METAMODELING FOR SIMULATION-BASED OPTIMIZATION OF STRATEGIES IN HEALTHCARE

Short Course for the Society for Medical Decision Making

H. Koffijberg, PhD, MSc
University of Twente
Enschede, The Netherlands

K. Degeling, PhD, MSc
University of Melbourne
Melbourne, Australia

Prof.dr. H. (Erik) Koffijberg

UNIVERSITY OF TWENTE.

Professor – *Technology assessment of digital health innovations*

Chair: Department of Health Technology & Services Research (HTSR)
Faculty of Behavioural, Management and Social sciences, Technical Medical Centre
University of Twente, Enschede, The Netherlands

HTSR ~ 30 researchers, with expertise in Health Technology Assessment, Health Economics, Preference Research, Epidemiology, and Data Science.


HEALTH SERVICES RESEARCH


HEALTH PREFERENCE RESEARCH


TECHNOLOGY ASSESSMENT FOR HEALTH SYSTEMS ENGINEERING

https://www.utwente.nl/en/bms/htsr/

Prof.dr. H. (Erik) Koffijberg

UNIVERSITY OF TWENTE.

HTSR

Education:

- MSc in Technical Computer Science
- PhD in Decision-analytical modeling with focus on neurology

Experience:

- Modeling projects since 2004: applied and methodological
- Member of scientific advisory committee of the National Health Care Institute, ConVOI
- Lecturing on health economic modeling: Health Sciences, Industrial Engineering, SMDM, ISPOR

Research:

- Model-based assessment of the impact of new imaging tests, biomarkers, prediction models, AI, supporting personalized care strategies – in cardiovascular disease & oncology

Dr. K. (Koen) Degeling

Scientific Director, *Healthcare Consultancy Group*

Honorary Fellow, Cancer Health Services Research,
Centre for Cancer Research & Centre for Health Policy,
Faculty of Medicine, Dentistry and Health Sciences,
University of Melbourne

Dr. K. (Koen) Degeling

Education:

- PhD in Simulation Modeling to Optimize Personalized Oncology
- BSc & MSc in Industrial Engineering and Management
  Health Care Technology and Management

Experience:

- Building (simulations) models for health care since 2014
- Lecturing on health economic (simulation) modeling

Research:

- Methods for individual-level modeling
- Health economics in oncology, incl. diagnostics

Participants

- Where you are from
- What your background is
- What experience you have regarding
    - (Simulation) modeling
    - Metamodeling
    - The use of R
- Why you enrolled in the short course

Learning goals

After this course you will

1. understand the concept of metamodeling and when its use can support computationally challenging model-based analyses.
2. understand the steps and design choices necessary for developing metamodels.
3. know how to distinguish between alternative metamodeling techniques and between alternative design of experiments, and select an appropriate technique and design of experiments based on specific study characteristics.
4. be able to perform a simple metamodeling study in R.

Short course overview

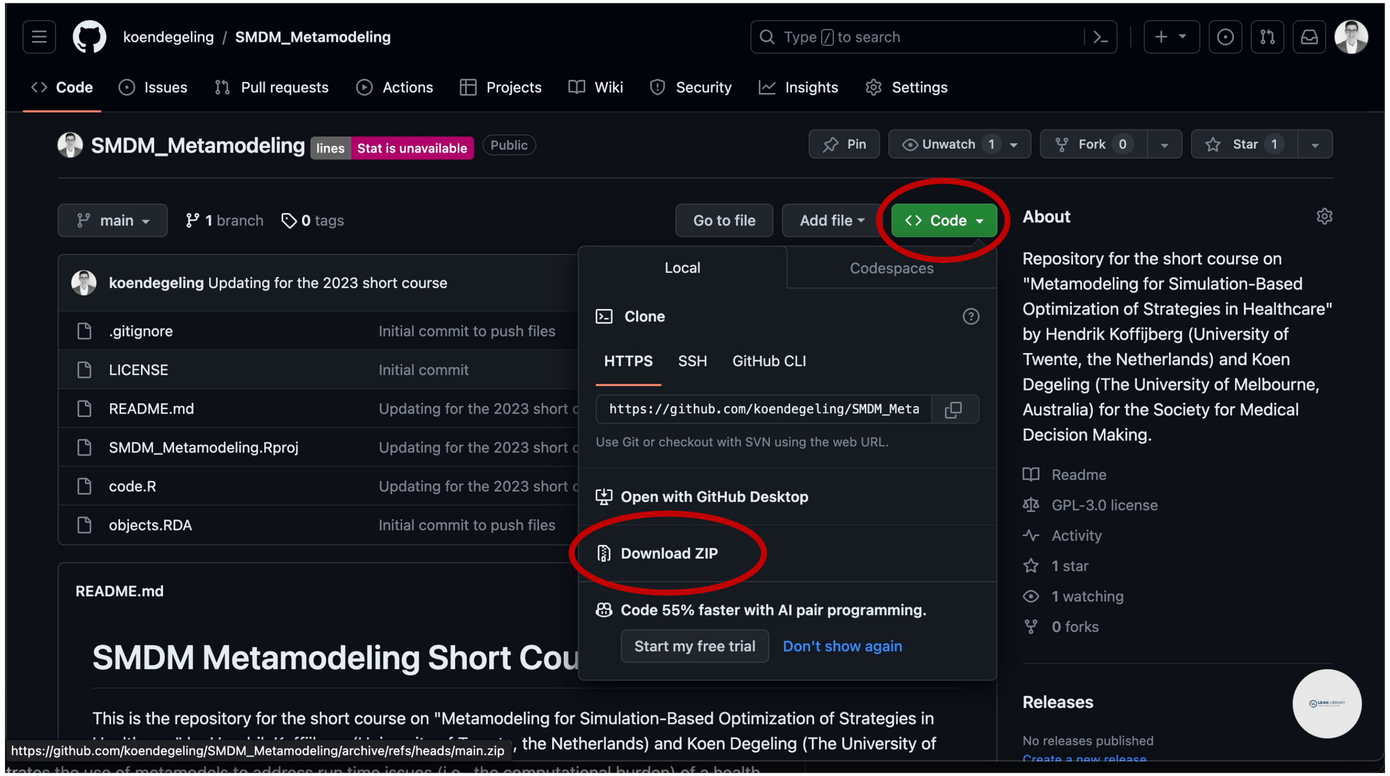
Block	Content
1	**Introduction and Course Overview**
2	**Content section 1:** Introduction to metamodeling and steps involved
3	**Content section 2:** Alternative metamodeling techniques
	BREAK *(10-15 minutes)*
4	**Breakout exercise:** Selecting a metamodeling technique
5	**Content section 3:** Alternative designs of experiments
6	**Content section 4:** Performance measures for assessing accuracy
	BREAK *(10-15 minutes)*
7	**Breakout exercise:** Performing a full metamodeling study
8	**Further Considerations**
9	**Overall Question & Answer** *(also questions throughout)*

Hands-on experience in R / R Studio

- Download files from GitHub:

  https://github.com/koendegeling/SMDM_Metamodeling

- Also see information in e-mail that was sent



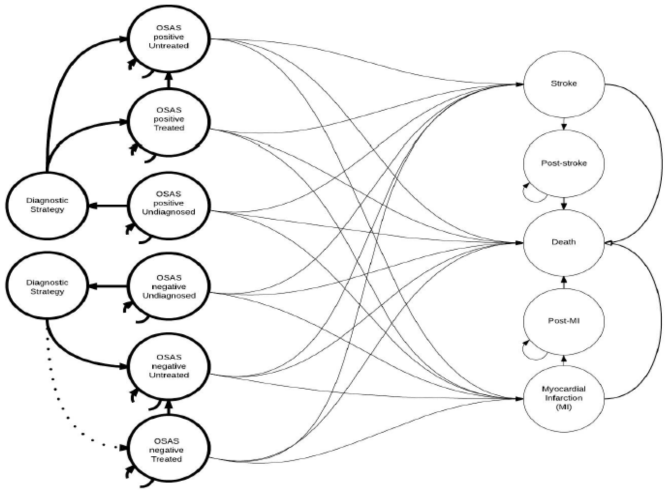
Content Section 1: Introduction to metamodeling and steps involved in using metamodeling

- Why do we need metamodeling?
- What is metamodeling?
- Use cases and examples
- Which steps to take?

Why do we need metamodeling?

- Value of health economic modeling has been recognized for over 25 years[1]
- Particularly relevant for evaluation of tests and biomarkers, and devices
  - *Test-treatment trials are quite rare*[2]
    - *Limited follow-up duration and high costs*
  - *Too many new tests to all directly evaluate in trials*
  - *Trials include only a limited number of comparators*
  - *Trial results reflect a highly controlled setting*

Decision analytic model



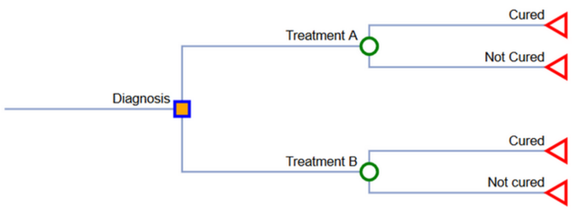
1 Buxton et al. Modeling in economic evaluation: an unavoidable fact of life. Health Economics. Vol 6: 217–227. 1997
2 Ferrante di Ruffano et al. A capture-recapture analysis demonstrated that randomized controlled trials evaluating the impact of diagnostic tests on patient outcomes are rare. J Clin Epi 2012;65:282–287.

Why do we need metamodeling?

- Health economic decision analytic models are
  - Relatively cheap and fast to develop
  - Reusable when new evidence becomes available
  - Able to incorporate multiple strategies/comparators

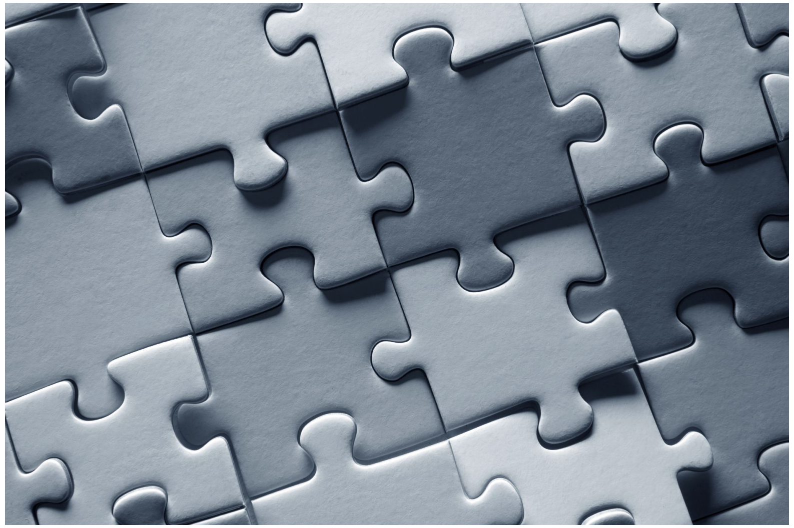
- Can be quite straightforward
  - Decision tree and Markov cohort models

Decision analytic model



Why do we need metamodeling?

- However, model complexity is increasing due to, amongst others
  - Desire to make full use of increasingly available detailed patient level data
  - Necessity to reflect increasingly complex clinical pathways
  - Need to evaluate personalized and precision medicine interventions
  - Desire to include (shared) decision making processes



Why do we need metamodeling?

- Test evaluation typically requires more complex models compared to drugs
  - *Indirect impact on treatment, costs & health outcomes*
  - *Proper incorporation of impact of incorrect test outcomes*
  - *Aspects of timing, threshold values, patient selection etc...*

- A checklist published in 2018 identified 44 potentially relevant aspects, in 6 major categories, to include in health economic models for test evaluation

Original Article

MDM
Medical Decision Making

Medical Decision Making
2018, Vol. 38(7) 778–788

DOI: 10.1177/0272989X18797590
journals.sagepub.com/home/mdm
SAGE

Toward Alignment in the Reporting of Economic Evaluations of Diagnostic Tests and Biomarkers: The AGREEDT Checklist

Michelle M.A. Kip, Maarten J. IJzerman, Martin Henriksson, Tracy Merlin, Milton C. Weinstein, Charles E. Phelps, Ron Kusters, and Hendrik Koffijberg

Kip MMA et al. Towards alignment in the reporting of economic evaluations of diagnostic tests and biomarkers: The AGREEDT checklist. MDM. 2018 38(7) 778-788.

Why do we need metamodeling?

- This means even (standard) probabilistic analysis may require huge computational resources
  - Detailed subgroup analyses

- Furthermore, more advanced model-based analyses are of increasing interest, for example
  - Model calibration algorithms
  - Value of information analyses
  - Optimization of interventions or strategies
    - *For tests: timing, threshold values, combinations,…*
    - *For screening: start/stop age, interval, subgroup-specific threshold values,*

Practical Illustration: Demonstrating how easily computational burden becomes an issue

Why do we need metamodeling?

Consider the computational burden when using a 'Simple patient-level discrete time state transition model in Excel'

Simple question: *which model parameters are might be relevant to collect additional evidence on to reduce decision uncertainty (VOI metric EVPPI)?*

- Typical PC takes ~1 min to run PA 500 samples, 2,000 patients
- ...need ≥ 20,000 patients for stable results (remove 1st order uncertainty) **x 10**
- ... need ≥ 5,000 Monte Carlo samples for proper PA **x 10**
- ... need ≥ 5,000 outer loop simulations (nesting for EVPPI) **x 5,000**
- ... need (10x10x5,000)/(60x24) ~ **347** days for a single EVPPI estimate

Yes, we should use R and not Excel, but we also want more complex metrics (e.g. EVSI)

Yes, approximation algorithms are increasingly available, but these are not under our control

We need methods to speed up this calculation

Why do we need metamodeling?

