Abstract The custom in surrogate-based modeling of complex engineering problems is to fit one or more surrogate models and select the one surrogate model that performs best. In this paper, we extend the utility of an ensemble of surrogates to (1) identify regions of possible high errors at locations where predictions of surrogates widely differ, and (2) provide a more robust approximation approach. We explore the possibility of using the best surrogate or a weighted average surrogate model instead of individual surrogate models. The weights associated with each surrogate model are determined based on the errors in surrogates. We demonstrate the advantages of an ensemble of surrogates using analytical problems and one engineering problem. We show that for a single problem the choice of test surrogate can depend on the design of experiments.

Keywords Multiple surrogate models · Polynomial response surfaces · Kriging · Radial basis neural networks

1 Introduction

Surrogate models have been extensively used in the design and optimization of computationally expensive problems. Different surrogate models have been shown to perform well in different conditions. Barthelemy and Haftka (1993) reviewed the application of meta-modeling techniques in structural optimization. Sobieszczanski-Sobieski and Haftka (1997) reviewed different surrogate modeling applications in multidisciplinary optimization. Giunta and Watson (1998) compared polynomial response surface approximations and Kriging on analytical example problems of varying dimensions. Simpson et al. (2001) reviewed different surrogates and gave recommendations on the usage of different surrogates for different problems. Jin et al. (2001) compared different surrogate models based on multiple performance criteria such as accuracy, robustness, efficiency, transparency, and conceptual simplicity. They recommended using radial basis function for high-order nonlinear problems, Kriging for low-order nonlinear problems in high dimension spaces, and polynomial response surfaces for low-order nonlinear problems. They also noted difficulties in constructing different surrogate models. Li and Padula (2005) and Queipo et al. (2005) recently reviewed different surrogate models used in the aerospace industry.

There are also a number of studies comparing different surrogates for specific applications. Papila et al. (2001), Shyy et al. (2001), Vaidyanathan et al. (2004), Mack et al. (2005b) presented studies comparing radial basis neural networks and response surfaces while designing the liquid rocket injector, supersonic turbines, shape of bluff body. For crashworthiness optimization, Stander et al. (2004) compared polynomial response surface approximation, Kriging, and neural networks, while Fang et al. (2005) compared polynomial response surface approximation and radial basis functions. Most researchers observed that no single surrogate model was found to be the most effective for all problems.

While most researchers have primarily been concerned with the choice among different surrogates, there has been relatively very little work about the use of an ensemble of surrogates. Zerpa et al. (2005) presented one application of using an ensemble of surrogates to construct weighted average surrogate model for the optimization of alkali–surfactant–polymer flooding process. They suggested that the weighted average surrogate model has better modeling capabilities than individual surrogates.

Typically, the cost of obtaining data required for developing surrogate models is high, and it is desired to extract as much information as possible from the data. Using an ensemble of surrogates, which can be constructed without a significant expense compared to the cost of acquiring data, can prove effective in distilling correct trends from the data and may protect against bad surrogate models. Averaging
surrogates is one approach motivated by our inability to find a unique solution to the nonlinear inverse problem of identifying the model from a limited set of data (Queipo et al. 2005). In this context, model averaging essentially serves as an approach to account for model uncertainty.

In this work, we explore methods to exploit the potential of use of an ensemble of surrogates. Specifically, we present the following two aspects:

1. Ensemble of surrogates can be used to identify regions where we expect large uncertainties (contrast).
2. Use of an ensemble of surrogates via weighted averaging (combination) or selection of best surrogate model based on error statistics for more robust approximation than individual surrogates.

This paper is organized as follows. In the next section, we present a method to use an ensemble of surrogates to identify the regions with large uncertainty, and the conceptual framework of constructing weighted average surrogate models. Thereafter, we discuss the test problems, numerical procedure, and results supporting our claims. We close the paper by recapitulating salient points presented.

2 Conceptual framework

2.1 Identification of region of large uncertainty

Surrogate models are used to predict the response in unsampled regions. There is an uncertainty associated with the predictions. An ensemble of surrogates can be used to identify the regions of large uncertainty. The concept is described as follows: Let there be \( N_{SM} \) surrogate models. We compute the standard deviation of the predictions at a design point \( \mathbf{x} \) as,

\[
s_{\text{exp}}(\hat{y}(\mathbf{x})) = \sqrt{\frac{\sum_{i=1}^{N_{SM}} (\hat{y}_i(\mathbf{x}) - \bar{y}_i(\mathbf{x}))^2}{N_{SM} - 1}}
\]

where \( \bar{y}_i(\mathbf{x}) = \frac{\sum_{i=1}^{N_{SM}} \hat{y}_i(\mathbf{x})}{N_{SM}} \) (1)

The standard deviation of the predictions will be high in regions where the surrogates differ greatly. A high standard deviation may indicate a region of high uncertainty in the predictions of any of the surrogates, and additional sampling points in this region can reduce that uncertainty. Note that while high standard deviation indicates high uncertainty, low standard deviation does not guarantee high accuracy. It is possible for all surrogate models to predict similar response (yielding low standard deviation) yet perform poorly in a region.

2.2 Weighted average surrogate model concept

We develop a weighted average surrogate model as

\[
\hat{y}_{\text{wt.avg}}(\mathbf{x}) = \sum_{i=1}^{N_{SM}} w_i(\mathbf{x}) \hat{y}_i(\mathbf{x})
\]

where, \( \hat{y}_{\text{wt.avg}}(\mathbf{x}) \) is the predicted response by the weighted average of surrogate models, \( \hat{y}_i(\mathbf{x}) \) is the predicted response by the \( i \)th surrogate model, and \( w_i(\mathbf{x}) \) is the weight associated with the \( i \)th surrogate model at design point \( \mathbf{x} \). Furthermore, the sum of the weights must be one \( \left( \sum_{i=1}^{N_{SM}} w_i = 1 \right) \), so that if all the surrogates agree, \( \hat{y}_{\text{wt.avg}}(\mathbf{x}) \) will also be the same.

A surrogate model that is deemed more accurate should be assigned a large weight, and the less accurate model should have less influence on the predictions. The confidence in surrogate models is given by different measures of “goodness” (quality of fit), which can be broadly characterized as (1) global vs local measures and (2) measures based on surrogate models vs measures based on data. Weights associated with each surrogate, based on the local measures of goodness, are functions of space \( w_i = w_i(\mathbf{x}) \). For example, weights, which are based on the pointwise model error, estimates like prediction variance, mean squared error (surrogate based), or weights based on the interpolated cross-validation errors (data based). When weights are selected based on the global measures of goodness, they are fixed in design space \( w_i = w_i(\mathbf{x}) \); examples are weights based on RMS error \( \hat{\sigma} \) for polynomial response surface approximation, process variance for Kriging (surrogate based), or weights based on cross-validation error (data based). While variable weights may capture local behavior better than constant weights, reasonable selection of weight functions is a formidable task.

Zerpa et al. (2005) constructed a local weighted average model from three surrogates (polynomial response surface approximation, Kriging, and radial basis functions) for the optimization of an alkali–surfactant–polymer flooding process. Their approach was based on the pointwise estimate of the variance predicted by the three surrogate models.

In this work, we propose a global weights selection scheme based on global data-based measure of goodness. There are many possible strategies of selecting weights. A few can be enumerated as follows:

1. WTA1:

Weights are a function of relative magnitude of (global data-based) errors. The weight associated with \( i \)th surrogate is given as:

\[
w_i = \frac{\sum_{j=1}^{N_{SM}} E_j}{(N_{SM} - 1) \sum_{j=1}^{N_{SM}} E_j}
\]

(3)