Inside-out cross-covariance
for spatial multivariate data

Michele Peruzzi
peruzzi@umich.edu
SCHOOL OF PUBLIC HEALTH
BIOSTATISTICS
UNIVERSITY OF MICHIGAN

Introduction: spatial multivariate data

Spatial multivariate data

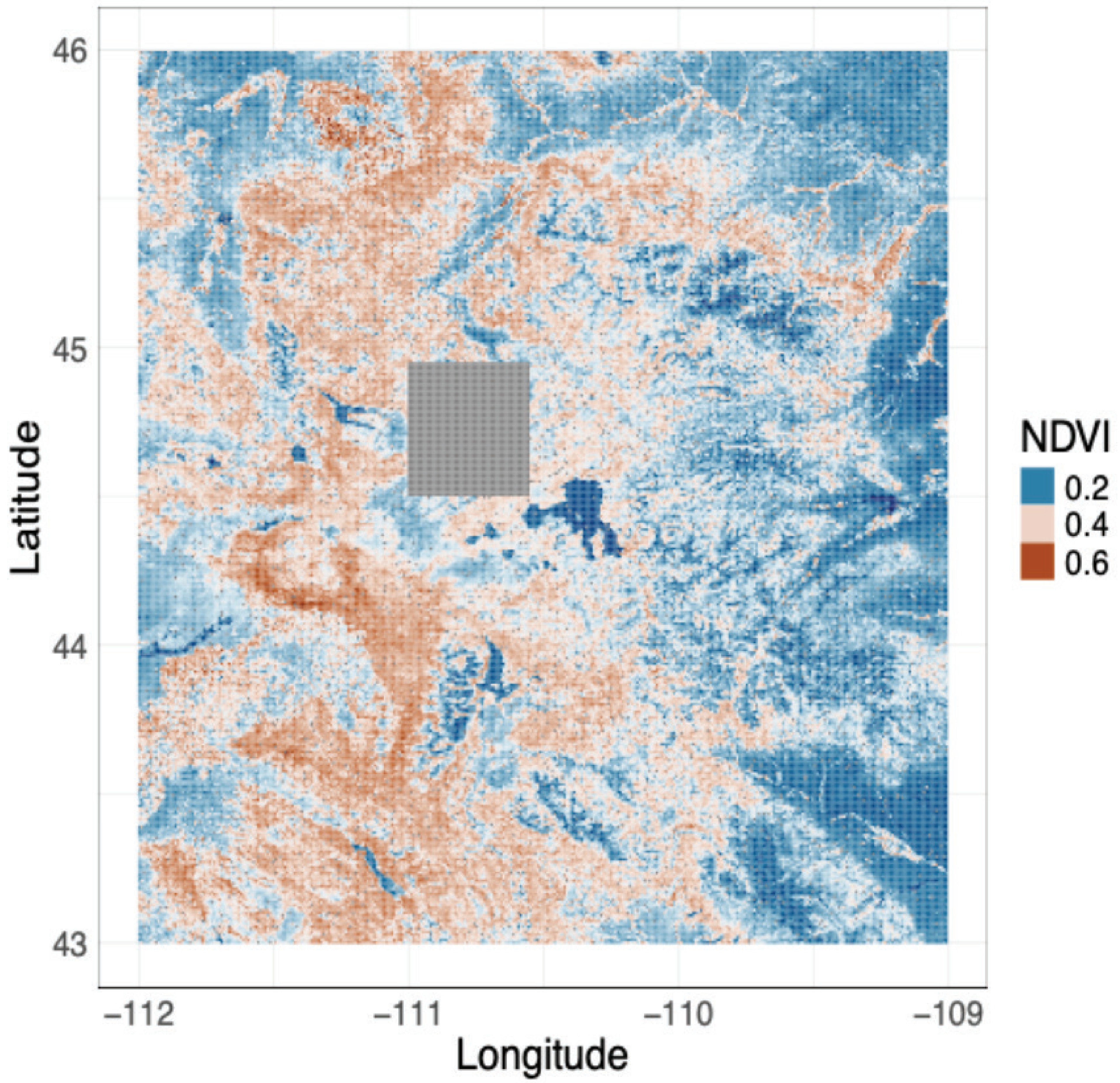
- at each location s we observe a random vector of dimension q
- spatial dependence and cross-variable dependence

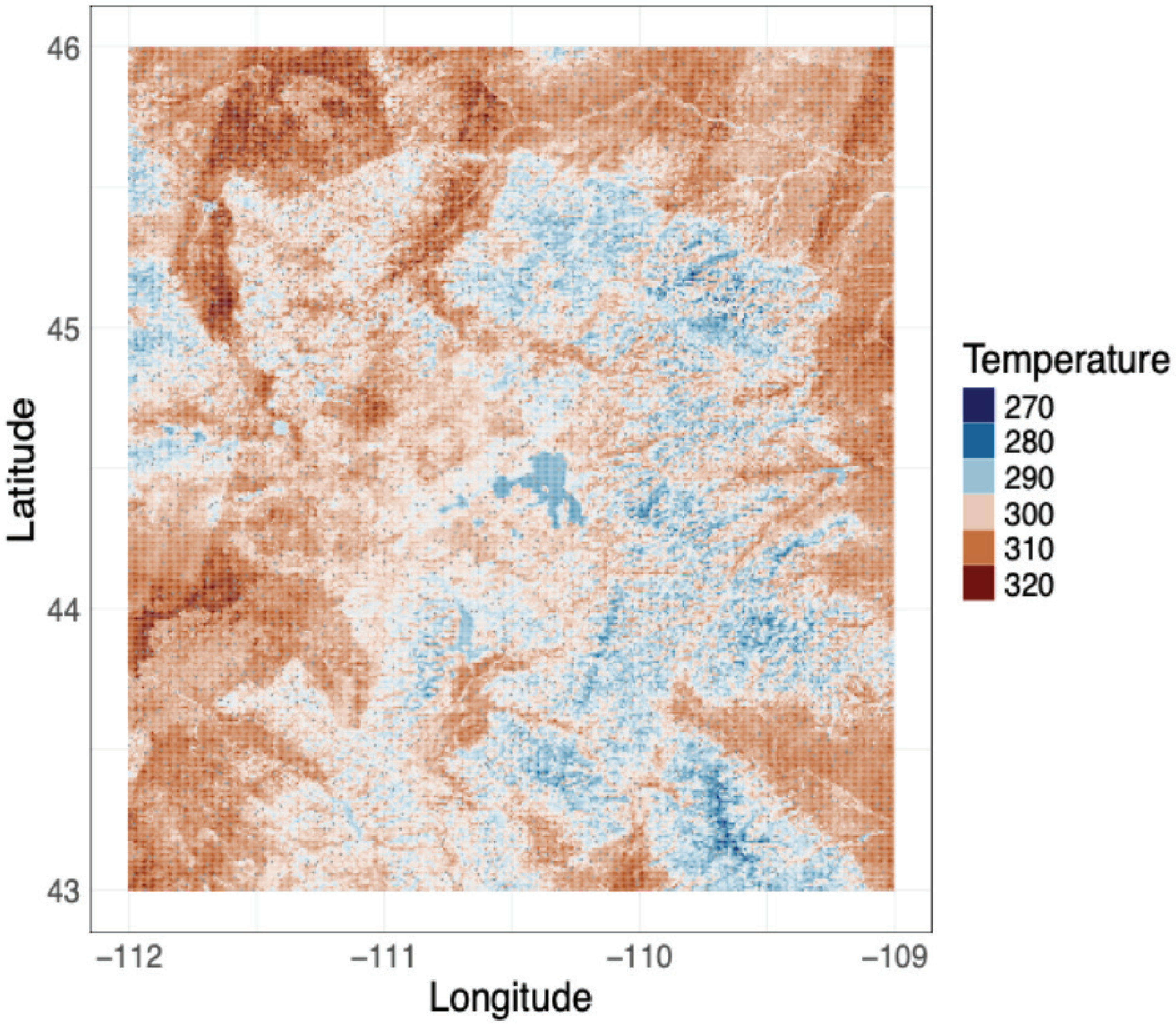
Examples

- community ecology
- remote sensing
- climate data
- multiplexed imaging data of tissue biopsy, “omics” data

