Supplemental Material

Supplemental Material

En-GARD: A statistical downscaling framework to produce and test large ensembles of climate projections

https://doi.org/10.1175/JHM-D-21-0142.1 *Journal of Hydrometeorology*

Ethan D. Gutmann[a], Joseph. J. Hamman[a,b], Martyn P. Clark[c],

Trude Eidhammer[a], Andrew W. Wood[a], Jeffrey R. Arnold[d]

a. National Center for Atmospheric Research, Boulder, Colorado

b. CarbonPlan, San Francisco, California

c. Centre for Hydrology, University of Saskatchewan Coldwater Laboratory, Canmore, Alberta, Canada

d. Responses to Climate Change Program, US Army Corps of Engineers

1. Quantile-Quantile plots for extreme precipitation events

Below we show quantile-quantile (QQ) plots between observed extreme precipitation amounts and En-GARD modeled precipitation amounts (Figure S1). In this instance, En-GARD is used to downscale the 50km WRF simulation driven by ERA-interim. This plot shows only the top 0.01% of daily precipitation amounts from across the entire domain. For the Analog-Regression (AR) and Pure-Analog (PA) algorithms, the QQ plot does not diverge substantially from the 1:1 line before approximately 300-400mm / day (corresponding to the 99.9999 percentile) except for the PA with all nine-predictors discussed below. This suggests that for most of the distribution, the algorithms in En-GARD realistically reproduce even moderately extreme precipitation amounts in the historical period. However, the statistical assumptions in this implementation breakdown for the most extreme cases. En-GARD is following the assumption that errors in the precipitation prediction are normally distributed after applying a cube-root transformation. Because there is no upper-bound to the normal distribution, in extremely rare instances, it is possible for En-GARD to predict precipitation amounts greater than 800 mm, while the maximum in the Maurer observational dataset is closer to 500 mm, although this is lower than the maximum ever recorded on Earth of 1825 mm (Holland, 1993) or North America of 1634 mm (Pasch et al., 2006). Adding an extreme value distribution to the capabilities of En-GARD might improve this fit in the future. As seen in figure 8 of the main manuscript, the PR algorithm consistently underestimates extreme precipitation events.

2. Sensitivity to a large predictor variable set

It is common for downscaling implementations to utilize a larger collection of predictor variables than has been done in this study. To put such an approach in context, we have provided

additional figures illustrating the use of nine different predictors with five upper-level variables, meridional "U", zonal "V", and vertical “W” winds, air temperature "T", and specific humidity "Q", as well as four near surface variables pressure "PSFC", 2m diurnal temperature range "DTR", precipitation from the cumulus scheme "PrC", and total precipitation "Pr".

a. Mean annual precipitation

The bias in mean annual precipitation when using all predictors is like the other predictor sets in most cases (Figure S2, S3). As noted in the main manuscript, the pure analog algorithm appears to have more grid cells with a larger negative bias when more predictors are used. The AR and PR algorithms do not exhibit this tendency, and the AR algorithm appears to have less of a negative bias averaged across the US, although the magnitude of some local biases increases (wetter in California, drier on the mid-Atlantic coast).

b. Extreme precipitation

The bias in extreme precipitation amounts is shown using quantile-quantile plots (Figure S1) and as maps for mean annual maximum (Figure S4) and histograms of all grid-cell biases (Figure S5). The results when using all nine predictors are not substantially different from the other datasets with the notable exception of the PA algorithm. As with mean annual precipitation, PA applied to nine predictors exhibits substantially more negatively biased grid cells.

c. Year-to-year variability

The year-to-year variability is shown for all methods (Figure S6). The use of all nine-predictors improves the correlation coefficient (r^2=0.92) and the root mean square error (RMSE=43mm) relative to the best performance with three predictors (r^2=0.87, RMSE=50mm). However, it remains the case that none of the En-GARD instantiations is able to match the wettest years in the record (1995, ‘96). To evaluate the role of stochastic uncertainty, we perform additional

post-processing on the raw GARD predictions to produce ten En-GARD ensemble members using different spatially correlated random fields. These ensemble members are shown in light grey lines (Figure S6). The ensemble encompasses all years except for 1980, '95, and '96.

d. Future

Finally, we show the effect that the use of nine predictor variables has on projections of future changes in mean annual precipitation (Figure S7) and mean annual maximum precipitation (Figure S8). The use of nine predictors modifies the change in mean annual precipitation, with many areas having larger magnitudes of increases and others having large magnitude decreases. As with other metrics, the PA algorithm has dampens the magnitude of these patterns, but it has a broadly similar spatial pattern (Figure S7). The pattern of changes when using all nine predictors is a mixture of the changes with the p and puv predictors (decreases in the non-coastal South) and changes with the uvq predictors (increases in the midwest and parts of the southwest). While agreeing in some areas the three prior predictor sets agreed (decreases along the west coast), but with a few areas in which the sign of the prediction differs from all other predictor sets (decreases along the coast of the northeast, and decreases throughout the interior western mountains).

Projected changes in mean annual maximum precipitation have larger differences across algorithm choices when using all nine predictors then with other predictor sets (Figure S8). Notably, in contrast to the primary results in the main manuscript, the AR algorithm predicts substantial increases in the mean annual maximum precipitation amounts. With the PR and PA algorithms, increases in mean annual maxima are smaller, and balanced by decreases in many regions.

3. Daily Variability in GARD relative to En-GARD

Here we illustrate the importance of the stochastic term in increasing the daily variability in precipitation. We use the standard deviation through time to quantify daily variability, and the ratio between the variability initially predicted by GARD, and the variability predicted by En-GARD after including the stochastic term is shown (Figure S9). This figure illustrates that for the AR algorithm and puv predictor set, only 37% of the daily variability is present in the GARD daily prediction relative to the En-GARD prediction. While more than 70% of the variability is present in GARD along much of the west coast, only 20-40% of the variability is present in much of the rest of the CONUS. In contrast, when using the AR algorithm applied to the predictor set with all nine predictors, GARD itself predicts 50% of the En-GARD daily variability averaged across the domain, with much of the CONUS west of -90° showing more than 60% of the variability predicted by GARD directly. The greatest fraction of variability is predicted by GARD for the 9-predictor set, followed by puv, uvq, and last p. The greatest fraction of variability is predicted by the AR algorithm, followed by PA, and PR algorithms.

