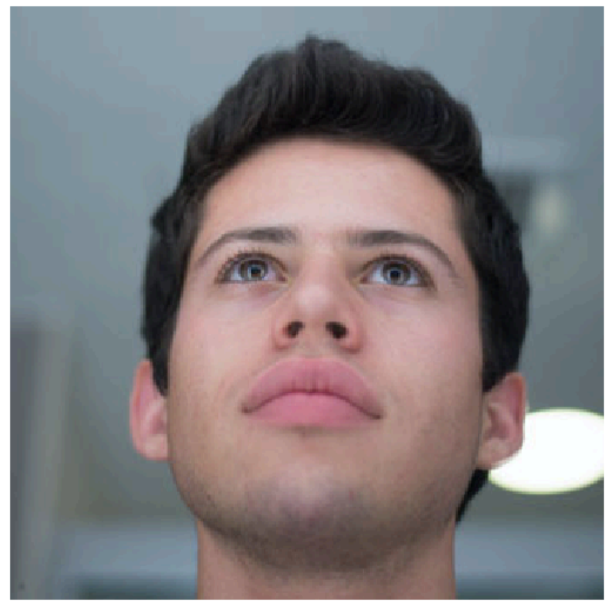
Inter-domain Deep Gaussian Processes

ICML 2020





Tim G. J. Rudner* Dino Sejdinovic Yarin Gal

University of Oxford

Project website: `http://bit.ly/inter-domain-dgp`

Bayesian Deep Learning

Classification

- Bayesian Neural Networks

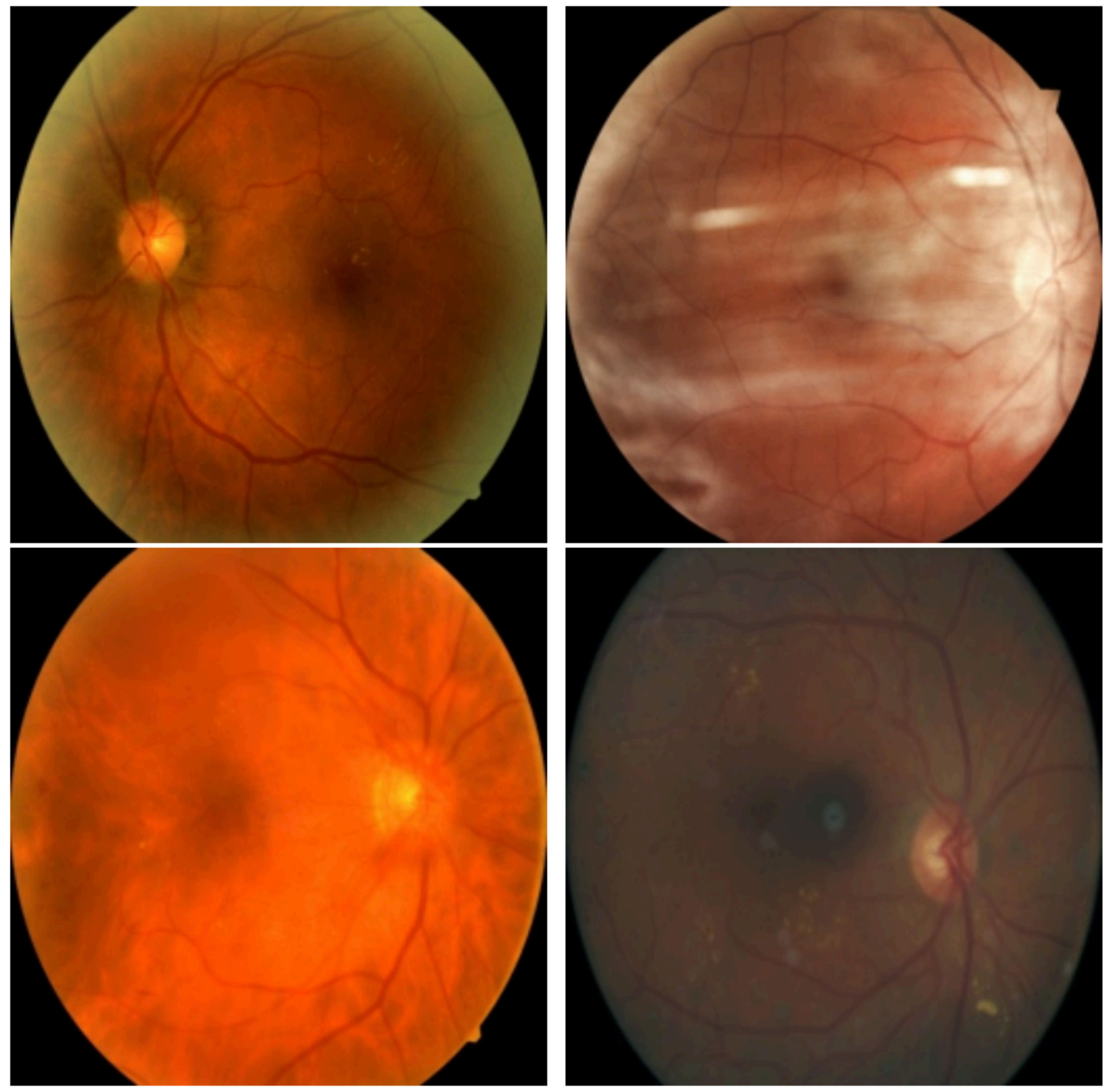


Figure 1. Retina scans for diabetic retinopathy diagnosis.[1]

Regression

- Deep Gaussian processes (DGPs)

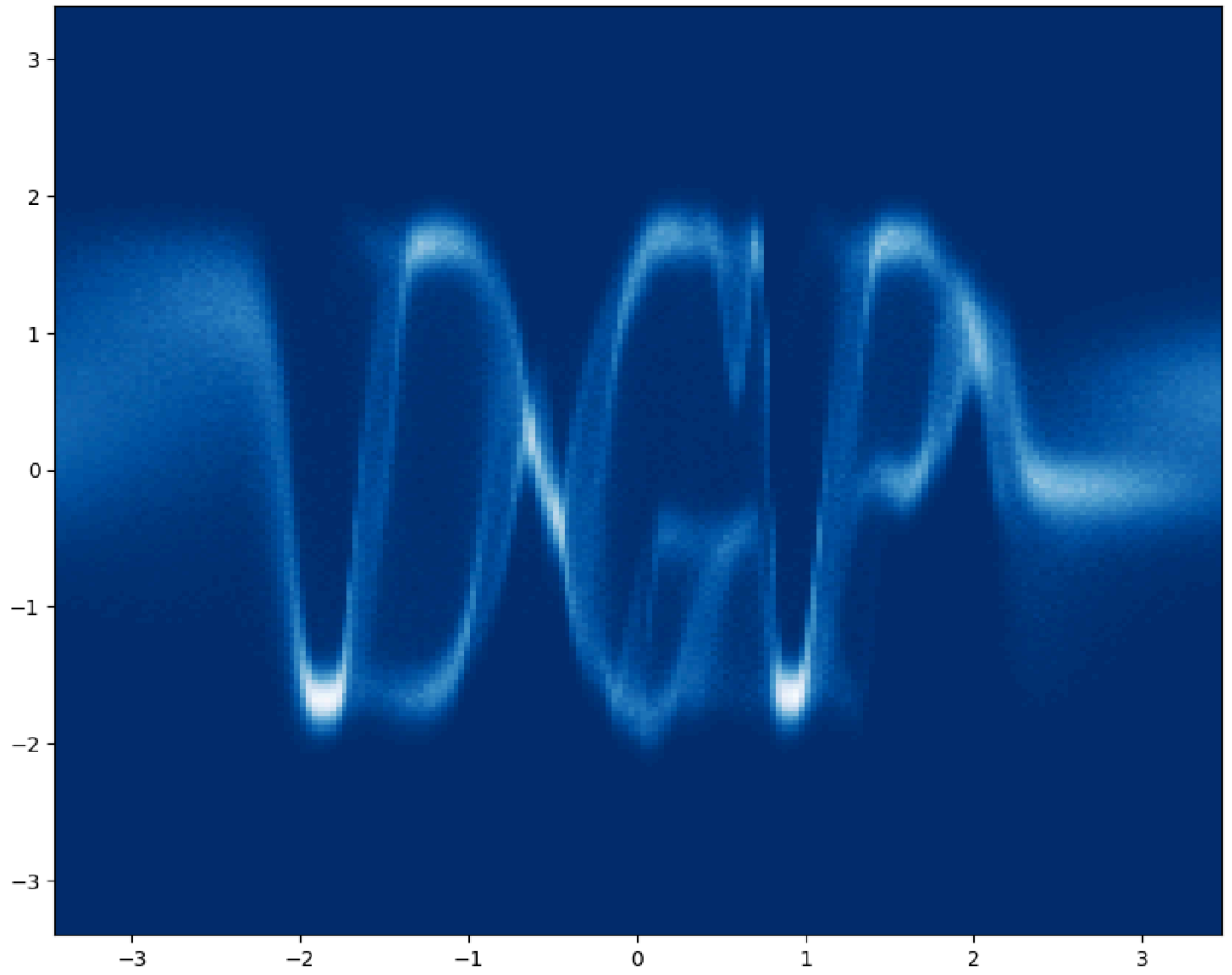


Figure 2. Complex, multi-modal deep GP posterior.[2]

[1] **Sebastian Farquhar, Michael Osborne, Yarin Gal.** Radial Bayesian Neural Networks: Beyond Discrete Support In Large-Scale Bayesian Deep Learning. In AISTATS 2020.

[2] **Hugh Salimbeni, Vincent Dutordoir, James Hensman, Marc P. Deisenroth.** Deep Gaussian Processes with Importance-Weighted Variational Inference. In ICML 2019.

Classification

- Bayesian Neural Networks

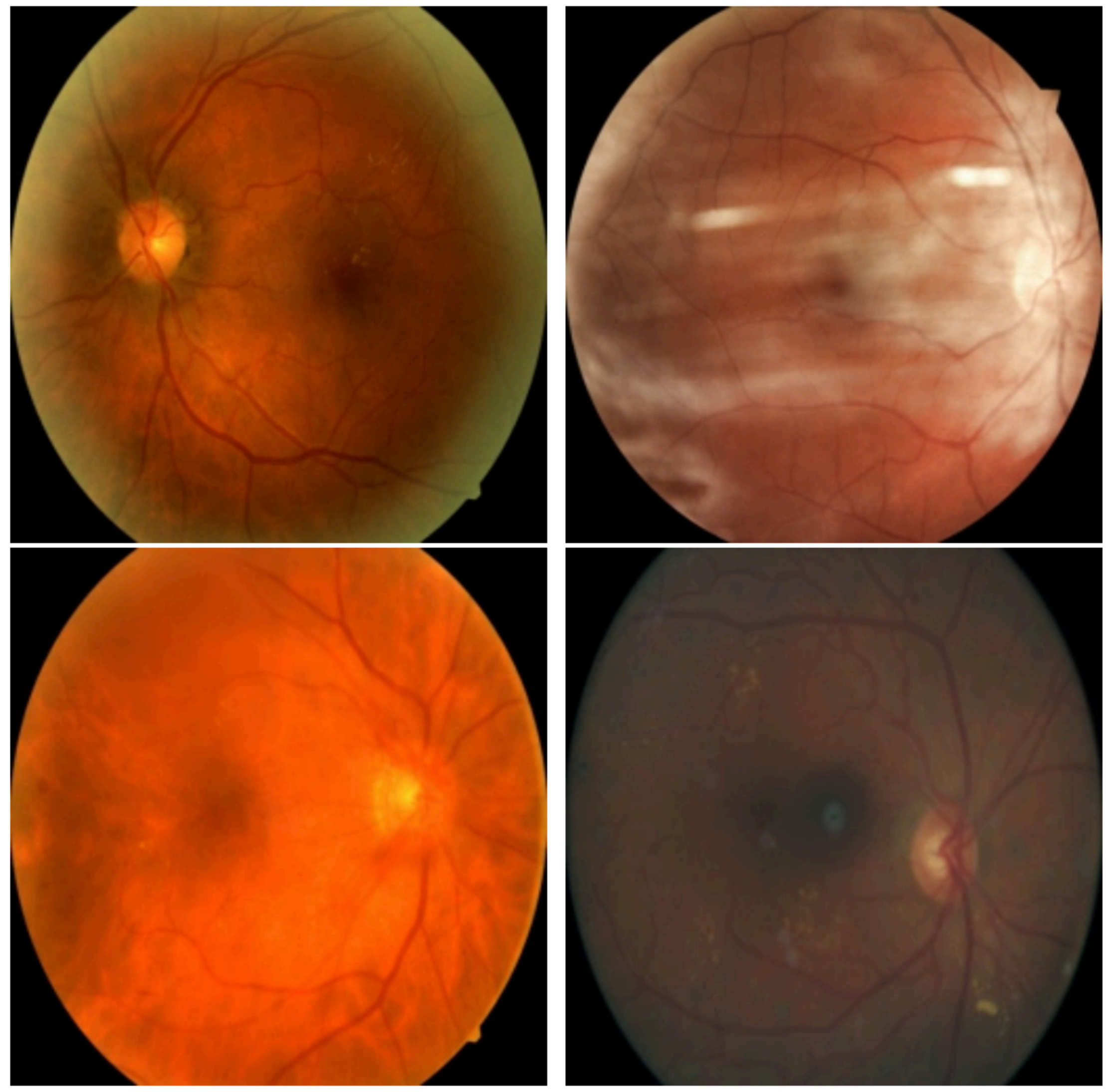


Figure 1. Retina scans for diabetic retinopathy diagnosis.[1]

Regression

- Deep Gaussian processes (DGPs)

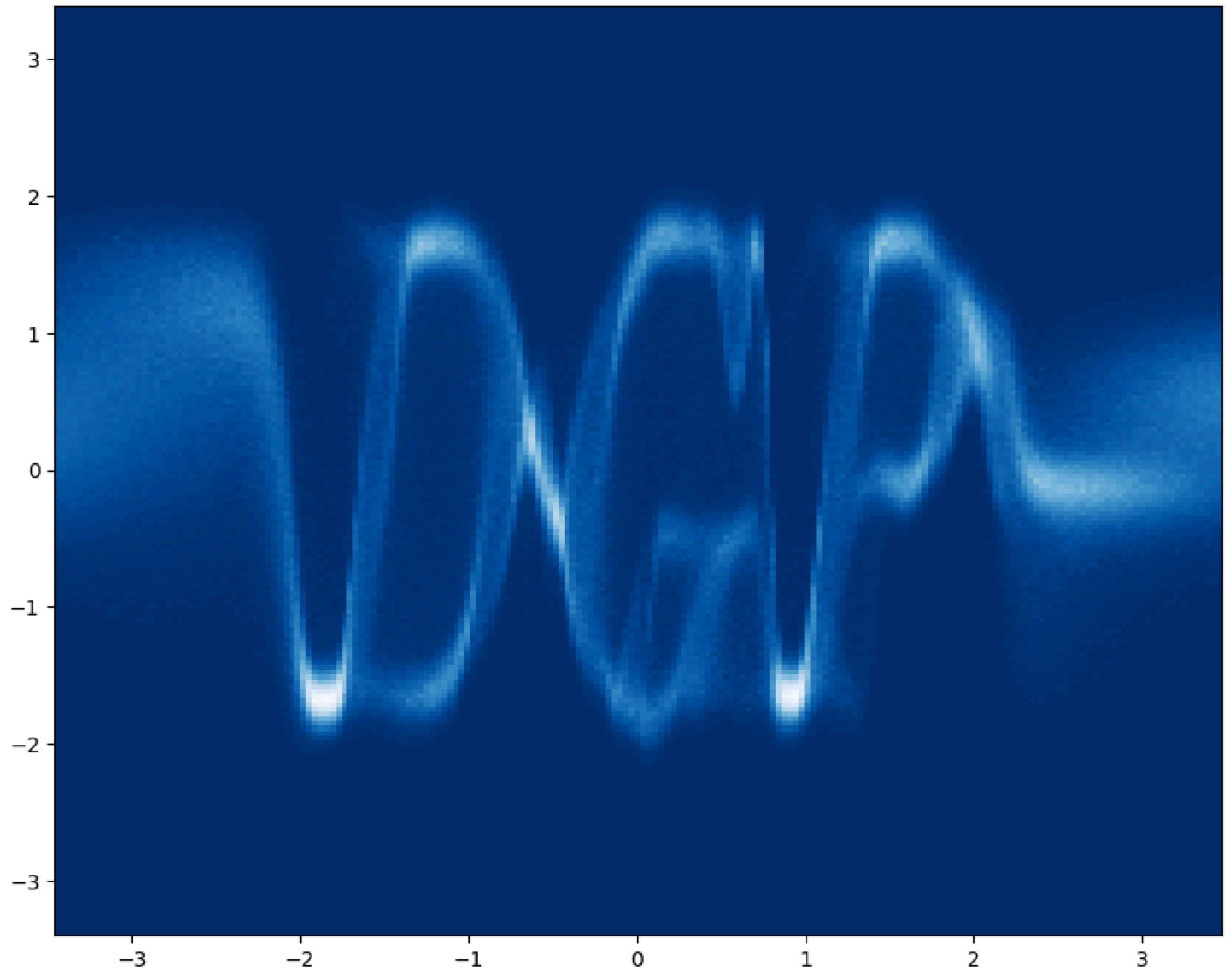


Figure 2. Complex, multi-modal deep GP posterior.[2]

[1] **Sebastian Farquhar, Michael Osborne, Yarin Gal.** Radial Bayesian Neural Networks: Beyond Discrete Support In Large-Scale Bayesian Deep Learning. In AISTATS 2020.

[2] **Hugh Salimbeni, Vincent Dutordoir, James Hensman, Marc P. Deisenroth.** Deep Gaussian Processes with Importance-Weighted Variational Inference. In ICML 2019.

Deep GP Regression

Shortcomings

- Reliant on "local" approximations based on "inducing points"
- Scales quadratically in the number of local approximations
- Does not capture any **global structure** in the data

Regression

- Deep Gaussian processes (DGPs)

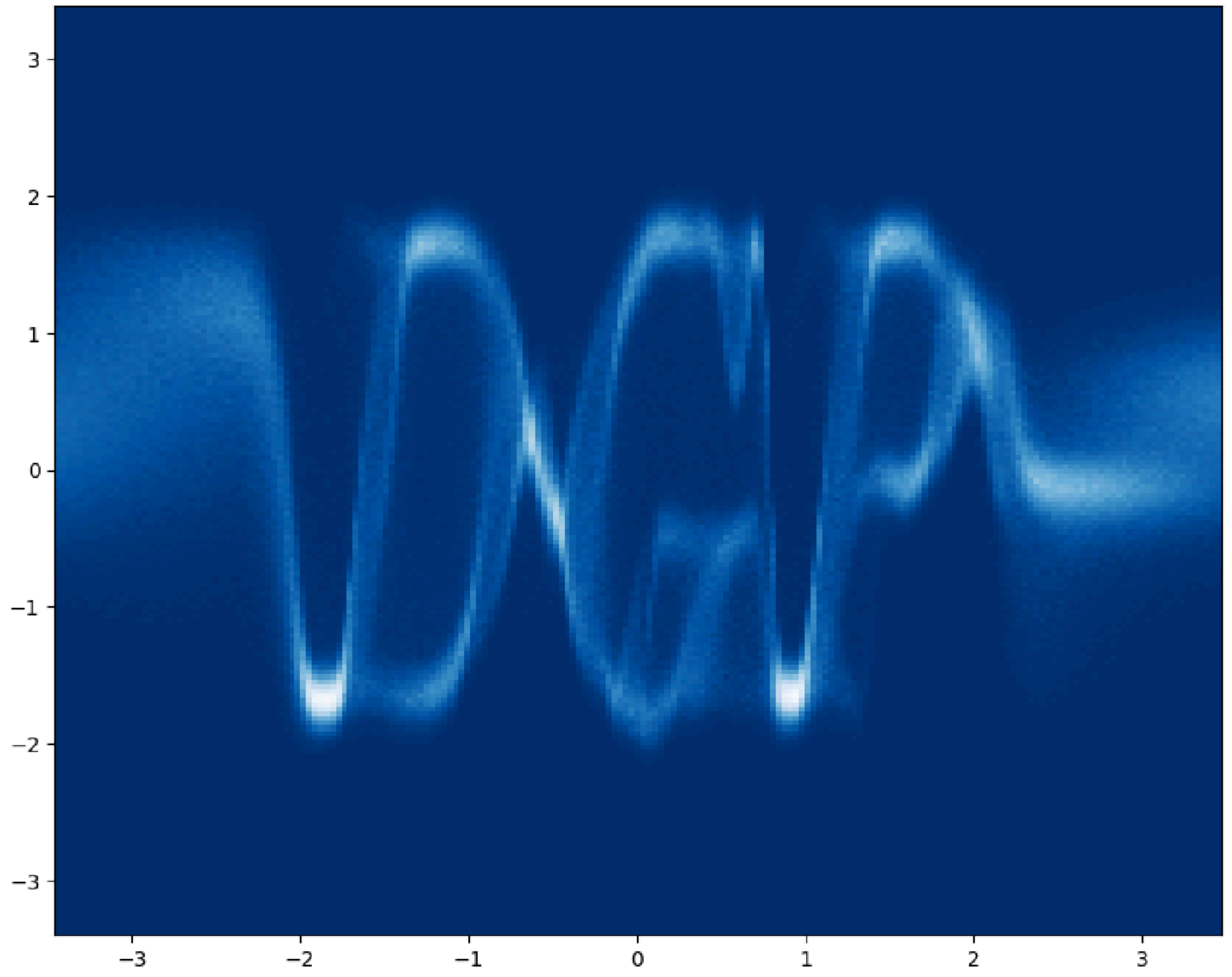


Figure 2. Complex, multi-modal deep GP posterior.[2]

[2] **Hugh Salimbeni, Vincent Dutordoir, James Hensman, Marc P. Deisenroth.** Deep Gaussian Processes with Importance-Weighted Variational Inference. In ICML 2019.

Deep GP Regression

Shortcomings

- Reliant on “local” approximations based on “inducing points”
- Scales quadratically in the number of local approximations
- Does not capture **global structure** in the data

Global structure in data

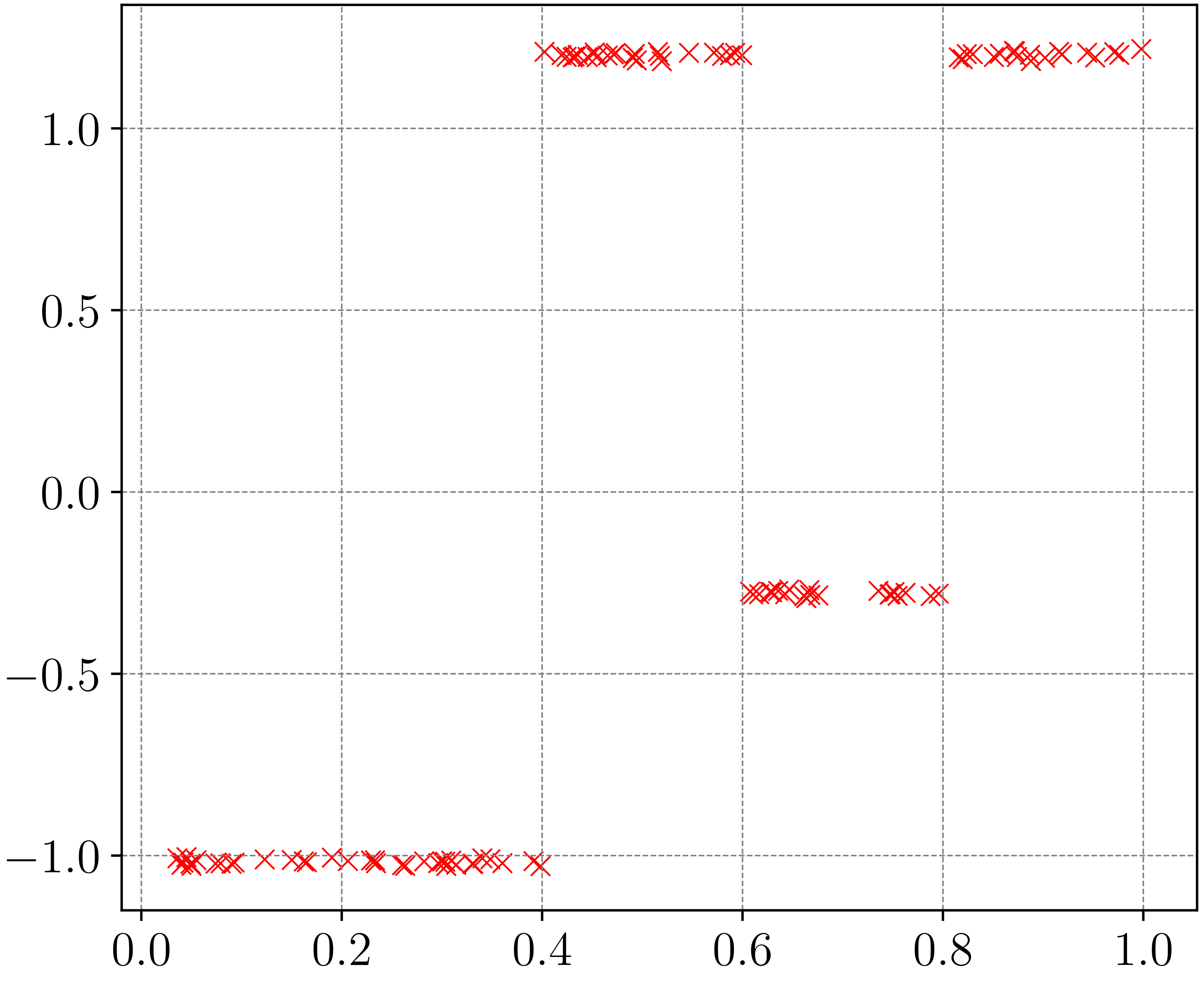


Figure 3. Multi-step function.

