FACULTEIT INGENIEURSWETENSCHAPPEN

Grid-Enabled Adaptive Surrogate Modeling for Computer-Based Design

SUMO Lab

INTEC Broadband Communication Networks Research Group (IBCN)

Outline

- Who are we ?
- Introduction
- Surrogate modeling
- SUMO Toolbox
- Examples
- Conclusions

Outline

- **Who are we ?**
  - who are we
  - what do we do
- **Introduction**
- **Surrogate modeling**
- **SUMO Toolbox**
- **Examples**
- **Conclusions**

Who are we ?

Ghent University

Faculty of Engineering

Department of Information Technology (INTEC)

INTEC Broadband Communication Networks (IBCN)

Surrogate Modeling Lab

Who are we ?

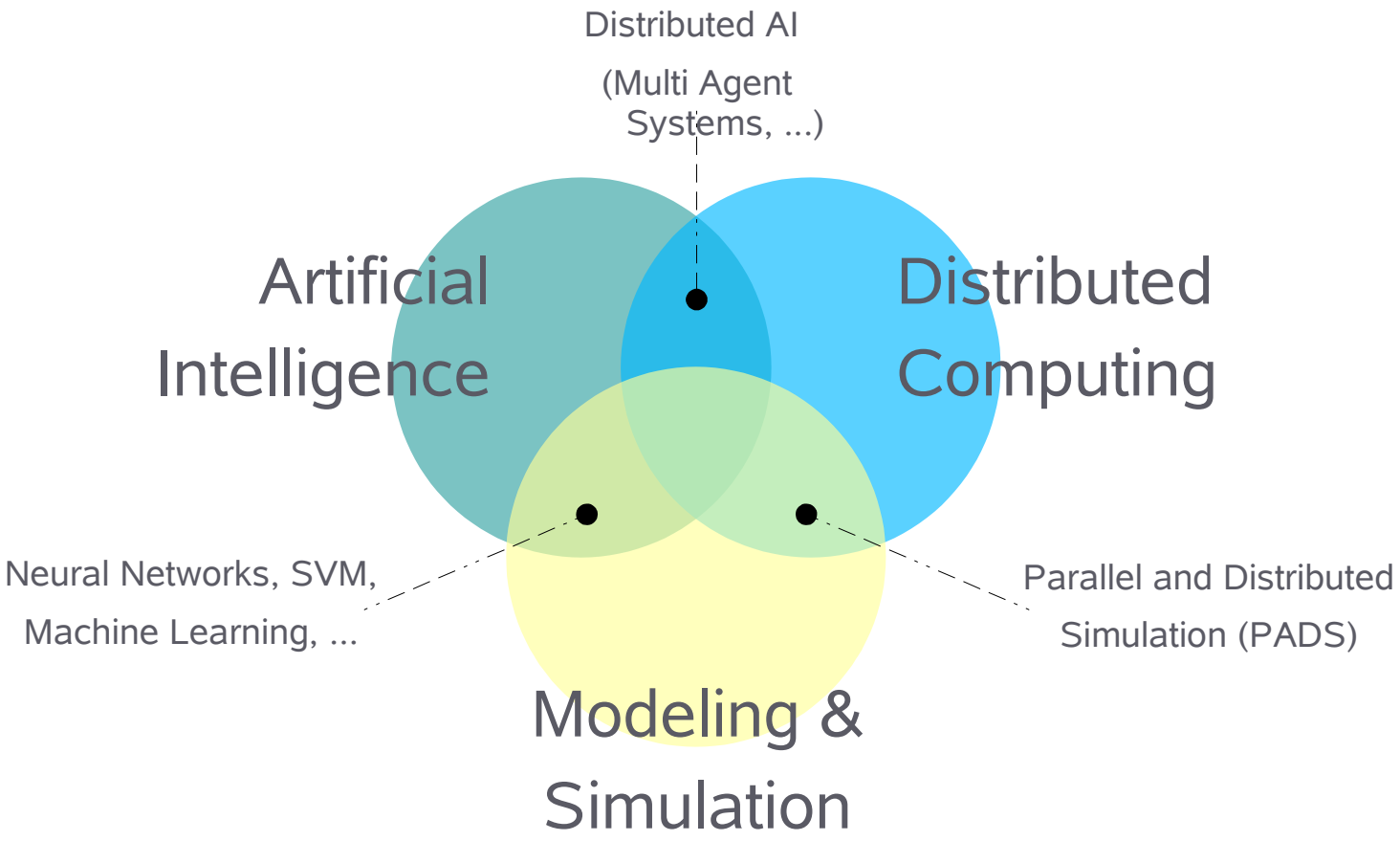

UNIVERSITEIT GENT
FACULTY OF ENGINEERING
INTEC
IBCN
Universiteit Antwerpen
PATS
KATHOLIEKE UNIVERSITEIT LEUVEN
DistriNet
Research Group
IBBT

Members

- **IBCN**
  - 8 professors
  - 7 postdocs
  - 84 research members

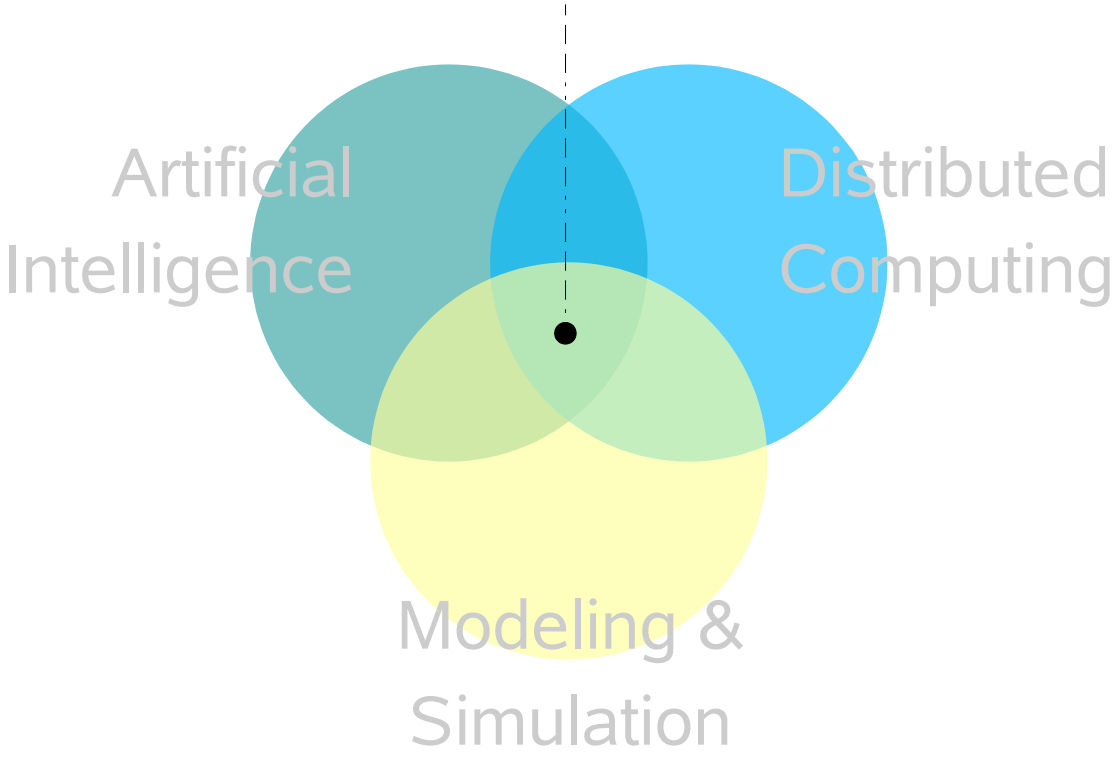
- **SUMO lab**
  - Professors
    - **Prof. dr. ir. Tom Dhane**
    - **Prof. dr. ir. Eric Laermans**
  - PhD students
    - **Dirk Gorissen**
    - **Karel Crombecq**
  - Postdocs
    - **Dirk Deschrijver**
    - **Ivo Couckuyt**
    - **Eng. Francesco Ferranti**

What do we do


Distributed AI
(Multi Agent
Systems, ...)
Artificial
Intelligence
Distributed
Computing
Neural Networks, SVM,
Machine Learning, ...
Parallel and Distributed
Simulation (PADS)
Modeling &
Simulation

Adaptive Surrogate Modeling

efficient and accurate characterization, modeling and simulation of complex systems in science and engineering



Outline

- **Who are we ?**
- **Introduction**
  - Surrogate model ?
  - What are we looking for ?
  - Existing approaches and techniques
- **Surrogate modeling**
- **SUMO Toolbox**
- **Examples**
- **Conclusions**

A New Science Paradigm

A New Science Paradigm

- thousand years ago : experimental science
  - description of natural phenomena

A New Science Paradigm

- thousand years ago : experimental science
  - description of natural phenomena
- last few hundred years : theoretical science
  - Newton's laws, Maxwell's equations …

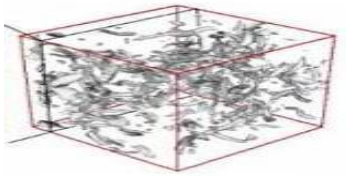
$$\left(\frac{\dot{a}}{a}\right)^2 = \frac{4\pi G\rho}{3} - K\frac{c^2}{a^2}$$

A New Science Paradigm

- thousand years ago : **experimental science**
  - description of natural phenomena
- last few hundred years : **theoretical science**
  - Newton's laws, Maxwell's equations …

$$\left(\frac{\dot{a}}{a}\right)^2 = \frac{4\pi G\rho}{3} - K\frac{c^2}{a^2}$$

- last few decades : **computational science**
  - simulation of complex phenomena



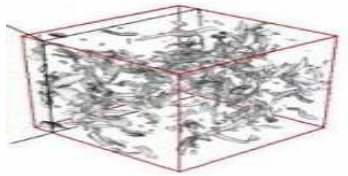
A New Science Paradigm

- thousand years ago : **experimental science**
  - description of natural phenomena

- last few hundred years : **theoretical science**
  - Newton's laws, Maxwell's equations …

$$\left(\frac{\dot{a}}{a}\right)^2 = \frac{4\pi G\rho}{3} - K\frac{c^2}{a^2}$$

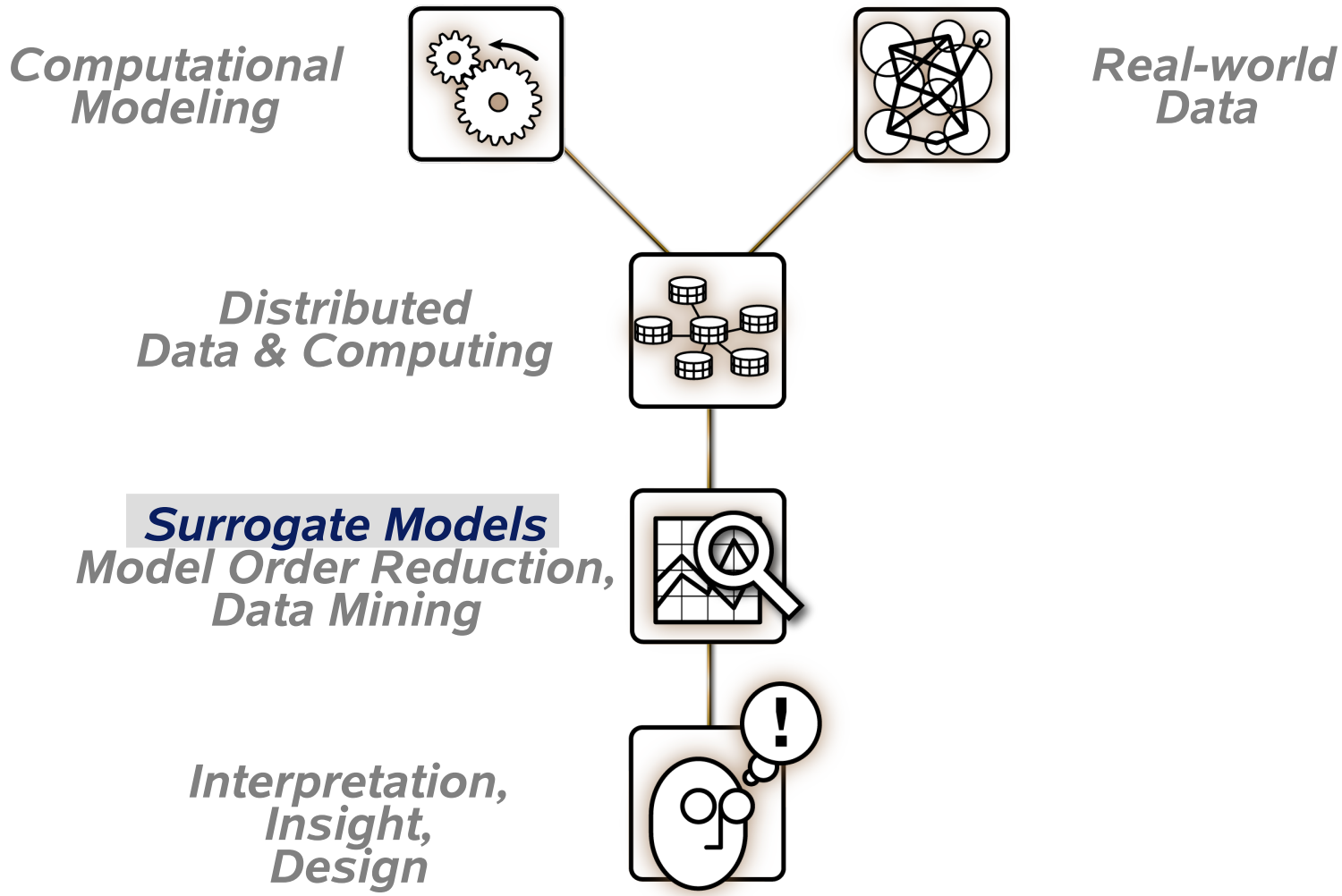
- last few decades : **computational science**
  - simulation of complex phenomena



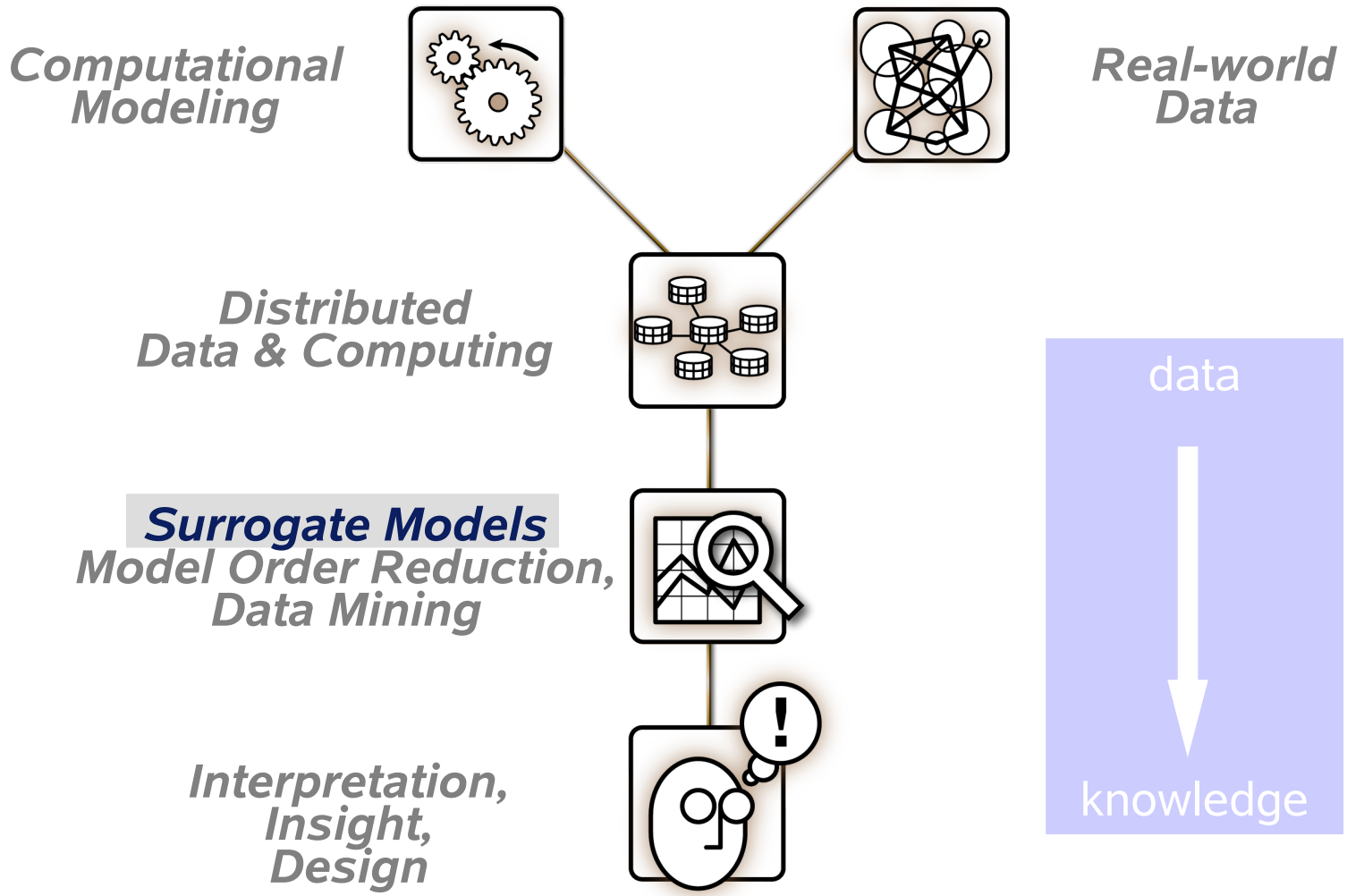
- today : **e-Science or data-centric science**
  - massive computing
  - large data exploration and mining
  - unify : theory, experiment, and simulation

(With thanks to Jim Gray)

A New Science Paradigm



A New Science Paradigm



Surrogate model ?

