Proceedings of the ASME 2017 International Design Engineering Technical Conferences & Computers and Information in Engineering Conference
IDETC/CIE 2017
August 6-9, 2017, Cleveland, Ohio, USA

DETC2017-67794

INVESTIGATING GREY-BOX MODELING FOR PREDICTIVE ANALYTICS IN SMART MANUFACTURING

Zhuo Yang, Douglas Eddy, Sundar Krishnamurty, Ian Grosse
University of Massachusetts Amherst
Department of Mechanical and Industrial Engineering
Amherst, MA 01003
Email: [zhuoyang, dceddy]@engin.umass.edu, [skrishna, grosse]@ecs.umass.edu

Peter Denno, Yan Lu, Paul Witherell
National Institute of Standards and Technology
Engineering Laboratory
Gaithersburg, MD 20899
Email: [peter.denno, yan.lu, paul.witherell]@nist.gov

ABSTRACT

This paper develops a two-stage grey-box modeling approach that combines manufacturing knowledge-based (white-box) models with statistical (black-box) metamodels to improve model reusability and predictability. A white-box model can use various types of existing knowledge such as physical theory, high fidelity simulation or empirical data to build the foundation of the general model. The residual between a white-box prediction and empirical data can be represented with a black-box model. The combination of the white-box and black-box models provides the parallel hybrid structure of a grey-box. For any new point prediction, the estimated residual from the black-box is combined with white-box knowledge to produce the final grey-box solution. This approach was developed for use with manufacturing processes, and applied to a powder bed fusion additive manufacturing process. It can be applied in other common modeling scenarios. Two illustrative case studies are brought into the work to test this grey-box modeling approach; first for pure mathematical rigor and second for manufacturing specifically. The results of the case studies suggest that the use of grey-box models can lower predictive errors. Moreover, the resulting black-box model that represents any residual is a usable, accurate metamodel.

1. INTRODUCTION

Smart manufacturing is becoming increasing possible as access to technology improves. Industry is, and will continue to be, increasingly reliant on data and predictive analytics to improve overall process efficiencies [1]. With this trend, industry is now collecting data at never before seen rates in hopes of gaining competitive advantages related to their products and processes. Often data are collected without regard for their interrelation, and it is not readily apparent how the collected information can be used to improve system efficiencies. To address this issue, we investigate a novel metamodeling technique based on the context from which the data was acquired and the domain in which it is relevant.

White-box modeling methods use knowledge such as rules and theories, to formulate models such as those that represent physical phenomena. Such classical white box modeling methods have been used for thousands of years. Newton's Law or Euler–Bernoulli bending theory [2] is a classic example of a traditional physical model. Such models usually require comprehensive knowledge of the target system and are usually represented by parametric formulation. For instance, in manufacturing, physics-based models are often derived from theoretical analysis that mostly focuses on individual sub-processes with idealized assumptions. In reality, the actual multi-physical system of manufacturing may involve numerous interactions among these sub-processes. Such complexity can be difficult to fully understand. For example, in additive manufacturing (AM), the isotherm migration method develops a thermal model to calculate the temperature on a powder surface being heated by a laser beam modeled as a point source

[3]. However, the complex inter-relationship between parameters of Powder Bed Fusion (PBF) processes renders these theoretical analyses insufficient for the needs of many practical applications [4].

Metamodels, also known as surrogate models, are statistical models that use a "black-box" approach to represent unknown systems and hence do not require detailed knowledge of the underlying physical phenomena [5]. Metamodeling focuses on the input/output parametric values while ignoring the complex inter-relationships within the unknown system (i.e. the black-box, which statistically approximates the relationships based purely upon data values). Rather than incorporating any physical knowledge, the predictability of metamodels completely relies upon statistical features such as sampling strategy and modeling algorithm. Metamodels can optimally reduce the inaccuracies that arise from incomplete knowledge. For example, an adaptive sampling method that iteratively updates the metamodel with a new sample data point can gradually refine the model's predictive power [6, 7].

Furthermore, predictability or efficiency of a metamodel can be improved by selecting the most suitable algorithm [8] and/or best combination of algorithms [9] in cases where it is not possible to acquire more data points.in the desired locations in the design space. Advanced modeling algorithms such as a dynamic Kriging method and artificial neural networks (ANN) [10] can significantly improve model efficiency and predictability for these types of inflexible datasets. However, such metamodels built by pure statistical approaches usually lack information about the model's physical meaning and assumptions due to a large degree of data-dependency. Moreover, modeling inaccuracies might accumulate during the model construction process due to the lack of physical knowledge about the critical features of the represented system [5]. Thus, both white and black box approaches alone have accuracy limitations due to different reasons.

Due to these intrinsic limitations in both approaches, a technique that can harness the advantages of both white and black box models while reducing their disadvantages is desirable for complex problems with understood subdomains. The modeling approach known as grey-box, or hybrid modeling, was invented to combine the benefits of domain knowledge and empirical information [11]. The models generated by this approach can obey general physical rules (white box) while optimizing the parameters from actual experimental data (black box).. Many of the newer and less established white box manufacturing physics-based models and numerical simulations may be founded on incomplete and/or inaccurate knowledge and idealized assumptions. For example, the previously mentioned isotherm migration model in AM does not account for the influence of powder particle size, part geometry, and environmental conditions [3]. The calculated solutions from current AM physics-based models are usually limited in the scope of what they describe, diminishing their predictive capability. Though AM metamodels can potentially avoid these errors, they usually require a large number of expensive samples and may not be reusable. Many examples in smart manufacturing have similar modeling challenges. These barriers potentially limit the adaptability of metamodeling in any manufacturing domain. Thus, neither approach can optimally construct robust, usable manufacturing models alone.

This paper aims to develop a grey-box modeling approach which combines the benefits of traditional serial methods (where black and white box knowledge is applied sequentially) and parallel methods (where knowledge from black and white boxes is composed before being applied). The result is a hybrid combination of knowledge of physical phenomena and statistical information. To address the challenge of combining knowledge of physical phenomena with statistical information, a two-stage approach is used. The first stage deploys a serial grey–box approach to build a statistical black-box model to estimate the errors caused by the inaccuracies in a white-box. The second stage uses the model from the first stage to estimate the basic solution and the residual solution. The final solution is a combination of these two.

