Adaptive Multi-fidelity Surrogate Modelling for High-quality Shape Optimization

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

Adaptive Multi-fidelity Surrogate Modelling for High-quality Shape Optimization

Jeroen Wackers
LHEEA Lab, Ecole Centrale de Nantes
Nantes
FRANCE

Hayriye Pehlivan Solak
Arkhe Marine Eng. Ltd. Co.
Istanbul
TURKEY

Riccardo Pellegrini, Andrea Serani, Matteo Diez
National Research Council - Institute of Marine Engineering
Rome
ITALY

ABSTRACT

Surrogate modeling and active learning methods for simulation-driven design optimization of innovative vehicles are negatively influenced by numerical noise, which is often unavoidable for numerical solvers. We propose a new approach for uncertainty estimation with noisy data, which models the uncertainty as separate contributions from the noise-affected training points, their interpolation, and the multi-fidelity corrections. The surrogate model and the associated interpolation uncertainty are reconstructed with Stochastic Radial Basis Functions (SRBF), which use a range of RBF fits with different kernels. The noise in the training points is filtered out by reconstructing filtered data in the training points, which are then interpolated with standard SRBF. For the filtering, several RBF surrogates with a number of kernels smaller than the number of training points are least-squares fitted through the data; the uncertainty contribution from the noise filtering is estimated as the variance of these reconstructions. Tests on analytical functions and airfoil shape optimization show that active learning based on this estimator is more efficient and more robust than existing approaches.

25.1 NOMENCLATURE

Operators:

$\widetilde{\cdot}$	Surrogate model
$\widehat{\cdot}$	Multi-fidelity surrogate
$\bar{\cdot}$	Filtering

Greek symbols:

α	Weight for baseline uncertainty
ε	Inter-level discrepancy
σ_{CV}	Estimation of σ_n based on LOOCV
$\sigma_{m,M}$	Mean-value uncertainty for M RBF centers
σ^{μ}	Mean-value uncertainty
σ_n	Training data noise
$\sigma^{\bar{s}}$	Noise-filtering uncertainty
τ	Stochastic power parameter
ψ	Acquisition function

Roman symbols:

$\mathbf{c}$	RBF kernel center
D	Design space dimension
$d_{\min}$	Distance to closest training point
$d_{\mathbf{x}_i}$	Characteristic radius for 2-point correlation
E_x	Design space error metric
E_p	Prediction error metric
e^M	Cross-validation error with M kernels

F^{-1}	Inverse cumulated chi-squared distribution
f	Objective function
$\widetilde{g}$	Individual RBF surrogate model
J	Training set size
J_{CV}	Number of points suitable for LOOCV
K	Two-point correlation function
$\mathcal{L}(M)$	Likelihood of M kernels
M	Number of RBF kernels
N	Number of fidelity levels
$\mathcal{N}$	Normal probability distribution
$\overline{n}$	Smoothed number of neighbor training points
P_x	Penalization in acquisition function centers
r	Distance between two arbitrary points
r_0	Estimated peak half width
r_i	Distance to training point i
s	(Noisy) training data
$\mathcal{T}, \mathcal{E}$	Training set
U^0	Baseline uncertainty
$U^{\widetilde{g}}$	SRBF interpolation uncertainty
U^I	Interpolation uncertainty
U^s	Training point uncertainty
w	RBF kernel weight
w_{cv}	Weight of training point for LOOCV
$\mathbf{x}$	Design variables

Indices and exponents:

i	Sample number
j	RBF kernel number
l	Fidelity level
$*$	Point to be sampled

Abbreviations:

CFD	Computational fluid dynamics
CC	Computational cost
EV	Expected value
HF	High fidelity
LCB	Lower confidence bound
LF	Low fidelity
LOOCV	Leave-one-out cross-validation
LS-SRBF	Least-squares SRBF
MF	Multi-fidelity
NACA	National Advisory Committee for Aeronautics
RBF	Radial basis functions
RMS	Root mean squared
RS	Reduced startset [1]
SRBF	Stochastic RBF

25.2 INTRODUCTION

Simulation-driven design optimization of complex engineering systems (such as naval or air vehicles) requires high-fidelity physics-based solvers to achieve sound and accurate performance assessment. The computational cost of such an optimization can become a technological challenge, especially for global optimum searches or high-dimensional design spaces. For this reason, the optimization step is often performed with surrogate models [2], i.e. approximate models fitted through a small dataset of simulation results. Furthermore, 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 active learning strategies, which add simulation points where they are most informative (i.e. in regions of the domain where it is most likely to discover the optimum) [3].

Uncertainty estimation of the surrogate model is crucial for efficient active learning, since it guides the choice of new sampling points. Existing surrogate models such as Gaussian process regression [4] and Stochastic Radial Basis Functions (SRBF) [5], provide uncertainty estimations associated to the prediction. Nevertheless, uncertainty estimation is such a fundamental task for successful adaptive surrogate modeling that a more thorough investigation seems warranted.

The active learning process is generally affected by two issues. First, computational-output noise [6] is often unavoidable when large systems of nonlinear partial differential equations are numerically solved. Active

learning methods may react to noise by adding many training points in noisy regions, rather than selecting new points in unseen regions [7]. And second, in the early stages of the process, the number of training points may be insufficient to show the true nature of the simulation response, which may mislead the surrogate creation.

This paper investigates efficient and effective active learning in the presence of numerical noise, based on a new model of the uncertainty associated to the multi-fidelity surrogate prediction. The uncertainty is modeled as the sum of three main components: the noise filtering in the training points, the surrogate model interpolation, and the propagation of these effects in a multi-fidelity case. Specifically:

- The training set noise, i.e. random errors without spatial correlation, can be filtered out by surrogate models, usually by modeling it as belonging to a normal distribution with zero mean. This introduces two separate uncertainties, since 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.

- Most of the 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.

- Finally, in multi-fidelity models, the low-fidelity training data are corrected by high-fidelity results, which could reduce the amount of uncertainty introduced by the low-fidelity data.

This paper extends the capabilities of a surrogate model using such uncertainty modeling [8], [9], by introducing and testing an active learning procedure.

25.3 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 $s_l(\mathbf{x})$ 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 [3]:

$$\widehat{f}_l(\mathbf{x}) = \widetilde{f}_N(\mathbf{x}) + \sum_{k=l}^{N-1} \widetilde{\varepsilon}_k(\mathbf{x}). \tag{25-1}$$

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 inter-level error training sets are defined as $\mathcal{E}_l = \{\mathbf{x}_i, \varepsilon_l(\mathbf{x}_i)\}_{i=1}^{J_l}$, where

$$\varepsilon_l(\mathbf{x}_i) = s_l(\mathbf{x}_i) - \widehat{f}_{l+1}(\mathbf{x}_i). \tag{25-2}$$

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 $\widetilde{g}_l$ [5]:

$$\widetilde{f}_l(\mathbf{x}) = \text{EV}\left[\widetilde{g}_l(\mathbf{x}, \tau)\right]_\tau, \quad \text{with} \quad \widetilde{g}_l(\mathbf{x}, \tau) = \text{EV}\left[\mathbf{s}_l\right] + \sum_{j=1}^{M_l} w_j^l ||\mathbf{x} - \mathbf{c}_j^l||^\tau, \tag{25-3}$$

where w_j^l are unknown coefficients, $||\cdot||$ is the Euclidean norm, $\mathbf{s}_l = \{s_l(\mathbf{x}_i)\}_{i=1}^{J_l}$ and $\mathbf{c}_j^l$ are the RBF centers, with $j = 1, \ldots, M_l$ and $M_l \leq J_l$.

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 $\mathbf{A}\mathbf{w}^l = (\mathbf{s}_l - \mathrm{EV}[\mathbf{s}_l])$, with $\mathbf{c}_j^l = \mathbf{x}_j$ (i.e. $M_l = J_l$). In the presence of noise, [3] 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 $\mathbf{A}^\mathsf{T}\mathbf{A}\mathbf{w}^l = \mathbf{A}^\mathsf{T}(\mathbf{s}_l - \mathrm{EV}[\mathbf{s}_l])$.

