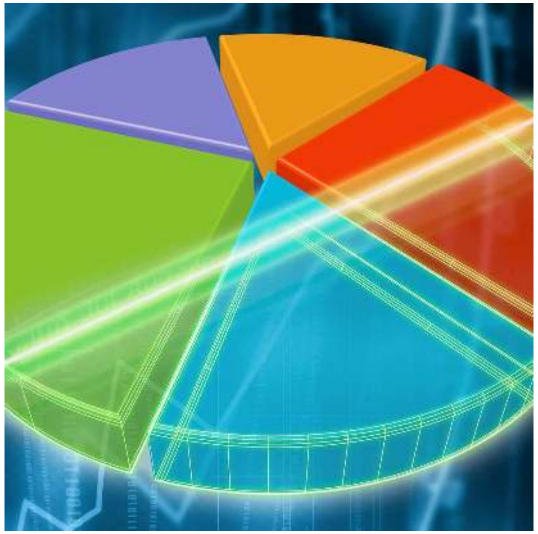
EUROPEAN COMMISSION
DIRECTORATE-GENERAL
Joint Research Centre

Joint Research Centre



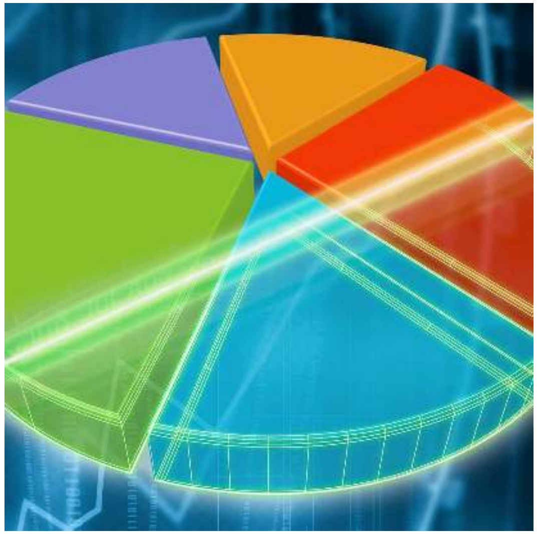
Sensitivity Analysis as an ingredient of modelling

Stefano Tarantola

Joint Research Centre
of the European Commission

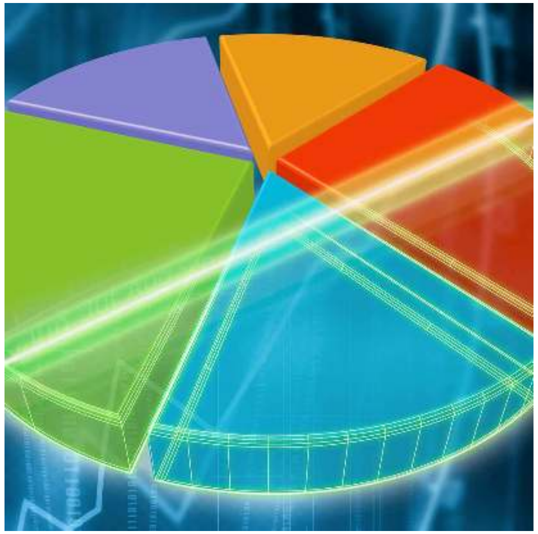
Cadarache
January, 12 2010

Plan of the presentation



- Models and uncertainties
- Can SA assist modellers?
- Variance-based methods

Plan of the presentation



•Models and uncertainties

Models and scientific uncertainty

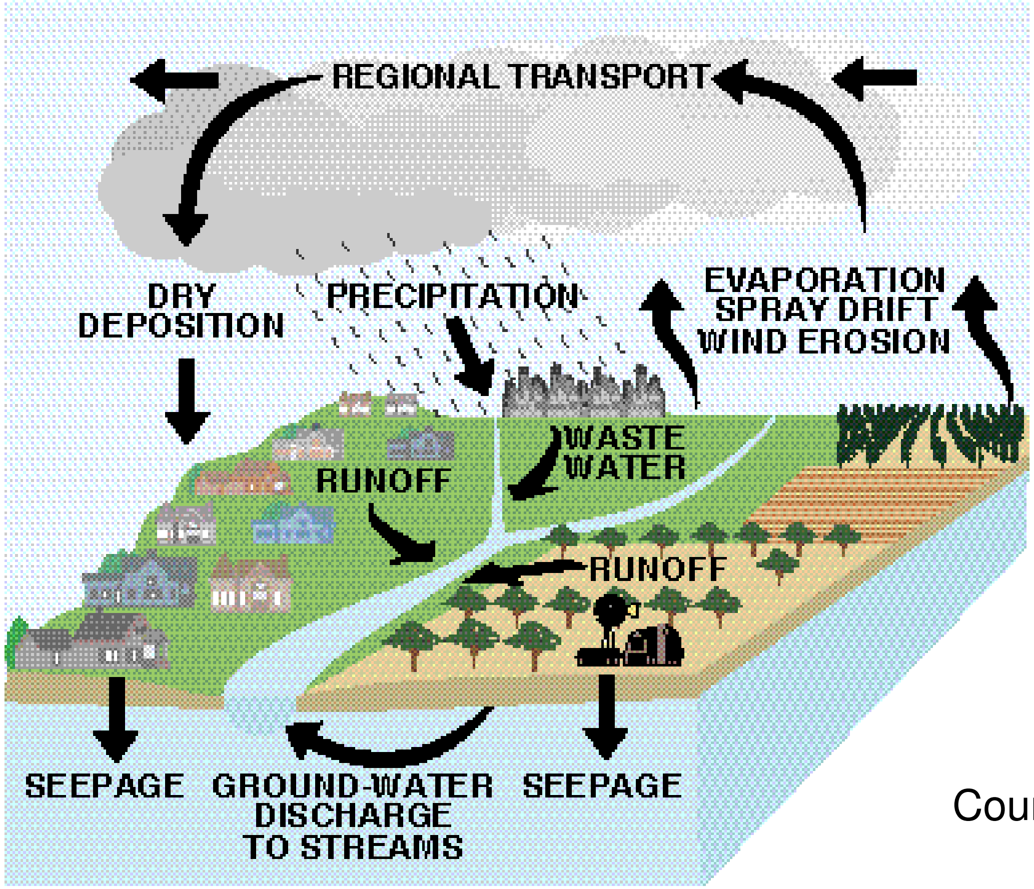
Simulation models are used (in diagnostic or prognostic fashion) in many fields to understand complex phenomena (natural or social) and consequently as tools to support decisions and policy.

Knowledge base is often flawed by uncertainties (partly irreducible, largely unquantifiable), imperfect understanding, subjective values.

A few examples …

Joint Research Centre

Models in hydrology



Courtesy of USGS

Uncertainties in model parameters that govern surface and ground water transport, …

Joint Research Centre

Models in biology

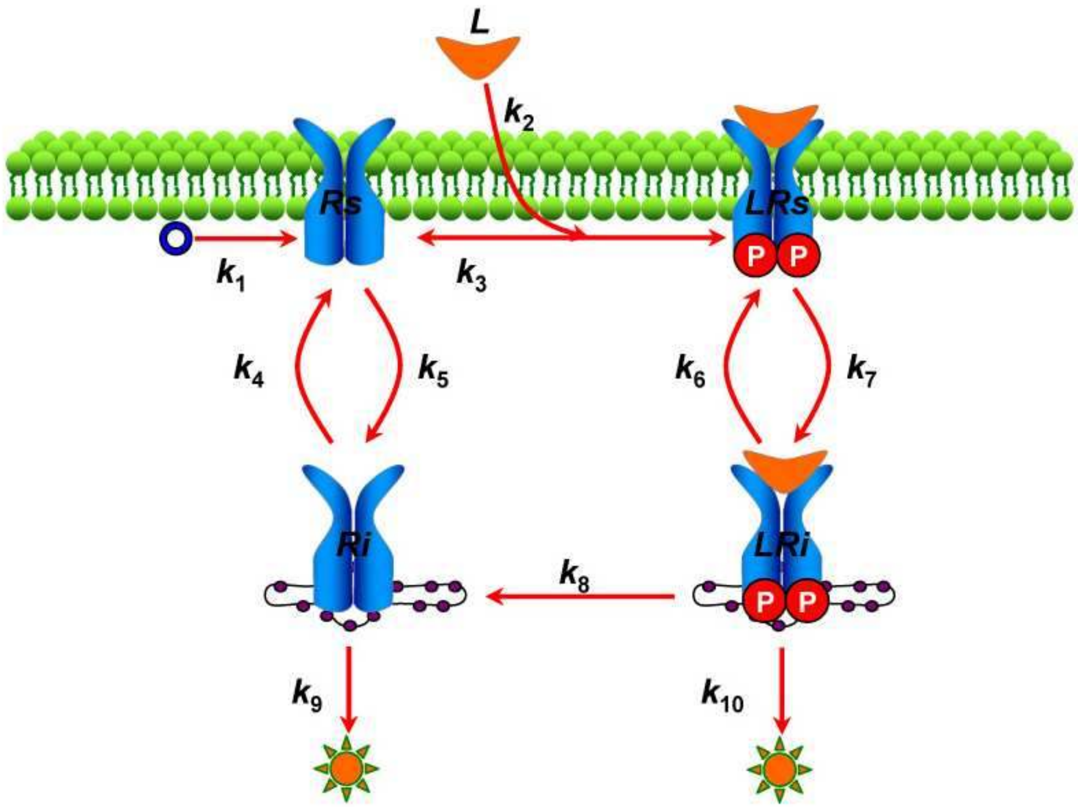
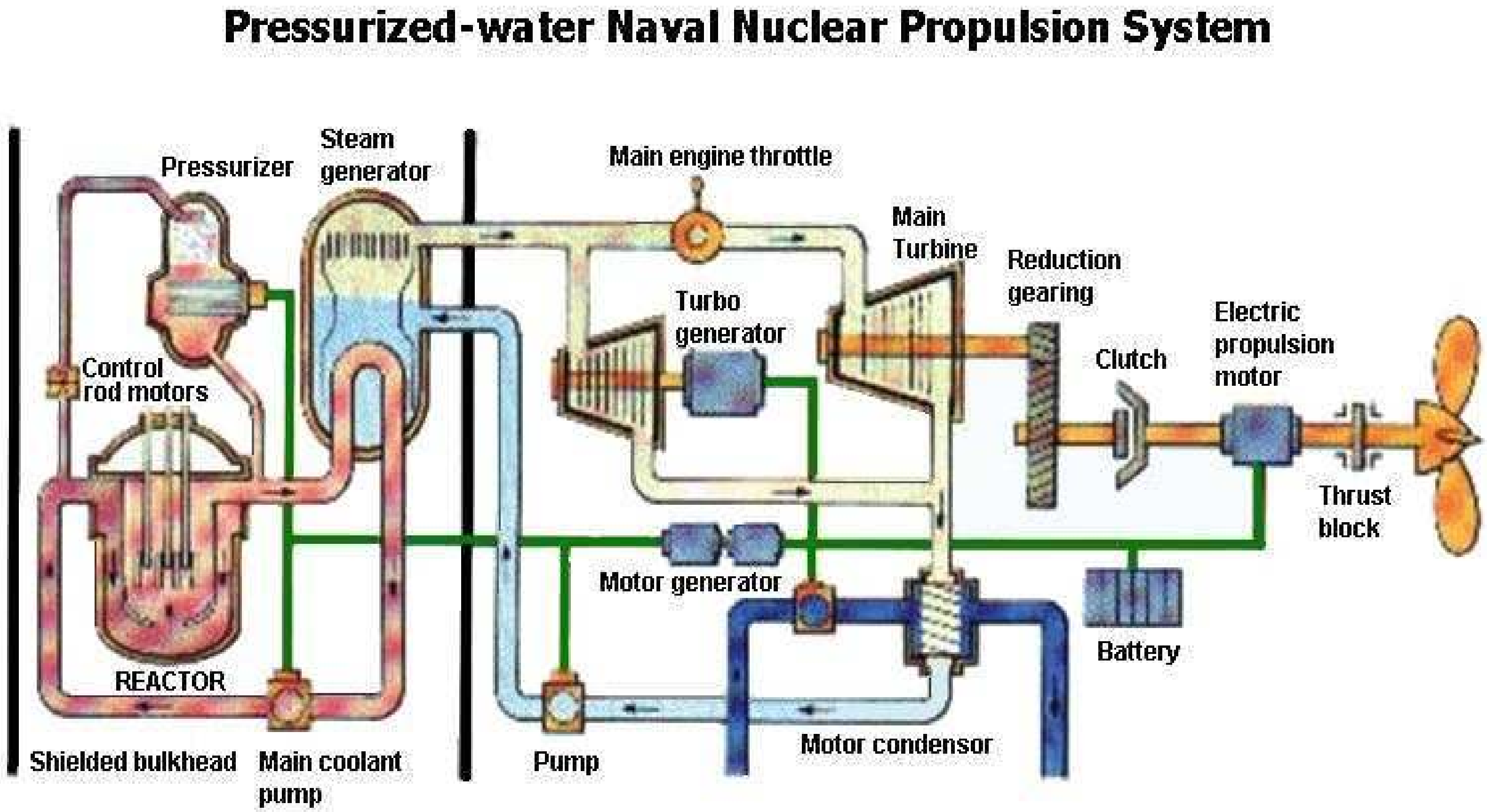


Figure 5
Scheme of receptor trafficking network model. Schematic description of the receptor trafficking network. The symbols *L, Rs, LRs, Ri, LRi* represent the ligand, unbound cell surface receptor, cell surface ligand-receptor complex, internalized unbound receptor and internalized ligand-receptor complex, respectively. The parameter information is listed in Table 1.

