Spatial logistic Gaussian process for density field modelling: application to stochastic inverse problems

Athénaïs Gautier   David Ginsbourger   Guillaume Pirot

This work has been supported by Swiss National Science Foundation project number 178858

Workshop on Stochastic Simulators
March 11, 2021

(Bayesian) inverse problems

Direct problem: find the response for a given x.

Parameter $x \in D \subset \mathbb{R}^d$ $\longrightarrow$ Forward model $\mathcal{F}$ $\longrightarrow$ Response Y

(Bayesian) inverse problems

Direct problem: find the response for a given x.

$$\underset{x \in D \subset \mathbb{R}^d}{\text{Parameter}} \longrightarrow \text{Forward model } \mathcal{F} \longrightarrow \underset{Y}{\text{Response}}$$

Inverse problem: find the x that generated the observed data y_{ref}.

$$\underset{x \in D \subset \mathbb{R}^d}{\text{Parameter}} \overset{?}{\longleftarrow\cdots\cdots\cdots} \underset{Y = y_{ref}}{\text{Response}}$$

(Bayesian) inverse problems

Direct problem: find the response for a given x.

Parameter $x \in D \subset \mathbb{R}^d$ $\longrightarrow$ Forward model $\mathcal{F}$ $\longrightarrow$ Response Y

Inverse problem: find the x that generated the observed data y_{ref}.

Parameter $x \in D \subset \mathbb{R}^d$ $\leftarrow \cdots ? \cdots$ Response $Y = y_{ref}$

Probabilistic approach:

- Introduce a prior distribution over the parameter: $\pi[x]$
- Use Bayes' theorem: $\pi[x|y_{ref}] \propto \pi[x]\pi[y_{ref}|x]$

(Bayesian) inverse problems

Direct problem: find the response for a given x.

Parameter $x \in D \subset \mathbb{R}^d$ ⟶ Forward model $\mathcal{F}$ ⟶ Response Y

Inverse problem: find the x that generated the observed data y_{ref}.

Parameter $x \in D \subset \mathbb{R}^d$ ⇠ ? ⇠ Response $Y = y_{ref}$

Probabilistic approach:

- Introduce a prior distribution over the parameter: $\pi[x]$
- Use Bayes' theorem: $\pi[x|y_{ref}] \propto \pi[x]\pi[y_{ref}|x]$

$\rightarrow$ requires the likelihood $\pi[y_{ref}|x]$ to get the posterior.

Approximate Bayesian Computation (ABC).

Assuming we can simulate a response Y_x at x and we have a dissimilarity measure $\Delta(y_{ref}, Y_x)$

Approximate Bayesian Computation (ABC).

Assuming we can simulate a response Y_x at x and we have a dissimilarity measure $\Delta(y_{ref}, Y_x)$

The approximation:

$$\pi\left[x|y_{ref}\right] \approx \pi\left[x|\Delta(y_{ref}, Y_x) \leq \epsilon\right] \text{ for } \epsilon \text{ small enough}$$

Approximate Bayesian Computation (ABC).

Assuming we can simulate a response Y_x at x and we have a dissimilarity measure $\Delta(y_{ref}, Y_x)$

The approximation:

$$\pi\left[x|y_{ref}\right] \approx \pi\left[x|\Delta(y_{ref}, Y_x) \leq \epsilon\right] \text{ for } \epsilon \text{ small enough}$$

ABC posterior:

$$\pi\left[x|\Delta(y_{ref}, Y_x) \leq \epsilon\right] \propto \pi[x]\pi\left[\Delta(y_{ref}, Y_x) \leq \epsilon|x\right]$$

Approximate Bayesian Computation (ABC).

Assuming we can simulate a response Y_x at x and we have a dissimilarity measure $\Delta(y_{ref}, Y_x)$

The approximation:

$$\pi\left[x|y_{ref}\right] \approx \pi\left[x|\Delta(y_{ref}, Y_x) \leq \epsilon\right] \text{ for } \epsilon \text{ small enough}$$

ABC posterior:

$$\pi\left[x|\Delta(y_{ref}, Y_x) \leq \epsilon\right] \propto \pi[x]\pi\left[\Delta(y_{ref}, Y_x) \leq \epsilon|x\right]$$

ABC rejection sampling algorithm: Uses simulation to approximate the ABC posterior.