Section 2 and 3 provide fundamental background knowledge relating to metamodeling and grey-box modeling techniques. Section 4 introduces this general algorithm of a two-stage grey-box modeling approach for additive or smart manufacturing. Case studies using a general mathematical example and a representative metal PBF AM problem are presented in Section 5. Section 6 discusses the results and identifies future work for this study.

2. OVERVIEW OF METAMODELING TECHNIQUES

This section briefly reviews the metamodeling techniques used in this work. Polynomial regression (PR) and Kriging methods were investigated in this study as they cover the spectrum of both parametric and non-parametric modeling algorithms [12]. PR is popular in that its model is represented by parametric formulation. The Kriging method, on the other hand, is an interpolation approach that uses positioning information for data estimation instead of conventional mathematical formulation. The general mathematical formulation of any metamodel can be expressed as:

$$y(\tilde{x}) = f(\tilde{x}) + \varepsilon \quad (1)$$

where $y(\tilde{x})$ represents the actual output for new point $\tilde{x}$ [13]. $f(\tilde{x})$ is a known function derived statistically from data that produces the model estimate as a function of $\tilde{x}$, ε is any residual error, and $\tilde{x}$ represents the set of the independent input variables. For different modeling methods, the composition of each of these elements could be different.

2.1 Polynomial Regression

PR is a higher order variation of linear regression in which an n^{th} order polynomial is used to formulate the relationship between the independent variables x and the outcome y [13]. It is popular in various engineering domains due to its efficiency. A second order quadratic polynomial function would have the form:

$$\hat{y} = \beta_0 + \sum_{i=1}^{k} \beta_i x_i + \sum_{i=1}^{k} \beta_{ii} x_i^2 + \sum_{i=1}^{k} \sum_{j=1}^{k} \beta_{ij} x_i x_j \quad (2)$$

where β_0, β_i, and β_{ij} are regression coefficients, and k is the number of design variables.

2.2 Kriging

Unlike parametric techniques that produce an actual formulation. Kriging methods are non-parametric methods that build their estimation based on the position of the samples. The basic assumption in kriging is that the estimating point (unknown point) can be represented by observed points (known points) based on spatial correlation [14]. The estimation process is completed by a variogram or so called spatial correlation functions [15, 16]. The general form of kriging estimation for an unknown predicted value of a point Z_E for a single outcome is:

$$Z_E = \bar{Z} + \sum_{i=1}^{n} \lambda_i (Z_i - \bar{Z}) \quad (3)$$

where $\bar{Z}$ represents the regional mean value of the response and λ_i is the distance-correlated weight value, which is determined by the computation of spatial correlation.

To approach the weight value, one should first compute the spatial correlation R between data points. The value of spatial correlation can be derived from:

$$R(\theta, x_i, x_j) = \prod_{l=1}^{n} \exp(-\theta_l (x_{i,l} - x_{j,l})^2) \quad (4)$$

where $x_{i,l}$ is the lth component of the i^{th} vector x_i [17]. $R(\theta, x_i, x_j)$ depends upon the location of points x_i and x_j, and the correlation parameter, θ. Kriging methods have multiple forms such as simple kriging, ordinary kriging, regression kriging, etc.[16]. In this paper, results from kriging models are obtained from the ordinary kriging method.

The correlation matrix can then be formulated. A problem with n given data points can be presented as:

$$\mathbf{C} = \begin{bmatrix} R(\theta, x_1, x_1) & R(\theta, x_1, x_2) & \dots & R(\theta, x_1, x_n) \\ R(\theta, x_2, x_1) & R(\theta, x_2, x_2) & \dots & R(\theta, x_2, x_n) \\ \vdots & \vdots & \ddots & \vdots \\ R(\theta, x_n, x_1) & R(\theta, x_n, x_2) & \dots & R(\theta, x_n, x_n) \end{bmatrix} \quad (5)$$

Similarly, the correlation vector B that presents the correlation between the new point x_E and all given points is formulated as:

$$\mathbf{B} = \begin{bmatrix} R(\theta, x_1, x_E) \\ R(\theta, x_2, x_E) \\ \vdots \\ R(\theta, x_n, x_E) \end{bmatrix} \quad (6)$$

The weight vector $\Lambda = [\lambda_1, \lambda_2, \dots \lambda_n]^T$ can now be calculated by C and B:

$$L = C^{-1} B \quad (7)$$

3. OVERVIEW OF GREY-BOX MODELS

A grey-box model is a hybrid model that combines different types of models such as physics-based models, numerical simulation models and statistical models [18]. The term “grey-box” stems from the mixture of white-box and black-box models. A conventional grey-box model uses a physical formulation to maintain the physical interpretation and uses data to estimate parameters [18]. In general, the basic structure of a grey-box model is inherited from knowledge and further improved by experimental data.

Grey-box model development can be summarized into three steps: 1) construct the foundation for the system with a simplified knowledge model; 2) determine the physical parameters from the description of the system behavior; 3) identify the value of model parameters from actual data [8]. The relationship between these three types of model and knowledge sources is shown in Figure 1. The proposed grey-box metamodeling method was developed based upon this viewpoint.

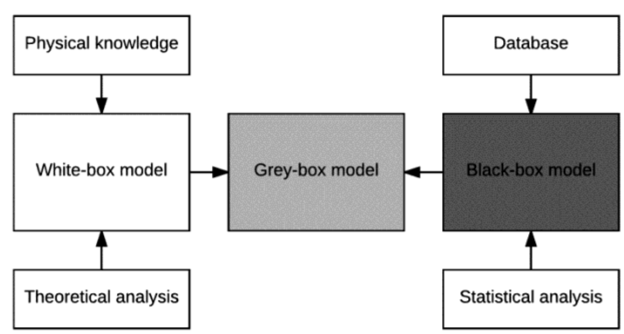

Figure 1. Relationships among physics-based white-box, statistics-based black-box, hybrid grey-box models and knowledge sources

Grey-box models can be generally classified into serial approach and parallel approaches [19, 20], which are shown in Figure 2. A serial approach aims to sequentially fill the gap between knowledge and experimental data. For example, the uncertainties raised from incomplete knowledge of a white-box model can be reduced by accompanying that model with actual data. A parallel approach, alternatively, aims to use both models together to estimate the correct results that would be difficult to approach by either a white or black-box model individually. A grey-box model with serial structure focuses on reducing the error between the prediction from physical model and actual result from experiments. For example, Duarte et al. [21] developed a hybrid modeling approach that combined knowledge and mechanistic, rather than statistical models, to improve traditional model performance. In this approach, the first model is built based on first-principles system behavior, and the second model estimates the residuals between real data and mechanistic predictions.

