Dynamic NNGP for Large Spatio-temporal Data

Abhi Datta[1], Sudipto Banerjee[2] and Andrew O. Finley[3]

July 31, 2017

[1]Department of Biostatistics, Bloomberg School of Public Health, Johns Hopkins University, Baltimore, Maryland.

[2]Department of Biostatistics, Fielding School of Public Health, University of California, Los Angeles.

[3]Departments of Forestry and Geography, Michigan State University, East Lansing, Michigan.

Spatio-temporal Data

- Data from multiple timepoints at some or each of the locations
- Example: Climate data, air pollution data, house prices data
- Time points can be regular (hourly, daily, weekly) or irregular

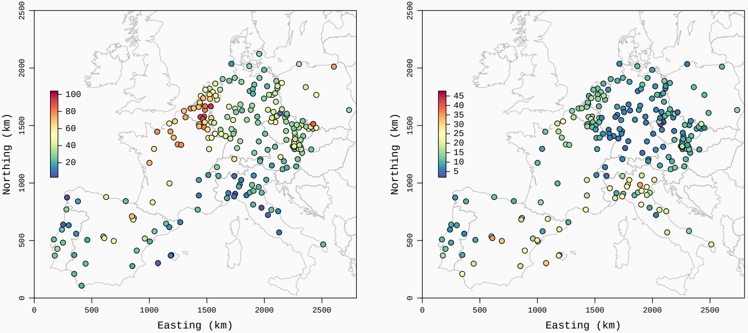


Figure: PM_{10} levels in Europe in March, 2009 (left) and June, 2009 (right)

Discrete Time Continuous Space Models

- $y(s,t) = \underbrace{x_t'(s)\beta_t}_{\text{mean}} \quad + \quad \underbrace{u_t(s)}_{\text{random effect}} \quad + \quad \underbrace{\epsilon_s(t)}_{\text{iid noise}}$
- $\beta_t = \beta_{t-1} + \eta_t$, $\eta_t \sim N(0, \Sigma_t)$
- $u_t(s) = u_{t-1}(s) + w_t(s)$
- $w_t = (w_t(s_1), \ldots, w_t(s_K))' \overset{ind}{\sim} N(0, C_S(\theta_t))$

Discrete Time Continuous Space Models

- $y(s,t) = \underbrace{x_t'(s)\beta_t}_{\text{mean}} \quad + \quad \underbrace{u_t(s)}_{\text{random effect}} \quad + \underbrace{\epsilon_s(t)}_{\text{iid noise}}$
- $\beta_t = \beta_{t-1} + \eta_t$, $\eta_t \sim N(0, \Sigma_t)$
- $u_t(s) = u_{t-1}(s) + w_t(s)$
- $w_t = (w_t(s_1), \ldots, w_t(s_K))' \overset{ind}{\sim} N(0, C_S(\theta_t))$
- Modeling large $C_S(\theta_t)$
  - Use the NNGP approximation $\tilde{C}_S(\theta_t)$
- Restricted to data observed at regular time intervals
- Interpolation at finer temporal resolution not possible

Continuous space-time models

- Spatio-temporal domain: $\mathscr{D} = S \times T$
- Every observation has a space and a time co-ordinate: $\ell = (s, t)$
- $y(s,t) = \underbrace{x(s,t)'\beta}_{\text{mean}} \quad + \quad \underbrace{w(s,t)}_{\text{random effect}} \quad + \ \underbrace{\epsilon(s,t)}_{\text{iid noise}}$

Continuous space-time models

- Spatio-temporal domain: $\mathscr{D} = S \times T$
- Every observation has a space and a time co-ordinate: $\ell = (s, t)$
- $y(s,t) = \underbrace{x(s,t)'\beta}_{\text{mean}} \quad + \quad \underbrace{w(s,t)}_{\text{random effect}} \quad + \underbrace{\epsilon(s,t)}_{\text{iid noise}}$
- Goals:
  - Identify association with the covariates
  - Predict the response at an arbitrary location and time-point
- $w(s,t)$ is often modeled as a spatio-temporal Gaussian Process

