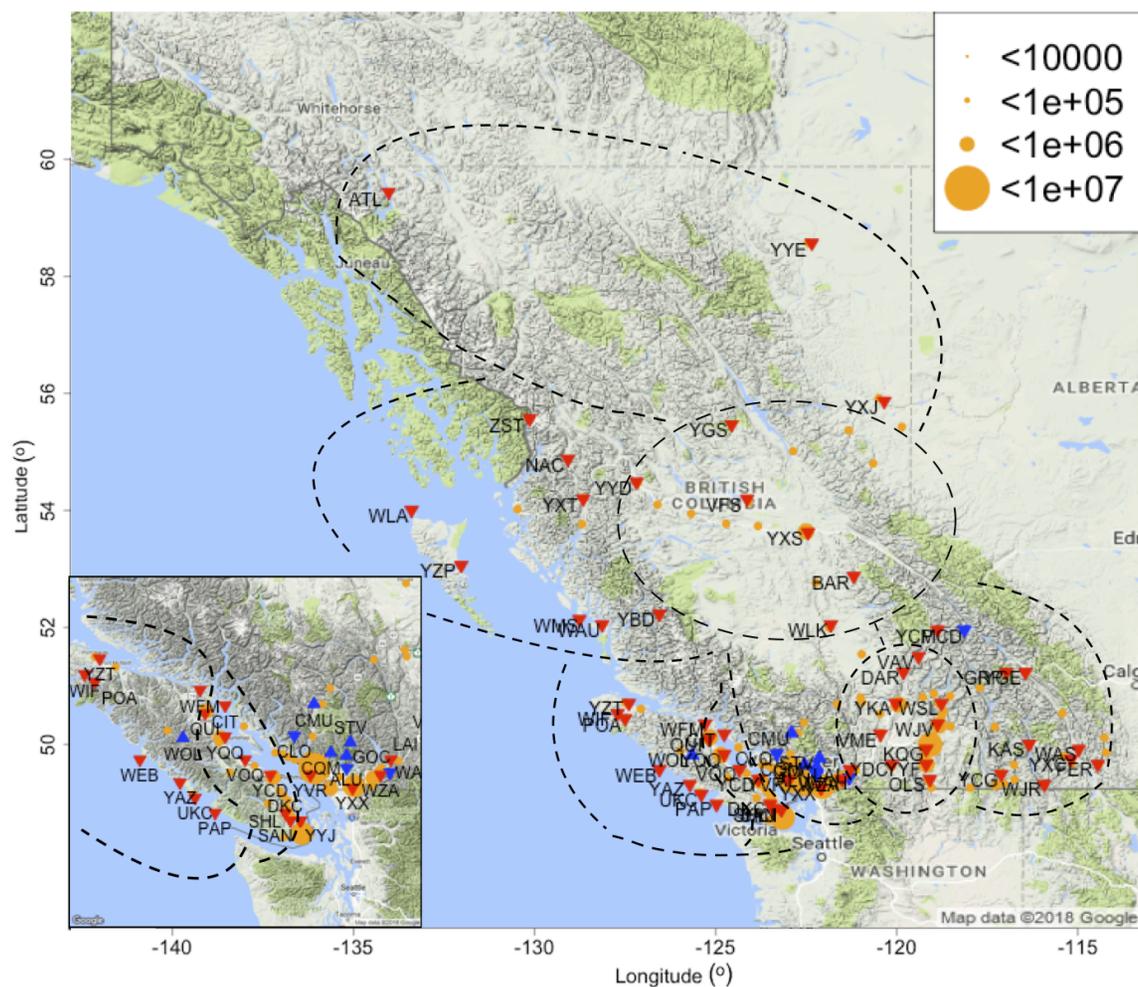


Figure 3.1: Location of ECCC (red) and BC Hydro (blue) weather stations used for precipitation analysis and British Columbia population distribution (orange). Upside-down triangles indicate valley stations; upright triangles indicate non-valley stations. Dashed lines delineate the dominant climate zones North, Northwest, Central, South Central, Maritime West, Maritime East, Southwest and Southeast.

3.2. Data and Methodology

Long-term time series often contain variations caused by changes in the environment surrounding the gauges, instrumentation, observing procedures including the time of observation, station location, or discontinuation of the station. As a result, variations unrelated to changes in weather and climate can be introduced into the time series. Different adjustment techniques for precipitation have been developed to address impacts on climate data homogenization [e.g., Jones et al. (1985); Peterson et al. (1998); Mekis and Hogg (1999); Vincent et al. (2002)].

The methods to adjust daily rainfall and snowfall for ECCC stations are described in Mekis and Vincent (2011). For each rain gauge type, corrections were implemented to account for undercatch due to wind, evaporation, and gauge-specific wetting losses. A complete description of gauges can be found in Metcalfe et al. (1997) and Devine and Mekis (2008). For snowfall, density corrections based on coincident ruler and Nipher measurements were applied to all snow ruler measurements (Mekis and Brown, 2010). A detailed description of trace (non-measurable precipitation amount) related issues and adjustments are given in Mekis (2005) and Mekis and Vincent (2011).

Daily total precipitation was calculated by adding a station's daily rain and snow observations together. In case of station relocation, a new identification number is often given to the new location and observations from the two stations are combined to create a longer time series. Adjustments are applied to join the two datasets, based on standardized ratios between the sites and neighbouring sites, or overlapping observation periods (Vincent et al., 2009).

Data homogeneity for BC Hydro stations was assessed by BC Hydro using Double-Mass Curves (DMC) (Searcy and Hardison, 1960). The theory of the double-mass curve is based on the graph of the precipitation at a station against precipitation of surrounding reference stations during the same period. A break in the slope of the double-mass curve means that a change in the constant of proportionality between the station and surrounding reference stations has occurred. The data before the date that the change occurred is modified to match the historic relationship between the station and its reference stations.

Reliable data is required in order to detect trends in daily and extreme PCP. When dealing with trends of daily data, it is important that the dataset is nearly complete during the studied period. Furthermore, when analyzing decade-long trends, it is important that years with many missing data, if they occur, are relatively few and not clustered during a certain time interval, as this period might have had an anomalous climate (Moberg and Jones,

2005; Vicente-Serrano et al., 2010). Finally, the reliability of frequency of extreme precipitation events is closely related to the sample size used during the study period (Hosking et al., 1985; Hosking, 1990; Porth et al., 2001; Cai and Hames, 2010). Stations with more than 1% missing data are excluded from our nonstationarity analysis, leaving twenty-nine ECCC and six BC Hydro stations.

3.2.2 CFSR

Precipitation is generated by the model during the direct assimilation of temperature and humidity information from satellite radiances (Saha et al., 2010). Then, the model-generated precipitation is corrected with the CPC (the National Oceanic and Atmospheric Administration (NOAA) Climate Prediction Center) Merged Analysis of Precipitation (CMAP) product (Xie and Arkin, 1997) — which defines 5-day mean precipitation on a $2.5^\circ \times 2.5^\circ$ latitude-longitude grid over the globe by merging information derived from gauge and satellite observations — and the CPC Unified Gauge-Based Analysis of Global Daily Precipitation (CPCU) product — which interpolates quality-controlled rain gauge reports collected from the Global Telecommunication System (GTS) and many other national and international archives (Saha et al., 2010) on a $0.5^\circ \times 0.5^\circ$ latitude-longitude grid over the globe. Finally, an algorithm accounts for orographic enhancements in precipitation (Xie et al., 2007).

3.2.3 ERA-Interim

Precipitation is generated by the model during the variational analysis of upper-air atmospheric fields such as temperature, wind, humidity and ozone, in combination with direct assimilation of 2-m temperature, 2-m relative humidity and 10-m winds from land stations, and upper-air temperatures, wind, and specific humidity from radiosonde data (Dee et al., 2011).

3.2.4 JRA-55

Precipitation is model-generated during direct assimilation of upper-air temperatures and humidity information from satellite radiances, and direct assimilation of surface pressure, 2-m temperature, 2-m relative humidity, 10-m winds from land stations, and upper-air temperatures, winds, and specific humidity from radiosonde data (Ebita et al., 2011).

3.2.5 MERRA-2

There are two kinds of precipitation fields in the MERRA-2 system. The precipitation generated by the model during the assimilation procedure (Bloom et al., 1996; Reichle et al., 2017) (PRECTOT is the variable name in the data product file), and the corrected precipitation that is seen by the MERRA-2 land surface and that modulates aerosol wet deposition over land and ocean (PRECTOTCORR is the variable name in the data product file). As mentioned, for a consistent independent evaluation of all reanalyses performance, the former is used in this study.

The precipitation is generated by the model during the direct assimilation of temperature and humidity information from satellite radiances (Bloom et al., 1996; Koster et al., 2016).

3.2.6 PRISM dataset

Precipitation is not evenly distributed over weather station areas in complex terrain (Taesombat and Sriwongsitanon, 2009). In order to estimate areal precipitation, it is preferable to have as many weather stations as possible. However, spatial and temporal coverage is a limiting factor (Karl et al., 1993; Odon et al., 2018), as is accuracy and reliability of precipitation records (Metcalf et al., 1997; Serreze et al., 2005). Additionally, in order to evaluate the agreement between observations and reanalyses, it is important to realize that each grid point in a reanalysis represents an average centered on the geographical coordinates of the grid point. By contrast, an observation represents a point measurement within a reanalysis grid-box, which may or may not be representative of the grid box average. Furthermore, grid resolution and location in each reanalysis dataset is different, and the location of a station can vary from the centre to the edge of the grid box. Interpolation techniques may produce inaccurate results because of the effects of topographical variation and the limited number of available rainfall stations (Taesombat and Sriwongsitanon, 2009). In order to identify regional and terrain biases and to improve the accuracy of areal rainfall estimation, the reanalyses are bilinearly interpolated to the same high-resolution grid as the Parameter-Elevation Regressions on Independent Slopes Model (PRISM; Daly et al. (1994, 1997, 2002)) climatology for grid comparison.

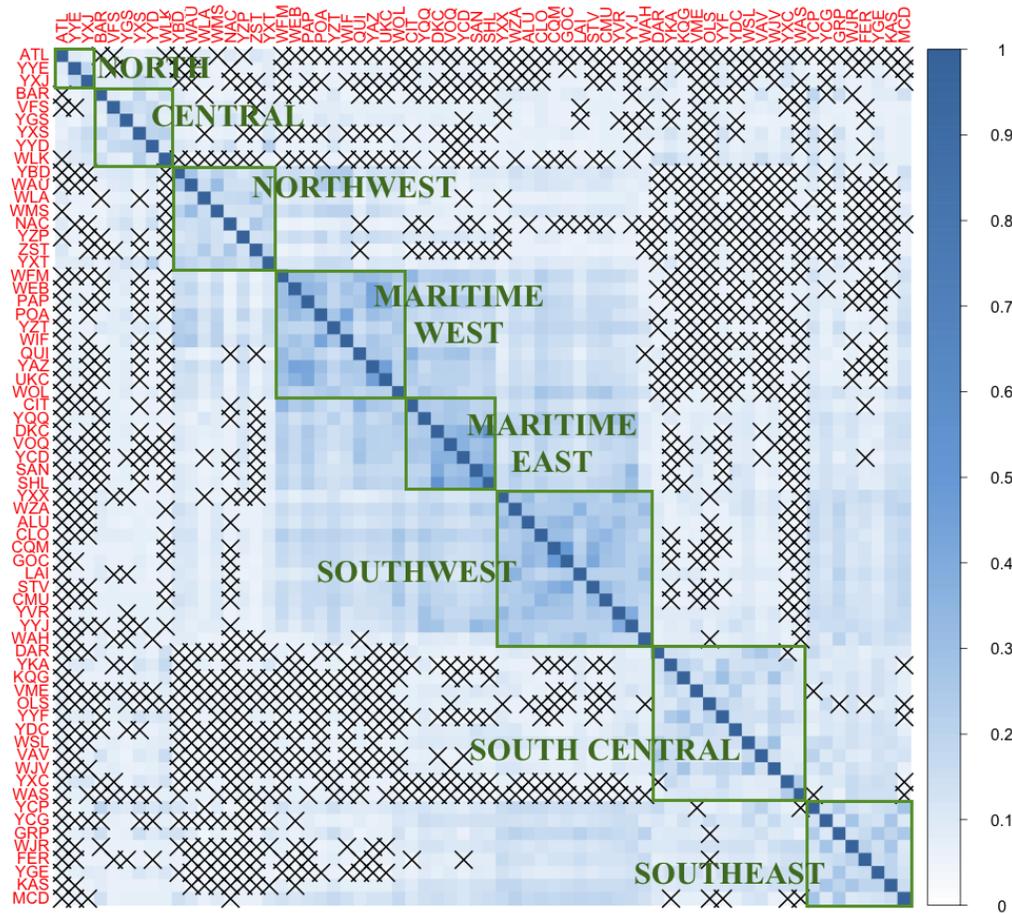
PRISM climatology, produced by the Pacific Climate Impact Consortium (PCIC), the Pacific Institute for Climate Solutions (PICS) and the BC Government provides access to a 30-arc-second (~ 800 m) gridded precipitation

climatology for the 1981-2010 climate normal period, for land-surface areas of BC (PCIC et al., 2014). PRISM has been tested and verified throughout the United States and has been applied in numerous countries across the globe including western Canada previously for the 1961-1990 period. In this study, the 1981-2010 climate period is used.

3.3 Climate zones

Principal component analysis (PCA) is employed to separate stations into groups with similar precipitation climates. Because precipitation has both frequency and magnitude, contingency tables are employed. Daily PCP (mm) is divided into seven non-overlapping bins for the entire study period: $[0, 0.25]$, $(0.25, 1.0]$, $(1.0, 2.5]$, $(2.5, 5.0]$, $(5.0, 10.0]$, $(10.0, 20.0]$ and $(20.0, \infty)$. The number of events falling into each bin is recorded in a contingency table for the entire study period. A χ^2 -test compares the pairwise difference between the distributions of the 66 stations. A sufficiently large difference between the distribution of the daily PCP amounts over the seven bins indicate the stations have different precipitation characteristics and therefore belong to different climate zones. Small differences suggest that the stations have similar climates. This critical difference is determined by the χ^2 -distribution. Due to the large number of stations pair comparisons (2141 pairs), the null hypothesis that the stations belong to the same climate zone is rejected at the $\alpha_{Walker} = 1 - (1 - \alpha_0)^{1/N_0} = 2.31 \times 10^{-5}$ level of significance, where $\alpha_0 = 0.05$ and $N_0 = 66$ (Fig. 3.2).

Cramer's correlation is computed for each pair χ^2 -statistic, and PCA is conducted on the Cramer's correlation matrix of daily PCP. The first 41 components explain 90% of the variability in the data, and are retained. The number of components is larger here than in Chapter 2 of this study because of the higher variability in precipitation. For temperature, where the Pearson correlation between the stations is high (see section 2.3 for more details), it is possible to capture most of their variance using a smaller number of principal components. For precipitation, where the Cramer correlation between the stations is lower (Fig. 3.2), more principal components are needed to capture most of their variance.



3.3. Climate zones

Figure 3.2: Cramer correlation matrix and eight dominant climate zone clusters in BC. Crosses indicate stations where differences in precipitation distribution are statistically significant.

A K-means clustering analysis is then performed on the components to arrange the data into groups. One to 10 clusters were tested and an eight-cluster solution was chosen (Fig. 3.2). This analysis yields 8 climate zones (North, Central, Northwest, Maritime West, Maritime East, Southwest, South Central, and Southeast) that roughly matched to those identified in Chapter 2.

The Maritime climate zone from Chapter 2 has been divided into Maritime East and Maritime West, while the Southeast climate zone has been divided into Southeast and South Central. These additional climate zones highlight the differences between windward and leeward regions, which see enhanced upslope precipitation and rain shadows, respectively.

3.4 Verification Metrics

The statistical behaviours of daily and extreme PCP are compared between observed weather station data, and their corresponding location in the reanalyses and PRISM. As in Chapter 2, four horizontal interpolation methods were trialed for reanalysis output. Also as in Chapter 2, the methods Nearest Neighbour, Inverse Distance Weighting (IDW), Bilinear and Bicubic interpolation perform very similarly; IDW is used. Furthermore, IDW from land-only grid points (omitting sea grid points) was also tested for coastal stations in the Maritime West, Maritime East, Southwest and Northwest climate zones (a grid point is classified as a land point based on each reanalysis land-sea mask). Some small variations between the two IDW methods were found, but overall the land-only IDW is similar in result to the IDW.

3.4.1 Daily PCP

The Canadian meteorological convention in defining a calendar day to be from 0601 UTC of that day to 0600 UTC of the following day is followed. For CFSR, JRA-55 and MERRA-2, precipitation accumulation intervals are summed over this window to get daily PCP (Table 3.1). For ERA-Interim, precipitation is accumulated in the forecast sense, i.e., reset to zero at 0000 and 1200 UTC. Six-hour accumulated precipitation for the 6-h periods *preceding* 0000, 0600, 1200 and 1800 UTC are obtained in the following manner: for 0600 UTC the 0000-UTC 6-h accumulated precipitation is used; for 1200 UTC the 0000-UTC 6-h accumulated precipitation is subtracted from the 0000-UTC 12-h accumulated precipitation; for 1800 UTC the 1200-UTC 6-h accumulated precipitation is used; and for 0000 UTC the 1200-UTC 6-h precipitation is subtracted from the 1200-UTC 12-h accumulated precipitation.

A 31-day centered rolling accumulation window is used to obtain smooth monthly mean precipitation for each calendar day. These 31-day accumulated values for each calendar day are then averaged over the 31-year evaluation period (1980-2010). This is done for each station and reanalysis data. The percentage bias (or systematic error) is then computed to estimate how accurately each reanalysis captures monthly precipitation (31-day precipitation totals).

For the same 31-day centered rolling window, days with daily PCP below 1.0 mm are classified (and hereafter referred to) as "dry" days, and days equal or above 1.0 mm are classified (and hereafter referred to) as "wet" days. In climate studies, this delineation is typically made at trace amounts or 1.0 mm (Vincent and Mekis, 2006; Werner and Cannon, 2016). In this study 1.0 mm was chosen because coarse resolution models tend to overforecast the frequency and spatial extent of light precipitation events (e.g., Zhu and Luo (2015), and shown later in this study). On every 31-day window, systematic error is computed from the number of wet days at a station location in the reanalysis with respect to the actual number of wet days observed at a station.

Finally, wet days are divided into the five non-overlapping intervals: [1.0 mm, 50th), [50th, 75th), [75th, 90th), [90th, 95th) and [95th, 100th]. The percentiles are calculated from the entire wet-day climatological distribution, centered on each calendar day, using each station's observed data.

This allows for an evaluation of a reanalysis' ability to correctly capture the frequency *and* distribution of precipitation intensities. If the reanalysis distribution is very close to that of the observation, the number of expected occurrences in each bin will be very close. This difference between the number of "light" ([1.0 mm, 50th)), "light" ([50th, 75th)), "moderate" ([75th, 90th)), "heavy" ([90th, 95th)) and "extreme" ([95th, 100th)) precipitation events in each dataset is given by the two-sample χ^2 -statistic.

3.4.2 Extreme PCP

The definition of an extreme precipitation event varies widely. One possibility is to define it as an event in which precipitation over some specified period exceeds some threshold, either at a point measured by a single rain gauge, or spatially averaged over some region. The choice of threshold also varies. Some studies use fixed absolute thresholds while others use a fixed percentile based on the distribution specific to a given location, so that it is specific to the location climatology. In this study, the 2- and 30-yr return levels for every station for each calendar day in the 1980-2010 study period

are estimated.

Furthermore, end users of precipitation forecasts, such as hydrologists, are concerned with both peak flows and total volume of flows, especially when they deviate from climatology. Correspondingly, they require accurate estimates of precipitation intensity and accumulation over a range of time scales. For the flashy reservoirs of the BC South Coast, the timescales of their total volume concerns typically range from 1 to 14 days. Hydropower facilities and their associated operating procedures are designed assuming estimated minimum and maximum volumes that could be possibly expected at various timescales. Extreme precipitation accumulations, whether accumulated over 1 or 14 days, will cause extreme total flow volumes over those time scales, pushing or possibly exceeding the limits of a hydropower facility. The effects of heavy or extreme precipitation events can be compounded if components of a hydropower facility (e.g., spill gates) are out of service for planned or unplanned maintenance. Given that flow volumes are important at a range of time scales, 2- and 30-year return levels of 1-, 3-, 7-, and 14-day accumulated precipitation are examined (2- and 30-year return levels are also known as 2- and 30-year recurrence intervals, or 0.5 and 0.03 Annual Exceedance Probabilities (AEP)).

To do this, the maximum 1-, 3-, 7-, and 14-day accumulated precipitation within a 31-day centered rolling window, for each calendar day are calculated (Fig. 3.3a). This is done for each year from 1980-2010 inclusive, yielding 31 annual maximum values of 1-, 3-, 7-, and 14-day accumulated precipitation for each calendar day. A 31-day window was chosen so that all values within the window are from the same time of year, and would have similar climatological precipitation distributions.

A Generalized Extreme Value distribution (GEV) fitted by the method of L moments dresses these 31 sample values of 1-, 3-, 7-, and 14-day accumulated precipitation for each calendar day (Fig. 3.3b-d). A GEV is chosen because of the interest in the statistical behaviour of the 31 annual maximum values of 1-, 3-, 7-, and 14-day accumulated precipitation at each calendar day. Estimates of the 2- and 30-year return levels of annual maximum are then obtained from the fitted GEV.

The Lilliefors (Wilks, 2011) test compares the largest difference, in absolute value, between the theoretical GEV cumulative distribution function (CDF) and the observed empirical cumulative distribution function (ECDF). The null hypothesis is that the observed data is drawn from the distribution being tested (i.e., the observation ECDF and GEV CDF are indistinguishable), and a sufficiently large critical difference results in the null hypothesis being rejected.

3.4. Verification Metrics

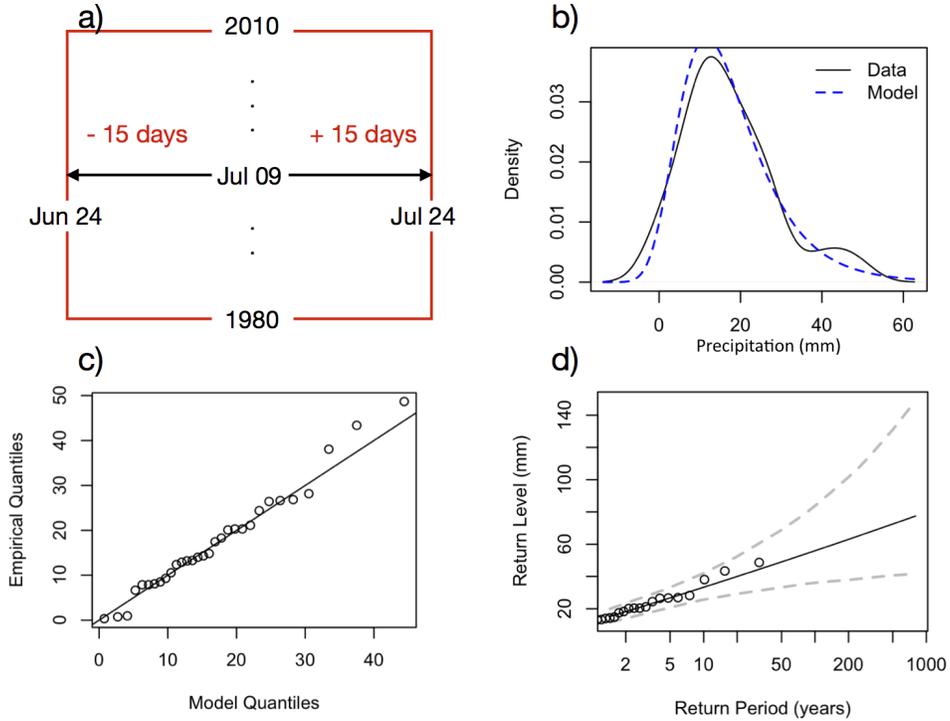


Figure 3.3: (a) Diagram illustrating the 31-day centered rolling window for July 9th, performed over the 31-year period; (b) GEV model fitted over the 31-day, 31-year centered window for 1-day precipitation total at Vancouver International Airport (YVR); (c) Quantile plot for the GEV fitted model for the same day, location, and variable; If the GEV is a reasonable model, the points on the quantile plot should lie close to the unit diagonal; (d) Return level, return period plot for the same day, location, and variable; showing the precipitation that corresponds to a given return period.

A parametric bootstrap procedure determines the critical value that results in the rejection of the null hypothesis. 100 samples of size 31 are generated from the fitted GEV distribution for each calendar day at each station, and a critical value is derived from each of the 100 generated samples. That is, the 90th percentile of the resulting collections of critical values is then used as the critical value for the rejection of the null hypothesis — that the sample originates from the GEV distribution at the 10% significance level ($\alpha_0 = 0.10$). However, there are 66 weather stations and 365 calendar days totaling $N_0 = 24090$ independent Lilliefors tests. Due to very large number of tests, the $\alpha_{Walker} = 1 - (1 - \alpha_0)^{1/N_0} = 4.37 \times 10^{-6}$ is instead regarded as significant, and the $(1 - \alpha_{Walker})$ th percentile of the resulting collections of critical values is then used as the critical difference (Wilks, 2016).

Less than 1% of the locations and calendar days where the null hypothesis is tested for all different accumulations totals were rejected during the 1980-2010 study period, suggesting the annual extremes in fact can be described by a GEV distribution.

The systematic error of the 2- and 30-year return levels were then calculated to estimate how each reanalysis captures annual extremes of 1-, 3-, 7-, and 14-day accumulated precipitation. These two return levels are chosen because less than 1% of their 90% confidence intervals overlap, indicating the difference between the two return levels is statistically significant, and because the 30-year return level is the most extreme verifiable value given the length of the data.

To calculate their 90% confidence intervals, 100 samples of size 31 are generated from the fitted GEV distribution for each calendar day at each station, and the 2- and 30-year return levels are estimated from each generated sample. Then, the 5th and 95th percentiles of the resulting collection of 2- and 30-year return levels are used as the lower and upper bounds of the 90% confidence intervals for the true 2- and 30-year return levels.

3.4.3 Kruskal-Wallis Analysis

The mean systematic error of monthly precipitation total, two-sample χ^2 -statistic, 31-day-window percentage of wet days, and of 2- and 30-year return levels of 1-, 3-, 7-, and 14-day accumulated precipitation are calculated for each station from all calendar days systematic errors. Comparisons between these eleven mean systematic errors of each reanalyses (CFSR, ERA-Interim, JRA-55 and MERRA-2) are performed using eleven independent Kruskal-Wallis nonparametric tests. Eleven independent Kruskal-Wallis tests are used because the mean systematic errors are skewed, and due to the differ-

ent magnitude and variability of each type of mean systematic error. Finally, Nemenyi's test (Hollander et al., 2013) is applied following statistical significance at the $\alpha_{Walker} = 9.53 \times 10^{-3}$ level ($\alpha_0 = 0.10$) in the Kruskal-Wallis results to identify significant performance differences in pairwise comparisons between the reanalyses mean systematic errors.

3.5 Results and discussion

3.5.1 Daily PCP

First, performance of daily PCP across the climate zones in the reanalyses are investigated. Reanalysis performance in the Northwest climate zone (Fig. 3.4) is representative of performance across the wetter climate zones (Northwest, Maritime West, Maritime East and Southwest; latter three not shown). The seasonal cycle and magnitudes of 31-day precipitation totals are fairly well captured. All four reanalyses exhibit similar seasonal cycles and differ mostly in magnitude of annual bias (Table 3.2). They show a positive (wet) bias all year long for the Northwest and Maritime East climate zones. In the Maritime West (the wettest climate zone in BC and in all of Canada) and Southwest zones, JRA-55, ERA-Interim and CFSR have a negative (dry) bias throughout the year (not shown).

Table 3.2: Averaged systematic error of: monthly precipitation total (M), two-sample χ^2 -statistic (χ^2), 31-day-window percentage of wet days (W), and 30-year return levels of 1- (1D30), 3- (3D30), 7- (7D30), and 14-day (14D30) accumulated precipitation across wetter climate zones Northwest, Maritime West, Maritime East and Southwest, drier climate zones North, South Central, Central and Southeast, and all climate zones in BC (all systematic errors but two-sample χ^2 -statistic in %)

	M	χ^2	W	1D30	3D30	7D30	14D30
Wet							
CFSR	8.57	26.16	6.44	-31.35	-21.24	-13.74	-8.71
ERA-Interim	14.98	26.44	5.57	-11.83	-3.74	0.80	2.57
JRA-55	11.50	18.92	4.29	-15.79	-6.51	-2.32	-0.03
MERRA-2	23.22	18.10	9.04	11.30	-4.80	0.79	3.46
Dry							
CFSR	98.56	15.90	20.41	-2.15	12.21	26.01	38.01
ERA-Interim	55.48	16.23	14.74	-11.99	-1.05	7.01	15.67
JRA-55	59.30	13.93	12.63	-6.53	5.59	13.06	19.98
MERRA-2	69.45	12.44	17.65	2.89	10.61	17.04	24.46
All							
CFSR	48.11	21.65	12.58	-18.52	-6.54	3.73	11.82
ERA-Interim	32.77	21.96	9.60	-11.90	-2.56	3.53	8.33
JRA-55	32.50	16.73	7.95	-11.72	-1.19	4.43	8.76
MERRA-2	43.53	15.61	12.83	-5.05	1.98	7.93	12.69

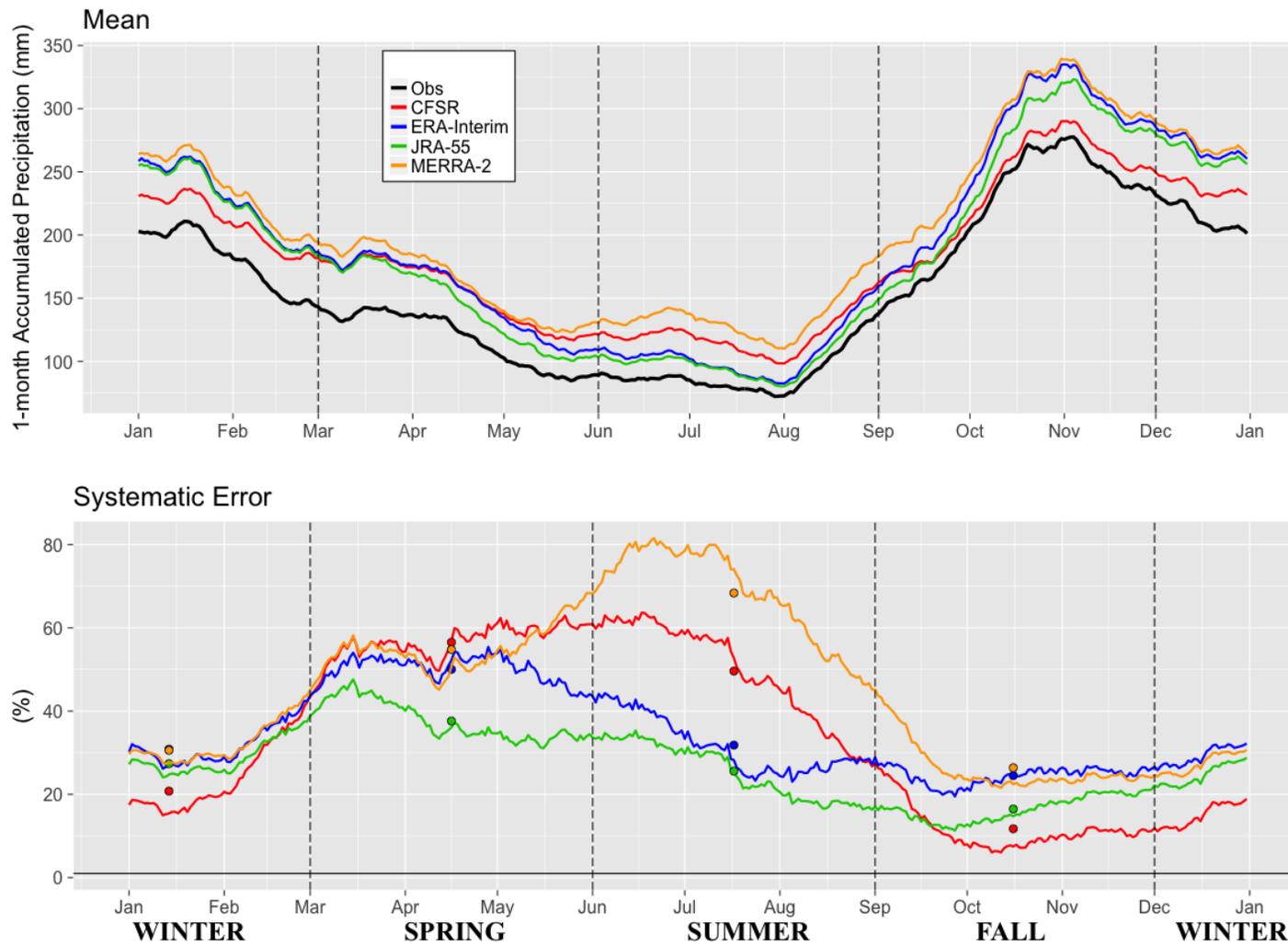


Figure 3.4: Observed and reanalysis running centered 31-day precipitation totals and systematic error averaged over stations in the Northwest climate zone. The vertical dashed lines indicate the change in seasons and the colored dots represent the seasonal average.

3.5. Results and discussion

Overall, JRA-55 and ERA-Interim outperform CFSR and MERRA-2. The latter two exhibit higher bias and higher variability in bias throughout the wetter climate zones and seasons.

In contrast, in the South Central zone, with the driest locations in all Canada, all reanalyses have a large wet bias all year long (not shown). In the remaining drier climate zones, North, Southeast and Central, reanalyses have a large wet bias for most of the year and the largest wet bias during spring (Fig. 3.5 for Central climate zone). Again, JRA-55 and ERA-Interim outperform CFSR, and to a lesser extent MERRA-2 (Table 3.2). The general wet bias in most zones could be a result of the low resolution of the reanalyses that tends to spread out precipitation into drier portions of grid cells, failing to capture the locally drier climates of lower (valley) elevations, where most stations are located.

