The 8th European Congress on Computational Methods in Applied Sciences and Engineering
ECCOMAS Congress 2022
5-–9 June 2022, Oslo, Norway

ADVANCED EXPERIMENTS ON GAUSSIAN PROCESS-BASED MULTI-FIDELITY METHODS OVER DIVERSE MATHEMATICAL CHARACTERISTICS

Sihmehmet Yildiz[1], Hayriye Pehlivan-Solak[1], Matteo Diez[2],Omer Goren[1] & Melike Nikbay[1]

[1] Istanbul Technical University, Istanbul, 34469, Turkey
{yildizsih, pehlivanha, ogoren, nikbay}@itu.edu.tr

[2]CNR-INM, National Research Council- Institute of Marine Engineering, Rome, Italy
matteo.diez@cnr.it

Key words: Multi-fidelity, Gaussian Process, CoKriging, NARGP

Abstract. Advanced applications of multi-fidelity surrogate modelling techniques provide significant improvements in optimization and uncertainty quantification studies in many engineering fields. Multi-fidelity surrogate modelling can efficiently save the design process from the computational time burden caused by the need for numerous computationally expensive simulations. However, no consensus exists about which multi-fidelity surrogate modelling technique usually exhibits superiority over the other methods given for certain conditions. Therefore, the present paper focuses on assessing the performances of the Gaussian Process-based multi-fidelity methods across selected benchmark problems, especially chosen to capture diverse mathematical characteristics, by experimenting with their learning processes concerning different performance criteria. In this study, a comparison of Linear-Autoregressive Gaussian Process and Nonlinear-Autoregressive Gaussian Process methods is presented by using benchmark problems that mimic the behaviour of real engineering problems such as localized behaviours, multi-modality, noise, discontinuous response, and different discrepancy types. Our results indicate that the considered methodologies were able to capture the behaviour of the actual function sufficiently within the limited amount of budget for 1-D cases. As the problem dimension increases, the required number of training data increases exponentially to construct an acceptable surrogate model. Especially in higher dimensions, i.e. more than 5-D, local error metrics reveal that more training data is needed to attain an efficient surrogate for Gaussian Process based strategies.

1 INTRODUCTION

The exploration of multi-fidelity design methods by the North Atlantic Treaty Organization (NATO) Science and Technology Organization (STO) AVT-331 Research Task Group (RTG) on "Goal-Driven, Multi-Fidelity Approaches for Military Vehicle System-Level Design" [1] has motivated the development of benchmark problems with varying levels of complex problems. AVT 331-Level 1 (L1) subgroup has focused on the development of simplified analytical benchmark problems to explore multi-fidelity methods [2]. This paper presents Gaussian Process

(GP) based multi-fidelity methods comparison using "Level 1" benchmark problems following the terminology of Beran et al.[1].

In recent years, there has been a significant development in data-based modelling approaches and as a result of this, surrogate modelling studies gained importance. Among these developments, the concept of multi-fidelity modelling is of great interest to the computational science and engineering communities. Multi-fidelity methods rely on mathematical algorithms to combine the responses obtained from different fidelity levels to reduce the computational load. Several survey studies were conducted for multi-fidelity methods [3, 4] to assess their advantages and disadvantages for diverse problems. However, the capabilities of multi-fidelity methods are still under discussion and their potential is being explored. In addition, no consensus exists about which multi-fidelity surrogate modelling technique usually exhibits superiority over the other methods given for certain conditions. This situation raises interest in benchmark problems that can support the comparative and rigorous evaluation of the methods and in sharing the application of methods to these benchmark problems.

The GP method is one of the popular surrogate modelling methods used in different engineering fields due to its predictive capability and the predicted variance estimation. Because of their similar advantages, GP-based multi-fidelity methods are popular methods used for data fusion between different fidelities. Many GP-based multi-fidelity applications are based on the linear-autoregressive data fusion scheme proposed by Kennedy and O'Hagan [5]. As observed in many studies in the literature, the study of Forrester et al. [6] and Perdikaris et al. [7] presents the success of the linear-autoregressive scheme-based multi-fidelity GP (ARGP) methods for low- and high-fidelity model outputs that exhibit strong linear correlation across the input domain. Recently, nonlinear fusion schemes based on GP (NARGP) have been introduced [8] to improve their prediction capabilities. These approaches can offer alternatives to standard ARGP methods and capture nonlinear relationships between fidelity models.

This paper explores GP-based multi-fidelity methods on "Level 1" benchmark functions that mimic features of real-world engineering problems. The aim of the paper is to present the advantages and disadvantages of GP-based multi-fidelity methods depending on the diverse mathematical characteristics of the selected benchmark functions. In addition, we would like to shed light on the cases for which GP-based multi-fidelity methods will give better results depending on the mathematical characteristics of the implemented case.

