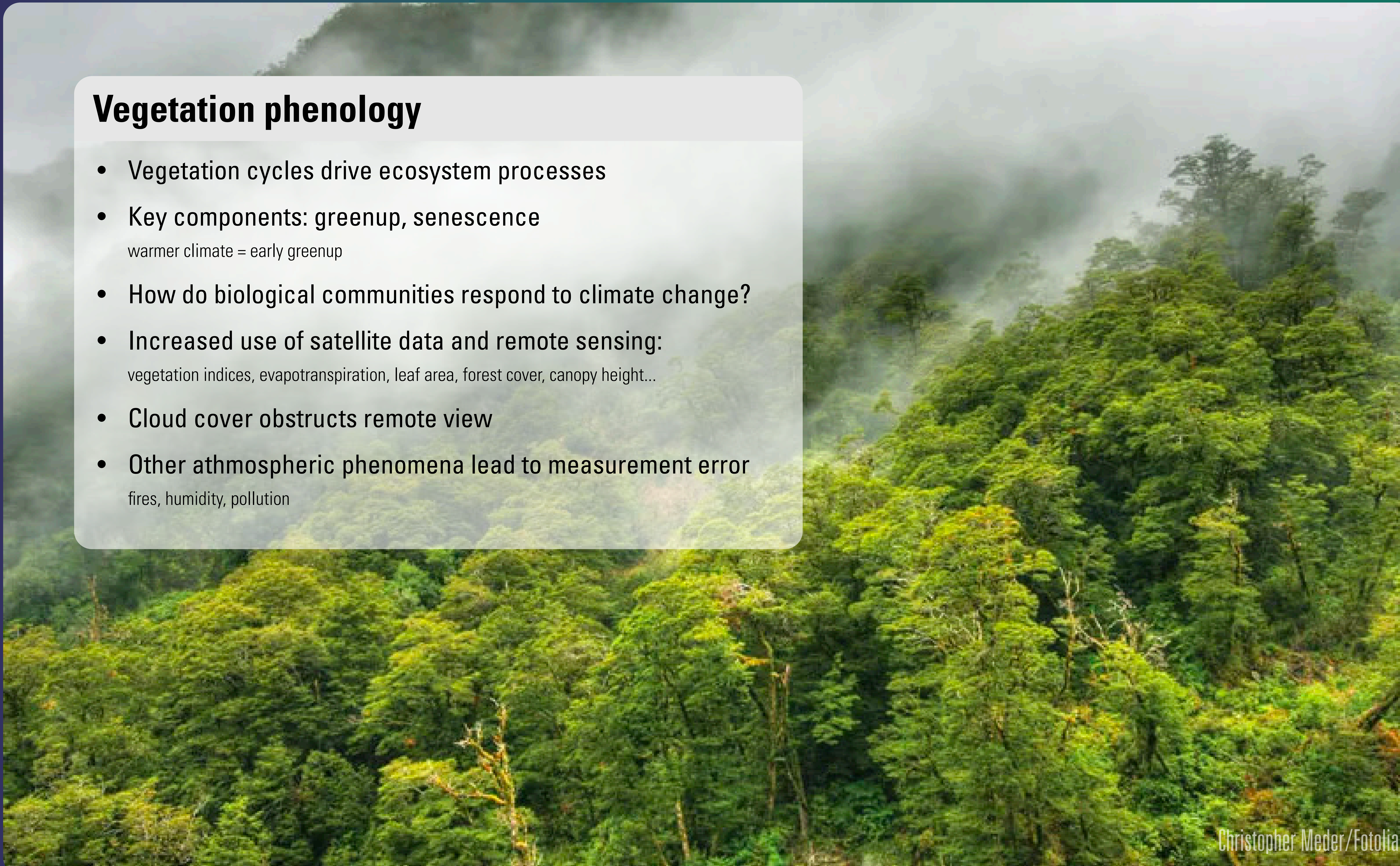
Designer DAGs:
Bayesian geostatistics with massive data

Michele Peruzzi
Postdoctoral Associate
Duke University
Dept of Statistical Science

presenting joint work with:
David Dunson (Duke)
Sudipto Banerjee (UCLA)
Andrew Finley (MSU)
Bora Jin (Duke)
Yichen Zhu (Duke)
Naresh Neupane (Georgetown)

Vegetation phenology

- Vegetation cycles drive ecosystem processes
- Key components: greenup, senescence
  warmer climate = early greenup
- How do biological communities respond to climate change?
- Increased use of satellite data and remote sensing:
  vegetation indices, evapotranspiration, leaf area, forest cover, canopy height...
- Cloud cover obstructs remote view
- Other athmospheric phenomena lead to measurement error
  fires, humidity, pollution



Christopher Meder/Fotolia

Vegetation phenology

- Vegetation cycles drive ecosystem processes
- Key components: greenup, senescence

  warmer climate = early greenup
- How do biological communities respond to climate change?
- Increased use of satellite data and remote sensing:

  vegetation indices, evapotranspiration, leaf area, forest cover, canopy height...
- Cloud cover obstructs remote view
- Other athmospheric phenomena lead to measurement error

  fires, humidity, pollution

Air quality & pollutant exposure monitoring

- Poor air quality linked to adverse health outcomes
- Acute vs chronic exposures
- Ground-level monitors vs satellite imaging
- What health effects? Interactions with other exposures?

Air quality & pollutant exposure monitoring

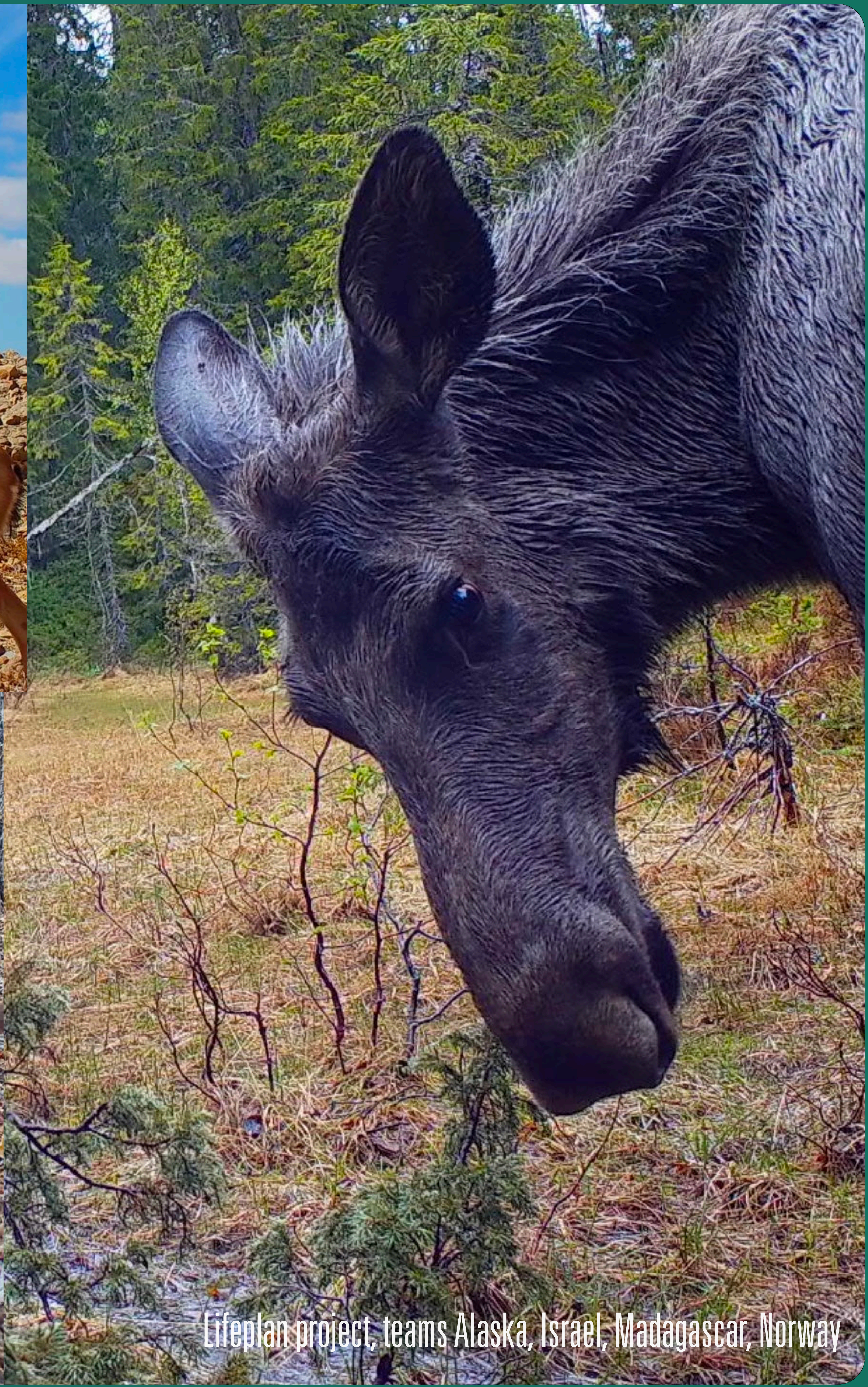
- Poor air quality linked to adverse health outcomes
- Acute vs chronic exposures
- Ground-level monitors vs satellite imaging
- What health effects? Interactions with other exposures?


Max Whittaker for The New York Times

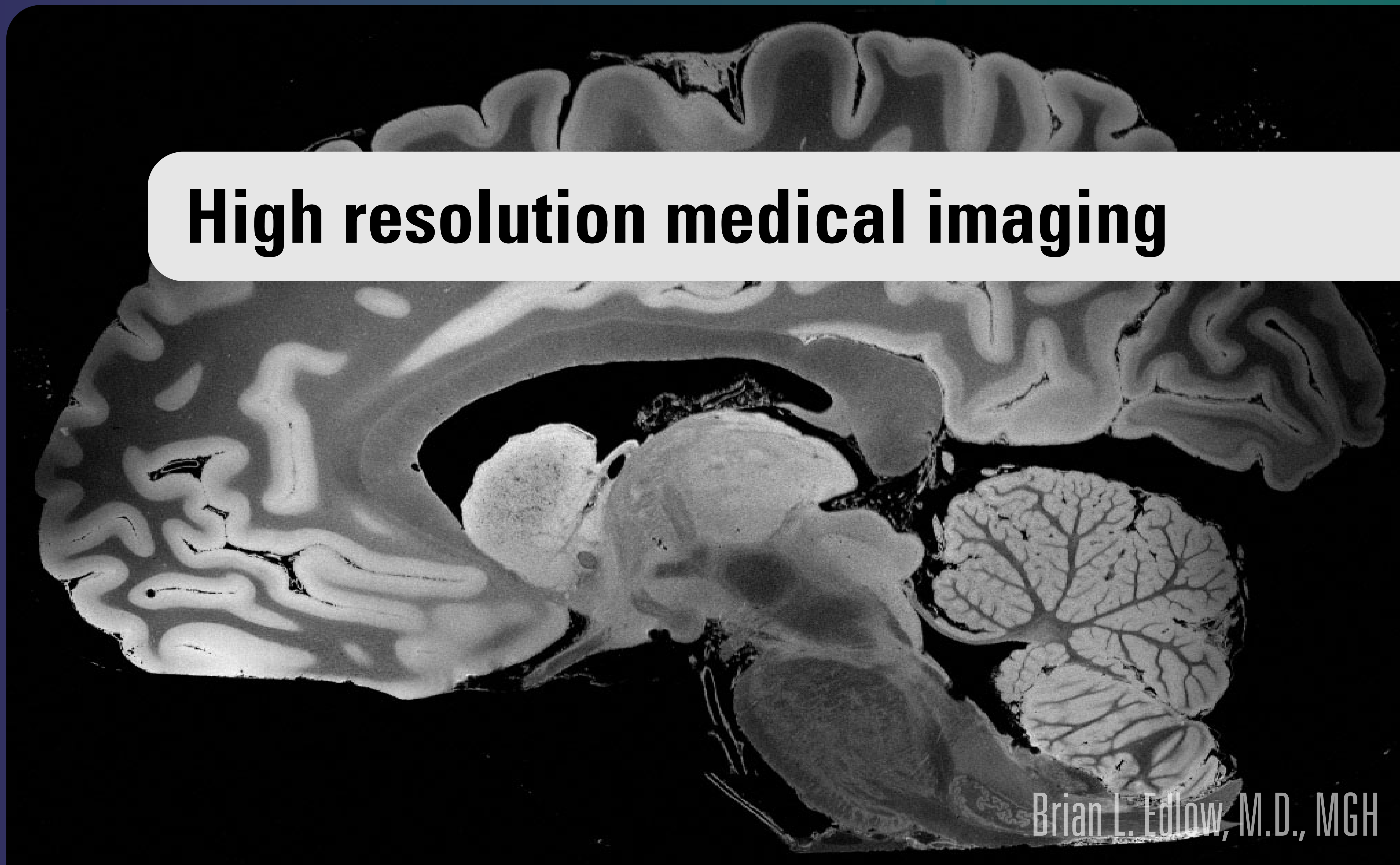

NASA

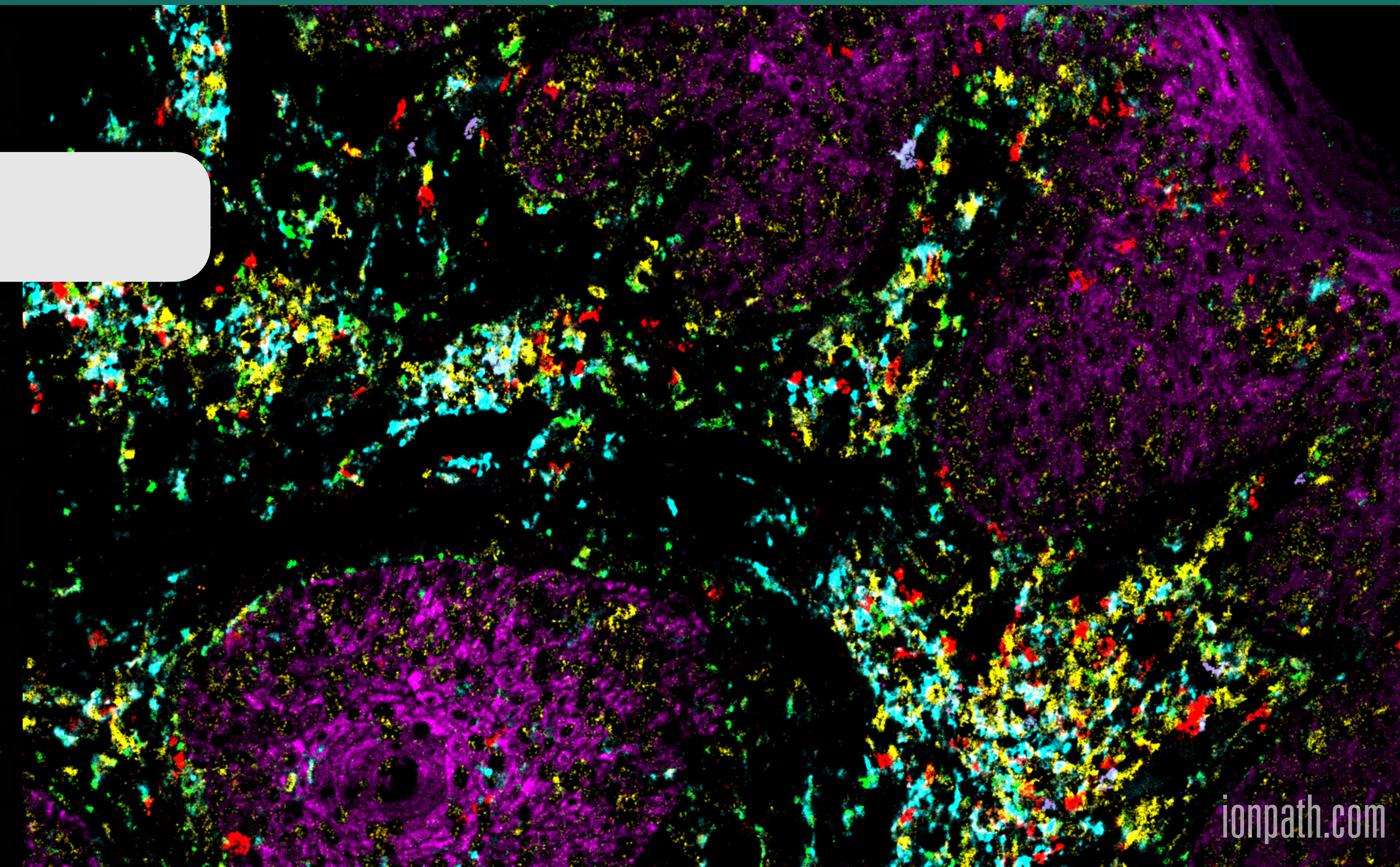
Species communities & interactions

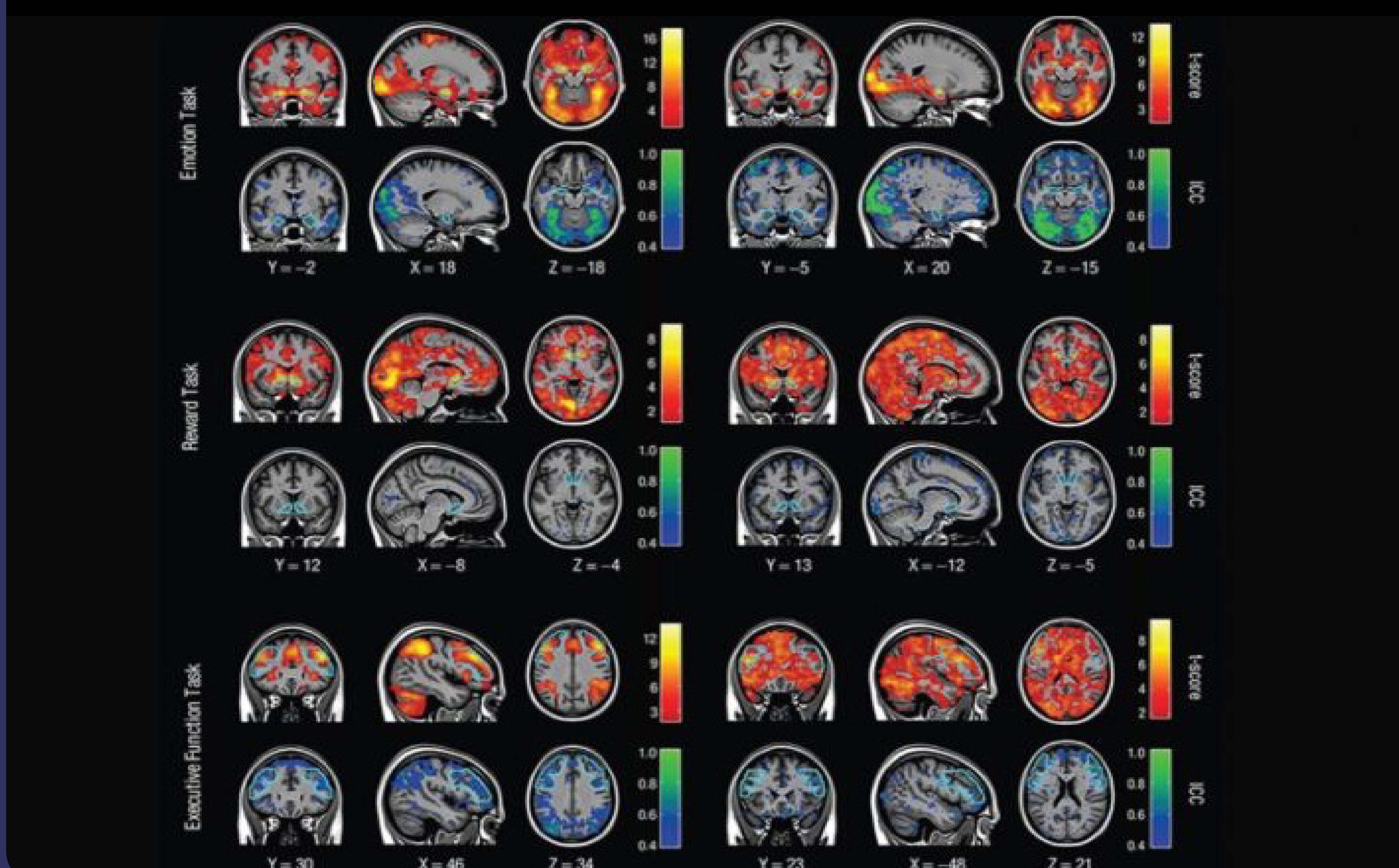
- Occurrence of species & their interactions depend on environmental factors
- How do species adapt to climate and environmental change?
- How does species richness vary spatially?
- What are the effects of climate change on richness?

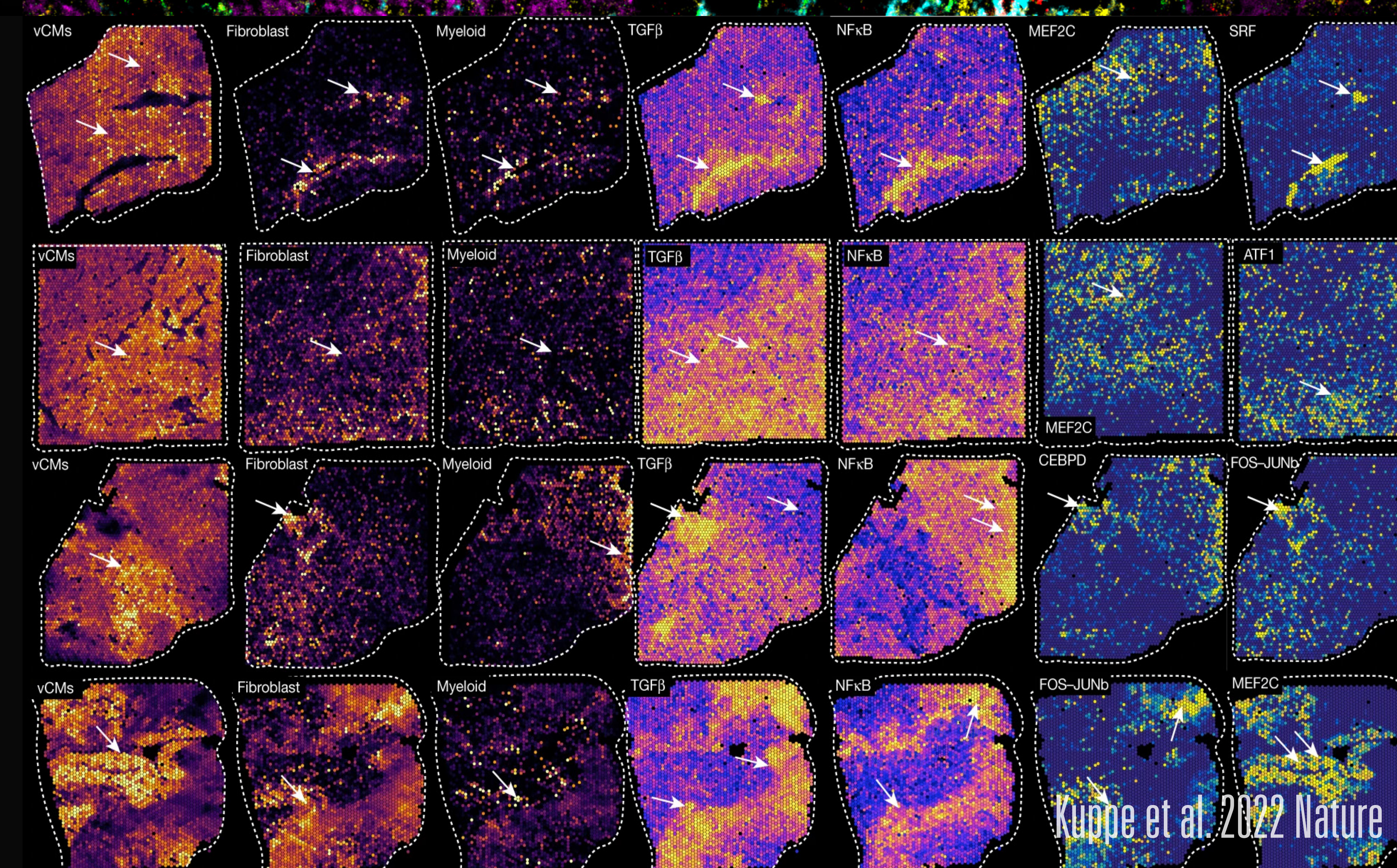


High resolution medical imaging









Plan of action

- Spatial latent effect modeling of multivariate outcomes
- Spatial process modeling with **Designer DAGs**
- Designer DAGs

  for images with gaps: *Cubic Meshed Gaussian Processes*

  for multi-source data: *Spatial Multivariate Trees*

  for multivariate non-Gaussian data: *Simplified Manifold Preconditioner Adaptation*

  for directional nonstationarity: *Bags of Directed Acyclic Graphs*

  for provable accuracy in approximating GPs: *Radial Neighbors Gaussian Process*
- Additional topics

  Multivariate geostatistics with R package `meshed`

  Gridding and parameter expansion for improving computations in challenging settings

  Graph Machine Regression & R package `gramar`

Setting up things: the Bayesian paradigm

obligatory Thomas Bayes picture

Bayesian paradigm:

- *probability model* for data
- *prior* distribution:
  uncertainty about model parameters
- *hyperprior* distribution(s):
  uncertainty about prior parameters

Setting up things: the Bayesian paradigm

obligatory Thomas Bayes picture

Bayesian paradigm:

- *probability model* for data
- *prior* distribution:
  uncertainty about model parameters
- *hyperprior* distribution(s):
  uncertainty about prior parameters

→ observe data

$$\boldsymbol{y}_i$$
$$i = 1, \ldots, n$$

Setting up things: the Bayesian paradigm

obligatory Thomas Beyes picture

Bayesian paradigm:

- *probability model* for data
- *prior* distribution:
  uncertainty about model parameters
- *hyperprior* distribution(s):
  uncertainty about prior parameters

→ observe data $\boldsymbol{y}_i$, $i = 1, \ldots, n$ → ***posterior*** uncertainty about model parameters

Bayesian hierarchical model with spatial random effects

Suppose we observe data at ***n*** locations (coordinates / pixels / voxels)

Probability model for outcome *j* $\quad y_j(\boldsymbol{\ell}) \mid \eta_j(\boldsymbol{\ell}), \tau_j \sim P_j(\eta_j(\boldsymbol{\ell}), \tau_j)$

$$\boldsymbol{\ell} \in \mathcal{D} \subset \Re^d \quad j = 1, \ldots, q$$

Linear predictor $\quad \eta_j(\boldsymbol{\ell}) = \boldsymbol{x}_j(\boldsymbol{\ell})^\top \boldsymbol{\beta}_j + w_j(\boldsymbol{\ell})$

features or observed covariates

latent/random/unobserved effects for dependent data

example: Gaussian outcome $\quad y_j(\boldsymbol{\ell}) = \eta_j(\boldsymbol{\ell}) + \varepsilon_j(\boldsymbol{\ell}) \quad \varepsilon_j(\boldsymbol{\ell}) \sim N(0, \tau_j^2)$

vector of random effects at location $\boldsymbol{\ell}$ $\quad \boldsymbol{w}(\boldsymbol{\ell}) = \begin{bmatrix} w_1(\boldsymbol{\ell}) \\ \vdots \\ w_q(\boldsymbol{\ell}) \end{bmatrix}$

What prior for $\boldsymbol{w}(\boldsymbol{\ell})$ at ***any*** set of locations $\boldsymbol{L} = \{\boldsymbol{\ell}_1, \ldots, \boldsymbol{\ell}_n\}$?

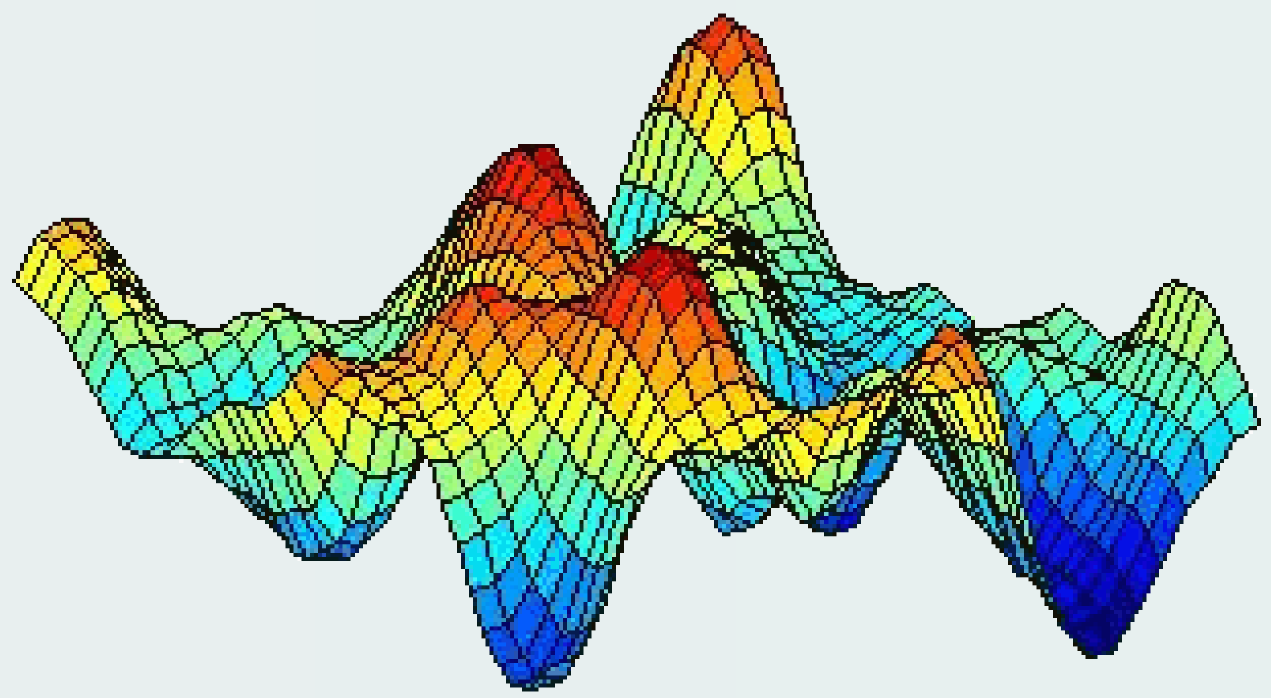
Gaussian process prior for spatial random effects

What prior for $\boldsymbol{w}(\boldsymbol{\ell})$ at ***any*** set of locations $\boldsymbol{L} = \{\boldsymbol{\ell}_1, \ldots, \boldsymbol{\ell}_n\}$?

q-variate Gaussian process
for correlating across space, time, outcomes

$$\boldsymbol{w}(\cdot) \sim GP(\mathbf{0}, \underline{\boldsymbol{C}_{\boldsymbol{\theta}}})$$

covariance function



for any $\boldsymbol{L}$, a GP gives $\boldsymbol{w}_L \sim N(\mathbf{0}, \boldsymbol{C}_L)$

$$\boldsymbol{C}_{\boldsymbol{L}}[i,j] = \boldsymbol{C}(\boldsymbol{\ell}_i, \boldsymbol{\ell}_j \mid \boldsymbol{\theta}) = \text{cov}(\boldsymbol{w}(\boldsymbol{\ell}_i), \boldsymbol{w}(\boldsymbol{\ell}_j))$$

Example with ***q=1***

Elements of $\boldsymbol{C}_{\boldsymbol{\theta}}$ matrix: Matérn model

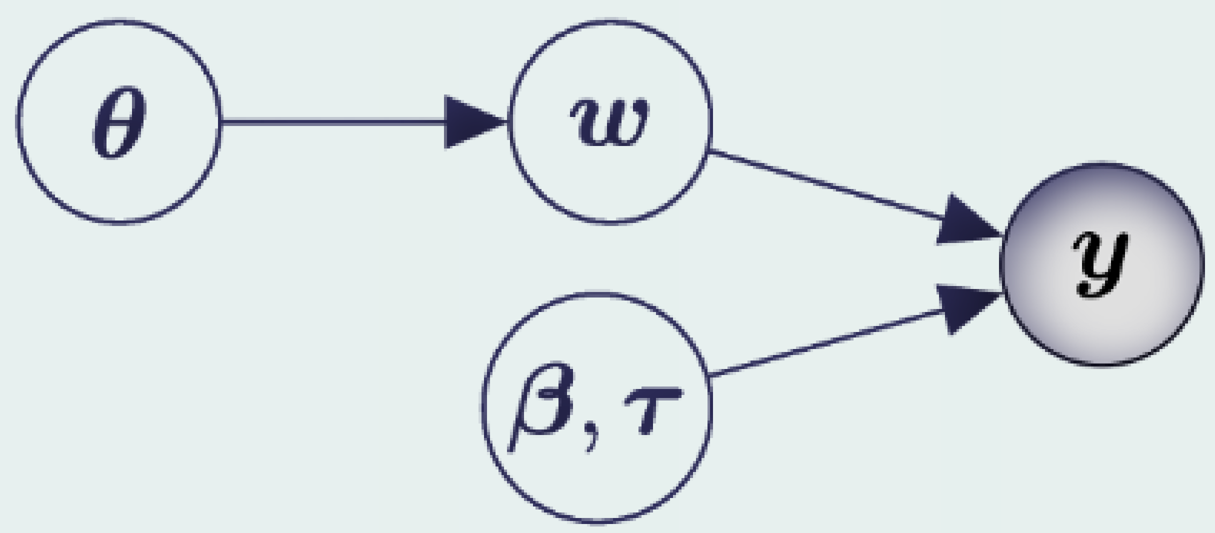
$$\boldsymbol{C}_{\boldsymbol{\theta}}(\boldsymbol{\ell}, \boldsymbol{\ell}') = \sigma^2 \frac{2^{1-\nu}}{\Gamma(\nu)} \phi^{\nu} \|\boldsymbol{\ell} - \boldsymbol{\ell}'\|^{\nu} K_{\nu}(\phi \|\boldsymbol{\ell} - \boldsymbol{\ell}'\|)$$

exponential covariance

$$(\nu = 0.5) \quad = \sigma^2 \exp\{-\phi \|\boldsymbol{\ell} - \boldsymbol{\ell}'\|\}$$

Gaussian Processes are slow: WHY?

Directed Acyclic Graph (DAG)
can be used to setup a **Gibbs sampler**
(procedure can even be automated, e.g. BUGS, JAGS & successors)



Gibbs sampler algorithm
requires knowledge about the *full conditional distributions*
outputs correlated samples from the joint posterior distribution

GIBBS SAMPLER

Cycle through these steps:

- sample $\boldsymbol{w} \mid \boldsymbol{y}, \boldsymbol{\theta}, \boldsymbol{\beta}, \boldsymbol{\tau}$
- sample $\boldsymbol{\theta} \mid \boldsymbol{w}$
- sample $\boldsymbol{\beta} \mid \boldsymbol{w}, \boldsymbol{y}, \boldsymbol{\tau}$
- sample $\boldsymbol{\tau} \mid \boldsymbol{y}, \boldsymbol{\theta}, \boldsymbol{w}$

With **Gaussian outcomes** we can also:

- Marginalize out $\boldsymbol{w}$ to get $\boldsymbol{y} \sim GP(\boldsymbol{X\beta}, \boldsymbol{C_\theta} + \boldsymbol{D})$
- Model $\boldsymbol{y}$ directly as a GP, i.e. $\boldsymbol{y} \sim GP(\boldsymbol{X\beta}, \widetilde{\boldsymbol{C}}_{\boldsymbol{\theta},\tau})$

COLLAPSED/RESPONSE GIBBS SAMPLER

Cycle through these steps:

- sample $\boldsymbol{\theta} \mid \boldsymbol{y}$
- sample $\boldsymbol{\beta} \mid \boldsymbol{y}, \boldsymbol{\tau}$
- sample $\boldsymbol{\tau} \mid \boldsymbol{y}, \boldsymbol{\beta}$

Gaussian Processes are slow: WHY?

WHEN UDPATING θ WE FIND:

$$p(\boldsymbol{\theta} \mid \boldsymbol{w}) \propto p(\boldsymbol{w} \mid \boldsymbol{\theta})p(\boldsymbol{\theta})$$

q-variate Gaussian process
for correlating across space, time, outcomes

$$\boldsymbol{w}(\cdot) \sim GP(\mathbf{0}, \boldsymbol{C}_{\boldsymbol{\theta}})$$

WE NEED TO EVALUATE:

$$p(\boldsymbol{w} \mid \boldsymbol{\theta}) = |2\pi \boldsymbol{C}_{\boldsymbol{\theta}}|^{-1/2} \exp\left\{-\frac{1}{2}\boldsymbol{w}^\top \boldsymbol{C}_{\boldsymbol{\theta}}^{-1}\boldsymbol{w}\right\}$$

Fill elements of matrix $\boldsymbol{C}_{\boldsymbol{\theta}}$ → Compute $\boldsymbol{C}_{\boldsymbol{\theta}}^{-1}$ and its determinant

complexity
$O(n^3q^3)$

LARGE DIMENSION
nq x nq

$n > 10^5$

A short summary...

- Gaussian Processes are **flexible**
- Gaussian Processes are **convenient**
- Gaussian Processes lead to **meaningful uncertainty quantification**

However...

- Gaussian Processes are **slow/do not scale** when:
  - the number of observed locations $\boldsymbol{n}$ is large
  - the number of observed outcomes $\boldsymbol{q}$ is large
- Need to use something
  - similarly flexible
  - scalable to big $\boldsymbol{n}$, big $\boldsymbol{q}$.

