

Application of tree-structured regression for regional precipitation prediction using general circulation model output

Xiangshang Li, David Sailor*

Department of Mechanical Engineering, 400 Lindy Boggs Center, Tulane University, New Orleans, Louisiana 70118, USA

ABSTRACT: This study presents a tree-structured regression (TSR) method to relate daily precipitation with a variety of free-atmosphere variables. Historical data were used to identify distinct weather patterns associated with differing types of precipitation events. Models were developed using 67 % of the data for training and the remaining data for model validation. Seasonal models were built for each of 2 US sites: San Francisco, California, and San Antonio, Texas. The average correlation between observed and simulated daily precipitation data series is 0.75 for the training set and 0.68 for the validation set. Relative humidity was found to be the dominant variable in these TSR models. Output from an NCAR CSM (climate system model) transient simulation of climate change were then used to drive the TSR models in the prediction of precipitation characteristics under climate change. A preliminary screening of the GCM output variables for current climate, however, revealed significant problems for the San Antonio site. Specifically, the CSM missed the annual trends in humidity for the grid cell containing this site. CSM output for the San Francisco site was found to be much more reliable. Therefore, we present future precipitation estimates only for the San Francisco site.

KEY WORDS: Climate change · Downscaling · Precipitation · Regional climate · General circulation models (GCMs)

Resale or republication not permitted without written consent of the publisher

1. INTRODUCTION

General circulation models (GCMs) are generally good at replicating large-scale circulation features of current climate (IPCC 1995). However the GCM surface predictions are not accurate on regional scales. Thus caution should be taken when applying them to climate change impact analyses. The limitations of GCMs have long been recognized, and several approaches have been suggested to remedy the problem. Downscaling is a technique that bridges the gap of GCM prediction skills over different scales. Giorgi & Mearns (1991) provided an early review of the various efforts in regional climate change simulation. The 3 categories of approaches discussed in this review were

empirical, semi-empirical, and nested modeling approaches. The semi-empirical (statistical downscaling) approaches use GCM output to represent large-scale forcing and develop empirical statistical relationships to account for mesoscale phenomena. Nested modeling, on the other hand, uses a higher resolution dynamical climate model to account for mesoscale forcing. A more recent review on downscaling by Wilby & Wigley (1997) divided downscaling into 4 categories: regression methods, weather pattern approaches, stochastic weather generators, and limited-area climate models. Newer developments include a statistical/dynamical approach (Fuentes & Heimann 1996) and a time-slicing approach (Cubasch et al. 1996). Therefore, a better taxonomy may break downscaling approaches into 3 types: statistical downscaling, hybrid statistical/dynamical downscaling and dynamical downscaling. The first 3 categories summarized by Wilby & Wigley

*Corresponding author. E-mail: sailor@mailhost.tcs.tulane.edu

actually fit within the general bounds of statistical downscaling.

In recent years, there has been growing interest in statistical downscaling, especially in the area of precipitation downscaling. This may be explained by the fact that statistical downscaling is computationally inexpensive, relatively quick to implement, and flexible. When properly established, statistical downscaling models can provide comparable results with the other 2 approaches. A discussion of various statistical downscaling methods was recently given by Zorita & von Storch (1999). Statistical downscaling is a 2-step procedure. First it uses statistical techniques to relate large-scale climate parameters to local surface variables such as temperature and precipitation. Then GCM output of perturbed climate is substituted into the models to generate predictions of future climate. These models are sometimes called transfer functions, since once established they can transfer large-scale GCM variables to local surface predictions. Statistical downscaling is built on 2 assumptions: the stationarity of the transfer functions under altered climate and the accuracy of GCM large-scale output. While it is very difficult to verify the stationarity assumption, the accuracy assumption is not hard to check.

Current statistical downscaling studies generally perform well when dealing with temporally continuous variables such as temperature that exhibit little random fluctuation in their time series. Explained variances in temperature downscaling models can be as high as 90% (Sailor & Li 1999). However, it is much more difficult to downscale discontinuous, highly intermittent variables such as precipitation. Of the precipitation downscaling studies found in the literature, only a few display their results in explained variances. Gyalistras et al. (1994) reported explained year-to-year variances of 29 to 55% in winter and 10 to 28% in summer. Noguera (1994) found a correlation ranging from -0.16 in the summer to 0.79 in the winter for monthly mean precipitation. Enke & Spekat (1997) gave a result of 11.5 to 25.4% for daily precipitation. In a more recent study, Weichert & Burger (1998) performed a comparative study of linear and non-linear techniques (neural network clustering and neural network function approximation) in downscaling. Correlation between simulated and observed normalized precipitation series was 39% for the linear model and 42% for the non-linear model. As expected all authors (except Noguera, who did not conduct temperature analysis) reported much higher explained variances in their temperature downscaling. It should be noted that monthly mean precipitation has different statistical characteristics than the daily time series and is less difficult to model.

Because statistical downscaling is rooted in the numerical-statistical methods (predominantly the

model output statistics [MOS] and perfect prognosis [PP] methods) used by the US National Weather Service, it is helpful to take a look at operational precipitation forecasting. Shuman (1989) showed that with the improvement of weather models over the years, the prediction of large-scale variables has significantly improved. However, these improvements did not translate into a concomitant increase in precipitation prediction skill. The difficulties in precipitation forecasting have been acknowledged by many researchers, and Anthes (1983) suggests this difficulty may result from the increasingly stochastic nature of smaller-scale phenomena. The precipitation caused by these small-scale phenomena is not captured by the current operational models. The relatively poor performance of current precipitation downscaling can be partly explained similarly—the few large-scale variables used in most current studies do not contain enough information for the prediction of precipitation events. Currently, there are 2 choices with regard to the selection of predictor variables: a wide area view, which uses gridded field variables, and a local area view, which uses grid point free-atmosphere variables. The majority of downscaling research uses gridded field variables. Sea level pressure (von Storch et al. 1993, Gyalistras et al. 1994, Zorita et al. 1995, Cubasch et al. 1996, Hewitson & Crane 1996) and geopotential heights (Matyasovszky et al. 1994, Enke & Spekat 1997, Weichert & Burger 1998) are the most widely used predictors for downscaling precipitation. Karl et al. (1990), Winkler et al. (1995) and Sailor & Li (1999) are among the few who took a local area view. While the gridded large-scale variables can give a good picture of the large-scale circulation of a broad region, using them as the only predictors fails to capture the role of smaller-scale features and phenomena. On the other hand, because the local area view approach takes GCM output from only 1 grid cell, it is more susceptible to limitations in the accuracy of the GCM output. Strictly speaking, the latter approach is not 'downscaling' per se. However, due to the similarity between the 2 approaches, we choose to refer to this second approach as one form of downscaling. This study follows the local area view approach by looking at relatively small regions (<100 km in diameter). It should be noted that an increase in the physical representation would be limited by the spatial and temporal scale of available observations. Therefore the problem of insufficient spatial and temporal data coverage is still an issue.

Statistical methods used in most precipitation downscaling studies generally fall into 2 categories. The first category treats the data pool as one unit and develops transfer functions for it. This approach includes principle component analysis (PCA)/canonical correlation

analysis (CCA) (Gyalistras et al. 1994, Noguer 1994, etc.), regression analysis (Karl et al. 1990) and artificial neural network (ANN) function approximation (Hewitson & Crane 1996 etc.). The second category separates the data pool into several classes, or weather patterns (WPs) as a first step. Then predictor-predictand relationships for each class are established. Cluster analysis (Bardossy et al. 1992, Enke & Spekat 1997), classification and regression trees (CART) (Hughes et al. 1993, Zorita et al. 1995) and radial-based function (RBF) neural networks (Weichert & Burger 1998) belong to this WP-based approach. The analog technique used by Zorita et al. (1995) and Cubasch et al. (1996) is somewhat similar to cluster analysis, although it does not explicitly define WPs/clusters. The first category can be viewed as a special case of the second when there is only 1 WP. Theoretically the statistical techniques used in the first category also apply to the second category for establishing the predictor-predictand relationships over each WP.

