

Supporting Information

Title: A closer look at novel climates: new methods and insights at continental to landscape scales.

Authors: Colin R. Mahony¹, Alex J. Cannon², Tongli Wang¹, and Sally N. Aitken¹

Institute of Origin:

¹Centre for Forest Conservation Genetics and Department of Forest and Conservation Sciences, University of British Columbia, 3041-2424 Main Mall, Vancouver, British Columbia, Canada V6T1Z4

²Environment and Climate Change Canada, 3800 Finnerty Rd, Victoria, BC V8P 5C2

Corresponding Author: Colin Mahony; c_mahony@alumni.ubc.ca; 778-288-4008

Contents

S1. Selection of reference interannual variability data	2
S2. PCA truncation threshold.....	7
S3. Subsampling the analog pool.....	9
S4. Equations for step 2 (PCA) of the Mahalanobis distance calculation.....	11
S5. CMIP5 Ensemble models	12
S6. Ensemble results	13
S7. Investigating analog outliers.....	15
S8. Accounting for non-normality in the distribution of ICV.....	17
S9. Null model analysis of elevation-novelty relationship	19
S10. Variation around mean novelty from multiple ICV proxies	20
S11. Bias due to nonrandom weather station placement.....	21
S12: Variable selection	22
S13: Comparison to standardized Euclidean distance.	30
S14: Assessment of error due to ICV sample size	34
S15: Basic code for calculating and mapping climatic novelty using the sigma dissimilarity metric.....	38
S16: Spatial data for Figure 7	38
References.....	39

S1. Selection of reference interannual variability data

Reference interannual variability is a foundation of the novelty analysis. Precipitation variability is particularly problematic because spatial correlation of precipitation is much lower than for temperature. In keeping with our use of observed historical climate normals instead of AOGCM/ESM-modeled historical normals, we required observational data for reference interannual variability in our study. We evaluated three sources of gridded precipitation variability data: two interpolated from station data—CRU TS3.23 (Harris et al. 2014) and GPCC v6 (Schneider et al. 2014)—and the JRA55 reanalysis (Kobayashi et al. 2015). A visual assessment (Figure S1) indicates that both of the station-based products (CRU and GPCC) are problematic in areas of low station density: the CRU dataset exhibits pronounced variance reduction associated with averaging between distant stations in Canada, as well as interpolation errors in the winter surface for Northeastern Alaska. The GPCC dataset, which assimilates many more stations than the CRU dataset, has several areas of anomalously high variability in the Arctic associated with individual station records that were excluded from the CRU dataset.

These artefacts suggest that the long-term JRA55 reanalysis might be preferable. However, the JRA55 dataset substantially underestimates variability relative to the CRU TS3.23 source station observations (Figure S2). This underestimation is expected, because the reanalysis provides an estimate of precipitation over the whole reanalysis grid cell, which is expected have lower variability than point observations at individual stations.

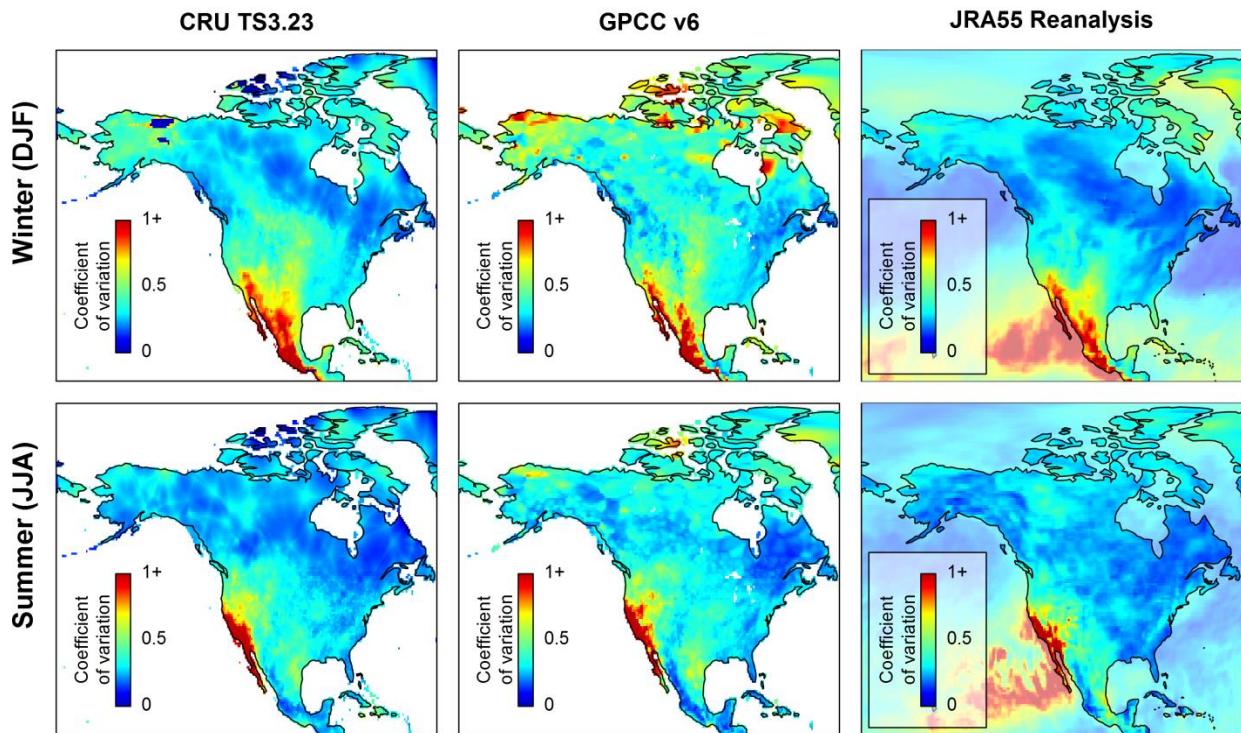


Figure S1: 1959-2013 interannual climatic variability of seasonal precipitation in North America, calculated from three gridded historical time series products: CRU TS3.23 (interpolated station data), GPCC v6 (interpolated station data), and the JRA55 reanalysis.

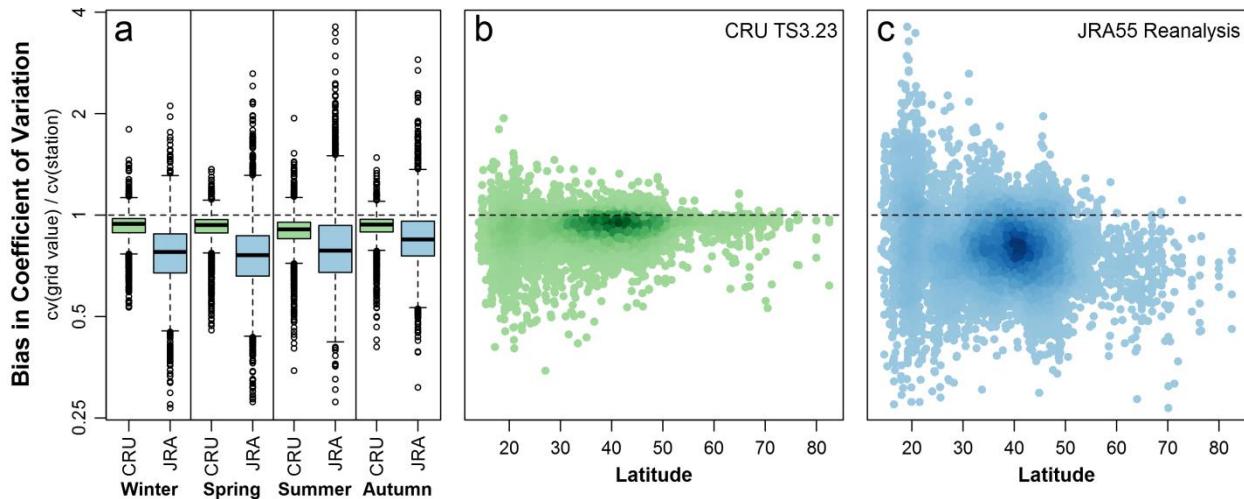


Figure S2: Assessment of bias in coefficient of variation in the CRU TS3.23 interpolated grid and the JRA55 reanalysis. Bias is calculated relative to the CRU TS3.23 source station observations. Each point in (b) and (c) represents a time series extracted from the grid at the station location. Each grid-based time series excludes years without a station observation.

As demonstrated above, gridded products have a recognized downward bias in interannual climatic variability due to variance reduction effects of interpolation and grid-box averaging (e.g., Jones et al. 2001, Director and Bornn 2015). To avoid these biases, we calculated interannual variability at the station level and used this variability in novelty calculations for nearby mapping grid cells. We selected the CRU TS3.23 source station observations for this purpose.

The first step in preparing the CRU station data is matching up station time series for tmn, tmx, and pre. Only 41% of precipitation stations in the North American study area have a station ID that matches the ID of the temperature stations (Table S1).

Table S1: Count of total North American stations in the CRU TS3.23 source observations for monthly mean daily minimum temperature (tmn) monthly mean daily maximum temperature (tmx) and total monthly precipitation (pre). Off-diagonals indicate number of stations with identification numbers matching a station in the corresponding climate element.

# of stations	tmn	tmx	pre
tmn	3166		
tmx	3116	3146	
pre	1059	1059	2557

A solution to the problem of poor station ID-matching between precipitation and temperature data is to assign the geographically closest temperature time series to each precipitation time series. If the search radius is kept small, then there is little risk of assigning an inappropriate temperature record. Even if the temperature record is not from the same station, the anomalies are likely to be applicable, due to high spatial autocorrelation of temperature anomalies documented in the CRU interpolation methods (Harris et al. 2014). This rationale does not hold for precipitation anomalies, which have low spatial autocorrelation. It follows that precipitation stations are an appropriate “base” for reference variability in the novelty analysis.

An appropriate distance cutoff needs to be selected for assigning nearest temperature stations. Figure S3 indicates that most precipitation stations in Canada and the US have a temperature station within 3km.

these can be considered effectively the same location. However, the southwestern US and Mexico have a large proportion of P stations that are 10-100 km away from the nearest T station. A cutoff of 60km is required to retain sufficient precipitation records for these areas, but potential errors associated with linking P & T records at this distance won't affect results for Canada and most of the USA.

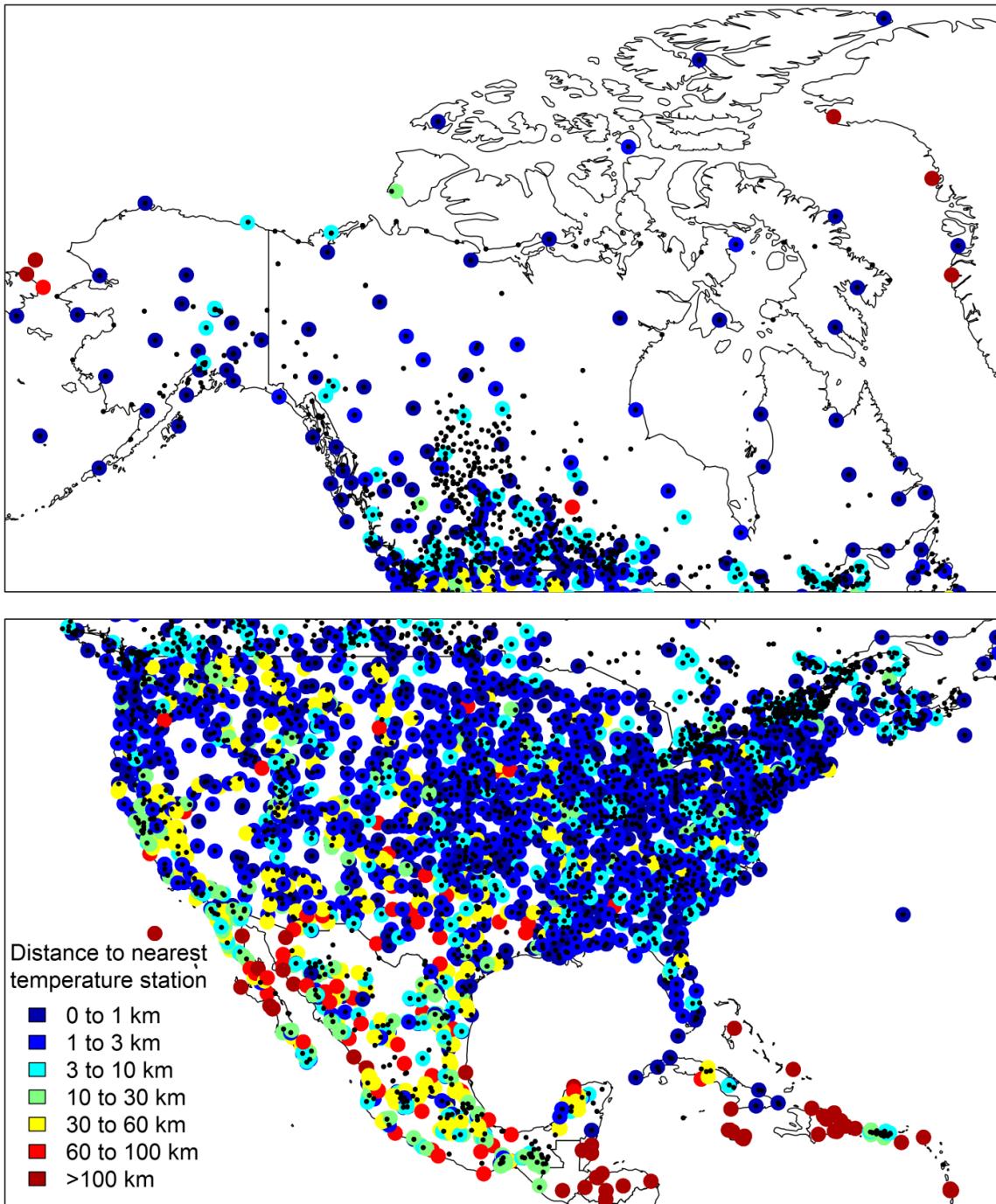


Figure S3: spatial distribution of CRU TS3.23 source stations for precipitation (coloured dots) and temperature (black dots). Precipitation stations are color coded by their proximity to the nearest temperature station.

Note in Figure S3 that there are also some clusters of precipitation stations that have only one temperature station between them. In this case, it is appropriate to assign that single temperature record to all

neighbouring precipitation stations, in recognition that the spatial autocorrelation of precipitation anomalies is much lower than temperature anomalies.

The number of stations with complete records is fairly stable between 1950 and 1985, after which there is a steep decline in precipitation records followed by a sharp decline in temperature records in 1990 (Figure S4, left). This decline is related to the CRU prioritization of stations with records in the 1961-1990 period. We used a reference variability period of 1951-1990. There are few stations with complete precipitation and temperature records (i.e., observations available for all seasons in all of pre, tmn, and tmx for a given year) for the full 1951-1990 period (Figure S4, right); 32% of stations have fewer than 31 complete years, and 9% have fewer than 21 complete years.

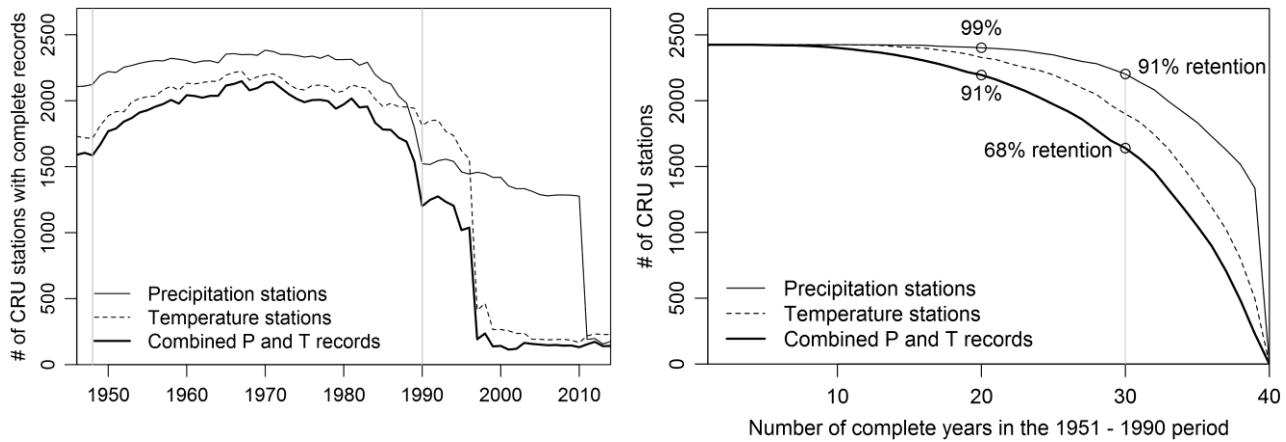


Figure S4: completeness of station data in the CRU TS3.23 source observations for north America. (left) Number of stations with complete records in the 1946-2013 period. (right) cumulative exclusion of CRU stations at increasing thresholds for number of complete years in the 1951-1990 period

Stations with insufficient records (<20 complete years) are predominantly in Mexico (Figure S5). There are no problem-free solutions to this data shortfall. To simplify the analysis, we used a low cut-off of 10 complete years for station exclusion south of 33°N , and 20 complete years north of that point. The caveat is that the novelty analysis is substantially less reliable for Mexico, since interannual variability is poorly sampled there, both spatially and temporally. For the purposes of this novelty analysis, Mexico should be seen largely as an analog pool for the United States and Canada.

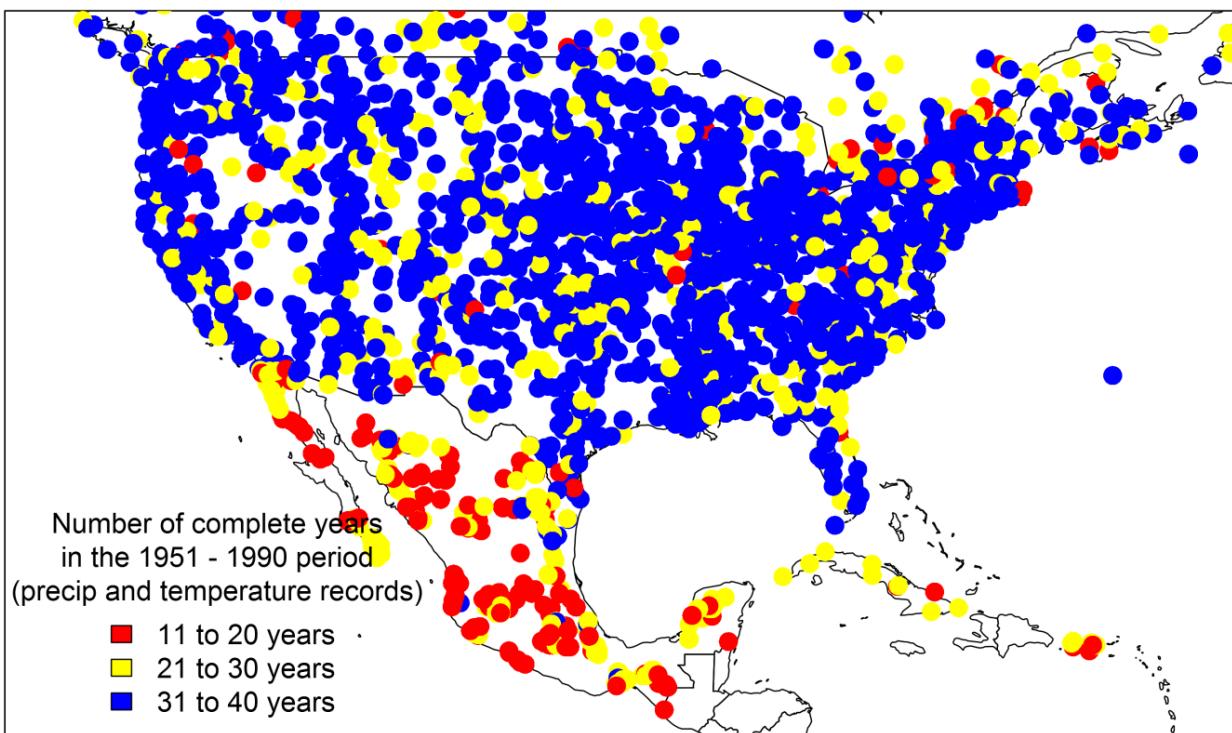
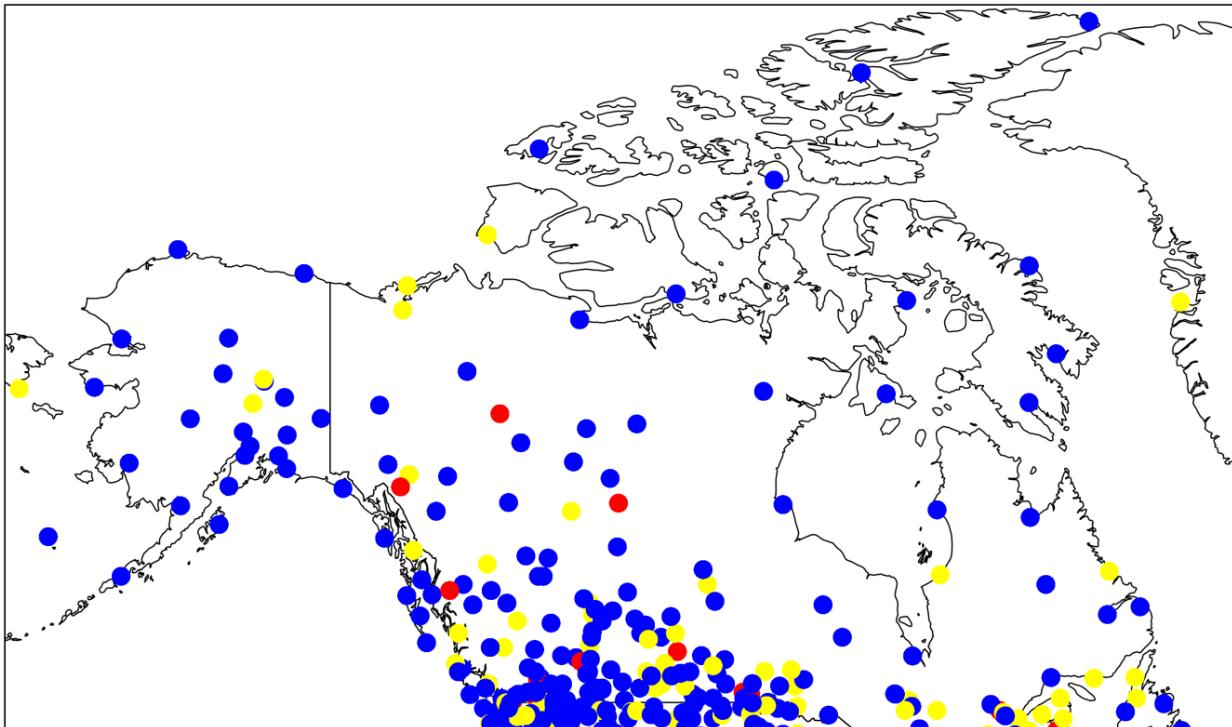


Figure S5: number of complete years (data present for all seasons in both temperature and precipitation) in the 1951-1990 period for North American CRU TS3.23 source stations.

