TUESDAY 2-3 (ALISON TOMLIN)

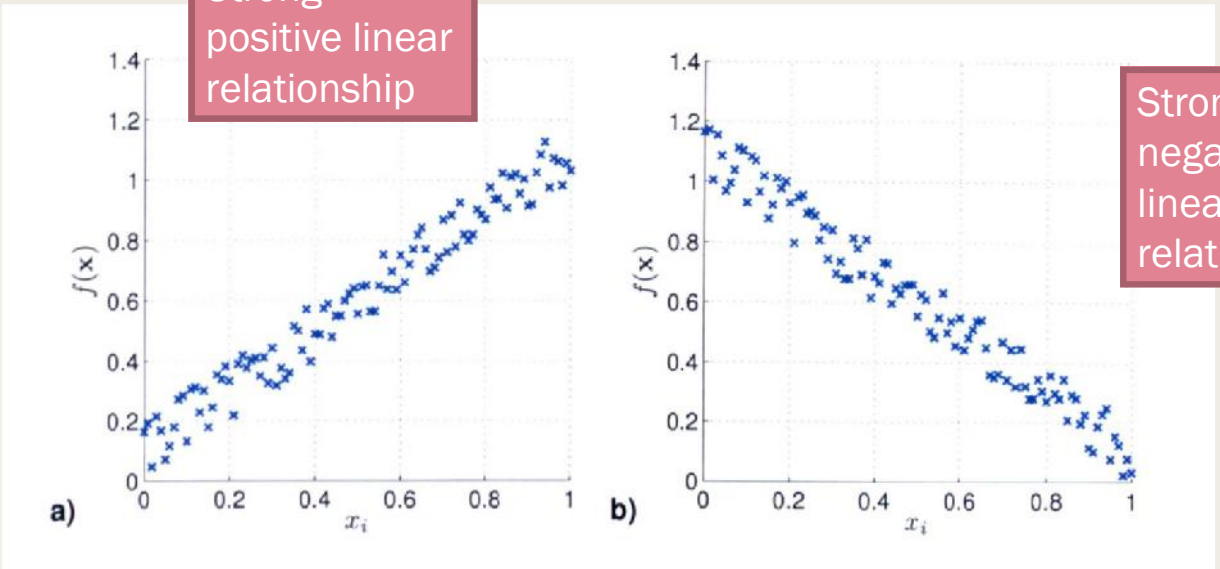
Global uncertainty analysis 2:

Sensitivity indices, FAST, HDMR

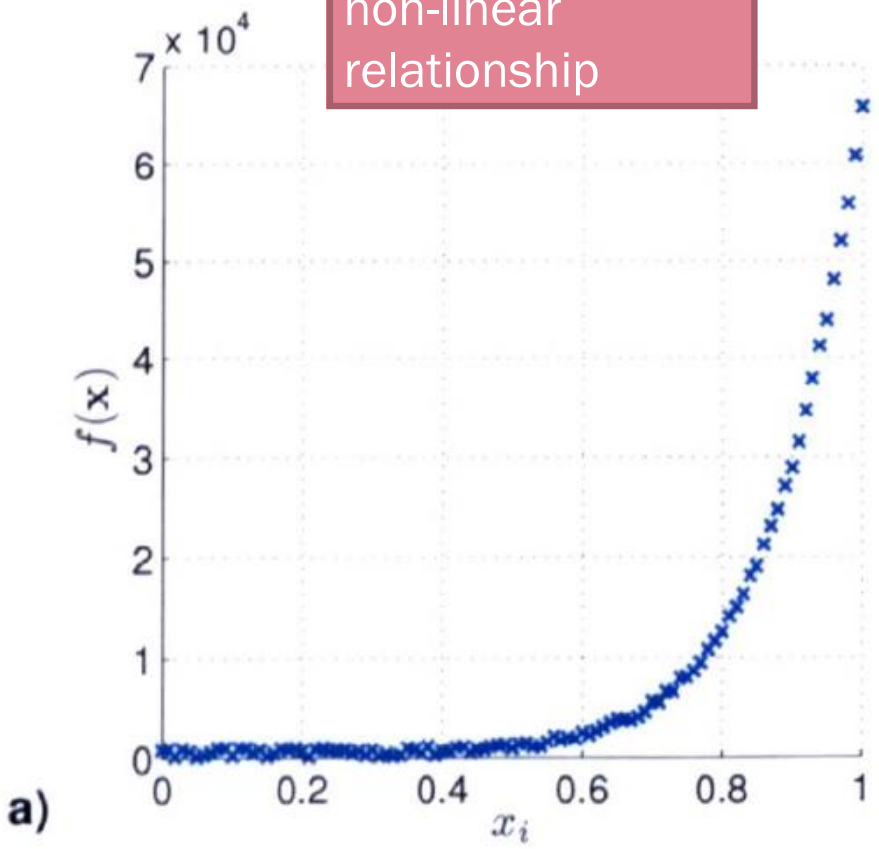
DETERMINING PARAMETER IMPORTANCE

How to determine parameter importance from random/quasi random samples

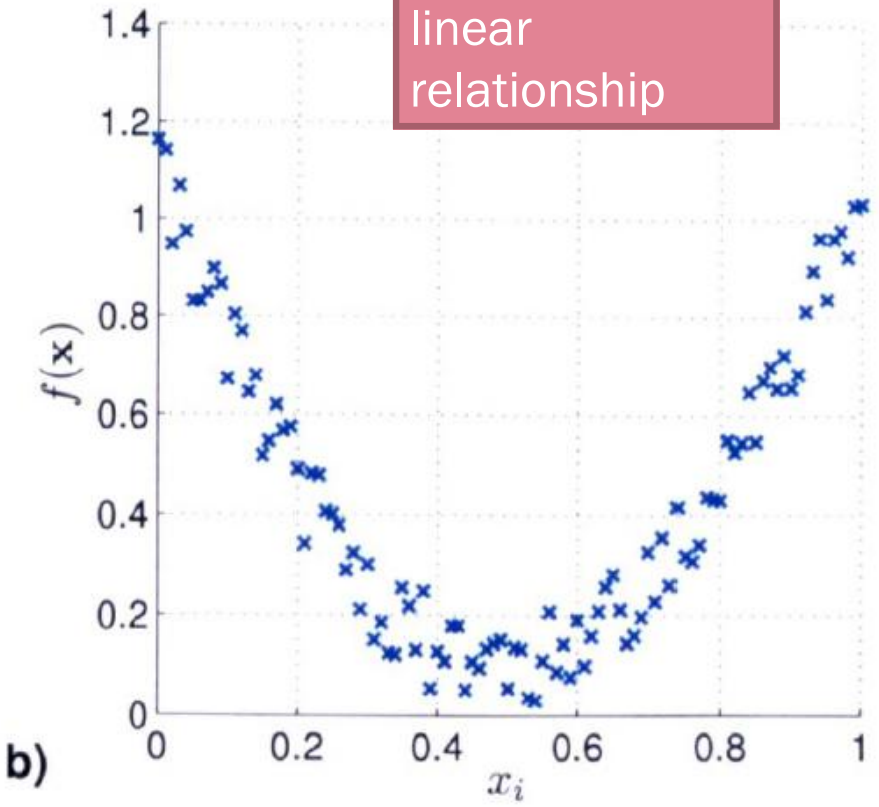
- Lots of different techniques with different levels of complexity.
- Easiest starting point is to generate scatter plots of model outputs vs. parameter input values.



Strong monotonic non-linear relationship



Strong non-monotonic non-linear relationship



Ranking parameters?

- For strong linear relationships the *Pearson correlation coefficient* would indicate the degree of parameter importance.
- Calculated by dividing the covariance of the variables by the square root of the product of their variances:

$$r_{xy} = \frac{\sum_{k=1}^{m}(x_k - \bar{x})(y_k - \bar{y})}{[\sum_{k=1}^{m}(x_k - \bar{x})^2]^{1/2}[\sum_{k=1}^{m}(y_k - \bar{y})^2]^{1/2}}$$

- High *r* values indicate a strong relationship between the input x_k and output y.
- Not as useful for nonlinear responses.

Spearman ranking

- In a rank transformation, data are replaced with their corresponding ranks and then correlation procedures are performed on these ranks.
- The smallest value of each variable/parameter is assigned rank 1, the next largest, rank 2, and so on up to sample size *m*.
- A correlation coefficient is then calculated using the rank values instead of the original values of the variables.
- The Spearman coefficient therefore assesses how well the relationship between two variables can be described using a monotonic function (see chapter 6 of (Saltelli et al. 2000)).
- A Spearman correlation of +1 or −1 therefore occurs when one variable is a perfect monotone function of the other.
- Not all relationships are monotonic and therefore more general methods are required for calculating parameter importance from sampling methods.

CALCULATION OF GLOBAL SENSITIVITY INDICES

VARIANCE BASED INDICES

Sobol's method

If the model result $Y_i = f_i(x_1, x_2, \ldots, x_N)$ is influenced by independent random parameters, then the joint *pdf* of the parameters $P(x_1, x_2, \ldots, x_N) = \prod_{j=1}^{N} p_j(x_j)$.

