

# Des échantillons épars aux champs de distributions discrètes

**Athénaïs Gautier**

DTIS, ONERA, Université Paris-Saclay, 91120, Palaiseau, France

`athenais.gautier@onera.fr`

**ETICS** 28 Sept. - 02 Oct. 2026

# La préparation des slides pour ETICS : deux salles, deux ambiances...



1 Modelling for probabilistic inference

2 Spatial Logistic Gaussian Process

3 Active learning

4 Applications and proof of concept

- Case 1: binomial distribution with fixed  $n_{bin}$  and varying  $p$
- Case 2: binomial distribution with varying  $n_{bin}$  and  $p$

# Why GPs are so popular for active learning ?

## Flexibility

GPs fits complex response surfaces with very few assumptions.

GPs can be defined over plenty of different inputs.

GPs can incorporate prior knowledge.

## Probabilistic prediction

A full posterior, hence **uncertainty** at every input - not just a point estimate.  
→ probabilistic prediction!

# Why GPs are so popular for active learning ?

## Flexibility

GPs fits complex response surfaces with very few assumptions.

GPs can be defined over plenty of different inputs.

GPs can incorporate prior knowledge.

## Probabilistic prediction

A full posterior, hence **uncertainty** at every input - not just a point estimate.  
→ probabilistic prediction!

→ That **probabilistic prediction** is exactly what is leveraged for active learning: level-set estimation, Bayesian optimisation, probabilistic inversion, ...

# Why GPs are so popular for active learning ?

## Flexibility

GPs fits complex response surfaces with very few assumptions.

GPs can be defined over plenty of different inputs.

GPs can incorporate prior knowledge.

## Probabilistic prediction

A full posterior, hence **uncertainty** at every input - not just a point estimate.  
→ probabilistic prediction!

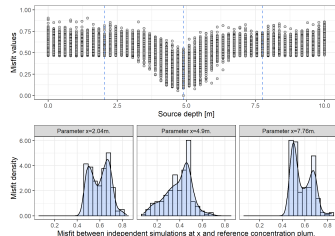
→ That **probabilistic prediction** is exactly what is leveraged for active learning: level-set estimation, Bayesian optimisation, probabilistic inversion, ...

**But:** a GP assumes a *Gaussian* output. What if the response is not Gaussian: a density, a class, a count?

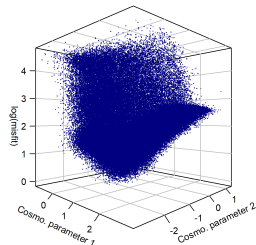
# PhD Motivation: Stochastic inverse problems & surrogate modelling

**The challenge:** Precise statistical inference in complex systems with index  $\mathbf{x}$  and response  $T_{\mathbf{x}}$ .

- Complex spatial dependence of  $T_{\mathbf{x}}$
- Costly model evaluations
- Heterogeneously scattered data



1D quantity of interest in an hydrogeological inverse problem



2D quantity of interest for a cosmological inverse problem

1 Modelling for probabilistic inference

2 Spatial Logistic Gaussian Process

3 Active learning

4 Applications and proof of concept

- Case 1: binomial distribution with fixed  $n_{bin}$  and varying  $p$
- Case 2: binomial distribution with varying  $n_{bin}$  and  $p$

# Spatial Logistic Gaussian Process : SLGP

Let the latent GP  $Z$  depend on the index  $\mathbf{x}$  as well as the response  $t \in [a, b]$ :

$$\left\{ \begin{array}{l} (Z_{\mathbf{x},t})_{\mathbf{x} \in D, t \in [a,b]} \sim \mathcal{GP} \\ Y_{\mathbf{x},t} := \frac{e^{Z_{\mathbf{x},t}}}{\int_{[a,b]} e^{Z_{\mathbf{x},u}} du} \sim SLGP \end{array} \right.$$

# Spatial Logistic Gaussian Process : SLGP

Let the latent GP  $Z$  depend on the index  $\mathbf{x}$  as well as the response  $t \in [a, b]$ :

$$\left\{ \begin{array}{l} (Z_{\mathbf{x},t})_{\mathbf{x} \in D, t \in [a,b]} \sim \mathcal{GP} \\ Y_{\mathbf{x},t} := \frac{e^{Z_{\mathbf{x},t}}}{\int_{[a,b]} e^{Z_{\mathbf{x},u}} du} \sim SLGP \end{array} \right.$$

At each  $\mathbf{x}$ ,  $Y_{\mathbf{x},\cdot}$  is, informally, a “random density” w.r.t. Lebesgues measure.

# Spatial Logistic Gaussian Process : SLGP

Let the latent GP  $Z$  depend on the index  $\mathbf{x}$  as well as the response  $t \in [a, b]$ :

$$\left\{ \begin{array}{l} (Z_{\mathbf{x},t})_{\mathbf{x} \in D, t \in [a,b]} \sim \mathcal{GP} \\ Y_{\mathbf{x},t} := \frac{e^{Z_{\mathbf{x},t}}}{\int_{[a,b]} e^{Z_{\mathbf{x},u}} du} \sim SLGP \end{array} \right.$$