4. Scatter plots

To illustrate the reasonability of an ordinary least-squares regression given the predictors used, a collection of "good" and "bad" predictor combinations, locations, and days are shown. For illustration, we select a "good" grid cell located at -121.354°E, 46.473°N and a "bad" grid cell located at 34.820°N, -99.669°E. The results from the pure regression algorithm as applied to the p predictor set, and as applied to the puv predictor set are shown first (Figure S10). Because regressions are calculated and applied in cube-root transformed space, the axes here are also transformed. Applying directly to the p predictor set has a coefficient of determination of 0.39,

while applying the regression to the puv predictor set increases the coefficient slightly to 0.42 for the “good” grid cell. For the “bad” grid cell, both coefficients are less than 0.1 and the effect of this poor fit on the slope of the regression line is evident in the relatively narrow distribution predicted by GARD. The AR algorithm is illustrated when predicting a “bad” day, 1980-01-21, and a “good” day, 1980-02-27 for both the puv and uvq predictor sets (Figures S11). In all plots, the data are close to normally distributed, though a statistical significance test (D’Agostino and Pearson, 1973) shows that the difference from normality is statistically significant ($p<0.01$). When the AR algorithm finds a collection of analog days with no strong relationship, the regression line will have a flat slope with an intercept equal to the distribution mean and the prediction will be nearly identical to that of the PA algorithm.

5. References

D’Agostino, R. and Pearson, E. S. 1973: Tests for departure from normality, *Biometrika*, 60, 613-622.

Holland, G.J. 1993: WMO/TC-No. 560, Report No. TCP-31, World Meteorological Organization; Geneva, Switzerland.

Pasch, R.J., Blake, E.S., Cobb, H.D., and Roberts, D.P. 2006: Tropical Cyclone Report: Hurricane Wilma. National Hurricane Center.

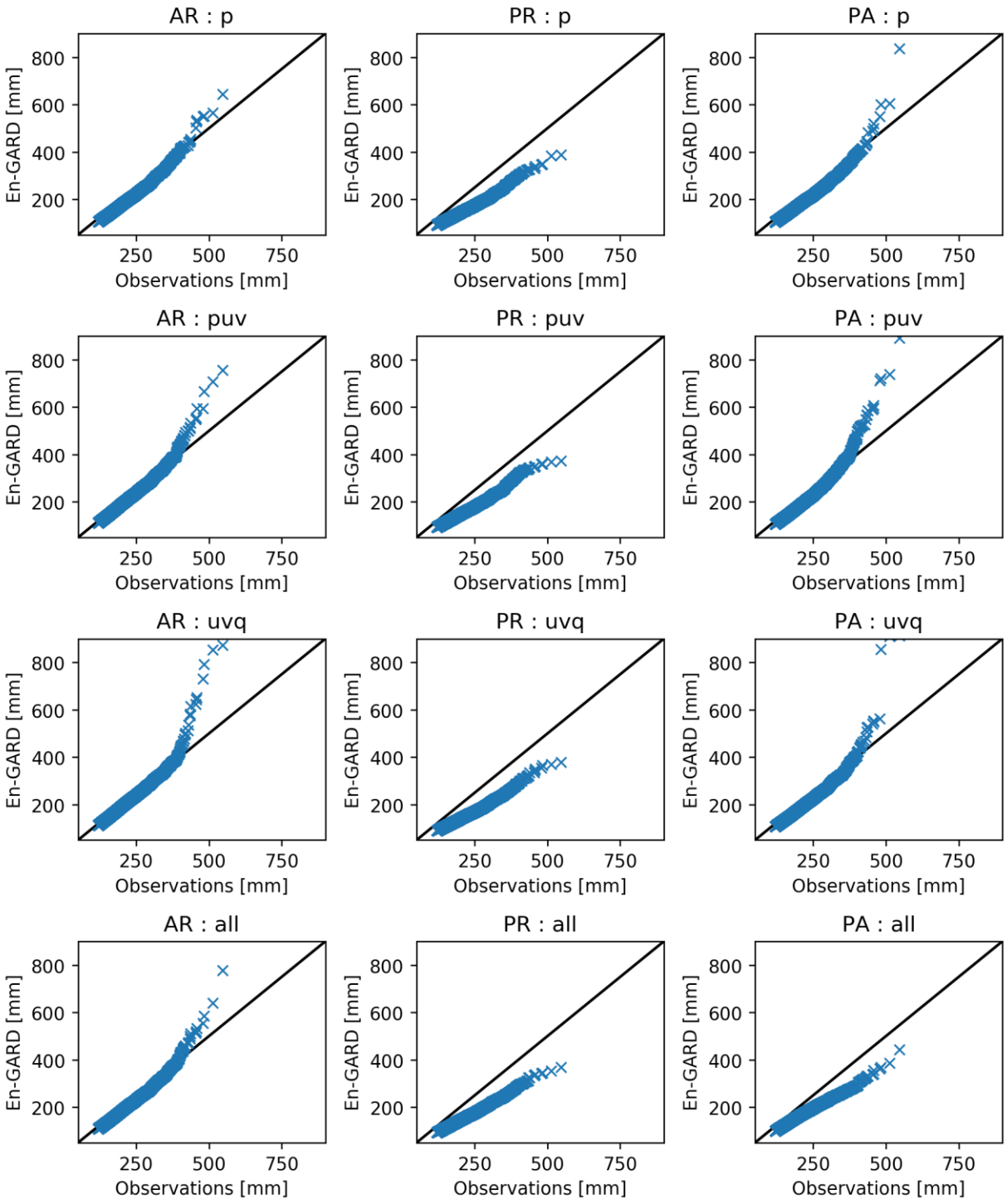

Figure S1. QQ plot of daily precipitation in observations (x-axis) and predicted by En-GARD (y-axis) for each of 12 different En-GARD instantiations noted in the title as combinations of algorithm: Analog-Regression (AR), Pure-Regression (PR), or Pure-Analog (PA); and predictor set: Precipitation (p), precipitation, U, and V (puv), specific-humidity, U, and V (uvq), or the set of nine variables discussed in this supplement (all).