Consider the computational burden when using a 'Complex patient-level discrete time state transition model in R' to optimize a testing strategy

Simple question: *which strategy is optimal in terms of NMB, considering test sequence and age group specific test thresholds for referral?*

- Three tests, 4 age groups (40-50 to 80-90), 25 discrete threshold values
- Typical PC takes ~1 min to run PA 500 samples, 20,000 patients **speedup**
- … need ≥ 5,000 Monte Carlo samples for proper PA **x 10**
- … have 6x(3x4)x25 = 1,800 possible strategies **x 1,800**
- … need (10x1,800)/(60x24) ~ **12.5** days to evaluate

And one run is never sufficient…

Why do we need metamodeling?

Metamodeling is one way to reduce the computational burden of simulation-based optimization of strategies in healthcare

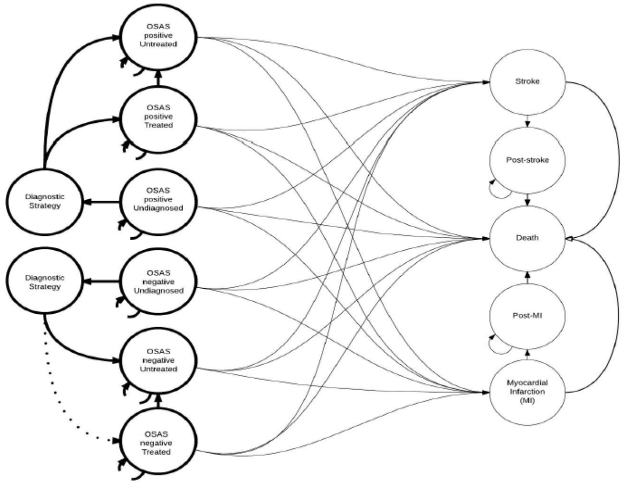
What is metamodeling?

A model of a model

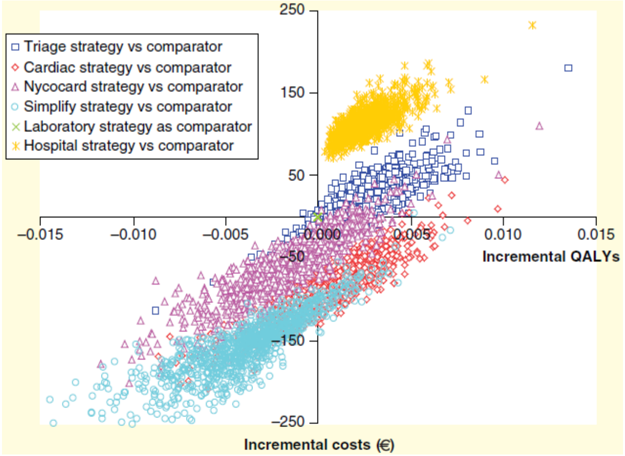
Input parameters (X)

Parameters	Base Case Value	Standard Deviation	Distribution	Source
Cohort characteristics				
Sex	Male			
Age	50			
OSAS prevalence (AHI≥15)	43%	0.06	Beta distribution	DiagnOSAS pilot*
Test characteristics				
Sensitivity PSG	1	Fixed		Gold Standard
Specificity PSG	1	Fixed		Gold Standard
Sensitivity DiagnOSAS	0.97	0.03	Beta distribution	DiagnOSAS pilot*
Specificity DiagnOSAS	0.47	0.07	Beta distribution	DiagnOSAS pilot*
Average time to diagnosis	5.35 (yr)	Fixed		[6]
Scenario time to diagnosis	2.68 (yr)	Fixed**		[6]
Probabilities				
Car accident	0.0461	0.00006	Beta distribution	[19,20]
Probability car accident is fatal	0.00035	0.00003	Beta distribution	[19,28]
Stroke (per 1000 person-yrs)	1.30	0.08	Normal distribution	[21]
MI (per 1000 person-yrs)	2.66	0.02	Normal distribution	[22]
Adjusted all-cause mortality (age- & gender-specific)	0.000574	Fixed		[15-18]
CPAP non-adherence (annual)	0.023	0.003	Normal Distribution	[27]
OSAS related Risks				
Untreated OSAS	*Compared to healthy individuals*			
Car accident (Relative Risk)	2.43	0.36	Log.Norm Distribution	[14]
Stroke (Hazard Ratio)	3.48	0.03	Log.Norm Distribution	[23]
MI (Hazard Ratio)	3.06	0.08	Log.Norm Distribution	[23]
CPAP-treated OSAS	*Compared to healthy individuals*			
Car accident (Risk Ratio)	1.29	0.43	Log.Norm Distribution	[26]
	Compared to untreated OSAS			
MI (Relative Risk)	0.54	0.17	Log.Norm Distribution	[24]
Stroke (Relative Risk)	0.27	0.10	Log.Norm Distribution	[25]
Stroke				
1-year case fatality (Probability)	0.21	0.04	Normal Distribution	[30]
All-cause mortality (Hazard ratio)	3.9	0.10	Log.Norm Distribution	[31]
MI				
1-year case fatality (Probability)	0.07	0.005	Normal Distribution	[29]
All-cause mortality (Hazard ratio)	1.47	0.16	Log.Norm Distribution	[32]

Decision analytic model (simulator)



Output estimates (Y)

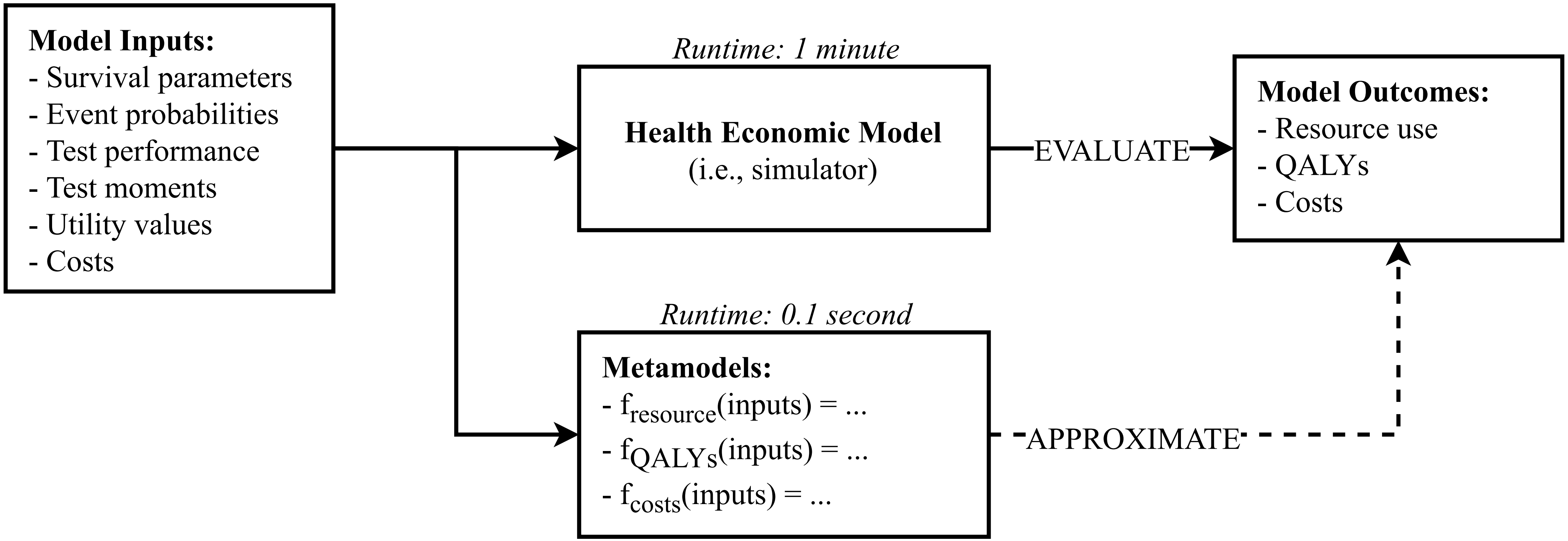


Metamodel

***F*(X) = Y**

What is metamodeling?

More concrete



What is metamodeling?

Terminology:

- Metamodel, surrogate model, emulator

Aim:

- Reduce runtime issues of complex models and analyses, by approximating the outcome of computationally demanding models within feasible time

Classification:

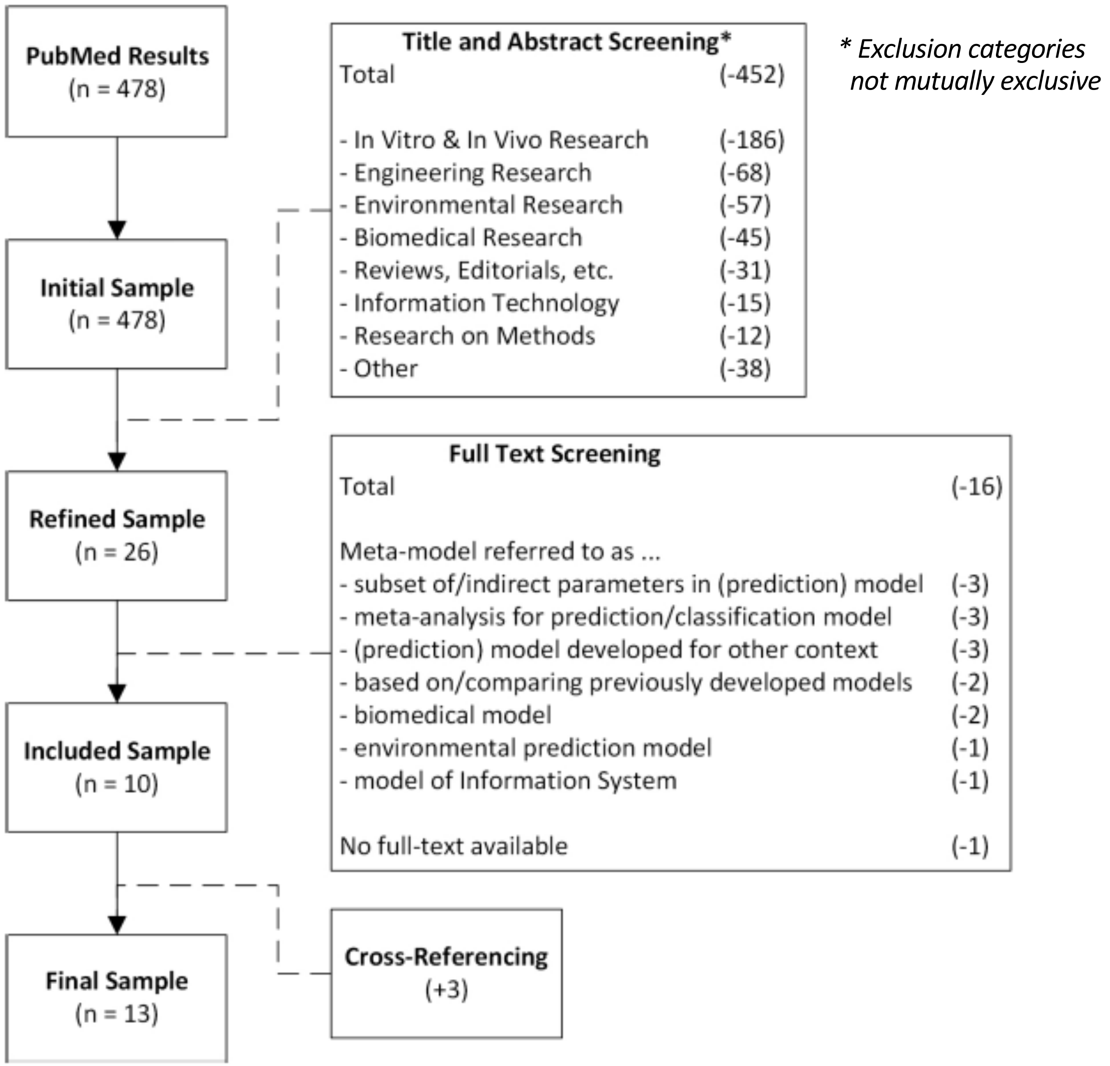
- *Statistical approach*, if used to obtaining insights into relations between simulator inputs and outputs
- *Simulation approach*, if used as substitute to perform additional simulation-based analyses

Use cases and examples of metamodeling

Widely used in some fields:

- Computer science, mechanical and aerospace systems engineering
- For high-fidelity engineering design, e.g. Computation Fluid Dynamics, Computational Structural Dynamics

Scoping review of meta-modelling applications and opportunities for advanced health economic analyses (2018)



Degeling K et al. A scoping review of metamodeling applications and opportunities for advanced health economic analyses. Expert Rev Pharmacoecon Outcomes Res. 2018 Nov 14:1-7.