Approximate Bayesian Computation (ABC).

Assuming we can simulate a response Y_x at x and we have a dissimilarity measure $\Delta(y_{ref}, Y_x)$

The approximation:

$$\pi\left[x|y_{ref}\right] \approx \pi\left[x|\Delta(y_{ref}, Y_x) \leq \epsilon\right] \text{ for } \epsilon \text{ small enough}$$

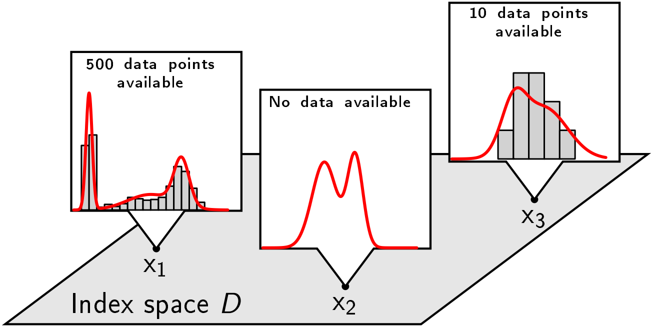
ABC posterior:

$$\pi\left[x|\Delta(y_{ref}, Y_x) \leq \epsilon\right] \propto \pi[x]\pi\left[\Delta(y_{ref}, Y_x) \leq \epsilon|x\right]$$

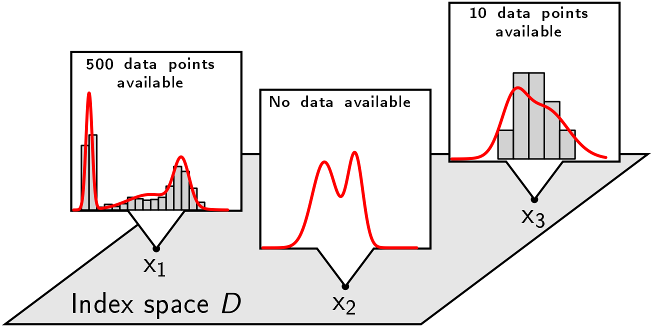
ABC rejection sampling algorithm: Uses simulation to approximate the ABC posterior.

- Requires many simulations, many are discarded.
- The sampling of x must be done with respect to the prior.

Distribution valued fields and ABC



Distribution valued fields and ABC



Some application fields for ABC

- Epidemiology
- Astrophysics and cosmology
- Geosciences

One application in hydrogeology 1/2

Setting: A contaminant, released at depth x (0m to 10m.) spreads through an aquifer with unknown geological structure. Given concentration breakthrough curves, we want to find x.

One application in hydrogeology 1/2

Setting: A contaminant, released at depth x (0m to 10m.) spreads through an aquifer with unknown geological structure. Given concentration breakthrough curves, we want to find x.

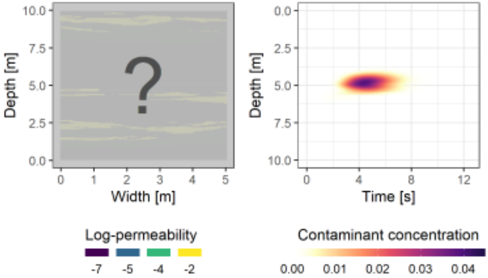


Figure: Reference curves for an unknown geology

One application in hydrogeology 2/2

New simulations for a source at depth x are obtained by:

One application in hydrogeology 2/2

New simulations for a source at depth x are obtained by:

1. Drawing a plausible geological structure.

One application in hydrogeology 2/2

New simulations for a source at depth x are obtained by:

1. Drawing a plausible geological structure.
2. Simulating the response for a source at depth x.

One application in hydrogeology 2/2

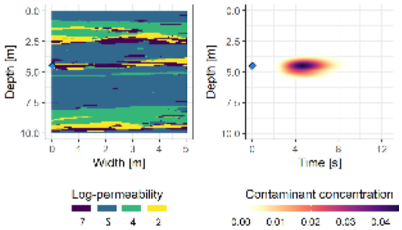
New simulations for a source at depth x are obtained by:

1. Drawing a plausible geological structure.
2. Simulating the response for a source at depth x.
3. Computing the misfit between simulated and reference curves.

