Leveraging spatial regularity in distribution field estimation to accelerate statistical inference for the sciences.

Athénaïs Gautier

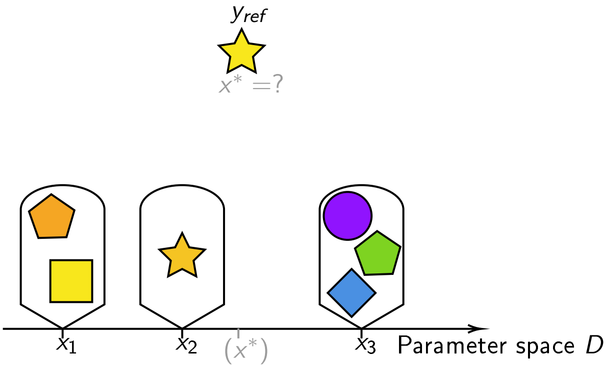
IMSV, University of Bern, Switzerland.
Ph.D. supervisor: David Ginsbourger
https://athenaisgautier.github.io/

MASCOT-NUM 2022, Clermont-Ferrand
Tuesday, 07 June 2022

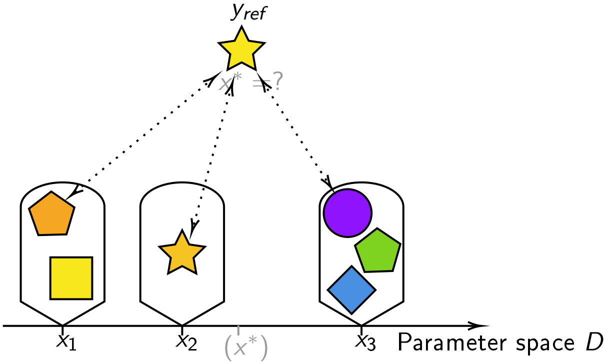
* Funded by : Swiss National Science Foundation project number 178858

Stochastic inverse problems and ABC... Without maths!

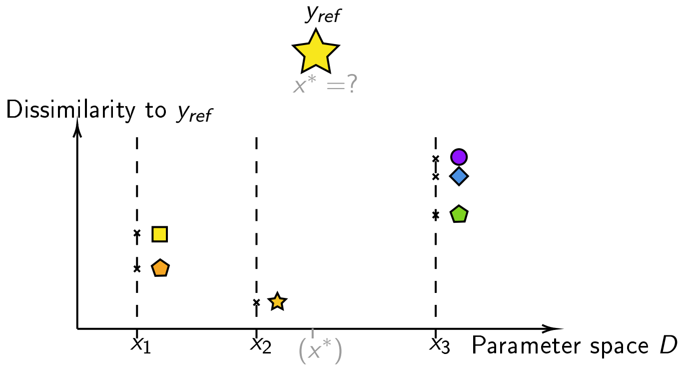
Stochastic inverse problems and ABC... Without maths!



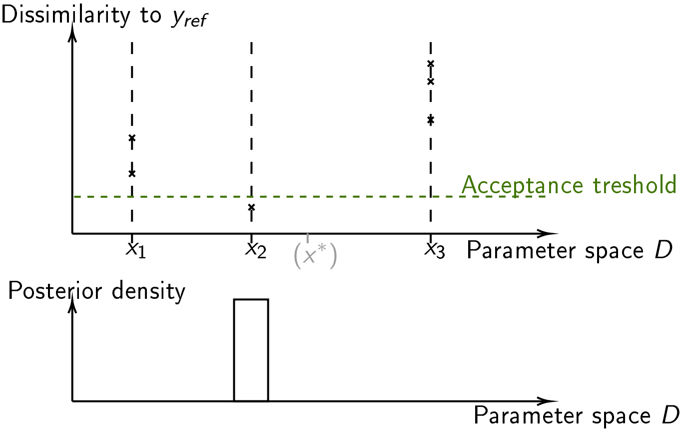
Stochastic inverse problems and ABC... Without maths!



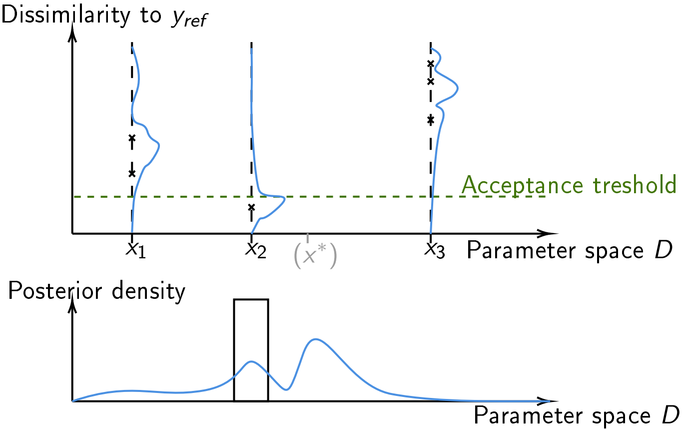
Stochastic inverse problems and ABC... Without maths!



