Combining Deterministic and Probabilistic Methods to Produce Gridded Climatologies

Michael Squires
Alan McNab
National Climatic Data Center (NCDC - NOAA)
Asheville, NC

Abstract

There are nearly 8,000 sites in the United States for which detailed climatologies of various atmospheric variables have been produced. However, spatial interpolation of these point climatologies via traditional techniques (IDW, kriging, etc.) is often not accurate, especially in mountainous regions, coastal zones, or data sparse areas. This paper discusses NCDC's preliminary efforts to combine deterministic and probabilistic methods to produce improved gridded climatologies. Two methods to combine multiple linear regression information with deterministic climatologies produced by the Advanced Climate Modeling and Environmental Simulation (ACMES) modeling system are described. GIS is used to visualize intermediate and final grids during the development of these techniques.

Background

Climatological information is traditionally created at point locations where systematic weather observations are recorded over a number of years. These observations are taken at airports and various other sites. The data at airports are collected hourly and include information about temperature, precipitation, clouds, winds and visibility. Other sites usually only have temperature and precipitation observations collected on a daily basis. These time series of meteorological data are summarized into climatologies that are used by various sectors of the economy (agriculture, energy, business, etc.) for decision making and planning activities.

While beneficial for these locations, the climate information is often needed at locations between observational sites. It is possible to create continuous surfaces from these point data by distributing the climate information everywhere in a region using various spatial interpolation techniques such as inverse distance weighting (IDW) and ordinary kriging (OK). However these methods may yield misleading results in mountainous areas, coastal zones, and data sparse regions.

In addition, problems related to reporting practices often occur at locations where meteorological observations are available. The period over which the observations have been recorded may be short or there may be missing data. Some sites have hourly data but most sites have only daily data with a reduced number of variables (usually temperature and precipitation). Since the hourly observations taken at airports are primarily designed to support aviation and day to day forecasting, information useful for climatological applications is not always measured. For example, cloud heights above 12,000 feet are not measured with the current generation of automated observing equipment. Therefore, changes in station location, instrumentation, and reporting practices all contribute to the difficulty of producing robust station climatologies.

The use of models to produce climate information on a grid addresses both the interpolation and data error problems. Although no model is perfect, the choice of a model should minimize the data problems and maximize the amount of useful information that can be mapped to a grid. In general there are two basic types of models that can be used in this context. One can either build a *probabilistic* or a *deterministic* model.

Probabilistic models have demonstrated an ability to create accurate, high-resolution precipitation and temperature gridded climatologies in complex terrain (Carrega, 1995; Daly et al., 1994). Models of this type use statistical relationships between a particular meteorological element and other predictor information that is available (elevation, aspect, etc.). It is difficult, however, to create climatologies for some meteorological variables. The vast majority of observation sites collect information only once per day and do not include elements such as winds, clouds, absolute humidity, and visibility. Also, regression models must be developed separately for each predictand. Since models for each predictand are developed separately, there may be inconsistencies between the predictions of different variables. Therefore, care must be taken to ensure predictions of different variables are physically consistent with each other.

If there is sufficient knowledge of the phenomena being estimated, it is possible to use a deterministic model to estimate information where there is no observed data (Isaacs and Srivastava, 1989). Advances in numerical weather prediction (NWP) and computer hardware have now made the deterministic approach possible (Cox, 1998). NWP uses three-dimensional point observations of the atmosphere (surface and aloft) as inputs into models which use the equations of motion and thermodynamics to describe the atmosphere. These equations are integrated over time to produce forecasts on spatially continuous grids. In addition to prediction over time, NWP models provide the additional benefit of interpolating point data over space. A long time series of 'forecasts' for past time periods can be used to produce spatially continuous climatologies.

There are several advantages to using an NWP approach to develop climatologies. NWP output contains many other elements in addition to temperature and precipitation such as winds, clouds, absolute humidity, and visibility. NWP wind, temperature, and moisture are dynamically and thermodynamically consistent over the output grid. The information is output at many levels in the vertical. NWP models can be run in parts of the world that have few observational sites. Finally, NWP explicitly considers the effects of complex terrain and its influence on the mesoscale environment.