Separable models

- $w(s,t) \sim GP(0, C(\cdot,\cdot \,|\, \theta))$
- $Cov(w(s_1,t_1), w(s_2,t_2)) = C_{S;\theta}(s_1,s_2)\; C_{T;\theta}(t_1,t_2)$
- $w = (w(s_1,t_1), w(s_2,t_1), \ldots, w(s_K,t_T))'$
- $Cov(w) = C_S(\theta) \otimes C_T(\theta)$
- Problem reduces to modeling large $C_S(\theta)$ and $C_T(\theta)$
  - Replace C_S and C_T by their NNGP analogues $\tilde{C}_S$ and $\tilde{C}_T$ respectively
- Separable models do not allow for space-time interaction (Cressie and Huang 1999)

Non-Separable models

- Non-separable covariance function (e.g. Gneiting (2001)):

$$C((s+h,t+u),(s,t)) = \frac{\sigma^2}{(\phi_1|u|^2+1)^k} \exp\left(\frac{-\phi_2||h||}{(\phi_1|u|^2+1)^{\frac{k}{2}}}\right)$$

- Allows space-time interaction
- $Cov(w) = C(\theta)$ is dense and cannot be decomposed

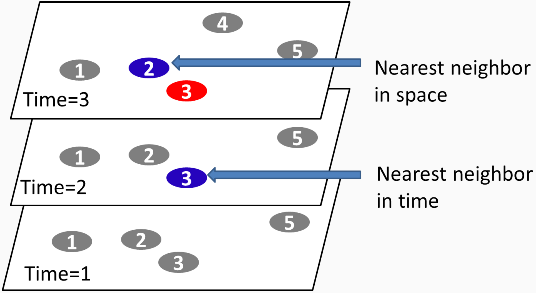
Non-separable NNGP for spatio-temporal data

- $w(s,t) \sim GP(0, C(\cdot,\cdot \,|\, \theta))$ (Non-separable covariance function)
- $R = \{(s_i, t_j) \,|\, i = 1,2,\ldots,K, j = 1,2,\ldots,T\}$ is the grid where the data is observed
- $w = (w(s_1,t_1), w(s_2,t_1),\ldots,w(s_K,t_T))'$ is the realizations of the GP over R

Non-separable NNGP for spatio-temporal data

- Let $H(s_i, t_j) = \{(s_{i'}, t_{j'}) \,|\, j' < j \text{ (or if } j' = j \text{ then } i' < i)\}$
- For the full GP, $p(w) = N(0, C) = \prod_{j=1}^{T} \prod_{i=1}^{K} p(w(s_i, t_j) \,|\, \{w(s, t) \,|\, (s, t) \in H_{ij}\})$
- A space-time NNGP can be derived by replacing the conditioning sets $H(s_i, t_j)$ with smaller neighbor sets $N(s_i, t_j) \subset H(s_i, t_j)$ of size m
- Storage and computational complexity similar to spatial NNGP, i.e., $O(n)$ where $n = KT$

Neighbors in Space and Time



- For spatial NNGP, neighbors were chosen based on Euclidean distances
- No universal definition of distance in a space-time domain

Adaptive neighbor sets

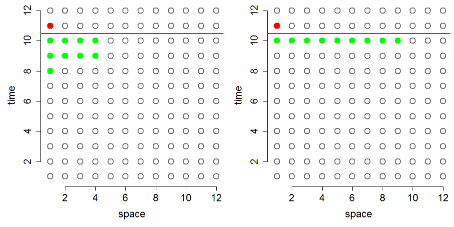
- Most popular spatial covariance functions decrease with increasing distance between locations
- So for spatial data, choosing nearest neighbors make sense as they correspond to locations with highest correlations with the given location

Adaptive neighbor sets

- Most popular spatial covariance functions decrease with increasing distance between locations
- So for spatial data, choosing nearest neighbors make sense as they correspond to locations with highest correlations with the given location
- For spatio-temporal data, if θ is known, we can use $C(\cdot, \cdot \mid \theta)$ directly to choose the neighbor sets
- Construct adaptive neighbor sets $N_\theta(s_i, t_j)$ using m-'nearest neighbors' based on $C_\theta(\cdot, \cdot)$ – Dynamic NNGP