Gaussian Processes are slow: a quick look at some relevant literature

Gaussian Predictive Process & Inducing points

Quiñonero-Candela and Rasmussen, 2005; Snelson and Ghahramani, 2007; Banerjee et al. 2008; Banerjee et al. 2010; Guhaniyogi et al. 2011; Finley, Banerjee, and Gelfand 2012; Low et al., 2015; Ambikasaran et al., 2016; Huang and Sun, 2018; Geoga et al., 2020

Exploit data structure

Gilboa et al., 2015; Moran and Wheeler, 2020; Loper et al., 2020

Fixed Rank Kriging

Cressie and Johannesson 2008

Multi-resolution approximations

Gramacy and Lee 2008; Fox and Dunson 2012; Katzfuss 2017

Covariance Tapering

Furrer, Genton, and Nychka 2006; Kaufman, Schervish, and Nychka 2008; Bevilacqua et al., 2019

Independent partitioning

Sang and Huang 2012; Stein 2014

Composite likelihood

Bai et al., 2012; Eidsvik et al., 2014; Bevilacqua and Gaetan, 2015

Gaussian Random Markov Fields

Cressie 1993; Rue 2001; Rue and Held 2005

Vecchia's approximation & extensions

Vecchia 1988; Stein et al. 2014; Gramacy and Apley 2015; Datta et al. 2016; Guinness 2018; Heaton et al. 2019; Katzfuss and Guinness 2019; Quiroz et al., 2019; Schafer et al. 2021

Vecchia's approximation & extensions

Fix reference set of locations $\mathcal{S}$ and order on it, then express joint density as product of conditionals

$$p(\boldsymbol{w} \mid \boldsymbol{\theta}) = p(\boldsymbol{w}_1 \mid \boldsymbol{\theta}) p(\boldsymbol{w}_2 \mid \boldsymbol{w}_1, \boldsymbol{\theta}) \cdots p(\boldsymbol{w}_n \mid \boldsymbol{w}_1, \dots, \boldsymbol{w}_{n-1}, \boldsymbol{\theta})$$

Approximate by limiting the size of conditioning sets picking *m* nearest neighbors

$$p(\boldsymbol{w} \mid \boldsymbol{\theta}) \approx \prod_{i=1}^{n} p(\boldsymbol{w}_i \mid \boldsymbol{w}_{N_i}, \boldsymbol{\theta}) \qquad N_i = \{j > 0 : i - m \leq j < i\}$$

Approximation leads to valid joint density (Vecchia 1988):

$$\prod_{i=1}^{n} p(\boldsymbol{w}_i \mid \boldsymbol{w}_{N_i}, \boldsymbol{\theta}) = \widetilde{p}(\boldsymbol{w} \mid \boldsymbol{\theta})$$

Extensible to valid stochastic process (Datta et al 2016). For any set of locations $\mathcal{V}$:

$$\widetilde{p}(\boldsymbol{w}_{\mathcal{V}}) = \int \widetilde{p}(\boldsymbol{w}_{\mathcal{U}} \mid \boldsymbol{w}_{\mathcal{S}}) \widetilde{p}(\boldsymbol{w}_{\mathcal{S}}) \prod_{\{\boldsymbol{s}_i \in \mathcal{S} \setminus \mathcal{V}\}} d(\boldsymbol{w}(\boldsymbol{s}_i))$$

Vecchia's approximation & extensions

Joint density

$$p(A, B, C)$$

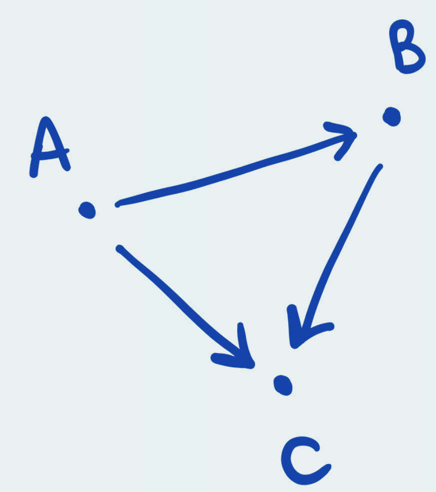
→

Product of conditional densities

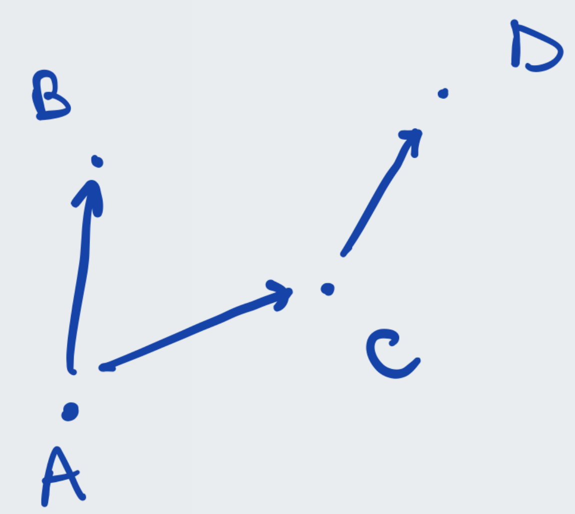
$$p(A)p(B \mid A)p(C \mid A, B)$$

→

Directed Acyclic Graph (DAG)



Why Vecchia/NNGP works: DAGs can be used to create new valid densities/processes



Directed Acyclic Graph (DAG)

→

Product of conditional densities

$$p(A)p(B \mid A)p(C \mid A)p(D \mid C)$$

→

Joint density

$$p(A, B, C, D)$$

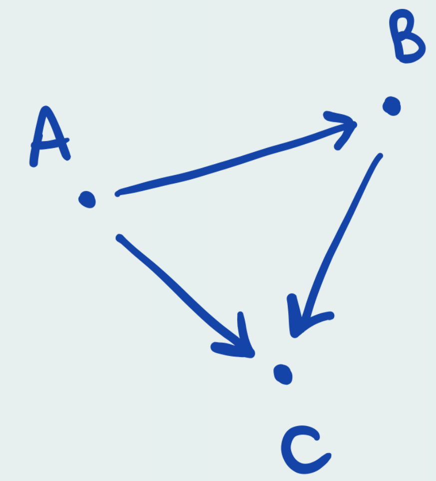
Neighbor-based constructions do not exploit full potential of DAG-based models:

- Challenges to **scalability** in fully Bayesian analyses of diverse data types
- Limited **flexibility** in modeling multivariate data with unusual configurations
- Lack of **theory** support of Vecchia approximations
- No clear path for **intepretable modeling** of certain nonstationary phenomena

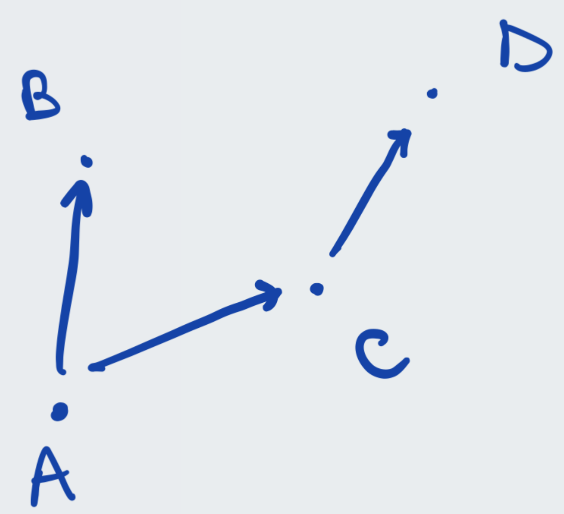
Vecchia's approximation & extensions

Joint density → Product of conditional densities → Directed Acyclic Graph (DAG)

$p(A, B, C)$ → $p(A)p(B \mid A)p(C \mid A, B)$



Why Vecchia/NNGP works: DAGs can be used to create new valid densities/processes



Directed Acyclic Graph (DAG) → Product of conditional densities → Joint density

$p(A)p(B \mid A)p(C \mid A)p(D \mid C)$ → $p(A, B, C, D)$

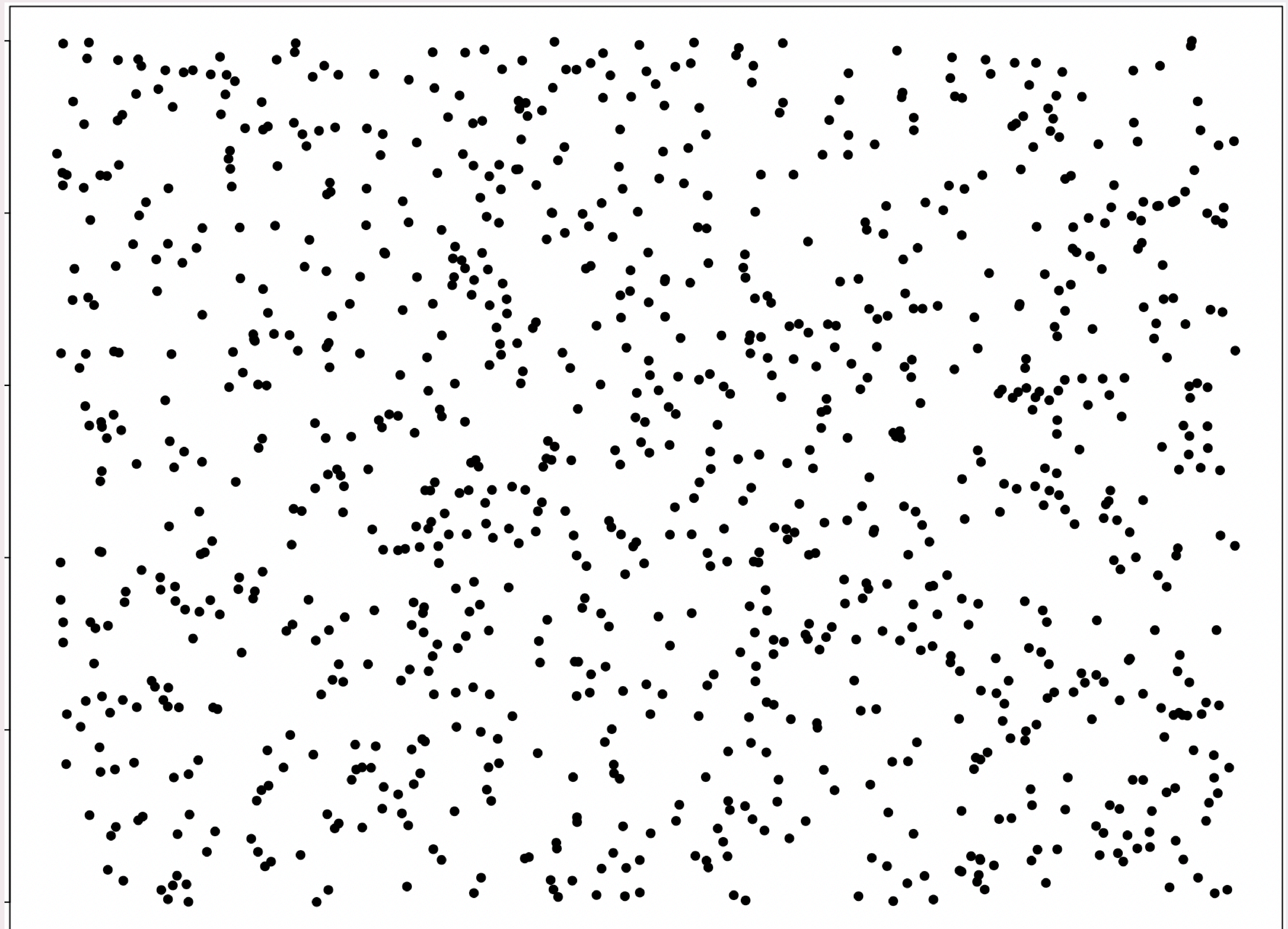
My contributions: carefully designing DAGs for

- Better **scalability** in fully Bayesian analyses of diverse data types
- Better **flexibility** in modeling multivariate data with unusual configurations
- New **theory** on quality of approximation of unrestricted GP
- Innovative **interpretable modeling** of certain nonstationary phenomena

Designer DAGs

Spatial meshing: spatial processes via designer DAGs

Defining a stochastic process = characterize the distribution of any finite set of random variables
Spatial process = each random variable is associated to a spatial coordinate (location)

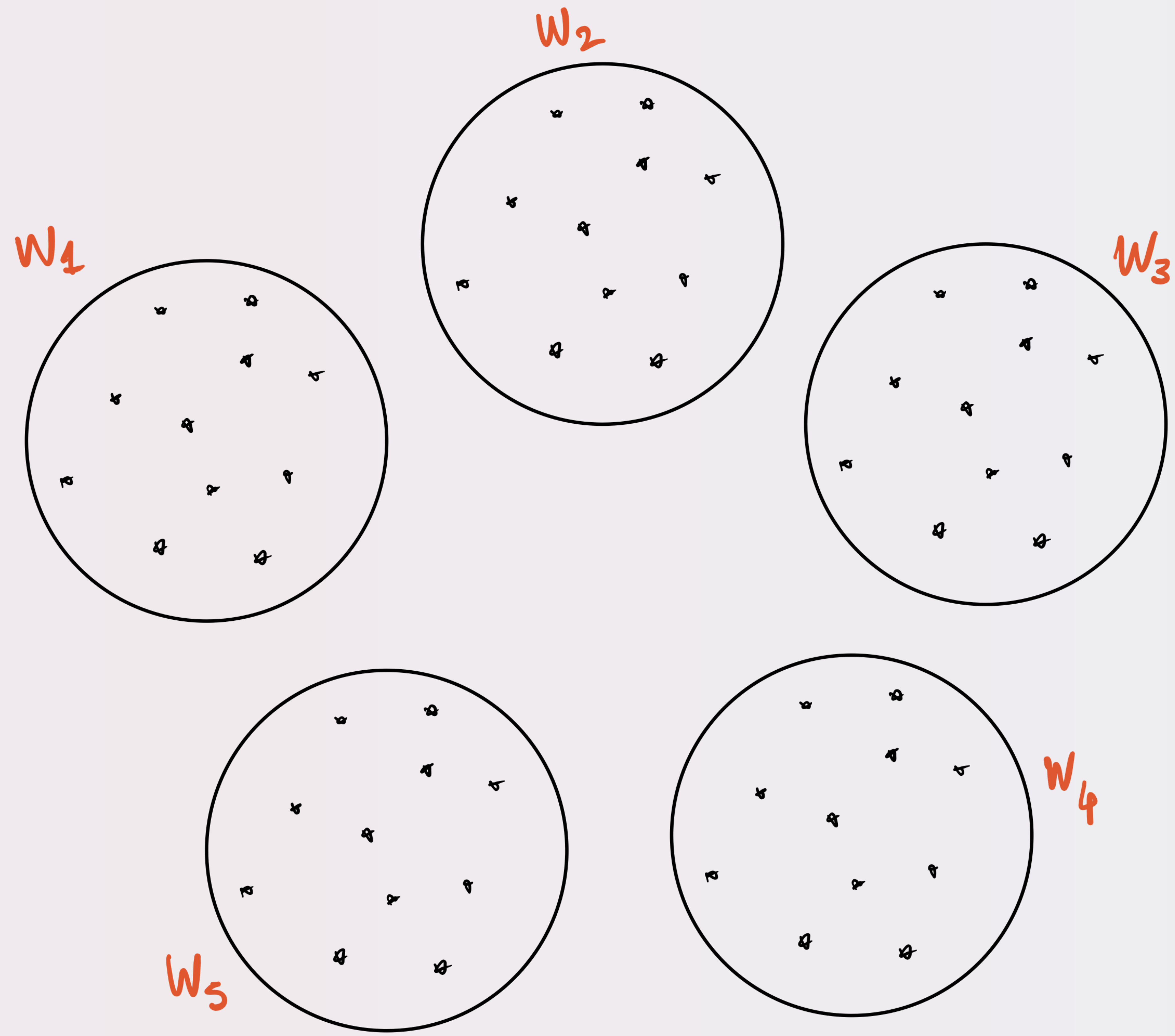


Data locations

Spatial meshing: spatial processes via designer DAGs

Fix a reference set of locations $\mathcal{S}$, then partition process realizations at $\mathcal{S}$ into $\boldsymbol{M}$ blocks

Random variables in each block are fully connected (edges hidden here)

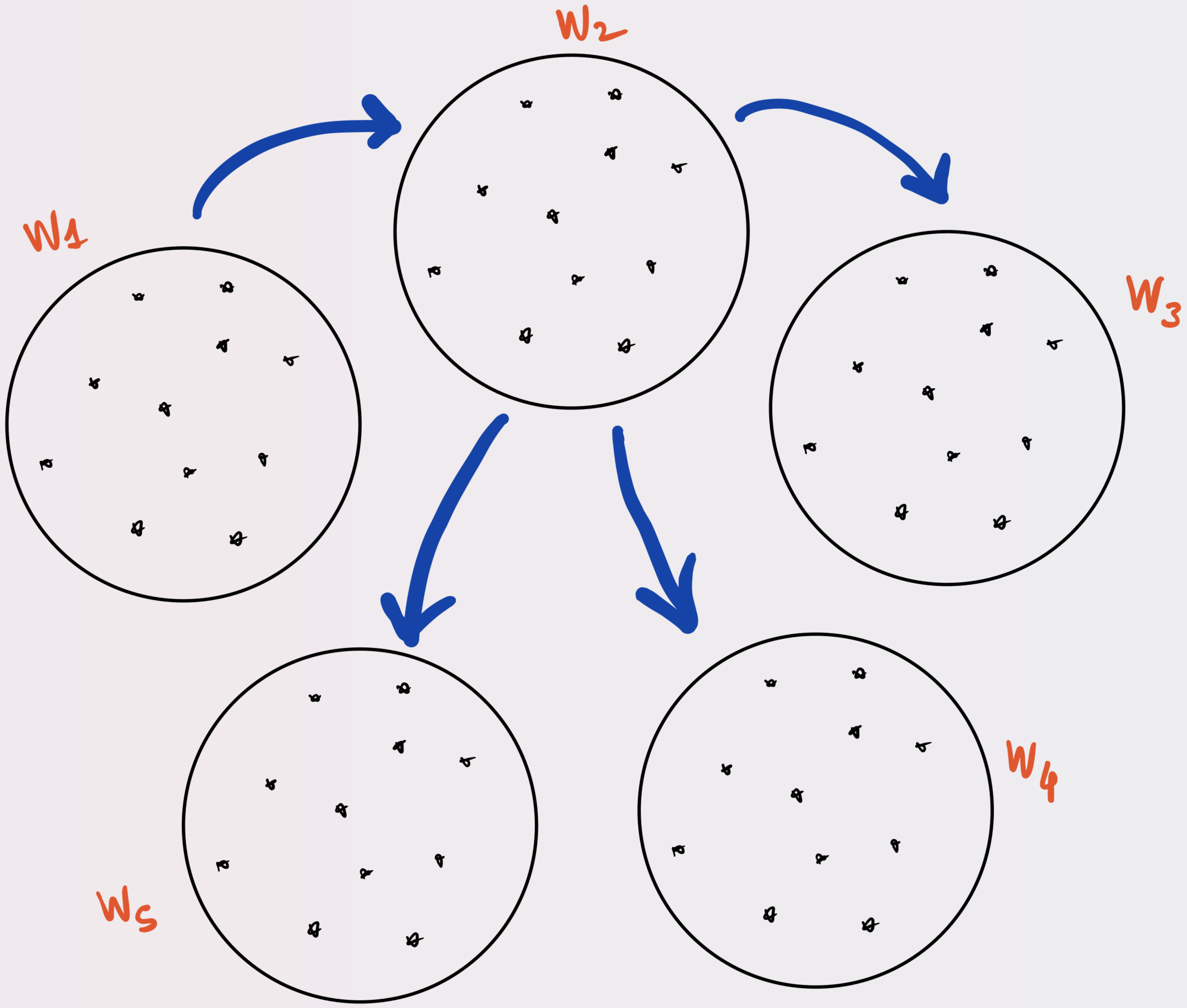


Spatial meshing: spatial processes via designer DAGs

Place edges between blocks to create a sparse DAG

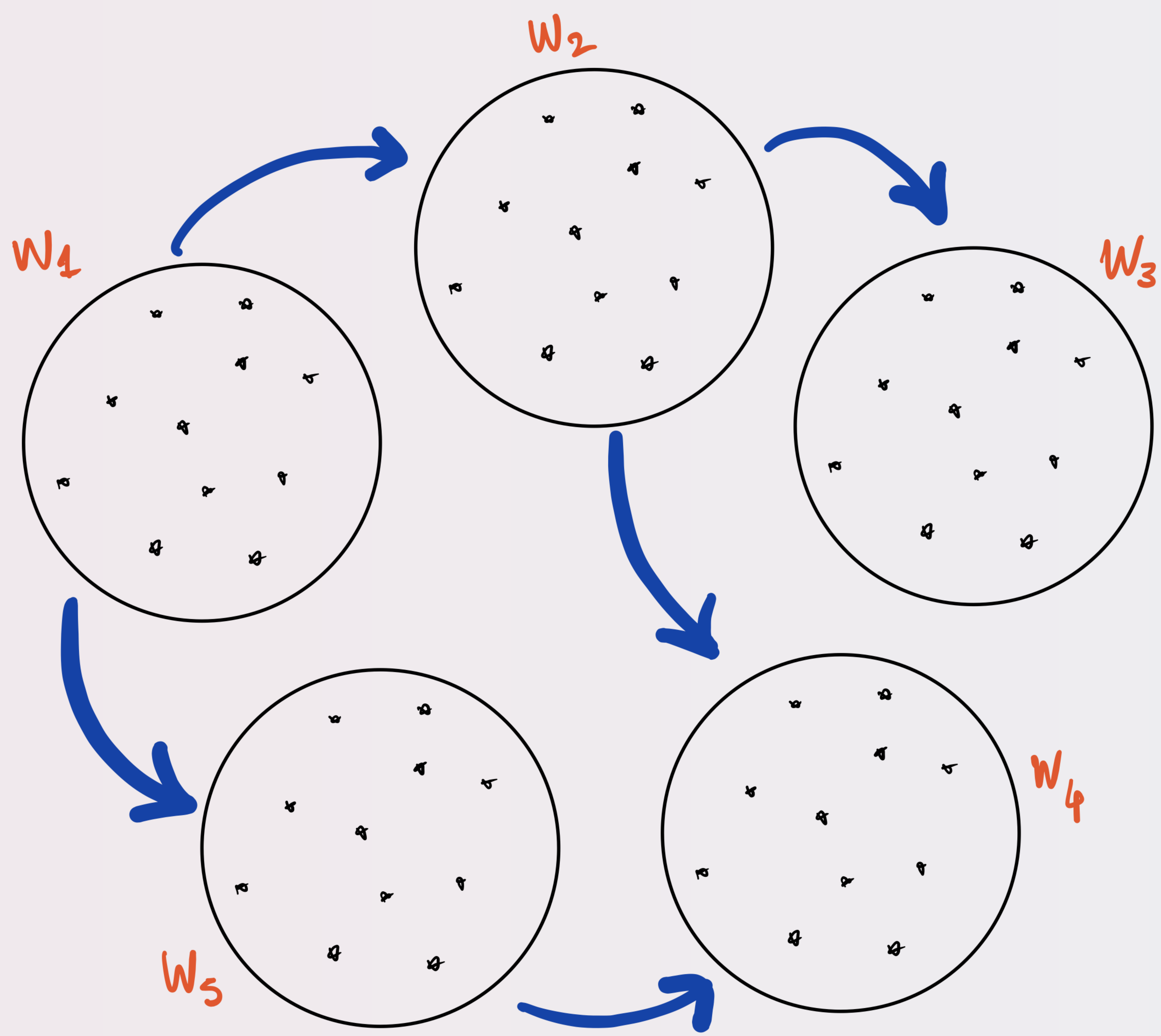
Example:
block 2 is the parent node of block 3

$$\boldsymbol{w}_{[3]} = \boldsymbol{w}_2$$



Spatial meshing: spatial processes via designer DAGs

Place edges between blocks to create a sparse DAG



Example: blocks {2,5} are the parent nodes of block 4

$$\boldsymbol{w}_{[4]} = \begin{bmatrix} \boldsymbol{w}_2 \\ \boldsymbol{w}_5 \end{bmatrix}$$

Spatial meshing: spatial processes via designer DAGs

Approximate by assuming joint density factorizes according to chosen DAG:

$$p(\boldsymbol{w} \mid \boldsymbol{\theta}) \approx \prod_{j=1}^{M} p(\boldsymbol{w}_j \mid \boldsymbol{w}_{[j]}, \boldsymbol{\theta}) \qquad [j] = \{i : i \to j \text{ in } \mathcal{G}\}$$

Approximation leads to valid density & can be extended to standalone stochastic process via Kolmogorov conditions (*P et al 2022 JASA*)

$$\prod_{j=1}^{M} p(\boldsymbol{w}_j \mid \boldsymbol{w}_{[j]}, \boldsymbol{\theta}) = \widetilde{p}(\boldsymbol{w} \mid \boldsymbol{\theta})$$

For any set $\mathcal{V}$ of locations, we have

$$\widetilde{p}(\boldsymbol{w}_{\mathcal{V}}) = \int \widetilde{p}(\boldsymbol{w}_{\mathcal{U}} \mid \boldsymbol{w}_{\mathcal{S}}) \widetilde{p}(\boldsymbol{w}_{\mathcal{S}}) \prod_{\{\boldsymbol{s}_i \in \mathcal{S} \setminus \mathcal{V}\}} d(\boldsymbol{w}(\boldsymbol{s}_i))$$

Where other locations $\mathcal{U}$ are similarly partitioned and can only have parents in $\mathcal{S}$.

Spatial meshing: spatial processes via designer DAGs

Assuming a Gaussian base density gives us

$$\boldsymbol{H}_i = \boldsymbol{C}_{i,[i]} \boldsymbol{C}_{i,[i]}^{-1}$$

$$\boldsymbol{R}_i = \boldsymbol{C}_i - \boldsymbol{C}_{i,[i]} \boldsymbol{C}_{i,[i]}^{-1} \boldsymbol{C}_{[i],i}$$

$$\prod_{i=1}^{M} N(\boldsymbol{w}_i; \boldsymbol{H}_i \boldsymbol{w}_{[i]}, \boldsymbol{R}_i) = N(\boldsymbol{w}; \boldsymbol{0}, \widetilde{\boldsymbol{C}}_{\boldsymbol{\theta}}) \approx N(\boldsymbol{w}; \boldsymbol{0}, \boldsymbol{C}_{\boldsymbol{\theta}})$$

Complexity (number of flops) for evaluating $N(\boldsymbol{w}; \boldsymbol{0}, \widetilde{\boldsymbol{C}}_{\boldsymbol{\theta}})$:

$$O(MJ^3m^3)$$

#blocks (→ M), max #parents in DAG (→ J), max block size (→ m)

Assuming blocks of equal size: $O(nJ^3m^2) \ll O(n^3)$

Spatial meshing: spatial processes via designer DAGs

Assuming a Gaussian base density gives us

$$\boldsymbol{H}_i = \boldsymbol{C}_{i,[j]}\boldsymbol{C}_{[j]}^{-1}$$

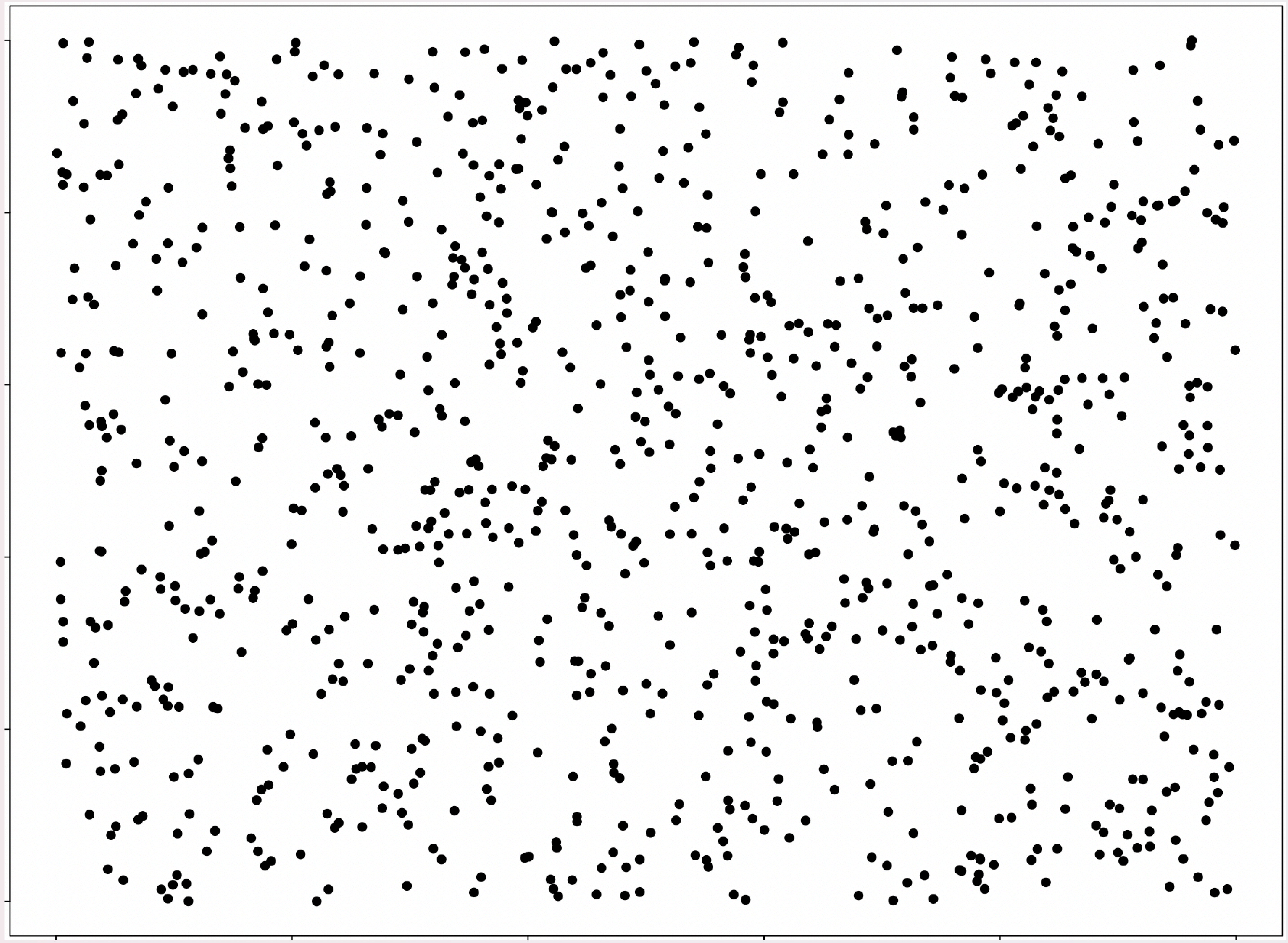
$$\boldsymbol{R}_i = \boldsymbol{C}_i - \boldsymbol{C}_{i,[j]}\boldsymbol{C}_{[j]}^{-1}\boldsymbol{C}_{[j],i}$$

$$\prod_{j=1}^{M} N(\boldsymbol{w}_j;\boldsymbol{H}_j\boldsymbol{w}_{[j]}, \boldsymbol{R}_j) = N(\boldsymbol{w};\boldsymbol{0},\widetilde{\boldsymbol{C}}_{\boldsymbol{\theta}}) \approx N(\boldsymbol{w};\boldsymbol{0},\boldsymbol{C}_{\boldsymbol{\theta}})$$

Extension to standalone stochastic process leads to **Meshed Gaussian Process**

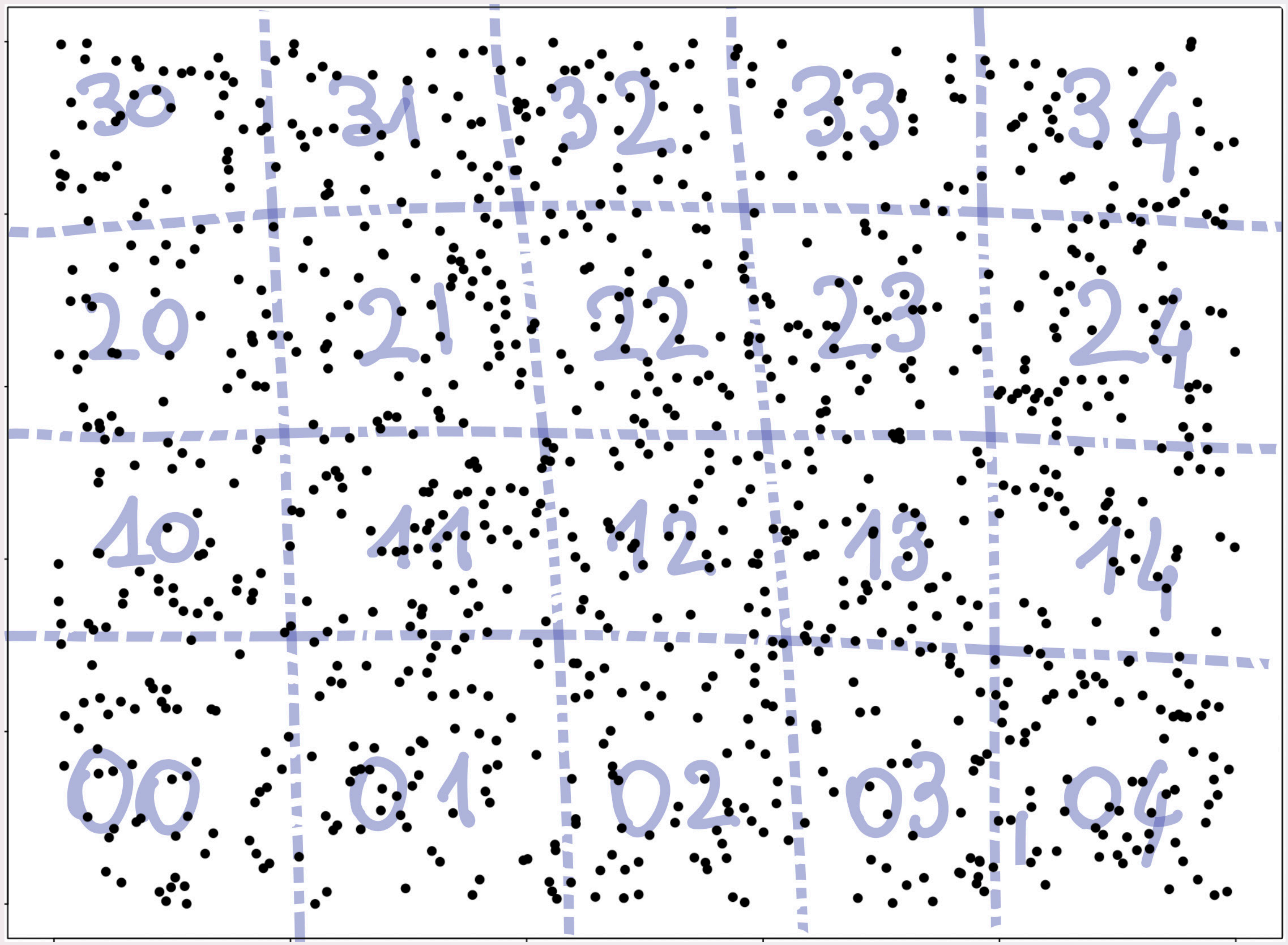
- Scalable replacement of GP in Bayesian hierarchical model
- Process based estimation and predictions at new locations
- Exact posterior sampling methods for Meshed GPs via Gibbs samplers
- Alternatively, Meshed GPs are interpretable as an approximation of $GP(\boldsymbol{0}, \boldsymbol{C}_{\boldsymbol{\theta}})$

QMGP: axis-parallel tessellation and 2-parent DAG



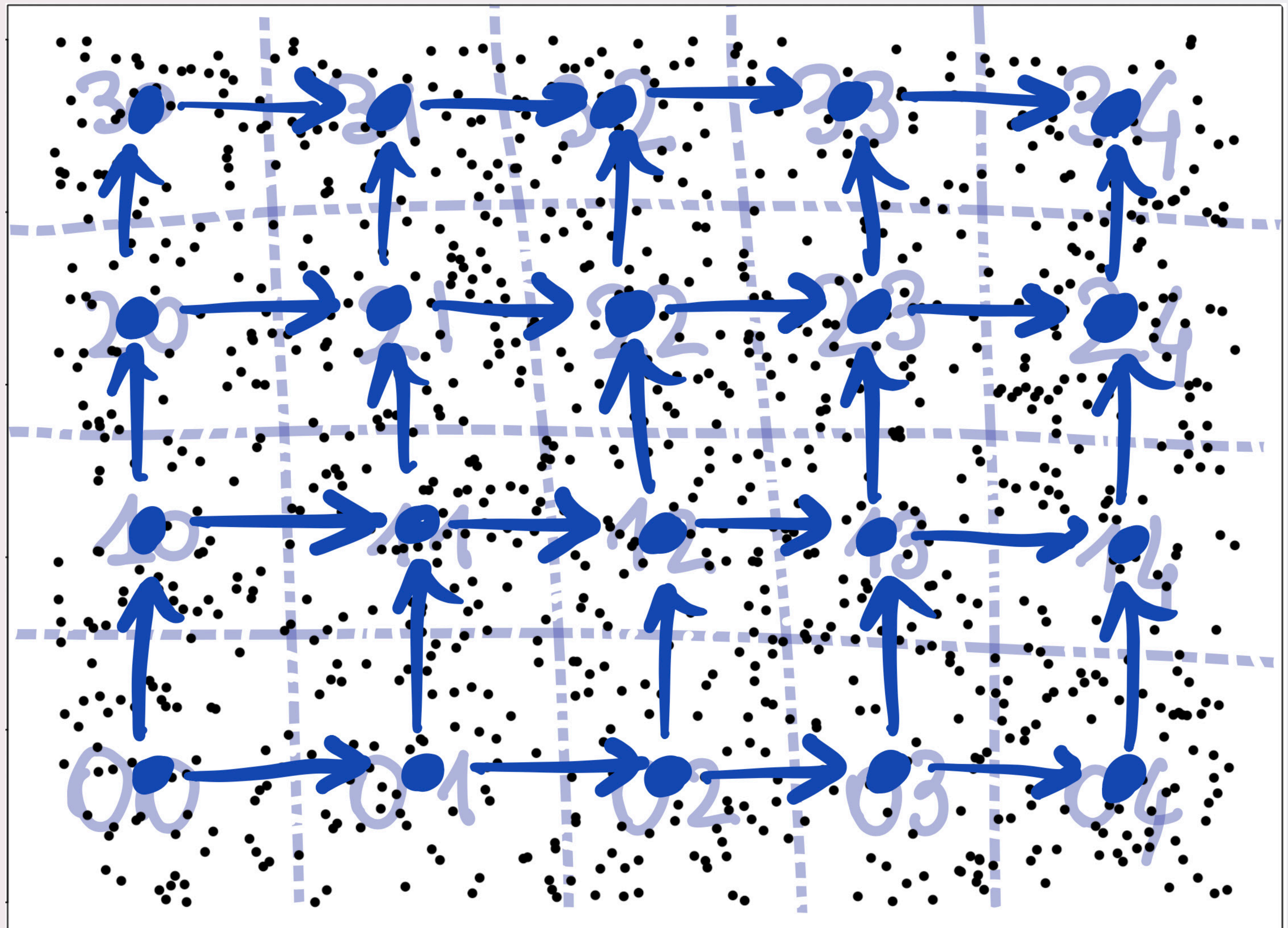
Data locations

QMGP: axis-parallel tessellation and 2-parent DAG



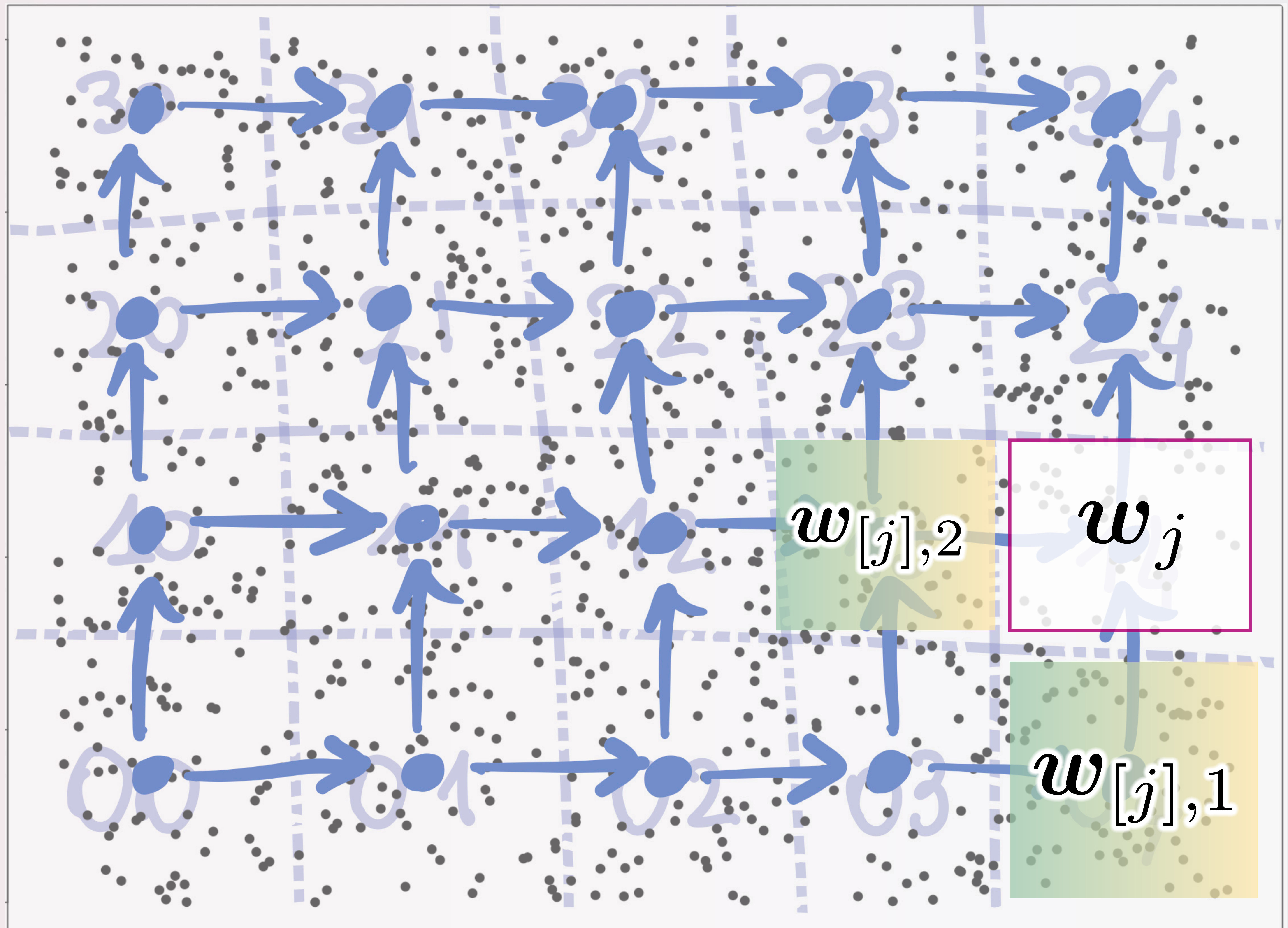
Domain partitioning

QMGP: axis-parallel tessellation and 2-parent DAG



DAG links domain partitions

QMGP: axis-parallel tessellation and 2-parent DAG



DAG links domain partitions

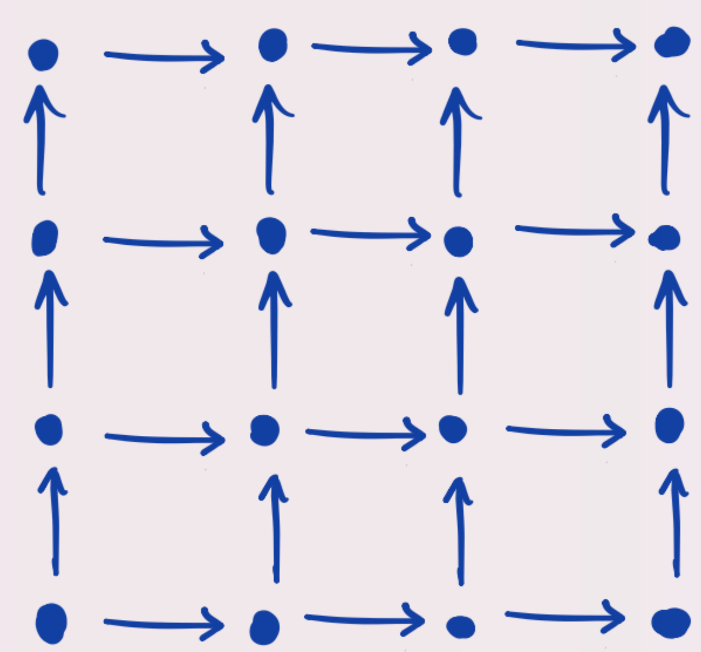
QMGP: axis-parallel tessellation and 2-parent DAG

