Multi-fidelity modeling using DGPs: Improvements and a generalization to varying input space dimensions

Ali Hebbal[1, 2], Mathieu Balesdent[1], Loic Brevault[1], El-Ghazali Talbi[2], Nouredine Melab[2]

[1]ONERA, DTIS, Université Paris Saclay | [2]Université de Lille, CNRS/CRIStAL, Inria Lille

Improvements of DGPs for multi-fidelity modeling

Context: using Gaussian Process for multi-fidelity modeling

- High-fidelity (HF) models → accurate data but on a limited dataset.
- Low-fidelity (LF) models → large amount of data of approximated data.
- Use of multi-fidelity model to exhibit correlations between datasets of s increasing levels of fidelity $(X^t, \mathbf{y}^t), \forall 1 \leq t \leq s$, to improve prediction accuracy.

- Multi-fidelity models based on Gaussian Processes (GPs):
  - **Auto-Regressive model (AR1)**:

$$f_t(\mathbf{x}) = \rho f_{t-1}(\mathbf{x}) + \gamma_t(\mathbf{x})$$

| Scaling factor ρ → capture only linear correlations between fidelities.

  - **Non-linear Autoregressive model (NARGP)**:

$$f_t(\mathbf{x}) = g_t\left(f^*_{t-1}(\mathbf{x})\right) + \gamma_t(\mathbf{x})$$

| The posterior GP f^*_{t-1} is used → Sequential optimization of each fidelity → overfitting.

  - **Multi-fidelity Deep Gaussian Process model (MF-DGP)**:

$$f_t(\mathbf{x}) = g_t\left(f_{t-1}(\mathbf{x})\right) + \gamma_t(\mathbf{x})$$

| Keeping the prior f_{t-1} comes back to a Deep Gaussian Process.

Use of stochastic variational inference to obtain the Evidence Lower Bound (ELBO):

$$\mathcal{L} = \sum_{t=1}^{s}\sum_{i=1}^{n_t} \mathbb{E}_{q(f_t^{(i),t})}\left[\log p(y^{(i),t}|f_t^{(i),t})\right] - \sum_{l=1}^{s} KL\left[q(\mathbf{u}_l)||p(\mathbf{u}_l; Z_{l-1})\right]$$

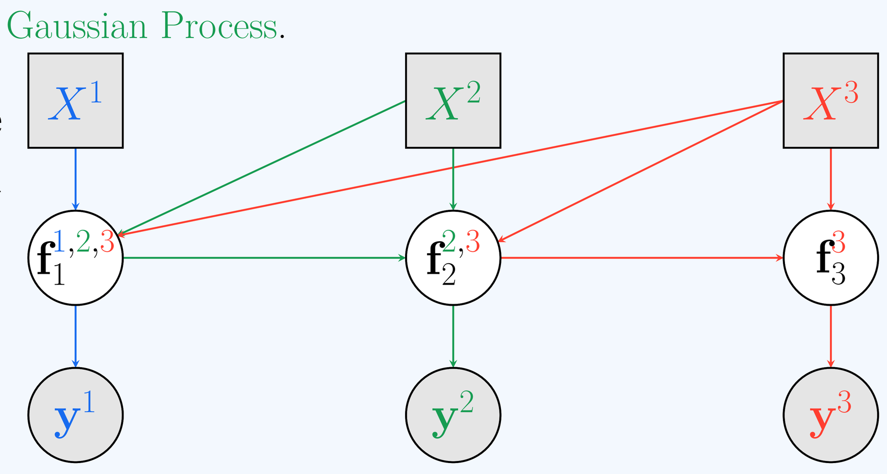

Figure 1:MF-DGP graphical representation

First improvement of MF-DGP: combining natural and stochastic gradient for training

✗ Issue: Parameter space not euclidian ($q(\mathbf{u})$) → stochastic ordinary gradient may not be adapted

✓ Proposed optimization process adapted to the parameter space:

Loop between:

- step using stochastic ordinary gradient (Adam Optimizer) with respect to $(\{\Theta\}_{l=1}^s, \{Z_l\}_{l=1}^s)$
- step using natural gradient with respect to all the variational distributions $(q(\mathbf{u}_l)\}_{l=1}^s)$.

