16th World Congress in Computational Mechanics (WCCM)
21-26 July 2024, Vancouver, Canada
A. Korobenko, M. Laforest, S. Prudhomme, R. Vaziri (Eds)

A NEW APPROACH TO COMBINE PHYSICS-BASED AND DATA-DRIVEN MODELS USING A LOCALISED TRUSTWORTHINESS METRIC

LARS BOGAERTS[1,3], MATTHIAS G.R. FAES[2] AND DAVID MOENS[1,3]

[1] Department of Mechanical Engineering - LMSD Division, KU Leuven
Jan De Nayerlaan 5, 2860 St-Katelijne-Waver, Belgium
https://iiw.kuleuven.be/onderzoek/R2D

[2] Chair for Reliability Engineering, TU Dortmund University
Leonhard-Euler-Strasse 5, 44227 Dortmund, Germany
https://lre.mb.tu-dortmund.de/

[3] Flanders Make@Kuleuven
Belgium

Key words: Grey-box model, Uncertainty propagation, multi-fidelity estimation, predictive process control

Abstract. Artificial Neural Networks (ANNs) can solve many (un)supervised learning tasks by virtue of the universal approximation theorem. In the context of on-line process control for manufacturing processes, ANNs are an ideal approach for e.g., on-line monitoring or prediction tasks. However, since they are trained on experimental input-output pairs, the governing physical relations are only implicitly included. This, for instance, can cause inaccuracies when extrapolating to out-of-sample data-points [1]. On the other hand, the numerical approximation of the governing physical laws via numerical methods holds strong potential for the accurate simulation of physical phenomena that occur during manufacturing processes. However, the corresponding computational effort is an impediment that arises with the need for numerous simulations [2]. This makes the application of such numerical schemes computationally intractable within an on-line monitoring context. As such, it is clear that ANNs and numerical simulation models have strong potential, but are fundamentally different models. However, their combination serves as a potentially efficient and accurate aggregated predictor, the so called grey-box model. Such grey-box model is based on highly efficient machine learning algorithms, the black-box member, and backed by validated data with respect to physics generated by the numerical model, the white-box member. A grey-box model capable of defining a trustworthy prediction, including a measurement of uncertainty on the estimator, remains challenging.

In this paper, we introduce an ensemble machine learning method based on the residual modelling approach [3] that is able to provide a best-estimate under scarce data. We do so by combining the predictions of the white-box, stemming from partial differential equations

implemented in a multi-physics numerical solver, with the black-box, representing the residual error between the white-box predictions and available experimental data. The novelty of this work is in the contribution of each counterpart in the grey-box. The weight of each counterpart in the ensemble model is based on distance metrics between available data in each model. The proposed grey box method indicates significant performance gains compared to either the performance of each counterpart, as well to the classic residual modelling approach.

1 INTRODUCTION

In the context of process monitoring, real-time quality estimation based on data-driven techniques are becoming ever more common [4, 5]. These approaches typically link process parameters and on-line measurements to product quality metrics in order to guard the process. In this respect, machine learning approaches yield very fast black-box models enabling on-line application for process control. However, multiple problems arise with these black-box models: (a) These models are known to have only limited value when extrapolation is required [6]. Their use is most relevant for well-confined and -controlled processes that enable the generation of a clear, industrially representative and comprehensive dataset for the model training. There are large discrepancies between a lab-environment or numerically-made dataset and an industrial dataset, which is prone to inaccuracy due to changing variables or boundary conditions that come with the large change in environment. (b) Quantitative measurement capabilities for the process response are often limited in industrial processes, hence eliminating the applicability of data-hungry algorithms such as most supervised deep-learning toolboxes [7].

On the other end of the spectrum, multi-physical models help to understand and give insight in the quality affecting phenomena. Defining a reliable 'white-box' model of a production process, however, proves to be extremely complex, as it often involves advanced non-linear models. The time to run one accurate, high resolution simulation typically takes several hours or even days. Consequently, the use of simulation nowadays remains limited to high-end products and processes that fail to meet specifications, but especially in large volume industry, experience on this topic remains scarce [8].