… Uncertainties of kinetic parameters in a chemical process…

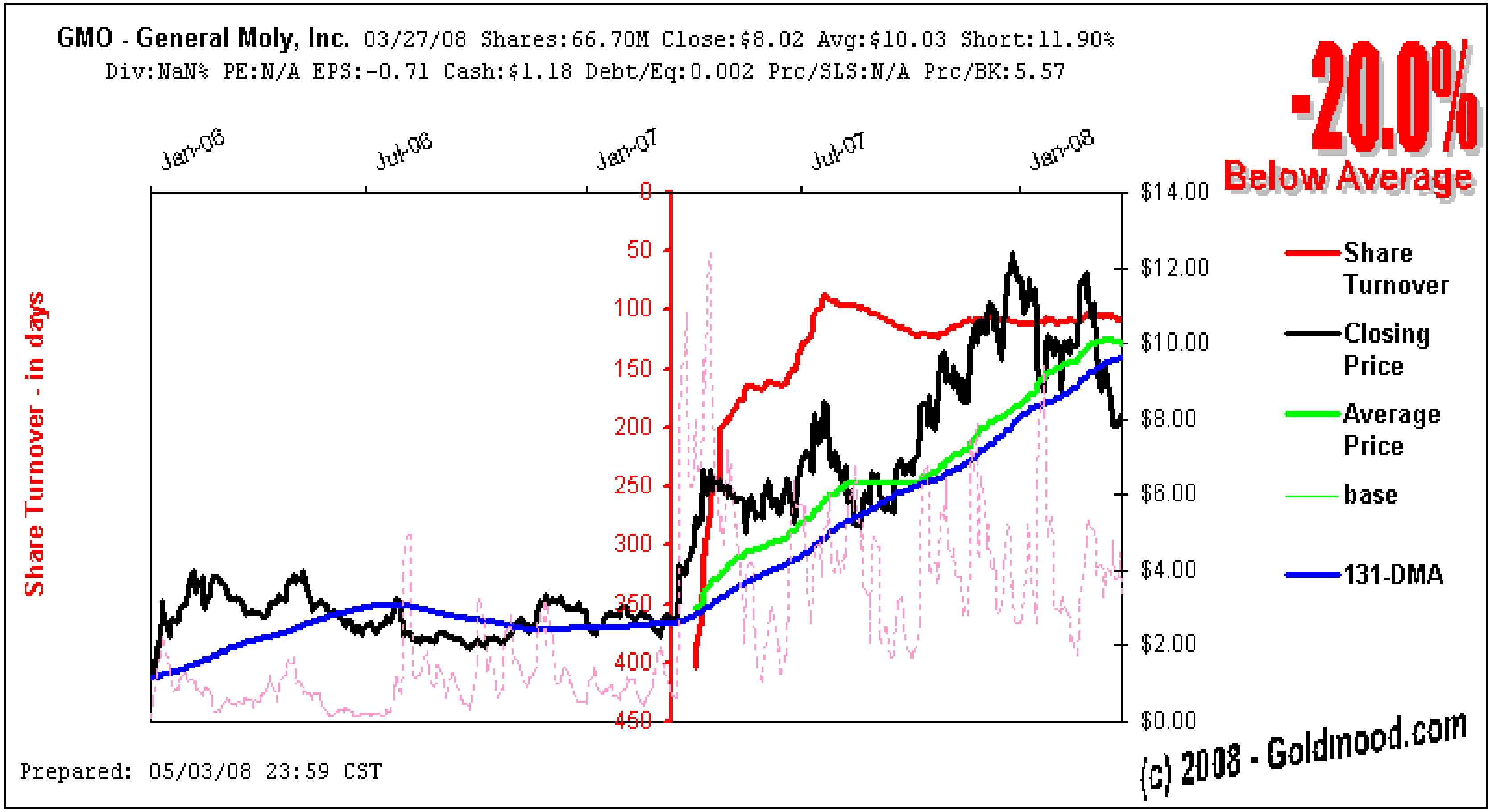
Ex: biological model
© 2008 Zi et al;
licensee BioMed Central Ltd.

Models in nuclear engineering



Uncertainties in reactor physics and thermodynamics of fluids …

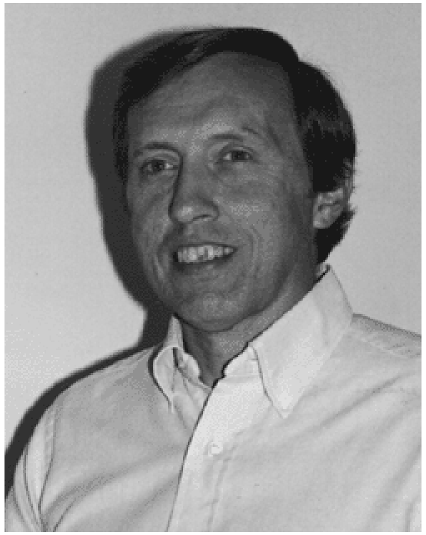
Models in finance



… or in the parameters regulating stock prices forecasts

The critique of models

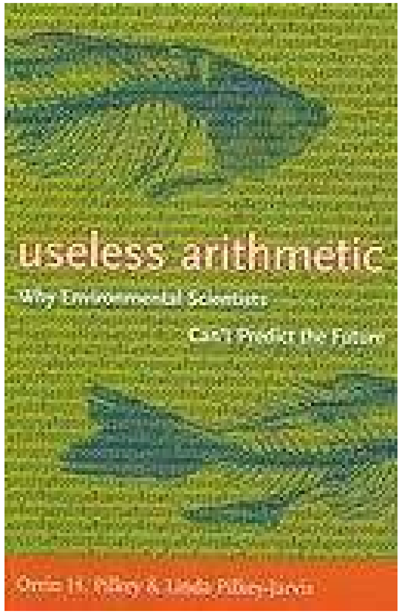
<<[…] most simulation models will be complex, with many parameters, state-variables and non linear relations. Under the best circumstances, such models have many degrees of freedom and, with judicious fiddling, can be made to produce virtually any desired behaviour, often with both plausible structure and parameter values.>>



HORNBERGER and Spear (1981).

Problem of credibility of models.

Specification of the model inputs: an example

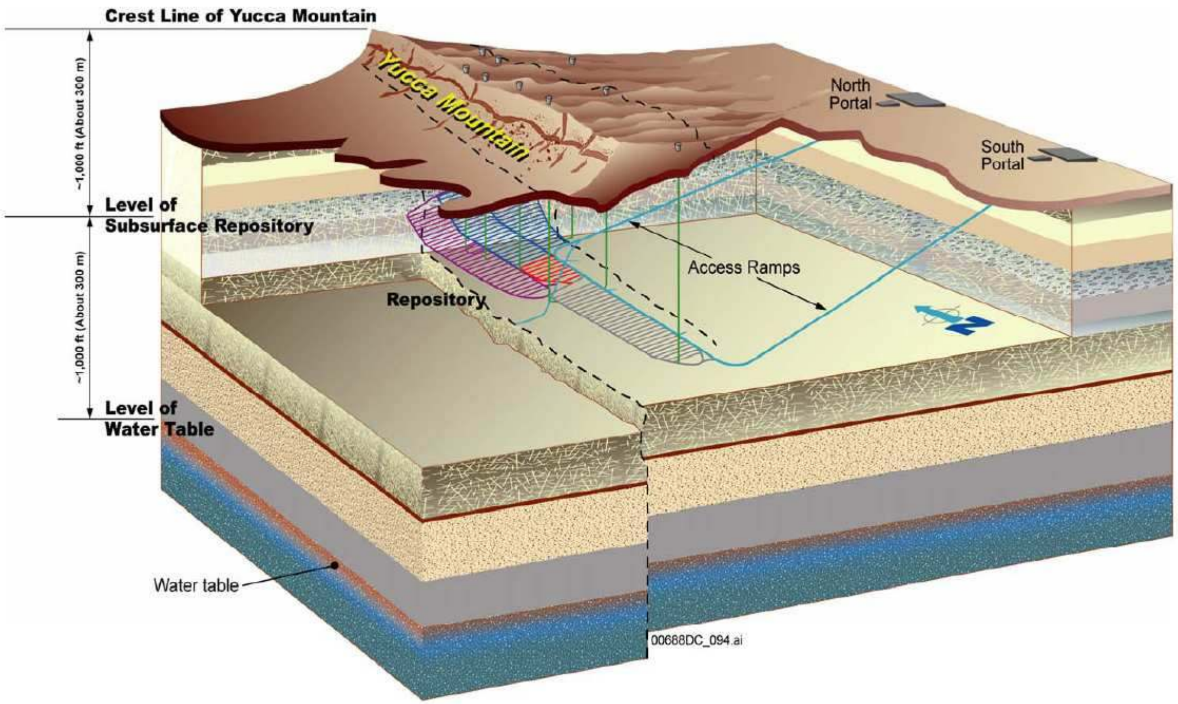


Useless Arithmetic: Why Environmental Scientists Can't Predict the Future
by Orrin H. Pilkey and Linda Pilkey-Jarvis

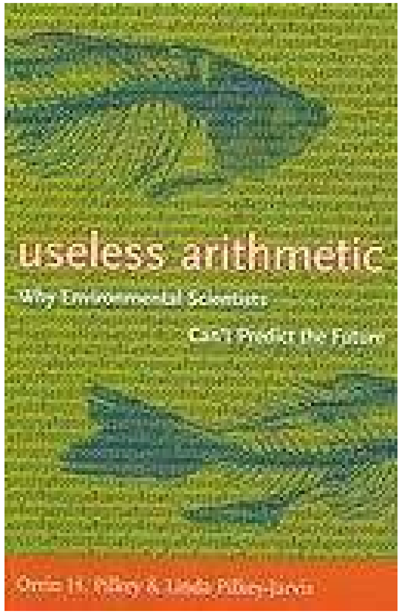
Mathematical models used by policy makers and government administrators
to shape environmental policies
are seriously flawed

Specification of the model inputs: an example

One of the examples discussed concerns the Yucca Mountain repository for radioactive waste disposal, where a very large model called TSPA (total system performance assessment) is used to guarantee the safe containment of the waste. TSPA is composed of 286 sub-models.



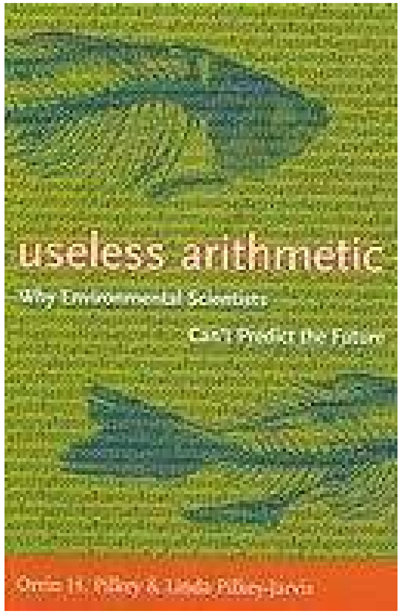
Specification of the model inputs: an example



TSPA (like any other model) relies on assumptions -- a crucial one being the low permeability of the geological formation and hence the long time needed for the water to percolate from the desert surface to the level of the underground disposal.

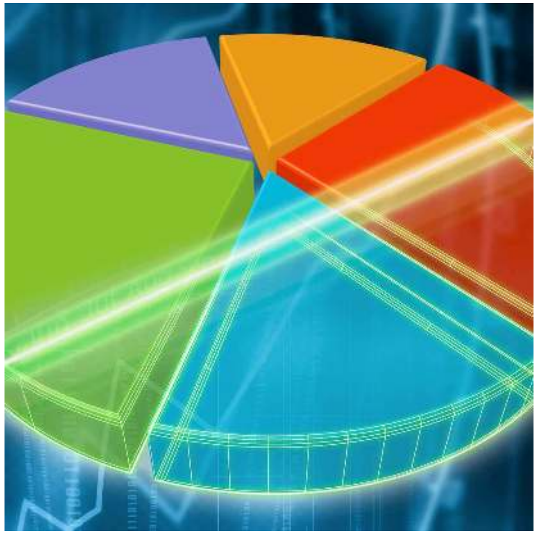
The confidence of the experts in TSPA was not helped when evidence was produced which could lead to an upward revision of 4 orders of magnitude of this parameter.

The critique of models



Stakeholders and media alike tend to expect or suspect instrumental use of computational models, amplification or dampening of uncertainty as a function of convenience.

Note: This book has a good state of the art on the sea level rise story.



•Can SA assist modellers?

Can SA assist modellers?

We need tools to scrutinize uncertainties in model inputs, assumptions, models structures, to see how they propagate and affect inferences (that are used for policy decisions).

Sensitivity analysis: “The study of how the uncertainty in the output of a model can be apportioned to different sources of uncertainty in the model input”.

Joint Research Centre

Can SA assist modellers?

Sensitivity Analysis decomposes the uncertainty in inference (policy conclusions) to uncertainty in inputs to identify which inputs are relevant for the prediction

and then investigate how their uncertainty can be reduced in order to improve the accuracy of the prediction.

Properly executed SA can help to …

Can SA assist modellers?

-gauge model adequacy and relevance (B. Beck),

-check whether policy options are distinguishable from one another given the uncertainties

-establish priorities for research,

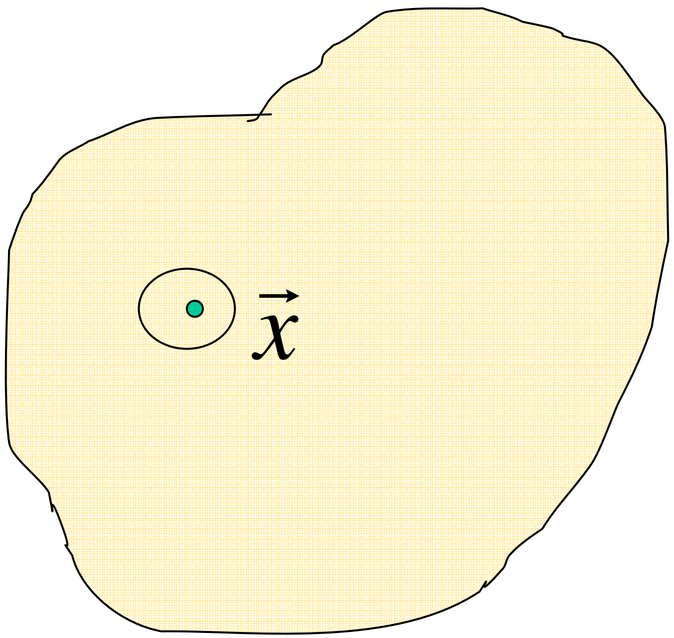
-discover interactions among inputs, simplify models

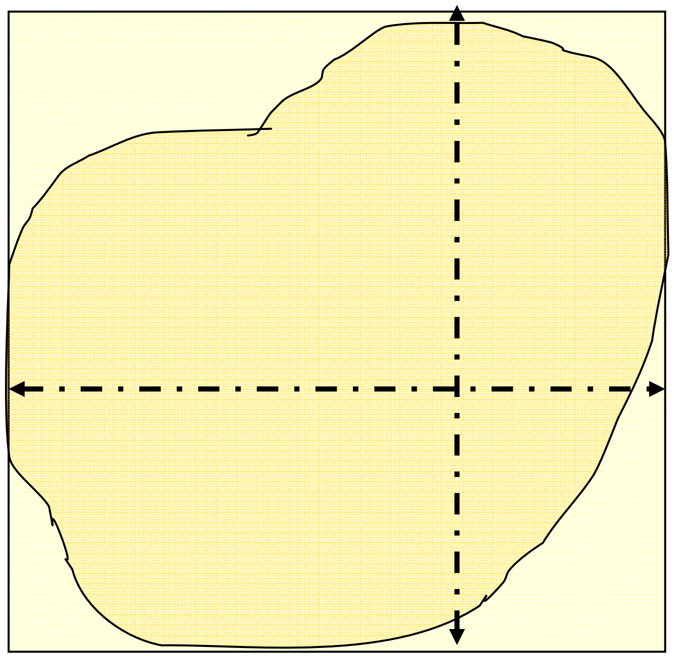
-facilitate parameter estimation and robust prediction

-identify critical regions in the input space

Joint Research Centre

Local techniques vs. Global approaches





$\vec{x}$ = nominal value

Prescriptions for sensitivity analysis…

Prescription #1: Use of global tools for SA

to vary the uncertain factors over a finite range;

to explore simultaneously all sources of uncertainty;

and ensure that we can capture possible interactions between factors.