A traditional grey-box model with parallel structure uses data to estimate the correct values of model responses which are difficult to approach given incomplete knowledge of phenomena [22]. For example, Psichogios et al. [19] hybrid neural network model utilizes a partial first principles model. This modeling approach combines available prior knowledge with an artificial neural network (ANN) to derive an estimator of unmeasured process parameters. This hybrid structure can interpolate and extrapolate much more accurately than a standard "black-box" ANN with significantly fewer training sample points to accompany the knowledge model.

The next section introduces the two-stage grey-box modeling approach developed for manufacturing problems in this study. To get the final prediction, the data serially flows into both types of grey-box models for the purpose of constructing the black-box model and estimating the residual. To demonstrate this approach, we chose a complex manufacturing process that we believe could particularly benefit. The AM-specific grey-box modeling approach that is introduced next in Section 4 is built upon both the Type II serial approach and parallel approach shown in Figure 2 based on current AM challenges.

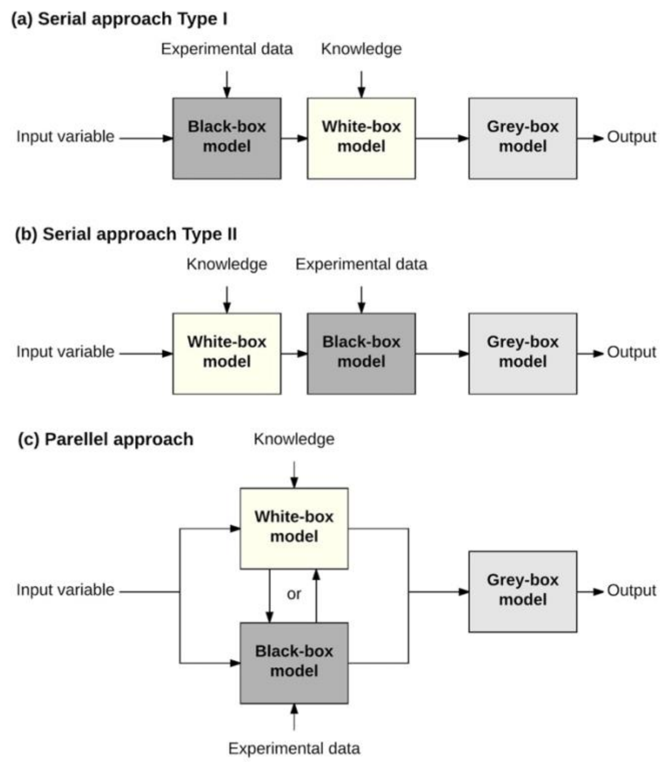

Figure 2. Basic grey-box modeling approaches

4. GREY-BOX METAMODELING FOR AM

Additive manufacturing (AM) and other smart manufacturing techniques are being widely used in various domains such as aerospace [23], medical devices [24] and energy systems [25]. However, many barriers and challenges, such as the large uncertainty of AM process results, have prevented its further adoption in industry [25]. The relationship between process parameters and mechanical properties are not fully understood for AM processes. For example, relative density, one of the major structural properties of the parts produced by metal PBF processes, depends upon multiple AM parameters such as laser power, scan speed, pulse frequency, and layer thickness [26]. Previous studies show that a typical metal PBF process consists of four general sub-systems classified by related physical phenomenon. Each sub-system can be further divided into multiple sub-processes [27]. A general AM process can involve more than fifty independent parameters [4]. For example, the melt pool sub-system is related to a number of factors that involve both thermal and fluid mechanics [28]. Though difficult, AM models built upon theoretical analysis, numerical simulation and statistical modeling have been developed for predictive purposes in recent years [7].

The general procedure to construct an AM grey-box model by using the approach introduced in this paper is shown in Figure 3 and Figure 4. First, the method builds the white-box model from available prior knowledge. If knowledge was derived from theoretical analysis, the white-box model can be directly represented by a parametric formulation. Alternatively, a parametric model can be derived through an approximation of a physics-based model using a formulation such as FEA for computational simplification. However, if the knowledge is based on a complex numerical simulation that requires high computational cost, the white-box model can be redesigned to be represented by some simplified parametric function such as a PR response surface model to more rapidly estimate a white box model using fewer sample points of that expensive data. The sampling data to construct that PR model could be collected by a technique such as space filling sampling (SFS) or sequential infilling sampling (SIS) [6] from data generated by an adequate number of simulations. It can be expected that the solution from the white-box model would contain large errors due to limited knowledge.

The next step is to build a black-box model from additional information. Potential sources of the additional information could be actual experimental data or a higher fidelity simulation used to generate the data. The black box model captures both the discrepancy between a lower fidelity FEA type of model and the real process as well as the discrepancy between a high fidelity model (FEA model) and the simplified model. The input values of this additional data are entered into the constructed white-box model to calculate the corresponding output responses. This computational output from the white-box model is next compared with the actual output values of the additional data to calculate each difference. This difference can be considered the white box model's actual residual at that data location as shown in Figure 3. Since each pair of the computational output and the actual output has the same input variables, the residual value directly represents the accumulated errors caused by incomplete and/or imperfect knowledge used to construct the white box model. The black-box model is used here to evaluate the relation between input variables and the estimated residual.

At this first stage, the serial grey-box structure is established based on the type II serial approach that is shown in Figure 2b: the output from the white-box becomes an intermediate input to the black-box. This serial grey-box approach is used to build a black-box model to estimate the residual value that cannot be derived from the white-box alone. That residual is the difference between the responses predicted by the white and black boxes. The inputs to the black box are those used to generate the white-box responses. The black-box model uses the kriging method to model the relationship between input variable values and residual response values. The kriging method is applied since this interpolation approach helps to avoid any significant intrinsic error in the resulting model [29]. Once the black-box is created, it can compute the estimated residual for any new data point. The approach, illustrated in Figure 3, establishes the black-box model used to derive the grey-box model created in the subsequent steps shown in Figure 4.