The mean or expected value $E(Y_i)$ of the calculated result Y_i is then given by:

$$E(Y_i) = \iint \ldots \int f_i(x_1, x_2, \ldots, x_N) \prod_{j=1}^{N} p_j(x_j)\, \mathrm{d}\, x_j$$

while the variance $V(Y_i)$ of the calculated result Y_i is specified as:

$$V(Y_i) = \iint \ldots \int (f_i(x_1, x_2, \ldots, x_N) - E(Y_i))^2 \prod_{j=1}^{N} p_j(x_j)\, \mathrm{d}\, x_j$$

$$= \iint \ldots \int f_i^{\,2}(x_1, x_2, \ldots, x_N) \prod_{j=1}^{N} p_j(x_j)\, \mathrm{d}\, x_j - E^2(Y_i)$$

- If this integral is calculated with a fixed value of a single parameter x_j, then the variance caused by all other parameters except for x_j , denoted by $V(Y_i|x_j)$ is obtained.
- If $V(Y_i|x_j)$ is calculated for many values of x_j, selected according to its *pdf*, then the expected value $E(V(Y_i|x_j))$ can be calculated.
- This requires the integration of $V(Y_i|x_j)$ over the *pdf* of x_j (Saltelli, 2002).
- The value $V(Y_i)$J$E(V(Y_i|x_j))$ is equal to the reduced variance of Y_i caused by fixing the value of x_j , and is equal to $V(E(Y_i|x_j))$.
- By dividing this conditional variance by the unconditional variance, the *first-order sensitivity index* for parameter x_j can be calculated:

$$S_{j(i)} = \frac{V\left(E\left(Y_i \middle| x_j\right)\right)}{V\left(Y_i\right)}$$

- This measure shows the fraction of the total variance of Y_i which is reduced when the value of x_j is held at a fixed value and is therefore a measure of the influence of uncertainty in x_j.

2nd order indices

- If the values of two parameters (*e.g.* x_j and x_k) are fixed, second-order sensitivity indices are obtained:

$$S_{kj(i)} = \frac{V\left(E\left(Y_i \middle| x_k, x_j\right)\right) - V\left(E\left(Y_i \middle| x_k\right)\right) - V\left(E\left(Y_i \middle| x_j\right)\right)}{V\left(Y_i\right)}$$

- The second-order sensitivity index characterizes the interaction of the corresponding parameters.
- Can be repeated up to higher and higher orders but the sample sizes required for the calculation of integrals using a Monte Carlo/sampling approach makes this prohibitive.

Cost

- The method provides sensitivity indices which are between 0 and 1, although sometimes this is multiplied by 100 yielding $S_{j(i)}$%.

- The calculation of these integrals is non-trivial and the use of a Monte Carlo sampling method is described in (Saltelli 2002) requiring N $(2m+1)$ model runs for first-order indices where N is the sample size chosen for the Monte Carlo estimates.
- The computational time requirement increases exponentially with the order of indices required.

Total effects and interactions

The *total effect* $S^{\text{tot}}_{j(i)}$ of parameter *j* can be defined as the sum of all sensitivity indices in which parameter *j* is present. Assume that we have three parameters *a, b* and *c*. The *total sensitivity index* of parameter *a* is defined as:

$$S^{\text{tot}}_{a(i)} = S_{a(i)} + S_{ab(i)} + S_{ac(i)} + S_{abc(i)}$$

If the parameters are totally additive, which means that there are no interactions at all between the parameters, then $\sum_{j=1}^{N} S_{j(i)} = 1$ and $S_{j(i)} = S^{\text{tot}}_{j(i)}$.

Then the variance of Y_i can be fully explained by first order effects.
If this is not the case then $S^{\text{tot}}_{j(i)} - S_{j(i)}$ is a measure of the level of interaction between the parameters.

The FAST (Fourier Amplitude Sensitivity) method

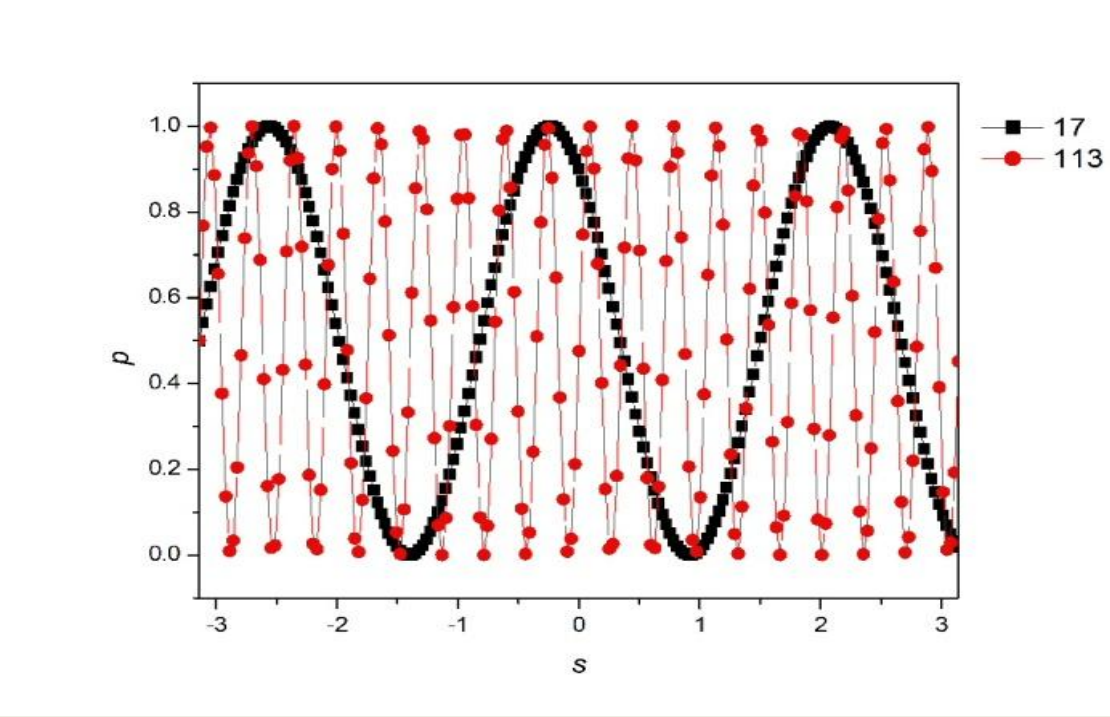
- Based on selecting *N* design points over a pre-described space-filling curve in the *m*-th dimensional input space, built so that each dimension (parameter) is investigated using a different frequency [ω_1,ω_2,...,ω_k] (Saltelli,Bolado 1998).
- In FAST, the *m*-dimensional integral in Sobol's method can be transformed to a one-dimensional integral using the following function:

$$x_j(s) = G_j(\sin \mu_j s)$$

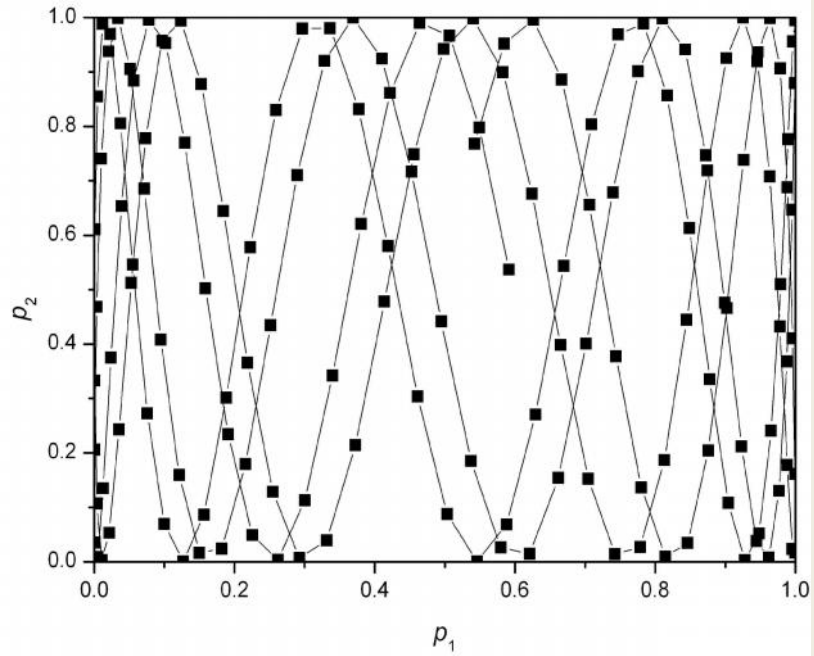
- The transformation function G_j depends on the probability density function of the corresponding parameter, the frequency ω_j belonging to parameter *j*, and the scalar search variable s.
- The values of all parameters become a periodic function of the search variable.

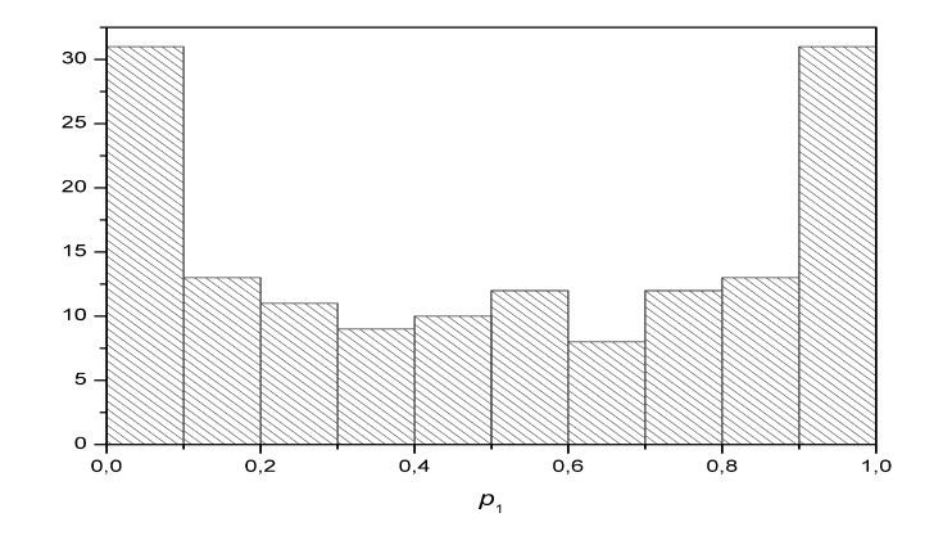
- If the frequencies ω_j are relative prime numbers, then the curve **x**(s) determined approaches all points of the parameter space in the rectangle of uncertainty of the parameters, while s is changing within the interval (-π, +π).

