Statistical downscaling of future climate projections for North America

Report on Contract No: KM040-131148/A

Prepared for: **Environment Canada**

Prepared by: Trevor Q. Murdock
Stephen R. Sobie
James Hiebert

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: 3 July 2014

Contents

Introduction

This report documents the production of statistically downscaled future climate projections by the Pacific Climate Impacts Consortium (PCIC) for Environment Canada under contract KM040-131148/A. Manuscripts are also in preparation for submission as journal publications that will describe our methods and results in further detail. Environment Canada will be invited to review all manuscripts related to this project before submission.

The goal of the project was to produce a set of statistically downscaled future climate projections for North America based on the latest available Global Climate Model (GCM) and Regional Climate Model (RCM) simulations suitable for driving hydrologic models that could facilitate furthering Environment Canada's study of changes in water availability. Attaining this goal involved carrying out five distinct tasks, described below and referred to throughout the report:

1. In consultation with Environment Canada, select an appropriate observational daily dataset to serve as a downscaling target for North America.
2. In consultation with Environment Canada, select a downscaling technique appropriate to the characteristics of the observational dataset.
3. In consultation with Environment Canada, downscale projections from CORDEX-NA (CanRCM4) and the final member of the NARCCAP ensemble for daily precipitation and temperature at 10 km resolution over North America.
4. Extend downscaling of selected CMIP5 simulations to North America for daily precipitation and temperature at a 10 km resolution over North America.
5. Use downscaled daily temperature and precipitation time series to produce projections of ETCCDI indices of extremes (Expert Team on Climate Change Detection and Indices, see Klein Tank et al. 2009) and conduct an extreme value analysis of 20-year return period temperature and precipitation events throughout North America.

The report consists of four sections. First, the selection and construction of the observational dataset that was developed to train the selected downscaling method is described. In section 2, results are provided that document the rationale for selecting BCCAQ as the primary downscaling method. In section 3, the GCM and RCM simulations downscaled are listed and notes on data access and meta-data information are provided. Finally, Section 4 is an overview of selected results of projected changes in ETCCDI extremes indices and the magnitude of 20-year return values over North America.

1 Training dataset

In consultation with Environment Canada, select an appropriate observational daily dataset to serve as a downscaling target for North America.

Canada-wide downscaling has already been carried out (Murdock et al. 2013) using two downscaling methods (BCSD and BCCAQ) that were trained against the ANUSPLIN gridded daily dataset produced by Natural Resources Canada (McKenney et al. 2011).

A comprehensive set of options for a North American statistical downscaling target was presented in the proposal. Upon confirmation with Environment Canada, we proceeded with ANUSPLIN daily for Canada and UW 1/16° (Livneh et al. 2013) for the US. We first re-gridded UW to the 300 arc second ANUSPLIN resolution using the *remap* command in CDO (climate data operators[1]).

Although the daily time series does not sync perfectly across the border, monthly average temperature records are quite consistent even without correction (not shown). However, considerable cross-border differences exist for precipitation. The US is wetter in the west during winter and prairies are drier year round while other seasons and locations are relatively similar (Figure 1 left column). Because of the dependence on location and time of year, we have not adjusted the climatology of the US training dataset in any way prior to downscaling. Rather, we carried out downscaling to the UW dataset re-gridded to ANUSPLIN resolution as our target dataset.

Another option for blending across the border includes correcting both Canada and the US daily values to WorldClim climatology. At this time, however, we recommend that harmonization of differences across the border be carried out on a project by project basis. The simplest adjustment is to bias correct the entire US domain to the monthly climatology of the ANUSPLIN monthly product. Though it reduces the magnitude of discrepancies, wetter western winters and drier prairies throughout the year remain (Figure 1 right column). Differences in the distribution of values across the border indicate that smooth stitching of precipitation may require a more complex method than bias correction. For example, Figure 2 shows that, during winter, there are many more days with small precipitation amounts on the US than Canadian side of the border. Note that the ANUSPLIN monthly and monthly averaged ANUSPLIN daily datasets themselves differ so a perfect match should not be expected. Further improvement is obtained by adjusting the entire domain to the ANUSPLIN monthly climatology, which would be an option for continent wide studies based on our downscaled projections.

[1] http://code.zmaw.de/projects/cdo

2 Select statistical downscaling method

In consultation with Environment Canada, select a downscaling technique appropriate to the characteristics of the observational dataset.

Before carrying out downscaling, it is essential to test different types of statistical downscaling methods for skill based on historical simulations, to ensure that the selected method is appropriate for use as intended. To do this, PCIC built upon statistical downscaling carried out over Canada (Murdock et al. 2013). That work in turn built upon previous downscaling evaluation projects (Bürger et al. 2012, 2013).

The methods assessed in this project for downscaling performance include:

1. DELTA is simply the delta method whereby time series of interpolated GCM anomalies are applied to historical fine-scale climatologies to produce future time series.
2. BCSD (Bias-Correction/Spatial Disaggregation; Werner 2011). We tested a modified version of BCSD that uses minimum and maximum temperature explicitly, because our previous assessment showed that not doing so led to a reduction in skill (Bürger et al. 2013). 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.
3. BCCA (Bias-Correction/Constructed Analogues; Maurer et al. 2010) 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. The quantile mapping is performed on large-scale daily simulations directly instead of on monthly aggregates.
4. QMAP (Quantile MAPping; Gudmundsson et al. 2012) 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). For this reason it is sometimes referred to as BCCI.
5. BCCAP (BCCA with Post-processing) applies a post-processing bias correction with quantile mapping to BCCA (Maurer et al. 2010).
6. BCCAQ (BCCA with Quantile MAPping reordering) is a modified version of BCCA that uses QMAP in addition to the post-processing correction used in BCCAP (Cannon et al., in prep). Additional details are provided on BCCAQ in section 2.2.

2.1 Performance

In the previous studies conducted by PCIC, a series of skill tests was used to measure the performance of each downscaling method with respect to its ability to simulate indices of extremes. There are two important reasons for evaluating the performance of downscaling methods with respect to extremes. First, the downscaling methods are generally tuned to reproduce the climatological properties of temperature and precipitation but training is usually not specifically focussed on extremes, so these indices provide a somewhat independent means of validation. Furthermore, in terms of providing information for adaptation planning or using downscaled fields to conduct further analysis of impacts, extremes are often more relevant than average quantities (Peterson 2005). Therefore, it is vital that the downscaling method that is selected be capable of demonstrating skill at simulating indices of extremes.

For these reasons, our previous studies evaluated the ability of downscaling methods to reproduce each of the widely used 27 ETCCDI indices (Klein Tank et al. 2009) in three different ways. We continued this approach to evaluation of the downscaling methods and considered diagnostics for three aspects of skill: (1) *sequencing* of events, (2) *distribution* of values, and (3) *spatial* structure.

To address *sequencing* diagnostic 1 is correlation; for *distribution* diagnostic 2 is the Kolmogorov-Smirnov D-statistic. These correspond to tests 1 and 2 in Bürger et al. (2012). We measure *spatial* skill in diagnostic 3 by considering the ability of methods to reproduce observed spatial autocorrelation: for each day we compared Moran's I (Moran 1950) between observed and downscaled fields for spatial neighbourhoods of 3 to 15 grid points. This is an extension of diagnostic 3 from Murdock et al. (2013) in which specific sites in British Columbia were considered.

We assessed each diagnostic in four ways:

1. Training on the ERA40 reanalysis in part of the historical period (1958-1990) over the Canada-wide domain and comparing to ANUSPLIN observations in a verification period (1991-2002).
2. Same as 1 but for NCEP1 reanalysis (with training period of 1950-1990 and verification period of 1991-2010).
3. 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 results to the future RCM projection. This was carried out for CGCM3/CRCM over the NARCCAP domain (2041-2070 SRES A2). 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.
4. A modified version of 3 in which CanRCM4 is used in a "perfect model" approach. Here, outputs are aggregated to the GCM scale for downscaling back to the RCM scale, over the CORDEX-NAM22 domain (2081-2100 RCP8.5). This setup also allows us to test in a changed future climate but is not influenced by any interactions between the dynamical and statistical downscaling steps.

Diagnostic 1 is shown in Figure 3a. Correlation ranges from 0 (dark blue) to 1 (dark red) for each of the ETCCDI indices. Correlations are generally highest in the perfect model setup and lowest in RCM emulation. Skill is lowest in indices related to rainfall or consecutive days, and generally highest in variables related to temperature. BCSD has the lowest skill of all methods in this measure because it uses historical months to obtain the daily temporal resolution. Indeed, it was this weakness of BCSD that inspired the search for an improved method. QMAP, BCCAQ, and to a lesser extent BCCA and DELTA perform well on *sequencing* because daily events come directly from the driving data.

However, BCSD maintains a realistic distribution as seen in Figure 3b, which displays results for diagnostic 2, where the Kolmogorov-Smirnov D statistic ranges from 1 for no skill (dark blue) to 0 for perfect skill (dark red). For most variables QMAP, BCCAQ, and BCSD have the highest skill and BCCA the lowest, with BCCAP and DELTA in between.

The poorest aspect of QMAP (and the DELTA method) performance is seen in diagnostic 3 (Moran's index; Moran 1950) shown in Figure 4 for the perfect model setup. Values range from −1 for perfect dispersion through 0 for random spatial pattern to +1 for perfect correlation. Neighbourhood refers to the number of grid points in each horizontal direction used. Spatial autocorrelation decreases with number of grid boxes included in the neighbourhood, as expected. Spatial autocorrelations are quite close to true values for each of BCCAQ, BCCAP, and BCCA in particular, with BCSD trailing these methods slightly. QMAP and DELTA are much worse than all other methods in this measure because their sub-grid scale differences between locations are essentially just interpolation. Results are similar for ERA40, NCEP, and RCM-emulation, though the separation between the high and low skill methods is most clear in the perfect model approach.

In addition to visual inspection of Figures 3, 4 and plots like Figure 4 for other variables and corresponding to ERA40, NCEP1, and RCM-emulation, the diagnostics for each variable / method pair were tested against threshold criteria to obtain a pass or fail against each of the three diagnostics. The test criteria for diagnostics 1 and 2 are identical to Bürger et al. (2012). The *spatial* test used comparative quartile analysis (Mearns et al. 2012) which checks whether there is any overlap between boxplots (i.e. between the ranges spanned by the 25th and 75th percentiles) of Moran's I for the downscaling method and the observed values.

In Figure 5 the percentage of all possible tests passed is shown for each method (for diagnostics 1 and 2 a method is considered to pass overall if it passes at over three-quarters of all grid points). The colours indicate the three types of tests and the lines denote the contributions from each of the four setups (starting from the bottom: ERA-40, NCEP1, RCM-emulation, and perfect model).

The percentage of tests passed is lowest for DELTA, as expected, at 53%. BCSD, QMAP, BCCA, and BCCAP have similar results, at 68%, 68%, 64%, and 69%, respectively. BCCAQ outperforms all other methods, passing 81% of tests.

Finally, all of these results are summarized in Table 1, where ranks are assigned according to the percentage of tests passed for each diagnostic. BCCAQ is first or tied for first in the sequencing and spatial diagnostics and second in distribution. Although QMAP beats BCCAQ slightly here and is tied for first in sequencing, it is only fifth on the spatial diagnostic. The average rank across all three tests for BCSD is fourth, better only than the DELTA method (the lowest average rank is 5 despite there being 6 methods due to some ties). The Table also shows a qualitative assessment of results as good or poor, denoted by a green check or red x, respectively. In this regard, DELTA is also the weakest – the only method poor in two areas (distribution and spatial). BCSD, BCCA, and QMAP are each poor in one area (sequencing, distribution, and spatial, respectively). Only BCCAP and BCCAQ are good across all three measures, with BCCAQ consistently scoring higher than BCCAP in general.

Thus, BCCAQ delivers the sequencing and distribution skill of QMAP along with the spatial skill of BCSD, and compared to the other gridded methods examined, demonstrates superior skill. For this reason, it was chosen as the primary downscaling method.

However, BCCAQ is a modification of a previously established method and as such should be subject to continued scrutiny. As it was also expressed during consultation with Environment Canada that further work on performance would be desired. Therefore, BCSD was run for a majority of the runs for which BCCAQ was run. In addition, both BCCA and QMAP are required inputs to BCCAQ so these are also available for detailed method comparisons as long as the disk space for retaining them is available. Therefore, while the main deliverable is BCCAQ downscaling, a considerable resource is available for conducting further analysis of statistical downscaling performance across North America.

2.1 Additional details on BCCAQ

Since BCCAQ was selected as the main downscaling method for proceeding, a more extensive description of the method than given in the previous section is provided here. BCCAQ is a hybrid downscaling algorithm that combines outputs from QMAP method, which we have shown performs well in terms of long-term temporal sequencing/distribution of extremes, but produces overly smoothed spatial fields on a day-by-day basis; and BCCA, which we have shown performs poorly in terms of the distribution of extremes, but produces daily fields with more realistic spatial structure.

First, the BCCA and QMAP algorithms are run independently (Figure 6), and then BCCAQ combines BCCA and QMAP as a post-processing step. The daily QMAP outputs at each fine-scale grid point are reordered within a given month according to the daily BCCA ranks. Reordering is done month-by-month, as in the final scale/shift step of BCSD, to prevent the downscaled outputs from drifting too far from the QMAP long-term trend. Because the optimal weights used to combine the analogues in BCCA are derived on a day-by-day basis, without reference to the full historical dataset, the algorithm is prone to "Huth's paradox", wherein models that are calibrated based on short-term variability fail to project realistic long-term trends (Huth 2004; Benestad et al. 2008). Reordering data for each fine-scale grid point within a month effectively breaks the overly smooth representation of sub GCM-grid scale spatial variability inherited from QMAP (Maraun 2013) thereby resulting in a more accurate representation of event-scale spatial gradients. Over longer time-scales, the spatial variability of BCCAQ converges to that of QMAP.

3 GCM and RCM runs downscaled

In consultation with Environment Canada, downscale projections from CORDEX-NA (CanRCM4) and the final member of the NARCCAP ensemble for daily precipitation and temperature at 10 km resolution over North America.