S2. PCA truncation threshold

Step 2 of the sigma similarity algorithm uses a truncation threshold of 0.1 standard deviations (eigenvalue > 0.01) for retention of principal components (PCs). This low truncation threshold is appropriate for the purpose of the principal components analysis (PCA) within this algorithm: to scale the data space, not to compress it. The PCA rotates the data space into alignment with the PCs of local reference interannual variability, providing the frame of reference and scaling of spatiotemporal climatic differences. The factors of interest are the spatial differences in climate (spatial variation) and the trajectory of climate change (the climate change signal). These two types of information are not expected to be aggregated into the high-eigenvalue PCs of interannual climatic variability. We are particularly interested in modes of spatial variation and climate change that are large relative to interannual variability. The basic rationale for Mahalanobis distance is that there is no intrinsic basis to assign relative importance to the modes of variation in data. In Mahalanobis distance, all PCs are given equal importance, i.e., unit variance. This rationale applies to interannual climatic variability. Since we are interested in conserving spatial variation and the climate change signal in the data space, there is a strong rationale for retaining most or all of the PCs of interannual variability.

The risk of retaining PCs with very low variance is that the spatial variation or climate change signal could be grossly and artificially amplified by standardizing to a trivial variance in interannual variability. However, the distributions of climate change signal and the North American spatial variation within the local climate space of the 2304 reference stations are similar across PCs (Figure S6), with the exception of increasing numbers of outliers at higher PCs. This result indicates that the potential for artificial amplification of climate change or spatial climatic differences is isolated to these outlier stations.

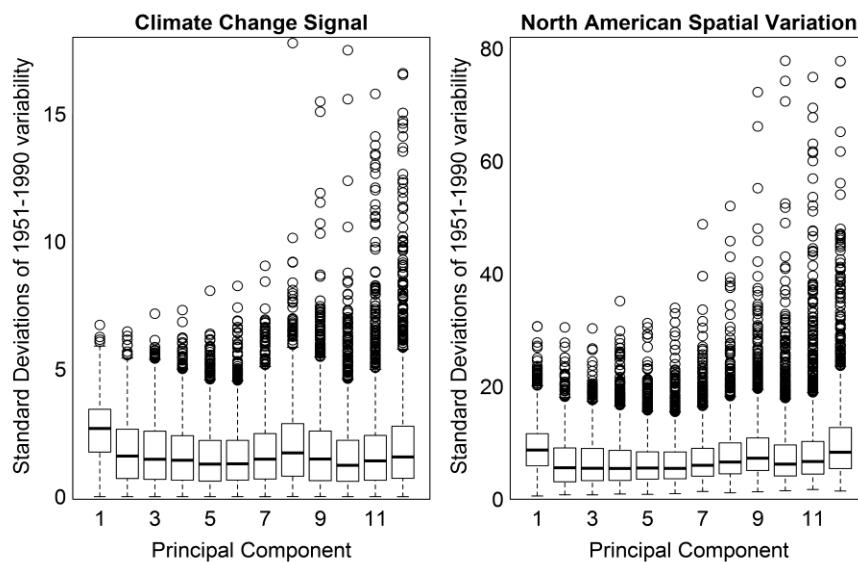


Figure S6: Local climate change signal (left) and North American spatial variation (right) within the localized data space of each reference station ($n=2304$ stations). Principal components are standardized to unit variance, which expresses the climate change signal and spatial variation as standardized anomalies of reference interannual variability. Note that these plots were made after truncating any PCs with reference variability less than 0.1SD.

The stations with outlier spatial and climate change signals are predominantly located in Mexico (Figure S7). The fact that the Mexican stations have large signals for both spatial variation and climate change suggests that this effect is most likely associated with weather stations with a low number of complete years. Alternatively, it could be a genuine aspect of the climates of Mexico (high precipitation variability

and low temperature variability). Due to the uncertainty with the cause of the large signals, removing the stations is not warranted. It is sufficient to note that the novelty analysis for the United States and Canada is not affected by signal amplification artefacts.

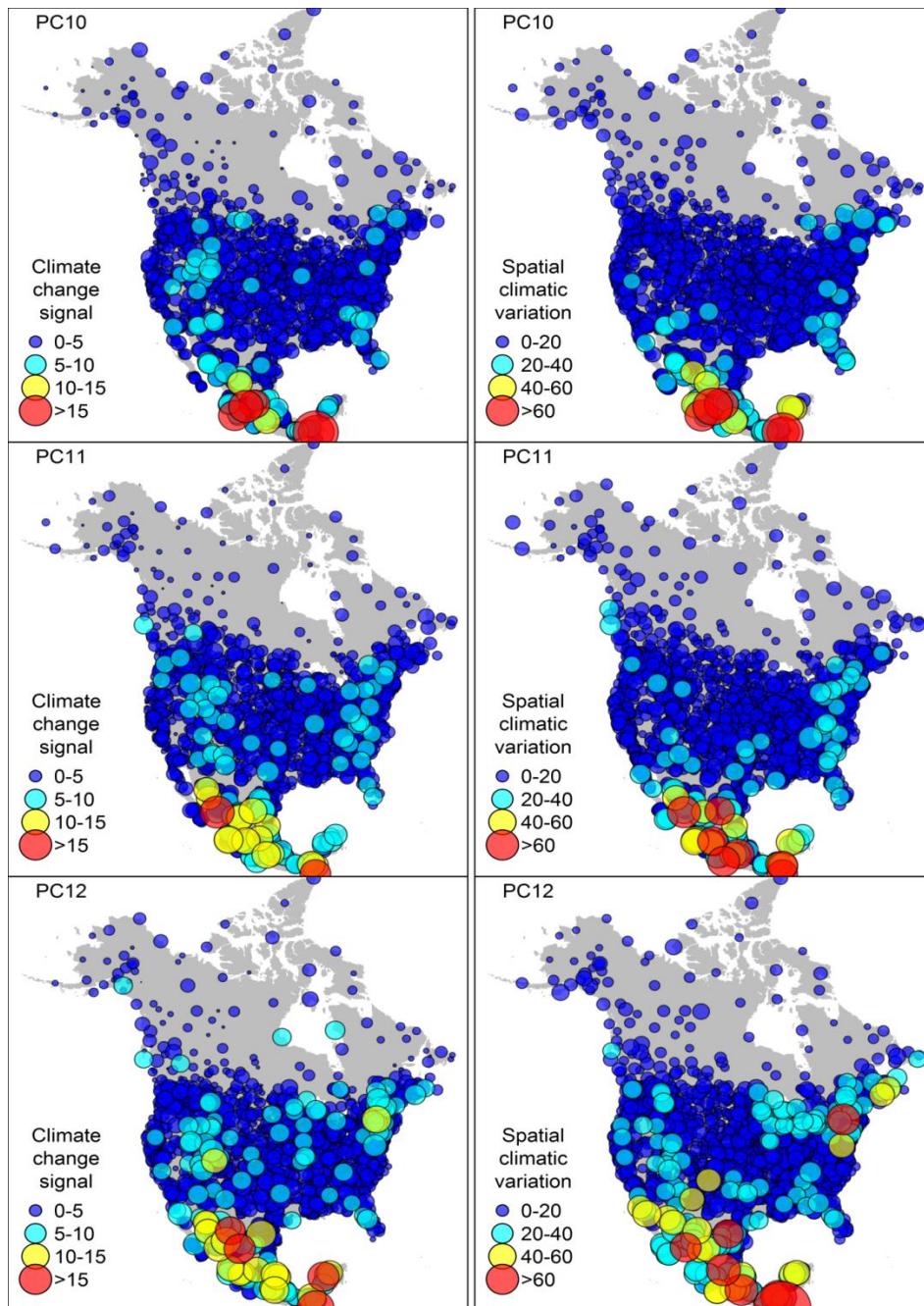


Figure S7: Local climate change signal (left) and North American spatial variation (right) expressed as standardized anomalies of the lesser principal components (PCs 10-12) of local interannual variability at each reference station ($n=2304$ stations). Stations with potential signal amplification artefacts are predominantly limited to Mexico.

S3. Subsampling the analog pool

The size of the analog pool is a major rate-limiting factor in the analog identification algorithm. To improve computational speed, we reduced the analog pool by subsampling the 4km map grid (1.3-million cells) in two steps. First, we subsampled the 30-arcsecond DEM grid into an 8km grid. Next, we applied a variable subsample inversely weighted on the standard deviation of elevation within WWF ecoregions (Olson et al. 2001). The purpose of this second subsample is to provide denser subsampling in areas of more complex terrain. Standard deviation was intentionally chosen over robust dispersion metrics in order to conserve the sensitivity of the method to elevation outliers. Sampling weights were calculated as standard deviation of the ecoregion divided by the 80th percentile of the distribution of ecoregion standard deviation across the continent, and were truncated at 0.2 and 1. This second subsample further reduced the number of analogs from about one-half from n=333,866 grid cells to n=161,032 (Figure S8), or 12% of the 4km map grid.

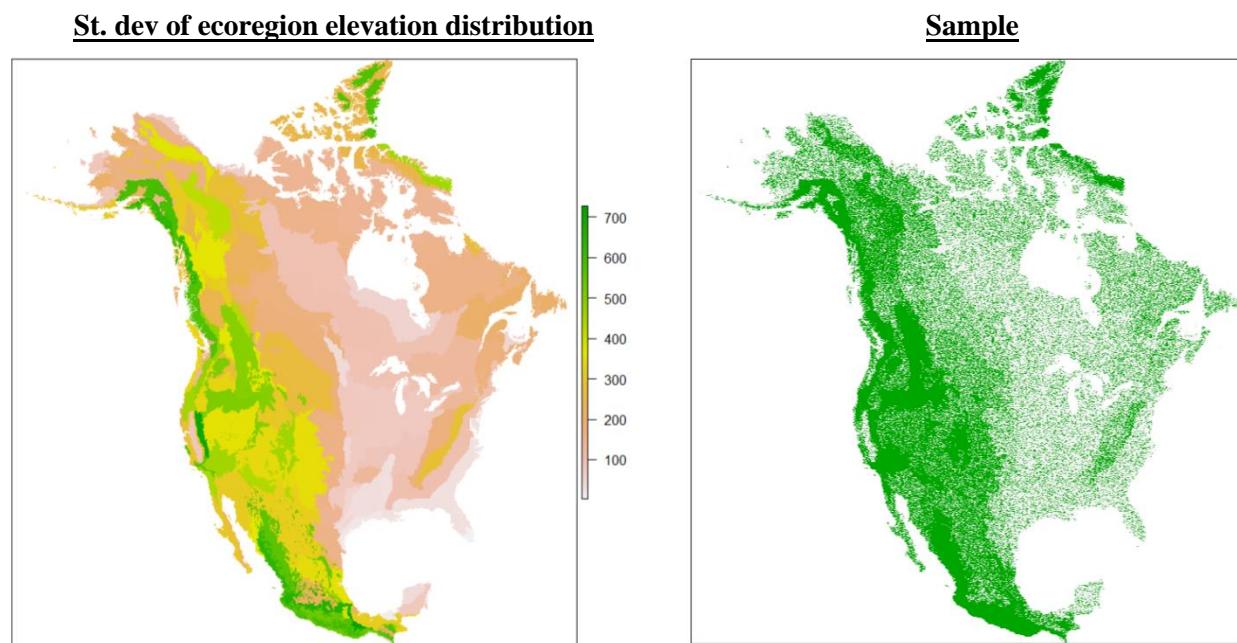


Figure S8: Development of a weighted subsample of 8-km North American grid cells (right), to reduce redundancy in the analog pool. Ecoregions were subsampled based on the standard deviation of their elevation distribution (left).

We conducted a novelty assessment for the reference period (1971-2000), providing the sigma dissimilarity between map grid cells and their best analog within the reduced analog pool (Figure S9). Only 0.02% and 0.003% of map grid cells have $>1\sigma$ and $>2\sigma$ dissimilarity, respectively, to their best analog in the reduced analog pool. These cells are dispersed, and are primarily located at high elevations. This assessment indicates that reduced analog pool adequately represents the diversity of climates present in the map grid, and results in minimal bias to the results of this study.

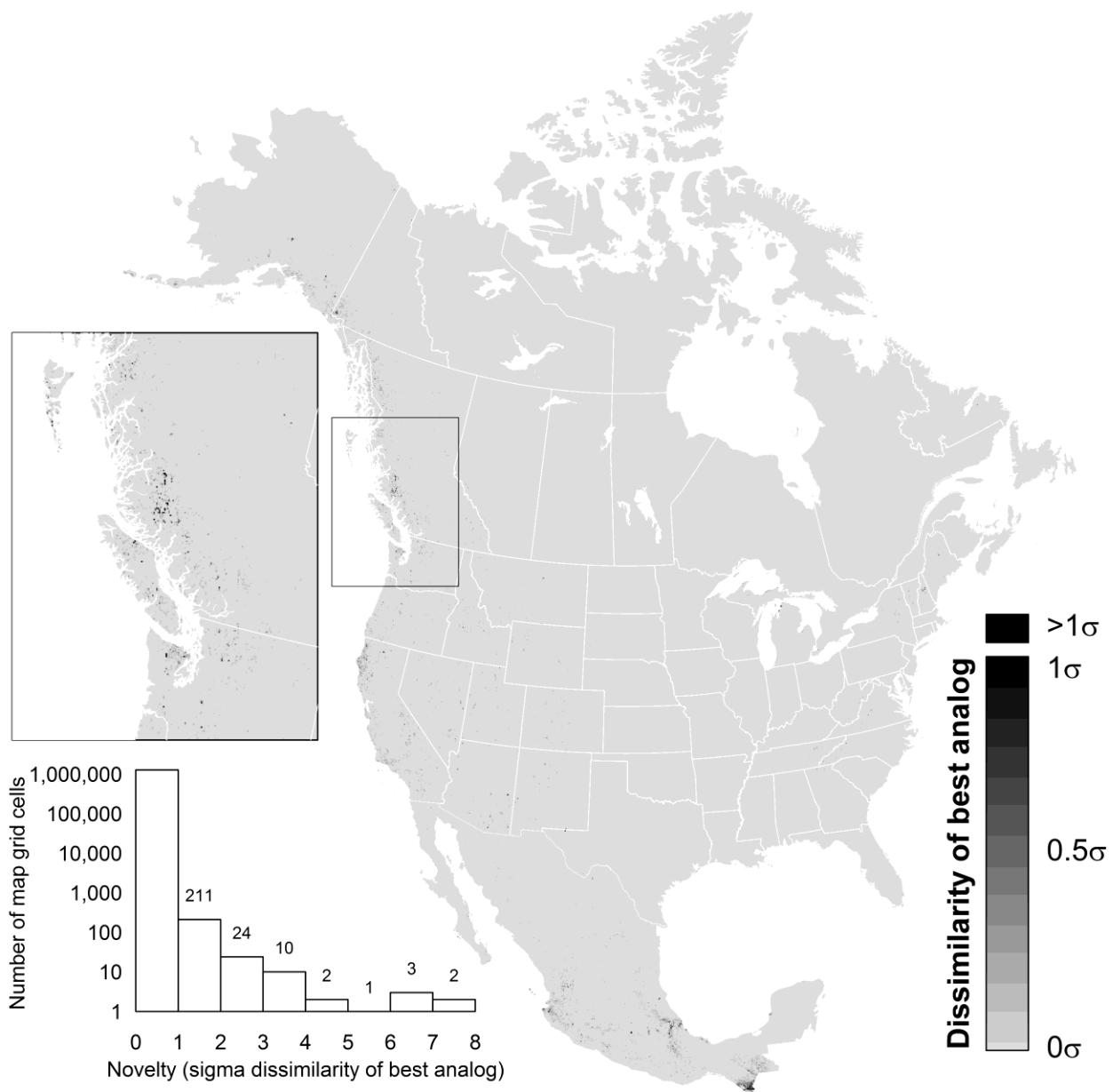


Figure S9: novelty assessment for the reference period (1971-2000), indicating locations whose climate is not adequately represented by the reduced analog pool. Note that the dissimilarity scale has been reduced from 4σ to 1σ . Only 0.02% of cells have $>1\sigma$ dissimilarity to their best analog in the reduced analog pool.

S4. Equations for step 2 (PCA) of the Mahalanobis distance calculation

The ($K \times K$) correlation matrix of $[C_j]$ is

$$[R_j] = \frac{1}{T-1} [C'_j]^T [C'_j] \quad (\text{S1})$$

A principal components analysis is completed on $[C'_j]$ by solving for the K ranked eigenvectors, \mathbf{e}_k , and corresponding K eigenvalues, λ_k , of $[R_j]$:

$$[R_j] \mathbf{e}_k = \lambda_k \mathbf{e}_k \quad (\text{S2})$$

The truncation criterion for retaining principal components is $\lambda_k > 0.01$ (standard deviation > 0.1). In other words, the dimensionality of the data space is reduced from K to M if any of the principal components of $[C'_j]$ have trivial variance. Eigenvectors meeting this criterion are assembled into $[E_j]$, a ($K \times M$) matrix whose columns are eigenvectors of $[R_j]$.

$$[E_j] = [\mathbf{e}_1 \ \mathbf{e}_2 \ \mathbf{e}_3 \dots \mathbf{e}_M] \quad (\text{S3})$$

$[A']$, $[B']$, and $[C'_j]$ are projected onto the eigenvectors of $[R_j]$, to produce linearly transformed matrices $[X]$, $[Y]$, and $[Z_j]$, respectively.

$$[X] = [A'][E_j] \quad (\text{S4})$$

$$[Y] = [B'][E_j] \quad (\text{S5})$$

$$[Z_j] = [C'_j][E_j] \quad (\text{S6})$$

S5. CMIP5 Ensemble models

Table S2: CMIP5 models included in the RCP8.5 and RCP4.5 ensemble mean projections. The model projection is an average of several model runs, as specified in the last column.

