Multi-output Gaussian Process Regression: Applications to Uncertainty Quantification

N. Zabaras[a], I. Bilionis[a] and G. Lin[b]

[a]Materials Process Design and Control Laboratory,
Mech. & Aero. Eng.,
Cornell Univ., Ithaca, NY 14853
zabaras@cornell.edu
http://mpdc.mae.cornell.edu

Cornell University

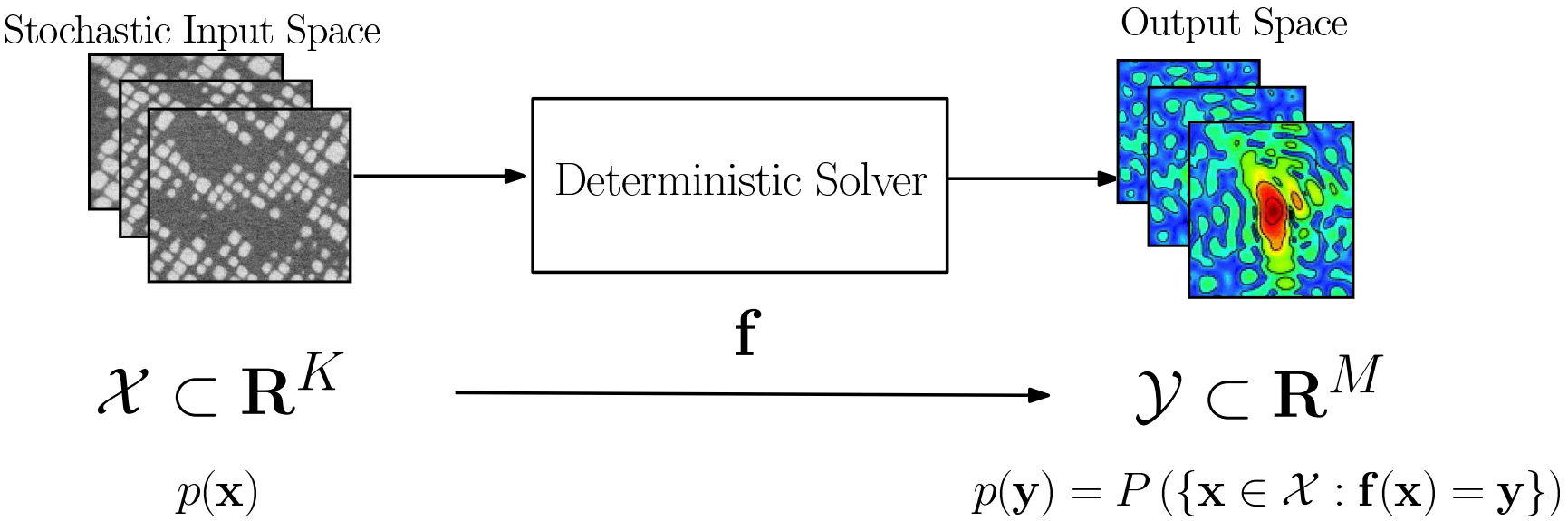
[b]Pacific Northwest National Laboratory,
Richland, WA

2011 DOE Applied Mathematics Program, 10/17/2011

Overview

- Problem definition
- Motivation
- Methodology
- Numerical Examples
- Conclusions

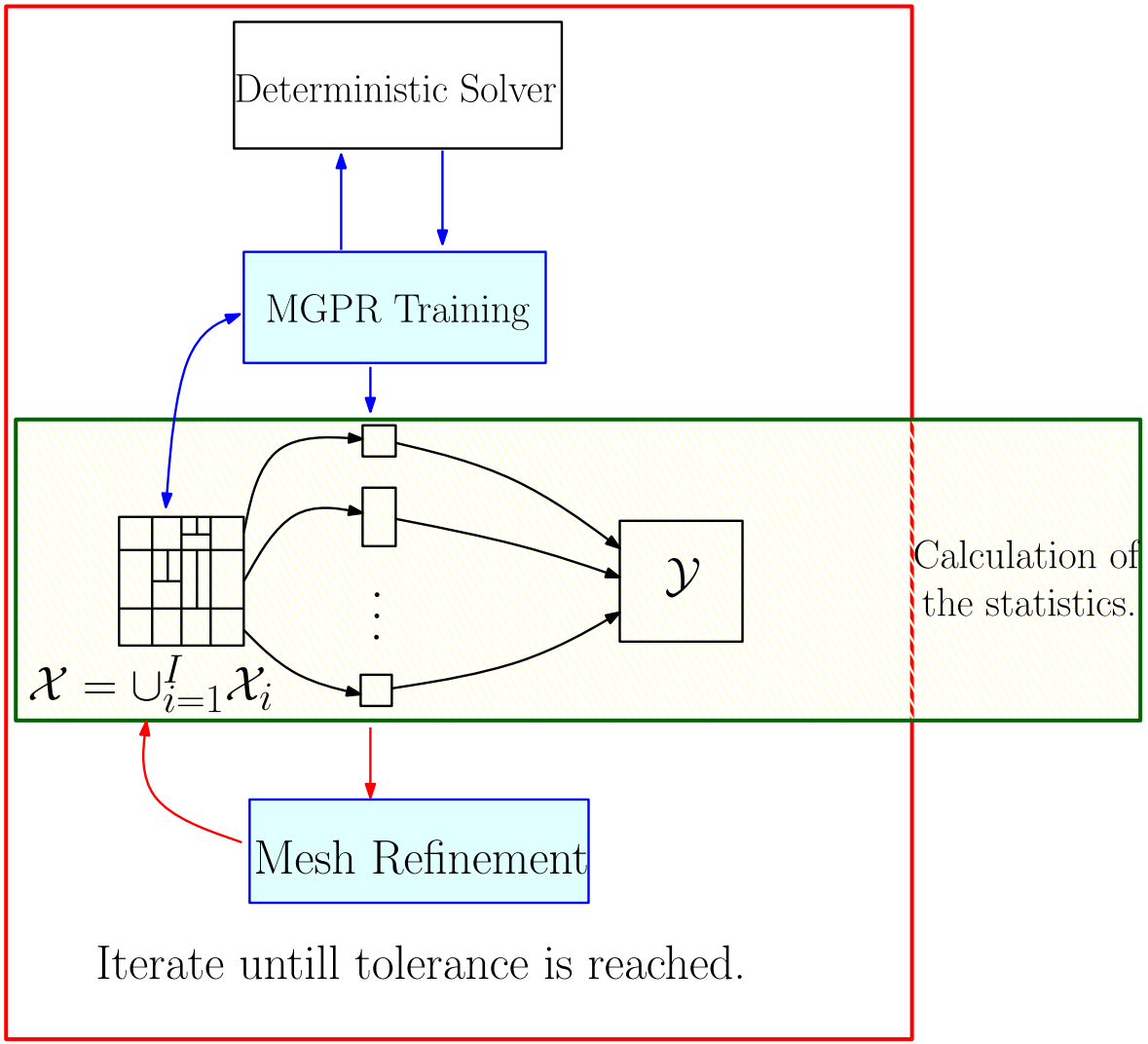
Problem Definition



Goal

Evaluate the statistics with as few calls to the deterministic solver as possible.

Two-Stage Solution Strategy



Desired Surrogate Properties

- Bayesian (Error Bars, Experimental Design).
- Non-Parametric (Expressive).
- Non-Isotropic (Localized Features).
- Cheap Evaluation of Statistics.

Desired Surrogate Properties

- Bayesian (Error Bars, Experimental Design).
- Non-Parametric (Expressive).
- Non-Isotropic (Localized Features).
- Cheap Evaluation of Statistics.

Multi-output Gaussian Process (MGP)

Desired Surrogate Properties

- Bayesian (Error Bars, Experimental Design).
- Non-Parametric (Expressive).
- Non-Isotropic (Localized Features).
- Cheap Evaluation of Statistics.

Patches of Isotropic MGP

Inspired from Multi-Element Method [Wan & Karniadakis (2005)]

Desired Surrogate Properties

- Bayesian (Error Bars, Experimental Design).
- Non-Parametric (Expressive).
- Non-Isotropic (Localized Features).
- Cheap Evaluation of Statistics.

Integrals of the Covariance Function

Analytical for uniform input distribution

Numerical integration for independent random variables

General Setting I

- Input space *adaptively* decomposed in *stochastic elements*:

$$\mathbf{X} = \cup_{i=1}^{I} \mathbf{X}_i \text{ and } \text{int}(\mathbf{X}_i) \cap \text{int}(\mathbf{X}_j) = \emptyset, \forall i, j \in I, i \neq j.$$

  We use rectangles:

$$\mathbf{X}_i = \times_{k=1}^{K} [a_k^i, b_k^i].$$

- Assume independent random variables (common case in UQ):

$$p(\mathbf{x}) = \prod_{k=1}^{K} p_k(x_k).$$

General Setting II

- Conditional input probability:

$$p_i(\mathbf{x}) := \frac{p(\mathbf{x})}{P(\mathbf{X}_i)} 1_{\mathbf{X}_i}(\mathbf{x}) = \prod_{k=1}^{K} \frac{p_k(x_k)}{P_k([a_k^i, b_k^i])} 1_{[a_k^i, b_k^i](x_k)}.$$

- N (fixed) data points drawn from $p_i(\mathbf{x})$ observed on $\mathbf{X}_i$:

$$\mathcal{D} := \left\{ (\mathbf{x}^{(n)}, \mathbf{y}^{(n)} \right\}_{n=1}^{N}$$

- A MGP is learnt on each element $\mathbf{f}_i$.
- The *global* surrogate:

$$\mathbf{f}(\mathbf{x}) := \sum_{i=1}^{I} \mathbf{f}^i(\mathbf{x}) 1_{\mathbf{X}_i}(\mathbf{x}).$$

Multi-output Gaussian Process I

- Scale the observed data:

$$\mu_{\text{obs},r} = \frac{1}{N}\sum_{n=1}^{N} y_r^{(n)},\ \sigma^2_{\text{obs},r} = \frac{1}{N}\sum_{n=1}^{N}(y_r - \mu_{\text{obs},r})^2.$$

- Define *scaled responses*: $g_r(\mathbf{x}) = \frac{f_r(\mathbf{x}) - \mu_{\text{obs},r}}{\sigma_{\text{obs},r}}$.
- Assume *Conditional independence* given the hyper-parameters:

$$g_r(\mathbf{x})|\boldsymbol{\theta} \sim \mathcal{GP}\left(0, k(\mathbf{x}, \mathbf{x}'; \boldsymbol{\theta})\right).$$

- Obtain $\boldsymbol{\theta}$ by maximizing the *marginal likelihood* of the data.

Multi-output Gaussian Process II

- Scaling back to the original responses, we can easily see that the predictive distribution of $f_r(\mathbf{x})$ is Gaussian:

$$f_r(\mathbf{x})|\mathcal{D}_r, \boldsymbol{\theta}^* \sim \mathcal{N}\left(\mu_{f_r}(\mathbf{x};\boldsymbol{\theta}^*), \sigma^2_{f_r}(\mathbf{x};\boldsymbol{\theta}^*)\right), \tag{1}$$

with mean

$$\mu_{f_r}(\mathbf{x};\boldsymbol{\theta}^*) = \sigma_{\text{obs},r}\mathbf{k}^T\mathbf{K}^{-1}\mathbf{z}_r + \mu_{\text{obs},r}, \tag{2}$$

and variance

$$\sigma^2_{f_r}(\mathbf{x};\boldsymbol{\theta}^*) = \sigma^2_{\text{obs},r}\left(k(\mathbf{x},\mathbf{x};\boldsymbol{\theta}^*) - \mathbf{k}^T\mathbf{K}^{-1}\mathbf{k}\right). \tag{3}$$

Adaptive Domain Decomposition

Assume given decomposition and current global surrogate learnt from data.

- Refine an element or not?
- What is the most important dimension?

Our Confidence on the Prediction I

- Let $\sigma^2_{f_r^i}(\mathbf{x})$ be the predictive variance of the $r = 1, \ldots, M$ output of the local surrogate of $\mathbf{f}^i$ at $\mathbf{x} \in \mathbf{X}_i$.
- The predictive variance of the global surrogate is:

$$\sigma^2_{f_r}(\mathbf{x}) = \sum_{i=1}^{I} \sigma^2_{f_r^i}(\mathbf{x}) 1_{\mathbf{X}_i}(\mathbf{x}).$$

- Uncertainty of prediction at test point:

$$\sigma^2_{\mathbf{f}}(\mathbf{x}) := \frac{1}{M} \sum_{r=1}^{M} \sigma^2_{f_r}(\mathbf{x}).$$

Our Confidence on the Prediction II

- Uncertainty of prediction over all the domain:

$$\sigma^2_{\mathbf{f},p} := \int_{\mathbf{X}} \sigma^2_{\mathbf{f}}(\mathbf{x}) p(\mathbf{x}) d\mathbf{x}.$$

- Element Contribution:

$$\sigma^2_{\mathbf{f},p} = \sum_{i=1}^{I} \sigma^2_{\mathbf{f},p_i} P(\mathbf{X}_i).$$

Refine or Not?

Yes, if

$$\sigma^2_{\mathbf{f},p_i} P(\mathbf{X}_i) > \delta.$$

The Most Important Dimension

- Squared Exponential covariance defines scales (unique to element):

$$k_{\text{SE}}(\mathbf{x}, \mathbf{x}') = s_f^2 \exp\left(-\frac{1}{2}\sum_{d=1}^{D}\frac{(x_k - x_k')^2}{\ell_k^2}\right),$$

- Probability of x_k falling inside $\mathbf{X}_i$:

$$p_k^i = \int_{a_k^i}^{b_k^i} p_k(x_k)dx_k.$$

Notice that if p_k is uniform, then

$$p_k^i \propto b_k^i - a_k^i,$$

is proportional to the extent of the along dimension k.

- Importance of dimension:

$$I_k^i = \frac{p_k^i}{\ell_k^i}.$$

How to Refine an Element?