Use cases and examples of metamodeling

First Author	Year of Publication	Journal of Publication	Clinical Context	Type of Study	Simulator Type
Merz	1992	Medical Decision Making	Deep vein thrombophlebitis	Health Economic Modeling	Decision Tree
Tappenden	2004	Health Technology Assessment	Multiple Sclerosis	Health Economic Modeling	Cohort State-transition
Stevenson	2004	Medical Decision Making	Osteoporosis	Health Economic Modeling	Microsimulation State-transition Model
Woodroffe	2005	Health Technology Assessment	Renal Transplantation	Health Economic Modeling	Microsimulation (unknown structure)
Rojnik	2008	Value in Health	Breast Cancer	Health Economic Modeling	Markov Cohort Model
Jalal	2013	Medical Decision Making	NA (Fictitious)	Health Economic Modeling	Markov Cohort Model
Willem	2014	PLoS Computational Biology	Influenza & Varicella	Infectious Disease Modeling	Agent-Based Model
Jalal	2015	Medical Decision Making	NA (Fictitious)	Health Economic Modeling	Decision Tree & Markov Cohort Model
Andrianakis	2015	PLoS Computational Biology	HIV	Infectious Disease Modeling	Agent-Based Model
Angus	2016	European Journal of Public Health	Excessive Alcohol Consumption	Health Economic Modeling	Mathematical Model (unknown structure)
Jutkowitz	2017	Pharmacoeconomics	Gout, Inflammatory Arthritis	Health Economic Modeling	Markov Cohort Model
Yousefi	2018	Artificial Intelligence in Medicine	Emergency Department	Health Care Logistics	Agent-Based Model
Jalal	2018	Medical Decision Making	NA (Fictitious)	Health Economic Modeling	Markov Cohort Model

Degeling K et al. A scoping review of metamodeling applications and opportunities for advanced health economic analyses. Expert Rev Pharmacoecon Outcomes Res. 2018 Nov 14:1-7.

Use cases and examples of metamodeling

Statistical approach

- Simplified example model
- Regressing input parameters on outcomes and using the metamodel for threshold analyses

Linear Regression Metamodeling as a Tool to Summarize and Present Simulation Model Results
Hawre Jalal, Bryan Dowd, François Sainfort and Karen M. Kuntz
Med Decis Making 2013 33: 880 originally published online 27 June 2013
DOI: 10.1177/0272989X13492014

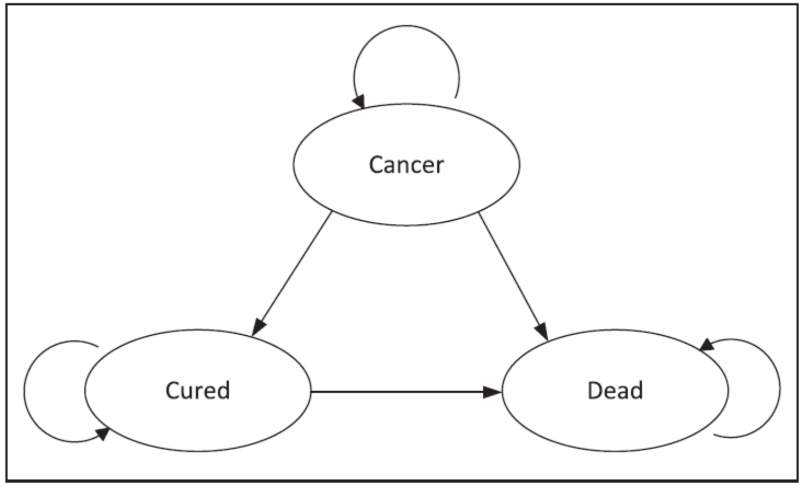


Table 4 Results of Regressing the Standardized Parameters and Their Interactions on the ΔNHB

Parameter	Chemo> Radio	Chemo> Surgery	Radio> Surgery
Intercept	0.320	0.120	−0.200
pFailChemo	−0.384	−0.384	0.000
pFailRadio	0.466	0.000	−0.466
pFailSurg	0.000	0.234	0.234
pDieSurg	−0.001	0.446	0.447
μCancer	0.054	−0.484	−0.538
cChemo	−0.041	−0.041	−0.001
cRadio	0.068	0.001	−0.067
cSurg	0.001	0.206	0.205
μCancer*pFailChemo	−0.065	−0.062	0.003
μCancer*pFailRadio	0.061	0.002	−0.059
μCancer*pFailSurg	0.000	0.017	0.017

Note: NHB = net health benefit.

Jalal H et al. Linear Regression Metamodeling as a Tool to Summarize and Present Simulation Model Results. Med Decis Making 2013 33: 880

Use cases and examples of metamodeling

Simulation approach

- Bayesian calibration in health decision sciences is challenging: program complex models with associated computational burden of applying Bayesian calibration
- BayCANN only uses a dataset of model inputs/outputs to obtain calibrated joint parameter distributions. It can be adapted to models of various levels of complexity with minor or no change to its structure. Its efficiency can be especially useful in computationally expensive models.

frontiers in Physiology

METHODS
published: 25 May 2021
doi: 10.3389/fphys.2021.662314

BayCANN: Streamlining Bayesian Calibration With Artificial Neural Network Metamodeling

Hawre Jalal[1], Thomas A. Trikalinos[2] and Fernando Alarid-Escudero[3]*

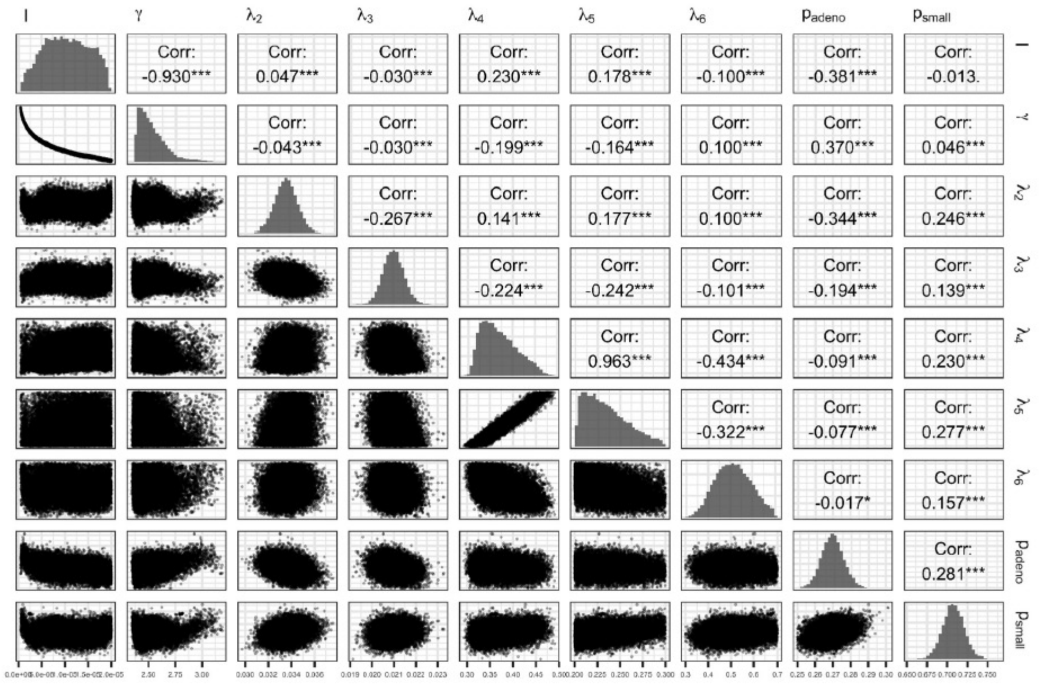


FIGURE 7 | Joint posterior distribution of the calibrated parameters of the case study using the ANN method.

Jalal H, et al. BayCANN: Streamlining Bayesian Calibration With Artificial Neural Network Metamodeling. Front Physiol. 2021;12:662314.

Use cases and examples of metamodeling

Simulation approach

- Microsimulation models are used extensively in cancer modeling
- Substantial uncertainty regarding estimates from these models is usually not thoroughly examined due to the high computational effort required
- *Objective*: To quantify uncertainty in model outcomes due to uncertainty in model parameters, using a computationally efficient emulator (Gaussian process regression) instead of the model.

Original Article

MDM
Medical Decision Making

Medical Decision Making
1–9
© The Author(s) 2019
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/0272989X19837631
journals.sagepub.com/home/mdm
SAGE

Evaluating Parameter Uncertainty in a Simulation Model of Cancer Using Emulators

Tiago M. de Carvalho, Eveline A. M. Heijnsdijk, Luc Coffeng, and Harry J. de Koning

"… instead of running MISCAN 1000 times, we would run MISCAN 100 times (to obtain data for the training of the emulator), plus about the computing time equivalent of 2 MISCAN runs. If we carry out an additional 30 runs for validation … this procedure will result in a reduction of more than 85% in computation time."

De Carvalho TM et al. Evaluating Parameter Uncertainty in a Simulation Model of Cancer Using Emulators. Med Decis Making. 2019 May;39(4):405-413.

Use cases and examples of metamodeling

Simulation approach

- Estimate the EVPI & EVPPI using a metamodel in the form of a Gaussian process
- Implemented and available online: https://savi.shef.ac.uk/SAVI/

Estimating Multiparameter Partial Expected Value of Perfect Information from a Probabilistic Sensitivity Analysis Sample: A Nonparametric Regression Approach

Mark Strong, PhD, Jeremy E. Oakley, PhD, Alan Brennan, PhD

Conclusion (copied)

"...With the increasing use of patient-level micro-simulation models, we envisage that obtaining partial EVPI via the traditional 2-level Monte Carlo approach will be considered just too time-consuming (in fact, experience suggests that the 2-level Monte Carlo procedure is considered too difficult for even moderately simple cohort models). In contrast, the regression methods we have presented provide a mechanism for rapidly estimating partial EVPI for any set of parameters in a model of any complexity."

Strong M et al. Estimating multiparameter partial expected value of perfect information from a probabilistic sensitivity analysis sample: a nonparametric regression approach. Med Decis Making. 2014 Apr;34(3):311-26.

Use cases and examples of metamodeling

Simulation approach

- *Goal*: To illustrate the potential advantages of using a metamodel to identify the optimal screening strategy for colorectal cancer (CRC), accounting for colonoscopy capacity constraints

ELSEVIER

Value in Health

ScienceDirect

Contents lists available at **sciencedirect.com**
Journal homepage: **www.elsevier.com/locate/jval**

Economic Evaluation

Using Metamodeling to Identify the Optimal Strategy for Colorectal Cancer Screening

Hendrik Koffijberg, PhD, Koen Degeling, PhD, Maarten J. IJzerman, PhD, Veerle M.H. Coupé, PhD, Marjolein J.E. Greuter, PhD

Leading short course example

Koffijberg H et al. Using Metamodeling to Identify the Optimal Strategy for Colorectal Cancer Screening.
Value Health. 2021 Feb;24(2):206-215.

Metamodeling to optimize CRC screening strategies

Common options for the analysis of screening strategies

1. Define and evaluate all possible strategies
    - Unlikely to be feasible in acceptable time frames for complex patient-level models including probabilistic analysis
    - Hard to report all results/comparisons

2. Evaluate a limited predefined set of strategies
    - Predefined set unlikely to include all potentially relevant strategies
    - Best strategy evaluated may be far from optimal

O'Mahony JF, et al. Beware of Kinked Frontiers: A Systematic Review of the Choice of Comparator Strategies in Cost-Effectiveness Analyses of Human Papillomavirus Testing in Cervical Screening. Value Health. 2015 Dec;18(8):1138-51.

Metamodeling to optimize CRC screening strategies

Focus: Primary screening program for CRC in the Netherlands

- Referral based on fecal immunochemical test (FIT)
- Currently available capacity for colonoscopies after referral is ~550 colonoscopies per 1,000 individuals lifelong
- The validated ASCCA model* was developed to evaluate screening strategies in terms of life years gained (LYG) and costs, compared to no screening
- Model programmed in C++, runtime still approx. 15 min/strategy due to internal calibration to Dutch observational data, without PA...

* Greuter MJ et al. Modeling the Adenoma and Serrated pathway to Colorectal CAncer (ASCCA). Risk Anal. 2014 May;34(5):889-910.

