Supplement of Adv. Stat. Clim. Meteorol. Oceanogr., 10, 123–141, 2024
https://doi.org/10.5194/ascmo-10-123-2024-supplement
© Author(s) 2024. CC BY 4.0 License.

Supplement of

Spatiotemporal functional permutation tests for comparing observed climate behavior to climate model projections

Joshua P. French et al.

Correspondence to: Joshua P. French (joshua.french@ucdenver.edu)

The copyright of individual parts of the supplement might differ from the article licence.

1 Secondary climate model evaluation results

Table S1 provides additional details about regional climate models (RMCs) discussed in the paper.

Table S1: Regional climate models used to create NA-CORDEX data sets used in this paper.

Acronym	Model Name	References
CanRCM4	Canadian Regional Climate Model, version 4	Scinocca et al. (2016)
HIRHAM5	High-Resolution Limited Area Model with ECHAM Physics, version 5	Christensen et al. (2007)
QCRCM5	Canadian Regional Climate Model, version 5 (contributed by University of Quebec at Montreal)	Martynov et al. (2013), Šeparović et al. (2013)
RCA4	Rossby Centre regional atmospheric model, version 4	Samuelsson et al. (2011)
RegCM4	Regional Climate Model, version 4	Giorgi and Anyah (2012)
WRF	Weather and Research Forecasting Model	Skamarock et al. (2008)

Climate model output produced by RCM-GCM combinations that use the same GCM for initial conditions could be considered correlated. To assess whether this has an impact on our previous analyses, we performed a secondary analysis using only NA-CORDEX climate model output for models using different GCMs. Table S2 indicates the combination of models used in the secondary analysis.

Table S2: RCM-GCM combinations used to produce the NA-CORDEX data sets used for the secondary analysis.

	CanRCM4	HIRHAM5	QCRCM5	RCA4	RegCM4	WRF
CanESM2	x			x		
EC-Earth		x				
GEMATM-Can			x			
GFDL-ESM2M					x	
HadGEM2-ES					x	
MPI-ESM-LR						x

We first examine distributional equality for the reanalysis and climate model data for the subset of climate models. Our results for both the gridMET and Daymet bias-corrected data sets are similar to the original analysis.