Modeling Center (or Group)	Institute ID	Model Name	# runs
Commonwealth Scientific and Industrial Research Organization (CSIRO) and Bureau of Meteorology (BOM), Australia	CSIRO-BOM	ACCESS1.0	1
Canadian Centre for Climate Modelling and Analysis	CCCMA	CanESM2	5
National Center for Atmospheric Research	NCAR	CCSM4	5
Community Earth System Model Contributors	NSF-DOE-NCAR	CESM1(CAM5)	3
Centre National de Recherches Météorologiques / Centre Européen de Recherche et Formation Avancée en Calcul Scientifique	CNRM-CERFACS	CNRM-CM5	1
Commonwealth Scientific and Industrial Research Organization in collaboration with Queensland Climate Change Centre of Excellence	CSIRO-QCCCE	CSIRO-Mk3.6.0	10
NOAA Geophysical Fluid Dynamics Laboratory	NOAA GFDL	GFDL-CM3	1
NASA Goddard Institute for Space Studies	NASA GISS	GISS-E2-R	5
Met Office Hadley Centre (additional HadGEM2-ES realizations contributed by Instituto Nacional de Pesquisas Espaciais)	MOHC	HadGEM2-ES	4
Institute for Numerical Mathematics	INM	INM-CM4	1
Institut Pierre-Simon Laplace	IPSL	IPSL-CM5A-LR	1
Japan Agency for Marine-Earth Science and Technology, Atmosphere and Ocean Research Institute (The University of Tokyo), and National Institute for Environmental Studies	MIROC	MIROC-ESM	1
Atmosphere and Ocean Research Institute (The University of Tokyo), National Institute for Environmental Studies, and Japan Agency for Marine-Earth Science and Technology	MIROC	MIROC5	3
Max-Planck-Institut für Meteorologie (Max Planck Institute for Meteorology)	MPI-M	MPI-ESM-LR	3
Meteorological Research Institute	MRI	MRI-CGCM3	1

S6. Ensemble results

Novelty assessments for individual global circulation models reveal a very large spread in overall extent of novelty projected for the RCP4.5 scenario (Figure S10). Differences between models are generally in magnitude rather than spatial pattern, except in the boreal and arctic regions.

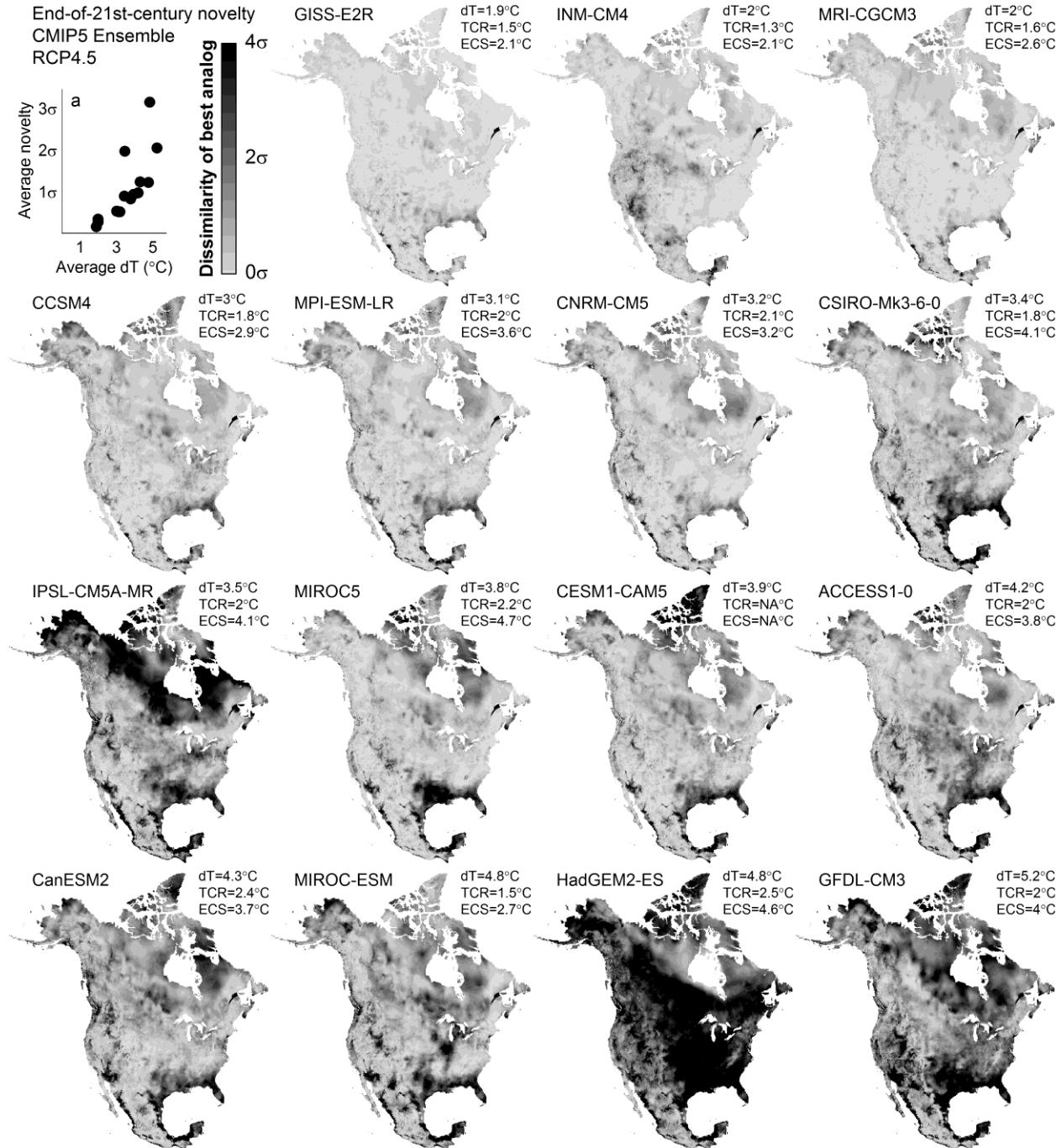


Figure S10: Novelty analyses for individual RCP4.5 projections of the 15-model CMIP5 ensemble. Models are ordered by the North American average dT, the difference in mean annual temperature between the 2071-2100 and 1971-2000 normal periods. Other model attributes are quoted from Forster et al. (2013): transient climate response (TCR) of global surface temperature to a 1%/yr increase in CO₂ and equilibrium climate sensitivity (ECS) to an instant quadrupling of CO₂. (a) plots the relationship between north American average dT and novelty for each of the 15 models.

Figure S11 demonstrates that the average of 15 separate novelty assessments for the individual ensemble model projections is very similar to the result obtained from a single analysis of the ensemble mean projection. The ensemble analysis results in a slightly higher level of novelty, but the spatial pattern is equivalent.

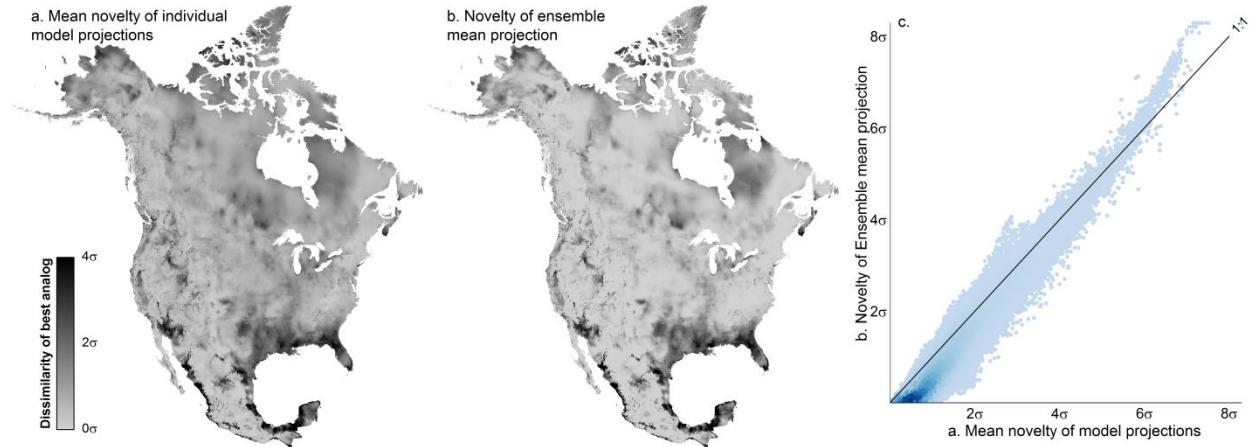


Figure S11: relationship (c) between novelty calculated from a single ensemble mean projection (b) and the average of 15 separate calculations for the 15 individual models in the ensemble (a).

S7. Investigating analog outliers

Novelty can be underestimated if outliers from the North American climate envelope are identified as analogs for a large number grid cells. Seasonal temperature-precipitation plots of the North American analog pool (Figure S12) show a few serious outliers in the analog pool. However, none of these visible outliers are acting as important analogs in the novelty analysis. In addition to the plots shown in Figure S12, there were no important outlier analogs in any combinations of seasonal Tmin/Tmax/PPT for either RCP.

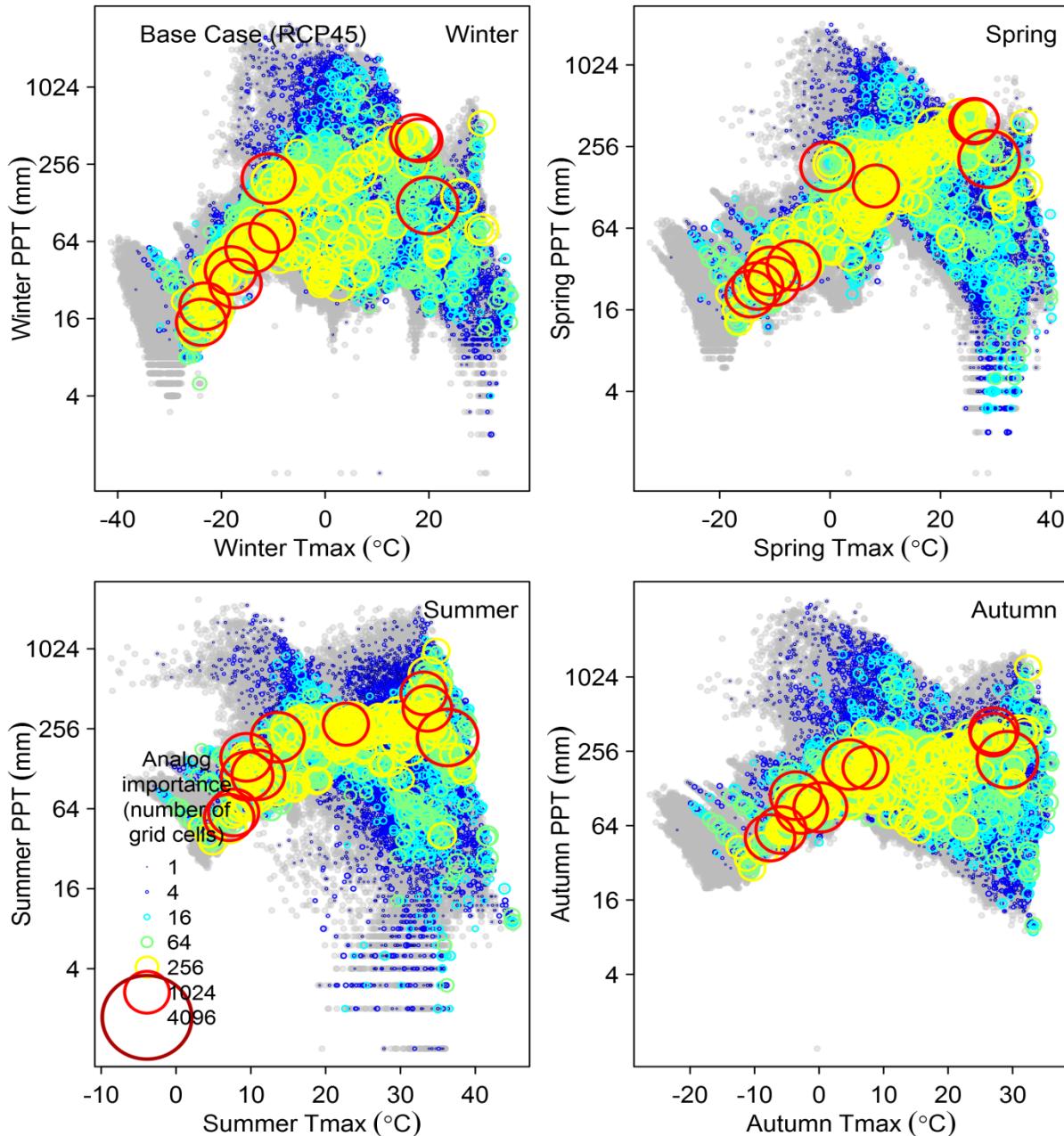


Figure S12: Seasonal temperature-precipitation plots of the North American analog pool. Analog importance is measured as the number of grid cells for which a candidate analog is the best analog in the novelty calculation. None of the visible outliers have high analog importance.

Maps of analog importance (Figure S13, left) and analog outliers (Figure S13, right) indicate that there are no outliers with high importance. A similar comparison was done for RCP8.5, with similar results (not shown). This strongly suggests that analog outliers do not have a major influence on the results of the novelty analysis.

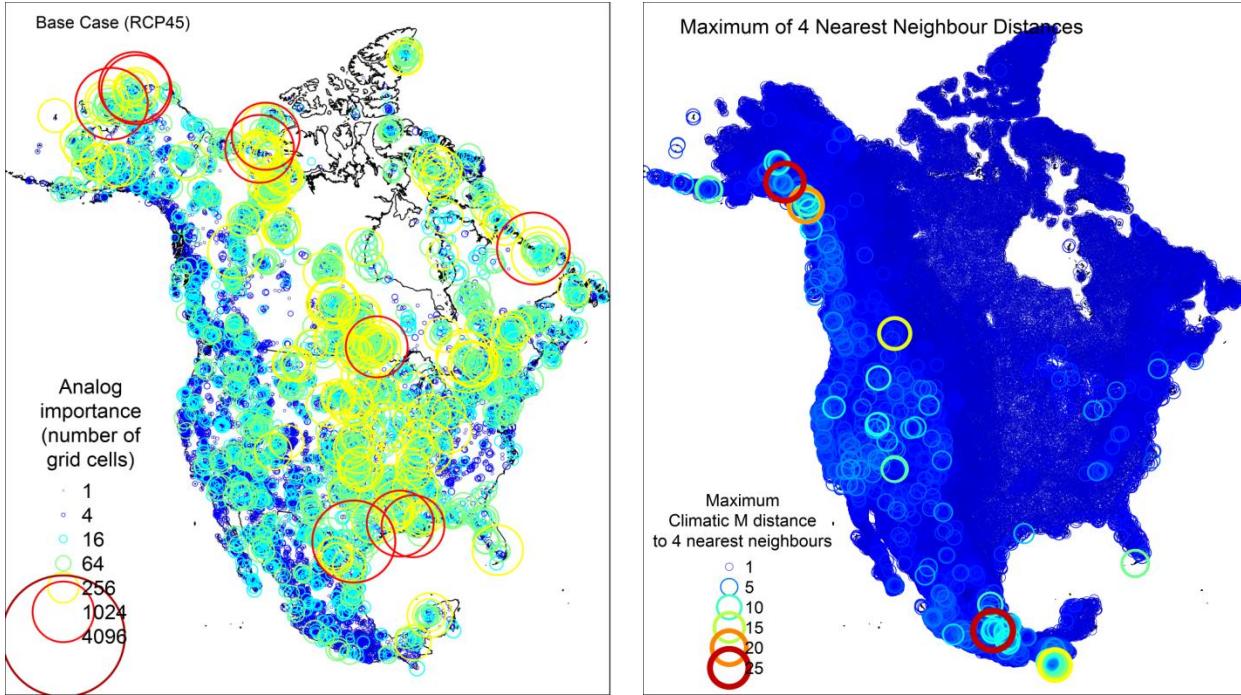


Figure S13: Outliers to the North American climate envelope do not have high analog importance. (left) Analog importance is measured as the number of grid cells for which a candidate analog is the best analog in the novelty calculation. (right) Outlier distance is measured as the localized climatic Mahalanobis (M) distance from each candidate analog to the furthest of its four nearest neighbours. This approach ensures that clusters of up to three outliers can be detected.

S8. Accounting for non-normality in the distribution of ICV

The novelty metric applied in this paper doesn't make any accommodations for non-normality in the distribution of ICV. The univariate ICV distributions could be normalized using the approach commonly used to calculate the standardized precipitation index (Guttman 1999): fit a PDF (e.g. 3-parameter gamma), and use quantile matching to the standard normal PDF to translate the variability into z-scores (Figure S14). This transformation would move the data towards multivariate normality by normalizing the marginal distributions, but would not necessarily achieve multivariate normality (which can't be convincingly detected/rejected anyways due to small sample size of ICV). This procedure could in theory be used as a substitute for z-standardization within the algorithm of the sigma dissimilarity metric.

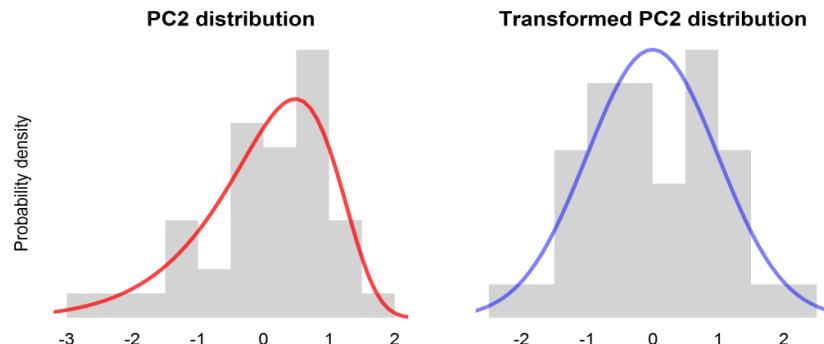


Figure S14: illustration of data transformation using quantile matching to a fitted PDF: (1) a generalized gamma PDF is fitted to the data; (2) the data are expressed as percentiles of the fitted PDF; and (3) these percentiles are then expressed as values of the standard normal PDF, i.e. as z-scores.

Unfortunately, the quantile matching approach doesn't work for this novelty analysis because the range of climates in the focal neighbourhood (grid cells associated with any given reference climate station) can be well beyond the range of decimal precision of the fitted PDF (Figure S15). Note that this precision issue would not be a limitation if the transformation was calculated at the cell level rather than the station level: values beyond the precision horizon of the cell's local ICV would be irrelevant to the calculation of novelty. The quantile-matching transformation approach is preferable to z-standardization for a multivariate standardized anomaly, but it is not feasible for this novelty analysis.

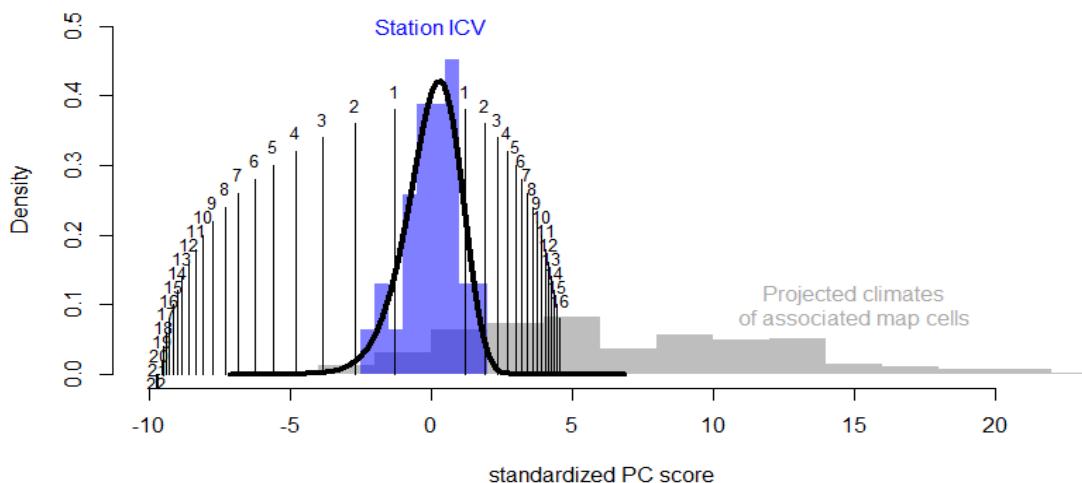


Figure S15: Example of decimal precision (black numbers in plot) required to perform quantile matching to a fitted gamma PDF. In this case, the range of projected climates of map cells associated with the reference station are beyond a feasible decimal precision (>16 decimal places). This problem precludes quantile matching in this novelty analysis, though it would not be a problem for an analysis in which reference variability was calculated separately for each grid cell.

Given that normalization of the reference variability is not feasible for this analysis, it is worthwhile to investigate the impact of the assumed distribution on the results. A sensitivity analysis using raw precipitation data (i.e. not log-transformed as in the base case) provides a bookend for the assumed distribution of precipitation data (Figure S16). With the exception of the Gulf Coast of the US, the general effect is to increase novelty. The results are not dramatic, however, and suggest that the assumed distribution is not of overriding importance.

