European Commission

On performing screening analysis with the Innovative Algorithm

Ivano Azzini, Thierry Mara, Rossana Rosati

European Commission, DG.JRC I.2

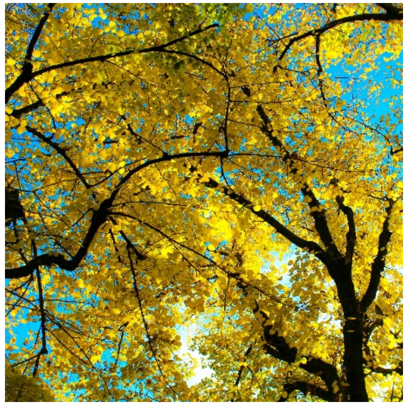
Competence Centre on Modelling

10th SAMO International Conference, 14-16 March 2022, Tallahassee Florida USA

Joint Research Centre

Outline

The IA-estimator



Illustration

Screening Methods

Sequential Bifurcation & IA-estimator

Screening

Let $y = f(\boldsymbol{x})$ be the scalar quantity of interest, function of D **independent** random variables $\boldsymbol{x} = (x_1,\ldots,x_D)$. Assume $f \in L_2$.

When: D is very large and number of model runs C limited

Scope: Identify the important variables with a minimum of model runs C

How: By relying on a qualitative sensitivity indicator

Qualitative sensitivity indicator means no ranking possible

Some Screening Methods

Design-driven

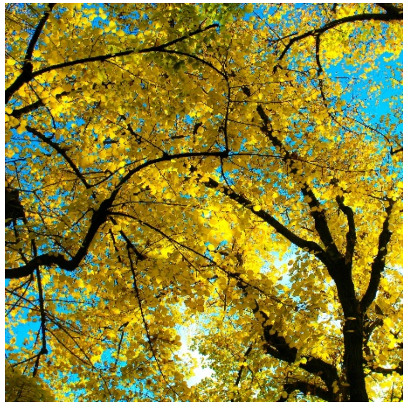
- Plackett-Burman design (1946, C ~ D, two levels, f additive and monotonic)
- Sequential bifurcation (1996, C<D, two levels, f monotonic, sign effect known)
- Morris Method (1991, C=rD+1, multi-level)

Data-driven

- Partitioning-based moment-dependent method (2006, C~hundreds, Monte Carlo, moment-dependent)
- Partitioning-based moment-independent methods (2013, 2015, C~thousands, Monte Carlo, moment-independent)
- Hilbert-Schmidt Information Criterion (2015, C~hundreds, Monte Carlo)

Outline

The IA-estimator



Illustration

Screening Methods

Sequential Bifurcation & IA-estimator

Sobol' Indices

Variance-based sensitivity indices = quantitative sensitivity indicators

- First-order Sobol' index $S_{x_i} = \frac{Var(E(y|x_i))}{Var(y)}$ ➡ Factor prioritization setting
- Total-order Sobol' index $ST_{x_i} = \frac{E(Var(y|\boldsymbol{x}_{\sim i}))}{Var(y)}$ ➡ Factor fixing setting

where $\boldsymbol{x}_{\sim i} = \boldsymbol{x}/x_i$ and with $1 \geq ST_{x_i} \geq S_{x_i} \geq 0$

$ST_{x_i} = S_{x_i}$ ➡ x_i does not interact with the other variables, i.e. with $\boldsymbol{x}_{\sim i}$

$ST_{x_i} \gg S_{x_i}$ ➡ x_i interacts strongly with $\boldsymbol{x}_{\sim i}$

The IA-estimator

The Innovative Algorithm was introduced by Azzini & Rosati (2021) and further studied in Azzini et al. (2021)

It relies on the following Monte Carlo estimators of first- and total-order Sobol' indices:

$$\hat{S}_{x_i}^{IA} = \frac{2\sum_{n=1}^{N}\left(y_n^{A} - y_n^{B_i}\right)\left(y_n^{A_i} - y_n^{B}\right)}{\sum_{n=1}^{N}\left[\left(y_n^{A} - y_n^{B}\right)^2 + \left(y_n^{A_i} - y_n^{B_i}\right)^2\right]}$$