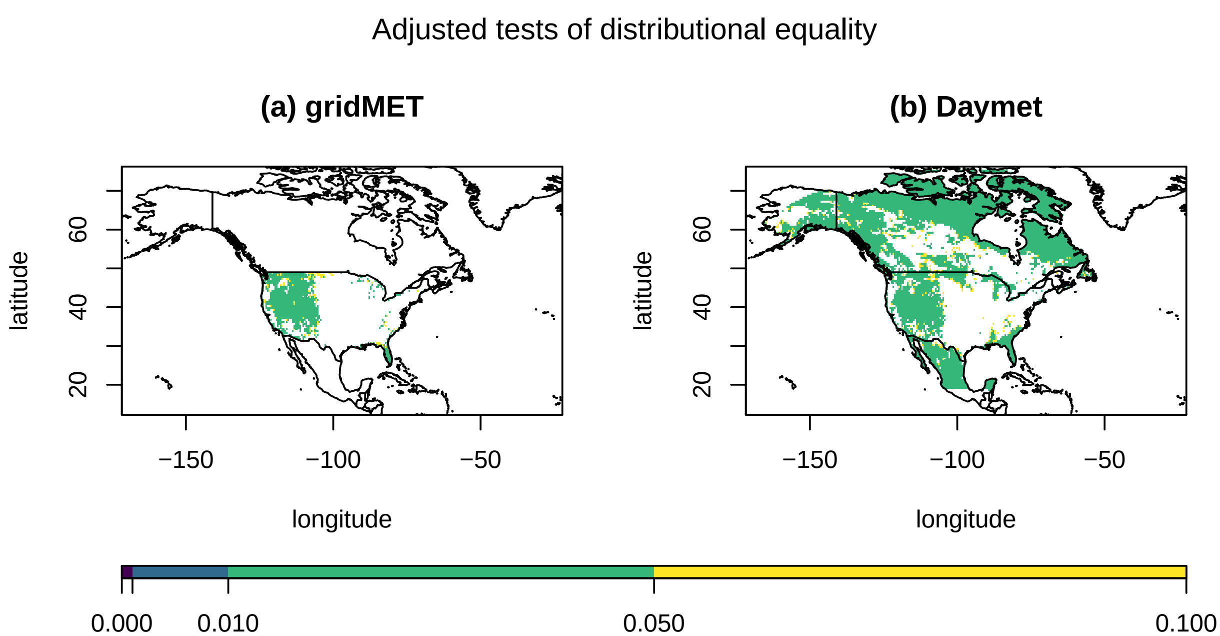

Figure S1: Heat maps of the p-values at significant locations ($\alpha = 0.10$) for the distributional equality test after using the FDR-controlling procedure proposed by Benjamini and Yekutieli (2001) for the (a) gridMET bias-corrected data, (b) Daymet bias-corrected data. Pixels with insignificant test statistics are not colored.

