SUPPLEMENTARY MATERIALS: Varying coefficient models and design choice for Bayes linear emulation of complex computer models with limited model evaluations *

Amy L. Wilson †, Michael Goldstein ‡, and Chris J. Dent*

SM1. Varying coefficient emulator: illustrative examples. In this section, the varying coefficient emulator (3.1) is compared to a fixed coefficient emulator for a series of illustrative examples. In the first two examples a simple function with just one input parameter is emulated. The third (more detailed) example uses a more complicated function with multiple inputs that is likely to be more representative of the types of functions encountered in practice.

SM1.1. Example 1 - a linear function. In this example, data are generated from the following linear function:

$$f(x) = \alpha_0 + \alpha_1 x, \tag{SM1.1}$$

where $x \in [-0.5, 0.5]$, $\alpha_0 = 0.3$ and $\alpha_1 = 0.5 + 0.3x$ so that α_1 varies with x. An emulator of form (3.1) is fitted to a set of 5 evaluations of the function f, where the inputs at which the function is evaluated are chosen using a maximin Latin hypercube design ([SM1], [SM4]).

The basis function in (3.1) is set to $\mathbf{h}(x) = (1, x)^T$ and β_1 and β_2 are chosen to be varying coefficients. The prior judgment for σ_1 was set to 0.02, corresponding approximately to the residual standard error of a standard regression fit to a design of size 5 and the prior judgment of σ_2 was set to 0.3. The correlation matrix $c_j(x, x')$ was set to the Gaussian correlation function given in (3.2) with $\delta_1 = 0.3$ and $\delta_2 = 1$ (here the subscript i has been dropped as the output is one-dimensional). Thus, the coefficient of x is a priori assumed to vary slowly throughout the space, whereas $\epsilon_{\beta_1}(x)$ is assumed to be more locally variable.

For the remainder of the prior assumptions, we set: $\mathbb{E}[\beta_j] = 0$; $\mathrm{Cov}[\beta_j, \beta_k] = 0$ for $j \neq k$ and $\mathrm{Cov}[\beta_j, \beta_k] = 1$ for $j = k$.

As $x \in [-0.5, 0.5]$, the prior variance of $f(x)$, given by $k(x, x)$ in (3.7), will be smaller for x closer to zero and larger as the absolute value of x increases. As in this case zero is in the centre of the range of x and a space-filling design has been used, this assumption is justified as we would expect the fit of the emulator to be better in the centre of the space and for the variance to increase the further away we move from the design points. In general, it may be the case that expert modellers are more confident in subjective judgments made in central regions of the input space as the edges of the input space may be less well explored.

For comparison, the fixed coefficient model described in (1.1) is also used to emulate the function (SM1.1) so that comparisons can be drawn between the two models. For this

*Submitted to the editors.

Funding: This work was supported by EPSRC grants EP/K03832X/1 and EP/K036211/1.

†School of Mathematics, University of Edinburgh, Edinburgh, UK (Amy.L.Wilson@ed.ac.uk, Chris.Dent@ed.ac.uk).

‡Department of Mathematical Sciences, Durham University, Durham, UK (Michael.Goldstein@durham.ac.uk).

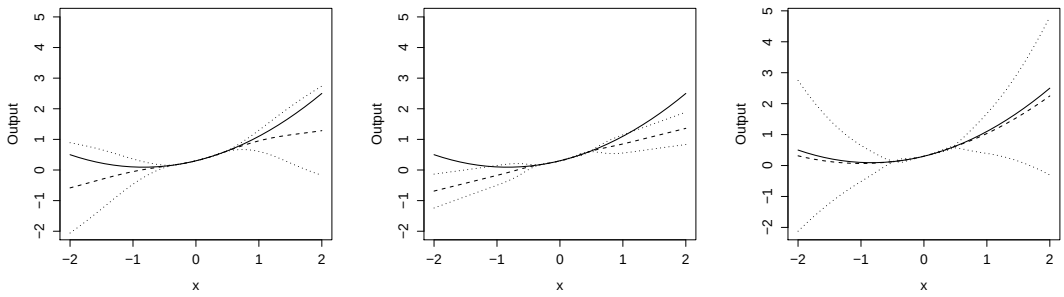

Figure SM1. *Example 1: Plots of the true function (solid line), the expectation of the emulator (dashed line) and the probability interval formed by taking two standard deviations either side of the mean (dotted line). The varying coefficient model (model (3.1)) is shown on the left hand side, the fixed coefficient model (model (1.1)) with basis functions* $\mathbf{h}(x) = (1, x)$ *in the centre and the fixed coefficient model with with basis functions* $\mathbf{h}(x) = (1, x, x^2)$ *on the right hand side.*

