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

Ali Hebbal
ONERA, Inria, France
ali.hebbal@onera.fr

Loic Brevault
ONERA, France
loic.brevault@onera.fr

Mathieu Balesdent
ONERA, France
mathieu.balesdent@onera.fr

El-Ghazali Talbi
Inria Lille, France
el-ghazali.talbi@univ-lille.fr

Nouredine Melab
Inria Lille, France
nouredine.melab@univ-lille.fr

Abstract

Multi-fidelity approaches improve the inference of a high-fidelity model which is constructed using a small set of accurate observations, by taking advantage of its correlations with a low-fidelity model built using a larger set of approximated data. Most existing multi-fidelity methods consider the inputs of the low and high fidelity models defined identically over the same input space. However, it happens that the low fidelity model variables are defined over a different space than the variables of the high fidelity model due to different modeling approaches *i.e.* input spaces with different dimensionality and different nature of the variables. Recently, Deep Gaussian Processes have been used to exhibit the correlations between the low and high fidelity models. In this paper, Deep Gaussian Processes for multi-fidelity (MF-DGP) are extended to the case where the input spaces of the low and high fidelity models are different. Moreover, the learning capacity of MF-DGP is improved by proposing an optimization approach for the inducing inputs and by using natural gradients for the variational distributions of the inducing variables which also allows time reduction in the training.

1 Introduction

High-fidelity (HF) models are constructed using accurate data of the inputs/outputs to be observed. However, obtaining a real observation is usually an expensive task, hence the HF models are based on a limited dataset. To reinforce these models, approximations of the real response to be observed are made in order to construct low-fidelity (LF) models with a larger set of data. The correlations between the LF and HF models are exhibited within a multi-fidelity model enabling the improvement of the high-fidelity prediction. Gaussian Processes (GPs) [1] are a popular approach for multi-fidelity modeling. The Auto-Regressive (AR1) [2][3] approach assigns a GP prior to each fidelity t, where the HF prior $f_t(\cdot)$ is equal to the LF prior $f_{t-1}(\cdot)$ multiplied by a scaling factor ρ plus an additive bias function $\gamma_t(\cdot)$:

$$f_t(\mathbf{x}) = \rho f_{t-1}(\mathbf{x}) + \gamma_t(\mathbf{x}), \forall \mathbf{x} \in \mathbb{R}^d \tag{1}$$

This formulation assumes only a linear relationship between the fidelities. A more global approach considers the HF prior equal to a non-linear transformation of the LF prior by a GP prior $g_t(\cdot)$ plus an additive bias function:

$$f_t(\mathbf{x}) = g_t\left(f_{t-1}(\mathbf{x})\right) + \gamma_t(\mathbf{x}), \forall \mathbf{x} \in \mathbb{R}^d \tag{2}$$

4th workshop on Bayesian Deep Learning (NeurIPS 2019), Vancouver, Canada.

This relationship has given rise to two approaches: the Non-linear Autoregressive multi-fidelity GP model (NARGP) [4] and the Multi-Fidelity Deep Gaussian Process model (MF-DGP) [5]. The NARGP simplifies the relationship by considering each fidelity as a transformation of the GP posterior of the lower fidelity, which enables the GPs to be trained sequentially. MF-DGP, on the other hand, keeps the exact relationship, which comes back to a Deep Gaussian Process (DGP) [6] where each layer corresponds to a fidelity level. MF-DGP proves to have better accuracy tthe interactions between the fidelities [5]. MF-DGP is based on the sparse DGP approximation proposed in [7]. However, since the inputs of the intermediate layers are the combination of the points in the original input space with their corresponding function evaluations, freely optimizing the inducing inputs is not adequate. In [5] the inducing inputs are fixed to arbitrary values. Moreover, the optimization of the variational distributions is performed using ordinary gradients, which can be inappropriate when optimizing a distribution as mentioned in [8]. The first contribution of this paper is a new optimization approach for the inducing inputs of the intermediate layers, and the use of natural gradients in the optimization of the variational distributions to get a better variational lower bound in less iterations.

