A Statistical Model of Climates in the Southwestern United States

Richard G. Craig and John F. Stamm

ABSTRACT: We describe an empirical-statistical model of climates of the southwestern United States. Boundary conditions include sea surface temperatures, atmospheric transmissivity, and topography. Independent variables are derived from the boundary conditions along 1000-km paths of atmospheric circulation. Upper (400 mb) and lower level (800 mb) atmospheric independent variables describe available moisture and heat. Lower level atmospheric variables also describe orographic controls. Other independent variables represent climatic controls at the surface, such as elevation and slope. Mean monthly temperature and total monthly precipitation are the predicted variables. Canonical regression is applied to avoid problems of co-linearity. Predictor equations are derived over a larger region than the application area to allow for the increased range of paleoclimate. This larger region is delimited by the autocorrelation properties of climatic data.

INTRODUCTION

The level of climatic information required to understand the hydrologic system of the southwestern United States is difficult to reconstruct from general circulation model (GCM) output because of the coarse scale compared to the phenomena of interest (Kutzbach, 1983). Such models are of interest because they offer the advantage of providing information for any period for which the requisite boundary conditions can be specified (COHMAP, 1988). A possible supplement to GMC solutions is a statistical analysis that explicitly computes climate in terms of controlling factors in that area (Roberts, Craig and Stamm, 1989). We describe here an extension of that approach that more fully expresses the influence of the boundary conditions and allows greater flexibility in its application. In this study, we use canonical regression (Glahn, 1968). Such a modeling procedure has been successfully applied to paleoclimatic studies (Fritts, et al., 1971; Webb and Bryson, 1974).

THE STUDY AREA

We differentiate two tasks for this modeling: (1) solution of the climate equations, and (2) calibration of the equations. The area used to calibrate the equations is much larger than the area in which the equations are solved to ensure a robust set of equations as explained below.

The climate equations are solved at all points in the **level II prediction area** (Figure 1). These predictions draw upon information (to compute independent variables) from an area up to 1000 km upwind (the **level I prediction area**).

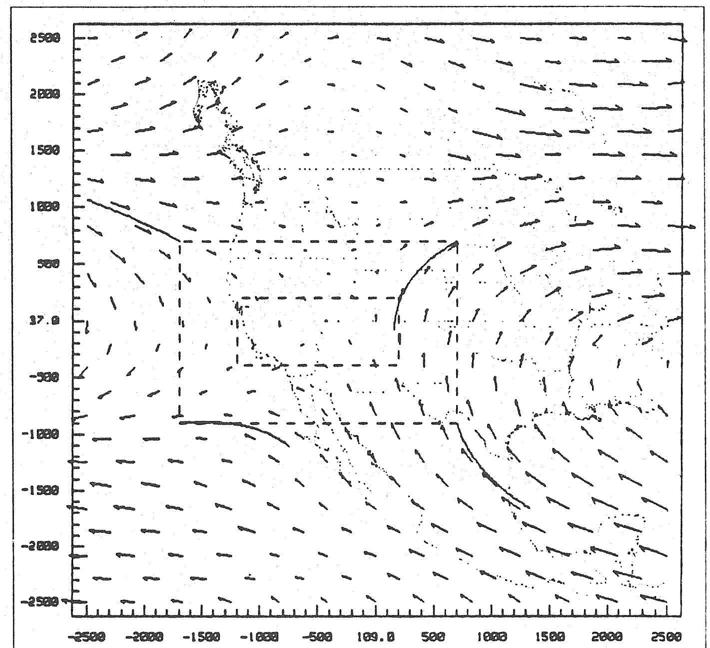

Figure 1. Level II calibration *(large dashed rectangle)* and solution *(small dashed rectangle)* areas. Arrows represent 800 mb July wind vectors interpolated from Schutz and Gates (1972). Curved lines represent wind trajectories along which independent variables are determined (axis units: km from central origin).