$$\widehat{ST}_{x_i}^{IA} = \frac{\sum_{n=1}^{N}\left[\left(y_n^{A} - y_n^{B_i}\right)^2 + \left(y_n^{A_i} - y_n^{B}\right)^2\right]}{\sum_{n=1}^{N}\left[\left(y_n^{A} - y_n^{B}\right)^2 + \left(y_n^{A_i} - y_n^{B_i}\right)^2\right]}$$

where

$$y_n^A = f(\boldsymbol{x}_n^A) = f(x_{1,n}^A, x_{2,n}^A, \dots, x_{D,n}^A)$$

$$y_n^B = f(\boldsymbol{x}_n^B) = f(x_{1,n}^B, x_{2,n}^B, \dots, x_{D,n}^B)$$

$$y_n^{B_i} = f(\boldsymbol{x}_n^{B_i}) = f(x_{1,n}^A, \dots, x_{i-1,n}^A, x_{i,n}^B, x_{i+1,n}^A, \dots, x_{D,n}^A)$$

$$y_n^{A_i} = f(\boldsymbol{x}_n^{A_i}) = f(x_{1,n}^B, \dots, x_{i-1,n}^B, x_{i,n}^A, x_{i+1,n}^B, \dots, x_{D,n}^B)$$

Computational cost to estimate (ST_{x_i}, S_{x_i}) for i=1,…,D is C=2N(D+1)

The IA-estimator

The Innovative Algorithm was introduced by Azzini & Rosati (2021) and further studied in Azzini et al. (2021)

It relies on the following Monte Carlo estimators of first- and total-order Sobol' indices:

$$\hat{S}_{x_i}^{IA} = \frac{2\sum_{n=1}^{N}\left(y_n^{A} - y_n^{B_i}\right)\left(y_n^{A_i} - y_n^{B}\right)}{\sum_{n=1}^{N}\left[\left(y_n^{A} - y_n^{B}\right)^2 + \left(y_n^{A_i} - y_n^{B_i}\right)^2\right]}$$

$$\widehat{ST}_{x_i}^{IA} = \frac{\sum_{n=1}^{N}\left[\left(y_n^{A} - y_n^{B_i}\right)^2 + \left(y_n^{A_i} - y_n^{B}\right)^2\right]}{\sum_{n=1}^{N}\left[\left(y_n^{A} - y_n^{B}\right)^2 + \left(y_n^{A_i} - y_n^{B_i}\right)^2\right]}$$

The IA-estimator possesses the following desired properties:

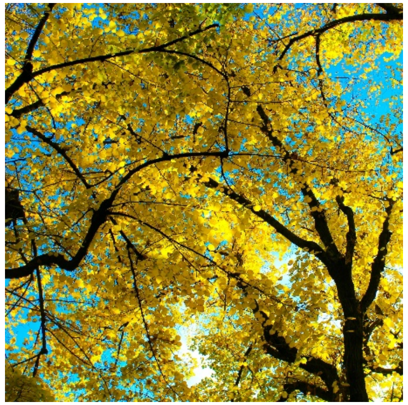
$$\hat{S}_{x_i}^{IA} + \widehat{ST}_{x_{\sim i}}^{IA} = 1$$

$$\widehat{ST}_{x_i}^{IA} \geq \hat{S}_{x_i}^{IA}$$

$\widehat{ST}_{x_i}^{IA} = \hat{S}_{x_i}^{IA}$ iif x_i does not interact with $\boldsymbol{x_{\sim i}}$

Outline

The IA-estimator



Illustration

Screening Methods

Sequential Bifurcation & IA-estimator

Sequential Bifurcation+IA-estimator

Let D=2^3 and consider the following scheme (for some sample of size N)

$$\boldsymbol{x}^{\boldsymbol{A}} = (x_1^A, x_2^A, x_3^A, x_4^A, x_5^A, x_6^A, x_7^A, x_8^A) \qquad \boldsymbol{x}^{\boldsymbol{B}} = (x_1^B, x_2^B, x_3^B, x_4^B, x_5^B, x_6^B, x_7^B, x_8^B)$$

$$\boldsymbol{x}^{\boldsymbol{A}_{1,\cdots,4}} = (x_1^A, x_2^A, x_3^A, x_4^A, x_5^B, x_6^B, x_7^B, x_8^B) \qquad \boldsymbol{x}^{\boldsymbol{B}_{1,\cdots,4}} = (x_1^B, x_2^B, x_3^B, x_4^B, x_5^A, x_6^A, x_7^A, x_8^A)$$

with this DOE we can infer: $\left(\hat{S}^{IA}_{x_1,x_2,x_3,x_4}, \widehat{ST}^{IA}_{x_1,x_2,x_3,x_4}, \hat{S}^{IA}_{x_5,x_6,x_7,x_8}, \widehat{ST}^{IA}_{x_5,x_6,x_7,x_8}\right)$