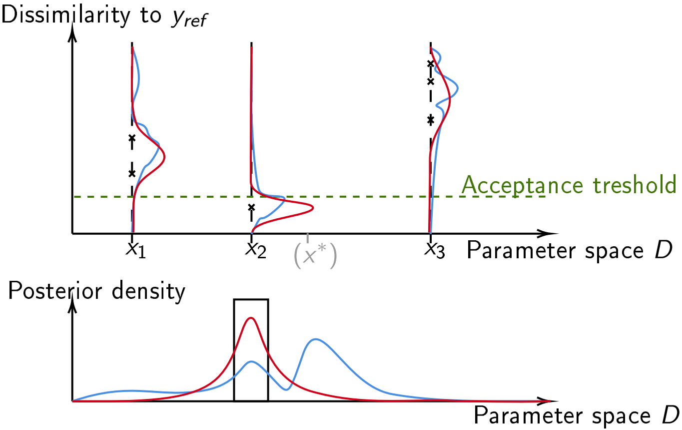
Stochastic inverse problems and ABC... Without maths!



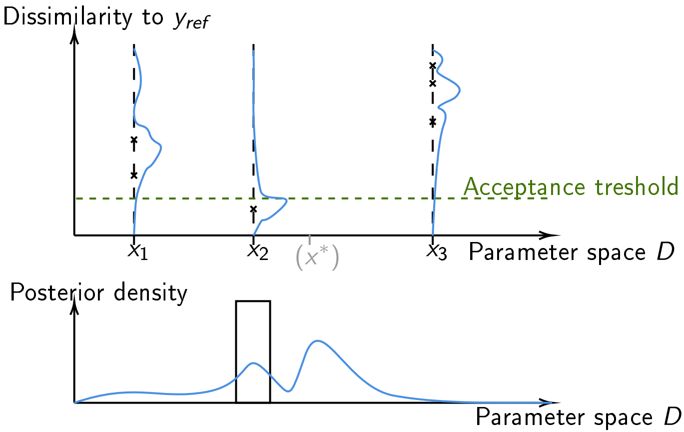
Stochastic inverse problems and ABC... Without maths!



Stochastic inverse problems and ABC... Without maths!



Stochastic inverse problems and ABC... Without maths!



Should we learn the misfit distribution?

Goal: Estimating spatially dependent probability distributions

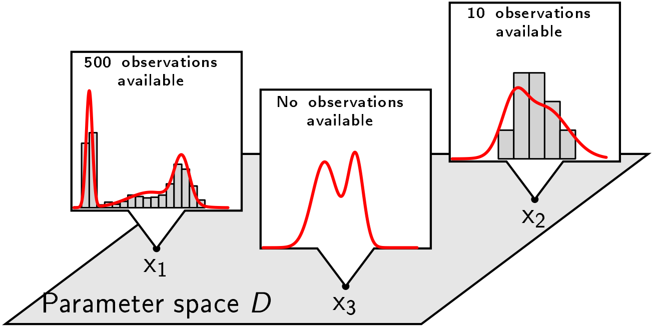


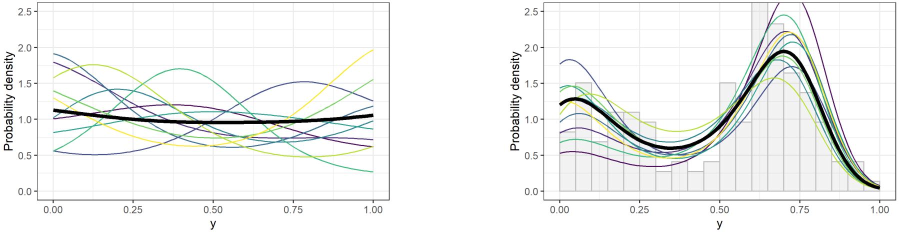
Figure: pdf (red curve) versus collected data (histogram)

- For various shapes or modalities
- Samples heterogeneously scattered across space + no replications.
- Uncertainty Quantification.

Building brick: density estimation with a LGP 1/2

Logistic Gaussian Process (LGP)

$$\begin{cases} (Z_y)_{y \in \mathcal{T}} \sim \mathcal{GP} \\ \frac{e^{Z_y}}{\int_{\mathcal{T}} e^{Z_u}\, du} \sim \mathcal{LGP} \end{cases}$$



(a) Prior of the LGP

(b) Posterior of the LGP

Figure: Sample paths (colours) and mean (black) of a LGP.

[1] Mean: 0; Kernel: Matérn 5/2 variance 1, lengthscale 0.1 in y.

Building brick: density estimation with a LGP 2/2

Main references on the LGP

Leonard, Tom (1978). "Density estimation, stochastic processes and prior information". In: *Journal of the Royal Statistical Society: Series B (Methodological)* 40.2, pp. 113–132.

Lenk, Peter J (1991). "Towards a practicable Bayesian nonparametric density estimator". In: *Biometrika* 78.3, pp. 531–543.

Tokdar, Surya T and Jayanta K Ghosh (2007). "Posterior consistency of logistic Gaussian process priors in density estimation". In: *Journal of statistical planning and inference* 137.1, pp. 34–42.

Vaart, Aad W van der and J Harry van Zanten (2008). "Rates of contraction of posterior distributions based on Gaussian process priors". In: *The Annals of Statistics* 36.3, pp. 1435–1463.

Gautier, Athénaïs and David Ginsbourger (2021). *Continuous logistic Gaussian random measure fields for spatial distributional modelling*. arXiv: 2110.02876 [math.ST].

