Title	Physics-Aware Machine Learning for Construction of Surrogate Models with Small Training Data
Author(s)	五十嵐, 一; Igarashi, Hajime
Citation	IEEE Transactions on Magnetics, 1-7 https://doi.org/10.1109/tmag.2025.3622416
Issue Date	2025-10-16
Doc URL	https://hdl.handle.net/2115/99254
Rights	© 2025 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works.
Type	journal article
File Information	igarashi-final.pdf

Physics-Aware Machine Learning for Construction of Surrogate Models with Small Training Data

Hajime Igarashi, *Member, IEEE*

Data-Driven Interdisciplinary Research Emergence Department: D-RED
Hokkaido University, Kita 13, Nish 10, Kita-ku, Sapporo 060-0813, Japan

This paper shows that physics-aware machine learning can construct accurate surrogate models for optimal design, uncertainty quantification, and data assimilation, even when limited training data is available. In this method, the surrogate model is trained to predict the ratio of the ground truth—provided by, e.g., field computations or measurements—to the prediction of a physics-based model. In this multiplicative correction, the surrogate model provides the compensation factor to the prediction of the physics-based model, for example, an analytical or empirical formula. After training, accurate predictions can be obtained using the physics-based model and trained machine learning system, both of which have a light computing burden. Additive correction is also tested, in which the difference between the ground truth and the prediction of a physics-based model is used rather than the ratio. Numerical experiments reveal that multiplicative correction based on the ratio is more accurate than the additive correction based on the difference. It is shown that the response surface of the ratio tends to be so flat that it can be represented by the small training data, while this is not the case for the difference-based method and direct prediction. The present method provides an explicit approximation formula for the ratio based on linear regression, allowing physical interpretation. The potential applications of this method, including data assimilation, are suggested.

Index Terms—**Physics-aware machine learning, surrogate model, kernel regression, data assimilation, wireless power transfer.**

I. Introduction

OPTIMAL DESIGN is widely performed in developing electrical and electronic devices, such as motors, transformers, inductors, and wireless power transfer (WPT) devices, to determine their shape, configuration, and material [1, 2]. Uncertainty quantification is also widely performed to evaluate how uncertain factors, such as changes in material properties over time and property variability, and manufacturing errors affect device characteristics. They involve computations with, for example, the finite element method (FEM) and circuit analysis, which require a large amount of computing resources, including computing time and the number of computing cores. An effective solution to this problem is replacing computationally intensive methods with less intensive surrogate models. Supervised learning methods, such as neural networks (NNs), kernel regression, and regression trees, are used to construct the surrogate models. However, supervised learning usually requires large training data, which again entails significant computational costs. To effectively use surrogate models, the total computational cost of constructing the models and performing optimization or uncertainty quantification must be significantly lower than the cost of conventional methods that do not use surrogate models. In this work, we incorporate physics into machine learning to reduce the computational cost of building surrogate models. This approach is here referred to as physics-aware machine learning.

A machine learning method based on physics called the Hamiltonian NN has been proposed. It uses the prior knowledge that the system obeys Hamiltonian mechanics [3]. This ensures that the prediction results of the NN satisfy the law of conservation of energy. Moreover, a physics-aware method has been proposed in which the prior predictions from a physics-based model along with the geometric parameters are input to a NN to predict device characteristics [4, 5]. Furthermore, knowledge-based NNs [6] and prior-knowledge incorporated NN [7] have been proposed to improve prediction accuracy by reflecting prior information to the NN.

There is another stream in the research activities aiming at effective use of the surrogate models, which is the multi-fidelity modeling [8]. This approach performs a fast computation using a low-fidelity model (LFM) for optimization, uncertainty quantification and inference, which is combined with a computationally intensive high-fidelity model (HFM) constructed by, e.g., FEM. The LFM is constructed using one of the following methods: simplifying HFM, fitting data, or creating a projection-based model [9]. The LFM is combined to HFM by either of multiplicative correction, additive correction and space mapping [10]. In the multiplicative correction, the ratio of the result computed by the LFM to that by the HFM is used for correction. In optimization, for example, the objective function is evaluated by the LFM and corrected by the precomputed correction ratio. On the other hand, in the additive correction, the difference between the LFM and HFM is used for correction. In [11], the additive and multiplicative corrections were compared for the optimization of honeycomb-type cellular material where FE models with coarse and fine meshes are used for the LFM and HFM, respectively, concluding that the multiplicative correction gives better results. In computational magnetics, an electromagnetic relay was optimized using the surrogate model with multiplicative correction based on the magnetic circuit and FEM, which correspond to the LFM and HFM, respectively [12]. However, it remained unclear how they would perform with limited learning data and why they would perform well if they did.