M. Peruzzi, S. Banerjee & A.O. Finley (2022)
Highly Scalable Bayesian Geostatistical Modeling via Meshed Gaussian Processes on Partitoned Domains.
Journal of the American Statistical Association *117(538): 969-982.*
https://www.tandfonline.com/doi/full/10.1080/01621459.2020.1833889

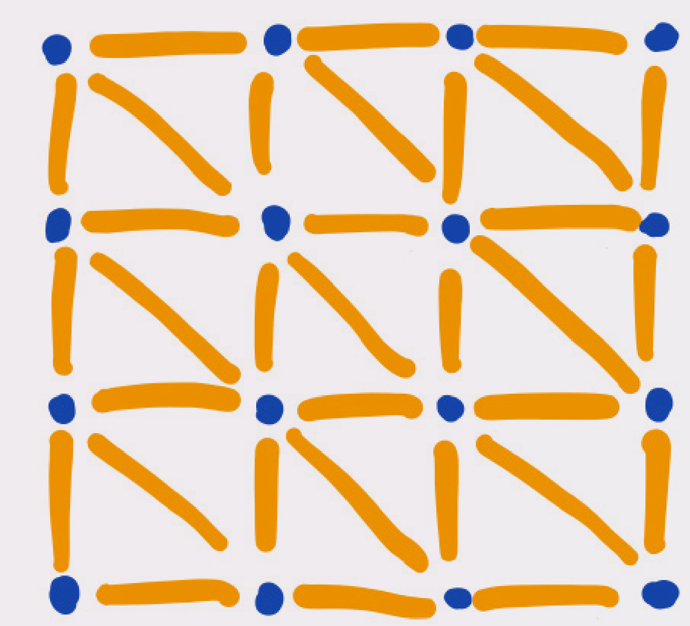
$$\prod_{j=1}^{M} N(\boldsymbol{w}_j; \boldsymbol{H}_j \boldsymbol{w}_{[j]}, \boldsymbol{R}_j) = N(\boldsymbol{w}; \boldsymbol{0}, \widetilde{\boldsymbol{C}}_{\boldsymbol{\theta}})$$

$$\boldsymbol{H}_j = \boldsymbol{C}_{j,[j]} \boldsymbol{C}_{[j]}^{-1}$$
$$\boldsymbol{R}_j = \boldsymbol{C}_{j,j} - \boldsymbol{H}_j \boldsymbol{C}_{[j],j}$$

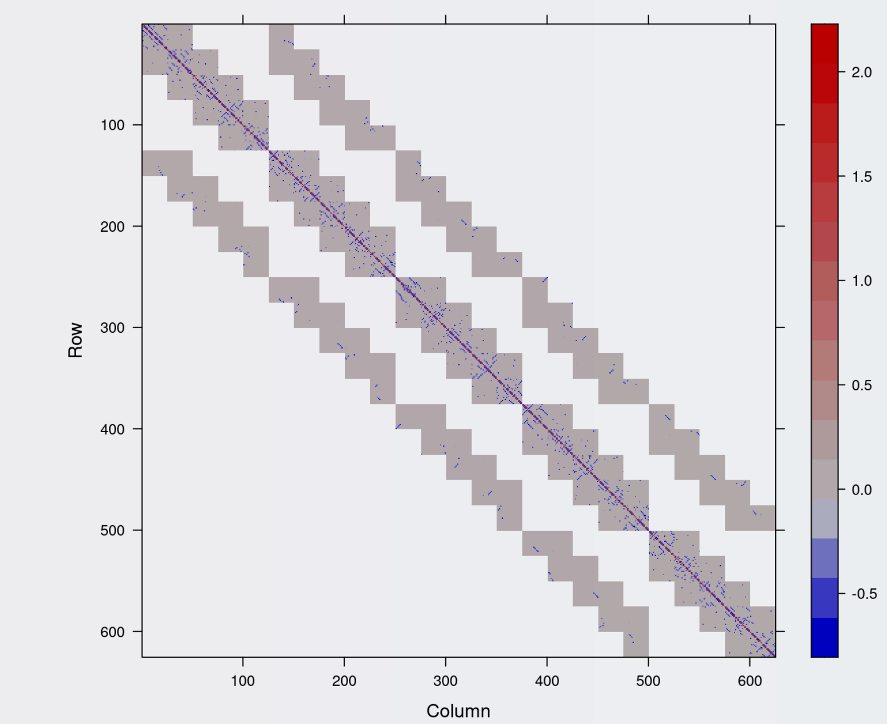
Sparse pattern DAG (mesh)



Undirected moral graph

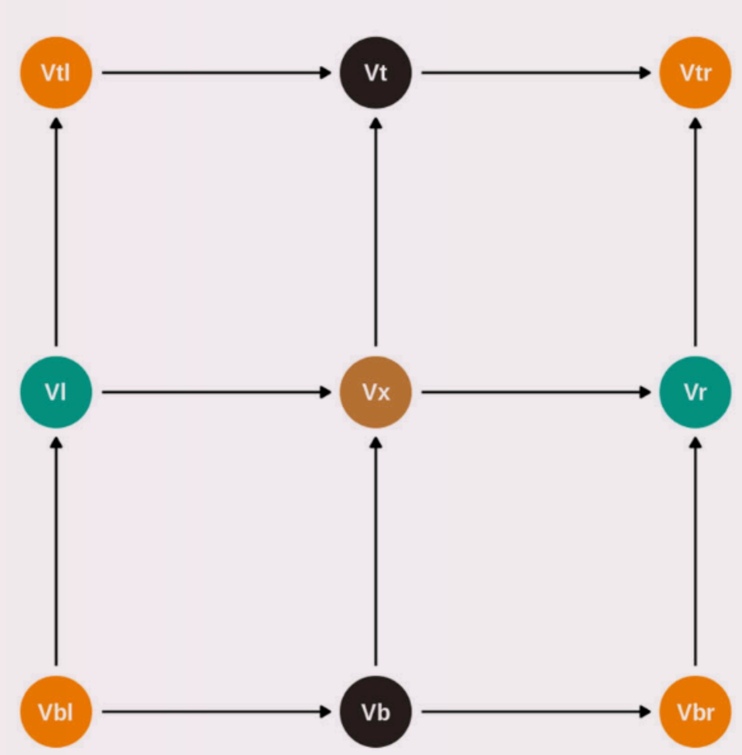


Sparsity pattern of $\widetilde{\boldsymbol{C}}_{\boldsymbol{\theta}}^{-1}$



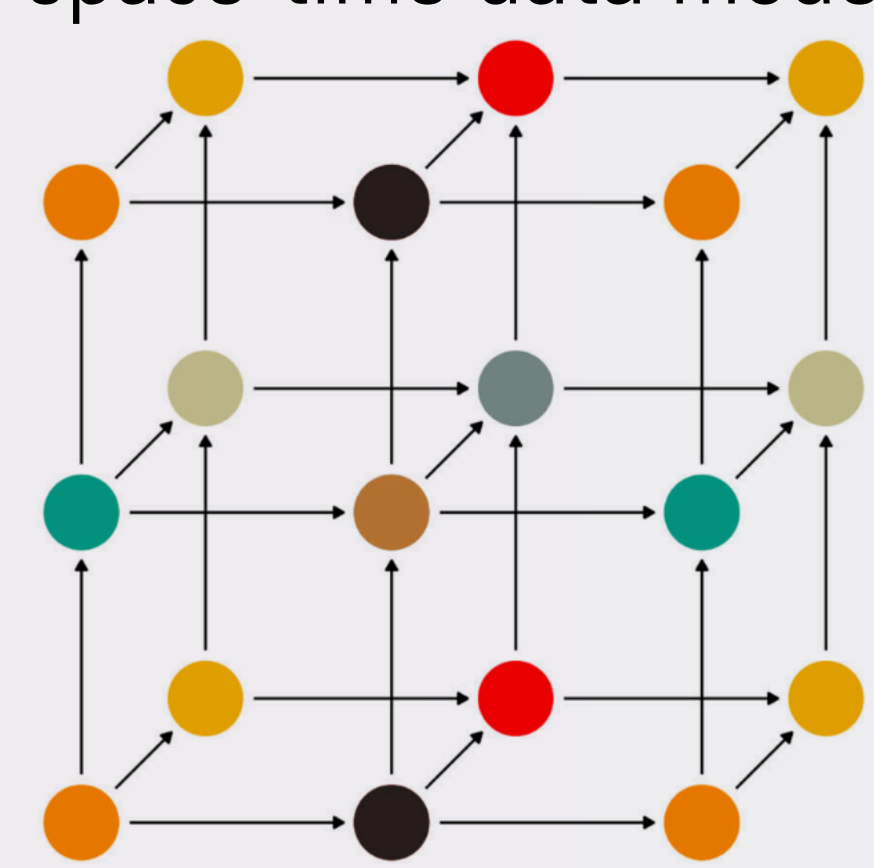
Graph coloring

parallel block Gibbs sampling



3D extension

space-time data models



QMGP: Bayesian hierarchical model & Gibbs sampler

M. Peruzzi, S. Banerjee & A.O. Finley (2022)
Highly Scalable Bayesian Geostatistical Modeling via Meshed Gaussian Processes on Partitoned Domains.
Journal of the American Statistical Association *117(538): 969-982.*
https://www.tandfonline.com/doi/full/10.1080/01621459.2020.1833889

Probability model for outcome *j* $$y_j(\boldsymbol{\ell}) \mid \eta_j(\boldsymbol{\ell}), \tau_j \sim P_j(\eta_j(\boldsymbol{\ell}), \tau_j)$$

$$\boldsymbol{\ell} \in \mathcal{D} \subset \Re^d \quad j = 1, \dots, q$$

Linear predictor $$\eta_j(\boldsymbol{\ell}) = \boldsymbol{x}_j(\boldsymbol{\ell})^\top \boldsymbol{\beta}_j + w_j(\boldsymbol{\ell})$$

Prior for spatial random effects

$$\boldsymbol{w}(\cdot) \sim \text{meshed}GP(\mathbf{0}, \boldsymbol{C}_{\boldsymbol{\theta}}, \mathcal{G}, \boldsymbol{T})$$

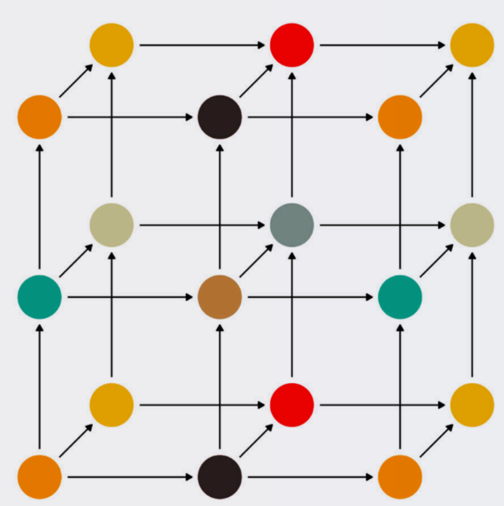
Gibbs sampler

Cycle through these steps:

- sample $\boldsymbol{w}_j \mid \boldsymbol{w}_{-j}, y, \boldsymbol{\theta}, \boldsymbol{\beta}, \boldsymbol{\tau}$
- update $\boldsymbol{\theta} \mid \boldsymbol{w}$ (MH accept/reject)
- sample $\boldsymbol{\beta} \mid \boldsymbol{w}, \boldsymbol{y}, \boldsymbol{\tau}$
- sample $\boldsymbol{\tau} \mid \boldsymbol{y}, \boldsymbol{\theta}, \boldsymbol{w}$

Cubic Meshed GP:
computationally **much cheaper**

Take advantage of graph coloring for **parallel sampling**



- Spatial DAG extends Bayesian model DAG
- Posterior computations via MCMC proceed straightforwardly

  (valid spatial process = no need to verify detailed balance conditions)
- Scalable to large data sets thanks to known properties of chosen DAG

Meshed Gaussian processes

M. Peruzzi, S. Banerjee & A.O. Finley (2022)
Highly Scalable Bayesian Geostatistical Modeling via Meshed Gaussian Processes on Partitoned Domains.
Journal of the American Statistical Association *117(538): 969-982.*
https://www.tandfonline.com/doi/full/10.1080/01621459.2020.1833889

Probability model for outcome *j* $y_j(\boldsymbol{\ell}) \mid \eta_j(\boldsymbol{\ell}), \tau_j \sim P_j(\eta_j(\boldsymbol{\ell}), \tau_j)$

$\boldsymbol{\ell} \in \mathcal{D} \subset \Re^d \quad j = 1, \dots, q$

Linear predictor $\eta_j(\boldsymbol{\ell}) = \boldsymbol{x}_j(\boldsymbol{\ell})^\top \boldsymbol{\beta}_j + w_j(\boldsymbol{\ell})$

Prior for spatial random effects

$\boldsymbol{w}(\cdot) \sim \text{meshed}GP(\mathbf{0}, \boldsymbol{C}_{\boldsymbol{\theta}}, \mathcal{G}, \boldsymbol{T})$

Gibbs sampler

Cycle through these steps:

- sample $\boldsymbol{w}_j \mid \boldsymbol{w}_{-j}, y, \boldsymbol{\theta}, \boldsymbol{\beta}, \boldsymbol{\tau}$
- update $\boldsymbol{\theta} \mid \boldsymbol{w}$ (MH accept/reject)
- sample $\boldsymbol{\beta} \mid \boldsymbol{w}, \boldsymbol{y}, \boldsymbol{\tau}$
- sample $\boldsymbol{\tau} \mid \boldsymbol{y}, \boldsymbol{\theta}, \boldsymbol{w}$

Cubic Meshed GP:
computationally **much cheaper**

$$p(\boldsymbol{\theta} \mid \boldsymbol{w}) \propto \prod_{j=1}^{M} p(\boldsymbol{w}_j \mid \boldsymbol{w}_{[j]}, \boldsymbol{\theta}) p(\boldsymbol{\theta})$$

and

$$p(\boldsymbol{w}_j \mid \boldsymbol{w}_{[j]}, \boldsymbol{\theta}) = N(\boldsymbol{w}_j; \boldsymbol{H}_j \boldsymbol{w}_{[j]}, \boldsymbol{R}_j)$$

where

$$\boldsymbol{H}_j = \boldsymbol{C}_{j,[j]} \boldsymbol{C}_{[j]}^{-1}$$

$$\boldsymbol{R}_j = \boldsymbol{C}_{j,j} - \boldsymbol{H}_j \boldsymbol{C}_{[j],j}$$

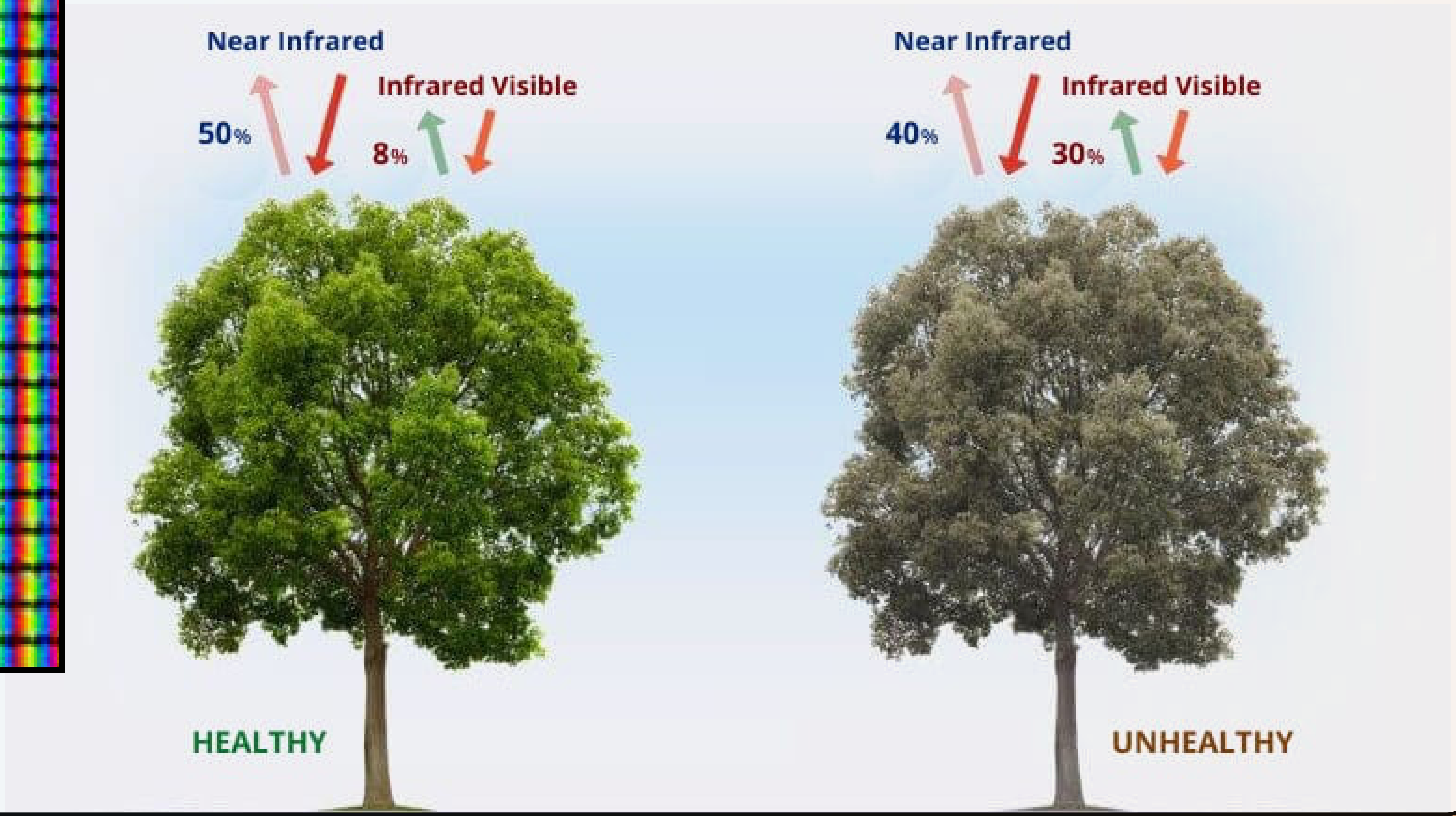
and $\boldsymbol{C}_{[j]}^{-1}$ is ***small for all j:***

$$O(nJ^3m^2) \ll O(n^3)$$

Application: NDVI imagery

M. Peruzzi, S. Banerjee & A.O. Finley (2022)
Highly Scalable Bayesian Geostatistical Modeling via Meshed Gaussian Processes on Partitoned Domains.
Journal of the American Statistical Association *117(538): 969-982.*
https://www.tandfonline.com/doi/full/10.1080/01621459.2020.1833889

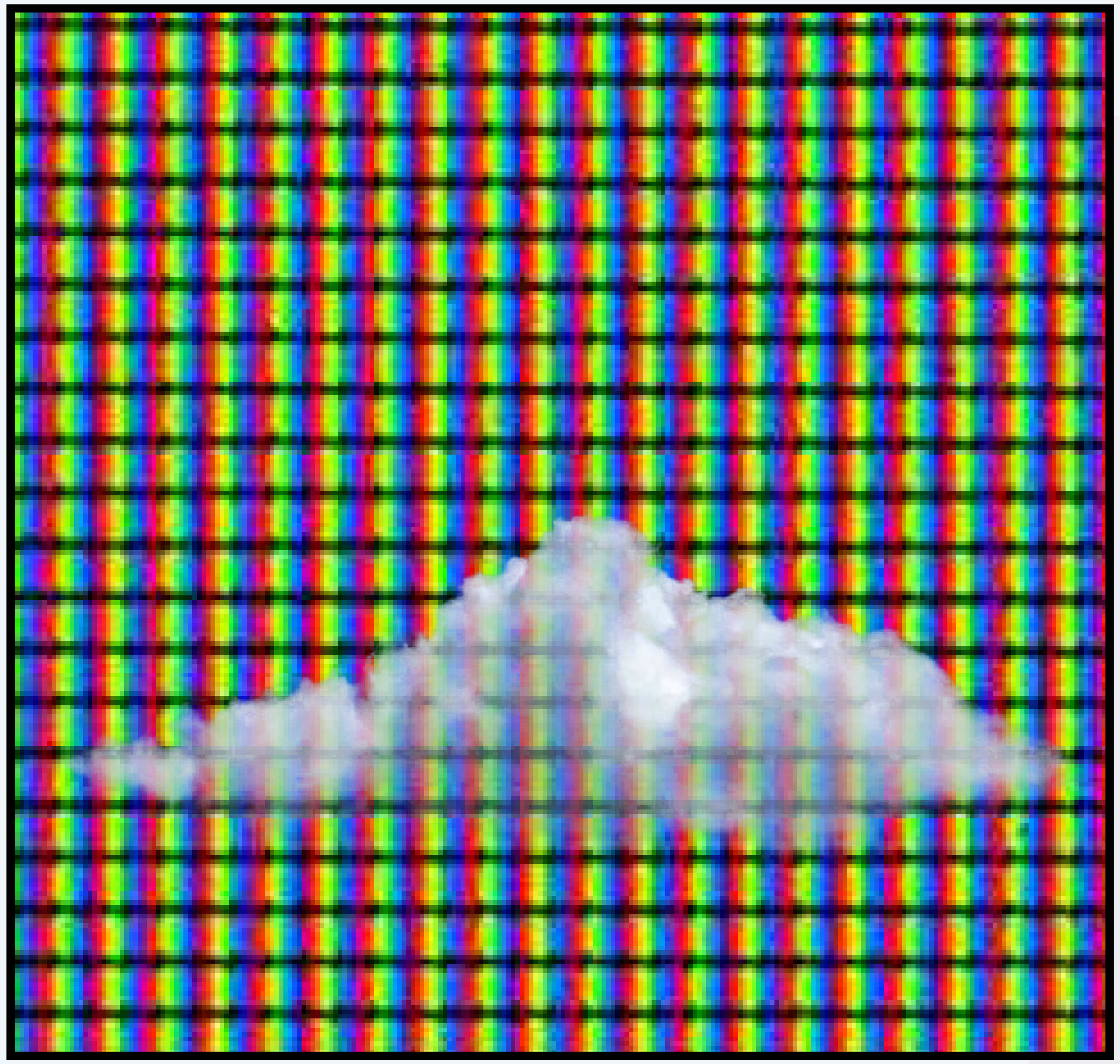
- Satellite images store different bands of the electromagnetic spectrum
- Using these bands, Normalized Difference Vegetation Index can be calculated
- Results in a NDVI “image” at regular grid of pixels

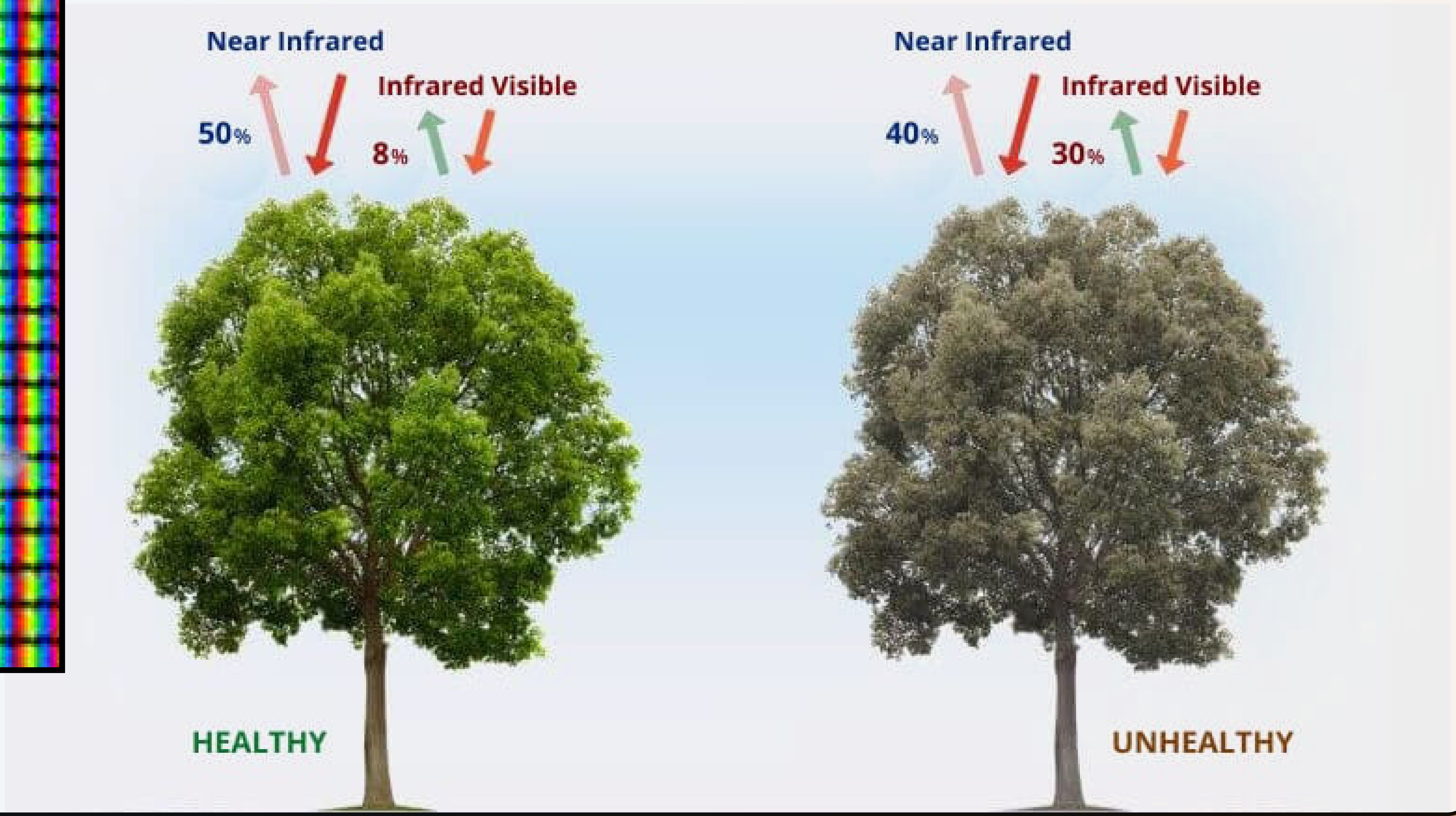


Application: NDVI imagery

M. Peruzzi, S. Banerjee & A.O. Finley (2022)
Highly Scalable Bayesian Geostatistical Modeling via Meshed Gaussian Processes on Partitoned Domains.
Journal of the American Statistical Association *117(538): 969-982.*
https://www.tandfonline.com/doi/full/10.1080/01621459.2020.1833889

- Satellite images store different bands of the electromagnetic spectrum
- Using these bands, Normalized Difference Vegetation Index can be calculated
- Results in a NDVI “image” at regular grid of pixels
- Clouds obstruct view and create gaps

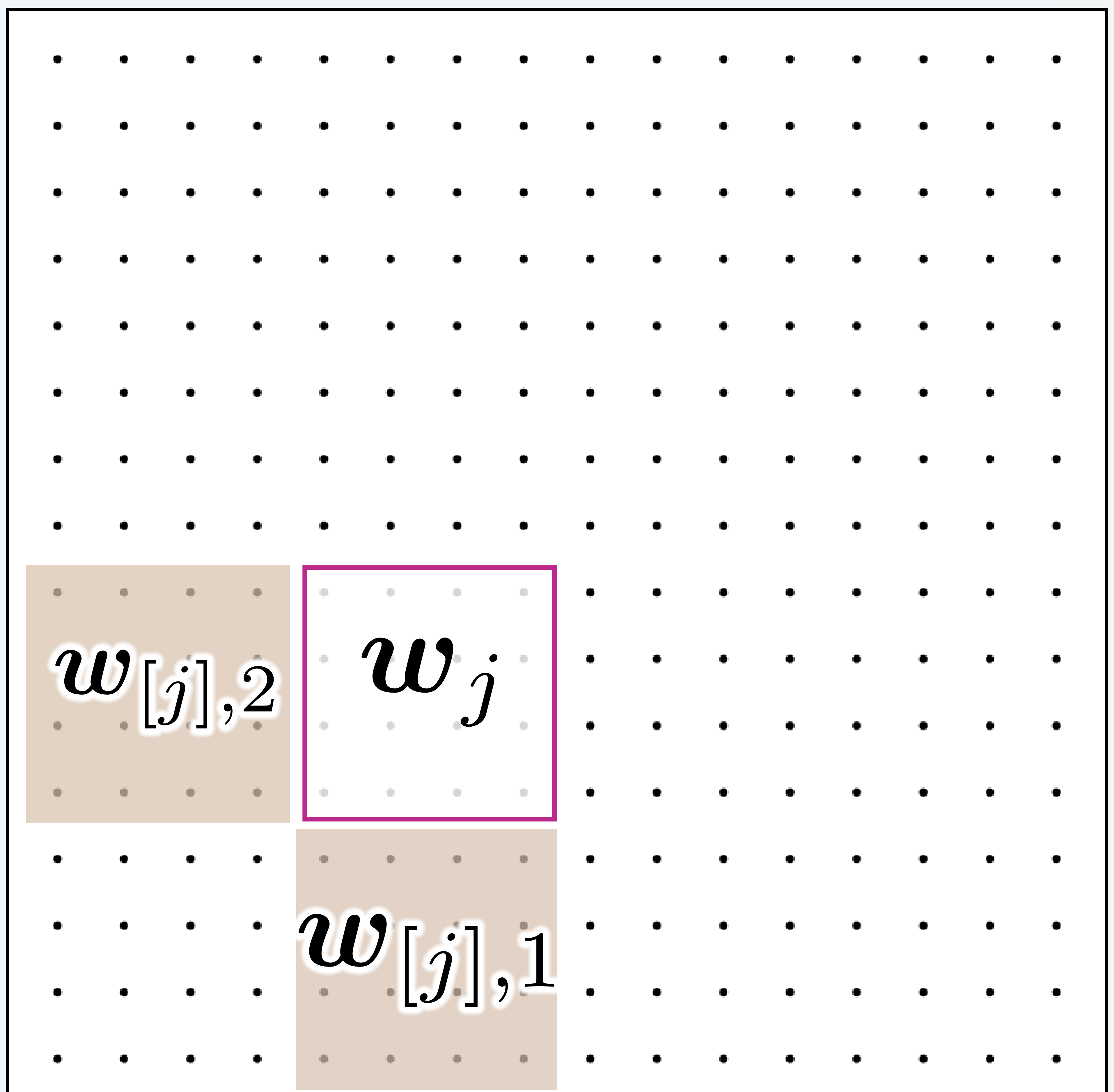




Application: NDVI imagery – why use QMGP?

M. Peruzzi, S. Banerjee & A.O. Finley (2022)
Highly Scalable Bayesian Geostatistical Modeling via Meshed Gaussian Processes on Partitoned Domains.
Journal of the American Statistical Association *117(538): 969-982.*
https://www.tandfonline.com/doi/full/10.1080/01621459.2020.1833889

- Satellite images store different bands of the electromagnetic spectrum
- Using these bands, Normalized Difference Vegetation Index can be calculated
- Results in a NDVI "image" at regular grid of pixels
- Clouds obstruct view and create gaps



$$p(\boldsymbol{\theta} \mid \boldsymbol{w}) \propto \prod_{j=1}^{M} p(\boldsymbol{w}_j \mid \boldsymbol{w}_{[j]}, \boldsymbol{\theta}) p(\boldsymbol{\theta})$$

$$p(\boldsymbol{w}_j \mid \boldsymbol{w}_{[j]}, \boldsymbol{\theta}) = N(\boldsymbol{w}_j; \boldsymbol{H}_j \boldsymbol{w}_{[j]}, \boldsymbol{R}_j)$$

where

$$\boldsymbol{H}_j = \boldsymbol{C}_{j,[j]} \boldsymbol{C}_{[j]}^{-1}$$

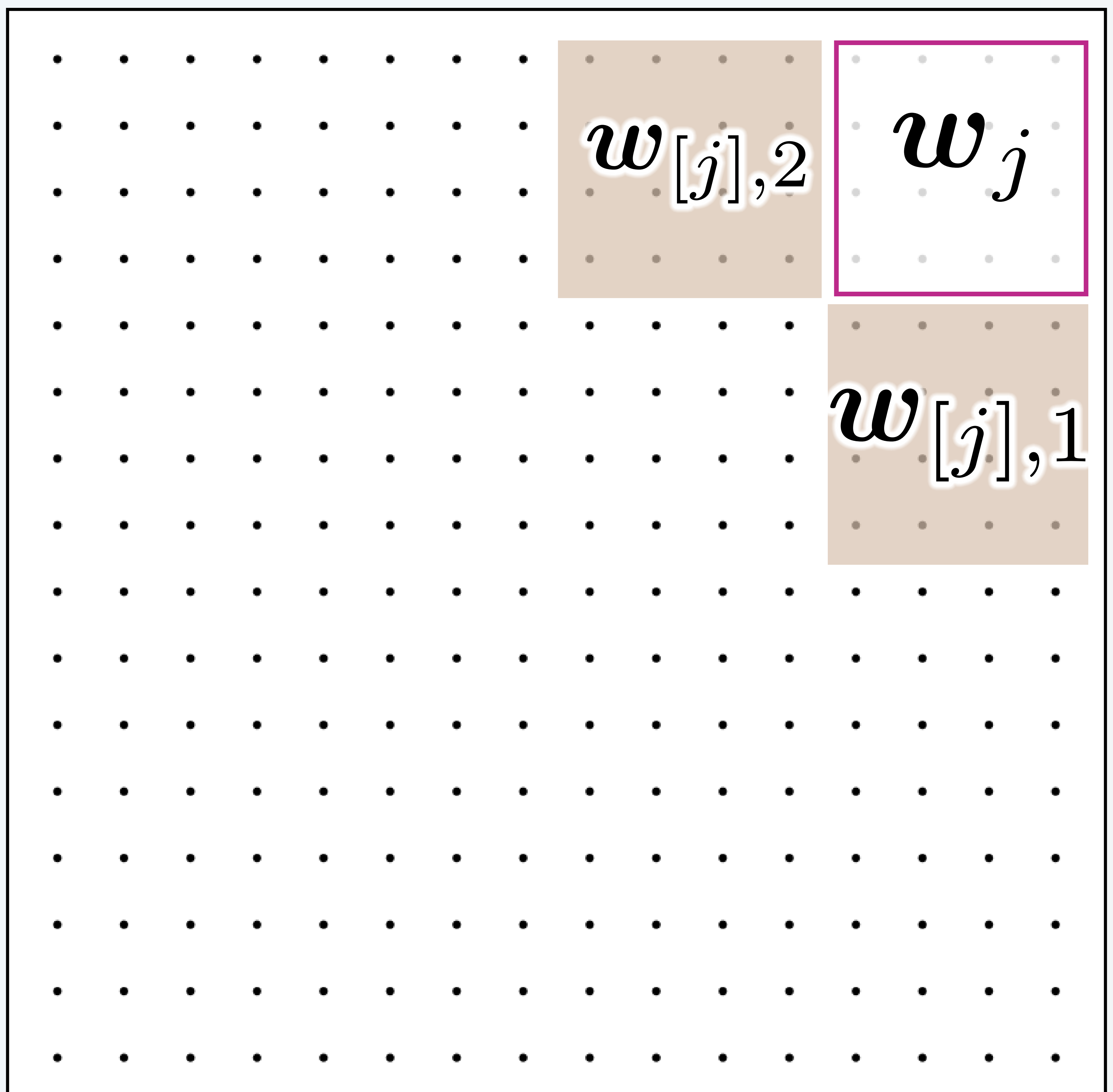
$$\boldsymbol{R}_j = \boldsymbol{C}_{j,j} - \boldsymbol{H}_j \boldsymbol{C}_{[j],j}$$

- These matrices only depend on relative distance between locations

Application: NDVI imagery – why use QMGP?