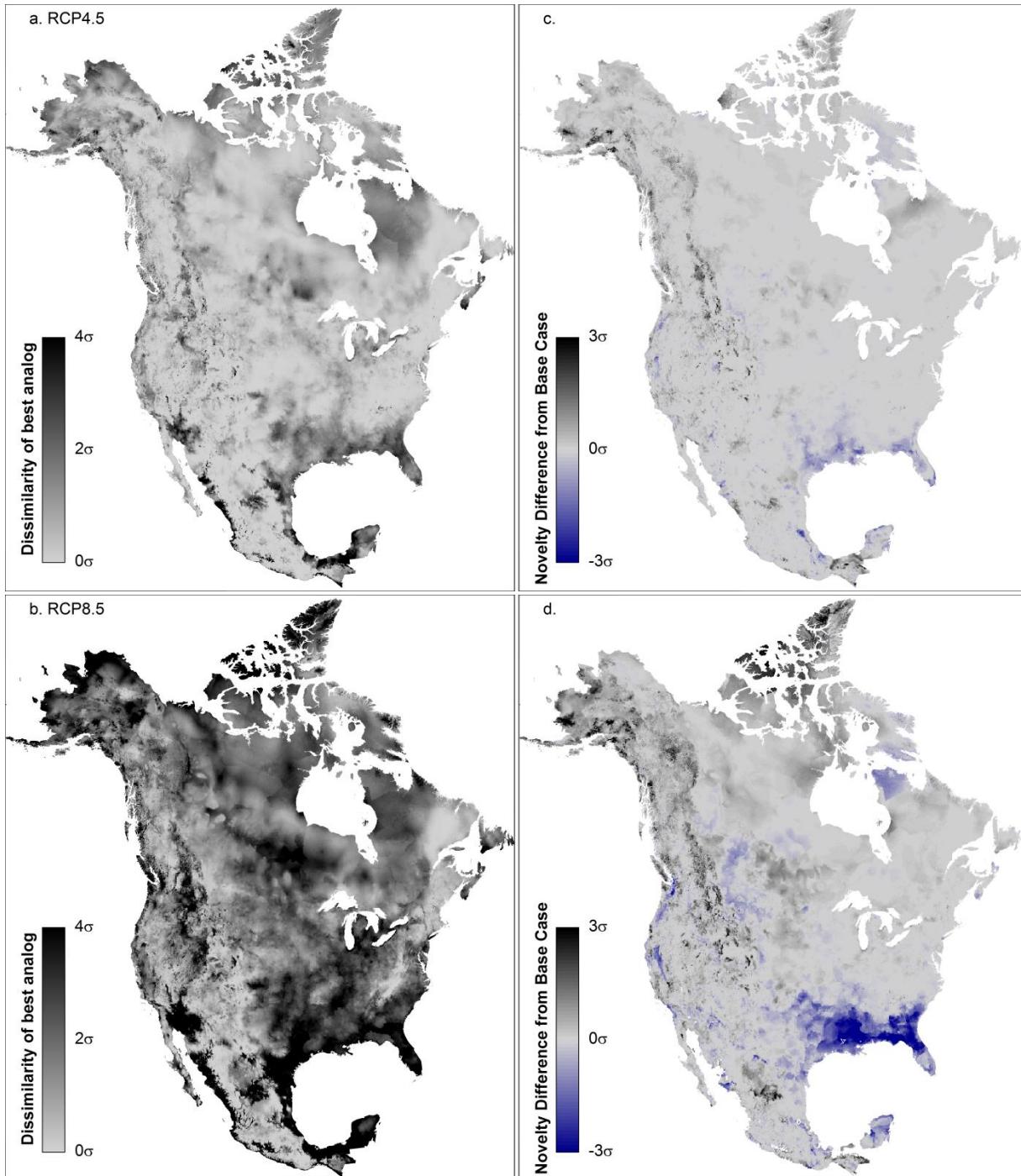


Figure S16: Sensitivity analysis using raw instead of log-transformed precipitation data, for (a) RCP4.5 and (c) RCP8.5. (b) and (d) are comparisons with the Base Case (log-transformed precipitation).

S9. Null model analysis of elevation-novelty relationship

Two factors complicate a statistical analysis of the relationship between topographic position (relative elevation) and novelty: (1) uneven sampling of relative elevation, and (2) high spatial autocorrelation. The problem of uneven sampling can be addressed by statistical analysis against a null model. However, spatial autocorrelation confounds statistical analysis by violating assumptions of independence (Gotelli and Ulrich 2012). We minimized the effect of spatial autocorrelation by conducting null model analysis in very small subsamples ($n < 25$) of the spatial data set ($N=199,059$).

Our method of null model analysis is:

1. Generate null data by randomizing relative elevation (the choice to randomize elevation instead of novelty is arbitrary; either approach produces identical results.)
2. for each sample size j between 5 and 25:
 - a. Randomly subsample the observed and null data to the specified sample size (e.g. $n=j$);
 - b. obtain a bootstrap distribution ($B=n$ replicates) of Spearman rank correlation in the subsampled observed and null data; and
 - c. conduct a paired t test between the bootstrap replicates of observed and null correlation.
3. Repeat step 2 100 times to obtain a boxplot of t -test p -values at each sample size j .

Estimates of spearman rank correlation converge on $\rho=-0.49$ and $\rho=0$ in the observed and null data, respectively, as sample size increases. The bootstrap uncertainty in estimates of ρ , and consequently the t test p -values, declines with increasing n .

Using this method, the observed relationship between relative elevation and novelty is statistically distinct ($p < \alpha = 0.05$) at $n > 15$ (Figure S17). Spatial autocorrelation can reasonably be assumed to be nil in such a small (<0.01%) subsample of the observed data. Obtaining a statistically significant result in this small subsample is strong evidence that the observed relationship between relative elevation and novelty is stronger than would be expected by chance.

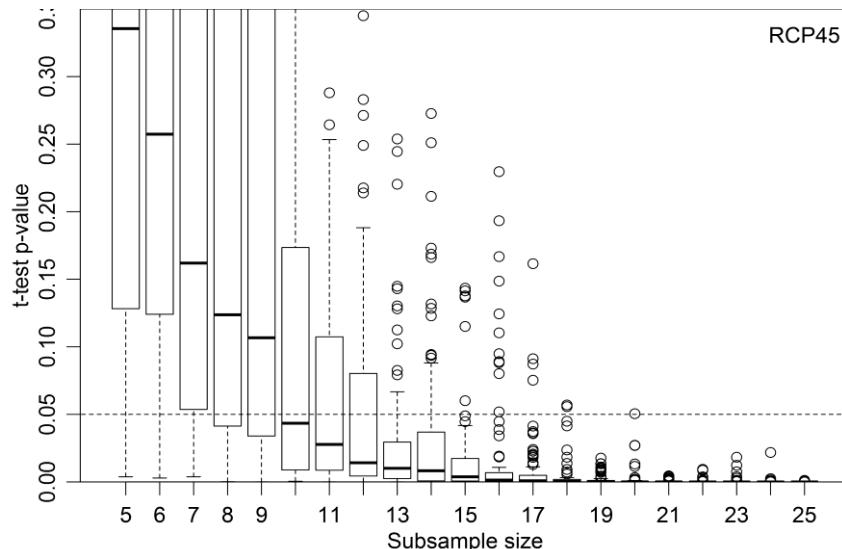


Figure S17: Null model analysis of the relationship between relative elevation and novelty. The plot shows boxplots distribution of p -values from 100 t tests of bootstrap estimates of Spearman rank correlation in the observed and null (randomized) relationship between relative elevation and novelty, at subsamples of the spatial data ($N=199,059$) ranging from $n=5$ to $n=25$. Statistical significance ($p < \alpha = 0.05$) at $n > 15$ provides strong support for the relationship between relative elevation and novelty.

S10. Variation around mean novelty from multiple ICV proxies

We calculate novelty at each grid cell as the distance-weighted average of novelty calculations for the four nearest climate stations used as proxies for the interannual climatic variability (ICV) of the grid cell. The range of variation in novelty calculated from the four ICV proxies is substantial in many grid cells, particularly in the RCP4.5 projection (Figure S18). This variation suggests some potential for local bias associated with averaging of ICV proxies. However, there are few locations with the potential for missed novelty or false novelty relative to the 2σ threshold (Figure S19). The main results of this paper do not appear to be sensitive to differences between the four ICV proxies for each grid cell.

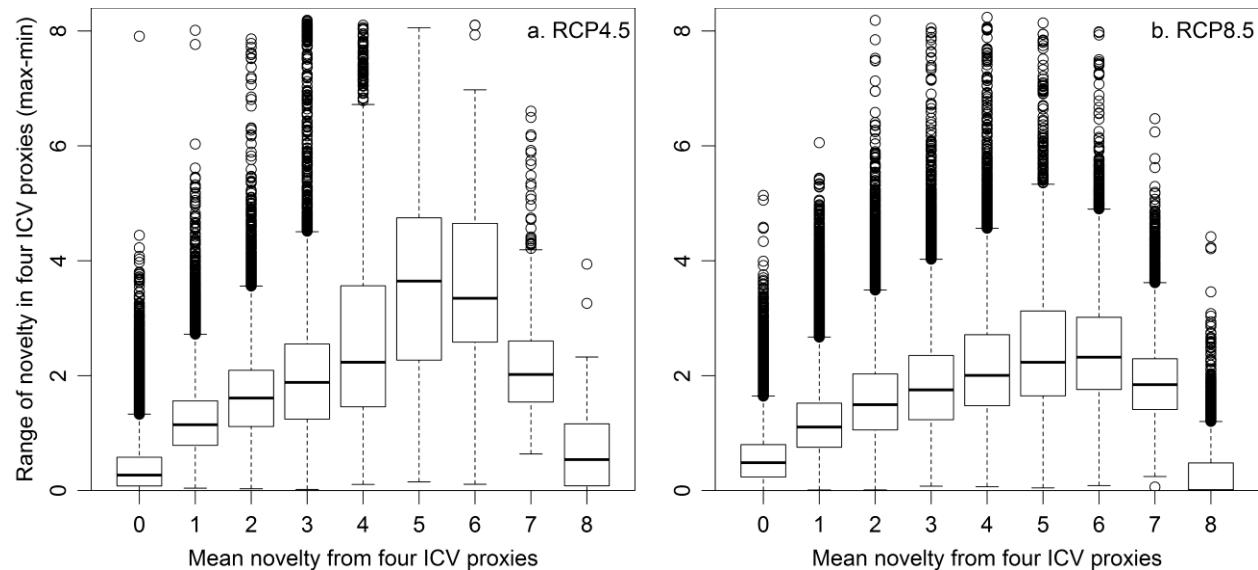


Figure S18: Range of novelty calculated from the four ICV proxies for each grid cell, relative to their mean, for RCP 4.5 (a) and RCP8.5 (b). Range decreases as mean approaches 8σ because the maximum detectable novelty is $\sim 8.2\sigma$.

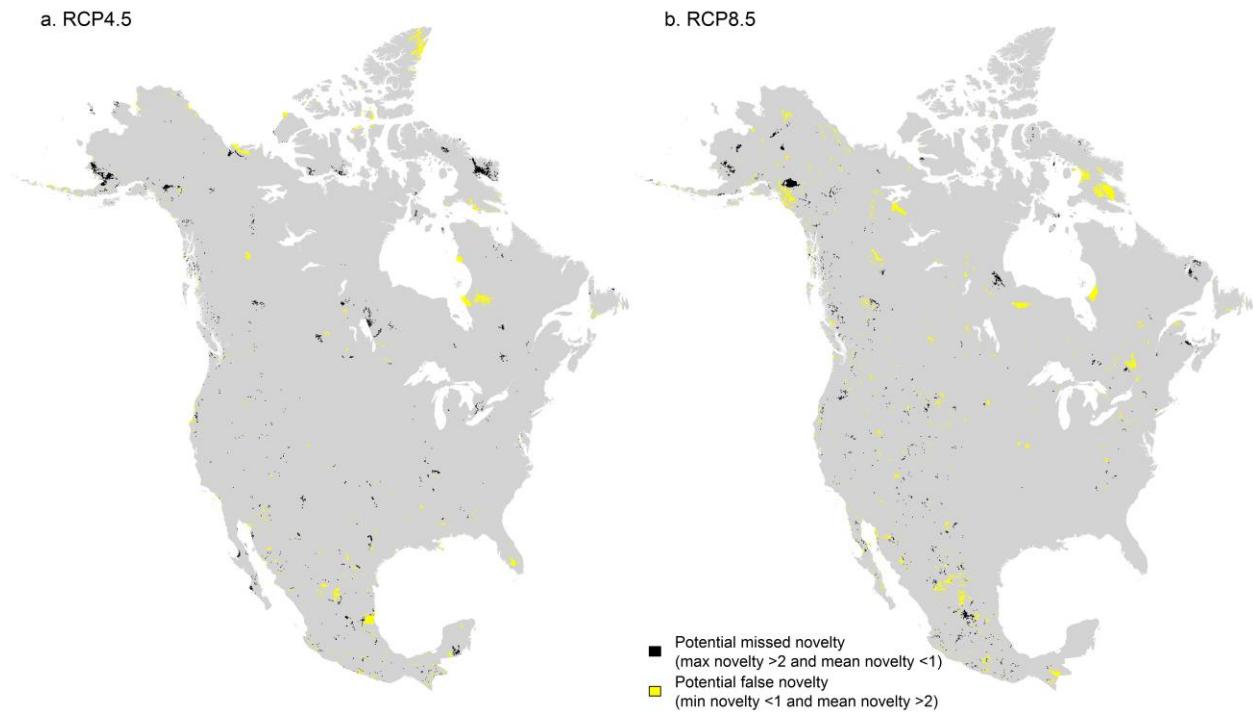


Figure S19: potential for missed novelty or false novelty relative to the 2σ threshold, for RCP 4.5 (a) and RCP8.5 (b).

S11. Bias due to nonrandom weather station placement

The potential for cross-contamination between the distinct ICV patterns of maritime and continental climates is a potential artefact of the use of weather stations as ICV proxies. Mexico, the contiguous US, and coastal British Columbia have sufficient station density in the coast-interior transition that conflation of distinct regional climates can reasonably be ruled out as a source of error. In contrast, low station density in the boreal and Arctic regions suggests the potential for conflation of maritime and continental ICV proxies over much of Canada and Alaska. Widespread systematic bias due to this conflation is not apparent when grid cells are given only a single station as an ICV proxy (Figure S20). Systematic bias would be indicated if coastal and interior areas had opposite colouring. Nevertheless, this bias cannot be definitively assessed without an independent sample of weather stations. The scarcity of weather stations in the boreal and Arctic regions highlights that results for these regions should be interpreted at a broad, regional scale.

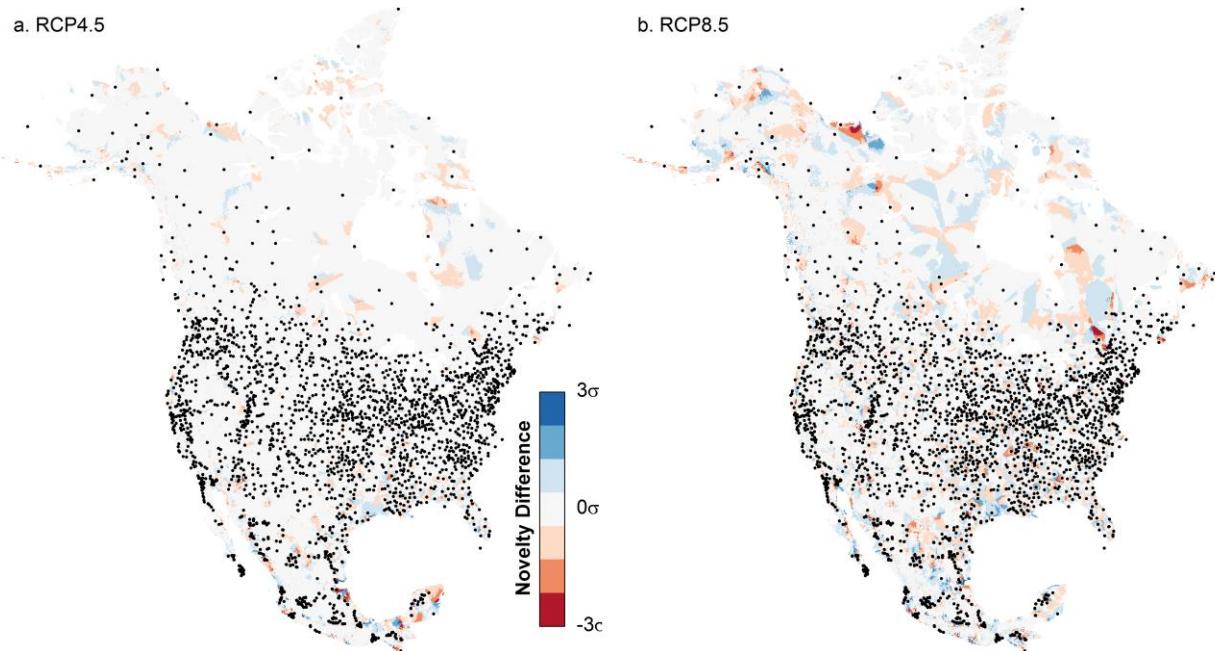


Figure S20: Change in novelty relative to the Base Case when novelty is calculated using the single nearest weather station as an ICV proxy, rather than the four nearest stations. Weather stations used as ICV proxies are displayed as black dots.

S12: Variable selection

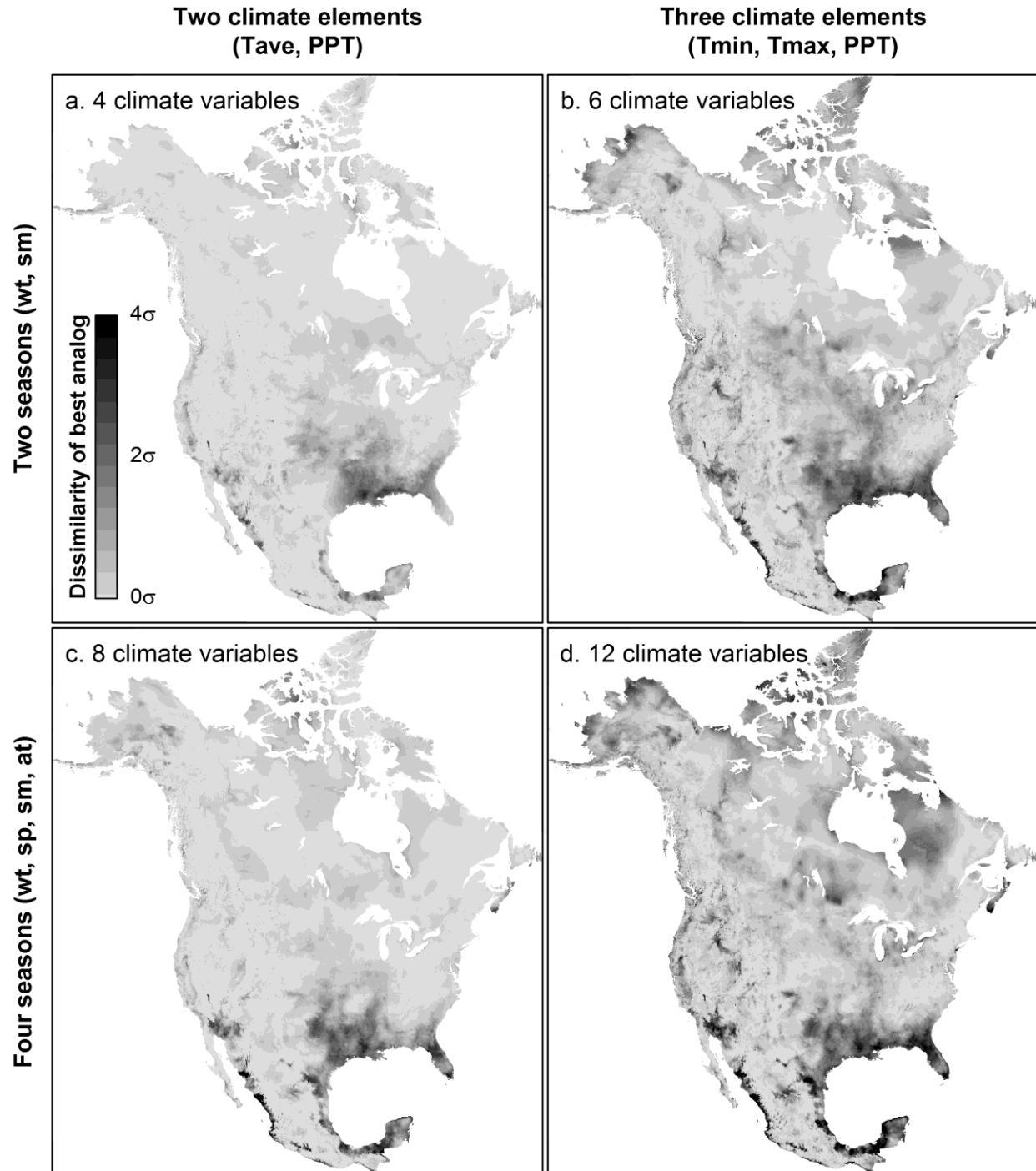


Figure S21: Effect of variable selection on novelty of the RCP4.5 ensemble mean projection. The 12-variable novelty (panel d) is the same as the main results presented in Figure 7 of the publication. Novelty is highly sensitive to the use of seasonal mean daily temperature (Tave) instead of seasonal mean minimum and maximum daily temperature (Tmin, Tmax). Novelty is less sensitive to the use of two seasons instead of four. Results for RCP8.5 are provided in Figure 8 of the publication.

