Adaptivity and uncertainty of multi-fidelity surrogate models for shape optimization

Jeroen Wackers, Hayriye Pehlivan Solak, Riccardo Pellegrini, Andrea Serani, Matteo Diez

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

ADAPTIVITY AND UNCERTAINTY OF MULTI-FIDELITY SURROGATE MODELS FOR SHAPE OPTIMIZATION

JEROEN WACKERS[†,*], HAYRIYE PEHLIVAN SOLAK[†], RICCARDO PELLEGRINI[‡], ANDREA SERANI[‡], AND MATTEO DIEZ[‡]

[†] LHEEA Lab, Ecole Centrale de Nantes / CNRS UMR 6598, F–44000 Nantes, France
e-mail: jeroen.wackers@ec-nantes.fr

[‡] CNR-INM, National Research Council-Institute of Marine Engineering
Via di Vallerano 139, 00128, Rome, Italy

Key words: Surrogate Modeling, Uncertainty Estimation, Noise Filtering, Small Data

1 INTRODUCTION

Engineering design optimization based on expensive simulation is increasingly performed with surrogate models [1], i.e. approximate models fitted through a small dataset of simulation results. To build surrogates within the lowest possible computational budget, modern approaches use multi-fidelity data (combinations of cheap low-fidelity and expensive high-fidelity simulation results) and adaptive sampling strategies, which add simulation points one by one where they are most likely to discover the optimum [2].

Uncertainty estimation of the surrogate model is crucial for efficient adaptive sampling, since it guides the choice of new sampling points. Thus, underestimation of the uncertainty may lead to sampling in suboptimal regions, missing the true optimum. Existing surrogate models such as Gaussian process regression [3] and Stochastic Radial Basis Functions (SRBF) [4], provide uncertainty estimations, which are often used. Nevertheless, uncertainty estimation is so important for a successful surrogate model, that a more thorough investigation seems warranted. This is the main objective of this paper.

This paper studies three issues with uncertainty estimation, in the context of multi-fidelity SRBF. First, most existing techniques rely on knowledge about the global behavior of the data, such as spatial correlations. However, the number of training points can be too small to reconstruct this global information from the data. We argue that in this situation, user-provided estimation of the function behavior is a better choice (section 3).

Furthermore, the dataset may contain noise, i.e. random errors without spatial correlation (section 4). Surrogate models can filter out this noise, usually by modeling it as belonging to a normal distribution with zero mean, but this introduces two separate uncertainties: the optimum amount of noise filtering is unknown, and for a small dataset the local mean of the noisy data may not correspond to the true simulation response.

Finally, in multi-fidelity models, the low-fidelity data are corrected by high-fidelity results, which could reduce the amount of uncertainty they introduce. Therefore, we argue

that the contribution of the low-fidelity uncertainty to the total multi-fidelity uncertainty is a conditional probability, considering the low-fidelity surrogate in the higher-fidelity training points as given. An approximation based on this principle is developed in section 5. This paper represents a multi-fidelity extension of [5], which introduces the single-fidelity estimations.

2 STOCHASTIC RBF MULTI-FIDELITY SURROGATE MODEL

Consider $\mathbf{x} \in \mathbb{R}^D$ as a design variable vector of dimension D. Let a true objective function $f(\mathbf{x})$ be assessed by N fidelity levels: the highest-fidelity level is $f_1(\mathbf{x})$, the lowest-fidelity is $f_N(\mathbf{x})$, and the intermediate fidelity levels are $\{f_l\}_{l=2}^{N-1}(\mathbf{x})$. Observations can be perturbed by (fidelity-dependent) Gaussian white noise: $s_l(\mathbf{x}) = f_l(\mathbf{x}) + \mathcal{N}(0, \sigma_{n_l})$. The multi-fidelity (MF) prediction $\widehat{f}_l(\mathbf{x})$ of $f_l(\mathbf{x})$ ($l = 1, \ldots, N-1$) is the sum of surrogates (denoted by $\widetilde{\cdot}$) for the lowest level and the inter-level errors [2]:

$$\widehat{f}_l(\mathbf{x}) = \widetilde{f}_N(\mathbf{x}) + \sum_{k=l}^{N-1} \widetilde{\varepsilon}_k(\mathbf{x}). \tag{1}$$

For each l-th fidelity level the training set is $\mathcal{T}_l = \{\mathbf{x}_i, s_l(\mathbf{x}_i)\}_{i=1}^{J_l}$, with J_l the training set size. Nested data are considered: $\{\mathbf{x}_i\}_{i=1}^{J_l} \subseteq \{\mathbf{x}_i\}_{i=1}^{J_{l+1}}$. The resulting inter-level error training sets are defined as $\mathcal{E}_l = \{\mathbf{x}_i, \varepsilon_l(\mathbf{x}_i)\}_{i=1}^{J_l}$, where

$$\varepsilon_l(\mathbf{x}_i) = s_l(\mathbf{x}_i) - \widehat{f}_{l+1}(\mathbf{x}_i). \tag{2}$$

Each prediction $\widetilde{\cdot}_l(\mathbf{x})$ is computed as the expected value (EV) over a stochastic tuning parameter $\tau \sim \text{unif}[1, 3]$ of a surrogate model $\widetilde{g}_l$ [4]:

$$\widetilde{f}_l(\mathbf{x}) = \text{EV}\left[\widetilde{g}_l(\mathbf{x}, \tau)\right]_\tau, \quad \text{with} \quad \widetilde{g}_l(\mathbf{x}, \tau) = \text{EV}\left[\mathbf{s}_l\right] + \sum_{j=1}^{M_l} w_j^l ||\mathbf{x} - \mathbf{c}_j^l||^\tau, \tag{3}$$

where w_j^l are unknown coefficients, $||\cdot||$ is the Euclidean norm and $\mathbf{c}_j^l$ are the RBF centers, with $j = 1, \ldots, M_l$ and $M_l \leq J_l$. The uncertainty $U^{\text{srbf}}(\mathbf{x})$ associated with the SRBF approach is quantified as the 95% confidence interval of the predictions $g_l(\mathbf{x}, \tau)$.

