**Supplementary Materials**

**MODERN ANALOG TECHNIQUE UNCERTAINTIES**

Constraining uncertainty of Modern Analog Technique (MAT) reconstructions is difficult because there are multiple ways to measure the statistical uncertainty, but no specific way to calibrate the reconstructions to past-observed climate. We use the root mean square error (RMSE) of the observed and predicted modern climate variables to estimate the uncertainty, and we propagate the error assuming each lake represents a measurement of the same regional mean (uncertainty=RMSE/$\sqrt{6}$). To test whether our assumption is correct we tested whether averaging the predicted value of 6 random sites decreased the uncertainty. We also tested whether spatial autocorrelation artificially improved our test statistics by ‘thinning’ the surface pollen data.

**METHODS**

**Averaging multiple sites**

We test the reduction in uncertainty that resulted from averaging multiple records by applying the multi-site averaging to the leave-one-out surface samples validation exercises, using the MAT to predict the climate variable for each modern sample and then randomly selecting six samples and averaging their observed and predicted climates. We repeated the random selection of six samples to produce the same number of six-sample means as there are samples in the modern analog dataset. We then calculate the RMSE and R2 of these new observed and predicted values. We repeat these steps 100 times to allow for the calculation of a mean and standard deviation to understand how the random selection of sites impacts the statistics.

**Spatial autocorrelation**

To test for the effect of spatial autocorrelation, we also ‘thinned’ the modern-surface sample data to determine whether climatic similarities in geographically close surface samples artificially inflate the calculated accuracy and precision of the reconstructions. Climate variables are spatially autocorrelated because the environment of neighboring samples is more likely to be similar than the environment of geographically distant samples. Temperature tends to be more spatially autocorrelated than precipitation because precipitation can rapidly change in space due to orographic effects (Shinker, 2010). One method used to account for autocorrelation is subsampling a larger dataset at regular intervals to decrease the redundancy of the residuals in the data (http://seismo.berkeley.edu/~kirchner/eps\_120/Toolkits/

Toolkit\_11.pdf). To do so, we modified code used to grid surface-pollen data (https://www.ncdc.noaa.gov/paleo-search/study/22992; Marsicek et al., 2018), and randomly selecteed only one sample as a possible analog per 0.5° latitude/longitude grid squares with 70 m vertical constraints. The highly variable topography in the Rocky Mountains necessitates a vertical constraint as climate space changes rapidly between nearby samples at different elevations. We compared the reconstructions and reconstruction statistics from the ‘thinned’ data to the results from the entire surface pollen data set to determine the impact of spatial autocorrelation.

**RESULTS**

**Averaging multiple sites**

Before error propagation, annual temperature has an RMSE of 1.7°C, annual precipitation has an RMSE of 130 mm, winter precipitation has an RMSE of 51 mm, and summer precipitation has an RMSE of 21 mm (Table S1B). After error propagation, when considering the reconstructions from the six lakes as multiple measurements of the same regional mean, the RMSEs decrease for annual temperature (0.7°C), annual precipitation (52 mm), winter precipitation (21 mm), and summer precipitation (14 mm) (Table S1A). We use the RMSE after error propagation as the uncertainty of our climate reconstructions.

To test whether it is appropriate to consider each lake as a replicate measurement of one regional mean, we use 100 iterations of modern validation comparisons to evaluate the effect of averaging multiple reconstructions to produce a single composite for six sites. For the 100 averages of six random sites, we find that comparison of the observed and predicted six-site average of mean annual temperature has a reduced RMSE of 0.70 ± 0.03°C (mean ± standard deviation) and a higher R2 of 0.79 ± 0.02 (Table SIC). Modern validations for averaging across six random sites for annual precipitation results in a decreased RMSE of 54 ± 1.9 mm (mean ± standard deviation) and a marginally decreased R2 of 0.67 ± 0.02 (mean ± standard deviation) (Table SIC). Winter precipitation has a decreased RMSE (22 ± 0.9mm) and a marginally reduced R2 (0.65 ± 0.03) (Table S1C). Summer precipitation has a decreased RMSE (15 ± 0.5mm) and a marginally decreased R2 (0.49 ± 0.02) (Table S1C). The RMSEs calculated from the random six-site averages are very similar to the RMSEs from single-site predictions after error propagation, confirming our use of RMSE/$\sqrt{6}$ as the uncertainty for our reconstructions.