The values of two parameters were changed between 0 and 1 so that 157 different parameter sets were produced.
The generation of the parameter sets were controlled by search scalar s with steps $\Delta s = 0.04$.
The following functions were used $p_1 = 0.5 \sin(17s) + 0.5$ and $p_2 = 0.5 \sin(113s) + 0.5$.



Space coverage?





- The left plot shows that the whole space is covered but from the right plot we see that points are focussed towards the edge of the domain.

Analysis in FAST

- N parameter sets are defined by selecting *N* points equidistantly in the interval -π< s < π, then the corresponding Y_i model results are calculated (once for each parameter set) and a Fourier analysis of the results is carried out.
- In this way the variance fraction of the total variance of Y_i can be obtained:

$$V(Y_i) = 2\sum_{l=1}^{+\infty}\left(A_{il}^2 + B_{il}^2\right)$$

where A_{il} and B_{il} are the Fourier coefficients:

$$A_{il} = \frac{1}{2\pi}\int_{-\pi}^{\pi} Y_i(s)\cos(ls)\,\mathrm{d}s, \quad l = 0,1,\ldots \qquad B_{il} = \frac{1}{2\pi}\int_{-\pi}^{\pi} Y_i(s)\sin(ls)\,\mathrm{d}s, \quad l = 1,2,\ldots$$

- If the Fourier coefficients and their harmonics are calculated at the frequency ω_j, then the partial variance caused by parameter x_j is calculated from:

$$V_j\left(Y_i\right) = 2\sum_{r=1}^{+\infty}\left(A^2_{i,r\mu_j} + B^2_{i,r\mu_j}\right)$$

$$S_{ij} = \frac{V_j\left(Y_i\right)}{V\left(Y_i\right)}$$

where index *r* refers to the *r*-th harmonics of the base frequency ω_j.

Cost of FAST

- For the analysis of a model with m parameters, $N = 1.2\ m^{2.5}$ model evaluations have to be carried out (Cukier et al. 1977).
- This means 21200 simulations for a 50 parameter system.
- If only a small number of important parameters exist (i.e. ones which influence the target model output) then it is possible that random or quasi-random sampling methods may converge using a smaller sample size than that required by FAST.
- These can be coupled with response surface techniques (see later) to give lower cost variance based methods.
- The classic FAST method is used to determine first-order sensitivity indices. Where the first-order sensitivities over all parameters sum to much less than 1, this approach may be insufficient.
- Extended FAST can also be applied to total indices but this adds to the computational cost.

RESPONSE SURFACE METHODS

General RSM approach

- For computationally intensive models with a large number of parameters, the full Sobol and even the FAST method can be computationally very expensive.
- RSM based methods attempt to reduce the computational cost of Variance based sensitivity methods by first developing a fitted meta-model that accurately represents the relationship between the model parameters and outputs.
- If such a meta-model can be fitted with a lower number of model runs then it can be used to calculate variance based indices at a lower cost than the previous methods discussed.
- Have some similarities with Monte Carlo approaches:
  - *first input parameter ranges must be selected*
  - *then a suitable sampling approach should be taken so that full model runs are obtained across a design which is suitable for the development of an accurate meta-model.*
- Cost of method therefore is driven by the cost of providing an accurate surrogate model.

Issues of dimensionality

- The high dimensionality of the parameter space does not always imply a complex functional relationship between the more influential model inputs and target outputs.
- Interaction effects among more than two parameters are fairly rare in models of chemical systems (Rabitz,Aliş 2000; Li et al. 2001).
- Sample size required to develop a suitable meta-model may be much lower than required for the full investigation of sensitivity indices using Sobol's method or the FAST method.
- In a global sensitivity study of sulphur chemistry within a doped methane flame, Ziehn and Tomlin (2008) found that only 5 of the 176 parameters tested in the SRM analysis were required to build a meta-model giving 99.05% of the tested samples within the 5% relative error range compared to full model.
  - *To achieve a relative error of 99.65%, required 51 of the possible 176 first-order terms and only 4 of the possible 15,400 second-order terms to be included within the RSM.*
  - *A sample size of N =1024 was sufficient to build an accurate RSM.*
- RSMs therefore offer a promising approach for large parameter systems, or systems with high computational cost associated with the full model simulations.

Different approaches to RSMs

- Polynomial chaos expansions (Balakrishnan et al. 2002; Reagan et al. 2004; Najm et al. 2009; Cheng,Sandu 2009; Blatman,Sudret 2010; Prager et al. 2013).
- Gaussian process emulators (Oakley,O'Hagan 2002).
- Orthonormal polynomials (Turányi 1994; Tomlin 2006)
- Splines (Storlie,Helton 2008)
- High-dimensional model representations (Sobol' 1995; Rabitz et al. 1999; Wang et al. 2001; Ziehn,Tomlin 2008b; Ziehn et al. 2009b; Skodje et al. 2010; Klippenstein et al. 2011; Tomlin,Ziehn 2011; Goldsmith et al. 2013).

Gaussian process models

- Gaussian process emulator methods develop meta-models based on the assumption that given a set of target outputs $Y = f(x)$, the value of Y at an unknown value of x follows a multivariate Gaussian distribution.
- Given a big enough sample size, it is possible to produce any general shape of response surface.
- However, according to Saltelli (2008), since Gaussian emulators attempt to interpolate the mapping from x to $f(x)$ by applying a Gaussian kernel of the same dimension as that of the input parameter space, the methods may suffer from over-parameterisation and the so-called curse of dimensionality.
- In practice, have mainly been used for systems with a low number of parameters.
- One advantage they do have over other methods is that they give an idea as to the quality of the sensitivity indices (i.e. accuracy) as well as the estimates themselves (Oakley,O'Hagan 2002) and (Oakley,O'Hagan 2004).

Polynomial chaos expansion methods

- Here an uncertainty factor $\underline{u}_i$ is first assigned to each input variable.
  - *Note that this uncertainty parameter u_i is related to uncertainty parameter f by $u_i = 10^f$.*
- Taking the example of rate coefficients, they are then normalised into factorial variables **x** as follows:

$$x_i = \frac{\ln k_i / k_{i,0}}{\ln u_i}$$

Nominal value

- Hence $x_i = 0$ gives the nominal value of the rate coefficient, and -1 and +1 represent the upper and lower bounds.
- A response surface of the predicted combustion properties is then generated with respect to **x**.

- Often restricted to a 2nd order polynomial expansion which for the r'th model response ηr(x) can be written as:

$$\eta_r(\mathbf{x}) = \eta_{r,0} + \sum_{i=1}^{m} a_{r,i} x_i + \sum_{i=1}^{m} \sum_{j \geq i}^{m} b_{r,i,j} x_i x_j$$

- The uncertainty in **x** may be expressed as a polynomial expansion of basis random variables $\boldsymbol{\xi}$:

$$\mathbf{x} = \mathbf{x}^{(0)} + \sum_{i=1}^{m} \boldsymbol{\alpha}_i \xi_i + \sum_{i=1}^{m} \sum_{j \geq i}^{m} \boldsymbol{\beta}_{ij} \xi_i \xi_j + \ldots,$$

where α and β are column vectors of expansion coefficients, *m* is the number of rate coefficients under consideration and $\mathbf{x}^{(0)}$ is a column vector of normalised rate coefficients which is a zero vector for the nominal reaction model.

- If the x's are independent of each other and normally distributed, then the usual choice for the form of ξ would be a set of unit-normal random variables.
- If ln u_i represents 2 times the standard deviation of ln k_i then α is ½ I_m, where I_m is the *m*-dimensional identity matrix. β and all higher order terms are zero (Sheen et al. 2009).
- In the general case, combining the above two equations and truncating the higher order terms gives:

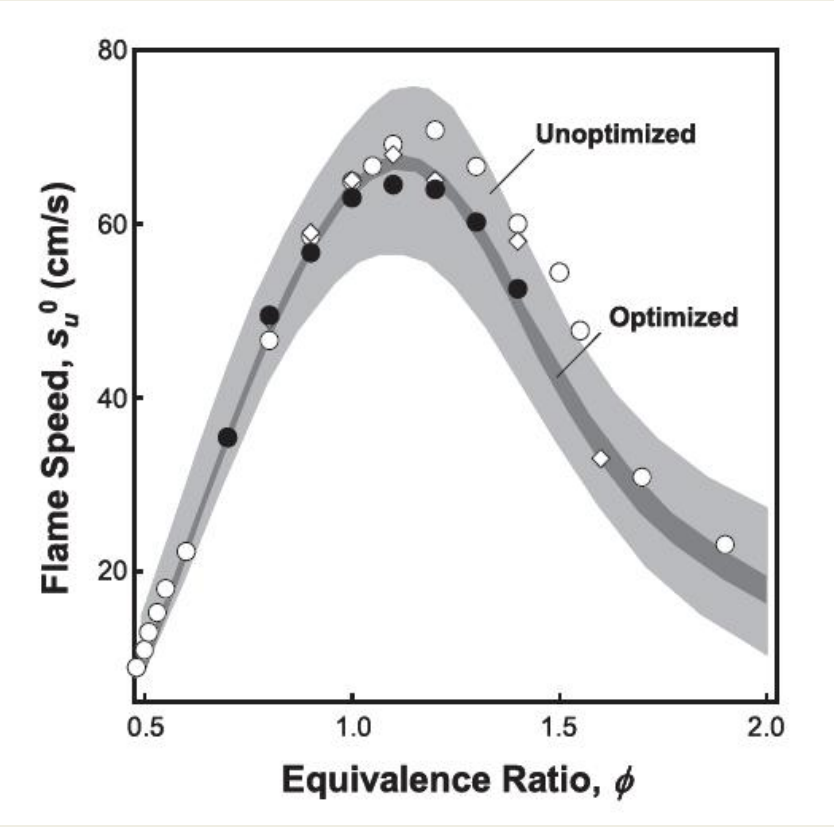
$$\ln r(\boldsymbol{\xi}) = \ln r(\mathbf{x}^{(0)}) + \sum_{i=1}^{m} \hat{\boldsymbol{\alpha}}_{r,i} \P_i + \sum_{i=1}^{m} \sum_{j \geq i}^{m} \hat{\boldsymbol{\beta}}_{r,ij} \P_i \P_j + \ldots,$$

$$\hat{\boldsymbol{\alpha}}_r = \tfrac{1}{2} \mathbf{I}_m \mathbf{a}_r \qquad \hat{\boldsymbol{\beta}}_r = \tfrac{1}{4} \mathbf{I}_m^T \mathbf{b}_r \mathbf{I}_m$$

