# Applications of Dynamic Trees to Sensitivity Analysis

# William Becker

Research Associate, Econometrics and Applied Statistics Unit, European Commission -Joint Research Centre, Ispra, Italy

ABSTRACT: A recent approach to surrogate modelling, called dynamic trees, uses regression trees to partition the input space, and fits simple constant or linear models in each "leaf" (region of the input space). This article aims to investigate the applicability of dynamic trees in sensitivity analysis, in particular on high dimensional problems at low sample size, to see whether they can be applied to dimensionalities usually out of the range of surrogate models. Comparisons are made with Gaussian processes, as well as three measures based on a radial sampling scheme: the Monte Carlo estimator of the total sensitivity index, an elementary effects measure, and a derivative-based sensitivity measure. The results show that the radial sampling measures generally outperform the surrogate models tested here, with the exception of response surfaces that feature discontinuities.

Uncertainty analysis (UA) and sensitivity analysis (SA) are now widely acknowledged as essential components of model-based engineering design and risk analysis. However, there are still many practical difficulties associated with accurately quantifying and propagating uncertainties through a model. Not the least of these problems is that of computational expense: given that most complex models cannot be represented in a closed form, UA and SA must be performed by sampling the model by running it a number of times at different values of its input variables. If it is possible to run the model a fairly large number of times, the Monte Carlo method can be used to estimate the variance of the model output and measures of sensitivity to a reasonable degree of precision.

Unfortunately, complex models may often take hours or even days to run for a single set of input variable values; in such cases the Monte Carlo method cannot provide accurate estimates. Surrogate models, known variously as "emulators" or "metamodels", have become widely-used tools that aim to overcome this problem by building a statistical approximation of the model based on a small number of model runs. The surrogate model can then be used to estimate quantities of interest via the Monte Carlo method (since the surrogate model can be run in a very small amount of time), or using

analytical integration if it is sufficiently tractable.

Data modelling approaches may also be used in the setting of "given data", i.e. when the available data points are arbitrarily placed and cannot be positioned according to a Monte Carlo design, for example. This occurs in the analysis of composite indicators, in which data is available for a fixed number of entities, such as countries, regions or universities, and no further points can be added, In such cases, nonlinear regression, such as local linear regression or penalised splines, can be used to estimate first-order sensitivity by regressing against each variable in turn (Paruolo et al., 2013). In general, however, interactions between variables may be significant and it is not sufficient to know only first-order sensitivity indices. Multidimensional surrogate models have the capability to estimate the effects of higher order interactions and can estimate the total sensitivity indices.

At least two major drawbacks of surrogate models are that first, they introduce further uncertainty in the approximation of the model by its surrogate, although this is acceptable if the uncertainty is sufficiently small. Second, they tend to scale poorly with dimensionality: as the number of input variables grows, the number of data points required to accurately fit the surrogate can grow beyond a feasible limit. Given these caveats, there is a substantial amount of research effort devoted to building surrogate models that can emulate possibly strongly nonlinear and nonstationary model responses, at high dimensionalities, for as few model runs as possible. Examples of such emulators include Gaussian processes (Oakley and O'Hagan, 2004), polynomial chaos expansions (Sudret, 2008), and high dimensional model representations (Rabitz and Aliş, 1999). A comparison of some of these methods in the context of sensitivity analysis can be found in Storlie and Helton (2008).

One class of surrogate model that has the potential to be applicable in high dimensions, and can demonstrate considerable flexibility, is the use of regression trees (RTs). RTs take the approach of dividing the input space of the model into a number of complementary regions, known as "leaves", such that each leaf has a regression model associated with it which approximates a particular region of the input space. The advantage of this approach is that a number of simple, stationary (possibly linear or even constant) regression models can be combined to make a surrogate model that, globally, is able to handle nonlinear and nonstationary model responses, even with discontinuities (Becker et al., 2013).