All the approaches presented previously assume that the input spaces of all the fidelities are defined identically in terms of input variables. However, this is not always the case. In fact, due to either different modeling approaches from one fidelity to another, or an omission of some variables in the lower fidelity models, the input spaces may differ in the form of the parametrization and also in the dimensionality. The second contribution of this paper is to generalize MF-DGP to the case with different input spaces. This is accomplished by a new model formulation of MF-DGP incorporating the mapping between the input spaces in a non-parametric way, based only on the nominal values of the mapping of the inputs data from one fidelity to another.

2 Improvements of MF-DGP

2.1 Description of Multi-Fidelity DGP

Let consider s levels of fidelity, and n_t inputs $X^t \in \mathcal{M}_{n_t,d}$ and their corresponding evaluations $\mathbf{y}^t \in \mathbb{R}^{n_t}$ at each fidelity $t \in \{1, \dots, s\}$. A DGP is considered where each layer corresponds to a fidelity. Moreover, the GP at each layer depends not only on the input data at this fidelity but also on all the previous fidelity evaluations for the same input data. To this end, $\mathbf{f}_l^t$ denotes the evaluation at the layer l of X^t the input data at the fidelity t (Fig. 2). This formulation of DGPs imposes the definition of a combination of covariance functions at each layer taking into account the correlation between the inputs as well as the correlation between the outputs:

$$k_l(\mathbf{x}^{(i)}, \mathbf{x}^{(j)}) = k_l^{\rho}(\mathbf{x}^{(i)}, \mathbf{x}^{(j)}) k_l^{f-1}\left(f_{l-1}^*(\mathbf{x}^{(i)}), f_{l-1}^*(\mathbf{x}^{(j)})\right) + k_l^{\gamma}(\mathbf{x}^{(i)}, \mathbf{x}^{(j)}) \tag{3}$$

where $f_{l-1}^*(\cdot)$ denotes the posterior of the GP at the layer $l-1$, k_l^{ρ} and k_l^{γ} are respectively an input space-dependent scaling factor and an input space-dependent bias, while k_l^{f-1} is the covariance between the evaluated outputs at the previous layer.

The DGP approximations follow the variational approximation presented in [7]. At each layer a set of inducing inputs / outputs $(Z_l, \mathbf{u}_l)$ are introduced and the following variational approximation is considered:

$$q\left(\{\{\mathbf{f}_l^t\}_{l=1}^t\}_{t=1}^s, \{\mathbf{u}_l\}_{l=1}^s\right) = \prod_{t=1}^{s}\prod_{l=1}^{t} p(\mathbf{f}_l^t | \mathbf{u}_l; \{\mathbf{f}_{l-1}^t, X^t\}, Z_{l-1}) \times \prod_{l=1}^{s} q(\mathbf{u}_l) \tag{4}$$

where $q(\mathbf{u}_l)$ is the approximated variational distribution of $\mathbf{u}_l$. The variational evidence lower bound (ELBO) is then obtained:

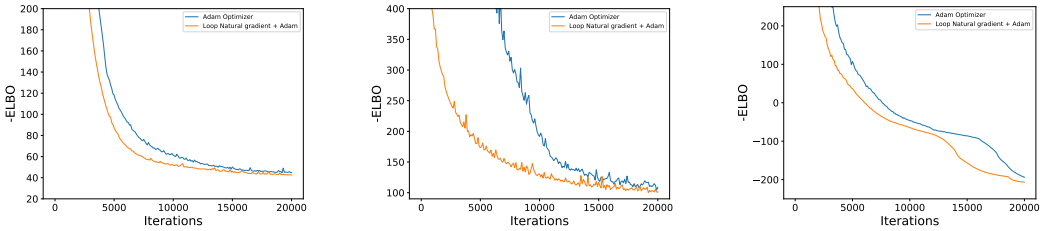
$$\mathcal{L}_{MF-DGP} = \sum_{t=1}^{s}\sum_{i=1}^{n_t} \mathbb{E}_{q(f_t^{(i),t})}\left[\log p(y^{(i),t} | f_t^{(i),t})\right] - \sum_{l=1}^{s} KL\left[q(\mathbf{u}_l) || p(\mathbf{u}_l; Z_{l-1})\right] \tag{5}$$