For each fidelity level, if the data are not affected by numerical noise ($\sigma_{n_l} = 0$), exact interpolation of the training set is imposed and the weights w_j^l are computed by solving $\mathbf{A}\mathbf{w}^l = (\mathbf{s}_l - \text{EV}\left[\mathbf{s}_l\right])$, with $\mathbf{c}_j^l = \mathbf{x}_j$ (i.e. $M_l = J_l$) and $\mathbf{s}_l = \{s_l(\mathbf{x}_i)\}_{i=1}^{J_l}$. In the presence of noise, [2] choose a number of RBF centers M_l smaller than the number of training points J_l, and $\mathbf{c}_j^l$ coordinates are defined via k-means clustering of the training point coordinates. The number of RBF kernels M_l ranges from $2D$ to $J-1$; least-squares regression provides more smoothing of the noisy training set when the number of centers M_l is small. The optimal number of stochastic RBF centers is defined by minimizing a leave-one-out cross-validation (LOOCV) metric. Finally, the weights w_j^l are determined with least-squares regression (LS) by solving $\mathbf{A}^\mathsf{T}\mathbf{A}\mathbf{w}^l = \mathbf{A}^\mathsf{T}(\mathbf{s}_l - \text{EV}\left[\mathbf{s}_l\right])$.

Sections 3 and 4 develop an uncertainty estimation for a single-level surrogate; the indices l are dropped there for clarity. Section 5 considers the multi-fidelity uncertainty.

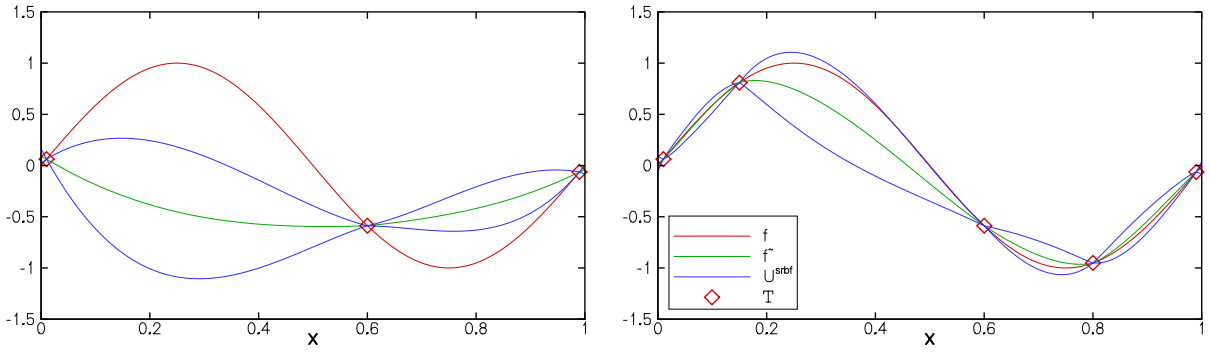

Figure 1: SRBF uncertainty estimation for a sine-wave function. 2 (left) and 3 (right) points per peak.

3 INTERPOLATION UNCERTAINTY IN SMALL-DATA CASES

The SRBF uncertainty estimation U^{srbf} is highly accurate in predicting the model error for surrogate models without noise, if enough training points are available to represent f correctly. The 95% confidence interval of $g(\mathbf{x}, \tau)$ is close to $g(\mathbf{x}, 3) - g(\mathbf{x}, 1)$, where $g(\mathbf{x}, 1)$ is $\mathcal{C}^0$ and piecewise linear, while $g(\mathbf{x}, 3)$ is piecewise cubic and $\mathcal{C}^2$. Our tests show that this difference is a good estimator for the missing above-cubic terms, as long as the second derivative of f is approximated correctly by the surrogate model (figure 1 right). When insufficient points are available to capture the second derivatives, the uncertainty estimation fails (figure 1 left).

We refer to a "small-data" situation when (a) the true function behavior cannot be estimated from the data, and (b) the data cannot indicate that the approximation of the true function is incorrect. In this case, the only way to evaluate the uncertainty is with user-provided estimations of the behavior of f as a supplement to the data. While reliance on user knowledge is a weakness for automatic surrogate model construction, we consider it as inevitable. This section presents a small-data uncertainty estimation.

Default uncertainty If not enough data are available to correctly model f, the uncertainty estimation must be based on assumptions about the function, rather than information from the training points. Since the minimum of f is being sought, the function is assumed to consist of peaks and valleys with a parabolic behavior, a typical peak width $2r_0$ and a typical peak height U_0. Both these parameters need to be estimated from another source than the training points. In the following, r_0 is estimated as $\frac{1}{4}$ times the domain size, while U_0 is taken as $\max_i s(\mathbf{x}_i) - \min_i s(\mathbf{x}_i)$. These are reasonable default choices for any function, but if more reliable estimates are available for a function, the uncertainty estimation will be better.

