Non-Gaussian Emulation of Climate Models via Scalable Bayesian Transport Maps

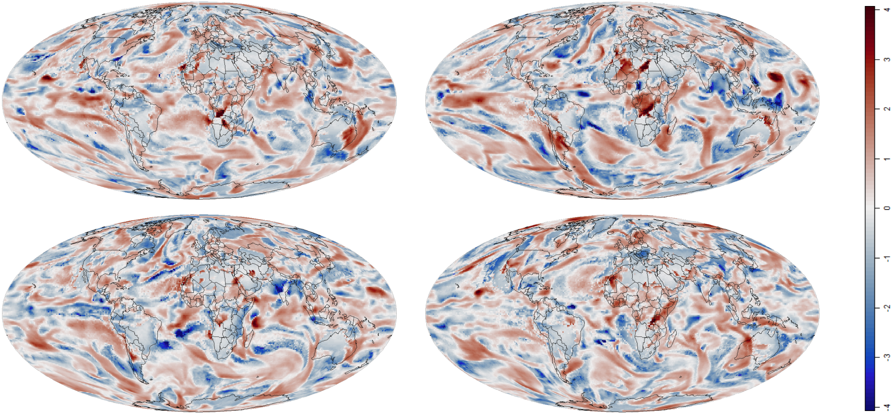
Matthias Katzfuss

Department of Statistics
University of Wisconsin–Madison

Outline

Outline

Motivating example: Climate-model output



- log precipitation rate on a grid of size $N = 288 \times 192 = 55{,}296$
- want to infer the N-dimensional distribution based on an ensemble of size $n < 100$

Challenges

- Dependence:
  - nonstationary
  - nonparametric
  - nonlinear (i.e., non-Gaussian)
- Computation (scalability to large N)

Many existing approaches address some of these challenges, but few address all.

Outline

Transport maps

- $\mathbf{y} = (y_1, \ldots, y_N)^\top$: zero-mean continuous random vector
  - for example, spatial field at N locations
- Transforming a Gaussian distribution to a standard Gaussian:

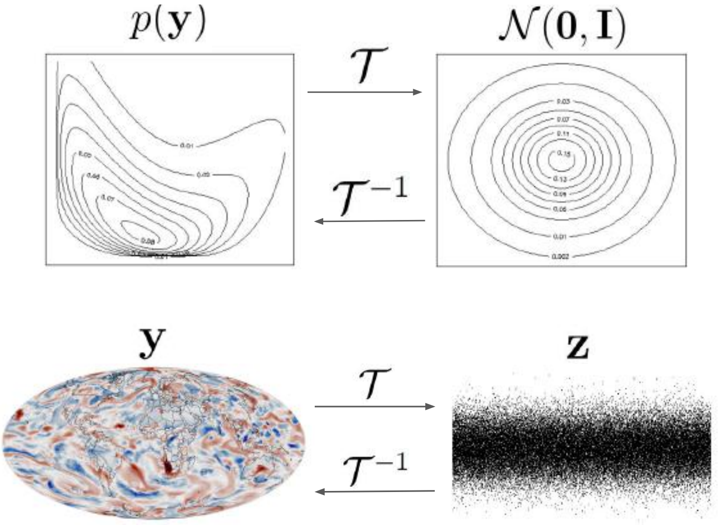
$$\mathbf{y} \sim \mathcal{N}_N(\mathbf{0}, \mathbf{\Sigma}) \text{ with } \mathbf{\Sigma}^{-1} = \mathbf{L}^\top \mathbf{L} \qquad \Longrightarrow \qquad \mathbf{L}\mathbf{y} \sim \mathcal{N}_N(\mathbf{0}, \mathbf{I}_N)$$

- Extension to non-Gaussian $p(\mathbf{y})$: nonlinear transport map $\mathcal{T} : \mathbb{R}^N \to \mathbb{R}^N$ (e.g., Marzouk et al, 2016) such that

$$\mathbf{y} \sim p(\mathbf{y}) \qquad \Longrightarrow \qquad \mathcal{T}(\mathbf{y}) \sim \mathcal{N}_N(\mathbf{0}, \mathbf{I}_N)$$