- **system modeling**
  - **real world**
    - I/O system
    - stimulus / response

    - examples: *mechanical, electrical, optical, electronic, chemical … systems*

Surrogate model ?

- **system modeling**
  - real world
    - I/O system
    - stimulus / response
  - **simulation model**
    - approximation
    - discretization

- model = *abstraction of a real system*
- simulation = *virtual experiment*

Surrogate model ?

- **system modeling**
  - real world
    - I/O system
    - stimulus / response
  - simulation model
    - approximation
    - discretization
  - **surrogate model**
    - metamodel, RSM, emulator
    - scalable analytical model
    - *"model of model"*

Surrogate model ?

- **simulation model** : widely used in engineering design

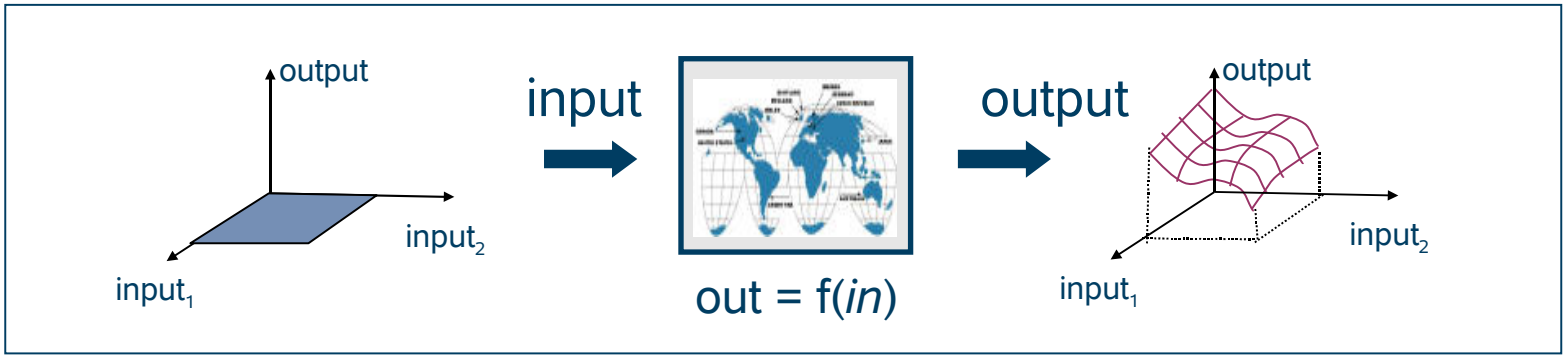
- each new sample in the input design space, requires new computer simulation
- accurate, high fidelity numerical model

Surrogate model ?

- **however, simulation models…**
  - …complex
  - …time consuming to run
  - …optimization is expensive
  - …not always available
  - …highly specialized
    - **scalability?**
    - **model chaining?**
    - **integration with other tools?**
    - **hardware / software requirements?**
    - **licensing?**
    - **…**

Surrogate model ?

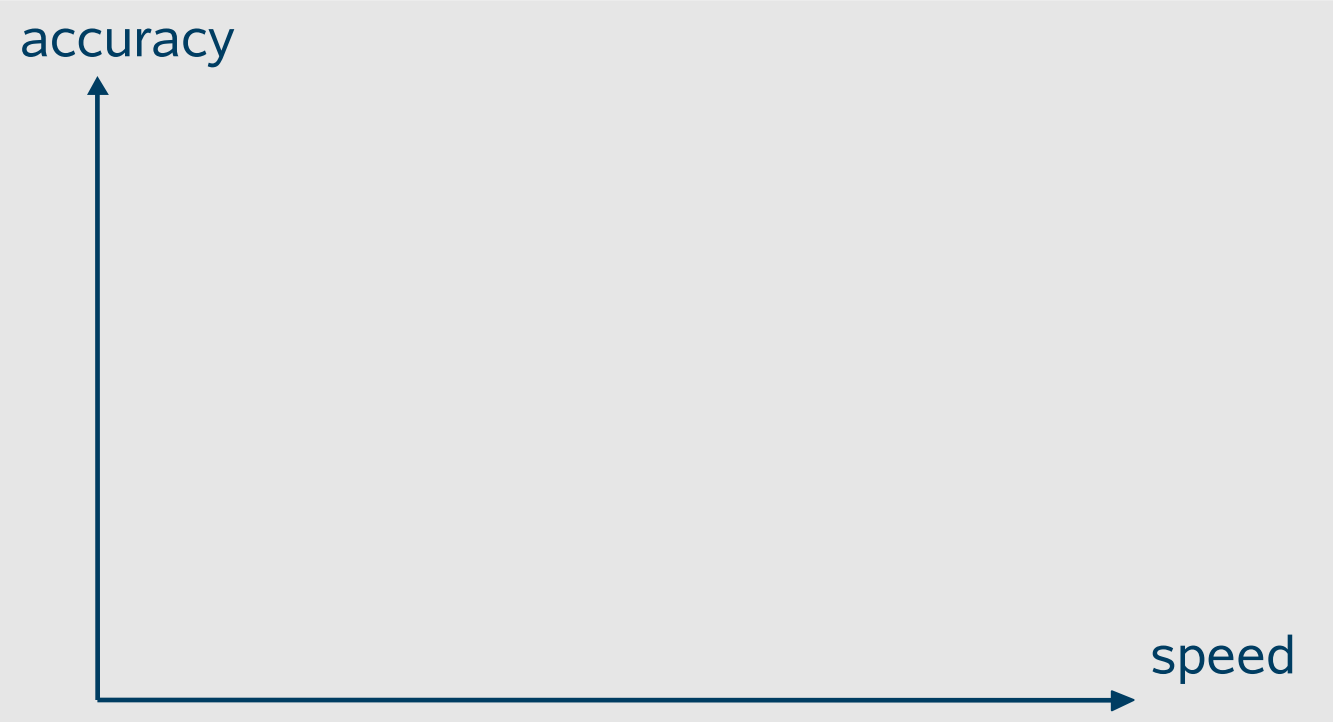
surrogate model



- analytical surrogate model
  - one-time upfront time investment
  - harness the power of the grid for simulation execution
  - adaptive sampling
- covers complete design space
  - design optimization, "*what-if*" analysis, sensitivity analysis

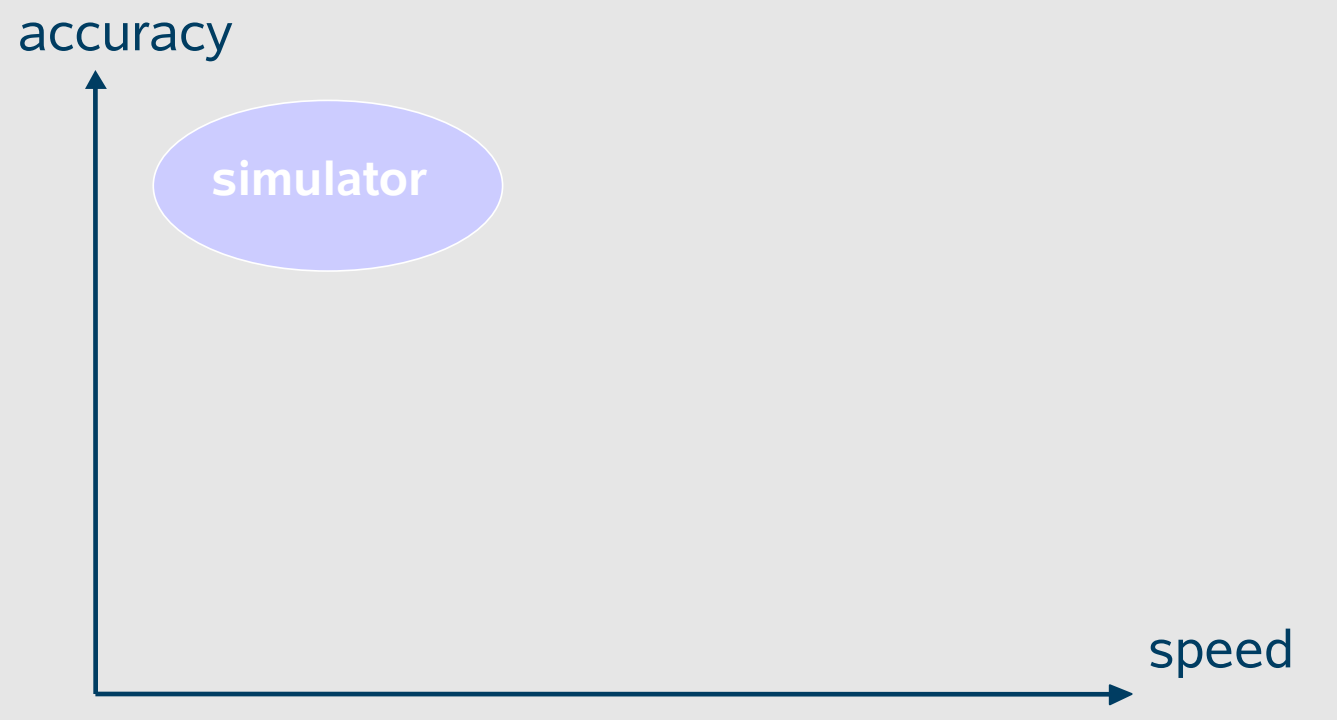
What are we looking for?

- **accuracy / speed trade-off**



What are we looking for?

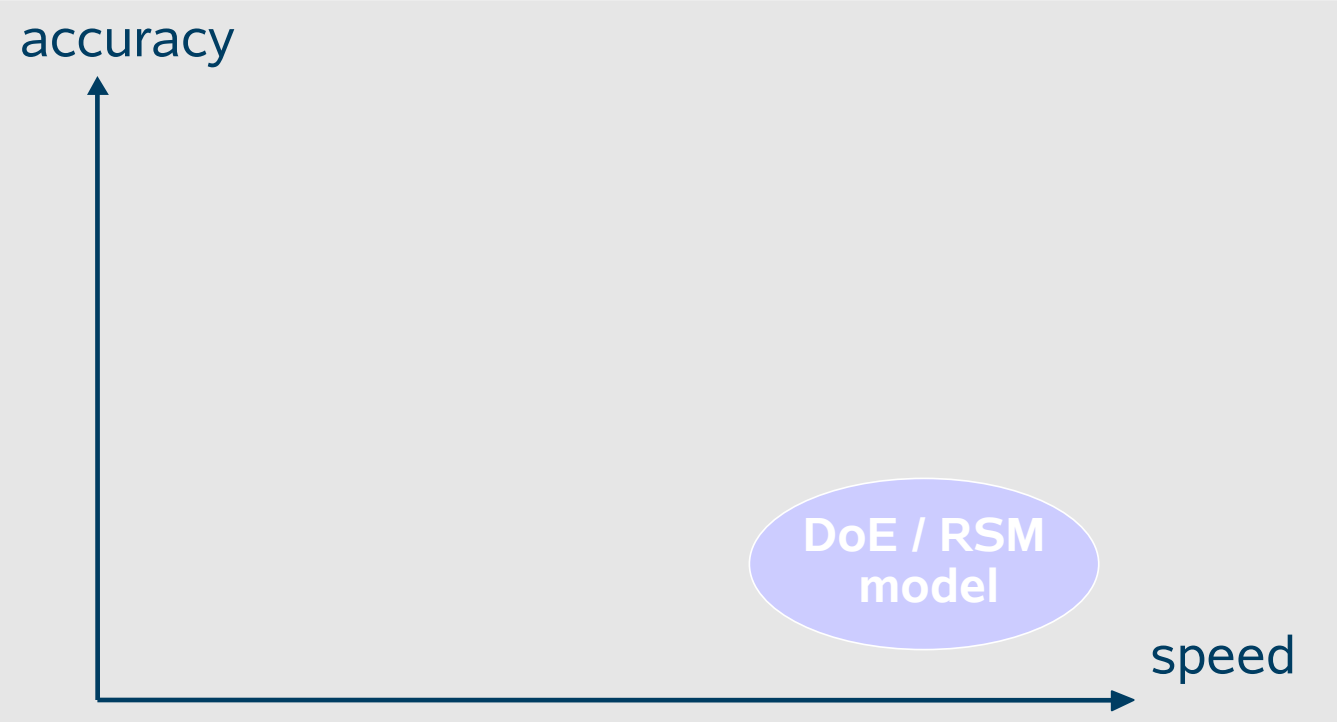
- **simulators**
  - domain-specific
  - high-accuracy



What are we looking for?

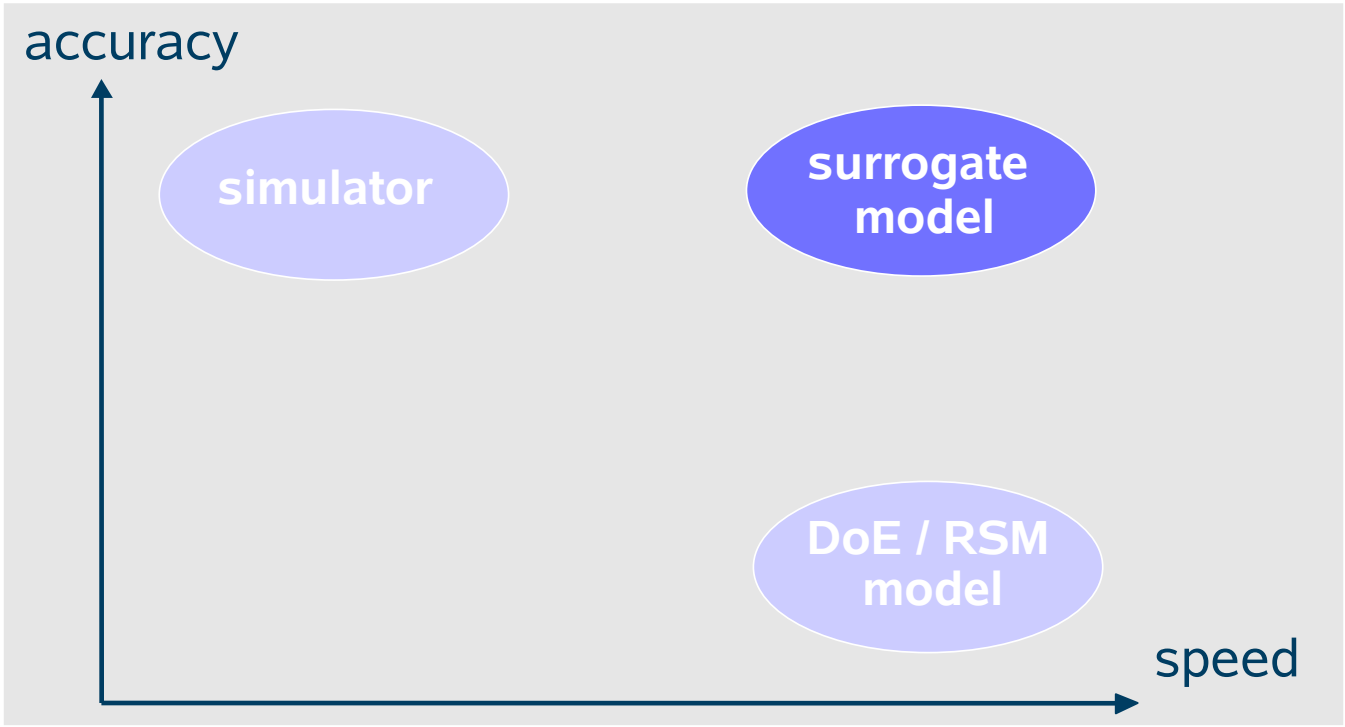
- **models**
  - 2nd order polynomial

    Response Surface Models (RSM)



What are we looking for?

