Task 1.2- Mining data for behavioral modelling.
Progresses Report January 14, 2023

From the project

Task 1.2: Mining data for behavioral modelling (M1-12, Lead: CNR, Other: UNINA, WMG, THAB). Causal relationships between external and human factors, and safety-critical driver behaviors from NDS and DSE data are sought at the level of specific driving situations. Data mining effort seeks to find (combinations of) features in a way that goes beyond first order correlations. Unsupervised learning tools are used and implemented to investigate about the data sets. Clustering techniques and (generalized) linear mixed effect models are used. Feature selection via regularized regression tools (e.g., LASSO, elastic net), Random Forests, subset selections, and global sensitivity analysis are used. Results are invaluable for both WP2 and WP4, and are part of the systematic sensitivity auditing approach we take in this project. This task supports Task 1.4 in terms of quantitative evidence (a) for which hypotheses may be a priori meaningful to test (and which not); and (b) for prioritizing specific modelling requirements.

Ongoing work

Based on existing experience at the consortium on data mining in the context of accident analysis (Montella et al. 2012) and on global sensitivity analysis (Saltelli et al. 2008; 2021; Puy et al. 2022) a combination of data mining techniques in being geared up for the analysis […data source… naturalistic driving data]. These include clustering techniques and (generalized) linear mixed effect models, feature selection via regularized regression tools (e.g., LASSO, elastic net), Random Forests, and subset selections. Additionally global sensitivity analysis methods are used for selecting features, following recent developments on the use of SA of and for data mining (Tunkiel, Sui, and Wiktorski 2020; Antoniadis, Lambert-Lacroix, and Poggi 2021), including using the concept of mean dimension (Hoyt and Owen 2021). Note also section 3.4 of a recent position paper on sensitivity analysis (Razavi et al. 2021). Interesting as a possible linkage sensitivity analysis - sensitivity auditing also this paper (Bénesse et al. 2021).

Specifically i4Driving will test on machine learning the total sensitivity indices (Homma and Saltelli 1996), that have already found use in an adjacent field – model selection in regression (Becker, Paruolo, and Saltelli 2021).

To be noted also the use of sensitivity analysis as a contribution to model interpretability (Iooss, Kenett, and Secchi 2022)

Additional avenues for research are the use of pre-integration techniques for SA (Liu and Owen 2022) based on Paul Constantine's active subspace decomposition (Constantine, Dow, and Wang 2014). We shall also test new measures of sensitivity analysis based on the concept of discrepancy (Puy and Saltelli 2022) for the machine learning investigation of i4Driving.

Works acknowledging i4Driving

The project is acknowledged in a published work (Di Fiore et al. 2022), in a preprint (Saltelli and Puy 2022) being revised for Humanities and Social Sciences Communication, and in a submitted work (Saltelli and Puy 2023).

Others

CNR is still in the process of recruiting two additional researchers to work on the project. According to the present procedures, these resources will join us in February 2023.

Sensitivity analysis: a short introduction

Summary

Sensitivity analysis (SA) studies how much the uncertainty of a model output depends upon its inputs. Though it is generally agreed in existing guidelines that uncertainty and sensitivity analyses are both crucial for the validation or verification of a model, their application is hampered by practical difficulties, scarce awareness, and at times reluctance to expose the weakness of a model.

We present here **global** sensitivity analysis, mainly through one class of global SA methods known as 'variance-based' – considered by most practitioners as a recommended practice - and offer pointers on additional methods. We also suggest several hints for a successful and effective use these of these techniques.

The topic

Consider the following definitions:

- Uncertainty analysis (UA): The quantification of the uncertainty in model output
- Sensitivity analysis: The study of the relative importance of different input factors on the model output uncertainty

As we shall discuss in this entry, the two analyses are linked.