Split perpendicular to the most *important dimension*:

$$k^* = \arg\max_k I_k^i = \arg\max_k \frac{p_k^i}{\ell_k^i}.$$

The element is split at the median of the conditional distribution $x_{k,\text{med}}^i$, so that both resulting sides are equally probable, i.e.

$$P_k^i([a_k^i, x_{k,\text{med}}^i]) = P_k^i([x_{k,\text{med}}^i, b_k^i) = \frac{1}{2}.$$

Notice that if p_k is uniform, then

$$x_{k,\text{med}}^i = \frac{b_k^i - a_k^i}{2}.$$

Collection of the Observations

We have to consider two distinct cases:

1. No data have been observed yet and we only have a single element (i.e. $\mathbf{X}$ itself).
2. We have obtained a fit of the response over an element $\mathbf{X}_i$ based on N_i observations

$$\mathcal{D}_i = \{(\mathbf{x}_i^{(n)}, \mathbf{y}_i^{(n)})\}_{n=1}^{N_i},$$

and we have decided to split it in two elements $\mathbf{X}_{i,1}$ and $\mathbf{X}_{i,2}$ so that

$$\mathbf{X}_i = \mathbf{X}_{i,1} \cup \mathbf{X}_{i,2} \text{ and } \mathbf{X}_{i,1} \cap \mathbf{X}_{i,2} = \emptyset.$$

Collection of the Observations I

- Given $\mathcal{D}_{i,1}^n$, observe the input point $\mathbf{x}^{\text{new},n+1} \in \mathbf{X}_{i,1}$ that maximizes the joint uncertainty of all outputs biased by the conditional distribution of the element:

$$\sigma_{\mathbf{f}}^2(\mathbf{x}; \boldsymbol{\theta}^*; \mathcal{D}_{i,1}^n) = \frac{1}{M} \sum_{r=1}^{M} \sigma_{f_r}^2(\mathbf{x}; \boldsymbol{\theta}^*, \mathcal{D}_{i,1}^n).$$

  That is,

$$\mathbf{x}^{\text{new},n+1} = \arg\max_{\mathbf{x} \in \mathbf{X}_{i,1}} \sigma_{\mathbf{f}}^2(\mathbf{x}; \boldsymbol{\theta}^*; \mathcal{D}_{i,1}^n) p_i(\mathbf{x}).$$

- We stop, either if N data points have been collected in $\mathcal{D}_{i,1}^n$, or if

$$\sigma_{\mathbf{f},p_{i,1}}^2(\mathcal{D}_{i,1}^n) P(\mathbf{X}_{i,1}) \leq \delta,$$

  where $\sigma_{\mathbf{f},p_{i,1}}^2(\mathcal{D}_{i,1}^n) P(\mathbf{X}_{i,1})$ is the expectation of $\sigma_{\mathbf{f}}^2(\mathbf{x}; \boldsymbol{\theta}^*; \mathcal{D}_{i,1}^n)$ with respect to the conditional probability $p_{i,1}(\mathbf{x})$ of $\mathbf{X}_{i,1}$.

Local Statistics I

- All quantities in this section are local to the element $\mathbf{X}_i$ but the notation is suppressed.
- We are interested in estimating all moments $\mathbf{m}^q = (m_1^q, \ldots, m_M^q)$, where

$$m_r^q = \int_{\mathbf{X}_i} f_r^q(\mathbf{x}) p_i(\mathbf{x}) d\mathbf{x}, \tag{4}$$

  and $q \geq 1$.
- Let us denote the predictive distribution for the q power of the response at $\mathbf{x} \in \mathbf{X}_i$ by

$$f_r^q(\mathbf{x}) | \mathcal{D}_r, \boldsymbol{\theta}^q \sim \mathcal{N}\left(\mu_{f_r^q}(\mathbf{x}; \boldsymbol{\theta}^q), \sigma^2_{f_r^q}(\mathbf{x}; \boldsymbol{\theta}^q)\right),$$

  where $\mu_{f_r^q}(\mathbf{x}; \boldsymbol{\theta}^q)$ is the predictive mean and $\sigma^2_{f_r^q}(\mathbf{x}; \boldsymbol{\theta}^q)$ the predictive variance for $r = 1, \ldots, M$.

Local Statistics II

- For convenience, let us write the predictive mean at $\mathbf{x}$ as

$$\mu_{f_r^q}(\mathbf{x};\boldsymbol{\theta}^q) = \sum_{n=1}^{N} \alpha_{rn}^q k(\mathbf{x}^{(n)},\mathbf{x};\boldsymbol{\theta}^q) + \mu_{\text{obs},r}^q,$$

and the predictive variance at $\mathbf{x}$ as

$$\sigma_{f_r^q}^2(\mathbf{x};\boldsymbol{\theta}^q) = (\sigma_{\text{obs},r}^q)^2 \left(k(\mathbf{x},\mathbf{x};\boldsymbol{\theta}^q) - \mathbf{k}^{q,T}(\mathbf{K}^q)^{-1}\mathbf{k}^q\right).$$

Local Statistics III

- To obtain analytic estimates of the predictive distribution of the statistics, let us make the simplifying assumption that the predictions at different input points **x** are conditionally independent given the data and the hyper-parameters. Then, by the additivity of independent normal variables, we arrive at the approximation:

$$m_r^q | \mathcal{D}_r, \boldsymbol{\theta}^q \sim \mathcal{N}\left(\mu_{m_r^q}, \sigma^2_{m_r^q}\right),$$

where the predictive mean of m_r^q is

$$\mu_{m_r^q} = \int_{\mathbf{X}_i} \mu_{f_r^q}(\mathbf{x}; \boldsymbol{\theta}^q) p_i(\mathbf{x}) d\mathbf{x},$$

and its predictive variance

$$\sigma^2_{m_r^q} = \int_{\mathbf{X}_i} \sigma^2_{f_r^q}(\mathbf{x}; \boldsymbol{\theta}^q) p_i(\mathbf{x}) d\mathbf{x}.$$

Local Statistics IV

- The integrals can be casted in terms of integrals of the covariance function:

$$\mu_{m_r^q} = \sum_{n=1}^{N} \alpha_{rn}^q \epsilon_n^q + \mu_r^q,$$

and

$$\sigma^2_{m_r^q} = (\sigma^q_{\text{obs},r})^2 \left(c^q - \sum_{n,l=1}^{N} (\mathbf{K}^q)^{-1}_{nl} \nu^q_{nl} \right),$$

where

$$\epsilon_n^q = \int_{\mathbf{x}_i} k(\mathbf{x}^{(n)}, \mathbf{x}; \boldsymbol{\theta}^q) p_i(\mathbf{x}) d\mathbf{x},$$

$$c^q = \int_{\mathbf{x}_i} k(\mathbf{x}, \mathbf{x}; \boldsymbol{\theta}^q) p_i(\mathbf{x}) d\mathbf{x},$$

Local Statistics V

$$\nu_{nl}^{q} = \int_{\mathbf{x}_i} k(\mathbf{x}, \mathbf{x}^{(n)}; \theta^q) k(\mathbf{x}, \mathbf{x}^{(l)}; \theta^q) p_i(\mathbf{x}) d\mathbf{x},$$

and $(\mathbf{K}^q)_{nl}^{-1}$ is the nl element of the inverse q covariance matrix $(\mathbf{K}^q)^{-1}$.

Global Statistics I

- Notice that m_r^q can be decomposed as

$$\begin{aligned} m_r^q &= \int_{\mathbf{X}} f_r^q(\mathbf{x}) p(\mathbf{x}) d\mathbf{x} \\ &= \sum_{i=1}^{I} \int_{\mathbf{X}_i} f_r^q(\mathbf{x}) \frac{p(\mathbf{x})}{P(\mathbf{X}_i)} d\mathbf{x} P(\mathbf{X}_i) \\ &= \sum_{i=1}^{I} \int_{\mathbf{X}_i} f_r^q(\mathbf{x}) p_i(\mathbf{x}) d\mathbf{x} P(\mathbf{X}_i), \end{aligned}$$

or

$$m_r^q = \sum_{i=1}^{I} m_r^{q,i} P(\mathbf{X}_i),$$

where $m_r^{q,i}$ is the local q moment over $\mathbf{X}_i$.

Global Statistics II

- Assuming conditional independence of the predictive distributions given the data and the hyper-parameters, we obtain that

$$m_r^q \,|\mathcal{D}_r, \boldsymbol{\theta}^q \sim \mathcal{N}\left(\mu_{m_r^q}, \sigma^2_{m_r^q}\right),$$

where the predictive mean is

$$\mu_{m_r^q} = \sum_{i=1}^{I} \mu_{m_r^{q,i}} P(\mathbf{X}_i),$$

and the variance

$$\sigma^2_{m_r^q} = \sum_{i=1}^{I} \sigma^2_{m_r^{q,i}} P(\mathbf{X}_i).$$

Global Statistics III

- Finally, we derive a normal approximation to the predictive distribution for the variance of the response $\mathbf{v} = (v_1, \ldots, v_M)$:

$$v_r \sim \mathcal{N}\left(\mu_{v_r}, \sigma^2_{v_r}\right).$$

Under the assumption of conditional independence of $m_r^q,\ q = 1, 2$, the predictive mean of v_r is given by

$$\mu_{v_r} = \mu_{m_r^2} - \mu^2_{m_r^1} - \sigma^2_{m_r^1}.$$

The predictive variance is

$$\sigma^2_{v_r} = \sigma^2_{m_r^2} + 4\mu^2_{m_r^1}\sigma^2_{\mu_r^1} + 2\sigma^4_{\mu_r^1}.$$

Numerical Examples

All examples are run on massively parallel computers at the National Energy Research Scientific Computing Center (NERSCC). The parallelization strategy is straightforward: each processor is assigned to work with a single element. The communication burden between the processes is minimal. The ultimate goal of the numerical examples is to demonstrate that the method can:

1. learn anisotropic surfaces,
2. deal with discontinuities,
3. identify localized features of the response and
4. reduce sampling frequency on unimportant input dimensions.

Numerical Examples

We test the method on the following problems:

- Simple validation examples (1, and input dimensions, 1 output dimension).
- Krainchnan-Orszag ODE (Identifying discontinuities).
- Elliptic PDE (Relatively high dimensions).
- A natural convection problem.
- Flow through porous media.

Simple Validation Examples

The purpose of this section is to demonstrate using simple, single-output functions, the claims $1-4$ made at the beginning of this section. The performance of each run is evaluated by comparing the predictive mean $\mu_{f_r}(\mathbf{x})$ to the true response. The error measure of choice here is the Mean Square Error (MSE) of $S = 10^5$ random samples drawn from $p(\mathbf{x})$. Specifically, MSE is defined to be

$$\text{MSE}\left(\mu_{f_r}(\cdot)\right) := \frac{1}{SM}\sum_{s=1}^{S}\sum_{r=1}^{M}\left(\mu_{f_r}(\mathbf{x}^{(s)}) - f_r(\mathbf{x}^{(s)})\right)^2,$$

where $\mathbf{x}^{(s)}, s = 1, \ldots, S$ are random samples from $p(\mathbf{x})$. Those sample were not used in the fitting procedure, hence MSE is a measure of the predictive capabilities of a regression method.

1D anisotropic, discontinuous function

Consider the real function:

$$f_1(x) = \begin{cases} \sin\left(\frac{\pi x}{5}\right) + \frac{1}{5}\cos\left(\frac{4\pi x}{5}\right), & x \leq 10 \\ \frac{x}{10} - 1, \text{otherwise} \end{cases}, \tag{5}$$

on the domain $\mathbf{X} = [0, 20]$. For $x \leq 10$, it varies with two different frequencies. For $x > 10$ it is linear and finally it has a discontinuity at $x = 10$.

1D anisotropic, discontinuous function

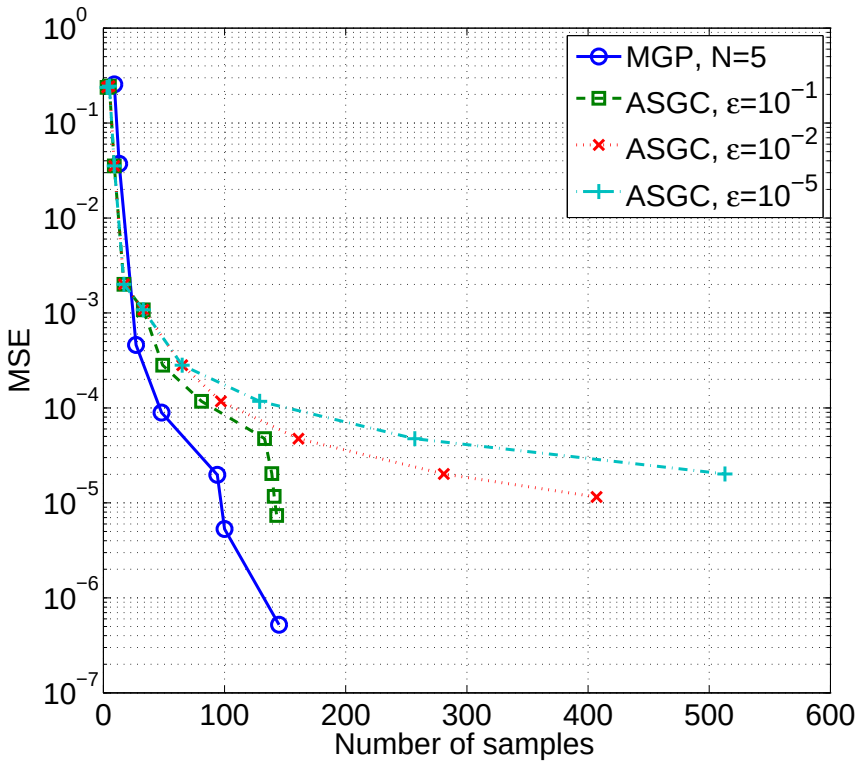


Figure: The MSE in the prediction of $f_1(x)$ as a function of the observed samples for MGP and ASGC for various ϵ.