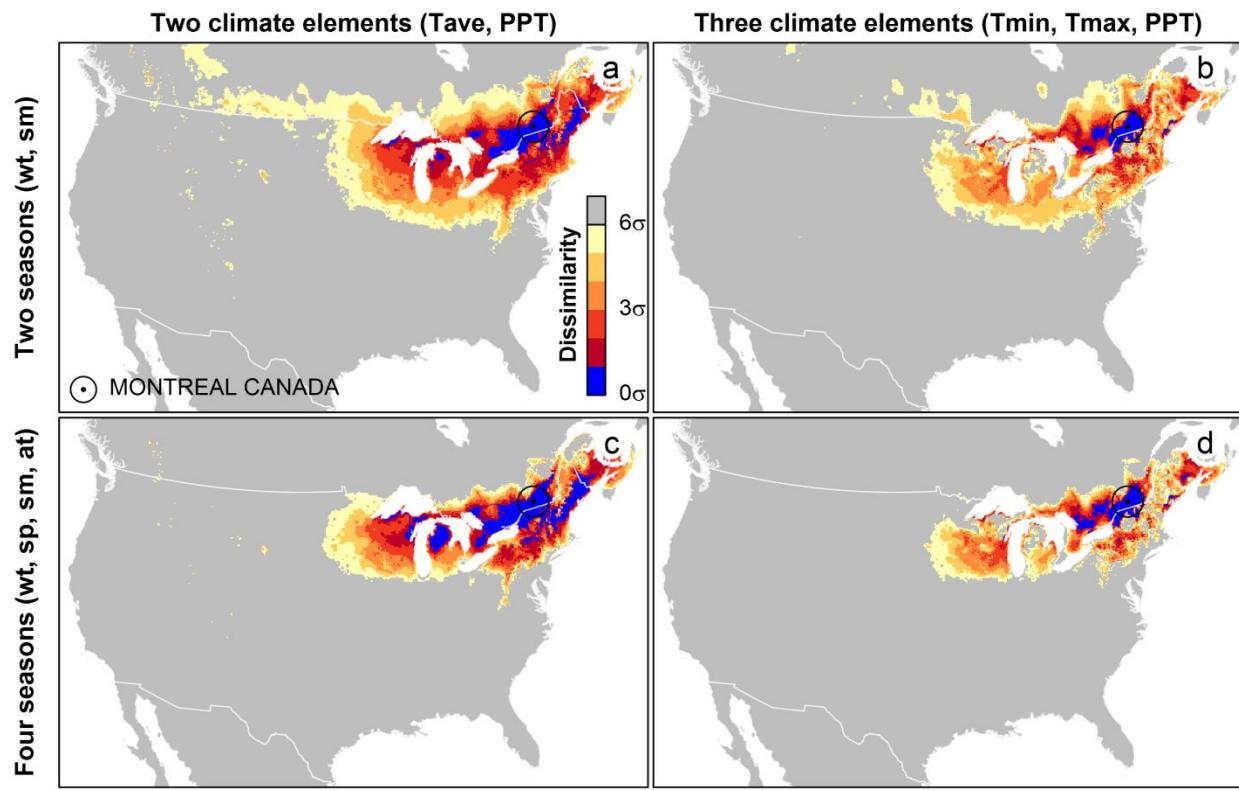


Figure S22: Effect of variable selection on reference period (1971-2000) dissimilarity to Montreal, Quebec.

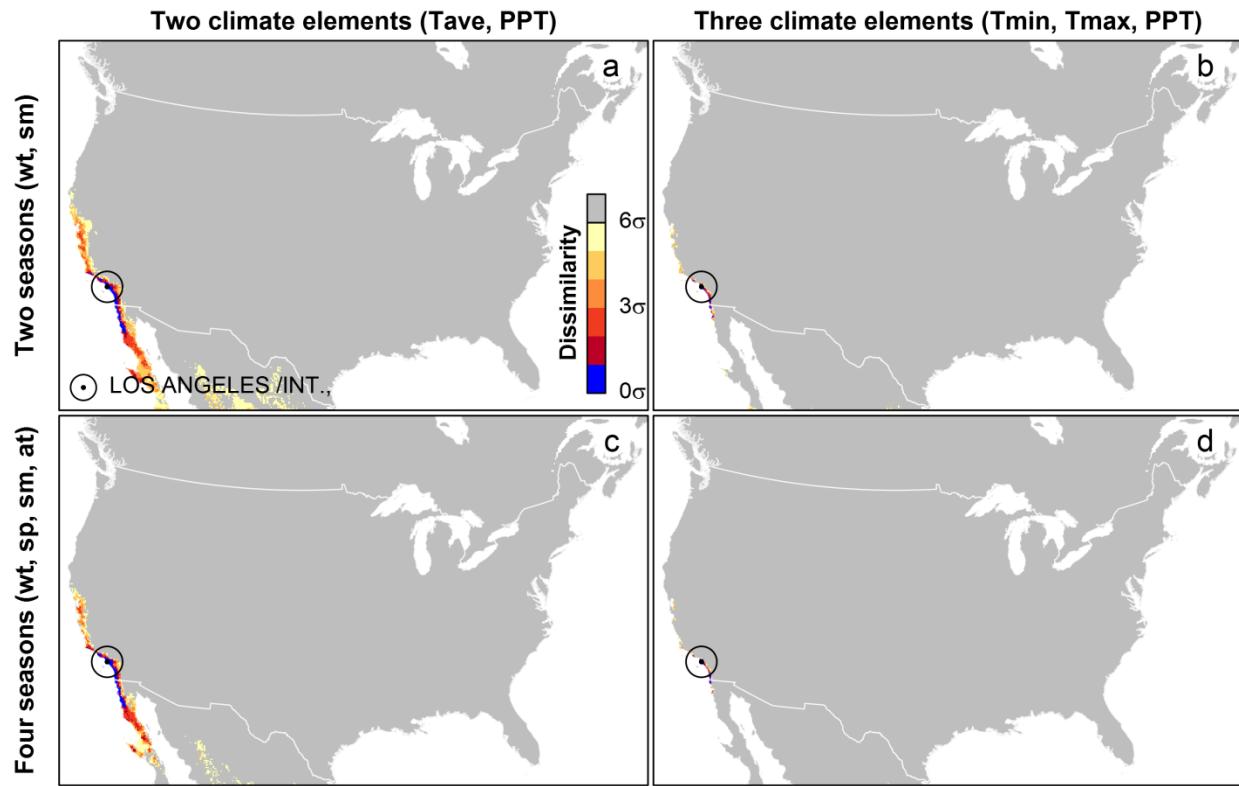


Figure S23: Effect of variable selection on reference period (1971-2000) dissimilarity to Los Angeles, California.

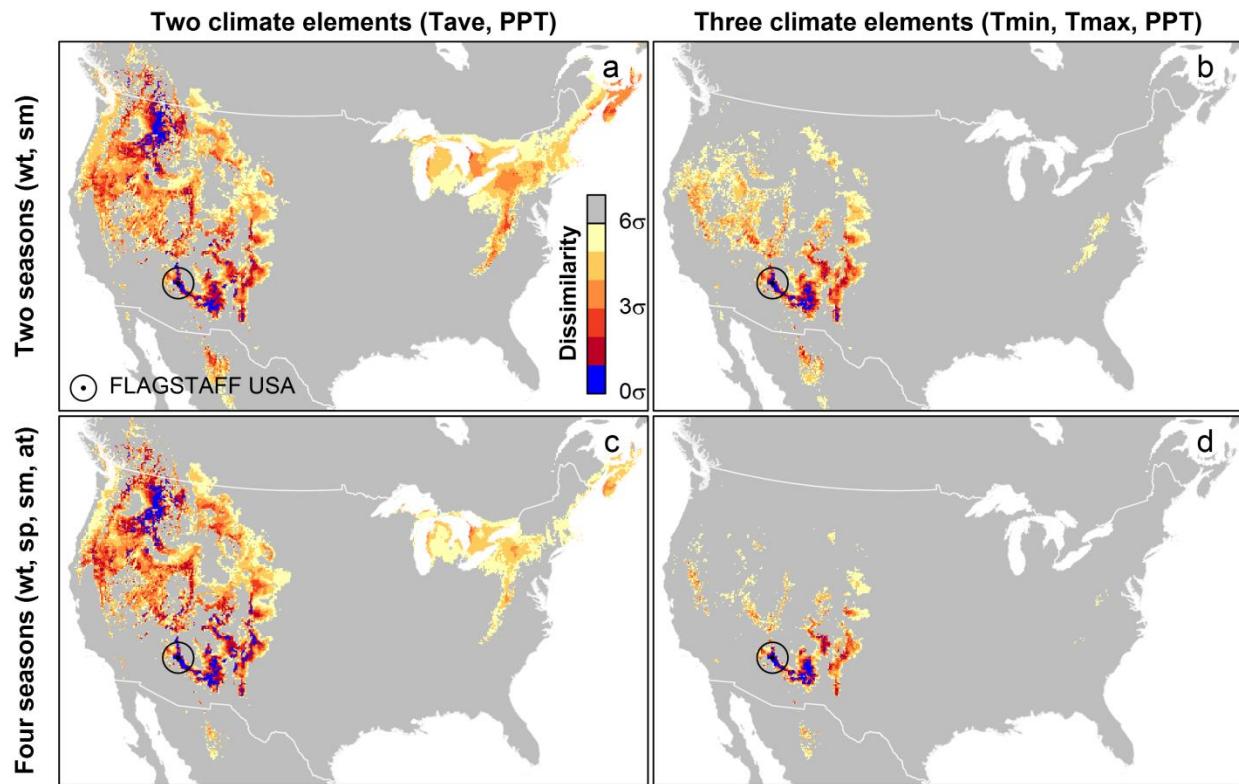


Figure S24: Effect of variable selection on reference period (1971-2000) dissimilarity to Flagstaff, Arizona.

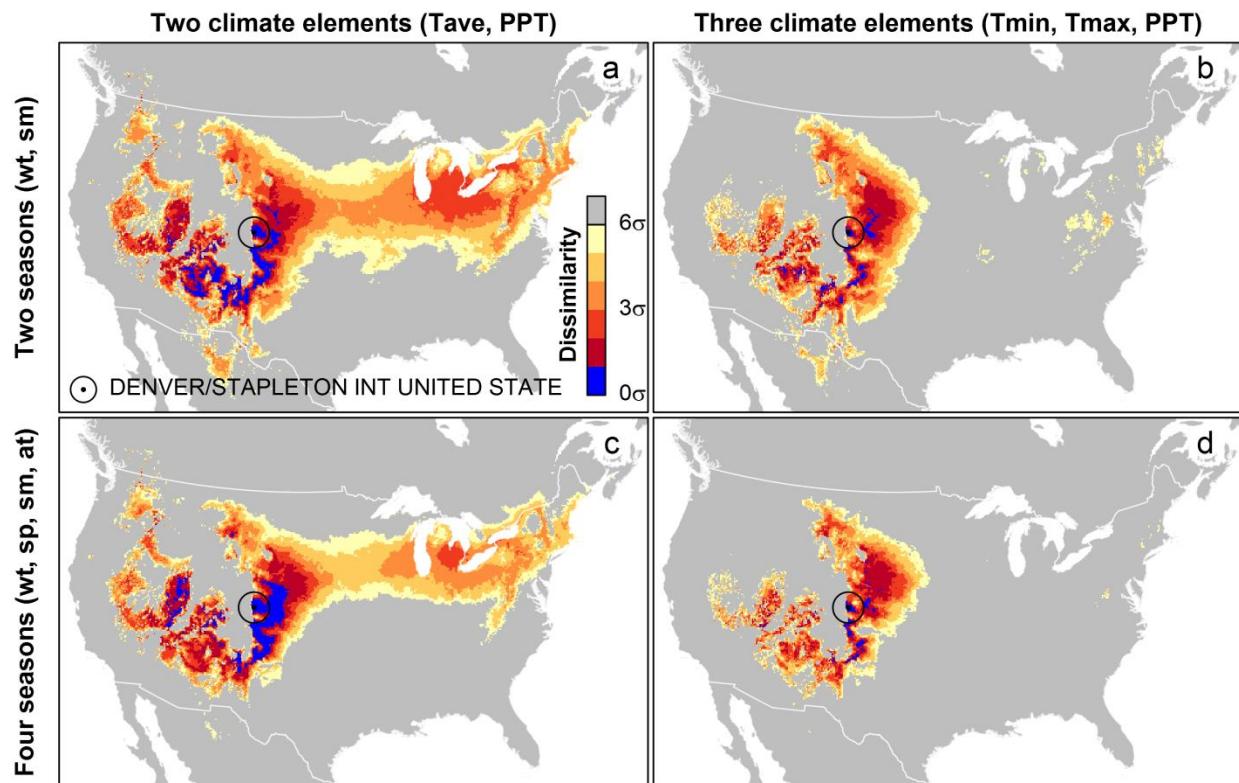


Figure S25: Effect of variable selection on reference period (1971-2000) dissimilarity to Denver, Colorado.

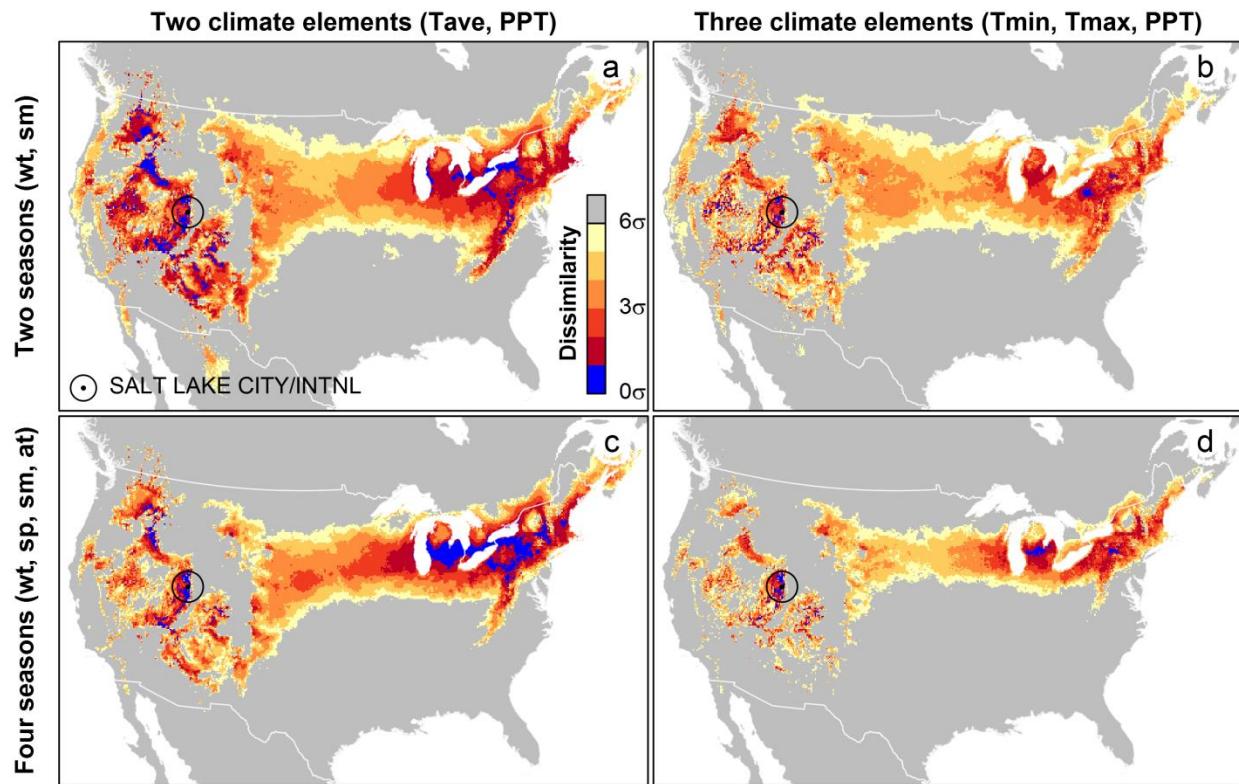


Figure S26: Effect of variable selection on reference period (1971-2000) dissimilarity to Salt Lake City, Utah.

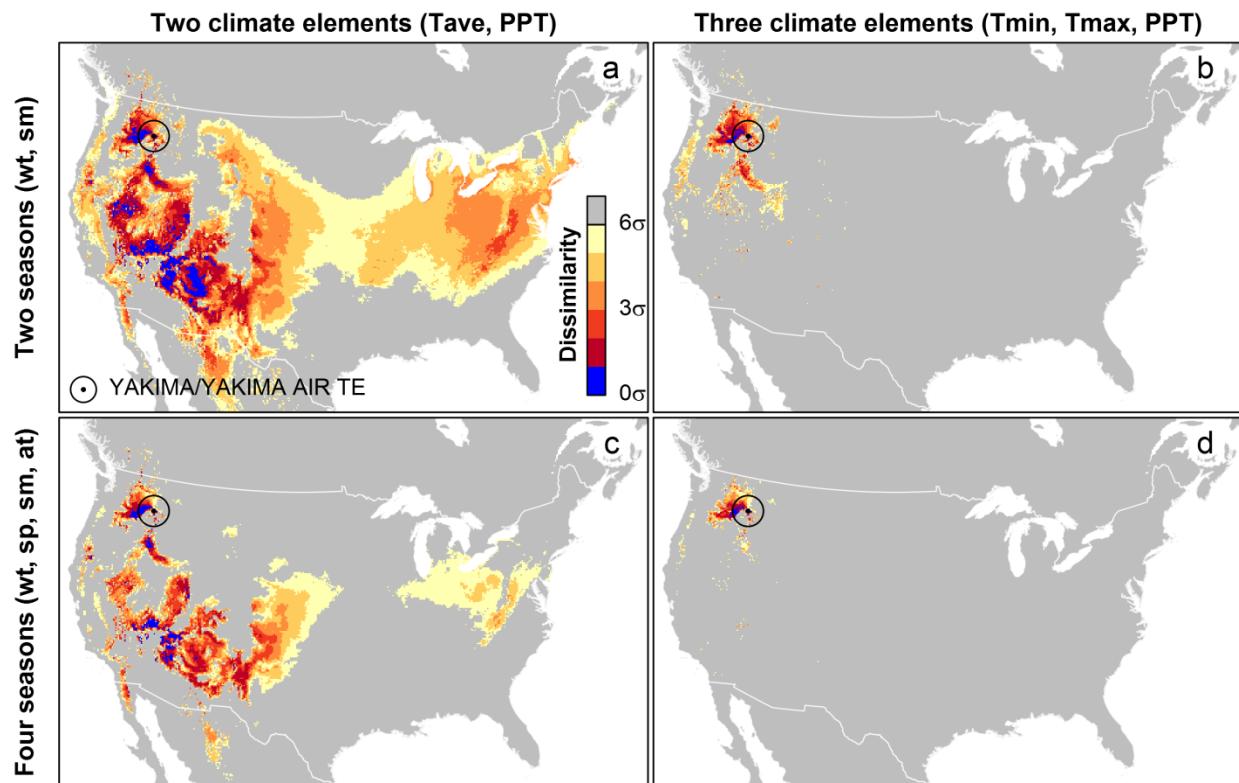


Figure S27: Effect of variable selection on reference period (1971-2000) dissimilarity to Yakima, Washington.