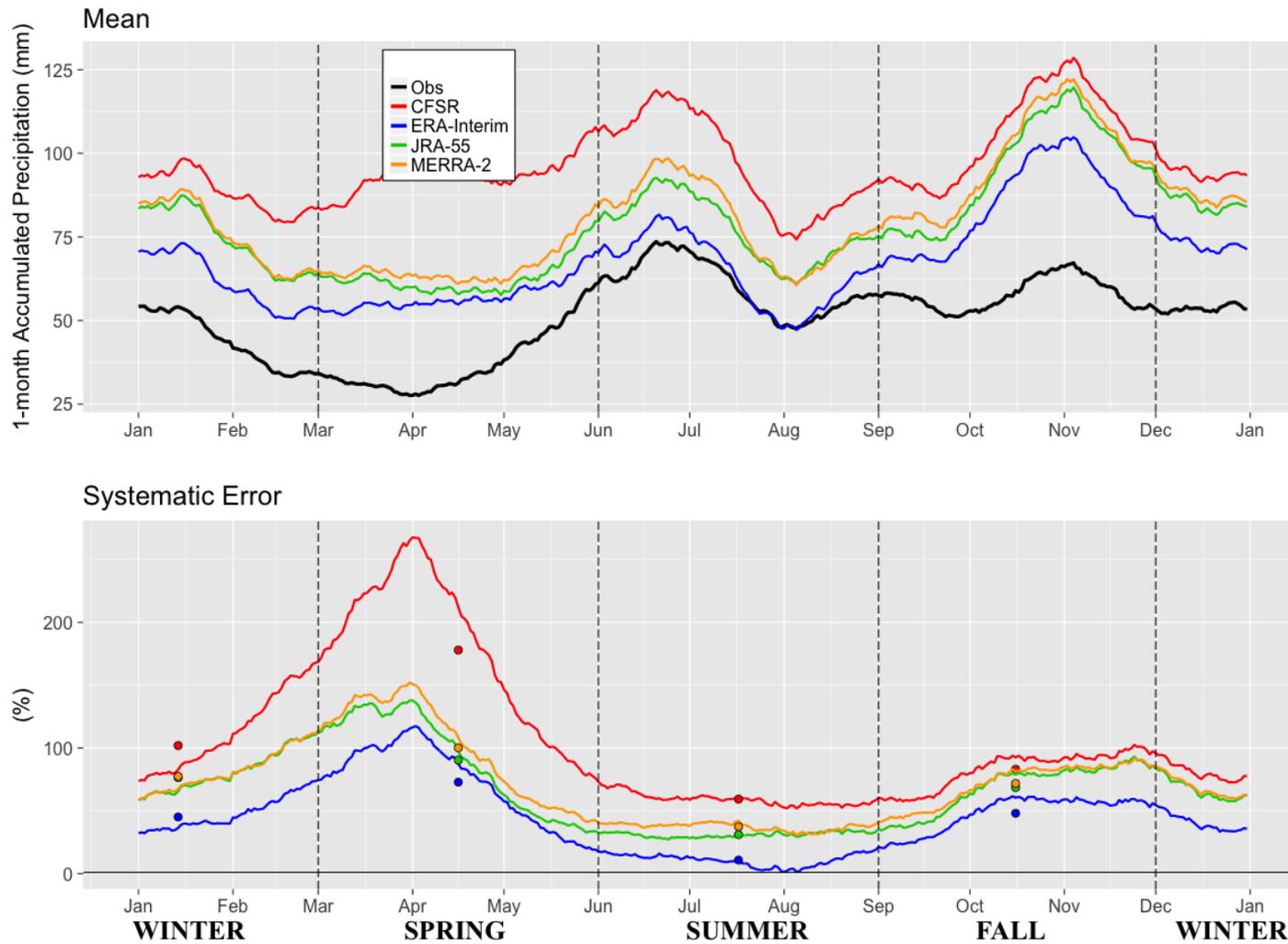
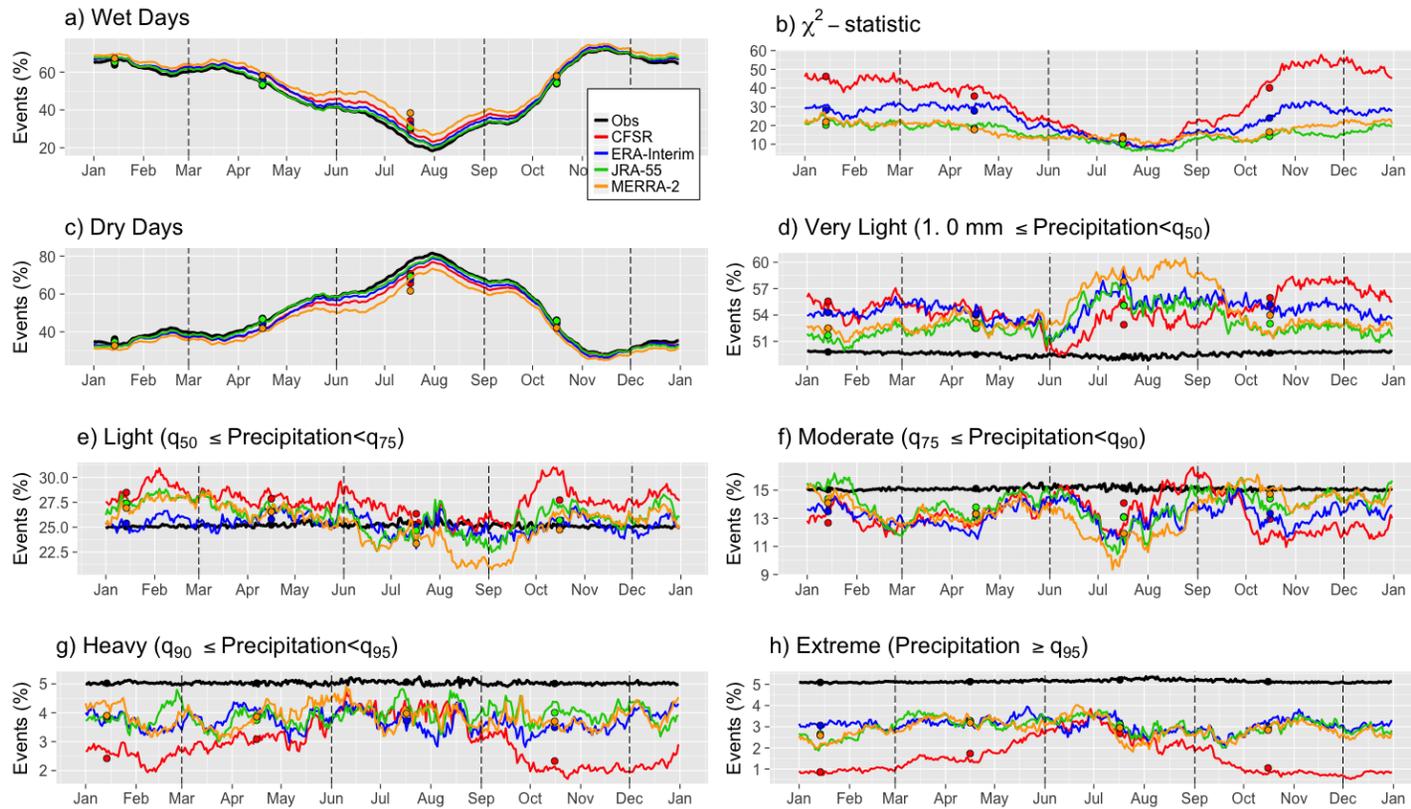


Figure 3.5: Observed and reanalysis running centered 31-day precipitation totals and systematic error averaged over stations in the Central climate zone. The vertical dashed lines indicate the change in seasons and the colored dots represent the seasonal average.

3.5. Results and discussion

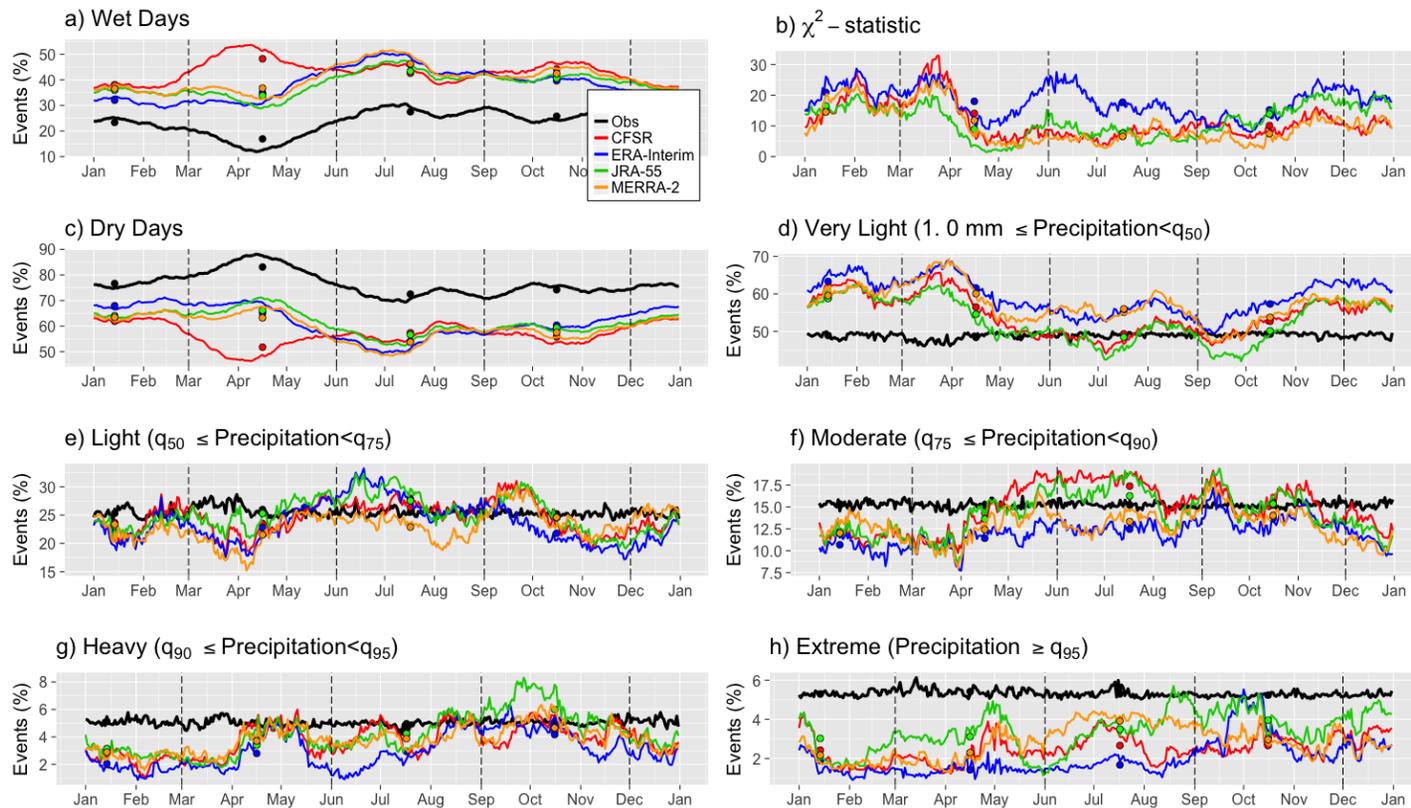
The percentages of “wet” and “dry” days, and the two-sample χ^2 -statistic are illustrated in Figure 3.6 (a, b, and c) for the Maritime West climate zone, which is representative of the wetter climate zones (Maritime West, Southwest, Northwest and Maritime East; latter three not shown). The reanalyses capture the annual cycle of precipitation frequency better than they do the precipitation amounts (Fig. 3.6a, c; Table 3.2). The two-sample χ^2 -statistic (Fig. 3.6b) is used to determine how well the reanalyses capture the histogram of observed precipitation events across all bins. Lower values are better, and 0 would indicate no difference between the observation histogram and that of the reanalysis. The MERRA-2 is the best due to its consistent low values of χ^2 -statistic across the wetter climate zones, followed closely by JRA-55 (Table 3.2). Finally, looking at the precipitation percentile bins, reanalyses overestimate the number of “very light” and “light” precipitation events (Fig. 3.6d, e), and underestimate the number of “moderate”, “heavy” and “extreme” precipitation events for the Maritime West and Southwest zones (Fig. 3.6f-h). The opposite is true for the Maritime East and Northwest (not shown).



8 Figure 3.6: Percentage of (a) wet days; (b) two-sample χ^2 -statistic; percentage of (c) dry days, (d) very light, (e) light (f) moderate, (g) heavy and (h) extreme precipitation events for Maritime West climate zone. The vertical dashed lines indicate the change in seasons and the colored dots represent the seasonal average.

Reanalysis performance in the North region (Fig. 3.7) is representative of the drier climate zones (North, South Central, Central and Southeast; latter three not shown). All reanalyses substantially overestimate the percentage of “wet” days, and the opposite for “dry” days (Fig. 3.7a, c; Table 3.2). MERRA-2 and JRA-55 outperform CFSR and ERA-Interim for two-sample χ^2 -statistic (Fig. 3.7b; table 3.2). Reanalyses overestimate the occurrence of “very light” precipitation events (Fig. 3.7d-h), somewhat captures “light”, “moderate”, and “heavy” events (Fig. 3.7d-g), and underestimates “extreme” precipitation events (Fig. 3.7h) for the North Southeast climate zones. For the Central and South Central all precipitation type events are well captured.

Overall, across all climate zones and reanalyses, “very light” and “light” precipitation events are overestimated. This is expected as lower-resolution models tend to overforecast such events; and extreme precipitation events are underestimated. Lower-resolution models and reanalyses typically are not able to resolve extreme precipitation maxima, especially in the wettest zones Maritime West and Southwest.



∞ Figure 3.7: Percentage of (a) ”wet days”; (b) Two-sample χ^2 -statistic; Percentage of (c) ”dry days”, (d) ”very light”, (e) ”light”, (f) ”moderate”, (g) ”heavy” and (h) ”extreme” precipitation events for North climate zone. The vertical dashed lines indicate the change in seasons and the colored dots represent the seasonal average.

3.5. Results and discussion

Long-term, homogenized stations in mountainous BC are mostly located in valleys (Fig. 3.1). To identify regional and upper-elevation biases, each reanalysis grid is bilinearly interpolated to the PRISM grid for comparison to PRISM's 30-year mean seasonal values. Similar to station-to-station comparison, JRA-55 and ERA-Interim outperform CFSR and MERRA-2 (Fig. 3.8). All four reanalyses show a dry bias during winter along the windward and upper elevations of the Islands, Coast Mountains and Rocky Mountains where the Maritime West, Northwest and Southeast climate zones are located; and a wet bias along the Interior Plateau and leeward, lower elevation regions of Vancouver Island and Lower Mainland where the Maritime East, Southwest and Central climate zones are located. The North climate zone has the lowest systematic error across the better reanalyses. Fall results are similar to Winter results (not shown). Summer has the lowest systematic errors across the better reanalyses (JRA-55 and ERA-Interim). MERRA-2 and CFSR exhibit a wet bias across the entire province (not shown). Finally, Spring has the largest systematic errors across all four reanalyses with wet biases in the Maritime East, Southwest, Central and North climate zones, and dry biases across the Maritime West, Northwest and Southeast climate zones. CFSR followed by MERRA-2 shows a larger wet bias across most of the province (not shown).

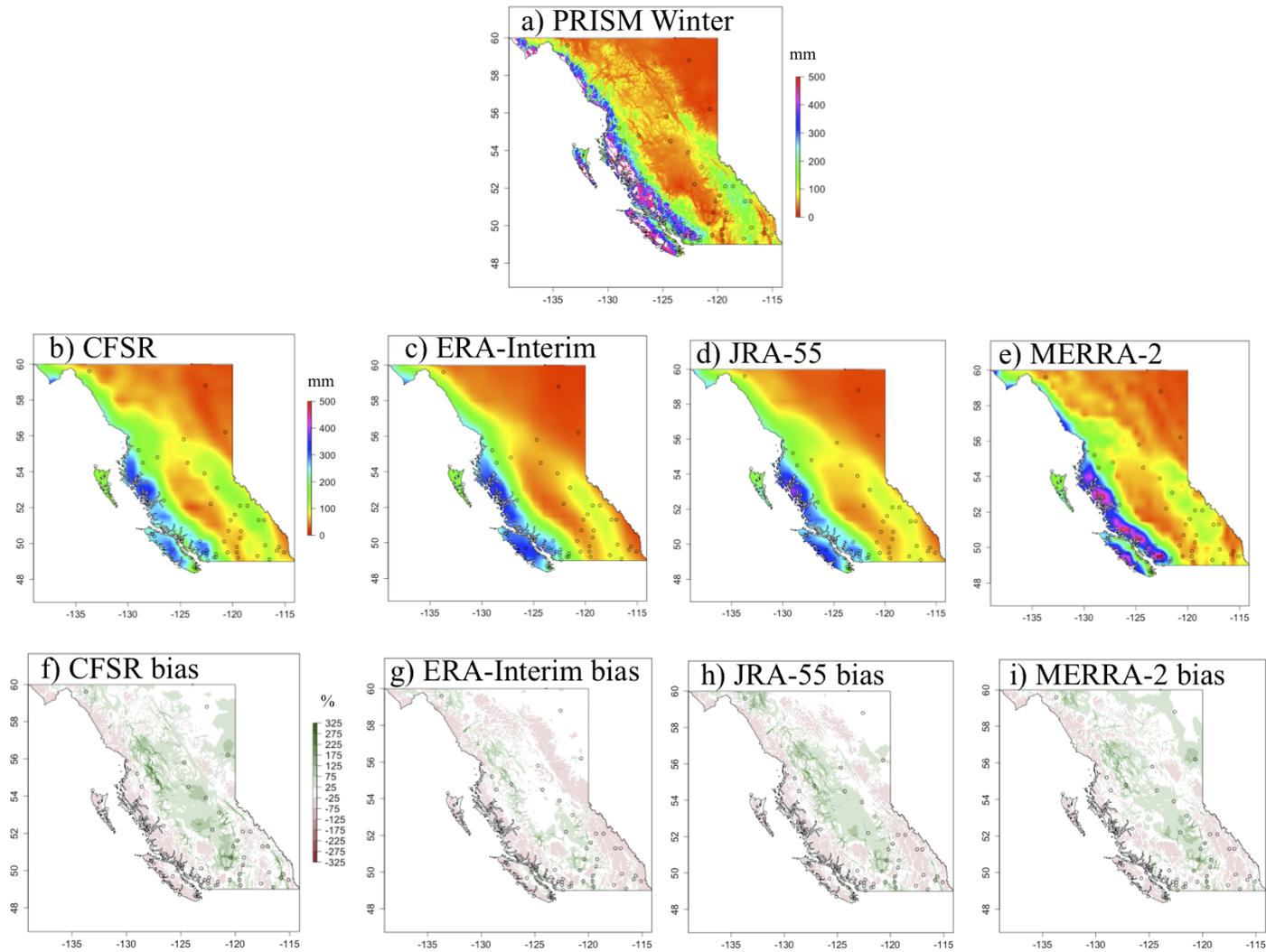


Figure 3.8: a) Winter precipitation totals of (a) PRISM. Winter precipitation totals bilinearly interpolated to PRISM grid of (b) CFSR, (c) ERA-Interim, (d) JRA-55 and (e) MERRA-2. Systematic error of (f) CFSR, (g) ERA-Interim, (h) JRA-55 and (i) MERRA-2. The dots represent weather station location.

An examination of the magnitudes of the biases of each metric at each station shows that reanalysis precipitation systematic error relative to weather station observations is highly correlated to reanalysis precipitation systematic error relative to PRISM (PRISM and all four reanalyses are bilinearly interpolated to station locations for station-to-station comparison). PRISM was developed to create a climatological precipitation on a regularly space grid that addresses spatial scales and patterns of orographic precipitation (Daly et al., 1994). This is an indication that the reanalysis biases in 31-day precipitation totals (shown for JRA-55 in Fig. 3.9) can be largely explained by topographic and synoptic parameters such as terrain steepness, exposure, elevation, location of barriers, and wind speed and direction, that are incorporated into PRISM. Strong correlations are also obtained for CFSR ($0.78 \leq R \leq 0.94$), ERA-Interim ($0.79 \leq R \leq 0.93$) and MERRA-2 ($0.77 \leq R \leq 0.91$; not shown). Hence, the PRISM accurately represents station observations, and can be used for bias correction and downscaling of reanalyses.

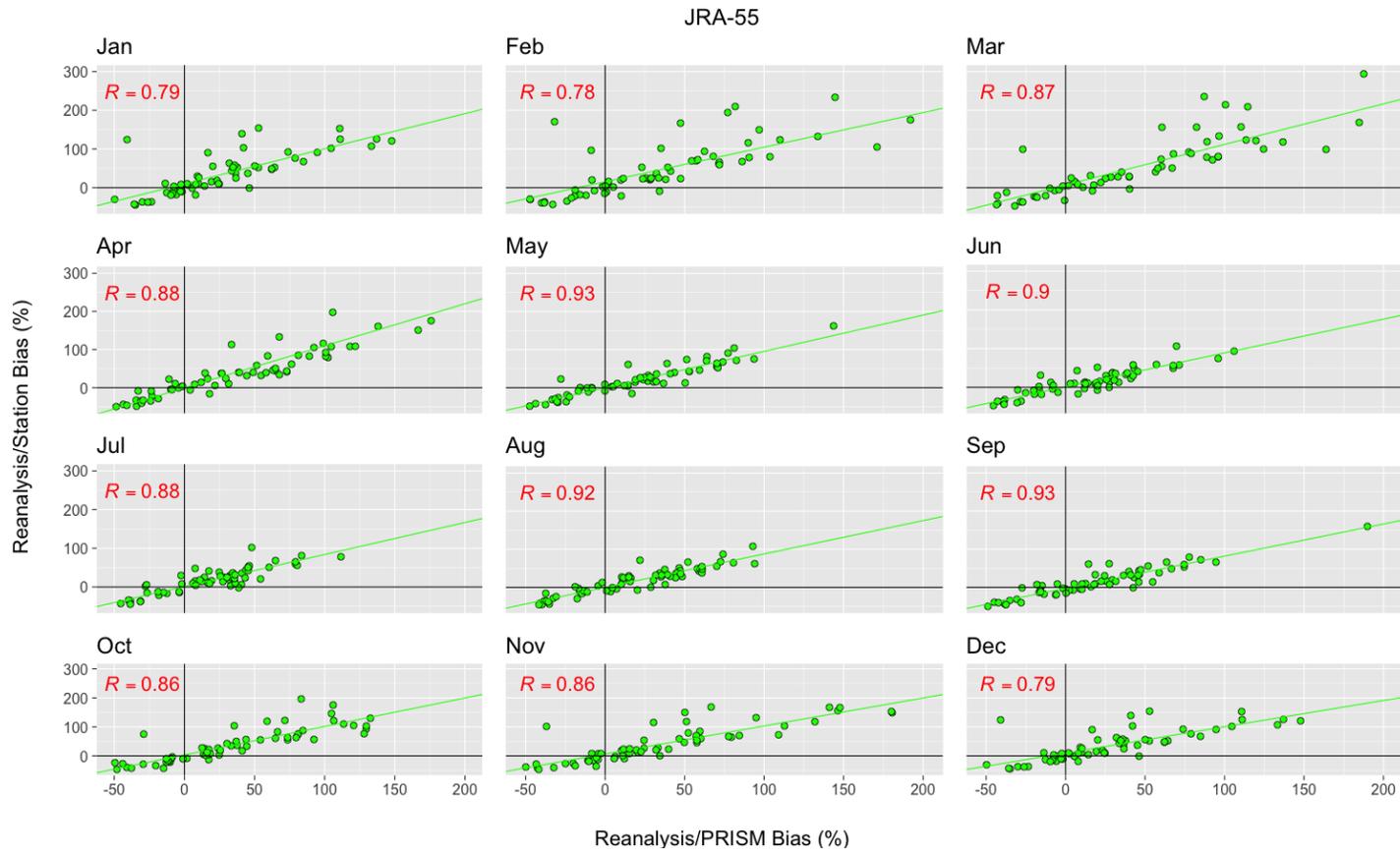


Figure 3.9: JRA-55 mean systematic error of 31-day precipitation totals for each of the 66 stations as a function of PRISM mean systematic error. The solid lines show the linear regression fits.

3.5.2 Extreme PCP

The Southwest climate zone has the highest population density, and reservoir sizes in the region are small relative to the magnitude of heavy and extreme precipitation events. These two factors make it among the most sensitive to extreme precipitation events. It is also one of the wettest regions — it has one of the highest 30-year return levels of 1-, 3-, 7-, and 14-day accumulated precipitation, second only to the Maritime West zone. Similar to the results of daily PCP for the wetter climate zones, the Southwest (Fig. 3.10) and Maritime West climate zones show that all reanalyses are typically too dry for extreme precipitation events, and too wet for the Northwest and Maritime East climate zones.

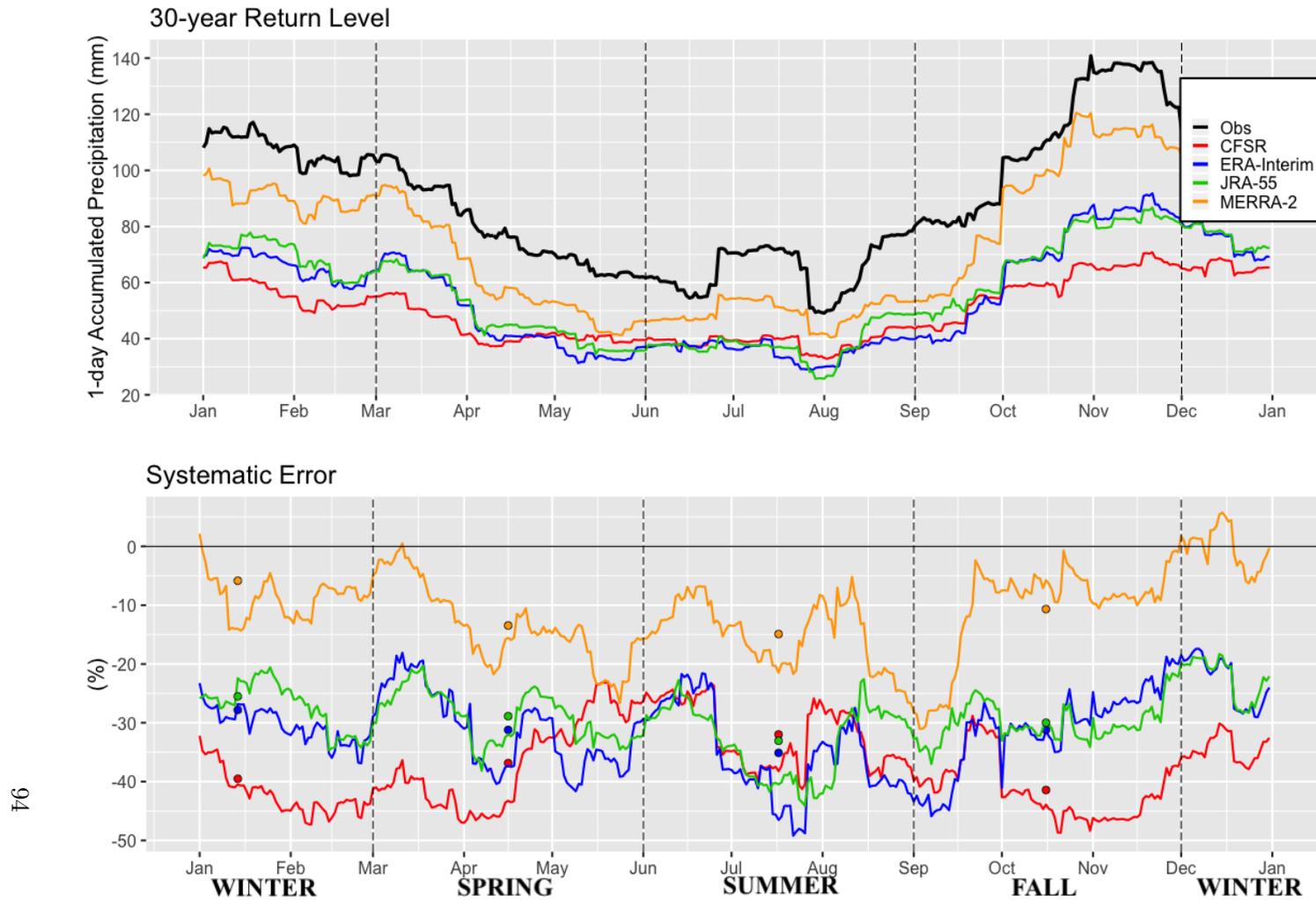


Figure 3.10: 30-year return level and systematic error of 1-day precipitation total, for all four reanalyses for Southwest climate zone. The vertical dashed lines indicate the change in seasons and the colored dots represent the seasonal average.

MERRA-2 performs best in this regard, followed by ERA-Interim and JRA-55, and then the CFSR. The percent magnitude of the 1-day dry biases in 30-year return levels are similar to those accumulated over 3, 7, and 14 days (not shown; Table 3.2).

For the wetter climate zones, the highest values of extreme precipitation occur during storm season (October to February for southwest BC; Fig.3.10). These tend to be associated with non-convective, synoptic systems.

For the drier climate zones the highest values of extreme precipitation occur during summer. Biases are smaller and typically associated with thunderstorm convection. In these zones, all reanalyses generally exhibit a dry bias (e.g., North, Fig. 3.11), with the exceptions that MERRA-2 has a wet bias during the summer peak, and JRA-55 a near zero bias. Furthermore, all reanalyses have smaller biases when compared to the wetter climate zones (cf. Figs. 3.10, 3.11), indicating that 30-year return levels of 1-, 3-, 7-, and 14-day precipitation totals are fairly well captured all year long for all accumulation periods (not shown) in drier zones. This is notably different from the inability of the reanalyses to capture the annual cycle of daily precipitation in drier zones (Fig. 3.5). This is somewhat surprising since one might expect relatively coarse-resolution reanalyses to capture monthly accumulated precipitation more accurately than extreme convective events.

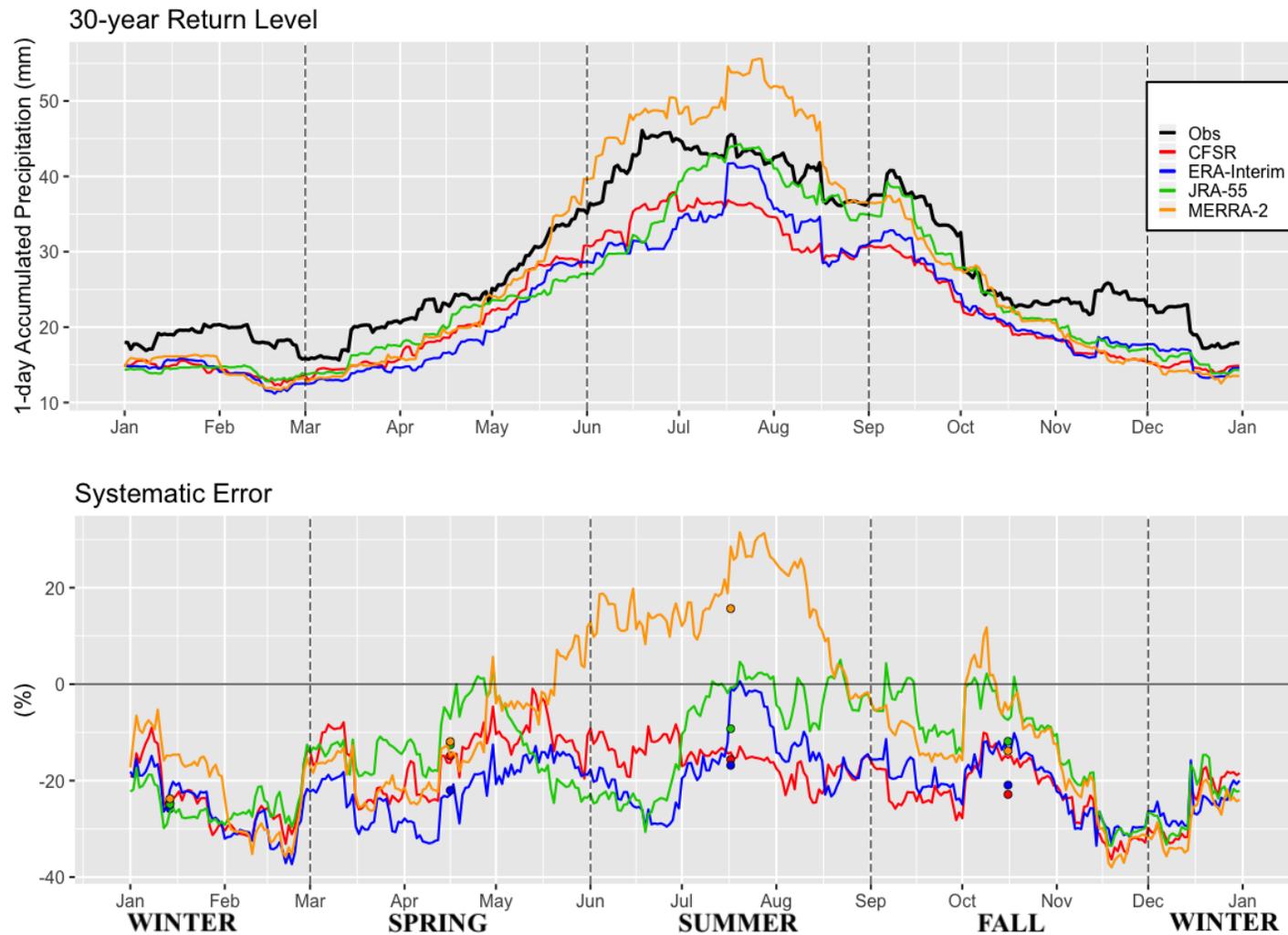
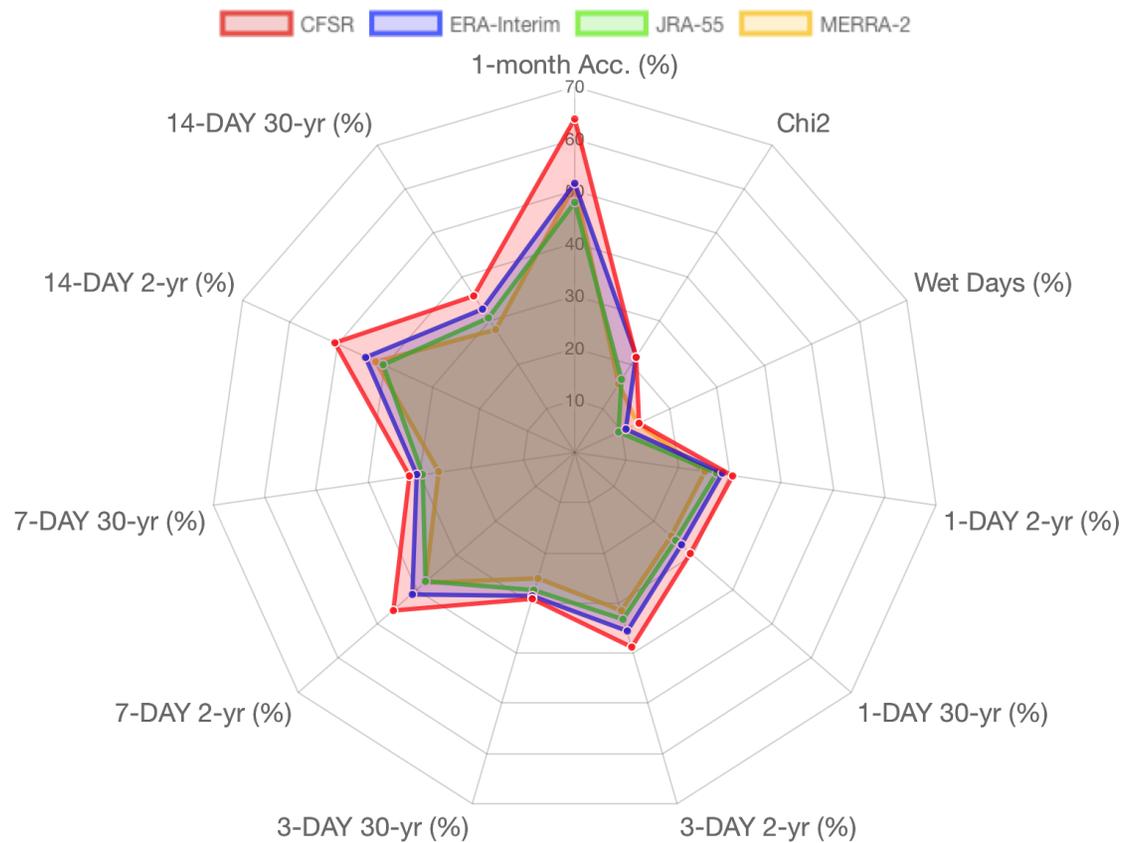


Figure 3.11: 30-year return level and systematic error of 1-day precipitation total, for all four reanalyses for North climate zone. The vertical dashed lines indicate the change in seasons and the colored dots represent the seasonal average.

3.5. Results and discussion

A Kruskal-Wallis analysis indicates significant differences in mean systematic error of daily and extreme PCP between the four reanalysis datasets at the $\alpha_{Walker} = 9.53 \times 10^{-3}$ level. After multiple comparisons by Nemenyi's test, for daily PCP the ERA-Interim, JRA-55 and MERRA-2 significantly outperform CFSR. For extreme PCP, the MERRA-2 and JRA-55 reanalyses significantly outperform ERA-Interim and CFSR.

All of these results are summarized in Fig. 3.12 where mean absolute error (MAE) of the reanalyses daily and extreme PCP are averaged over the entire study period and all stations. The closer the value is to 0 for a given reanalysis, the better its performance. MERRA-2 and JRA-55 are the better reanalyses, outperforming CFSR for all metrics, and to a lesser extent, ERA-Interim for daily PCP. This averaging also hides the greater variability in bias of the poorer performing reanalyses, which is harder to correct for. For extreme PCP, the difference between MERRA-2 and JRA-55, and ERA-Interim is more noticeable with the former two outperforming the latter. Additionally, the errors of 30-year return levels are smaller than those of 2-year return levels. Although MERRA-2 outperforms JRA-55 in most metrics on average, the differences are not significant.