Computational Roadblock

- Neighbor sets now depend on θ



$N_{\theta_1}(s,t)$ $N_{\theta_2}(s,t)$

Figure: Adaptive neighbor sets (green) of the red point for different choices of θ

- Need to be updated after every update of θ in the MCMC
- Computationally inefficient:
    - Need to calculate pairwise correlations for all locations
    - $O(n^2)$ flops

Updating Neighbor Sets

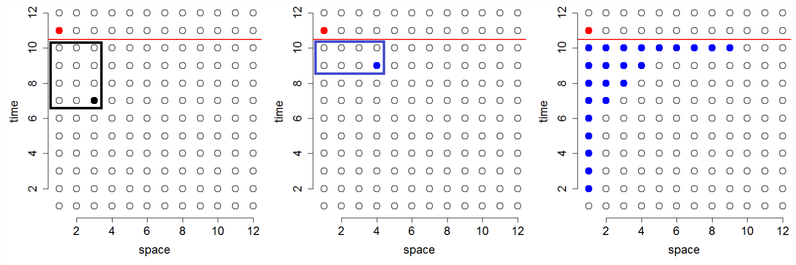
Eligible sets

For every (s_i, t_j) in R one can construct an *'eligible set'* $E(s_i, t_j)$ such that

(a) $E(s_i, t_j)$ does not depend on θ

(b) For every value of θ, $N_\theta(s_i, t_j) \in E(s_i, t_j)$

(c) For $m \sim 20$, $|E(s_i, t_j)| \sim 4m$ for every i, j



Not eligible Eligible Eligible set $(m = 9)$

Updating Neighbor Sets

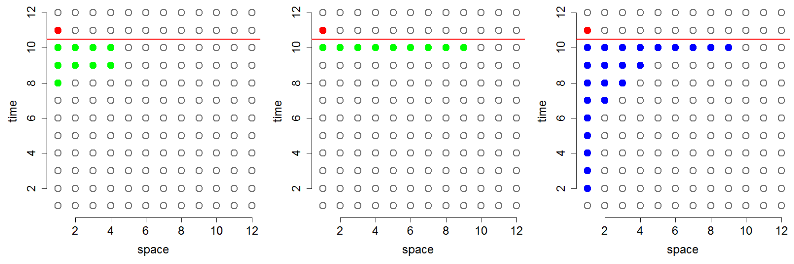
Eligible sets

For every (s_i, t_j) in R one can construct an *'eligible set'* $E(s_i, t_j)$ such that

(a) $E(s_i, t_j)$ does not depend on θ

(b) For every value of θ, $N_\theta(s_i, t_j) \in E(s_i, t_j)$

(c) For $m \sim 20$, $|E(s_i, t_j)| \sim 4m$ for every i, j



$N_{\theta_1}(s, t)$ $N_{\theta_2}(s, t)$ Eligible set $(m = 9)$

Updating Neighbor Sets

Eligible sets

For every (s_i, t_j) in R one can construct an *'eligible set'* $E(s_i, t_j)$ such that

(a) $E(s_i, t_j)$ does not depend on θ

(b) For every value of θ, $N_\theta(s_i, t_j) \in E(s_i, t_j)$

(c) For $m \sim 20$, $|E(s_i, t_j)| \sim 4m$ for every i, j

- $E(s_i, t_j)$ needs to be constructed only once
- For every update of θ, search for $N_\theta(s_i, t_j)$ only in $E(s_i, t_j)$
- Total computational cost for this step $= O(4nm)$ i.e., at par with the rest of the sampler – Datta et al., AOAS, (2016)

Simulation Experiments

- 225 locations on a 15×15 grid within a unit square
- 20 time steps within a 0 to 1 range
- $y(s, t) = \beta_0 + \beta_1 x(s, t) + w(s, t) + \epsilon(s, t)$
- $w(\mathbf{s}, t) \sim$ GP with non-separable space time covariance function

$$K(h; u) = \frac{\sigma^2}{(\phi_1|u|^2 + 1)} \exp\left(\frac{-\phi_2||h||}{(\phi_1|u|^2 + 1)^{0.5}}\right)$$