- **best of both worlds**
  - combining accuracy & generality of **simulators**, with the speed & flexibility of **models**



Surrogate model ?

- **advantages**
  - instant evaluation
  - compact formulation (few 100 parameters)

- **applications**
  - prototyping
  - design space exploration
  - design optimization
  - sensitivity analysis
  - *what-if* analysis
  - ...

Surrogate model ?

- **surrogate modeling challenges**
  - experimental design?
  - sample selection?
  - model type?
  - model tuning?
  - black box – grey box – white box?
  - …
- **only as good as the available data / designer**
- **surrogate models are still models**
  - model assessment & model selection are crucial

What are we looking for?

- **Grid-enabled adaptive algorithm for automatic surrogate model construction**
  - fully automated
  - minimize prior, problem specific knowledge
    - **trade-off**
  - minimal number of samples
    - **computationally expensive**
  - support for distributed computing
  - pre-defined accuracy
  - pluggable / extensible
    - **no one-size-fits-all**
  - integrate easily into the design process

Existing approaches

- **existing approaches**
  - discrete model library
    - **Database**
  - look-up tables with local curve fitting
  - hand made analytical models
  - …

Existing approaches

- **common drawbacks**
  - oversampling / undersampling
    - waste of resources / important details missed
  - overmodeling / undermodeling
  - accuracy unknown
  - prior knowledge required
  - problem specific
  - "*not invented here*" syndrome

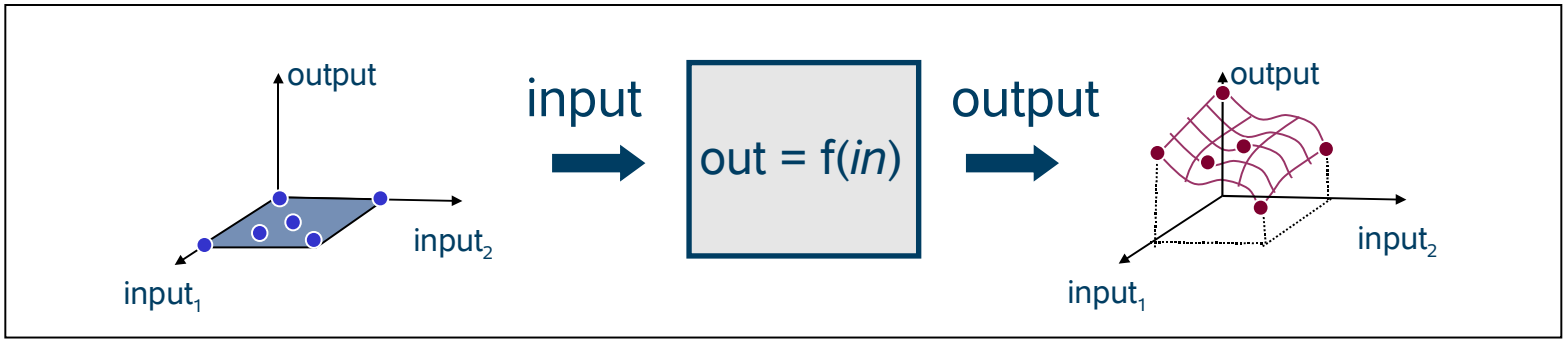
highly skilled modeler
several months of work

Outline

- **Who are we ?**
- **Introduction**
- **Surrogate modeling**
  - adaptive modeling
  - adaptive sampling
  - distributed computing
  - adaptive surrogate modeling
- **SUMO Toolbox**
- **Examples**
- **Conclusions**

Surrogate modeling

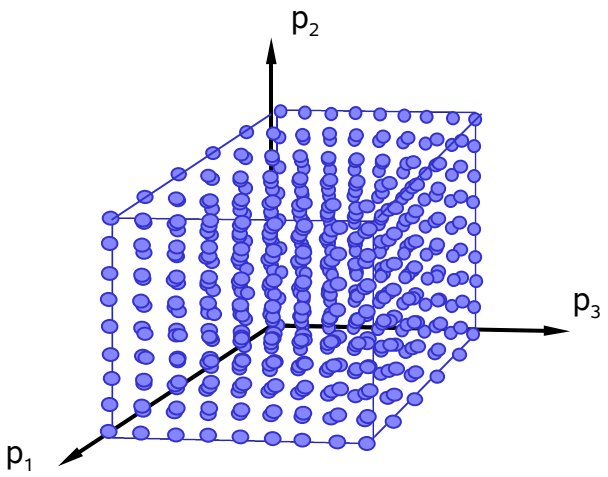
scalable **surrogate model**, valid over design space



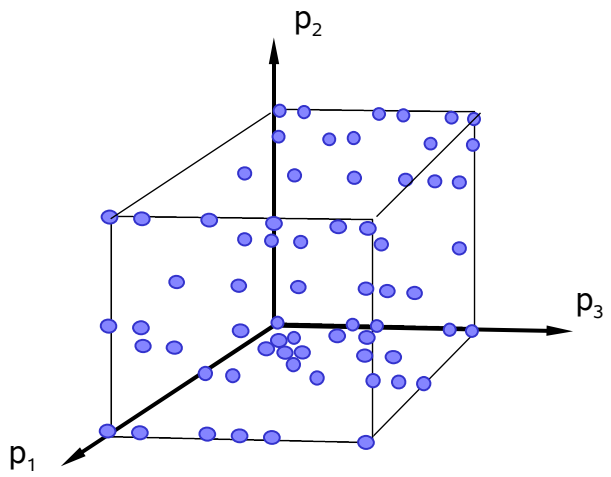
- 3 key technologies (+ 1 in development)
  - adaptive data sampling
  - adaptive model building
  - distributed computing
  - optimization

Surrogates – *adaptive sampling*

- traditional approach
  - uniform sampling
  - oversampling
  - undersampling

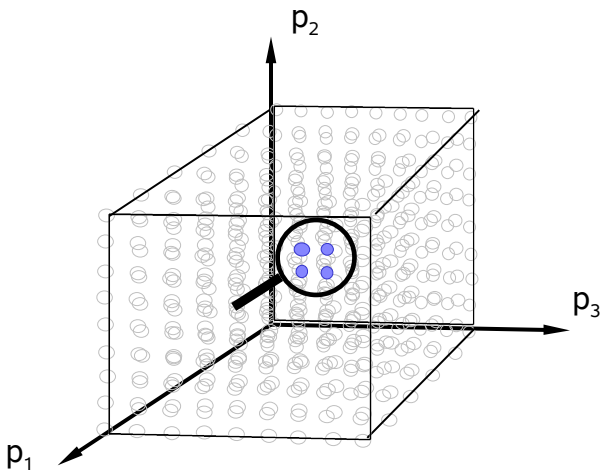


- adaptive sampling
  - optimal sample distribution
  - *Reflective Exploration*

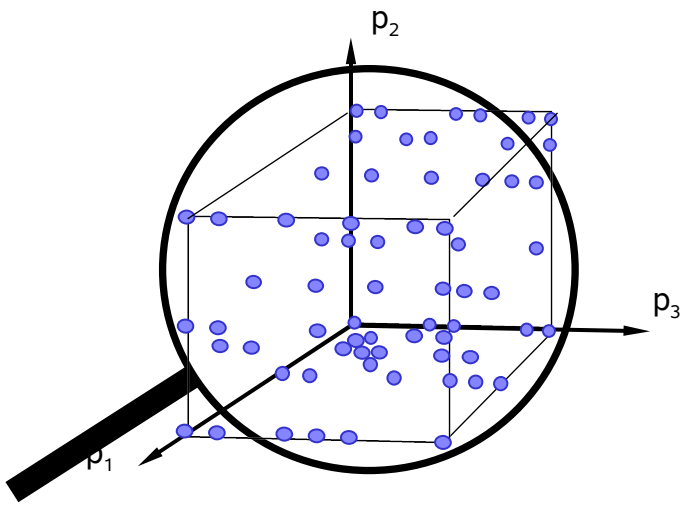


Surrogates – *adaptive modeling*

- traditional approach
  - local approximation
  - overmodeling
  - undermodeling



- adaptive modeling
  - global approximation
  - optimal model complexity



Surrogates – *distributed computing*

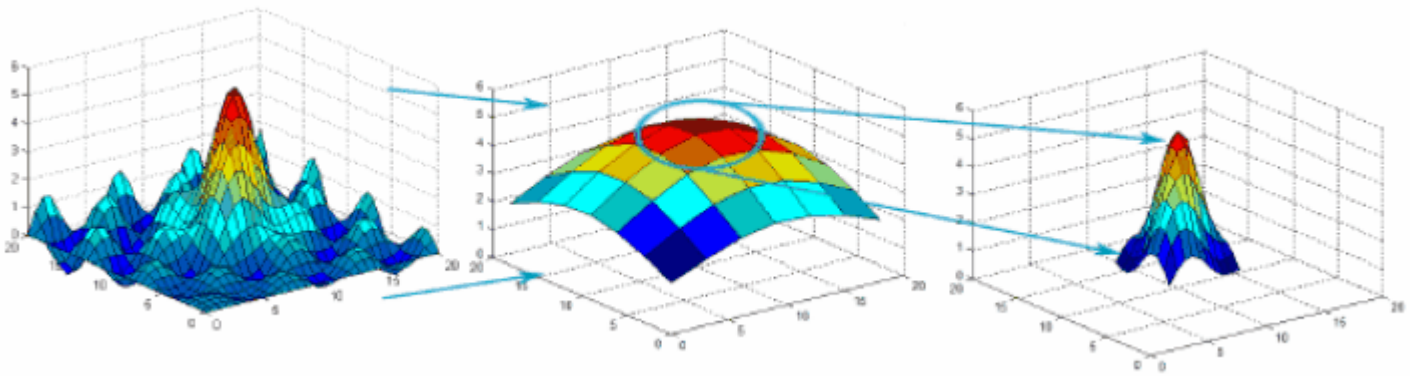
- traditional approach
  - sequential computing

- distributed computing
  - cluster
  - grid

Surrogates – *optimization*

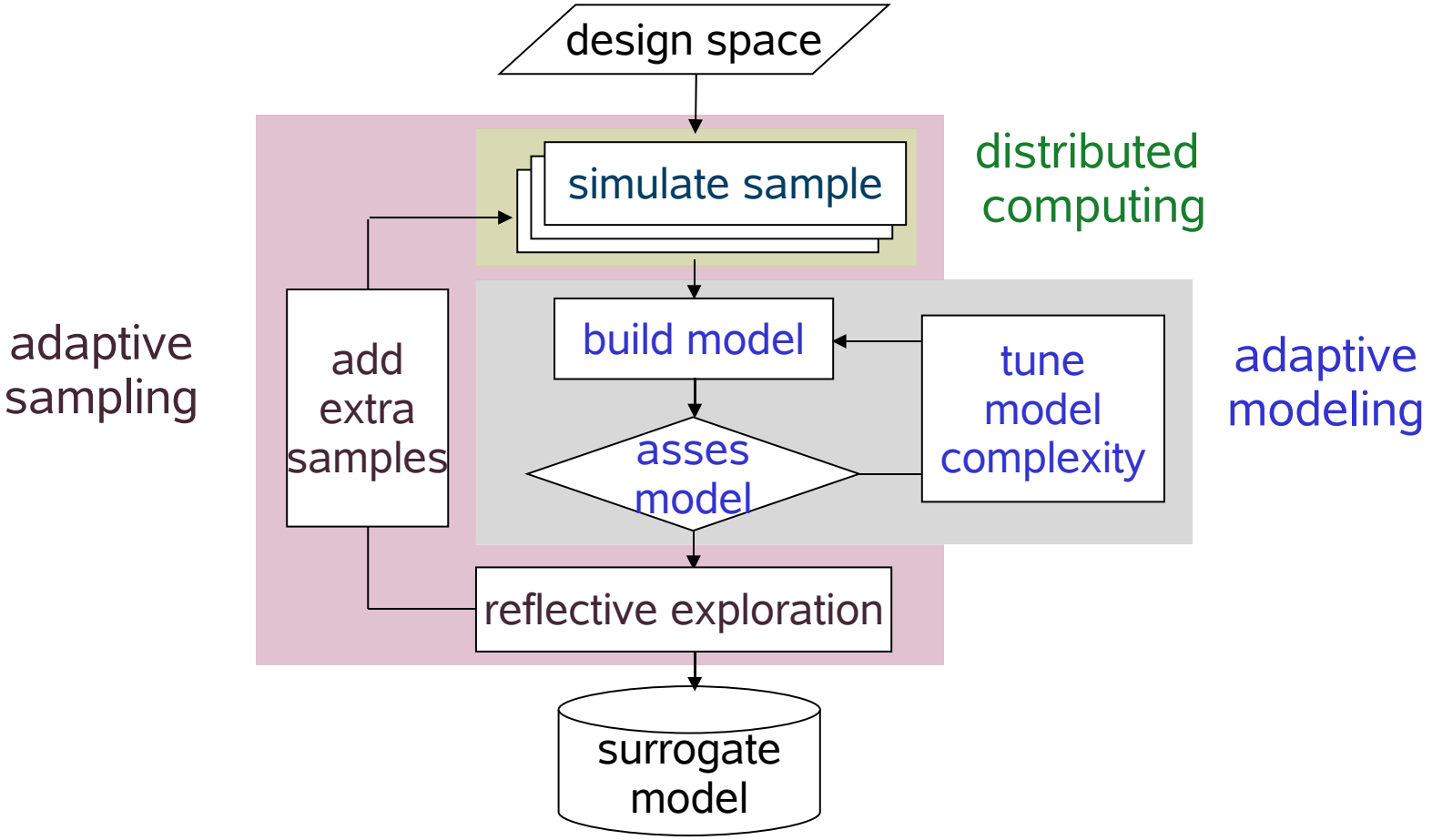
- traditional approach
  - classic optimization
    - not well suited for computational expense simulations

- Optimization
  - surrogate-assisted optimization
    - global surrogate model
    - intermediate surrogate models & zoom-in