DEEP GP REGRESSION

Shortcomings

- Reliant on “local” approximations based on “inducing points”
- Scales quadratically in the number of local approximations
- Does not capture **global structure** in the data

Global structure in data

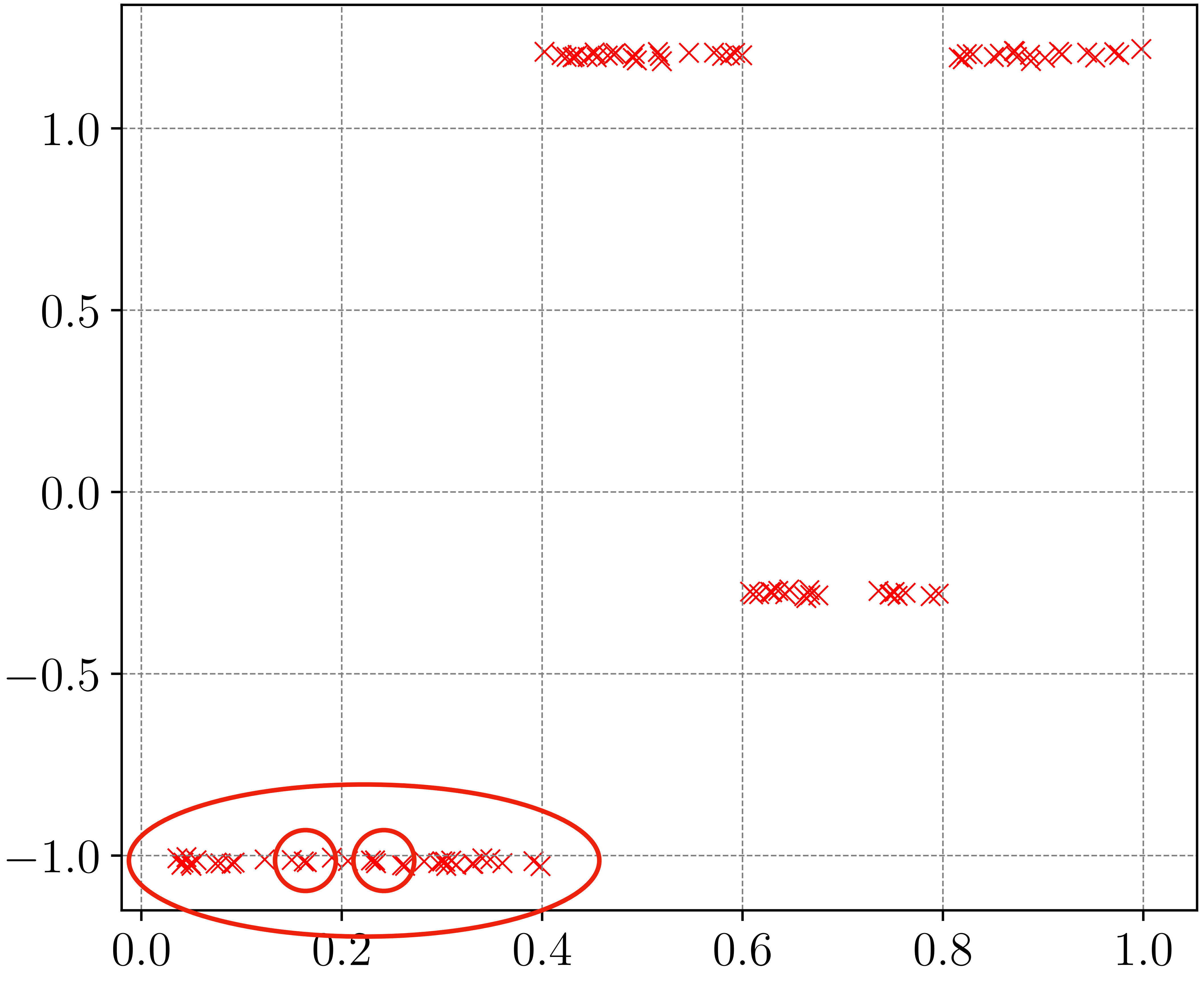


Figure 3. Multi-step function.

DEEP GP REGRESSION

Shortcomings

- Reliant on “local” approximations based on “inducing points”
- Scales quadratically in the number of local approximations
- Does not capture **global structure** in the data

Global structure in data

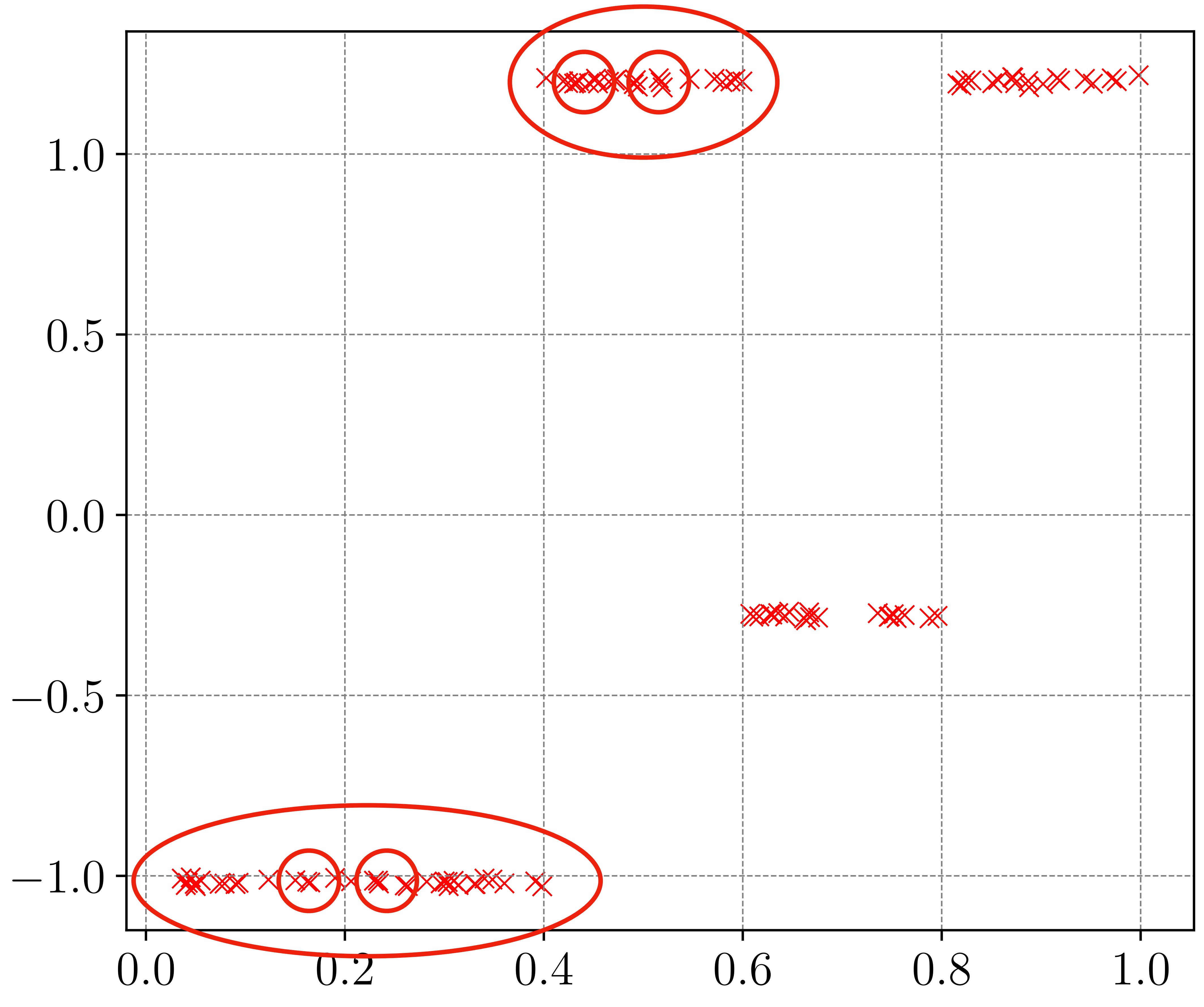


Figure 3. Multi-step function.

Deep GP Regression

Shortcomings

- Reliant on “local” approximations based on “inducing points”
- Scales quadratically in the number of local approximations
- Does not capture **global structure** in the data

Global structure in data

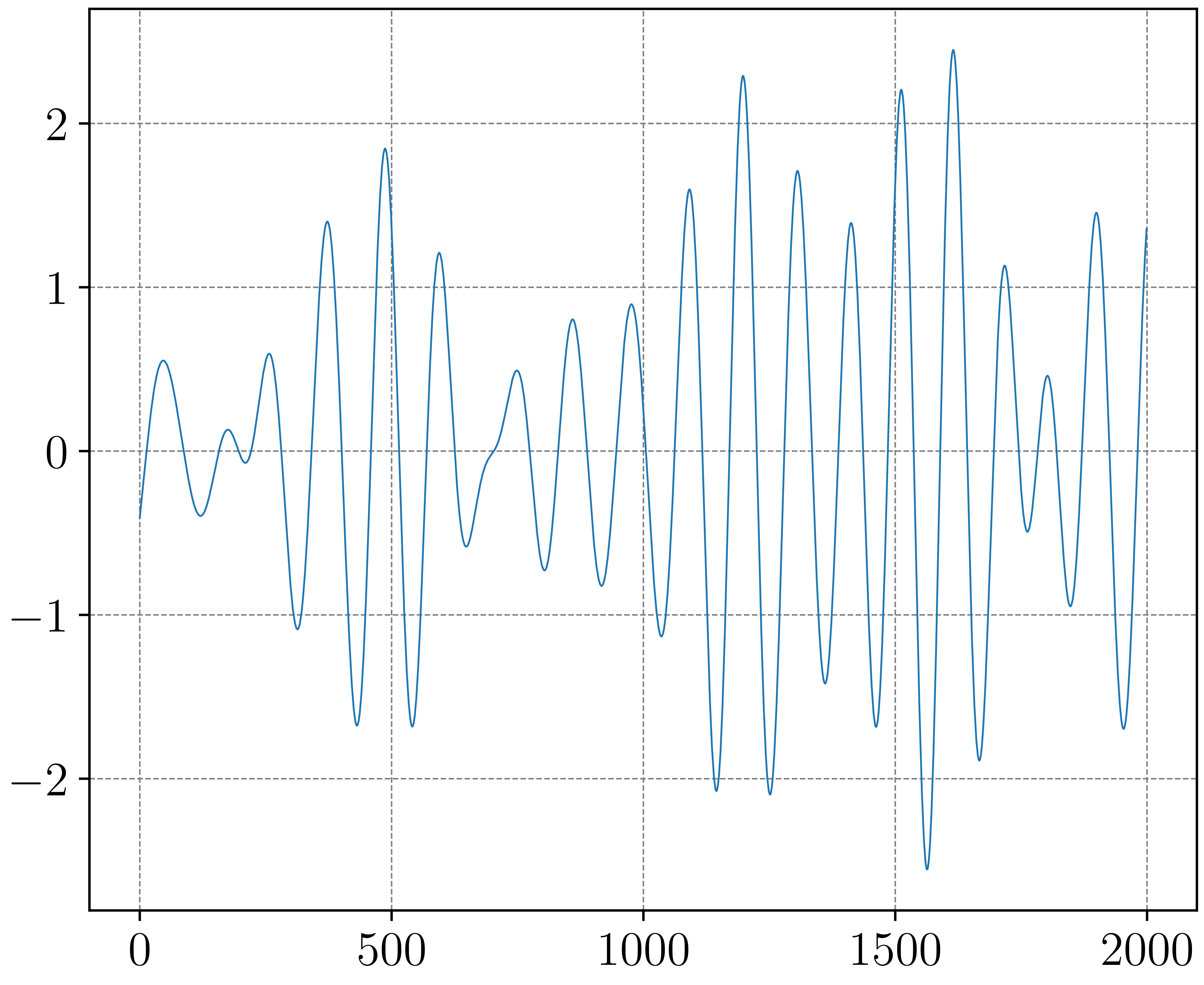


Figure 5. Audio signal.

Deep GP Regression

Shortcomings

- Reliant on “local” approximations based on “inducing points”
- Scales quadratically in the number of local approximations
- Does not capture **global structure** in the data

Global structure in data

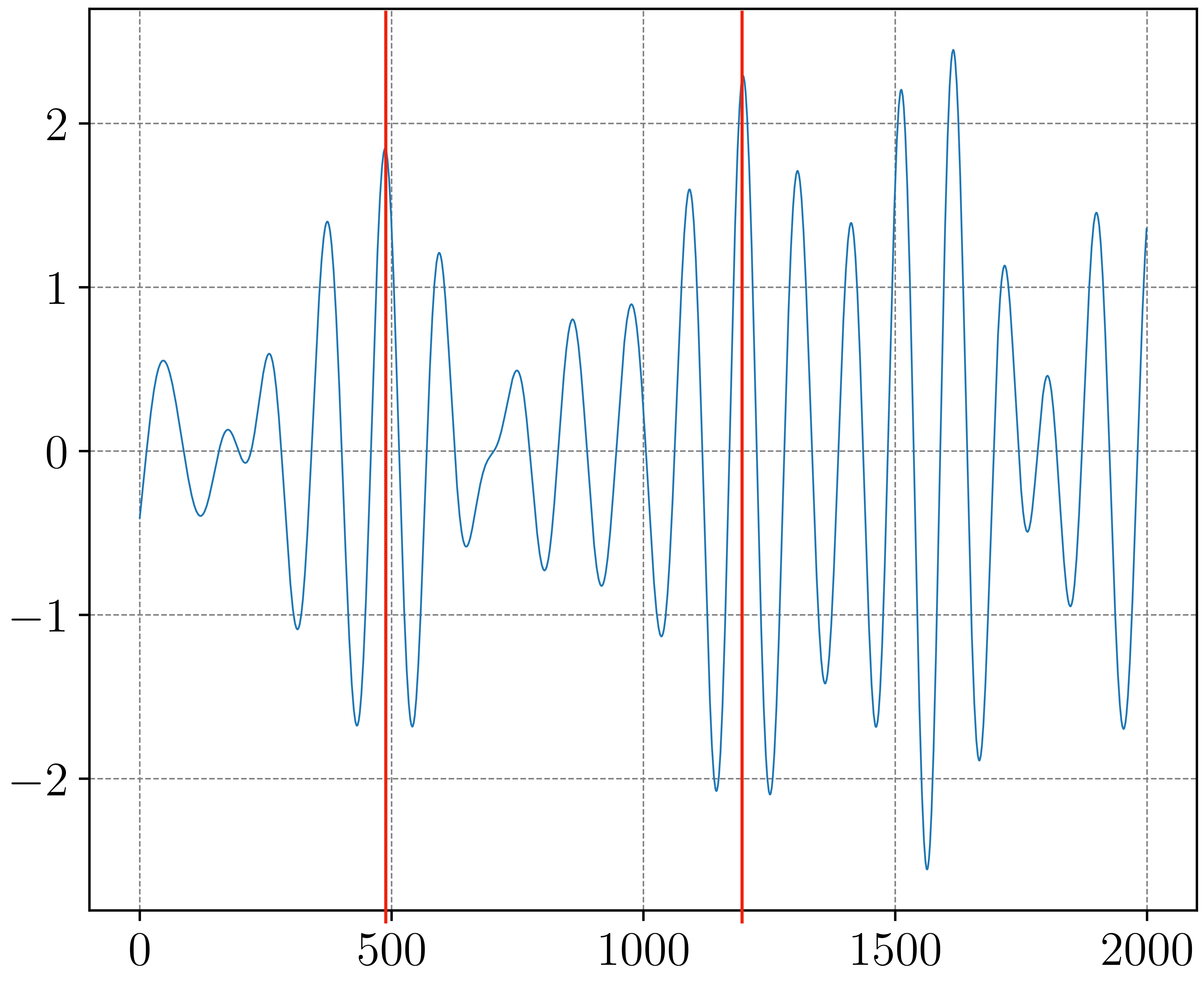


Figure 5. Audio signal.

Deep GP Regression

Shortcomings

- Reliant on "local" approximations based on "inducing points"
- Scales quadratically in the number of local approximations
- Does not capture **global structure** in the data

Global structure in data

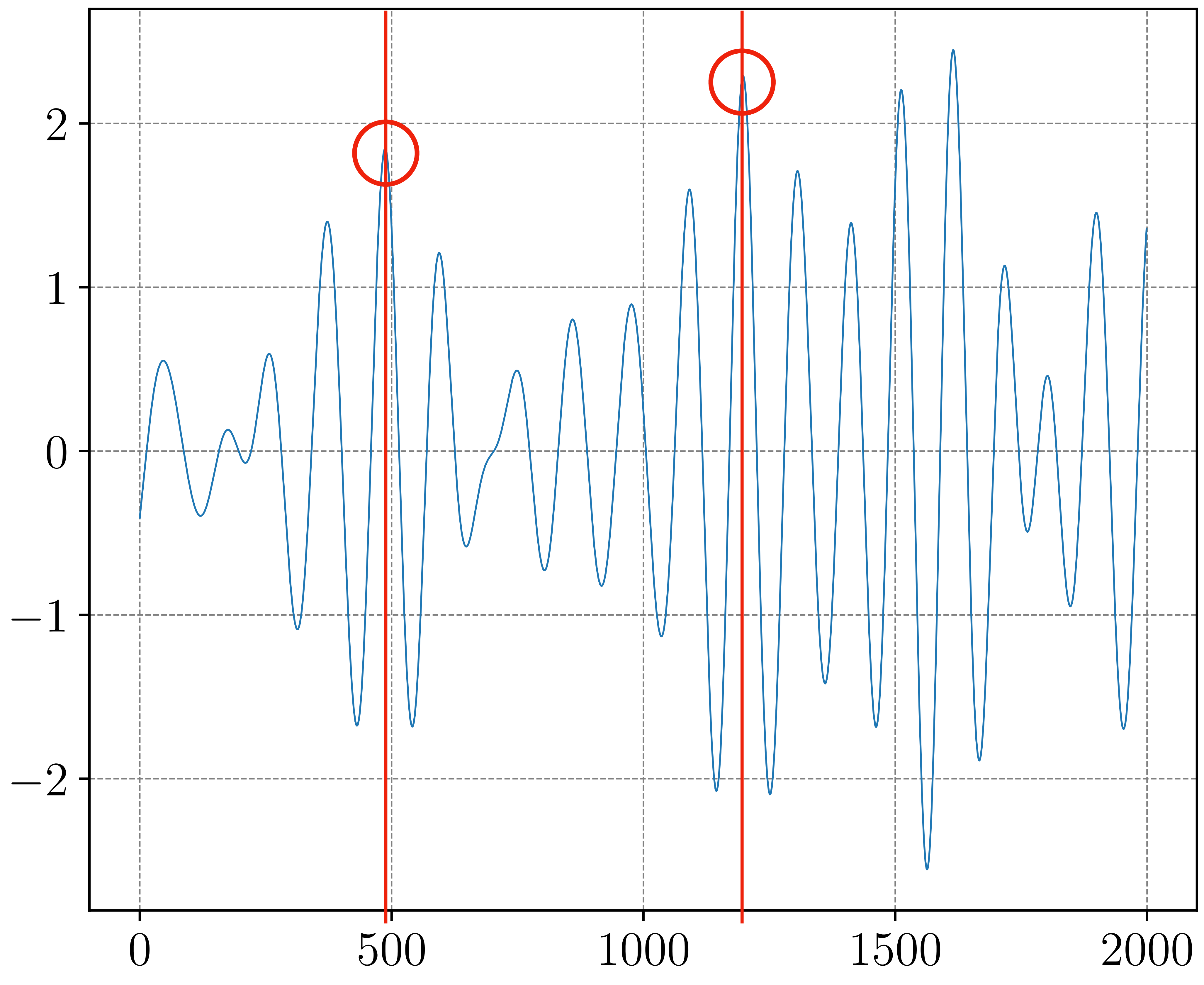


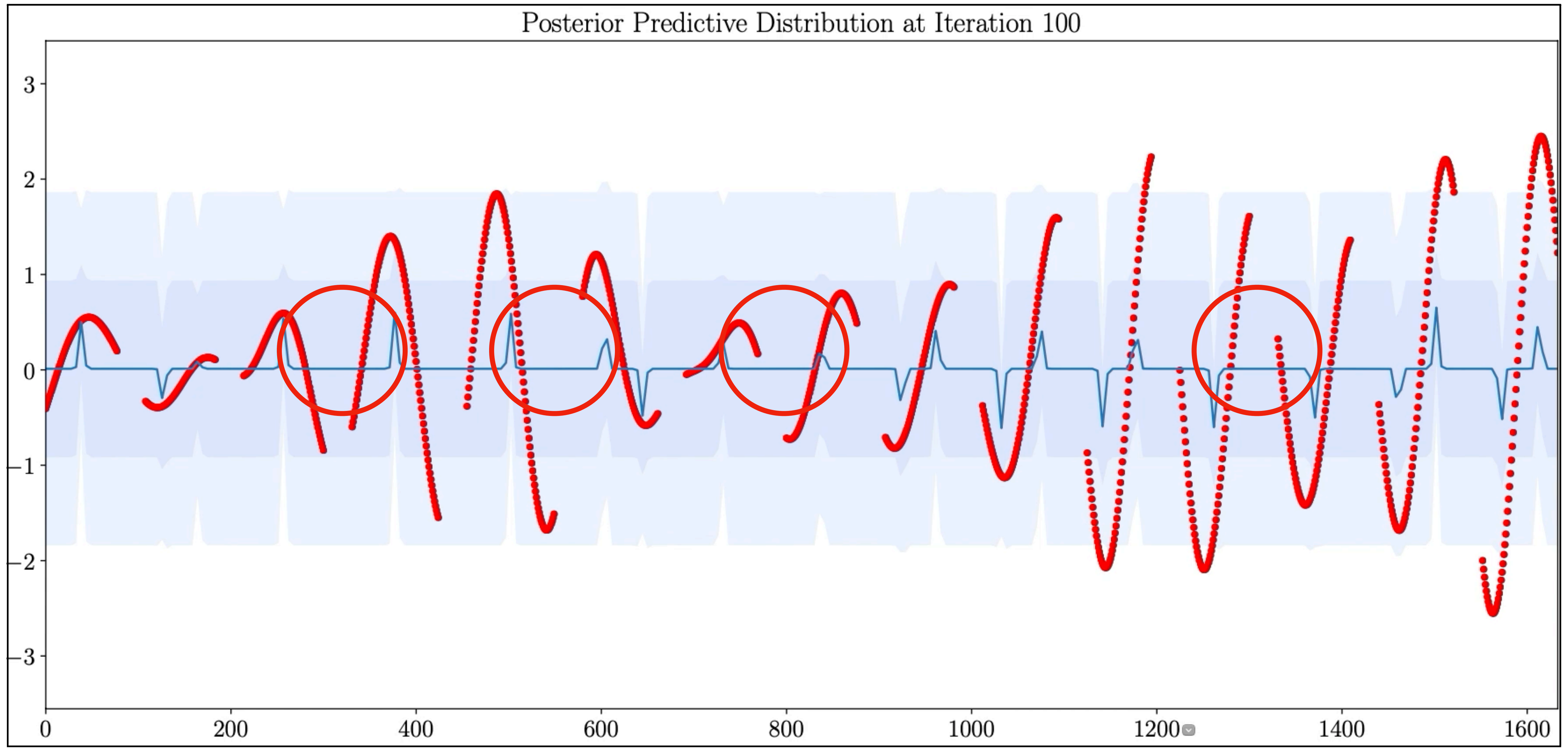
Figure 5. Audio signal.

