Training to DGs on Statistical and Participatory Tools for Impact Assessment

Brussels, 28 – 29 April, 2016

Lecture 1
An introduction to Sensitivity Analysis of Model Output

Stefano Tarantola
Institute for Energy and Transport
Joint Research Centre of the European Commission

Stefano.tarantola@jrc.ec.europa.eu

The Joint Research Centre: the European Commission's science and knowledge service

"As the science and knowledge service of the Commission our mission is to support EU policies with independent evidence throughout the whole policy cycle"

www.ec.europa.eu/jrc/

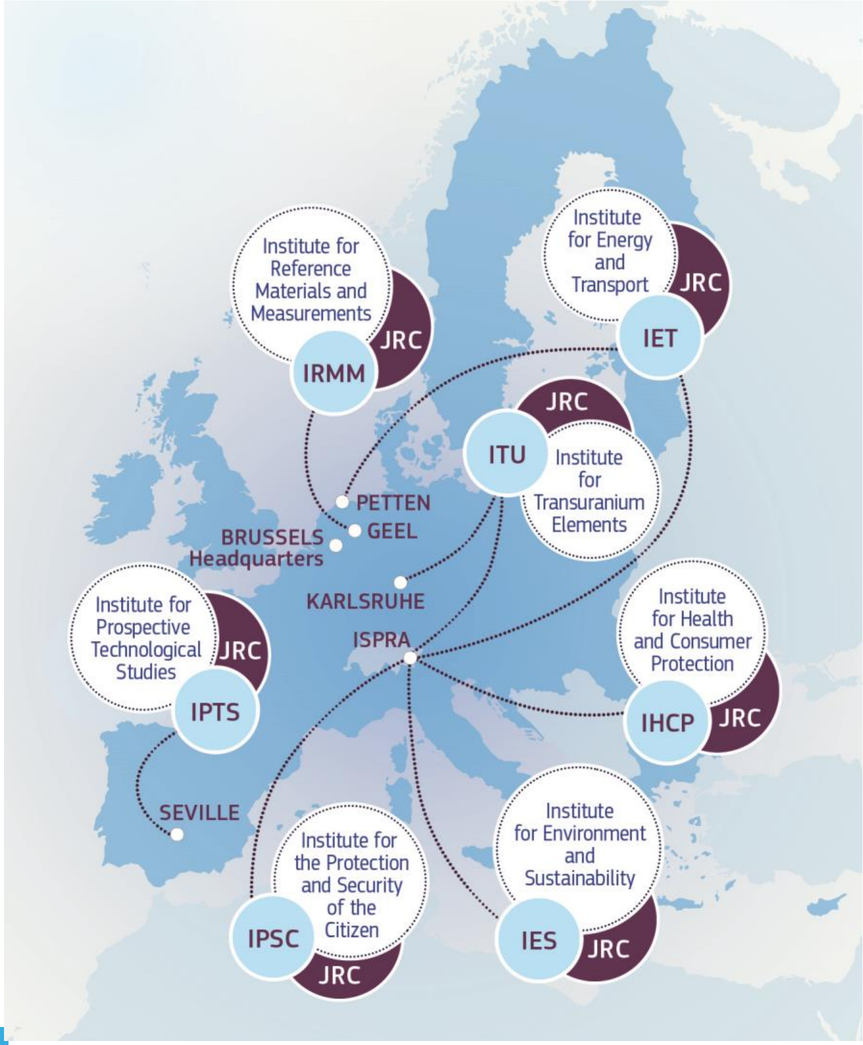
The JRC plays a central role in creating, managing and making sense of collective scientific knowledge for better EU policies.

Its core function is twofold:

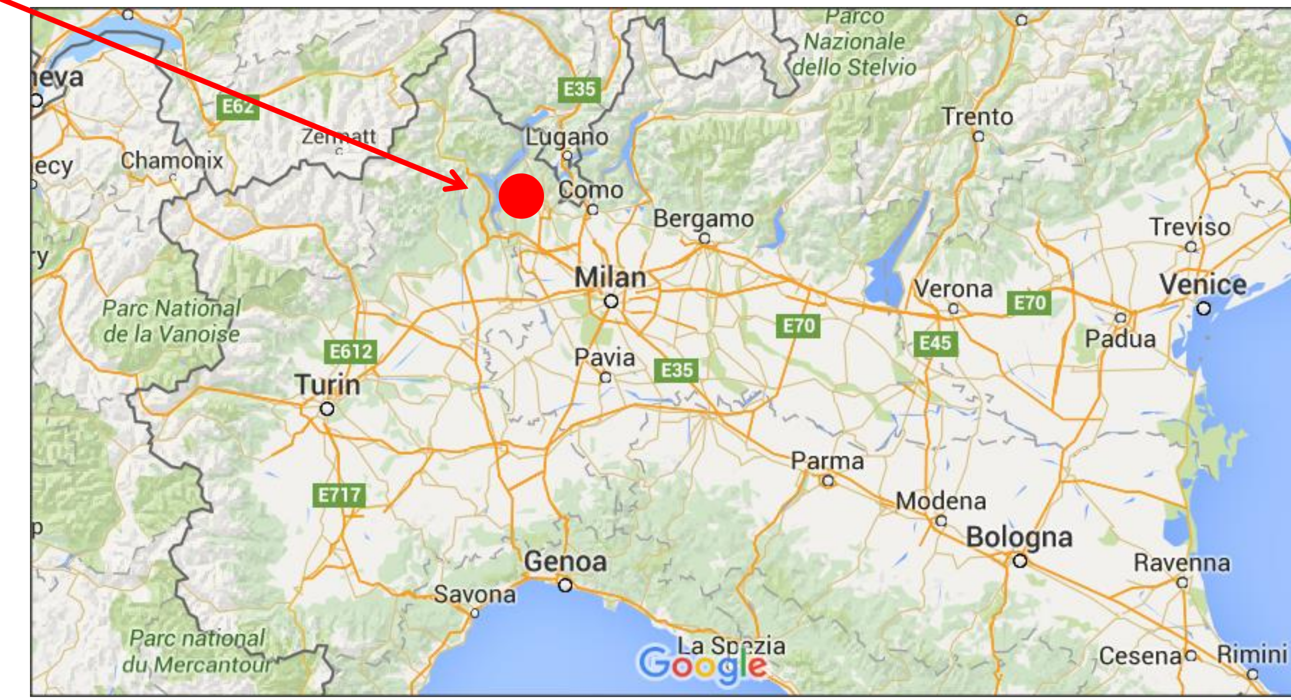
1. to create new scientific knowledge by carrying out research work
2. to manage knowledge from other sources (collate, analyse and communicate knowledge to policy makers).

JRC's output reflects all EU's overarching objectives:

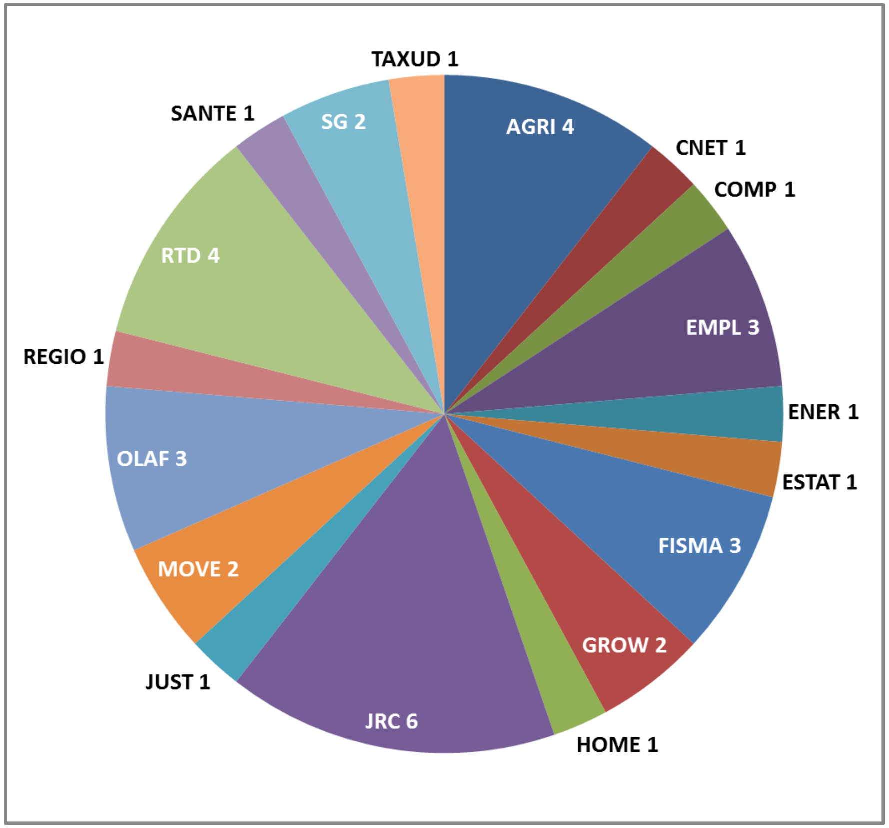
Economy, finance and markets
Energy and transport
Education, skills and employment
Food, nutrition and health
Environment and climate change
Civil security
Migration, data and digital transformations



The JRC Ispra ...



Your presence today
38 EC staff from 15 DGs



Modelling in Impact Assessment

It is increasingly recognised that modelling is an important tool to take into account all the main impacts of a policy, and make the best possible use of all the relevant data

However, modelling is not yet widely used to produce impact assessments for policy making in the Commission; in the period 2009-2014, roughly 15% of all Impact Assessments (IA) (486) used evidence from modelling, and some DGs make very little or no use of modelling in their IAs.

Modelling in Impact Assessment

The JRC is a major source of support to the Commission (roughly 50% of the IAs in the period 2009-2014 that used model results included modelling by the JRC).

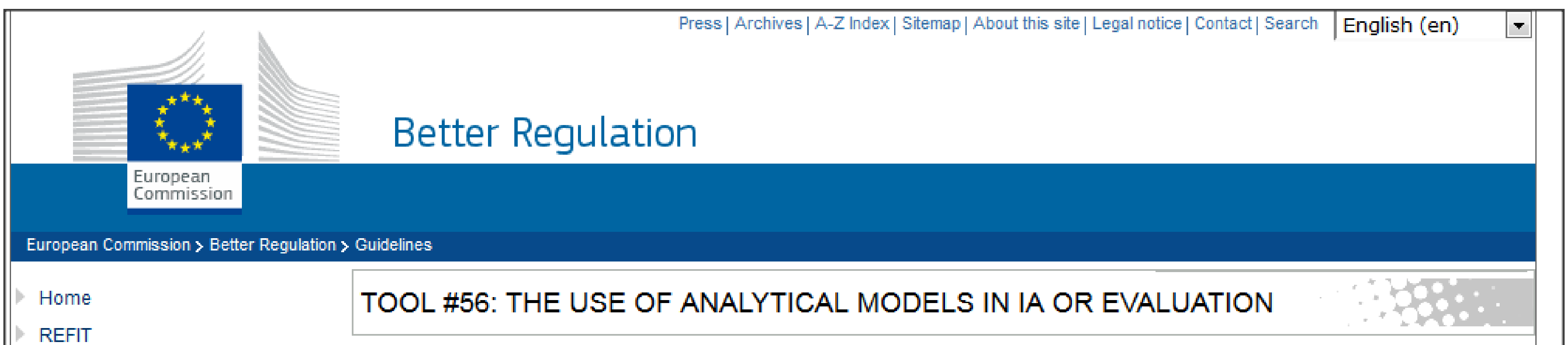
Some other DGs, notably ECFIN, REGIO and AGRI, also have their own modelling capacity.

