Multiresolution Gaussian Process Models for Spatial Data

Paper review

March 14, 2024

Table of contents

Intro

Motivation

- Current geoscience data is characterized by high spatial resolution and scale dependence.
- Applying standard techniques can be non-feasible due to computational costs.
- **Goal:** Introduce a new spatial model that does well with large datasets and can be implemented in a relatively low cost manner

The model

- Describe the observations $\vec{y}$ according to the model

$$y_i = Z_i^T \vec{d} + g(x_i) + \epsilon_i \tag{1}$$

  where Z are the covariates, g is a Gaussian process, and ϵ_i are mean zero errors.
- Determine g using a multiresolution Markov Random Field model

Markov Random fields

Markov Random Fields

- Write
$$g(\vec{x}) = \sum_{l=1}^{L} g_l(\vec{x}) \tag{2}$$
with g_l independent spatial processes with increasing resolution.
- Each g_l has an expansion in terms of a multiresolution basis (i.e. radial basis functions, wavelets, needlets, etc):
$$g_l(\vec{x}) = \sum_{j=1}^{m(l)} c_{j,l}\phi_{j,l}(\vec{x}) \tag{3}$$
- Apply the Markov Random Field model: Place lattice points at the peaks of the $\phi_{j,l}$ and model the coefficients for each level as a random process on the lattice.

Lattice points

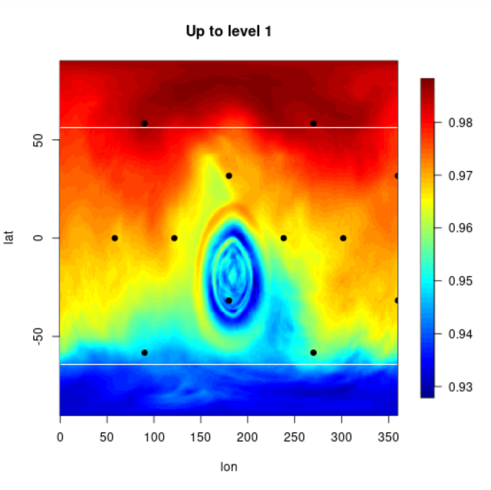


Figure 1: Lattice points, First resolution level

Lattice Points

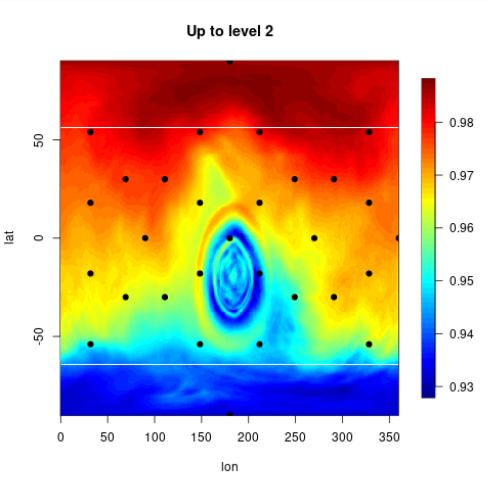


Figure 2: Lattice points, second resolution level

Lattice Points

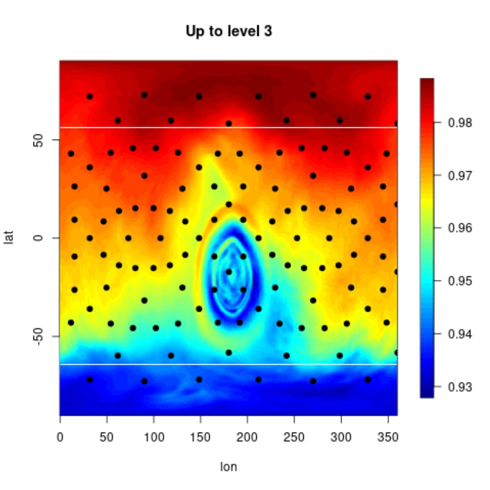


Figure 3: Lattice points, third resolution level

Markov Random Fields

- Key insight: We can enforce a simple spatial structure to compute the precision matrix Q rather than worry about the covariance matrix if we use Gauss-Markov Random fields.
- Given autoregression matrix B_l and $\vec{e} \sim N(0, \rho I)$ we have that the coefficients $\vec{c_l} = B_l^{-1}\vec{e} \implies \text{cov}(\vec{c_l}) = \rho B_l^{-1} B_l^{-T}$
  - Do this because B is easy to generate.
- So: We can just calculate the precision matrix of $\vec{c_l}$ as $Q_l = \frac{1}{\rho} B_l^T B_l$ and can use these to construct a block-diagonal process precision matrix Q

Markov Random Fields

- Generate B using a spatial autoregression model: How?
- Idea: Each coefficient is a combination of its neighbors plus some noise: B should respect this structure: If the lattice point i has neighbors N_i, we should have

$$\begin{aligned} B_{i,i} &= 1 + \kappa^2 \\ \sum_{j \in N_i} B_{i,j} &= -1 \end{aligned} \tag{4}$$

- This enforces sparsity in B, and therefore in Q.
- This results in normally distributed coefficients with a Matern covariance with scale parameter κ and smoothness $\nu = 1$.