Introduction: spatial multivariate data

Example: satellite imaging



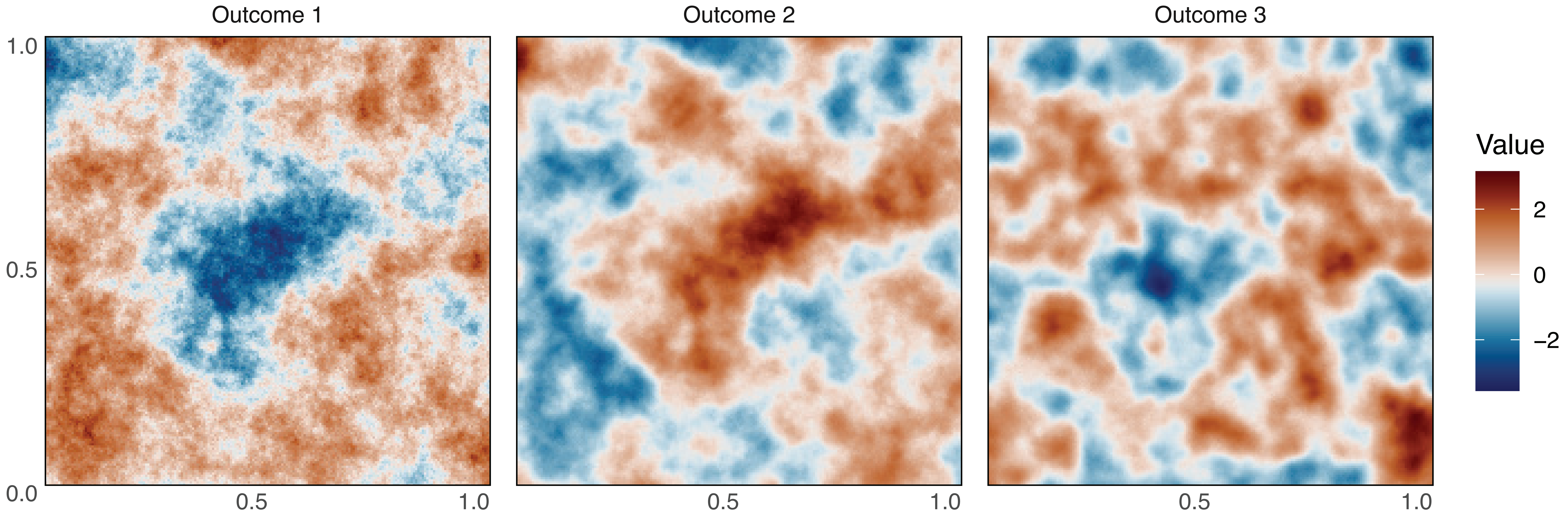


Introduction: spatial multivariate data

Example: microimmunofluorescence of tissue biopsies

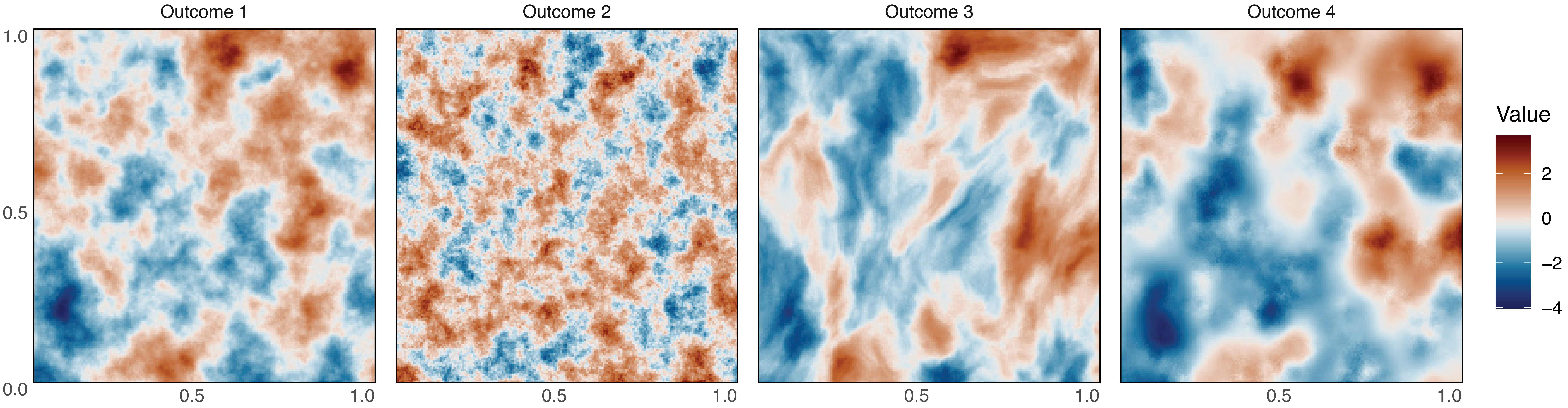
Introduction: example

Example: simulated data



Introduction: example

Example: simulated data but more complicated



Introduction: covariance modeling for GPs

Spatial multivariate data

- at each location s we observe a random vector of dimension q

$$\mathbf{Y}(s) = \begin{bmatrix} y_1(s) \\ \vdots \\ y_q(s) \end{bmatrix}$$

- Gaussian assumption on the process $\{\boldsymbol{Y}(s) : s \in \mathcal{D}\}$ leads to **multivariate Gaussian Process (GP)** model
- the **cross-covariance matrix function** fully characterizes a (zero-mean) multivariate GP (Genton & Kleiber 2015):

$$C(\cdot,\cdot) : \Re^d \times \Re^d \to \mathcal{M}$$

  where $\mathcal{M}$ is the space of all positive semidefinite matrices of size $q \times q$
- this is our covariance model: we are modeling $\mathrm{cov}(\boldsymbol{Y}(s), \boldsymbol{Y}(s')) = C(s, s')$
- via C we model all combinations of $\mathrm{cov}(y_r(s), y_c(s'))$ for $r, c = 1, \ldots, q$

Introduction: covariance modeling for GPs

Cross-covariance matrix function

- the **cross-covariance matrix function** fully characterizes a (zero-mean) multivariate GP (Genton & Kleiber 2015):

$$C(\cdot, \cdot) : \Re^d \times \Re^d \to \mathcal{M}$$

where $\mathcal{M}$ is the space of all positive semidefinite matrices of size $q \times q$

- via C we model all combinations of $\text{cov}(y_r(s), y_c(s'))$ for $r, c = 1, \ldots, q$

Desired features

- **parsimony** for large q
- computational **tractability** via exploitable structure of sample covariance
- **easy-to-interpret** parameters