Extend downscaling of selected CMIP5 simulations to North America for daily precipitation and temperature at a 10 km resolution over North America.

Downscaled projections of daily minimum and maximum temperature, daily precipitation from Global Climate Model (GCM) and Regional Climate Model (RCM) simulations at daily time resolution and 300 arc-second (~10 km) spatial resolution have been produced over North America. In collaboration with Environment Canada, a subset of the 88 potential historical and future time periods described in the proposal were selected that could be completed in the time frame. Table 2 lists the downscaled runs completed for BCCAQ. The downscaled runs include (for RCP8.5) all of priorities 1 through 3, and none of priority 4 as outlined in Tables 2 through 5 of the proposal.

In addition to daily temperature and precipitation time series, annual ETCCDI indices of extremes (Sillmann et al. 2013a; Sillmann et al. 2013b) were computed for each run at the same spatial resolution using the climdex.pcic R-package[2]. Finally, 20 year return intervals were estimated by fitting the annual maxima of downscaled values to the Generalized Extreme Value (GEV) distribution for three standard future periods (2011-2040, 2041-2070, 2071-2100) using the ismev package for the R software programming language which uses the method of L-moments (Hosking et al. 1985). Uncertainty was assessed based on the standard errors and confidence intervals of estimates of the three GEV parameters. These are approaches that are frequently used in climate research, e.g. (Kharin et al. 2007; Kharin et al. 2013), and with which PCIC has substantial expertise.

For CMIP5 runs (Taylor et al. 2012) priority was given to the highest emissions scenario RCP 8.5 (Moss et al. 2010) as it was determined in conjunction with Environment Canada that more RCPs provide the least additional information compared to ensuring as large an ensemble size as possible for a given RCP. The NCARCCAP MM5I RCM simulation driven by the HadCM3 global climate model was the only NARCCAP simulation that had not yet been downscaled for Canada (Murdock et al. 2013a) and is now complete for all of North America. The first run from CORDEX-NA (CanRCM4 0.22°) has also been downscaled for all of North America.

As BCCAQ was chosen as the primary downscaling method, and its computation requires BCCA and QMAP, the available suite of downscaled products over Canada has now been expanded to include all six methods listed in section 2, providing a greater ability to assess the impact of the choice of downscaling method on fine-scale climate change projections.

3.1 Data file information

All of the data produced are stored as NetCDF4/HDF5 files on PCIC's RAID 6 arrays and are backed up to

[2] http://cran.r-project.org/web/packages/climdex.pcic/index.html

a tape archive. Access is available at https://www.pacificclimate.org/data/statistically-downscaled-climate-scenarios for GCMs over Canada only. Downloading of results from this location may be automated using a DAP request as described at http://tools.pacificclimate.org/dataportal/docs/user.html#power-user-howto or by using wget. The full set of results is also available from http://pacificclimate.org/~tmurdock/ecdownscaling/ which includes three sub-directories: CMIP5, NARCCAP, and CONUS and may be browsed by http as well as accessed by wget.

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. 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).

3.1.1 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 but PCIC has developed a coherent standard for use with files on our data portal and these files all follow that standard.

3.1.2 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 = e.g., BCCAQ 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)

4 Analysis of selected results

Use downscaled daily temperature and precipitation time series to produce projections of ETCCDI indices of extremes (Expert Team on Climate Change Detection and Indices, see Klein Tank et al. 2009) and conduct an extreme value analysis of 20-year return period temperature and precipitation events throughout North America.

We provide below an overview of projected changes in R95p, RX1day, TX90p, and TN10p indices, the annual minima and maxima of daily minimum and maximum temperature, respectively, (i.e. TNn and TXx), and of annual maximum daily precipitation accumulations (RX1day). In addition we consider the 20-year return values of precipitation, cold events for daily minimum temperature and hot events for daily maximum temperature.

The focus of the analysis for annual temperature and precipitation is on the ensemble median of the 12 RCP 8.5 runs (Figures 7 to 12). In addition to the ensemble median, results from downscaling CanESM2 are also presented. This helps give a sense of the amount of smoothing that occurs in the ensemble average by visualizing a single run.

We present extremes results for CanESM2 only (Figures 13 to 21). A sense of how the CanESM2 results for the ETCCDI indices may differ from the ensemble median may then be obtained by referring to the differences between the ensemble median and CanESM2 for annual temperature and precipitation (Figures 7 to 12). Note that the CanESM2 results do not exist in a small part of the far north but the ensemble median results do – in those few locations that are not included in the domain of all 12 models, the median is computed from those models with results only.

Annual precipitation is projected to increase Canada-wide (Figure 7), with a south to north gradient of larger increases, and with larger relative increases at the coasts. In addition, we see some very large increases of over 50% in the far North. This is due in part to low historical precipitation in the area but should also be treated with caution due to the sparse station network in the region upon which that historical climatology is based. The overall latitudinal gradient in projected precipitation change also generally exists in the United States (Figure 8), with the exception of increases throughout most of the west. Precipitation increases are generally small near the Canada-US border.

Comparing CanESM2 to the CMIP5 ensemble median, we see quite similar results overall except that CanESM2 projects larger precipitation increases over much of the prairies, north-central British Columbia, and all of the Western US. The projected precipitation decrease in the Southeastern US is also more intense in the CanESM2 run, however.

Night-time low temperatures show warming across Canada (Figure 9), with a gradient from about 2°C to 3°C warming in the south to well over 5°C in much of the north. This south to north gradient is also present in the central US (Figure 10), but with less warming on the west and southeast coasts. Day-time high temperatures (Figures 11 and 12) follow similar patterns, with slightly less warming projected than for night-time low temperatures in most locations (compare with Figures 9 and 10).

The individual CanESM2 run generally projects more warming in both night-time low and day-time high temperatures than the ensemble median, particularly in the far north, parts of BC and most of the western and northern US (Figures 9 to 12).

Moving to the indices of extremes, annual precipitation during very wet days, R95p (Figure 13), is projected to increase over most of the continent. The only decreases projected lie in a swath in the centre of the continent from Texas to south-eastern Manitoba/south-western Ontario. The largest increases are projected for the west coast, far North, and the eastern edge of the Maritimes including many areas with considerable increases of over 80%. At a small number of locations, increases of close to 400% are projected. Preliminary investigation into the cause of the very large projected increases appears to indicate that inflation is occurring, a phenomenon that quantile mapping methods are prone to (Maraun 2013). This may warrant investigation of further modifications to BCCAQ to avoid inflation. Note that if such modifications were made, BCCA would not need to be re-computed, only QMAP and the combination of the two methods into BCCAQ, meaning that computing a modified version of BCCAQ would be less onerous than computing it the first time.

The annual maximum one-day precipitation (RX1day – roughly comparable to the 1-year return period; Figure 14) is also projected to decrease in a few locations in the mid-continental swath where R95p is projected to decrease, but in this case there are also some locations further to the west with decreases, which do tend to coincide with locations with the smallest R95p increases (compare with Figure 13). The largest increases in RX1day are also along all three coasts. While some locations display increases of over 80%, generally the percentage increases in RX1day are less than those in R95p, as may be expected.

Note that the largest increases in R95p and RX1day in the US west are also located where CanESM2 precipitation projections are larger than the ensemble median (Figure 7 and 8), so these increases may be exaggerated in this single run compared to the ensemble as a whole.

For temperature extremes, we first consider the change in the annual minima TNn (Figure 15) and annual maxima TXx (Figure 16) – roughly akin to 1-year return period events. The largest warming in TNn is >15°C at the southern tip of Baffin island, with increases between 10°C to 15°C in northern Quebec, parts of southern BC, the US Pacific Northwest, and Yukon/Northwest Territories. Increases in TXx are considerably smaller than for TNn at most locations, though much of BC and parts of Quebec show increases of over 6°C. No increases smaller than 1°C are indicated for either index in the US, and none smaller than 3°C for TNn in Canada though TXx does have some locations in the far north with warming close to zero projected.

The other indices of temperature extremes that we considered are the change in percentage of cold days TN10p (Figure 17) and warm days TX90p (Figure 18). These figures indicate that cold days are projected to occur during 5% to 10% fewer days (darker blue indicates larger decrease in number of days thus larger warming) than in the past (by definition 10% in the past). In other words, the previous TN10p events are projected to occur between half as often to not at all. The near disappearance of cold days is primarily in the western US and most of Canada with the exception of some parts of the

Maritimes, southern Ontario, Hudson Bay region, and Northwest Territories. Even in most of these locations however, a reduction by at least 5% of days (half of past occurrence) is projected. Increases in warm days are projected over the entire domain, with the largest increases (60% to 70%) in Florida and Haida Gwaii. What was the 10% warmest days in past are projected to occur around 1/2 of the time in future (increases of about 40%) over most of the US, northern Canada, and parts of BC and the Maritimes, and around 1/3 of the time (increases of about 20% to 25%) in the remainder of the continent.

Moving to the return periods, for 20-year return period one-day precipitation events (Figure 19), the spatial pattern is less coherent than for annual precipitation (Figures 7 and 8) and R95p (Figure 13), with a pattern quite similar to that for RX1day (Figure 14) but with more variability (note the different legends between Figures 14 and 19). Indeed, the patchy nature of the projected change suggests caution in interpreting these results. The largest changes here are also most likely affected by inflation and correction of inflation may result in a less patchy future projection.

For temperature events we consider the 20-year return period for minima of night-time low temperatures (Tmin-RP20; Figure 20) and maxima of day-time high temperatures (Tmax-RP20; Figure 21). At most locations, the 20-year return periods warm by considerably more than the corresponding annual average warming in Tmin (Figures 9 and 10) and Tmax (Figures 11 and 12). Warming in Tmin-RP20 and Tmax-20 tends to also exceed that of the annual minimum TNn (Figure 15) and annual maximum (TXx), and follow a somewhat similar spatial pattern, but with larger variability. Tmin-RP20 warming is largest throughout the Canadian and US rockies, prairies, northern New England, Maritimes, Quebec, and southern Baffin Island while Tmax-RP20 warming is considerably less than for Tmin-RP20 in the US, Rockies, and prairies but larger in northwest BC and southern Yukon.

In summary, the downscaled projections indicate increased precipitation across almost all of North America, with larger relative increases at many locations when considering more extreme indices, though also with increasing spatial variability. Some locations display alarmingly large relative increases, suggesting a potential need for a modification to BCCAQ to avoid inflation. Warming is projected in all aspects of temperature, again with generally larger increases (and spatial variability) for more extreme indices. The CanESM2 run used to display the indices of extremes and return periods (Figures 13 to 21) is generally similar to the ensemble median projected change except with slightly more warming and precipitation increase in many locations, particularly the US west (Figures 7 to 12).

5 Tables

Table 1: Performance of methods with rank based on number of tests passed as well as qualitative assessment based on visual inspection (e.g. Figures 1 and 2) as strong (green check) or weak (red X) for each diagnostic: sequencing (correlation), distribution (K-S test D statistics), and spatial (Moran's I). Final column shows average rank and overall qualitative assessment across all three measures.

Method	Sequencing		Distribution		Spatial		Average	
DELTA	4	✓	5	X	6	X	5	X
BCSD	6	X	3	✓	4	✓	4	OK
BCCA	3	✓	6	X	1	✓	3	OK
QMAP	1	✓	1	✓	5	X	2	OK
BCCAP	4	✓	4	✓	1	✓	3	✓
BCCAQ	1	✓	2	✓	1	✓	**1**	✓

Table 2: Runs downscaled with BCCAQ for North America

Project, model, run	Forcing (for GCM and RCM simulations) and/or source of driving data (RCMs)	Time slice
CMIP5 RCP 8.5 (x12)*	20c3m	1951-2005
	RCP 8.5 (x12)	2006-2100
CORDEX-NA CanRCM4 (0.22° resolution)	20c3m	1951-2005
	RCP 8.5 + CanESM RCP 8.5	2006-2100
NARCCAP: MM5i	NCEP	1980-2000
	20c3m	1971-2000
	A2 + HadCM3 A2	2041-2070

* CNRM-CM5-r1, CanESM2-r1, CSIRO-Mk3-6-0-r1, CCSM4-r2, MIROC5-r3, MPI-ESM-LR-r3, MRI-CGCM3-r1, GFDL-ESM2G-r1, HadGEM2-ES-r1, ACCESS1-0-r1, inmcm4-r1, and HadGEM2-CC-r1

6a Figures – training dataset and downscaling method

The figures in this section describe results of to sections 1 and 2 of the text regarding training dataset and statistical downscaling method selection, respectively.