25.4 UNCERTAINTY MODELLING

The uncertainty $U^{\tilde{g}}(\mathbf{x})$ associated with the SRBF approach is quantified as the 95% confidence interval of the predictions $g_l(\mathbf{x}, \tau)$ in (25-3). This estimation is highly accurate in predicting the error for surrogate models without noise, if enough training points are available to represent f correctly. Sections 25.4.1 extends this uncertainty estimation for small training sets, and section 25.4.2 for noise. Since these sections concern a single-level surrogate, the indices l are dropped for clarity. Section 25.4.3 considers the multi-fidelity uncertainty.

25.4.1 Interpolation Uncertainty in Small-Data Cases

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 (i.e. the true function is effectively enclosed by the uncertainty), as long as the second derivative of f is approximated correctly by the surrogate model (figure 25-1 right). When insufficient points are available to capture the second derivatives, the uncertainty estimation fails (figure 25-1 left).

We thus 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.

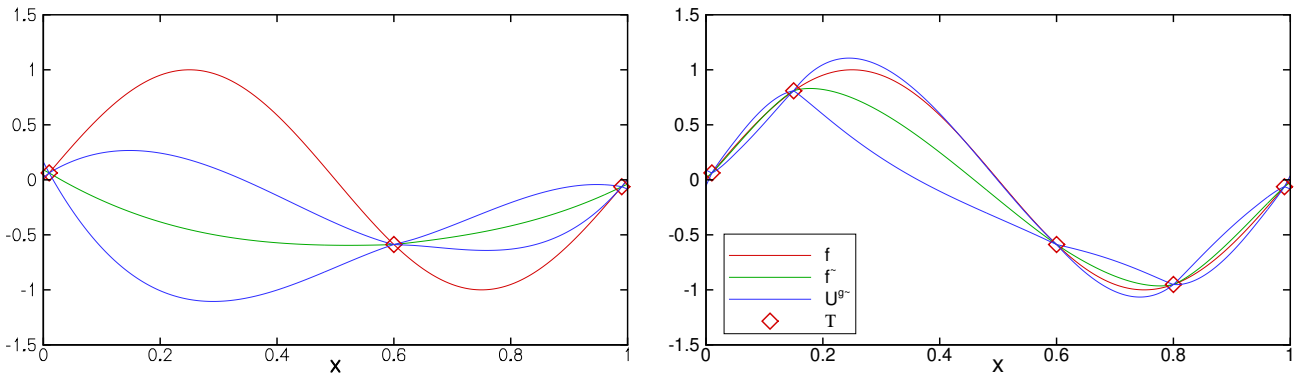

Figure 25-1: SRBF uncertainty estimation $U^{\tilde{g}}(\mathbf{x})$ for a sine-wave function. 2 (left) and 3 (right) points per peak.

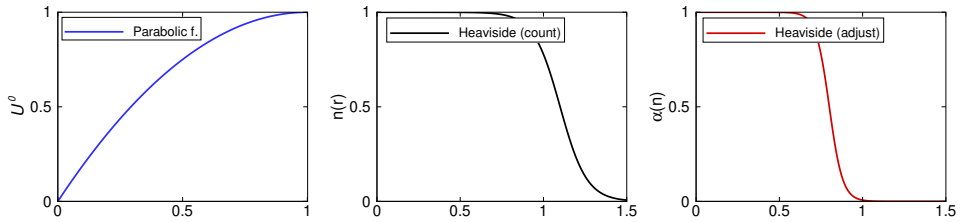

Figure 25-2: Adjustment functions: baseline uncertainty U^0, counting function $\bar{n}(r)$, weight $\alpha(\bar{n})$.

Baseline 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 sought, it is assumed that the function consists of peaks and valleys with a parabolic behavior having a characteristic peak width and height of $2r_0$ and P_{U^0}, respectively. Both these parameters need to be estimated from another source than the training points. In the following, r_0 is estimated as 0.25, considering the domain as normalized, while P_{U^0} is taken as the data range, i.e. $P_{U^0} = \max_i s(\mathbf{x}_i) - \min_i s(\mathbf{x}_i)$.

The assumed parabolic behavior between data points is used to define a baseline uncertainty U^0, based only on the distance to the closest training point $d_{\min}(\mathbf{x}) = \min_{i=1\ldots J} \|\mathbf{x} - \mathbf{x}_i\|$:

$$U^0(\mathbf{x}) = \begin{cases} P_{U^0}\left[1 - \left(\frac{d_{\min}(\mathbf{x})}{r_0} - 1\right)^2\right] & d_{\min}(\mathbf{x})/r_0 \leq 1, \\ P_{U^0} & d_{\min}(\mathbf{x})/r_0 > 1. \end{cases} \tag{25-4}$$

Blending the uncertainties As noted above, the SRBF uncertainty $U^{\tilde{g}}$ 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 baseline 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, a smoothed number ($\bar{n}$) of training points in the sphere r_0 is counted as:

$$\bar{n}(\mathbf{x}) = \sum_{i=1}^{n} \left(1 - \frac{1}{1 + e^{-c_1\left(\frac{r_i(\mathbf{x})}{r_0} - c_2\right)}}\right). \tag{25-5}$$

with $c_1 = 12.5$ and $c_2 = 1.1$. A weight α for the baseline uncertainty is evaluated based on $\bar{n}$:

$$\alpha(\bar{n}) = 1 - \frac{1}{1 + e^{-c_3\left(\frac{\bar{n}}{2D+1} - c_4\right)}}. \tag{25-6}$$

with $c_3 = 25$ and $c_4 = 0.8$. Using α, the modified interpolation uncertainty estimation is defined as:

$$U^{\mathrm{I}}(\mathbf{x}) = U^{\tilde{g}}(\mathbf{x})\left[1 - \alpha\left(\bar{n}(\mathbf{x})\right] + U^0(\mathbf{x})\,\alpha(\bar{n}(\mathbf{x})\right). \tag{25-7}$$

The adjustment functions used in the modified uncertainty are shown in figure 25-2.

25.4.2 Noise Canceling and Training-Point Uncertainty

Least-squares regression of the training set (section 25.3) is effective for filtering noise, but has three disadvantages: (i) the interpolation error estimator $U^{\tilde{g}}$ is ill founded for non-interpolating surrogate models, (ii) the uncertainty $U^{\tilde{g}}$ 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 [1].

Therefore, we define a new noise canceling approach which consists of separating the noise-filtering and interpolation steps. To compute the surrogate, noise-filtered training point values $\bar{s}_i(\mathbf{x}_i)$ are first reconstructed in the training points $\mathbf{x}_i$. Analogous to SRBF, these values are a weighted average of fits with different noise levels; this reduces the risk of overfitting caused by one extreme fit. Then, interpolating SRBF is performed on $\{\mathbf{x}_i, \bar{s}_i(\mathbf{x}_i)\}$ with the error estimation of section 25.4.1, which eliminates the problems (i) and (ii) above.

The uncertainty stemming from the noise-filtering step is evaluated at the training points, interpolated, and added to the interpolation uncertainty. It has two components: (1) the unknown 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 Noise-filtered training point values are derived from least-squares fits of the noisy training set with a number of RBF kernels $M < J$. Depending on the noise that is actually present, not all M are equally likely to provide an accurate regression. 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:

$$e_i^M = \left| s(\mathbf{x}_i) - \widetilde{g}^M_{-i}(\mathbf{x}_i) \right|. \tag{25-8}$$

However, LOOCV requires interpolation towards the point i, which is not reliable if too few training points are available (section 25.4.1). Therefore, the training points which are suitable for cross-validation are selected by assigning weights $\omega_{cv,i} = 1$ to all the training points which have at least $2D+1$ neighbor points within a distance r_0 (the point itself is not counted since it is not available for LOOCV reconstruction). All other points receive $\omega_{CV,i} = 0$.