A Bayesian approach to building RTs was developed by Chipman et al. (1998), which builds a posterior distribution over trees from a prior distribution conditioned on training data (model runs). Inferences can then be made by averaging over a large number of possible tree models. This concept was extended by Gramacy and Lee (2008) by using Gaussian processes as the regression model at each leaf. Most recently, the idea of "dynamic trees" (DTs) was developed by Gramacy and Polson (2011), which uses a particle learning algorithm to allow sequential updating of the RT. Additionally, an approach for variable selection was proposed based on the variance reduction due to tree nodes using each variable. These approaches are encoded in the R packages tgp (Gramacy, 2007) and dynaTree (Gramacy et al., 2013), which are used as the basis for these experiments.

In this work the use of DTs, and RTs in general,

is investigated in the context of sensitivity analysis, in particular to see whether RT surrogate models can be used to perform sensitivity analysis and variable screening in high dimensions, at low sample sizes. RTs are used to estimate variance-based sensitivity indices as well as an alternative sensitivity measure based on variance reduction caused by splits on a particular variable, described in Section 2. The performance is compared with samplingbased measures described in Section 3.

#### 1. BAYESIAN REGRESSION TREES

Consider a model f, which has k uncertain inputs denoted  $\{X_i\}_{i=1}^k$ , and a univariate output Y, such that Y = f(X). It will be assumed that  $X \in \mathscr{X}$ , where  $\mathscr{X} = [0,1]^k$ . A regression tree works by recursively dividing the input space  $\mathscr{X}$  into nonoverlapping partitions using rules of the form  $x_i \leq s$ , i.e. splitting the data using a single input variable at a time. An example of a simple RT is shown in Figure 1. The data are sorted by the splitting rules into the terminal nodes, or "leaves". Each leaf defines a region of the input space and has its own regression model assigned to it – this may be as simple as a constant or linear regression, or more sophisticated, such as a Gaussian process. Although a single re-



Figure 1: An example of a regression tree.

gression tree will have discontinuities between each leaf region, by adopting a Bayesian approach which averages over many possible trees, the discontinuities can potentially be smoothed out. In this sense the Bayesian approach shares some similarities to the random forests method (Breiman, 2001): a large number of "'weak learners" which can model complex data when combined. In the following, a very brief overview of Bayesian tree-based models is given.

In order to build regression trees using the Bayesian framework, it is necessary to create a prior distribution over tree models,  $p(T, \Theta)$ , where *T* represents the tree-structure random variable, and  $\Theta$  the random vector of unknown regression parameters defining the regression models at each leaf of the tree. This can conveniently be divided such that

$$p(\mathbf{\Theta}, T) = p(\mathbf{\Theta}|T)p(T) \tag{1}$$

which allows the tree prior to specified independently of the regression parameters. It is not possible to specify an analytical prior over trees, but a prior distribution may be specified indirectly via a number of rules which dictate the probability of moving from one tree to another, by adding and removing nodes, as well as swapping and changing splitting rules. This fits naturally into the framework of Markov chain Monte Carlo (MCMC) sampling, which is used to sample the posterior distribution.

The prior may now be combined with the model likelihood,  $(Y|X, T, \Theta)$ , given the data, which is dependent on the type of model used at each leaf. If the parameter prior has a carefully chosen form, it is possible to analytically marginalise the model parameters, i.e.,

$$p(Y|X,T) = \int p(Y|X,T,\mathbf{\Theta})p(\mathbf{\Theta}|T)d\mathbf{\Theta}.$$
 (2)

Now using Bayes' theorem the posterior distribution over trees can be found up to a proportional constant,

$$p(T|X,Y) \propto p(Y|X,T)p(T)$$
(3)

which can now be explored via the Metropolis Hastings (MH) algorithm. When the regression model at the leaves is more complex, for example when it is a Gaussian process, the parameters cannot be analytically marginalised. As a result, the MH algorithm cannot be exclusively used since the dimensionality of the parameter space changes from one step to the next. In this situation, Gramacy and Lee (2008) use reversible-jump MCMC, for jumps between tree models, with a combination of Gibbs sampling and the MH algorithm for sampling from the posterior parameter distribution.