This paper elaborates the modelling methodology proposed for a robust 'grey-box' under scarce data. The idea for a method for scarce data arises from a bottom up approach. To achieve such a grey-box, an efficient approach is required to switch or incorporate two or more models. Current state-of-the-art literature to include physics, albeit in the form of ground truth data, partial differential equations or boundary conditions, ranges from very basic data augmentation models [9] to very complex systems, e.g., Physics Informed Neural Networks (PINN) [10], NeuralPDEs [11] or Neural Operators [12]. For data augmentation, the main principal is to include both data from an experimental campaign and from a validated physics based white-box, and to train a model on the union set of both subsets. On the other hand, a PINN model can extend a black-box architecture based on experimental data by adding the governing partial differential equations to include physics in the overarching model. Ultimately, both fundamentally

different models can be assigned the label of 'grey-box'. While very efficient methods have been studied in literature, there is often the assumption of a large stream of measurement data. This assumption is often required, e.g., with the use of neural network approaches, requiring a rather high amount of unique data-points. However, from a practical perspective, there are many arguments that obtaining such a dataset is not trivial. This is often the case where generating data is exceeding a threshold on a cost function, including, e.g., labour, destructive tests, materials, environmental aspects. Consequently, in this study, the 'black-box' model is required to operate with a restricted dataset. Furthermore, the specific case arises that the white-box is a rather expensive model, thus also providing data that is sparse in the time domain. The effect of both aforementioned assumptions are investigated in this paper. The rest of this paper is organised as follows: Section 2 starts with elaborating the general concept, and introduces the weighted residual model based on the shortcomings of the classic residual modelling approach. Section 3 provides a summary of the results on a 1D numerical case study. Finally, section 4 provides the main conclusions in this work.

2 WEIGHTED RESIDUAL MODELLING

This section introduces an approach based on the existing residual modelling approach [13] for an effective predictor by combining the predictions of the white-box, stemming from partial differential equations implemented in a multi-physics numerical solver, with the black-box, representing the residual error between the white-box predictions and available experimental data.

2.1 General bias correction

An efficient, comprehensive approach for a grey-box model is the use of residual modelling. The residual model consists of two terms, where one is adding knowledge from the white-box and the other from observed data. Thus, the resulting grey-box is a summation equivalent to the predictions of a white-box and the offset between the white-box predictions and any collected experimental data, mapped as the black-box model:

$$f_{GB}(x) = f_{WB}(x|x^s) + r(x|x^e, x^s), \tag{1}$$

where (x^s) represents numerical sample points, (x^e) represents experimental sample points, f_{WB} represents the white-box and $r(x|x^e, x^s)$ the residual between simulated and experimental data, modelled by a black-box that projects and interpolates the residual error. The magnitude of the black-box term can be interpreted as being the extent to which the measurements verify the prediction of the white-box counterpart. This is defined as:

$$r(x|x^e, x^s) = f_{BB}(x|x^e) - f_{WB}(x|x^s). \tag{2}$$

This modelling architecture is often applied in cases where the white-box part is only a rough estimation, often due to multiple assumptions to lower the computation cost or due to lack of knowledge on the process. Furthermore, it is vital that the available data is continuous,

enabling a very fine discretisation over the parameter domain. In this regard, the residual can correct the rough estimation of the white-box towards known observations. However, in the scenario where the white-box is representing a complex high dimensional, non-linear model, it becomes infeasible to establish a dense sampling in the overall parameter domain. As such, in the scenario of limited, discontinuous data, the white-box is presented as a surrogate model by means of a Gaussian process (GP) model:

$$f_{WB}(x^s) \sim GP_w(m(x^s), k(x^s, x'^s)), \tag{3}$$

where $m(x^s)$ is the mean function and $k(x^s, x'^s)$ a positive definite kernel function. For simplicity in this work, the subscript $_w$ is used to refer to the black box, e.g., μ_w, σ_w, respectively the mean and standard deviation of the Gaussian process. The GP structure has become a standard in many applications because of its powerful, non-parametric regression nature which has been developed considerably within the machine learning community in the last 15 years, thereby having a pedigree in terms of response surface modelling and sensitivity analysis [14]. There is extended literature on applications [15, 16], as well as further developments of the method, such as PCE-kriging [17, 18], co-kriging [19] or structural reliability analysis [20]. The main reasons why GP are of great value in this work, are the minimal requirement for prior knowledge, its capabilities under the presence of noise, and the fact that GP does not only provide an interpolation technique, but it automatically provides confidence intervals for the model predictions as well. It is worth mentioning that although GP is ideal for this work, there are drawbacks, such as it's difficulty in handling high dimensional data and large datasets, where none of the aforementioned criteria is relevant in the context of this work.

