Exploratory modeling

Me:

I am good in C language.

Interviewer:

Then write "Hello World" using C.

Me:

```
c   c ccccc c     c     ccccc
c   c c     c     c     c   c
c   c c     c     c     c   c
ccccc ccc   c     c     c   c
c   c c     c     c     c   c
c   c c     c     c     c   c
c   c ccccc ccccc ccccc ccccc

c   c ccccc ccccc c     cccc
c   c c   c c   c c     c   c
c c c c   c ccccc c     c   c
cc cc c   c cc    c     c   c
c   c c   c c c   c     c   c
c   c c   c c  c  c     c   c
c   c ccccc c   c ccccc cccc
```

Session outline

1. Short recap on Python
2. Exploratory modeling concept
3. EMA Workbench
4. Live coding and hands-on


I HAVE QUESTIONS
LOTS OF QUESTIONS

Part 1 Short recap on Python

Short recap

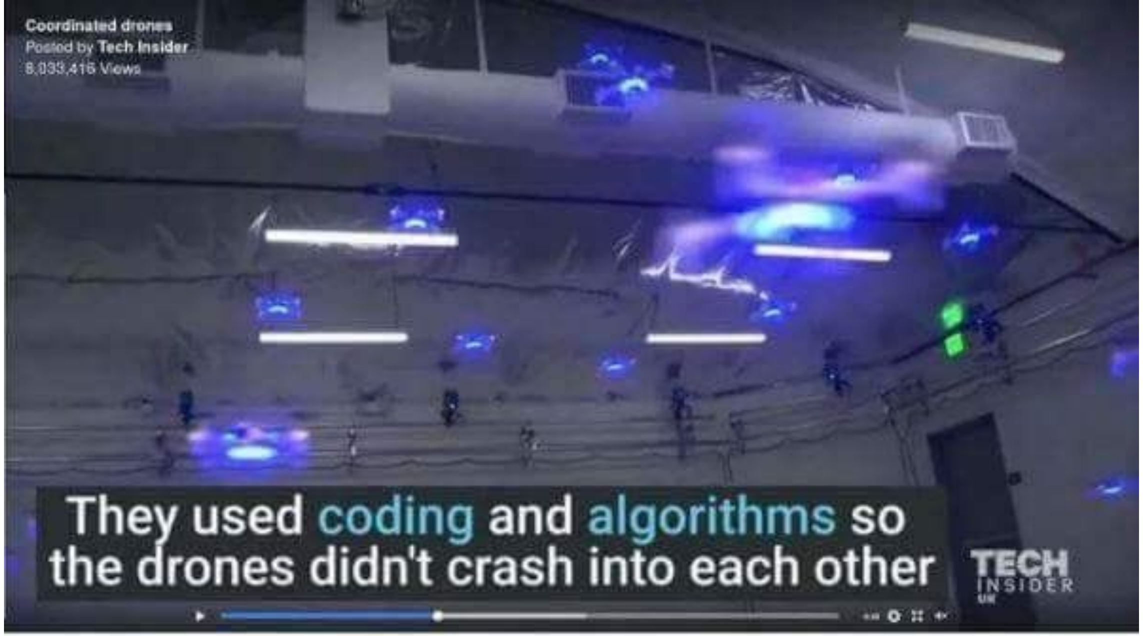
- Anaconda?
- Python?
- Jupyter Notebook?
- Modules and Package?

Anaconda distribution

- Anaconda is a **distribution** of the **Python** language for data science and machine learning related applications **that aims to simplify package management** and **deployment**.

Python Programming Language

- **Python** is an interpreted high-level **programming language** for general-purpose programming.



16k 485 Share

BEST

u/Skizm • 2mo

```
if(goingToCrashIntoEachOther)
{ dont(); }
```

Jupyter Notebook

- The **Jupyter Notebook** is an open-source **web application** that **allows** you to **create** and share documents that contain live **code**, **equations**, **visualizations** and **narrative text**.

Python Modules and Packages

Module is a **piece of software** that has a specific functionality. For example, if you want to have a module that will draw lines, then run following:

from matplotlib import pyplot

Packages are **namespaces** which **contain multiple packages** and **modules** themselves. If you want to have instruments to work with data frames, then run following:

import pandas

How to install packages?

- Globally
  a. Find and run Anaconda Prompt
  b. Type **conda install <package_name>**
     - For example, conda install geopandas
  c. Click enter
- Locally
  a. Find and run command line
  b. Type **cd anaconda3**
  c. Type **python -m pip install <package_name>**
     - For example, python -m pip install geopandas

Where to look for help?

- https://stackoverflow.com/
- https://github.com
- Just Google It!

Let’s practice!

- Try to install following package: ema_workbench

“Answer”

- Find and open command line
- Type cd anaconda3
- Type python -m pip install ema_workbench

Now you know!

- What is Anaconda
- What is Python
- What modules and packages
- How to install packages

Part 2 Exploratory modeling

Problem 1

- Imagine you are a flood risk manager. You need to decide on an investment in dikes for the coming ten years. You have been given the following information
  - Chance of a severe flood is 1 in 6
  - Damage in case of a flood is 1 million
  - Costs of investment in dikes is 1.3 million
  - If dikes are built, chance of flood falls below 1 in 50
- Will you invest in the dikes?

Problem 2

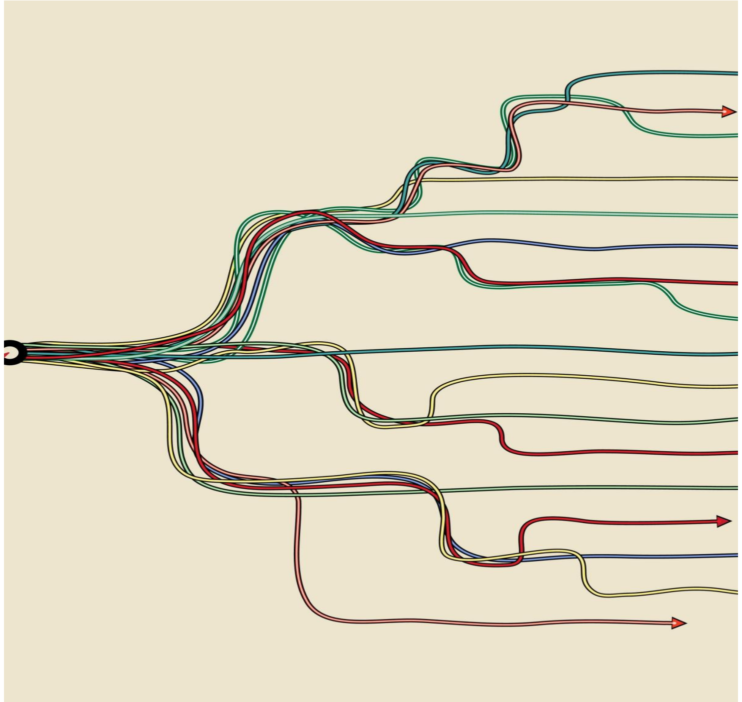
- Imagine you are a flood risk manager. You need to decide on an investment in dikes for the coming ten years. You have been given the following information
    - According to a first group of experts, the chance of a severe flood is 1 in 6
    - According to a second group of experts, the chance of a severe flood is 1 in 10
    - Damage in case of a flood is 1 million
    - Costs of investment in dikes is 1.3 million
    - If dikes are built, chance of flood falls below 1 in 50
- Will you invest in the dikes?