Then, the errors e_i^M are used to achieve a likelihood estimation for each M. This requires an approximation of the noise level σ_n. Since the reconstruction with the lowest error is probably closest to the true function, the minimum of the LOOCV errors is taken as the closest approximation of the noise:

$$\sigma_n^2 \approx \sigma_{CV}^2 = \min_M \frac{1}{J_{cv}} \sum_{i=1}^{J} \omega_{CV,i} \left(e_i^M\right)^2, \quad \text{with} \quad J_{cv} = \sum_{i=1}^{J} \omega_{CV,i}. \tag{25-9}$$

The likelihood $\mathcal{L}(M)$ that each M provides an accurate regression of the noisy data is the probability of the training point values $s(\mathbf{x}_i)$ given the surrogate models $\widetilde{g}^M_{-i}$ and white noise $\sim \mathcal{N}(0, \sigma_{CV})$ [4]:

$$\mathcal{L}(M) = \prod_{i=1}^{J} p\left(s(\mathbf{x}_i) \,\middle|\, \widetilde{g}^M_{-i}(\mathbf{x}_i)\right) = \prod_{i=1}^{J} \frac{1}{\sigma_{CV}\sqrt{2\pi}} \exp\left(-\tfrac{1}{2}\left(\frac{e_i^M}{\sigma_{CV}}\right)^2\right). \tag{25-10}$$

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

$$\bar{s}_i = \omega_{CV,i} \sum_M \mathcal{L}(M) \widetilde{g}^M(\mathbf{x}_i, 3) + (1 - \omega_{CV,i}) s(\mathbf{x}_i), \qquad \left(\sigma_i^{\bar{s}}\right)^2 = \sum_M \mathcal{L}(M) \left(\widetilde{g}^M(\mathbf{x}_i, 3) - \bar{s}_i\right)^2. \quad (25\text{-}11)$$

Thus, the noise-filtering uncertainty ($\sigma^{\bar{s}}$) 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 points $\bar{s}_i$ are based on the average of $\widetilde{g}^M(\mathbf{x}_i, 3)$.

Mean-value uncertainty Noise filtering is a smoothing, so each $\bar{s}_i$ is a local mean of the data around $\mathbf{x}_i$. However, the local mean of sparse data does not necessarily correspond to the true function $f(\mathbf{x})$. According to the central limit theorem, the standard deviation for the mean of n realizations of a function with stochastic noise σ_n is $\sigma_n/\sqrt{n}$. Applying this theorem to the local mean of spatially varying data $s(\mathbf{x}_i)$ requires (i) an estimation of the noise level and (ii) an indication of how many training points contribute to each local mean value $\bar{s}_i$; both of these vary with M. Following a conservative approach, it is preferable to overestimate the noise, therefore a different estimate than (25-9) is selected: the highest noise level for which the outcome $s(\mathbf{x}_i)$ is in the 95% confidence interval:

$$\sigma_{n,M}^2 = \frac{1}{F^{-1}(0.025, J)} \sum_{i=1}^{J} \left(s(\mathbf{x}_i) - \widetilde{g}^M(\mathbf{x}_i)\right)^2, \quad (25\text{-}12)$$

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 estimated number of training points which contribute to each local mean is the number of training points n_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:

$$\sigma_{m,M}(\mathbf{x}_i)^2 = \frac{\sigma_{n,M}^2}{n_i^M}, \quad (25\text{-}13)$$

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

$$(\sigma_i^{\mu})^2 = \sum_M \mathcal{L}(M) (\sigma_{m,M}(\mathbf{x}_i))^2. \quad (25\text{-}14)$$

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:

$$(\sigma_i^s)^2 = \left(\sigma_i^{\bar{s}}\right)^2 + (\sigma_i^{\mu})^2. \quad (25\text{-}15)$$

The 95% confidence interval U_i^s is assumed equal to $2\sigma_i^s$, it is interpolated with SRBF ($M = J$) and added to U^{I}, as shown in the following section, equation (25-17).

25.4.3 Multi-Fidelity Extension

The uncertainty of the multi-fidelity surrogate $\widehat{f}_l(\mathbf{x})$ in equation (25-1) combines the uncertainties of all its surrogate components. However, thanks to the definition (25-2), in the training points $\{\mathbf{x}_i\}_{i=1}^{J_l}$ the error training data cancel the influence of all lower-fidelity surrogates. Therefore, each multi-fidelity model $\widehat{f}_l$ is a fit through its own data $\mathcal{T}_l$, using the lower-fidelity models for interpolation. The uncertainty induced by the lower-fidelity surrogates can therefore be seen as conditional with respect to their values in the points of $\mathcal{T}_l$.

Writing (25-1) as $\widehat{f}_l = \widetilde{\varepsilon}_l + \widehat{f}_{l+1}$ allows its uncertainty to be decomposed recursively as:

$$\left(U_{\widehat{f}_l}\right)^2 = \left(U^{\mathrm{I}}_{\widetilde{\epsilon}_l}\right)^2 + \left(U^{s}_{\widetilde{\epsilon}_l}\right)^2 + \left(\left(U_{\widehat{f}_{l+1}}\right)^2 \middle| \widehat{\mathbf{f}}^{l}_{l+1}\right), \tag{25-16}$$

where $\widehat{\mathbf{f}}^{l}_{l+1} = \{\widehat{f}_{l+1}(\mathbf{x}_i)\}_{i=1}^{J_l}$. The recursion can be removed by substituting $U_{\widehat{f}_{l+1}}$ evaluated using (25-16) itself, and simplifying. The interpolation error is zero in training points, so $(U^{\mathrm{I}}_{\widetilde{\epsilon}_{l+1}} \,|\, \widehat{\mathbf{f}}^{l}_{l+1}) = U^{\mathrm{I}}_{\widetilde{\epsilon}_{l+1}}$. Furthermore, the data are nested, i.e. $\{\mathbf{x}_i\}_{i=1}^{J_l} \subseteq \{\mathbf{x}_i\}_{i=1}^{J_{l+1}}$, so $\left((U_{\hat{f}_{l+2}} \,|\, \widehat{\mathbf{f}}^{l+1}_{l+2}) \,\middle|\, \widehat{\mathbf{f}}^{l}_{l+2}\right) = (U_{\hat{f}_{l+2}} \,|\, \widehat{\mathbf{f}}^{l+1}_{l+2})$. Thus, equation (25-16) becomes:

$$U^2_{\widehat{f}_l} = \left[\left(U^{\mathrm{I}}_{\widetilde{f}_N}\right)^2 + \sum_{k=N-1}^{l} \left(U^{\mathrm{I}}_{\widetilde{\epsilon}_k}\right)^2\right] + \left[\left(U^{s}_{\widetilde{f}_N} \middle| \widehat{\mathbf{f}}^{N-1}_{N}\right)^2 + \sum_{k=N-1}^{l} \left(U^{s}_{\widetilde{\epsilon}_{k+1}} \middle| \widehat{\mathbf{f}}^{k}_{k+1}\right)^2\right] + \left(U^{s}_{\widetilde{\epsilon}_l}\right)^2. \tag{25-17}$$

Estimation Since no stochastic process is known for the training point values, the conditional training point uncertainties in (25-17) have to be estimated. We suppose that they are 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 $f_{l+1}(\mathbf{x})$. The two-point correlation of the noise-filtered data can be estimated from the local distances between RBF kernels $\mathbf{c}_j^{l+1}$, since these represent data which the noise canceling deems independent. However, the kernel point distance depends on the number of kernels M (section 25.4.2). The following estimation has been implemented:

1. In each training point $\mathbf{x}_i$, the characteristic radius $d_{\mathbf{x}_i}^{l+1,M}$ for each M is the minimum distance of the kernel center into which the training point is clustered to any other kernel center. Like (25-14), the final characteristic radius $d_{\mathbf{x}_i}^{l+1}$ of the point 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:

$$K^{l+1}_{\max}(\mathbf{x}) = \max_{i \in J_l} K\left(\mathbf{x}, \mathbf{x}_i, d^{l+1}_{\mathbf{x}_i}\right). \tag{25-18}$$