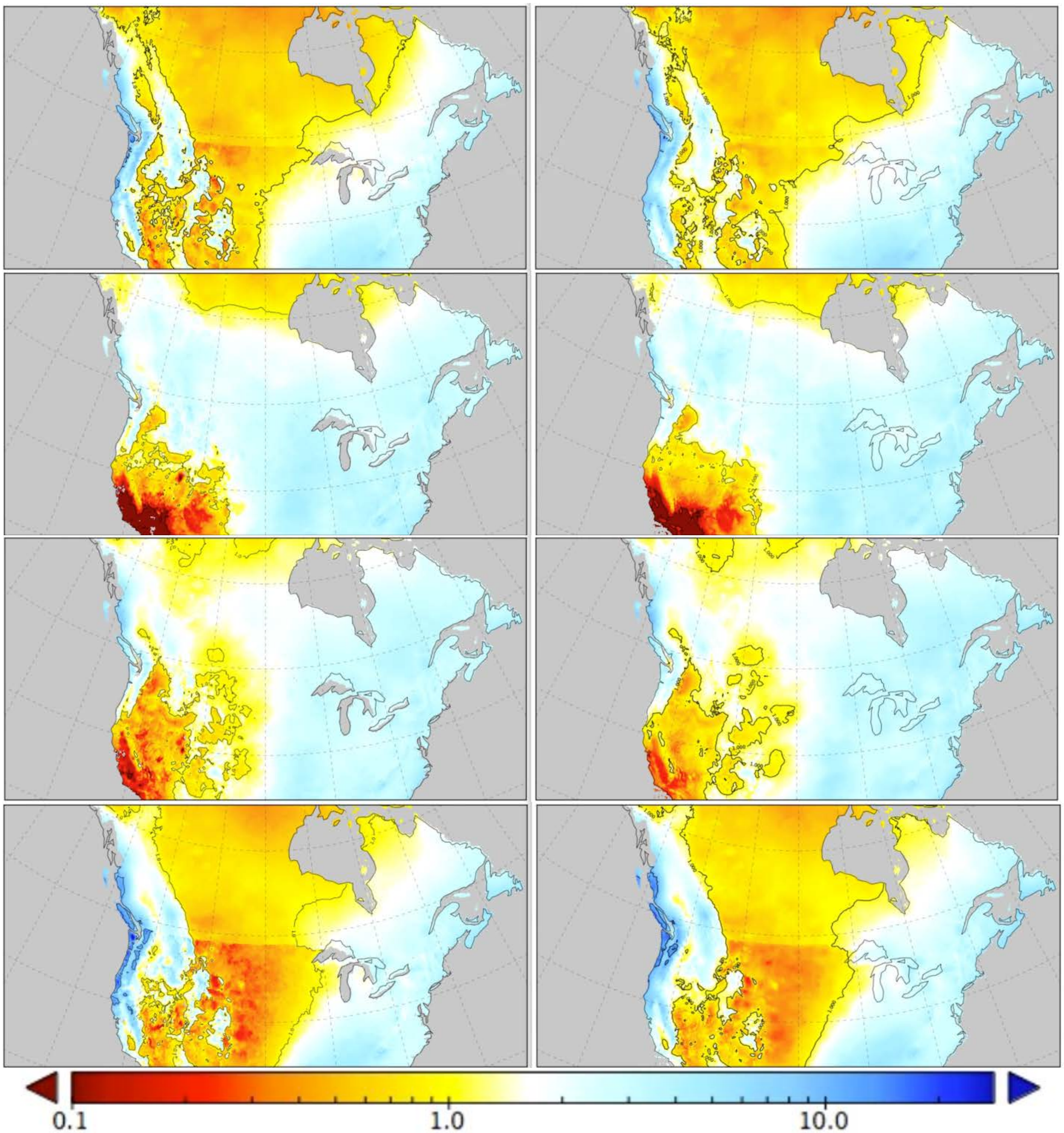

Figure 1: Monthly average precipitation climatologies for March (top row), June (second row), September (third row) and December (bottom row) for raw stitched driving data (left) and bias corrected stitched (right).

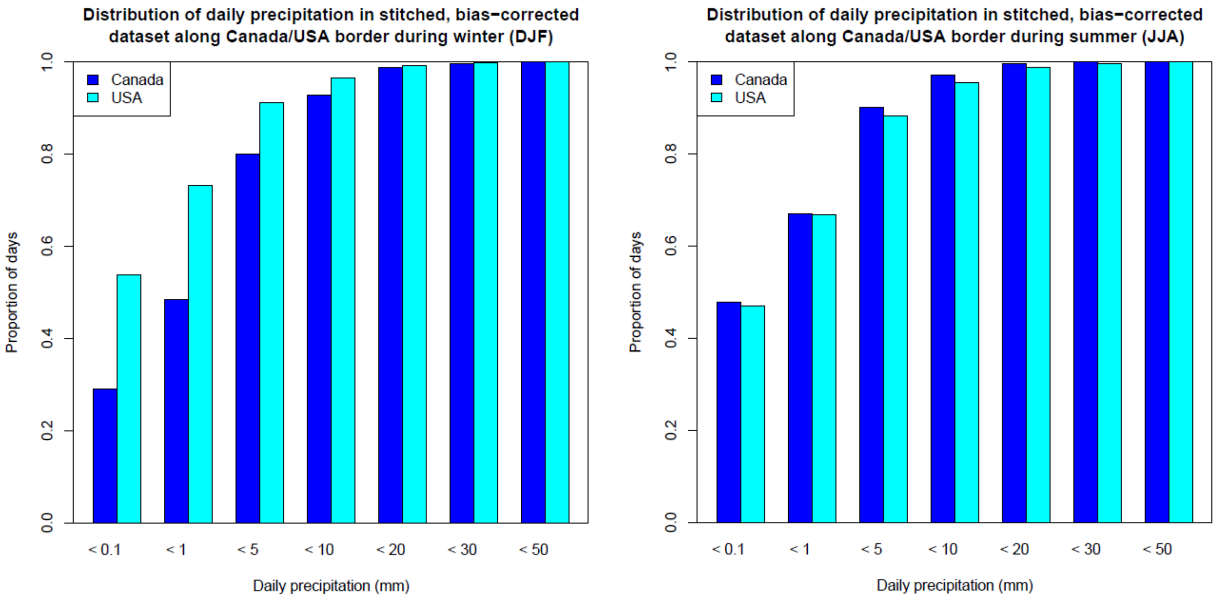

Figure 2: Distribution of daily precipitation for strip along Canada-United States border (two grid boxes wide on each side) for winter (left) and summer (right) in bias-corrected stitched dataset.

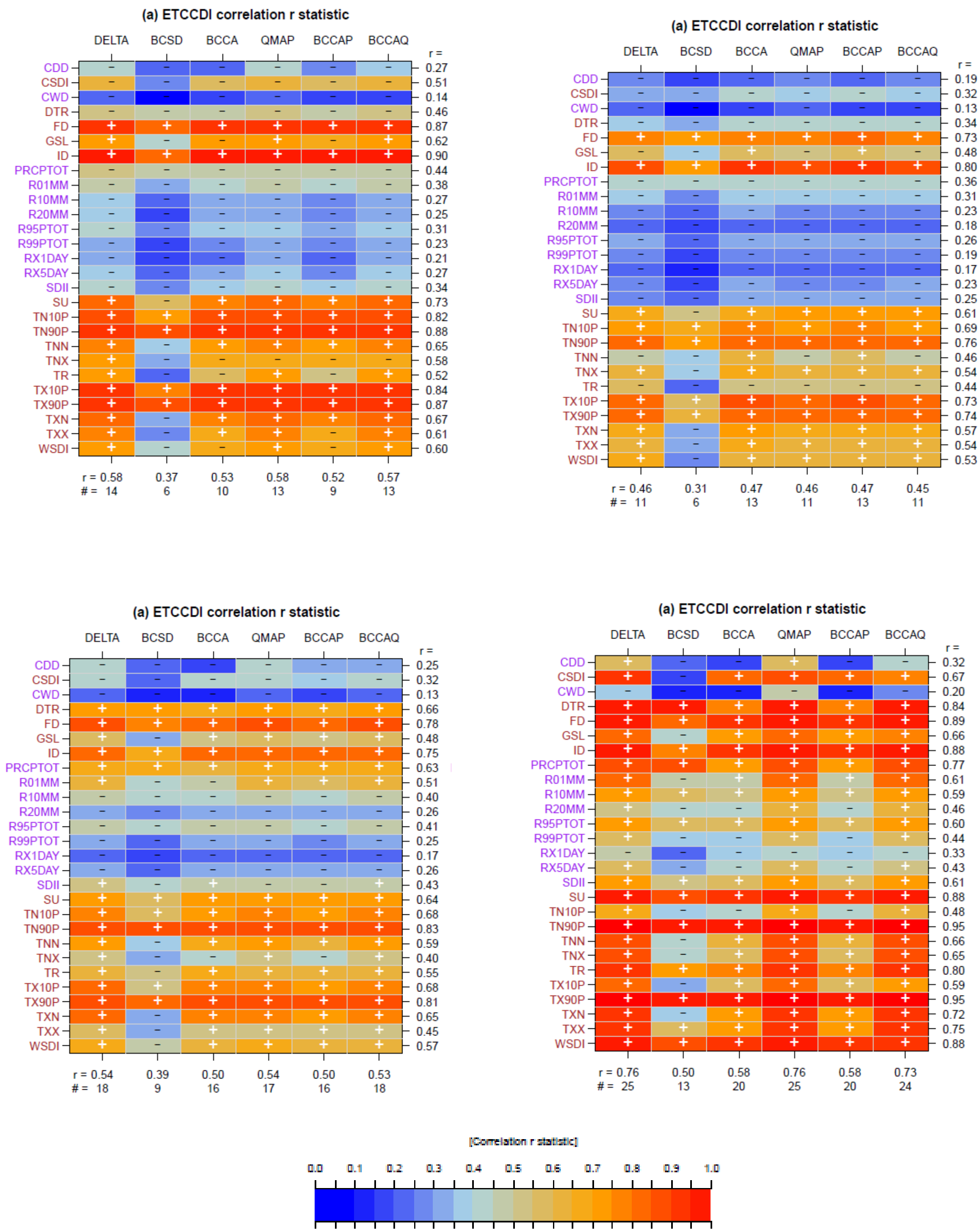

Figure 3a: Diagnostic 1 (sequencing): correlation r statistic for CLIMDEX indices for ERA40 (top left), NCEP1 (top right), RCM-emulation (bottom left), and perfect model (bottom right).

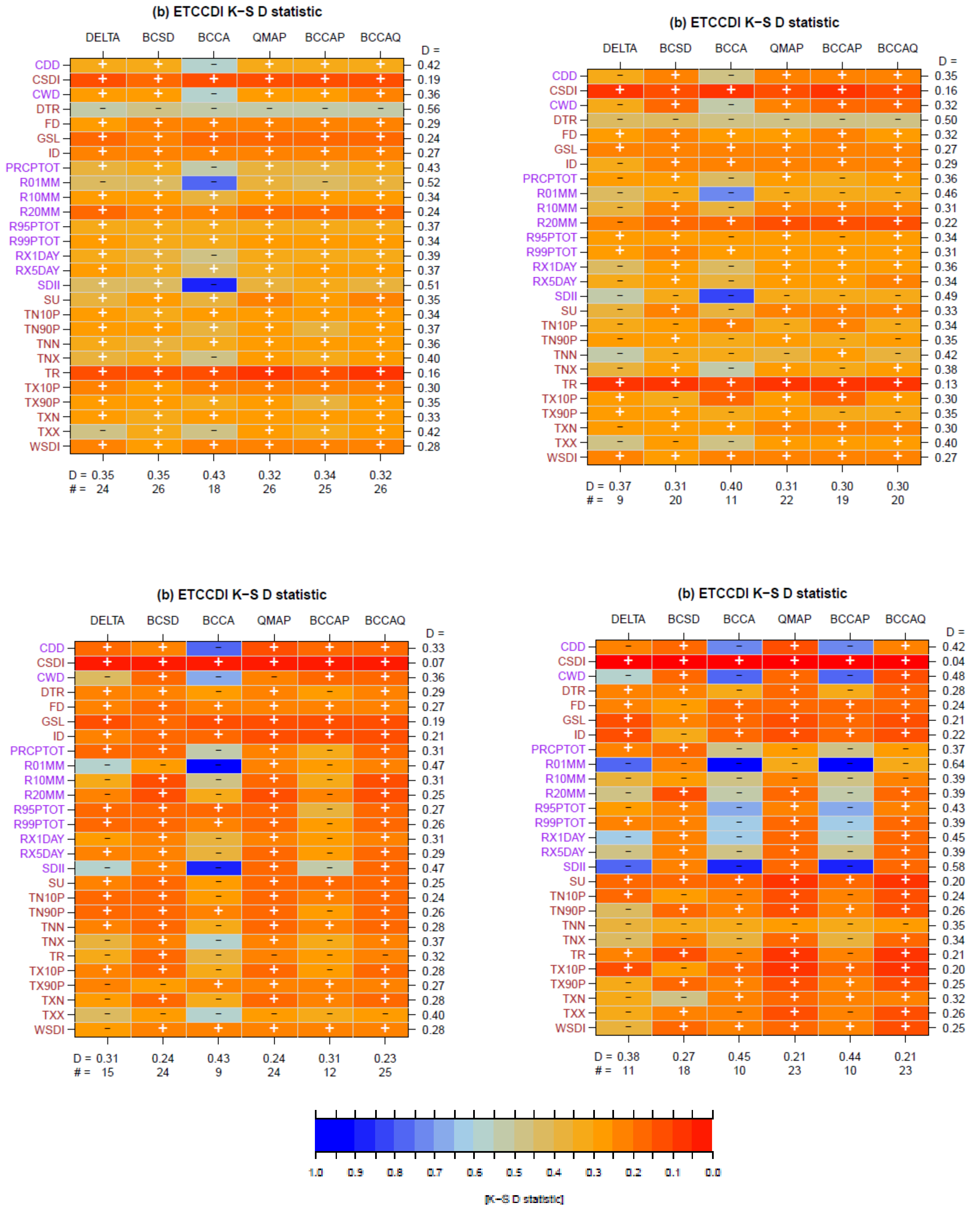

Figure 3b: Diagnostic 2 (distribution): K-S D statistic for CLIMDEX indices for ERA40 (top left), NCEP1 (top right), RCM-emulation (bottom left), and perfect model (bottom right).