Dynamic trees (DTs) are an extension to RTs in general, being designed to dynamically adapt and update on the observation of new data, using a series of rules on how a tree can change with the arrival of new data points, and a particle learning approach. The DTs used here use either constant or linear models at each tree leaf: in this sense they are simpler than the treed Gaussian process models described in (Gramacy and Lee, 2008). However, in the context of low sample sizes, the simpler leaf models might have an advantage.

DTs are designed in particular to dynamically update on the arrival of new data – they incorporate rules which change the structure of the tree locally with each new observation. The experiments here do not involve dynamic updating of the model, however the DT approach can equally be applied to batch data by running the data through the learning algorithm several times in a different order and using the averaged results. The details of the construction of the tree prior and the updating rules must be referred to Gramacy and Polson (2011) due to space limitations, however the essence of the DT approach is the same as any other Bayesian regression tree.

# 2. SENSITIVITY ANALYSIS WITH REGRESSION TREES

RTs can be used to generate large surrogate-model samples which can be used to estimate measures of sensitivity, specifically the first-order sensitivity indices (Cukier et al., 1973), and the total-order sensitivity indices (Homma and Saltelli, 1996). The experiments in this work will focus on the latter, defined as,

$$S_{Ti} = \frac{E_{\boldsymbol{X}_{\sim i}}[V_{X_i}\left(Y \mid \boldsymbol{X}_{\sim i}\right)]}{V(Y)} \tag{4}$$

where  $V(\cdot)$  is the variance operator, and  $\sim i$  denotes the set of indices except *i*. Estimation is performed using the Monte Carlo approach, described in Jansen (1999). In fact estimates can be returned

as distributions rather than point estimates, since they are estimated at every tree model visited in the MCMC search (after burn in). These distributions therefore account for both the uncertainty in the tree structure and model parameters.

A more novel approach to measuring sensitivity was also proposed by Gramacy et al. (2013), in which the "relevance" of a variable  $x_i$  is measured as the sum of the reductions in predictive variance due to splits involving  $x_i$ . Let the reduction in predictive variance for a given node (split) be  $\Delta(\eta)$ . The relevance is defined as,

$$J_i(T) = \sum_{\eta \in I_T} \Delta \eta \mathbf{1}_{[\nu(\eta)=i]},$$
(5)

where  $v(\eta)$  is the variable index of the split of  $\eta$ , and  $I_T$  is the set of all internal (non-terminal) tree nodes. When the regression model is a simple constant at each leaf, dependencies of y on  $x_i$  are captured exclusively by splits on  $x_i$ , and each split contributes to a reduction in variance if  $x_i$  affects the output in some way. Intuitively then, if a variable has no influence on the output, any splits on that variable will give no reduction in variance; conversely splitting on an influential variable will decrease the predictive variance. In the case where the leaf model is anything other than constant, the relevance measure cannot be used as a reliable measure of sensitivity because reductions in predictive variance will be due to both splits on variables and the specification of the regression model at the leaf. Although the requirement of a constant leaf model is somewhat restrictive, this work aims to see whether it can be used to perform "low-resolution" screening analyses in for models with high dimensionality.

### 3. SAMPLING-BASED MEASURES

In order to put the performance of the DT approach into perspective, the measures from the DTs ( $S_T$ and J) are compared here with some other sensitivity measures. To compare against the performance of a more "conventional" surrogate model, the fully Bayesian Gaussian process is used from the TGP package. Additionally, three much simpler "sampling-based" approaches are used which do not use surrogate models: first, the Monte Carlo estimator of  $S_T$ , which is based on a so-called "radial" experimental design. Letting  $\mathbf{x}_j^{(i)}$  and  $\mathbf{x}_j^{(i')}$  be, respectively, a point in the input space, and a point that differs from  $\mathbf{x}_j^{(i)}$  only in the value of  $x_i$ , the estimator of the numerator of  $S_{Ti}$  (see (4)) is as follows (Jansen, 1999),

$$\hat{V}_{Ti} = \frac{1}{2N} \sum_{j=1}^{N} \left| f(\mathbf{x}_{j}^{(i')}) - f(\mathbf{x}_{j}) \right|^{2}.$$
 (6)

