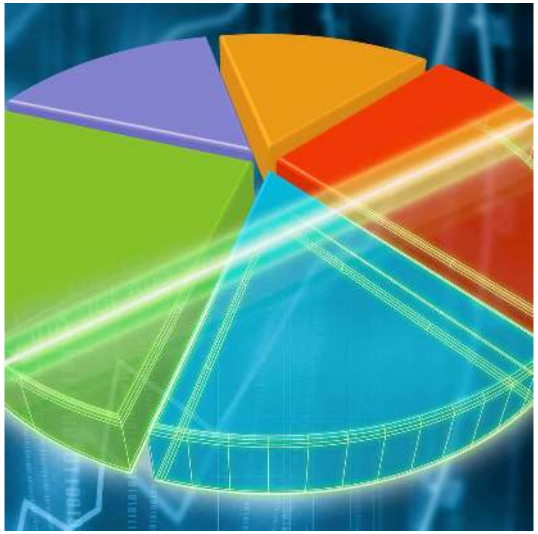
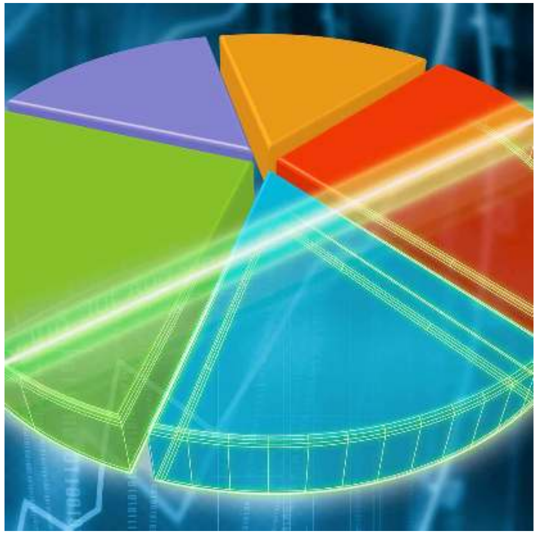


Innovative SA methods

Stefano Tarantola

Joint Research Centre
of the European Commission

Cadarache
January, 12 2010



Sobol's formula for first order indices

Classic Sobol' formula (1993) $x = (y, z)$

$$V_y = \int f(x) f(y, z') dx dz' - f_0^2$$

$$S_y = V_y / V$$

For the r-th MC trial two independent random points x_r and x'_r are used:

$$f_0 \approx \frac{1}{N} \sum_{r=1}^{N} f(x_r)$$

$$V \approx \frac{1}{N} \sum_{r=1}^{N} f^2(x_r) - f_0^2$$

$$V_y \approx \frac{1}{N} \sum_{r=1}^{N} f(x_r) f(y_r, z'_r) - f_0^2$$

When $V_y \ll f_0^2$ can loose accuracy

Sobol', Tara, Gatelli, Kuche (2007) $x = (y, z)$

$$f_0^2 = \int f(x) f(x') dx dx'$$ ← Saltelli (2002)

$$V_y = \int f(x) f(y, z') dx dz' - f_0^2$$

$$V_y = \int f(x) \left[f(y, z') - f(x') \right] dx dx'$$

The MC algorithm:

$$V_y \approx \frac{1}{N} \sum_{r=1}^{N} f(x_r) \left[f(y_r, z'_r) - f(x'_r) \right]$$

More stable than the original algorithm

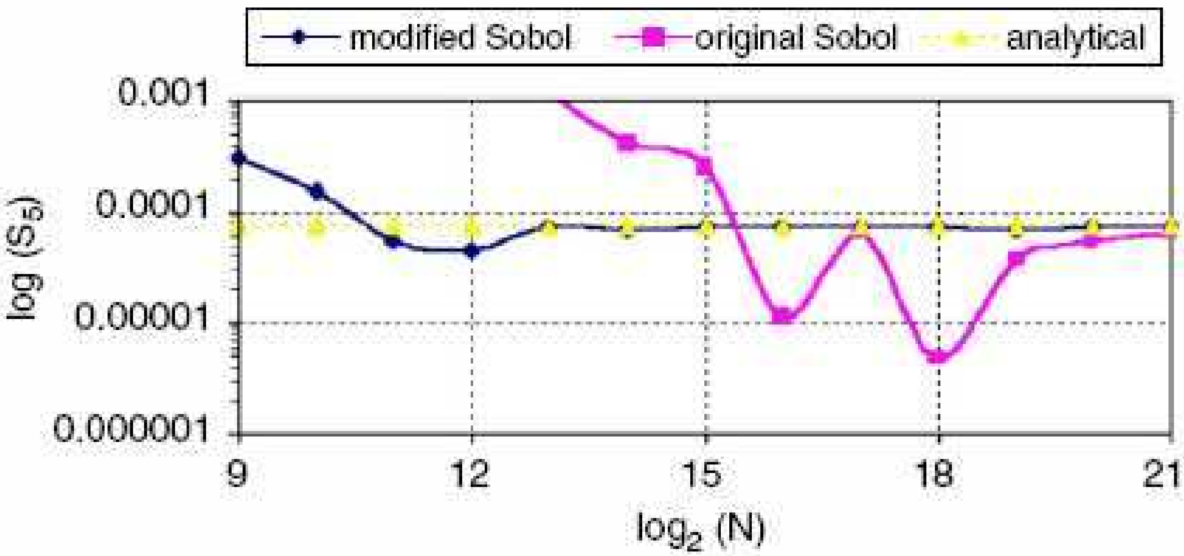
Classic Sobol' (1993)

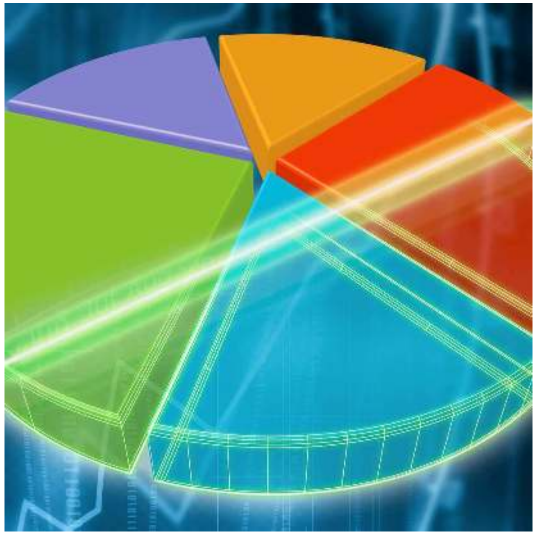
$$V_y \approx \frac{1}{N}\sum_{r=1}^{N} f(x_r) f(y_r, z'_r) - f_0^2$$

$$f_0 \approx \frac{1}{N}\sum_{r=1}^{N} f(x_r)$$

Sobol', Tara, Gatelli, Kuche (2007)

$$V_y \approx \frac{1}{N}\sum_{r=1}^{N} f(x_r)\left[f(y_r, z'_r) - f(x'_r)\right]$$