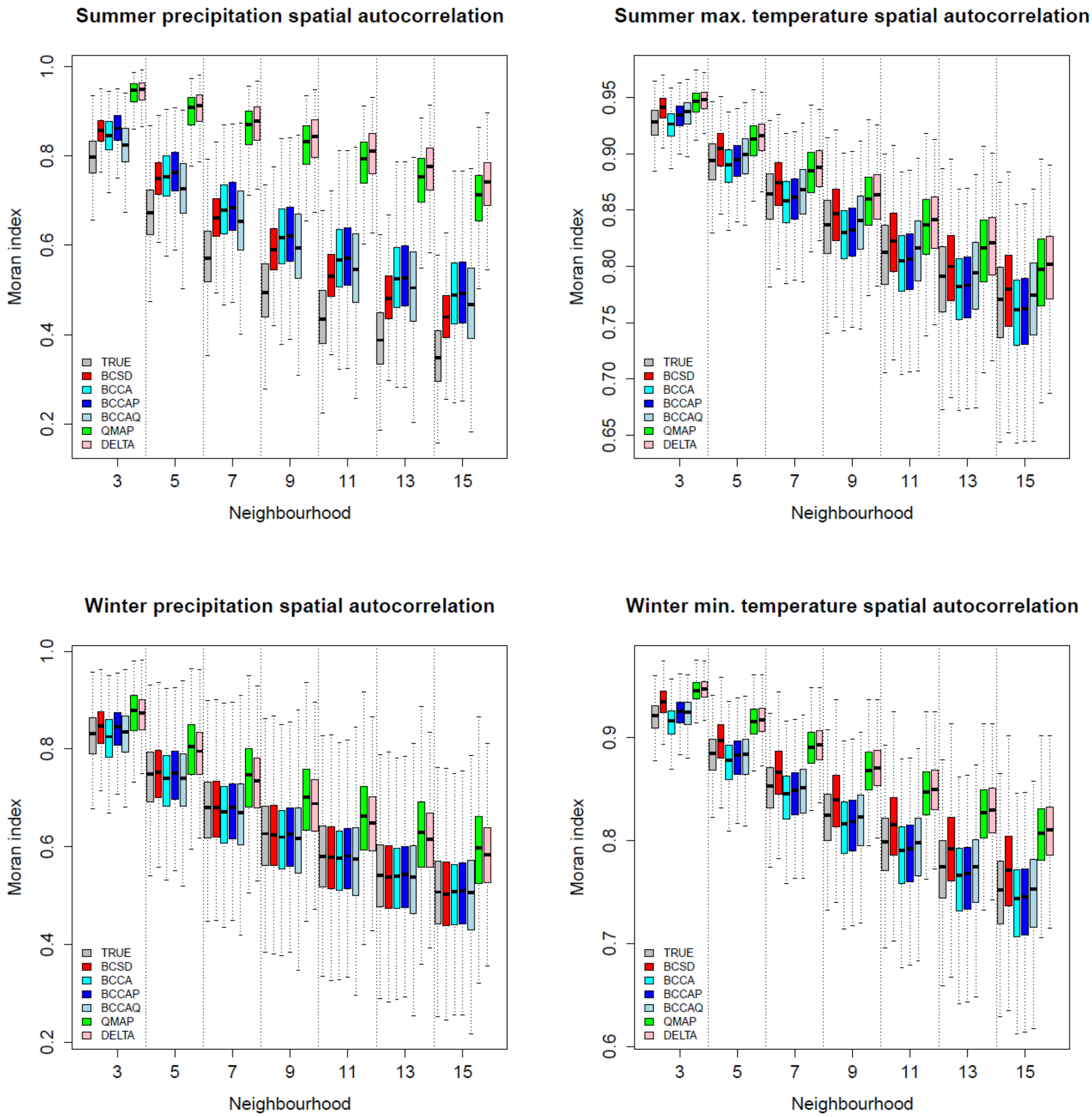

Figure 4: Diagnostic 3: spatial autocorrelation (Moran index) for ERA40 with BCSD, BCCA, BCCAP, BCCAQ, QMAP, and DELTA for "perfect model" setup.

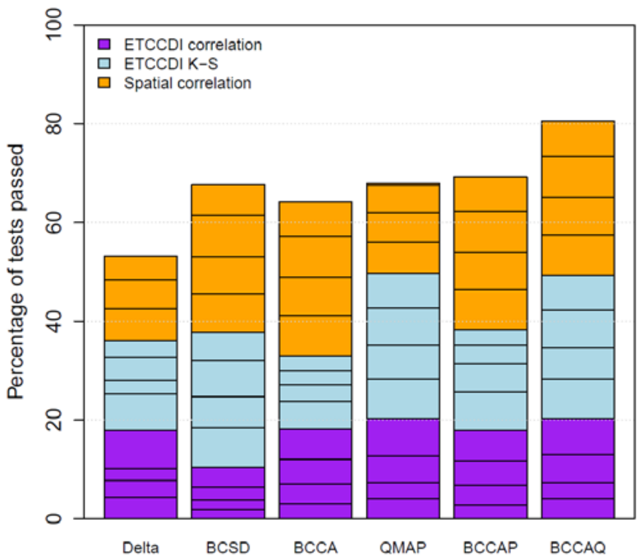

Figure 5: Summary of diagnostics 1-3 for DELTA, BCSD, BCCA, QMAP, BCCAP, and BCCAQ. See text for descriptions of the colours and lines denoted in the bars.

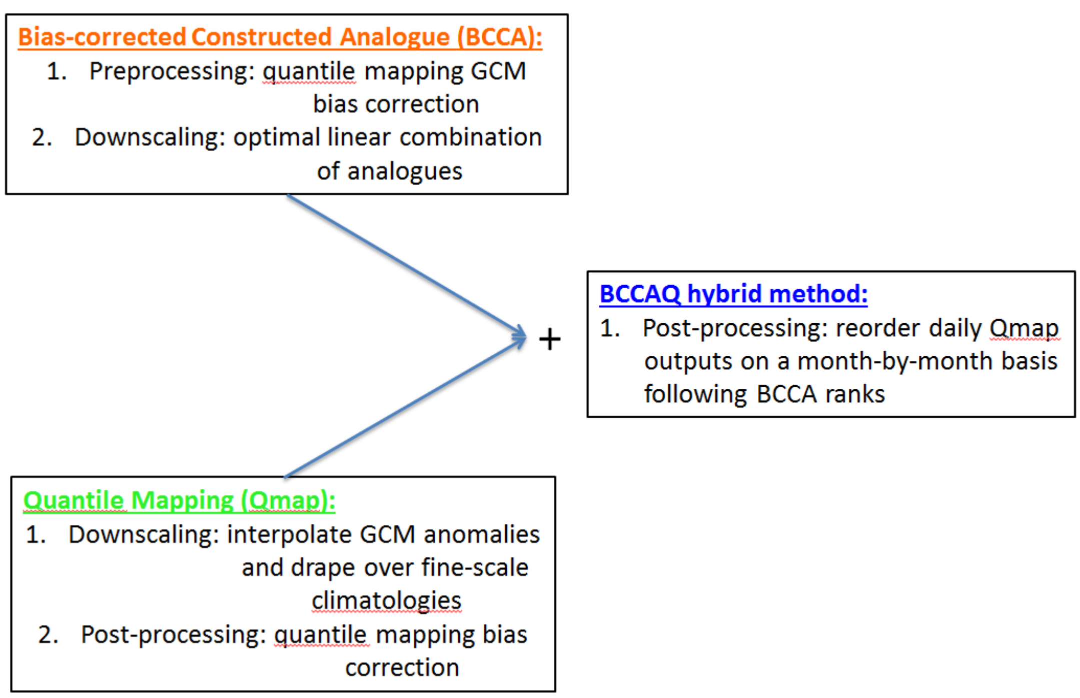

Figure 6: Schematic BCCAQ methodology

6b Figures – change in annual means

The figures in this section show projected change in annual means based on the 12 downscaled CMIP5 RCP 8.5 runs, as described in sections 3 and 4.

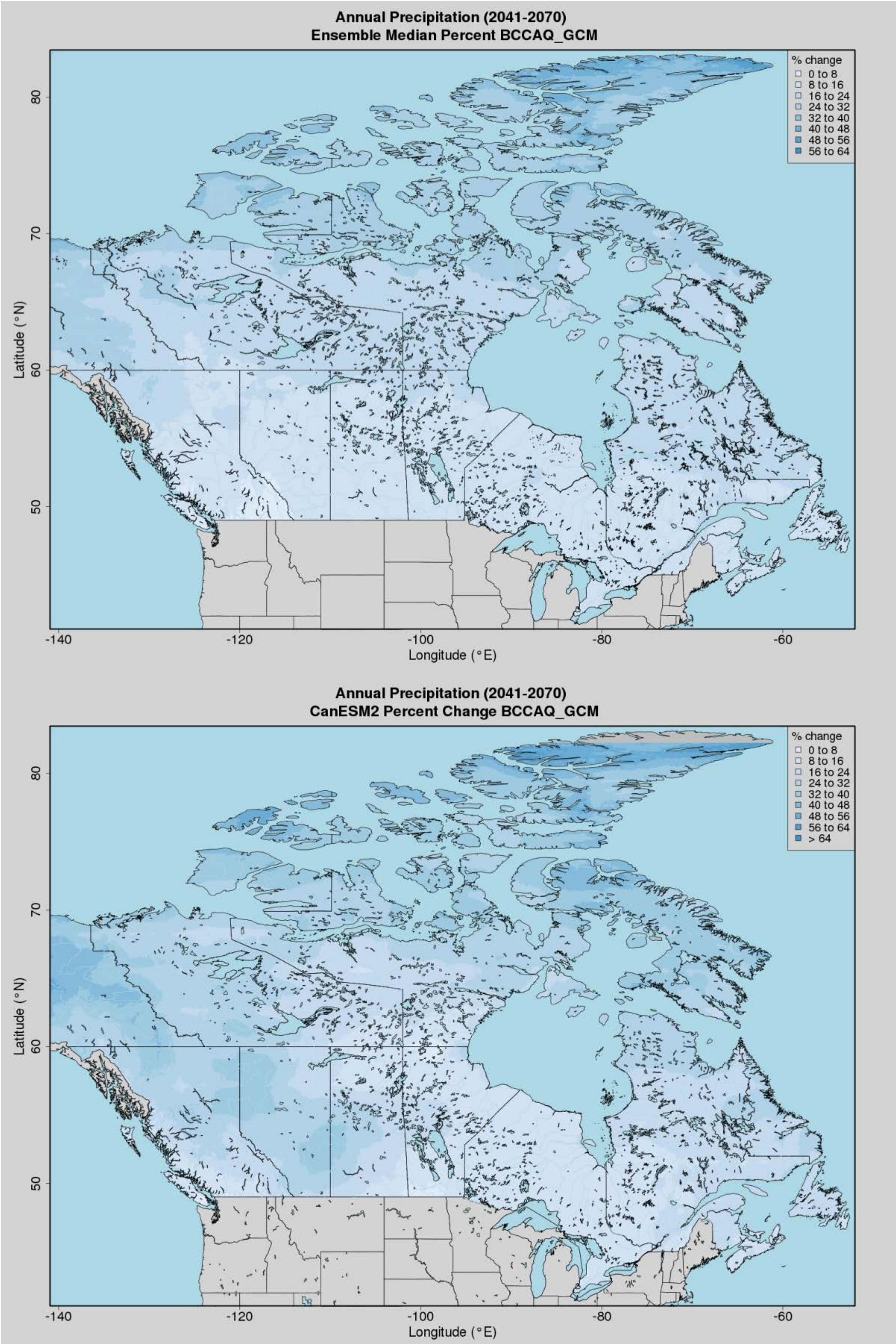

Figure 7: 2050s percent change in annual precipitation for ensemble median (top) and CanESM2 (bottom) for RCP8.5 over Canada. Ensemble median is computed at each grid point independently.

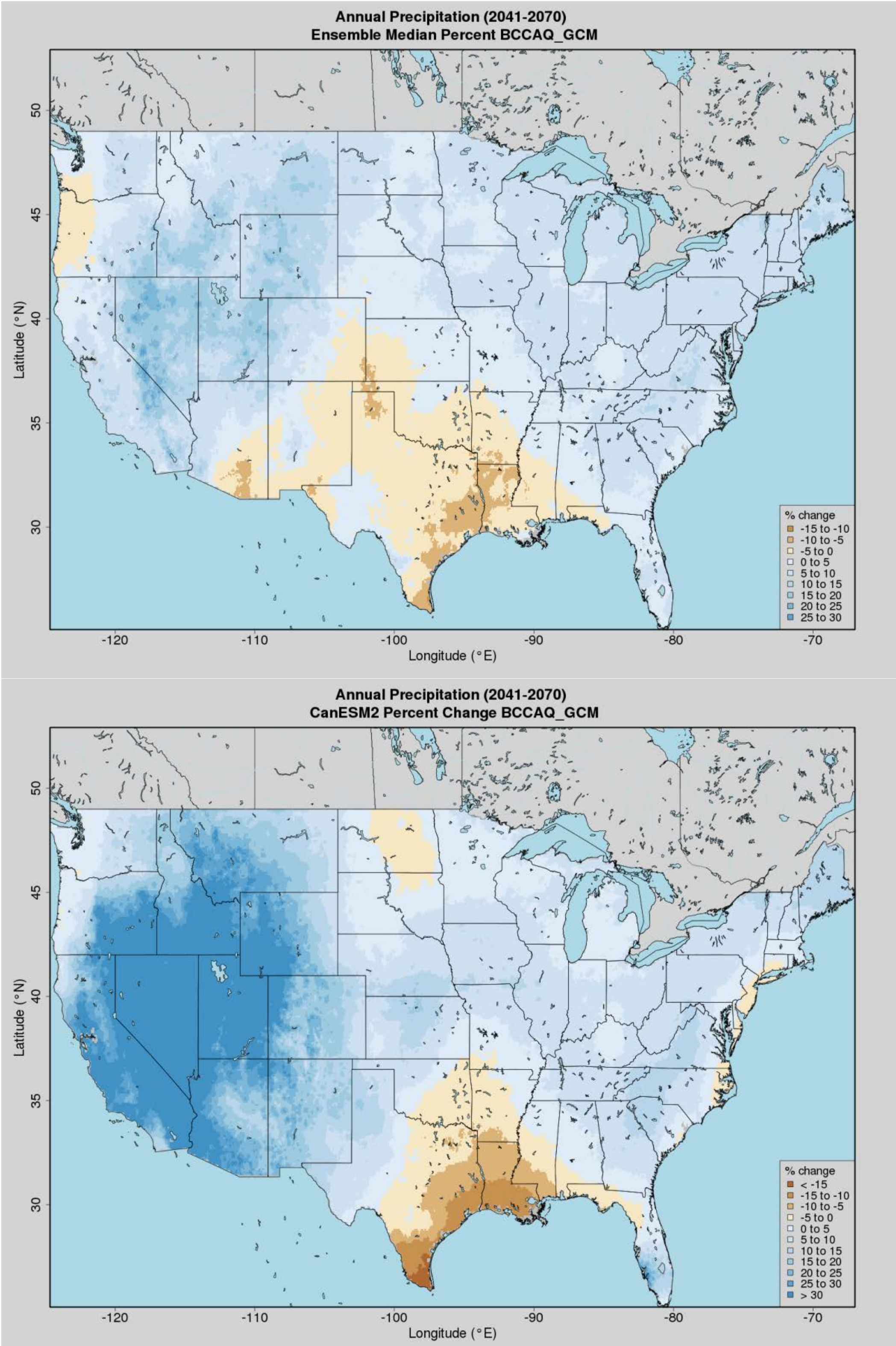

Figure 8: 2050s percent change in annual precipitation for ensemble median (top) and CanESM2 (bottom) for RCP8.5 over the United States. Ensemble median is computed at each grid point independently.