Problem 3

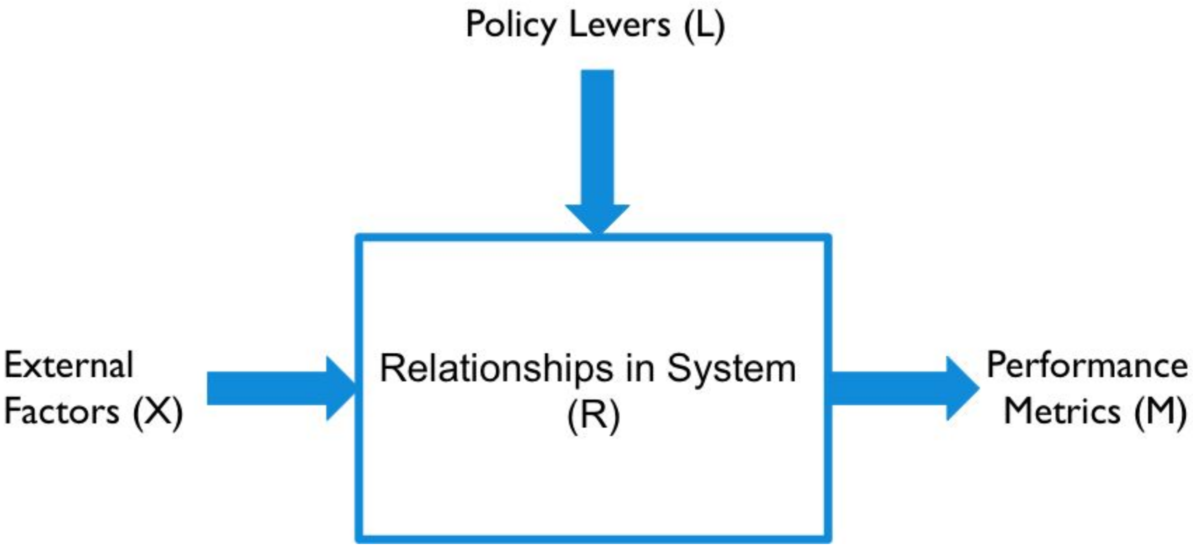
- The two groups of experts worked together on making a model of the occurrence of floods. They came back with a cone. If it lands on its base, there is no flood, if it lands on the small side there is a flood, and they don't know whether it can land on its side.
- Will you invest in the dikes?

Deep uncertainty

- Implication: set of plausible models of the system, set of outcomes of interest without a priori weighting, and sets of scenarios
- Decision makers and stakeholders do not know or cannot agree on the outcomes of interest, the system under study, or future developments



XLRM Framework



Decision-making under deep uncertainty

Key ideas:

1. **Exploratory modeling** – systems are complex and their context is deeply uncertain, human reasoning alone is incapable of handling this. We need computer assisted reasoning.
2. **Adaptive planning** – plans should be designed from the outset to be adapted over time in response to how the future is actually unfolding
3. **Supporting decision-making** – the aim of decision advice is to facilitate learning about a problem and potential courses of action, not to dictate the right solution. This entails a shift from a priori to a posteriori decision analysis.

Exploratory modeling

More formally **Exploratory Modeling** is a **research method** that uses computational experimentation for analyzing complex and uncertain systems (Bankes, 1993; Bankes et al., 2013).

Exploratory modeling approach

- Agnostic about modeling paradigm: SD, agent-based, etc.
- EMA is not merely a post-processing step after a model has already been build
  - in ***conceptualization***, identify key uncertainties;
  - in ***specification***, design model to explore over uncertainties;
  - in ***verification*** and ***validation*** assess comprehensiveness of exploratory character;
  - etc.
- The iterative analysis of results is the most time consuming phase
  - Risk of information overload.

Now you know!

- What is exploratory modeling?
- How the model is represented in exploratory modeling concept?
- What is deep uncertainty?

Part 3 EMA Workbench

EMA Workbench

- EMA Workbench is a Python package that provide you with an opportunity to use exploratory modeling techniques, for instance:
    - Open exploration (sampling)
    - Sensitivity analysis (Sobol indices, Feature Scoring, etc.)
    - Scenario discovery (PRIM)
    - Multi-objective robust optimization
- Was developed by TU Delft associate professor Jan Kwakkel
- It allows you to “connect” a Vensim model to Python

What is EMA Workbench?

- Model: Python, Vensim and Excel
- Inputs (levers) what kind of means you will use to influence your model?
- Outputs (outcomes) what are the outcomes of interest?
- Uncertainties about what you’re not sure?

How to use EMA Workbench?

1. Step 1 Import the packages: ema_workbnech
2. Step 2 Load a model: Python, Vensim, Excel or even discrete-event
3. Step 3 Specify levers, outcomes and uncertainties: What are your policies (combination of levers)?
4. Step 4 Perform experiments
5. Step 5 Run the method
6. Step 6 Visualize the results

Where to look for help?

https://emaworkbench.readthedocs.io/en/latest/

Part 4 Live coding and hands-on

What to should we do?

1. Together, we reprogram the scripts step-by-step;
2. You will check the “basic” scripts to understand how to do it;
3. You might do “exercises” by yourself;
4. Finally, you check “advanced” scripts.

What files to use?

1. If option 1 is chosen, then:
   a. lake_model_open_exploration_draft.ipynb
   b. lake_model_scenario_discovery_draft.ipynb
   c. lake_model_sensitivity_analysis_draft.ipynb
2. If option 2 is chosen, then:
   a. lake_model_open_exploration.ipynb
   b. lake_model_scenario_discovery.ipynb
   c. lake_model_sensitivity_analysis.ipynb

What files to use?

1. If option 3 is chosen, then:
    a. lake_model_open_exploration_exercise.ipynb
    b. lake_model_open_exploration_answer.ipynb
2. If option 4 is chosen, then:
    a. dike_model_open_exploration_demo.ipynb
    b. dike_model_sensitivity_analysis_demo.ipynb
    c. dike_model_directed_search_worst_case_demo.ipynb

Download the scripts

Go to https://github.com/kgb101/sd_summer_school and click the green button called “Clone or download”, “Download ZIP”

The case - “Lake problem”

- The lake problem is a stylized and hypothetical decision problem where the **population** of a city **has to decide** on the **amount of annual pollution** it will put into a lake.
- If the **pollution** in the lake **passes a threshold**, it will suffer **irreversible eutrophication**.