Manuscript received April 1, 2015; revised May 15, 2015 and June 1, 2015; accepted July 1, 2015. Date of publication July 10, 2015; date of current version July 31, 2015.. Corresponding author: H. Igarashi (e-mail: igarashi.hokudai@gmail.com).

Digital Object Identifier (inserted by IEEE).

The present work introduces physics-based models for the additive and multiplicative corrections. There are overlaps between the physics-aware machine learning and multi-fidelity models, while the former contains use of theoretical or empirical formula, and use of experimental data for data assimilation, and also addresses the prediction of characteristics of the modified model that have the same fidelity as the original model. They will be discussed in section IV.

The following will be shown in this paper : (i) the accurate surrogate model can be constructed even with small training data with multiplicative correction, (ii) the multiplicative method gives flat response surfaces, which is shown to be important to have an accurate prediction by the surrogate model constructed with small data, (iii) it is possible to have an explicit prediction formula by the multiplicative correction due to the nature in (ii), (iv) a light surrogate model can be constructed with an analytical formula for the mutual inductance of a WPT system, (v) the magnetic saturation gives a significant impact on the prediction accuracy of the surrogate model. The novelty of this study lies in the points (i) through (v) mentioned above.

This paper is organized as follows: the next section formulates the kernel ridge regression which is used to construct surrogate models with small training data. The third section reports numerical results for WPT and electromagnet models. The potential applications of the present method are suggested in the fourth section. The last section summarizes conclusions of this work.

II. Formulation

A. Linear and kernel ridge regression

Although we mainly use the kernel ridge regression in this work, we begin with the least square method for linear multiple regression because it will also be used in the next section. We assume that we have data $\mathcal{D} = \{(\boldsymbol{x}_i, y(\boldsymbol{x}_i))\}$, $i = 1,2,\dots,m$, $\boldsymbol{x}_i \in \mathbb{R}^n$, $y \in \mathbb{R}$ and $m \geq n+1$. We introduce the data matrix

$$\mathrm{X} = \begin{bmatrix} 1 & \boldsymbol{x}_1^t \\ \vdots & \vdots \\ 1 & \boldsymbol{x}_m^t \end{bmatrix} \tag{1}$$

and a vector $\boldsymbol{y} = (y(\boldsymbol{x}_1), y(\boldsymbol{x}_2), \dots, y(\boldsymbol{x}_m))^t$. Now we consider minimizing the sum of the residual norm and a regularization term with respect to the coefficient vector $\boldsymbol{w}$ as follows:

$$\frac{1}{2}(\boldsymbol{y} - \mathrm{X}\boldsymbol{w})^t(\boldsymbol{y} - \mathrm{X}\boldsymbol{w}) + \frac{\lambda}{2}\boldsymbol{w}^t\boldsymbol{w} \rightarrow \text{min.} \tag{2}$$

where λ is a regularization constant. From (2) we obtain

$$(\mathrm{X}^t\mathrm{X} + \lambda\mathrm{I})\boldsymbol{w} = \mathrm{X}^t\boldsymbol{y} \tag{3}$$

which is the regularized normal equation. Then, we decompose $\boldsymbol{w}$ in such a way that $\boldsymbol{w} = \boldsymbol{w}_1 + \boldsymbol{w}_2$, where $\boldsymbol{w}_1 \in \mathrm{Ker}(\mathrm{X})^\perp$ and $\boldsymbol{w}_2 \in \mathrm{Ker}(\mathrm{X})$. We substitute this expression into (2) to find that $\boldsymbol{w}_2$ must be zero due to the regularization term. Because it can be shown that $\boldsymbol{w}_1 = \mathrm{X}^t\boldsymbol{a}$, we have so called the dual form to (3) [13] as follows:

$$(\mathrm{X}\mathrm{X}^t\mathrm{X}\mathrm{X}^t + \lambda\mathrm{X}\mathrm{X}^t)\boldsymbol{a} = \mathrm{X}\mathrm{X}^t\boldsymbol{y} \tag{4}$$

Now assuming that $m = n+1$ and $\mathrm{X}\mathrm{X}^t$ is regular in (4), we obtain the equation for the kernel ridge regression given by

$$(\mathrm{X}\mathrm{X}^t + \lambda\mathrm{I})\boldsymbol{a} = \boldsymbol{y} \tag{5}$$

