Working paper: Spatial precipitation modeling based on a directional random field

Oleg Makhnin and Anandkumar Shetiya *

July 4, 2007

Abstract

We present a precipitation modeling approach based on "moisture flux direction" (MFD) random field. A Markov chain Monte-Carlo (MCMC) method is proposed to estimate this directional random field. The precipitation data used are point (gauge) measurements in a mountainous terrain in northern New Mexico. We analyze the spatial distribution of moisture flux direction, as inferred from the data.

1 Introduction

Henceforth, s will denote a spatial location, $s \in \mathcal{S}$, t will be typically time. Sections marked by * may be omitted at the first reading.

1.1 Radar and gauge data

In the original proposal of Wilson et al [4], the radar data are taken on a 4x4-km grid and the task of downscaling is performed by estimating the influence of orographic features (elevation and terrain aspect) as covariates, and using the values of these covariates on a finer (1x1-km) grid in the downscaling (interpolation) process. The gauge data are then used for validation purposes.

Regressions that take into account orographic features are fitted locally, within

*Department of Mathematics, New Mexico Tech, Socorro, NM 87801

a specified cluster. The results therefore depend on the clustering algorithm.

In this work, we attempt to create a smooth random field of orographic features that will obviate the need for a clustering algorithm, and allow efficient borrowing of information between the neighboring groups of grid points.

Although this paper is not necessarily geared towards using radar observations (we focus on the methodology of estimating MFD random field), but portions of it (for example, the next section) are done with an eye on radar downscaling as possible application.

2 Model description

2.1 Grid modeling approach

A traditional geostatistical approach for assessing spatial relationships is to model the data as multivariate normal, with variance matrix based on a parametric variogram model (exponential, spherical, Matérn etc. etc.). Such models may be difficult to work with. For example, one might need to invert large matrices in order to estimate variogram parameters and use the model for prediction (kriging).

In case when the data are collected over a regular (e.g. rectangular) grid, another approach can be used that is computationally more effective. Instead of postulating an explicit variogram model, one can instead consider a hidden smooth process $X(s)$, and in case when the observations $Y(\cdot)$ are made at the same locations (e.g. radar data), we assume that

$$Y(s) = X(s) + \varepsilon(s), \tag{1}$$

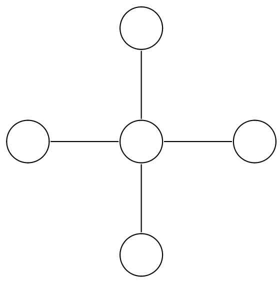
and $\varepsilon(s)$ is a white noise ("pure nugget") process. The degree of smoothness of $X(s)$ is determined by its prior distribution

$$p(x) \propto \exp(-\lambda_x \, x^T W x) \tag{2}$$

with the "locally linear" precision matrix W defined by

$$W_{ij} = \begin{cases} -1, & i \sim j \\ \text{number of neighbors of } i, & i = j \\ 0, & \text{otherwise} \end{cases}$$

and λ_x is the smoothness parameter.
Here, $i \sim j$ denotes neighboring locations on the grid. One possible choice of neighboring relation on a rectangular grid is shown below (see [5]).



When the data do not come from a regular grid (e.g. rain gauge data), it is still possible to estimate the hidden process X at the grid locations $s_j, j = 1, ..., m$ by using the convolution

$$Y(s) = \sum_j X(s_j)\phi(s - s_j) + \varepsilon(s) \tag{3}$$

Locations s are now not assumed to be on the grid. Above, ϕ is some prespecified kernel function, we use the Gaussian kernel $\phi(x) = exp(-x^2/2\sigma^2)$, and the parameter σ can be taken equal to the grid spacing. Higdon (see [5]) discusses the issues with choice of σ.

Computationally, the equation (3) simplifies to

$$\mathbf{y} = K\mathbf{x} + \boldsymbol{\varepsilon}$$

where $\mathbf{y}$, $\mathbf{x}$ are vectors containing all the values of Y, X and K is a matrix depending on the function ϕ and locations $s \in \mathcal{S}$ and s_j, $j = 1, ..., m$.