Prescriptions for sensitivity analysis…

Prescription #2: Use of model-free tools

Real models are often non linear and/or non additive

Local SA, linear regression, correlation techniques capture only part of the information

Prescriptions for sensitivity analysis…

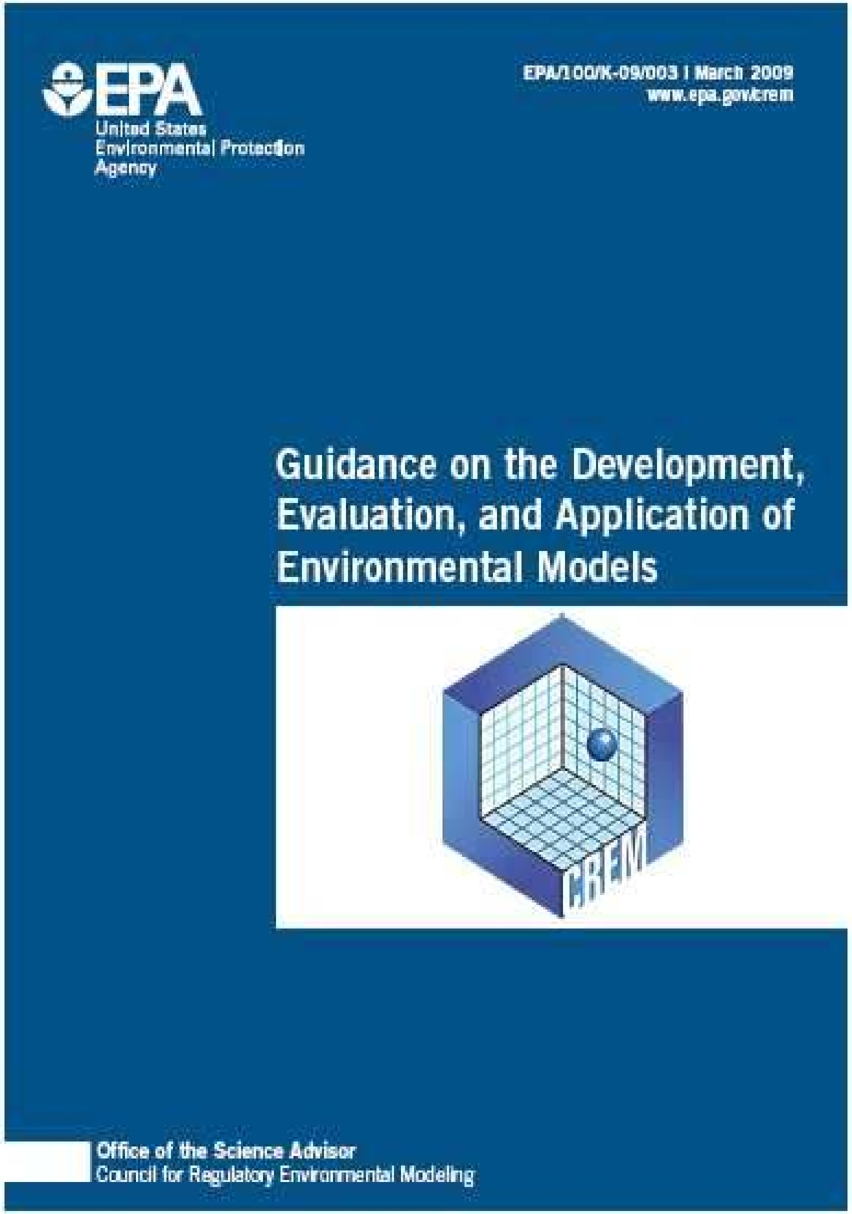
European Commission

IMPACT ASSESSMENT GUIDELINES

15 January 2009
SEC(2009) 92

"**…** Sensitivity analysis can be used to explore how the impacts of the options you are analysing would change in response to variations in key parameters and how they interact."

Prescriptions for sensitivity analysis…



" [SA] methods should preferably be able to deal with a model regardless of assumptions about a model's linearity and additivity, consider interaction effects among input uncertainties, […], and evaluate the effect of an input while all other inputs are allowed to vary as well."

Can SA assist modellers?

Unfortunately, most of the sensitivity analysis that we encounter in practice are local or one-at-a-time (OAT):

J. Campbell, *et al.*, *Science* 322, 1085 (2008).
R. Bailis, M. Ezzati, D. Kammen, *Science* 308, 98 (2005).
E. Stites, P. Trampont, Z. Ma, K. Ravichandran, *Science* 318, 463 (2007).
J. Murphy, et al., *Nature* 430, 768-772 (2004).
J. Coggan, *et al.*, *Science* 309, 446 (2005).

These designs poorly explore the input space.

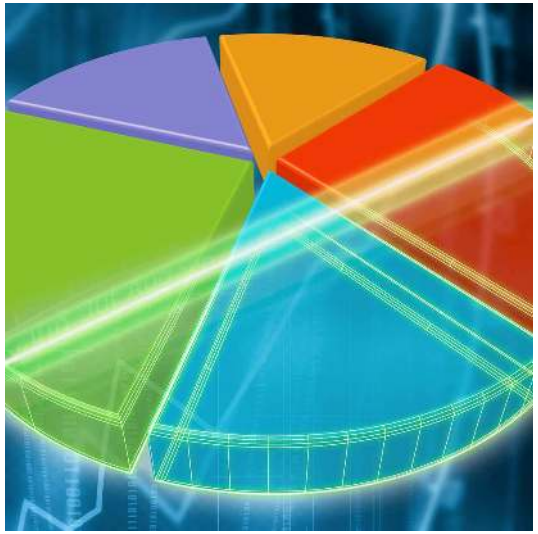
Can SA assist modellers?

I will introduce one broad class
of global sensitivity methods capable of:

-Exploring uncertain inputs over a range

-highlighting interactions among uncertain inputs

-Applicable to any kind of model



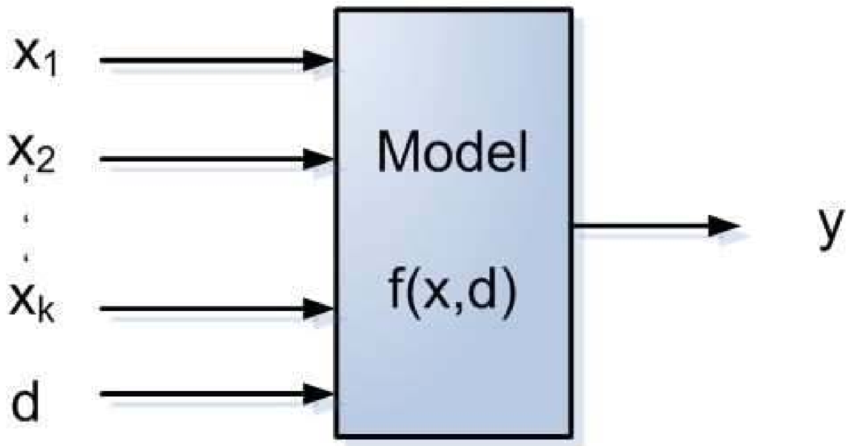
•Variance-based methods

Consider a deterministic model $y = f(\mathbf{x}, \mathbf{d})$ where:

y is the model output,

$\mathbf{x} = (x_1, x_2, \ldots, x_k)$ is a vector of (uncertain) model inputs (k is the dimensionality of the space)

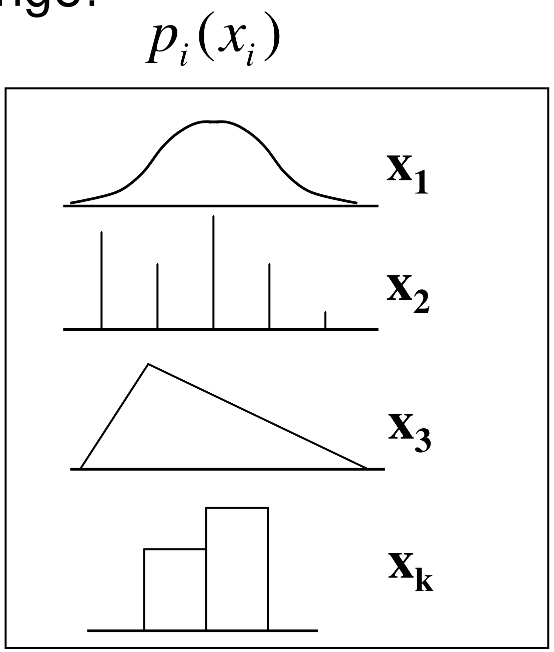
d is a vector of known (fixed) model inputs.



Each model input x_i is treated as if it were a random variable, i.e its uncertainty is characterized through a probability density function (pdf).

Characterizing uncertainties is a challenge!

- Scientific literature
- Physical bounds
- Experiments
- Expert judgement
- Opinion polls, surveys



Ref. Helton et al. (1993) RESS

Example

Table 1. List of Input Factors for the Level E Model

Notation	Definition	Distribution	Range	Units
T	Containment time	Uniform	[100, 1000]	yr
k_I	Leach rate for iodine	Log-uniform	$[10^{-3}, 10^{-2}]$	mols/yr
k_C	Leach rate for *Np* chain nuclides	Log-uniform	$[10^{-6}, 10^{-5}]$	mols/yr
$v^{(1)}$	Water velocity in geosphere's 1st layer	Log-uniform	$[10^{-3}, 10^{-1}]$	m/yr
$l^{(1)}$	Length of geosphere's 1st layer	Uniform	[100, 500]	m
$R_I^{(1)}$	Retention factor for I (1st layer)	Uniform	[1, 5]	—
$R_C^{(1)}$	Factor to compute retention coefficients for *Np* (1st layer)	Uniform	[3, 30]	—
$v^{(2)}$	Water velocity in geosphere's 2nd layer	Log-uniform	$[10^{-2}, 10^{-1}]$	m/yr
$l^{(2)}$	Length of geosphere's 2nd layer	Uniform	[50, 200]	m
$R_I^{(2)}$	Retention factor for I (2nd layer)	Uniform	[1, 5]	—
$R_C^{(2)}$	Factor to compute retention coefficients for *Np* (2nd layer)	Uniform	[3, 30]	—
W	Stream flow rate	Log-uniform	$[10^{5}, 10^{7}]$	m^3/yr
C_I^0	Initial inventory for ^{129}I	Constant	100	mols
C_{Np}^0	Initial inventory for ^{237}Np	Constant	1000	mols
C_U^0	Initial inventory for ^{233}U	Constant	100	mols
C_{Th}^0	Initial inventory for ^{229}Th	Constant	1000	mols
w	Water ingestion rate	Constant	.73	m^3/yr
β_I	Ingestion-dose factor for ^{129}I	Constant	56	Sv/mols
β_{Np}	Ingestion-dose factor for ^{237}Np	Constant	6.8×10^3	Sv/mols
β_U	Ingestion-dose factor for ^{233}U	Constant	5.9×10^3	Sv/mols
β_{Th}	Ingestion-dose factor for ^{229}Th	Constant	1.8×10^6	Sv/mols

We assume that each input is uniformly distributed over a unit interval. Hence:

$$p(x_i) = 1$$

All the integrals can be written without integration limits

$$\int_{\Omega} x_i p(x_i) dx_i \equiv \int x_i dx_i$$

$$y = f(\mathbf{x}) = f_0 + \sum_{i=1}^{k} f_i(x_i) + \sum_i \sum_{j>i} f_{ij}(x_i, x_j) + \ldots + f_{1,2,\ldots,k}(x_1, x_2, \ldots, x_k)$$

with: $f(\mathbf{x}) \in L^2(\mathbf{x}) \quad \mathbf{x} \in [0;1]^k$

An example in 3 dimensions:

$$\begin{aligned} y = f_0 &+ f_1(x_1) + f_2(x_2) + f_3(x_3) \\ &+ f_{12}(x_1, x_2) + f_{13}(x_1, x_3) + f_{23}(x_2, x_3) \\ &+ f_{123}(x_1, x_2, x_3) \end{aligned}$$

There are infinite ways to build such expansion →

An example in 2 dimensions: $f(\mathbf{x}) = 3 + \log(1 + x_1) + x_2^3$

a) $f_0 = 3;\quad f_1(x_1) = \log(1 + x_1)\quad f_2(x_2) = x_2^3\quad f_{12}(x_1, x_2) = 0$

b) $f_0 = 5;$

$$f_1(x_1) = \log(1 + x_1) - 2x_1$$

$$f_2(x_2) = x_2^3 - \sqrt{x_2}$$

$$f_{12}(x_1, x_2) = 2x_1 + \sqrt{x_2} - 2$$

Properties

$$y = f(\mathbf{x}) = f_0 + \sum_{i=1}^{k} f_i(x_i) + \sum_i \sum_{j>i} f_{ij}(x_i, x_j) + \ldots + f_{1,2,\ldots,k}(x_1, x_2, \ldots, x_k)$$