There are 2 distinct ways to classify WPs; a self-organizing (unsupervised) classification that includes cluster analysis and some ANNs, and a guided (supervised) classification scheme such as CART. Currently, most WP-based downscaling studies use cluster analysis, especially the *k*-means clustering algorithm. Cluster analysis finds cluster structure inside a data pool. Resulting clusters consist of points separated by small distances, relative to the distances between clusters. The distances are calculated by using only the predictor variables. Due to the difficulties in rescaling different types of large-scale variables to calculate distances, self-organizing classification usually uses only 1 or 2 types of large-scale variables. Typically the primary variable is sea level pressure or the 700 hPa geopotential height. Usually WPs generated by self-organized classification are 'universal', since, once built, they are assumed to be good for predicting any kind of surface variable. On the other hand, guided classification uses a predictand for guidance when generating clusters (or WPs). The 'distance' between clusters is measured by the value of the predictand variable. Therefore there is no limitation in selecting predictors. Because the WPs created by guided classification are predictand-sensitive, they are only valid for predicting the same kind of variable. Another difference between the guided and self-organizing classification is how to decide on the number of final clusters (WPs). All self-organizing classification schemes require subjective determination of the number of clusters. Some guided classification schemes such as tree-structured regression (TSR) do not require this subjective determination.

CART is a data-driven, rule-based classification technique pioneered by Breiman et al. (1984). There

are 2 different though related components in CART: the classification tree and the regression tree (also called Tree-Structured Regression, or TSR). The primary difference between the classification tree and the regression tree is that the former seeks to find a set of rules that classify the data into pre-defined groups, while in the latter approach there are no pre-defined groups. In regression trees the groups are generated automatically during the classification process. In other words, the predictand in the classification tree method is categorical, while in the regression tree approach it is continuous. Hughes et al. (1993) first experimented with classification trees to derive WPs associated with the occurrence or absence of precipitation. Zorita et al. (1995) applied a similar approach to 2 US regions. After identifying a few weather states that are closely related to the wet/dry states of weather stations, they used a precipitation generator to simulate the mean precipitation, storm interval times and daily precipitation probability distributions. One issue with this approach is that when the precipitation patterns of several stations are not closely related the CART procedure may not generate any weather categories. This can happen when small-scale precipitation events constitute a significant portion of the total precipitation.

As mentioned in Weichert & Burger (1998), the daily precipitation distribution in most places exhibits a highly skewed distribution with a peak at zero precipitation. Direct modeling of precipitation is difficult because most predictors are fairly normally distributed. Even powered with the ANN's universal function approximation ability, it is still difficult to obtain strong models (Hu 1998). Some level of preliminary data processing seems indispensable. For example, Weichert & Burger (1998) conducted a normalizing step and Hewitson & Crane (1996) used a multiple-day running average before applying an ANN.

The statistical approach used in this paper is based on the TSR component of CART. This approach is fundamentally similar to the probability of precipitation (PoP) forecasting procedure of the original MOS (Glahn & Lowry 1972). While the original MOS used a 100 pre-defined questions (rules) to predict PoP, TSR constructs the rules objectively during the classification process so each rule must be effective in separating the data. Also the overall skewness of the precipitation distribution becomes less of a problem when the raw data pool is broken into a series of clusters. This is because the datum points inside a cluster have similar precipitation magnitudes. In contrast to TSR, classification trees need to define precipitation 'bins' subjectively. When downscaling is conducted on different regions the bins should be adjusted to reflect characteristics of regional precipitation. Inappropriate bin definition can reduce model strength.

2. METHODOLOGY

Before developing downscaling relationships for any particular region there are several rules that should be followed. First, one must identify a set of physically relevant GCM parameters that are strongly correlated with the parameter to be downscaled. The GCM output of each of the candidate large-scale parameters must then be tested against observational records to ensure that the parameters are well-simulated by the climate model to be used. The statistical relationships developed must then be validated on an independent data set. In this paper this philosophy has been applied to develop a downscaling approach that is similar to the perfect prognosis (PP) procedure for daily precipitation.

PP was developed to take advantage of the dynamical forecasts from the numerical weather prediction (NWP) models. In PP, historical surface observations are related to observational large-scale variables to develop the forecasting equations. Then the forecasted large-scale variables from NWP models are substituted into the equations to calculate surface predictions. The performance of PP relies on 2 conditions: the strength of the forecasting equations and the accuracy of the NWP-generated large-scale predictors. In this paper, the same general procedure is followed except that TSR is used instead of regression analysis and GCM output has replaced that of an NWP model.

Among the various statistical techniques mentioned earlier, TSR was selected as the modeling tool because of its distinctive features. TSR exerts no limitation in selecting predictors. Any number and any kind (categorical or continuous) of predictors can be used. Its capability in handling categorical predictors makes it possible to include month, season, or Julian day in the model. In some cases this could lead to a better seasonal stratification scheme. TSR analysis is relatively simple to implement. The algorithm described in the original Breiman et al. (1984) book is not very complicated. It consists of only a few elements—a splitting rule, a stopping rule and an assigning rule. Being a component of CART, TSR shares the same strengths with respect to accuracy. According to Steinberg & Colla (1995), when automatic CART analyses are compared with stepwise logistic regressions or discriminant analyses, CART typically performs about 10 to 15% better on the training sample. When automatic CART analyses are compared with the best parametric models of sophisticated teams of statisticians, CART is still competitive. CART can often generate models in an hour or two that are only slightly worse in predictive accuracy than models that may take specialists several days to develop. TSR models are easy to interpret and can be simply displayed using a number of descriptive rules.

In order that the established TSR models generate reliable precipitation projections, one must ensure that GCM-predicted large-scale variables under an altered climate are reliable. To account for the systematic component of GCM internal errors, the difference of GCM $2 \times \text{CO}_2$ (doubling point) and $1 \times \text{CO}_2$ predictions were used instead of the direct $2 \times \text{CO}_2$ output. The differences were then added to the $1 \times \text{CO}_2$ baseline observations. Although this procedure may reduce the systematic error within the GCM, other errors such as the incorrect representation of physical processes are not affected. It is still important to make sure that no poorly simulated GCM variables enter the TSR models. In this analysis 19 large-scale variables are used, 16 base variables and 3 derived variables. Since the strength of the GCM to simulate different kinds of large-scale variables varies, a quality check of the 16 GCM and observed base variables was conducted.