Second improvement of MF-DGP: Optimization of inducing points

✗ Issue: The augmented input space induces a dependency between $Z_{l,1:d}$ and $Z_{l,d+1}$ → Free optimization of the inducing inputs is not appropriate.

$\{Z_l\}_{l=2}^s$ are constrained as follows:

$$Z_l = \left[Z_{l,1:d}, f^*_{l-1}(Z_{l,1:d})\right]; \forall 2 \leq l \leq s$$

where $f^*_{l-1}(\cdot)$ corresponds to the prediction at the previous layer.

✓ Optimization with respect to $Z_{l,1:d}$ instead of Z_l.

Results MF-DGP improved

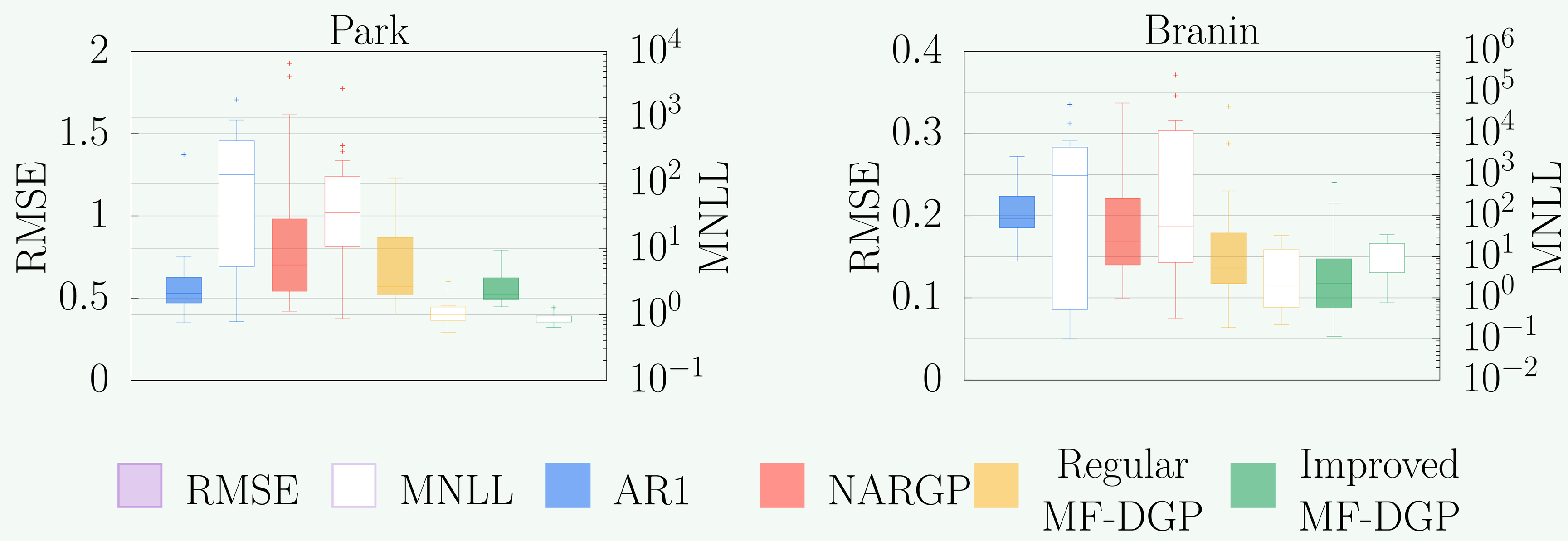

Figure 2:Performance of the different multi-fidelity models with different DoE. RMSE refers to the root mean squared error and MNLL to the mean negative test log likelihood.