One application in hydrogeology 2/2

New simulations for a source at depth x are obtained by:

1. Drawing a plausible geological structure.
2. Simulating the response for a source at depth x.
3. Computing the misfit between simulated and reference curves.

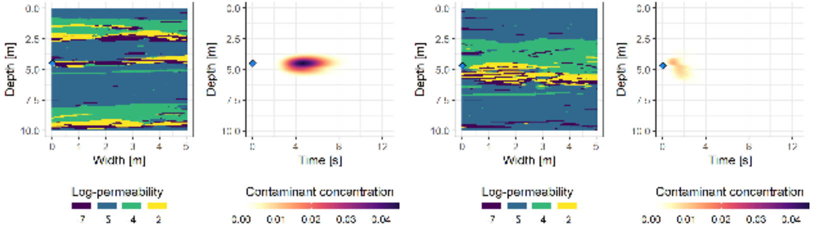


(a) $x = 4.49$m, misfit ≈ 0.09

One application in hydrogeology 2/2

New simulations for a source at depth x are obtained by:

1. Drawing a plausible geological structure.
2. Simulating the response for a source at depth x.
3. Computing the misfit between simulated and reference curves.

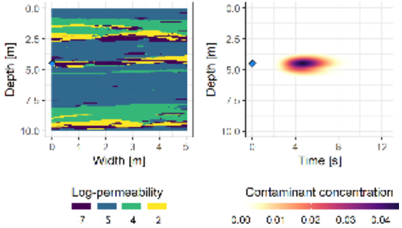


(a) $x = 4.49$m, misfit ≈ 0.09 (b) $x = 4.69$m, misfit ≈ 0.35

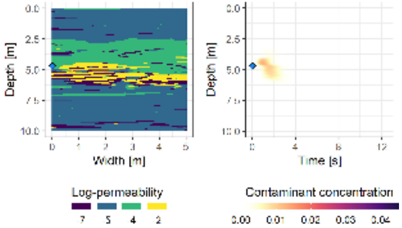
One application in hydrogeology 2/2

New simulations for a source at depth x are obtained by:

1. Drawing a plausible geological structure.
2. Simulating the response for a source at depth x.
3. Computing the misfit between simulated and reference curves.



(a) $x = 4.49$m, misfit ≈ 0.09



(b) $x = 4.69$m, misfit ≈ 0.35

$\rightarrow$ Want to learn the misfit distribution

Requirements

- Suitable for small sample size and arbitrary designs

Requirements

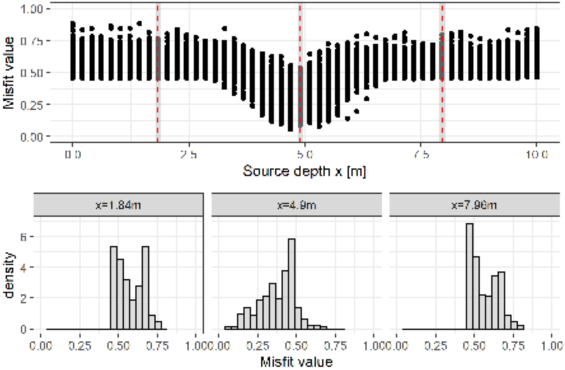
- Suitable for small sample size and arbitrary designs
- Provides uncertainty quantification

Requirements

- Suitable for small sample size and arbitrary designs
- Provides uncertainty quantification
- Accounts for shape (uni VS multi-modality, skewness, etc.)

Requirements

- Suitable for small sample size and arbitrary designs
- Provides uncertainty quantification
- Accounts for shape (uni VS multi-modality, skewness, etc.)



Spatial Logistic Gaussian Process (SLGP)

We assume the target distribution field admits a density at each location x.

Spatial Logistic Gaussian Process (SLGP)

We assume the target distribution field admits a density at each location x.

We consider the hierarchical model:

$$\begin{cases} W \sim \mathcal{GP}(m, k) \text{ on } D \times \mathcal{T} \\ p_x(t) | W = \dfrac{e^{W_{x,t}}}{\int_{\mathcal{T}} e^{W_{x,u}} \, du} \end{cases} \tag{1}$$

Spatial Logistic Gaussian Process (SLGP)

We assume the target distribution field admits a density at each location x.