At each  $\mathbf{x}$ ,  $Y_{\mathbf{x},\cdot}$  is, informally, a “random density” w.r.t. Lebesgues measure.  
Discrete-valued SLGP, for a response  $k$  in  $\{1, \dots, K\}$ :

$$\left\{ \begin{array}{l} (Z_{\mathbf{x},k})_{\mathbf{x} \in D, k \in \{1, \dots, K\}} \sim \mathcal{GP} \\ Y_{\mathbf{x},k} := \frac{e^{Z_{\mathbf{x},k}}}{\sum_{i=1}^K e^{Z_{\mathbf{x},i}}} \sim SLGP \end{array} \right.$$

## Likelihood of the observations

For locations  $(x_i)_{i=1}^n \in D^n$ , observations  $\mathbf{k}_n := (k_i)_{i=1}^n \in \{1, \dots, K\}^n$ , assuming independent  $k_i$ 's we have the likelihood:

$$\pi(\mathbf{k}_n | Y) = \prod_{i=1}^n Y_{\mathbf{x}_i, k_i} = \prod_{i=1}^n \frac{e^{Z_{\mathbf{x}_i, k_i}}}{\sum_{j=1}^K e^{Z_{\mathbf{x}_i, j}}} \quad (1)$$

## Likelihood of the observations

For locations  $(\mathbf{x}_i)_{i=1}^n \in D^n$ , observations  $\mathbf{k}_n := (k_i)_{i=1}^n \in \{1, \dots, K\}^n$ , assuming independent  $k_i$ 's we have the likelihood:

$$\pi(\mathbf{k}_n | Y) = \prod_{i=1}^n Y_{\mathbf{x}_i, k_i} = \prod_{i=1}^n \frac{e^{Z_{\mathbf{x}_i, k_i}}}{\sum_{j=1}^K e^{Z_{\mathbf{x}_i, j}}} \quad (1)$$

We focus on *finite-rank* GPs:

$$Z(\mathbf{x}, k) = \mu(\mathbf{x}, k) + \sum_{j=1}^p f_j(\mathbf{x}, k) \varepsilon_j = \boldsymbol{\varepsilon}^\top \mathbf{F}(\mathbf{x}, k), \quad \forall \mathbf{x} \in D, \quad k \in \{0, \dots, K\}, \quad (2)$$

where  $\boldsymbol{\varepsilon} = (\varepsilon_1, \dots, \varepsilon_p)^\top$  is a random vector of  $p$  i.i.d.  $\mathcal{N}(0, \sigma^2)$  random variables, and  $\mu$  is a deterministic known trend function.

## Posterior of a finite-rank SLGP

For locations  $(x_i)_{i=1}^n \in D^n$ , observations  $\mathbf{k}_n := (k_i)_{i=1}^n \in \{1, \dots, K\}^n$ , assuming independent  $k_i$ 's and a finite-rank SLGP, we have the posterior:

$$\pi[\epsilon|\mathbf{k}_n] \propto \phi_p(\epsilon) \prod_{i=1}^n \frac{e^{\mu(\mathbf{x}_i, k_i) + \sum_{j=1}^p \epsilon_j f_j(\mathbf{x}_i, k_i)}}{\sum_{\ell=1}^K e^{\mu(\mathbf{x}_i, \ell) + \sum_{j=1}^p \epsilon_j f_j(\mathbf{x}_i, \ell)}} \quad (3)$$

where  $\phi_p$  the pdf of the  $p$ -variate standard normal distribution

## Posterior of a finite-rank SLGP

For locations  $(x_i)_{i=1}^n \in D^n$ , observations  $\mathbf{k}_n := (k_i)_{i=1}^n \in \{1, \dots, K\}^n$ , assuming independent  $k_i$ 's and a finite-rank SLGP, we have the posterior:

$$\pi[\epsilon|\mathbf{k}_n] \propto \phi_p(\epsilon) \prod_{i=1}^n \frac{e^{\mu(\mathbf{x}_i, k_i) + \sum_{j=1}^p \epsilon_j f_j(\mathbf{x}_i, k_i)}}{\sum_{\ell=1}^K e^{\mu(\mathbf{x}_i, \ell) + \sum_{j=1}^p \epsilon_j f_j(\mathbf{x}_i, \ell)}} \quad (3)$$

where  $\phi_p$  the pdf of the  $p$ -variate standard normal distribution

This posterior is **log-concave**.

- Fast optimization (MAP)
- Good Laplace approximation
- Fast MCMC sampling. (Dwivedi et al., 2018)

# The main properties of discrete SLGP

Inference yields  $N$  posterior draws  $\epsilon^{(1)}, \dots, \epsilon^{(N)}$  (via Laplace or MCMC), each inducing a pmf field  $\hat{Y}_{\mathbf{x},k}^{(i)}$ . The average is  $\bar{Y}_{\mathbf{x},k} = \frac{1}{N} \sum_{i=1}^N \hat{Y}_{\mathbf{x},k}^{(i)}$ .

# The main properties of discrete SLGP

Inference yields  $N$  posterior draws  $\epsilon^{(1)}, \dots, \epsilon^{(N)}$  (via Laplace or MCMC), each inducing a pmf field  $\hat{Y}_{\mathbf{x},k}^{(i)}$ . The average is  $\bar{Y}_{\mathbf{x},k} = \frac{1}{N} \sum_{i=1}^N \hat{Y}_{\mathbf{x},k}^{(i)}$ .

## ✓ Exact

No quadrature error; the normaliser is a finite softmax sum. Inference is *cheaper* than in the continuous case.

## 📍 Can be evaluated at any $\mathbf{x}$

The finite-rank representation allows the surrogate can be *queried at unseen indices* - essential for design.