IF we choose each term in the HDMR such that ...

$$\int f_{i_1 i_2 \ldots i_s}(x_{i_1}, x_{i_2}, \ldots, x_{i_s})\, dx_j = 0 \quad \forall j = i_1, i_2, \ldots, i_s$$

Properties

THEN the HDMR has the following properties:

1) $$\int f(\mathbf{x})\, d\mathbf{x} = f_0$$

f_0 = mean value of f(**x**)

2) Any pair of terms in the HDMR is orthogonal:

$$\int f_{i_1,\ldots,i_s} f_{j_1,\ldots,j_l}\, d\mathbf{x} = 0 \qquad \text{for } (i_1,\ldots,i_s) \neq (j_1,\ldots,j_l)$$

3) the HDMR is unique →

Computing the terms

$$
\begin{aligned}
y = f(\mathbf{x}) = f_0 &+ f_1(x_1) + f_2(x_2) + f_3(x_3) \\
&+ f_{12}(x_1, x_2) + f_{13}(x_1, x_3) + f_{23}(x_2, x_3) \\
&+ f_{123}(x_1, x_2, x_3)
\end{aligned}
$$

$$
\begin{aligned}
f_1(x_1) &= \int f(\mathbf{x})dx_2 dx_3 - f_0 \\
&= E(y \mid x_1) - f_0
\end{aligned}
$$

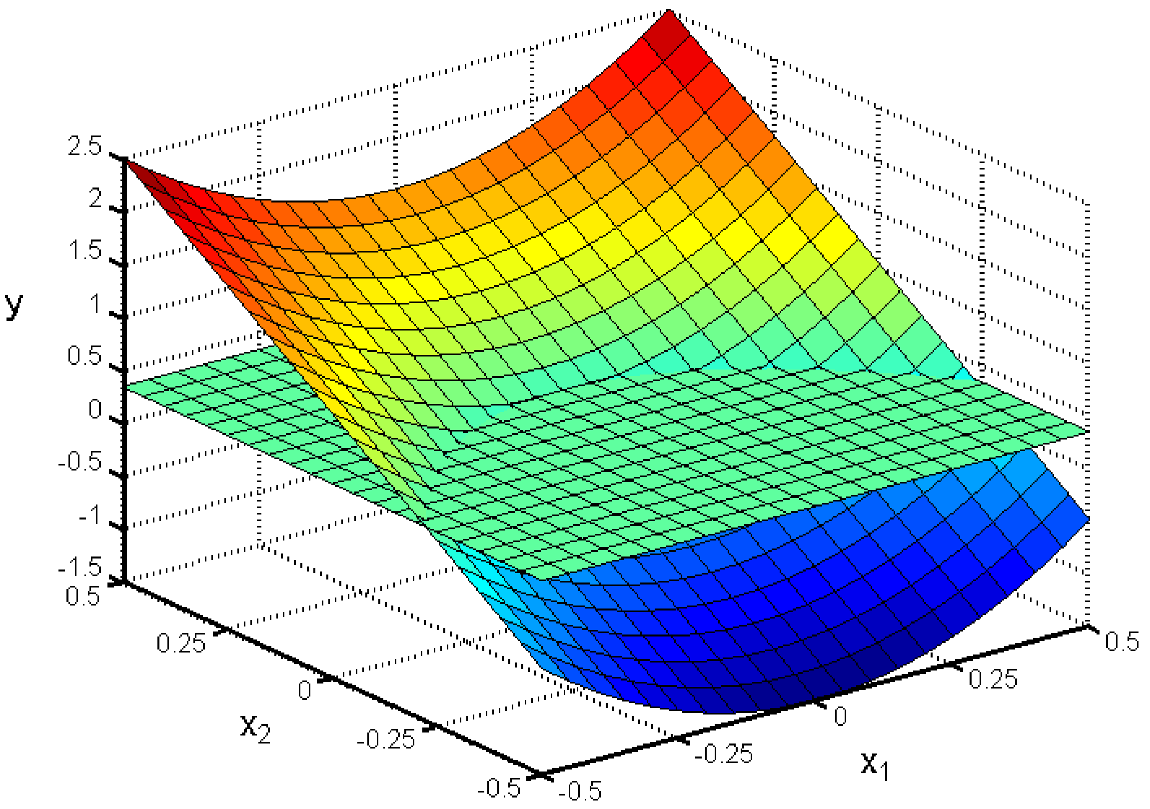
$$
f_2(x_2) = E(y \mid x_2) - f_0
$$

$$
\begin{aligned}
f_{12}(x_1, x_2) &= \int f(\mathbf{x})dx_3 - f_1(x_1) - f_2(x_2) - f_0 \\
&= E(y \mid x_1, x_2) - E(y \mid x_1) - E(y \mid x_2) + f_0
\end{aligned}
$$

Example →

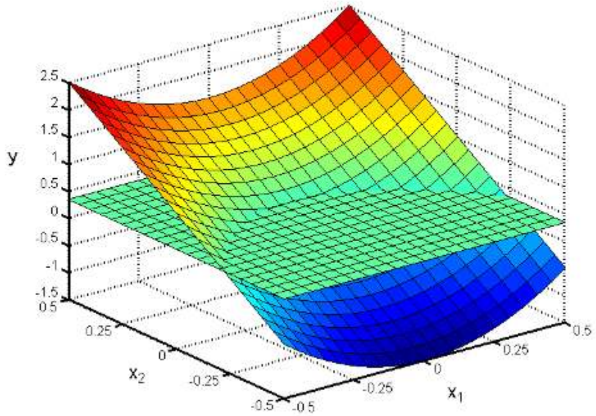
An example

$$f(x_1, x_2) = 4x_1^2 + 3x_2 \qquad x_1, x_2 \in U[-1/2;1/2]$$



HDMR

An example



$$f(x_1, x_2) = 4x_1^2 + 3x_2$$

$$x_1, x_2 \in U[-1/2; 1/2]$$

~~$f_0 = 0$~~

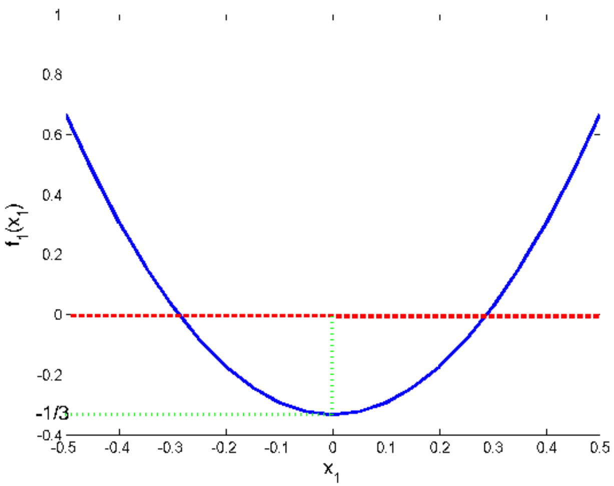
~~$f_1(x_1) = 4x_1^2$~~

~~$f_2(x_2) = 3x_2$~~

~~$f_{12}(x_1, x_2) = 0$~~

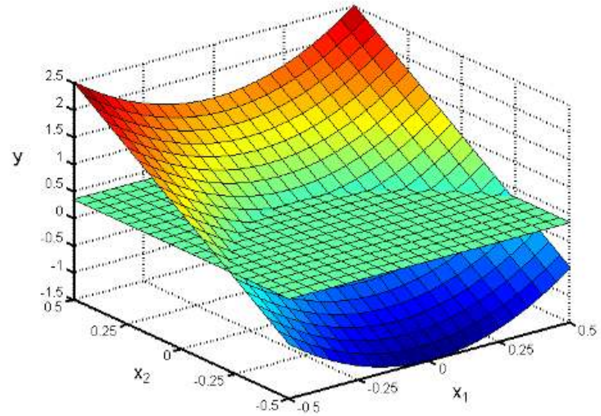
$$f_0 = E(y) = \int_{-1/2}^{1/2}\int_{-1/2}^{1/2}(4x_1^2 + 3x_2)dx_1dx_2 = \frac{1}{3}$$

$$f_1(x_1) = E(y \mid x_1) - f_0 = \int_{-1/2}^{1/2}(4x_1^2 + 3x_2)dx_2 = 4x_1^2 - \frac{1}{3}$$



HDMR

An example



$$f(x_1, x_2) = 4x_1^2 + 3x_2$$

$$x_1, x_2 \in U[-1/2; 1/2]$$

$$f_2(x_2) = E(y \mid x_2) - f_0 = 3x_2$$

$$f_{12}(x_1, x_2) = 0$$

The terms of the HDMR are the building blocks for the SA

Decomposition of variance

$$f(\mathbf{x}) = f_0 + f_1(x_1) + f_2(x_2) + f_{12}(x_1, x_2)$$

Let us define the following quantities:

$$V_1 = \int f_1^2(x_1)dx_1 = Var[f_1(x_1)] = Var[E(y \mid x_1)]$$

$$V_2 = \int f_2^2(x_2)dx_2 = Var[f_2(x_2)] = Var[E(y \mid x_2)]$$

$$V_{12} = \int f_{12}^2(x_1, x_2)dx_1dx_2 = Var[f_{12}(x_1, x_2)] = Var[E(y \mid x_1, x_2)] - V_1 - V_2$$

$$V = \int f^2(\mathbf{x})dx_1dx_2 - f_0^2$$

If x_1 and x_2 are independent: $\Longrightarrow$ $V = V_1 + V_2 + V_{12}$

Sensitivity measures

In k dimensions…

$$V = \sum_{i=1}^{k} V_i + \sum_i \sum_j V_{ij} + \sum_i \sum_j \sum_k V_{ijk} \dots + V_{1,2,\dots,k}$$

Dividing by V we obtain the sensitivity indices

$$1 = \sum_{i=1}^{k} S_i + \sum_i \sum_j S_{ij} + \sum_i \sum_j \sum_k S_{ijk} \dots + S_{1,2,\dots,k}$$

An example with k=4 …

An example with k=4

$$S_1 + S_2 + S_3 + S_4 + S_{12} + S_{13} + S_{14} + S_{23} + S_{24} + S_{34} +$$
$$S_{123} + S_{124} + S_{134} + S_{234} + S_{1234} = 1$$

There are:

4 first order sensitivity measures

6 second order measures

4 third order measures

1 fourth order measure

In total there are 15 terms.

In general for k factors there are 2^k-1 terms

Properties of sensitivity measures

$$1 = \sum_{i=1}^{k} S_i + \sum_i \sum_j S_{ij} + \sum_i \sum_j \sum_k S_{ijk} \ldots + S_{1,2,\ldots,k}$$

$\sum_i S_i \leq 1$ Always

$\sum_i S_i = 1$ The model is additive

$1 - \sum_i S_i$ Indicator of the presence of interactions

Example of an additive model

$$y = f(x_1, x_2) = 4x_1^2 + 3x_2 \qquad x_1, x_2 \in U[-1/2;1/2]$$

Remember that:

$$f_0 = E(y) = \frac{1}{3}$$

$$S_1 = \frac{Var[f_1(x_1)]}{V} = \frac{0.0\overline{8}}{0.83\overline{8}} = 0.106$$

$$f_1(x_1) = E(y \mid x_1) - f_0 = 4x_1^2 - \frac{1}{3}$$

$$S_2 = \frac{Var[f_2(x_2)]}{V} = \frac{0.75}{0.83\overline{8}} = 0.894$$

$$f_2(x_2) = E(y \mid x_2) - f_0 = 3x_2$$

$$f_{12}(x_1, x_2) = 0$$

Joint Research Centre

Total effects

$$S_1 + \cancel{S_2} + \cancel{S_3} + \cancel{S_4} + S_{12} + S_{13} + S_{14} + \cancel{S_{23}} + \cancel{S_{24}} + \cancel{S_{34}} +$$
$$S_{123} + S_{124} + S_{134} + \cancel{S_{234}} + S_{1234} = 1$$

$$S_{T1} = S_1 + S_{12} + S_{13} + S_{14} + S_{123} + S_{124} + S_{134} + S_{1234}$$

$$S_{T2} = S_2 + S_{12} + S_{23} + S_{24} + S_{123} + S_{124} + S_{234} + S_{1234}$$

$$S_{Ti} \geq S_i$$

Total effects

$$S_{T1} = S_1 + S_{12} + S_{13} + S_{14} + S_{123} + S_{124} + S_{134} + S_{1234}$$

$$S_{T2} = S_2 + S_{12} + S_{23} + S_{24} + S_{123} + S_{124} + S_{234} + S_{1234}$$

$$S_{T3} = S_3 + S_{13} + S_{23} + S_{34} + S_{123} + S_{134} + S_{234} + S_{1234}$$

$$S_{T4} = S_4 + S_{14} + S_{24} + S_{34} + S_{124} + S_{134} + S_{234} + S_{1234}$$

$$\sum_i S_{Ti} = S_1 + S_2 + S_3 + S_4 + 2\left(S_{12} + S_{13} + S_{14} + S_{23} + S_{24} + S_{34}\right) + 3\left(S_{123} + S_{124} + S_{134} + S_{234}\right) + 4S_{1234}$$

Another way to look at sensitivities

$$Var(y) = \underbrace{Var[E(y \mid \bullet)]}_{\text{Explained variance}} + \underbrace{E[Var(y \mid \bullet)]}_{\text{Residual variance}}$$

$$Var(y) = \underbrace{Var[E(y \mid x_i)]}_{} + E[Var(y \mid x_i)]$$

↓

$$V_i$$

Joint effects

$$Var(y) = Var[E(y \mid \bullet)] + E[Var(y \mid \bullet)]$$

$$Var(y) = \underbrace{Var[E(y \mid x_i, x_j)]}_{V_i + V_j + V_{ij}} + E[Var(y \mid x_i, x_j)]$$

$$V_i + V_j + V_{ij}$$

This equality does hold only for independent inputs

Total effects

$$Var(y) = Var[E(y \mid \bullet)] + E[Var(y \mid \bullet)]$$

$$Var(y) = \underbrace{Var[E(y \mid x_{\sim i})]}_{V_{\sim i}} + \underbrace{E[Var(y \mid x_{\sim i})]}_{V_{Ti}}$$

How can we estimate first order measures?

Numerical evaluation of sensitivities

All previous formulas are analytical.
In practice we can evaluate the model at a number of sample points obtaining a set of model outputs.

