Statistical downscaling of future climate projections

Report on Contract No: KM170-12-1236

Prepared for: **Environment Canada**

Prepared by: Trevor Q. Murdock
Alex. J. Cannon
Stephen R. Sobie

Pacific Climate Impacts Consortium
University of Victoria
PO Box 1700 STN CSC
Victoria, BC V8W 2Y2

PCIC Contact: Dr. Francis Zwiers
Phone: 250-721-7223

Email: fwzwiers@uvic.ca

Date: 28 March 2013

PACIFIC CLIMATE IMPACTS CONSORTIUM

Contents

1 Introduction

This report documents the production of statistically downscaled future climate projections by the Pacific Climate Impacts Consortium (PCIC) for Environment Canada under contract number KM170-12-1236. A manuscript is also in preparation for submission as a journal publication that will describe our methods in detail. Section 2 describes the selection of a subset of GCM scenarios for the CMIP5 ensemble based on an objective set of selection criteria. The criteria included hemispheric skill assessment based on the CLIMDEX indices (Sillmann et al. 2013) historical criteria used previously at PCIC (Werner 2011), and refinement based on a modified clustering algorithm (Houle et al. 2012; Katsavounidis et al. 1994). In section 3, results are summarized from an intercomparison of three downscaling techniques based on methods used in a previous intercomparison conducted by PCIC (Bürger et al. 2012a). Finally, the deliverables produced (downscaled GCM and RCM projections for Canada using two methods) are described and information about accessing them is given in section 4. Tables and figures are provided in sections 5 and 6, respectively.

2 GCM selection

2.1. Screening

We ranked GCMs according to mean ARMSE over Northern Hemisphere land in simulating all CLIMDEX indices during 1981-2000 for four reanalyses based on the results of Sillmann et al. (2013). Of the 31 models available at the time that ranking was performed, 26 had future projections for both RCP 4.5 and 8.5 (Table 1). Although there is uncertainty in the reanalyses, rankings changed very little if only the two reanalyses in which we have the most confidence (NCEP2 and ERA-interim) were used. The set of 7 least skillful models was identical, for example.

Because there are many ways to measure past model performance (Gleckler et al. 2008) and historical skill does not necessarily imply future skill (Tebaldi and Knutti 2007), we screened out models below a cut-off level rather than forming an ensemble from the 12 models with highest skill. To choose a cut-off below which to screen models out of our selection process, we turned to historical precedent to maintain consistency with methods used to choose the ensemble of CMIP3 runs that have been used extensively at PCIC. Seven criteria were considered by Werner (2011) for CMIP3 with only three models meeting all criteria: CGCM3.1(T47), GFDL-CM2.1, and UKMO-HadCM3. Each of these have been widely used for climate scenarios, impacts and adaptation studies in Canada and tended to score at or above the middle on most skill measures (Werner 2011; Gleckler et al. 2008). To keep the CMIP5 versions of each of those models (CanESM2, GFDL-CM3 and HadCM3) we retained the top 19 models based on our measure of historical skill in simulating extremes, and drop the 7 models with skill below this cut-off.

To understand the implications of removing the bottom 7 models, we compared their future projected changes to those from all possible ensembles with 7 members (Figure 1). For each model, RCP4.5 projected changes for the 2050s and 2080s were calculated for all CLIMDEX indices. The standardized

distance of the projected changes for each model from the centroid of all models was then calculated. The model with projected changes most similar to that of the centroid was assigned a rank of 1 and so on. The lowest mean rank possible for a 7 member ensemble (i.e., the 7 that are farthest from the centroid on average) is 23. The screened-out ensemble of 7 models (dashed red line) is quite different from its peers, with a mean projected rank of 19. Less than 1% of ensembles were greater outliers than this. This suggests that screening out the bottom 7 models has resulted in a narrowing in the range of change that will be projected according to runs selected from the remainder.

2.2 Ensemble selection

To select a subset of runs from the 19 remaining models, we used a technique adapted from Houle et al. (2012). We replaced k-means clustering with the KKZ algorithm (Katsavounidis et al. 1994) since the former tends to result in a narrow ensemble by favouring groupings of results and we wanted to better represent the range of climate sensitivities represented by the remaining ensemble. The KKZ algorithm recursively selects members that best span the centre and edges rather than breaking the ensemble into clusters. This allowed us to better represent the range of variation across the ensemble while still favouring clusters of results, and produced an ordered sequence of solutions.

To ensure that the selected ensemble is representative of a range of change in extremes, we considered not only projected seasonal mean temperature (minimum and maximum) and precipitation but extremes as well. Although we assigned equal weight to each variable, similar results are obtained by weighting 27 CLIMDEX indices for each variable by a factor of 1/27 and each season of minimum temperature, maximum temperature, and precipitation by 1/4.

2.2.1 Selected ensemble

Table 2 shows the ordered ensemble obtained in this way based on RCP 4.5 projections for both the 2050s and 2080s for each of the 5 Giorgi regions (Figure 2) that overlap Canada. The twelve selected members explain nearly 90% of variation in 90% of the indices and seasonal means for all variables combined over all regions considered. However, users selecting a smaller subset of the full ensemble will want to choose one that represents as wide a range as possible over their sub-region. To accommodate smaller subsets that are more regionally relevant, we ran the method separately for each region using only the 12 runs in Table 2. This produces unique ordering for each of the 5 regions (Table 3) but maintains the consistency of the full ensemble across the country.