Adaptive surrogate modeling

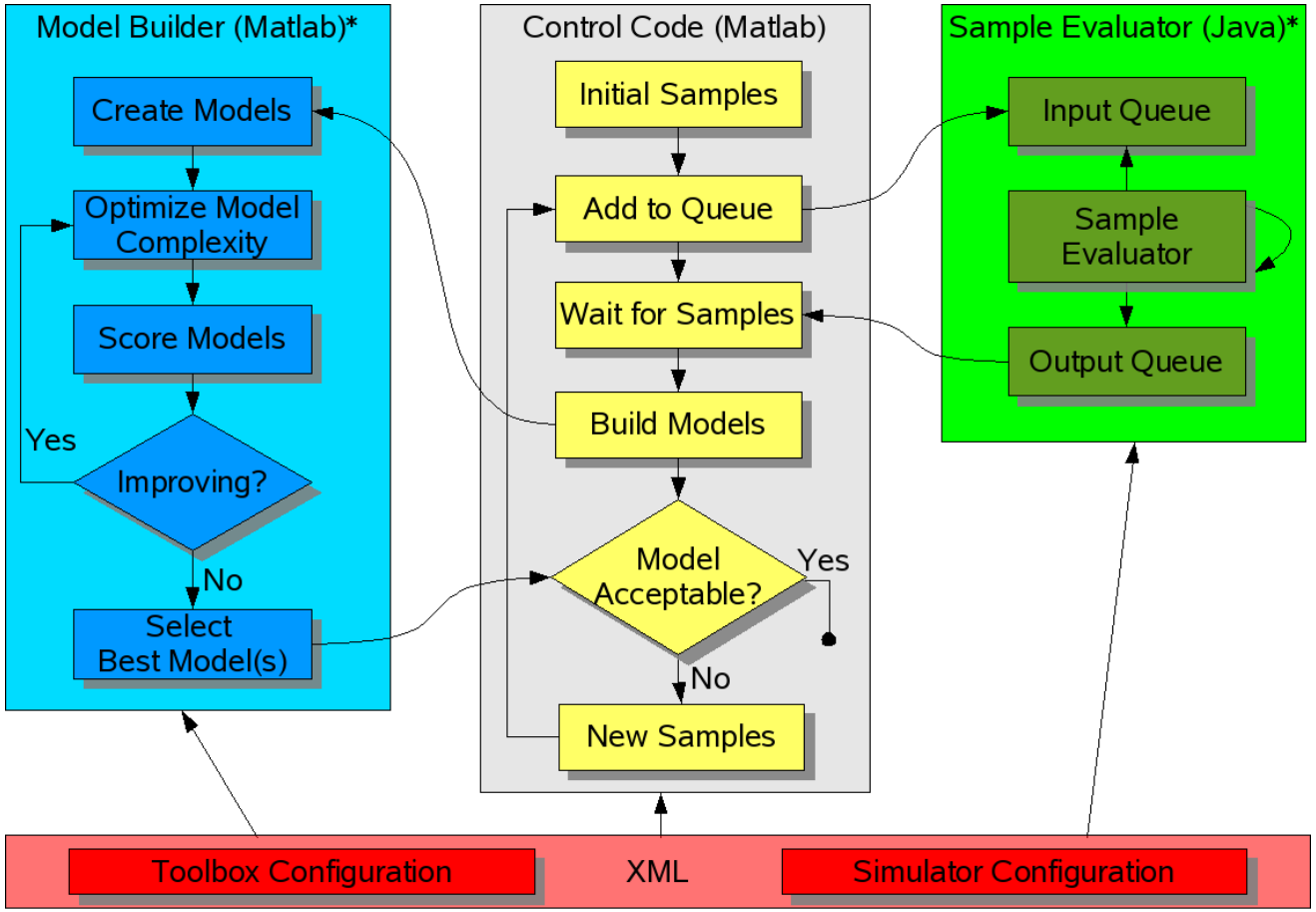
- flow chart



Outline

- **Who are we ?**
- **Introduction**
- **Surrogate modeling**
- **SUMO Toolbox**
  - control flow & design
  - automatic model type selection
  - integrating gridcomputing
- **Examples**
- **Conclusions**

- SUrrogate MOdeling (SUMO) Toolbox



* The Model Builder and Sample Evaluator run in parallel (non blocking)

levels of pluggability

adaptive modeling

- supports multiple model types
  - Polynomial/Rational functions
  - Artificial Neural Networks
  - RBF models
  - Support Vector Machines (LS-SVM, epsilon-SVR, nu-SVR)
  - Kriging models
  - Splines

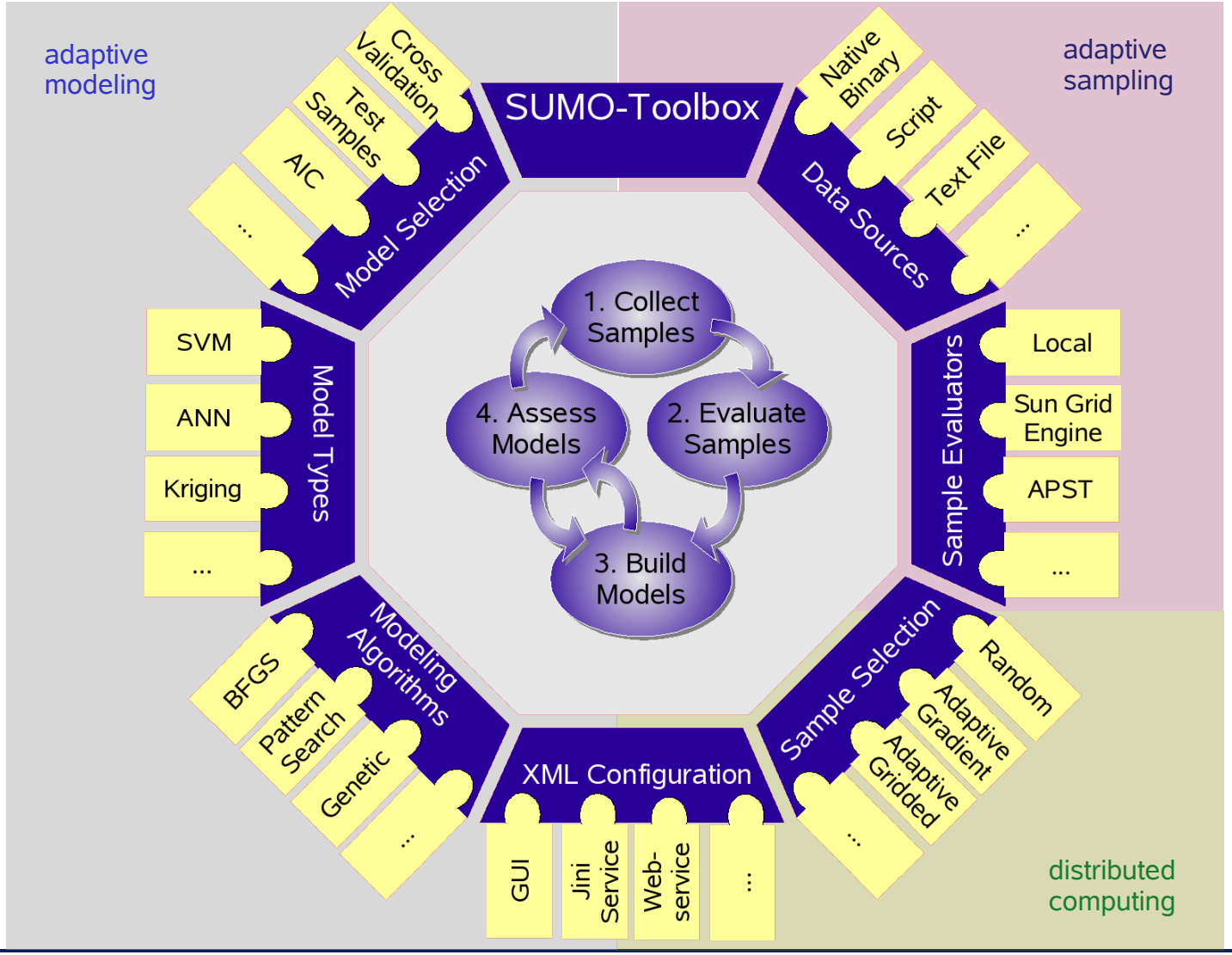
adaptive sampling

- modeling algorithm (BFGS, pattern search, GA, PSO, ...)
- initial experimental design (central composite, LHS, ...)
- sequential design (error-based, density-based, hybrid, ...)
- model selection (crossvalidation, R^2, AIC, ...)

distributed computing

- sample evaluation (local, distributed)

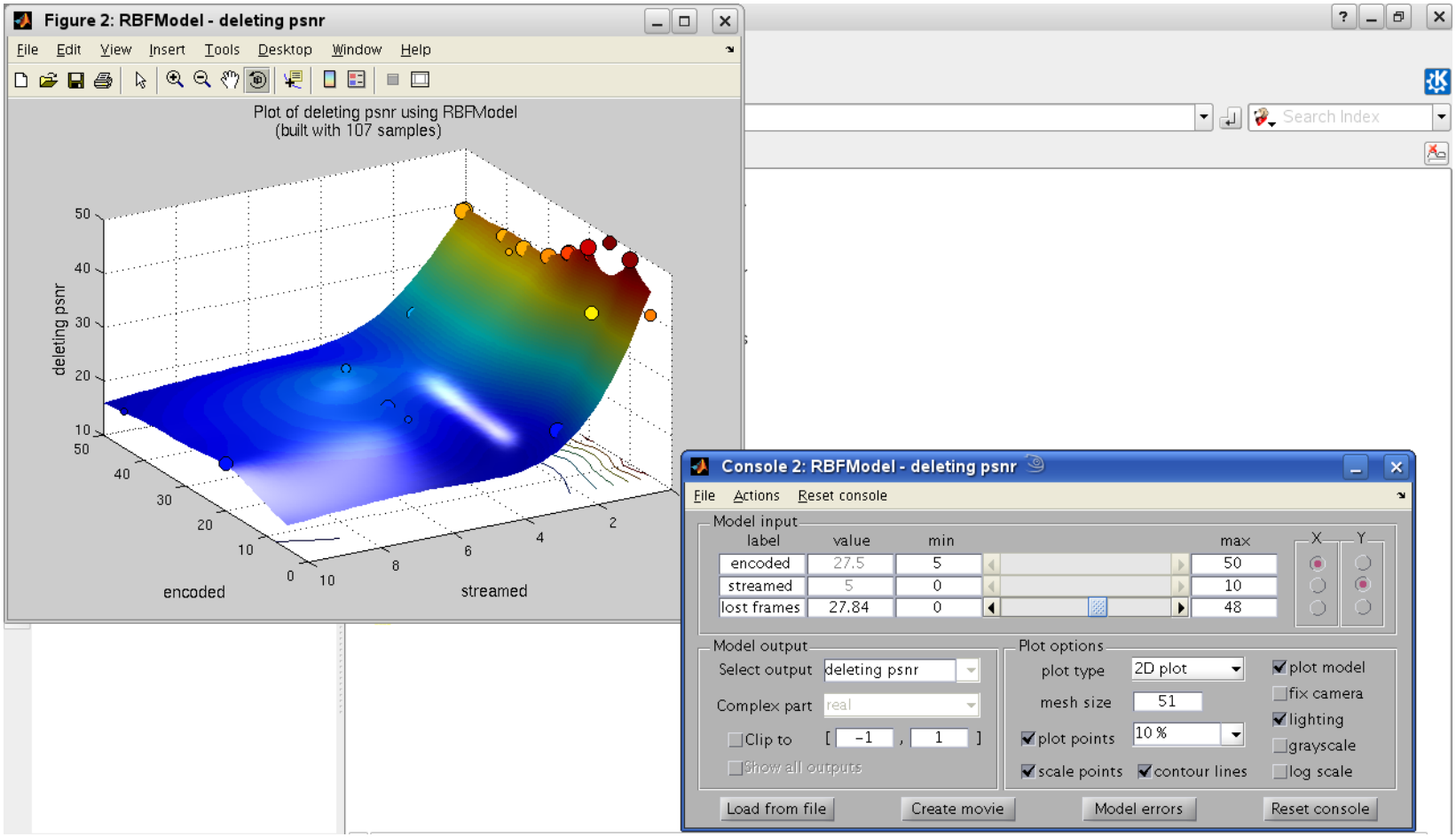
SUMO Toolbox – *Pluggability*



SUMO Toolbox

- straightforward configuration in XML
- modeling primitives can be combined in many ways
- sensible defaults but many 'expert' options available
  - **user remains in control**
- modular design to allow 3rd party extensions
- extensive logging of what is going on
- intermediate models (and plots) stored for further reference
- profiling framework to track modeling progress
- GUI Tool for easy visualization and data exploration

- **Available from v5.1**



Model type selection

- **however, which plugins to use?**
  - most important within adaptive modeling
- **many surrogate model types available:**
  - Rational functions, RBF models, Kriging, MLP, RBFNN, SVM, LS-SVM, regresison trees, splines, ...
- **which type to use?**
  - problem & data dependent
  - little theory available
    - **e.g., rational functions and EM data**
  - usually pragmatic
  - impossible to solve in general

Model type selection

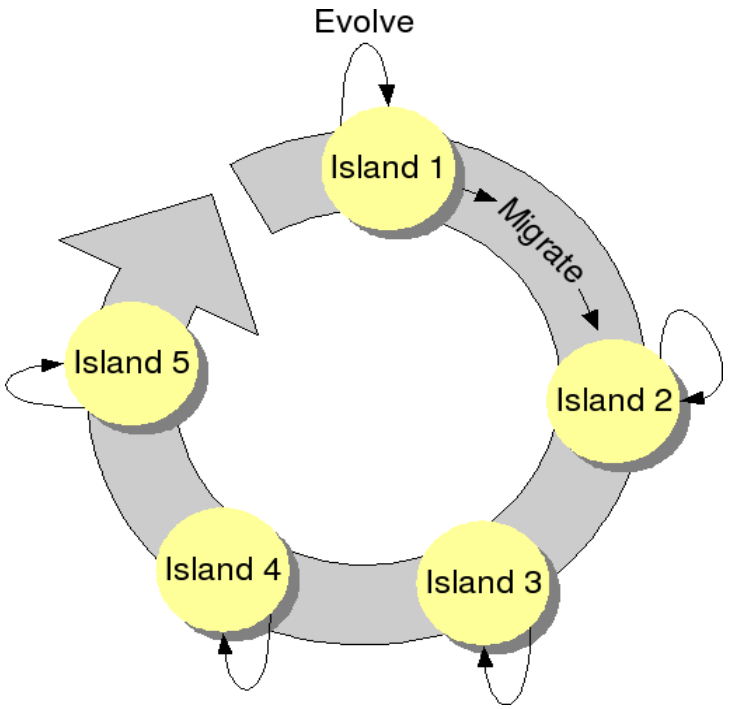
- **each model is characterized by parameter set θ**
- **how to select θ_i?**
  - by hand?
  - rule of thumb?
  - optimization algorithm?
    - **BFGS, GA, pattern search, simulated annealing, PSO, …**
- **optimization landscape is dynamic!**
  - cfr. adaptive sampling

Model type selection

- **SUMO Toolbox makes it trivial to run and compare different methods**
- **however, an idea…**
  - Tackle the model type selection and model parameter optimization problem in one speciated evolutionary algorithm
- **let evolution decide**
  - survival of the fittest
  - multiple final solutions possible
  - hybrid solutions possible (cfr. ensembles)
- **interesting population dynamics?**

Model type selection

- **island model (migration model)**
  - most natural
  - ring topology with different migration directions
  - NB: inter-model speciation, not intra-model