The JRC has several decades of experience in sensitivity and uncertainty analysis to improve the robustness of modelling.

The JRC also has recently set up an inventory (MIDAS) of its own models; this is currently being expanded to be a Commission-wide inventory.

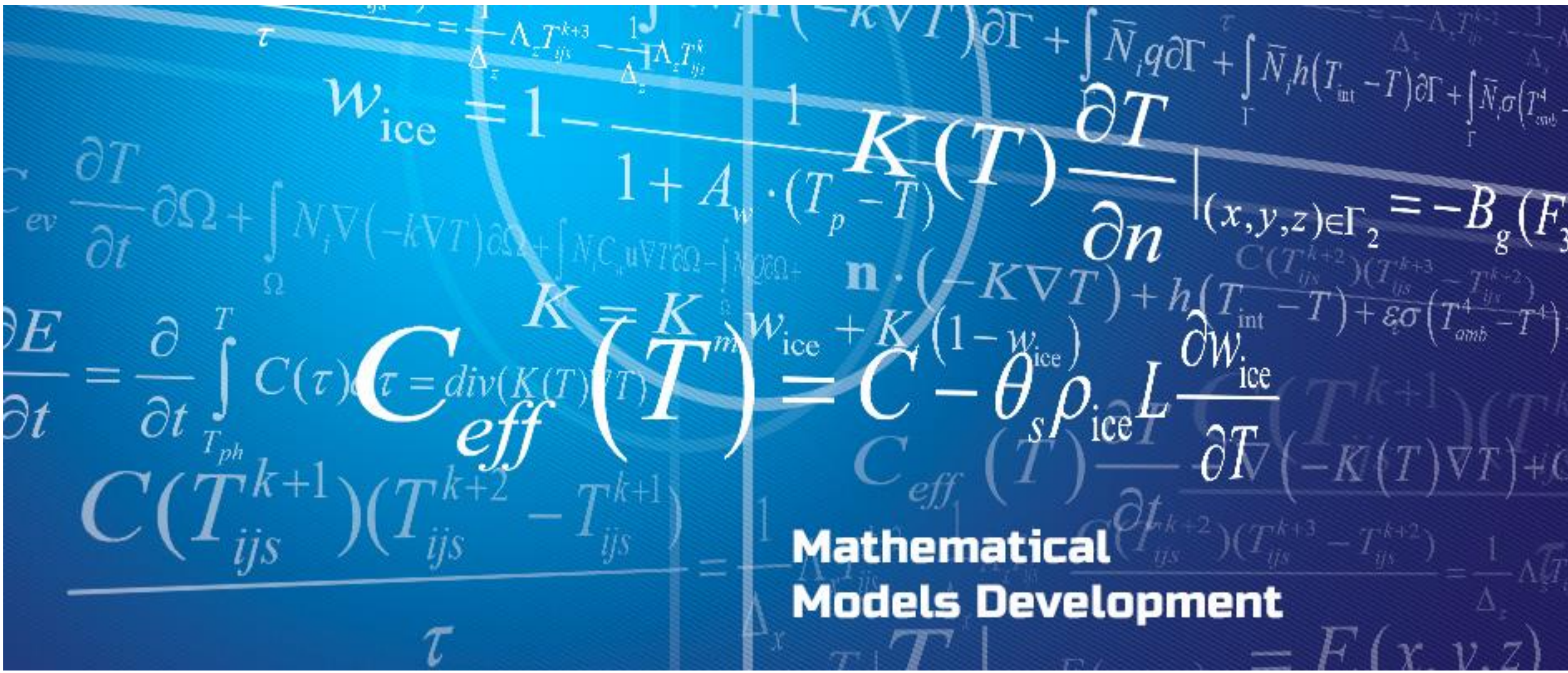
All officials involved in the preparation of impact assessments are responsible for ensuring their quality.

Guidance is provided by the **better regulation guidelines**. These are accompanied by a ***toolbox*** structured around various chapters. One of these chapters is:

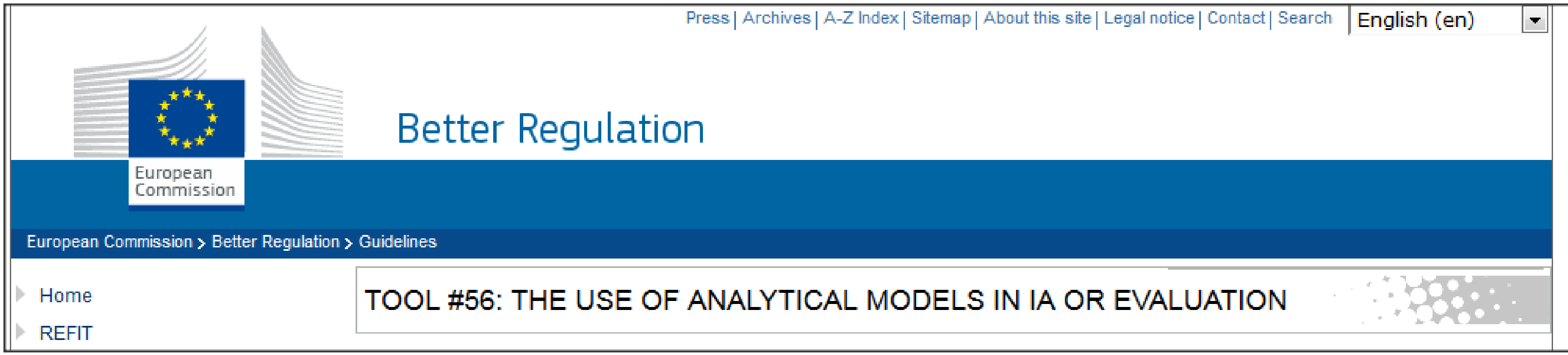


European
Commission

Models in Impact Assessment



Joint
Research
Centre

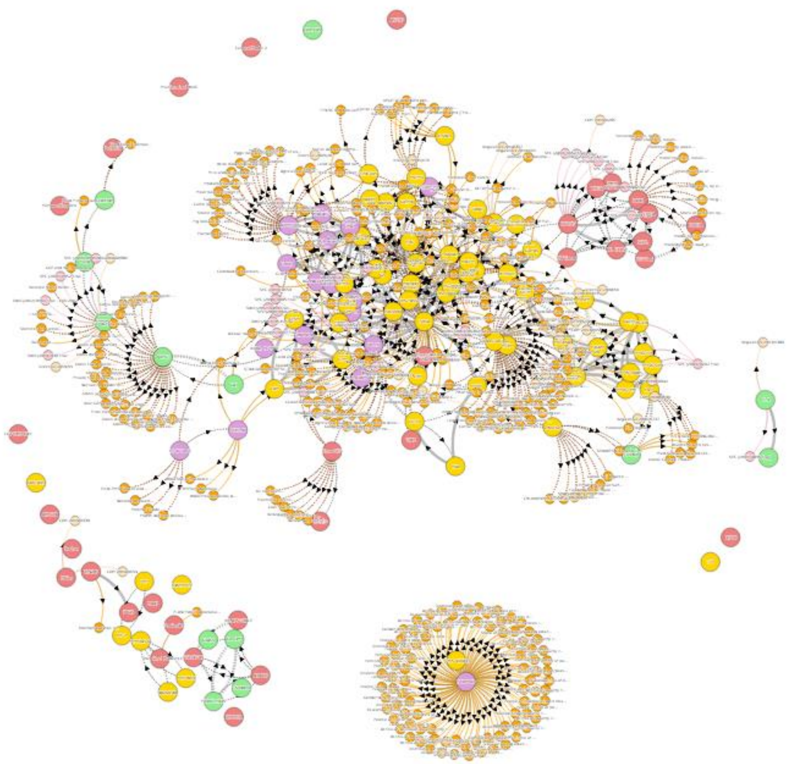


"All models are simplifications, but good models provide insights and understanding if used correctly."

"It is important to ensure that the right model is selected and used in a manner to deliver policy relevant results of the requisite quality."

MIDAS

Modelling Inventory Database & Access Services



Linking together …

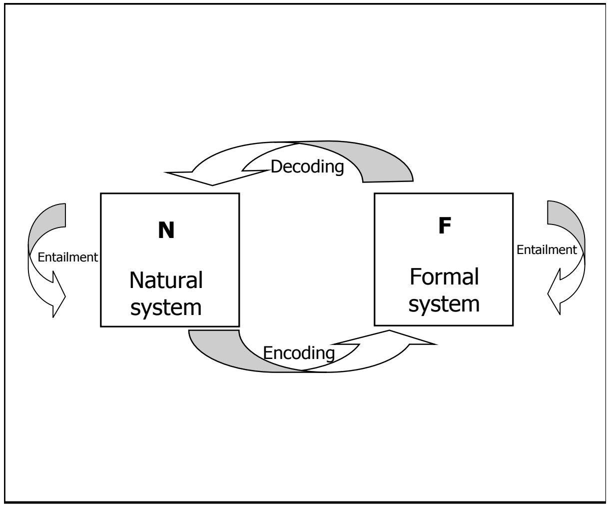
- Models
- Data
- Policies
- Experts
- Publications

http:

Accessible from within the Commission Network

European Commission

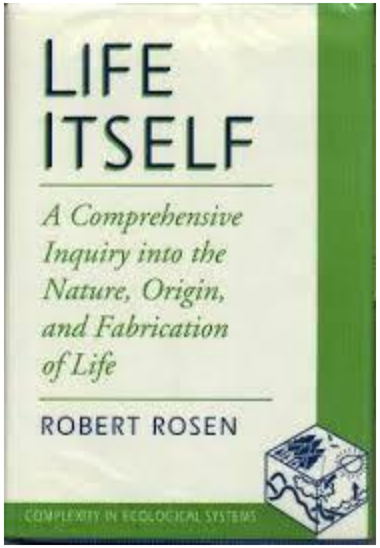
In many cases quantification of impacts relies on assumptions (e.g. monetised value of a fatality, levels of inflation, return on investment, etc) and is affected by uncertainty in input data, model parameters and model components.