Deep GP Regression

Shortcomings

- Reliant on "local" approximations based on "inducing points"
- Scales quadratically in the number of local approximations
- Does not capture **global structure** in the data

Global structure in data

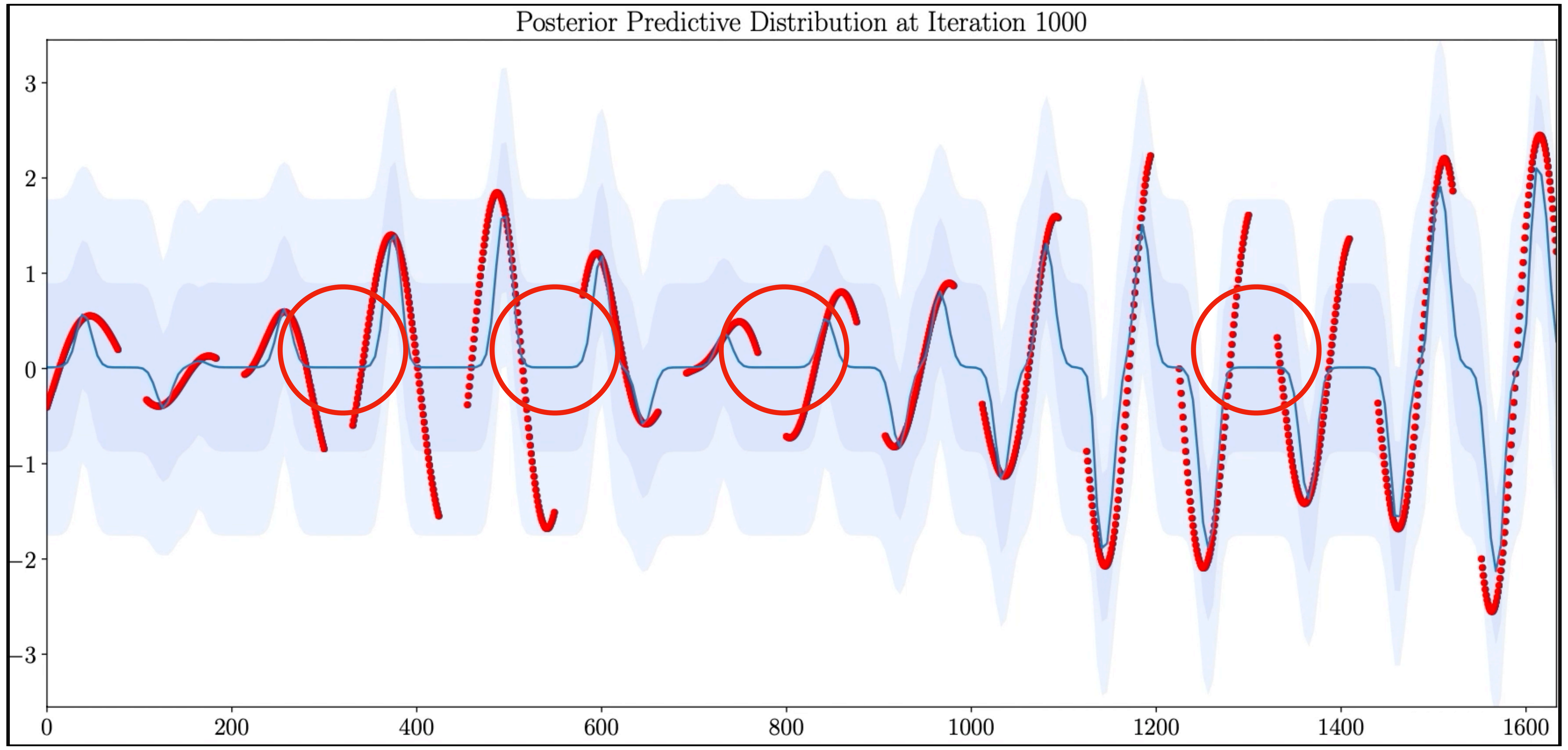


Deep GP Regression

Shortcomings

- Reliant on “local” approximations based on “inducing points”
- Scales quadratically in the number of local approximations
- Does not capture **global structure** in the data

Global structure in data

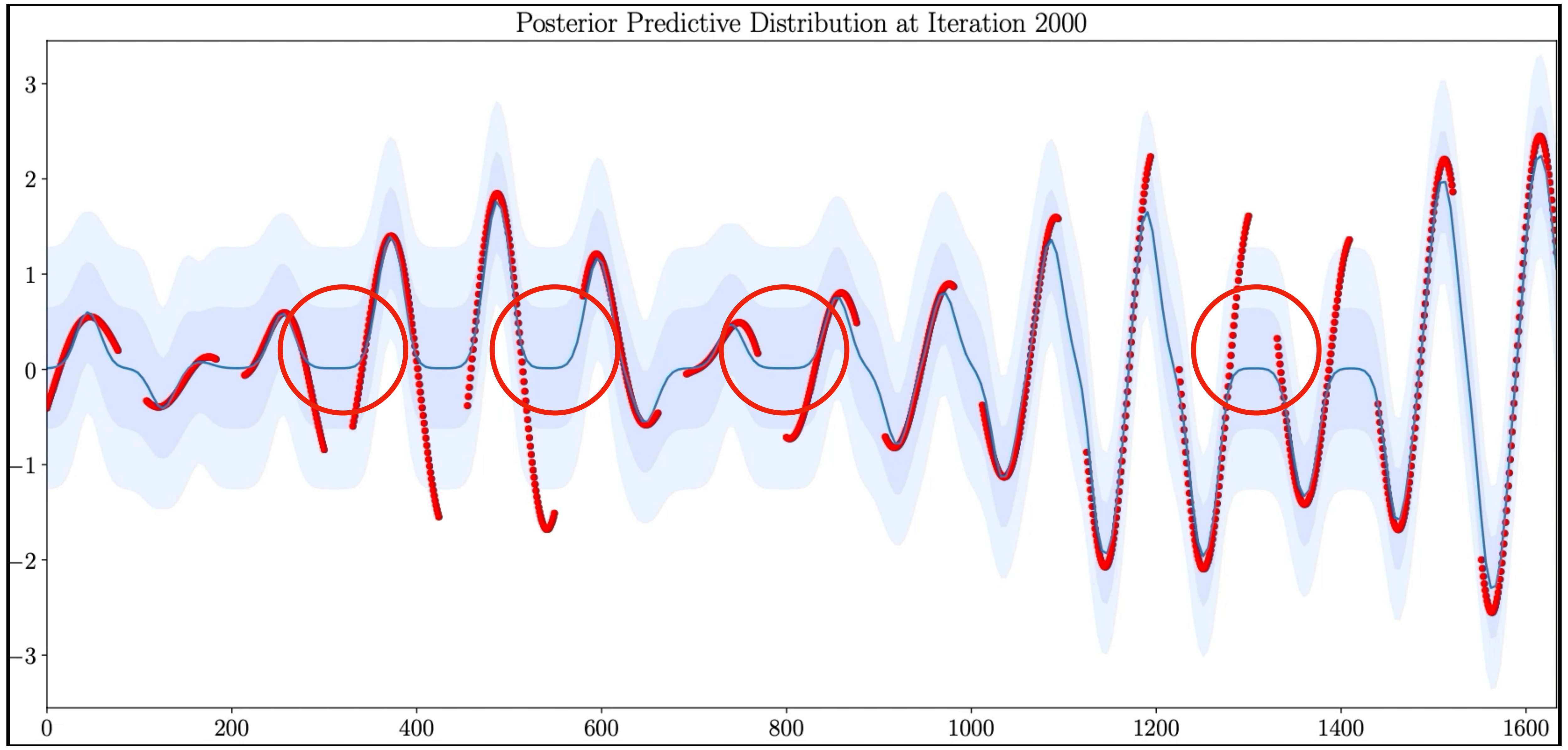


DEEP GP REGRESSION

Shortcomings

- Reliant on "local" approximations based on "inducing points"
- Scales quadratically in the number of local approximations
- Does not capture **global structure** in the data

Global structure in data

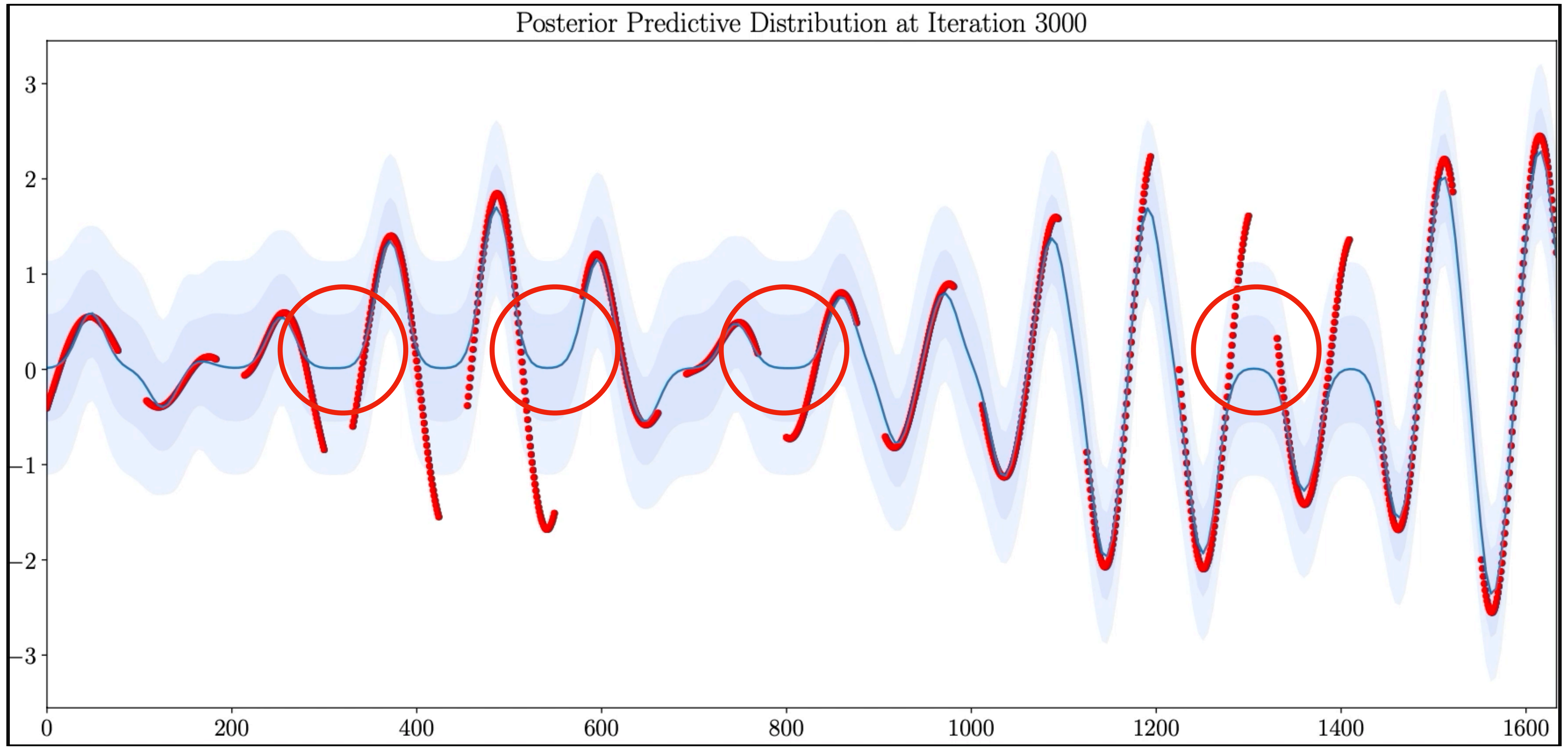


Deep GP Regression

Shortcomings

- Reliant on "local" approximations based on "inducing points"
- Scales quadratically in the number of local approximations
- Does not capture **global structure** in the data

Global structure in data



DEEP GP REGRESSION

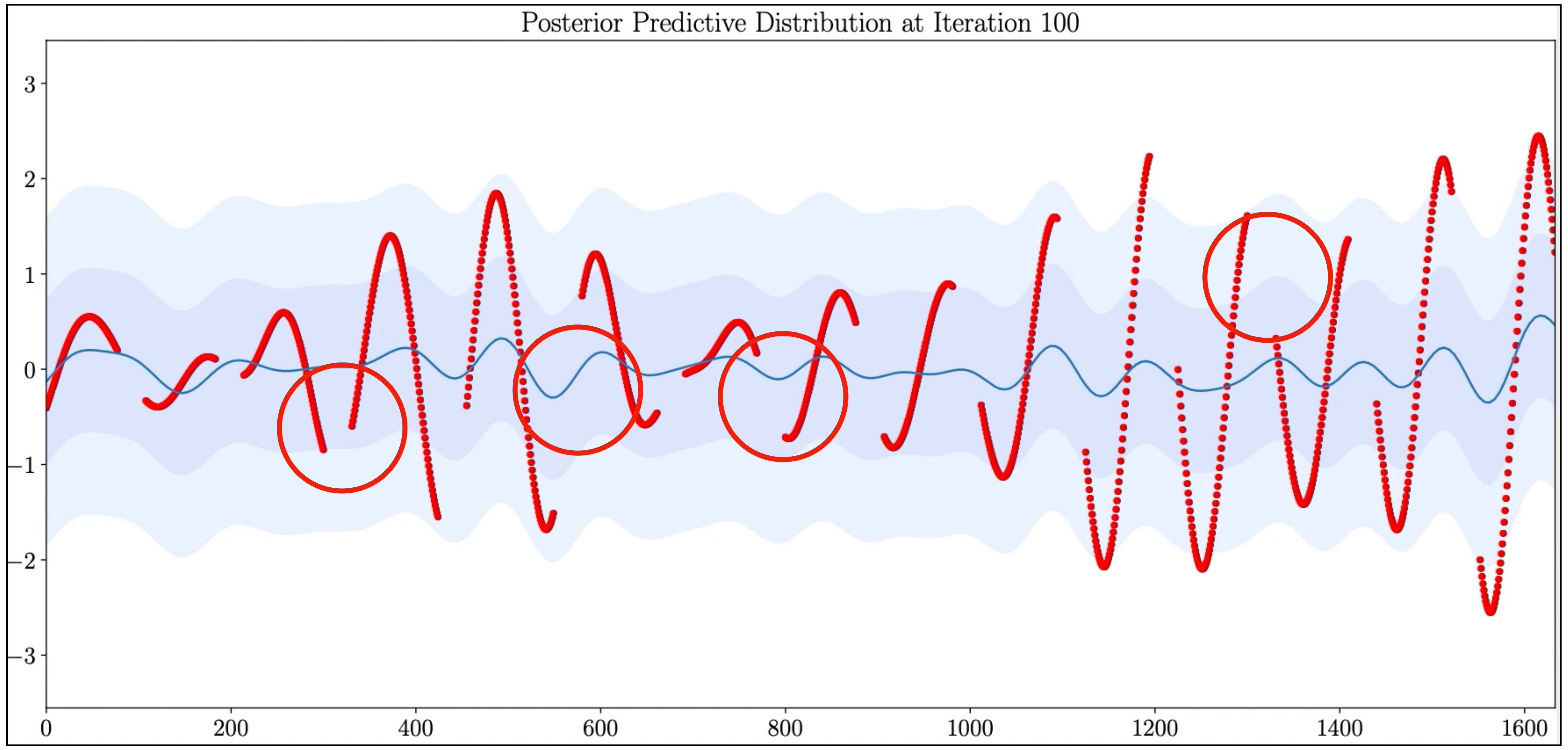
Shortcomings

- Reliant on “local” approximations based on “inducing points”
- Scales quadratically in the number of local approximations
- Does not capture **global structure** in the data

Solution?

- Incorporate **global structure** into approximation

Global structure in data



DEEP GP REGRESSION

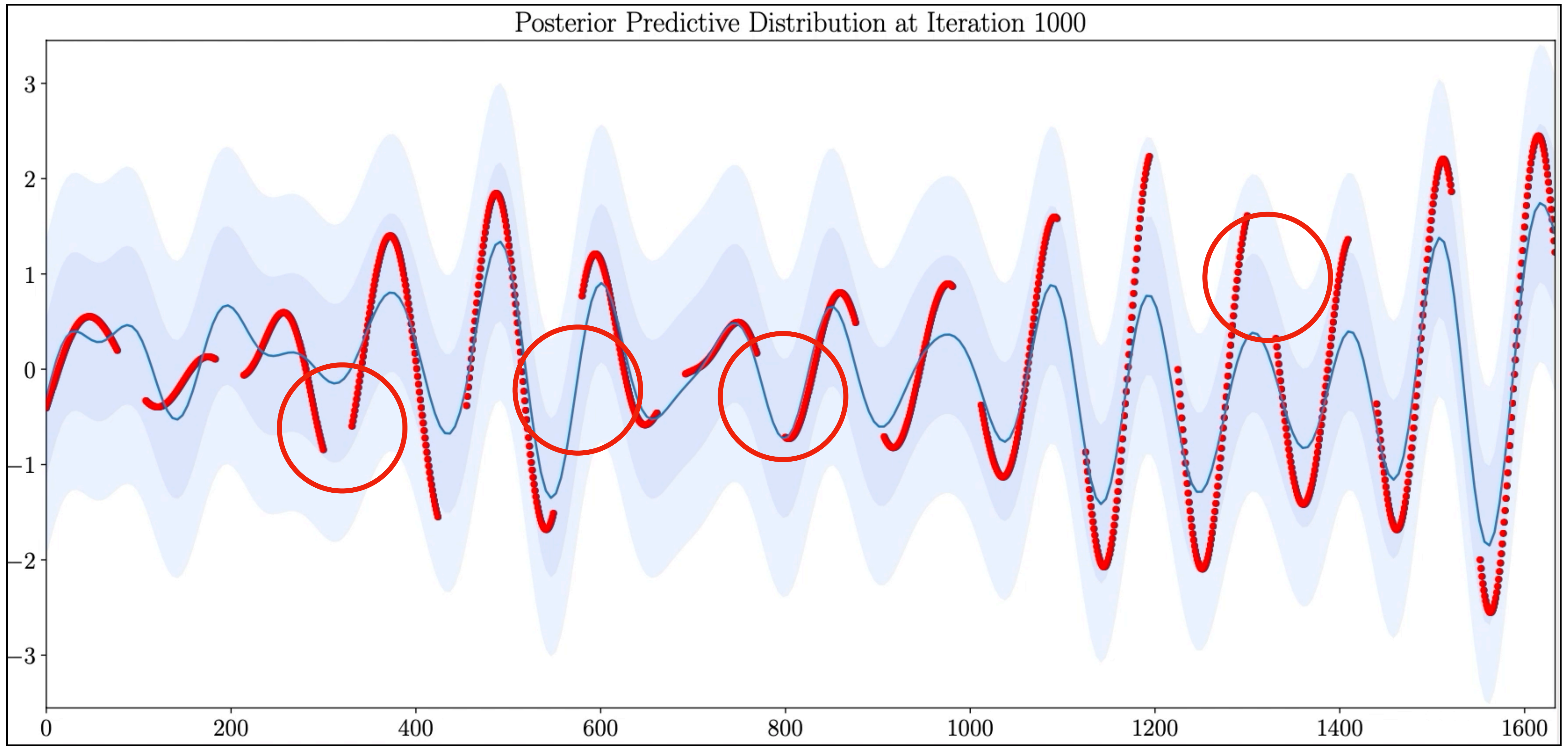
Shortcomings

- Reliant on “local” approximations based on “inducing points”
- Scales quadratically in the number of local approximations
- Does not capture **global structure** in the data

Solution?

- Incorporate **global structure** into approximation

Global structure in data



Deep GP Regression

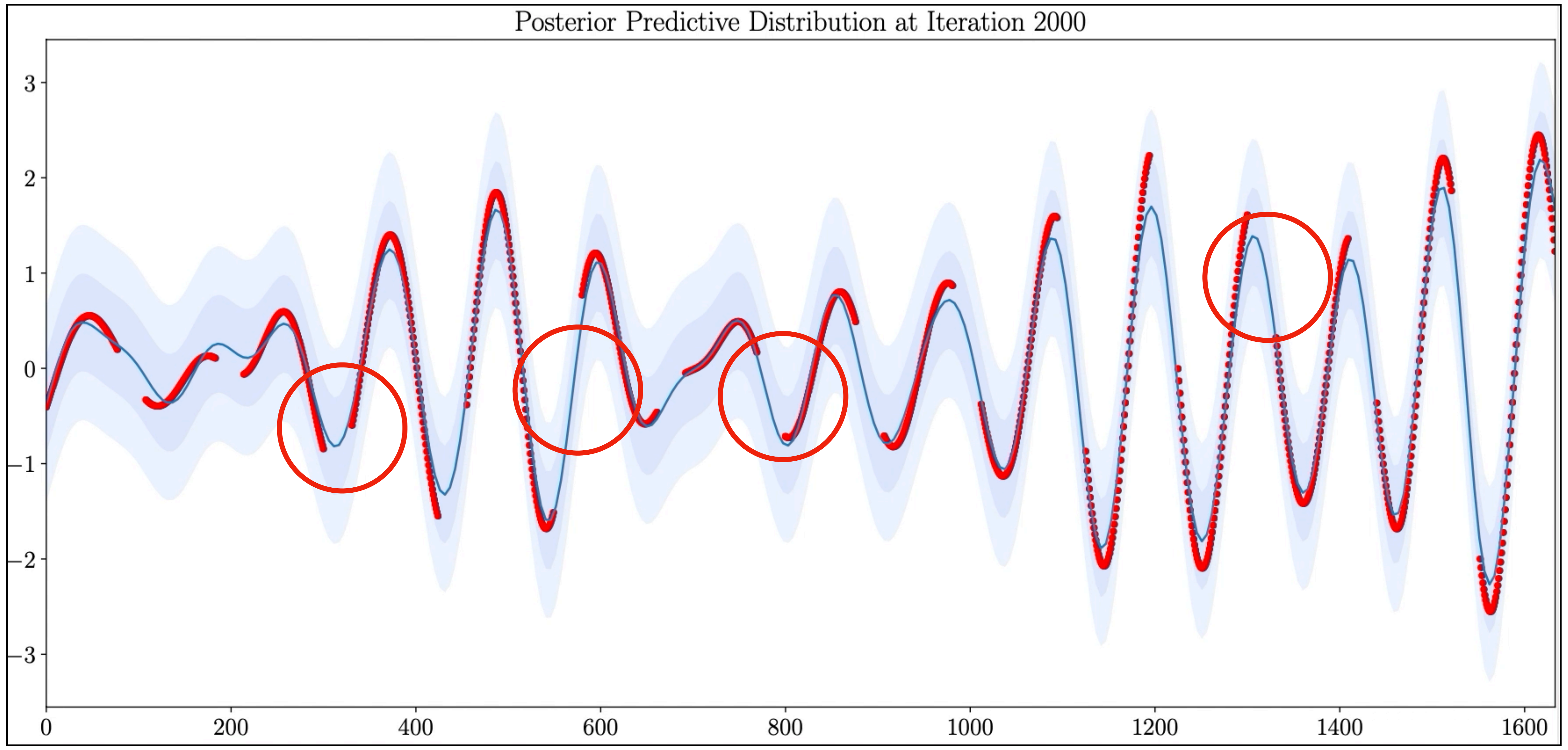
Shortcomings

- Reliant on "local" approximations based on "inducing points"
- Scales quadratically in the number of local approximations
- Does not capture **global structure** in the data

Solution?

- Incorporate **global structure** into approximation

Global structure in data

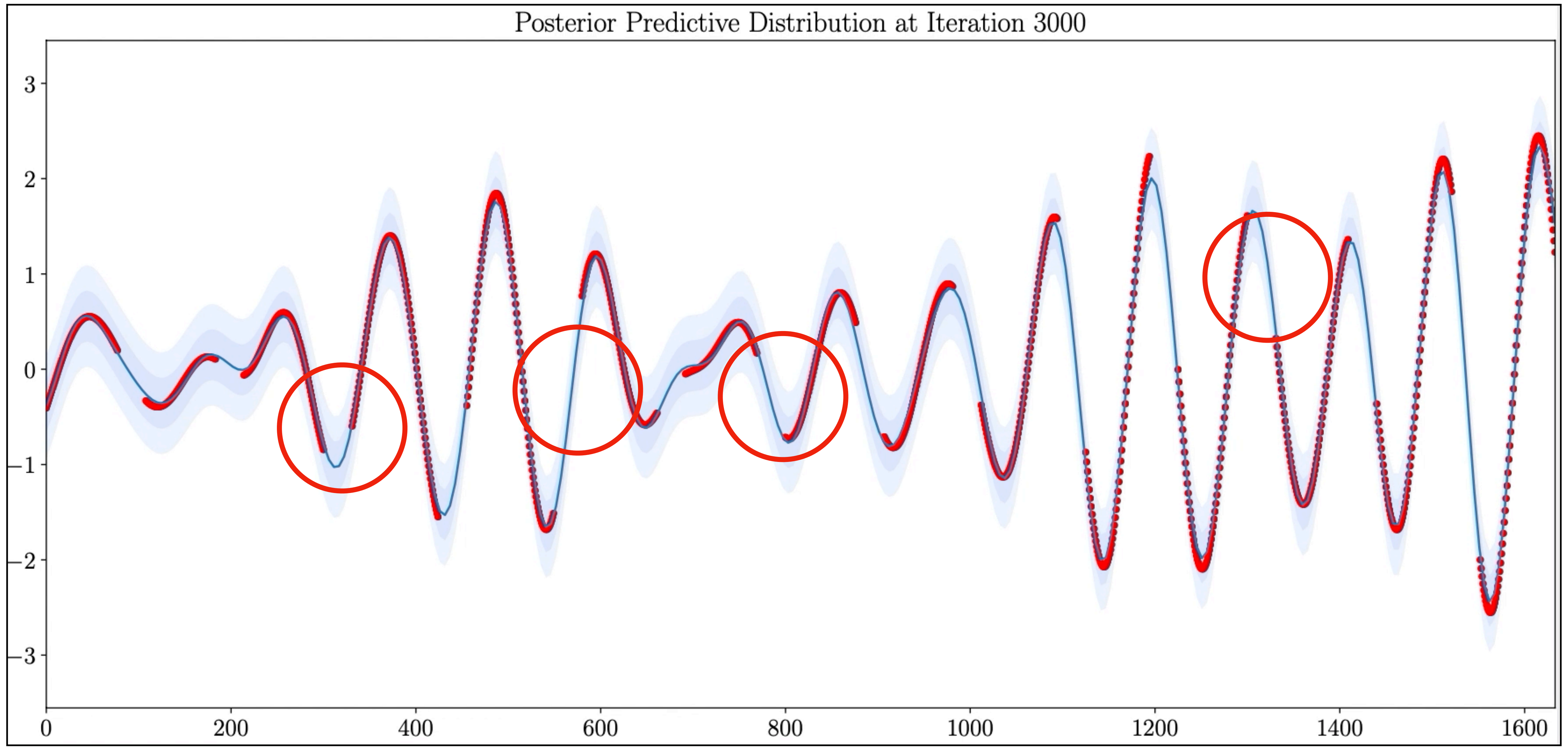


Deep GP Regression

Shortcomings

- Reliant on "local" approximations based on "inducing points"
- Scales quadratically in the number of local approximations
- Does not capture **global structure** in the data

Global structure in data



Solution?