Metamodeling to optimize CRC screening strategies

Basic relevant screening characteristics (parameters)

- Starting age (years, 30-90)
- Screening interval (years, 1-60)
- Number of screening rounds (1-30)
- FIT cutoff (discrete: 50, 75, 100, 150)

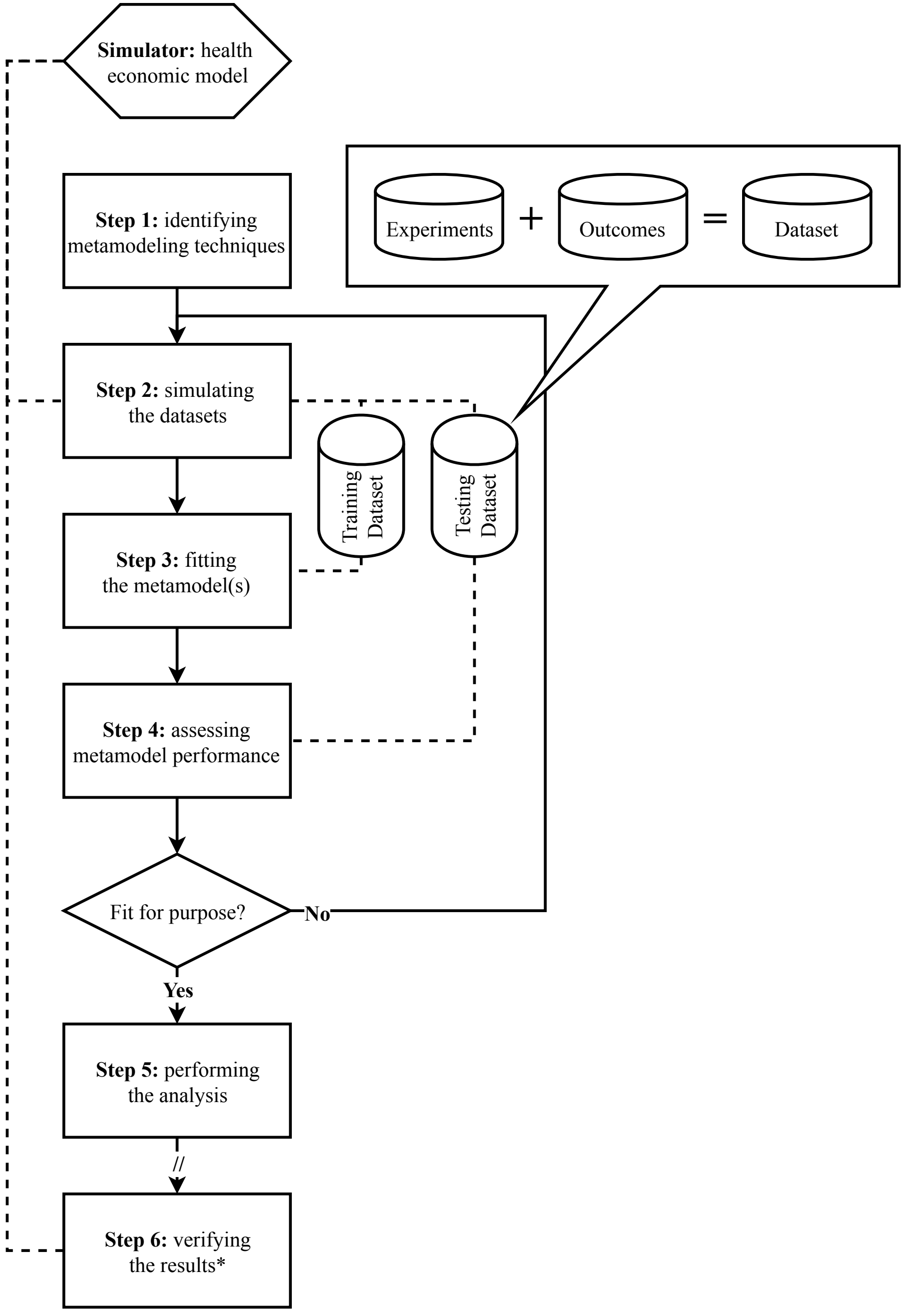
Practical constraint: max screening age = 90 yrs

- Number of strategies > 450,000, plausible = **40,864**
- Evaluation of all plausible strategies with ASCCA model unfeasible
  - A metamodel is needed to allow optimization!

Koffijberg H et al. Using Metamodeling to Identify the Optimal Strategy for Colorectal Cancer Screening.
Value Health. 2021 Feb;24(2):206-215.

Metamodeling process

- Starting point: validated model
- Metamodeling techniques
- Training and testing datasets
- Fitting metamodels
- Metamodel validation
- Update technique/data/specification
- Use the metamodel
- Verify the results



* Optional, depending on the analysis performed in Step 5.

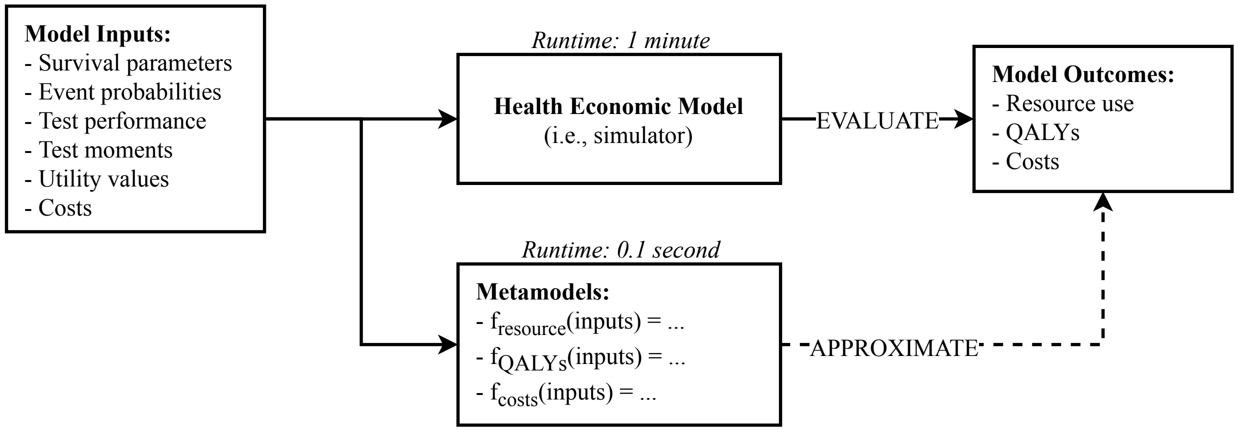
Degeling et al. (2020) Introduction to Metamodeling for Reducing Computational Burden of Advanced Analyses with Health Economic Models: A Structured Overview of Metamodeling Methods in a 6-Step Application Process. *Med Decis Making.*

Content Section 2: Alternative metamodeling techniques

- What metamodeling techniques are available?
- What are the characteristics of these techniques?
- When might these techniques be suitable for use?

Metamodeling techniques

- Simple linear regression
- Response surface methodology
- Symbolic regression
- Multivariate adaptive regression splines
- Generalized additive models
- Gaussian processes
- Neural networks

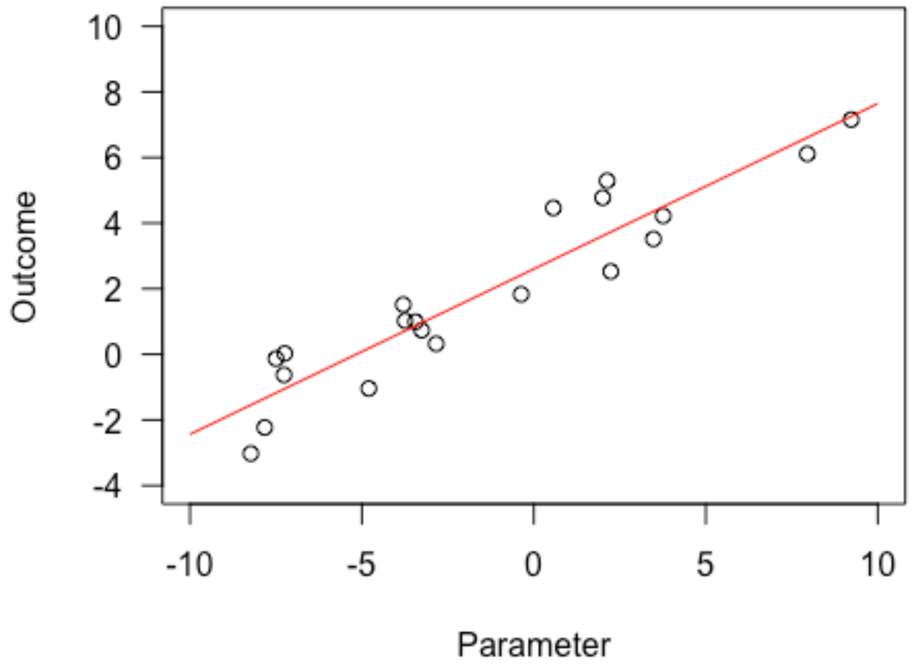


- Selection:
  - Required number of experiments
  - Number of inputs/parameters
  - Interpretability

Simple linear regression

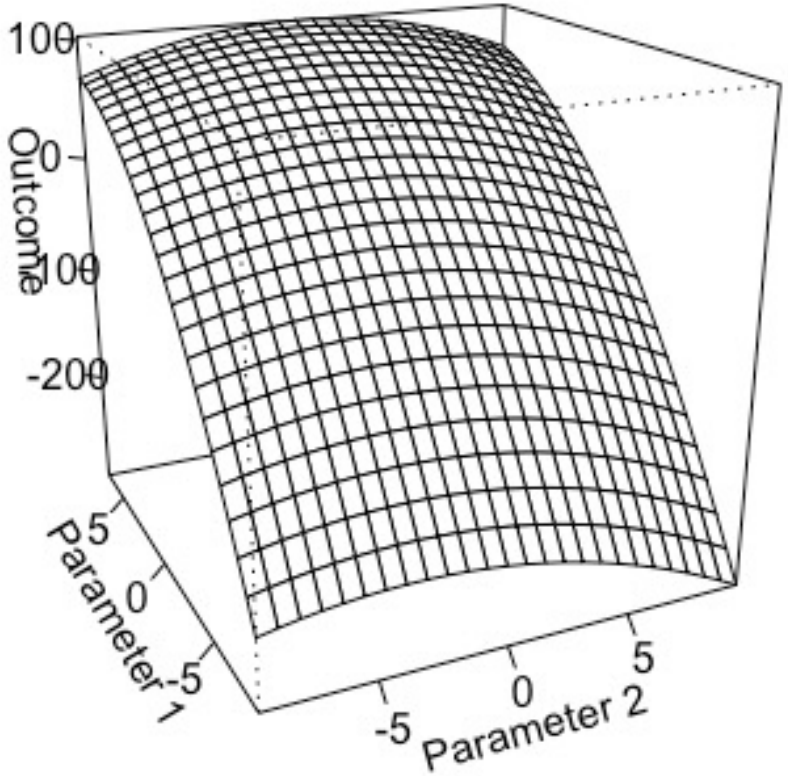
- Assumes a linear relationship between independent variables (i.e., input parameters) and the dependent variable (i.e., outcome of interest) and is linear in the regression model parameters

$$y = \beta_0 + \beta_1 X + \varepsilon$$



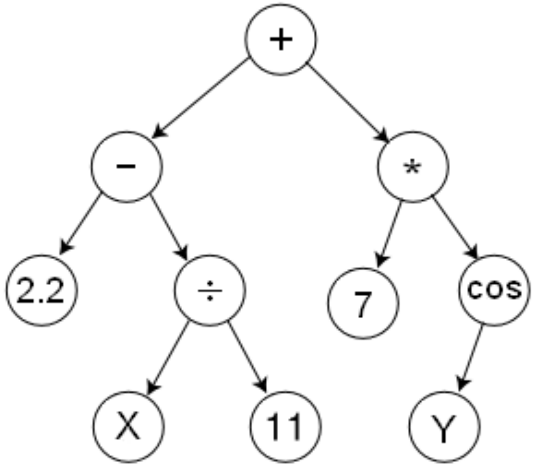
Response surface methodology

- Also linear in the regression model parameters but does not assume a linear input-output relationship, and it fits polynomial regression models to predict responses (i.e., outcomes)



Symbolic regression

- Uses genetic programming to construct a mathematical expression from elementary operators (e.g., "+" and "×") and elementary functions (e.g., "log"), accurately describing the relation between input parameters and the outcome of interest, without making any priori assumption about this relationship
- Can be challenging to fit because of large number of options