Winter 500 mb geopotential height departures from transient means (Figure 2) clearly show a decrease in the correlation between stations as the distance increases (Thiebaux and Pedder, 1987, p. 143). At about 1500 km distance, the averaged correlation values become negative. Because of the spread of the data, it is not clear that points more distant than about 1000 km are positively correlated. Therefore, trajectories of 1000 km are used to compile the independent variables.

To calibrate the climate equations, we select stations whose range of independent variables is considerably larger than that currently observed in the solution area. We assume that the most extreme climates to be

In J.L. Betancourt and A.M. MacKay, editors, 1990. Proceedings of the Sixth Annual Pacific Climate (PACLIM) Workshop, March 5-8, 1989: California Department of Water Resources, Interagency Ecological Studies Program Technical Report 23.

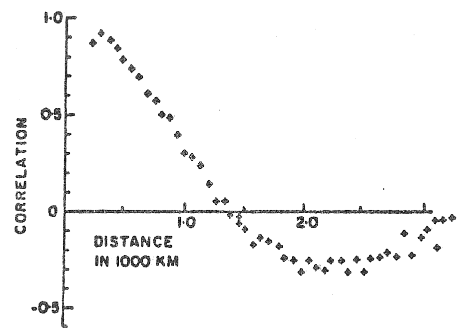

Figure 2. Correlation between geopotential height transient means as a function of distance of separation of stations *(from Thiebaux and Pedder, 1987).*

predicted are represented at stations within 1000 km upwind of the solution area. The area containing selected climate stations is called the **level II calibration area.**

The level I calibration area extends 1000 km upwind from the level II calibration area and is of sufficient size to provide boundary conditions needed to calculate independent variables in the level II calibration area. Figure 1 shows wind paths up to 1000 km upwind from the level II calibration area corners during July.

BOUNDARY CONDITIONS FOR CLIMATIC MODELING

Independent variables used in the climate equations are based on one or more of the following boundary conditions:

- Elevation
- Upper level (400 mb) winds, winter
- Upper level winds, summer
- Surface (800 mb) winds, winter
- Surface winds, summer
- Sea surface temperatures, winter
- Sea surface temperatures, summer
- Solar isolation

Elevation is represented by a digital elevation model (DEM) with a 10-km-square grid. This grid was interpolated from elevation and bathymetric data at a spacing of 5 minutes of latitude and longitude (National Geophysical Data Center, 1988).

Windfields for modern winter and summer are interpolated from first order U.S. Weather Bureau instrumental stations. We employ winds at two levels: near the surface (800 mb) and in the upper atmosphere (400 mb).

The windfield of the local climate model is interpolated with a 3-dimensional, objective-analysis mesoscale windfield model. This model was originally part of a meteorological data processor for an air quality monitoring program, MELSAR (MEsoscale Location Specific Air Resources) (Allwine and Whiteman, 1985). Given the horizontal velocity components at discrete points in the domain, the model generates an orthogonal polynomial that represents horizontal velocity components continuously through the domain. Horizontal velocity components are then computed under constraint of conservation of mass. Figure 1 is an example of the output of the model.

Instrumental records allow direct calibration of the climatic equations with SST data representing the same years as the climatic data. We use those data to calibrate four polynomial equations. Two seasons, winter and summer, are used. Two functional surfaces are computed for each season, one for the Pacific Ocean and one for the Gulf of Mexico.

The polynomial equations provide a compact representation of SST, which minimizes storage requirements while accurately reproducing the original data. An R^2 of 99 percent is achieved, and the patterns of variability are well described. They allow estimation of SST at any point. This is important for applications along wind vector paths that do not lie at grid point locations. Sea surface temperatures are represented with CLIMAP (McIntyre, et al., 1981) and other data sources.

We limit our estimates of insolation to the top of the atmosphere and solve for daily insolation from orbital solutions based on the work of Berger (1978a). Inclusion of this term provides greater flexibility in considering climatic solutions for boundary conditions representing other climatic states.

