Abstract: A simple, reliable, and accurate surrogate modelling technique for microwave structures is presented. A hierarchical surrogate model is constructed using response surface approximations (RSAs) of suitable response features with embedded ‘visible’ knowledge. The authors describe and discuss a novel scanning algorithm for capturing these desired response features. They demonstrate that the dependence of the selected features on the designable parameters is much less non-linear than that of the original responses taken as functions of the design parameters. They illustrate the steps of their algorithm using novel diagrams. They simplify the description of their modelling process by exploiting an operator notation. They discuss relationships between their method and other feature-based approaches such as shape-preserving response prediction (SPRP). They provide a wideband microstrip bandstop filter, a fourth-order ring resonator bandpass filter and a microstrip bandpass filter with open stub inverter examples to demonstrate their approach. Using these examples, they compare their approach with two very different direct RSA methods (kriging and radial basis function) and an SPRP method. The models of all the filter examples are validated using design optimisation.

1 Introduction

Accurate and computationally cheap models are fundamental components of the contemporary microwave design process. Reliable performance evaluation of microwave components and structures can be obtained by means of full-wave electromagnetic (EM) analysis at fine discretisation. As a matter of fact, in many situations, EM simulation might be the only way to account for various effects such as couplings between components or interactions between the structure itself and its environment (e.g. antenna housing and/or installation fixtures). This is particularly true for miniaturised structures (e.g. [1, 2]) where traditional equivalent circuit modelling becomes inadequate.

Despite its benefits, EM simulation is computationally expensive, especially for complex structures. As a result, EM-driven design is usually a challenging process. In particular, direct use of high-fidelity EM models may be prohibitive when multiple structure evaluation is required, for example, to perform parametric optimisation, statistical analysis, or tolerance-aware design.

There are two basic ways of creating fast replacement models (surrogates): (i) response surface approximation (RSA) of samples high-fidelity model data [3] and (ii) physics-based surrogate modelling [4]. Numerous techniques for RSA modelling are available, including polynomial regression [3, 6], radial-basis function interpolation [7], kriging [8, 9], support vector regression [10, 11], fuzzy systems [12–14], multi-dimensional Cauchy approximation [15], and neural networks [14, 16–18] which is probably the most popular RSA approach in microwave engineering. The most important advantage of RSA models is their speed. The fundamental drawback is poor scalability: the number of training samples necessary to ensure usable accuracy grows rapidly with the design space dimensionality and size (in terms of parameter ranges). Acquisition of hundreds and thousands of samples may be justifiable for creating multiple-use library models but it is normally a too expensive option for solving one-time design problems such as parametric optimisation.

Physics-based surrogates offer a partial solution to this problem by exploiting cheaper representations of the structure of interest, referred to as low-fidelity or coarse models (e.g. equivalent circuits or coarse-mesh EM simulations). To create a surrogate, the low-fidelity model is corrected using a limited number of high-fidelity training points [19]. Perhaps the most popular technique of this kind in microwave engineering is space mapping (SM) [14, 20, 21]. One of the issues of physics-based modelling is the rather complex implementation (additional model and implementation of interactions between models of different fidelity are required). Moreover, the applicability of physics-based modelling is limited to cases when fast low-fidelity models are available (e.g. filters).

A shape-preserving response prediction (SPRP) technique introduced in [22], as well as its later developments (e.g. generalised SPRP (GSPRP) [23]) alleviate – to some extent – the drawbacks of RSA-based modelling, by conducting the modelling process in the domain of suitably selected response features [23]. SPRP is a physics-based modelling technique so that it requires a fast underlying low-fidelity model. On the other hand, GSPRP is a one-level approach so formally it falls into the category of RSA surrogates.