However, there are several problems with NWP forecasts (Wilks, 1995). First, although NWP has made considerable progress over the last three decades, the models must severely simplify the atmosphere and the processes that affect it. Figure 1 is an example of the manner in which an NWP model can oversimplify the environment. The image shows a semi-transparent 40 km terrain grid draped over a 1 km terrain grid in western North Carolina. Although the terrain grid was not produced by the NWP, it is used by the model to describe the environment. The 40 km grid resolution is typical of what is found in NWP models. The 1 km grid shows the extent to which a single 40 km average value simplifies the 'real' elevation variations within an individual 40 km cell. Secondly, NWP models often exhibit some type of systematic error. That is, a model may have a bias not apparent in an individual forecast, but detectable in the long term error statistics. Also, NWP forecasts give no information about uncertainty. The model makes specific forecasts for gridpoints, and there is no indication about possible error or confidence level. Finally, high-resolution (10 km and less) NWP forecasts over climatological periods (10-30 years) are computationally intensive and require extremely long run times.

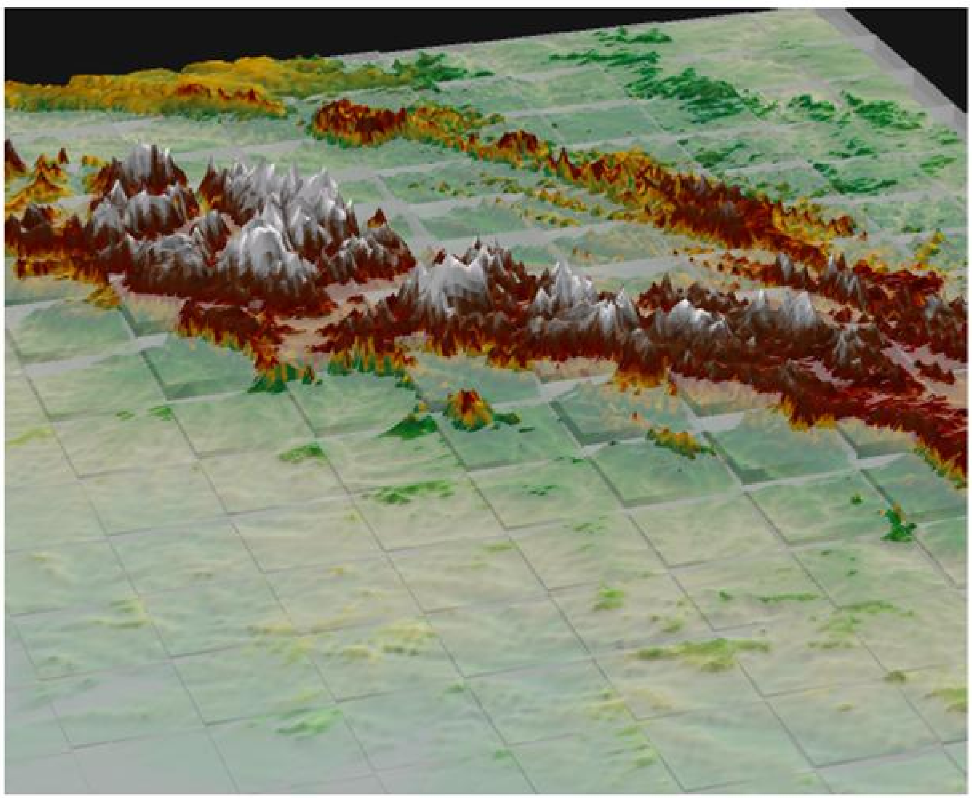

Figure 1. Three dimensional comparison of a 40 km terrain grid (semi-transparent) with a 1 km terrain grid over western North Carolina.

The purpose of this paper is to show some of the preliminary efforts at the National Climatic Data Center to produce gridded climatologies of temperature, precipitation, and other meteorological variables using output from NWP models and then refining that output with various statistical methods. The intent is to use the strength of each method and minimize weak points of each method.

Advanced Climate Modeling and Environmental Simulation (ACMES)

The Air Force Combat Climatology Center (AFCCC), which is collocated with NCDC, has used an NWP approach to generate gridded climatologies over the last several years. The ability to generate climatologies for many different elements at various levels throughout the atmosphere (surface, 1000 ft., 2000 ft., etc.) for any location in the world is the primary reason AFCCC chose to use an NWP model to produce its high-resolution gridded climatologies. The Advanced Climate Modeling and Environmental Simulation (ACMES) uses observational data as inputs into an NWP model. Figure 2 is a schematic diagram of the ACMES process. The model is run over a ten year period, typically 1987 to 1996 (Van Knowe, 1999).