The remaining of the paper is as follows. In Section 2, Gaussian Process and multi-fidelity implementations are explained. In Section 3, benchmark problems are introduced and then surrogate model assessment metrics and experimental setups used for comparisons are explained. In Section 4, the surrogate model metrics obtained for different benchmark problems are presented. In the final section, the results are discussed.

2 METHODOLOGY

In this section, the concepts and basic terminology of a Gaussian Process are explained and then the multi-fidelity implementations, Linear-Autoregressive and Nonlinear-Autoregressive Gaussian Processes, are presented. For a more detailed introduction and description of the models, the following references can be consulted [6, 8, 9].

2.1 Gaussian Process

The Gaussian process regression is used to approximate a latent function $f(\boldsymbol{x})$ with compromising input-output pairs $D = \{x_i, y_i\}$ $(i = 1, ..., N)$. Modelling the latent function with the Gaussian process instead of an analysis program can be expressed as in Eq. (1).

$$\mathbf{y} = f(\boldsymbol{x}) + \epsilon, \quad \epsilon \sim \mathcal{N}\left(\mathbf{0}, \sigma^2\mathbf{I}\right) \tag{1}$$

The latent function $f(\boldsymbol{x})$ is modeled by a Gaussian process equation with 0 mean and specific r variances and can be expressed as in Eq. (2).

$$f(\boldsymbol{x}) \sim \mathcal{GP}\left(\mathbf{0}, r\left(\boldsymbol{x}, \boldsymbol{x'}; \boldsymbol{\theta}\right)\right) \tag{2}$$

In Eq. (2), r is the covariance function while typically the squared exponential kernel covariance function is expressed as in Eq. (3), where $\boldsymbol{\theta} = (\boldsymbol{\alpha}, \boldsymbol{\beta})$ are the hyper-parameters.

$$r\left(\boldsymbol{x}, \boldsymbol{x'}; \boldsymbol{\theta}\right) = \boldsymbol{\alpha}^2 \exp\left(-\frac{1}{2}\sum_{d=1}^{D}\frac{\left(x_d - x'_d\right)^2}{\beta_d^2}\right) \tag{3}$$

The hyper-parameter, $\boldsymbol{\theta}$, and noise variance, σ^2, are obtained by minimizing the negative log marginal likelihood equation given in Eq. (4), where $\mathbf{R} = r(\boldsymbol{x}, \boldsymbol{x}; \boldsymbol{\theta}) + \boldsymbol{\sigma}^2\mathbf{I}$.

$$\mathcal{NLML}(\boldsymbol{\theta}) = \frac{1}{2}\boldsymbol{y}^T\mathbf{R}^{-1}\boldsymbol{y} + \frac{1}{2}\log|\mathbf{R}| + \frac{N}{2}\log(2\pi) \tag{4}$$

It is difficult to find optimum hyper-parameters by taking the partial derivative of Eq. (4), so the study finds optimum hyper-parameters using an optimization algorithm. With the obtained optimum hyper-parameters, the predictive value and variance value at a new test point $(\boldsymbol{x}^*)$ are obtained using Eq. (5) and (6).

$$\mathrm{f}\left(\boldsymbol{x}^*\right) = r\left(\boldsymbol{x}^*, \boldsymbol{x}\right)\mathbf{R}^{-1}\boldsymbol{y} \tag{5}$$

$$\sigma^2(\boldsymbol{x}^*) = r\left(\boldsymbol{x}^*, \boldsymbol{x}^*\right) - r\left(\boldsymbol{x}^*, \boldsymbol{x}\right)\mathbf{R}^{-1}k\left(\boldsymbol{x}, \boldsymbol{x}^*\right) \tag{6}$$

2.2 Linear-Autoregressive Gaussian Process

The Gaussian Process can be further constructed to enable systematic combination of variable fidelity models. The Linear-Autoregressive method is the most widely used in engineering design problems for multi-fidelity modelling of GP. It is based on a Linear-Autoregressive information fusion scheme introduced by Kennedy and O'Hagan [5], assuming a linear dependence between different model fidelities. According to this approach, the prediction in the highest fidelity model can be written as in Eq. (7).

$$f_{hf}(\boldsymbol{x}) = \rho f_{lf}(\boldsymbol{x}) + f_d(\boldsymbol{x}) \tag{7}$$

In this equation, $f_{hf}(\boldsymbol{x})$ and $f_{lf}(\boldsymbol{x})$ denote the high- and low-fidelity models at multi- and low-fidelity levels, ρ is a scaling constant that measures the correlation between different fidelity responses. Next, $f_d(\boldsymbol{x})$ is the GP model that represent discrepancy between $f_{hf}(\boldsymbol{x})$ and $\rho f_{lf}(\boldsymbol{x})$.