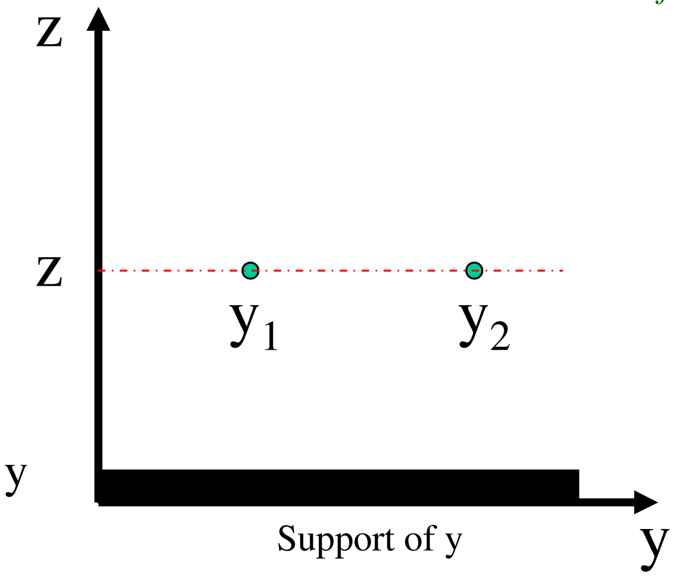
Jansen's formula for total indices

Jansen's formula

$$f(x) = f(y, z) \qquad S_{Ty} = \frac{E_z\left[Var_y(f \mid z)\right]}{Var_x(f)}$$

$$d \equiv f(y_1, z) - f(y_2, z)$$

y_1, y_2 are two independent realizations of y



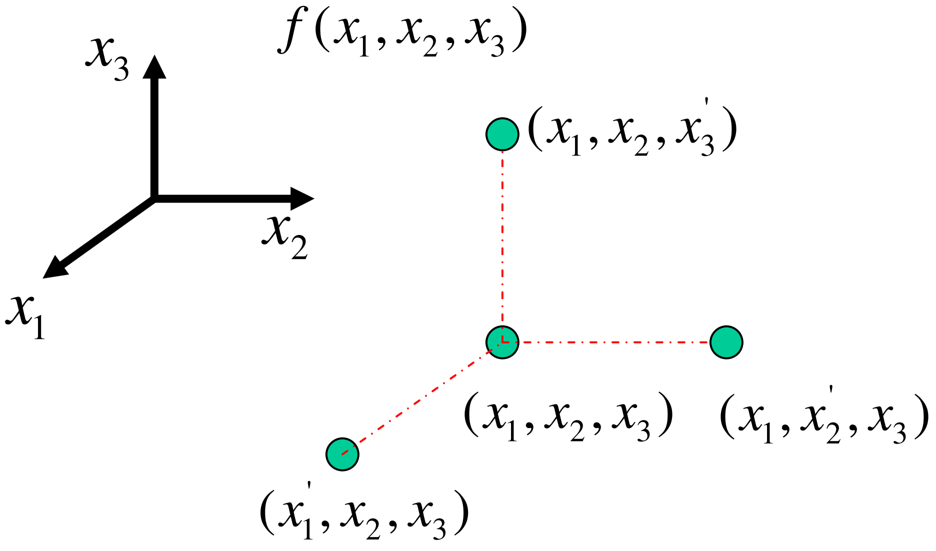
$$E_y(f(y_1, z)) = E_y(f(y_2, z)) \Rightarrow E_y(d) = 0$$

$$Var(d) = E(d^2) = 2Var_y(f \mid z)$$

$$\tfrac{1}{2} E_z E_y(d^2) = E_z[Var(f \mid z)]$$

$$E_z[Var(f \mid z)] = \tfrac{1}{2} E(d^2)$$

$$E_z[Var(f \mid z)] \approx \frac{1}{2N} \sum_{r=1}^{N} \left[f(y_r, z_r) - f(y_r^{'}, z_r)\right]^2$$



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

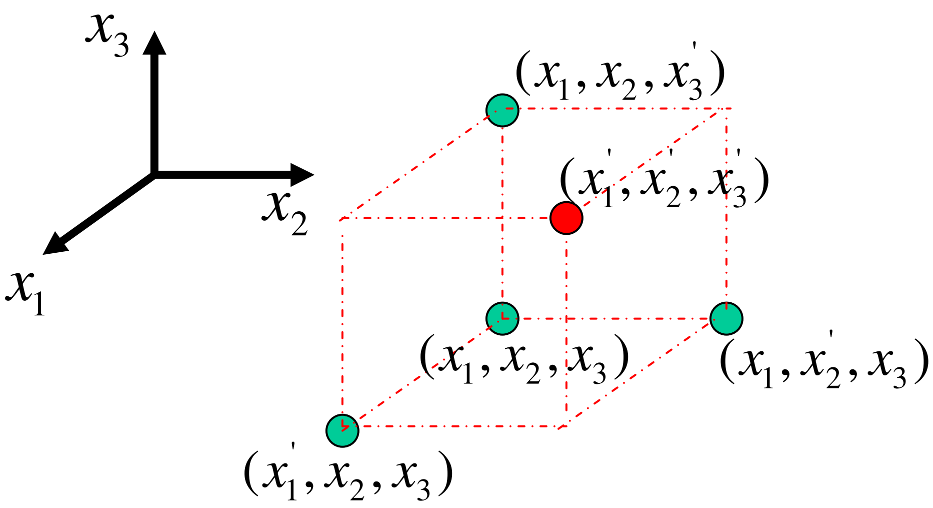
$$S_{T2} = \frac{E[Var(f \mid x_1, x_3)]}{Var(y)}$$