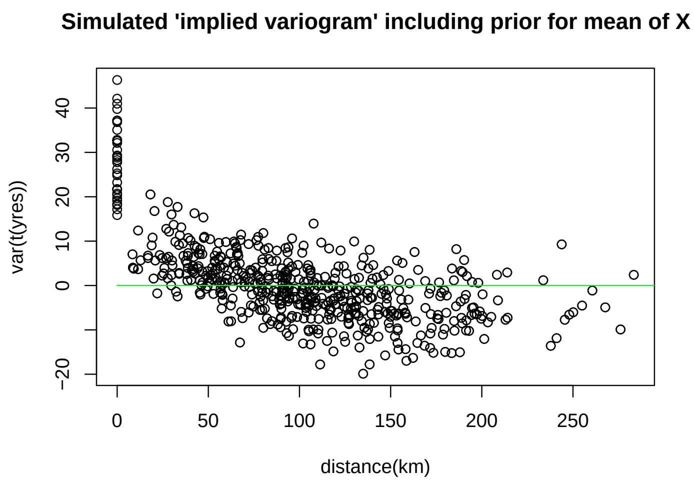
The above equation is then inverted, using Bayesian methods, to obtain posterior distribution $p(\mathbf{x} \,|\, \mathbf{y})$.

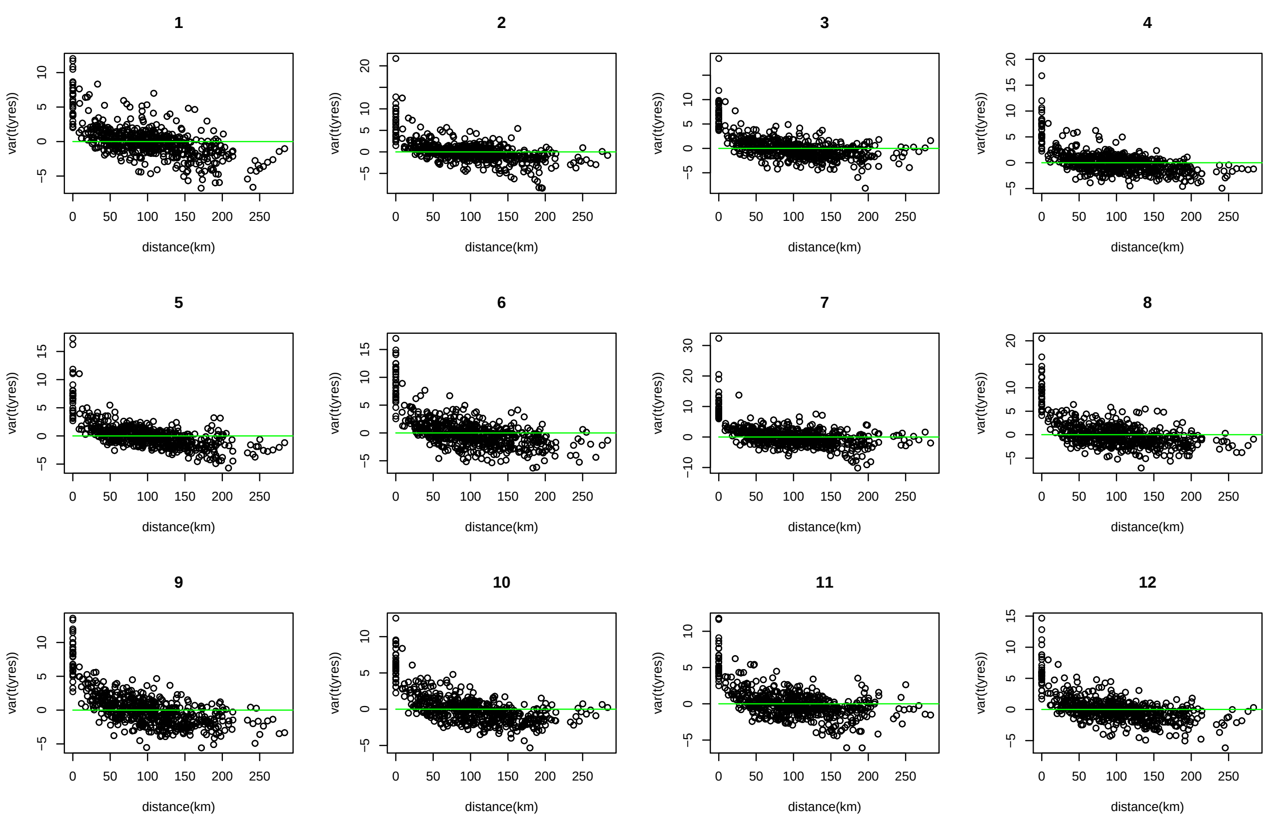
The approach of estimating the random field X on a grid is also beneficial when combining gauge data with some type of regular grid data (radar, satellite).