- Incorporate **global structure** into approximation

Inter-domain Deep Gaussian Processes

Deep Gaussian Processes

- Deep Gaussian processes with L layers

$$\mathbf{y} = \mathbf{f}^{(L)} + \boldsymbol{\epsilon} = f^{(L)}\left(f^{(L-1)}(\ldots f(\mathbf{X}))\ldots\right) + \boldsymbol{\epsilon}$$

- Exact inference in this model is **intractable**
- Requires **approximate inference**

Variational Inference via Local Approximations

- Define inducing variables

$$u(\mathbf{z}) = f(\mathbf{z})$$

- Construct operators

$$\mathbf{K_{fu}} = \mathbf{K}(\mathbf{X}, \mathbf{Z}) \quad \mathbf{K_{uu}} \equiv \mathbf{K}(\mathbf{Z}, \mathbf{Z})$$

- Compute posterior predictive distribution

[1] **Michalis K. Titsias.** Variational Learning of Inducing Variables in Sparse Gaussian Processes. In AISTATS 2009.

Variational Inference via Local Approximations

- Define inducing variables

$$u(\mathbf{z}) = f(\mathbf{z})$$

- Construct operators

$$\mathbf{K}_{\mathbf{fu}} = \mathbf{K}(\mathbf{X}, \mathbf{Z}) \quad \boxed{\mathbf{K}_{\mathbf{uu}} = \mathbf{K}(\mathbf{Z}, \mathbf{Z})}$$

- Compute posterior predictive distribution

[1] **Michalis K. Titsias.** Variational Learning of Inducing Variables in Sparse Gaussian Processes. In AISTATS 2009.

Variational Inference via Local Approximations

- Define inducing variables

$$u(\mathbf{z}) = f(\mathbf{z})$$

- Construct operators

$$\boxed{\mathbf{K_{fu}} = \mathbf{K}(\mathbf{X}, \mathbf{Z})} \quad \boxed{\mathbf{K_{uu}} \equiv \mathbf{K}(\mathbf{Z}, \mathbf{Z})}$$

- Compute posterior predictive distribution

[1] **Michalis K. Titsias.** Variational Learning of Inducing Variables in Sparse Gaussian Processes. In AISTATS 2009.

Local vs. Global Approximations

- **Local** inducing variables

$$u(\mathbf{z}) = f(\mathbf{z})$$

- **Global** (inter-domain) inducing variables

$$u(\mathbf{z}) = \int_{\mathbb{R}^D} f(\mathbf{x}) g(\mathbf{x}, \mathbf{z}) \mathrm{d}\mathbf{x} \longrightarrow \mathbf{K}_{\mathbf{fu}}^{\phi} \quad \mathbf{K}_{\mathbf{uu}}^{\phi}$$

Local vs. Global Approximations

- **Local** inducing variables

$$u(\mathbf{z}) = f(\mathbf{z})$$

- **Global** (inter-domain) inducing variables

$$\boxed{u(\mathbf{z}) = \int_{\mathbb{R}^D} f(\mathbf{x}) g(\mathbf{x}, \mathbf{z}) \mathrm{d}\mathbf{x}} \longrightarrow \mathbf{K}^{\phi}_{\mathbf{fu}} \quad \mathbf{K}^{\phi}_{\mathbf{uu}}$$

Local vs. Global Approximations

- **Local** inducing variables

$$u(\mathbf{z}) = f(\mathbf{z})$$

- **Global** (inter-domain) inducing variables

$$u(\mathbf{z}) = \int_{\mathbb{R}^D} f(\mathbf{x}) g(\mathbf{x}, \mathbf{z}) \mathrm{d}\mathbf{x} \longrightarrow \mathbf{K}^{\phi}_{\mathbf{fu}} \quad \boxed{\mathbf{K}^{\phi}_{\mathbf{uu}}}$$

Local vs. Global Approximations

- **Local** inducing variables

$$u(\mathbf{z}) = f(\mathbf{z})$$

- **Global** (inter-domain) inducing variables

$$u(\mathbf{z}) = \int_{\mathbb{R}^D} f(\mathbf{x}) g(\mathbf{x}, \mathbf{z}) \mathrm{d}\mathbf{x} \longrightarrow \mathbf{K}^{\phi}_{\mathbf{fu}} \quad \mathbf{K}^{\phi}_{\mathbf{uu}}$$

RKHS Fourier Features

- Define truncated Fourier basis:

$$\phi(x) = \left[1, \cos\left(\omega_1(x-a)\right), \ldots, \cos\left(\omega_M(x-a)\right), \sin\left(\omega_1(x-a)\right), \ldots, \sin\left(\omega_M(x-a)\right)\right]^\top$$

- Inter-domain operators

$$\mathbf{K}_{\mathbf{fu}}^{\phi} = \phi(\mathbf{X})$$

$$\mathbf{K}_{\mathbf{uu}}^{\phi} = \langle \phi, \phi \rangle_{\mathcal{H}}$$

contain **information** about **global structure**[4]

[4] **James Hensman, Nicolas Durrande, and Arno Solin.** Variational Fourier features for Gaussian processes. Journal of Machine Learning Research 2018.

How can we use inter-domain transformations?

- Damianou & Lawrence (2013)[2] construct posterior from:

$$\int \mathbf{K}_{\mathbf{f}^{\ell}\mathbf{u}^{\ell}}^{\phi\top} q\left(\mathbf{f}_n^{(\ell)}\right) \mathrm{d}\mathbf{f}_n^{(\ell)}$$

[2] **Andreas Damianou and Neil Lawrence.** Deep Gaussian processes. In AISTATS 2013.

How can we use inter-domain transformations?

- Damianou & Lawrence (2013)[2] construct posterior from:

$$\int \mathbf{K}_{\mathbf{f}^\ell \mathbf{u}^\ell}^{\phi\,\top} q\left(\mathbf{f}_n^{(\ell)}\right) \mathrm{d}\mathbf{f}_n^{(\ell)}$$

[2] **Andreas Damianou and Neil Lawrence.** Deep Gaussian processes. In AISTATS 2013.

How can we use inter-domain transformations?

- Damianou & Lawrence (2013)[2] construct posterior from:

$$\int \mathbf{K}_{\mathbf{f}^{\ell}\mathbf{u}^{\ell}}^{\phi\top} q\left(\mathbf{f}_n^{(\ell)}\right) \mathrm{d}\mathbf{f}_n^{(\ell)}$$

Hard to compute!

[2] **Andreas Damianou and Neil Lawrence**. Deep Gaussian processes. In AISTATS 2013.

INFERENCE VIA GLOBAL APPROXIMATION

Can we do better? Yes!

- Use **different factorization** of variational posterior
- Doubly stochastic variational inference (DSVI)[3]:

$$q\left(\left\{\mathbf{F}^{(\ell)}\right\}_{\ell=1}^{L}\right) = \prod_{\ell=1}^{L} \mathcal{N}\left(\mathbf{F}^{(\ell)} | \widetilde{\mathbf{m}}^{(\ell)}, \widetilde{\mathbf{S}}^{(\ell)}\right)$$

with

$$\widetilde{\mathbf{m}}\left(\mathbf{F}^{(\ell)}\right) \equiv \mathbf{m}_{\mathbf{f}^{\ell}} - \mathbf{K}_{\mathbf{f}^{\ell}\mathbf{u}^{\ell}}\mathbf{K}_{\mathbf{u}^{\ell}\mathbf{u}^{\ell}}^{-1}\left(\boldsymbol{\mu}^{(\ell)} - \mathbf{m}_{\mathbf{u}^{\ell}}\right)$$

$$\widetilde{\mathbf{S}}\left(\mathbf{F}^{(\ell)}, \mathbf{F}^{(\ell)}\right) \equiv \mathbf{K}_{\mathbf{f}^{\ell}\mathbf{f}^{\ell}} - \mathbf{K}_{\mathbf{f}^{\ell}\mathbf{u}^{\ell}}\mathbf{K}_{\mathbf{u}^{\ell}\mathbf{u}^{\ell}}^{-1}\left(\mathbf{K}_{\mathbf{u}^{\ell}\mathbf{u}^{\ell}} - \boldsymbol{\Sigma}^{(\ell)}\right)\mathbf{K}_{\mathbf{u}^{\ell}\mathbf{u}^{\ell}}^{-1}\mathbf{K}_{\mathbf{u}^{\ell}\mathbf{f}^{\ell}}$$

[3] **Hugh Salimbeni and Marc Deisenroth**. Doubly Stochastic Variational Inference for Deep Gaussian Processes. In NeurIPS 2017.

Inference via Global Approximation

Can we do better? Yes!

- Use **different factorization** of variational posterior
- Doubly stochastic variational inference (DSVI)[3]:

$$q\left(\left\{\mathbf{F}^{(\ell)}\right\}_{\ell=1}^{L}\right) = \prod_{\ell=1}^{L} \mathcal{N}\left(\mathbf{F}^{(\ell)} | \widetilde{\mathbf{m}}^{(\ell)}, \widetilde{\mathbf{S}}^{(\ell)}\right)$$

with

$$\widetilde{\mathbf{m}}\left(\mathbf{F}^{(\ell)}\right) \equiv \mathbf{m}_{\mathbf{f}^{\ell}} - \boxed{\mathbf{K}_{\mathbf{f}^{\ell}\mathbf{u}^{\ell}} \mathbf{K}_{\mathbf{u}^{\ell}\mathbf{u}^{\ell}}^{-1}} \left(\boldsymbol{\mu}^{(\ell)} - \mathbf{m}_{\mathbf{u}^{\ell}}\right)$$

$$\widetilde{\mathbf{S}}\left(\mathbf{F}^{(\ell)}, \mathbf{F}^{(\ell)}\right) \equiv \mathbf{K}_{\mathbf{f}^{\ell}\mathbf{f}^{\ell}} - \boxed{\mathbf{K}_{\mathbf{f}^{\ell}\mathbf{u}^{\ell}} \mathbf{K}_{\mathbf{u}^{\ell}\mathbf{u}^{\ell}}^{-1}} \left(\boxed{\mathbf{K}_{\mathbf{u}^{\ell}\mathbf{u}^{\ell}}} - \boldsymbol{\Sigma}^{(\ell)}\right) \boxed{\mathbf{K}_{\mathbf{u}^{\ell}\mathbf{u}^{\ell}}^{-1} \mathbf{K}_{\mathbf{u}^{\ell}\mathbf{f}^{\ell}}}$$

[3] **Hugh Salimbeni and Marc Deisenroth**. Doubly Stochastic Variational Inference for Deep Gaussian Processes. In NeurIPS 2017.

Inter-domain Deep GP approximate posterior

- Use inter-domain operators $\mathbf{K}_{\mathbf{fu}}^{\phi}$ and $\mathbf{K}_{\mathbf{uu}}^{\phi}$ as **drop-in replacements**
- Posterior predictive distribution under DSVI[3]:

$$q\left(\left\{\mathbf{F}^{(\ell)}\right\}_{\ell=1}^{L}\right)=\prod_{\ell=1}^{L}\mathcal{N}\left(\mathbf{F}^{(\ell)}|\widetilde{\mathbf{m}}^{(\ell)},\widetilde{\mathbf{S}}^{(\ell)}\right)$$

with

$$\widetilde{\mathbf{m}}\left(\mathbf{F}^{(\ell)}\right)\equiv\mathbf{m}_{\mathbf{f}^{\ell}}-\mathbf{K}_{\mathbf{f}'\mathbf{u}^{\ell}}^{\phi}\mathbf{K}_{\mathbf{u}^{\ell}\mathbf{u}^{\ell}}^{\phi^{-1}}\left(\boldsymbol{\mu}^{(\ell)}-\mathbf{m}_{\mathbf{u}^{\ell}}\right)$$

$$\widetilde{\mathbf{S}}\left(\mathbf{F}^{(\ell)},\mathbf{F}^{(\ell)}\right)\equiv\mathbf{K}_{\mathbf{f}^{\ell}\mathbf{f}^{\ell}}-\mathbf{K}_{\mathbf{f}^{\ell}\mathbf{u}^{\ell}}^{\phi}\mathbf{K}_{\mathbf{u}^{\ell}\mathbf{u}^{\ell}}^{\phi^{-1}}\left(\mathbf{K}_{\mathbf{u}^{\ell}\mathbf{u}^{\ell}}^{\phi}-\boldsymbol{\Sigma}^{(\ell)}\right)\mathbf{K}_{\mathbf{u}^{\ell}\mathbf{u}^{\ell}}^{\phi^{-1}}\mathbf{K}_{\mathbf{u}^{\ell}\mathbf{f}^{\ell}}^{\phi}$$

[3] **Hugh Salimbeni and Marc Deisenroth**. Doubly Stochastic Variational Inference for Deep Gaussian Processes. In NeurIPS 2017.

Inter-domain Deep GP approximate posterior

- Use inter-domain operators $\mathbf{K}^{\phi}_{\mathbf{fu}}$ and $\mathbf{K}^{\phi}_{\mathbf{uu}}$ as **drop-in replacements**
- Posterior predictive distribution under DSVI[3]:

$$q\left(\left\{\mathbf{F}^{(\ell)}\right\}_{\ell=1}^{L}\right)=\prod_{\ell=1}^{L}\mathcal{N}\left(\mathbf{F}^{(\ell)}|\widetilde{\mathbf{m}}^{(\ell)},\widetilde{\mathbf{S}}^{(\ell)}\right)$$

with

$$\widetilde{\mathbf{m}}\left(\mathbf{F}^{(\ell)}\right)\equiv\mathbf{m}_{\mathbf{f}^{\ell}}-\mathbf{K}^{\phi}_{\mathbf{f}'\mathbf{u}^{\ell}}\mathbf{K}^{\phi^{-1}}_{\mathbf{u}^{\ell}\mathbf{u}^{\ell}}\left(\boldsymbol{\mu}^{(\ell)}-\mathbf{m}_{\mathbf{u}^{\ell}}\right)$$

$$\widetilde{\mathbf{S}}\left(\mathbf{F}^{(\ell)},\mathbf{F}^{(\ell)}\right)\equiv\mathbf{K}_{\mathbf{f}^{\ell}\mathbf{f}^{\ell}}-\mathbf{K}^{\phi}_{\mathbf{f}^{\ell}\mathbf{u}^{\ell}}\mathbf{K}^{\phi^{-1}}_{\mathbf{u}^{\ell}\mathbf{u}^{\ell}}\left(\mathbf{K}^{\phi}_{\mathbf{u}^{\ell}\mathbf{u}^{\ell}}-\boldsymbol{\Sigma}^{(\ell)}\right)\mathbf{K}^{\phi^{-1}}_{\mathbf{u}^{\ell}\mathbf{u}^{\ell}}\mathbf{K}^{\phi}_{\mathbf{u}^{\ell}\mathbf{f}^{\ell}}$$

[3] **Hugh Salimbeni and Marc Deisenroth**. Doubly Stochastic Variational Inference for Deep Gaussian Processes. In NeurIPS 2017.

Inference via Global Approximation

Inter-domain Deep GP approximate posterior

- Use inter-domain operators $\boxed{\mathbf{K}^{\phi}_{\mathbf{fu}}}$ and $\boxed{\mathbf{K}^{\phi}_{\mathbf{uu}}}$ as **drop-in replacements**
- Posterior predictive distribution under DSVI[3]:

$$q\left(\left\{\mathbf{F}^{(\ell)}\right\}_{\ell=1}^{L}\right) = \prod_{\ell=1}^{L} \mathcal{N}\left(\mathbf{F}^{(\ell)} | \widetilde{\mathbf{m}}^{(\ell)}, \widetilde{\mathbf{S}}^{(\ell)}\right)$$

Drop-in replacements

with

$$\widetilde{\mathbf{m}}\left(\mathbf{F}^{(\ell)}\right) \equiv \mathbf{m}_{\mathbf{f}^{\ell}} - \boxed{\mathbf{K}^{\phi}_{\mathbf{f}'\mathbf{u}^{\ell}} \mathbf{K}^{\phi^{-1}}_{\mathbf{u}^{\ell}\mathbf{u}^{\ell}}} \left(\boldsymbol{\mu}^{(\ell)} - \mathbf{m}_{\mathbf{u}^{\ell}}\right)$$

$$\widetilde{\mathbf{S}}\left(\mathbf{F}^{(\ell)}, \mathbf{F}^{(\ell)}\right) \equiv \mathbf{K}_{\mathbf{f}^{\ell}\mathbf{f}^{\ell}} - \boxed{\mathbf{K}^{\phi}_{\mathbf{f}^{\ell}\mathbf{u}^{\ell}} \mathbf{K}^{\phi^{-1}}_{\mathbf{u}^{\ell}\mathbf{u}^{\ell}}} \left(\boxed{\mathbf{K}^{\phi}_{\mathbf{u}^{\ell}\mathbf{u}^{\ell}}} - \boldsymbol{\Sigma}^{(\ell)}\right) \boxed{\mathbf{K}^{\phi^{-1}}_{\mathbf{u}^{\ell}\mathbf{u}^{\ell}} \mathbf{K}^{\phi}_{\mathbf{u}^{\ell}\mathbf{f}^{\ell}}}$$

[3] **Hugh Salimbeni and Marc Deisenroth.** Doubly Stochastic Variational Inference for Deep Gaussian Processes. In NeurIPS 2017.