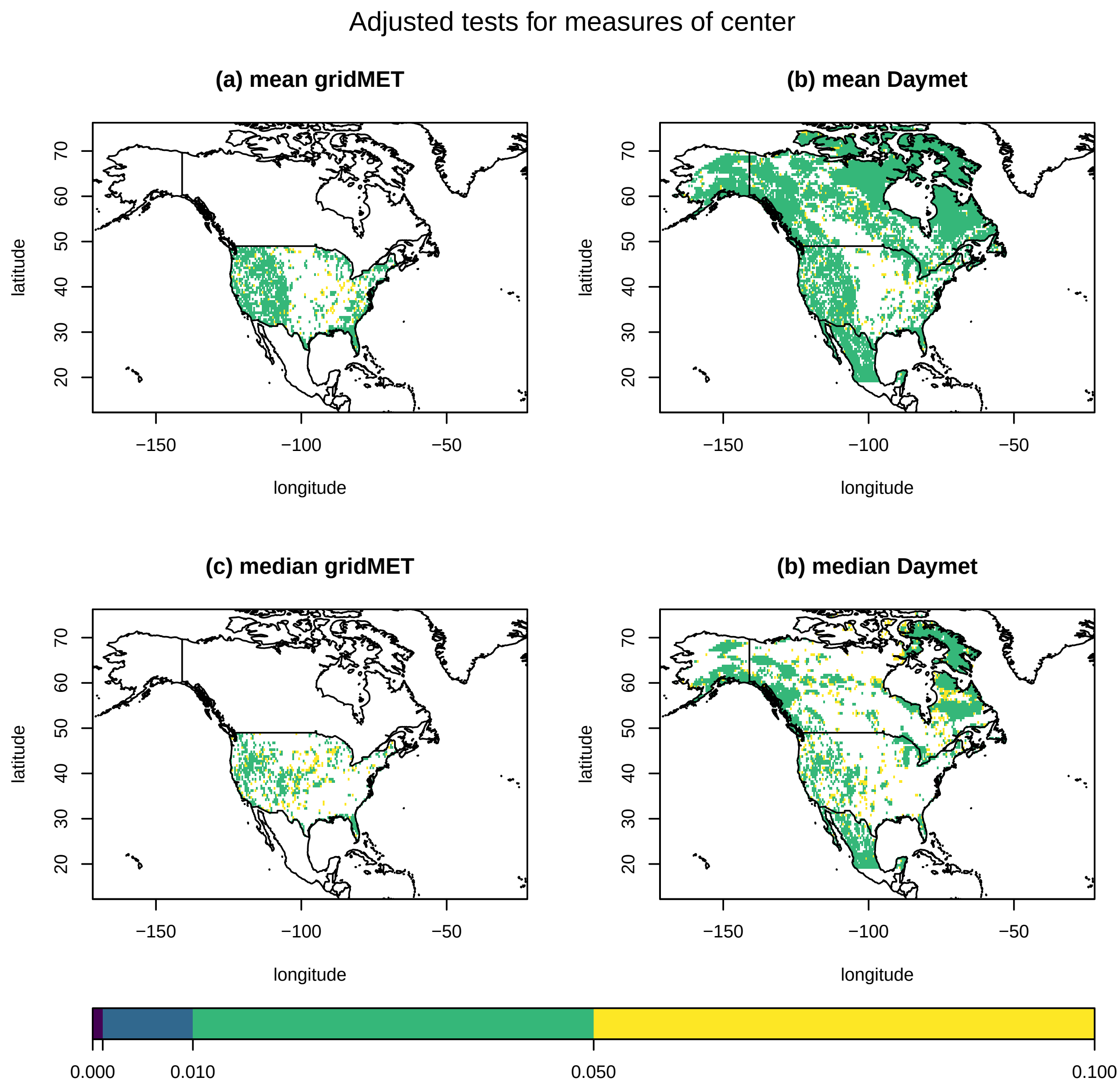

Figure S2: Heat maps of the p-values at significant locations ($\alpha = 0.10$) after using the FDR-controlling procedure proposed by Benjamini and Yekutieli (2001) for the (a) gridMET bias-corrected data with a test of 55-year mean temperature equality, (b) Daymet bias-corrected data with a test of 55-year mean temperature equality, (c) gridMET bias-corrected data with a test of 55-year median temperature equality, (d) Daymet bias-corrected data with a test of 55-year median temperature equality. Pixels with insignificant test statistics are not colored.

Next, we compare measures of center for the reanalysis and climate model data for the subset of climate models. Our results for both the gridMET and Daymet bias-corrected data sets are similar to the original analysis.