Scatterplots of winter and summer temperature (Figure 3) for two of the Giorgi regions using the regional ordering in Table 3 indicate that the ensemble of 12 members (shown by the blue numbers) spans a fairly wide range of winter and summer temperature and precipitation projections. Runs from the 26 models considered for selection but not among the 12 selected members are shown as blue dots. Runs from the 7 models that were screened out based on historical skill are shown as red dots.

3 Downscaling evaluation

3.1 Statistical downscaling methods

The three methods considered for downscaling include Bias-Correction/Spatial Disaggregation (BCSD; Werner 2011), Bias-Correction/Constructed Analogues (BCCA; Maurer et al. 2010), and Bias-Correction/Climate Imprint method (BCCI; Hunter and Meentemeyer 2005). We downscaled each to a target resolution of 300 arc seconds (roughly 10 km x 10 km) using the ANUSPLIN historical dataset (McKenney et al. 2011).

First we modified BCSD to use minimum and maximum temperature explicitly, because our previous assessment showed that not doing so led to a reduction in skill (Bürger et al. 2012a). BCSD bias-corrects monthly mean GCM/RCM precipitation and temperature via quantile mapping onto gridded observed data aggregated to the scale of the GCM/RCM. Daily results at high spatial resolution are obtained by temporal and spatial disaggregation using rescaled randomly sampled historical observations. As a result, historical day to day variability and sequencing of events is imposed on the downscaled future projections.

BCCA is a hybrid method that combines the spatial aggregation and quantile mapping steps from BCSD with spatial information from a linear combination of historical analogues for daily large scale anomalies (Maurer et al. 2010). The quantile mapping is performed on large-scale daily simulations directly instead of on monthly aggregates.

BCCI produces spatially disaggregated results by applying quantile mapping to daily GCM/RCM outputs that have been interpolated to the high-resolution grid using the climate imprint method of Hunter and Meentemeyer (2005; Maurer and Hidalgo 2008).

3.2 Performance

We relied on three diagnostics for our evaluation of the downscaling methods. Each diagnostic is labelled by the aspect of downscaled results we aim to assess: (1) *sequencing* of events, (2) *distribution* of values, and (3) *spatial* structure. The first two diagnostics correspond to “test 1 and test 2” used in our previous downscaling intercomparison (Bürger et al., 2012a). Diagnostic 1 *(sequencing)* is based on correlation (e.g., Figure 4) while diagnostic 2 *(distribution)* is based on the Kolmogorov-Smirnov D-statistic (e.g., Figure 5). We assessed both of these diagnostics in two ways. The first was by training on four different reanalyses over British Columbia and comparing to observations in a verification period (1991-2005). The second way was by training each method using an RCM historical simulation as the target then downscaling the driving GCM’s future projection to the RCM grid and comparing to the future RCM projection. This RCM emulation setup allows us to test the downscaling methods in a changed future climate where the relationship between the model and fine scales could differ.

Diagnostic 3 *(spatial)* measures the ability of methods to retain spatial covariance: for the 20 stations used in Bürger et al. (2012b) and the three directly downscaled variables daily precipitation, minimum

and maximum temperature, we considered correlations between all station-variable combinations observed and compared to downscaled results (e.g., Figure 6) for the verification period (1991-2010).

The main findings of our comparison include:

- Neither BCCA nor BCSD strictly surpass BCCI in skill for most measures.
- For *sequencing*, BCCI considerably outperformed BCSD and slightly outperformed BCCA.
- BCCA had the worst performance on *distribution*.
- BCSD was generally superior to the other two methods for *distribution* except tor temperature percentiles and extreme minima/maxima where it fared worse than BCCI.
- BCCA's poor performance for *distribution* is more prominent for precipitation than temperature variables.
- RCM emulation indicates that results over BC are generally consistent with Canada-wide results except for the CSDI and SU indices.
- RCM emulation results generally do not depend on the GCM-RCM combination used, with the exception of precipitation skill in western Ontario/eastern Manitoba.
- BCSD has superior *spatial* performance.

In Table 4, we summarize these findings in two ways. In the first column, we assign a numerical rank based on quantitative skill. The rank is consistent across almost all variables except for diagnostic 2 *(distribution)* where BCSD was best for many variables but BCCI best for temperature percentiles and extreme minima/maxima. In the second column we provide a qualitative assessment based on expert judgment of the overall performance of the method as strong, medium or weak.

From these results we conclude that BCCA should not be considered for further use at this time. The method shows promise in the sense of an improvement over BCSD on *sequencing* and BCCI *spatially*, but at the cost of *distribution* skill. The overall rank of BCCI and BCSD is identical, but the two methods have complementary strengths and weaknesses. BCCI shines on *sequencing* because daily events come directly from the climate model being downscaled. BCSD cannot do very well in this regard because it uses historical months to obtain the daily temporal resolution. However, BCSD maintains good *spatial* correlation between stations and variables whereas BCCI is very poor at this measure because its sub-grid scale differences between locations are essentially just interpolation. RCM emulation results show higher *distribution* skill for BCSD than BCCI for all indices except temperature percentiles and extreme minima/maxima where the opposite is true. Due to the trade-offs between different aspects of performance between the two methods, we concluded that it would be preferred to downscale both.

4 Deliverables

Downscaled climate scenarios were produced for Canada between longitudes from 50°W to 143°W and latitudes from 40°N to 84°N for all past and future time slices at daily time resolution and 300 arc second (~10 km) spatial resolution. Runs downscaled are shown in Table 5 and include the NARCCAP ensemble

(NCEP-driven for the period 1980-2000 as well as GCM-driven for 1971-2000 and 2041-2070) as well as CMIP5 (historical runs for 1950-2005 and from RCP 2.6, 4.5, and 8.5 projections for 2006-2100).