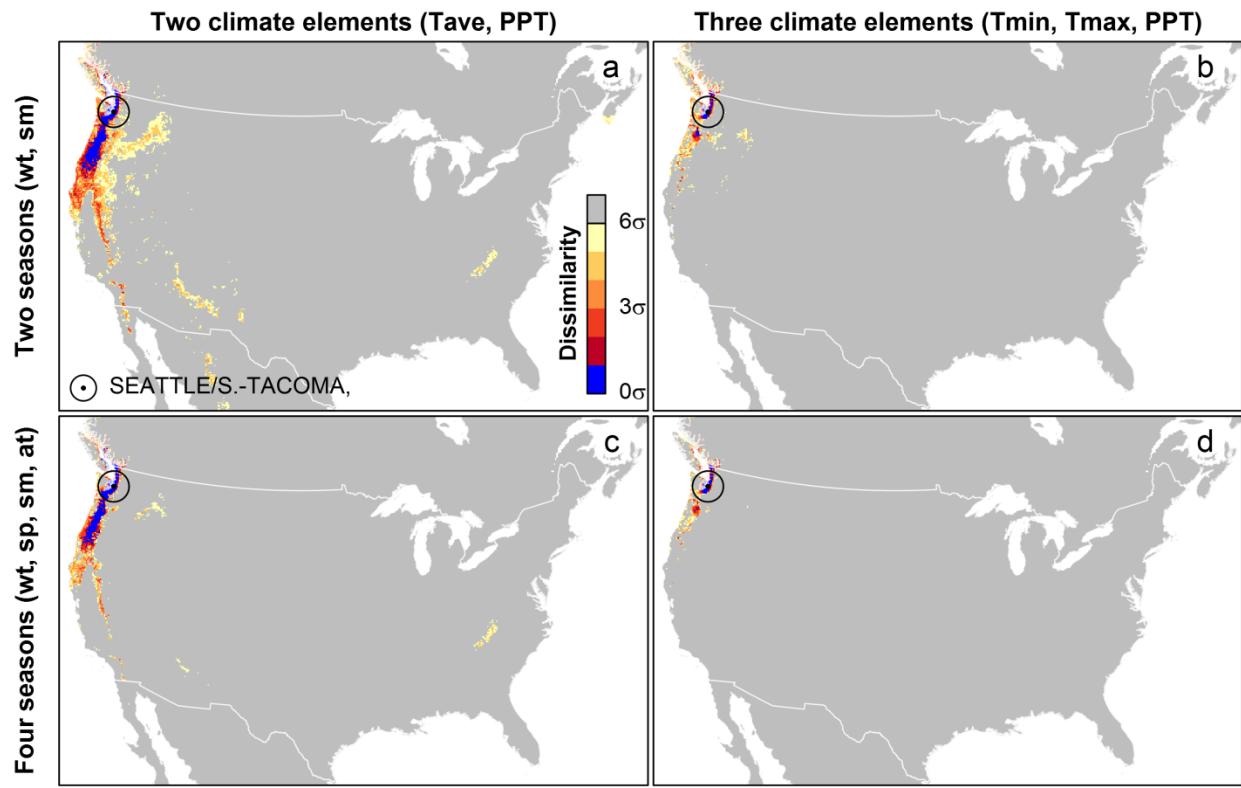


Figure S28: Effect of variable selection on reference period (1971-2000) dissimilarity to Seattle, Washington.

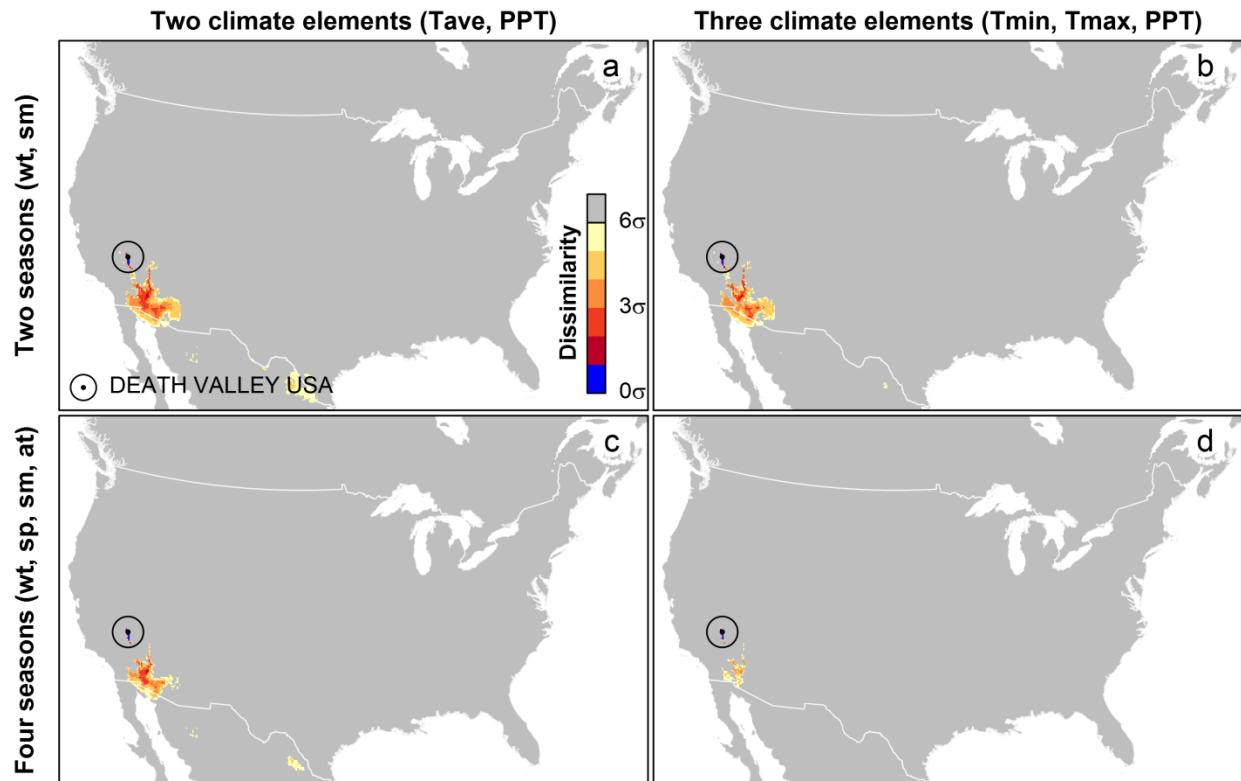


Figure S29: Effect of variable selection on reference period (1971-2000) dissimilarity to Death Valley, California

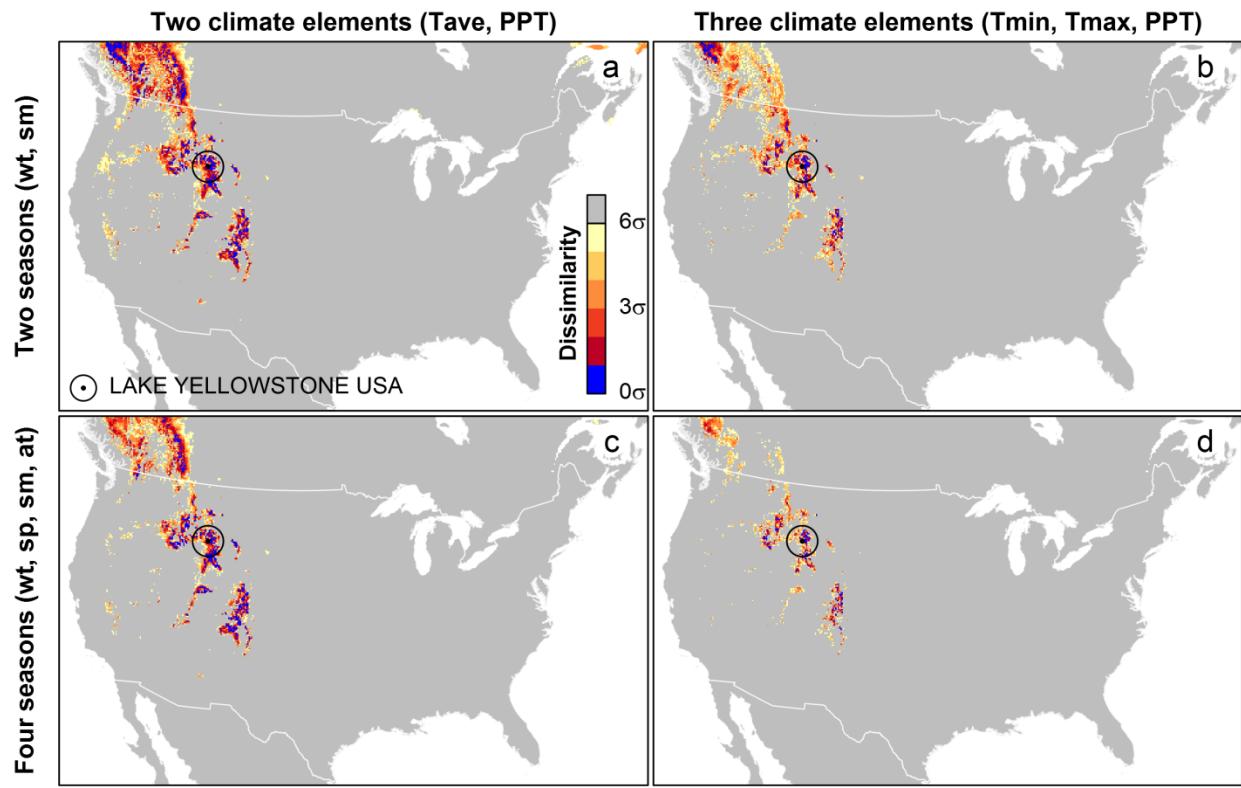


Figure S30: Effect of variable selection on reference period (1971-2000) dissimilarity to Lake Yellowstone, Wyoming.

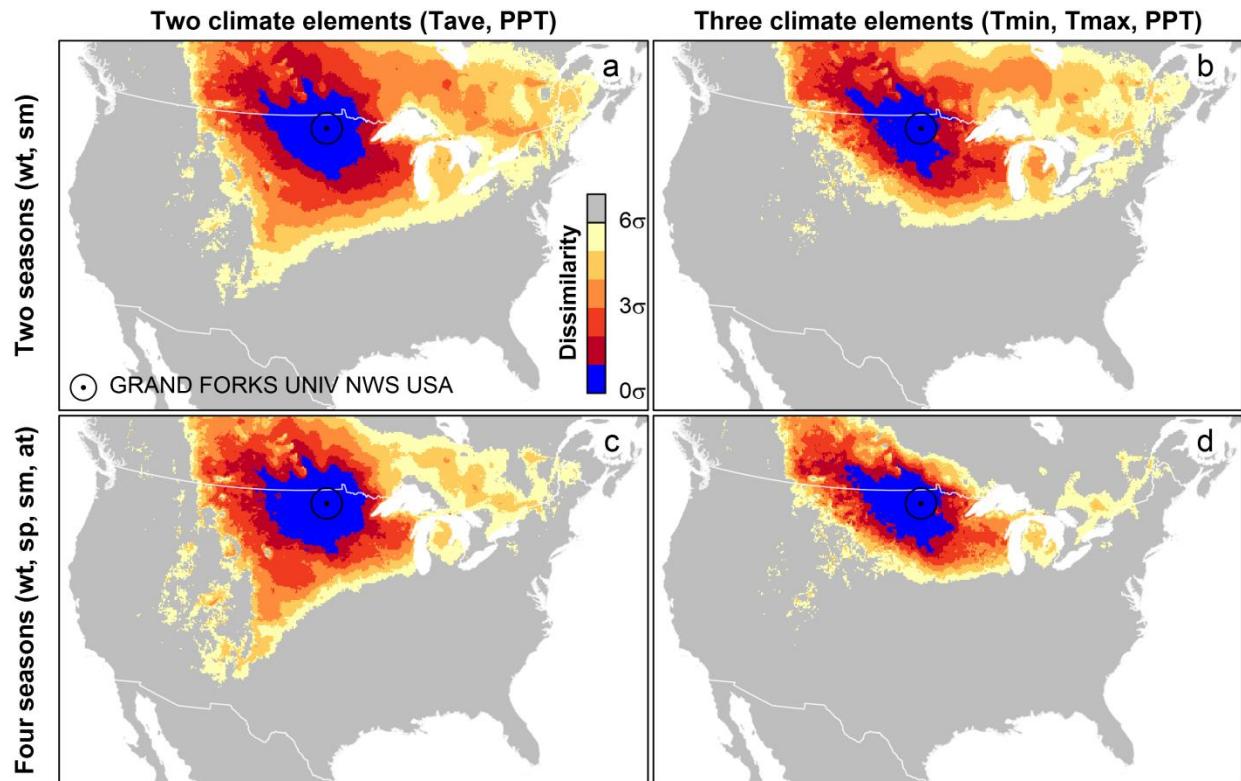


Figure S31: Effect of variable selection on reference period (1971-2000) dissimilarity to Grand Forks, North Dakota.

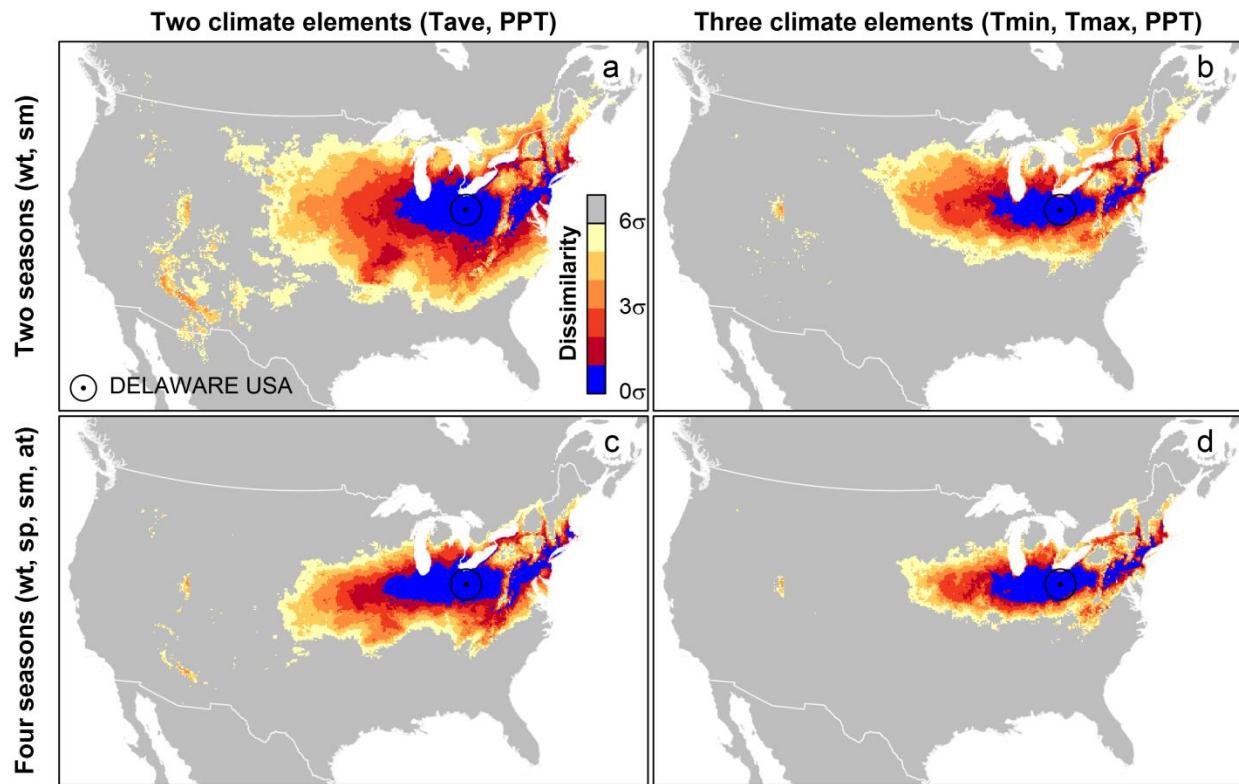


Figure S32: Effect of variable selection on reference period (1971-2000) dissimilarity to Delaware, Ohio.

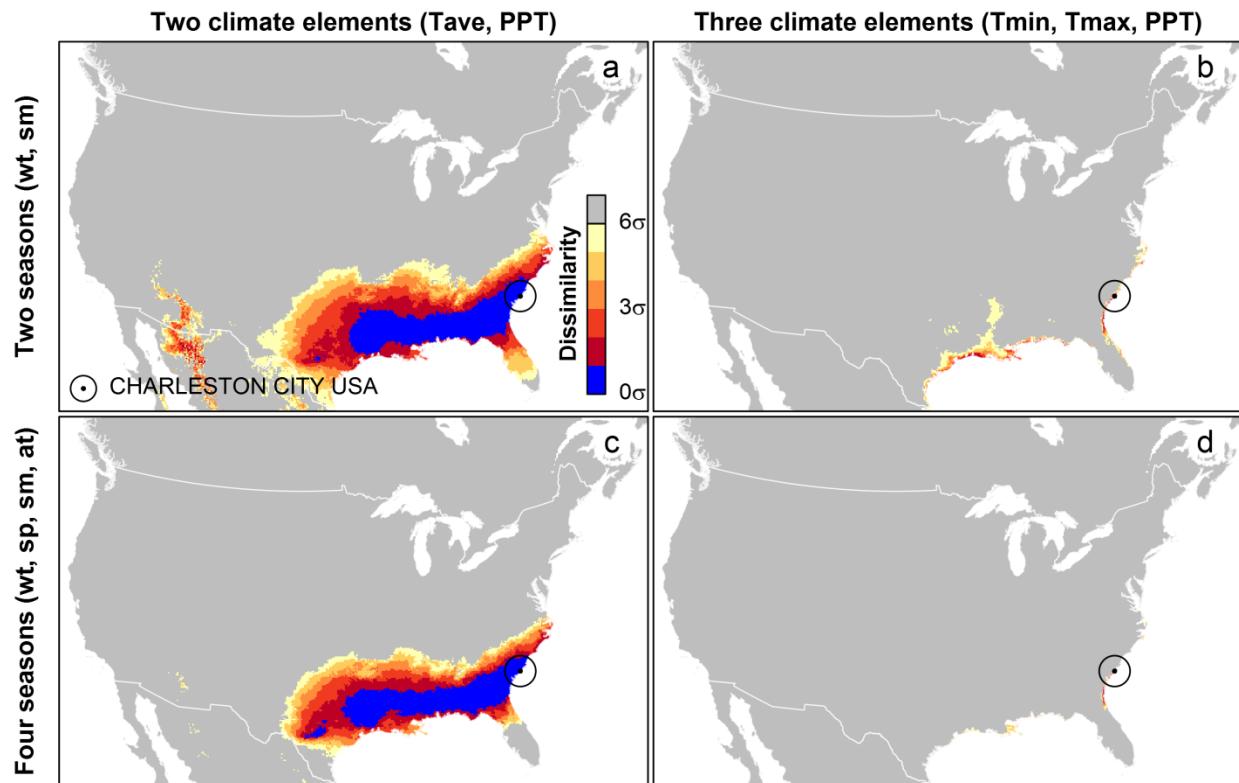


Figure S33: Effect of variable selection on reference period (1971-2000) dissimilarity to Charleston City, South Carolina.