If one of the two subsets contains solely non-important variables, their Sobol' indices must be zero. Then, split in two each subset with **non-zero** Sobol' indices as follows:

Sequential Bifurcation+IA-estimator

Let D=2^3 and consider the following scheme (for some sample of size N)

$\boldsymbol{x}^A = (x_1^A, x_2^A, x_3^A, x_4^A, x_5^A, x_6^A, x_7^A, x_8^A)$

$\boldsymbol{x}^{A_{1,\cdots,4}} = (x_1^A, x_2^A, x_3^A, x_4^A, x_5^B, x_6^B, x_7^B, x_8^B)$

$\boldsymbol{x}^{A_{3,4}} = (x_1^B, x_2^B, x_3^A, x_4^A, x_5^B, x_6^B, x_7^B, x_8^B)$

$\boldsymbol{x}^{A_{1,\cdots,6}} = (x_1^A, x_2^A, x_3^A, x_4^A, x_5^A, x_6^A, x_7^B, x_8^B)$

$\boldsymbol{x}^B = (x_1^B, x_2^B, x_3^B, x_4^B, x_5^B, x_6^B, x_7^B, x_8^B)$

$\boldsymbol{x}^{B_{1,\cdots,4}} = (x_1^B, x_2^B, x_3^B, x_4^B, x_5^A, x_6^A, x_7^A, x_8^A)$

$\boldsymbol{x}^{B_{3,4}} = (x_1^A, x_2^A, x_3^B, x_4^B, x_5^A, x_6^A, x_7^A, x_8^A)$

$\boldsymbol{x}^{B_{1,\cdots,6}} = (x_1^B, x_2^B, x_3^B, x_4^B, x_5^B, x_6^B, x_7^A, x_8^A)$

$(\boldsymbol{x}^{A_{1,\cdots,4}}, \boldsymbol{x}^{B_{1,\cdots,4}}, \boldsymbol{x}^{A_{3,4}}, \boldsymbol{x}^{B_{3,4}}) \longrightarrow (\hat{S}^{IA}_{x_1,x_2}, \widehat{ST}^{IA}_{x_1,x_2})$

$(\boldsymbol{x}^{A}, \boldsymbol{x}^{B}, \boldsymbol{x}^{A_{3,4}}, \boldsymbol{x}^{B_{3,4}}) \longrightarrow (\hat{S}^{IA}_{x_3,x_4}, \widehat{ST}^{IA}_{x_3,x_4})$

$(\boldsymbol{x}^{A_{1,\cdots,4}}, \boldsymbol{x}^{B_{1,\cdots,4}}, \boldsymbol{x}^{A_{1,\ldots,6}}, \boldsymbol{x}^{B_{1,\ldots,6}}) \longrightarrow (\hat{S}^{IA}_{x_5,x_6}, \widehat{ST}^{IA}_{x_5,x_6})$

$(\boldsymbol{x}^{A}, \boldsymbol{x}^{B}, \boldsymbol{x}^{A_{1,\ldots,6}}, \boldsymbol{x}^{B_{1,\ldots,6}}) \longrightarrow (\hat{S}^{IA}_{x_7,x_8}, \widehat{ST}^{IA}_{x_7,x_8})$

Split again the groups with non-zero Sobol' indices

European Commission

Sequential Bifurcation+IA-estimator

Let D=2^3 and consider the following scheme (for some sample of size N)

$$\boldsymbol{x}^{\boldsymbol{A}} = (x_1^A, x_2^A, x_3^A, x_4^A, x_5^A, x_6^A, x_7^A, x_8^A) \qquad \boldsymbol{x}^{\boldsymbol{B}} = (x_1^B, x_2^B, x_3^B, x_4^B, x_5^B, x_6^B, x_7^B, x_8^B)$$

$$\boldsymbol{x}^{\boldsymbol{A}_{1,\cdots,4}} = (x_1^A, x_2^A, x_3^A, x_4^A, x_5^B, x_6^B, x_7^B, x_8^B) \qquad \boldsymbol{x}^{\boldsymbol{B}_{1,\cdots,4}} = (x_1^B, x_2^B, x_3^B, x_4^B, x_5^A, x_6^A, x_7^A, x_8^A)$$

