

Globalna Analiza Wrażliwości metodą Sobola

Zastosowanie do analiz RAMI



NATIONAL
CENTRE
FOR NUCLEAR
RESEARCH
ŚWIERK



Paweł Stano, Ph.D.

Department of Complex
Systems

Pawel.Stano@ncbj.gov.pl



Rozważany model

Model matematyczny jest przedstawiony wzorem:

$$Y = F_{\theta}(X, \omega)$$

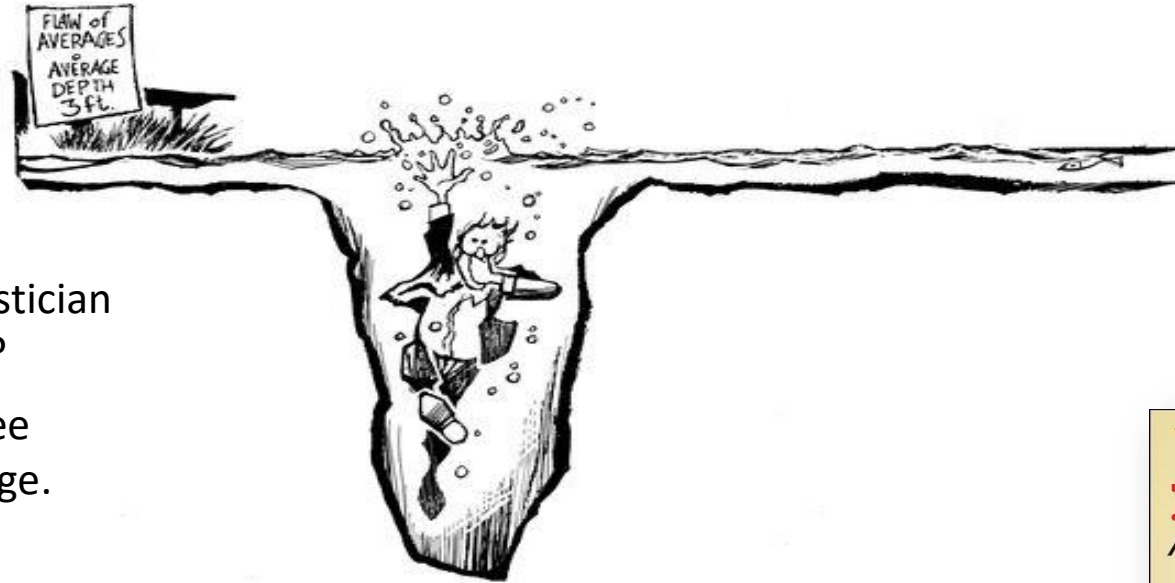
gdzie:

- Y – zmienne wyjściowe systemu,
- X – zmienne wejściowe systemu,
- ω – zewnętrzny szum środowiskowy,
- θ – strukturalna niepewność modelu,
- F – model systemu (analityczny, black-box, etc.)

Dla uproszczenia:

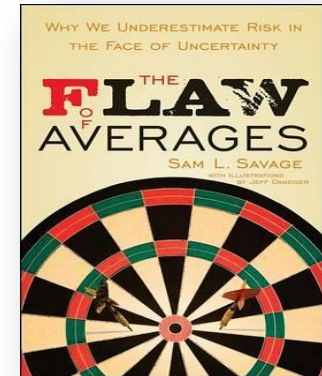
$$Y = F(X)$$

Czemu służą analizy niepewności i wrażliwości?