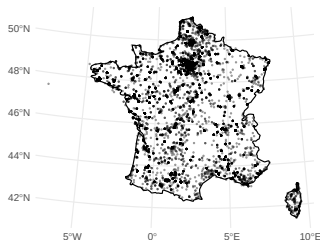
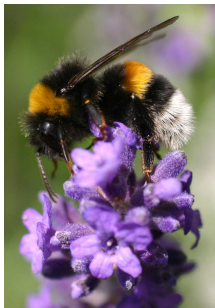
## 🔗 Ordinal for free

One *shared* latent process on  $D \times \{1, \dots, K\}$  couples neighbouring categories - natural for counts.

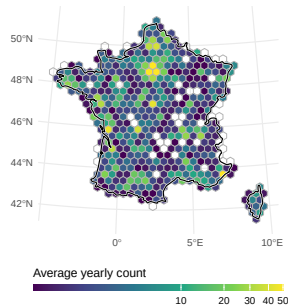
# Real data: *Bombus terrestris* occurrences in France

**Data.** GBIF occurrence records of *Bombus terrestris* in metropolitan France, 2019–2023, with valid coordinates and dates. Spatial index  $\mathbf{x} = (\text{lon}, \text{lat}) \in D \subset \mathbb{R}^2$ .

- Regular **hexagonal tessellation** (isotropic neighbourhoods, less directional bias).
- Per cell  $h$  and year  $y$ , a truncated count response:  
 $t_{h,y} = \min(50, \#\{\text{occurrences in } h \text{ during year } y\})$ .
- Truncation limits the influence of densely sampled urban cells.



Raw occurrence points.

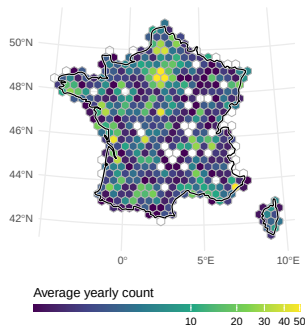


Hexagonal tessellation.

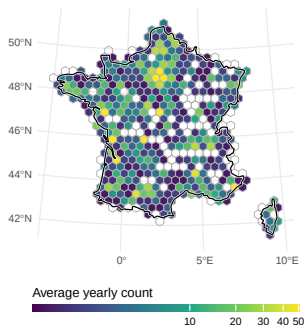
# Real data: training protocol and predictive summary

**Protocol.** Train on counts from **2020–2021** (each year an independent spatial replication), RFF Matérn-5/2, 200 frequencies  $\rightarrow$  400 basis functions.

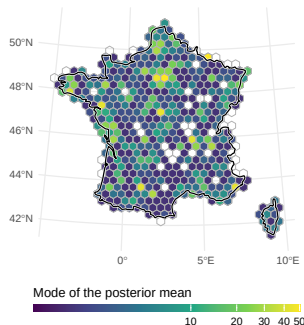
- Reference: empirical average yearly count over the longer window 2019–2023,  
$$\bar{t}_h^{\text{ref}} = \frac{1}{5} \sum_{y=2019}^{2023} t_{h,y}.$$
- Compared to the **mode of the SLGP posterior predictive** fitted on 2020–2021.



Full reference (2019–2023).



Training data (2020–2021).



SLGP posterior mode.

## Real data: fitting a model in R with SLGP

```
library(SLGP)
fit <- slgp(new_n ~ lon + lat, data = train, basisFunctionsUsed = "RFF",
            discrete = TRUE, nIntegral = 51, responseRange = c(0, 50),
            method = "Laplace", opts = list(ndraws = 100), seed = 1,
            opts_BasisFun = list(nFreq = 250, MatParam = 5/2),
            hyperparams = list(lengthscale = c(0.15, 0.15, 0.15), sigma2 = 1))
pmf <- predict(fit, newdata = grid)           # full pmf at every cell
cdf <- predict(fit, newdata = test, type = "cdf")
```

# Real data: assessing the calibration

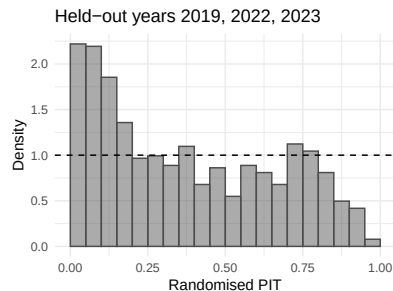
**Calibration assessment for count data** (model fitted on 2020–2021).

Czado, Gneiting & Held (2009). Predictive model assessment for count data. *Biometrics*, 65(4), 1254–1261.

For a count  $t$  at location  $\mathbf{x}$ , with  $F$  the posterior predictive CDF:

$$u = F(t - 1 \mid \mathbf{x}) + V (F(t \mid \mathbf{x}) - F(t - 1 \mid \mathbf{x})), \quad V \sim \mathcal{U}(0, 1).$$

Randomised PIT: If the model is calibrated,  $u \sim \mathcal{U}(0, 1)$ .