2.3 Nonlinear-Autoregressive Gaussian Process

The linear mapping between different fidelities may not be appropriate for some engineering design problems. In order to generalize the linear-autoregressive approach, Perdikaris et al. [8] proposed a non-linear mapping between the fidelities called NARGP. According to this approach, the prediction in the highest fidelity model can be written as in Eq. (8).

$$f_{hf}(\boldsymbol{x}) = z(f_{lf}(\boldsymbol{x})) + f_d(\boldsymbol{x}) \tag{8}$$

where, $f_{hf}(\boldsymbol{x})$ and $f_{lf}(\boldsymbol{x})$ denote the high- and low-fidelity models and $f_d(\boldsymbol{x})$ is the GP model that represents the discrepancy between $f_{hf}(\boldsymbol{x})$ and $z(f_{lf}(\boldsymbol{x}))$. Unlike the linear model, in this method, a mapping is applied with the low-fidelity model with the stochastic GP method ($z()$) to capture the nonlinear relationships between the models.

3 NUMERICAL EXPERIMENT SETUP

This chapter covers the recommended setup for the experiments to assess and compare the considered multi-fidelity methods for the selected set of benchmark functions. The following sections provide an overview of assessment criteria, introducing benchmark set and associated mathematical characteristics, experimental setup such as; computational budget and fidelity cost assignment, sampling, termination criteria and test data set.

3.1 Assessment Metrics

In order to compare the accuracy of the surrogate models, some error metrics are calculated. The proposed metrics are selected to offer a comprehensive assessment of the surrogates in terms of local, global and goal-sensitive characteristics of the surrogate modelling ability. These accuracy metrics used in the presented study are root-mean-square error (ε_{RMSE}), maximum absolute error (ε_{MAE}) and goal-sensitive error (ε_t) (optimization accuracy metric).

- Root-Mean Square Error
  The root-mean square error is a global error metric that is used to measure the global success of surrogate models. The formulation can be expressed for the test data-set as follows,

$$\begin{aligned} \varepsilon_{RMSE} &= \sqrt{\frac{1}{N_{test}} \sum_{i=1}^{N_{test}} \left(f(x_i) - \hat{f}(x_i) \right)^2} \\ \varepsilon_{NRMSE} &= \frac{\varepsilon_{RMSE}}{(f_{\max} - f_{\min})} \end{aligned} \tag{9}$$

  where $f(x_i)$ is the exact value of the function and $\hat{f}(x_i)$ represents the surrogate model prediction. f_{min} and f_{max} represent the minimum and maximum values of the analytical function, respectively.

- Maximum Absolute Error
  The maximum absolute error is a local error metric that is used to measure the maximum

approximation error of surrogate models. The formulation can be expressed for the test data-set as follows,

$$\begin{aligned} \varepsilon_{MAE} &= \max\left(\left|f(x_i) - \hat{f}(x_i)\right|\right), \quad i = 1, 2, ..., N_{test} \\ \varepsilon_{NMAE} &= \frac{\varepsilon_{MAE}}{(f_{\max} - f_{\min})} \end{aligned} \tag{10}$$

- Goal Sensitive Error
  The distance-based metric (ε_t) defined by Serani et al. [10] is used to evaluate the optimization performance of the methods.

$$\varepsilon_t \equiv \sqrt{\frac{\varepsilon_x^2 + \varepsilon_f^2}{2}} \quad \text{where} \quad \varepsilon_x \equiv \sqrt{\frac{1}{D}\sum_{j=1}^{D}\left(\frac{g_j - x^*_{j,min}}{Z_j}\right)} \quad \textit{and} \quad \varepsilon_f \equiv \frac{f(\boldsymbol{g}) - f_{min}}{f_{max} - f_{min}} \tag{11}$$

  where ε_x is the normalized distance between the point representing the optimization variables at the global minimum found by the surrogate model ($\boldsymbol{g}$) and the point representing the optimization variables at the global minimum found by the analytical function (x^*_{min}). ε_f is the normalized difference between the analytical function evaluated at the optimum value of the design variables found by the surrogate model ($f(\boldsymbol{g})$) and the minimum of the analytical function (f_{min}). $Z_j = |u_j - l_j|$ is the range of j^{th} design variable.

3.2 Benchmark Problems

The multi-fidelity benchmark functions are selected to capture fundamental mathematical characteristics and properties which mimic real-world engineering problems. The distinctive mathematical features of the selected benchmark problems can be listed as; multi-modality, localized behaviours, scalability, discrepancy type and noise. The multi-modality behaviour of the considered benchmark function represents, whether it is uni/multi-modal, or shows localized behaviour and having discontinuity within its domain. If a function has multi-modality property, then the surrogate representation of the function becomes more challenging in terms of exploration of multi-modal regions. The discrepancy feature identifies the linear and nonlinear analytical correlation between the fidelity levels. Noise distribution is also taken into consideration within the scope of this study. Noise is expected in any kind of engineering response within a wide range of distribution due to reasons such as multiple physical models and solvers, space/time discretizations, levels of physical coupling to construct multi-fidelity approximations. For this reason, relative noise terms are added to both low and high-fidelity models. Considering all these mathematical characteristics over the benchmark problems will allow users to assess superiorities and weaknesses of GP-based multi-fidelity methods depending on the main mathematical characteristics of the problem. The benchmark functions determined by the AVT-331 RTG [2] and their main features are presented in Table 1.