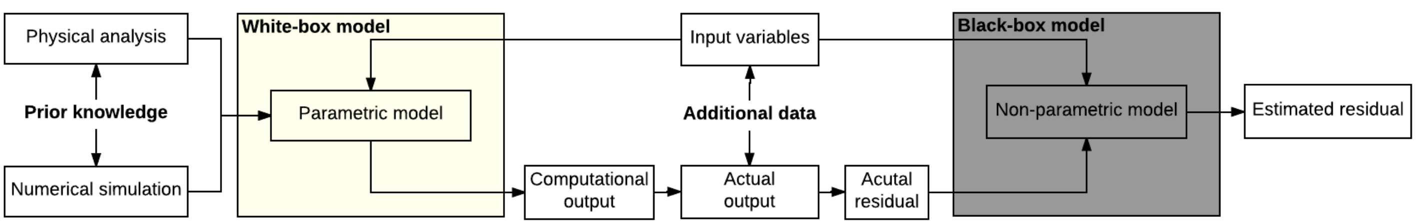

Figure 3. General workflow of the first stage

Figure 4 depicts the second stage of the process, wherein the white-box and black-box of the residual built in the first stage are composed to a parallel structure. It is considered a parallel structure because the given values of input variables are entered into white-box and black-box models simultaneously. The white-box in Figure 4 is the same as the one in Figure 3. However, at this grey-box modeling stage, the output from the white-box directly estimates the final solution in concert with the estimated residual from the black-box. For each new data point prediction, the output from the white-box is used as the basic solution. The residual solution from the black-box is the estimated residual for that same new data point. The final solution is the combination of basic solution with its estimated residual, or the results from both stages.

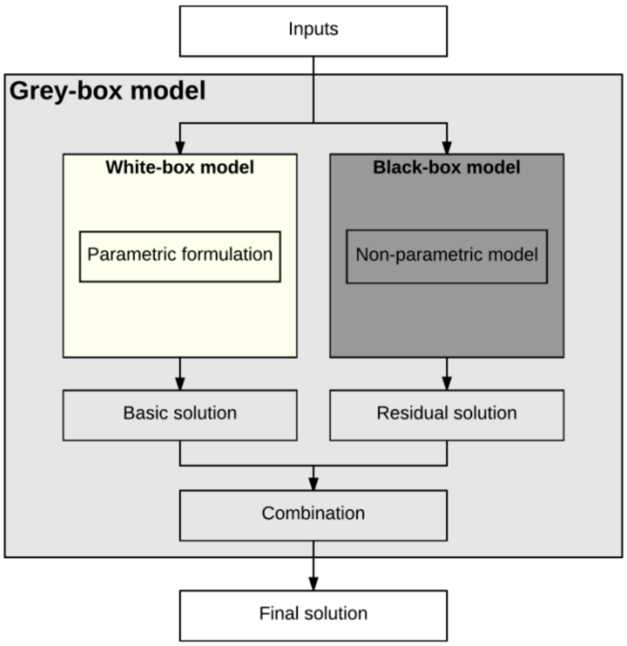

Figure 4. General workflow of the second stage

To illustrate the proposed grey-box modeling method, two case studies are presented in the next section. The first, a classical mathematical example of a mystery function [5, 30] demonstrates the process of constructing grey-box models from pre-existing knowledge that can be expressed numerically. The second example illustrates the use of this grey-box modeling technique to predict the relative density resulting from an AM process and represented by actual experimental data.

5. ILLUSTRATIVE CASE STUDIES

To illustrate the method introduced in the prior section, two case studies are presented in this section. The mystery function examples in Section 5.1 illustrate the process of grey-box model construction for different types of knowledge.

Maximum relative error magnitude (MREM) and average relative error magnitude (AREM) are used to represent the model predictability [7]. These two metrics are used to evaluate metamodel predictability because the combination of both metrics reveals both the overall predictability and the worst case of predictability of a metamodel in its design space. The formulation of MREM and AREM are:

$$MREM = \max\left(\frac{|y_i-\hat{y}_i|}{y_i}\right) \quad (y_i \neq 0) \tag{8}$$

$$AREM = \frac{1}{m}\left(\frac{\sum_{i=1}^{m}|y_i-\hat{y}_i|}{y_i}\right) \quad (y_i \neq 0) \tag{9}$$

where y_i is the observed value from given data, $\hat{y}$ is the value predicted by the metamodel of the data points that were not selected to construct the metamodel, and m is the number of data points.

5.1 Case studies: mystery function problem

A classical mystery function [5, 30] is brought into this study to mimic a complex unknown system. The function $f(x_1, x_2)$ that represents a nonlinear and complex system is used to generate experimental results used for model creation and assessment. The original equation of this mystery function is:

$$Y = f(x_1, x_2) = 2 + 0.01(x_2 - x_1^2)^2 + (1 - x_1) + 2(2 - x_2)^2 + 7\,sin(\tfrac{x_1}{2})\,sin(\tfrac{7x_1x_2}{10})$$

(10)

x_1 and x_2 are two input variables and Y is the actual output. The true surface and contour plots of the original mystery function are shown in Figure 5. To illustrate the effectiveness of this method in Section 5.1.1, the original equation is manipulated to illustrate a scenario similar to that of an inaccurate white-box model representing model construction with incomplete prior knowledge. In this situation, a parametric formulation is accessible before constructing the grey-box. However, Section 5.1.2 simulates a situation where the parametric white-box model cannot be directly derived from current knowledge. In that case, the prior knowledge was delivered by running a hypothetical simulation-based model, i.e. the manipulated function $f_k(x_1, x_2)$. These two examples illustrate how to use this grey-box modeling approach for different types of problems.

