PAPER • OPEN ACCESS

Sensitivity analysis and machine learning modelling for the output characteristics of rotary HTS flux pumps

To cite this article: Zezhao Wen *et al* 2021 *Supercond. Sci. Technol.* **34** 125019

View the article online for updates and enhancements.

Sensitivity analysis and machine learning modelling for the output characteristics of rotary HTS flux pumps

Zezhao Wen[*] ，**Hongye Zhang** and **Markus Mueller**

School of Engineering, University of Edinburgh, Edinburgh EH9 3JL, United Kingdom

E-mail: Zezhao.wen@ed.ac.uk, hongye.zhang@ed.ac.uk and Markus.Mueller@ed.ac.uk

CrossMark

Abstract

High temperature superconducting (HTS) rotatory flux pump, or so called HTS dynamo, can output none-zero time-averaged DC voltage and charge the rest of the circuit if a closed loop has been formed. This type of flux pump is often employed together with HTS coils, where the HTS coils can potentially work in persistent current mode, and act like electromagnets with considerable magnetic field, having wide range of applications in industry. The output characteristics of HTS rotary flux pumps have been extensively explored through experiments and finite element method (FEM) simulations, yet the work on constructing statistical models as an alternative approach to capture key characteristics has not been studied and published. A 2D FEM program has been used to model the HTS rotatory flux pumps and evaluate the effects of different factors upon the output voltage through parameter sweeping and analysis of variance. Typical design considerations, including operation frequency, air gap, HTS tape width and remanent flux density have been investigated, in particular the bilateral effect of HTS tape width has been explained by looking at the averaged integration of the electric field over the tape. Based on the data obtained from various simulations, regression analysis has been conducted through a collection of machine learning methods and demonstrated that the output voltage of a rotary flux pump can be obtained promptly with satisfactory accuracy via Gaussian process regression, aiming to provide a novel approach for future research and powerful design tool for industrial applications using HTS rotary flux pump devices.

Keywords: HTS dynamo, flux pump, T–A formulation, machine learning, regression

(Some figures may appear in colour only in the online journal)

1. Introduction

In order to enable the advance of a variety of applications in the future [1–5], such as wind turbines, all electric aircraft propulsion systems and energy storage devices, the clear trend in designing electric machines is to maintain the 'power density' ratio as high as possible [6–8]. High temperature superconducting (HTS) magnets have been expected to be the optimal option, owing to their superior ability to carry considerable current and hence generate very high magnetic field with the cost of dramatically reduced consumption of space and materials, compared to their common counterparts [9–11]. Nevertheless, the applicability of HTS magnets in industrial applications has been severely restricted. One significant challenge is that the current leads bridge the cryostat and the ambient environment in room temperature plus the inherent Joule losses, imposing crucial thermal load on the cooling system [12–14].

---

[*] Author to whom any correspondence should be addressed.

In addition, the essential bulky power supplies accompanied with sophisticated electronic devices are expensive and introduce extra complexity [15].

Apart from the traditional approach to feed current through current leads, it is possible to energize HTS magnets without any physical connections, using the so called 'flux pump'. Flux pumping is a fully contactless method to inject current into a closed-circuit and generate the corresponding magnetic field within the loop. Up to now, there has been worldwide research regarding different types of HTS flux pumps, and the perspectives have been summarized in [16]. In particular, the HTS rotary flux pumps [17] that utilize rotating permanent magnets to provide a time and space varied magnetic waveform and hence induce a DC biased voltage across the HTS stator wire, has drawn extensive attention, especially in the field of HTS machine design. The simple and straightforward structure of the rotary flux pumps, and the benefits they offer to eliminate the brushes and slip rings, make them a very promising candidate for exciters in future HTS machines.

Flux pumping effects in HTS rotary flux pumps have been investigated experimentally for different influential design parameters, such as the air-gap distance [18], the operating frequency [19], the width of the stator wire [20] and the geometry of the magnet [21], and typical curves for output characteristics have been obtained. Both qualitative and quantitative analysis [22–29] have been done previously for the output characteristics of HTS dynamo, but little attention has been paid to integrate all design parameters together and then derive a model that can capture the output characteristics in a fast and accurate manner. Although experiments and finite element method (FEM) simulations are capable of extracting essential information for quantitative analysis, the frequent change of a variety of parameters is costly in terms of time due to the complexity of models and experimental arrangements. In order to address this challenge, a statistical model has been applied to an engineering formulation to quantify the relationship between the open circuit voltage V_{oc} versus variable operating frequency f, air gap g, stator width W_s and remanent flux density B_r within a rational range.

2. Methodologies

The methodology can be divided into two parts, in which the 1st step was to establish a numerical model to simulate the operation of HTS rotary flux pumps. Simulation results are then collected for different input parameters to go through a series of machine learning processes, forming the 2nd step, from which the best machine learning model can be selected to predict the output of HTS rotary flux pumps accurately without having to simulate the complete system design.

2.1. Numerical model

Various formulations have been implemented into commercial software, e.g. COMSOL Multiphysics, for HTS modelling,

including coupled *H–A* formulation [30], *H*-formulation with shell current [25–27], segregated *H*-formulation [31], minimum electromagnetic entropy production [32, 33], coupled *T–A* formulation [34, 35], integral equation [36], volume integral equation-based equivalent circuit [37]. Adequate details about comparison for each formulation can be found in recent benchmark research [28]. The unique feature of coupled *T–A* formulation is that it simulates the superconductor (SC) as a 1D object in this 2D problem, and it was chosen to simplify the model complexity, as illustrated in figure 1.

The coupled *T–A* formulation was firstly proposed by Zhang *et al* [34], and validated to calculate the electromagnetic characteristics of HTS tape stacks and coils. For the superconductor domain, the current vector potential is chosen as the solved variable, and the governing equations are:

$$J = \nabla \times T \tag{1}$$

$$\nabla \times (\rho \nabla \times T) = -\frac{\mathrm{d}B}{\mathrm{d}t} \tag{2}$$

where *J, T, ρ* and *B* represent the current density, current vector potential, resistivity, and magnetic flux density, respectively. According to the thin strip approximation, only the superconductor layer of HTS coated conductors is taken into account and represented by a line without thickness, which assumes that the distribution of current density in the horizontal (J_x) and vertical (J_y) direction are zero, simplifying the calculation to a 1D problem. Therefore, (1) and (2) can be simplified as follows:

$$J_z = \frac{\partial (T \cdot n_y)}{\partial x} \tag{3}$$

$$-\frac{\rho \cdot \partial (T \cdot n_y)}{\partial_x{}^2} = -\frac{\partial B_y}{\partial t}. \tag{4}$$

For the whole space, the magnetic vector potential *A* is introduced to link the distribution of sheet current density in the superconductors, and hence comes to the *A*-formulation part:

$$\nabla \times \left(\frac{1}{\mu}\nabla \times A\right) = J \tag{5}$$

$$B = \nabla \times A. \tag{6}$$

The current density in the non-superconducting region is not of interest, since that would not affect the open circuit voltage induced on the SC tape in this model. Instead of introducing the magnetic vector potential *A* in the non-superconducting parts, the magnetic scalar potential V_m can be utilized:

$$\nabla \times (-\nabla V_m) = J. \tag{7}$$

By replacing *A* with V_m, the solution time can be dramatically reduced, because the computational cost of scalar is much less compared to vector.
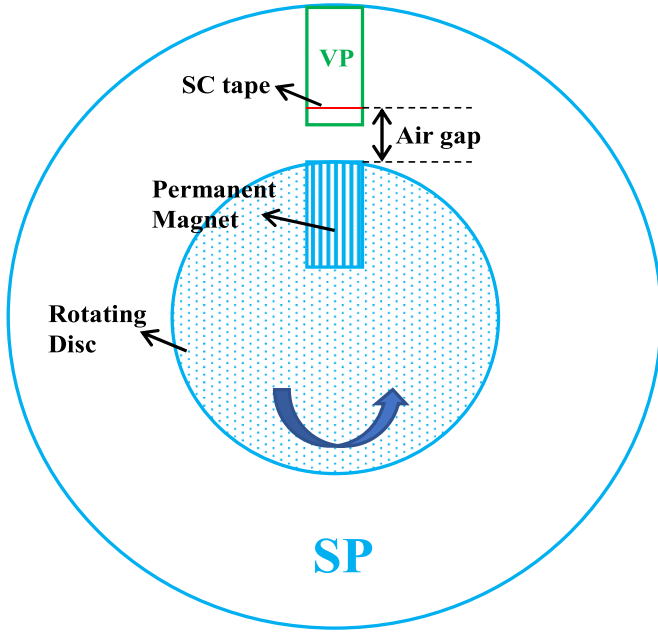
Figure 1. Schematic illustration of the basic geometry of the 2D *H*-formulation model used in this work, showing the modelled cross-section of the device with a 12 mm wide stator tape. Different model domains are indicated respectively, namely the magnetic vector potential (VP) region, the magnetic scalar potential (SP) region, and the *T*-formulation (superconductor) region. Note that the geometry is presented in the *x*–*y* plane, and the $+z$ orientation aligns with direction that points into the plane.