86 Figure 3.12: MAE of 31-day precipitation totals, χ^2 -statistic, "wet days", and 2-year and 30-year return levels of 1-, 3-, 7-, and 14-day precipitation totals. MAE is averaged across all 66 stations. Values closer to 0 at the origin of the plot are better.

3.6 Stationarity

3.6.1 Daily PCP

Stationarity in daily and extreme PCP distributions is assessed to determine if temporal changes are significant. If there are significant temporal changes, that would mean a traditional, stationary distribution (based on the 1981-2010 period) would not be appropriate to represent the present day expected precipitation distribution. First, variations through time — for each calendar day at each station — for season precipitation totals, number of "wet" days, "very light", "light", "moderate", "heavy" and "extreme" precipitation events are modeled using linear regression to identify patterns.

Due to large year-to-year background variability in precipitation, the study period is divided into three decades (1981-1990, 1991-2000 and 2001-2010). A 91-day centered rolling window is used to obtain seasonal precipitation totals for each calendar day; one value per decade per calendar day, yielding 3 values for each calendar day. The seasonal precipitation total for a given calendar day is assumed to vary with time as $\mu(t) = \mu_0 + \mu_1(t - t_0)$, where the slope coefficient μ_1 represents the average change from one decade to the next (Fig. 3.13).

Over the same 91-day centered rolling window, the number of "wet" days is assumed to vary with time as $\mu(t) = \mu_0 + \mu_1(t - t_0)$, where μ_1 represents the 10-year rate of change of the number of wet days in the 91-day centered rolling window. Namely, the number of "wet" days is recorded per decade per calendar day, yielding 3 values for each calendar day. The delineation of "wet" days is lowered from 1.0 mm to trace because stationarity is being assessed on weather station data only (excessive occurrences of very light model precipitation are not an issue).

Finally, still over the same 91-day centered rolling window "wet" days are recorded into five non-overlapping intervals: [*Trace*, *50th*), [*50th*, *75th*), [*75th*, *90th*), [*90th*, *95th*) and [*95th*, *100th*]th for "very light", "light", "moderate", "heavy" and "extreme" precipitation events, respectively. The number of events in each bin is assumed to vary with time as $\mu(t) = \mu_0 + \mu_1(t - t_0)$, where the slope coefficient μ_1 represents the average change in the number of events from one decade to the next.

Second, confidence intervals (CI) (Sun et al., 2018a) are implemented to assess significant changes of 10-year means from one decade to the next of season precipitation totals, number of "wet" days, "very light", "light", "moderate", "heavy" and "extreme" precipitation events.

A sufficiently large change in means from one decade to the next indicates

3.6. Stationarity

there is a trend, and therefore nonstationarity is required to characterize daily PCP. Small changes suggest a simpler, stationary model is accurate enough to represent precipitation. This critical difference is determined by the $(1 - \alpha_{Walker}) \times 100\%$ CI due to multiple testing, where $\alpha_{Walker} = 4.37 \times 10^{-6}$ and $\alpha_0 = 0.10$. Namely, variations in the mean over successive decades are large enough to be considered statistically significant if they fall outside the lower and upper bounds of the $(1 - \alpha_{Walker})\%$ CI of the resulting difference between the means of the 10-year periods and the 30-year period (Fig. 3.13). The individual 90% CI is adopted instead of the 95% CI to reduce the probability of making a type II error — that is, to reduce the probability of failing to see that there is a trend.

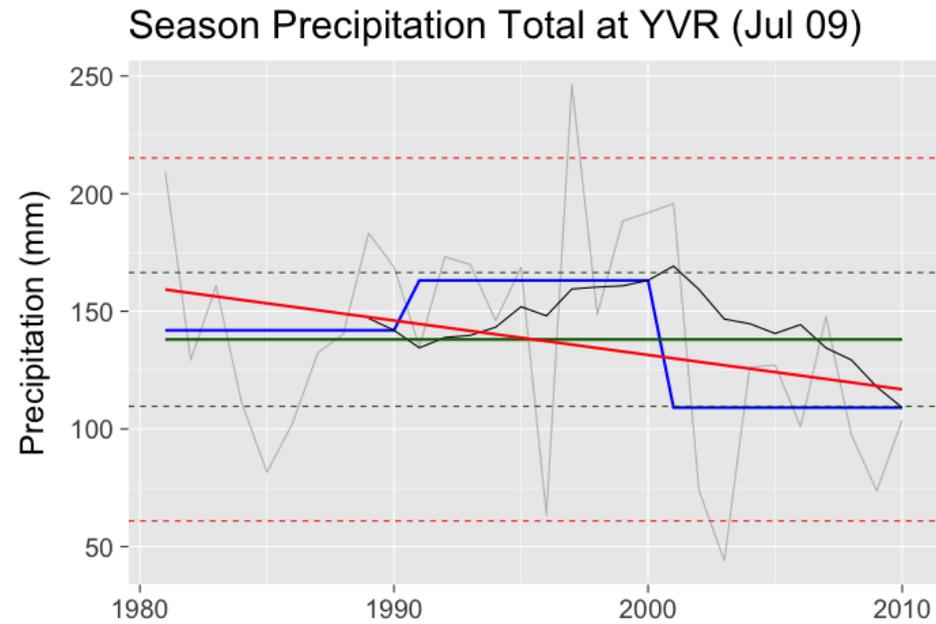


Figure 3.13: Seasonal precipitation total at Vancouver International Airport (YVR) for July 9th (grey), 10-year left moving average (black), decadal averages of seasonal precipitation total (1981-1990, 1991-2000 and 2001-2010) (blue) and 30-year season precipitation total averaged over the study period 1981-2010 (green). The red line shows the linear regression fit to the three decadal mean values. Individual 90% confidence intervals (CI; green dotted line) and multiple $(1 - \alpha_{Walker}) \times 100\%$ CI (red dotted line) are drawn; $\alpha_0 = 0.10$. This example indicates that a stationary distribution is appropriate for Jul 9th because the three decadal average values all fall within both CI.

3.6. Stationarity

Figure 3.14a indicates a noticeable drying trend of seasonal precipitation totals across southern BC during spring and summer (late summer in South Central zone), and a weaker wet trend during late summer, fall and early winter across most of BC. The weather stations are organized by climate zones with North on top, followed by Central, Northwest, and southern regions Maritime West, Maritime East, Southwest, South Central and Southeast at the bottom. Similarly, precipitation frequency (Fig. 3.14b) suggests an increase in the number of dry days during spring and summer across most of BC (particularly for the South Central zone in summer), and a weaker increase in the frequency of "wet" days during fall and winter across southern BC. None of the trends are significant at the multiple $(1 - \alpha_{Walker}) \times 100\%$ CI. At the individual 90% CI, the trends are generally not statistically significant, with exceptions in the Northwest and Maritime West climate zone during summer.

Furthermore, Figure 3.15a, b and c, indicates a weak increase in the number of "very light", "light" and "moderate" precipitation events across Maritime West, East and Southwest BC during fall (and winter for "very light" events), and a weak decrease in such events during spring and summer (substantially stronger for "very light" events in the South Central zone). Again they are not significant — none of the trends fall outside either the multiple $(1 - \alpha_{Walker}) \times 100\%$ CI or the individual 90% CI. "heavy" (Fig. 3.15d) and in particular "extreme" (not shown) precipitation events do not show any clear pattern.

3.6. Stationarity

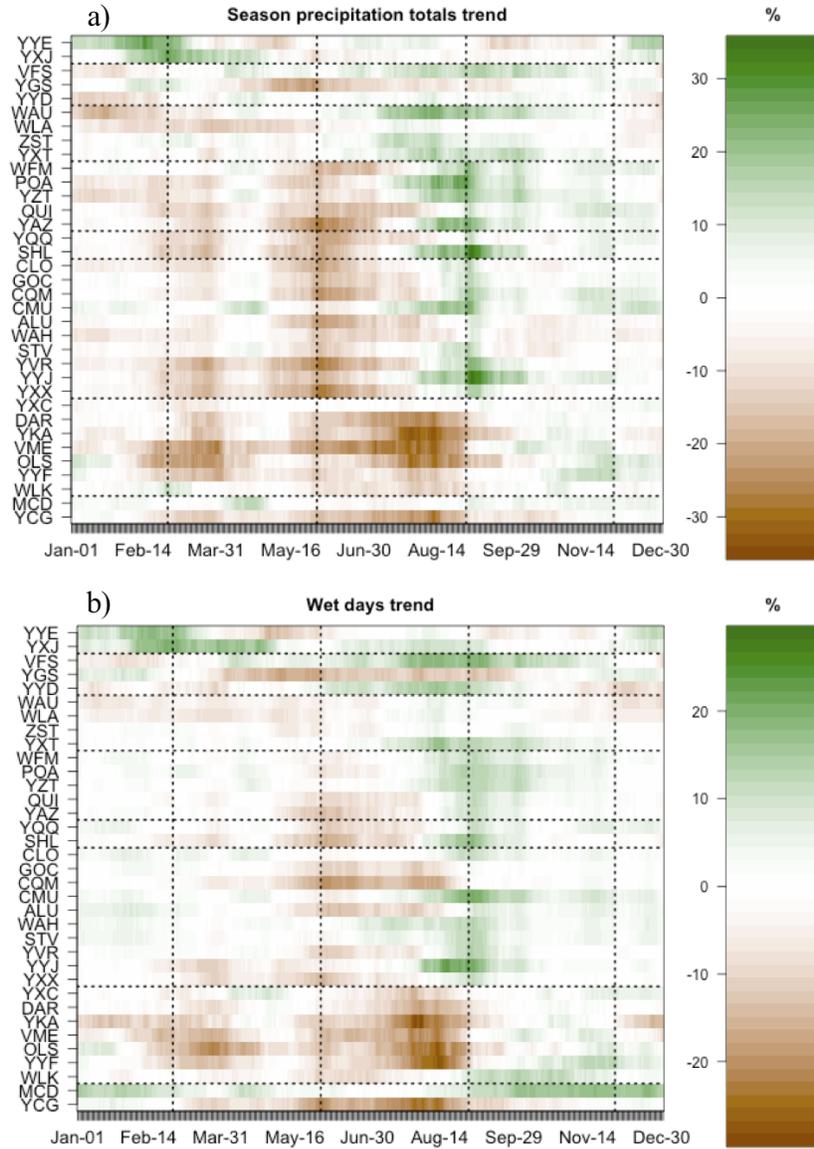


Figure 3.14: (a) Mean linear trend of season accumulated precipitation; (b) Mean linear trend of days with precipitation. The vertical dashed lines indicate the change in seasons and the horizontal dashed lines delineate from top to bottom the North, Central, Northwest, Maritime West, Maritime East, Southwest, South Central and Southeast climate zones respectively. Units are percent change over the 30-year period.

3.7. Conclusion

Despite some clear trends, none of the changes of 10-year means from one decade to the next are statistically significant at the multiple $(1 - \alpha_{Walker}) \times 100\%$ CI, and only a few isolated cases are statistically significant at the individual 90% CI. During the 1981-2010 study period, a stationary 30-year mean is accurate enough to represent mean values of season precipitation totals, number of "wet" days, "very light", "light", "heavy", "heavy" and "extreme" precipitation events.

3.6.2 Extreme PCP

For extreme PCP, the rolling 91-day centered rolling window is maintained for 1-, 3-, 7-, and 14-day precipitation totals. A GEV dresses the 31 annual maximum values for each calendar day by the method of maximum likelihood. A nonstationary GEV distribution is compared, where only the location parameter is allowed to exhibit trend, with a stationary GEV distribution with constant location, scale and shape parameters.

The GEV distribution location parameter is assumed to vary with time as $\mu(t) = \mu_0 + \mu_1(t - t_0)$, where the slope coefficient μ_1 represent the 10-year rate of change in location values of extreme 1-, 3-, 7-, and 14-day precipitation totals.

None of the variations in the average of annual maxima over successive decades fall outside either the $(1 - \alpha_{Walker}) \times 100\%$ CI, or the 90% CI. Furthermore, no clear trend in extreme PCP is discernible (not shown). During the 31 years of the study period, a stationary GEV distribution with model $M_0 = GEV(\mu, \sigma, \kappa)$ is accurate enough to represent 1-, 3-, 7-, and 14-day precipitation totals.

3.7 Conclusion

Reanalysis performance for daily and extreme precipitation (PCP) is evaluated across British Columbia (BC) during the 1980-2010 study period. To compare daily PCP among CFSR, ERA-Interim, JRA-55 and MERRA-2, the systematic error of 31-day precipitation total, wet days, and two-sample χ^2 -statistic are calculated. To identify performance of extreme PCP, the systematic error of 2- and 30-year return levels of 1-, 3-, 7- and 14- day accumulated precipitation are compared.

In reanalyses, precipitation is generally better represented over areas well-covered with accurate, complete, and coherent observations of all numerical forecast variables, which are used to correct the reanalyses. In relatively data-sparse areas such as BC (Hacker et al., 2003; Spagnol, 2005),

3.7. Conclusion

reanalysis precipitation relies mostly on the underlying model output rather than observations (Dee et al., 2011; Lindsay et al., 2014). Namely, despite model estimation of precipitation being based on temperature and humidity information derived from the assimilated observations, approximations used in the models representation of moist processes over data-sparse areas strongly affect the quality and consistency of the hydrological cycle (Dee et al., 2011).

The model generated-precipitation in CFSR traditionally has a wet bias (Saha et al., 2010). In this study, CFSR wasn't the best reanalysis in any of the metrics. Results suggest a wet bias across all climate zones in BC matching previous results which indicate a wet bias over the western Pacific and in mid-high latitudes (Saha et al., 2010; Wang et al., 2011).

JRA-55 has the most consistent, and among the smallest, systematic error throughout the year and across the different climate zones in BC. Previous studies concluded that the quality of the JRA-55 improved significantly when compared with that of JRA-25 (Kobayashi et al., 2015; Harada et al., 2016), especially in the Pacific Ocean north of 30° N (Harada, 2018).

ERA-Interim exhibits the largest variation in performance throughout the calendar year and across the different climate zones, with lower systematic errors across the drier climates zones than the wetter climate zones; capturing the wettest months in the dry climate zones, and missing the correct amount of precipitation during storm season in the wet climate zones. However, ERA-Interim performed fairly well for daily PCP across BC. Uppala et al. (2005) explains the various difficulties encountered in ERA-40 with the assimilation of humidity information, which led to a generally poor representation of the global transport of moisture in the atmosphere. According to our results, those problems seem to have been corrected on ERA-Interim. Additionally, Uppala et al. (2005) concludes that there was an improvement over previous generation reanalysis ERA-40 to ERA-Interim in precipitation over higher latitudes.

A previous MERRA study (Bosilovich et al., 2015) suggests that the sparse coverage of precipitation gauges in high latitudes may lead to significant biases. Studies have documented the difficulty of conserving atmospheric dry mass while guaranteeing that the net source of water from precipitation and surface evaporation equals the change in the total atmospheric water (Trenberth and Smith, 2005; Bosilovich et al., 2008; Berrisford et al., 2011). Reconsideration of these issues were taken into account during the development of MERRA-2. In this study, MERRA-2 performed well, particularly for extreme PCP.

In summary, JRA-55 and MERRA-2 better capture precipitation distri-

3.7. Conclusion

bution across BC all year, and have the lowest systematic errors across the wet climate zones during storm season. This makes them the better choices for a gridded climatological dataset of daily precipitation over BC. For daily PCP, MERRA-2 and JRA-55 are the better reanalyses followed closely by ERA-Interim. For extreme PCP, MERRA-2 and JRA-55 are the better reanalyses, with the lowest systematic errors throughout the year and across different climate zones.

According to Chapter 2 and this study, ERA-Interim performs better for daily and extreme 2-m temperature than it does for daily and extreme PCP, even though the two fields should influence one another. A possible explanation is that many reanalyses do not directly assimilate 2-m air temperature observations, whereas ERA-Interim does. In contrast, PCP is a model-produced field, influenced indirectly by surface and upper-air temperature and humidity observations (Dee et al., 2011; Lader et al., 2016).

There is a noticeable drying trend in precipitation total during spring and summer months across southern BC, and a wet trend during early fall for northern and southwestern BC. These patterns also manifest themselves in dry- and wet-day frequencies. The strongest signal is drying in the South Central zone in summer. These findings add more information to previous studies. Vincent and Mekis (2006) showed that the the number of days with precipitation per year also have significantly increased from 1950-2003 across BC, and Zhang et al. (2000) showed a distinct drying pattern in the southern regions of BC during summer and spring during the second half of the twentieth century. Finally, our analysis shows that spring and summer have been getting drier for much of BC, which is in line with future climate projections (Haughian et al., 2012).

The number of light and moderate precipitation events has generally increased during fall and winter months, and decreased mostly during spring and summer across BC. Despite the clear evidence of a dry trend for spring and summer months, and a wet trend during fall and winter months, the trends are not as discernible for precipitation intensity. Finally, there is no discernible pattern in changes in return levels of extreme PCP, or frequency of heavy and extreme precipitation events. A different study also showed no consistent trends in the number of precipitation extremes during the last century (Zhang et al., 2001). By contrast, Groisman et al. (2005) showed an increase in heavy and very heavy precipitation events south of 55°N from 1910-2001 across BC. The lack of consistency between periods and methodology for computing the trends have made it difficult to compare results across different studies. One possibility is that changes in precipitation were occurring too slowly to be discerned in the 31-year study period, given

3.7. Conclusion

the the considerable year-to-year variability in precipitation. Therefore, it is possible that with a longer record discernible trends may be found. For this 31-year study period, apart from some isolated cases, no statistically significant trends are found. Thus, a stationary distribution is sufficient to represent daily and extreme PCP over BC.

Chapter 2 shows that, across mountainous BC, ERA-Interim and JRA-55 are the most consistent and accurate reanalyses for daily and extreme temperature. This chapter shows that JRA-55 and MERRA-2 are the most consistent reanalyses for daily and extreme precipitation. More consistent biases are favoured, as they are more easily removed by bias correction. It is important to note that, as expected, the results for daily and extreme temperatures are more conclusive than the results for daily and extreme precipitation since models do not simulate precipitation as well as they do temperature (Kendon et al., 2014; Ravishankar et al., 2016), and have difficulties to represent extreme precipitation (Zhu et al., 2014). Due to higher variability in precipitation across BC, there is a large variation in performance—even for the better reanalyses JRA-55 and MERRA-2—across the different climate zones and seasons.

Furthermore, the longer JRA-55 record is advantageous in that the standard errors (of the estimated parameters used in extremes modelling and the resulting return levels) are expected to decrease as the sample size increases (Hosking et al., 1985; Hosking, 1990; Cai and Hames, 2010).

Based on these findings, and the temperature findings in Chapter 2, the JRA-55 is recommended as the most accurate reanalysis over BC. This chapter concludes that for daily PCP, the JRA-55 systematic error relative to weather station observations is highly correlated to JRA-55 systematic error relative to PRISM. It suggests that the biases can be explained by topographic and synoptic parameters—parameters that were implemented in the development of PRISM. In the next chapter the bias corrections based on error dependencies found in Chapter 2 and this study will be implemented in the JRA-55 to create an even more accurate gridded climatological dataset. This will then be used, in conjunction with a probabilistic forecast dataset to create an extreme temperature and precipitation forecast index.

Chapter 4

Analysis and Forecast of High-resolution Extreme Weather

4.1 Introduction

This Chapter has four objectives. First, the high-resolution PRISM climatology, the JRA-55 and the homogenized weather station dataset are combined to downscale and bias correct the JRA-55 and create a 30-arc-second ($\sim 800\text{ m}$) very-high-resolution surface analysis (VHRSA) of daily maximum and minimum 2-m temperature, and 1-day accumulated precipitation.

VHRSA temperature and precipitation datasets have the potential benefit of rendering a feasible solution to the paucity of observational data across BC due to its inherent spatial and temporal completeness. Additionally, such a dataset has the ability to resolve fine-scale topographic features that are important for a wide variety meteorological, climatological, and hydrological studies.

The accuracy of the VHRSA is compared to the JRA-55 — the best performing reanalyses across BC. Performance is assessed with respect to daily maximum and minimum 2-m temperature, and 1-day accumulated precipitation over mountainous BC (hereafter daily and extreme T2M as defined in section 2.3; daily and extreme PCP as defined in section 3.2).

Then, the VHRSA is used to statistically downscale and bias correct North American Ensemble Forecast System (NAEFS) forecasts, creating a very-high-resolution probabilistic forecast across BC.

Finally, trends in extreme T2M and PCP are examined to determine whether or not a stationary climatological distribution is appropriate to represent present-day extreme distributions. The stationarity results are used to inform the creation of a new extreme weather forecast index. The objective of the index to improve upon existing tools (that heighten operational awareness of potentially extreme events) by more accurately calculating the

extremity of the event, and producing fewer false alarms.

In section 4.2, a brief description of the different datasets and of the weather station observations are given. In section 4.3, the methodology to statistically downscale and bias correct the JRA-55 reanalysis, and the various metrics used to evaluate daily and extreme T2M and PCP are presented. In section 4.4, the methodology to statistically downscale and bias correct the NAEFS forecast is presented. The metrics used to evaluate daily T2M and PCP with respect to the raw NAEFS forecast are also presented. In section 4.5, the methods for assessing statistical nonstationarity are introduced, and trends of extreme T2M and PCP are examined. Section 4.6 introduces the Parametric Extreme Index (PEI) and compares it to Standardized Anomalies (SA). Results are summarized in the conclusion.

4.2 Data

Observational data were obtained from 62 surface weather stations for daily maximum and minimum T2M, and from 69 stations for daily accumulated PCP, for the period 1 Jan 1958 to 31 Dec 2017. These data are used in conjunction with the PRISM dataset to bias-correct and downscale the JRA-55, which in turn is used to bias-correct and downscale the North American Ensemble Forecast System (NAEFS) forecast.

A detailed description of the T2M and PCP weather station datasets are given in chapter 2 and 3, respectively. Information on the JRA-55 atmospheric models and configurations are presented in tables 2.1 and 3.1. A more detailed description of the JRA-55 can be found in Chapter 2 or in Ebita et al. (2009, 2011) and Takeuchi et al. (2013).

4.2.1 Weather station data

Of the 72 and 118 geographically-dispersed T2M and PCP stations initially selected for this study respectively, 10 T2M and 49 PCP stations with more than 10% missing data were excluded. The upper bound threshold was raised to 10% in this Chapter to strike a balance between having enough stations with a complete temporal record, while still having representative data in each climate zone in the province. A broad description of climate zones can be found in Chilton (1981) and Moore et al. (2008). A more detailed description and formal statistical derivation of T2M and PCP climate zones can be found in Chapters 2 and 3, respectively.

Figures 4.1 and 4.2 shows the locations of all Environment and Climate Change Canada (ECCC) and BC Hydro stations and their corresponding

4.2. Data

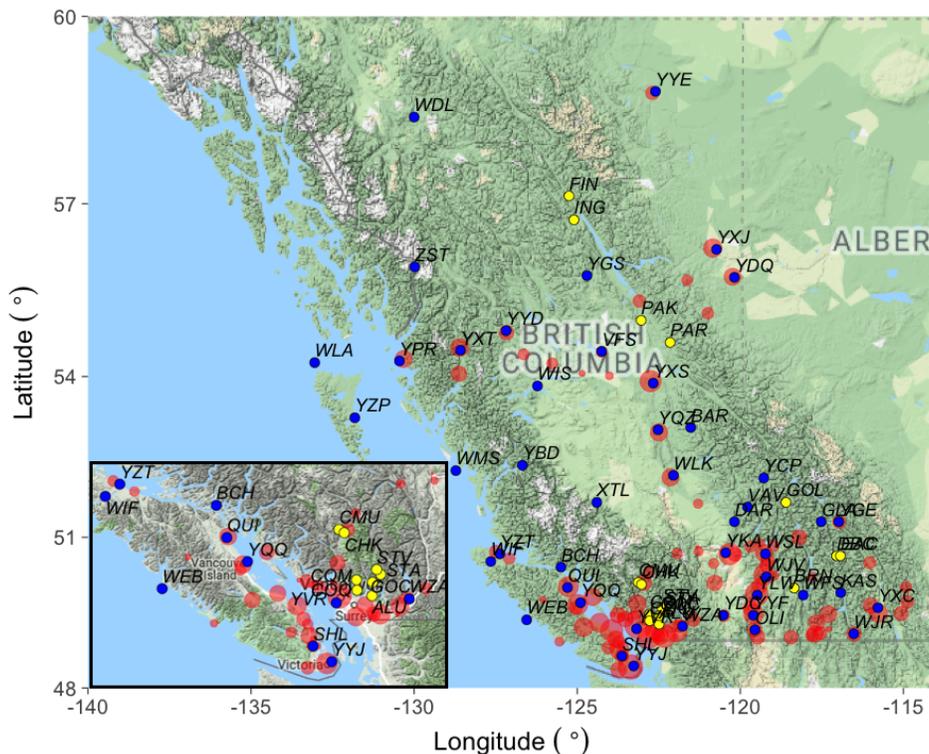


Figure 4.1: Location of ECCC (blue) and BC Hydro (yellow) T2M weather stations and British Columbia population distribution (red).

three-letter abbreviations, overlaid with the population distribution across BC (these maps are shown again because they have a different selection of stations from those shown in Chapters 2 and 3). Station elevation varies from sea-level to alpine, with most stations located in valleys with different orientations, slopes and elevations, and some stations located on mountain slopes with different slope angles (see Appendix A for more details on stations).

For nonstationary analysis, stations with more than 1% missing data are excluded, leaving eighteen ECCC T2M stations and eight ECCC PCP stations. BC Hydro stations were excluded due to apparently spurious temporal trends. A detailed discussion of station homogeneity can be found in Chapter 2 for T2M and Chapter 3 for PCP.

4.2. Data

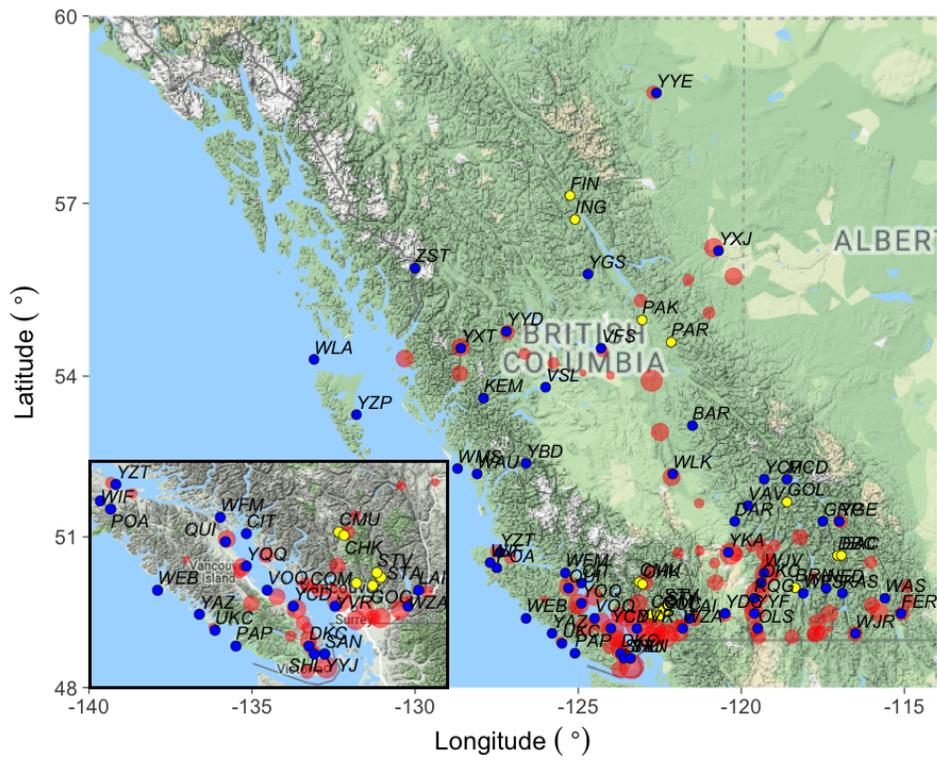


Figure 4.2: Location of ECCC (blue) and BC Hydro (yellow) PCP weather stations and British Columbia population distribution (red).

4.2.2 PRISM

In this study, gridded daily T2M and PCP climatology for the 1981-2010 climate normal period are used to downscale and improve the accuracy of the JRA-55 by identifying high-resolution regional and terrain biases in the JRA-55 daily T2M and PCP. A more detailed description of the PRISM can be found in Chapter 3 or in Daly et al. (1994, 1997, 2002).

4.2.3 NAEFS

The North American Ensemble Forecast System (NAEFS) is the amalgamation of the Meteorological Service of Canada (MSC) Global Ensemble Prediction System (GEPS) (Charron et al., 2010) and the United States National Centers for Environmental Prediction (NCEP) Global Ensemble Forecast System (GEFS) (Toth and Kalnay, 1997). The GEPS and GEFS are each comprised of 20 perturbed and 1 control members. Both ensemble systems employ ensemble Kalman filtering to generate perturbed initial conditions, and perturbed physical parameterizations (Wang et al., 2013; Hou et al., 2015; Wei et al., 2008; Zhou et al., 2017). When combined, the NAEFS provides a 42-member forecasts out to 16 days that is of higher quality than either ensemble alone (Zhu et al., 2014). The available archive of NAEFS data was downloaded at a horizontal grid spacing of $1.0^\circ \times 1.0^\circ$ every 6 hours (0000, 0600, 1200 and 1800 UTC) out to 14 days.

4.3 The VHRSA: Downscaling and bias-correcting the JRA-55

The statistical behaviour of extreme T2M and PCP are compared between observed weather station data, and the JRA-55 and PRISM interpolated to the stations locations. As in Chapters 2 and 3, the Inverse Distance Weighting (IDW) is used to interpolate reanalysis and PRISM output. A more detailed description of each interpolation method can be found in Mooney et al. (2011) and Stahl et al. (2006).

For JRA-55, precipitation accumulation intervals are summed over the 0601UTC-0600UTC window to get daily PCP (Table 3.1). The daily maximum (minimum) T2M is defined as the highest (lowest) value of the six-hourly T2M outputs in the same window (Table 2.1).

The highest (lowest) value of daily maximum (minimum) T2M within a 31-day centered rolling window is selected for each calendar day, for each year in the 60-year dataset. Thus, each calendar day has 60 values of annual

maximum (minimum) T2M and PCP. This is done for both the station data and JRA-55 interpolated to the stations.

A Generalized Extreme Value distribution is fitted over the 60 sample values for each calendar day by the method of L moments. As in previous chapter, estimates of return levels derived by the method of L moments are more reliable than the method of maximum likelihood (Hosking et al., 1985; Hosking, 1990).

A Lilliefors test is conducted to evaluate the goodness-of-fit test of the fitted GEV distribution to the observed data. The null hypothesis that the observed data is drawn from a GEV distribution is rejected in favour of the alternate hypothesis if a large difference between the fitted GEV cumulative distribution function (CDF) and the observation empirical cumulative distribution function (ECDF) is detected.

This large critical difference is determined by a parametric bootstrap procedure. Namely, 100 samples of size 60 are generated from the fitted GEV distribution for each calendar day at each station. Next, 100 critical differences are computed from the comparison of each generated sample ECDF and the fitted GEV CDF. Accepting or rejecting the null hypothesis at the 10% significance level is equivalent to comparing the 90th percentile of the resulting collections of critical differences with the actual difference between the fitted GEV CDF and the observation ECDF. At this 10% significance level, 10% of the tests are expected to exceed a critical difference, rejecting the null hypothesis. However, to reduce the probability of incorrectly rejecting one or more of the true null hypotheses due to multiple testing, the global 10% significance level is regarded as significant. Namely, the null hypothesis is rejected when the critical difference between the fitted GEV CDF and the observations ECDF exceeds the global 90th percentile of the resulting collections of critical differences between the fitted GEV CDF and the generated ECDF of the samples (Wilks, 2016).

At the global 10% significance level, less than 8% of the stations and calendar days are rejected during the 1958-2017 study period. It indicates extreme T2M and PCP values can be described by a GEV distribution. The 2-year return level is then estimated from the GEV distribution for each calendar day and station.

Next, the monthly mean PRISM values are assumed to be valid on the 15th of each month. The twelve monthly values are linearly interpolated in time to obtain 365 calendar day values of monthly means (example shown for maximum T2M, Fig. 4.3a).

For the JRA-55, the same 31-day centered rolling window is used to accumulate PCP values and obtain monthly precipitation for each calendar

4.3. The VHRSA: Downscaling and bias-correcting the JRA-55

day. These 31-day accumulated values are then averaged over the 30-year PRISM climate period (1981-2010). For T2M, the 31-day window is used to obtain monthly mean values of daily maximum and minimum T2M for each calendar day, which are also averaged over the 30-year PRISM climate period.

Then, for each calendar day and variable, a linear regression summarizes the relationship between the bias of the JRA-55 2-year return level value relative to the station 2-year return level value, and the bias of the monthly mean JRA-55 value relative to the monthly mean PRISM value (additive bias for T2M, multiplicative bias for PCP). This yields 365 regression models and 730 parameters — two parameters per calendar day (the slope and y-intercept; Fig. 4.3b). The 2-year return level is chosen because it is the median of a GEV distribution and therefore a robust representation of the centre of the data. Two-thirds of the stations are randomly selected to train the linear regression model and the remaining 1/3 of the stations are used to test it (see Appendix B for a list of train and test stations). This cross validation is needed to test how well the model will generalize to grid points throughout BC.

Finally, for each calendar day, the JRA-55 is bilinearly interpolated (Fig. 4.3e) to the PRISM grid (Fig. 4.3a) and the monthly bias between the JRA-55 and the PRISM is calculated (additive bias for T2M, multiplicative bias for PCP; Fig. 4.3c).

The VHRSA is generated by starting with the JRA-55 field (Fig. 4.3d), and bias correcting it applying the linear regression equation (Fig. ??b) to the monthly mean bias (Fig. 4.3c). The final result is the VHRSA field (Fig. 4.3f). That is, the VHRSA value is computed by subtracting (multiplying for PCP) the downscaling difference for T2M (ratio for PCP) from the JRA-55 value.

The downscaling differences for T2M in Figure 4.3c are typically less than 0°C in valleys and greater than 0°C in mountainous regions, leading to a dataset with larger vertical temperature differences. Similarly, for PCP the downscaling ratio is typically greater than 1 in valleys and less than 1 in mountainous regions, leading to a dataset that is drier in valleys and wetter in ridges and upper elevation regions (see Figure 3.9 and subsection 3.5.1 for more details).