Modelling is a craftmanship …
(R. Rosen, Life itself 1991)

Modelling is an encoding process from the phenomenon to the formal system: theories, laws, subjective choices, values, assumptions

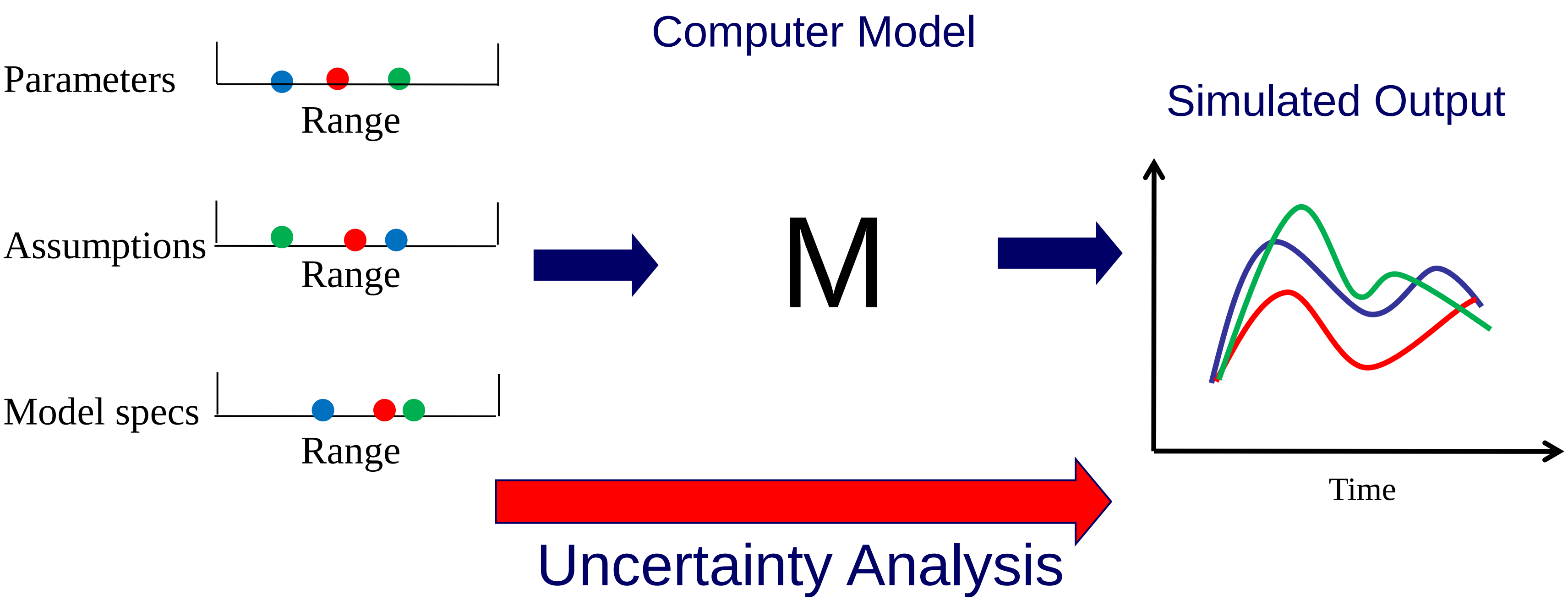
Different assumptions can lead to different predictions



These sources of uncertainty must be made explicit in the impact assessment and it is important to quantify how much they affect the resulting model output.

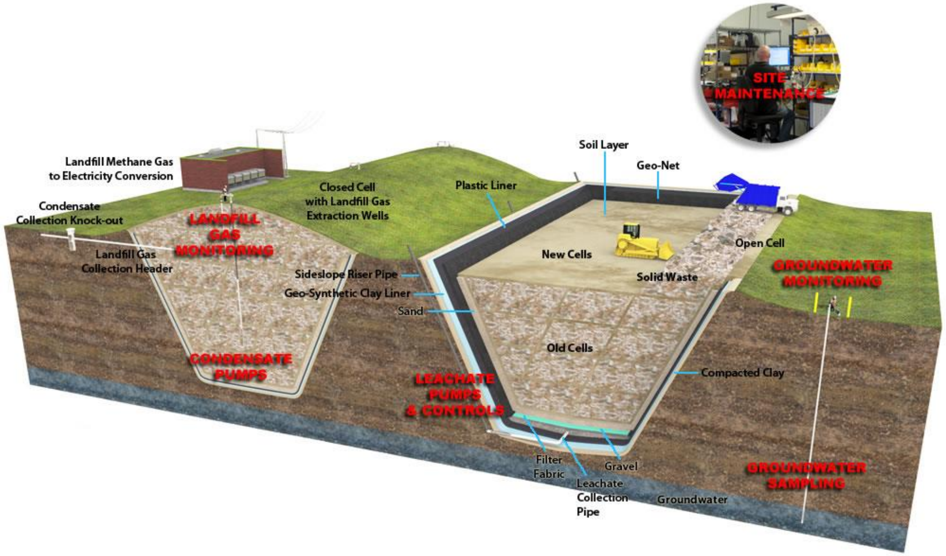
Ranges, or confidence intervals, for the model output should be provided rather than analytical fixes.

Uncertainty analysis: the 'forward process'



Example of waste disposal policy: a model is used to evaluate the environmental impact of two alternative policy options for waste disposal

Incineration



Landfill

Joint Research Centre

The model computes an environmental impact index by aggregating several indicators of atmospheric /soil emissions

NOx, SOx, PM, CO, Hg, Cd, CO2, CH4, HM, phenols, benzene, ammonia, dioxines, etc

Emissions (Kg NOx) =

Activity rates (Tons waste) * Emission Coeffs (Kg NOx /Tons waste)

Environmental Impact = function (weights * normalized emissions)

A typical impact assessment would supply one value per policy:

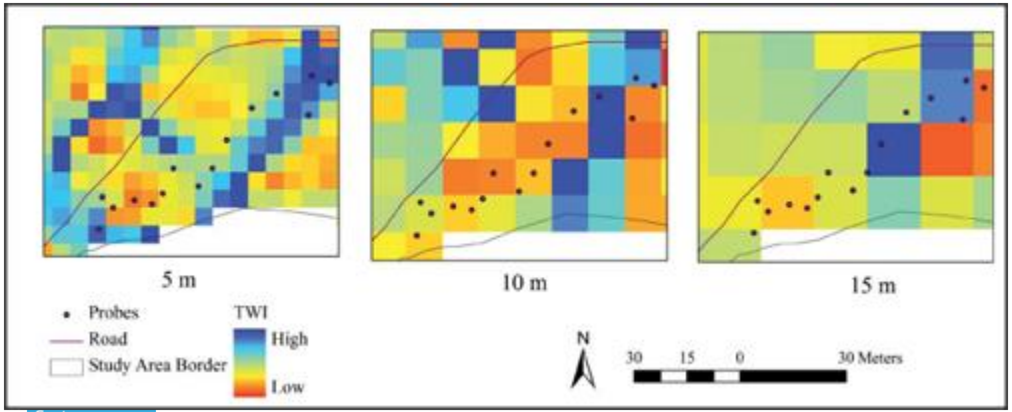
Impact (incineration) = 29

Impact (landfill) = 22

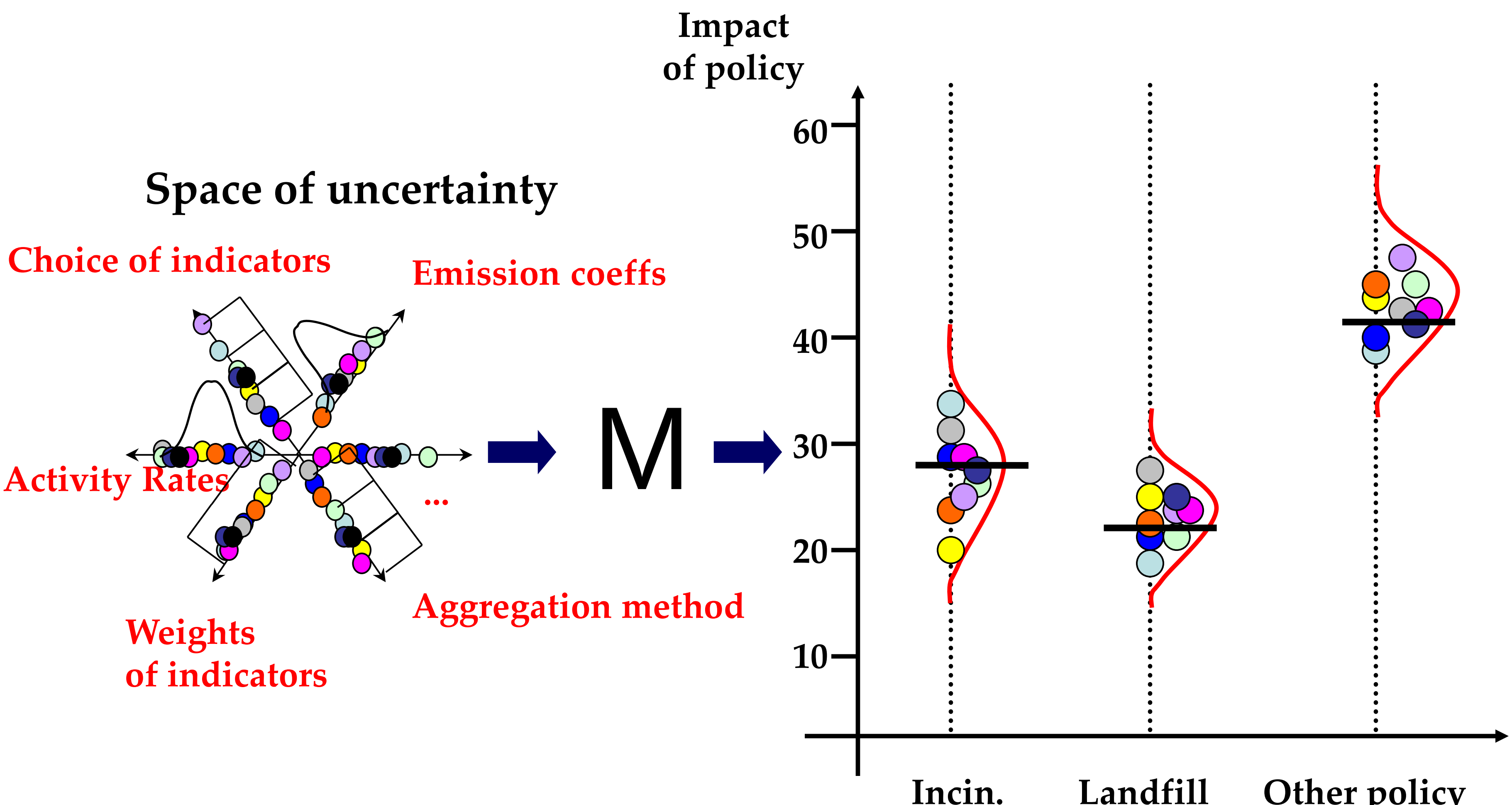
Landfill better than incineration

Various uncertainties are present in the model:

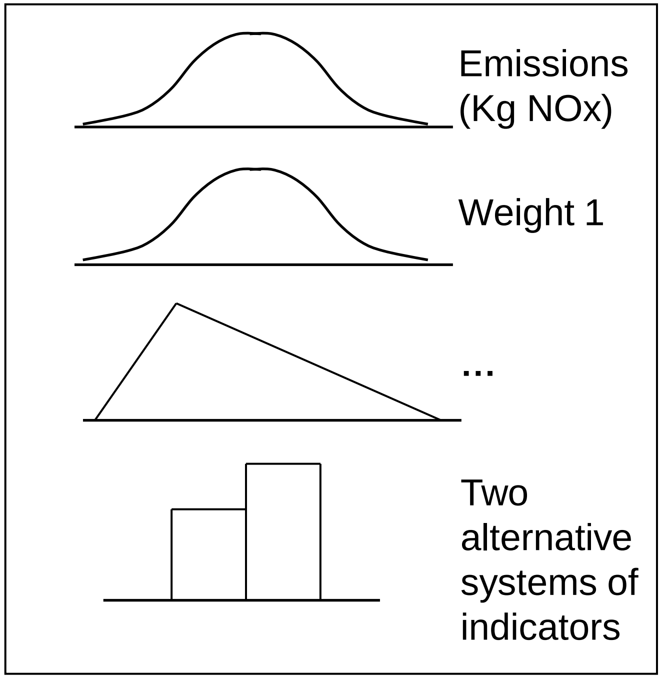
- Modelling assumptions (e.g. choice of impact indicators),
- Activity rates (Tons of waste),
- Emission Coefficients (Kg pollutant / Ton waste)
- Normalization of the indicators
- Weights of the indicators
- Type of aggregation of the indicators
- Data at different resolution levels



Probabilistic approach to uncertainty analysis



Specification of the model inputs: a delicate step



The following information can assist:

Available data / measurements

Scientific literature

Physical bounds

Expert judgement

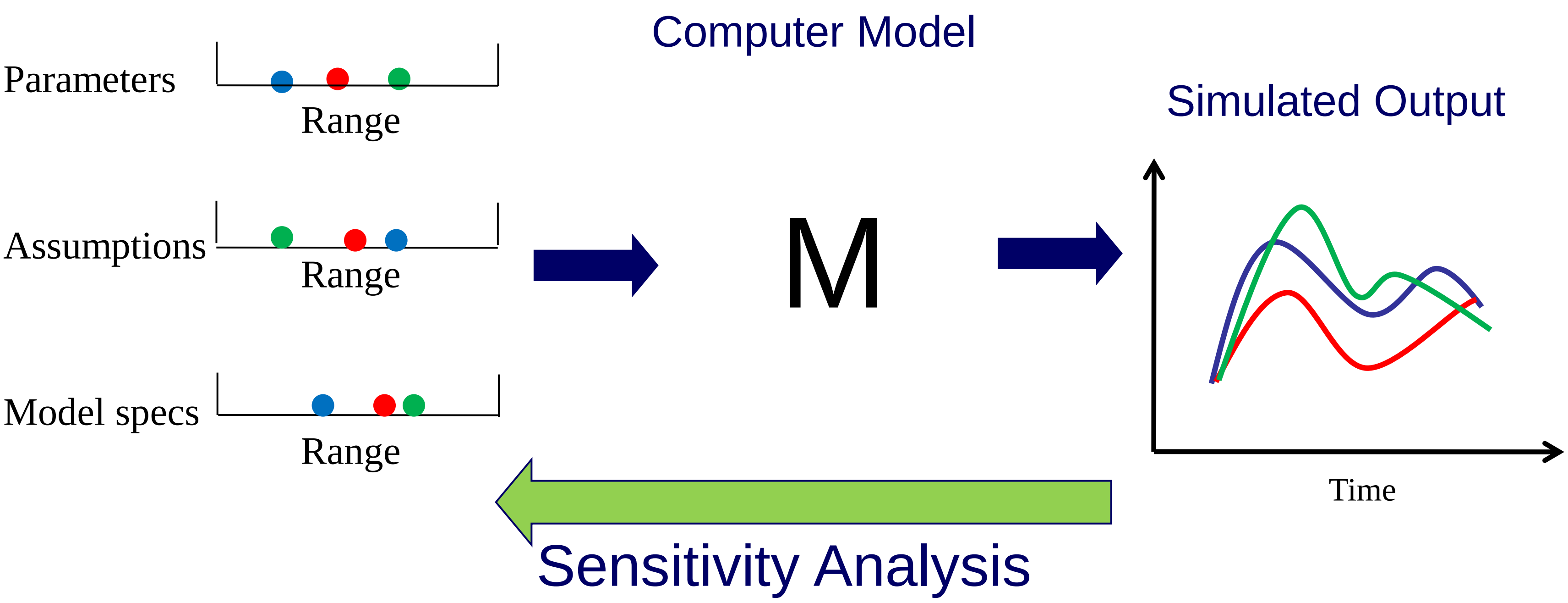
Opinion polls, surveys

How much does a change in model factors (parameters, assumptions, model spec's) affect the model response?

Would a change of model specifications, eg the use of a different set of indicators (NOx instead of SOx, or PM instead of CH4) influence the model output?

Would a change of assumptions, eg the choice of different set of weights for the indicators, influence the resulting environmental impact?

Sensitivity analysis: the ‘backward process’



Sensitivity analysis is the backward process. It provides users of mathematical and simulation models with tools to appreciate the dependency of the model output from model input, and to investigate how important is each model input in determining its output.

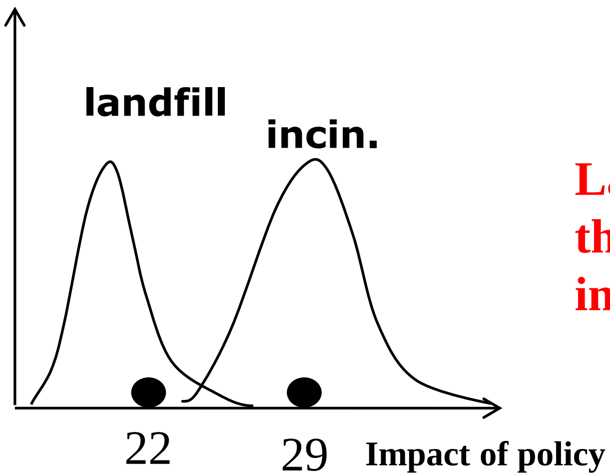
Sensitivity Analysis: a necessary ingredient for modelling

- Identify inputs causing most uncertainty – direct research or information gathering
- Model simplification - identify inputs that do not affect the output, therefore redundant
- anticipate (prepare against) falsifications of the IA results
- Model calibration: identification of optimal parameter sets

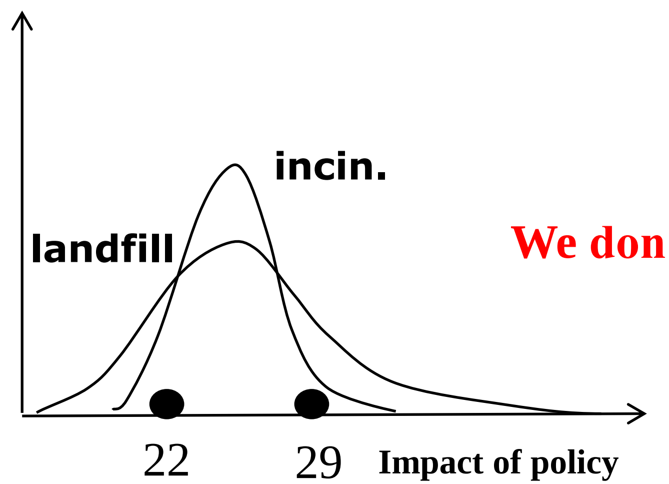
Sensitivity Analysis: a necessary ingredient for modelling

- Check the effect of model assumptions on model predictions
- Identify critical regions in the space of the inputs, e.g. instabilities of the model
- Surprise the analyst – finding unexpected relationships between inputs and outputs

Probabilistic approach to uncertainty analysis

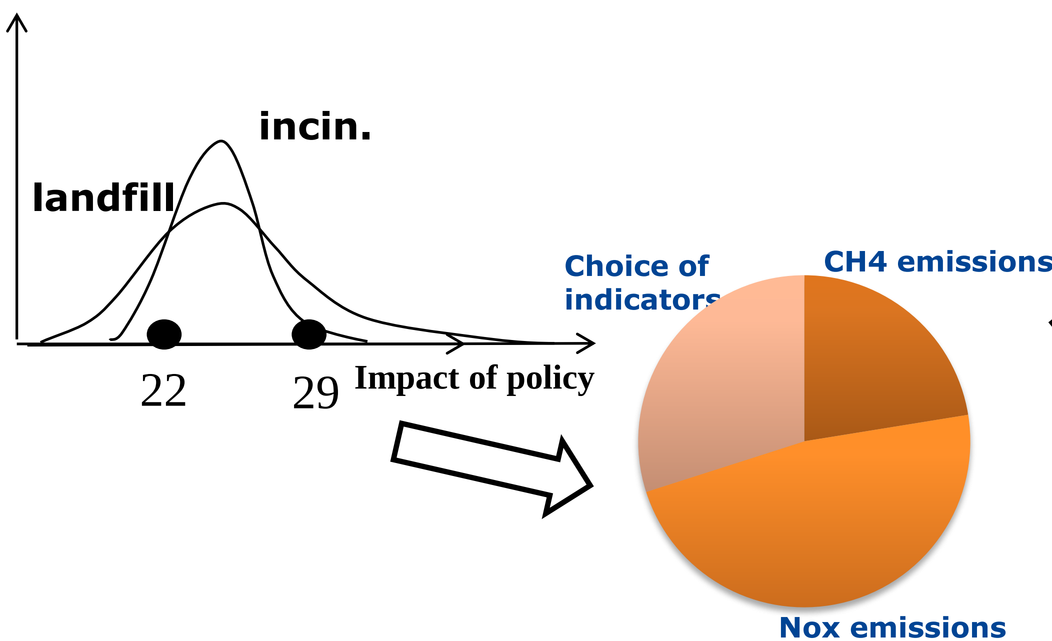


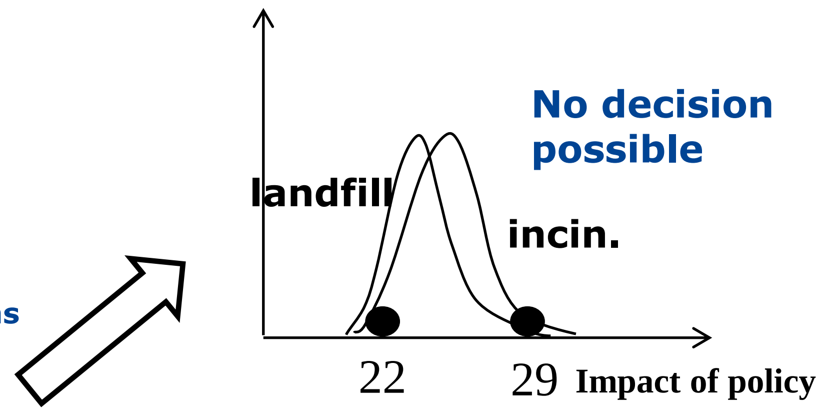
Landfill better than incineration

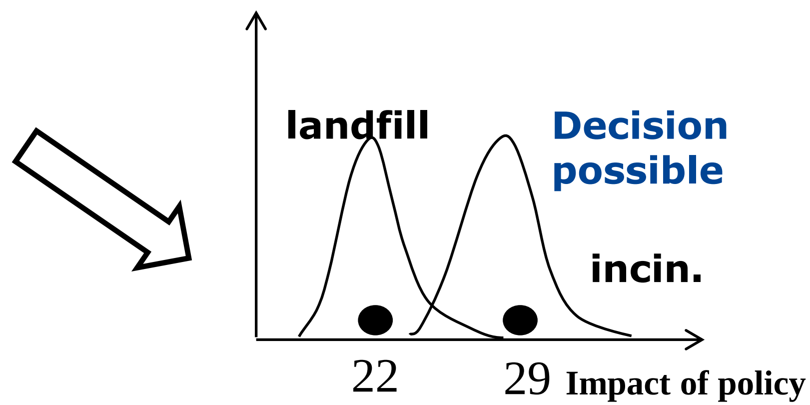


We don't know

- Perform SA
- reduce uncertainty of most important inputs if possible (eg, NOx emissions)
- recalculate the impacts
- this could make the decision possible







The role of uncertainty/sensitivity analysis when comparing options is to facilitate the identification of the preferred option, if possible.
If the identification is not possible, this is a useful result anyway as we know that we should not waste time in discussing about which is the better option.