$$\boldsymbol{x}^{\boldsymbol{A}_{3,4}} = (x_1^B, x_2^B, x_3^A, x_4^A, x_5^B, x_6^B, x_7^B, x_8^B) \qquad \boldsymbol{x}^{\boldsymbol{B}_{3,4}} = (x_1^A, x_2^A, x_3^B, x_4^B, x_5^A, x_6^A, x_7^A, x_8^A)$$

$$\boldsymbol{x}^{\boldsymbol{A}_{1,\cdots,6}} = (x_1^A, x_2^A, x_3^A, x_4^A, x_5^A, x_6^A, x_7^B, x_8^B) \qquad \boldsymbol{x}^{\boldsymbol{B}_{1,\cdots,6}} = (x_1^B, x_2^B, x_3^B, x_4^B, x_5^B, x_6^B, x_7^A, x_8^A)$$

$$\boldsymbol{x}^{\boldsymbol{A}_{1,3,4}} = (\boxed{x_1^A}, x_2^B, x_3^A, x_4^A, x_5^B, x_6^B, x_7^B, x_8^B) \quad \left(\hat{S}_{x_1}^{IA}, \widehat{ST}_{x_1}^{IA}\right) \quad \boldsymbol{x}^{\boldsymbol{B}_{1,3,4}} = (\boxed{x_1^B}, x_2^A, x_3^B, x_4^B, x_5^A, x_6^A, x_7^A, x_8^A)$$

$$\boldsymbol{x}^{\boldsymbol{A}_{4}} = (x_1^B, x_2^B, \boxed{x_3^B}, x_4^A, x_5^B, x_6^B, x_7^B, x_8^B) \quad \left(\hat{S}_{x_3}^{IA}, \widehat{ST}_{x_3}^{IA}\right) \quad \boldsymbol{x}^{\boldsymbol{B}_{4}} = (x_1^A, x_2^A, \boxed{x_3^A}, x_4^B, x_5^A, x_6^A, x_7^A, x_8^A)$$

$$\boldsymbol{x}^{\boldsymbol{A}_{1,\cdots 4,6}} = (x_1^A, x_2^A, x_3^A, x_4^A, \boxed{x_5^B}, x_6^A, x_7^B, x_8^B) \quad \left(\hat{S}_{x_5}^{IA}, \widehat{ST}_{x_5}^{IA}\right) \quad \boldsymbol{x}^{\boldsymbol{B}_{1,\cdots 4,6}} = (x_1^B, x_2^B, x_3^B, x_4^B, \boxed{x_5^A}, x_6^B, x_7^A, x_8^A)$$

$$\boldsymbol{x}^{\boldsymbol{A}_{1,\cdots,7}} = (x_1^A, x_2^A, x_3^A, x_4^A, x_5^A, x_6^A, \boxed{x_7^A}, x_8^B) \quad \left(\hat{S}_{x_7}^{IA}, \widehat{ST}_{x_7}^{IA}\right) \quad \boldsymbol{x}^{\boldsymbol{B}_{1,\cdots,7}} = (x_1^B, x_2^B, x_3^B, x_4^B, x_5^B, x_6^B, \boxed{x_7^B}, x_8^A)$$

Sequential Bifurcation+IA-estimator

Let $D=2^3$ and consider the following scheme (for some sample of size N)

$$\boldsymbol{x}^{A} = (x_1^A, x_2^A, x_3^A, x_4^A, x_5^A, x_6^A, x_7^A, x_8^A) \qquad \boldsymbol{x}^{B} = (x_1^B, x_2^B, x_3^B, x_4^B, x_5^B, x_6^B, x_7^B, x_8^B)$$

$$\boldsymbol{x}^{A_{1,\cdots,4}} = (x_1^A, x_2^A, x_3^A, x_4^A, x_5^B, x_6^B, x_7^B, x_8^B) \qquad \boldsymbol{x}^{B_{1,\cdots,4}} = (x_1^B, x_2^B, x_3^B, x_4^B, x_5^A, x_6^A, x_7^A, x_8^A)$$