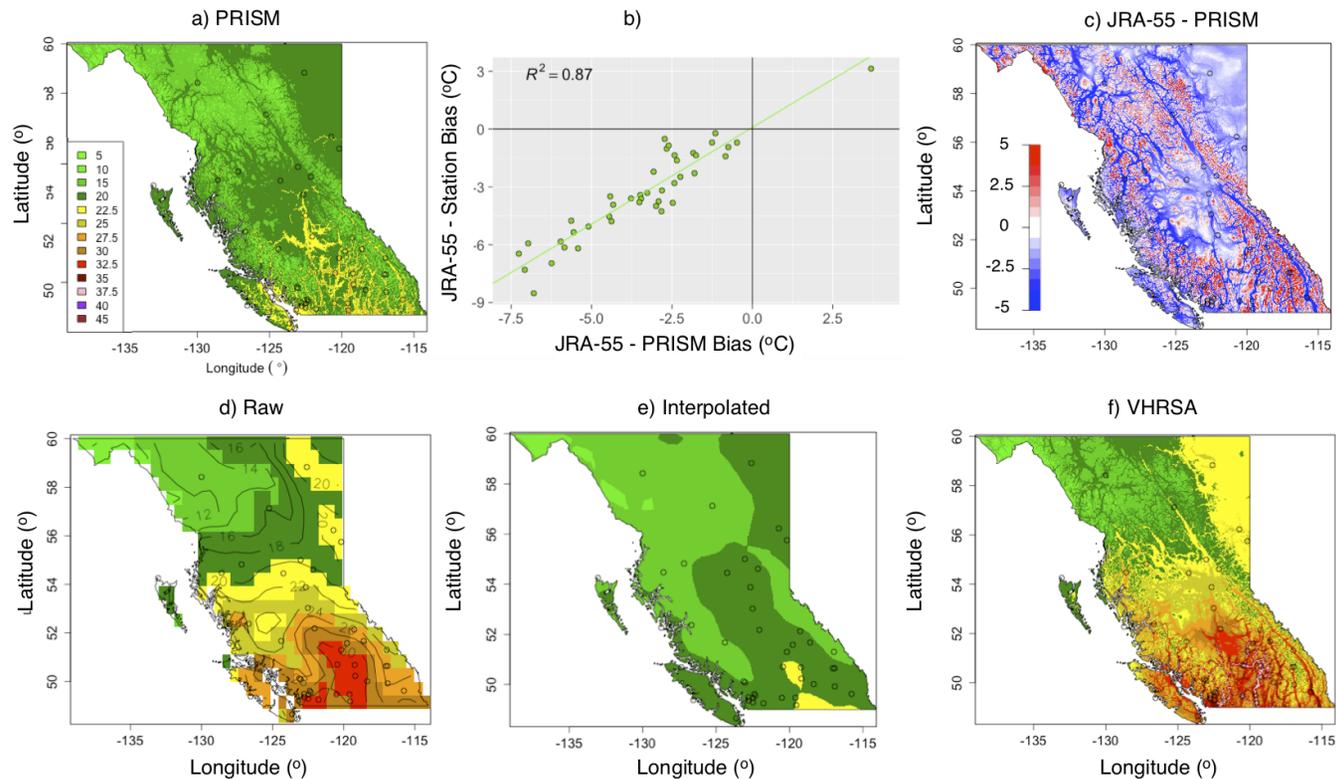


Figure 4.3: (a) 30-year monthly mean of daily maximum T2M PRISM climate on August 28th. (b) 2-year return level systematic error between training stations and reanalysis interpolated to training stations as a function of monthly mean daily maximum T2M bias between JRA-55 and PRISM interpolated to training stations. The solid line shows the linear regression fit. (c) Downscaling difference derived by subtracting (a) from (e). (d) JRA-55 daily maximum T2M on 28 Aug 2017. (e) Monthly mean of daily maximum T2M JRA-55 climate bilinearly interpolated to PRISM grid. (f) VHRSA daily maximum T2M on 28 Aug 2017.

To evaluate whether the VHRSA is statistically better as a gridded climatological dataset than the JRA-55, the monthly mean bias (or systematic error) and mean absolute error (MAE; or random error) are computed to estimate how accurately each dataset captures T2M and PCP. The bias measures the average difference between the datasets and the observation. The MAE measures the mean of the absolute differences between the datasets and observations. The two-sample Kolmogorov-Smirnov (KS) and χ^2 -statistics are computed to estimate how accurately each dataset captures the distribution of T2M and PCP respectively. The two-sample KS statistic determines the largest absolute difference between the gridded dataset values of the T2M ECDF, and the observed ECDF values. The χ^2 -statistic measures the difference between the number of “very light” ([1.0 mm, 50th]), “moderate” ([50th, 75th]), “moderate” ([75th, 90th]), “heavy” ([90th, 95th]) and “extreme” ([95th, 100th]) PCP events in each dataset.

Finally, the bias and MAE of the 2-year return levels are calculated to estimate how well each gridded analysis captures extreme T2M and PCP. For more details on these metrics, refer to sections 2.4 and 3.4.

Given that the test stations belong to different climate zones, the median is used to average the monthly mean and 2-year return levels biases, MAE, KS and χ^2 -statistics across all test stations on each calendar day (Figures 4.4, 4.5 and 4.6).

4.3.1 Verification of the VHRSA Daily and Extreme T2M

For daily and extreme maximum T2M, the seasonal cycle of temperature is very well captured and maintained over the VHRSA (Fig. 4.4a,b). Additionally, the VHRSA clearly outperforms the JRA-55 throughout the entire calendar year for both bias and MAE of daily and extreme T2M, and KS-statistic (Fig. 4.4c-g). The cold bias in the JRA-55 has mostly been removed for both daily and extreme T2M (Fig. 4.4c,d), and the VHRSA better captures the observations ECDF (Fig. 4.4g). The coloured points represent the seasonal error of each metric. They indicate that the errors in the VHRSA are consistently substantially smaller than those of the JRA-55 across all metrics evaluated.

The results for daily and extreme minimum T2M are nearly as good as those for daily maximum T2M. Both the JRA-55 and the VHRSA capture the seasonal cycle of daily and extreme minimum T2M well (Fig. 4.5a,b). The JRA-55 exhibits high variability in bias throughout the seasons for both mean and 2-year return levels of daily minimum T2M (Fig. 4.5c,d). The cold bias from mid-spring to mid-fall, and the warm bias during winter have

4.3. The VHRSA: Downscaling and bias-correcting the JRA-55

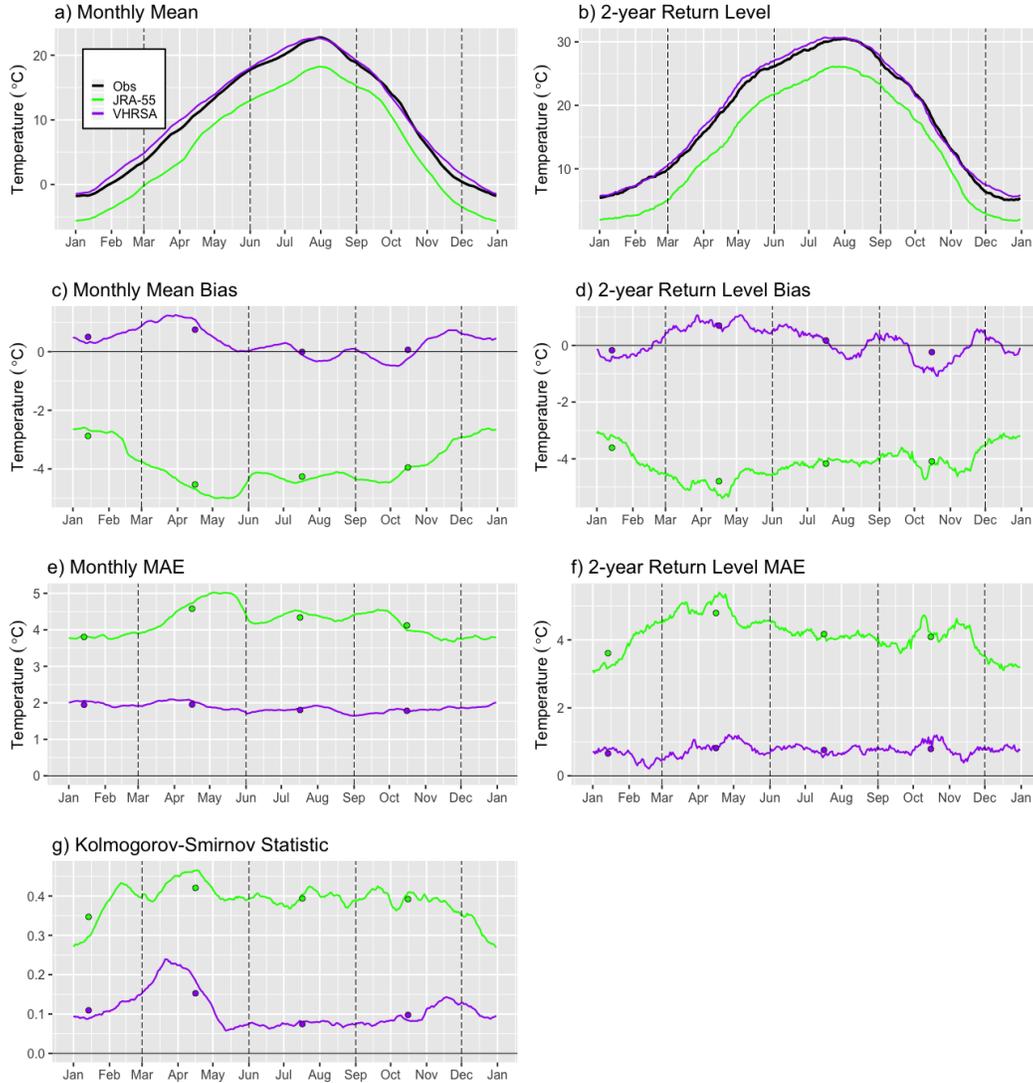


Figure 4.4: (a) Observed, JRA-55 and VHRSA monthly mean daily maximum T2M averaged over test stations. (b) 2-year return levels of observed, JRA-55 and VHRSA maximum T2M averaged over test stations. (c) Monthly mean bias of JRA-55 and VHRSA daily maximum T2M averaged over test stations. (d) As in (c) but for 2-year return levels. (e) As in (c) but for MAE. (f) As in (e) but for 2-year return levels. (g) KS statistic for daily maximum T2M. The vertical dashed lines indicate the change in seasons and the coloured dots represent the errors seasonal averages. In (a) and (b), values closer to observations (black line) are better. In (c)-(g), values closer to zero are better.

been reduced in the VHRSA. The VHRSA consistently outperforms the JRA-55 for daily MAE, averaging about 2° C (Fig. 4.5e,f), similar to that of the VHRSA maximum T2M. Since the VHRSA is corrected relative to 2-year return values, the MAE is even lower here, around 1° C annually — again similar to that of the VHRSA maximum T2M. Finally, the VHRSA better captures the observed ECDF throughout the entire calendar year (Fig. 4.5g).

4.3.2 Verification of the VHRSA Daily and Extreme PCP

The results for PCP, while not quite as good as those for T2M, still generally show substantial improvements after statistical downscaling and bias correction. The seasonal cycle is well captured and the wet bias is reduced over the VHRSA (Fig. 4.6a,b). The VHRSA consistently outperforms the JRA-55 throughout the seasons for monthly bias and MAE (Fig 4.6c,e). During the most important part of the year, storm season, the VHRSA better captures the observations distribution of precipitation (Fig 4.6g). Although the bias in extreme PCP is not improved (Fig. 4.6d), the VHRSA consistently outperforms the JRA-55 for extreme PCP MAE (Fig. 4.6f). One possible explanation is that the biases across BC are cancelling each other resulting in a smaller bias but not a smaller MAE.

The monthly mean systematic errors and MAE of daily maximum and minimum T2M and PCP, and the 2-year return level systematic errors and MAE are calculated from all calendar day systematic errors (Table 4.1). Similarly, the mean KS statistics and χ^2 -statistic are computed for daily maximum and minimum T2M, and PCP respectively. Comparisons between the fifteen mean systematic errors of the JRA-55 and the VHRSA are made using fifteen independent Mann-Whitney nonparametric tests (Hollander et al., 2013). Fifteen independent Mann-Whitney tests are used because the MAE is inherently skewed as it is left-bounded at 0, and due to the differences in magnitude and variability of each type of error.

A sufficiently large difference between the fifteen mean systematic errors of the JRA-55 and the VHRSA indicate that daily and extreme T2M and PCP are more accurate in the VHRSA than in the JRA-55. Small differences in the mean systematic errors suggest that the datasets have similar accuracy. Due to the fact that there are fifteen error pair comparisons between the JRA-55 and the VHRSA, the null hypothesis that the datasets have similar accuracy is rejected at the $\alpha_{Walker} = 1 - (1 - \alpha_0)^{1/N_0} = 0.0034$ level of significance, where $\alpha_0 = 0.05$ and $N_0 = 15$.

The Mann-Whitney tests indicate that all errors of daily and extreme

4.3. The VHRSA: Downscaling and bias-correcting the JRA-55

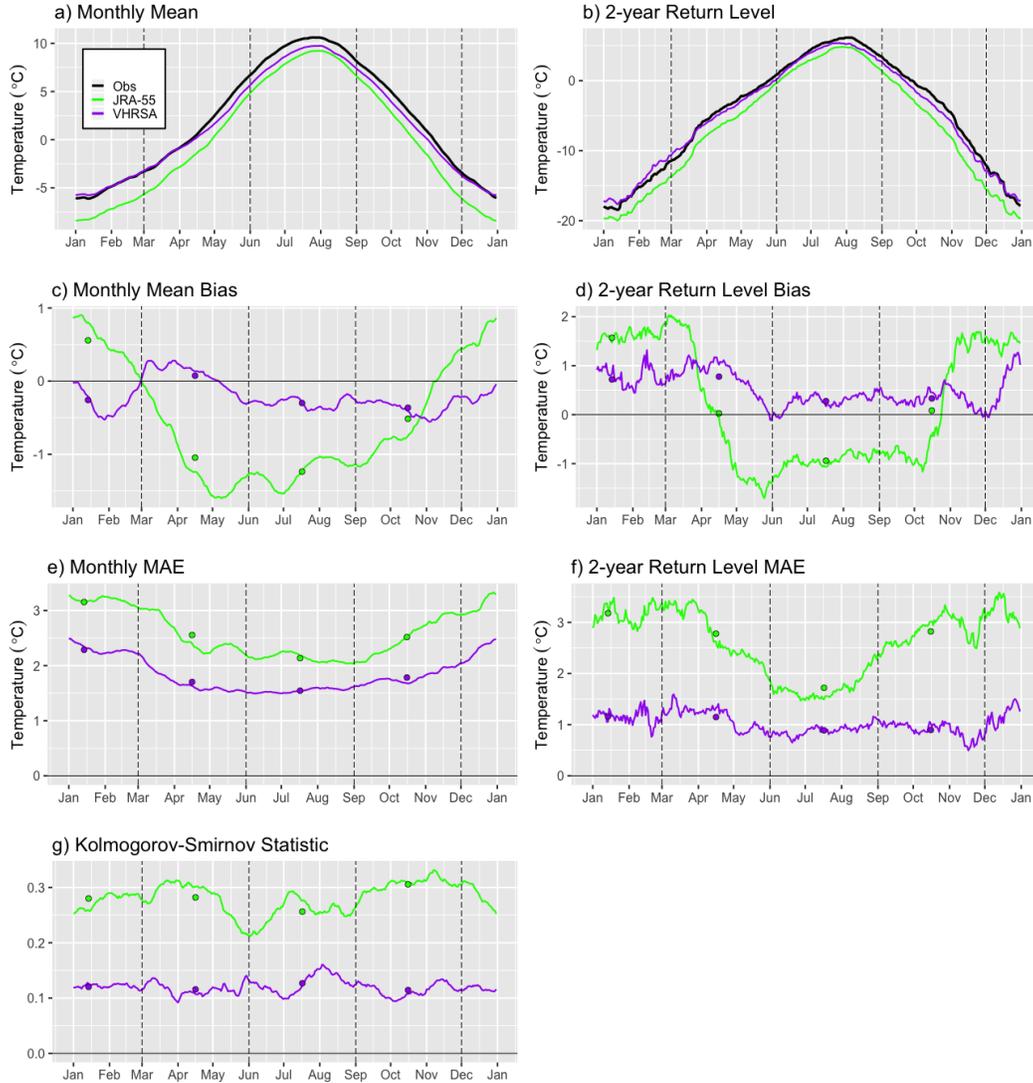


Figure 4.5: (a) Observed, JRA-55 and VHRSA monthly mean daily minimum T2M averaged over test stations. (b) 2-year return levels of observed, JRA-55 and VHRSA minimum T2M averaged over test stations. (c) Monthly mean bias of JRA-55 and VHRSA daily minimum T2M averaged over test stations. (d) Same as in (c) but for 2-year return levels. (e) Same as in (c) but for MAE. (f) Same as in (e) but for 2-year return levels. (g) KS statistic for daily minimum T2M. The vertical dashed lines indicate the change in seasons and the coloured dots represent the errors seasonal averages. Values of bias, MAE, and KS Statistic closer to zero are better.

4.3. The VHRSA: Downscaling and bias-correcting the JRA-55

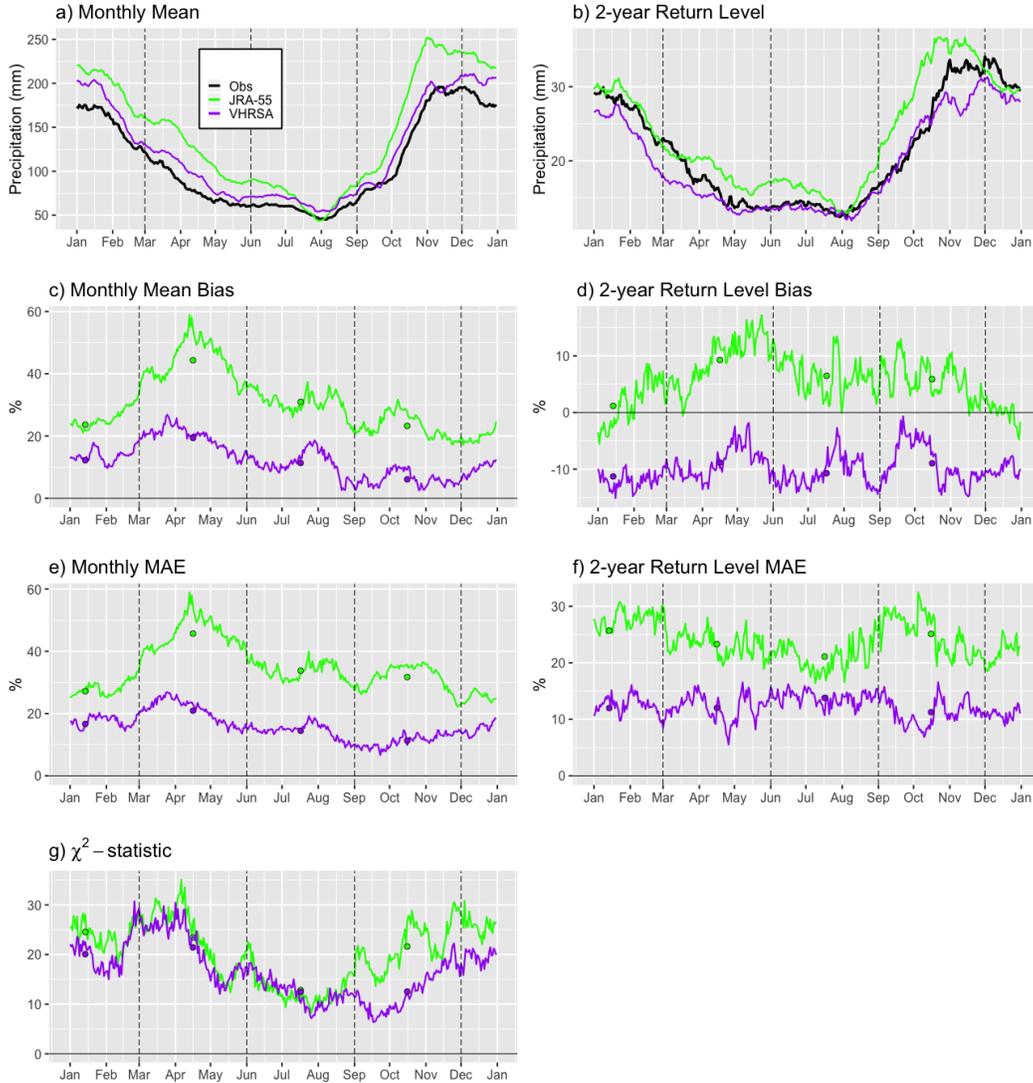


Figure 4.6: (a) Observed, JRA-55 and VHRSA mean of monthly PCP totals averaged over test stations. b) 2-year return levels of observed, JRA-55 and VHRSA PCP averaged over test stations. (c) Monthly mean bias of bias-corrected JRA-55 PCP averaged over test stations. (d) Same as in (c) but for 2-year return levels. (e) Same as in (c) but for MAE. (f) Same as in (e) but for 2-year return levels. (g) χ^2 -statistic of PCP. The vertical dashed lines indicate the change in seasons and the coloured dots represent the errors seasonal averages. Values of bias, MAE, and χ^2 -statistic closer to zero are better.

4.3. The VHRSA: Downscaling and bias-correcting the JRA-55

Table 4.1: Averaged systematic error and MAE of monthly precipitation total and monthly mean daily maximum and minimum T2M (MSE and MMAE respectively), of 2-year return levels (SE2 and MAE2 respectively), averaged systematic error of two-sample χ^2 -statistic (χ^2) and KS statistics, across all test stations (all T2M errors but KS statistic in °C; all PCP errors but two-sample χ^2 -statistic in %)

	MSE	MMAE	SE2	MAE2	KS	χ^2
T2M MAX						
VHRSA	0.32	1.87	0.12	0.76	0.11	
JRA-55	-3.91	4.21	-4.17	4.17	0.39	
T2M MIN						
VHRSA	-0.21	1.83	0.53	1.02	0.12	
JRA-55	-0.57	2.59	0.18	2.62	0.28	
PCP						
VHRSA	12.26	15.84	-9.94	12.26		16.62
JRA-55	30.59	34.63	5.72	23.79		20.58

T2M and PCP of the VHRSA are smaller than those of the JRA-55 at the global 5% significance level, with the exception of extreme PCP bias as shown in Figure 4.6d).

In general, the results of maximum T2M are better than those of minimum T2M, which in turn are better than the results of PCP. One possible explanation is the coefficient of determination R^2 . The R^2 is a standard measure of the goodness-of-fit of a regression. It can be interpreted as the proportion of the variation of the bias of the JRA-55 2-year return level value relative to the station 2-year return level value that is described or accounted for by the regression (Fig. 4.3b). Higher R^2 values are obtained for maximum T2M ($0.62 \leq R^2 \leq 0.91$) than for minimum T2M, ($0.69 \leq R^2 \leq 0.85$) which in turn are higher than the R^2 values of PCP ($0.58 \leq R^2 \leq 0.81$) throughout the year.

Finally, looking at the different climate zones across all variables, results are very similar, there is always a clear improvement in the VHRSA (not shown). For instance, the large cold systematic errors of daily and extreme maximum T2M across the Southeast climate zone found in Chapter 2 (Tables 2.2 and 2.3) have been removed, and the large wet biases in systematic errors of daily PCP across the drier climate zones found in Chapter 3 (North,

South Central, Central and Southeast; Table 3.2) have been reduced.

4.4 Bias-corrected NAEFS

A rolling window of the 30 most recent values (as would be used in real-time forecasting) is used to bias correct and downscale each member’s daily maximum and minimum T2M, and 1-day accumulated PCP forecast towards the VHRSA. Given that the daily maximum and minimum T2M were not available across in the NAEFS archive used, the daily maximum (minimum) T2M is defined as the highest (lowest) value of the six-hourly T2M outputs (for the calendar day 0601-0600 UTC). The 1-day accumulated PCP is summed over the four six-hourly outputs for each day.

The NAEFS forecast F_t is bias corrected using an additive (multiplicative) degree-of-mass-balance bias-correction factor DMB_t for T2M (PCP) (Grubišić et al., 2005; McCollor and Stull, 2008a; Bourdin et al., 2014). First, the raw NAEFS forecast (~ 100 -km horizontal grid spacing) is bilinearly interpolated to the VHRSA grid (~ 800 -m grid spacing). Then, the bias-correction factor is calculated and updated daily, and applied to the raw interpolated NAEFS forecast to generate a bias-corrected, downscaled NAEFS forecast \hat{F}_t . The factor DMB_t is a combination of the previous-day DMB_{t-1} and the difference (ratio for PCP) of the previous day forecast-observation pair ($F_{t-1} - O_{t-1}$ for T2M; F_{t-1}/O_{t-1} for PCP), weighted by a time parameter $\tau = 30$:

$$DMB_t = \frac{\tau - 1}{\tau} DMB_{t-1} + \frac{1}{\tau} (F_{t-1} - O_{t-1}). \quad (4.1)$$

The factor functions as a 30-day exponentially decaying rolling window. Namely, the influence of the forecast-observation pairs decreases with an e-folding time of $\tau = 30$ days from most recent to least recent. The value of $\tau = 30$ is chosen because McCollor and Stull (2008a) have shown it has the optimal minimum bias between the bias-corrected forecast and the observation for T2M and PCP forecasts over BC. The bias correction

$$\hat{F}_t = F_t - DMB_t \quad (4.2)$$

is done separately for each NAEFS member and then averaged to generate the bias-corrected ensemble mean ($\hat{F}_t = F_t/DMB_t$ for PCP; Figures 4.7, 4.8 and 4.9).

The degree-of-mass-balance method is chosen for three reasons. (1) Due to the 30-arc-second (~ 800 m) grid resolution of the VHRSA, an algorithm

is needed to improve the forecast in an operational setting (with associated time constraints) without exceeding computing resources. (2) This bias correction has been well studied, used and proven to work well. (3) It efficiently updates daily and retains a memory without creating an ever-growing training dataset.

Typically, operational raw ensemble forecasts are found to exhibit a low ensemble spread (Buizza, 1997; Hamill, 2001; Hamill et al., 2008; McCol-
 lator and Stull, 2008b), which in turn leads to overconfidence in probability
 assessment. Namely, ensemble forecasts are underdispersive and produce
 uncalibrated probabilistic forecasts (that is, the empirically derived proba-
 bilities are not accurate). A Nonhomogeneous Gaussian regression (NGR) is
 used to calibrate the forecast \hat{F}_t at the stations for T2M and PCP (Gneiting
 et al., 2005; Hagedorn et al., 2008). The NGR assumes the forecast errors
 form of a Gaussian distribution dressed about the ensemble mean:

$$N_t(\bar{F}_t, s_{ens}^2). \tag{4.3}$$

Here, \bar{F}_t is the bias-corrected ensemble mean calculated from 4.1 and
 4.2, where each member is equally weighted. The variance of the residuals
 — which are assumed to be Gaussian distributed— is computed as a linear
 function of the ensemble variance s_{ens}^2 . The regression parameters are chosen
 to minimize the continuous ranked probability score (CRPS).

Previous studies have reported good results calibrating surface temper-
 ature error distributions (which are approximately Gaussian) using NGR
 (Hamill et al., 2008; Hagedorn et al., 2008). Thorarinsdottir and Gneiting
 (2010) extend NGR to handle non-Gaussian distributed variables such as
 wind speeds. Once again, this method has been chosen to strike a balance
 between computing demands and well reported results.

The bias-corrected, calibrated NAEFS is evaluated for daily PCP and
 minimum T2M during fall (defined here as October, November and De-
 cember) and winter (defined here as December, January and February) of
 2016/17, which were abnormally wet and cold seasons (e.g., Figs. 4.7 and
 4.8); and daily maximum T2M during summer of 2017 (defined here as June,
 July and August), which was abnormally hot (e.g., Fig. 4.9). For more de-
 tails on the severity and impacts refer to subsection 1.1. The month prior
 to each season is used to spin up the bias correction and calibration of the
 raw forecast. Again, a 30-day training period is chosen to strike a balance
 between computing demands and reported results. According to McCol-
 lator and Stull (2008a), the DMB requires a 40 day training period for precipi-
 tation and a shorter 14 day training period for temperature to accomplish

4.4. *Bias-corrected NAEFS*

bias reduction. Hagedorn et al. (2008) and Hamill et al. (2008) suggest NGR can calibrate 2-m temperature and lighter precipitation events with a 30-day training dataset. A total of 3588, 4140, 4416 observations are used to evaluate forecast performance during fall, winter and summer respectively.

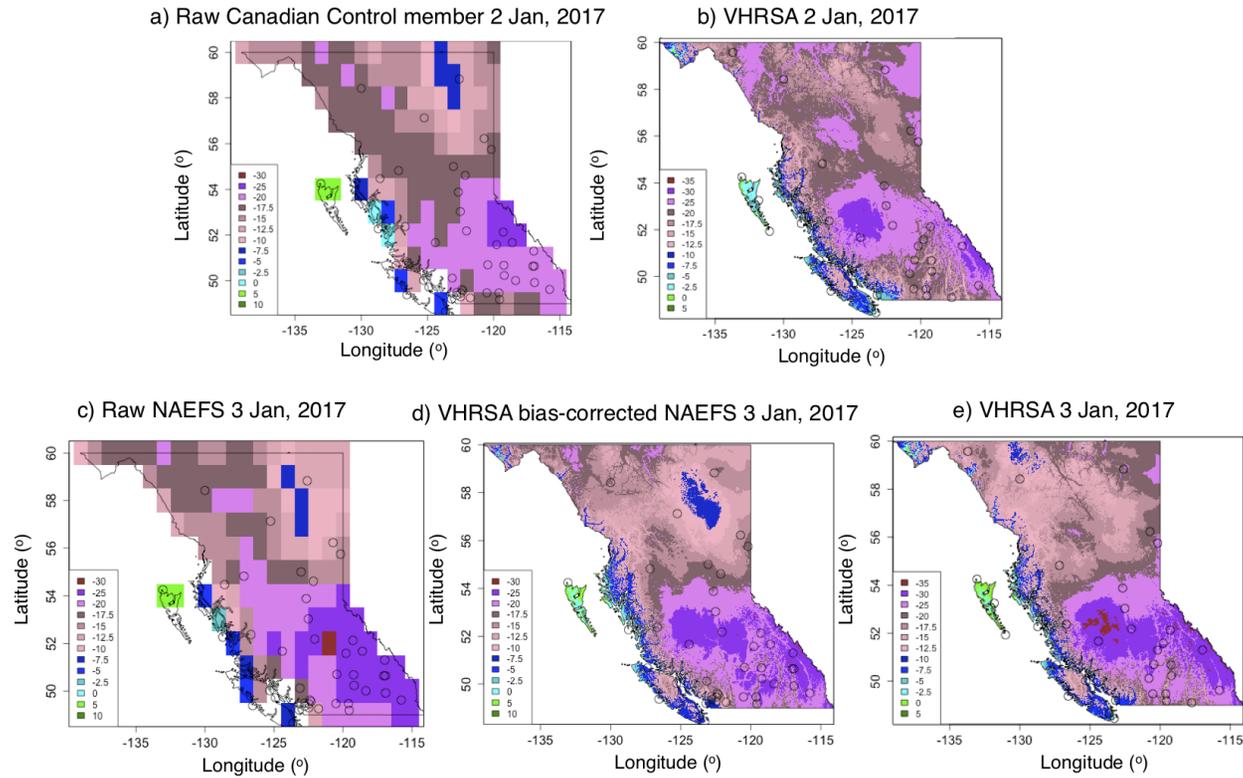


Figure 4.7: (a) NAEFS raw Canadian control member forecast of daily minimum T2M on previous day 2 Jan, 2017; (b) Same as in (a) but for VHRSA; (c) Ensemble mean raw NAEFS forecast on current day 3 Jan, 2017; (d) Ensemble mean bias-corrected NAEFS forecast on current day 3 Jan, 2017. DMB method done separately for each NAEFS member and then averaged to generate the bias-corrected ensemble mean. (e) VHRSA daily minimum T2M on 3 Jan, 2017 presented for comparison with (c) and (d). The dots represent weather station location.

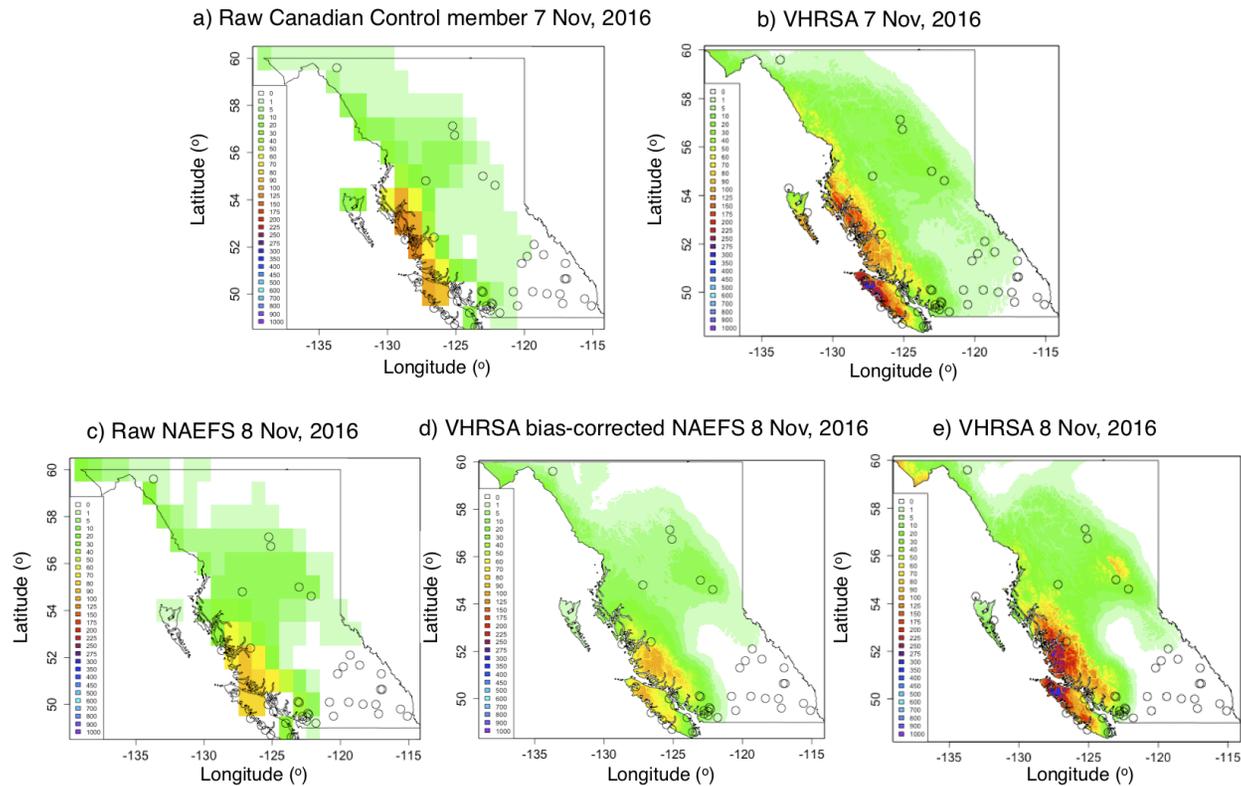


Figure 4.8: (a) NAEFS raw Canadian control member forecast of 1-day accumulated PCP on previous day 7 Nov, 2016; (b) Same as in (a) but for VHRSA; (c) Ensemble mean raw NAEFS forecast on current day 8 Nov, 2016; (d) Ensemble mean bias-corrected NAEFS forecast on current day 8 Nov, 2016. DMB method done separately for each NAEFS member and then averaged to generate the bias-corrected ensemble mean. (e) VHRSA 1-day accumulated PCP on 8 Nov, 2016 presented for comparison with (c) and (d). The dots represent weather station location.