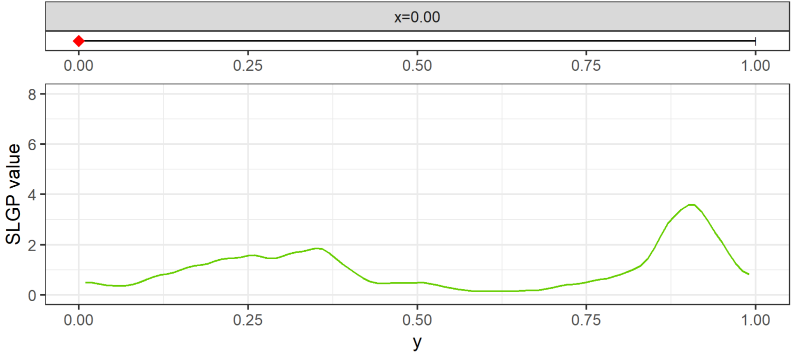
Spatial extension of the LGP

Spatial Logistic Gaussian Process (SLGP)

$$\begin{cases} (Z_{x,y})_{x \in D,\ y \in \mathcal{T}} \sim \mathcal{GP} \\ W_{x,y} := \dfrac{e^{Z_{x,y}}}{\int_{\mathcal{T}} e^{Z_{x,u}}\, du} \sim \mathcal{SLGP} \end{cases}$$

A look at the prior induced by the SLGP

Drawing one realisation of the SLGP prior, and plotting it for varying x.

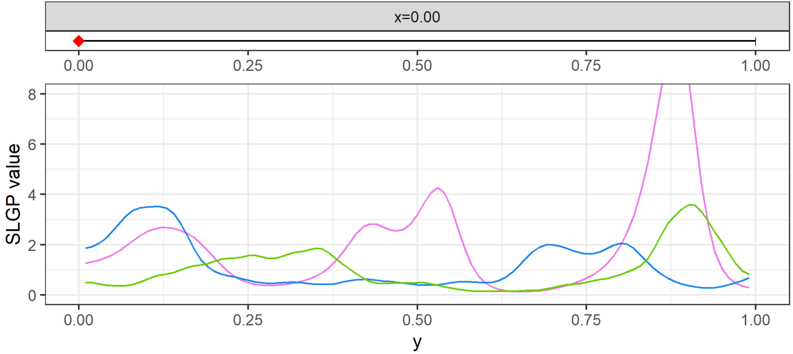


▸▸ Not showing the animation ? Click here!

[1] Mean: 0; Kernel: Matérn 5/2 variance 1, lengthscale 0.1 in x et y.

A look at the prior induced by the SLGP

Drawing three realisations of the SLGP prior, and plotting them for varying x.



Not showing the animation ? Click here!

[1]Mean: 0; Kernel: Matérn 5/2 variance 1, lengthscale 0.1 in x et y.

What is the SLGP ?

$$\begin{cases} (Z_{x,y})_{x \in D \;\; y \in \mathcal{T}} \sim \mathcal{GP} \\ W_{x,y} := \frac{e^{Z_{x,y}}}{\int_{\mathcal{T}} e^{Z_{x,u}} \, du} \sim \mathcal{SLGP} \end{cases}$$

Gautier, Athénaïs and David Ginsbourger (2021). *Continuous logistic Gaussian random measure fields for spatial distributional modelling*. arXiv: 2110.02876 [math.ST].

- Random measure at each x $\rightarrow$ Random field of measures.
- Distribution characterized by $Z_{x,y} - Z_{x,y'} = \log W_{x,y} - \log W_{x,y'}$ $(x \in D, y, y' \in \mathcal{T})$.

Spatial regularity of the SLGP

Spatial regularity of the SLGP

For W a SLGP, if it satisfies for all $(\mathrm{x}, \mathrm{x}') \in D^2, (y, y') \in \mathcal{T}^2$:

$$\mathbb{V}\mathrm{ar}\left[\log W_{\mathrm{x},y} - \log W_{\mathrm{x}',y'}\right] \leq C \cdot \max(\|\mathrm{x}-\mathrm{x}'\|_\infty^{\alpha_1}, \|y-y'\|_\infty^{\alpha_2}) \tag{1}$$

for constants $C, \alpha_1, \alpha_2 > 0$, then for all $\gamma > 0, \delta > 0$, there exists a constant $K_{\gamma,\delta}$ such that for all $\mathrm{x}, \mathrm{x}' \in D^2$

$$\begin{aligned} \mathbb{E}\left[d_H(W_{\mathrm{x},\cdot}, W_{\mathrm{x}',\cdot})^\gamma\right] &\leq K_{\gamma,\delta}\|\mathrm{x}-\mathrm{x}'\|_\infty^{\gamma\alpha_1/2-\delta} \\ \mathbb{E}\left[KL(W_{\mathrm{x},\cdot}, W_{\mathrm{x}',\cdot})^\gamma\right] &\leq K_{\gamma,\delta}\|\mathrm{x}-\mathrm{x}'\|_\infty^{\gamma\alpha_1-\delta} \\ \mathbb{E}\left[d_{TV}(W_{\mathrm{x},\cdot}, W_{\mathrm{x}',\cdot})^\gamma\right] &\leq K_{\gamma,\delta}\|\mathrm{x}-\mathrm{x}'\|_\infty^{\gamma\alpha_1-\delta} \end{aligned} \tag{2}$$

where d_H is the Hellinger distance, KL the Kulback-Leibler divergence, d_{TV} the total variation distance.