In addition, the *E*–*J* power law was employed to represent the electrical characteristics of superconductors:

$$E_z = \rho J_z = E_c \left(\frac{J_z}{J_c(B)} \right) \left(\frac{|J_z|}{J_c(B)} \right)^{(n-1)} \tag{8}$$

where E_c and *n* were chosen as 1×10^{-4} V m^{-1} and 20, respectively. The critical current density $J_c(B)$ is a magnetic field dependent variable, which is commonly described by an anisotropic Kim-like model [38]. In order to compare our results to the work presented in [25], the magnetic field dependence of the critical current density is represented by a practical interpolation function directly obtained from the experimental measurements in [25]. The instantaneous equivalent voltage V_{eq} and open circuit voltage V_{oc} are defined as follows:

$$V_{eq}(t) = -L \cdot \frac{1}{W_s} \int_0^{W_s} E_Z(x,t) \, dl \tag{9}$$

$$V_{oc}(t) = \frac{1}{T} \int_T^{2T} V_{eq}(t) \, dt. \tag{10}$$

Note that *L* is the active length of the dynamo, and the 2nd cycle was chosen to perform the calculations, avoiding any initial transient effect that may be present in the 1st cycle.

2.2. Machine learning

Regression is a popular method for data processing, which is found to be very effective in exploring the underlying relations between a set of predictor variables and corresponding response variables. The aim of regression analysis is to predict outputs by feeding relevant inputs, which can be achieved either by parametric regression or non-parametric regression. Parametric regression is the process to describe the correlations among all inputs and outputs by mathematical expressions, and a typical model for a polynomial regression can be expressed as:

$$\widehat{y^i} = \alpha + \beta_1 x_1^i + \beta_2 x_2^i + \cdots + \beta_n x_n^i + \epsilon^i \tag{11}$$

where α is the intercept to describe a constant factor that can be inherently included in the expression, β_n is the coefficient for *n*th independent predictor variable x_n, $\widehat{y^i}$ and ϵ^i are the predictive response variable and relative error for *i*th observation, respectively. In most cases, parametric regression is preferred, because it gives intuitive hints about what roles each of the inputs play in determining the output and provides full details about the numerical relations. Nevertheless, parametric regression is feasible only if the investigated data correlations can be described by explicit formulations. When the underlying relations among a set of input and output data involves highly complex mathematical transformations, it is challenging and inefficient to obtain analytical formulations. Fortunately, a modern artificial intelligence-based approach enables non-parametric regression, which is the process to summarize the key features about the dependence of outputs on the inputs in a certain manner that can be understood by computers, and hence the prediction of outputs by feeding relevant inputs is achieved without deriving explicit expressions.

The five popular artificial intelligence-based methods have been employed via Python Scikit-learn 5.1 environment [39]: multi-layer perception neuron network (MLP), support vector machine (SVM), Gaussian process regression (GPR), decision tree (DT) and k-nearest neighbours (KNNs). A typical MLP neuron network comprises three layers at least: the input layer, hidden layer and output layer. The number of neurons in the input layer and output layer are predefined by the features of given problem, namely the number of predictor and independent variables. It is the process that takes place in the hidden layers links the output layer to the input layer, where each neuron takes values from the previous layer and conducts a weighted linear summation by a transformation with specific activation functions. An identity function is adopted as the activation function in the output layer to receive results from the last hidden layer, and MLP uses the squared error loss function given as:

$$\text{Loss}(\hat{y}, \, y, \, W) = \frac{1}{2} \|\hat{y} - y\|_2^2 + \frac{\alpha}{2} \|W\|_2^2 \tag{12}$$

where $\hat{y}$ signifies the predicted values, *W* is the weight coefficients matrix and α is a non-negative hyperparameter that puts penalty on complex models and hence suppresses overfitting.

In the field of non-parametric regression, GPR is considered as a highly effective approach because it can provide a framework to approximate sophisticated nonlinear functions with probabilistic estimates of the unseen uncertainty. The most important step in GPR is to train the specified data and find the optimized best hyperparameters θ comprised by parameters in the kernel function and a manually editable noise level σ, by the means of maximizing the log-marginal-likelihood for a GP as expressed in [40]:

$$\log p\left(y|X,\theta\right) = -\frac{1}{2}y^{\mathrm{T}}\left(K\left(X,X\right) + \sigma^2 I_N\right)^{-1}y$$
$$-\frac{1}{2}\log\left|K\left(X,X\right) + \sigma^2 I_N\right| - \frac{N}{2}\log 2\pi \quad (13)$$

where I_N denotes the $N \times N$ identity matrix, and N is the number of inputs.

SVM allows the declaration of an allowable error rate ε that tolerates all predicted values that are deviated from real values less than ε. In order to fairly take every data sample into consideration, one can introduce the slack variable ξ and ξ^*, which effectively releases the error rate so that all data samples can be potentially included in the safety margin. Meanwhile, any data point that lies beyond the margin bounds are undesired, so a coefficient C is necessary to impose penalties on such data points to minimize their numbers. The object of this regression process is now to ensure the flatness of the hyperplane, or alternatively speaking that the coefficient vector w^{T} must be constrained to restrict its impacts on the regression results and hence generalize the problem, which can be formulated as [41]:

$$\mathrm{MIN}\ \frac{1}{2}\left\|w\right\|_2^2 + C\sum_{i=1}^{l}\left(\xi + \xi^*\right) \quad (14)$$

$$\mathrm{st}\ \begin{cases} y - \langle w,x\rangle - b \leqslant \varepsilon + \xi \\ \langle w,x\rangle + b - y \leqslant \varepsilon + \xi^* \\ \xi, \xi^* \geqslant 0 \end{cases} . \quad (15)$$

The main idea of DT algorithm is to divide all data samples into smaller subsets by certain threshold criteria subsequently until decisions can be made on each data sample with enough accuracy. The finalized structure of model assembles a bottom-up tree consists of a root, certain amount of branch nodes and leaves. The critical step in DT regression is to select proper threshold criteria, which must ensure every split of the data samples pushes the model closer to the optimal. Assume there are N_m samples at node m represented by Q_m, a possible split $\theta\left(j, t_m\right)$, where j is the feature dimension and t_m is the threshold value, can partition Q_m into the left subset Q_m^{left} and right subset Q_m^{right}:

$$Q_m^{\mathrm{left}}\left(\theta\right) = \left\{(x,y)\left|x_j \leqslant t_m\right.\right\} \quad (16)$$

$$Q_m^{\mathrm{right}}\left(\theta\right) = Q_m\backslash Q_m^{\mathrm{left}}\left(\theta\right). \quad (17)$$

Then the quality of this split can be evaluated by employing the impurity function $P()$, which depends on the task to be solved:

$$M\left(Q_m,\theta\right) = \frac{N_m^{\mathrm{left}}}{N_m}P\left(Q_m^{\mathrm{left}}\left(\theta\right)\right) + \frac{N_m^{\mathrm{right}}}{N_m}P\left(Q_m^{\mathrm{right}}\left(\theta\right)\right). \quad (18)$$

The algorithms mentioned above all implement supervised machine learning process that relies on a specific objective function in those models, i.e. the loss function in MLP and impurity function in DT, which essentially converts the regression to an optimization problem of such a function. On the contrary, KNN model does not count on an objective function to converge the regression problem to an optimization process. Instead, KNN strictly relates the dependent variables to their corresponding inputs through an intuitive convention: any data samples that have the same values of dependent variables will also share the same values of independent variables.

In this paper, firstly it will be shown that parametric regression is not suitable for relating the predicted variable V_{oc} to the four parameters investigated here, namely f, g, W_{s} and, and then secondly conduct non-parametric regression via each of the approaches mentioned above. The adapted Nash–Sutcliffe model efficiency coefficient (NSE) is employed to evaluate the goodness-of-fit, and other top three most popular metrics [42], namely root mean square error (RMSE), mean absolute error (MAE) and mean absolute percentage error (MAPE) are included as complementary statistics to assess the regression results from different aspects. These numerical criteria are centred on prediction error, whose values can be interpreted mathematically and translated to the effectiveness of a certain regression. For example, the best possible score for NSE is 1, which implies that the regression model reaches the highest correctness level so that any unseen observations can be predicted with full confidence. In addition, graphical criteria proposed in [43] have been implemented, such as prediction distribution versus observations and the prediction rate curves, to illustrate the broader and more complex aspects of the relationship between the models and data, from which the most effective one can be selected.

$$\mathrm{NSE} = 1 - \frac{\sum_{i=1}^{n}\left(y_i - \widehat{y_i}\right)^2}{\sum_{i=1}^{n}\left(y_i - \bar{y}\right)^2} \quad (19)$$

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(y_i - \widehat{y_i}\right)^2} \quad (20)$$

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}\left|y_i - \widehat{y_i}\right| \quad (21)$$

$$\mathrm{MAPE} = \frac{100}{n}\sum_{i=1}^{n}\left|\frac{y_i - \widehat{y_i}}{y_i}\right|. \quad (22)$$