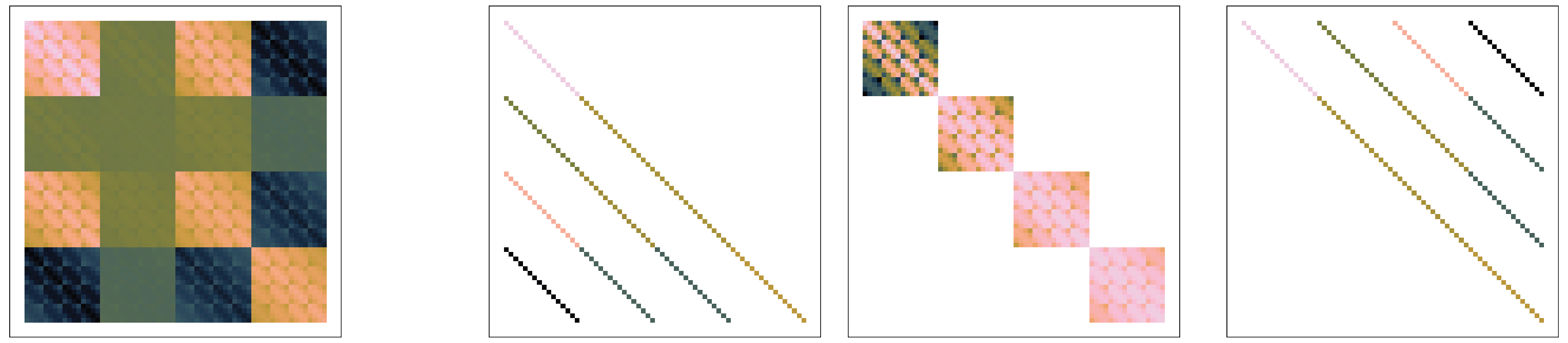
Linear model of coregionalization / spatial factor model (LMC)

- introduce a matrix A of dimension $q \times k$ with elements $a_{rc}, r = 1, \dots, q; c = 1, \dots, k$
- introduce k correlation functions $\rho_j(\cdot, \cdot)$
- LMC models all covariances as linear combinations:

$$\text{cov}\{y_r(s), y_c(s')\} = \sum_{j=1}^{k} a_{rj} a_{jc} \rho_j(s, s')$$

- suppose S is the set of observed locations. the sample covariance for the $nq \times 1$ vector is

$$\text{cov}\{\boldsymbol{y}\} = (\boldsymbol{A} \otimes \boldsymbol{I}_n)\{\oplus \boldsymbol{R}_j\}(\boldsymbol{A}^\top \otimes \boldsymbol{I}_n) \qquad \boldsymbol{R}_j = \rho_j(S, S)$$



- **parsimonious** for large q
- computationally **tractable** via exploitable structure of sample covariance
- **easy-to-interpret** parameters???

LMC pros

LMC is the **most popular model** for multivariate spatial data:

- extend for some form of nonstationarity (Gelfand et al. 2004)
- spatially-varying regression coefficients, typically via separability assumptions (Gelfand et al. 2003 and Reich et al. 2010)
- space-time data (Berrocal et al. 2010, De Iaco et al. 2019)
- used for latent process models for non-Gaussian data (Peruzzi & Dunson 2024)
- dimension reduction tool if q is large (Taylor-Rodriguez et al. 2019, Zhang & Banerjee 2022)
- popular in many fields
  (see, e.g., Teh et al. 2005, Finley et al. 2008, Álvarez & Lawrence 2011, Fricker et al. 2013, Moreno-Muñoz et al. 2018, Liu et al. 2022, Townes & Engelhardt 2023)
- Software packages typically use LMCs
  (Pebesma 2004, Finley et al. 2015, Tikhonov et al. 2020, Finazzi & Fassò 2014, Krainski et al. 2019, Peruzzi 2022)

LMC cons

LMC has a few important drawbacks:

- cannot model outcomes with **different smoothness**
- parameters of $\rho_j(\cdot)$ are **not directly interpretable**
- **specifying priors** is difficult
- cross covariances $C_{rc}(\cdot), r \neq c$ are “as important as” marginal covariances $C_{rr}(\cdot)$
- difficult to introduce nugget effects in the $k = q$ case
- poorly understood infill asymptotics
- lack of easy pipeline for introducing **outcome-specific features**

LMC alternatives

Multivariate Matérn (Gneiting 2010):

- each $C_{rc}(\cdot), r \neq c$ and $C_{rr}(\cdot)$ is Matérn
- **validity conditions** restrict parameter space (Apanasovich & Genton 2012, Emery et all 2022)
- need more flexible extensions? validity conditions become a huge burden
- lack of structure in sample covariance matrices
- most useful for the small q regime

Latent dimensions (Apanasovich & Genton 2010):

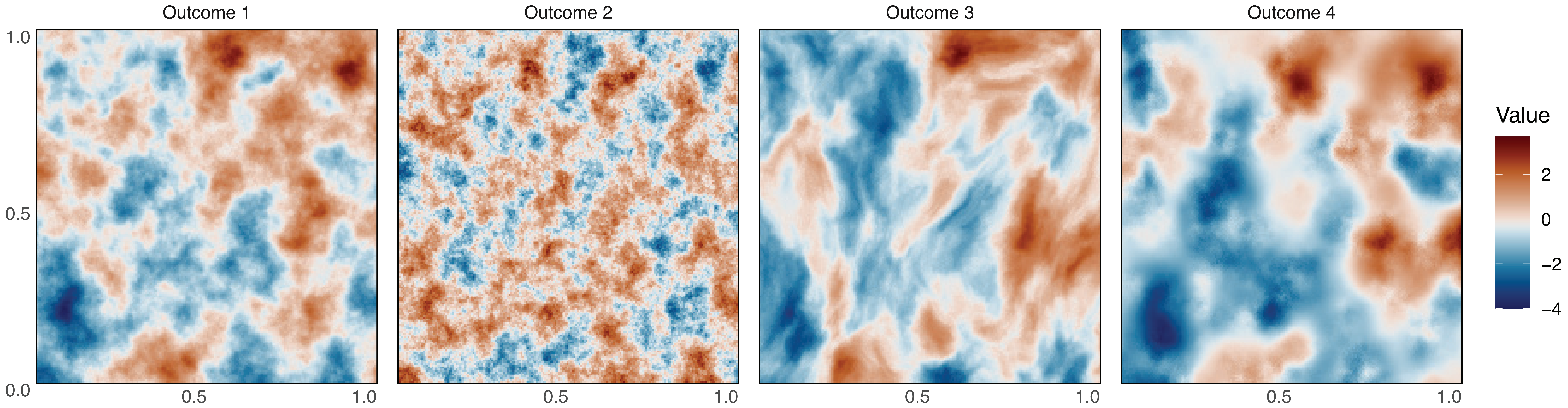
- elegant construction
- lack of **structure** in sample covariance matrices
- most useful for the small q regime

Convolution methods (Gaspari & Cohn 1999, Majumdar & Gelfand 2007):

- **computationally prohibitive**
- most useful for the small q regime

What covariance model for this simulated example?