The original plan was to choose from one of three different downscaling methods (BCCI, BCCA, and BCSD – described briefly in section 3.1). The results of the comparison, however, indicated such different strengths and weaknesses between BCCI and BCSD that the size of the final ensemble was doubled and the final ensemble includes both. Guidelines to users regarding use of these scenarios will need to include information on when to use BCCI only, BCSD only, or both.

All CMIP5 runs listed in Tables 2 and 3 have been downscaled using BCCI and BCSD. As the GCM downscaling was completed ahead of schedule, BCCA downscaling was also carried out, although based on section 3 we do not consider these results part of our final ensemble. We have downscaled the 10 NARCCAP runs for which output is available (Table 6). All are complete for both BCCI and BCSD. Due to the longer time to process the RCMs than GCMs, some post-processing steps remain for both BCCI and BCSD runs from RCMs that may not be complete until shortly after March 31, 2013. We did not downscale BCCA for the RCMs as we had completed our performance assessment which ruled it out before these runs would have begun. Results may be accessed as described in the next section.

4.1 Data file information

All of the data produced are stored as netcdf files on one of PCIC's servers which is backed up using a RAID system as well as tape archive. Access is available at http://pacificclimate.org/~tmurdock/ecdownscaling/ which includes two sub-directories: CMIP5 and NARCCAP. A directory tree and additional information on using the files is included in a ReadMe text file in each of the sub-directories.

Historical and future simulations for each model/downscaling method combination have been grouped together into a single netcdf file. Units are mm/day for precipitation and degrees Celsius for temperature. Time resolution is daily. Each grid point normally contains 55115 data points for CMIP5 and 21900 data points for NARCCAP files. Note that start and end dates given by the file name do not indicate that simulations are continuous (NARCCAP time series jump from 2000 to 2041, for example). Note also that the end dates of different GCM-RCM pairs are not identical.

4.1.1 Meta-data

The global netcdf file attributes provide detailed information about the driving models and experimental setup. These can be viewed using: ncdump -h <filename>. These meta-data are mostly carried through from the downscaled GCM or RCM run. There is no standard convention for statistical downscaling meta-data and to preserve information about methods and driving models/runs. PCIC hopes to develop a coherent standard for downscaled products and will revise the metadata of the current files when that is complete.

4.1.2 Naming conventions

All files follow a standard naming convention:
variable_time.resolution_downscaling.method+target.dataset+GCM+RCM_run+forcing_start-end.nc

Each item in the naming convention is defined as follows:

variable = tasmin, tasmax, or pr for minimum temperature, maximum temperature and precipitation respectively

time.resolution = day for daily time resolution

downscaling.method = BCCI or BCSD

target.dataset = ANUSPLIN300 (ANUSPLIN 300 arc-second dataset)

GCM = GCM or driving GCM (in the case of downscaling from an RCM run), e.g. CCSM

RCM (optional) = RCM, e.g. CRCM

run = GCM or RCM run name or number (e.g., historical, 1, 2, etc.)

forcing = greenhouse gas forcing (e.g. sresa2, rcp45, etc.)

start / end = date of first / last element in time series in YYYYMMDD format (e.g., 19710101)

5 Tables

Table 1: Rank based on mean ARMSE of all CLIMDEX indices during 1981-2000 compared to four reanalyses as in Sillmann et al. (2013). Rank is shown both for models with RCPs 4.5 and 8.5 future runs. The cutoff above which models were retained is indicated by the bold line (the top 19 models).

Rank	Model
1	MPI-ESM-LR
2	CSIRO-Mk3-6-0
3	MRI-CGCM3
4	EC-EARTH
5	MPI-ESM-MR
6	HadGEM2-ES
7	CESM1-BGC
8	CCSM4
9	GFDL-ESM2G
10	CNRM-CM5
11	CMCC-CM
12	GFDL-ESM2M
13	HadGEM2-CC
14	ACCESS1-0
15	NorESM1-M
16	PSL-CM5A-MR
17	MIROC5
18	NMCM4
19	CanESM2
20	PSL-CM5A-LR
21	MIROC-ESM-C
22	BNU-ESM
23	BCC-CSM1-1
24	FGOALS-s2
25	MIROC-ESM
26	PSL-CM5B-LR

Table 2: Runs selected for the CMIP5 ensemble. All selected models have historical, RCP4.5, and RCP8.5 runs. Models with runs for RCP2.6 are denoted with an X.

Order	Model	Run	RCP2.6
1	MPI-ESM-LR	3	X
2	inmcm4	1	
3	HadGEM2-ES	1	X
4	CanESM2	1	X
5	MIROC5	1	X
6	CSIRO-Mk3-6-0	1	X
7	MRI-CGCM3	1	X
8	ACCESS1-0	1	
9	CNRM-CM5	1	X
10	CCSM4	2	X
11	HadGEM2-CC	1	
12	GFDL-ESM2G	1	X

Table 3: Regional ordering of selected ensemble (based on RCP4.5).