The correlation function $K(\mathbf{x}, \mathbf{x}', d)$ has to be estimated. Since the baseline interpolation uncertainty (section 25.4.1) is based on parabolas we use:

$$K\left(r, d\right) = \begin{cases} (r/d - 1)^2 & r/d \leq 1, \\ 0 & r/d > 1, \end{cases} \tag{25-19}$$

where $r = \|\mathbf{x} - \mathbf{x}'\|$. However, this choice is arbitrary; a Gaussian kernel function could also be used. The conditional uncertainty in (25-17) then becomes:

$$\left(U^{s}_{\widetilde{\epsilon}_{l+1}} \middle| \widehat{\mathbf{f}}^{l}_{l+1}\right) \approx U^{s}_{\widetilde{\epsilon}_{l+1}}(\mathbf{x}) \left(1 - K^{l+1}_{\max}(\mathbf{x})\right). \tag{25-20}$$

25.5 ACTIVE LEARNING

The uncertainty estimation of section 25.4 is intended to pilot the dynamic addition of new training points to the multi-fidelity surrogate model. For this, first a new training point $\mathbf{x}^\star$ is identified based on the lower confidence bounding (LCB) with equal weight for the objective function and the uncertainty [10]. It aims to find the lowest point of the uncertainty band around the objective function, i.e. the point where the smallest possible minimum could exist. Since the current uncertainty estimation is designed to overestimate $U_{\hat{f}}$ in case of doubt, there is no need for ad-hoc multiplication of the uncertainty in the acquisition function, which is why equal weights are used. Accordingly, LCB identifies a new training point by minimizing the acquisition function $\psi(\mathbf{x})$:

$$\begin{aligned} \mathbf{x}^\star &= \underset{\mathbf{x}}{\operatorname{argmin}} \left[\psi(\mathbf{x})\right], \qquad \psi(\mathbf{x}) = \hat{f}(\mathbf{x}) - U_{\hat{f}}(\mathbf{x}) + P_x(\mathbf{x}) \\ P_x(\mathbf{x}) &= \begin{cases} \frac{1}{\epsilon}\frac{d_0 - d_{\min}(\mathbf{x})}{d_0} & \text{if } \; d_{\min}(\mathbf{x}) < d_0, \\ 0 & \text{elseif } \; d_{\min}(\mathbf{x}) \geq d_0, \end{cases} \end{aligned} \tag{25-21}$$

where the penalization $P_x(\mathbf{x})$ is added to the main function, to prevent the sampling of already sampled points. ϵ is a coefficient here set equal to 0.1, $d_{\min}(\mathbf{x})$ is the distance of the point $\mathbf{x}$ to the closest training point in any set, and d_0 is the minimum acceptable distance to an existing point.

The fidelity to be evaluated $l^\star$ is chosen as the one which gives the highest expected reduction in the uncertainty, divided by the total cost of sampling the level (which includes sampling all coarser levels $l+1 \ldots N$ since the data are nested). In the error estimation (25-17), sampling a level implies eliminating all interpolation errors up to and including that level, and all the conditional training point uncertainties on the coarser levels. However, we assume that sampling a level does not decrease the training point uncertainty for that level $U^s_{\tilde{\epsilon}_l}$. Thus:

$$l^* = \underset{l=1\ldots N}{\operatorname{argmax}} \frac{\left[U^{\text{interp}}_{\tilde{f}_N} + \sum_{k=N-1}^{l} U^{\text{interp}}_{\tilde{\epsilon}_k}\right] + \left[\left(U^{\text{data}}_{\tilde{f}_N} \,\middle|\, \tilde{f}_N(\mathbf{x}_j^{N-1})\right) + \sum_{k=N-1}^{l}\left(U^{\text{data}}_{\tilde{\epsilon}_{k+1}} \,\middle|\, \hat{f}_{k+1}(\mathbf{x}_j^k)\right)\right]}{\sum_{k=l}^{N} \beta_k}. \tag{25-22}$$

Here β_l represents the evaluation cost of level l divided by the cost for the highest fidelity. While ignoring $U^s_{\tilde{\epsilon}_l}$ is a pessimistic estimation, it discourages the sampling of levels where the uncertainty is dominated by its own training point uncertainty, promoting instead the sampling of higher levels with less noise. This has the same effect as the RS-MSE approach of [3].

25.6 NUMERICAL RESULTS

25.6.1 Uncertainty Estimation

The uncertainty estimation is tested on static training sets, using the 1D MF Forrester function from [11], whose LF function f_2 is modified such that the difference function is a simple straight line. For single-fidelity tests, only the HF function f_1 is used.

$$\begin{aligned} f_1(x) &= (6x-2)^2 \sin(12x-4), \\ f_2(x) &= (6x-2)^2 \sin(12x-4) - 10(x-0.5) - 5. \end{aligned} \tag{25-23}$$

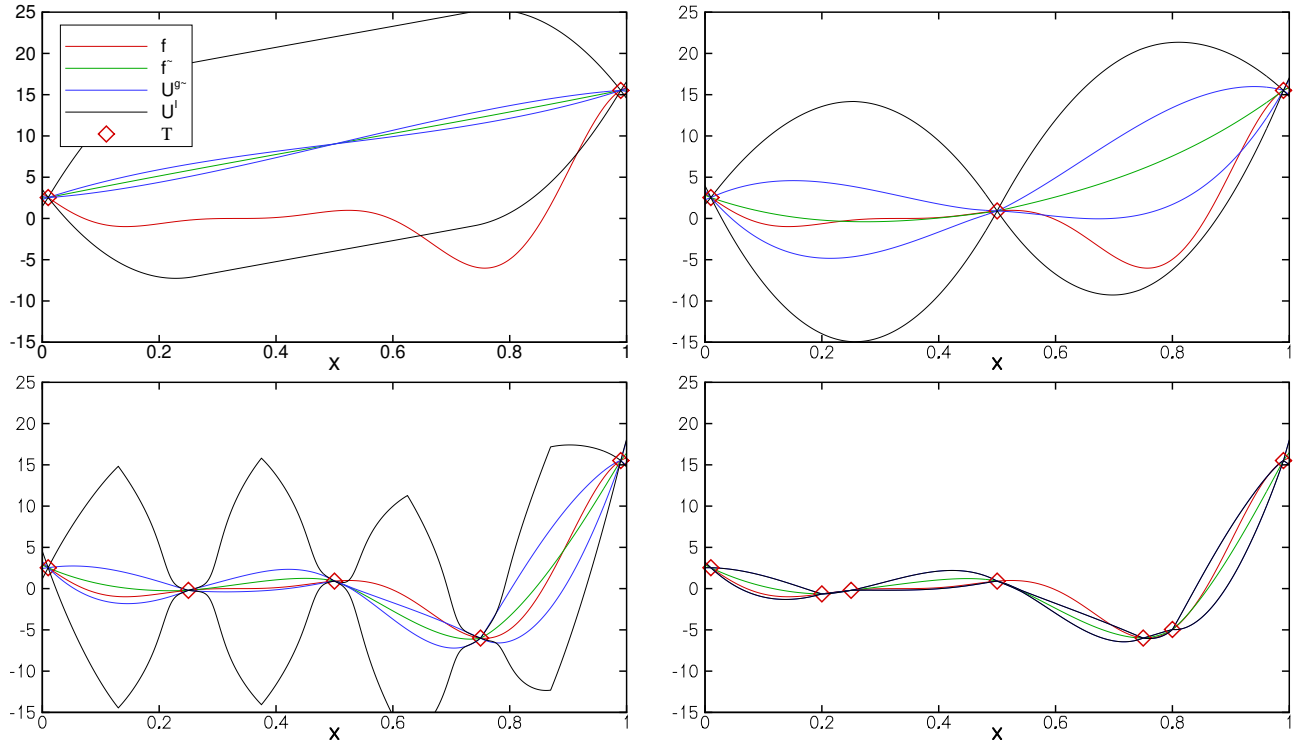

Figure 25-3: Forrester without noise: interpolation uncertainty for 2, 3, 5 and 7 training points.