Uncertainties

Deep uncertainty is presented by uncertainty about the:

- mean μ and standard deviation σ of the lognormal distribution characterizing the natural inflow
- b: lake's natural recycling rate
- q: rate of recycling phosphor from the sediment
- δ: discount rate

Parameter	Range	Default value
μ	0.01 – 0.05	0.02
σ	0.001 – 0.005	0.0017
b	0.1 – 0.45	0.42
q	2 – 4.5	2
δ	0.93 – 0.99	0.98

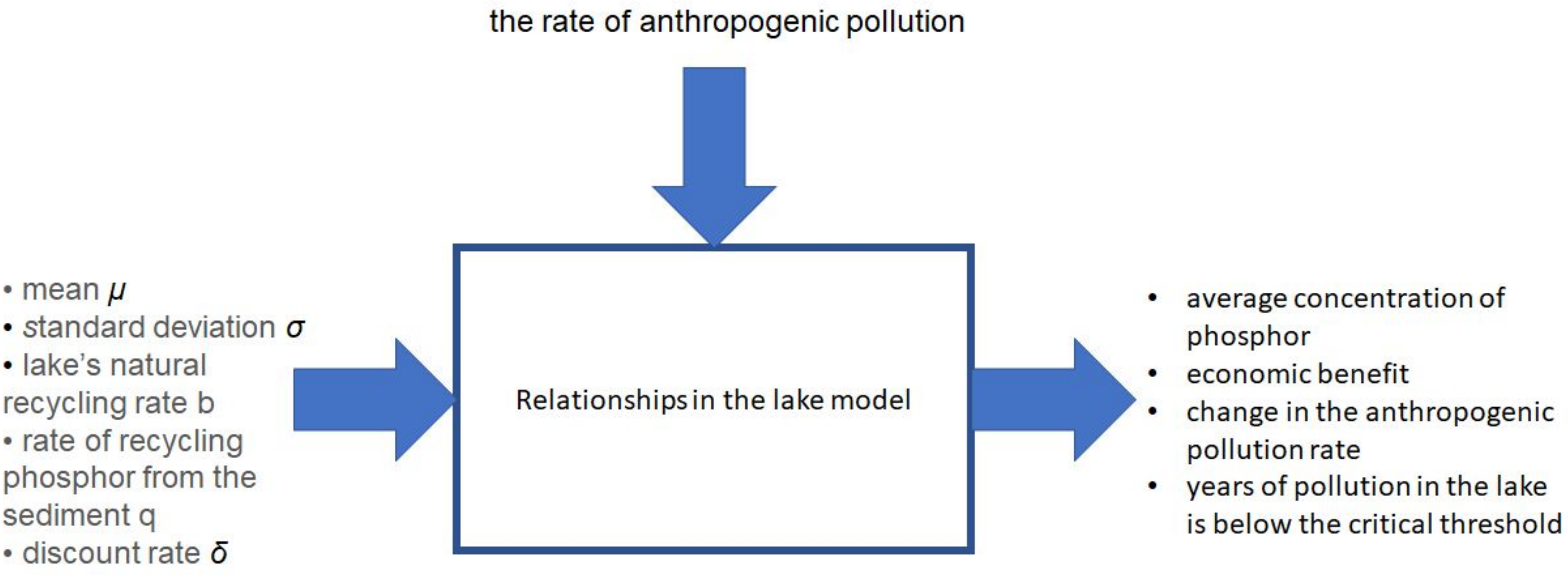
Outcomes

The outcome of interest is:

- average concentration of phosphor in the lake
- economic benefit derived from polluting the lake
- year over year change in the anthropogenic pollution rate
- fraction of years where the pollution in the lake is below the critical threshold

The lever/decision variable is the rate of anthropogenic pollution which is somewhere between 0 and 0.1. The decision maker in our model decide on them at every time step (100).

Lake model (XLRM - Framework)