To a natural scientist trained in calculus, sensitivity analysis may evoke the derivative of a function of interest with respect to its inputs. So if the function or model has the form $y = f(x_1, x_2, \dots\ x_k)$, where y is the scalar output and $x_1, x_2, \dots\ x_k$ are input factors, the sensitivity of x_i is simply $\partial y / \partial x_i$. Economists likewise use elasticities such as $\frac{x_{i0}}{y_0} \partial y / \partial x_i$ where x_{i0} and y_0 are averages. In most ecological studies, the factors may vary considerably, from a few percent to orders of magnitude, and likewise will do the output, because of error propagation. Hence, to an ecologist, what happens to y in a single point of the multidimensional space of existence of $x_1, x_2, \dots\ x_k$ may be uninformative; ecologists will want sensitivity measures that are global, i.e. concerned with the whole space of variability of the inputs.

When the overall uncertainty in y is modest, it is not so important to ascertain where this is coming from. Conversely, if y spans orders of magnitude, then SA becomes indispensable to understand the system studied and pinpoint the factor(s) that convey the most uncertainty. Such information might help guiding further research by highlighting where efforts on data collection should focus to maximize the reduction of uncertainty in the output.

Variance-based sensitivity analysis

The starting point for UA is the analytic or computer-coded form of $f(x_1, x_2, \dots\ x_k)$, the model, and the probability distributions of the inputs, $p_1(x_1), p_2(x_2), \dots p_k(x_k)$. Although determining these probability distributions is preliminary to any analysis, it is often the most important and expensive part of the work. This stage of elicitation may involve experts from several disciplines and/or the collection of a considerable amount of data. Monte Carlo based UA consists in a series of simulations. In each of these simulations the value of each input factor is sampled from its distribution. The corresponding output value produced is recorded and the statistical properties of the output distribution are finally analysed.

We now concisely describe the SA variance-based methods. A handbook is available for a more detailed treatment of the topic (Saltelli et al. 2008). These measures are mostly due to the work of Russian mathematician Ilya M Sobol' (Sobol' 1993)

We take the variance of the output as the target of the analysis, and following statistical theory, we decompose it following the ANOVA (ANalysis Of VAriance) scheme:

$$(1) \qquad V(y) = \sum_{i_1=1}^{k} V_{i_1} + \sum_{i_1 i_2, i_1<i_2}^{k} V_{ij} + \cdots + V_{i_1 i_2 \ldots i_k, i_1<i_2<\cdots<i_k}$$

In plain English this says that the unconditional variance of y is decomposed in terms relative to an individual factor, plus terms (named interactions) pertaining to two factors, three factors, and so on until the last single term due to interactions among all factors. Hence, we have k first order terms, $\binom{k}{2} = \frac{k(k-1)}{2}$ second-order terms (corresponding to all the possible combinations of interactions between two factors), $\binom{k}{3} = \frac{k(k-1)(k-2)}{6}$ third-order terms and so on until a single term of order k. A model with $k = 3$ factors has three first order terms, three second-order ones and one third-order one. The total number of terms is $2^k - 1 = 7$ for this model. Dividing Equation 1 by $V(y)$ and swapping the two terms of equation (1) one obtains:

$$(2) \qquad \sum_{i_1=1}^{k} S_{i_1} + \sum_{i_1 i_2, i_1<i_2}^{k} S_{ij} + \cdots + S_{i_1 i_2 \ldots i_k, i_1<i_2<\cdots<i_k} = 1$$

The S's are the desired sensitivity coefficients, the objective of the estimation procedures of variance based SA methods. Without proof, we add here a few notes:

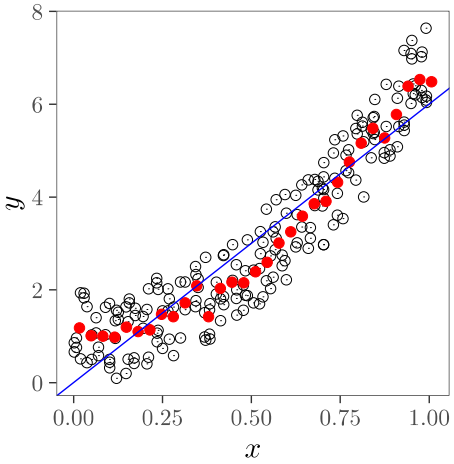
- The first order term S_i is identical to the Pearson correlation ratio η_i^2:

$$(3) \qquad S_i = \eta_i^2 = \left(\frac{V\big(E(y|x_i)\big)}{V(y)} \right)$$

In this expression the mean variance is taken over all factors but x_i, which is kept fixed, while the outer variance is taken over all possible values of x_i, e.g. in a more verbose notation

$$(3a) \qquad S_i = \eta_i^2 = \left(\frac{V_{x_i}\left(E_{x_{\sim i}}(y|x_i = x_i^*)\right)}{V(y)} \right)$$

In other words, Equation (3) computes the mean of a moving average, see Figure 1.



[Figure 1 here]: Scatterplots with moving averages (red). The straight line is the standardised regression coefficient of y on x_i, the discontinuous line is the moving average $E_{x_{\sim i}}(y|x_i)$.

- While one generally estimates all the first order terms, the higher order terms are not often computed. For a system with ten factors, there would be 45 second order terms and 1,023 in total, simply too many to look at. Practitioners compute a total sensitivity index T_i which is written as

$$(4) \qquad T_i = \left(\frac{E\left(V(y|x_{\sim i})\right)}{V(y)}\right)$$

T_i can be computed with a single estimate, without knowing all terms Equation (2). In Equation (4) the inner mean is taken over x_i while keeping all other factors $x_{\sim i}$ fixed, and the outer mean is taken over $x_{\sim i}$. T_i is the sum of all terms – from the first to the k^{th}, that include factor x_i. While the sum of the S_i is always less or equal to one, that of the T_i can be larger than one. For a model with three factors $T_1 = S_1 + S_{12} + S_{13} + S_{123}$ with similar formulae for T_2 and T_3.

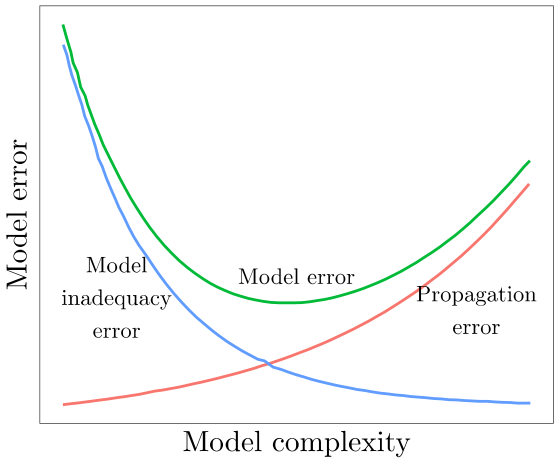
- While Equations (3) and (4) seem to suggest a complex Monte Carlo estimation based e.g. on nested loops, Monte Carlo estimation of both indices is straightforward and based on a single Monte Carlo loop. These formulae can be found in Saltelli *et al.* (2008), and a discussion is in Saltelli *et al.* (2010).
- For most applications in ecology, it is sufficient to compute the S_i and the T_i.
  If all S_i's are equal to the corresponding T_i's the model is said to be additive, i.e. without interactions.
- This information is considerably superior to that offered by the derivatives $\partial y / \partial x_i$ because it captures interactions among factors.
- One see often in the literature SA done via incremental rations such as $\frac{\Delta y}{\Delta x_i}$ where the increment Δx_i are taken one at a time (OAT), always stating from a central (nominal) point in the space of the $x_1, x_2, \ldots x_k$. This approach appears self-evident and 'safe' in the sense that, if a model crashes during the SA, one knows which factor is the responsible. However, OAT has many flaws both in terms of design (one should not always move from the same point but try many points to fully explore the range of variability of the model) and statistical property (it completely misses interactions), and hence it should not be used (Saltelli et al. 2019; Saltelli and Annoni 2010). A local SA – run without a Monte Carlo UA, may offer an illusory reassuring image of the predictive capacity of a model.
- Variance based methods offer the following advantages:
  - They are grounded in statistical theory,
  - They decompose the variance into sets of factors, i.e. one can rewrite Equation (2) as $S_u + S_v + S_{uv} = 1$ where factors $x_1, x_2, \ldots x_k$ have been partitioned in two groups u and v,
  - They are easy to interpret; e.g. Equation (4) is the fractional variance that would be left on average if all factors but x_i could be fixed, and
  - S_i's and T_i's can be linked to well-defined experimental settings – e.g. in order to decide if a variable can be fixed one need to use T_i and not S_i. All this is treated in Saltelli *et al.* (2008).

