GC43F
1016

NARCCAP

IMAGe NCAR

Effects of Spatial Interpolation Algorithm Choice on Regional Climate Model Data Analysis

Seth A. McGinnis, Linda O. Mearns, and Larry McDaniel — National Center for Atmospheric Research, Boulder, CO

email: mcginnis@ucar.edu
http://www.narccap.ucar.edu

Comparative analysis of Regional Climate Model (RCM) output usually requires spatial interpolation. The question is:

DOES IT MATTER WHICH ALGORITHM YOU USE?

We regridded NCEP-driven NARCCAP output using four different algorithms with varying degrees of mathematical sophistication. Shown here are typical results from the MM5I model, which is unexceptional in its performance, with overall biases that are neither particularly good nor bad.

Biases are evaluated against two sets of observations: the half-degree climatology from Wilmott and Matsuura, *et al* at University of Delaware ("UDEL") and the NCEP Reanalysis-II data used to drive the regional models ("NCEP")

ABSTRACT

The analysis of regional climate model (RCM) outputs frequently requires spatial interpolation of the data from the model's native grid to another set of locations: a different grid is needed for comparison with other models, a set of station locations for modeling of dependent processes or comparison with raw observations, specific points of interest for impacts studies, and so on. Different interpolation algorithms will produce results with different spatial characteristics, such as smoothness, synoptic patterning, and distribution of extremes.

To explore the importance of these differences in the NARCCAP context, we regrid model output from six different RCMs driven with NCEP boundary conditions using several interpolation methods of varying mathematical sophistication: nearest-neighbor, bilinear, inverse-distance weighting, and thin-plate spline interpolation. For each algorithm, the results are compared with observations, driving data, and source model data to determine what the magnitude of the artifacts due to interpolation is and whether these effects are likely to be significant for inter-model comparison, impacts modeling and analysis, and other uses popular in the NARCCAP community.

1-D Analysis

The one-dimensional analogs of each interpolation algorithm are useful for conceptualizing their character.

Note that spline interpolation is the only method that can produce values outside the range of the inputs. (Good for peaks, bad for variables that floor at zero.)

Response Plots

Regridding simple data to higher resolution gives a sense of how the algorithm spreads values.

Note: IDW response looks anisotropic because it uses great circle distance.

0	0	0	0	0
0	**3**	0	**4**	0
0	0	0	0	0
0	**1**	0	**2**	0
0	0	0	0	0

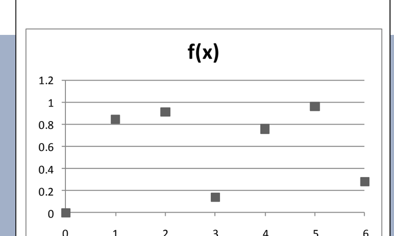
Doubleback Regridding Error

To assess error for each algorithm, we performed a round-trip or "doubleback" regridding: we interpolated the data from its native curvilinear grid to the regular lat/lon grid used for intercomparison, and then reinterpolated it back to the original grid. Subtracting the original values shows where regridding creates significant deviations in the data.

NN error is entirely a function of geometry, determined by alignment between the grids. For BL and IDW, which use weighted averaging, errors are largest where the field changes rapidly: mountainous and coastal regions for temperature, and wherever it was raining heavily that month for precipitation. TPS performs best, with very little error in either field.

RAW

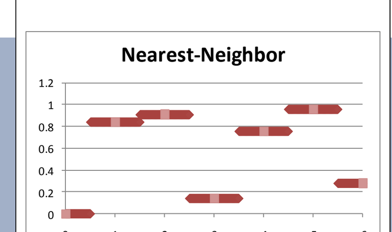
Uninterpolated data shown for comparison. Roughness matters, so temperatures are multi-year seasonal average for winter, a smooth field, while precipitation is monthly average for July, 1993, a comparatively rough field.

