Reprinted from:
Preprints AMS 15th Conference on Applied Climatology, June 20-24, 2005, Savannah, GA.

**JP 1.6**

# BUILDING A GRIDDED CLIMATOLOGICAL DATASET FOR USE IN THE STATISTICAL INTERPRETATION OF NUMERICAL WEATHER PREDICTION MODELS

Rachel A. Trimarco, Kari L. Sheets, and Kathryn K. Hughes
Meteorological Development Laboratory
Office of Science and Technology
National Weather Service/NOAA
Silver Spring, Maryland

## 1. INTRODUCTION

The Meteorological Development Laboratory (MDL) of NOAA's National Weather Service (NWS) currently produces forecast guidance using the Model Output Statistics (MOS) technique (Glahn and Lowry 1972) for around 1700 sites across the United States. Recently, development of a new generation of statistical guidance has begun for grids with the resolution of the National Digital Forecast Database (NDFD) (Glahn and Ruth 2003). These grids will become part of the National Digital Guidance Database (NDGD), which will be used in the NWS Weather Forecast Office (WFO) forecast process.

The gridded MOS guidance is produced by performing a sophisticated objective analysis of the traditional MOS station forecasts. The first guess forecast for this analysis is a MOS forecast valid at the NDFD 5-km gridpoints, which is created by a generalized operator regression equation. Observations used in the forecast process are neither available every 5 km, nor are they uniformly distributed. Therefore, a gridded climatological dataset is needed to supplement the available coarser resolution observational data and to provide greater detail.

Grids containing monthly normal maximum and minimum temperatures were created by using Geographic Information Systems (GIS) techniques to combine data from multiple sources and format the data to match the appropriate grid characteristics. The gridded normals are 30-year climatologies obtained by merging Oregon State University's Parameter-elevation Regressions on Independent Slopes Model (PRISM) data with the National Center for Atmospheric Research's (NCAR) International Comprehensive Ocean - Atmosphere Data Set (ICOADS). The PRISM data were based on observations from 1971-2000 and covered the conterminous United States (CONUS). The ICOADS data were valid for the same time period, but covered the oceans and the Great Lakes. The merged GIS data were then used to create normal temperatures every fifth day by applying a cubic spline interpolation algorithm to fit the monthly normals and then interpolate to the appropriate day. These gridded datasets will be used as predictors to develop gridded MOS forecast guidance for temperature, which is derived from the National Centers for Environmental Prediction's (NCEP) Global Forecast System (GFS) numerical weather prediction model (Dallavalle et al. 2004). This paper focuses on the techniques used to generate the climatological datasets, as well as their importance to the gridded MOS system.

**Corresponding author address:*
Rachel A. Trimarco, 1325 East-West Highway, Station 11303, Silver Spring, MD 20910-3283; e-mail:Rachel.Trimarco@noaa.gov

## 2. THE PRISM DATA

The monthly normal maximum and minimum temperatures for the CONUS were obtained from the Spatial Climate Analysis Service at Oregon State University. The PRISM system that was used to generate the normals is a knowledge-based interpolation system. PRISM uses a linear climate-elevation regression model that takes into account such factors as distance between stations, elevation, and coastal proximity (Daly and Johnson 1999). The observations used in the PRISM analysis were obtained from reports at approximately 8000 sites. These sites were part of multiple networks. The networks included NWS cooperative observer stations (COOP), National Resources Conservation Service (NRCS) Snow Telemetry (SNOTEL) sites, NRCS agricultural climate stations for southeastern Washington, Midwest Climate Data Center data, Historical Climate Network data, and miscellaneous sites, including storage gauges, snow courses, and manually estimated points (Gibson et al. 2002). The original PRISM dataset was based on a 103-year observational dataset covering the years

1895-1997. However, the data we obtained for the purposes of our project only spanned the years 1971-2000, though this PRISM climatology was created by using the same techniques as the 103-year dataset. The data were obtained in a gridded format available as Environmental Systems Research Institute's (ESRI) ArcInfo ASCII grids. This format allowed for easy import into ESRI's ArcGIS software system. The grids have a 2.5-min (4-km) resolution with the units of temperature in degrees Celsius times 100.