The parabolic behavior between data points is used to define a default uncertainty U^{def}, based only on the distance to the closest training point $r_{i_{min}}(\mathbf{x}) = \min_{i=1\ldots J} \|\mathbf{x} - \mathbf{x}_i\|$:

$$U^{\text{def}}(\mathbf{x}) = \begin{cases} U_0(1 - (r_{i_{min}}(\mathbf{x})/r_0 - 1)^2) & r_{i_{min}}(\mathbf{x})/r_0 \leq 1, \\ U_0 & r_{i_{min}}(\mathbf{x})/r_0 > 1. \end{cases} \tag{4}$$

Blending the uncertainties As noted above, the SRBF uncertainty is reliable when the second derivatives of f are well represented by the data. In each peak or valley, ignoring cross-derivatives, a central point plus 2 points for each dimension are needed to capture the second derivatives. Thus, the original SRBF uncertainty can be considered as valid in points $\mathbf{x}$ where at least $2D+1$ training points are found in a sphere of radius r_0 around $\mathbf{x}$. If fewer training points are close to $\mathbf{x}$, the default uncertainty should be used.

A smooth transition between the two uncertainty estimations is obtained with two smeared Heaviside functions. To prevent a sharp distinction between points on the inside of the region and just outside, the number of training points in the sphere r_0 is counted as:

$$\bar{d}(\mathbf{x}) = \sum_{i=1}^{n} \left(1 - \frac{1}{1 + e^{-12.5\left(\frac{r_i(\mathbf{x})}{r_0} - 1.1\right)}} \right). \tag{5}$$

A weight w for the default uncertainty is evaluated based on $\bar{d}$:

$$w(\bar{d}) = 1 - \frac{1}{1 + e^{-25\left(\frac{\bar{d}}{2D+1} - 0.8\right)}}. \tag{6}$$

Using w, the modified interpolation uncertainty estimation is defined as:

$$U^{\text{interp}}(\mathbf{x}) = U^{\text{srbf}}(\mathbf{x}) \left(1 - w\left(\bar{d}(\mathbf{x})\right)\right) + U^{\text{def}}(\mathbf{x})\, w(\bar{d}(\mathbf{x})). \tag{7}$$

The adjustment functions used in the modified uncertainty are shown in figure 2.

4 NOISE CANCELING AND TRAINING-POINT UNCERTAINTY

Least-squares regression of the training set (section 2) is effective for filtering noise, but has three disadvantages: (i) the interpolation error estimator is ill founded for non-interpolating surrogate models, (ii) the uncertainty associated with the surrogate model prediction may be smaller between the training points than in the points themselves, which is illogical, and (iii) if the LOOCV procedure chooses a high number of kernels, overfitting can occur [6].

Therefore, we define a new noise canceling approach which consists of separate noise-filtering and interpolation steps. To compute the surrogate, noise-filtered data $\bar{f}_i(\mathbf{x}_i)$ are first reconstructed in the training points. Analogous to SRBF, these data are a weighted average of fits with different noise levels; this reduces the risk of overfitting caused by one

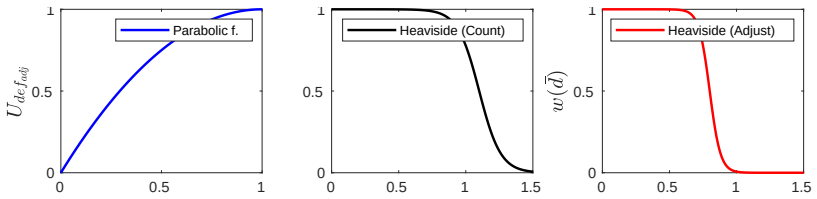

Figure 2: Adjustment functions: default uncertainty U^{def}, counting function $\bar{d}(r)$, and weight $w(\bar{d})$.

extreme fit. Then, standard SRBF interpolation of $\bar{f}_i(\mathbf{x}_i)$ is performed, which eliminates the problem (i) of the interpolation error estimation.

To associate an uncertainty to the noise-filtered surrogate model, it is assumed that the training data uncertainty has two components: (1) the unknown amount of noise in the data introduces an uncertainty in the required amount of noise filtering, which can be estimated from the variance of the different fits, and (2) for a small dataset, even with perfect noise filtering, the actual mean of the function in the neighborhood of a training point may not correspond to the true simulation response; this mean-value uncertainty can be estimated with the central limit theorem.

Training set reconstruction and associated uncertainty The noise in the training points is removed with least-squares fitting, varying the number of RBF kernels M to represent different noise levels. Thus, the noise-filtering uncertainty in the training points is evaluated as the variance associated to the different fits of $\widetilde{g}^M(\mathbf{x}_i, \tau)$ due to varying M with a fixed value of $\tau = 3$. The noise-filtered training data $\bar{f}_i$ are based on the average of $\widetilde{g}^M(\mathbf{x}_i, 3)$.

Depending on the noise that is actually present, not all M are equally likely. Therefore, LOOCV is used to identify the most likely $\widetilde{g}^M(\mathbf{x}_i, 3)$ fits. Let $\widetilde{g}^M_{-i}(\mathbf{x})$ be a surrogate model trained with all the noisy data $\mathcal{T} = \{\mathbf{x}_i, s(\mathbf{x}_i)\}_{i=1}^J$ except the i-th point, using M centers. Then the LOOCV error in the point i is:

$$e_i^M = \left| s(\mathbf{x}_i) - \widetilde{g}^M_{-i}(\mathbf{x}_i) \right|. \tag{8}$$