where $\mathrm{X}\mathrm{X}^t$ is the Gram matrix whose entities are

$$k_{ij} = 1 + \boldsymbol{x}_i^t\boldsymbol{x}_j \tag{6}$$

It is concluded from the above discussion that the multiple linear regression (3) is equivalent to the kernel regression with the linear kernel (6) when $m = n+1$. To extend (6) for nonlinear regression, we replace (6) with an inner product $k_{ij} = \boldsymbol{\Phi}(\boldsymbol{x}_i)^t\boldsymbol{\Phi}(\boldsymbol{x}_j)$ for a nonlinear function $\boldsymbol{\Phi}(\boldsymbol{x})$. Furthermore, it is shown that this inner product can be represented by the kernel function $k_{ij} = k(\boldsymbol{x}_i, \boldsymbol{x}_j)$ that is symmetric, $k(\boldsymbol{x}_i, \boldsymbol{x}_j) = k(\boldsymbol{x}_j, \boldsymbol{x}_i)$, and semi-positive definite, that is, for any $c_i, c_j \in \mathbb{R}$

$$\sum_i \sum_j c_i c_i k(\boldsymbol{x}_i, \boldsymbol{x}_j) \geq 0 \tag{7}$$

holds [13]. After solving (5), we have a response for a new given point $\boldsymbol{x}$

$$\hat{y}(\boldsymbol{x}) \approx \sum_i k(\boldsymbol{x}, \boldsymbol{x}_i)a_i \tag{8}$$

which forms a response surface. In this work, we use the Gaussian kernel given by

$$k(\boldsymbol{x}, \boldsymbol{x}_i) = \exp\left(-\frac{|\boldsymbol{x} - \boldsymbol{x}_i|^2}{\theta}\right) \tag{9}$$

B. Additive and multiplicative corrections

We can use the kernel ridge regression formulated above to have the predicted value, $\hat{y}(\boldsymbol{x})$, based on the training data $\mathcal{D}$. This method will be referred to as the direct method. However, as mentioned in Introduction, the direct method would be computationally expensive to prepare for sufficient training data $\mathcal{D}$ to construct a surrogate model for optimization or uncertainty quantification. To overcome this difficulty, we incorporate physics into machine learning to build a surrogate model. To do so, we prepare for two data $\mathcal{D}_p$ and $\mathcal{D}$ composed of $y_p(\boldsymbol{x}_i)$ and $y(\boldsymbol{x}_i)$, $i = 1,2,\dots,m$, which are computed by a physics-based model with a small computing cost and accurate model based on, e.g., FEM, respectively. In the multiplicative correction [8, 9], we introduce the ratio defined by

$$r(\boldsymbol{x}) = \frac{y(\boldsymbol{x})}{y_p(\boldsymbol{x})} \tag{10a}$$

and construct the training data $\mathcal{D}_r = \{(\boldsymbol{x}_i, r(\boldsymbol{x}_i))\}$, $i = 1,2,\dots,m$. Then based on $\mathcal{D}_r$, the kernel ridge regression (8) is performed to have the prediction $\hat{r}(\boldsymbol{x})$ instead of $\hat{y}(\boldsymbol{x})$. Now the predicted response for a new input $\boldsymbol{x}$ is given by

$$\hat{y}(\boldsymbol{x}) = y_p(\boldsymbol{x})\hat{r}(\boldsymbol{x}) \tag{10b}$$

It will be shown in the next section that the training for prediction of $r(\boldsymbol{x})$ needs much smaller data in comparison with the direct method. Because (10b) needs a small computing cost, the total computing cost can be much reduced.

Similarly, the additive correction introduces the difference

$$d(\boldsymbol{x}) = y(\boldsymbol{x}) - y_p(\boldsymbol{x}) \tag{11a}$$

and constructs the training data $\mathcal{D}_d = \{(\boldsymbol{x}_i, d(\boldsymbol{x}_i))\}$. Then we can have prediction $\hat{d}(\boldsymbol{x})$ and predicted response given by

$$\hat{y}(\boldsymbol{x}) = y_p(\boldsymbol{x}) + \hat{d}(\boldsymbol{x}) \tag{11b}$$

In this work, we use the FEMM code [14] to compute $y(\boldsymbol{x})$.

III. NUMERICAL RESULTS

A. WPT model

Let us first consider the axisymmetric WPT model shown in Fig.1 which includes the parameters $\boldsymbol{x} \in \mathbb{R}^8$ [15]. There are magnetic layers near the coils for shielding and increase in the magnetic coupling. We build a surrogate model for the mutual inductance $M(\boldsymbol{x})$ as well as the ratio $r(\boldsymbol{x}) = M(\boldsymbol{x})/M_p(\boldsymbol{x})$ and difference $d(\boldsymbol{x}) = M(\boldsymbol{x}) - M_p(\boldsymbol{x})$ of this device using the kernel ridge regression. For the physics-based model, we use the analytical formula for M given by

$$M_p(\boldsymbol{x}) = \sum_{i=1}^{n_1}\sum_{j=1}^{n_2} M_{ij} \tag{12a}$$

$$M_{ij} = \frac{2\mu_0\sqrt{r_i r_j}}{k}\left[\left(1-\frac{k^2}{2}\right)K(k) - E(k)\right] \tag{12b}$$

$$k = \sqrt{\frac{4r_i r_j}{\left(r_i + r_j\right)^2 + d^2}} \tag{12c}$$

where K and E denotes the complete elliptic integral of the first and second kind, r_i, r_j are the coil radii and d is the distance between the coil pair. Note that Eq.(12) does not account for the existence of magnetic layers and coil volume, but it does include the essential properties that M is proportional to the number of windings n_1, n_2 and increases with coil distance, d. The error in (12) is corrected by the machine leaning.