1 Modelling for probabilistic inference

2 Spatial Logistic Gaussian Process

3 Active learning

4 Applications and proof of concept

- Case 1: binomial distribution with fixed  $n_{bin}$  and varying  $p$
- Case 2: binomial distribution with varying  $n_{bin}$  and  $p$

# Active data acquisition with SLGP

**Goal.** Select new design points  $\mathbf{x} \in D$  that are informative for the entire conditional probability mass function  $(Y_{\mathbf{x},1}, \dots, Y_{\mathbf{x},K}) \mid \mathcal{D}_n$ .

$$p(T_{\mathbf{x}} = k \mid \mathcal{D}_n) = \int Y_{\mathbf{x},k}(\boldsymbol{\varepsilon}) p(\boldsymbol{\varepsilon} \mid \mathcal{D}_n) d\boldsymbol{\varepsilon}. \quad (4)$$

# Active data acquisition with SLGP

**Goal.** Select new design points  $\mathbf{x} \in D$  that are informative for the entire conditional probability mass function  $(Y_{\mathbf{x},1}, \dots, Y_{\mathbf{x},K}) \mid \mathcal{D}_n$ .

$$p(T_{\mathbf{x}} = k \mid \mathcal{D}_n) = \int Y_{\mathbf{x},k}(\boldsymbol{\varepsilon}) p(\boldsymbol{\varepsilon} \mid \mathcal{D}_n) d\boldsymbol{\varepsilon}. \quad (4)$$

## Posterior-draw approximation

Given posterior draws  $\hat{Y}_{\mathbf{x},k}^{(i)}$ ,  $i = 1, \dots, N$ :

$$p(T_{\mathbf{x}} = k \mid \mathcal{D}_n) \approx \bar{Y}_{\mathbf{x},k} = \frac{1}{N} \sum_{i=1}^N \hat{Y}_{\mathbf{x},k}^{(i)}.$$

# Active data acquisition with SLGP

**Goal.** Select new design points  $\mathbf{x} \in D$  that are informative for the entire conditional probability mass function  $(Y_{\mathbf{x},1}, \dots, Y_{\mathbf{x},K}) \mid \mathcal{D}_n$ .

$$p(T_{\mathbf{x}} = k \mid \mathcal{D}_n) = \int Y_{\mathbf{x},k}(\boldsymbol{\varepsilon}) p(\boldsymbol{\varepsilon} \mid \mathcal{D}_n) d\boldsymbol{\varepsilon}. \quad (4)$$

## Posterior-draw approximation

Given posterior draws  $\hat{Y}_{\mathbf{x},k}^{(i)}$ ,  $i = 1, \dots, N$ :

$$p(T_{\mathbf{x}} = k \mid \mathcal{D}_n) \approx \bar{Y}_{\mathbf{x},k} = \frac{1}{N} \sum_{i=1}^N \hat{Y}_{\mathbf{x},k}^{(i)}.$$

$$\mathbf{x}_{n+1} \in \arg \max_{\mathbf{x} \in D} a(\mathbf{x})$$

## Considered strategies

- **Random sampling** (*baseline*).

## Considered strategies

- **Random sampling** (*baseline*).
- **Probability variance:**

$$\hat{a}_{\text{VarProb}}(\mathbf{x}) = \sum_{k=1}^K \frac{1}{N-1} \sum_{i=1}^N \left( \hat{Y}_{\mathbf{x},k}^{(i)} - \bar{Y}_{\mathbf{x},k} \right)^2$$

Direct  $\ell_2$ -type dispersion in probability space.

## Considered strategies

- **Random sampling** (*baseline*).
- **Probability variance:**

$$\hat{a}_{\text{VarProb}}(\mathbf{x}) = \sum_{k=1}^K \frac{1}{N-1} \sum_{i=1}^N \left( \hat{Y}_{\mathbf{x},k}^{(i)} - \bar{Y}_{\mathbf{x},k} \right)^2$$

Direct  $\ell_2$ -type dispersion in probability space.

- **Predictive entropy** (MacKay, 1992; Cohn et al., 1994):

$$\hat{a}_{\text{Ent}}(\mathbf{x}) = - \sum_{k=1}^K \bar{Y}_{\mathbf{x},k} \log \bar{Y}_{\mathbf{x},k}$$

Total predictive uncertainty, mixes epistemic and aleatoric.

## Considered strategies

- **Random sampling** (*baseline*).
- **Probability variance:**

$$\hat{a}_{\text{VarProb}}(\mathbf{x}) = \sum_{k=1}^K \frac{1}{N-1} \sum_{i=1}^N \left( \hat{Y}_{\mathbf{x},k}^{(i)} - \bar{Y}_{\mathbf{x},k} \right)^2$$

Direct  $\ell_2$ -type dispersion in probability space.

- **Predictive entropy** (MacKay, 1992; Cohn et al., 1994):

$$\hat{a}_{\text{Ent}}(\mathbf{x}) = - \sum_{k=1}^K \bar{Y}_{\mathbf{x},k} \log \bar{Y}_{\mathbf{x},k}$$

Total predictive uncertainty, mixes epistemic and aleatoric.

- **BALD** (Houlsby et al., 2011):

$$\hat{a}_{\text{BALD}}(\mathbf{x}) = - \sum_{k=1}^K \bar{Y}_{\mathbf{x},k} \log \bar{Y}_{\mathbf{x},k} + \frac{1}{N} \sum_{i=1}^N \sum_{k=1}^K \hat{Y}_{\mathbf{x},k}^{(i)} \log \hat{Y}_{\mathbf{x},k}^{(i)}$$

Mutual information; targets disagreement between pmf draws.