We also use the orbital parameter calculations of Berger (1978b) to compute sun angle (declination and azimuth) and combine this information with slope derived from the DEM to determine the about of solar energy received at each point of the level II area.

Heat energy added to an air mass upwind of a chosen point is approximated with an energy budget approach (Pease, 1987). The algorithm takes into account the incoming short-wave solar energy and long-wave energy emitted from the earth. This information is used to estimate the average airmass temperature along a 1000 km trajectory upwind of the point in question.

For this energy budget computation, certain assumptions must be made about the emissivity of the earth, the transmissivity of the atmosphere, certain partitioning fractions, and the density of cloud cover. A complete discussion is provided by Pease (1987). This model has the advantage of allowing adjustment for changes in the emissivity of the atmosphere such as could happen with a change in the concentration of carbon dioxide. This concentration has been shown to have varied in the Late Quaternary Period (Neftel et al., 1988). This concentration and the other partitioning fractions become parameters of the model.

Thermal energy input to the air mass from the upper level of the ocean is modeled by setting the surface energy in the energy balance model to the sea surface temperature. Planetary temperature and the non-radiative flux parameter, gamma, are adjusted to that SST.

INDEPENDENT VARIABLES

The canonical regression equation is based on independent variables representing the physical factors that can influence climate in the area while satisfying the constraint that they must be available for times in the geologic past. This constraint is exploited in other applications not discussed in this paper. We further constrain the independent variables to be calculated from boundary conditions within 1000 km of the calibration or solution point.

The independent variables are divided into three groups:

- Surface variables - calculated at the calibration or solution point.
- Lower-air variables - representing the influence of the boundary conditions from points up to 1000 km upwind from the calibration or solution point along the lowest conformal surface ($\sim$800 mb).
- Upper-air variables - representing the influence of the boundary conditions from points up to 1000 km upwind from the calibration or solution point along the upper conformal surface ($\sim$400 mb).

Considering that some variables are computed for each month, there are presently 72 independent variables used in the canonical regression procedure. The variables are described below.

Surface Variables:

a. Climate station elevation interpolated from DEM.
b. Maximum slope of the terrain surface.
c. Normalized horizontal components of the maximum slope of the terrain surface.
d. The angle between the horizontal wind direction and the azimuth of the maximum slope of the terrain surface.
e. Vertical component of wind velocity.
f. Surface temperature, from energy balance model.
g. Planetary temperature, from energy balance model.
h. Day length (sunrise to noon) at mid-month.
i. Mid-month insolation.

Variables *d* and *e* are computed for both summer and winter, and variables *f* through *i* are computed for each month, yielding 54 independent variables.

Variable *a* represents the important effects of elevation. Variables *b* and *c* represent the microclimatic influence of slope. Variable *d* represents the angle of approach of a storm on terrain and the associated efficiency of precipitation. Variable *e* represents the potential for convective storms. Variables *f* through *i* represent the local effects of insolation.

Lower-Air Variables:

a. Maximum [elevation/ln (distance to the elevation)] along upwind trajectory from the calibration/solution point.
b. Maximum elevation along upwind trajectory from the calibration/solution point.
c. Minimum elevation along upwind trajectory from the calibration/solution point and downwind from the location of variable *b*.
d. Distance from calibration/solution point to variable *b*.
e. Distance from calibration/solution point to variable *c*.
f. Average surface temperature along trajectory using energy budget model.
g. Percent distance over oceans/lakes along trajectory.

These variables are computed for both winter and summer, yielding 14 independent variables.

Variables *a*, *b*, and *d* represent the controls on orographic precipitation. Variables *c* and *e* represent adiabatic cooling that an air mass undergoes after it has passed an orographic depression. Variable *a* represents effects of local barriers, while variables *b* through *e* represent effects of distant barriers. Variable *f* represents possible horizontal advection of heat. Variable *g* represents the amount of moisture input from points upwind.