The errors e_i^M are used to achieve a likelihood estimation for each M based on an approximation of the noise level σ_n:

$$\sigma_n^2 \approx \sigma_{CV}^2 = \min_M \frac{1}{J} \sum_{i=1}^{J} \left(e_i^M\right)^2, \tag{9}$$

where the minimum of the LOOCV errors is taken as the closest approximation of the noise, since the reconstruction with the lowest error is probably closest to the true function. The likelihood that each M provides an accurate regression of the noisy data, $\mathcal{L}(M)$, is the probability of the training point data $s(\mathbf{x}_i)$ given the surrogate model $\widetilde{g}^M_{-i}$ and white noise $\sim \mathcal{N}(0, \sigma_{CV})$ [3]:

$$\mathcal{L}(M) = \prod_{i=1}^{J} p\left(s(\mathbf{x}_i) \,\middle|\, \widetilde{g}^M_{-i}(\mathbf{x}_i)\right) = \prod_{i=1}^{J} \frac{1}{\sigma_{CV}\sqrt{2\pi}} \exp\left(-\tfrac{1}{2}\left(\frac{e_i^M}{\sigma_{CV}}\right)^2\right). \tag{10}$$

The likelihoods $\mathcal{L}(M)$ for $M = 2D \ldots J-1$ are finally scaled to form a partition of unity. $\mathcal{L}$ serves to reconstruct the data $\bar{f}_i$ with a weighted average of the non-LOOCV fits $\widetilde{g}^M$, while the variance provides an uncertainty for the noise filtering:

$$\bar{f}_i = \sum_M \mathcal{L}(M)\widetilde{g}^M(\mathbf{x}_i, 3), \qquad \left(\sigma_i^{\text{filt}}\right)^2 = \sum_M \mathcal{L}(M)\left(\widetilde{g}^M(\mathbf{x}_i, 3) - \bar{f}_i\right)^2. \tag{11}$$

Finally, since LOOCV uses interpolation towards the excluded point, it does not provide reliable results if too few training points are available (section 3). Therefore, the training points which have fewer than $2D + 1$ neighbors within a distance r_0 are ignored for the LOOCV and their training data $s(\mathbf{x}_i)$ are kept unmodified.

Mean-value uncertainty According to the central limit theorem, the mean of n realizations of a function with stochastic noise is another stochastic variable whose standard deviation is the noise standard deviation divided by $\sqrt{n}$. Therefore, estimating the uncertainty in the mean value of $s(\mathbf{x}_i)$ requires an estimation of the noise level and an indication of how many training points contribute to the local mean value; both vary with M. Following a conservative approach, it is preferable to overestimate the noise, therefore a different estimate than (9) is selected, i.e. the highest noise level for which the outcome $s(\mathbf{x}_i)$ is in the 95% confidence interval:

$$\sigma_{n,M}^2 = \frac{1}{F^{-1}(0.025, J)} \sum_{i=1}^{J} \left(s(\mathbf{x}_i) - \widetilde{g}^M(\mathbf{x}_i)\right)^2, \tag{12}$$

where $F^{-1}(0.025, J)$ is the inverse of the cumulated distribution function for the chi-squared distribution with J samples, evaluated at a probability of 2.5%.

The (probably pessimistic) estimated number of training points which contribute to each local minimum is the number of training points I_i^M which are k-means clustered into the same RBF center as point i, when M centers are used. The mean-value uncertainty for M centers is then:

$$\sigma_{m,M}(\mathbf{x}_i)^2 = \frac{\sigma_{n,M}^2}{I_i^M}, \tag{13}$$

The final estimated mean-value variance is weighted like the noise-filtering variance:

$$(\sigma_i^{\text{mean}})^2 = \sum_M \mathcal{L}(M)(\sigma_{m,M}(\mathbf{x}_i))^2. \tag{14}$$

Total training point uncertainty Assuming that the mean-value and noise-canceling uncertainties are independent, the standard deviation of the total training point uncertainty in the i-th point is:

$$\left(\sigma_i^{\text{data}}\right)^2 = \left(\sigma_i^{\text{filt}}\right)^2 + (\sigma_i^{\text{mean}})^2. \tag{15}$$

The 95% confidence interval U_i^{data} is assumed equal to $2\sigma_i^{\text{data}}$, it is interpolated with SRBF ($M = J$) and added to U^{interp}, as shown in the following section.

5 MULTI-FIDELITY

Thanks to the definition (2), the error training data cancel the influence of LF data in HF training points, so each multi-fidelity model $\widehat{f_l}$ is a fit through its own data $\mathcal{T}_l$. Hence, the full multi-fidelity surrogate model can be seen as a (highly sophisticated) interpolation

through the highest-fidelity data. Following the principles of the preceding sections, the MF uncertainty therefore consists of HF data uncertainty and an interpolation uncertainty caused by the use of lower-fidelity models between the high-fidelity data.

Recursively writing (1) as $\widehat{f}_l = \widetilde{\varepsilon}_l + \widehat{f}_{l+1}$ allows its uncertainty to be decomposed as

$$U^2_{\widehat{f}_l} = (U^{\text{data}}_{\widetilde{\epsilon}_l})^2 + (U^{\text{interp}}_{\widetilde{\epsilon}_l})^2 + \left(U^2_{\widehat{f}_{l+1}}\right)^{\text{interp}}. \tag{16}$$

When using the low-fidelity surrogate $\widehat{f}_{l+1}$ for interpolation, the error induced in the HF training points $\{\mathbf{x}_i\}_{i=1}^{J_l}$ is zero, since any deviation is corrected by the error surrogate $\widetilde{\varepsilon}_l$. The MF interpolation uncertainty induced by $\widehat{f}_{l+1}(\mathbf{x})$ can therefore be seen as conditional w.r.t its value in those points:

$$\left(U^2_{\widehat{f}_{l+1}}\right)^{\text{interp}} = \left(U^2_{\widehat{f}_{l+1}} \,\middle|\, \widehat{\mathbf{f}}^l_{l+1}\right), \tag{17}$$