Posterior consistency of the SLGP

Result adapted from:

Pati, Debdeep, David B Dunson, and Surya T Tokdar (2013a). "Posterior consistency in conditional distribution estimation". In: *Journal of multivariate analysis* 116, pp. 456–472.

Posterior consistency

If the model is well-specified, then we achieve posterior consistency.

Important aspects of the implementation

$$\begin{cases} (Z_{x,y})_{x \in D \ y \in \mathcal{T}} \sim \mathcal{GP} \\ W_{x,y} := \frac{e^{Z_{x,y}}}{\int_{\mathcal{T}} e^{Z_{x,u}} du} \sim \mathcal{SLGP} \end{cases}$$

We focus on *finite-rank* GPs $Z_{x,y} = \sum_{j=1}^{p} f_j(x, y)\varepsilon_j$

Considered functions

In the upcoming R package:

- Multivariate discrete Fourier Features
- Random Fourier Features
- Space-filling Random Fourier Features
- Inducing points
- User-defined functions (!)

Conditioning on data

Conditioning on data

For locations $(x_i)_i \in D^n$, observations $(y_i)_{i=1}^n \in \mathcal{T}^n$ and a finite rank SLGP, assuming independent Y_i's we have the posterior:

$$\pi(\varepsilon|T_n) \propto \pi(\varepsilon) \prod_{i=1}^{n} \frac{e^{\sum_{j=1}^{p} f_j(x_i,y_i)\varepsilon_j}}{\int_{\mathcal{T}} e^{\sum_{j=1}^{p} f_j(x_i,u)\varepsilon_j}\,du} \tag{3}$$

This posterior is **log-concave**.

- Fast optimization (MAP)
- fast MCMC sampling.[1]

[1] Dwivedi et al. 2018.

SLGP: pros and cons.

The SLGP: $\dfrac{e^{Z_{x,y}}}{\int_{\mathcal{T}} e^{Z_{x,u}}\, du}$

Pros:

- Flexible model, non parametric
- Can be conditioned on data (with little to no replications)
- Bayesian: probabilistic response (UQ + sequential)
- Encoding prior knowledge in the GP

Cons:

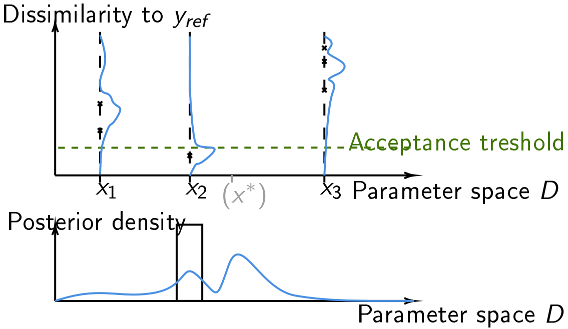
- Computationally expensive

SLGP-ABC

Idea presented in:

Gautier, Athénaïs, 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*.

Models directly the misfit distribution:



SLGP-ABC

Idea presented in:

Gautier, Athénaïs, 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*.

Main elements about SLGP-ABC

- Works in low data regime.
- Little assumption on the misfit likelihood.
- Can use any data-set and any prior on x.
- Posterior consistency for the ABC-posterior.

Distribution field estimation & speeding-up inference
A. Gautier
Motivation
SLGP
LGP
SLGP
Theory
Implementation
Application
SLGP-ABC
Application
Sequential
Criteria used
Application

An hydrogeological application: setup

- Contaminant released at depth x (0 to 10m.), aquifer with unknown structure
- We observe a contaminant plum and want to recover x.[2]

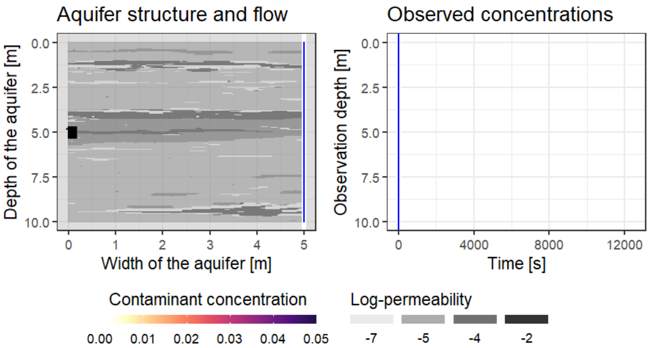
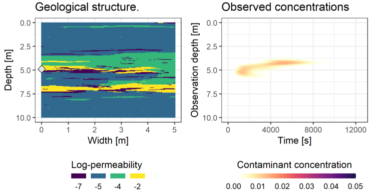


Figure: Flow through the aquifer [left panel, **not observed**], contaminant concentration at 100 boreholes [right panel, **observed**]

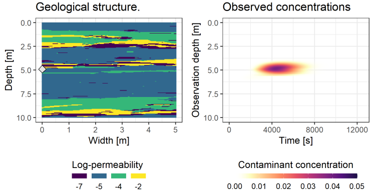
Not showing the animation ? Click here!

