Fast non-stationary geospatial modelling with multiresolution (wavelet) Gaussian processes

Talay M Cheema (tmc49@cam.ac.uk) Carl Edward Rasmussen

UNIVERSITY OF CAMBRIDGE

computational &biological learning laboratory

Background

- Goal: learn a spatial field f with uncertainty estimates from noisy measurements $y \approxeq f(x)$
- Gaussian process (GP) regression is a good tool, but the learning objective costs $O(N^3)$ to evaluate for N datapoints – and needs to be evaluated many times in learning.
- Precomputable sparse methods like variational Fourier features reduce the cost to $O(M^3)$ for $M \ll N$ but only for stationary priors.
- Lots of geospatial phenomena in weather and climate are highly non-stationary. Using a stationary prior leads to learning too smooth an approximation or a globally low lengthscale.
- There are good ways to construct non-stationary GP priors – like using Gibbs' covariance function, but these are not compatible with precomputable methods, and scale poorly.

$$y_n = f(x_n) + \sigma\rho_n \quad n \in \{1 : N\}; \quad f \sim \mathcal{GP}(0, k)$$

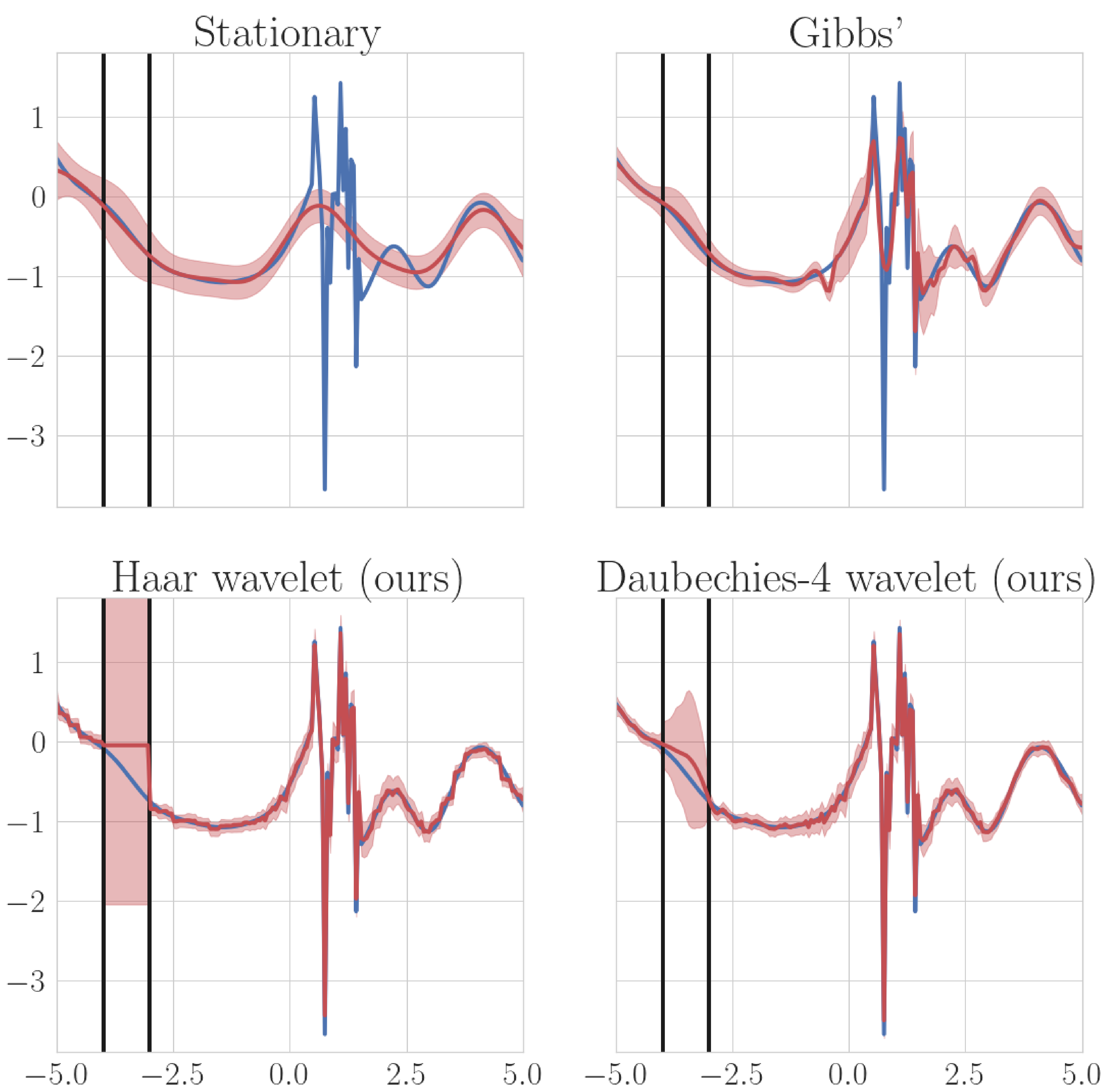
where each ρ_n has an independent standard normal distribution.