The next measure is the mean of absolute elementary effects,  $\hat{\mu}_i^*$ , which is estimated as (Campolongo et al., 2011),

$$\hat{\mu}_{i}^{*} = \frac{1}{N} \sum_{j=1}^{N} \frac{\left| f(\boldsymbol{x}_{j}^{(i')}) - f(\boldsymbol{x}_{j}) \right|}{|\boldsymbol{x}_{ji}^{(i')} - \boldsymbol{x}_{ji}|}.$$
(7)

Here,  $x_{ji}$  denotes the *i*th coordinate of  $\mathbf{x}_j$ , so that the denominator of (7) is equal to the difference in  $x_i$  between  $\mathbf{x}_j^{(i)}$  and  $\mathbf{x}_j^{(i')}$ . The final measure used in this study is part of a set of sensitivity measures called "derivative-based global sensitivity measures" (DGSM). The measure is the integral of squared derivatives, i.e.  $v_i = \int_{\mathscr{H}} (\partial y / \partial x_i)^2 d\mathbf{x}$ . This may be estimated as (Sobol and Kucherenko, 2009),

$$\hat{\mathbf{v}}_{i} = \frac{1}{N} \sum_{j=1}^{N} \frac{\left| f(\mathbf{x}_{j}^{(i'')}) - f(\mathbf{x}_{j}) \right|^{2}}{|\mathbf{x}_{ji}^{(i'')} - \mathbf{x}_{ji}|}, \quad (8)$$

where  $\mathbf{x}_{j}^{(i'')}$  is a point that differs from  $\mathbf{x}_{j}$  only by a small increment  $\delta$  of  $x_{i}$ , in order to give an estimate of  $\frac{\partial y}{\partial x_{i}}$  at each point  $\mathbf{x}_{j}$ . This increment is kept fixed for all j, and is recommended as  $\delta = 1 \times 10^{-5}$  when sampling with respect to the unit hypercube.

#### 4. EXPERIMENTS

In order to assess the performance of the tree models and their associated measures, experiments were performed on test functions rather than physical models. Test functions represent the possible behaviour of complex physical models, but have the advantage that the sensitivity indices are known *a priori* via analytical expressions. The performance of the methods and measures here is of course conditional on the test function and may be different for other functions, however the same is true for physical models.

The focus of this work was on the applicability of DTs to problems of high dimensionality, at low sample sizes. In such cases it is unrealistic to expect precise estimates of sensitivity indices — usually one is interested in identifying input variables which have a significant influence on the model output, and similarly identifying those that have little or no effect. This setting is often referred to as "screening". Accordingly, for each test function, a certain fraction  $\gamma$  of the input variables was set to be of higher influence, and the remaining fraction  $1 - \gamma$  of variables to be of lower influence (the exact sensitivity is controlled by the parameter values in each case).

The experiments are then set as follows. Let  $k_{\text{high}} = \lfloor \gamma k \rfloor$ , i.e. the number of input variables that are set as high influence, and  $k_{\text{low}} = k - k_{\text{high}}$ . In each test function, the variables are set such that  $\{S_{T1} = S_{T2} = ... = S_{Tk_{\text{high}}}\} \gg \{S_{Tk_{\text{high}}+1} = S_{Tk_{\text{high}}+2} = ... = S_{Tk}\}$ . In other words, the first  $k_{\text{high}}$  variables are set to have equal and high sensitivities, and the remainder to have equal and low sensitivities.

Now let  $r_i$  be the ranking of the *i*th variable by one of the sensitivity measures defined previously, where ranking runs in descending order, i.e.  $r_i = 1$ is ranked as the most influential variable, and  $r_i = k$ is the least. The measure of error, *Z*, is as follows,

$$Z = \frac{1}{k_{\text{high}}} \sum_{i=1}^{k_{\text{high}}} \mathbb{1}(r_i > k_{\text{high}}), \tag{9}$$

where  $1(\cdot)$  is the count function. This metric therefore measures the fraction of influential variables that are ranked outside the top  $k_{high}$  variables by the sensitivity measure. This is purely a measure of sorting the variables into high and low importance groups, and gives no regard to precise rankings or possible cutoff values that might be used to select high importance from low importance variables, since what is a "high importance" variable is usually subjective and problem-dependent.