Interpolation The estimation of the interpolation uncertainty U^{I} is first tested on f_1 without noise. With 2 and 3 training points (figure 25-3) the baseline uncertainty is used everywhere, which is the right choice: unlike $U^{\tilde{g}}$, the modified uncertainty interval contains most of the true function. For 5 data points, $U^{\tilde{g}}$ becomes reliable and U^0 starts to be switched off, while for 7 points, $U^{\tilde{g}}$ 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 25-4 shows three surrogate models for the f_1 Forrester function with noise. The first one has $\sigma_n = 1.5$ (see section 25.3) and clustered data. The interpolation uncertainty varies with the distance between sampling points and even U^0 is used. The training point uncertainty is reduced around the cluster, thanks to the lower mean-value uncertainty. The right figure shows that 3 values of M are the most likely. These coincide both with the minimum of the LOOCV error $\sigma_{CV,M}$ and with the minimum true RMS error err, 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 training points uniformly distributed. For this point density, U^{I} 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.

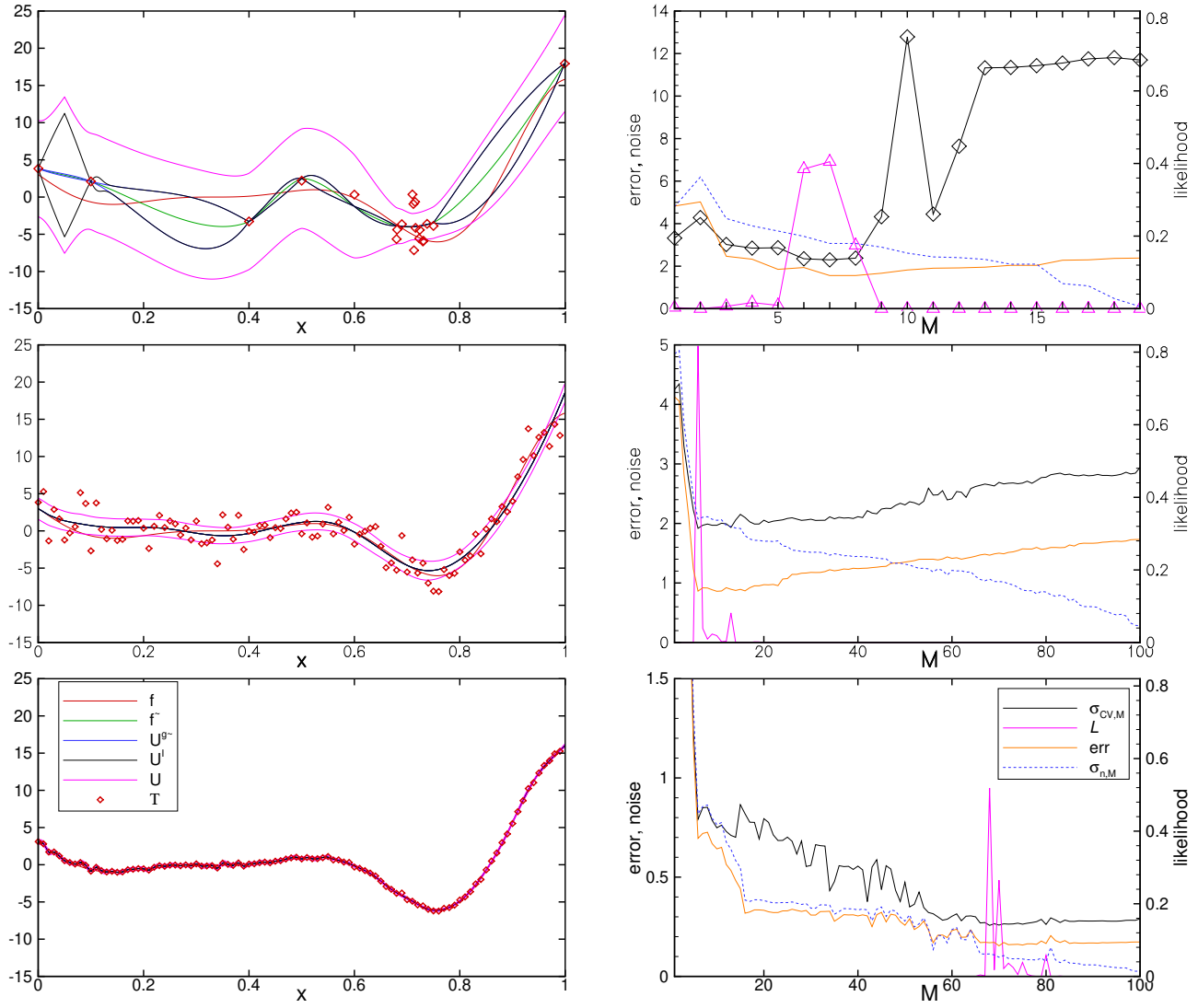

Figure 25-4: Surrogate models for Forrester with noise: 20 points, $\sigma_n = 1.5$ (top), 100 points, $\sigma_n = 1.5$ (middle), and 100 points, $\sigma_n = 0.15$ (bottom). Right figures: the noise and errors of the fits with different M. $\sigma_{CV,M}$ is the function minimized in (25-9), err the RMS difference between the true f and each fit.

Multi-fidelity Finally, a multi-fidelity case is performed with 21 uniformly distributed LF points, 6 HF points, and noise $\sigma_{n_2} = 1.5$ / $\sigma_{n_1} = 0.15$. Figure 25-5 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^{I} is negligible for 21 training points. The 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 25-6.

The weighted uncertainty of the low-fidelity surrogate (figure 25-5 top) is low compared with the error model uncertainty, which therefore makes up most of the MF uncertainty. Since the error model must compensate

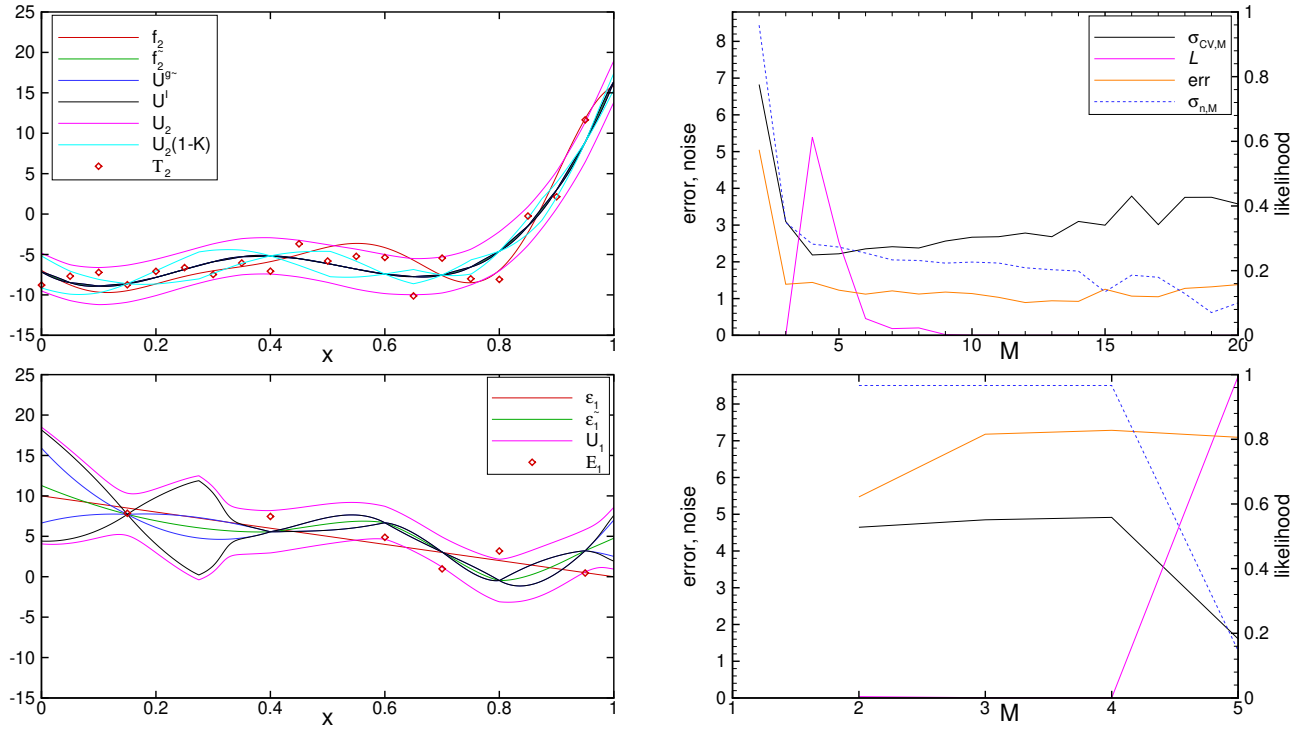

Figure 25-5: Individual surrogate model predictions and uncertainty components for MF Forrester with noise: LF 21 points, $\sigma_{n_2} = 1.5$ (left), HF 6 points, $\sigma_{n_1} = 0.15$ (right). Right figures: the noise and errors of the fits with different M.