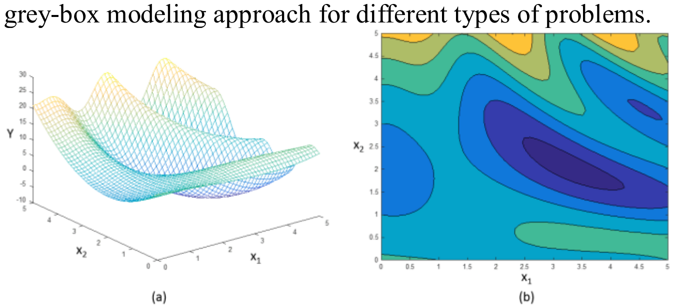

Figure 5. (a) True 3D surface plot and (b) contour plot of the original mystery function

5.1.1 Case study: theoretical physics-based formulation

In this example, the available prior knowledge is assumed derived from theoretical analysis and represented by an inaccurate parametric formulation. In this paper, it is assumed that the white-box models are reasonable representations of the manufacturing phenomena being modeled, and no validation step was included in our approach. Thus, to mimic this condition, the original mystery function is manually manipulated to represent a white-box model as:

$$\tilde{Y} = f_k(x_1, x_2) = 2 + 0.01(x_1 - x_1^2)^2 + (1 - x_1) + 2(2 - x_2)^2 + 7\,sin(\tfrac{x_1}{2})\,sin\left(\frac{5x_1x_2}{10}\right) - 0.4x_1\sin(2x_1)\cos(x_2) \tag{11}$$

where subscript k indicates a function derived from knowledge. $\tilde{Y}$ is computational output from the white-box model $f_k(x_1, x_2)$. Plots of the white-box model are shown Figure 6 (earlier stage of grey-box modeling). After the manipulation, the 3D surface maintains its general shape but several characteristics are changed, which can be observed in the figure. For example, the original local minima and maxima have shifted and the original sharp ridges became flatter. These changes result from the inaccurate white-box model. If we use the correct data from the original function to test current white-box model, the MREM and AREM are equal to 942.43 and 2.74, respectively. The large error indicates that the white-box model has very low fidelity and large predictive errors.

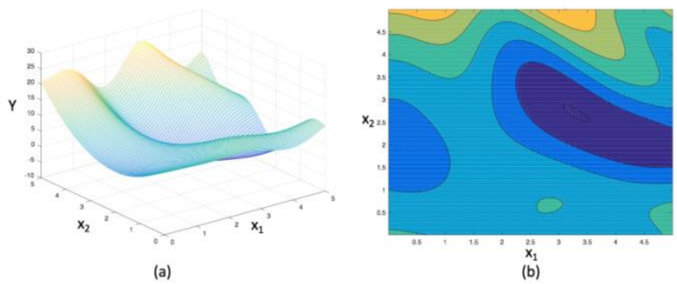

Figure 6. (a) 3D surface and (b) contour plots of manipulated mystery function

Though the white-box model has defects, it can contribute to a grey-box model. As mentioned in the last paragraph, the general information delivered from this white-box model is a reasonable representation of the level of knowledge to be expected from a white-box since the plots are generally similar to its original shape in that the local optima are still located close to their original positions. The next step is to add additional information to the initial, low fidelity prediction obtained from the white-box model. This high fidelity data is used alongside the low fidelity white-box prediction to build the black-box model. The additional information was generated from the original function using Latin Hypercube Sampling (LHS) [31] to generate 100 new data points. These additional data points represent an experimental or high fidelity model result as they were generated from the original function $f(x_1, x_2)$, which is defined as a high fidelity system without significant error.

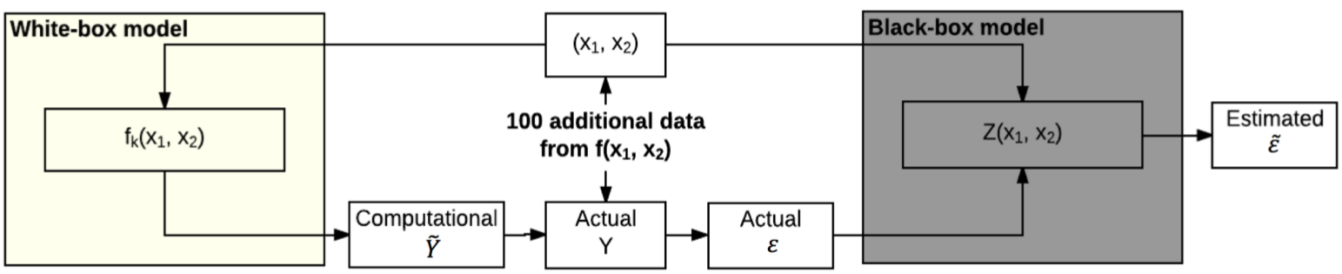

Figure 7. Black-box model construction to estimate residual

Figure 7 shows the process to construct the black-box model using the Type II serial approach (second stage of grey-box modeling). The input variables x_1 and x_2 are first entered into the white-box model $f_k(x_1, x_2)$ and used to calculate the computational output $\tilde{Y}$. The residual ε is the difference between $\tilde{Y}$ and actual output Y. The input variables (x_1, x_2) and the residual ε are next used to construct the black-box model $Z(x_1, x_2)$ using the Ordinary Kriging method. The black box model is used to compute the estimated error $\tilde{\varepsilon}$ in subsequent steps. Table 1 shows some examples from the 100 data points for illustration of the process shown in Figure 8. For example, one of the additional data points (3.73, 0.98) has an actual output 6.1881. This input when entered into the white-box model yields a prediction of 8.1416. The actual residual ε is next derived based on $\varepsilon = Y - \tilde{Y}$, which is equal to -1.9535. Once the black-box model is built from the residual values, it can estimate the residual of any unknown point from its input variable values. This estimated residual represents an expected difference between the white-box prediction and an unknown actual output. As a result, the final grey-box solution $\tilde{Y}_{final}$ should at any point be equal to $\tilde{Y} - \tilde{\varepsilon}$. This value combines the results from both stages, as shown in Figure 8. The white-box model in the parallel grey-box structure is the same one used in the prior serial approach. For any new point, the grey-box would combine basic solution $\tilde{Y}_{new}$ and the estimated residual $\tilde{\varepsilon}_{new}$ to get the final solution at that point location.