1D anisotropic, discontinuous function

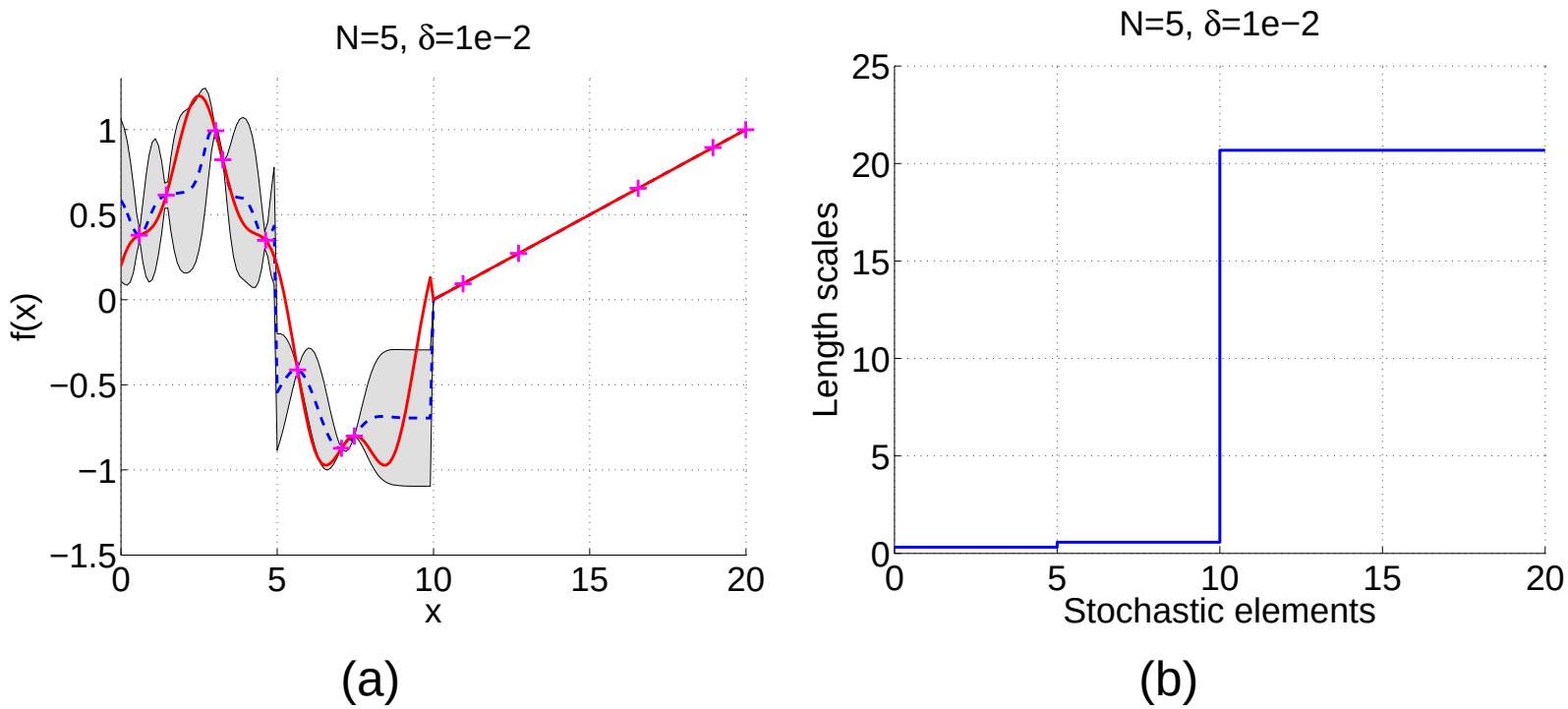


Figure: Left (a): comparison of the predictive mean $\mu_{f_1}(x)$ (dashed blue line) and 95% error bars (shaded grey area) with true response $f_1(x)$ (solid red line), where the symbols mark the observed data. Right (b): predicted length scale across the domain. Tolerances $\delta = 10^{-2}$ with 13 samples gathered.

1D anisotropic, discontinuous function

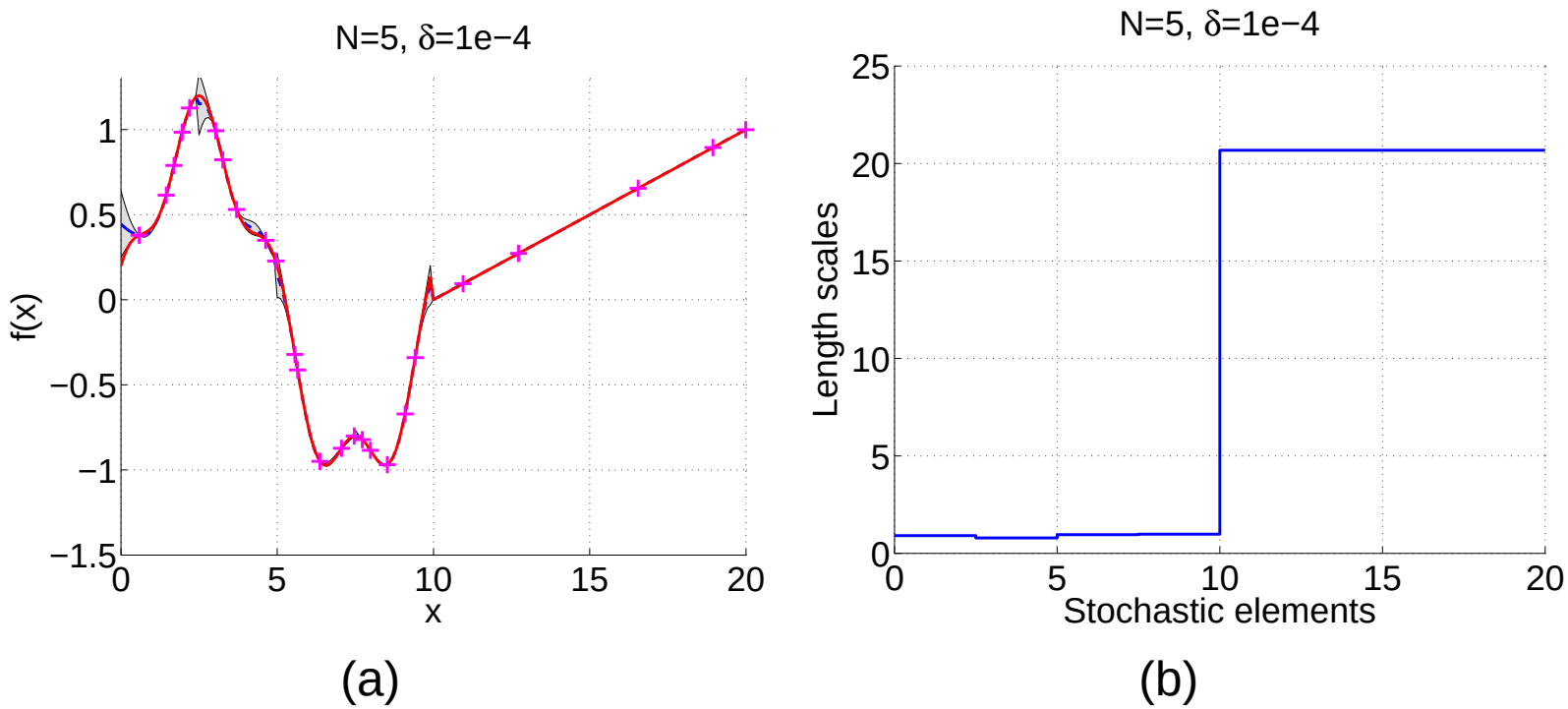


Figure: Left (a): comparison of the predictive mean $\mu_{f_1}(x)$ (dashed blue line) and 95% error bars (shaded grey area) with true response $f_1(x)$ (solid red line), where the symbols mark the observed data. Right (b): predicted length scale across the domain. Tolerances $\delta = 10^{-4}$ with 25 samples gathered.

1D anisotropic, discontinuous function

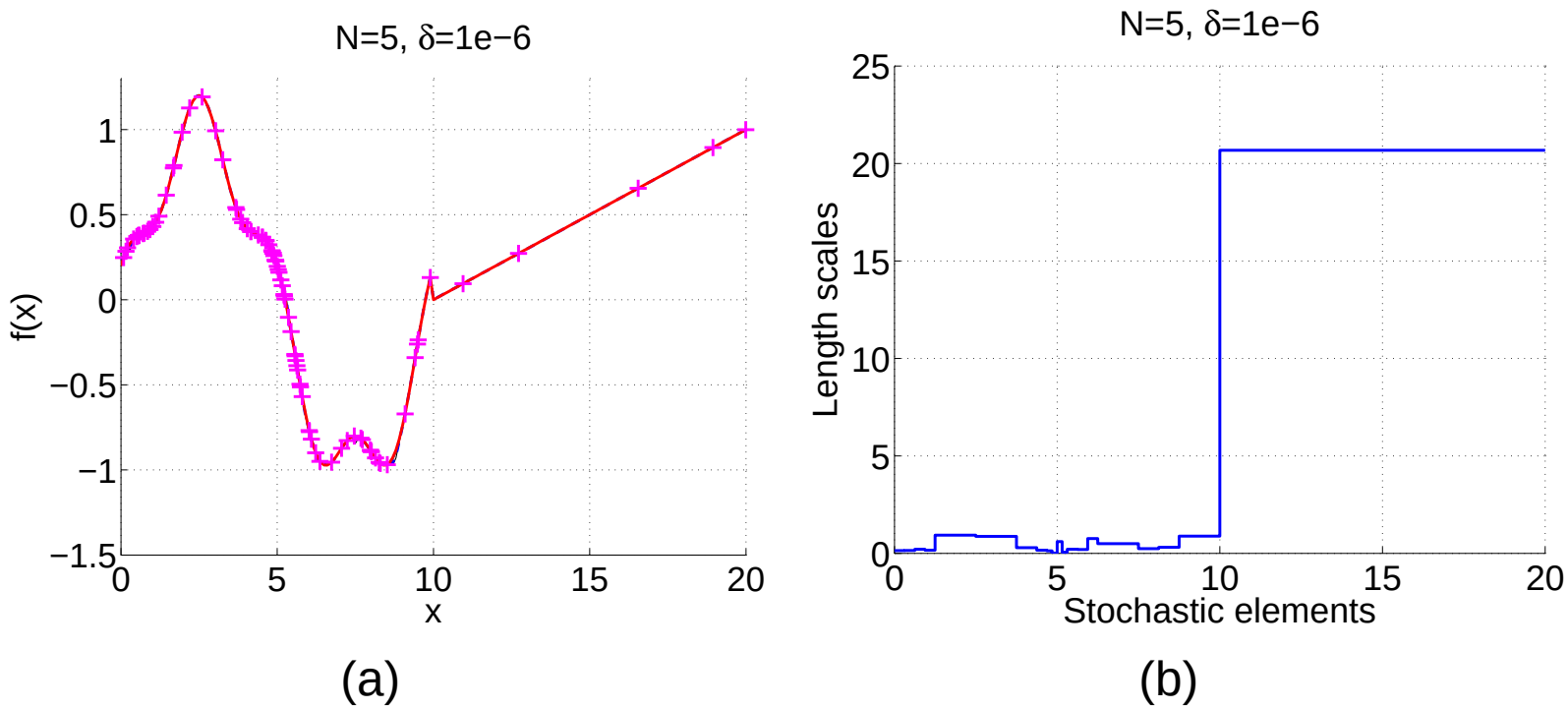


Figure: Left (a): comparison of the predictive mean $\mu_{f_1}(x)$ (dashed blue line) and 95% error bars (shaded grey area) with true response $f_1(x)$ (solid red line), where the symbols mark the observed data. Right (b): predicted length scale across the domain. Tolerances $\delta = 10^{-6}$ with 94 samples gathered.

2D function with local features

Let us now consider a two-dimensional real function:

$$f_2(x_1, x_2) = x_1 \exp\left\{-x_1^2 - x_2^2\right\}, \tag{6}$$

on $\mathbf{X} = [-2, 6]^2$.

2D function with local features

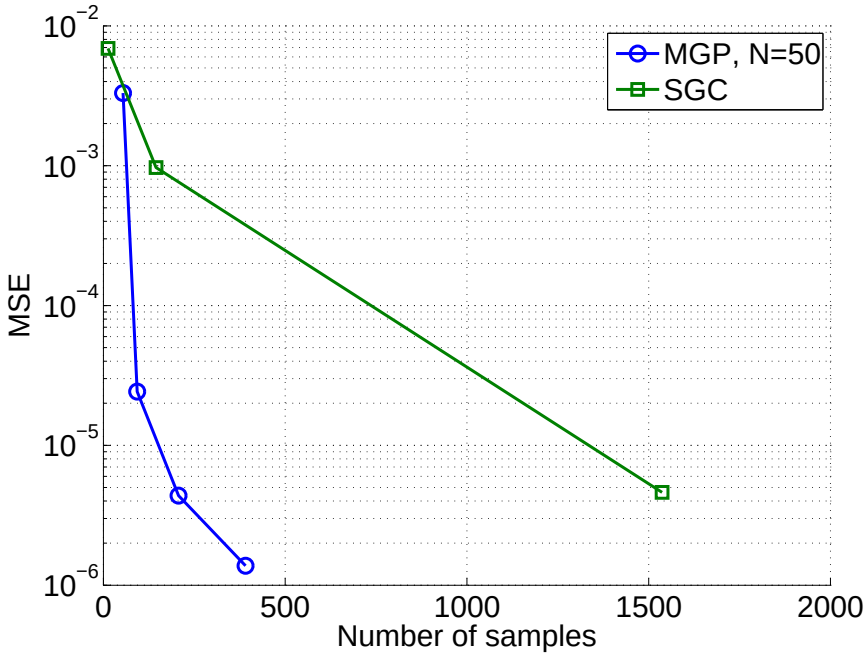


Figure: The MSE in the prediction of $f_2(\mathbf{x})$ as a function of the observed samples for MGP and SGC. ASGC ($\epsilon = 10^{-3}$) is not reported since it fails to identify the localized features.