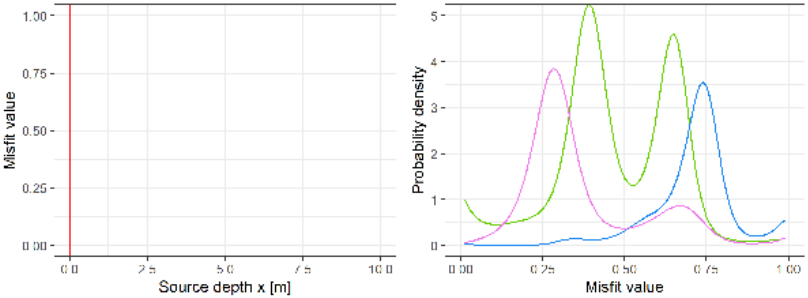
We consider the hierarchical model:

$$\begin{cases} W \sim \mathcal{GP}(m, k) \text{ on } D \times \mathcal{T} \\ p_x(t)|W = \dfrac{e^{W_{x,t}}}{\int_{\mathcal{T}} e^{W_{x,u}}\, du} \end{cases} \tag{1}$$

Under sufficient measurability assumptions on W, the Spatial Logistic Gaussian Process (SLGP) $p_x(t)$ is a random pdf field.

Prior induced by a SLGP[1]

Visualisation of $p_x(t;\omega_1), p_x(t;\omega_2), p_x(t;\omega_3)$, with x varying.



[1] Mean: 0
Kernel: Random Fourier approximation (order 100) of a Matérn 5/2 kernel.
With variance 0.01^2, lengthscale 0.2 in t (range $[0,1]$) and 2 in x (range $[0,10]$)

Some references

- Tom Leonard (1978). "Density Estimation, Stochastic Processes and Prior Information". In: *Journal of the Royal Statistical Society. Series B (Methodological)*. URL: http://www.jstor.org/stable/2984749
- Surya Tokdar and Jayanta K. Ghosh (2007). "Posterior consistency of logistic Gaussian process priors in density estimation". In: *Journal of Statistical Planning and Inference*. DOI: 10.1016/j.jspi.2005.09.005
- Debdeep Pati, David Dunson, and Surya Tokdar (2013). "Posterior consistency in conditional distribution estimation". In: *Journal of Multivariate Analysis*. DOI: 10.1016/j.jmva.2013.01.011
- Athénaïs Gautier, David Ginsbourger, and Guillaume Pirot (2021). *Goal-oriented adaptive sampling under random field modelling of response probability distributions*. arXiv: 2102.07612 [stat.ME]
- Athénaïs Gautier, David Ginsbourger, and Guillaume Pirot (2020). "Probabilistic ABC with Spatial Logistic Gaussian Process modelling". In: *NeurIPS 2020. Third Workshop on Machine Learning and the Physical Sciences*. URL: https://ml4physicalsciences.github.io/2020/files/NeurIPS_ML4PS_2020_112.pdf

Challenge for implementation

The SLGP involves values of W over the whole domain:

$$p_x(t) = \frac{e^{W_{x,t}}}{\int_{\mathcal{T}} e^{W_{x,u}} \, du} \tag{2}$$

Challenge for implementation

The SLGP involves values of W over the whole domain:

$$p_x(t) = \frac{e^{W_{x,t}}}{\int_{\mathcal{T}} e^{W_{x,u}}\, du} \tag{2}$$

We consider only finite rank Gaussian Processes:

$$\forall \mathrm{x} \in D,\ t \in \mathcal{T},\ W_{\mathrm{x},t} = \sum_{j=1}^{p} \sqrt{\lambda_j} f_j(\mathrm{x}, t)\varepsilon_j$$

$p \in \mathbb{N}$, $\lambda_i > 0$, f_i's are functions on $D \times \mathcal{T}$ and ε_i's are i.i.d. $\mathcal{N}(0, 1)$.