The low-dimensional versions of the benchmark problems are represented in Figure 1. See [2] for the analytical definitions of the benchmark problems and for more detailed information. The fidelity number and the problem dimension of the benchmark problems used in this study are presented in Table 2.

Table 1: Main features of benchmark problems

Function	Behaviours	Scalability	Discrepancy	Noise
Forrester	Local / (Dis)continous	-	linear	no
Rosenbrock	Local	Parametric	nonlinear	no
Shifted-rotated Rastrigin	Multi-modal	Parametric / Fidelity	nonlinear	no
Heterogeneous	Local / Multi-modal	Parametric	nonlinear	no
Spring-Mass system	Multi-modal	Parametric / Fidelity	nonlinear	no
Paciorek	Multi-modal	Fidelity	nonlinear	yes

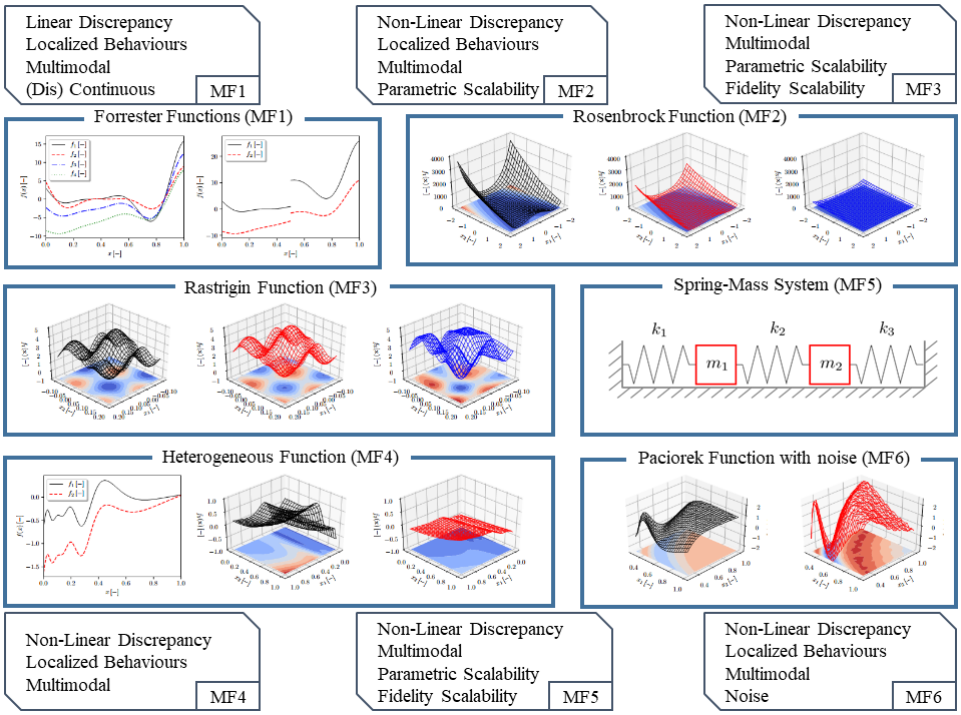

Figure 1: Benchmark problems.

3.3 Experimental Setup

Other experimental setups used in the comparison study are presented below.

- **Budget:** The maximum total budget is determined depending on the dimension of the selected benchmark function. It is given in terms of the highest fidelity simulation cost. The assigned budget values for the benchmark functions are presented in Table 2.
- **Fidelity cost assignment:** Fidelity cost assignment criteria are described as follows; high fidelity equations are considered as the unitary cost and lower fidelity models are

assigned as fractions of high fidelity costs. The cost for different fidelity of the benchmark functions is presented in Table 2.

- **Sampling method:** The Halton sequence distributes the samples homogeneously over the design area. As the number of samples increases, Halton sampling adds new sample points homogeneously distributed in the vacant regions of the design area while keeping the previous sample set the same [11]. For this reason, the Halton sampling method is used due to the success in filling the experimental space in the presented study.

- **Termination criteria:** Termination criteria is attained as the experiments are reached at the maximum assigned budget.

- **Surrogate model performance evaluation:** Fixed test points are determined by the Halton sampling method to assess the surrogate models. The number of test points used to test the surrogate models depends on the dimension, and the number of test points is determined as $N_{test} = 10000 \times D$.

Table 2: Summary of multi-fidelity benchmark problems.