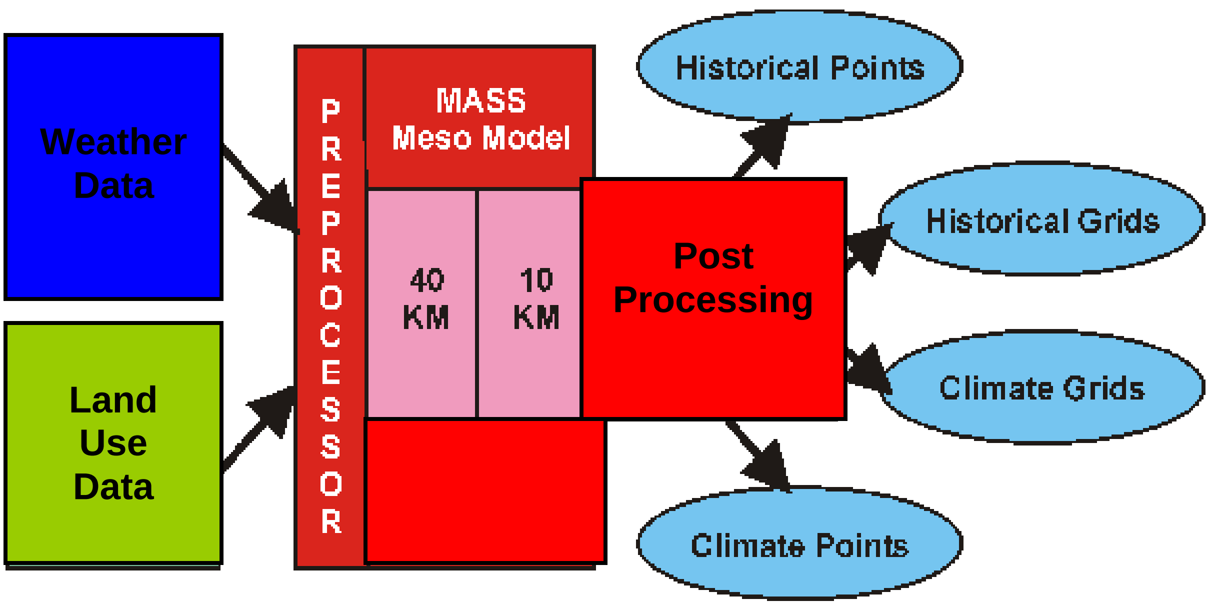

Figure 2. Process used to produce gridded climatologies with ACMES.

The model is updated with the analysis grids every six hours and observational data every 12 hours, so it never has a chance to stray significantly from the observed atmosphere. However the model produces hourly output at 40 and 10 km, which is a much higher resolution than the input data. These hourly grids are then summarized to produce gridded climatologies.

Figure 3 is an example of an ACMES climatology with 40 km grid resolution over eastern Asia. In this case the model has produced the mean maximum daily temperature for April. There is information over ocean and poorly observed land regions; however more detail would be useful over mountainous areas. The 10 km grids (not shown) did have more detail, but still did not resolve some of the small-scale temperature features caused by topography. An NWP model only resolves features at approximately twice its grid size. Therefore topographic features smaller than about 80 km have little affect on the output in Fig. 3. A method is needed to create more fine-scale detail in the final gridded climatology.

