Multi-output Gaussian processes for inverse uncertainty quantification in random neutronics

Paul Lartaud[1,2], **Josselin Garnier**[2], **Philippe Humbert**[1]

[1] CEA/DAM/DIF, [2]Ecole polytechnique (CMAP)

Outline

Introduction

Context

- Identification of a fissile material based on neutron correlation measurements $\longrightarrow$ inverse problem resolution.
- Applications : nuclear safeguards, criticality accidents detections, waste identification ..

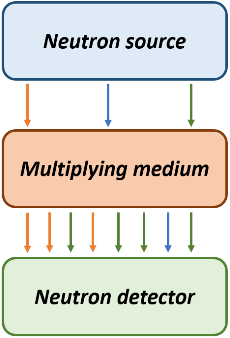
Methodology

- Direct model : **too costly** (MC simulations) or **biased** (analytical model).
- Surrogate model for the direct model using Gaussian process regression.
- Bayesian resolution of the inverse problem, including the predictive covariance of the multi-output.

Neutron correlations

Definitions

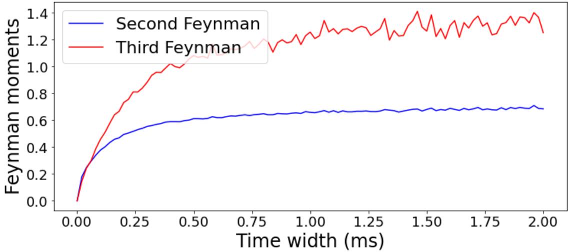
- **Excess of variance** in a multiplying medium compared to a standard Poisson process.
- **True correlations** = coincident detections of neutrons from the same fission chain.
- Measuring the **correlated detections** provide information on the multiplying medium.



Feynman moments

Definitions

- R is the average count rate (in neutrons.s^{-1}) during a detection window of size T.
- $Y(T)$ and $X(T)$ are respectively the average number of double and triple true correlations.
- Observations : (R, Y_∞, X_∞) where $Y_\infty = \lim_{T\to+\infty} Y(T)$ and $X_\infty = \lim_{T\to+\infty} X(T)$.



Point model

Problem description

- Prompt multiplication factor $0 < k_p < 1$.
- Detector efficiency $\varepsilon_F = \frac{\text{nb of counts}}{\text{nb of induced fissions}}$
- Source intensity (in $n.s^{-1}$)

Assumptions

- Homogeneous infinite material
- Mono-energetic neutrons
- Only fission and capture reactions

$$y = (R, Y_\infty, X_\infty)^T = f_{\text{PM}}(x) \text{ with } x = (k_p, \varepsilon_F, S) \quad (1)$$

Observations

Direct model outputs

- Analog Monte-Carlo simulations with MCNP6.
- $N = 20$ independent noisy observations $\mathbf{y} = (y^{(k)})_{1 \leq k \leq N}$.

$$y^{(k)} = f(x) + \varepsilon^{(k)} \text{ with } \varepsilon^{(k)} \sim \mathcal{N}(0, \mathbf{C}_m), \quad \mathbf{C_m} : \text{Noise error}$$

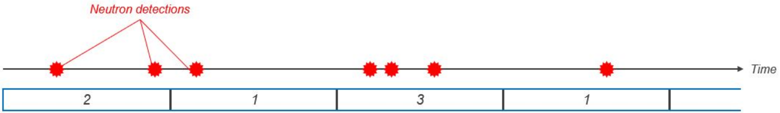
- The direct model f is identified to the approximated point model f_{PM}.
- Observation covariance $\mathbf{C}_m$ estimated by bootstrap.

Sequential binning

- Simulation split into N_T time windows of width T, with n_k the number of detections in window k.
- Asymptotically normal estimators based on central moment estimators

$\widehat{\mu_p} = \frac{1}{N_T} \sum_{k=1}^{N_T} (n_k - \widehat{\mu_1})^p$ for $p > 1$ and $\widehat{\mu_1} = \frac{1}{N_T} \sum_{k=1}^{N_T} n_k$

$$\widehat{Y_T} = \frac{\widehat{\mu_2}}{\widehat{\mu_1}} - 1 \text{ and } \widehat{X_T} = \frac{\widehat{\mu_3}}{\widehat{\mu_1}} - 1 - 3\widehat{Y_T}$$



Bayesian inverse problem

- **Goal** : uncertainty quantification for decision-making.
- **Ill-posed inverse** with very noisy observations, solved by Bayesian inference.