fixed coefficient model two basis functions are tested: $\mathbf{h}(x) = (1, x)^T$ (as above) and $\mathbf{h}(x) = (1, x, x^2)^T$. The priors used are largely the same as those used above, with the exception of σ_1, which is set to 0.08. This value is chosen so that $\mathrm{Var}[f(x)] = k(x, x)$ is approximately the same for both the fixed coefficient and varying coefficient models when $x = 0.25$ and $x = -0.25$. When $\mid x_1 \mid < 0.25$, the prior variance of $f(\mathbf{x})$ is larger for the fixed coefficient model, and when $\mid x_1 \mid > 0.25$ the prior variance is smaller for the fixed coefficient model. The correlation length δ is set to 0.3, to correspond to δ_1 in the varying coefficient model.

Figure SM1 compares the varying coefficient emulator to the fixed coefficient emulator. It is clear that the variance of the fixed coefficient emulator with basis function $\mathbf{h}(x) = (1, x)^T$ (centre) is too small to capture the variation in the true function when extrapolating outside $x \in [-0.5, 0.5]$ (i.e. the range of the design). When the correct basis function $\mathbf{h}(x) = (1, x, x^2)^T$ is used (right hand side) the true function lies within the dotted lines but the posterior variance is much larger than that of the varying coefficient emulator (note the difference in scales on the y-axis). In practice, the correct basis function will not be known and as both basis functions appear to fit the data well within $x \in [-0.5, 0.5]$ there is a risk that uncertainty will be underestimated when extrapolating with fixed coefficients.

SM1.2. Example 2 - a trigonometric function. For the second example, data are generated from the following function:

$$\begin{aligned} f(x) &= \alpha_0 + \alpha_1 x \quad \text{for} \quad x \leq 0 \\ f(x) &= \alpha_0 + \sin(\alpha_1 x) \quad \text{for} \quad x > 0 \end{aligned} \tag{SM1.2}$$

where $x \in [-0.5, 0.5]$, $\alpha_0 = 0.3$ and $\alpha_1 = 1 + 0.8x$ so that α_1 varies with x. As before, an emulator of form (3.1) with basis function $\mathbf{h}(x) = (1, x)^T$ is fitted to a set of 5 evaluations of the function f. Priors are chosen as for the first example, with $\sigma_1 = 0.06$ and $\sigma_2 = 0.8$.

Figure SM2 compares the varying coefficient model to two fixed coefficient models with basis functions $\mathbf{h}(x) = (1, x)^T$ (centre) and $\mathbf{h}(x) = (1, x, x^2)^T$ (right hand side). We see again the same pattern as for the first example - the variance of the fixed coefficient emulator with

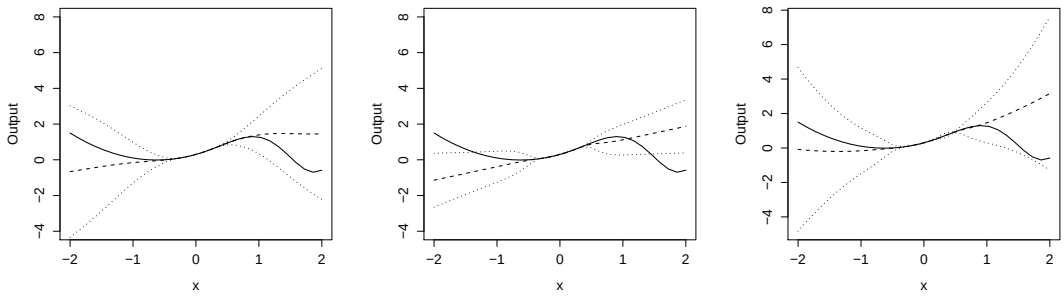

Figure SM2. *Example 2: Plots of the true function (solid line), the expectation of the emulator (dashed line) and the probability interval formed by taking two standard deviations either side of the mean (dotted line). The varying coefficient model (model (3.1)) is shown on the left hand side, the fixed coefficient model (model (1.1)) with basis functions* $\mathbf{h}(x) = (1, x)$ *in the centre and the fixed coefficient model with with basis functions* $\mathbf{h}(x) = (1, x, x^2)$ *on the right hand side.*

basis function $\mathbf{h}(x) = (1, x)^T$ is too small to capture the variation in the true function when extrapolating outside the range of the design. In the right hand plot, the fit is much better but again the variance is larger than for the varying coefficient emulator and it is not clear that this basis function could be correctly chosen based on the design alone.