$$\boldsymbol{x}^{A_{3,4}} = (x_1^B, x_2^B, x_3^A, x_4^A, x_5^B, x_6^B, x_7^B, x_8^B) \qquad \boldsymbol{x}^{B_{3,4}} = (x_1^A, x_2^A, x_3^B, x_4^B, x_5^A, x_6^A, x_7^A, x_8^A)$$

$$\boldsymbol{x}^{A_{1,\cdots,6}} = (x_1^A, x_2^A, x_3^A, x_4^A, x_5^A, x_6^A, x_7^B, x_8^B) \qquad \boldsymbol{x}^{B_{1,\cdots,6}} = (x_1^B, x_2^B, x_3^B, x_4^B, x_5^B, x_6^B, x_7^A, x_8^A)$$

$$\boldsymbol{x}^{A_{1,3,4}} = (x_1^A, x_2^B, x_3^A, x_4^A, x_5^B, x_6^B, x_7^B, x_8^B) \quad \left(\hat{S}_{x_2}^{IA}, \widehat{ST}_{x_2}^{IA}\right) \quad \boldsymbol{x}^{B_{1,3,4}} = (x_1^B, x_2^A, x_3^B, x_4^B, x_5^A, x_6^A, x_7^A, x_8^A)$$

$$\boldsymbol{x}^{A_{4}} = (x_1^B, x_2^B, x_3^B, x_4^A, x_5^B, x_6^B, x_7^B, x_8^B) \quad \left(\hat{S}_{x_4}^{IA}, \widehat{ST}_{x_4}^{IA}\right) \quad \boldsymbol{x}^{B_{4}} = (x_1^A, x_2^A, x_3^A, x_4^B, x_5^A, x_6^A, x_7^A, x_8^A)$$

$$\boldsymbol{x}^{A_{1,\cdots4,6}} = (x_1^A, x_2^A, x_3^A, x_4^A, x_5^B, x_6^A, x_7^B, x_8^B) \quad \left(\hat{S}_{x_6}^{IA}, \widehat{ST}_{x_6}^{IA}\right) \quad \boldsymbol{x}^{B_{1,\cdots4,6}} = (x_1^B, x_2^B, x_3^B, x_4^B, x_5^A, x_6^B, x_7^A, x_8^A)$$

$$\boldsymbol{x}^{A_{1,\cdots,7}} = (x_1^A, x_2^A, x_3^A, x_4^A, x_5^A, x_6^A, x_7^A, x_8^B) \quad \left(\hat{S}_{x_8}^{IA}, \widehat{ST}_{x_8}^{IA}\right) \quad \boldsymbol{x}^{B_{1,\cdots,7}} = (x_1^B, x_2^B, x_3^B, x_4^B, x_5^B, x_6^B, x_7^B, x_8^A)$$

Computational cost (worst case): $C=2ND<2N(D+1)$

European Commission

Sequential Bifurcation+IA-estimator

Assume K out of D important variables

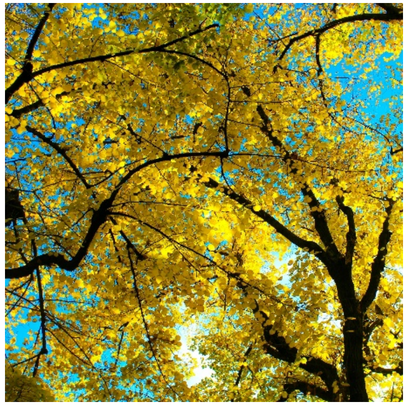
The number of steps of SBIA ranges from: $Ns=[(K+\log_2(D/K)), D]$

So, the computational cost of SBIA ranges from: $C\epsilon[2N(K+\log_2(D/K)), 2ND]$