where KL corresponds to Kulblack-Leibler divergence. More details in this derivation can be found in [5], [7]. This bound is optimized with respect to $\{Z_l, q(\mathbf{u}_l)\}_{l=1}^s$ and the GP hyperparameters at each layer $\{\Theta\}_{l=1}^s$. However, optimizing the variational distributions $q(\mathbf{u}_l)\}_{l=1}^s$ using ordinary gradient can be not appropriate. Moreover, in the case of MF-DGP, the inputs at each layer are a combination of inputs in the original input space with the outputs of the previous layer, hence, optimizing freely the inducing inputs is not adequate.

2.2 First improvement of MF-DGP

The inference in MF-DGP using the sparse variational GP approximation comes back to maximizing the ELBO with respect to the hyperparameters of the GPs $\{\Theta\}_{l=1}^{s}$ the induced inputs $\{Z_l\}_{l=1}^{s}$ and also the variational distributions $\{q(\mathbf{u}_l)\}_{l=1}^{s}$. Including the variational distributions makes the parameter space not Euclidian, hence, the ordinary gradient is not the suitable direction to follow. In fact, the variational distribution parameter space has a Riemannian structure defined by the Fisher information [8]. In this case, the natural gradient which comes back to the ordinary gradient rescaled by the inverse Fisher information matrix, is the steepest descent direction. Natural gradients were used in the case of conjugate variational inference in GP [9] and also in the non-conjugate case [10] where an efficient computation has been proposed. A generalization to MF-DGP is proposed in this work. Specifically, an optimization procedure is used, consisting of a loop between an optimization step using a stochastic ordinary gradient (Adam Optimizer [11]) with respect to the Euclidian space parameters $(\{\Theta\}_{l=1}^{s}, \{Z_l\}_{l=1}^{s})$ and an optimization step using the natural gradient with respect to all the variational distributions $(q(\mathbf{u}_l)\}_{l=1}^{s})$ (lines 8,9 of Algorithm 1) .

This optimization approach was compared to the classic training method using only Adam Optimizer [5] on three different multi-fidelity problems (Currin [12], Park[12], Branin [4]). Using natural gradient shows faster convergence of the ELBO than the former approach (Fig 1).



(a). Currin (b). Park (c). Branin

Figure 1: Comparison of the evolution of the optimization of the ELBO on three different multi-fidelity problems using the Adam Optimizer and the proposed approach based on natural gradient

2.3 Second improvement of MF-DGP

One of the major difficulties in MF-DGP is the optimization of the inducing inputs $\{Z_l\}_{l=1}^{s}$. In [5] the inducing inputs were arbitrary fixed and not optimized. In fact, except for the first layer, the inducing inputs in MF-DGP do not play the same role as in classic DGPs, where they are defined in the original input space. Specifically, the input space of the inner layers of the MF-DGP is augmented with the output of the previous layer, inducing a non-linear dependence between the d first components and the $d+1$ component of each element in this augmented input space. Hence, freely optimizing Z_l (with $2 \leq l \leq s$) as vectors with independent components is no longer suitable.

To overcome this issue, $\{Z_l\}_{l=2}^{s}$ are constrained as follows:

$$Z_l = \left[Z_{l,1:d}, f_{l-1}^{*}(Z_{l,1:d})\right]; \forall 2 \leq l \leq s \tag{6}$$

where $f_{l-1}^{*}(\cdot)$ corresponds to the prediction at the previous layer. This constraint keeps a dependency between $Z_{l,d+1}$ and $Z_{l,1:d}$, allowing to remove $Z_{l,d+1}$ from the expression of the ELBO. Hence the optimization is done with respect to $Z_{l,1:d}$ instead of Z_l (lines 7,8 of Algorithm 1) .