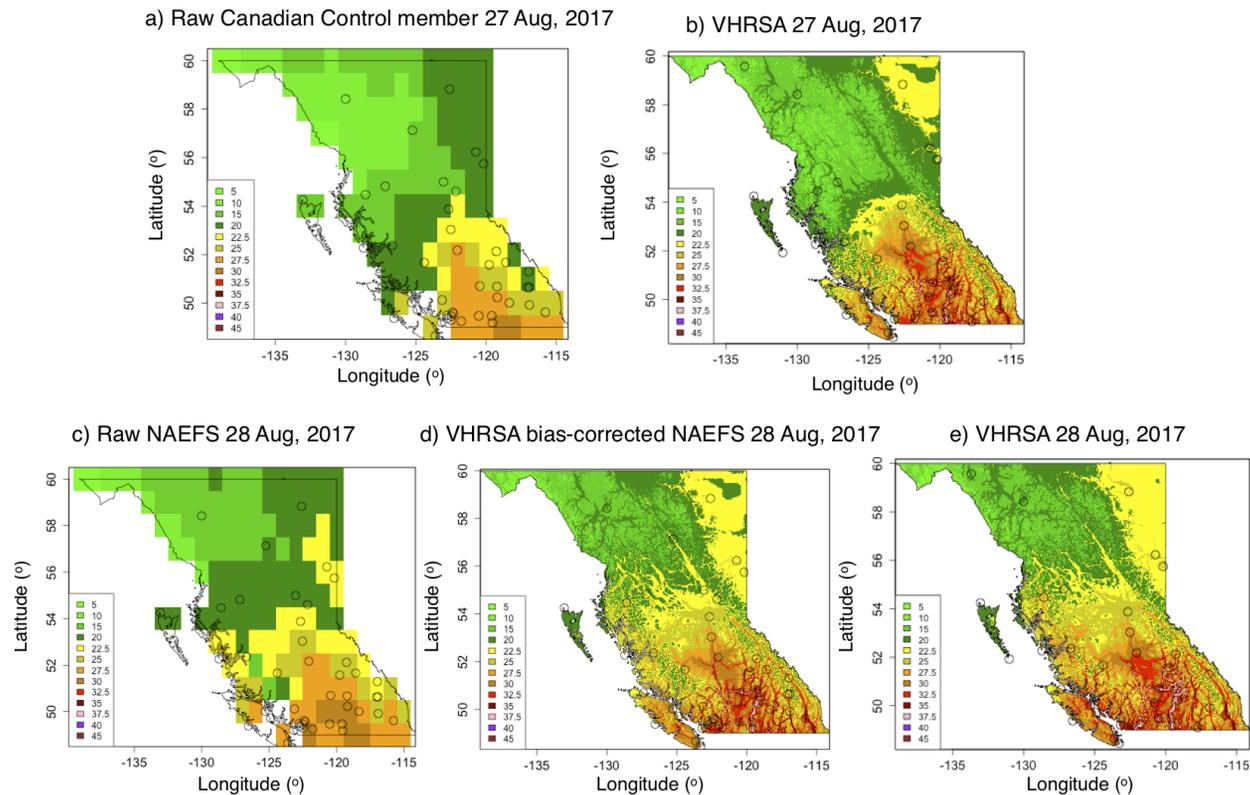


Figure 4.9: (a) NAEFS raw Canadian control member forecast of daily maximum T2M on previous day 27 Aug, 2017; (b) Same as in (a) but for VHRSA; (c) Ensemble mean raw NAEFS forecast on current day 28 Aug, 2017; (d) Ensemble mean bias-corrected NAEFS forecast on current day 28 Aug, 2017. DMB method done separately for each NAEFS member and then averaged to generate the bias-corrected ensemble mean. (e) VHRSA daily maximum T2M on 28 Aug, 2017 presented for comparison with (c) and (d). The dots represent weather station location.

4.4.1 Verification of the Forecast

Most forecast verification metrics have weaknesses (Murphy and Winkler, 1987). In order to obtain an informed picture of the skill of the forecast, different forecast attributes need to be analyzed using different metrics.

The bias (or systematic error) simply measures the difference between the average forecast and average observation (additive for T2M, multiplicative for PCP), and therefore expresses the bias of the forecasts. A positive bias indicates that forecasted value was larger than the observed, which is called overforecasting. Conversely, a negative bias indicates that the event was underforecast, namely, the forecast value was less than the observed.

Metrics based on a 2×2 contingency table are also used for T2M and PCP verification (Table 4.2). Stephenson (2000) suggests the hit (H) and false alarm (F) rates should be used in combination with the threat score (TS). This is helpful for extreme T2M and PCP values because the TS gives a more reasonable comparison between H and F due to the large number of correct rejections ignored.

Forecast	Observed	
	Yes	No
Yes	Hit (a)	False alarm (b)
No	Miss (c)	Correct rejection (d)

Table 4.2: Contingency table.

H is the ratio of correct forecasts to the number of times the event has been observed:

$$H = \frac{a}{a + b}. \quad (4.4)$$

F is the ratio of false alarms to the total number of nonoccurrences of the event:

$$F = \frac{b}{b + d}. \quad (4.5)$$

TS is the number of hits divided by the total number of occasions on which that event was either forecast and/or observed:

$$TS = \frac{a}{a + b + c}. \quad (4.6)$$

The worst possible forecast has high systematic error, $H=0$, $F=1$ and $TS=0$. Conversely, the best possible forecast has systematic error equals to zero, $H = 1$, $F = 0$ and $TS = 1$. In order to evaluate higher impact T2M and PCP events, the 90th percentile of daily maximum T2M and daily PCP, and the 10th percentile of daily minimum T2M are calculated within a 31-day centered rolling window for each calendar day. A hit is issued when both the observed and the forecast ensemble mean values are above/below that calendar day 90th/10th percentile on any given day.

For probabilist forecast verification, the quantile Brier Score (QBS) is used to measure the reliability of the ensemble 90th percentile of daily T2M and daily PCP (Bentzien and Friederichs, 2014). The QBS is connected to the reliability diagram. A reliable or well-calibrated forecast with $QBS = 0$ has points falling closer to the 1:1 perfect-reliability line. That is, if the observations fall at or below the 90th percentile forecast 90% of the time, then the probabilistic forecast is reliable for that percentile.

For overall reliability of any T2M and PCP quantile, Candille et al. (2007) suggest the bias score (b) is combined with the dispersion score (d). The scores b and d are connected to the rank or PIT histogram (Anderson, 1996; Talagrand et al., 1999). A perfectly reliable forecast with $b = 0$ and dispersion equal to 1 ($d = 1$) has a flat rank histogram. A large negative (positive) value of b indicates a negative (positive) bias, where a PIT histogram would slope down towards the right (left). A value of d greater (smaller) than 1 characterizes an underdispersed (overdispersed) ensemble with U-shaped (bell-shaped) PIT histogram.

Finally, the CRPS is analyzed to evaluate forecast sharpness. Forecasts that are frequently much different from climatology are sharp forecasts. The best possible forecast has $CRPS=0$.

A bootstrap procedure calculates the 95% confidence interval for each metric. A metric is estimated from 100 generated samples of the ensemble forecast-observation pairs (Candille et al., 2007). Then, the 2.5th and 97.5th percentiles of the resulting collection of metrics are used as the lower and upper bounds of the 95% confidence intervals for the true metric value. The 95% confidence interval measures the uncertainty around the metric and determines significant statistical differences between the raw and the post-processed forecasts.

4.4.2 Verification of Daily PCP

For daily PCP, the post-processed NAEFS exhibits a statistically significantly better (higher) TS and H than the raw NAEFS out to a forecast lead

time of 7 days (Fig. 4.10a,b). However, the post-processed NAEFS also produces more F than the raw NAEFS (Fig. 4.10c). This is likely due to general inability of the coarse-resolution NAEFS to produce more extreme values. As expected, as lead time increases, TS and H worsen (decrease) and F improves (decreases). One possible explanation is that the number of misses increase as the systematic error increases with increasing lead time (Fig. 4.10d). Further, ensemble member forecast solutions become more random in nature as lead time increases and the ensemble mean typically trends towards climatology, decreasing the chances of forecasting (H or F) a 90th event.

The systematic error of the post-processed NAEFS is also statistically significantly smaller than those of the raw NAEFS as far out as day 9 (Fig 4.10d), indicating the efficacy of the bias correction. As expected, the 95% confidence interval widens with lead time across all metrics. It suggests the post-processed NAEFS has statistically significantly less bias than the raw NAEFS as far out as day 7 across most metrics. Beyond this lead time, although the post-processed NAEFS still outperforms the raw NAEFS, the uncertainty around the metric is too high to determine whether statistically significant differences exist.

The post-processed NAEFS exhibits a much lower QBS, b , d and CRPS values (Fig. 4.10e-h) than the raw NAEFS. This indicates a statistically significantly more reliable, calibrated and sharp forecast. The combined b and d scores (Fig. 4.10e,f) match the results of the QBS (Fig. 4.10g). The post-processed NAEFS is still underdispersed. Namely, too many observations fall in the low and high percentiles. Finally, the CRPS indicates the CDF of the post-processed NAEFS is sharper than the raw NAEFS out to day 4, suggesting forecast are frequently much different than climatology.

Although the downscaled NAEFS is statistically significantly more reliable, calibrated and sharp than the raw NAEFS, there is room for improvement. QBS, b and d values remain high, suggesting a different calibration method may lead to a more skillful forecast.

A Gamma-distributed calibration method was also attempted for PCP (Scheuerer and Hamill, 2015; Baran and Lerch, 2016). Although the versatility in shape of the gamma distribution makes it an attractive candidate for representing precipitation error distributions, the results are unsatisfying in an operational setting due to erratic results. A gamma distribution is more difficult to work with than a Gaussian distribution, and obtaining parameter estimates from data is not as straightforward Wilks (2011). More research is needed in this area.

4.4. Bias-corrected NAEFS

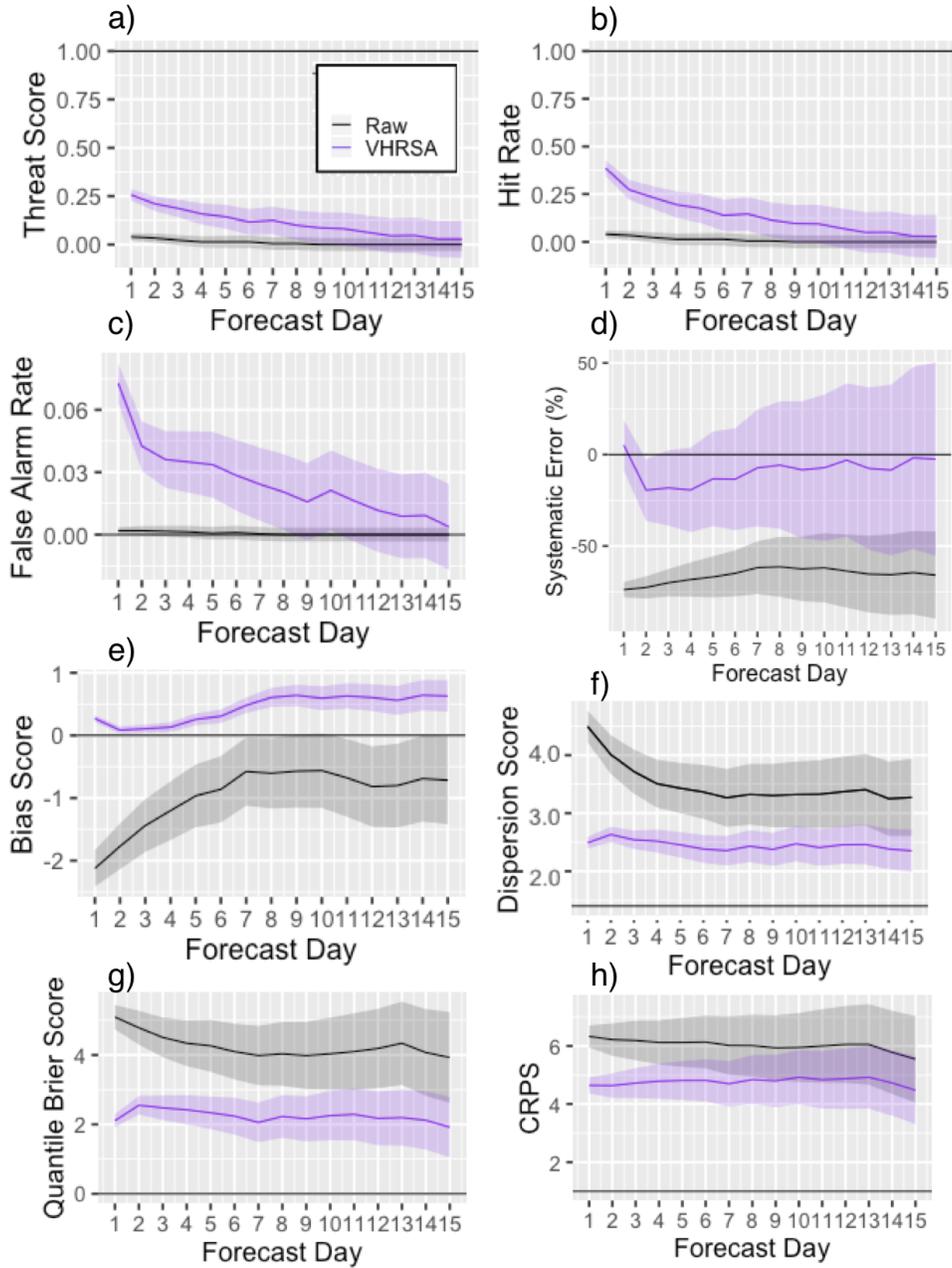


Figure 4.10: a) TS for events above the 95th percentile during Fall 2016; (b) Same as in a) but for H; (c) Same as in a) but for F; (d) Systematic error of daily PCP; (e) Same as in a) but for b ; (f) Same as in a) but for d ; (g) Same as in a) but for QBS; (h) Same as in a) but for CRPS. Values of TS and H closer to one are better. Values of F and systematic error closer to zero are better. Combined values of $b = 0$ and $d = 1$ are better. Values of QBS and CRPS closer to zero are better.

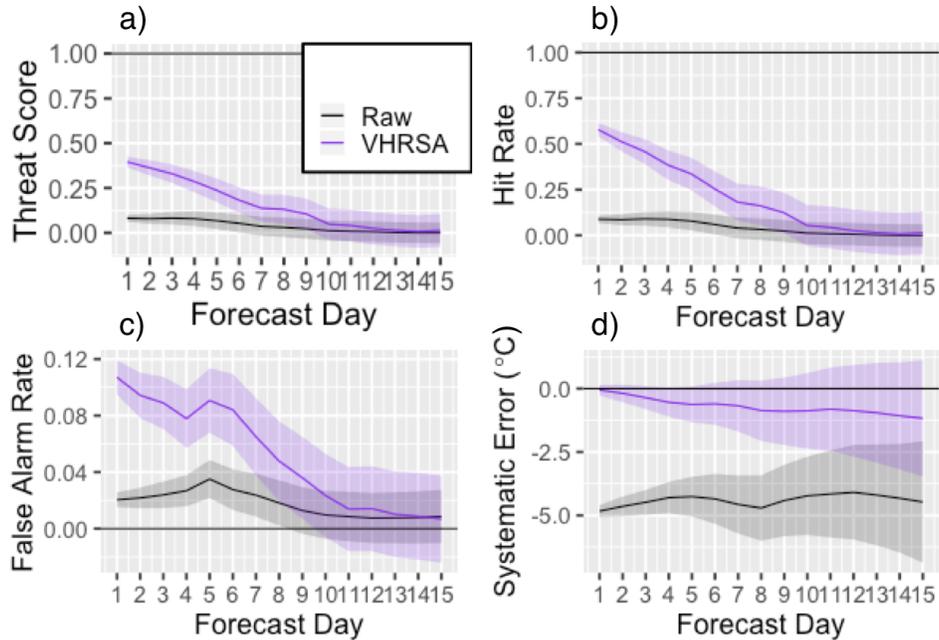


Figure 4.11: a) TS for events above the 90th percentile during Summer 2017; (b) Same as in a) but for H; (c) Same as in a) but for F ; d) Systematic error of daily maximum T2M. Values of TS and H closer to one are better. Values of F and systematic error closer to zero are better.

4.4.3 Verification of Daily Maximum and Minimum T2M

For daily maximum T2M, the post-processed NAEFS exhibits a statistically significantly better (higher) TS and H than the raw NAEFS out to a forecast lead time of 7 days (Fig. 4.11a,b), at which point the uncertainty around the metric is too high to determine significant differences in skill. Similar to daily PCP, the post-processed NAEFS also produces more false alarms (F) than the raw NAEFS (Fig. 4.11c).

The systematic error of the post-processed NAEFS is also statistically significantly smaller than those of the raw NAEFS as far out as day 10 (Fig. 4.11d).

Finally, the post-processed NAEFS exhibits a much better QBS, *b*, *d*, and CRPS than the raw NAEFS for daily maximum T2M (not shown), suggesting a more reliable, calibrated and sharp forecast.

The results for daily minimum T2M are not nearly as good as those for daily maximum T2M and PCP for TS, F and H. This is largely because the

raw NAEFS daily minimum T2M have much less systematic error than the raw NAEFS daily maximum T2M (higher H, TS and F values; not shown), so it is not as easy for the bias correction to improve upon them. The VHRSA post-processed NAEFS exhibits lower TS, H and F than the raw NAEFS. However, the systematic error of daily minimum T2M is statistically significantly better (lower) than those of the raw NAEFS as far as day 10 (Fig 4.12a).

The values of QBS, b , d , and CRPS of daily minimum T2M are of similar magnitude to those of daily maximum T2M (Fig. 4.11b-e). This indicates a statistically significantly more reliable, calibrated and sharp forecast. The combined b and d scores (Fig. 4.11c,d) match the results of the QBS (Fig. 4.11b). The post-processed NAEFS is slightly underdispersed. Namely, a few observations fall in the low and high percentiles. Finally, the CRPS indicates the CDF of the post-processed NAEFS is more accurate than the raw NAEFS out to day 7. As expected, as the lead time increases, the QBS and the CRPS increase as the ensemble mean and distribution becomes less reliable and sharp respectively. b increases in concert with systematic bias, indicating that the mean and distribution follow similar trends. The probabilistic spread goes from being too narrow to well-calibrated, as indicated by d — the raw and calibrated spreads increase quickly with lead time.

The downscaled NAEFS is statistically significantly more reliable, calibrated and sharp than the raw NAEFS for both T2M and PCP. However, QBS, b and d values for T2M suggest the NGR calibration method provides better results for T2M which is approximately Gaussian distributed.

4.5 Stationarity

Many studies using either homogenized station datasets or gridded General Circulation Models (GCM) have shown that climate has undergone changes on a multidecadal time-scale (Mekis and Vincent, 2011; Kharin and Zwiers, 2000; Zwiers and Kharin, 2005; Zhang et al., 2001; Groisman et al., 2005; Odon et al., 2018).

Analysis of multidecadal trends of extremes are an important aspect of climate research. Changes in the magnitude and frequency of extremes have environmental and socioeconomical consequences. It is therefore of great interest to evaluate changes in extreme T2M and PCP in the VHRSA since climate trends at the regional scale are not easy to detect in a station-based study, as it was done in Chapters 2 and 3, due to large natural variability in regional climate (Dettinger et al., 1998; Odon et al.).

4.5. Stationarity

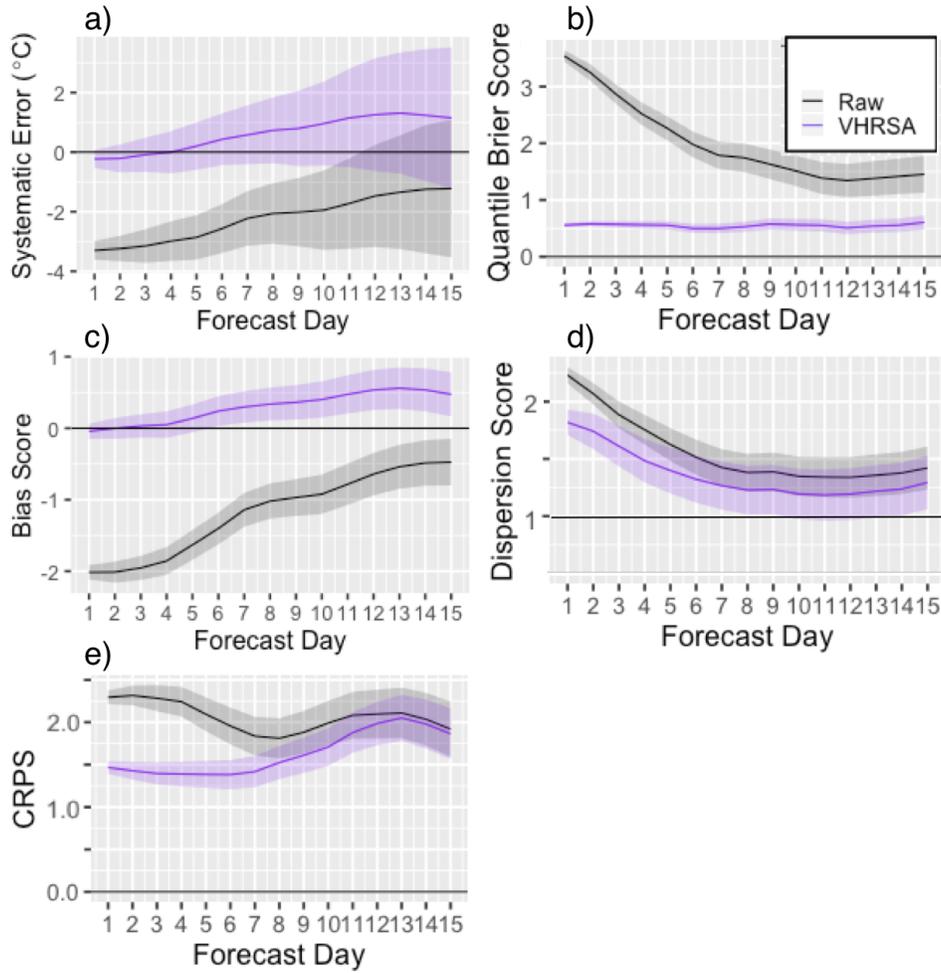


Figure 4.12: a) Systematic error of daily minimum T2M during Winter 2016/17; (b) Same as in a) but for QBS; (c) Same as in a) but for b ; (d) Same as in a) but for d ; (e) Same as in a) but for CRPS. Values of systematic error, QBS and CRPS closer to zero are better. Combined values of $b = 0$ and $d = 1$ are better.

Generally, T2M is Gaussian distributed while PCP is not. Accordingly, parametric and nonparametric methods were presented in Chapters 2 and 3, respectively. In this Chapter, nonparametric methods are presented as they are better suited to compare T2M and PCP on equal footing.

In order to detect spatial climate signals, a nonparametric robust regression is used in the VHRSA to determine the trends in extreme T2M and PCP. This method has the following advantages when compared to the linear regression used in Chapter 2: 1) the regression is robust against outliers and 2) the data doesn't need to be Gaussian distributed.

The trend is assumed to vary with time as $\mu(t) = \mu_0 + \mu_1(t - t_0)$, where the slope coefficient μ_1 represents the annual rate of change in extreme T2M and PCP. The magnitude of μ_1 is calculated by Theil-Sen single median (Theil, 1950) method which computes the slopes of all possible combinations of pairs of the 60 sample values. After calculating these 1475 slopes, the median is taken as the magnitude of μ_1 .

Such an analysis is computationally expensive at every grid point in the VHRSA. Hence, the trend is performed only on the 15th of each month which is assumed to represent the trend of each month.

The trend statistical significance is determined using the nonparametric Mann-Kendall test (Mann, 1945; Wilks, 2011) which has been applied in hydroclimate trend studies (Kumar et al., 2013). The Mann-Kendall test is robust against outliers, independent of the T2M or PCP distribution, and provides a more powerful analysis for non-Gaussian distributed data such as PCP (Yue et al., 2002; Onoz and Bayazit, 2003).

The null hypothesis $\mu_1 = 0$ (no trend) is tested against the alternate hypothesis $\mu_1 \neq 0$ that there is a trend at the $\alpha_{FDR} = 0.10$ level of significance. Namely, a sufficiently large change in extremes throughout the study period indicates there is a trend, and therefore nonstationarity is required to characterize extreme T2M and PCP. Because the Mann-Kendall test requires serially independent data, the effective sample size is used rather than the actual sample size.

Finally, the rolling 31-day centered rolling window is maintained and a GEV dresses these 60 annual values for the 15th of each month by the method of maximum likelihood. A nonstationary GEV distribution is compared, where only the location parameter is allowed to exhibit trend, with a stationary GEV distribution with constant location, scale and shape parameters. The GEV distribution location parameter is assumed to vary with time as $\mu(t) = \mu_0 + \mu_1(t - t_0)$, where the slope coefficient μ_1 represents the annual rate of change in extreme T2M and PCP.

With models $GEV(\mu(t), \sigma, \kappa)$ and $GEV(\mu, \sigma, \kappa)$, the alternate hypothe-

sis is tested against the null hypothesis that extreme values of T2M and PCP are drawn from the same GEV distribution using a LRT at the $\alpha_{FDR} = 0.10$ level of significance. Namely, the alternate hypothesis suggests that a non-stationary model explains more of the variation in the time series and that consequently changes in return levels should be accounted for. For more details on this methodology, refer to subsections 2.5.2 and 3.6.2.

4.5.1 Extreme T2M

Figures 4.13a,b indicate a statistically significant warming trend of extreme values of minimum T2M across most of BC during July at the $\alpha_{FDR} = 0.10$ significance level, with exceptions in the Central climate zone. Similar results are obtained during January, May, June and August (not shown). A non-statistically-significant warming trend of extreme minimum T2M is present throughout the rest of the months across all of BC (not shown).

Under models $GEV(\mu, \sigma, \kappa)$ and $GEV(\mu(t), \sigma, \kappa)$, the LRT suggests the evidence supporting a trend in July is strong, implying a nonstationary model brings significant improvements over a stationary model at the $\alpha_{FDR} = 0.10$ significance level (Fig. 4.13c). Namely, during the 60 years of the study period, changes in return levels of extreme minimum T2M significant, and nonstationarity is required to characterize extreme levels. Again, similar results are obtained during January, May, June and August (not shown).

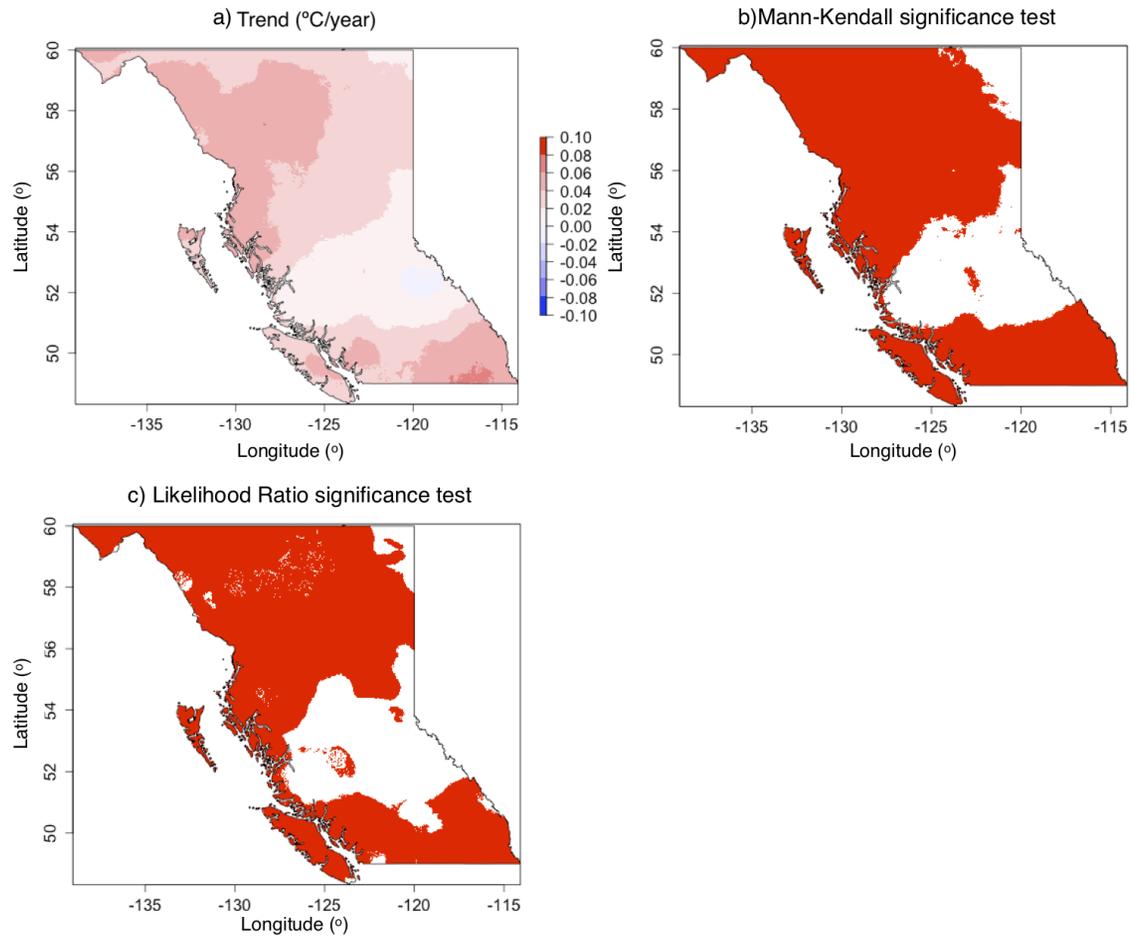


Figure 4.13: (a) Trend of extreme minimum T2M in July; (b) Locations where trend is statistically significant; (c) Locations where $GEV(\mu(t), \sigma, \kappa)$ for extreme minimum T2M is statistically significant.

Analogously, figures 4.14a,b indicate a statistically significant warming trend of extreme values of maximum T2M across the South Central climate zone throughout the year (shown for August). None of the LRT are statistically significant (not shown). A stationary GEV distribution with model $GEV(\mu, \sigma, \kappa)$ is accurate enough to represent extreme maximum T2M.

4.5.2 Extreme PCP

None of the trends in extreme PCP are statistically significant. Furthermore, no clear trend in extreme PCP is discernible across the months (not shown). During the 60 years of the study period, a stationary GEV distribution with model $GEV(\mu, \sigma, \kappa)$ is accurate enough to represent extreme PCP.

4.6 The Parametric Extreme Index

A new Parametric Extreme Index (PEI) is herein developed and presented, with the goal of alerting forecasters of extreme events. It does this by providing a concise, single index value derived from comparing the downscaled NAEFS forecast cumulative distribution function (hereafter NCDF) to the GEV cumulative distribution function (derived from the VHRSA; hereafter GCDF). Namely, on any given day, the NCDF is compared to the GCDF for a given location and calendar day. The 31-day rolling window approach is useful for operational forecasting because it allows for assessment of extreme weather on days within a similar climate. For instance, comparing the true annual maximum T2M (which likely occurs in July) to a winter-time NCDF is of no use since it has no chance of occurring during the winter months. Finally, the GCDF is a distribution of climatological extremes only, which in turn means that the PEI will alert only if the NCDF forecasts an extreme event, rather than just anomalous events, which occur frequently. The PEI ranges from 0 (no forecasted chance of extreme weather) to 1 (high forecasted probability of very extreme weather).

Figure 4.15 illustrates the NCDF and GCDF for the PEI calculation on 28 August 2017 at one grid point. The PEI value is low because there is a low probability of exceeding low extreme values. For reference, the 2-year return level, which represents the median of the GCDF, is far to the right. The PEI is calculated according to the formula

$$PEI = \frac{1}{2} \int_0^{F_c(F_f^{-1}(1))} \frac{1 - F_f(F_c^{-1}(p_c))}{\sqrt{1 - p_c}} dp_c, \quad (4.7)$$

4.6. The Parametric Extreme Index

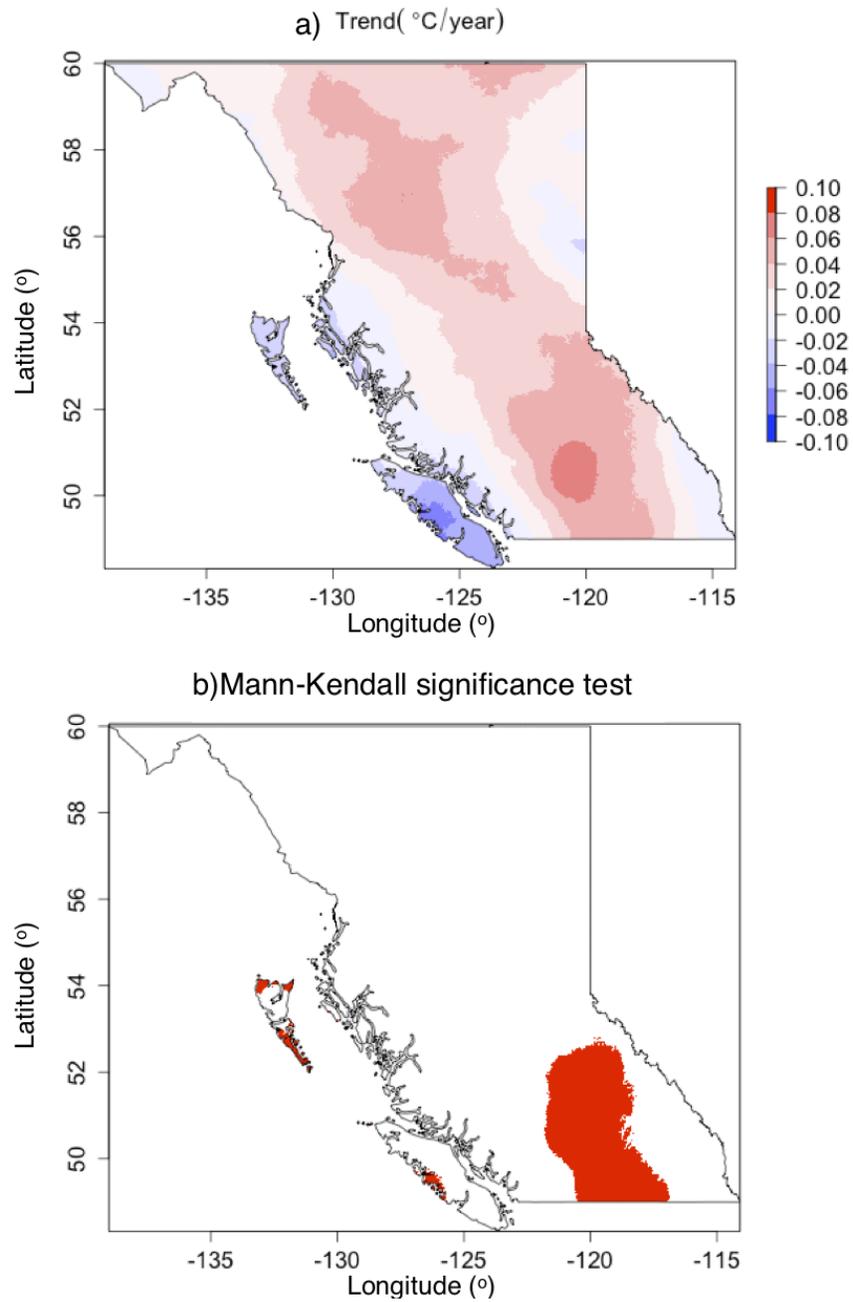


Figure 4.14: (a) Trend of extreme maximum T2M in August; (b) Locations where trend is statistically significant.

where p_c is the probability that an extreme value ($F_c^{-1}(p_c)$) will not be exceeded in the extreme climate record, the inverse CDF F_f^{-1} and F_c^{-1} are simply the quantile functions of the forecast and of the extreme climate distributions respectively, and $1 - F_f(F_c^{-1}(p_c))$ denotes the complement of the NCDF. Namely, if $F_f(F_c^{-1}(p_c))$ represents the NAEFS-derived probability that the extreme value $F_c^{-1}(p_c)$ will not be exceeded, $1 - F_f(F_c^{-1}(p_c))$ represents the probability that the value will be exceeded. $p_c = 0$ is the lower integral limit. It represents the lowest extreme value in the GCDF ($F_c^{-1}(0)$), or the 0th percentile. $p_c = F_c(F_f^{-1}(1))$ is the upper integral limit. It represents the probability that the highest forecast value issued $F_f^{-1}(1)$ will not be exceeded in the extreme climate distribution. In essence, the integral is bounded by the lowest extreme value in the GCDF and the highest forecast issued in the NCDF. The calculations are made so that more weight is given to the right tail of the distributions (where rarer, more extreme values are); that is, as $p_c \rightarrow 1$, $\sqrt{1 - p_c} \rightarrow 0$ and $\frac{1}{\sqrt{1 - p_c}} \rightarrow \infty$.

If the NCDF lies below the lowest extreme value in the GCDF ($F_c^{-1}(0)$), the PEI computes its lowest value: $PEI = 0$. As the magnitude of the highest forecast value increases ($F_f^{-1}(1) \rightarrow \infty$) and the less frequent, more extreme that value is in the GCDF ($F_c(F_f^{-1}(1)) \rightarrow 1$), the PEI approaches its highest value: $PEI = +1$.