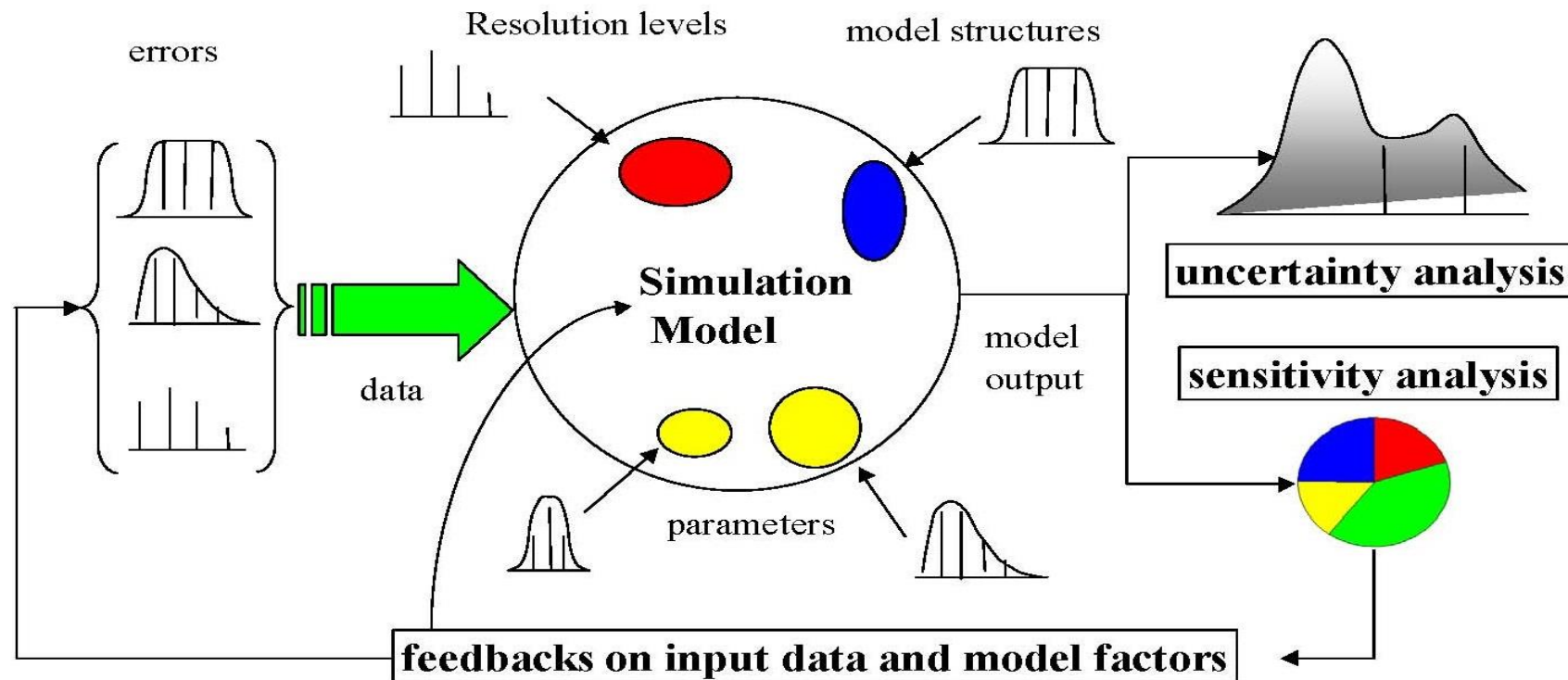


- Why did the statistician drown in the river?
- The river was three feet deep on average.

© Sam L. Savage, *The Flaw of Averages*



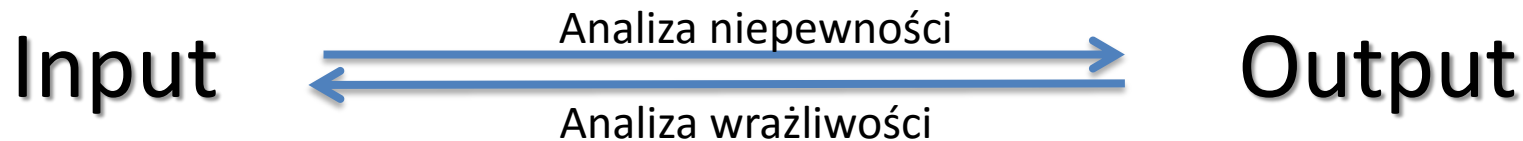
Struktura analizy



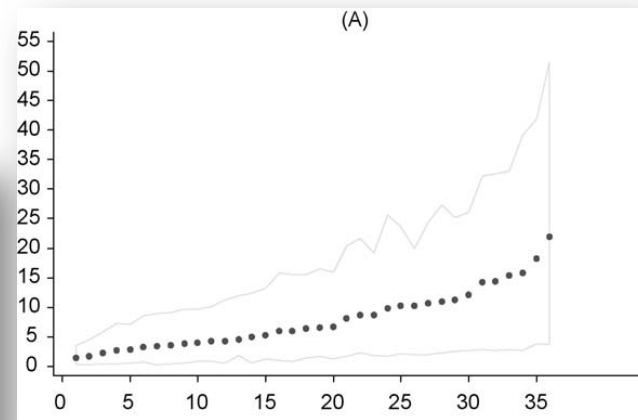
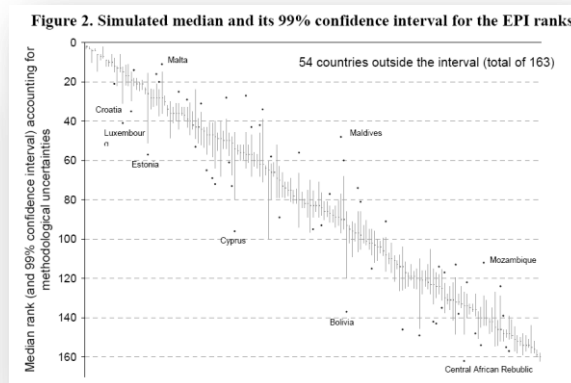
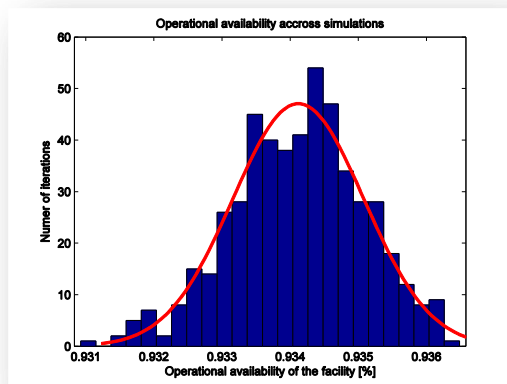
© Andrea Saltelli