Table 1. Results at some sample data point locations

Input (x_1, x_2)	Actual output (y)	Computational output $(\tilde{Y})$	Actual residual (ε)
(3.73, 0.98)	6.1881	8.1416	-1.9535
(4.48, 4.03)	9.3751	11.9010	-2.5258
(0.23, 1.83)	3.0593	3.0066	0.0527
(4.33, 2.98)	5.0025	4.7074	0.2951

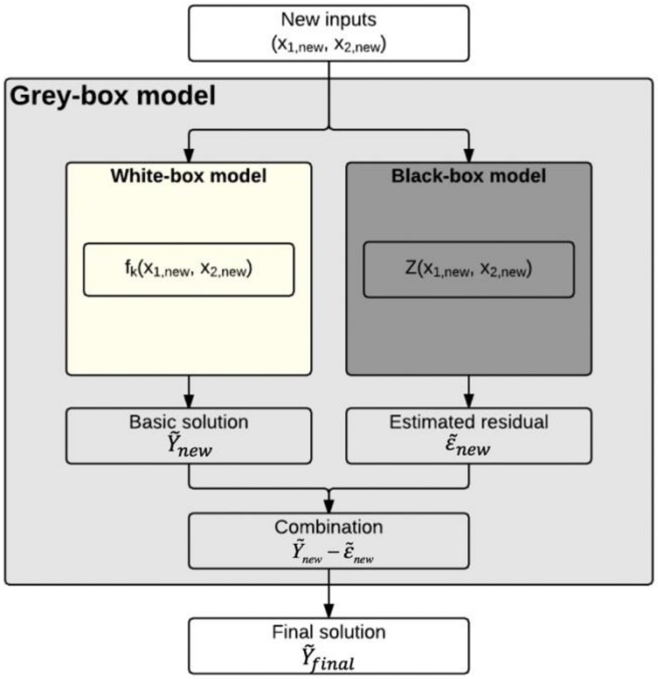

Figure 8. Grey-box model construction

One thousand randomly generated data points from the original mystery function were used to validate the resulting grey-box model. The grey-box model has reduced the initial white-box MREM from 942.24 to 2.7452 and AREM from 2.74 to 0.0359. The 3D surface and contour plots shown in Figure 9 are significantly improved and very close to the true plots of the original mystery function (Figure 5).

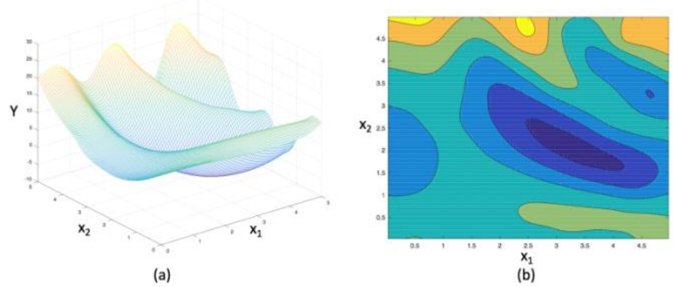

Figure 9. (a) 3D surface and (b) contour plots for grey-box model built based on 100 additional data points.

The grey-box constructed with 100 additional data points improved the initial white-box model globally. However, the MREM which represents the local error of the model remains higher than expected. It may be that the information provided by the additional dataset is insufficient. To further test the proposed method, Table 2 shows the MREM and AREM of grey-box models that are constructed with different numbers of additional data points. The top number of zero data points is a case where no data is available and the results are derived from the white-box model only. As shown, the model performance can decrease exponentially with more points in this process. Convergence criteria can be established to determine the desired accuracy.

Table 2. Model performance of additional quantity of data

Number of additional data	MREM	AREM
0	942.4327	2.7451
100	2.7452	0.0359
200	0.9372	0.0030
500	0.0019	0.0001

5.1.2. Case study: simulation-based knowledge

Many times, a theoretical physics-based parametric model is hard to access for complex problems. Simulation-based models have become more and more popular as basic reference points. Here, the initial knowledge-based parametric model f_k is assumed to be no longer available. Instead, a hypothetical simulation model replaces the former parametric white-box model. As a result, the function $f_k(x_1, x_2)$ cannot be used to generate the data needed to directly construct a black-box model and a subsequent grey-box model. Thus, a simplified white-box model is necessary since it is costly to run a high fidelity simulation for each point. To address this issue, a PR model was built to represent the white-box model. The manipulated function in Section 5.1.1 was assumed to be the simulation model. 1000 simulated data points were generated from function $f_k(x_1, x_2)$ and were used to create the PR model using LHS. The reason that the manipulated function f_k was employed instead of directly using the original mystery function is because the simulation-based model is also assumed to be low fidelity. The white-box model in PR form was generated as:

$$\tilde{Y}_{PR} = f_{PR}(x_1, x_2) = -0.3048 - 2.753x_1 - 0.228x_2 + 3.543x_1^2 + 1.973x_1x_2 + 8.582x_2 + 0.5552x_1^3 + 0.9604x_1^2x_2 + 0.9191x_1x_2^2 + 0.6433x_2^3 - 1.103x_1^4 - 0.7201x_1^3x_2 - 0.3864x_1^2x_2^2 - 0.5393x_1x_2^3 - 1.435x_2^4 \quad (12)$$

The R^2 value of this PR model is 0.7296. Comparing the PR white-box model to the original function yields an MREM of 846.6876 and an AREM of 1.9458. This indicates that this white-box model has poor predictability. This finding is reflected visually in the 3D surface and contour plots shown in Figure 10. In this figure, the shape is completely different from the original model (Figure 5). The ridges on the original surface disappeared. Thus, it is necessary to use the additional data points to build the grey-box model.

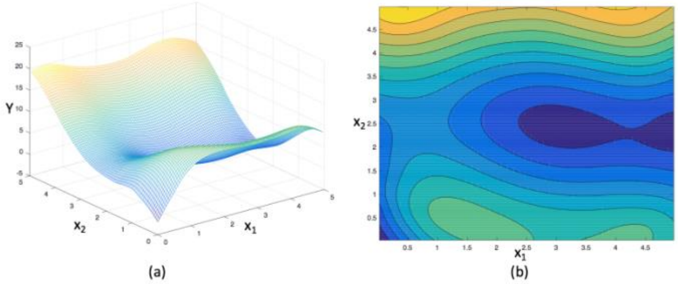

Figure 10. (a) 3D surface and (b) contour plots of the initial white-box model

The general updating process is similar to that shown in Section 5.1.1. The same 100 additional data points were used in this example compare the difference between results from Section 5.1.1 and Section 5.1.2. The Kriging black-box model was formulated by input variables and the corresponding residual values using the same procedure that was described in Section 2.2. The grey-box model was then developed by combining the PR and Kriging models by the process shown in Figure 8. The same validation process was executed to evaluate the model performance with the same validation dataset that was used in Section 5.1.1. The final MREM and AREM of this grey-box model are 3.4318 and 0.0506, which is slightly higher than using physics-based white-box model with same amount of additional data. The plots for this grey-box model are shown in Figure 2. The results from using additional data points are listed in Table 3.