Here, we discuss a feature-based modelling technique that is based on modelling of suitably selected ‘visible’ response features (e.g. 3 dB bandwidth, frequency location of −20 dB response level, local ‘peaks’, and/or ‘valleys’) rather than the responses (e.g. S-parameters versus frequency) themselves. Using conventional design space sampling, RSA models are constructed for extracted features using kriging interpolation. The surrogate model response is then calculated by interpolating predicted features (equivalent to characteristic points of the SPRP technique [23]) obtained through evaluation of these RSA models. Our methodology exploits the fact that the coordinates (both level and frequency) of the features are much less non-linearly dependent on the designable parameters of the structure of interest than the original responses. This results in considerable computational savings as compared with conventional RSA modelling.
The proposed methodology can be considered a hierarchical (multilevel) surrogate modelling technique (e.g. [24, 25]). Our hierarchy has two levels: approximation of frequency responses and approximation of features.

An initial version was introduced in [26]. In this paper, we use an operator notation [27] to simplify the description of the process of extracting features and restoring responses. We use new diagrams to illustrate the concepts of our method and to explain the steps of our approach. We provide the details of a scanning algorithm to find the feature points. We provide a new comprehensive numerical validation based on each of the three examples of microstrip filters. We add new comparisons to show our method with a similar SPRP [22] and GSPRP [23] approach in theory and in numerical examples. Comparisons indicate that our approach is similar to or better than GSPRP [23] in performance while being much simpler to implement. Application to design optimisation of filter structures is also demonstrated. We discuss and compare our approach with traditional RSA approaches (kriging and radial basis function (RBF) interpolation).

2 Surrogate modelling using response features

In this section, we define a surrogate modelling problem and formulate the feature-based modelling technique. We also provide an intuitive explanation of its efficiency. Numerical verification using specific filter examples will be provided in Section 3.

2.1 Surrogate modelling problem

Let \( R_{\omega} X \rightarrow R^m X \subseteq R^m \) denote the response vector of the microwave device of interest. For example, \( R_{\omega}(x) \) may represent \( S_{\omega} \) at \( m \) chosen frequencies, \( \omega_a \) to \( \omega_b \), that is, \( R_{\omega}(x) = [R_{\omega}(x, \omega_1), \ldots, R_{\omega}(x, \omega_m)]^T \). \( R_{\omega} \) is assumed to be evaluated using high-fidelity EM simulation and, therefore, computationally expensive. The task is to build a fast surrogate model \( R_{\tilde{\omega}} \) that represents \( R_{\omega} \) in \( X \) (compare Fig. 1).

Let \( X = \{x^1, x^2, \ldots, x^K\} \subseteq X \) be the training set so that the responses of the high-fidelity model \( R_{\omega}(x^k) \) at \( x^k \) are known. Conventional response surface modelling attempts to directly model \( R_{\omega}(x, \omega_j), j = 1, \ldots, m \). In many cases, the surrogate is created as an ensemble of RSA models constructed for individual frequencies, that is, obtained by approximating the data sets \( \{x^k, R_{\omega}(x^k, \omega_j)\}_{k=1, \ldots, K} \) for \( j = 1, \ldots, m \). Sometimes [9], frequency is treated as an additional designable parameter, so that the RSA surrogate is constructed by approximating the data pairs \( \{(x^k)^T, \omega_j\}, R_{\omega}(x^k, \omega_j)\}_{k=1, \ldots, K, \omega_j(1, \ldots, m)} \).

2.2 Response features

Construction of RSA models in a conventional way, for example, for individual frequencies as explained in Section 2.1 is a challenging task, particularly in the case of filter structures. The reason is the non-linearity of the responses of interest, usually \( S \)-parameters. Fig. 2a shows the family of responses of a microstrip bandpass filter (considered in Section 3) evaluated along a selected line segment in the design space \((1-t)x_a^k + tx_b^k\), where \( 0 \leq t \leq 1 \) \((x_a^k \text{ and } x_b^k \text{ are the two arbitrarily selected points in the design space})\). The circles mark the responses of design \( x^k \) at selected frequencies 1.9, 2, and 2.1 GHz. Now we plot responses with respect to \((w.r.t.) \ t \) (i.e. design value changes from \( x_a^k \) to \( x_b^k \)).