Modeling and Estimation

Spatial modeling and estimation

- Let Φ be a matrix with columns indexing basis functions, rows locations. Then this all implies that the field $\vec{y} = Z\vec{d} + \Phi\vec{c} + \epsilon$ and

$$\vec{y} \sim N\left(Z\vec{d}, \rho\Phi Q^{-1}\Phi^{T} + \sigma^2 W^{-1}\right) \tag{5}$$

  where $\epsilon \sim N(0, \sigma^2 W^{-1})$
- Reparameterize the model: let $M_\lambda = \Phi Q^{-1}\Phi^{T} + \lambda W^{-1}$, then

$$\vec{y} \sim N(Z\vec{d}, \rho M_\lambda) \tag{6}$$

  where λ is inverse SNR.
- Now we can use MLE!

Spatial modeling and estimation

- The full likelihood function is given by

$$\ell\left(\vec{y}|\rho,\lambda,Q^{-1},\vec{d}\right) \sim -\frac{1}{2\rho}\left\langle \vec{y}-Z\vec{d}, M_\lambda^{-1}\left(\vec{y}-Z\vec{d}\right)\right\rangle - \frac{1}{2}\log\det\rho M_\lambda \tag{7}$$

- For fixed λ and Q^{-1} the MLE of $\vec{d}$ is given by GLS:

$$\hat{d} = \left(Z^T M_\lambda^{-1} Z\right)^{-1} Z^T M_\lambda^{-1}\vec{y} \tag{8}$$

and $\hat{\rho}$ can be determined analytically.

- Now we have what we need to determine the coefficients $\hat{c}$.

Spatial modeling and estimation

- Given all the estimates on the previous slide, we can determine the conditional distribution of $\vec{c}$:

$$\begin{aligned} \vec{c}|\vec{y},\vec{d},\lambda,\rho,Q^{-1} \sim N\left(\hat{c},\rho Q^{-1}\left(I-\Phi^T M_\lambda^{-1}\Phi Q^{-1}\right)\right) \\ \hat{c} = Q^{-1}\Phi^T M_\lambda^{-1}\left(\vec{y}-Z\vec{d}\right) \end{aligned} \tag{9}$$

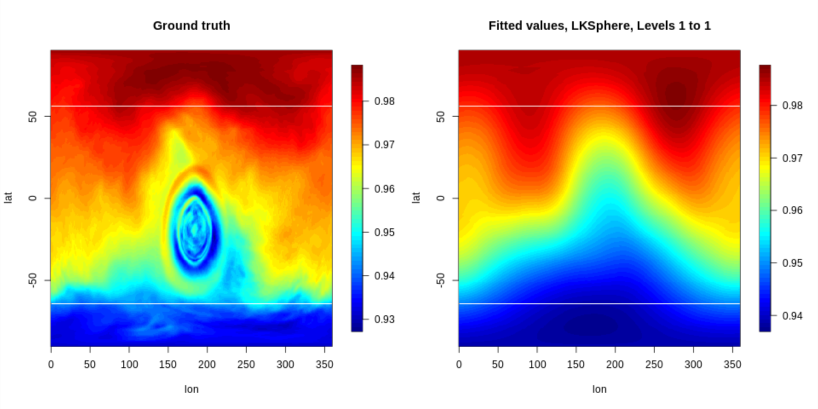
- Take $\hat{c}$ to be the coefficients, then at a point $\vec{x}$ the estimate of the field is

$$\hat{y} = z(\vec{x})^T\hat{d} + \sum_{j=1}^{m}\hat{c}_j\phi_j(\vec{x}) \tag{10}$$

Application to data

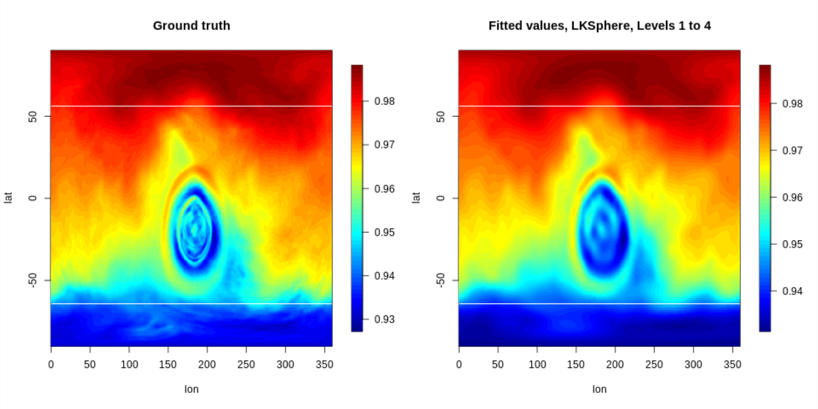
Application to WACCMX Tonga eruption data

- The algorithm was applied to WACCM-X oxygen mixing ratio data from the Tonga eruption.