3. Results

3.1. Numerical model validation

In this paper, a numerical model was built to be solved in the commercially available FEM software COMSOL 5.5. Different state variables are chosen to solve Maxwell equations in
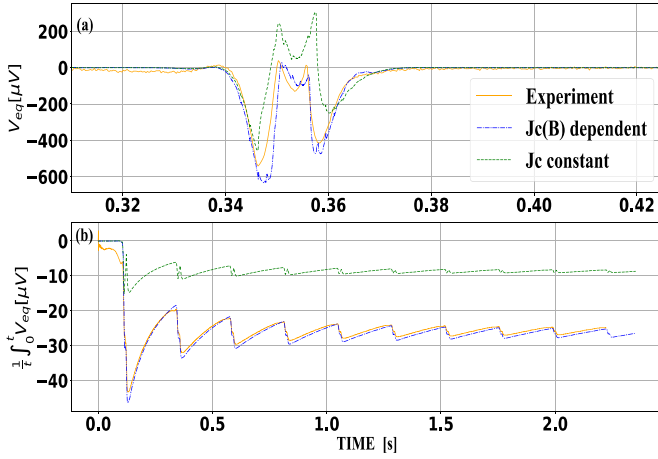
Figure 2. (a) Instantaneous open circuit equivalent voltage waveforms for the 2nd transit of the permanent magnet past the HTS tape calculated by our model used in this paper, ignoring any initial transient effects that may be present in the 1st cycle. (b) Cumulated voltage across the tape over ten cycles. The experimental results are drawn from [25].

different regions: the current vector potential A was solved only in the superconductor domain, and the magnetic scalar potential V_m was solved in the whole space. However, due to the internal software settings that forbids existence of current within any closed loop in the V_m region, A was solved in a small area with close proximity to the superconducting domain. In order to verify the applicability, this model was adapted to represent the experimental set up described in [25]: a 12 mm wide ReBCO tape (Superpower SF12050F) is separated from a NdFeB magnet by an air gap of 3.7 mm. The magnet has a remanent flux density of 1.25 T and is embedded in a cylindrical disc with radius 35 mm. The HTS dynamo is operated at low frequency 4.25 Hz. The simulation results have been compared with data measured in experiments as shown in figure 2, in which excellent correlation can be observed. It should be pointed out that the distinct four peaks and noticeable left-to-right asymmetry observed in experiments have been reproduced qualitatively, and the quantitative agreement between the experimental results and simulation results from the model that takes the critical current density dependence on magnetic field density into account is very good. The up-biased open circuit voltage obtained from the model with constant critical current density can be readily understood by considering the dynamic resistance effect [44–46]. As stated in [25, 44], the open circuit voltage of HTS dynamo originates from the actions of circulating current within the tape, which can be equivalently reflected by the voltage division on the backward and forward eddy current regions, and it is the electric field generated on the forward eddy current region that essentially forms the DC output voltage. The backward region, aligned beneath the PM position, would demonstrate a significantly higher dynamic resistance if the field dependence is accounted, and hence more proportions of the conserved EMF (by Faraday's law) are distributed on this region. As a combinational result, higher DC output voltage is generated eventually.
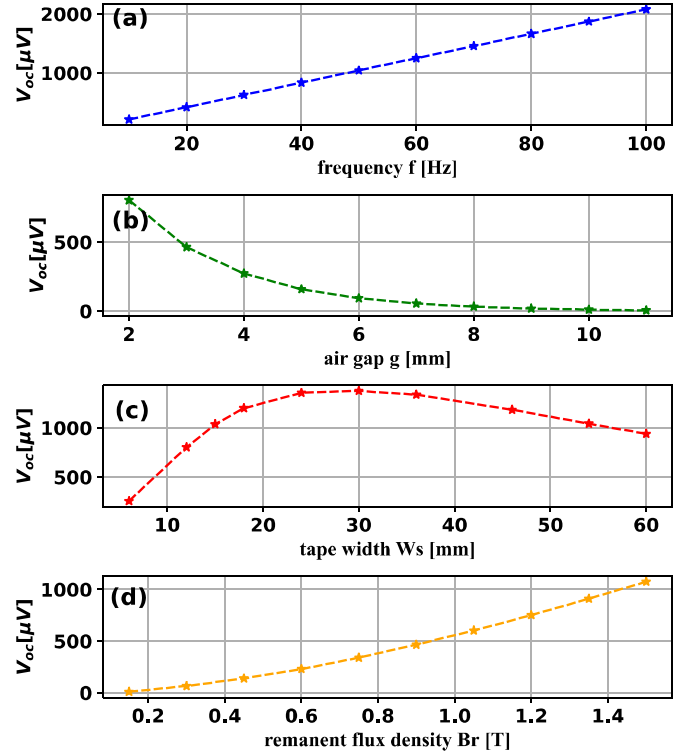

Figure 3. Calculated open circuit voltage V_{oc} versus design parameters: (a) frequency, (b) air gap, (c) stator width and (d) remanent flux density.

Four groups of simulations were arranged to primarily illustrate the relations between the controllable design parameters and the open circuit voltage of the rotary HTS flux pump, where in each group only one of the four aforementioned parameters is altered while the others remain constant. Figure 3(a) shows the frequency response within the input range from 10 to 100 Hz, aligning with the frequency range used in experimental results in [19]. It is worthwhile mentioning that a perfect linear relation was observed in [19] as the frequency increased from zero until a turning point was captured at approximately 100 Hz, which can attribute to the current interactions between different layers of the HTS tape [46–49] under high frequency. The single layer numerical model is not suitable for manifesting the multi-layer current interactions related phenomenon, hence the frequency is limited to 100 Hz to avoid this issue. Figures 3(b) and (c) illustrate how the output voltage varies with air gap and width of superconductor tape respectively, namely the output voltage decreases towards zero along the increase of air gap and increase from zero along the increase superconductor tape width, and both of them show acceptable agreement with previous work in [18, 20], which further validates the model. The impact of the remanent flux density of the permanent magnets upon the open circuit voltage is included, and initial results implies it has a clear positive correlation with the open circuit voltage. In summary, within the ranges investigated in our work, it was found the open circuit voltage tends to only monotonically decrease with the air gap, whilst increasing with the other parameters. In particular, the response variable of open circuit voltage demonstrates obvious linearity with frequency,
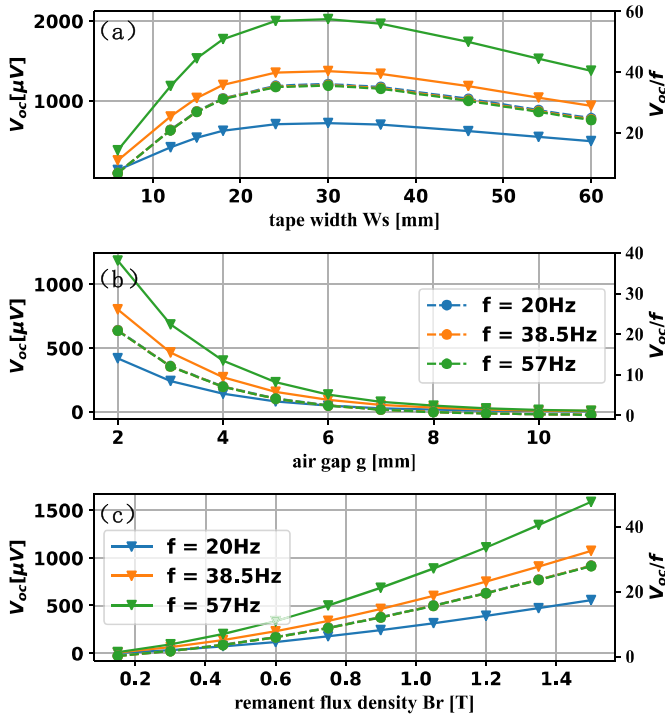
Figure 4. V_{oc} response for each parameter (solid lines are referred to the vertical axis on the left, and the right vertical axis are for the dash lines).

and non-linearity was observed for all the rest parameters. As for the variations of SC tape width, it appears that there exists an optimal point, after which its positive correlation against the open circuit voltage becomes negative, which will be discussed in more detail later.

3.2. Impact of f on V$_{oc}$

Inspired by the observation in figure 3 that V_{oc} increases linearly with frequency, it is reasonable to believe that the function that describes V_{oc} about can be easily formulated by a straight line passing through the origin (based on the fact that no output is expected if the device is not operating at all) under the condition that all other parameters remain the same. Hence, it leads to a critical question: does this linear relation hold through the variation of other parameters? In order to investigate this problem, each of the other three parameters were varied under different frequencies, so that the frequency-normalized output profile V_{oc}/f can be plotted for each of them. It can be seen from figure 4 that the response level of V_{oc} against each of the parameters does vary under different frequencies as demonstrated by the separated solid lines, whilst the frequency-normalized output profiles collapse together as indicated by the dash lines in each subplot. Taking figure 3(a) into consideration, the curve behaviour mentioned above can happen only if the frequency acts as an 'amplifier' in the determination of V_{oc}, which implies that the linear relation between V_{oc} and f holds despite any changes from other parameters. That is to say, the four-variable problem concerned in this paper can now be reduced to a three-variable problem by a simple frequency normalization as shown in following:

$$V_{oc}(f,g,W_s,B_r) = f\frac{V_{oc}(f,g,W_s,B_r)}{f}$$
$$= f \cdot V_{oc}(g,W_s,B_r). \quad (23)$$

3.3. Impact of g and B$_r$ on V$_{oc}$

The consequence of adjusting g or B_r will be reflected by the changing of peak flux density on the surface of the tape, as well as the flux passing through it. More precisely, increasing g or decreasing B_r will lead to a smaller magnetic field seen by the tape surface, and less flux can be captured by the close loop within the tape, e.g. if the magnet is moved away from the tape or replaced by another one with smaller remanent flux density, the tape is then effectively exposed to a smaller magnetic field, and vice versa. According to Faraday's law, the electromotive force around a closed path is equal to the negative of the time rate of change of the magnetic flux enclosed by the path. Under the condition of fixed frequency, the magnetic flux transversing the tape per unit time is determined by the remanent flux density as well as air gap. Hence, the open circuit voltage V_{oc}, which is time integration of the electromotive force, should drop according to an increase in g or decrease in, due to the reduction of flux along with its changing rate. This is clearly revealed by the results in figures 3(b)–(d), where it was shown that V_{oc} drops from 800 μV with a 2 mm air gap and approaching zero after the gap distance exceeds 8 mm, while V_{oc} climbs from approximately 0–1000 μV for values of B_r varying between 0.15 and 1.5 T. It is worthwhile mentioning that both extreme cases, where the gap distance goes to infinity and the remanent flux density reaches zero (the magnet loses magnetism) are well demonstrated in the results by confidently implying a zero output in those conditions.