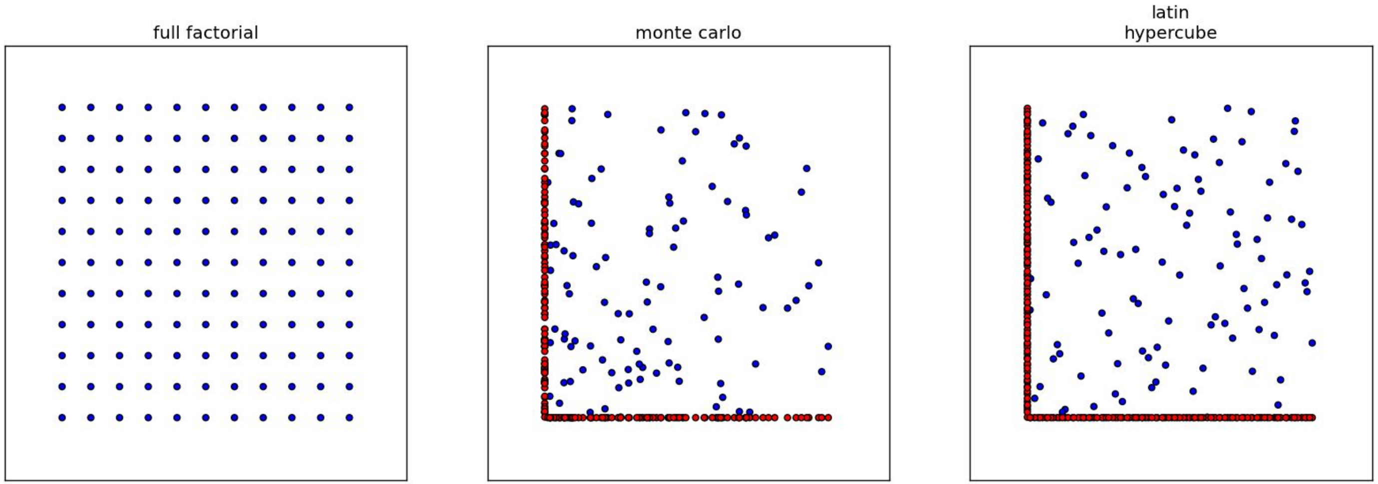


Method 1 Open Exploration

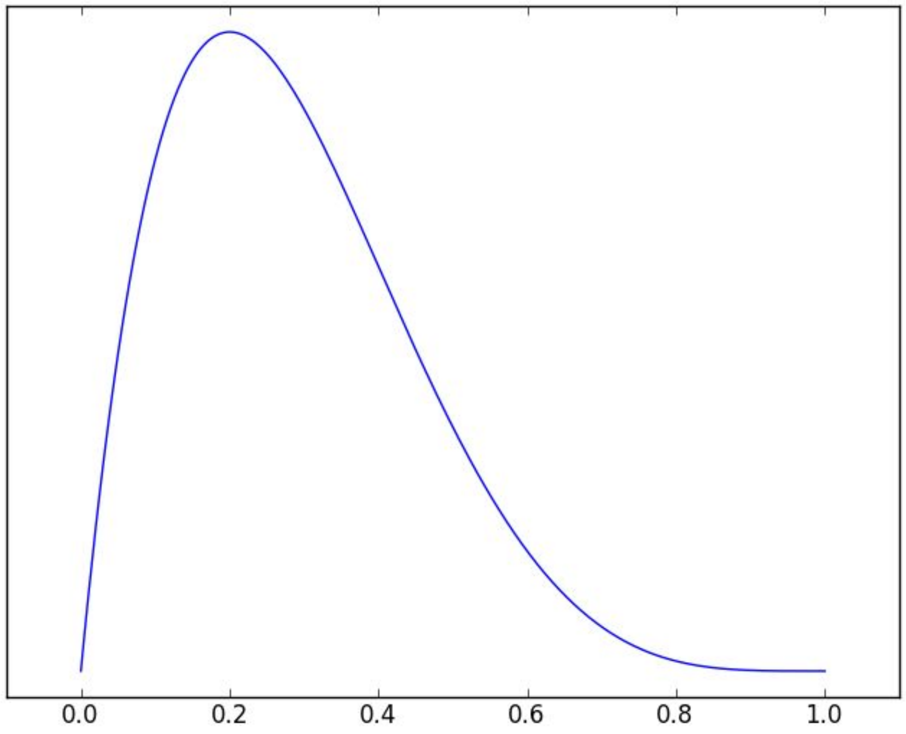
Open exploration

- **Design of experiments**
  - Factorial methods
  - Monte Carlo sampling
  - Latin Hypercube sampling
  - Etc.
- **Used for**
  - Identification of bandwidth outcomes
  - Identification of types of behavior
- **Subsequent analysis**
  - Global sensitivity analysis
  - Subspace partitioning

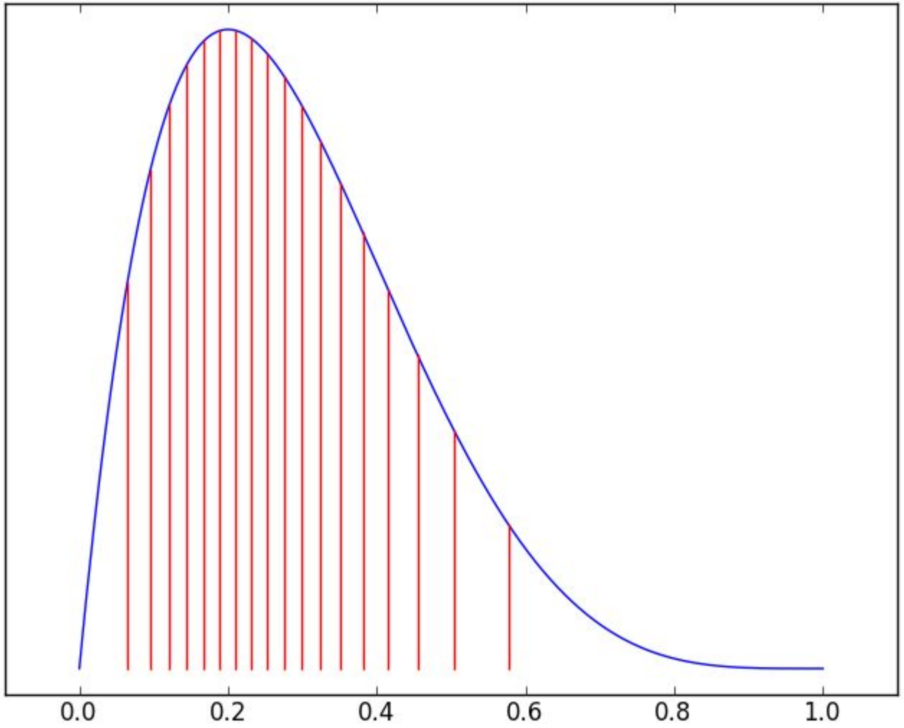
Sampling techniques



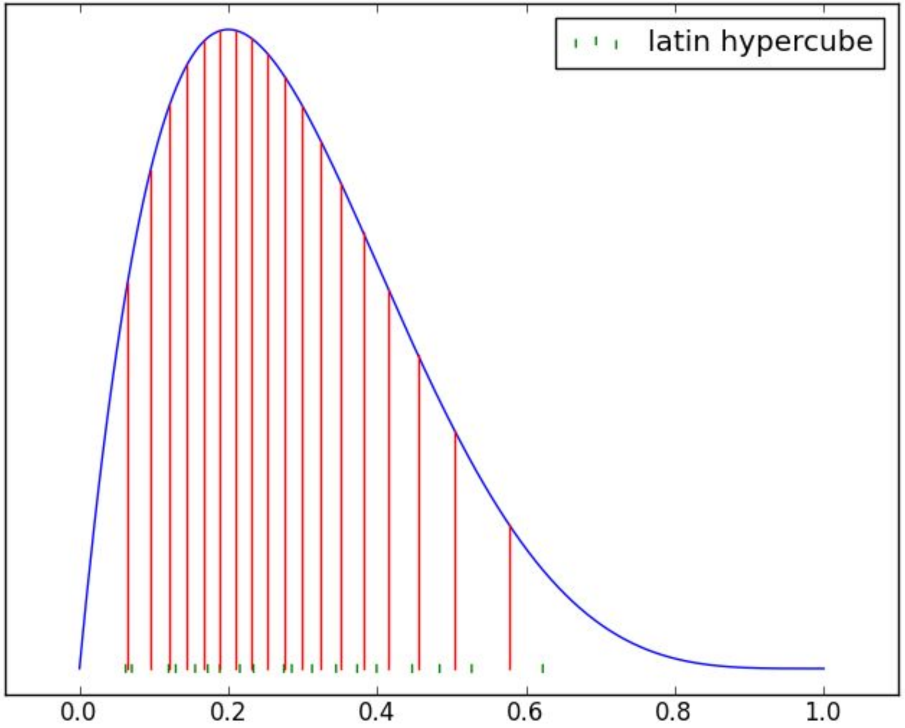
How does a Latin Hypercube work?



