Sensitivity Analysis

How to perform comprehensive, efficient, and robust Sensitivity Analysis?

Dr. Debraj Roy

June 11, 2024

Computational Science Lab

Table of contents

1. Introduction

2. Local Sensitivity analysis

Introduction

"Essentially, all models are wrong, but some are useful."

— Box, George E. P.; Norman R. Draper (1987). Empirical Model-Building and Response Surfaces, p. 424, Wiley. ISBN 0471810339.

Case study of Covid

The Telegraph

Coding that led to lockdown was 'totally unreliable' and a 'buggy mess', say experts

The code, written by Professor Neil Ferguson and his team at Imperial College London, was impossible to read, scientists claim

By Hannah Boland *and* Ellie Zolfagharifard

16 May 2020 • 1:32pm

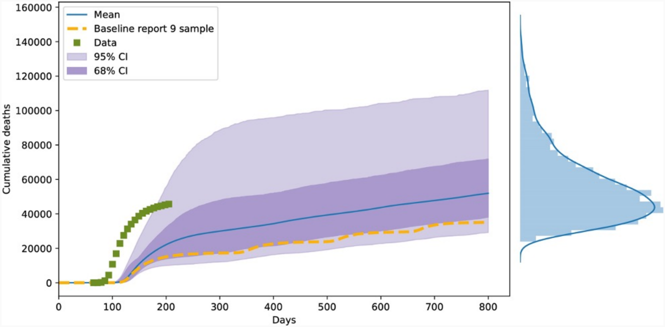
NewScientist

UK's scientific advice on coronavirus is a cause for concern

By Jessica Hamzelou

23 March 2020

Case study of Covid



nature computational science

Article | Published: 22 February 2021

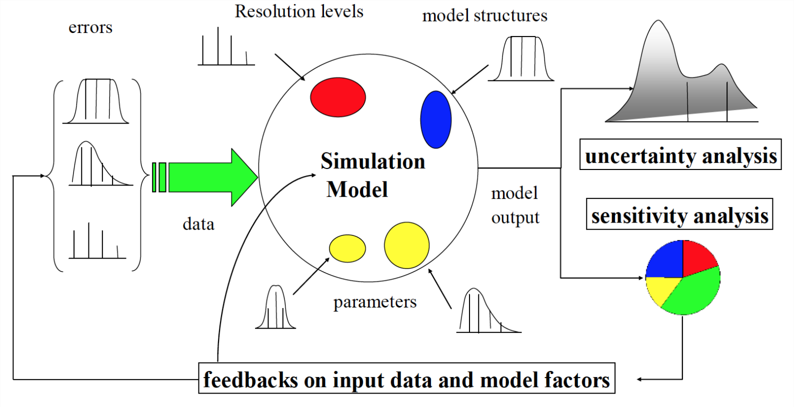
The impact of uncertainty on predictions of the CovidSim epidemiological code

Wouter Edeling, Hamid Arabnejad, Robbie Sinclair, Diana Suleimenova, Krishnakumar Gopalakrishnan, Bartosz Bosak, Derek Groen, Imran Mahmood, Daan Crommelin & Peter V. Coveney

Modelling & Simulation

Uncertainty analysis: Focuses on just quantifying the uncertainty in model output.
Sensitivity analysis: The study of the relative importance of different input factors on the model output.

Modelling & Simulation



Applications of sensitivity analysis

Parameter prioritization
which factors determine output the most

Parameter fixing
which factors can be removed from the model

Variance cutting
which factors, made more certain, would make output more certain

Parameter mapping
which factors are most important for causing good/bad outputs

Types of sensitivity analysis

- Sensitivity analysis (SA) : quantification of the effects of changes or uncertainties in a models input parameters on the model output
  - Local SA: Consider only the effect of changes in individual parameters
  - Global SA: Consider the effects of changes in multiple parameters simultaneously

Local Sensitivity analysis

Which parameters **when nudged** will change the output most?

$$\left.\frac{\partial f}{\partial \theta_i}\right|_{\theta^0}$$

Local sensitivity analysis

Two notable deficiencies of this definition of sensitivity are:

1. First, if f is nonlinear with respect to, then its partial derivative $\frac{\partial f}{\partial \theta_i}$ will change depending on where in the range of you choose to measure.
2. Second, if there are interactions between model inputs, then $\frac{\partial f}{\partial \theta_i}$ will change depending on the values of the remaining input factors as well.