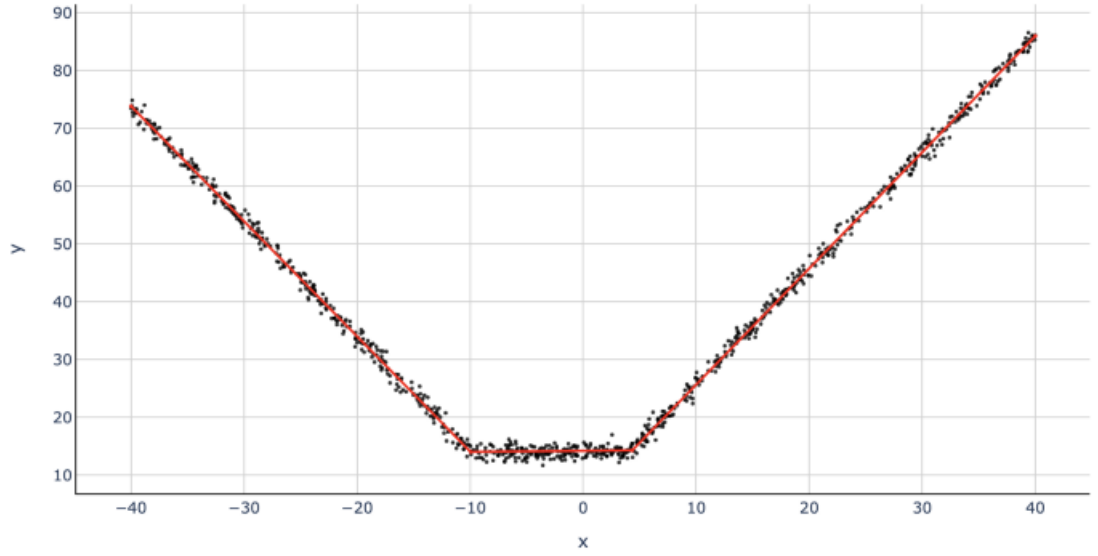


$$\left(2.2 - \left(\frac{X}{11}\right)\right) + \left(7 * \cos(Y)\right)$$

https://towardsdatascience.com/ml-approaches-for-time-series

Multivariate adaptive regression splines

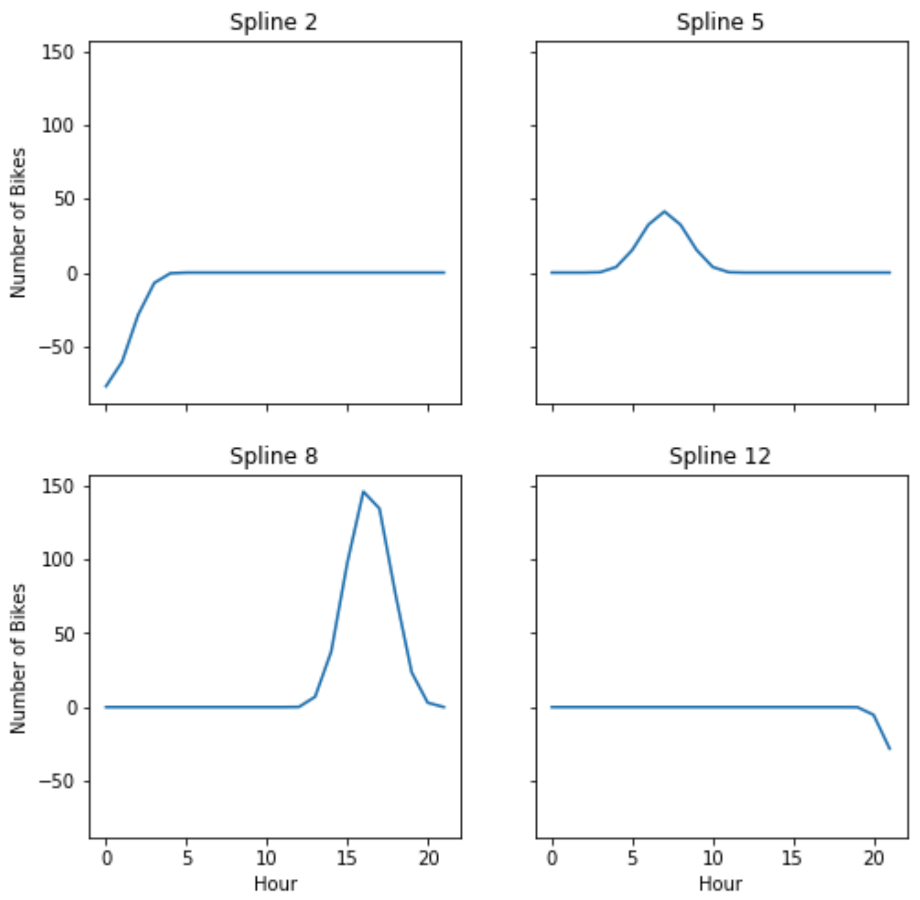
- Divide the outcome domain into intervals and then estimates an equation, typically a low-order polynomial, for each interval
- Different types of splines can be distinguished, based on how the number of intervals and level of smoothness are defined
- Prone to overfitting



https://towardsdatascience.com/mars-multivariate-adaptive-regression-splines-how-to-improve-on-linear-regression

Generalized additive models

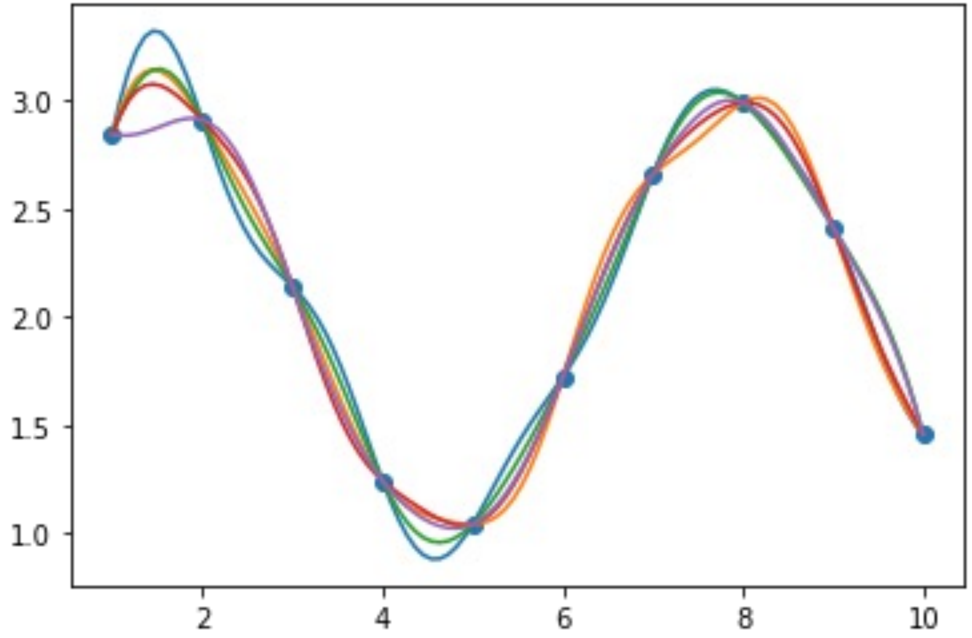
- Assume that the dependent variable is a smooth function of the independent variables, usually represented using splines
- Can also be represented as the weighted sum of a series of predetermined “basis functions” that extend over the whole range of the function input



https://towardsdatascience.com/generalised-additive-models

Gaussian processes

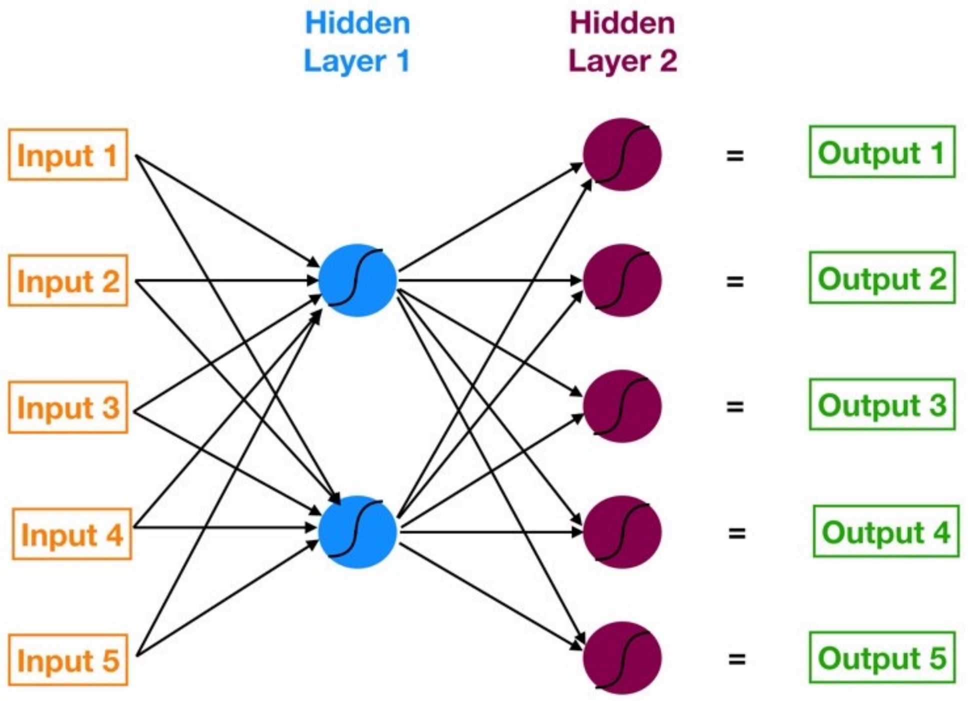
- Nonparametric regression method also known as Kriging, which uses information on neighbor experiments for new predictions while directly providing information on the uncertainty in these predictions



https://towardsdatascience.com/what-on-earth-is-a-gaussian-process

Neural networks

- Nonparametric models consisting of networks of nodes (called neurons) and layers, which learn about relationships between inputs and outputs, typically using large data sets
- Commonly used for classification, but also able to predict continuous outcomes

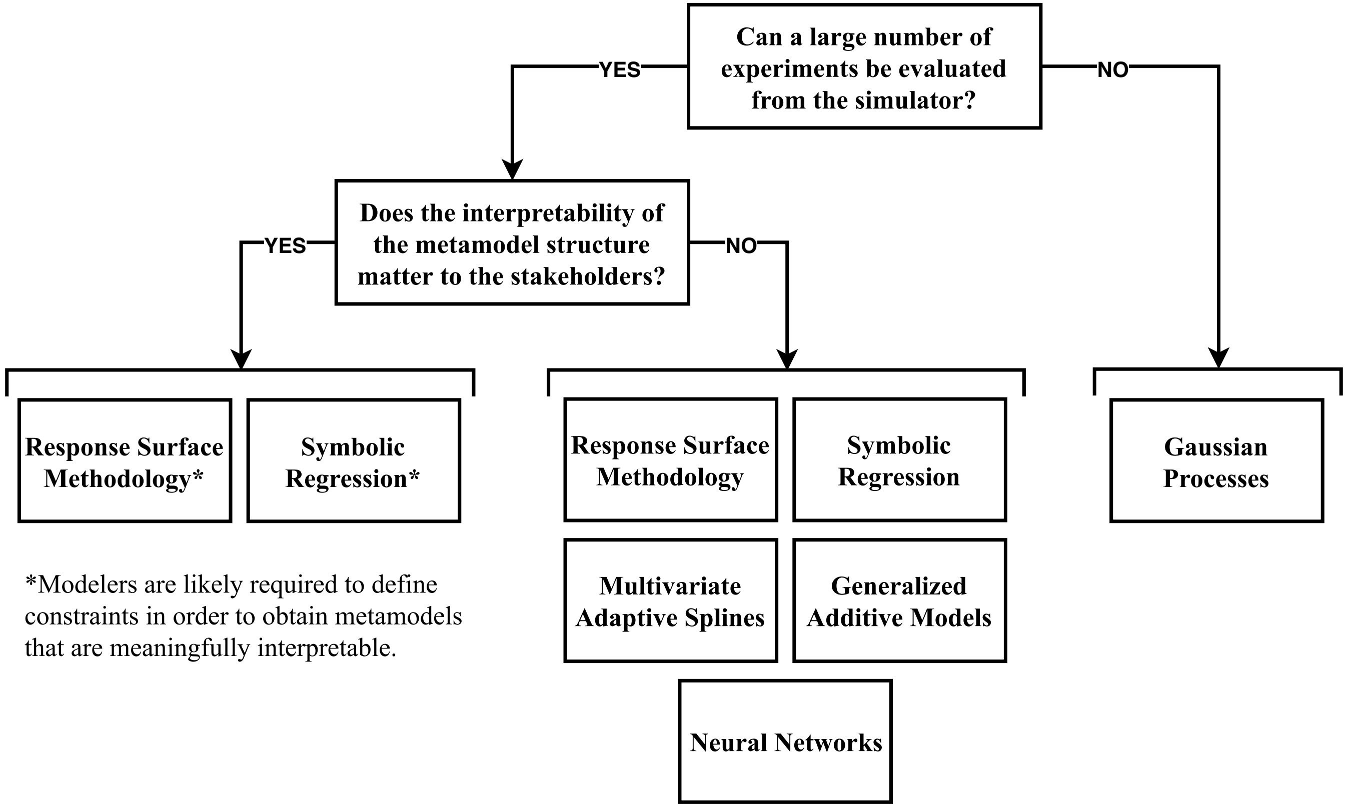


https://towardsdatascience.com/understanding-neural-networks

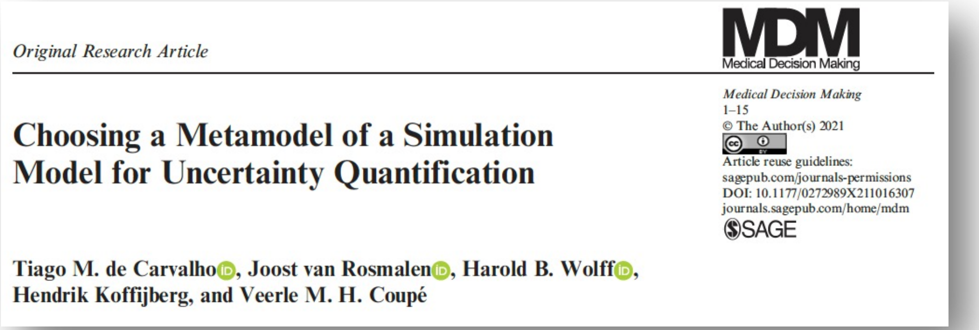
Overview of metamodeling techniques

Technique	(Required) Number of Experiments	Number of Inputs	Interpretability
Linear Regression	High	Large	High
Response Surface Methodology	High	Large	Moderate
Symbolic Regression	High	Large	Moderate
Multivariate Adaptive Regression Splines	High	Large	Low
Generalized Additive Models	High	Large	Low
Gaussian Processes	Low	Low	Low
Neural Networks	High	Large	Low

Selecting a metamodeling technique



*Modelers are likely required to define constraints in order to obtain metamodels that are meaningfully interpretable.



Carvalho et al. (2021) Choosing a Metamodel of a Simulation Model for Uncertainty Quantification. *Med Decis Making.*

Degeling et al. (2020) Introduction to Metamodeling for Reducing Computational Burden of Advanced Analyses with Health Economic Models: A Structured Overview of Metamodeling Methods in a 6-Step Application Process. *Med Decis Making.*


coffee
BREAK

Breakout Exercise: Selecting a metamodeling technique based on case study characteristics

What is an appropriate metamodeling technique?