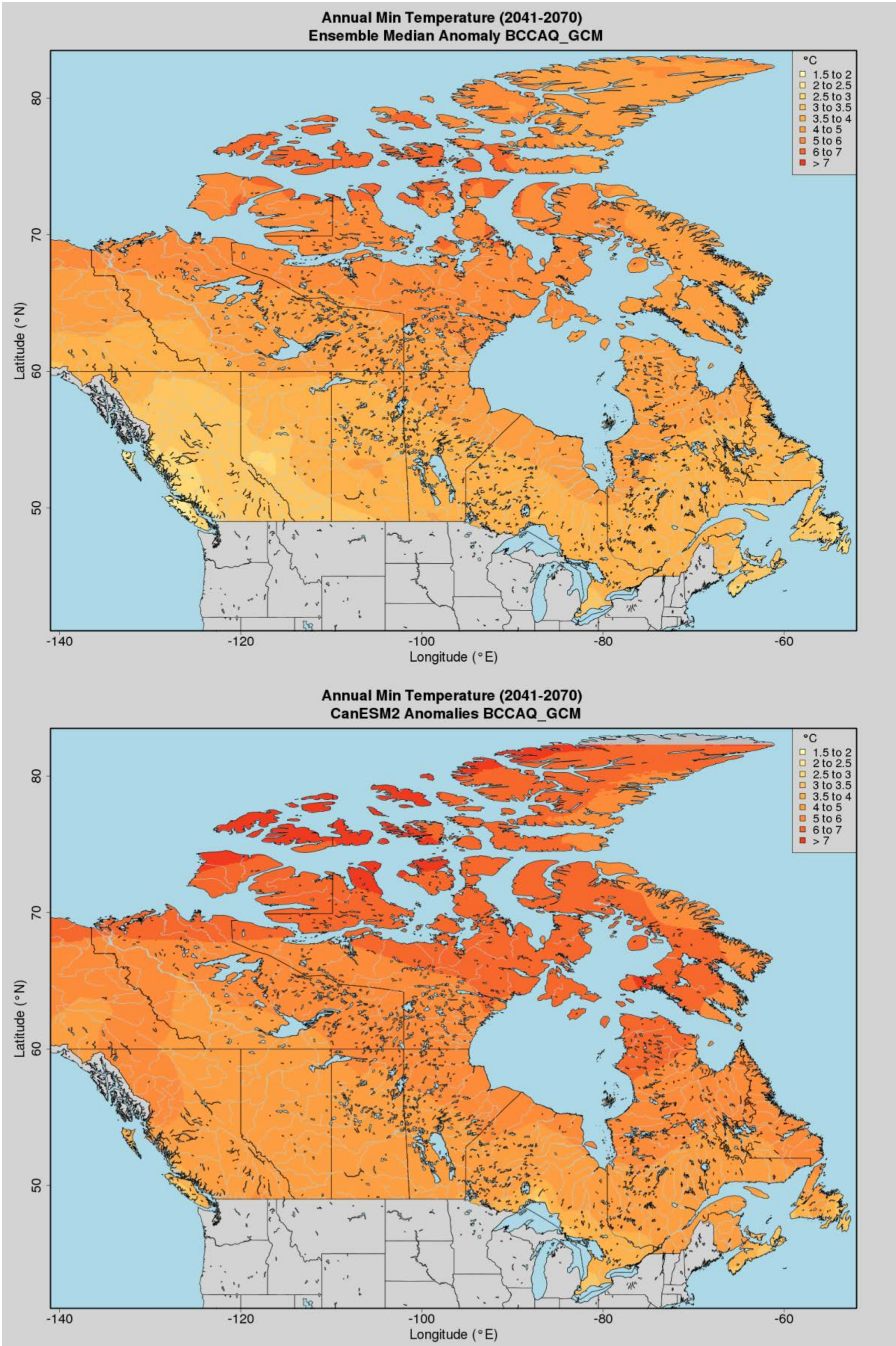

Figure 9: 2050s change in annual average tmin for ensemble median (top) and CanESM2 (bottom) for RCP8.5 over Canada. Ensemble median is computed at each grid point independently.

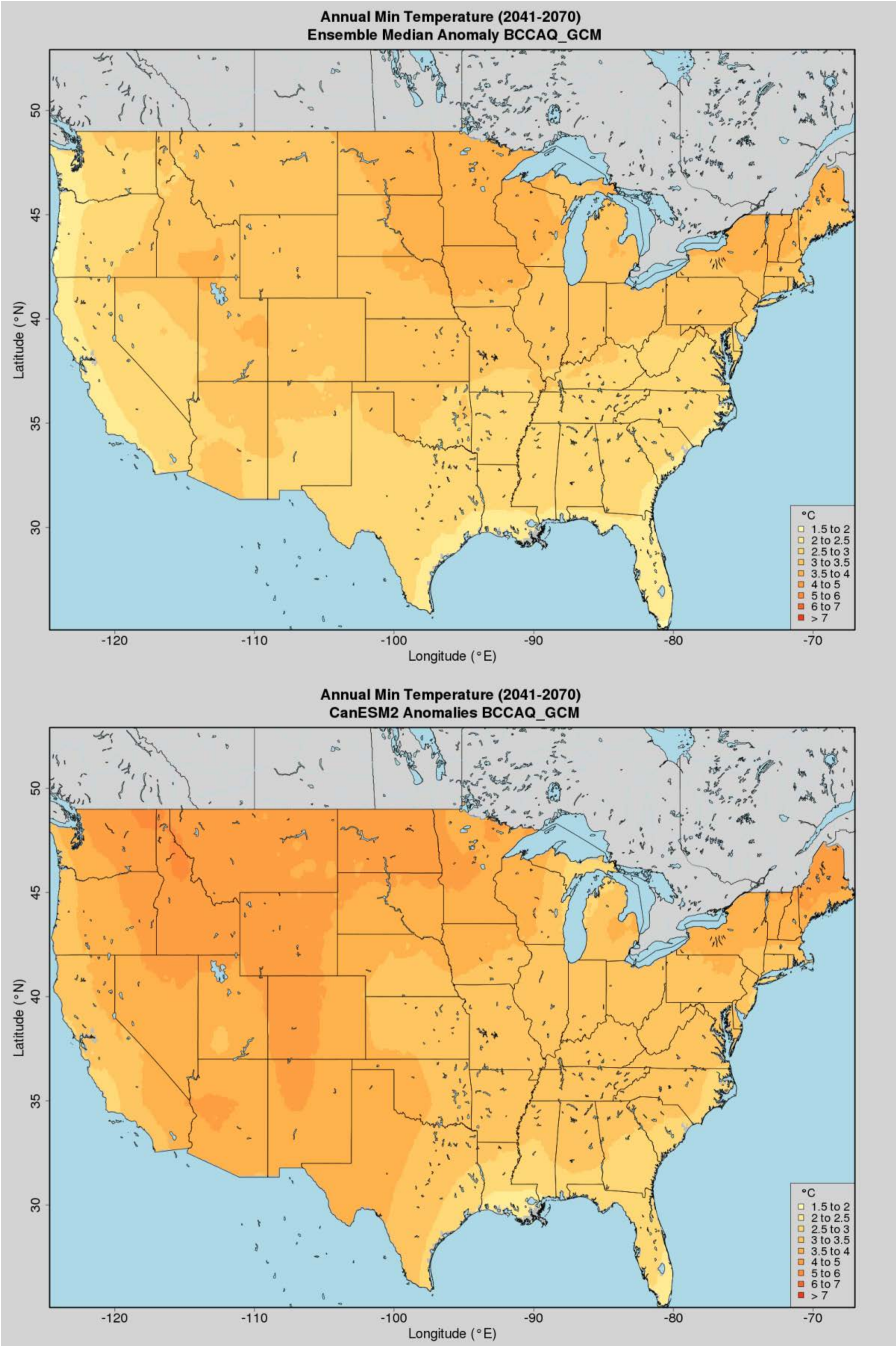

Figure 10: 2050s change in annual average tmin for ensemble median (top) and CanESM2 (bottom) for RCP8.5 over the United States. Ensemble median is computed at each grid point independently.

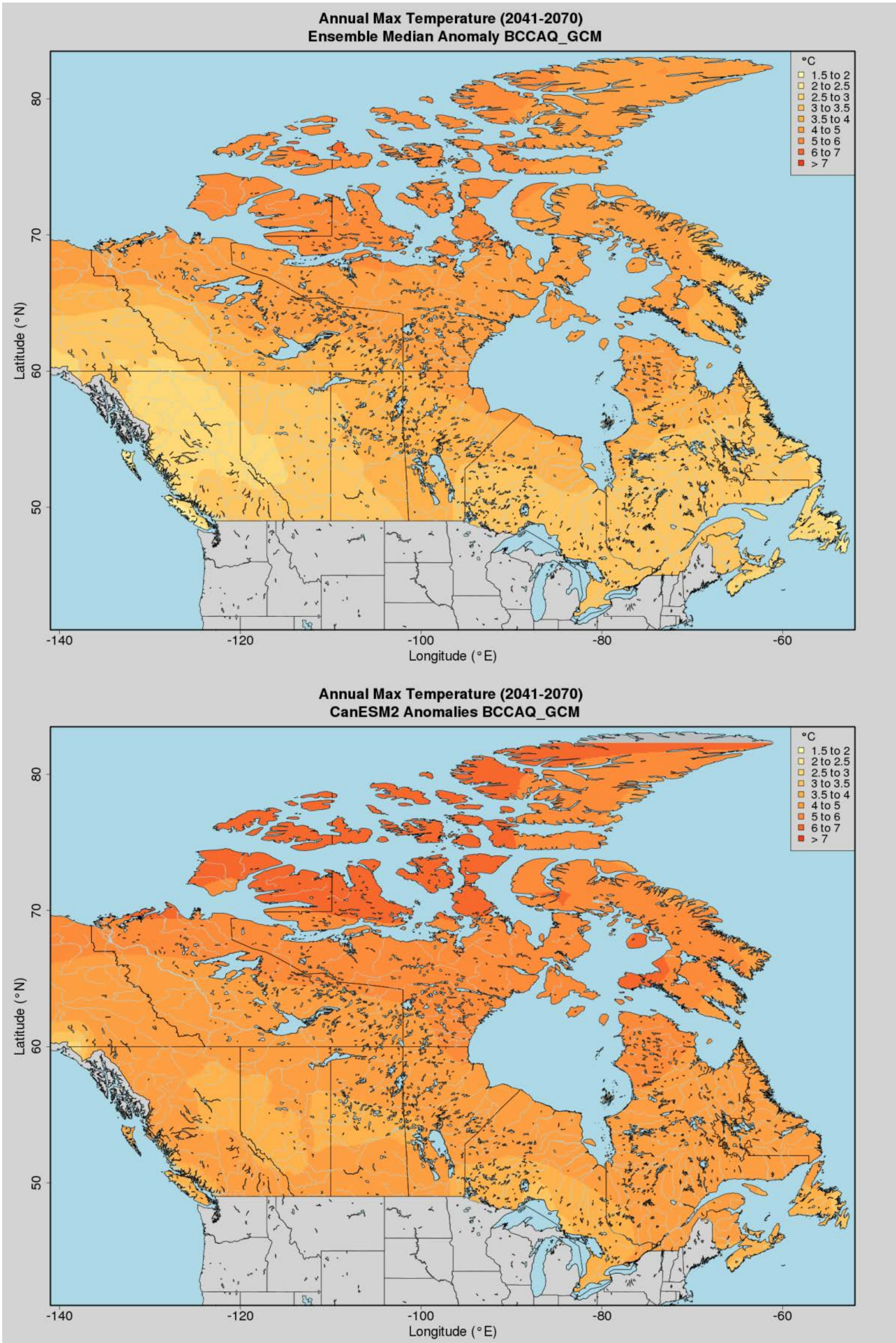

Figure 11: 2050s change in annual average tmax for ensemble median (top) and CanESM2 (bottom) for RCP8.5 over Canada. Ensemble median is computed at each grid point independently.