Next, the residual is modelled as the error between the surrogate of the white-box and the black-box counterpart, whereas the latter is based on incomplete or unevenly distributed measurement data, which is also interpolated by means of a GP model:

$$f_{BB}(x^e) \sim GP_b(m(x^e), k(x^e, x'^e)). \tag{4}$$

For simplicity in this work, the subscript $_b$ is used to refer to the black box, e.g., μ_b, σ_b, respectively the mean and standard deviation of the Gaussian process. In the context of FE modelling, residual modelling is often known as 'bias correction', which tries to improve the fitting of a white-box model, that shows discrepancies between model simulations and observations. This definition acknowledges that there is likely some error in the complex FE representation, possibly originating from wrong or conservative assumptions, often in the case of epistemic uncertainty in the corresponding structure.

In the case where a white-box can provide continuous data, e.g. by relatively simple, yet inaccurate analytical equations, the bias correction is an improvement of the solution space, by correcting the solution with known observations.

However, in the context of limited observations, the accuracy of a bias correction approach might suffer from inter- or extrapolation error stemming from the black-box. Imagine the extreme scenario where only two deterministic experimental observations are available. The best

fit for a possible interpolation method, without any additional physical knowledge on the structure, is a linear combination between the points. Suppose a white-box counterpart, which does include the physical knowledge on the structure, is calibrated on the known observations. Additional points are propagated through the white-box model and a well-fit meta model is constructed to interpolate the results from the model, thereby showing that there is a clear nonlinear trend between the two known experimental observations. Following the definition of a bias correction, the continuous function $f_{WB}(x^s)$ is enriched with the knowledge from $f_{BB}(x^e)$, being the linear function between the set of two point, thereby neglecting the extra information generated by the white-box. As a result, by applying the residual model, the additional information is lost. This proves that the problem arises that a well performing residual model is neglecting the additional value of the extra propagation of the white-box in the case of scarce, incomplete or unevenly distributed data.

2.2 A weighted residual

As a response to the problem described in section 2.1, where scarce data can cause potential inter- or extrapolation errors, a solution is required that enables the traditional residual modelling to comply with the assumption of limited data. In this paper, we propose to achieve this by assigning a weight to both the response surface on the limited experimental and numerical data in the total predictor. Therefore, we propose to extend Eq.(1) with a control parameter α such that

$$f_{GB}(x) = f_{WB}(x|x^s) + \alpha(r(x|x^e, x^s)). \tag{5}$$

Substituting with Eq.(2) yields:

$$f_{GB}(x) = (1-\alpha) f_{WB}(x|x^s) + \alpha f_{BB}(x|x^e). \tag{6}$$

We propose to add the the control parameter α, which is a measure defined by the variance of the GP representing f_{BB}. The idea is inspired by the intuition of Kalman filters, where the state update equation is based on a Kalman gain K_n [21]:

$$\hat{x}_{n,n} = (1-K_n)\hat{x}_{n,n-1} + K_n z_n, \tag{7}$$

where K_n is the measurement weight and $(1-K_n)$ a weight for the current state estimate. To quote the definition of the Kalman gain: *"The Kalman Gain is close to zero when the measurement uncertainty is high and the estimate uncertainty is low. Hence we give a significant weight to the estimate and a small weight to the measurement. On the other hand, when the measurement uncertainty is low, and the estimate uncertainty is high, the Kalman Gain is close to one. Hence we give a low weight to the estimate and a significant weight to the measurement." [21].* It is clear that the proposed α in this work is analogue to K_n in the state space update equation used in Kalman filters. However, while Kalman filters aim at providing a next step prediction in the presence of uncertainty, our work aims at giving a prediction in the presence of uncertainty based on a mix of two surrogate models over a predefined parameter domain. To define