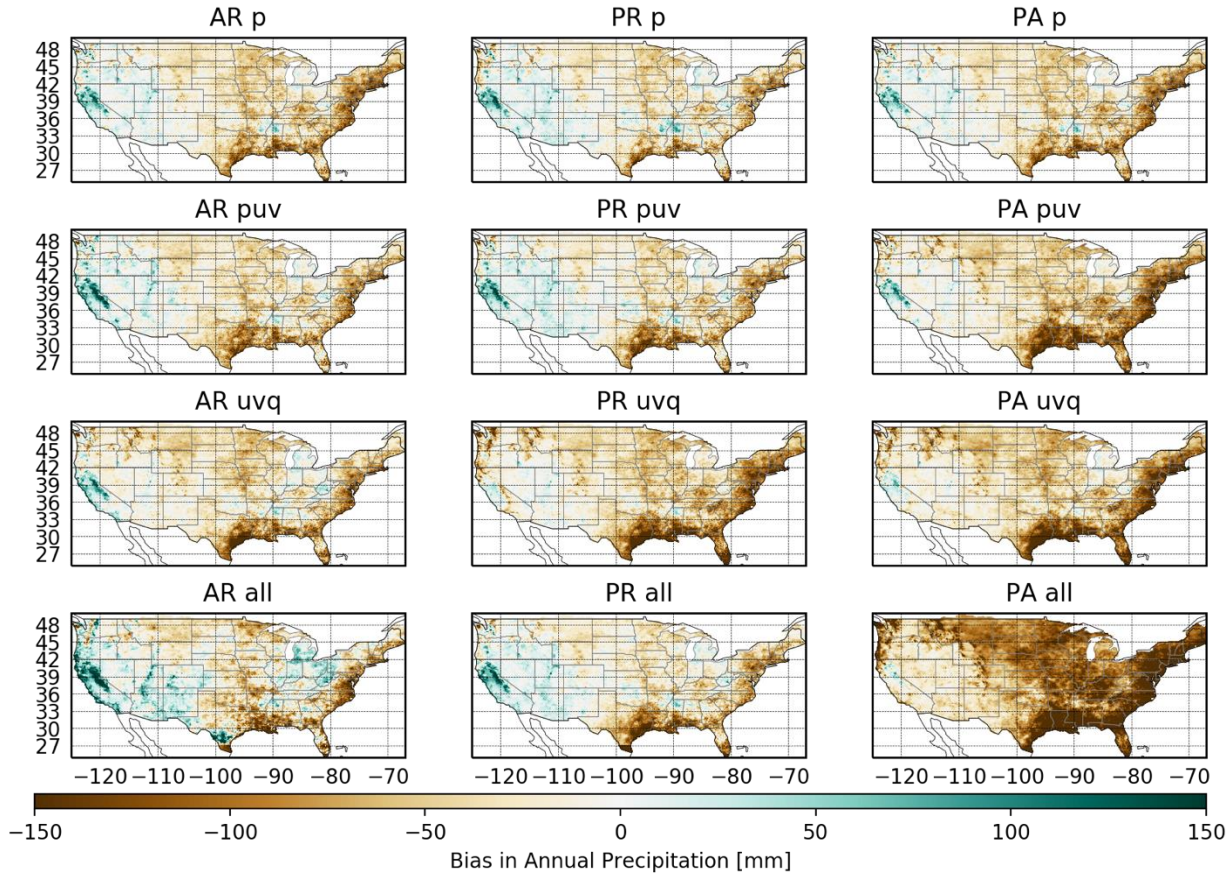

Figure S2. Bias in mean annual precipitation by algorithm (column) and predictor set (row), labels as in S1.

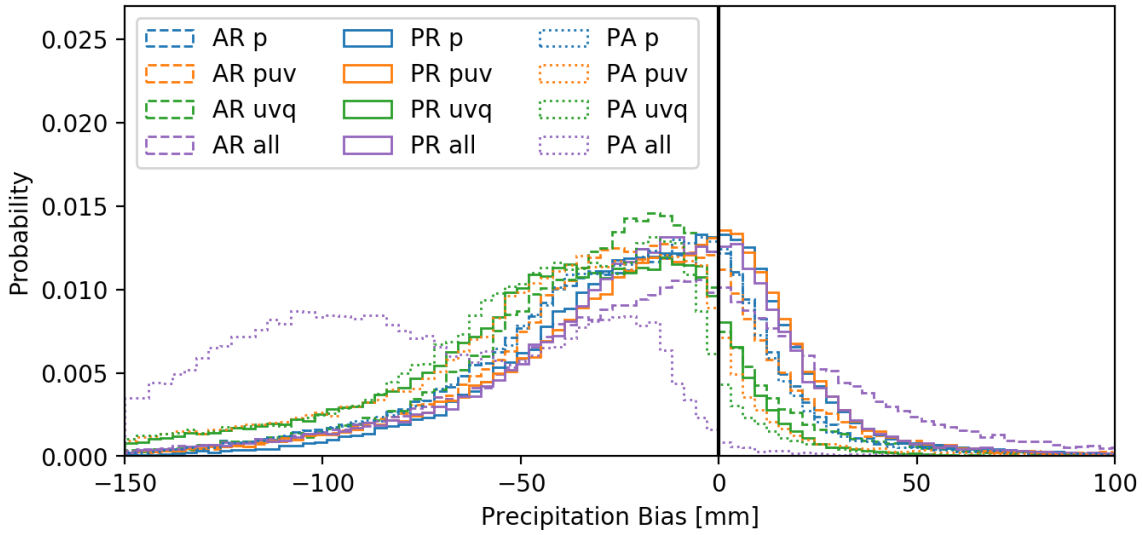

Figure S3. Histogram of grid-cell biases for all algorithms (line styles) and predictor sets (colors), labels as in S1.