Fig. 2b shows the non-linear characteristics of \( \{R_{\omega}(1-t)x^k + tx^m, \omega_j\} \), at the selected frequencies, 1.9, 2, and 2.1 GHz. Adequate modelling of such non-linear responses across a multi-dimensional design space requires a large quantity of training data.

The modelling technique presented here is based on approximating sets of suitably selected response features (compare [22]) rather than directly working with the original responses. The features may include points corresponding to specific response levels (e.g. \(-20, -10, \text{ and } -3 \text{ dB}\)), as well as those allocated in between fixed-level points (e.g. uniformly on the frequency scale).

Fig. 3 shows a few responses of the bandpass filter used to create the plots in Fig. 2, with the selected feature points marked as circles. We will use the notation \( f_k^i = [\omega_j^i, l_j^i] \), \( j = 1, \ldots, K \), and \( k = 1, \ldots, l \).
Fig. 3 Selected feature point progress plots between two designs $x^a (t = 0)$ and $x^b (t = 1)$. The lines correspond to selected responses for various values of parameter $t$ between 0 and 1. Three selected groups of feature points (corresponding to $-20$ dB level on the left-hand side of the passband, right-hand side $-3$ dB frequency; and centre frequency of the filter) are marked with circles. {A, B, C} are $-20$ dB level feature points corresponding to $t = 0, 0.5$, and $1$; {D, E, F} are centre frequency feature points corresponding to $t = 0, 0.5$, and $1$; {G, H, I} are the respective $-3$ dB level feature points corresponding to $t = 0, 0.5$, and $1$. Selected groups of corresponding features are marked with dotted-line ellipses.

Fig. 4 Selected feature point progress plots between designs $x^a (t = 0)$ and $x^b (t = 1)$. Levels. They correspond to three feature points: $-20$ dB level on the left-hand side of the passband (continuous lines), right-hand side $-3$ dB frequency (dashed dotted lines), and centre frequency of the filter (dotted-line ellipses). Letters A, B, …, G mark the feature points corresponding to those in Fig. 3.

2.3 Construction of feature-based surrogate model

The feature-based surrogate is formulated using the operator notation [27]. We use the following two operators: (i) $C(\cdot)$ that extracts $K$ feature points from frequency responses of the structure of interest and (ii) $F(\cdot)$ that carries out the inverse process, that is, it converts the feature points back to frequency responses, by means of appropriate interpolation schemes. In practice, the operator $C$ is implemented through ‘scanning’ of the frequency response of the filter of interest using an auxiliary algorithm (here, implemented in MATLAB, see Section 2.4). The operator $F$ uses spline-based interpolation (also implemented in the MATLAB built-in function interp1). Using the operator notation, our feature-based modelling procedure can be described as follows:

Step 1: Create feature points using the operator $C$ for all training designs $x^i$, $i = 1, \ldots, N$

$$\{f^1_i, \ldots, f^K_i\} = C[R(x^i), \omega]$$

where $\omega = [\omega_1 \ldots \omega_m]$ are the frequency samples corresponding to $R_i$. Step 2: Construct kriging interpolation models $R_{ij}^f = 1, \ldots, K_i$ of the feature points using the features $(f^k_i)$, $i = 1, \ldots, N$, corresponding to all training designs $x^i$. The model $R_{ij}^f$ is used to predict the feature point $f^j(x)$ at the evaluation design $x$ as follows

$$f^j(x) = R_{ij}^f(x)$$

Step 3: The feature-based surrogate model (in the form of the corresponding frequency response) is then defined as

$$R(x) = F(\{f^1(x), \ldots, f^K(x)\})$$

$$= F(\{R_{ij}^f(x), R_{ij}^f(x), \ldots, R_{ij}^f(x)\})$$

Fig. 5 shows conceptual illustrations of the feature-based surrogate modelling process. Steps 1 and 2 correspond to surrogate model construction; they are performed only once for a given modelling problem. Step 3 corresponds to the use of the surrogate model and is performed for each model evaluation.