1 Modelling for probabilistic inference

2 Spatial Logistic Gaussian Process

3 Active learning

4 Applications and proof of concept

- Case 1: binomial distribution with fixed  $n_{bin}$  and varying  $p$
- Case 2: binomial distribution with varying  $n_{bin}$  and  $p$

1 Modelling for probabilistic inference

2 Spatial Logistic Gaussian Process

3 Active learning

4 Applications and proof of concept

- Case 1: binomial distribution with fixed  $n_{bin}$  and varying  $p$

- Case 2: binomial distribution with varying  $n_{bin}$  and  $p$

## Case 1: binomial conditional pmf

**Controlled benchmark.** We consider a one-dimensional conditional distribution with known ground truth:

$$T_p \mid p \sim \text{Bin}(10, p), \quad p \in (0, 1),$$

so that  $K = 11$  categories and

$$\mathbb{P}(T_p = k) = \binom{10}{k} p^k (1 - p)^{10-k}, \quad k = 0, \dots, 10.$$

## Case 1: binomial conditional pmf

**Controlled benchmark.** We consider a one-dimensional conditional distribution with known ground truth:

$$T_p \mid p \sim \text{Bin}(10, p), \quad p \in (0, 1),$$

so that  $K = 11$  categories and

$$\mathbb{P}(T_p = k) = \binom{10}{k} p^k (1 - p)^{10-k}, \quad k = 0, \dots, 10.$$

**Evaluation metric.** Prediction quality is measured using integrated KL divergence:

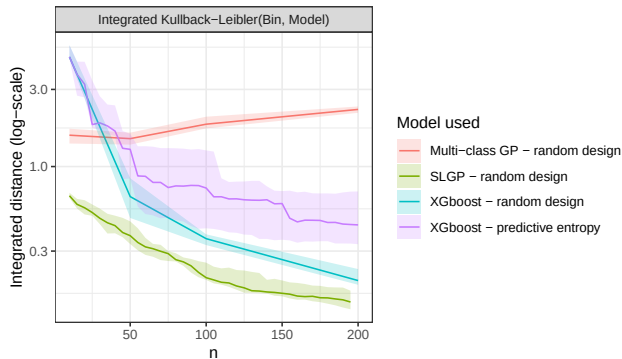
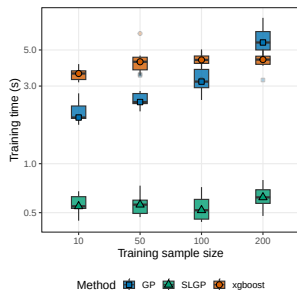
$$\text{IKL}(f, g) = \int_0^1 \sum_{k=0}^{10} f_p(k) \log \frac{f_p(k)}{g_p(k)} dp,$$

approximated numerically on a dense grid.

# Case 1: binomial conditional pmf

## Compared methods

- SLGP for conditional pmf estimation;
- GP-based multiclass model;
- random forest baseline.



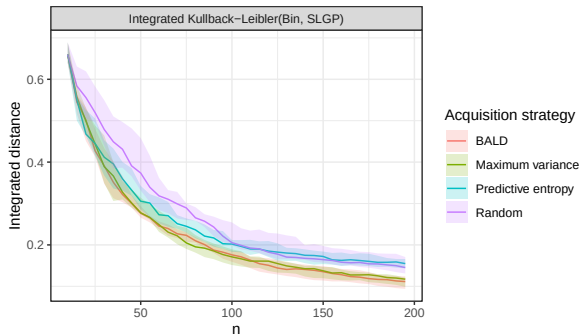
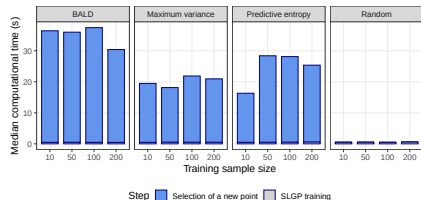
Predictive performance and computational time over 10 repetitions.

# Sequential design: learning the pmf field efficiently

**Protocol.** The SLGP is initialized with  $n_0 = 10$  random observations. At each iteration, a new location  $p_{\text{new}}$  is selected and 5 independent observations are generated from  $\text{Bin}(10, p_{\text{new}})$ .

## Acquisition strategies

- random sampling;
- predictive entropy;
- BALD;
- posterior variance of class probabilities.



Median IKL across 10 repetitions; ribbons show interquartile ranges.

Finite-rank SLGP (RFF, Matérn-5/2, 100 frequencies → 200 basis functions).

1 Modelling for probabilistic inference

2 Spatial Logistic Gaussian Process

3 Active learning

4 Applications and proof of concept

- Case 1: binomial distribution with fixed  $n_{bin}$  and varying  $p$

- Case 2: binomial distribution with varying  $n_{bin}$  and  $p$

## Case 2: binomial with varying $n_{bin}$ and $p$

**Two-dimensional benchmark.** We now index by  $\mathbf{x} = (n_{bin}, p) \in \{5, \dots, 20\} \times (0, 1)$ , with

$$T_{n_{bin}, p} \mid \mathbf{x} \sim \text{Bin}(n_{bin}, p),$$

so that  $K = 21$  categories and the *support itself varies* with  $n_{bin}$ :

$$\mathbb{P}(T_{n_{bin}, p} = k) = \mathbb{1}_{k \leq n_{bin}} \binom{n_{bin}}{k} p^k (1 - p)^{n_{bin} - k}, \quad k = 0, \dots, 20.$$