2.1. Data

The 2 regions selected for this analysis are San Antonio, Texas, and San Francisco, California. Fig. 1 shows the location of the sites as well as the extent of the corresponding GCM grid cells. There is a radiosonde station and a number of precipitation stations within each region (grid cell). The GCM used in this study is the climate system model (CSM) of the National Center for Atmospheric Research (NCAR). The T42 CSM grid has a spatial resolution of $2.8 \times 2.8^\circ$, which translates to a 310 km (S-N) \times 270 km (E-W) grid over the San Antonio area and a 310 km (S-N) \times 250 km (E-W) grid over the San Francisco area. While the large-scale observations from the radiosonde station are considered valid over an area the size of a CSM grid cell, the observations from precipitation stations are representative of much smaller regions. To build a meaningful transfer

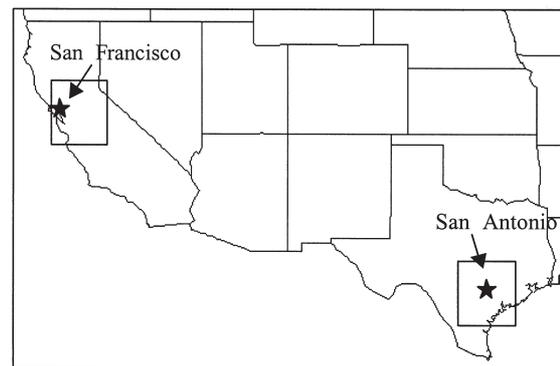


Fig. 1. Location and corresponding GCM grid cell for the 2 sites modeled in this paper

function, the precipitation data must be aggregated to better match the scale of large-scale observations. For this purpose, 15 precipitation stations were selected for San Antonio and 13 for San Francisco. A number of studies have investigated the factors that affect the spatial distribution of rainfall. Singh & Chowdhury (1986) performed a comparison analysis of 13 different methods of estimating mean areal rainfall on a daily, monthly, and yearly basis for 3 different hydrologic environments. They concluded that there is no basis to claim that one method is significantly better than the others. Therefore, in this study the site data were aggregated by taking simple algebraic averages. The precipitation sites were chosen based on proximity to the radiosonde data site for upper air observations. In the case of San Antonio the radiosonde site had been moved approximately 150 km during the period of record. All precipitation sites for San Antonio are therefore located within a 75 km radius of the center of the radiosonde sites. For San Francisco all precipitation data sites are located within 40 km of the radiosonde site.

2.1.1. Observational data set

Two kinds of observational data sets are necessary to build a transfer function, the large-scale variables and local surface variables. Fifteen free-atmosphere variables were extracted from the NCDC (National Climate Data Center) Radiosonde Data of North America for the period of 1946 to 1992: temperature (T), geopotential height (HT), dew point, wind speed and direction at the 850, 700, and 500 hPa levels. These large-scale variables represent all the possible observational variables from the 3 lowest Radiosonde Observation (RAO) levels. Dew point, wind speed, and direction of wind speed were then converted to relative humidity (RH), U velocity (zonal component of wind speed) and V velocity (meridian component of wind speed). Many of these variables are traditionally used in precipitation weather forecasting. To be consistent with the PP/MOS methods in precipitation forecasting the K -index (KI) and T -totals (TT) derived variables, known to be important for precipitation prediction, were added to the variable list:

$$KI = (T_{850} - T_{500}) + (Td_{850} - Tdd_{700}) \quad (1)$$

$$TT = (T_{850} + Td_{850}) - 2T_{500} \quad (2)$$

In these equations Td is the dewpoint temperature at the pressure level denoted by the subscript, and Tdd is the dew point depression. The values of KI and TT are usually very high during severe weather. Sea level pressure (SLP) for the 2 regions is taken from the

NCDC Surface Airways data set. The backward 24 h change of SLP ($SLP-$) was also calculated. The forward and backward 24 h changes of other free-atmosphere variables were not included in the variable list here for 2 reasons. First, the large number of missing data would be amplified by the requirement of having valid sequential data for calculating differences. Furthermore, some experimentation with forward and backward differences on other variables resulted in minimal impact on model quality.

The RAO data extracted from the 2 sites both spanned from 1958 to 1992 and were sampled twice daily (0Z and 12Z), with daily averages calculated as an average of the 2 observations. Daily precipitation data for the same period (1958 to 1992) were obtained from the NCDC Summary of the Day data set which contains over 6000 cooperating precipitation stations in the US. Around 20 precipitation stations were initially selected for each region. After a screening process to ensure good spatial and temporal data coverage, 15 stations were selected for San Antonio and 13 stations for San Francisco.

2.1.2. GCM output data set

The daily GCM output were taken from a CSM transient run (b006) completed in 1997 (NCAR Science Briefing 1997). It is a 130 yr coupled simulation with CO_2 levels increasing $1\% \text{ yr}^{-1}$. Its atmospheric component is the Community Climate Model (CCM 3.2) with a T42 resolution. The ocean component is the NCAR CSM Ocean Model (NCOM 1.1). Other components include the Land Surface Model (LSM 1), the CSM Sea Ice Model (CSIM 3.5.3) and the CSM Flux Coupler. The run was initialized from year 15 of the b003 simulation, which is a 300 yr CSM control run. For the first 10 yr (Years 15 to 24), the CO_2 level was held constant at the present day CO_2 level (355 ppm). At Year 25, a compounded $1\% \text{ yr}^{-1}$ increase in the CO_2 level was initiated. Doubling of atmospheric CO_2 concentrations was achieved in Year 95 of the run and a tripling occurred in Year 135. The current climate output for the present study was taken from the first 10 yr (Year 15 to Year 24), while future climate output was from a 10 yr period beginning in the Year 95, representing the doubling point for CO_2 concentrations. To match the observational data, 16 base free-atmosphere variables were extracted. These variables include HT , T , RH , U and V for each of the 850, 700, and 500 hPa, levels as well as SLP . For comparison purposes, total precipitation output was also extracted. All GCM output were obtained using the CCM processor through the Internet Remote Job Entry (IRJE) system of NCAR.

2.2. Statistical techniques

Similar to the classification component of CART, the TSR builds a binary-tree-like structure with nodes. Originally all the data points in the training sample reside in a single root node. By continuously posing and answering binary (yes/no) questions, each data point flows down to next level of nodes. The binary questions are constructed to maximize the difference of predictand (Y) in the 2 daughter nodes. Finally each data point attaches to a terminal node. Unlike classification, each data point in the root node is not pre-assigned to any group. All the nodes (mid-tier and terminal) are formed automatically during the process. The key elements of TSR involve developing the questions (splitting rule), determining when the tree is large enough (stopping rule), and deciding how to characterize data points that reside in any terminal node (assigning rule).

A basic TSR will only give discrete categories for the predictand. All the data points in 1 terminal share 1 predicted value: either the mean of the data points (least-square difference for assigning rule) or the median (least absolute deviation for assigning rule). Although sometimes a basic TSR is sufficient, more often it is desirable to conduct an in-node regression to achieve better accuracy. Usually a multiple linear regression (MLR) within the terminal nodes is good enough. However, a more sophisticated neural network model may also be used. Because TSR alone tends to be less accurate on linearly structured data, a hybrid combination of TSR and MLR (or neural network) will make the downscaling procedure more robust.

2.3. Building TSR models

Seasonal aggregation has been used in most previous statistical downscaling studies to reflect the fact that the transfer functions can be different in different seasons. Usually the seasonal division (also called stratification) separates the year into 4 (spring, summer, fall and winter) or 2 (winter and summer) seasons. In this analysis the year was divided into 6 seasons, each of 2 mo duration, starting from January. Because precipitation can take several forms (rainfall, snow, sleet, etc.) and its amount can vary significantly during the year, the stratification of 6 seasons can better capture the different precipitation types.