Even though the individual effect of g and B_r are clearly revealed and can be well explained as stated above, it is challenging to test whether or not the same effect can be influenced by changing other parameters. Unlike frequency, its effect on other parameters can be verified by a simple normalizing transformation about response versus each of the other parameters. The non-linear relationship between V_{oc} and g and B_r make it difficult to achieve the same validation easily, and hence an analytical formulation that assembles (23) for g and B_r is not obtainable. As a result, parametric regressions on g and B_r are not preferred here, leading to the consideration of non-parametric regressions as an alternative.

3.4. Impact of W$_s$ on V$_{oc}$

As indicated in figure 3(b), V_{oc} increases gradually as the tape width increases from 6 to 30 mm, after which it starts to decrease if W_s is further increased. This matches the experimental results presented in [20], where the V–I curve intercept on the x-axis firstly became more positive and then less positive for the tape width increasing from 6 mm to 46 mm. Yet the V–I curve from our model plotted in figure 5 differs from the one presented in [20], in which a straight line is interpolated for each value of tape width. In fact, in their experiments, the net transport current of the tape is constrained by the critical current of the superconducting coil connected to the tape,
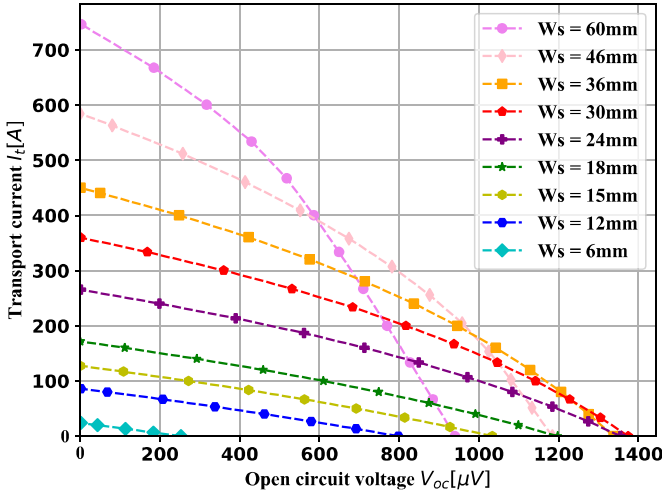
Figure 5. *V–I* curve (short circuit current versus open circuit voltage) obtained from a range of SC tape widths.

Figure 6. Derived parameters from *V–I* curve: (a) short circuit current I_{sc}, (b) calculated *f*-ratios (the values of α and β); (c) derived resistance in open circuit condition R_{oc} and short circuit condition R_{sc}; versus the SC tape width.

so for the section of curve with the range of current which exceeds the critical current of the connected coils, it can only be interpolated by 1st few measurements where the transport current is below coil critical current. In figure 5, the limits of coil critical current have been removed so that the full profile of *V–I* curve for each tape width can be obtained. The intercept on the *y*-axis of each curve represents the theoretical maximum current that can be pumped if the connected coil has infinite critical current. It can be seen from figure 5 that the *V–I* curves tend to show non-linearity for large tape width, which is in alignment with the results in [26].

The characteristics of this *V–I* curve can be further explored by analysing I_{sc} and V_{oc} separately. It is not surprising that the short circuit current I_{sc} increases with the tape width, because wider tape tends to have higher critical current. Two *f*-ratios are defined to measure the proportion of the capacity of the tape occupied by the transporting current, α and β given as:

$$\alpha = \frac{I_{sc}}{I_{c,self}} \tag{24}$$

$$\beta = \frac{I_{sc}}{I_{c,min}} \tag{25}$$

where $I_{c,self}$ denotes the critical current when the tape is not exposed to any external magnetic field, and $I_{c,min}$ refers to the minimum value of immediate critical current as the applied magnetic field varies. Figure 6(b) illustrated that I_{sc} can reach up to about 55% of the self-field critical current and 75% of the minimum critical current for a 36 mm wide tape, and then stay at this value. Moreover, if we consider this dynamo as an electric circuit comprised by a voltage source with an internal resistor, then its resistance can be derived from the slope of the *V–I* curve. Please note that due to the existence of non-linearity for some curves, a data point based differential slope has been used to calculate the resistance for each curve. Figure 6(c) plots the resistance R_{sc} and R_{oc} for each tape width when the tape is shorted (e.g. there is no voltage drop across the tape $V_{oc} = 0$) and when the tape is isolated (e.g. the net transport
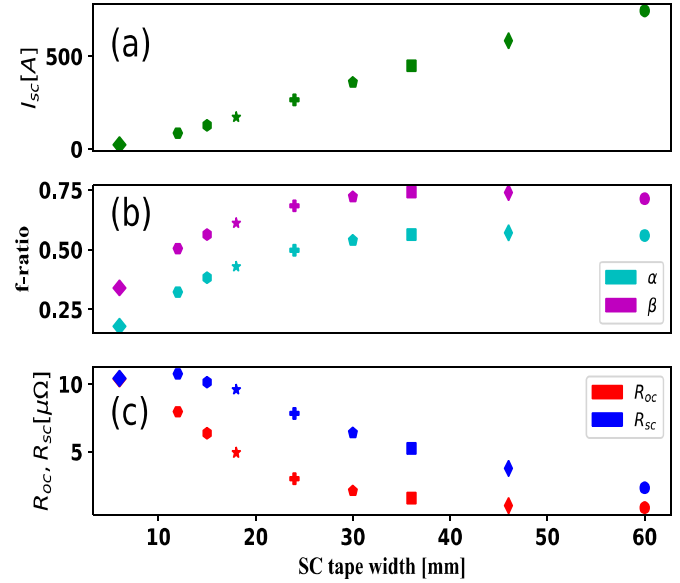
current $I = 0$), whose values indicate the resistive properties of the tape under high and low current level respectively. It can be seen from figure 6 that R_{sc} and R_{oc} both drop as the tape width increases, meaning that the wider the tape the more superconductivity can be utilized.

The most noticeable feature of V_{oc} versus tape width curve is that it increases first and then decreases as the tape gets wider, which makes W_s the only parameter that has a non-monotonical relationship with V_{oc} out of the four parameters investigated in this paper. In order to understand why this occurs, we need to go back to the origin of V_{oc}. The uneven and nonlinear resistivity distribution of the tape can produce an asymmetric electric field induced by the eddy current under a changing magnetic field as shown in figure 2(a), which can result in a nonzero time integration and hence an averaged dc voltage output. That is to say, the instantaneous electric field E_z distribution, or equivalently the averaged instantaneous voltage V_{eq}, has direct impact on the determination of V_{oc}. Based on this idea, it is reasonable to deduce that the value of V_{oc} can be roughly reflected by the magnitude of the 1st peak of V_{eq} curve. As it tells in equation (9), for each instant V_{eq} is calculated by averaging the integration of the electric field along the tape, and it is the net remanent integration of the negative electric filed that determines the final value of V_{eq}. Thus, we will need to know how the electric field is distributed along the tape to investigate the formation of V_{eq}. The distribution profiles have been captured for selected tape width in figure 7 when the 1st peak appears. It can be seen from figures 7(a)–(c) that the integration over the range for negative electric field (the region marked as N_1 in figure 7(d)) increases significantly as the tape becomes wider, which contributes to a larger value of V_{eq}. As illustrated from figures 7(d)–(f), once the tape reaches a certain width, the integration for electric field in region N_1 remains approximately constant, and
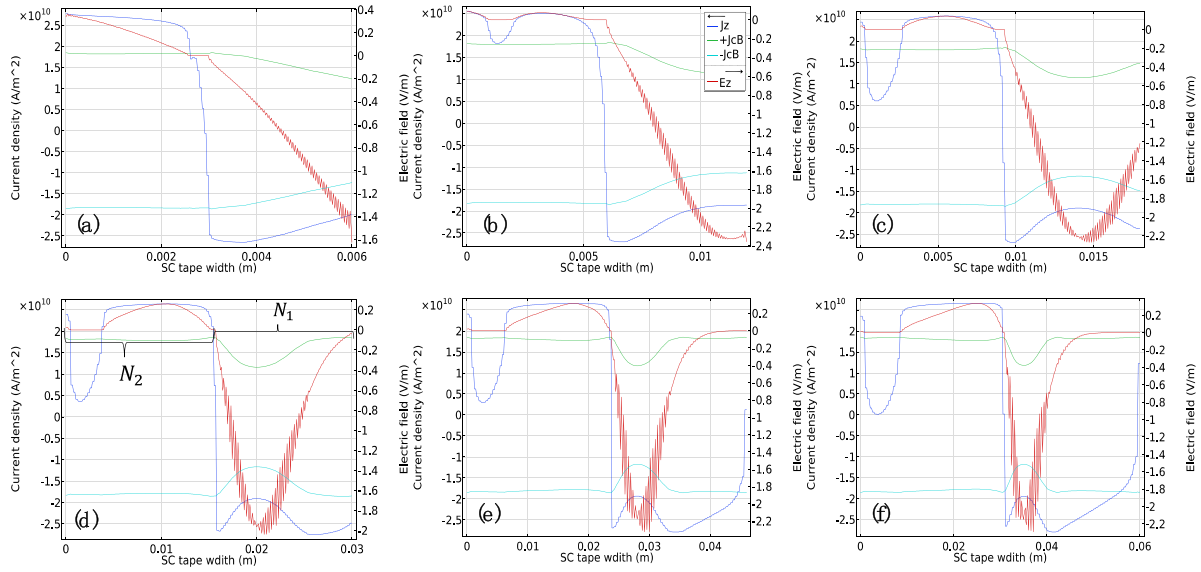
Figure 7. The current and electric field distribution at 1st peak of V_{eq} for selected tapes (6, 12, 18, 30, 46, 60 mm respectively).