Order	WNA	ALA	CNA	ENA	GRL
1	CNRM-CM5-r1	CSIRO-Mk3-6-0-r1	CanESM2-r1	MPI-ESM-LR-r3	MPI-ESM-LR-r3
2	CanESM2-r1	HadGEM2-ES-r1	ACCESS1-0-r1	inmcm4-r1	inmcm4-r1
3	ACCESS1-0-r1	inmcm4-r1	inmcm4-r1	CNRM-CM5-r1	CanESM2-r1
4	inmcm4-r1	CanESM2-r1	CSIRO-Mk3-6-0-r1	CSIRO-Mk3-6-0-r1	CNRM-CM5-r1
5	CSIRO-Mk3-6-0-r1	ACCESS1-0-r1	MIROC5-r3	HadGEM2-ES-r1	ACCESS1-0-r1
6	CCSM4-r2	MIROC5-r3	HadGEM2-ES-r1	CanESM2-r1	CSIRO-Mk3-6-0-r1
7	MIROC5-r3	HadGEM2-CC-r1	MPI-ESM-LR-r3	MRI-CGCM3-r1	HadGEM2-ES-r1
8	MPI-ESM-LR-r3	MRI-CGCM3-r1	CNRM-CM5-r1	CCSM4-r2	MIROC5-r3
9	HadGEM2-CC-r1	CCSM4-r2	CCSM4-r2	MIROC5-r3	HadGEM2-CC-r1
10	MRI-CGCM3-r1	CNRM-CM5-r1	GFDL-ESM2G-r1	ACCESS1-0-r1	CCSM4-r2
11	GFDL-ESM2G-r1	MPI-ESM-LR-r3	HadGEM2-CC-r1	HadGEM2-CC-r1	MRI-CGCM3-r1
12	HadGEM2-ES-r1	GFDL-ESM2G-r1	MRI-CGCM3-r1	GFDL-ESM2G-r1	GFDL-ESM2G-r1

Table 4: Ranking of methods and overall assessment of performance as strong (✓), medium (OK), or weak (X) of tested methods as assessed by diagnostic 1 (correlation), diagnostic 2 (K-S test: similarity of distribution to observations), and diagnostic 3 (inter-station/variable correlation). See text for additional explanation.

Method	Diagnostic 1- Sequencing		Diagnostic 2 - Distribution		Diagnostic 3 - Spatial	
	Rank	Performance	Rank	Performance	Rank	Performance
BCCI	**1**	✓	**1** / 2	✓	3	X
BCCA	2	✓	3	X	2	OK
BCSD	3	X	**1** / 2	✓	**1**	✓

Table 5: Downscaled climate scenario deliverables for each downscaling method.

RCMs from NARCCAP		
Time slice	**Forcing**	**# of scenarios**
1980-2000	NCEP	6 BCCI + 6 BCSD = 12
1971-2000	Historical (20c3m)	10 BCCI + 10 BCSD = 20
2041-2070	A2	10 BCCI + 10 BCSD = 20
GCMs from CMIP5		
Time slice	**Forcing**	**# of scenarios**
1950-2005	Historical	12 BCCI + 12 BCSD = 24
2006-2100	RCP 2.6	9 BCCI + 9 BCSD = 18
2006-2100	RCP 4.5	12 BCCI + 12 BCSD = 24
2006-2100	RCP 8.5	12 BCCI + 12 BCSD = 24

Table 6: GCM-driven NARCCAP ensemble members. NARCCAP uses a factorial experimental design with completed and planned runs indicated in the table with an X (completed runs) or a U (unavailable as of January 2013). See Mearns et al., 2007 and references therein as well as http://narccap.ucar.edu for RCM and driving GCM details.

		RCM run						
		CRCM	**ECP2**	**HRM3**	**MM5I**	**RCM3**	**WRFG**	***Total***
Driving GCM	**CGCM3**	X				X	X	3
	CCSM	X			X		X	3
	GFDL		X	X		X		3
	HadCM3		U	U	X			1 (3)
	Total	2	1 (2)	1 (2)	2	2	2	10 (12)

6 Figures

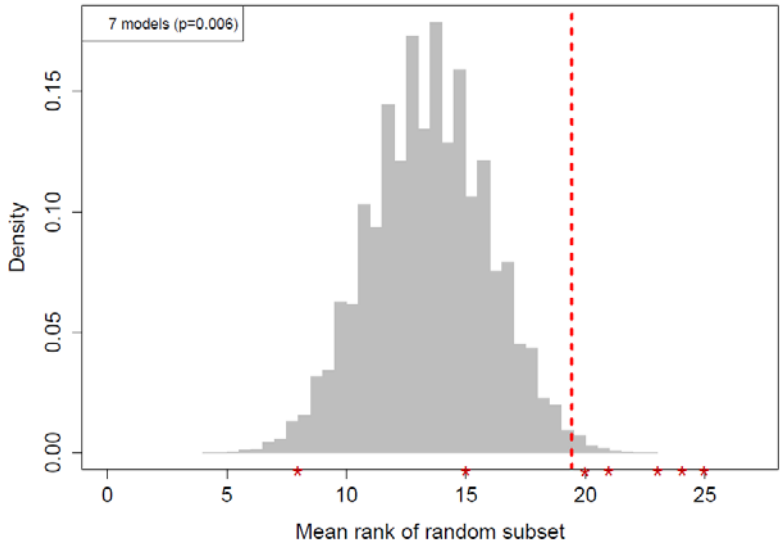

Figure 1: Mean rank of all 7 member ensembles based on comparison of RCP4.5 projected change for all CLIMDEX indices during the 2050s and 2080s periods in the 14 Giorgi and Francisco (2000) regions in the Northern Hemisphere. The histogram shows all combinations of ensembles with 7 members. The dashed red line shows models ranked 20-26 in Table 1. Red asterisks show individual runs making up that ensemble.