- Given $\mathcal{T}$ and $\mathcal{T}^{-1}$, we can:
  - Sample $\mathbf{y}^\star = \mathcal{T}^{-1}(\mathbf{z}^\star) \sim p(\mathbf{y})$ based on $\mathbf{z}^\star \sim \mathcal{N}_N(\mathbf{0}, \mathbf{I}_N)$
  - Transform the non-Gaussian $\mathbf{y}$ to $\mathbf{z} = \mathcal{T}(\mathbf{y})$, for which simple linear operations (e.g., averaging) may be more meaningful

Illustration of transport maps



Transport maps as regressions

- WLOG, $\mathcal{T}$ is lower-triangular:

$$\mathcal{T}(\mathbf{y}) = \begin{bmatrix} \mathcal{T}_1(y_1) \\ \mathcal{T}_2(y_1, y_2) \\ \vdots \\ \mathcal{T}_N(y_1, y_2, \ldots, y_N) \end{bmatrix}$$

- Assumption: each $\mathcal{T}_i$ is linearly additive in its ith argument
- Then the target distribution has the form:

$$p(\mathbf{y}) = \prod_{i=1}^{N} \mathcal{N}(y_i | f_i(\mathbf{y}_{1:i-1}), d_i^2)$$

for some $f_i : \mathbb{R}^{i-1} \to \mathbb{R}$, $d_i \in \mathbb{R}^+$

- Thus, the difficult problem of inferring $p(\mathbf{y})$ has turned into N independent regressions of y_i on $\mathbf{y}_{1:i-1}$ of the form

$$y_i = f_i(\mathbf{y}_{1:i-1}) + \epsilon_i, \quad \epsilon_i \overset{ind.}{\sim} \mathcal{N}(0, d_i^2), \qquad i = 1, \ldots, N$$

Bayesian transport maps

Most existing transport-map approaches (e.g., Marzouk et al, 2016): estimate parameters in parametric map, often without quantifying uncertainty.

Our Bayesian approach:

- Independent Gaussian-process-inverse-Gamma priors for the f_i and d_i^2, $i = 1, \ldots, N$
- GP prior on f_i is nonparametric and shrinks toward a linear function
- Induces prior distributions on the $\mathcal{T}_i$ and thus on the entire map $\mathcal{T}$
- Priors are conjugate, which leads to closed-form expressions

Posterior map

- Training data: $\mathbf{Y} = (\mathbf{y}^{(1)}, \ldots, \mathbf{y}^{(n)})$ such that $\mathbf{y}^{(j)} \overset{iid}{\sim} p(\mathbf{y}|\mathbf{f}, \mathbf{d})$ with $\mathcal{T}(\mathbf{y}^{(j)}) \,|\, \mathbf{f}, \mathbf{d} \sim \mathcal{N}_N(\mathbf{0}, \mathbf{I}_N)$, $j = 1, \ldots, n$
- Transport map $\widetilde{\mathcal{T}}$ for posterior predictive distribution:

$$\mathbf{y}^\star \sim p(\mathbf{y}|\mathbf{Y}) \qquad \Longrightarrow \qquad \widetilde{\mathcal{T}}(\mathbf{y}^\star) \sim \mathcal{N}_N(\mathbf{0}, \mathbf{I}_N)$$

is also triangular with components obtained using GP prediction:

$$\widetilde{\mathcal{T}}_i(y_1^\star, \ldots, y_i^\star) = \Phi^{-1}\Big(F_{2\tilde{\alpha}_i}\Big(\hat{d}_i^{-1}\big(v_i(\mathbf{y}_{1:i-1}^\star)+1\big)^{-1/2}\big(y_i^\star - \hat{f}_i(\mathbf{y}_{1:i-1}^\star)\big)\Big)\Big)$$

- $\widetilde{\mathcal{T}}^{-1}$ can be evaluated at $\mathbf{z}^\star$ by recursively solving $\widetilde{\mathcal{T}}(\mathbf{y}^\star) = \mathbf{z}^\star$ for $\mathbf{y}^\star$:

$$y_i^\star = \hat{f}_i(\mathbf{y}_{1:i-1}^\star) + F_{2\tilde{\alpha}_i}^{-1}\big(\Phi(z_i^\star)\big)\, \hat{d}_i\big(v_i(\mathbf{y}_{1:i-1}^\star) + 1\big)^{1/2}, \quad i = 1, \ldots, N$$