Function	**D**	**Budget**	**Fidelity cost**			
			f_1	f_2	f_3	f_4
Forrester	1	100	1.00000E-0	5.00000E-1	1.00000E-1	5.0000E-2
Dis. Forrester	1	100	1.00000E-0	2.00000E-1	-	-
Rosenbrock	2	200	1.00000E-0	5.00000E-1	1.00000E-1	-
	5	500	1.00000E-0	5.00000E-1	1.00000E-1	-
	10	1000	1.00000E-0	5.00000E-1	1.00000E-1	-
Shifted-rotated Rastrigin	2	200	1.00000E-0	6.25000E-2	3.90625E-3	-
	5	500	1.00000E-0	6.25000E-2	3.90625E-3	-
	10	1000	1.00000E-0	6.25000E-2	3.90625E-3	-
Heterogeneous	1	100	1.00000E-0	2.00000E-1	-	-
	2	200	1.00000E-0	2.00000E-1	-	-
	3	300	1.00000E-0	2.00000E-1	-	-
Springs	2	200	1.00000E-0	1.66667E-2	-	-
Springs-masses	4	400	1.00000E-0	1.66667E-2	-	-
Paciorek	2	200	1.00000E-0	2.00000E-1	-	-

4 Results

The surrogate modelling performance of the GP method is assessed over a selected set of benchmark functions which covers quite challenging diverse mathematical characteristics. In order to give more precise comparison in terms of the advantages/disadvantages of the multi-fidelity strategies, GP methods are also compared with the single-fidelity versions. Error metrics are presented for variable computational budget values.

Among others, here, Forrester function is selected to present the results more explicitly with respect to the increasing computational budget. The surrogate model performance metrics obtained for the Forrester equation are presented in Figure 2. In this figure, multi-fidelity applications are performed using two fidelities, such that; MF2 shows surrogate model performance where f_1 is used as high fidelity and f_2 is used as low fidelity. MF3 shows surrogate model

performance where f_1 is used as high fidelity and f_3 is used as low fidelity. MF4 shows surrogate model performance where f_1 is used as high fidelity and f_4 is used as low fidelity. Here, it should be noted that MF2 will employ less number of LF samples when compared to MF3 and MF4 since its LF samples are still expensive compared to MF3 and MF4 ones. In Figure 2; in plot (f), the number of high fidelity samples with respect to the total number of samples is depicted in order to discriminate the ratio of the used high- and low-fidelity training data among the total computational budget for different multi-fidelity cases. It is clearly seen that the number of HF data used in MF2 case is much less compared to the other multi-fidelity cases and this dampens the advantage of flexible budget allocation between two fidelities. Plots (a), (b) and (c) stand for optimization accuracies (ε_x, ε_f ε_t), whereas plots (d) and (e) represent global error metrics lastly. In Figure 2, for MF3 and MF4, it is seen that as a higher computational budget is employed, the accuracy of the multi-fidelity approximation increases very sharply in the beginning as compared to the single-fidelity surrogate model and all models saturate at the very high computational budgets. In most of the multi-fidelity cases for the Forrester function, an accurate multi-fidelity surrogate is achieved even for low computational costs for MF3 and MF4. This shows the benefit of establishing a multi-fidelity surrogate model instead of a SF model when the computational budget is very limited.