- What this equation shows is that the overall model prediction is given by its nominal value plus uncertainty contributions from each rate coefficient.

$$\pm_r(\boldsymbol{\xi})^2 = \sum_{i=1}^{m} \hat{\boldsymbol{\alpha}}^2{}_{r,i} + 2 \sum_{i=1}^{m} \hat{\boldsymbol{\beta}}^2{}_{r,ij} + \sum_{i=1}^{m} \sum_{j>i}^{m} \hat{\boldsymbol{\beta}}^2{}_{r,ij}$$

Examples of application



Experimental data and computed 2σ uncertainty bands for the laminar flame speed of ethylene-air mixtures at p = 1 atm. (Sheen et al. 2009).

Note that following the application of an optimization procedure, the uncertainty bounds are much narrower.

The polynomial chaos expansion is used within the optimisation procedure.

ANOVA (ANALYSIS OF VARIANCES) DECOMPOSITION AND HDMR METHODS

Variance decomposition

For independent inputs (*i.e.* no correlations exist between inputs), a unique decomposition of the unconditional variance *V(Y)* can be obtained (Li et al. 2010):

$$V(Y) = \sum_{i=1}^{n} V_i + \sum_{1 \le i < j \le n} V_{ij} + \ldots + V_{12\ldots n} = \sum_{j=1}^{2^n-1} V_{x_j}$$

$$\sum_{j=1}^{2^n-1} \frac{V_{x_j}}{V(Y)} = \sum_{j=1}^{2^n-1} S_{x_j} = 1$$

The approach is therefore analogous to the classical approaches described above but instead of directly calculating the conditional variances using *e.g.* FAST or Monte Carlo samples, now a meta-model is developed first and the sensitivity indices are calculated using the meta-model.

RSM approaches to ANOVA decomposition

- Polynomial chaos expansions were one method to achieve this ANOVA decomposition.
- Other methods are based on High Dimensional Model Representations (HDMR).
- HDMR originally developed to provide a straightforward approach to explore the input-output mapping of a model without requiring large numbers of model runs (Sobol' 1995; Rabitz et al. 1999; Li et al. 2001).
- The use of truncated expansions is possible because usually only low-order correlations between inputs have a significant effect on the outputs.
- Because of the *hierarchical form* of HDMR component functions, sensitivity indices can be determined from them in an automatic way in order to rank the importance of input parameters and to explore the influence of parameter interactions.

Basic mapping

- The mapping between the inputs $x_1,\ldots, x_n$ and the output variable $Y(\mathbf{x}) = f(x_1,\ldots,x_n)$ can be written in the following hierarchical form:

$$Y(\mathbf{x}) = f_0 + \sum_{i=1}^{n} f_i(x_i) + \sum_{1\le i<j\le n} f_{ij}(x_i, x_j) + \ldots + f_{12\ldots n}(x_1, x_2, \ldots, x_n)$$

- Here the zeroth-order component f_0 denotes the mean effect, which is the expected value of the model output $f_0 = E(Y)$.
- The first-order component functions $f_i(x_i)$ give the effect of variable x_i acting independently (although generally nonlinearly) upon the output $Y(\mathbf{x})$:

$$f_i(x_i) = E(Y|x_i) - f_0$$

- The function $f_{ij}(x_i,x_j)$ is a second-order term describing the cooperative effects of the variables x_i and x_j upon the output $Y(\mathbf{x})$:

$$f_{ij}(x_i, x_j) = E(Y|x_{i,}, x_j) - f_i - f_j - f_0$$

HDMR and ANOVA

- So we can see that the basic HDMR expansion is equivalent to the ANOVA decomposition.
- This means that if we can find an ***accurate meta-model*** with which to represent the HDMR expansion, we can provide an accurate estimate of the partial variances and therefore the global sensitivity indices.
- The ANOVA decomposition has several special properties:
  - *The expected value of all non-constant component functions is zero and the terms are orthogonal (Sobol 2001).*
  - *The notation of zeroth-, first-, second-order, etc. in the HDMR expansion should not be confused with the terminology of a Taylor series since the HDMR expansion is always of finite order (Rabitz,Aliş 2000).*
- The higher-order terms reflect the ***cooperative effects*** of increasing numbers of input variables acting together to influence the output Y(x).
- HDMR is computationally very efficient if higher-order input variable interactions are weak and can therefore be neglected. Li et al. suggest (2001) that for many models, an HDMR expansion up to second-order gives a good approximation to the function Y(x). Where not, Tomlin and Ziehn (2011) showed that transformations of the outputs can be used to help build a low-order HDMR model and to therefore identify the important parameters.

QRS-HDMR

- We showed earlier that a quasi-random sequence such as a Sobol sequence had better convergence properties than other sampling approaches.
- Therefore we expect the Sobol' sequence to be a better choice of sampling strategy for fitting an HDMR meta model.
    1. *A quasi-random sample would therefore be developed for the chosen input parameter space.*
    2. *The full model would be run for each sample (e.g. 1024, 2048, etc) and target outputs stored.*
    3. *A meta-model would be fitted to the input-output relationships for each target output.*
    4. *The fitted HDMR meta-model would be used to derive global sensitivity indices.*
- The accuracy of the meta-model determines the accuracy of the calculated indices and needs to be checked carefully.

QRS-HDMR fitting procedure

- Using the RS-HDMR method the zeroth-order term f_0 is approximated by the average value of $Y(\mathbf{x})$ for all $\mathbf{x}^{(s)} = (x_1^{(s)}, x_2^{(s)}, \ldots, x_n^{(s)})$, s = 1,2,...,*N*

$$f_0 \approx \frac{1}{N}\sum_{s=1}^{N} Y(\mathbf{x}^{(s)})$$

where *N* is the sample size.

- To reduce the sampling effort, the higher-order component functions are approximated by expansions in terms of suitable basis functions which may include polynomials, splines *etc.* For example, expansion in terms of orthonormal polynomials is given by:

$$f_i(x_i) \approx \sum_{r=1}^{k} \sim_r^i \ddot{}_r(x_i) \qquad f_{ij}(x_i, x_j) \approx \sum_{p=1}^{l}\sum_{q=1}^{l'} {}_{pq}^{ij} \ddot{}_p(x_i) \ddot{}_q(x_j), \ldots$$

where *k,l,l'* represent the order of the polynomial expansion, $\sim_r^i$ and ${}_{pq}^{ij}$ are constant coefficients to be determined, and $\ddot{}_r(x_i)$, $\ddot{}_p(x_i)$ and $\ddot{}_q(x_j)$ are the orthonormal basis functions (Li et al. 2002a).

Sample size

- The coefficients are determined using Monte Carlo integration over the chosen input sample (Li et al. 2002a).
- The approximation of the component functions reduces the sampling effort dramatically so that only one set of quasi-random samples *N* is necessary in order to determine all RS-HDMR component functions and subsequently the sensitivity indices.
- For first-order indices this sample can usually be quite small (e.g. 1024).
- If significant second-order effects are present then the sample size will need to be bigger.