Analiza niepewności vs analiza wrażliwości

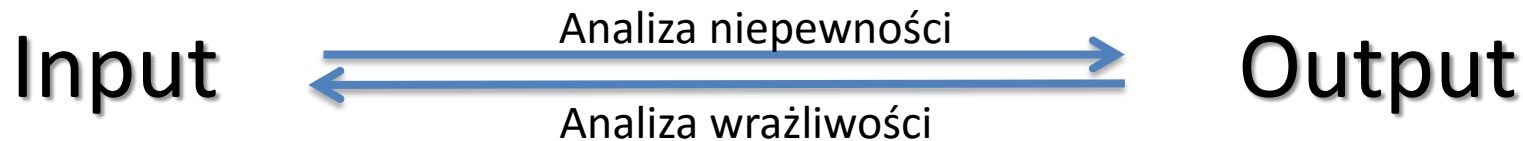


Typowe rezultaty analiz niepewności:

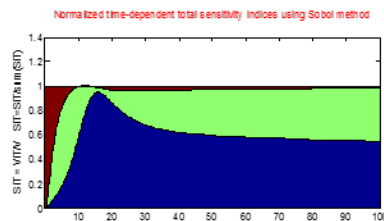
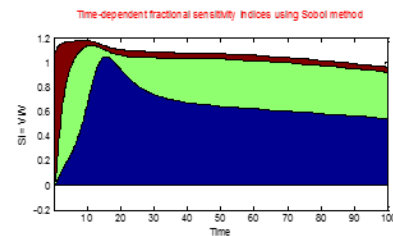
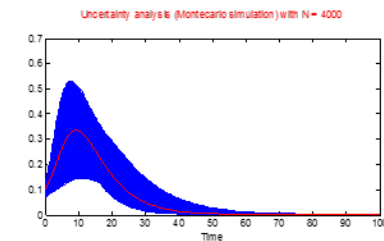
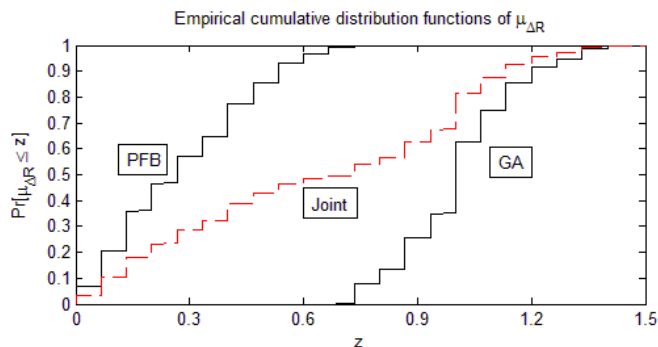




Analiza niepewności vs analiza wrażliwości



Typowe rezultaty analiz wrażliwości:



Pie chart of sensitivity indices for MSE function by Sobol method





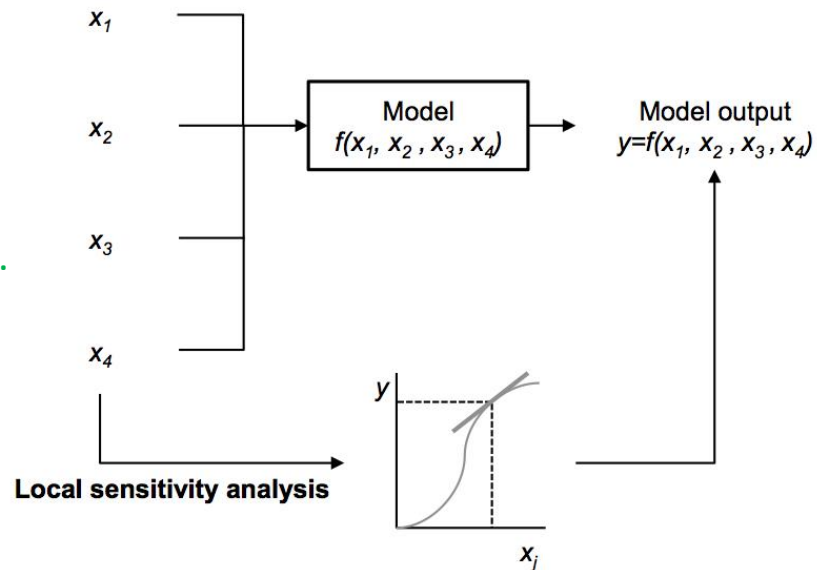
Najpopularniejsza (ciągła) metoda: One Factor At Time (OFAT)

Zalety:

- prostota,
- dobre rezultaty dla modeli liniowych,
- dobre rezultaty dla małych zbiorów danych.