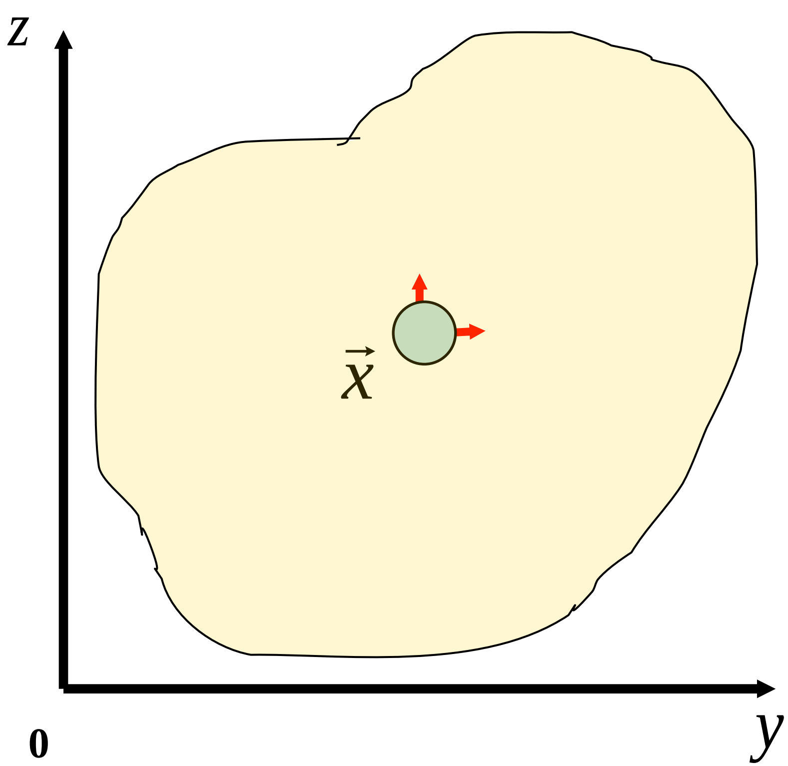
The first historical approach to SA is known as the local approach. The impact of small input perturbations on the model output is studied.

These small perturbations occur around nominal values (the mean of an input, for instance).

This deterministic approach consists of calculating or estimating the partial derivatives of the model output at a specific point of the input space

Local Sensitivity Analysis

$$\boldsymbol{x} \rightarrow \mathrm{M} \rightarrow f(x)$$



$$\boxed{\left.\frac{\partial f}{\partial y}\right|_{x=\vec{x}}} \qquad \boxed{\left.\frac{\partial f}{\partial z}\right|_{x=\vec{x}}}$$

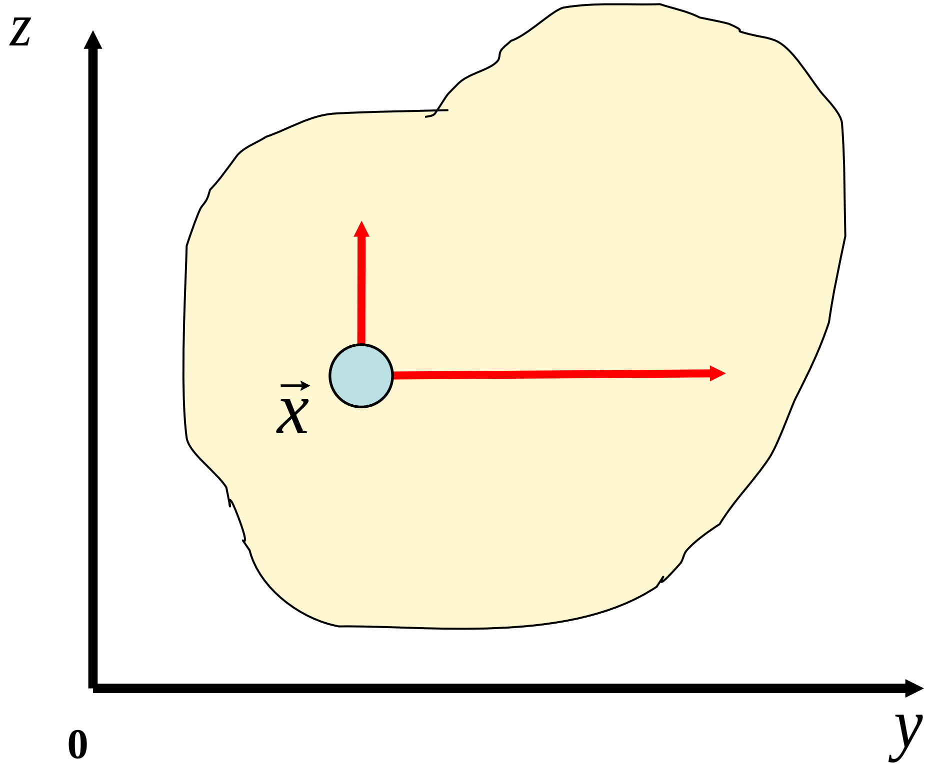
$x = (y, z)$ **= space of input**

$\vec{x}$ **= nominal value**

One at a time SA

- evaluation of model output at two different points for a given input

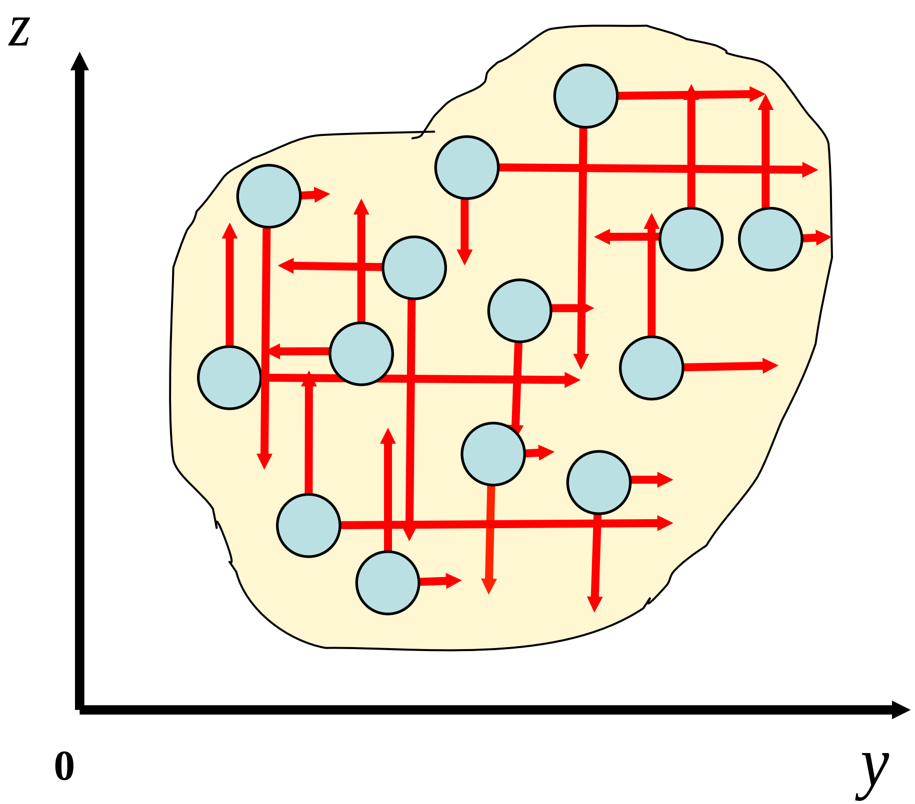
- the other inputs are kept fixed



$x = (y, z)$ = **space of input**

$\vec{x}$ = **nominal value**

Global SA



$x = (y, z)$ = **space of input**

- full exploration of uncertainty

- Monte Carlo methods to generate samples

Regression / correlation
Screening techniques
Variance decomposition
Moment- independent
Statistical tests
Graphical tools

European Commission

Preference of using OAT

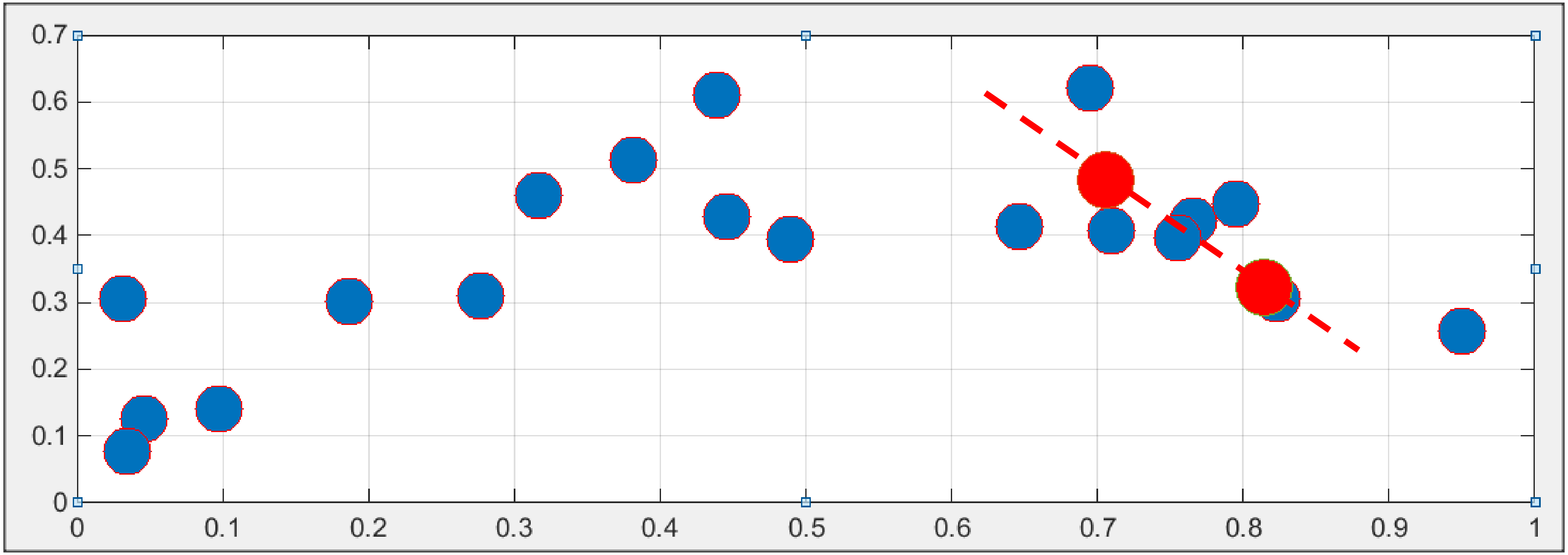
Global SA explores the boundaries of the input space and the modeller is afraid that the model will crash.

That’s the main reason why most of the sensitivity analysis that we encounter today are still local or 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).

OAT vs Global analysis

Model prediction



Generic model input

Which approach to use ("one at a time‟ or "global”) depends on the complexity of the model and the available resources (computing, time, personnel).

“One-at-a-time” is most commonly used in Commission IAs.

The “global” approach allows for the simultaneous exploration of all sources of known uncertainty and captures nonlinearities and interactions between model inputs. It is more expensive to conduct

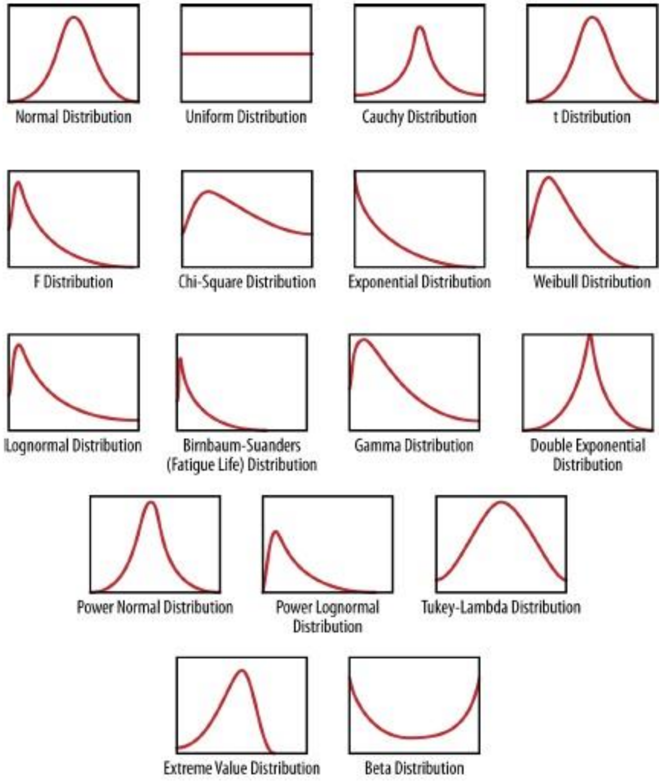
The basic steps to performing GSA are as follows:

A. Define a variable of interest for the analysis. It should be the main model output of interest in the impact assessment, and can be an aggregation of spatially distributed or time-dependent model outputs. Examples: the total cost of damage due to an accident at a chemical plant; the width of the area where a pollutant is above a given limit, the maximum radioactive dose in a given time span.