Figure 4.16 shows the PEI during an arctic outbreak on 3 January 2017 (VHRSA for this date shown in Fig. 4.7e). Since the event happened during January, the GCDF is given by $GEV(\mu(t), \sigma, \kappa)$. Namely, a nonstationary distribution is used to characterize extreme levels. According to the results of section 4.5, current extreme levels are warmer than a stationary distribution based on the 60-year record would indicate. If a stationary distribution had been used, the PEI would yield a lower index (alert) value, as the extreme minimum T2M levels would have been colder and therefore harder to reach (not shown). Figures 4.17 and 4.18 illustrate the PEI during an atmospheric river event on 8 November 2016 (Fig. 4.8e), and a heat wave on 28 August 2017 (Fig. 4.9e). Both days had large socioeconomical impacts (Odon et al., 2017) (see Chapter 1 for more details).

4.6.1 Verification of the PEI

Existing situational/extreme awareness tools, such as the Standardized Anomaly (SA) and EFI, do not handle extreme values in a proper manner, something that this dissertation has sought to improve upon. Previous methods detect anomalous weather events by comparing them with historical means and/or

4.6. The Parametric Extreme Index

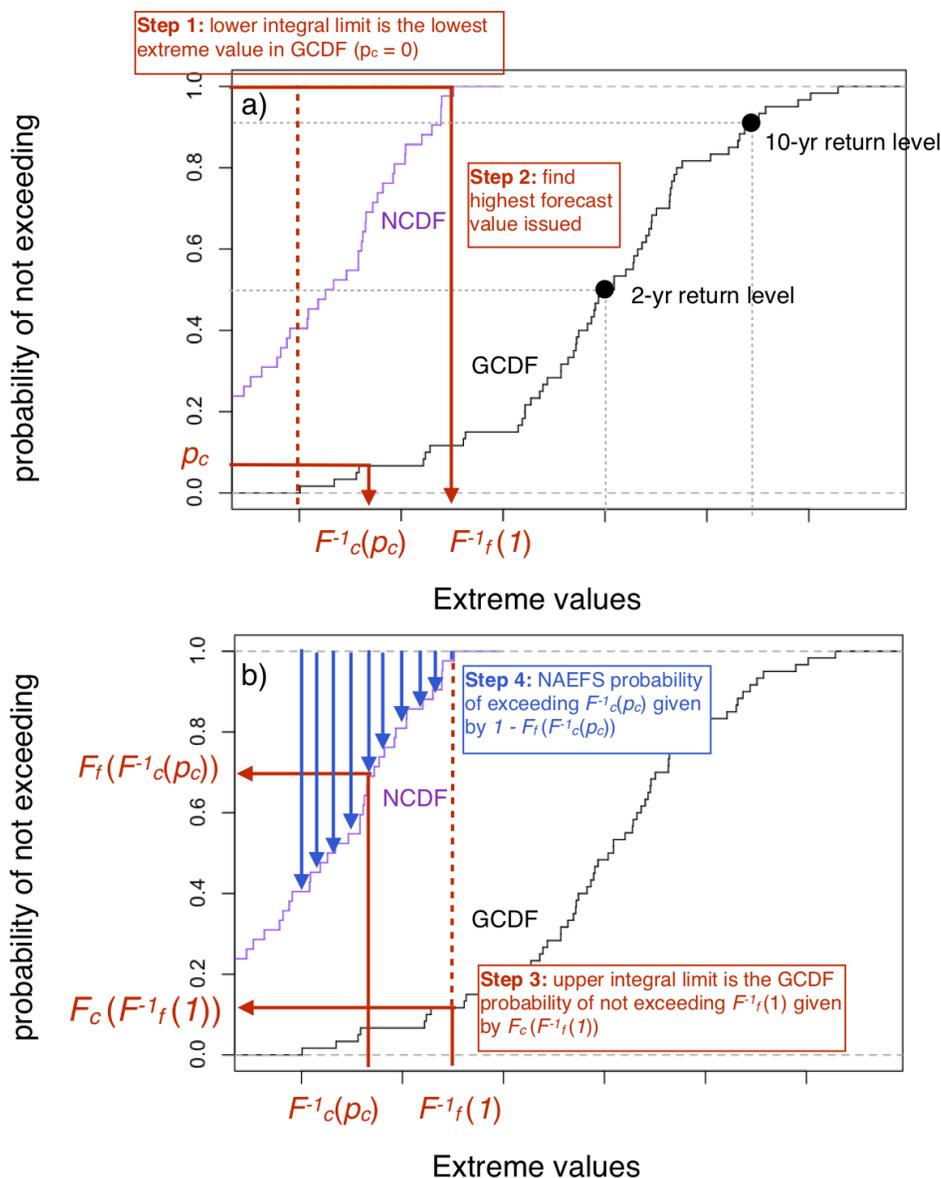


Figure 4.15: (a) The NCDF (purple) and the GCDF (black) show the probability (y-axis) that the extreme value ($F_c^{-1}(p_c)$; x-axis) will not be exceeded. (b) The NAEFS-derived probability that $F_c^{-1}(p_c)$ will not be exceeded is given by $F_f(F_c^{-1}(p_c))$. The complement of the NCDF, given by $1 - F_f(F_c^{-1}(p_c))$, computes the NAEFS-derived probability that $F_c^{-1}(p_c)$ will be exceeded. The PEI is related to the area above the NCDF bounded by $p_c = 0$ and $p_c = F_c(F_f^{-1}(1))$ ($p_c = 0.12$ in the figure). PEI=0.02 indicating low extreme values have a low probability of exceedance. 142

4.6. The Parametric Extreme Index

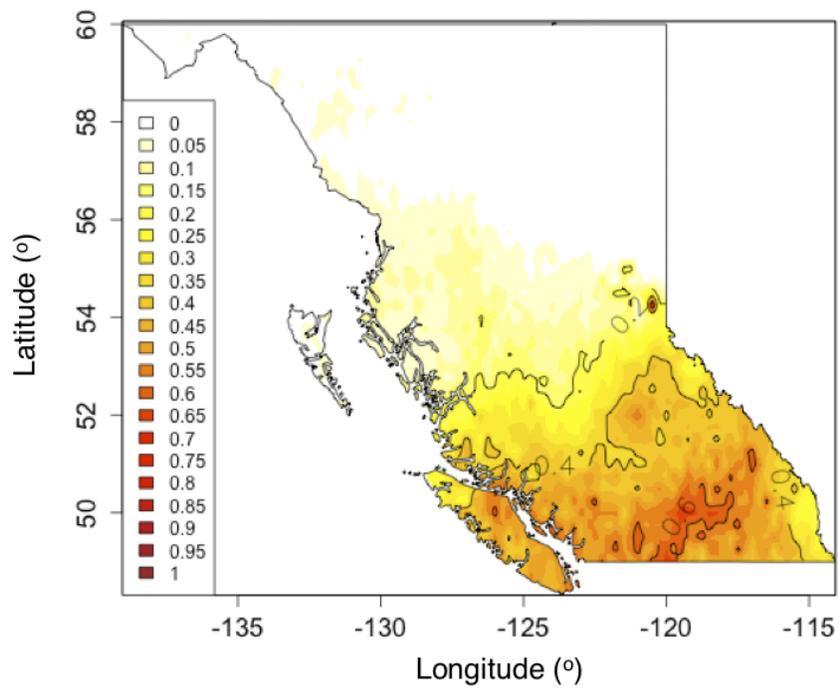


Figure 4.16: Day 1 PEI over BC on 3 Jan 2017.

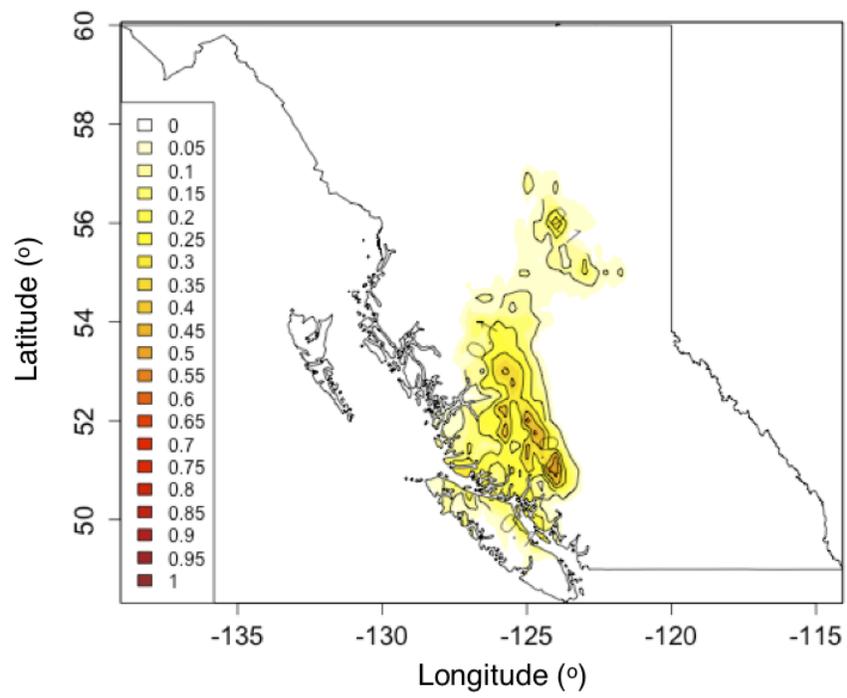


Figure 4.17: Day 1 PEI over BC on 8 Nov 2016.

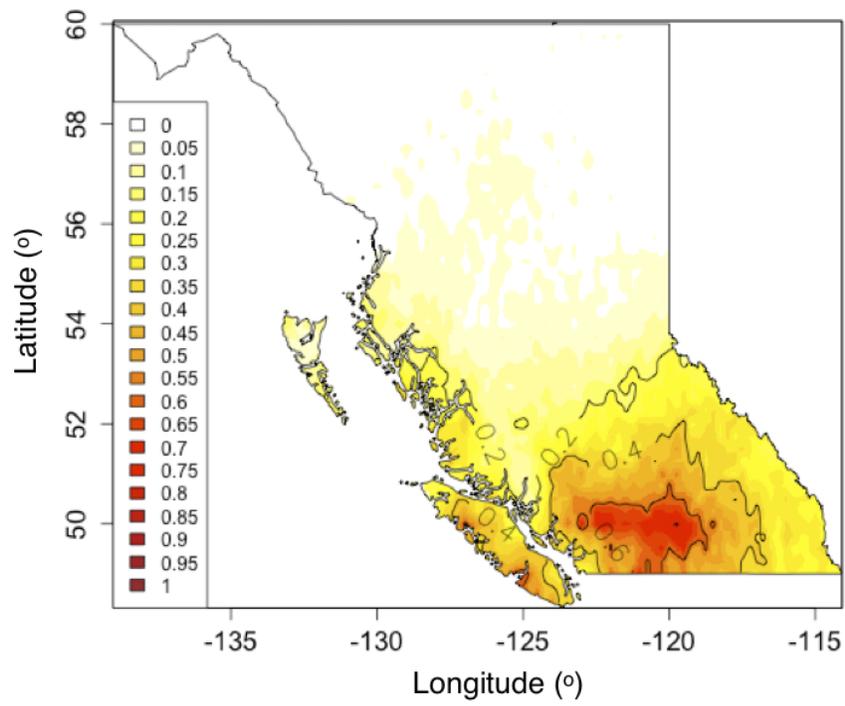


Figure 4.18: Day 1 PEI over BC on 28 Aug 2017.

climate distributions (Lalauette, 2003; Dutra et al., 2013) rather than extreme value distributions as defined above. Little is found in the literature on the skill of either of these existing indices. Generally, SAs are associated with relatively high false alarm rates for medium- and long-range forecasts (Alcott, 2014). In contrast, another study suggests short-range forecasts SAs perform well (Qian et al., 2016). The EFI shows good skill for forecasts between the 10th and 90th percentiles (?).

It is worth mentioning that the climatological mean and standard deviation of the SA are neither robust nor resistant. If the variable distributions are asymmetric or non-Gaussian distributed (like precipitation or integrated precipitable water), the mean and the standard deviation will misrepresent the centre and spread of the data respectively.

Caution is advised when verifying high-impact events. Studies have shown that the skill of the verification metric of choice is proportional to the frequency of extreme weather events (Schaefer, 1990). This creates the misleading impression that rare events cannot be skillfully forecast (Ferro, 2007). For instance, the hit rate converges to zero with increasing rarity of the event (Stephenson et al., 2008).

As most of the scores vanish with increasing rarity of the event (Ghelli and Primo, 2009; Stephenson et al., 2008), the Extreme Dependency Score (EDS) is proposed as a non-vanishing alternative to verify rare events and compare the PEI with the SA during the summer of 2017. The bias-corrected uncalibrated NAEFS is used so that both indices are compared on equal footing since the SA only uses the ensemble mean and cannot fully benefit from a calibrated forecast. The EDS is given by:

$$2 \frac{\ln\left(\frac{a+c}{n}\right)}{\ln\left(\frac{a}{n}\right)} - 1, \quad (4.8)$$

where $n = a + b + c + d$ (see Table 4.2 for contingency table). The EDS is presented with the base rate (BR). BR represents the probability that an event occurs:

$$BR = \frac{a + c}{n} \quad (4.9)$$

In Figure 4.19 the EDS is plotted as a function of $1 - BR$. Thus, large values of $1 - BR$ represent small BR which are interpreted as rare events. The probability $1 - BR$ varies from 0.4 to 0.95 to capture very rare events. Each point in the graph represents BR and EDS values calculated in a contingency table generated with observations varying from above the 95th percentile to below the 99th percentile. That is, as the rarity of the event increases and

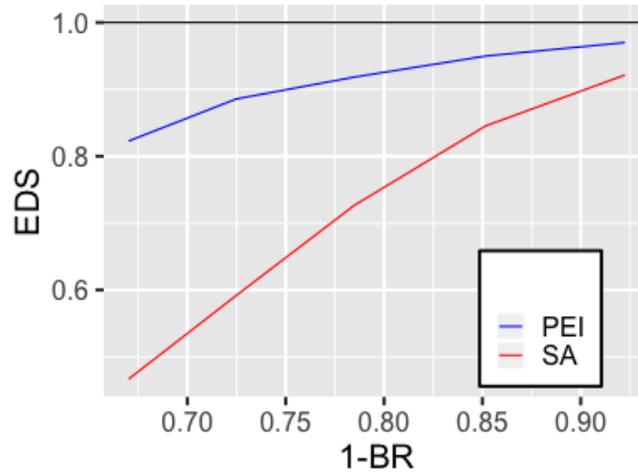


Figure 4.19: EDS as function of $1 - BR$ for Day 1 PEI and SA across all stations during summer of 2017. Rarer events are to the right. Values of EDS closer to 1 are better.

the observations fall above the 95th percentile to below the 99th percentile, a hit is issued if the SA falls above +1 (or below -1 for minimum T2M); for the PEI, a hit is issued if the PEI falls above 0. SA values above +1 (below -1) are considered anomalous because they are at least one standard deviation away from the mean.

The best possible forecast has $EDS=1$. An $EDS=1$ happens if the forecast is either perfect or if it tends to overforecast the extreme event by avoiding any miss ($c=0$). In essence, as the number of hits (a) increase, and the number of misses (c) decrease, the EDS approaches 1, the best possible forecast. Conversely, the worst possible forecast has negative EDS values. That is, as the number of hits decrease, and the number of misses increase, the EDS computes negative values.

4.6.2 PEI Performance across T2M and PCP

In Figure 4.19, the PEI outperforms the SA for all BR values for Summer 2017. The shape of the curve determines the degree of dependency of the EDS on the BR. It implies that the number of misses (c) decreases faster than the number of hits (a) since the EDS approaches 1 as BR approaches less frequent events (Ghelli and Primo, 2009).

Similar results are obtained during fall and winter (not shown). Based

on these findings, the PEI clearly outperforms the SA for alerting users to extreme weather events, because the higher EDS values indicate a higher number of hits and a lower number of misses across a range of extreme events (ranging in rarity, or, in other words, return intervals).

4.7 Conclusion

A new very-high-resolution surface analyses (VHRSA) of daily maximum and minimum 2-m temperature (T2M), and 1-day accumulated precipitation (PCP) dataset that spans the period 1958-2017 was created by downscaling and bias correcting the best performing reanalysis, the JRA-55. The VHRSA was subsequently evaluated against the JRA-55 across the complex terrain of British Columbia (BC). In the evaluation systematic error, MAE and KS statistics were used to compare daily maximum and minimum T2M. To evaluate extreme maximum and minimum T2M, systematic error of 2-year return levels were compared. To compare daily PCP, the systematic error of 31-day precipitation total, and two-sample χ^2 -statistic were calculated. For extreme PCP, the systematic error of 2-year return levels of 1-day accumulated precipitation were compared.

The VHRSA consistently exhibited better scores across all metrics throughout the year and across BC for both daily and extreme T2M. The cold bias of daily maximum T2M in the JRA-55 was mostly removed in the VHRSA. Although the difference in performance in daily and extreme minimum T2M is not as apparent as it is in daily and extreme maximum T2M, the VHRSA still outperforms the JRA-55 throughout the year and across all metrics evaluated. The cold (warm) bias during fall and winter (spring and summer) in the JRA-55 has been reduced in VHRSA.

The results for PCP, while not quite as good as those for T2M, still generally show substantial improvements in the VHRSA. For monthly and extreme PCP, all metrics indicate the VHRSA consistently outperforms the JRA-55 throughout the seasons. In general, the downscaling and bias-correcting leads to a dataset with higher temperatures in valleys and lower temperatures in mountainous regions. Similarly, for PCP the VHRSA is drier in valleys and wetter in ridges and upper elevation regions. This is an important improvement since different studies have shown that NARR under-predicts precipitation and temperature in mountainous regions across BC (Trubilowicz et al., 2016; Hunter et al., 2019).

Thus, the VHRSA will serve as a valuable new spatially and temporally complete, high-resolution, 60-year climatological dataset for BC. Although

4.7. Conclusion

in the present study the VHRSA was developed for BC, the same methodology could be used to improve upon reanalysis datasets wherever observations and PRISM or PRISM-like datasets are available — namely, the United States. The VHRSA has a wide range of potential applications in meteorology, climatology, and hydrology. This study developed the VHRSA with the intention of using it for some such applications.

The VHRSA was first used to bias-correct and downscale the NAEFS ensemble forecast. In order to conduct a thorough evaluation of the forecast, the systematic error, hit rate, false alarm rate, threat score, quantile Brier Score, bias and dispersion scores, and CRPS were used to compare the post-processed NAEFS to the raw NAEFS forecast.

The post-processed NAEFS is generally statistically significantly more skilful than the raw NAEFS forecast for forecast lead times out to 10 days for both T2M and PCP according to the metrics evaluated. Some previous studies have concluded that higher-resolution models were better or at least equal in performance to lower-resolution models across mountainous regions (Schirmer and Jamieson, 2015; Weusthoff et al., 2010; Garvert et al., 2005; Ikeda et al., 2010). A skilful very-high-resolution forecast is useful across the complex terrain of BC as many studies have shown that mountain ranges play an important role in the regional and synoptic scale weather features (Deng et al., 2005; Aptsatryan et al., 2015), and influence the distribution and intensity of precipitation and temperature (Junker et al., 1992; Kunz and Kottmeier, 2006; Smith et al., 2010; Haren et al., 2015).

The Nonhomogeneous Gaussian regression (NGR) probabilistic calibration method resulted in better results for temperature than precipitation. Previous studies have also shown that NGR results in better calibration for temperature (Hagedorn et al., 2008) than precipitation (Hamill et al., 2008). Baran and Lerch (2016) and Hamill et al. (2008) suggest different calibration techniques for non-Gaussian distributed fields such as precipitation and wind speeds. However, finding new robust methods to improve the forecast calibration using such distributions in an operational setting without exacerbating computing demands is an ongoing issue. Nipen and Stull (2011), Baran and Lerch (2016) and Hamill et al. (2008) present different calibration techniques in datasets of the order of 10^3 as it was done here calibrating point forecasts for station locations. Nipen and Stull (2011) evaluated different calibration methods on 1225 grid points across North America between 2001 and 2004 for 2-m temperature, mean sea level pressure, wind speed, precipitable water and relative humidity. Baran and Lerch (2016) verified wind speed forecast across Germany between 2010 and 2011 on 83220 grid points, and Hamill et al. (2008) verified precipitation skill at

4.7. Conclusion

a 32-km horizontal resolution grid across North America from 1982 to 2001. However, conducting such a calibration for the number of grid points in the downscaled NAEFS, which are on the order of 10^6 , is problematic. In Siuta et al. (2017), which also used a dataset three orders of magnitude smaller, a regression method was reported to be computationally cost effective, and gave improved post-calibrated results for wind, which is a non-Gaussian distributed variable. They argue this likely worked because the wind-forecast error distribution was quasi-Gaussian.

The VHRSA is then used to estimate the magnitude and the statistical significance of trends extreme T2M and PCP. There is a noticeable and statistically significant warming trend in the extreme maximum T2M across the South Central climate zone throughout the year, and in extreme values of minimum T2M across most of BC during Summer and late January. The results for extreme T2M support the findings of Chapter 2, that there is a noticeable and statistically significant warming trend in the Southwest and Southeast climate zones of BC during summer months for extreme maximum T2M; during winter months for extreme minimum T2M across BC. Vincent et al. (2012) performed a trend study during a similar study period (1950-2010) for annual maximum and minimum T2M and also reported the results for minimum T2M were more significant than the results of maximum T2M.

Finally, the new Parametric Extreme Index (PEI) outperformed an existing situational/extremes awareness index, the Standardized Anomaly (SA). Its substantially superior EDS value across events with a range of rareness (extremity), indicated that the new index had more hits and less misses when detecting such events. The SA tool is already widely used and liked by operational forecasters. Since the PEI is substantially more useful, it has great potential to see even more widespread use than the SA, for alerting forecasters to the potential for extreme weather.

Although the PEI was developed using the VHRSA over BC, a similar methodology could be employed to create such an index in other parts of the world. The formulation of the PEI alone would likely lead to improved extremes alerting. Having more accurate and reliable probabilistic forecasts, however, will further improve PEI performance.

Future work should examine the performance of alternative calibration methods to further improve the PEI. One possibility is to upscale the VHRSA and probabilistic forecast somewhat, to perhaps 3 km, to reduce computational demands. ? suggests decreasing grid spacing below 4 km provides more detail and structure but has only a limited impact on accuracy. This may make calibration of the entire gridded forecast over BC operationally feasible without losing accuracy. Trends in mean values of

4.7. Conclusion

T2M and PCP using the VHRSA should be evaluated as well.

Chapter 5

Conclusion

The behaviour of extreme and daily values are inherently different. Additionally, the distribution of daily and extreme maximum and minimum 2-m temperature (T2M), and 1-day accumulated precipitation (PCP) are also different. This difference in behaviour and distribution motivated the development of new methodologies to statistically analyze and forecast them. The end goal is to better alert end users to forecasted extremes.

First, in order to identify forecasted values as extreme (i.e., being on the tails of the climatological distribution), the best possible source for climatological data needed to be identified. The paucity of surface weather station data outside of southwest BC population centres motivated the search for the best gridded reanalysis dataset to replace observations as a surrogate climatological dataset. Performance of the latest-generation reanalyses, the Climate Forecast System Reanalysis (CFSR), the ECMWF interim reanalysis (ERA-Interim), the Japanese Meteorological Agency (JMA), and the Modern Era Retrospective-Analysis for Research and Applications (MERRA-2), were rigorously evaluated with respect to daily and extreme maximum and minimum T2M, and daily and extreme PCP over the complex terrain of BC.

New methodologies were developed for statistical evaluation of the reanalyses. These included evaluations based on a daily rolling window, breaking results down by climate zone (which were also determined via a statistically rigorous method), and applying metrics that have rarely if ever been used in previous meteorological studies. In addition to bringing improved methodologies into the field, an effort was made to account for the inherent assumptions in these methods, assumptions which are sometimes ignored. For example, statistical significance had to be determined accounting for multiple testing on a rolling window. Additional efforts were made to find methods and metrics that would be appropriate specifically for extreme values, such as the Generalized Extreme Value distribution (GEV).

The result of the reanalysis evaluation showed the JRA-55 best represents the climatological means, distributions, and distributions of extreme T2M and PCP over BC. While the ERA-Interim showed similar consistent and accurate results for daily and extreme T2M, and the MERRA-2 showed

better results for daily and extreme PCP, the JRA-55 was clearly the superior overall reanalysis with consistent and accurate results across all metrics evaluated.

In the reanalysis evaluation, it became clear that the coarse-resolution JRA-55 had biases related to its inability to resolve the complex terrain of BC. Thus, a methodology was developed to downscale and bias correct T2M and PCP from the JRA-55. The temporal resolution of the JRA-55 (6-hourly), the very high spatial resolution of the PRISM dataset (~ 800 m), and the homogeneity and ground truth from surface weather stations, were combined to create a spatially and temporally complete very-high-resolution surface analysis (VHRSA). Again, this new VHRSA was rigorously evaluated, and was generally found to be significantly less biased, and more accurate than the JRA-55. Thus, a new, significantly improved, very-high-resolution 60-year daily surface analysis dataset was created for BC. This should prove to be an immensely valuable dataset for research and operational use by the meteorological, climatological, and hydrological communities. For the purposes of this study, it also rendered a feasible solution to the paucity of observational data across BC.

Probabilistic forecasts of extremes requires a probabilistic forecast dataset. The North American Forecast System (NAEFS) is one such widely used ensemble forecast dataset. However, it is relatively low resolution, like the reanalysis datasets. Downscaling and bias correction were needed here as well. The VHRSA was used to statistically downscale and bias correct the NAEFS. Computational restraints prohibited use of the VHRSA to probabilistically calibrate the forecasts, however, calibration was done for point forecasts at weather station locations. Calibration is needed so that the forecasted probabilities are more accurate (that is, an event with a forecasted 30% probability of occurrence is observed on average 30 out of 100 times). While the bias correction was largely effective at significantly reducing bias, the calibration delivered mixed results. It reported better results calibrating surface temperature — which is approximately Gaussian — than precipitation, a non-Gaussian distributed variable. Nonetheless, overall the post-processed NAEFS showed significant improvement over the raw, in most cases out to a forecast lead time of 10 days.

The VHRSA was also used to analyze trends in extreme T2M and PCP during the 1958-2017 period. A significant warming trend in extreme values of minimum T2M across most of BC during Summer and January were found — that is, extreme daily minimum temperatures are getting less cold at some times of year. Additionally, there was a significant increasing trend in extreme maximum T2M values across the South Central climate zone

throughout the year — extremes of daily maximum temperatures are getting hotter. While there were some indications of warming trends in other months, they were not statistically significant. For extreme PCP, no increasing or decreasing trends were apparent.

These post-processed probabilistic forecast distributions were then used to create a new extreme forecast index (sometimes referred to as a situational awareness tool). This so-called Parametric Extreme Index (PEI) differentiates itself and improves upon existing such indexes in that:

1. It accounts for nonstationarity when appropriate. This accounts for time-changing climate distributions, which were herein shown to be significant for some variables, regions and times of year. It yields a more appropriate gauge of how extreme a value is based on the current climate.
2. It is based on forecast distributions that are bias corrected and down-scaled onto a very-high-resolution grid. Further, point forecast values of the index are probabilistically calibrated. This improves the forecast accuracy and bias of the index.
3. It uses a more appropriate extreme values climate distribution, rather than a mean climate distribution. A mean climate distribution is inappropriate for non-Gaussian distributed variables such as precipitation. The mean and the standard deviation will misrepresent the centre and spread of the data respectively leading to meaningless Standardized Anomalies values. Additionally, it leads to more false alarms as forecasts far away from the climate mean occur frequently.

Indeed, the PEI was shown to substantially outperform the well known Standardized Anomalies index across a range of extreme events ranging in rarity, with a higher number of hits and a lower number of misses. Both are desirable to alert forecast users of future extreme weather events.

Collectively, this dissertation developed appropriate statistical methods to analyze and forecast extreme weather events. These new techniques lead to the creation of a new dataset, and the ability to better forecast extreme weather events. This methodology can be used to improve reanalyses wherever observations and PRISM or PRISM-like datasets are available, and be used for modelling in fields like hydrology, ecology, and agriculture besides meteorology and climatology. A tool to better forecast extreme weather events provides earlier and more accurate detection of such events, which in turn can help community responders, emergency managers, regional plan-

ners, government and the media to take appropriate action to mitigate damage and reduce casualties.

Future work will examine the performance of alternative calibrating methods to post-process the very-high-resolution NAEFS forecast in a true operational setting without exacerbating computing demands. The PEI will be evaluated across different seasons and forecast lead times. Finally, non-linear trends will be tested for precipitation to evaluate whether a more complex nonstationary model is more accurate to represent extreme precipitation than the ones studied here. Trends in mean values of temperature and precipitation using the VHRSA will also be evaluated.

Bibliography

- J. T. Abatzoglou. Development of gridded surface meteorological data for ecological applications and modelling. *International Journal of Climatology*, 33(1):121–131, 2013. ISSN 08998418. doi: 10.1002/joc.3413.
- R. F. Adler, G. J. Huffman, A. Chang, R. Ferraro, P.-P. Xie, J. Janowiak, B. Rudolf, U. Schneider, S. Curtis, D. Bolvin, A. Gruber, J. Susskind, P. Arkin, #. And, and E. Nelkin. The Version-2 Global Precipitation Climatology Project (GPCP) Monthly Precipitation Analysis (1979-Present). *J. Hydrometeor*, 4:1147–1167, 2003. URL <http://precip.gsfc.nasa.gov>.
- T. Alcott. How I Learned to Stop Worrying and Love the Global Ensembles, 2014.
- J. Anderson. A method for producing and evaluating probabilistic forecasts from ensemble model integrations. *Journal of Climate*, 9:1518–1530, 1996.
- H. Atsatryan, A. Shakhnazaryan, V. Sahakyan, Y. Shoukourian, V. Kotroni, Z. Petrosyan, R. Abrahamyan, and H. Melkonyan. WRF-ARW model for prediction of high temperatures in south and south east regions of Armenia. In *Proceedings - 11th IEEE International Conference on eScience, eScience 2015*, pages 207–213, 2015. ISBN 9781467393256. doi: 10.1109/eScience.2015.82.
- X. Bao and F. Zhang. Evaluation of NCEP-CFSR, NCEP-NCAR, ERA-Interim, and ERA-40 reanalysis datasets against independent sounding observations over the Tibetan Plateau. *Journal of Climate*, 26:206–214, 2013. ISSN 08948755. doi: 10.1175/JCLI-D-12-00056.1.
- S. Baran and S. Lerch. Mixture EMOS model for calibrating ensemble forecasts of wind speed. *Environmetrics*, 27(2):116–130, 2016. ISSN 1099095X. doi: 10.1002/env.2380.
- BC Hydro. Record inflows, but limited downstream flooding, thanks to BC Hydro operations and some luck, 2016.

Bibliography

URL <https://www.bchydro.com/news/conservation/2016/comox-campbell-river-watersheds-break-records.html>.

- S. Bentzien and P. Friederichs. Decomposition and graphical portrayal of the quantile score. *Quarterly Journal of the Royal Meteorological Society*, 140(683):1924–1934, 2014. ISSN 1477870X. doi: 10.1002/qj.2284.
- A. A. Berg, J. Famiglietti, J. Walker, and P. Houser. Impact of bias correction to reanalysis products on simulations of North American soil moisture and hydrological fluxes. *Journal of Geophysical Research*, 108(4490), 2003. ISSN 0148-0227. doi: 10.1029/2002JD003334.
- P. Berrisford, D. Dee, K. Fielding, M. Fuentes, P. Kallberg, S. Kobayashi, and S. Uppala. The ERA-Interim Archive Version 2.0. *ERA report series 1*, 1(1):1–27, 2011.
- A. K. Betts, M. Köhler, and Y. Zhang. Comparison of river basin hydrometeorology in ERA-Interim and ERA-40 reanalyses with observations. *Journal of Geophysical Research Atmospheres*, 114(2), 2009. ISSN 01480227. doi: 10.1029/2008JD010761.
- J. Bhend and P. Whetton. Consistency of simulated and observed regional changes in temperature, sea level pressure and precipitation. *Climatic Change*, 118(3-4):799–810, jun 2013. ISSN 0165-0009. doi: 10.1007/s10584-012-0691-2. URL <http://link.springer.com/10.1007/s10584-012-0691-2>.
- S. C. Bloom, L. L. Takacs, A. M. da Silva, D. Ledvina, S. C. Bloom, L. L. Takacs, A. M. da Silva, and D. Ledvina. Data Assimilation Using Incremental Analysis Updates. *Monthly Weather Review*, 124(6):1256–1271, jun 1996. ISSN 0027-0644. doi: 10.1175/1520-0493(1996)124(1256:DAUIAU)2.0.CO;2.
- M. G. Bosilovich, J. Chen, F. R. Robertson, and R. F. Adler. Evaluation of global precipitation in reanalyses. *Journal of Applied Meteorology and Climatology*, 47(9):2279–2299, 2008. ISSN 15588424. doi: 10.1175/2008JAMC1921.1.
- M. G. Bosilovich, S. Akella, L. Coy, R. Cullather, C. Draper, R. Gelaro, R. Kovach, Q. Liu, A. Molod, P. Norris, K. Wargan, W. Chao, R. Reichle, L. Takacs, Y. Vikhliayev, S. Bloom, A. Collow, S. Firth, G. Labow,

Bibliography

- G. Partyka, S. Pawson, O. Reale, S. D. Schubert, and M. Suarez. MERRA-2: Initial Evaluation of the Climate. *Technical Report Series on Global Modeling and Data Assimilation*, 43(September):1–139, 2015.
- D. R. Bourdin, T. N. Nipen, and R. B. Stull. Reliable probabilistic forecasts from an ensemble reservoir inflow forecasting system. *Water Resources Research*, 50:3108–3130, 2014. doi: 10.1002/2014WR015462.Received.
- Y. Brend. BC Ferries cancels sailing due to 'severe weather' as winter chill hits — CBC News, 2016. URL <https://www.cbc.ca/news/canada/british-columbia/weather-ferry-arctic-outbreak-1.3880910>.
- R. Buizza. Potential Forecast Skill of Ensemble Prediction and Spread and Skill Distributions of the ECMWF Ensemble Prediction System. *Monthly Weather Review*, 125(1):99–119, 1997. ISSN 0027-0644. doi: 10.1175/1520-0493(1997)125<0099:PFSOEP>2.0.CO;2. URL <http://journals.ametsoc.org/doi/abs/10.1175/1520-0493%7D281997%7D29125-%7D3C0099%7D3APFSOEP%7D3E2.0.CO%7D3B2>.
- Y. Cai and D. Hames. Minimum sample size determination for generalized extreme value distribution. *Communications in Statistics: Simulation and Computation*, 40(1):87–98, 2010. ISSN 15324141. doi: 10.1080/03610918.2010.530368.
- G. Candille, C. Côté, P. L. Houtekamer, and G. Pellerin. Verification of an Ensemble Prediction System against Observations. *Monthly Weather Review*, 135(7):2688–2699, 2007. ISSN 0027-0644. doi: 10.1175/MWR3414.1. URL <http://journals.ametsoc.org/doi/abs/10.1175/MWR3414.1>.
- M. Charron, G. Pellerin, L. Spacek, P. L. Houtekamer, N. Gagnon, H. L. Mitchell, and L. Michelin. Toward Random Sampling of Model Error in the Canadian Ensemble Prediction System. *Monthly Weather Review*, 138(5):1877–1901, 2010. ISSN 0027-0644. doi: 10.1175/2009MWR3187.1. URL <http://journals.ametsoc.org/doi/abs/10.1175/2009MWR3187.1>.
- R. R. Chilton. *A Summary of Climatic Regimes of British Columbia*. B.C. Min. Environ., Victoria, B.C., 1981.
- C. Correia. Winter conditions testing patience of Lower Mainland residents — CBC News, 2016. URL <https://www.cbc.ca/news/canada/british-columbia/lower-mainland-snow-issues-1.3913674>.