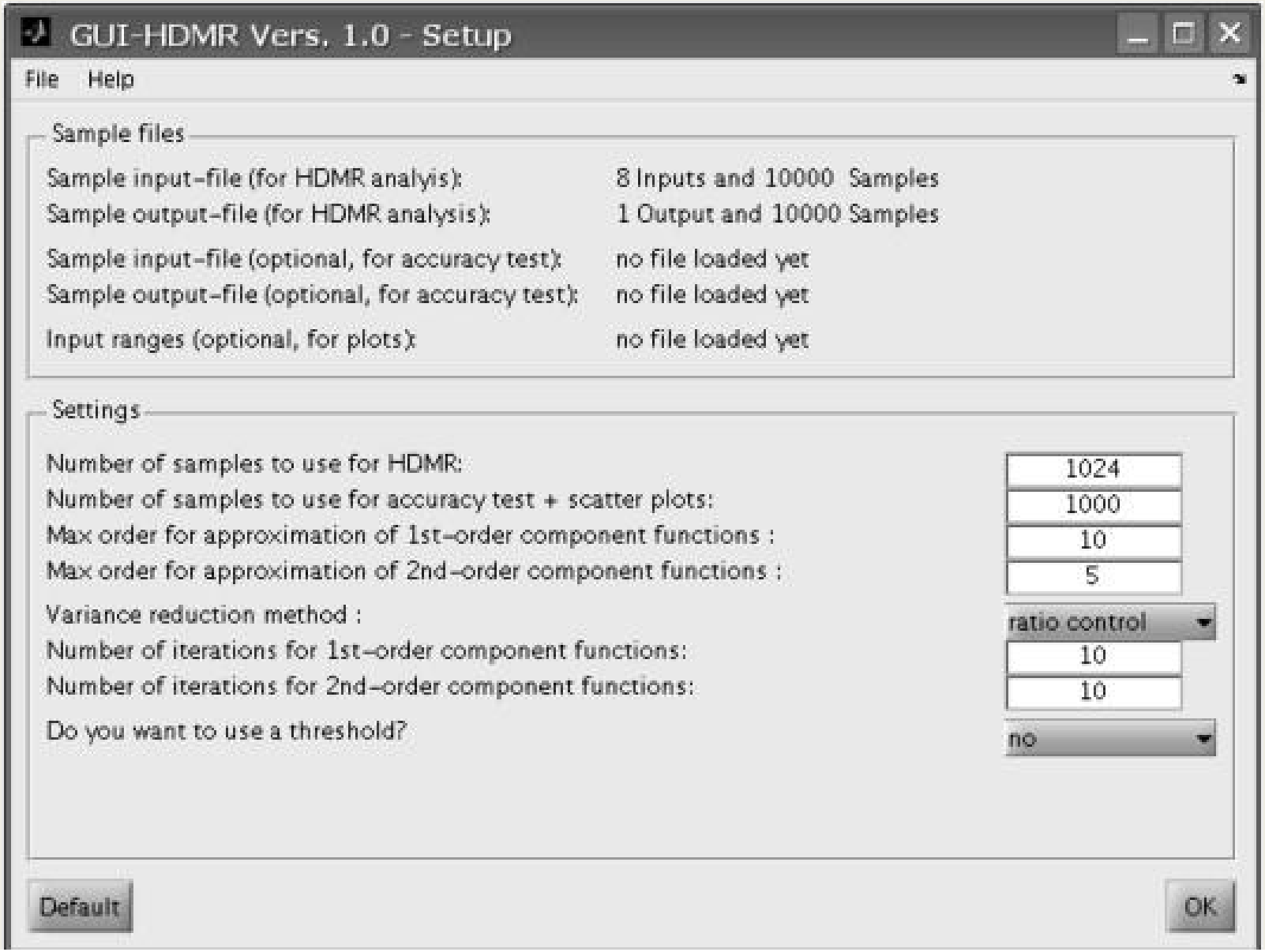
- Remember – base 2 system so sample size increases as 2^{Ns}
  - *2, 4, 8, 16, 32, 64, 128, 256, 512, 1024, 2048 etc!*

Improving accuracy

- The standard RS-HDMR approach was extended by an optimisation method (Ziehn,Tomlin 2008a), which automatically chooses the *best polynomial order* for the approximation of each of the component functions.
- Component functions can also be excluded from the HDMR expansion if they do not make a significant contribution to the modelled output value via the *use of a threshold*.
- The aim is to reduce the number of component functions to be approximated by polynomials and therefore to achieve automatic complexity reduction without the use of prior screening methods such as the Morris method (Morris 2006).
- For a second-order HDMR expansion a separate threshold can be defined for the exclusion of the first and second-order component functions.
- It is important to use thresholds for large parameter systems! (see practical class).

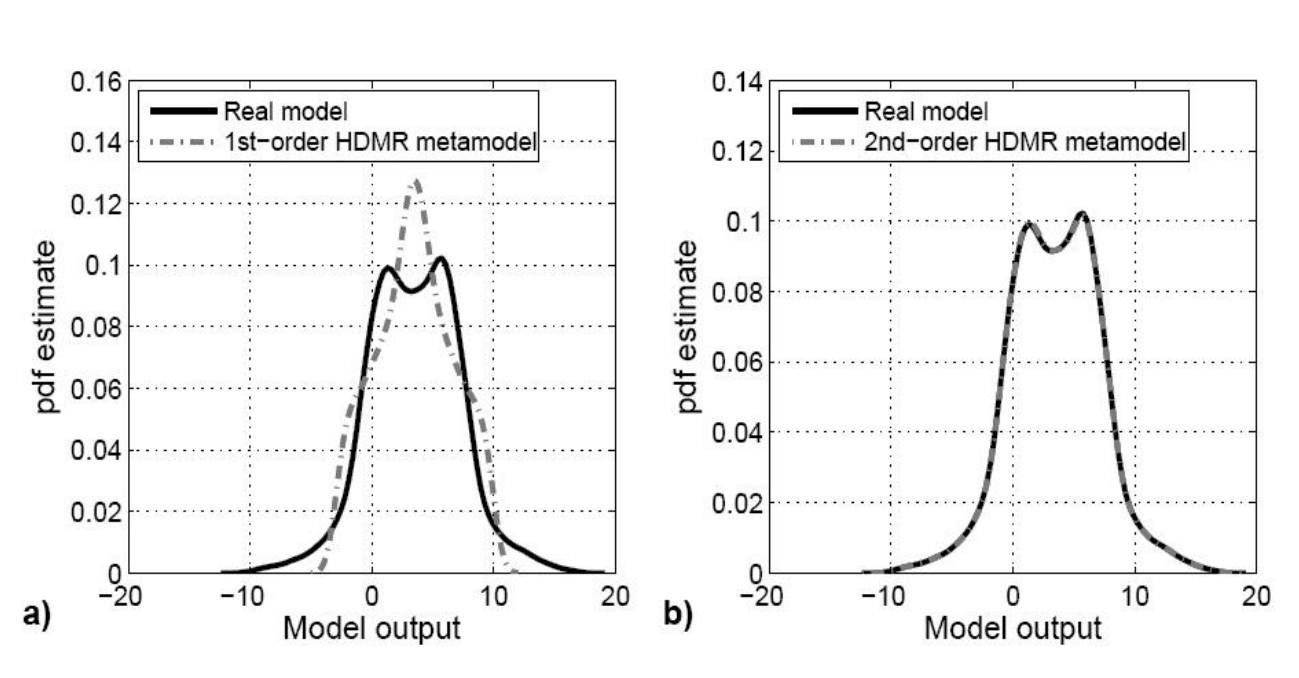
Variance reduction methods

- When using sample based methods such as Monte Carlo or quasi-random sampling we are approximating the integrals using a discrete sample.
- Increasing sample size will of course reduce the Monte Carlo integration errors.
- Often we cannot afford this from the point of view of computational cost.
- Another way to improve the accuracy of the Monte Carlo integration is to reduce the variance of the integrand.
- Two methods used in GUI-HDMR
  - *the correlation method*
  - *the ratio control variate method.*
- In both cases the determination of the expansion coefficients becomes an iterative process and requires an analytical reference function $h(\mathbf{x})$.
- This function has to be similar to $f(\mathbf{x})$ and as shown by Li et al. (2003) and Li and Rabitz (2006) a truncated RS-HDMR expansion can be used as a reference function whose expansion coefficients were calculated by direct MC integration.


GUI-HDMR Vers. 1.0 - Setup
File
Help
Sample files
Sample input-file (for HDMR analyis):
8 Inputs and 10000 Samples
Sample output-file (for HDMR analysis):
1 Output and 10000 Samples
Sample input-file (optional, for accuracy test):
no file loaded yet
Sample output-file (optional, for accuracy test):
no file loaded yet
Input ranges (optional, for plots):
no file loaded yet
Settings
Number of samples to use for HDMR:
1024
Number of samples to use for accuracy test + scatter plots:
1000
Max order for approximation of 1st-order component functions :
10
Max order for approximation of 2nd-order component functions :
5
Variance reduction method :
ratio control
Number of iterations for 1st-order component functions:
10
Number of iterations for 2nd-order component functions:
10
Do you want to use a threshold?
no
Default
OK

Testing accuracy

- Several methods could be used. What is important is to ensure that the output distribution of the meta-model is well matched to that of the distribution from the full model runs.

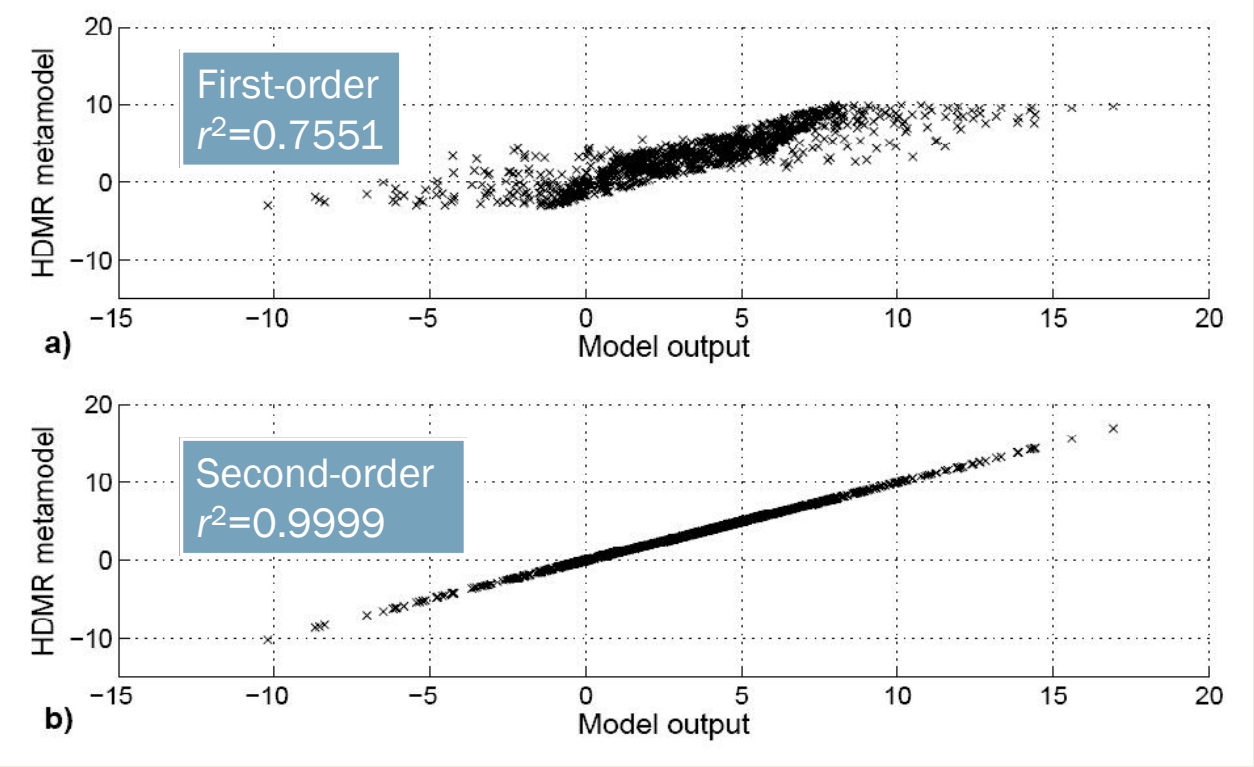


Here the first-order model is poor but adding second-order terms gives a good fit.