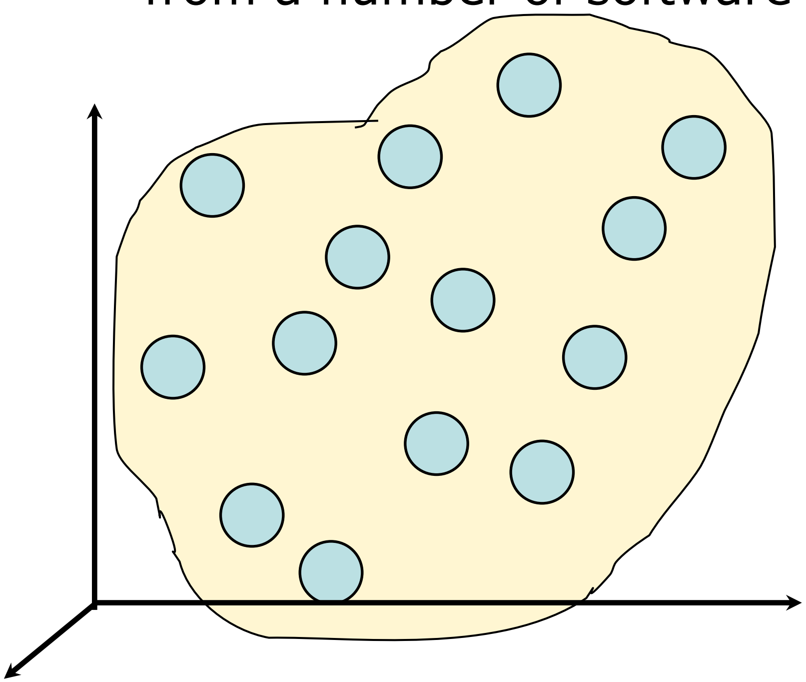
B. Identify all model variables which are affected by uncertainty in consultation with experts and stakeholders as appropriate. Inputs can be of various natures, i.e. scalar variables, time series or spatially distributed maps.

European
Commission

C. Characterise the uncertainty for each selected input by assigning a probability distribution using all available information such as experiments, estimations, physical bounds considerations and expert opinion. This step may require significant resources. Extended peer-review should be considered to ensure quality in the treatment of uncertainty.

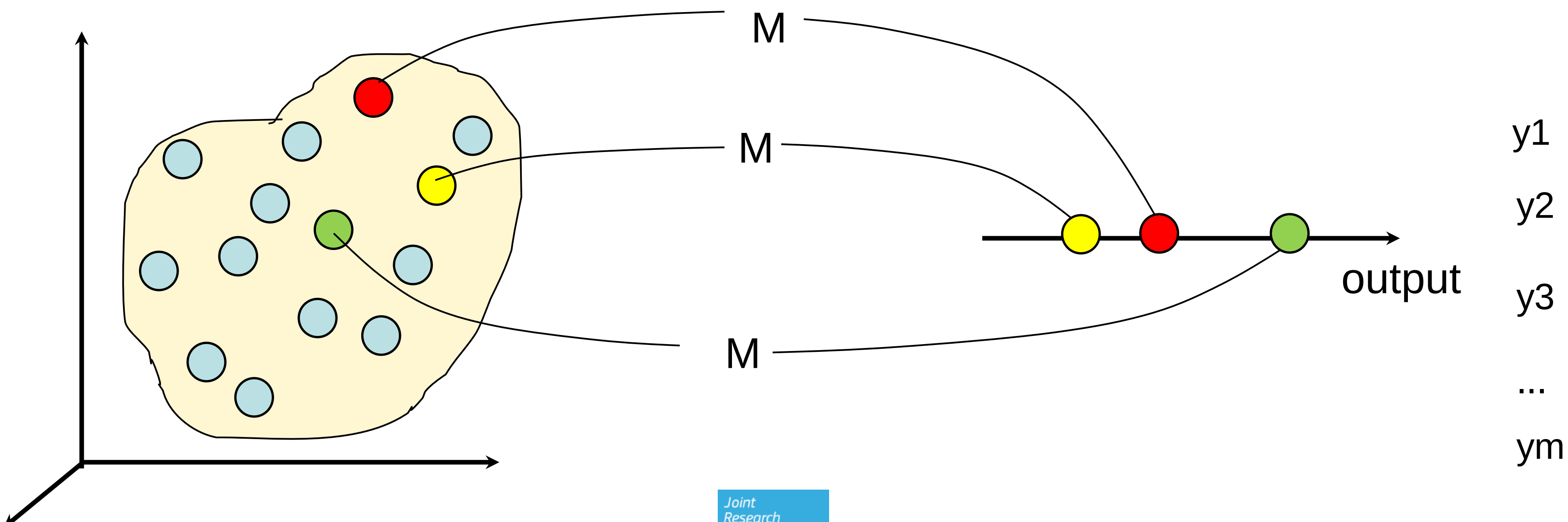


D. Generate a sample from the previously defined probability distributions. The sample is a matrix which specifies the input values to use for each model run and is designed to allow the calculation of sensitivity. The sample is generated so as to explore the full extent of uncertainty. Such samples can be generated from a number of software packages.

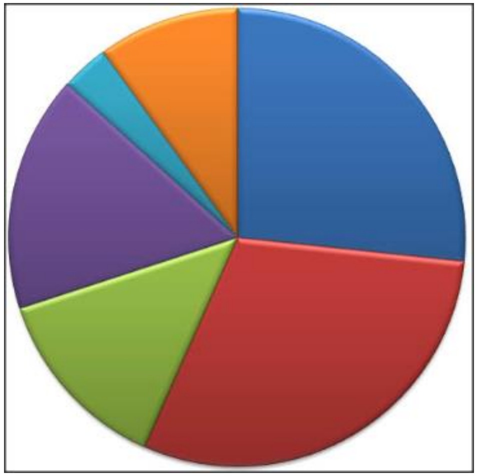


$$\begin{array}{c} 1 \\ 2 \\ 3 \\ \cdots \\ m \end{array} \begin{vmatrix} x11 & x12 & x13 \\ x21 & x22 & x23 \\ x31 & x32 & x33 \\ \cdots & \cdots & \cdots \\ xm1 & xm2 & xm3 \end{vmatrix}$$

E. The model is run many times using the sampled input variables for each model run as identified in the previous step. For each run, the value of the output variable of interest is recorded. This process is usually accomplished automatically using computer software.

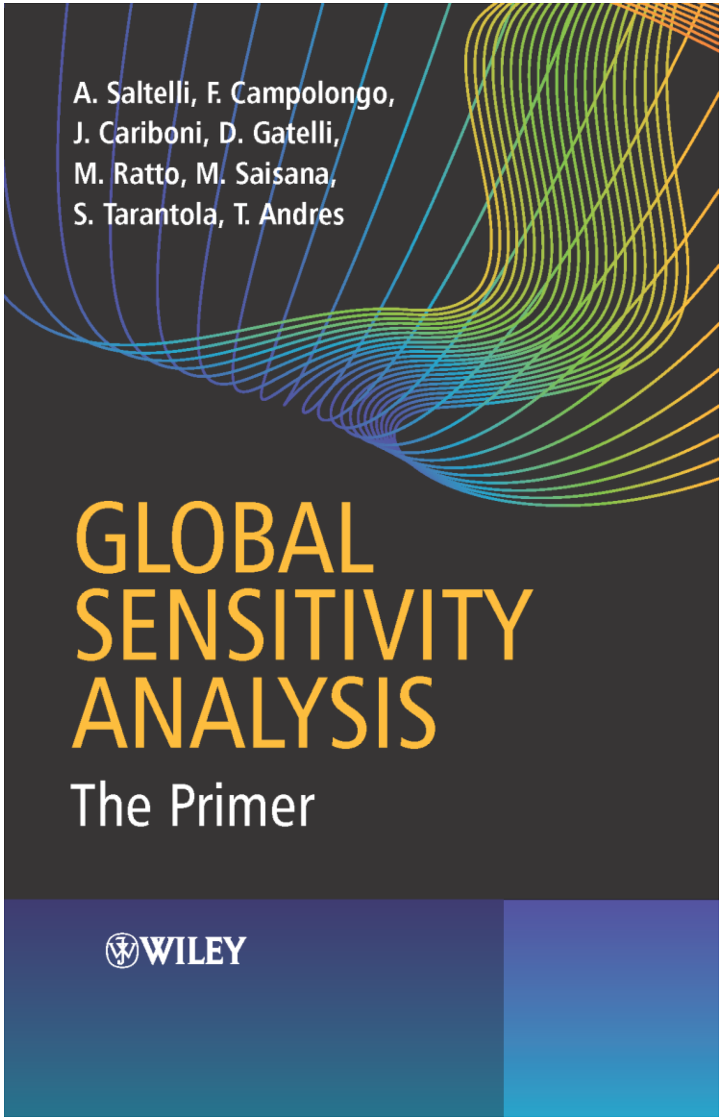


F. The results of the model runs are then used to estimate sensitivity, as well as the uncertainty in the model output. The sensitivity formulas yield the fractional contribution of each input to the output variance.



$$S_y = \frac{1}{V(f)}\left[\int p(y)\,dy\left[\int f(y,z)\;p(z)\,dz\right]^2 - f_0^{\,2}\right]$$

How to conduct an SA?



Global sensitivity analysis. The Primer

A textbook of methods to evidence how model-based inference depends upon model specifications and assumptions,

John Wiley, 2008

Saltelli, A., Ratto M., Andres, T., Campolongo, F., Cariboni J., Gatelli D., Ratto, M., Saisana, M., Tarantola, S.