2D function with local features

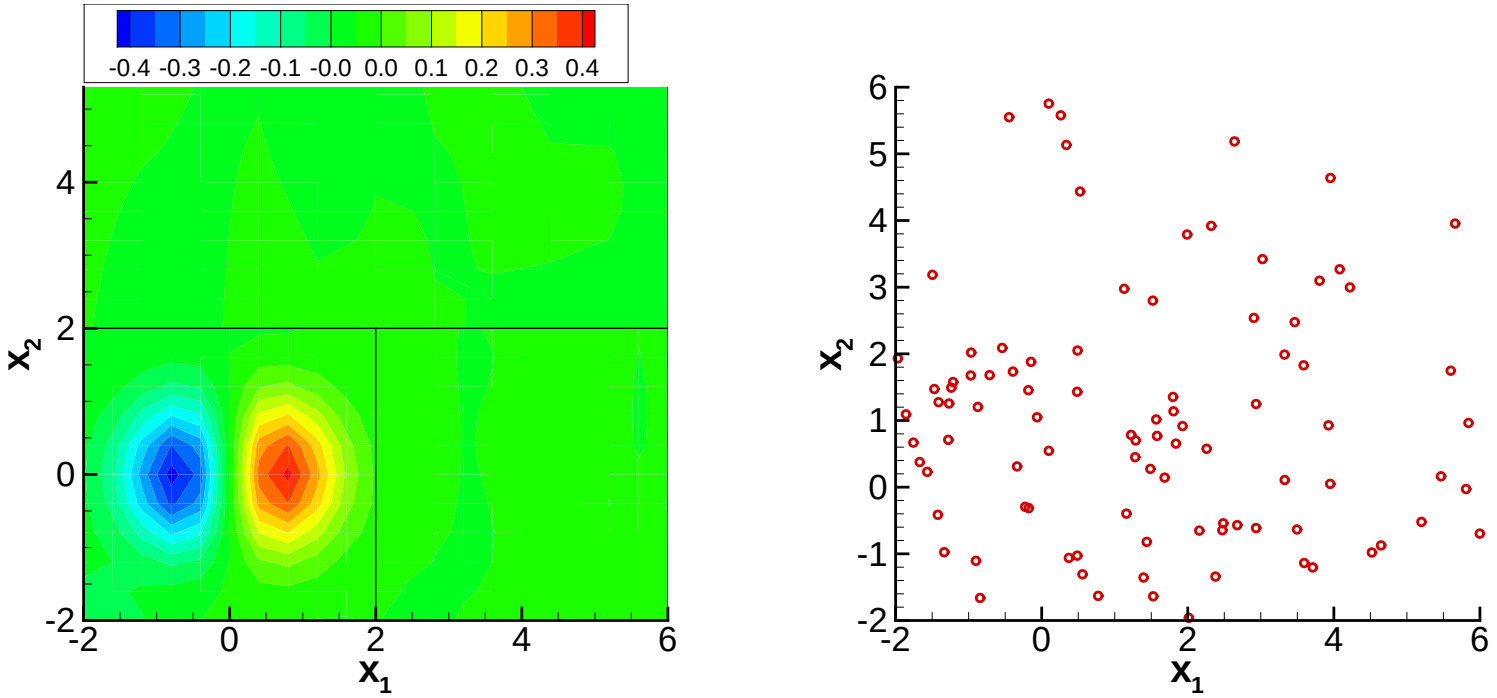


Figure: Left (a): predictive mean for $f_2(\mathbf{x})$ of the MGP ($N = 50$) and decomposition of the domain for $\delta = 10^{-4}$. Right (b): observations made for the same δ.

2D function with local features

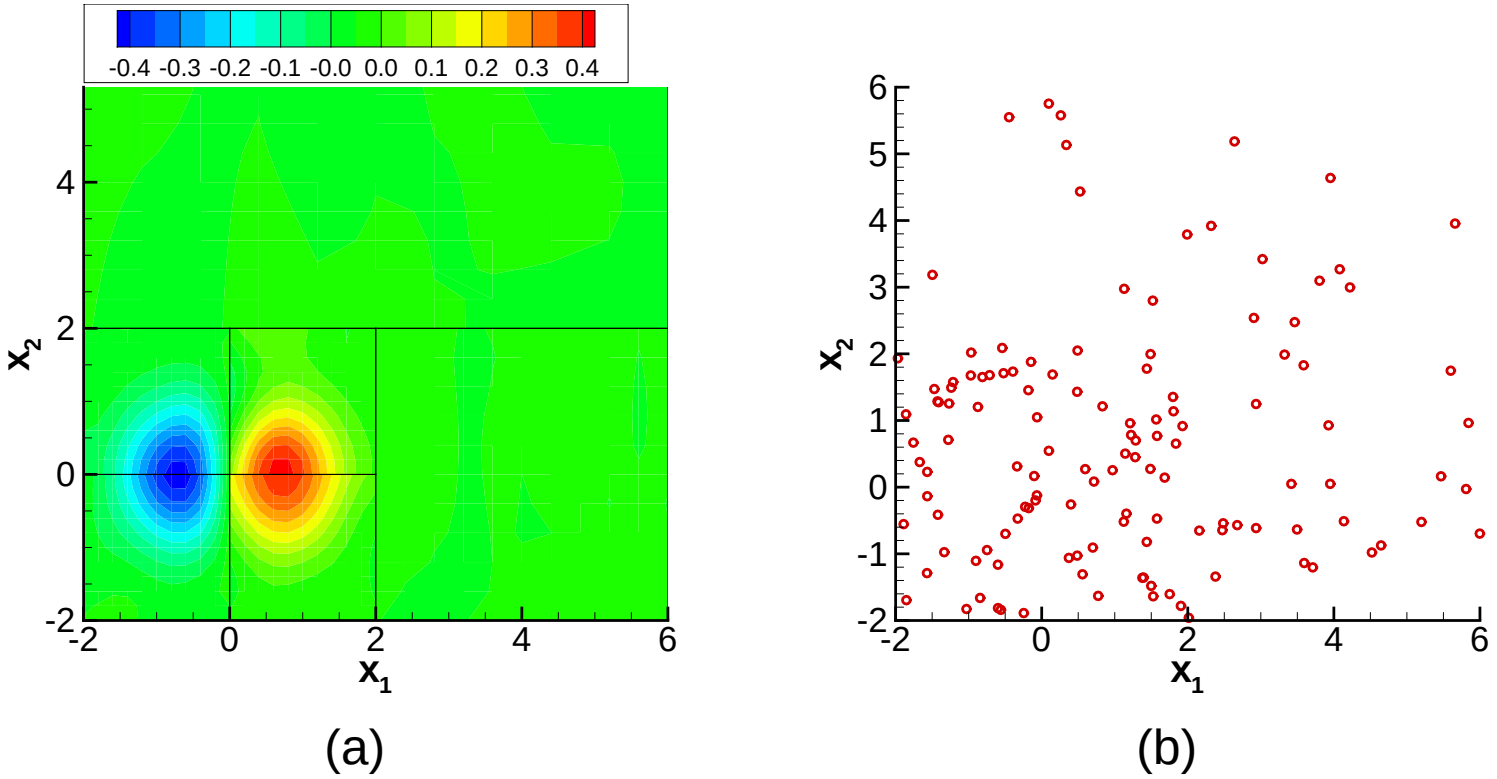


Figure: Left (a): predictive mean for $f_2(\mathbf{x})$ of the MGP ($N = 50$) and decomposition of the domain for $\delta = 10^{-5}$. Right (b): observations made for the same δ.

2D function with local features

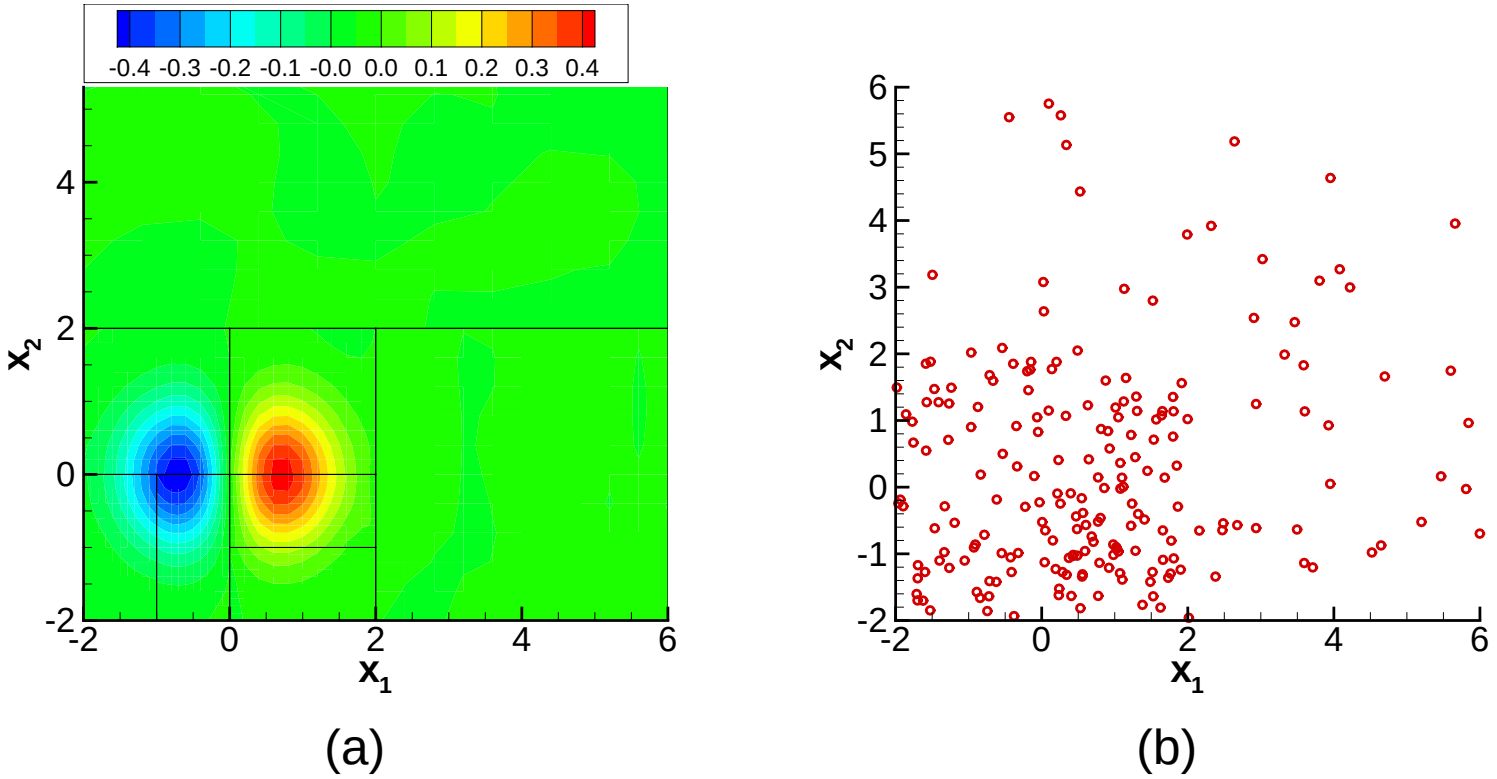


Figure: Left (a): predictive mean for $f_2(\mathbf{x})$ of the MGP ($N = 50$) and decomposition of the domain for $\delta = 10^{-6}$. Right (b): observations made for the same δ.

Numerical Examples: Krainchnan-Orszag

Consider the system of ordinary differential equations [Wan and Karniadakis (2005)]

$$\begin{aligned} \frac{dy_1}{dt} &= y_1 y_3, \\ \frac{dy_2}{dt} &= -y_2 y_3, \\ \frac{dy_3}{dt} &= -y_1^2 + y_2^2, \end{aligned}$$

subject to random initial conditions at $t = 0$.

$$y_1(0) = x_1,\ y_2(0) = x_2,\ y_3(0) = x_3,$$

where $x_i, i = 1, 2, 3$ are random parameters.

Numerical Examples: Krainchnan-Orszag

KO-1:

$$y_1(0) = 1,\ y_2(0) = 0.1x,\ y_3(0) = 0,$$

KO-2:

$$y_1(0) = 1,\ y_2(0) = 0.1x_1,\ y_3(0) = x_2,$$

KO-3:

$$y_1(0) = x_1,\ y_2(0) = x_2,\ y_3(0) = x_3,$$

where

$$x_i \sim U([-1, 1]), i = 1, 2, 3.$$

The error of the statistics will be evaluated using the (normalized) L_2 *norm of the error in variance* defined by:

$$\mathrm{E}_{\mathrm{L2}} = \frac{1}{M} \sum_{r=1}^{M} \left(v_{r,\mathrm{MC}} - \mu_{v_r} \right)^2, \tag{7}$$

where μ_{v_r} is the predictive mean of v_r (Eq. (19)). The results are compared with SGC and ASGC.

Numerical Examples: Krainchnan-Orszag

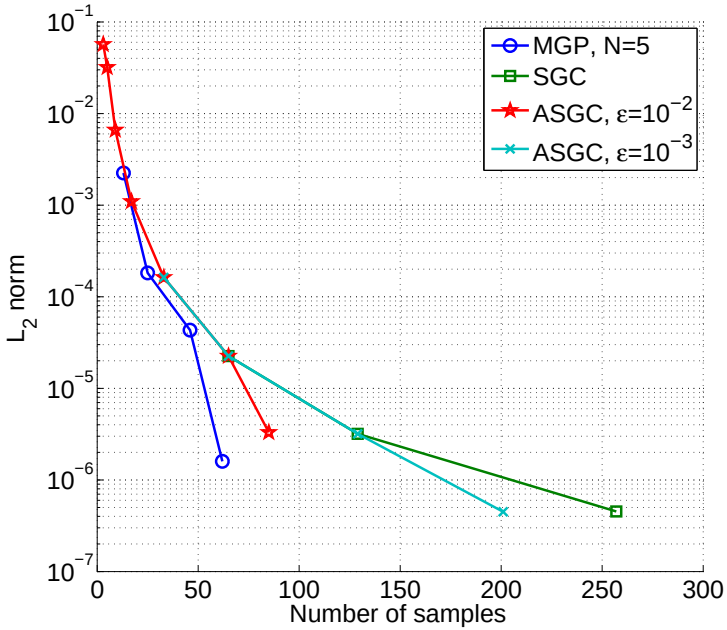
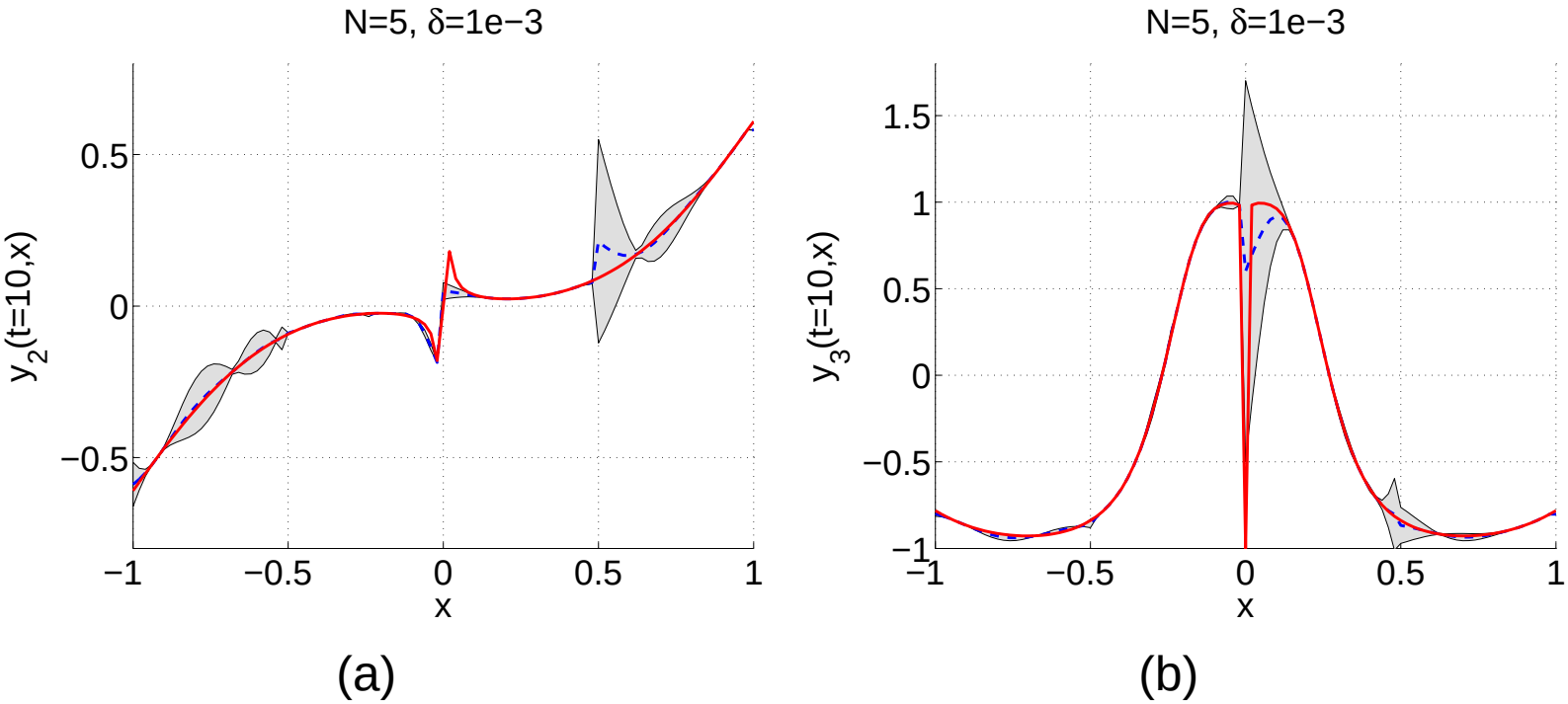