2.2 Experimental*

To show that the hidden field modeling is adequate for our needs, we compare a simulated covariogram implied by model (3), with the experimental ones (plotting the covariance of the monthly precipitation data at N stations *vs* distance between the stations, this is done separately for each of the Months 1-12 –

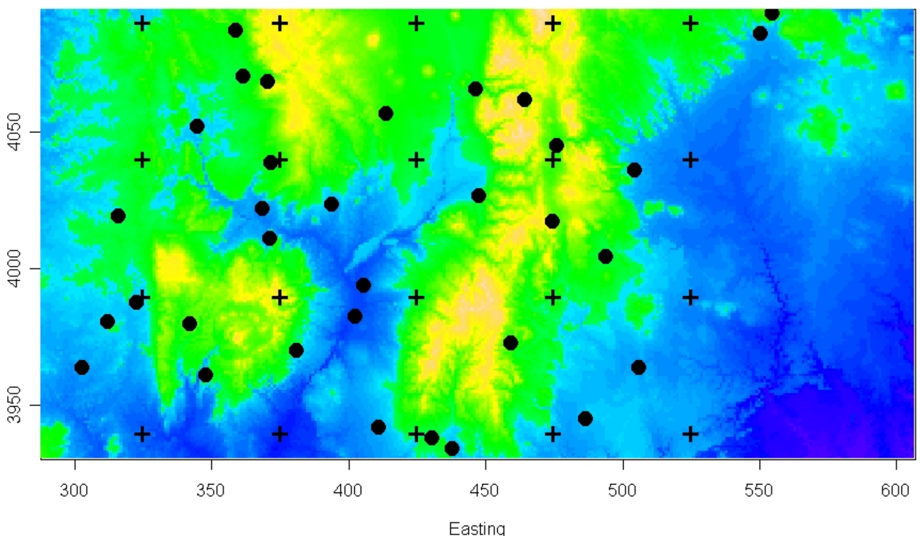
explain this in more detail).





2.3 Data and basic model

In the current work, we use the monthly precipitation totals for 33 stations in northern New Mexico, for the period of 1971-2005. The station locations are shown below, the grid locations are shown as "+".



For each month, we fit the regression equation

$$R_k = \beta_0 + \beta_1 Z_k + \beta_2 \cos(W - A_k) + Y_k, \tag{4}$$

where R_k is total monthly precipitation at the site k, Z_k is the elevation at the site, and the Y_k are spatially correlated residuals described in section 2.1. In fact, we will be using square-root transformed precipitation R_k to achieve near-normality of the results.

The value of "moisture flux direction" (MFD) W can be taken constant, then the approach coincides with ASOADeK, or can be a random field (MFDRF). For the latter, see Section 2.4.
A_k is the terrain *aspect*, computed using a 9x9-km window.

Note that, unlike ASOADeK, we do not use any spatial trend terms in this equation. Any existing trend will be reflected in the fitted values of the gridded field X and can be recovered from these, if necessary.

In order to increase the amount of data available, we assume that the MFDRF

W is the same for each calendar month of different years. Thus, we have 34 replicates, according to the number of years. (In the original ASOADeK paper, the data used were averages for each calendar month, so there was only one replicate.)