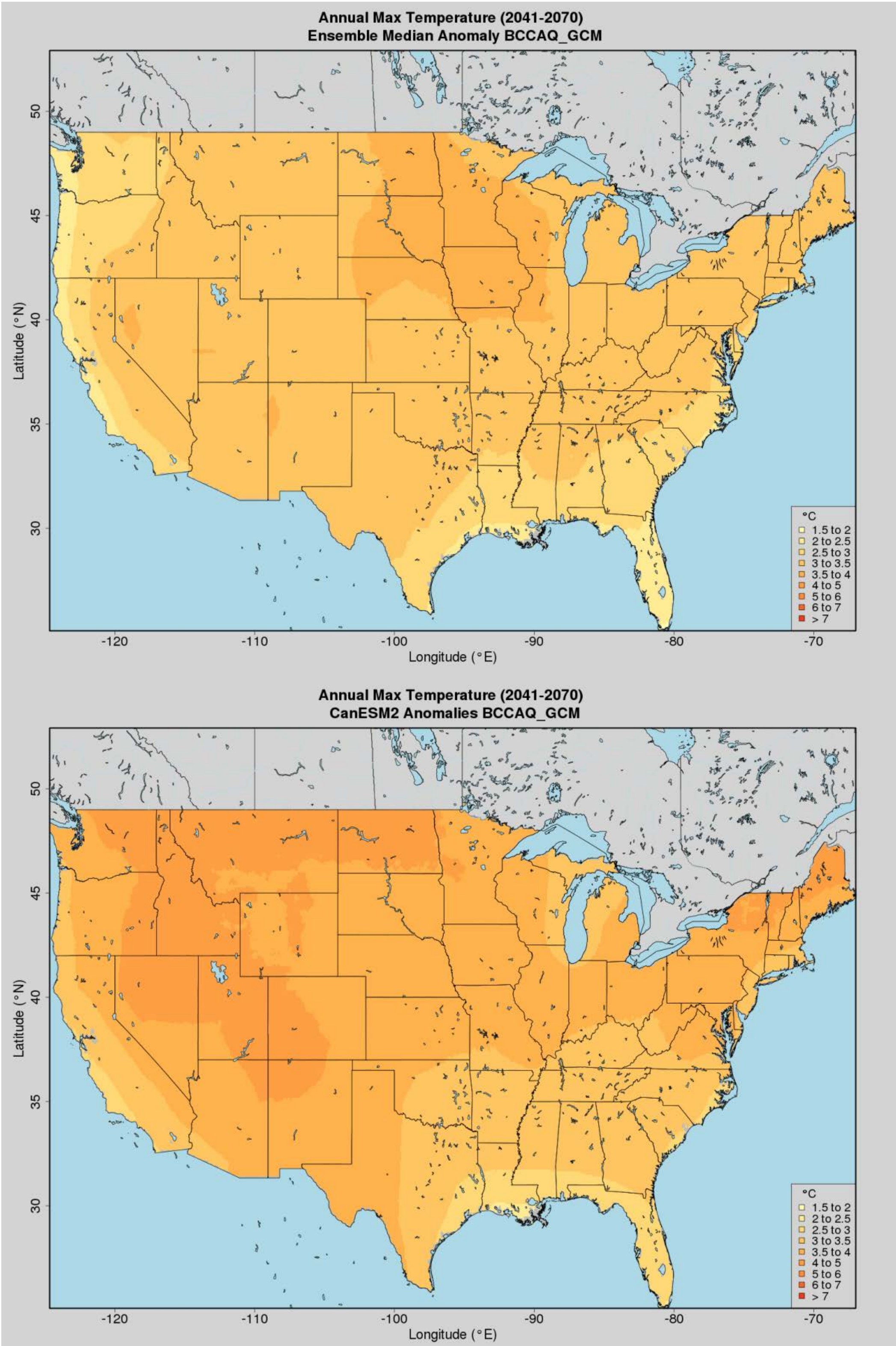

Figure 12: 2050s change in annual average tmax for ensemble median (top) and CanESM2 (bottom) for RCP8.5 over the United States. Ensemble median is computed at each grid point independently.

6c Figures – change in extremes

The figures in this section show projected change in indices of extremes and return periods based on the 12 downscaled CMIP5 RCP 8.5 runs, as described in section 4.

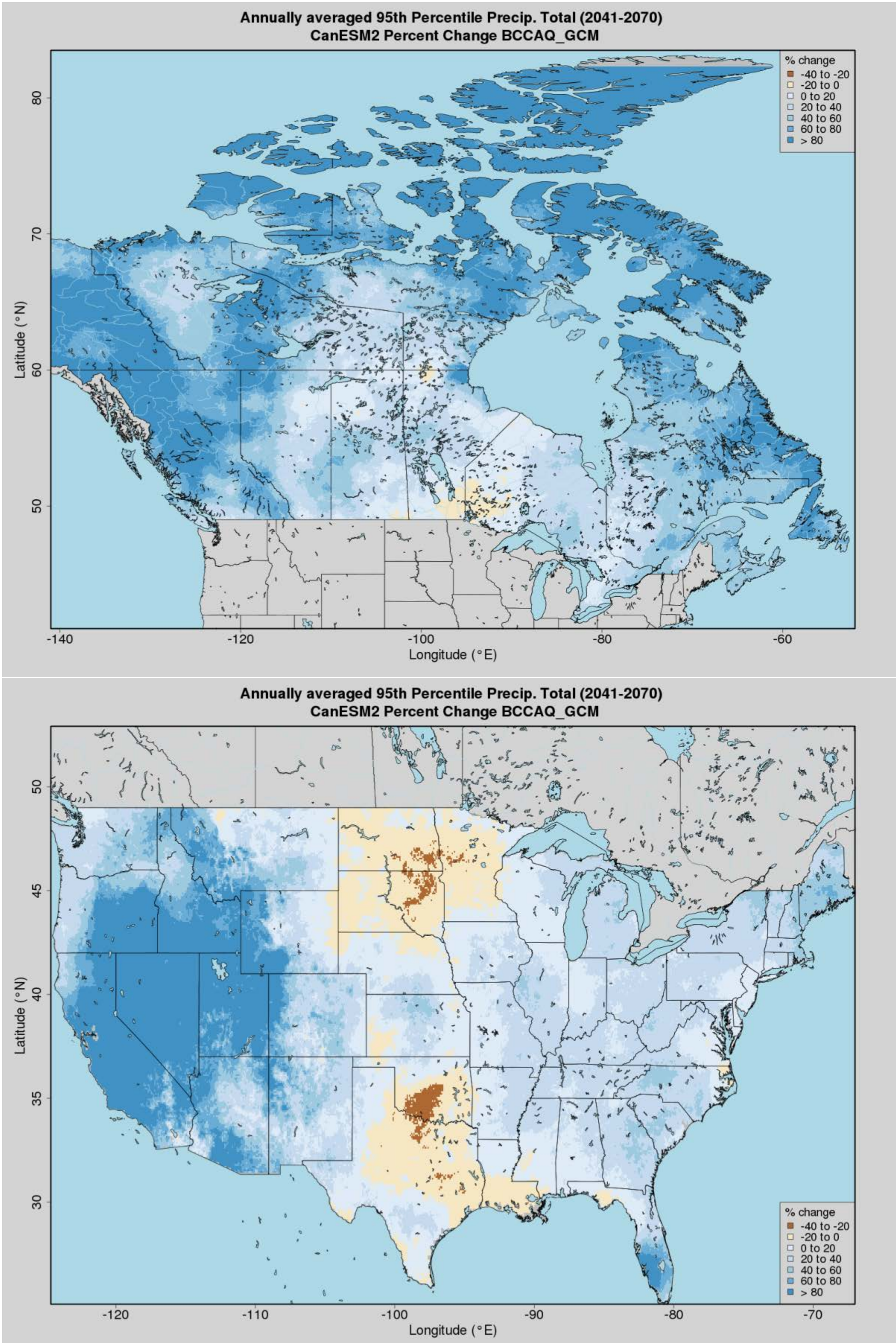

Figure 13: 2050s change in R95p for CanESM2 RCP8.5 over Canada (top) and United States (bottom).

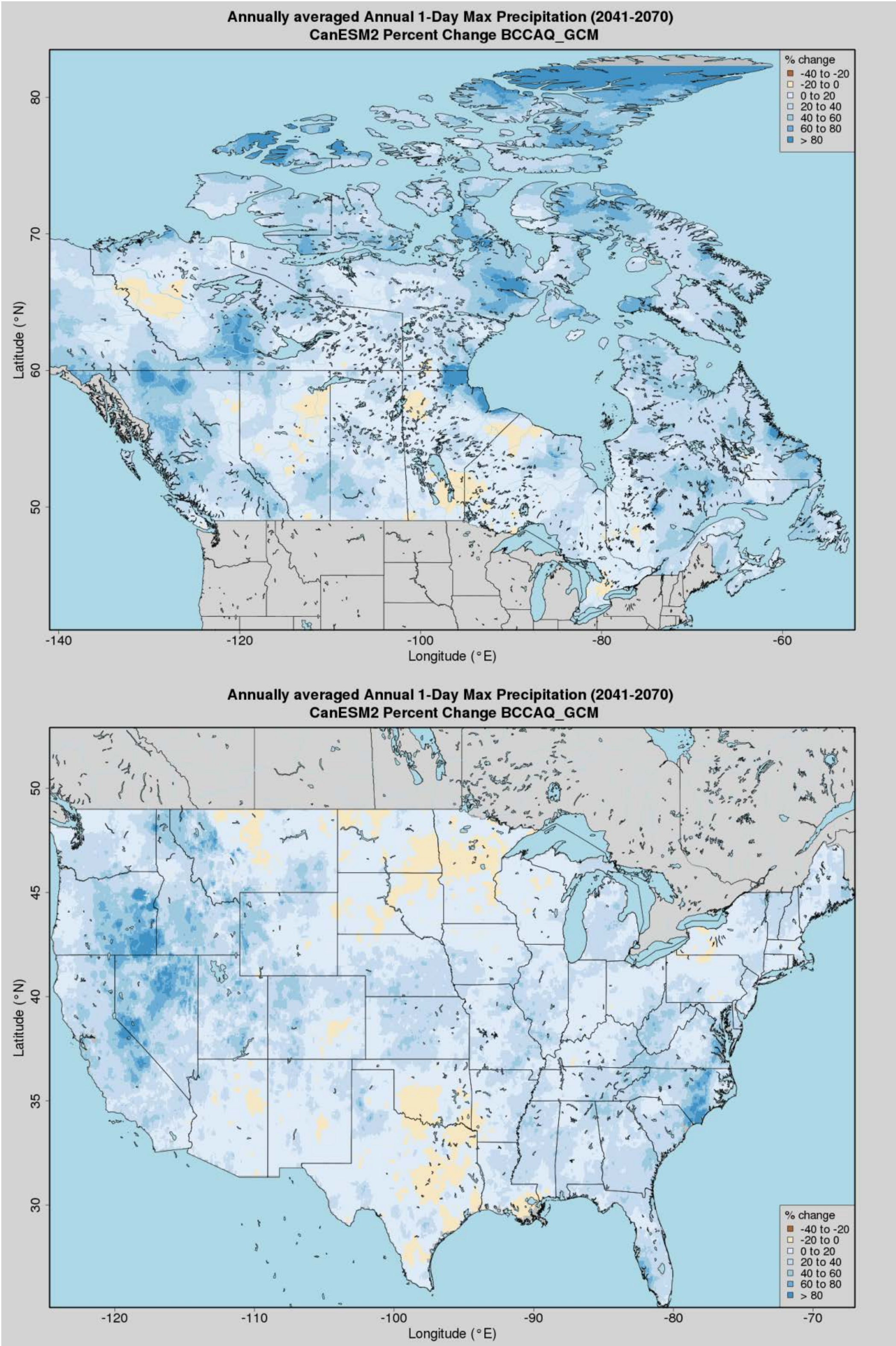

Figure 14: 2050s change in annual max of RX1day for CanESM2 RCP8.5 over Canada (top) and United States (bottom).

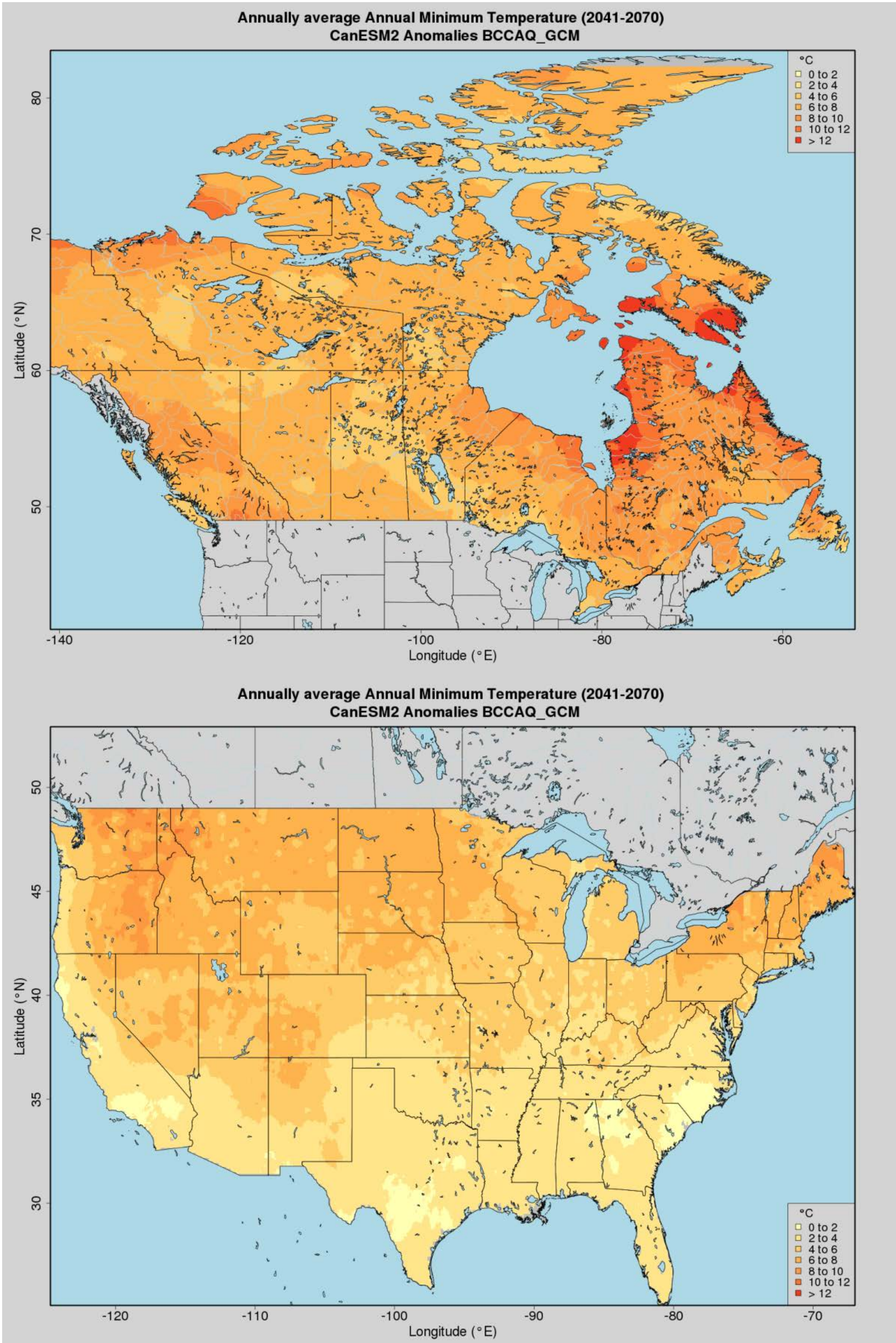

Figure 15: 2050s change in TNn (annual minima of tmin) for CanESM2 RCP8.5 over Canada (top) and United States (bottom).

