Gaussian Markov Random Fields for Big-scale Spatio-Temporal Data

Ron Sarafian

Ben-Gurion University of the Negev

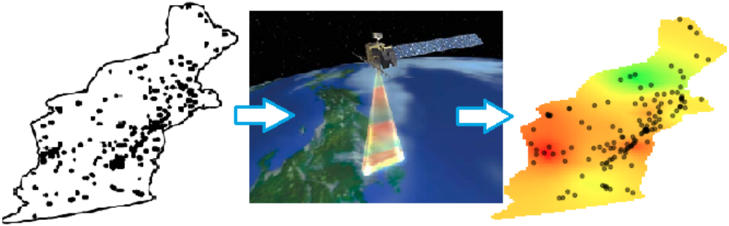
ronsar@post.bgu.ac.il

Motivation

Motivating example: Predicting Air pollution

- **Requires:** Air pollution levels over a dense spatial domain
- **Exists:** Measurements from spatially limited monitoring stations
- **Also Exists:** Unlimited Geographic and Atmospheric data

Goal: Use spatio-temporal data to predict air pollution in space and time



Applications: Predictions are used as covariates in future researches (e.g., epidemiology).

Challenges: Dependencies and Scale

Spatial and temporal structure of the predictors raises some complex issues:

- **Statistical modeling:** How to account for space-time dependencies?
  - Spatio-temporally correlated prediction errors
  - $\Rightarrow$ Epidemiological error in variables
  - $\Rightarrow$ Biased epidemiological results
- **Learning large and complex dependencies:** When learning the correlation structure between large amount of spatial units, The computational problem scales very fast

Our talk

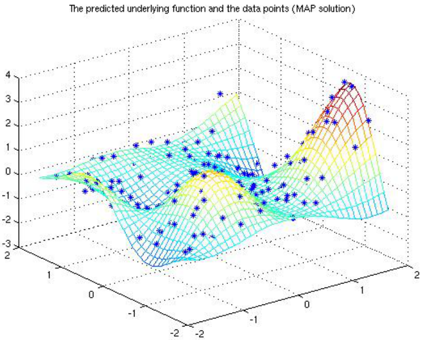
1. **Gaussian Random Field (GRF):** Reach but limited
   - Definition
   - Computational limitations
2. **Gaussian Markov Random Field (GMRF):**
   - Conditional Independence via Markov property
   - How to fit (almost) a GRF on a large-scale data
   - Comparing GMRF and LMM: discrete vs. continuous spatial random effects

Gaussian Random Field

The classic model: Gaussian Random Field (GRF)

GRF is one of the fundamental and most common approach for analysis of **spatio-temporal dependency** over **continuous domains**

A GRF is completely determined through its **mean** and **covariance function**



Source: https://research.cs.aalto.fi/pml/software/gpstuff/demo_regression1_r.shtml

GRF definition

A **spatio-temporal process** $Y(s,t), \quad s \in S; t \in T,$ **is a GRF if** $Y = (y(s_1,t_1), ..., y((s_{n_s}, t_{n_t}))$ has a $N = n_s \cdot n_t$ multivariate Gaussian density function:

$$f_Y(y) = \left(\frac{1}{\sqrt{2\pi}}\right)^N |\Sigma|^{-\frac{1}{2}} \exp\{-\frac{1}{2}(y-\mu)\Sigma^{-1}(y-\mu)'\}$$

with **mean** $\mu = \mu(s,t) \in \mathbb{R}^N$
and **variance**: $\Sigma = C\{(s,t)_i, (s,t)_j)\}_{ij} \in \mathbb{R}^N \times \mathbb{R}^N$.

GRF's Covariance

The covariance C is usually a stationary, spatially isotropic function

Stationarity means that

C can be written as: $C((s,t),(s',t')) = C(s-s', t-t')$

(spatial) Isotropy means that

C is only depends on symmetric distance (e.g. Euclidean: $||s-s'||$)

Can we choose any covariance function?