In order to capture the average performance, for is little better than random noise (consider that if each test function investigated, 20 repetitions are made (this limit was imposed by the significant variable, the value of Z would on average be 0.8).

computational cost of constructing a large number of surrogate models). The experimental designs here are all based on the Sobol' sequence, which is a low-discrepancy sequence suitable for both Monte Carlo estimation and surrogate model training. The sample is randomised by applying a random shift in each dimension for each replication, following the approach of Owen (1998). Additionally, each function is tested at sample sizes from  $N_T = 62$  to  $N_T = 248$ , representing the sizes of samples that might be available when the model is very computationally expensive. These particular values were chosen because the measures based on radial sampling require a structured sample of size  $N_T = N(k+1)$ , where N is a positive integer. At the chosen dimensionality of k = 30,  $N_T = 62$ when N = 2, for example. The surrogate models were built at the same sample sizes to make a fair comparison.

### 4.1. Polynomial additive function

The first function used for comparison was a simple polynomial additive function, of the form,

$$h(\mathbf{x}) = \sum_{i=1}^{k} a_i x_i^p, \tag{10}$$

where p is the order of the polynomial, and the  $a_i$ are weighting coefficients. In this function there are no interactions between variables, so  $\sum S_i = 1$ . The parameters were set as follows: p = 2,  $a_{high} = 3$ and  $a_{\text{low}} = 1$ , and  $\gamma = 0.2$ , with k = 30. This means that 20% of variables are set to have high sensitivities, i.e. by setting  $a_1 = a_2 = \dots = a_6 = 3$ , and  $a_7 = a_8 = \dots = a_30 = 1$ . The results are as shown in Figure 2. The polynomial function is a smooth additive function, which would tend to favour surrogate models which rely on assumptions of smoothness. However the results show that the performance of the DTs is rather poor. By far the worst performer is the relevance measure of the dynamic trees surrogate model, which does start converge to a reasonable level of error as the sample size increases above 200 points, but at lower sample sizes is little better than random noise (consider that if random sensitivity measures were assigned to each



Figure 2: Results of sensitivity measures applied to polynomial additive function.

The Monte Carlo estimator of  $S_T$  performs better, but still with a considerable margin of error. A much better performance is given by the  $S_T$  estimation via the DT surrogate, and even better via the Gaussian process. However, the best performance of all is given by the elementary effects and DGSM measures, which do not rely on surrogate models, and correctly identify the group of significant variables at every sample size, and for every data replication tested. This seems to suggest that at the sample sizes tested, when the objective is simply to identify important variables, surrogate models do not offer any improvement over measures that estimate directly from the sample (at least on this function).

### 4.2. $G^*$ function

The second test function is a widely-used benchmark function in sensitivity analysis studies – the " $G^*$  function". It has the following form:

$$G^{*} = \prod_{i=1}^{k} g_{i}^{*}$$
$$g_{i}^{*} = \frac{(1+\alpha) |2(x_{i}+\delta_{i}-\mathbf{I}[x_{i}+\delta_{i}])-1|^{\alpha} + a_{i}}{1+a_{i}}$$
(11)

where  $a_i, \delta_i$  and  $\alpha_i$  are parameters which can be chosen to obtain different behaviours of the function.  $\mathbf{I}[x_i + \delta_i]$  is the integer part of  $(x_i + \delta_i)$ . The relative importance of the inputs  $(x_1, x_2, ..., x_k)$  in the  $G^*$  function is controlled by the magnitude of  $a_i$ , and the nonlinearity by  $\alpha_i$ . The parameter  $\delta_i$  is a "shift" parameter which moves the position of the function in the input space, without having any effect on the sensitivities. This is set to zero in this work since for each replication, the Sobol' sample is already randomly shifted, achieving exactly the same effect. In this experiment, the parameters are set as  $a_{high} = 1$  and  $a_{low} = 2$ , which are chosen to result in a function with strong interactions: with k = 30 and  $\alpha = 2$ , the sensitivity indices can be analytically calculated as  $\sum S_i = 0.151$ .



