Geometric Gaussian Processes

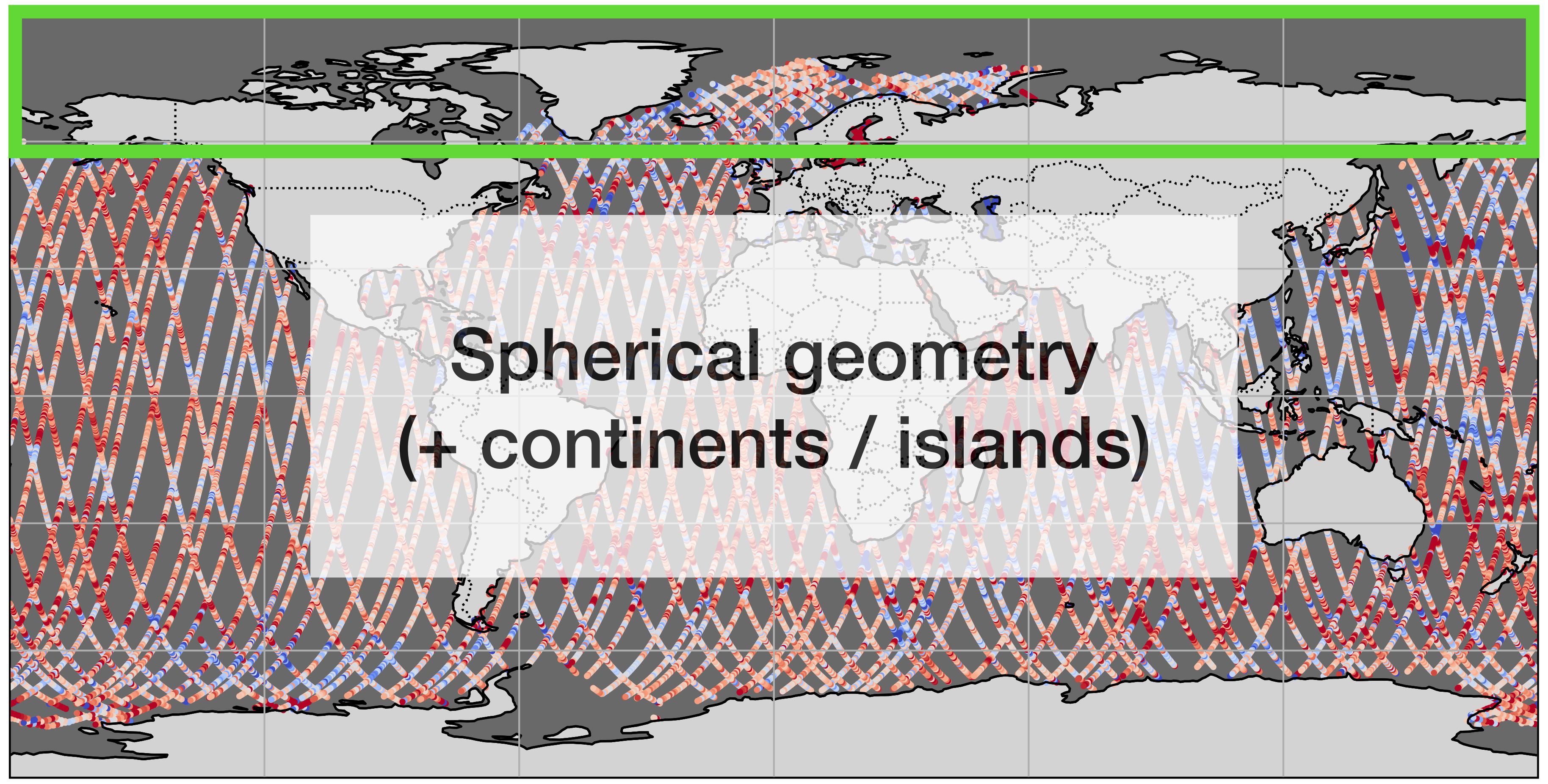
With applications to remote sensing observations

So Takao
Caltech

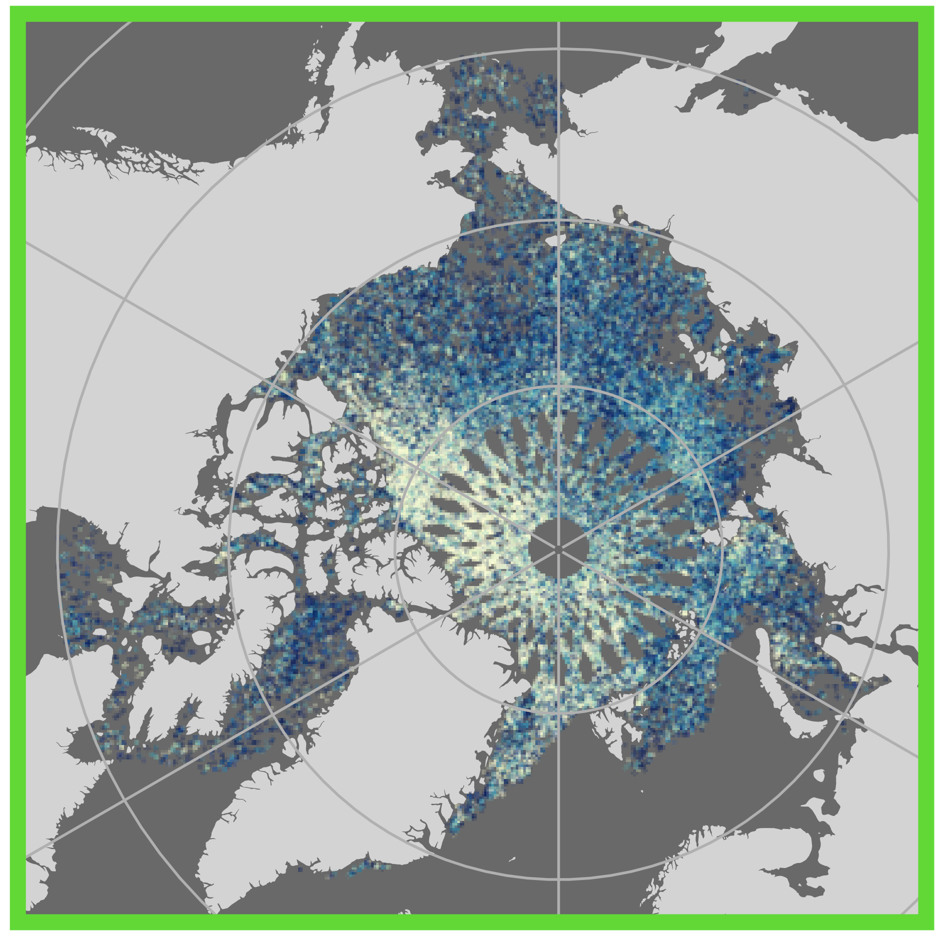
New era for EO

Innovation in satellite products calls for new approaches to extract information from them

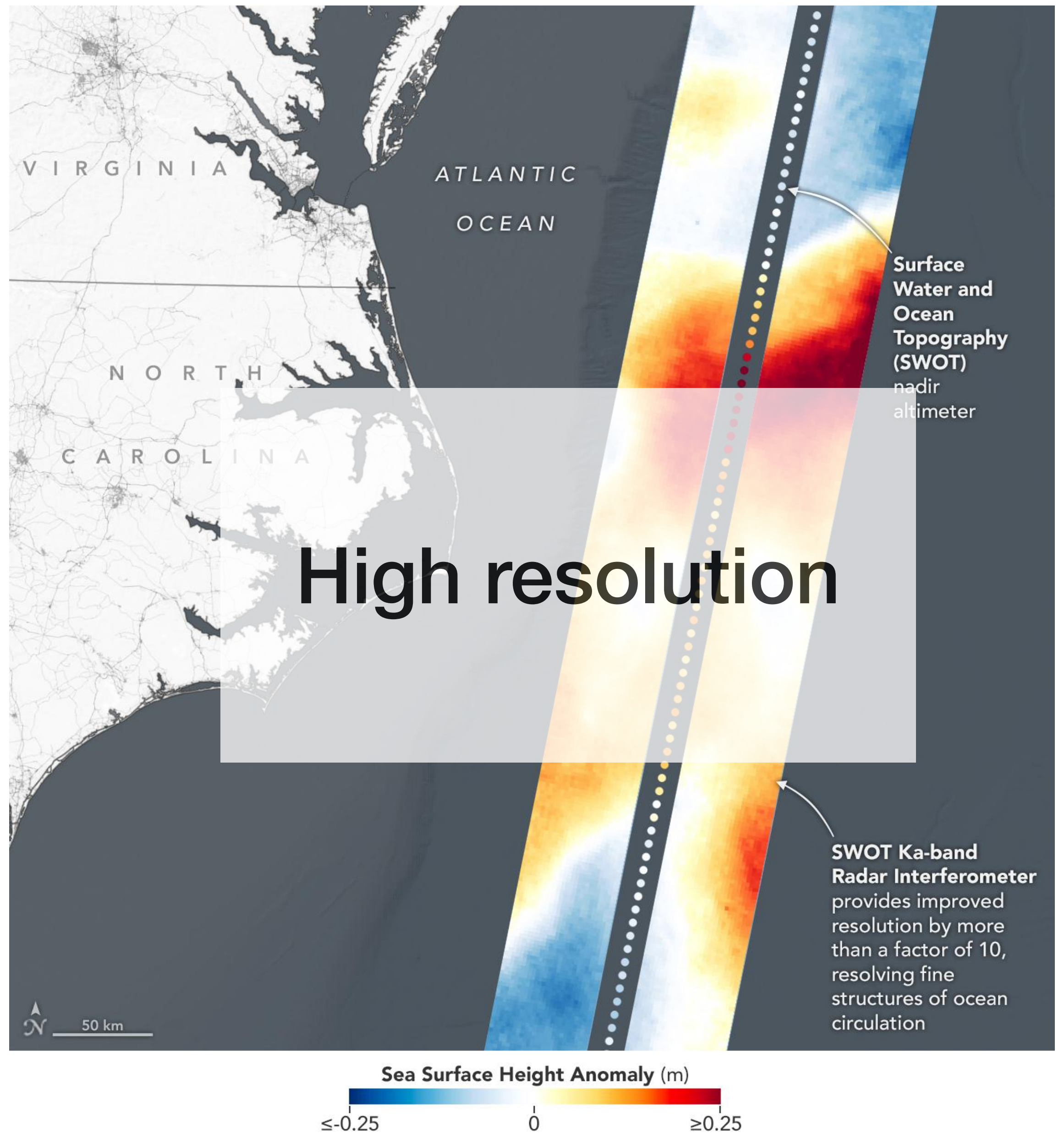
Aeolus project



Sentinel 3A and 3B

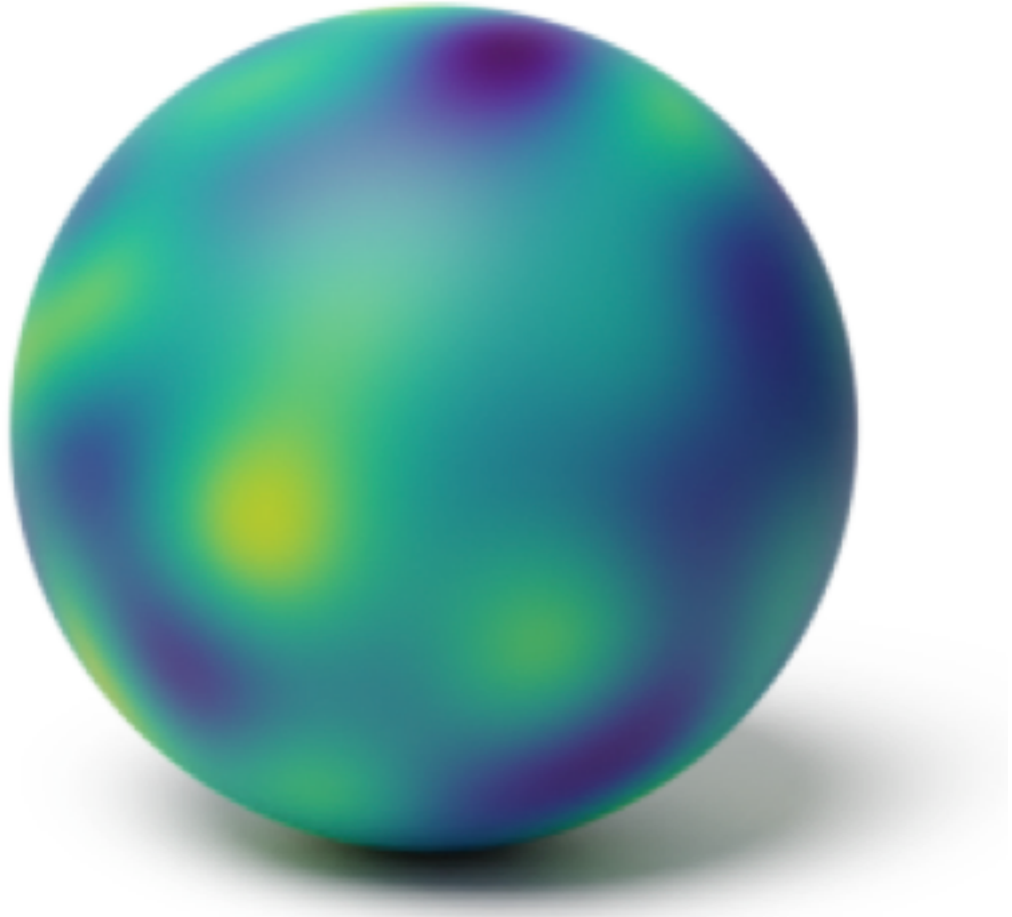


CryoSat 2

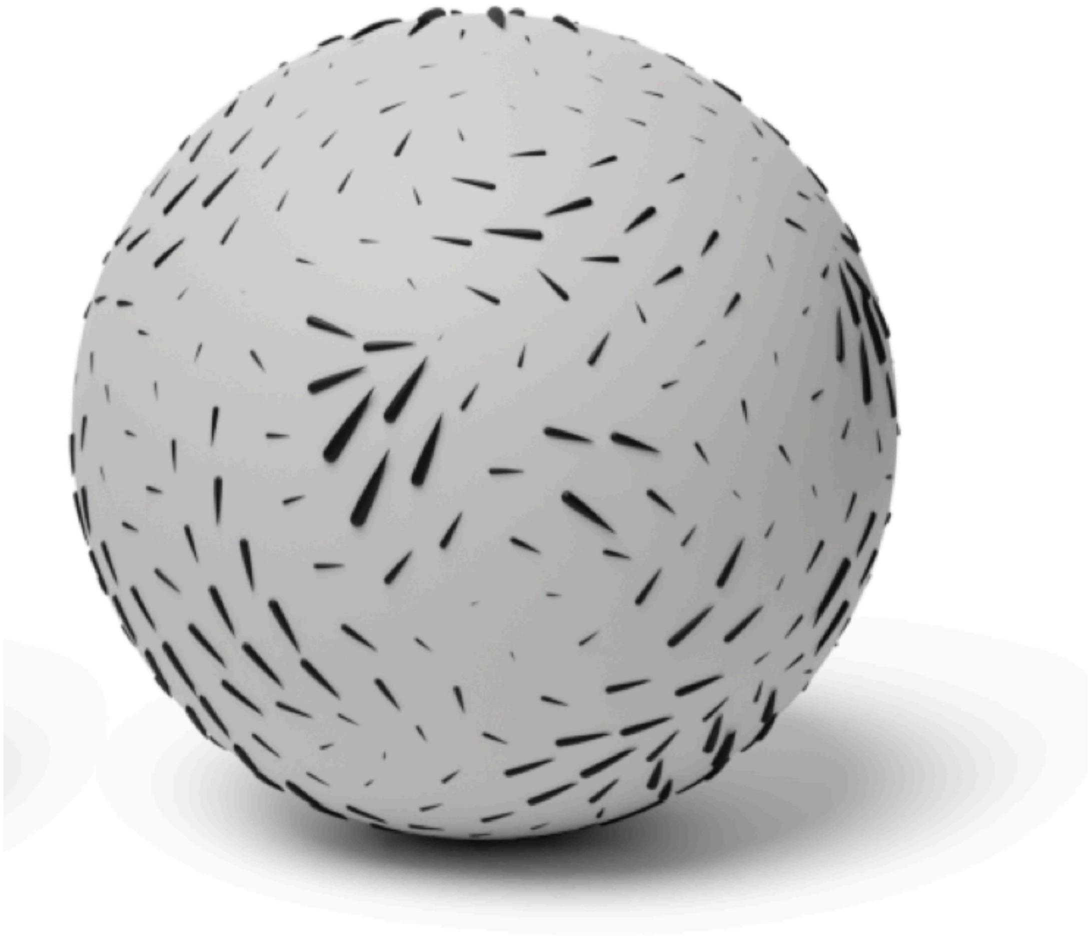


SWOT mission

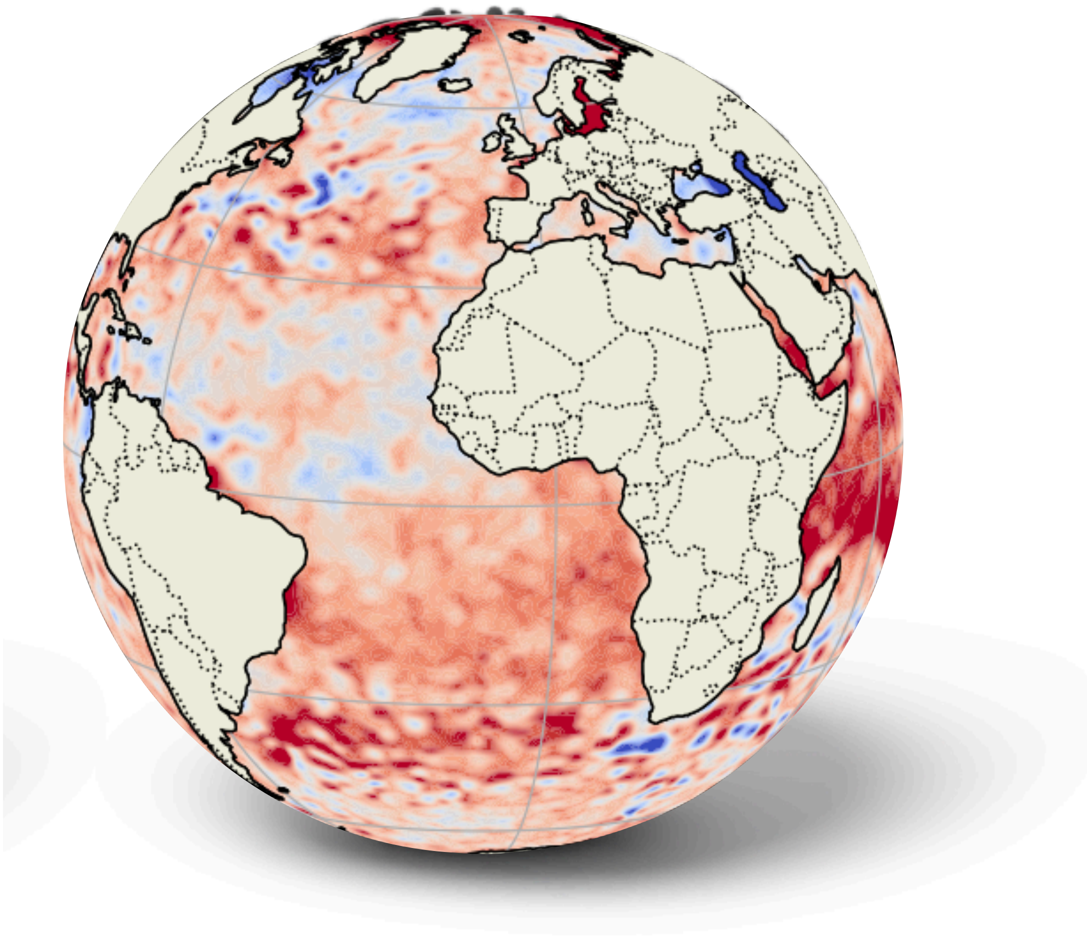
Part 1



Part 2



Part 3



0.0. Gaussian processes

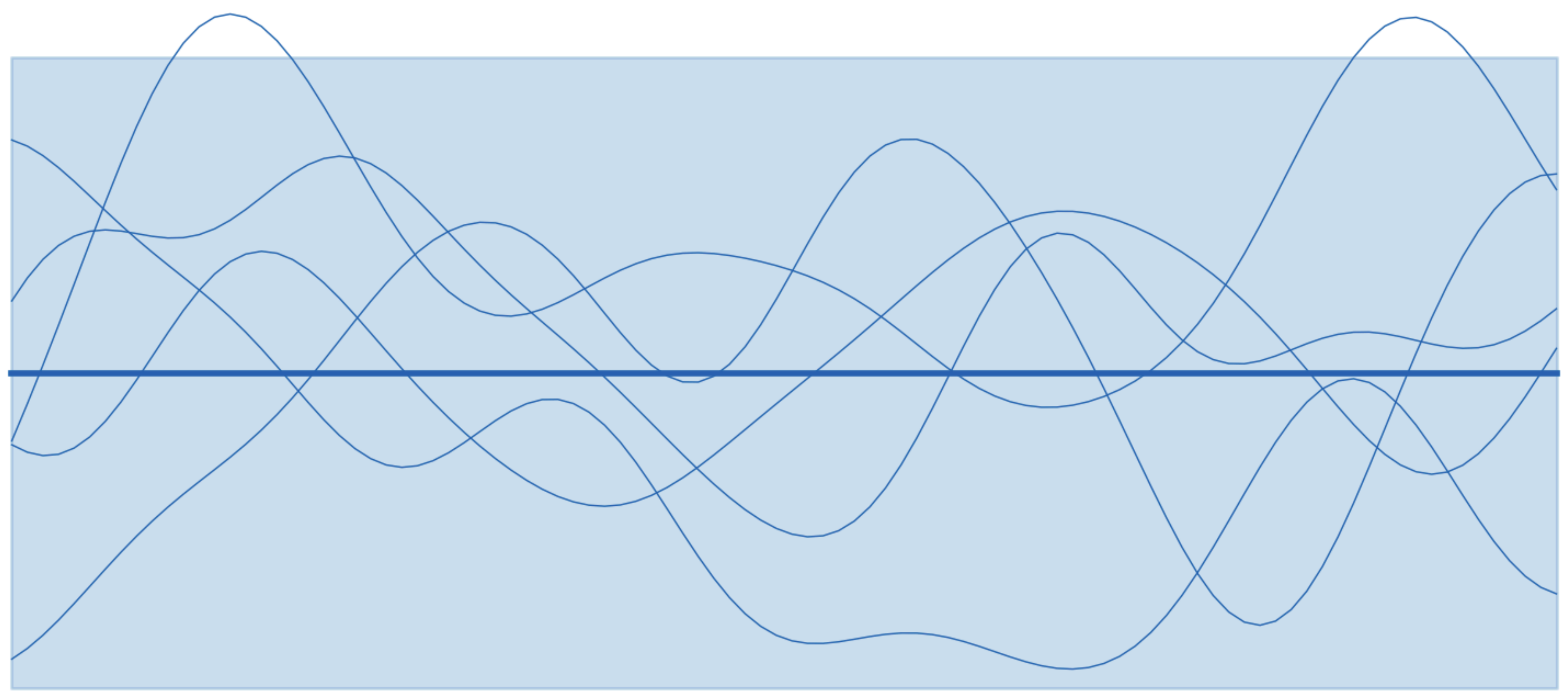
Gaussian processes

Gaussian processes (GP) are random variables on the *space of functions*

A GP $f : X \to \mathbb{R}$ is characterised by:

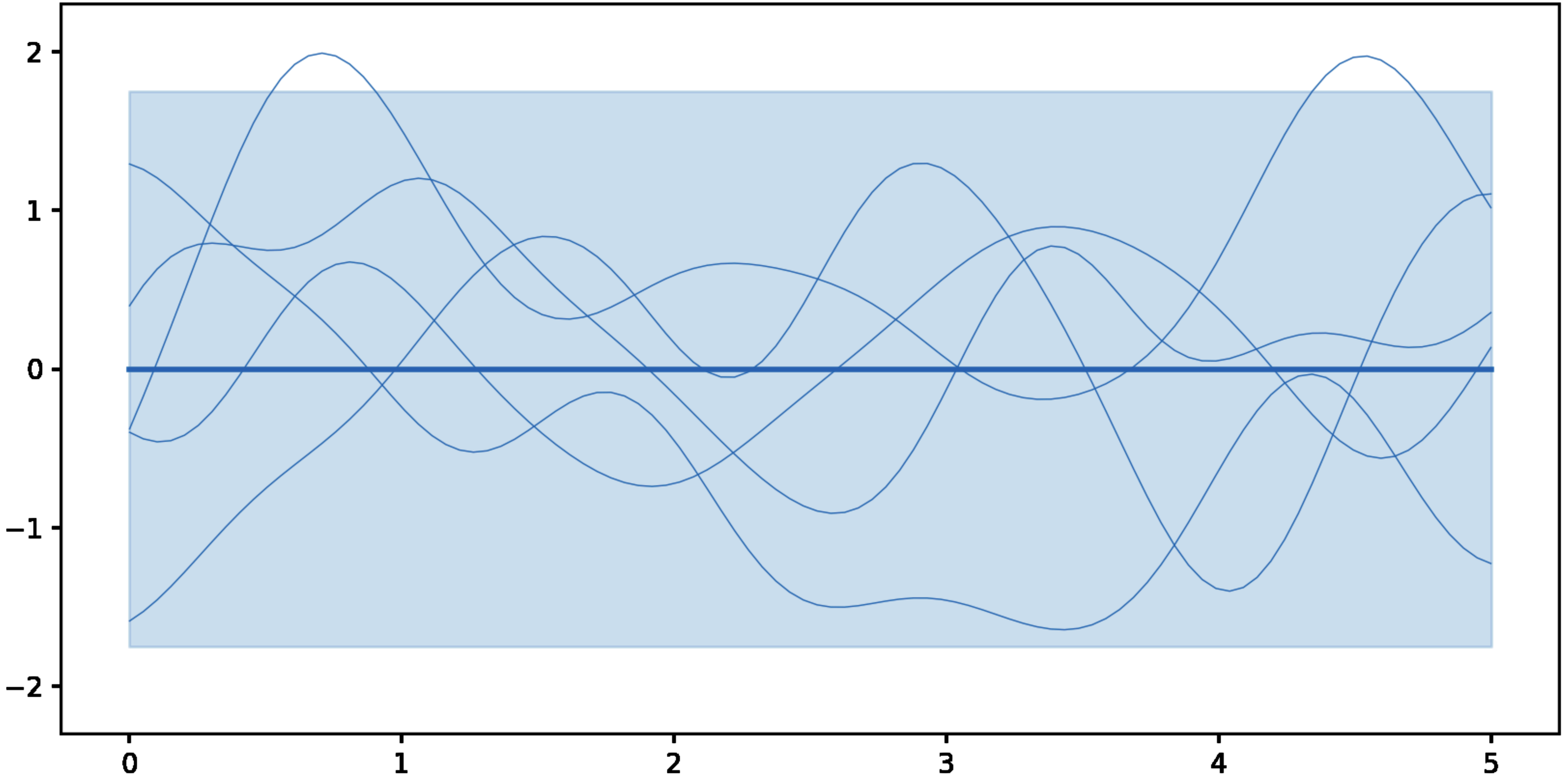
1. A mean function $m(x) := \mathbb{E}[f(x)]$, and
2. A kernel $k(x, x') = \mathrm{Cov}[f(x), f(x')]$,

   which is *positive semidefinite*.



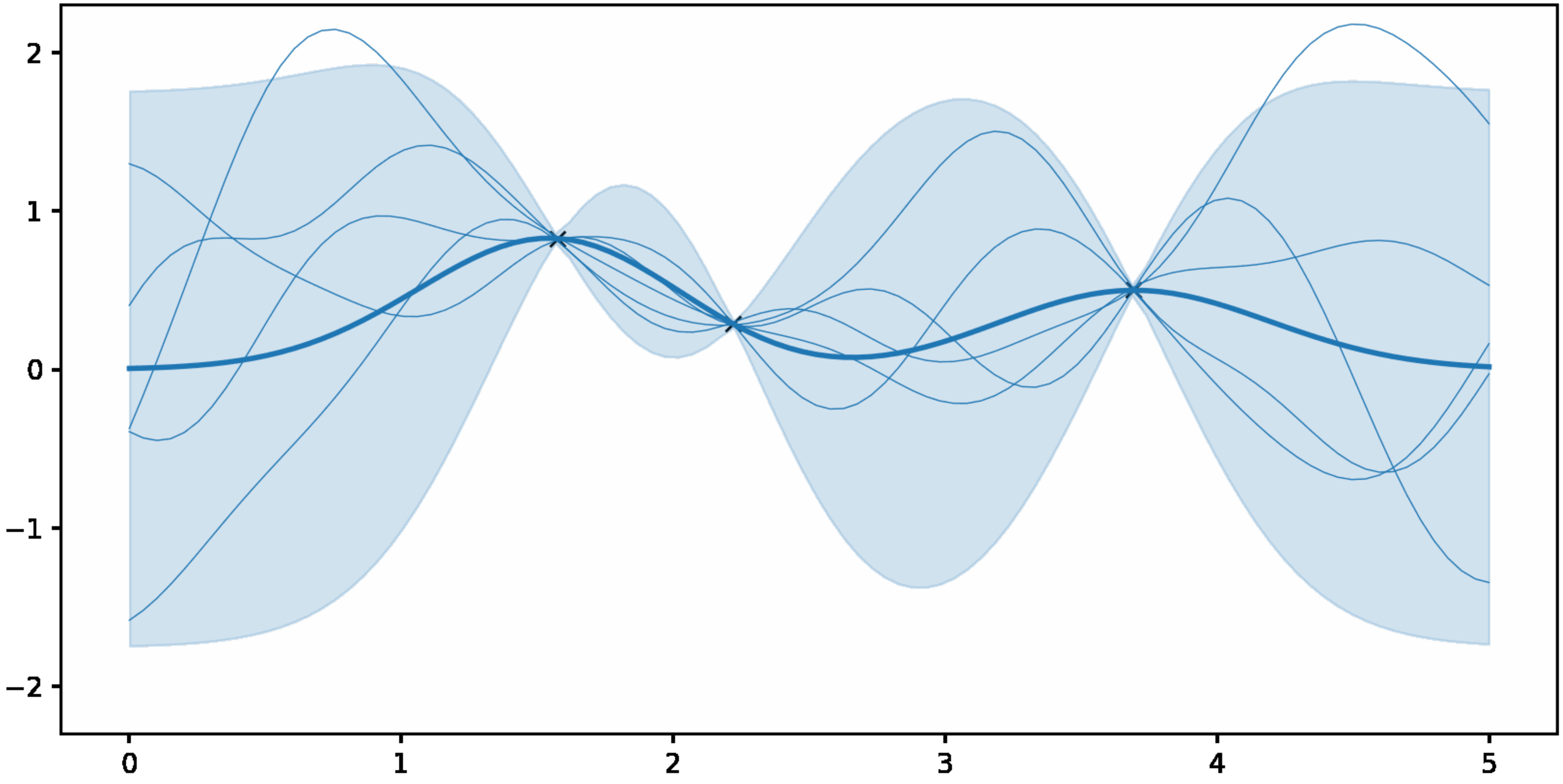
A zero-mean GP $f : \mathbb{R} \to \mathbb{R}$

$$p(f\,|\,y) = \frac{p(y\,|\,f)p(f)}{\int p(y\,|\,f)p(f)\mathrm{d}f}$$



A zero-mean GP $f : \mathbb{R} \to \mathbb{R}$

$$p(f \mid y) = \frac{p(y \mid f)p(f)}{\int p(y \mid f)p(f)\mathrm{d}f}$$

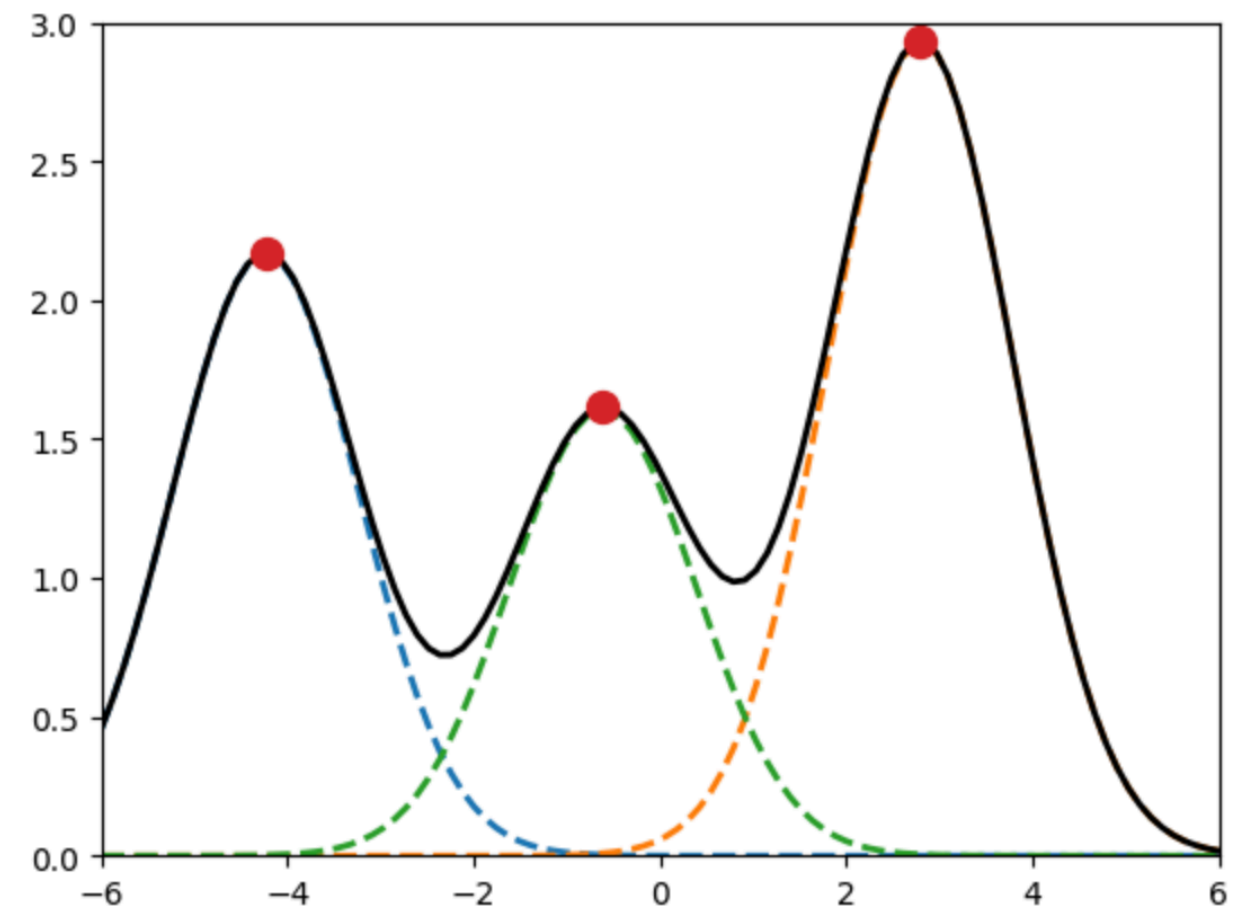


Function-space vs weight-space view

GP regression is a *kernel method*

$$\hat{f}(x) = \sum_{n=1}^{N} k(x, x_n)\hat{y}_n$$

Scales with number of data points N.



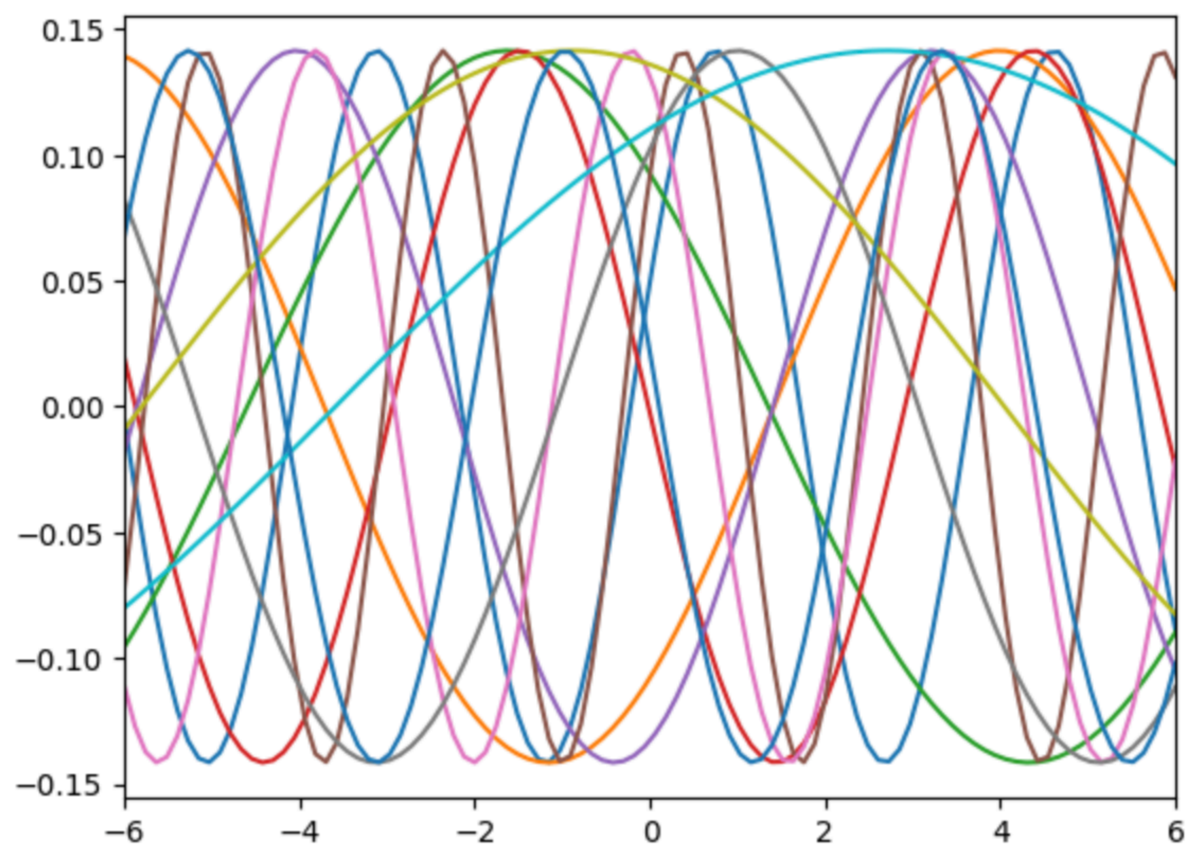
Predictive mean with $k(x, x') = \exp(-|x - x'|^2)$

Weight-space **view:**

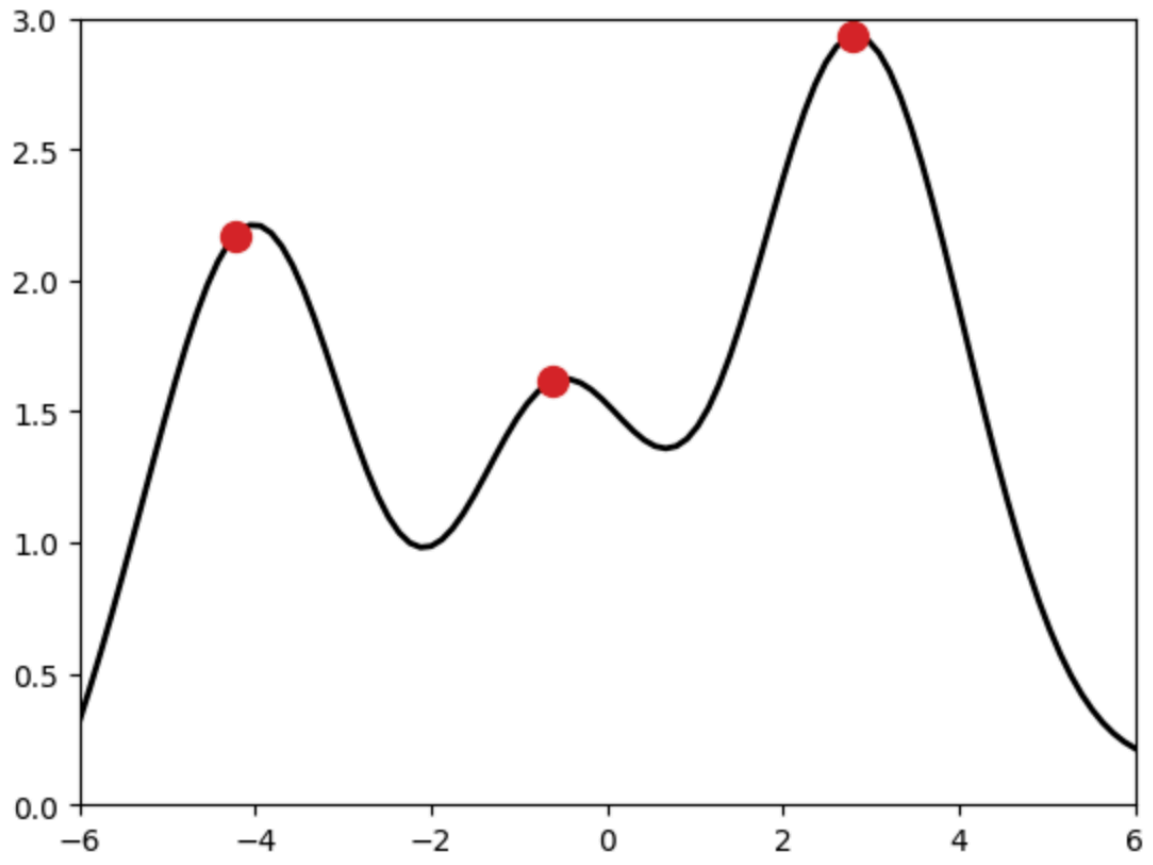
$$\hat{f}(x) \approx \sum_{m=1}^{M} \theta_m \psi_m(x)$$

where $k(x, x') \approx \sum \psi_m(x)\psi_m(x')$

Fit data by learning weights $\{\theta_m\}_{m=1}^{M}$.