where $\widehat{\mathbf{f}}^l_{l+1} = \{\widehat{f}_{l+1}(\mathbf{x}_i)\}_{i=1}^{J_l}$. The recursion in (16) can be removed by substitution in (17) and simplifying. The interpolation error is zero in data points, so $(U^{\text{interp}}_{\widetilde{\epsilon}_{l+1}} \,|\, \widehat{\mathbf{f}}^l_{l+1}) = U^{\text{interp}}_{\widetilde{\epsilon}_{l+1}}$. Furthermore, the data are nested, so $\left((U_{\hat{f}_{l+2}} \,|\, \widehat{\mathbf{f}}^{l+1}_{l+2}) \,\middle|\, \widehat{\mathbf{f}}^l_{l+2}\right) = (U_{\hat{f}_{l+2}} \,|\, \widehat{\mathbf{f}}^{l+1}_{l+2})$ since $\{\mathbf{x}_i\}_{i=1}^{J_l} \subseteq \{\mathbf{x}_i\}_{i=1}^{J_{l+1}}$. Thus, equation (16) becomes:

$$U^2_{\widehat{f}_l} = \left[(U^{\text{interp}}_{\widetilde{f}_N})^2 + \sum_{k=N-1}^{l} (U^{\text{interp}}_{\widetilde{\epsilon}_k})^2\right] + \left[\left(U^{\text{data}}_{\widetilde{f}_N} \,\middle|\, \widehat{\mathbf{f}}^{N-1}_N\right)^2 + \sum_{k=N-1}^{l} \left(U^{\text{data}}_{\widetilde{\epsilon}_{k+1}} \,\middle|\, \widehat{\mathbf{f}}^k_{k+1}\right)^2\right] + (U^{\text{data}}_{\widetilde{\epsilon}_l})^2. \tag{18}$$

Estimation Since no stochastic process is known for the data, the conditional training point uncertainty has to be estimated. We suppose that it is linked to the two-point correlation of the data: if a training point $\mathbf{x}_i$ and an arbitrary point $\mathbf{x}$ are uncorrelated, fixing $f_{l+1}(\mathbf{x}_i)$ has no influence on $\mathbf{x}$. The two-point correlation of the noise-filtered data can be estimated from the local distances between RBF kernels $\mathbf{c}^{l+1}_j$, since these represent data which the noise canceling deems independent. However, the kernel point distance depends on the number of kernels M_{l+1} (section 4). The following estimation has been implemented:

1. In each training point $\mathbf{x}_i$, the characteristic radius $d^{l+1,M}_{\mathbf{x}_i}$ for each M_{l+1} is the minimum distance of the kernel center into which the training point is clustered to any other kernel center. Like (14), the final characteristic radius $d^{l+1}_{\mathbf{x}_i}$ is a weighted mean of these distances.

2. Then, for any point $\mathbf{x}$, the largest two-point correlation with any training point in $l+1$ which is also available on level l, is:

$$K^{l+1}_{\max}(\mathbf{x}) = \max_{i \in J_l} K\left(\mathbf{x}, \mathbf{x}_i, d^{l+1}_{\mathbf{x}_i}\right). \tag{19}$$

The correlation function $K(\mathbf{x}, \mathbf{x}', d)$ has to be estimated. Since the default interpolation uncertainty (section 3) is based on parabolas we use:

$$K\left(r,d\right) = \begin{cases} \left(r/d-1\right)^2 & r/d \leq 1, \\ 0 & r/d > 1, \end{cases} \tag{20}$$

where $r = \|\mathbf{x} - \mathbf{x}'\|$. However, this choice is arbitrary; a Gaussian kernel function could also be used. The conditional uncertainty in (18) then becomes:

$$\left(U^{\text{data}}_{\tilde{\epsilon}_{l+1}} \middle| \widehat{\mathbf{f}}^{\,l}_{l+1}\right) \approx U^{\text{data}}_{\tilde{\epsilon}_{l+1}}(\mathbf{x}) \left(1 - K^{l+1}_{\max}(\mathbf{x})\right). \tag{21}$$

6 TEST CASES

Interpolation The interpolation uncertainty estimation is tested on the 1D MF Forrester function [7]. For the moment, only the HF function f_1 is used:

$$\begin{aligned} f_1(x) &= \left(6x-2\right)^2 \sin(12x-4), \\ f_2(x) &= \left(6x-2\right)^2 \sin(12x-4) - 10(x-0.5) - 5. \end{aligned} \tag{22}$$

With 2 and 3 training points (figure 3) the default uncertainty is used everywhere, which is the right choice: unlike U^{srbf}, the modified uncertainty interval contains most of the true function. For 5 data points, U^{srbf} becomes reliable and U^{def} starts to be switched