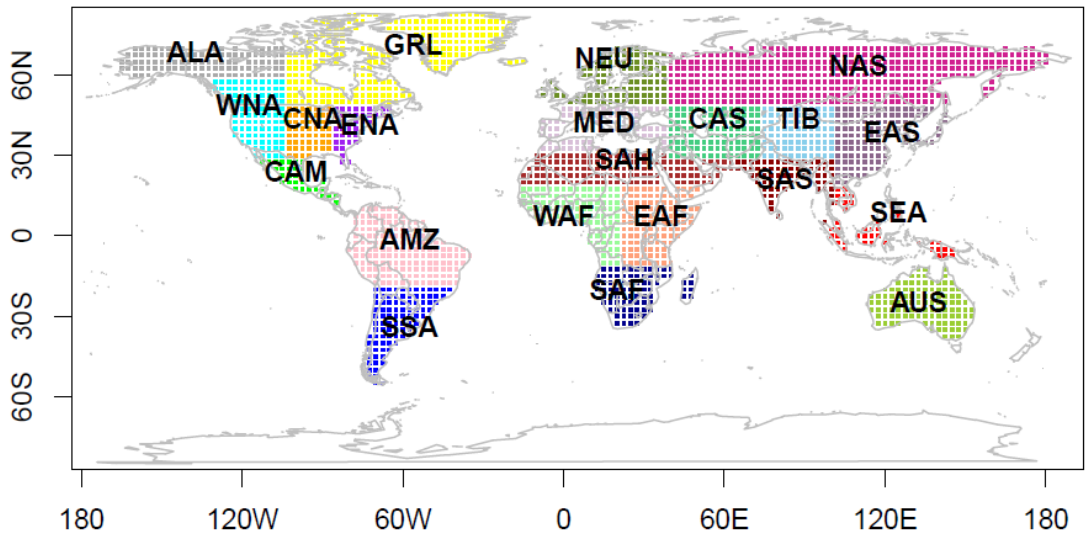

Figure 2: Giorgi Regions. The regions that intersect with Canada are ALA, WNA, CAN, GRL, and ENA.

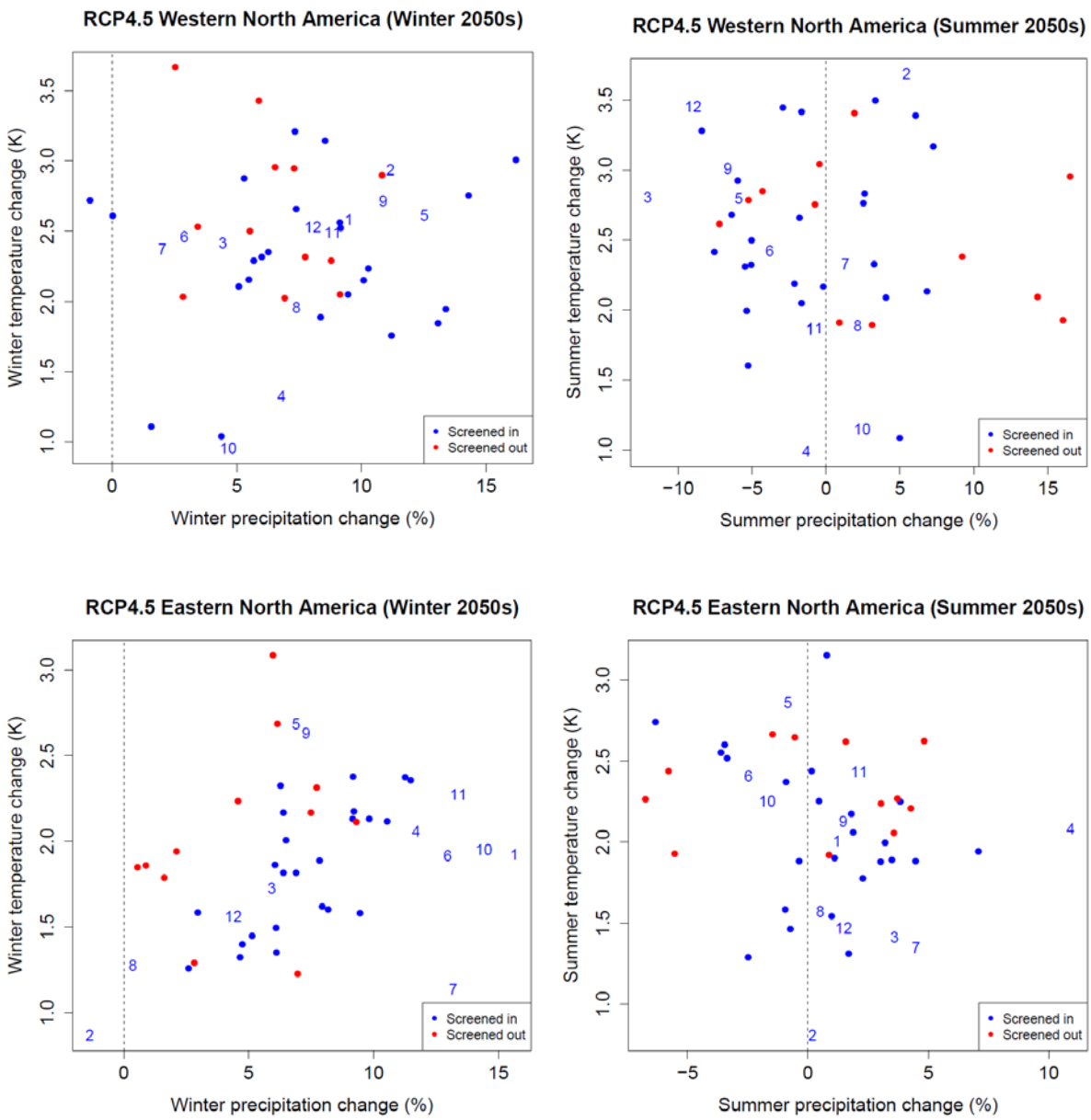

Figure 3: Scatterplots of 2050s temperature vs. precipitation for winter (left column) and summer (right column) in Western North America (top row) and Eastern North America (bottom row). The blue numbers show the 12 members of the ensemble with the number denoting the ordering, blue dots show runs with rank larger than 12, and red dots show runs from models that were screened out.