M. Peruzzi, S. Banerjee & A.O. Finley (2022)
Highly Scalable Bayesian Geostatistical Modeling via Meshed Gaussian Processes on Partitoned Domains.
Journal of the American Statistical Association *117(538): 969-982.*
https://www.tandfonline.com/doi/full/10.1080/01621459.2020.1833889

- Satellite images store different bands of the electromagnetic spectrum
- Using these bands, Normalized Difference Vegetation Index can be calculated
- Results in a NDVI "image" at regular grid of pixels
- Clouds obstruct view and create gaps



$$p(\boldsymbol{\theta} \mid \boldsymbol{w}) \propto \prod_{j=1}^{M} p(\boldsymbol{w}_j \mid \boldsymbol{w}_{[j]}, \boldsymbol{\theta}) p(\boldsymbol{\theta})$$

$$p(\boldsymbol{w}_j \mid \boldsymbol{w}_{[j]}, \boldsymbol{\theta}) = N(\boldsymbol{w}_j; \boldsymbol{H}_j \boldsymbol{w}_{[j]}, \boldsymbol{R}_j)$$

where

$$\boldsymbol{H}_j = \boldsymbol{C}_{j,[j]} \boldsymbol{C}_{[j]}^{-1}$$

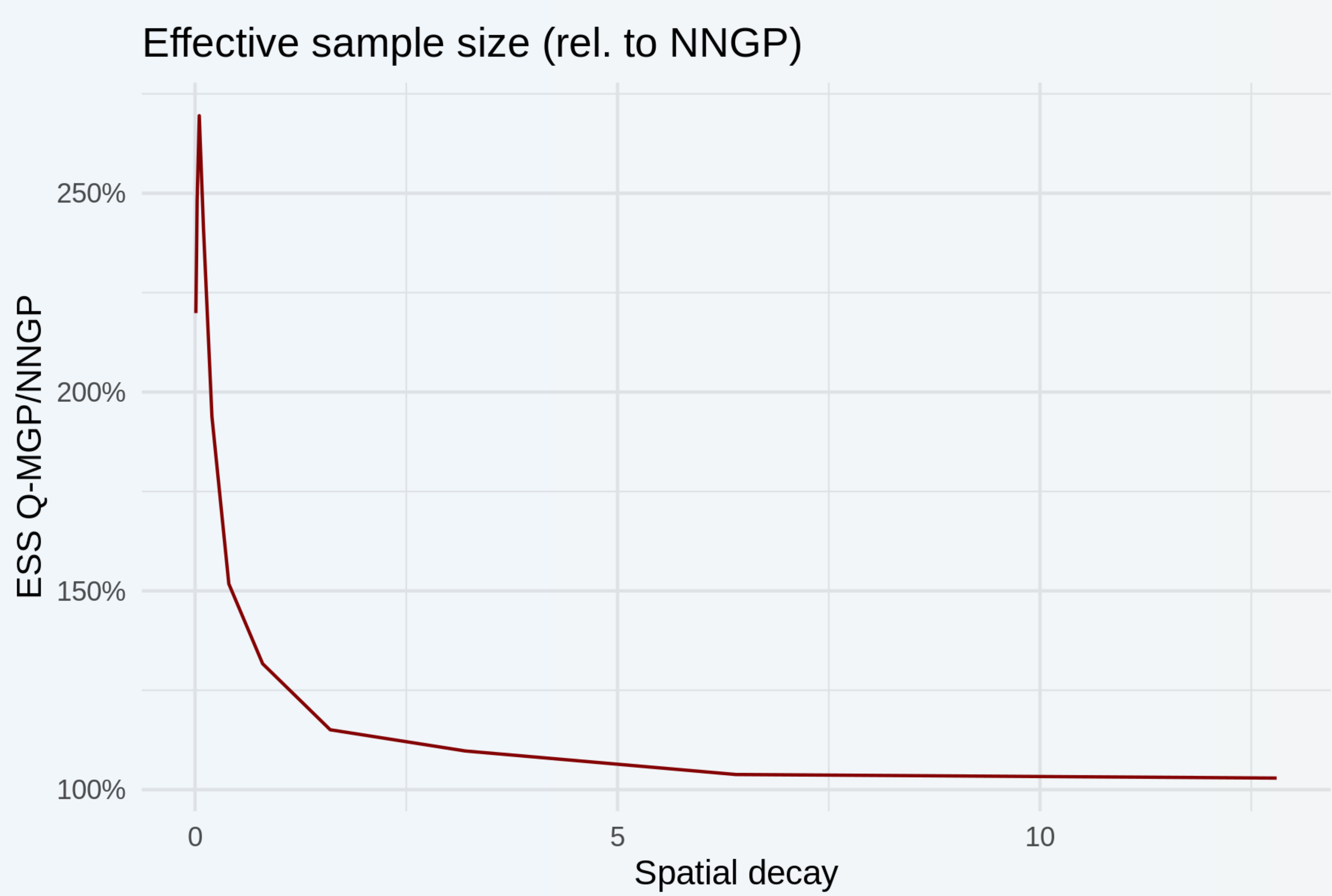
$$\boldsymbol{R}_j = \boldsymbol{C}_{j,j} - \boldsymbol{H}_j \boldsymbol{C}_{[j],j}$$

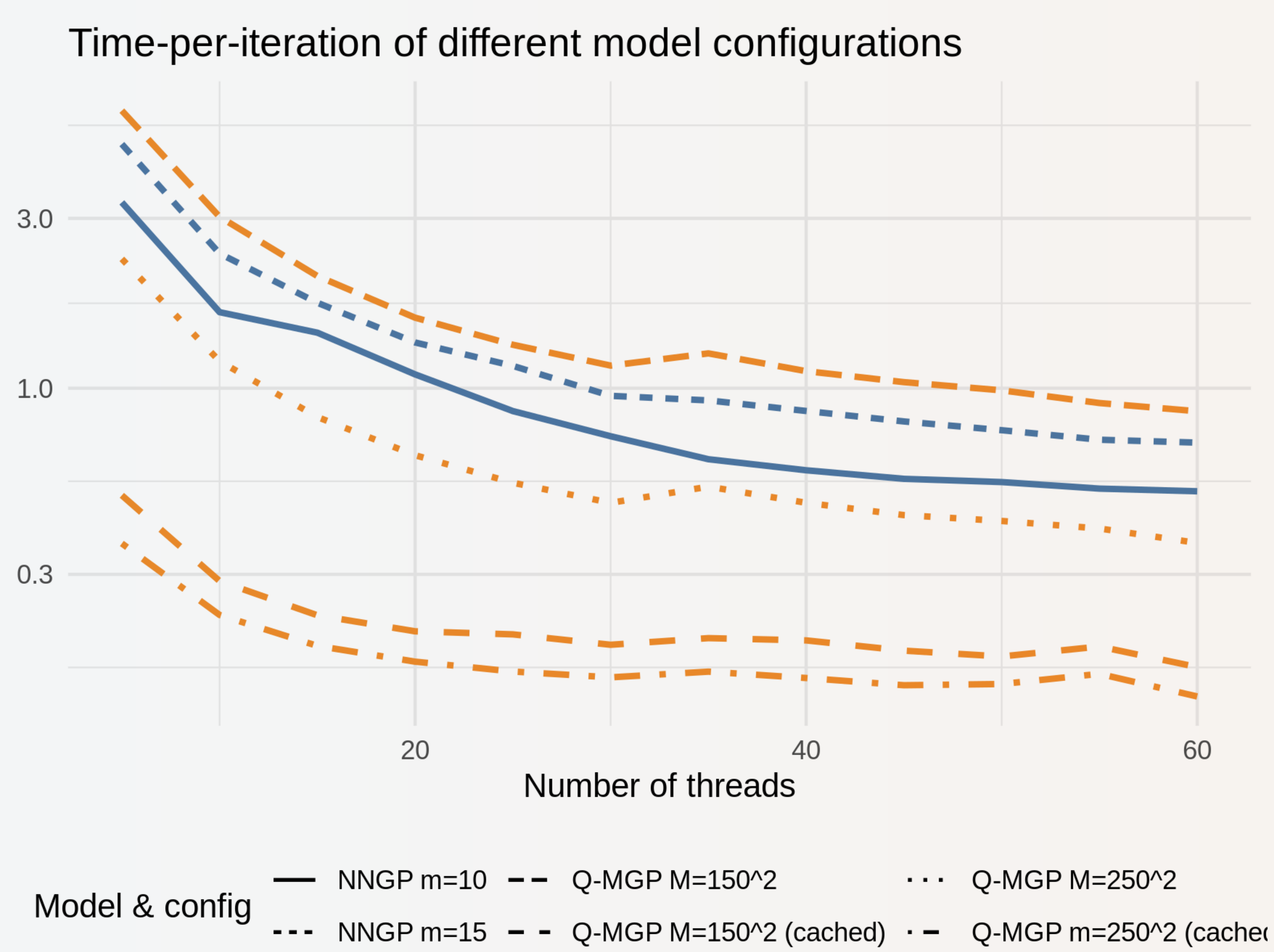
- These matrices only depend on relative distance between locations
- We only need to calculate ***M**** of them, where ***M* = O(1)***
- **Density evaluation cost down to** $O(nJm)$ **from** $O(nJ^3m^2)$
- Cost dominated by sampling: $O(nm^2)$

Cubic MGPs compared to Nearest-neighbor GPs

M. Peruzzi, S. Banerjee & A.O. Finley (2022)
Highly Scalable Bayesian Geostatistical Modeling via Meshed Gaussian Processes on Partitoned Domains.
Journal of the American Statistical Association *117(538): 969-982.*
https://www.tandfonline.com/doi/full/10.1080/01621459.2020.1833889

- Favorable comparison with state-of-the-art alternatives
- Up to **10x faster** (wall clock time)
- Up to **2.5x more efficient** Markov-chain Monte Carlo
- Total **25x** improvement
- Improvements are due to the ability to design purpose-made DAGs



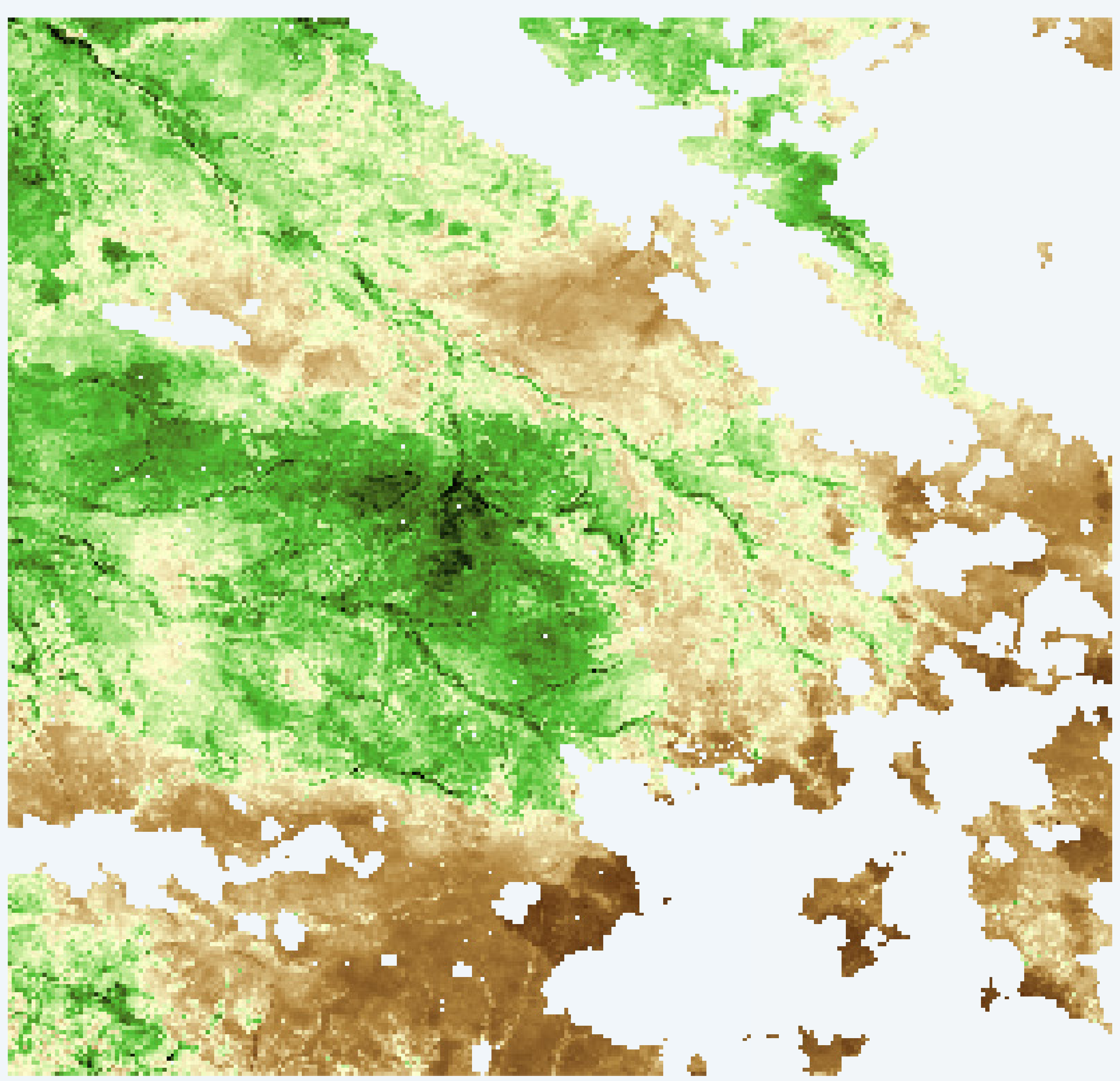


Application: Serengeti NDVI probabilistic gap filling

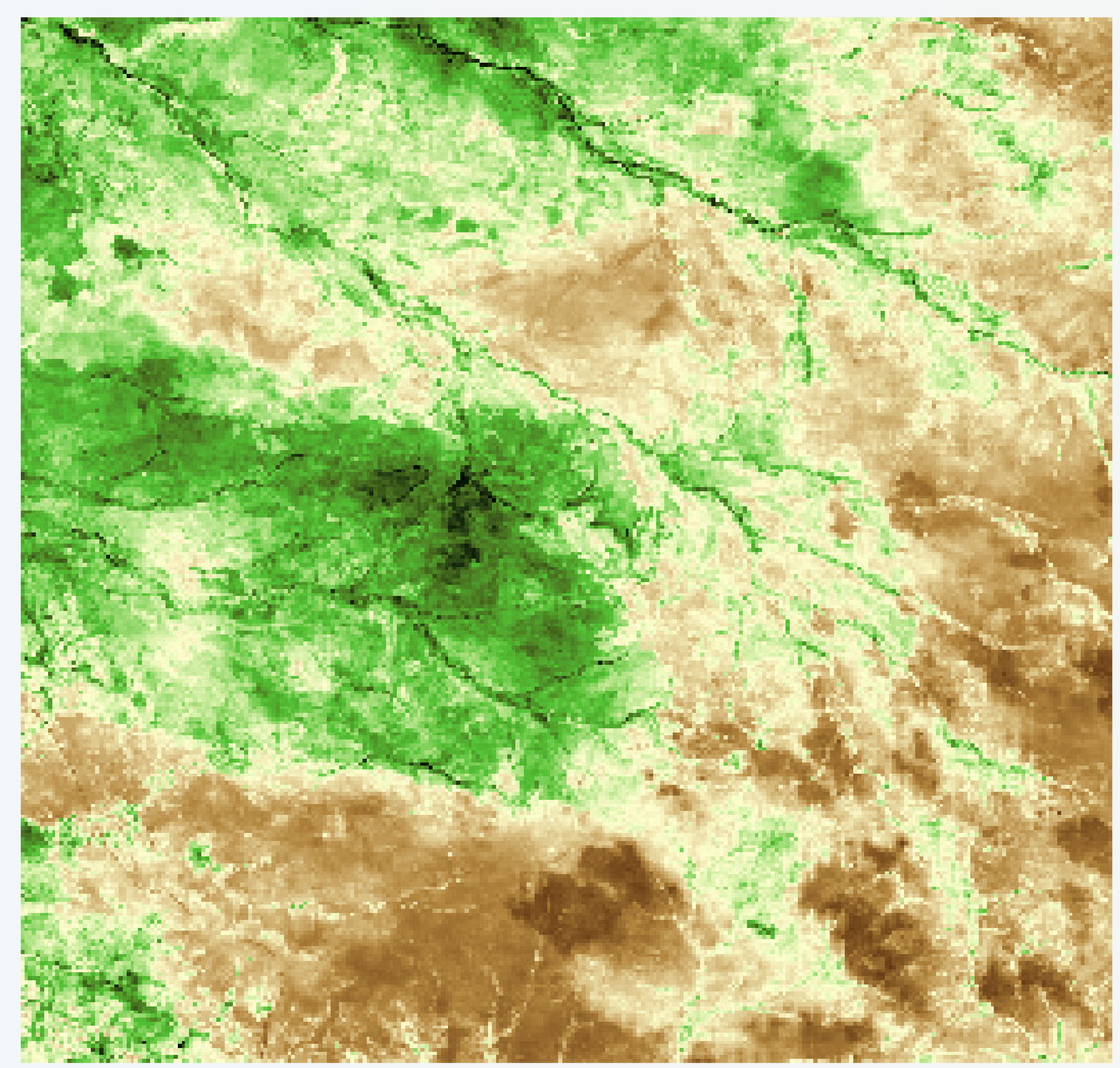
M. Peruzzi, S. Banerjee & A.O. Finley (2022)
Highly Scalable Bayesian Geostatistical Modeling via Meshed Gaussian Processes on Partitioned Domains.
Journal of the American Statistical Association *117(538): 969-982.*
https://www.tandfonline.com/doi/full/10.1080/01621459.2020.1833889

- Normalized Difference Vegetation Index (NDVI) measured by LANDSAT over **Serengeti park area**
- **16 million locations** in space & time
- Frequent **cloud cover** obfuscates images
- Bayesian model for probabilistic recovery of data behind clouds
- Model univariate NDVI outcome using spatiotemporally varying coefficient regression on 2 regressors (intercept, elevation)
- Posterior sampling for a **bivariate latent MGP**
- Results in less than 2 days (this is fast)

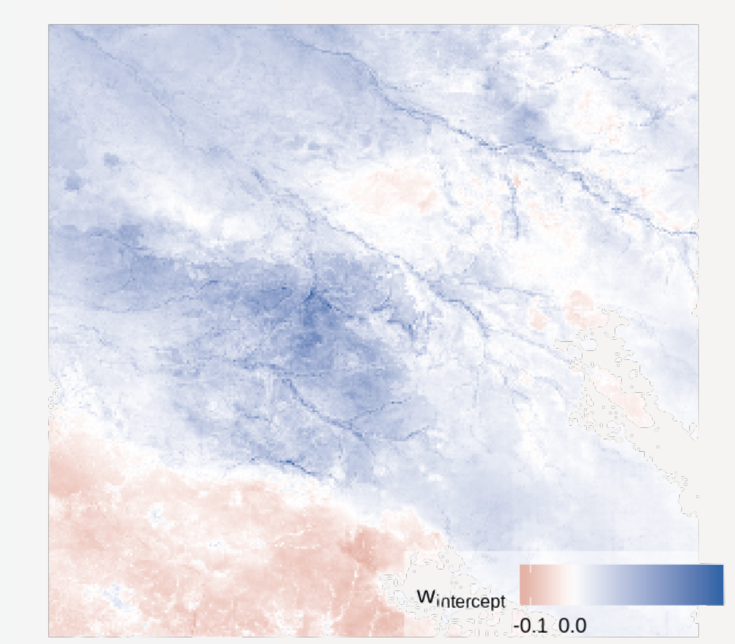
Observed NDVI



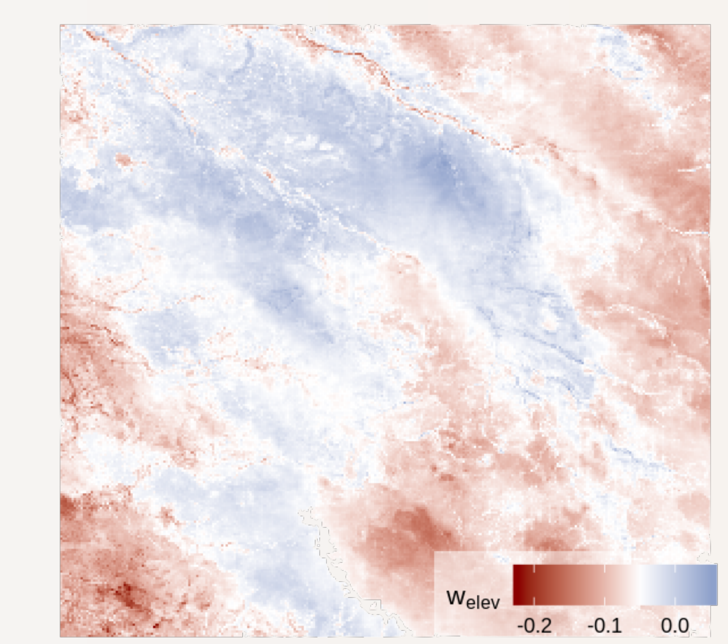
Predicted NDVI



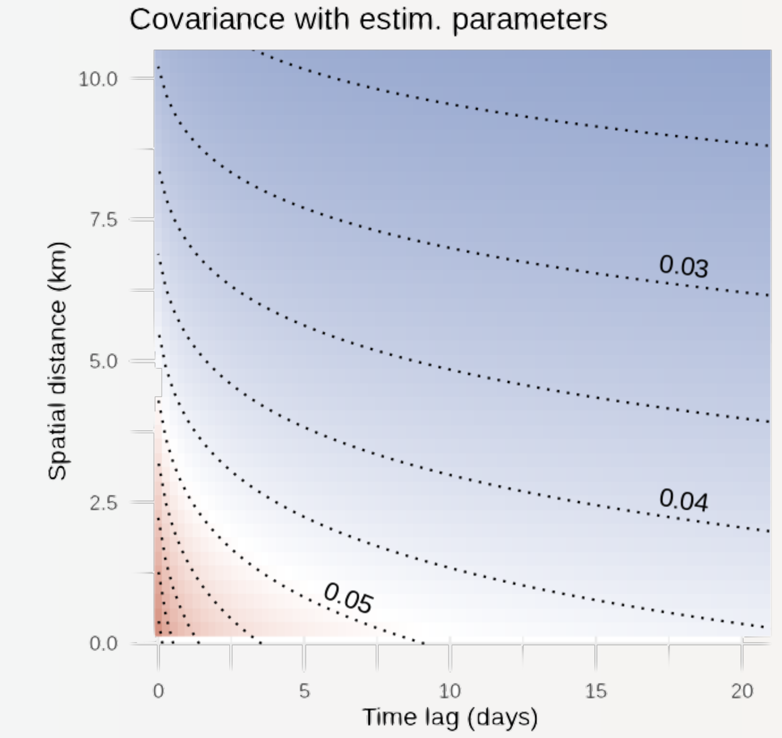
Residual spatial effect



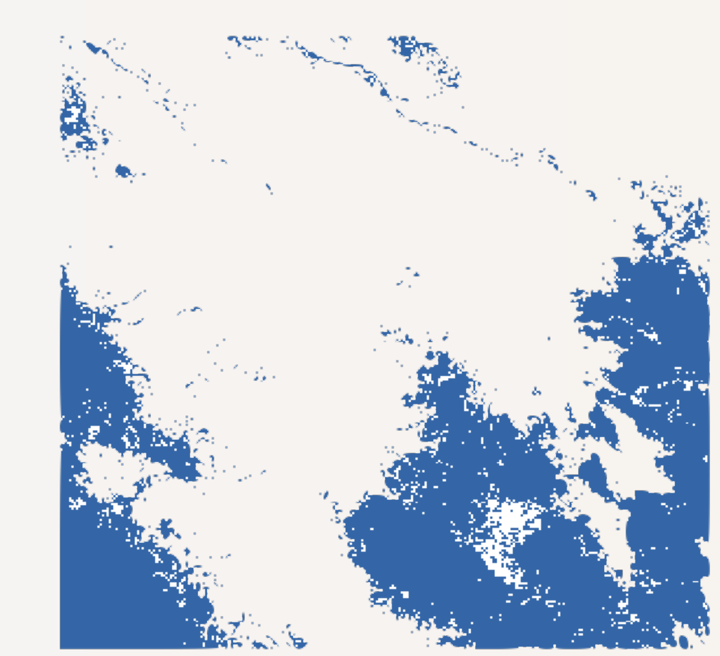
Elevation: effect on NDVI



Covariance with estim. parameters

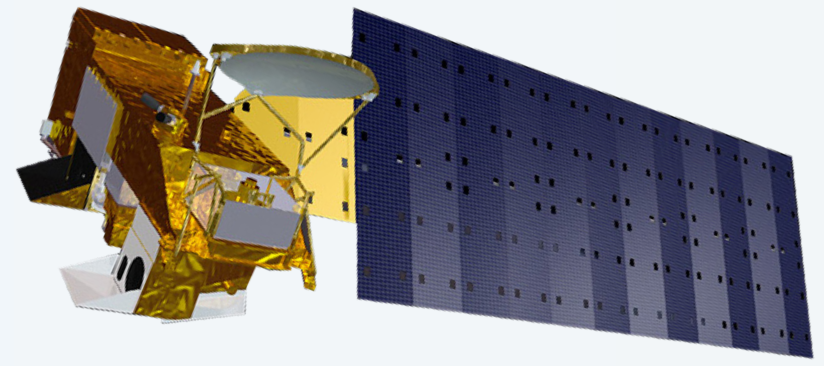


Elevation: nonzero effect locations (95% cred. int.)

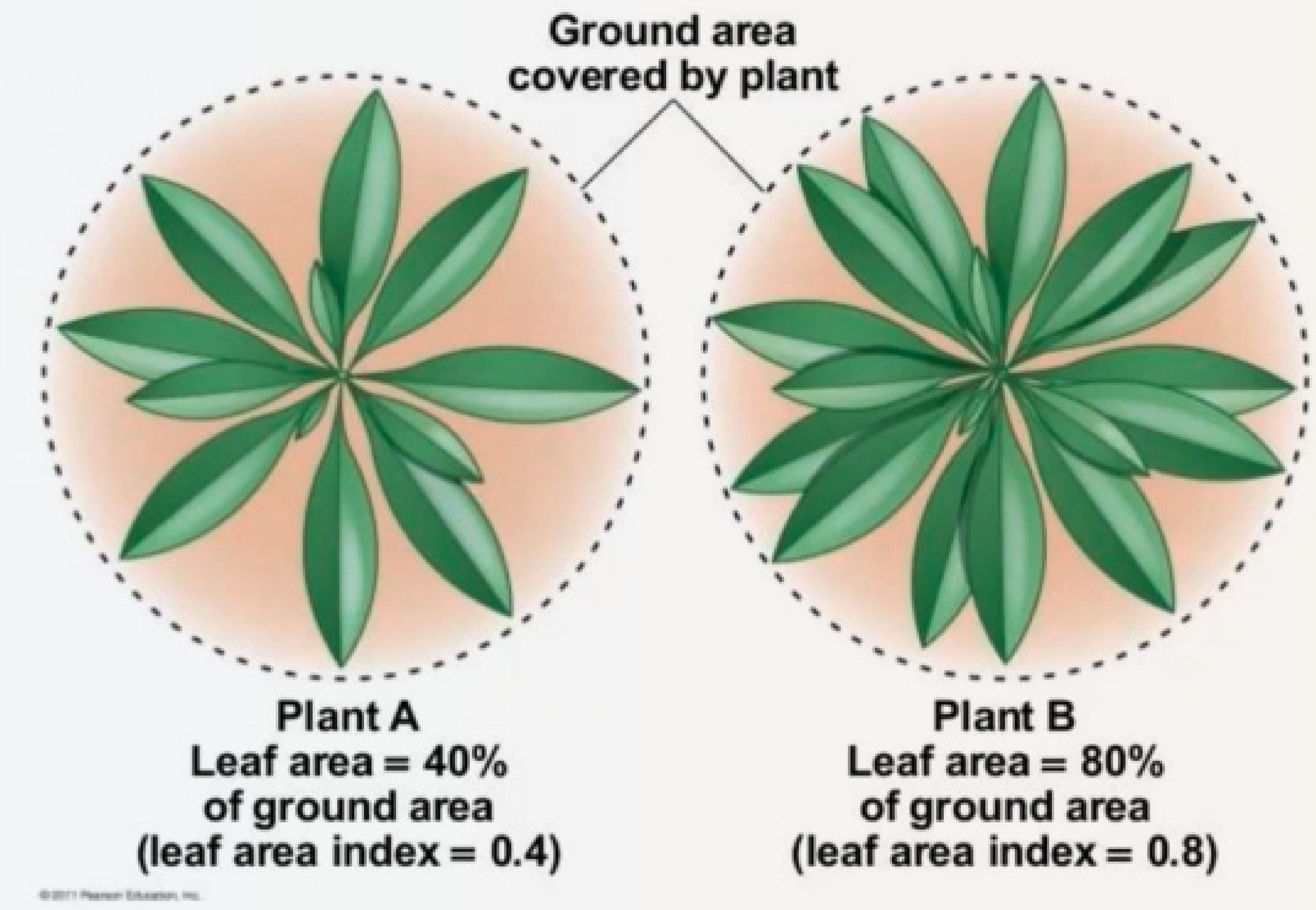


Application: MODIS Alpine Snow/Tree cover data

MODIS satellite data



- **Snow cover**: number of days with snow within 8-day period
- **Leaf area index**: leaf surface relative to ground surface (an integer)



High performance methods for non-Gaussian data

- In many cases, Gaussian assumption is inappropriate
- Latent Gaussian process models still useful with non-Gaussian first stage
- Use-case: **multivariate multi-type data** using spatial factor model

$$y_j(\boldsymbol{\ell}) \sim P_j(\eta_j(\boldsymbol{\ell}), \tau_j) \qquad j = 1, \dots, q$$

$$\boldsymbol{\eta} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\Lambda}\boldsymbol{w}$$

$$\boldsymbol{w}(\cdot) \sim \text{meshed}GP(\mathbf{0}, \boldsymbol{C}_{\boldsymbol{\theta}})$$

Gibbs sampler

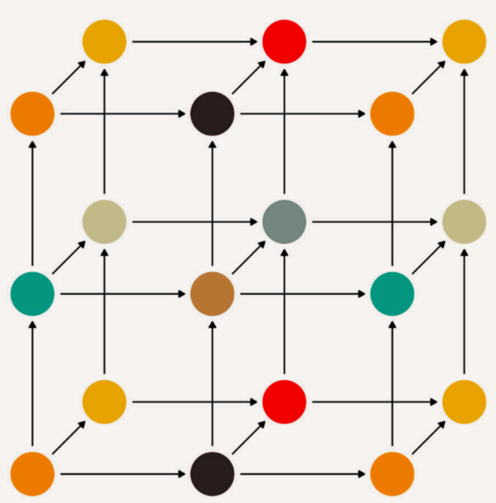
Cycle through these steps:

- sample $\boldsymbol{w}_j \mid \boldsymbol{w}_{-j}, y, \boldsymbol{\theta}, \boldsymbol{\Lambda}, \boldsymbol{\beta}, \boldsymbol{\tau}$
- sample $\boldsymbol{\theta} \mid \boldsymbol{w}$
- sample $\boldsymbol{\beta} \mid \boldsymbol{w}, \boldsymbol{y}, \boldsymbol{\tau}$
- sample $\boldsymbol{\tau} \mid \boldsymbol{y}, \boldsymbol{\theta}, \boldsymbol{w}$

Meshed GP: computationally cheap

Cubic Meshed GP: computationally **much cheaper**

Take advantage of graph coloring for **parallel sampling**



High performance methods for non-Gaussian data

- In many cases, Gaussian assumption is inappropriate
- Latent Gaussian process models still useful with non-Gaussian first stage
- Use-case: **multivariate multi-type data** using spatial factor model

$$y_j(\boldsymbol{\ell}) \sim P_j(\eta_j(\boldsymbol{\ell}), \tau_j) \qquad j = 1, \dots, q$$

$$\boldsymbol{\eta} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\Lambda}\boldsymbol{w}$$

$$\boldsymbol{w}(\cdot) \sim \text{meshed}GP(\boldsymbol{0}, \boldsymbol{C}_{\boldsymbol{\theta}})$$

Gibbs sampler

Cycle through these steps:

- sample $\boldsymbol{w}_j \mid \boldsymbol{w}_{-j}, y, \boldsymbol{\theta}, \boldsymbol{\Lambda}, \boldsymbol{\beta}, \boldsymbol{\tau}$

  Lack of conjugacy: How can we update $\boldsymbol{w}_j$ efficiently?

- sample $\boldsymbol{\theta} \mid \boldsymbol{w}$

  Meshed GP: computationally cheap

- sample $\boldsymbol{\beta} \mid \boldsymbol{w}, \boldsymbol{y}, \boldsymbol{\tau}$
- sample $\boldsymbol{\tau} \mid \boldsymbol{y}, \boldsymbol{\theta}, \boldsymbol{w}$

SiMPA: Simplified Manifold Preconditioner Adaptation

M. Peruzzi & D.B. Dunson (2022)
Spatial Meshing for General Bayesian Multivariate Models.
https://arxiv.org/abs/2201.10080

- Target update: preconditioned MALA (simple, fast, efficient)
- Adaptively & quickly build the preconditioner using **simplified Manifold MALA**

1 – Propose a new value for $\boldsymbol{w}_j$

$$\boldsymbol{w}_j^* \mid \boldsymbol{w}_{[j]}, \boldsymbol{w}_{[j\to i]}, \boldsymbol{\theta}, \boldsymbol{\beta}, \boldsymbol{y}, \boldsymbol{\tau} \sim N\left(\boldsymbol{w}_j + \frac{\epsilon^2}{2} \boldsymbol{M}^*_{(m)} \boldsymbol{g}, \epsilon^2 \boldsymbol{M}^*_{(m)}\right)$$

Where $\boldsymbol{w}_{[j]}$ are the parents and $\boldsymbol{w}_{[i\to j]}$ the children in the DAG, ϵ the step size, and

$$\boldsymbol{M}^*_{(m)} = \boldsymbol{M}_{(m-1)} + \kappa\left(\boldsymbol{G}_{\boldsymbol{w}_j} - \boldsymbol{M}_{(m-1)}\right)$$ **Adaptation of preconditioner**

This is what Simplified Manifold MALA would use!

$$\boldsymbol{g} = \boldsymbol{R}_j^{-1} \boldsymbol{H}_j \boldsymbol{w}_{[j]} + \sum_{i \in [j\to i]} \boldsymbol{H}_{[i]\setminus j}^\top \boldsymbol{R}_i^{-1} \boldsymbol{w}_{[i]\setminus j} + \boldsymbol{f}_{\boldsymbol{y},\boldsymbol{\beta},\boldsymbol{\tau}}$$

2 – Accept/Reject based on Metropolis ratio

3 – With probability $\gamma_m \downarrow 0$, Update $\boldsymbol{M}_{(m)}$ based on choice at step 2.