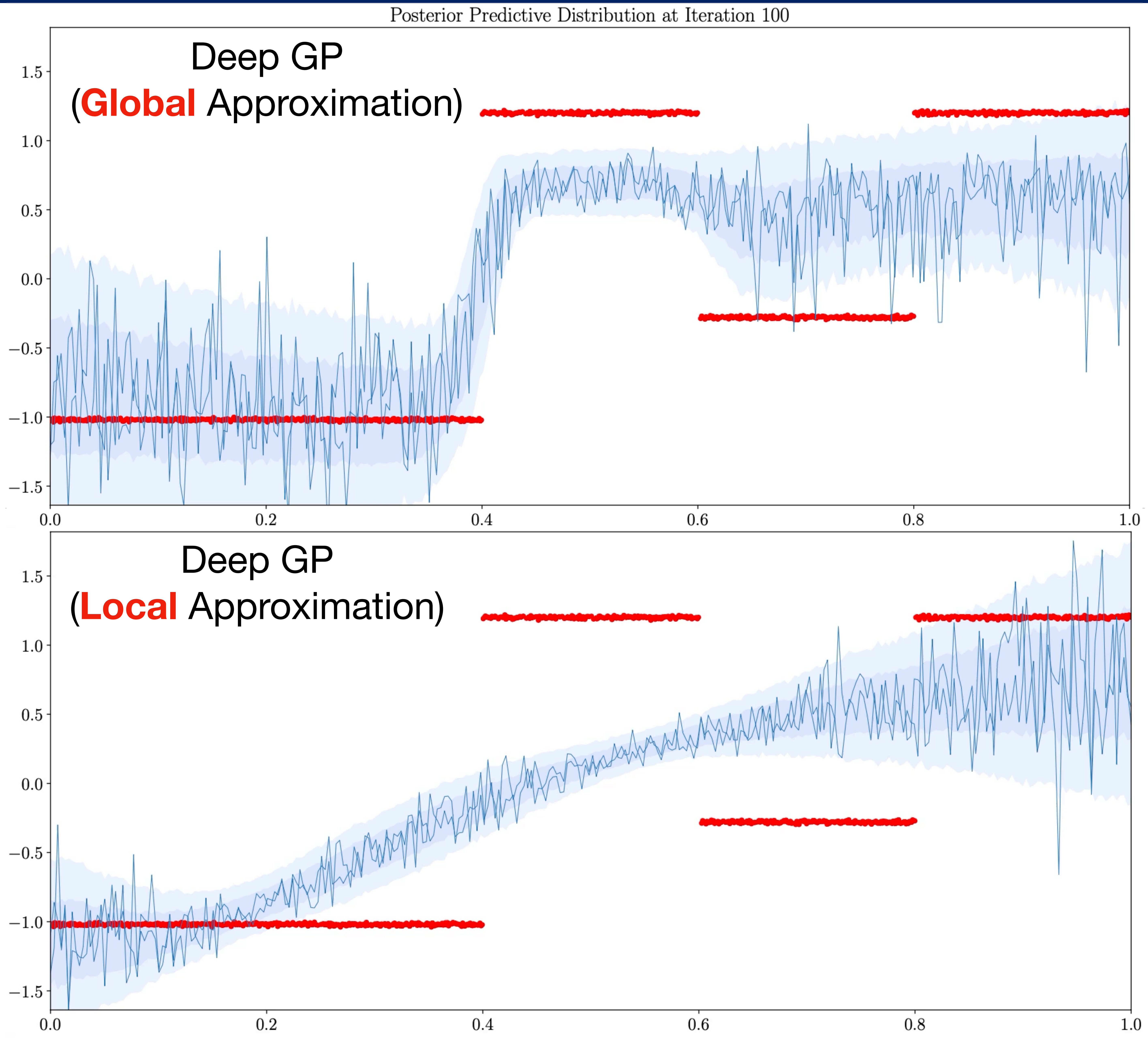
Empirical Evaluation

Experiment: Multi-step Function

Convergence & Fit

Converges within 500 iterations

Never attains a good fit

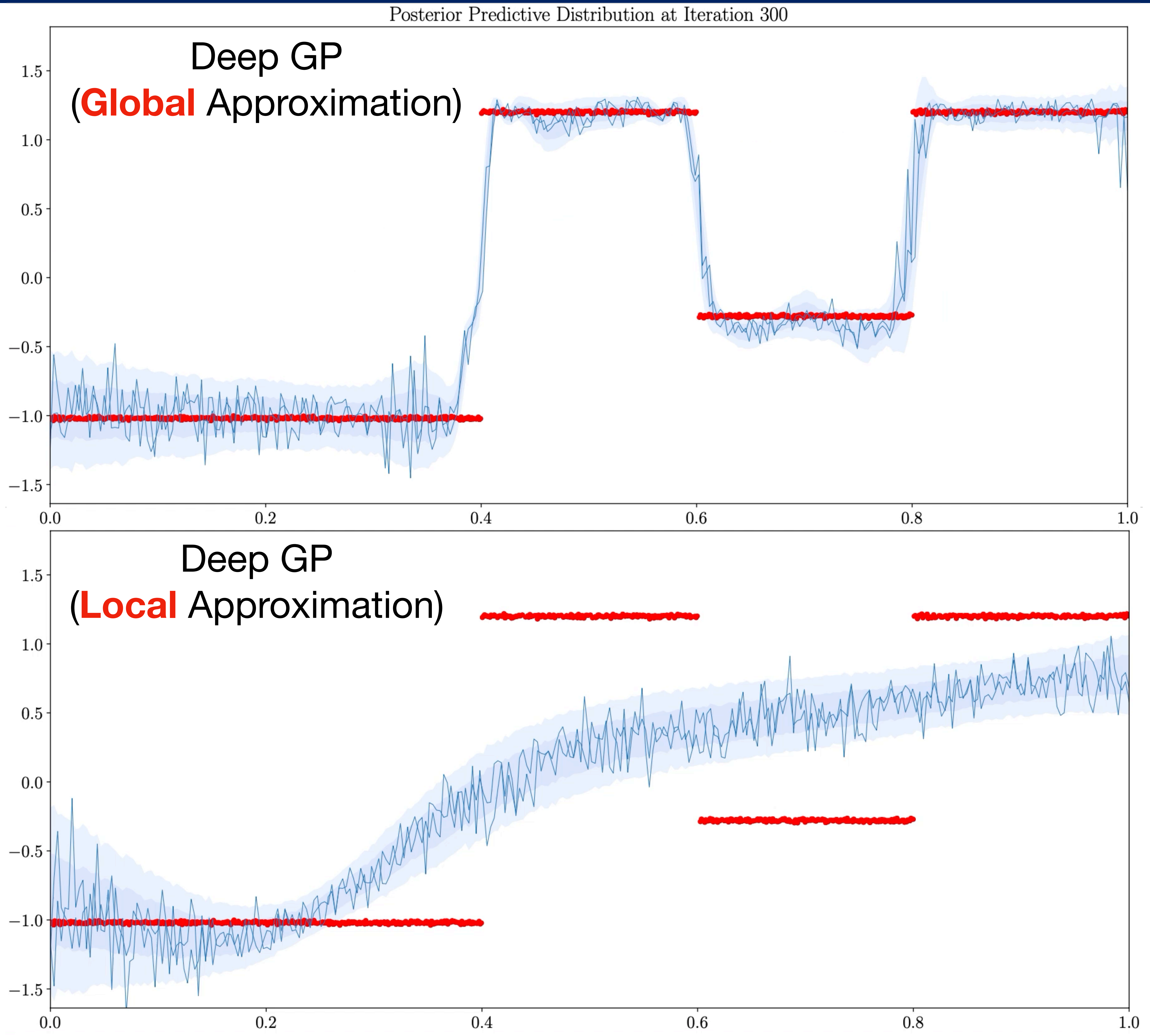


EXPERIMENT: MULTI-STEP FUNCTION

Convergence & Fit

Converges within 500 iterations

Never attains a good fit

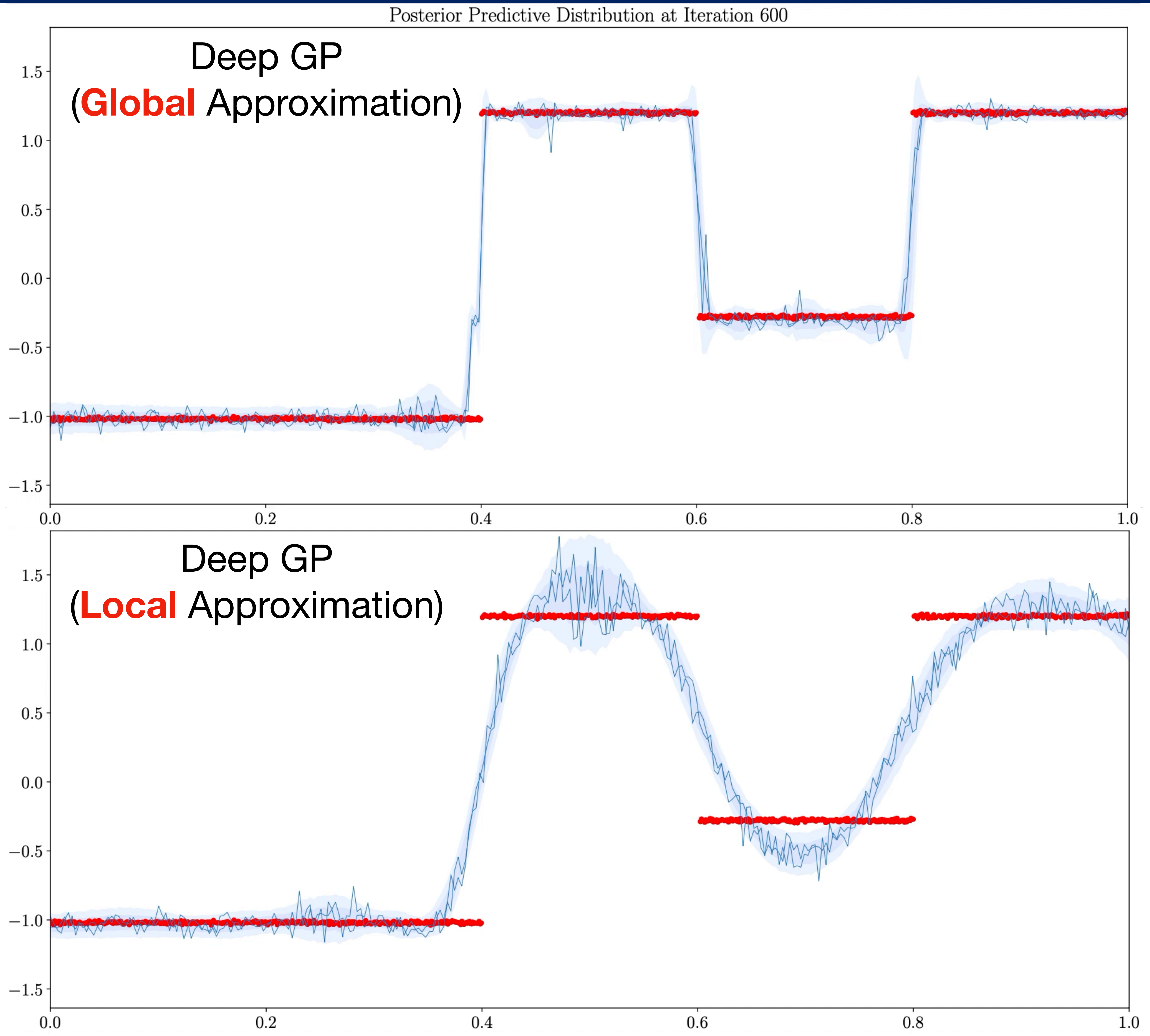


EXPERIMENT: MULTI-STEP FUNCTION

Convergence & Fit

Converges within 500 iterations

Never attains a good fit

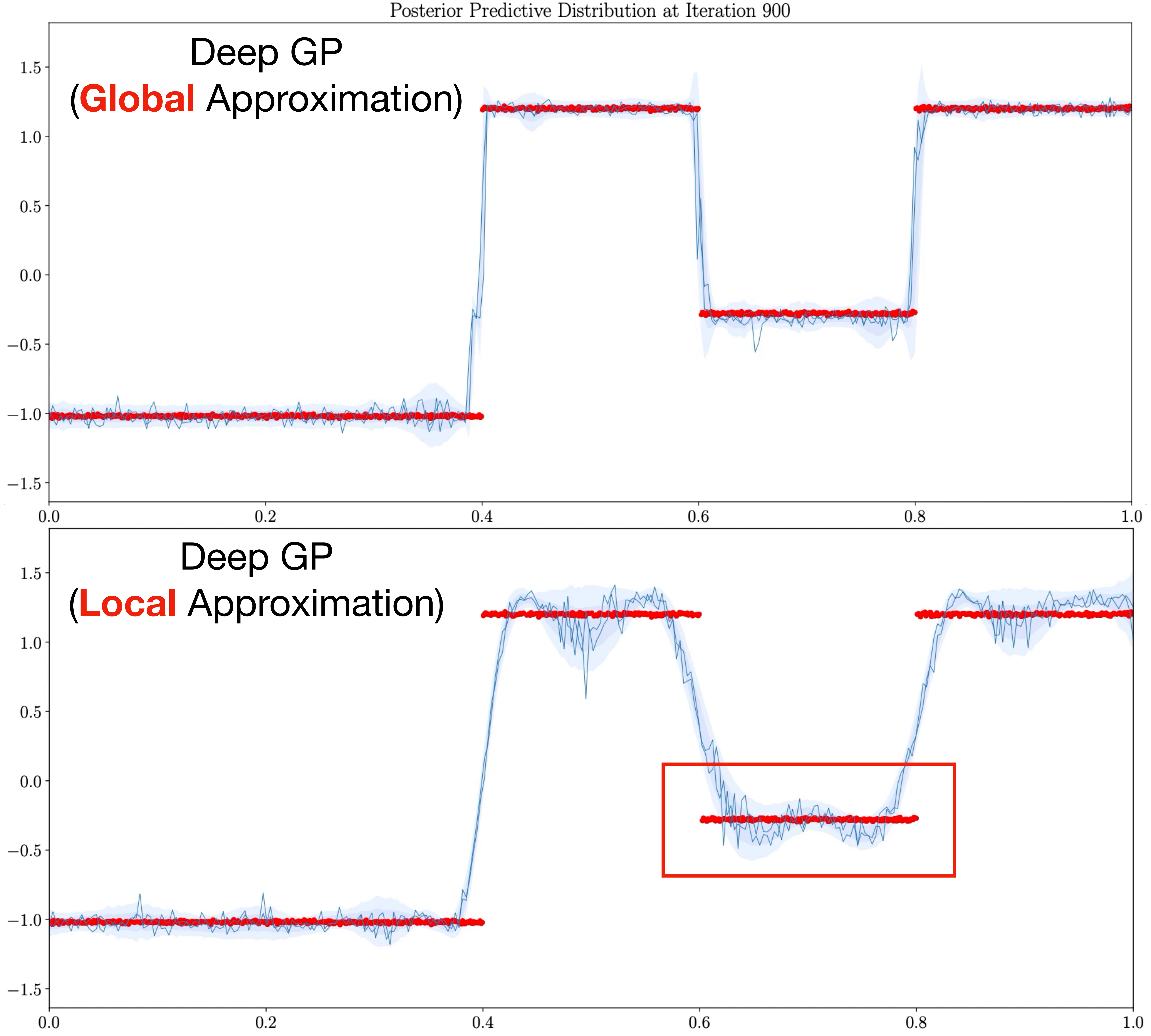


Experiment: Multi-step Function

Convergence & Fit

Converges within 500 iterations

Never attains a good fit

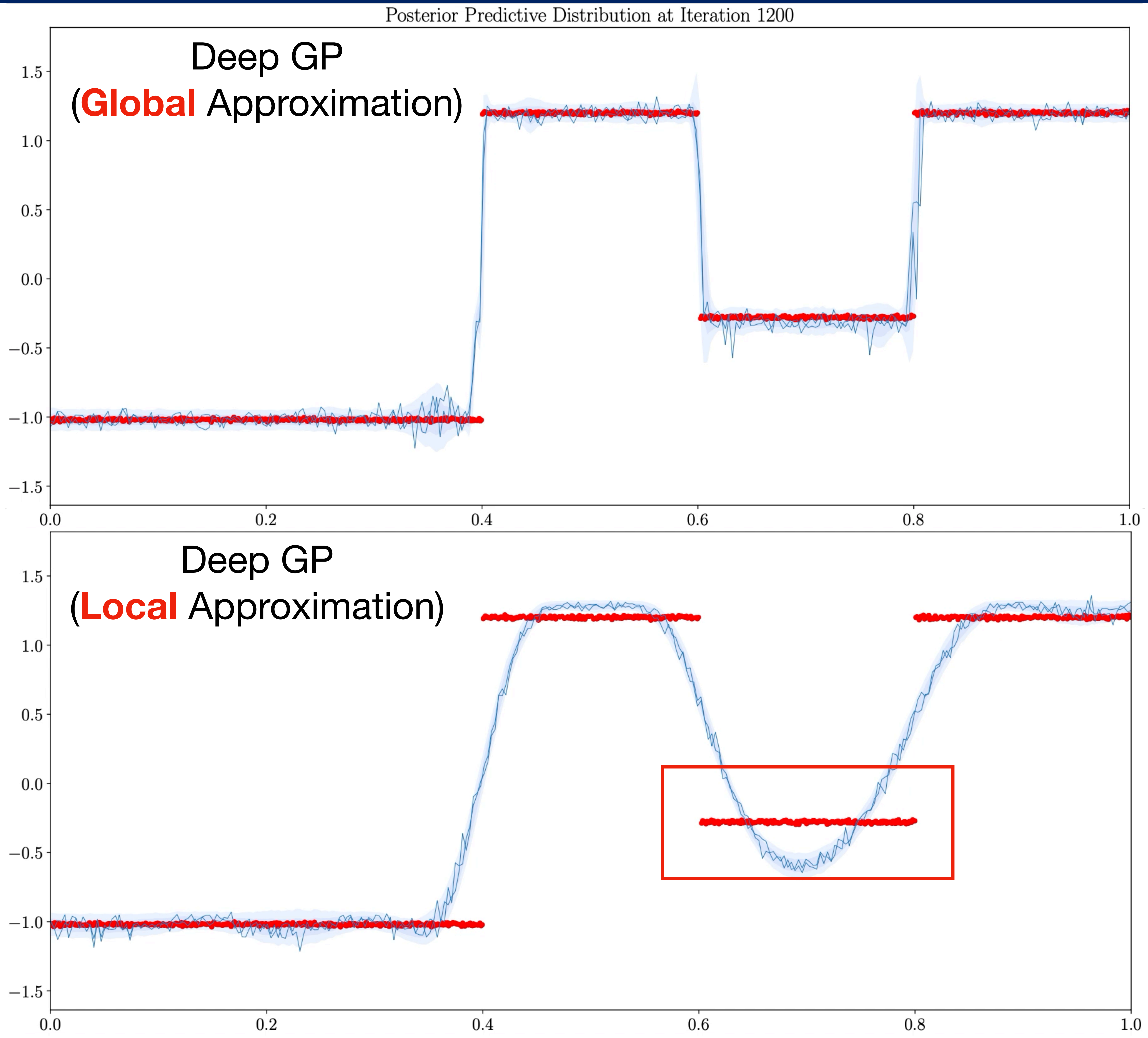


EXPERIMENT: MULTI-STEP FUNCTION

Convergence & Fit

Converges within 500 iterations

Never attains a good fit

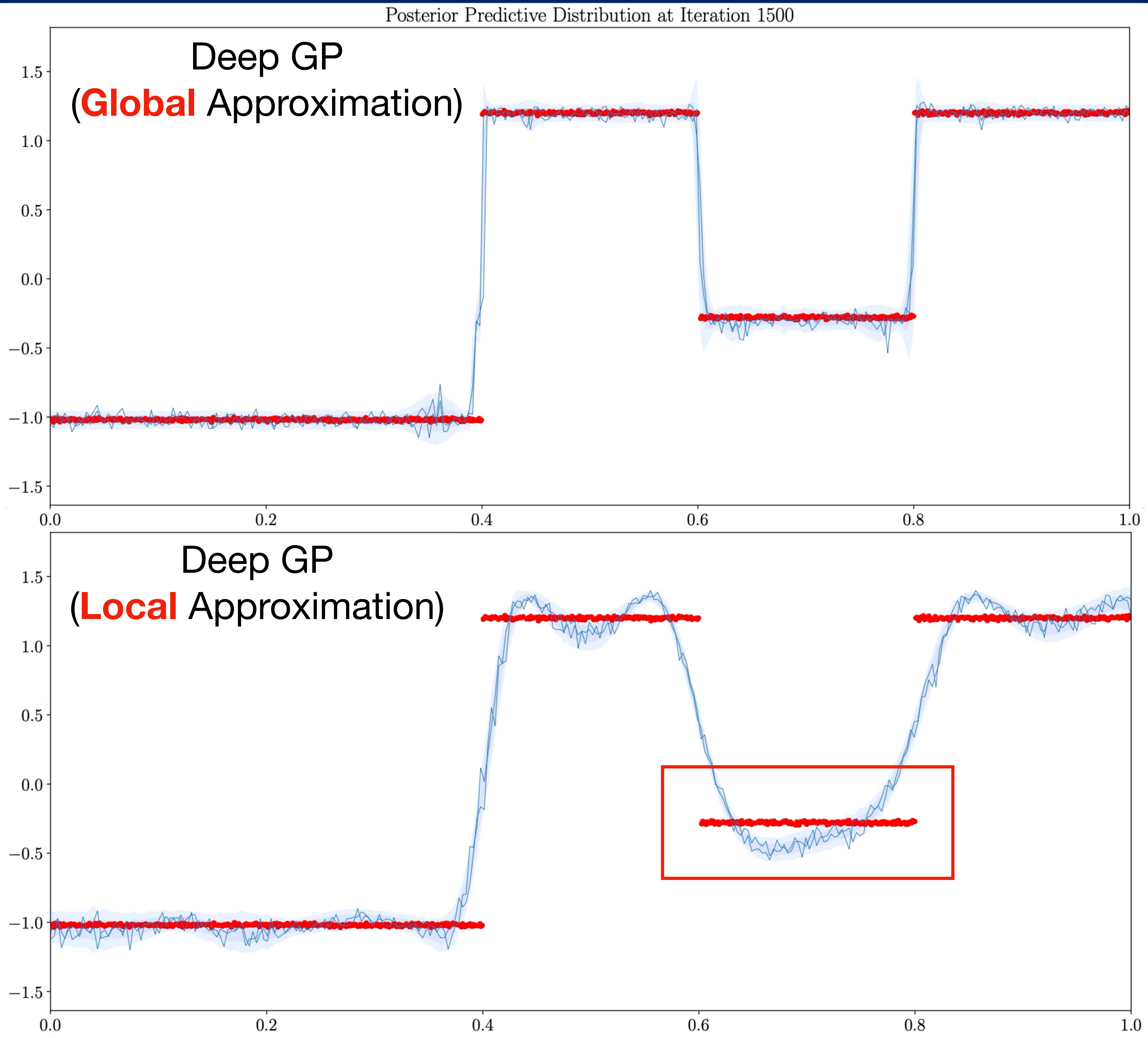


EXPERIMENT: MULTI-STEP FUNCTION

Convergence & Fit

Converges within 500 iterations

Never attains a good fit

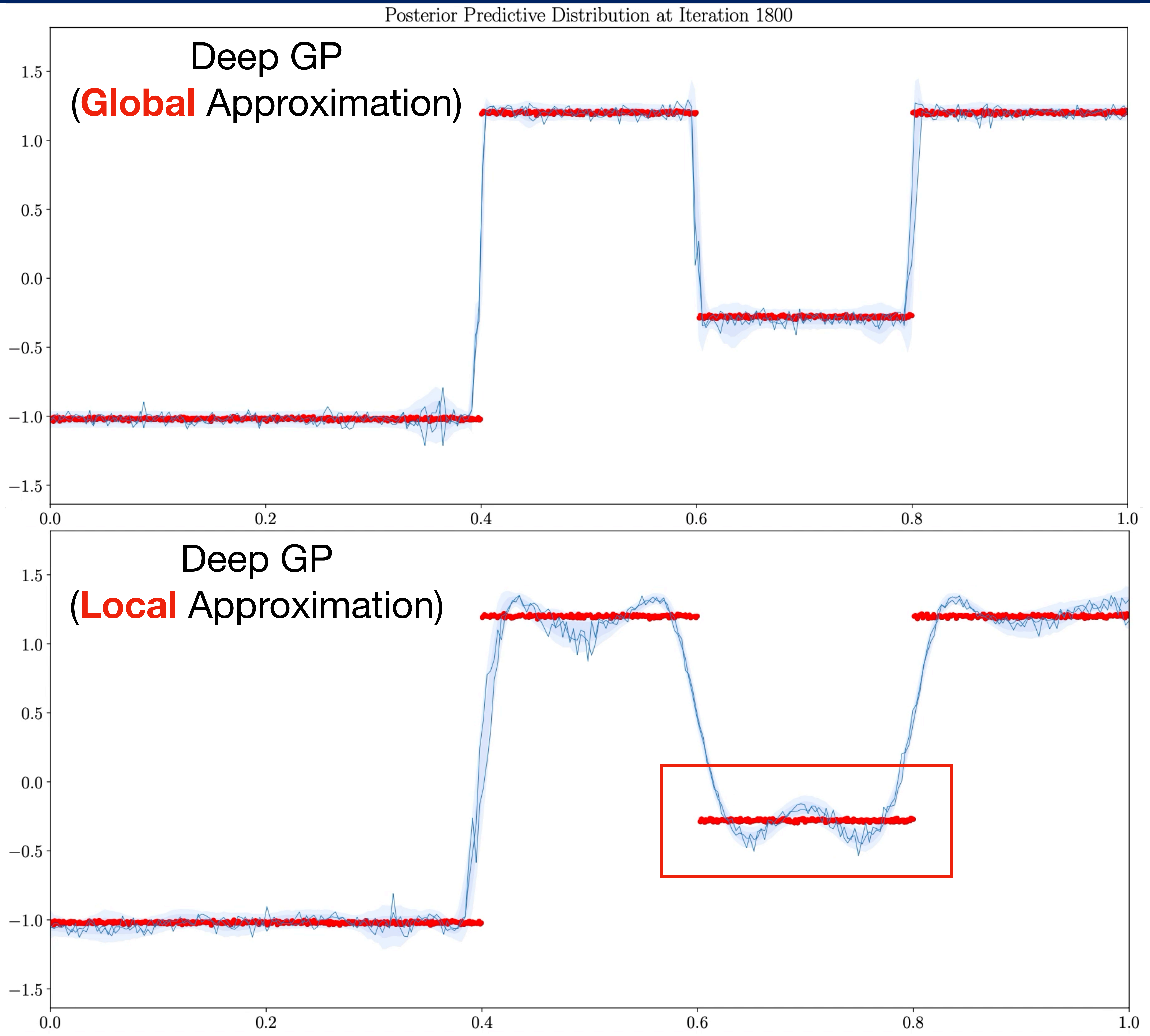


Experiment: Multi-step Function

Convergence & Fit

Converges within 500 iterations

Never attains a good fit



EXPERIMENT: MULTI-STEP FUNCTION

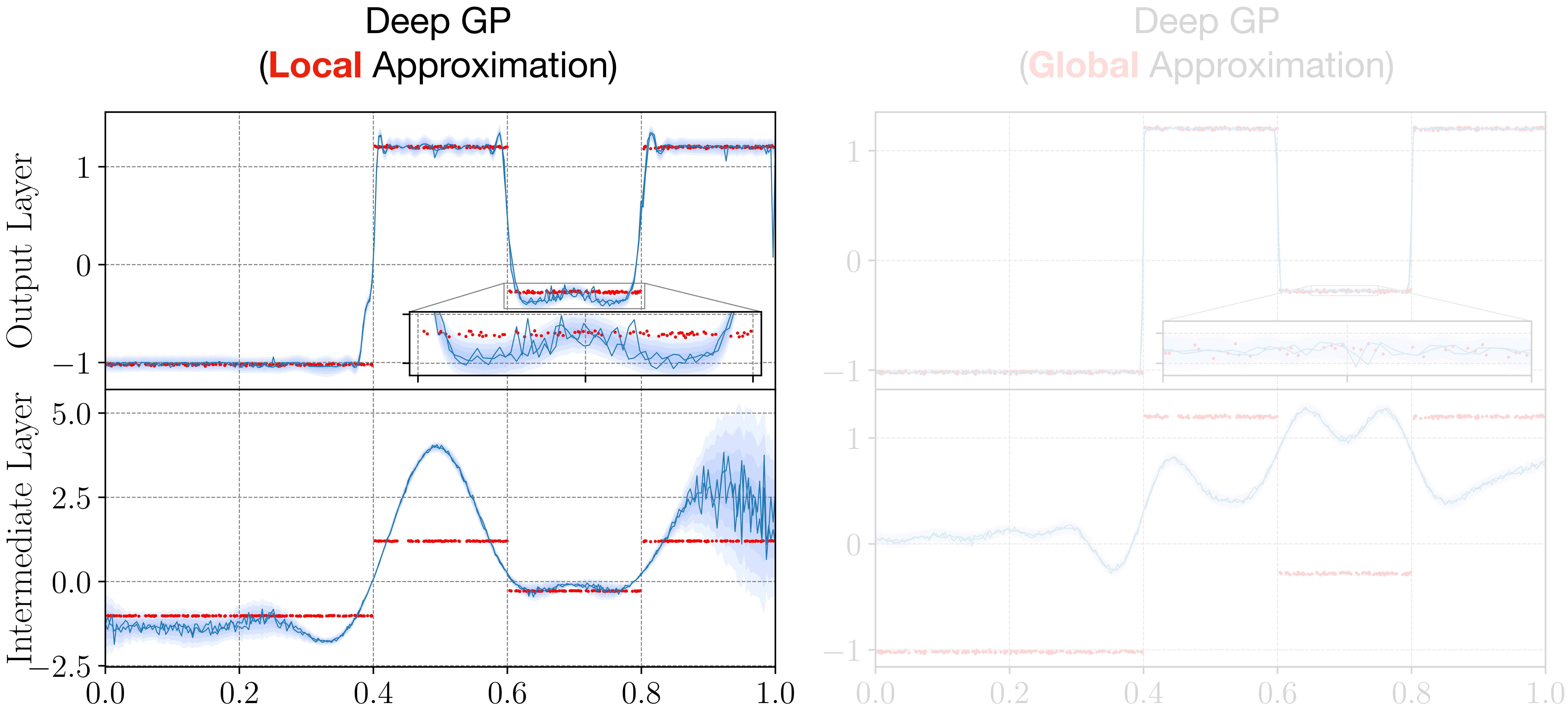


Figure 6. Comparison of posterior predictive distributions. The global approximation (right) captures the global structure, whereas the local approximation (left) is not.