These two proposed improvements are summarized in Algorithm 1. Combining this optimization approach of the inducing inputs along with the use of natural gradients for the variational distributions, improves the learning capacity of MF-DGP as shown in Table 1. The proposed approach has been compared with AR1, NARGP and regular MF-DGP in the same benchmark of 4 analytical functions as used in [5] (Currin, Park, Borehole [12], Branin). For AR1, NARGP and regular MF-DGP, the 'emukit' library [13] is used. 20 repetitions on different Design of Experiments (DoE) have been performed. The improved MF-DGP shows the best results in prediction accuracy and in uncertainty quantification with a large robustness to different DoE. On the Borehole problem , it gives comparable results to the AR1. This is explained by the fact that the Borehole problem shows strong linearity

Algorithm 1 ELBO Optimization

Initialization of the number of maximum iterations. *maxiter*
Initialization of the hyperparameters of the kernels $\{\Theta^{(0)}\}_{l=1}^{s}$.
$q(\mathbf{u}_l^{(0)}) \leftarrow \mathcal{N}\left(\mathbf{y}^l, \Sigma_l^{(0)}\right), \forall 1 \leq l \leq s$. For optimization stability $\Sigma_l^{(0)}$ is fixed at low values.
$Z_{l,1:d}^{(0)} = X^l, \forall 1 \leq l \leq s$
$i \leftarrow 0$
while $0 \leq i \leq$ *maxiter* **do**
  $Z_l^{(i)} \leftarrow \left[Z_{l,1:d}^{(i)}, f_{l-1}^{*}(Z_{l,1:d}^{(i)})\right]; \forall 2 \leq l \leq s$
  $ELBO^{(i+1)}, \{Z_{l,1:d}^{(i+1)}\}_{l=1}^{s}, \{\Theta^{(i+1)}\}_{l=1}^{s} \leftarrow$ Adam step $\left(ELBO^{(i)}, \{\Theta^{(i)}\}_{l=1}^{s}, \{Z_{l,1:d}^{(i)}\}_{l=1}^{s}\}\right)$
  $ELBO^{(i+1)}, \{q(\mathbf{u}_l^{(i+1)})\}_{l=1}^{s} \leftarrow$ Natural gradient step $\left(ELBO^{(i+1)}, \{q(\mathbf{u}_l^{(i)})\}_{l=1}^{s}\right)$
  $i \leftarrow i + 1$
end while
return $ELBO^{(i)}, \{Z_{l,1:d}^{(i)}\}_{l=1}^{s}, \{\Theta^{(i)}\}_{l=1}^{s}, \{q(\mathbf{u}_l^{(i)})\}_{l=1}^{s}$

Table 1: Performance of the different multi-fidelity models on 4 different problems using 20 repetitions with different DoE. R^2 refers to the R squared error, MNLL to the mean negative test log likelihood, RMSE to the root mean squared error, and std to the standard deviation. Currin and Park ($D_{in} = 2$) problems are modeled with 12 inputs data on the LF and 5 inputs data on the HF. Borehole ($D_{in} = 8$) is modeled with 60 inputs data on the LF and 5 inputs data on the HF. Branin ($D_{in} = 2$) is used with 80 inputs data on the lower fidelity, 30 on the medium fidelity and 5 inputs data on the higher fidelity.