Plots available in GUI-HDMR.

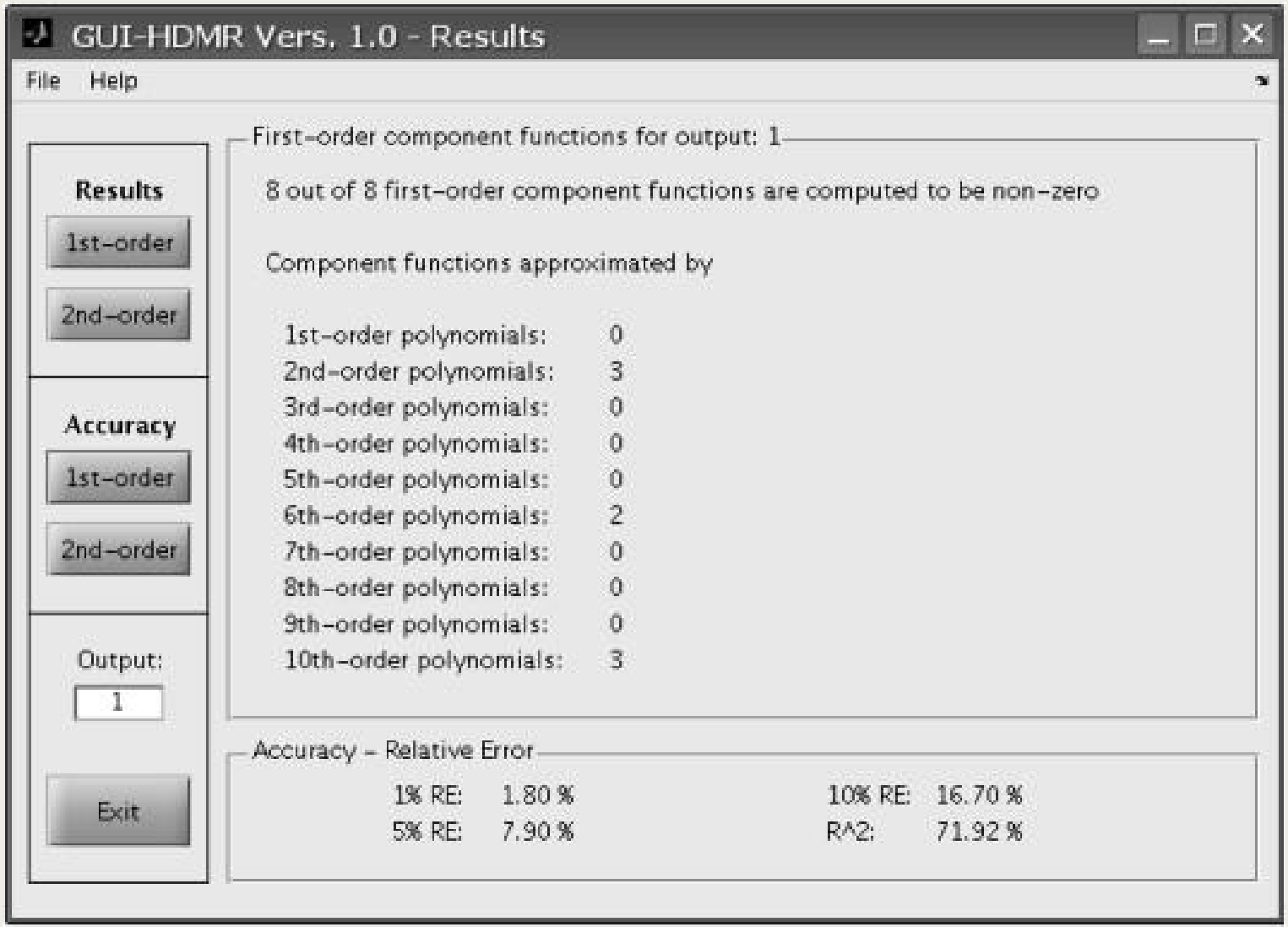
Testing accuracy

- Scatter plots of full model vs. meta-model are also available.



A trend in the residuals for the scatter plot for the first-order model is often a sign that higher order effects are present.

$$\mathrm{RE} = \left| \frac{f(\mathbf{x}^{(s)}) - \hat{f}(\mathbf{x}^{(s)})}{f(\mathbf{x}^{(s)})} \right|$$


GUI-HDMR Vers. 1.0 - Results
File
Help
Results
1st-order
2nd-order
Accuracy
1st-order
2nd-order
Output:
1
Exit
First-order component functions for output: 1
8 out of 8 first-order component functions are computed to be non-zero
Component functions approximated by
1st-order polynomials: 0
2nd-order polynomials: 3
3rd-order polynomials: 0
4th-order polynomials: 0
5th-order polynomials: 0
6th-order polynomials: 2
7th-order polynomials: 0
8th-order polynomials: 0
9th-order polynomials: 0
10th-order polynomials: 3
Accuracy - Relative Error
1% RE: 1.80 %
5% RE: 7.90 %
10% RE: 16.70 %
R^2: 71.92 %

Calculation of global sensitivity indices

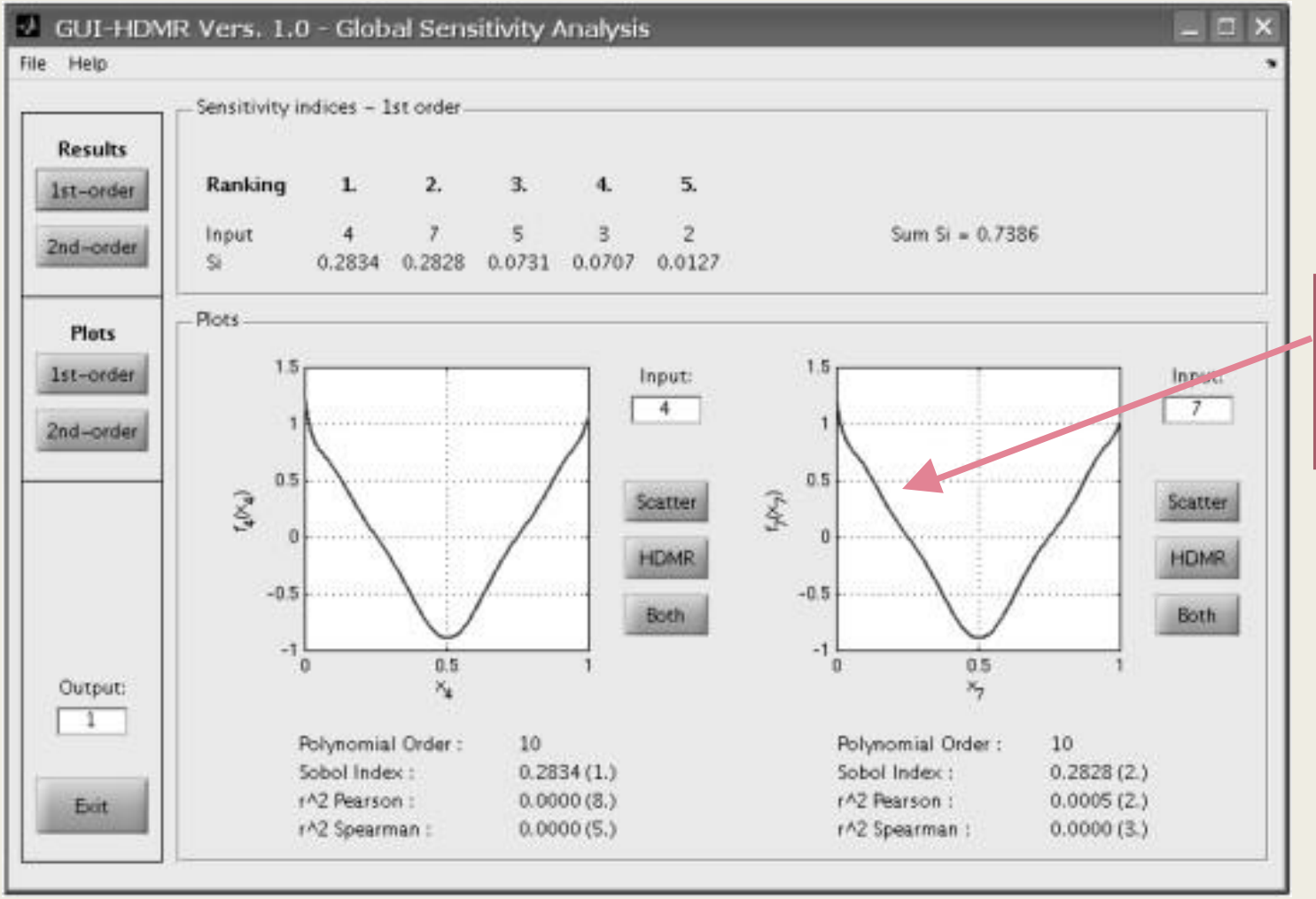
- The partial variances D_i and D_{ij} can be calculated (Li et al. 2002b; Feng et al. 2004):