Case study A

Goal: Generating insight into the factors influencing the impact (iNHB) of imposing a sugar tax on beverages compared with usual care.

Context: Simulation model can generate 1000s of input/output samples

- **A** Simple Linear Regression
- **B** Response Surface Methodology
- **C** Symbolic Regression
- **D** Multivariate Adaptive Splines
- **E** Generalized Additive Models
- **F** Neural Networks
- **G** Gaussian Processes

What is an appropriate metamodeling technique?

Case study B

Goal: Performing a probabilistic analysis to assess uncertainty in outcomes for a strategy imposing a sugar tax on beverages compared with usual care

Context: Simulation model can generate 1000s of input/output samples

A Simple Linear Regression

B Response Surface Methodology

C Symbolic Regression

D Multivariate Adaptive Splines

E Generalized Additive Models

F Neural Networks

G Gaussian Processes

What is an appropriate metamodeling technique?

Case study C

Goal: Assessing the budget impact of imposing a sugar tax on beverages compared with usual care based on a microsimulation of all Dutch citizens (~18 million)

Context: Simulation model can generate 100-200 input/output samples

- **A** Simple Linear Regression
- **B** Response Surface Methodology
- **C** Symbolic Regression
- **D** Multivariate Adaptive Splines
- **E** Generalized Additive Models
- **F** Neural Networks
- **G** Gaussian Processes

Content Section 3: Alternative designs of experiments for simulating training and testing data

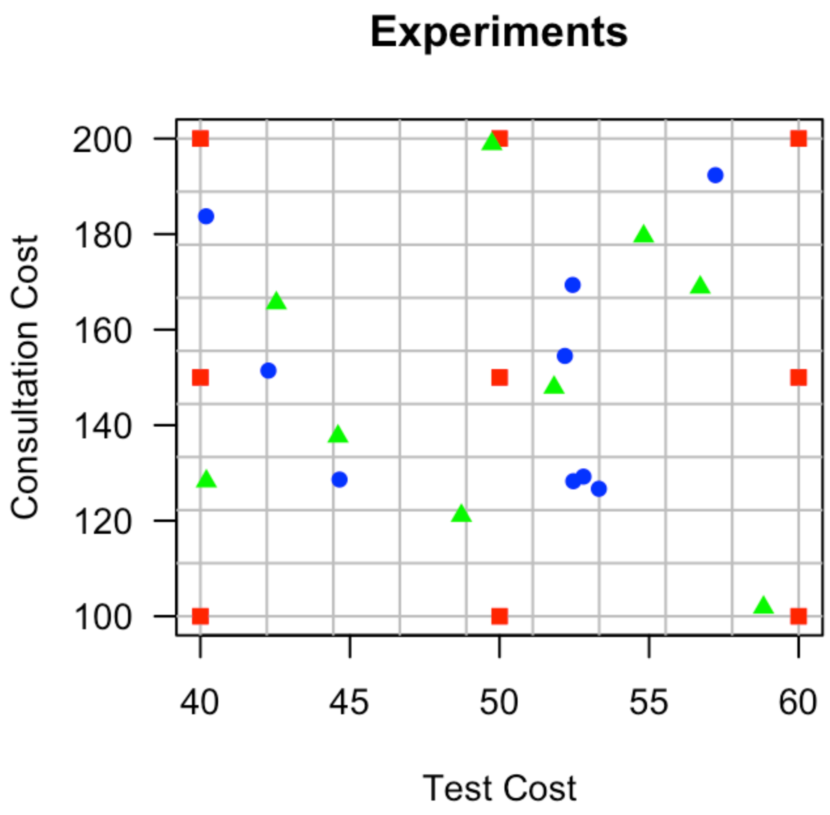
- Objective of experiment generation
- Different designs of experiments
- Parameter types and constraints
- When to use which design

Experiments are combinations of inputs and outcomes

Experiment	Parameter 1	Parameter 2	Parameter 3	Outcome
1	81	29	11	-822
2	53	49	20	124
3	69	44	24	146
4	46	19	8	-635
5	79	46	8	-341
6	49	39	21	202
7	44	23	13	-905
…	…	…	…	…

Design of experiments for training and testing

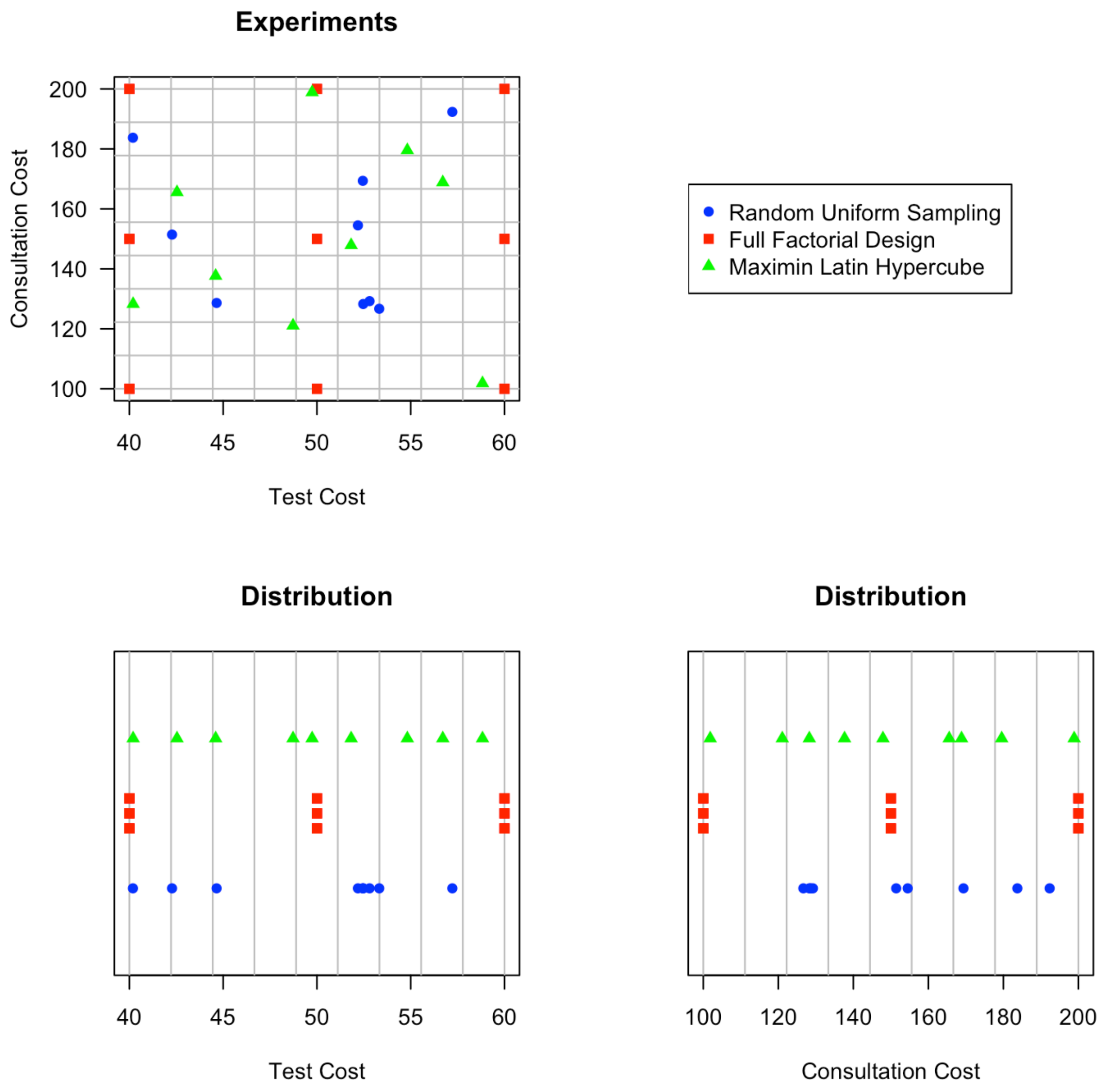
- Efficiently generating data is important, as running the simulator takes a lot of time (that is why we need a metamodel!)
- Experiments need to provide the maximum amount of information:
  - Input/parameter and output relations
  - Whole parameter range of interest
  - Interactions between parameters
- Three common designs:
  - Random sampling
  - Full factorial
  - Latin Hypercube sampling



Random sampling

- Randomly select values for the parameters independently of the other parameters

- Easy to apply, but inefficient – typically many samples required for good coverage