In short, first partial derivatives are only a valid measure of sensitivity when the model is linear, in which case $\frac{\partial f}{\partial \theta_i}$ will remain constant for any θ_i

Local sensitivity analysis: Ricker Model

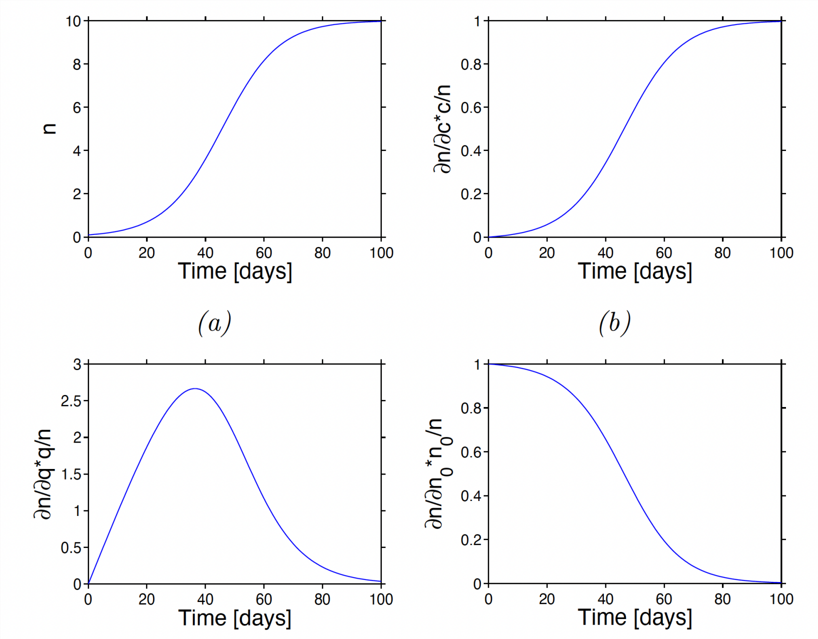
The Ricker model is given by:

$$N_{t+1} = N_t e^{q\left(1-\frac{N_t}{c}\right)}$$

where:

- N_t is the population size at time t.
- q is the intrinsic growth rate.
- c is the carrying capacity.

Local sensitivity analysis: Ricker Model

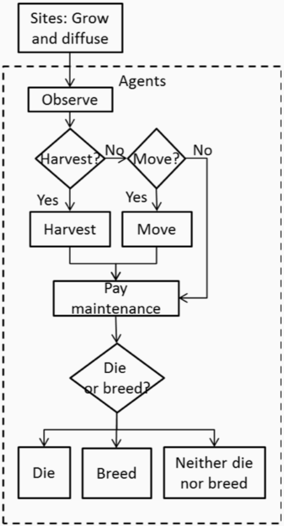


Local sensitivity analysis: Ricker Model

- The elasticities show that on long timescales, c is the only influential parameter. On shorter time scales q and n_0 are more influential.
- q does not influence the steady state,but influences strongly the transient dynamics,especially at intermediat epopulation sizes

Test Case Description

- All decisions are stochastic
- Agents tend to harvest if :
  - internal energy is low
  - resource is abundant
  - other agents are close
- Agents move to sites with :
  - abundant resource
  - few agents



Flow chart for one time step

Input Parameters

- recommended to include all input parameters
- parameters may have different dimensions

Parameter	Description
c	Efficiency
D	Diffusion coefficient
E_b	Birth energy
E_h	Harvest cost
E_m	Cost of energy maintenance
E_{move}	Move cost
K	Carrying capacity
n_0	Initial number of agents
r	Growth rate
R_0	Initial resource
R_{max}	Maximum harvest
R_{unc}	Uncertainty of resource estimations
v_b	Birth coefficient
v_d	Mortality coefficient
z	Variation in offspring traits

Output Parameters

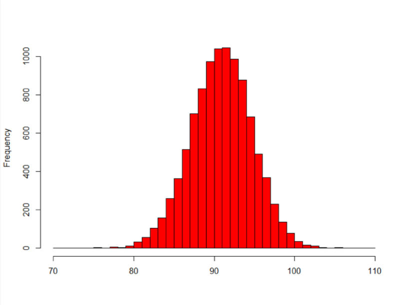
- As main output, we consider the number of agents n
- ABMs produce large amounts of outputs on different levels (e.g., system level, agent level)
- Multiple outputs may be considered separately during sensitivity analysis