The monthly maximum and minimum temperature normals for each month were downloaded, imported into the ArcGIS system, and were then converted from ASCII grids to raster layers. During the conversion, these raster layers were resampled to the geographic extent of the NDFD grid, but with a cell size and geographic projection that are compatible with the GFS model. The data were converted to a grid with the NDFD extent but with a GFS compatible cell size and projection because the gridded MOS is only being developed to comply with the geographic domain of the NDFD. The normals were resampled by using a nearest neighbor interpolation, from the 4-km PRISM grid to a 5.953-km grid with a North Polar Stereographic projection. Figure 1 shows an example of the PRISM data after the data had been resampled with the new grid characteristics.

## 3. THE ICOADS DATA

Normals over water bodies are not included in the PRISM datasets, so supplementary data were obtained from the ICOADS. The ICOADS dataset is provided to the public by the three-way cooperation among NOAA's Climate Diagnostics Center (CDC), NOAA's National Climatic Data Center (NCDC), and the National Science Foundation (NSF) National Center for Atmospheric Research (NCAR). ICOADS data include global marine

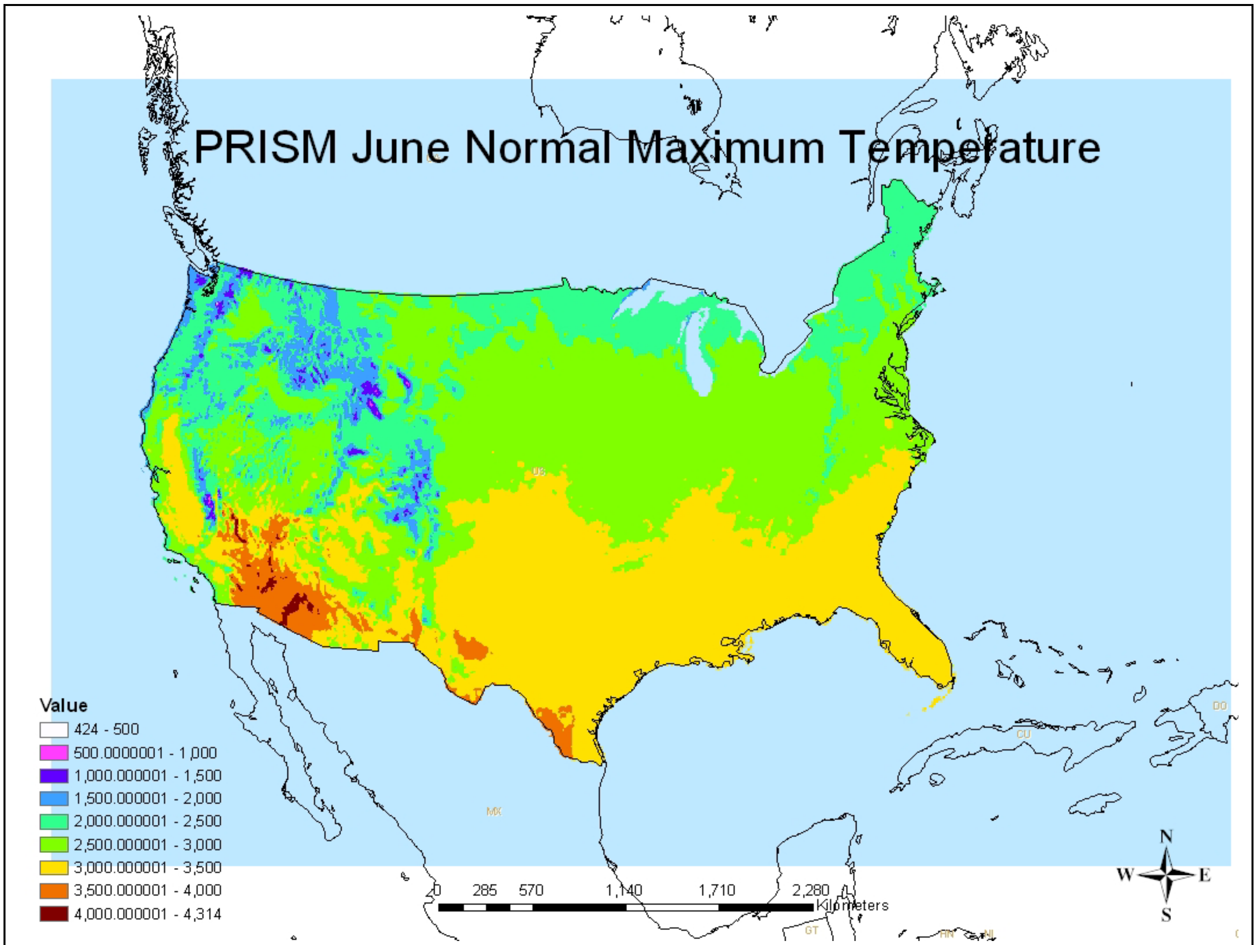


**Figure 1.** The normal maximum temperature for June from the PRISM data. The grid has been resampled to the geographic extent of the NDFD with a cell size of 5.953 km. The units are degrees Celsius times 100.

meteorological observations that come from a variety of sources such as merchant, research, fishing, and naval vessels, as well as moored and drifting buoys (Worley et al. 2004). The ICOADS dataset is comprised of various meteorological parameters and spans a time period from 1784-2002. For the purposes of this project, we obtained an air temperature dataset that provided mean temperatures by month, for the years 1960-2002, with a grid resolution of 1-degree latitude and longitude. An example of the raw ICOADS data is shown in Fig. 2. Although we had data from 1960-2002, we only used the years 1971-2000 so that we would be consistent with the PRISM data. It is important to note that records exist only for months and locations that contain at least one observation. Therefore, some locations do not contain records for every month or year. In addition, there were no data available for October, November, or December 1973.

The format of the ICOADS data differed from the format of the PRISM data, resulting in a significant amount of preprocessing before the ICOADS could be imported into ArcGIS. First, the data were separated by month, creating 30 text files for each month (29 text files for October, November, and December). Then, the sign of the longitude in each record had to be changed to conform to GIS standards. Finally, the mean temperatures were converted from degrees Celsius to degrees Celsius times 100 to be consistent with the PRISM data. At this point, in order to maintain a certain degree of organization with such a large sample of data, separate ArcGIS ArcMap projects were created for each of the 12 months. The text