For screening purposes, N can be chosen small, say: $N\epsilon[8,12]$

Outline

The IA-estimator



Illustration

Screening Methods

Sequential Bifurcation & IA-estimator

Example: The Morris function

Introduced by Morris (1991) to test screening methods, D=20, K=10

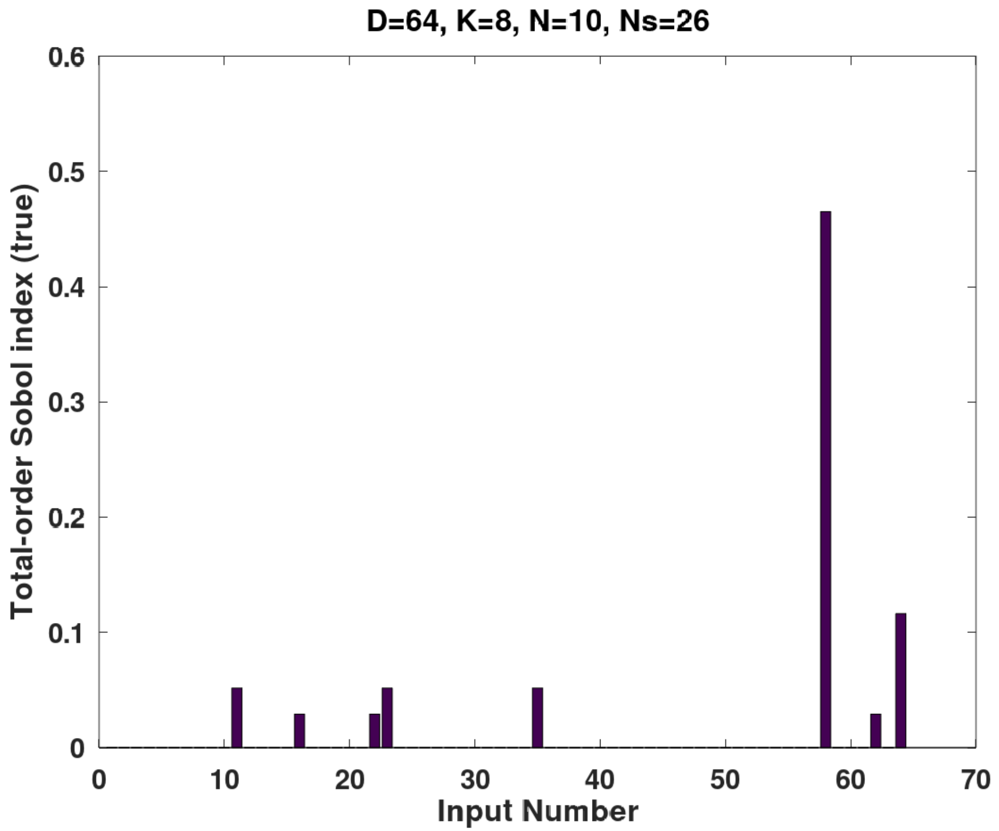
- $y = \beta_0 + \sum_{i=1}^{20} \beta_i x_i + \sum_{i<j}^{20} \beta_{ij} x_i x_j + \sum_{i<j<k}^{20} \beta_{ijk} x_i x_j\, x_k + \sum_{i<j<k<l}^{20} \beta_{ijkl} x_i x_j x_k x_l$
- with $\begin{cases} \beta_i = 20 & i = 1, \dots, 10 \\ \beta_{ij} = -15 & i,j = 1, \dots, 6 \\ \beta_{ijk} = -10 & i,j,k = 1, \dots, 5 \\ \beta_{ijkl} = 5 & i,j,k,l = 1, \dots, 4 \end{cases}$
- zero otherwise

The input set is ordered such that $(x_1, \dots, x_{10})$ are the important variables, then it would require C=2N(K+$\log_2$(D/K))=2N(10+1) function evaluations

