1. Introduction

Downscaling is the methods that are used to infer local- or station-scale climate information from large-scale atmospheric predictors of coarsely resolved climate models or reanalysis. Downscaling precipitation is very important since accurate estimates of local precipitation and its variability are valuable for informing local policy dicesions and estimating potential impacts on areas such as hydrology, water resource management, agriculture, forestry, floods, droughts, soil erosion, land use change, groundwater, enfironment trourism, human and animal population and their health. There are a significant amount of downscaling techniques in literature, but two major approaches can be categorized to dynamical downscaling and statistical downscaling(Wilgy et al.2002; Anandhi et al.2008) The dynamic downscaling approach extract local-scale information by developing and using limited-area models or regional climate models, with coarse GCM data used as boundary conditions. However, their complicated design and high computation cost can be considered the drwabacks of these methods.The basic steos are then to use a GCM to simulate the response of the global circulation to large scale forcings, such as complex topographical features and land cover inhomogeneity in a phusically based way.(xu 1999)

2. material and methodology

Study area and data

The study region is the Oklahoma region. We use observation of daily precipitation(millimeters) from local weather stations as local variable. Daily precipitation data over period 1979-2010 from 10 stations (Table1) were obtained from the National Weather Service’s (NWS) Cooperative Observer Data (COD). The stations selected with primary station located in Norman (N35.18, W97.43) and 3 other stations close to primary station and 3 stations relatively father away from the primary station. The primary station was the statistic downscaling model develops from and the additional stations were chosen for spatial validation purpose.

|  |  |  |
| --- | --- | --- |
|  | latitude | longitude |
| primary | 35.18 | -97.43 |
| close | 35.12 | -97.67 |
| 35.38 | -97.6 |
| 35.17 | -97.9 |
| far | 34.78 | -96.68 |
| 34.4 | -96.13 |
| 35.8 | -96.67 |

Large-scale predictor variables were derived from three different reanalysis: 1, Modern-era Retrospective Analysis for Resrach and Applications(MERRA) from NASA. 2, ERA-Interim produced by the European Centre for Mdedium-Range Weather Forecasts(ECMWF). 3, North American Regional Reanalysis(NARR) provided by NOAA. Predictor variables from these reanalysis are extract over a period 1979-2010.

The strength of predictors is time and space-dependent. Wilby and Wigley (2000) found that predictor’s domain is crucial factor that determines reliability and robustness of precipitation downscaling models. So in this research the potential predictors will cover a large quantity of variables as well as a wide range of area that will cover all the selected weather stations. The relative predictors’ grid points of these three reanalysis and the primary stations are shown in Figure 1.

In the following sections, the predictor selection procedure is illutstrated and the modeling method are described. The results of the models are then followed by the dicccussion on the proposed methodology.

methodology structure

The whole downscaling procedure are divided into four separate steps, first analysis all the potential predictors from all three different reanalysis. Second, select predictors from different reanalysis, third step is decide the bias correction technique, the final step is modeling and forecasting.

imbalance sample bias correction techniques

a dataset is imbalanced if the classification categories are not approximately equally represented.

Synthetic Minority Over-sampling technique(smote)

*SMOTE* is an approach to the construction of classifiers from imbalanced datasets. Smote over-sample the minority class by creating “synthetic” examples rather than by over-sampling with replacement.

Synthetic samples are generated in the following way: Take the difference between the feature vector (sample) under consideration and its nearest neighbor. Multiply this difference by a random number between 0 and 1, and add it to the feature vector under consideration. This causes the selection of a random point along the line segment between two specific features. This approach effectively forces the decision region of the minority class to become more general.

statistic models

modeling select metric

absolute PSS is used to select model while model training. The reason we use absolute PSS rather then PSS is that the PSS only gives information about the maximum achievable value, and gives no information about the value for a specific user having a cost/loss ratio differen to the base rate. Strictly the maximum value should be the absolute value of the perice skill score since forecasts with negative Peirce skill scores can always be recalibrated (by relabeling the forecasted event as non-event, and vice versa)to have positive skill. (jolliffe,& Stephenson ,2012).