Figure: KO-1: the L_2 norm of the error in variance as a function of the observed samples for MGP, SGC and ASGC.

Numerical Examples: Krainchnan-Orszag



(a) (b)

Figure: KO-1: prediction (dashed blue) with 95% error bounds for tolerances (top to bottom) $\delta = 10^{-3}$ versus the true response (solid red) for y_2 $(t = 10)$ (left, (a)) and y_3 $(t = 10)$ (right, (b)).

Numerical Examples: Krainchnan-Orszag

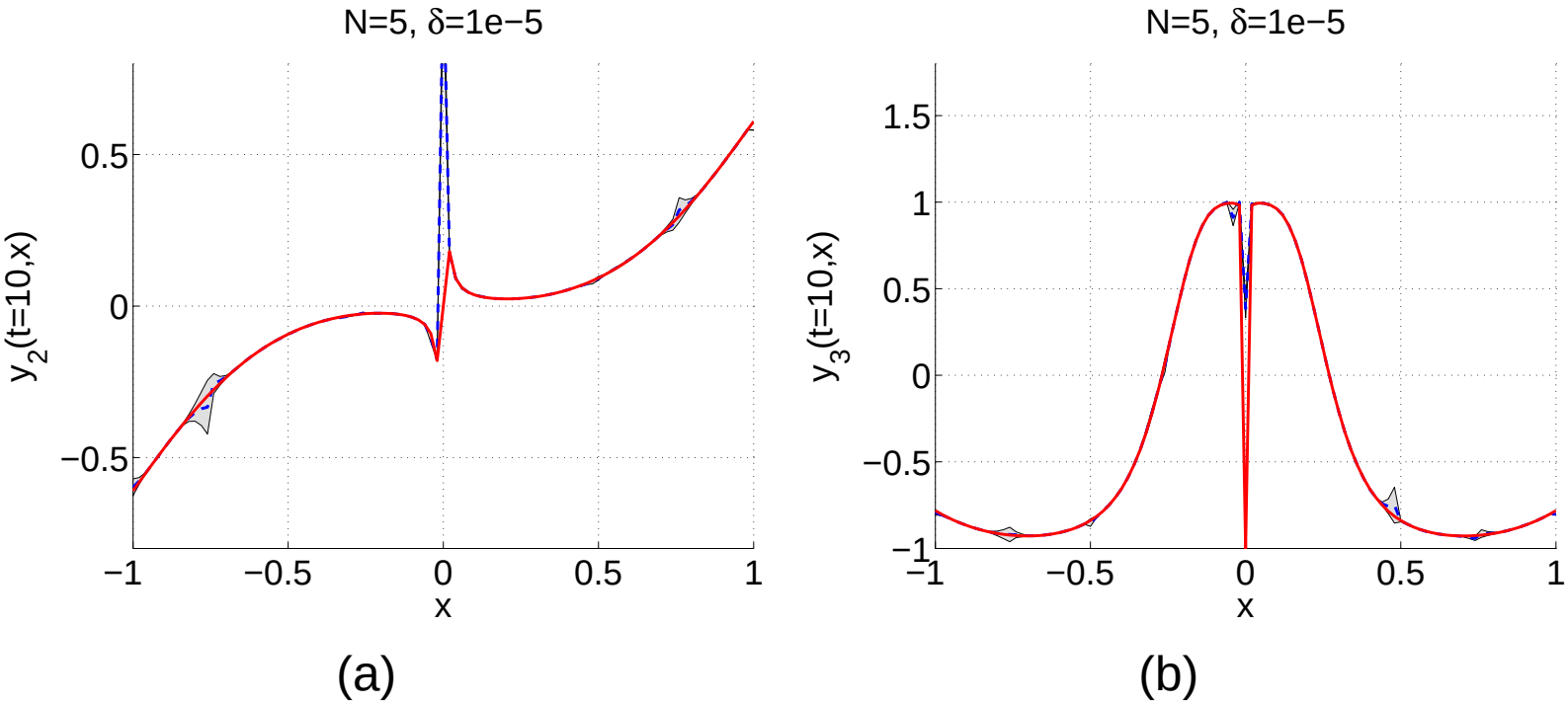


Figure: KO-1: prediction (dashed blue) with 95% error bounds for tolerances (top to bottom) $\delta = 10^{-5}$ versus the true response (solid red) for y_2 ($t = 10$) (left, (a)) and y_3 ($t = 10$) (right, (b)).

Numerical Examples: Krainchnan-Orszag

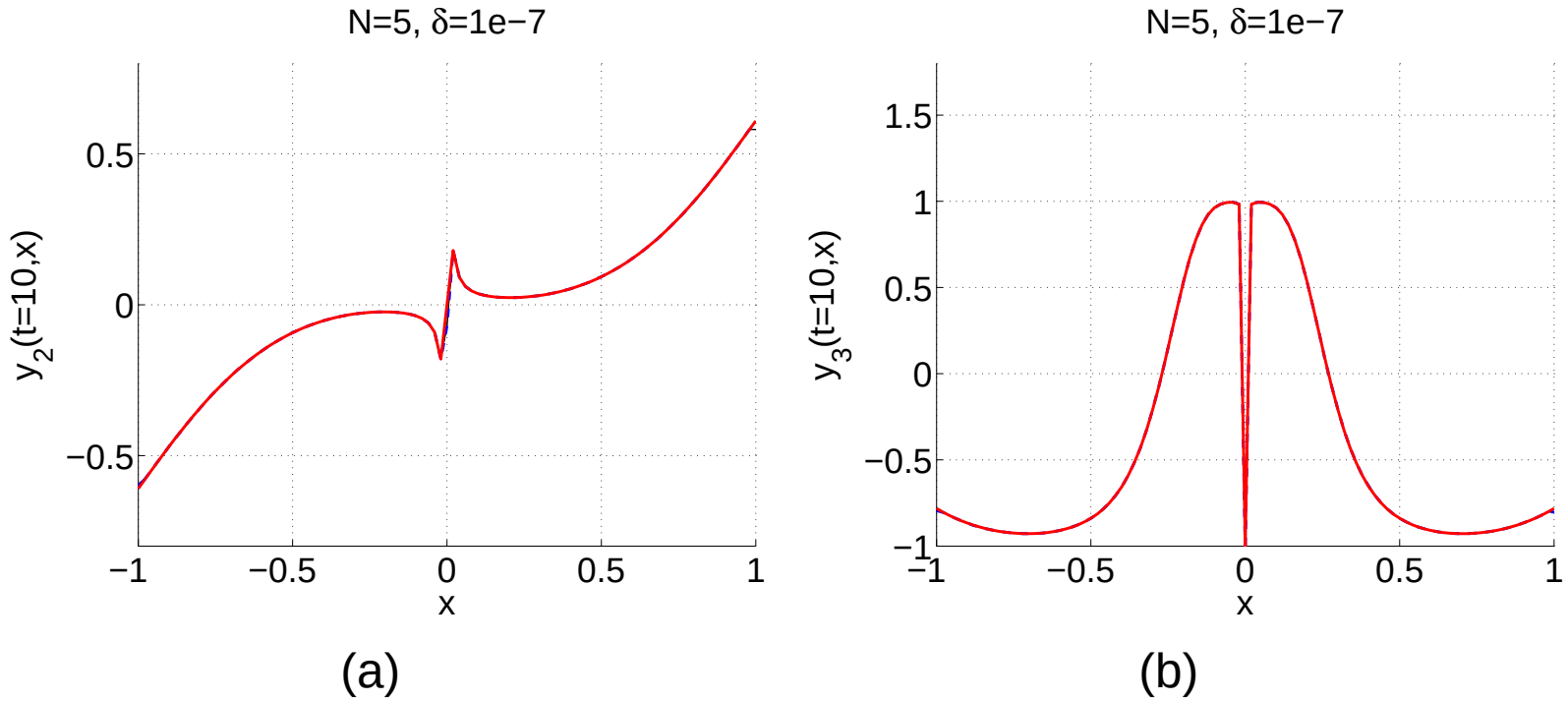


Figure: KO-1: prediction (dashed blue) with 95% error bounds for tolerances (top to bottom) $\delta = 10^{-7}$ versus the true response (solid red) for y_2 ($t = 10$) (left, (a)) and y_3 ($t = 10$) (right, (b)).

Numerical Examples: Krainchnan-Orszag

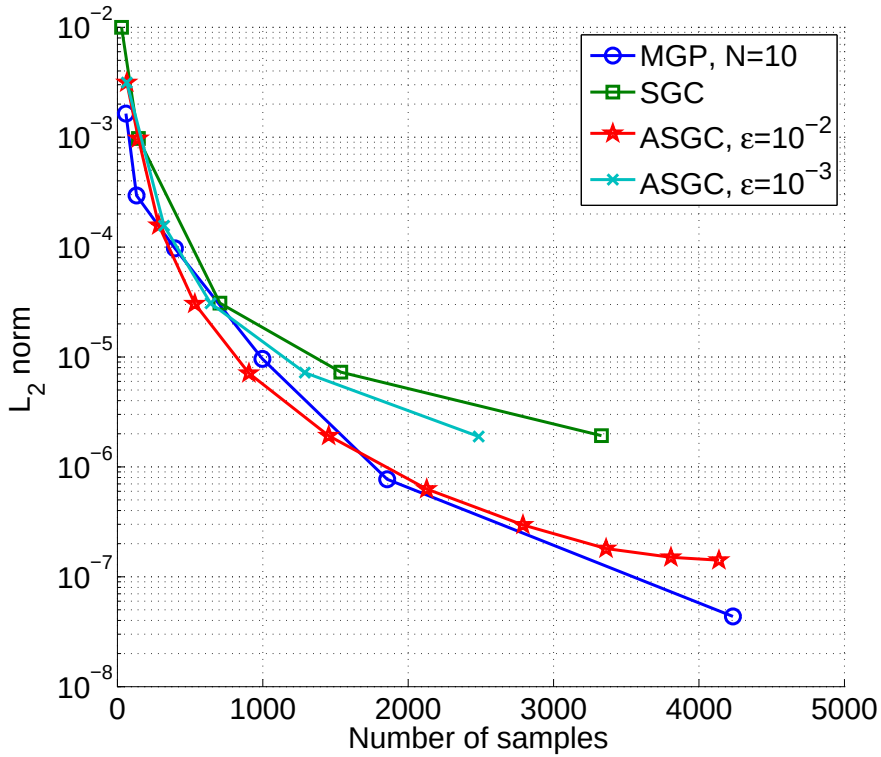


Figure: KO-2: the L_2 norm of the error in variance as a function of the observed samples for MGP, SGC and ASGC.