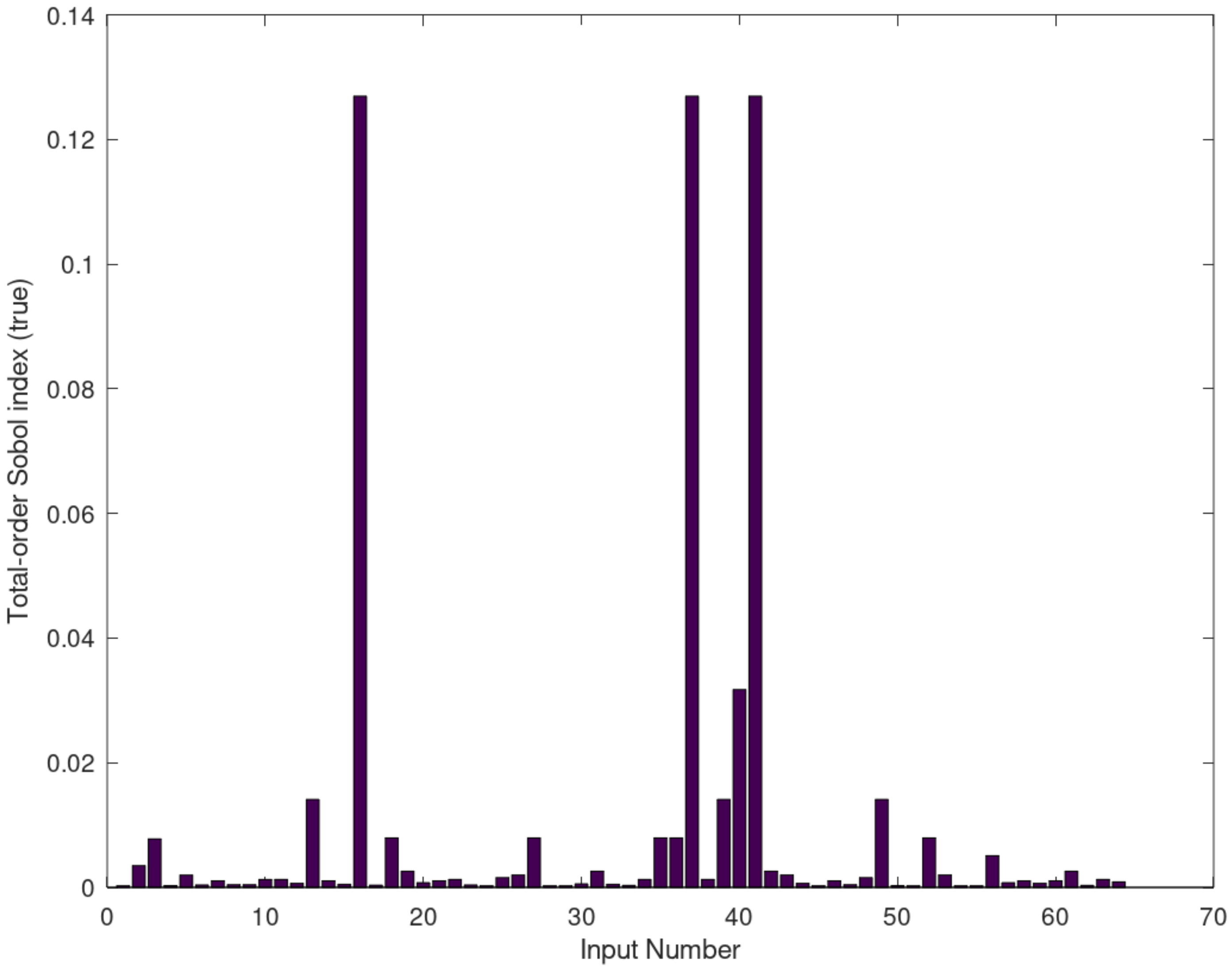
If the input set is ordered such that $(x_2, x_4, \dots, x_{18}, x_{20})$ are the important variables, then C=2ND=40N