Figure 3: Results of sensitivity measures applied to  $G^*$  function. The line styles are the same as in Figure 2.

The results of the RTs and other measures applied to the  $G^*$  function show a similar story to that of the polynomial function, but with a clearer division. The  $G^*$  function is strongly nonlinear and has strong interactions between variables, so is naturally a more challenging subject for sensitivity analysis. Referring to Figure 3, one can see that the RT surrogate models, and the Gaussian process, all show poor performance at the sample sizes tested, with variable ordering little better than random. The most successful measure is clearly the DGSM measure v, which effectively orders the variables even at the lowest sample size. Similar to the other measures, the error does not decrease very significantly with increasing sample size.

The DGSM measure appears to have an advan-



*Figure 4: Results of sensitivity measures applied to step function.* 

tage because it relies on small steps, which capture the partial derivative of the function at each point visited. In the case of the  $G^*$  function this captures the sensitivity quite well. Measures such as  $S_T$ and  $\mu^*$ , on the other hand, take large steps between samples, which in the case of a non-monotonic function, can underestimate sensitivity at low sample sizes. The surrogate models simply do not have enough training data to characterise the function at this sample size.

#### 4.3. Step function

The final test function is a simple function with a near-discontinuity, of the form,

$$s(\mathbf{x}) = \sum_{i=1}^{k} a_i \operatorname{erf}(15(x_i - 0.5))$$
(12)

where erf is the error function. This function has a gradient of zero in most places, except around  $x_i=0.5$ , at which point the gradient is very steep. For the numerical experiments,  $a_{high} = 2$  and  $a_{low} = 1$ , with a fraction  $\gamma$  set to 0.2.

The step function was in fact chosen as a counterexample to show the limitations of the DGSM measures. This is clearly shown in Figure 4, where the three DGSM measures perform quite poorly. This is very likely due to the fact that DGSM measures use small steps to approximate the pointwise gradient, however in the step function the gradient is zero in most places. So the DGSM measures require a fairly large sample size to sample a point in which the gradient is non-zero. On the other hand, the emulator approaches build a response surface from all the points simultaneously, so the large steps in each x direction are identified, even at low sample sizes. The relevance measure J still does not perform well, but the  $S_T$  estimate of the dynamic tree model performs the best on average of all the methods considered here. This suggests that in the presence of discontinuities, the dynamic tree approach might be the preferred option.

#### 5. DISCUSSION AND CONCLUSIONS

The conclusion of this work is that surrogate models did not really help in identifying significant variables at low sample sizes, when the dimensionality was reasonably high, with the exception of the neardiscontinuous step function. The hope was that dynamic trees, being relatively simple surrogate models, might be fruitfully applied in the screening context. However, it seems that surrogate models (including DTs) tend to be restricted to a particular domain of application: problems with low dimensionality and sufficient sample size. The fact that gradient-based measures easily outperform the surrogates in many of the experiments here demonstrates that when the sample size is low, surrogate models impose assumptions which cannot be justified - this is likely due to the surrogate model's attempts to extrapolate into unsampled regions of the input space with very little sample data to estimate the behaviour of the true model. Even in the case of the additive polynomial function, a very smooth function which would tend to favour smooth surrogate models such as Gaussian process, the surrogate models did not perform as well as the simpler elementary effects and DGSM measures.

One case in which the dynamic trees did perform well was that of the discontinuous step function, which is a setting that is unsuitable for gradientbased approaches. This also favours a model based on linear or constant regressions. However (in the experience of the author), most physical models do not exhibit this kind of behaviour.

In particular, the relevance measure based on dynamic trees was not very successful in the experiments performed here. This could be due to the fact that it relies on regression trees with constants at each leaf, which were unable to effectively model the nonlinear test functions considered here.