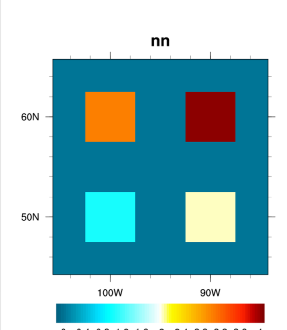


NN

Nearest Neighbor (NN) interpolation is the simplest possible method: find the closest grid point, and use that value.

Implementation: NCL function `getind_latlon2d()`

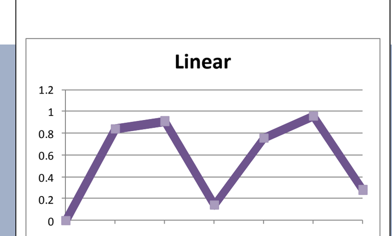


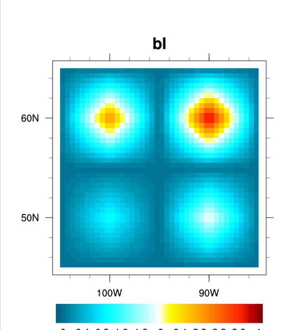


BL

Bilinear (BL) Interpolation extends linear interpolation to a 2-D grid by interpolating in two dimensions successively.

Implementation: NCL function `rcm2points(opt=2)`

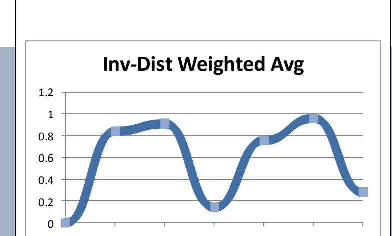


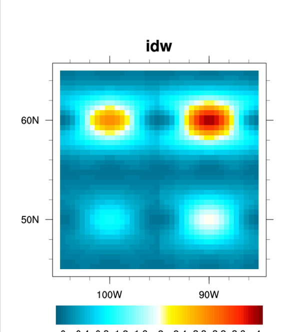


IDW

Inverse-Distance Weighting (IDW) averages neighboring values with weights proportional to the inverse of the distance from the interpolation point, squared ($1/R^2$).

