CornellBowers Statistics + Data Science

WashU

Nonstationary GPs with Neural Network Parameters

Zachary James[1], Joseph Guinness[2]

[1]Cornell University, Department of Statistics and Data Science
[2]Washington University in Saint Louis, Department of Statistics and Data Science

Objective

Given observations $Y \in \mathbb{R}^n$ at locations $X \in \mathbb{R}^{n \times d}$, we learn the relationship

$$y_i = f(x_i) + \epsilon_i \tag{1}$$

where y_i is the ith element of Y, x_i is the ith row of X, and $\epsilon_i \sim \mathcal{N}(0, \tau_i^2)$ is independent noise. We first place a GP prior on the function $f \sim \mathcal{GP}(\mu(\cdot), k(\cdot, \cdot \mid \theta))$, where $\mu(x_i)$ is the expected value of $f(x_i)$, and $k(x_i, x_j \mid \theta)$ is the covariance between $f(x_i)$ and $f(x_j)$. A frequent assumption is that the covariance is stationary

$$k(x, x' \mid \theta) = k^{stat}(x - x' \mid \theta) \tag{2}$$

Nonstationary Parameters

Stationarity may be too strict for some datasets. A large class of covariance functions relax these assumptions and instead assume that the parameters vary across the spatial domain. [1] developed the covariance function

$$k(x, x' \mid \theta, \sigma(\cdot), \Sigma(\cdot)) = \sigma(x)\sigma(x')|\Sigma(x)|^{\frac{1}{4}}|\Sigma(x')|^{\frac{1}{4}} \left| \frac{\Sigma(x) + \Sigma(x')}{2} \right|^{-\frac{1}{2}} k^{iso}\left(\sqrt{Q(x, x')} \mid \theta\right) \tag{3}$$

$$Q(x, x') = (x - x')^T \left(\frac{\Sigma(x) + \Sigma(x')}{2} \right)^{-1} (x - x') \tag{4}$$

where $\sigma(\cdot)$ and $\Sigma(\cdot)$ represent local variances and anisotropies and k^{iso} is any isotropic covariance function.

Neural Nets as Nonstationary Parameters

We propose the nonstationary covariance $k(\cdot, \cdot \mid \theta, \tilde{\theta}(\cdot))$ that extends on the class first described in [1], where

- θ are stationary parameters
- $\tilde{\theta}(\cdot) = g(\cdot \mid w)$ are nonstationary parameters expressed as a function of a neural network g with weights w.

Consider the nonstationary variance kernel from [1]

$$k(x, x' \mid \theta, \sigma(\cdot)) = \sigma(x)\sigma(x')k^{stat}(x, x' \mid \theta) \tag{5}$$

where k^{stat} is any stationary covariance and $\sigma^2(x) = Var[f(x)]$. The corresponding covariance function under our framework is

$$k(x, x' \mid \theta, w) = g(x \mid w)g(x' \mid w)k^{stat}(x, x' \mid \theta) \tag{6}$$

where g is a neural network with positive output of dimension one. The noise can also be modeled as a nonstationary parameter by setting $\tau_i^2 = \tau(x_i)^2$. We then use a neural network to model $\tilde{\theta}(\cdot) = \tau(\cdot)$.

Interpretation as Learned Basis Functions

A nonstationary parameter $\tilde{\theta}$ (or any spatial process) observed at location s can be expressed in terms of its Karhunen–Loève expansion

$$\tilde{\theta}(s) = \sum_{i=1}^{\infty} \phi_i(s)\lambda_i \tag{7}$$

where ϕ_i are orthonormal eigenfunctions and λ_i are uncorrelated random coefficients. Truncating the series and relaxing the orthogonality produces a basis function approximation. Our framework is equivalent to learning a basis function expansion of a transformed parameter. A neural network of depth d is a composition of d nonlinear functions $g = g_d(g_{d-1}(\ldots \mid w_{d-1}) \mid w_d)$. Each layer $g_i : \mathbb{R}^{m_i} \to \mathbb{R}^{m_{i+1}}$ computes m_{i+1} linear combinations of the m_i inputs and applies an activation function element-wise. The final layer of the neural network $g_d : \mathbb{R}^{m_d} \to \tilde{\Theta}$ can be interpreted as taking a linear combination of m_d learned basis functions and then applying a link function.