- B. A. Cosgrove. Real-time and retrospective forcing in the North American Land Data Assimilation System (NLDAS) project. *Journal of Geophysical Research*, 108(D22):8842, 2003. ISSN 0148-0227. doi: 10.1029/2002JD003118.
- F. C. Curriero, J. a. Patz, J. B. Rose, and S. Lele. The Association Between Extreme Precipitation and Waterborne Disease Outbreaks in the United States , 1948 – 1994. *American Journal of Public Health*, 91(8):1194–1199, 2001. ISSN 0090-0036. doi: 10.2105/AJPH.91.8.1194.
- C. Daly, R. Neilson, and D. Phillips. A Statistical-Topographic Model for Mapping Climatological Precipitation over Mountainous Terrain. *Journal of Applied Meteorology*, 33:140–158, 1994.
- C. Daly, G. Taylor, and W. Gibson. The Prism approach to mapping precipitation and temperature. In *10th AMS Conference on Applied Climatology*, number 1, pages 10–12, Reno, NV, 1997.
- C. Daly, G. H. Taylor, W. P. Gibson, T. W. Parzybok, G. L. Johnson, and P. A. Pasteris. Climate division normals derived from topographically-sensitive climate grids. In *13th AMS Conf. on Applied Climatology*, pages 177–180, Portland, OR, 2002.
- R. A. Davis. The Rate of Convergence in Distribution of the Maxima. *Statistica Neerlandica*, 36(1):31–35, 1982.
- L. De Haan. Sample Extremes: An Elementary Introduction. *Statistica Neerlandica*, 30(4):161–172, 1976. doi: 10.1111/j.1467-9574.1976.tb00275.x.
- M. Decker, M. A. Brunke, Z. Wang, K. Sakaguchi, X. Zeng, and M. G. Bosilovich. Evaluation of the Reanalysis Products from GSFC, NCEP, and ECMWF Using Flux Tower Observations. *Journal of Climate*, 25: 1916–1944, 2012.
- D. P. Dee, S. M. Uppala, A. J. Simmons, P. Berrisford, P. Poli, S. Kobayashi, U. Andrae, M. A. Balmaseda, G. Balsamo, P. Bauer, P. Bechtold, A. C. M. Beljaars, L. van de Berg, J. Bidlot, N. Bormann, C. Delsol, R. Dragani, M. Fuentes, A. J. Geer, L. Haimberger, S. B. Healy, H. Hersbach, E. V. Hólm, L. Isaksen, P. Kållberg, M. Köhler, M. Matricardi, A. P. McNally, B. M. Monge-Sanz, J.-J. Morcrette, B.-K. Park, C. Peubey, P. de Rosnay, C. Tavolato, J.-N. Thépaut, and F. Vitart. The ERA-Interim reanalysis: configuration and performance of the data assimilation system. *Quarterly*

Bibliography

Journal of the Royal Meteorological Society, 137(656):553–597, apr 2011. ISSN 00359009. doi: 10.1002/qj.828.

Demography Division. Canadian Demographics at a Glance. *Statistics Canada*, 91(003):1–81, 2016. ISSN 1916-1832. doi: 91-003-XWE.

X. Deng, R. Stull, and B. Columbia. A Mesoscale Analysis Method for Surface Potential Temperature in Mountainous and Coastal Terrain. *Monthly weather review*, 133(2):389–408, 2005. ISSN 0027-0644. doi: 10.1175/MWR-2859.1. URL <http://journals.ametsoc.org/doi/pdf/10.1175/MWR-2859.1><http://journals.ametsoc.org/doi/abs/10.1175/MWR-2859.1>.

M. D. Dettinger, D. R. Cayan, H. F. Diaz, and D. M. Meko. North-South precipitation patterns in western North America on interannual-to-decadal timescales. *Journal of Climate*, 11(12):3095–3111, 1998. ISSN 08948755. doi: 10.1175/1520-0442(1998)011<3095:NSPPIW>2.0.CO;2.

K. a. Devine and É. Mekis. Field accuracy of Canadian rain measurements. *Atmosphere - Ocean*, 46(2):213–227, 2008. ISSN 07055900. doi: 10.3137/ao.460202.

M. G. Donat, L. V. Alexander, H. Yang, I. Durre, R. Vose, R. J. H. Dunn, K. M. Willett, E. Aguilar, M. Brunet, J. Caesar, B. Hewitson, C. Jack, A. M. G. Klein Tank, A. C. Kruger, J. Marengo, T. C. Peterson, M. Renom, C. Oria Rojas, M. Rusticucci, J. Salinger, A. S. Elrayah, S. S. Sekele, A. K. Srivastava, B. Trewin, C. Villarreal, L. A. Vincent, P. Zhai, X. Zhang, and S. Kitching. Updated analyses of temperature and precipitation extreme indices since the beginning of the twentieth century: The HadEX2 dataset. *Journal of Geophysical Research: Atmospheres*, 118(5):2098–2118, mar 2013. ISSN 2169897X. doi: 10.1002/jgrd.50150.

E. Dutra, M. Diamantakis, I. Tsonevsky, E. Zsoter, F. Wetterhall, T. Stockdale, D. Richardson, and F. Pappenberger. The extreme forecast index at the seasonal scale. *Atmospheric Science Letters*, 14(4):256–262, 2013. ISSN 1530261X. doi: 10.1002/asl2.448.

A. Ebita, S. Kobayashi, Y. Ota, and M. Moriya. The new Japanese reanalysis project: JRA55. In *Proc. 15th GCOS/WCRP Atmospheric Observation Panel for Climate*, volume Global Cli, page 12.2, Geneva, Switzerland, 2009. Climate Prediction Division, JMA.

- A. Ebita, S. Kobayashi, Y. Ota, M. Moriya, R. Kumabe, O. Kazutoshi, Y. Harada, S. Yasui, K. Miyaoka, K. Takahashi, H. Kamahori, C. Kobayashi, H. Endo, M. Soma, Y. Oikawa, and T. Ishimizu. The Japanese 55-year Reanalysis “JRA-55”: An Interim Report. *SOLA: Scientific Online Letters of the Atmosphere*, 7:149–152, 2011.
- C. a. T. Ferro. A Probability Model for Verifying Deterministic Forecasts of Extreme Events. *Weather and Forecasting*, 22(5):1089–1100, 2007. ISSN 0882-8156. doi: 10.1175/WAF1036.1.
- W. Y. Fritz. Winter is no reason to mind your manners in @WinnieYeo . *Twitter*, (Dec 20, 2016):Retrieved from <https://twitter.com/winnieyeo?lang=en>, 2016.
- M. F. Garvert, B. a. Colle, and C. F. Mass. The 13–14 December 2001 IMPROVE-2 Event. Part I: Synoptic and Mesoscale Evolution and Comparison with a Mesoscale Model Simulation. *Journal of the Atmospheric Sciences*, 62(10):3474–3492, 2005. ISSN 0022-4928. doi: 10.1175/JAS3549.1. URL <http://journals.ametsoc.org/doi/abs/10.1175/JAS3549.1>.
- R. Gelaro. MERRA-2 and future reanalysis at GMAO. In *NOAA Climate Reanalysis Task Force Technical Workshop*, page 15pp., College Park, MD, 2015.
- R. Gelaro, W. McCarty, M. J. Suárez, R. Todling, A. Molod, L. Takacs, C. a. Randles, A. Darmenov, M. G. Bosilovich, R. Reichle, K. Wargan, L. Coy, R. Cullather, C. Draper, S. Akella, V. Buchard, A. Conaty, A. M. da Silva, W. Gu, G. K. Kim, R. Koster, R. Lucchesi, D. Merkova, J. E. Nielsen, G. Partyka, S. Pawson, W. Putman, M. Rienecker, S. D. Schubert, M. Sienkiewicz, and B. Zhao. The modern-era retrospective analysis for research and applications, version 2 (MERRA-2). *Journal of Climate*, 30(14):5419–5454, 2017. ISSN 08948755. doi: 10.1175/JCLI-D-16-0758.1.
- A. Ghelli and C. Primo. On the use of the extreme dependency score to investigate the performance of an NWP model for rare events. *Meteorological Applications*, 16:537–544, 2009.
- B. Gnedenko. On the Limiting Distribution of a Supercritical Branching Process in a Random Environment. *Annals of Mathematics*, 44:423–453, 1943.

- T. Gneiting, A. E. Raftery, A. H. Westveld, and T. Goldman. Calibrated Probabilistic Forecasting Using Ensemble Model Output Statistics and Minimum CRPS Estimation. *Monthly Weather Review*, 133(5): 1098–1118, 2005. ISSN 0027-0644. doi: 10.1175/MWR2904.1. URL <http://journals.ametsoc.org/doi/abs/10.1175/MWR2904.1>.
- R. Graham, T. Alcott, N. Hosenfeld, and R. Grumm. Anticipating a rare event utilizing forecast anomalies and a situational awareness display: The Western U.S. Storms of 18-23 January 2010. *Bulletin of the American Meteorological Society*, 94(12):1827–1836, 2013. ISSN 00030007. doi: 10.1175/BAMS-D-11-00181.1.
- P. Groisman, R. Knight, D. Easterling, T. R. Karl, G. C. Hegerl, and V. N. Razuvaev. Trends in precipitation intensity in the climate record. *Journal of Climate*, 18:1326–1350, 2005. ISSN 08948755. doi: 10.1175/JCLI3339.1.
- V. Grubišić, R. K. Vellore, and A. W. Huggins. Quantitative Precipitation Forecasting of Wintertime Storms in the Sierra Nevada: Sensitivity to the Microphysical Parameterization and Horizontal Resolution. *Monthly Weather Review*, 133(10):2834–2859, 2005. ISSN 0027-0644. doi: 10.1175/MWR3004.1. URL <http://journals.ametsoc.org/doi/abs/10.1175/MWR3004.1>.
- R. Grumm and R. Hart. Standardized anomalies applied to significant cold season weather events: Preliminary findings. *Weather & Forecasting*, 16: 736–754, 2001. ISSN 0882-8156. doi: 10.1175/1520-0434(2001)016<0736:SAATSC>2.0.CO;2.
- A. M. Gunn. *Climate and Weather Extremes*. Greenwood, 2010. ISBN 9780313355684. doi: 10.1016/B978-0-12-386917-3.00010-5.
- J. P. Hacker, E. Scott, K. And, and R. B. Stull. Ensemble Experiments on Numerical Weather Prediction Error and Uncertainty for a North Pacific Forecast Failure. *Weather and Forecasting*, 18:12–31, 2003.
- R. Hagedorn, T. M. Hamill, and J. S. Whitaker. Probabilistic Forecast Calibration Using ECMWF and GFS Ensemble Reforecasts. Part I: Two-Meter Temperatures. *Monthly Weather Review*, 136(7):2608–2619, 2008. ISSN 0027-0644. doi: 10.1175/2007MWR2410.1. URL <http://journals.ametsoc.org/doi/abs/10.1175/2007MWR2410.1>.
- T. M. Hamill. NOTES AND CORRESPONDENCE Interpretation of Rank Histograms for Verifying Ensemble Forecasts. *Mon. Wea. Rev.*, 129:

Bibliography

- 550–560, 2001. URL [https://journals.ametsoc.org/doi/pdf/10.1175/1520-0493\(2001\)29129:3C0550:3AIORHFV:3E2.0.CO;3B2](https://journals.ametsoc.org/doi/pdf/10.1175/1520-0493(2001)29129:3C0550:3AIORHFV:3E2.0.CO;3B2).
- T. M. Hamill, R. Hagedorn, and J. S. Whitaker. Probabilistic Forecast Calibration Using ECMWF and GFS Ensemble Reforecasts. Part II: Precipitation. *Monthly Weather Review*, 136(7):2620–2632, 2008. ISSN 0027-0644. doi: 10.1175/2007MWR2411.1. URL <http://journals.ametsoc.org/doi/abs/10.1175/2007MWR2411.1>.
- M. Harada. Introduction to reanalysis and JRA-55. In *TCC Training Seminar on Seasonal Forecast, Tokyo Climate Center - Japan Meteorological Agency*, number Tokyo Climate Center - Japan Meteorological Agency, pages 17–26, Tokyo, Japan, 2018.
- Y. Harada, H. Kamahori, C. Kobayashi, H. Endo, S. Kobayashi, Y. Ota, H. Onoda, K. Onogi, K. Miyaoka, and K. Takahashi. The JRA-55 Reanalysis: Representation of Atmospheric Circulation and Climate Variability. *Journal of the Meteorological Society of Japan. Ser. II*, 94(3):269–302, 2016. ISSN 0026-1165. doi: 10.2151/jmsj.2016-015.
- R. V. Haren, R. J. Haarsma, G. J. V. Oldenborgh, and W. Hazeleger. Resolution dependence of European precipitation in a state-of-the-art atmospheric general circulation model. *Journal of Climate*, 28(13):5134–5149, 2015. ISSN 08948755. doi: 10.1175/JCLI-D-14-00279.1.
- R. E. Hart and R. H. Grumm. Using Normalized Climatological Anomalies to Rank Synoptic-Scale Events Objectively. *Monthly Weather Review*, 129:2426–2442, 2001. ISSN 0027-0644. doi: 10.1175/1520-0493(2001)129(2426:UNCATR)2.0.CO;2.
- S. R. Haughian, P. J. Burton, S. W. Taylor, and C. L. Curry. Expected effects of climate change on forest disturbance regimes in British Columbia. *BC Journal of Ecosystems and Management*, 13(1):16–39, 2012. ISSN 15945685.
- S. C. Herring, N. Christidis, A. Hoell, J. P. Kossin, C. J. S. Iii, and P. a. Stott. Explaining Extreme Events of 2016 From a Climate Perspective. *Bulletin of the American Meteorological Society*, 98(12):S1–S157, 2017. ISSN 0003-0007. doi: 10.1175/BAMS-D-17-0118.1. URL http://www.ametsoc.net/eee/2016/2016_bams_eee_high_res.pdf.

- K. I. Hodges, R. W. Lee, and L. Bengtsson. A comparison of extratropical cyclones in recent reanalyses ERA-Interim, NASA MERRA, NCEP CFSR, and JRA-25. *Journal of Climate*, 24:4888–4906, 2011. ISSN 08948755. doi: 10.1175/2011JCLI4097.1.
- M. Hollander, D. A. Wolfe, and E. Chicken. *Nonparametric Statistical Methods*. John Wiley and Sons, Hoboken, New Jersey, 3rd edition, 2013. ISBN 978-0-471-19045-5.
- J. Hosking. L-Moments : Analysis and Estimation of Distributions Using Linear Combinations of Order Statistics. *Journal of the Royal Statistical Society - Series B*, 52(1):105–124, 1990.
- J. Hosking, J. Wallis, and E. Woo. Estimation of the generalized extreme value distribution by the method of probability weighted moments. *Technometrics*, 27(3):251–261, 1985. ISSN 0040-1706. doi: dx.doi.org/10.1080/00401706.1985.10488049.
- D. Hou, Y. Zhu, X. Zhou, R. Wobus, J. Peng, Y. Luo, and B. Cui. The 2015 Upgrade of NCEP’s Global Ensemble Forecast System (GEFS). In *27th Conf. on Weather Analysis and Forecasting/23rd Conf. on Numerical Weather Prediction*, pages Amer. Meteor. Soc., 2A.6., Chicago, IL, jun 2015. AMS. URL <https://ams.confex.com/ams/27WAF23NWP/webprogram/Paper273406.html>.
- C. Hunter, R. Moore, and I. McKendry. Evaluation of the north american regional reanalysis (narr) precipitation fields in a topographically complex domain. *Hydrological Sciences Journal*, 0(ja):null, 2019.
- K. Ikeda, R. Rasmussen, C. Liu, D. Gochis, D. Yates, F. Chen, M. Tewari, M. Barlage, J. Dudhia, K. Miller, K. Arsenault, V. Grubišić, G. Thompson, and E. Guttman. Simulation of seasonal snowfall over Colorado. *Atmospheric Research*, 97(4):462–477, sep 2010. ISSN 0169-8095. doi: 10.1016/J.ATMOSRES.2010.04.010. URL <https://www.sciencedirect.com/science/article/pii/S0169809510001043>.
- P. L. Jackson. Surface winds during an intense outbreak of arctic air in Southwestern British Columbia. *Atmosphere-Ocean*, 34(2):285–311, jun 1996. ISSN 0705-5900. doi: 10.1080/07055900.1996.9649566.
- A. F. Jenkinson. The frequency distribution of the annual maximum (or minimum) values of meteorological elements. *Quarterly Journal of the*

Bibliography

- Royal Meteorological Society*, 81(348):158–171, apr 1955. ISSN 00359009. doi: 10.1002/qj.49708134804.
- P. Jones, S. Raper, B. Santer, B. Cherry, C. Goodess, R. Bradley, H. Diaz, P. Kelly, and T. Wigley. A grid point surface air temperature data set for the Northern Hemisphere. Technical Report DOE/EV/10098-2, United States Department of Energy, Washington, D.C., 1985.
- P. D. Jones, M. New, D. Parker, S. Martin, and I. G. Rigor. Surface air temperature and its changes over the past 150 years. *Reviews of Geophysics*, 37(2):173–199, 1999.
- R. W. Jones, I. Renfrew, A. Orr, B. G. M. Webber, D. M. Holland, and M. A. Lazzara. Evaluation of four global reanalysis products using in situ observations in the Amundsen Sea Embayment, Antarctica. *J. Geophys. Res. Atmos.*, 121, 2016. ISSN 2169897X. doi: 10.1002/2016JD025476.
- H.-m. H. Juang and M. Kanamitsu. The NMC Nested Regional Spectral Model. *Monthly Weather Review*, 122(January):3–26, 1994.
- N. W. Junker, J. E. Hoke, B. E. Sullivan, K. F. Brill, and F. J. Hughes. Seasonal geographic variations in quantitative precipitation prediction by NMC’s nested-grid model and medium-range forecast model. *Wea. Forecasting*, 7:410–429, 1992.
- T. R. Karl, P. Y. Groisman, R. W. Knight, and R. R. Heim. Recent variations of snow cover and snowfall in North America and their relation to precipitation and temperature variations. *Journal of Climate*, 6(7):1327–1344, 1993. ISSN 08948755. doi: 10.1175/1520-0442(1993)006<1327:RVOSCA>2.0.CO;2.
- E. J. Kendon, N. M. Roberts, H. J. Fowler, M. J. Roberts, S. C. Chan, and C. A. Senior. Heavier summer downpours with climate change revealed by weather forecast resolution model. *Nature Climate Change*, 4(7):570–576, jul 2014. ISSN 1758-678X. doi: 10.1038/nclimate2258.
- V. V. Kharin and F. W. Zwiers. Changes in the extremes in an ensemble of transient climate simulations with a coupled atmosphere-ocean GCM. *Journal of Climate*, 13(21):3760–3788, 2000. ISSN 08948755. doi: 10.1175/1520-0442(2000)013<3760:CITEIA>2.0.CO;2.
- S. Kobayashi, Y. Ota, Y. Harada, and A. Ebata. The JRA-55 Reanalysis: General Specifications and Basic Characteristics. *Journal of the Meteorological Society of Japan*, 93(1):5–48, 2015.

Bibliography

- R. D. Koster, W. McCarty, L. Coy, R. Gelaro, A. Huang, D. Merkova, E. B. Smith, M. Sienkiewicz, and K. Wargan. MERRA-2 Input Observations: Summary and Assessment. Technical Report NASA/TM-2016-104606, National Aeronautics and Space Administration - NASA, Greenbelt, 2016.
- C. K. B. Krishnamurthy, U. Lall, H.-H. Kwon, C. K. B. Krishnamurthy, U. Lall, and H.-H. Kwon. Changing Frequency and Intensity of Rainfall Extremes over India from 1951 to 2003. *Journal of Climate*, 22(18):4737–4746, sep 2009. ISSN 0894-8755. doi: 10.1175/2009JCLI2896.1.
- S. Kumar, V. Merwade, J. L. Kinter, and D. Niyogi. Evaluation of temperature and precipitation trends and long-term persistence in CMIP5 twentieth-century climate simulations. *Journal of Climate*, 26(12):4168–4185, 2013. ISSN 08948755. doi: 10.1175/JCLI-D-12-00259.1.
- M. Kunz and C. Kottmeier. Orographic Enhancement of Precipitation over Low Mountain Ranges. Part II: Simulations of Heavy Precipitation Events over Southwest Germany. *Journal Of Applied Meteorology And Climatology*, 45:1041–1055, 2006. ISSN 1558-8424. doi: 10.1175/JAM2390.1.
- M. Laanela. Vancouver mayor absent as councillor calls for snow removal inquiry — CBC News, 2016. URL <https://www.cbc.ca/news/canada/british-columbia/vancouver-snow-removal-1.3922282>.
- R. Lader, U. S. Bhatt, J. E. Walsh, T. S. Rupp, and P. a. Bieniek. Two-meter temperature and precipitation from atmospheric reanalysis evaluated for Alaska. *Journal of Applied Meteorology and Climatology*, 55(4):901–922, 2016. ISSN 15588432. doi: 10.1175/JAMC-D-15-0162.1.
- F. Lalaurette. Early detection of abnormal weather conditions using a probabilistic extreme forecast index. *Quarterly Journal of the Royal Meteorological Society*, 129(594):3037–3057, 2003. ISSN 00359009. doi: 10.1256/qj.02.152.
- M. Leadbetter, G. Lindgren, and H. Rootzen. *Extremes and Related Properties of Random Sequences and Processes*. Springer Verlag, 1983. ISBN 9780387775005. doi: 10.1007/978-0-387-98135-2.
- R. Lindsay, M. Wensnahan, a. Schweiger, and J. Zhang. Evaluation of seven different atmospheric reanalysis products in the arctic. *Journal of Climate*, 27(7):2588–2606, 2014. ISSN 08948755. doi: 10.1175/JCLI-D-13-00014.

Bibliography

- G. Luymes. B.C. weather: Winter storm blows over in time for Family Day — Vancouver Sun, 2017. URL <https://vancouver.sun.com/news/local-news/winter-storm-blows-over-in-time-for-family-day>.
- M. MacMahon. Fraser Valley hit hardest by snowfall, more on the way - NEWS 1130, 2017. URL <https://www.citynews1130.com/2017/02/06/fraser-valley-hit-hardest-snowfall-way/>.
- H. B. Mann. Non-parametric tests against trend. *Econometrica*, 13(3): 245–259, 1945. ISSN 0168-6054. doi: 10.1016/j.annrmp.2004.07.001.
- M. E. Mann and K. A. Emanuel. Atlantic hurricane trends linked to climate change. *Eos, Transactions American Geophysical Union*, 87(24):233, jun 2006. ISSN 0096-3941. doi: 10.1029/2006EO240001.
- D. McCollor and R. Stull. Hydrometeorological Accuracy Enhancement via Postprocessing of Numerical Weather Forecasts in Complex Terrain. *Weather and Forecasting*, 23:131–144, 2008a. ISSN 0882-8156. doi: 10.1175/2008WAF2222177.1. URL <http://journals.ametsoc.org/doi/abs/10.1175/2008WAF2222177.1>.
- D. McCollor and R. Stull. Hydrometeorological Short-Range Ensemble Forecasts in Complex Terrain. Part I: Meteorological Evaluation. *Weather and Forecasting*, 23(4):533–556, 2008b. ISSN 0882-8156. doi: 10.1175/2008WAF2007063.1. URL <http://journals.ametsoc.org/doi/abs/10.1175/2008WAF2007063.1>.
- J. McElroy. Extended cold snap putting a strain on services throughout Metro Vancouver — CBC News, 2016. URL <https://www.cbc.ca/news/canada/british-columbia/extended-cold-snap-putting-a-strain-on-services-throughout-metro-vancouver-1.3897474>.
- J. McElroy. Vancouver in its longest cold snap in over 30 years — CBC News, 2017a. URL <https://www.cbc.ca/news/canada/british-columbia/when-vancouver-had-winter-1.3918910>.
- J. McElroy. Weather closes every highway linking Lower Mainland to rest of B.C. — CBC News, 2017b. URL <https://www.cbc.ca/news/canada/british-columbia/weather-closes-every-highway-linking-lower-mainland-to-rest-of-b-c-1.3975497>.

- É. Mekis. J3.7 Adjustments for trace measurements in Canada. In *15th Conference on Applied Climatology, Savannah, Georgia, USA, 20-24 June 2005*, 2005.
- É. Mekis and R. Brown. Derivation of an adjustment factor map for the estimation of the water equivalent of snowfall from ruler measurements in Canada. *Atmosphere - Ocean*, 48(4):284–293, 2010. ISSN 07055900. doi: 10.3137/AO1104.2010.
- E. Mekis and W. D. Hogg. Rehabilitation and analysis of Canadian daily precipitation time series. *Atmosphere-Ocean*, 37(1):53–85, 1999. ISSN 0705-5900. doi: 10.1080/07055900.1999.9649621.
- É. Mekis and L. a. Vincent. An overview of the second generation adjusted daily precipitation dataset for trend analysis in Canada. *Atmosphere - Ocean*, 49(2):163–177, 2011. ISSN 07055900. doi: 10.1080/07055900.2011.583910.
- J. R. Metcalfe, B. Routledge, and K. Devine. Rainfall measurement in Canada: Changing observational methods and archive adjustment procedures. *Journal of Climate*, 10(1):92–101, 1997. ISSN 08948755. doi: 10.1175/1520-0442(1997)010<0092:RMICCO>2.0.CO;2.
- Meteorological Service of Canada. *MANOBS - manual of surface weather observations*. 2015. ISBN 9781100254456.
- M. Meuse. Fraser Valley weather the worst veteran CBC cameraman has ever seen — CBC News, 2017. URL <https://www.cbc.ca/news/canada/british-columbia/freezing-rain-worst-ever-1.3974244>.
- A. Moberg and P. D. Jones. Trends in indices for extremes in daily temperature and precipitation in central and western Europe, 1901-99. *International Journal of Climatology*, 25(9):1149–1171, 2005. ISSN 08998418. doi: 10.1002/joc.1163.
- P. a. Mooney, F. J. Mulligan, and R. Fealy. Comparison of ERA-40, ERA-Interim and NCEP/NCAR reanalysis data with observed surface air temperatures over Ireland. *International Journal of Climatology*, 31(4):545–557, 2011. ISSN 08998418. doi: 10.1002/joc.2098.
- R. D. Moore, D. L. Spittlehouse, P. H. Whitfield, and K. Stahl. *Chapter 3 - Weather and Climate*. Number November. Ministry of Forests and Range, Research Branch, Victoria, B.C. and FORREX Forest Research Extension Partnership, Kamloops, BC, 2008.

Bibliography

- P. W. Mote, A. F. Hamlet, M. P. Clark, and D. P. Lettenmaier. Declining mountain snowpack in western north America. *Bulletin of the American Meteorological Society*, 86(1):39–49, 2005. ISSN 00030007. doi: 10.1175/BAMS-86-1-39.
- A. H. Murphy and R. L. Winkler. A General Framework for Forecast Verification. *Monthly Weather Review*, 115(7):1330–1338, jul 1987. ISSN 0027-0644. doi: 10.1175/1520-0493(1987)115<1330:AGFFV>2.0.CO;2. URL [http://dx.doi.org/10.1175/1520-0493\(1987\)115{ }3C1330:AGFFV{ }3E2.0.CO2](http://dx.doi.org/10.1175/1520-0493(1987)115{ }3C1330:AGFFV{ }3E2.0.CO2).
- T. Nipen and R. Stull. Calibrating probabilistic forecasts from an NWP ensemble. *Tellus, Series A: Dynamic Meteorology and Oceanography*, 63(5):858–875, 2011. ISSN 02806495. doi: 10.1111/j.1600-0870.2011.00535.x.
- P. Odon, G. West, and R. Stull. Reanalyses Performance across British Columbia . Part II : Evaluation of Extreme Precipitation. *Journal of Applied Meteorology and Climatology*, (in press).
- P. Odon, G. West, and R. Stull. Vancouver Fall and Winter 2016/17: How Bad Was It? *CMOS Bulletin SCMO*, 45(4):9–12, 2017.
- P. Odon, G. West, and R. Stull. Evaluation of reanalyses over British Columbia. Part I: Daily and Extreme 2-m temperature. *Journal of Applied Meteorology and Climatology*, 57(9):2091–2112, 2018. ISSN 15588432. doi: 10.1175/JAMC-D-18-0058.1.
- C. of Vancouver. City crews continue work to remove remaining snow and ice — City of Vancouver, 2017. URL <https://vancouver.ca/news-calendar/city-crews-continue-work-to-remove-remaining-snow-and-ice-jan-2017.aspx>.
- B. Onoz and M. Bayazit. The power of statistical tests for trend detection. *Turkish Journal of Engineering and Environmental Sciences*, 27(4):247–251, 2003. ISSN 13000160. doi: 10.1016/j.ica.2004.10.030.
- J. Paetkau. Record breaking power usage as cold snap continues — CBC News, 2017. URL <https://www.cbc.ca/news/canada/british-columbia/bc-hydro-cold-winter-power-ted-olynek-1.3921852>.

Bibliography

- C. Parmesan, T. L. Root, and M. R. Willig. Impacts of extreme weather and climate on terrestrial biota. *Bulletin of the American Meteorological Society*, 81(3):443–450, 2000. ISSN 0003-0007. doi: 10.1175/1520-0477(2000)081<0443>
- J. A. Patz, D. Campbell-Lendrum, T. Holloway, and J. A. Foley. Impact of regional climate change on human health. *Nature*, 438(7066):310–317, 2005. ISSN 14764687. doi: 10.1038/nature04188.
- PCIC, C. Pacific Climate Impacts, V. University of, P. C. Group, and O. S. University. PCIC and Prism Climate Group, 2014: High Resolution Climatology, 2014. URL <https://pacificclimate.org/>.
- M. C. Peel, B. L. Finlayson, and T. a. McMahon. Updated world map of the Köppen-Geiger climate classification. *Hydrology and Earth System Sciences*, 11(5):1633–1644, 2007. ISSN 16077938. doi: 10.5194/hess-11-1633-2007.
- T. C. Peterson and R. S. Vose. An Overview of the Global Historical Climatology Network Temperature Database. *Bulletin of the American Meteorological Society*, 78(12):2837–2849, 1997. ISSN 00030007. doi: 10.1175/1520-0477(1997)078<2837:AOOTGH>2.0.CO;2.
- T. C. Peterson, D. R. Easterling, T. R. Karl, P. Groisman, N. Nicholls, N. Plummer, S. Torok, I. Auer, R. Boehm, D. Gullett, L. Vincent, R. Heino, H. Tuomenvirta, O. Mestre, T. Szentimrey, J. Salinger, E. J. Forland, I. Hanssen-Bauer, H. Alexandersson, P. Jones, and D. Parker. Homogeneity Adjustments of in situ Atmospheric Climate Data: A Review. *International Journal of Climatology*, 18(13):1493–1517, 1998. ISSN 08998418. doi: 10.1002/(SICI)1097-0088(19981115)18:13<1493::AID-JOC329>3.0.CO;2-T.
- D. Phillips. Canada’s Top Ten Weather Stories 2017. *CMOS Bulletin SCMO*, 46(1):13–17, 2018.
- L. S. Porth, D. C. Boes, R. a. Davis, C. a. Troendle, and R. M. King. Development of a technique to determine adequate sample size using subsampling and return interval estimation. *Journal of Hydrology*, 251(1-2):110–116, 2001. ISSN 00221694. doi: 10.1016/S0022-1694(01)00442-5.
- W. Qian, N. Jiang, and J. Du. Anomaly-Based Weather Analysis versus Traditional Total-Field-Based Weather Analysis for Depicting Regional Heavy Rain Events. *Weather and Forecasting*, 31(1):71–93, 2016. ISSN

Bibliography

0882-8156. doi: 10.1175/WAF-D-15-0074.1. URL <http://journals.ametsoc.org/doi/abs/10.1175/WAF-D-15-0074.1>.

- F. M. Ralph, P. J. Neiman, and G. a. Wick. Satellite and CALJET Aircraft Observations of Atmospheric Rivers over the Eastern North Pacific Ocean during the Winter of 1997/98. *Monthly Weather Review*, 132(7):1721–1745, 2004. ISSN 0027-0644. doi: 10.1175/1520-0493(2004)132(1721:SACAOO)2.0.CO;2.
- F. M. Ralph, P. J. Neiman, and R. Rotunno. Dropsonde Observations in Low-Level Jets over the Northeastern Pacific Ocean from CALJET-1998 and PACJET-2001: Mean Vertical-Profile and Atmospheric-River Characteristics. *Monthly Weather Review*, 133(4):889–910, 2005. ISSN 0027-0644. doi: 10.1175/MWR2896.1.
- R. Rasmussen, C. Liu, K. Ikeda, D. Gochis, D. Yates, F. Chen, M. Tewari, M. Barlage, J. Dudhia, W. Yu, K. Miller, K. Arsenault, V. Grubišić, G. Thompson, and E. Gutmann. High-resolution coupled climate runoff simulations of seasonal snowfall over Colorado: A process study of current and warmer climate. *Journal of Climate*, 24(12):3015–3048, 2011. ISSN 08948755. doi: 10.1175/2010JCLI3985.1.
- A. Ravishankar, G. Meehl, and L. F. Bosart. Attribution of Extreme Weather Events in the Context of Climate Change. *National Academy of Sciences*, page 200, 2016. doi: 10.17226/21852.
- R. H. Reichle, Q. Liu, R. D. Koster, C. S. Draper, S. P. Mahanama, and G. S. Partyka. Land Surface Precipitation in MERRA-2. *Journal of Climate*, 30(5):1643–1664, 2017. ISSN 08948755. doi: 10.1175/JCLI-D-16-0570.1.
- M. M. Rienecker, M. J. Suarez, R. Gelaro, R. Todling, and J. Bacmeister. MERRA: NASA’s Modern-Era Retrospective Analysis for Research and Applications. *Journal of Climate*, 24:3624–3648, 2011.
- C. Rosenzweig, a. Iglesias, X. B. Yang, P. R. Epstein, and E. Chivian. Climate change and extreme weather events: Implications for food production, plant diseases, and pests. *Global Change and Human Health*, 2: 90–104, 2001.
- A. Ruiz-Barradas and S. Nigam. Warm season rainfall variability over the U.S. Great Plains in observations, NCEP and ERA-40 reanalyses, and NCAR and NASA atmospheric model simulations. *Journal of Climate*, 18(11):1808–1830, 2005. ISSN 08948755. doi: 10.1175/JCLI3343.1.

Bibliography

- S. Saha, S. Nadiga, C. Thiaw, J. Wang, W. Wang, Q. Zhang, H. M. V. D. Dool, H.-L. Pan, S. Moorthi, D. Behringer, D. Stokes, M. Pena, S. Lord, G. White, W. Ebisuzaki, P. Peng, and P. Xie. The NCEP Climate Forecast System. *Journal of Climate*, 19:3483–3517, 2006. doi: 10.1175/2010Bams3001.1.
- S. Saha, S. Moorthi, H.-L. Pan, X. Wu, J. Wang, S. Nadiga, P. Tripp, R. Kistler, J. Woollen, D. Behringer, H. Liu, D. Stokes, R. Grumbine, G. Gayno, J. Wang, Y.-T. Hou, H.-Y. Chuang, H.-M. H. Juang, J. Sela, M. Iredell, R. Treadon, D. Kleist, P. Van Delst, D. Keyser, J. Derber, M. Ek, J. Meng, H. Wei, R. Yang, S. Lord, H. Van Den Dool, A. Kumar, W. Wang, C. Long, M. Chelliah, Y. Xue, B. Huang, J.-K. Schemm, W. Ebisuzaki, R. Lin, P. Xie, M. Chen, S. Zhou, W. Higgins, C.-Z. Zou, Q. Liu, Y. Chen, Y. Han, L. Cucurull, R. W. Reynolds, G. Rutledge, and M. Goldberg. The NCEP Climate Forecast System Reanalysis. *Bulletin of the American Meteorological Society*, 91(8):1015–1057, aug 2010. ISSN 0003-0007. doi: 10.1175/2010BAMS3001.1.
- S. Saha, S. Moorthi, X. Wu, J. Wang, S. Nadiga, P. Tripp, D. Behringer, Y. T. Hou, H. Y. Chuang, M. Iredell, M. Ek, J. Meng, R. Yang, M. P. Mendez, H. Van Den Dool, Q. Zhang, W. Wang, M. Chen, and E. Becker. The NCEP climate forecast system version 2. *Journal of Climate*, 27(6): 2185–2208, 2014. ISSN 08948755. doi: 10.1175/JCLI-D-12-00823.1.
- J. Saltman. BY THE NUMBERS: Winter’s impact costly on Alex Fraser and Port Mann bridges, 2017. URL <https://vancouver.sun.com/news/local-news/by-the-numbers-winters-costly-impact-on-alex-fraser-and-port-mann-bridges>.
- J. Schaefer. The Critical Success as an Indicator of Warning Skill. *Weather and Forecasting*, 5:570–575, 1990.
- M. Scheuerer and T. M. Hamill. Statistical Postprocessing of Ensemble Precipitation Forecasts by Fitting Censored, Shifted Gamma Distributions*. *Monthly Weather Review*, 143(11):4578–4596, 2015. ISSN 0027-0644. doi: 10.1175/MWR-D-15-0061.1. URL <http://journals.ametsoc.org/doi/10.1175/MWR-D-15-0061.1>.
- M. Schirmer and B. Jamieson. Verification of analysed and forecasted winter precipitation in complex terrain. *Cryosphere*, 9(2):587–601, 2015. ISSN 19940424. doi: 10.5194/tc-9-587-2015.