In any regression analysis, the condition that the training and testing data sets should come from the same population must be met. Due to the high interannual variability of precipitation, an arbitrary sampling procedure may result in an imbalance between train-

ing and testing data sets. For example, if one picks the first 20 yr for training and the remainder for testing, there is a chance that the training set may contain too many drought years and the testing set contain too many wet years. In this analysis, the training and testing data sets were created using a random sampling procedure with 67% of data points going to training data set. By putting all the data points in the same pool, the training set can be well mixed (can include data points from any month and any year for the season). Because the sampling procedure redraws data points each time a model is constructed, a different training data set is created. To achieve the best model, the procedure was repeated numerous times with different training and testing data sets. The best model is simply defined as the one that makes the best prediction for the testing set. Some choices we made in our TSR analysis included no cross-validation, 1% minimum percentage of cover (each daughter node must contain at least 1% of its parent node), and 10% extrapolation rate for predicting unseen cases. For any regression models, over-fitting can be a problem. In TSR analysis, cross-validation is often used to avoid over-fitting. It sets aside a percentage of training cases to test model accuracy during the model building process. However, because we use extra testing cases to validate, cross-validation can be thought of as an option rather than a requirement.

The TSR analysis was performed independently for each site and for each bimonthly season, resulting in a total of 12 tree models (summarized in Table 1). The correlation between observed daily precipitation and simulated precipitation falls in the range of 0.65 to 0.86 for the training set and 0.40 to 0.82 for the testing set. The average correlation for the San Antonio site is 0.70 for the training set and 0.64 for testing. For the San Francisco site, the values are 0.80 and 0.71 respectively. In most cases, the correlation for the training set is higher than that of testing set because the models are built to maximize the correlation for the training set. However, there is an exception in the May-June model for San Francisco. The worst model is the July-August model for San Francisco, with a testing data set correlation of 0.40. The problem with this model is that there are very few days with significant precipitation during this season. As a result the precipitation distribution is highly skewed, making it very difficult to generate a reliable model. In the other 5 San Francisco models this is not an issue and the next weakest model has a correlation of 0.74.

Most of the tree models varied significantly from season to season. The number of terminal nodes ranges from 4 to 11 for San Francisco and 5 to 12 for San Antonio. Fig. 2 shows a sample tree model of moderate complexity for San Francisco. The rectangular nodes are the

terminal nodes and the elliptic nodes are intermediate nodes. The number in a node represents the number of data points within the node and the number below a terminal node is the monthly average precipitation in centimeters. In this example the root node contains 1034 data points for training. The TSR process results in a total of 6 terminal nodes, representing seasonal precipitation ranges from 0.03 to 1.2 cm d⁻¹. Table 2 displays the details of the tree model. Since the ratio of training data points to testing datum points in generating the models is 2 to 1, a well-behaved model should roughly preserve this same ratio in each of its terminal nodes. The testing set generally contains half the datum points of the training set for an individual terminal node. The average precipitation of each node also agrees well. These results suggest no evidence of an over-fitting problem within this model. Fig. 3 compares the model prediction with the observations for the San Francisco site. Without the in-node regression, the correlation was found to be 0.70 for the training set as opposed to 0.82 with in-node regression. This indicates that the in-node regression only provides a modest increase (0.12 in this case) in correlation. Hence, while the model strength is most dependent upon the classification step, further improvement is obtained via in-node regression.

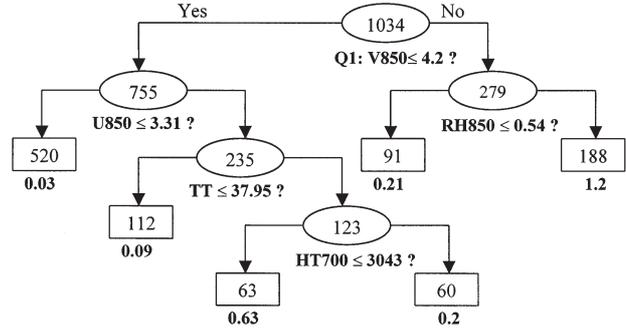


Fig. 2. Sample regression tree for the San Francisco site (January-February). Ovals and rectangles represent intermediate and terminal nodes, respectively. The binary questions are listed below intermediate nodes. The mean daily precipitation associated with each terminal node is given below the node (cm)

To compare the importance of the different independent variables, the Level 1 split variables are listed in Table 3. Being the first variable selected for splitting in a tree, the Level 1 split variable is the most influential in determining the precipitation amount. Table 3 clearly shows that the relative humidity, espe-

Table 1. Summary of TSR models including correlation results for training (Trn) and testing (Tst) data sets

	Jan-Feb		Mar-Apr		May-Jun		Jul-Aug		Sep-Oct		Nov-Dec	
	Trn	Tst										
San Antonio												
Cases	976	480	1031	508	1102	543	1223	603	1136	560	996	491
Avg. error (cm d ⁻¹)	0.15	0.16	0.17	0.18	0.27	0.28	0.16	0.17	0.25	0.28	0.16	0.18
Correlation	0.72	0.72	0.65	0.64	0.75	0.56	0.70	0.64	0.70	0.67	0.70	0.59
San Francisco												
Cases	1034	509	1093	538	1125	554	1115	549	1091	537	1087	536
Avg. error (cm d ⁻¹)	0.18	0.20	0.10	0.12	0.01	0.02	0.00	0.01	0.04	0.06	0.13	0.16
Correlation	0.82	0.75	0.83	0.82	0.69	0.74	0.78	0.40	0.84	0.77	0.86	0.79

Table 2. TSR model detail for the San Francisco January-February model tree

	Node 1	Node 2	Node 3	Node 4	Node 5	Node 6	Total
Training set							
Data points	520	112	60	91	63	188	1034
Precipitation avg. (cm d ⁻¹)							
Observation	0.02	0.09	0.2	0.21	0.63	1.25	0.32
Prediction	0.02	0.09	0.2	0.21	0.65	1.25	0.32
Testing set							
Data points	256	57	28	48	21	99	509
Precipitation avg. (cm d ⁻¹)							
Observation	0.02	0.17	0.2	0.25	0.46	1.03	0.28
Prediction	0.02	0.17	0.2	0.25	0.49	1.05	0.29

cially at the 850 and 700 hPa levels is the dominant variable. While the K -index and T -totals are traditionally considered to be important in predicting precipitation, neither appeared as top level split variables in the TSR models. Actually they seldom appeared as split variables for other levels within the trees. The models suggest that the relative humidity is more effective than these derived variables.

2.3.1. Quality checking of GCM output variables

As shown by various researchers (Portman et al. 1992, Sailor & Li 1999) GCMs generally predict free-atmosphere variables better than surface variables. However, the quality of the large-scale variables can differ significantly from variable to variable and region

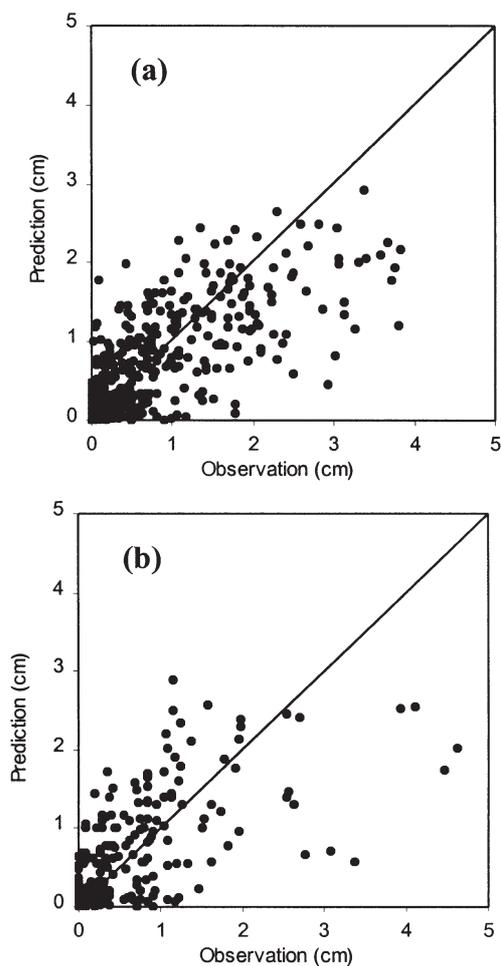


Fig. 3. QQ plots of model prediction versus observation for the January-February model of San Francisco. (a) Training set and (b) testing set

Table 3. List of top level (most significant) split variables

	Jan-Feb	Mar-Apr	May-Jun	Jul-Aug	Sep-Oct	Nov-Dec
San Antonio	<i>RH850</i>	<i>RH700</i>	<i>RH700</i>	<i>RH700</i>	<i>RH700</i>	<i>RH700</i>
San Francisco	<i>V850</i>	<i>RH850</i>	<i>HT700</i>	<i>RH850</i>	<i>RH850</i>	<i>RH850</i>

to region. For example, GCMs may predict surface temperature better than 700 hPa relative humidity, which is the case for the San Antonio site. Therefore, a careful comparison between GCM and observed parameters is necessary.