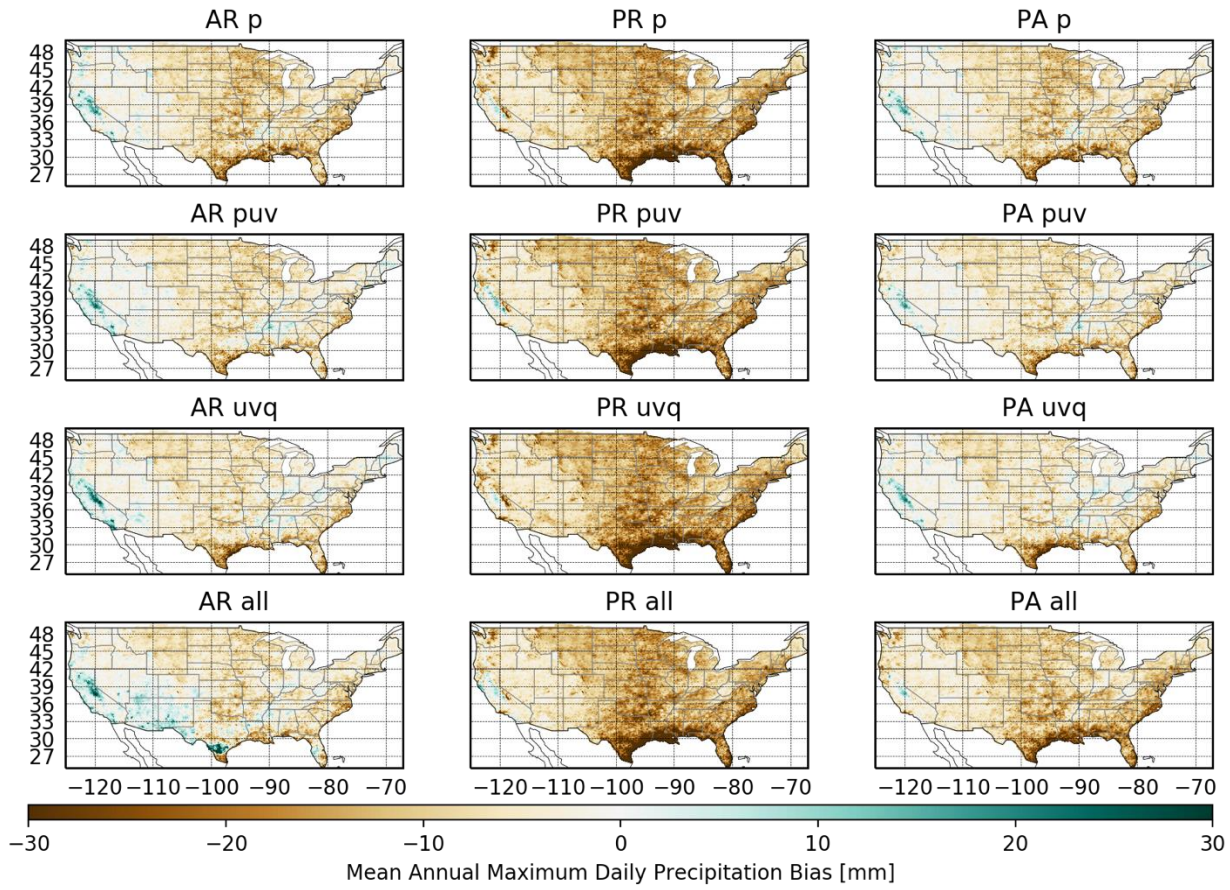

Figure S4. Bias in mean annual maximum precipitation amounts for all algorithms (columns) and predictor sets (rows), labels as in S1.

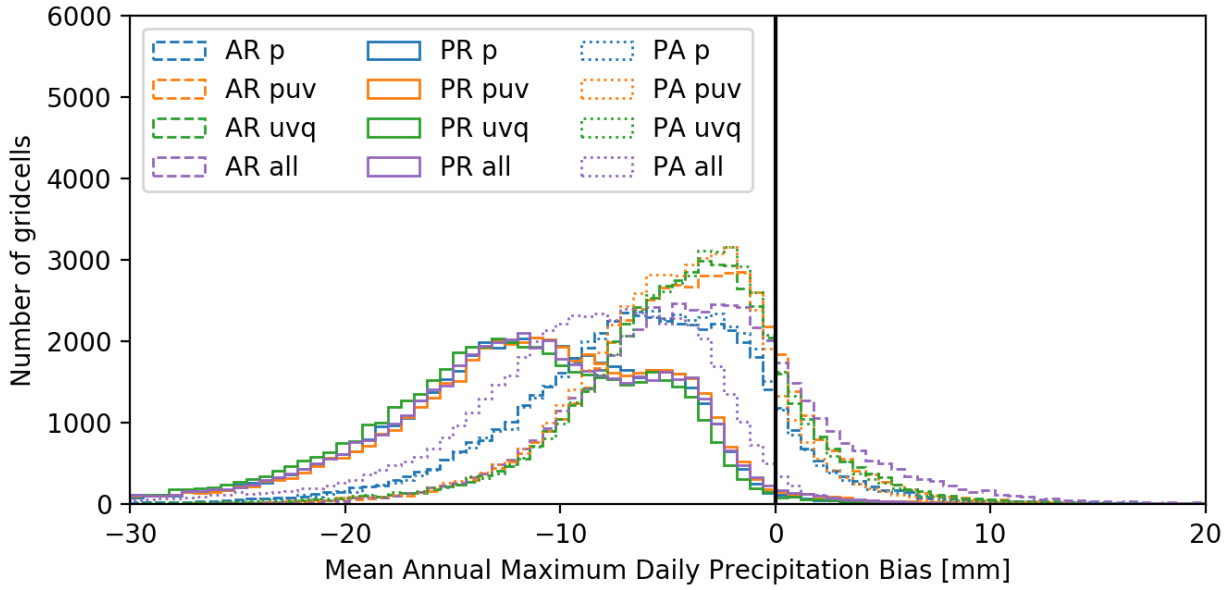

Figure S5. Histogram of grid-cell biases in mean annual maximum precipitation for all algorithms (line styles) and predictor sets (colors), labels as in S1.

