Spatio-temporal precipitation modeling based on time-varying regressions

Oleg Makhnin
Department of Mathematics, New Mexico Tech, Socorro, NM 87801

Abstract

A time-varying regression model is considered, based on monthly precipitation data from gauge measurements. The model accounts for orographic effects, that is elevation and *aspect* of the terrain. The study area is NCDC climate division 2 in a mountainous area in northern New Mexico. We assess spatio-temporal variability and also trace the dependence of precipitation on El Niño/Southern Oscillation (ENSO) index.

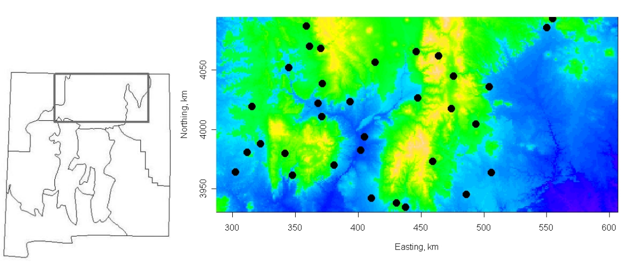
1 Introduction

In many studies (see, e.g. Gershunov and Barnett (1998)), a question was raised of teleconnections of ENSO (El Niño/Southern Oscillation) with precipitation in Southwestern US. In Guan, Vivoni and Wilson (2005), an instant of such teleconnection was reported. In particular, they looked at three categories of years (ENSO High, Low and Neutral) and observed, for example, a positive precipitation anomaly for High and Neutral ENSO in the winter.

A question was also raised about the relationship of precipitation to PDO (Pacific Decadal Oscillation). However, at the time scales for the PDO (decades) we don't have enough data to reliably assess this relationship.

This work attempts to assess ENSO influence on a more statistical footing by capturing the spatiotemporal variability of precipitation in the mountainous region in northern New Mexico. The region is chosen because it has a great impact on the water supply in the state of New Mexico. Also, mountainous terrain affects precipitation in a certain way.

The study region corresponding to New Mexico NCDC climate Division 2 is shown below. The NCDC rain gauges provide direct measurement of precipitation over various locations in the area. We picked 33 stations for which mostly uninterrupted records are available from 1970 to 2003. The data are total monthly precipitation measurements at these stations.



2 Model

We fit time-varying regressions to the square-root transformed values of precipitation P_{jt} for Station j and Month t:

$$P_{jt} = \beta_0^t + \beta_1^t E_j + \beta_2^t N_j + \beta_3^t Z_j + \beta_4^t cos(A_j) + \\ + \beta_5^t sin(A_j) + S_{mod(t,12)} + \beta_{SOI} L_t + \tau_t + \varepsilon_{jt}, \\ j = 1, ..., N, \quad t = 1, ..., T \quad (1)$$

Here, E_j and N_j are easting and northing coordinates of a station (in km), Z_j is the elevation of a station, $S_{mod(t,12)}$ are seasonal corrections (January through December, constant for each month throughout the study period). The term L_t contains the value of SOI (Southern Oscillation Index, reported at http://www.bom.gov.au/climate/current/soihtm1.shtml) which will serve as a proxy for ENSO.

The terms $\beta_4\, cos(A_j)$ and $\beta_5\, sin(A_j)$ account for the "moisture flux direction" (MFD) effect (see Guan, Wilson and Makhnin (2005)). They provide auto-search for the MFD W with

$$\beta_4\, cos(A_j) + \beta_5\, sin(A_j) = \beta_{\mathtt{MFD}}\, cos(A_j - W) \quad (2)$$

The term $\beta_{\mathtt{MFD}}\, cos(A_j - W)$ captures interaction of the MFD with the terrain aspect A_j, which is the gradient direction of the terrain at station j, averaged in a $5km$ window. When the moisture is coming up slope (the directions of A_j and W coincide), this results in extra precipitation. The term W inferred in our model is a statistical average over potentially many precipitation events.

In Anandkumar (2005), the random field $W(\cdot)$ was introduced, varying over the region. However, for our purposes, the region is fairly small, therefore we assume here that W is constant, and the relation (2) is used instead. However, for larger regions, working with the random field MFD will be critical.