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

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

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

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



$$S_1 = \frac{Var[E(f \mid x_1)]}{Var(f)}$$

$$S_2 = \frac{Var[E(f \mid x_2)]}{Var(f)}$$

$$S_3 = \frac{Var[E(f \mid x_3)]}{Var(f)}$$

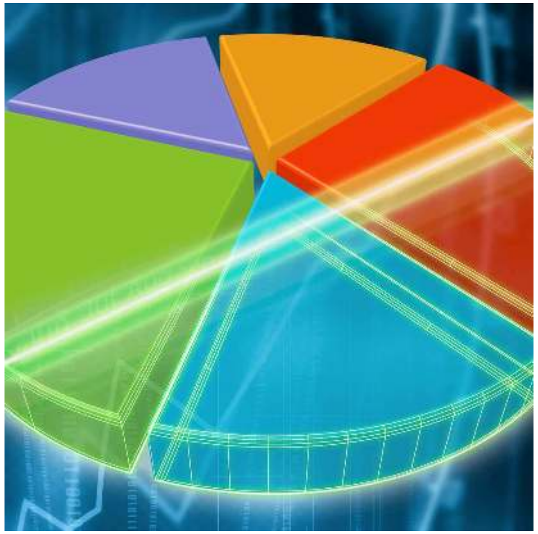
$$Var[E(f \mid x_1)] = \frac{1}{N-1}\sum_{r=1}^{N} f(x_{r1}, x_{r2}, x_{r3})[f(x_{r1}, x'_{r2}, x'_{r3}) - f(x'_{r1}, x'_{r2}, x'_{r3})]$$

? ?

Swap x with x’

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

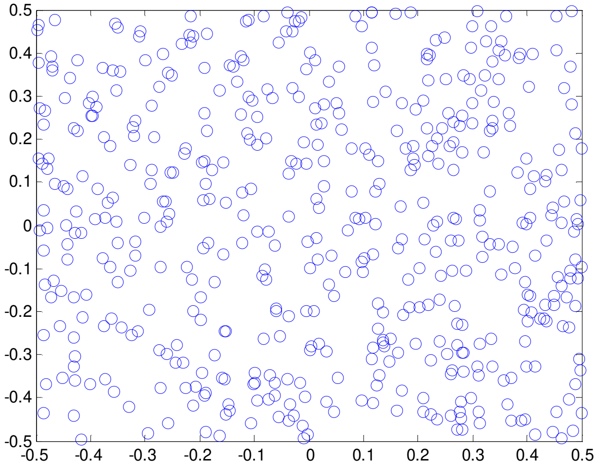
?