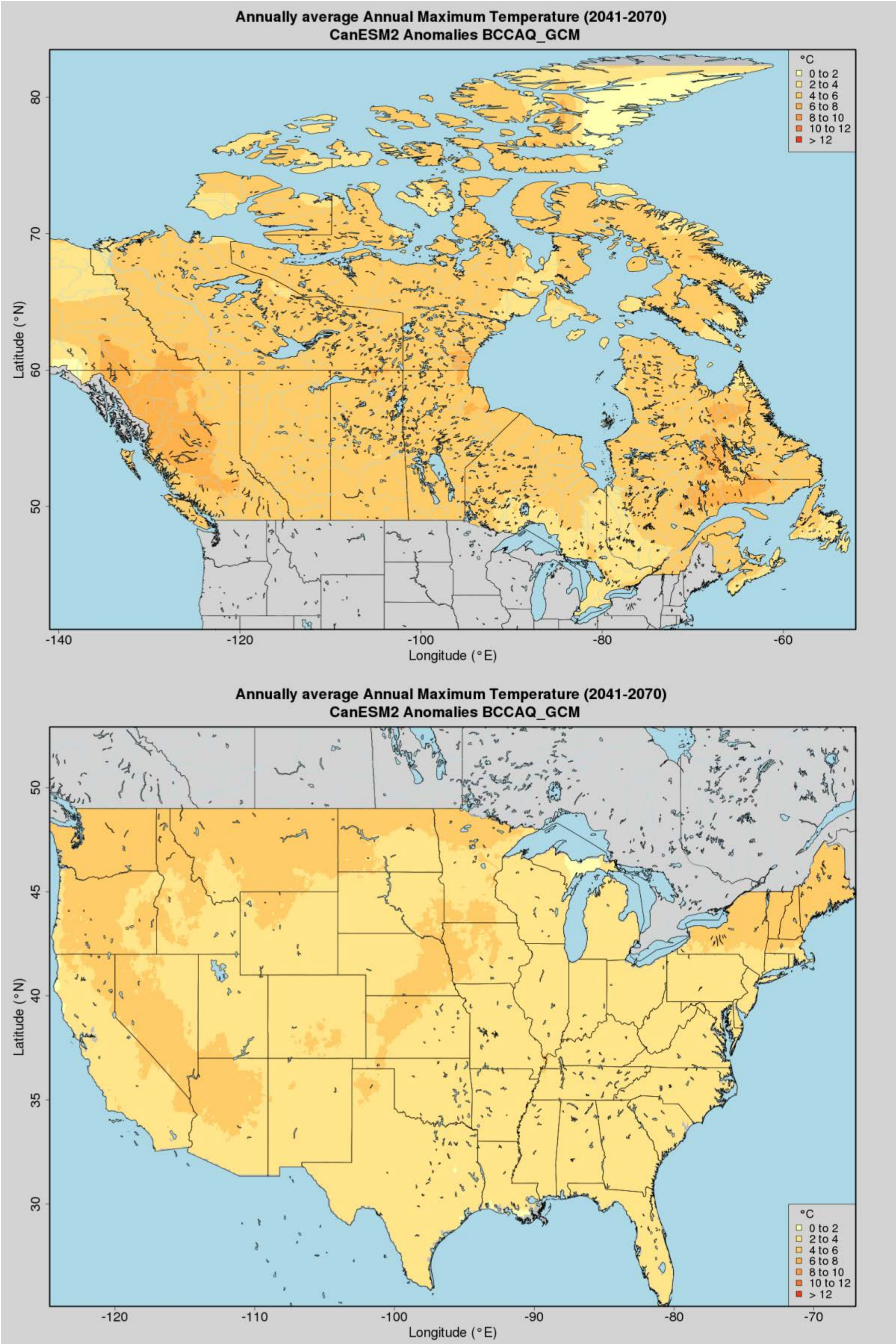

Figure 16: 2050s change in TXx (annual maxima of Tmax) for CanESM2 RCP8.5 over Canada (top) and United States (bottom).

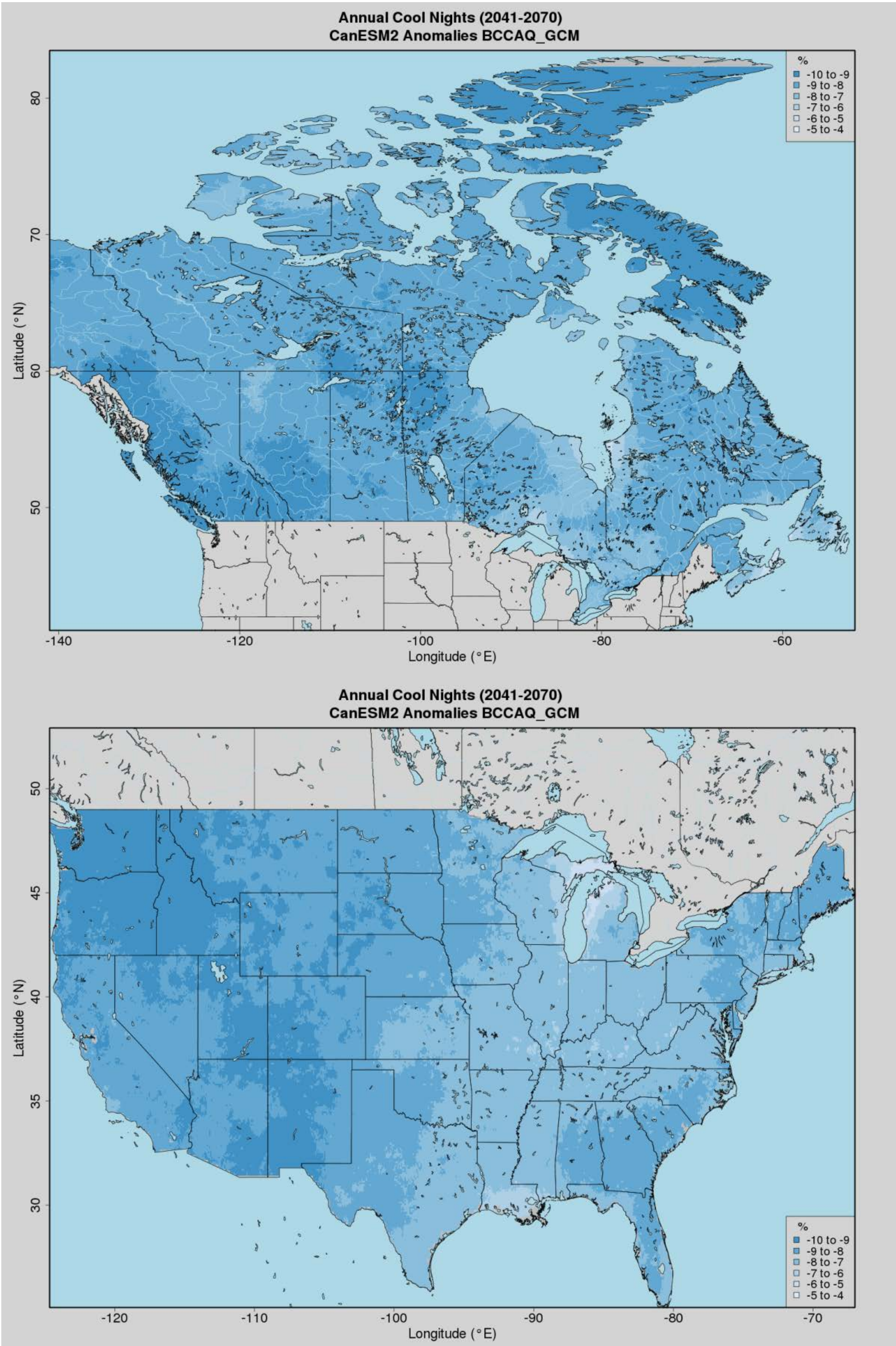

Figure 17: 2050s change in percentage of cold days TN10p for CanESM2 RCP8.5 over Canada (top) and United States (bottom).

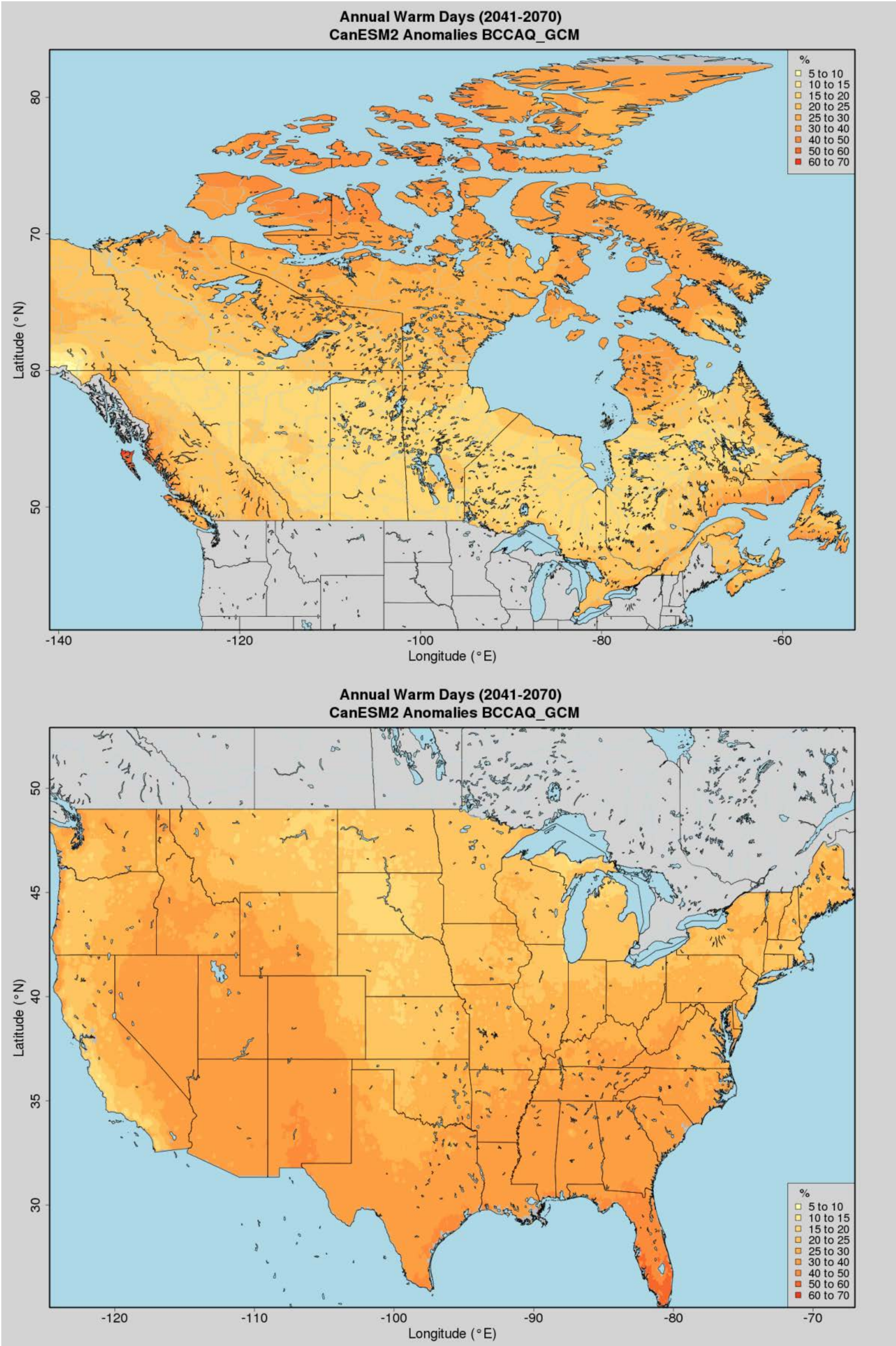

Figure 18: 2050s change in the percentage of warm days TX90p for CanESM2 RCP8.5 over Canada (top) and United States (bottom).

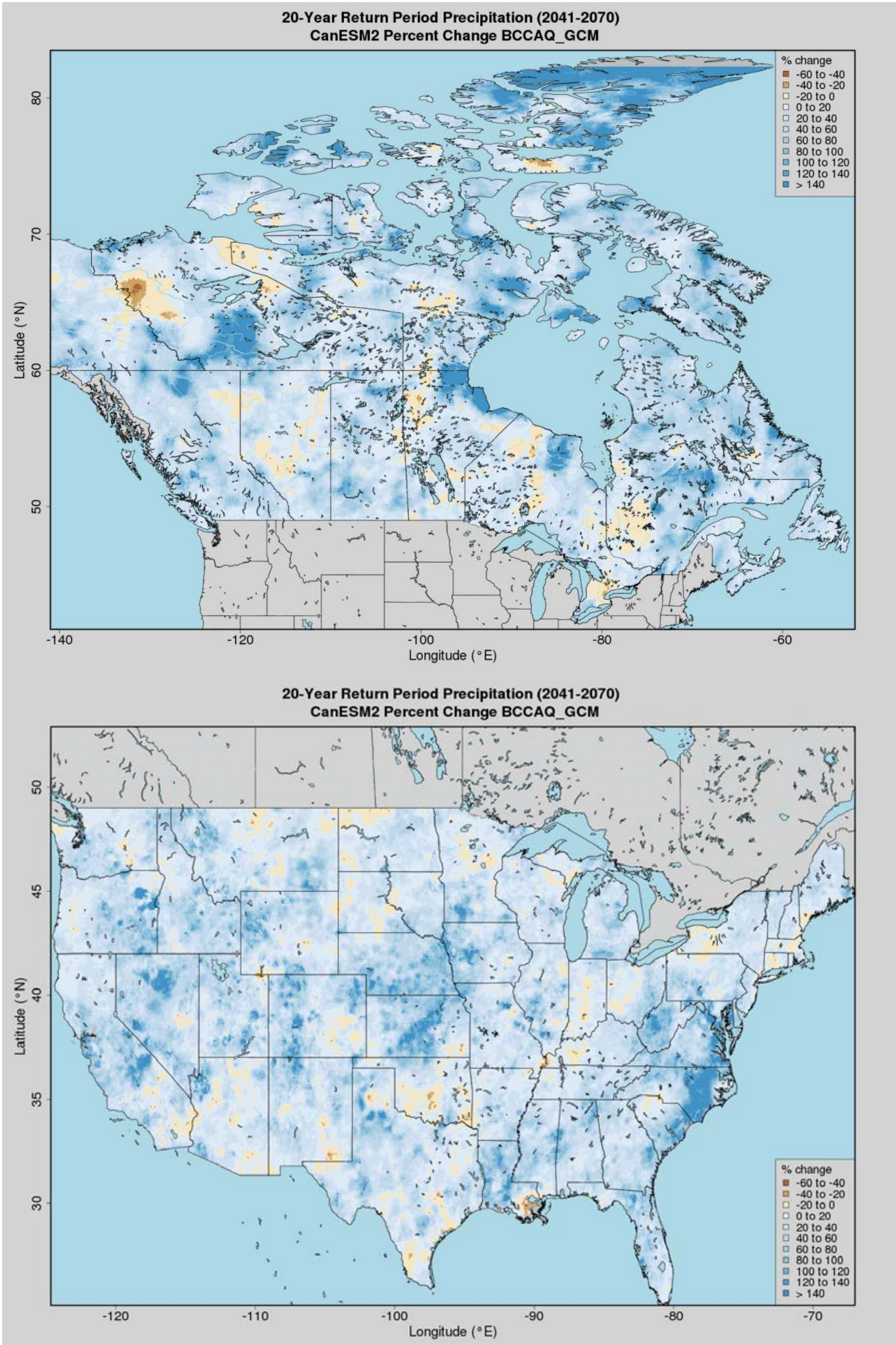

Figure 19: 2050s change in P-RP20, the daily precipitation 20-year return period event.