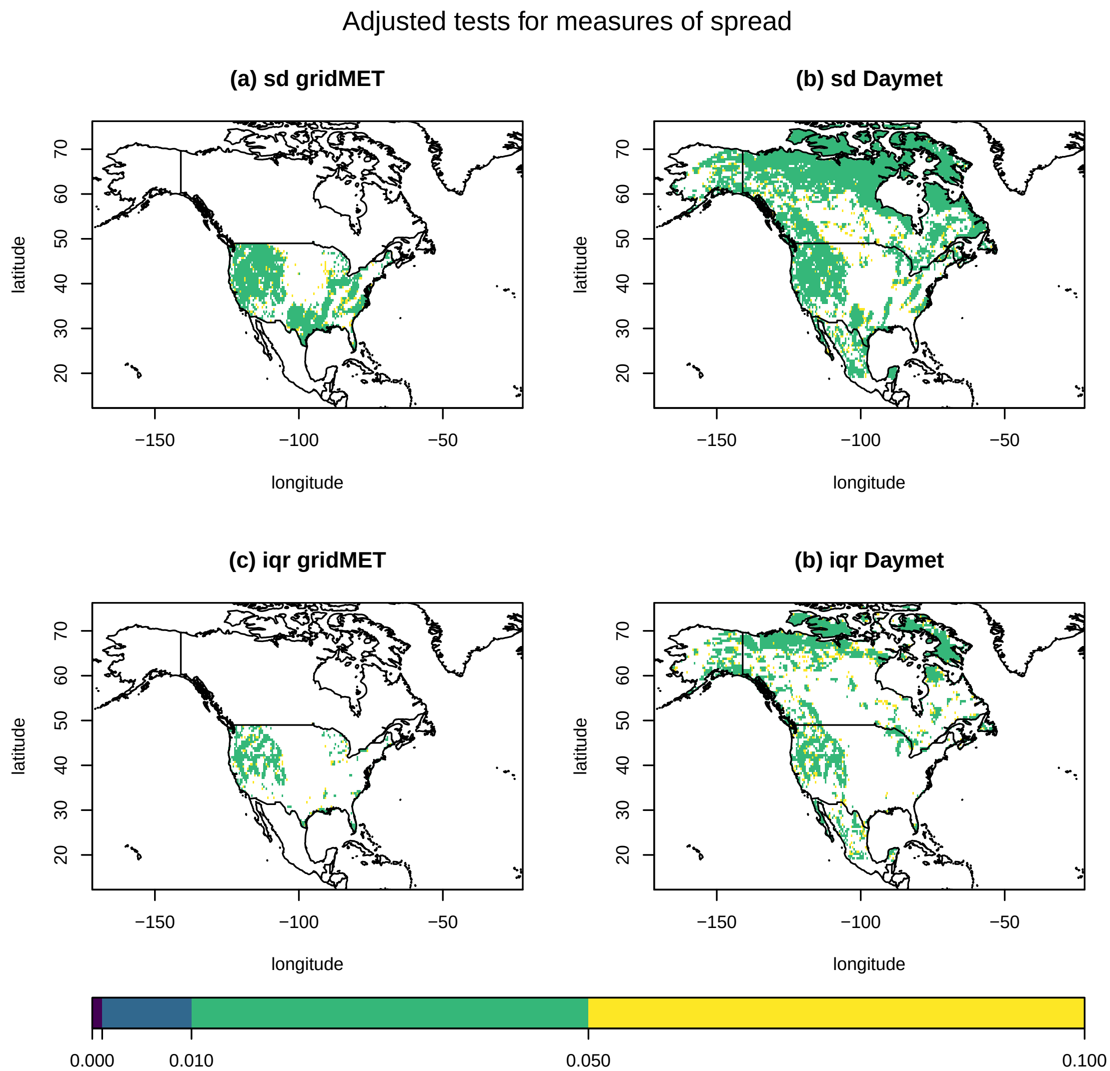

Figure S3: Heat maps of the p-values at significant locations ($\alpha = 0.10$) after using the FDR-controlling procedure proposed by Benjamini and Yekutieli (2001) for the (a) gridMET bias-corrected data with a test of 55-year standard deviation temperature equality, (b) Daymet bias-corrected data with a test of 55-year standard temperature equality, (c) gridMET bias-corrected data with a test of 55-year interquartile temperature equality, (d) Daymet bias-corrected data with a test of 55-year interquartile range temperature equality. Pixels with insignificant test statistics are not colored.

Lastly, we compared functional characteristics of the reanalysis and climate data over time at each spatial location. Specifically, we compared coefficients for b-splines fit to the data available at each spatial location over time for both the reanalysis data and the subset of climate model output. Our results for both the gridMET and Daymet bias-corrected data sets are similar to the original analysis, though the p-values tend to be larger for the subset of climate models.

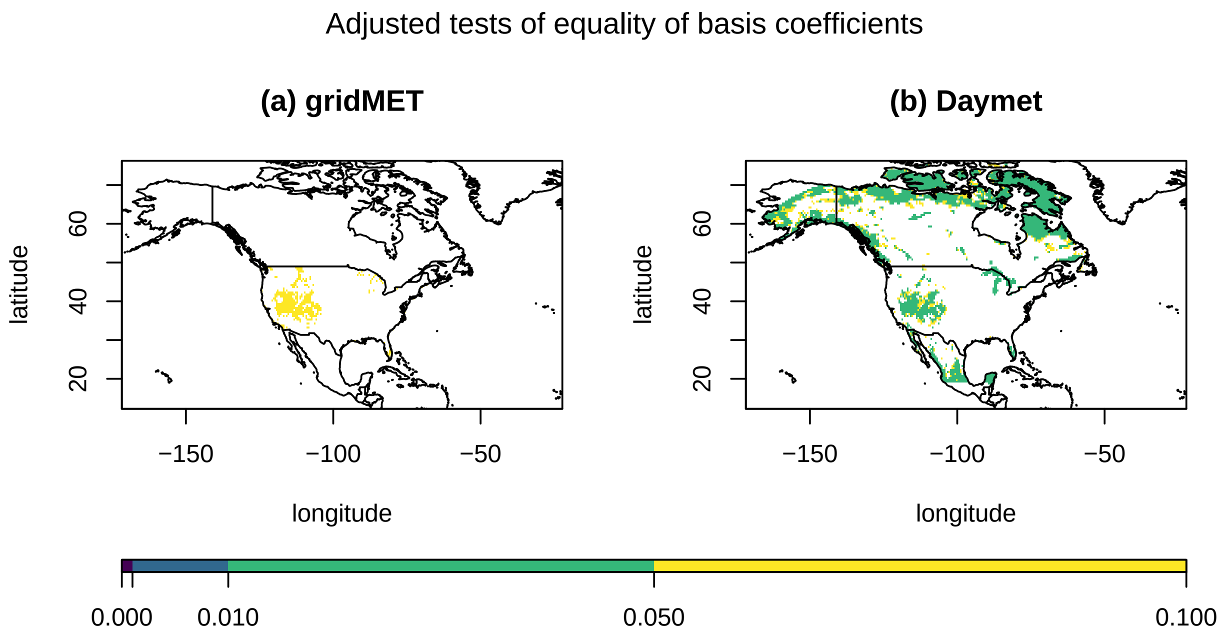

Figure S4: Heat maps of the p-values at significant locations ($\alpha = 0.10$) after using the FDR-controlling procedure proposed by Benjamini and Yekutieli (2001) for the (a) gridMET bias-corrected data for a test of coefficient equality, (b) Daymet bias-corrected data for a test of coefficient equality. Pixels with insignificant test statistics are not colored.