Sampling strategies

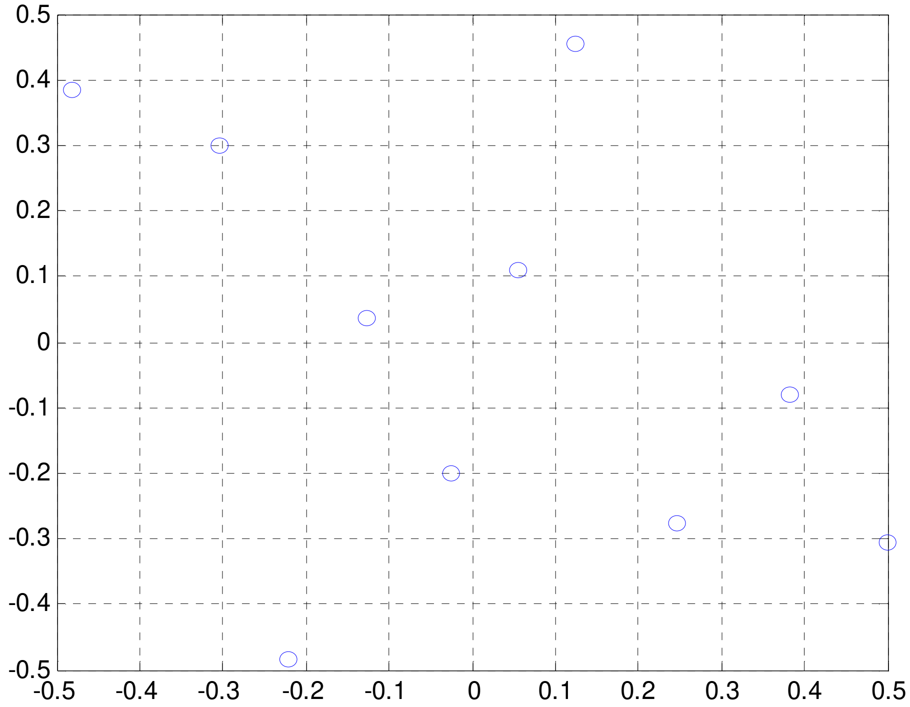
Sampling strategies

Different strategies can be used to generate the input sample



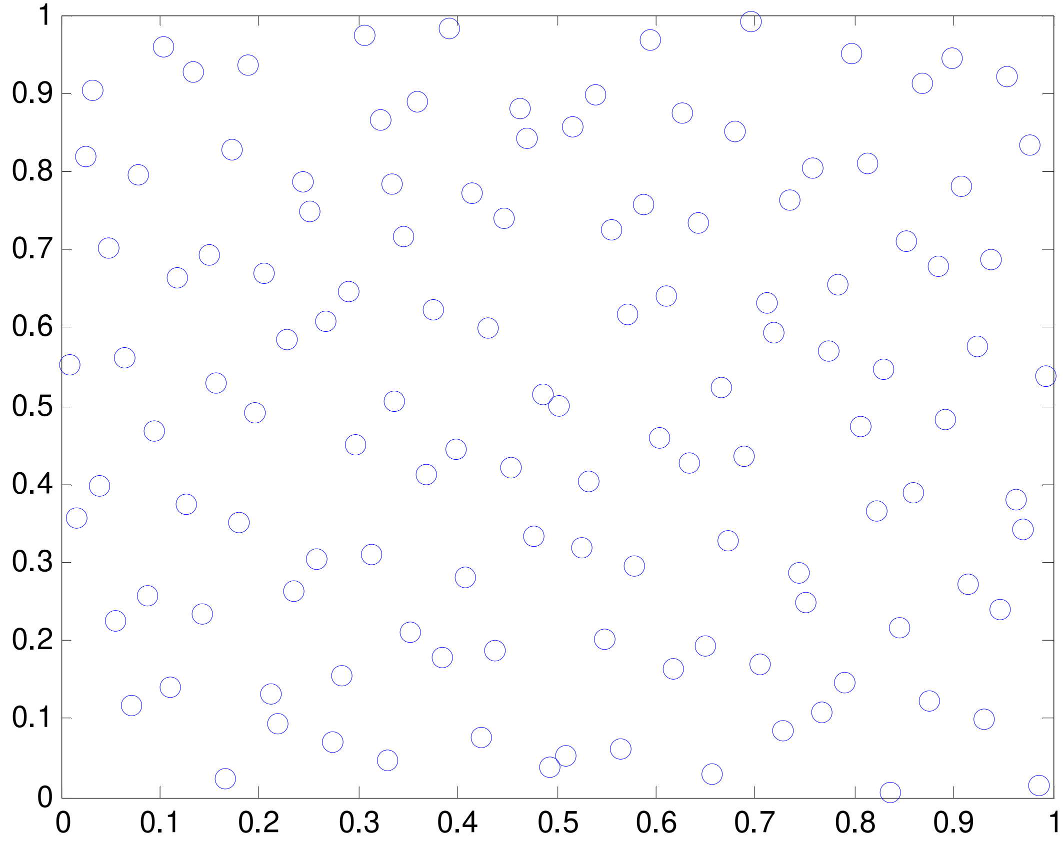
Pure random

Latin Hypercube

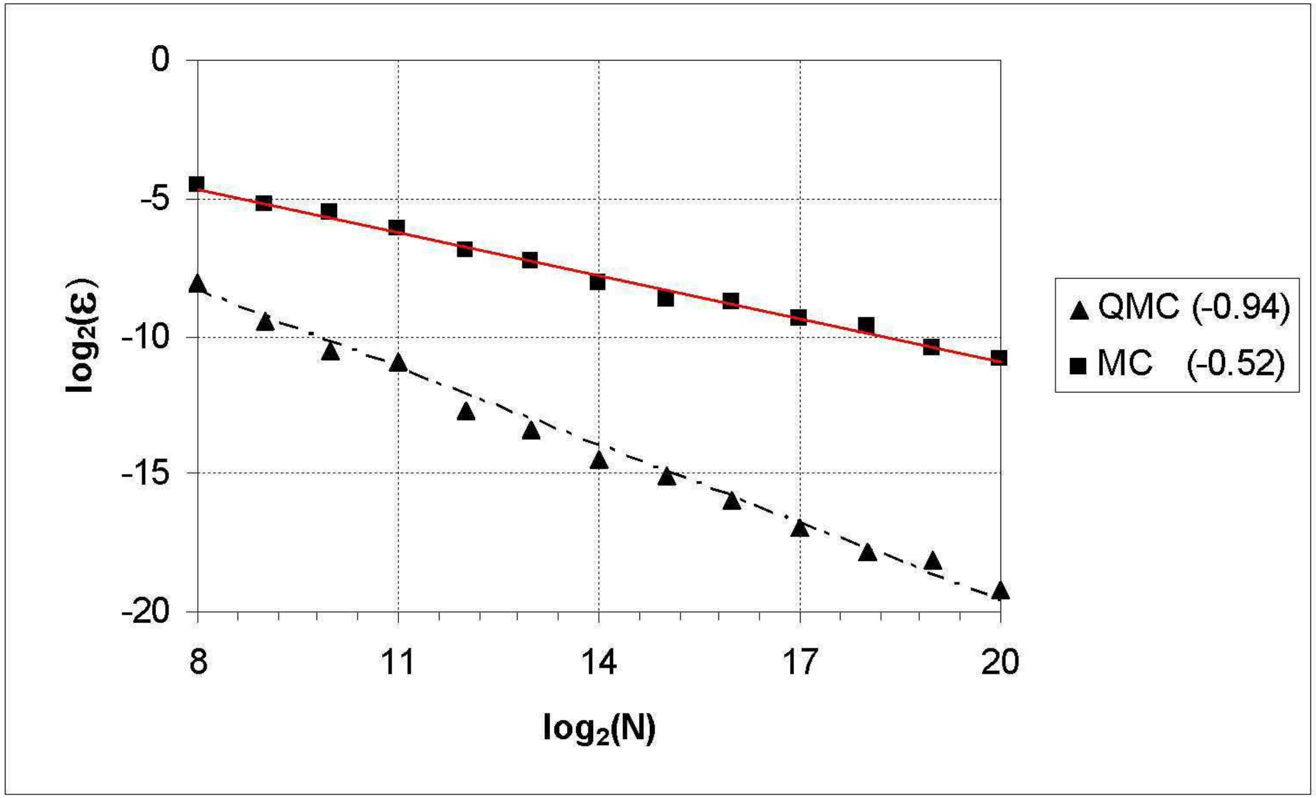


Sampling strategies

…or quasi random sequences

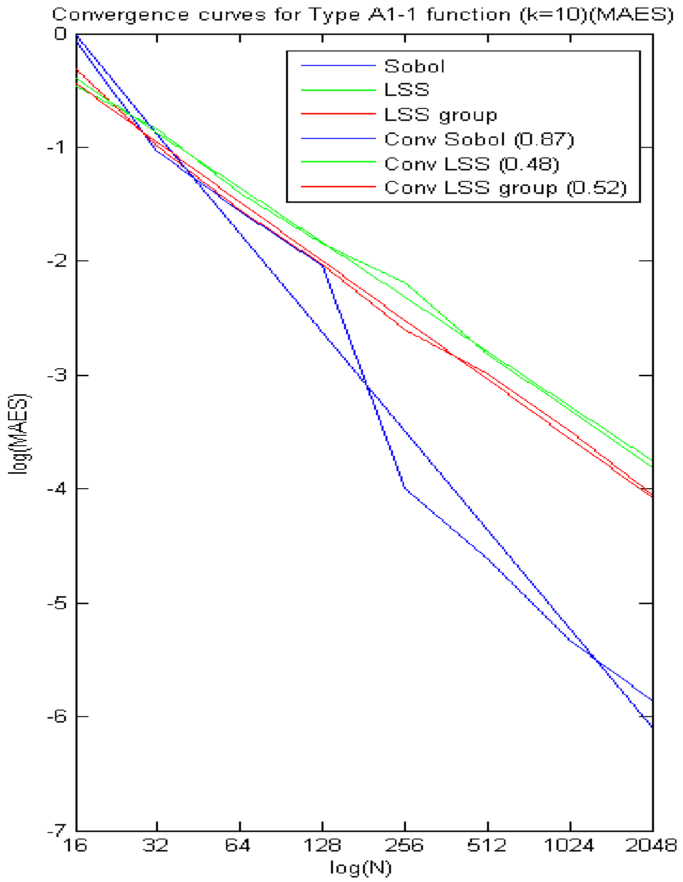


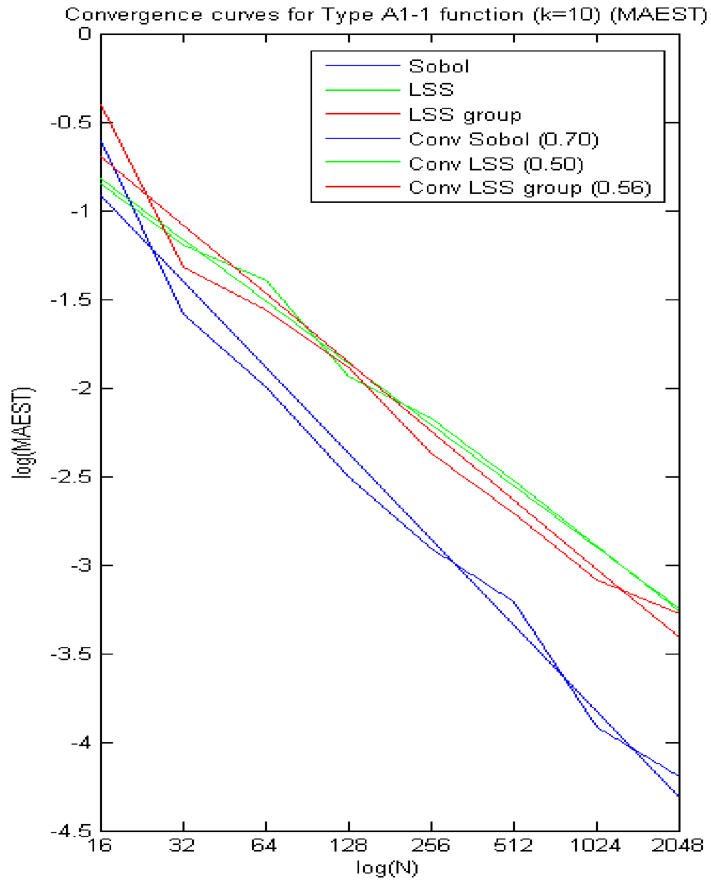