Sensitivity analysis in practice

We pause now with the mathematics of SA and move to its philosophy with a series of suggestions for the ecologist practitioner.

Chose one and only one output of interest. Since a model may produce many outputs (e.g., time-series, spatially-distributed), we suggest to run SA only on the output that helps answering the question posed by the analysis. E.g. if one is examining soil contamination, a summary measure such as the fraction of the area where a threshold is exceeded might do the job better that many tables of sensitivity indices.

Be open to the possibility that the model produces uncertainties so wide as to make its predictions irrelevant. If this happens, it could simply mean that the quality of the evidence feeding into the model does not allow meaningful estimates to be produced. One should then change the model or the question asked from it. We recommend this approach to tame 'modelling hubris', e.g. the temptation to develop larger and larger models (Saltelli et al. 2020), see Figure 2.



[Figure 2 here]. The conjecture of O'Neil (O'Neill 1971; Turner and Gardner 2015) suggests that while developing a model into more and more complex formulation the error initially decreases as the model becomes a better match to data, then it worsens due to the uncertainty cascade effect, i.e. the accumulation of error as more and more factors are included. In data analysis this corresponds to the two opposing effects of under- and over-parameterization. See also (Saltelli 2019).

Consider extending the set of input factors using triggers. If one is uncertain about epistemic features of the model – e.g. what formula to use for a particular phenomenon in the model, a trigger may allow one to select two or more formulae 'at runtime' – e.g. if $x_i < 0.5$ then choose formula A, if $x_i \geq 0.5$ use formula B. The same may apply to different grid resolutions, choice of algorithms in the model, and so on. The effect of triggers on the model output should be examined jointly with parametric uncertainties to capture possible interaction effects.

Why running a model just once? In the process of building a model, time and effort can be minimized by running systematically the model in Monte Carlo simulations: instead of executing the model once, execute it one hundred times, or even maybe only ten. Interesting discoveries or question may arise:

- Bugs can be detected more quickly and fixed, instead of carrying them forward in the model building.
- An addition to a model makes no change to the output in none of the points tested; is the addition necessary?
- An addition makes a change which exceed expectation; why was this the case?

Avoid lying with SA. It is said that one can lie with statistics. One can lie with SA by varying only some factors, implicitly assuming that all other are perfectly known. In an adversarial setting, this risk being exposed by the opposing party. Scarce attention to uncertainties ultimately erodes trust in modelling. It happens frequently that models run to produce point estimate are revealed as non-conservative when uncertainties are properly plugged in (Puy, Lo Piano, and Saltelli 2020). An OAT approach is also vulnerable

to deconstruction for the reasons discussed above. Another way of making a perfunctory SA is to bypass the stage of careful appraisal of the $p_1(x_1), p_2(x_2), \dots p_k(x_k)$ and perform an analysis where all factors have the same uncertainty, e.g. 5% or 10%. These analyses are a case of GIGO, garbage in, garbage out, as instances where all factors are equally uncertain are possibly non-existent in ecology.