$$D_i = \sum_{r=1}^{k_i} \left(\sim_r^i \right)^2$$

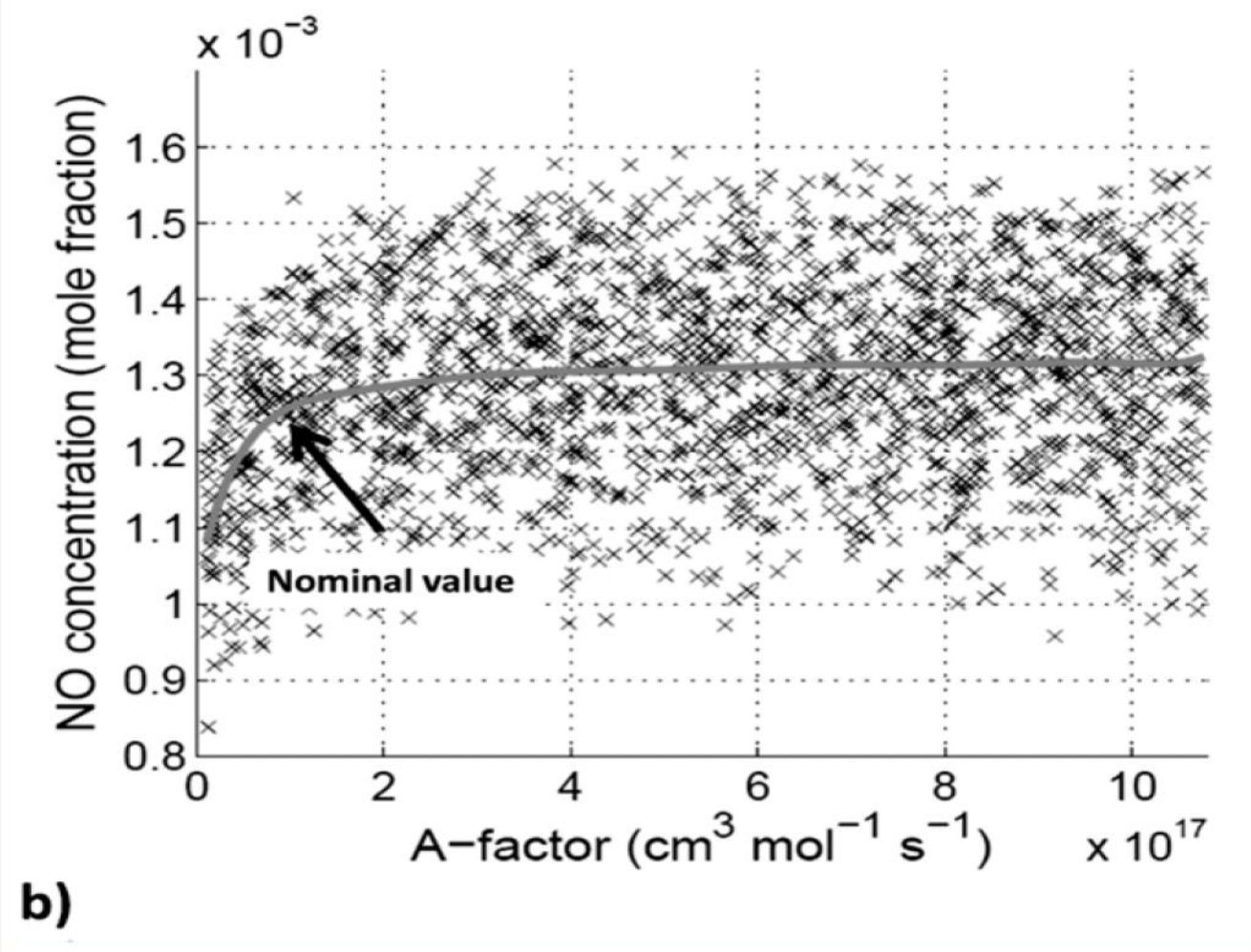
$$D_{ij} = \sum_{r=1}^{l_i} \sum_{q=1}^{l_j'} \left(\begin{smallmatrix} ij \\ rq \end{smallmatrix} \right)$$

Dividing these partial variances by the total variance of the model output, sensitivity indices equivalent to the Sobol' indices can be calculated:

$$S_{i_1,\ldots,i_s} = \frac{D_{i_1,\ldots,i_s}}{D}, \qquad 1 \le i_1 < \cdots < i_s \le m$$



Component functions in the expansion indicate the individual effect of each parameter



Here the component function shows the impact of a particular reaction A-factor within the scatter caused by uncertainties in other important parameters.

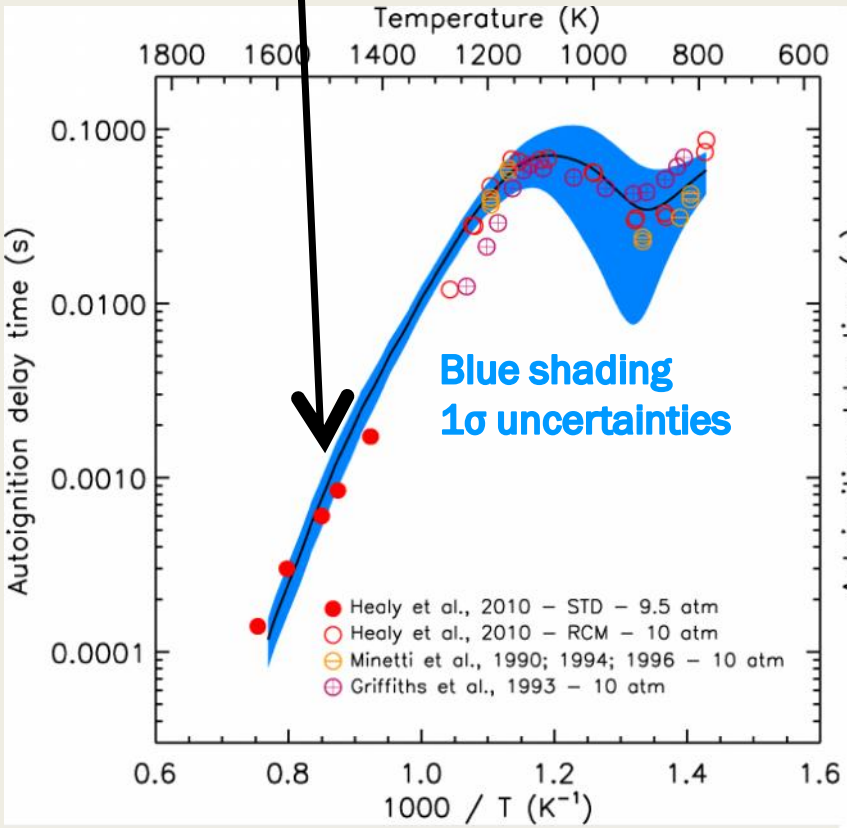
If S_i for a parameter was 1 then there would be no scatter.

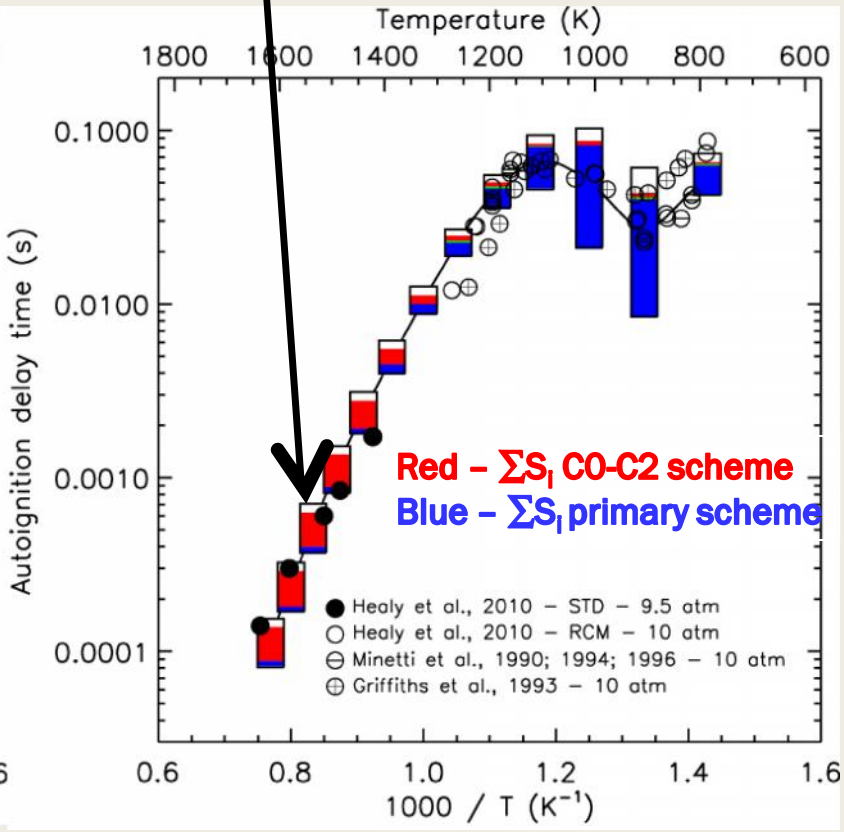
Very rare – but indicates a very high information content for a particular experimental target measurement.