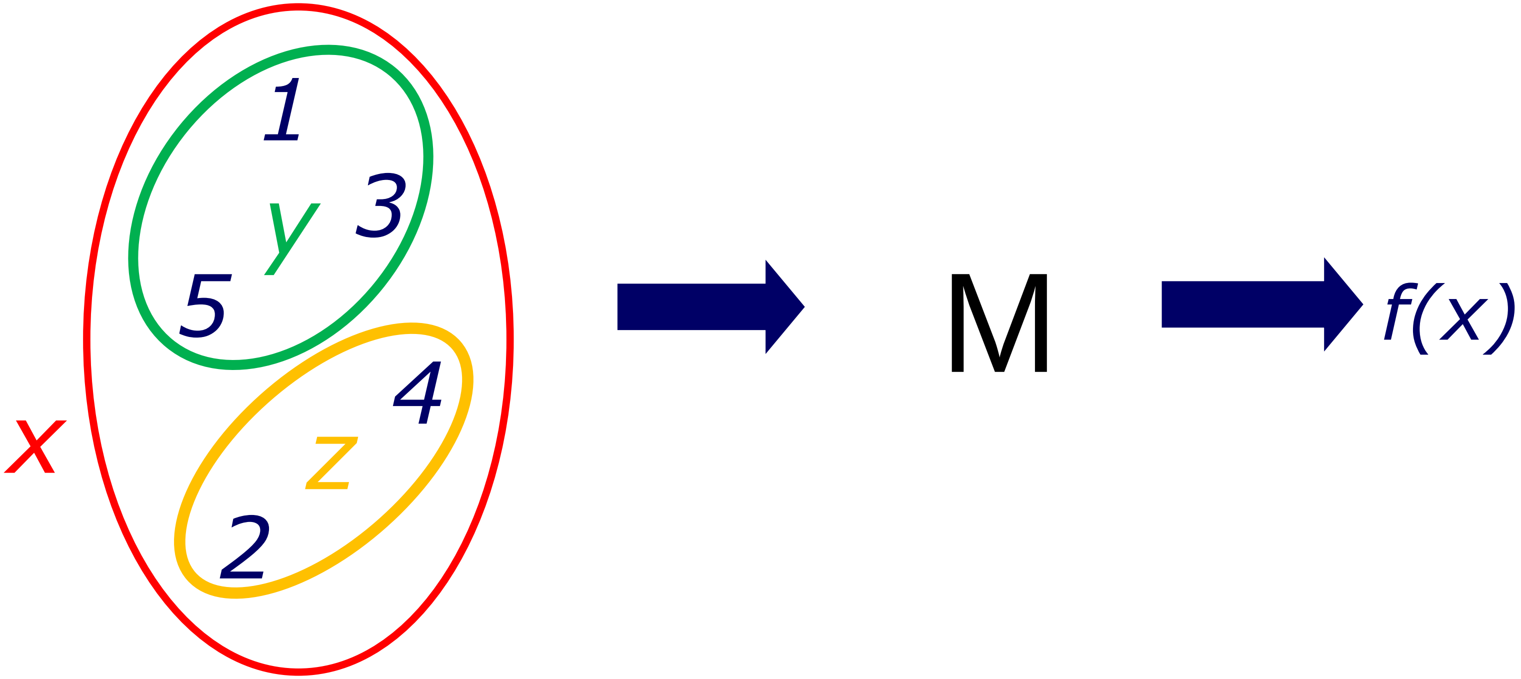
Variance-based techniques

And Saltelli/Sobol'/Jansen's formulas

Joint Research Centre

Variance-based method’s best formalization is based on the work of Ilya M. Sobol’(1990)

Some Notation ...