2.4 Feature scanning algorithm

The feature point allocation is a two-step process: (i) identify the core feature points (e.g. those corresponding to critical levels such as $-3$ dB and/or $-20$ dB, as well as response local minima/maxima, if relevant), (ii) assign additional points, usually allocated uniformly (either in frequency or level wise) between the core features. We define all possible core features in a set $\{f_1, \ldots, f_N\}$ through a visual inspection of the responses. We use the following pseudocode to scan and find the features for all the responses $R(x^k)$, $k = 1, \ldots, N$ (see Fig. 6).

It is important to note that the number of feature points should be sufficient so that the interpolation operation $F$ has enough data for accurate response prediction. It is also important that the feature points corresponding to any two training designs are in one-to-one correspondence, and that the number of feature points is the same for all training designs. It should be mentioned, however, that if a one-to-one correspondence between the ‘visible’ core feature points throughout the training set is not achievable, the entire concept would still formally work (in which case, the corresponding core feature points are omitted from all training designs while constructing the surrogate model). This is not the case for other feature-based methods such as SPRP [22] or GSPRP [23], where the characteristic point correspondence is of fundamental importance. In [23], some ways have been described of defining the feature points (or ‘characteristic’ points in the SPRP context) to work around the possible lack of their one-to-one correspondence throughout the design space.

2.5 Relation to SPRP and GSPRP

The concept of feature points was originally introduced in the SPRP technique [22] (there, referred to as characteristic points). It should be emphasised, however, that the use of feature points is the only common element of feature-based modelling and SPRP. The fundamental components of SPRP are translation vectors that describe the change...
of the characteristic points w.r.t. a so-called reference design (normally, the training data point closest to a design of interest $x$).

This is the basis for making predictions about the high-fidelity model response at $x$, both for the original SPRP [22] as well as the GSPRP [23]. While GSPRP allows an arbitrary training set, it still requires a reference design $x^{(i)}$. The technique proposed here does not use translation vectors and does not need reference designs. As a result, it is substantially simpler to implement than SPRP [22, 23].

No requirement of reference designs is actually the biggest advantage of the feature-based modelling compared with SPRP/GSPRP, particularly from the point of view of further design cost reduction. More specifically, feature-based modelling can be realised in a multi-fidelity framework with the surrogate constructed primarily from the low-fidelity EM simulation data and sparse high-fidelity EM data sets utilised for model correction (using, e.g. SM [19] or co-kriging [29]). The details exceed the scope of this paper and will be treated in a separate paper.

### 3 Verification examples

In this section, we provide a comprehensive numerical validation of the feature-based modelling technique. We consider three examples of microstrip filters and compare the performance of our approach with direct RSA modelling of the high-fidelity model as well as with GSPRP [23]. Applications of feature-based surrogates for filter optimisation are discussed in Section 4.

#### 3.1 Filter structures

Our first example (Filter 1) is the wideband microstrip bandstop filter [30] shown in Fig. 7a. The filter is simulated in FEKO [31] using 642 triangular meshes (simulation time: 5 s per frequency on 2.53 GHz Intel Xeon central processing unit with 6 GB random access memory). The substrate parameters are thickness $h = 0.508$ mm and permittivity $\varepsilon_r = 3.38$. The designable parameters are $x = [L_r \ W_r \ L_c \ W_c \ G_c]_T$. The region of interest is defined as an interval $[x_0 - d, x_0 + d]$ with $x_0 = [7.0 \ 1.2 \ 9.0 \ 0.075 \ 0.10]_T$ and $d = [1.0 \ 0.4 \ 1.0 \ 0.025 \ 0.05]_T$. The frequency range is 1–11 GHz with 100 MHz step size.