Example: The g-function

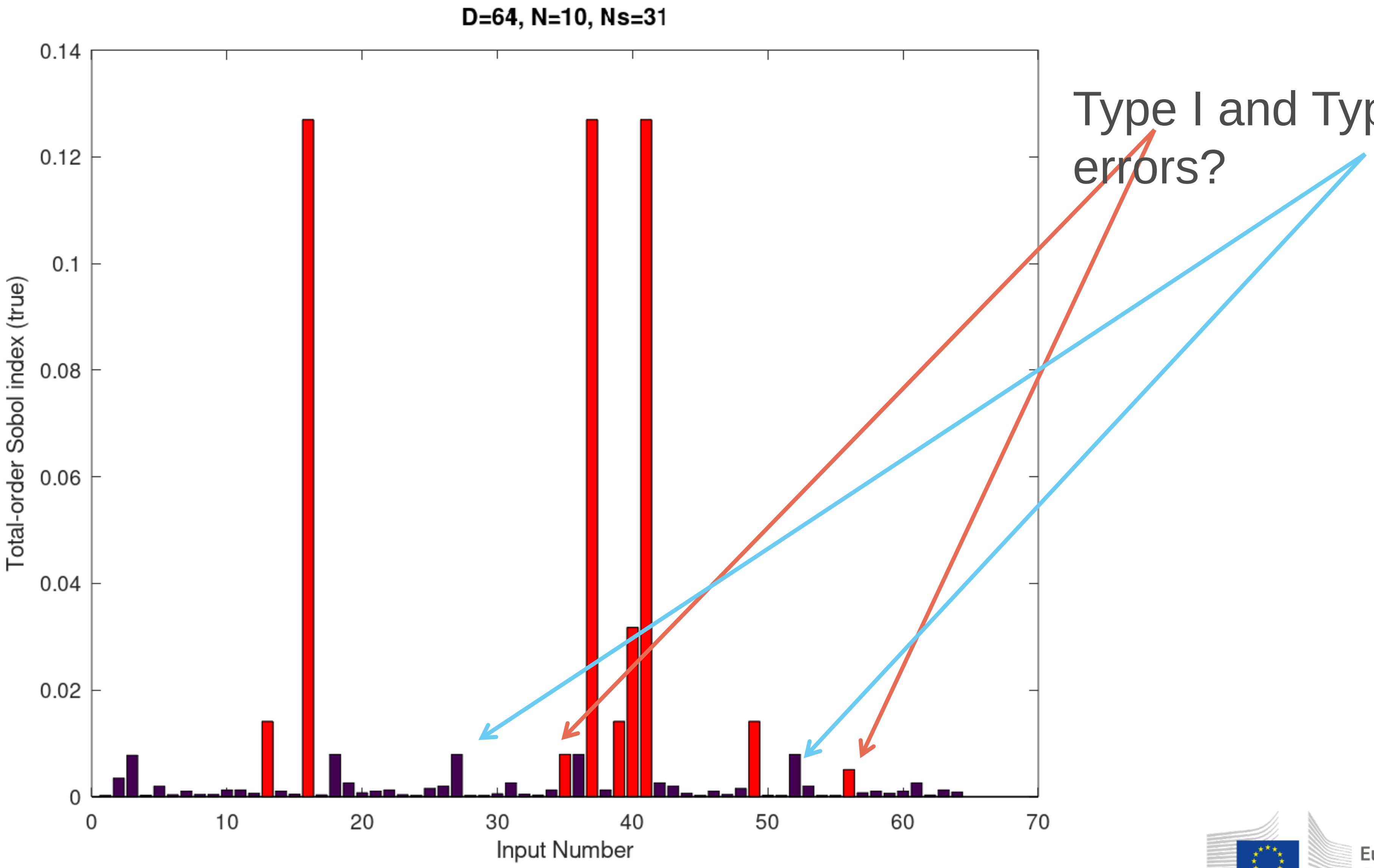
Numerical experiments show that to reduce Type I (erroneously concluding that x_i is important) and Type II (erroneously concluding that x_i is not-important) errors, better set N ϵ [8,12]



Example: The g-function



Example: The g-function



Type I and Type II errors?

Conclusion

MC estimators of Sobol’ indices (IA, Sobol’-Saltelli, Sobol’Jansen, Monod-Janon,…) are computationally demanding

IA-estimator = best MC estimator of Sobol’ indices (personal *biased* opinion)

Classical pick-freeze approaches waste model runs to estimate accurately Sobol’ indices of non-important variables

SBIA allows for significant reduction of the computational burden

SBIA can be used for screening purposes by choosing N small,…

but is a little tricky to implement

References

Morris, M.D., **Technometrics**, Vol. 33, pp. 161-174, 1991

Bettonvil, M. & Kleijnen, J.P.C., **EJOR**, Vol. 96, pp. 180-194,1996

Wu, Y.T. & Mohanty, S., **RESS**, Vol. 91, pp. 634-6472006)

Plischke, E., Borgonovo, E., Smith, S.L., **EJOR**, Vol. 226, pp. 536-550, 2013

Pianosi, F., Wagener, T., **EMS**, Vol. 67, pp. 1-11, 2015)

Azzini, I. & Rosati, R., **RESS**, Vol. 213, pp. 107647, 2021

Azzini, I., Mara, T., Rosati, R., **EMS**, Vol. 144, pp. 105167, 2021

Da Veiga, S., **JSCS**, Vol. 85, pp. 1283-1305, 2015

Thank you for your attention

Visit our Public Webpage: https://ec.europa.eu/knowledge4policy/modelling/topic/sensitivity-analysis-models_en

Our functional mailbox: ec-sensitivity-analysis@ec.europa.eu

JRC is recruiting: https://recruitment.jrc.ec.europa.eu
3(+3) years CA position on SAMO for policy

Sim@b: https://web.jrc.ec.europa.eu/rapps/pub/simlab/
An online tool to perform SAMO