No, the induced covariance matrix has to be nonsingular for real solution

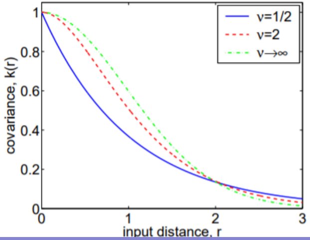
The Matérn Covariance

A typical spatial covariance functional form is the Matérn:

Matérn

$$C(s,s') = \sigma^2 \frac{2^{1-\nu}}{\Gamma(\nu)} \left(\sqrt{2\nu} \frac{\|s-s'\|}{\rho} \right)^{\nu} K_{\nu} \left(\sqrt{2\nu} \frac{\|s-s'\|}{\rho} \right)$$

The Matérn allows a general structure of stationary covariances with parameters: ν, p associated with the smoothness and range of the process



Is GRF perfect?

Advantages:

- Specify **continuous dependence patterns** (both space and time)
- **Good analytic properties**
- Usually **very accurate** spatio-temporal predictions

a fly in the ointment...

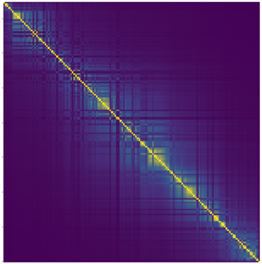
Disadvantages:

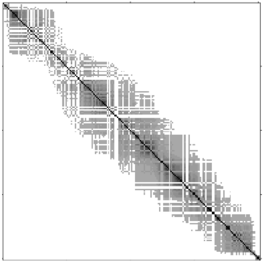
- Dense $N \times N$ structure of the covariance matrix:
  Fitting a GRF requires $\mathcal{O}(N^3)$ operations (due to $|\Sigma|$ and Σ^{-1})

Unfortunately, GRF computation is **infeasible for large space-time datasets**

Can we do something with GRF's *"Big N"*?

Yes, the answer is **Sparsity!**





Sparsity enables using particular numerical algorithms that allows doing mathematical operations with **less memory and computing time**

Sparse covariance matrix is the base behind most of the strategies developed to overcome GRF's computational bottleneck
(more precisely - precision's sparsity)

Examples for sparse covariance based approach for space-time prediction

- **Covariance Tapering:** location pairs associated with near-zero entries in Σ are considered independent
- **Mixed Models:** specifying correlation structures using multilevel effects (discrete clusters). Then learning cluster's unique distribution parameters. Results in a sparse *Block-diagonal* Σ.
  Advantage: Computationally efficient. Disadvantage: discrete dependency structure.
- **Markovian assumptions:** Inducing conditional Independence

Gaussian Markov Random Field

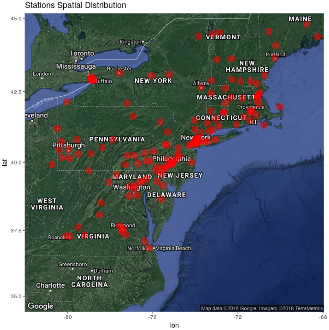
The Gaussian Markov Random Field (GMRF)

One approach for avoiding its computational hurdle is by **approximating** a GRF **with a GMRF**

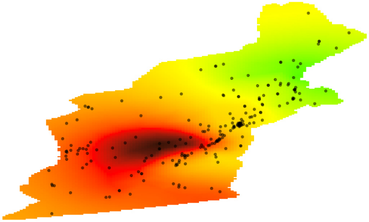
- GMRFs are **discretely indexed random fields** involving multivariate Gaussian distribution with **Markov property**
- Markov property induce **conditional independence** between random variables, so that a data point in a GMRF **depends only on its neighbors** (in space/time)

We can "fit" a reach GRF model with continously covariance structure, but actually solve it by GMRF approximation.