Bibliography

- J. K. Searcy and C. H. Hardison. Double-Mass Curves. *Manual of Hydrology: Part 1. General Surface-Water Techniques*, page 41, 1960.
- M. C. Serreze, A. P. Barrett, and F. Lo. Northern High-Latitude Precipitation as Depicted by Atmospheric Reanalyses and Satellite Retrievals. *Monthly Weather Review*, 133(12):3407–3430, 2005. ISSN 0027-0644. doi: 10.1175/MWR3047.1. URL <https://journals.ametsoc.org/doi/pdf/10.1175/MWR3047.1>.
- D. Siuta, G. West, R. Stull, and T. Nipen. Calibrated Probabilistic Hub-Height Wind Forecasts in Complex Terrain. *Weather and Forecasting*, 32(2):555–577, 2017. ISSN 0882-8156. doi: 10.1175/WAF-D-16-0137.1. URL <http://journals.ametsoc.org/doi/10.1175/WAF-D-16-0137.1>.
- B. L. Smith, S. E. Yuter, P. J. Neiman, and D. E. Kingsmill. Water Vapor Fluxes and Orographic Precipitation over Northern California Associated with a Landfalling Atmospheric River. *Monthly Weather Review*, 138(1): 74–100, 2010. ISSN 0027-0644. doi: 10.1175/2009MWR2939.1. URL <http://journals.ametsoc.org/doi/abs/10.1175/2009MWR2939.1>.
- R. B. Smith, Q. Jiang, M. G. Fearon, P. Tabary, M. Dorninger, J. D. Doyle, and R. Benoit. Orographic precipitation and air mass transformation: An Alpine example. *Quarterly Journal of the Royal Meteorological Society*, 129(588 PART B):433–454, 2003. ISSN 00359009. doi: 10.1256/qj.01.212.
- G. J. Spagnol. Rocketsonde buoy system : observing system simulation experiments. *Retrospective Theses and Dissertations, 1919-2007*, 2005. doi: 10.14288/1.0052550.
- K. Stahl, R. D. Moore, J. a. Floyer, M. G. Asplin, and I. G. McKendry. Comparison of approaches for spatial interpolation of daily air temperature in a large region with complex topography and highly variable station density. *Agricultural and Forest Meteorology*, 139(3-4):224–236, 2006. ISSN 01681923. doi: 10.1016/j.agrformet.2006.07.004.
- L. Stefanova, V. Misra, S. Chan, M. Griffin, J. J. O'Brien, and T. J. Smith III. A proxy for high-resolution regional reanalysis for the Southeast United States: assessment of precipitation variability in dynamically downscaled reanalyses. *Climate Dynamics*, 38(11-12):2449–2466, jun 2012. ISSN 0930-7575. doi: 10.1007/s00382-011-1230-y. URL <http://link.springer.com/10.1007/s00382-011-1230-y>.

- D. B. Stephenson. Use of the “Odds Ratio” for Diagnosing Forecast Skill. *Weather and Forecasting*, 15(2):221–232, 2000. ISSN 0882-8156. doi: 10.1175/1520-0434(2000)015<0221:UOTORF>2.0.CO;2. URL <http://journals.ametsoc.org/doi/abs/10.1175/1520-0434%7D282000%7D29015%7D3C0221%7D3AUOTORF%7D3E2.0.CO%7D3B2>.
- D. B. Stephenson, B. Casati, C. Ferro, and C. Wilson. The extreme dependency score: a non-vanishing measure for forecast of rare events. *Meteorological Applications*, 15:41–50, 2008. ISSN 13504827. doi: 10.1002/met.
- J. Stoklosa, C. Daly, S. D. Foster, M. B. Ashcroft, and D. I. Warton. A climate of uncertainty: Accounting for error in climate variables for species distribution models. *Methods in Ecology and Evolution*, 6(4):412–423, 2015. ISSN 2041210X. doi: 10.1111/2041-210X.12217.
- F. Sun, M. L. Roderick, and G. D. Farquhar. Rainfall statistics, stationarity, and climate change. *Proceedings of the National Academy of Sciences*, (23):201705349, 2018a. ISSN 0027-8424. doi: 10.1073/pnas.1705349115. URL <http://www.pnas.org/lookup/doi/10.1073/pnas.1705349115>.
- Q. Sun, C. Miao, Q. Sun, and C. Miao. Extreme Rainfall (R20mm, RX5day) in Yangtze–Huai, China, in June–July 2016: The Role of ENSO and Anthropogenic Climate Change. *Bulletin of the American Meteorological Society*, 99(1):S102–S106, jan 2018b. ISSN 0003-0007. doi: 10.1175/BAMS-D-17-0091.1. URL <http://journals.ametsoc.org/doi/10.1175/BAMS-D-17-0091.1>.
- W. Taesombat and N. Sriwongsitanon. Areal rainfall estimation using spatial interpolation techniques. *ScienceAsia*, 35(3):268–275, 2009. ISSN 15131874. doi: 10.2306/scienceasia1513-1874.2009.35.268.
- Y. Takeuchi, H. Eito, T. Fujieda, and T. Fujita. Outline of the operational numerical weather prediction of the Japan Meteorological Agency. Technical Report March, Japan Meteorological Agency, 2013.
- O. Talagrand, B. Vautard, and B. Strauss. Evaluation of probabilistic prediction systems. In *Proc. Workshop on Predictability*, pages ECMWF, 1–25, Reading, UK, 1999.
- H. Theil. A rank-invariant method of linear and polynomial regression analysis. *Proc. Ned. Akad. Wet*, 53(2):1397–1412, 1950. ISSN 2008-8140 (Print). doi: 10.1007/978-94-011-2546-8.

- J.-N. Thépaut. Assimilating only surface pressure observations in 3D- and 4D-Var. In *Proceedings of the ECMWF workshop on atmospheric reanalysis*, number June, pages 1–44, Reading, UK, 2006.
- T. L. Thorarinsdottir and T. Gneiting. Probabilistic Forecasts of Wind Speed : Ensemble Model Output Statistics using Heteroskedastic Censored Regression. *J.R. Statist. Soc. A*, 173:371–388, 2010.
- P. E. Thornton, S. W. Running, and M. A. White. Generating surfaces of daily meteorological variables in complex terrain. *Journal of Hydrology*, 190:214–251, 1997.
- Z. Toth and E. Kalnay. Ensemble Forecasting at NCEP and the Breeding Method. *Monthly Weather Review*, 125(12):3297–3319, 1997. ISSN 0027-0644. doi: 10.1175/1520-0493(1997)125<3297:EFANAT>2.0.CO;2. URL <http://journals.ametsoc.org/doi/abs/10.1175/1520-0493%281997%29125%3C3297%3A%EFANAT%3E2.0.CO%3B2>.
- K. E. Trenberth and L. Smith. The mass of the atmosphere: A constraint on global analyses. *Journal of Climate*, 18(6):864–875, 2005. ISSN 08948755. doi: 10.1175/JCLI-3299.1.
- J. W. Trubilowicz, J. M. Shea, G. Jost, and R. D. Moore. Suitability of North American Regional Reanalysis (NARR) output for hydrologic modelling and analysis in mountainous terrain. 2347(March):2332–2347, 2016. doi: 10.1002/hyp.10795.
- S. M. Uppala, P. W. Kållberg, A. J. Simmons, U. Andrae, V. da Costa Bechtold, M. Fiorino, J. K. Gibson, J. Haseler, a. Hernandez, G. a. Kelly, X. Li, K. Onogi, S. Saarinen, N. Sokka, R. P. Allan, E. Andersson, K. Arpe, M. a. Balmaseda, a. C. Beljaars, L. van de Berg, J. Bidlot, N. Bormann, S. Caires, F. Chevallier, a. Dethof, M. Dragosavac, M. Fisher, M. Fuentes, S. Hagemann, E. Hólm, B. J. Hoskins, L. Isaksen, P. a.E.M. Janssen, R. Jenne, a. P. McNally, J. F. Mahfouf, J. J. Morcrette, N. a. Rayner, R. W. Saunders, P. Simon, a. Sterl, K. E. Trenberth, a. Untch, D. Vasiljevic, P. Viterbo, and J. Woollen. The ERA-40 re-analysis. *Quarterly Journal of the Royal Meteorological Society*, 131(612):2961–3012, 2005. ISSN 00359009. doi: 10.1256/qj.04.176.
- G. J. van Oldenborgh, F. J. Doblas Reyes, S. S. Drijfhout, and E. Hawkins. Reliability of regional climate model trends. *Environmental Research Letters*, 8(1):014055, mar 2013. ISSN 1748-9326. doi: 10.1088/1748-9326/8/

Bibliography

1/014055. URL <http://stacks.iop.org/1748-9326/8/i=1/a=014055?key=crossref.2aa3fd77fc365abd77e9d8773bb66738>.

- C. Vancouver. Wicked winter leaves behind 15,000 potholes in Vancouver — CTV Vancouver News, 2017. URL <https://bc.ctvnews.ca/wicked-winter-leaves-behind-15-000-potholes-in-vancouver-1.3348547?autoPlay=true>.
- S. M. Vicente-Serrano, S. Beguería, J. I. López-Moreno, M. a. García-Vera, and P. Stepanek. A complete daily precipitation database for northeast Spain: Reconstruction, quality control, and homogeneity. *International Journal of Climatology*, 30(8):1146–1163, 2010. ISSN 08998418. doi: 10.1002/joc.1850.
- L. a. Vincent. A technique for the identification of inhomogeneities in Canadian temperature series. *Journal of Climate*, 11(5):1094–1104, 1998. ISSN 08948755. doi: 10.1175/1520-0442(1998)011<1094:ATFTIO>2.0.CO;2.
- L. a. Vincent and D. W. Gullett. Canadian historical and homogeneous temperature datasets for climate change analyses. *International Journal of Climatology*, 19(12):1375–1388, 1999. ISSN 08998418. doi: 10.1002/(SICI)1097-0088(199910)19:12<1375::AID-JOC427>3.0.CO;2-0.
- L. a. Vincent and É. Mekis. Changes in Daily and Extreme Temperature and Precipitation Indices for Canada over the Twentieth Century. *Atmosphere-Ocean*, 44(2):177–193, 2006. ISSN 0705-5900. doi: 10.3137/ao.440205.
- L. a. Vincent, X. Zhang, B. R. Bonsal, and W. D. Hogg. Homogenization of daily temperatures over Canada. *Journal of Climate*, 15(11):1322–1334, 2002. ISSN 08948755. doi: 10.1175/1520-0442(2002)015<1322:HODTOC>2.0.CO;2.
- L. a. Vincent, E. J. Milewska, R. Hopkinson, and L. Malone. Bias in minimum temperature introduced by a redefinition of the climatological day at the canadian synoptic stations. *Journal of Applied Meteorology and Climatology*, 48(10):2160–2168, 2009. ISSN 15588424. doi: 10.1175/2009JAMC2191.1.
- L. a. Vincent, X. L. Wang, E. J. Milewska, H. Wan, F. Yang, and V. Swail. A second generation of homogenized Canadian monthly surface air temperature for climate trend analysis. *Journal of Geophysical Research Atmospheres*, 117(17):1–13, 2012. ISSN 01480227. doi: 10.1029/2012JD017859.

- W. Wang, P. Xie, S. H. Yoo, Y. Xue, A. Kumar, and X. Wu. An assessment of the surface climate in the NCEP climate forecast system reanalysis. *Climate Dynamics*, 37(7-8):1601–1620, 2011. ISSN 09307575. doi: 10.1007/s00382-010-0935-7.
- X. Wang, D. Parrish, D. Kleist, and J. Whitaker. GSI 3DVar-Based Ensemble-Variational Hybrid Data Assimilation for NCEP Global Forecast System: Single-Resolution Experiments. *Monthly Weather Review*, 141(11):4098–4117, 2013. ISSN 0027-0644. doi: 10.1175/MWR-D-12-00141.1. URL <http://journals.ametsoc.org/doi/abs/10.1175/MWR-D-12-00141.1>.
- X. L. Wang, Q. H. Wen, and Y. Wu. Penalized maximal t test for detecting undocumented mean change in climate data series. *Journal of Applied Meteorology and Climatology*, 46(6):916–931, 2007. ISSN 15588424. doi: 10.1175/JAM2504.1.
- M. Wei, Z. Toth, R. Wobus, and Y. Zhu. Initial perturbations based on the ensemble transform (ET) technique in the NCEP global operational forecast system. *Tellus, Series A: Dynamic Meteorology and Oceanography*, 60 A(1):62–79, 2008. ISSN 02806495. doi: 10.1111/j.1600-0870.2007.00273.x.
- A. T. Werner and A. J. Cannon. Hydrologic extremes - An intercomparison of multiple gridded statistical downscaling methods. *Hydrology and Earth System Sciences*, 20(4):1483–1508, 2016. ISSN 16077938. doi: 10.5194/hess-20-1483-2016.
- S. Westra, L. V. Alexander, F. W. Zwiers, S. Westra, L. V. Alexander, and F. W. Zwiers. Global Increasing Trends in Annual Maximum Daily Precipitation. *Journal of Climate*, 26(11):3904–3918, jun 2013. ISSN 0894-8755. doi: 10.1175/JCLI-D-12-00502.1. URL <http://journals.ametsoc.org/doi/abs/10.1175/JCLI-D-12-00502.1>.
- T. Weusthoff, F. Ament, M. Arpagaus, and M. W. Rotach. Assessing the Benefits of Convection-Permitting Models by Neighborhood Verification: Examples from MAP D-PHASE. *Monthly Weather Review*, 138(9):3418–3433, 2010. ISSN 0027-0644. doi: 10.1175/2010MWR3380.1. URL <http://journals.ametsoc.org/doi/abs/10.1175/2010MWR3380.1>.
- J. S. Whitaker, G. P. Compo, and J.-N. Thépaut. A Comparison of Variational and Ensemble-Based Data Assimilation Systems for Reanalysis of Sparse Observations. *Monthly Weather Review*, 137(6):1991–1999, 2009. ISSN 0027-0644. doi: 10.1175/2008MWR2781.1.

- T. White, J. Wolf, F. Anslow, A. Werner, and R. Creative. Indicators of Climate Change for British Columbia: 2016 Update. Technical report, Ministry of Environment - BC, 2016.
- D. Wilks. *Statistical Methods in the Atmospheric Sciences*, volume 100. Elsevier-International geophysics series, 3rd edition, 2011. ISBN 9780123850225. doi: 10.1016/B978-0-12-385022-5.00022-1.
- D. Wilks. The Stippling Shows Statistically Significant Grid Points: How Research Results are Routinely Overstated and Overinterpreted, and What to Do about It. *Bulletin of the American Meteorological Society*, 97:2263–2274, 2016. doi: 10.1175/BAMS-D-00267.1.
- P. Xie and P. a. Arkin. Global Precipitation: A 17-Year Monthly Analysis Based on Gauge Observations, Satellite Estimates, and Numerical Model Outputs. *Bulletin of the American Meteorological Society*, 78(11): 2539–2558, 1997. ISSN 00030007. doi: 10.1175/1520-0477(1997)078<2539:GPAYMA>2.0.CO;2.
- P. Xie, M. Chen, S. Yang, A. Yatagai, T. Hayasaka, Y. Fukushima, and C. Liu. A Gauge-Based Analysis of Daily Precipitation over East Asia. *Journal of Hydrometeorology*, 8(3):607–626, 2007. ISSN 1525-755X. doi: 10.1175/JHM583.1.
- S. Yue, P. Pilon, B. Phinney, and G. Cavadias. The influence of autocorrelation on the ability to detect trend in hydrological series. *Hydrological Processes*, 16(9):1807–1829, 2002. ISSN 08856087. doi: 10.1002/hyp.1095.
- X. Zhang, L. a. Vincent, W. D. Hogg, and A. Niitsoo. Temperature and precipitation trends in Canada during the 20th century. *Atmosphere - Ocean*, 38(3):395–429, 2000. ISSN 07055900. doi: 10.1080/07055900.2000.9649654.
- X. Zhang, W. D. Hogg, and É. Mekis. Spatial and Temporal Characteristics of Heavy Precipitation Events over Canada. *Journal of Climate*, 14(9):1923–1936, 2001. ISSN 0894-8755. doi: 10.1175/1520-0442(2001)014<1923:SATCOH>2.0.CO;2.
- X. Zhou, Y. Zhu, D. Hou, Y. Luo, J. Peng, and R. Wobus. Performance of the New NCEP Global Ensemble Forecast System in a Parallel Experiment. *Weather and Forecasting*, pages WAF–D–17–0023.1, 2017. ISSN 0882-8156. doi: 10.1175/WAF-D-17-0023.1. URL <http://journals.ametsoc.org/doi/10.1175/WAF-D-17-0023.1>.

- H. Zhu, M. C. Wheeler, A. H. Sobel, D. Hudson, H. Zhu, M. C. Wheeler, A. H. Sobel, and D. Hudson. Seamless Precipitation Prediction Skill in the Tropics and Extratropics from a Global Model. *Monthly Weather Review*, 142(4):1556–1569, apr 2014. ISSN 0027-0644. doi: 10.1175/MWR-D-13-00222.1.
- Y. Zhu and Y. Luo. Precipitation Calibration Based on the Frequency-Matching Method. *Weather and Forecasting*, 30(5):1109–1124, 2015. ISSN 0882-8156. doi: 10.1175/WAF-D-13-00049.1.
- E. Zsoter. Recent developments in extreme weather forecasting. *ECMWF Newsletter*, Spring(107):8–17, 2006.
- F. W. Zwiers and V. V. Kharin. Estimating extremes in transient climate change simulations. *Journal of Climate*, (18):43, 2005. URL http://www.cccma.ec.gc.ca/papers/fzwiers/kz_{_}2003.pdf.
- V. V. K. F. W. Zwiers and X. Z. M. Wehner. Changes in temperature and precipitation extremes in the CMIP5 ensemble. *Climate Change*, 119: 345–357, 2013. doi: 10.1007/s10584-013-0705-8.

Appendix A

Surface weather stations

Table A.1: Surface weather stations description. Station name and abbreviation, longitude and latitude in degrees, elevation in meters, variable as either 2-m temperature (T) or 1-day accumulated precipitation (p), network and whether it is a SYNOP station.

	Station	Abbrev	Lon	Lat	Elev	Var	Network
1	Abbotsford	YXX	-122.4	49.0	59	T, p	ECCC(S)
2	Agassiz	WZA	-121.8	49.2	15	T, p	ECCC(S)
3	Alouette	ALU	-122.5	49.3	125	T, p	BCH
4	Atlin	ATL	-133.70	59.6	74	T, p	ECCC
5	Barkerville	BAR	-121.5	53.1	128	p	ECCC
6	Bella Coola	YBD	-126.6	52.4	36	T, p	ECCC(S)
7	Blind Channel	BCH	-125.5	50.4	23	T	ECCC
8	Blue River	YCP	-119.3	52.1	68	T, p	ECCC(S)
9	Castlegar	YCG	-117.6	49.3	496	T, p	ECCC(S)
10	Chatham Point	WFM	-125.4	50.3	23	p	ECCC
11	Clowhom Falls	CLO	-123.5	49.7	10	T, p	BCH
12	Comox	YQQ	-124.9	49.7	26	T, p	ECCC(S)
13	Coquitlam	CQM	-122.8	49.5	290	T, p	BCH
14	Cortes Island Tiber Bay	CIT	-124.9	50.1	15	p	ECCC
15	Cranbrook	YXC	-115.8	49.6	939	T, p	ECCC(S)
16	Creston	WJR	116.5	49.1	597	p	ECCC
17	Darfield	DAR	-120.2	51.3	412	T, p	ECCC
18	Dawson's Creek	YDQ	-120.2	55.8	655	T	ECCC(S)
19	Dease Lake	WDL	-130.0	58.4	807	T	ECCC(S)
20	Dryad Point	DAU	-128.1	52.2	4	p	ECCC
21	Duncan Kelvin Creek	DKC	-123.7	48.7	103	p	ECCC
22	Estevan Point	WEB	-126.6	49.4	7	T, p	ECCC
23	Fernie	FER	-115.1	49.5	1001	p	ECCC

Continued on next page

Appendix A. Surface weather stations

Table A.1 – continued from previous page

	Station	Abbrev	Lon	Lat	Elev	Var	Network
24	Fort Nelson	YYE	-122.6	58.8	382	T, p	ECCC(S)
25	Fort St James	VFS	-124.3	54.5	686	T, p	ECCC
26	Fort St John's	YXJ	-120.7	56.2	695	T, p	ECCC(S)
27	Germansen Landing	YGS	-124.7	55.8	766	p	ECCC
28	Glacier	GLA	-117.5	51.3	1323	T,p	ECCC
29	Gold Creek	GOC	-122.5	49.45	794	T, p	BCH
30	Golden	YGE	-117.0	51.3	785	T, p	ECCC
31	Grand Forks	GRF	-118.5	49.0	532	T	ECCC
32	Kamloops	YKA	-120.4	50.7	345	T, p	ECCC(S)
33	Kaslo	KAS	-116.9	49.9	591	T	ECCC
34	Kelowna	YLW	-119.6	49.8	417	T	ECCC(S)
35	Kelowna Quails Gate	KQG	-119.6	49.8	417	p	ECCC
36	Laidlaw	LAI	-121.6	49.4	27	p	ECCC
37	Langara	WLA	-133.1	54.3	41	T, p	ECCC
38	Little Qualicum Hatchery	VOQ	-124.5	49.4	30	p	ECCC
39	McInnes Island	WMS	-128.7	52.3	26	T , p	ECCC(S)
40	Merritt	VME	-120.8	50.1	609	T, p	ECCC
41	Mica Dam	MCD	-118.6	52.1	579	T, p	BCH
42	Nanaimo City Yard	YCD	-124.0	49.2	114	p	ECCC
43	Nass Camp	NAC	-129.0	55.2	290	p	ECCC
44	Oliver	OLI	-119.5	49.2	297	T, p	ECCC
45	Pachena Point	PAP	-125.1	48.7	37	p	ECCC
46	Penticton	YYF	-119.6	49.5	344	T, p	ECCC(S)
47	Port Alice	POA	-127.5	50.4	21	p	ECCC
48	Port Hardy	YZT	-127.4	50.7	22	T, p	ECCC(S)
49	Prince George	YXS	-122.7	53.9	691	T, p	ECCC(S)
50	Prince Rupert	YPR	-130.5	54.3	35	T	ECCC(S)
51	Princeton	YDC	-120.5	49.5	700	T, p	ECCC(S)
52	Quesnel	YQZ	-122.5	53.0	545	T	ECCC(S)
53	Quatsino	WIF	-127.7	50.5	8	T, p	ECCC
54	Quinsam River Hatchery	QUI	-125.3	50	46	T, p	ECCC

Continued on next page

Appendix A. Surface weather stations

Table A.1 – continued from previous page

	Station	Abbrev	Lon	Lat	Elev	Var	Network
55	Saanichton	SAN	-123.4	48.6	61	p	ECCC
56	Salmon Arm	WSL	-119.2	50.7	527	T, p	ECCC
57	Sandspit	YZP	-131.8	53.3	6	T, p	ECCC(S)
58	Shawnigan Lake	SHL	-123.6	48.6	138	T, p	ECCC
59	Smithers	YYD	-127.2	54.8	522	T, p	ECCC(S)
60	Stave Ridge Upper	STV	-122.4	49.6	930	T, p	BCH
61	Stewart	ZST	-130	55.9	7	T, p	ECCC(S)
62	Tatlayoko Lake	XTL	-124.4	51.7	870	T	ECCC
63	Terrace	YXT	-128.6	54.5	7	T, p	ECCC(S)
64	Tofino	YAZ	-125.8	49.1	24	p	ECCC(S)
65	Ucluelet Kennedy Camp	UKC	-125.5	48.9	30	p	ECCC
66	Vancouver	YVR	-123.2	49.2	4	T, p	ECCC(S)
67	Cheakamus Upper	CMU CMU	-123.1	50.1	880	T, p	BCH
68	Vavenby	VAV	-119.8	51.6	445	T, p	ECCC
69	Vernon	WJV	-119.3	50.3	427	T, p	ECCC(S)
70	Victoria	YYJ	-123.4	48.6	19	T, p	ECCC(S)
71	Wahleach Jones Reservoir	WAH	-121.6	49.2	641	T, p	BCH
72	Wasa	WAS	-115.6	49.8	930	p	ECCC
73	Wistaria	WIS	-126.2	53.8	863	T	ECCC
74	William's Lake	WLK	-122.1	52.2	940	T, p	ECCC(S)
75	Wolf Ridge Upper	WOL	-125.7	49.7	1490	T, p	BCH

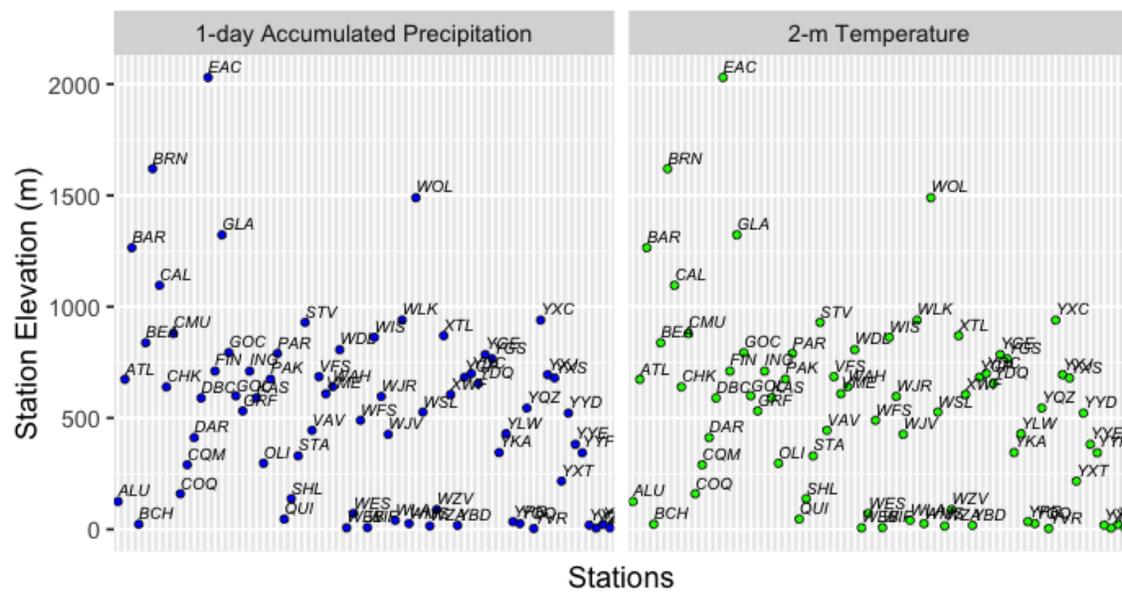


Figure A.1: Elevation of 2-m temperature and 1-day accumulated precipitation stations.

Appendix B

Train and test stations

Table B.1: Train surface weather stations description. Station name and abbreviation, longitude and latitude in degrees, elevation in meters, variable as either 2-m temperature (T) or 1-day accumulated precipitation (p), network station.

	Station	Abbrev	Lon	Lat	Elev	Var	Network
1	Alouette	ALU	-122.5	49.3	125	T	BCH
2	Barkerville	BAR	-121.5	53.1	128	p	ECCC
3	Bella Coola	YBD	-126.6	52.4	36	T, p	ECCC
4	Blue River	YCP	-119.3	52.1	68	p	ECCC
5	Chatham Point	WFM	-125.4	50.3	23	p	ECCC
6	Coquitlam	CQM	-122.8	49.5	290	T, p	BCH
7	Cortes Island Tiber Bay	CIT	-124.9	50.1	15	p	ECCC
8	Creston	WJR	116.5	49.1	597	T	ECCC
8	Darfield	DAR	-120.2	51.3	412	T, p	ECCC
10	Dawson's Creek	YDQ	-120.2	55.8	655	T	ECCC
11	Duncan Kelvin Creek	DKC	-123.7	48.7	103	p	ECCC
12	Fernie	FER	-115.1	49.5	1001	p	ECCC
13	Fort Nelson	YYE	-122.6	58.8	382	p	ECCC
14	Fort St James	VFS	-124.3	54.5	686	T, p	ECCC
15	Fort St John's	YXJ	-120.7	56.2	695	T, p	ECCC
16	Germansen Landing	YGS	-124.7	55.8	766	T	ECCC
17	Glacier	GLA	-117.5	51.3	1323	T	ECCC
18	Gold Creek	GOC	-122.5	49.45	794	T, p	BCH
19	Golden	YGE	-117.0	51.3	785	p	ECCC
19	Ingenika	ING	-125.1	56.73	711	T	BCH
20	Kamloops	YKA	-120.4	50.7	345	T, p	ECCC
21	Kaslo	KAS	-116.9	49.9	591	T	ECCC

Continued on next page

Appendix B. Train and test stations

Table B.1 – continued from previous page

	Station	Abbrev	Lon	Lat	Elev	Var	Network
22	Kelowna	YLW	-119.6	49.8	417	T	ECCC
23	Langara	WLA	-133.1	54.3	41	T, p	ECCC
24	McInnes Island	WMS	-128.7	52.3	26	T, p	ECCC
25	Mica Dam	MCD	-118.6	52.1	579	p	BCH
26	Nanaimo City Yard	YCD	-124.0	49.2	114	p	ECCC
27	Oliver	OLI	-119.5	49.2	297	T	ECCC
28	Pachena Point	PAP	-125.1	48.7	37	p	ECCC
29	Penticton	YYF	-119.6	49.5	344	T	ECCC
30	Port Alice	POA	-127.5	50.4	21	p	ECCC
31	Port Hardy	YZT	-127.4	50.7	22	T	ECCC
32	Prince Rupert	YPR	-130.5	54.3	35	T	ECCC
33	Princeton	YDC	-120.5	49.5	700	T	ECCC
34	Quesnel	YQZ	-122.5	53.0	545	T	ECCC
35	Quatsino	WIF	-127.7	50.5	8	p	ECCC
36	Quinsam River Hatchery	QUI	-125.3	50	46	T	ECCC
37	Saanichton	SAN	-123.4	48.6	61	p	ECCC
38	Salmon Arm	WSL	-119.2	50.7	527	T	ECCC
39	Sandspit	YZP	-131.8	53.3	6	p	ECCC
40	Shawnigan Lake	SHL	-123.6	48.6	138	p	ECCC
41	Smithers	YYD	-127.2	54.8	522	T, p	ECCC
42	Stave Ridge Upper	STV	-122.4	49.6	930	T, p	BCH
43	Stewart	ZST	-130	55.9	7	p	ECCC
44	Tatlayoko Lake	XTL	-124.4	51.7	870	T	ECCC
45	Ucluelet Kennedy Camp	UKC	-125.5	48.9	30	p	ECCC
46	Vancouver	YVR	-123.2	49.2	4	T	ECCC
47	Cheakamus Upper	CMU CMU	-123.1	50.1	880	p	BCH
48	Vavenby	VAV	-119.8	51.6	445	T, p	ECCC
49	Vernon	WJV	-119.3	50.3	427	T, p	ECCC
50	Victoria	YYJ	-123.4	48.6	19	T, p	ECCC
51	Wasa	WAS	-115.6	49.8	930	p	ECCC

Continued on next page

Appendix B. Train and test stations

Table B.1 – continued from previous page

	Station	Abbrev	Lon	Lat	Elev	Var	Network
52	Wistaria	WIS	-126.2	53.8	863	T	ECCC
53	William’s Lake	WLK	-122.1	52.2	940	T, p	ECCC

Table B.2: Test surface weather stations description. Station name and abbreviation, longitude and latitude in degrees, elevation in meters, variable as either 2-m temperature (T) or 1-day accumulated precipitation (p), network station.

	Station	Abbrev	Lon	Lat	Elev	Var	Network
1	Agassiz	WZA	-121.8	49.2	15	T,p	ECCC
2	Barnes Creek	BRN	-118.35	50.01	1620	T	BCH
3	Blind Channel	BCH	-125.5	50.4	23	T	ECCC
4	Blue River	YCP	-119.3	52.1	68	T	ECCC
5	Cheakamus Creek	CHK	-123.03	50.08	640	T,p	BCH
6	Comox	YQQ	-124.9	49.7	26	T,p	ECCC
7	Cranbrook	YXC	-115.8	49.6	939	T	ECCC
8	Creston	WJR	116.5	49.1	597	p	ECCC
9	Dease Lake	WDL	-130.0	58.4	807	T	ECCC
10	Duncan Kelvin Creek	DKC	-123.7	48.7	103	T,p	ECCC
11	Estevan Point	WEB	-126.6	49.4	7	T,p	ECCC
12	Fort Nelson	YYE	-122.6	58.8	382	T	ECCC
13	Golden	YGE	-117.0	51.3	785	T	ECCC
14	Goldstream	GOL	-118.6	51.67	600	p	BCH
15	Germansen Landing	YGS	-124.7	55.8	766	p	ECCC
16	Glacier	GLA	-117.5	51.3	1323	T,p	ECCC
17	Grand Forks	GRF	-118.5	49.0	532	T	ECCC
18	Kemano	KEM	-127.9	53.6	87	p	ECCC
19	Kelowna	KQG	-119.6	49.8	417	p	ECCC
20	Laidlaw	LAI	-121.6	49.4	27	p	ECCC
21	Little Qualicum Hatchery	VOQ	-124.5	49.4	30	p	ECCC
22	Little Qualicum	VOQ	-124.5	49.4	30	p	ECCC

Continued on next page

Appendix B. Train and test stations

Table B.2 – continued from previous page

	Station	Abbrev	Lon	Lat	Elev	Var	Network
	Hatchery						
23	Oliver	OLI	-119.5	49.2	297	p	ECCC
24	Ootsa	VSL	-126	53.8	861	p	ECCC
25	Pack Lake	PAK	-123.04	55	675	p	ECCC
26	Penticton	YYF	-119.6	49.5	344	p	ECCC
27	Port Hardy	YZT	-127.4	50.7	22	p	ECCC
28	Prince George	YXS	-122.7	53.9	691	T	ECCC
29	Princeton	YDC	-120.5	49.5	700	p	ECCC
30	Quatsino	WIF	-127.7	50.5	8	T	ECCC
31	Quinsam River	QUI	-125.3	50	46	p	ECCC
	Hatchery						
32	Sandspit	YZP	-131.8	53.3	6	T	ECCC
33	Shawnigan Lake	SHL	-123.6	48.6	138	T	ECCC
34	Stave Ridge	STV	-122.4	49.6	930	p	BCH
35	Stewart	ZST	-130	55.9	7	T,	ECCC
36	Terrace	YXT	-128.6	54.5	7	T, p	ECCC
37	Tofino	YAZ	-125.8	49.1	24	p	ECCC
38	Vancouver	YVR	-123.2	49.2	4	p	ECCC