Basis functions $\{\psi_m(x)\}_{m=1}^{M}$



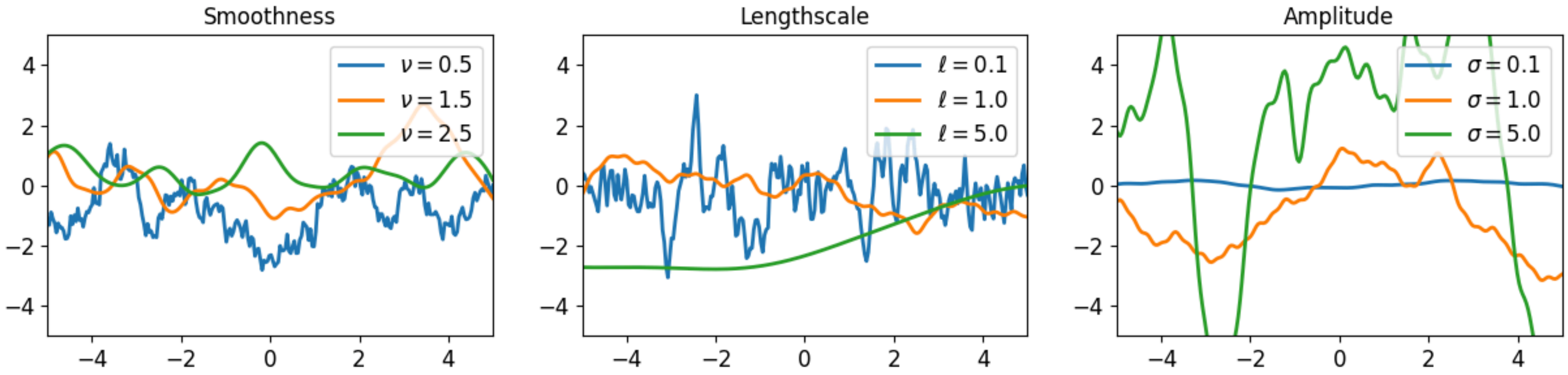
$\hat{f}(x) = \sum \hat{\theta}_m \psi_m(x)$

0.1. Matérn GPs

GPs of the Matérn family

Bertil Matérn

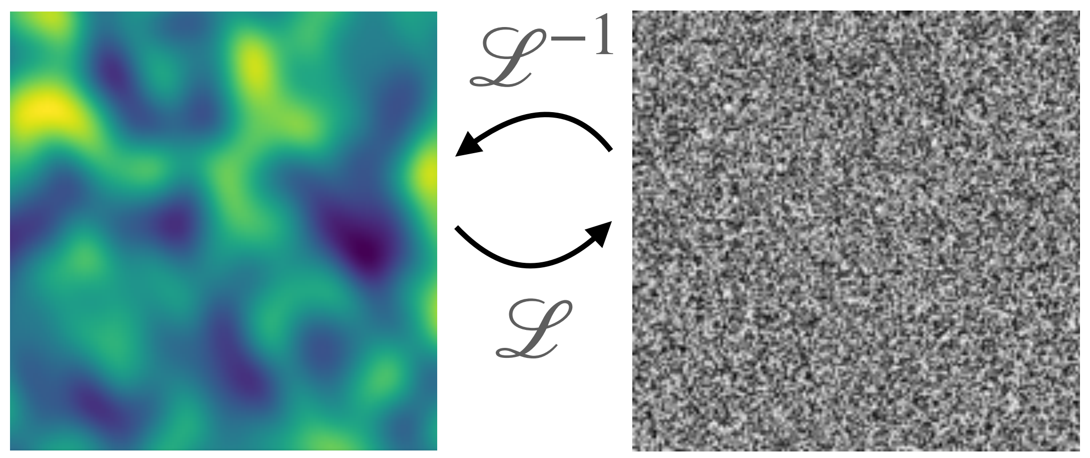
$$k_{\nu,\ell,\sigma}(x,x') = \sigma^2 \frac{2^{1-\nu}}{\Gamma(\nu)} \left(\sqrt{2\nu} \frac{|x-x'|}{\ell} \right)^{\nu} K_{\nu} \left(\sqrt{2\nu} \frac{|x-x'|}{\ell} \right)$$

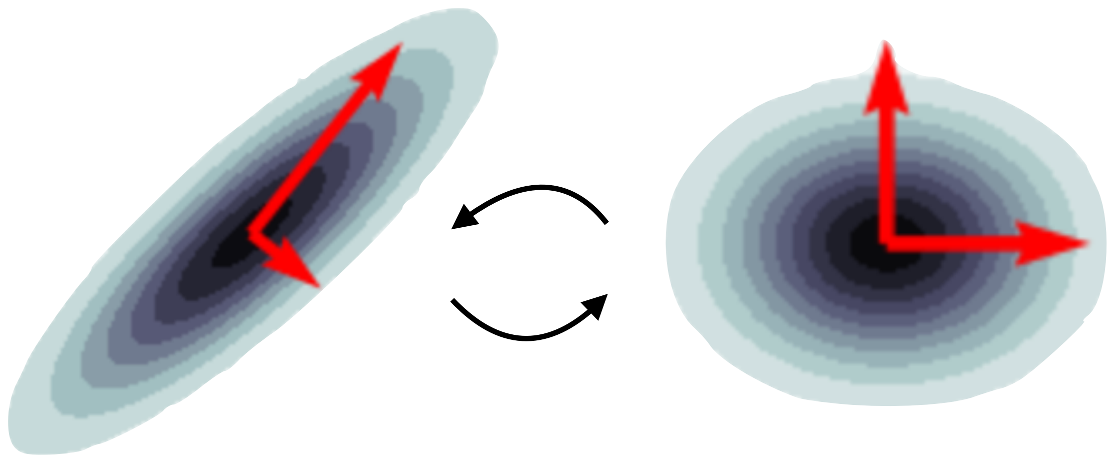


SPDE reformulation of Matérn GPs

Matérn GPs can also be derived as a solution to the stochastic PDE:

$$\underbrace{\left(\frac{2\nu}{\ell^2} - \Delta\right)}_{=:\,\mathscr{L}}{}^{\frac{\nu + d/2}{2}} f = \dot{\mathscr{W}}_\sigma$$





This is a “white noise” on *Sobolev spaces* over $\mathbb{R}^d$

$$H^{\nu+d/2}(\mathbb{R}^d) := \left\{ f \in L^2(\mathbb{R}^d) : \int_{\mathbb{R}^d} \left(1 + |\omega|^2\right)^{\nu+d/2} |\hat{f}(\omega)|^2 \mathrm{d}\omega < \infty \right\}.$$

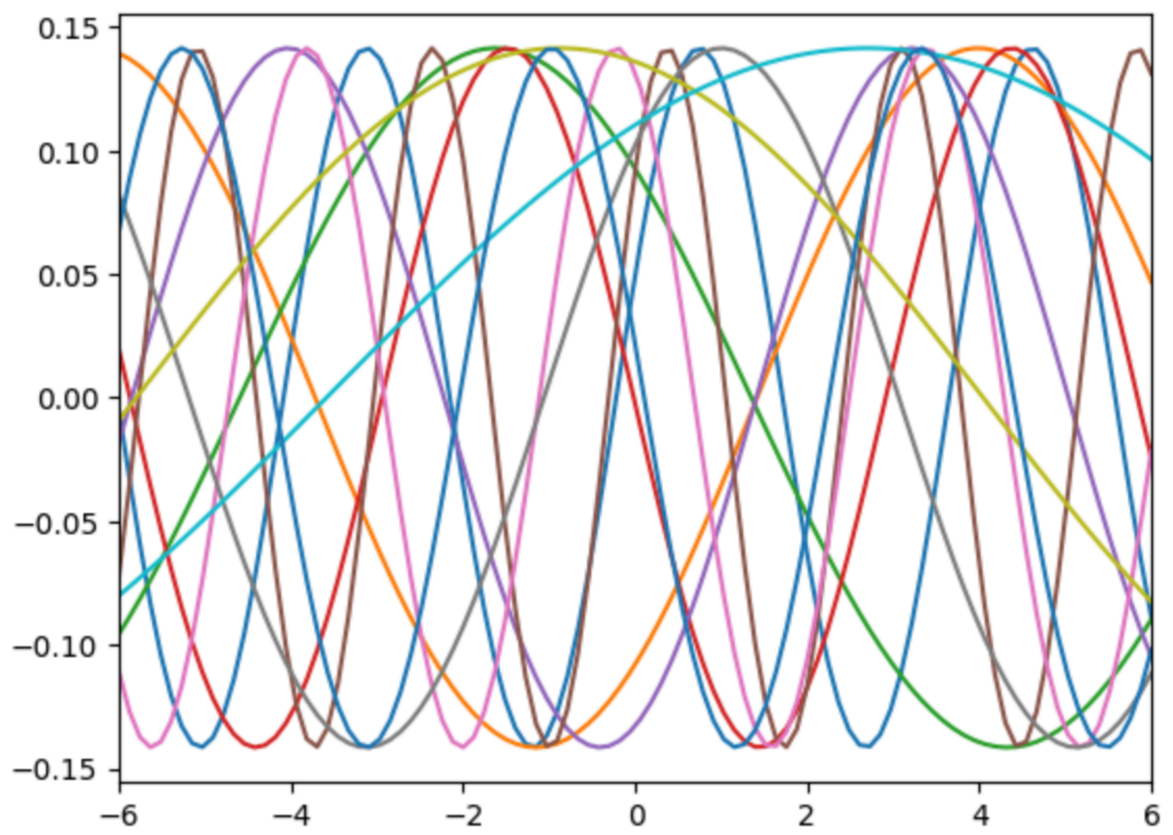
Random feature expansion of Matérn GPs

The Matérn kernel over $\mathbb{R}^d$ can also be expressed as

$$k_{\nu,\ell,\sigma}(x,x') = 2\sigma^2 \mathbb{E}_{\omega,b}[\cos(\omega^\top x + b)\cos(\omega^\top x' + b)], \quad \omega \sim t_{2\nu}(0,\ell^{-2}),$$

$$\approx \sum_{m=1}^{M} \sqrt{\frac{2\sigma^2}{M}} \cos(\omega_m^\top x + b_m) \sqrt{\frac{2\sigma^2}{M}} \cos(\omega_m^\top x' + b_m),$$

$$\omega_m \sim t_{2\nu}(0,\ell^{-2}), \quad b_m \sim U([0,2\pi]) \,.$$

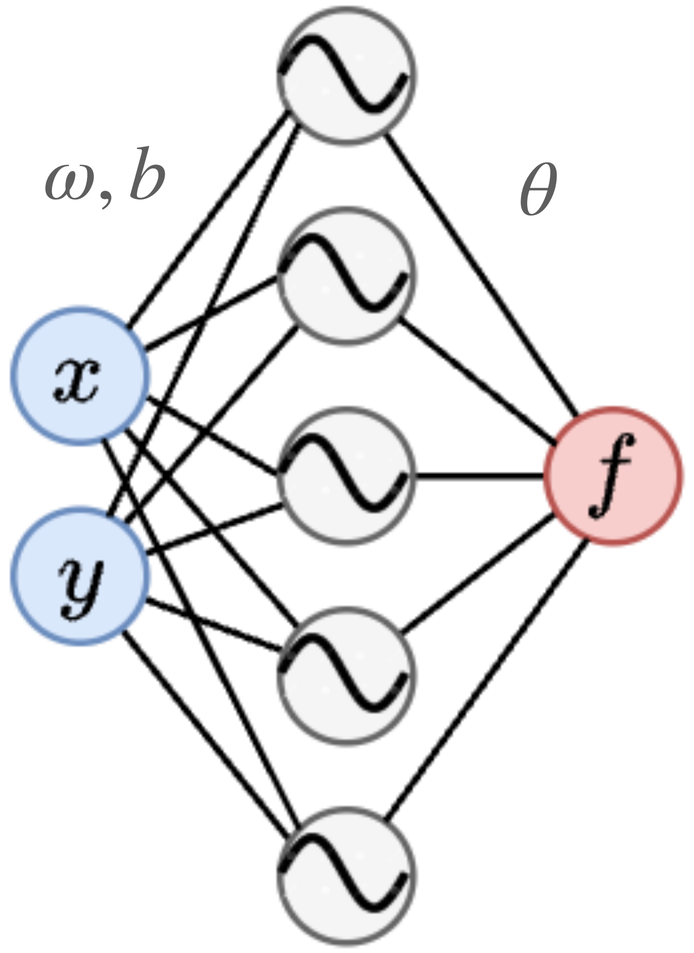


Random features!

This gives us the weight-space representation of Matérn GPs*

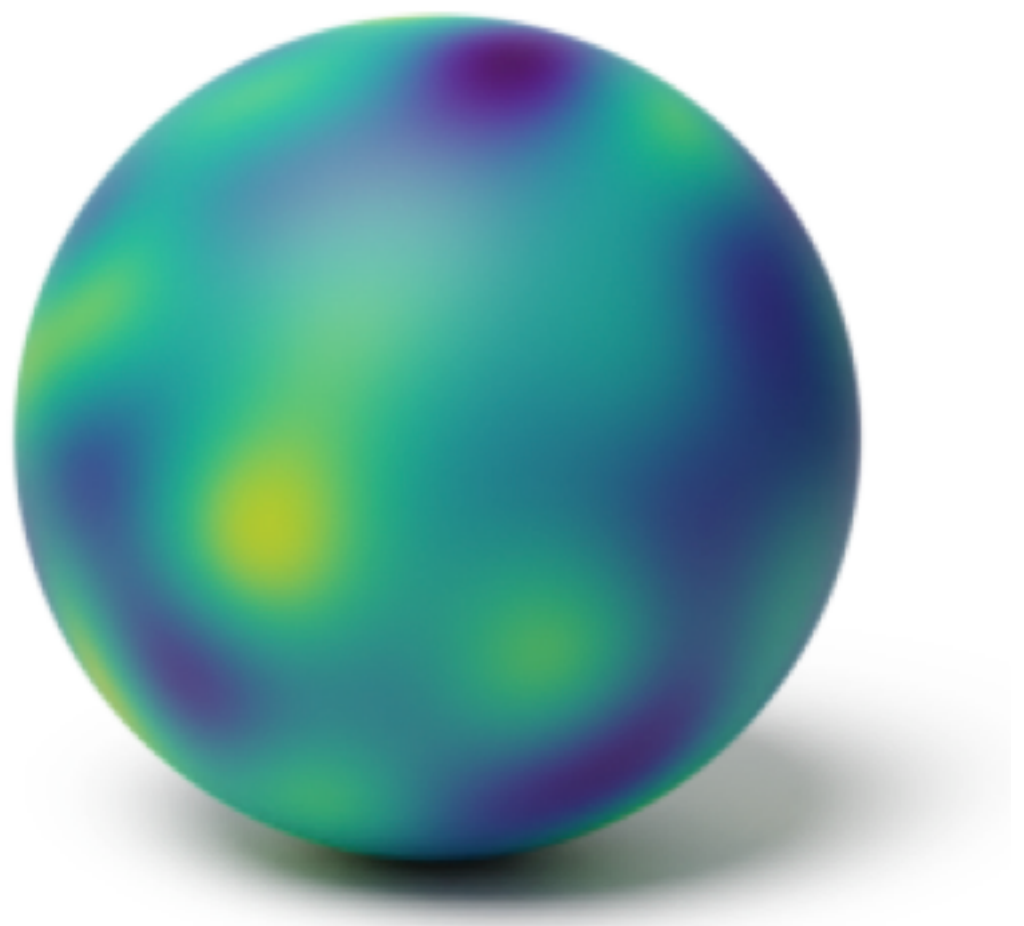
$$f(x) \approx \sum_{m=1}^{M} \theta_m \psi_m(x), \quad \text{where} \quad \psi_m(x) = \sqrt{\frac{2\sigma^2}{M}} \cos(\omega_m^\top x + b_m)$$

$$\text{and} \quad \theta_m \sim \mathcal{N}(0,1)\,.$$



*Rahimi et al. "Random features for large-scale kernel machines." *NeurIPS 2007.*

1. Riemannian Matérn GPs



A no-go result

Recall the Matérn kernel on $\mathbb{R}^d$

$$k_{\nu,\ell,\sigma}(x,x') = \sigma^2 \frac{2^{1-\nu}}{\Gamma(\nu)} \left(\sqrt{2\nu} \frac{|x-x'|}{\ell} \right)^{\nu} K_\nu \left(\sqrt{2\nu} \frac{|x-x'|}{\ell} \right)$$

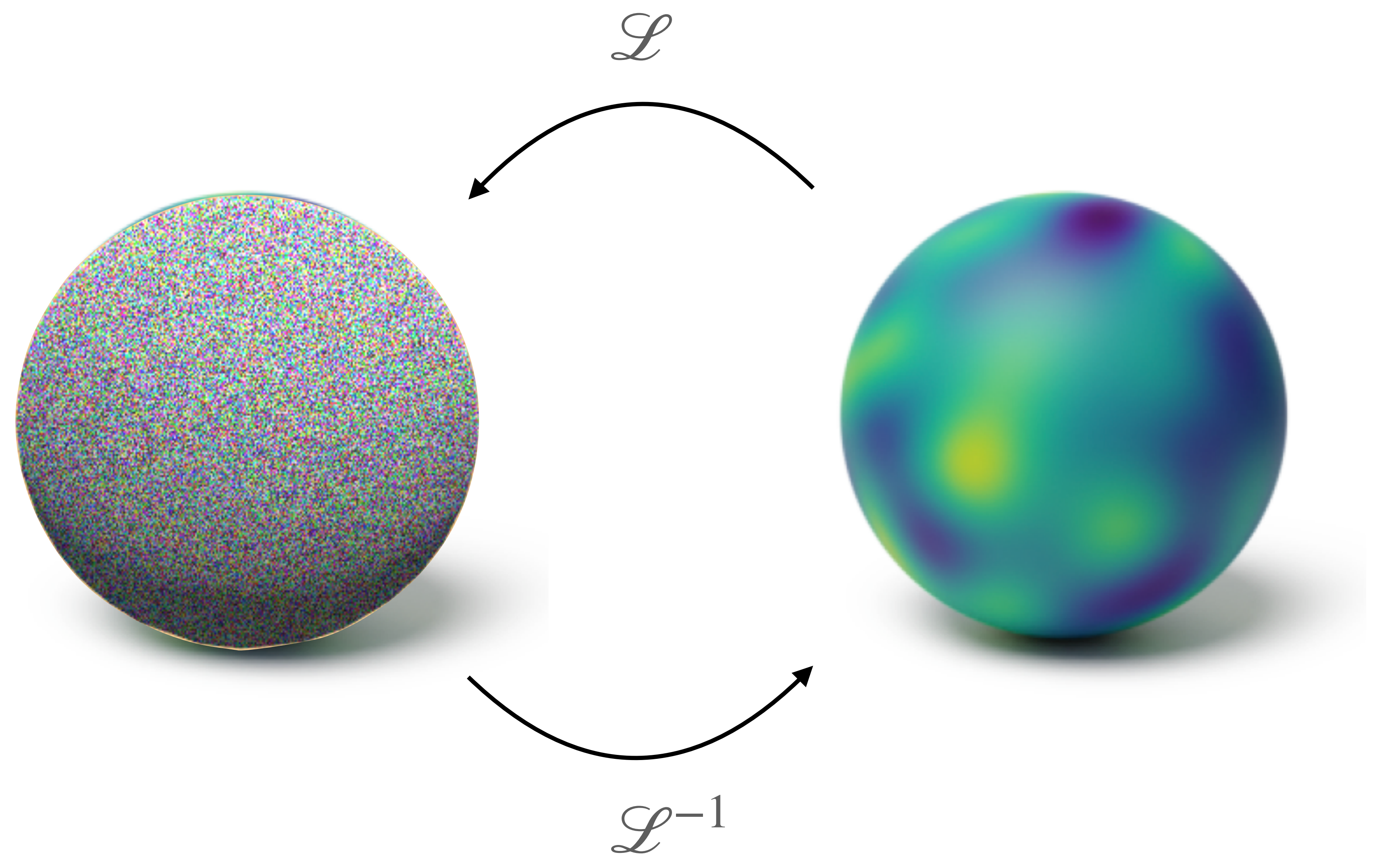
On manifolds, can we just replace $|x - x'|$ by the geodesic distance?

No-go Theorem*:

The resulting kernel is positive semidefinite if and only if the manifold is flat.

*Feragen et al.. "Geodesic exponential kernels: When curvature and linearity conflict." *CVPR*. 2015.

SPDE Approach



Recall the SPDE formulation of Matérn GP:

$$\underbrace{\left(\frac{2\nu}{\ell^2} - \Delta\right)}_{=:\,\mathscr{L}}^{\frac{\nu+d/2}{2}} f = \dot{\mathscr{W}}_\sigma$$

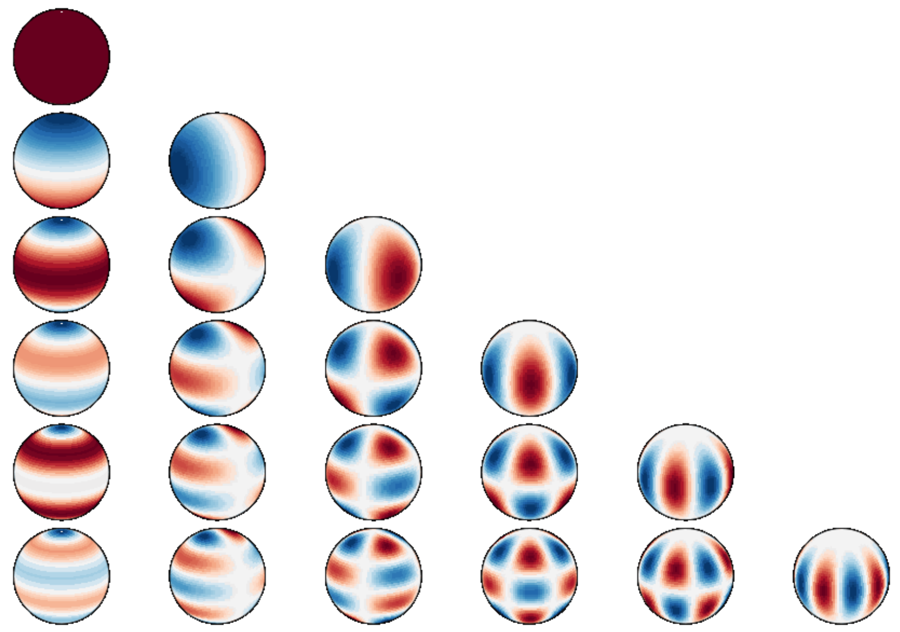
Extend to Riemannian manifolds by using the Laplace-Beltrami operator Δ?*

*Lindgren et al. "An explicit link between Gaussian fields and Gaussian Markov random fields: the stochastic partial differential equation approach." *J. Royal. Stat. Soc. B. 2011*.