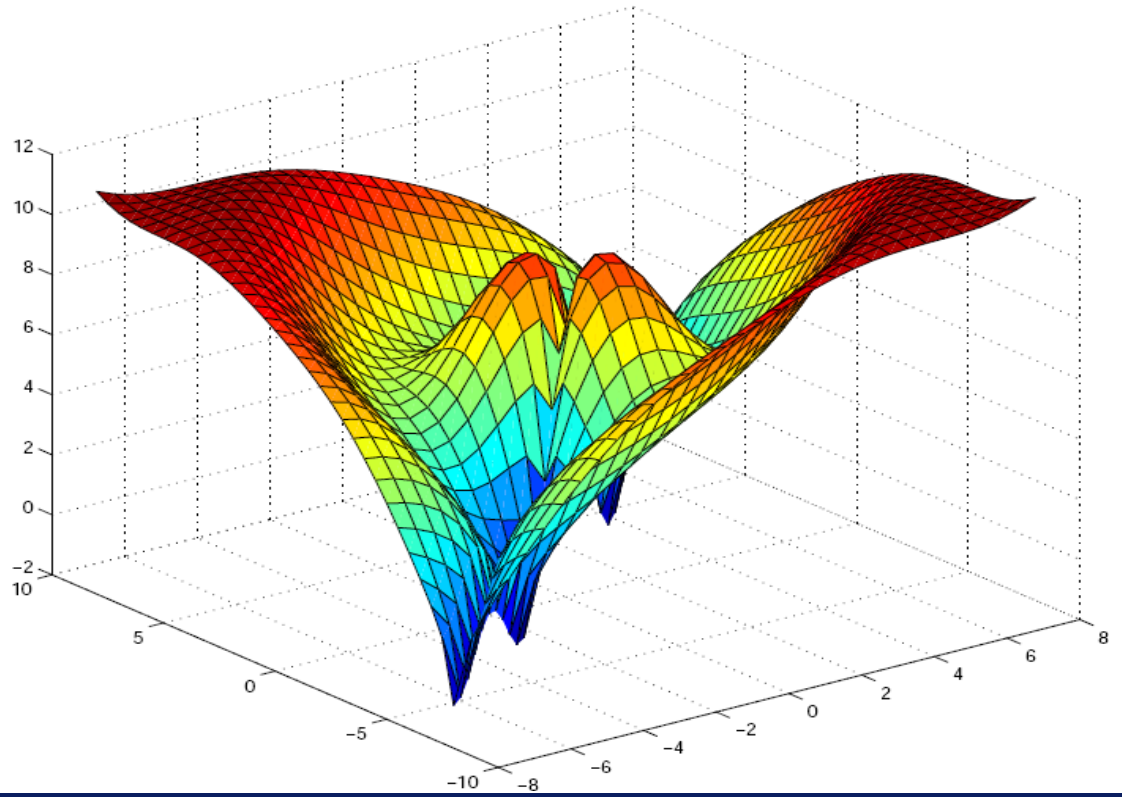
Model type selection

- **heterogeneous recombination**
  - Rational model x SVM = ???
- **use ensembles**
  - phenotypic (behavioral) recombination
  - avoid when possible
  - many ensemble methods
    - **use simple average**
    - **others can easily be used instead**

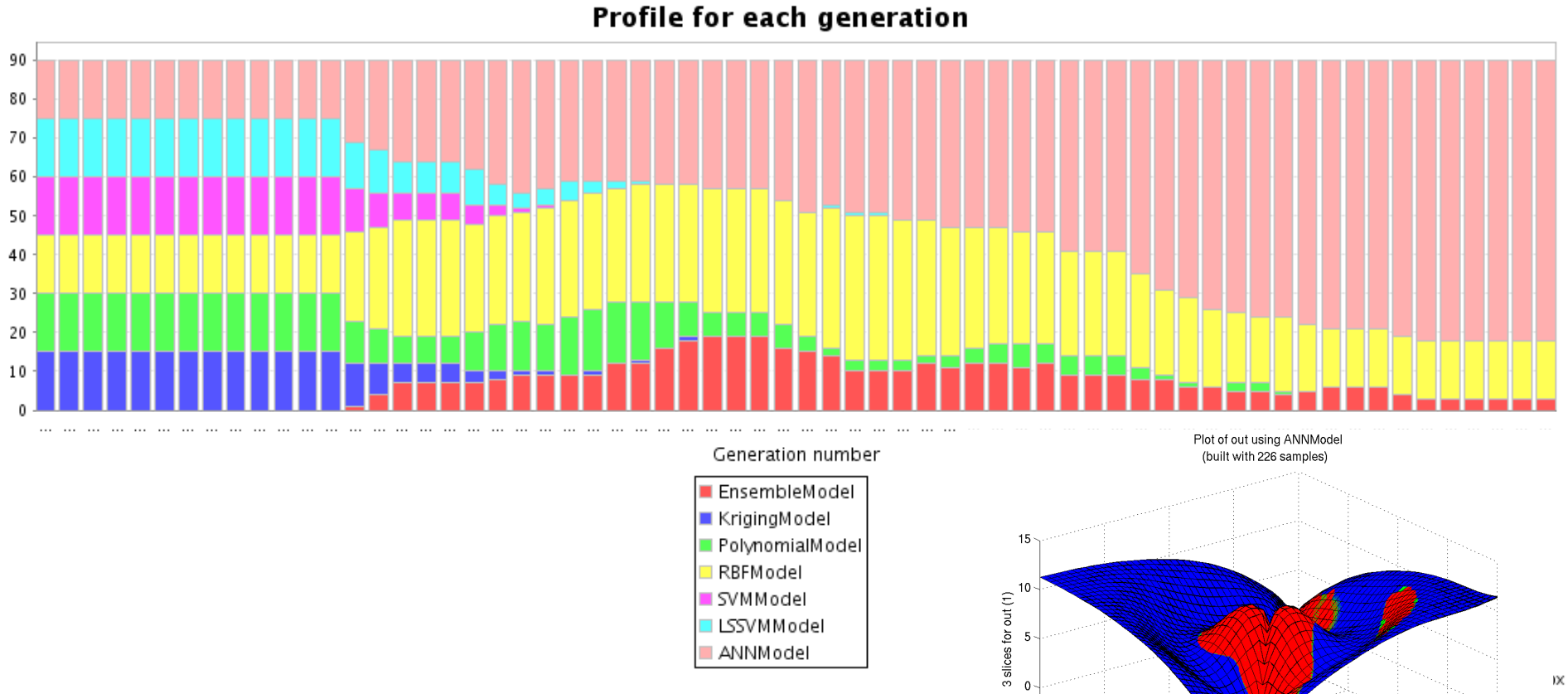
Model type selection

- **3D example** *(z=0)*

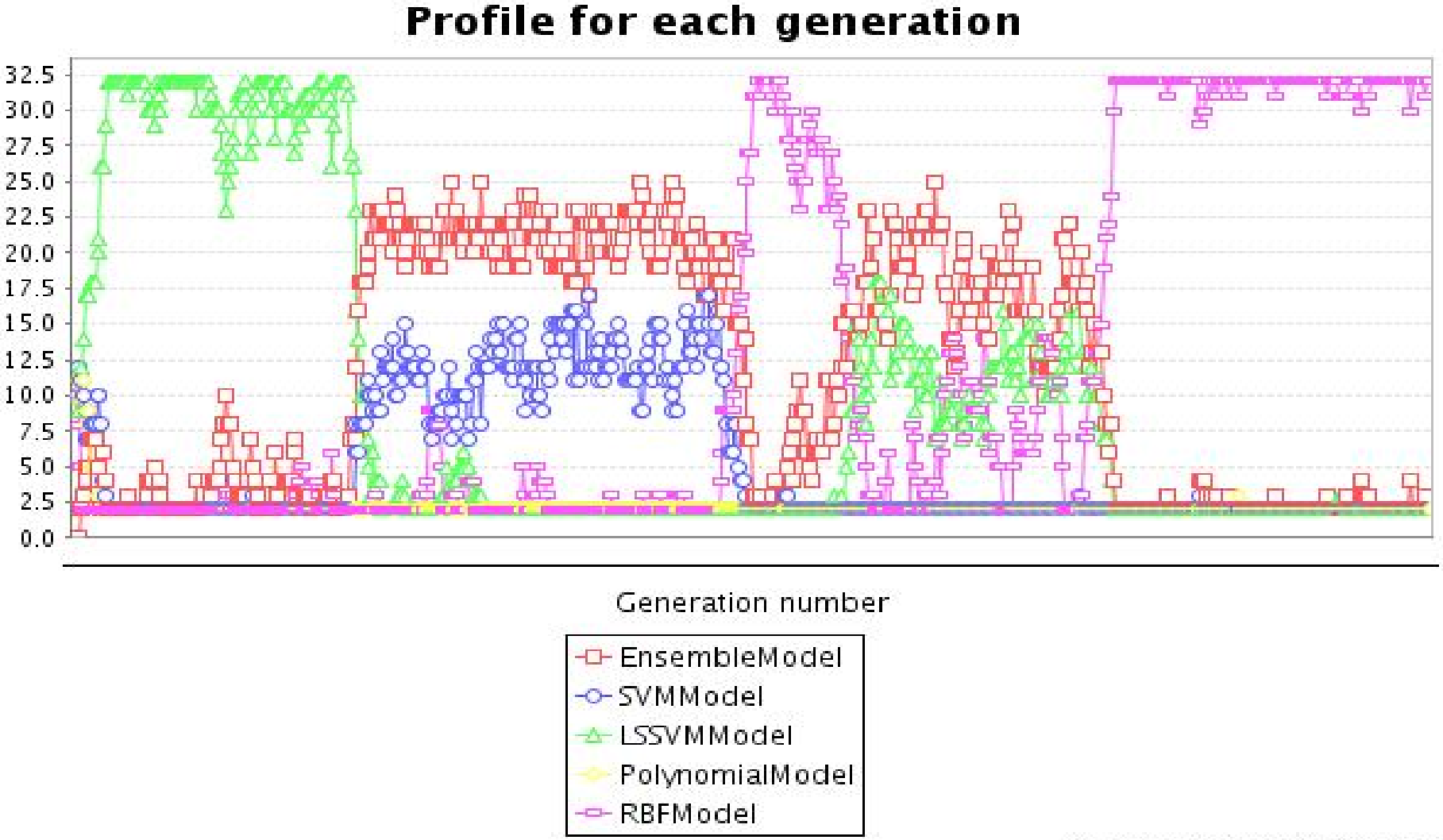
$$f(x,y,z) = 7\,\frac{\sin(\sqrt{x^2+y^2})+\epsilon}{\sqrt{x^2+y^2}} + 3|x-y|^{1/2} + 0.01z$$



Model type selection



Model type selection



Generated using the M3-toolbox

Model type selection

- **promising results**
- **computation time ≤ pure sequential**
- **delivers more insight**
- **however**
  - model type selection is not solved absolutely
    - **theoretically impossible without assumptions**
    - **GA meta parameters expected to be more robust**
  - sensitivity to migration/selection parameters?
  - constraints on reproducibility?

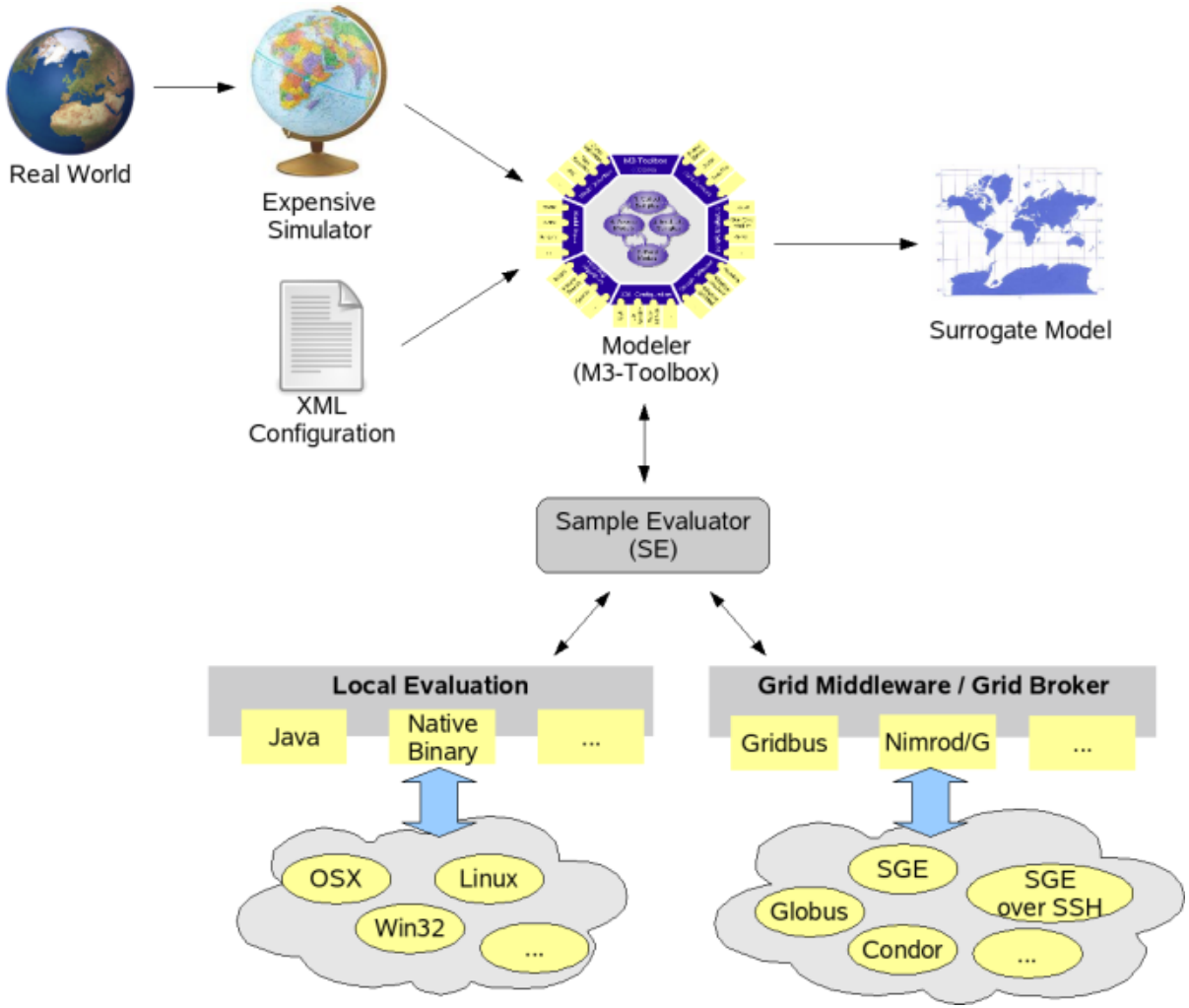
Integrating gridcomputing

- **simulations are expensive**
  - adaptive sampling
  - 1-time up front investment
  - provide interface to the grid
- **goal**
  - transparent integration
  - avoid middleware lock-in
  - hide grid details
- **integration on 3 levels**
  - resource level
  - scheduling level
  - service level

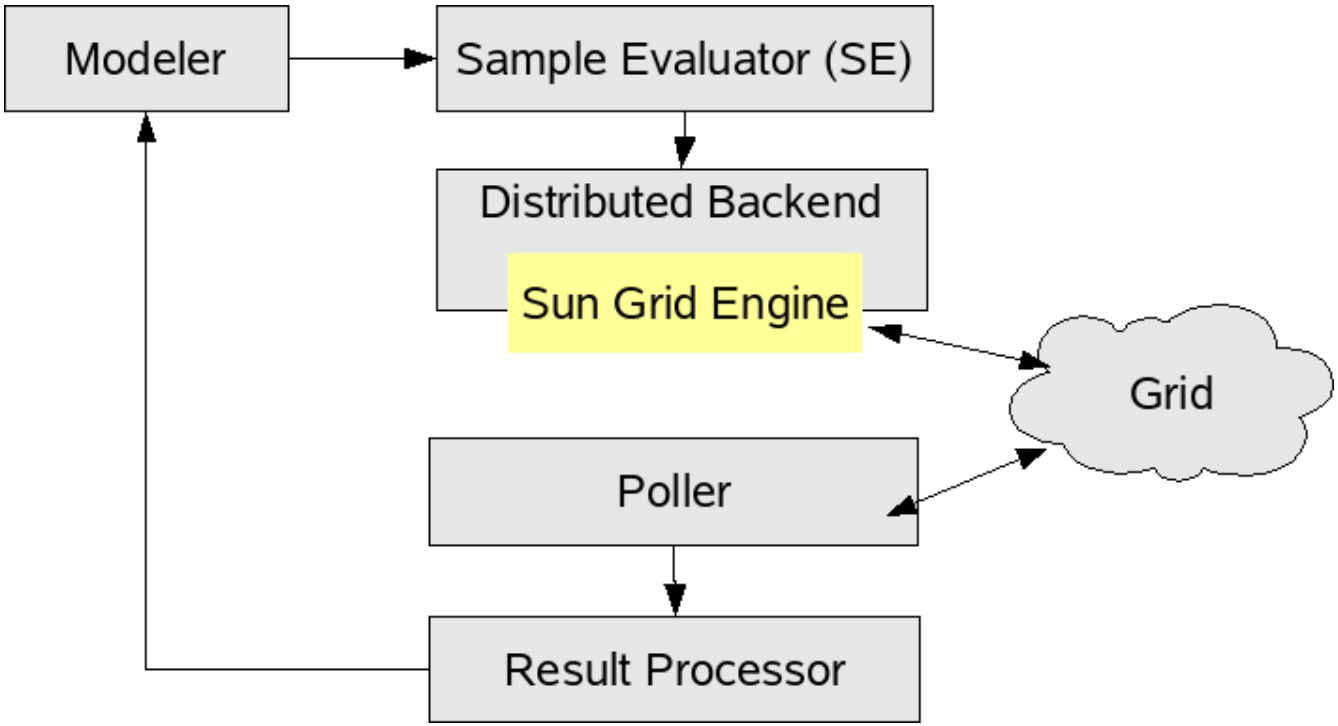
Integrating gridcomputing

- **resource level**
  - raw distribution
  - un simulations in parallel
- **SampleEvaluator abstraction**
  - cfr. flow chart
  - clean object oriented interface
  - translates modeler requests into middleware specific jobs
  - support multiple backends
    - **Sun Grid Engine**
    - **LCG middleware**
    - **APST**
    - **...**