Numerical Examples: Krainchnan-Orszag

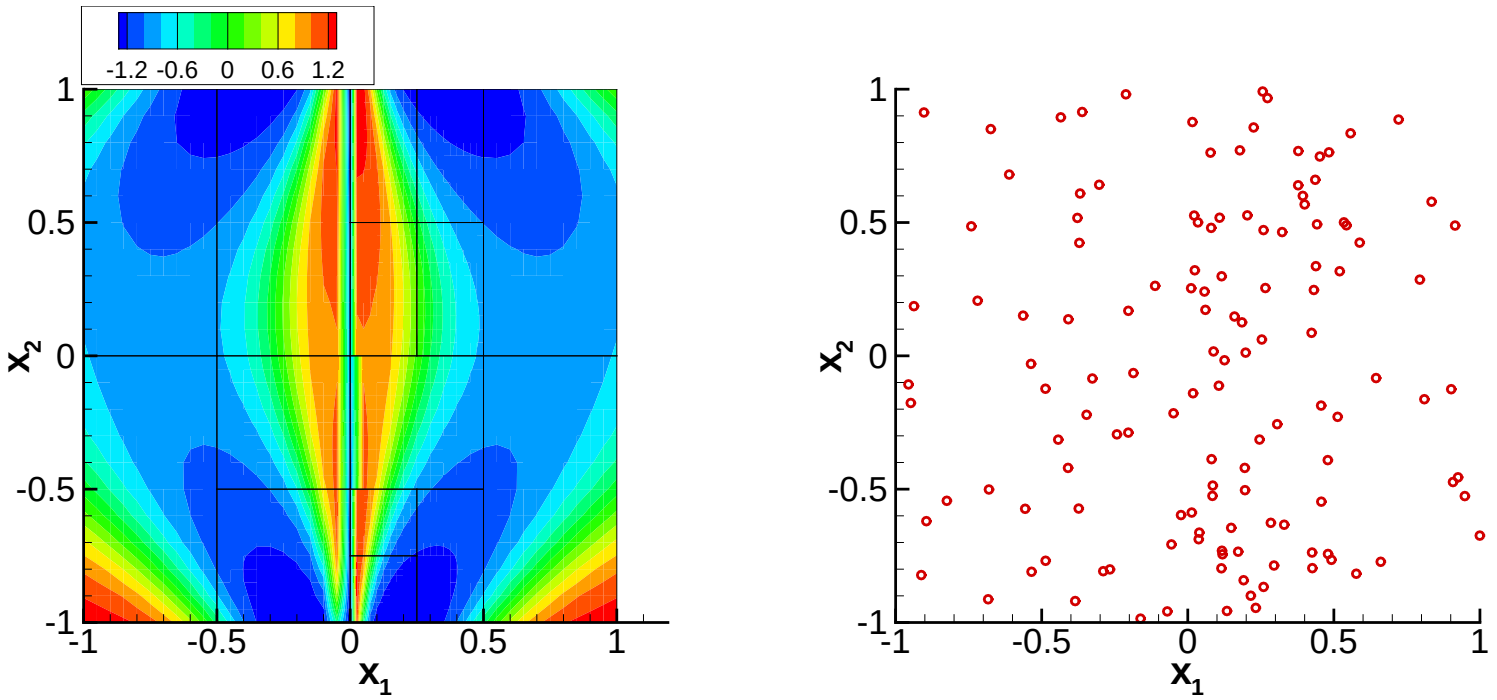


Figure: KO-2: The prediction at y_3 ($t = 10$) with the stochastic elements (left, (a)) and the observed samples (right, (b)) for tolerances $\delta = 10^{-3}$.

Numerical Examples: Krainchnan-Orszag

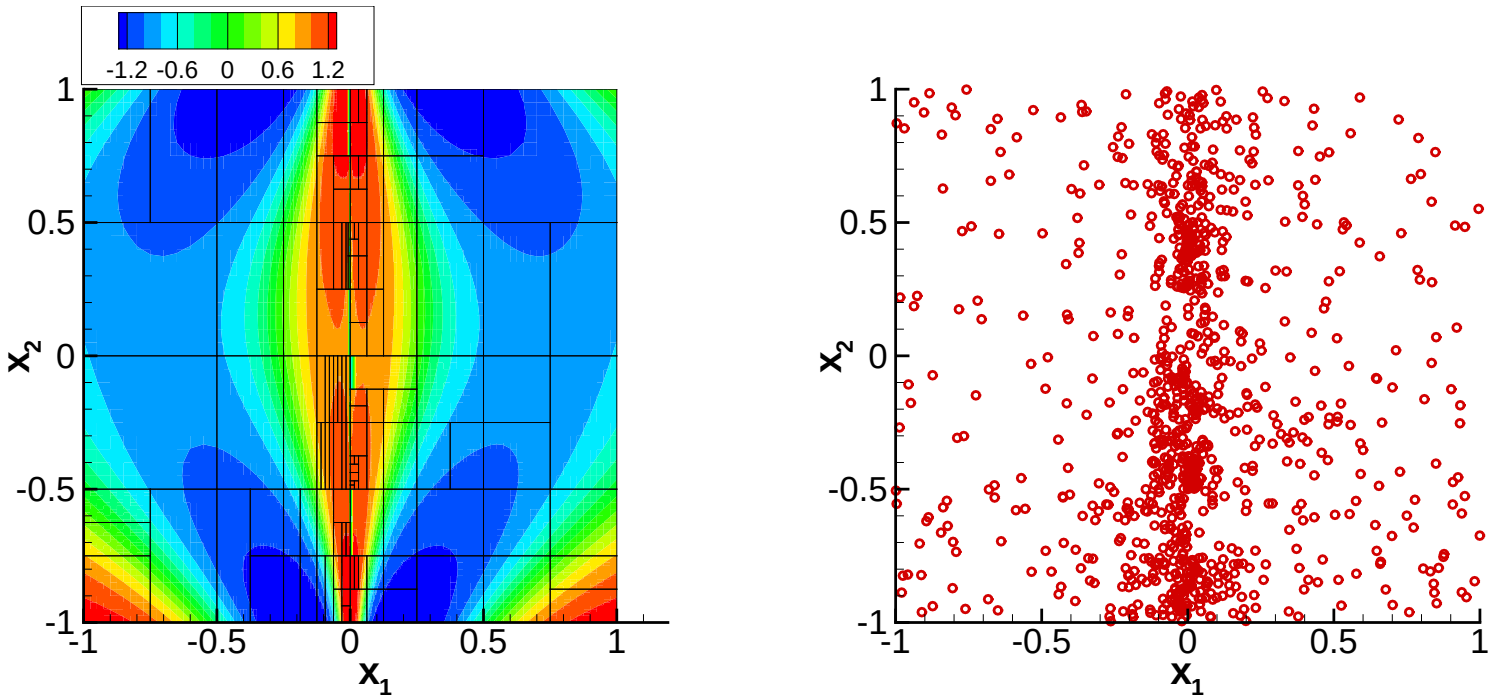


Figure: KO-2: The prediction at y_3 ($t = 10$) with the stochastic elements (left, (a)) and the observed samples (right, (b)) for tolerances $\delta = 10^{-5}$.

Numerical Examples: Krainchnan-Orszag

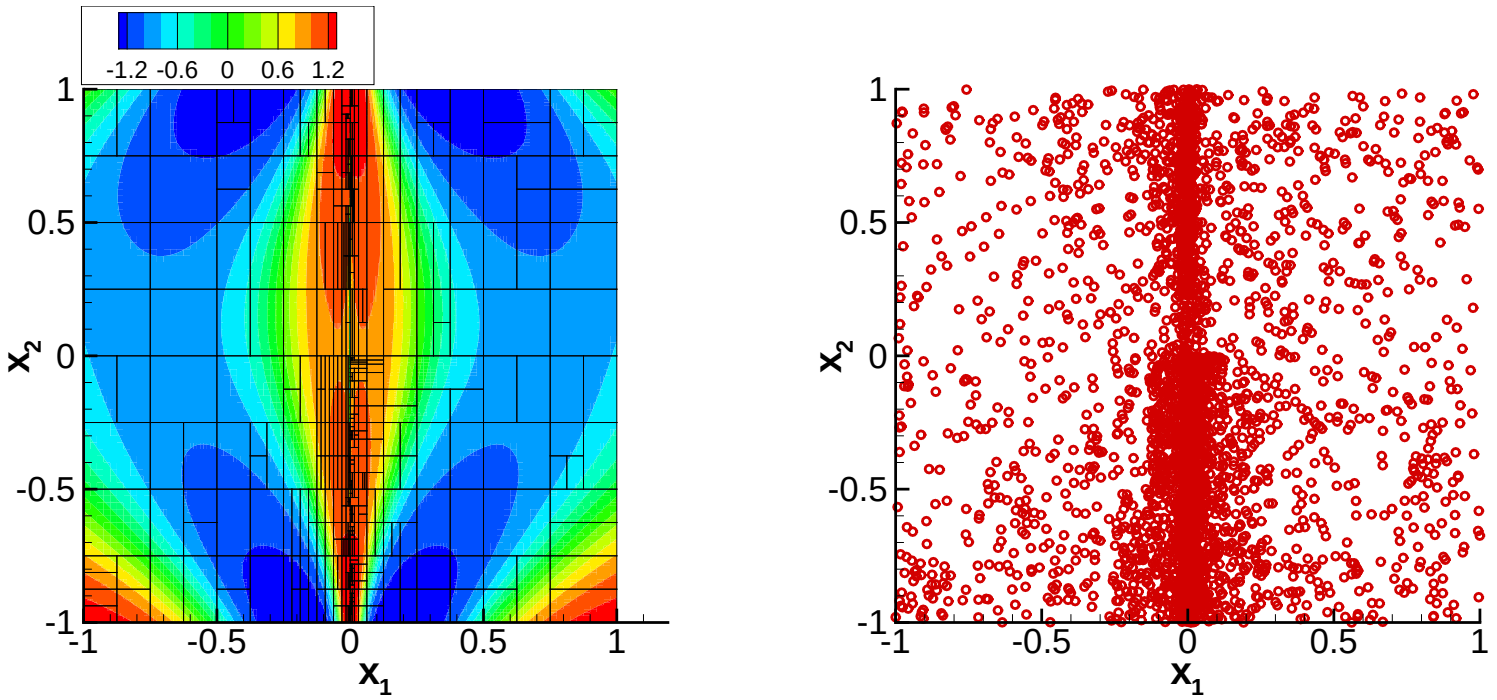
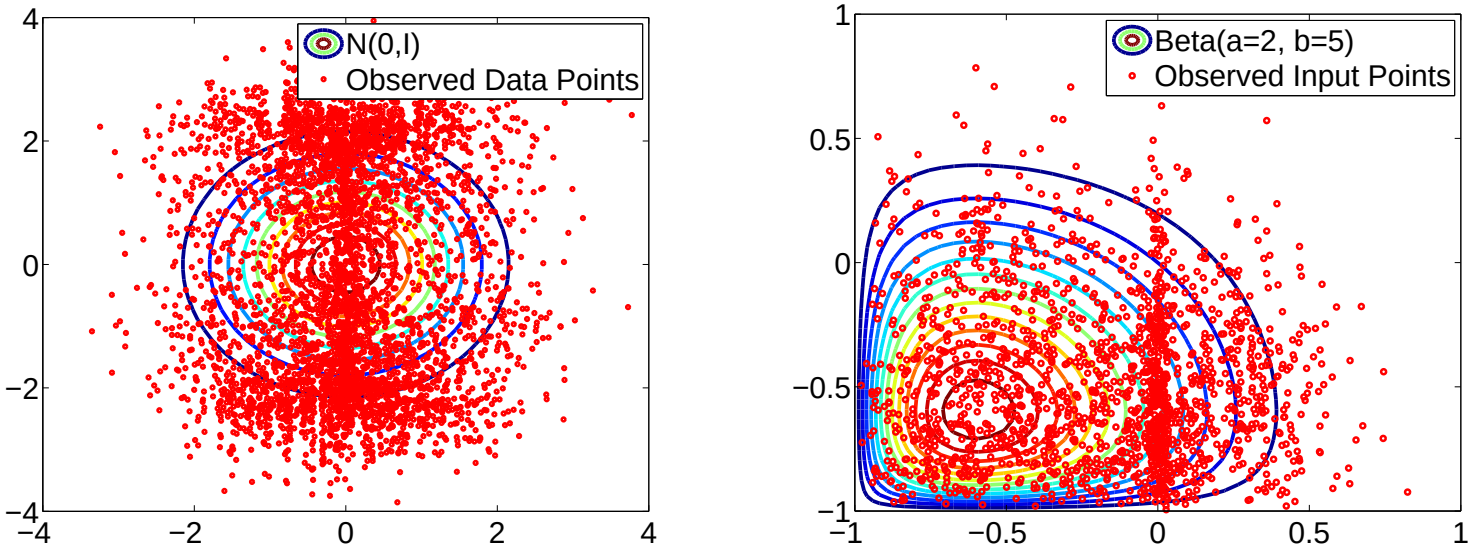


Figure: KO-2: The prediction at y_3 ($t = 10$) with the stochastic elements (left, (a)) and the observed samples (right, (b)) for tolerances $\delta = 10^{-7}$.

Numerical Examples: Krainchnan-Orszag



(a) (b)

Figure: KO-2: The observed input points for (a) Gaussian distribution with zero mean and unit variance and (b) beta distribution with $a = 2, b = 5$. for tolerance $\delta = 10^{-4}$.

Numerical Examples: Krainchnan-Orszag

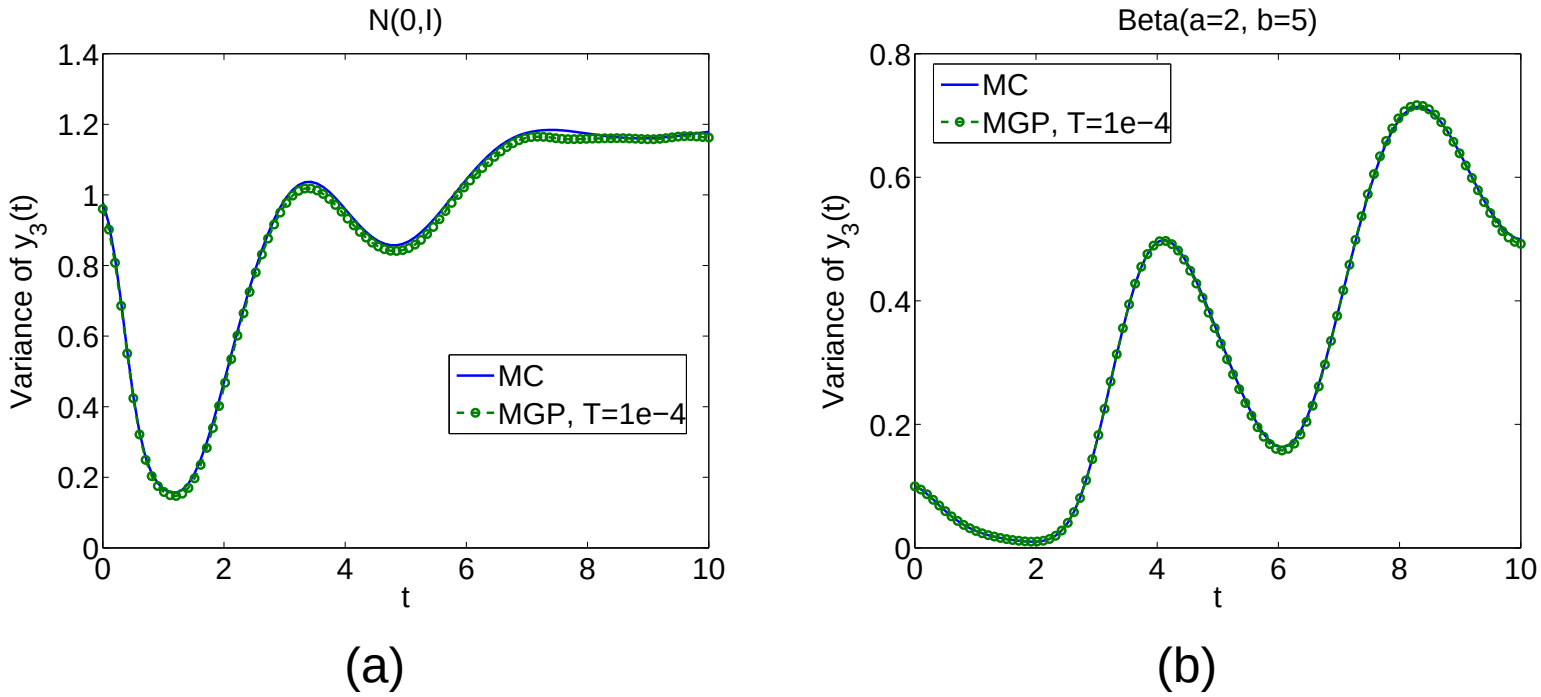


Figure: KO-2: The variance of $y_3(t)$ for (a) Gaussian distribution with zero mean and unit variance and (b) beta distribution with $a = 2, b = 5$. for tolerance $\delta = 10^{-4}$.