Kernel construction

Let $\{(\lambda_n, \phi_n)\}_{n=0}^{\infty}$ be eigendecomposition of $-\Delta$ over a compact Riemannian manifold

- On $\mathbb{T}^d$, we have $\lambda_n = n^2$ and $\phi_n(x) = \{\sin(nx), \cos(nx)\}$
- On $\mathbb{S}^2$, we have $\lambda_n = n(n+1)$ and $\phi_n(x) = \{Y_n^1(x), \ldots, Y_n^{2n+1}(x)\}$



Theorem:*

$f \sim GP(0, k(\cdot, \cdot))$, where

$$k(x, x') = \frac{\sigma^2}{C_{\nu,\ell}} \sum_{n=0}^{\infty} \left(\frac{2\nu}{\ell^2} + \lambda_n \right)^{-\nu-d/2} \phi_n(x)\phi_n(x')$$

is a solution to the SPDE $\left(2\nu/\ell^2 - \Delta\right)^{\frac{\nu+d/2}{2}} f = \dot{\mathscr{W}}_\sigma$

*Borovitskiy et al. "Matérn Gaussian processes on Riemannian manifolds." *NeurIPS. 2020.*

Concrete expressions*

Matérn-ν on $\mathbb{R}^d$:

$$k_{\nu,\ell,\sigma}(x,x') = \sigma^2 \frac{2^{1-\nu}}{\Gamma(\nu)} \left(\sqrt{2\nu}\frac{|x-x'|}{\ell}\right)^{\nu} K_\nu\left(\sqrt{2\nu}\frac{|x-x'|}{\ell}\right)$$

Matérn-$1/2$ on $\mathbb{T}^1$:

$$k_{1/2,\ell,\sigma}(x,x') = \frac{\sigma^2}{\cosh(1/2\ell)} \cosh\left(\frac{|x-x'-1/2|}{\ell}\right)$$

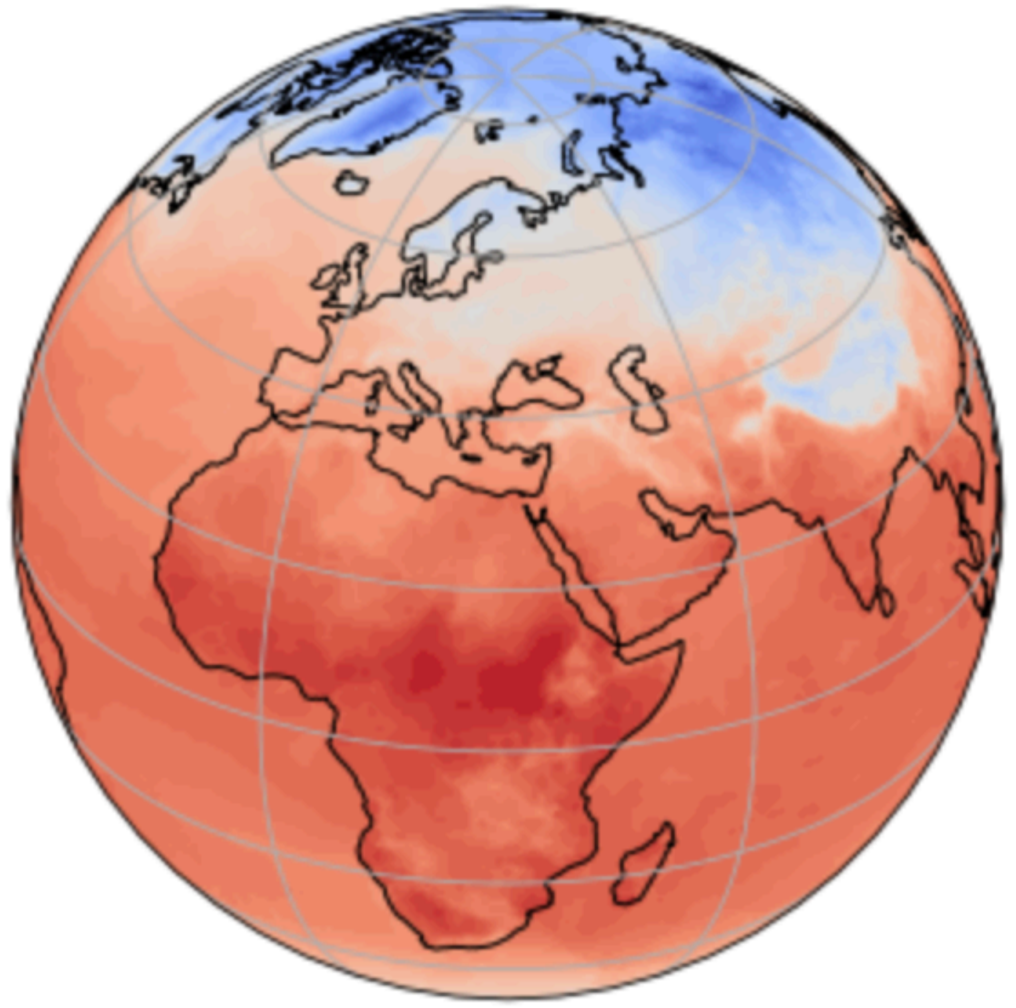
Matérn-ν on $\mathbb{S}^2$:

Gegenbauer polynomial

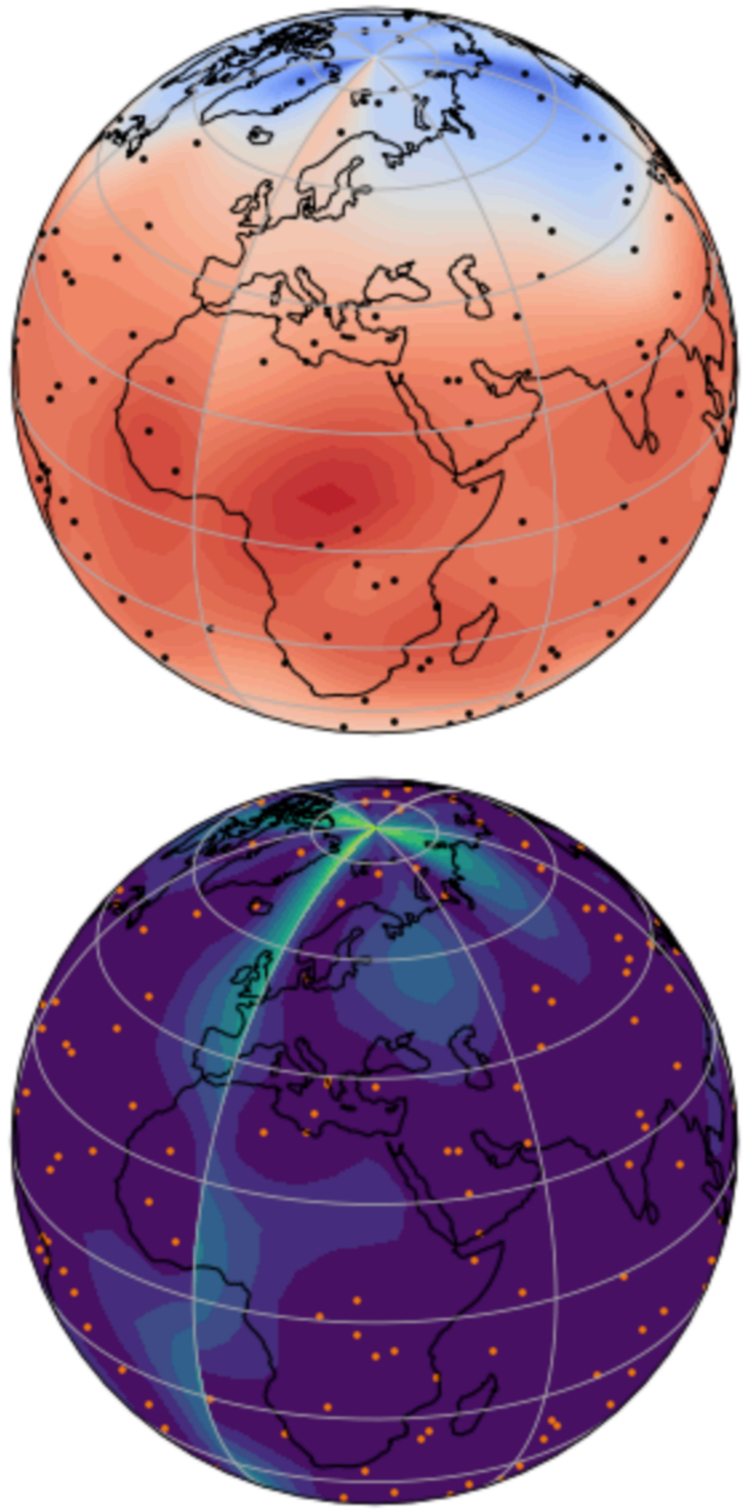
$$k_{\nu,\ell,\sigma}(x,x') = \sigma^2 \sum_{n=0}^{\infty} c_n \rho_\nu(n) \mathcal{G}_n^{1/2}(\cos(\langle x, x'\rangle))$$

*Borovitskiy et al. "Matérn Gaussian processes on Riemannian manifolds." *NeurIPS. 2020.*

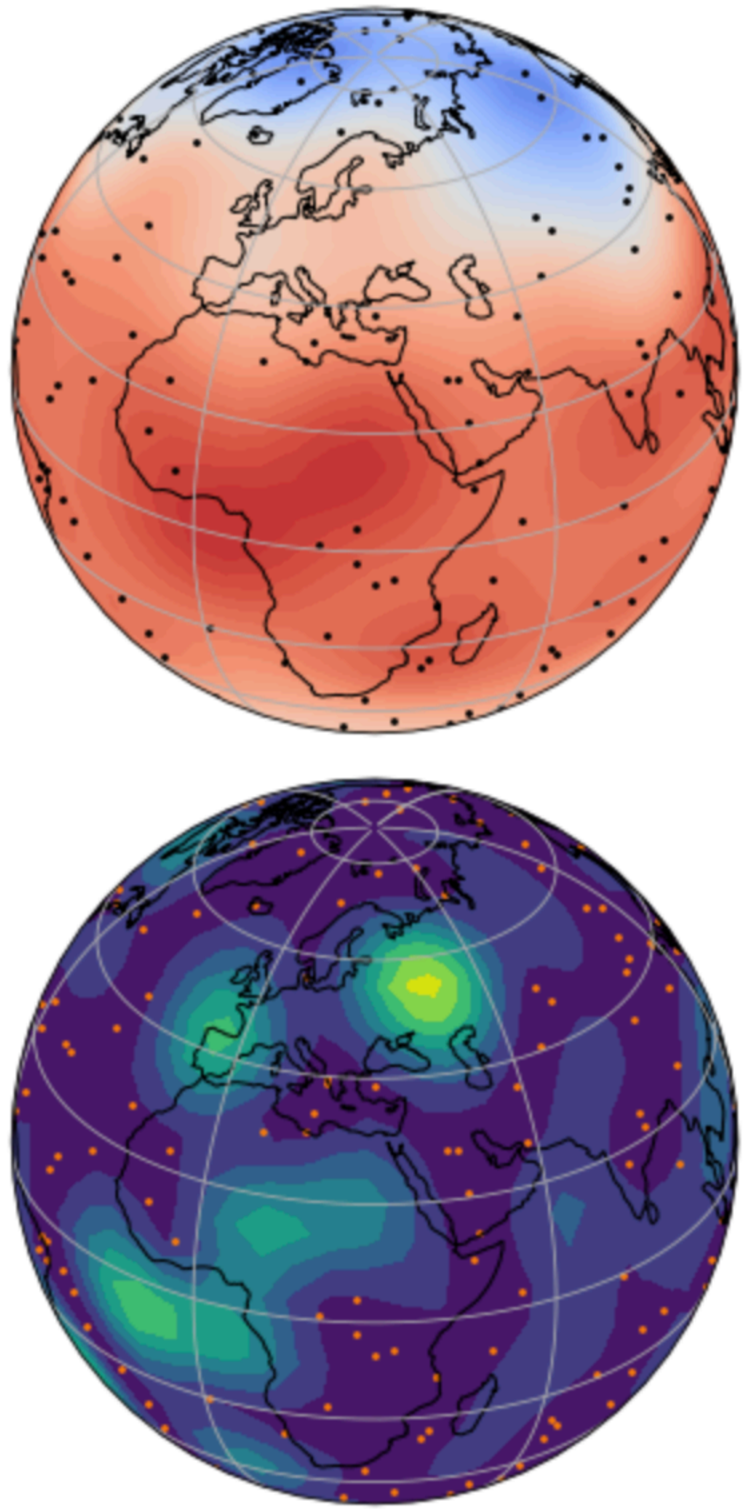
Global temperature interpolation



2m temperature from ERA5



Euclidean Matérn GP



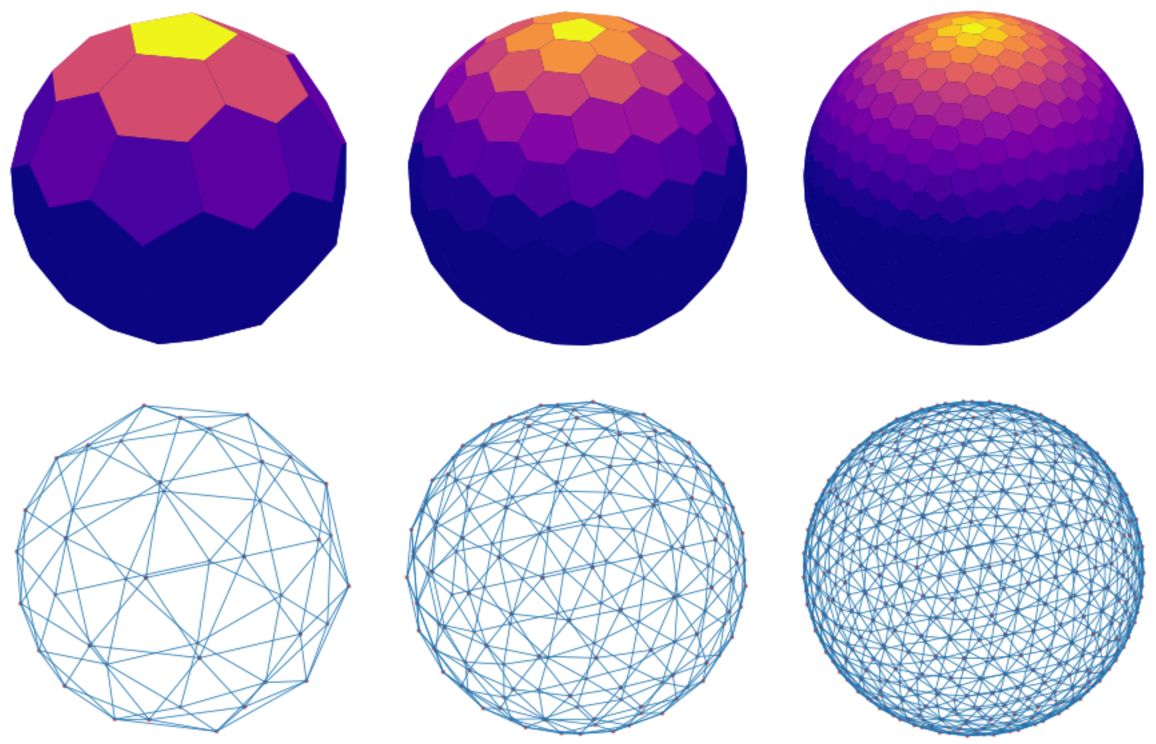
Spherical Matérn GP

Graph approximations

Recall expression for Riemannian Matérn kernel:

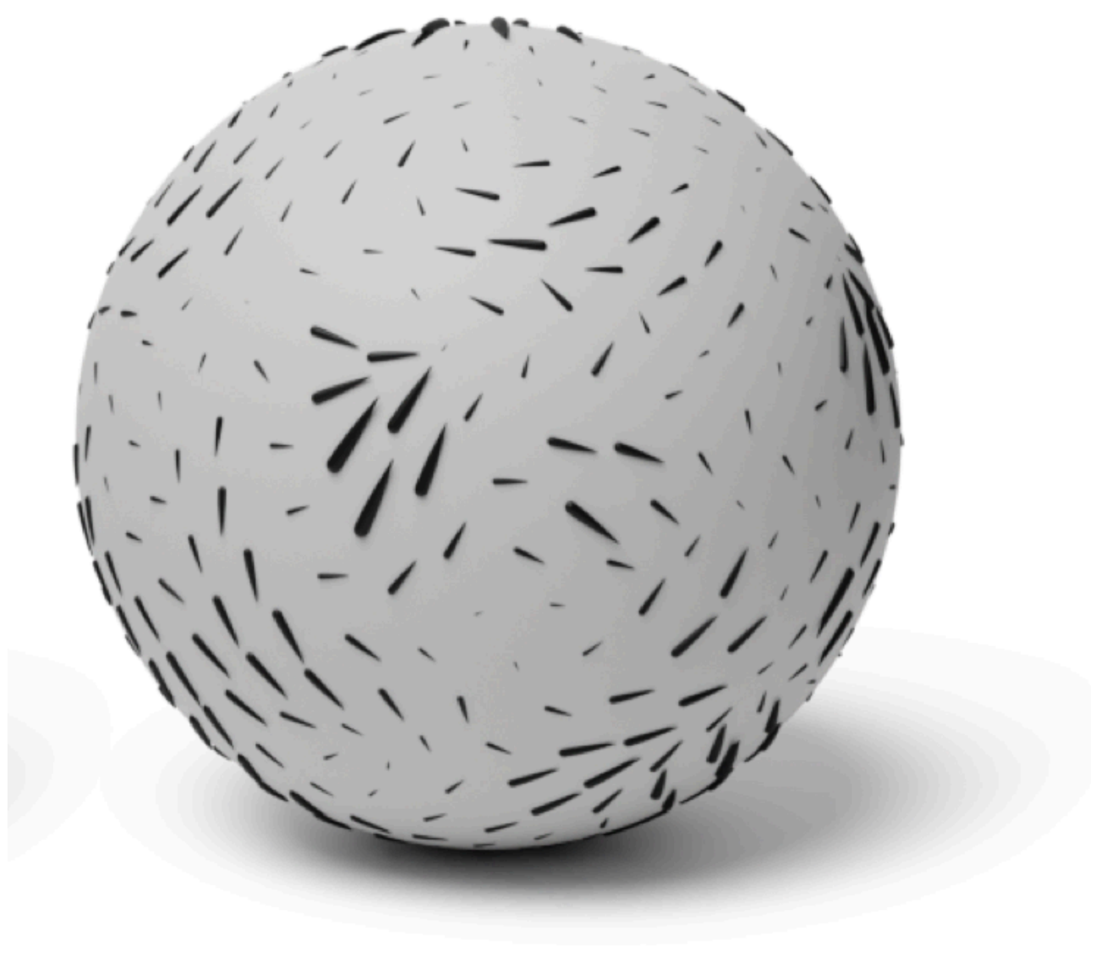
$$k(x, x') = \frac{\sigma^2}{C_{\nu,\ell}} \sum_{n=0}^{\infty} \left(\frac{2\nu}{\ell^2} + \lambda_n \right)^{-\nu-d/2} \phi_n(x)\phi_n(x')$$

We can use $\{(\lambda_n, \phi_n)\}_{n=0}^{\infty}$ of the *graph Laplacian* instead!

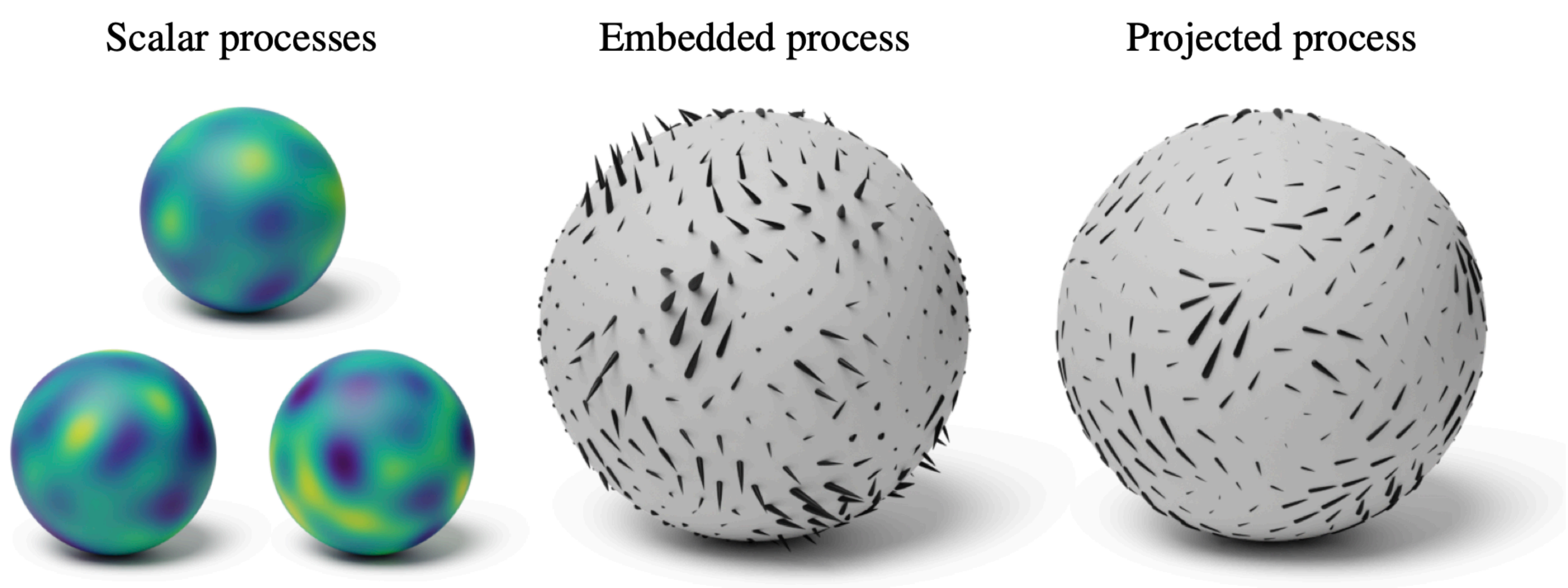


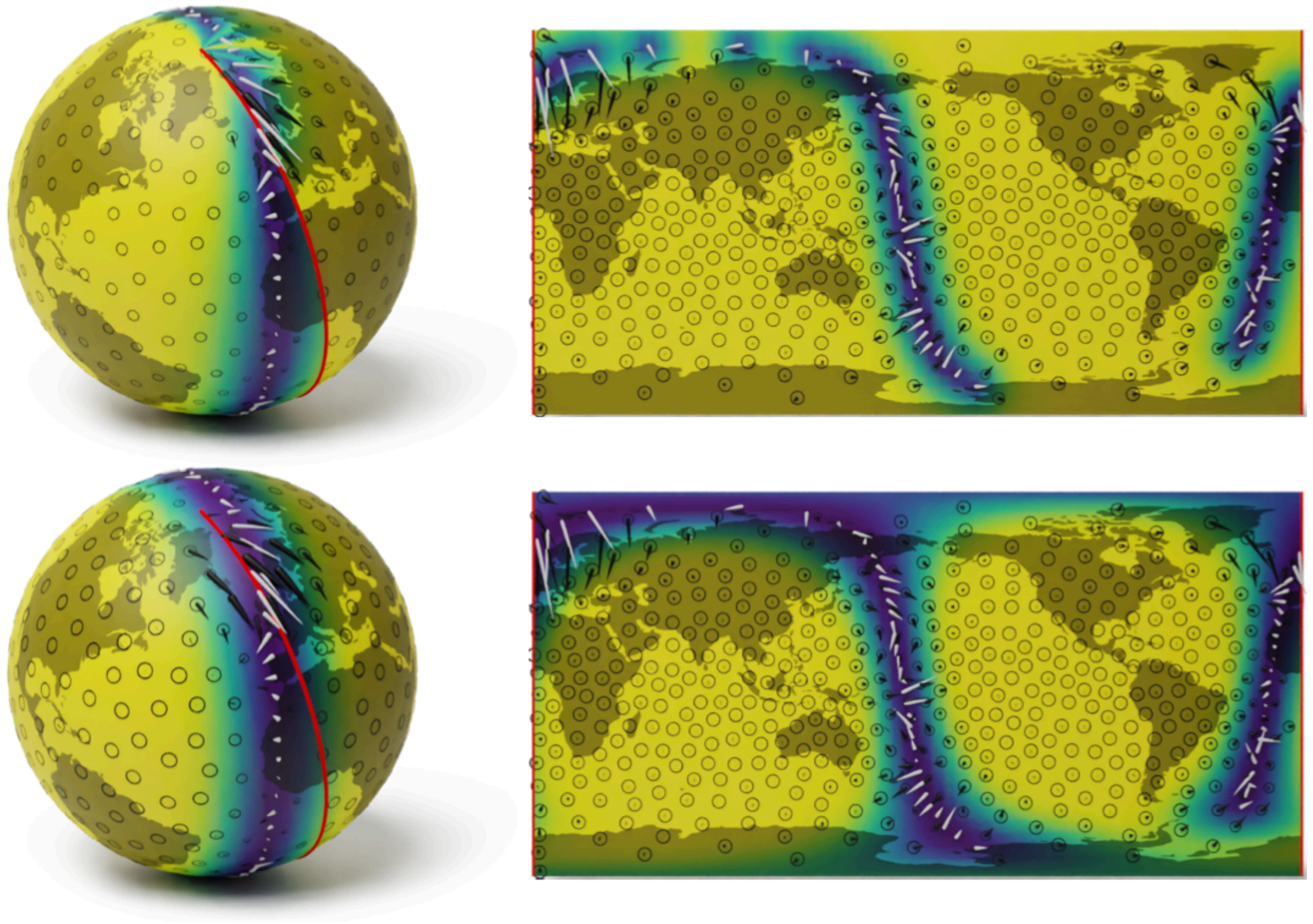
*Borovitskiy et. al. "Matérn Gaussian processes on Riemannian manifolds." *NeurIPS* (2020)