further width increment will only raise the integration over the range for positive electric field (the regions marked as N_2 in figure 7(d)). Consequently, a fixed integration over the negative region will be cancelled more and more by an increasing integration over the positive region, leading to a small V_{eq} as a combinational effect. This smaller V_{eq} peak value implies that less V_{oc} is obtainable through the time integration of V_{eq} over the whole cycle, which explains why V_{oc} can be promoted by the tape width but suppressed if the tape becomes too wide. Besides, the definition of 'too wide' here is restricted by the sole variation of tape width, but the effects described above can be also influenced by the geometry of the device, e.g. the airgap between the magnet and tape, the dimension of the magnet and even the radius of the rotary disc. More comprehensive analysis is required to fully investigate the 'width effect' and complete the definition of 'too wide' by cooperating all relevant parameters, which is beyond the scope of this paper and will be addressed in the future.

3.5. Regression analysis

Although the HTS dynamo can now be simulated by various numerical models via FEM programs, including the *T–A* formulation based model proposed in this paper, which enables us to obtain considerable knowledge of the physical mechanism that takes place during the operation, but in terms of industrial design, the ultimate goal is to make use of the functionality of certain devices. As for this HTS dynamo, it is expected to be utilized as an energisation source that can pump current into load coils and hence form a HTS magnet. From this point of view, it is the averaged output dc voltage V_{oc} that matters. There is no doubt that V_{oc} can be measured from experiments, but these are often time-consuming and expensive and not accessible to all occasions. Valid FEM simulations are considered to be a promising substitution for setting up real experiments to get comprehensive results in less time and saved cost.

However, these FEM models can sometimes require relatively long time for hours, and the applicability is also limited by user knowledge and accessibility to a sophisticated FEM program, which is not guaranteed. Artificial intelligence methods are powerful tools to accurately model certain data behaviour, which have been recently involved in solving applied superconductivity problems [50]. For instance, artificial neural network (ANN) has been utilized in [51–54] to estimate the calculation of AC losses in specific SC system and aid the design of SC facilities. As an alternative, we have explored the feasibility of constructing a statistical model, which can give accurate V_{oc} values according to the given controllable parameters within seconds, and hence provides primary guidance for the design of such devices in practical applications.

Thousand sets of different parameters, in which each of them lies in the rational range for a practical design process, have been randomly chosen and simulated by the numerical model described above to derive V_{oc} for each set of parameters. As suggested by equation (23), all output voltage variations that are solely affected by frequency can be easily quantified by simple normalization calculation, so we have precluded the frequency parameter *f* in following investigations, e.g. all parameter settings share the same value of frequency (50 Hz), in order to simplify the modelling work. According to the conventional 8–2 rule, data samples are evenly divided into five folders for cross validation, in which four folders would be utilized as training group and keep the remaining one as a testing group for each validation.

3.5.1. Neuron network model. In this MLP model, there are several tuning parameters that have direct impact on the regression performance. In order to figure out the best combination of parameter settings, a grid search was performed over possible combinations for different hidden layer sizes, activation functions and weight coefficients update solvers. The determination of the hidden layer size to be tested remains

Table 1. Parameter setting for the best MPL model.

Hidden layer size	Activation function	Weight coefficients solver
1×6	Tanh	Adam

Table 2. Cross-validated metrics for the best MPL model.

RMSE	MAE	MAPE	NSE
58.56	39.73	9447.38	0.975268

Table 3. Parameter setting for the best GPR model.

Kernel function	μ	Number of optimisations
Matern	2.5	20

Table 4. Cross-validated metrics for the best GPR model.

RMSE	MAE	MAPE	NSE
2.68	1.25	0.0684	0.999949

Table 5. Parameter setting for the best SVM model.

Kernel function	Precision level	Penalty coefficient
RBF	3.237613×10^{-05}	87 472

Table 6. Cross-validated metrics for the best SVM model.

RMSE	MAE	MAPE	NSE
3.07	1.51	0.0743	0.999936

the critical step to perform the grid search, as it needs manual declaration. Firstly, following the basis that the least hidden layer should be used to save the training time and complexity [55], only one hidden layer is considered. On the other hand, it is stated in [56] that the number of hidden layer neurons should be less than twice of the number of neurons in input layer, hence the number of neurons was set from 1 to 7. Therefore, in total 84 combinations of parameter settings have been tested, from which the one with highest NSE score can be selected as the best model.

Note that the maximum iteration limitation for each model setting has been adjusted carefully to provide them enough space to reach a convergence. As can be seen from table 1 and table 2, the MLP model that has the Tanh activation function, Adam weight coefficients update solver and one hidden layer with six neurons performs the best. A further increase of the hidden layer size can possibly enhance the model performance, yet the model complexity and consequently the solution time will increase exponentially with the size of hidden layers. Considering the already obtained predictions accuracy, it is reasonable to utilize the existing model as the optimal one.

3.5.2. Gaussian process model. The controllable port that can be tuned in GPR is the kernel functions to be applied on each data sample to transform initial non-linear observations into a higher dimensional feature space, which almost determines the generalization properties of the model. Given a specific type of kernel function, there are inherent parameters that can be modified to adjust the characteristics of the function and hence influence the regression performance.

Common kernel functions for GPR models, include the squared-exponential kernel (RBF), the rational quadratic kernel, the periodic kernel, the dot-product kernel as well as the Matern kernel, which is a generalization of the RBF with an additional parameter μ to control the smoothness of the resulting function. Since μ is exclusive of the hyperparameters, the most important intermediate values of 1.5 and 2.5 for machine learning as stated in [57] has been manually chosen. Moreover, the results of a GPR model are highly dependent on the optimization of hyperparameters, so the number of allowable optimizations were also taken into account. As shown in table 3 and table 4, it is found that the GPR model that has the Matern kernel with μ of 2.5 and 20 runs of hyperparameter optimizations achieved the highest NSE score. It should be pointed out that it is possible to combine different kernel functions and hence form a new one or even manually construct a novel function

that is specially designed for certain problems. Nevertheless, this requires significant mathematical work and computation resources, so it is a good start point to look at the established popular kernel functions.

3.5.3. SVM model. Similar to GPR models, SVM models also employ kernel functions to conduct the regression process. Yet the tested functions are different from those applied in GPR models, due to the availability in Python package [39], the basic idea remains the same: further adjustment on specific kernel function can potentially provide altered results. Once again, the default settings for each kernel function were adopted to avoid the complexity of deep optimization. Compared with GPR models, SVM models have extra flexibility because they rely on not only kernel functions, but also the precision level ε and penalty coefficient C. Generally, small absolute value of ε and large value of C are required to obtain accurate predictions, e.g. the best model should be achieved when ε is close to zero and C goes to infinity. However, it is much more difficult to reach a convergence in such extreme condition. We have arranged randomized search for values of ε and C within the bounds of 10^{-5} to 1 and 1 to 10^5 and compare the performance of SVM models with different kernels functions. As illustrated in table 5 and table 6, the RBF kernel function with the precision level and penalty coefficient set to be 3.237613×10^{-05} and 87 472, respectively demonstrates remarkable capability of predicting the output from given input with minor errors.

3.5.4. DT model. From regression point of view, the general object is to minimize the residuals between predictions and observations. Therefore, any metrics that measure the residuals level can be utilized as the threshold criteria to be optimized at each split. Based on the availability of Python package, we have tested the effects of mean squared error (MSE),

Table 7. Parameter setting for the best DT model.

Threshold criterion	Maximum depth
F-MSE	10

Table 8. Cross-validated metrics for the best DT model.

RMSE	MAE	MAPE	NSE
79.07	49.28	0.5356	0.957424

Table 9. Parameter setting for the best KNN model.