Consider via negativa. Some authors, including us, recommend using models also to disprove rather than to prove a give thesis (Oreskes 2010; Oreskes, Shrader-Frechette, and Belitz 1994; Saltelli and Giampietro 2017). Via negativa can provide valuable insights because:

- 'Wrongs' are more evident than 'rights'
- Knowledge grows by subtracting what cannot be
- "Actions that remove are more robust than those that add because addition may have unseen, complicated feedback loops" (Taleb 2012).

Other methods**:** The literature offers several other interesting methods for sensitivity analysis. When for some reasons one is not interested in the variance of the output, e.g. because its distribution is very skewed or long-tailed, then one may resort to moment-independent measures. These permit e.g. ranking factors based on how – fixing them – affects the entire probability distribution function - rather than just its variance. These measures are named moment-independent (Borgonovo and Iooss 2016).

Shapley coefficients used by economists can be related to the sensitivity coefficients just discussed (Owen 2014). Many practitioners use the method of Morris (Morris 1991), which is also close to the total sensitivity index T_i and is recommended when only few simulations can be performed. Morris needs more modelling assumptions than T_i and is more cumbersome to interpret as it produces two measures for each factor. For this we would rather suggest T_i at low sample size rather than Morris (Campolongo, Saltelli, and Cariboni 2011).

Large, CPU-intensive models: Variance-based indices are rather expensive to compute in terms of number of simulations; computing all the S_i's and all the T_i's may come to a cost of $N(k+2)$ where N may be of the order of hundreds or thousands. When the model cannot afford this number, one may use emulators, replacement models that run cheaply. See for an example (Schöbi, Sudret, and Wiart 2015).

Other readings: Razavi *et al.* (2021) describes future orientations for SA. Recent reviews are Norton (2015) and Wei *et al.* (2015).

A worked example: We conclude this voice with a worked example of UA and SA. The Bateman equations are a simple model describing radioactive chain decay – or ecological transfer of a pollutant between compartments without feedback terms. Here x_i is the concentration of the pollutant in the *i*-th compartment:

(5) $\frac{dx_1}{dt} = -\lambda_1 x_1$

$\frac{dx_2}{dt} = -\lambda_2 x_2 + \lambda_1 x_1$

…

$\frac{dx_k}{dt} = -\lambda_k x_k + \lambda_{k-1} x_{k-1}$

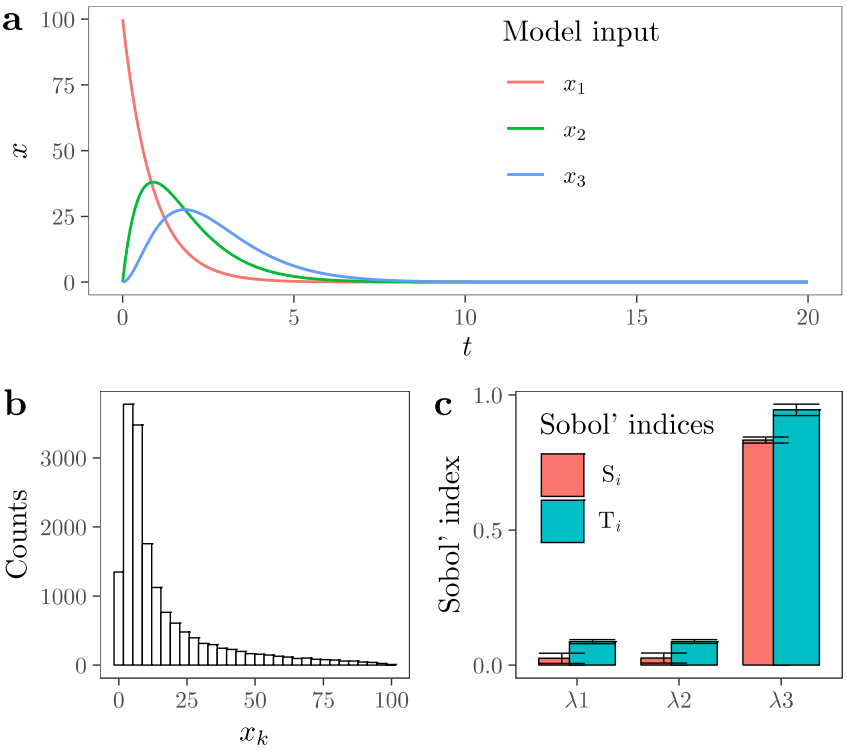
Whose solution e.g. for the last term in the chain is