SM1.3. Example 3.

SM1.3.1. Model and prior assumptions. For the third example, an emulator is fitted to data generated by the following function:

$$\begin{aligned} f(x_1, x_2, x_3, x_4) =& \alpha_0 + 2\cos\left(-\frac{\pi}{15} + \alpha_1 x_1\right) - \alpha_2(x_2 + \alpha_3)^2 \\ & - x_3 \sin(\alpha_4 x_3) - \alpha_5 x_4^2, \end{aligned} \tag{SM1.3}$$

where $x_1, x_2, x_3, x_4 \in [-0.5, 0.5]$. Set $\alpha_0 = 0.3$, $\alpha_2 = 1.5$, $\alpha_3 = 0.1$, $\alpha_4 = 0.6$ and $\alpha_5 = 0.75$. Let $\alpha_1 = \exp(0.4(x_1 + x_2 + x_3))$ so that α_1 varies with $\mathbf{x}$. An emulator of form (3.1) is fitted to a set of N evaluations (where N is small relative to the number of inputs) of the function f. This time, the aim is to use the emulator to find the inputs that correspond to the maximum of $f(\mathbf{x})$.

For this example, the basis functions in (3.1) are set to

$$\mathbf{h}(\mathbf{x}) = (1, x_1, x_2, x_1^2, x_2^2, x_3^2, x_4^2)^T. \tag{SM1.4}$$

These basis functions have been chosen to correspond to the Taylor expansion of $f(x_1, x_2, x_3, x_4)$.

The parameter α_1 varies with x_1, x_2 and x_3. As α_1 appears as a coefficient of x_1 and x_1^2 in the Taylor expansion of $f(\cdot)$, β_2 and β_4 are chosen to be the varying coefficients in model (3.1). To model the local variation arising from the error term in the Taylor expansion, β_1 is also allowed to vary throughout the input space. As β_3, β_5, β_6 and β_7 are to be held constant, we set $\sigma_3 = \sigma_5 = \sigma_6 = \sigma_7 = 0$.

The prior judgments for the variance parameters are set as for the first two examples, so that $\sigma_1 = 0.04$. Prior judgments for σ_2 and σ_4 were obtained by assessing the size of the coefficients of x_1 and x_1^2 in a standard regression fit and by considering the range of values taken by the coefficients of x_1 and x_1^2 in the Taylor expansion of $f(\cdot)$. Based on this, σ_2 was set to 0.2 and σ_4 was set to 0.7. In practice, a Taylor expansion is unlikely to be available for setting prior judgments but subjective views from expert modellers, who may have an idea as to the relative effects of different input parameters on the output in different regions of the space, can be used to set these parameters. Prior judgments can also be formed by using a test set or small initial set of model evaluations to assess the variation in coefficients if regression models are fitted to different regions of the space.

The correlation functions $c_j(\mathbf{x}, \mathbf{x}')$ are again set to the Gaussian correlation function given in (3.2). The correlation lengths are set to $\delta_{1k} = 0.3$, $\delta_{2k} = 1$ and $\delta_{4k} = 1$ for $k = 1, 2, 3$. Thus, the coefficients of x_1 and x_1^2 are a priori assumed to vary slowly throughout the space, whereas $\epsilon_{\beta_1}(\mathbf{x})$ is assumed to be more locally variable. The remainder of the prior assumptions are the same as the previous two examples.

SM1.3.2. Comparison to fixed coefficient model. The fixed coefficient model described in (1.1) is also used to emulate the function (SM1.3) so that comparisons can be drawn between the two models. The same basis functions (as given in (SM1.4)) are used for this fixed coefficient model. The priors used are largely the same as those used above, with the exception of σ_1, which is set to 0.07755 (chosen as in the first example).