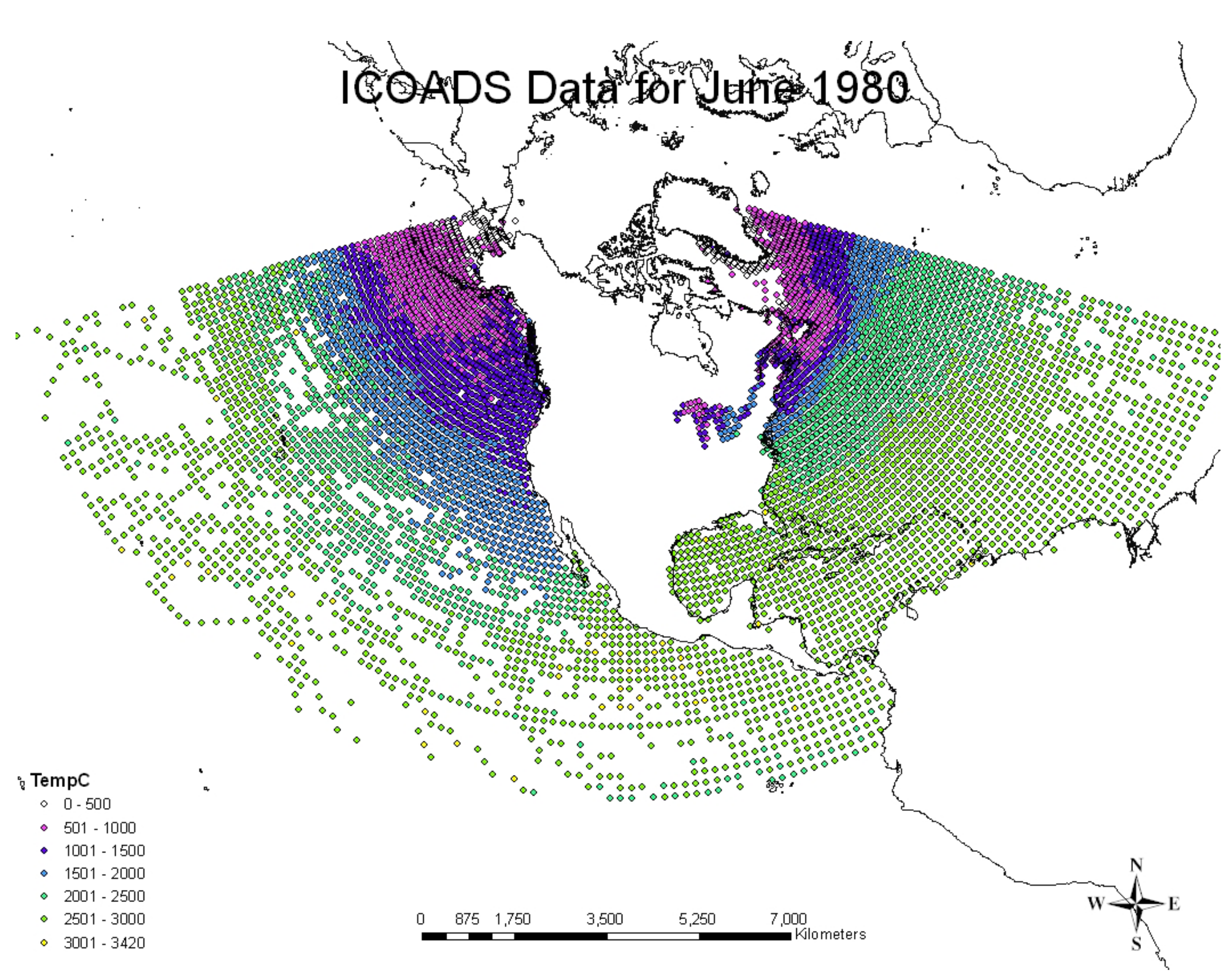


**Figure 2.** The June 1980 monthly temperature average, in degrees Celsius times 100, from the ICOADS data. Data points are every 1-degree falling on lines of whole degrees latitude and longitude. This example shows that data were not always available in every grid box.

files containing the monthly ICOADS data were then imported into ArcGIS where the data could be mapped and analyzed.

## 4. MERGING PRISM AND ICOADS DATA

In order to merge the ICOADS data with the PRISM data, further preprocessing had to be done after the data were imported into ArcGIS. First, a kriging interpolation was used to convert the ICOADS point data to a raster grid with the same cell size and extent as the already resampled PRISM grids. This was done for each month of each year. The ArcGIS Spatial Analyst cell statistics function was then used to average together the 30 years, thus creating one mean temperature for each month. Next, because the interpolation method produced a floating point gridded dataset, though the original values were integers, the ArcGIS Spatial Analyst extension was once again utilized to perform a raster calculation that rounded the decimal values to create an integer grid. At this point, both the ICOADS data and the PRISM data were in the same format and nearly ready to be merged.

As a result of the interpolation applied to the ICOADS data, not only was the area covering the oceans and Great Lakes converted to a grid, but the data were also interpolated across continental North America. The area covering the CONUS needed to be erased before merging could take place so that the PRISM values covering that area would not be affected when the two grids were combined. In order for this to be done, both the ICOADS and PRISM grids for each month were converted to vector shapefiles. The ArcGIS erase function was then used to cut out an area from the ICOADS data based on the area that was covered by the PRISM data. This created a data-void hole that was the shape of the CONUS in the middle of the ICOADS data. The ICOADS datasets for each month were then converted back to grids. In order to add the two grids, the null data values in both the PRISM and ICOADS grids were changed from -9999 to zeros by using a Spatial Analyst raster calculator function. The raster calculator was then used one final time to add the values of both grids together and resample their extent to that of the GFS model, which is a larger domain than the previous NDFD extent. Fig. 3 shows the final grid.