We can generate a sample of size N from r.v. $\mathbf{x} = (x_1, x_2, \ldots, x_k)$ using Monte Carlo techniques:

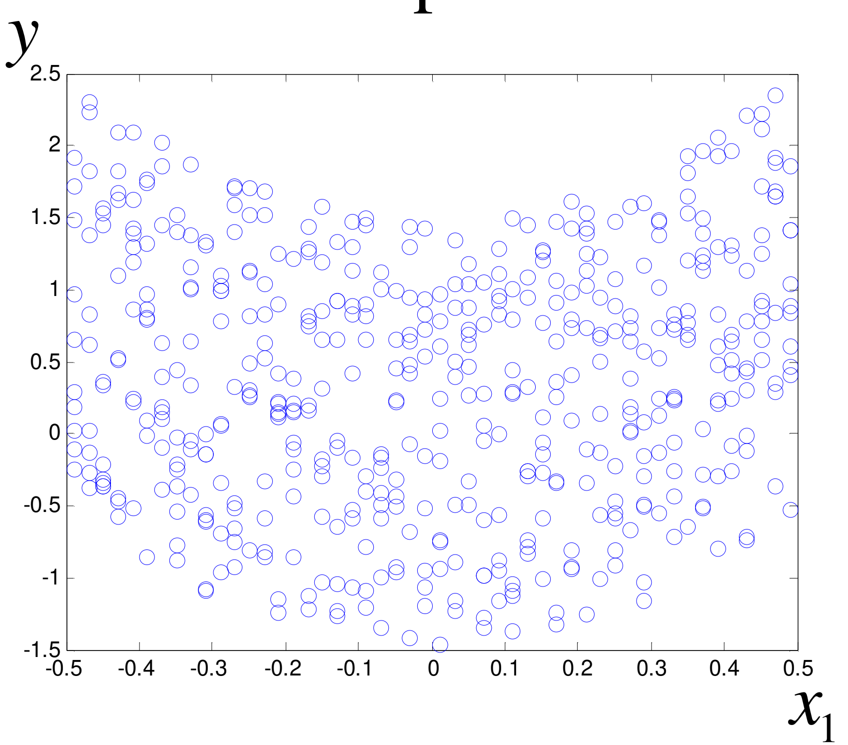
$$\begin{vmatrix} x_1^1, x_2^1, \ldots, x_k^1 \\ x_1^2, x_2^2, \ldots, x_k^2 \\ \ldots \\ x_1^N, x_2^N, \ldots, x_k^N \end{vmatrix} \xrightarrow{y = f(\mathbf{x})} \begin{vmatrix} y^1 \\ y^2 \\ \ldots \\ y^N \end{vmatrix}$$

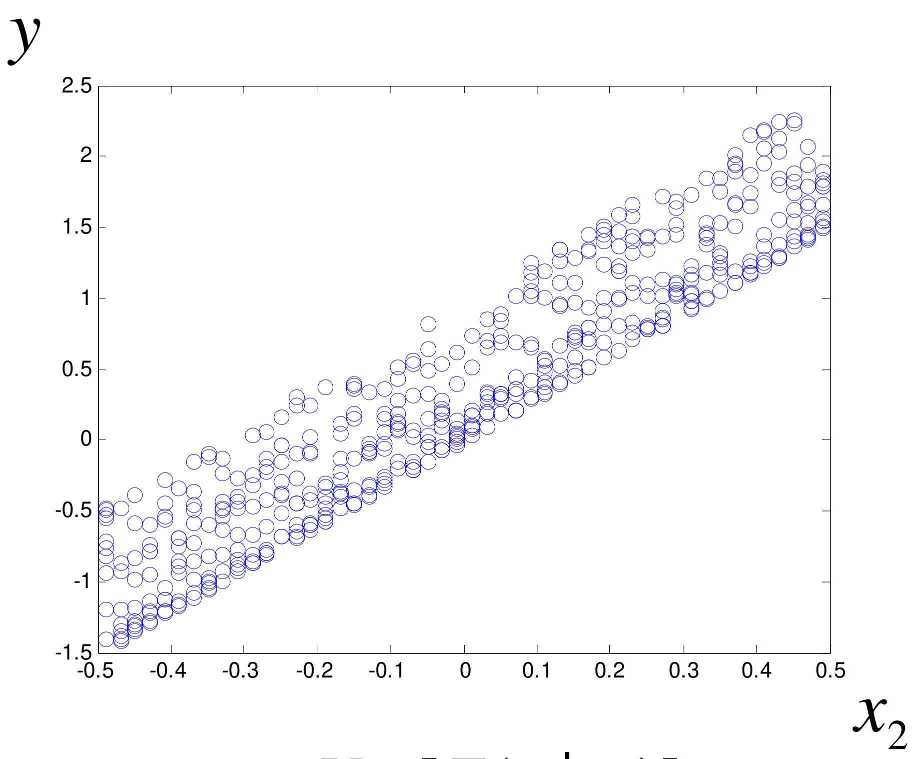
We can now represent y vs. x_i.
An example →

Scatter-plots of y vs. x_i

$$y = 4x_1^2 + 3x_2 \qquad x_1, x_2 \in U[-1/2;1/2]$$

N=500 points



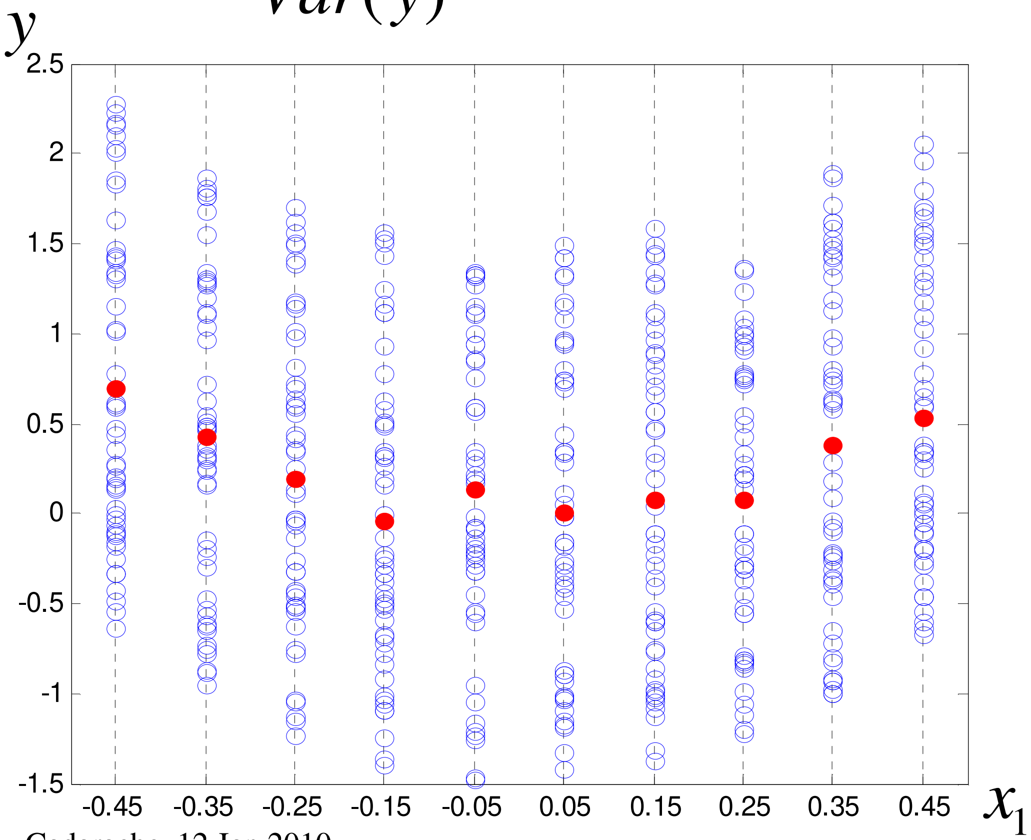


How can we compute $S_i = \dfrac{Var[E(y\,|\,x_i)]}{Var(y)}$

Calculating sensitivities: brute-force method

$$y = 4x_1^2 + 3x_2 \qquad x_1, x_2 \in U[-1/2; 1/2]$$

$$S_i = \frac{Var[E(y|x_i)]}{Var(y)} \qquad f_1(x_1) = E(y|x_1) - f_0 = 4x_1^2 - \frac{1}{3}$$