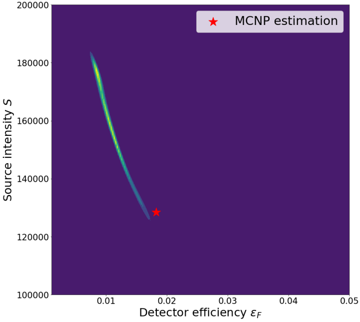
$$y = f(x) + \varepsilon \text{ with } \varepsilon \sim \mathcal{N}(0, \mathbf{C}_m), \quad \mathbf{C_m} : \text{Noise error}$$

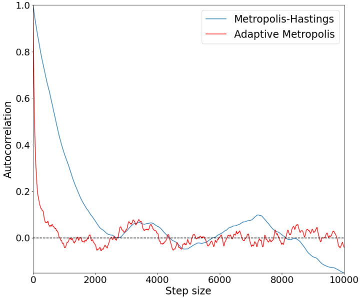
- Posterior distribution $p(x|\mathbf{y})$ sampled by MCMC methods.

$$p(x|\mathbf{y}) \propto p(x) \exp\left(\prod_{k=1}^{N} (y^{(k)} - f(x))^T \mathbf{C}_m{}^{-1} (y^{(k)} - f(x))\right)$$

Posterior distribution

- Sampling with Adaptive Metropolis instead of standard Metropolis-Hastings.
- Thin posterior distribution but true point outside the support.
- Model bias not accounted for in the Bayesian resolution.





Which direct model ?

Limitations

- **Point model** : analytical model describing neutron correlations. Fast but biased. UQ not reliable.
- **Monte-Carlo simulations** : accurate but too costly (30 minutes per run).

Solution

- **GP Surrogate model** : fast predictions and improved compared to point model.

$$f_s(x) \sim \mathcal{N}\left(\overline{f_s}(x), \mathbf{C_s}(x)\right), \quad \mathbf{C_s}(x) : \text{Model error}$$

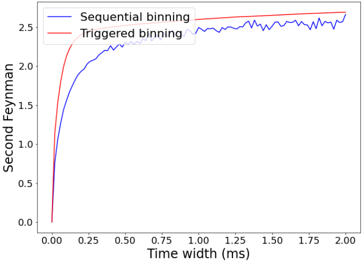
- Training on a dataset based on MC simulations. Analytical model used to improve training [1].

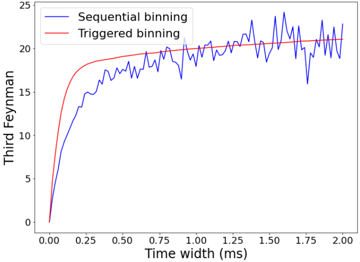
1. M. C. Kennedy et A. O'Hagan, « Bayesian calibration of computer models, » *Journal of the Royal Statistical Society : Series B (Statistical Methodology)*, t. 63, n° 3, p. 425-464, 2001

Training dataset (1)

Filtered triggered binning

- Neutron history number are known. Accidental correlations can be filtered out.
- Resulting training dataset much less noisy.
- Observation noise is kept identical to represent experimental results.

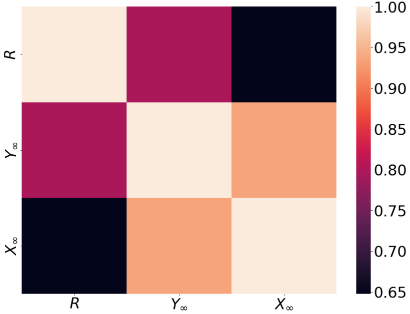




Training dataset (2)

- 1125 independent simulations with MCNP6. Feature dimension $I = 7$, observation dimension $D = 3$.
- Density, geometries, compositions randomly chosen for each simulation.
- Strongly correlated outputs

Output correlations



Standard scalar GP

Standard GP training and inference

- For $f \sim \mathcal{GP}(0, k(x, x'))$ and training dataset $(\mathbf{X}, \mathbf{f})$. Posterior predictive law on a test set $(\mathbf{X}_*, \mathbf{f}_*)$:

$$p(\mathbf{f}_*|\mathbf{f}, \mathbf{X}_*, \mathbf{X}) \sim \mathcal{N}(\boldsymbol{\mu}, \Sigma)$$

$$\boldsymbol{\mu} = K(\mathbf{X}, \mathbf{X}_*)^T K(\mathbf{X}, \mathbf{X})^{-1}\mathbf{f}$$

$$\Sigma = K(\mathbf{X}_*, \mathbf{X}_*) - K(\mathbf{X}, \mathbf{X}_*)^T (K(\mathbf{X}, \mathbf{X}) + \sigma_n^2 \mathcal{I})^{-1} K(\mathbf{X}, \mathbf{X}_*)$$

- Hyperparameter selection by maximization of marginal likelihood.

$$\log p(\mathbf{f}|\mathbf{X}) = -\frac{1}{2}\mathbf{f}^T K_\sigma^{-1}\mathbf{f} - \frac{1}{2}\log|K_\sigma| - \frac{I}{2}\log(2\pi) \text{ with } K_\sigma = K(\mathbf{X}, \mathbf{X}) + \sigma_n^2 \mathcal{I}$$