Since a GCM simulation does not represent any specific historical period, a direct comparison of the daily GCM and observed data is not practical. So an averaged data set of both GCM and observational variables was generated. To do this, 10 yr of observational data (1983 to 1992) were used to create an average meteorological year (AMY) data set. This was accomplished by taking a day-by-day average of the variables over the period. The GCM AMY data set was also calculated using the 10 yr $1 \times \text{CO}_2$ output. Then a series of scatter plots was generated from the 2 AMY data sets. As a sample of these the plots of 700 hPa level variables as well as SLP are presented in Fig. 4. To quantify the agreement between the GCM and observations the mean, median and standard deviation of the predictor variables were calculated for a sample season (Table 4). To investigate the changes that may happen during the observation years (1958 to 1992), the observational data in the first decade (1958 to 1967) was also used to calculate the AMY data sets. It was found that the differences in the 2 observational AMY data sets are minimal compared to the differences between the GCM and observational AMY data sets.

Because our TSR models show that the 850 and 700 hPa relative humidity are the dominant variables for both sites, they are examined in further detail. The scatter plots show good agreement between most GCM variables and observations. The relative humidity for San Antonio is an exception. Although it has about the same yearly average as the observations, it over-predicts in the beginning and at the end of the year and under-predicts in the middle of the year. In short, the annual cycle in the GCM is reversed. This could indicate that the GCM has serious internal flaws when predicting the moisture over the San Antonio region. The poor precipitation prediction over the San Antonio region (Fig. 5a) can be similarly explained. For San Francisco, relative humidity at 700 hPa has correct but exaggerated swings in the trends. The plots of the 850 and 500 hPa variables generally have a similar shape to their 700 hPa counterparts. The quality of the 700 hPa variables usually falls between those of 850

Table 4. Comparison of variable statistics for 700 hPa level and SLP from from GCM and observations

Variable	T_{700} ($^{\circ}\text{C}$)		RH_{700}		U_{700} (m s^{-1})		V_{700} (m s^{-1})		HT_{700} (m)		SLP (hPa)	
	Obs	GCM	Obs	GCM	Obs	GCM	Obs	GCM	Obs	GCM	Obs	GCM
San Francisco												
Mean	2.99	-0.26	0.30	0.38	3.91	4.37	-0.43	-1.48	3107.4	3093.7	1017.1	1019.5
Median	1.99	-3.01	0.30	0.42	3.71	3.97	-0.27	-0.96	3104.7	3088.3	1016.7	1019.4
SD	4.75	5.59	0.07	0.16	2.35	2.29	3.31	2.76	38.20	38.40	2.60	3.60
San Antonio												
Mean	6.69	4.32	0.41	0.41	4.21	4.00	1.36	1.48	3144.5	3139.3	1016.1	1019.2
Median	7.18	4.33	0.40	0.42	4.43	5.72	1.12	1.63	3145.2	3145.1	1015.6	1018.3
SD	2.68	4.70	0.10	0.15	4.43	5.81	2.09	2.06	33.80	35.70	3.10	4.00

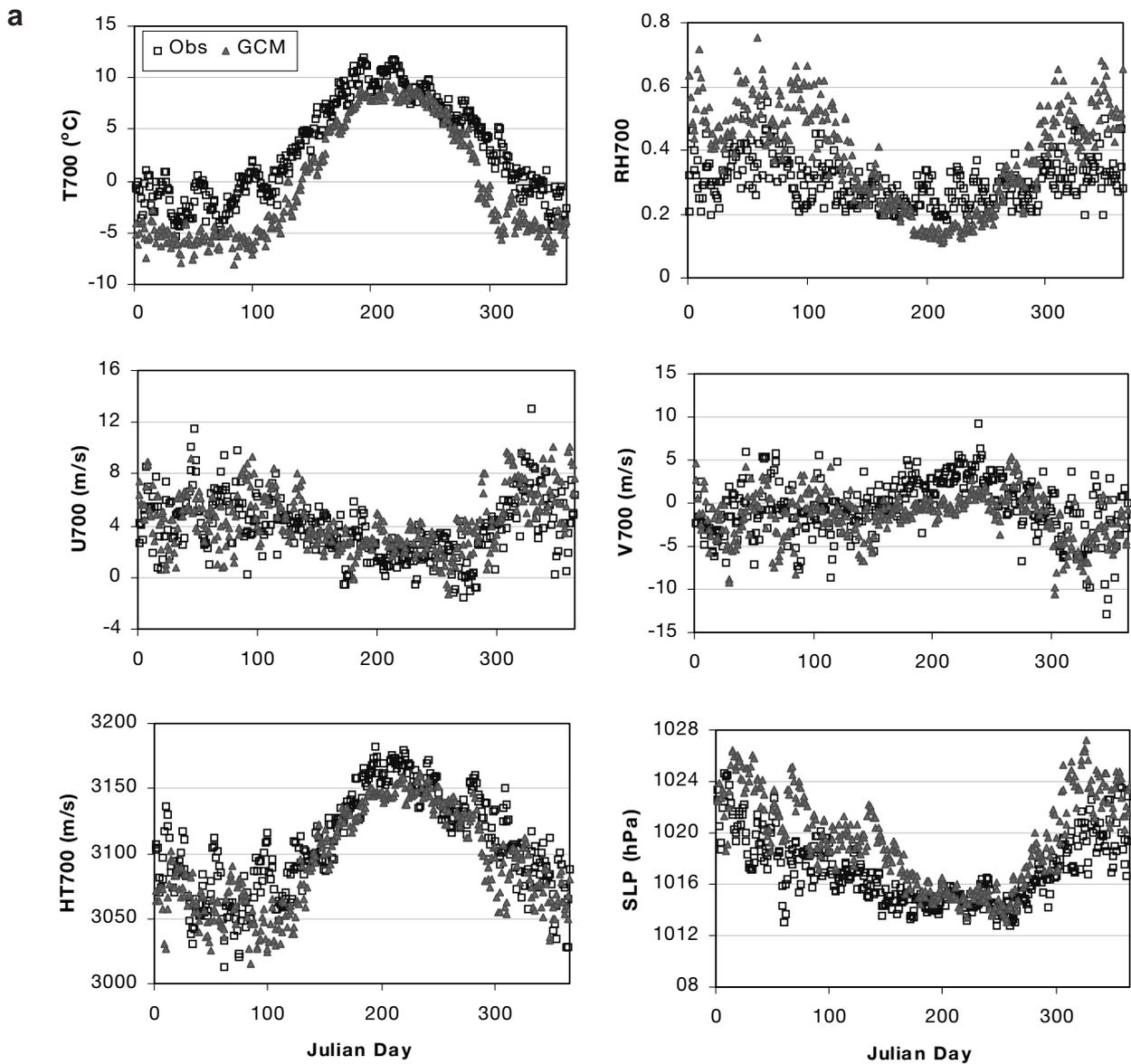


Fig. 4. (Above and following page) A comparison of the 700 hPa variables and sea level pressure (SLP) for GCM quality checking with observations for (a) the San Francisco site and (b) the San Antonio site