Condition on Gaussian inducing variables $u \sim p(u)$ with approximate posterior $q(u)$. Then the variational learning objective and optimum posterior are available in closed form.

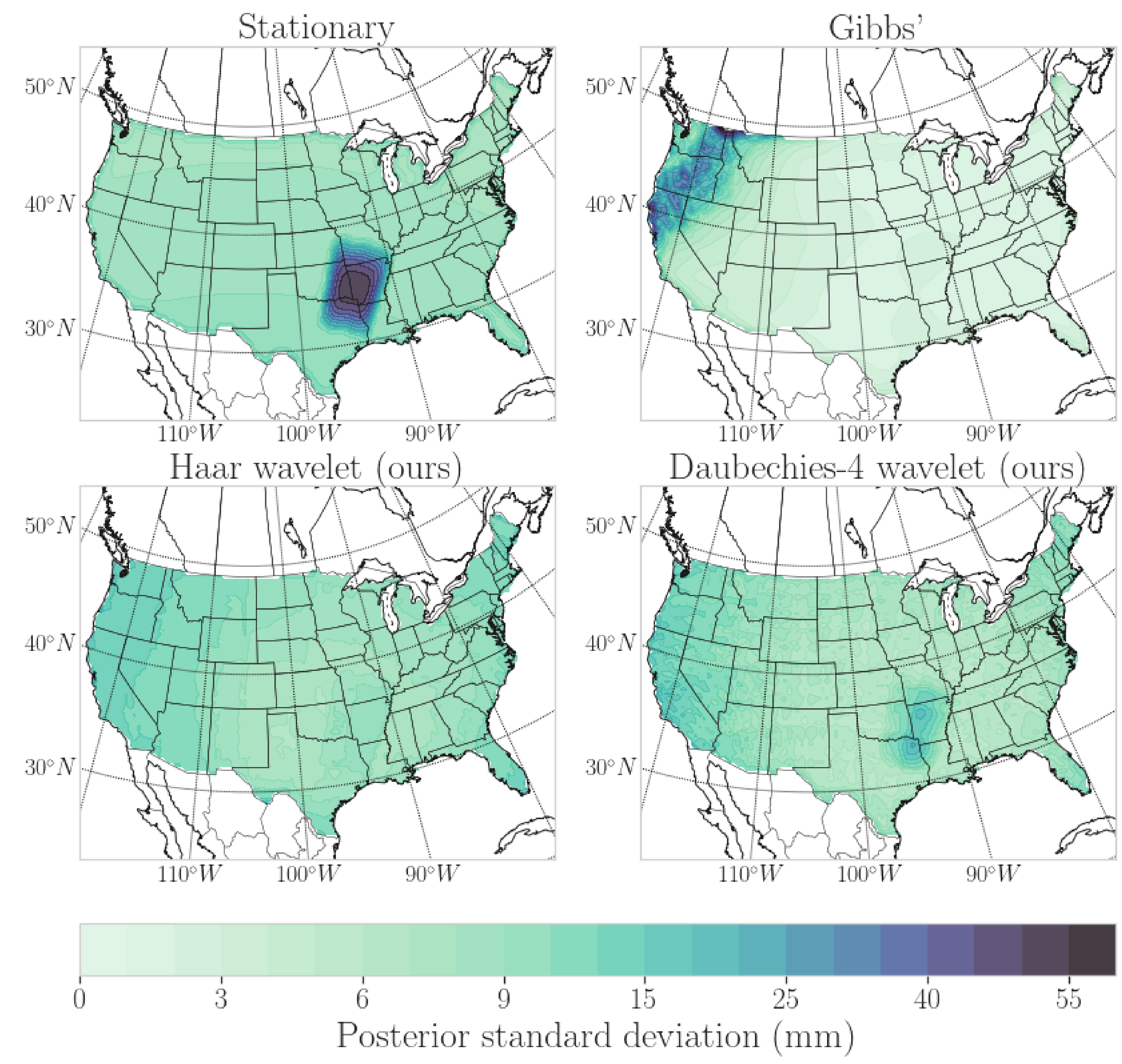
Choosing u carefully leads to much faster learning (for some k ...)