Table SM1 compares the fit of the fixed coefficient model and the varying coefficient model using the mean squared error as the metric (the mean squared error is calculated on a leave-one-out basis). The results presented are the average values obtained using 50 randomly selected Latin hypercube designs. Under this metric, the fit of the varying coefficient model is better than the fit of the fixed coefficient model for $N > 5$. The fixed coefficient model has a stochastic process term, which should be capable of modelling the local variation in β_2 and β_4 given enough design points, but even with 100 design points, the fixed coefficient model still has a larger mean squared error than the varying coefficient model.

Number of Design points	Average reduction in MSE	Average MSE Model (3.1)	Average MSE Model (1.1)
5	-0.0020 (19 positive)	0.0722	0.0702
10	0.0025 (41 positive)	0.0093	0.0118
15	0.0019 (49 positive)	0.0031	0.0050
20	0.0016 (49 positive)	0.0019	0.0035
25	0.0014 (50 positive)	0.0011	0.0025
50	0.0012 (50 positive)	0.0004	0.0016
100	0.0009 (50 positive)	0.0001	0.0010

Table SM1

Assessment of average model fit over 50 Latin Hypercube samples with varying size of design. The average mean squared error is shown for Model (3.1) and Model (1.1) .

Figure SM3 compares the true $f(\mathbf{x})$ to the estimates obtained using each of the emulators

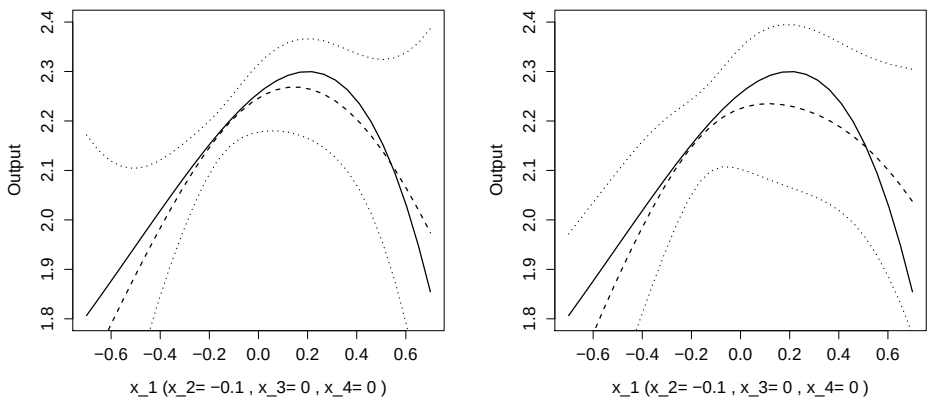

Figure SM3. *Plots of the true function (solid line), the expectation of the emulator (dashed line) and the probability interval formed by taking two standard deviations either side of the mean (dotted line) for fixed* $x_2 = -0.1$*,* $x_3 = 0$*,* $x_4 = 0$ *and* $x_1 \in (-0.75, 0.75)$ *for a design of size 15. The varying coefficient model (model (3.1)) is shown on the left hand side and the fixed coefficient model (model (1.1)) is shown on the right hand side.*

when varying x_1, for a design of size 15 with $x_i \in [-0.5, 0.5]$ for $i \in \{1, 2, 3, 4\}$. The mean of the emulator tracks the true function more closely when the varying coefficient model is used. The updated variance of the emulator is also generally smaller for the varying coefficient model for $x_1 \in [-0.5, 0.5]$. Extrapolating outside the design, we see the same as in the first two examples: the probability interval associated with the varying coefficient model is larger than the probability interval associated with the fixed coefficient model. This finding is confirmed in Figure SM4 (which considers values of x_1 between -2 and 2). In this range, the error term in the Taylor expansion of $f(\cdot)$ is much larger, and the varying coefficient α_1 takes a wider range of values. As seen in Figure SM4, the varying coefficient model is better at capturing this increased uncertainty outside $x_1 \in [-0.5, 0.5]$. Using the fixed coefficient model, the true function shown in bold is outside the probability interval for large values of x_1.

Table SM2 assesses the two models against the objective of the study, which was to find $\mathbf{x}$ that maximises $f(\mathbf{x})$. The Table gives the loss (against the maximum found using the true function) for each of the emulators for $N = 10, 15, 20$. The search for the maximum is performed over a Latin hypercube sample of size 50,000. The loss presented is averaged over ten draws of a Latin hypercube design for fitting the emulator and over ten draws from the grid over which the search is performed. We see that for all N tested, the optimum found using the varying coefficient model is closer to the truth than the optimum found using the fixed coefficient model.