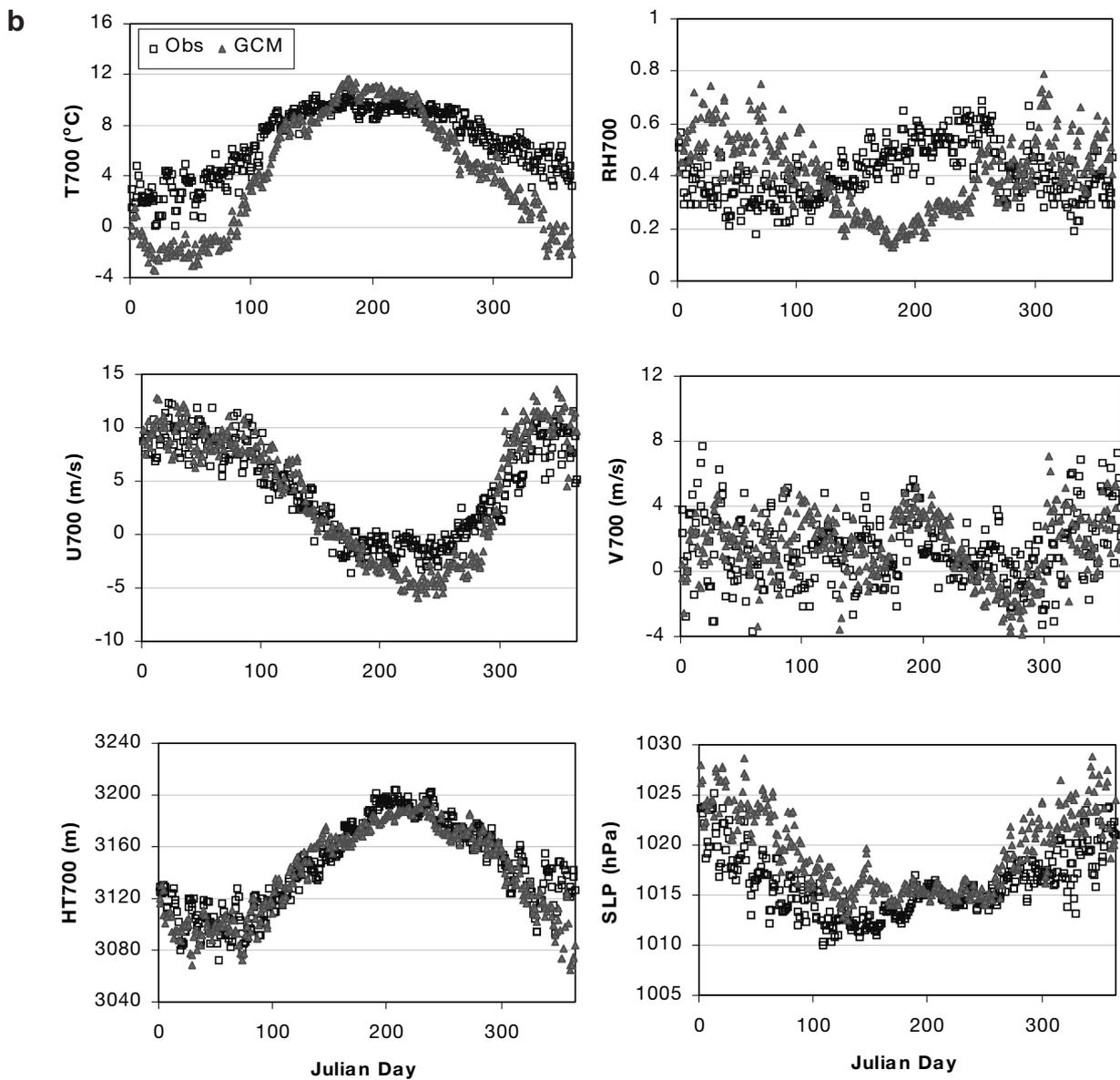


Fig. 4 (continued)

Table 5. GCM predicted change for San Francisco (January-February model)

Variable	Mean	SD	Variable	Mean	SD
<i>T</i> 850 (°C)	1.53	0.99	<i>V</i> 700 (m s ⁻¹)	3.02	3.57
<i>RH</i> 850	0.006	0.10	<i>HT</i> 700 (m)	-3.37	25.25
<i>U</i> 850 (m s ⁻¹)	0.81	2.08	<i>T</i> 500 (°C)	1.32	1.09
<i>V</i> 850 (m s ⁻¹)	2.63	3.05	<i>RH</i> 500	0.006	0.10
<i>HT</i> 850 (m)	-10.95	22.09	<i>U</i> 500 (m s ⁻¹)	1.79	3.54
<i>T</i> 700 (°C)	1.17	1.27	<i>V</i> 500 (m s ⁻¹)	3.02	4.93
<i>RH</i> 700	0.021	0.11	<i>HT</i> 500 (m)	9.07	33.83
<i>U</i> 700 (m s ⁻¹)	1.09	2.59	<i>SLP</i> (hPa)	-2.46	2.83