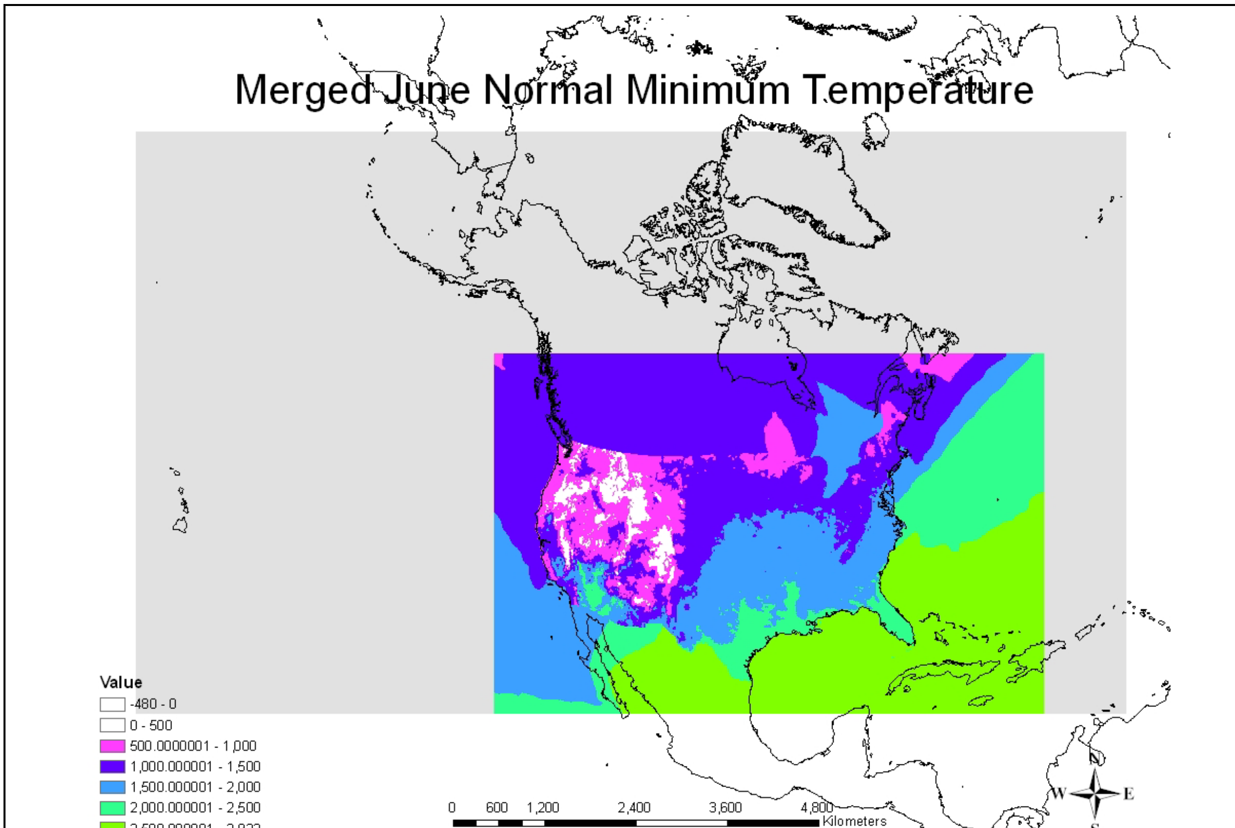


**Figure 3.** This image shows the final merged June normal minimum temperatures in degrees Celsius times 100. The area in grey is the geographical extent of the GFS with null data values of -9999.

Although the ICOADS data were available only as monthly mean temperatures, and the PRISM data contained monthly maximum and minimum temperature normals, we decided that both datasets could still be merged. Each monthly mean temperature set provided by the ICOADS data was merged with both the maximum and minimum temperature normals provided by the PRISM data. Because data over the oceans are often sparse, and because the diurnal temperature variation over the ocean is small, we felt this was the best available compromise in order to create a dataset that would cover the entire area of interest.