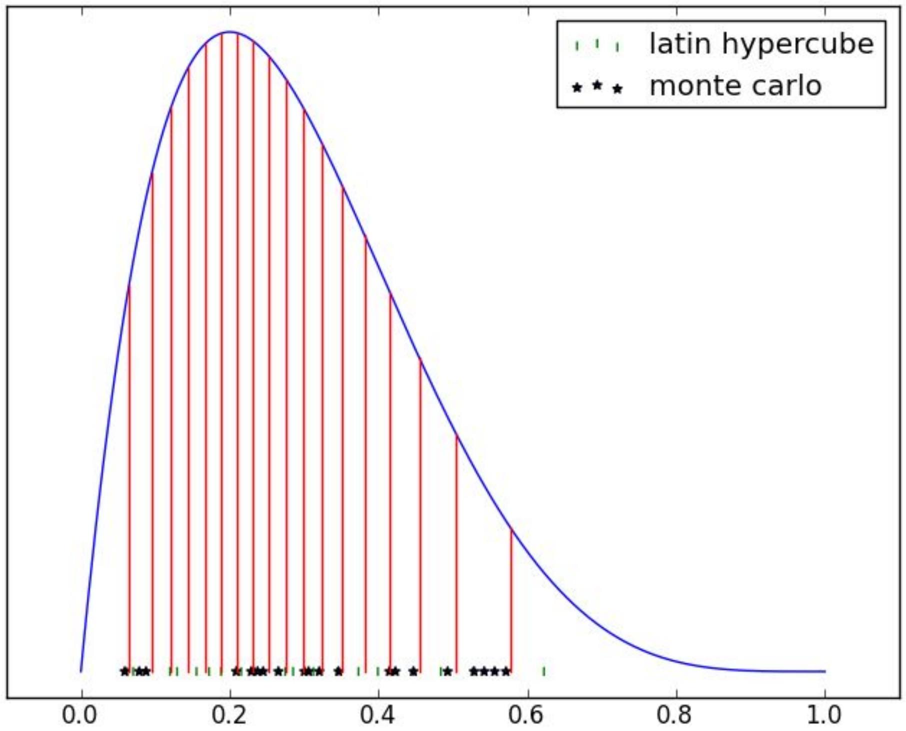
How does a Latin Hypercube work?



How does a Latin Hypercube work?

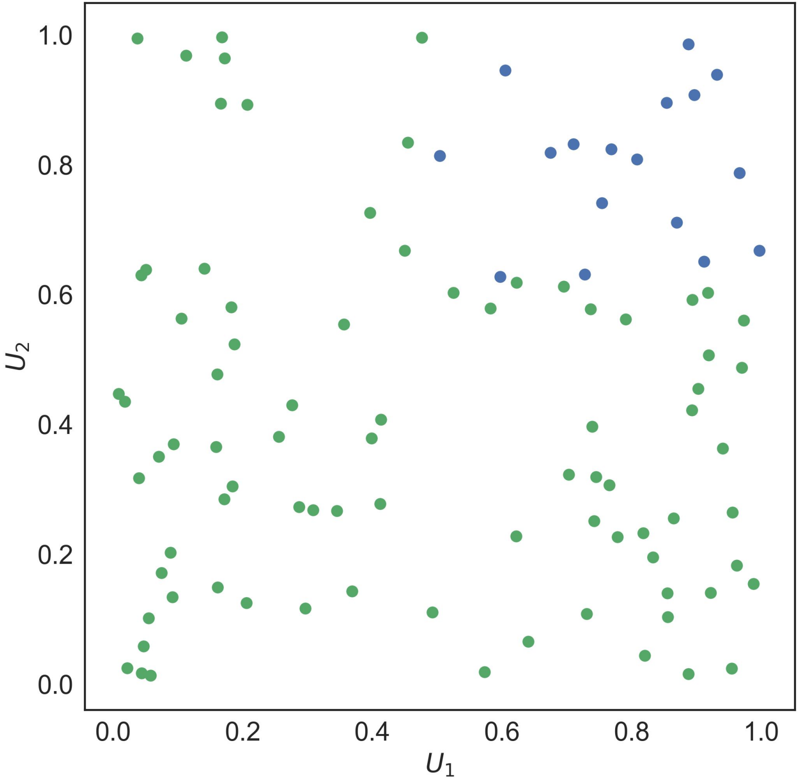


How does a Latin Hypercube work?

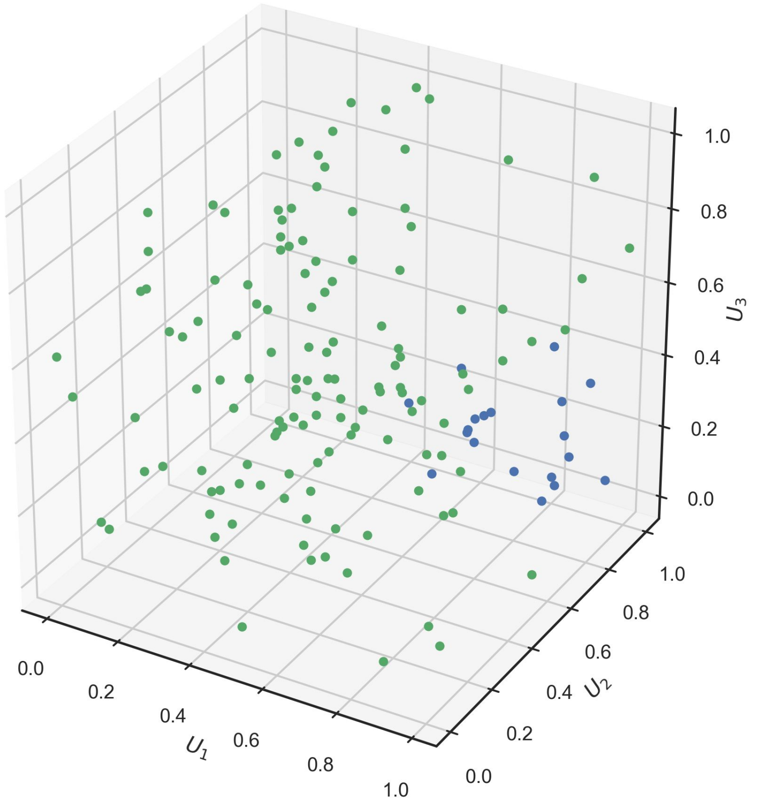


Method 2 Scenario Discovery

Subspace partitioning 2d



Subspace partitioning 3d



Subspace partitioning

- **Problem**: find an (orthogonal) subspace in the model input space, which has a high concentration of cases of interest