Integrating gridcomputing



Integrating gridcomputing



Integrating gridcomputing

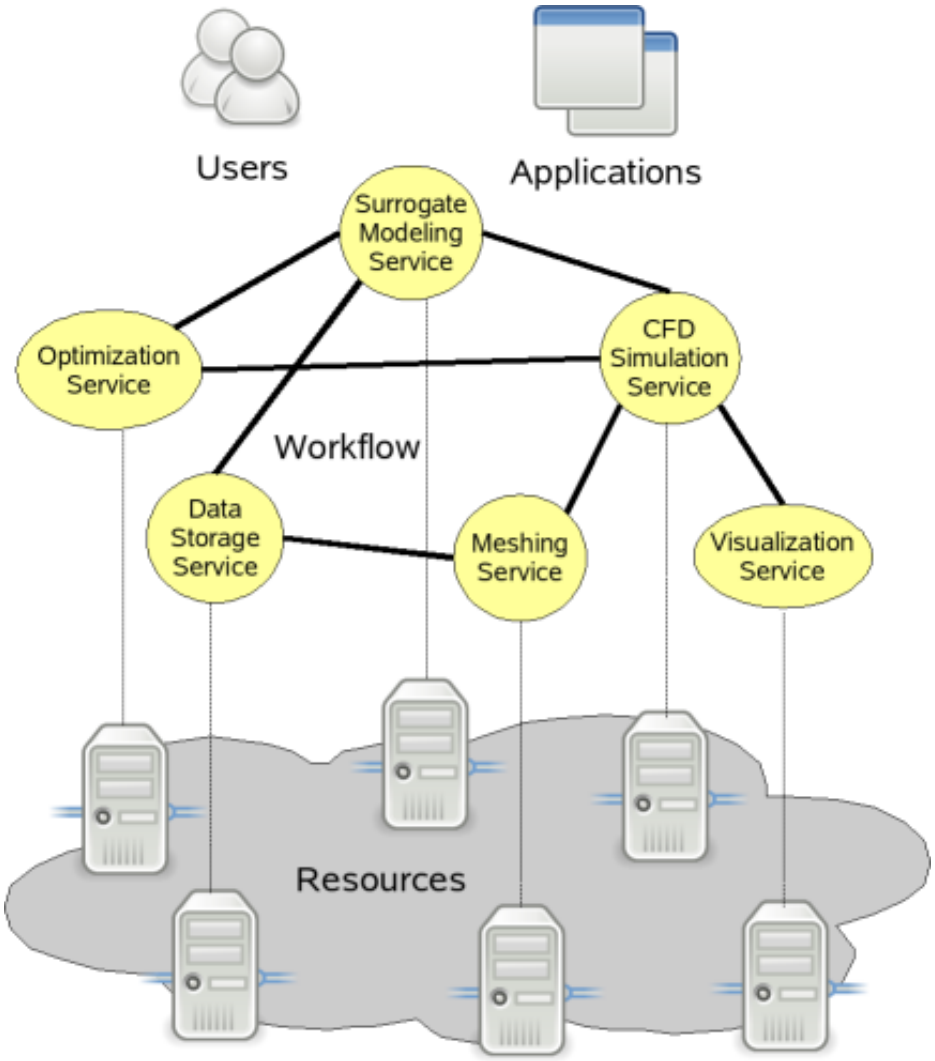
- **scheduling level**
  - data points have different priorities
    - **e.g., domain borders, optima, sparse regions, …**
  - compute resources are heterogeneous
  - resources are shared (dynamic!)
  - integrate grid resource information and modeling information into scheduling decisions

→ Application and resource aware scheduling

Integrating gridcomputing

- **service level**
  - integration as part of a larger service oriented architecture (SOA)
  - easy access and integration into the design process
    - **web browser, Jini, SOAP, …**
  - complicated workflows possible

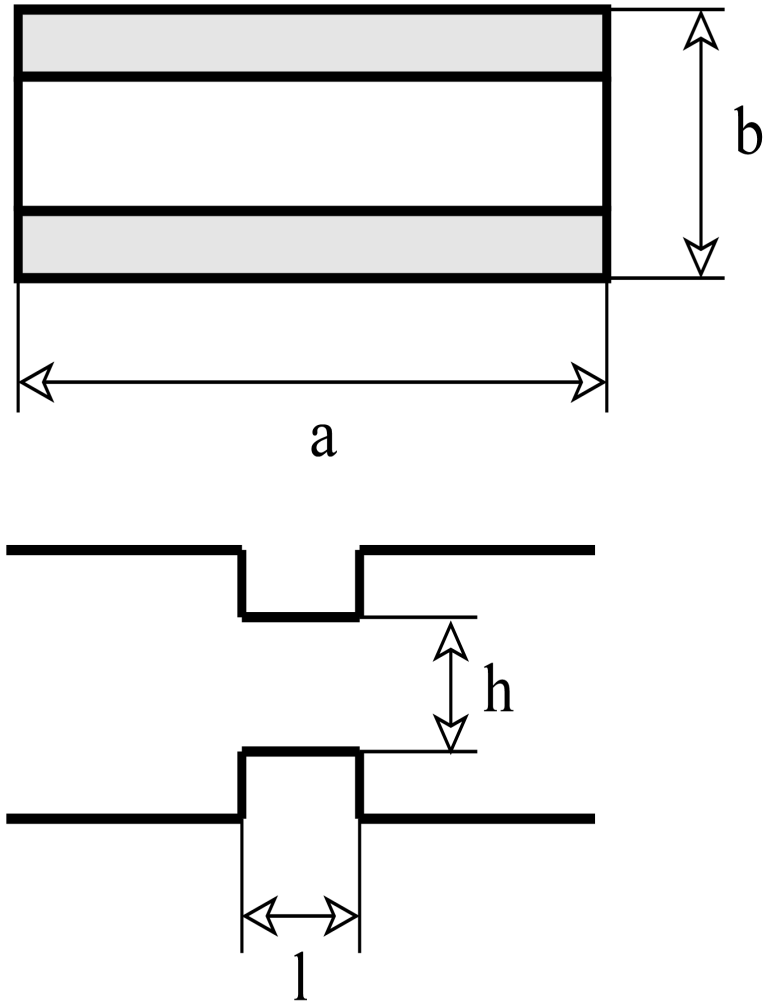
Integrating gridcomputing



Outline

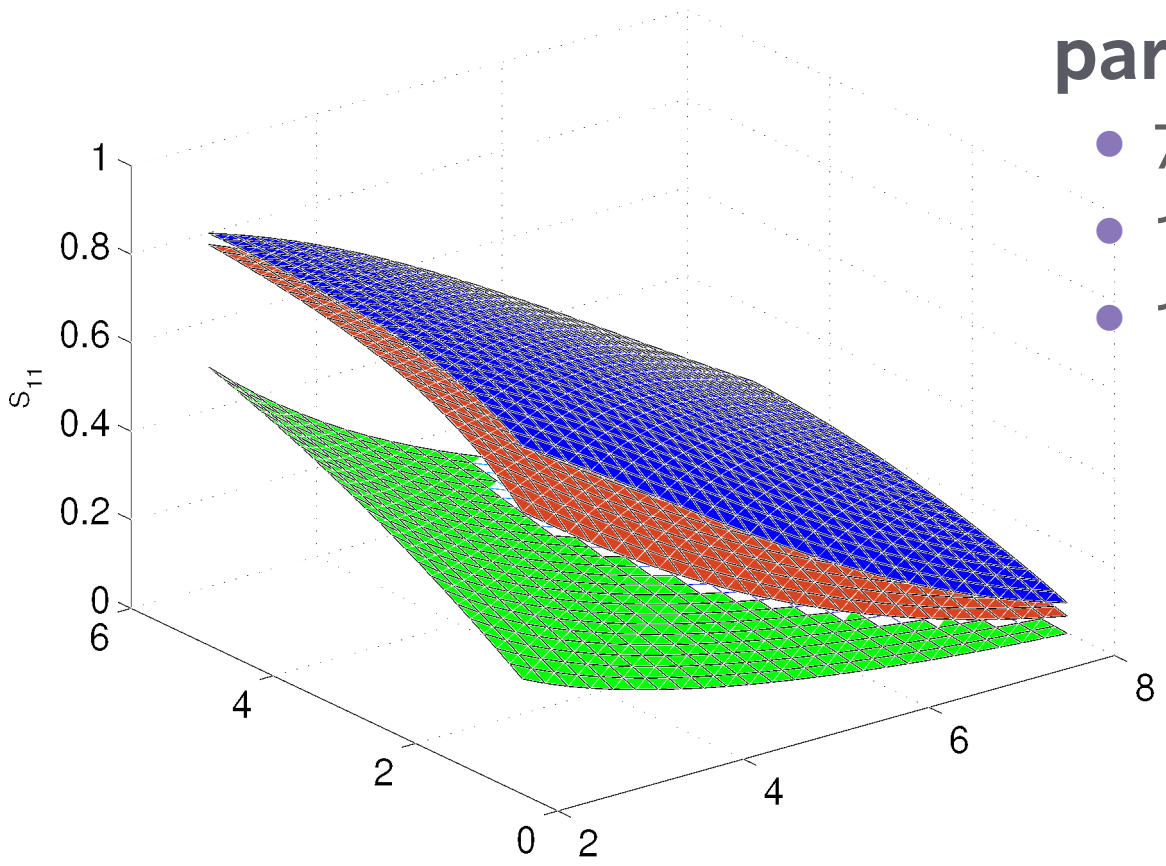
- Who are we ?
- Introduction
- Surrogate modeling
- SUMO Toolbox
- Examples
- Conclusions

Example - Step Discontinuity (SD)



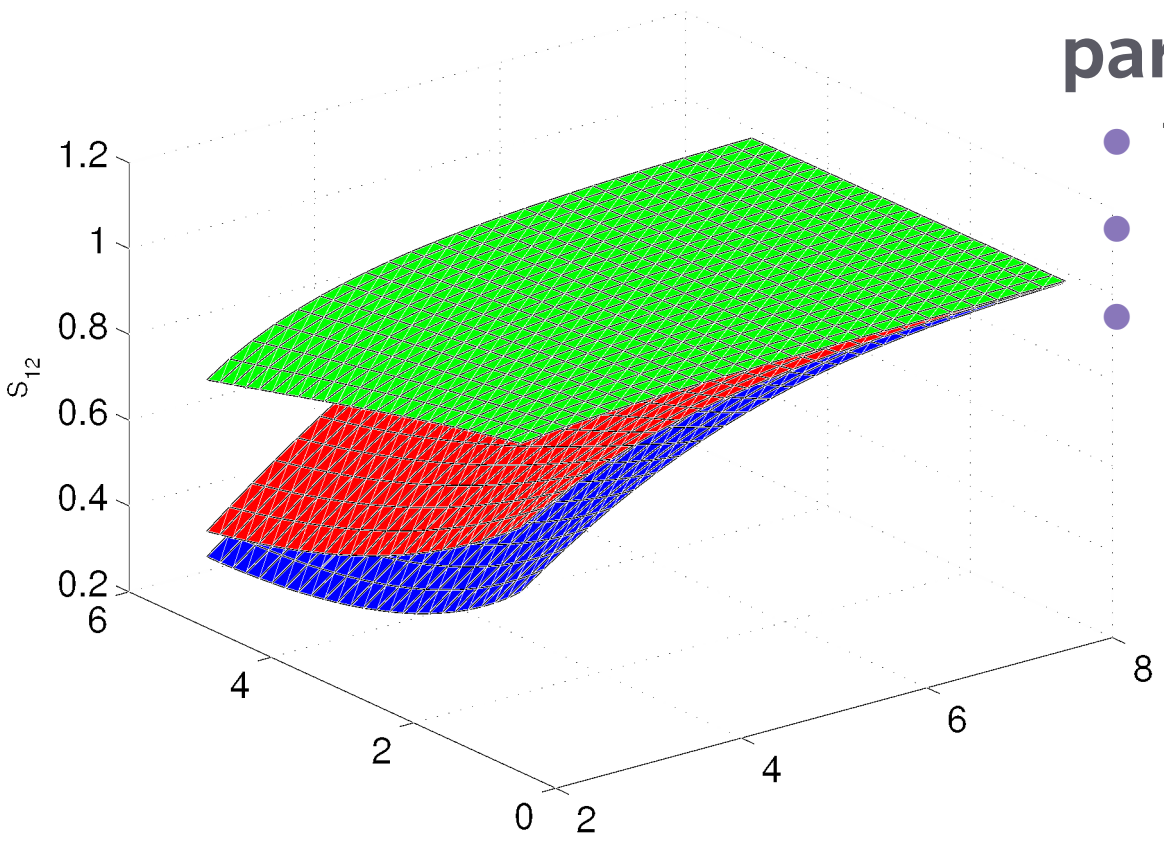
- **Step discontinuity in a rectangular waveguide**
  - Frequency : 7-13 GHz
  - Step length [l] : 2-8 mm
  - Gap height [h] : 0.5-5 mm
  - Waveguide width [a] : 22.86 mm
  - Waveguide height [b] : 10.16 mm
- **Distributed backend:**
  - Remote 256 node SGE cluster

SD : First output



- **S11 scattering parameter**
  - 7GHz (green)
  - 10GHz (red)
  - 13GHz (blue)

SD : Second output



- **S12 scattering parameter**
  - 7GHz (green)
  - 10GHz (red)
  - 13GHz (blue)