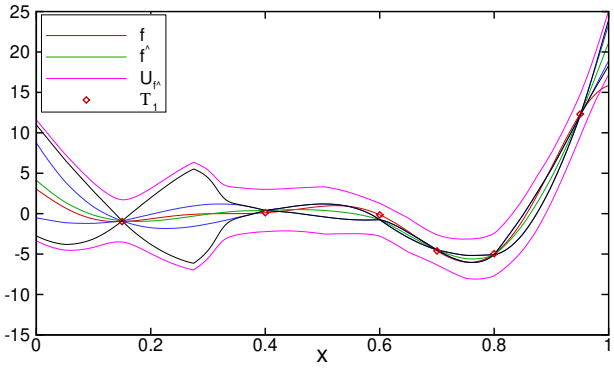

Figure 25-6: Multi-fidelity prediction and uncertainty components for MF Forrester.

for the deficiencies of the low-fidelity model, its complexity is higher than the analytical error function, so it requires more training points for sufficient accuracy. In this situation, the uncertainty estimation shows that further training points should mainly be added to the high-fidelity training set.

25.6.2 Active Learning

The active learning is tested on the optimization of a single-fidelity Forrester function and a NACA 4-digit airfoil with multi-fidelity. The computational cost (CC) is expressed as the equivalent number of HF simulations.

Single-fidelity Forrester For this benchmark, the identification of the global minimum by active learning is performed using the proposed uncertainty model and compared with our earlier work: the least-squares fitted SRBF (LS-SRBF) [3] and interpolated SRBF (see for example [12]). The sampling is started from a set of 3 uniformly distributed points, and the penalization radius d_0 in (25-21) is 10^{-5}, i.e. the penalization is negligible. Computational cost ratios are $\beta_1 = 1$ and $\beta_2 = 0.1$.

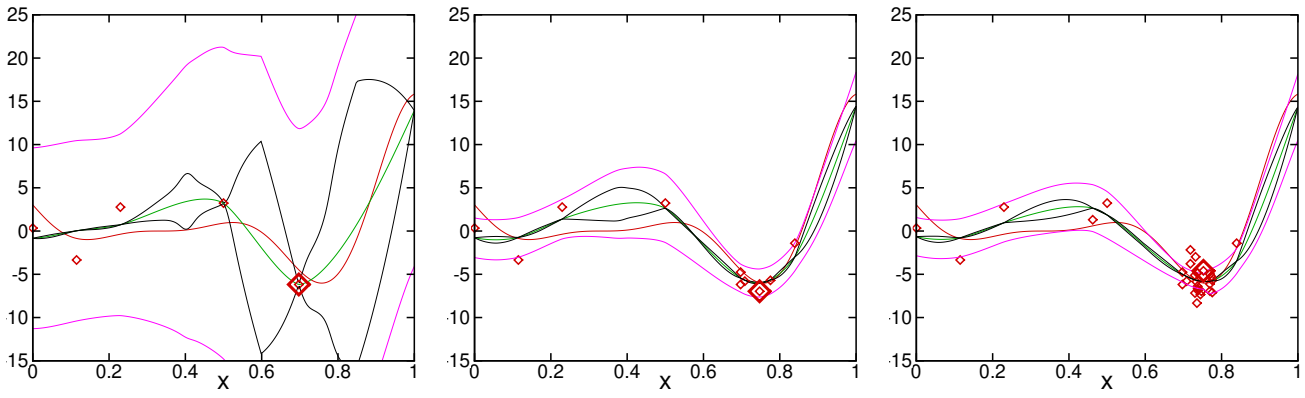

Figure 25-7: Evolution of the Forrester sampling for $\sigma_n = 1.5$, using the new uncertainty estimation. Left to right: CC = 7, 12, and 32. See figure 25-4 for the legend.

Figure 25-7 shows the convergence of the adaptive sampling, using the proposed uncertainty model. In the left image, the dominant interpolation uncertainty improves the exploration of the domain. In the middle image, the estimation has determined the right amount of noise and the error band narrows down. In the last one, the global optimum region has been found and is filled with points, to accurately determine the mean function f.

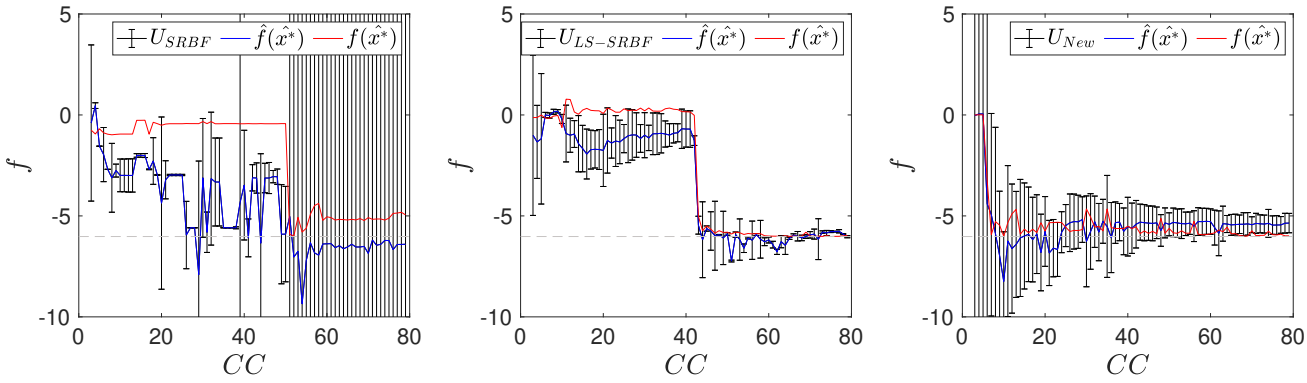

Figure 25-8: Convergence of the predicted uncertainty, the predicted value $\hat{f}$, and the actual value f in the predicted optimum $\mathbf{x}^*$, for $\sigma_n = 1.5$. Left to right: interpolated SRBF, least-squares SRBF, and the new uncertainty estimation.

Figure 25-8 compares the convergence of the three approaches for $\sigma_n = 1.5$. Interpolating SRBF is clearly not suited for this amount of noise: while the optimization converges to roughly the right value, the surrogate model never predicts the true function accurately and the estimated uncertainty has no link with the actual errors. LS-SRBF provides much better surrogate predictions and uncertainty, but the good convergence is a matter of chance: the procedure has no knowledge of the true optimum until CC=40, when one sampled point completely changes the perceived optimum. The new approach is much more reliable: it quickly converges to the true optimum, the surrogate predictions are accurate, and the estimated uncertainty overestimates the actual error. In

the end, the procedure and the uncertainty slowly converge as more points are added around the optimum.