Example: Air-pollution prediction in USA



GMRF Model



Study domain: Area: 450,000 km^2, Period: years 2000-2015
Prediction resolution: 1 km^2 , daily

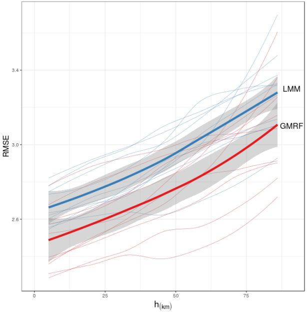
Comparing learners results

Northern USA Air-pollution prediction
$\sim$250K space-time samples
8 cores machine

Algorithm	RMSE	Training time (hours)	
RF	2.75	5.5	including tuning
LMM	2.68	0.16	
XGBoost	2.61	10	including tuning
GMRF	2.42	1.5	

Comparing GMRF to LMM prediction performance in different extrapolation levels (Sarafian et. al, 2019)

- **Increasing** h - predict in remoter areas, i.e., more extrapolation.
- With more extrapollation both performance decline (RMSE increase)
- In any level the **GMRF dominance** is significant



GMRF is more accurate at any distance in USA air-pollution prediction

Whether the goal is predictions in **remote areas**, or whether it is accuracy in areas where **stations are crowded**

So how does this magic work?

GMRF conditional independence

Let the **neighbors** $\mathcal{N}_i$ of a point x_i be the points $\{x_j | j \in \mathcal{N}_i\}$ that are "close" to x_i

Gaussian Markov Random Field

A GRF $x \sim N(\mu, \Sigma)$ that satisfies

$$f\left(x_i | \{x_j : j \neq i\}\right) = f\left(x_i | \{x_j : j \in \mathcal{N}_i\}\right)$$

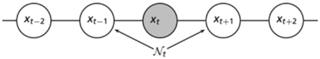
is a GMRF

Can you think of a famous GMRF?

GMRF Examples

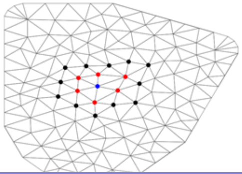
AR(1)

$$x_t = a x_{t-1} + \epsilon_t, \qquad t = 1, 2, ..., \qquad \epsilon_t \sim N(0, \sigma^2)$$



Neighbors on a mesh

All information for the **blue** point is:
in the **red** points (1st order) or **red** + **black** points (2nd order)



Conditional independence (CI)

Conditional independence does not necessarily mean sparse covariance matrix, rather a **sparse precision matrix**

Now, instead of $y(s,t) \sim N(\mu, \Sigma)$, let us write:

$$y(s,t) \sim N(\mu, Q^{-1})$$

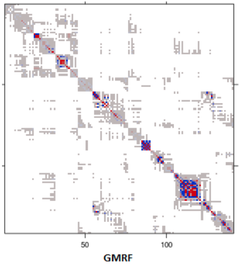
The precision matrix is sparse

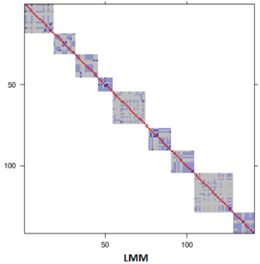
Elements in the precision matrix of a GMRF are non-zero only for neighbors and diagonal elements:

$$j \neq \{i, \mathcal{N}_i\} \iff Q_{ij} = 0$$

GMRF vs. LMM: Learned precision matrices

Part (one day) of the precision matrices. data is ordered by spatial regions (LMM's random effect level).





In GMRF, the correlation between spatial units is **not limited to a specific region**

Hence, precision matrix is not subject to discrete spatial definitions

Aproximating GRF with GMRF

- Lindgren et al. (2011) provide an explicit **link** between **GRF with Matérn** covariance and **GMRF** using stochastic partial differential equation (SPDE).
- It is based on the relationship that a **GRF with a Matérn covariance** is a solution to the linear fractional **SPDE** with Gaussian white noise inovation process.

- The SPDE approach allows **fitting a GRF** with a continuously and smoothly decaying covariance function, while enjoying the sparse precision matrix of a **GMRF representation**!
- It also has a great implementation in R: `R-INLA`

From continuous to discrete

Wait, but the space is continuous and GMRFs are discrete!