Brute-force method
r = 10 n° of cond values
N = 50
Cost = N*r = 500 runs

$$\hat{S}_1 = \frac{0.0610}{0.7692} = 0.0793$$

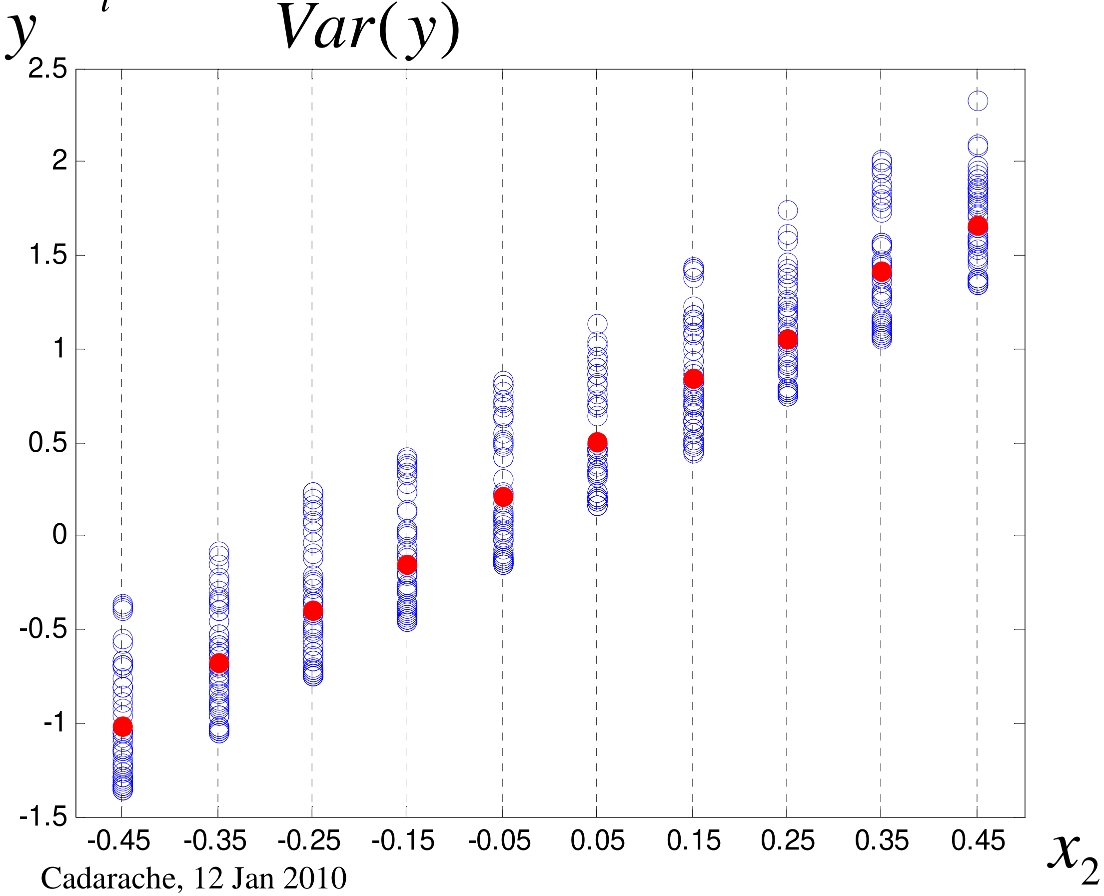
$$S_1 = 0.106$$

Calculating sensitivities: brute-force method

$$y = 4x_1^2 + 3x_2 \qquad x_1, x_2 \in U[-1/2; 1/2]$$

$$S_i = \frac{Var[E(y|x_i)]}{Var(y)}$$

$$f_2(x_2) = E(y|x_2) - f_0 = 3x_2$$



Brute-force method
r = 10 n° of cond values
N = 50
Cost = N*r = 500 runs

$$\hat{S}_2 = \frac{0.8182}{0.8240} = 0.9929$$

$$S_2 = 0.894$$

Calculating sensitivities: brute-force method

$$S_1 = 0.106$$

r = 10; N = 50

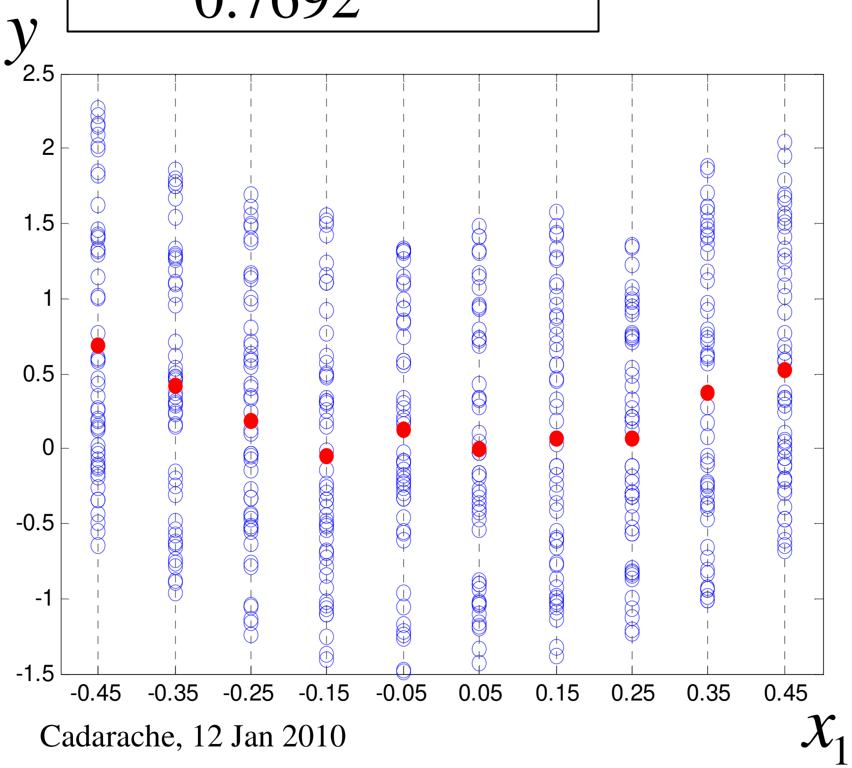
Cost = N*r = 500 runs

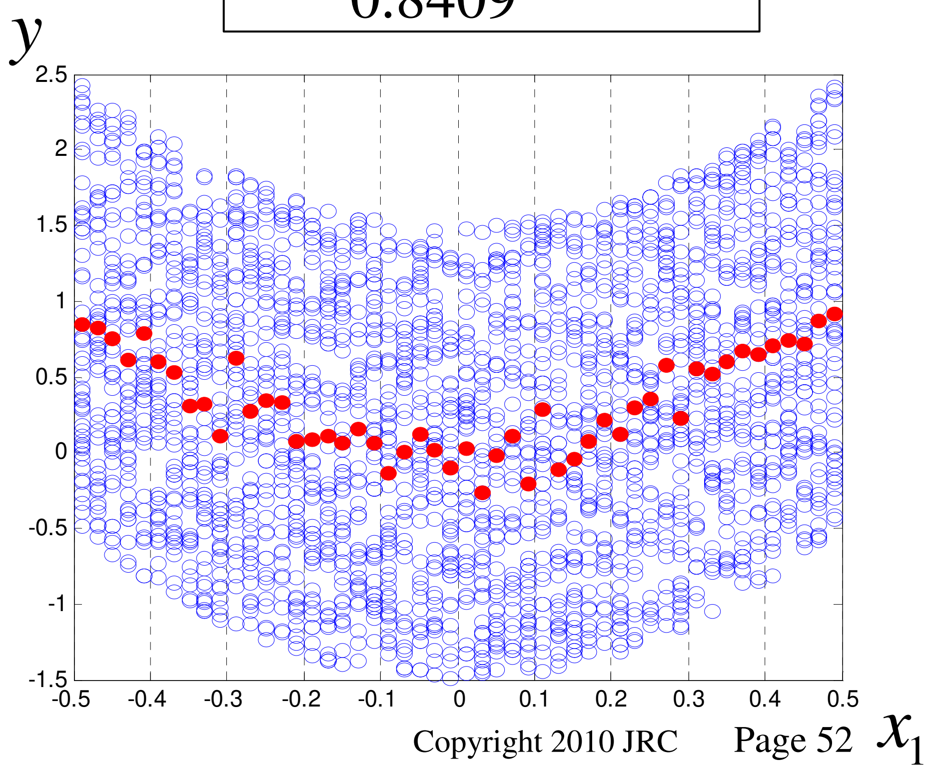
$$\hat{S}_1 = \frac{0.0610}{0.7692} = 0.0793$$

r = 50; N = 50

Cost = N*r = 2,500 runs

$$\hat{S}_1 = \frac{0.1048}{0.8409} = 0.1247$$





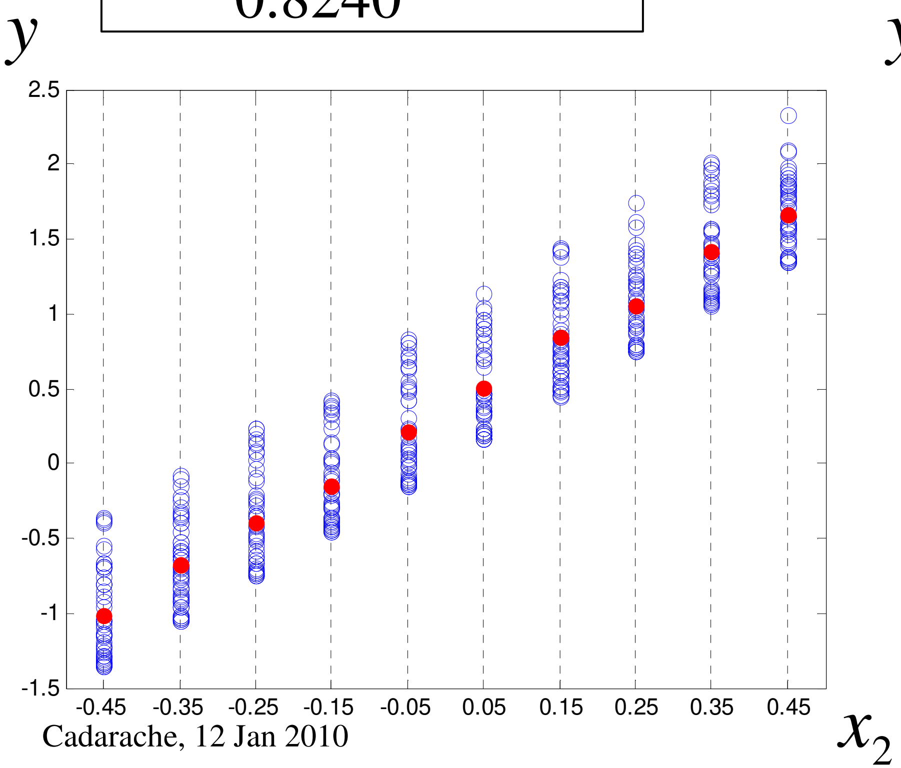
Calculating sensitivities: brute-force method

$$S_2 = 0.894$$

r = 10; N = 50

Cost = N*r = 500 runs

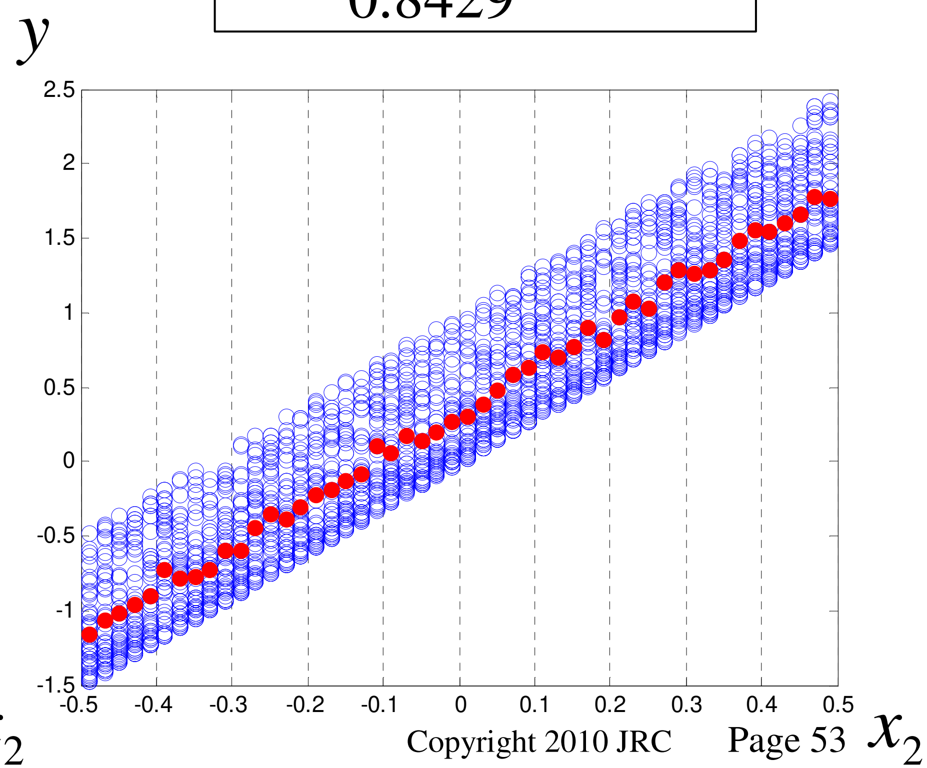
$$\hat{S}_2 = \frac{0.8182}{0.8240} = 0.9929$$



r = 50; N = 50

Cost = N*r = 2,500 runs

$$\hat{S}_2 = \frac{0.7694}{0.8429} = 0.9128$$



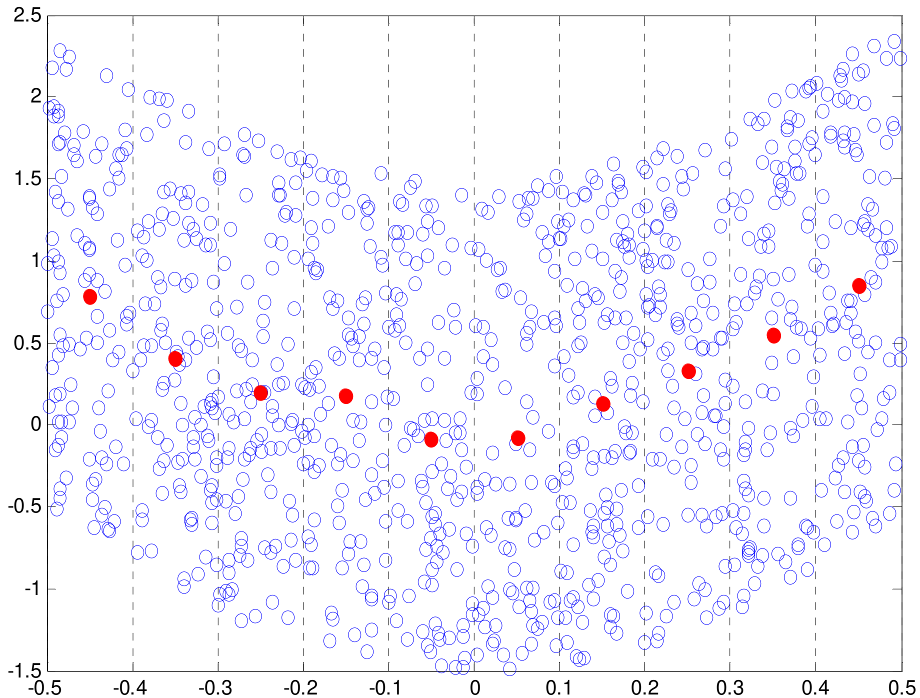
Joint Research Centre

Calculating sensitivities: approximate method

Approximate method:

1) choose total n° of points (cost=1,000);

2) calculate $E(Y|X_1)$ within stripes (r = n°. of stripes)