Challenge for implementation

The SLGP involves values of W over the whole domain:

$$p_x(t) = \frac{e^{W_{x,t}}}{\int_{\mathcal{T}} e^{W_{x,u}} \, du} \tag{2}$$

We consider only finite rank Gaussian Processes:

$$\forall \mathrm{x} \in D, \ t \in \mathcal{T}, \ W_{\mathrm{x},t} = \sum_{j=1}^{p} \sqrt{\lambda_j} f_j(\mathrm{x}, t) \varepsilon_j$$

$p \in \mathbb{N}$, $\lambda_i > 0$, f_i's are functions on $D \times \mathcal{T}$ and ε_i's are i.i.d. $\mathcal{N}(0, 1)$.
$\rightarrow$ Can approximate the integral with a quadrature scheme

Conditioning on observations

Data: the design $(\mathrm{x}_i)_i \in D^n$, and the corresponding observed data $(t_i)_i \in \mathcal{T}^n$.

Conditioning on observations

Data: the design $(\mathrm{x}_i)_i \in D^n$, and the corresponding observed data $(t_i)_i \in \mathcal{T}^n$.

Assuming the observations stem independently from the model and applying Bayes theorem yields the posterior:

$$\pi(\varepsilon|\mathrm{T}_n) \propto \pi(\varepsilon) \prod_{i=1}^{n} \frac{e^{\sum_{j=1}^{p} \sqrt{\lambda_j} f_j(\mathrm{x}_i, t_i) \varepsilon_j}}{\int_{\mathcal{T}} e^{\sum_{j=1}^{p} \sqrt{\lambda_j} f_j(\mathrm{x}_i, u) \varepsilon_j} \, du} \tag{3}$$

Conditioning on observations

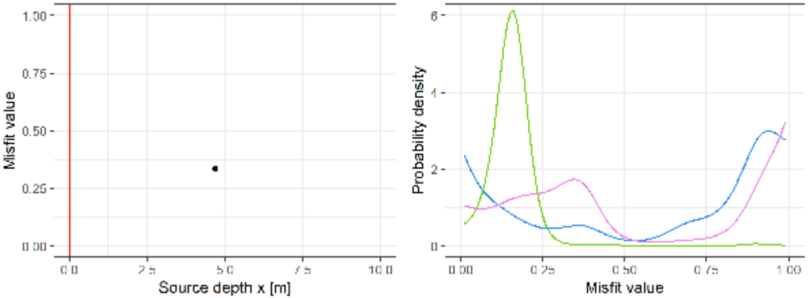
Data: the design $(\mathrm{x}_i)_i \in D^n$, and the corresponding observed data $(t_i)_i \in \mathcal{T}^n$.

Assuming the observations stem independently from the model and applying Bayes theorem yields the posterior:

$$\pi(\varepsilon|\mathrm{T}_n) \propto \pi(\varepsilon) \prod_{i=1}^{n} \frac{e^{\sum_{j=1}^{p} \sqrt{\lambda_j} f_j(\mathrm{x}_i, t_i) \varepsilon_j}}{\int_{\mathcal{T}} e^{\sum_{j=1}^{p} \sqrt{\lambda_j} f_j(\mathrm{x}_i, u) \varepsilon_j} \, du} \tag{3}$$

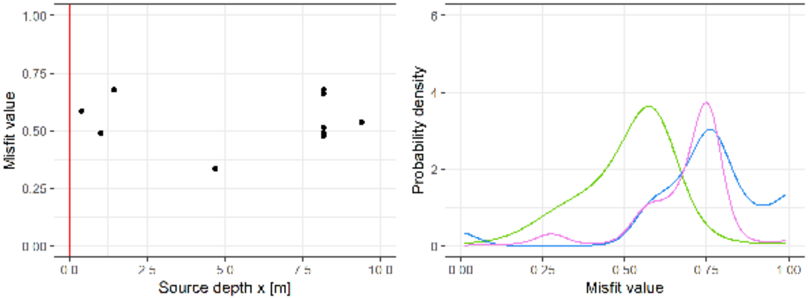
$\rightarrow$ Could be implemented within a MCMC framework.