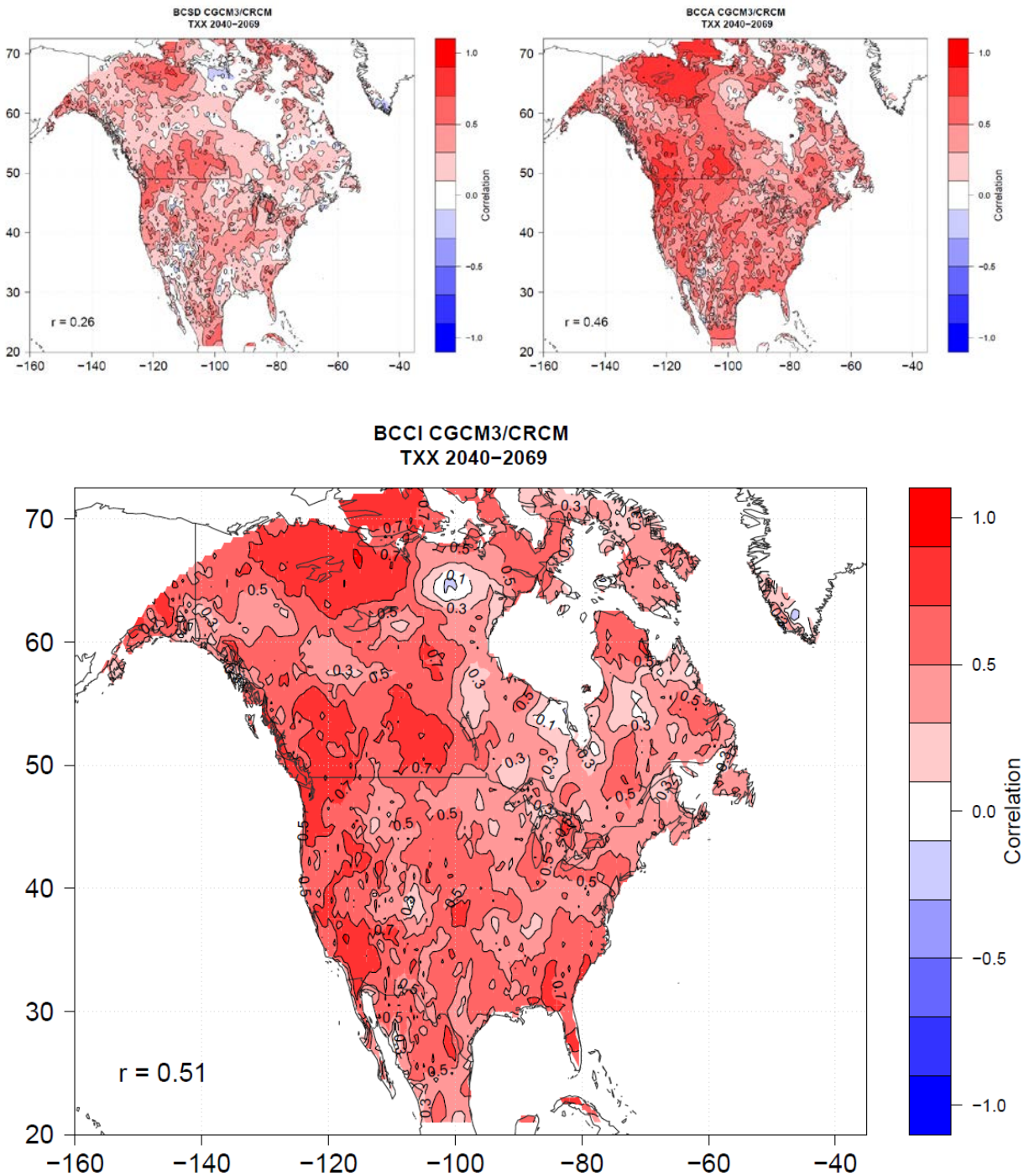

Figure 4: RCM emulation maps of diagnostic 1 – *sequencing* (correlation) between 2050s TXX from CRCM and statistical downscaling from the driving CGCM3 run.