The terms with β_1 and β_2 account for linear Moisture Gradient (MG) throughout the region. It does not necessarily coincide with MFD.

The terms τ_t are random effects for the month t, and ε_{jt} are residual errors (possibly correlated).

Instead of actual precipitation measurement we have used square-root transformed precipitation P_{jt}. It is a popular choice of transformation and attains near normality of the transformed values. The covariates E, N, Z are coded; that is, they are scaled to have mean 0. This helps eliminate unwanted correlations between regression coefficients.

2.1 Time-varying regression coefficients

The coefficients $\beta_k^t,\ k = 0, 1, ..., 5$ depend on t. However, we allow for some degree of smoothing by introducing the autoregressive evolution equations, for each k:

$$\beta_k^{t+1} = \mu_k + r_k(\beta_k^t - \mu_k) + \Delta\beta_k^t \qquad t = 1, ..., T-1 \quad (3)$$

(note that β_{SOI} is not time-varying).
The increments $\Delta\beta_k^t$ are assumed to be $\mathcal{N}(0, q_k^2)$ with additional variance parameters q_k^2, $k = 0, ..., 5$.

Note that $r_0 = 1$ and $\mu_0 = 0$, for identifiability purposes. This way, the mean precipitation value for the entire region for a given month t is split into a slow-varying component β_0^t and a random perturbation τ_t.

2.2 Other parameters

The seasonals $S_{mod(t,12)}$ are currently computed as a simple average of the P_{jt} values. Later, we hope to include them into the MCMC iterative scheme.

The residuals ε_{jt} are assumed to be temporally independent. It is well known that the spatial dependence exists (e.g. Guan, Wilson and Makhnin (2005)). We describe spatial dependence for ε_{jt} based on exponential covariance model

$$Cov(\varepsilon_{it}, \varepsilon_{jt}) = \sigma^2[exp(-dist(i,j)/\phi) + w^2]$$

where $dist(i,j)$ is the Euclidean distance between stations i and j. Currently we fit the values of the range $\phi = 30\,km$ and relative nugget $w^2 = 1/3$. Later we will introduce the estimation of ϕ, w^2 into the MCMC sampler.

Some data were missing. In the MCMC framework, it is straightforward to impute the missing data using full conditional posteriors, through equation (1).

3 MCMC fit

The parameters in the model are fitted using Markov Chain Monte Carlo approach. It is implemented via Gibbs sampler. The full conditional posteriors (FCP) for the parameters are indicated below.

The FCP for the entire block of $\{\beta_t^k\}$, $\quad t = 1, ..., T, \quad k = 0, ..., 5$, given all the other parameters from equations (1) and (3), can be computed using forward-filtering backward-sampling (FFBS) approach described in West and Harrison (1997).

The variance parameters q_k^2 are fitted using inverse chi-square (conjugate) prior with the location parameter ζ_k and ν_k degrees of freedom, similarly to Kim et al. (1998). Then, the FCP distribution of q_k^2 is inverse chi-square with the location parameter $\left(\zeta_k + \sum_{t=1}^{T-1}(\Delta\beta_t^k)^2\right)/(\nu_k + T - 1)$ and $\nu_k + T - 1$ degrees of freedom. We can choose informative priors for q_k^2 if the shrinking of regression parameters is desired.

Similar analysis can be done for residual variance σ^2 and random-effect variance σ_τ^2.

Sampling of AR coefficients r_k from equation (3) was done using a Metropolis step (see Kim et al. (1998), also for fitting μ_k).

4 Results

The results of an MCMC simulation are presented below. We used the burn-in of 0 iterations and number of MC replicates $M = 50,000$, with every 50th selected for the output.