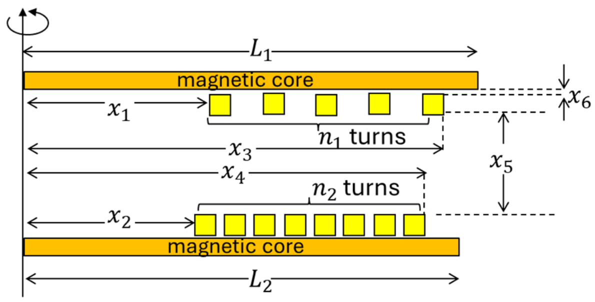
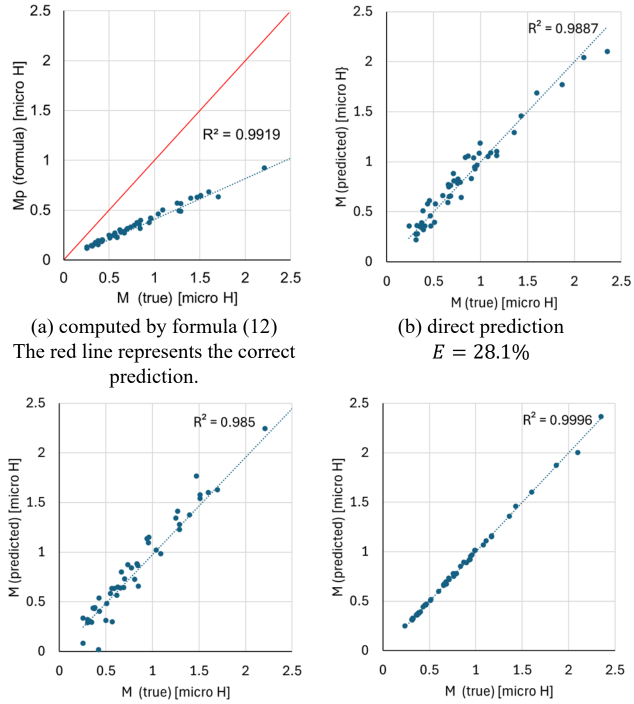


Fig.1 Axisymmetric WPT model
$\boldsymbol{x} = (x_1, x_2, \dots x_6, n_1, n_2)^t$



(a) computed by formula (12)
The red line represents the correct prediction.

(b) direct prediction
$E = 28.1\%$

(c) additive correction (11)
$E = 22.0\ \%$

(d) multiplicative correction (10)
$E = 1.88\ \%$

Fig.2 Correlation between the M-values computed by FEM (truth) and M-values obtained by formula (12) and the kernel ridge regression for $m = 50$. R^2 stands for the coefficient of determination.

Table I
Specification for WPT model [15]
$L_1 = x_3 + 4, L_2 = x_4 + 4$, geometric unit: mm
Relative permeability of the magnetic cores: $\mu_r = 1000$

parameters	range	parameters	range
x_1	$(x_3/3, x_3/2)$	x_5	$(4.0, 16.0)$
x_2	$(x_4/3, x_4/2)$	x_6	$(0.2, 1.0)$
x_3	$(30.0\ 50.0)$	n_1	$\{4,5, \dots, 9\}$
x_4	$(30.0\ 50.0)$	n_2	$\{4,5, \dots, 9\}$

The specification of the WPT model is summarized in Table I. When building $\mathcal{D}$, $\mathcal{D}_r$ and $\mathcal{D}_d$, the parameters are changed to follow a uniform distribution over the range shown in Table I. Before the regression, all the data are standardized by $p_i' = (p_i - \mu_i)/\sigma_i$ where μ_i, σ_i are the mean and standard deviation.