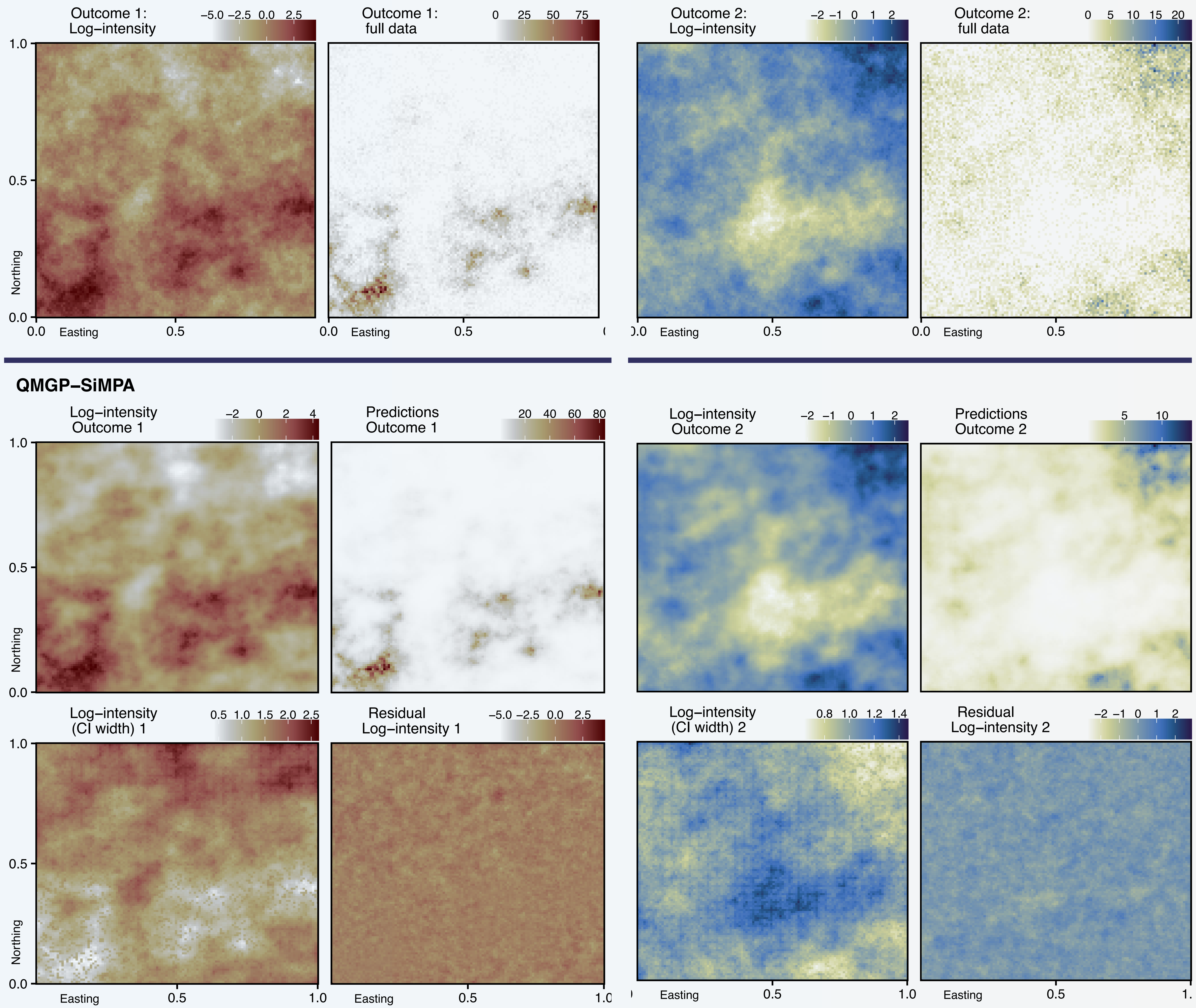
- **Cost** $\boldsymbol{O(q^2 n_j^2)}$ after adaptation period
- **Efficient move** due to using second order info

SiMPA: Simplified Manifold Preconditioner Adaptation

M. Peruzzi & D.B. Dunson (2022)
Spatial Meshing for General Bayesian Multivariate Models.
https://arxiv.org/abs/2201.10080

Poisson outcome 1

Poisson outcome 2

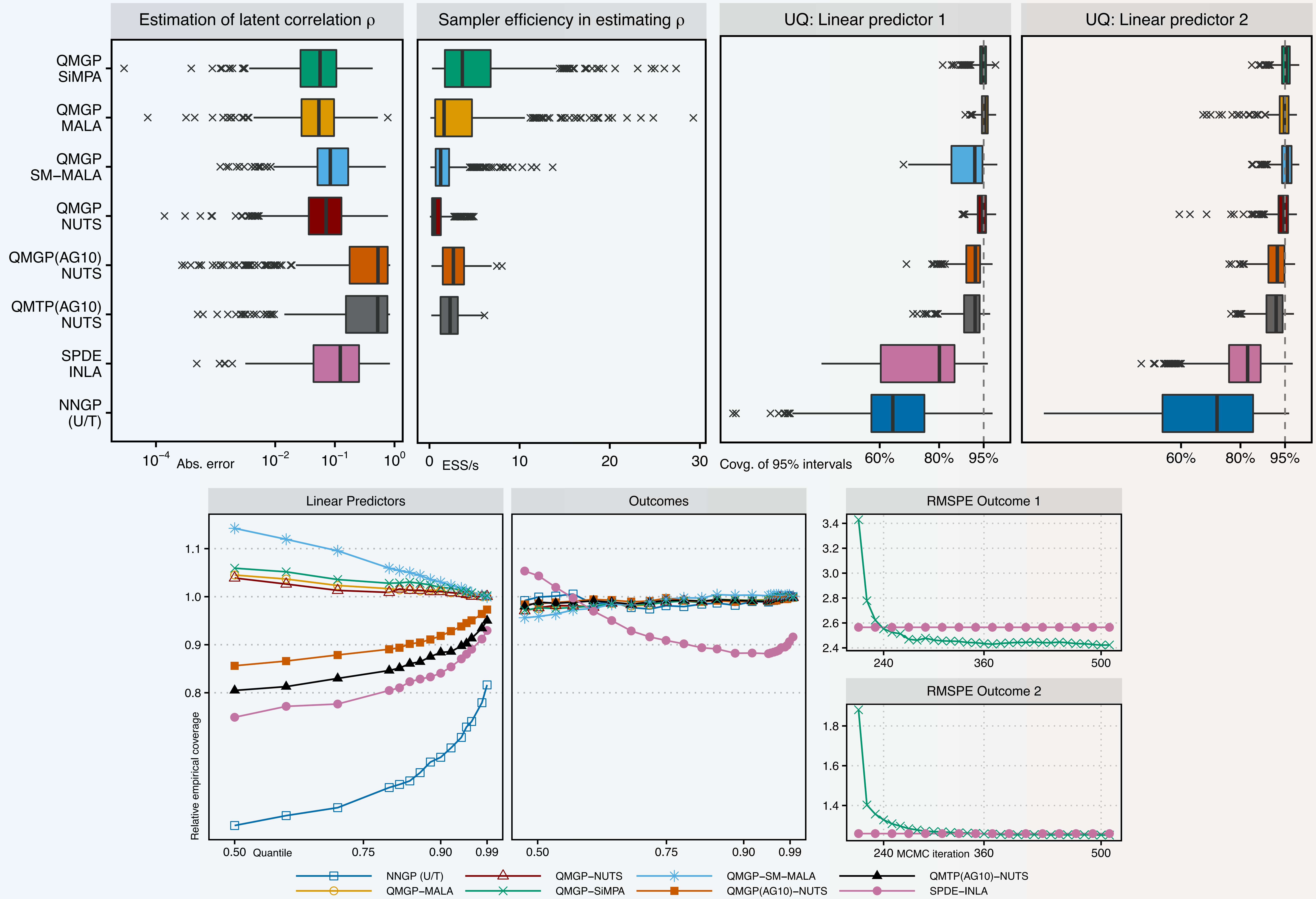


- Two related count outcomes
  - *n=3600*
  - leave out 20% of data
- Goals
  - prediction
  - recovery of latent log-intensity
  - uncertainty quant. about log-intensity

- meshed GPs with SiMPA
  - results in <10s
  - orders of magnitude faster than full GP

SiMPA: Simplified Manifold Preconditioner Adaptation

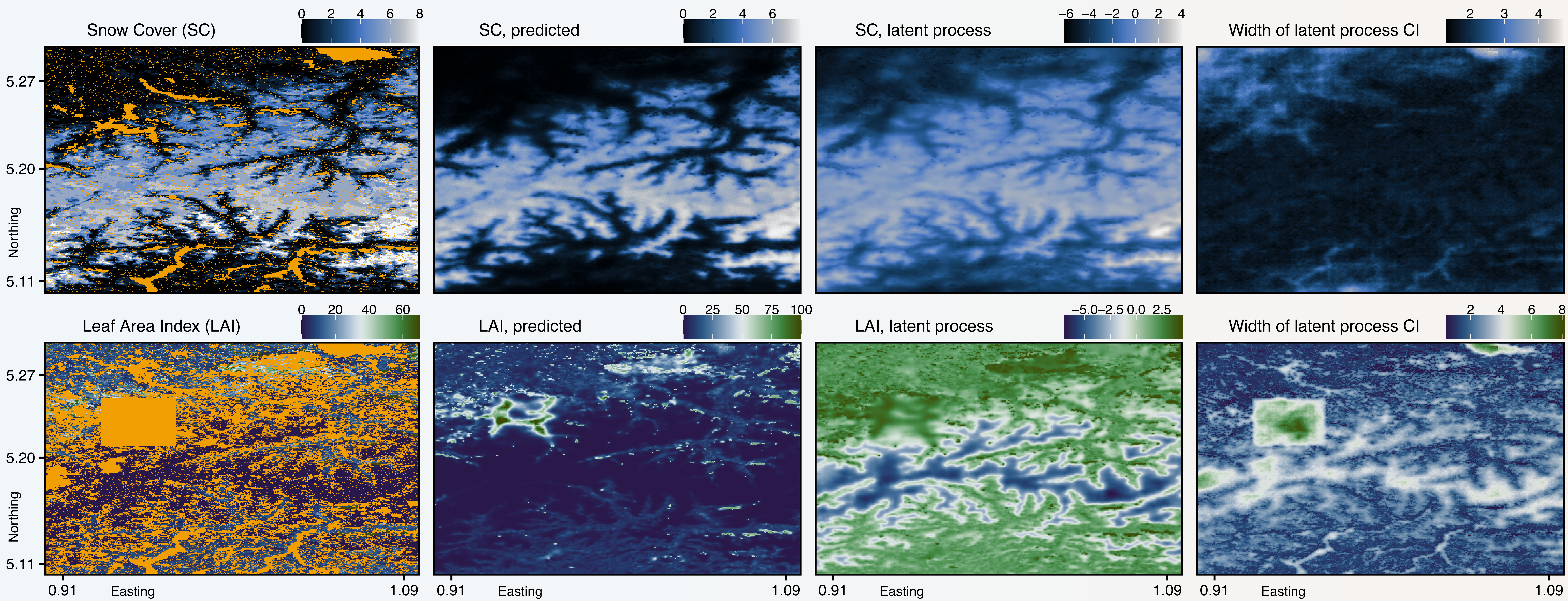
M. Peruzzi & D.B. Dunson (2022)
Spatial Meshing for General Bayesian Multivariate Models.
https://arxiv.org/abs/2201.10080



SiMPA: Simplified Manifold Preconditioner Adaptation

M. Peruzzi & D.B. Dunson (2022)
Spatial Meshing for General Bayesian Multivariate Models.
https://arxiv.org/abs/2201.10080

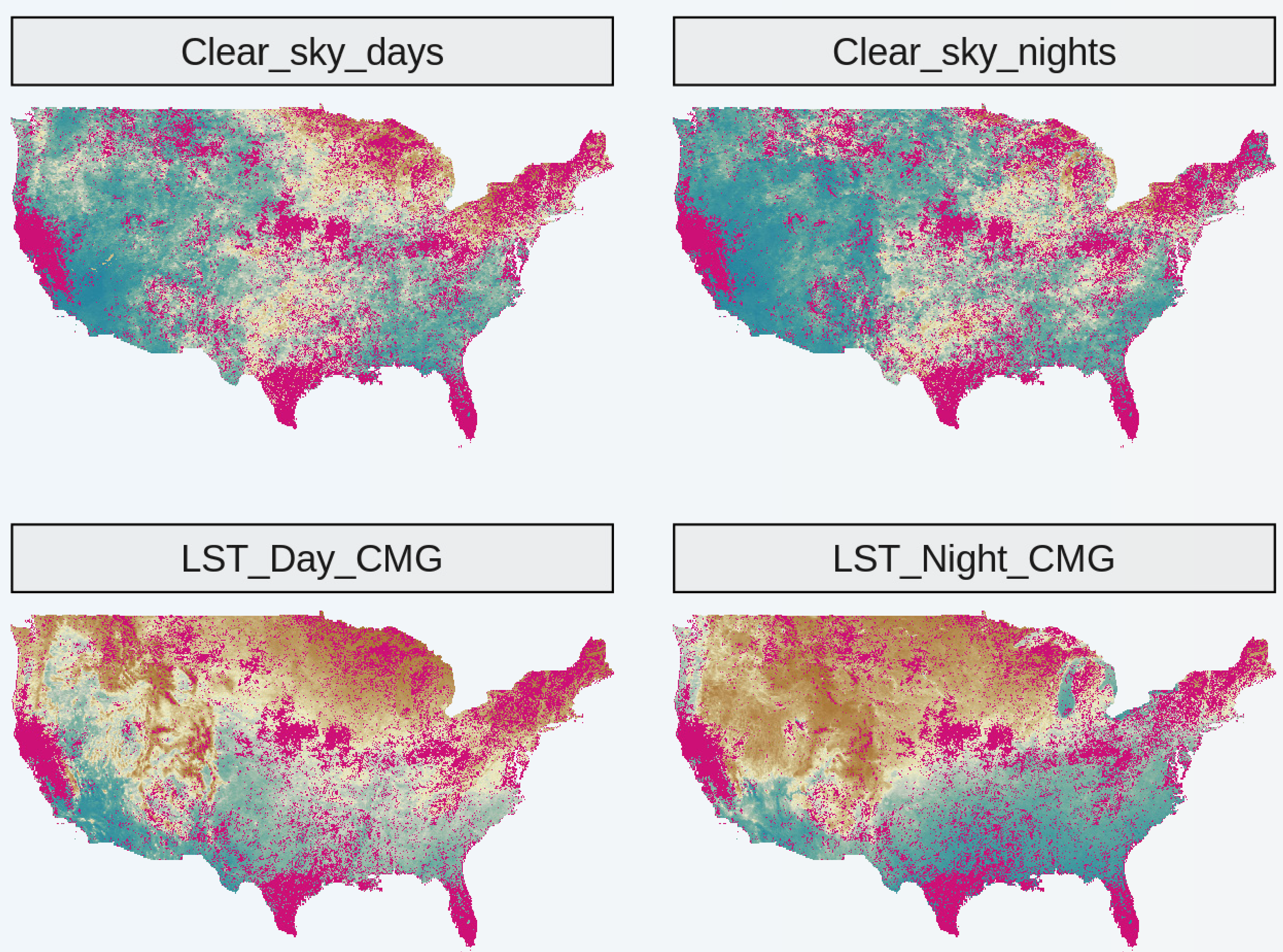
- MODIS data on Snow cover and Leaf Area Index over the central Alps
- Data size is about 250,000
- **Snow cover**: number of days with snow within 8-day period
- **Leaf area index**: leaf surface relative to ground surface (an integer)



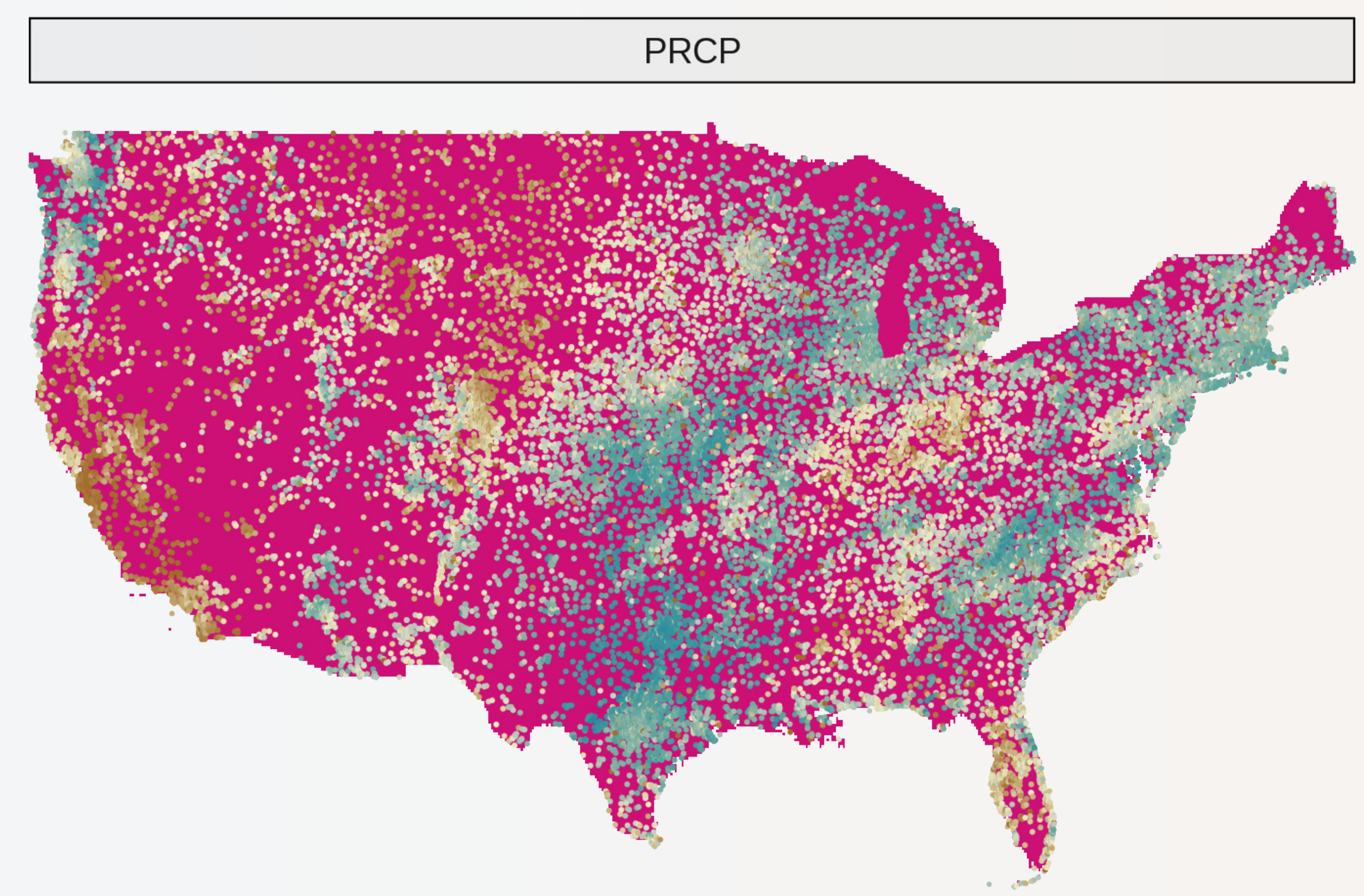
Multivariate misaligned data from MODIS (satellite) **and GHCN** (land based)

- Multivariate **misaligned data** difficult for methods based on neighbor search (QMGP, NNGP, Vecchia)
- Need DAG to produce reasonable spatial conditional independence
- The set of “neighbors” might include no data from some variable
- Develop new method to account for **different resolutions** of multiple measured outcomes

Missing data: MODIS

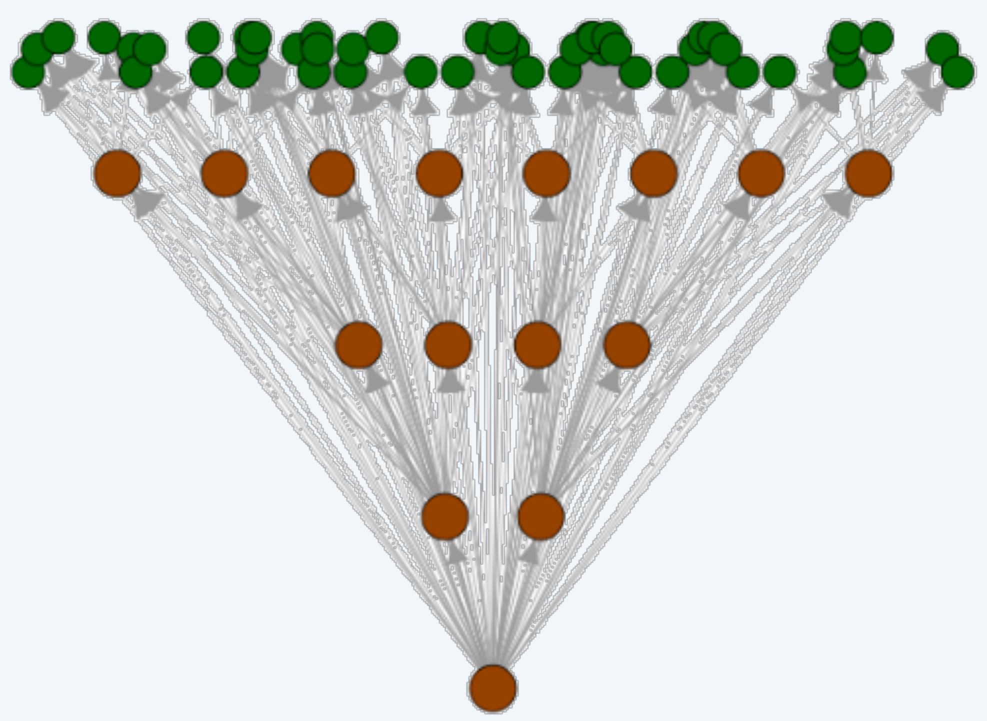


Missing data: GHCN



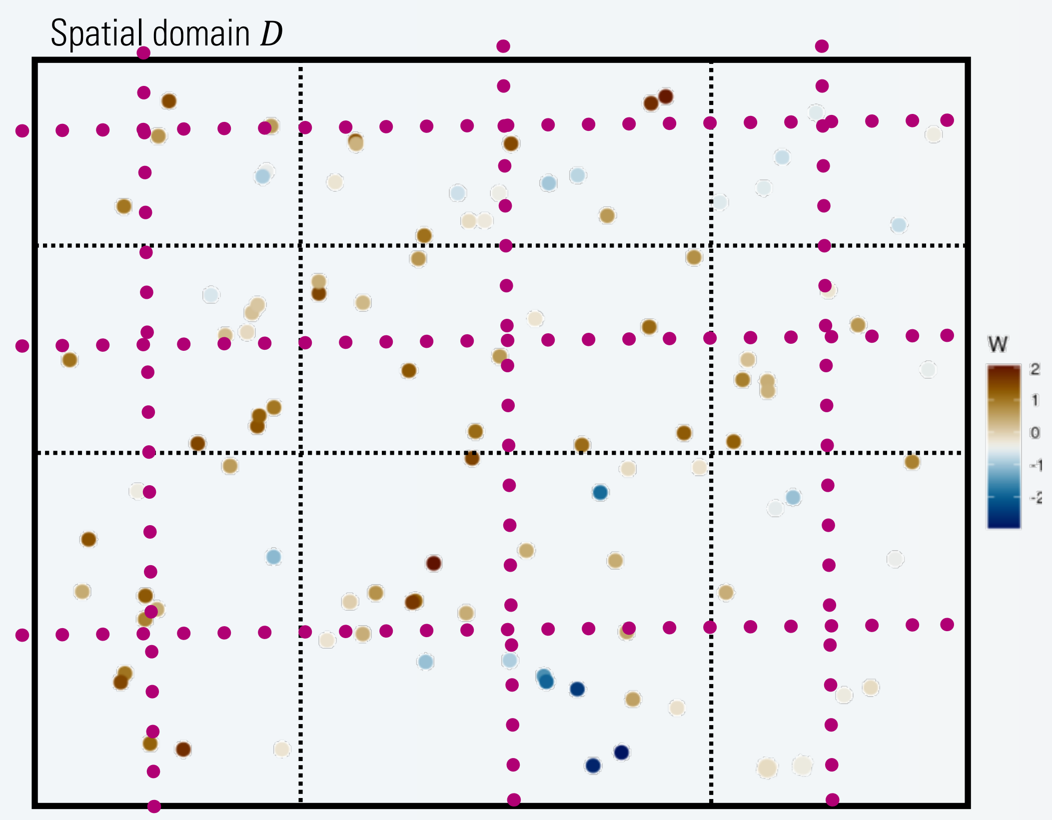
Spatial Multivariate Trees

M. Peruzzi & D.B. Dunson (2022)
Spatial Multivariate Trees for Big Data Bayesian Regression.
Journal of Machine Learning Research *23(17):1–40.*
https://www.jmlr.org/papers/v23/20-1361.html



Treed DAG

+



Recursive partitioning

+

Multivariate outcomes:
variable index j used as coordinate

= **Spatial Multivariate Tree**

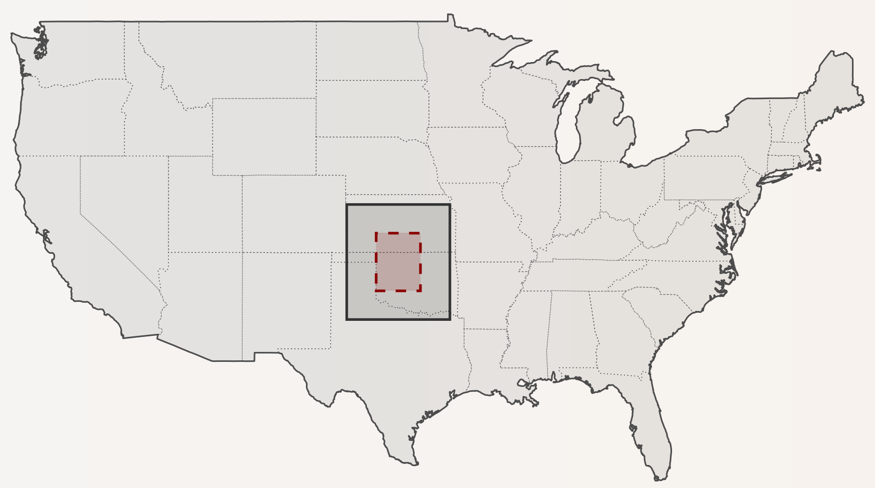
- Recursive domain partitioning & recursive treed DAG
- Outcomes at low resolution are placed **near the tree root**: reasonable **within-outcomes** conditional independence restriction
- Outcomes at high resolution fill the tree to ensure reasonable **between-outcomes** conditional independence restriction
- Large conditioning sets for variables at top levels ("information never hurts" principle)
- Reduced cost of building large $C_{[j]}^{-1}$: if ***j*** is at level ***T>0*** of the tree, ***[j]*** has size ***Tm*** but cost is $O(\cancel{T}^3 m^3)$

$$C_{[j]}^{-1} = \begin{bmatrix} C_{[i]}^{-1} + H_i^\top R_i^{-1} H_i & -H_i^\top R_i^{-1} \\ -R_i^{-1} H_i & R_i^{-1} \end{bmatrix}$$

Spatial Multivariate Trees: MODIS and GHCN weather data

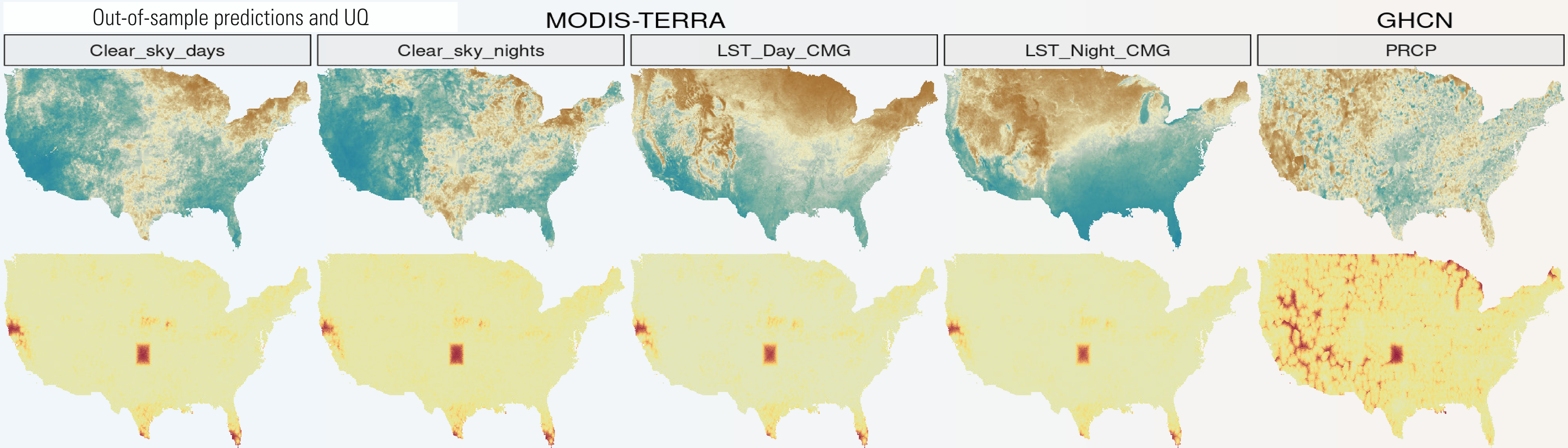
M. Peruzzi & D.B. Dunson (2022)
Spatial Multivariate Trees for Big Data Bayesian Regression.
Journal of Machine Learning Research *23(17):1–40.*
https://www.jmlr.org/papers/v23/20-1361.html

- 5 outcomes from 2 different sources
- MODIS **satellite** data
- GHCN **land-based station** data
- PRCP more sparsely observed
- Misalignment
- Data size **~ 1M**
- Compute time 16 hours or less



Test set

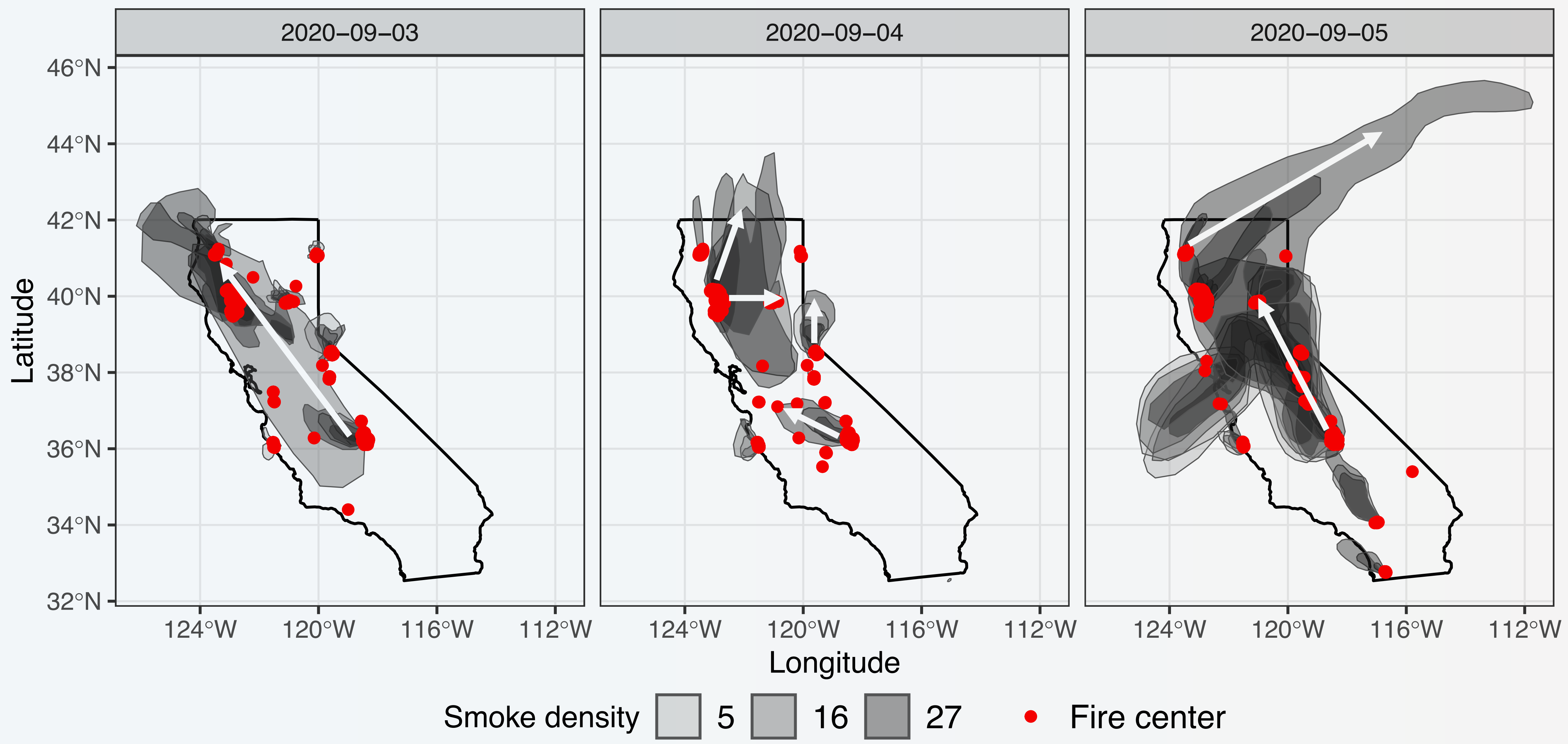
Out-of-sample predictions and UQ



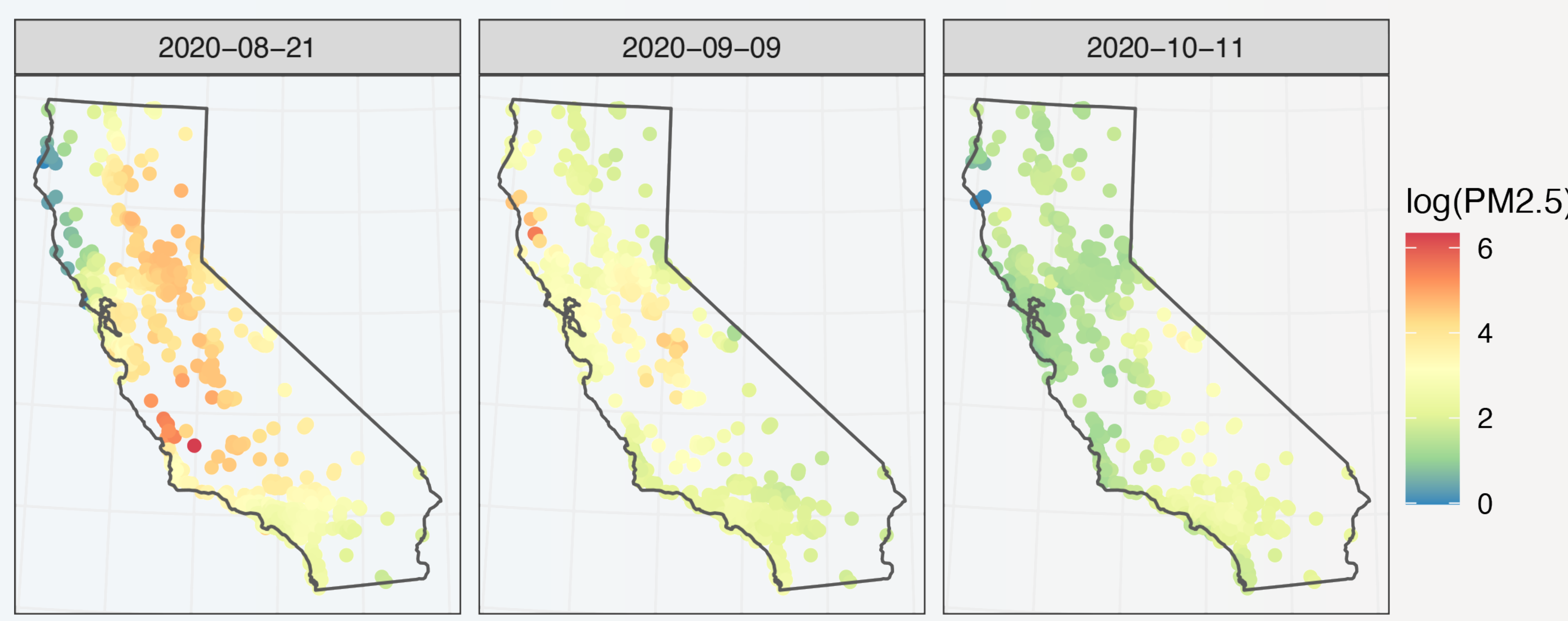
BAGs: Bags of directed Acyclic Graphs

B. Jin, **M. Peruzzi**, D.B. Dunson (2021)
Bag of DAGs: Flexible & Scalable Modeling of Spatiotemporal Dependence.
https://arxiv.org/abs/2112.11870

- **Application**: PM 2.5 due to forest fires depends on winds
- Measure PM2.5 using network of PurpleAir monitors
- Forest fires in California cause acute exposure to PM2.5