Multi-output Gaussian processes (1)

Linear Model of Coregionalization (LMC)

- Linear combination of Q latent independent scalar GPs $u_q \sim \mathcal{GP}\left(0, k_q(x, x')\right)$ for $q \leq Q$ with $\mathbf{W} \in \mathcal{M}_{D,Q}(\mathbb{R})$.

$$f_d(x) = \sum_{q=1}^{Q} w_{d,q} u_q(x)$$

- Standard GP inference with multi-output covariance kernel.

$$\text{Cov}(f_d(x), f_{d'}(x_*)) = \sum_{q=1}^{Q} w_{d,q} w_{d',q} k_q(x, x_*)$$

Surrogate models performance (1)

- Dataset of 1125 instances. Normalized mean absolute error on the validation set (30% of dataset).
  - *IGP* : independent scalar GP for each output.
  - *LMC2* : LMC GP with 2 latent GPs.
  - *LMC3* : LMC GP with 3 latent GPs.
  - *PM* : analytical point model.

	IGP	*LMC2*	*LMC3*	*PM*
R	0.0083	**0.0081**	0.0082	0.1933
Y_∞	0.029	0.028	**0.027**	0.0654
X_∞	0.155	**0.073**	0.084	0.151

- Improved predictions compared to point model. Comparable performance for the different surrogates. But IGP assume independent channels.

Surrogate models performance (2)

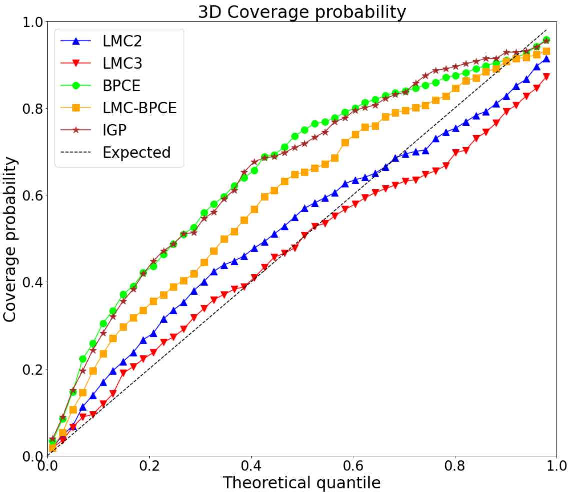
- **Reliability** of predictive covariances ?
- **Multi-output coverage probabilities** : for a given test instance (X_*, f_*), fraction of test inputs in the credible regions $I_\alpha(f_*)$ of confidence level α

$$D_M(f_*, f_s(X_*))^2 = \left(f_* - \overline{f_s(X_*)}\right)^T \mathbf{C}_s(X_*)^{-1} \left(f_* - \overline{f_s(X_*)}\right)$$

$$I_\alpha(f_*) = \left\{x \in \mathcal{X} | D_M(f_*, f_s(x))^2 \leq q_\alpha\right\}$$

with q_α the quantile of confidence level α for the χ^2 distribution with $D = 3$ levels of freedom.

Surrogate models performance (3)



Posterior distribution sampling

- Bayesian resolution of the inverse problem.

$$y = \overline{f_s(x)} + \eta(x) + \varepsilon \text{ with } \varepsilon \sim \mathcal{N}(0, \mathbf{C}_m) \text{ and } \eta(x) \sim \mathcal{N}(0, \mathbf{C}_s(x))$$

- Gaussian likelihood with model error and noise error [2].

$$L(\mathbf{y}|x) \propto \exp\left(-\frac{1}{2}\left(\mathbf{y} - \overline{\mathbf{f_s}(x)}\right)^T \mathbf{C}_{\text{tot}}(x)^{-1}\left(\mathbf{y} - \overline{\mathbf{f_s}(x)}\right)\right)$$

where $\overline{\mathbf{f_s}(x)} = \left(\overline{f_s(x)}, \ldots, \overline{f_s(x)}\right)^T \in \mathbb{R}^{DN}$

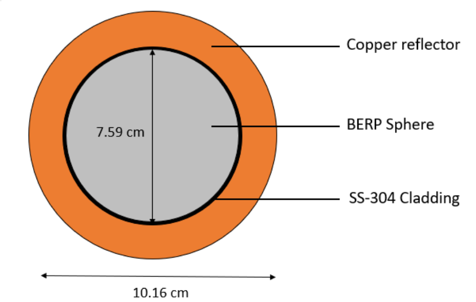
$$\mathbf{C}_{\text{tot}}(x) = \begin{pmatrix} \mathbf{C}_s(x) + \mathbf{C}_m & \cdots & \mathbf{C}_s(x) \\ \vdots & \ddots & \vdots \\ \mathbf{C}_s(x) & \cdots & \mathbf{C}_s(x) + \mathbf{C}_m \end{pmatrix} \in \mathbb{R}^{DN \times DN}$$