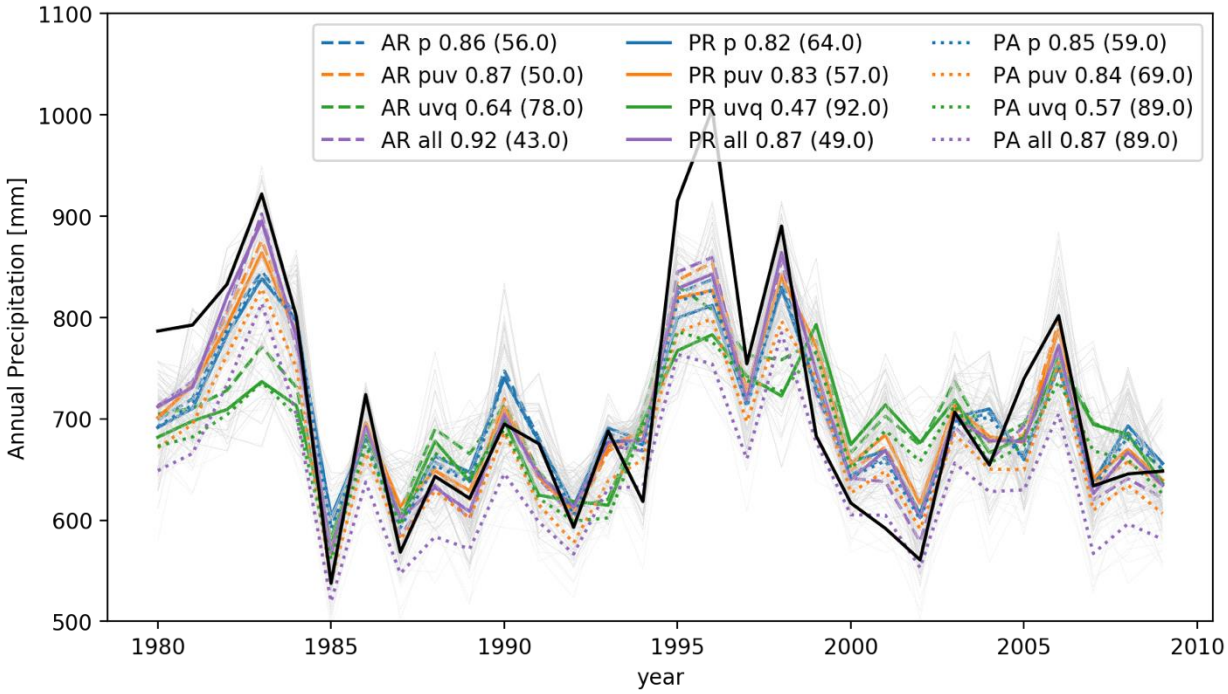

Figure S6. Year to year variability for all algorithms (line styles) and predictor sets (colors), labels as in S1, as well as observations (black). Correlation (r^2) shown in the legend along with root mean square error (in parenthesis). Ten stochastic perturbations (SCRFs) are applied for each method (thin grey lines).