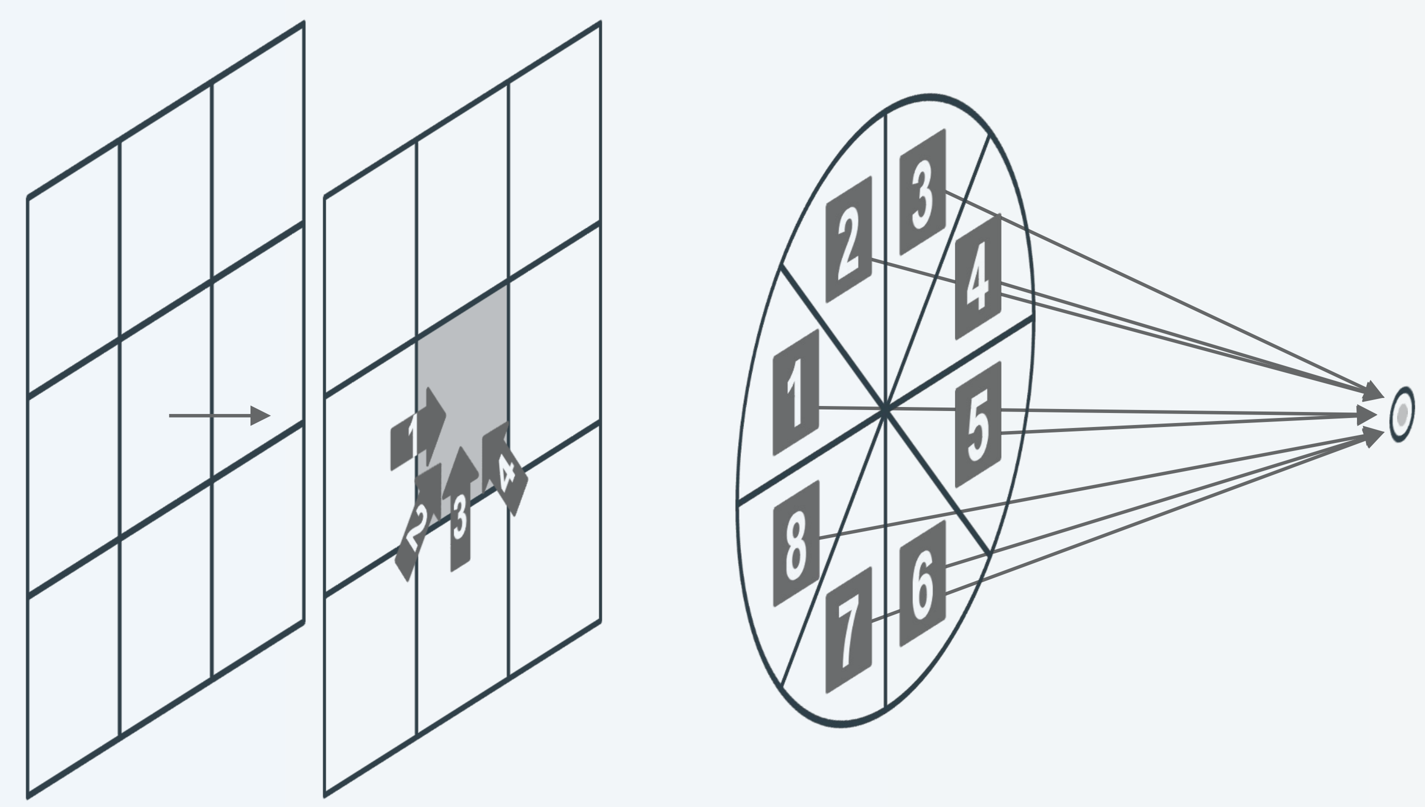
PM2.5 data

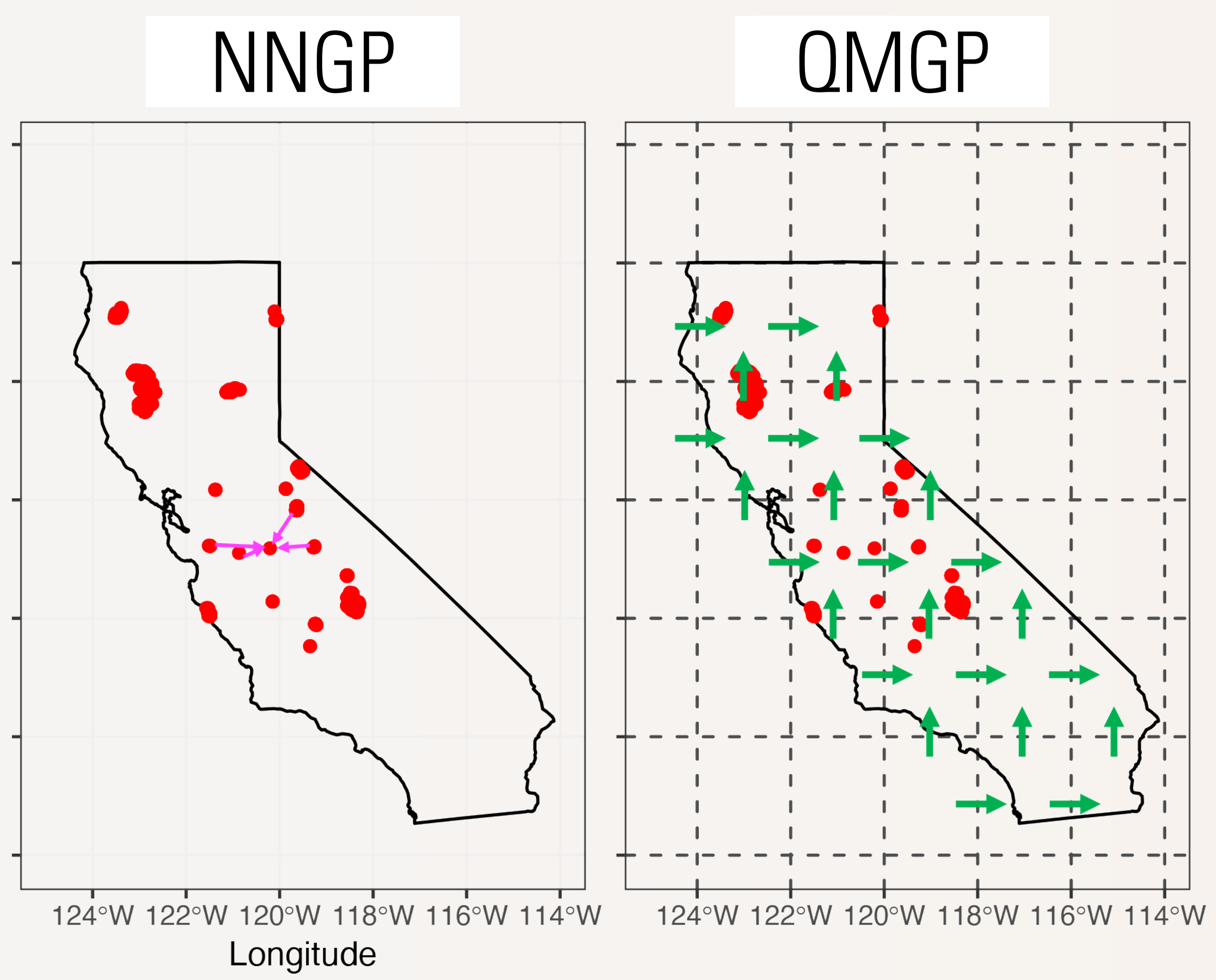


BAGs: Bags of directed Acyclic Graphs

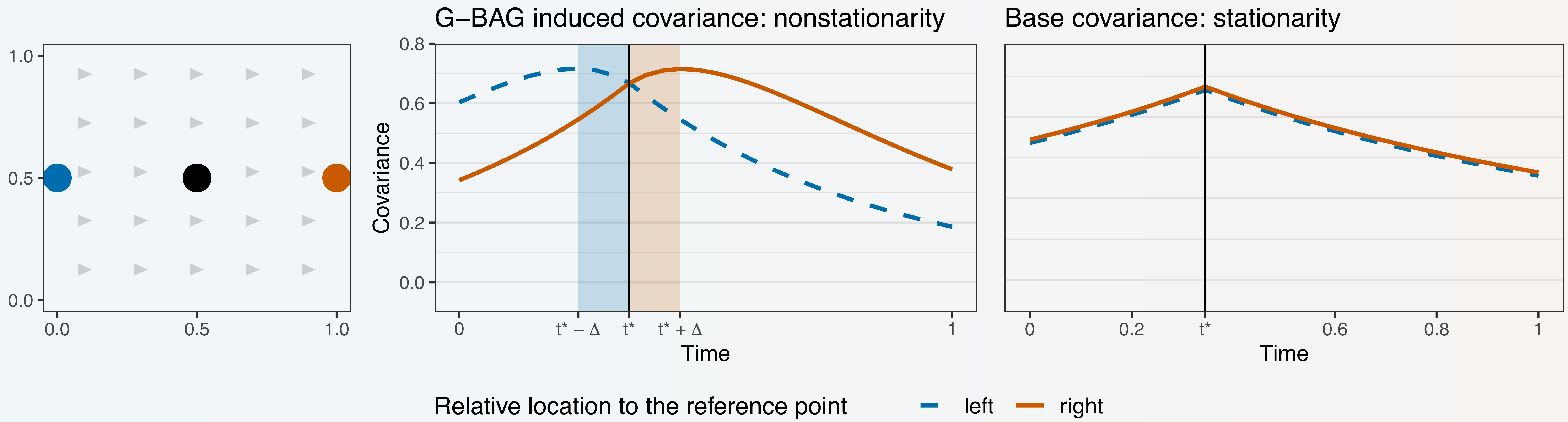
B. Jin, **M. Peruzzi**, D.B. Dunson (2021)
Bag of DAGs: Flexible & Scalable Modeling of Spatiotemporal Dependence.
https://arxiv.org/abs/2112.11870

- Directed graphical models have **causal interpretation**
- Fixed DAG may lead to **wrong** conditional independence assumptions



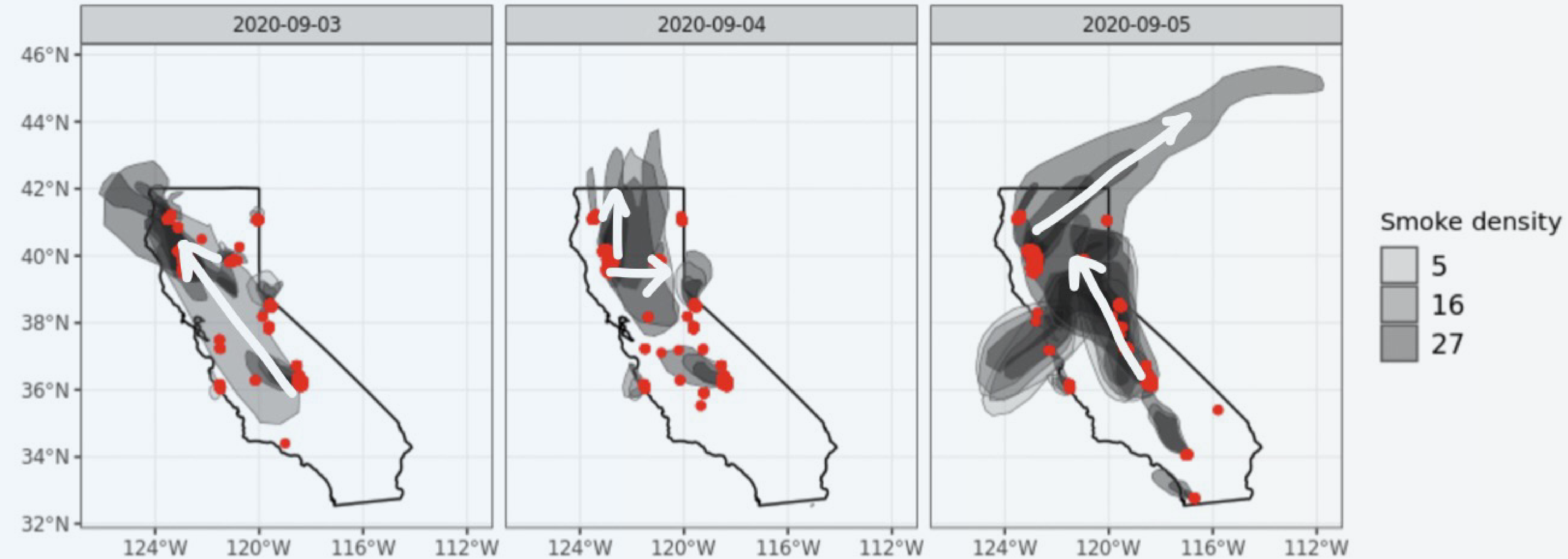


- Reduce sensitivity by **stochastically searching reasonable DAG edges** from a bag of allowed directions
- Inferred DAG edges inform about prevalent wind directions
- Interpretable, process based inference of directional dependence
- Scalable because DAG is sparse

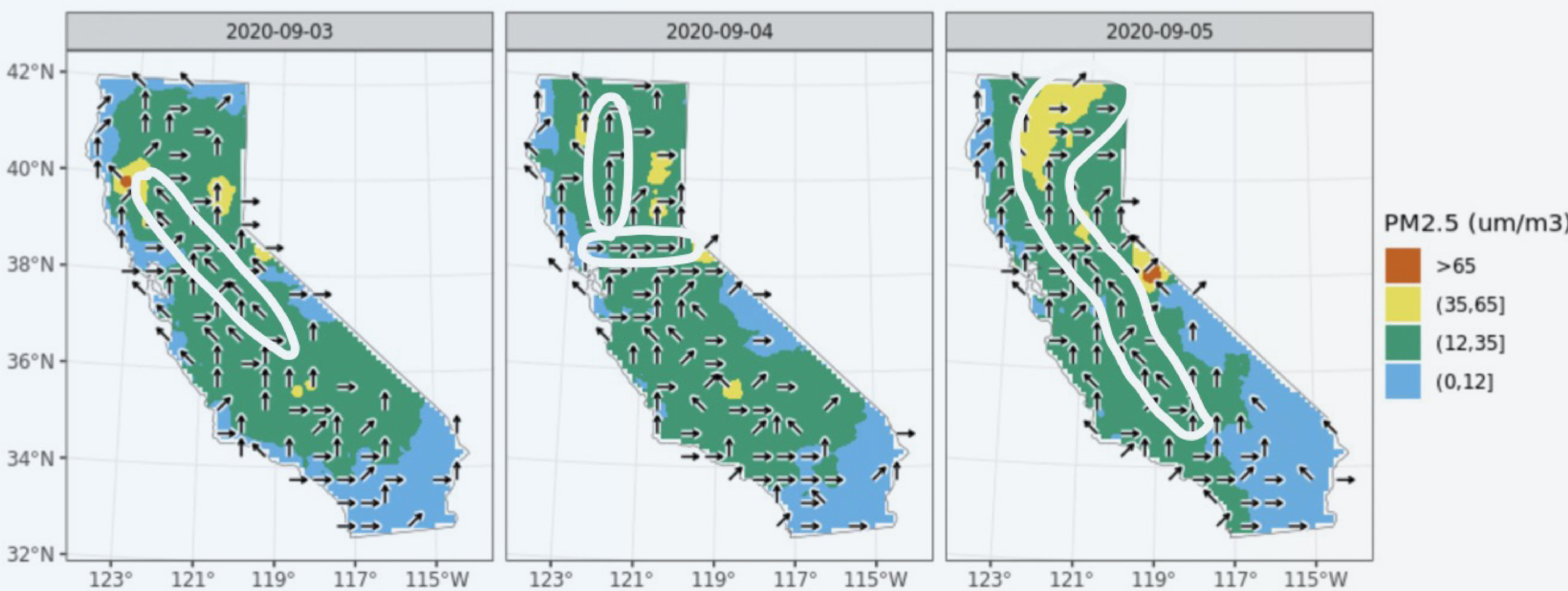


BAGs: Bags of directed Acyclic Graphs

B. Jin, **M. Peruzzi**, D.B. Dunson (2021)
Bag of DAGs: Flexible & Scalable Modeling of Spatiotemporal Dependence.
https://arxiv.org/abs/2112.11870

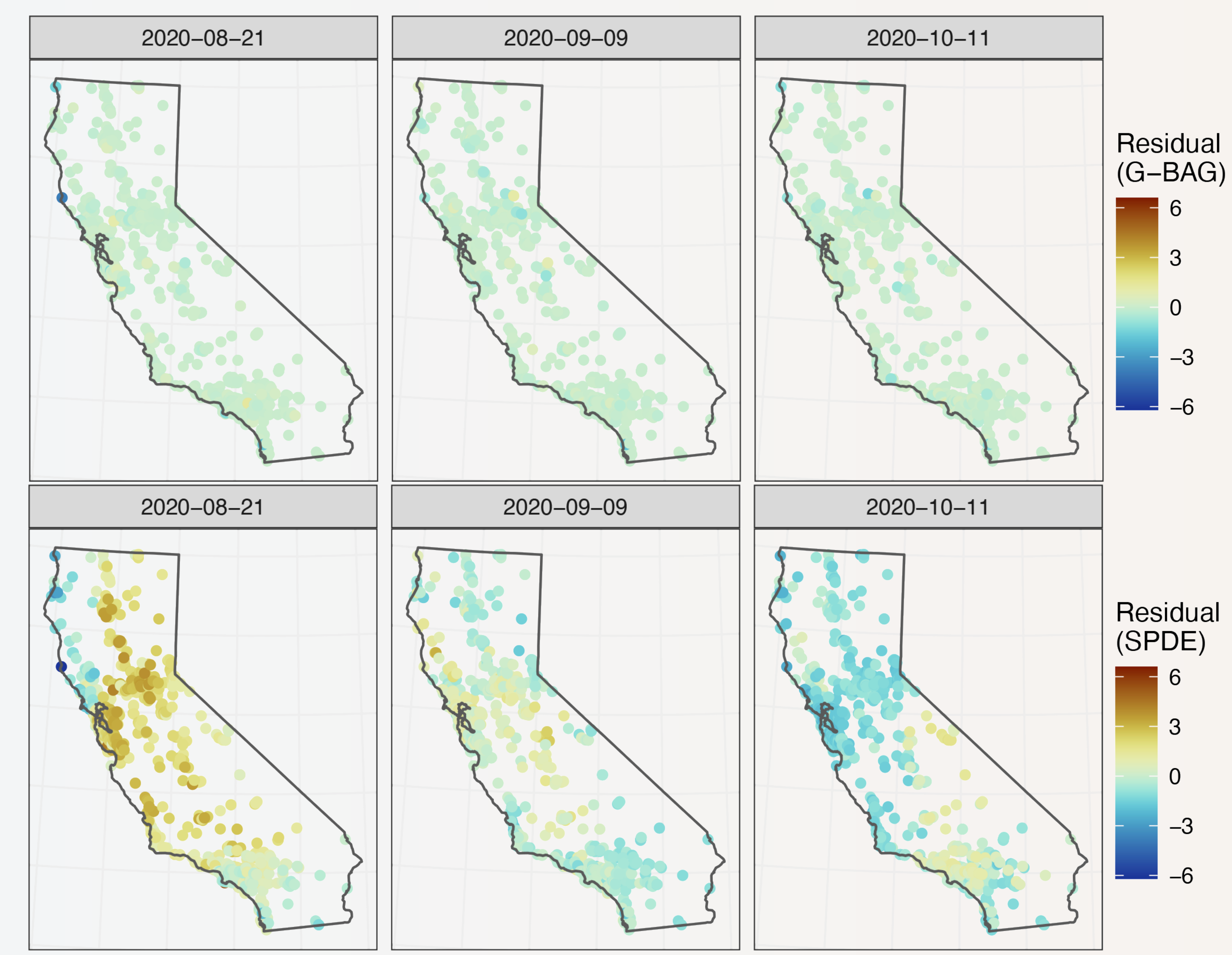


	G-BAG	Fixed DAG	SPDE-nonstationary
β	0.003 (−0.011, 0.016)	0.008 (−0.003, 0.021)	−0.038 (−0.054, −0.021)
τ^2	0.011 (0.011, 0.011)	0.011 (0.011, 0.011)	0.079 (0.076, 0.081)
σ^2	3.781 (3.600, 3.990)	4.410 (4.410, 4.410)	–
a	3.099 (2.963, 3.241)	1.262 (1.262, 1.262)	–
c	0.009 (0.008, 0.009)	0.010 (0.010, 0.010)	–
κ	0.011 (0.000, 0.041)	0.152 (0.152, 0.152)	–
RMSPE	0.296	0.343	0.343
MAPE	0.154	0.213	0.174
95% CI coverage	0.963	0.969	0.961
95% CI width	1.504	1.794	1.216



Inferred winds and levels of PM2.5

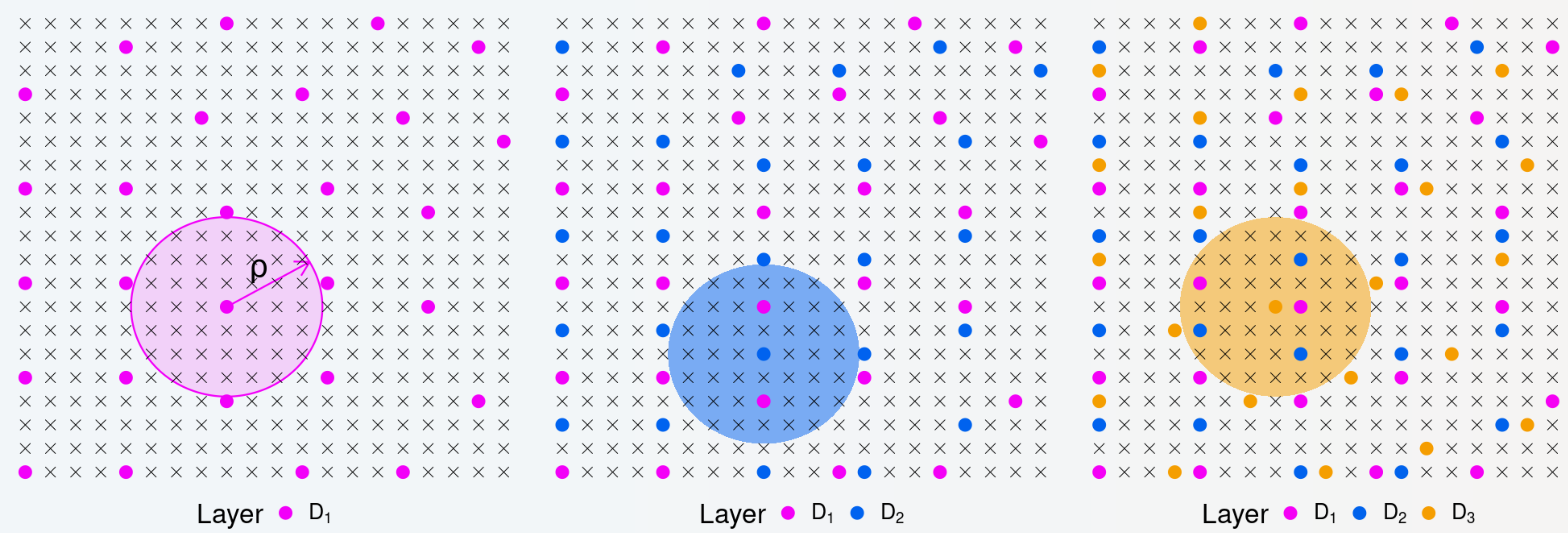
BAGs vs SPDE residuals

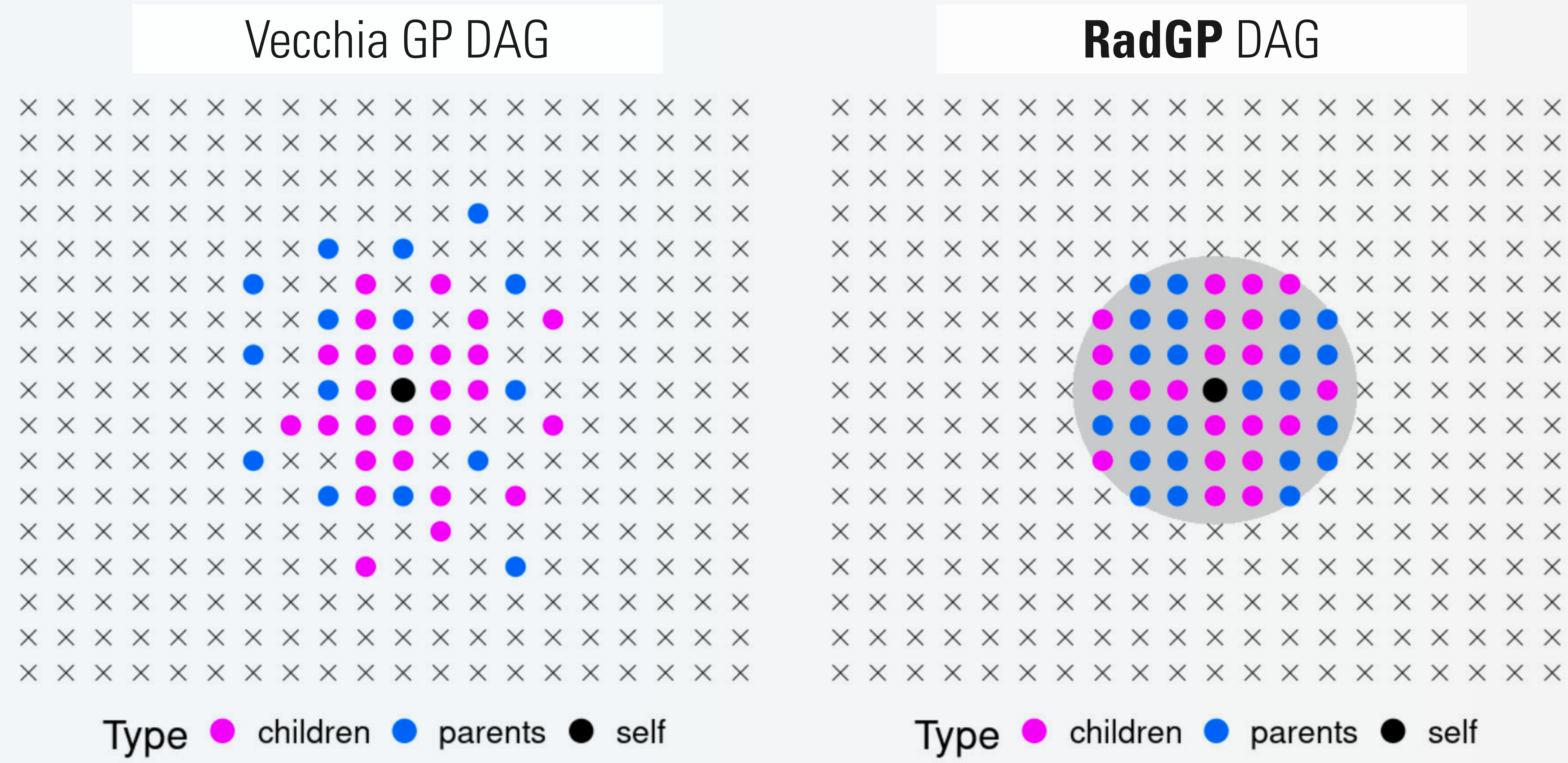


RadGP: provable approximation accuracy

Y. Zhu, **M. Peruzzi**, C. Li & D.B. Dunson (2022)
Radial Neighborss for Provably Accurate Scalable Approximations of Gaussian Processes
https://arxiv.org/abs/2211.14692

- Dual interpretation: approximation to a GPs or standalone processes based on a parent GP
- NNGP / Vecchia GP / MGP lack theoretical results on **approximation quality**
- Alternating partition construction leads to **Radial neighbors Gaussian process**
- Draw a radius around location: all other locations within radius are either parents or children in DAG





RadGP: provable approximation accuracy

Y. Zhu, **M. Peruzzi**, C. Li & D.B. Dunson (2022)
Radial Neighborss for Provably Accurate Scalable Approximations of Gaussian Processes
https://arxiv.org/abs/2211.14692

Intuition:

- If covariance matrix entries decay "fast enough" then its inverse inherits such property
- Take advantage of theory on norm-controlled inversion (Grochenig and Klotz 2014, Fang and Shin 2020)

The Gaussian process $Z(\cdot)$ has the isotropic covariance function $\mathrm{Cov}(Z(s_1), Z(s_2)) = K_0(\|s_1 - s_2\|_2)$.

Case 1 If the covariance function decays faster than any polynomials:

- Define the rate function $v_r(x) = \sum_{k=0}^{+\infty} x^k/(k!)^r$;
- Define the family $\mathscr{Z}_{v_r} = \left\{Z = (Z_s : s \in \Omega) : K_0(\|s_1 - s_2\|_2) \leq \frac{1}{v_r(\|s_1-s_2\|_2)(1+\|s_1-s_2\|_2^{d+1})}\right\}$.

Theorem 1 *For the family $\mathscr{Z}_{v_r}$ with $r > 1$, if $0 < q < 1$, then*

$$\sup_{Z \in \mathscr{Z}_{v_r}} W_2^2(Z_{\mathcal{D}}, \hat{Z}_{\mathcal{D}}) \lesssim \frac{n}{v_r(\rho/\sqrt{d})} \{\phi_0(c_2/q)\}^{-5} q^{-d} v_{r-1}(c_3\{\phi_0(c_2/q)\}^{-1}).$$

Else if $q \geq 1$, then $\sup_{Z \in \mathscr{Z}_{v_r}} W_2^2(Z_{\mathcal{D}}, \hat{Z}_{\mathcal{D}}) \lesssim n/v_r(\rho/\sqrt{d})$.

Case 2 Else if the covariance function decays no faster than some polynomials:

- Define the rate function $c_r(x) = (1 + |x|)^r$;
- Define the family $\mathscr{Z}_{c_r} = \left\{Z = (Z_s : s \in \Omega) : K_0(\|s_1 - s_2\|_2) \leq \frac{1}{c_r(\|s_1-s_2\|_2)}\right\}$.

Theorem 2 *For the family $\mathscr{Z}_{c_r}$ with $r \geq d + 1$, if $0 < q < 1$, then,*

$$\sup_{Z \in \mathscr{Z}_{c_r}} W_2^2(Z_{\mathcal{D}}, \hat{Z}_{\mathcal{D}}) \lesssim \frac{n}{(1 + \rho/\sqrt{d})^{-(r-d-1)}} q^{(r-8)d} \{\phi_0(c_2/q)\}^{-(r+9/2)} (c_1 c_5 d 2^{d-1} \pi/\sqrt{6})^r.$$

Else if $q \geq 1$, then $\sup_{Z \in \mathscr{Z}_{c_r}} W_2^2(Z_{\mathcal{D}}, \hat{Z}_{\mathcal{D}}) \lesssim n(1 + \rho/\sqrt{d})^{-(r-d-1)} \{c_1 c_5 d 2^{d-1} \phi_0(c_2/q) \pi/\sqrt{6}\}^r$.

RadGP: provable approximation accuracy

Y. Zhu, **M. Peruzzi**, C. Li & D.B. Dunson (2022)
Radial Neighborss for Provably Accurate Scalable Approximations of Gaussian Processes
https://arxiv.org/abs/2211.14692

- If minimal separation distance ***q*** and approximation radius are fixed,

  then approximation accuracy will not improve with increasing ***n***
- When minimal separation distance ***q*** is fixed, if the GP covariance decays fast enough

  then we can be arbitrarily accurate for any n by increasing the approximation radius
- If minimal separation distance ***q*** decreases with increasing ***n***,

  then a larger approximation radius is required to compensate for near-singular cov. matrix

Examples

Covariance function $K_0(\|\Delta s\|_2)$	Lower bounds for ρ
Matérn: $\frac{\sigma^2 2^{1-\nu}}{\Gamma(\nu)}(\alpha\|\Delta s\|)^{\nu}\mathcal{K}_{\nu}(\alpha\|\Delta s\|_2)$	$\rho \gtrsim \frac{\sqrt{d}}{\alpha}\left[c_{m,1}\left(1+\frac{c_2^2}{\alpha^2 q^2}\right)^{\nu+\frac{d}{2}} \ln\left\{c_{m,1} n q^{-d}\left(1+\frac{c_2^2}{\alpha^2 q^2}\right)^{5(\nu+\frac{d}{2})}\right\}\right]^3$
Gaussian: $\exp(-a\|\Delta s\|_2^2)$	$\rho \gtrsim \frac{\sqrt{d}}{\alpha}\left[e^{\frac{c_2^2}{4aq^2}}\left\{\ln(nq^{-d})+\frac{c^2}{4aq^2}\right\}\right]^3$
G-Cauchy: $\sigma^2\left\{1+(\|s_1-s_2\|/\alpha)^{\delta}\right\}^{-\lambda/\delta}$	$\rho \gtrsim q^{-\left\{\frac{25}{2}\lambda d+\delta(\lambda+\frac{9}{2})\right\}/\{\lambda-(d+1)\}}$

R package `meshed`

`meshed::spmeshed` targets the following **spatial factor model** for multivariate outcomes:

$$y_j(\boldsymbol{\ell}) \sim P_j(\eta_j(\boldsymbol{\ell}), \tau_j) \qquad j = 1, \dots, q$$
$$\boldsymbol{\eta}(\boldsymbol{\ell}) = \boldsymbol{X}\boldsymbol{B} + \boldsymbol{\Lambda}\boldsymbol{v}$$
$$\boldsymbol{v}(\cdot) \sim \text{meshed}GP(\boldsymbol{0}, \boldsymbol{C}_{\boldsymbol{\theta}})$$

where $\boldsymbol{C}_{\boldsymbol{\theta}}$ is a cross-correlation function with indep. Matérn margins, $\boldsymbol{\Lambda}$ a $q \times k$ lower triangular matrix with positive diagonal

Accepting any combination of:

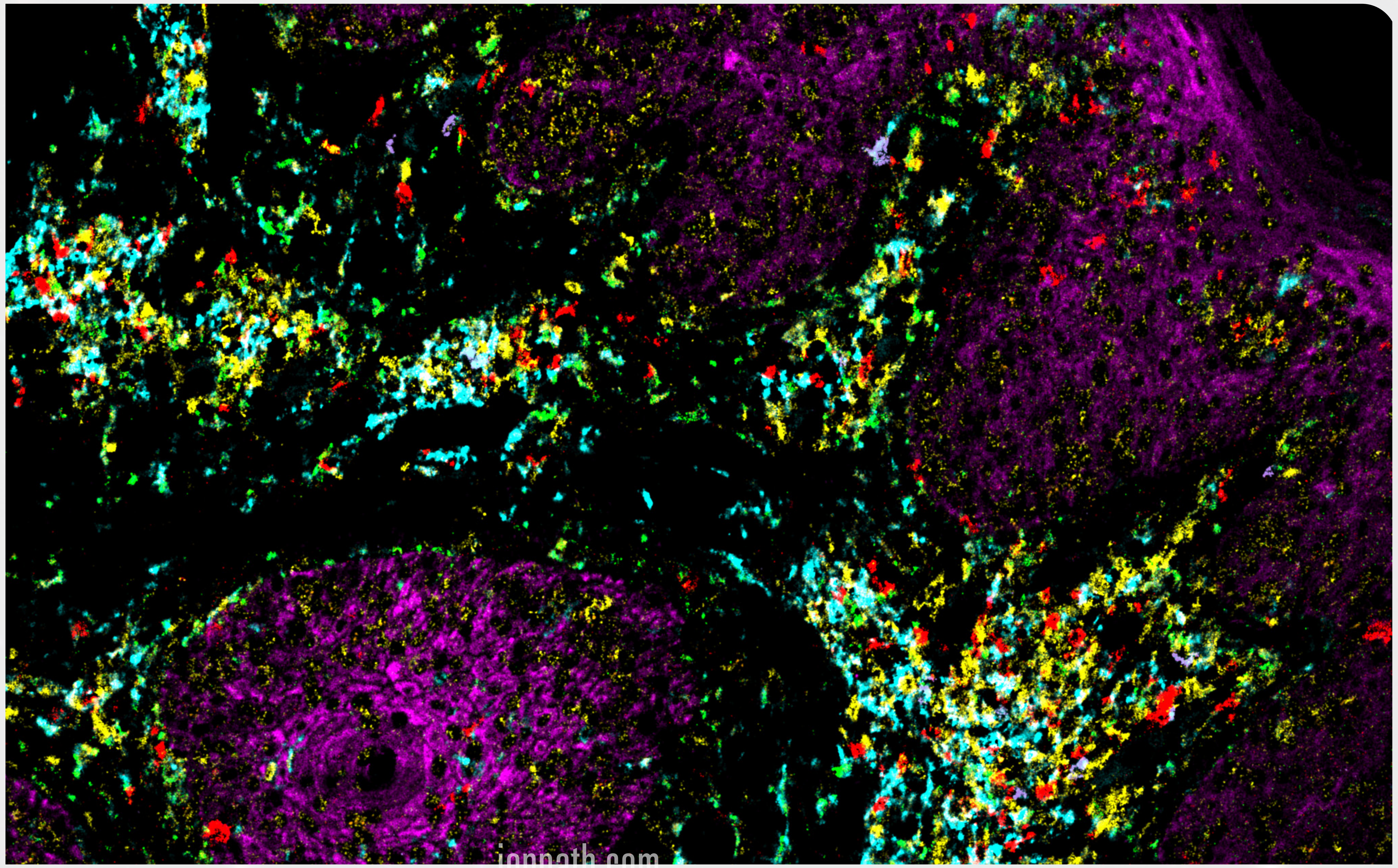
- gridded data, or data at irregular spatial locations
- spatial or spatiotemporal data (using Gneiting's nonseparable space-time correlation)
- univariate or multivariate outcomes
- spatial misalignment of multivariate outcomes
- outcomes of different types (Gaussian, binomial, Poisson, negative binomial, beta)

- Works with fast **BLAS/Lapack** libraries and in parallel using **OpenMP** if you let it
- Compares favorably to `spNNGP` in univariate settings (`spNNGP` does not do multivariate)

Looking ahead

Some ideas: application/modeling/computational challenges

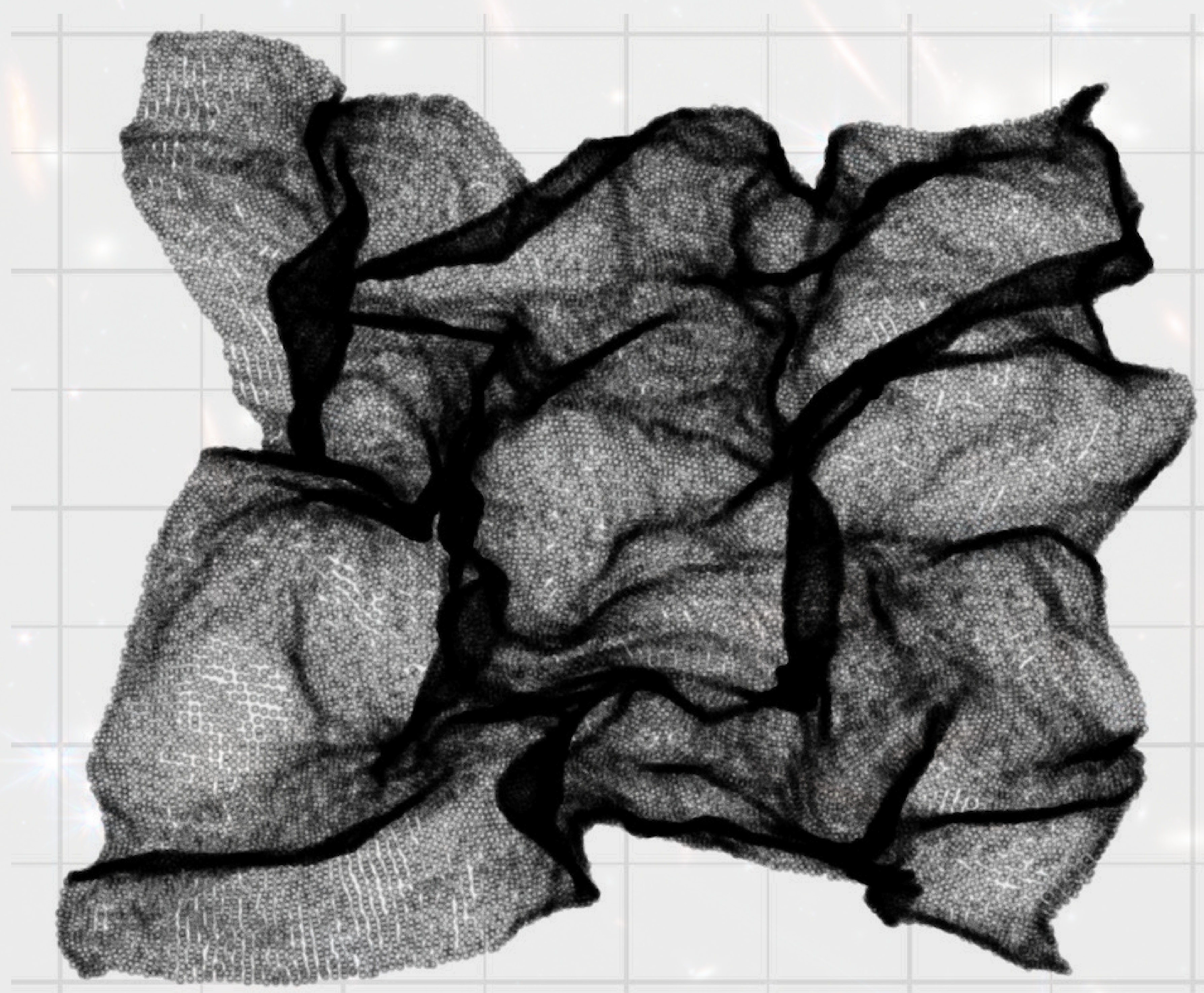
- Interest in modeling ***spatial*** co-variability of different protein types?
- Spatial occurrence of proteins in cells = multivariate spatial latent factor model
- Multiple cells or multiple subjects = common factor loadings, possibly hierarchically to borrow strength
- Multiple subjects (healthy vs not?): characterize groups via differences in spatial variability. Early detection?
- Odd “shape” of cell structures? Use spatial deformation/deep GP.