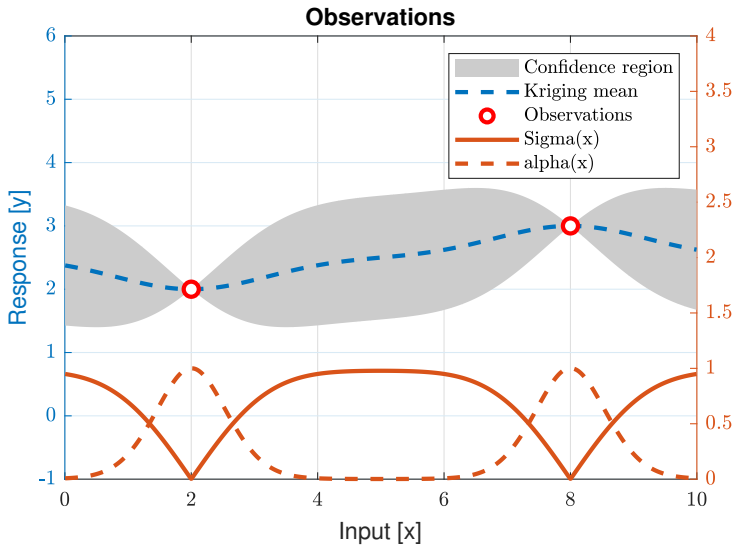

Figure 1: Introduction of alpha, based on the distance measure derived from a kriging interpolator

α, we propose a measure for the uncertainty of the experimental data based on the estimate given by the GP representing f_{BB}. The variance of the GP is defined in terms of a correlation function R and its hyper-parameters θ. The correlation function $R = R(x, x'; \theta)$ describes the correlation between two sample points in the output space, that depends on x, x', and the hyper-parameters θ. Therefore α becomes a distance measure, since the variance of the GP model describes "similarity" between observations and new points, i.e., how similar such points are, depending on the distance between the input points. E.g., regions of denser observations yield a high confidence, allowing the residual model to add more of the residual term. Regions with a lack of observations, both inter- or extrapolating, yield a lower level of confidence, declining the residual model to add its residual to the white-box information. Figure 1 illustrates a 1D example with two observations, on which a GP model is fit. The grey area is the predicted confidence region, stemming from the correlation between the sample points in the output space. The variance is extracted and visualised as an orange curve, to clearly visualise effect on the proposed α weight, visualised as an orange dashed curve.

2.3 Properties

We propose the hypothesis that the variance, resulting from the GP as a surrogate model on the measurement data, is the best estimator for a fidelity of the black-box. Albeit the estimator has to be selected carefully, as the GP model has variation stemming from the correlation function R. Eq.(8) describes a smooth-step function, ranging from $\sigma^2_{b,min}$ to $\sigma^2_{b,max}$, and acting as a

power law for σ_b^2 and α:

$$\alpha(\sigma_b^2) = \begin{cases} \alpha_{max} & \sigma_b^2 < \sigma_{b,min}^2 \\ \mathcal{S}_m(\alpha_{max} - \alpha_{min}) + \alpha_{min} & \sigma_b^2 \in \left[\sigma_{b,min}^2, \sigma_{b,max}^2\right] \\ \alpha_{min} & \sigma_b^2 > \sigma_{b,max}^2 \end{cases} \tag{8}$$

with $\mathcal{S}_m$ a sigmoid-like interpolation and α ranges from α_{min} to α_{max}. $\sigma_{b,min}^2$ and $\sigma_{b,max}^2$ represent the parameters that define the bounds to fine tune the trade-off between the models. Without prior knowledge on the process or system, the value of $\sigma_{b,min}^2$ and $\sigma_{b,max}^2$ is set equal to the upper and lower limit of the variance in $f_{BB}(x^e)$. The hypothesis, Eq.(6) and Eq.(8) allow to set up some ground rules for the concept:

1. $\sigma_b^2 = \sigma_{b,min}^2 \Rightarrow f_{GB}(x) = f_{BB}(x|x^e), \alpha = 1$
2. $\sigma_b^2 = \sigma_{b,max}^2 \Rightarrow f_{GB}(x) = f_{WB}(x|x^s), \alpha = 0$
3. $\sigma_g^2 \geq \sigma_{b,min}^2$
4. $\alpha \in [0, 1]$

where rule $n.1$ & $n.2$ describe the extreme scenario's where the weight is either minimal or maximal and neglects information from $f_{WB}(x|x^s)$ in rule $n.1$ and vice versa. Rule $n.3$ & $n.4$ specifically aim at a Gaussian process regression (GPR) model, similar to the described GP in this work, however, including an extra term for added noise in the system. For a simple GP problem, as provided in section 3, α is set to range between zero and one. However, in more complex scenario's, one might prefer one model over the other, such that one model, regardless of the availability of measurement data, is part of the grey box. This can be made feasible by diminishing the range of α by increasing the lower or reducing the upper bound. By doing so, either rule $n.1$ or $n.2$ is unachievable, thereby the condition that the grey box takes the form of the corresponding white or black box, becomes excluded.

In the case of sparse data, where the proposed grey-box largely depends on the interpolation performance of both surrogate models, it is vital for the proposed approach to make a distinction between ground truth and artificial data, that, although generated by the governing physical laws of the process, might drift due to extrapolation on the deterministically validated model. To incorporate this effect, we propose to add a 'confidence index' to the confidence bounds on the grey box model. this confidence index makes the distinction between a single model that is trained with a dataset consisting of the combination of experimental and numerical data (augmented model) and the proposed weighted combination of two surrogate models. The confidence index acts as a safety, where the confidence region enlarges according to the already present weight parameter α. Thus, the variance in the resulting grey model is a trade-off between confidence on measurement data and uncertainty arising on the extrapolation of a well calibrated white-box model:

$$\sigma_g^2 = (1 - \alpha)(\sigma_w^2 + \gamma) + \alpha.\sigma_b^2, \tag{9}$$

with Eq.(9) analogue to Eq.(6), but including an extra term with γ, a confidence index based on the trustworthiness in the white-box. The confidence index, according to Eq.(9) an absolute combination with the weighted variance from both models, requires a specific value in the same magnitude as the predicted variance over the response. It should be noted that the confidence is strictly case dependent and requires expert knowledge on the subject. According to the authors, a good baseline for this value is the mean variance over the complete parameter domain. This effect is depicted in figure 2. Two individual models, referred to as the white surrogate model and the black surrogate model are both fitted on several deterministic points. The black surrogate model is fitted on p points $x_i = \{x|0 < x \leq 11, i = 1, \ldots, p\}$ and the white surrogate model is fitted on q points $x_i = \{x|11 \leq x < 15, i = 1, \ldots, q\}$ with $x = 11$ the intersection of both sets. It is noted that the variance increases in function of the distance between points. When extrapolating, e.g., the black-box above $x = 11$ and vice versa, the variance reaches a maximum. The confidence of the grey-box, based on Eq.(9) is a hybrid form of both models, thus resulting in a variance that considers all data-points. However, due to the confidence index, based on the knowledge that the FE-model accounts for possible drift, there is a noticeable disadvantageous effect when interpolation is also exceeding α_{max}, due to γ acting as a confidence index of the white-box. When comparing to the 'augmented model', it can be noted that the model treats all points as equal, hence not assigning a weight to each model.

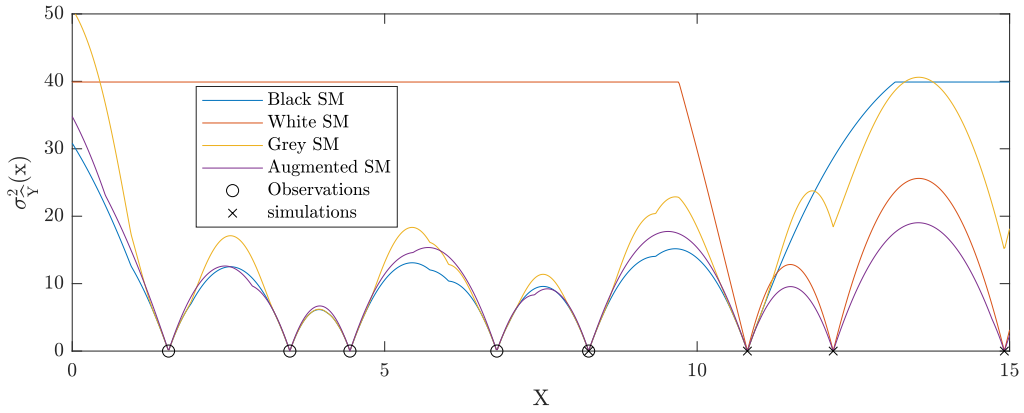

Figure 2: The expected variance, given by a two independent GP models and their hybrid form