Algorithm	Number of closest points
Ball tree	6

Table 10. Cross-validated metrics for the best KNN model.

RMSE	MAE	MAPE	NSE
60.40	35.74	0.5627	0.975355

Friedman-MSE (F-MSE), MAE as the threshold criterion with the maximum depth being tuned in the range of 5–200 to give the optimal results. As can be seen from tables 7 and 8, the DT model that has the maximum depth of ten and F-MSE threshold criterion achieved the highest NSE score and demonstrated the best prediction performance.

3.5.5. KNNs model. Based on its rather straightforward principle, KNN can predict independent variables for a new set of dependent variables by sorting the k closest data points in the given data sample and then averaging their correspondingly independent values. This makes the computation of the nearest neighbours a critical problem. Popular algorithms used to compute the nearest neighbours, including the ball tree, KD tree, and brute-force search have been tested. Besides, another important parameter that determines the performance of a KNN model is the number of the closest points selected to estimate the new predictions. The values of k from 1 to 50 have been tried in order to find out the optimal KNN model. Our findings are presented in tables 9 and 10, which concludes that the best KNN model was obtained by utilizing the ball tree algorithm to estimate new predictions by six closest points.

3.5.6. Model selection. Five machine learning based regression models have been primarily tested under various settings, and each of them have also been evaluated by numerical metrics to judge the goodness of such a model in predicting the HTS output voltage based on the four design parameters. Yet, numerical criteria just show the general aspect of a model, which are not enough in practice. As stated earlier, in order to illustrate the broader and more complicated aspects of the relationships between models and data, graphical criteria are indispensable. Let us define the samples with relative prediction errors below a given level as 'effective-predicted'
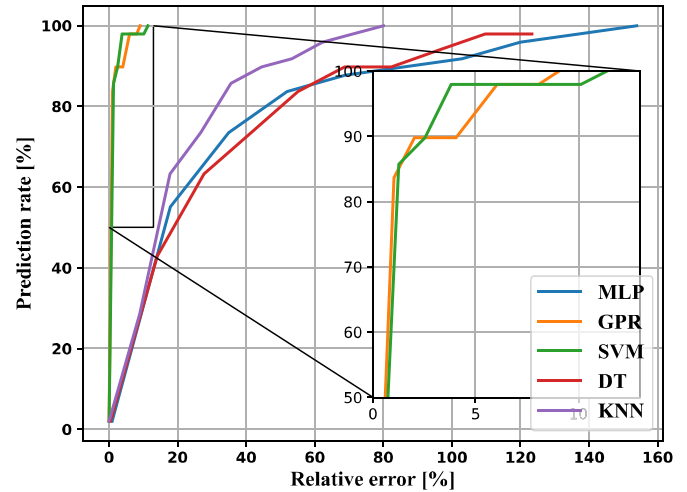
Figure 8. Prediction rate curve.

samples, and we can evaluate the predicting effectiveness of different models by the proportion of the effective-predicted samples with respect to the relative prediction error, which is referred as 'prediction rate'. In order to achieve this, we define $\delta_i = \left| \frac{y_i - \widehat{y_i}}{y_i} \right|$ and then plot the additive rate of n samples, e.g. from 0 to 1 by step of $1/n$ with respect to the sorted δ_i from zero to maximum relative error. The prediction rate curve of a model is obtained by joining those points. The ratio of effective-predicted samples can be accessed from vertical axis as relative error being horizontal axis. The introduction of this graph allows us to examine more details about the performance of each model. The area between the prediction rate curve and the vertical axis, which aggregates the relative errors, provides a linkage to the numerical cross-validation criteria formerly discussed. In general, a model can be considered good if this area is smaller, and it is visually easy to compare models with this plot. According to former analysis, the best candidate from each model group are chosen to plot their prediction rate curve.

As shown in figure 8, the prediction rate curves for MLP and DT and KNN models lie completely below those for GPR and SVM, which means that for any given error level, less samples can be effectively predicted by MLP, DT and KNN models than the others. Besides, MLP, DT and KNN curves cover a much wider range of errors than the other two curves, indicating that they have worse error containing ability. For instance, the maximum relative error generated by MLP, DT and KNN models are around 154%, 123% and 80%, respectively, and this strengthens the conclusion that MLP, DT and KNN models are not effective in providing stable and trustable predictions (note that the actual V_{oc} values for those samples being predicted range from 0 to 1971 μV). As for the GPR and SVM models, though they both demonstrated good capability of maintaining the relative errors well below 12%, the GPR model is believed to restrict the error level under a lower level as it arrives the full proportions slightly faster than the SVM model. Meanwhile, the curve of SVM almost consistently
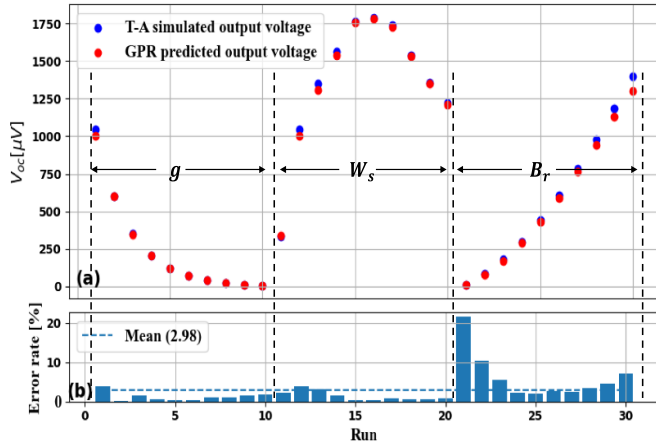
Figure 9. Re-produced results by our GPR model.

grows beneath that of GPR, which can be interpreted as afore-mentioned: more samples can be accurately predicted by GPR model than SVM model given a certain error threshold level. It is true that the capability of producing minor error predictions is not significantly different for GPR and SVM models, yet their solution time are relatively distinct. The SVM model requires more than 2 min to be solved, while the GPR model only takes few seconds, which is ten times less than the former one. Therefore, it is reasonable to select the GPR model as the optimal candidate to model the HTS dynamo output voltage, not only for its excellent accuracy but also due to the promise of saving considerable solution cost.

To verify this hypothesis, we tried to apply this GPR model to reproduce figure 3 by feeding three sets of parameters (except for the frequency *f*) as inputs and extract the values of V_{oc} as outputs. The results are presented in figure 9(a), where the blue dots V_{oc} values come from real simulations, e.g. those plotted in figure 3, and the red dots include V_{oc} values predicted by our GPR model. It is clear that the output voltage dependence against all three parameters are well described by our predictions, more precisely, V_{oc} drops over the increase of air gap (from 1st to 10th run), V_{oc} firstly increases then decreases over the increase of SC tape width (from 11th to 20th run), and V_{oc} climbs over the increase of the remanent flux density of the permanent magnet (from 21st to 30th run). As further illustrated in figure 9(b) that the average error rate for all predictions is 2.98%, which demonstrates the accuracy of our model. It should be noted that there are several predictions that have noticeable errors, such as the 21st, 22nd runs, which are corresponding to extreme conditions, e.g. B_r equals to 0.15 T and 0.3 T. Under such extreme conditions, weak output voltages are expected and hence those small values are highly sensitive to any mismatch. However, in practice B_r is commonly around 1 T, so that the incompetence of our GPR model for predicting extreme conditions will not undermine its performance. It can then be concluded that our GPR model is capable of predicting accurate output characteristics for HTS rotary flux pumps in general, and hence can be utilized as a reliable design tool for such devices.

Table 11. ANOVA results.

Source	D_f	F-value	p-value
g	2	23.668	0.000437
W_s	2	10.136	0.006
B_r	2	25.342	0.000345
$g \times W_s$	4	3.261	0.073
$W_s \times B_r$	4	3.835	0.050
$B_r \times g$	4	8.039	0.007