Wady:

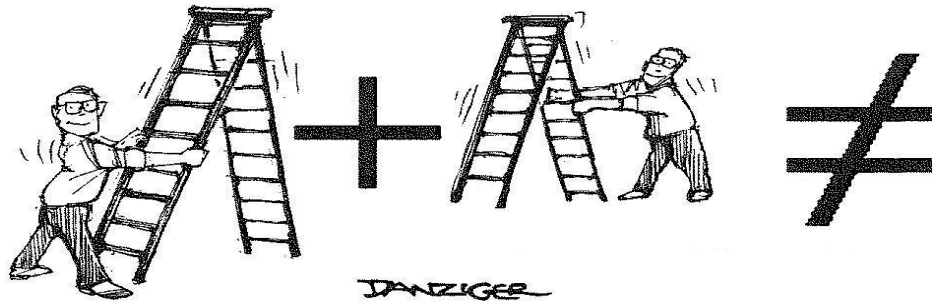
- duże błędy dla modeli nieliniowych,
- nie wykrywają zależności między skorelowanymi zmiennymi,
- nie eksplorują całej przestrzeni stanów



© Evelyne Groen, *An uncertain climate: the value of uncertainty and sensitivity analysis in environmental impact assessment of food*

Globalna analiza wrażliwości vs lokalna analiza wrażliwości

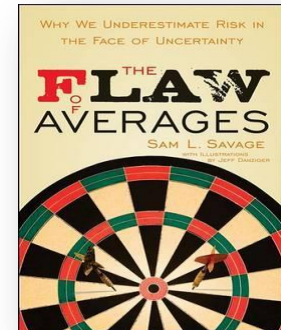
Lokalna analiza wrażliwości OFAT



Globalna Analiza Wrażliwości



© Sam L. Savage, *The Flaw of Averages*

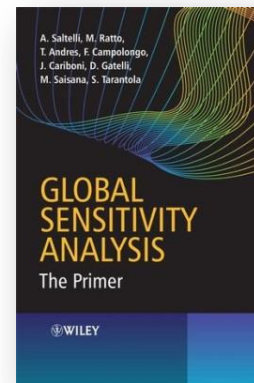




Dlaczego Globalna Analiza Wrażliwości?

- Analizowanie czynników pojedynczo (OFAT) nie pozwala na wykrycie wpływu nieliniowych interakcji pomiędzy skorelowanymi zmiennymi
- Statystyczna odporność (robustness) systemu nie może być zapewniona bez odpowiednio gęstej eksploracji wielowymiarowej przestrzeni wygenerowanej przez zmienne wejściowe
- Metody lokalne nie nadają się do badania niepewności epistemicznej

© Saltelli et al., *Global Sensitivity Analysis, The Primer*





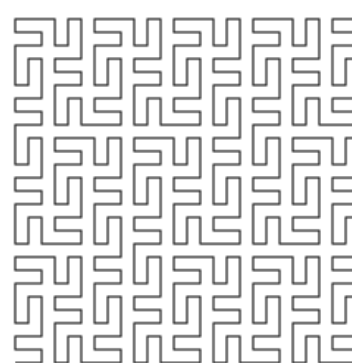
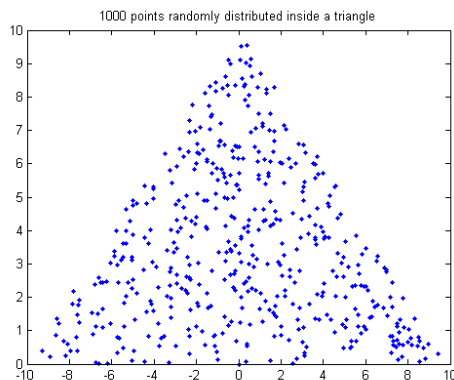
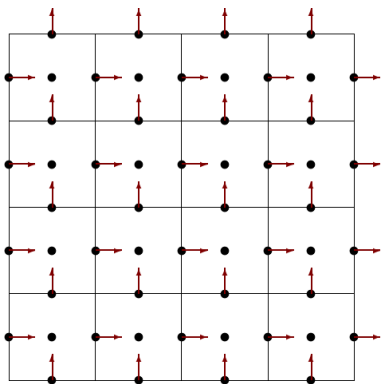
Curse of dimensionality

Czynniki generujące wysoką złożoność numeryczną GSA:

- Złożoność obliczeniowa modelu

$$Y = F(X)$$

- Gęstość i struktura siatki pokrywającej przestrzeń stanów wejścia





Metoda Sobola: dekompozycja wariancji

Ortogonalna dekompozycja
modelu:

$$Y = F(X_1, \dots, X_n) \\ = f_0 + \sum_{i=1}^n f_i(X_i) + \sum_{i < j}^n f_{ij}(X_i, X_j) + \dots + f_{1,2,\dots,n}(X_1, X_2, \dots, X_n)$$

gdzie:

$$f_0 = E(Y) \\ f_i(X_i) = E(Y | X_i) - f_0 \\ f_{ij}(X_i, X_j) = E(Y | X_i, X_j) - f_0 - f_i - f_j$$

tak że:

$$\int_0^1 f^2(X) dX - f_0^2 = \sum_{s=1}^n \sum_{i_1 < \dots < i_s}^n \int f^2_{i_1 \dots i_s} dX_{i_1} \dots dX_{i_s}$$



Metoda Sobola: dekompozycja wariancji

Ze wzoru:

$$\int_0^1 f^2(X) dX - f_0^2 = \sum_{s=1}^n \sum_{i_1 < \dots < i_s} \int f^2_{i_1 \dots i_s} dX_{i_1} \dots dX_{i_s}$$

dekompozycja wariancji:

$$Var(Y) = \sum_{i=1}^n V_i + \sum_{i < j} V_{ij} + \dots + V_{12 \dots n}$$

gdzie:

$$V_i = Var_{X_i} (E_{X \sim i} (Y | X_i))$$
$$V_{ij} = Var_{X_{ij}} (E_{X \sim ij} (Y | X_i, X_j)) - V_i - V_j$$