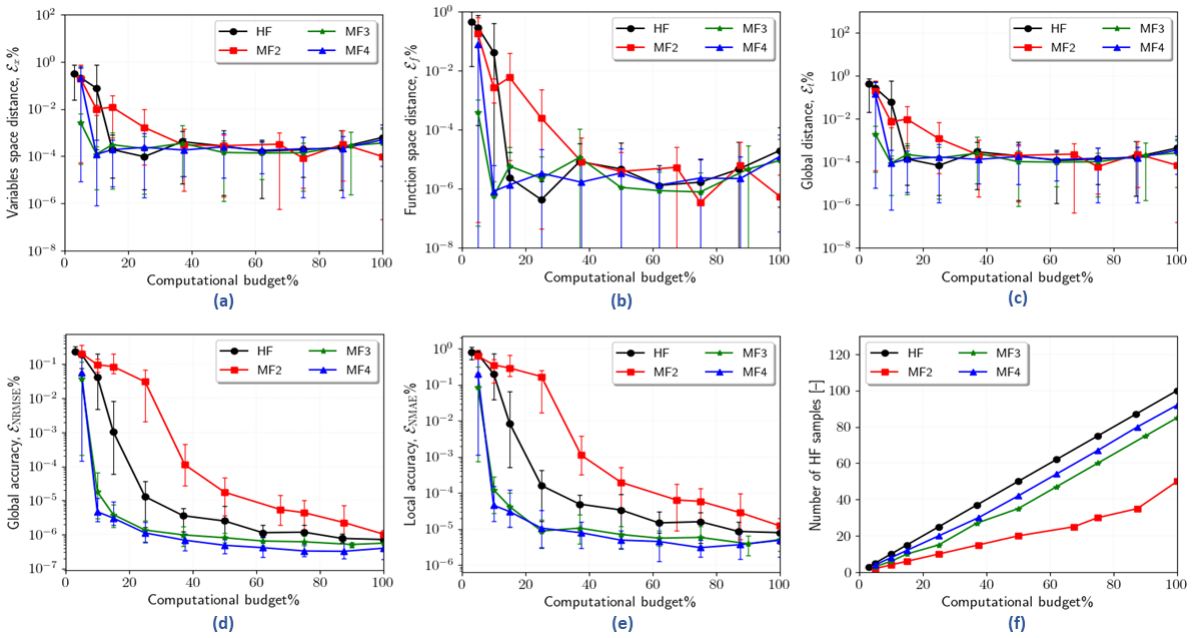

Figure 2: ARGP result for Forrester function.

The rest of the implementations are presented in tabulated form in Figure 3 and 4 covering all the individual applications within the study. Since the plots shared in Figure 2 are difficult to interpret, error metrics are tabulated for certain budget values such as; 5, 15, 25, 50, 75 and 100 with different colors.

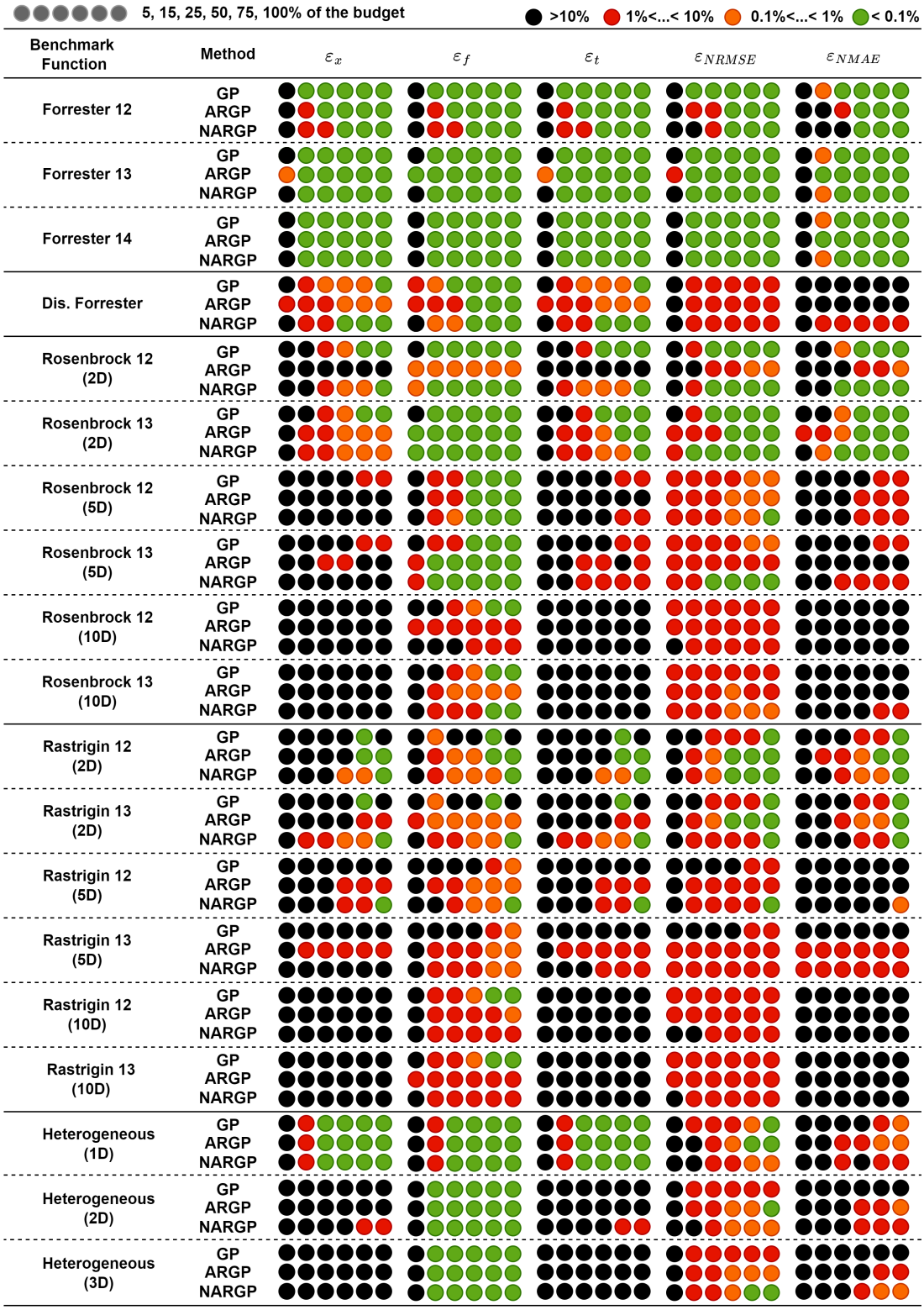

Figure 3: Comparison of surrogate model local/global modeling performances.