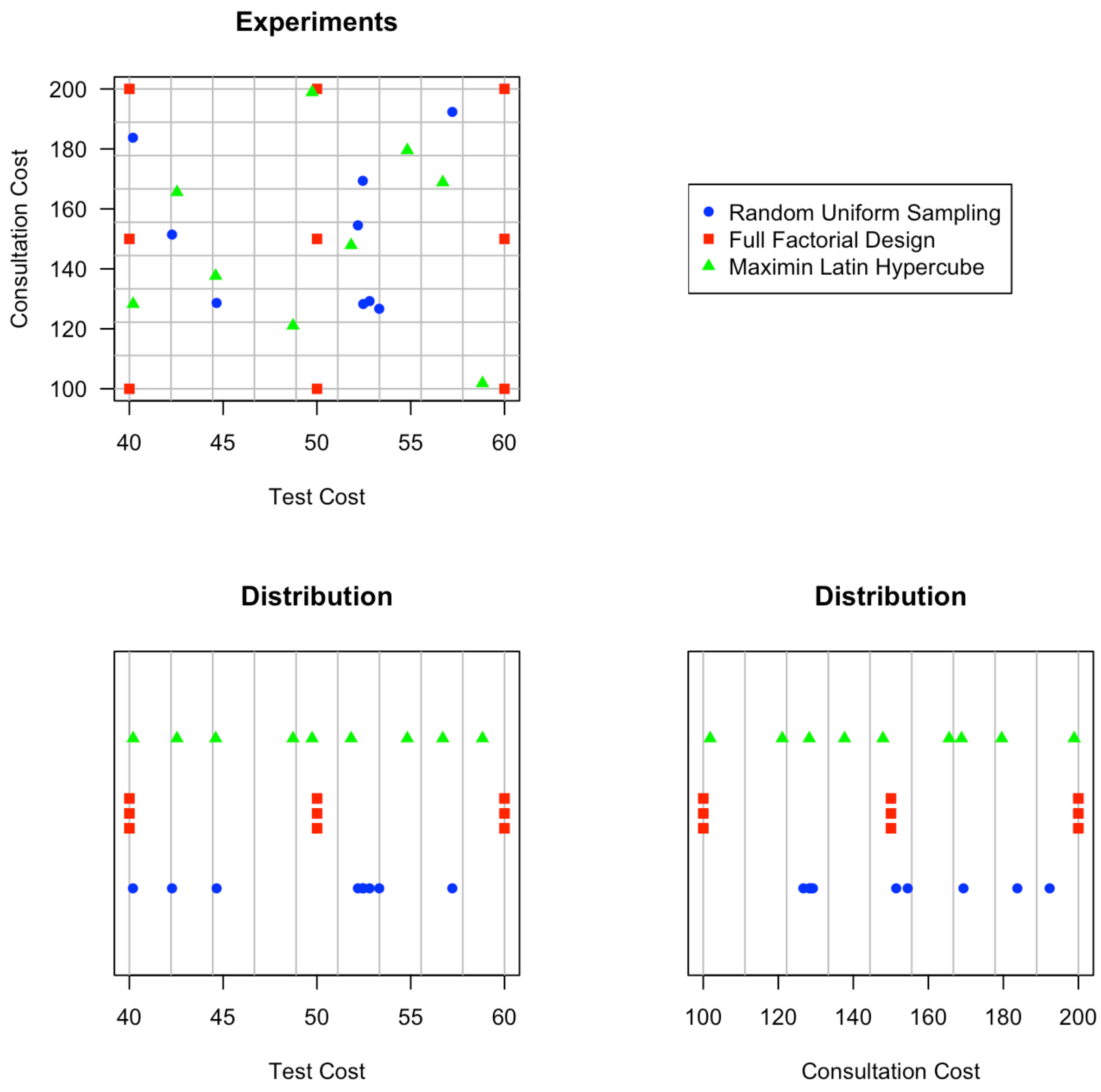


Full factorial

- Enumerates all possible parameter values and parameter interactions

- Easy to apply and full coverage, but requires a large number of samples: n^k

- Any continuous parameters need to be discretized

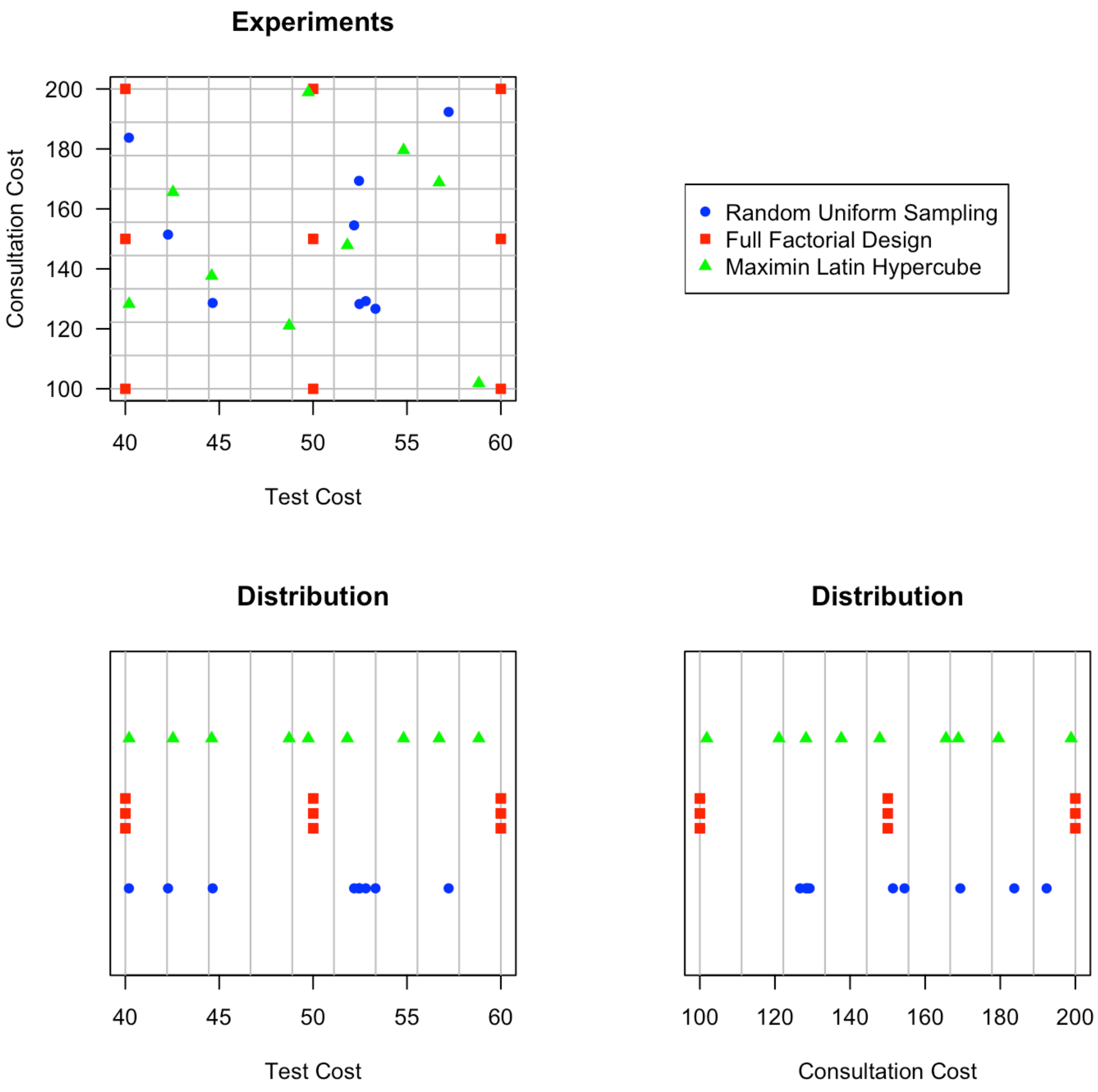


Latin Hypercube

- Divides the parameter space in bins and randomly samples one value in each bin

- Very efficient design and most commonly used

- Further optimized designs further improve the performance of the original Latin Hypercube sampling design
  - For example, maximin designs



Further considerations regarding DoE

- Use of Uniform distributions
- Continuous vs. discrete parameters
- Normalization may be required after generating the experiments
- Training and testing datasets need to be generated separately
  - Except for random designs
- Constraints on interactions between parameters may be needed
  - Not easily taken into account for Latin Hypercube sampling
  - Oversampling an option, but not formal Latin Hypercube anymore
- How many experiments to generate?
  - As many as possible! Iterative process…

Content Section 4: Performance measures for assessing metamodel accuracy

- Validation
- Error measures
- Calibration plots

Metamodel validation

- Metamodels are approximations , so validation is a crucial part of the metamodeling process
- Similar to validation of regression models
- Options for validation / performance assessment:
  - Error measures (mean, min, max), for example:
    - Error
    - Relative error
    - Absolute error
    - Relative absolute error
    - (Root) squared error
  - Calibration plots

Error measures

- **Error:** shows the amount with which the metamodel systematically underestimates/overestimates the observed outcomes (i.e., from the simulator)

  $$E = predicted - observed$$

- **Relative error:** shows the relative amount with which the metamodel systematically underestimates/overestimates the observed outcomes

  $$RE = (predicted - observed)/observed$$

- **Absolute error:** shows the distance between the metamodel predictions and the observed outcomes

  $$AE = |predicted - observed|$$

- **Relative absolute error:** shows the distance between the metamodel predictions and the observed outcomes relative to the observed outcomes

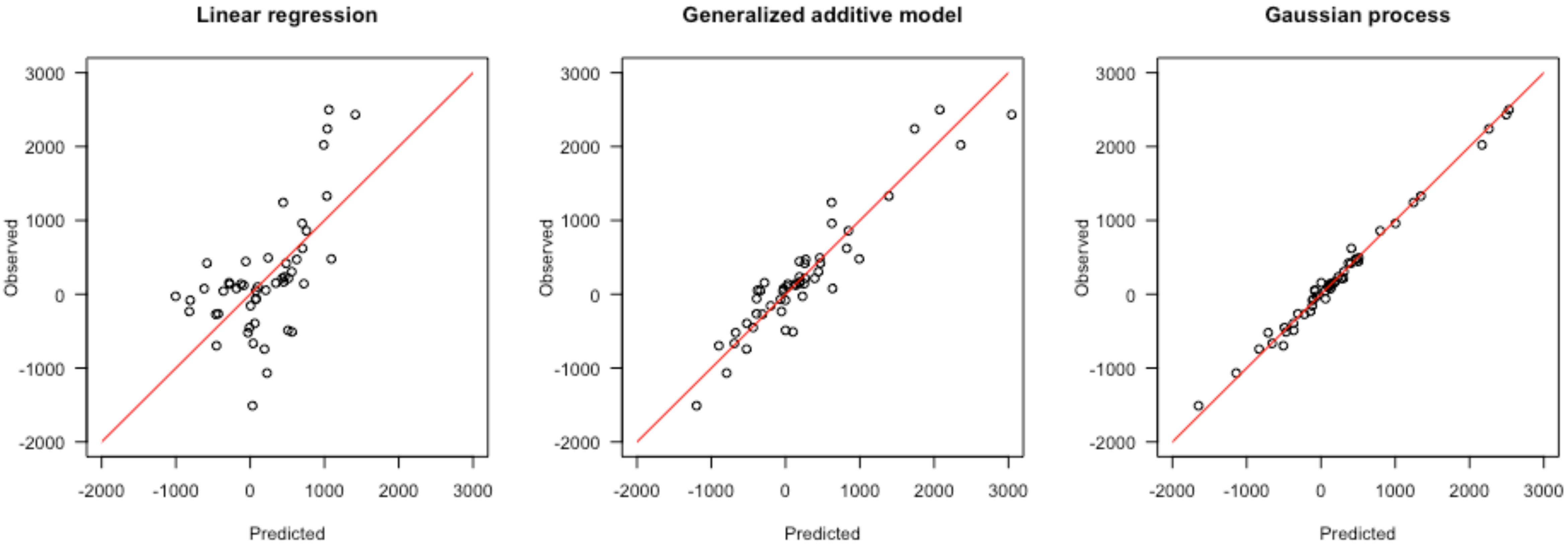
  $$RAE = |predicted - observed|/observed$$

- **Squared errors*:** increased weight of outliers

  $$RMSE = \sqrt{\sum(predicted - observed)^2/_{n}}$$

Information criteria might be less relevant in this context

Calibration plots



Error measures - interpretation

Which error value is acceptable?

- There is no, and cannot be, a general threshold for acceptability

- *The metamodel(s) should be fit for purpose*
  - Stakeholders should agree that the (remaining) error in outcomes is acceptable
  - In the simulation approach: using the metamodel instead of the simulator should not result in different conclusions

When errors are (too) large or when in doubt

Try and Try Again – Further considerations

Performance in practice

In general, metamodel accuracy can be quite good

J. Opl Res. Soc. Vol. 39, No. 10, pp. 939–948, 1988

0160-5682/88 $3.00 + 0.00

The Metamodel in Simulation Analysis: Can It Be Trusted?

LINDA WEISER FRIEDMAN and ISRAEL PRESSMAN

YES

Conclusion based on 30 simulation experiments with 3 simulation models

Weiser Friedman et al. The Metamodel in Simulation Analysis: Can It be Trusted?
The Journal of the Operational Research Society. Vol. 39, No. 10 (Oct., 1988), pp. 939-948

Performance in practice

Why can you expect prediction accuracy to be better for metamodels than for statistical models fitted on real-world data?

Compared with regression on real-world data, regression on simulated outcomes

1. Has no missing data or censored data
2. Has no measurement error
3. Can make use of large sample sizes
4. Has independent X's (may be correlated) and independent Y's, with a causal relationship between X and Y induced by the simulator

coffee
BREAK

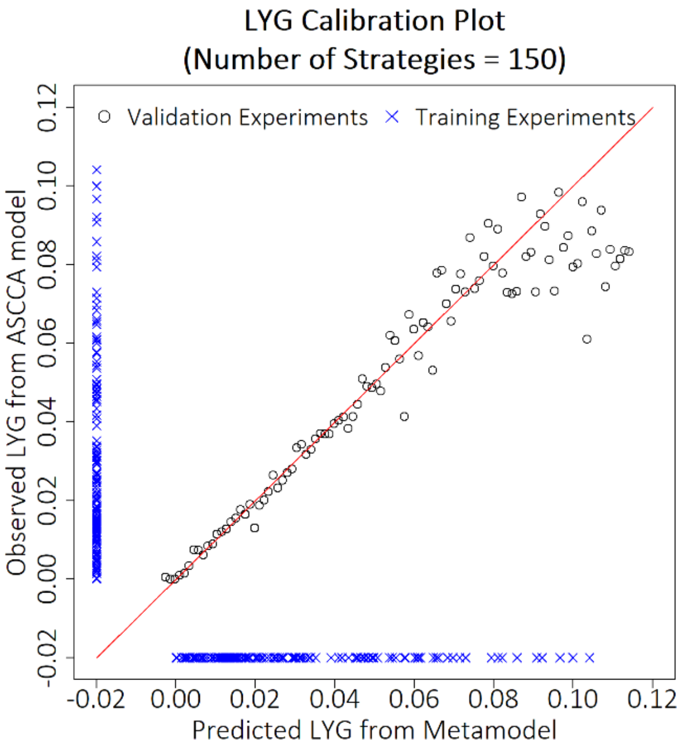
Breakout Exercise: Performing a full metamodeling study

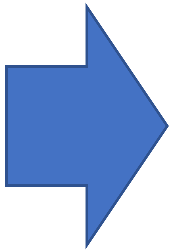
Further Considerations

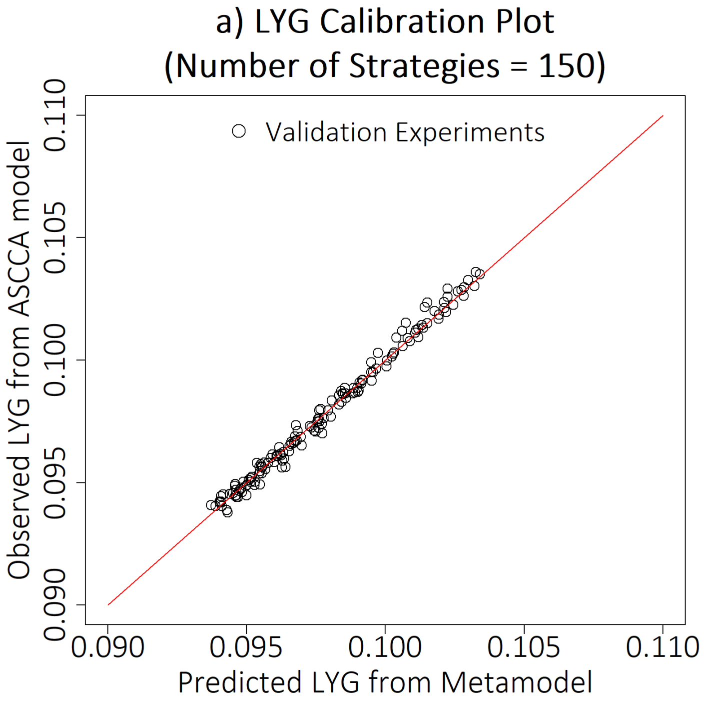
- Iterative nature of metamodeling process
- Extrapolation beyond simulator support
- General limitations

Iterative nature of metamodeling process

- Select technique and resample + refit to improve performance
- In optimization iteration is always needed: wide -> focused sampling