Approach	AR1				NARGP			
Functions	R^2	MNLL	RMSE	std RMSE	R^2	MNLL	RMSE	std RMSE
Currin	0.8994	46.400	0.7355	0.2131	0.8743	123.398	0.8147	0.2572
Park	0.9831	299.463	0.5809	0.2100	0.8792	213.759	1.0986	1.2420
Borehole	**0.9998**	**-3.777**	**0.0047**	**0.00097**	0.9968	-2.503	0.0226	0.0120
Branin	0.1944	15810.3	0.1765	0.0658	0.0241	4285.26	0.2034	0.0344
Approach	MF-DGP				MF-DGP improved			
Functions	R^2	MNLL	RMSE	std RMSE	R^2	MNLL	RMSE	std RMSE
Currin	0.8856	1.6834	0.7427	0.3398	**0.9148**	**1.4165**	**0.6735**	**0.2056**
Park	0.8436	1.1616	1.1364	1.496	**0.9852**	**0.8807**	**0.5693**	**0.0969**
Borehole	0.9986	-2.006	0.0168	0.0032	0.9994	-2.733	0.0107	0.0016
Branin	0.3592	3.5977	0.1541	0.0665	**0.5865**	**5.0382**	**0.1256**	**0.0480**

between the fidelities. The results given by the improved MF-DGP are better than the regular MF-DGP on the four problems, in prediction accuracy, uncertainty quantification and robustness to DoE.

3 MF-DGP with different input spaces

3.1 Problematic

To handle varying input spaces for the different fidelities, a nominal mapping (based on theoritical insight of the multi-fidelity problem) is required that expresses the relationship between the different input spaces. It is possible to define a more adapted mapping between fidelities in order to exhibit missing physics between fidelities. The Input Mapping Calibration (IMC) is a recent approach [14] consisting in finding such a mapping $g_{\boldsymbol{\beta}}(\cdot)$ that is parametric. The parameters are obtained by minimizing the difference between the LF and HF model outputs on the HF data points. A regularization

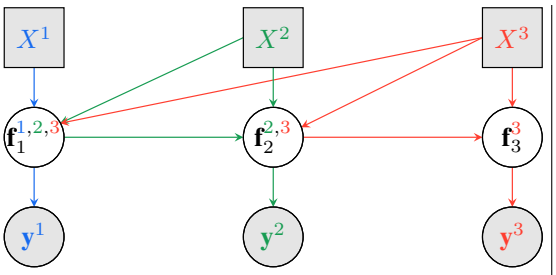


Figure 2: Graphical representation of MF-DGP model.

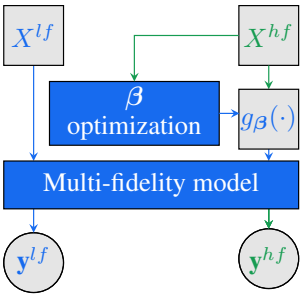


Figure 3: Graphical representation of the IMC approach.

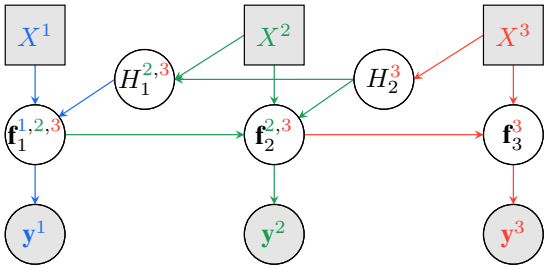


Figure 4: Graphical representation of MF-DGP with different input spaces.

term $R(\boldsymbol{\beta}, \boldsymbol{\beta}_0)$ can be taken into consideration based on the nominal mapping parameters $\boldsymbol{\beta}_0$:

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

where n_{hf} is the number of high fidelity training data points. The high-fidelity data is then projected with the obtained mapping on the low-fidelity input space, and a multi-fidelity model with the same input spaces can be used (Fig 3). However, this optimization of the mapping parameters is done previously to the training of the multi-fidelity model, which prevents the parameters of the mapping to be updated, once the model is optimized. Moreover, the correlations over the original HF input space are not taken into account, since the multi-fidelity model is trained only on the lower fidelity input space.