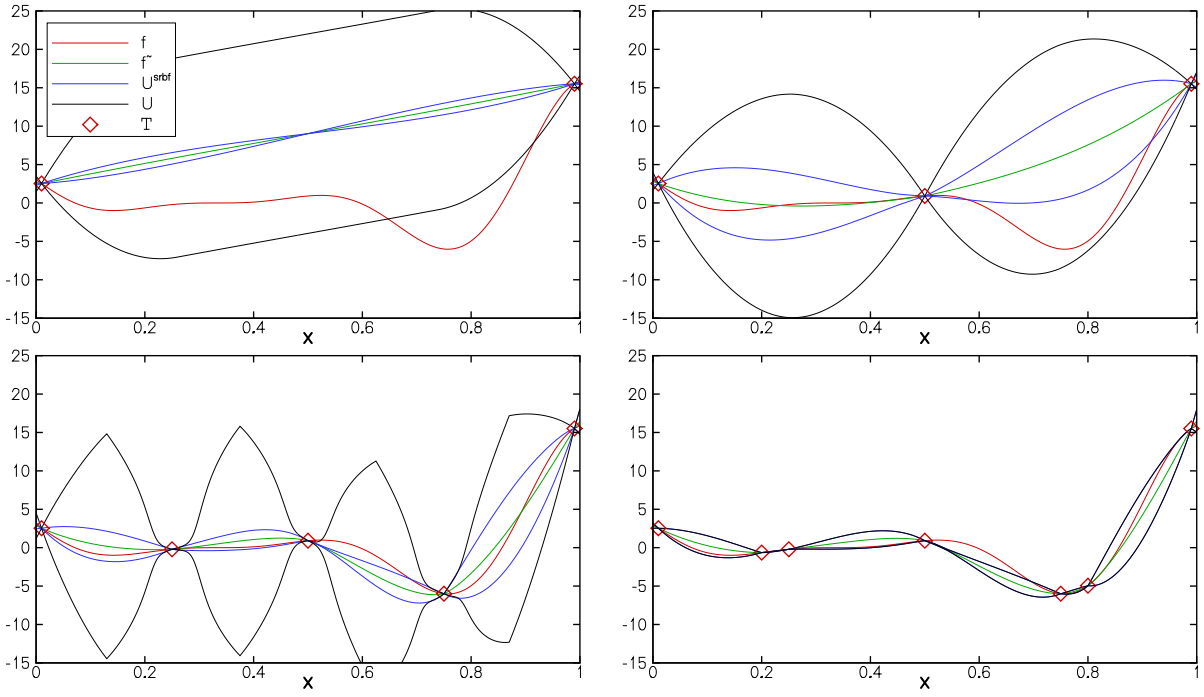

Figure 3: HF Forrester (f_1) without noise: interpolation uncertainty with 2, 3, 5 and 7 training points.

off, while for 7 points, U^{srbf} is selected everywhere. Looking in detail, the estimation is pessimistic for 5 points, while for 7 points the true function leaves the uncertainty domain once. This is because the peak width r_0 which was chosen as a compromise to fit many different functions, does not correspond to the actual peak widths for Forrester. Given this limitation, the new estimation predicts the uncertainty with a reasonable accuracy.

Noise filtering Figure 4 shows three surrogate models for the HF Forrester function with noise. The first one has $\sigma_n = 1.5$ and clustered data. $RMSE$ is the function minimized in (9), err the RMS difference between the true f and each fit. The interpolation uncertainty varies with the distance between sampling points and even U^{def} is used. The training point uncertainty is reduced around the cluster, thanks to the lower mean-value