References

[1] H. Salimbeni and M. Deisenroth, "Doubly stochastic variational inference for deep gaussian processes," in *Advances in Neural Information Processing Systems*, pp. 4588–4599, 2017.
[2] K. Cutajar, M. Pullin, A. Damianou, N. Lawrence, and J. González, "Deep gaussian processes for multi-fidelity modeling," *arXiv preprint arXiv:1903.07320*, 2019.
[3] P. Perdikaris, M. Raissi, A. Damianou, N. Lawrence, and G. E. Karniadakis, "Nonlinear information fusion algorithms for data-efficient multi-fidelity modelling," *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 473, no. 2198, p. 20160751, 2017.
[4] A. Paleyes, M. Pullin, M. Mahsereci, N. Lawrence, and J. González, "Emulation of physical processes with emukit," in *Second Workshop on Machine Learning and the Physical Sciences, NeurIPS*, 2019.
[5] M. C. Kennedy and A. O'Hagan, "Bayesian calibration of computer models," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 63, no. 3, pp. 425–464, 2001.
[6] S. Tao, D. W. Apley, W. Chen, A. Garbo, D. J. Pate, and B. J. German, "Input mapping for model calibration with application to wing aerodynamics," *AIAA Journal*, pp. 2734–2745, 2019.

Multi-fidelity with varying input spaces

In some industrial problems each fidelity can be characterized by its own input space due to:

- Different modeling approaches in each fidelity,
- Omission of some variables in the lower fidelities.

Example: A launch vehicle thrust frame can be modeled in the low fidelity with one average thickness parameter for all the thrust frame, while in high fidelity, a thickness for each element of the system can be considered.

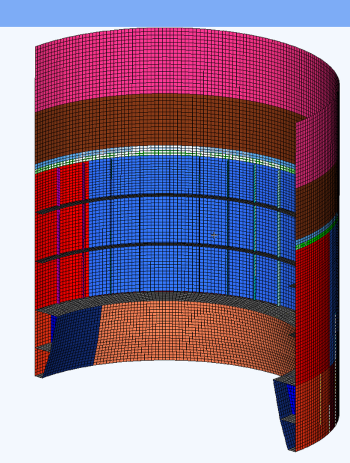

Figure 3:Model of launch vehicle thrust frame.

Existing approaches

- Nominal mapping between input spaces based on physics of the problem.
  ✗ Potentially better calibrated input mapping,
  ✗ Not applicable to black-box problems.
- Input mapping calibration (IMC) finds a parametric mapping $g_{\boldsymbol{\beta}}(\cdot)$ with the parameters $\boldsymbol{\beta}$ estimated as follows:

$$\boldsymbol{\beta}^* = \underset{\boldsymbol{\beta}}{\operatorname{argmin}}\left(\sum_{i=1}^{n_{hf}}\left(f_{hf}(\mathbf{x}^{hf,(i)}) - f_{lf}\left(g_{\boldsymbol{\beta}}(\mathbf{x}^{hf,(i)})\right)\right)^2 + R_{\boldsymbol{\beta}_0}(\boldsymbol{\beta})\right)$$

✗ Disjoint optimization of the mapping parameters and the multi-fidelity model,
✗ Training only on the lower fidelity input space.

Proposed method: MF-DGP integrated projection

- MF-DGP integrated projection (MF-DGP-IP) is a two levels DGP with:
  - First level contains multi-output GPs $\{H_l(\cdot)\}_{l=1}^{s-1}$ that are used as non-linear mappings between the different fidelity input spaces,
  - Second level contains single-output GPs $\{f_l(\cdot)\}_{l=1}^s$ used to propagate the fidelities evaluations.

✓ Uncertainty quantification on the input space mapping,
✓ Joint optimization of the mapping and the multi-fidelity model,
✓ Correlation computed in the original input space of each fidelity.