(6) $x_k = \frac{x_k^0}{\lambda_k} \sum_{i=1}^{k} \lambda_i \alpha_i e^{-\lambda_i t}$; x_k^0 is the amount at time equal zero, $x_j^0 = 0$ for $j < k$ and

$$\alpha_i = \prod_{j=1, j\neq i}^{k} \frac{\lambda_j}{\lambda_j - \lambda_i}$$

The concentration x_k of the pollutant in the final compartment thus depends on the uncertainty in lambda, which we describe with a log-uniform distribution with support $(10^2, 10)$ to make all rates equally uncertain. We set the number of compartments at $k = 3$ and $x_1 = 100$.

[Figure 3 here]: The Bateman equations; a) Dynamics for $k = 3$ and$t = 20$. b) Uncertainty analysis. c) Sensitivity analysis.



References

Antoniadis, Anestis, Sophie Lambert-Lacroix, and Jean-Michel Poggi. 2021. "Random Forests for Global Sensitivity Analysis: A Selective Review." *Reliability Engineering & System Safety* 206 (February): 107312. https://doi.org/10.1016/j.ress.2020.107312.

Becker, William, Paolo Paruolo, and Andrea Saltelli. 2021. "Variable Selection in Regression Models Using Global Sensitivity Analysis." *Journal of Time Series Econometrics* 13 (2): 187–233. https://doi.org/10.1515/jtse-2018-0025.

Bénesse, Clément, Fabrice Gamboa, Jean-Michel Loubes, and Thibaut Boissin. 2021. "Fairness Seen as Global Sensitivity Analysis." arXiv. https://doi.org/10.48550/arXiv.2103.04613.

Borgonovo, Emanuele, and Bertrand Iooss. 2016. "Moment-Independent and Reliability-Based Importance Measures." In *Handbook of Uncertainty Quantification*, edited by Ghanem R., Higdon D., and Owhadi H., 1–23. Cham: Springer.

Campolongo, Francesca, Andrea Saltelli, and Jessica Cariboni. 2011. "From Screening to Quantitative Sensitivity Analysis. A Unified Approach." *Computer Physics Communications* 182 (4): 978–88. https://doi.org/10.1016/j.cpc.2010.12.039.

Constantine, Paul G., Eric Dow, and Qiqi Wang. 2014. “Active Subspace Methods in Theory and Practice: Applications to Kriging Surfaces.” *SIAM Journal on Scientific Computing* 36 (4): A1500–1524. https://doi.org/10.1137/130916138.

Di Fiore, Monica, Marta Kuc Czarnecka, Samuele Lo Piano, Arnald Puy, and Andrea Saltelli. 2022. “The Challenge of Quantification: An Interdisciplinary Reading.” *Minerva*, December. https://doi.org/10.1007/s11024-022-09481-w.

Homma, Toshimitsu, and Andrea Saltelli. 1996. “Importance Measures in Global Sensitivity Analysis of Nonlinear Models.” *Reliability Engineering & System Safety* 52 (1): 1–17. https://doi.org/10.1016/0951-8320(96)00002-6.

Hoyt, Christopher, and Art B. Owen. 2021. “Efficient Estimation of the ANOVA Mean Dimension, with an Application to Neural Net Classification.” *SIAM/ASA Journal on Uncertainty Quantification* 9 (2): 708–30. https://doi.org/10.1137/20M1350236.