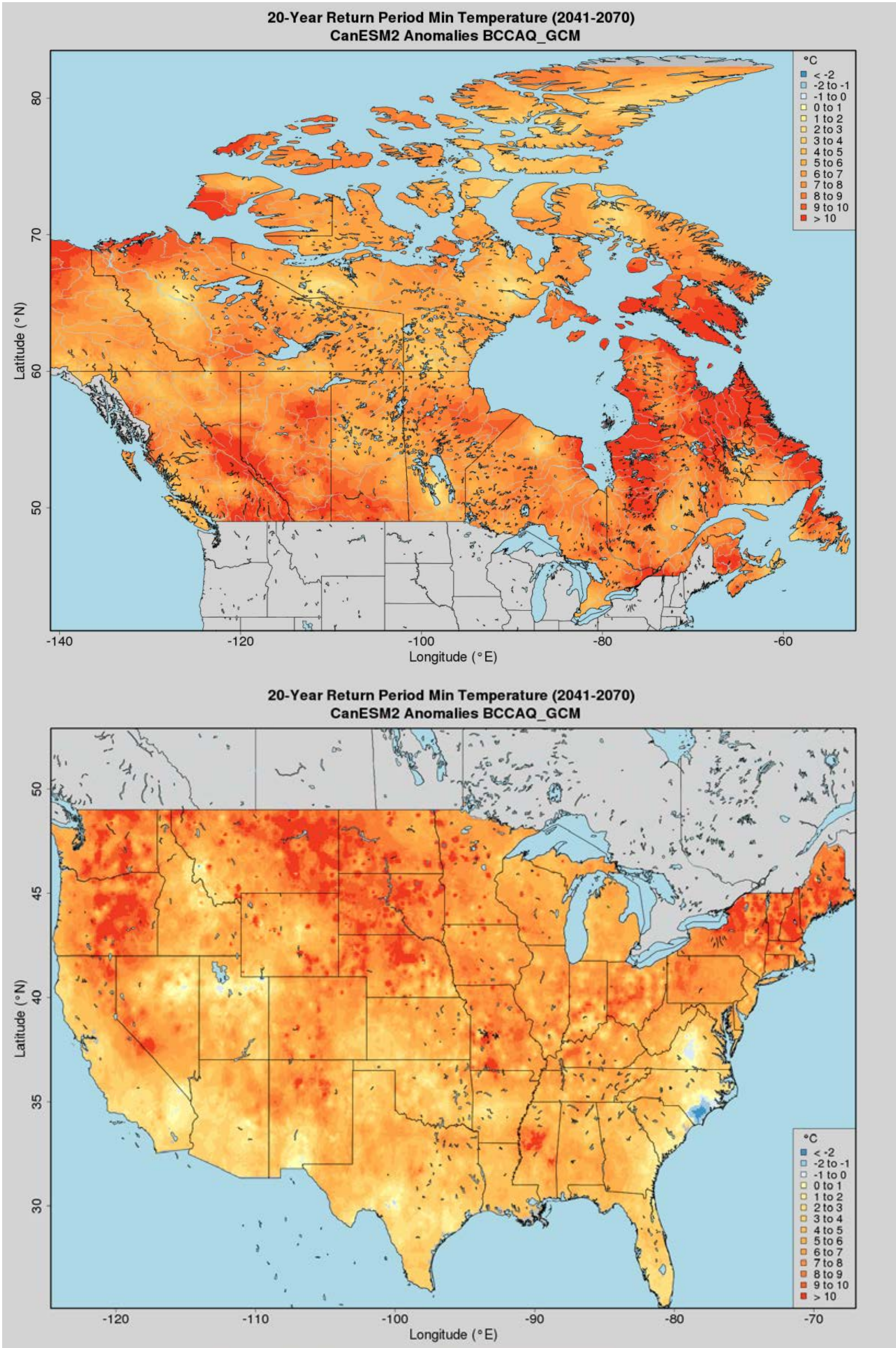

Figure 20: 2050s change in Tmin-RP20, the daily night-time low temperature 20-year return period cold event.

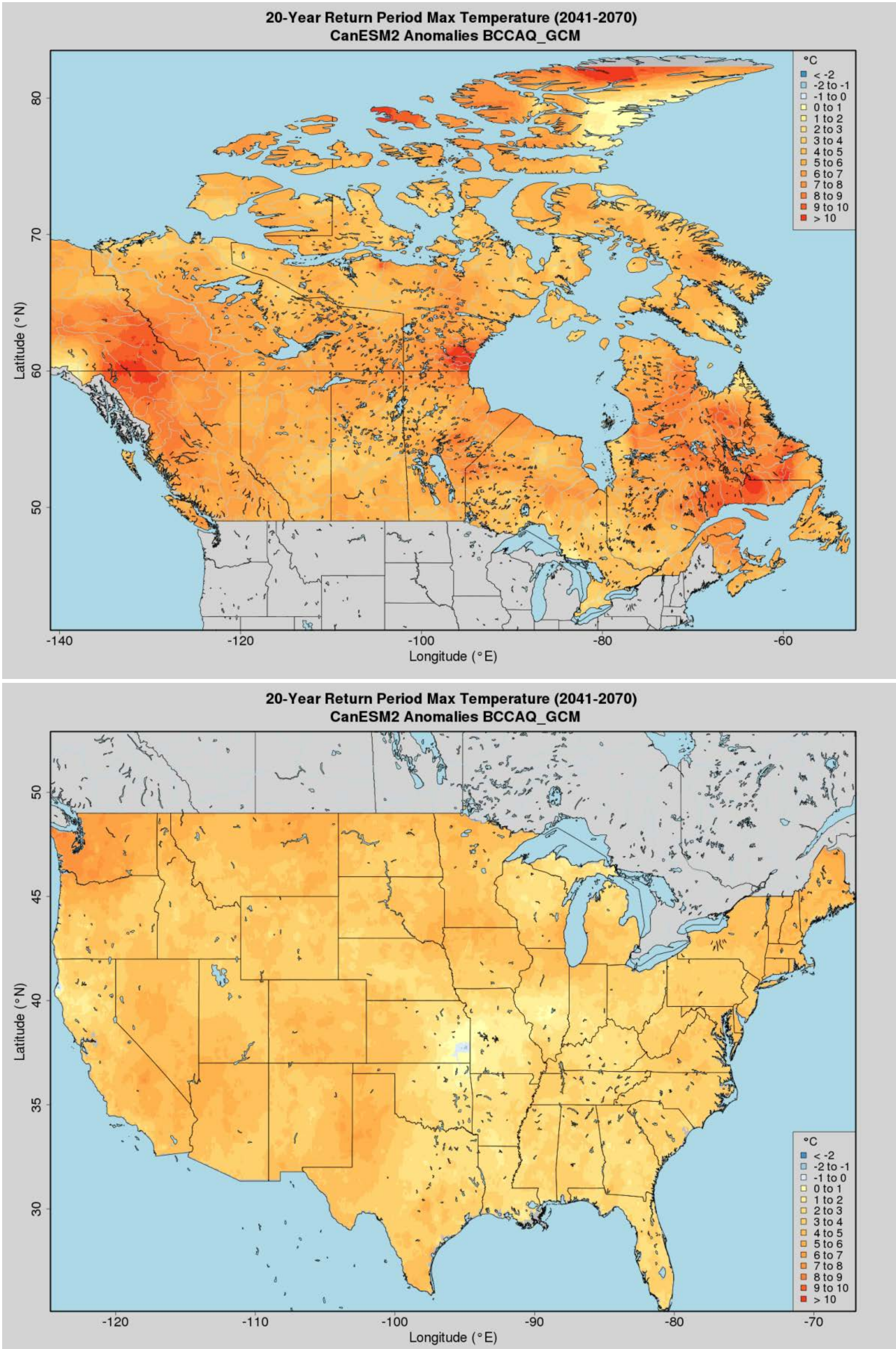

Figure 21: 2050s change in Tmax-RP20, the daily day-time high temperature 20-year return period hot event.

7 Works cited

Benestad, R. E., I. Hanssen-Bauer, and C. Deliang, 2008: *Empirical-statistical downscaling*. World Scientific, Singapore,.

Bürger, G., S. R. Sobie, A. J. Cannon, A. T. Werner, and T. Q. Murdock, 2013: Downscaling Extremes: An Intercomparison of Multiple Methods for Future Climate. *Journal of Climate*, **26**, 3429–3449, doi:10.1175/JCLI-D-12-00249.1.

Cannon, A. J., S. R. Sobie, and T. Q. Murdock, 2014: Downscaling extremes - an intercomparison of multiple gridded methods.

Gudmundsson, L., J. B. Bremnes, J. E. Haugen, and T. Engen-Skaugen, 2012: Technical Note: Downscaling RCM precipitation to the station scale using statistical transformations - a comparison of methods. *Hydrology and Earth System Sciences*, **16**, 3383–3390, doi:10.5194/hess-16-3383-2012.

Hosking, J. R., J. R. Wallis, and E. F. Wood, 1985: Estimation of the generalized extreme-value distribution by the method of probability-weighted moments. *Technometrics*, **27**, pp. 251–261.

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.

Huth, R., 2004: Sensitivity of Local Daily Temperature Change Estimates to the Selection of Downscaling Models and Predictors. *Journal of Climate*, **17**, 640–652.

Kharin, V. V., F. W. Zwiers, X. Zhang, and G. C. Hegerl, 2007: Changes in Temperature and Precipitation Extremes in the IPCC Ensemble of Global Coupled Model Simulations. *Journal of Climate*, **20**, 1419–1444.

——, ——, ——, and M. Wehner, 2013: Changes in temperature and precipitation extremes in the CMIP5 ensemble. *Climatic Change*, **119**, 345–357, doi:10.1007/s10584-013-0705-8.

Klein Tank, A. M. G., F. W. Zwiers, and X. Zhang, 2009: *Guidelines on analysis of extremes in a changing climate in support of informed decisions for adaptation*. World Meteorological Organization,.

Livneh, B., E. A. Rosenberg, C. Lin, B. Nijssen, V. Mishra, K. M. Andreadis, E. P. Maurer, and D. P. Lettenmaier, 2013: A Long-Term Hydrologically Based Dataset of Land Surface Fluxes and States for the Conterminous United States: Update and Extensions* ,#. *Journal of Climate*, **26**, 9384–9392, doi:10.1175/JCLI-D-12-00508.1.

Maraun, D., 2013: Bias Correction, Quantile Mapping, and Downscaling: Revisiting the Inflation Issue. *J. Climate*, **26**, 2137–2143, doi:10.1175/JCLI-D-12-00821.1.

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., and Coauthors, 2012: The North American Regional Climate Change Assessment Program: Overview of Phase I Results. *Bulletin of the American Meteorological Society*, **93**, 1337–1362, doi:10.1175/BAMS-D-11-00223.1.

Moran, P. A. P., 1950: Notes on Continuous Stochastic Phenomena. *Biometrika*, **37**, 17–23, doi:10.2307/2332142.

Moss, R. H., and Coauthors, 2010: The next generation of scenarios for climate change research and assessment. *Nature*, **463**, 747–756, doi:http://www.nature.com/nature/journal/v463/n7282/suppinfo/nature08823_S1.html.

Murdock, T. Q., A. J. Cannon, and S. R. Sobie, 2013: *Statistical downscaling of future climate projections*. University of Victoria, Victoria BC.

Peterson, T. C., 2005: Climate Change Indices. *WMO Bulletin*, **54**, 83–86.

Sillmann, J., V. V. Kharin, X. Zhang, F. W. Zwiers, and D. Bronaugh, 2013a: Climate extremes indices in the CMIP5 multimodel ensemble: Part 1. Model evaluation in the present climate. *Journal of Geophysical Research: Atmospheres*, doi:10.1002/jgrd.50203. http://onlinelibrary.wiley.com/doi/10.1002/jgrd.50203/abstract (Accessed March 28, 2013).

——, ——, F. W. Zwiers, X. Zhang, and D. Bronaugh, 2013b: Climate extremes indices in the CMIP5 multimodel ensemble: Part 2. Future climate projections. *Journal of Geophysical Research: Atmospheres*, **118**, 2473–2493, doi:10.1002/jgrd.50188.

Taylor, K., R. Stouffer, and G. Meehl, 2012: An overview of CMIP5 and the experiment design. *Bulletin of the American Meteorological Society*, **93**, 485–498, doi:10.1175/BAMS-D-11-00094.1.

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