Choosing k carefully leads to much better learning.

We want to do both for geospatial data.



Groundtruth and posterior fit. No training data between the black lines.



Precipitation data. No training data in a southeastern rectangular patch.

Wavelet features

- Discrete wavelets $\{w_{j\ell}\}_{j\in\mathbb{Z},\ell\in\mathbb{Z}}$ form an orthonormal basis of $L^2(\mathbb{R})$ which captures local variations in scale.
- The span of coarser wavelets of scale less than ι, say, has an orthonormal basis given by the scaling function $\{v_{\iota\ell}\}_{\ell\in\mathbb{Z}}$.
- Stationary covariance functions are a sum of independent global (Fourier) features:

$$k_{\text{stationary}}(x, x') = \int s(2\pi\xi) e^{i2\pi\xi}\, d(2\pi\xi)$$

- To get a non-stationary covariance function, use the wavelet basis

$$k_{\text{multiresolution}}(x, x') = \sum_{\ell} \alpha_\ell v_{\iota\ell}(x) v_{\iota\ell}(x') + \sum_{j=\iota}^{\infty} \sum_{\ell} \beta_{j\ell} w_{j\ell}(x) w_{j\ell}(x')$$

- To get precomputable features, define the features $u_{j,\ell} = \langle f, w_{j'+\iota,\ell} \rangle / \beta_{j'+\iota}$.
- We design the parametric form of the coefficients to be suitable to learn local variations in scale and try Haar and Daubechies-4 wavelets.

- These lead to faster learning
- The performance avoids the failure mode of stationary priors
- But there is still a gap to the performance of existing, more computationally demanding non-stationary methods.

Table 1: Precipitation dataset evaluation results: mean and standard error over five runs. The test set is either a uniformly random 20% of points, or a rectangular patch (as plotted above).

Test		Stationary	Gibbs'	Wavelet–Haar	Wavelet–db4
Uniform	RMSE	21.19 (0.87)	23.52 (1.33)	26.03 (0.72)	22.27 (0.78)
	NLPD	4.484 (0.04)	4.07 (0.016)	4.60 (0.02)	4.47 (0.03)
Patch	RMSE	21.38 (0.00)	7.606 (0.135)	18.15 (0.88)	8.287 (0.282)
	NLPD	4.69 (0.000)	3.734 (0.043)	4.430 (0.025)	4.212 (0.013)