- 4 spatially indexed variables with different degrees of **spatial cross-correlation**
- each variable has specific features
- simulated at a very large number of spatial locations



marginally stationary | unequal smoothness | nonstationary via warp | spatially-varying sill

Sampling spatial data

Univariate case:

- choose sampling locations S
- sample $\boldsymbol{u} \sim N(\boldsymbol{0}, \boldsymbol{I}_n)$
- compute $\boldsymbol{L} = \texttt{chol}\{\boldsymbol{R}\}$ where $\boldsymbol{R} = \rho(S, S)$
- finally, $\boldsymbol{y} = \sigma \boldsymbol{L} \boldsymbol{u}$

LMC:

- choose sampling locations S
- sample $\boldsymbol{u}_j \sim N(\boldsymbol{0}, \boldsymbol{I}_n), j = 1, \ldots, k$
- compute $\boldsymbol{L}_j = \texttt{chol}\{\boldsymbol{R}_j\}$ where $\boldsymbol{R}_j = \rho_j(S, S)$
- compute $\boldsymbol{v}_j = \boldsymbol{L}_j \boldsymbol{u}_j$ and stack into matrix $\boldsymbol{V}$
- finally, $\boldsymbol{Y} = \boldsymbol{V} \boldsymbol{A}^\top$ where $\boldsymbol{A}$ is the factor loadings matrix

Sampling spatial data

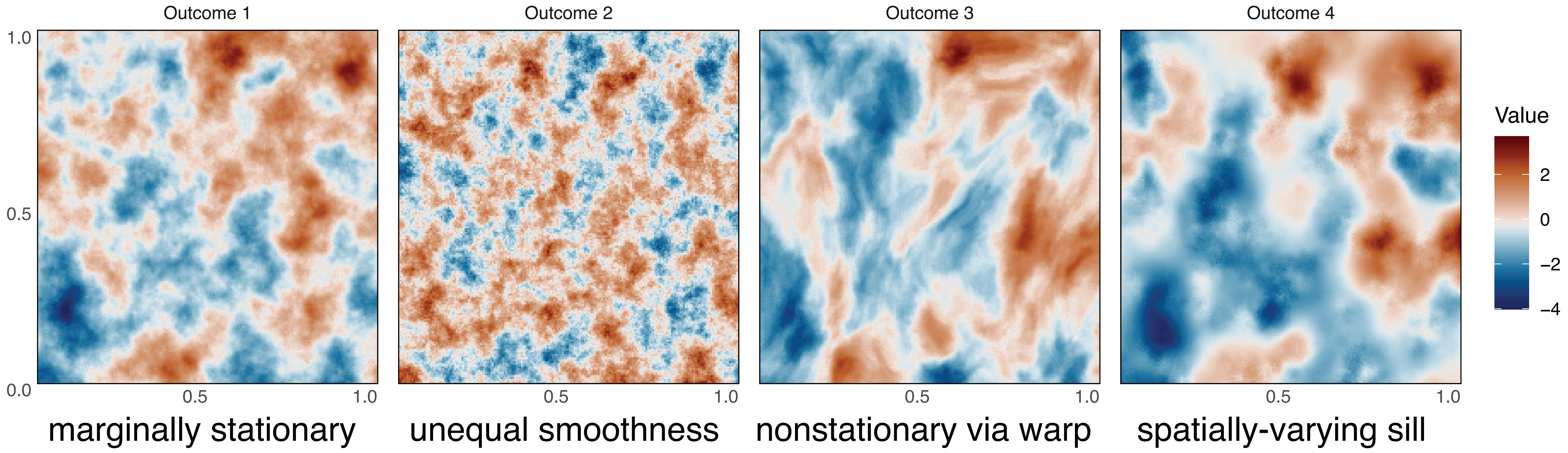
Univariate case:

- choose sampling locations S
- sample $\boldsymbol{u} \sim N(\boldsymbol{0}, \boldsymbol{I}_n)$ — generate iid data
- compute $\boldsymbol{L} = \texttt{chol}\{\boldsymbol{R}\}$ where $\boldsymbol{R} = \rho(S, S)$ — introduce spatial correlation
- finally, $\boldsymbol{y} = \sigma \boldsymbol{L}\boldsymbol{u}$

LMC:

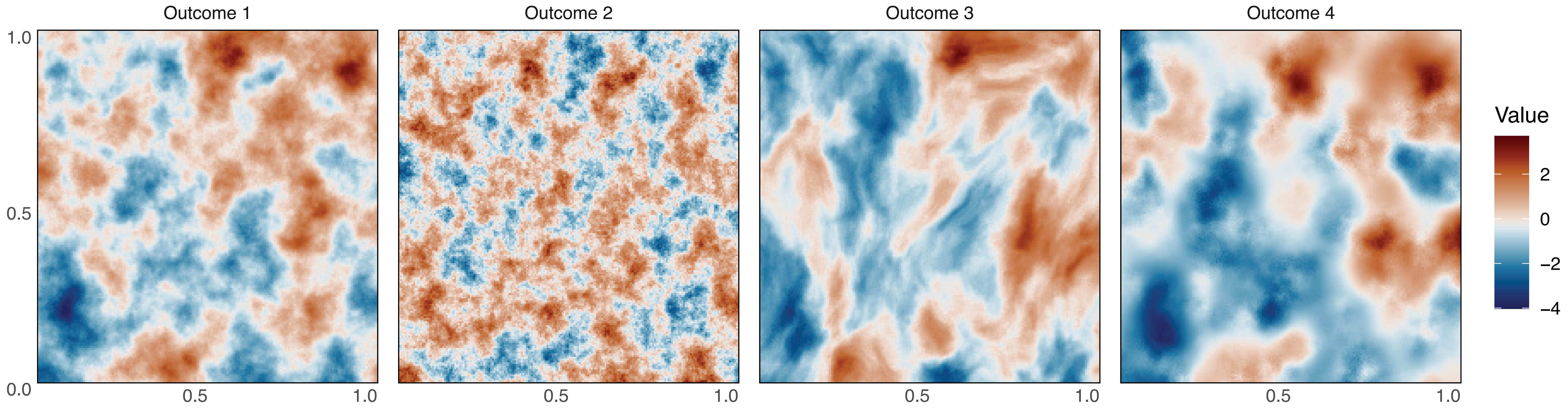
- choose sampling locations S
- sample $\boldsymbol{u}_j \sim N(\boldsymbol{0}, \boldsymbol{I}_n), j = 1, \ldots, k$ — generate iid data
- compute $\boldsymbol{L}_j = \texttt{chol}\{\boldsymbol{R}_j\}$ where $\boldsymbol{R}_j = \rho_j(S, S)$
- compute $\boldsymbol{v}_j = \boldsymbol{L}_j \boldsymbol{u}_j$ and stack into matrix $\boldsymbol{V}$ — introduce spatial correlation
- finally, $\boldsymbol{Y} = \boldsymbol{V}\boldsymbol{A}^\top$ where $\boldsymbol{A}$ is the factor loadings matrix — introduce cross-correlation

Sampling the example data



- choose sampling locations S
- sample $\boldsymbol{u}_j \sim N(\mathbf{0}, \boldsymbol{I}_n), j = 1, \ldots, k$ and stack into $\boldsymbol{U}$
- compute $\boldsymbol{V} = \boldsymbol{U}\boldsymbol{A}^\top$
- compute $\boldsymbol{L}_j = \texttt{chol}\{\boldsymbol{R}_j\}$ where $\boldsymbol{R}_j = \rho_j(S, S)$
- finally, $\boldsymbol{y}_j = \boldsymbol{L}_j \boldsymbol{v}_j$ (by column) and stack into $\boldsymbol{Y}$

Sampling the example data



marginally stationary | unequal smoothness | nonstationary via warp | spatially-varying sill

- choose sampling locations S
- sample $\boldsymbol{u}_j \sim N(\boldsymbol{0}, \boldsymbol{I}_n), j = 1, \ldots, k$ and stack into $\boldsymbol{U}$ — generate iid data
- compute $\boldsymbol{V} = \boldsymbol{U}\boldsymbol{A}^\top$ — introduce cross-correlation
- compute $\boldsymbol{L}_j = \texttt{chol}\{\boldsymbol{R}_j\}$ where $\boldsymbol{R}_j = \rho_j(S, S)$
- finally, $\boldsymbol{y}_j = \boldsymbol{L}_j \boldsymbol{v}_j$ (by column) and stack into $\boldsymbol{Y}$ — introduce spatial correlation

How is this different from a LMC?