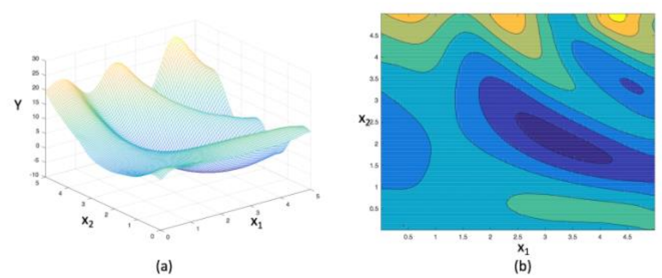

Figure 11. (a) 3D surface and (b) contour plots of the grey-box model built by simulation-based knowledge and additional actual data.

Table 3. Model performance of additional data

Number of additional data	MREM	AREM
0	846.6876	1.9458
100	3.4318	0.0506
200	1.5988	0.0055
500	0.0052	0.0001

5.2 Case study: metal PBF problem

This example uses the proposed method to build a grey-box model for a realistic PBF problem. Louvis and associates' experiments with a PBF process measured the relative density produced by different scan speed (v) and hatch spacing (d) for different aluminum alloy powders [32]. Relative density is the ratio of the actual part density to that of a completely filled solid with no porosity. The experimental results indicate higher relative density is generally produced by lower scan speed and closer hatch spacing. However, the relative density has unique behavior for specific powders and machines. For example, the density of AlSi12 and 6061 aluminum powder produced by the same parameters in different machines have different results [32], which indicates the a model built based on 6061's data may not be accurate for AlSi12. Instead of building an expensive new model, this study uses the findings from 6061's data to construct a grey-box model for AlSi12 for illustrative purposes.

In this case, there is no available physics-based knowledge to build the white-box model since the only available prior knowledge is from historical experiments. Therefore, the prior experimental knowledge of 6061 powder [32] was used to build a PR model to serve as the white-box model for illustrative purposes just as was done in Section 5.1.2. The reported measurements from AlSi12 powder were used as the additional information. First, the 177 data points from the 6061 powder experiment were used to build the PR based white-box model. The resulting quadratic model was generated as:

$$\tilde{Y}_{PR} = f_{PR}(v,d) = 81.54 + 116.83v - 0.0127d - 0.079vd - 332.39v^2 + 7082600d^2 \quad (13)$$

The initial R^2 value of this model is 0.953. The set of 36 data points from the AlSi12 PBF experiment was divided into two sets. 80% (29 points) of the data was extracted from the initial dataset to use as additional information for grey-box construction. The remaining 20% (7 points) of the data set was used to validate the models. Table 4 lists the MREM and AREM for different types of models based on the data. The 7 validation data points from the AlSi12 experiment were entered into all three types of models to evaluate and compare the predictive accuracy. The pure white-box is the PR model built using the 6061 powder experiment. The pure black-box model represents the model built with the 29 AlSi12 data points with kriging method. The grey-box represents the model built with input from both experiments using the same method presented in the prior sections. As shown, even though the pure white-box model has low predictive error, the model can be further improved by the grey-box modeling approach. The MREM of the original model is reduced from 0.0375 to 0.0238, which is a 37% improvement. Compared to the pure black-box model, the MREM of the grey-box model reduced from 0.0485 to 0.0238, which is a 51% improvement after the completion of both modeling stages.

Table 4. The performance of different types of models

	MREM	AREM
Pure white-box	0.0375	0.0170
Pure black-box	0.0485	0.0169
Grey-box	0.0238	0.0134

6. DISCUSSION

Effectively deploying predictive analytics in smart manufacturing is a challenge that many now face. This challenge is highlighted in AM, where current AM models often lack comprehensive information, and where information could be either knowledge-based or statistically generated. The lack of the former is typically the result of an incomplete understanding of the physical processes of AM. The lack of empirical data, on the other hand, may be caused by the difficulty of instrumenting AM processes, and more generally, the expense of producing AM parts. Even as more and more empirical data is available in the coming years, it is still difficult to duplicate all the conditions and the model predictability for all data sets. The experimental results can exhibit noticeable differences even where the experiments are operated in similar AM processes with comparable process parameters. It is also very difficult to build the connection between simulations and actual experimental data. These difficulties result from the uncertainties in and complexities of AM processes. The uncertainties significantly reduce the utility of AM predictive models since a well-validated model from one dataset may be difficult to apply to other experimental conditions.

The highlight of the two-stage grey-box modeling approach developed in this paper is that it can combine disparate knowledge and information together to produce an accurate hybrid model. To further extract the information from limited knowledge, this two-stage grey-box structure can functionally improve the predictive accuracy. In the example in Section 5.1, the original white-box model produced a large global predicative error, with an AREM over 200%. However, when updated with 100 additional data points, the grey-box approach reduced that same global error to less than 5%. These results suggest that the grey-box approach can improve the model even if the original knowledge-based model is very weak. Moreover, the results were derived from a highly nonlinear mathematical example which is even more complex than common AM problems. The 100 additional data points can sufficiently update the initial white-box model to a higher accuracy. However, a smaller sample size is expected in actual AM experiments. It is thus desirable to further reduce that sample size needed to achieve the higher model predictability.

The involved metamodeling algorithms were used as primary candidates to build the grey-box in this paper. However, the general modeling process should have no bias to other black-box modeling techniques. Any suitable algorithm that can improve the model predictability might be introduced in future work.

In grey-box modeling, the reusability of a model built on prior knowledge can be improved when combined with additional information about problem specific conditions. In Section 5.3, the case study investigated the performance of a grey-box model with two different PBF datasets generated with similar experimental conditions. The two experiments were completed in different PBF machines, with different metal powders, and unknown experimental conditions such as chamber temperature and layer thickness. Both the pure white-box and black-box models resulted in a higher MREM than the combined grey-box model when predicting the other dataset. The grey-box model reduces the MREM by about 37% compared to the white-box model and 51% compared to the black-box model. Thus, the grey-box approach provides a reliable way to reuse the knowledge from one manufacturing case to another.

ACKNOWLEGEMENT

This material is based upon work supported by the National Science Foundation (NSF) under Grant No. 1439683, the National Institute of Standards and Technology (NIST) under Cooperative Agreement number NIST 70NANB15H320, and industry members of the NSF Center for e-Design.