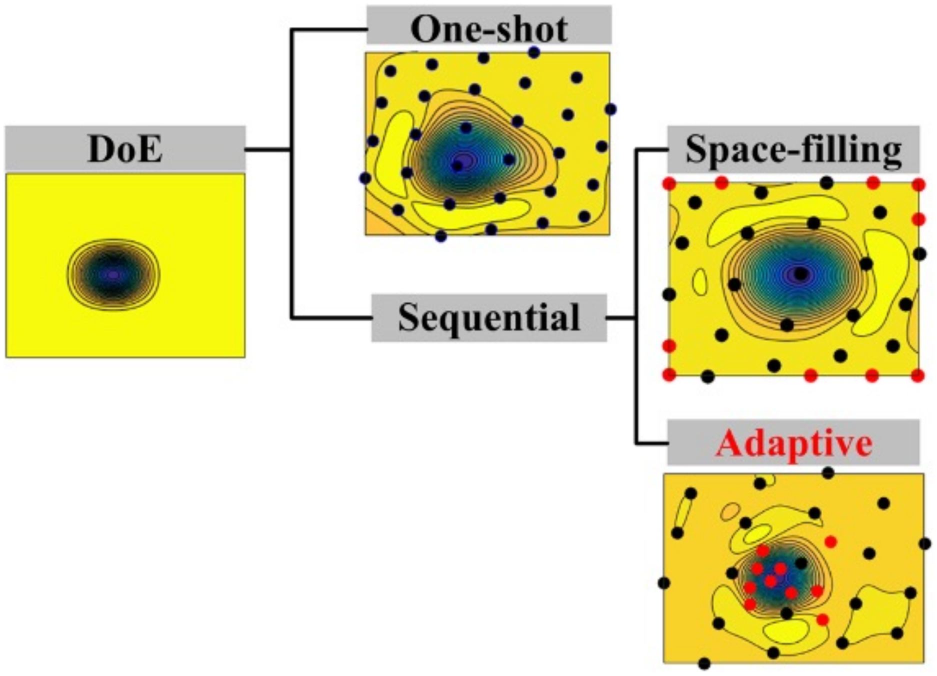


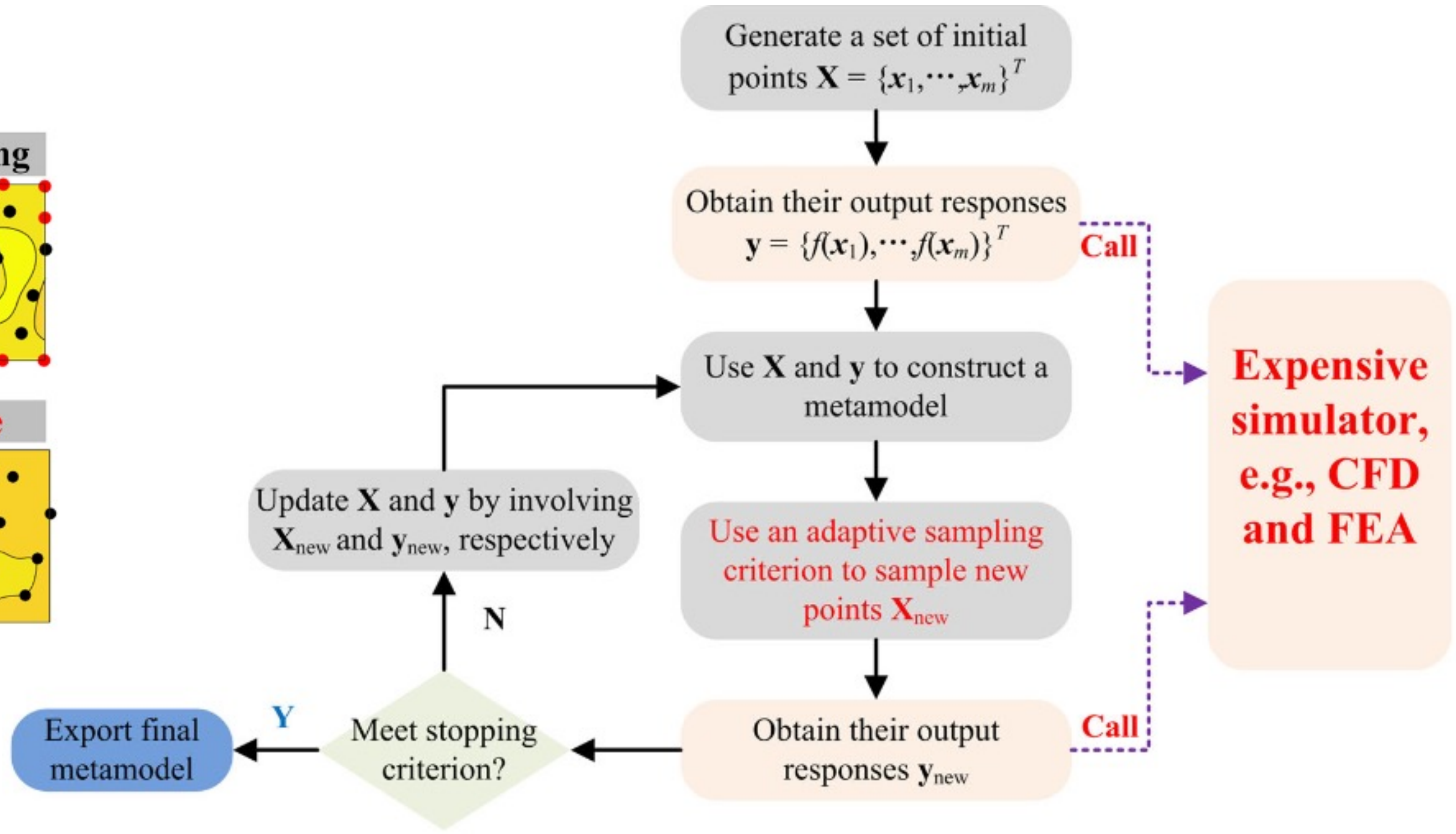




Adaptive sampling in metamodeling

- Global metamodeling vs optimization





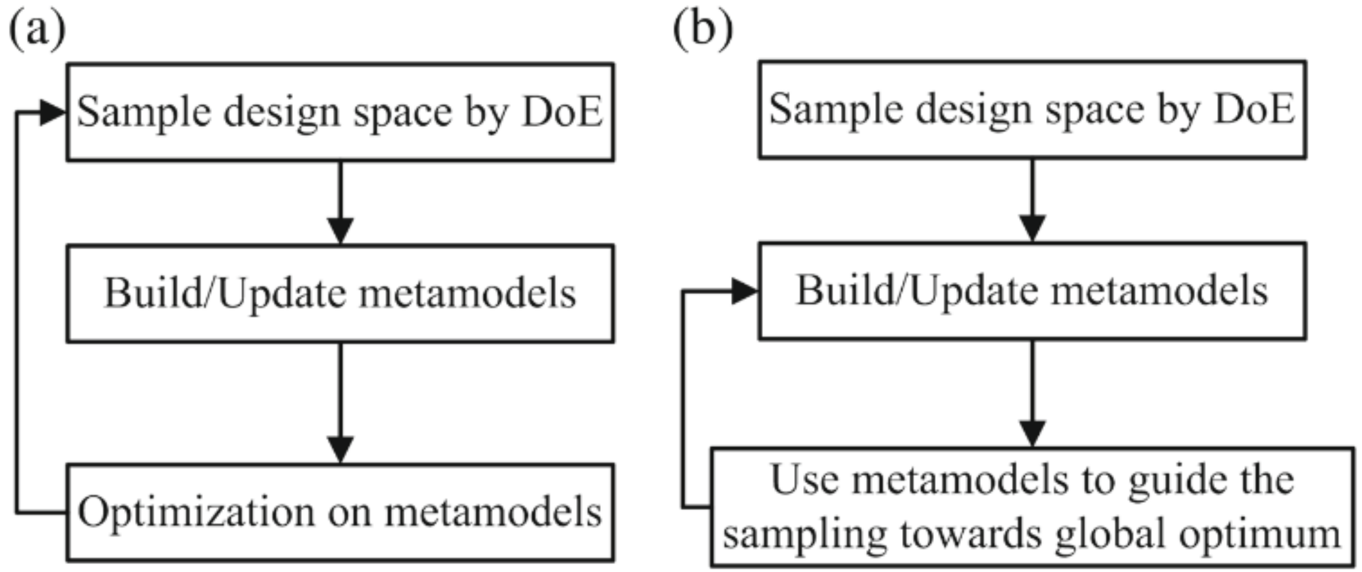
Figures 2 & 3 from Liu et al. (2018) A survey of adaptive sampling for global metamodeling in support of simulation-based complex engineering design. Struct Multidisc Optim.

Adaptive sampling in metamodeling

- Sequential (a) vs adaptive sampling (b)

(a): Interesting region(s) to explore *(best outcomes – optimization)*

(b): Interesting next sample point to assess *(largest error - global best fit)*



In health economics

- Direct model access?
- Goal of metamodeling?
  - Optimization
  - Exploration

Figure 8 from Liu et al. (2018) A survey of adaptive sampling for global metamodeling in support of simulation-based complex engineering design. Struct Multidisc Optim.

Extrapolation beyond simulator

- Metamodels can generate outcomes for parameter values that may be beyond the support of the simulator (i.e., original model)

Table 2. Definitions and outcomes of screening strategies expected to be optimal in LYG.

Constraint	Screening strategy identified as optimal				Predicted strategy outcomes*			
Maximum colonoscopy capacity†	Start age screening (years)	Screening interval (years)	Number of screening rounds	FIT cutoff‡ for referral (ng/mL)	Number of colonoscopies†	LYG	Incremental costs (€)	NMB§
300	32	1	18	150	296	0.058	-266	1430
450	40	1	20	150	441	0.081	-316	1941
550	33	2	21	150	546	0.092	-361	2202
650	34	2	21	100	647	0.097	-363	2294
800	35	2	21	75	738	0.100	-380	2374

Koffijberg et al. (2021) Using Metamodeling to Identify the Optimal Strategy for Colorectal Cancer Screening. *Value Health.*

- Current Dutch screening guideline:
  - Start age: 55, Interval: 2, Number of rounds: 11, FIT cut-off: 75 ng/ml

General limitations

- Extra step in the overall modeling process
- Metamodels are approximations, so verification may be valuable (is essential for optimization)

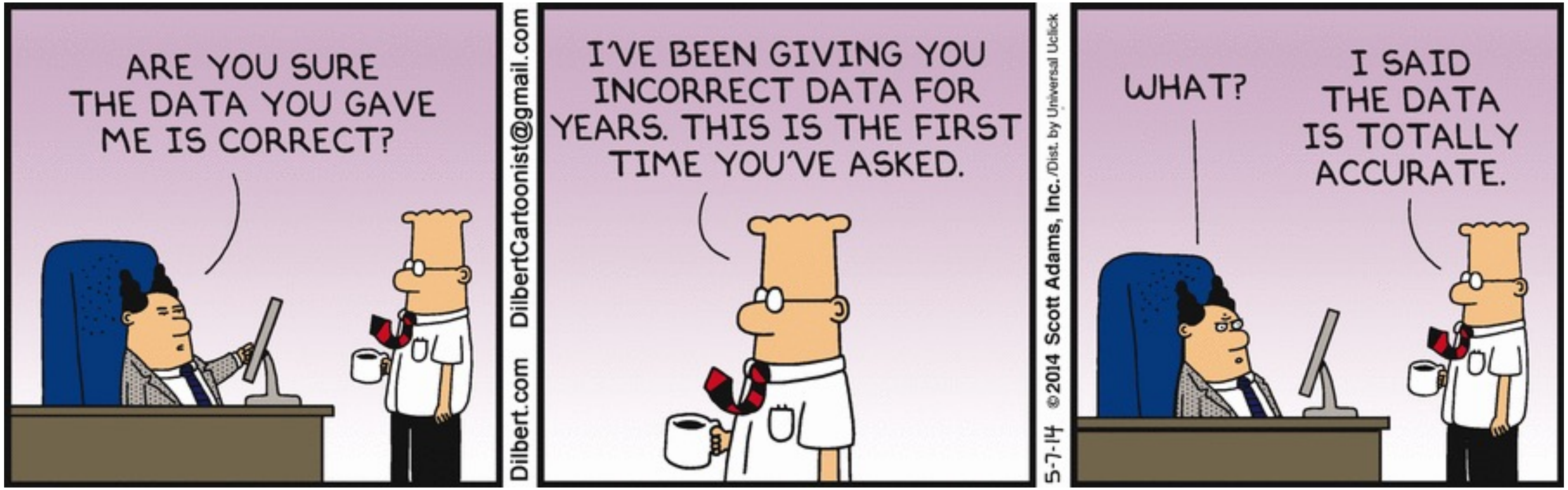
Table 3. Top 10 screening strategies in terms of LYG for a colonoscopy capacity of 550.

Screening strategy identified as optimal				Predicted strategy outcomes*			
Start age screening (years)	Screening interval (years)	Number of screening rounds	FIT cutoff[‡] for referral (ng/mL)	Number of colonoscopies[†]	LYG	Incremental costs (€)	NMB[§]
33	2	21	150	546	0.0920	-361	2202
34	2	20	150	545	0.0919	-348	2185
36	2	19	150	548	0.0917	-348	2183
35	2	19	150	541	0.0916	-361	2193
33	2	20	150	537	0.0914	-359	2186
32	2	21	150	538	0.0912	-343	2166
34	2	19	150	533	0.0911	-343	2166
37	2	18	150	543	0.0911	-361	2183
36	2	18	150	536	0.0910	-343	2163
38	2	18	150	549	0.0909	-347	2165

Koffijberg et al. (2021) Using Metamodeling to Identify the Optimal Strategy for Colorectal Cancer Screening. *Value Health.*

General limitations

- An accurate metamodel may be just as valuable as the original model
- It also inherits the original model’s validity, limitations and plausibility
- Your metamodel can (at most) be as good (or bad) as the original model



Start with a validated model you understand!

Overall Question & Answer

Thank you!

Hendrik Koffijberg
h.koffijberg@utwente.nl

Koen Degeling
info@koendegeling.nl