Convergence properties



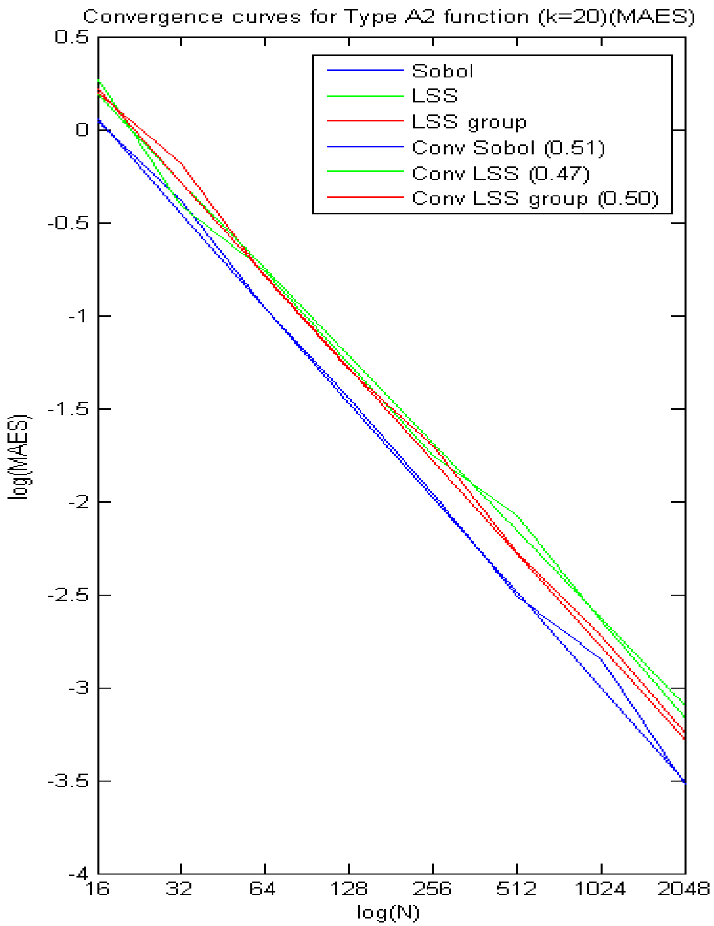
Source: Mauntz and Kucherenko, 2005

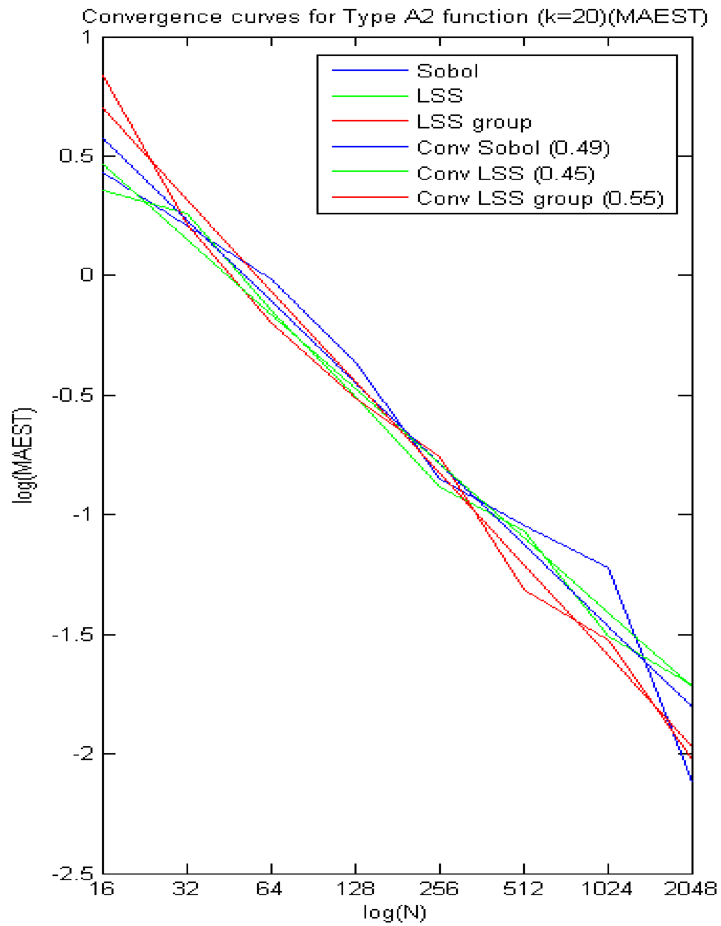
Convergence properties

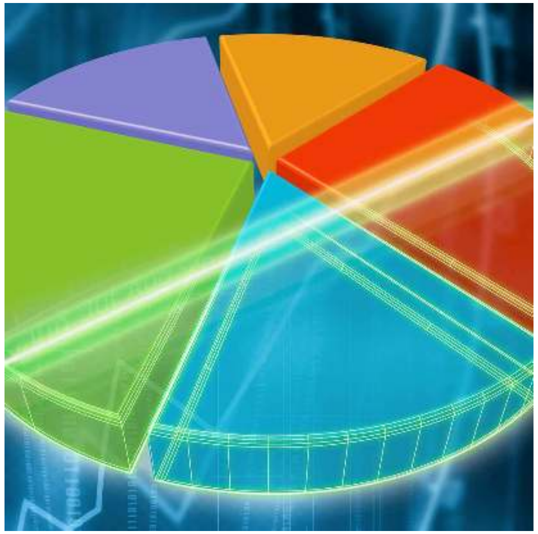




Convergence properties







Screening methods

Joint Research Centre

Screening designs “highlight” few important inputs among many (hundreds, thousands)