We are **inverting the order of operations**:

first, cross-variable dependence. second, spatial dependence

- choose sampling locations S
- sample $\boldsymbol{u}_j \sim N(\boldsymbol{0}, \boldsymbol{I}_n), j = 1, \ldots, k$ and stack into $\boldsymbol{U}$
- compute $\boldsymbol{V} = \boldsymbol{U}\boldsymbol{A}^\top$
- compute $\boldsymbol{L}_j = \texttt{chol}\{\boldsymbol{R}_j\}$ where $\boldsymbol{R}_j = \rho_j(S, S)$
- finally, $\boldsymbol{y}_j = \boldsymbol{L}_j \boldsymbol{v}_j$ (by column) and stack into $\boldsymbol{Y}$

LMC:

- choose sampling locations S
- sample $\boldsymbol{u}_j \sim N(\boldsymbol{0}, \boldsymbol{I}_n), j = 1, \ldots, k$
- compute $\boldsymbol{L}_j = \texttt{chol}\{\boldsymbol{R}_j\}$ where $\boldsymbol{R}_j = \rho_j(S, S)$
- compute $\boldsymbol{v}_j = \boldsymbol{L}_j \boldsymbol{u}_j$ and stack into matrix $\boldsymbol{V}$
- finally, $\boldsymbol{Y} = \boldsymbol{V}\boldsymbol{A}^\top$ where $\boldsymbol{A}$ is the factor loadings matrix

How is this different from a LMC?

We are **inverting the order of operations**:
first, cross-variable dependence. second, spatial dependence

Inside-out construction:

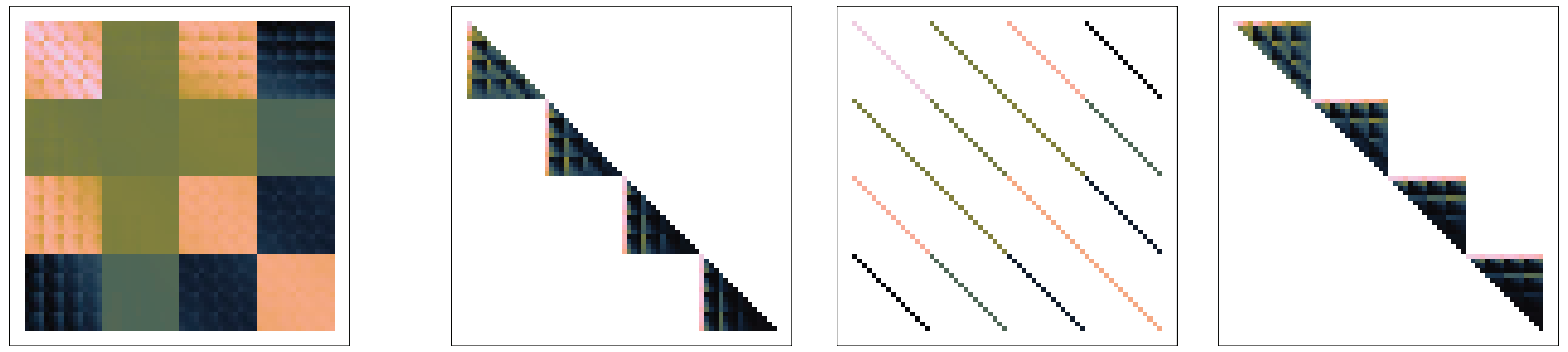
- choose sampling locations S
- sample $\boldsymbol{u}_j \sim N(\boldsymbol{0}, \boldsymbol{I}_n), j = 1, \dots, k$ and stack into $\boldsymbol{U}$
- compute $\boldsymbol{V} = \boldsymbol{U}\boldsymbol{A}^\top$
- compute $\boldsymbol{L}_j = \texttt{chol}\{\boldsymbol{R}_j\}$ where $\boldsymbol{R}_j = \rho_j(S, S)$
- finally, $\boldsymbol{y}_j = \boldsymbol{L}_j \boldsymbol{v}_j$ (by column) and stack into $\boldsymbol{Y}$

LMC:

- choose sampling locations S
- sample $\boldsymbol{u}_j \sim N(\boldsymbol{0}, \boldsymbol{I}_n), j = 1, \dots, k$
- compute $\boldsymbol{L}_j = \texttt{chol}\{\boldsymbol{R}_j\}$ where $\boldsymbol{R}_j = \rho_j(S, S)$
- compute $\boldsymbol{v}_j = \boldsymbol{L}_j \boldsymbol{u}_j$ and stack into matrix $\boldsymbol{V}$
- finally, $\boldsymbol{Y} = \boldsymbol{V}\boldsymbol{A}^\top$ where $\boldsymbol{A}$ is the factor loadings matrix

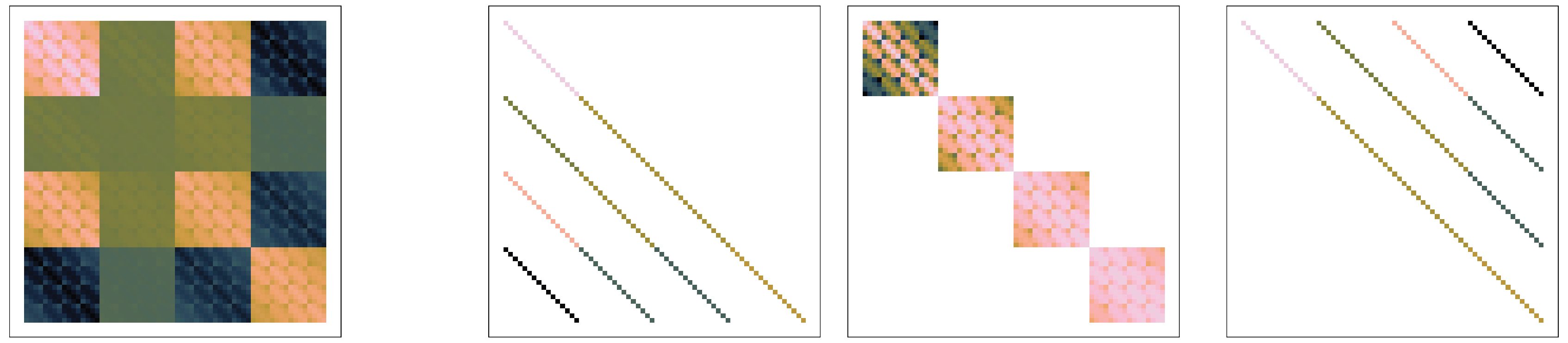
$\boldsymbol{y} = \mathrm{vec}\{\boldsymbol{Y}\}$

$$\mathrm{cov}\{\boldsymbol{y}\} = \{\oplus \boldsymbol{L}_j\}(\boldsymbol{\Sigma} \otimes \boldsymbol{I}_n)\{\oplus \boldsymbol{L}_j^\top\}$$



$$\boldsymbol{\Sigma} = \boldsymbol{A}\boldsymbol{A}^\top$$

$$\mathrm{cov}\{\boldsymbol{y}\} = (\boldsymbol{A} \otimes \boldsymbol{I}_n)\{\oplus \boldsymbol{R}_j\}(\boldsymbol{A}^\top \otimes \boldsymbol{I}_n)$$



$$\boldsymbol{R}_j = \boldsymbol{L}_j \boldsymbol{L}_j^\top$$

...but can this lead to a valid cross-covariance matrix function?

...but can this lead to a valid cross-covariance matrix function?

YES!