Simulator configuration

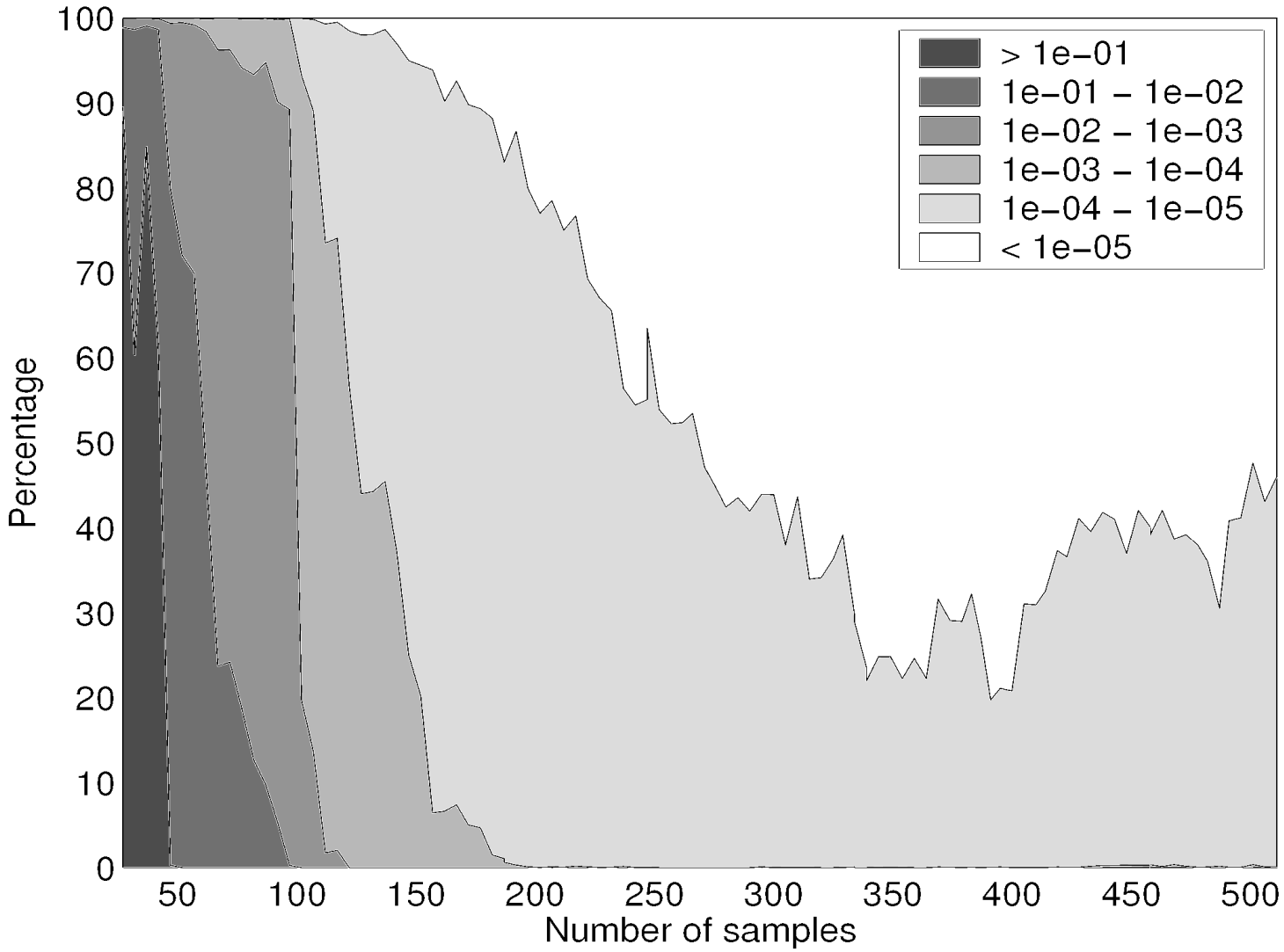
```
<Simulator>
    <Name>Step Discontinuity</Name>
    <InputParameters>
        <Parameter name="frequency" type="real"/>
        <Parameter name="gapHeight" type="real"/>
        <Parameter name="stepLength" type="real"/>
    </InputParameters>

    <OutputParameters>
        <Parameter name="S11" type="complex"/>
        <Parameter name="S12" type="complex"/>
    </OutputParameters>

    <Implementation>
        <Executable platform="unix" arch="amd64">StepDiscontinuity</Executable>
        <DataFiles>...</DataFiles>
    </Implementation>
</Simulator>
```

Toolbox configuration

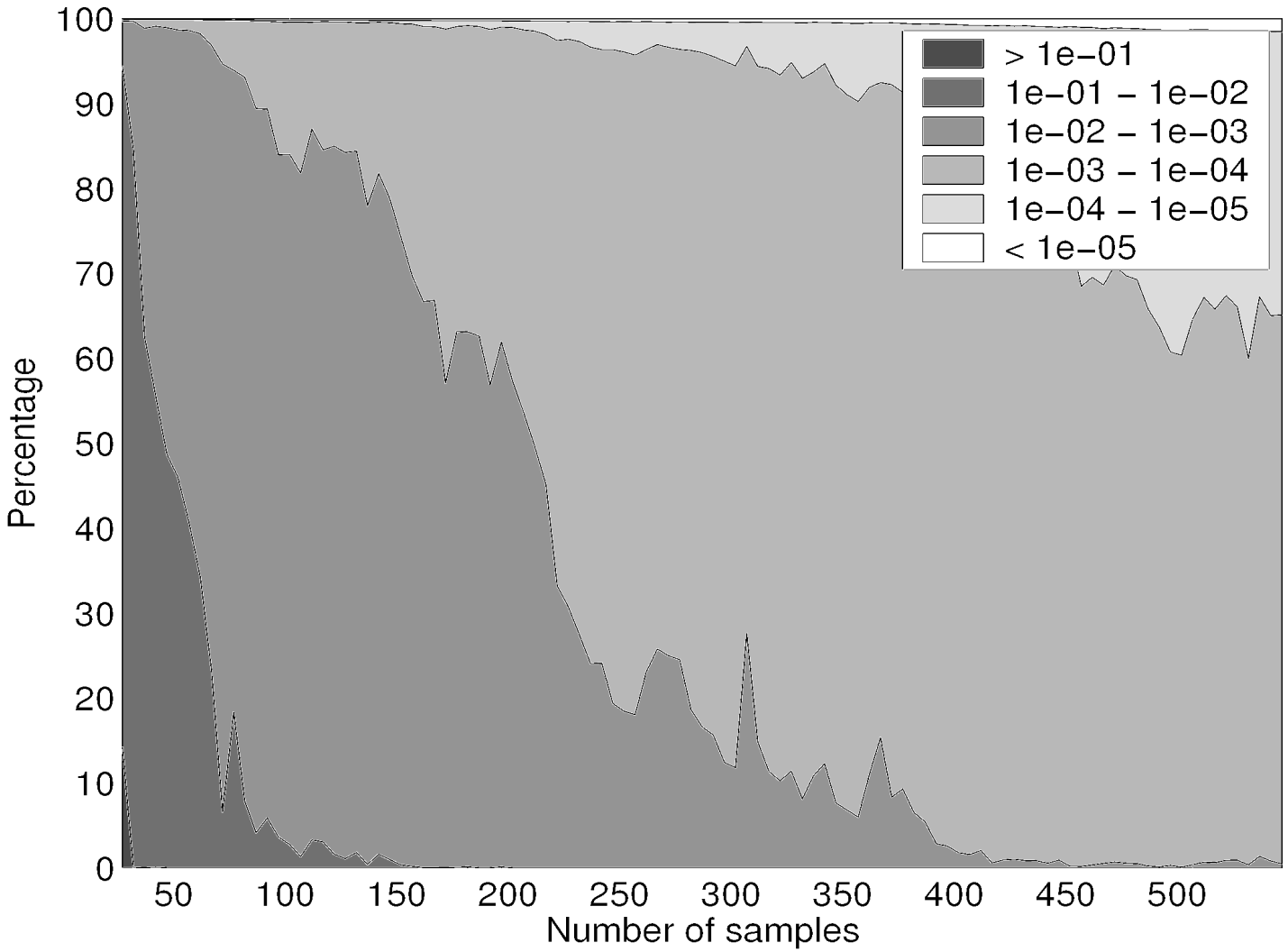
```
<ToolboxConfiguration version="5.1">
    <Plan>
        ...
        <SampleSelector>gradient</SampleSelector>
        <Measure type="CrossValidation" target=".0001" errorFcn=”absoluteRMS” use="on" />
        ...
        <Run>
            <Simulator>StepDiscontinuity.xml</Simulator>
            <SampleEvaluator>sge</SampleEvaluator>

            <Outputs>
                <Output name="S11" complexHandling=”complex”>
                    <AdaptiveModelBuilder>poly</AdaptiveModelBuilder>
                </Output>
                <Output name="S12" complexHandling=”split”>
                    <AdaptiveModelBuilder>kriging</AdaptiveModelBuilder>
                </Output>
                <Output name="S11,S12" complexHandling=”modulus”>
                    <AdaptiveModelBuilder>anngenetic</AdaptiveModelBuilder>
                </Output>
            </Outputs>
```

Modeling progres - Movie

Results – Rational models

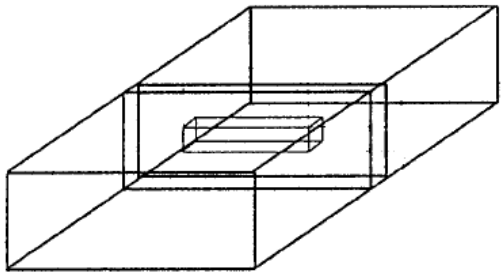


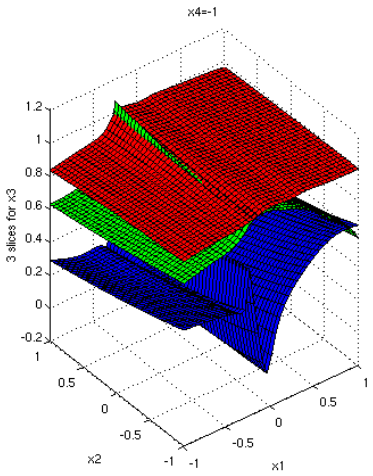
Results – RBF models

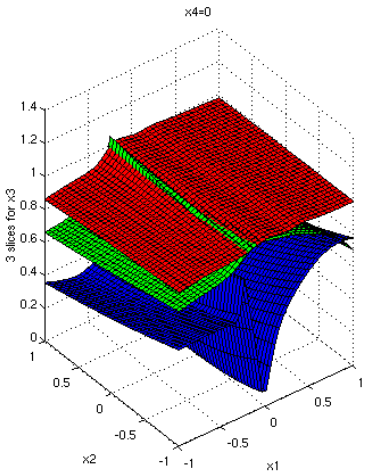


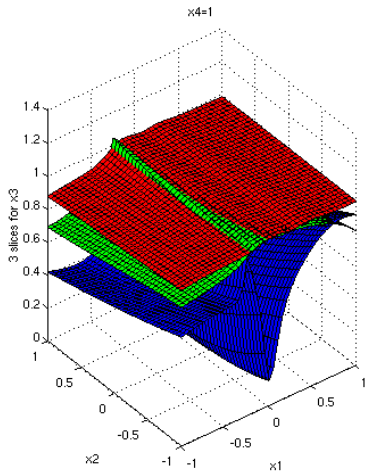
Electromagnetic example

- iris in rectangular waveguide (From Lamecki 2005)
  - Simulation of scattering parameters
    - Input : frequency, iris height, length, width,
    - Output : S11, S12









Chemistry example

- methane – air combustion (From Ihme 2007)
  - Simulation of temperature
    - Input : mixture fraction variable z, reaction progress variable c
    - Output : temperature

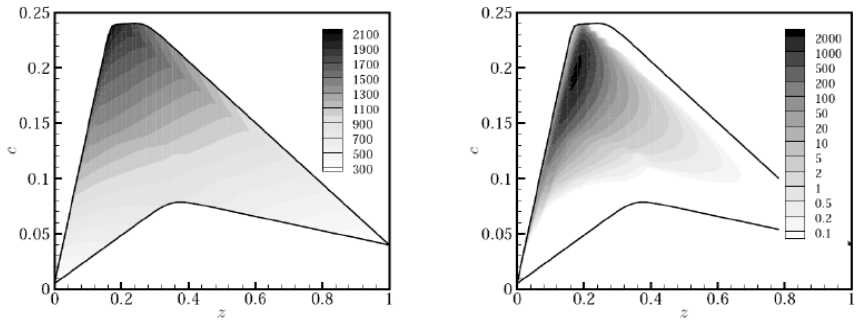
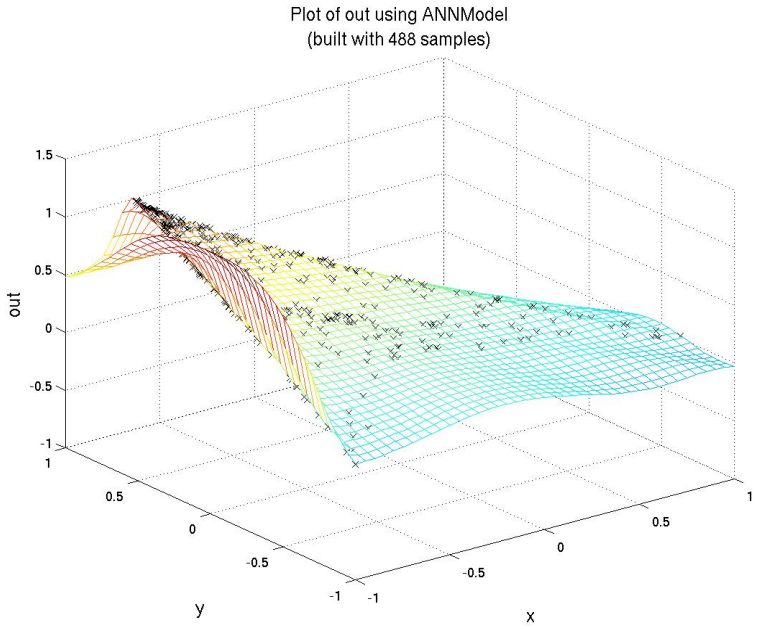


Figure 5.3: Solution of the steady laminar flamelet equations as a function of mixture fraction z and progress variable c; (a) temperature (K) and (b) chemical source term $(kg/(m^3 s)$ (Source: [77])

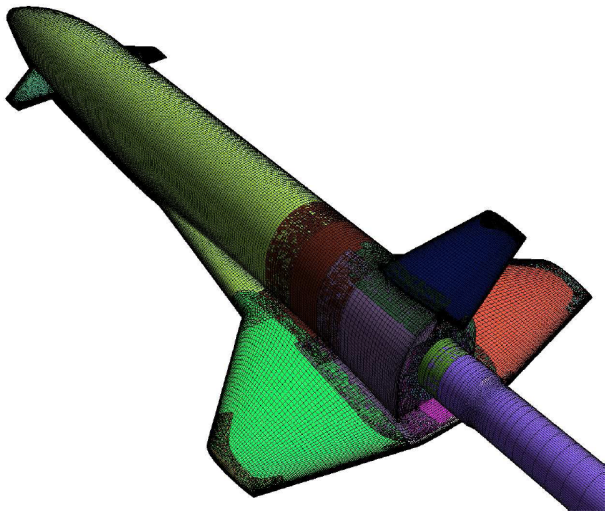


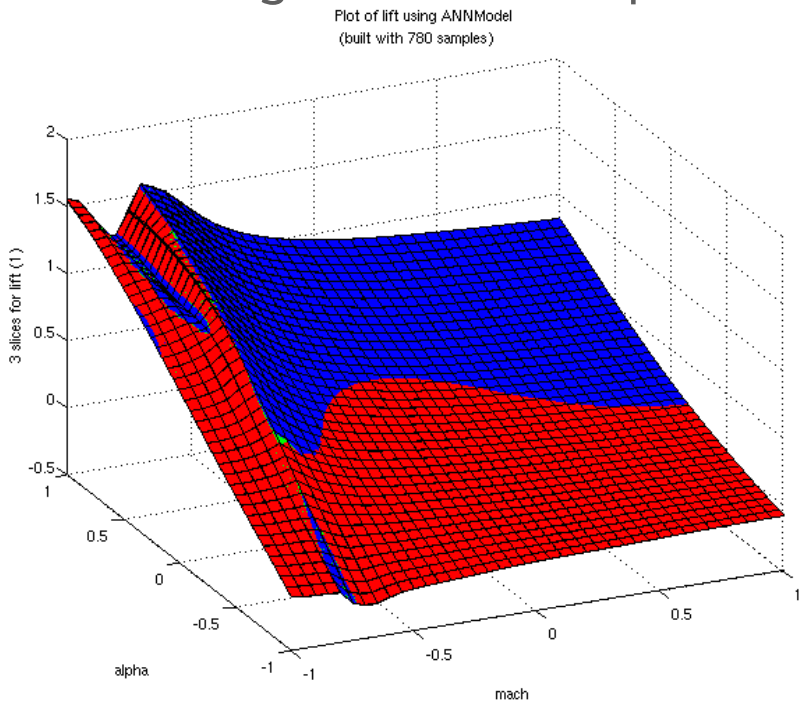
Aerodynamics example

- re-usable Langly Glide Back Booster (LGBB)

  (From Gramancy 2004 / NASA)