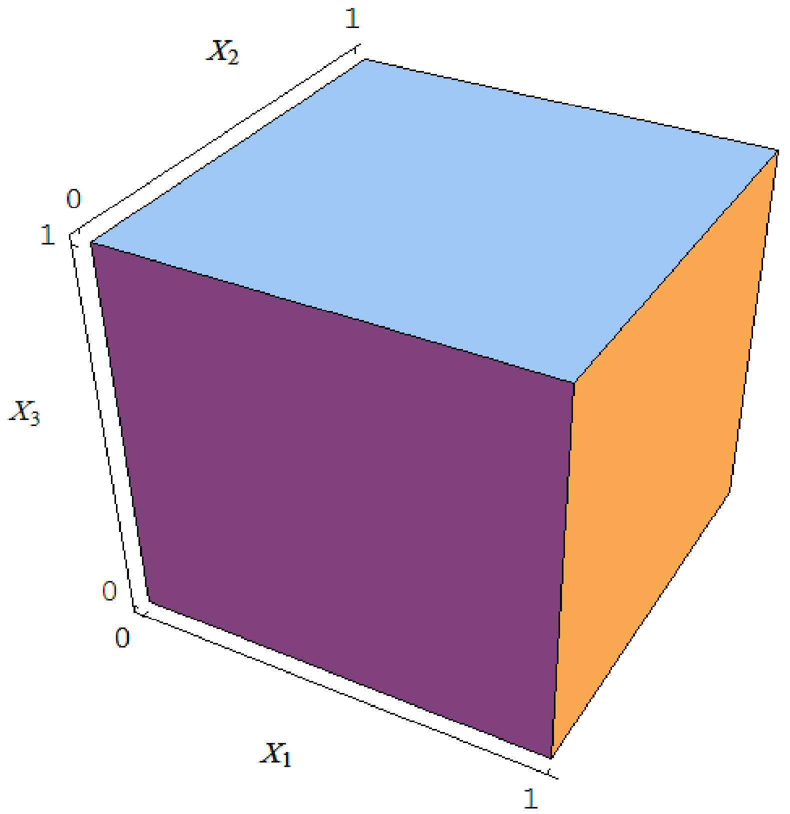
Goal: Model simplification

Screening designs derive from
Design of Experiments (DOE)

The Full Factorial Design

A 2-Level Full Factorial Design for 3 parameters

X_1	X_2	X_3
1	1	1
-1	1	1
1	-1	1
-1	-1	1
1	1	-1
-1	1	-1
1	-1	-1
-1	-1	-1



The full design requires an enormous number of points to get all combinations of levels.

Using 2 levels: 10 parameters $\rightarrow$ $2^{10} = 1024$ points

20 parameters $\rightarrow$ more than a million!!

Solution

Select only a sub-set of these points to generate a smaller design that can still produce useful results

A Fractional Factorial design can be generated using Hadamard matrices.

Def. A Hadamard matrix is a matrix of 1's and –1's that is orthogonal such that:

$$H_n H_n^T = nI_n$$

$$H_2 = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \qquad H_2 H_2^T = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix} = 2I_2$$

A $2^k \times 2^k$ Hadamard matrix can be generated recursively. For example, for k=2:

$$H_4 = \left[\begin{array}{c|c} H_2 & H_2 \\ \hline H_2 & -H_2 \end{array}\right] = \left[\begin{array}{cc|cc} 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 \\ \hline 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \end{array}\right]$$

Removing the first column, a Hadamard matrix makes a balanced Fractional Factorial design. Indeed …

A Hadamard matrix makes a design of Resolution III:
any 2 columns are such that the 4 combinations
(1, 1), (1, –1), (–1, 1) and (–1, –1) appear equally often in those columns

$$\left[\begin{array}{c|cc} 1 & 1 & 1 \\ -1 & 1 & -1 \\ \hline 1 & -1 & -1 \\ -1 & -1 & 1 \end{array}\right]$$

[It is possible to make Resolution IV designs where any 3 columns are such that the 8 combinations:
(1, 1, 1), (1, 1, –1), (1, –1, 1), (1, –1, –1), (–1, 1, 1), (–1, 1, –1), (–1, –1, 1) and (–1, –1, –1) appear equally often in those columns]

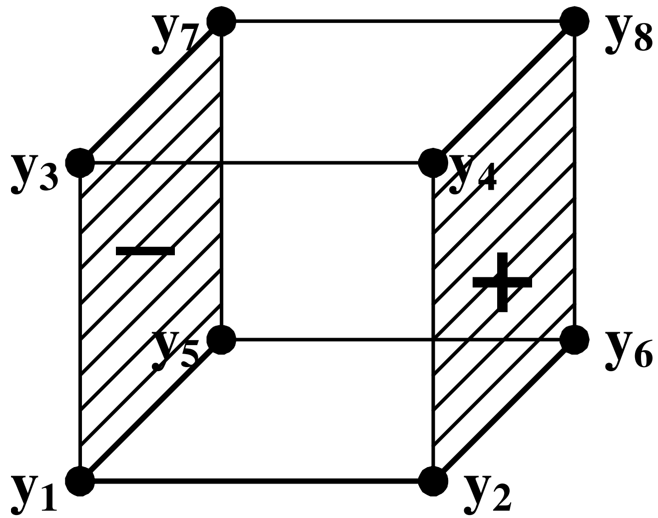
$$H_8 = \left[\begin{array}{c|c} H_4 & H_4 \\ \hline H_4 & -H_4 \end{array} \right]$$

Resolution III design with 7 factors (H_8):