Experiment: Multi-step Function

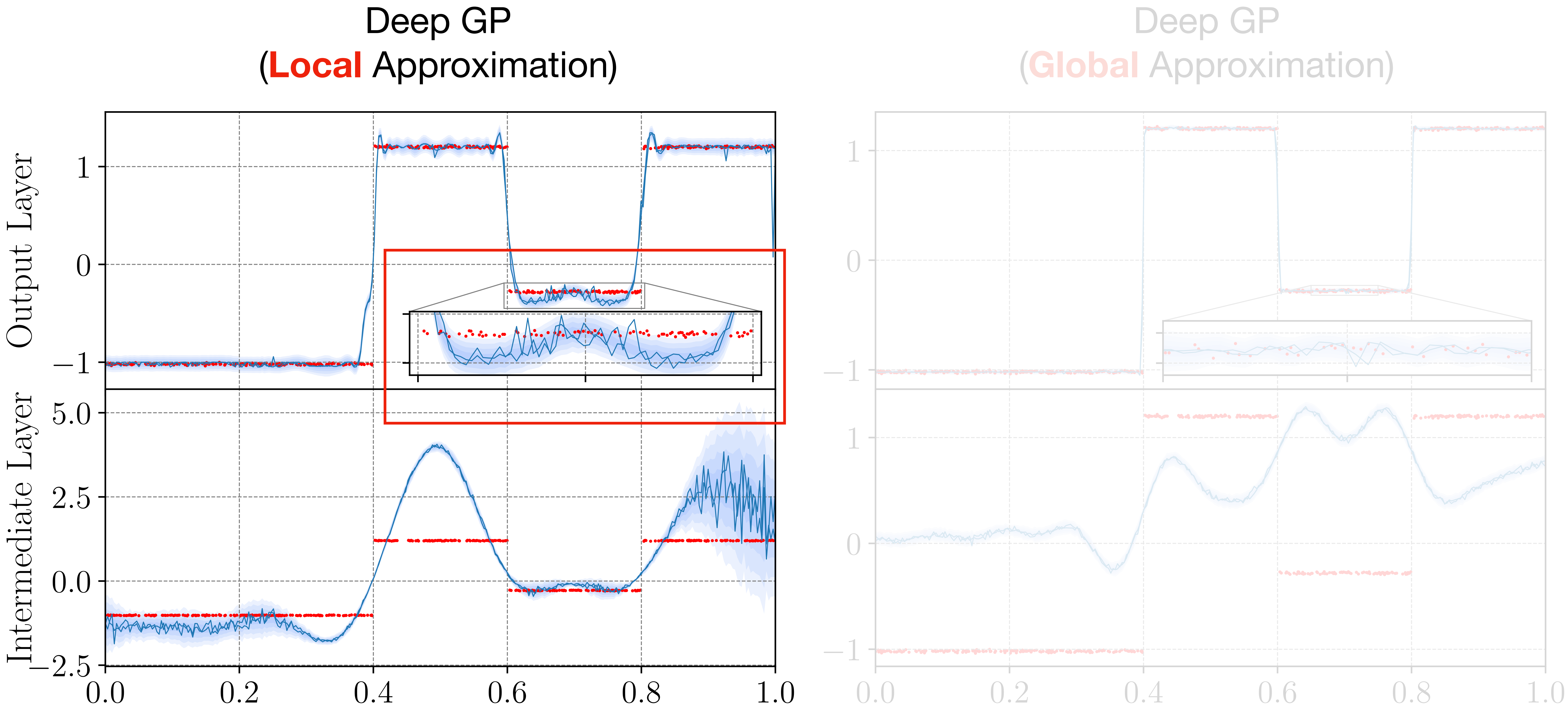


Figure 6. Comparison of posterior predictive distributions. The global approximation (right) captures the global structure, whereas the local approximation (left) is not.

Experiment: Multi-step Function

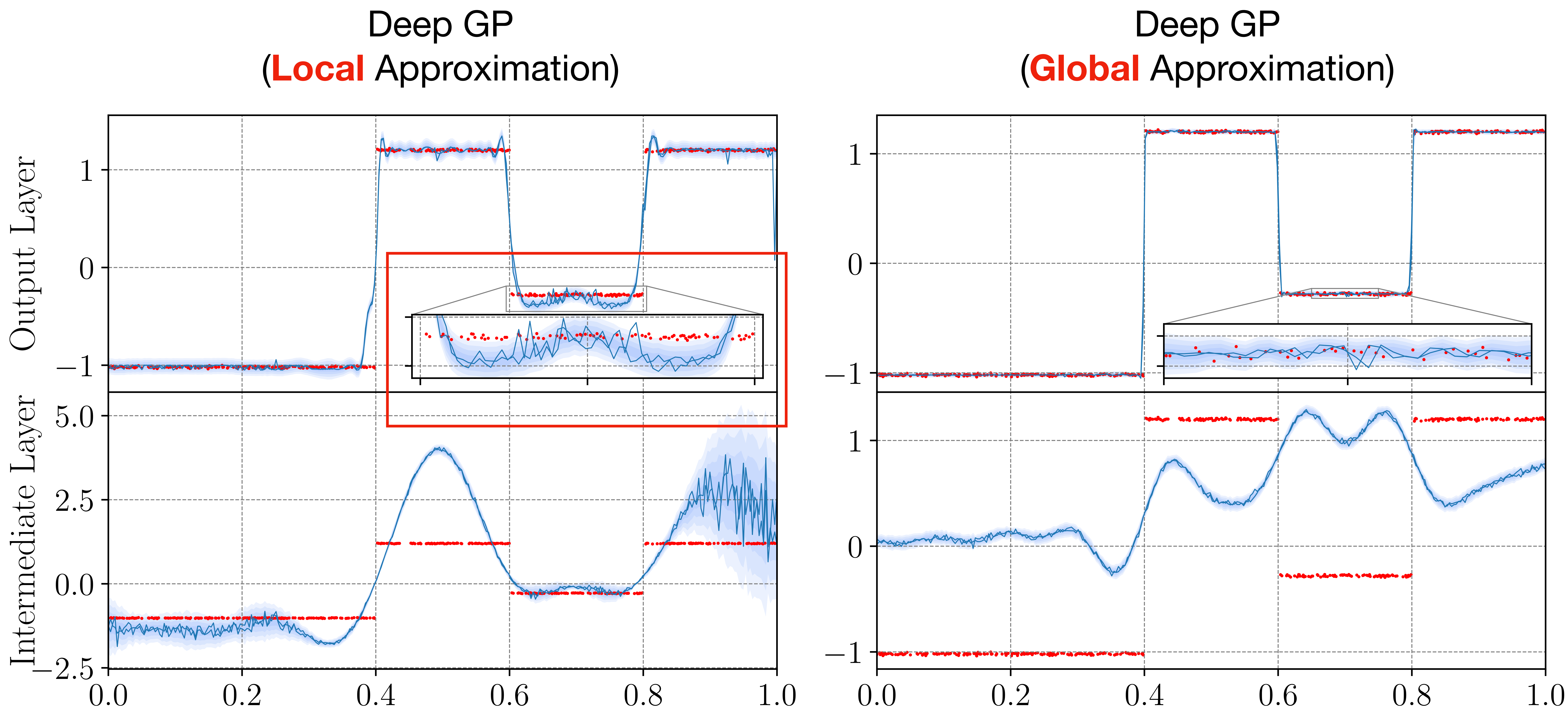


Figure 6. Comparison of posterior predictive distributions. The global approximation (right) captures the global structure, whereas the local approximation (left) is not.

Experiment: Multi-step Function

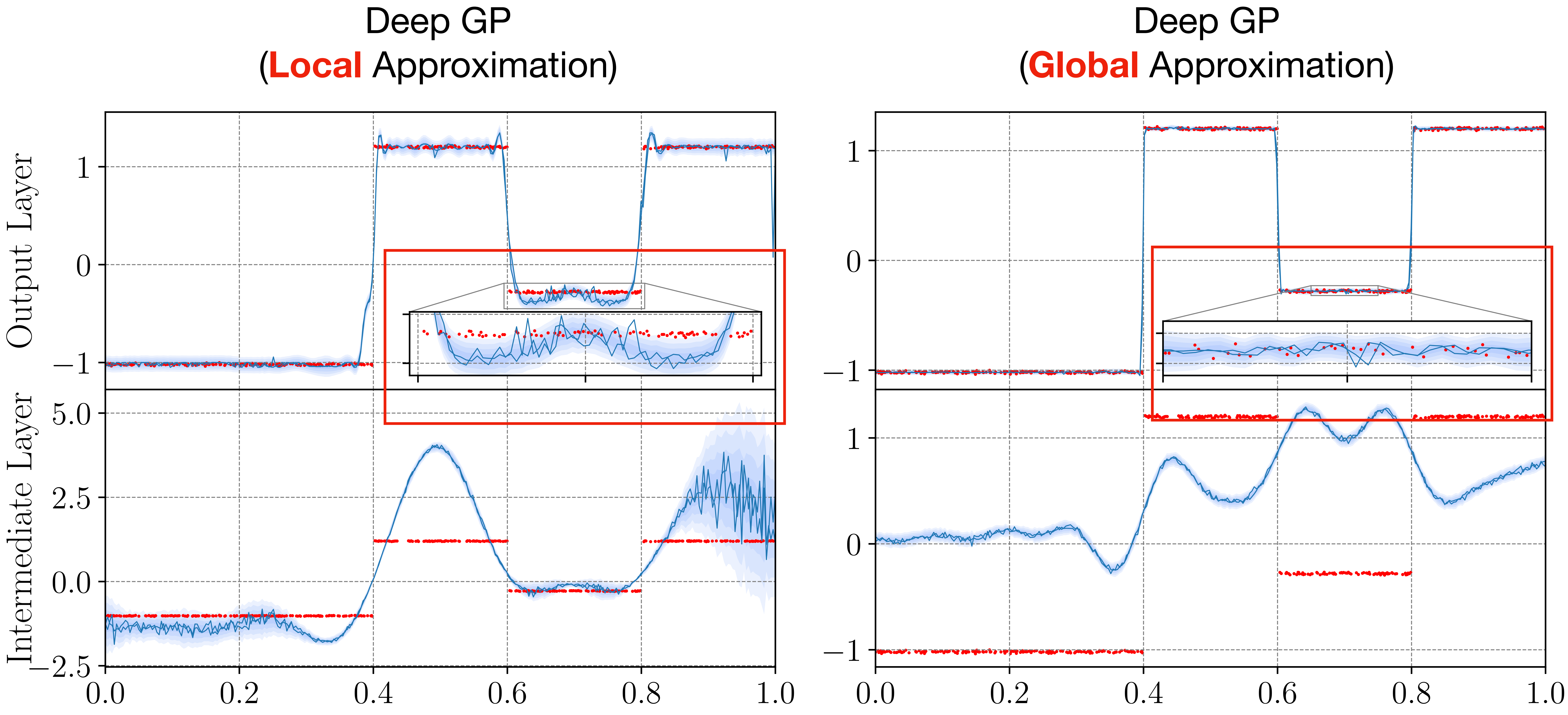


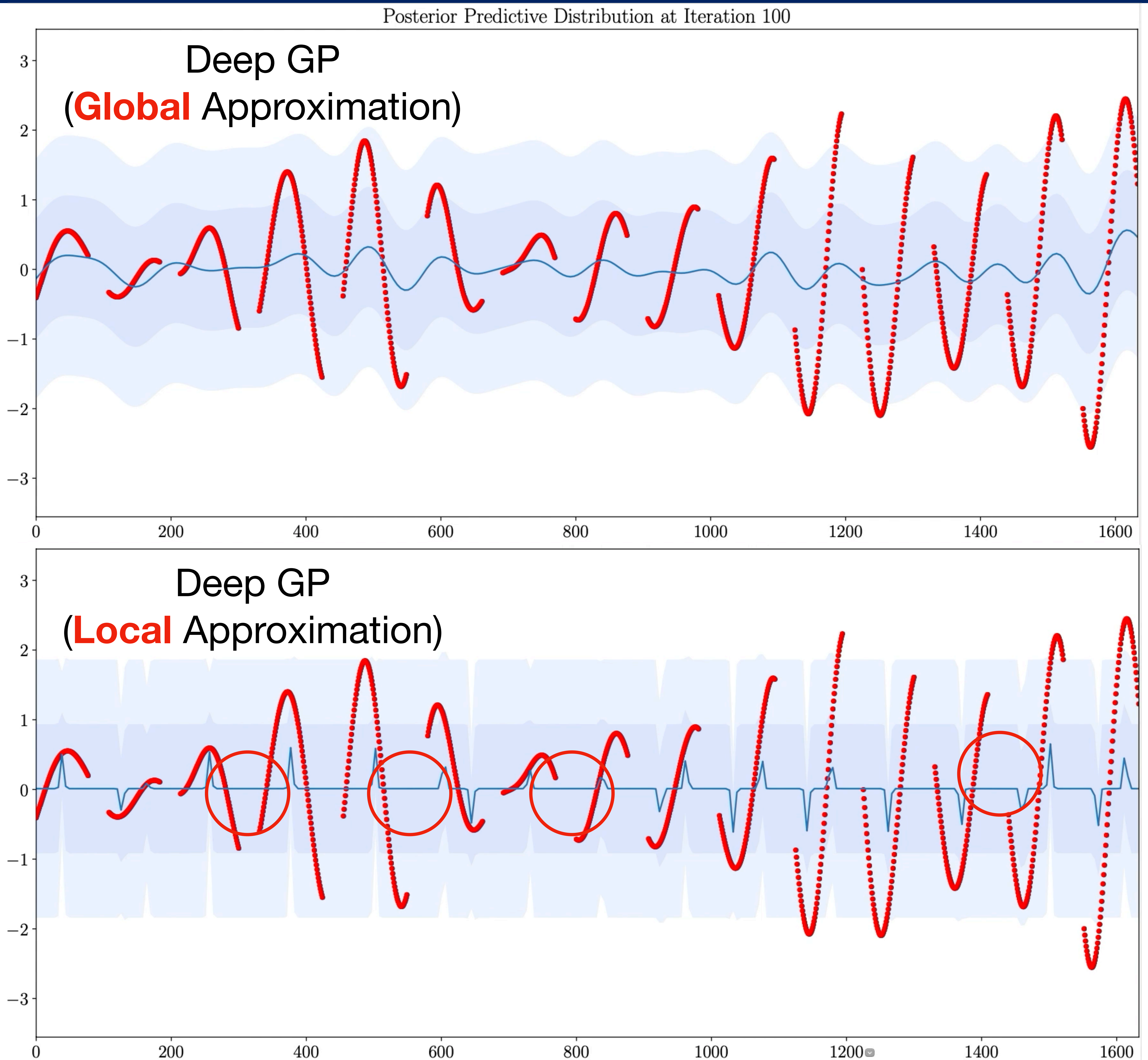
Figure 6. Comparison of posterior predictive distributions. The global approximation (right) captures the global structure, whereas the local approximation (left) is not.

