Sensitivity and uncertainty analyses

Mark Szalai

ENDURE Summer School 2016
The role of IPM in mitigating pest development under climate change–modelling approaches
Hungary

Gödöllő

Szent István University
Desired modelling result

Clear answer(s)
Considering X the result will be Y

Your modell will be uncertain
= variability in the output
Terminology

What is sensitivity analysis?

and what is uncertainty analysis?
Disclaimer

Andrea Saltelli
Disclaimer
Terminology

Sensitivity analysis (investigating uncertainty to answer)

What's important to model or system development?

What's important to calculated measures of uncertainty?

uncertainty analysis
Sources of uncertainty

1. Some aspects of the system are not exactly known

   Weak point(s) of model strucute
   Garbage in garbage out?
   Remember the model ensembles (Daniel)

2. Uncertainty in the input

   Inputs: parameters and explanatory variables
What to do?

[Activity]
Input?

Literature, spread of a pest :
  input of 3 and 4 (km/year)
The number of generations:
  input of 3 and 4

Investigate the input space! (What is input space?)
What to do?

[Activity]
Investigate the input space! HOW?

Brute force method:
run the model with all possible inputs

It is often time consuming (CPU time, cost)

SAMPLE the input space!
Note

It is often time consuming (CPU time, cost) advise can be: do a more efficient code

It is better to code in 3 hours and run it during your lunch break than code it in 3 days and run it in 10 minutes
What to do?

[Activity]
Investigate the input space! HOW?

Brute force method:
run the modell with all possible inputs

It is often time consuming (CPU time, cost)

SAMPLE the input space!
How to sample the input space?

[Activity] [in pairs?]

two input variables, good sampling

(Remember: we want to do sensitivity analysis)

Problems of one-at-a-time (OAT)

Global SA: the entire input space is interesting

Local SA?

sometimes it is also interesting, but
Global SA - sampling

We can be happy... running our model not once not twice but...

Sampling often means randomness

t'aute randomness'

www.random.org

randomness comes from atmospheric noise
Discrepancy

our desire: a sequence fills the input space leaving no gaps

plot
cost effective?

```r
N=50
dimen=8
x.rand=matrix(runif(N*dimen), nrow=N, ncol=dimen)
par(mfrow=c(2,2))
plot(x.rand[,3:4])
plot(x.rand[,5:6])
plot(x.rand[,7:8])
plot(x.rand[,4:5])
par(mfrow=c(1,1))
```
Quasi random sampling

Low-discrepancy sequences are also called quasi-random or sub-random sequences
Quasi random sampling

Sobol sequence 😊

randtoolbox package --R

\[
\begin{align*}
N &= 50 \\
dimen &= 8 \\
x.sob &= \text{sobol}(N, dimen) \\
x.rand &= \text{matrix}(\text{runif}(N \times dimen), nrow = N)
\end{align*}
\]

\[
\begin{align*}
\text{par(mfrow=c(2,2))} \\
\text{plot}(x.sob[,5:6]) \\
\text{plot}(x.rand[,5:6]) \\
\text{plot}(x.sob[,3:4]) \\
\text{plot}(x.rand[,3:4]) \\
\text{par(mfrow=c(1,1))}
\end{align*}
\]
Quasi random sampling

Sobol sequence  😊
randtoolbox package  --R

hint: correlation in the Sobol sequence
If you have N input factors, generate N+2 dimensions and drop the first two
Quasi random sampling

Sobol sequence 😊, but it is [0,1]

How to translate this to the real inputs? Can I use it for categories such as scenario RCP 2.6, RCP 4.5? Uniform distribution vs non-uniform
Global SA

Problems of one-at-a-time (OAT) approach

We have a nice (Sobol) sequence

That means we first have the inputs
then run the model several times
then ??? Where does SA come???
Global SA

we first have the inputs
then run the model several times
then ??? Where does SA come???

Output?
Let’s assume
the output is a single numerical value
Global SA

we first have the inputs
then run the model several times
then calculate sensitivity metrics

variance based SA
variance decomposition
ANOVA-like

!!! independent inputs !!! e.g. % of maize, % of cont. maize
Variance based global SA

Main effect index (or "first-order sensitivity index"): effect of varying $X_i$ alone, but averaged over variations in other input parameters.

Total-effect index: contribution to the output variance of $X_i$, including all variance caused by its interactions
Variance based global SA

Often
There will be a few influential input
Many inputs have hardly any effect on the output
Variance based global SA

Often
There will be a few influential input
Many inputs have hardly any effect on the output

Simplify?
Where to do more effort?
Global SA

Problem: Too many variables

Variance based method still costs a lot

Try to simplify (always can be suggested by others)

Ok this is the fruit of your work

Screening to identify non-influential inputs

Morris method (this is basically a OAT-way)
Desired modelling result

Clear answer(s)
Considering X the result will be Y
BUT your modell will be uncertain
= variability in the output

How to communicate this to non-modellers?
How to communicate this to non-modellers?

Plots - simple plots

particularly: ggplot2

metamodels (emulators) as you have only a few really influential inputs
How to communicate this to non-modellers?

Plots - simple plots

Interactive plots? Shiny?