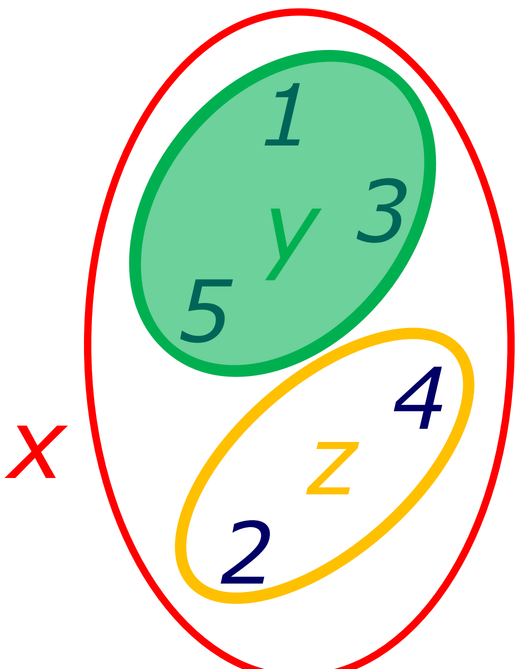


$$x = (y, z)$$

First-order sensitivity indices

$$V(f) = V[E(f \mid \circ)] + E[V(f \mid \circ)]$$

Explained variance + Residual variance



$$V(f) = V[E(f \mid y)] + E[V(f \mid y)]$$

$$S_y = \frac{V[E(f \mid y)]}{V(f)}$$

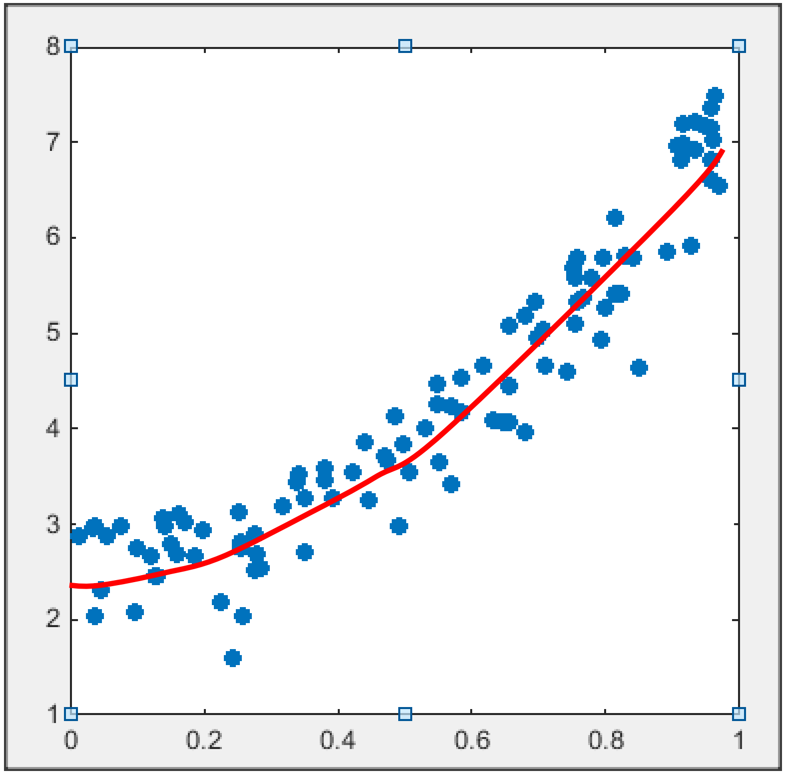
$$S_z = \frac{V[E(f \mid z)]}{V(f)}$$

European Commission

$S_y \uparrow$

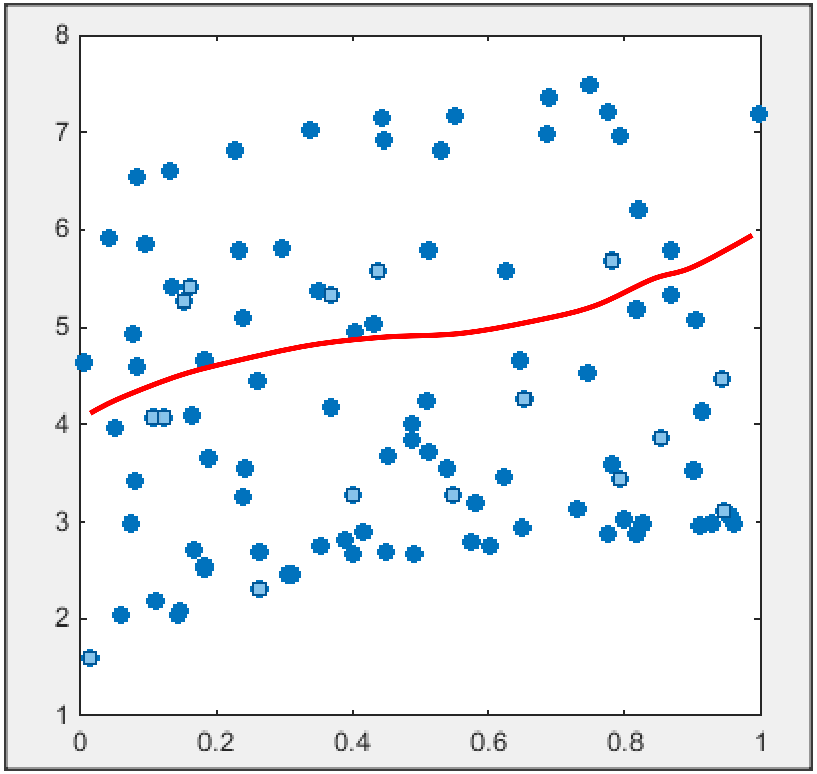
$S_z \downarrow$

f



y

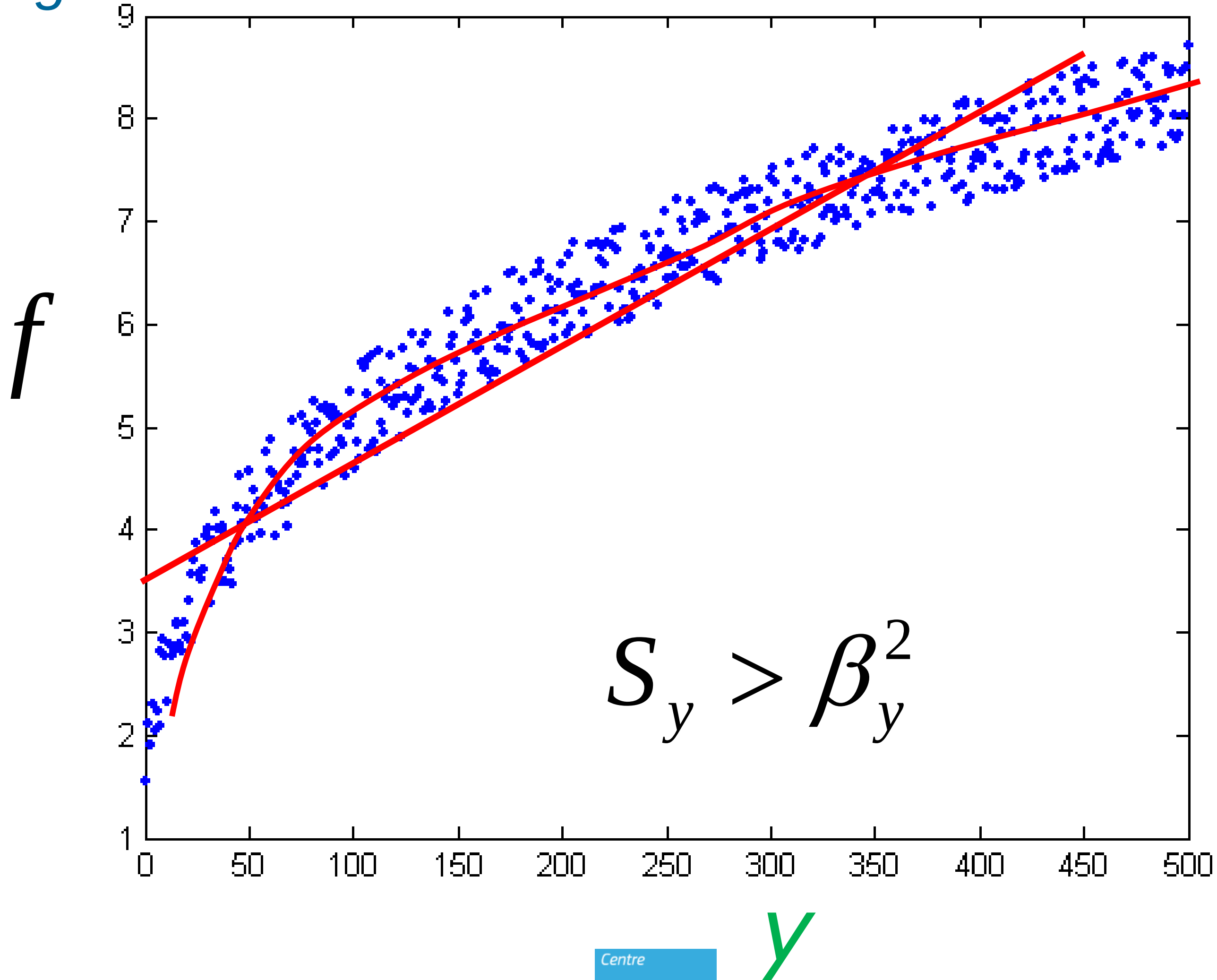
f



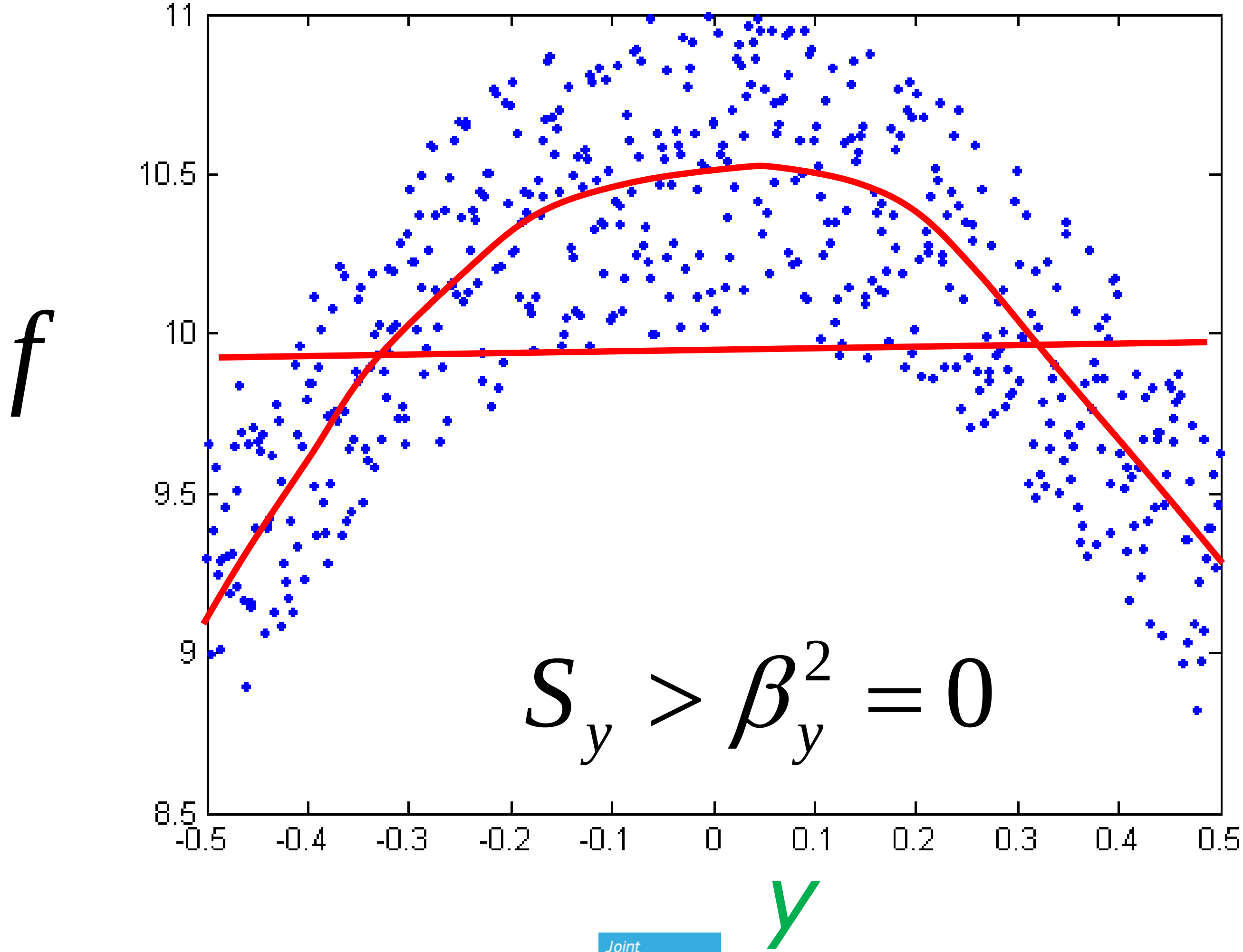
z

Joint Research Centre

First order indices explain more output uncertainty than linear regression



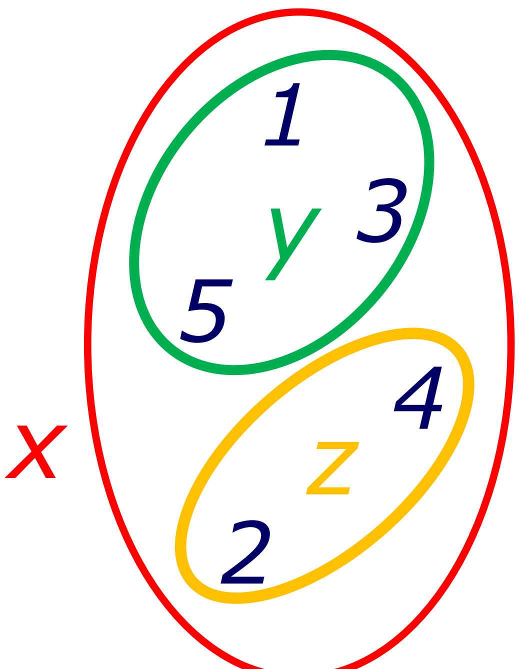
Here linear regression does not explain anything



Properties

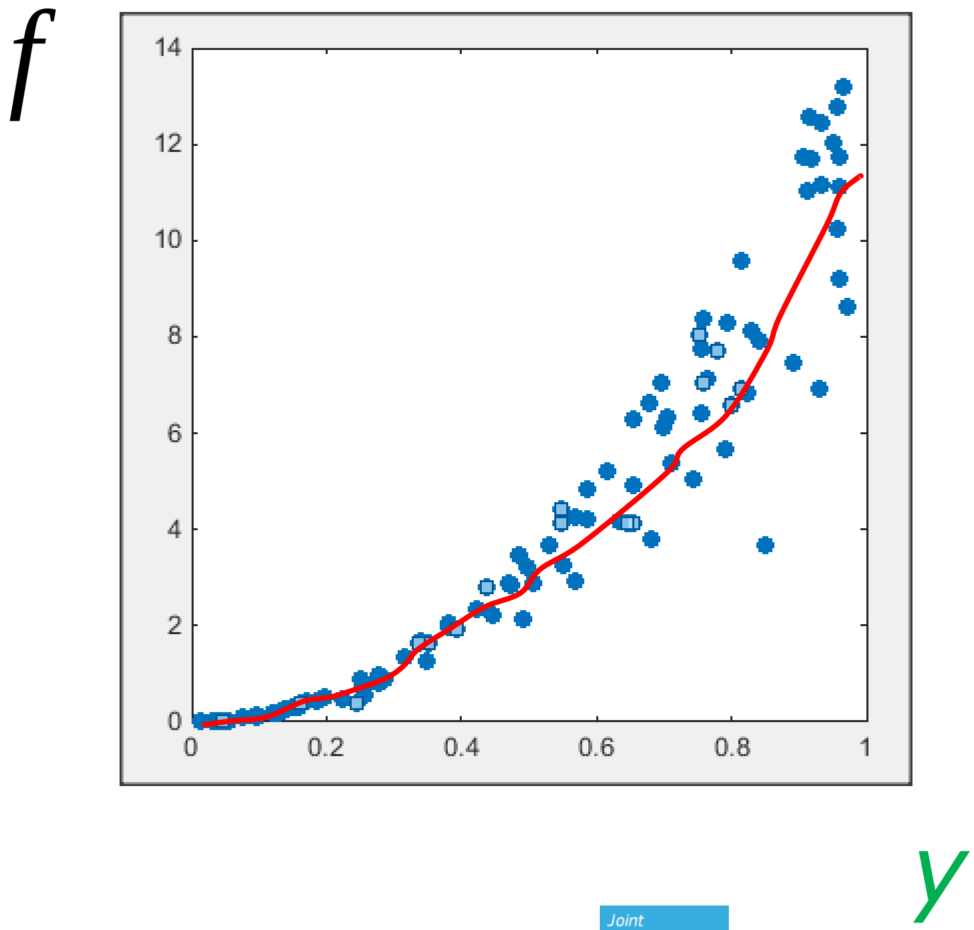
$$S_y + S_z + S_{yz} = 1$$

interactions

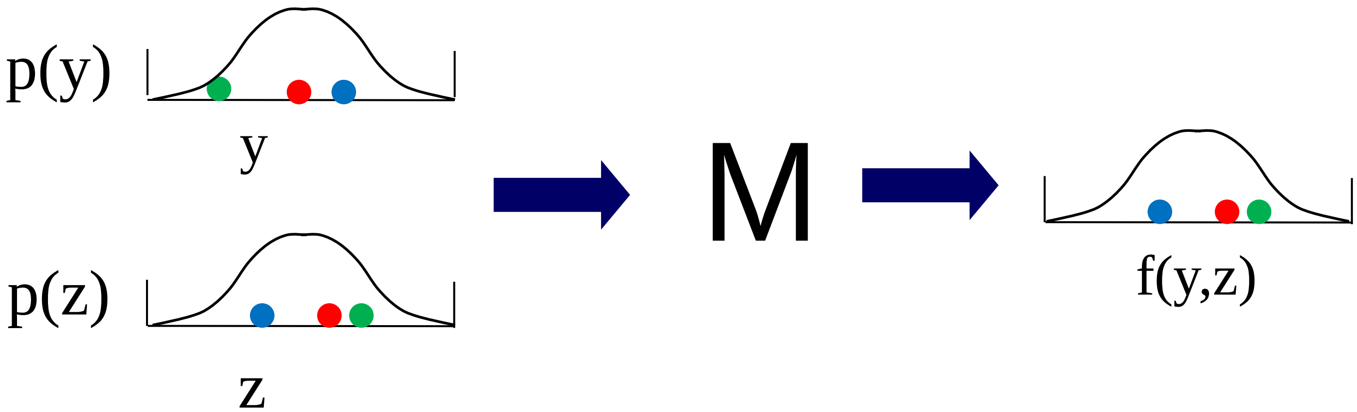


$$\sum_i S_i \leq 1$$

The effects of interactions



The theory



$$S_y = \frac{V[E(f \mid y)]}{V(f)} = \frac{1}{V(f)}\left[\int p(y)\,dy\left[\int f(y,z)\,p(z)\,dz\right]^2 - f_0^{\,2}\right]$$

For independent inputs

First-order indices

$$S_y = \frac{V[E(f \mid y)]}{V(f)}$$

The expected reduction of variance that is obtained when y is fixed over its uncertainty range.

Use: for R&D allocation, for parameter estimation

Total effects

$$T_y = \frac{E[V(f \mid z)]}{V(f)}$$

The expected variance that would be left when y is free to vary over its uncertainty range.

Use: for model simplification, to identify unessential inputs in the model, which are not important neither singularly nor in combination with others.

An input with a small value of its total effect sensitivity index can be frozen to any value within its range.

First and Total effects

$$S_y \leq T_y$$

$$T_y - S_y$$

It measures the interactions which y is involved in

$$S_y = T_y$$

It means that y is not involved in any interactions with other inputs

Thank you.
Any questions?