## Case 2: binomial with varying $n_{bin}$ and $p$

**Two-dimensional benchmark.** We now index by  $\mathbf{x} = (n_{bin}, p) \in \{5, \dots, 20\} \times (0, 1)$ , with

$$T_{n_{bin}, p} \mid \mathbf{x} \sim \text{Bin}(n_{bin}, p),$$

so that  $K = 21$  categories and the *support itself varies* with  $n_{bin}$ :

$$\mathbb{P}(T_{n_{bin}, p} = k) = \mathbb{1}_{k \leq n_{bin}} \binom{n_{bin}}{k} p^k (1 - p)^{n_{bin} - k}, \quad k = 0, \dots, 20.$$

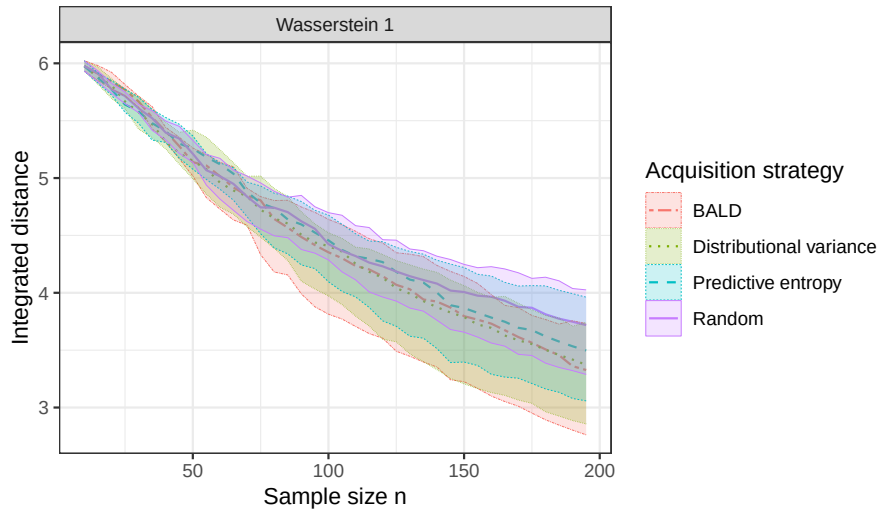
**Evaluation metric.** KL is unstable under support shifts, so we use the Wasserstein-1 distance:

$$W_1(F, G) = \sum_{k=0}^K |F(k) - G(k)|,$$

on the CDFs  $F, G$ , integrated over the input domain.

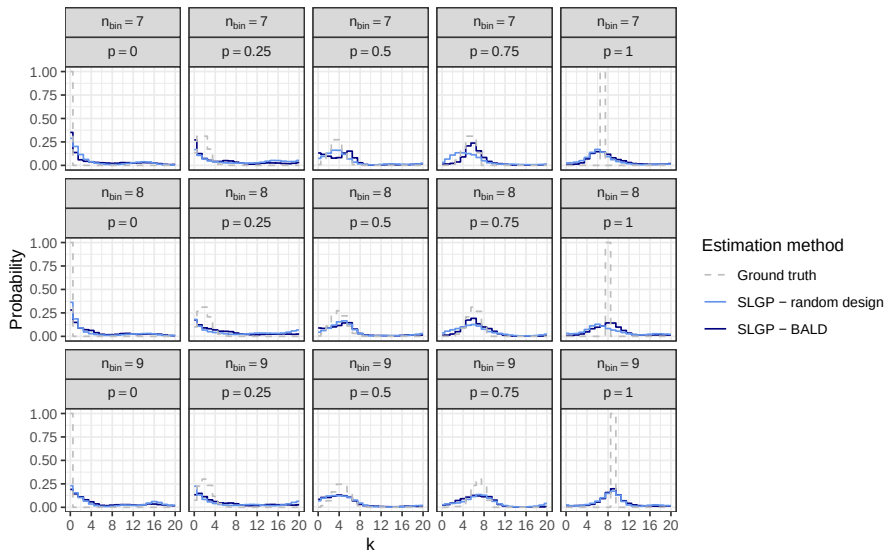
Same finite-rank SLGP as before (RFF, Matérn-5/2, 100 frequencies  $\rightarrow$  200 basis functions). The varying support makes inference and design more involved.

## Case 2: results without structural trend



vs. dataset size; ribbons show interquartile ranges (10 repetitions).

## Case 2: results without structural trend



Truth vs. SLGP (random) vs. SLGP (BALD).  $\Rightarrow$  purely smooth latent representations adapt poorly to

# Encoding support with a structured trend

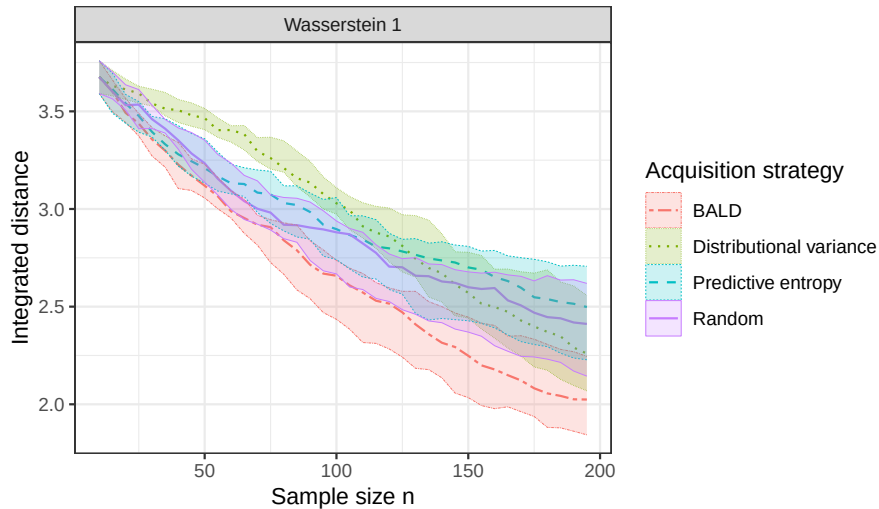
**Idea.** Inject structural prior knowledge into the latent field via a deterministic trend that penalises out-of-support categories:

$$\mu(n_{bin}, p, k) = -10 \cdot \mathbb{1}_{\{k > n_{bin}\}}.$$

Plugging into the discrete SLGP gives probabilities:

$$Y_{n_{bin}, p}(k) \propto \begin{cases} \exp\{\varepsilon^\top \mathbf{F}(n_{bin}, p, k)\}, & k \leq n_{bin}, \\ 4.5 \times 10^{-5} \exp\{\varepsilon^\top \mathbf{F}(n_{bin}, p, k)\}, & k > n_{bin}. \end{cases}$$

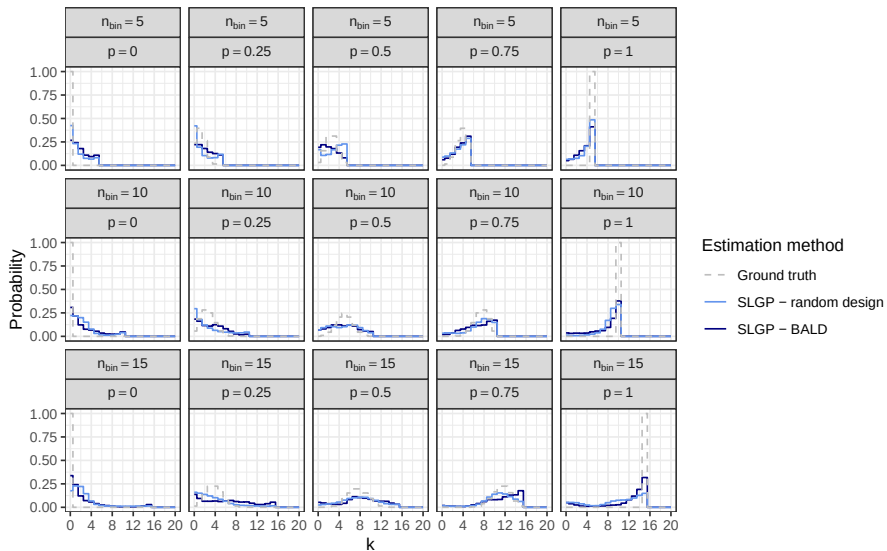
## Case 2: results with structural trend



Integrated  $W_1$

vs. dataset size; ribbons show interquartile ranges (10 repetitions).

## Case 2: results with structural trend



Truth vs. SLGP (random) vs. SLGP (BALD).

**What we did.** A discrete-output **SLGP** for spatially indexed categorical/count responses  $\rightarrow$  full posterior on the pmf at any location, with coherent uncertainty.

## Takeaways.

- The continuous SLGP extends to finite support  $\rightarrow$  normalized w.r.t. atomic reference measure, normaliser is a finite softmax (no quadrature error).
- Epistemic-uncertainty criteria (especially **BALD**) beat random sampling for distributional reconstruction.
- A **structured trend** encodes support constraints and markedly improves the varying- $n$  case.

**Next.** Estimating the trend jointly, variational / augmentation-based inference for large  $K$ , SUR-type criteria on distributional functionals, proper scoring rules when ground truth is unknown.

## Previous SLGP literature

-  Tokdar, Surya T. et al. (2010). “Bayesian density regression with logistic Gaussian process and subspace projection”. In: *Bayesian analysis* 5.2, pp. 319–344.
-  Pati, Debdeep et al. (2013). “Posterior consistency in conditional distribution estimation”. In: *Journal of multivariate analysis* 116, pp. 456–472.

## Theory and methodology

-  Gautier, Athénaïs (2023). “Modelling and predicting distribution-valued fields with applications to inversion under uncertainty”. PhD Thesis. Institute of Mathematical Statistics and Actuarial Science, University of Bern.
-  Gautier, Athénaïs and David Ginsbourger (2025). *Continuous Logistic Gaussian Random Measure Fields for Spatial Distributional Modelling*. To appear in Annals of the Institute of Statistical Mathematics. arXiv: 2110.02876 [stat.ME].
-  Gautier, Athenaïs (2026). *SLGP-based surrogates for spatially dependent discrete outputs: modelling, uncertainty quantification, and data acquisition*. Under review for the Proceedings of mODa 14.

## References for the criterion

-  MacKay, David JC (1992). “Information-based objective functions for active data selection”. In: *Neural computation* 4.4, pp. 590–604.
-  Cohn, David et al. (1994). “Active learning with statistical models”. In: *Advances in Neural Information Processing Systems* 7.
-  Houlisby, Neil et al. (2011). *Bayesian Active Learning for Classification and Preference Learning*. arXiv: 1112.5745 [stat.ML].