Ignition delays butane: sources of uncertainties

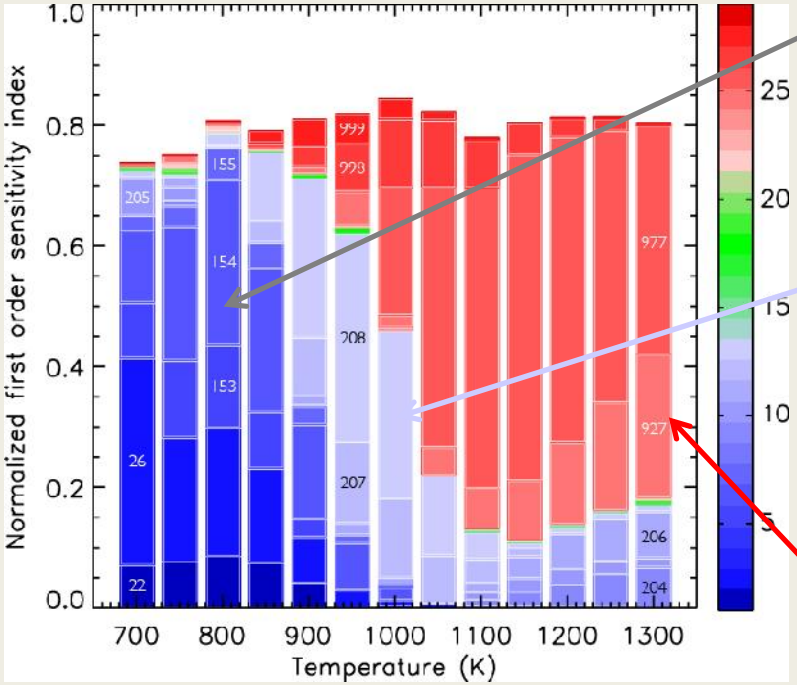
Uncertainties lowest at higher temperatures.

Dominated by sensitivity to reactions in C0-C2 base scheme





Ignition delays: 1st-order global sensitivities



Key reactions at low temperatures

$C_4H_9\text{-}1\text{-}OO \rightarrow C_4H_8\text{-}1\text{-}OOH$

$C_4H_9\text{-}1\text{-}OO \rightarrow C_4H_8\text{-}2\text{-}OOH$

$1\text{-}C_4H_9 + O_2 \rightarrow 1\text{-}C_4H_8 + HO_2$

$2\text{-}C_4H_9 + O_2 \rightarrow 1\text{-}C_4H_8 + HO_2$

Key reactions at intermediate temperatures

$C_4H_{10} + HO_2 \rightarrow H_2O_2 + 1\text{-}C_4H_9$

$C_4H_{10} + HO_2 \rightarrow H_2O_2 + 2\text{-}C_4H_9$

Key reactions at high temperatures

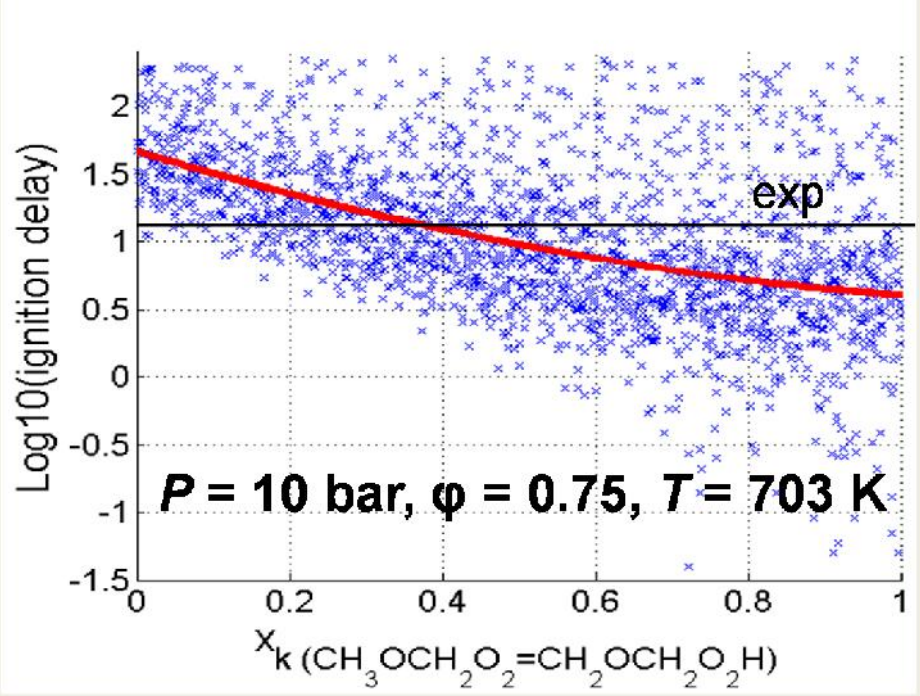
$HO_2 + CH_3 \rightarrow CH_3O + OH$

$O_2 + H \rightarrow OH + O$

$C_4H_{10} + H \rightarrow H_2 + 2\text{-}C_4H_9$

$C_4H_{10} + OH \rightarrow H_2O + 2\text{-}C_4H_9$

What do global sampling based sensitivities tell us about model tuning? Example of ignition delays for DME.

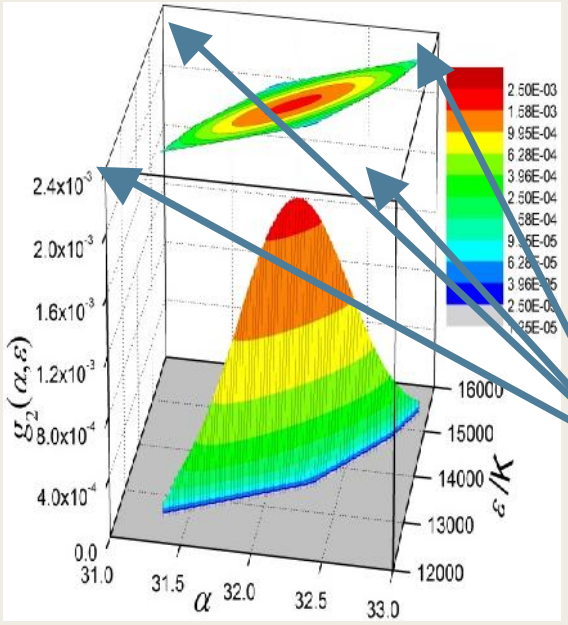


- Temptation to tune isomerisation route to give long enough ignition delays.
- However any rate constant within a factor of 10 would match experiment given other uncertainties.
- Further studies required for branching channels.

CORRELATIONS IN INPUTS

Optimised mechanisms

- The output from a mechanism optimisation procedure based on a wide range of data sets would be the joint pdf of the Arrhenius parameters (see Zsely lectures).
- When propagating uncertainties these need to be accounted for otherwise the output uncertainty can be over exaggerated.



- Correlation matrix (output from optimisation procedure) can be used to generate a probabilistic sample.

If whole space was used then uncertainties would be too large.

Sampling procedure

- Perform a probabilistic sample based on optimised model.
- Need to *de-correlate* the parameters in order to perform HDMR global sensitivity analysis.
- Rosenblatt transformation used to generate uncorrelated sample (Mara and Tarantola, 2012).
- Perform HDMR and generate sensitivity indices.
- HOWEVER, these now represent *marginal indices* with varying degrees of inter-correlation amongst the parameters.
- Valko et al. (2016) have shown for a hydrogen oxidation model that the correlated effects of parameters dominate.
  - *Effects of final marginal indices for uncorrelated effects of parameters are very small.*

	reaction	parameter	S_i^{corr}	$S_i^{corr\_total}$	S_i^{uncorr}	$S_i^{uncorr\_total}$
1	$H+O_2=O+OH$	ln *A*	0.088	0.089	0.110	0.110
2		*n*	0.035	0.036	0.109	0.109
3		*E*/*R*	0.259	0.260	0.099	0.099
4	$H+O_2(+M)=HO_2(+M)$	LP ln *A*	0.012	0.012	0.000	0.000
5		LP *n*	0.000	0.001	0.000	0.000
6		*m*(H_2)	0.171	0.172	0.000	0.000
7		*m*(H_2O)	0.039	0.041	0.000	0.000
8		*m*(Ar)	0.044	0.044	0.001	0.001
9	$O+H_2=H+OH$	ln *A*	0.236	0.237	0.024	0.024
10		*n*	0.229	0.229	0.023	0.023
11		*E*/*R*	0.258	0.258	0.014	0.014
12	$OH+H_2=H+H_2O$	ln *A*	0.158	0.159	0.004	0.004
13		*n*	0.167	0.168	0.005	0.005
14		*E*/*R*	0.190	0.191	0.002	0.002
15	$H+HO_2=H_2+O_2$	ln *A*	0.149	0.150	0.011	0.011
16		*n*	0.279	0.280	0.012	0.012
17		*E*/*R*	0.131	0.133	0.004	0.004

Also note effects are not additive. i.e. don't sum to 1.

What does this mean?

- It means that if new experiments are performed which might help to better constrain a hydrogen oxidation scheme then they should be folded into a new optimisation analysis.
  - *i.e. not used to update parameters individually.*
- A parallel with the Active Thermochemical Tables approach can be drawn.
  - *The whole TN is updated as new data is added.*
- Time consuming for reaction model since MANY simulations go into an optimisation procedure.

Summary of lecture 2-3

- Various methods for estimating global sensitivity indices have been reviewed.
- Each estimates the contribution of each parameter to some measure of the predicted output distribution (often variance).
- Response surface based methods tend to be the most efficient for large parameter systems.
  - *Even so, if higher-order effects are present the required sample size to achieve good accuracy may be large. Much larger than estimating the variance itself.*
- Quasi-random samples tend to converge better.
  - *Convergence can be tested by calculating indices for increasing sample sizes and testing when they converge.*
- The methods allow a ranking of important parameters.
- They also inform us about the information content of experiments i.e. how experiments can help us to better constrain our models.