Inside-Out Cross-covariance (IOX)

Ingredients:

- q valid correlation functions $\rho_j(\cdot, \cdot)$
- $\boldsymbol{\Sigma}$ symmetric positive semidefinite
- a set of "reference" locations S

no additional constraints on parameter space

For **any** pair of locations:

$$\mathrm{cov}\{y_i(s), y_j(s')\} = C_{ij}(s, s') = \sigma_{ij} \left[\boldsymbol{h}_i(s) \boldsymbol{L}_i \boldsymbol{L}_j^\top \boldsymbol{h}_j(s') + \varepsilon_{ij}(s, s') \right]$$

where:

$$\boldsymbol{h}_i(s) = \rho_i(s, S)\rho_i(S, S)^{-1} \qquad r_i(s, s') = \rho_i(s, s') - \boldsymbol{h}_i(s)\rho_s(S, s') \qquad \varepsilon_{ij}(s, s') = \mathbb{1}_{\{s=s'\}} \sqrt{r_i(s, s) r_j(s, s)}$$

Inside-Out Cross-covariance: key properties

$$C_{ii}(s, s') = \begin{cases} \sigma_{ii}\rho_i(s, s') & \text{if } s \in S \text{ or } s' \in S \text{ or } s = s', \\ \sigma_{ii}\rho_i(s, S)\rho_i(S)^{-1}\rho_i(S, s') & \text{if } s, s' \in S^c \text{ and } s \neq s'. \end{cases}$$

- marginal covariance only depends on $\rho_i(\cdot)$
- like a "predictive process" (Banerjee et al. 2009) with knots S when both s and s' are not in S
- **easy to interpret, easy to assign priors**
- cross-covariances are not parametrized directly and $C_{ij}(s, s') \leq \sigma_{ij}$
- non-stationarity induced by dependence on S
- choice of S? default to observed locations
- **outcome-specific features** introduced via $\rho_i(\cdot)$ (eg. nugget effects)
- GP with IOX lead to **efficient Gibbs samplers** for response models and latent models
- new ways to define **spatial factor models**

GPs with IOX

- suppose we use IOX as the **covariance model for a multivariate GP**
- in GP-IOX, $\boldsymbol{y}(s)$ and $\boldsymbol{y}(s')$ are conditionally independent given $\boldsymbol{y}$ (i.e. the data at S)
- let Y be the matrix of observed variables (one per column) and
  V the matrix obtained by "spatial whitening" of each column of Y, i.e. $\boldsymbol{v}_j = \boldsymbol{L}_j^{-1}\boldsymbol{y}_j$
- likelihood and full conditional densities have convenient structure:

$$\log p(\boldsymbol{y} \mid \boldsymbol{\Sigma}) = \text{const} - \frac{n}{2}\log\det(\boldsymbol{\Sigma}) + \sum_{ij}\log \boldsymbol{L}_i^{-1}[j,j] - \frac{1}{2}\text{Tr}\left(\boldsymbol{V}\boldsymbol{\Sigma}^{-1}\boldsymbol{V}^\top\right)$$

$$\log p(\boldsymbol{y}_j \mid \boldsymbol{y}_{-j}) = \text{const} + \frac{1}{2}\log\det\{Q_{jj}\rho_j(\mathcal{S})^{-1}\} - \frac{1}{2Q_{jj}}\boldsymbol{Q}_{j\cdot}\boldsymbol{V}^\top\boldsymbol{V}\boldsymbol{Q}_{j\cdot}^\top$$

where $\boldsymbol{Q} = \boldsymbol{\Sigma}^{-1}$

- if n is large, we can use a Vecchia-style approximation to sparsify $\boldsymbol{L}_i^{-1}$
- the entirety of GP-IOX depends on $\boldsymbol{L}_i^{-1}$, we never work with $\boldsymbol{L}_i$ in practice
- factor models target $\boldsymbol{\Sigma}$ directly: **seamlessly plug-in any (non-spatial) factor model** (unlike LMC!)

GPs with IOX: models and algorithms

Response model

$$\boldsymbol{Y}(\cdot) \sim \text{GP-IOX}$$

- dimension reduction via clustering of $\rho_j(\cdot)$
- update covariance parameters $\boldsymbol{\theta}$ as a block or $\boldsymbol{\theta}_j \mid \boldsymbol{\theta}_{-j}$ Metropolis-within-Gibbs
- conditionally conjugate updates for $\boldsymbol{\Sigma}$ available

Latent model

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{B} + \boldsymbol{W} + \boldsymbol{E}$$

$$\boldsymbol{W}(\cdot) \sim \text{GP-IOX}$$

- dimension reduction via low-rank assumption on $\boldsymbol{\Sigma}$
- block sampler for $\boldsymbol{W}$ may be slow if n and/or q large
- better: block-sample $\boldsymbol{W}_j \mid \boldsymbol{W}_{-j}$ or single-site sampler

Application 1: simulated data - setup

- each dataset $n = 2{,}500$ locations, $q = 3$ outcomes, dimension $nq = 7{,}500$
- 60 datasets generated with IOX with Matérn components
- 60 datasets generated with multivariate Matérn

targets:

- estimation of $\mathrm{corr}\{\boldsymbol{Y}(s), \boldsymbol{Y}(s)\}$ (correlation at zero spatial distance)
- estimation of smoothness, spatial decay, and nuggets for each component

Application 1: simulated data - results

- each dataset $n = 2{,}500$ locations, $q = 3$ outcomes, dimension $nq = 7{,}500$
- 60 datasets generated with **IOX with Matérn components**
- 60 datasets generated with multivariate Matérn

targets:

- estimation of $\text{corr}\{\boldsymbol{Y}(s), \boldsymbol{Y}(s)\}$ (correlation at zero spatial distance)
- estimation of smoothness, spatial decay, and nuggets for each component

results:

- GP-IOX models outperform others in all tasks

IOX data	ρ_{21}	ρ_{31}	ρ_{32}	ν_1	ν_2	ν_3	ϕ_1	ϕ_2	ϕ_3	τ_1^2	τ_2^2	τ_3^2	Time
IOX Response	**0.0045**	0.0208	0.0188	0.1100	0.0335	**0.0474**	**2.89**	**2.01**	**2.28**	0.0089	0.0007	**0.0005**	12
IOX Latent Sequential single-site	0.0065	**0.0198**	0.0187	0.0803	0.0293	0.0836	3.25	2.11	3.90	0.0013	**0.0004**	**0.0005**	22
IOX Latent Sequential single-outcome	0.0058	0.0197	**0.0184**	**0.0763**	**0.0285**	0.0842	3.36	2.13	3.87	**0.0006**	0.0005	**0.0005**	41
Mult. Matérn	0.0098	0.0246	0.0226	0.1170	0.0620	0.0616	7.53	3.31	2.32	0.0209	0.0026	0.0006	3
LMC	0.0936	0.3510	0.4020							0.0252	0.0025	0.0020	13

Application 1: simulated data - results

- each dataset $n = 2{,}500$ locations, $q = 3$ outcomes, dimension $nq = 7{,}500$
- 60 datasets generated with IOX with Matérn components
- 60 datasets generated with **multivariate Matérn**

targets:

- estimation of $\text{corr}\{\boldsymbol{Y}(s), \boldsymbol{Y}(s)\}$ (correlation at zero spatial distance)
- estimation of smoothness, spatial decay, and nuggets for each component