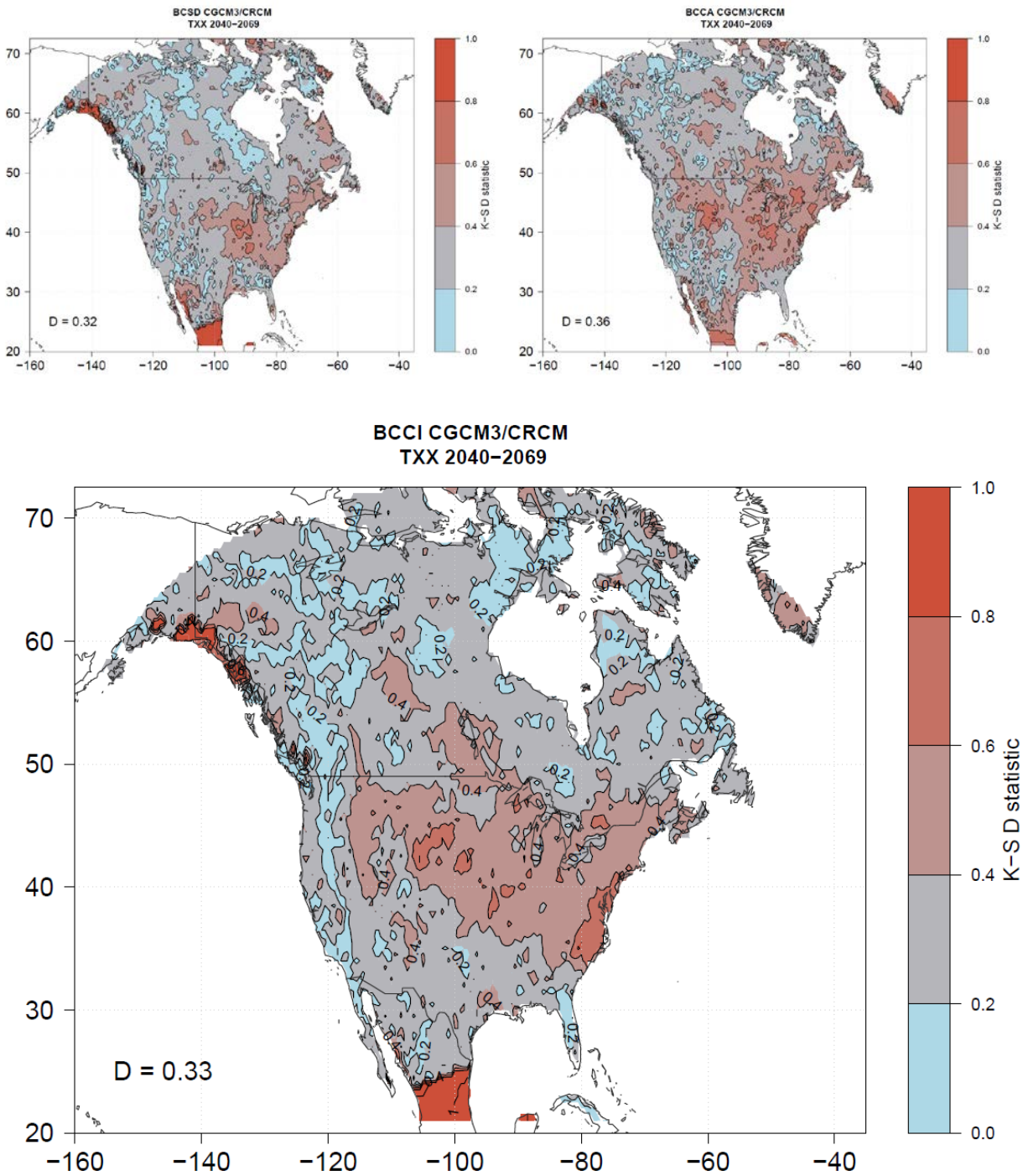

Figure 5: RCM emulation maps of diagnostic 2 – *distribution* (Kolmogorov-Smirnov D statistic) between 2050s TXX from CRCM and statistical downscaling from the driving CGCM3 run.