Numerical Examples: Krainchnan-Orszag

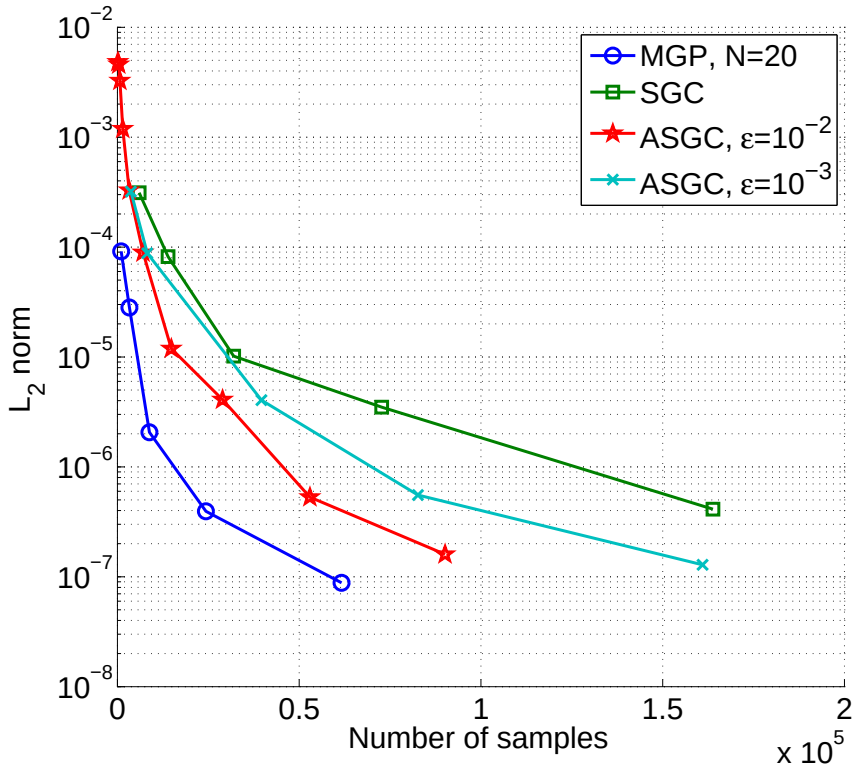


Figure: KO-3: the L_2 norm of the error in variance as a function of the observed samples for MGP, SGC and ASGC.

Numerical Examples: Krainchnan-Orszag

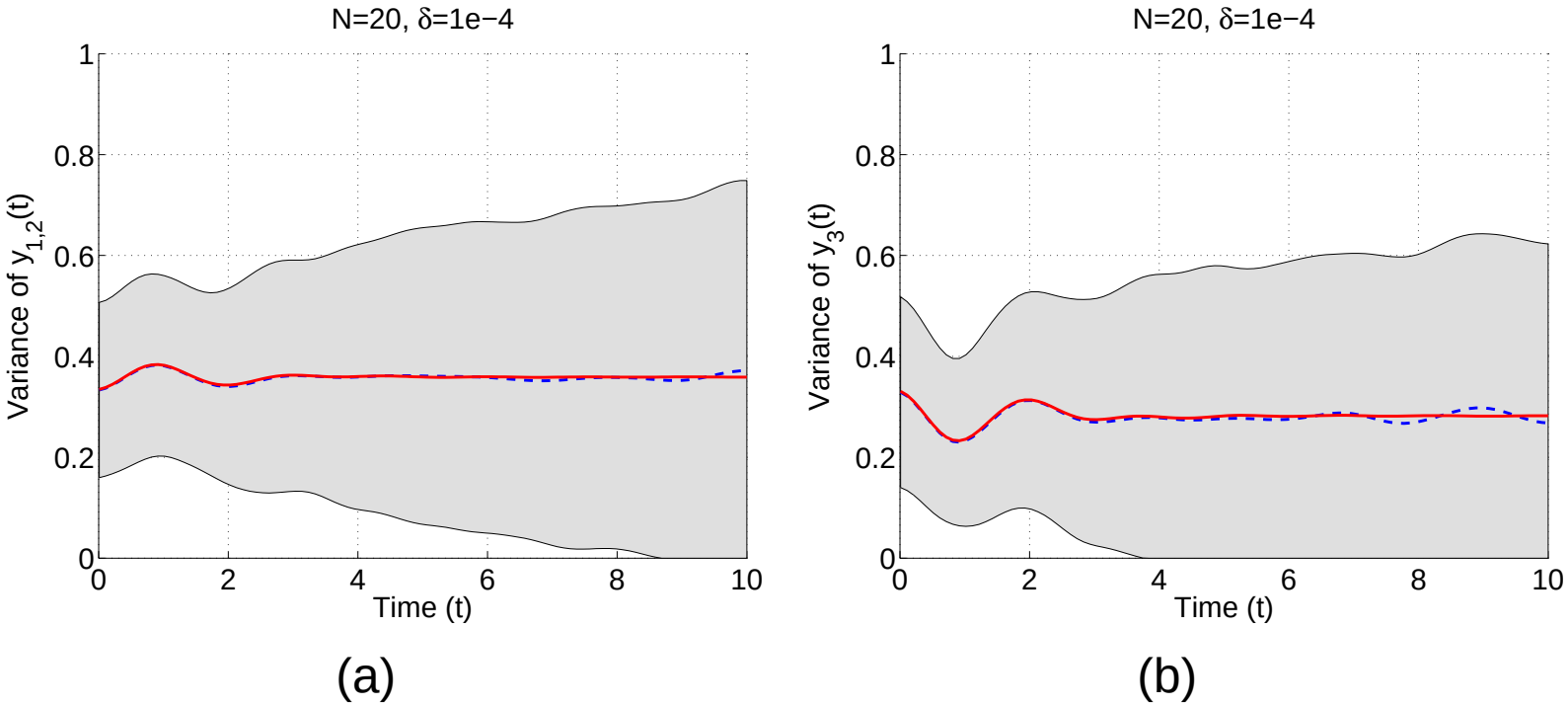


Figure: KO-3: predictive mean (dashed blue) versus the MC estimate (solid red) of the variance of $y_1(t)$ (left, (a)) and $y_2(t)$ (right, (b)) with 95% error bars for tolerances (top to bottom) $\delta = 10^{-4}$.

Numerical Examples: Krainchnan-Orszag

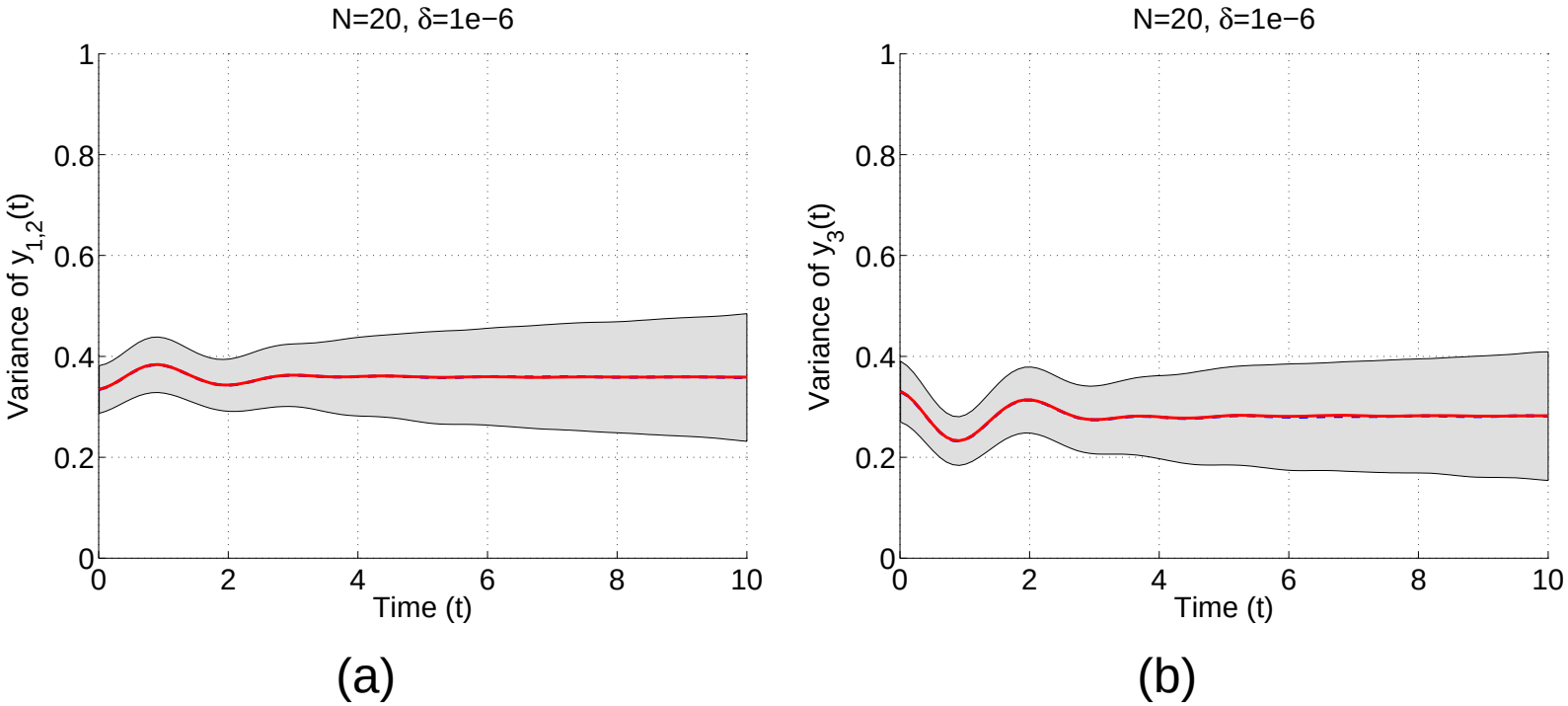


Figure: KO-3: predictive mean (dashed blue) versus the MC estimate (solid red) of the variance of $y_1(t)$ (left, (a)) and $y_2(t)$ (right, (b)) with 95% error bars for tolerances (top to bottom) $\delta = 10^{-6}$.