2. Matérn Gaussian vector fields



Gauge-equivariant vector GP



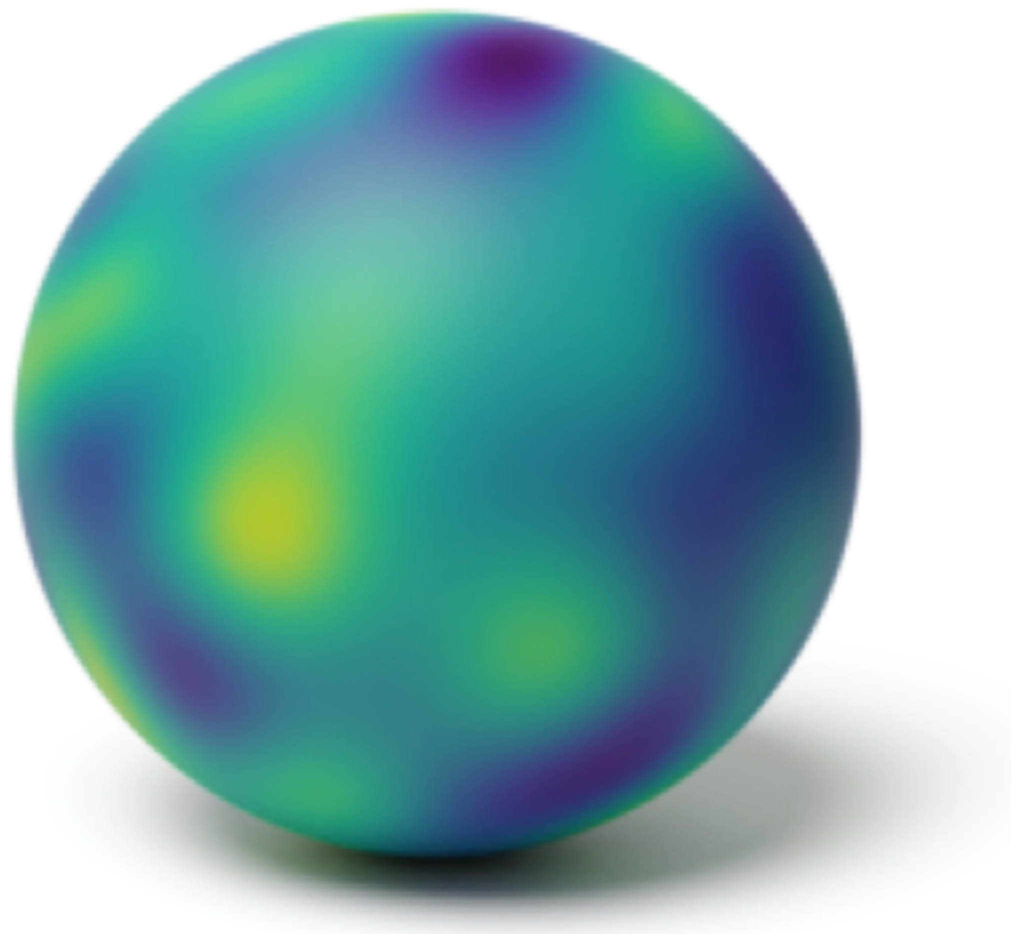


Combine multiple scalar-valued GPs and project onto tangent spaces to get Gaussian random vector fields.

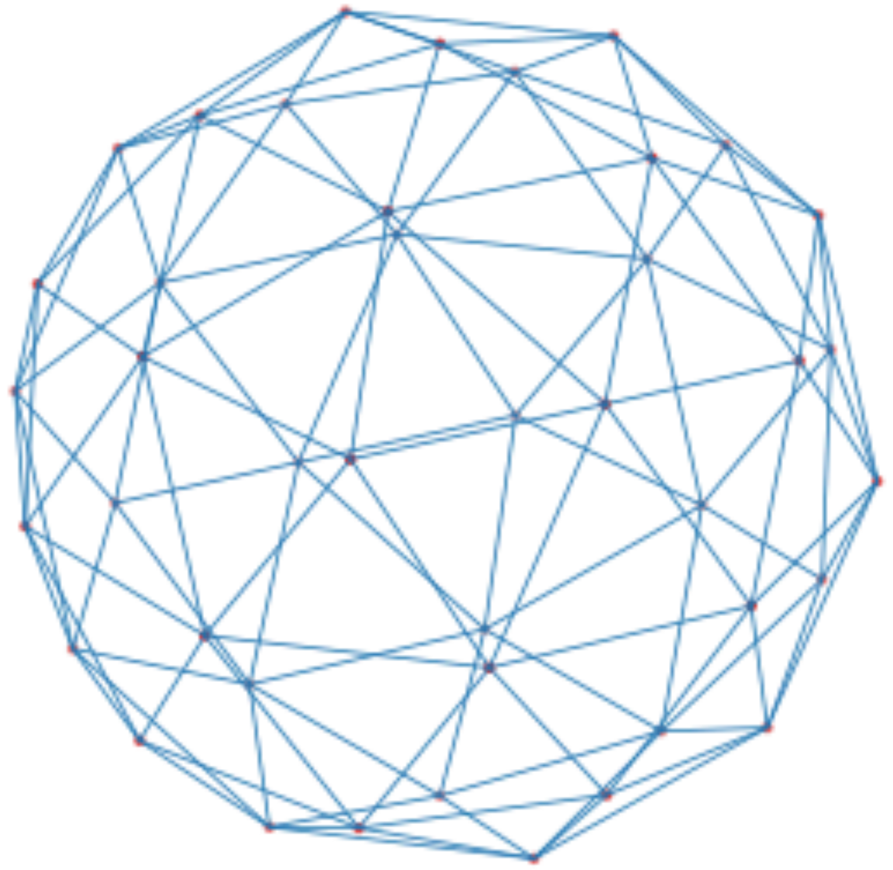
However, requires
1. An isometric embedding to an ambient space
2. A choice of *gauge field*

Hutchinson et al. "Vector-valued Gaussian Processes on Riemannian Manifolds via Gauge Equivariant Projected Kernels." *NeurIPS 2021*

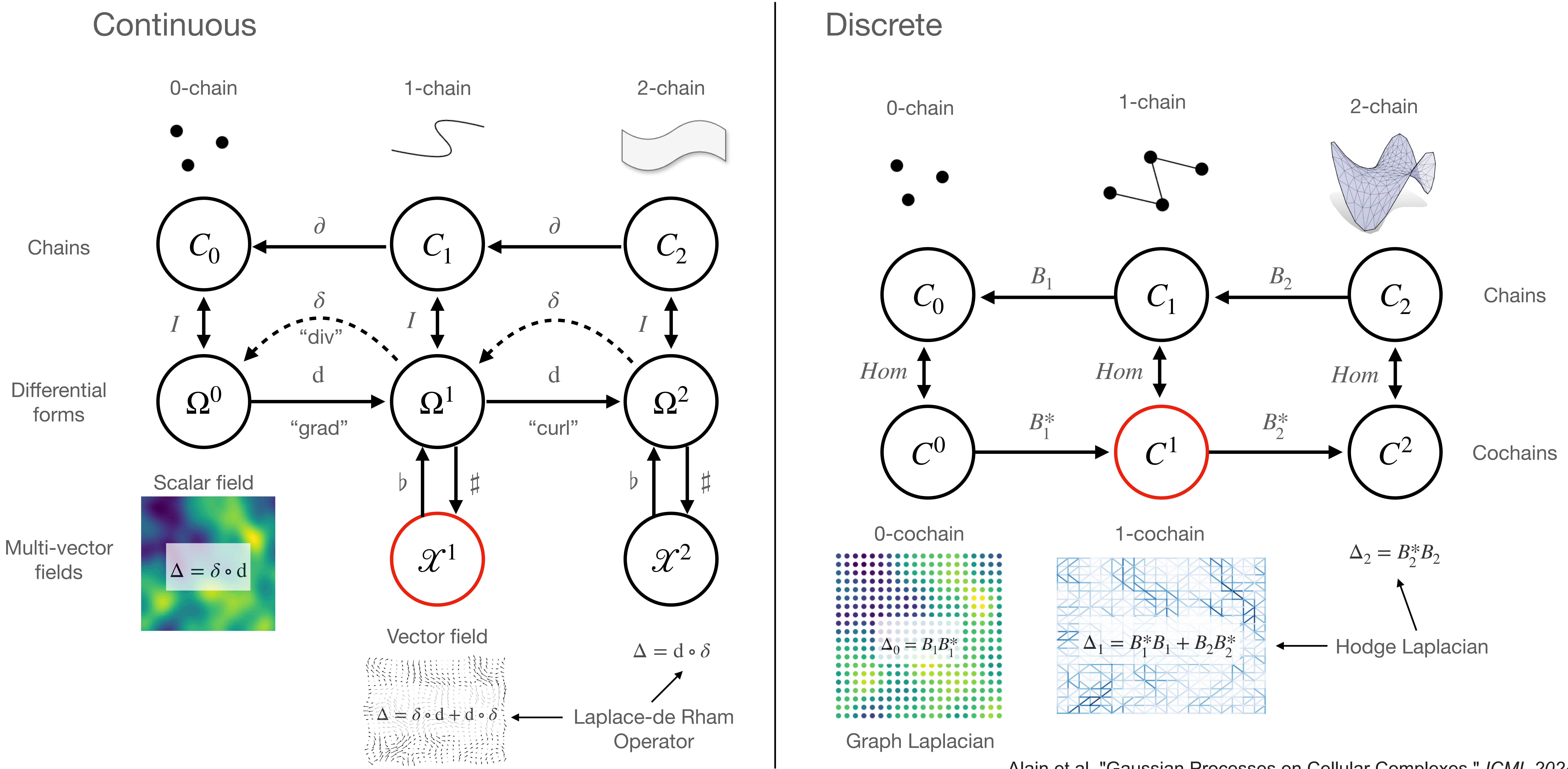
Manifolds



Graphs



Chasin' the gram: some homological algebra



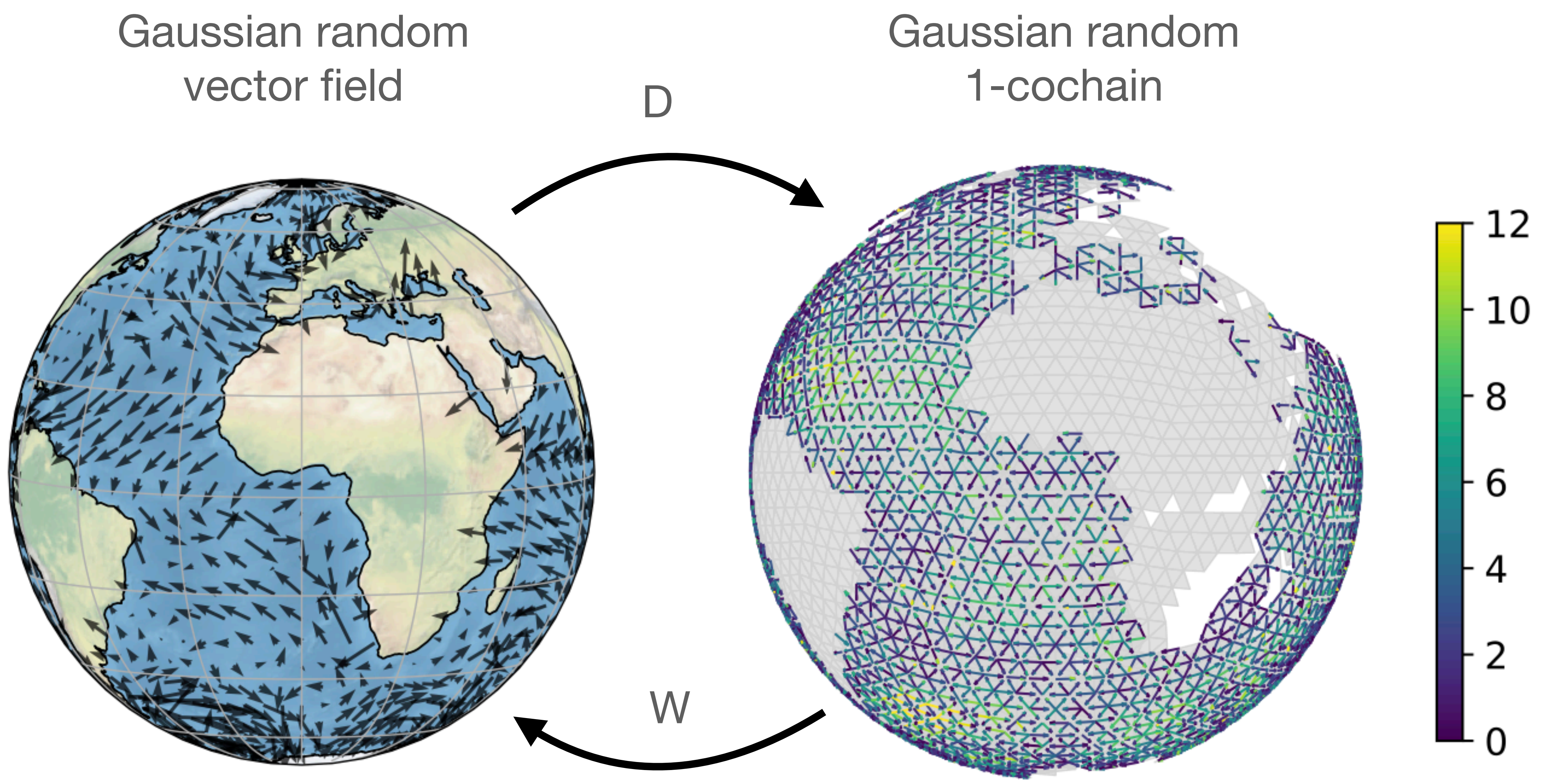
Alain et al. "Gaussian Processes on Cellular Complexes." *ICML 2024*

Gaussian random cochains

We define a Gaussian random k-cochain as a random group homomorphism

$$f : C_k \to \mathbb{R},$$

such that $f(c)$ is Gaussian for all $c \in C_k$.



Project vector field on edges of a graph

Alain et al. "Gaussian Processes on Cellular Complexes." *ICML 2024*

Matérn random cochain

A Gaussian random k-cochain $f : C_k \to \mathbb{R}$ is of the Matérn class if it satisfies

$$\left(\frac{2\nu}{\ell^2} + \Delta_k\right)^{\frac{\nu + d/2}{2}} f = \dot{\mathscr{W}}_\sigma$$

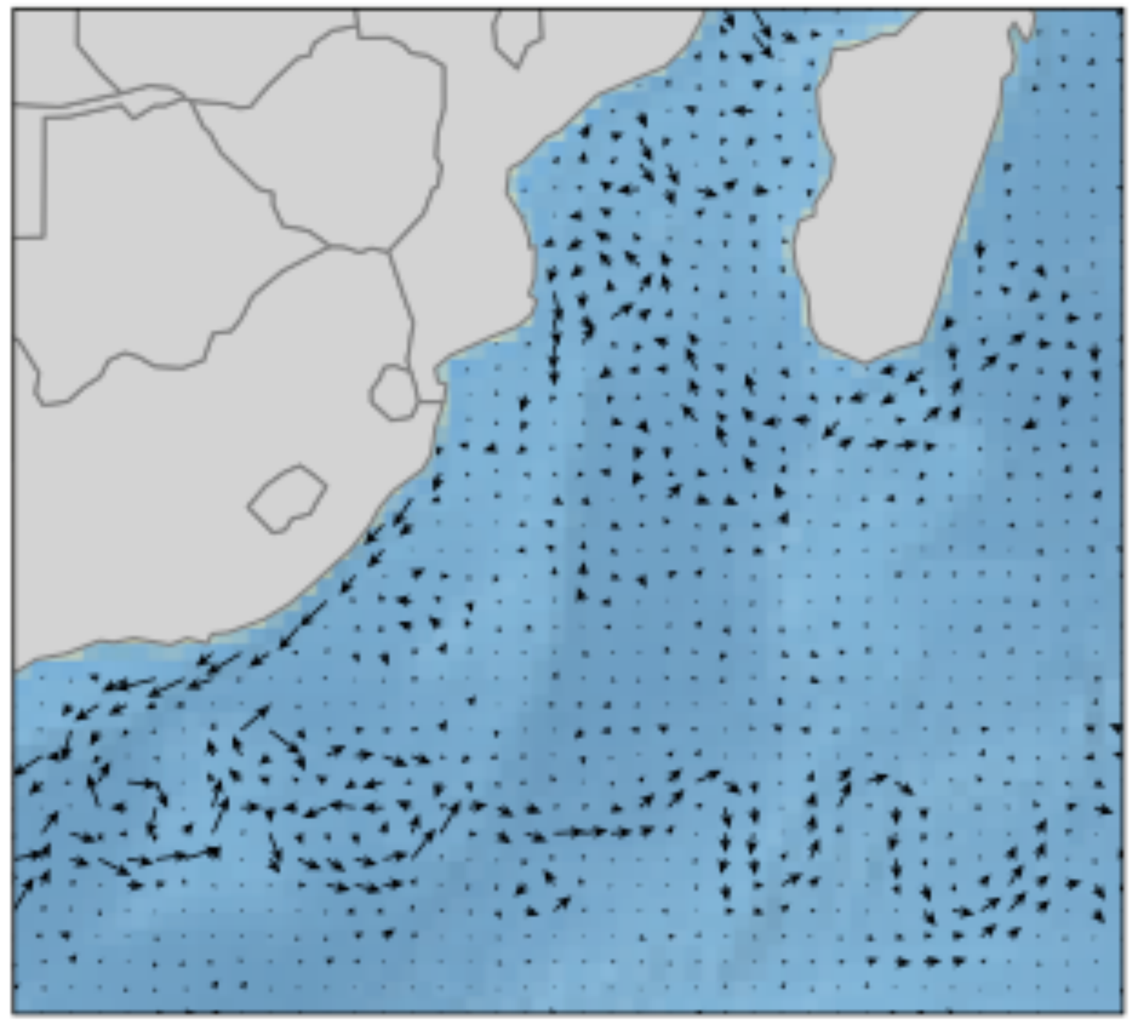
Kernel has the numerical representation

$$\mathbf{K} = \sigma^2 \mathbf{U} \left(\frac{2\nu}{\ell^2}\mathbf{I} + \mathbf{\Lambda}\right)^{-\nu - d/2} \mathbf{U}^\top,$$