REFERENCES

[1] Lu, Y., Witherell, P., Lopez, F., 2016, "Digital Solutions for Integrated and Collaborative Additive Manufacturing," ASME 2016 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference, American Society of Mechanical Engineers, pp. V01BT02A033-V01BT02A033.

[2] Bauchau, O., and Craig, J., 2009, "Structural Analysis,"Springer, pp. 173-221.

[3] Devesse, W., De Baere, D., and Guillaume, P., 2014, "The Isotherm Migration Method in Spherical Coordinates with a Moving Heat Source," International Journal of Heat and Mass Transfer, 75pp. 726-735.

[4] Frazier, W. E., 2014, "Metal Additive Manufacturing: A Review," Journal of Materials Engineering and Performance, 23(6) pp. 1917-1928.

[5] Shao, T., 2007. Toward a structured approach to simulation-based engineering design under uncertainty. University of Massachusetts Amherst.

[6] Shao, T., and Krishnamurty, S., 2008, "A Clustering-Based Surrogate Model Updating Approach to Simulation-Based Engineering Design," Journal of Mechanical Design, 130(4) pp. 041101.

[7] Yang, Z., Eddy, D., Krishnamurty, S., 2016, "Investigating Predictive Metamodeling for Additive Manufacturing," ASME 2016 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference, American Society of Mechanical Engineers, pp. V01AT02A020-V01AT02A020.

[8] Cui, C., 2016. Building Energy Modeling: A Data-Driven Approach (Doctoral dissertation, ARIZONA STATE UNIVERSITY).

[9] Zhao, D., and Xue, D., 2011, "A Multi-Surrogate Approximation Method for Metamodeling," Engineering with Computers, 27(2) pp. 139-153.

[10] Shanmuganathan, S., and Samarasinghe, S., 2016, "Artificial neural network modelling," Springer, .

[11] Kristensen, N. R., Madsen, H., and Jørgensen, S. B., 2004, "A Method for Systematic Improvement of Stochastic Grey-Box Models," Computers & Chemical Engineering, 28(8) pp. 1431-1449.

[12] Jin, R., Chen, W., and Sudjianto, A., 2004, "Analytical Metamodel-Based Global Sensitivity Analysis and Uncertainty Propagation for Robust Design," Sae Sp, 14(429) pp. 47-54.

[13] Simpson, T., Mistree, F., Korte, J. and Mauery, T., 1998, September. Comparison of response surface and kriging models for multidisciplinary design optimization. In 7th AIAA/USAF/NASA/ISSMO Symposium on Multidisciplinary Analysis and Optimization (p. 4755).

[14] Simpson, T. W., Booker, A. J., Ghosh, D., 2004, "Approximation Methods in Multidisciplinary Analysis and Optimization: A Panel Discussion," Structural and Multidisciplinary Optimization, 27(5) pp. 302-313.

[15] Shan, S., and Wang, G. G., 2010, "Survey of Modeling and Optimization Strategies to Solve High-Dimensional Design Problems with Computationally-Expensive Black-Box Functions," Structural and Multidisciplinary Optimization, 41(2) pp. 219-241.

[16] Cressie, N., 2015, "Statistics for spatial data," John Wiley & Sons, .

[17] Sacks, J., Welch, W. J., Mitchell, T. J., 1989, "Design and Analysis of Computer Experiments," Statistical Science, pp. 409-423.
[18] Bohlin, T.P., 2006, "Practical grey-box process identification: theory and applications," Springer Science & Business Media, .
[19] Psichogios, D. C., and Ungar, L. H., 1992, "A Hybrid Neural Network-first Principles Approach to Process Modeling," AIChE Journal, 38(10) pp. 1499-1511.
[20] Schubert, J., Simutis, R., Dors, M., 1994, "Hybrid Modelling of Yeast Production processes–Combination of a Priori Knowledge on Different Levels of Sophistication," Chemical Engineering & Technology, 17(1) pp. 10-20.
[21] Duarte, B. P., and Saraiva, P. M., 2003, "Hybrid Models Combining Mechanistic Models with Adaptive Regression Splines and Local Stepwise Regression," Industrial & Engineering Chemistry Research, 42(1) pp. 99-107.
[22] Thibault, J., Van Breusegem, V., and Chéruy, A., 1990, "On-line Prediction of Fermentation Variables using Neural Networks," Biotechnology and Bioengineering, 36(10) pp. 1041-1048.
[23] Gibson, I., Rosen, D.W., and Stucker, B., 2010, "Additive manufacturing technologies," Springer, .
[24] Melchels, F. P., Domingos, M. A., Klein, T. J., 2012, "Additive Manufacturing of Tissues and Organs," Progress in Polymer Science, 37(8) pp. 1079-1104.
[25] Bourell, D. L., Beaman, J., Leu, M. C., 2009, "A Brief History of Additive Manufacturing and the 2009 Roadmap for Additive Manufacturing: Looking Back and Looking Ahead," Proceedings of RapidTech, pp. 24-25.
[26] Murr, L. E., Gaytan, S. M., Ramirez, D. A., 2012, "Metal Fabrication by Additive Manufacturing using Laser and Electron Beam Melting Technologies," Journal of Materials Science & Technology, 28(1) pp. 1-14.
[27] Witherell, P., Feng, S., Simpson, T. W., 2014, "Toward Metamodels for Composable and Reusable Additive Manufacturing Process Models," Journal of Manufacturing Science and Engineering, 136(6) pp. 061025.
[28] Agarwala, M., Bourell, D., Beaman, J., 1995, "Direct Selective Laser Sintering of Metals," Rapid Prototyping Journal, 1(1) pp. 26-36.
[29] Cressie, N., 1993, "Statistics for Spatial Data: Wiley Series in Probability and Statistics," Wiley-Interscience, New York, 15pp. 105-209.
[30] Martin, J. D., 2009, "Computational Improvements to Estimating Kriging Metamodel Parameters," Journal of Mechanical Design, 131(8) pp. 084501.
[31] McKay, M. D., Beckman, R. J., and Conover, W. J., 2000, "A Comparison of Three Methods for Selecting Values of Input Variables in the Analysis of Output from a Computer Code," Technometrics, 42(1) pp. 55-61.
[32] Louvis, E., Fox, P., and Sutcliffe, C. J., 2011, "Selective Laser Melting of Aluminium Components," Journal of Materials Processing Technology, 211(2) pp. 275-284.