2 Additional comments about multiple comparisons and power

The standard permutation procedure lacks testing power. One implementation of the Benjamini–Yekutieli (BY) procedure (Benjamini and Yekutieli, 2001) takes the standard p-values and adjusts them upward so that testing can be performed at a fixed significance level while addressing the multiple comparisons problem. Since the BY-adjusted p-values are uniformly larger than the unadjusted p-values, any location significant after the BY p-value adjustment is automatically significant at the same significance level for the unadjusted p-values. However, we would also expect additional significant locations when using the unadjusted p-values; the locations significant for the unadjusted p-values will extend from the locations that are significant for the adjusted p-values. However, the standard permutation test may lack power to detect any significant locations after adjusting for multiple comparisons.

Figure S5 displays a comparison of the p-value spatial distributions for a test of distributional equality based on the statistic in Eq. (3) of the main paper for both the gridMET and Daymet biased corrected data sets using both standard and stratified permutation tests and unadjusted and BY-adjusted p-values. The standard permutation tests do not have enough power to detect any significant locations after adjusting for multiple comparisons. When considering the stratified permutation tests, there are more significant locations when using the unadjusted p-values than the BY-adjusted p-values, but the overall pattern of the significant locations is similar.

3 Additional discussion about independence assumptions

We have provided additional details about some of the independence assumptions related to the proposed methodology.

We assume that the observations in year n follow the model $X_n(\mathbf{s}, t) = \mu_n(\mathbf{s}, t) + \varepsilon_n(\mathbf{s}, t)$, where t is time within a year. In our application, the variable t is a calendar month, $t = 1, 2, \ldots, 12$, starting with January. The independence assumption means that the error surfaces $\varepsilon_n(\cdot, \cdot)$ are independent and identically distributed across n. The mean surfaces $\mu_n(\cdot, \cdot)$ look similar for each year n and dominate the shape of the observations, cf. Figure 3 of the original paper.

We note that the division into years starting with January is arbitrary. The key is to use any interval that includes the whole year to account for the annual periodicity. We also note that we do not assume that the errors, say in January and July have the same distribution or are independent. We only assume that the whole annual error curves are i.i.d.

We provide a more detailed explanation here. Suppose we have a multivariate time series $x_{i,k}$, where i indexes time and k indexes the component. In our case, we have $i = 1, 2, \ldots, 55$ years, $k = 1, 2, \ldots 12$ months. The theory to be presented applies to stationary time series, so we first transform the temperature data (at a fixed location) to approximate stationarity and work with

$$y_{i,k} = x_{i,k} - \bar{x}_k, \quad \bar{x}_k = \frac{1}{N} \sum_{i=1}^{N} x_{i,k}, \quad N = 55.$$

For $h = 1, 2, 3, 4 \approx \ln 55$ and $k, l \in \{1, 2, \ldots, 12\}$, we define the cross-correlations

$$\hat{\rho}_{kl}(h) = \operatorname{corr}\left(y_{i+h,k}, y_{i,l}\right).$$

These are just the usual sample correlations. For example, if $h = 2$, $k = 1$ (January) and $l = 7$ (July), we compute the correlation coefficient of