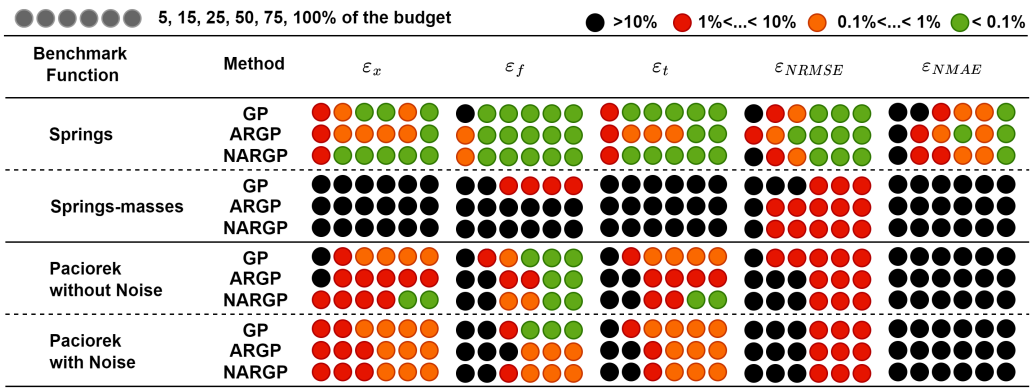

Figure 4: Comparison of surrogate model local/global modeling performances.

As seen in Figure 3, the Forrester function has linear discrepancy among different fidelity levels, which makes ARGP more appropriate to model with. Especially in Forrester 12 case, ARGP outperforms when it is compared to NARGP even though very little computational budget is used (see Figure 1 to evaluate the similarity between fidelity levels visually).

As seen from Figure 3, the Forrester function is approximated well with both single- and multi-fidelity GP whereas for the discontinuous case; optimization accuracy and global error metrics are reduced. This result shows that discontinuity could not be explored within the assigned computational budget. Moreover, as known from the related literature the recognition of a discontinuity strictly depends on the relevance of sampling locations and the location of the discontinuity. Therefore, the discontinuity topic will be further investigated comprehensively within a future study for different sampling types to exclude sampling dependency from the learning process. All in all, NARGP better performs in capturing the behaviour of discontinuity in the response of the function for all computational budget values compared to ARGP or single-fidelity GP.

For Rosenbrock, the challenge in constructing an approximation function is quite different because of modality. Here, when the dimension of the benchmark problem increases mathematical characteristics differ with the problem dimension where the problem shifts from uni-modal to multi-modal behavior. Accordingly, this characteristic difference has a major impact on the surrogate performance and it is seen from the results that are presented for different dimensions of Rosenbrock function. It is seen from Figure 3, 2-D and 5-D performance results are quite different due to the change in mathematical behaviour. In addition, for all multi-fidelity cases when an easier/cheaper function is used as low-fidelity (see Figure 1 to compare 2^{nd} and 3^{rd} fidelity levels) more accurate metrics are accomplished. As 3^{rd} fidelity is easier to approximate, the multi-fidelity implementations with this fidelity level attain more accurate metrics compared to 2^{nd} fidelity level. Moreover, the Rosenbrock function has localized behaviours, therefore the optimum location can be explored more easily as seen from the optimization metric ε_x. On the other hand, the optimum value of the function is not estimated sufficiently since the function has a valley-type optimum range (see Figure 1 for the function behaviour). According to the results, as the dimension increases both optimization accuracy and global error metrics are reduced. In addition, for an overall evaluation of the considered methods for the Rosenbrock function,

NARGP produces more accurate performance metrics due to the non-linear discrepancy between the fidelity levels.

Rastrigin function is very challenging to approximate with any kind of surrogate since it is multi-modal and has steep changes in its domain. Accordingly, it is very difficult to conclude that one surrogate is more successful than the others, however, it can be concluded that NARGP performs slightly better in capturing the global behaviour of the function. In addition, optimization accuracy metric in terms of the optimum value of the function (ε_f) is approximated sufficiently, whereas the optimum location of the function (represented by metric ε_x) is not resolved well except in some minor cases.

Heterogenous function results show that as the dimension increases optimization accuracy is reduced but further increase in dimension has no effect on capturing the optimum value of the function. For the global error metrics, it is very difficult to interpret a comprehensive evaluation. Since the function has steep changes in its domain, all performance results are similar.