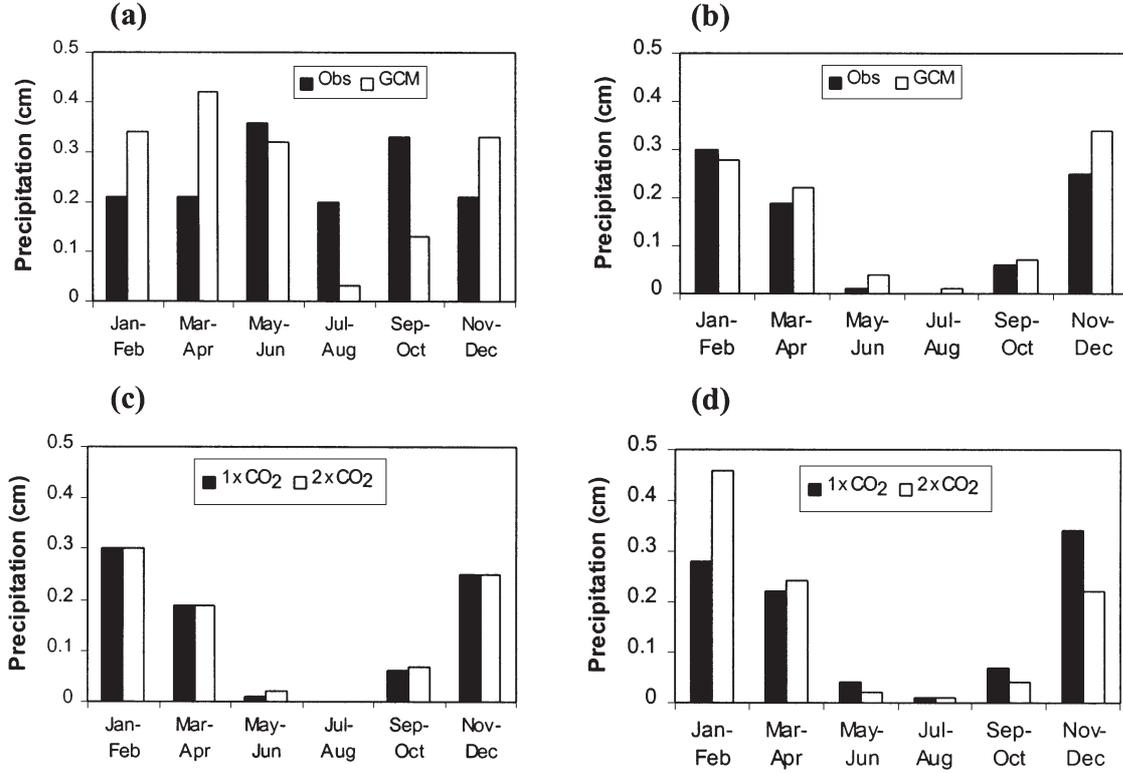


Fig. 5. Summary of daily mean precipitation by season for current climate of (a) San Antonio and (b) San Francisco. Modeled future precipitation in San Francisco is presented in (c) for the TSR model and in (d) for direct GCM output

and 500 hPa. The U , V , T , and HT variables are generally well simulated. For San Antonio, the GCM predicts $RH850$ well in the cold season but still under-predicts in the warm season. The annual cycle of $RH850$ is well simulated in San Francisco, although the GCM over-predicts somewhat in the cold season. Since our TSR models rely heavily on the relative humidity, we choose to apply our models to the future climate scenario for only the San Francisco site.

2.3.2. Predicting $2 \times \text{CO}_2$ large-scale variables

Climate and weather models are known to have systematic errors when predicting large-scale variables. One statistical technique to address these errors is the well-known MOS method. To reduce the systematic error the difference of GCM output ($2 \times \text{CO}_2$ minus $1 \times \text{CO}_2$) was used in place of the direct $2 \times \text{CO}_2$ output. For predicting precipitation under doubled CO_2 the required large-scale variables were calculated by adding the GCM-predicted differences in these variables to the observed $1 \times \text{CO}_2$ variables.

To calculate the $2 \times \text{CO}_2$ large-scale variables, the difference of the $2 \times \text{CO}_2$ and $1 \times \text{CO}_2$ GCM AMY data set was first calculated. The difference data set was

then stratified into 6 seasonal data sets. In the next step, the means and standard deviations of the seasonal data sets were calculated. An examination shows that the means tend to be smaller than the standard deviation, of the difference data set. To investigate the role of the standard deviation, we calculated $2 \times \text{CO}_2$ large-scale variables by adding the mean difference and a random fluctuation component to the observed seasonal data sets:

$$\text{Pred}_{2 \times \text{CO}_2} = \text{Obs}_{1 \times \text{CO}_2} + (\bar{\Delta}) + N(0,1) \cdot \sigma_{\Delta} \quad (3)$$

where

$$\Delta \equiv (\text{GCM}_{2 \times \text{CO}_2} - \text{GCM}_{1 \times \text{CO}_2}) \quad (4)$$

and $N(0,1)$ is a normally-distributed random series with a zero mean and a unit standard deviation. An alternate $2 \times \text{CO}_2$ prediction was calculated by removing the random term in Eq. (3) to test the influence of this term. Table 5 shows a sample of the mean and the standard deviation of the seasonal difference data set for the San Francisco site.

2.3.3. Downscaled precipitation results

Seasonal precipitation scenarios after CO_2 doubling for San Francisco were obtained by substituting the

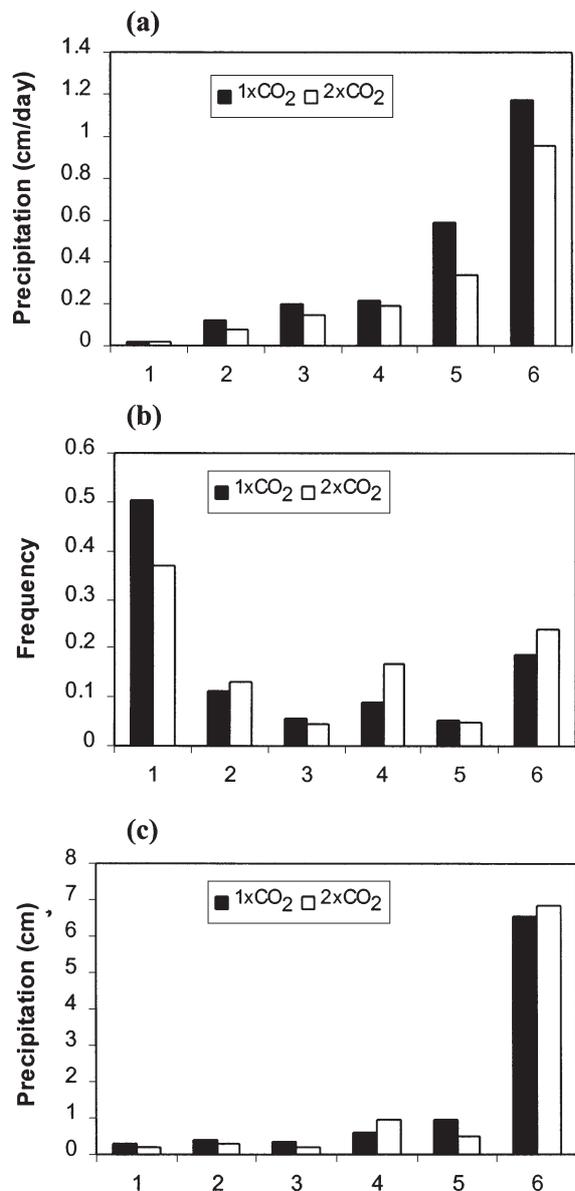


Fig. 6. Precipitation prediction from TSR models for the January-February model of San Francisco. (a) Precipitation intensity of weather patterns (WPs), (b) precipitation frequency, and (c) contribution to total precipitation of each WP

calculated $2 \times \text{CO}_2$ large-scale variables into the 6 seasonal models. The seasonal summaries are plotted in Fig. 5c. Direct GCM output and observations are plotted Fig. 5d for comparison. These plots show that the GCM captures the seasonal characteristics such as the very dry summer and wet winter. However, the GCM does show some deviations with respect to the amount of precipitation. For example, it under-predicts the November-December precipitation by about 36%. Because the TSR precipitation results for $1 \times \text{CO}_2$ represent a sorting of observational data into categories,

the TSR mean precipitation results for $1 \times \text{CO}_2$ are identical to the mean observations. As for the predicted change in seasonal mean precipitation under climate change, the TSR models suggest almost no change at all, while the GCM predicts discernible changes for most seasonal mean precipitation. For example, the GCM predicts a 64% increase in the January-February season and a 35% decrease in the November-December season. Because the tree models are so strongly dependent upon relative humidity, the minimal TSR-predicted impacts on precipitation can be traced to small changes in *RH* predicted by the GCM. For yearly mean precipitation, both GCM and TSR models show a small increase.