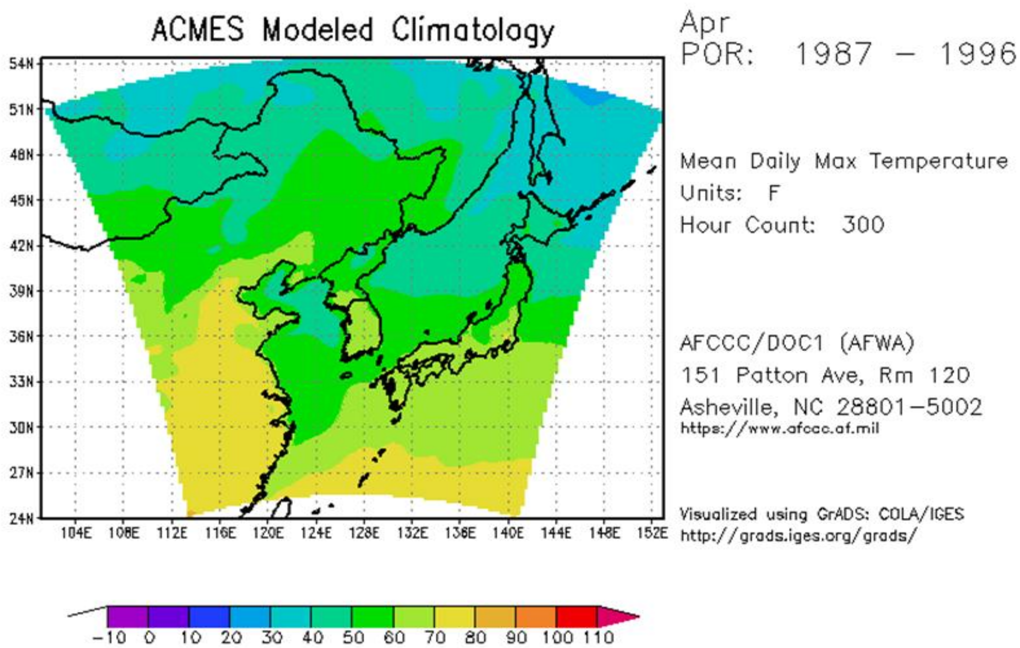

Figure 3. April mean daily maximum temperature for eastern Asia produced with ACMES. The climatology is based on data from 1987-1996.

Downscaling

One approach to improve the resolution of the ACMES output (or any other low resolution NWP output) is called 'downscaling'. The ACMES output is directly correlated with information that is available at finer spatial resolution. A multiple linear regression model was developed to downscale 40 km ACMES temperature climatologies to 1km and 10 km. Elevation, dew point, and latitude were chosen as the predictors. Temperature and elevation are strongly correlated and elevation is available globally at 1 km resolution. Dew point was used because atmospheric moisture tends to moderate temperatures. This can be important when the spatial domain being modeled contains coastal areas and interior regions. The predictor dew points were obtained from ACMES and interpolated down to the final output resolution. While this technique does not produce high-resolution moisture information, it provides the regression with a "spatial trend" of the moisture field and allows the regression to be run at the 1 km and 10 km resolution. Finally, latitude was used because the spatial domains were large enough that the latitudinal variation of temperature was important. Latitude and high resolution elevation have little statistical dependence over most regions so they make excellent co-predictors in the regression model. An example of the downscaled climatology is shown in Figure 4.

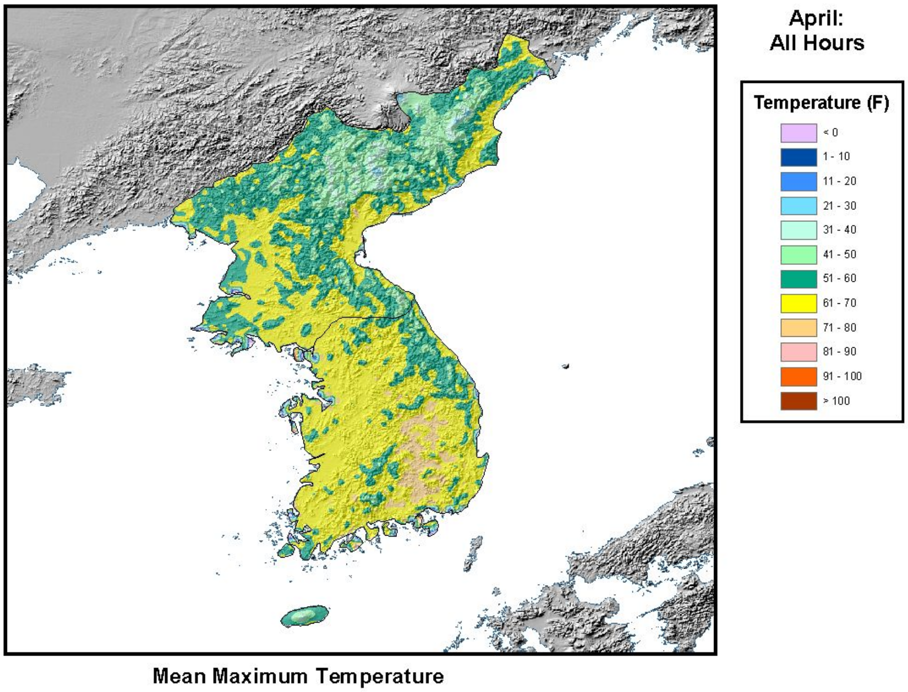

Figure 4. April mean daily maximum temperature for eastern Asia produced with ACMES and downscaled using multiple linear regression.