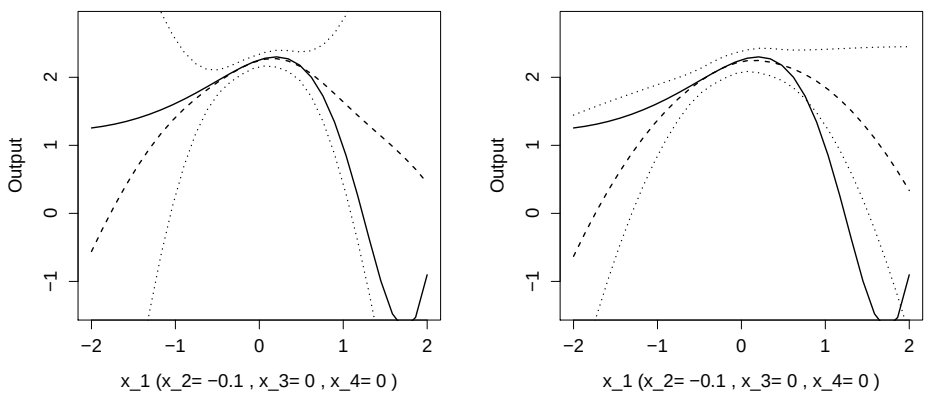

Figure SM4. *Plots of the true function (solid line), the expectation of the emulator (dashed line) and the probability interval formed by taking two standard deviations either side of the mean (dotted line) for fixed* $x_2 = -0.1$, $x_3 = 0$, $x_4 = 0$ *and varying* $x_1 \in (-2, 2)$ *for a design of size 15. The varying coefficient model (model (3.1)) is shown on the left hand side and the fixed coefficient model (model (1.1)) is shown on the right hand side.*

Number of Design points	Average loss Varying coefficient model (3.1)	Average loss Fixed coefficient model (1.1)
10	0.031	0.054
15	0.007	0.017
20	0.001	0.004

Table SM2

The average loss (over ten designs and ten grids) when using an emulator instead of the true function to find the maximum of $f(\mathbf{x})$*, searching over a Latin hypercube sample of size 50,000. The loss is defined as* $f(\mathbf{x}_{opt}) - f(\mathbf{x}_e)$*, where* $\mathbf{x}_e$ *is the input that maximises* $\mathbb{E}[f(\mathbf{x})]$ *(as estimated using the emulator) and* $\mathbf{x}_{opt}$ *is the input that maximises* $f(\mathbf{x})$*.*

SM2. Further details on the inputs of the motivating example. In this section, further details are given on the parametrisation of the inputs of the energy policy computer model. In the main manuscript, a brief discussion of these inputs is given in subsection 3.3.1.

Strike prices were parametrised as following an exponential decay (with each price rounded to the nearest £5). For each technology considered, the strike price in 2016 and the rate of decay over time were varied, with the remaining parameters of the exponential decay set to constants over all model runs. The strike price for each technology in year t (for $t \in \{2016, \ldots, 2035\}$) was given by

$$s_t = s_{min} + (s_{max} - s_{min}) \exp\left(-\frac{r(t-2016)}{T}\right),$$

for r the rate of decay over time and s_{max} the strike price in 2016 for the technology in question. The parameters s_{min} and T were set to be constant over all model runs. For offshore wind

$s_{min} = 90$ and $T = 19$; for onshore wind, $s_{min} = 75$ and $T = 4$; and for solar power, $s_{min} = 75$ and $T = 4$. The strike prices for round three offshore wind (which is a scheme for larger scale offshore wind development) were determined by setting s_{max} equal to that of standard offshore wind, and r equal to half that of standard offshore wind. The other parameters for round three offshore wind were set to $s_{min} = 100$ and $T = 19$. Values of the parameters s_{min} and T for each technology were set by considering the amount of support required and the expected timescales for further cost-reducing technological developments. The parameters s_{max} and r were varied in the design to investigate the impact on the output of changing these parameters. For all technologies, $r \in [0, 4]$. For offshore wind, $s_{max} \in [100, 160]$ and for onshore wind and solar, $s_{max} \in [75, 130]$.