## 5. CREATING FIVE-DAY TEMPERATURE NORMALS FROM THE MONTHLY DATA

After the monthly maximum and minimum temperature normal grids were generated, they were used to create temperature normals valid every fifth day instead of just once a month. The monthly normals were output from ArcGIS as ASCII grids that were ingested into a specially designed software package that could perform more complicated calculations than the GIS. The software employed a cubic spline interpolation algorithm. This algorithm computes and evaluates cubic spline interpolation polynomials for a given set of data points. In this case, the data points are the monthly temperature values. Different cubic splines are available for every consecutive pair of data points, or months (Burden and Faires 1989). Hence, for a specific day, the appropriate polynomial can be used to interpolate the monthly normal to that day. Different splines are available for each grid point. From this method, maximum and minimum temperature normals were computed for every fifth day of the year starting at day 5, which is January 5, through day 365. The normals were computed every fifth day, rather than every day, for efficiency in the MOS system.

## 6. FUTURE WORK

Five-day temperature normals were created so they could be used as potential predictors in the MOS forecast system, particularly for the enhancement of the gridded MOS guidance. The gridded temperature normals will be interpolated in both time and space to match the dates and stations (or grid points) at which guidance is desired. The normals will then be offered as predictors in the multiple regression analysis. Once equations have been developed from the observations at the stations, these equations will be applied to points that are evenly spaced every 5 km, according to the specifications of the NDFD grid. Maximum and minimum temperature forecasts will be generated, and then the point data will be converted to a gridded format that will become the first guess field necessary for analysis. Additional experimentation will be needed to determine the best method for using this valuable dataset to improve the gridded MOS guidance.

## 7. CONCLUSIONS

Until now, gridded temperature climatologies did not exist for use in the MOS system, although other elements, such as the thunderstorm probabilities, have long relied on high resolution lightning relative frequencies as a predictor (Hughes and Trimarco 2004). By combining data from multiple sources, we hope to improve the gridded temperature guidance. This project also exemplified how useful GIS tools are for analyzing and manipulating spatial data. Without the GIS tools, we would not have so easily been able to create monthly temperature normals in a usable format.

## 8. ACKNOWLEDGMENTS

PRISM data are provided by The Spatial Climate Analysis Service and the Oregon Climate Service at Oregon State University. They are currently provided as a free public service and can be obtained online at http://www.ocs.orst.edu/prism/. ICOADS data are provided by the NOAA-CIRES Climate Diagnostics Center. They are also provided as a free public service and can be obtained online at http://www.cdc.noaa.gov/coads/.

## 9. REFERENCES

Burden, R. L., and J. D. Faires, 1989. Numerical Analysis Fourth Edition. PWS-KENT Publishing Company, Boston, MA.

Dallavalle, J. P., M. C. Erickson, and J. C. Maloney, 2004: Model Output Statistics (MOS) guidance for short-range projections. *Preprints 20th Conference on Weather Analysis and Forecasting,* Seattle, WA, Amer. Meteor. Soc., CD-ROM, P6.1.

Daly, C., and G. Johnson, 1999. PRISM Guide Book. Climate Mapping with PRISM: PRISM Spatial Climate Layers. [Available online at: http://www.ocs.orst.edu/pub/prism/docs/prisguid.pdf.]

Gibson, W., C. Daly, T. Kittel, D. Nychka, C. Johns, N. Rosenbloom, A. McNab, and G. Taylor, 2002: Development of a 103-year high-resolution climate data set for the conterminous United States. *Proceedings*, *13th Conf. on Applied Climatology,* Portland, OR, Amer. Meteor. Soc., 181-183.

Glahn, H. R., and D. A. Lowry, 1972: The use of Model Output Statistics (MOS) in objective weather forecasting. *J. Appl. Meteor.,* **11,** 1203-1211.

_____, and D. P. Ruth, 2003: The new digital forecast database of the National Weather Service. *Bull. Amer. Meteor. Soc.,* **84**, 195-201.

Hughes, K. K., and R. A. Trimarco, 2004: Impacts of resolution on gridded probability thunderstorm forecast guidance. *Preprints 20th Conference on Weather Analysis and Forecasting/16th Conference on Numerical Weather Prediction,* Seattle, WA, Amer. Meteor. Soc., CD-ROM, J6.5.

Worley, S. J., S. D. Woodruff, R. W. Reynolds, S. J. Lubker, and N. Lott, 2004: ICOADS Release 2.1 Data and Products. Submitted to *International Journal of Climatology*, CLIMAR-II special issue. [Available online at: http://www.cdc.noaa.gov/coads/icoads-ijc.pdf.]