The second structure (Filter 2) is the fourth-order ring resonator bandpass filter [32] shown in Fig. 7b. The filter is simulated in FEKO [31] using 952 triangular meshes (simulation time: 8 s per frequency). The substrate parameters are thickness $h = 1.52$ mm and permittivity $\varepsilon_r = 4.32$. The designable parameters are $x = [L_1 \ L_2 \ L_3 \ S_1 \ S_2 \ W_1 \ W_2]_T$. The region of interest is defined as an interval $[x_0 - d, x_0 + d]$ with $x_0 = [24.0 \ 20.0 \ 25.0 \ 0.2 \ 0.2 \ 1.0 \ 0.5]_T$ and $d = [1.0 \ 1.0 \ 1.0 \ 0.1 \ 0.1 \ 0.2 \ 0.2]_T$. The frequency range is 1–3 GHz with 25 MHz step size.
The last structure considered here (Filter 3) is the microstrip bandpass filter with open stub inverter [33] shown in Fig. 7c. The filter is simulated in FEKO [31] using 432 triangular meshes (simulation time: 2 s per frequency). The substrate parameters are thickness $h = 0.508$ mm and permittivity $\epsilon_r = 2.6$. The designable parameters are $x = \{L_1, L_2, L_3, S_1, S_2, W_1, W_2\}^T$. The region of interest is defined as an interval $[x^0 - d, x^0 + d]$ with $x^0 = [24.0 100 2.0 0.6 0.2 0.5]^T$ and $d = [2.0 2.0 1.0 0.4 0.1 0.4]^T$. The frequency range is 1.5–2.5 GHz with 12.5 MHz step size.

Filters 2 and 3 are bandpass filters. Therefore, for these cases, the feature points are selected to capture the (frequency) location of the passband. In particular, we have: (i) two points at the $-10$ dB levels (i.e. determining the $-10$ dB passband; we do not use the $-3$ dB feature because, for certain training points, the $-3$ dB level might not be achieved at all), (ii) additional points uniformly distributed in frequency and placed between the $-10$ dB features (20–40 points depending on the example), and (iii) additional points at $-12$, $-14$, $-16$ dB etc., allocated on both sides of the $-10$ dB passband.

The particular choice of the $-10$ dB level is, of course, not critical (e.g. $-11$ dB will work equally well, as long as it exists in all the responses). The same principle applies to the choice of the number and levels of the additional points. For Filter 1 (bandstop filter), the feature points are selected at $-20$ dB level, 50 points in between these two (allocated uniformly in frequency), the two local maxima (close to the edges of the stopband), and additional points between the maxima and the $-20$ dB level points.

### 3.2 Experimental setup and results

The model accuracy is verified using the relative error measure $\frac{||R(x) - R(x)||}{||R(x)||}$ expressed in per cent and averaged over 100 random test designs. The feature-based model is compared with direct kriging interpolation of high-fidelity data, direct RBF interpolation of $R$ data, as well as GSPRP [23]. The kriging model utilises a Gaussian correlation function [34], whereas the RBF model uses Gaussian basis functions [5]. The length-scale parameter of the latter is optimised using cross-validation [5]. For all four modelling methods, five different cases are considered with the number of training points varying between $N = 20$ and 400. The results are gathered in Table 1.

It can be observed that for all considered problems, the accuracy of the feature-based surrogate is comparable or better than GSPRP and better than the accuracy of both the kriging and RBF surrogates for the corresponding number of training points. Fig. 8 shows the high-fidelity and feature-based model responses at selected test points for Filters 1–3. It should also be noted that feature-based model (similarly as GSPRP) ensures good accuracy even for a small number of training samples ($N = 50$ or 100), therefore, ‘low cost’, which is not the case for the ‘direct’ RSA approaches. Our examples also indicate that our approach is capable of modelling components of various types (bandstop, bandpass, narrowband, wideband etc.), as long as the corresponding feature sets are consistent across the design space.

**Table 1 Modelling results for Filters 1–3**

<table>
<thead>
<tr>
<th>Filter</th>
<th>Model setup cost, min*</th>
<th>Modelling method</th>
<th>Average error</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$N = 20, %$</td>
<td>$N = 50, %$</td>
<td>$N = 100, %$</td>
</tr>
<tr>
<td>1</td>
<td>8-N</td>
<td>Feature-basedb</td>
<td>9.1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>GSPRP</td>
<td>8.5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Direct krigingc</td>
<td>16.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Direct RBFsd</td>
<td>18.9</td>
</tr>
<tr>
<td>2</td>
<td>11-N</td>
<td>Feature-basedb</td>
<td>11.6</td>
</tr>
<tr>
<td></td>
<td></td>
<td>GSPRP</td>
<td>8.8</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Direct krigingc</td>
<td>13.5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Direct RBFsd</td>
<td>21.2</td>
</tr>
<tr>
<td>3</td>
<td>3-N</td>
<td>Feature-basedb</td>
<td>7.5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>GSPRP</td>
<td>7.7</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Direct krigingc</td>
<td>11.6</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Direct RBFsd</td>
<td>16.7</td>
</tr>
</tbody>
</table>