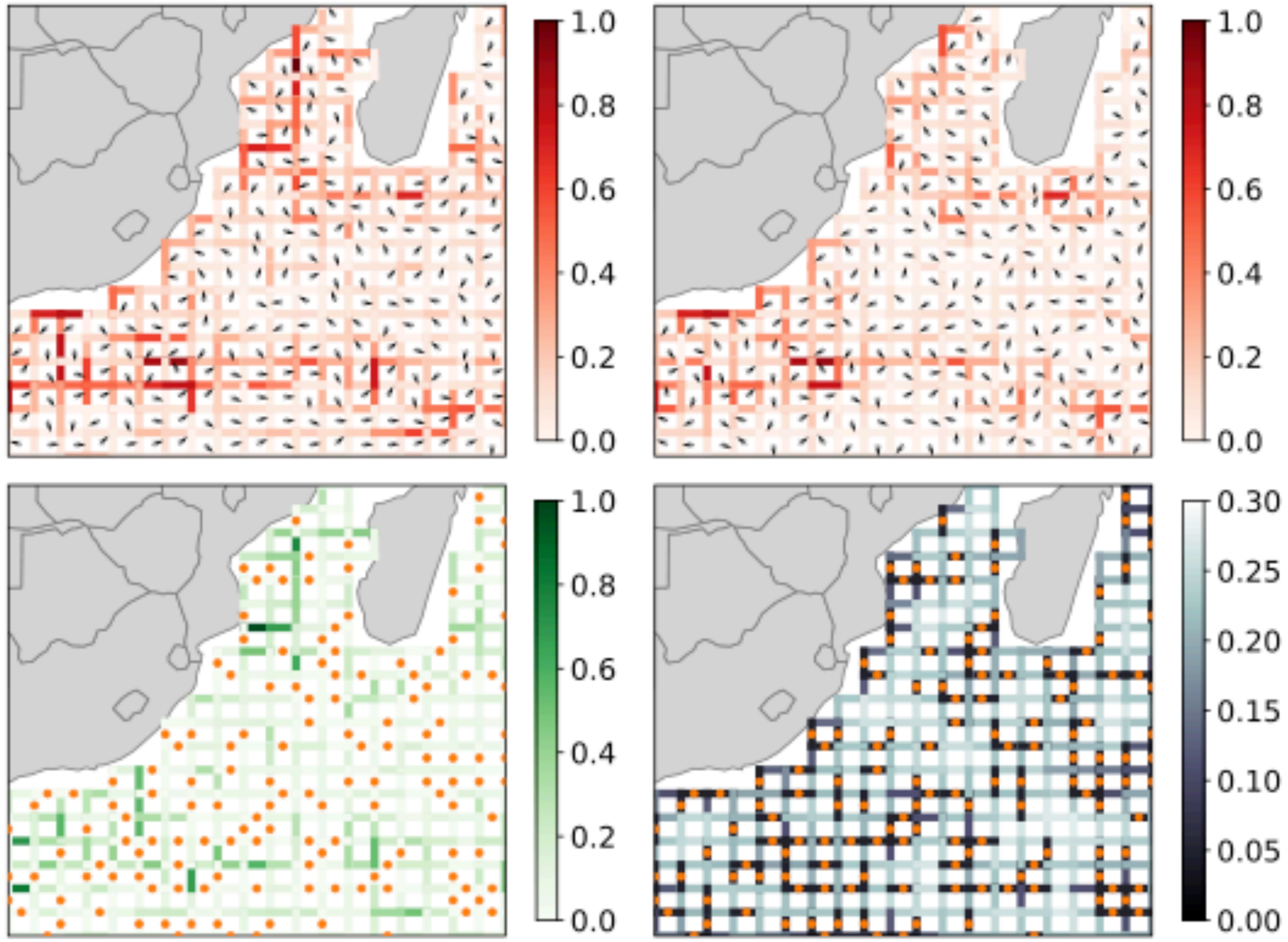
where $\Delta_k = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^\top$ is the eigen-decomposition of the Hodge Laplacian.

$$k(x, x') = \sigma^2 \sum_{n=0}^{\infty} \left(\frac{2\nu}{\ell^2} + \lambda_n\right)^{-\nu - d/2} \phi_n(x)\phi_n(x')$$

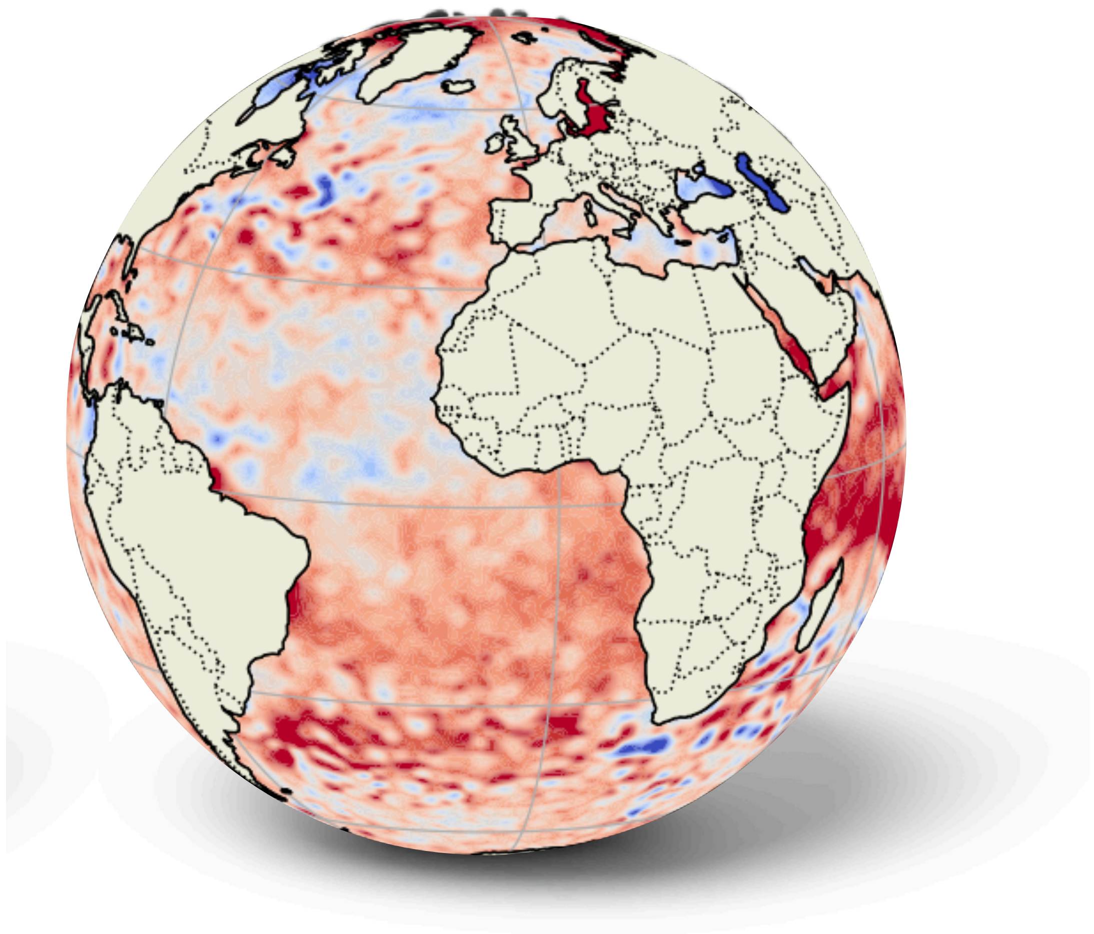
Ocean current interpolation



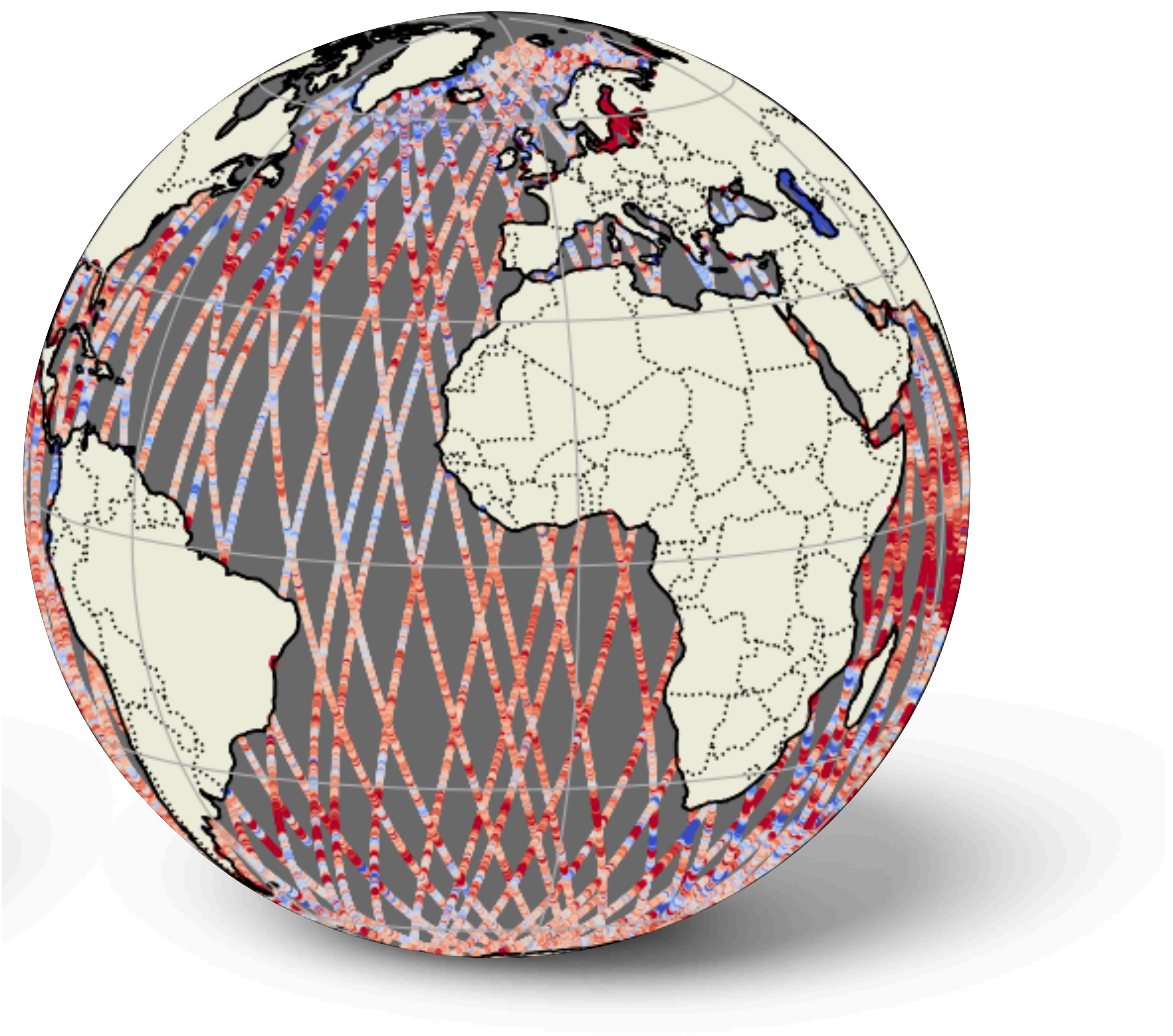
(a) Geostrophic current



3. Deep Random Features



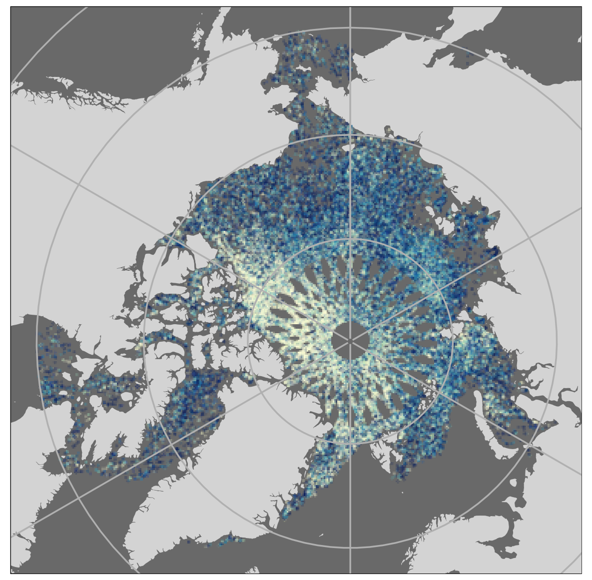
Limitations of Gaussian processes



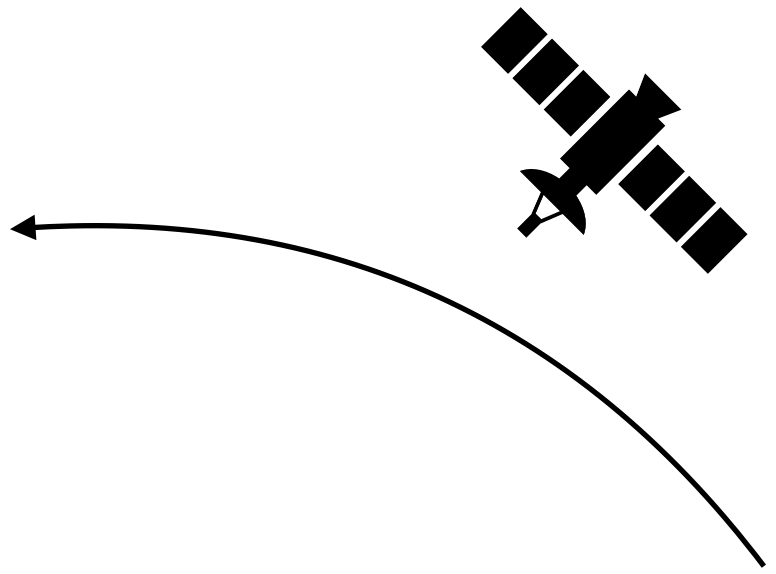
Sea-level observations from Sentinel 3A and 3B
8,094,569 data points

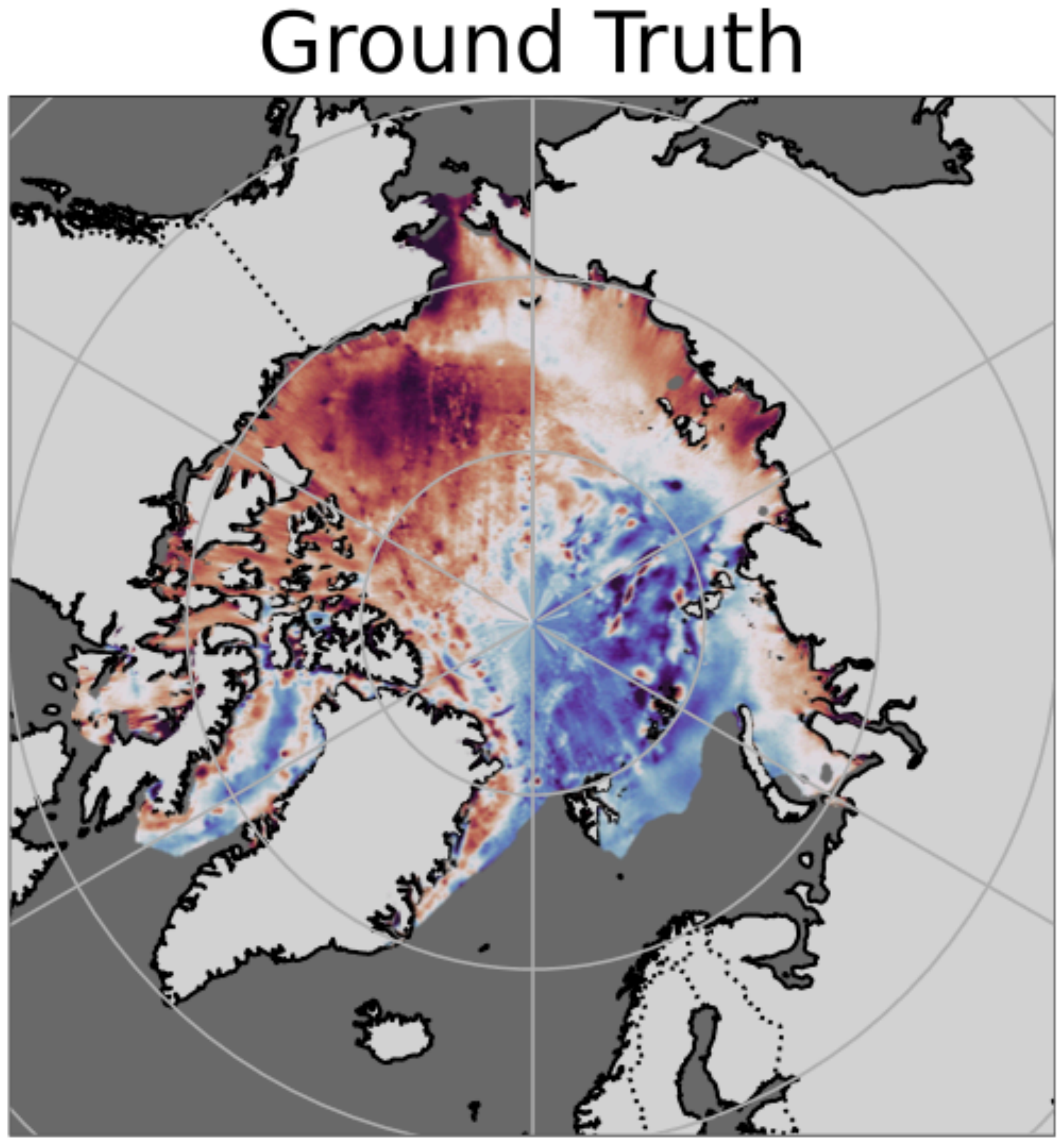
- Cost increases cubically with data
- Stationarity often assumed
- It’s Gaussian

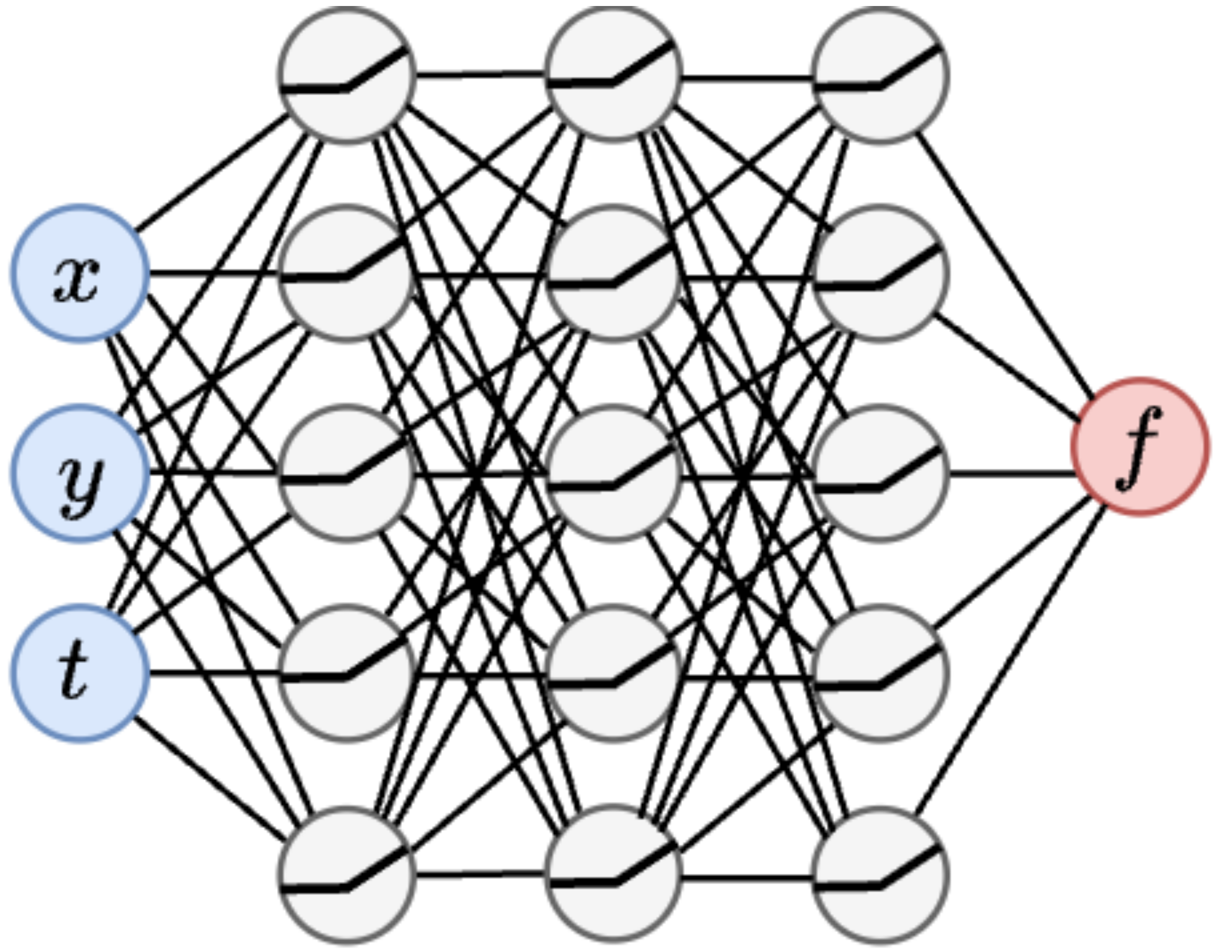
Can we use neural networks instead?