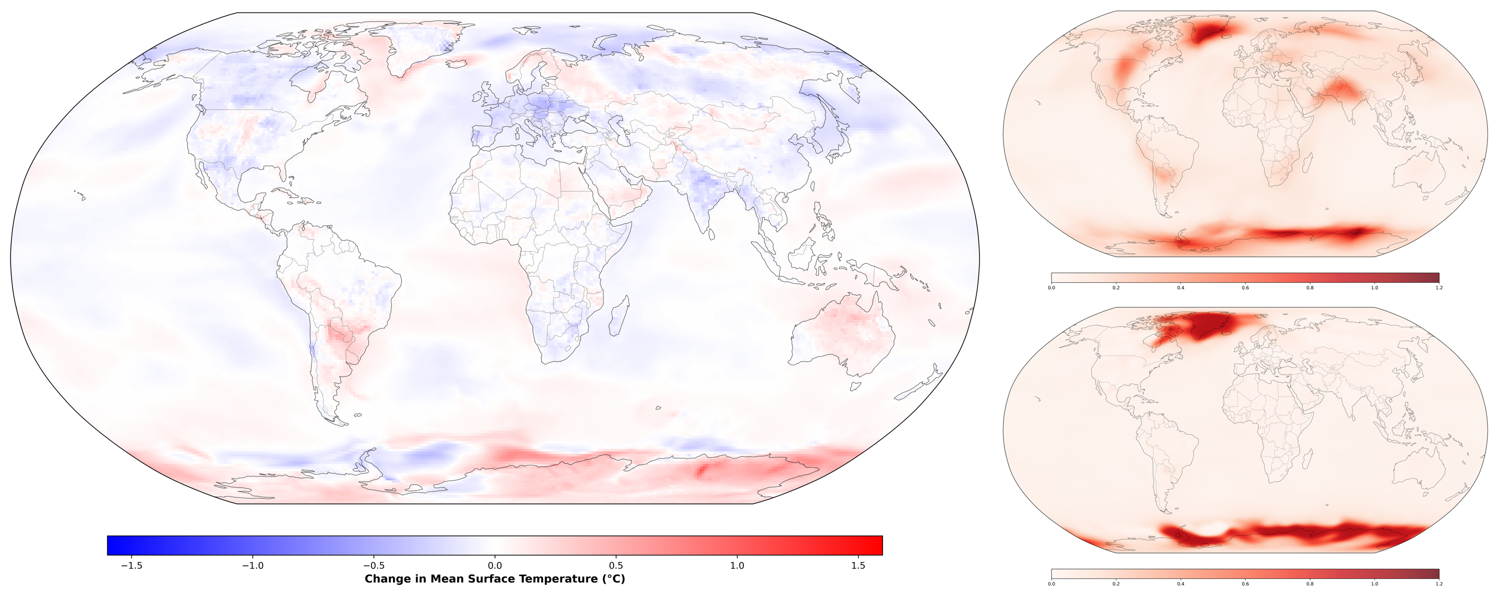

Figure: Surface temperature pattern scaling field (left), neural network standard deviation prediction (top right), basis approximation standard deviation prediction (bottom right)