Hyperparameters

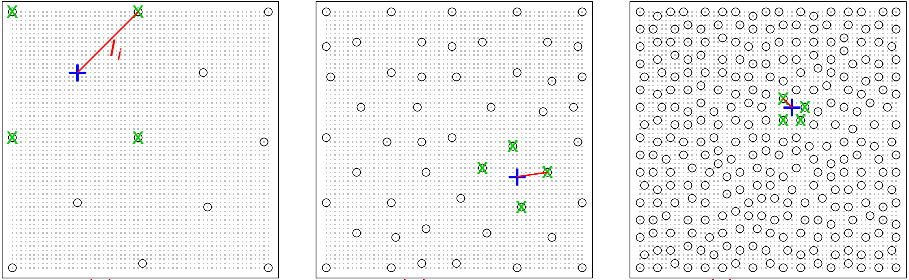
- The priors on the f_i and d_i^2 may depend on hyperparameters $\boldsymbol{\theta}$
- Closed-form expression of the integrated likelihood $p(\mathbf{Y})$ (with the f_i and d_i^2 integrated out) as a product over N closed-form terms
- Use integrated likelihood for empirical (optimization over $\boldsymbol{\theta}$) or fully Bayesian (sampling $\boldsymbol{\theta}$) inference
- Fast optimization via stochastic gradient descent with autodiff

Outline

Maximin ordering

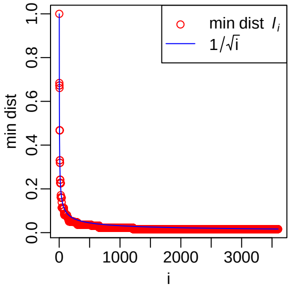
Spatial field: $\mathbf{y} = (y_1, \ldots, y_N)^\top$ at locations $\mathbf{s}_1, \ldots, \mathbf{s}_N$, with $y_i = y(\mathbf{s}_i)$

Maximum-minimum-distance ordering of $\mathbf{s}_1, \ldots, \mathbf{s}_N$ and hence $y_1, \ldots, y_N$:

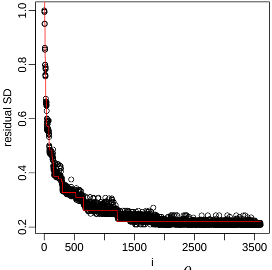


(a) $i = 13$ (b) $i = 51$ (c) $i = 290$

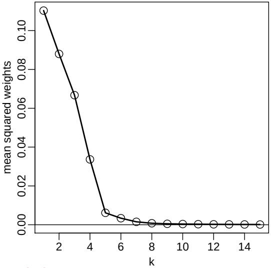
Prior distributions based on scale decay



(a) scale decay



(b) d_i ($\circ$), $e^{\theta_{d,1}} \ell_i^{\theta_{d,2}}$ (—)



(c) Effect of kth NN

- Spatial predictions on systematically decaying scale
- Conditional near-Gaussianity on finer scales for stochastic processes with quasiquadratic loglikelihoods
- Priors on d_i and f_i motivated by GP with Matérn-type covariance:
  - d_i decay polynomially with ℓ_i
  - Influence of kth-NN decays rapidly
    $\rightarrow$ only consider m NNs for sparsity ($\rightarrow N \times \mathcal{O}(n^3 + mn^2)$, parallel)

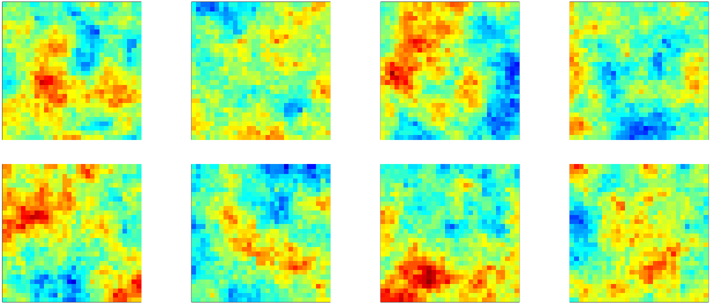
Increased marginal flexibility

- So far, focus on nonlinear (i.e., non-Gaussian) dependence
- Two possible extensions to increase marginal flexibility
  - Marginal transformation: Observe $\tilde{\mathbf{y}} = \mathcal{G}(\mathbf{y})$ such that $\tilde{y}_i = g_i(y_i)$
  - Bayesian nonparametrics: Model regression errors using Dirichlet process mixtures with Gaussian base measures
    (Requires MCMC; no closed-form transport map)