Iooss, Bertrand, Ron S. Kenett, and Piercesare Secchi. 2022. “Different Views of Interpretability.” In *Interpretability for Industry 4.0 : Statistical and Machine Learning Approaches*, Antonio Lepore, Boagio Palumbo, Jean-Michel Poggi. Springer International Publishing.

Liu, Sifan, and Art B. Owen. 2022. “Pre-Integration via Active Subspaces.” arXiv. https://doi.org/10.48550/arXiv.2202.02682.

Montella, Alfonso, Massimo Aria, Antonio D’Ambrosio, and Filomena Mauriello. 2012. “Analysis of Powered Two-Wheeler Crashes in Italy by Classification Trees and Rules Discovery.” *Accident Analysis & Prevention*, PTW + Cognitive impairment and Driving Safety, 49 (November): 58–72. https://doi.org/10.1016/j.aap.2011.04.025.

Morris, Max D. 1991. “Factorial Sampling Plans for Preliminary Computational Experiments.” *Technometrics* 33 (2): 161. https://doi.org/10.2307/1269043.

Norton, J.P. 2015. “An Introduction to Sensitivity Assessment of Simulation Models.” *Environmental Modelling & Software* 69 (C): 166–74. https://doi.org/10.1016/j.envsoft.2015.03.020.

O’Neill, R. V. 1971. “Error Analysis of Ecological Models.” In *Radionuclides in Ecosystems*. Oak Ridge - Tenn.

Oreskes, Naomi. 2010. “Why Predict? Historical Perspectives on Prediction in Earth Sciences.” In *Prediction, Science, Decision Making and the Future of Nature*, edited by Daniel R.; Sarewitz, Roger A.; Pielke, and Radford. Byerly. Island Press. http://trove.nla.gov.au/work/6815576?selectedversion=NBD21322691.

Oreskes, Naomi, Kristin Shrader-Frechette, and Kenneth Belitz. 1994. “Verification, Validation, and Confirmation of Numerical Models in the Earth Sciences.” *Science* 263 (5147). http://science.sciencemag.org/content/263/5147/641.

Owen, Art B. 2014. “Sobol’ Indices and Shapley Value.” *SIAM/ASA Journal on Uncertainty Quantification* 2 (1): 245–51. https://doi.org/10.1137/130936233.

Puy, Arnald, Pierfrancesco Beneventano, Simon A. Levin, Samuele Lo Piano, Tommaso Portaluri, and Andrea Saltelli. 2022. “Models with Higher Effective Dimensions Tend to Produce More Uncertain Estimates.” *Science Advances* 8 (eabn9450).

Puy, Arnald, Samuele Lo Piano, and Andrea Saltelli. 2020. “Current Models Underestimate Future Irrigated Areas.” *Geophysical Research Letters* 47 (8): e2020GL087360).

Puy, Arnald, and Andrea Saltelli. 2022. “Discrepancy Measures for Sensitivity Analysis in Mathematical Modeling, ArXiv:2206.13470.” arXiv. https://doi.org/10.48550/arXiv.2206.13470.

Razavi, Saman, Anthony Jakeman, Andrea Saltelli, Clémentine Prieur, Bertrand Iooss, Emanuele Borgonovo, Elmar Plischke, et al. 2021. “The Future of Sensitivity Analysis: An Essential Discipline for Systems Modeling and Policy Support.” *Environmental Modelling & Software*, 104954–104954. https://doi.org/10.1016/j.envsoft.2020.104954.

Saltelli, Andrea. 2019. “Statistical versus Mathematical Modelling: A Short Comment.” *Nature Communications* 10: 1–3. https://doi.org/10.1038/s41467-019-11865-8.

Saltelli, Andrea, Ksenia Aleksankina, William Becker, Pamela Fennell, Federico Ferretti, Niels Holst, Sushan Li, and Qiongli Wu. 2019. “Why so Many Published Sensitivity Analyses Are False: A Systematic Review of Sensitivity Analysis Practices.” *Environmental Modelling & Software* 114 (April): 29–39. https://doi.org/10.1016/J.ENVSOFT.2019.01.012.