$$y_{3,1}, y_{4,1}, y_{5,1}, \ldots, y_{54,1}, y_{55,1},$$

and

$$y_{1,7}, y_{2,7}, y_{3,7}, \ldots, y_{52,7}, y_{53,7}.$$

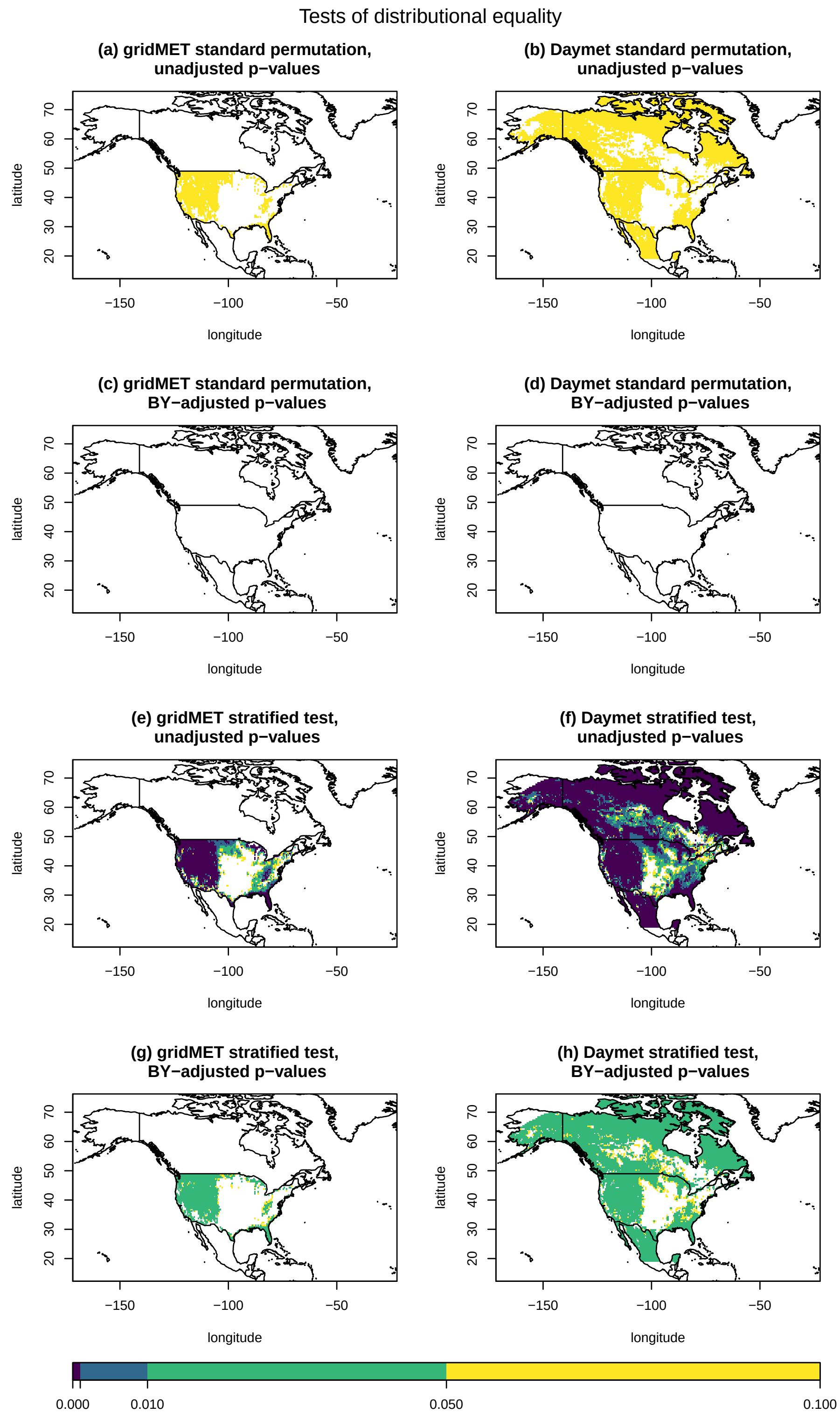

Figure S5: Heat maps of the p-values ≤ 0.10 for the distributional equality tests using combinations of data set (gridMET or Daymet bias-corrected data), testing procedure (standard or stratified permutation test), and p-value (unadjusted or BY-adjusted). The combination of data set, testing procedure, and p-value used for inference is specified in the panel label.