Outline

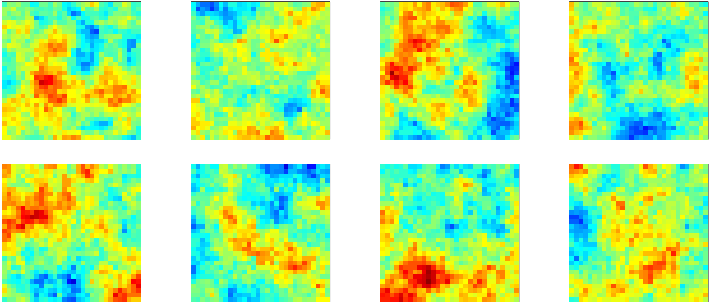
Samples from a Gaussian distribution

Gaussian distribution with exponential cov. on grid of size $N = 30 \times 30$



Samples from a Gaussian distribution

Gaussian distribution with exponential cov. on grid of size $N = 30 \times 30$

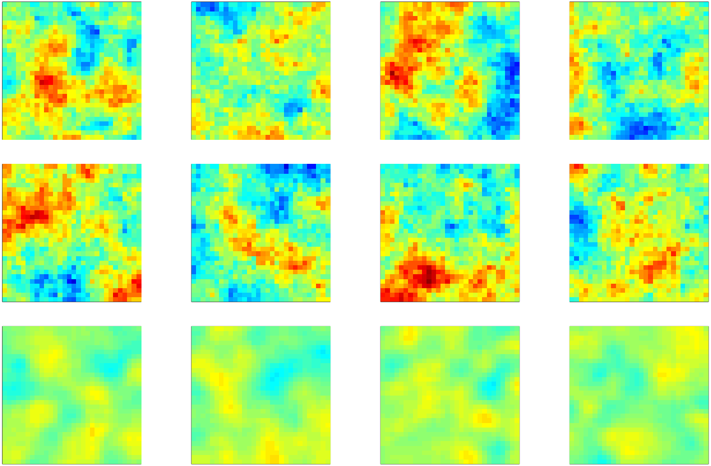


Training data

Transport map

Samples from a Gaussian distribution

Gaussian distribution with exponential cov. on grid of size $N = 30 \times 30$



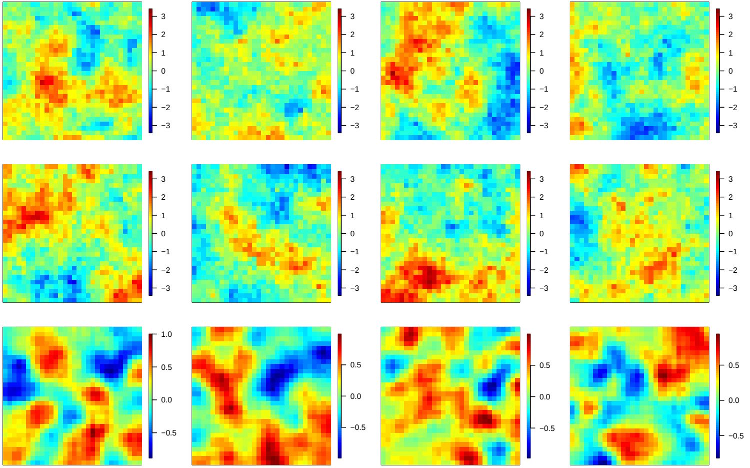
Training data

Transport map

VAE

Samples from a Gaussian distribution

Gaussian distribution with exponential cov. on grid of size $N = 30 \times 30$



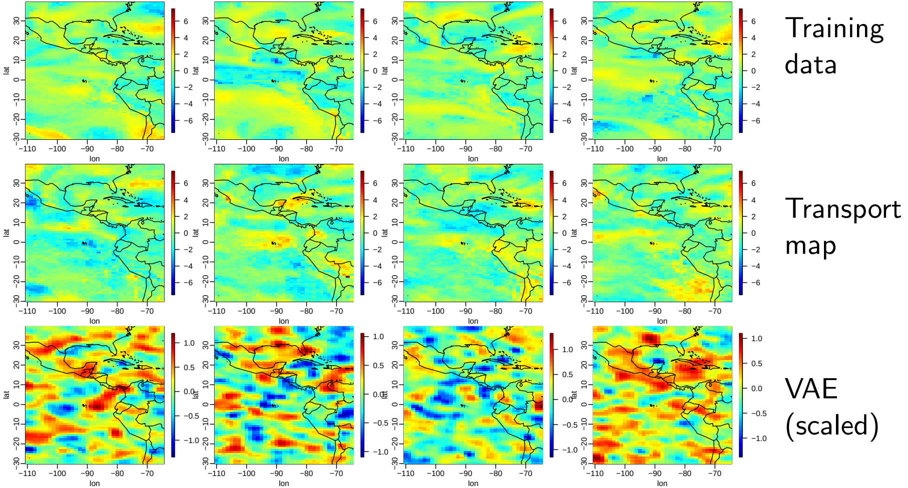
Training data

Transport map

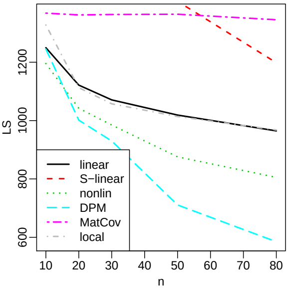
VAE (scaled)

Precipitation over Americas subregion

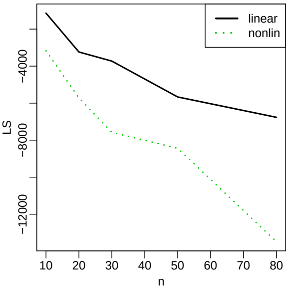