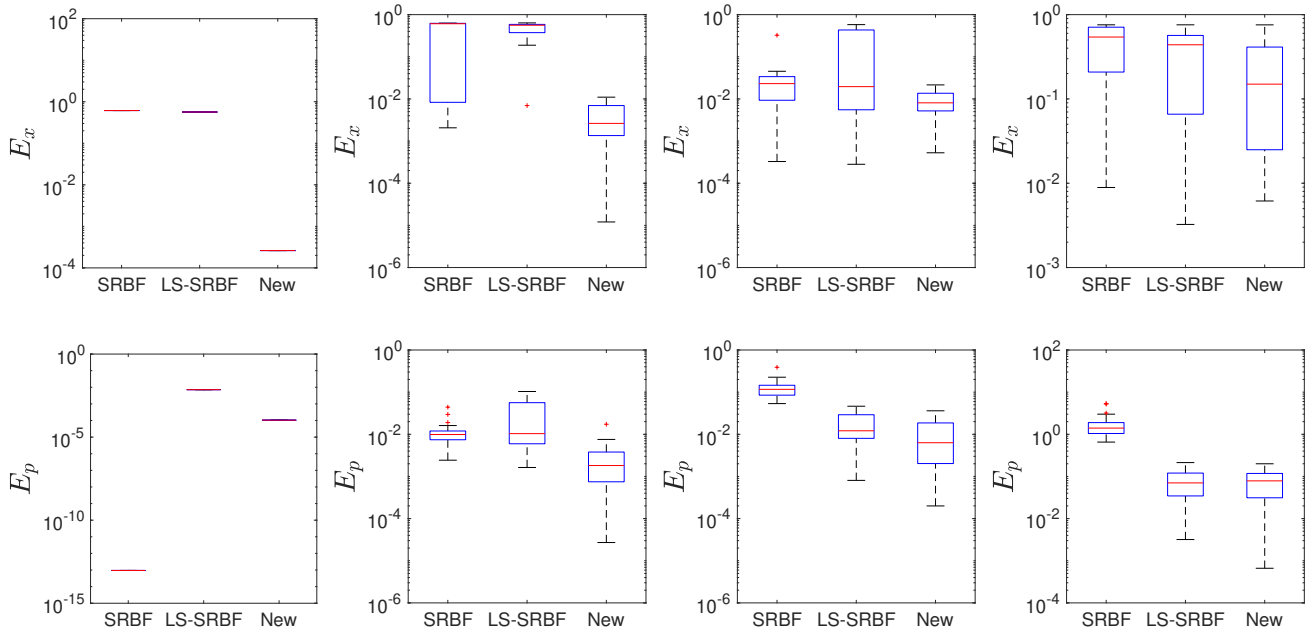

Figure 25-9: Box plots of E_x (top) and E_p (bottom) for the single-fidelity Forrester case with different noise levels. Left to right: $\sigma_n = 0.0$ (CC 20), $\sigma_n = 0.15$ (CC 40), $\sigma_n = 1.5$ (CC 80), and $\sigma_n = 15.0$ (CC 160).

Finally, figure 25-9 shows the result of 25 optimizations with different realizations of the noise, for 4 different noise levels $\sigma_n = 0$, $\sigma_n = 0.15$ (corresponding to $\sim 1\%$ noise), $\sigma_n = 1.5$ ($\sim 10\%$ noise) and $\sigma_n = 15$ ($\sim 100\%$ noise), in terms of the metrics $E_x = \|\hat{\mathbf{x}}^* - \mathbf{x}^*\|/\sqrt{D}$, i.e. the distance between the predicted and true optimum [11], and the prediction error $E_p = \left|\hat{f}_1(\mathbf{x}^*) - f(\mathbf{x}^*)\right| / (f_{\max} - f_{\min})$.

For $\sigma_n = 0$, the surrogates around the optimum are of excellent quality, as shown by the small values of E_p. However, E_x shows that both existing approaches fail to find $\mathbf{x}^*$ since the adaptive sampling is misguided by the three-point startset (c.f. figure 25-1) and never explores the right-hand side of the domain. Thanks to the baseline uncertainty, the new approach explores the entire domain and therefore finds the optimum with high accuracy. The same is true for $\sigma_n = 0.15$, except that the small amount of noise sometimes makes the SRBF approach sample the right-hand side of the domain, leading to good solutions in terms of E_x.

When the noise level increases ($\sigma_n = 1.5$), the noise filtering of LS-SRBF shows its importance, producing better surrogates than SRBF (lower E_p). In terms of E_x however, LS-SRBF is sometimes effective but unreliable. The new uncertainty solves this problem and always leads to acceptable optima. Finally, $\sigma_n = 15$ is an extreme case, for which the optima are of low quality. Still, the new estimation gives a clearly lower median E_x than the other approaches.

Overall, the new approach outperforms the others in both E_x and E_p for all cases (except for E_p at $\sigma_n = 0$, for which all models are essentially perfect). Furthermore, the new approach is reliable, approximating $\mathbf{x}^*$ to within 1% or less for all cases with $\sigma_n \leq 1.5$.

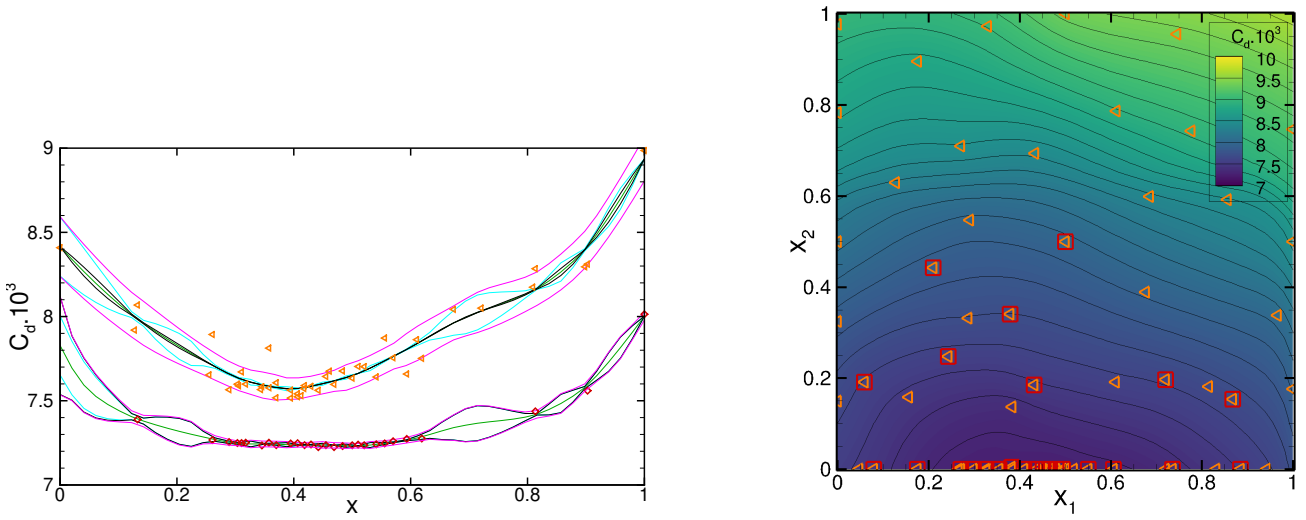

Figure 25-10: NACA surrogates for $D = 1$ (left, see figure 25-6 for the legend), and $D = 2$ (right). The 1D figure combines the LF (top) and MF model (bottom). LF training points are plotted in orange. $CC = 32.6$ for $D = 1$ and 32.3 for $D = 2$.

NACA 4-digit multi-fidelity optimization Aerodynamic shape optimization is performed over the NACA 4-digit parametrized series of airfoils, for $D = 1$ (variation of camber) and $D = 2$ (camber and thickness). While no stochastic noise is added to the data, this optimization is challenging since the simulation results themselves are very noisy, especially for LF. Thus, it is an interesting test to see how the procedure works for non-Gaussian noise.

Simulations are performed with ISIS-CFD developed at ECN/CNRS and two fidelity levels are obtained by varying the mesh size, with LF being four times coarser than HF. The full case definition is given by [3]. Here, the initial training set is the Reduced Startset (RS) from [1], which has only one HF point in the center of the domain. Cost ratios are $\beta_1 = 1$ and $\beta_2 = 0.1$. The penalization radius d_0 is $5 \cdot 10^{-3}$, since the noise for CFD simulations is not random: repeating the same computation twice produces the same result. Therefore, simulations must be placed far enough apart to get different results each time.