Implementation: NCL function `rcm2points(opt=1)`

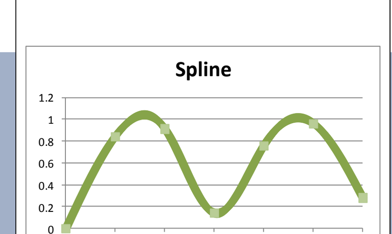


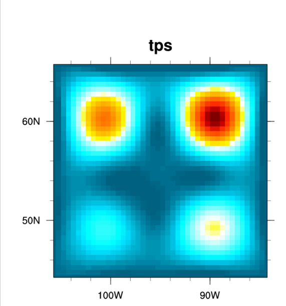


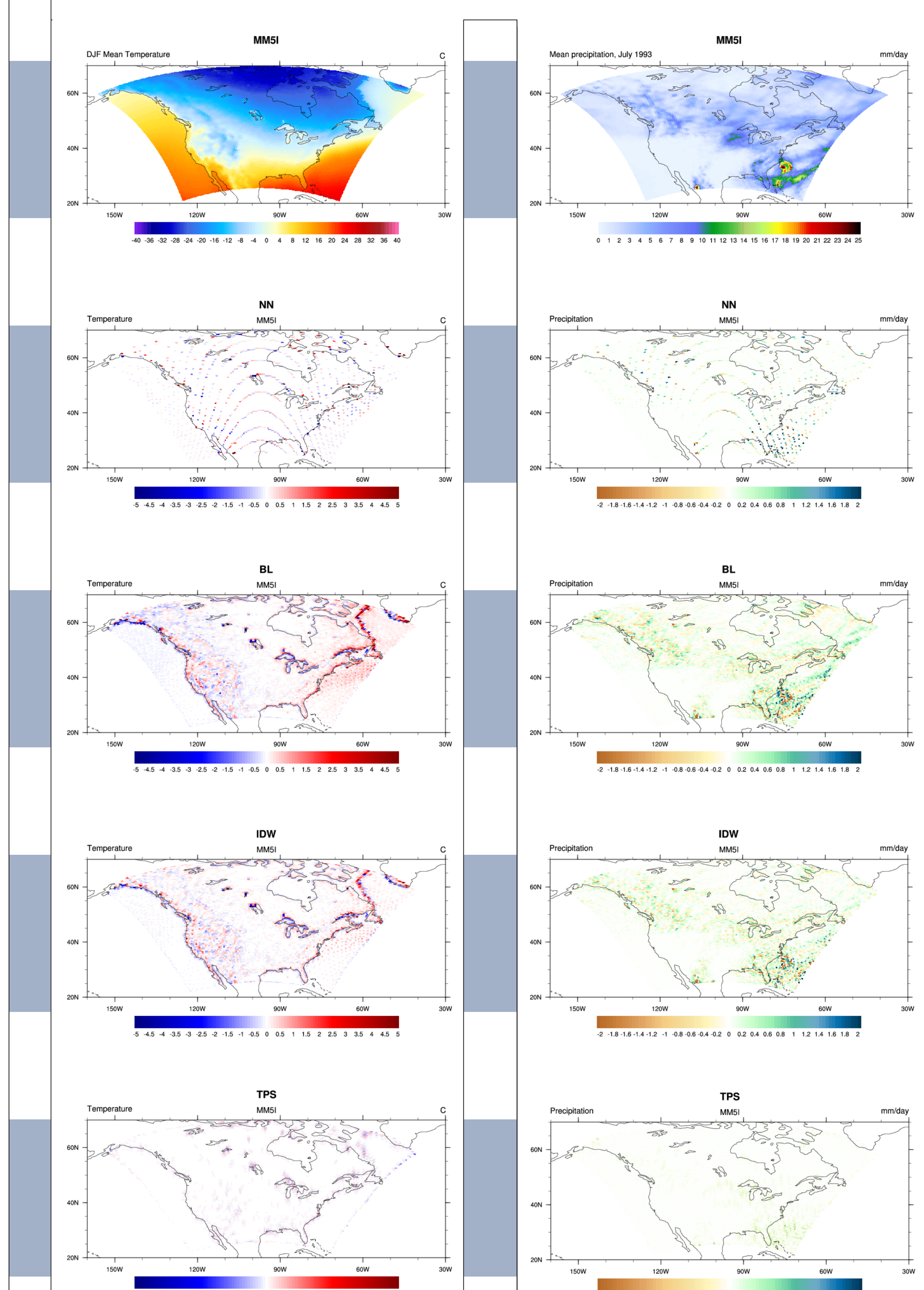
TPS

Thin-Plate Spline (TPS) interpolation uses radial basis functions analogous to the low-order piecewise polynomials used in 1-D spline interpolation.