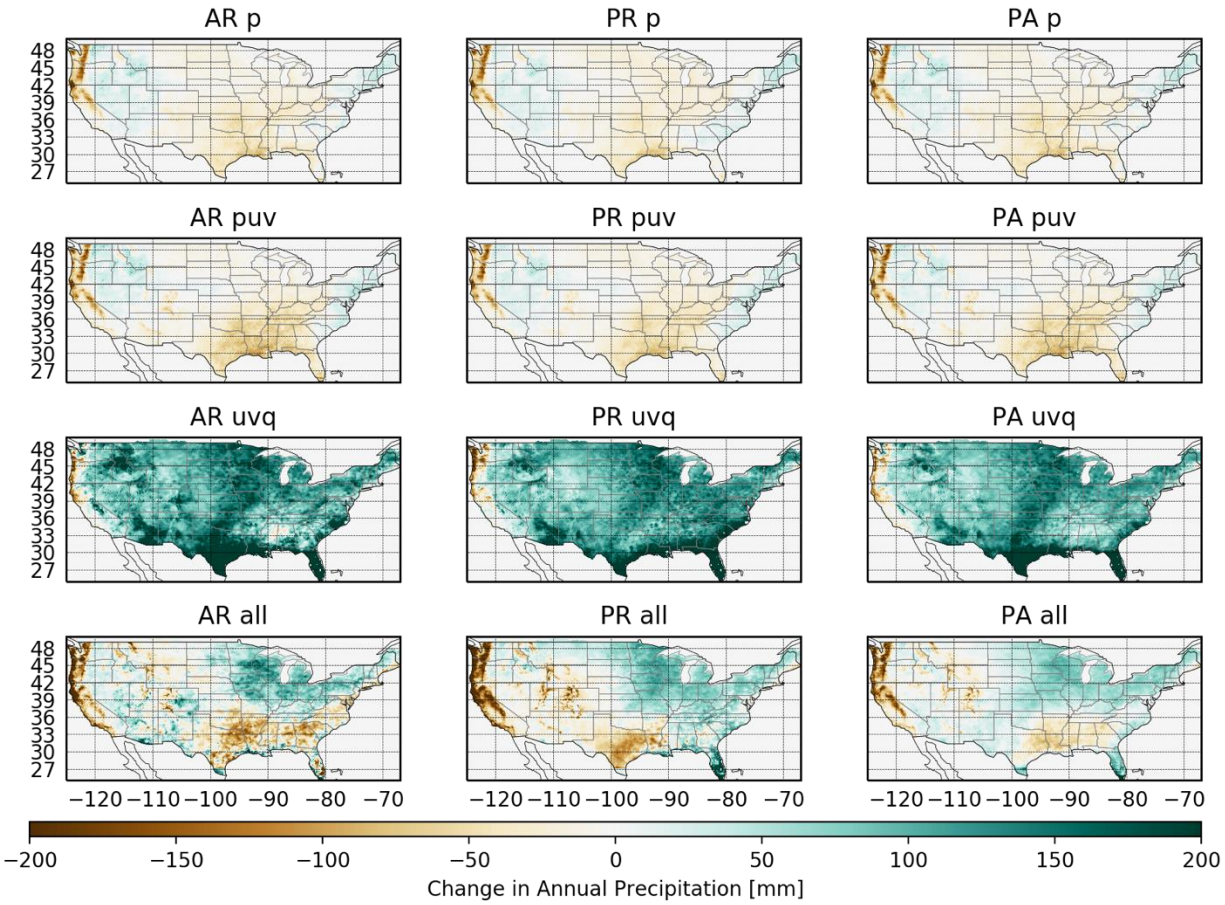

Figure S7. Change in mean annual precipitation for all algorithms (columns) and predictor sets (rows), labels as in S1.

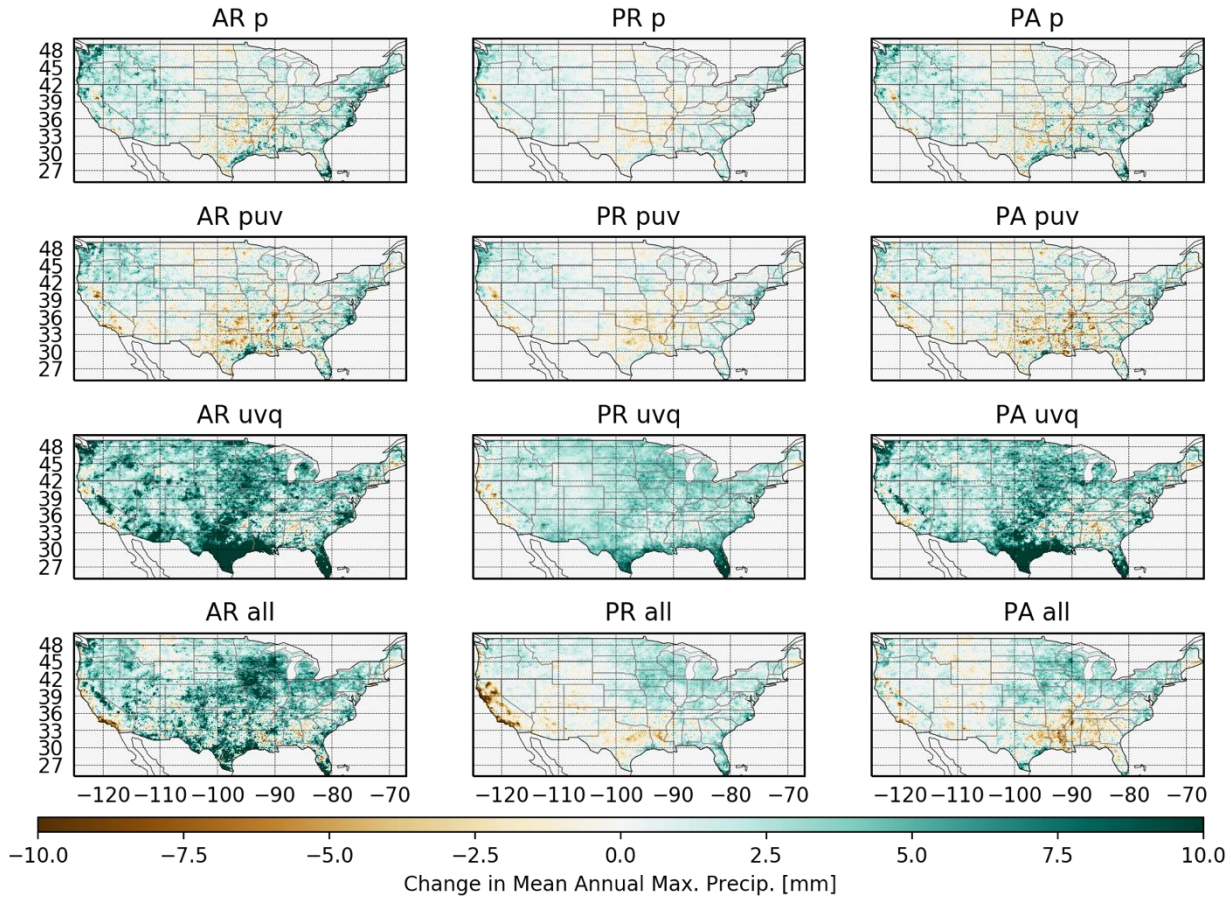

Figure S8. Change in mean annual maximum daily precipitation for all algorithms (columns) and predictor sets (rows), labels as in S1.