3.2 Proposed model

Let $(X^t, \mathbf{y}^t)$ be the couple of inputs/outputs of each fidelity $t \in \{1, \dots, s\}$, where s is the number of fidelities. Let D_t be the dimension of the input data of fidelity t. Since each fidelity is defined on its own input space, MF-DGP can not be used directly. To overcome this issue, a GP mapping between the input spaces is considered. Specifically, a GP $H_t(\cdot)$ is used as a mapping between the input spaces of two successive fidelities t and $t+1$. The input mapping GPs have parametric mean functions which are initialized on the nominal mapping functions. The model obtained is a two-level DGP, where the first level maps between the different fidelity input spaces and the second level propagates the fidelities evaluations (Fig 4). Hence, the mapping between the input spaces of the fidelities is defined within the multi-fidelity model. This allows a joint optimization of the mapping and of the multi-fidelity model. Moreover, using a GP as a mapping induces an uncertainty quantification and avoids over-fitting compared to parametric mapping. Finally, this model keeps the original input space correlations, since X^l is used as input for f_l, unlike IMC where the projection of X^l on the lower fidelity input space is used.

The latent variables involved in this two-level DGP are $\{\{\mathbf{f}_l^t\}_{l=1}^t\}_{t=1}^s, \{\mathbf{u}_l\}_{l=1}^s$ and $\{\{H_l^t\}_{l=1}^t\}_{t=1}^s, \{V_l\}_{l=1}^{s-1}$, where $H_t^t = X^t$ and V_l are the inducing variables introduced at the mapping

layer l, and W_{l+1} their corresponding inducing inputs. A similar sparse variational approximation to the one used in MF-DGP is followed:

$$q\left(\{\{\mathbf{f}_l^t\}_{l=1}^t\}_{t=1}^s, \{\mathbf{u}_l\}_{l=1}^s, \{\{H_l^t\}_{l=1}^t\}_{t=1}^{s-1}, \{V_l\}_{l=1}^{s-1}\right) = \prod_{t=1}^{s}\prod_{l=1}^{t-1}\left[p(\mathbf{f}_l^t|\mathbf{u}_l; \{\mathbf{f}_{l-1}^t, H_l^t\}, Z_{l-1})\right.$$
$$\left. p\left(H_l^t|V_l; H_{l+1}^t, W_{l+1}\right)\right] \times \prod_{t=1}^{s} p(\mathbf{f}_t^t|u_t; \{\mathbf{f}_{t-1}^t, X^t\}, Z_{t-1}) \times \prod_{l=1}^{s} q(\mathbf{u}_l) \times \prod_{l=1}^{s-1} q(V_l) \tag{7}$$

The variational lower bound on the marginal likelihood is then obtained:

$$\begin{aligned}\mathcal{L} = &\sum_{t=1}^{s}\sum_{i=1}^{n_t} \mathbb{E}_{q(f_t^{(i),t})}\left[\log p(y^{(i),t}|f_t^{(i),t})\right] - \sum_{l=1}^{s} KL\left[q(\mathbf{u}_l)||p(\mathbf{u}_l; Z_{l-1})\right] \\ &- \sum_{l=1}^{s-1} KL\left[q(V_l)||p(V_l; Z_{l+1})\right]\end{aligned} \tag{8}$$

The prediction of a test data X^{*,t_0} belonging to the input space of fidelity t using the two-level MF-DGP is straightforward. This consists in propagating the test data X^{*,t_0} through the first level of the DGP allowing the projection of the test data on the lower fidelity inputs spaces to obtain $H_{t_0-1}^{*,t_0}, \ldots, H_1^{*,t_0}$, then a propagation through the second level to propagate the evaluation at the different fidelities. Hence, a prediction of X^{*,t_0} with fidelity t is:

$$q(\mathbf{f}_t^*) = \frac{1}{k}\sum_{j=1}^{k} q\left(\mathbf{f}_t^{j,*}|q(\mathbf{u}_t); \{\mathbf{f}_{t-1}^{j,*}, H_t^{j,*,t_0}\}, Z_{t-1}\right) \tag{9}$$

Where k is the number of samples propagated using a Monte Carlo sampling.