3.6. Sensitivity analysis

It has so far been revealed that the operating frequency *f*, air gap *g*, SC tape width W_s and remanent flux density of the permanent magnet B_r all have influences on the HTS dynamo output voltage V_{oc}, after which a GPR based statistical model was proposed to accurately predict V_{oc} by feeding each of parameters as inputs. However, to what extent does each parameter impact the output voltage still remains unclear. In order to fulfil the aim of this paper, which is to provide compressive guidance for HTS rotary flux pump design, the sensitivity of each parameter was further explored through an analysis of variance (ANOVA). As before, the operating frequency *f* is excluded from the analysis list to reduce the workload. As a as result, the number of second order interaction terms is reduced from C_4^2 to C_3^2, namely $g \times W_s$, $g \times B_r$ and $W_s \times B_r$. We have set the values of each parameter to three levels, because the relation between V_{oc} and W_s is not monotonic so that at least three values are required to represent the variation trend, which comprises of both increasing and decreasing parts. The orthogonal tables are believed to effectively reduce the cost of experimental schedules and mathematical calculations, by the means of only considering the representative values of each parameter in a proper way. The $L_{27}(3^{13})$ Taguchi array has been adapted to include the three parameters and their corresponding interaction terms. The ANOVA results are represented in table 11, which interprets each of variable via F-test and the statistical hypothesis (under a 95% confidence level) test. It is shown that all three parameters are significant as their F-values are far greater than the criterion $F_{0.05}(2,8) = 4.46$. The remanent flux density seems to be the prior dominant factor for determining the output voltage, and the importance of air gap is slightly lower while the importance of SC tape width is dramatically less with only 40% of the F-value for the remanent flux density. As for the interaction terms, it is believed that the interactions between g and B_r are very significant as its F-value suppresses the criterion $F_{0.05}(4,8) = 3.84$. Meanwhile, the other two interaction terms are relatively significant, because their F-values are located between the criterion interval $F_{0.05}(4,8) = 3.84$ and $F_{0.1}(4,8) = 2.81$. In a summary, the importance rank for each parameter is: $B_r > g > W_s$, and all 2nd order interaction effects between them are found to be influential as well. This can be interpreted as that the individual effect of each parameter should be considered first during the design and optimization of HTS rotary flux pumps, whilst the designer should be aware of the existence of interaction
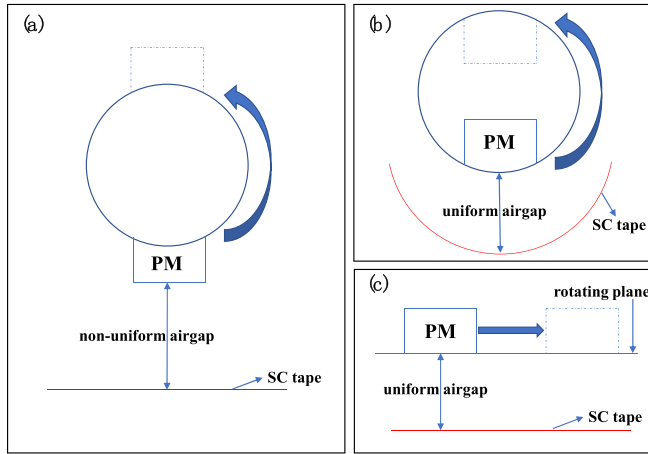
Figure 10. Schematic (cross-section) diagrams for different HTS dynamo configurations. (a) Radial flux arrangement with flat tape and PM surface. (b) Radial flux arrangement with curved tape and PM surface. (c) Axial flux arrangement with flat tape and PM surface.

effects since they possess unignorable influences upon the output characteristics.

4. Discussions

It should be noted that the simulation model utilized in this paper has adopted the radial flux geometry, in which the SC tape and PM surfaces are flat such that the airgap between the magnet and tape varies as the magnet rotates. This radial flux arrangement has been investigated in a number of previous studies [17, 18, 21, 25, 58, 59]. Alternatively, the superconducting tape and magnet could be curved so that their surfaces are parallel resulting in a constant airgap between tape and PM, but still in a radial flux configuration. Such an arrangement has been used in [26]. In an axial flux arrangement, a PM and superconducting tape both with flat surfaces can be mounted on flat discs, such that the airgap is uniform [19, 20]. Figure 10 illustrates these different geometries. The use of flat surfaces in a radial flux configuration resulting in a non-uniform airgap accounts for the difference in the results obtained in this paper compared to those in [26], in which the airgap is uniform. As can be seen from section 3.3, the open circuit voltage V_{oc} tends to increase versus the SC tape width W_s until it reaches an optimal value, after which it starts to decrease. However, it was found in [26], where the uniform airgap geometry was modelled, that V_{oc} would increase with W_s at the beginning, and it gradually saturates to a maximum value, which implies that the uniform airgap type HTS dynamo will not experience the bilateral width effect discovered in this paper. Besides, the results presented in figure 6 shows that the maximum short circuit current I_{sc} that can be possibly pumped will increase as the SC tape width and then saturate at 75% of the minimum critical current $I_{c,min}$ of the tape, whilst in [26] this proportion reached 100%. The reason for the distinct SC tape width response can be attributed to whether the airgap has been

maintained constant or not, though future work is required to further explore this phenomenon and provide clearer explanation for the underlying physics.

The machine learning techniques introduced in this paper have been only utilized to predict the output characteristics for the radial flux type HTS dynamo, but they can be confidently applied to uniform airgap radial and axial flux type HTS dynamo for similar analysis, despite the existence of different SC tape width response. Moreover, in order to fully investigate the bilateral effect reported in this paper, more design parameters of the HTS rotary flux pump are required to be considered, which will lead to an enhanced complex situation where more systematic interactions are involved. The algorithms and analysis approach adopted in this paper are ideal for such more complex interactions, as the techniques do not require any knowledge of the relationships involved.

5. Conclusion

In this paper the 2D T–A formulation has been applied to simulate HTS dynamo via FEM models, validated with experimental results in [25]. The difference between the results from a constant J_c model and a field dependent $J_c(B)$ model have been explained by considering dynamic resistance. It was pointed out that the $J_c(B)$ dependence tends to increase the dynamic resistance for the tape region beneath the permanent magnet, where the forward eddy current flows, and elimination of such dependence, e.g. assume constant J_c, will cause less of the conserved EMF to be distributed to the forward eddy current region and consequently result in a smaller open circuit voltage, from which we demonstrated the importance of considering field dependence of J_c in such models. Based on this numerical model, a comprehensive sensitivity analysis was conducted for the open circuit voltage V_{oc} versus the operating frequency f, air gap g, SC tape width W_s and remanent flux density of the permanent magnet B_r. The individual effect of each parameter on the open circuit voltage was investigated, in particular it was found that the tape width W_s has a bilateral impact on V_{oc}, and this is because the integration over negative electric field saturates when the tape is wide enough, yet further increasing the tape width will boost the integration over positive electric field to cancel the integration over negative electric field that essentially forms V_{oc}. Lastly, we successfully proved the feasibility of constructing statistic models as a substitution to FEM models by utilizing a GPR regression model to reproduce V_{oc} characteristic curves for each parameter, which has an average accuracy as high as 97%. Besides, the ANOVA results for our Taguchi tables have revealed the significance rank as: $B_r > g > W_s$, while the interaction effects have also been investigated. The authors believe this work can provide more insight into research and industrial applications for HTS flux pumps, because these machine learning based models can be powerful design tools to confidently estimate key characteristics of the investigated devices almost instantaneously, and hence avoid excessive reliance on complex numerical models, allowing rapid design.

Data availability statement

The data that support the findings of this study are available upon reasonable request from the authors.

Acknowledgments

ORCID iDs

Zezhao Wen ⬤ https://orcid.org/0000-0001-6498-3438
Hongye Zhang ⬤ https://orcid.org/0000-0002-8960-4614

References

[1] Abrahamsen A B, Mijatovic N, Seiler E, Zirngibl T, Træholt C, Nørgård P B, Pedersen N F, Andersen N H and Østergård J 2010 Superconducting wind turbine generators *Supercond. Sci. Technol.* **23** 034019

[2] Sivasubramaniam K, Zhang T, Lokhandwalla M, Laskaris E T, Bray J W, Gerstler B, Shah M R and Alexander J P 2009 Development of a high speed HTS generator for airborne applications *IEEE Trans. Appl. Supercond.* **19** 1656–61

[3] Miki M *et al* 2006 Development of a synchronous motor with Gd–Ba–Cu–O bulk superconductors as pole-field magnets for propulsion system *Supercond. Sci. Technol.* **19** S494

[4] Umemoto K *et al* 2010 Development of 1 MW-class HTS motor for podded ship propulsion system **234** 032060

[5] Nick W, Frank M, Klaus G, Frauenhofer J and Neumuller H-W 2007 Operational experience with the world's first 3600 rpm 4 MVA generator at Siemens *IEEE Trans. Appl. Supercond.* **17** 2030–3

[6] Sung H-J, Kim G-H, Kim K, Jung S-J, Park M, Yu I-K, Kim Y-G, Lee H and Kim A-R 2013 Practical design of a 10 MW superconducting wind power generator considering weight issue *IEEE Trans. Appl. Supercond.* **23** 5201805

[7] Karmaker H *et al* Direct Drive HTS Wind Generator Design for Commercial Applications pp 491–5

[8] Sung H-J, Kim G-H, Kim K, Park M, Yu I-K and Kim J-Y 2013 Design and comparative analysis of 10 MW class superconducting wind power generators according to different types of superconducting wires *Physica* C **494** 255–61

[9] Neumüller H, Nick W, Wacker B, Frank M, Nerowski G, Frauenhofer J, Rzadki W and Hartig R 2006 Advances in and prospects for development of high-temperature superconductor rotating machines at Siemens *Supercond. Sci. Technol.* **19** S114

[10] Kalsi S S 2011 *Applications of High Temperature Superconductors to Electric Power Equipment* (New York: Wiley)

[11] Barnes P N, Sumption M D and Rhoads G L 2005 Review of high power density superconducting generators: present state and prospects for incorporating YBCO windings *Cryogenics* **45** 670–86

[12] Bumby C W, Badcock R A, Sung H-J, Kim K-M, Jiang Z, Pantoja A E, Bernardo P, Park M and Buckley R G 2016 Development of a brushless HTS exciter for a 10 kW HTS synchronous generator *Supercond. Sci. Technol.* **29** 024008

[13] Sung H *et al* 2016 Design and heat load analysis of a 12 MW HTS wind power generator module employing a brushless HTS exciter *IEEE Trans. Appl. Supercond.* **26** 1–4