3 ILLUSTRATIVE EXAMPLE

The performance of the proposed grey-box methodology is assessed based on a numerical example, representing a random generated higher order polynomial sampled between zero and 3π. Figure 3 illustrates the steps in generating the grey-box and visually compares the proposed grey-box methodology with the original residual modelling approach. The underlying true function is given in semi-opaque as a dotted line. The first figure illustrates the example where 10 samples are used to fit a black-box model, where the samples are only selected until

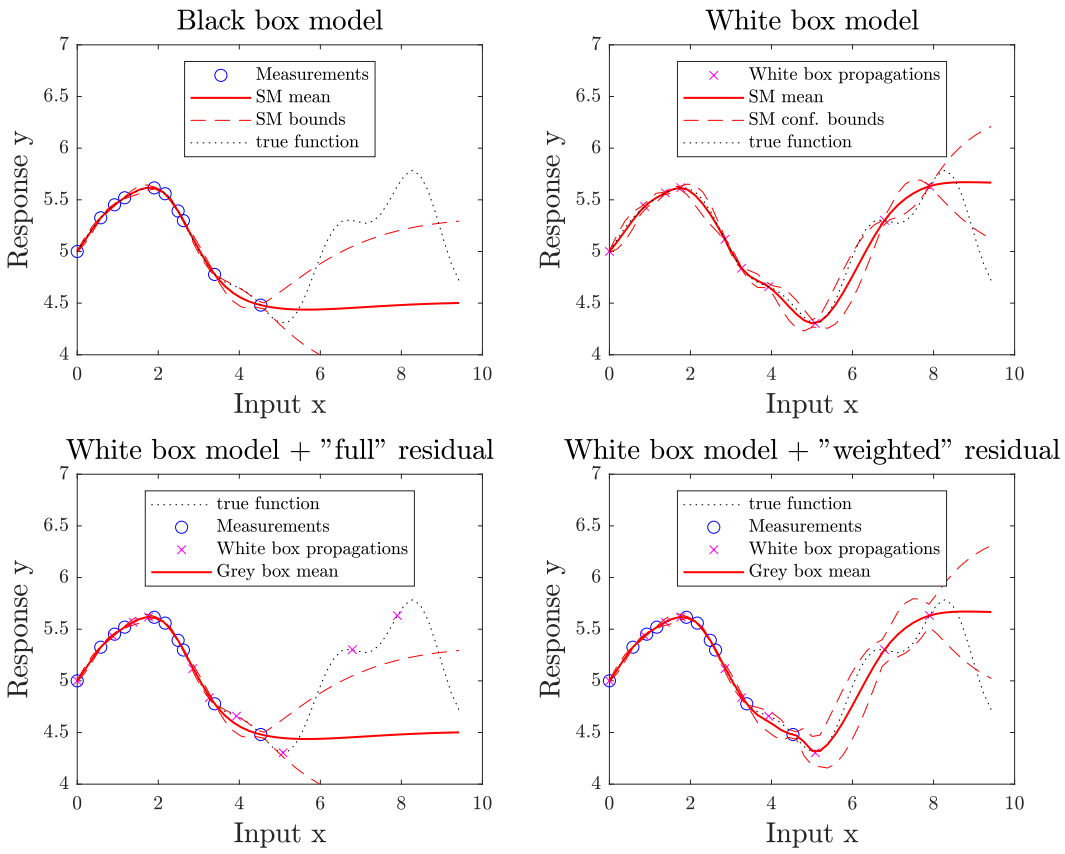

Figure 3: Surrogate model in terms of: observations [$n = 10$] (top-left), white-box results [$n = 10$] (top-right), residual model (bottom-left), weighted residual model (bottom-right)