2. P. Lartaud, P. Humbert, Garnier et al., « Multi-output Gaussian processes for inverse uncertainty quantification in neutron noise analysis, » *Nuclear Science and Engineering*, p. 1-24, 2023

Copper-reflected Pu sphere

ICSBEP Benchmark

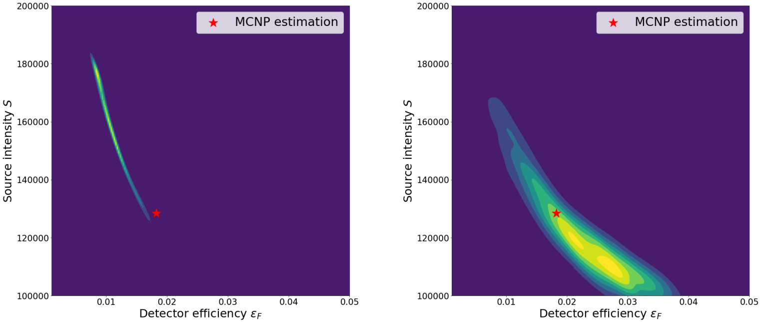
- Test example extracted from ICSBEP Handbook [3]. Metallic Pu sphere with copper reflector and polyethylene-moderated ^{3}He detector.
- 20 independent observations $y^{(k)}$ with sequential binning in MCNP6.



3. J. B. Briggs, L. Scott et A. Nouri, « The international criticality safety benchmark evaluation project, » *Nuclear science and engineering*, t. 145, n° 1,

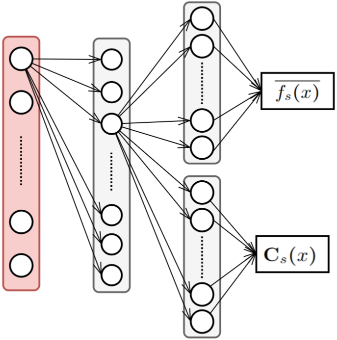
Copper-reflected Pu sphere (2)

- MCMC sampling (Adaptive Metropolis) with likelihood including model and observation noise.
- Thin distribution support, long decorrelation time $\widehat{\tau} \simeq 3000$. Large number of MCMC iterations $K = 5 \times 10^6$ MCMC iterations.



Probabilistic BNN

- Assume Gaussian training outputs and learn directly the **output distribution**.
- Learn mean $\overline{f_s(x)}$ and covariance $\mathbf{C}_s(x)$ with **negative log-likelihood loss**.
- Predictive covariances included in inverse problem likelihood.



Heteroscedastic GP - two approaches

Covariance modeling with auxiliary GP

- Covariance is learnt with an **auxiliary homoscedastic GP** based on residuals dataset.
- Very data hungry. Larger dataset and sparse approximations required [4].

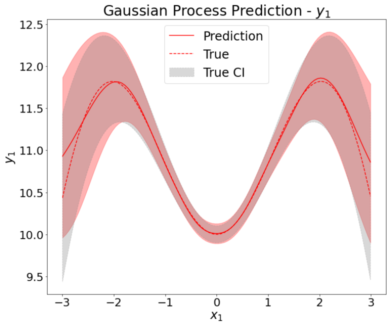
GP with PCE covariance noise kernel

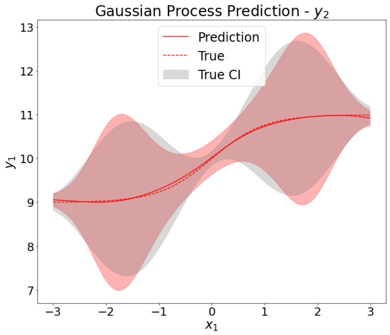
- **Polynomial Chaos Expansion** (PCE) modeling of the covariance.
- Polynomial coefficients treated as kernel hyperparameters.
- Dramatic increase of nb of hyperparameters. Optimization more difficult.

4. J. Quinonero-Candela et C. E. Rasmussen, « A unifying view of sparse approximate Gaussian process regression, » *The Journal of Machine Learning Research*, t. 6, p. 1939-1959, 2005

Heteroscedastic GP - a toy case

Toy case example of an heteroscedastic GP regression with 2 outputs.





Conclusion

Contributions

- Development of a general methodology for reliable inverse UQ with surrogate models.
- Application to a neutron multiplication benchmark.

Limitations

- Heteroscedasticity not treated as of yet.
- Generalization to other problems in neutron noise analysis ?

Thank you for your attention.