Nonlinear TM for precip anomalies on subregion ($N = 37 \times 74 = 2{,}738$)



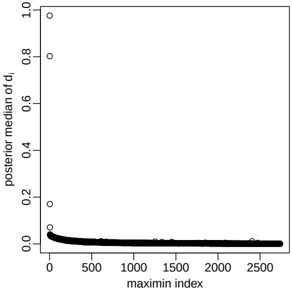
Results for precipitation anomalies



(a) LS (Americas)



(b) LS (global)



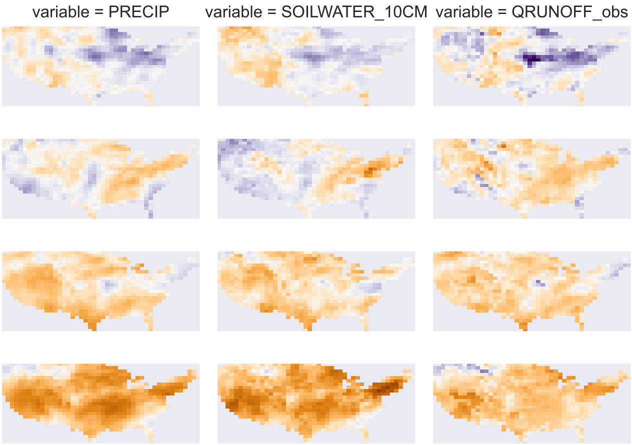
(c) Posterior median of d_i

- Log score (LS) for Americas subregion ($N = 37 \times 74 = 2{,}738$) and entire globe ($N = 288 \times 192 = 55{,}296$) — lower is better
- Early map coefficients captured much more variation than later-ordered coefficients
  $\rightarrow$ “nonlinear principal components” for non-Gaussian fields

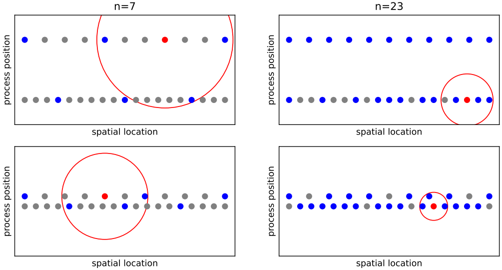
Outline

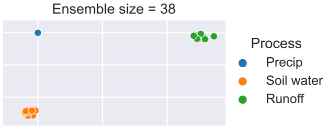
Multivariate spatial fields (Wiemann & K, 2023)

Example: Precipitation, soil water content, and runoff anomalies (columns) from 4 samples (rows) from the CESM large ensemble project:



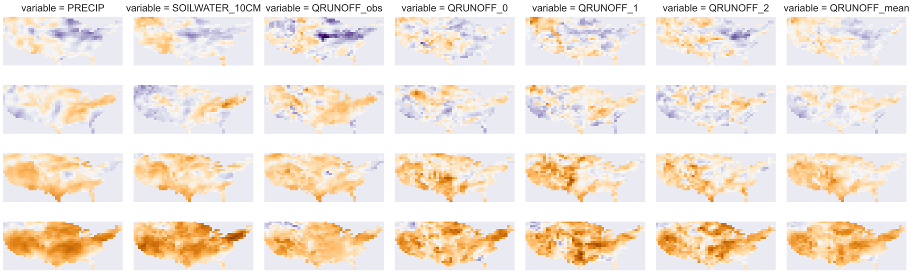
Multivariate spatial fields (Wiemann & K, 2023)





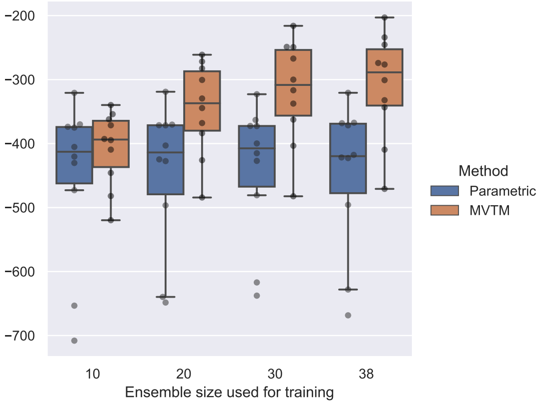
Left: Illustration of MM ordering and NN in a toy example in 1-D space for two processes with weak (top) and strong (bottom) dependence
Right: Latent process positions estimated from the climate data over 10 CV splits

Multivariate spatial fields (Wiemann & K, 2023)



Posterior distribution of held-out runoff given precipitation and soil water ($n = 38$): 3 samples and mean

Multivariate spatial fields (Wiemann & K, 2023)

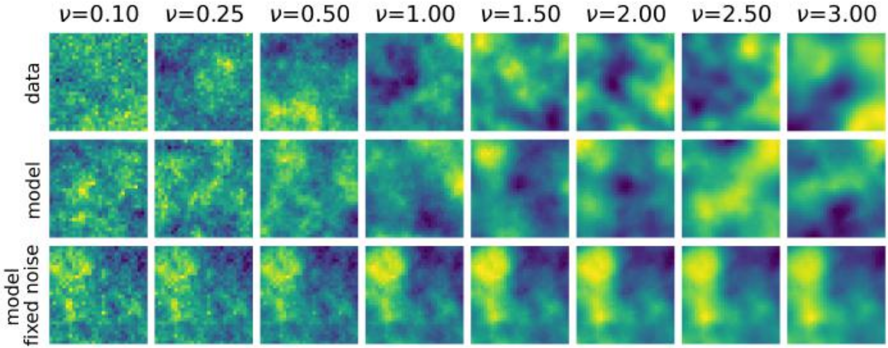


Log scores (higher is better) on held-out runoff test samples conditional on precip and soil water, for our multivariate transport map (MVTM) and a Matérn GP (Parametric), as a function of training ensemble size

Conditional distributions (Drennan et al, in prep)

Goal: Learn conditional distribution $p(\mathbf{y}|\mathbf{x})$ given covariates $\mathbf{x}$ (e.g., emission scenarios, input parameters), by including $\mathbf{x}$ in TM regressions.

Example: $p(\mathbf{y}|x) = \mathcal{N}(\mathbf{y}|\mathbf{0}, \mathbf{K}_x)$, where $\mathbf{K}_x$ is a Matérn covariance with smoothness $\nu = \exp(x)$:



Bottom: $\tilde{\mathcal{T}}_x^{-1}(\mathbf{z})$ for fixed $\mathbf{z}$, different x, using learned conditional TM $\tilde{\mathcal{T}}_\mathbf{x}$

Climate-model calibration (Chakraborty et al, in prep)

For climate model $p_{\mathbf{x}}(\mathbf{y})$, want to find the parameter value $\mathbf{x}^*$ that most closely matches $p_{\mathbf{x}}(\mathbf{y})$ to the true climate $\pi(\mathbf{y})$:

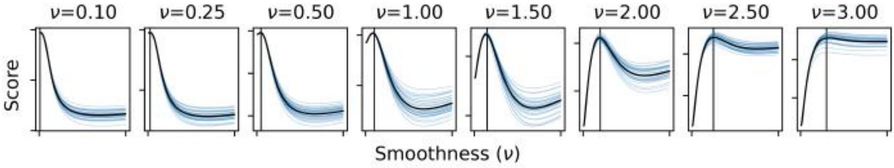
$$\mathbf{x}^* = \underset{\mathbf{x}}{\operatorname{argmin}} \operatorname{KL}\big(\pi(\mathbf{y}) \| p_{\mathbf{x}}(\mathbf{y})\big) = \underset{\mathbf{x}}{\operatorname{argmax}} \, \mathbb{E}_{\pi}[\log p_{\mathbf{x}}(\mathbf{y})] \approx \underset{\mathbf{x}}{\operatorname{argmax}} \log p_{\mathbf{x}}(\mathbf{y}^*),$$

given the observation $\mathbf{y}^* \sim \pi(\mathbf{y})$.

Idea: As $p_{\mathbf{x}}(\mathbf{y})$ is unknown (and sampling from it is expensive), use TM $\tilde{p}(\mathbf{y}|\mathbf{x})$ as a surrogate.

Example: Calibration using TM surrogate

- Train TM on $n \approx 20$ Matérn fields as before.
- New observation $\mathbf{y}^* \sim p(\mathbf{y}|x^*) = \mathcal{N}(\mathbf{0}, \mathbf{K}_{x^*})$ from "true" distribution
- Learned log score $\log \tilde{p}(\mathbf{y}^*|x)$ is maximized approximately at the true value x^* (vertical lines) for different $(x^*, \mathbf{y}^*)$ pairs

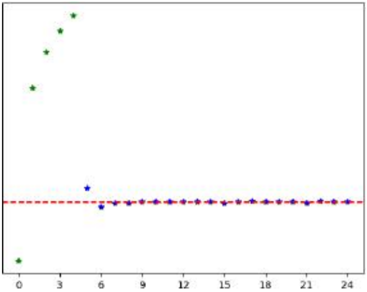


Idea for calibration via active learning

Sequentially for each $k = 1, 2, \ldots$,

1. train TM $\tilde{p}^{(k)}(\mathbf{y}|\mathbf{x})$ on training data $(\mathbf{x}^{(1)}, \mathbf{y}^{(1)}), \ldots, (\mathbf{x}^{(k)}, \mathbf{y}^{(k)})$
2. obtain a new training sample $\mathbf{y}^{(k+1)} \sim p_{\mathbf{x}}(\mathbf{y})$ at the value of $\mathbf{x}$ that
   Exploitation: maximizes $\tilde{p}^{(k)}(\mathbf{y}^*|\mathbf{x})$, or that
   Exploration: maximizes the discrepancy between $\tilde{p}^{(k)}(\mathbf{y}|\mathbf{x})$ and $p(\mathbf{y}|\mathbf{x})$

Exploitation-only results on Matérn example:



Outline

Conclusions

- Infer non-Gaussian distribution via a Bayesian transport map:
  - flexible and nonparametric
  - probabilistic regularization and UQ
- For spatial fields:
  - data-dependent sparsity and thus scalability to high dimensions
  - Bayesian, nonparametric, non-Gaussian extension of Vecchia approx.
- Supported by NSF DMS–1654083, DMS–1953005, NASA AIST–21
- Extensions and applications:
  - More complicated input domains: multivariate, space-time (in prep.), ...
  - Learn conditional distribution $p(\mathbf{y}|\mathbf{x})$ given covariates $\mathbf{x}$
    - interpolate between observed $\mathbf{x}$ (e.g., emission or intervention scenarios)
    - use as a surrogate for calibrating parameters $\mathbf{x}$

References

Main: Katzfuss, M., & Schäfer, F. (2023). Scalable Bayesian transport maps for high-dimensional non-Gaussian spatial fields. *Journal of the American Statistical Association*, accepted.

Linear/Gaussian case: Kidd, B., & Katzfuss, M. (2022). Bayesian nonstationary and nonparametric covariance estimation for large spatial data (with discussion). *Bayesian Analysis*, 17(1), 291–351.

Multivariate extension: Wiemann, P. F. V., & Katzfuss, M. (2023). Bayesian nonparametric generative modeling of large multivariate non-Gaussian spatial fields. *Journal of Agricultural, Biological, and Environmental Statistics*, 28, 597–617.

Use in EnKF: Boyles, W., & Katzfuss, M. (2021). Ensemble Kalman filter updates based on regularized sparse inverse Cholesky factors. *Monthly Weather Review*, 149(7), 2231–2238.