Default parameter setting

- First step in most SA methods is to choose a default parameter setting
- This setting acts as a reference point to assess the effects of parameter changes
- First, we run the model a number of times in the default setting

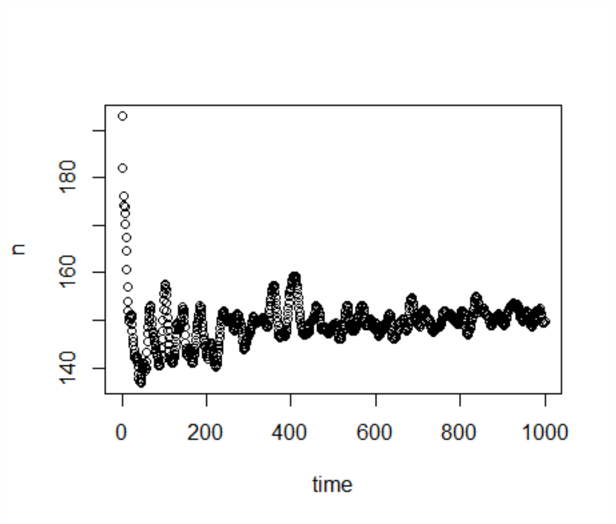
Histogram model runs

- The distribution resembles a normal distribution
- A large number of replicates is needed to estimate this distribution

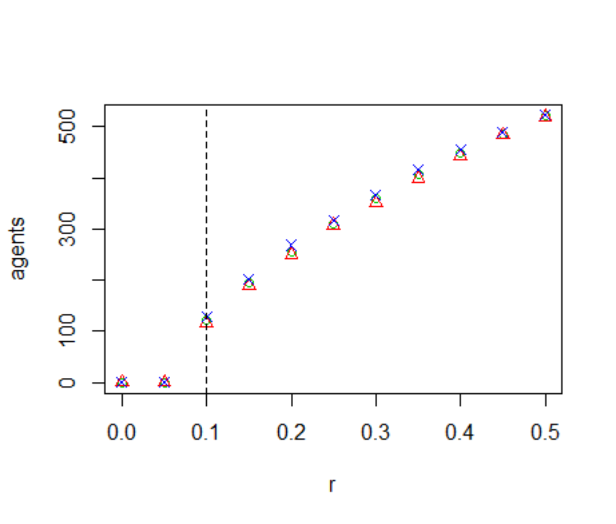


One-factor-at-a-time (OFAT) SA

Model Output



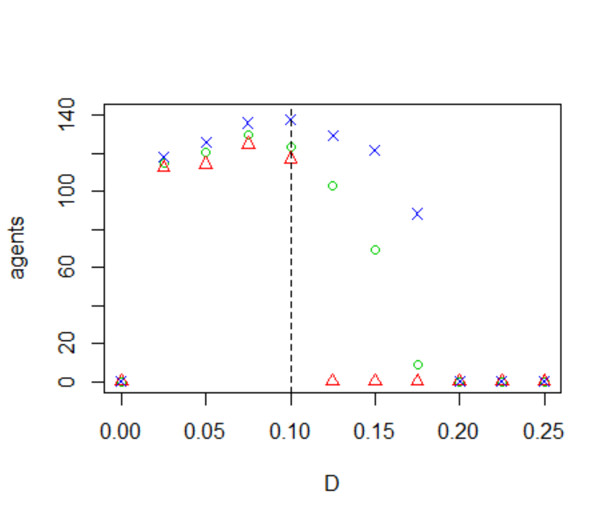
OAT results of r (resource growth rate)



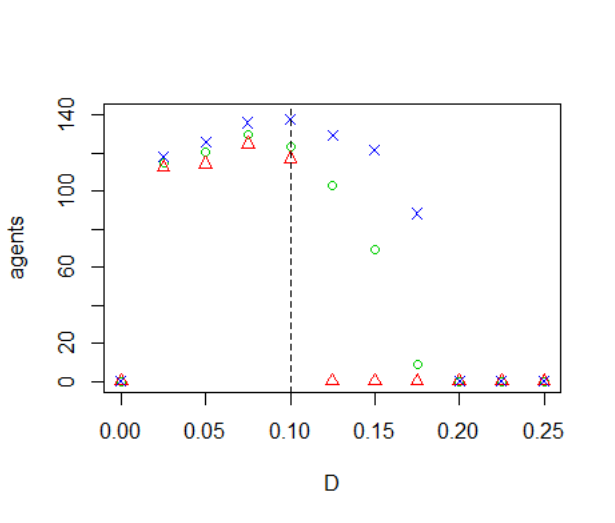
OAT results of r (resource growth rate)