Figure 25-10 shows the surrogate models for $D = 1$ and $D = 2$, while figure 25-11 presents the convergence and uncertainty of the optimum. These figures show early exploration of the domain with low-fidelity samples (i.e. points lying closely together in figure 25-11), followed by a switch to HF exploitation once the optimum region is identified. No effort is made to fully cancel the LF noise around the optimum: for this case, the CFD numerical error on the coarse grids is known to be more or less independent of the geometry, but the current results have too few LF points to show this. Instead, the active learning covers the optimum region with HF points. This is highly effective: the points are placed efficiently, the optimization converges quickly and the uncertainty estimations are sensible. With respect to our earlier work (e.g. [3]) the surrogates are much smoother, which indicates effective noise filtering.

CONCLUSIONS AND DISCUSSION

This paper presents an active learning approach for multi-fidelity surrogate models, based on a new uncertainty estimation. This estimation breaks down the uncertainty into contributions from the interpolation, the training

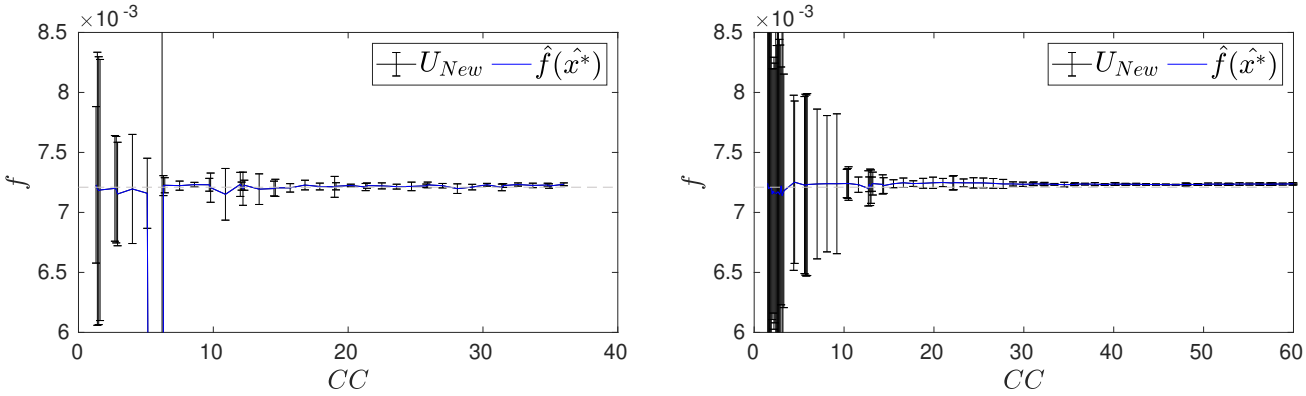

Figure 25-11: NACA optimization: convergence of the predicted optimum and uncertainty for $D = 1$ (left) and $D = 2$ (right).

points, and the conditional uncertainty coming from the low-fidelity models. Furthermore, it accounts for the uncertainty created when the number of training points is too low to accurately represent the objective function. However, since the function behavior cannot be conclusively determined from a finite training set, the estimations necessarily contain ad-hoc elements.

All the main components of our previous work, such as the SRBF surrogate, the least-squares noise filtering, and the LCB active learning are still essential for the new approach. However, compared with our earlier methods, the new uncertainty estimation provides greater reliability. The tests in this paper show consistent good behavior over several test cases and a range of noise conditions. The model is able to automatically adapt itself to the amount of noise that is present.

For the active learning, the main open question concerns the exclusion of the highest level training point uncertainty $U^s_{\tilde{c}_l}$ from the level-selection acquisition function (25-22). As a result, the current approach uses low-fidelity mostly for the initial exploration. As soon as the optimum region is identified, instead of increasing the accuracy for the noisy low-fidelity surrogate, it switches to high-fidelity sampling. While this can be seen as contrary to the spirit of multi-fidelity surrogates, it is a reliable way to achieve highly accurate optimization. Whether this approach is optimal, has to be determined.

Overall, while the current method is open for discussion and further improvement, we consider that the idea of dividing the surrogate uncertainty into several components with different origins and different modeling strategies is validated. More generally, this work shows that for efficient and reliable surrogate modeling based on active learning, the detailed study of surrogate uncertainty is a necessity.

ACKNOWLEDGMENTS

The work at ECN is funded by the Institut Carnot MERS in the ORUP project. CNR-INM is partially supported by the Italian Ministry of University and Research (MUR), through the National Recovery and Resilience Plan (PNRR), Sustainable Mobility Center (CNMS), Spoke 3 Waterways, CN00000023 - CUP B43C22000440001. The work is conducted in collaboration with NATO STO AVT task group on ”Enhanced Design Processes of Military Vehicles through Machine Learning Methods” (AVT-404).

REFERENCES

[1] J. Wackers, R. Pellegrini, A. Serani, M. Visonneau, and M. Diez, "Efficient initialization for multi-fidelity surrogate-based optimization," *Journal of Ocean Engineering and Marine Energy*, vol. 9, no. 2, pp. 291–307, 2023.

[2] A. Serani, T. P. Scholcz, and V. Vanzi, "A scoping review on simulation-based design optimization in marine engineering: Trends, best practices, and gaps," *Archives of Computational Methods in Engineering*, pp. 1–29, 2024.

[3] R. Pellegrini, J. Wackers, R. Broglia, A. Serani, M. Visonneau, and M. Diez, "A multi-fidelity active learning method for global design optimization problems with noisy evaluations," *Engineering with Computers*, vol. 39, no. 5, pp. 3183–3206, 2023.

[4] C. E. Rasmussen and C. K. I. Williams, *Gaussian Processes for Machine Learning*. The MIT Press, 2005. [Online]. Available: https://doi.org/10.7551/mitpress/3206.001.0001.

[5] S. Volpi, M. Diez, N. J. Gaul, *et al.*, "Development and validation of a dynamic metamodel based on stochastic radial basis functions and uncertainty quantification," *Structural and Multidisciplinary Optimization*, vol. 51, pp. 347–368, 2015.

[6] H. Liu, Y.-S. Ong, and J. Cai, "A survey of adaptive sampling for global metamodeling in support of simulation-based complex engineering design," *Structural and Multidisciplinary Optimization*, vol. 57, pp. 393–416, 2018.

[7] J. Wackers, A. Serani, R. Pellegrini, M. Diez, and M. Visonneau, "Adaptive multifidelity shape optimization based on noisy CFD data," in *Adaptive Modeling and Simulation (ADMOS 2019)*, 2019.

[8] J. Wackers, H. Pehlivan Solak, R. Pellegrini, A. Serani, and M. Diez, "Error estimation for surrogate models with noisy small-sized training sets," in *Adaptive Modelling and Simulation (ADMOS 2023)*, 2023.

[9] J. Wackers, H. Pehlivan Solak, R. Pellegrini, A. Serani, and M. Diez, "Adaptivity and uncertainty of multi-fidelity surrogate models for shape optimization," in *16th World Congress in Computational Mechanics (WCCM-PANACM)*, 2024.

[10] D. D. Cox and S. John, "A statistical method for global optimization," in *IEEE International Conference on Systems, Man, and Cybernetics*, Chicago, IL, 1992, pp. 1241–1246.

[11] L. Mainini, A. Serani, M. P. Rumpfkeil, *et al.*, "Analytical benchmark problems for multifidelity optimization methods," *arXiv preprint arXiv:2204.07867*, 2022.

[12] A. Serani, R. Pellegrini, J. Wackers, *et al.*, "Adaptive multi-fidelity sampling for CFD-based optimisation via radial basis function metamodels," *International Journal of Computational Fluid Dynamics*, vol. 33, no. 6-7, pp. 237–255, 2019.