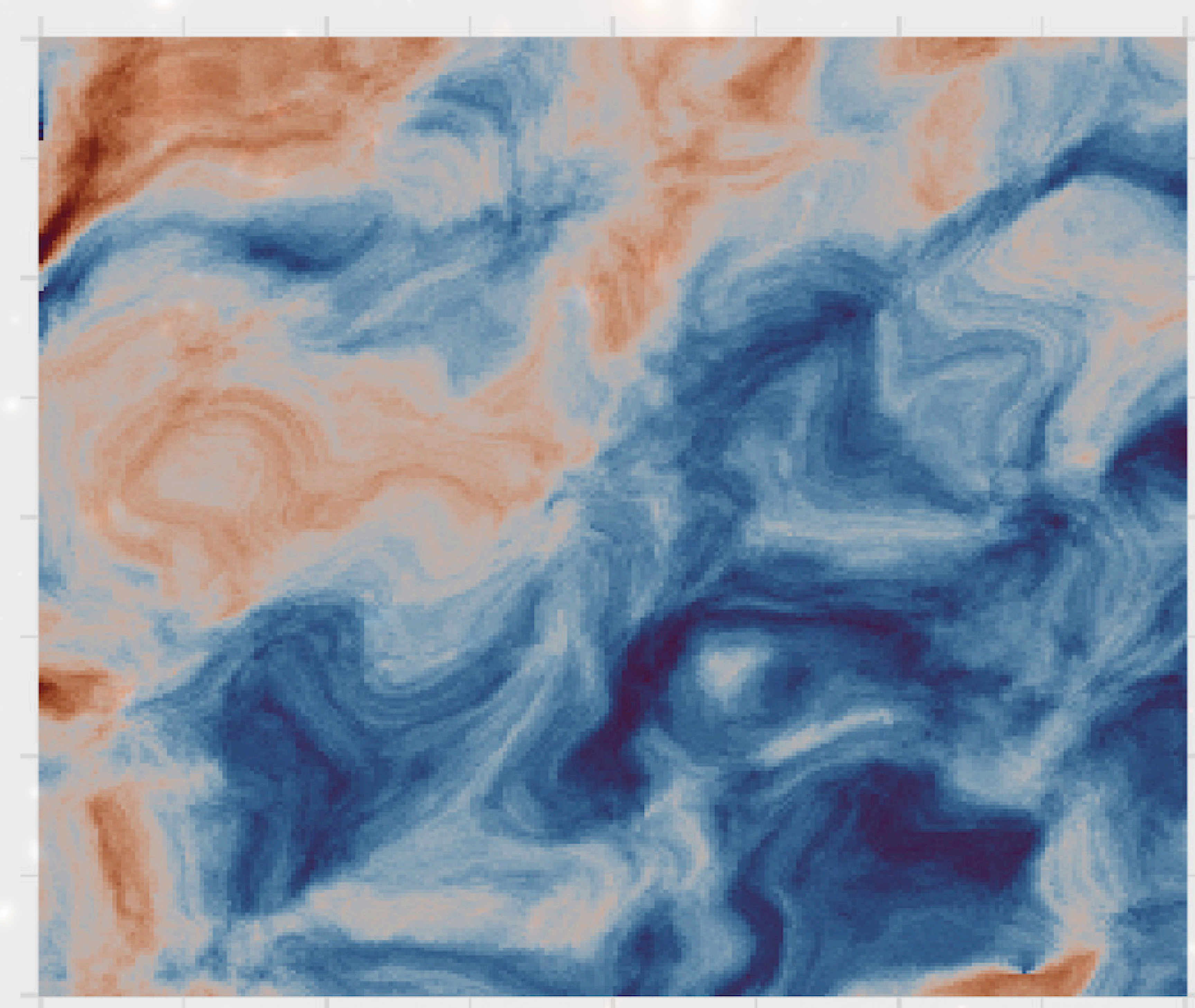


ionpath.com

Some ideas: application/modeling/computational challenges

- Interest in modeling ***spatial*** co-variability of different protein types?
- Spatial occurrence of proteins in cells = multivariate spatial latent factor model
- Multiple cells or multiple subjects = common factor loadings, possibly hierarchically to borrow strength
- Multiple subjects (healthy vs not?): characterize groups via differences in spatial variability. Early detection?
- Odd “shape” of cell structures? Use spatial deformation/deep GP. Scalability to high resolution cell data? TBD.





Thank you!

contact: **michele.peruzzi@duke.edu**

M. Peruzzi, S. Banerjee & A.O. Finley (2022)
Highly Scalable Bayesian Geostatistical Modeling via Meshed Gaussian Processes on Partitoned Domains.
Journal of the American Statistical Association *117(538): 969-982.*
https://www.tandfonline.com/doi/full/10.1080/01621459.2020.1833889

M. Peruzzi & D.B. Dunson (2022)
Spatial Multivariate Trees for Big Data Bayesian Regression.
Journal of Machine Learning Research *23(17):1–40.*
https://www.jmlr.org/papers/v23/20-1361.html

M. Peruzzi & D.B. Dunson (2022)
Spatial Meshing for General Bayesian Multivariate Models.
https://arxiv.org/abs/2201.10080

M. Peruzzi, S. Banerjee, D.B. Dunson & A.O. Finley (2021)
Grid-Parametrize-Split (GriPS) for Improved Scalable Inference in Spatial Big Data Analysis.
https://arxiv.org/abs/2101.03579

Y. Zhu, **M. Peruzzi**, C. Li, D.B. Dunson (2022)
Radial Neighbors for Provably Accurate Scalable Approximations of Gaussian Processes.
https://arxiv.org/abs/2211.14692

B. Jin, **M. Peruzzi**, D.B. Dunson (2021)
Bag of DAGs: Flexible & Scalable Modeling of Spatiotemporal Dependence.
https://arxiv.org/abs/2112.11870

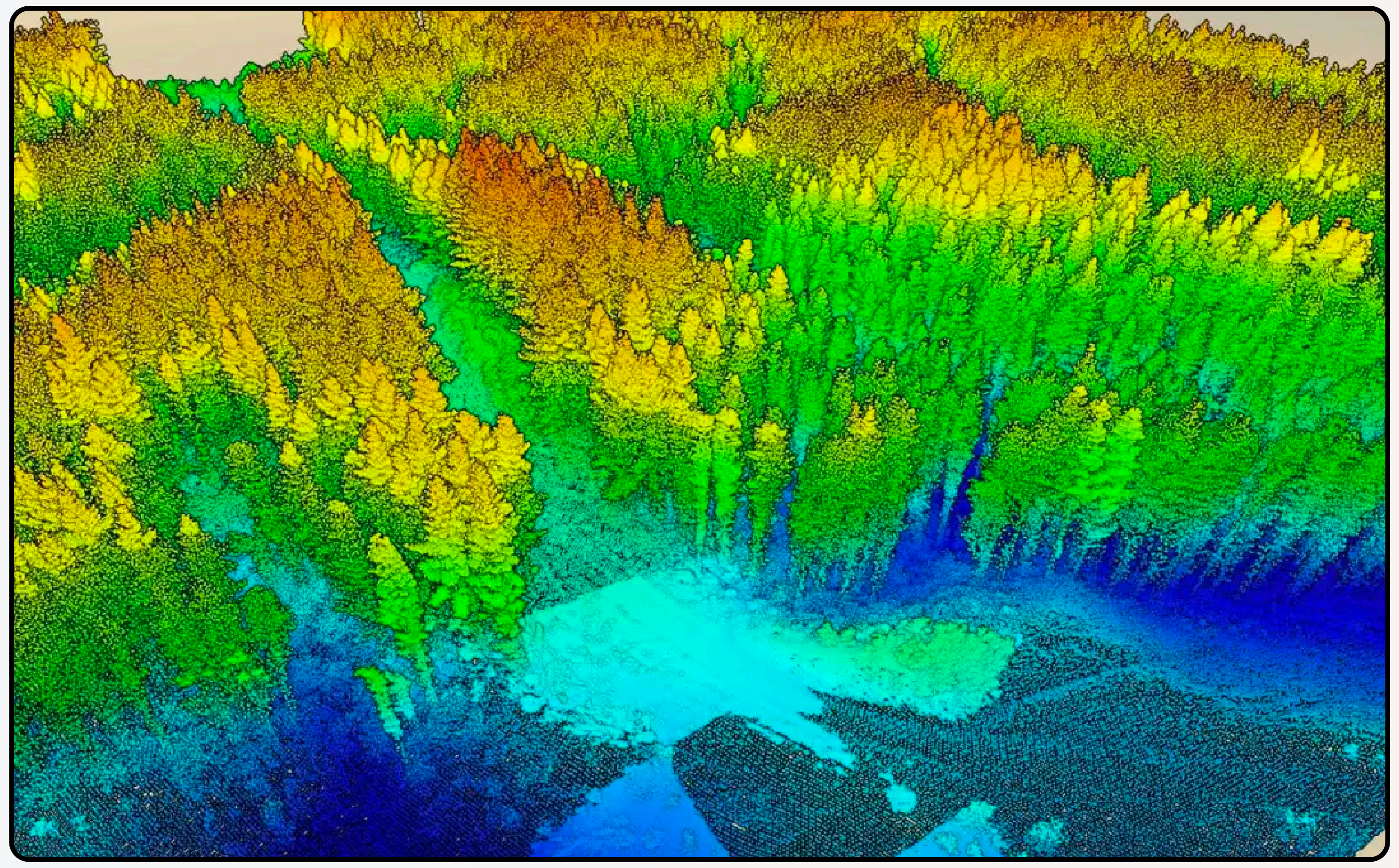
N. Neupane, **M. Peruzzi**, L. Ries, A. Arab, S. J. Mayor, J. C. Withey, A. O. Finley (2022)
A novel model to accurately predict continental-scale green-up timing.
International J. of Applied Earth Observation and Geoinformation *108:102747*
https://doi.org/10.1016/j.jag.2022.102747

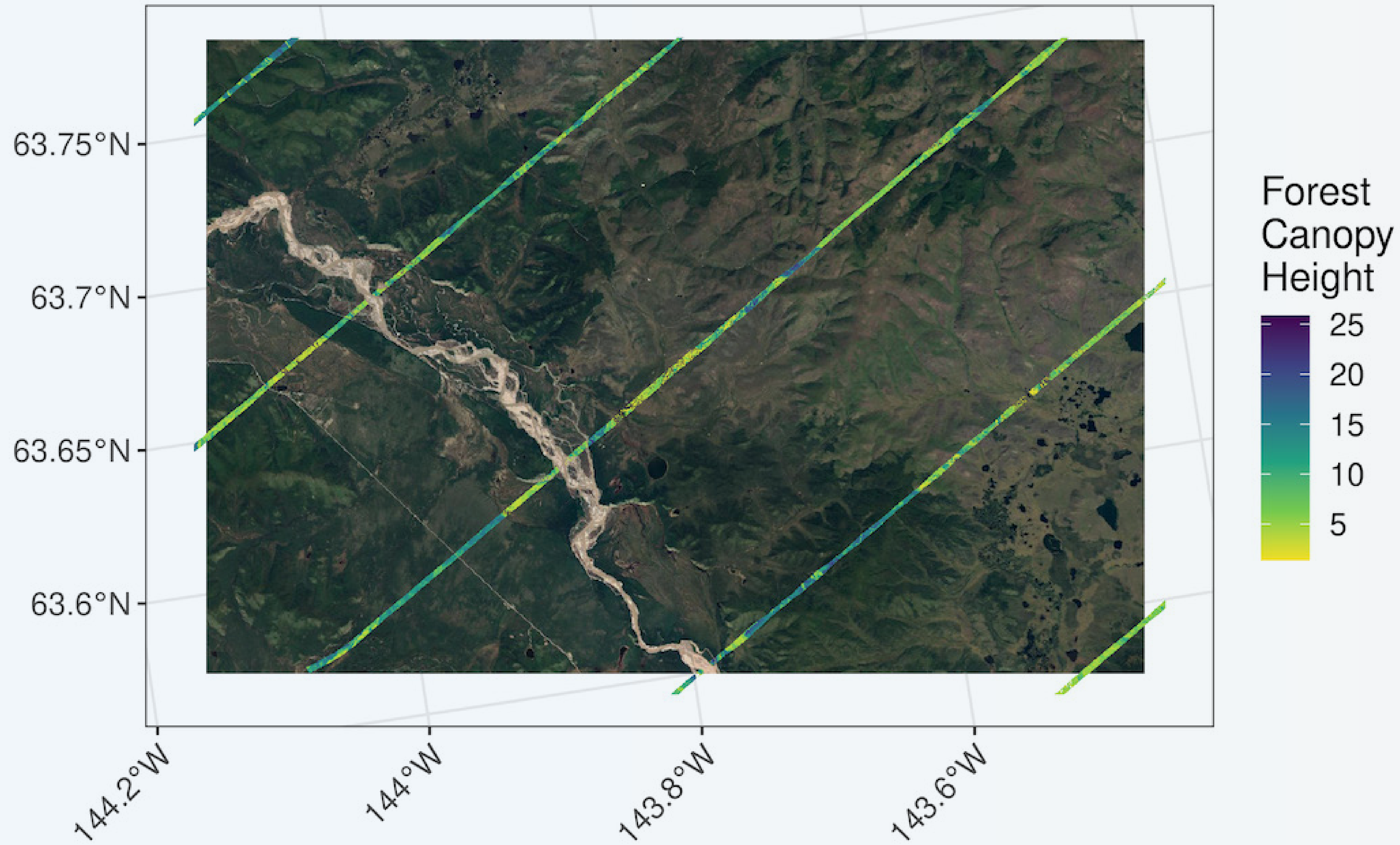
GriPS: Gridding and parameter expansion for better MCMC

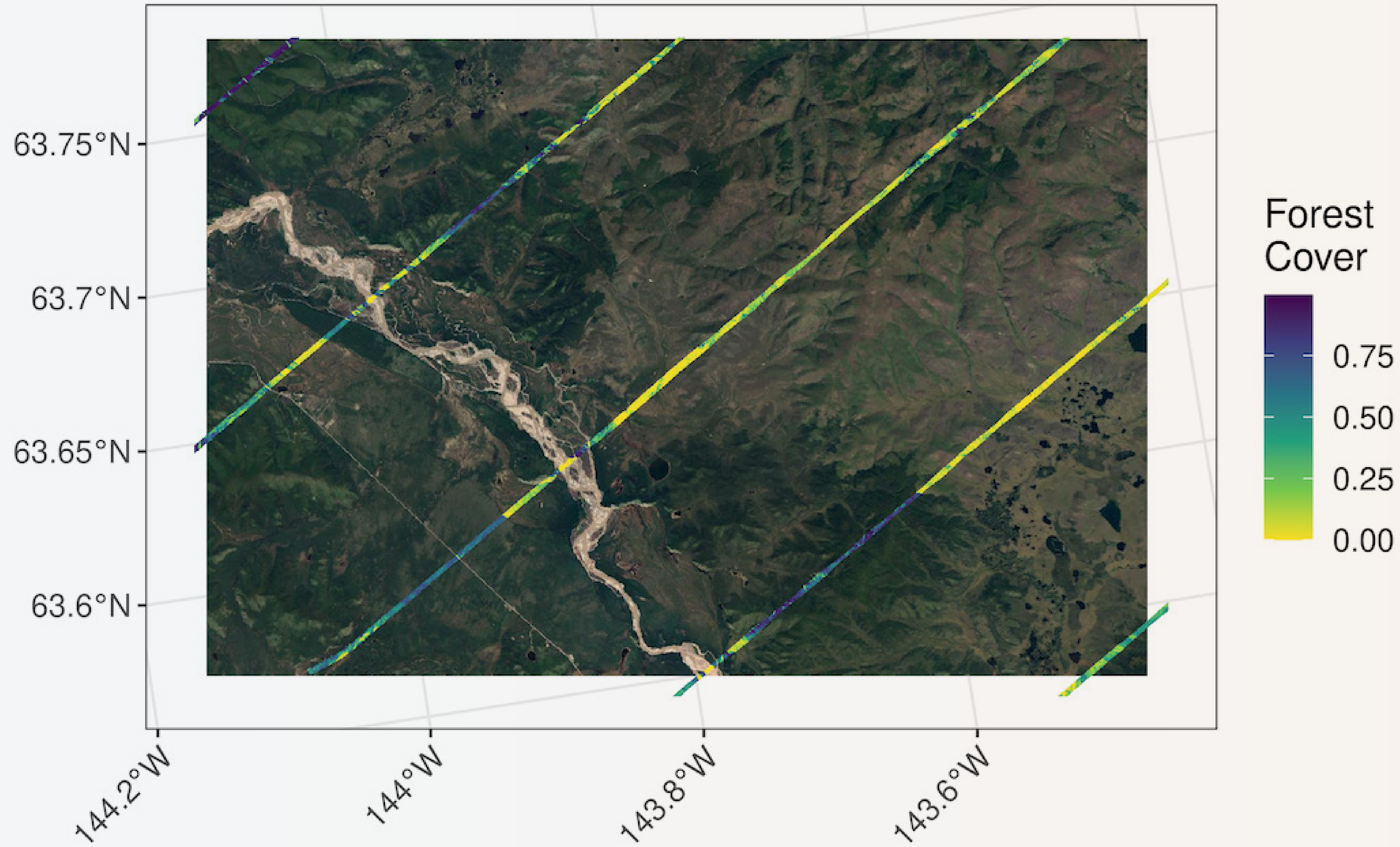
M. Peruzzi, S. Banerjee, D.B. Dunson & A.O. Finley (2021)
Grid-Parametrize-Split (GriPS) for Improved Scalable Inference in Spatial Big Data Analysis.
https://arxiv.org/abs/2101.03579

Application: LiDAR data

- Tanana forest, Alaska: 2 outcomes at **2.5M locations**
- Forest cover and canopy height
- Data measured at thin strips, 9km apart
- Difficult to set up usual NNGP or a standard QMGP
- Take advantage of **modeling flexibility of MGPs**
- Highly customized setup for MGPs for under 48h compute time
- Custom grid





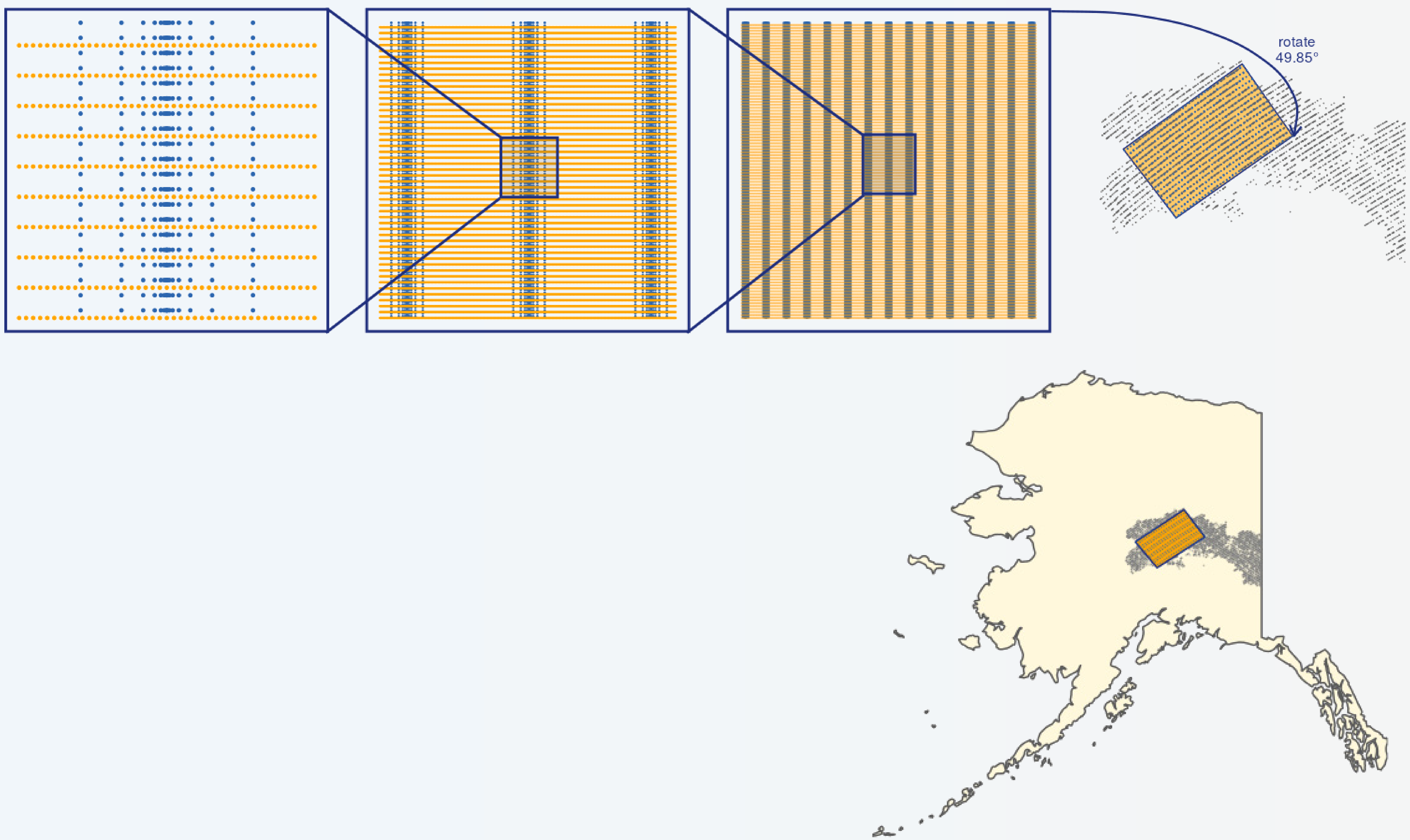


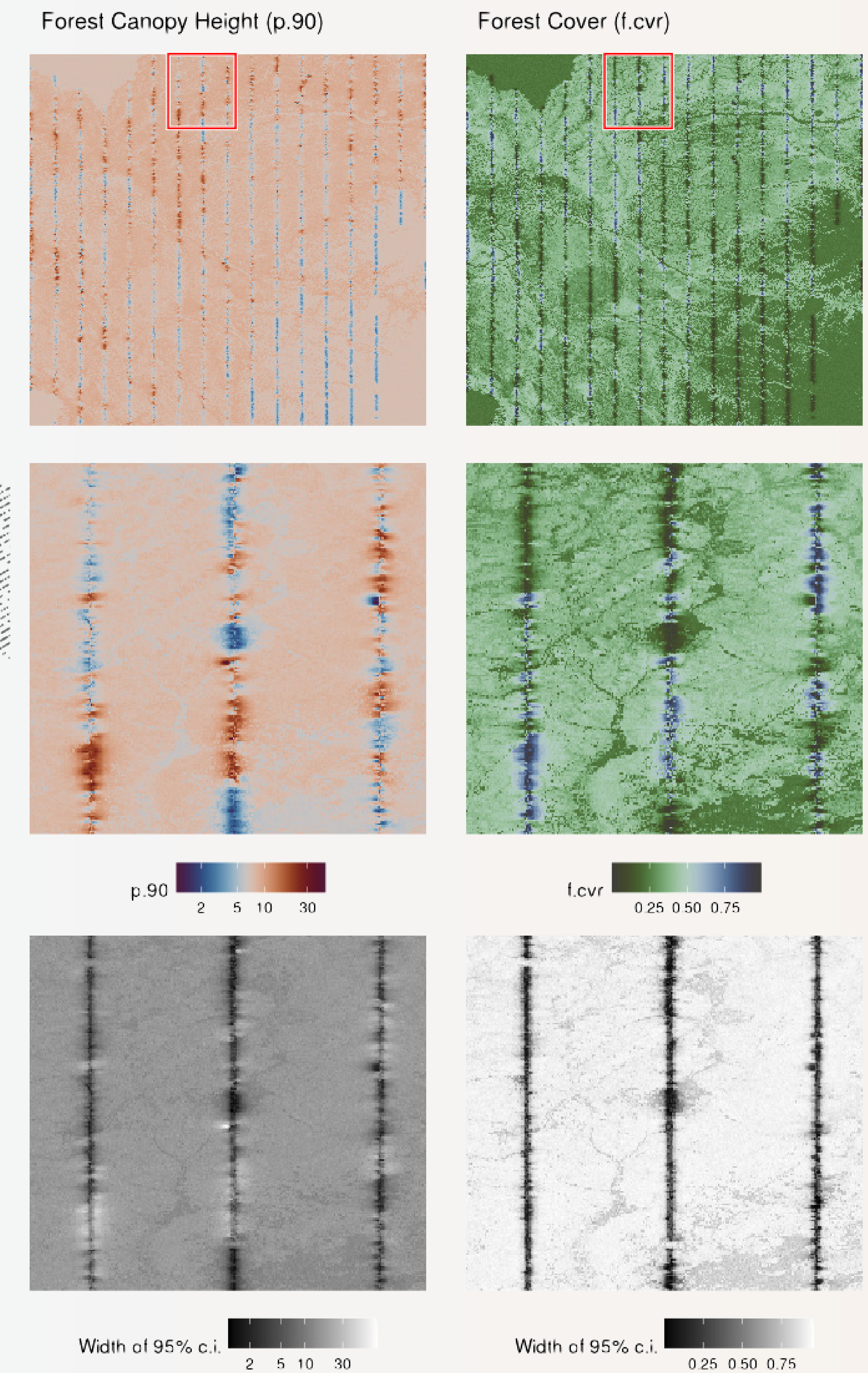
GriPS: Gridding and parameter expansion for better MCMC

M. Peruzzi, S. Banerjee, D.B. Dunson & A.O. Finley (2021)
Grid-Parametrize-Split (GriPS) for Improved Scalable Inference in Spatial Big Data Analysis.
https://arxiv.org/abs/2101.03579

Application: LiDAR data

- Tanana forest, Alaska: 2 outcomes at **2.5M locations**
- Forest cover and canopy height
- Highly customized setup for MGPs for under 48h compute time
- Custom grid





GriPS: Gridding and parameter expansion for better MCMC

M. Peruzzi, S. Banerjee, D.B. Dunson & A.O. Finley (2021)
Grid-Parametrize-Split (GriPS) for Improved Scalable Inference in Spatial Big Data Analysis.
https://arxiv.org/abs/2101.03579

- Matérn covariance **weakly identifiable**: high posterior dependence between σ^2, ϕ, ν
- Targeting the usual identifiable LMC model $\boldsymbol{y}(\boldsymbol{\ell}) = \boldsymbol{X}(\boldsymbol{\ell})^\top \boldsymbol{\beta} + \boldsymbol{\Lambda} \boldsymbol{w}(\boldsymbol{\ell}) + \boldsymbol{\varepsilon}(\boldsymbol{\ell})$ we actually run MCMC on the expanded model:

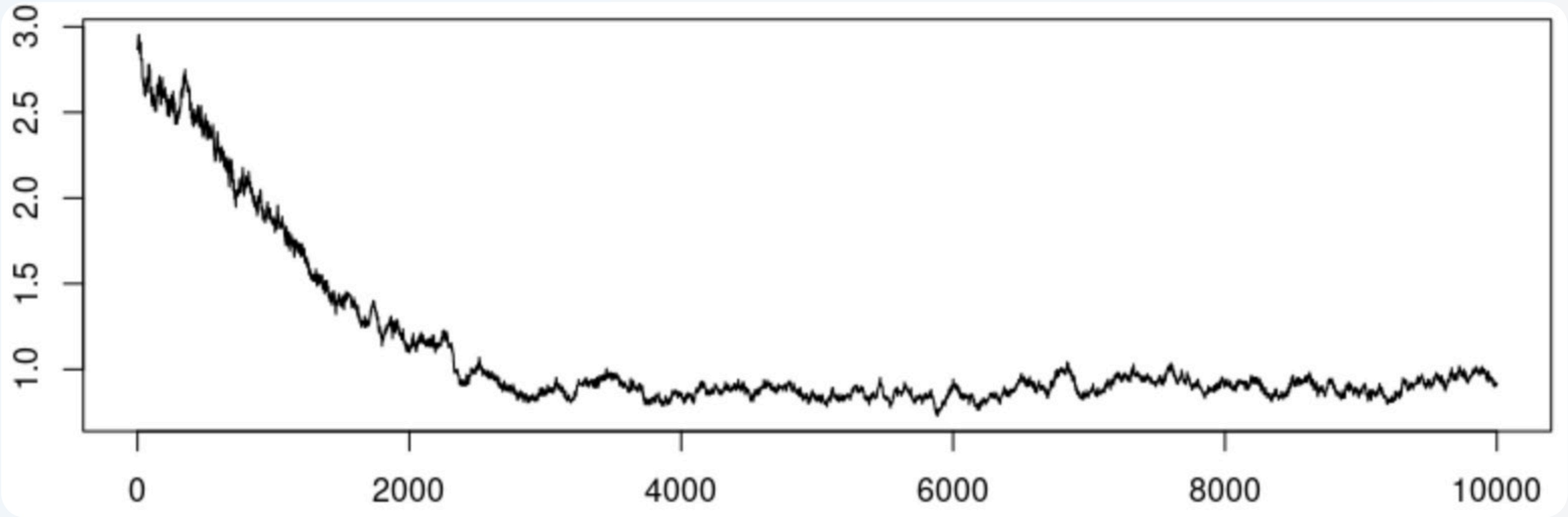
$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\mathcal{Z}}\boldsymbol{r} + \boldsymbol{\varepsilon}' \; ; \quad \boldsymbol{\varepsilon}' \sim N(\boldsymbol{0}, \boldsymbol{D}_n + \boldsymbol{\Sigma})$$

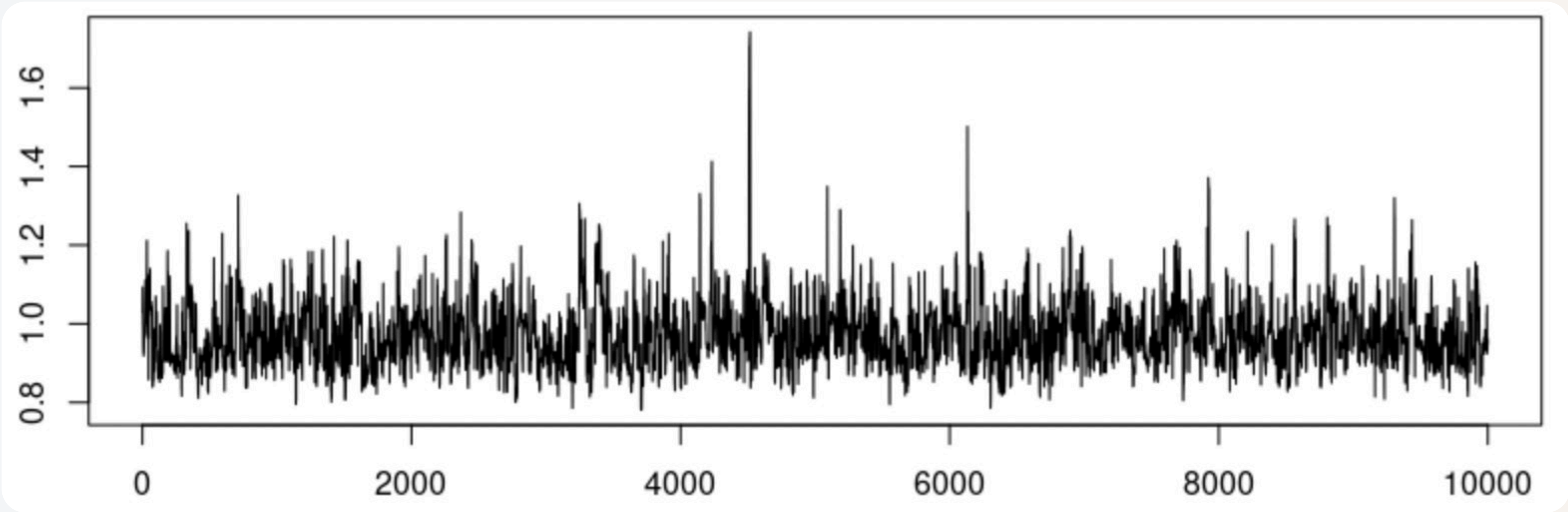
$$\boldsymbol{\mathcal{Z}} = (\boldsymbol{I}_n \otimes \boldsymbol{\mathcal{A}})\boldsymbol{\mathcal{H}} \qquad \boldsymbol{\Sigma} = (\boldsymbol{I}_n \otimes \boldsymbol{\mathcal{A}})\boldsymbol{\mathcal{R}}(\boldsymbol{I}_n \otimes \boldsymbol{\mathcal{A}}^\top)$$

$\boldsymbol{\mathcal{A}}$ is lower triangular with positive diagonal
$\boldsymbol{r}$ is a k-variate MGP with independent margins.
The j-th base covariance is $\boldsymbol{C}_j(\cdot,\cdot) = \sigma_j^2/\phi_j^{2\nu}\rho_{\phi_j,\nu}(\cdot,\cdot)$, where $\rho_{\phi_j,\nu}(\cdot,\cdot)$ is Matérn correlation
$\boldsymbol{\mathcal{H}}$ and $\boldsymbol{\mathcal{R}}$ are derived from kriging relations and allow $\boldsymbol{r}$ to be located on a regular grid

- After running MCMC, postprocess to get original model parameters
- Over **10x efficiency** in estimating σ_j^2, ϕ_j in some settings
- Takes advantage of the multiple **parametrizations** of the hierarchical model with latent effects

Extreme case: Noncentered parametrization vs GriPS. No thinning.

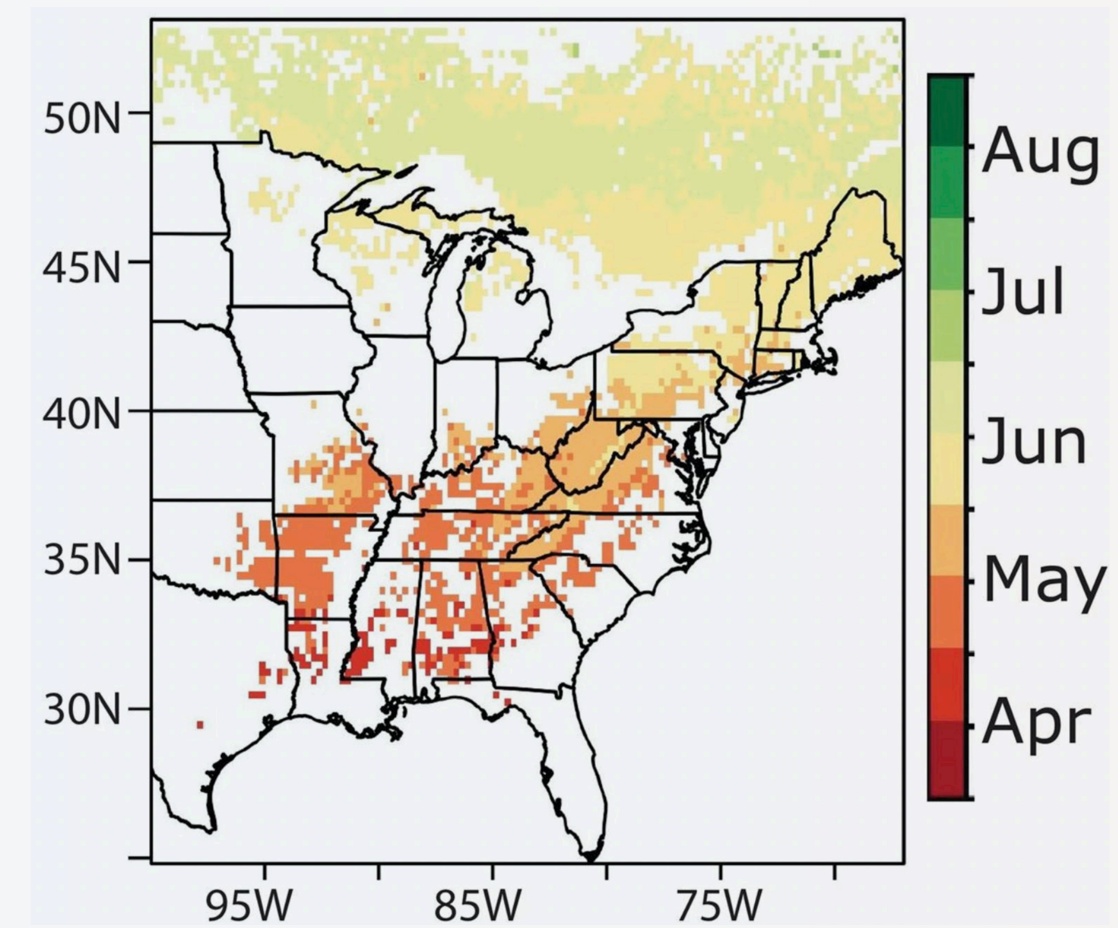




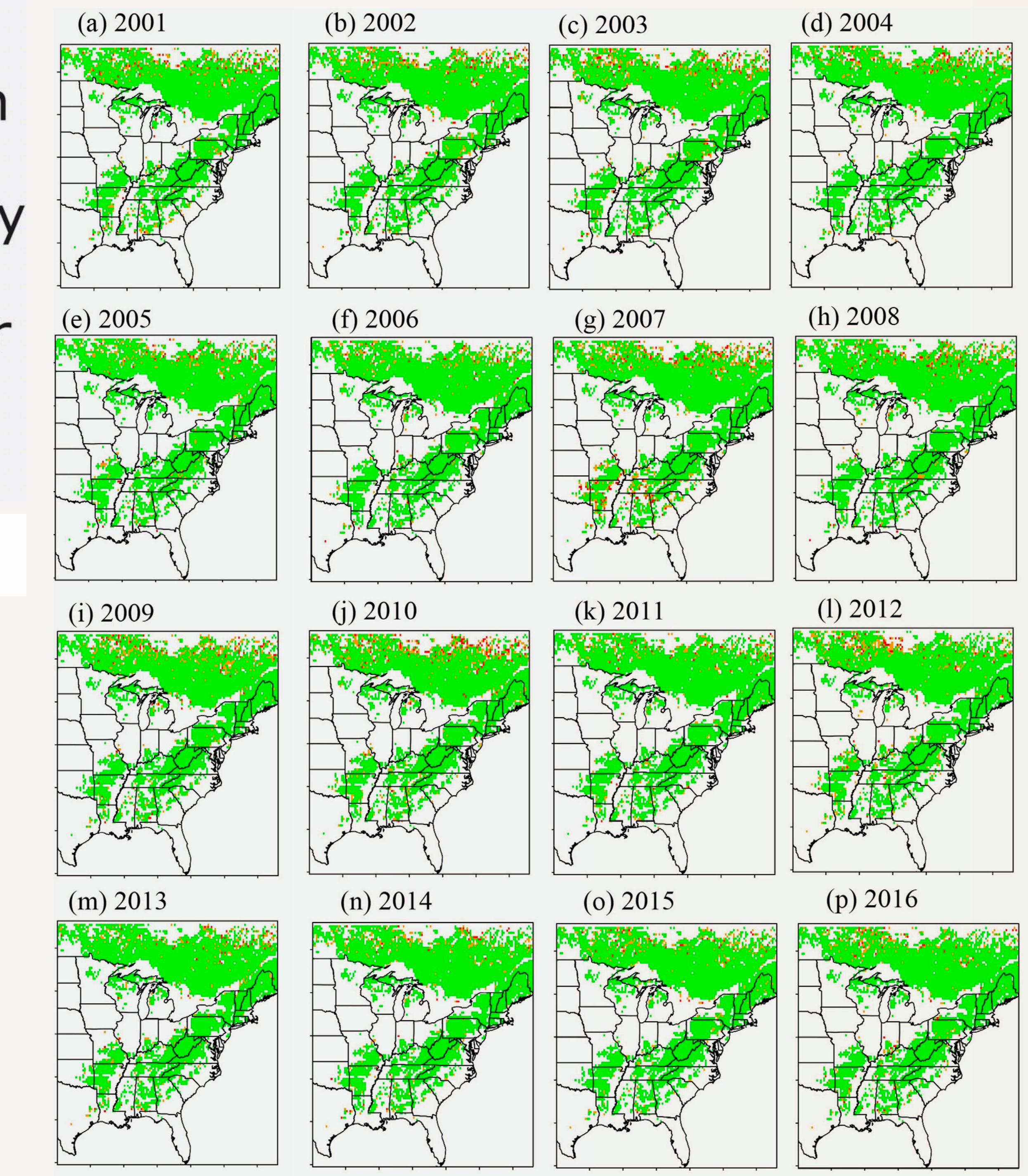
Application: Greenup timing study using MODIS data

N. Neupane, **M. Peruzzi**, L. Ries, A. Arab, S. J. Mayor, J. C. Withey, A. O. Finley (2022)
A novel model to accurately predict continental-scale green-up timing.
International J. of Applied Earth Observation and Geoinformation *108:102747*
https://doi.org/10.1016/j.jag.2022.102747