Upper-Air Variables:

a. Average surface temperature along trajectory using energy budget model.
b. Percent distance over oceans/lakes along the trajectory.

These variables are computed for both winter and summer, yielding four independent variables. These variables have the same physical representation as lower-air variables *f* and *g*.

STATUS

We have used a data set consisting of monthly mean maximum temperature and total monthly precipitation for 180 stations in the Great Basin (Wernstedt, 1972) to compute a canonical regression. For these stations, elevations ranged from -16 m to 2746 m, with a mean of 786 m and standard deviation of 734 m. Climatic variables, except for summertime precipitation, are highly correlated with one another (Table 1). We infer that at least two canonical variates must be employed.

Table 1. Squared multiple correlation of variables in the first set with all other variables in the first set.

	Temperature	Precipitation
January	0.99	0.99
February	0.99	0.98
March	0.99	0.98
April	0.99	0.97
May	0.99	0.97
June	0.99	0.92
July	0.99	0.89
August	0.99	0.88
September	0.99	0.82
October	0.99	0.98
November	0.99	0.98
December	0.99	0.98

Seventy-two independent variables were available for analysis. Because of co-linearity, only 26 independent variables were used in the canonical correlation procedure. Slope-related variables were not used, except for slope itself. Some mid-month day lengths were not used; summertime values were important. All insolation values except December were excluded. All planetary temperature variables were removed. Percent over ocean for surface variables were not needed.

Tests of the eigenvalues are reported in Table 2. Using Bartlett's test of sphericity, we found that seven canonical variates are significant (at the 0.01 level). Three were retained to develop regression coefficients, scores and loadings.

Table 2. Bartlett's test of number of eigenvalues needed.

	Chi-Square	d.f.	Tail Probability
	1999.34	624	0.0000
1	1518.60	575	0.0000
2	1102.34	528	0.0000
3	817.17	483	0.0000
4	643.39	440	0.0000
5	524.33	399	0.0000
6	426.37	360	0.0091
7	346.66	323	0.1748

Adjusted squared multiple correlations of temperature and precipitation with the chosen canonical variates (Table 3) ranged from 17 percent (September precipitation) to 82 percent (January and December temperatures). All adjusted R^2 values are significant at the 0.0005 probability level.

Table 3. Adjusted squared multiple correlations of each variable in the first set with chosen canonical variables of the second set. For all variables, the degrees of freedom are 26 and 153, and the F-statistic is significant at the 0.0005 level.

Variable	Adj. R^2	F-stat.	Variable	Adj. R^2	F-stat.
TJAN	0.82	33.22	PJAN	0.49	7.83
TFEB	0.79	27.59	PFEB	0.49	7.80
TMAR	0.75	21.86	PMAR	0.50	8.05
TAPR	0.70	17.85	PAPR	0.47	7.23
TMAY	0.66	14.89	PMAY	0.55	9.52
TJUN	0.62	12.33	PJUN	0.63	13.07
TJUL	0.53	8.82	PJUL	0.54	9.12
TAUG	0.54	9.33	PAUG	0.54	9.11
TSEP	0.58	10.67	PSEP	0.17	2.46
TOCT	0.68	15.84	POCT	0.38	5.38
TNOV	0.78	25.54	PNOV	0.44	6.59
TDEC	0.82	33.01	PDEC	0.50	8.11

Loadings of the independent variables on the canonical variates indicate that the first canonical variate has high loadings from nearly all variables. The second canonical variate has highest loadings from the day-length variables.

Wintertime temperatures tend to load highest on the first canonical variate. Of the precipitation variables, summer months tend to load highest on this variate (with negative loadings). Summer temperatures and all precipitations except June, July, and September load highest on the second canonical variate. Loadings on the third canonical variate are also high for summer temperatures and spring precipitation. Temperature loadings on that variate are negative and positive for all precipitation variables.