SLGP[1] conditioned on observations, n=1



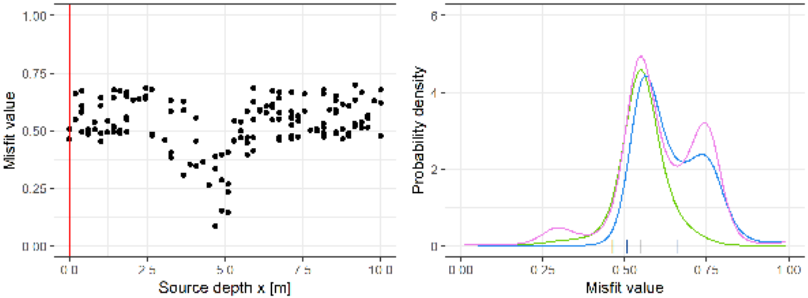
[1] Mean: 0
Kernel: Random Fourier approximation (order 100) of a Matérn 5/2 kernel.
With variance 0.01^2, lengthscale 0.2 in t (range $[0, 1]$) and 2 in x (range $[0, 10]$)

SLGP[1] conditioned on observations, n=10



[1]Mean: 0
Kernel: Random Fourier approximation (order 100) of a Matern 5/2 kernel.
With variance 0.01^2, lengthscale 0.2 in t (range $[0, 1]$) and 2 in x (range $[0, 10]$)

SLGP[1] conditioned on observations, n=150



[1]Mean: 0
Kernel: Random Fourier approximation (order 100) of a Matérn 5/2 kernel.
With variance 0.01^2, lengthscale 0.2 in t (range $[0, 1]$) and 2 in x (range $[0, 10]$)

Back to inverse problems

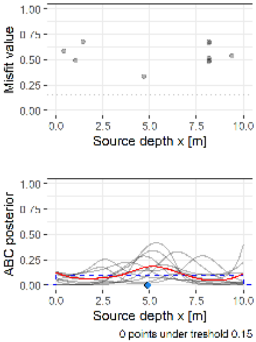
Goal: find the source that yielded our reference curves

Back to inverse problems

Goal: find the source that yielded our reference curves
ABC posterior: $\pi\left[x|\Delta(y_{ref}, Y_x) \leq \epsilon\right] \propto \pi[x]\pi\left[\Delta(y_{ref}, Y_x) \leq \epsilon|x\right]$

Back to inverse problems

Goal: find the source that yielded our reference curves
ABC posterior: $\pi\left[x|\Delta(y_{ref}, Y_x) \leq \epsilon\right] \propto \pi[x]\pi\left[\Delta(y_{ref}, Y_x) \leq \epsilon|x\right]$

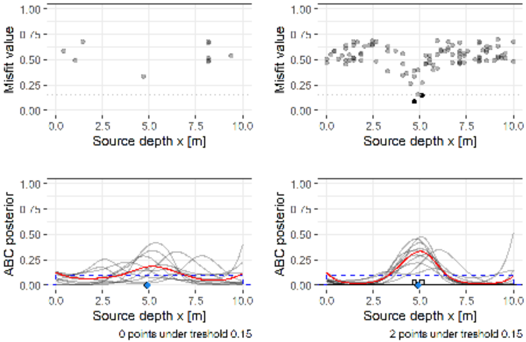


(a) n=10

Back to inverse problems

Goal: find the source that yielded our reference curves

ABC posterior: $\pi\left[x|\Delta(y_{ref}, Y_x) \leq \epsilon\right] \propto \pi[x]\pi\left[\Delta(y_{ref}, Y_x) \leq \epsilon|x\right]$

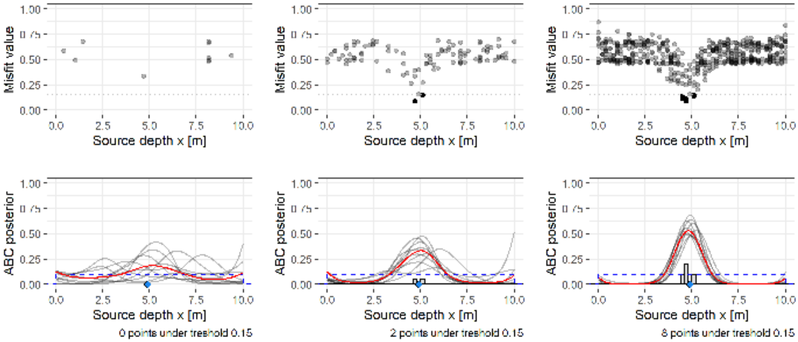


(a) n=10 (b) n=100

Back to inverse problems

Goal: find the source that yielded our reference curves

ABC posterior: $\pi\left[x|\Delta(y_{ref}, Y_x) \leq \epsilon\right] \propto \pi[x]\pi\left[\Delta(y_{ref}, Y_x) \leq \epsilon|x\right]$