On the other hand, it is revealing that DGSM measures can be applied successfully even at very low sample sizes, when the aim is to screen significant variables from insignificant ones. A possible drawback of this approach could however be that the requirement of small perturbations may present problems in real models, because results may not be available to a sufficient number of significant figures to accurately estimate partial derivatives. This could possibly be overcome by tuning the perturbation size to the smallest value that can result in a reasonable estimate.

Further work that could stem from this study would be to understand under what circumstances surrogate models in general perform better than DGSM and elementary effects measures, and thus to guide practitioners to know whether to use a surrogate or sampling-based measures for a particular problem, perhaps based on the dimensionality of the problem and the available number of sample points.

## 6. References

- Becker, W., Worden, K., and Rowson, J. (2013). "Bayesian sensitivity analysis of bifurcating nonlinear models." *Mechanical Systems and Signal Processing*, 34(1), 57–75.
- Breiman, L. (2001). "Random forests." *Machine learn-ing*, 45(1), 5–32.
- Campolongo, F., Saltelli, A., and Campolongo, J. (2011). "From screening to quantitative sensitivity analysis. a unified approach." *Computer Physics Communication*, 182(4), 978–988.
- Chipman, H. A., George, E. I., and McCulloch, R. E. (1998). "Bayesian CART model search." *Journal of the American Statistical Association*, 93(443), 935– 948.
- Cukier, R. I., Fortuin, C., Schuler, K. E., Petschek, A. G., and Schaibly, J. (1973). "Study of the sensitivity of coupled reaction systems to uncertainties in rate coefficients. i theory." *The Journal of Chemical Physics*, 59, 3873–3878.
- Gramacy, R. B. (2007). "tgp: An R package for Bayesian nonstationary, semiparametric nonlinear re-

gression and design by treed Gaussian process models." *Journal of Statistical Software*, 19(9).

- Gramacy, R. B. and Lee, H. K. H. (2008). "Bayesian treed Gaussian process models with an application to computer modeling." *Journal of the American Statistical Association*, 103(483), 1119–1130.
- Gramacy, R. B. and Polson, N. G. (2011). "Particle learning of gaussian process models for sequential design and optimization." *Journal of Computational and Graphical Statistics*, 20(1).
- Gramacy, R. B., Taddy, M., Wild, S. M., et al. (2013). "Variable selection and sensitivity analysis using dynamic trees, with an application to computer code performance tuning." *The Annals of Applied Statistics*, 7(1), 51–80.
- Homma, T. and Saltelli, A. (1996). "Importance measures in global sensitivity analysis of nonlinear models." *Reliability Engineering & System Safety*, 52(1), 1–17.
- Jansen, M. (1999). "Analysis of variance designs for model output." Computer Physics Communications, 117, 35–43.
- Oakley, J. and O'Hagan, A. (2004). "Probabilistic sensitivity analysis of complex models: a Bayesian approach." *Journal of the Royal Statistical Society B*, 66, 751–769.
- Owen, A. B. (1998). "Monte carlo, quasi-monte carlo, and randomized quasi-monte carlo." *Monte Carlo and Quasi-Monte Carlo Methods*, 2000, 86–97.
- Paruolo, P., Saisana, M., and Saltelli, A. (2013). "Ratings and rankings: voodoo or science?." *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 176(3), 609–634.
- Rabitz, H. and Aliş, Ö. F. (1999). "General foundations of high-dimensional model representations." *Journal of Mathematical Chemistry*, 25(2-3), 197–233.
- Sobol, I. M. and Kucherenko, S. (2009). "Derivative based global sensitivity measures and their link with global sensitivity indices." *Mathematics and Computers in Simulation*, 79(10), 3009–3017.
- Storlie, C. and Helton, J. (2008). "Multiple predictor smoothing methods for sensitivity analysis: Description of techniques." *Reliability Engineering & System Safety*, 93(1), 28–54.
- Sudret, B. (2008). "Global sensitivity analysis using polynomial chaos expansions." *Reliability Engineering & System Safety*, 93(7), 964–979.