- **Rule induction problem**

  Regression vs. (binary) classification

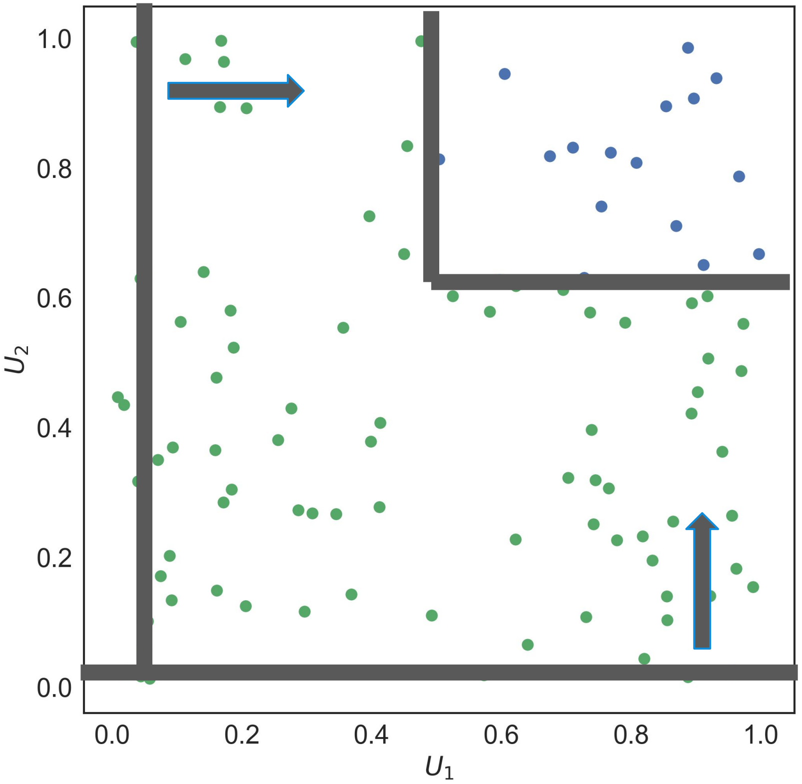
- **Rule induction algorithms**

  Classification and Regression Trees (CART)

  Patient Rule Induction Algorithm (PRIM)

  Various other more specialized possibilities

PRIM

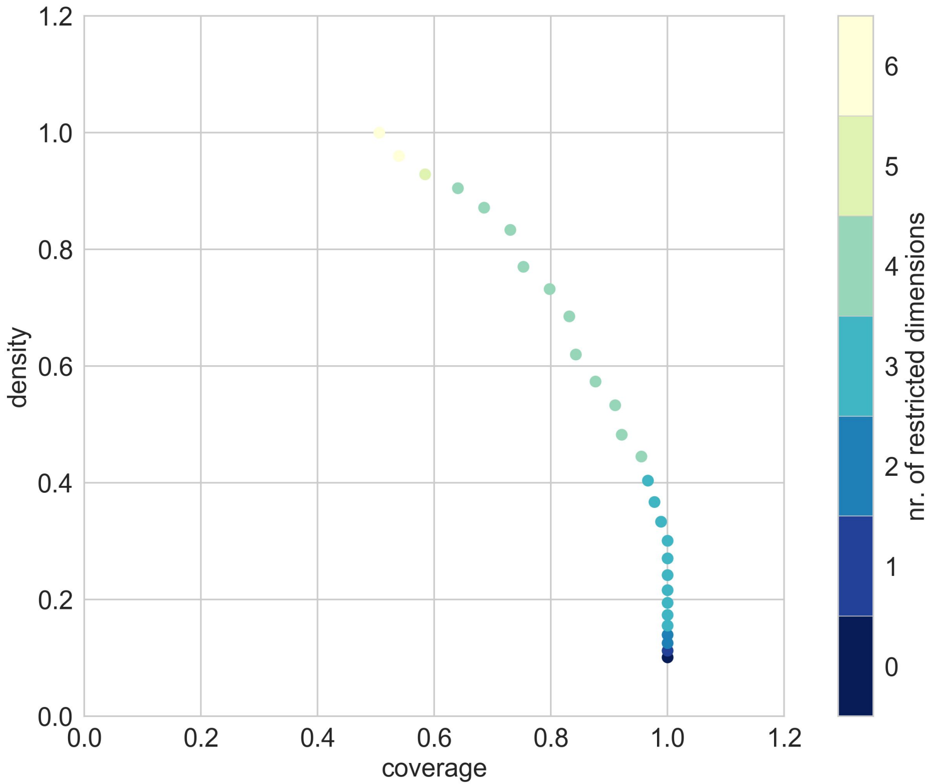


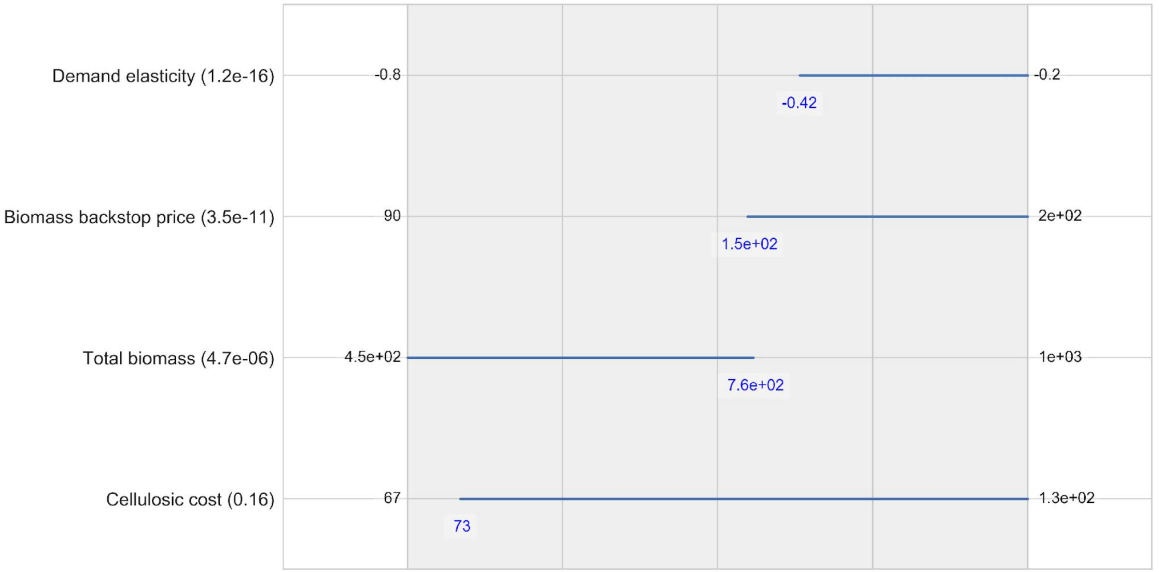
PRIM

- Lenient hill climbing optimization algorithm

- Each step in the optimization is stored → peeling trajectory

- **Coverage**: fraction of cases of interest within the box
- **Density**: fraction of cases in the box which is of interest
- Interpretability: number of restricted uncertainties
- **Quasi p-values**: one sided binomial test, proxy for statistical significance of each restricted uncertainty in isolation


density
coverage
nr. of restricted dimensions
0.0
0.2
0.4
0.6
0.8
1.0
1.2
0
1
2
3
4
5
6



coverage	0.753
density	0.77

Method 3 Sensitivity Analysis

Sensitivity analysis and scenario discovery

UA/SA: What impact do my uncertain inputs have on output?

$$f(x) = Y$$

SD: What are the uncertain inputs that cause an output (region) of interest?

- In practice: complementary
- SA and scenario discovery are usually both iterative

Sensitivity analysis in the modelling cycle

A for model analysis and evaluation

- Which uncertain inputs are the most influential on output?
- How much of the uncertainty is epistemic; how much is irreducible?
- Which uncertain inputs should be a priority for research?
- Can some inputs be left out to simplify the model?