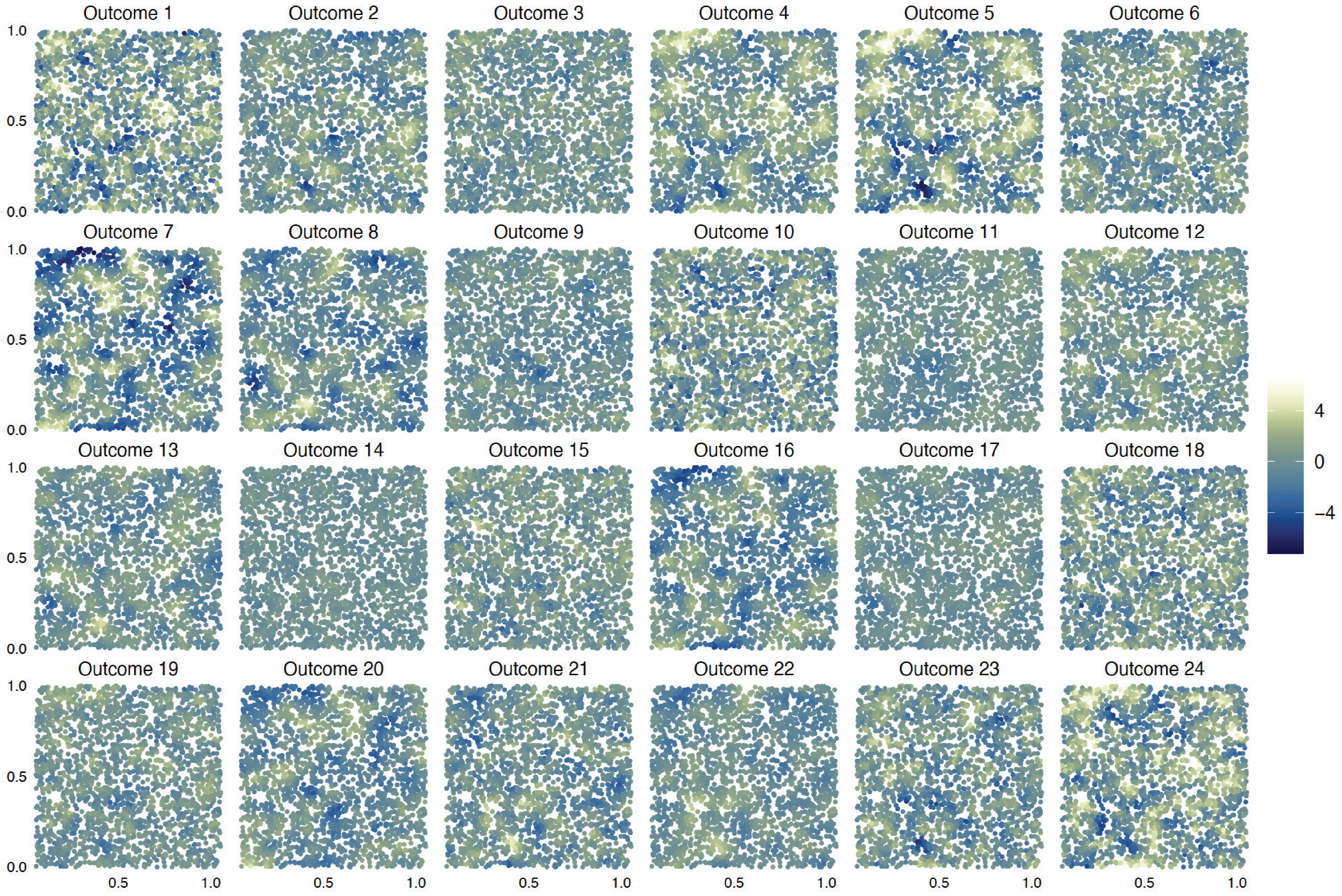
results:

- GP-IOX models (mispecified) competitive with the (well specified) multivariate Matérn

Mult. Matérn data	ρ_{21}	ρ_{31}	ρ_{32}	ν_1	ν_2	ν_3	ϕ_1	ϕ_2	ϕ_3	τ_1^2	τ_2^2	τ_3^2	Time
IOX Response	0.0228	0.0533	0.0506	0.1030	0.0465	0.0539	**2.84**	**2.23**	2.21	0.0219	0.0016	0.0009	11
IOX Latent Sequential single-site	0.0100	0.0431	0.0440	0.0351	0.0462	0.0998	3.92	2.34	3.45	0.0056	0.0002	**0.0004**	21
IOX Latent Sequential single-outcome	0.0129	0.0452	0.0454	**0.0258**	0.0551	0.1050	4.29	2.53	3.43	**0.0037**	**0.0001**	**0.0004**	40
Mult. Matérn	**0.0074**	**0.0180**	**0.0234**	0.0527	**0.0436**	**0.0473**	4.03	2.92	**2.10**	0.0109	0.0013	**0.0004**	3
LMC	0.0643	0.3450	0.3920							0.0269	0.0032	0.0024	12

Application 2: simulated data - setup

- each dataset n = 2,500 locations, q = 24 outcomes, dimension nq = 60,000
- 20 datasets generated with IOX
- 20 datasets generated with LMC (k=8)
- target estimating $\mathrm{corr}\{\boldsymbol{Y}(s), \boldsymbol{Y}(s)\}$ (correlation at zero spatial distance)
- target predictions at 400 out-of-sample locations