(a) n=10 (b) n=100 (c) n=500

Conclusion and upcoming work

In this work we:

- Presented a Bayesian non parametric density field model
- Gave one possible implementation
- Showed applicability on an inverse problem from geosciences

Conclusion and upcoming work

In this work we:

- Presented a Bayesian non parametric density field model
- Gave one possible implementation
- Showed applicability on an inverse problem from geosciences

Upcoming work:

- Quantify performances of the model, benchmark it
- Leverage the probabilistic nature to guide data acquisition
- Study the theoretical properties of the SLGP
- Improve the implementation
- Work on more complex problems

Conclusion and upcoming work

In this work we:

- Presented a Bayesian non parametric density field model
- Gave one possible implementation
- Showed applicability on an inverse problem from geosciences

Upcoming work:

- Quantify performances of the model, benchmark it
- Leverage the probabilistic nature to guide data acquisition
- Study the theoretical properties of the SLGP
- Improve the implementation
- Work on more complex problems

Thank you for your attention!

Random Fourier Features

Assume k is stationary, we approximate k with k_{RFF} defined as:

$$k_{RFF}(x,y) := \frac{\sigma^2}{2\pi p}\sum_{i=1}^{p}\cos(\omega_i^T x + u_i)\cos(\omega_i^T y + u_i) \qquad (4)$$

where $\sigma^2 = k(0,0)$, the ω_i's are draws of independent random variables that have a density equal to spectral density associated to k and the u_i's are draws of i.i.d. $\mathcal{U}(0, 2\pi)$.

Random Fourier Features

Assume k is stationary, we approximate k with k_{RFF} defined as:

$$k_{RFF}(x, y) := \frac{\sigma^2}{2\pi p} \sum_{i=1}^{p} \cos(\omega_i^T x + u_i) \cos(\omega_i^T y + u_i) \qquad (4)$$

where $\sigma^2 = k(0, 0)$, the ω_i's are draws of independent random variables that have a density equal to spectral density associated to k and the u_i's are draws of i.i.d. $\mathcal{U}(0, 2\pi)$.
It satisfies $\mathbb{E}[k_{RFF}] = k$
$W_{RFF}(x) := \frac{\sigma}{\sqrt{2\pi p}} \sum_{i=1}^{p} \epsilon_i \cos(\omega_i^T x + u_i)$ where ϵ_i are i.i.d. $\mathcal{N}(0, 1)$, is a $\mathrm{GP}(0, k_{RFF}(x, y))$.

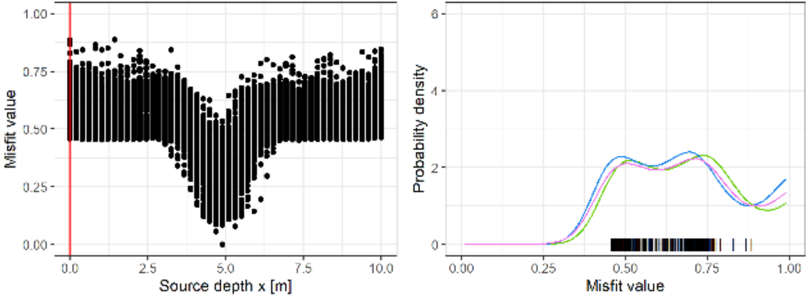
Estimating the hyperparameters of the GP

Assume k depends on two types of parameters, a variance parameter $\sigma^2 > 0$ and a some strictly positive length-scale θ:

$$\forall x, y \in (D \times \mathcal{T})^2,\ k(x, y) = \sigma^2 k_0(D_{1/\theta}(x - y)) \tag{5}$$

where $D_{1/\theta}$ is the diagonal matrix with diagonal $1/\theta$.
Setting a prior distribution Λ over (σ^2, θ) enables us to do a Bayesian estimation of the parameters.

SLGP[1] conditioned on the full data set, n=10000



[1] Mean: 0
Kernel: Random Fourier approximation (order 100) of a Matérn 5/2 kernel.
With variance and lengthscale sampled

Fit on the full data set: parameters estimated