Subrata Pal, PhD 2024 (expected).

Winner of ASA Student Paper Competition award “Fast matrix-free methods for model-based personalized synthetic MR imaging” (Wednesday 10:55AM, CC 713A)
Motivating dataset: Groundwater mineral concentration in Bangladesh

- Concentration of 19 total minerals.
- Dropped 4 minerals because of very low variation.
- Main goal: understand the association among the mineral concentrations after accounting for spatial autocorrelation.
Factor models are often useful for characterizing the dependence among several variables using a few latent factors (Lawley, 1940).

BIC selects 4 factors assuming the factor scores are iid standard normals.

Factor scores are found to be spatially autocorrelated (variograms shown).

The variograms suggest long range dependence/non-stationary.

Fig: Empirical variograms of the first four factor scores.
Fractional Gaussian fields factors

Dutta and Mondal (2016)

Arise as the solution to fractional Laplacian differencing equation on sub-lattice of spacing $1/m$:

$$\Delta_m^{\nu/2} \psi^{(m)}(u, v) = \xi^{(m)}_{u,v}, \quad \nu \geq 0,$$

where

$$\Delta_m w(u, v) = w(u, v) - [\alpha\{w(u + \frac{1}{m}, v) + w(u - \frac{1}{m}, v)\} + (\frac{1}{2} - \alpha)\{w(u, v + \frac{1}{m}) + w(u, v - \frac{1}{m})\}],$$

where $0 \leq \alpha_m \leq 1/2$ is an anisotropy parameter, and $\xi^{(m)}_{u,v}$ is a Gaussian white noise on the sub-lattice $\mathbb{Z}_m^2$ with

$$\text{Var}\xi^{(m)}_{u,v} = \sigma^2_m/m^2.$$

Scaling limits these models contain a large class of intrinsic random fields:

- $\nu = 1$ : de Wijs process
- $1 < \nu < 2$ : Power variogram models.
- $\nu = 2$ : thin plate spline on 2D.
- $\nu = 1.5$ : Lévy-Brownian motion in 2D.
Assume observations are embedded in an \( r \times c \) array:

\[
Y = 1\mu^T + FZ\Lambda^T + E
\]

- The \( i \)th column of \( Z \), is a fractional Gaussian field on the \( r \times c \) array with parameters \( \nu_i \) (dependence) and \( \alpha_i \) (anisotropy).

- \( F : n \times rc \) incidence matrix; \( f_{ij} = 1 \) iff \( i \)th observation belongs to \( j \)th grid cell \((1 \leq j \leq rc; 1 \leq i \leq n)\): sparse binary matrix.

- Rows of \( E \) are iid \( N(0, \Psi) \), \( \Psi \) arbitrary positive definite matrix.

- The idea dates backs to at least Matheron (1982).

- Recent works include Zhang and Banerjee (2022), Taylor-Rodriguez et al. (2019): Scalable, Bayesian, based on NNGP.

- Our focus: flexible class of models accommodating long range dependence.
Exploratory analysis: Fit theoretical cross-variograms to empirical cross-variograms.

Matrix-free Stochastic EM algorithm for maximum likelihood estimation.

Information matrix estimated through matrix-free conditional simulation.

Computational requirements:
- Storage required $O(n)$,
- Computational complexity: $O(n \log n^2)$ through the use of discrete cosine transformations.
Parameters: $\Lambda$, $\Psi$, $\{\nu_j\}$, and $\{\alpha_j\}$ ($1 \leq j \leq q$).

- Each factor score $Z_i$ follows a (generalized) multivariate normal distribution:

$$p(z_j) \propto |W_j|_{+}^{\nu_j/2} \exp \left( -\frac{1}{2} z_j^\top W_j^{\nu_j} z_j \right)$$

Where $W_j^{\nu_j} = M' D(\alpha_j)^{\nu_j} M$, $D(\alpha)$ is analytically known diagonal matrix, and $M$ is the matrix of two-dimensional DCT.
Parameters: $\Lambda$, $\Psi$, $\{\nu_j\}$, and $\{\alpha_j\}$ ($1 \leq j \leq q$).

- Each factor score $Z_i$ follow a (generalized) multivariate normal distribution:

$$p(z_j) \propto |W_{\nu_j}|^{\nu_j/2} \exp \left(-\frac{1}{2}z_j^T W_{\nu_j}^j z_j\right)$$

Where $W_{\nu_j}^j = M^\prime D(\alpha_j)^{\nu_j} M$, $D(\alpha)$ is analytically known diagonal matrix, and $M$ is the matrix of two-dimensional DCT.

- Given $Y$, the vec$(Z)$ is Gaussian with mean $\Omega^{-1}\mu$ and precision matrix $\Omega$, where

$$\Omega = (\Lambda^\prime \Psi^{-1} \Lambda) \otimes (F^\prime F) + W$$

$$\mu = \text{vec}(F^\prime (Y - X\beta)\Psi^{-1} \Lambda)$$

where $W$ is the block diagonal matrix consisting of $W_{\nu_j}^j$. 
EM-algorithm

Parameters: $\Lambda$, $\Psi$, $\{\nu_j\}$, and $\{\alpha_j\} \ (1 \leq j \leq q)$.