EXPERIMENT: AUDIO SUB-BAND RECONSTRUCTION

Convergence & Fit

Converges within 3000 iterations

Never attains a good fit

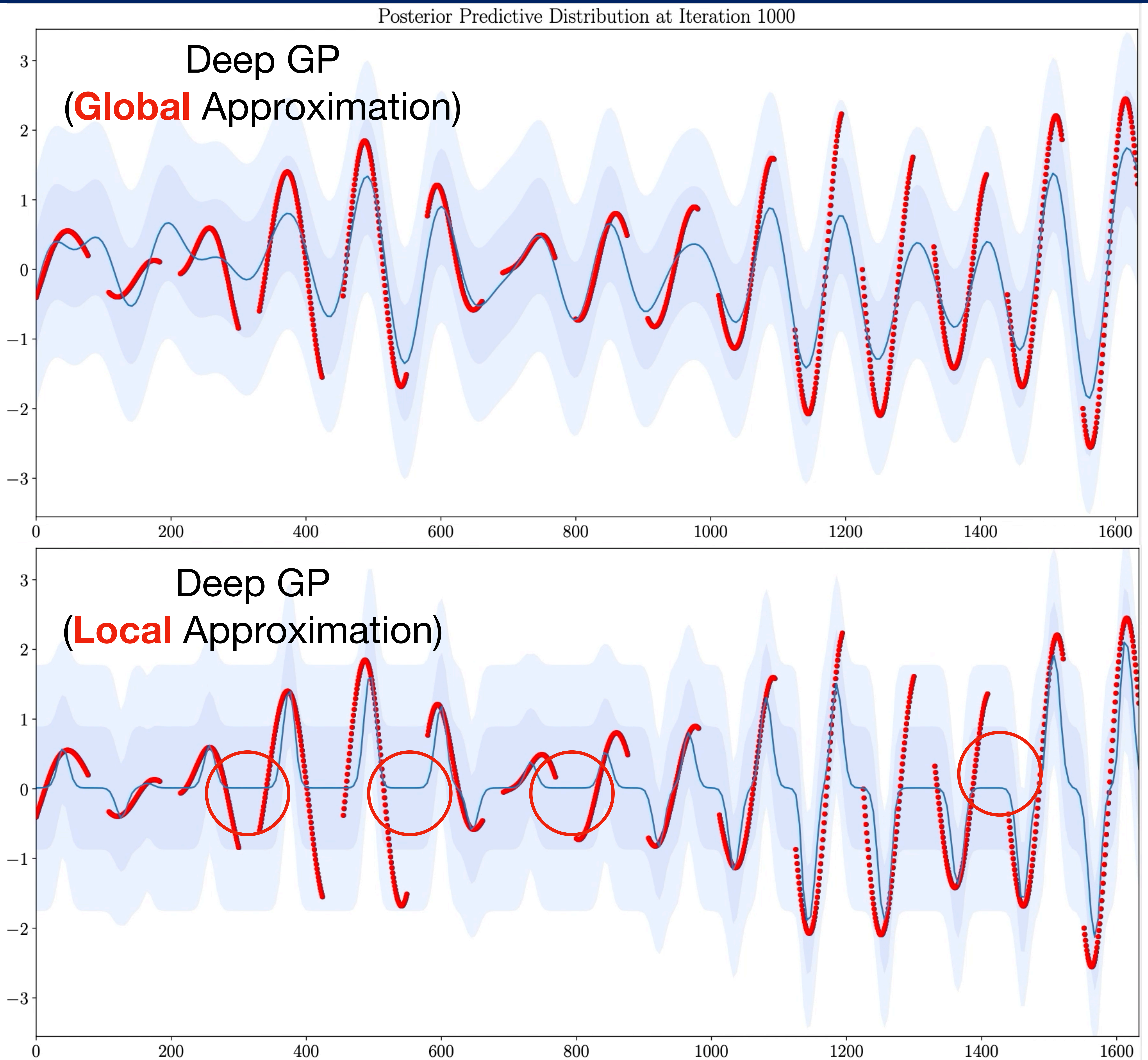


EXPERIMENT: AUDIO SUB-BAND RECONSTRUCTION

Convergence & Fit

Converges within 3000 iterations

Never attains a good fit

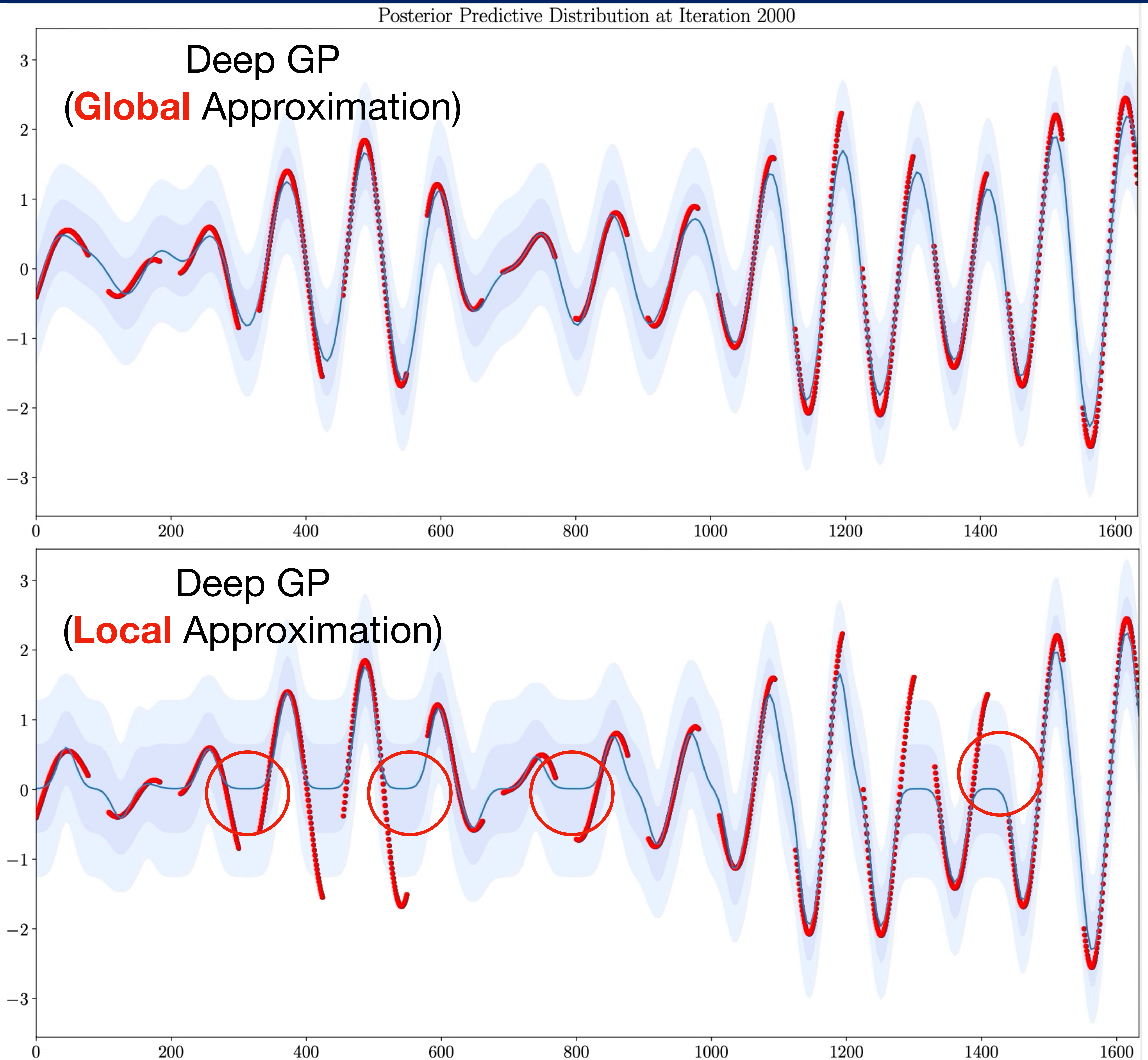


Experiment: Audio sub-band reconstruction

Convergence & Fit

Converges within 3000 iterations

Never attains a good fit

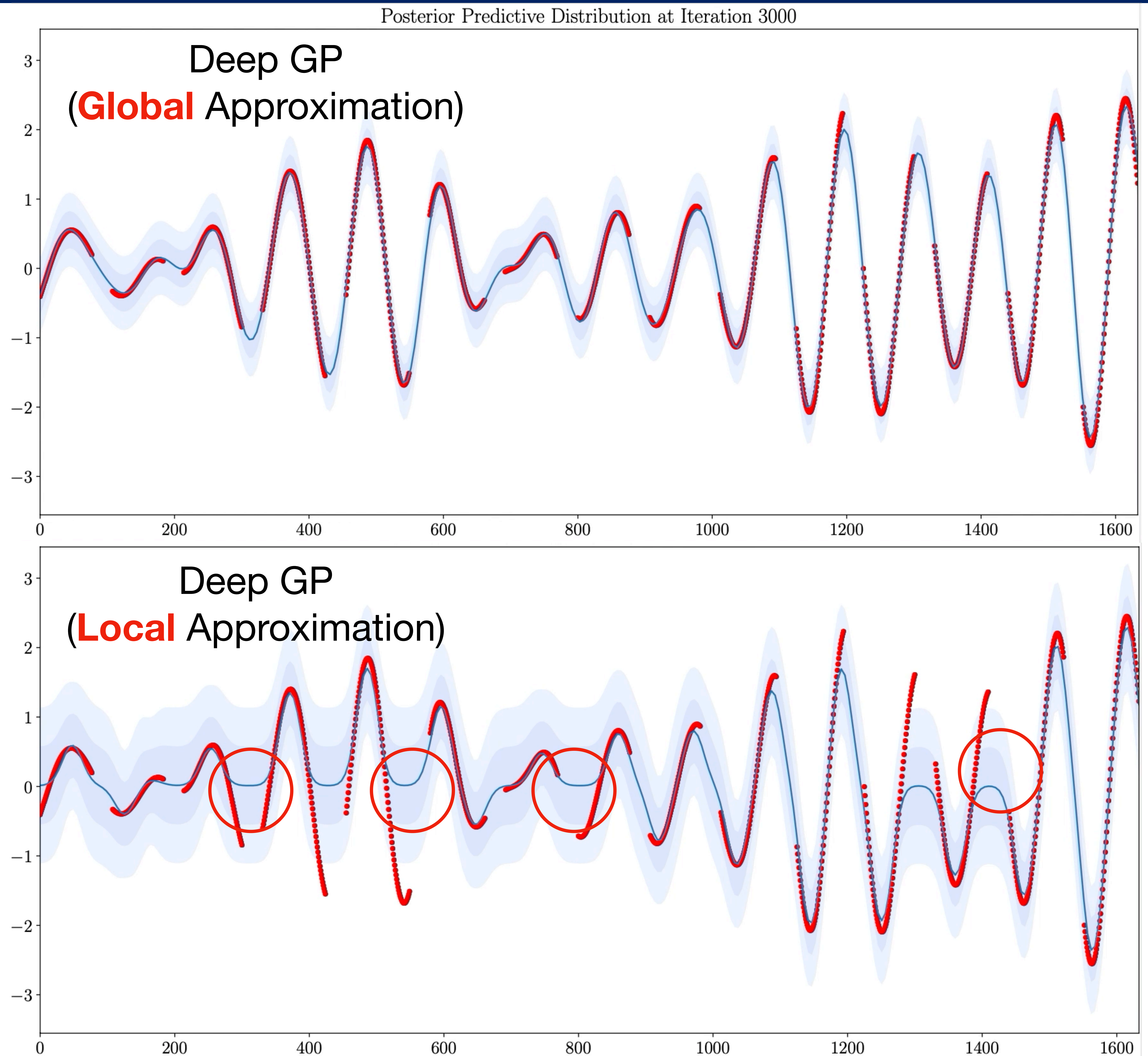


Experiment: Audio sub-band reconstruction

Convergence & Fit

Converges within 3000 iterations

Never attains a good fit



Experiment: Audio sub-band reconstruction

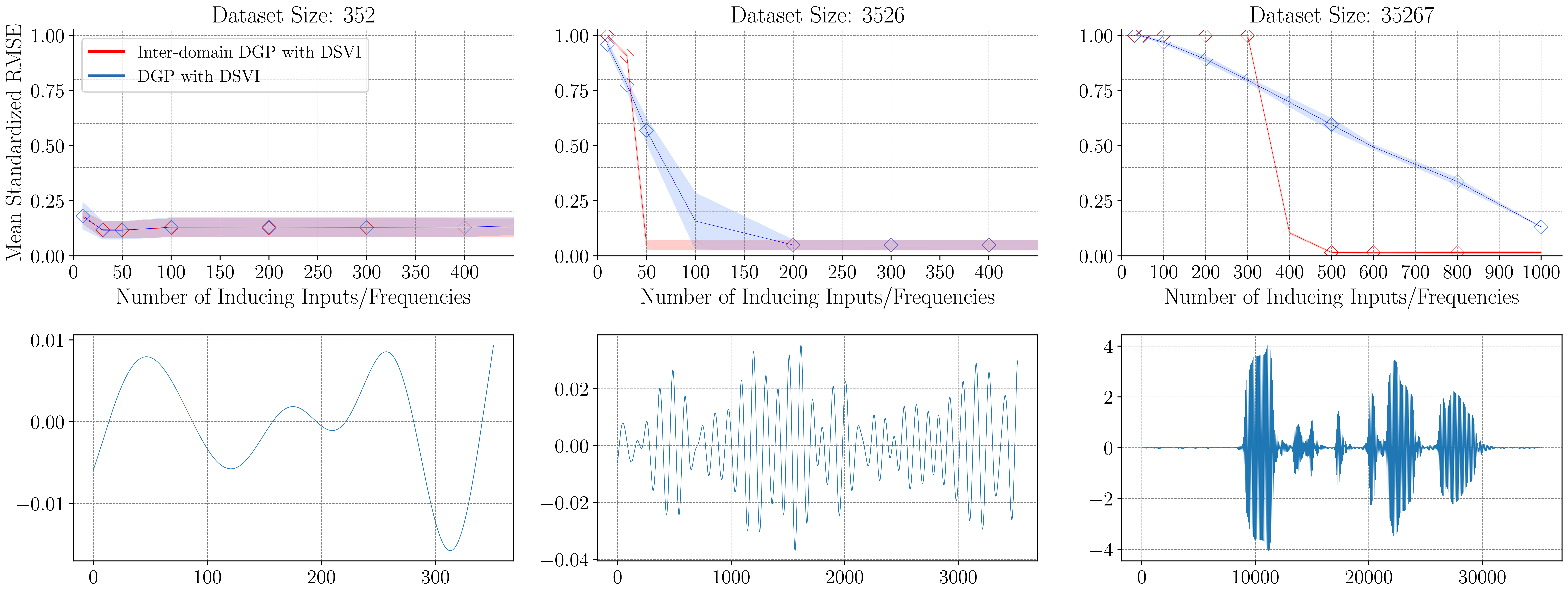


Figure 7. Comparison of mean standardized RMSEs on audio sub-band reconstruction tasks. The local approximation (**blue**) requires a much larger number of inducing inputs than the global approximation (**red**) to reconstruct the data well.

Experiment: Audio sub-band reconstruction

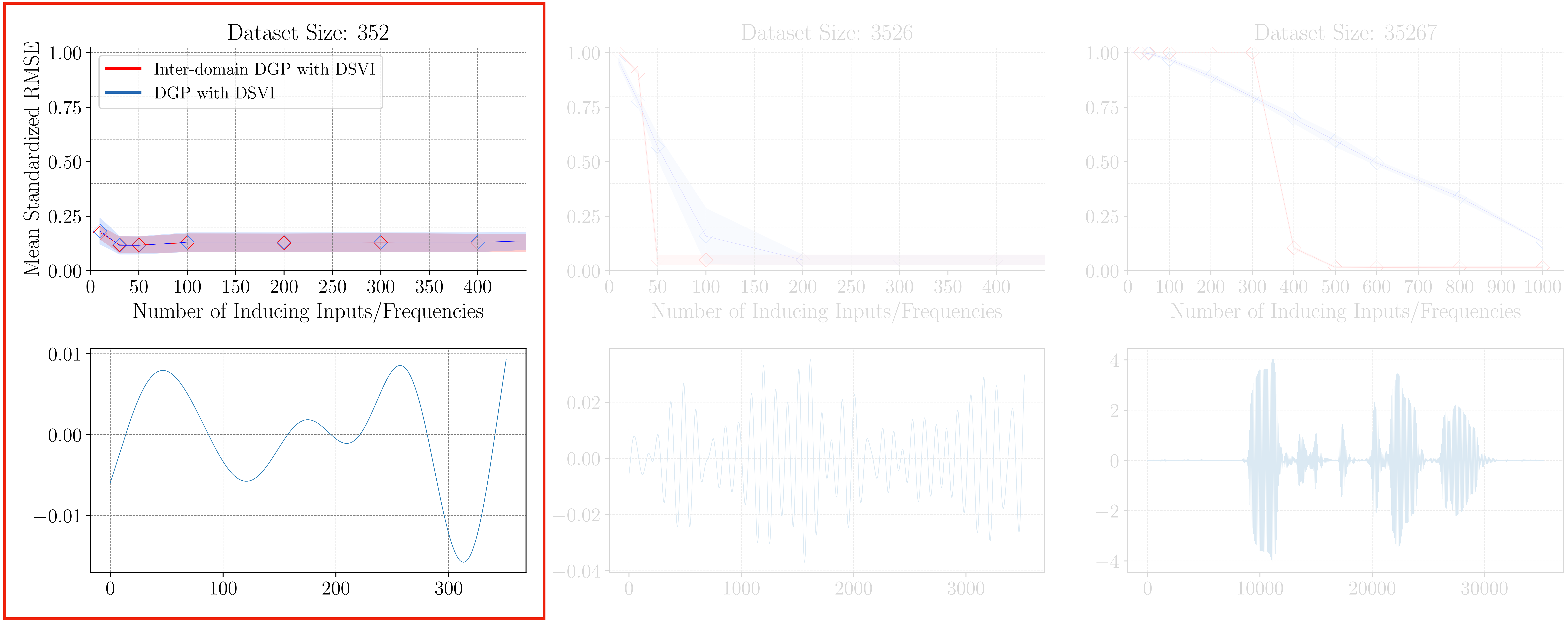


Figure 7. Comparison of mean standardized RMSEs on audio sub-band reconstruction tasks. The local approximation (**blue**) requires a much larger number of inducing inputs than the global approximation (**red**) to reconstruct the data well.

Experiment: Audio sub-band reconstruction

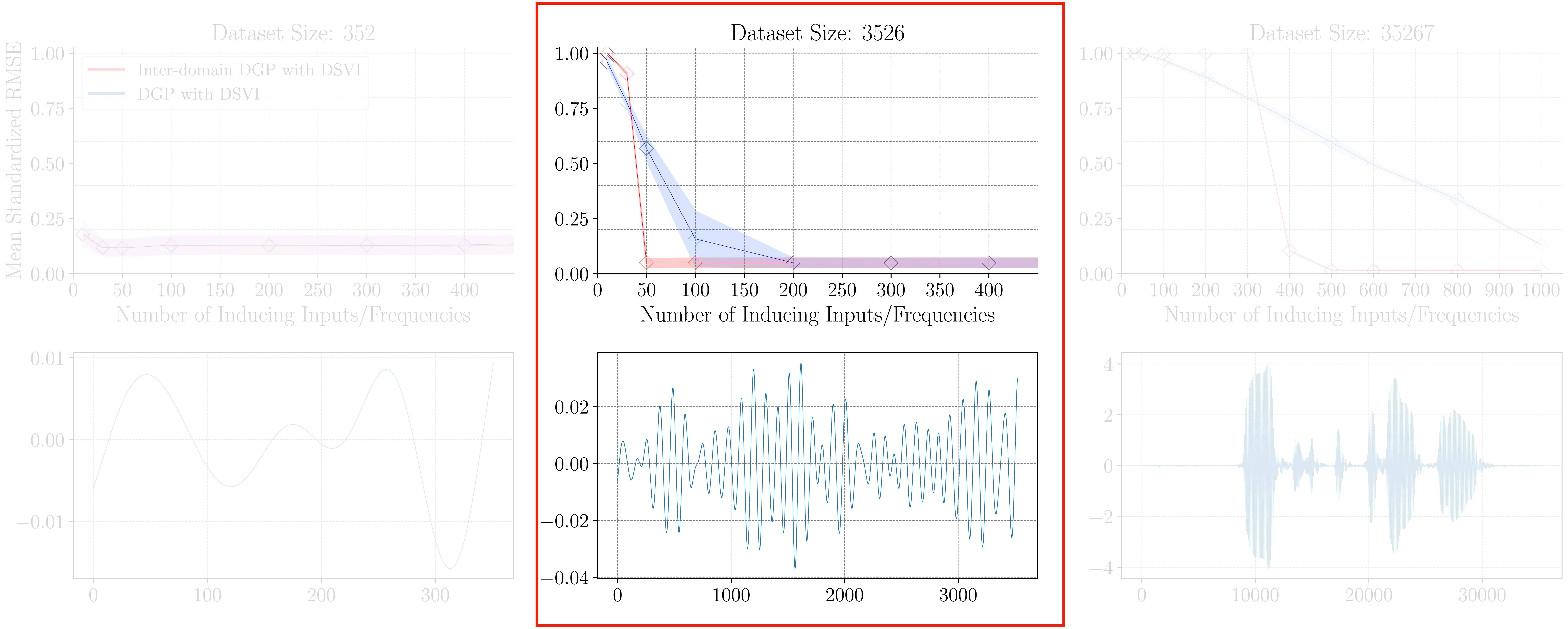


Figure 7. Comparison of mean standardized RMSEs on audio sub-band reconstruction tasks. The local approximation (**blue**) requires a much larger number of inducing inputs than the global approximation (**red**) to reconstruct the data well.

Experiment: Audio sub-band reconstruction

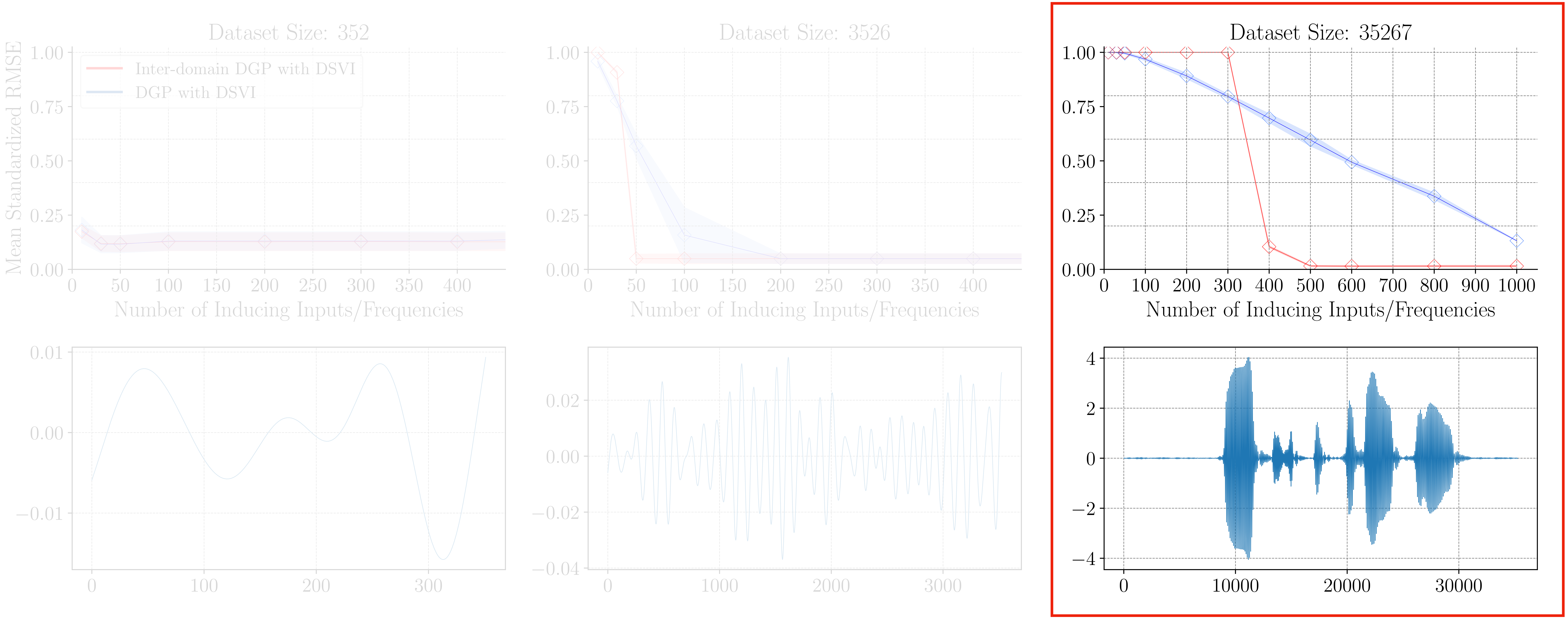


Figure 7. Comparison of mean standardized RMSEs on audio sub-band reconstruction tasks. The local approximation (**blue**) requires a much larger number of inducing inputs than the global approximation (**red**) to reconstruct the data well.

EXPERIMENT: AUDIO SUB-BAND RECONSTRUCTION

Global approximation yields **better predictive performance** at **same computational cost**
same predictive performance at **lower computational cost**

Figure 7. Comparison of mean standardized RMSEs on audio sub-band reconstruction tasks. The local approximation (**blue**) requires a much larger number of inducing inputs than the global approximation (**red**) to reconstruct the data well.

Experiment: Real-World Data with Global Structure

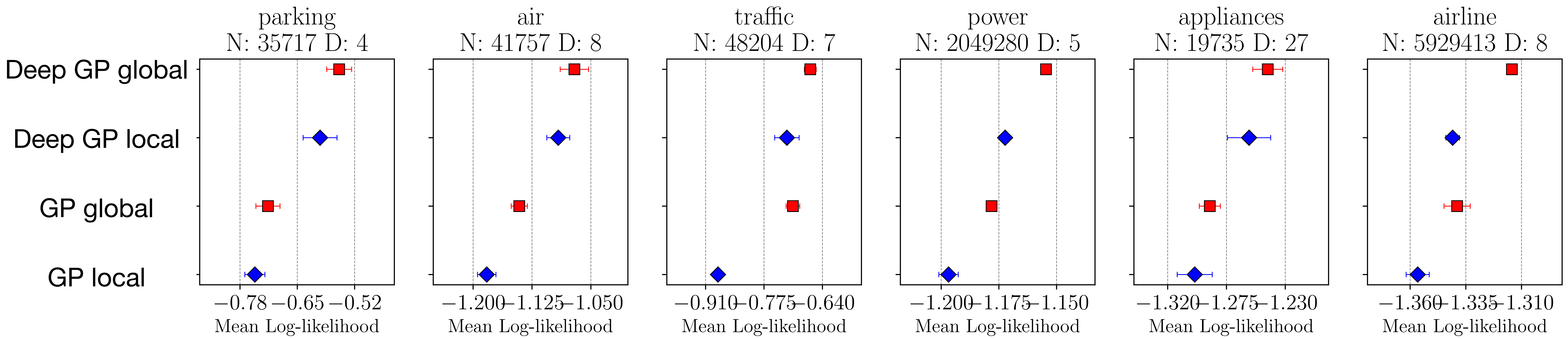


Figure 8. Comparison of mean test log-likelihoods on real-world prediction tasks (higher is better). All datasets exhibit global structure. Global approximations (**red**) outperform local approximations (**blue**) throughout.

EXPERIMENT: REAL-WORLD DATA WITH GLOBAL STRUCTURE

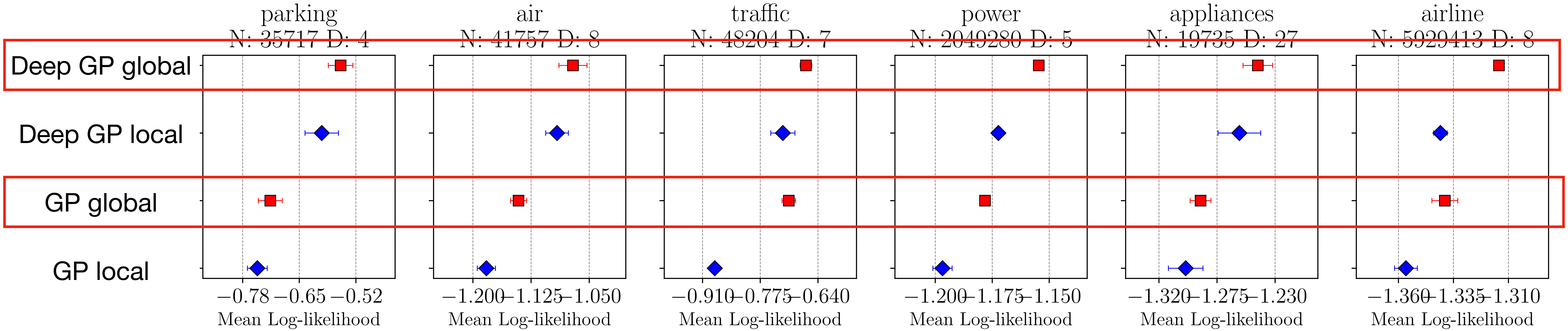


Figure 8. Comparison of mean test log-likelihoods on real-world prediction tasks (higher is better). All datasets exhibit global structure. Global approximations (**red**) outperform local approximations (**blue**) throughout.

Experiment: Real-World Data with Global Structure

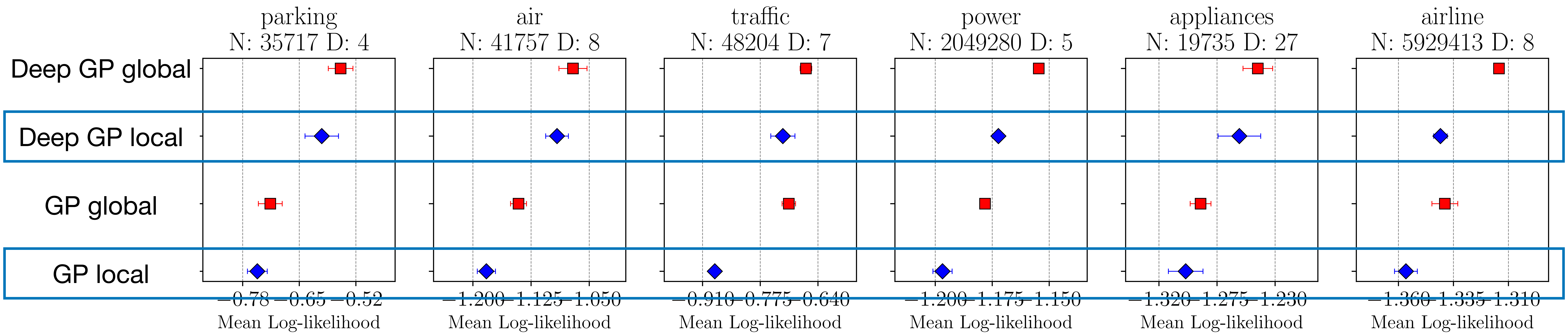


Figure 8. Comparison of mean test log-likelihoods on real-world prediction tasks (higher is better). All datasets exhibit global structure. Global approximations (**red**) outperform local approximations (**blue**) throughout.

Experiment: Real-World Data with Global Structure

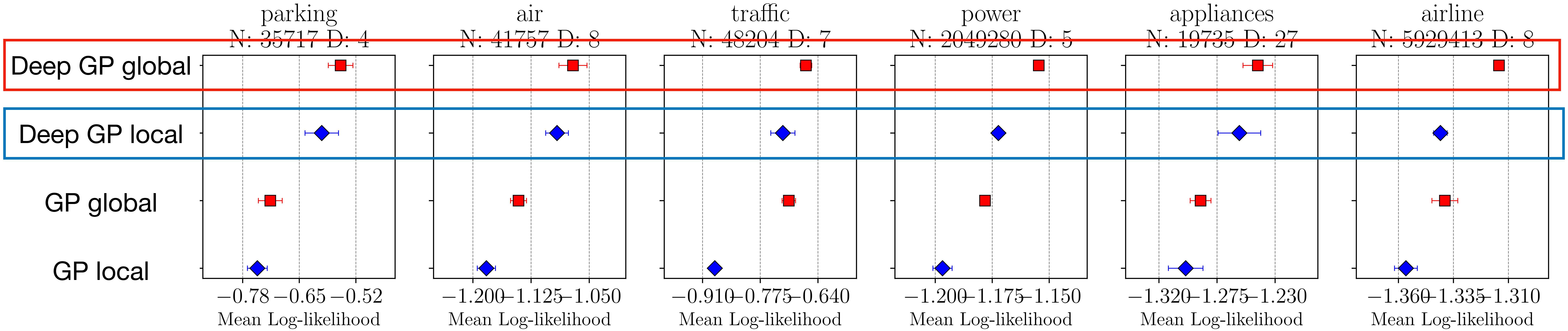
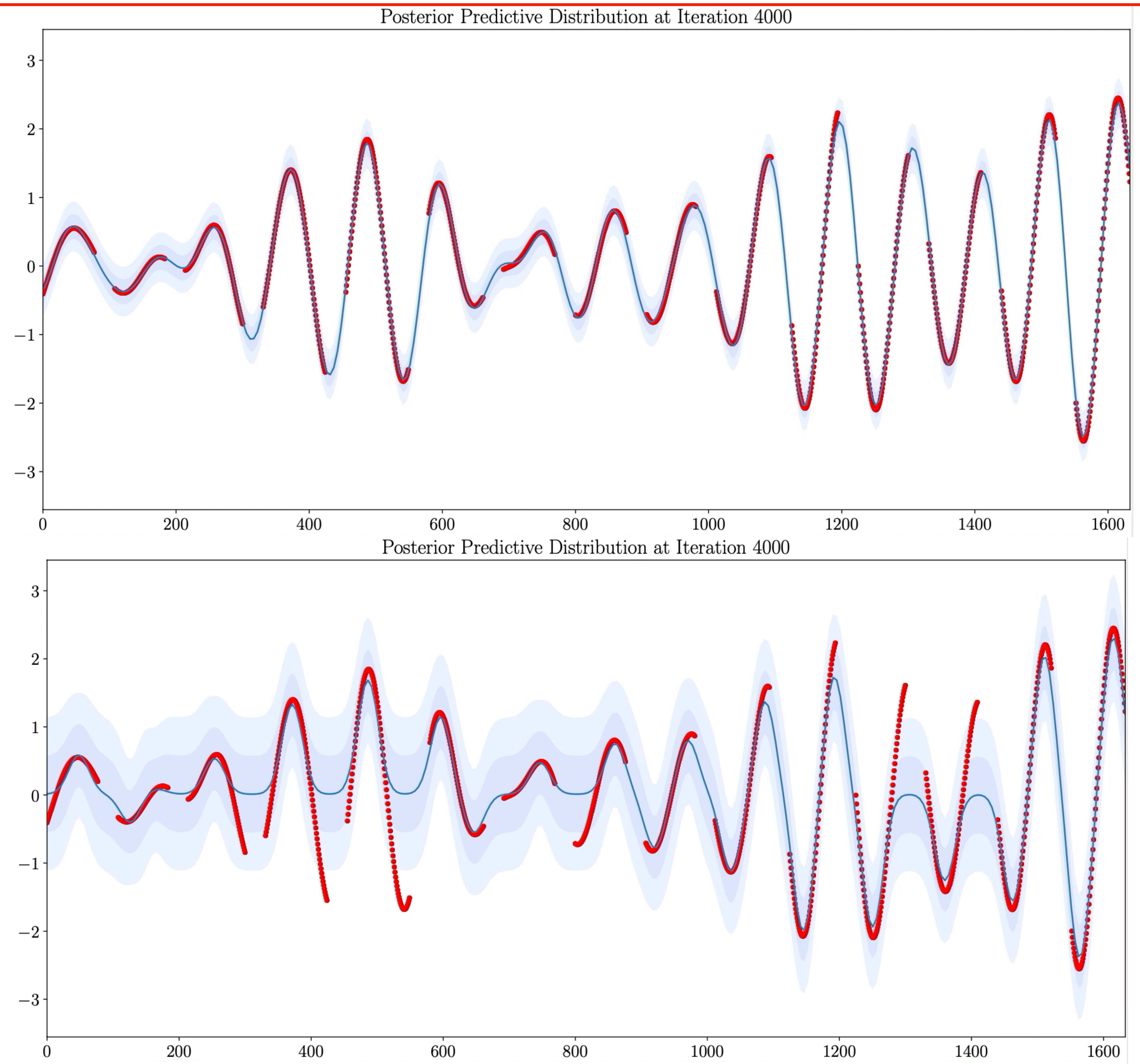


Figure 8. Comparison of mean test log-likelihoods on real-world prediction tasks (higher is better). All datasets exhibit global structure. Global approximations (**red**) outperform local approximations (**blue**) throughout.