Cost = 1,000; r = 10 x_1

$\hat{S}_1 = 0.069$

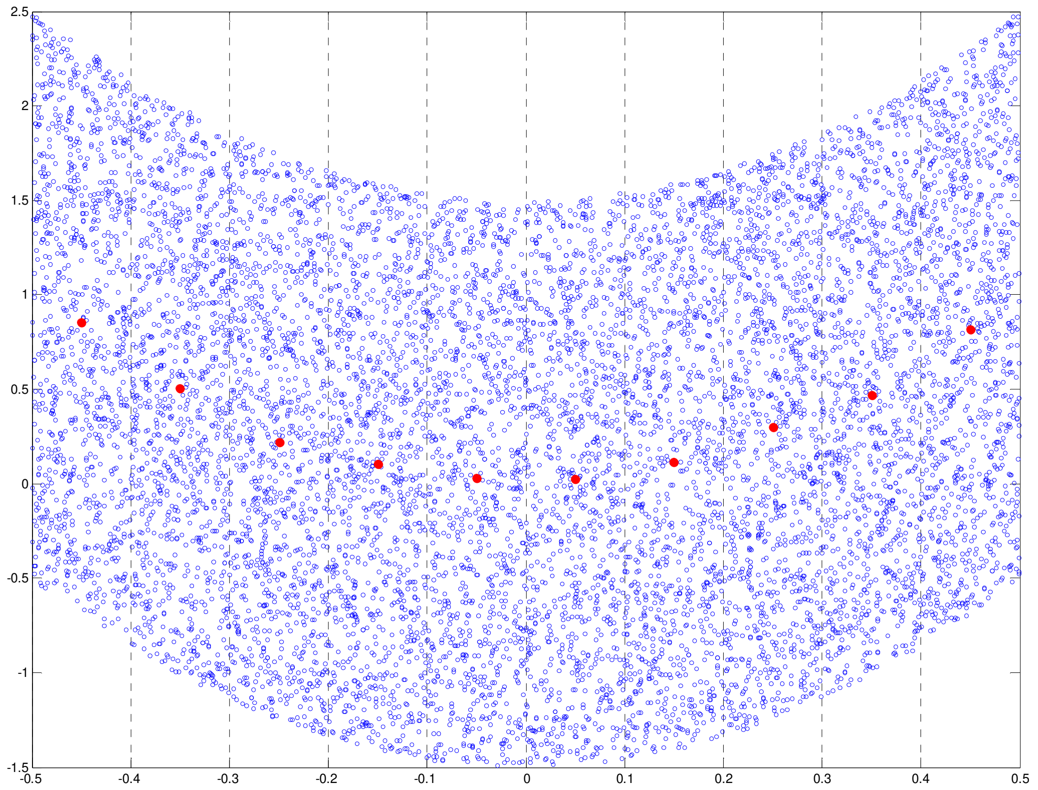
$S_1 = 0.106$

$\hat{S}_2 = 0.797$

$S_2 = 0.894$

Calculating sensitivities: approximate method

Increasing Cost
r = costant



Cost = 10,000; r = 10 x_1

$$\hat{S}_1 = 0.099$$

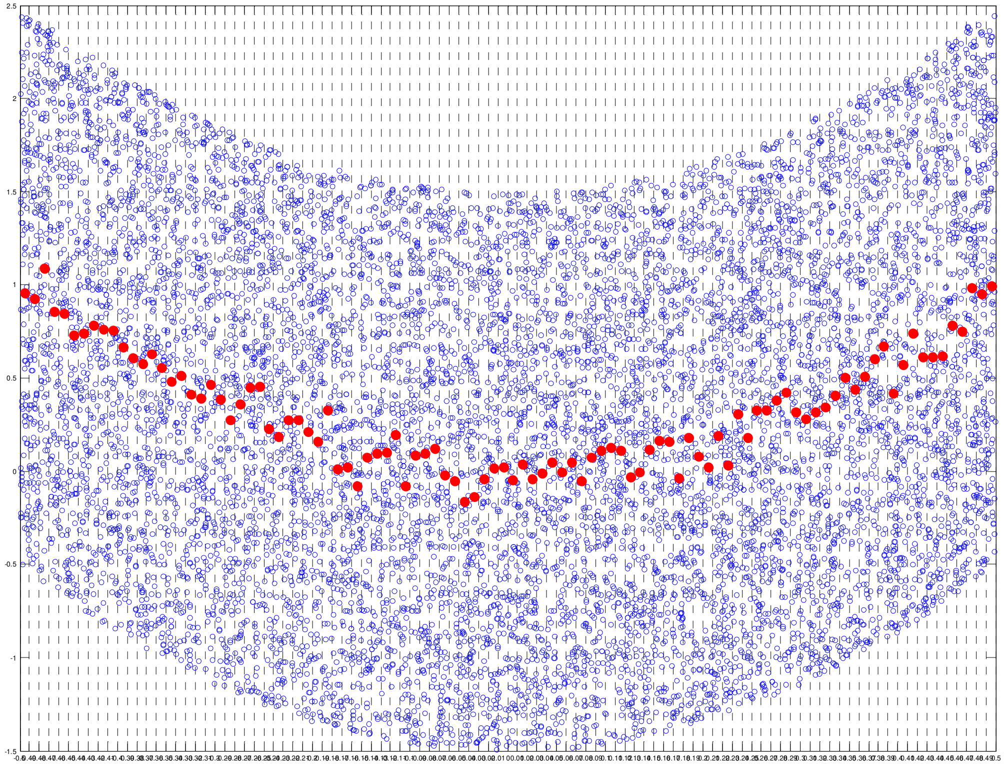
$$S_1 = 0.106$$

$$\hat{S}_2 = 0.831$$

$$S_2 = 0.894$$

Calculating sensitivities: approximate method

Increase r



Cost = 10,000; r = 100

$$\hat{S}_1 = 0.099 \quad \approx$$

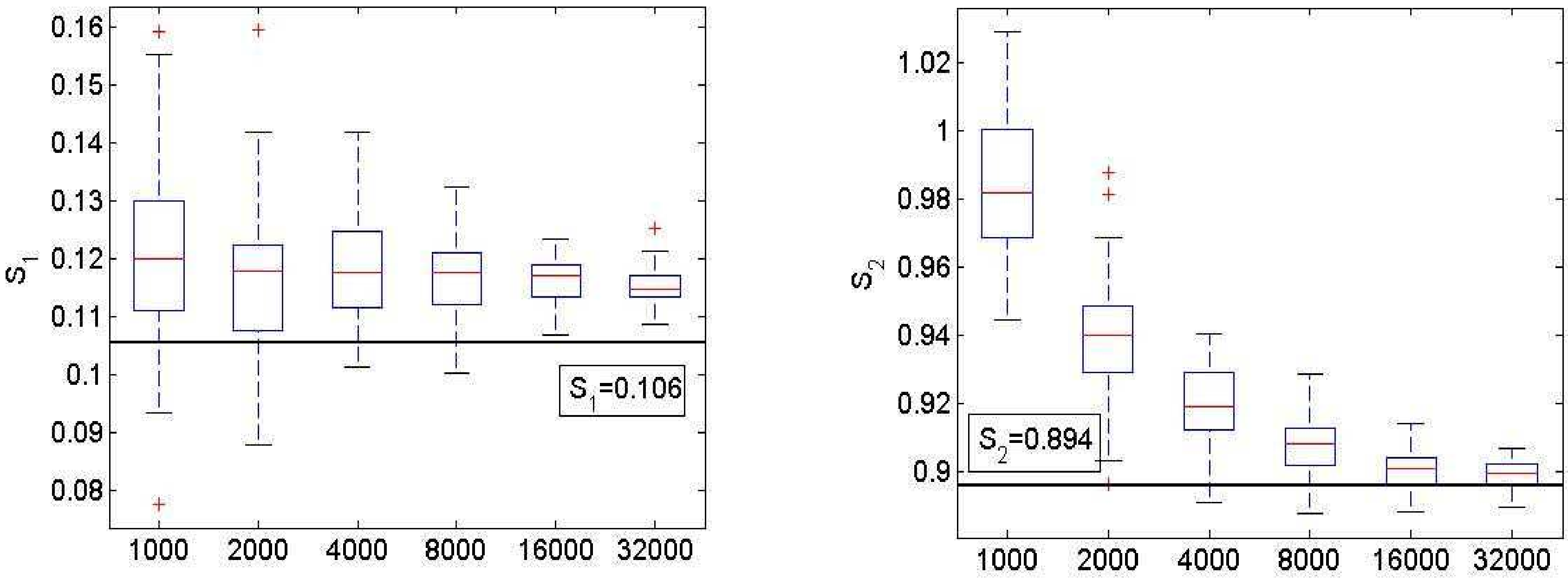
$$S_1 = 0.106$$

$$\hat{S}_2 = 0.760$$

$$S_2 = 0.894$$

Joint Research Centre

Convergence curve at increasing n. of model runs (50 replicates)



Chronology of variance-based methods

FAST (Fourier Amplitude Sensitivity Test), (Cukier et al., 1973)

The method of Sobol', (Sobol', 1993)

The method of Jansen, (Jansen, 1994)

EFAST (Extended FAST), (Saltelli, Tarantola and Chan, 1999)

An improvement of the Sobol' method, (Saltelli, 2002)

Random Balance Designs (Tarantola, Gatelli, Mara, 2006)

Improvement of Sobol' for small indices (Sobol', Tarantola, et al.,2007)

Best practices for first and total indices (Saltelli et al., 2009)

First order indices

$$V_1 = Var[E(y|x_1)] = \int \left(E(y|x_1) - f_0\right)^2 dx_1 =$$
$$\int E^2(y|x_1)dx_1 - 2f_0 \int E(y|x_1)dx_1 + f_0^2 =$$
$$\int E^2(y|x_1)dx_1 - f_0^2$$

Let (x_1', x_2',…, x_k') be an independent realization of (x_1, x_2,…, x_k).

$$E^2(y|x_1) = E(y|x_1)E(y|x_1) = \int f(x_1, x_2, x_3)dx_2 dx_3 \int f(x_1, x_2', x_3')dx_2' dx_3'$$

The integral over 3 (k) dimensions is transformed into an integral over 5 (2k-1) dimensions [Ishigami and Homma,1990],

$$\int E^2(y|x_1)dx_1 = \int f(x_1, x_2, x_3)\underbrace{dx_1 dx_2 dx_3}_{k} \int f(x_1, x_2', x_3')\underbrace{dx_2' dx_3'}_{k-1}$$

First order indices

$$Var[E(y|x_1)] = \int E^2(y|x_1)dx_1 - f_0^2 =$$

$$\int f(x_1,x_2,x_3)f(x_1,x_2',x_3')dx_1dx_2dx_3dx_2'dx_3' - \int f(x_1,x_2,x_3)dx_1dx_2dx_3 f(x_1',x_2',x_3')dx_1'dx_2'dx_3' =$$

$$\int f(x_1,x_2,x_3)\left[f(x_1,x_2',x_3') - f(x_1',x_2',x_3')\right]dx_1dx_2dx_3dx_1'dx_2'dx_3'$$

Monte Carlo algorithm:

$$Var[E(Y|\underline{x_1})] = \frac{1}{N-1}\sum_{r=1}^{N} f(x_{r1},x_{r2},x_{r3})[f(\underline{x_{r1}},x'_{r2},x'_{r3}) - f(x'_{r1},x'_{r2},x'_{r3})]$$

$$Var[E(Y|\underline{x_2})] = \frac{1}{N-1}\sum_{r=1}^{N} f(x_{r1},x_{r2},x_{r3})[f(x'_{r1},\underline{x_{r2}},x'_{r3}) - f(x'_{r1},x'_{r2},x'_{r3})]$$

$$Var[E(Y|\underline{x_3})] = \frac{1}{N-1}\sum_{r=1}^{N} f(x_{r1},x_{r2},x_{r3})[f(x'_{r1},x'_{r2},\underline{x_{r3}}) - f(x'_{r1},x'_{r2},x'_{r3})]$$

First order indices

$$Var[E(Y|x_1)] = \frac{1}{N-1}\sum_{r=1}^{N} f(x_{r1},x_{r2},x_{r3})[f(x_{r1},x'_{r2},x'_{r3}) - f(x'_{r1},x'_{r2},x'_{r3})]$$

$$A = \begin{bmatrix} x_{11} & x_{12} & x_{13} \\ x_{21} & x_{22} & x_{23} \\ \vdots & & \vdots \\ x_{N1} & x_{N2} & x_{N3} \end{bmatrix} \xrightarrow{N} \begin{bmatrix} f(x_{11},x_{12},x_{13}) \\ f(x_{21},x_{22},x_{23}) \\ \vdots \\ f(x_{N1},x_{N2},x_{N3}) \end{bmatrix}$$

$$B = \begin{bmatrix} x'_{11} & x'_{12} & x'_{13} \\ x'_{21} & x'_{22} & x'_{23} \\ \vdots & & \vdots \\ x'_{N1} & x'_{N2} & x'_{N3} \end{bmatrix} \xrightarrow{N} \begin{bmatrix} f(x'_{11},x'_{12},x'_{13}) \\ f(x'_{21},x'_{22},x'_{23}) \\ \vdots \\ f(x'_{N1},x'_{N2},x'_{N3}) \end{bmatrix}$$

$$B_A^1 = \begin{bmatrix} x_{11} & x'_{12} & x'_{13} \\ x_{21} & x'_{22} & x'_{23} \\ \vdots & & \vdots \\ x_{N1} & x'_{N2} & x'_{N3} \end{bmatrix} \xrightarrow{N} \begin{bmatrix} f(x_{11},x'_{12},x'_{13}) \\ f(x_{21},x'_{22},x'_{23}) \\ \vdots \\ f(x_{N1},x'_{N2},x'_{N3}) \end{bmatrix}$$

In this ex: N*(3+2)

In general: N*(k+2)

Joint Research Centre

Total order indices (Jansen's)

$$S_{T1} = \frac{E[Var(y \mid x_2, x_3)]}{Var(y)}$$

$$E[Var(y \mid x_2, x_3)] = \frac{1}{2(N-1)} \sum_{r=1}^{N} [f(x'_{r1}, x_{r2}, x_{r3}) - f(x_{r1}, x_{r2}, x_{r3})]^2$$

$$S_{T2} = \frac{E[Var(y \mid x_1, x_3)]}{Var(y)}$$

$$E[Var(y \mid x_1, x_3)] = \frac{1}{2(N-1)} \sum_{r=1}^{N} [f(x_{r1}, x'_{r2}, x_{r3}) - f(x_{r1}, x_{r2}, x_{r3})]^2$$

$$S_{T3} = \frac{E[Var(y \mid x_1, x_2)]}{Var(y)}$$

$$E[Var(y \mid x_1, x_2)] = \frac{1}{2(N-1)} \sum_{r=1}^{N} [f(x_{r1}, x_{r2}, x'_{r3}) - f(x_{r1}, x_{r2}, x_{r3})]^2$$

Best practice

Hence, both first order and total effects…

can be estimated in a single shot at the cost of N(k+2) model runs, better than brute-force (yields only first order measures at N*r*k).

Example: N=512 k=2 N(k+2)=2,048

$$\boxed{S_1 = 0.106}$$

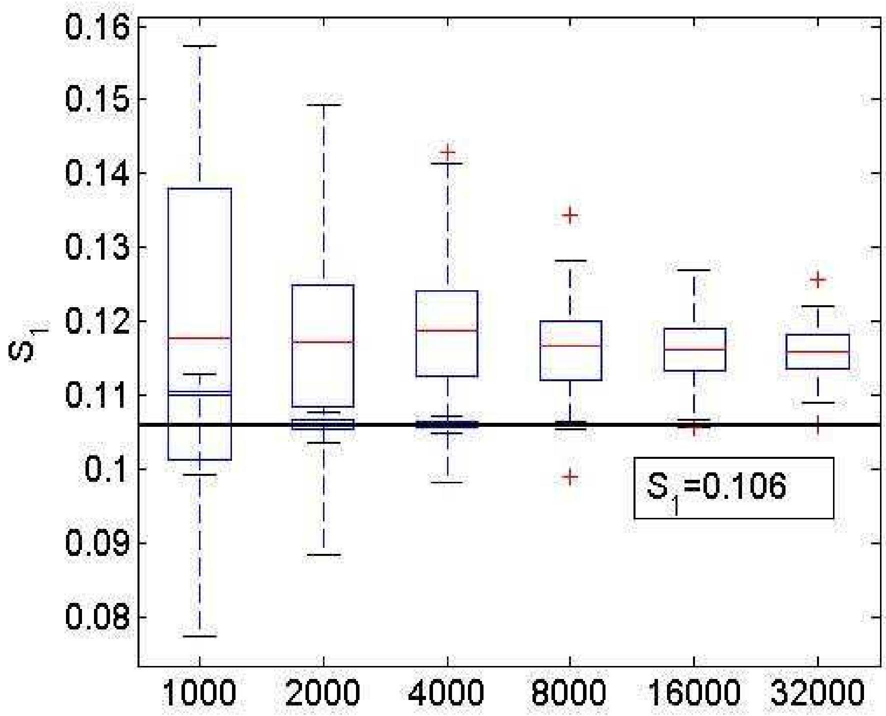
$$\hat{S}_1 = 0.106$$

$$\boxed{S_2 = 0.894}$$

$$\hat{S}_2 = 0.897$$

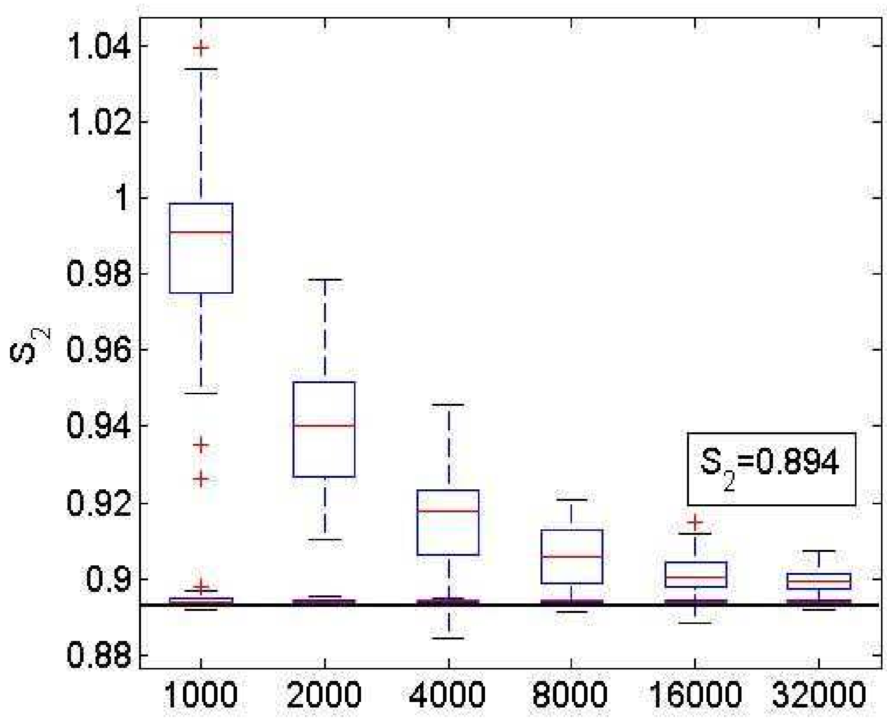
Calculating sensitivities: Sobol' vs. approximated