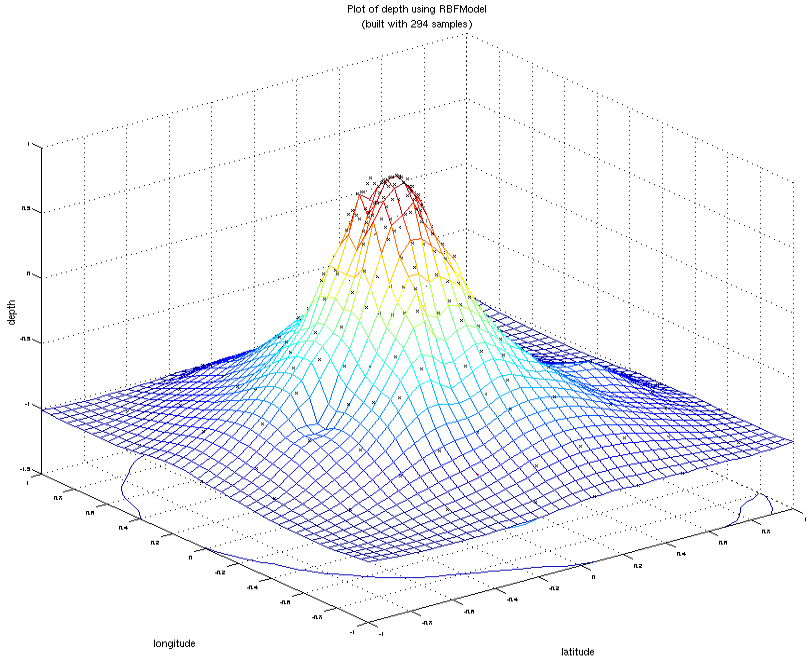
  - Simulation of lift
    - Input : mach number, angle of attack, slip slide angle
    - Output : lift





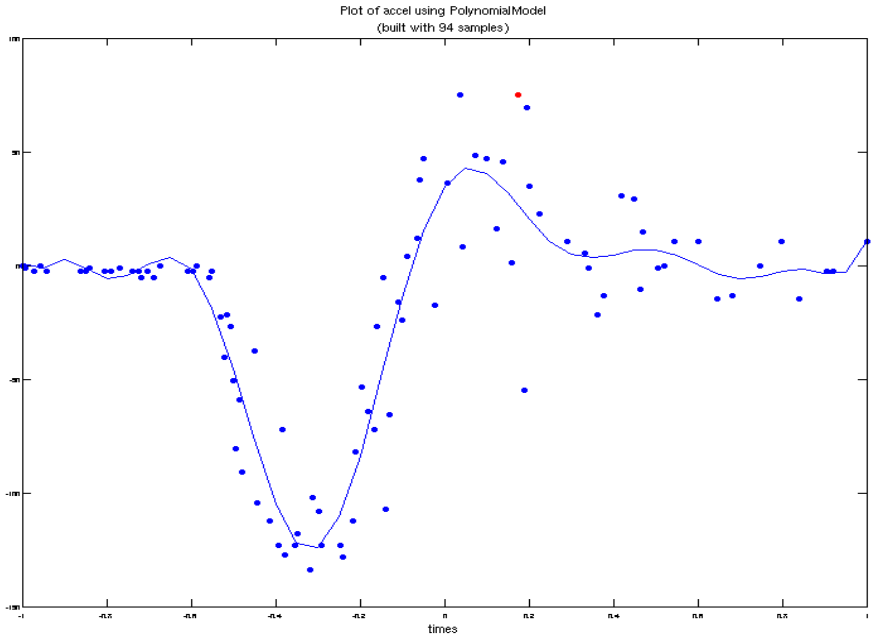
Geology example

- seamount (From Parker 1987)
  - Elevation data from a submerged mountain
    - Input : latitude, longitude
    - Output : depth



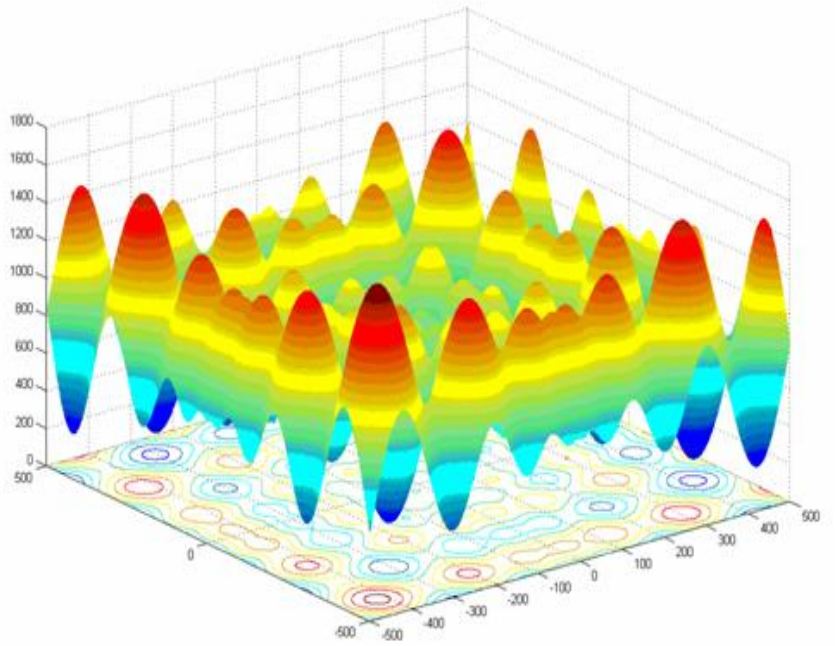
Automotive example

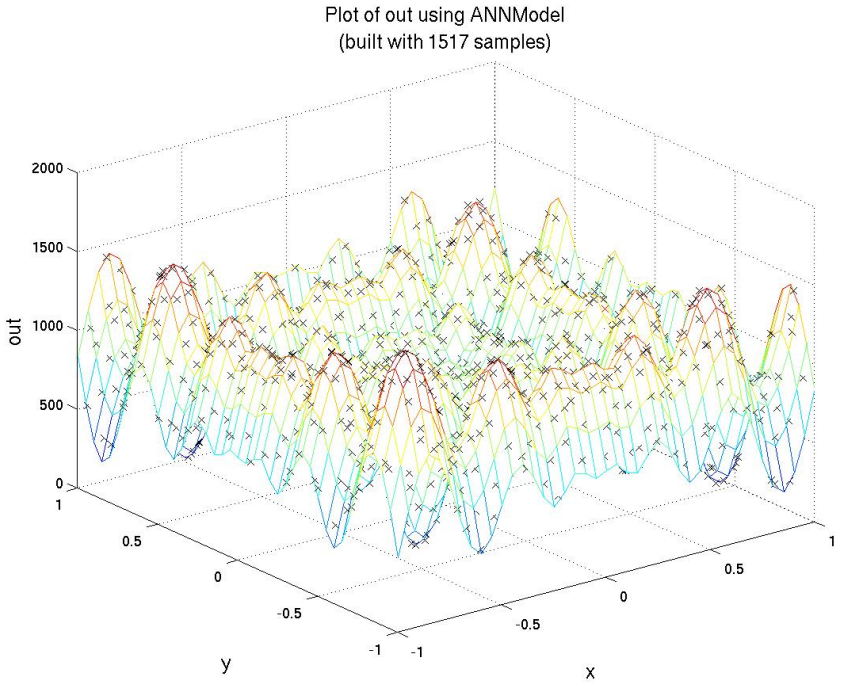
- motorcycle accident (From Silverman 1987)
  - Simulate a motorcycle crash against a wall
    - Input : time in milliseconds since impact.
    - Output : the recorded head acceleration (in g)



Synthetic example

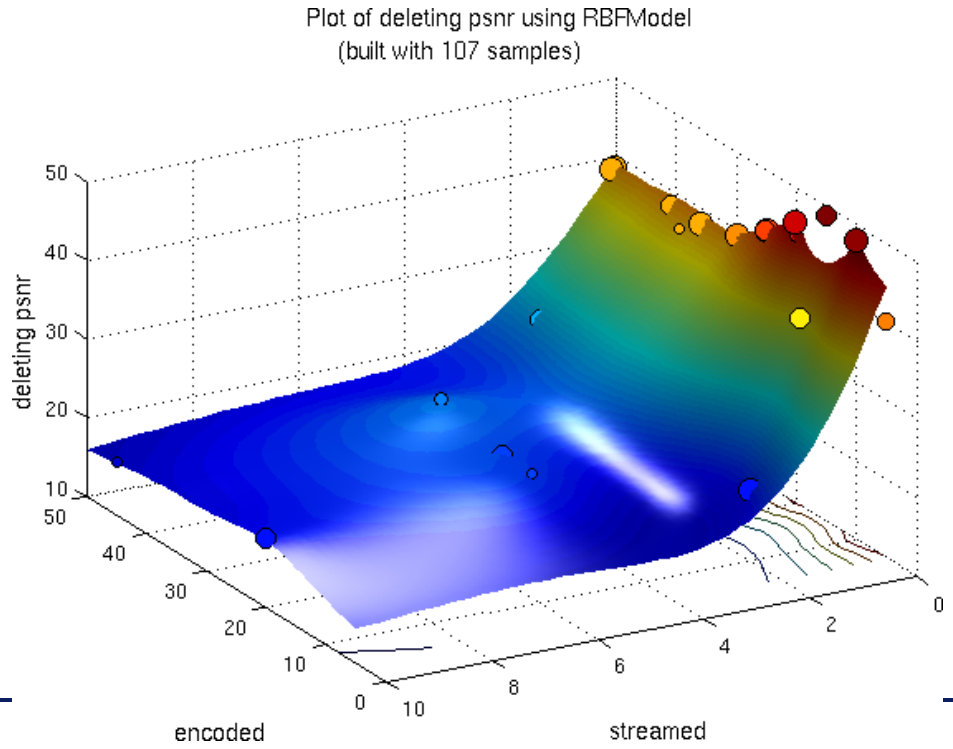
- Schwefel Function
  - Classic 2D test function form optimization





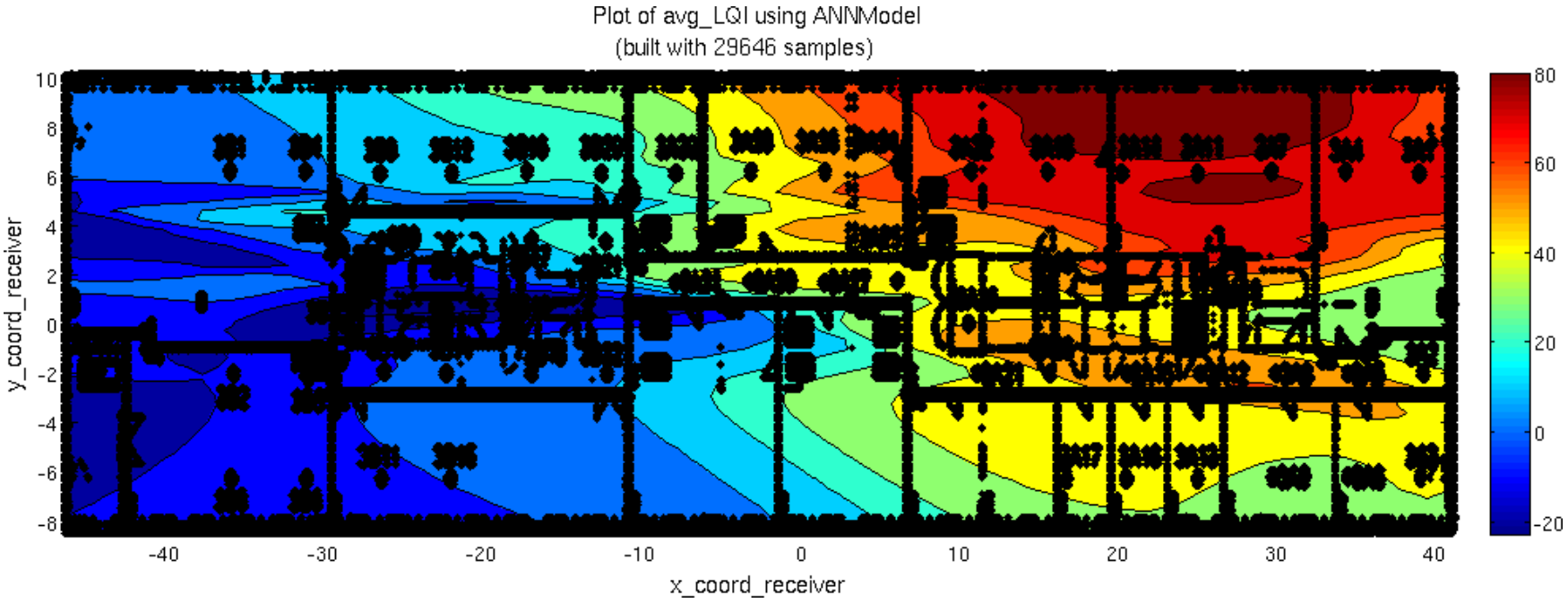
Multimedia example

- **Video quality data** (From Nick Vercammen, IBBT)
  - How does streaming/encoding affect quality
    - Input : encoding, transmission parameters
    - Output : quality metric



Networking example

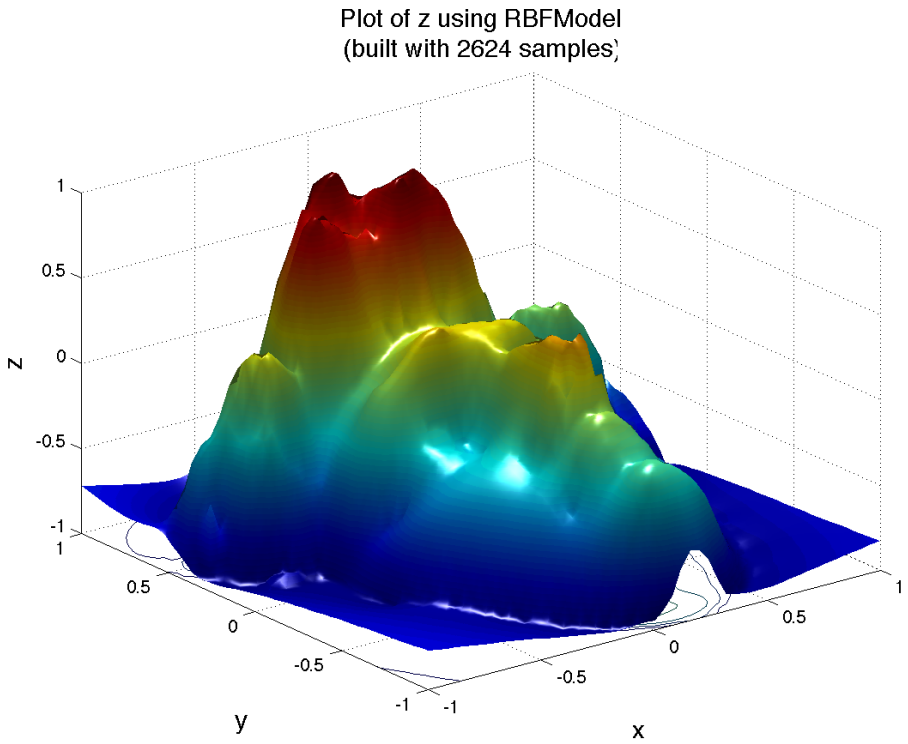
- **Wireless sensor data** (From Sensor Lab, IBBT)
  - Model reception quality
    - Input : sender/receiver coordinates
    - Output : reception quality metric



David

- **David data**

  (From the Digital Michalangelo project, Stanford University)



Outline

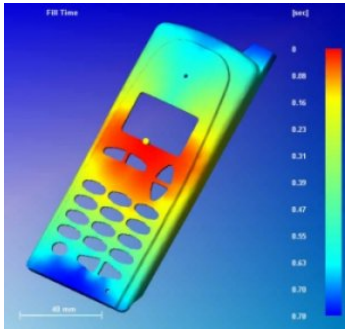
- Who are we ?
- Introduction
- Surrogate modeling
- SUMO Toolbox
- Examples
- Conclusions

Conclusions

- compact scalable surrogate models
  - metamodels, response surface models

SUMO

- fully automated
  - adaptive model selection
  - adaptive sample selection
  - distributed computing
  - (optimization)



- SUrrogate MOdeling (SUMO) Toolbox
  - easy to setup and run different modeling experiments
  - natural platform for benchmarking different techniques
  - download from http://www.sumo.intec.ugent.be

Questions

- **Questions ?**