To evaluate the prediction error, we use the 5-fold cross validation. For example, when $m = 50$, we select 40 data for training and use 10 data for validation and this evaluation is performed five times with different combinations of data. The prediction error is characterized by the root mean square percent error (RMSPE) defined by

$$E = 100\sqrt{\sum_{i=1}^{m}\left(\frac{y_i - \hat{y}_i}{y_i}\right)^2} \tag{13}$$

Fig.2 shows the correlation between the M-values computed by FEM, which is regarded as the ground truth, and M-values obtained by formula (12) and the kernel ridge regressions for $m = 50$. Note that the amount of training data, 40 , is very small because we need 2^8 data even for two-level design. Because the effect of the magnetic layer is neglected in (12), we observe large error in Fig.2 (a) while the correlation is rather high. It can be seen in (b) and (c) that the predictions by the direct method and additive corrections have errors greater than 20% and the correlation is lower than that in (a). On the other hand, the multiplicative correction has error smaller than 2% as shown in (d). Fig.3 plots the convergence of the prediction errors against the test data size, showing the excellent accuracy of the multiplicative correction over the whole range.

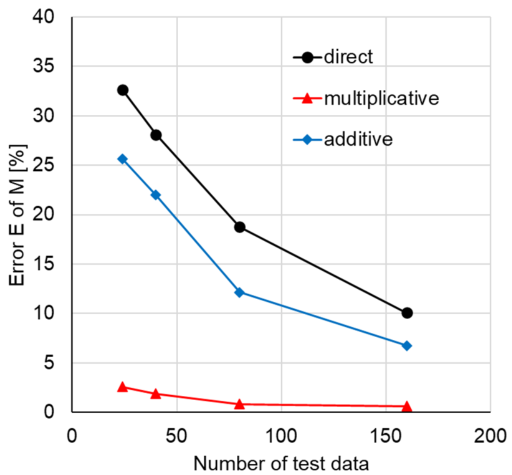
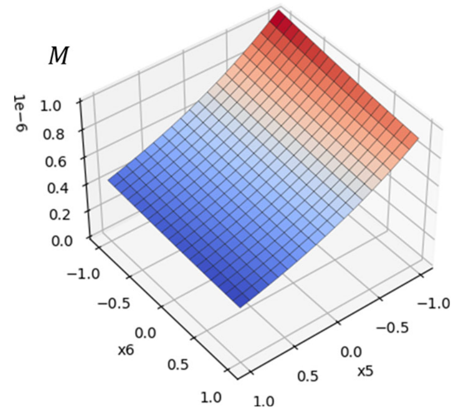


Fig. 3 Convergence of prediction errors

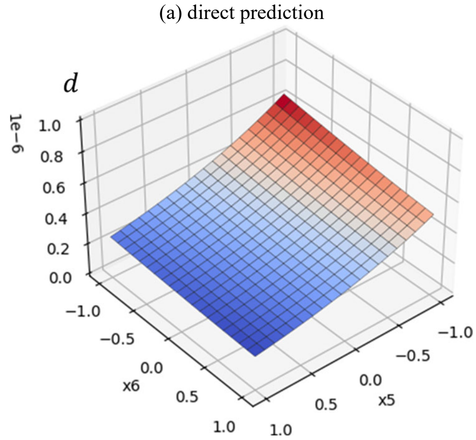
B. Discussions on results of WPT model

To find the reason why there are significant differences in the prediction error of the three methods, the response surfaces for x_5, x_6 are plotted in Fig.4. It can be found that the response surface for the multiplicative correction is rather flat while they are curved in the direct method and additive correction. Due to the limited amount of training data, the curved surface cannot be accurately approximated using kernel ridge regression or other machine learning methods. This suggests that the high accuracy of the multiplicative method is due to its flat response surface. To test this hypothesis, linear regression with a linear kernel (6) is performed on the three methods. The results are summarized in Table II. It can be seen that there is little difference in the errors in the regressions with the Gaussian and linear kernels for the multiplicative correction. This means that the response surface is flat for this method. On the other hand, there are significant differences in the two other methods. This is due to the fact that their response surfaces are curved, and therefore, it is difficult to make accurate prediction with small training data.

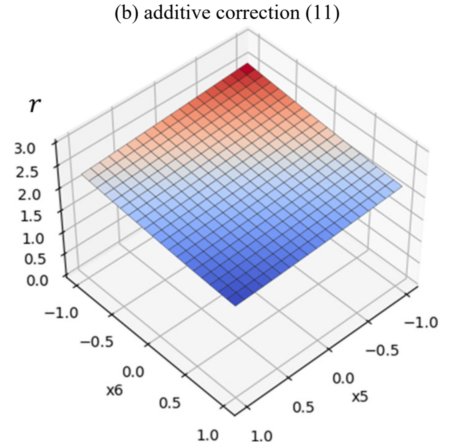
Since $y(\boldsymbol{x})$ is positively correlated with $y_p(\boldsymbol{x})$, we expect both $r(\boldsymbol{x})$ and $d(\boldsymbol{x})$ to be flat. However, as mentioned above, $d(\boldsymbol{x})$ is actually curved in contrast to $r(\boldsymbol{x})$. To intuitively understand this difference, we assume that the response surfaces for $y(\boldsymbol{x})$ and $y_p(\boldsymbol{x})$ are represented by an M-degree polynomial with two variables as follows:



(a) direct prediction



(b) additive correction (11)



(c) multiplicative correction (10)

Fig. 4 Response surfaces

Table II

Error E[%] of regression with Gaussian and linear kernels

kernel	direct	multiplicative	additive
Gauss (9)	28.1	1.88	22.0
Linear (6)	52.8	2.13	34.1

$$y(\boldsymbol{x}) = \sum_{i=0}^{M}\sum_{j=0}^{M} c_{ij} x^i x^j \qquad (14a)$$

$$y_p(\boldsymbol{x}) = \sum_{i=0}^{M}\sum_{j=0}^{M}(c_{ij} \mp \delta_{ij})x^i x^j \tag{14b}$$

where $|\delta_{ij}| \ll |c_{ij}|$. The ratio and difference are then given by

$$r(\boldsymbol{x}) \approx 1 \pm \sum_{i=0}^{M}\sum_{j=0}^{M}\delta_{ij}x^i x^j \Big/ \sum_{i=0}^{M}\sum_{j=0}^{M}c_{ij}x^i x^j \tag{15a}$$

$$d(\boldsymbol{x}) = \pm \sum_{i=0}^{M}\sum_{j=0}^{M}\delta_{ij}x^i x^j \tag{15b}$$

Hence, the first constant term is dominant in $r(\boldsymbol{x})$ while $d(\boldsymbol{x})$ is still an M degree polynomial. Theoretically explaining these properties in a more general case remained an open problem.

As shown above, the response surface for the multiplicative correction is nearly flat provided that its prediction accuracy is high for the small training data. Under this situation, the response surface is approximately represented by a hyperplane using the linear regression. Thanks to this property, we can express $y(\boldsymbol{x})$ by an explicit formula, allowing for a physical interpretation. For the mutual inductance, we have

$$M(\boldsymbol{x}) \approx M_p(\boldsymbol{x})\hat{r}(\boldsymbol{x}) \tag{16a}$$

$$\hat{r}(\boldsymbol{x}) = \sigma\hat{r}'(\boldsymbol{x}) + \mu \tag{16b}$$

$$\hat{r}'(\boldsymbol{x}) \approx w_3x_3' + w_4x_4' + w_5x_5' + w_6x_6' \tag{16c}$$

where the quantities with prime denote their standardized values. The values of σ, μ, and w_i are summarized in Table III. We neglected w_i of small values. From Table III, we can see that x_5, x_6, which represent the distance between the coils and the distance between the coil and the magnetic layer, respectively, have the largest impact on the correction factor r. Their negative sign means that r increases as they decrease. The sloping plane as a function of x_5, x_6, can be seen in Fig. 4(c). Having an explicit formula like (16) would be useful for quickly estimating device characteristics. Moreover, we can identify the important and less important variables in $r(\boldsymbol{x})$.

Table III
Coefficient values in (16)

σ	μ	w_3	w_4	w_5	w_6
0.20	2.28	2.60 $\times 10^{-1}$	3.42 $\times 10^{-1}$	-6.27 $\times 10^{-1}$	-4.47 $\times 10^{-1}$

C. Electromagnet model

As a second numerical example, we consider the electromagnet model shown in Fig. 5. This model has seven geometrical parameters $\boldsymbol{x} \in \mathbb{R}^7$, which are subjected to a uniform distribution over the range shown in Table IV. The prediction is performed for the magnetic force $F(\boldsymbol{x})$ acting on the I-shaped core. The material of the C and I-shaped magnetic cores are assumed to be M50 whose BH characteristic is provided by FEMM. For the physics-based model, we use here

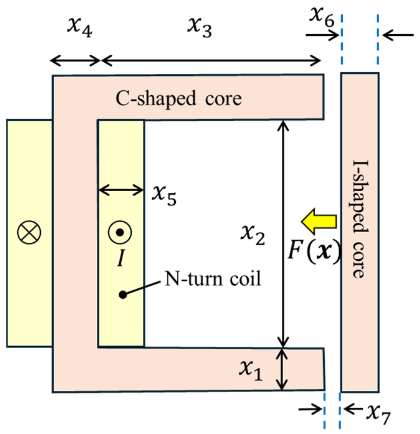


Fig.5 Two-dimensional electromagnet model
Magnetic cores have BH characteristic of M50.
$\boldsymbol{x} = (x_1, x_2, \ldots, x_7)^t$.

a magnetic circuit with constant permeability μ for the C and I-shaped cores neglecting the leakage flux. We compute $F(\boldsymbol{x})$ from the following circuit equations

$$\left(\frac{x_1+x_2}{x_4\mu} + \frac{2x_3+x_4}{x_1\mu} + \frac{2x_7}{x_1\mu_0} + \frac{2x_6}{x_1\mu} + \frac{x_1+x_2}{x_6\mu}\right)\Phi = NI \tag{17a}$$

$$F = \frac{\Phi^2}{x_1\mu_0} \tag{17b}$$

where N, I denote the number of turns and current amplitude.