2.4 MFD random field

The gridding approach of Section 2.1 can also be carried out in case of MFD random field. We assume that W varies smoothly with the location, with values between 0 and 2π. What makes this problem non-trivial is that the values of W possess circular topology, $0 = 2\pi$, thus the traditional methods based on normal distribution are not applicable.

We will try to replicate the above approach that involved the grid process X and data Y. To this end, we will define $\tilde{W}$ = the values of MFD at the grid points, and W = the values of MFD co-located with the data Y. In case when data are observed on a regular grid to begin with, we do not need to consider W separately.

The analog of prior (2) is the prior for $\tilde{W}$,

$$p(\tilde{W}) \propto \exp(\gamma \sum_{i \sim j} \cos(\tilde{W}_i - \tilde{W}_j))$$

The values of $\tilde{W}$ that are close (in circular topology) will receive a high prior weight, and diametrically opposite values of $\tilde{W}$ receive a low prior weight. Thus, $\gamma \geq 0$ plays the role of a smoothness parameter (for example, $\gamma = 0$ implies prior values of W's completely independent of each other – no smoothing required, and very large values of γ demand the values of W to be almost constant throughout the whole region).

After this, it's easy to define W through an analog of convolution formula (cf. (3)):

$$W(s) = \sum_j \phi(s - s_j)\tilde{W}_j$$

however the sum taken above is understood in the circular topology, that is, through weighted averages

$$\sin(W(s)) = \frac{\sum_j \phi(s - s_j)\sin(\tilde{W}_j)}{\sum_j \phi(s - s_j)},$$

$$\cos(W(s)) = \frac{\sum_j \phi(s - s_j)\cos(\tilde{W}_j)}{\sum_j \phi(s - s_j)}.$$

These relations define a smooth random field W over the entire region. (We do not introduce a "nugget" value the way we did for random field Y.)

2.5 MCMC calculation

The model is fitted using Markov chain Monte-Carlo (MCMC) simulation involving Gibbs sampler. The latter is based on *full-conditional posteriors* (FCP). FCP's for all model parameters and unknown quantities are derived below.

The results of a MCMC simulation are the posterior distributions of the parameters of interest. From these, we can obtain point estimates (means, or, in case of highly skewed distributions, medians), confidence intervals etc.

2.5.1 Full conditional posteriors for model parameters*

To be added later.

3 Results

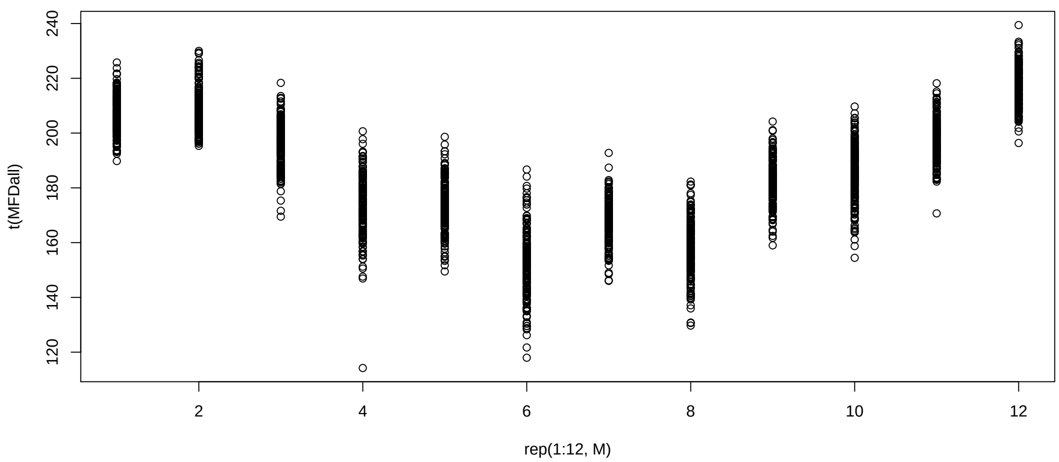
The results are obtained for each calendar month (1, ..., 12) separately (in line with the original ASOADeK paper).