Metoda Sobola: indeksy Sobola

Indeksy pierwszego stopnia

$$S_i = \frac{V_i}{\text{Var}(Y)}$$

Indeksy wyższych stopni

$$S_{ijk} = \frac{V_{ijk}}{\text{Var}(Y)}$$

Dekompozycja wrażliwości
($2^n - 1$ obliczeń)

$$\sum_{i=1}^n S_i + \sum_{i < j}^n S_{ij} + \dots + S_{12\dots n} = 1$$

Efekty całkowite
(n obliczeń)

$$S_{Ti} = 1 - \frac{\text{Var}_{X_{\sim i}}(E_{X_i}(Y | X_{\sim i}))}{\text{Var}(Y)}$$

$$\sum_{i=1}^n S_{Ti} \geq 1$$

- Aproksymacja Sobola

$$V_i \approx \frac{1}{N} \sum_{k=1}^N F(x_{k,i}, x_{k,\sim i}) F(x_{k,i}, x'_{k,\sim i}) - \frac{1}{N} \left(\sum_{k=1}^N F(x_{k,i}, x_{k,\sim i}) \right)^2$$

- Aproksymacja Sobola-Kucherenko

$$V_i \approx \frac{1}{N} \sum_{k=1}^N F(x_{k,i}, x_{k,\sim i}) (F(x_{k,i}, x'_{k,\sim i}) - F(x'_{k,i}, x'_{k,\sim i}))$$

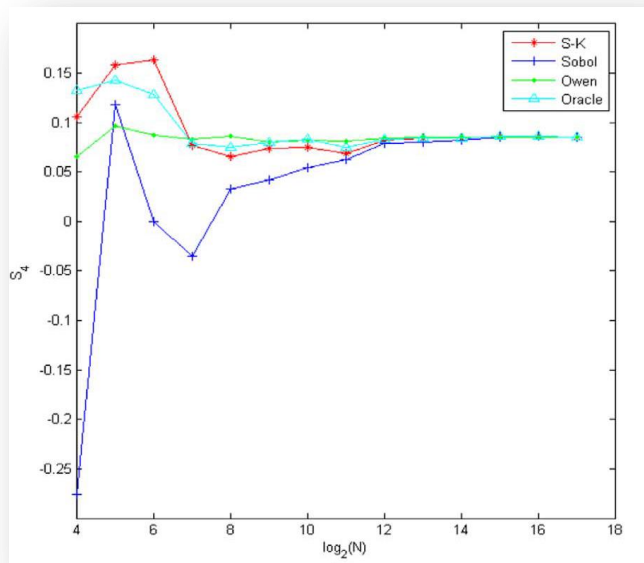
- Aproksymacja Owena

$$V_i \approx \frac{1}{N} \sum_{k=1}^N (F(x_{k,i}, x_{k,\sim i}) - F(x''_{k,i}, x_{k,\sim i})) (F(x_{k,i}, x'_{k,\sim i}) - F(x'_{k,i}, x'_{k,\sim i}))$$

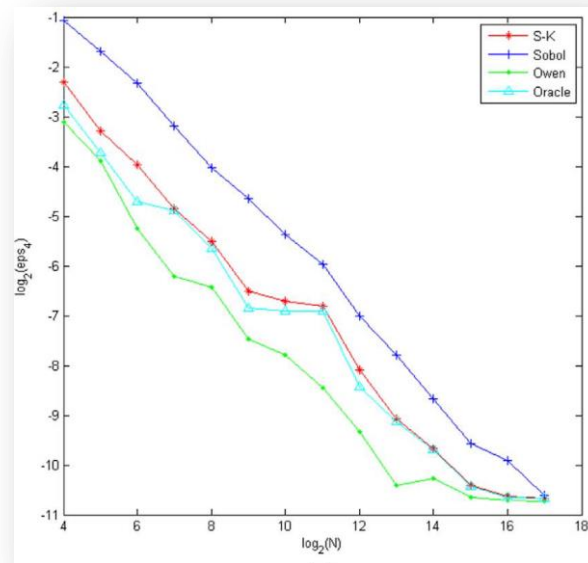
- Aproksymacja „Oracle”

$$V_i \approx \frac{1}{N} \sum_{k=1}^N (F(x_{k,i}, x_{k,\sim i}) - f_0) (F(x_{k,i}, x'_{k,\sim i}) - F(x'_{k,i}, x'_{k,\sim i}))$$

Zbieżność estymatora



RMSE estymatora



© Sergei Kucherenko, “Different numerical estimators for main effect global sensitivity indices”, Reliability Engineering & System Safety (2017)



Metoda Sobola: quasi-losowe ciągi Sobola

Motywacja: skonstruować ciąg

$$\{x_i\}_{i=1}^{\infty} \in I^s$$

tak by zachodziło

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n F(x_i) = \int_{I^s} F$$

i by zbieżność była jak najszybsza.