Comparing Figs. 3 and 4, it is apparent that the climatology which combines NWP and statistical methods yields much more detail. The information about elevation, moisture, and latitude allows the final climatology to be much more realistic. While a strictly formal verification was not done for this method, the regression diagnostics were all favorable (high coefficient of determination and low mean square error). Preliminary comparison with observed climatologies indicated low RMS errors and low bias.

Two Component Synoptic Region Method

The NCDC is also developing a 'Two Component Synoptic Region' (TCSR) mapping technique to produce gridded climatologies. The method uses daily station measurements, daily NWP analyses, and high resolution topography to produce high resolution temperature, precipitation, and surface pressure climate grids. TCSR mapping can produce high resolution maps of other climate parameters such as wind, visibility, and humidity if user-defined 'co-predictor' grids are added to the input. The TCSR method, including the use of co-predictor grids, is described below.

Station measurements are grouped into 'synoptic regions' and then modeled within each region as the sum of a '*predicted component*' and a '*residual component*'. Synoptic regions are moving, contiguous collections of grid cells that exhibit a common meteorological property (possible examples are areas defined by air mass source regions, areas defined by frontal characteristics, etc).

The *predicted component* is estimated from a statistical model that relates the quantity to be mapped to two sets of high resolution daily grids. The first set of these predictor grids consists of directions defined by the angle between the low level wind and the gradient of elevation, i.e., 'barrier angle' grids. The cosine of the barrier angles adds upslope and downslope information to the statistical model. See Figure 5. In this example, westerly winds over Colorado represent upslope conditions on western slopes and downslope conditions on the eastern slopes.

The second set of predictor grids is user-defined, high resolution grids that complement the upslope/downslope information within synoptic regions. These user-defined grids are denoted as 'co-predictor' grids. Elevation can be used as the co-predictor grid to map temperature, precipitation, and surface pressure. User-defined co-predictor grids for climate parameters such as wind, visibility, and humidity must be constructed outside of the TCSR method and then used as input to the TSCR mapping.

Differences between the predicted component and point observations in a region are not considered to be 'errors', but rather are considered to be point residuals. *Residual component* grids are estimated from a spatial distribution model. The predicted component and residual components are combined to create the final daily grid. Figure 6 gives an overview of the TCSR process.