- Each factor score $Z_i$ follow a (generalized) multivariate normal distribution:
  \[
p(Z_j) \propto |W_j|^\nu_j/2 \exp \left( -\frac{1}{2} z_j^T W_j^{\nu_j} z_j \right)\]
  Where $W_j^{\nu_j} = M'D(\alpha_j)^{\nu_j} M$, $D(\alpha)$ is analytically known diagonal matrix, and $M$ is the matrix of two-dimensional DCT.

- Given $Y$, the $\text{vec}(Z)$ is Gaussian with mean $\Omega^{-1}\mu$ and precision matrix $\Omega$, where
  \[
  \Omega = (\Lambda'\Psi^{-1}\Lambda) \otimes (F'F) + W
  \]
  \[
  \mu = \text{vec}(F'(Y - X\beta)\Psi^{-1}\Lambda)
  \]
  where $W$ is the block diagonal matrix consisting of $W_j^{\nu_j}$.

- M-step is conceptually tractable. After iteration $t$:
  \[
  \Lambda_t^{'} = \left( E_{t}[Z'F'FZ] \right)^{-1} E_{t}[Z'F'Y]
  \]
  \[
  \Psi_{t+1} = \frac{1}{n} E_{t} [(Y - FZ\Lambda_{t+1})'(Y - FZ\Lambda_t')]
  \]
  \[
  (\nu_j, \alpha_j)_{t+1} = \arg \max_{\nu_j, \alpha_j} \left[ \nu_j \log |D(\alpha_j)|_+ - E_{t}[Z_j'M'D(\alpha_j)^{\nu_j}MZ_j] \right]
  \]
  \[
  E_t \text{ denotes the conditional expectation with respect to } Z \text{ given } Y \text{ and parameter values at } t\text{th iteration: computationally untractable.}
  \]
Matrix-free stochastic E-Step: Simulate $E_t$

Need to sample from $N(\Omega^{-1}\mu, \Omega^{-1})$.

- **Challenge:** $\Omega$ is $rcq \times rcq$ large dense matrix ($r = 500$, $c = 300$, $q = 4$, requires 2.7TB RAM). Cannot use traditional methods for conditional simulation.

- $\Omega$ admits a matrix-free rectangular root $\Omega = B_1B_1' + B_2B_2'$

$$B_1 = \Lambda' \Psi^{-1/2} \otimes F'$$

$$B_2 = (I_q \otimes M)D^{1/2}$$

where $D^{1/2} = \left\{ D_{1}^{\nu_1/2}, D_{2}^{\nu_2/2}, \ldots, D_{q}^{\nu_q/2} \right\}$.

- $B_1$, $B_2$, and $\Omega$ are **matrix-free operators**: Can use DCT to compute matrix-vector multiplication without storing the matrix.

- Simulate $x \sim N(0, \Omega)$: $x = B_1v_1 + B_2v_2$ where $v_1$ and $v_2$ are multivariate standard normal vectors.

- Solve: $\Omega z = \mu + x$ using matrix-free Lanczos algorithm.

- Then $z \sim N(\Omega^{-1}\mu, \Omega^{-1})$. 
Simulation studies

- 128 × 128 grid, \( q = 2 \) factors with \( \nu = 1.95 \) and 1.2, anisotropy parameters \( \alpha = 0.2 \) and 0.25:

- Missing rate: 0, 10%, 30%, 50%, 70%, and 90%.

- Estimation accuracies of anisotropy parameters and nugget covariance matrix are more affected by more missing data.
Simulation studies

Standard errors from information matrix versus empirical standard deviations (200 reps):

- No practical loss in efficiency in estimating $\nu$ and nugget covariance $\Psi$.
- Some loss of efficiency in estimating anisotropy and factor loading parameters at high missing rate.
We embed the region in a $500 \times 300$ array from 20 to 27° N and 88 to 93° E.

Analyze log concentration of 15 minerals.

Estimates of factor loadings are shown below (values < 0.1 are not shown).

<table>
<thead>
<tr>
<th>Chemicals</th>
<th>Z1</th>
<th>Z2</th>
<th>Z3</th>
<th>Z4</th>
</tr>
</thead>
<tbody>
<tr>
<td>As</td>
<td>0.11</td>
<td>0.14</td>
<td>0.16</td>
<td></td>
</tr>
<tr>
<td>Al</td>
<td>0.11</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B</td>
<td></td>
<td>-0.11</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ba</td>
<td>0.19</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ca</td>
<td>0.29</td>
<td></td>
<td>0.10</td>
<td></td>
</tr>
<tr>
<td>Fe</td>
<td></td>
<td>0.12</td>
<td>0.19</td>
<td></td>
</tr>
<tr>
<td>K</td>
<td>0.12</td>
<td>0.12</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mg</td>
<td>0.28</td>
<td>0.12</td>
<td>0.10</td>
<td></td>
</tr>
<tr>
<td>Mn</td>
<td>0.12</td>
<td></td>
<td>0.10</td>
<td></td>
</tr>
<tr>
<td>Na</td>
<td></td>
<td>-0.11</td>
<td></td>
<td></td>
</tr>
<tr>
<td>P</td>
<td></td>
<td>0.16</td>
<td>0.10</td>
<td></td>
</tr>
<tr>
<td>Si</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SO4</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sr</td>
<td>0.28</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Zn</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Conclusions

- Fast matrix-free maximum likelihood methods for spatial factor models on lattice.
- Model accommodates spatial misalignment.
- Estimates are practically efficient.
- Future works include extensions to functional spatial data and spatio-temporal data.

Choice of the number of factors?

Thank you!