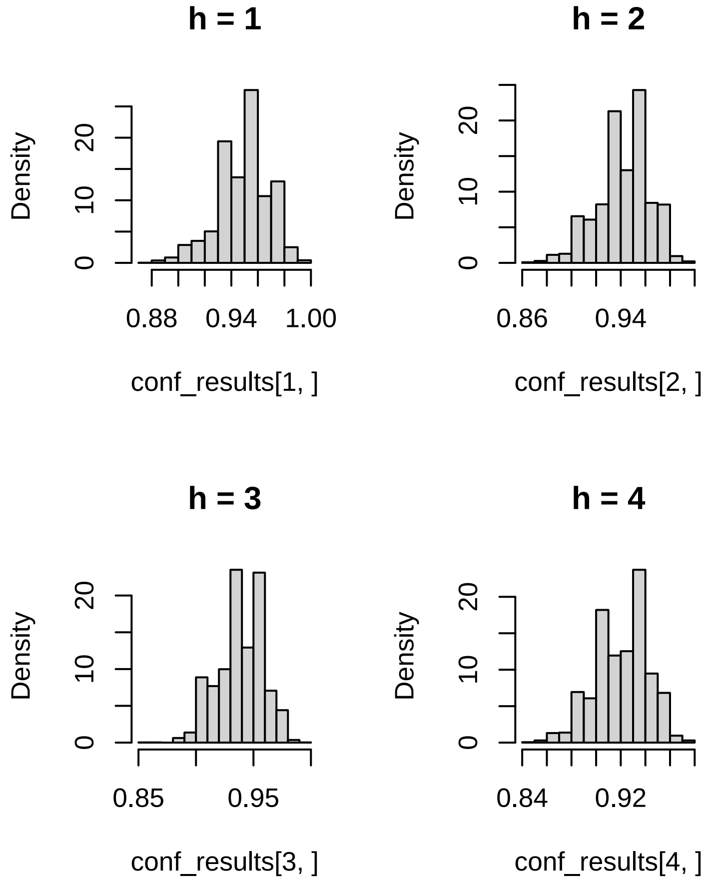

Figure S6: Histograms of the proportion of the 12×12 cross-correlation values less than or equal to $2/\sqrt{55}$, where the proportions are computed for 3,357 spatial locations, for lags $h = 1$, 2, 3, and 4.

Thus, for each h we have $12 \times 12 = 144$ correlations.

If the vectors $\mathbf{y}_1, \mathbf{y}_2, \mathbf{y}_3, \ldots$ are independent, each $\hat{\rho}_{kl}(h)$ is asymptotically normal with mean zero and variance $1/\sqrt{N}$. This follows, for example, from Theorem 11.2.2 of Brockwell and Davis (1991). Thus, if the years are independent, about 95% of the 144 $\hat{\rho}_{kl}(1)$ should be smaller than $2/\sqrt{55} \approx 0.27$ (i.e., approximately 137 of the 144 sample correlations). The same should be true for $h = 2, 3, 4$.

The above argument applies to a fixed location $\mathbf{s}$. Since in the study area we have over 3,300 locations, for some of them, due to random variability, the proportion will be lower than 95% for others higher. Figure S6 shows histograms of the proportion of cross-correlations $\hat{\rho}_{i,j}(\mathbf{s}, h) \leq 2/\sqrt{55}$. Without random variability, in an infinite sample, and under perfect independence, for every location the proportion should be 0.95. In our finite samples, the vast majority are close 0.95.

We provide additional discussion regarding the yearly distributions of the models to support the assumption that $X^{M_j} \overset{i.i.d}{\sim} F^M$. The independence of the fields X^{M_j} and $X^{M_{j'}}$

for $j \neq j'$ means that any functional computed from X^{M_j} is independent from any functional computed from $X^{M_{j'}}$. An analogous statement is true for the equality of the distributions of these fields. The i.i.d. assumption cannot thus be fully verified. However, it is possible to provide some evidence to support it.

One can proceed as follows. Choose at random $K = 100$ locations and for each year n compute

$$G_{j,n} = \frac{1}{K} \sum_{k=1}^{K} \frac{1}{12} \sum_{i=1}^{12} X_n^{M_j}(\mathbf{s}_k, t_i), \quad j = 1, 2, \ldots 15,$$

i.e., we average the temperature values in year n across 100 randomly selected spatial locations for all 12 months in year n. The function $G_{j,n}$ is an example of a relevant functional of the field X^{M_j} our of infinitely many possible functionals. Next, we compute the 15×15 correlation matrix for the above variables. If 95% of the off-diagonal entries are smaller in absolute value than $2/\sqrt{N_n} \approx 0.27$, then there is evidence to support the assumption of independence.

We created the 15×15 correlation matrix below. We see that almost all off-diagonal values are within $\pm 2/\sqrt{n}$.