*a$N$ stands for the number of training points.

*bFeature-based modelling using the procedure proposed in Section 2.

*cDirect kriging interpolation of high-fidelity model ($I_{\Sigma_1}$) responses.

*dDirect interpolation of high-fidelity model ($I_{\Sigma_1}$) responses using RBFs.

*eCost reported in this table refers to training data acquisition. The cost of surrogate model identification can be neglected.
Fig. 8  Fine (continuous lines) and feature-based model (1)-(4) responses (o) obtained for 100 base points at the selected test designs

(a) Filter 1  
(b) Filter 2  
(c) Filter 3

Fig. 9  Feature-based surrogate model responses at the initial design (dashed lines), at the optimised design (continuous lines), and the corresponding verification high-fidelity model responses (o)

(a) Filter 1  
(b) Filter 2  
(c) Filter 3
3.3 Application examples: filter optimisation

For the sake of illustration, the feature-based model of Section 2 was used to carry out parametric optimisation of the filter structures considered in Section 3.1. The design specifications are as follows:

- Filter 1: $|S_{12}| \leq -20$ dB for $3.0$ GHz $\leq \omega \leq 9.0$ GHz, $|S_{21}| \geq -3$ dB for $\omega \leq 2$ GHz, and $\omega \geq 10$ GHz.
- Filter 2: $|S_{12}| \geq -1$ dB for $1.75$ GHz $\leq \omega \leq 2.25$ GHz, $|S_{21}| \leq -20$ dB for $\omega \leq 1.5$ GHz, and $\omega \geq 2.5$ GHz.
- Filter 3: $|S_{12}| \leq -1$ dB for $1.95$ GHz $\leq \omega \leq 2.05$ GHz, $|S_{21}| \leq -20$ dB for $\omega \leq 1.8$ GHz, and $\omega \geq 2.2$ GHz.

For all three cases, the feature-based surrogate created with $N = 200$ training samples was considered. Fig. 9 shows the initial and final filter responses obtained by optimising the feature-based surrogate model (the final design is verified by the high-fidelity model). Design specifications are marked using thick horizontal lines. Owing to the very good accuracy of the surrogates, no further design tuning is necessary. In all cases, the surrogate models were optimised using a sequential quadratic programming algorithm (implemented in MATLAB’s `fmincon` routine).

4 Conclusions

In this paper, a simple, reliable, and low-cost technique for modelling of microwave structures has been introduced. Our notation and suitable illustrations simplify our presentation. Frequency responses are converted to so-called feature points through exploiting visible knowledge. By means of RSAs of these feature points, our approach allows us to construct accurate surrogates at low cost (a fraction of the cost required by direct RSA methods such as kriging and RBF interpolation). The implementation of our method is more straightforward than that of GSPRF, while maintaining similar or better accuracy. The effectiveness of our new methodology is demonstrated by modelling and design optimisation of three filters, including comparisons with the results from other modelling techniques. Future work will focus on extending feature-based modelling to a variable-fidelity framework, aiming at further reduction of the surrogate model setup cost without compromising its predictive power. Moreover, the method will be applied for modelling of other types of structures such as narrow-band antennas etc.

5 Acknowledgment

This work was supported in part by the Icelandic Centre for Research (RANNIS) grants 130450051 and 14127051, the National Natural Science Foundation of China grant 61417258, by the Natural Sciences and Engineering Research Council of Canada under grants RGPIN7239-11 and STGPGP447367-13, and by Bandler Corporation.

6 References