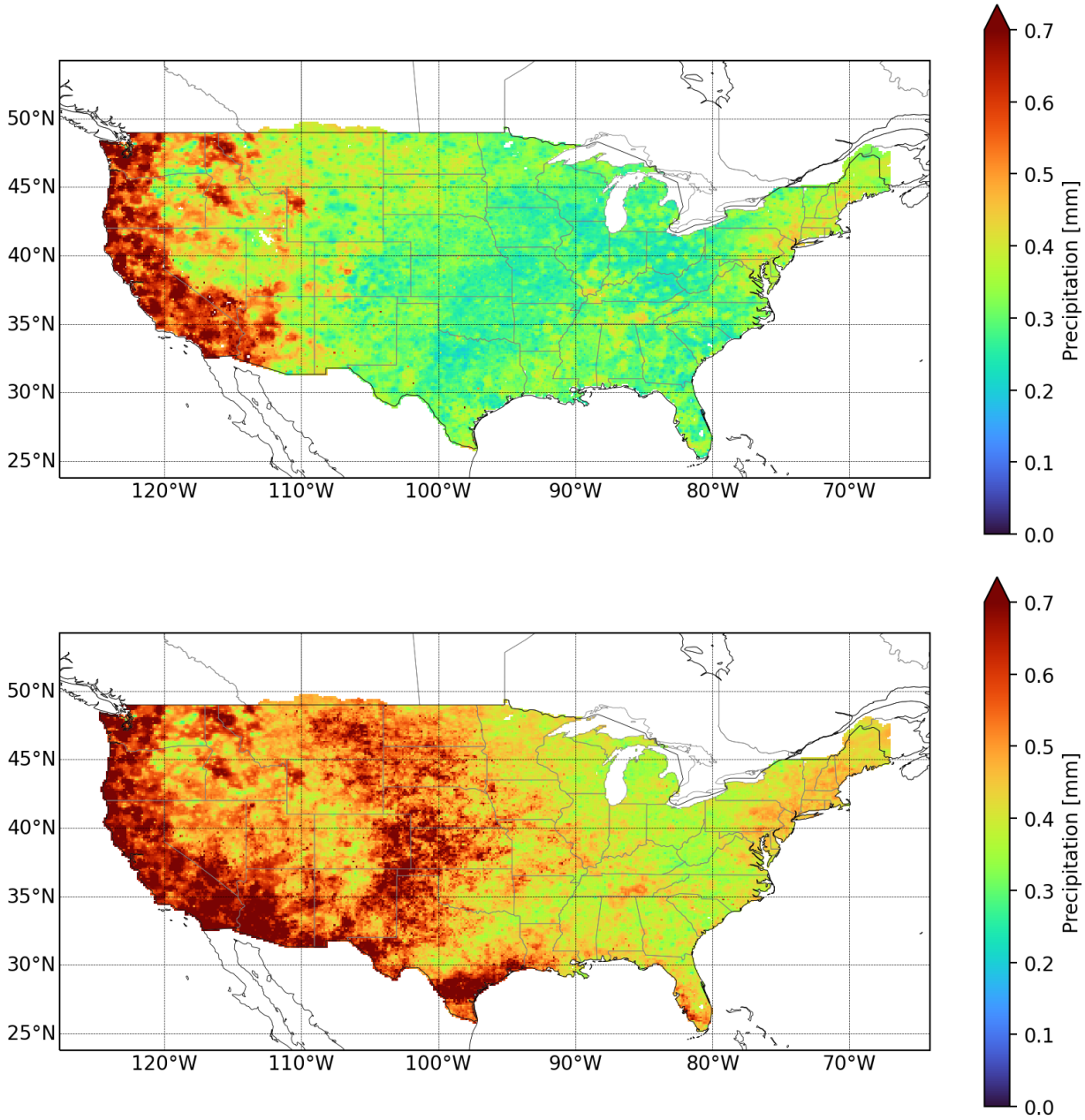

Figure S9. Fraction of total variance explained by the GARD mean prediction relative to the En-GARD prediction including stochastic term from the AR-puv projection (top) and the AR-all projection (bottom).