The other inputs varied in the study were: electricity demand, coal price, gas price, the construction costs of different plants, the load factors for onshore and offshore wind power and the hurdle rates for onshore and offshore wind power. The assumptions used in a government study ([SM5], [SM2]) of this problem were available and were used to parametrise each of the inputs.

For annual total electricity demand from 2010 to 2080, three time series of assumptions were tested in the government study: $\mathbf{d}_c$ (the central assumption), $\mathbf{d}_l$ (the low assumption) and $\mathbf{d}_h$ (the high assumption). For emulation, the demand time series was parametrised as a shift away from the central assumption, setting demand to

$$\mathbf{d}_c + \frac{z_1}{2}\left((\mathbf{d}_c - \mathbf{d}_l) + (\mathbf{d}_h - \mathbf{d}_c)\right),$$

where z_1 is some scalar shift parameter. The high and low assumptions $\mathbf{d}_h$ and $\mathbf{d}_h$ are such that the range of demands tested increases through time. The shift parameter z_1 was allowed to range from -1.5 to 1.5 in the design.

The same approach was used to parametrise the time series of annual gas and coal prices. The price for each fossil fuel was set to the central government assumption plus some proportion of the average shift tested in the government study. Coal and gas were each given an independent scalar shift parameter, which varied between -1.5 and 1.5 in the design.

The construction costs for different plants are modelled with three different costs (low, medium and high) for each different plant and for each year. The supply curve used for each technology in each year consists of straight lines linking the low cost to the medium cost and the medium cost to the high cost. The central government assumptions for these costs were varied using one scalar multiplier, so that all costs for all technologies in all years are shifted by this multiplier. The multiplier varied between 0.8 and 1.2 in the design.

A hurdle rate for a plant is the return on an investment that would be required by an investor in this plant. The computer model incorporates a hurdle rate for each type of plant for each year. To vary the central hurdle rate assumptions used in the government study for onshore and offshore wind farms we took the central government assumptions for each type of wind farm and shifted them independently using two scalar multipliers (one for onshore wind and one for offshore wind). Each multiplier was set to between 0.8 and 1.2 in the design. Hurdle rates for all other plants were set to the central government assumption.

A similar approach was used to vary annual load factors for onshore and offshore wind. The load factor of a plant is the energy produced over a year divided by the theoretical amount

Parameter	Description	Minimum	Maximum
θ_1	Offshore strike price rate of decay	0.0	4.0
θ_2	Offshore strike price starting price	100.0	160.0
θ_3	Solar strike price rate of decay	0.0	4.0
θ_4	Solar strike price starting price	75.0	130.0
θ_5	Onshore strike price rate of decay	0.0	4.0
θ_6	Onshore strike price starting price	75.0	120.0
z_1	Demand	-1.5	1.5
z_2	Coal	-1.5	1.5
z_3	Gas	-1.5	1.5
z_4	Technology costs	0.8	1.2
z_5	Offshore hurdle rate	0.8	1.2
z_6	Onshore hurdle rate	0.8	1.2
z_7	Offshore load factor	0.8	1.2
z_8	Onshore load factor	0.8	1.2

Table SM3

Description of parameters with ranges explored when selecting Latin Hypercube sample

of energy that could be produced if the plant were operating at its maximum capacity. The computer model takes the load factor of a plant to be constant over the years of operation. Load factors for onshore and offshore wind farms were allowed to vary independently in the design between 0.8 and 1.2 of the central government assumption.

A summary of the parameters included in the study, with associated ranges is given in Table SM3. As described in the main manuscript, as an initial design, a maximin Latin hypercube sample was used to select 40 design points over the fourteen dimensional input space. After completing the initial design some computing time was still available so a further 16 design points were run. These 16 design points were chosen by generating 32 potential design points and calculating the expected value and standard deviation of the emissions, renewable generation and cost output for each of these potential design points. These values were calculated using a fixed coefficient emulator. By comparing the expected value and variance for each design point, 16 were subjectively selected as having a higher probability of meeting the three objectives. Time constraints prevented a more detailed analysis at this stage.

SM3. Further details on selecting the design for the motivating example. This section gives further technical details on the selection of the design for the motivating example, expanding on details in subsection 4.5.