Model evaluation

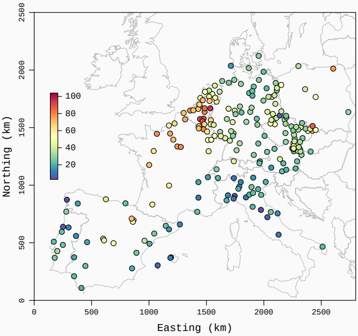
	DNNGP $m = 25$	Predictive Process 64 knots	Full Gaussian Process
DIC score	3866	7012	3988
RMSPE	0.53	0.71	0.53
Run time (Hours)	8	14	127

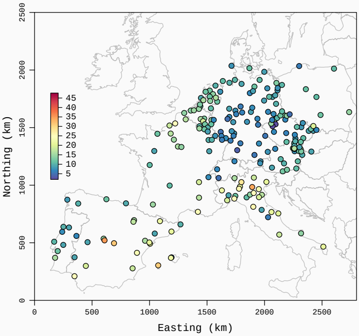
- DNNGP performs at par with Full GP, PP performs worse
- DNNGP yields huge computational gains

European Particulate Matter Dataset

- Particulate matter (PM)
  - Environmental pollutant associated with increased human morbidity and mortality
  - PM_{10} (PM with diameter $< 10\mu m$)
- EU countries face legal action if PM_{10} exceeds 50 $\mu g \ m^{-3}$ for more than 35 days per year
- Accurate high-resolution regional space-time PM maps required for monitoring compliance

European PM_{10} Dataset



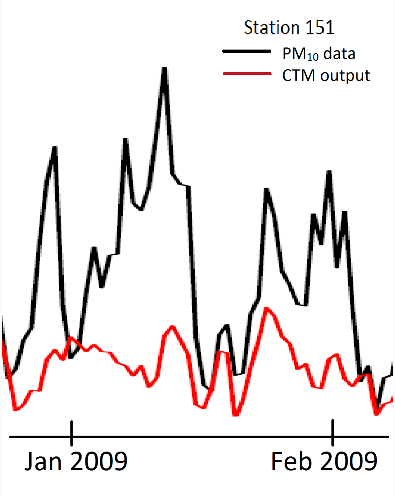


PM_{10} levels in March, 2009

PM_{10} levels in June, 2009

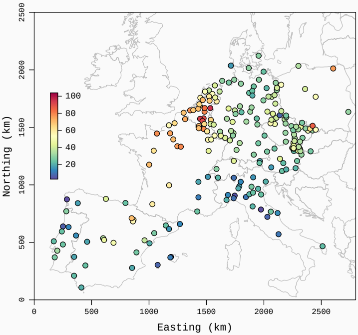
- Significant variation across space and time
- Daily observations at 308 stations for 2 years i.e., $n = 308 \times 730 = 224,840$

European PM_{10} Dataset

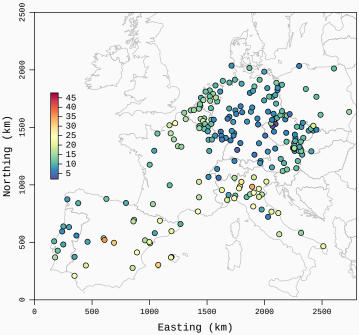


- Computer models like Chemistry Transport Model (CTM) consistently underestimate PM_{10} levels
- CTM outputs used as covariates to improve fits $\log(PM_{10})(s,t) = \beta_0 + \beta_1 CTM(s,t) + \epsilon(s,t)$

European PM_{10} data



PM_{10} levels in March, 2009

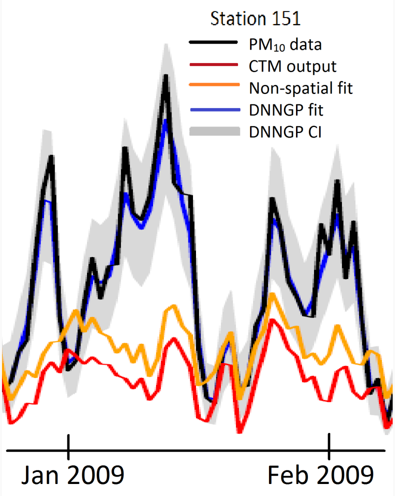


PM_{10} levels in June, 2009

Model

- $\log(PM_{10})(s,t) = \beta_0 + \beta_1 CTM(s,t) + w(s,t) + \epsilon(s,t)$
- $w(s,t) \sim \color{red}{DNNGP}(0, \tilde{K}_\theta)$