Table S3: The matrix of estimated correlations between the $G_{j,n}$ for $j = 1, 2, \ldots, 15$. About 95% of the values should be less than 0.27 if the X^{M_j} are independent, $j = 1, 2, \ldots, 15$.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	1.00	0.11	-0.06	0.04	0.05	0.28	-0.05	0.18	0.20	0.06	0.02	-0.16	0.06	0.13	0.32
2	0.11	1.00	-0.01	0.28	-0.03	0.22	0.18	-0.03	0.38	-0.03	0.20	0.07	0.04	0.22	0.24
3	-0.06	-0.01	1.00	0.32	0.10	0.25	-0.10	0.66	0.24	-0.09	0.19	0.15	0.19	0.19	0.01
4	0.04	0.28	0.32	1.00	0.17	0.46	0.10	0.30	0.23	-0.03	0.35	0.01	0.34	0.27	0.30
5	0.05	-0.03	0.10	0.17	1.00	0.23	0.42	0.06	0.31	0.10	0.21	-0.05	0.15	0.02	0.09
6	0.28	0.22	0.25	0.46	0.23	1.00	-0.09	0.21	0.21	0.06	0.21	0.04	0.10	0.23	0.68
7	-0.05	0.18	-0.10	0.10	0.42	-0.09	1.00	-0.10	0.10	0.05	0.28	-0.09	0.19	-0.01	0.04
8	0.18	-0.03	0.66	0.30	0.06	0.21	-0.10	1.00	0.26	0.09	0.11	0.04	0.09	0.23	0.03
9	0.20	0.38	0.24	0.23	0.31	0.21	0.10	0.26	1.00	0.08	0.15	0.11	0.17	0.37	0.09
10	0.06	-0.03	-0.09	-0.03	0.10	0.06	0.05	0.09	0.08	1.00	0.18	0.49	0.10	0.01	0.00
11	0.02	0.20	0.19	0.35	0.21	0.21	0.28	0.11	0.15	0.18	1.00	0.16	0.56	0.06	0.24
12	-0.16	0.07	0.15	0.01	-0.05	0.04	-0.09	0.04	0.11	0.49	0.16	1.00	0.20	0.05	0.02
13	0.06	0.04	0.19	0.34	0.15	0.10	0.19	0.09	0.17	0.10	0.56	0.20	1.00	0.17	0.10
14	0.13	0.22	0.19	0.27	0.02	0.23	-0.01	0.23	0.37	0.01	0.06	0.05	0.17	1.00	0.16
15	0.32	0.24	0.01	0.30	0.09	0.68	0.04	0.03	0.09	0.00	0.24	0.02	0.10	0.16	1.00

To verify the assumption of equality in distribution between the X^{M_j}, $j = 1, 2, \ldots, 15$, one can perform $(15 \times 14)/2 = 105$ Cramér-von Mises tests of the equality in distribution. Without any multiple testing adjustment, we expect 5% of them to reject the null hypothesis of identical distribution. We performed $(15 \times 14)/2 = 105$ Cramér-von Mises tests; 15% of the p-values were less than 0.05, so the evidence for an identical distribution is weaker.

References

Benjamini, Y. and Yekutieli, D. (2001). The control of the false discovery rate in multiple testing under dependency. *The Annals of Statistics*, 29(4):1165 – 1188.

Brockwell, P. J. and Davis, R. A. (1991). *Time Series: Theory and Methods.* Springer Science & Business media.

Christensen, O. B., Drews, M., Christensen, J. H., Dethloff, K., Ketelsen, K., Hebestadt, I., and Rinke, A. (2007). The hirham regional climate model. version 5 (beta).

Giorgi, F. and Anyah, R. (2012). The road towards regcm4. *Climate Research*, 52:3–6.

Martynov, A., Laprise, R., Sushama, L., Winger, K., Šeparović, L., and Dugas, B. (2013). Reanalysis-driven climate simulation over cordex north america domain using the canadian regional climate model, version 5: model performance evaluation. *Climate Dynamics*, 41:2973–3005.

Samuelsson, P., Jones, C. G., Willén, U., Ullerstig, A., Gollvik, S., Hansson, U., Jansson, C., Kjellström, E., Nikulin, G., and Wyser, K. (2011). The rossby centre regional climate model rca3: model description and performance. *Tellus A: Dynamic Meteorology and Oceanography*.

Scinocca, J. F., Kharin, V. V., Jiao, Y., Qian, M. W., Lazare, M., Solheim, L., Flato, G. M., Biner, S., Desgagne, M., and Dugas, B. (2016). Coordinated global and regional climate modeling. *Journal of Climate*, 29(1):17 – 35.

Šeparović, L., Alexandru, A., Laprise, R., Martynov, A., Sushama, L., Winger, K., Tete, K., and Valin, M. (2013). Present climate and climate change over north america as simulated by the fifth-generation canadian regional climate model. *Climate Dynamics*, 41(11-12):3167–3201.

Skamarock, W. C., Klemp, J. B., Dudhia, J., Gill, D. O., Barker, D. M., Duda, M. G., Huang, X.-Y., Wang, W., Powers, J. G., et al. (2008). A description of the advanced research wrf version 3. *NCAR technical note*, 475:113.