For this model, we pay attention to the effect of the magnetic saturation on the prediction accuracy of $F(\boldsymbol{x})$. When $I = 10$A, the cores have little magnetic saturation. This makes F change sensitively depending on $\boldsymbol{x}$. In this situation, the circuit model (17) is rather accurate, which has $E = 5.41\%$. However, when $I \geq 40$A, E is greater than 35% due to the magnetic saturation even if μ is adjusted to have highest accuracy. Note that we did not use the nonlinear permeability in the magnetic circuit because we found that this did not improve the prediction accuracy of the machine learning.

The prediction accuracy for the three methods is plotted in Fig.6. We can see that the direct method gives large prediction error when $= 10$A while it is improved as I increases. This can be understood from the fact that F depends insensitively on $\boldsymbol{x}$ for the deeply saturated core. Hence, the response surface becomes flatter as I increases. The multiplicative correction gives the best prediction accuracy among the three methods. The additive correction is less accurate than the direct method at two points. The numerical results revealed that the reason is the same as in the WPT model; the response surface for $r(\boldsymbol{x})$ is nearly flat whereas that for $d(\boldsymbol{x})$ is curved.

Table IV Specification for electromagnet model
geometric unit: cm

parameters	range	parameters	range
x_1	$(1.5, 2.5)$	x_5	$(1.5, 2.5)$
x_2	$(8.0, 12.0)$	x_6	$(0.5, 1.5)$
x_3	$(8.0, 12.0)$	x_7	$(0.1, 0.5)$
x_4	$(1.5, 2.5)$	N	100 turns

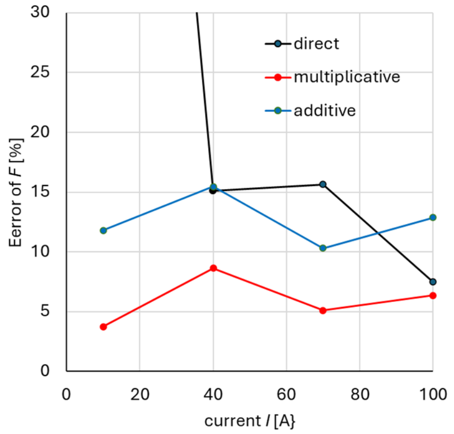


Fig. 6 Dependence of prediction error on current.
The error of the direct prediction is 115% when $I = 10$A.

IV. Possible Applications of Present Method

Table V lists potential applications of the proposed method. We have good analytical models for, e.g., planar inductors [16] as well as surface permanent magnet (SPM) motors [17] and synchronous reluctance motors (SynRMs) [18] as written in (a), which can be used as the physics-based models as in the WPT model discussed in this paper. Cases (b) through (d) can also be regarded as in the categories of multi-fidelity or space mapping models. In case (b) a surrogate model is constructed from the analysis results of a 2D and 3D FE models. After the construction of a surrogate model, it becomes possible to make fast and accurate predictions of the characteristics of a 3D model for given any input. In particular, this method would be effective for the analysis and optimization of radial and axial flux motors considering the end effect.

In case (e), it is assumed that we have sufficient training data to accurately predict the characteristics of the original model. Then, we make a prediction for a modified model. This situation often arises in research and development of devices. What we need is just to build a small training dataset for the modified model in order to build a surrogate model based on the proposed method.

Table V Possible applications of the proposed method

cases	Physics-based models	Target model examples
(a)	analytical or empirical formula	FE models of inductors, SPM motor SynRM
(b)	magnetic circuit, reluctance network model PEEC model	FE models of SPM and switched reluctance motors, SynRM, induction machine, reactor, transformer, superconducting device, microwave device
(c)	2D FE model	3D FE model
(d)	linear/rough FE model	nonlinear/fine FE model
(e)	original model	modified model
(f)	numerical model	real device (experiment)

In case (f), we build a surrogate model for a real device where $y_p(\boldsymbol{x})$ and $y(\boldsymbol{x})$ correspond to the computed results and measured characteristics, respectively. For instance, the machine loss computed by FEM correspond to $y_p(\boldsymbol{x})$, while the measured losses correspond to $y(\boldsymbol{x})$. Quantities such as geometric parameters, duty ratio, frequency, moving or rotation speed, and bias voltage can form $\boldsymbol{x}$. This enables us to make fast and accurate corrections for $y_p(\boldsymbol{x})$ to account for effects not considered in the numerical model. One example is correcting computed efficiency maps for rotating machines based on small amount of measured data. Case (f) can be recognized as data assimilation in a novel sense between numerical and measured results. Note that different levels of fidelity are not used in (a), which is not a numerical model, nor in (e) or (f).