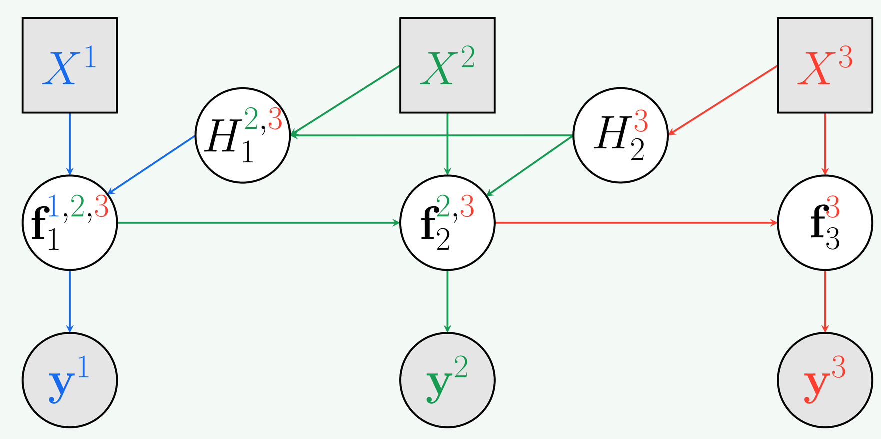

Figure 4:MF-DGP-IP graphical representation

The ELBO is obtained using a similar variational approximation as in MF-DGP:

$$\mathcal{L} = \sum_{t=1}^{s}\sum_{i=1}^{n_t} \mathbb{E}_{q(f_t^{(i),t})}\left[\log p(y^{(i),t}|f_t^{(i),t})\right] - \sum_{l=1}^{s} KL\left[q(\mathbf{u}_l)||p(\mathbf{u}_l; Z_{l-1})\right] - \sum_{l=1}^{s-1} KL\left[q(V_l)||p(V_l; Z_{l+1})\right]$$

Results MF-DGP-IP

Problem 1:

- Four-dimensions HF function $f_{hf}(x_1, x_2, x_3, x_4)$: Park HF function
- Two-dimensions LF function $f_{lf}(x_1, x_2)$: Park LF function with $x_3 = 0.5, x_4 = 0.5$.

Problem 2:

- Three-dimensions HF function $f_{hf}(r, \theta, \phi)$:

$$3.5\left(r\cos\left(\frac{\pi}{2}\theta\right)\right) + 2.2\left(r\sin\left(\frac{\pi}{2}\theta\right)\right) + 0.85\left(\left|r\cos\left(\frac{\pi}{2}\theta\right) - 2r\sin\left(\frac{\pi}{2}\theta\right)\right|\right)^{2.2} + \frac{2\cos(\pi\phi)}{1 + 3r^2 + 10\theta^2}$$

- Two-dimensions LF function $f_{lf}(x_1, x_2)$:

$$3x_1 + 2x_2 + 0.7(|x_1 - 1.7x_2|)^{2.35}$$

Table 1:Performance using 20 repetitions with different DoE (30 inputs data on LF and 5 inputs data on HF)

	Problem 1			Problem 2		
Algorithms	MNLL	RMSE	std RMSE	MNLL	RMSE	std RMSE
HF model	2.1e3	2.785	1.242	86.96	0.829	0.382
MF-DGP-IP	**16.84**	**1.787**	**0.692**	**1.02**	**0.663**	**0.11**
MF-DGP IMC	33.7	2.54	0.9560	5.562	0.7632	0.167
MF-DGP nominal	40.034	2.367	1.248	2.701	0.8996	0.196
AR1	1.6e5	2.7801	0.8975	98.31	0.736	0.148
NARGP	1383	2.6754	1.023	2221.	0.81	0.272

Contributions

- Improvement of MF-DGP in term of prediction accuracy and robustness to DoE,
- Proposition of a MF-DGP model for varying input spaces,
- Numerical simulations to assess the performance with respect to classical methods.