Saltelli, Andrea, and Paola Annoni. 2010. “How to Avoid a Perfunctory Sensitivity Analysis.” *Environmental Modelling & Software* 25 (12): 1508–17. https://doi.org/10.1016/j.envsoft.2010.04.012.

Saltelli, Andrea, Paola Annoni, Ivano Azzini, Francesca Campolongo, Marco Ratto, and Stefano Tarantola. 2010. “Variance Based Sensitivity Analysis of Model Output. Design and Estimator for the Total Sensitivity Index.” *Computer Physics Communications* 181 (2): 259–70. https://doi.org/10.1016/j.cpc.2009.09.018.

Saltelli, Andrea, Gabriele Bammer, Isabelle Bruno, Erica Charters, Monica Di Fiore, Emmanuel Didier, Wendy Nelson Espeland, et al. 2020. “Five Ways to Ensure That Models Serve Society: A Manifesto.” *Nature* 582: 482–84.

Saltelli, Andrea, and M. Giampietro. 2017. “What Is Wrong with Evidence Based Policy, and How Can It Be Improved?” *Futures* 91 (February): 62–71.

Saltelli, Andrea, Anthony J. Jakeman, Saman Razavi, and Qiongli Wu. 2021. “Sensitivity Analysis: A Discipline Coming of Age.” *Environmental Modelling and Software* 146, (105226). https://doi.org/10.1016/j.envsoft.2021.105226.

Saltelli, Andrea, and Arnald Puy. 2022. “What Can Mathematical Modelling Contribute to a Sociology of Quantification?” SSRN Scholarly Paper. Rochester, NY. https://doi.org/10.2139/ssrn.4212453.

———. 2023. “Should Statistics Rescue Mathematical Modelling?” *Simulation Modelling Practice and Theory (Submitted)*.

Saltelli, Andrea, M. Ratto, T. H. Andres, F. Campolongo, J. Cariboni, D. Gatelli, M. Saisana, and S. Tarantola. 2008. *Global Sensitivity Analysis : The Primer*. John Wiley. https://doi.org/10.1002/9780470725184.

Schöbi, Roland, Bruno Sudret, and Joe Wiart. 2015. “Polynomial-Chaos-Based Kriging.” *International Journal for Uncertainty Quantification* 5 (2): 171–93. https://doi.org/10.1615/Int.J.UncertaintyQuantification.2015012467.

Sobol’, I.M. 1993. “Sensitivity Analysis for Non-Linear Mathematical Models,.” *Mathematical Modelling and Computational Experiment (Translated from Russian: I.M. Sobol’, Sensitivity Estimates for Nonlinear Mathematical Models, Matematicheskoe Modelirovanie 2 (1990) 112–118)* 1 (4): 407–14.

Taleb, N N. 2012. *Antifragile: Things That Gain from Disorder*. Incerto. Random House Publishing Group. https://books.google.es/books?id=5fqbz_qGi0AC.

Tunkiel, Andrzej T., Dan Sui, and Tomasz Wiktorski. 2020. “Data-Driven Sensitivity Analysis of Complex Machine Learning Models: A Case Study of Directional Drilling.” *Journal of Petroleum Science and Engineering* 195 (December): 107630. https://doi.org/10.1016/j.petrol.2020.107630.

Turner, Monica G., and Robert H. Gardner. 2015. “Introduction to Models.” In *Landscape Ecology in Theory and Practice*, 63–95. New York, NY: Springer New York. https://doi.org/10.1007/978-1-4939-2794-4_3.

Wei, Pengfei, Zhenzhou Lu, and Jingwen Song. 2015. “Variable Importance Analysis: A Comprehensive Review.” *Reliability Engineering & System Safety* 142: 399–432. https://doi.org/10.1016/j.ress.2015.05.018.