Przykładowy ciąg Sobola:

(0 0 0)	(0.5 0.5 0.5)	(0.75 0.25 0.25)
(0.25 0.75 0.75)	(0.375 0.375 0.625)	(0.875 0.875 0.125)
(0.625 0.125 0.875)	(0.125 0.625 0.375)	(0.1875 0.3125 0.9375)
(0.6875 0.8125 0.4375)		

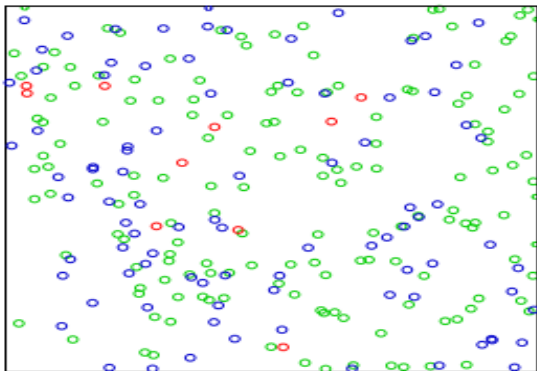


Metoda Sobola: miara rozbieżności

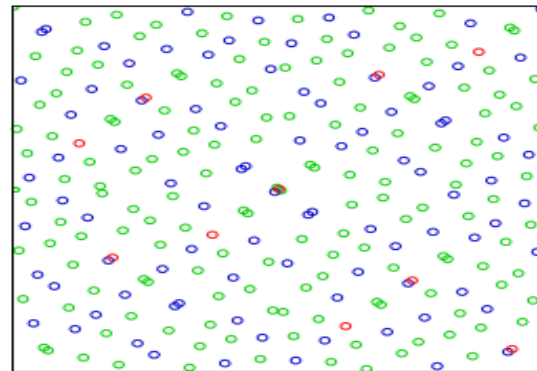
Miara rozbieżności (discrepancy measure):

$$h(t) = \frac{N_{Q(t)}}{N} - m(Q(t))$$
$$D^{L^2}_N = \left(\int_{I^s} h^2(t) dt \right)^{1/2}$$

Pseudo-losowe pokrycie metodą Monte Carlo

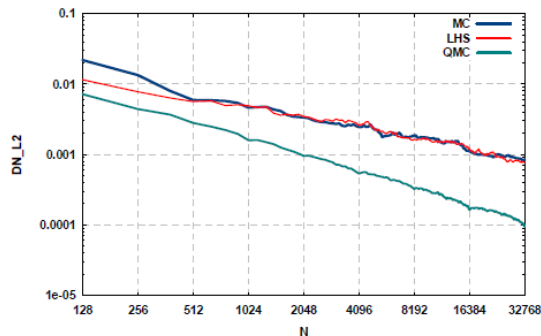


Quasi-losowe pokrycie ciągiem Sobola



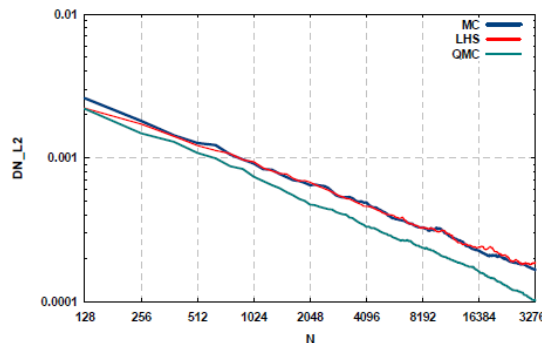
Metoda Sobola: porównanie pokrycia

$s = 5$



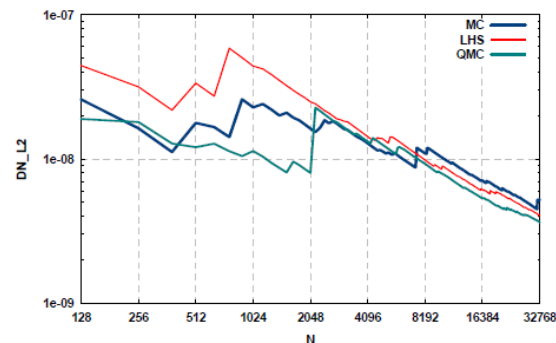
(a)

$s = 10$



(b)

$s = 40$



(c)

© Kucherenko, Albrecht, Saltelli, “Exploring multi-dimensional spaces: a Comparison of Latin Hypercube and Quasi Monte Carlo Sampling Techniques” (2015)



RAMI: Reliability, Availability, Maintainability and Inspectability

- **Reliability:** continuity of correct operation;
- **Availability:** readiness for correct operation;
- **Maintainability:** ability to undergo repairs and modifications;
- **Inspectability:** ability to undergo visits and controls.