V. Conclusion

This paper presents surrogate models constructed using physics-aware machine learning. The multiplicative correction provides more accurate predictions because, unlike the direct prediction and additive correction, the response surface is planar. The surrogate model can be obtained with a small amount of training data using the present method with multiplicative correction. This method also allows us to obtain an explicit formula based on linear regression which allows us physical interpretation. We discuss the possible applications of the present method.

It remains an open problem to theoretically explain why the multiplicative and additive corrections tend to produce planar and curved response surfaces, respectively, so that the former is more accurate.

Acknowledgment

The author would like to thank Profs. S. Clénet and T. Sato for their comments on the magnetic field analysis and space mapping, respectively. This work was supported in part by JSPS KAKENHI number 25H00723.

References

[1] H. Igarashi, Topology optimization and AI-based design of power electronic and electric devices, Academic Press, 2024.
[2] P. D. Barba, Multiobjective shape design in electricity and magnetism, Springer Science+Business Media, 2009.
[3] S. Greydanus, M. Dzamba, J. Yosinski, Hamiltonian neural networks. Advances in neural information processing systems, 32, 2019.
[4] Q. Xu, Y. Shi, J. Bamber, Y. Tuo, R. Ludwig, X. X. Zhu, Physics-aware machine learning revolutionizes scientific paradigm for machine learning and process-based hydrology, arXiv:2310.05227, 2023.
[5] Y Sakamoto, Y Xu, B Wang, T Yamamoto, Y Nishimura, 2023 IEEE Int. Electric Machines & Drives Conference (IEMDC), 2023.
[6] Y. F. Liu, L. Peng, W. Shao, An efficient knowledge-based artificial neural network for the design of circularly polarized 3-D-printed lens antenna, IEEE Trans. Antennas Propagation, 70(7), 5007-5014, 2022.
[7] D. Sarkar, T. Khan, S. R. Rengarajan, Prior-knowledge incorporated forward and reverse neural network modeling for estimating performance parameters of band-notched UWB MIMO antennas: smart estimation strategies, IEEE Antennas Propagation Magazine, 67(1), 34-47, 2023.
[8] M. G. Fernández-Godino, Review of multi-fidelity models, arXiv preprint arXiv:1609.07196, 2016.
[9] B. Peherstorfer, K. Willcox, M. Gunzburger, Survey of multifidelity methods in uncertainty propagation, inference, and optimization, Siam Review, 60(3), 550-591, 2018.
[10] J.W. Bandler, Q.S. Cheng, S.A. Dakroury, A.S. Mohamed, M.H. Bakr, K. Madsen, Space mapping: the state of the art, IEEE Trans. Microwave Theory Tech., 52(1), 337-361, 2004.
[11] G. Sun, G. Li, M. Stone, M., Q. Li, A two-stage multi-fidelity optimization procedure for honeycomb-type cellular materials, Computational Materials Science, 49(3), 500-511, 2010.

[12] Z. Guofo, W. Qiya, R. Wanbin, An output space-mapping algorithm to optimize the dimensional parameter of electromagnetic relay, IEEE Trans. Magn., 47(9), 2194-2199, 2011.
[13] C. M. Bishop, Pattern recognition and machine learning, Springer, 2006.
[14] https://www.femm.info/wiki/HomePage
[15] K. Sato, T. Kanamoto, R. Kudo, K. Hachiya, A. Kurokawa, Bayesian neural network based inductance calculations of wireless power transfer systems, IEICE Electronics Express, 20(5), 2023.
[16] H. M. Greenhouse, Design of planar rectangular microelectronic inductors, IEEE Transactions on parts, hybrids, and packaging, 10(2), 101-109, 1974.
[17] T. Lubin, S. Mezani, A. Rezzoug, 2-D exact analytical model for surface-mounted permanent-magnet motors with semi-closed slots, IEEE Trans. Magn., 47(2), 479-492, 2010.
[18] N. Bianchi, S. Bolognani, D. Bon, M. D. Pre, Rotor flux-barrier design for torque ripple reduction in synchronous reluctance and PM-assisted synchronous reluctance motors, IEEE Trans. Industry Appl., 45(3), 921-928, 2009.

H. Igarashi (M'94-) received the Ph.D. degree in engineering from Hokkaido University in 1992. He has been a professor at Graduate School of Information Science and Technology, Hokkaido University, since 2004. He is now a professor emeritus and a specially appointed professor at D-RED, Hokkaido University. His research area is computational electromagnetism, optimal design and AI-based design of electric machines. He has authored and coauthored more than 300 peer-reviewed journal papers.