European PM_{10} Dataset

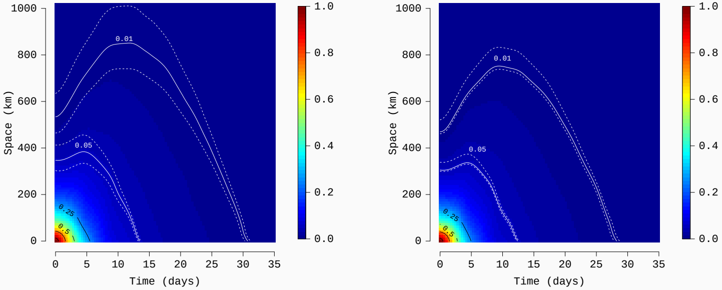


- Significantly improved fit

	OLS	DNNGP
RMSPE	12.8	8.2

- Total time 24 hrs

European PM10 Dataset

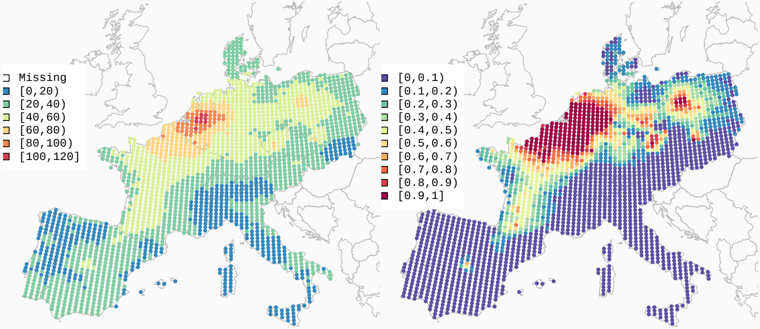


m=10 m=20

Figure: Space-time correlation posterior distribution median surfaces. Median (white lines) and associated 95% credible intervals (dotted white lines) for correlations of 0.05 and 0.01.

- Posterior distribution of the space-time covariance parameters provides insight into the residual spatio-temporal structure

European PM_{10} Dataset

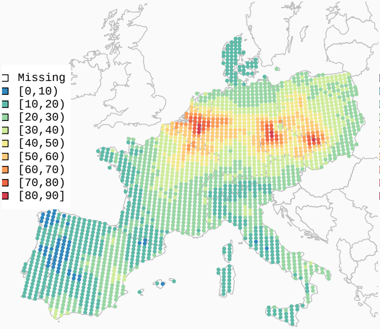


$\widehat{\mathrm{PM}_{10}}$ for 04.03.2009

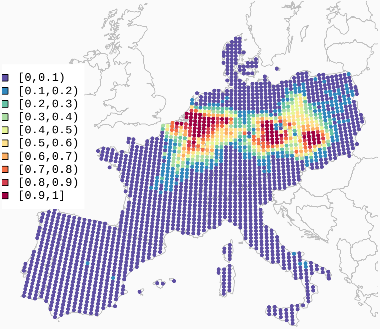
$Pr(\widehat{\mathrm{PM}_{10}} > 50\mu gm^{-3})$

Figure: Posterior predictve maps of PM_{10} and of probability that PM_{10} exceeds the legal threshold

European PM_{10} Dataset



$\widehat{\mathrm{PM}_{10}}$ for 04.05.2009



$Pr(\widehat{\mathrm{PM}_{10}} > 50\mu gm^{-3})$

Figure: Posterior predictve maps of PM_{10} and of probability that PM_{10} exceeds the legal threshold

Summary

- Spatio-temporal regression models – discrete time continuous space, separable and non-separable continuous space time models
- Dynamic NNGP for non-separable models for large spatio-temporal data
- Neighbor sets chosen based on strength of spatio-temporal covariance
- Fast algorithm to update the neighbor sets
- Retains all computational advantages of spatial NNGP: total requirements is $O(n)$
- Performs at par with the original non-separable GP
- Proper Gaussian process: Fully Bayesian inference, produces a variety of space-time forecast maps at arbitrary resolution