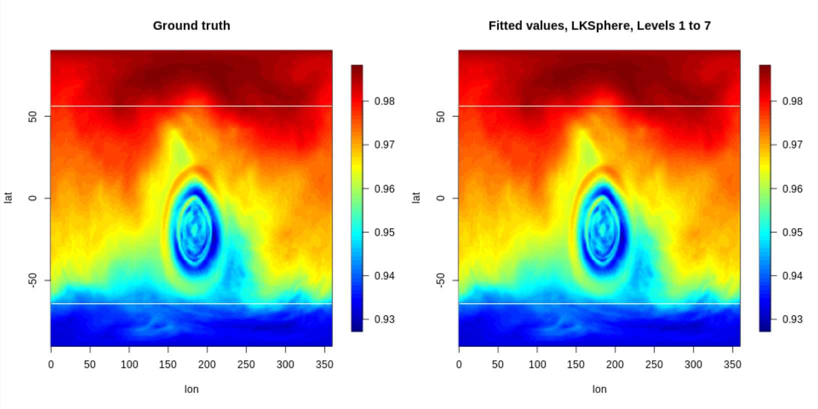
Application to WACCMX Tonga eruption data

- The algorithm was applied to WACCM-X oxygen mixing ratio data from the Tonga eruption.



Application to WACCMX Tonga eruption data

- The algorithm was applied to WACCM-X oxygen mixing ratio data from the Tonga eruption.



Computational considerations

Computational condiderations

- Number of basis functions increases very quickly: Last plot above was generated with > 40000 basis functions.
- The method relies heavily on matrices being sparse: Structure already guarantees that Q is sparse, but we have to worry about M_λ^{-1} as well, which may not be.
- Apply the Shannon-Woodbury formula to M_λ^{-1}:

$$M_\lambda^{-1} = \left(\Phi Q^{-1}\Phi^T + \lambda W^{-1}\right)^{-1} = \frac{1}{\lambda}\left(W - W\Phi G^{-1}\Phi^T W\right) \qquad (11)$$
$$G = \Phi^T W \Phi + \lambda Q$$

Computational Considerations: Matrix solves

- Since W and Q are sparse by construction, G will be sparse if Φ is. G will also be positive definite, and therefore invertible.
    - Basis functions need to be zero on a large portion of the domain
- If the above is true, a sparse Cholesky decomposition can handle the system

$$\begin{gathered} G\vec{v} = \Phi^T W\vec{u} \implies \\ M_\lambda^{-1}\vec{u} = \frac{1}{\lambda}\left(W\vec{u} - W\Phi\vec{v}\right) \end{gathered} \tag{12}$$

Computational considerations: Determinants

- Recall: Evaluating the log-likelihood means finding $\det M_\lambda$: Can be very hard!
- Utilize Sylvester's Theorem: If $U \in R^{n \times m}$

$$\det\left(UU^T + I_n\right) = det\left(U^T U + I_m\right) \tag{13}$$

- Can use this to show that

$$\det(M_\lambda) = \lambda^{n-m} \frac{\det(G)}{\det(W)\det(Q)} \tag{14}$$

which can all be computed from the sparse Cholesky decompositions!

Conclusion

Conslusion

- We can use a Gaussian Markov Random Field model applied to multiresolution basis coefficients to represent geospatial data.
- These models have tractable likelihood functions that make estimating and predicting relatively straightforward
- By using compactly supported basis functions we can induce sparsity in the relevant matrices and do these computations quickly, which makes these models scalable.

References

References

- Nychka, Douglas, et al. "A Multiresolution Gaussian Process Model for the Analysis of Large Spatial Datasets." Journal of Computational and Graphical Statistics, vol. 24, no. 2, 2015, pp. 579–99. JSTOR, http://www.jstor.org/stable/24737282.
- Liu, H.-L., Wang, W., Huba, J. D., Lauritzen, P. H., Vitt, F. (2023). Atmospheric and ionospheric responses to Hunga-Tonga volcano eruption simulated by WACCM-X. Geophysical Research Letters, 50, e2023GL103682. https://doi.org/10.1029/2023GL103682