3.1 Constant MFD results

First, we fit the equation

$$R_k^t = \beta_0 + \beta_1 Z_k + \beta_2 \cos(W - A_k) + Y_k^t,$$

for each Station k and each year t, but neither the coefficients β nor MFD W are dependent on t.
As a result, we get a sample from the posterior distribution of W plotted below for each calendar month.



They show the MFD drifting in a fairly regular pattern, between south-westerly in the winter and southerly in the summer.

3.2 MFD random field results

Next, we fit the equation

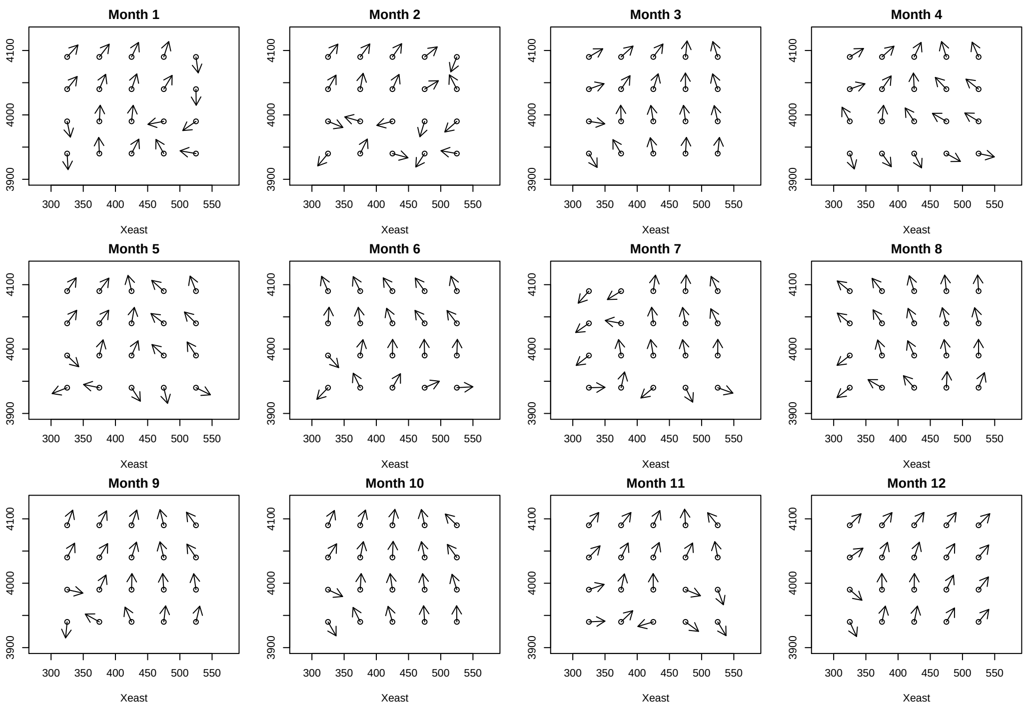
$$R_k^t = \beta_0^t + \beta_1^t Z_k + \beta_2^t \cos(W_k - A_k) + Y_k^t,$$

Regression coefficients β are now dependent on the year t, however, MFD random field W is assumed to be constant (for the same calendar month of different years).