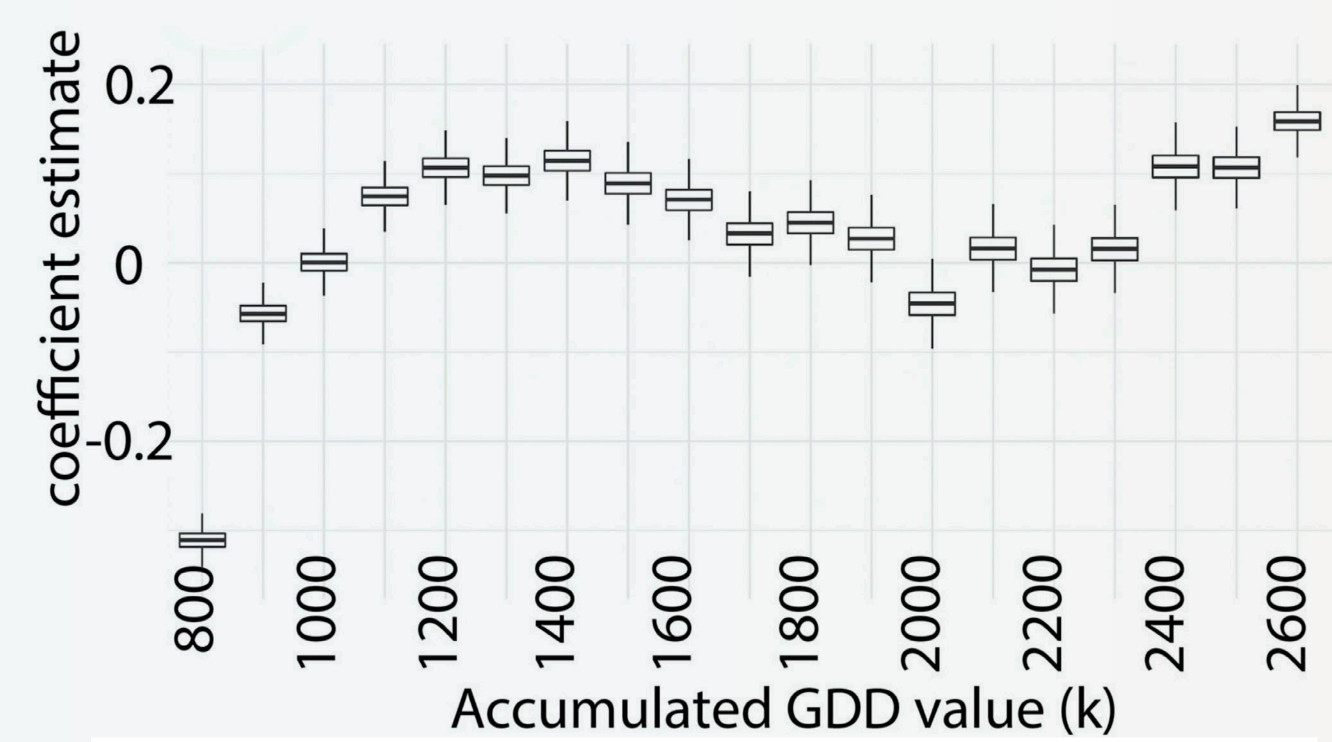
- **Vegetation phenology** study. "Greenness cycles"
- Interest: **when** there is peak greenness
- Vegetation buildup depends on **temperature**
- We consider North American data east of 100W
- With high-res satellite data, can make **continental-level** predictions
- Use MGP + temperature accumulation "speed"
- **1.7 million** space-time locations
- Predict next 2 years
- **Large gains** relative to same model without spatial random effects



Mean peak onset of greenness timing



In-sample residuals



Temperature accumulation effects

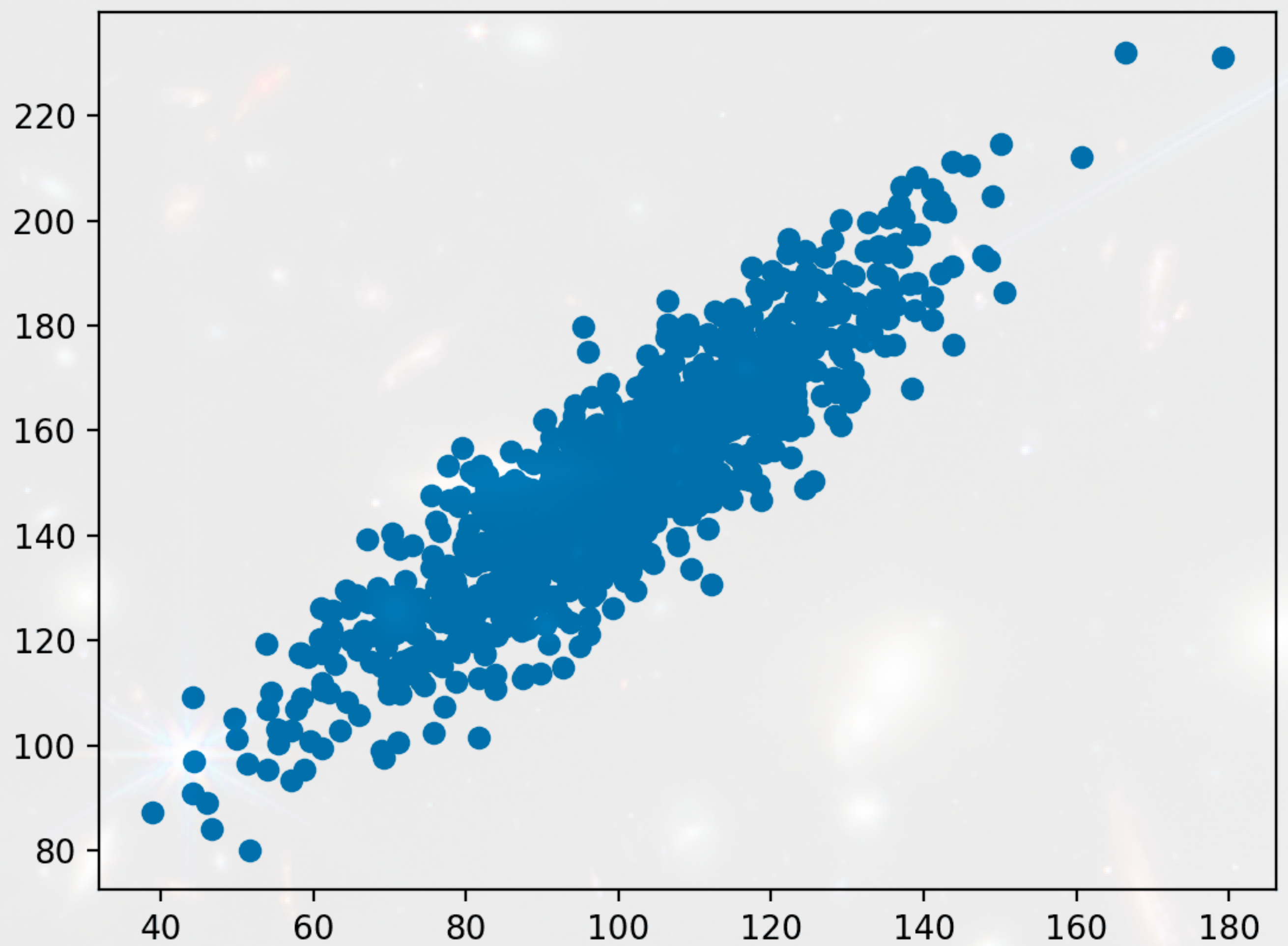
- **Future**: how does climate change affect greenness cycles?

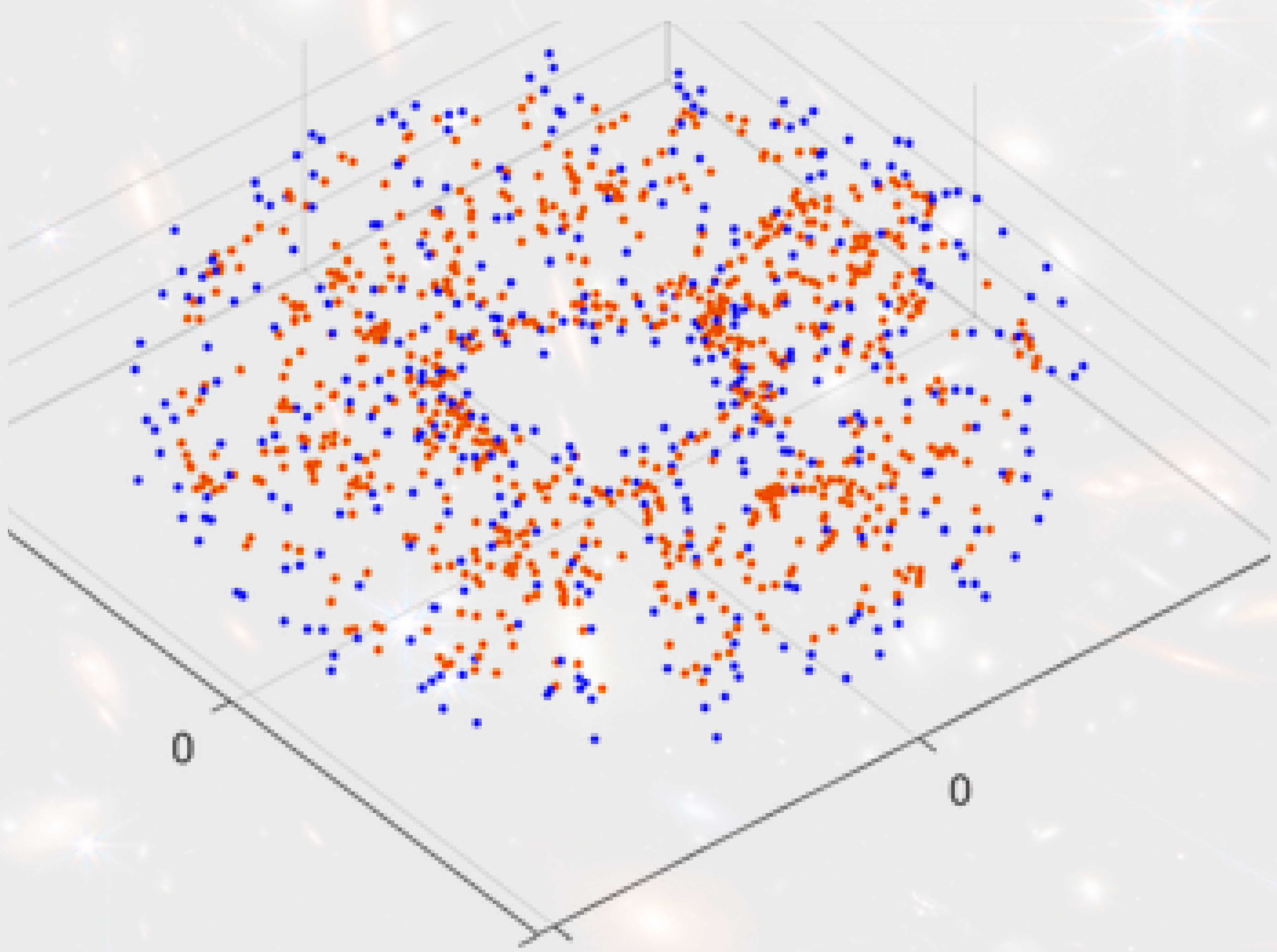
Out-of-sample predictions vs linear regression

Model	Forest	Validation set MAPE	RMSE	95% Covg	Width (days)	2017 prediction MAPE	RMSE	95% Covg	Width (days)	2018 prediction MAPE	RMSE	95% Covg	Width (days)
SVI	no	4.841	8.2384	95.70%	33.8	**6.4687**	**10.1459**	**95.20%**	**40.89**	**5.9737**	**9.156**	**97.98%**	**46.92**
SVI	yes	**4.8174**	**8.2227**	95.32%	**32.7**	6.5488	10.1969	95.71%	43.43	6.1102	9.2742	98.42%	51.68
SLR	yes	10.5255	15.4441	93.35%	60.7	10.7402	17.0653	91.10%	60.7	8.1921	12.9737	95.78%	60.68


Looking ahead:
beyond spatial data

Higher dimensional input spaces





REALITY:

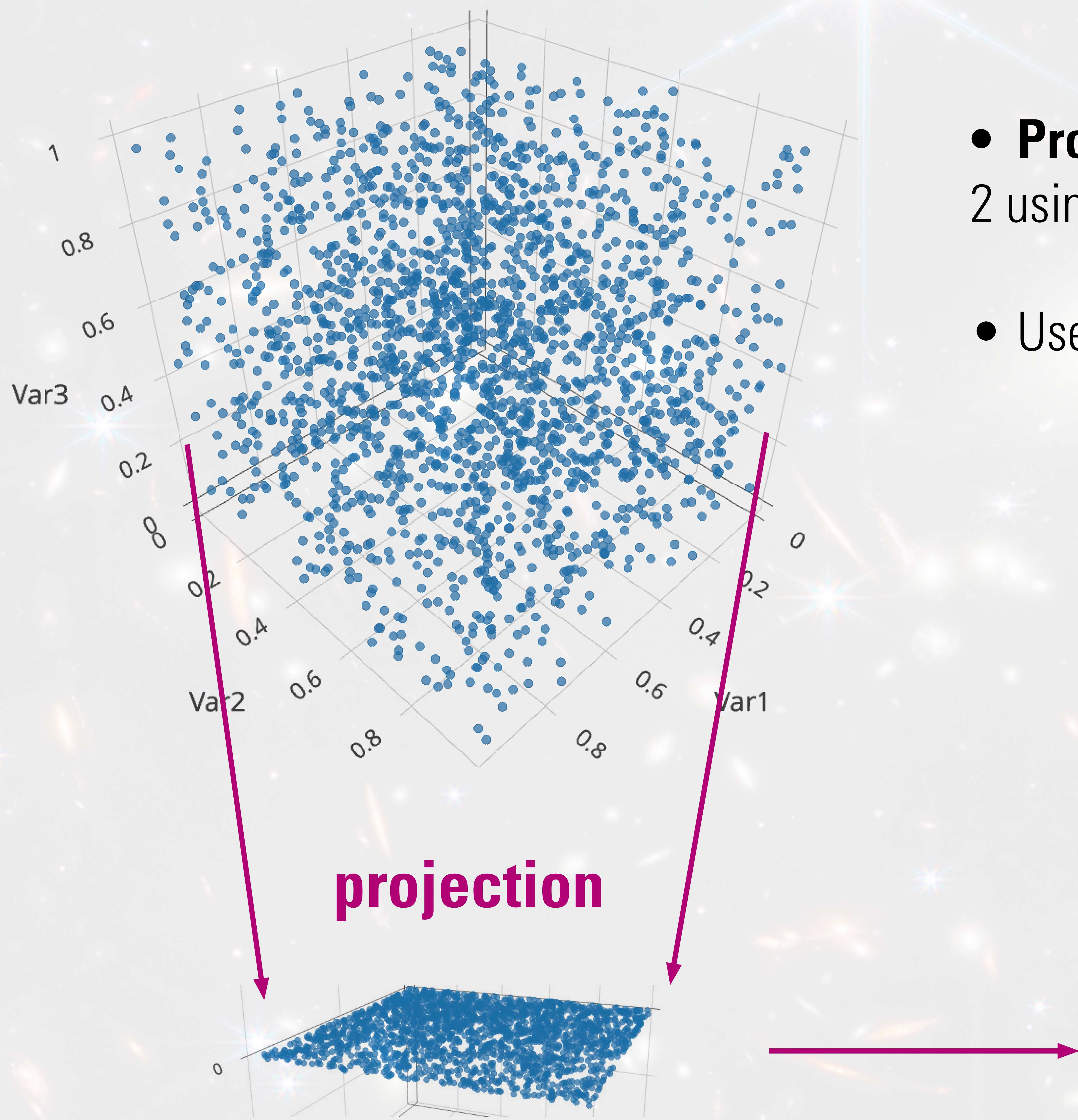
- Correlated inputs (e.g. chemical exposures)
- Extreme case: some inputs on lower dimensional manifold
- Scalability depends on notion of neighbor
- Difficult to find “good” neighbors with high dimension input

IDEA:

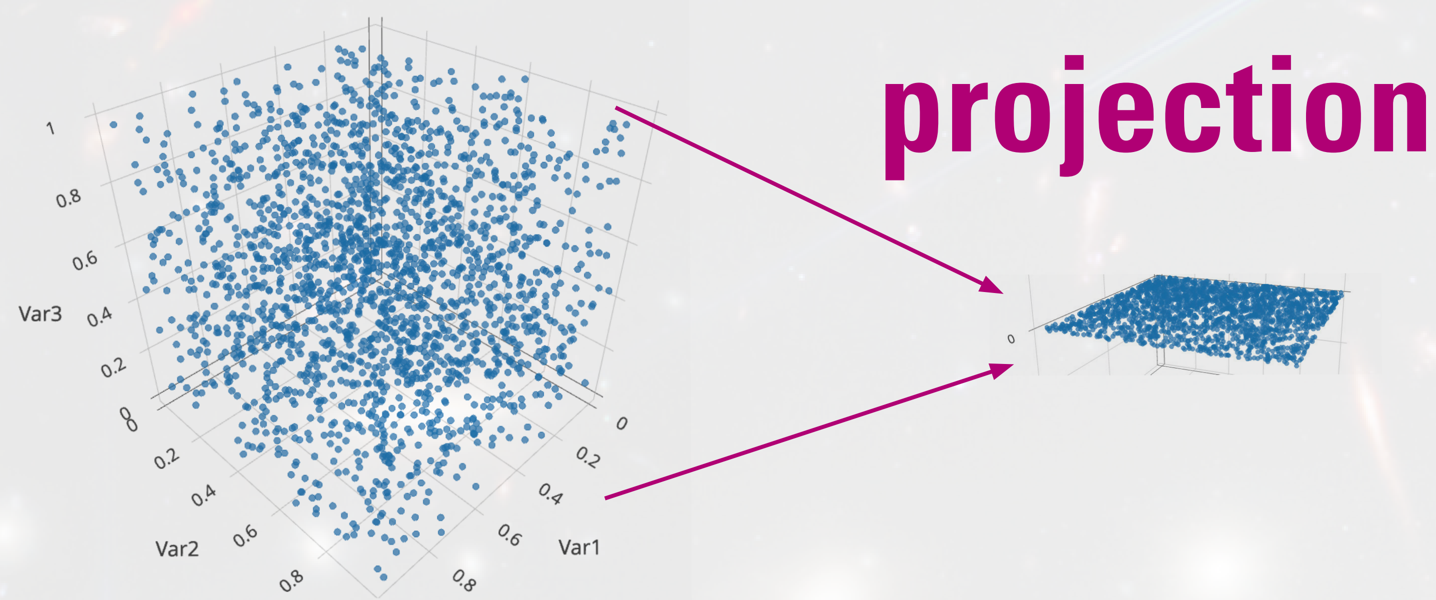
- **Take advantage** of input structure
- **Create** a 2D input space when a natural one does not exist

Graph Machine Regression

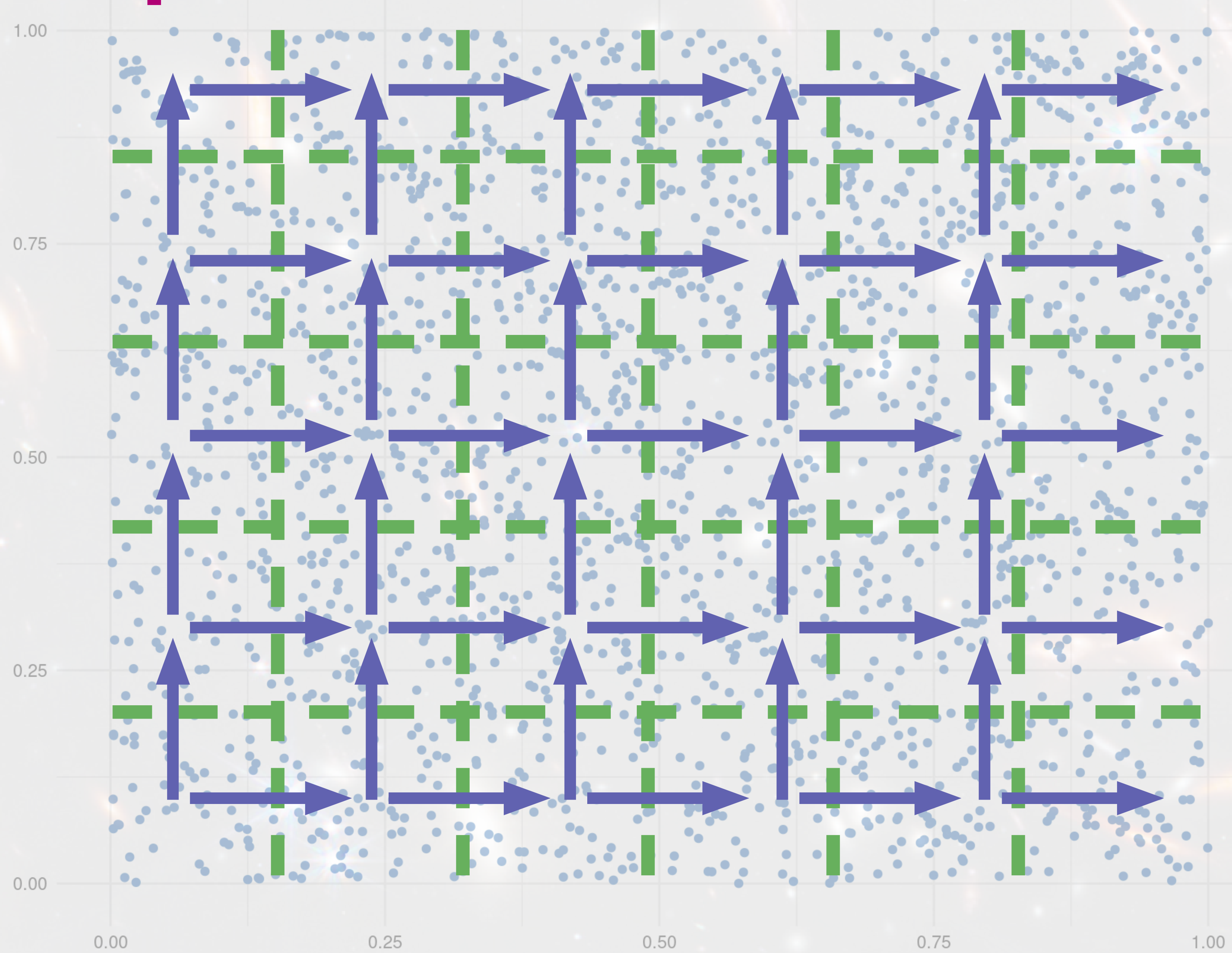
- **Project** original input space onto a new one of dimension 2 using PCA (easy!), Laplacian Eigenmaps?, other?
- Use projected space as a **spatial coordinate system**



Graph Machine Regression



+ partition + DAG

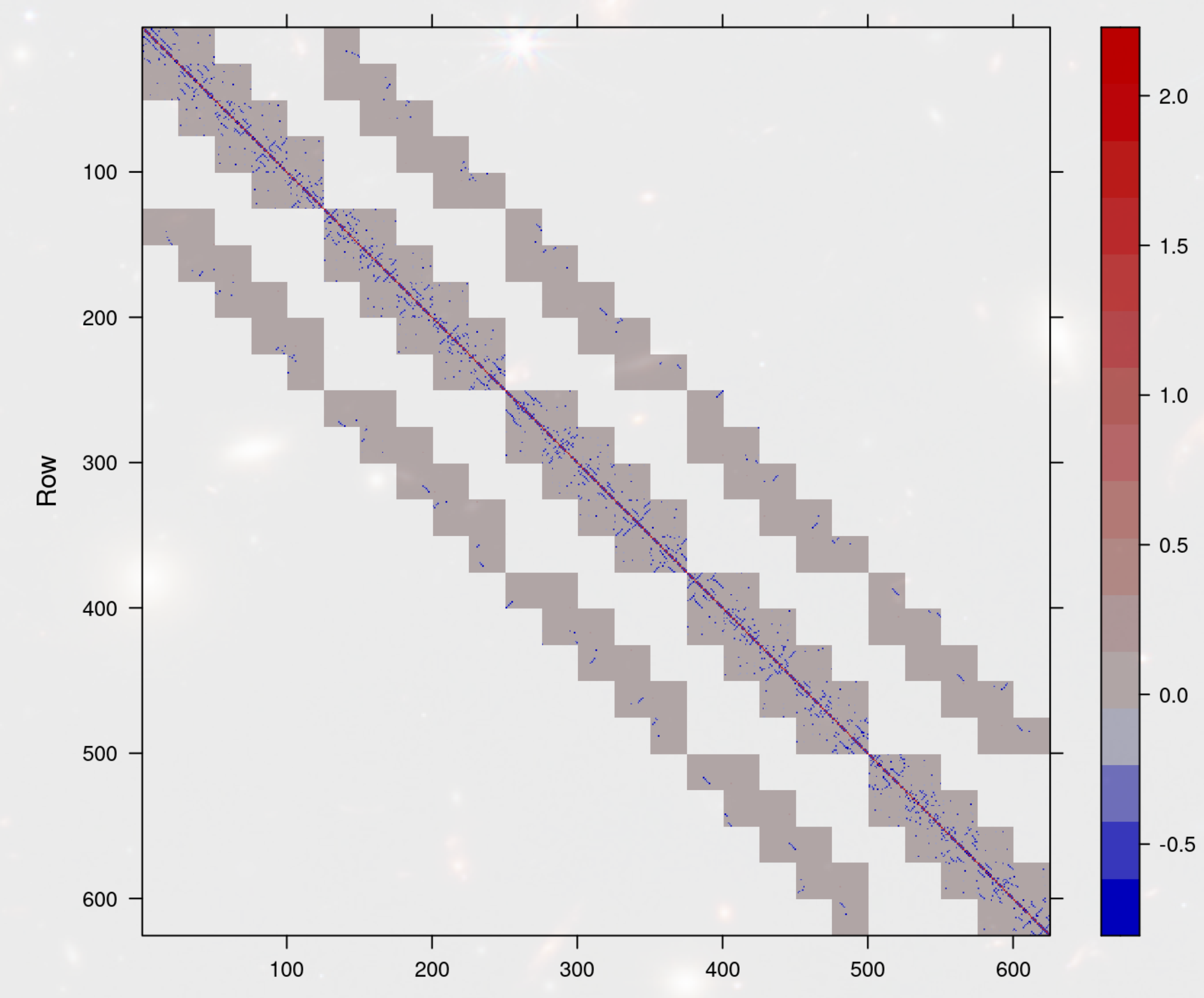


y_i	health outcome/phenotype for subject *i*
$\boldsymbol{x}_i$	vector of predictors, dimension *p*
$f(\cdot)$	unknown function

$$y_i = f(\boldsymbol{x}_i) + \varepsilon_i \qquad \begin{array}{l} \varepsilon_i \overset{iid}{\sim} N(0, \sigma^2) \\ i = 1, \dots, n \end{array}$$

$$p_{\text{meshed}}(\boldsymbol{f}) \propto |\widetilde{\boldsymbol{K}}|^{-\frac{1}{2}} \exp\left\{ -\frac{1}{2} \boldsymbol{f}^\top \widetilde{\boldsymbol{K}}^{-1} \boldsymbol{f} \right\}$$

the **precision matrix is sparse** with pattern:

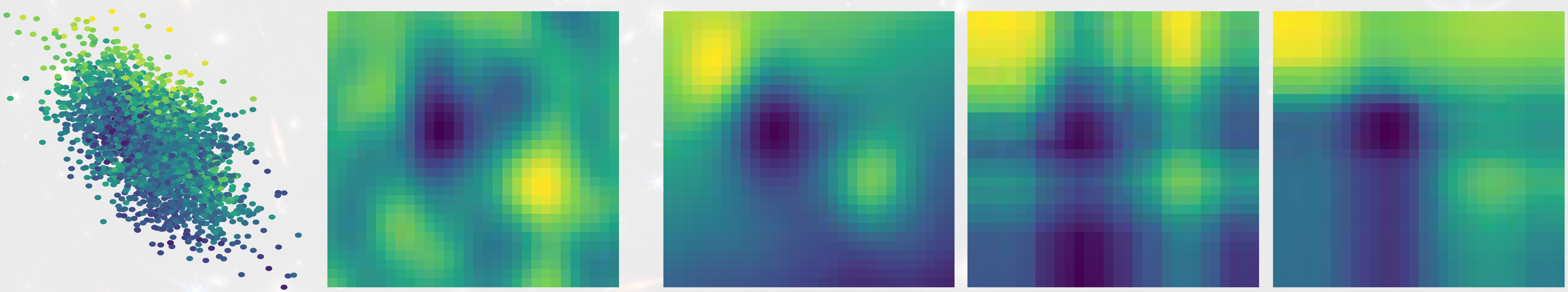


Graph Machine Regression

- 1 outcome
- data size ~3000
- **15 correlated inputs**

using new package `gramar` (github.com/mkln/gramar)

- for univariate outcomes
- uses a collapsed sampler
- **0.017** seconds/iteration (compare with BKMR implementing full GP: 0.838 seconds/iteration, **50x slower**)

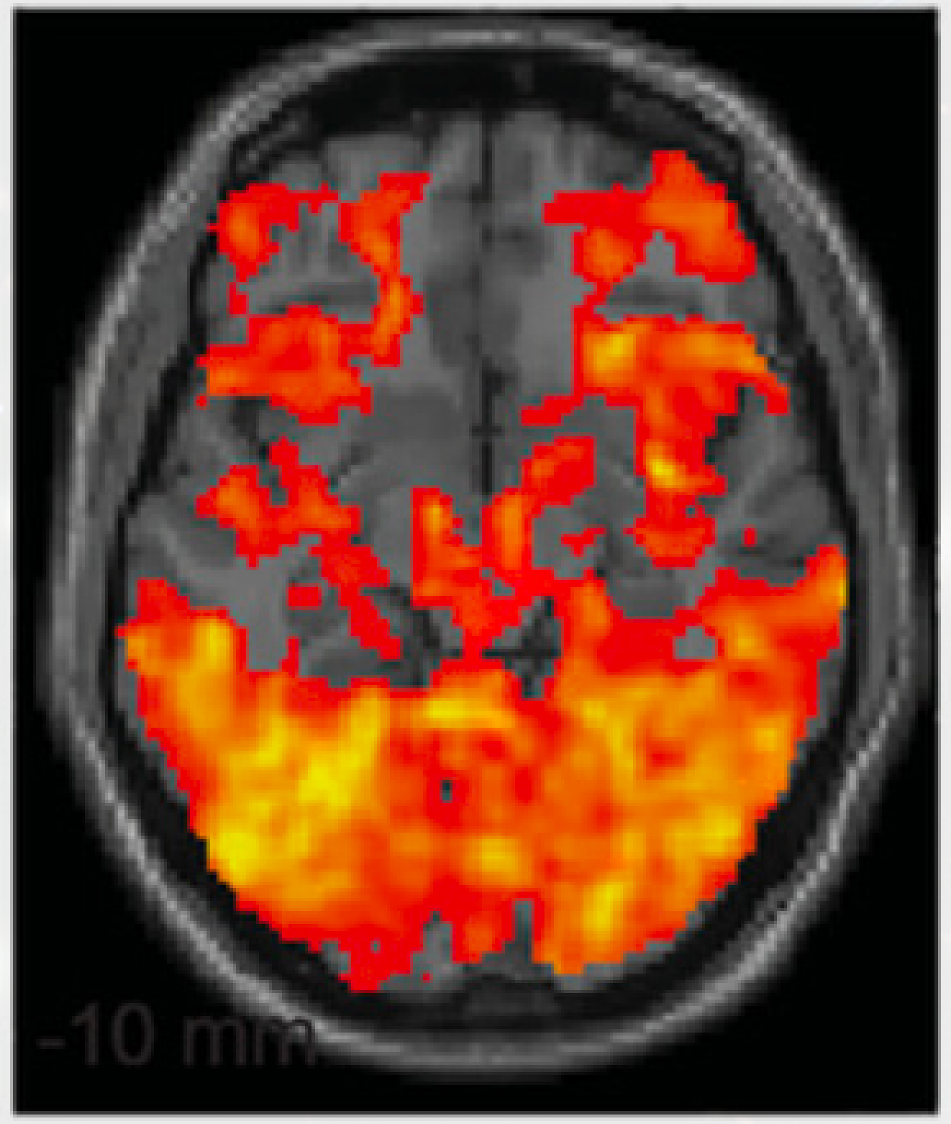


Observed data | Latent X1-X2 surface | GRAMAR | RandomForest | BART

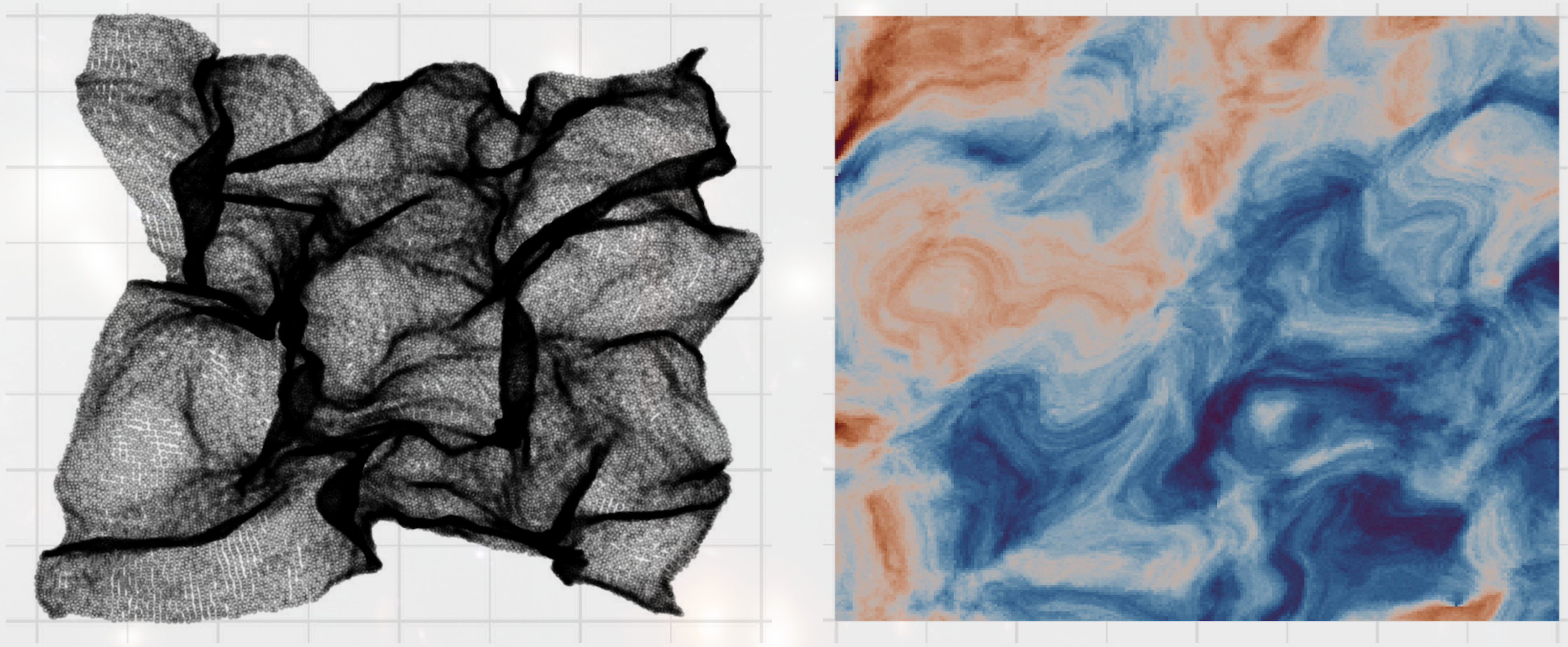
Recovered surfaces

Some future directions

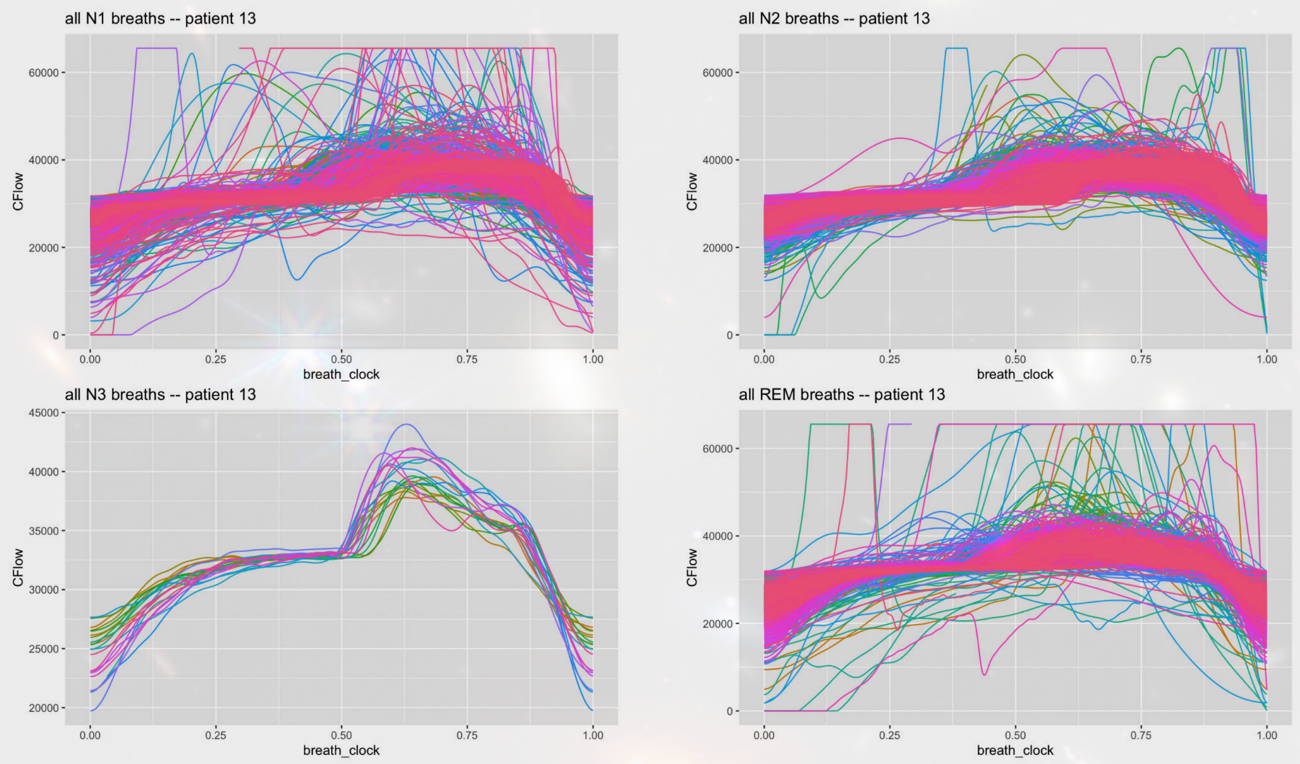
Other kinds of images



Nonstationary data via complex warping



Sleep studies



Activity tracking