Implementation: R function `fastTps()` in package *fields.*



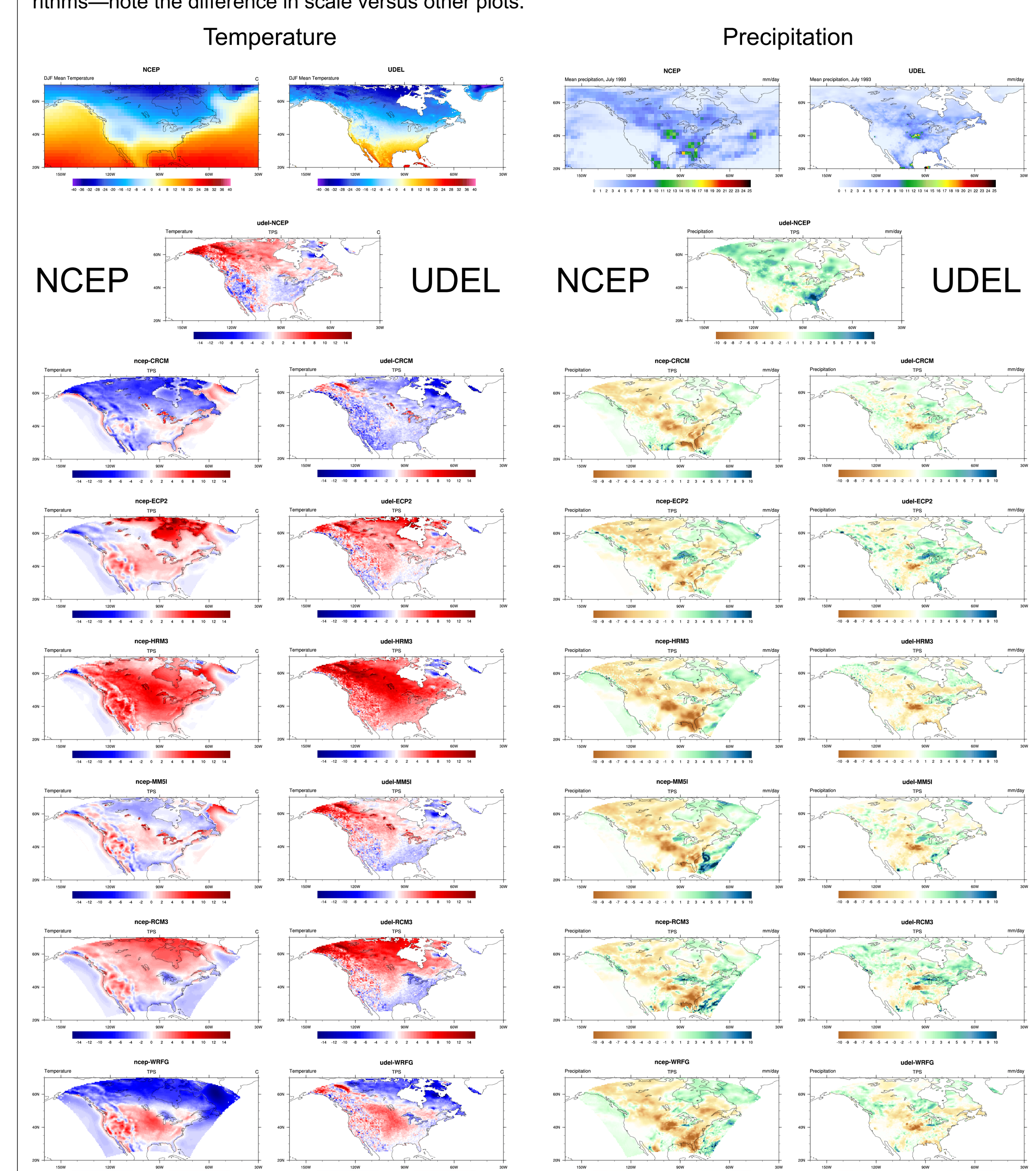




Model Bias Relative to Observations

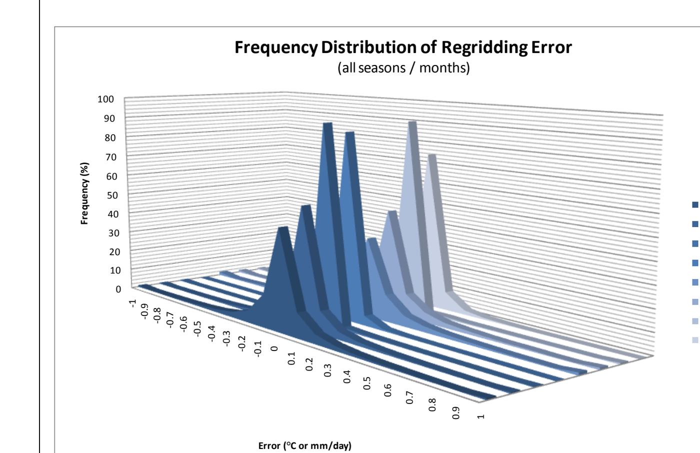
To evaluate the significance of regridding error, we need to compare it to other biases present in the model outputs. The contour maps below show temperature and precipitation biases relative to two baselines: the NCEP reanalysis data used to drive the RCMs, and the UDEL historical observations dataset.