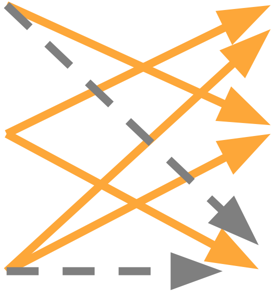
SA for policy design

- Can I intervene in the system, starting from the parameters to which the model is most sensitive?

Techniques for global sensitivity analysis

An ideal SA technique would be:

- Accurate
- Fast
- User-friendly



Linear regression

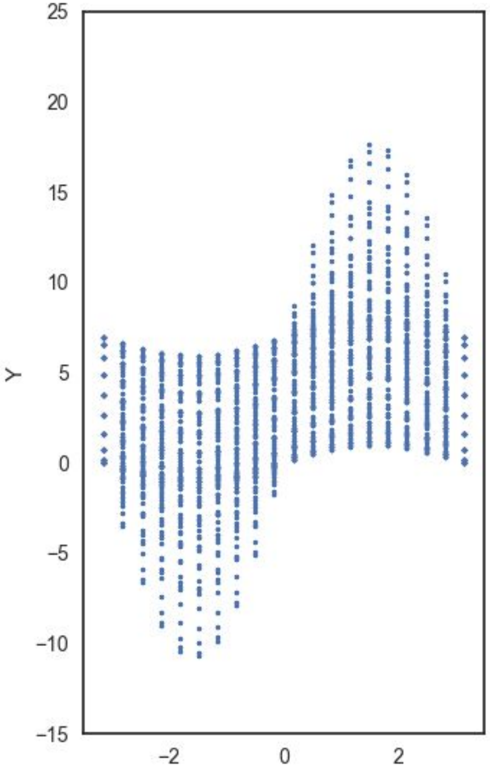
Sobol indices

Random forests /
Extra-Trees

Usually: pick two…

Interpretation of Sobol indices

- Based on variance decomposition – tells us the fraction of total variance added by each variable



x1

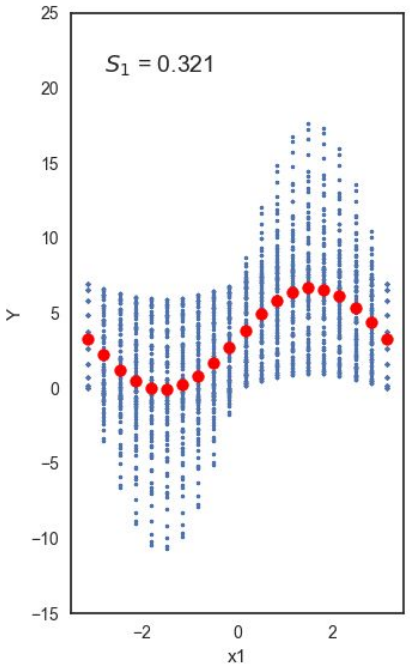
- First-order effect (S1): e.g. how much does x1 add to the variance of Y on its own?

$$S1_{x1} = V_{x1}\,[E_{X\sim x1}(Y|x1)]\;/\;V(Y)$$

V = variance
E = mean
$X{\sim}x_i$ = Set of inputs except x_i

Interpretation of Sobol indices

- Based on variance decomposition – tells us the fraction of total variance added by each variable



- First-order effect (S1): e.g. how much does x1 add to the variance of Y on its own?

$$S1_{x1} = V_{x1}\,[E_{X\sim x1}(Y|x1)]\ /\ V(Y)$$

- Total effect (ST): e.g. how much does x1 add to the variance of Y, including all its interactions?

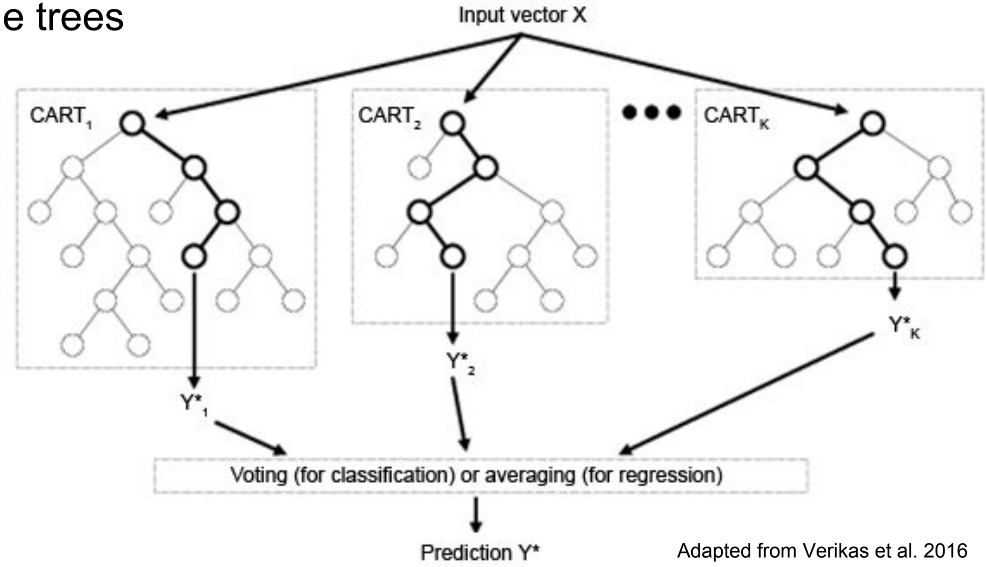
$$ST_{x1} = E_{X\sim x1}\,[V_{x1}(Y|X\sim x1)]\ /\ V(Y)$$

- Second-order effects (S2): e.g. how much specific
interactions between x1 and x2 add to variance of Y

$$S2_{x1,x2} = E_{X\sim x1,x2}\,[V_{x1,x2}(Y|X\sim x1,x2)]\ /\ V(Y) - S1_{x1} - S1_{x2}$$