[14] Jeon H *et al* 2018 Methods for increasing the saturation current and charging speed of a rotary HTS flux-pump to charge the field coil of a synchronous motor *IEEE Trans. Appl. Supercond.* **28** 1–5

[15] Coombs T 2019 Superconducting flux pumps *J. Appl. Phys.* **125** 230902

[16] Zhai Y, Tan Z, Liu X, Shen B, Coombs T A and Wang F 2020 Research progress of contactless magnetization technology: HTS flux pumps *IEEE Trans. Appl. Supercond.* **30** 4602905

[17] Hoffmann C, Pooke D and Caplin A D 2010 Flux pump for HTS magnets *IEEE Trans. Appl. Supercond.* **21** 1628–31

[18] Jiang Z, Bumby C W, Badcock R A, Sung H-J, Long N J and Amemiya N 2015 Impact of flux gap upon dynamic resistance of a rotating HTS flux pump *Supercond. Sci. Technol.* **28** 115008

[19] Bumby C W, Phang S, Pantoja A E, Jiang Z, Storey J G, Sung H-J, Park M and Badcock R A 2016 Frequency dependent behavior of a dynamo-type HTS flux pump *IEEE Trans. Appl. Supercond.* **27** 1–5

[20] Pantoja A E, Jiang Z, Badcock R A and Bumby C W 2016 Impact of stator wire width on output of a dynamo-type HTS flux pump *IEEE Trans. Appl. Supercond.* **26** 4805208

[21] Badcock R A, Phang S, Pantoja A E, Jiang Z, Storey J G, Sung H-J, Park M and Bumby C W 2016 Impact of magnet geometry on output of a dynamo-type HTS flux pump *IEEE Trans. Appl. Supercond.* **27** 5200905

[22] Ghabeli A and Pardo E 2020 Modeling of airgap influence on DC voltage generation in a dynamo-type flux pump *Supercond. Sci. Technol.* **33** 035008

[23] Ghabeli A, Pardo E and Kapolka M 2021 3D modeling of a superconducting dynamo-type flux pump *Sci. Rep.* **11** 1–12

[24] Ainslie M D, Quéval L, Mataira R C and Bumby C W 2021 Modelling the frequency dependence of the open-circuit voltage of a high-T_c superconducting dynamo *IEEE Trans. Appl. Supercond.* **31** 4900407

[25] Mataira R, Ainslie M, Badcock R and Bumby C W 2019 Origin of the DC output voltage from a high-T_c superconducting dynamo *Appl. Phys. Lett.* **114** 162601

[26] Mataira R, Ainslie M D, Badcock R and Bumby C W 2020 Modeling of stator versus magnet width effects in high-T_c superconducting dynamos *IEEE Trans. Appl. Supercond.* **30** 5204406

[27] Mataira R, Ainslie M, Pantoja A, Badcock R and Bumby C 2020 Mechanism of the high-T_c superconducting dynamo: models and experiment *Phys. Rev. Appl.* **14** 024012

[28] Ainslie M, Grilli F, Quéval L, Pardo E, Perez-Mendez F, Mataira R, Morandi A, Ghabeli A, Bumby C and Brambilla R 2020 A new benchmark problem for electromagnetic modelling of superconductors: the high-T_c superconducting dynamo *Supercond. Sci. Technol.* **33** 105009

[29] Bumby C, Jiang Z, Storey J, Pantoja A E and Badcock R A 2016 Anomalous open-circuit voltage from a high-T_c superconducting dynamo *Appl. Phys. Lett.* **108** 122601

[30] Brambilla R, Grilli F, Martini L, Bocchi M and Angeli G 2018 A finite-element method framework for modeling rotating machines with superconducting windings *IEEE Trans. Appl. Supercond.* **28** 5207511

[31] Quéval L, Liu K, Yang W, Zermeño V M R and Ma G 2018 Superconducting magnetic bearings simulation using an H-formulation finite element model *Supercond. Sci. Technol.* **31** 084001

[32] Pardo E, Šouc J and Frolek L 2015 Electromagnetic modelling of superconductors with a smooth current–voltage relation: variational principle and coils from a few turns to large magnets *Supercond. Sci. Technol.* **28** 044003

[33] Pardo E and Kapolka M 2017 3D computation of non-linear eddy currents: variational method and superconducting cubic bulk *J. Comput. Phys.* **344** 339–63

[34] Zhang H, Zhang M and Yuan W 2016 An efficient 3D finite element method model based on the *T–A* formulation for superconducting coated conductors *Supercond. Sci. Technol.* **30** 024005

[35] Benkel T, Lao M, Liu Y, Pardo E, Wolfstadter S, Reis T and Grilli F 2020 *T–A*-formulation to model electrical machines with HTS coated conductor coils *IEEE Trans. Appl. Supercond.* **30** 5205807

[36] Brambilla R, Grilli F, Nguyen D N, Martini L and Sirois F 2009 AC losses in thin superconductors: the integral equation method applied to stacks and windings *Supercond. Sci. Technol.* **22** 075018

[37] Morandi A and Fabbri M 2014 A unified approach to the power law and the critical state modeling of superconductors in 2D *Supercond. Sci. Technol.* **28** 024004

[38] Grilli F, Pardo E, Stenvall A, Nguyen D N, Yuan W and Gomory F 2013 Computation of losses in HTS under the action of varying magnetic fields and currents *IEEE Trans. Appl. Supercond.* **24** 78–110

[39] Pedregosa F *et al* 2011 Scikit-learn: machine learning in Python *J. Mach. Learn. Res.* **12** 2825–30

[40] Swiler L P *et al* 2020 A survey of constrained Gaussian process regression: approaches and implementation challenges *J. Mach. Learn. Model. Comput.* **1** 119–56

[41] Smola A J and Schölkopf B 2004 A tutorial on support vector regression *Stat. Comput.* **14** 199–222

[42] Botchkarev A 2018 Performance metrics (error measures) in machine learning regression, forecasting and prognostics: properties and typology (arXiv:1809.03006 [stat.ME])

[43] Gao X, Asami Y and Chung C-J F 2006 An empirical evaluation of spatial regression models *Comput. Geosci.* **32** 1040–51

[44] Jiang Z, Hamilton K, Amemiya N, Badcock R A and Bumby C W 2014 Dynamic resistance of a high-T_c superconducting flux pump *Appl. Phys. Lett.* **105** 112601

[45] Zhang H *et al* 2020 Demarcation currents and corner field for dynamic resistance of HTS-coated conductors *IEEE Trans. Appl. Supercond.* **30** 6601305

[46] Zhang H, Chen H, Jiang Z, Yang T, Xin Y, Mueller M and Li Q 2020 A full-range formulation for dynamic loss of high-temperature superconductor coated conductors *Supercond. Sci. Technol.* **33** 05LT01

[47] Zhang H, Yao M, Jiang Z, Xin Y and Li Q 2019 Dependence of dynamic loss on critical current and *n*-value of HTS coated conductors *IEEE Trans. Appl. Supercond.* **29** 1–7

[48] Zhang H, Yao M, Kails K, Machura P, Mueller M, Jiang Z, Xin Y and Li Q 2020 Modelling of electromagnetic loss in HTS coated conductors over a wide frequency band *Supercond. Sci. Technol.* **33** 025004

[49] Zhang H, Machura P, Kails K, Chen H and Mueller M 2020 Dynamic loss and magnetization loss of HTS coated conductors, stacks, and coils for high-speed synchronous machines *Supercond. Sci. Technol.* **33** 084008

[50] Yazdani-Asrami M, Taghipour-Gorjikolaie M, Song W, Zhang M and Yuan W 2020 Prediction of nonsinusoidal AC loss of superconducting tapes using artificial Intelligence-based models *IEEE Access* **8** 207287–97

[51] Leclerc J, Hell L M, Lorin C and Masson P J 2016 Artificial neural networks for AC losses prediction in superconducting round filaments *Supercond. Sci. Technol.* **29** 065008

[52] Zhang Z *et al* 2018 AC loss prediction model of a 150 kJ HTS SMES based on multi-scale model and artificial neural networks *IEEE Trans. Magn.* **54** 7206005

[53] Pan H and Shen G X 2003 Application of artificial neural network methods in HTS RF coil design for MRI *Concepts Magn. Reson.* B **18** 9–14

[54] Richard L S, Bonifetto R, Carli S, Froio A, Foussat A and Zanino R 2014 Artificial neural network (ANN) modeling of the pulsed heat load during ITER CS magnet operation *Cryogenics* **63** 231–40

[55] Karsoliya S 2012 Approximating number of hidden layer neurons in multiple hidden layer BPNN architecture *Int. J. Eng. Trends Technol.* **3** 714–7

[56] Berry M J and Linoff G S 2004 *Data Mining Techniques: For Marketing, Sales, and Customer Relationship Management* (New York: Wiley)

[57] Rasmussen C E and Williams C K 2006 *Gaussian Processes for Machine Learning Cambridge [Google Scholar]* (MA: MIT Press)

[58] Coombs T, Fagnard J-F and Matsuda K 2014 Magnetization of 2G coils and artificial bulks *IEEE Trans. Appl. Supercond.* **24** 8201005

[59] Ghabeli A *et al* 2021 Modeling the charging process of a coil by an HTS dynamo-type flux pump (arXiv:2105.00510 [physics.app-ph])