Definicja za *ITER* (www.iter.org)



Zastosowanie GSA do analiz RAMI: IFMIF akcelerator

International Fusion Materials
Irradiation Facility (www.ifmif.org)

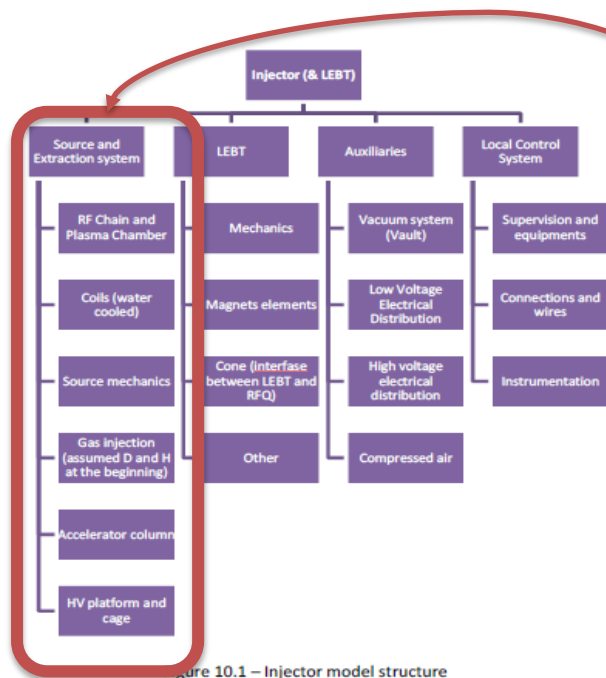
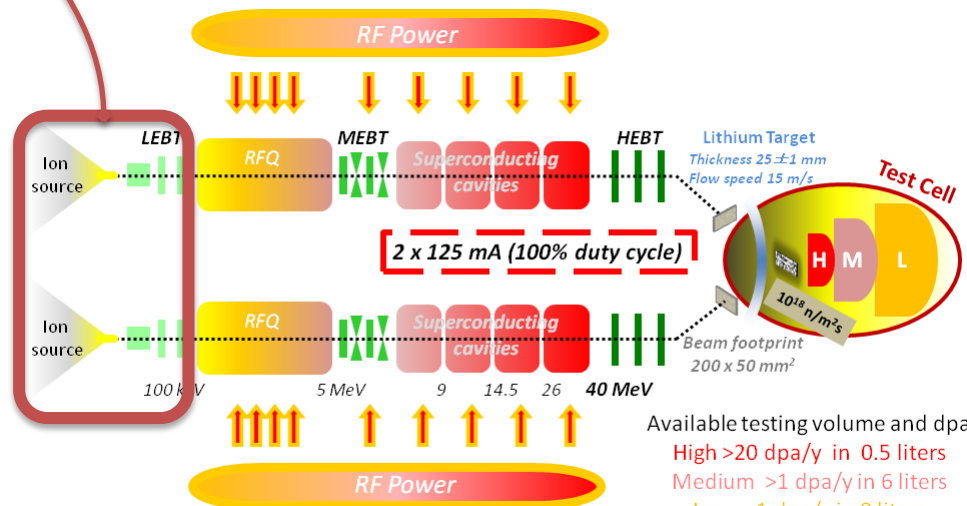


Figure 10.1 – Injector model structure



1. Operational Availability (OA):

$$OA = \frac{\text{up hours}}{\text{total hours}}$$

2. Hardware Availability (HA):

$$HA = \frac{\text{up hours}}{(\text{up hours}) + (\text{down hours})}$$

gdzie:

- 'up hours' – total time the facility is running,
- 'down hours' – total time the facility is shut due to system failures,
- 'maintenance hours' - total time the facility is shut down due to routine maintenance,
- 'total hours' - total life time of the facility, which includes 'up hours', 'down hours', and 'maintenance hours', so that:

$$\text{total hours} = (\text{up hours}) + (\text{down hours}) + (\text{maintenance hours})$$

Model matematyczny:

$$Y = F(X)$$

- Model F jest zadany przez program AvailSim symulujący dostępność akceleratora, który jest traktowany jak black-box,
- Zmienna wyjścia Y jest zadana przez wskaźnik HA,
- Zdarzenia początkowe są modelowane jako procesy odnawialne (renewal process) o rozkładzie wykładniczym z parametrami λ zdeterminowanymi przez częstotści awarii. Zmienne wejścia X odpowiadają niepewnościom wokół λ i mają rozkłady logarytmicznie normalne.



Zastosowanie GSA do analiz RAMI: źródła niepewności

- Główne źródła niepewności w układach wejścia pochodzą z obliczania współczynników **częstości awarii (Failure Rates - FR)**.
- Współczynniki FR są uzyskiwane w wyniku analizy statystycznej występowania awarii, więc charakteryzują się naturalną niepewnością.
- Współczynniki FR są modelowane jako zmienne losowe o rozkładach logarytmicznie normalnych o parametrach podanych w tabelach niezawodności.
- Przedziały niepewności wylicza się korzystając ze wzorów:

$$\text{uncertainty interval} = [\text{median}, 95\text{th percentile}]$$

$$\text{sigma} = \frac{\log(EF)}{1.65}$$

$$\text{median} = \text{mean} \times \exp\left(-\frac{(\text{sigma})^2}{2}\right)$$

$$(95\text{th percentile}) = EF \times \text{median}$$

Złożoność obliczeniowa problemu to:

$$n^d \times N_{sim} \times (D_{availsim})$$

gdzie:

- ' n ' definiuje gęstość siatki ,
- ' d ' ilość zdarzeń początkowych [ponad 12000 dla całego akceleratora],
- ' N_{sim} ' ilość symulacji probabilistycznych [od 80 do 1000],
- ' $D_{availsim}$ ' złożoność obliczeniowa pojedynczej symulacji na silniku Availsim [1-200 sekund dla analizowanego podsystemu na standardowej maszynie]

Złożoność obliczeniową można zredukować w następujący sposób:

1. Zmniejszyć ilość analizowanych zdarzeń początkowych z ' d ' do ' $d_{imp} \ll d$ ' poprzez zidentyfikowanie ' d_{imp} ' najważniejszych zdarzeń, tj. takich które charakteryzują się najwyższymi współczynnikami Fractional Contribution (FC):

$$FC(e) = \frac{\text{unavailability due to event 'e'}}{\text{total unavailability}}$$

2. Zmniejszenie gęstości siatki ' n ' przez pokrycie d_{imp} -wymiarowej przestrzeni ciągami niskiej rozbieżności (Sobol sequences, Latin-hypercube, etc.);
3. Paralellizację systemu.



Zastosowanie GSA do analiz RAMI: identyfikacja najważniejszych zdarzeń



13 zdarzeń, które mają
największy wkład w
awaryjność systemu

Event number	name	FC [%]	Event number	name	FC [%]
1	Magnetron	34,85%	54	Support	0,13%
32	Motor Valve (Controlled)	6,57%	15	Braze joints	0,12%
22	Motor Valve (Controlled)	6,44%	5	Cooling channels	0,06%
37	Extraction electrodes	6,29%	39	Cooling channels	0,05%
36	Extraction electrodes	6,18%	53	Support	0,05%
34	Extraction electrodes	5,93%	38	Cooling channels	0,05%
35	Extraction electrodes	5,70%	4	Cooling channels	0,05%
33	Extraction electrodes	5,47%	14	Braze joints	0,05%
3	RF window	5,05%	46	Platform	0,03%
2	Waveguides	3,66%	52	Insulators	0,03%
10	Step motor	2,85%	9	Screw	0,03%
44	Electrical feedthrough	2,60%	47	Insulators	0,03%
45	Thermocoupler (2/3)	1,89%	13	Braze joints	0,02%
40	Switch Flow	0,84%	51	Insulators	0,02%
41	Switch Flow	0,79%	50	Insulators	0,02%
6	Switch Flow	0,75%	31	Tank Deuterium	0,02%
43	Switch temperature	0,45%	49	Insulators	0,01%
42	Switch temperature	0,44%	16	Braze joints	0,01%
23	Motor Valve (Controlled)	0,34%	21	Platform	0,01%
8	Plunger	0,34%	48	Insulators	0,01%
7	Switch temperature	0,30%	24	Gas pipes 2m	0,00%
20	Hoses 1m	0,29%	25	Gas pipes 2m	0,00%
17	Hoses 1m	0,27%	26	Gas pipes 2m	0,00%
18	Hoses 1m	0,24%	27	Gas pipes 2m	0,00%
19	Hoses 1m	0,24%	28	Gas pipes 2m	0,00%
12	Coils	0,23%	29	Gas pipes 2m	0,00%
11	Coils	0,19%	30	Gas pipes 2m	0,00%



Zastosowanie GSA do analiz RAMI: wyniki

Event number	name	FC [%]	Sobol first-order index Si	Sobol total effect St
1	Magnetron	34,85%	0,338	0,376
32	Motor Valve (Controlled)	6,57%	0,030	0,095
22	Motor Valve (Controlled)	6,44%	0,028	0,106
37	Extraction electrodes	6,29%	0,029	0,078
36	Extraction electrodes	6,18%	0,010	0,102
34	Extraction electrodes	5,93%	0,009	0,069
35	Extraction electrodes	5,70%	0,030	0,086
33	Extraction electrodes	5,47%	0,015	0,067
3	RF window	5,05%	0,045	0,121
2	Waveguides	3,66%	0,025	0,129
10	Step motor	2,85%	0,013	0,040
44	Electrical feedthrough	2,60%	0,006	0,084
45	Thermocoupler (2/3)	1,89%	0,005	0,063
Sum of first-order effects Si			0.584	
Sum of second-order effects Sij			0.416	

5 zdarzeń o największych efektach całkowitych.

First-order effects plus second-order effects

Second-order approximation including interactions between the inputs

First-order approximation of the input impacts



- Modelowany system jest silnie nieliniowy przez co indeksy wrażliwości pierwszego stopnia są niewystarczającą miarą wrażliwości systemu,
- W przypadku wielu zmiennych ich wpływ na wrażliwość objawia się głównie poprzez interakcje z innymi zmiennymi,
- Ranking wrażliwości zmiennych wejścia nie jest tożsamy z rankingiem awaryjności (zadany przez współczynniki FC)

Thank you for attention



NATIONAL
CENTRE
FOR NUCLEAR
RESEARCH
ŚWIERK



Paweł Stano, Ph.D.

Department of Complex
Systems

Pawel.Stano@ncbj.gov.pl