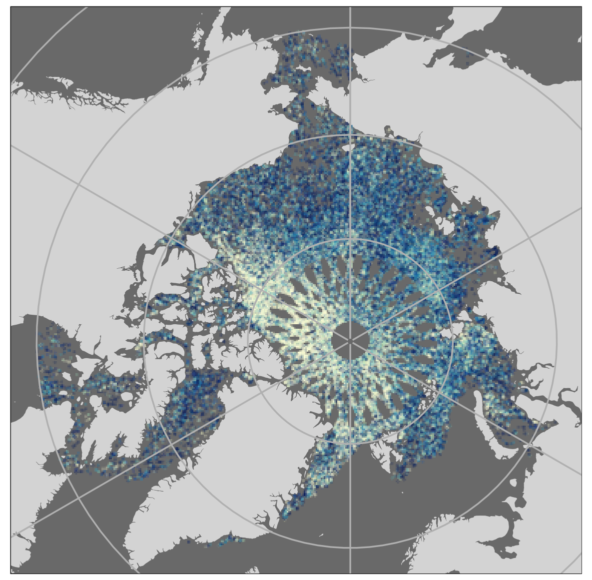


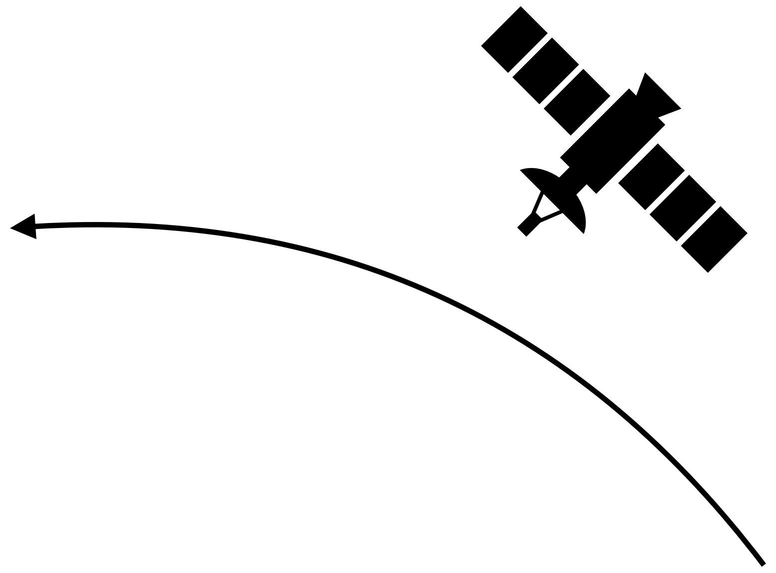
Inputs: (x, y, t)
Outputs: $f(x, y, t) + \epsilon$





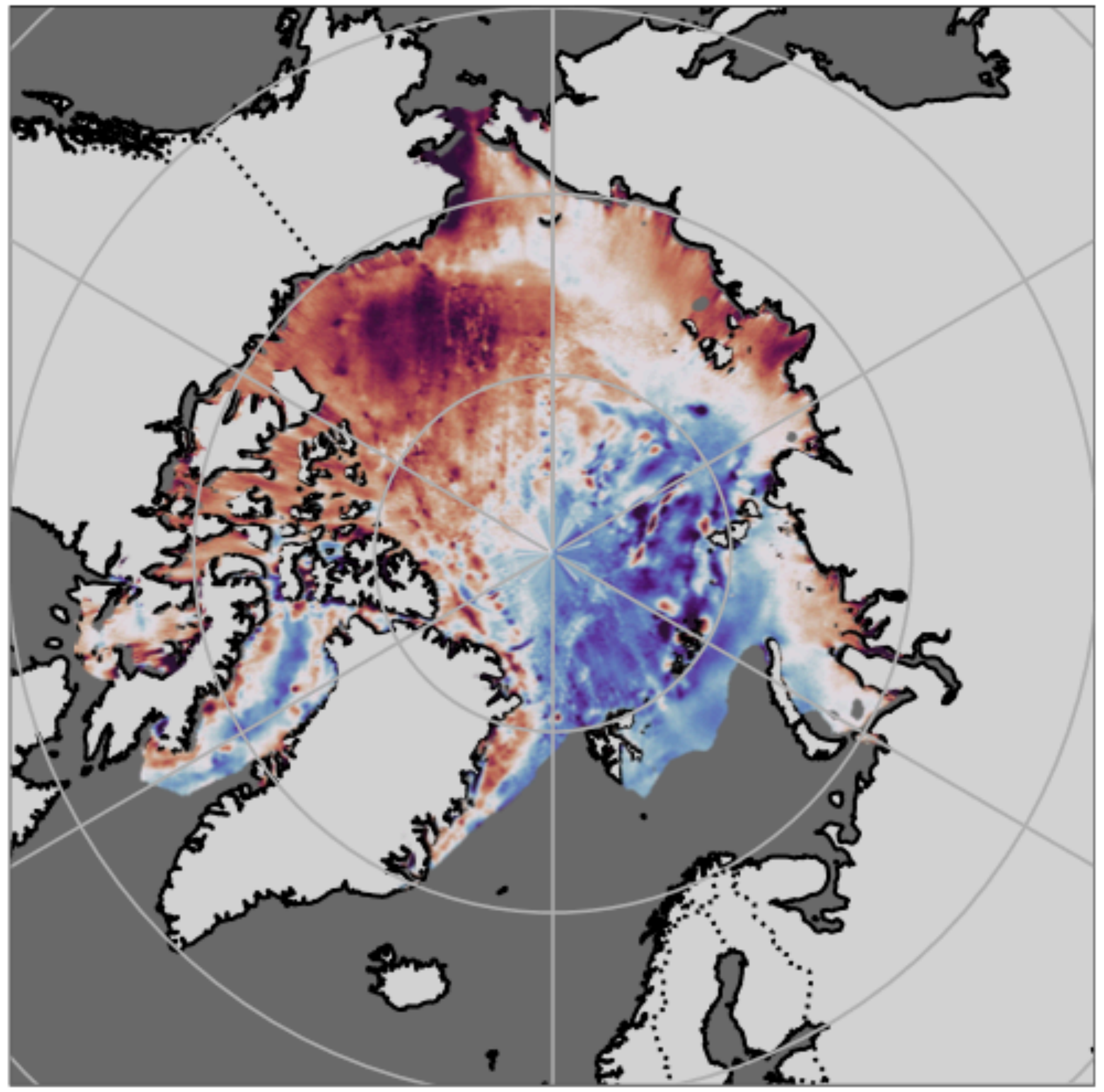




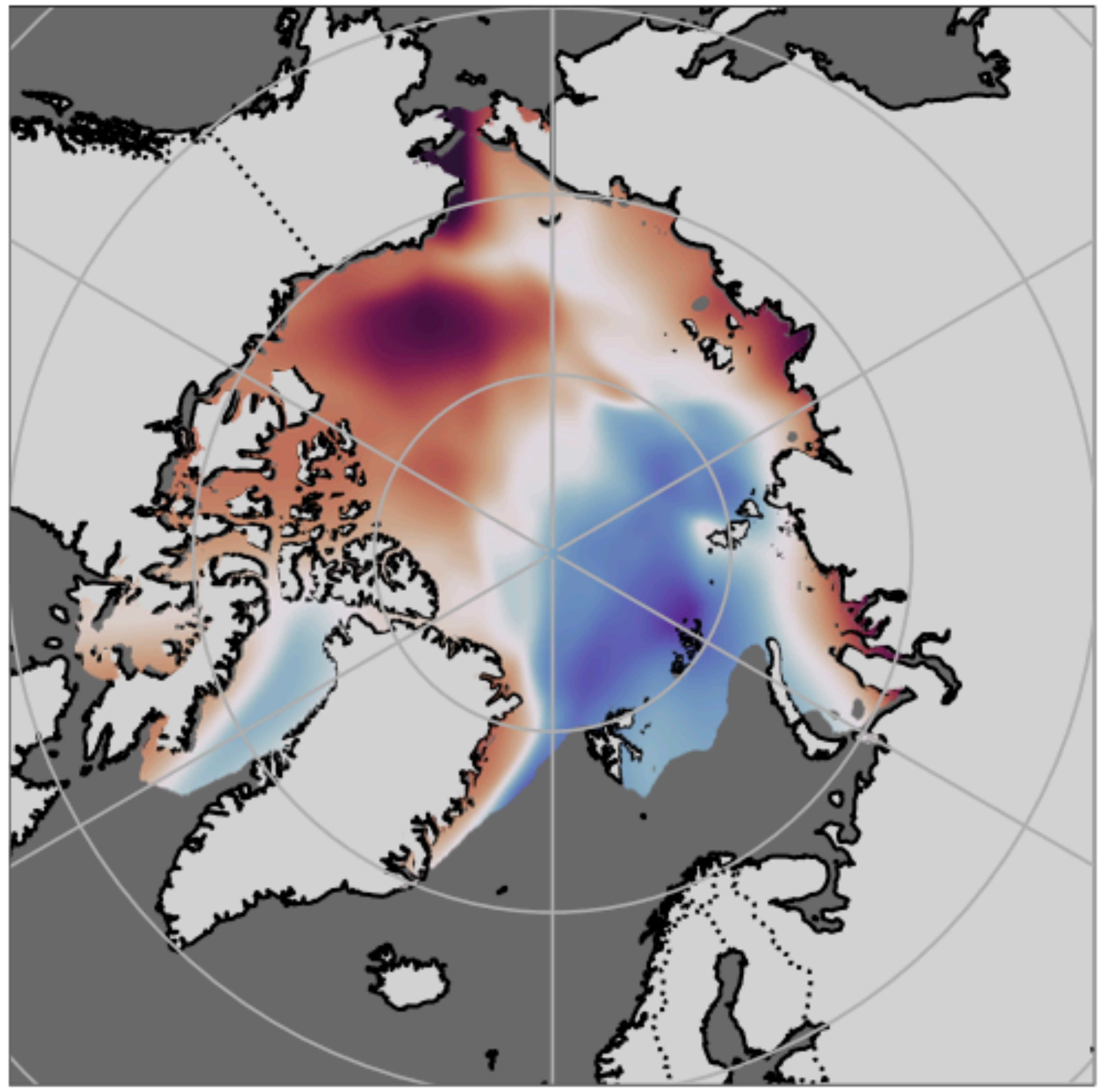


Inputs: (x, y, t)
Outputs: $f(x, y, t) + \epsilon$

Ground Truth



Mean: ReLU MLP

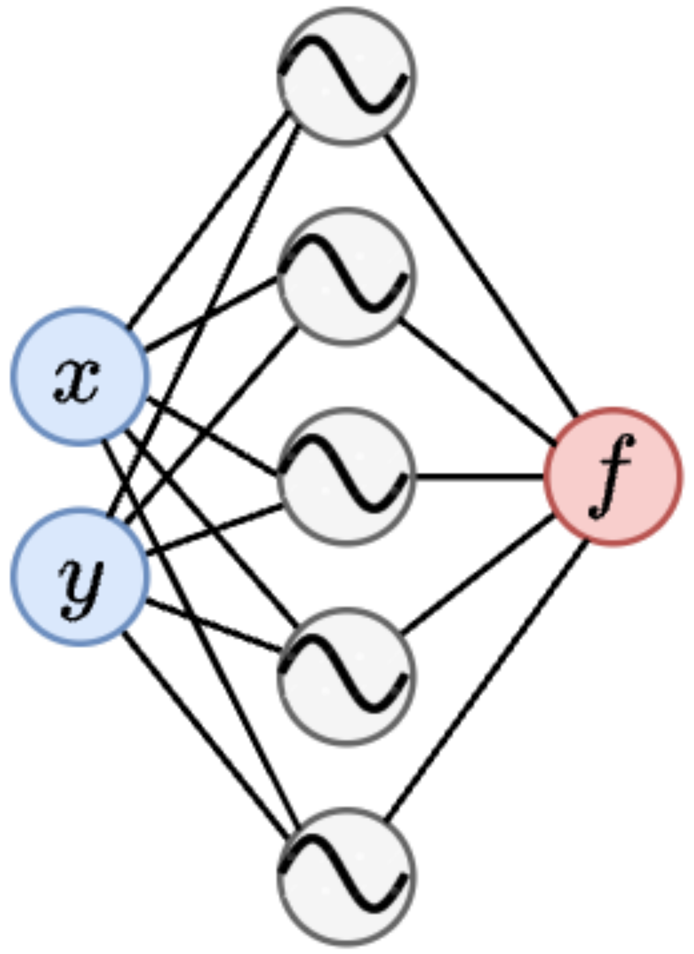


Random features revisited

Recall that a Matérn GP f has the random feature representation:

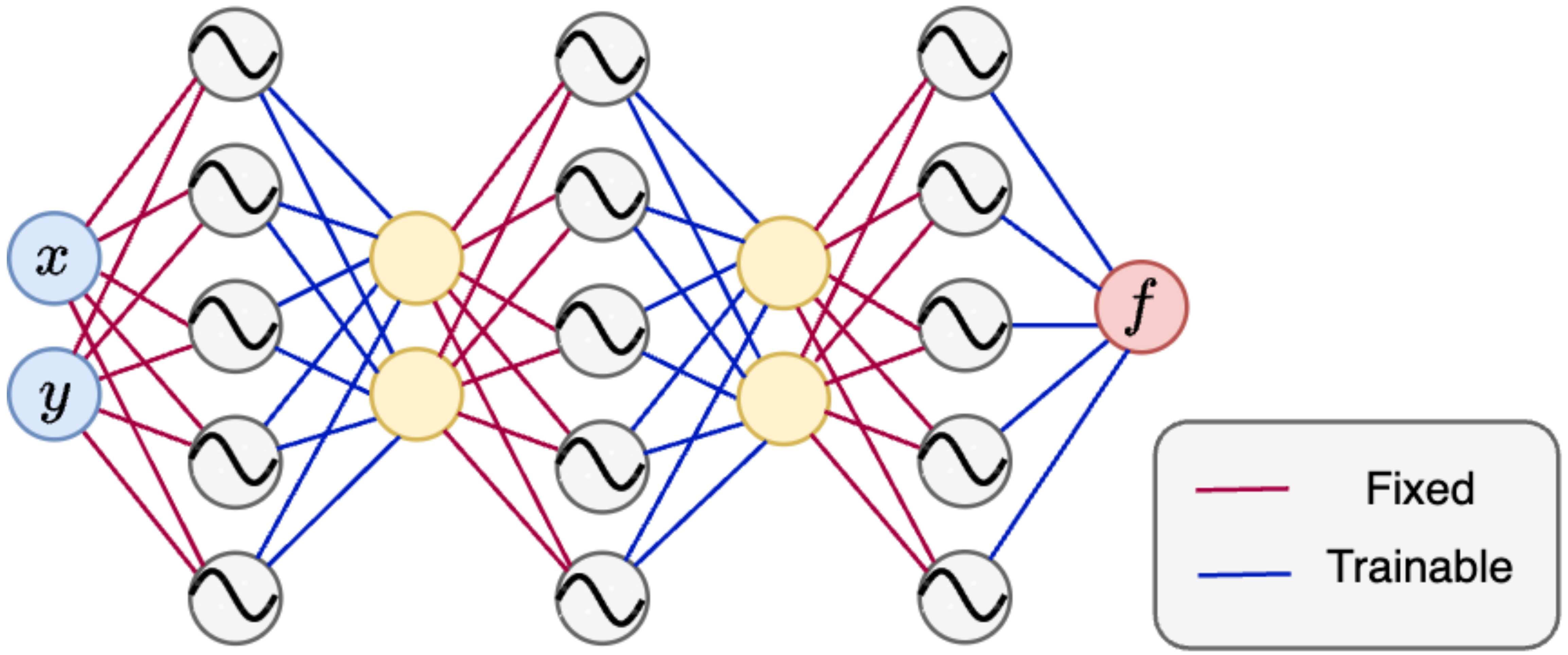
$$f(x) \approx \sum_{m=1}^{M} \theta_m \psi_m(x), \quad \text{where} \quad \psi_m(x) = \sqrt{\frac{2\sigma^2}{M}} \cos(\omega_m^\top x + b_m)\,.$$

$$\omega_m \sim t_{2\nu}(0, \ell^{-2}), \quad b_m \sim U([0, 2\pi]) \quad \text{and} \quad \theta_m \sim \mathcal{N}(0,1)\,.$$



Idea:

Make it deep!



Similar to the SIREN architecture.

Spherical Random Features

Recall the Riemannian Matérn kernel on $\mathbb{S}^2$:

$$k(x, x') = \frac{\sigma^2}{C_{\nu,\ell}} \sum_{n=0}^{\infty} \left(\frac{2\nu}{\ell^2} + \lambda_n \right)^{-\nu - d/2} \phi_n(x) \phi_n(x'),$$

where $\lambda_n = n(n+1)$ and $\phi_n(x) = \{Y_n^1(x), \ldots, Y_n^{2n+1}(x)\}$.

This implies *deterministic feature maps*

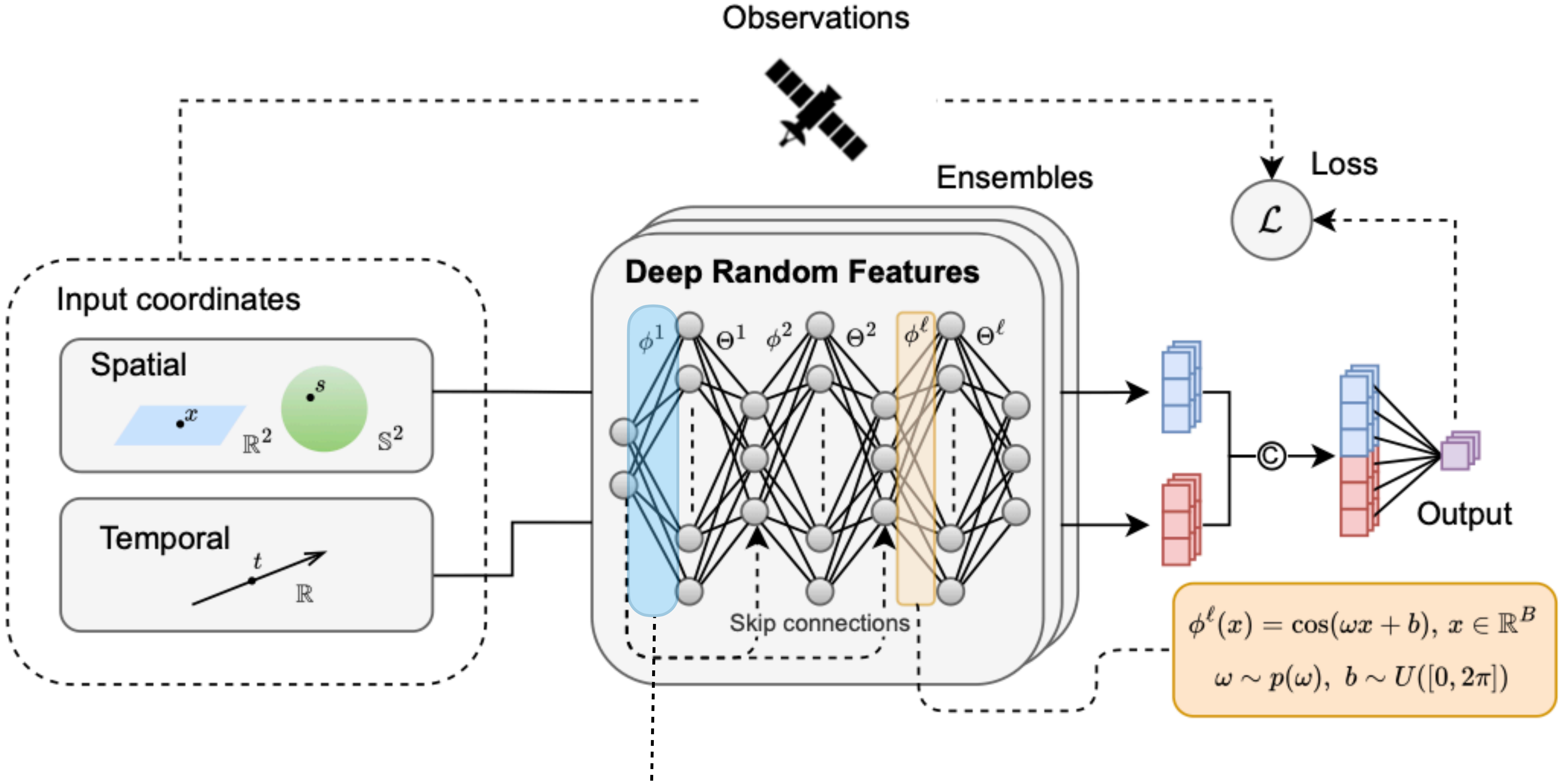
$$\psi_{n,m}(x) = \sqrt{\frac{\sigma^2}{C_{\nu,\ell}}} \left(\frac{2\nu}{\ell^2} + n(n+1) \right)^{\frac{-\nu - d/2}{2}} Y_n^m(x).$$

$$
\begin{aligned}
k(x,x') &\approx \frac{Z_N\sigma^2}{C_{\nu,\ell}} \sum_{n=0}^{N} \sum_{m=1}^{2n+1} \rho(n) Y_n^m(x) Y_n^m(x'), \quad \text{where} \quad \rho(n) := \frac{1}{Z_N}\left(\frac{2\nu}{\ell^2} + n(n+1)\right)^{-\nu-d/2} \\
&= \frac{Z_N\sigma^2}{C_{\nu,\ell}} \sum_{n=0}^{N} \rho(n) \left(\sum_{m=1}^{2n+1} Y_n^m(x) Y_n^m(x') \right) \\
&\propto \sum_{n=0}^{N} \rho(n) \mathcal{G}_n^{1/2}(\cos(\langle x, x' \rangle)) \\
&= \sum_{n=0}^{N} \rho(n) \int_{\mathbb{S}^2} \mathcal{G}_n^{1/2}(\cos(\langle x, b \rangle)) \mathcal{G}_n^{1/2}(\cos(\langle x', b \rangle)) \, \mathrm{d}b \\
&= \mathbb{E}_{\omega,b}\left[\mathcal{G}_\omega^{1/2}(\cos(\langle x, b \rangle)) \mathcal{G}_\omega^{1/2}(\cos(\langle x', b \rangle))\right], \quad \omega \sim \text{Multinomial}(\rho(1), \dots, \rho(N)), \quad b \sim U(\mathbb{S}^2) \\
&\approx \frac{1}{J} \sum_{j=1}^{J} \mathcal{G}_{\omega_j}^{1/2}(\cos(\langle x, b_j \rangle)) \mathcal{G}_{\omega_j}^{1/2}(\cos(\langle x', b_j \rangle)), \quad \omega_j \sim \text{Multinomial}(\rho(1), \dots, \rho(N)), \quad b_j \sim U(\mathbb{S}^2)
\end{aligned}
$$

This implies spherical random features

$$
\psi_j(x) \propto \mathcal{G}_{\omega_j}^{1/2}(\cos(\langle x', b_j \rangle)).
$$