3.3 Numerical experiments

To evaluate the efficiency of the proposed model and to compare it with the reference approaches (IMC coupled to AR1, NARGP, MF-DGP), two analytical test cases have been defined.
Problem 1: The first test case is the Park multi-fidelity problem [12] where the low fidelity is considered only with two variables (Eq. 10 and Eq. 11). This problem describes the case where the low fidelity does not take into account some variables in the modeling process. The nominal mapping is naturally the identity mapping of the HF variables (Eq. 12).
The high-fidelity function is four dimensional with an input domain $[0, 1]^4$:

$$f_{hf}(x_1, x_2, x_3, x_4) = \frac{x_1}{2}\left(\sqrt{1 + (x_2 + x_3^2)\frac{x_4}{x_1^2}} - 1\right) + (x_1 + 3x_4)\exp\left(1 + \sin(x_3)\right) \tag{10}$$

The low-fidelity function is two dimensional with an input domain $[0, 1]^2$:

$$f_{lf}(x_1, x_2) = \left(1 + \frac{\sin(x_1)}{10}\right) f_{hf}(x_1, x_2, 0.5, 0.5) - 2x_1 + +x_2^2 + 0.75 \tag{11}$$

The used nominal mapping is a linear mapping $X^T A_0 + \mathbf{b}_0$ with:

$$A_0 = \begin{bmatrix}1 & 0\\ 0 & 1\\ 0 & 0\\ 0 & 0\end{bmatrix} \text{ and } \mathbf{b}_0 = [0, 0] \tag{12}$$

Problem 2 : The second test case is a problem describing the case where the input spaces have different parametrization defining variables with different nature (cartesian and polar parametrization), in addition to different dimensionality (Eq. 13 and Eq. 14). The nominal mapping is a linear transformation based on first order Taylor series of the HF variables (Eq. 15):

The high-fidelity function is three dimensional with an input domain $[0,1]^3$:

$$f_{hf}(r,\theta,\phi) = 3.5\left(r\cos\left(\frac{\pi}{2}\theta\right)\right) + 2.2\left(r\sin\left(\frac{\pi}{2}\theta\right)\right) + 0.85\left(\left|r\cos\left(\frac{\pi}{2}\theta\right) - 2r\sin\left(\frac{\pi}{2}\theta\right)\right|\right)^{2.2} + \frac{2\cos(\pi\phi)}{1+3r^2+10\theta^2} \tag{13}$$

The low-fidelity function is two dimensional with an input domain $[0,1]^2$:

$$f_{lf}(x_1,x_2) = 3x_1 + 2x_2 + 0.7(|x_1 - 1.7x_2|)^{2.35} \tag{14}$$

The used nominal mapping is a linear mapping $X^T A_0 + \mathbf{b}_0$ with:

$$A_0 = \begin{bmatrix} \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \\ \frac{-\pi\sqrt{2}}{8} & \frac{\pi\sqrt{2}}{8} \\ 0 & 0 \end{bmatrix} \text{ and } \mathbf{b}_0 = [\frac{\pi\sqrt{2}}{16}, -\frac{\pi\sqrt{2}}{16}] \tag{15}$$

Table 2 presents the results obtained by the proposed approach (MF-DGP Integrated Projection *MF-DGP-IP*) and a comparison to the results obtained by AR1, NARGP, MF-DGP coupled with the IMC approach and also with MF-DGP using only the nominal mapping. The experimentation has been repeated on 20 different DoE. On the two problems, MF-DGP-IP is more efficient than the IMC approach on the prediction and also in uncertainty quantification and is also more robust to the DoE. Moreover, it is interesting to observe that MF-DGP-IP enables the discovery of hidden physics which were not described in the nominal mapping since the results obtained are better than using only MF-DGP with only the nominal mapping.

Table 2: Performance of the different multi-fidelity models on the test problems using 20 repetitions with different DoE. R^2 refers to the R squared error, MNLL to the mean negative test log likelihood, RMSE to the root mean squared error. 30 inputs data on the LF and 5 inputs data on the HF are used.