Convergence curve at increasing n. of model runs (50 replicates)



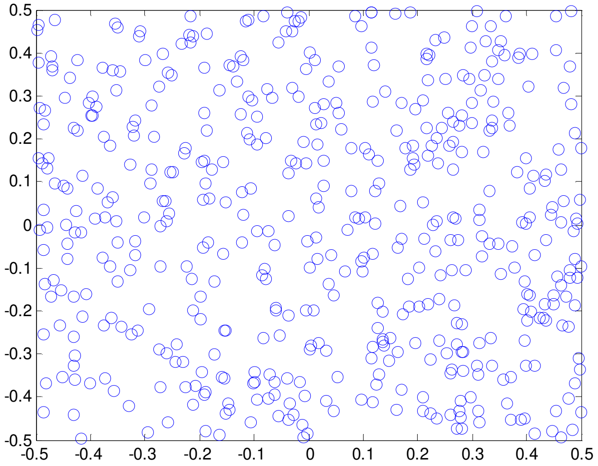
Calculating sensitivities: Sobol' vs. approximated

Convergence curve at increasing n. of model runs (50 replicates)



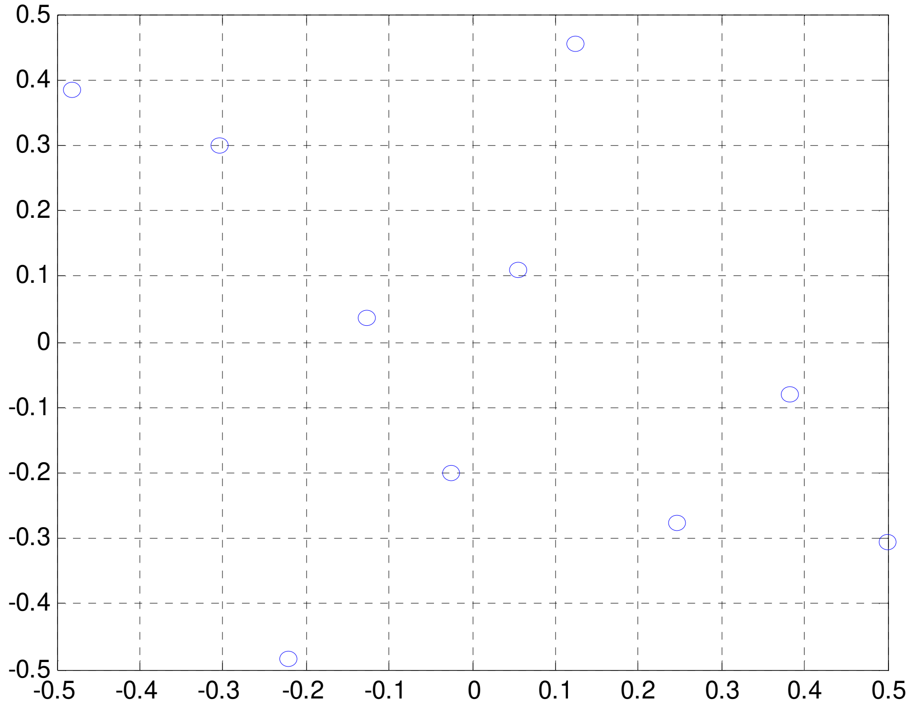
Sampling strategies

Different strategies can be used to generate the input sample



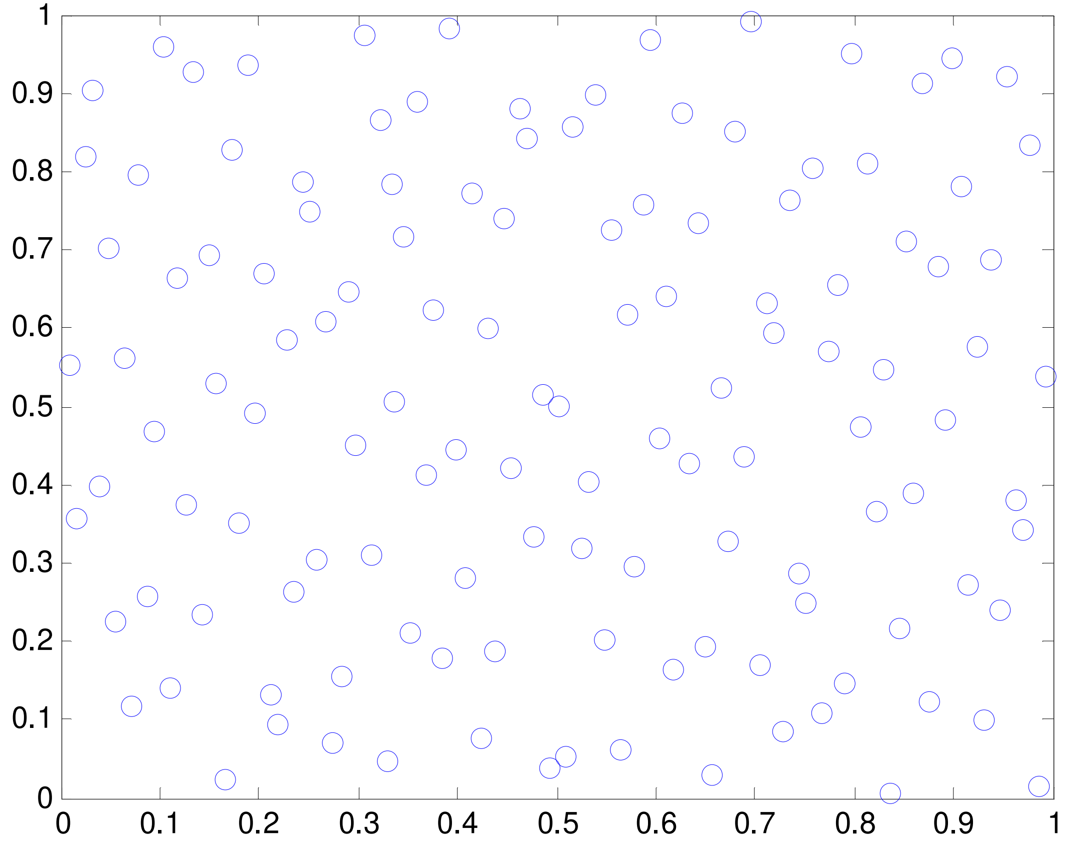
Pure random

Latin Hypercube

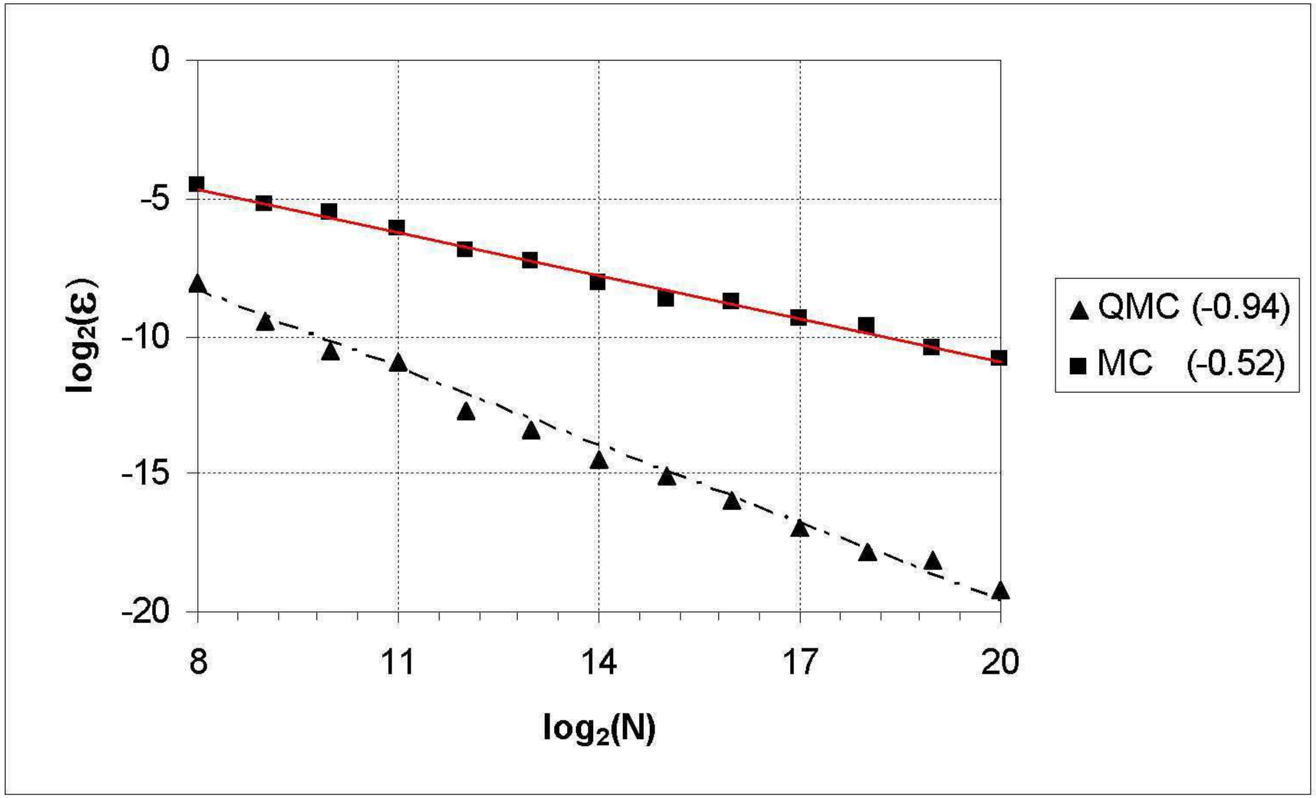


Sampling strategies

…or quasi random sequences



Convergence properties



Source: Mauntz and Kucherenko, 2005

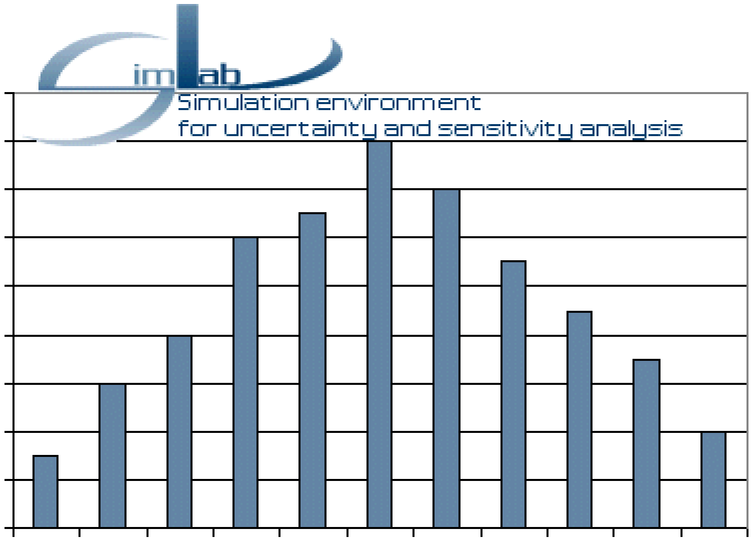
Limitations of the variance-based methods

(1) They are more expensive to compute than regression coefficients or screening techniques

(2) They are elaborate to code

(3) They assume that all the information about the uncertainty of Y is captured by its variance

Software for Sensitivity Analysis



http://simlab.jrc.it

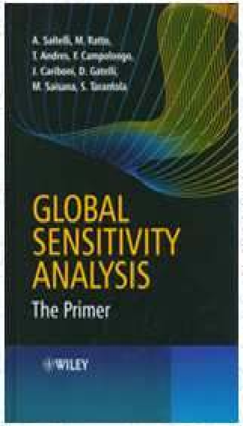
Version 2.2 → Desktop application

Version 3.2.6 → Development Environment

JRC
EUROPEAN COMMISSION

http://sensitivity-analysis.jrc.ec.europa.eu/

Sensitivity Analysis



Global Sensitivity Analysis **The Primer**
Saltelli, A., Ratto, M., Andres, T., Campolongo, F., Cariboni, J., Gatelli, D. Saisana, M., and Tarantola, S., 2008, John Wiley & Sons (ISBN: 978-0-470-05997-5)

Who needs Sensitivity Analysis

Tutorial on Sensitivity Analysis

SimLab Software for Sensitivity Analysis

What's New

- Sixth **International Conference** on Sensitivity Analysis of Model Output, Bocconi University of Milan, 19-22 July 2010
- Sixth **Summer School** on Sensitivity Analysis of Model Output, Villa La Stella, Fiesole - Florence, 21-24 September 2010

Contact Information:

Econometrics and Applied Statistics Unit (EAS)
Via Enrico Fermi 2749 , I-21027 ISPRA (VA) ITALY
Tel.:+39(0332)789686
Fax:+39(0332)785733
Email: Andrea Saltelli , Stefano Tarantola

Joint Research Centre

http://samo2010.unibocconi.it/

Bocconi

- Home
- Mission
- People
- Research Areas
- Research Projects
- Seminars
- Sponsors

Samo 2010

- Scope
- Call for papers
- Important dates
- Conference Proceedings and Special Issues
- Abstract Submission
- Registration
- Keynote Speakers
- Scientific committee
- Local Organizing Committee
- Programme
- Social dinner
- Travel advice
- Conference venue
- Accommodation

ELEUSI Centre for Research on Analysis and Systematic Use of Information

Contacts | Site Map | Search | OK

Samo 2010

SAMO

6th International Conference on Sensitivity Analysis of Model Output

19-22 July, 2010
Bocconi University, Milano, Italy
http://samo2010.unibocconi.it/

Last updated September 04, 2009

© ELEUSI - Università Bocconi | Print