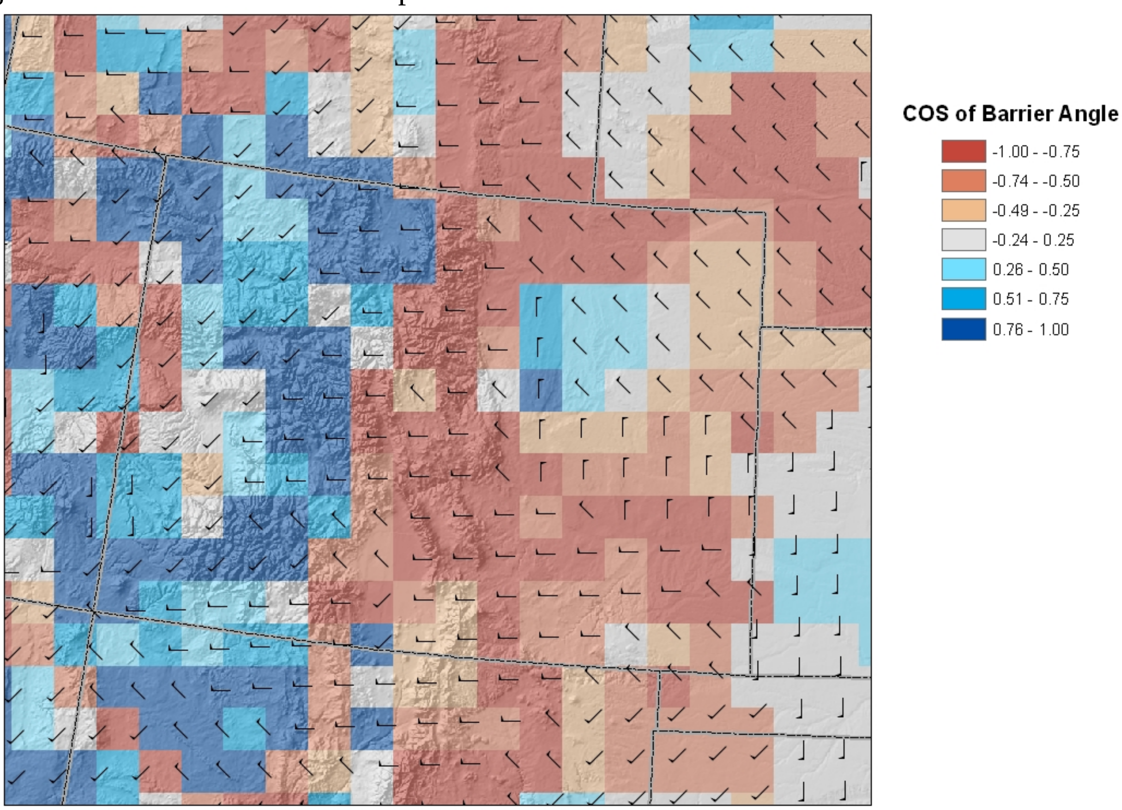

Figure 5. An example of terrain, daily winds, and cosine of the daily barrier angle over Colorado. The winds are blowing away from the “barbs”. The orange colors represent downslope winds and blue colors represent upslope winds.

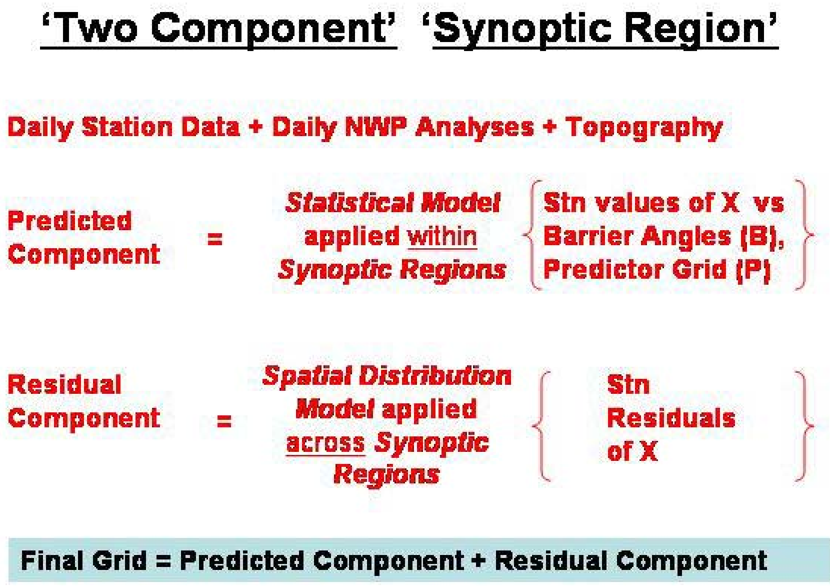

Figure 6. Overview of the Two Component Synoptic Region method.

Monthly and seasonal climatological statistics such as average, extremes, number of days exceeding a threshold, etc. are computed directly from the long series of daily grids produced within the TCSR method.

Conclusion

There is a need for detailed climatology information in locations where no systematic observations are taken. This paper has described two promising methods that the NCDC is investigating to generate gridded climatologies. Our strategy is to combine output from NWP models with various statistical techniques to take advantage of both methods.

References

Carrega, P, 1995: A method for the Reconstruction of Mountain Air Temperatures with Automatic Cartographic Applications. *Theor. and Appl. Clima.*, **52**, 69-84.

Cox, R., B. Bauer, T. Smith, 1998: A Mesoscale Model Intercomparison. *Bull. Amer. Meteor. Soc.*, **79**, 265-283.

Daly, C., R. Neilson, D. Phillips, 1994: A Statistical-Topographical Model for Mapping Climatological Precipitation over Mountainous Terrain. *J. Appl. Meteor.*, **33**, 140-158.

Isaaks, E., R. Srivastava, 1989: *Applied Geostatistics*. Oxford University Press, 561 pp.

Van Knowe, G. E. , J. W. Zack, S. Young, M. D. Bousquet, P..E. Price, C. E. Graves, K. Poage, 1999: An evaluation of the sensitivity of local climate statistics generated from the output of a 3-d mesoscale atmospheric model to model configuration, data assimilation, resolution and subgrid parameterization schemes. Preprints, 11th Conference on Applied Climatology, Dallas, TX, Amer. Meteor. Soc.

Wilks, D. S., 1995: Statistical Methods in the Atmospheric Sciences. Academic Press, 467 pages.