First, a plot of data and the model fit are shown for the first 120 months

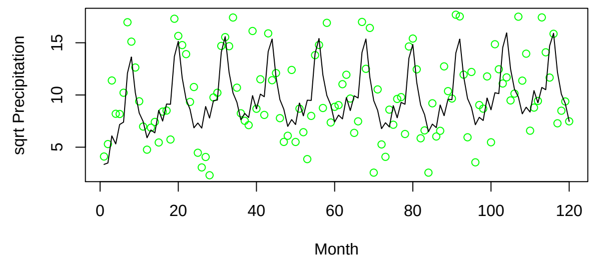


Figure 1: The data and model fit

Next, the MCMC output for $\beta_{\mathtt{SOI}}$ is shown. The graphs shown include a “trace plot" of MC samples, an autocorrelation plot and a histogram of MC samples.

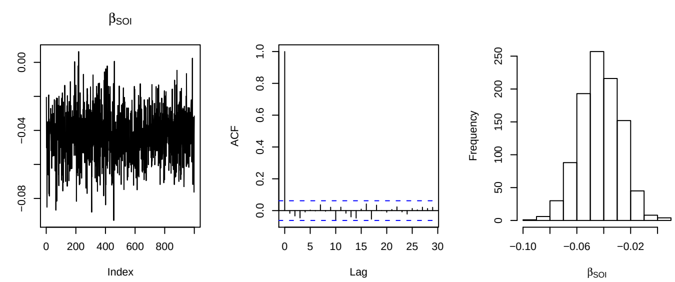


Figure 2: The MCMC output for $\beta_{\mathtt{SOI}}$

These clearly indicate the significant negative relationship between the SOI index values and precipitation.

Next, you can see the posterior means for coefficients β_3^t (elevation), for all months.

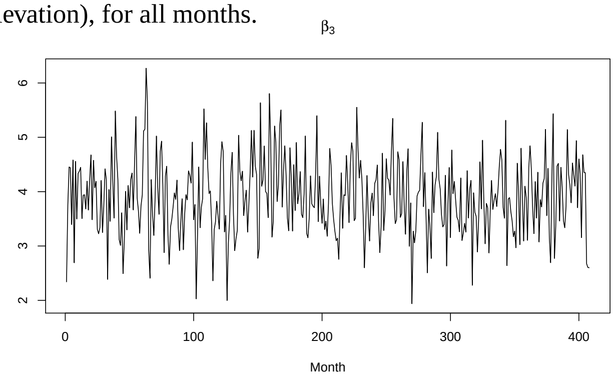


Figure 3: Posterior means of β_3^t, for all Months t.

Consistently positive values for β_3 are indicative of the well-documented relationship between elevation and precipitation.

A clearly expressed seasonal behavior is observed for β_3, as well as MG and MFD:

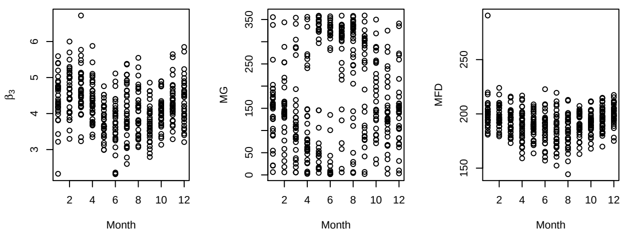


Figure 4: Seasonal behavior of coefficient β_3^t (elevation), MG and MFD

For example, the elevation effects are more significant during the Winter months (higher regression coefficient). Moisture Flux Direction fluctuates in a narrow band between southerly, for Summer months, to south-westerly, for Winter months.

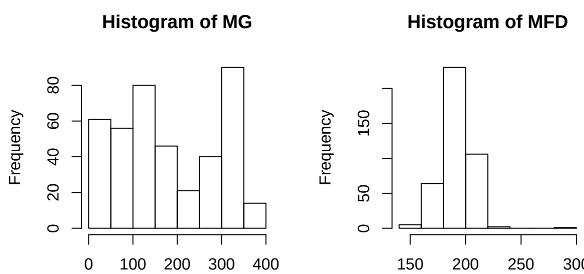


Figure 5: Histograms of posterior means for Moisture Flux Direction (MFD) and Moisture Gradient (MG), all months.

The results indicate a consistent near-southerly MFD for most months (about $180°$, clockwise with $0°$ pointing North), and a somewhat less consistent Moisture Gradient. It would be interesting to further investigate the dependence of MG on the SOI phases.