An hydrogeology application: simulations

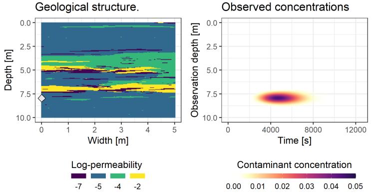
New simulations and computed misfits at a source depth x.



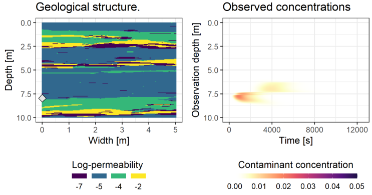
(a) $x = 4.90$m, misfit ≈ 0.035



(b) $x = 4.90$m, misfit ≈ 0.35



(c) $x = 7.96$m, misfit ≈ 0.63



(d) $x = 7.96$m, misfit ≈ 0.22

An hydrogeology application: misfits

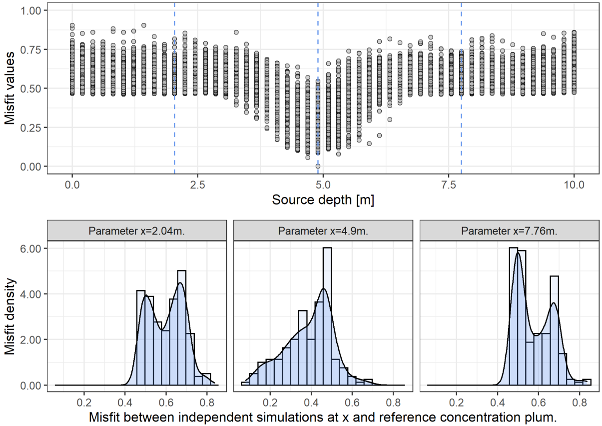


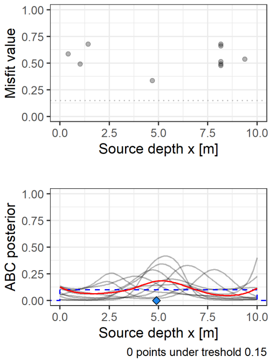
Figure: 10000 misfits obtained when running simulations for **200 geological structures** and **50 source depths**.

SLGP-ABC for hydrogeology.

Setting: Find the source that yielded reference curves, ABC-threshold of 0.15 (1% of simulations below it)

SLGP-ABC for hydrogeology.

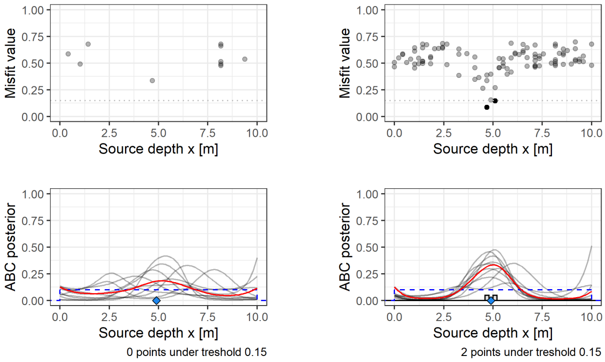
Setting: Find the source that yielded reference curves, ABC-threshold of 0.15 (1% of simulations below it)



(a) n=10

SLGP-ABC for hydrogeology.

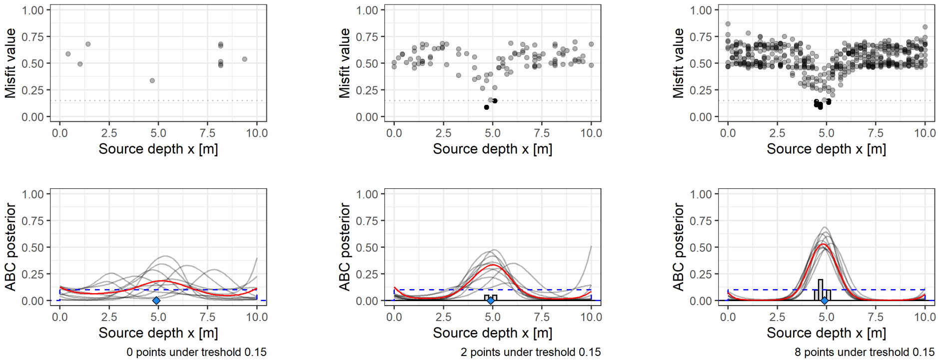
Setting: Find the source that yielded reference curves, ABC-threshold of 0.15 (1% of simulations below it)



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

SLGP-ABC for hydrogeology.

Setting: Find the source that yielded reference curves, ABC-threshold of 0.15 (1% of simulations below it)



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

Sequential designs within GP-ABC

Inspired from batch methods in GP-ABC:

Järvenpää, Marko, Aki Vehtari, and Pekka Marttinen (2019). "Batch simulations and uncertainty quantification in Gaussian process surrogate-based approximate Bayesian computation". In: *arXiv preprint arXiv:1910.06121*.