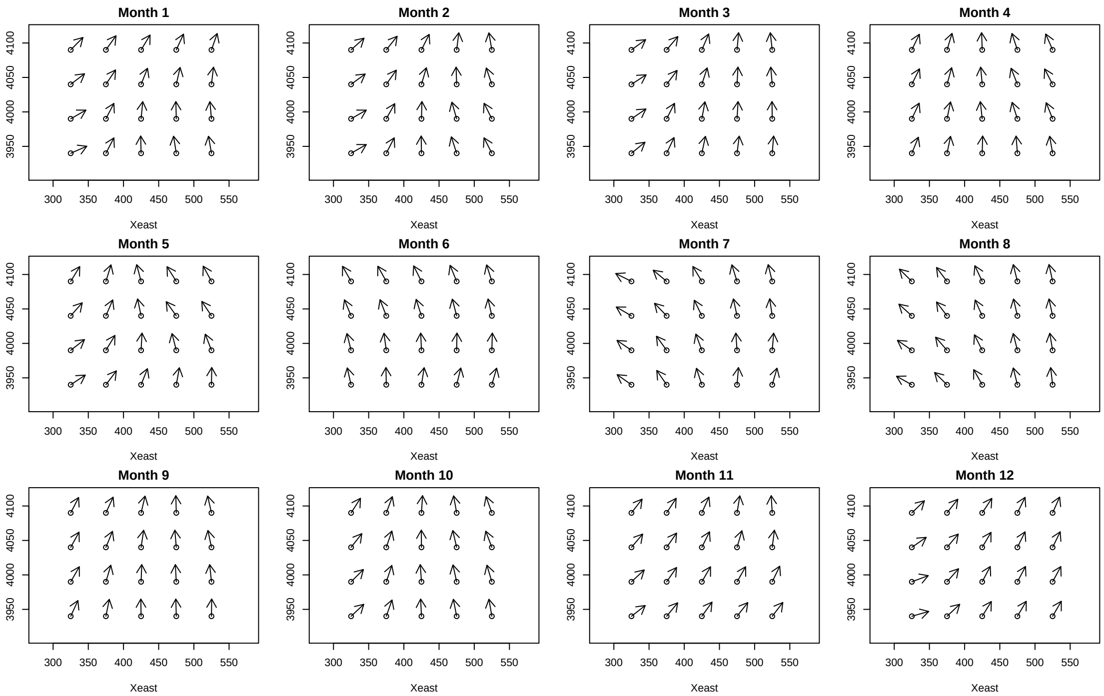
The results depend on the smoothing parameter γ. [Estimating γ is a separate issue.]

MFD RF estimates for $\gamma = 2$ (at the grid points, $\tilde{W}_j$ in the notation of Section

2.4):


Month 1
Month 2
Month 3
Month 4
Month 5
Month 6
Month 7
Month 8
Month 9
Month 10
Month 11
Month 12
Xeast
3900
4000
4100
300
350
400
450
500
550

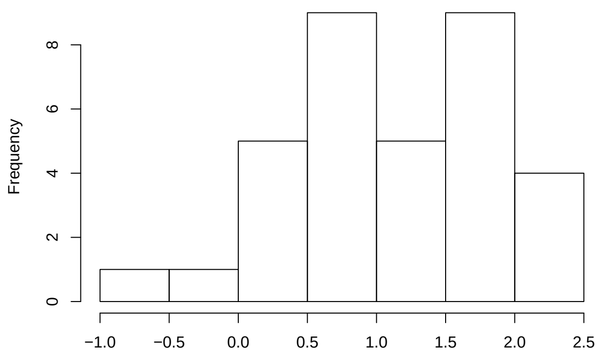
MFD RF estimates for $\gamma = 5$:



As seen from above, the value of $\gamma = 5$ results in a fairly smooth random field patterns, that follow closely those of the constant MFD, and also are very similar for all the Winter months, and for the monsoon months (July and August). They also suggest different behaviors of MFD in the western and eastern parts of the study region.

The MFD RF estimates using $\gamma = 2$ are more chaotic and are harder to interpret.

The estimates of β_2 are constrained to be positive, otherwise MFD W changes to diametrically opposite direction. Below, we show a histogram of the β_2 estimates for different years, in the month of December.



Other parameters: later

4 Conclusions

In conclusion, MFD RF modeling is a tool to assess MFD that changes over space, and may uncover additional climatic/orographic information.

References

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

[2] Brown, P.E., P. Diggle, M. Lord and P. Young (2001) *Space-Time Calibration of Radar Rainfall Data,* Applied Statistics, vol. 50, no. 2, 221-241

[3] 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, 2005.

[4] Guan, H., H. Xie and J. Wilson (2006) *A physically-based multivariate-regression approach for downscaling NEXRAD precipitation in mountainous terrain*, Manuscript.

[5] Higdon, D. (2007) *A primer on space-time modeling from a Baeysian perspective*, in *Statistical Methods for Spatio-Temporal Systems*, Chapman & Hall.