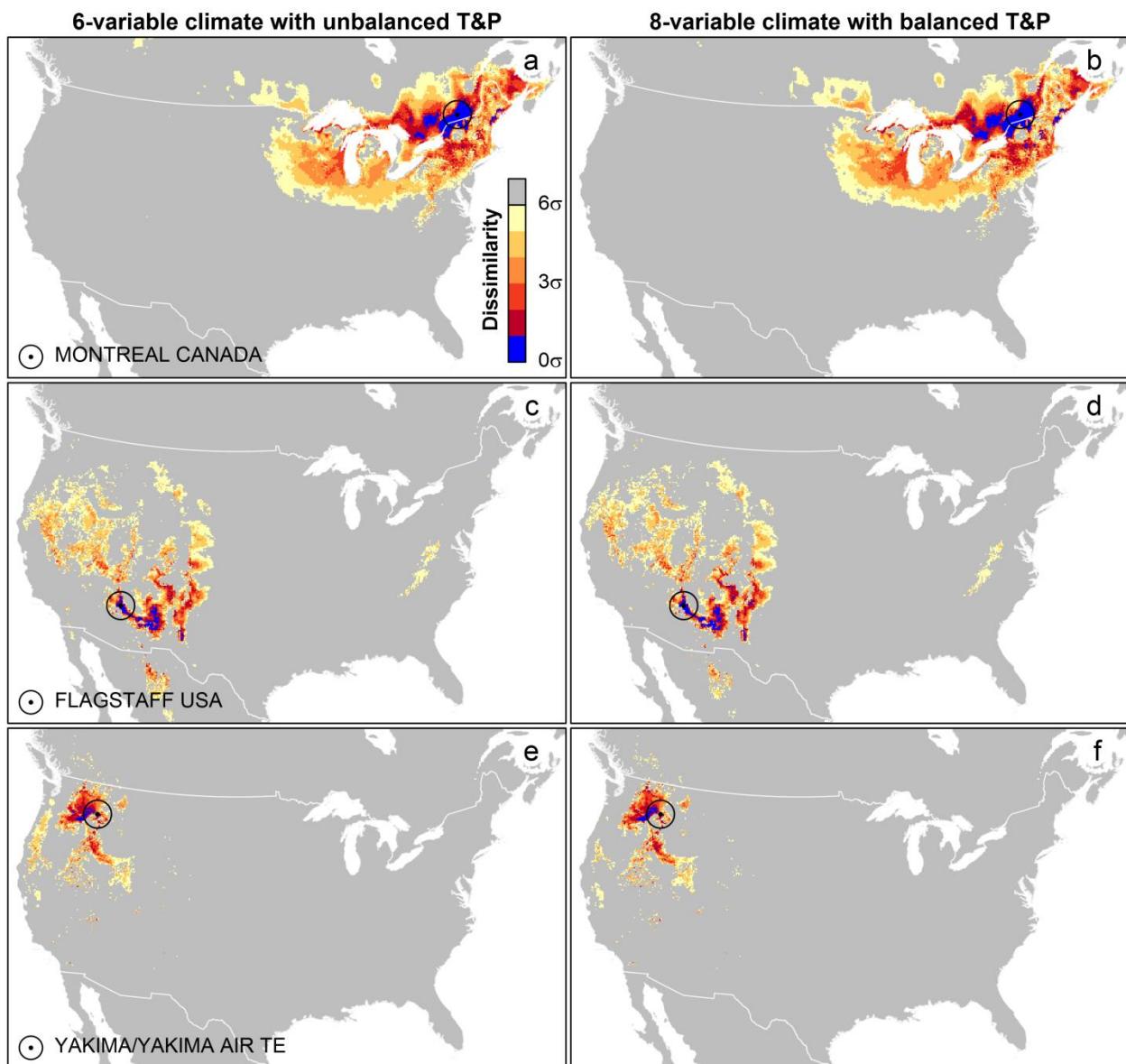


Figure S34: Effect of balancing the number of temperature (T) and precipitation (P) variables on reference period (1971–2000) dissimilarity to Montreal (a,b), Flagstaff (c,d), and Yakima (e,f). The 6-variable climate (a,c,e) is composed of Tmin, Tmax, and PPT for winter (DJF) and summer (JJA). Precipitation in spring (MAM) and autumn (SON) are added to these variables to define the 8-variable climate (b,d,f). Balancing the number of T&P variables has minimal effect on climatic dissimilarity.

S13: Comparison to standardized Euclidean distance.

Our study builds on Williams et al. (2007) by making several methodological modifications: (1) high spatial resolution; (2) observational data for baseline climatology; (3) a 12-variable description of climate, instead of 4; and (4) a Mahalanobian sigma dissimilarity metric instead of SED. In this section, we perform a stepwise comparison of our results to those of Williams et al., focusing on sensitivity to the 12-variable climate and the sigma dissimilarity metric.

In the first step of this factorialized comparison, we performed an SED novelty calculation on our dataset using four climate variables similar to those used by Williams et al. (2007), and attempted to match the colour scheme to Williams et al. (2007) (Figure S35). Despite several differences in data (Table S3), our benchmark results are broadly similar in pattern to those of Williams et al. (2007). The higher novelty present in our benchmark in the Gulf states of the USA and the west coast of Mexico can likely be attributed to the continental rather than global analog search. Lower novelty throughout the western cordillera and arctic can likely be attributed to a higher-resolution representation of terrain.

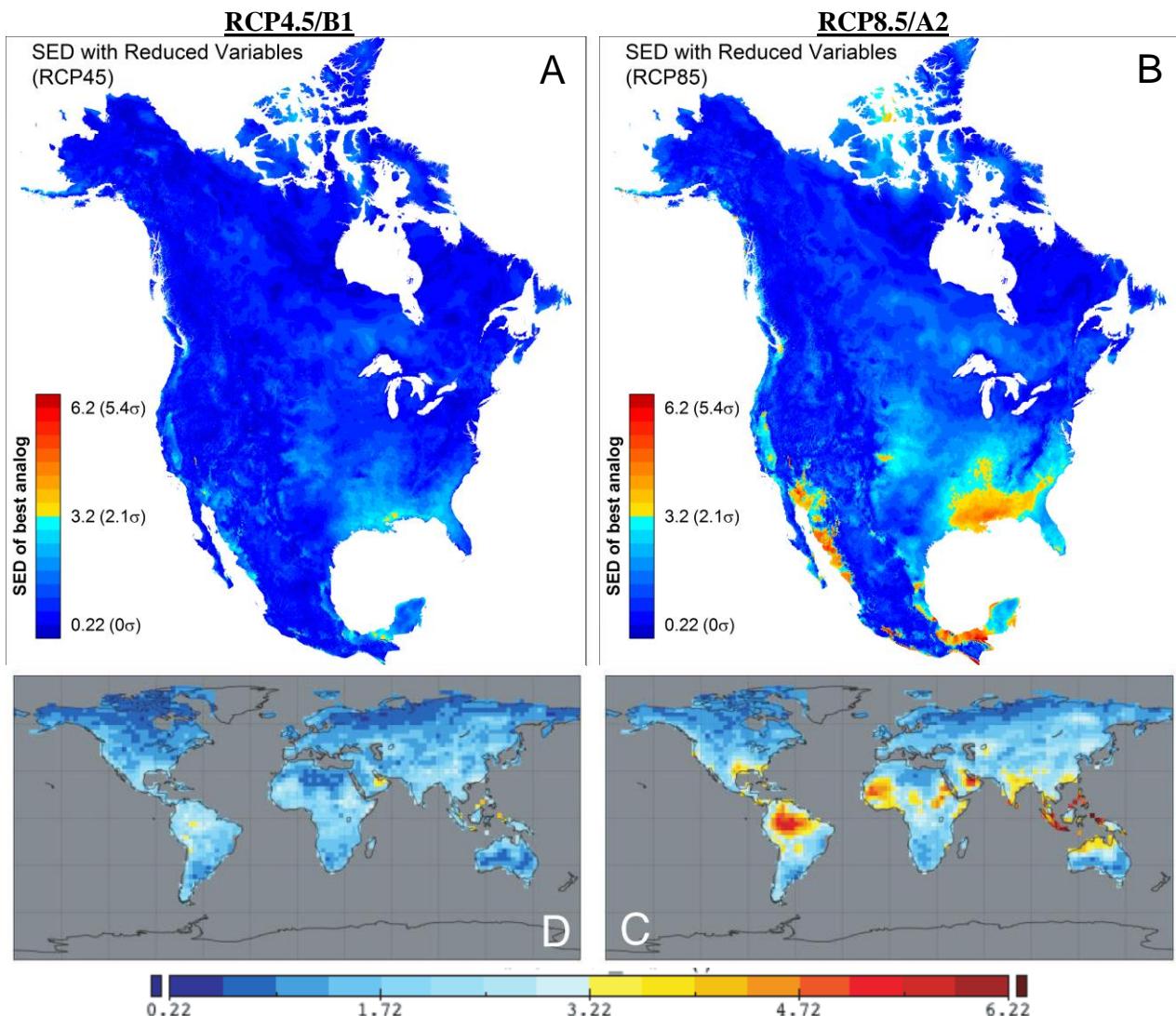


Figure S35: Approximate benchmarking of our dataset to the Williams et al. (2007) results. (A) and (B) are SED novelty based on the four climate variables used by Williams et al. (2007) for RCP4.5 and RCP8.5, respectively. (C) and (D) are excerpted from Williams et al. (2007) for the SRES B1 and A2 scenarios, respectively.

Table S3: Differences between the data used in Williams et al. (2007) and the benchmark scenario presented in Figure S35, including probable effect on novelty.

Williams et al. (2007)	Benchmark scenario	Effect
Global analog search	Continental analog search	+
1980-1999 GCM variability	Detrended 1951-2000 station variability	-
CMIP3 Ensemble (A2/B1)	CMIP5 mean projection (RCP8.5/4.5)	?
250km grid (i.e. poor sampling of analog pool)	8km grid (i.e. good sampling of analog pool)	-

Converting SED distances to approximate sigma dissimilarity allows comparison between SED novelty (Figure S36) and Mahalanobian novelty (Figure S21). Sigma dissimilarity based on SED is an approximation because it does not account for correlations between variables, and therefore is biased towards underestimation of statistical dissimilarity. For this reason, the observed reduction in novelty detected by SED (Figure S36) relative to Mahalanobis distance (Figure S21) is expected. Note, however, that there are only subtle differences between SED and Mahalanobis distance in the 4-variable climate used by Williams *et al.* (2012). The underlying differences in spatial climatic dissimilarity associated with SED and Mahalanobis distance are illustrated in Figure S37.

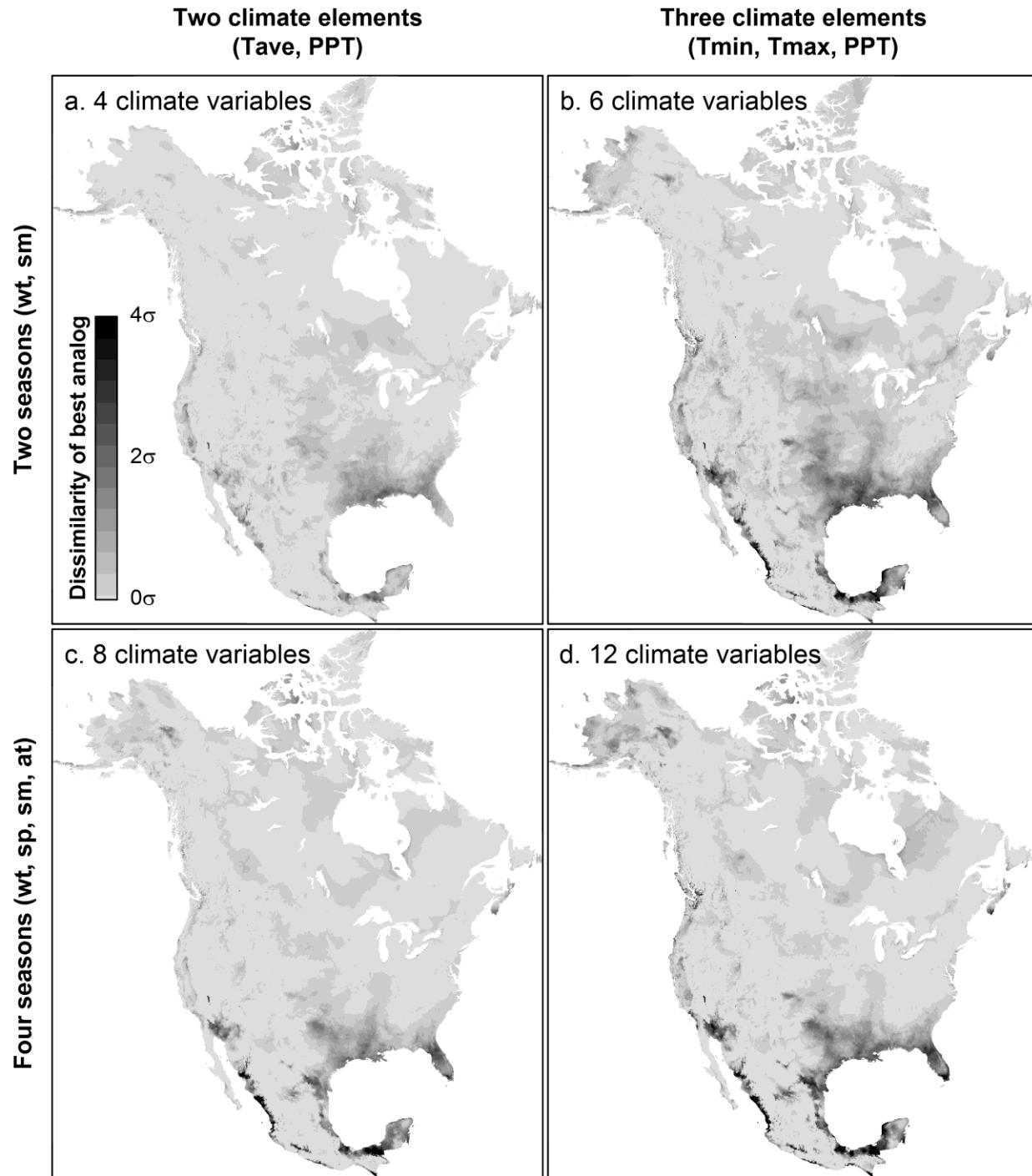


Figure S36: Effect of variable selection on SED novelty in the RCP4.5 ensemble mean projection. Distances have been converted to sigma levels to account for dimensionality effects

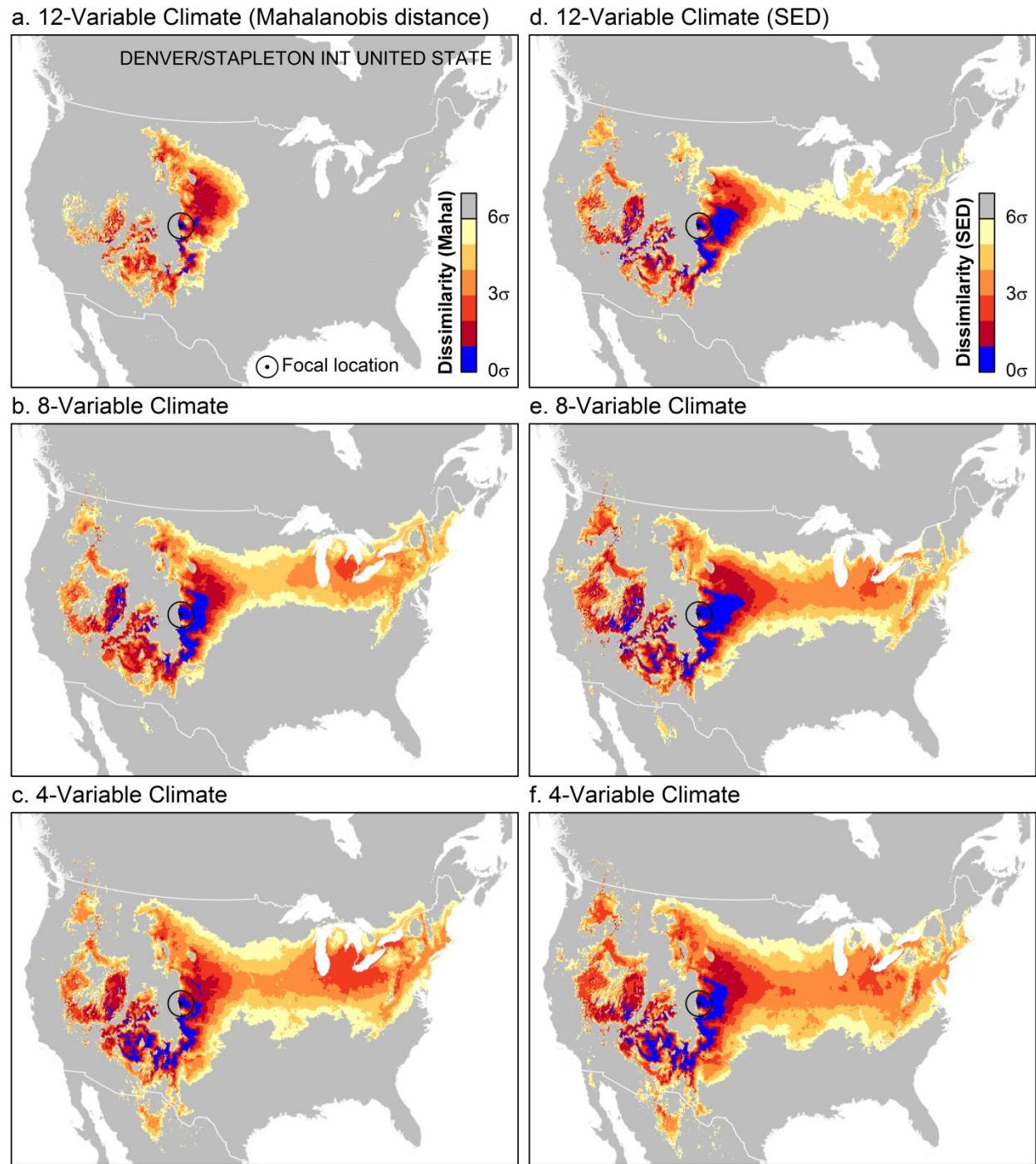


Figure S37: reference period (1971-2000) sigma dissimilarity to Denver, CO, using Mahalanobis distance (a,b,c) and SED (d,e,f). Results are shown for three different variable sets: 12-variables (a,d; Tmin/Tmax/PPP for 4 seasons), 8 variables (b,e; Tave and PPT for 4 seasons), and 4 variables (c,f; Tave and PPT for winter and summer). SED generally reduces spatial climatic dissimilarity relative to Mahalanobis distance, particularly in the 12-variable climate.

S14: Assessment of error due to ICV sample size

Some of the ICV proxies—weather station records used as proxies for local interannual climatic variability (ICV) in each map cell—contain missing data that reduce the sample size contributing to sigma dissimilarity. In general, most map cells have few missing years: 50% of cells outside of Mexico are represented by ICV proxies with ≥ 37 complete years; 75% have ≥ 32 complete years; and 95% of have ≥ 24 complete years (Figure S38). Despite the completeness of the 1951-1990 record for most of the study area, the sufficiency of these samples for calculation of sigma dissimilarity deserves consideration.

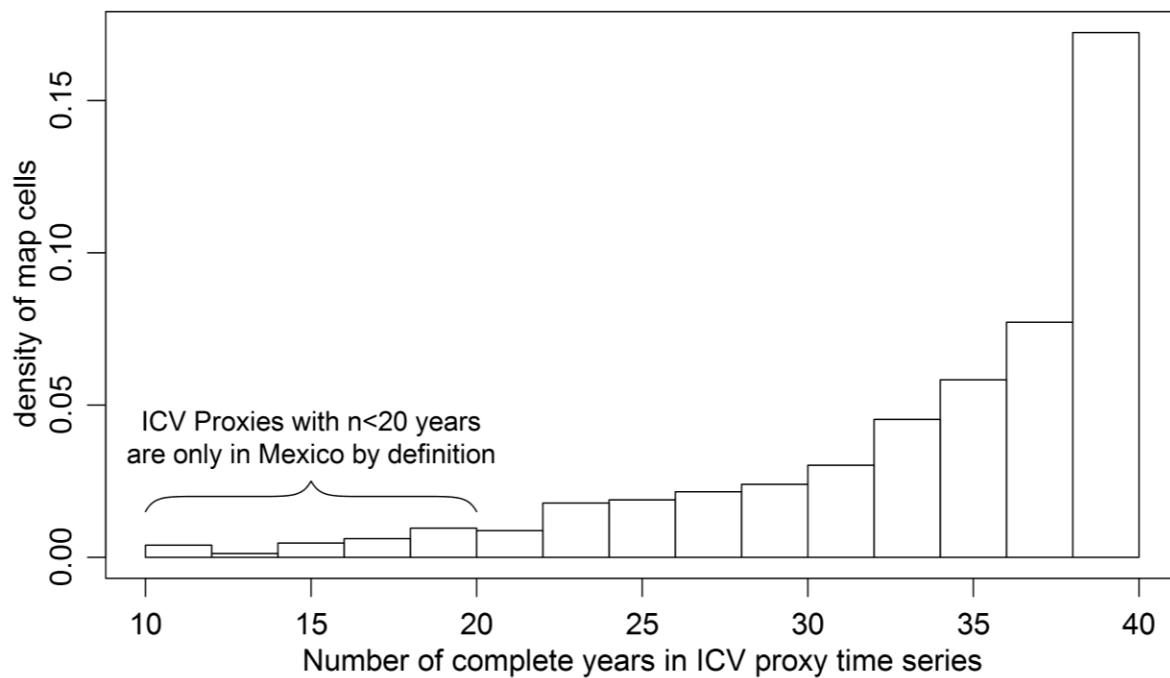


Figure S38: ICV proxy sample size for map cells of the North American study area.

We have determined that sample size has a negligible effect on the z-standardization component (steps 1 and 3) of the calculation of sigma dissimilarity (bootstrap analysis, not shown). However, the effect of reduced sample size is an important consideration for the principal components analysis (PCA; step 2) because even the maximum sample size of 40 years is small for a 12-dimensional PCA. Despite the commonly-cited rule of thumb that sample size should be greater than five times the dimensionality of the PCA, empirical analyses have demonstrated that the instability associated with low sample size is highly dependent on the covariance structure of the sample (Osborne and Costello 2004), and hence that a sufficient sample size cannot reliably be determined *a priori*.