- Current model trained on n simulations.
- Want to add k simulations
- Reduce the SLGP-ABC posterior's uncertainty

Two criteria used

Minimize the "future uncertainty" for adding k simulations at x*:

- **Random posterior at x**

$$\hat{\pi}_{SLGP-ABC}(x)$$

Two criteria used

Minimize the "future uncertainty" for adding k simulations at x*:

- **Variance**

$$\mathrm{Var}_{\mathrm{SLGP}|\mathcal{D}_{\mathbf{1}:n}\cup\mathcal{D}^*}\left(\hat{\pi}_{SLGP-ABC}(x)\right)$$

- **Median Interquartile Range**

$$\mathrm{IQR}_{\mathrm{SLGP}|\mathcal{D}_{\mathbf{1}:n}\cup\mathcal{D}^*}\left(\hat{\pi}_{SLGP-ABC}(x)\right)$$

where IQR is the inter-quartile range.

Two criteria used

Minimize the "future uncertainty" for adding k simulations at x*:

- **Integrated Variance**

$$\int_X \mathrm{Var}_{\mathrm{SLGP}|\mathcal{D}_{\mathbf{1}:n}\cup\mathcal{D}^*}\left(\hat{\pi}_{SLGP-ABC}(x)\right)\,dx$$

- **Integrated Median Interquartile Range**

$$\int_X \mathrm{IQR}_{\mathrm{SLGP}|\mathcal{D}_{\mathbf{1}:n}\cup\mathcal{D}^*}\left(\hat{\pi}_{SLGP-ABC}(x)\right)\,dx$$

where IQR is the inter-quartile range.

Two criteria used

Minimize the “future uncertainty” for adding k simulations at x^*:

- **Expected Integrated Variance (EIV)**

$$L_{EIV}(x^*) := \mathbb{E}_{y^*|x^*}\left[\int_X \mathrm{Var}_{SLGP|\mathcal{D}_{1:n}\cup\mathcal{D}^*}\left(\hat{\pi}_{SLGP-ABC}(x)\right) dx\right]$$

- **Integrated Median Interquartile Range (IMIQR)**

$$L_{IMIQR}(x^*) := \mathbb{E}_{y^*|x^*}\left[\int_X \mathrm{IQR}_{SLGP|\mathcal{D}_{1:n}\cup\mathcal{D}^*}\left(\hat{\pi}_{SLGP-ABC}(x)\right) dx\right]$$

where IQR is the inter-quartile range.

Computing criteria by simulation

No closed-form formula $\rightarrow$ simulations.

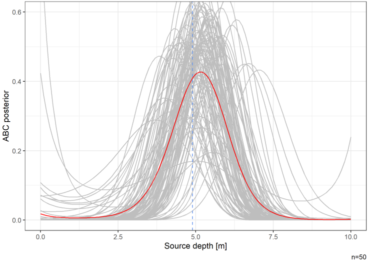
A reference with simulation for computing criteria in SLGP-based Bayesian Optimisation:

Gautier, Athénaïs, Ginsbourger, David, and Pirot, Guillaume (2021). "Goal-oriented adaptive sampling under random field modelling of response probability distributions". In: *ESAIM: ProcS* 71, pp. 89–100.

Go to the pseudo-algorithm we use

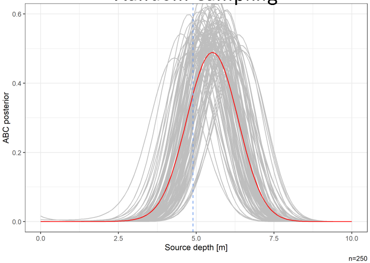
A qualitative look at the uncertainty on the SLGP-ABC posterior

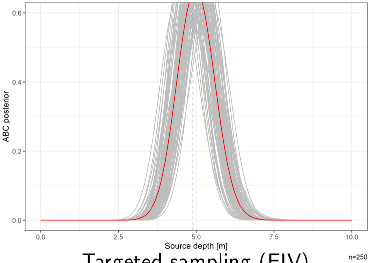
Same initial design ($n = 50$)



Adding 8 batches of 25

Random sampling





Targeted sampling (EIV)

Conclusion and upcoming work

What we saw:

- Leverage random fields for density field estimation
- Application to inverse problems: accelerate inference in low data regime
- Application to inverse problems: guide data acquisition

Perspectives and related works:

- Continuing implementation and package release
- Evaluating predictions (look into probabilistic forecasting for scoring of fields of density)
- Benchmark the approach (compare to other estimation methods, use other test cases)

Thank you for your attention!

Questions?

Using simulations to compute the criterions

Data: $x^*,\ k \geq 1,\ \mathcal{D}_{1:n},\ (\varepsilon^{(i)})_{1\leq i\leq M},\ n_{sim}$
Result: $\hat{L}(x^*)$
Use $\sum_{i=1}^{M} \frac{1}{M}\varepsilon^{(i)}$ as approx. of $\pi[\varepsilon|\mathcal{D}_{1:n}]$;
for $i \leftarrow 1$ **to** n_{sim} **do**

Generate k new observations at x^* from the current mean field: $y_1^*, ..., y_k^*$;

Reweight the $\varepsilon^{(i)}$: $\omega_i \propto \prod_{i=1}^{k} \frac{e^{\sum_{j=1}^{p} f_j(x^*, y_i^*)\varepsilon_j}}{\int_{\mathcal{T}} e^{\sum_{j=1}^{p} f_j(x^*, u)\varepsilon_j}\, du}$;

Use $\sum_{i=1}^{M} \omega_i \varepsilon^{(i)}$ as approx. of $\pi[\varepsilon|\mathcal{D}_{1:n} \cup \mathcal{D}^*]$;

Compute and store the criterion of interest;

end
Average the stored values of the criterion.