Summary

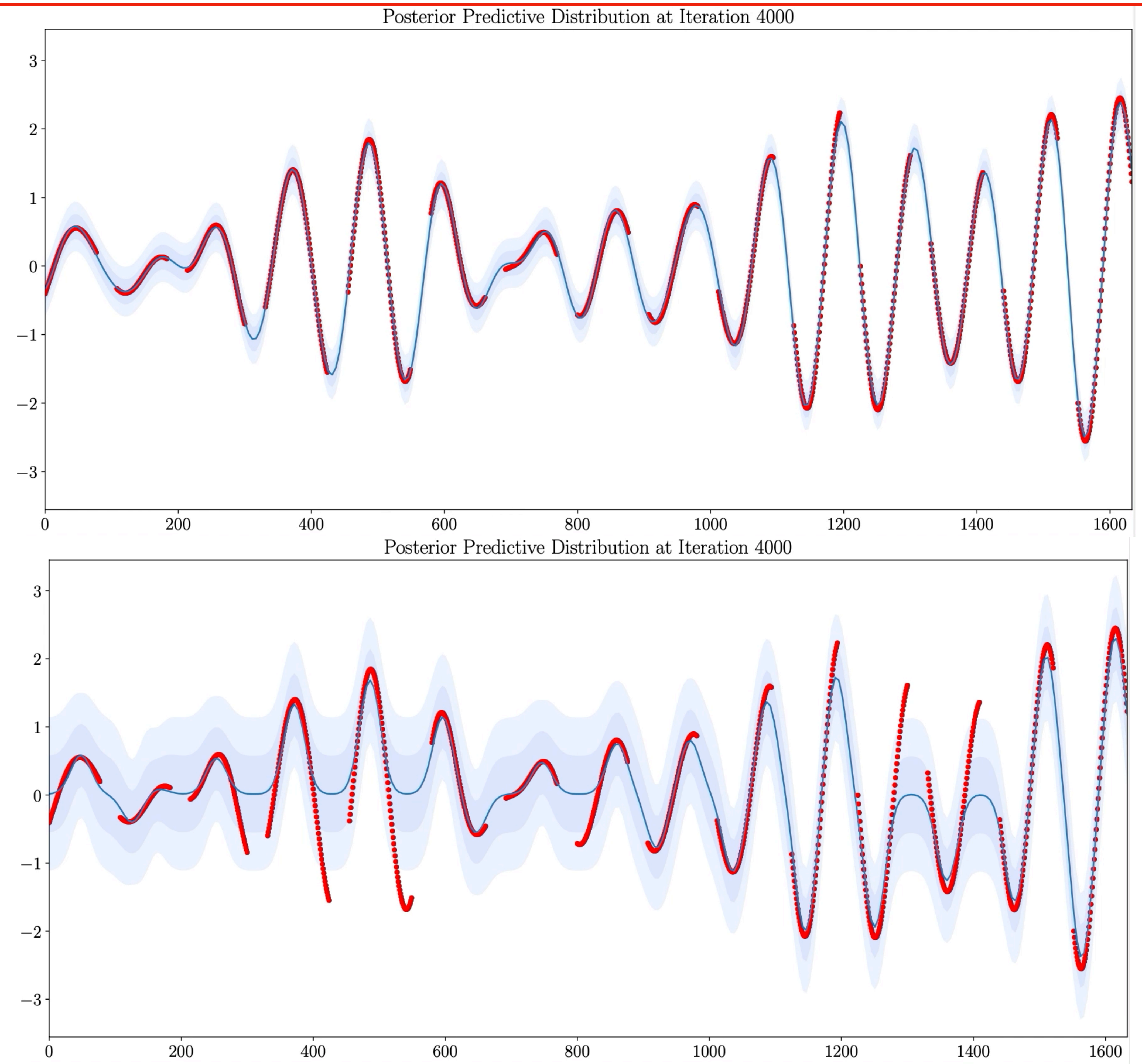
Benefits of Inter-domain Deep GPs

- Exploit global structure in the data
- Better predictive performance
- Higher computational efficiency
- Simple drop-in replacement



Benefits of Inter-domain Deep GPs

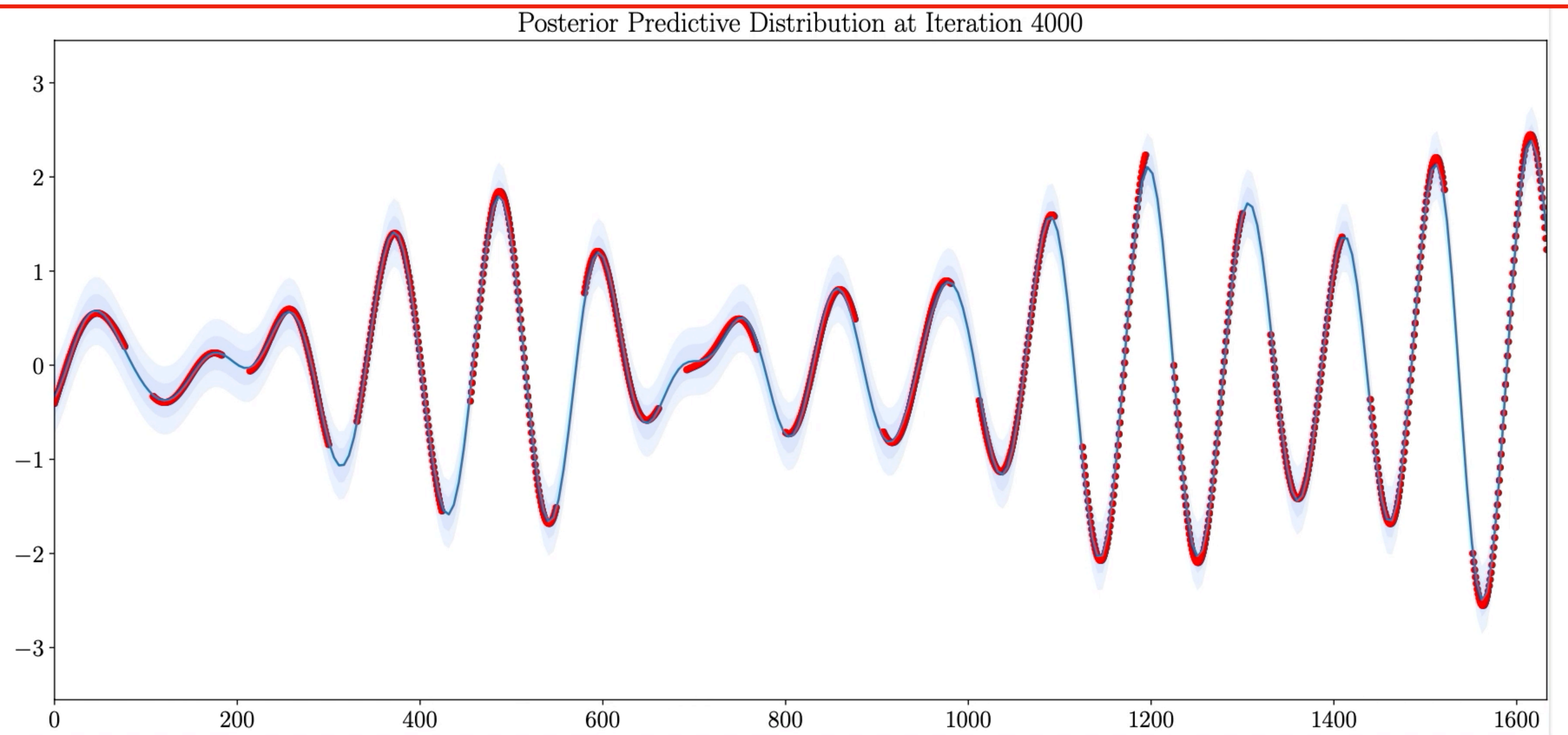
- Exploit **global structure in the data**
- Better predictive performance
- Higher computational efficiency
- Simple drop-in replacement

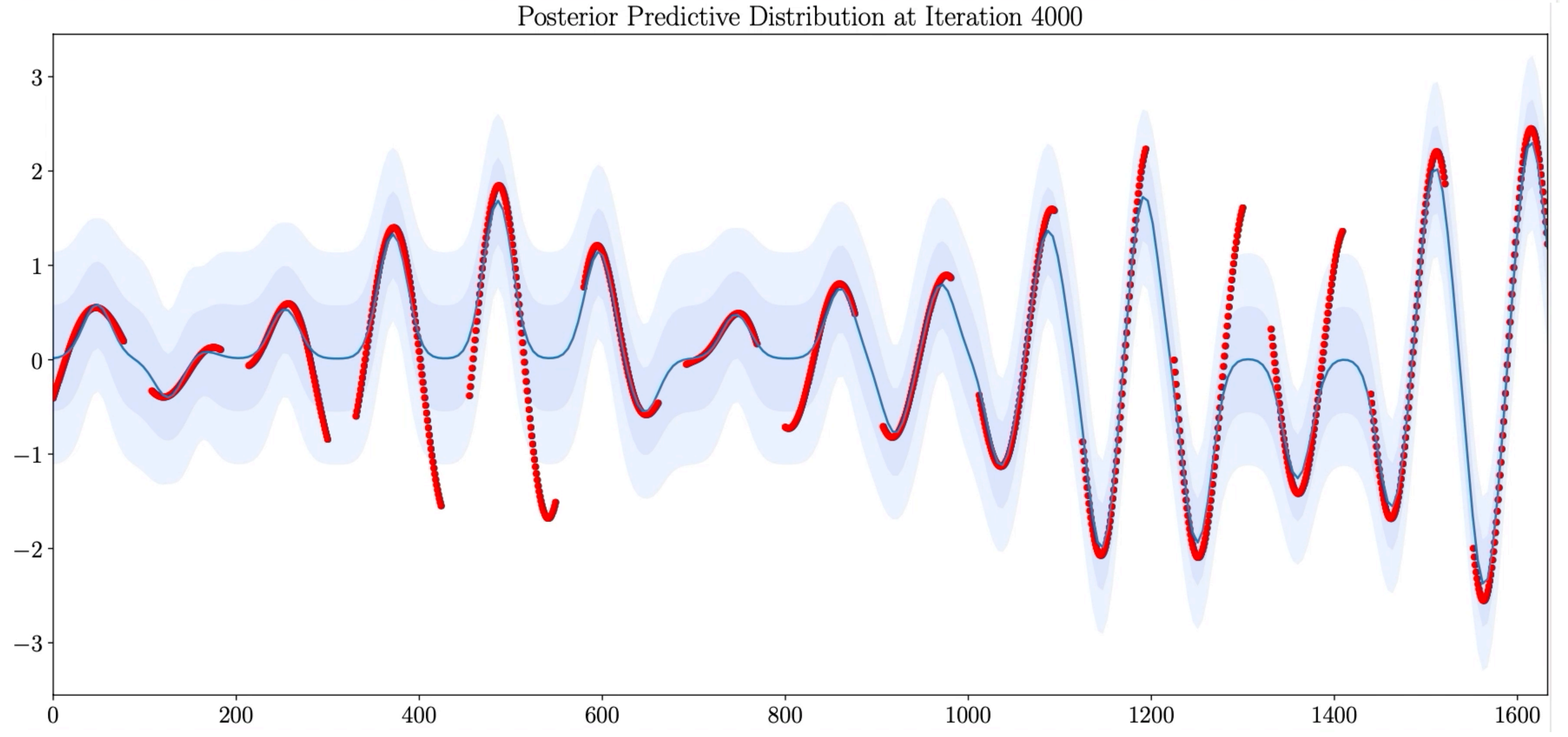


SUMMARY

Benefits of Inter-domain Deep GPs

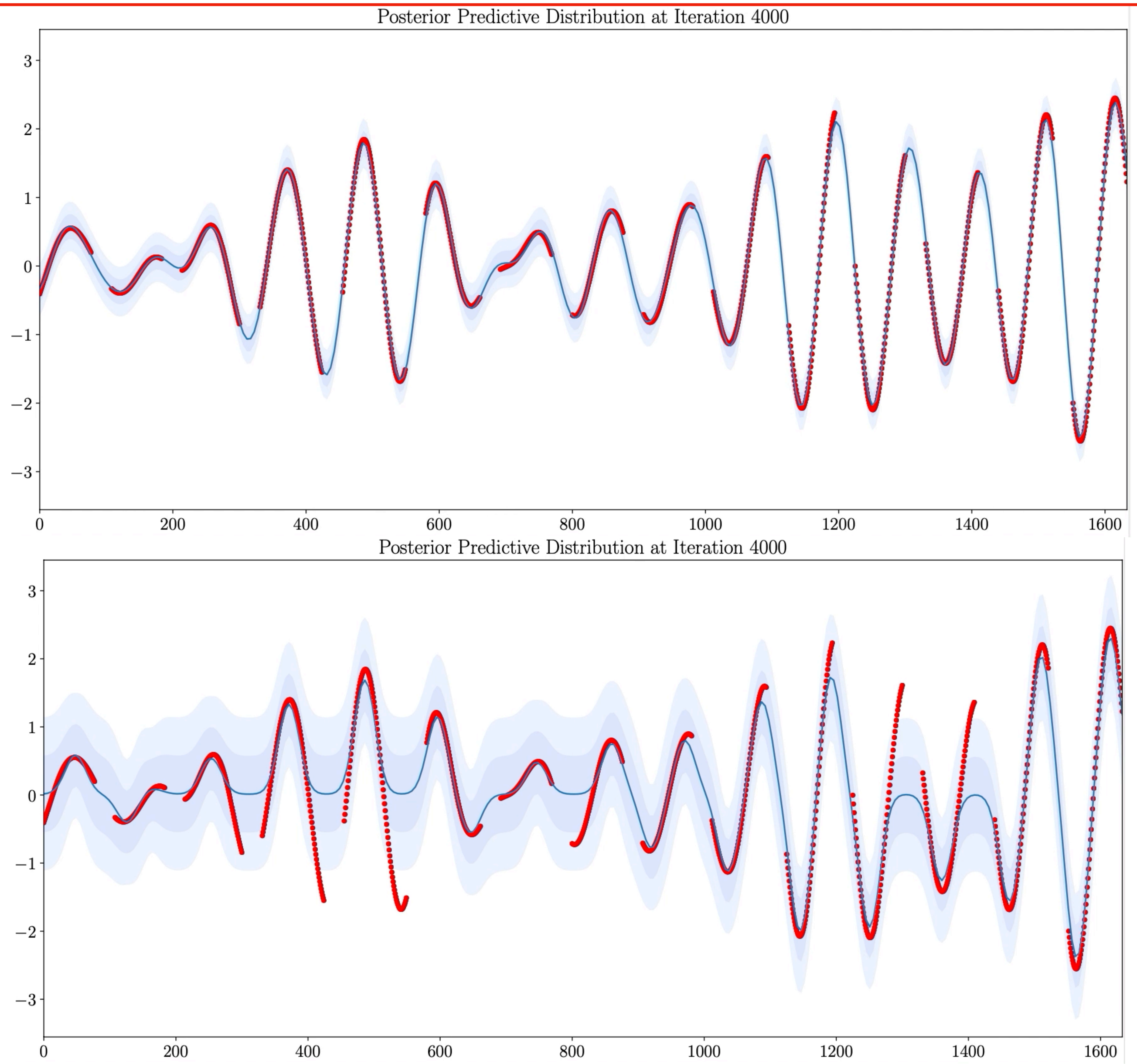
- **Exploit global structure in the data**
- **Better predictive performance**
- **Higher computational efficiency**
- **Simple drop-in replacement**





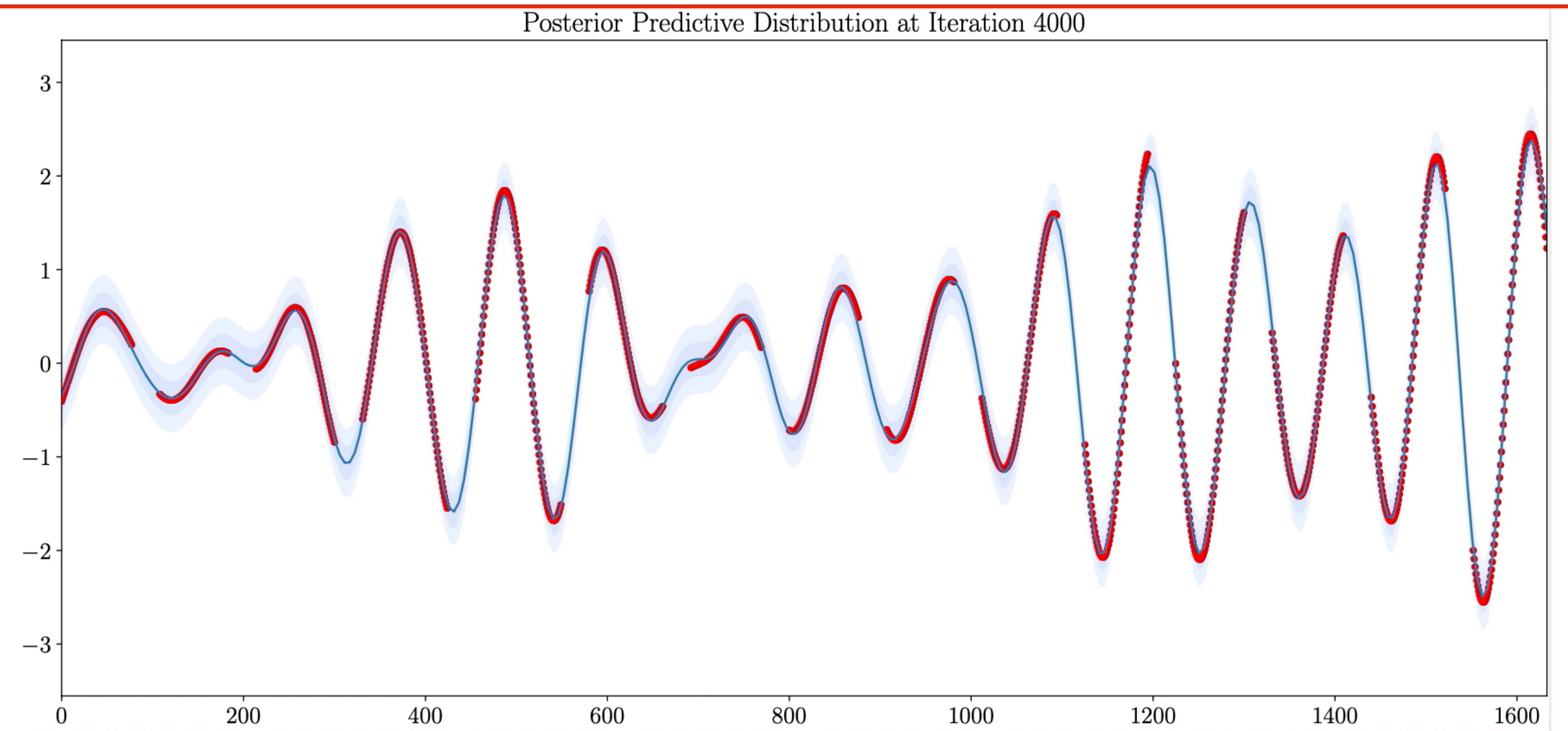
Benefits of Inter-domain Deep GPs

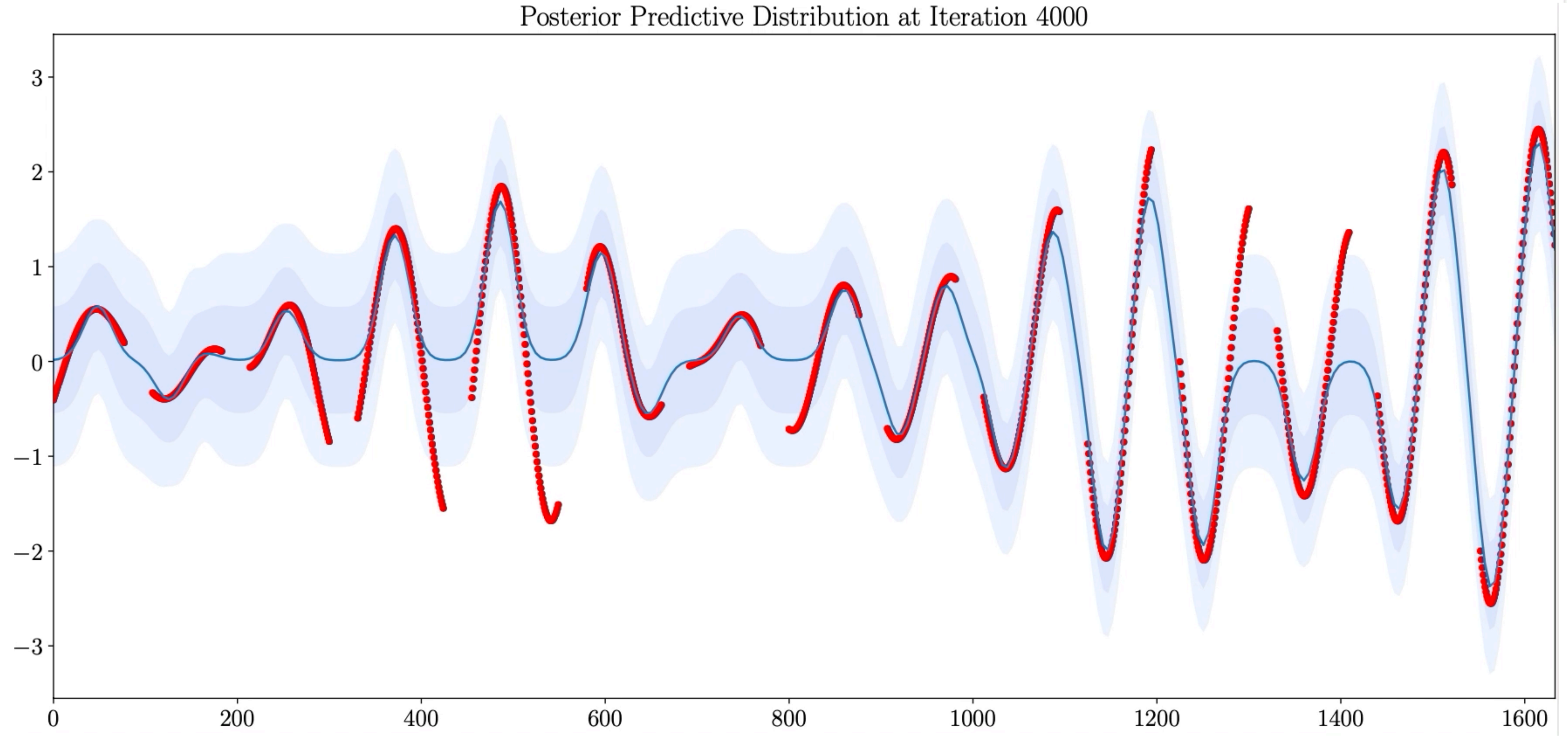
- **Exploit global structure in the data**
- **Better predictive performance**
- **Higher computational efficiency**
- **Simple drop-in replacement**



Benefits of Inter-domain Deep GPs

- **Exploit global structure in the data**
- **Better predictive performance**
- **Higher computational efficiency**
- **Simple drop-in replacement**





Thank You!

Project Website:

`http://bit.ly/inter-domain-dgp`

Email: tim.rudner@cs.ox.ac.uk

Website: http://timrudner.com

Twitter: @timrudner