The top row of plots shows the baseline data, and the second row shows the bias of these two baselines relative to each other. It is noteworthy that there are significant differences between the two. The following rows show bias plots against the two baselines for each of the six NARCCAP regional models. The differences from model to model are striking, and considerably larger than the differences due to interpolation algorithms—note the difference in scale versus other plots.



Distribution of Regridding Error

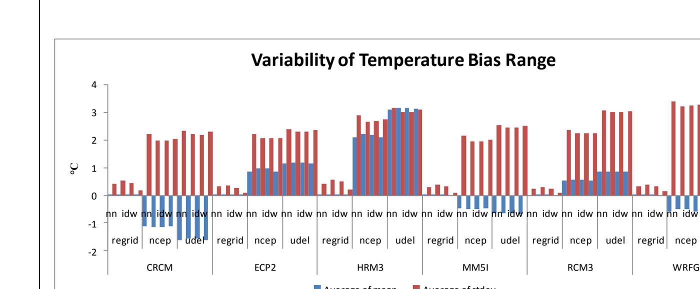
We calculated frequency distributions of the doubleback regridding error for each algorithm. Unsurprisingly, the averaging methods (BL and IDW) display considerably larger average bias because they 'leak'. The simplistic NN method performs surprisingly well by this metric, but although it is tightly clustered near zero, it has significant power far out in the tails compared to TPS.

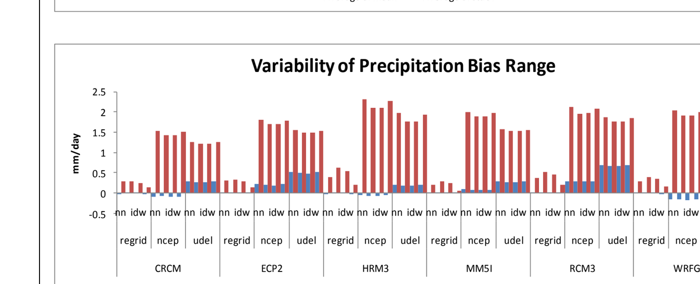


Comparative Bias Ranges

We also calculated the means and standard deviations of both the regridding (doubleback) error and the model bias relative to the NCEP and UDEL baselines to compare their ranges. As shown below, the ranges of error values due to regridding are not insignificant, but they are small compared to both the model biases relative to either baseline and to the differences between the models, no matter which algorithm is used.

Note: algorithms are sorted in the order: NN, BL, IDW, TPS. Values are calculated from data for all months / seasons in the entire 25-year simulation period.





ANSWER: IT DEPENDS (BUT MOSTLY YES)

Spatial interpolation error can be minimized with the proper choice of algorithm. Of those we studied, the Thin-Plate Spline had the best performance, while methods based on averaging had significantly worse performance. Errors are greatest when the data field is changing rapidly, as in mountainous and coastal regions. However, regridding bias is small compared to differences between models. Researchers using detailed local results should choose interpolation algorithms carefully, while those more interested in the big-picture behavior of multiple models can be less concerned. That said, just because other errors are large doesn't mean one should add to them when it's avoidable. If using a mathematically sophisticated interpolation method like the Thin-Plate Spline is feasible, in general it would be good practice to do so.

Special thanks to Doug Nychka and Steve Sain for all their aid