Go back to the main slide

Proper scoring rule for (random) pdf

Positively oriented **proper scoring rule**, inspired from probabilistic forecasting:

Steinwart, Ingo and Johanna F Ziegel (2021). "Strictly proper kernel scores and characteristic kernels on compact spaces". In: *Applied and Computational Harmonic Analysis* 51, pp. 510–542.

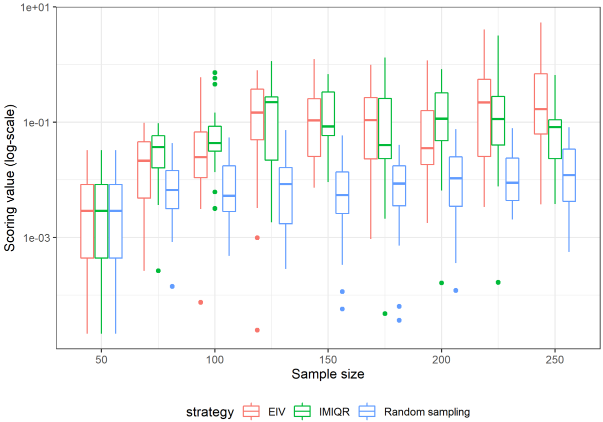
Compare random distribution Π (SLGP-ABC posterior) to the reference p_0 (true source):

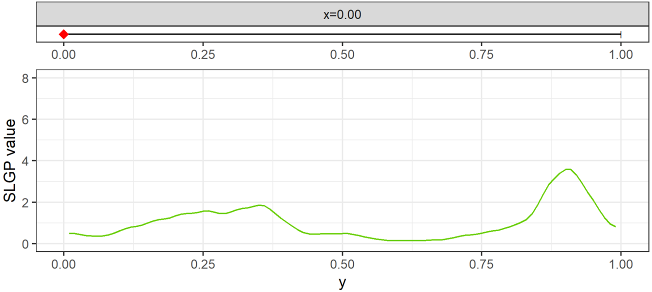
$$S(\Pi, p_0) := -\mathbb{E}_{P\sim\Pi}\left[\gamma(P, p_0)\right] + \frac{1}{2}\mathbb{E}_{P,P'\sim\Pi}\left[\gamma(P, P')\right] \tag{4}$$

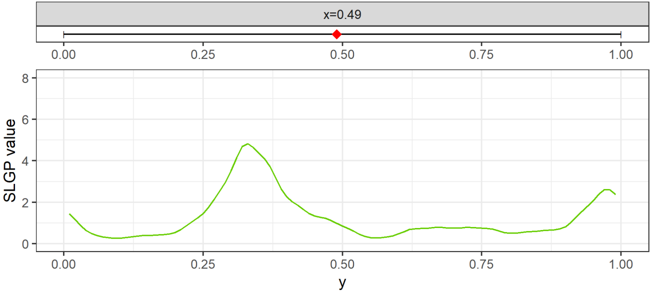
where $\gamma(P, P') := \mathbb{E}\left[(X - X')^2\right] + \mathbb{E}'\left[(Y - Y')^2\right] - 2\mathbb{E}\left[(X - Y)^2\right]$ with $X, X' \sim P$ and $Y, Y' \sim P'$.

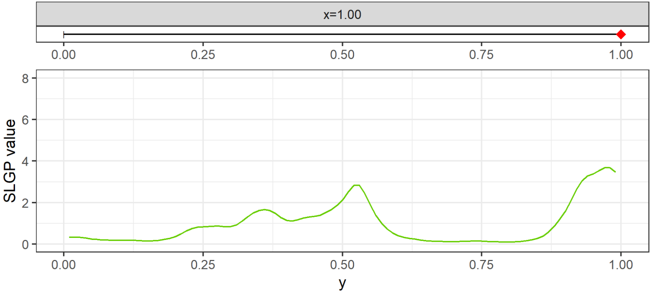
Scoring (random) densities

Is adaptive SLGP-ABC better than random-ABC ? Using a (positively oriented) **proper scoring rule**:

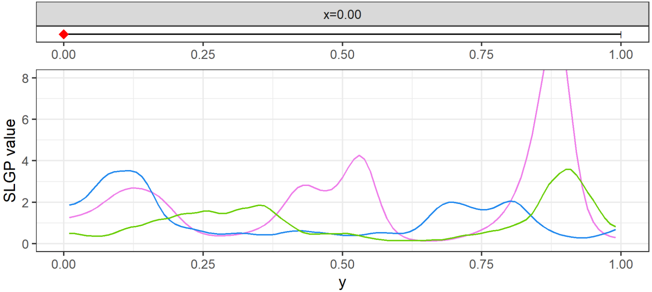


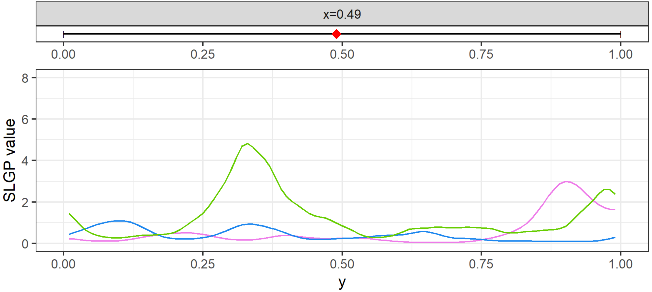

x=0.00
0.00
0.25
0.50
0.75
1.00
SLGP value
8
6
4
2
0
0.00
0.25
0.50
0.75
1.00
y

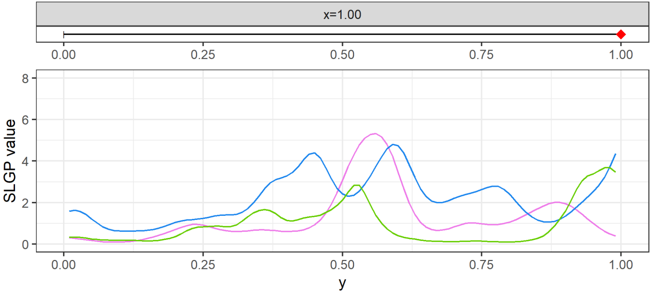

x=0.49
0.00
0.25
0.50
0.75
1.00
SLGP value
8
6
4
2
0
y



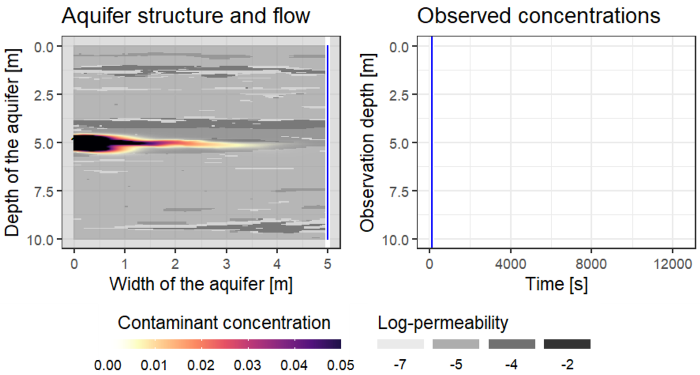
Go back to the main slide

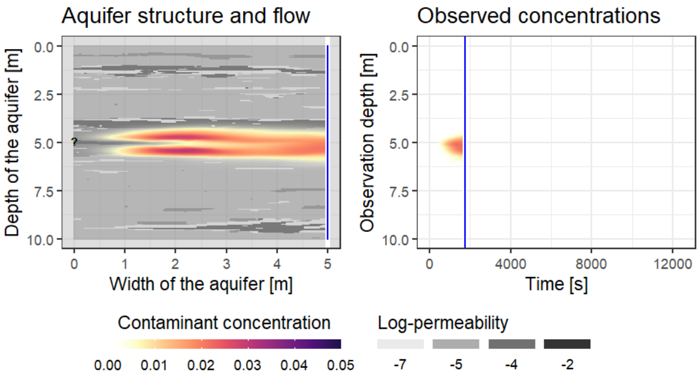

x=0.00
0.00
0.25
0.50
0.75
1.00
SLGP value
8
6
4
2
0
0.00
0.25
0.50
0.75
1.00
y

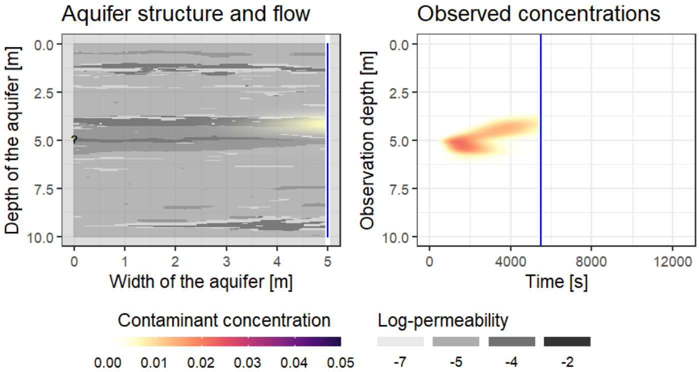

x=0.49
0.00
0.25
0.50
0.75
1.00
SLGP value
8
6
4
2
0
y



Go back to the main slide


Aquifer structure and flow
Depth of the aquifer [m]
0.0
2.5
5.0
7.5
10.0
0
1
2
3
4
5
Width of the aquifer [m]
Observed concentrations
Observation depth [m]
0.0
2.5
5.0
7.5
10.0
0
4000
8000
12000
Time [s]
Contaminant concentration
0.00
0.01
0.02
0.03
0.04
0.05
Log-permeability
-7
-5
-4
-2


Aquifer structure and flow
Observed concentrations
Depth of the aquifer [m]
0.0
2.5
5.0
7.5
10.0
0
1
2
3
4
5
Width of the aquifer [m]
Observation depth [m]
0
4000
8000
12000
Time [s]
Contaminant concentration
0.00
0.01
0.02
0.03
0.04
0.05
Log-permeability
-7
-5
-4
-2



Go back to the main slide