## References for the SLGP implementation, and code

-  Gautier, Athénaïs (2026a). *Github repository - Code for the numerical experiments*. URL: <https://github.com/AthenaisGautier/SLGP-discrete> (visited on 05/30/2026).
-  — (2026b). *SLGP: An R Package for Spatial Conditional Density Estimation*. arXiv: 2609.33827 [stat.ME]. URL: <https://arxiv.org/abs/2609.33827>.
-  — (2026c). *SLGP: Spatial Logistic Gaussian Process for field density estimation*. R package version 2.0.0. URL: <https://CRAN.R-project.org/package=SLGP>.

### Other application of the SLGP for goal-oriented learning

-  Gautier, Athénaïs et al. (2020). “Probabilistic ABC with Spatial Logistic Gaussian Process modelling”. In: *Third Workshop on Machine Learning and the Physical Sciences*. URL: [https://ml4physicalsciences.github.io/2020/files/NeurIPS\\_ML4PS\\_2020\\_112.pdf](https://ml4physicalsciences.github.io/2020/files/NeurIPS_ML4PS_2020_112.pdf).
-  — (2021). “Goal-oriented adaptive sampling under random field modelling of response probability distributions”. In: *ESAIM: Proceedings and Surveys* 71, pp. 89–100. DOI: 10.1051/proc/202171108. URL: <https://doi.org/10.1051/proc/202171108>.
-  Gautier, Athénaïs (2023). “Modelling and predicting distribution-valued fields with applications to inversion under uncertainty”. PhD Thesis. Institute of Mathematical Statistics and Actuarial Science, University of Bern.

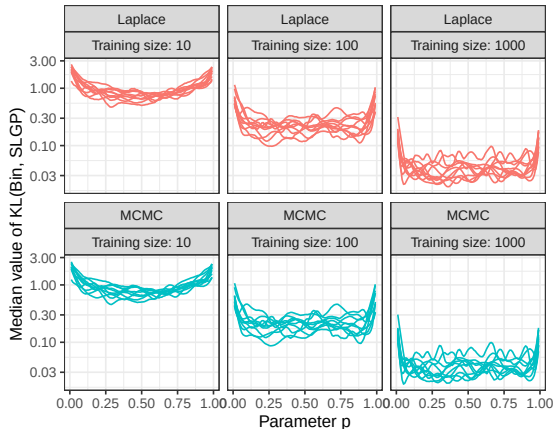
**Thank you for your attention!**

**Questions?**

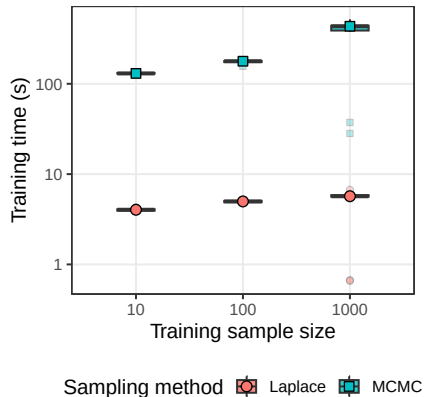
## 5 Appendix

# Posterior approximation: Laplace vs. MCMC

The SLGP likelihood is log-concave and the prior Gaussian  $\Rightarrow$  the posterior is log-concave, so a Laplace approximation around the mode is well justified.



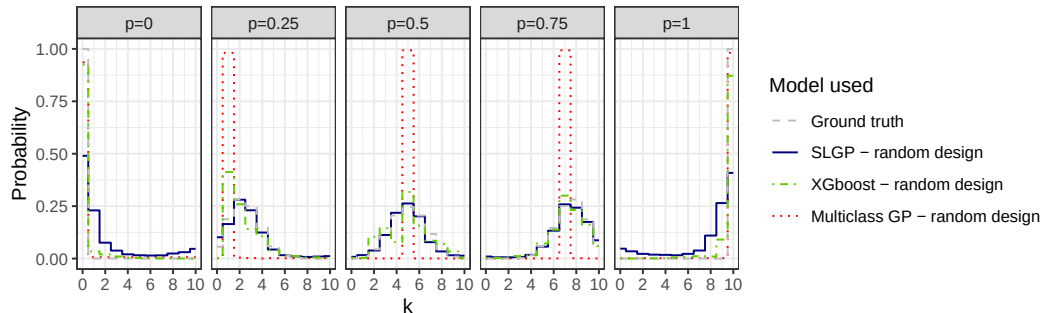
Pointwise KL to the ground-truth pmf,  $n \in \{10, 100, 1000\}$ , 10 repetitions.



Training time on a laptop.

# Posterior approximation: SLGP vs GP VS XGBoost

Case : Binomial with varying  $p$ , estimation of SLGP vs of multi-class GP VS XGboost



# Connections to familiar models

## Multinomial logistic regression

- feature map  $\tilde{\mathbf{F}}(\mathbf{x})$  *shared* across classes
- coefficients  $\tilde{\theta}_k$  vary with  $k$
- up to  $K \times P$  free parameters

## Multiclass GP classifier

- one latent GP *per class*
- no intrinsic ordering of labels

## Discrete-output SLGP

- *single* shared  $\epsilon$
- class-specific *basis*  $\mathbf{F}(\mathbf{x}, k)$
- one process on  $D \times \{1, \dots, K\}$
- structured, input-dependent cross-class correlation
- ordinal structure built in
- latent dimension fixed