X_1	X_2	X_3	X_4	X_5	X_6	X_7
1	1	1	1	1	1	1
–1	1	–1	1	–1	1	–1
1	–1	–1	1	1	–1	–1
–1	–1	1	1	–1	–1	1
1	1	1	–1	–1	–1	–1
–1	1	–1	–1	1	–1	1
1	–1	–1	–1	–1	1	1
–1	–1	1	–1	1	1	–1

The main effect of factor X_i on Y in a FF design:

$$ME_i(y) = \frac{1}{2n}\left(\sum_{x_{ji}=1} y_j - \sum_{x_{ji}=-1} y_j\right)$$



For a 2-level FF design, the variance explained by parameter X_i alone is just $\hat{V}_i(y) = \left(ME_i(y)\right)^2$

(orthogonality of the Hadamard matrix)

Method of elementary effects: is able to “screen” a subset of few important inputs among the many (tenths, hundreds) often contained in models

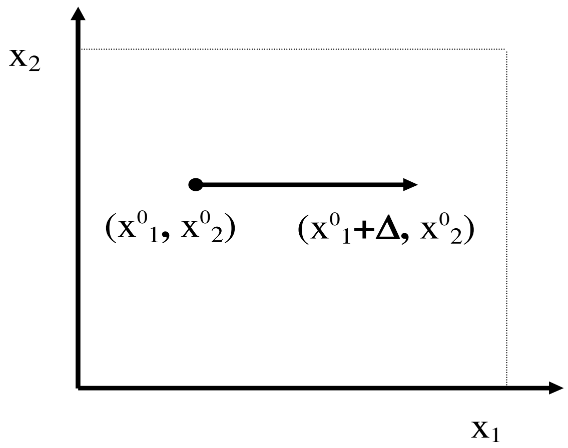
The method of elementary effects is simply a derivative (in the form of incremental ratios) computed at different points in the space of the input and averaged over the same space.

(Morris, 1991)

Model $y = y(x_1,..., x_k)$

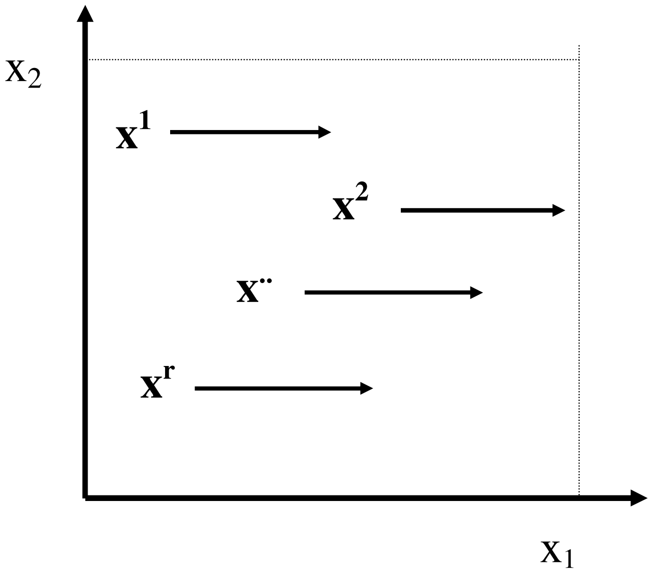
Elementary Effect for the i-th input in a point X°

$$EE_i(x_1^0,..., x_k^0) = \frac{y(x_1^0, x_2^0,..., x_{i-1}^0, x_i^0 + \Delta, x_{i+1}^0,..., x_k^0) - y(x_1^0,..., x_k^0)}{\Delta}$$



Δ is larger than in local methods

The average of several EE_i is taken.

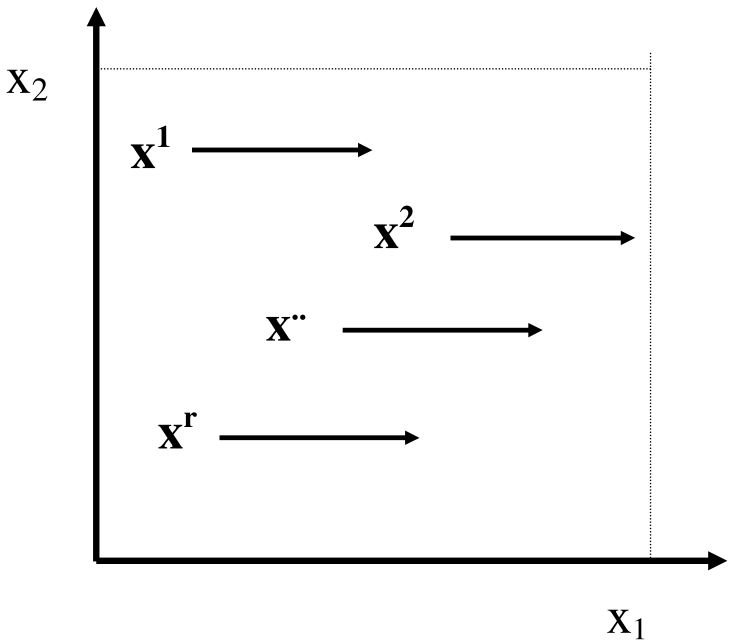


r elementary effects EE^1_i EE^2_i … EE^r_i are computed at X_1 , … , X_r

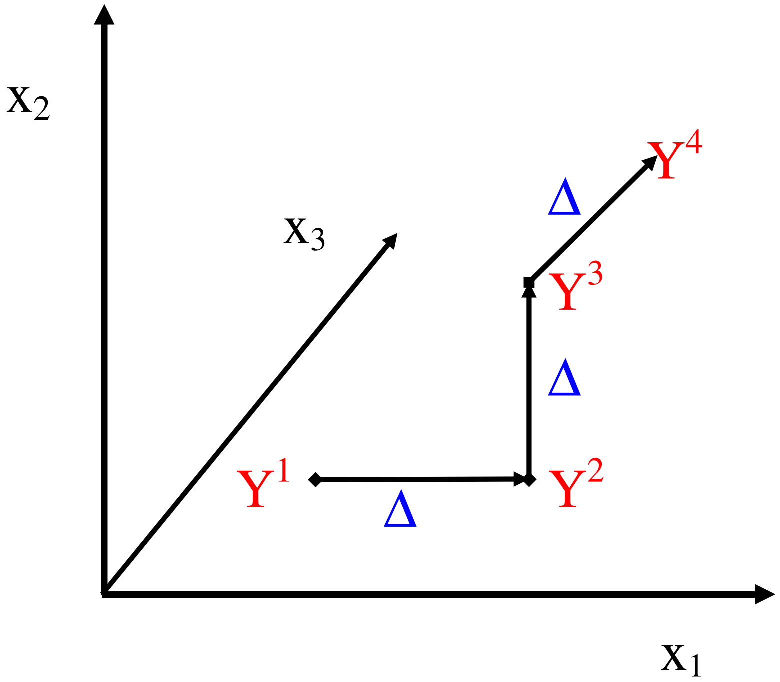
Average of $|EE_i|$'s $\rightarrow$ $\mu^*(x_i)$

$\mu^*(x_i)$ is useful to identify inputs that are not responsible for output variation

The standard deviation σ of the EE's is a measure of the sum of all interactions of x_i with other inputs



A clever sample design is used to build r trajectories of (k+1) sample points each providing one EE per input. The total cost of the screening is (k+1)r instead of 2kr.