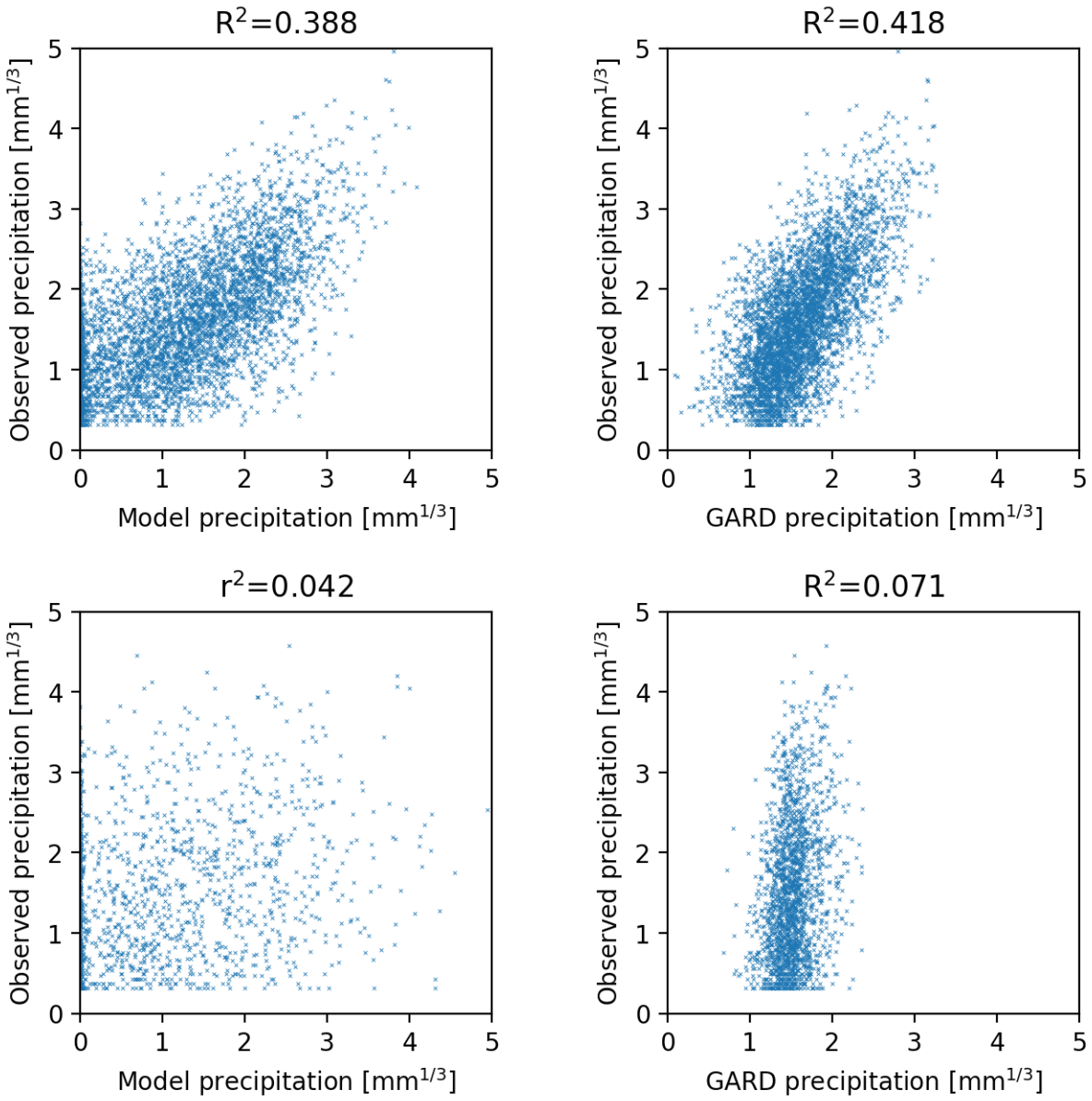

Figure S10. Scatter plots of WRF daily precipitation vs observed precipitation (left) and GARD predicted precipitation using PR-puv (right) for a grid cell located at 46.473°N -121.354°E (top) 34.820°N -99.669°E (bottom).

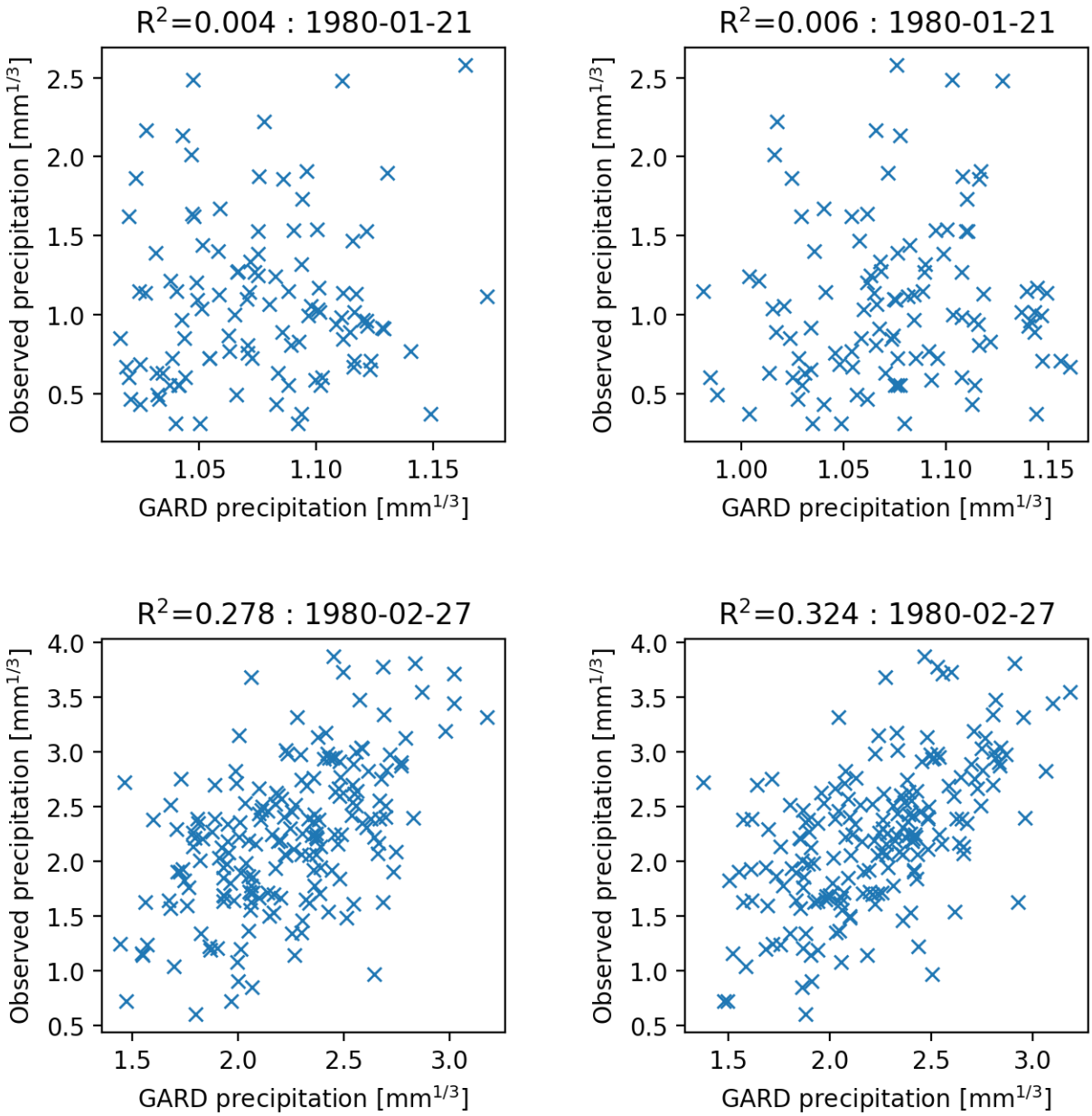

Figure S11. Scatter plots of GARD predicted precipitation vs observed precipitation when applying the AR algorithm to the puv predictor set (left) and the uvq predictor set (right) to predict precipitation on a "bad" day, 1980-01-21 (top) and on a "good" day, 1980-02-27 (bottom).