$x = 2\pi$. It can be noted that the model has a dense region of data-points, hence the very small confidence intervals. When the interpolation error increases, e.g., between $x = 3.7$ and $x = 4.9$, and especially during extrapolation, $x > 2\pi$, the confidence bounds indicate a build-up of lack of confidence. Furthermore, when no data are present, GP revert to its prior trend, in this case a constant mean function. The second graph visualises the surrogate model acting as the white-box, which is fitted using another 10 samples, sampled more uniformly between $x = 0$ and $x = 3\pi$. The bottom left graph represents a classic residual model, according to Eq.(1). Due to the lack of experimental data, the residual term that is mapped by the black-box revert to it's prior knowledge, however it does not incorporate the fact that the surrogate model reverted to its mean prior, hence the lack of accurate predictions at $x = 5$ and higher. To evaluate the performance of a surrogate model, the Mean square error (MSE) is a widely adopted metric, expressed as:

$$\left(y_* - f(\mathbf{x}_*)\right)^2, \tag{10}$$

Table 1: Effect of sample-sizes

No. BB.	No. WB	MSE BB	MSE WB	MSE GB	MSE αGB	MSLL GB	MSLL αGB
10	10	0.3075	0.0453	0.3075	0.0449	-0.4684	-0.9978
5	30	0.4244	0.00394	0.4244	0.00002724	5.91	-1.25
30	20	0.3583	0.0034	0.3583	0.0033	13.8479	-2.2402
10	200	0.3075	0.0020	0.3075	0.0003396	22.8580	-2.0682
200	10	0.3451	0.1198	0.3451	0.0892	7.4435	0.3283

with y_* the measured signal and $f(\mathbf{x}_*)$ the model model prediction. The MSE of the model results in 0.3075, which is actually worse than the MSE of the SM representing the white-box (MSE WB = 0.0453) In contrast, the last figure illustrates the performance of the proposed weighted residual model, clearly indicating a mix of the black and white-box surrogate model according to Eq.(6). In conditions where experimental data are available, they serve as a ground truth. As soon as the confidence interval increases, the information of the white-box receive a higher fidelity, albeit with a wider confidence interval compared to the surrogate model that is fitted on the data of the individual model, due to the added term consisting of the confidence index to the grey-box. The proposed model results in a MSE of 0.0449, only a slight improvement over the white-box (MSE WB = 0.0453). However, since MSE does only provide information on the mean value, a second metric, the Mean Standardised Log Loss (MSLL), introduced by Rasmussen et al. [6] provides a probabilistic measure based on both the mean and confidence level of a surrogate model.

$$-\log p\left(y_*|\mathcal{D}, x_*\right) = \frac{1}{2}\log\left(2\pi\sigma_*^2\right) + \frac{\left(y_* - \bar{f}(\mathbf{x}_*)\right)^2}{2\sigma_*^2}, \tag{11}$$

where the predictive variance σ_*^2 is computed as $\sigma_*^2 = \mathbb{V}(f_*) + \sigma_n^2$, and $\mathbb{V}(f_*)$ the expected variance, given by the key predictive equations of GPR [6]. Comparing the proposed architecture with the classic residual model lowers the MSLL from -0.4684 to -0.9978, thus a better score for overall fit and confidence level. Table 1 gives an overview of the scores, both MSE and MSLL, for different scenarios where the amount of samples in each surrogate model varies. This clearly indicates that the weighted residual model repeatedly dominates. Furthermore it can be noted that even in the case where 200 samples are present, the MSE of the weighted residual model outperforms the surrogate model that is trained on one out of two models.

4 CONCLUSIONS

The main goal of this work is to address the main bottleneck of black-box models: their unreliable and often poor performance in scarce sampled regions of the spacial representation of the data. One of the possible advantages of the proposed grey-box methodology is the expected improvement of the grey-box in sparsely sampled datasets by means of enriching a black-box, experimentally based surrogate model with validated data with respect to physics generated by a numerical model, the white-box counterpart.

This paper introduces an ensemble machine learning method based on the residual modelling approach to achieve a trustworthy estimation, effectively addressing challenges posed by scarce data availability. The method leverages the available experimental data by utilising a distance metric to assign a weight to both the white and black-box part. Furthermore, a confidence index is added, based on the hypothesis that a well calibrated physical model might drift over a parameter domain due to epistemic uncertainty or assumptions in the physical model. To adress this, the authors would argue that a case specific value has to be determined based on expert knowledge or a defined measure of uncertainty on the model. We use mean-squared-error and standardised-log-loss metrics to evaluate the method, which indicates a superior performance compared of the proposed methodology to either the performance of the white or black-box model, as well to the classic residual modelling approach.