We conducted a bootstrap analysis to estimate the error associated with the sample size of the ICV proxy time series. We ran 120 iterations of the novelty analysis, each using a bootstrap resample of the complete years contributing to the PCA. To reduce computation times, we conducted the analysis only on map cells with > 10 novelty in the main 12-variable analysis, and only for a single ICV proxy per map cell rather than four. 95% confidence intervals (CIs) for each cell are the 2.5th and 97.5th percentiles of these 120 novelty estimates. The “relative 95% CI” is the CI divided by the median. As expected, error is high at small sample sizes (Figure S39). A large relative error is observed at a sample size of 30 years, suggesting the complicating role of covariance structure in generating instability in the PCA. The median relative confidence interval is approximately one at most sample sizes $n > 24$ years, indicating that bootstrap error

is moderate for most of the map cells with novelty greater than 1σ . Figure S40 shows the spatial distribution of PCA instability as indicated by bootstrap error.

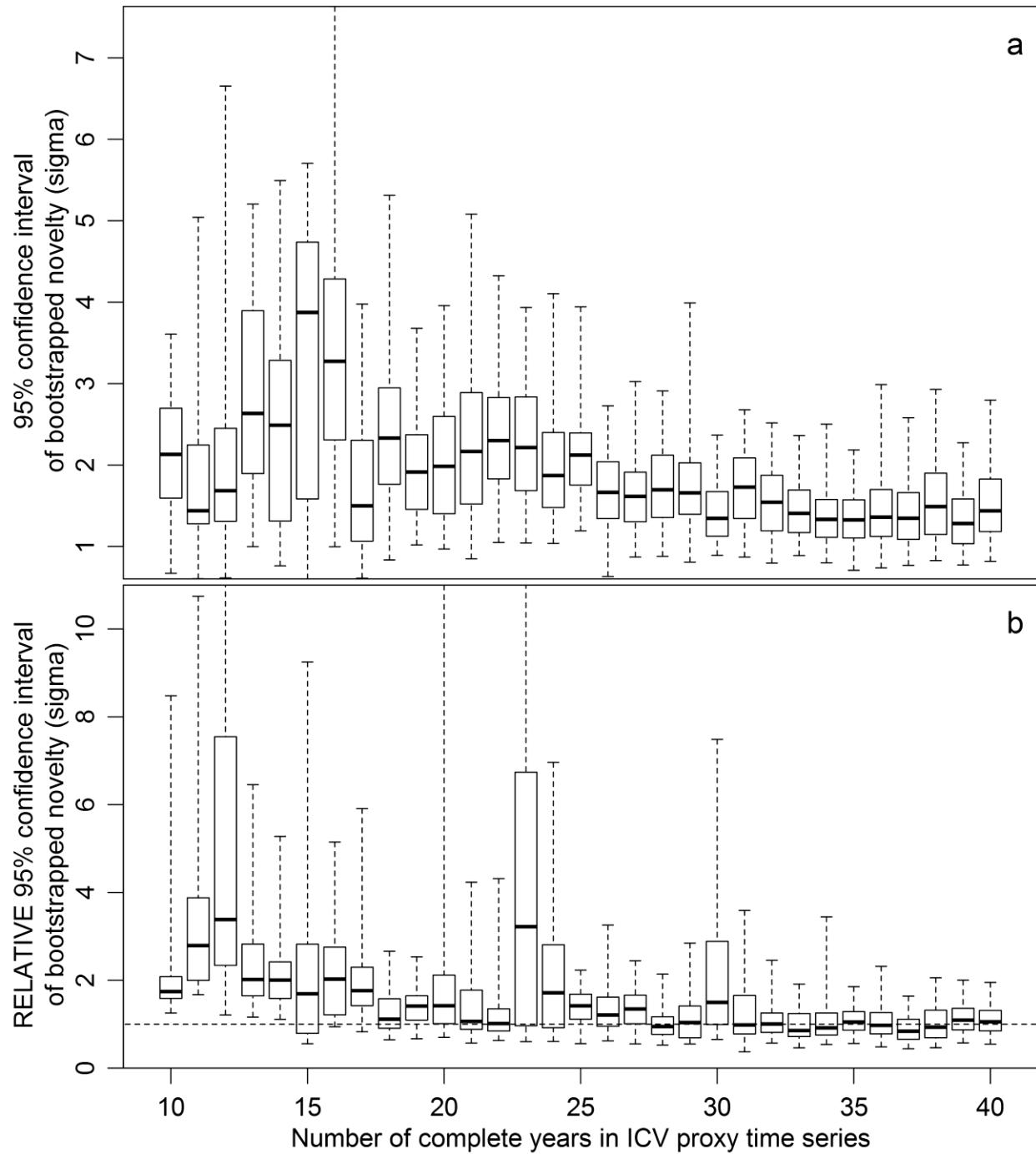


Figure S39: (a) Distribution of the 95% confidence interval of 120 bootstrapped novelty calculations for each map cell with novelty $> 1\sigma$. Distributions are stratified by the number of complete years in the proxy ICV time series representing the map cell. (b) Same as (a) but showing 95% confidence interval as a ratio of the median of the $n=120$ bootstrap distribution.

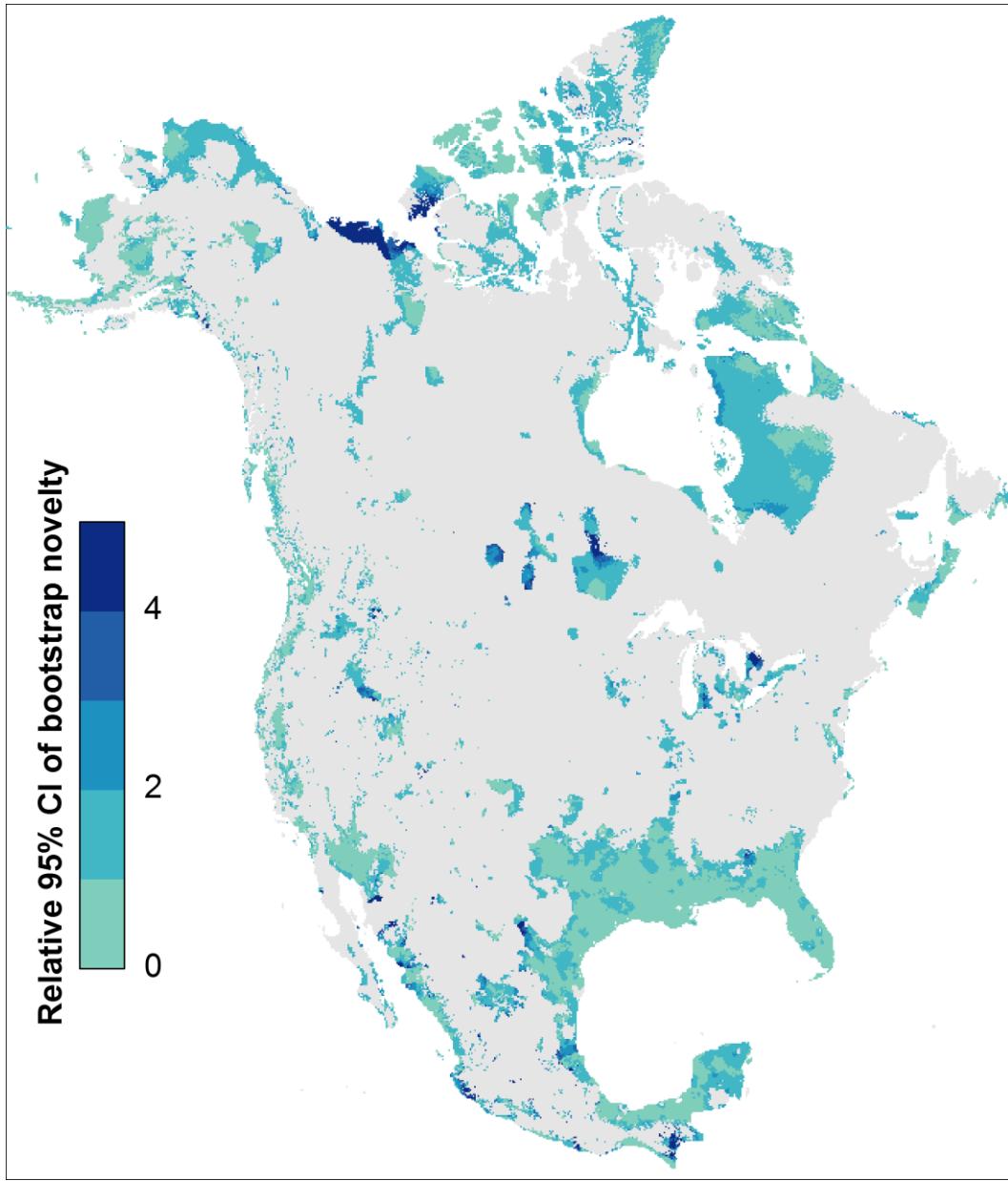


Figure S40: Map of the relative 95% confidence interval (a ratio of the median) of $n=120$ novelty calculations for each cell with bootstrap resampling of the ICV time series contributing to the PCA. Map cells with novelty $<1\sigma$ in the main analysis are excluded and shown as grey. Locations with high relative CI indicate areas where novelty estimates are potentially unreliable due to an unstable PCA.

The bootstrap error estimation above provides an indication of the relative effect of sample size on novelty error and locations where novelty results may be less reliable. The absolute confidence intervals are not reliable, however, because bootstrapping intrinsically reduces the stability of the PCA.

Duplication of observations by bootstrap resampling reduces the effective sample size of the covariance matrix. Hence PCA errors can be expected to be higher for all bootstrap resamples than for the full sample. At the level of an individual map cell (Figure S41), variance of novelty calculated by $n=39$ bootstrap resampling is much greater than $n=39$ subsamples of an $N=40$ ICV time series. In this case, bootstrap resampling at $n=39$ produces the error equivalent to an $n=20$ subsample. For this one map cell, bootstrapping in small samples ($n \leq 20$) produces a substantial underestimation of novelty. This

underestimation can be seen across the full population of map cells with novelty $>1\sigma$ (Figure S42), particularly in that the majority median bootstrap novelty, and substantial proportion of the 97.5th percentile of bootstrap novelty, are located below the 1:1 line. These results indicate that bootstrap resampling produces a biased and exaggerated estimate of uncertainty in the novelty analysis.

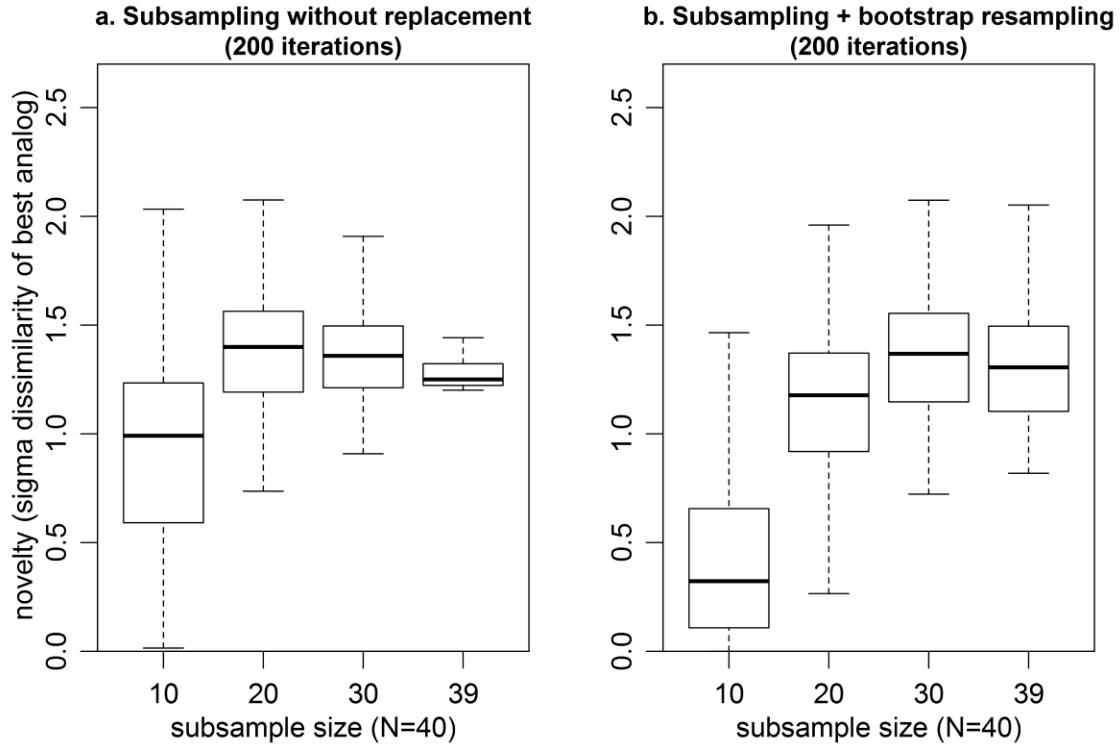


Figure S41: Effect of bootstrapping on estimates of error at various sample sizes at a single map cell. (a) The distribution of novelty calculated from subsamples of the 40-year time series converges as the sample size ($n=10, 20, 30$, & 39) approaches the population size ($N=40$). (b) Bootstrap resamples of the subsamples do not converge because the effective sample size of the PCA is reduced. Bootstrap estimates of novelty have a downward bias in small subsamples ($n \leq 20$).

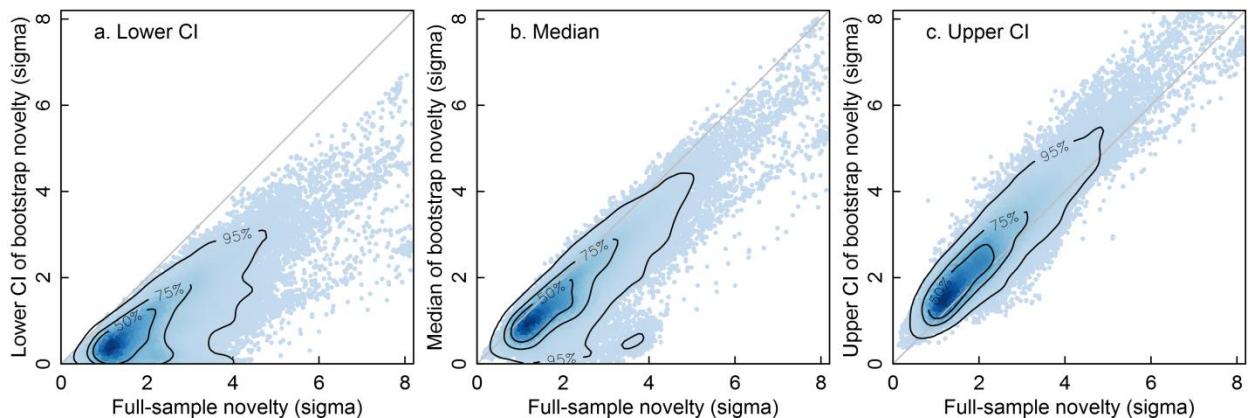


Figure S42: (a) Lower bounds of the 95% confidence interval for the $n=120$ bootstrapped novelty calculations for each map cell, plotted against the novelty for each map cell calculated from the full time series of the ICV proxy. Contours show the regions containing 50%, 75%, and 95% of the points. The grey 1:1 line indicates equivalence between the full-sample and bootstrapped novelty calculation. (b) and (c) show the medians and upper 95% CI bounds, respectively. These plots demonstrate that bootstrapping introduces a downward bias into calculations of sigma dissimilarity.

In conclusion, the ICV proxy time series with <40 complete years should be considered to be small samples in the context of a 12-dimensional principal components analysis. Despite the potential for small samples to produce unstable novelty results, the similarity between the 12-dimensional results and the 6-dimensional results (Figures 8 and S21) suggests that limited sample size is not a major source of error in the main novelty results presented in this publication. Areas with very small sample size (i.e., sample size less than double the dimensionality) are limited to 5% of the area outside Mexico. The bootstrap analysis indicates some isolated locations where novelty results may be unreliable due to PCA instability. However, bootstrapping overestimates novelty error because it reduces the effective sample size of the PCA. Estimation of error in sigma dissimilarity is non-trivial and an area for future research. In the absence of a reliable estimate of error, the ratio of time series observations to the dimensionality of the analysis should be maximized wherever possible. This could be achieved by parsimonious variable selection, higher thresholds for complete years in ICV proxies, and even pooling of monthly observations into each seasonal variable.

S15: Basic code for calculating and mapping climatic novelty using the sigma dissimilarity metric.

The Supporting Information repository for this publication includes a zip folder (“gcb13645-sup-0002-AppendixS15.txt”) containing the R script and data required for calculating and mapping climatic novelty on an 8km grid of North America. The R script also provides the code to prepare additional ClimateNA data for other emissions scenarios, time periods, and climate models. Download the file, change the extension of the file from “txt” to “zip” and open as a windows zip file.

S16: Spatial data for Figure 7

The Supporting Information repository for this publication includes a zip folder (“gcb13645-sup-0003-AppendixS16.zip”) containing GeoTIFF files of the novelty results shown in Figure 7 of the publication. Download the file, change the extension of the file from “txt” to “zip” and open as a windows zip file.

The zip folder contains the following files:

- SigmaNovelty.RCP45.NAnaec4.tif: RCP4.5 novelty for a 4-km grid of North America
- SigmaNovelty.RCP45.WNAnaec2.tif: RCP4.5 novelty for a 2-km grid of the Western Cordillera
- SigmaNovelty.RCP85.NAnaec4.tif: RCP8.5 novelty for a 4-km grid of North America
- SigmaNovelty.RCP85.WNAnaec2.tif: RCP8.5 novelty for a 2-km grid of the Western Cordillera

References

- Director, H., and L. Bornn. 2015. Connecting point-level and gridded moments in the analysis of climate data. *Journal of Climate* 28:3496–3510.
- Forster, P. M., T. Andrews, P. Good, J. M. Gregory, L. S. Jackson, and M. Zelinka. 2013. Evaluating adjusted forcing and model spread for historical and future scenarios in the CMIP5 generation of climate models. *Journal of Geophysical Research Atmospheres* 118:1139–1150.
- Gotelli, N. J., and W. Ulrich. 2012. Statistical challenges in null model analysis. *Oikos* 121:171–180.
- Guttman, N. B. 1999. Accepting the Standardized Precipitation Index: a Calculation Algorithm. *JAWRA Journal of the American Water Resources Association* 35:311–322.
- Harris, I., P. D. Jones, T. J. Osborn, and D. H. Lister. 2014. Updated high-resolution grids of monthly climatic observations - the CRU TS3.10 Dataset. *International Journal of Climatology* 34:623–642.
- Jones, P. D., T. J. Osborn, K. R. Briffa, C. K. Folland, B. Horton, L. V Alexander, D. E. Parker, and N. a Rayner. 2001. Adjusting for sampling density in grid-box land and ocean surface temperature time series. *Journal of Geophysical Research* 106:3371–3380.
- Kobayashi, S., Y. Ota, Y. Harada, A. Ebita, M. Moriya, H. Onoda, K. Onogi, H. Kamahori, C. Kobayashi, H. Endo, K. Miyaoka, and K. Takahashi. 2015. The JRA-55 Reanalysis: General Specifications and Basic Characteristics. *Journal of the Meteorological Society of Japan. Ser. II* 93:5–48.
- Olson, D. M., E. Dinerstein, E. D. Wikramanayake, N. D. Burgess, G. V. N. Powell, E. C. Underwood, J. a. D'amico, I. Itoua, H. E. Strand, J. C. Morrison, C. J. Loucks, T. F. Allnutt, T. H. Ricketts, Y. Kura, J. F. Lamoreux, W. W. Wettenberg, P. Hedao, and K. R. Kassem. 2001. Terrestrial ecoregions of the world: A new map of life on Earth. *BioScience* 51:933–938.
- Osborne, J. W., and A. B. Costello. 2004. Sample size and subject to item ratio in principal components analysis and exploratory factor analysis. *Practical Assessment, Research & Evaluation* 9:1–13.
- Schneider, U., A. Becker, P. Finger, A. Meyer-Christoffer, M. Ziese, and B. Rudolf. 2014. GPCC's new land surface precipitation climatology based on quality-controlled in situ data and its role in quantifying the global water cycle. *Theoretical and Applied Climatology* 115:15–40.
- Williams, J. W., S. T. Jackson, and J. E. Kutzbach. 2007. Projected distributions of novel and disappearing climates by 2100 AD. *Proceedings of the National Academy of Sciences of the United States of America* 104:5738–5742.