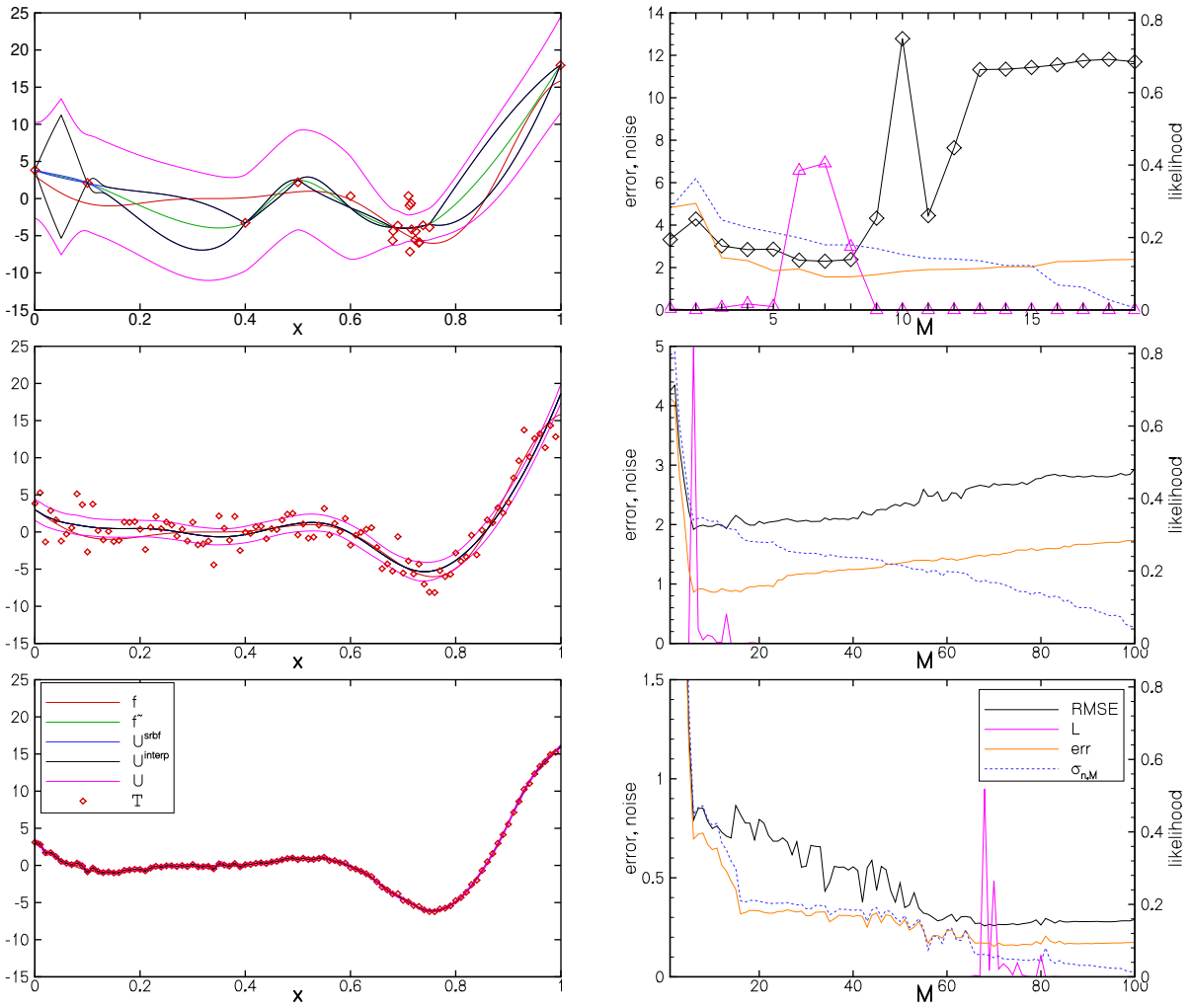

Figure 4: Surrogate models for HF Forrester with noise: 20 points, $\sigma_n = 1.5$ (top), 100 points, $\sigma_n = 1.5$ (middle), and 100 points, $\sigma_n = 0.15$ (bottom). Right figures: the noise and errors of the fits with different M. $RMSE$ is the function minimized in (9), err the RMS difference between the true f and each fit.

uncertainty. The right figure shows that 3 values of M are the most likely. These coincide both with the minimum of the RMSE and with the minimum true error, showing the efficiency of the likelihood estimator. Finally, the noise $\sigma_{n,M}$ is overestimated w.r.t. σ_n as desired, but the order of magnitude is correct. These observations hold for all three tests.

The middle image retains $\sigma_n = 1.5$ but has 100 equidistributed training points. For this point density, U^{interp} is negligible. The total uncertainty interval is smaller than the spread of the data, indicating effective noise filtering. Also, the uncertainty is smaller than in the cluster for the first case, although the local point density is lower. Thus, the data uncertainty in a given point is non-local; it depends on the data in a region around the point. With a noise level $\sigma_n = 0.15$ (bottom row) the LOOCV automatically detects that less smoothed fits (higher M) are more likely and changes the chosen fits.

The last test (figure 5) uses the 145-point low-fidelity dataset from the 2-parameter two-fidelity airfoil optimization of [2], which has a valley-like response shape with a minimum around [0.3, 0] and at least 10% noise. The new approach is compared with the LS-SRBF uncertainty estimation we presented in [2] (see section 2). For the new approach, the neighbor count of equation (5) varies abruptly since the data are highly clustered, which explains the rapid change to the default uncertainty in the top half of the domain. The separation of training point and interpolation uncertainty ensures that the uncertainty minima are in the training points. Also, the clustered data reduce the mean-value uncertainty, which leads to minimum zones around the clusters. The LS-SRBF approach however, predicts the minimum uncertainty in positions next to the data, which likely coincide with the RBF center positions. Altogether, the new uncertainty estimation seems more credible and appears to be a good basis for adaptive sampling.

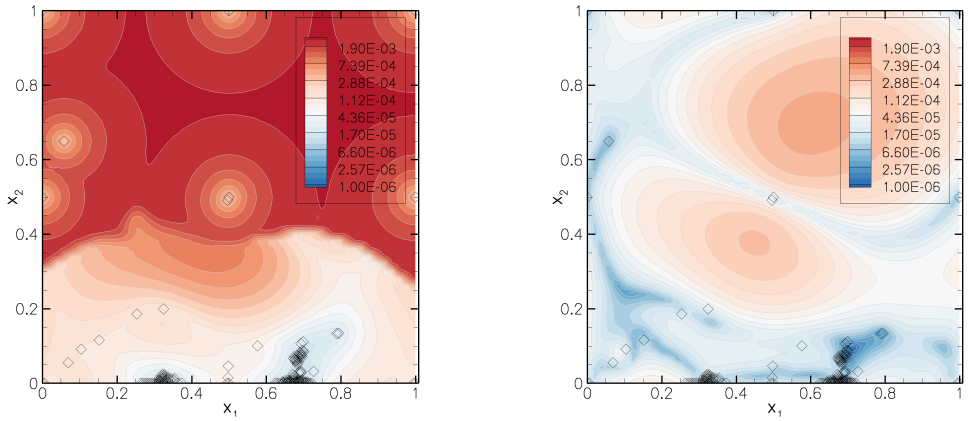

Figure 5: NACA airfoil 2D uncertainty with the new approach (left) and with LS-SRBF from [2] (right).

Multi-fidelity Finally, a multi-fidelity case is performed with 21 equidistributed LF points, 6 HF points, and noise $\sigma_{n_2} = 1.5$ / $\sigma_{n_1} = 0.15$. Figure 6 shows the surrogate model predictions and uncertainty components and the LOOCV error and noise estimation for the low-fidelity and error models. It is worth noting how the new uncertainty U_2 mostly encloses the true function f_2, whereas the U^{interp} is negligible for 21 training points. The