	Problem 1			
Algorithms	R^2	MNLL	RMSE	std RMSE
HF model	0.5729	2082.116	2.7852	1.2424
MF-DGP-IP	**0.8312**	**16.8401**	**1.7868**	**0.69224**
MF-DGP IMC	0.661398	33.702	2.5399	0.95609
MF-DGP nominal	0.67055	40.034	2.367	1.2485
AR1	0.6076	161220	2.7801	0.89725
NARGP	0.62281	1383	2.6754	1.02314
	Problem 2			
Algorithms	R^2	MNLL	RMSE	std RMSE
HF model	0.67619	86.9156	0.82865	0.38172
MF-DGP-IP	**0.82173**	**1.02519**	**0.66339**	**0.11**
MF-DGP IMC	0.76628	5.56258	0.76232	0.1672
MF-DGP nominal	0.6756	2.7010	0.8996	0.1958
AR1	0.78467	98.3183	0.73618	0.14806
NARGP	0.73274	2221.3	0.80993	0.272

4 Conclusion

The contribution of this paper is twofold: Firstly, improvements of MF-DGP were presented, overcoming its previous limitations. Experiments on analytical problems have demonstrated the improvements of its prediction accuracy, uncertainty quantification and robustness to DoE. Secondly, a generalization of MF-DGP to different definitions of input spaces, was accomplished by proposing a two-level MF-DGP. Experimentations on analytical test cases show promising results of the proposed approach. Next steps will consist in the application to real world engineering problems to assess the performance in complex cases and also to problems with more than two fidelities.

References

[1] Carl Rasmussen and Christopher KI Williams. *Gaussian processes for machine learning*, volume 1. MIT press Cambridge, 2006.

[2] Marc C Kennedy and Anthony O'Hagan. Bayesian calibration of computer models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 63(3):425–464, 2001.

[3] Loic Le Gratiet and Josselin Garnier. Recursive co-kriging model for design of computer experiments with multiple levels of fidelity. *International Journal for Uncertainty Quantification*, 4(5), 2014.

[4] Paris Perdikaris, Maziar Raissi, Andreas Damianou, ND Lawrence, and George Em Karniadakis. Nonlinear information fusion algorithms for data-efficient multi-fidelity modelling. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 473(2198):20160751, 2017.

[5] Kurt Cutajar, Mark Pullin, Andreas Damianou, Neil Lawrence, and Javier González. Deep gaussian processes for multi-fidelity modeling. *arXiv preprint arXiv:1903.07320*, 2019.

[6] Andreas Damianou and Neil Lawrence. Deep gaussian processes. In *Artificial Intelligence and Statistics*, pages 207–215, 2013.

[7] Hugh Salimbeni and Marc Deisenroth. Doubly stochastic variational inference for deep gaussian processes. In *Advances in Neural Information Processing Systems*, pages 4588–4599, 2017.

[8] Shun-Ichi Amari and Scott C Douglas. Why natural gradient? In *Proceedings of the 1998 IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP'98 (Cat. No. 98CH36181)*, volume 2, pages 1213–1216. IEEE, 1998.

[9] James Hensman, Nicolo Fusi, and Neil D Lawrence. Gaussian processes for big data. *arXiv preprint arXiv:1309.6835*, 2013.

[10] Hugh Salimbeni, Stefanos Eleftheriadis, and James Hensman. Natural gradients in practice: Non-conjugate variational inference in gaussian process models. In *Artificial Intelligence and Statistics*, 2018.

[11] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

[12] Shifeng Xiong, Peter ZG Qian, and CF Jeff Wu. Sequential design and analysis of high-accuracy and low-accuracy computer codes. *Technometrics*, 55(1):37–46, 2013.

[13] Andrei Paleyes, Mark Pullin, Maren Mahsereci, Neil Lawrence, and Javier González. Emulation of physical processes with emukit. In *Second Workshop on Machine Learning and the Physical Sciences, NeurIPS*, 2019.

[14] Siyu Tao, Daniel W Apley, Wei Chen, Andrea Garbo, David J Pate, and Brian J German. Input mapping for model calibration with application to wing aerodynamics. *AIAA Journal*, pages 2734–2745, 2019.