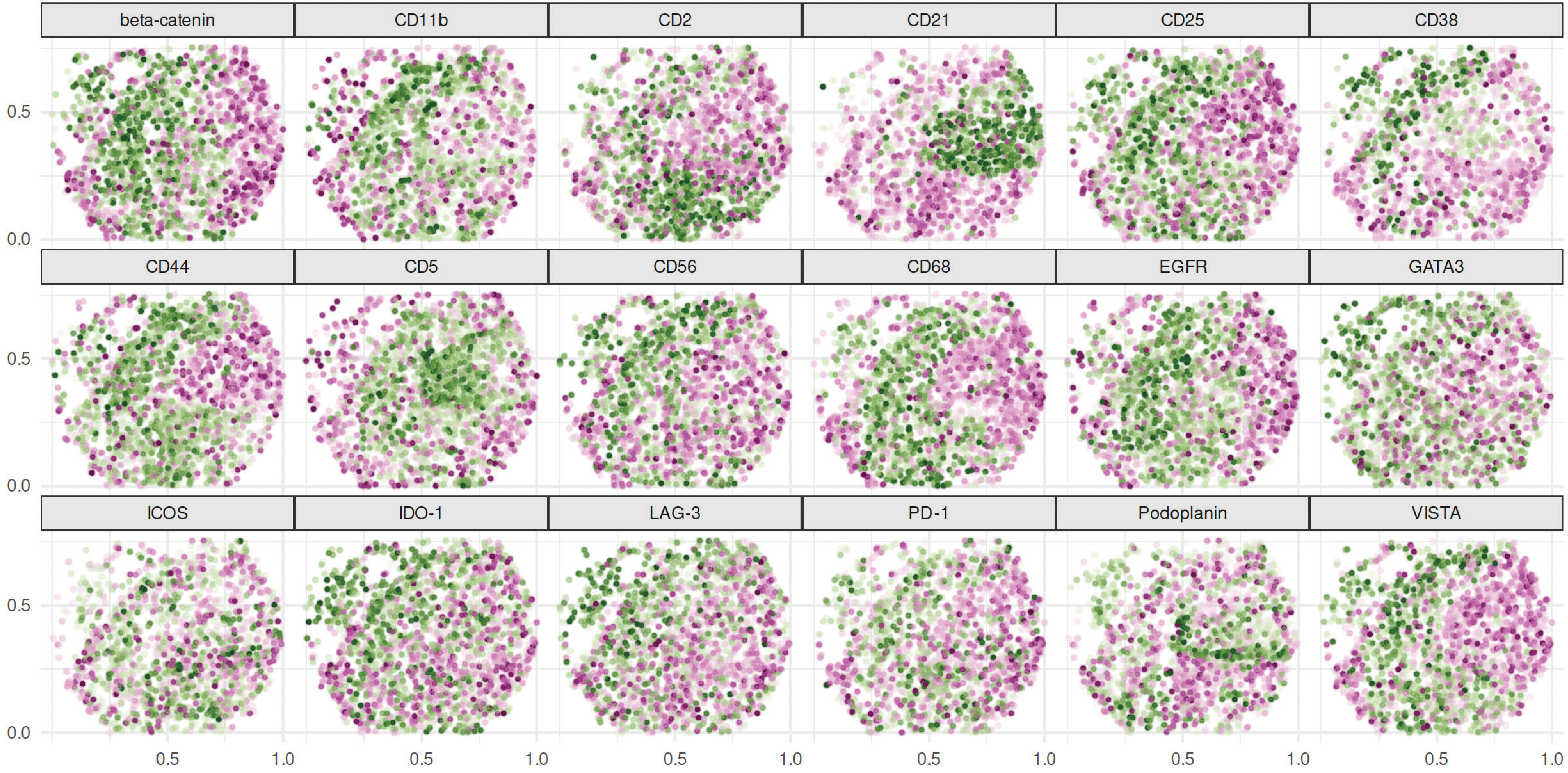
Application 2: simulated data - results

	IOX data					LMC data			
Method	ρ_{ij}	ν_j	Predictions (full)	Predictions (partial)	Time	ρ_{ij}	Predictions (full)	Predictions (partial)	Time
IOX Full	**0.0167**	**0.0692**	**0.482**	**0.123**	40	0.162	1.22	1.15	66
IOX Grid	0.0250	0.169	0.490	0.140	4.1	0.234	1.45	1.46	11
IOX Cluster	0.0191	0.250	0.493	0.138	12	0.163	1.23	1.16	20
LMC	0.270		0.685	0.631	15	0.312	**1.15**	**1.08**	34
NNGP Indep. univariate	0.106	0.124	0.483		76	0.123	1.21		55
Non-spatial model	0.0610			0.386	3	**0.0921**		1.27	3

- GP-IOX outperforms all others in the 20 IOX datasets
- GP-LMC does not outperform a non-spatial model in the 20 LMC datasets

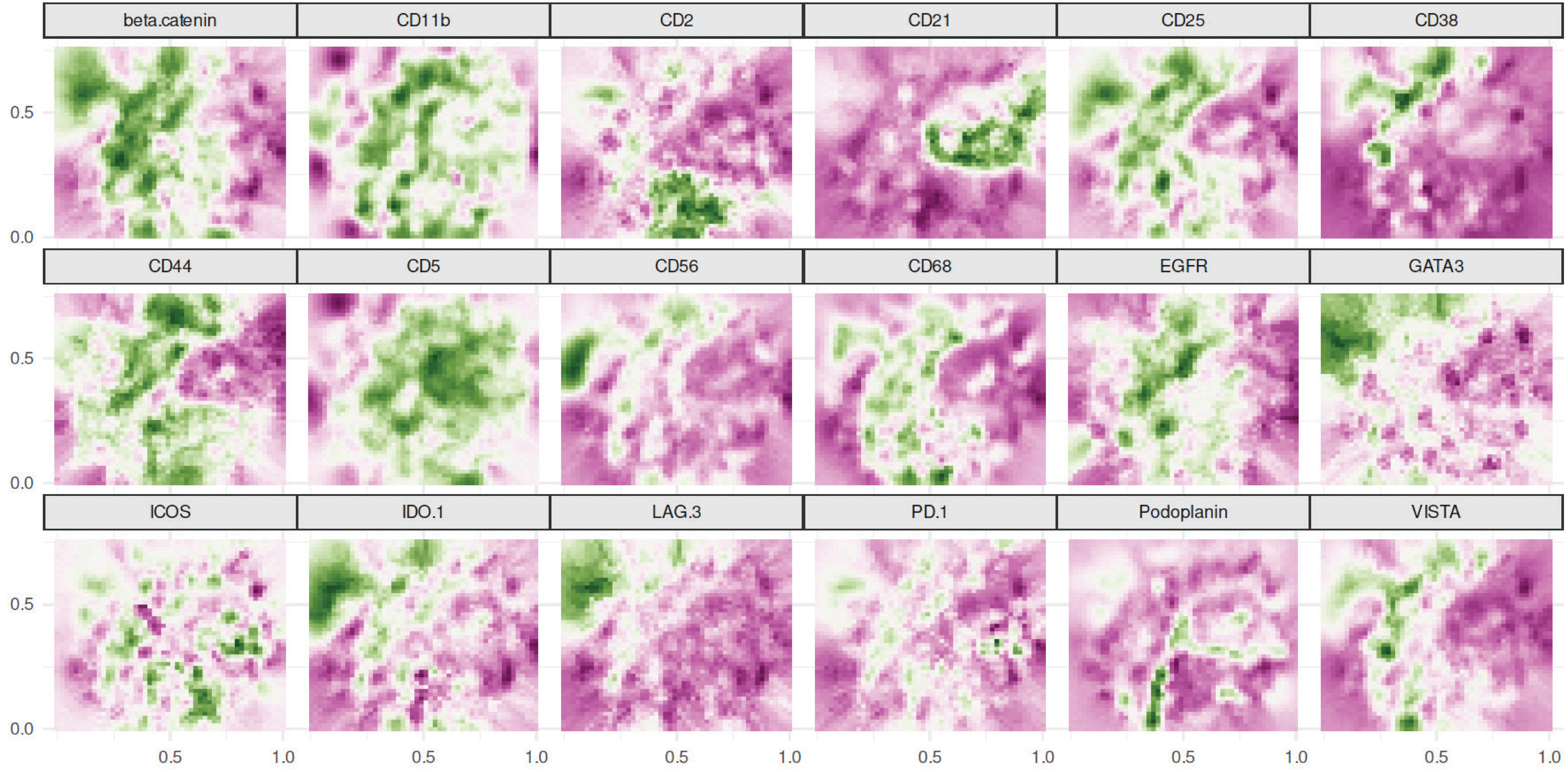
Application 2: colorectal cancer data - setup

- 18 protein markers on tissue biopsy slide from 1 patient
- detection intensity varies in space
- $n = 2{,}873$ spatial locations. $nq = 51{,}714$
- apply severeal GP-IOX models, LMC, and a non-spatial model



Application 2: colorectal cancer data - results

- Intensity maps reflect varying ranges, smoothness, variance



Application 2: colorectal cancer data - results

- Average percentage error in out-of-sample prediction of 2 variables given all others at the same location

Method	APE	Time(s)
IOX Full	0.0639	47
IOX Cluster	**0.0638**	22
LMC $k = 6$	0.0704	37
LMC $k = 8$	0.0679	52
Non-spatial model	0.0687	1

- IOX outperforms others while maintaining good scalability profile
- LMC must increase number of factors to outperform a simple non-spatial model

Conclusions

- IOX offers a new way to model multivariate spatial data
- structured covariance and precision matrices yield scalable algorithms
- flexibility in modeling outcome-specific features
- interpretable and direct parameter inference for marginal covariances
- competitive with multivariate Matérn in small dimensional settings, but can extend to higher-dimensional data
- competitive with LMC while being more flexible and interpretable
- software for fitting response & latent GP-IOX via MCMC at **`github.com/mkln/spiox`**
- for more info and references: https://arxiv.org/abs/2412.12407