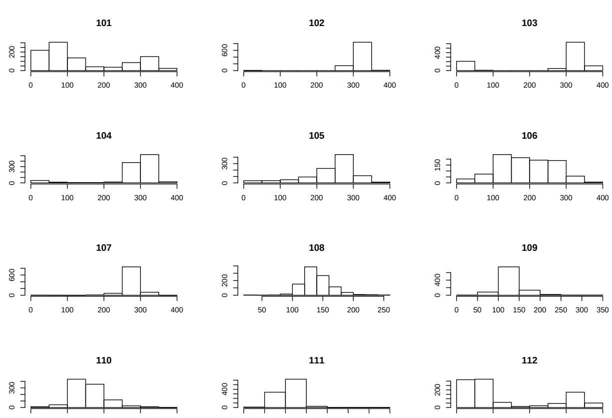


Figure 6: Posterior histograms of Moisture Gradient (MG) for select months

Posterior quantiles of variance parameters:

parameter	5%	25%	50%	75%	95%
σ	2.828	2.848	2.863	2.878	2.897
σ_τ	2.994	3.108	3.182	3.266	3.402
q_0	0.294	0.348	0.387	0.450	0.578
q_1	0.062	0.064	0.065	0.067	0.069
q_2	0.063	0.065	0.066	0.068	0.070
q_3	0.846	0.958	1.062	1.149	1.281
q_4	0.236	0.260	0.281	0.306	0.349
q_5	0.244	0.270	0.294	0.319	0.360

Posterior quantiles of mean parameters:

parameter	5%	25%	50%	75%	95%
μ_1	-0.004	0.002	0.005	0.009	0.015
μ_2	-0.005	-0.001	0.002	0.004	0.009
μ_3	3.951	4.082	4.174	4.276	4.394
μ_4	-0.637	-0.591	-0.560	-0.530	-0.488
μ_5	-0.200	-0.153	-0.121	-0.090	-0.048

The values of μ_4 and μ_5 are indicative of the MFD value staying in a narow band (see Figure 4 above)

Autoregression coefficients r_3 and r_5 were significantly different from 0. This indicates predictability of β regression coefficients for elevation and MFD, likely due to seasonality of the these effects (see Figure 4).

R code and the data used are available from http://www.nmt.edu/˜olegm/JSM06/

5 Conclusions

A time-varying regression model was introduced, describing spatial and temporal variability of the precipitation in a given area. A fairly regular seasonal behavior is observed for some elements in our model, in particular, Moisture Flux Direction.

There is a significant negative influence of SOI values on the average monthly precipitation. Thus, it confirms the hypothesis of teleconnections between ENSO and the climate in northern New Mexico. This has a potential significance for predicting water supply, especially in semi-arid Southwestern US.

6 Acknowledgments

Thanks to Devon MacAllister for collaborating on R programs, and Huade Guan for useful discussions.

References

Anandkumar, S. (2005) *Hidden Random Field Modeling of Orographic Effects on Mountainous Precipitation*, Independent Study Report, New Mexico Tech.

Gershunov, A., and T. P. Barnett (1998), “Interdecadal modulation of ENSO teleconnections", *Bull. Am. Meteorol. Soc.*, **79**, 2715-2725.

Guan, H., Wilson, J., and Makhnin, O. (2005) “Geostatistical Mapping of Mountain Precipitation Incorporating Auto-Searched Effects of Terrain and Climatic Characteristics", *Journal of Hydrometeorology*, **6**, 1018-31.

Guan, H., E. R. Vivoni, and J. L. Wilson (2005), “Effects of atmospheric teleconnections on seasonal precipitation in mountainous regions of the southwestern U.S.: A case study in northern New Mexico", *Geophys. Res. Lett.*, **32**, L23701

Kim, S., Shephard, N., and Chib, S. (1998) “Stochastic Volatility: Likelihood Inference and Comparison with ARCH Models." *Review of Economic Studies*, **65**, 361-394

West, M., and Harrison, J. (1997) *Bayesian forecasting and dynamic models*, Springer-Verlag, New York.