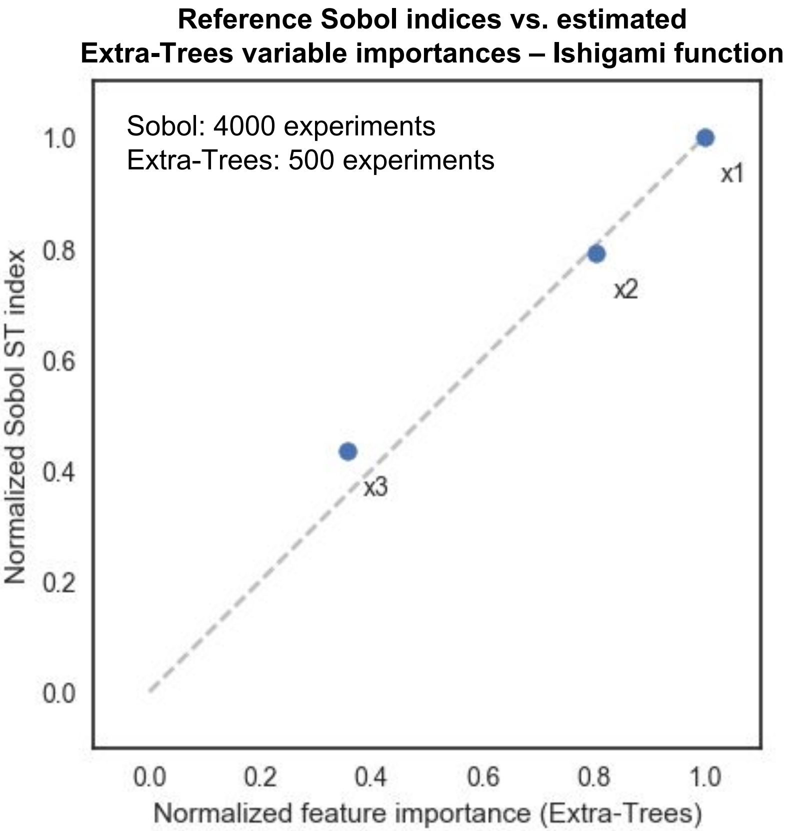
Random forests/Extra-Trees

- Common machine learning method for non-linear regression – based on ensembles (“forests”) of classification and regression trees
- Can be used to estimate variable importances (≈ Sobol ST)
- Extra-Trees (ET): variant with additional randomization of the trees



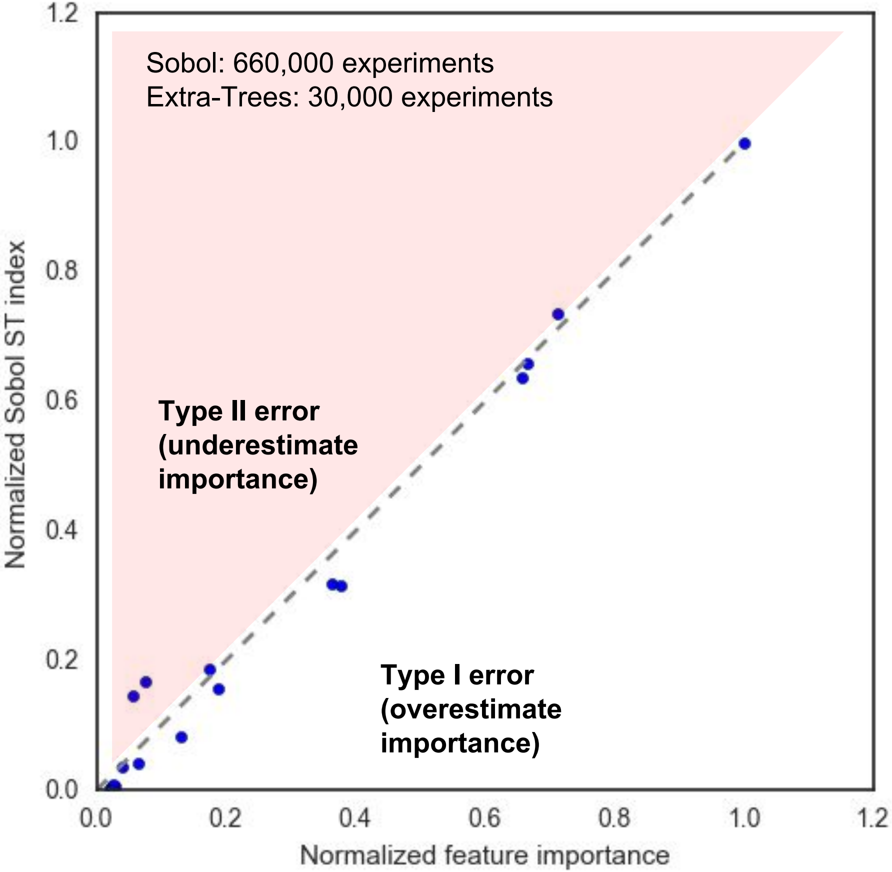
Adapted from Verikas et al. 2016

Estimation of relative variable importances

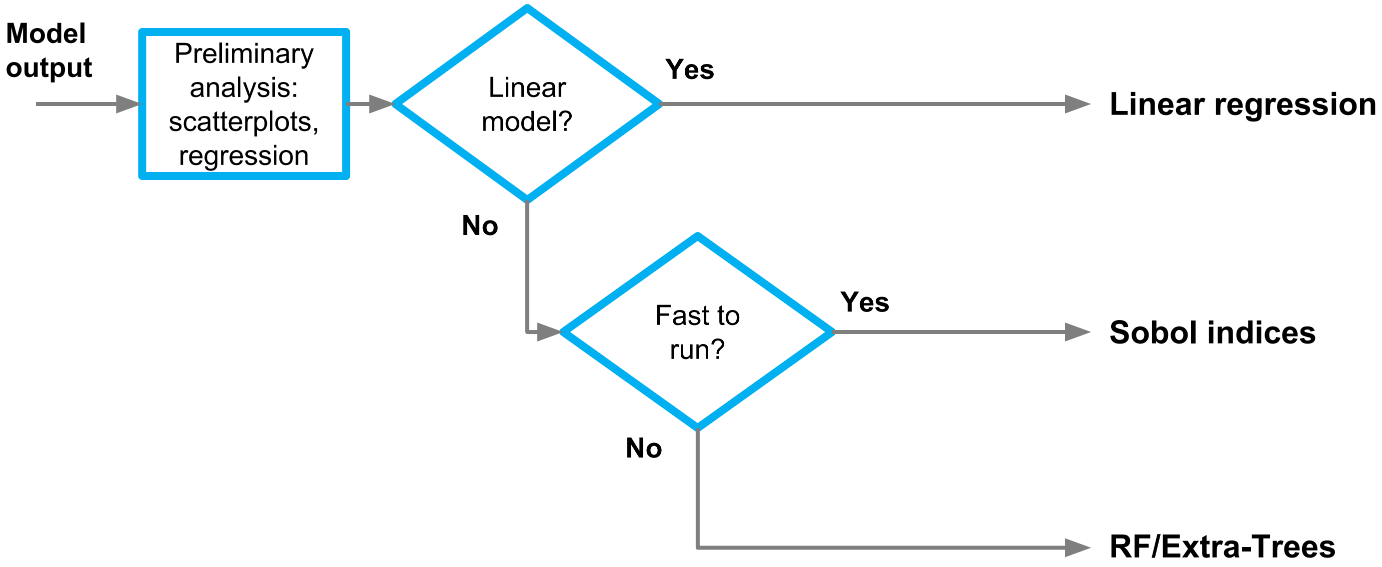


Estimation of relative variable importances

Reference Sobol indices vs. Extra-Trees variable importances – CDICE model



Summary: Sensitivity analysis techniques



Good job!