- The spread within replicates is small compared to the spread between replicates (i.e. stochasticity has little influence).
- The effect of the parameter on the output is approximately linear

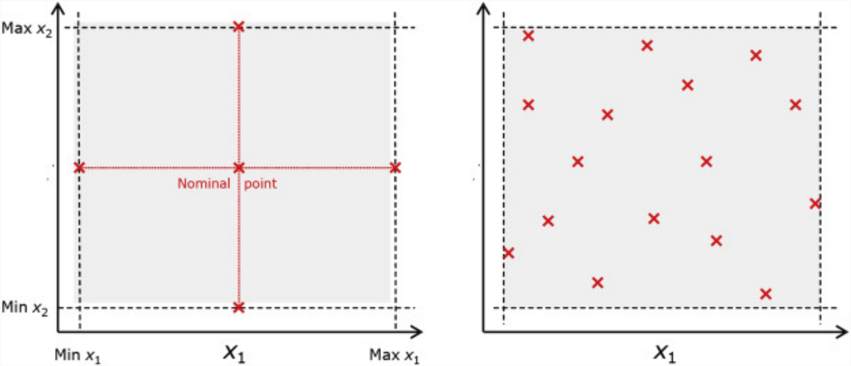
OAT results of D (Diffusion coefficient)



OAT results of C (Carrying capacity)



OAT in 2D: Analyse This?



Example

Two Boolean Parameters

X	Y
0	0
0	1
1	0
1	1

OFAT:

Keep Y fixed at 0 and vary X:

X	Y
0	0
1	0

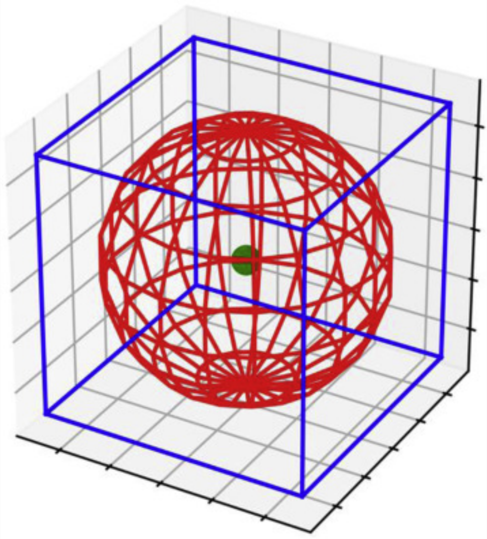
Keep X fixed at 0 and vary Y:

X	Y
0	0
0	1

At the end:

X	Y
0	0
1	0
X	Y
0	0
0	1

OAT in 3D: Analyse This?

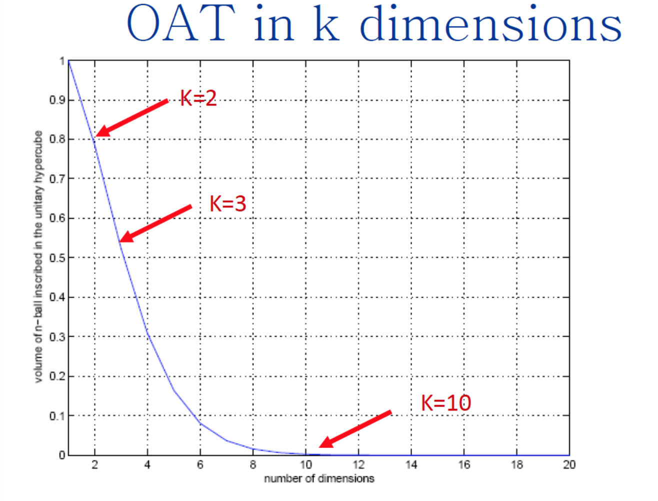


Analyse This?

OAT in 10 dimensions

Volume hypersphere / volume ten dimensional hypercube =? ~ 0.0025

Analyse This?



One-at-a-time (OAT) designs

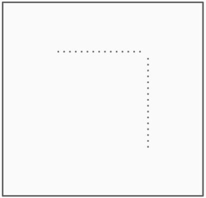
One parameter *changed* at a time

Pros:

In every pair, it's that parameter.