ACKNOWLEDGEMENTS

This work was funded by KU Leuven, grant #C24E/21/026.

REFERENCES

[1] Cross, E. J., and Rogers, T. J. (2021). Physics-derived covariance functions for machine learning in structural dynamics. IFAC-PapersOnLine, 54(7), 168-173.

[2] Faes, M. and Moens, D. (2020) Recent trends in the modeling and quantification of non-probabilistic uncertainty. Archives of Computational Methods in Engineering, 27, 633–671.

[3] Pitchforth, D. J., Rogers, T. J., Tygesen, U. T., and Cross, E. J. (2021). Grey-box models for wave loading prediction. Mechanical Systems and Signal Processing, 159, 107741.

[4] Brunton, S. L., and Kutz, J. N. (2022). Data-driven science and engineering: Machine learning, dynamical systems, and control. Cambridge University Press.

[5] Bogaerts, L., Dejans, A., Faes, M. G., and Moens, D. (2023). A machine learning approach for efficient and robust resistance spot welding monitoring. Welding in the World, 67(8), 1923-1935.

[6] Williams, C. K. I., and Rasmussen C. E. (2006) Gaussian processes for machine learning. Vol. 2. No. 3. Cambridge, MA: MIT press.

[7] Saxe, A. M., Bansal, Y., Dapello, J., Advani, M., Kolchinsky, A., Tracey, B. D., and Cox, D. D. (2019). On the information bottleneck theory of deep learning. Journal of Statistical Mechanics: Theory and Experiment, 124020.

[8] Faes, M., Imholz, M., Vandepitte, D., and Moens, D. (2021). A review of interval field approaches for uncertainty quantification in numerical models. Modern Trends in Structural and Solid Mechanics 3: Non-deterministic Mechanics, 95-110.

[9] Sohlberg, B., and Sernfält M. (2002). Grey box modelling for river control. Journal of Hydroinformatics 4.4: 265-280.

[10] Bonnet-Eymard, D., Persoons, A., Faes, M. G., and Moens, D. (2023) quantifying uncertainty of physics-informed neural networks for continuum mechanics applications.

[11] Hsieh, J. T., Zhao, S., Eismann, S., Mirabella, L., and Ermon, S. (2019). Learning neural PDE solvers with convergence guarantees. arXiv preprint arXiv:1906.01200.

[12] Kovachki, N., Li, Z., Liu, B., Azizzadenesheli, K., Bhattacharya, K., Stuart, A., and Anandkumar, A. (2023). Neural operator: Learning maps between function spaces with applications to pdes. Journal of Machine Learning Research, 24(89), 1-97.

[13] Nelles, O. Nonlinear System Identification [electronic resource]: From Classical Approaches to Neural Networks and Fuzzy Models.

[14] Kennedy, M. C., Anderson, C. W., Conti, S., and O'Hagan, A. (2006). Case studies in Gaussian process modelling of computer codes. Reliability Engineering & System Safety, 91(10-11), 1301-1309.

[15] Worden, K., and E. J. Cross. (2018). On switching response surface models, with applications to the structural health monitoring of bridges. Mechanical Systems and Signal Processing 98: 139-156.

[16] Faes, M. G., Broggi, M., Spanos, P. D., and Beer, M. (2022). Elucidating appealing features of differentiable auto-correlation functions: A study on the modified exponential kernel. Probabilistic Engineering Mechanics, 69, 103269.

[17] Wang, Y., and Chaib-Draa. B. (2013). A KNN based Kalman filter Gaussian process regression. Twenty-Third International Joint Conference on Artificial Intelligence.

[18] Parussini, L., Venturi, D., Perdikaris, P., and Karniadakis, G. E. (2017). Multi-fidelity Gaussian process regression for prediction of random fields. Journal of Computational Physics 336: 36-50.

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

[20] Dang, C., Valdebenito, M. A., Faes, M. G., Wei, P., and Beer, M. (2022). Structural reliability analysis: A Bayesian perspective. Structural Safety, 99, 102259.

[21] Becker, A. (2022). Kalman Filter from the Ground Up. Independently Published. ISBN: 9798811172753.