**Spatial autocorrelation**

Thinning the surface-pollen data to account for the effects of spatial autocorrelation decreases the R2 from 0.80, when including all surface samples, to 0.74 (Table S1B and S1D). The RMSE after error propagation increases from 0.7°C to 0.8°C (Table S1A and S1D). The reconstruction from the thinned data, like the reconstruction from all surface-pollen data, does not exhibit significant temperature changes during the past 2500 years (p=0.09; Fig. S1A).

Results using the thinned data also display the same increase in annual precipitation from the first millennium of the Common Era to the second (Fig. S1). The thinned data results in a minor increase in RMSE to 65 mm and a minor decrease in R2 to 0.66 (Table S1D). These worsened test statistics indicate that the RMSE and R2 calculated from all the modern surface samples included in our reconstruction have inflated precision and accuracy due to spatial autocorrelation. Although our reconstructions have marginally inflated test statistics, the cross-validation using the thinned data still has acceptable precision and accuracy. The millennial-scale trend and the two peaks at 600 and 100 BP are still significant (Fig. S1B).

Winter precipitation has a slightly decreased R2 (0.63) and increased RMSE (29 mm). Summer precipitation also has a slightly decreased R2 (0.55) and increased RMSE (33 mm). Both winter and summer precipitation reconstructions explain significantly (p<0.01) more variation than random reconstructions.

**CONCLUSIONS**

Modern cross-validations indicate that it is appropriate to propagate the error, assuming each lake provides a replicate measurement of the same mean for all four climate variables. Because wind-blown pollen derives from large source areas, the use of multiple pollen records from the same landscape could facilitate the goal of sampling from the regional mean climate in a way that would not be true of locally sensitive weather stations.

Thinning the surface-pollen data to include one sample per 0.5° latitude/longitude by 70 m rectangular prisms results in R2 values ranging from 0.55 to 0.74 (Table S1D), which is a small decrease from the range of 0.56 to 0.80 that results from the inclusion of all surface samples. After error propagation, the RMSE of mean annual temperature is 0.8°C (Table S1D), which is slightly higher than the 0.7°C RMSE resulting from the inclusion of all surface samples. The RMSE for annual, winter, and summer precipitation ranges from 29-65 mm (Table S1D), compared to 14-52 mm from including all surface samples. These slightly worsened statistics indicate spatial autocorrelation artificially improves R2 and RMSE, but only marginally.

**REFERENCES CITED**

Marsicek, J., Shuman, B.N., Bartlein, P.J., Shafer, S.L. and Brewer, S., 2018. Reconciling divergent trends and millennial variations in Holocene temperatures. *Nature*, *554*(7690), p.92.

Shinker, J.J., 2010. Visualizing spatial heterogeneity of western US climate variability. *Earth Interactions*, *14*(10), pp.1-15.

Table S1 – (A) The uncertainty used for the reconstructions, which is the RMSE after error propagation; and the RMSE and R2 of (B) leave-one-out cross-validations including all surface samples from the North American Modern Pollen Database (Whitmore et al., 2005) between 105°W and 117°W longitude, between 25°N and 60°N latitude, and above 1600 m a.s.l; (C) leave-one-out cross-validations for averaging the observed and predicted of the average of six sites; and (D) leave-one-out cross-validations including only one sample per 0.5° latitude x 0.5° longitude x 70 m elevation window (the error was propagated for the RMSE).

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | A. Reconstruction uncertainty | B. Modern cross-validation for single site  | C. Modern cross-validation accounting for six-site averages | D. Modern cross-validation thinned data |
|  | RMSE | RMSE | R2 | RMSE | R2 | RMSE | R2 |
| Mean annual temperature | 0.7°C | 1.7°C | 0.80 | 0.7± 0.03°C | 0.79±0.02 | 0.8°C | 0.74 |
| Total annual precipitation | 52 mm | 130 mm | 0.70 | 54±1.9 mm | 0.67±0.02 | 65 mm | 0.66 |
| Total winter precipitation | 21 mm | 51 mm | 0.67 | 22±0.9mm | 0.65±0.03 | 29 mm | 0.63 |
| Total summer precipitation | 14 mm | 34 mm | 0.56 | 15±0.5 mm | 0.49±0.02 | 33 mm | 0.55 |

****

Figure S1 – (A) Median of the mean annual temperature from the six lakes using ‘thinned’ surface pollen samples with only one surface sample per 0.5° latitude x 0.5° longitude x 70 m elevation window, (B) the median annual precipitation, (C) the median winter precipitation, (D) the median summer precipitation.