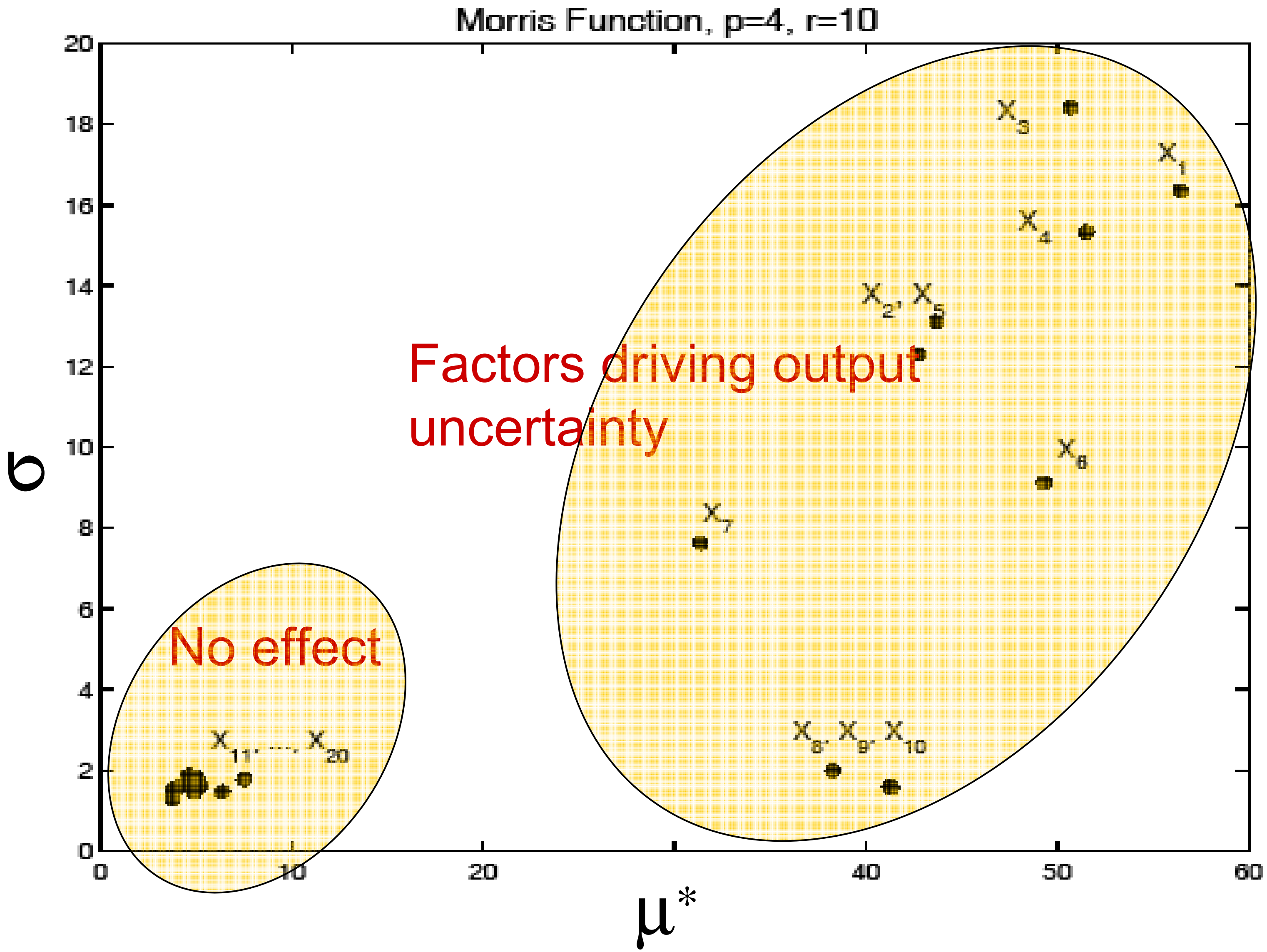


$$EE^1_1 = (Y^2 - Y^1) / \Delta$$

$$EE^1_2 = (Y^3 - Y^2) / \Delta$$

$$EE^1_3 = (Y^4 - Y^3) / \Delta$$

A graphical representation of results



EUROPEAN COMMISSION
DIRECTORATE-GENERAL
Joint Research Centre

Joint Research Centre

Saltelli, A., Ratto, M., Andres, T., Campolongo, F., Cariboni, J., Gatelli, D. Saisana, M., and Tarantola, S., 2008, Global Sensitivity Analysis The Primer John Wiley & Sons

Campolongo, F., Cariboni, J., Saltelli, A., 2007, An effective screening design for sensitivity analysis of large models, Environmental Modelling & Software, 22, 1509-1518

Saltelli, A., Tarantola, S., Campolongo, F., and Ratto, M., 2005, Sensitivity Analysis for Chemical Models, Chemical Reviews, 105(7), pp 2811 - 2828

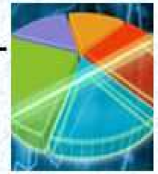
Saltelli, A., Tarantola, S., Campolongo, F., and Ratto, M., 2004, Sensitivity Analysis in Practice. A Guide to Assessing Scientific Models, John Wiley & Sons publishers, Probability and Statistics series

Saltelli A., Chan K., Scott E.M., 2000, Sensitivity Analysis, John Wiley & Sons publishers, Probability and Statistics series, 355-365

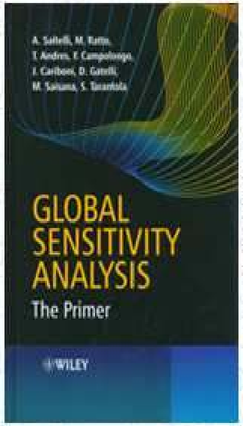
Morris M. D., 1991, Factorial sampling plans for preliminary computational experiments, Technometrics, 33(2): 161-174

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

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

 Print