True, but a continuous field can be approximated with basis functions using Finite Element Method (FEM)

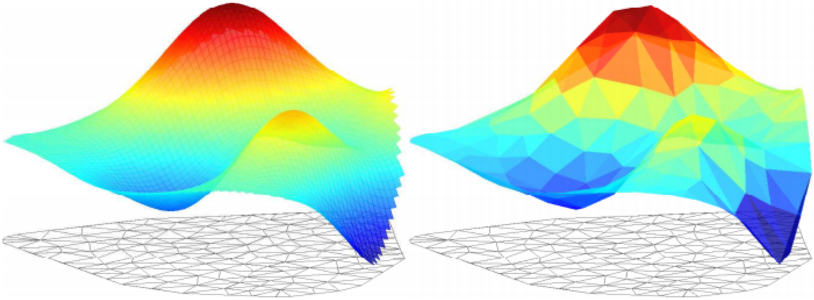
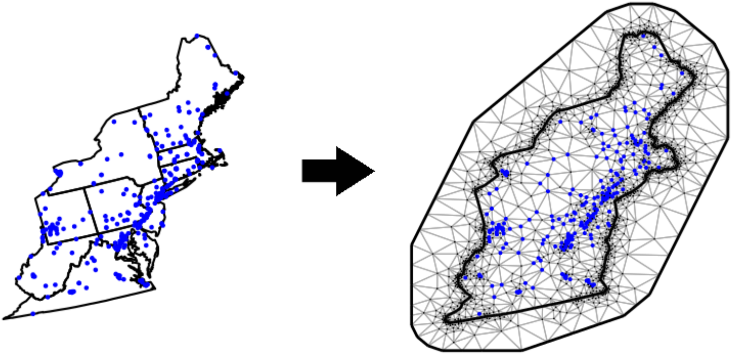


image source: Johan Lindstrom, GMRF presentation

Approximating a continuous field with a mesh



Is CI reasonable?

How reasonable is the Markovian CI assumption?

Remember, we wanted to obtain **sparse precision matrix** and not sparse covariance

Sparse covariance matrices imply marginal independence: strong and generally unreasonable assumption. However, conditional independence (via the Markov property) is **often a very reasonable** assumption

Comparing GMRF and LMM

The only difference in terms of our model formulation between the LMM and GMRF lies in the form of the **spatial random effects**

Although we were able to achieve better results also by considering continuous temporal patterns, this was not our research goal

In both models spatial effects include **intercept** and **slopes** of satellite data, within each **day** (so we catch unique effect that are changing over space-time)

Spatial random effects:

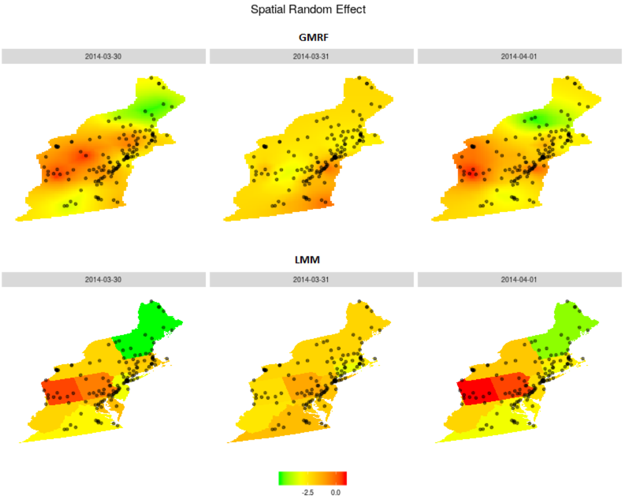
LMM

Region-wise discrete effect

GMRF

Matérn field

Spatial random effects: Discrete vs. Continuous



Color-scale indicate the estimated value of the spatial random effect (intercept)

References

Lindgren, Finn, Hvard Rue, and Johan Lindstrm. "An explicit link between Gaussian fields and Gaussian Markov random fields: the stochastic partial differential equation approach." Journal of the Royal Statistical Society: Series B (Statistical Methodology) 73.4 (2011): 423-498.

Lindgren, Finn, and Hvard Rue. "Bayesian spatial modelling with R-INLA." Journal of Statistical Software 63.19 (2015).

Sarafian, Ron, et al. "Gaussian Markov Random Fields versus Linear Mixed Models for satellite-based PM2.5 assessment: Evidence from the Northeastern USA." Atmospheric Environment (2019).

Thank you!