CONCLUSIONS

Empirical-statistical models capable of describing a significant portion of the variability of average monthly temperature and precipitation in the southwestern United States are feasible. These models can define the importance of various orographic and synoptic variables that can be solved from a fundamental set of boundary conditions, which can be specified by a GCM. The availability of a local climate model allows a meaningful comparison of climate forecasts to field observations.

ACKNOWLEDGEMENTS

This research has been supported by the Desert Research Institute of the University of Nevada, Reno through a grant from the National Science Foundation. We appreciate the support of personnel at DRI, including Cynthia Irwin-Williams, Jonathan Davis, Gil Cochran, and Peter Wigand.

REFERENCES

Allwine, K.J. and Whiteman, C.D., 1985, MELSAR: A Mesoscale Air Quality Model for Complex Terrain. Volume 1 - Overview, Technical Description and User's Guide, Pacific Northwest laboratories, Battelle Memorial Institute. PNL-5460 v.1.

Berger, A.L., 1978a, Long-term Variations of Daily Insolation and Quaternary Climatic Changes, Journal of Atmospheric Sciences, v.35, p.2362-2367.

Berger, A.L., 1978b, A Simple Algorithm to Compute Long-Term Variations of Daily or Monthly Insolation (Contribution degree 18), Universite Catholique De Louvain, v.18.

COHMAP Members, 1988, Climatic Changes of the Last 18,000 Years: Observations and Model Simulations, Science, v.241, p.1043-1052.

Fritts, H.C., Blasing, T.J., Hayden, B.P., and Kutzbach, J.E., 1971, Multivariate Techniques for Specifying Tree-Growth and Climate Relationships and for Reconstructing Anomalies in Paleoclimate, Journal of Applied Meteorology, v.10, p.645-864.

Glahn, J.R., 1968, Canonical Correlation and its Relation to Discriminat Analysis and Multiple Regression, Journal of Atmospheric Sciences, v.25, p.23-31.

Kutzbach, J.E., 1983, Modeling of Holocene Climates, in Late-Quarternary Environments of the United States, ed., Wright, H.E. Jr., v.2, The Holocene, p.271-277.

McIntyre, A., Cline, R. (Editor), Hayes, J., Prell, W., Moore, T., Kipp, N., Molfino, B., Denton, G., Kukla, G., Matthews, R., Imbrie, J., and Hutson, W. (CLIMAP Project Members), 1981, Seasonal Reconstructions of the Earth's Surface at the Last Glacial Maximum, The Geological Society of America Map and Chart Series, MC-36.

National Geophysical Data Center, 1988, Bathymetry/Topography Data, Newly Revised. Data Announcement 88-MGG-02.

Neftel, A., Oeschgeer, H., Staffelback, T. and Stauffer, B., 1988, CO_2 record in the Byrd ice core 50,000-5,000 years BP, Nature, v.331, p.609-611.

Pease, R.W., 1987, The Average Surface Temperature of the Earth: An Energy Balance Approach, Anals of the Association of American Geographers, v.77, no.3, p.450-461.

Roberts, B.L., Craig, R.G., and Stamm, J.F., 1989, A Microcomputer Reconstruction of Paleoclimates, <u>in</u> Microcomputer Applications in Geology, v.II, J.T. Hanley and D.F. Merriam, editors, Pergamon Press.

Schutz, R.P. and Gates, W.L., 1972, Global Climatic Data for Surface, 800mb, 400mb: July, ARPA, RAND 1029-ARPA, 180p.

Thiebaux, H.J. and Pedder, M.A., eds., 1987, Spatial Objective Analysis: With Applications in Atmospheric Science: Academic Press, 299p.

Webb, T. and Bryson, R.A., 1974, Late- and Postglacial Climatic Change in the Northern Midwest, USA: Quantitative Estimates Derived from Fossil Pollen Spectra by Multivariate Statistical Analysis, Quaternary Research, v.2, p.70-115.

Wernstedt, F.L., 1972, World Climatic Data, VI. North America: Lemont, PA, Climatic Data Press.