The steps listed in subsection 4.5 add design points in batches of four. This process could be improved by the addition of stepwise delete steps, and by adding points to the design individually. By adding points to the design in batches of four and not performing stepwise deletes, it is not possible to mitigate against situations where points added to the design later reduce the importance of points selected earlier in the process. These improvements were difficult in practice because even with Bayes linear methods, the calculation for estimating the

criterion was expensive. The computing time needed to estimate the criterion also meant that only 50 points were selected as candidate design points initially. By using a Latin hypercube sample to select these 50 points we aimed to ensure that the candidate design points were well spread throughout the space. The final step above of choosing four points from a new design was used to reduce dependence on the single initial choice of Latin hypercube sample.

The government target for emissions was to reach 100gCO2/kWh in 2030. Based on the assumed parametric uncertainty and model evaluations chosen in the first wave, the probability of meeting this target was found to be much smaller than the probability of meeting the targets on spend and renewables. The target for emissions occurs ten years after the targets on renewable generation and spend and so the small probability of reaching 100gCO2/kWh by 2030 is not necessarily an issue. There is time to adjust future policy to meet 100gCO2/kWh when more information about the evolution of the electricity system through time is known. To account for this inequality in treatment of targets, each stage of the design selection process was repeated for an emissions target of 120gCO2/kWh. The selection of design points at each stage was made with reference to the criterion both for an emissions target of 120gCO2/kWh and an emissions target of 100gCO2/kWh (this only resulted in changes to the design in step 4). Another approach would be to adjust the weights in the criterion (4.4).

The grid over which the criterion was estimated was chosen by sampling 100 possible grids of size 25 and then finding which grid had the highest probability that one or more of the points in the grid met the three objectives associated with cost, emissions and renewable technologies. The aim of this process was to reduce time spent computing the criterion at grid points with very low probabilities of meeting the objectives as these points will have a weight close to zero and so any reduction in emulator variance at these points will not contribute to the overall criterion. By estimating the probability that one or more points in each selection of 25 meets the objectives we aimed to capture grid points across multiple modes in the space. If instead, the 25 grid points with highest individual probability were selected, there is a higher risk that all of these points could belong to one mode.

The design points included in the final design generally had high strike prices for offshore wind, having either a high starting price or a mid-range starting price with a low rate of decay. The strike prices for onshore wind and solar chosen in the design were spread throughout the range, although points with a very high strike price for offshore wind tended to have mid-range rates of decay for solar plant, possibly reflecting the increased cost that would result in later years from having high strike prices for both solar and offshore wind. Offshore wind has a larger effect on both the cost of government support and the emissions than onshore wind and solar because offshore wind is currently a more expensive technology and so requires higher strike prices to incentivise investment. The chosen design points also tended not to have high or low values for the construction cost of plant or the load factor for offshore wind. The computer model output is sensitive to these two inputs, so it is important to reduce the emulator variance in regions where these two inputs have a higher chance of occurring.

SM4. Further details on fitting the emulator using the full design. This section gives additional details on the fitting of the emulator to the full design, expanding on the text in subsection 5.1 in the main manuscript.

Table SM4 shows the linear terms included as basis functions of each element of the

emulator along with the estimates of the coefficients for these linear terms. Additional non-linear terms are listed beneath Table SM4. Intuitive explanations for the estimates of the coefficients can be found in subsection 3.3.2 of the main manuscript. Prior judgments for the variance and correlation lengths for the stochastic processes associated with each varying coefficient are listed in Table SM5. As in subsection 3.3.2, these prior judgments were set by considering the expected size of variation of each coefficient, and the speed over which the coefficient might vary. As described in the main manuscript, prior judgments were adjusted so that the variance term $\sum h_j^2(\mathbf{x})\sigma_j^2$ was approximately equal to the residual variance of a linear regression fit when $x_k = 0.5$ for all $k \in \{1, \ldots, 14\}$. These residual variances were 0.13^2 (renewables model), 0.13^2 (emissions model) and 0.15^2 (spend model). The remainder of the prior assumptions were set as described in subsection 3.3.2. Unlike in subsection 3.3.2, values for δ_{jk} were set to the same value for all k. Although it was expected that inputs included in the mean function would have a bigger effect on the correlation (and hence be associated with a smaller δ_{jk}), this was not found to make a difference in practice. To improve the fit of the emulator, the emissions outputs for all design points were multiplied by the demand in 2030 at that design point. This scaling was done so that the data used corresponded to the original computer model output of total emissions. This output is later post-processed by dividing by demand to give emissions per kWh. The input and output data were also scaled to lie between -1 and 1.