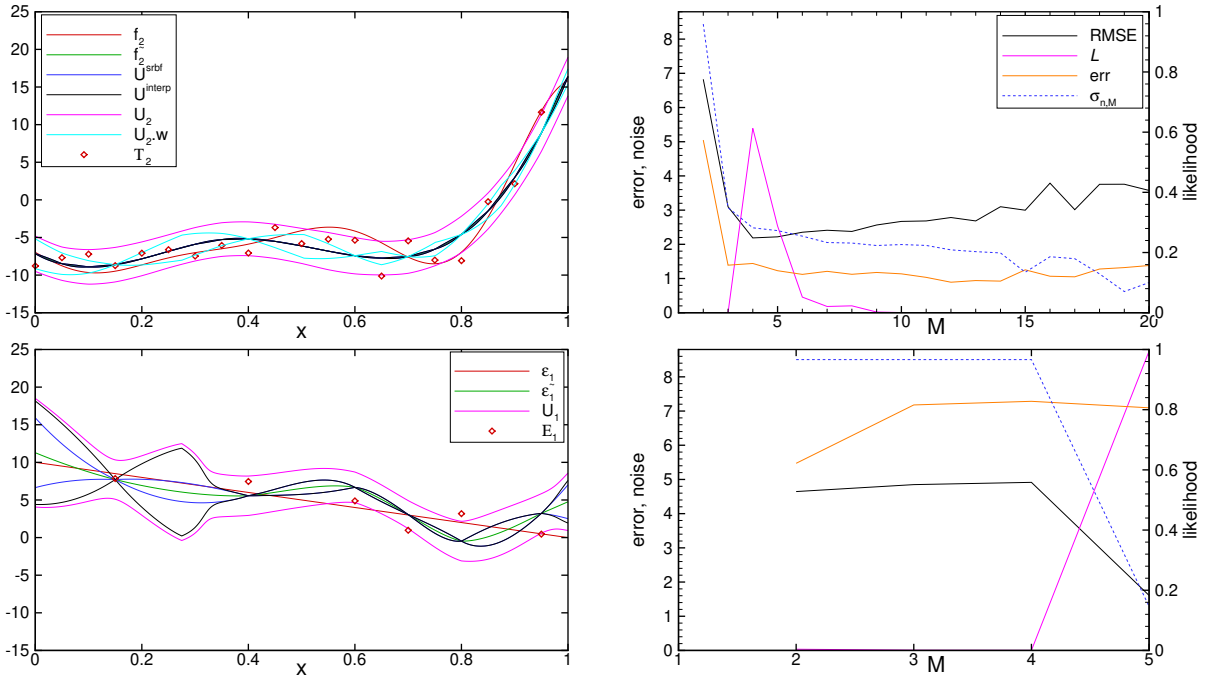

Figure 6: Individual surrogate model predictions and uncertainty components for MF Forrester with noise: LF 21 points, $\sigma_{n_2} = 1.5$ (left), HF 6 points, $\sigma_{n_1} = 0.15$ (right). Right figures: the noise and errors of the fits with different M.

low-fidelity surrogate model is not accurate in modeling the actual function, likely due to the high noise level which leads to low numbers of centers, see the top right figure. As a consequence, the error surrogate model does not agree with the true error function, since it compensates for the inaccuracy of the low-fidelity model. Thus, the final MF surrogate model and the uncertainty estimation are accurate, see Figure 7.

The weighted uncertainty of the low-fidelity surrogate (figure 6 top) is low compared with the error model uncertainty, which therefore makes up most of the MF uncertainty. Since the error model must compensate for the deficiencies of the low-fidelity model, its complexity is higher than the analytical error function, so it requires more training points

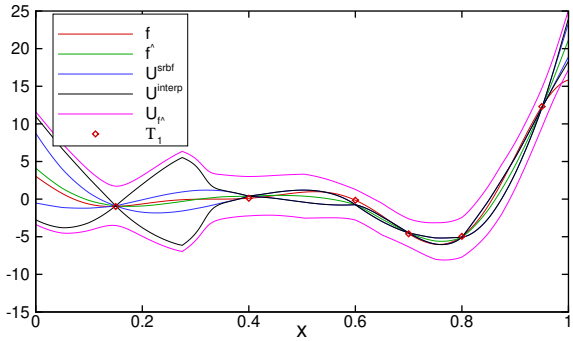

Figure 7: Multi-fidelity prediction and uncertainty components for MF Forrester.

for sufficient accuracy. In this situation, the uncertainty estimation shows that further training points should mainly be added to the high-fidelity training set.

7 CONCLUSIONS

This paper presents several additions to SRBF uncertainty estimation for surrogate models: a default estimation when few training data are available, and a noise canceling procedure which estimates both the uncertainty in the estimated noise level and the deviations due to the uncertain local means of the data. Finally, it is argued that for multi-fidelity, LF uncertainty should be treated as conditional to the HF training data.

A common conclusion from all these additions is, that the training data do not contain enough information to accurately compute both the surrogate model and its uncertainty. Instead, the uncertainty has to be estimated using assumptions (such as default uncertainty values, influence radii, etc.) which have to be provided separately from the data. While the estimation methods provided in this paper are open to discussion, we think this work shows that surrogate uncertainty estimation can, and should, be studied in detail.

ACKNOWLEDGEMENTS

The work at ECN is funded by the Institut Carnot MERS in the ORUP project. CNR-INM is partially supported by the Horizon Europe "RETROFIT55 - Retrofit solutions to achieve 55% GHG reduction by 2030", grant agreement 101096068.

REFERENCES

[1] A. Serani et al., "A scoping review on simulation-based design optimization in marine engineering: trends, best practices, and gaps". *Arch. Comput. Meth. Eng.* 1–29 (2024).

[2] R. Pellegrini et al., "A multi-fidelity active learning method for global design optimization problems with noisy evaluations". *Eng. with Comput.* **39**, 3183–3206 (2023).

[3] C.E. Rasmussen and C.K.I. Williams, *Gaussian processes for machine learning*. The MIT Press, (2006).

[4] S. Volpi et al., "Development and validation of a dynamic metamodel based on stochastic radial basis functions and uncertainty quantification". *Struct. Multidiscipl. Optim.* **51**(2), 347–368, (2015).

[5] J. Wackers et al., "Error estimation for surrogate models with noisy small-sized training sets". In *Adaptive Modelling and Simuation (ADMOS 2023)*, Gothenburg, Sweden.

[6] J. Wackers et al., "Efficient initialization for multi-fidelity surrogate-based optimization". *J. Ocean Eng. Marine Energy* **9**, 291–307 (2023).

[7] L. Mainini et al., "Analytical benchmark problems for multifidelity optimization methods". *arXiv preprint arXiv:2204.07867*, (2022).