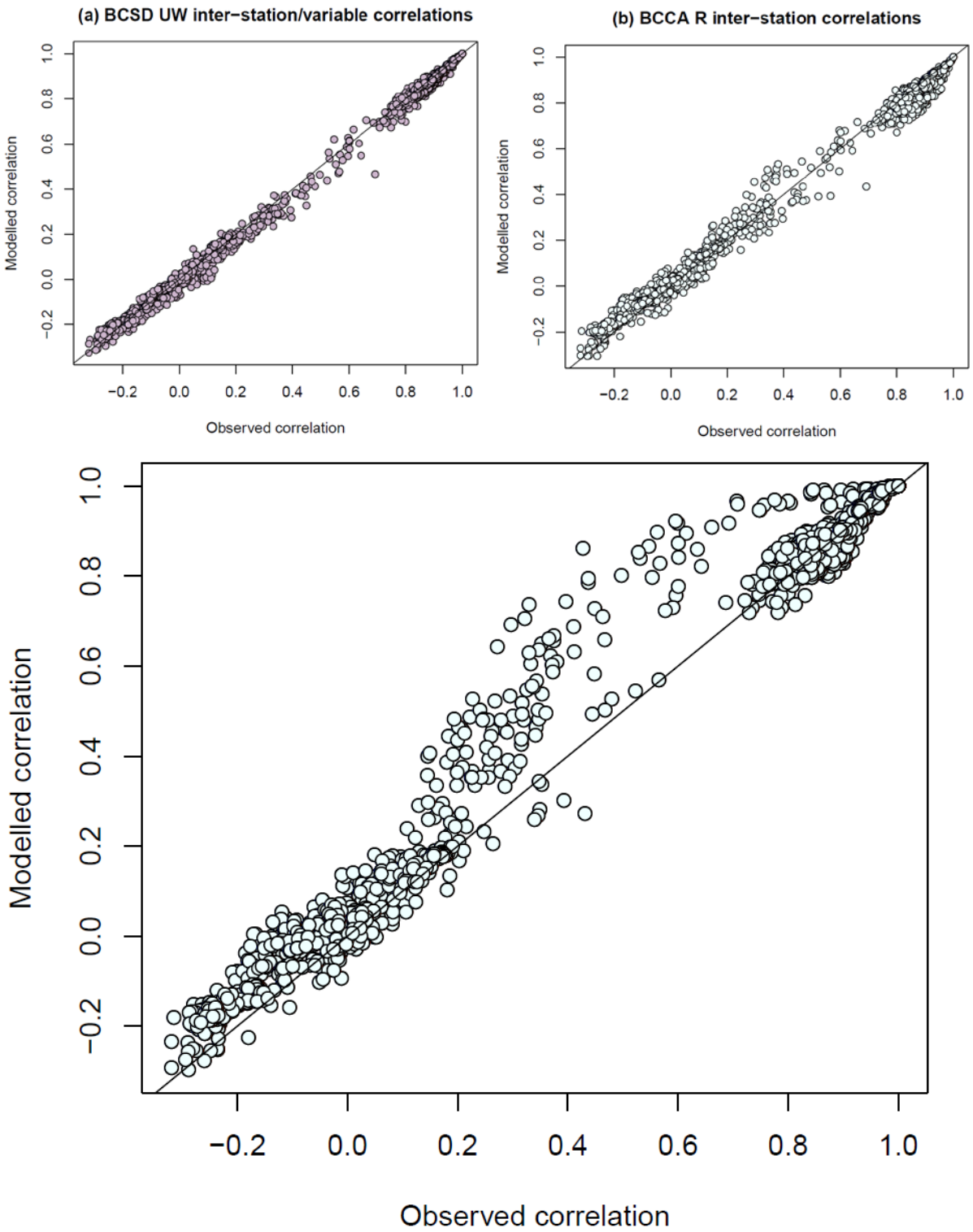

Figure 6: Plots of diagnostic 3 – *spatial* (inter-station/variable correlations) for BCSD, BCCA, and BCCI for all stations in Bürger et al. (2012b).

7 Works cited

Bürger, G., T. Q. Murdock, A. T. Werner, S. R. Sobie, and A. J. Cannon, 2012a: Downscaling Extremes—An Intercomparison of Multiple Statistical Methods for Present Climate. *Journal of Climate*, **25**, 4366–4388, doi:10.1175/JCLI-D-11-00408.1.

Bürger, G., T. Q. Murdock, A. T. Werner, S. R. Sobie, and A. J. Cannon, 2012b: Downscaling extremes - an intercomparison of multiple methods for future climate. *Journal of Climate*, 121207115945000, doi:10.1175/JCLI-D-12-00249.1.

Giorgi, F., and R. Francisco, 2000: Evaluating uncertainties in the prediction of regional climate change. *Geophysical Research Letters*, **27**, 1295–1298, doi:10.1029/1999GL011016.

Gleckler, P. J., K. E. Taylor, and C. Doutriaux, 2008: Performance metrics for climate models. *Journal of Geophysical Research-Atmospheres*, **113**, D06104–D06104.

Houle, D., A. Bouffard, L. Duchesne, T. Logan, and R. Harvey, 2012: Projections of future soil temperature and water content for three southern Quebec forested sites. *Journal of Climate*, doi:10.1175/JCLI-D-11-00440.1.

Hunter, R. D., and R. K. Meentemeyer, 2005: Climatologically Aided Mapping of Daily Precipitation and Temperature. *Journal of Applied Meteorology*, **44**, 1501–1510, doi:10.1175/JAM2295.1.

Katsavounidis, I., C.-C. Jay Kuo, and Z. Zhang, 1994: A new initialization technique for generalized Lloyd iteration. *IEEE Signal Processing Letters*, **1**, 144–146, doi:10.1109/97.329844.

Maurer, E. P., and H. G. Hidalgo, 2008: Utility of daily vs. monthly large-scale climate data: an intercomparison of two statistical downscaling methods. *Hydrological Earth System Sciences*, **12**, 551–563.

Maurer, E. P., H. G. Hidalgo, T. Das, M. D. Dettinger, and D. R. Cayan, 2010: The utility of daily large-scale climate data in the assessment of climate change impacts on daily streamflow in California. *Hydrol. Earth Syst. Sci.*, **14**, 1125–1138, doi:10.5194/hess-14-1125-2010.

McKenney, D. W. and Coauthors, 2011: Customized Spatial Climate Models for North America. *Bulletin of the American Meteorological Society*, **92**, 1611–1622, doi:10.1175/2011BAMS3132.1.

Mearns, L.O. et al., 2007: *The North American Regional Climate Change Assessment Program dataset, National Center for Atmospheric Research Earth System Grid data portal, Boulder, CO. Data downloaded 2009 to 2011.* http://www.earthsystemgrid.org/project/NARCCAP.html.

Sillmann, J., V. V. Kharin, X. Zhang, F. W. Zwiers, and D. Bronaugh, 2013: Climate extremes indices in the CMIP5 multimodel ensemble: Part 1. Model evaluation in the present climate. *Journal of Geophysical Research: Atmospheres*, n/a–n/a, doi:10.1002/jgrd.50203.

Tebaldi, C., and R. Knutti, 2007: The use of the multi-model ensemble in probabilistic climate projections. *Phil. Trans. R. Soc. A*, **365**, 2053–2075, doi:10.1098/rsta.2007.2076.

Werner, A. T., 2011: *BCSD Downscaled Transient Climate Projections for Eight Select GCMs over British Columbia, Canada*. Pacific Climate Impacts Consortium, Victoria, British Columbia.