Input	Renewables	Emissions	Spend
1. Offshore strike price rate of decay	-0.11	0.16	-0.11
2. Offshore strike price starting price	0.35	-0.27	0.47
3. Onshore strike price rate of decay			
4. Onshore strike price starting price	0.07	0.04	0.07
5. Solar strike price rate of decay			
6. Solar strike price starting price			
7. Demand	-0.40	0.43	
8. Coal		-0.06	
9. Gas	0.07	0.08	-0.30
10. Technology costs	-0.33	0.62	-0.36
11. Hurdle rate offshore	-0.15	0.16	-0.19
12. Hurdle rate onshore	-0.07		
13. Load factor offshore	0.47	-0.29	0.53
14. Load factor onshore	0.23		0.10

Table SM4

Coefficient estimates for inputs included as linear terms in the vector of basis functions for each emulator are shown in the table. Interaction and squared terms also included were: renewables -(2,10), (2,13), (10,13), (2,10,13); emissions - (1,2), (1,7), (2,7), (4,7), (9,13), (11,11); spend - (1,2), (2,9), (2,10), (2,13), (9,10), (9,13), (1,1), (2,2).

The coefficient estimates in Table SM4 correspond to intuitive explanations for the relationships between the inputs and outputs of the computer model. When the starting strike price increases or the rate of decay of the strike price decreases, the support available to

Renewable model			Emissions model			Spend model		
Basis function	σ_j^2	δ_{jk}	Basis function	σ_j^2	δ_{jk}	Basis function	σ_j^2	δ_{jk}
Constant	0.07^2	0.3	Constant	0.05^2	0.3	Constant	0.05^2	0.3
2	0.11^2	1.5	10	0.1^2	1	9	0.15^2	2
4	0.11^2	1	13	0.22^2	2	11	0.15^2	3
7	0.1^2	1	1,7	0.05^2	1	14	0.15^2	1.5
10	0.05^2	1				2,9	0.25^2	4
2,10	0.2^2	1.5						
2,10,13	0.2^2	1						

Table SM5

Prior judgments for the covariance function associated with each basis function . The basis function number corresponds to numbers given in Table SM4.

renewable generation is greater and so there is more investment in renewable technologies, resulting in lower emissions and a greater cost to the government. Increasing demand results in a lower proportion of that demand being met by renewable sources and hence increasing emissions. Greater construction costs will result in less investment in renewable technology and hence greater emissions and a lower cost to the government (as there is less take-up of the support mechanism). An increase in hurdle rate or a decrease in load factor will also both result in lower investment in renewable plant and hence greater emissions. An increase in the price of gas will result in an increase in the wholesale price of electricity. If the wholesale electricity price rises, the difference between the strike price and the wholesale price will decrease, meaning that the cost of the government support decreases. The effect that increasing the gas price has on emissions is slightly less intuitive. The estimate of the coefficient is positive, whereas it might be expected that an increase in gas price would result in an increase in investment in renewable technology and hence a reduction in emissions, as plant can expect increased revenue through increased electricity prices. Offsetting this, decreasing gas prices can result in gas generation overtaking coal generation, resulting in a decrease in emissions (which would give the positive coefficient seen in Table SM4). A discussion of these issues in relation to the USA can be seen in [SM3].

REFERENCES

[1] R. Carnell, *lhs: Latin Hypercube Samples*, 2016, https://CRAN.R-project.org/package=lhs. R package version 0.13.

[2] DECC, *Electricity Market Reform - contracts for difference (department of energy and climate change)*. https://www.gov.uk/government/collections/electricity-market-reform-contracts-for-difference, 2015. Accessed: 2018-03-12.

[3] C. Knittel, K. Metaxoglou, and A. Trindade, *Are we fracked? The impact of falling gas prices and the implications for coal-to-gas switching and carbon emissions*, Oxford Review of Economic Policy, 32 (2016), pp. 241–259.

[4] M. D. Morris and T. J. Mitchell, *Exploratory designs for computational experiments*, Journal of Statistical Planning and Inference, 43 (1995), pp. 381 – 402.

[5] National Grid, *National Grid Electricity Market Reform analytical report*. https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/267614/Annex_D_-_National_Grid_EMR_

Report.pdf, 2013. Accessed: 2018-03-12.