The Springs function is a 2-D benchmark that is comparatively easy to model with surrogate approaches compared to the springs-masses version. On the one hand, the springs benchmark function is mostly well approximated both single- and multi-fidelity GP surrogates, on the other hand, the springs-masses version could not be modelled as in 2-D case. Even in the case of the whole computational budget is attained for springs-masses the performance metrics may not regard as acceptable to be represented as a successful surrogate.

Paciorek function has a non-linear discrepancy, therefore the results show that NARGP approximated it more accurately. Global error metrics produce similar performances for both ARGP and NARGP methods. The artificial noise has an effect on both the optimum location and the value of the function, whereas global accuracy metrics are not changed. The noise distribution change will be varied within a future study in order to present the noise-capturing performance of considered surrogates.

5 Conclusion

A comprehensive assessment of multi-fidelity strategies over selected set of benchmark problems, in parallel with the efforts in NATO STO AVT-331 Subgroup L1, is presented and discussed. Multi-fidelity implementations of GP which are called ARGP and NARGP are taken into consideration and their comparative performances are presented and discussed in terms of local, global, and optimization accuracy metrics. The results show that, when different fidelity levels have linear-discrepancy ARGP performs better in terms of both optimization and global error metrics. In addition, when the function has multi-modal property, it is difficult to generalize which method outperforms over another. The effect of artificial noise could not be sufficiently explored over a wide range of benchmark problems, this needs to be further studied in future research.

Acknowledgements

All authors would like to acknowledge the collaboration established among members of the NATO STO AVT-331 Research Task Group titled "Goal-driven, multi-fidelity approaches for military vehicle system-level design" co-chaired by Dr.Phil Beran and Dr.Matteo Diez under the framework of NATO Science and Technology Organization Applied Vehicle Technology Panel.

First and last authors would like to express their gratitude to TUBITAK for the research grant provided under the 218M471 TUBITAK 1001 project titled as "Development of Multi-fidelity and Multidisciplinary Methodologies Integrating Sonic Boom, Aeroelasticity and Propulsion System for Supersonic Aircraft Design". Second and fourth authors would like to acknowledge the financial support of TUBITAK for Energy Saving Techniques for Energy Efficient Vessels and Emission Reduction towards Green Shipping Project with grant number 218M487.

REFERENCES

[1] P. Beran, D. Bryson, A.S. Thelen, M. Diez, A. Serani, Comparison of Multi-Fidelity Approaches for Military Vehicle Design. In Proceedings of the AIAA AVIATION 2020 FORUM, Virtual Event, 15–19 June 2020.

[2] L. Mainini, A. Serani, M.P. Rumpfkeil, E. Minisci, D. Quagliarella, H. Pehlivan, S. Yildiz, S. Ficini, R. Pellegrini, F.D. Fiore, D. Bryson, M. Nikbay, M. Diez and P. Beran, Analytical Benchmark Problems for Multifidelity Optimization Methods, ArXiv preprint, *ArXiv: 2204.07867*, 2022.

[3] M. Giselle Fernandez-Godino, C. Park, N.H. Kim and R.T. Haftka, Issues in deciding whether to use multifidelity surrogates, AIAA Journal, 57(5):2039–2054, 2019.

[4] B. Peherstorfer, K. Willcox and M. Gunzburger, Survey of multifidelity methods in uncertainty propagation, inference, and optimization. Siam Review, 60(3):550–591, 2018.

[5] M. C. Kennedy and A. O' Hagan, Predicting the output from a complex computer code when fast approximations are available, *Biometrika*, Vol. **87**, pp. 1-13, 2000.

[6] A. Forrester, A. Sobester and A. Keane, *Engineering design via surrogate modelling: a practical guide*, John Wiley & Sons, 2008.

[7] P. Perdikaris, D. Venturi and G.E. Karniadakis, Multifidelity information fusion algorithms for high-dimensional systems and massice data sets, SIAM J. Sci. Comput. 38, B521–B538, 2016.

[8] P. Predikaris, M. Raissi, A. Damianou, N.D. Lawrence and G.E. Kardinadakis, Nonlinear information fusion algorithms for data-efficient multi-fidelity modelling, *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, Vol. **473**, 2017.

[9] L. Brevault, M. Balesdent and A. Hebbal, Overview of gaussian process based multi-fidelity techniques with variable relationship between fidelities *Aerospace Science and Technology*, Vol. **107**, 2020.

[10] A. Serani, C. Leotardi, U. Iemma, E.F. Campana, G. Fasano and M. Diez, Parameter selection in synchronous and asynchronous deterministic particleswarm optimization for ship hydrodynamics problems, *Applied Soft Computing, 49:313–334.*, 2016.

[11] S. Yildiz, H. Pehlivan Solak and M. Nikbay, Multi-Fidelity Low-Rank Approximations for Uncertainty Quantification of a Supersonic Aircraft Design, *Algorithms 2022, 15, 250. https://doi.org/10.3390/a15070250*