While the TSR models predict virtually no change in mean precipitation for all the seasons, they do show some important changes in the precipitation patterns for most seasons. Fig. 6 shows a sample of the change in the seasonal precipitation patterns for San Francisco. Fig. 6a is the precipitation intensity (cm d^{-1}) from 6 different WPs, with WP1 having the lowest average precipitation. It shows that the precipitation intensity decreases for almost all WPs after doubling CO_2 . Fig. 6b shows the frequency of the 6 WPs, with WP1 having the highest occurrence probability. It can be seen that WP2, WP4 and WP6 occur more often after doubling CO_2 . Fig. 6c shows the monthly (30 d) total precipitation contribution of the different WPs. It shows that precipitation from WP6 constitutes most of the precipitation in the January-February season for San Francisco, about 72% for the current climate. The contribution of WP6 is roughly unchanged after doubling CO_2 , suggesting the increase in its frequency offsets the decrease in its intensity. Overall, the net result is no significant change in total precipitation, but the changes in the distribution of precipitation events from larger to smaller events could have significant implications for flooding, runoff, and the general hydrology of the region.

The alternate $2 \times \text{CO}_2$ prediction approach (without the stochastic term) was also applied to the TSR models, with no significant impact on the modeling predictions. This indicates that the TSR models are relatively insensitive to random noise.

3. DISCUSSION AND CONCLUSIONS

This paper has presented a methodology of using TSR to downscale daily precipitation from large-scale GCM output. Relatively strong models were obtained compared with other published downscaling studies. Possible reasons behind the modeling improvement may be a combination of the extensive predictors included in the model, resulting in improved physical representation and the effectiveness of the TSR tech-

nique. It should be noted that this improvement is limited due to the spatial and temporal resolution of currently available data. Further improvement in physical representation can be achieved when higher resolution observational data become available. Although cluster analysis and TSR are both able to objectively generate WPs, there is a distinctive difference in that the WPs generated by cluster analysis are not optimized to predict any surface variable, including precipitation. A good judgement of how many clusters are sufficient is also required for cluster analysis.

The TSR models presented here rely heavily on the relative humidity at the 850 and 700 hPa levels. Because relative humidity at these 2 levels tends to be high when precipitation occurs, it is reasonable that they were found to be important indicators for precipitation events. Although the GCM used in this study performed well for predicting most other variables over the San Antonio region, it did not do well with respect to predicting relative humidity over the region. Hence it is not worthwhile downscaling this GCM simulation to predict $2 \times \text{CO}_2$ precipitation over San Antonio. This suggests that great caution should be taken when applying downscaling models for climate change. Specifically, predictors entering the final downscaling models must be carefully screened for accuracy. A future GCM simulation with better relative humidity prediction would lend more credibility to the TSR model results under a perturbed climate scenario. Nevertheless, the TSR models generated and presented here are based solely on observational data and are very strong. Hence, when improved GCM model runs do become available, the present models can be applied to the new GCM output in a straightforward fashion.

It is somewhat surprising that the models predict almost no change for all the seasons. One possible explanation may be that the change in large-scale variables predicted by the GCM is rather small. The average change in relative humidity is about -0.03 for the 3 levels for both sites. A small change in the predictors, however, did alter the precipitation patterns.

Acknowledgements. This research was funded in part by the US Department of Energy's (DOE) National Institute for Global Environmental Change under DOE cooperative agreement No. DE-FC03-90ER61010. Financial support does not constitute an endorsement by DOE of the views expressed in this article.

LITERATURE CITED

- Anthes RA (1983) Regional models of the atmosphere in middle latitudes. *Mon Weather Rev* 111:1306–1335
- Bardossy A, Plate EJ (1992) Space-time model for daily rainfall using atmospheric circulation patterns. *Water Resour Res* 28:1247–1259
- Breiman L, Friedman JH, Olsen RA, Stone JC (1984) Classification and regression trees. Wadsworth, Belmont, CA
- Cubasch U, von Storch H, Waszkewitz J, Zorita E (1996) Estimates of climate change in southern Europe derived from dynamical climate model output. *Clim Res* 7:129–149
- Enke W, Spekat A (1997) Downscaling climate model outputs into local and regional weather elements by classification and regression. *Clim Res* 8(10):195–207
- Fuentes U, Heimann D (1996) Verification of statistical-dynamical downscaling in the Alpine region. *Clim Res* 7: 151–168
- Giorgi F, Mearns LO (1991) Approaches to the simulation of regional climate change: a review. *J Geophys Res* 29: 191–216
- Glahn HR, Lowry DA (1972) The use of model output statistics (MOS) in objective weather forecasting. *J Appl Meteorol* 11:1203–1211
- Gyalistras D, Von Storch H, Fischlin A, Beniston M (1994) Linking GCM-simulated climatic changes to ecosystem models. *Clim Res* 4:167–189
- Hewitson BC, Crane RG (1996) Climate downscaling: techniques and application. *Clim Res* 7:85–95
- Hu T (1998) A study of applying artificial neural network techniques to regional climate downscaling under global climate change. Master's thesis, School of Engineering, Tulane University, New Orleans
- Hughes J, Lettenmaier D, Guttorp P (1993) A stochastic approach for assessing the effect of changes in regional circulation patterns on local precipitation. *Water Resour Res* 29:3303–3315
- IPCC (1995) Climate change 1995. Zinyowera MC, Watson RT Moss RH (eds) Cambridge University Press, Cambridge
- Karl TR, Wang WC, Schlesinger ME, Knight RW, Portman D (1990) A method of relating General Circulation Model simulated climate to the observed local climate. Part I: seasonal statistics. *J Climate* 3(10):1053–1079
- Matyasovszky I, Bogardi I, Duckstein L (1994) Comparison of two general circulation models to downscale temperature and precipitation under climate change. *Water Resour Res* 30(12):3437–3448
- NCAR Science Briefing (1997) NCAR Staff Notes Monthly, Vol 32, No. 5, May 1997. National Center for Atmospheric Research, Boulder
- Noguer M (1994) Using statistical techniques to deduce local climate distributions. An application for model validation *Meteorol Appl* 1:277–287
- Portman DA, Wang W-C, Karl TR (1992) Comparison of general circulation model and observed regional climates: daily and seasonal variability. *J Clim* 5:343–353
- Sailor DJ, Li X (1999) A semi-empirical downscaling approach for predicting regional temperature impacts associated with climatic change. *J Clim* 12(1):103–114
- Shuman FG (1989) History of numerical weather prediction at the National Meteorological Center. *Weather Forecast* 4: 286–296
- Singh VP, Chowdhury PK (1986) Comparing some methods of estimating mean areal rainfall. *Water Resour Bull* 22: 275–281
- Steinberg D, Colla PL (1995) CART: tree-structured nonparametric data analysis. Salford Systems, San Diego, CA
- von Storch H, Zorita E, Cubasch U (1993) Downscaling of climate change estimates to regional scales: application to winter rainfall in the Iberian Peninsula. *J Clim* 6:1161–1171
- Weichert A, Burger G (1998) Linear versus nonlinear techniques in downscaling. *Clim Res* 10(8):83–93
- Wilby RL, Wigley TML (1997) Downscaling general circulation model output: a review of methods and limitations.

Prog Phys Geogr 21:530–548

Winkler JA, Palutikof JP, Andresen JA, Goodess CM (1995) The simulation of daily time series from GCM output. Part II: Development of local scenarios for maximum and minimum temperature using statistical transfer functions. In: 6th Symposium on Global Change Studies, Dallas. American Meteorological Society, Dallas, TX

*Editorial responsibility: Hans von Storch,
Geesthacht, Germany*

Zorita E, von Storch H (1999) The analog method as a simple statistical downscaling technique: comparison with complicated methods. J Clim 12(8):2474–2489

Zorita E, Hughes JP, Lettenmaier DP, von Storch H (1995) Stochastic characterization of regional circulation patterns for climate model diagnosis and estimation of local precipitation. J Clim 8:1023–1042

*Submitted: September 1, 1999; Accepted: March 16, 2000
Proofs received from author(s): August 25, 2000*