Deep random features

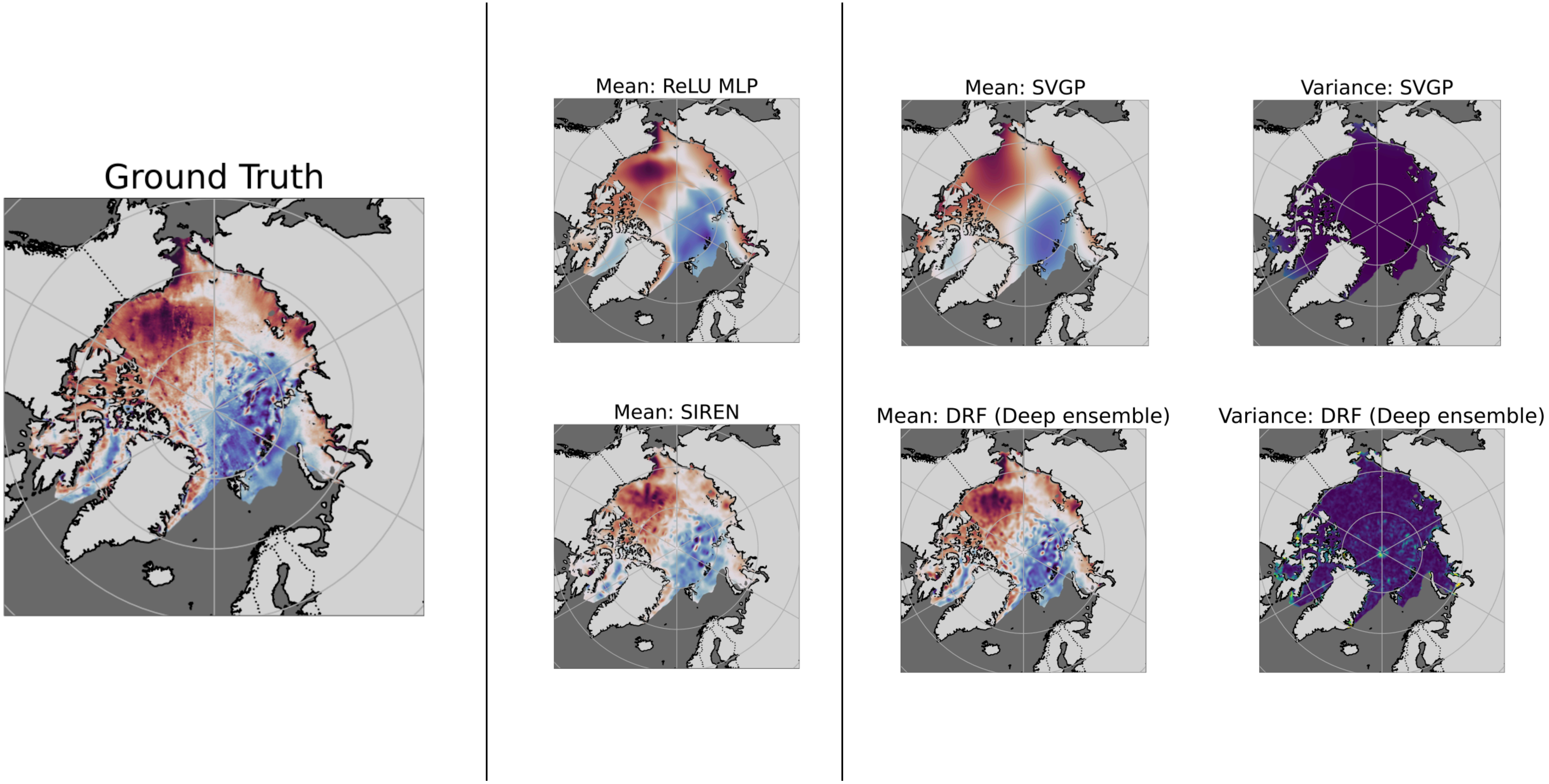


Notes:

- Use deep ensembles to estimate uncertainty
- Use batched stochastic gradient optimisation
- Tune kernel hyperparameters on validation set

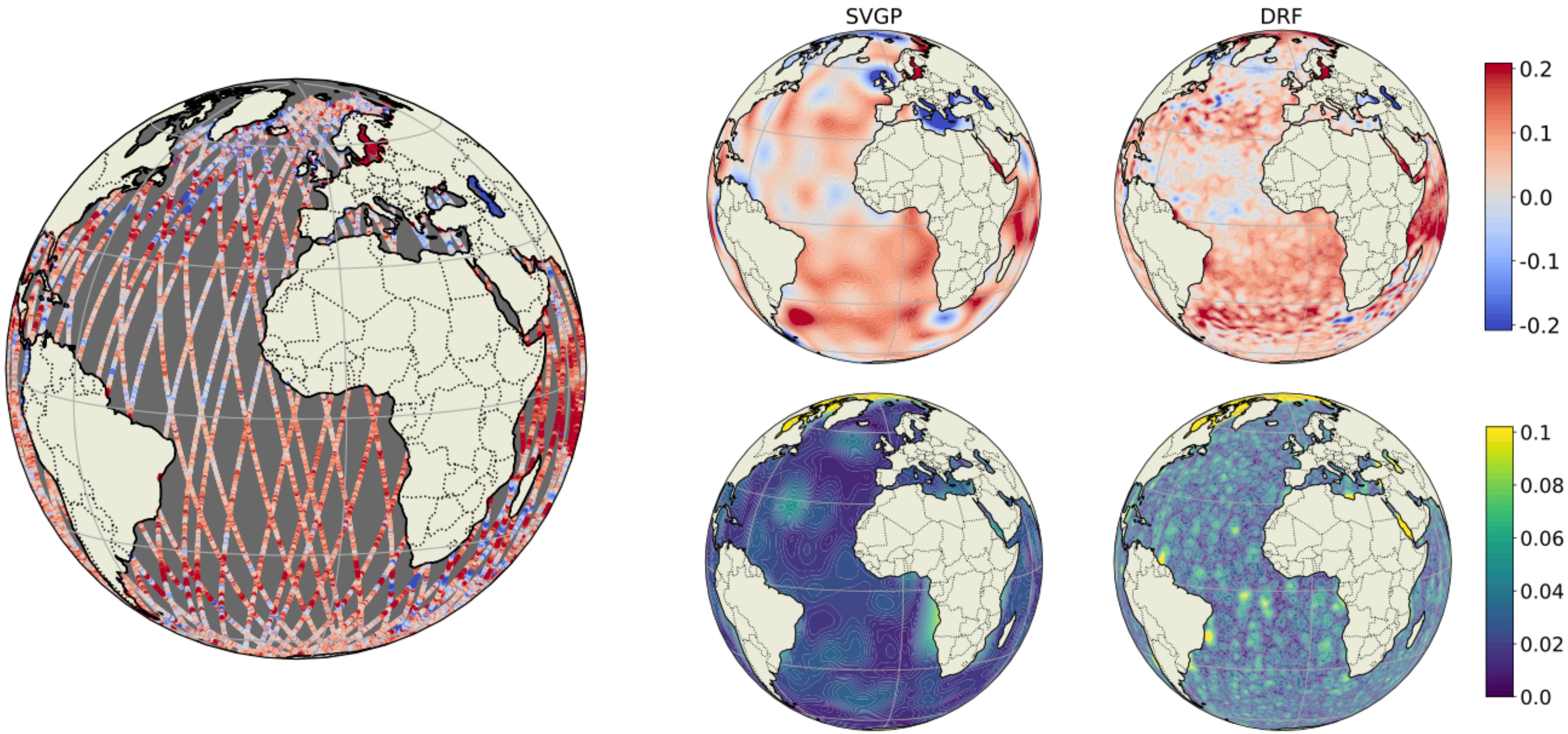
If $x \in \mathbb{R}^d$, $\phi^1(x) = \cos(\omega x + b)$, where $\omega \sim t_{2\nu}(0, \ell^{-2})$, $b \sim U([0, 2\pi])$

If $x \in \mathbb{S}^2$, $\phi^1(x) = \mathcal{G}_\omega^{1/2}(\cos(\langle x, b \rangle))$, where $\omega \sim$ Multinomial$(\rho(1), \ldots, \rho(N))$, $b \sim U(\mathbb{S}^2)$



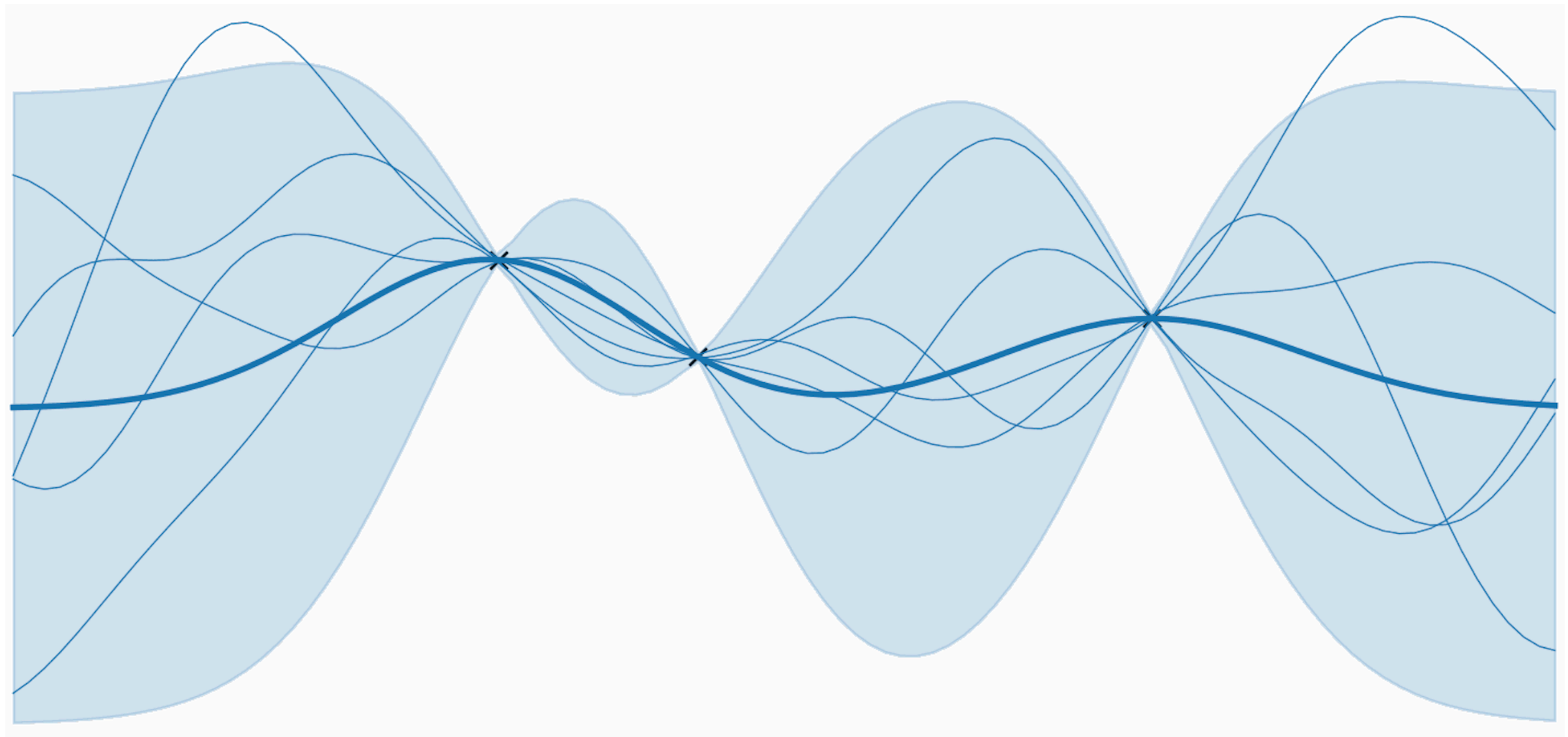
Ground Truth
Mean: ReLU MLP
Mean: SIREN
Mean: SVGP
Variance: SVGP
Mean: DRF (Deep ensemble)
Variance: DRF (Deep ensemble)

Global sea level interpolation



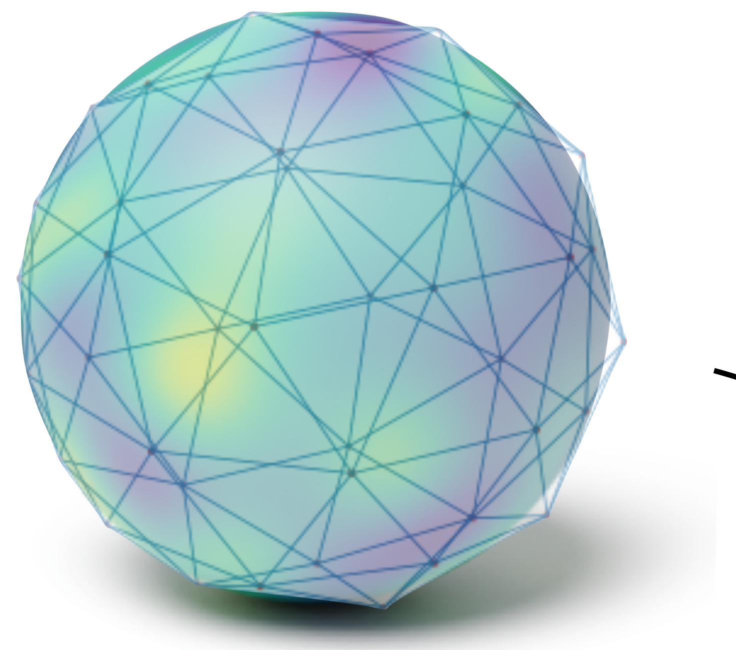
Summary

Summary

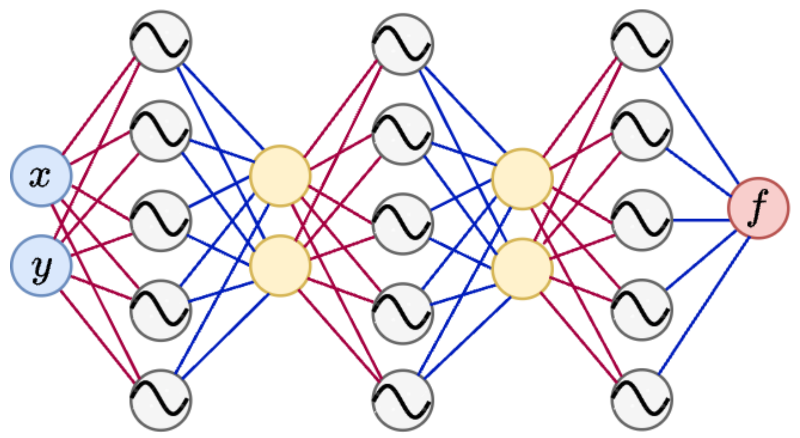


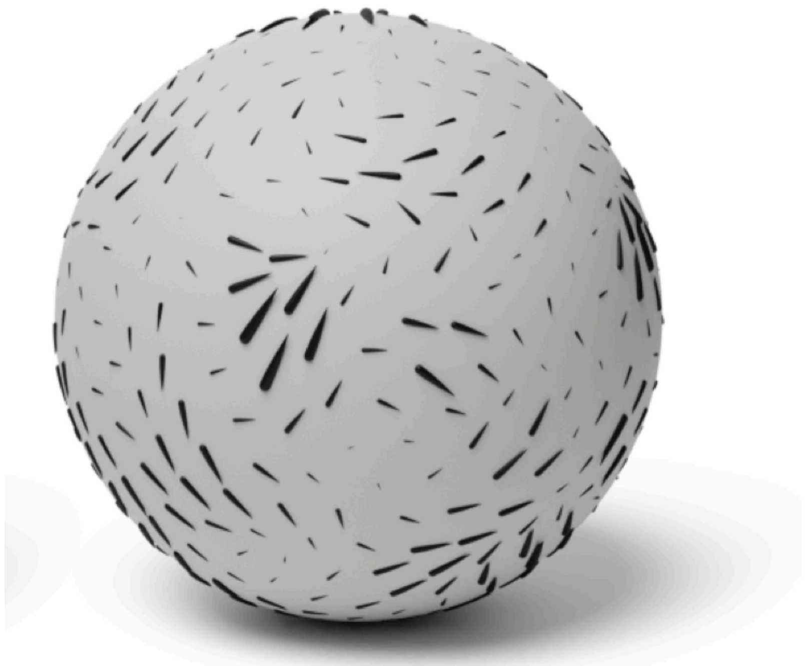
Introduced Matérn Gaussian processes
1. SPDE formulation
2. Weight-space perspective / random features

$$\left(\frac{2\nu}{\ell^2} - \Delta\right)^{\frac{\nu+d/2}{2}} f = \dot{\mathscr{W}}_\sigma$$

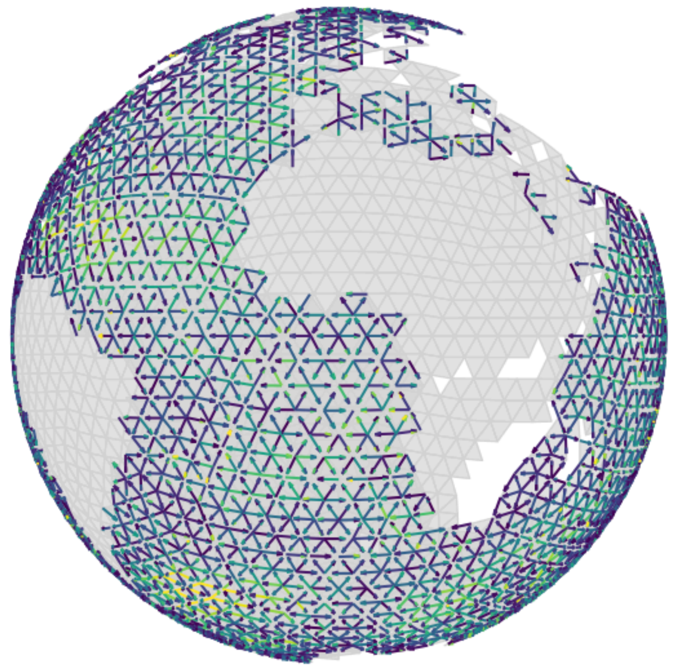


Extend to Riemannian manifolds / graphs via the SPDE framework

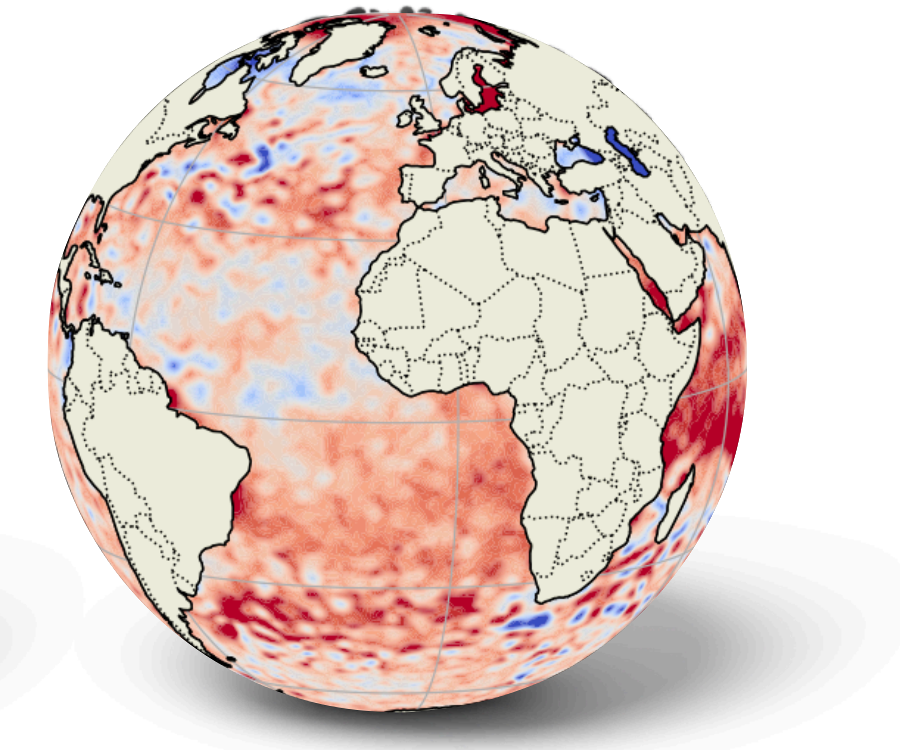




Extended to vector fields by combining multiple Riemannian Matérn GPs in embedding space



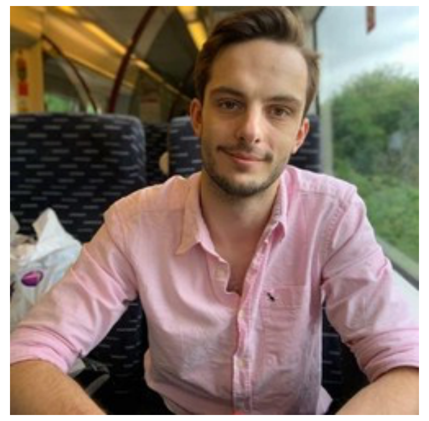
Embedding-free approach to modelling vector fields using Matérn cochains



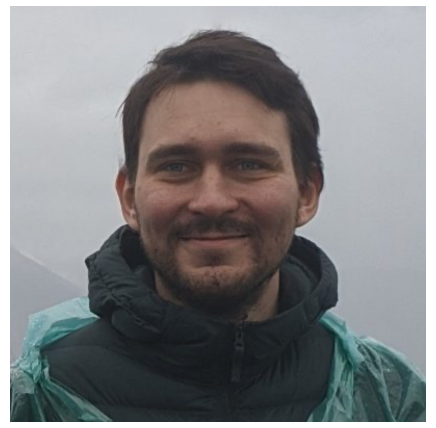
Stacked random feature representations of Matérn GPs

Thank you collaborators

Hutchinson et al. "Vector-valued Gaussian Processes on Riemannian Manifolds via Gauge Equivariant Projected Kernels." *NeurIPS 2021*


Michael Hutchinson


Alex Terenin


Slava Borovitskiy


Yee Whye Teh

Alain et al. "Gaussian Processes on Cellular Complexes." *ICML 2024*


Mathieu Alain


Brooks Paige


Marc Deisenroth

Chen et al. “Deep Random Features for Scalable Interpolation of Spatiotemporal Data.” *Submitted to ICLR 2025*


Weibin Chen


Azhir Mahmood


Michel Tsamados