Easy.

Only 2 points per parameter (fixable - Elementary Effects method)



(hypercube in general)

OAT is inefficient

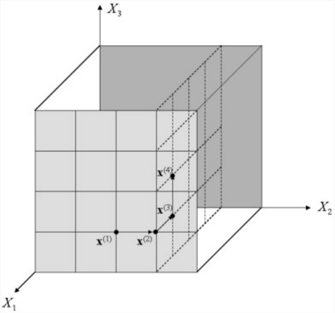
Can we analyse influence of many parameters using few points?

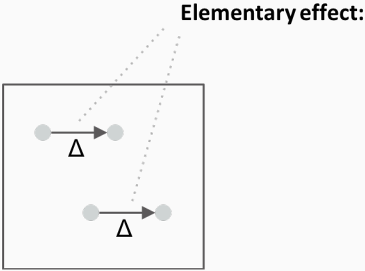
3 points for each of 2 parameters. Parameter effects are mixed (but maybe we can un-mix them)

Conclusions on OFAT

- OFAT is a good method for analysing the qualitative behaviour of your ABM
- It can detect tipping points, and other non-linearities
- OFAT presents a good trade-off between costs (time) and gains (model insight) for ABM goals
    - Parameter effects are not readily comparable due to differences in dimensions and units
    - No interaction effects covered

The elementary effect method





The elementary effect method

$$EE_i = \frac{Y(x_1, \ldots, x_{i-1}, x_i + \Delta_i, x_{i+1}, \ldots, x_k) - Y(x_1, \ldots, x_k)}{\Delta_i}.$$

$$\mu_i = \frac{\sum_{i=1}^{r} EE_i}{r}.$$

$$\sigma_i = \frac{\sum_{i=1}^{r} (EE_i - \mu_i)^2}{r}.$$

The elementary effect method

1. if this term is non null, then X_j has an influence on the output,
2. if this term is non null and does not vary as X_j varies, therefore, X_j has a linear influence on the output and has no interactions with other input factors,
3. if this term varies as X_j varies, then X_j affects non linearly the output with or without interactions.

The elementary effect method

It is not possible with Morris method to distinguish between non linearity and the interactions with other input factors. its drawback is that it is not possible to distinguish non linearity from interactions, which might be essential for the designer

Regression based

$$R^2 = 1 - \frac{\sum (y_j - y_j^r)^2}{\sum (y_j - y_{av})^2}.$$

Regression based

The first-order sensitivity index of parameter i is computed as the value of R^2 when excluding that parameter from the fit,

$$S_{r,i} = 1 - \frac{\sum (y_j - y_j^{r*})^2}{\sum (y_j - y_{av})^2}$$

Regression based

Similarly, the total-order sensitivity index is computed as the explained variance when keeping parameter i, but leaving out all other parameters,

$$S_{tot,r,i} = 1 - \frac{\sum(y_j - y_j^{r'})^2}{\sum(y_j - y_{av})^2}$$

where $y_j^{r'}$ is the predicted outcome by the regression function including only parameter i.

Law of total variance

Derivation of the law of total variance

$$\boxed{\mathrm{var}(X) = \mathbf{E}\big[\mathrm{var}(X \mid Y)\big] + \mathrm{var}\big(\mathbf{E}[X \mid Y]\big)}$$

- $\mathrm{var}(X) = \mathbf{E}[X^2] - (\mathbf{E}[X])^2$

$$\mathrm{var}(X \mid Y = y) = E[X^2 \mid Y=y] - (E[X \mid Y=y])^2 \quad \text{for all } y$$

$$\mathrm{var}(X \mid Y) = E[X^2 \mid Y] - (E[X \mid Y])^2$$

$$\mathbf{E}\big[\mathrm{var}(X \mid Y)\big] = E[X^2] - \cancel{E[(E[X \mid Y])^2]}$$

$$+$$

$$\mathrm{var}\big(\mathbf{E}[X \mid Y]\big) = \cancel{E[(E[X \mid Y])^2]} - (E[E[X \mid Y]])^2$$

$$(E[X])^2$$

Reading I

Ten Broeke, G., Van Voorn, G., & Ligtenberg, A. (2016). Which sensitivity analysis method should I use for my agent-based model?. Journal of Artificial Societies and Social Simulation, 19(1).