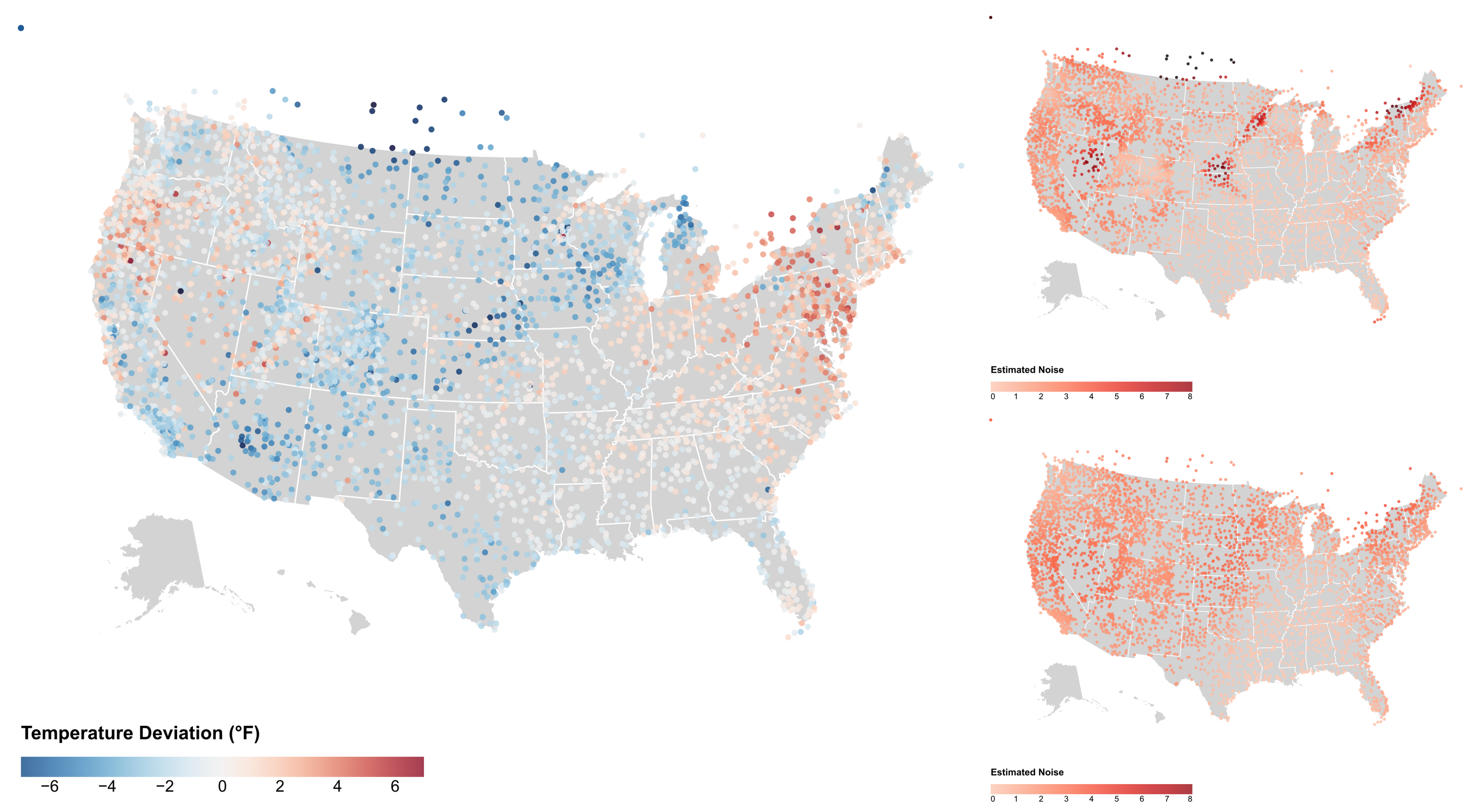

Figure: Surface temperature pattern scaling field (left), neural network standard deviation prediction (top right), basis approximation standard deviation prediction (bottom right)

Results

We test our framework on a surface temperature pattern scaling field that exhibits nonstationary variance and a weather station dataset that exhibits nonstationary noise. We fit three models for each: a stationary model, a nonstationary model with neural network parameter, and a nonstationary model approximated with Gaussian basis functions. We evaluate the models on RMSE, log-score, and computation time, with results in Table 1 and Table 2.

Model	RMSE ($\times 10^{-2}$)	Log-score ($\times 10^{4}$)	Time (s)
Stationary	3.65 ± 0.04	2.04 ± 0.011	73 ± 2.0
Basis Nonstat. Var.	3.65 ± 0.03	2.03 ± 0.005	179 ± 2.6
Neural Nonstat. Var.	3.61 ± 0.25	$\mathbf{2.12 \pm 0.079}$	77 ± 6.0

Table: Log-score, RMSE, and runtime of three GP models fit on the pattern scaling field. The basis approximation uses $\ell = 2500$ basis functions.

The neural network model outperforms on log-score for the pattern scaling, but not RMSE. All models perform similarly on the weather station dataset. The estimated variances of the neural net and basis models are similar, though the neural net model appears more sensitive. The neural net noise model is able to identify specific areas of high noise, while the basis model is not.

Model	RMSE ($\times 10^{-2}$)	Log-score ($\times 10^{4}$)	Time (s)
Stationary	1.457 ± 0.045	-1.643 ± 0.027	53.31 ± 2.63
Basis Nonstat. Noise	1.471 ± 0.041	-1.650 ± 0.026	51.73 ± 2.45
Neural Nonstat Noise	1.467 ± 0.045	-1.608 ± 0.037	51.71 ± 3.56

Table: Log-score, RMSE, and runtime of three GP models fit on the weather station dataset. The basis approximation uses $\ell = 1160$ basis functions.

Conclusion

We present a new framework for modeling nonstationary GPs. We hope our framework enables practitioners to efficiently fit a wide range of nonstationary GP models without being hindered by model setup and training.

References

[1] C. J. Paciorek and M. J. Schervish. Spatial modelling using a new class of nonstationary covariance functions. *Environmetrics*, 17(5):483–506, 2006. doi: https://doi.org/10.1002/env.785.