Numerical Examples: Krainchnan-Orszag

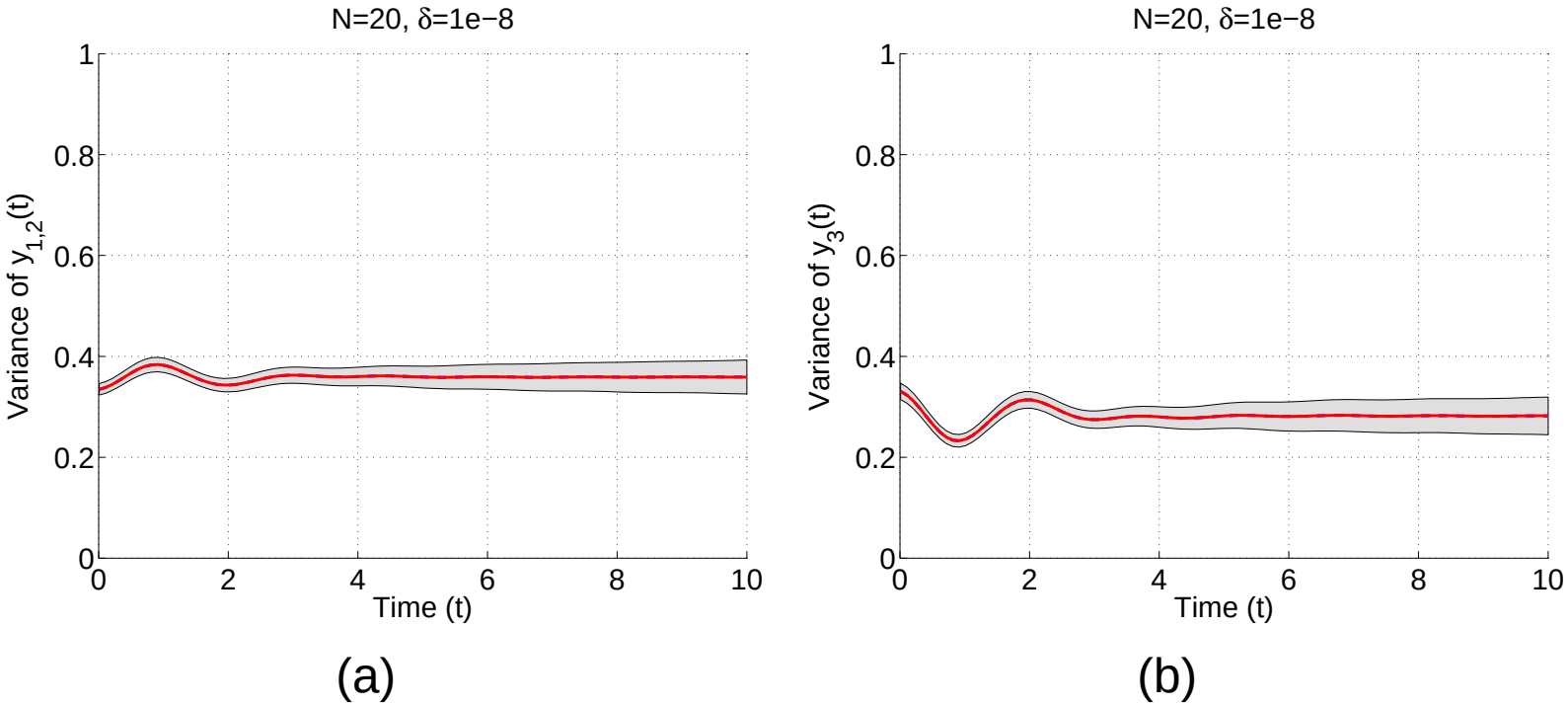
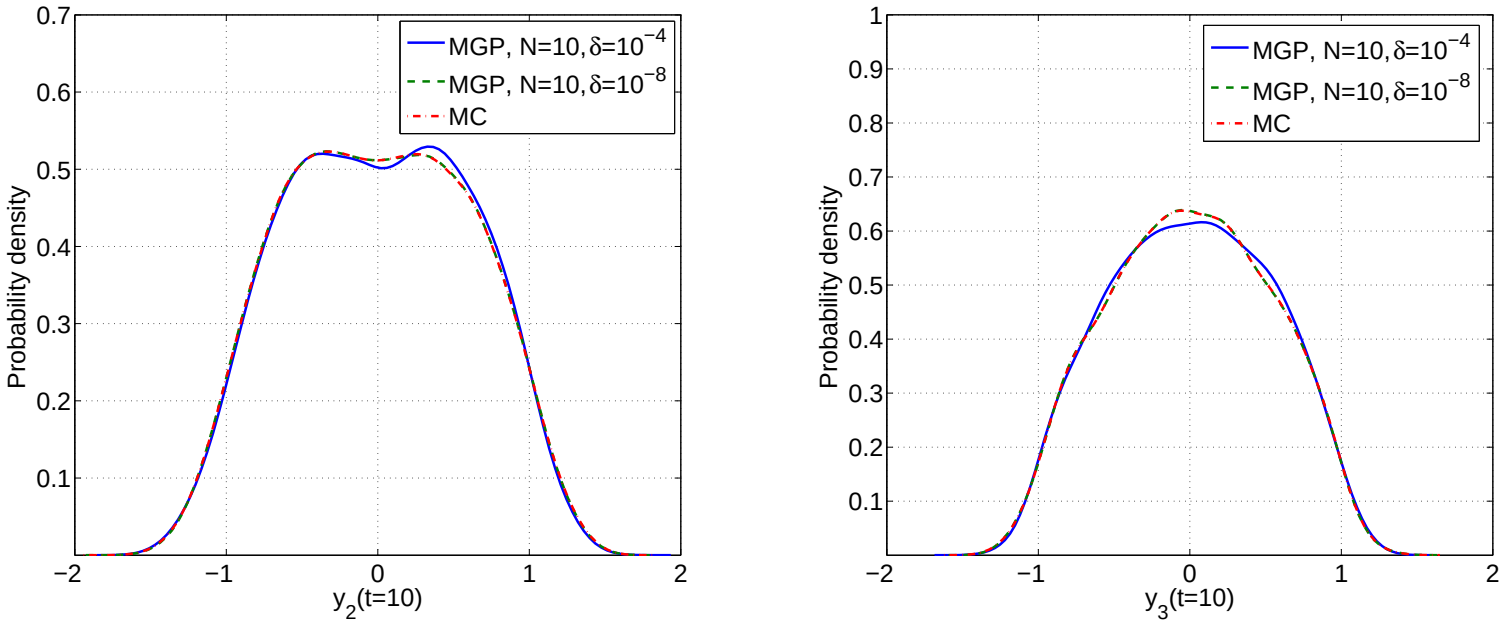


Figure: KO-3: predictive mean (dashed blue) versus the MC estimate (solid red) of the variance of $y_1(t)$ (left, (a)) and $y_2(t)$ (right, (b)) with 95% error bars for tolerances (top to bottom) $\delta = 10^{-8}$.

Numerical Examples: Krainchnan-Orszag



(a) (b)

Figure: KO-3: kernel density estimation of the PDF of y_2 $(t = 10)$ (left) and y_3 $(t = 10)$ (right) using 10^5 samples.

Numerical Examples: Flow through porous media I

Water and oil, ignore gravity effects and capillary forces, assume porosity is a constant.

$$\begin{aligned} \nabla \cdot \mathbf{u} &= 0 \\ \mathbf{u} &= -K(\mathbf{x},\omega)\nabla p, \forall x \in D, \\ \frac{\partial S(\mathbf{x},t,\omega)}{\partial t} + \mathbf{u}\cdot\nabla S(\mathbf{x},t,\omega) &= 0, \forall x \in D, t \in [0,T], \\ p &= \bar{p}, \text{ on } \partial D_p, \quad \mathbf{u}\cdot\mathbf{n} = 0, \text{ on } \partial D_u. \end{aligned}$$

The permability is modeled as follows:

$$G(\mathbf{x},\omega) = \log K(\mathbf{x},\omega),$$

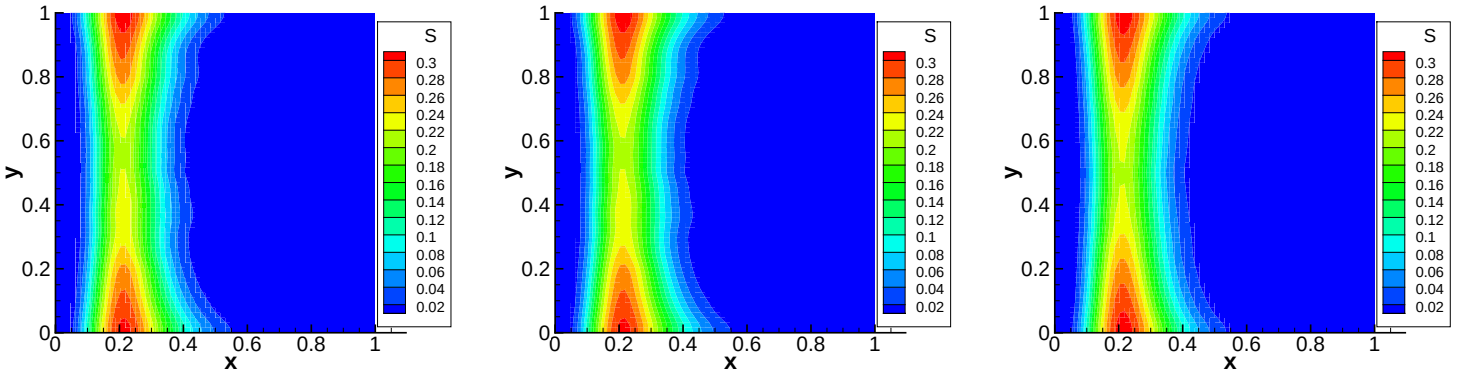
Numerical Examples: Flow through porous media II

where G is a Gaussian random field with exponential covariance function:

$$\text{Cov}[G]((x_1, y_1), (x_2, y_2)) = \exp\left(-\frac{(x_1 - x_2)^2}{L_c^2}\right).$$

In the numerical examples we use a correlation length $L_c = 1$ and take a 33 term KL expansion corresponding to approx. 95% of the field's energy.

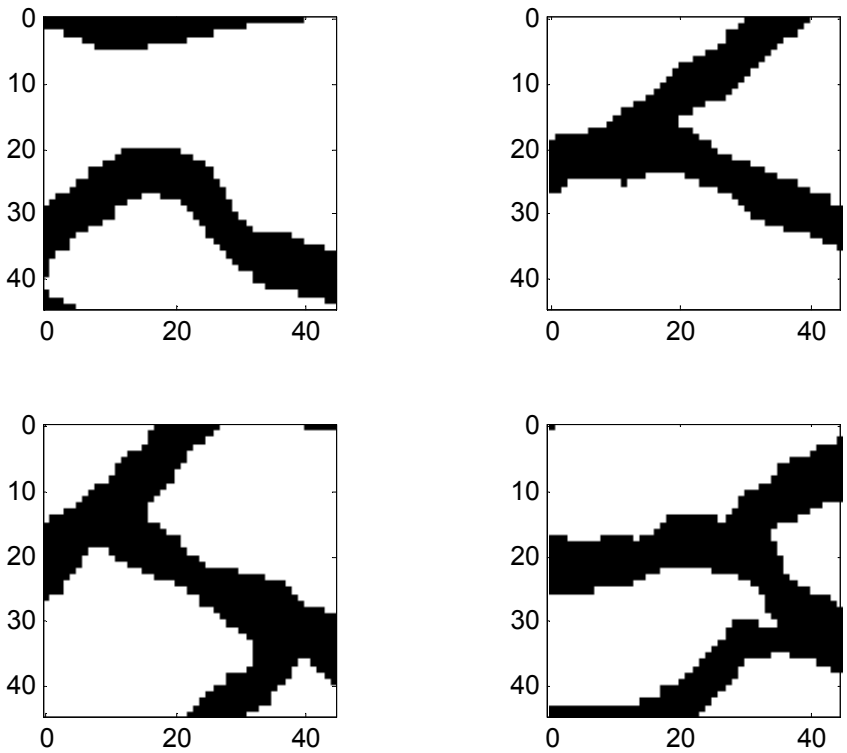
Numerical Examples: Flow through porous media



(a) MGP, $\delta = 10^{-3}$ (b) MGP, $\delta = 10^{-4}$ (c) MC

Figure: Comparison of standard deviations of the saturation S using MGP with approx. 1,000 (a) and 8,000 (b) samples with a MC simulation using 10^6 samples at 0.5 PVI.

Numerical Examples: Flow through porous media (KPCA)



The input permeability field is learnt from data. For details see *Xiang Ma and N. Zabaras, "Kernel Principal Component Analysis for Stochastic Input Model Reduction", Journal of Computational Physics, 2011.*

Numerical Examples: Flow through porous media (KPCA)

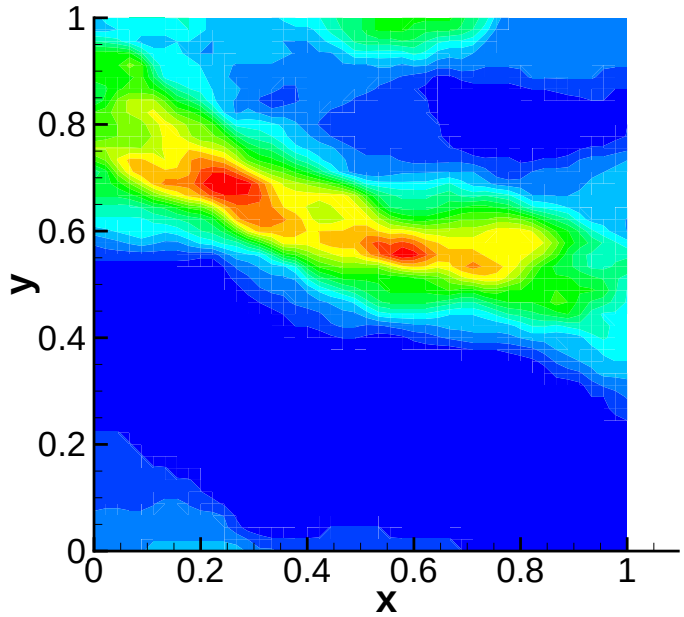
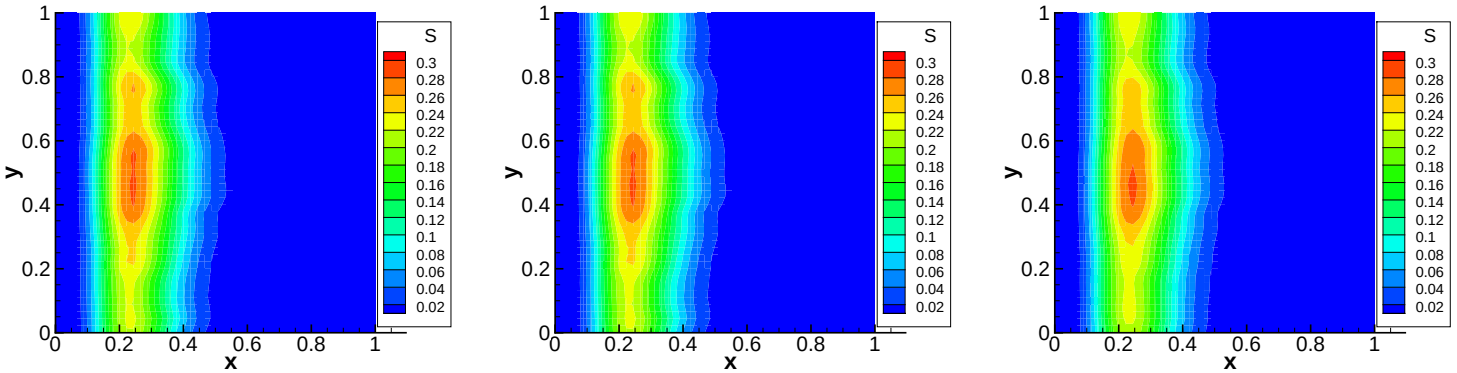


Figure: A sample from the learnt permeability field.

Numerical Examples: Flow through porous media with KPCA input



(a) MGP, $\delta = 10^{-3}$ (b) MGP, $\delta = 10^{-4}$ (c) MC

Figure: Comparison of standard deviations of the saturation S using MGP with approx. 800 (a) and 6,500 (b) samples with a MC simulation using 10^6 samples at 0.2 PVI.

Conclusions

- Faster convergence rate than MC.
- Identification of local features (discontinuities, fast variations).
- Decreased sampling density along unimportant dimensions.
- Semi-analytic calculation of the statistics.
- Fully functioning surrogate model with error bars.
- Easy coupling with HDMR (Anova) as a subroutine (high-dimensions).

On-going research

- Find tighter confidence intervals for the various statistics using the Bayesian variance (few data points).
- Model correlation between outputs (more info out of existing data points $\longrightarrow$ high dimensions).
- Use Bayesian Experimental Design [Mackey (1992)] to select the number observations N on each element (faster convergence, the analogue of p-convergence).
- Hierarchical (Sparse-Grid-like) nested GPs (utilizing more information, smooth prediction along element boundaries).