

Visualization and machine learning

Thomas Torsney-Weir

About me

About me

- Born: Allentown, PA



About me

- Born: Allentown, PA
- Undergrad: Georgetown



About me



- Born: Allentown, PA
- Undergrad: Georgetown
- Finance: NYC



About me



- Born: Allentown, PA
- Undergrad: Georgetown
- Finance: NYC
- Master's: Simon Fraser



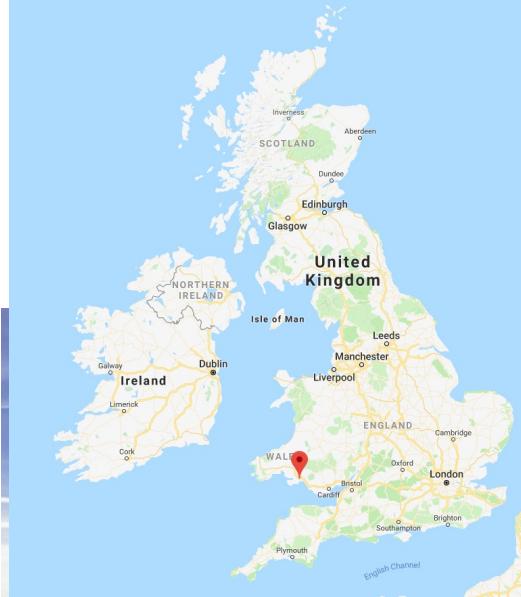
About me



- Born: Allentown, PA
- Undergrad: Georgetown
- Finance: NYC
- Master's: Simon Fraser
- PhD/Postdoc: University of Vienna

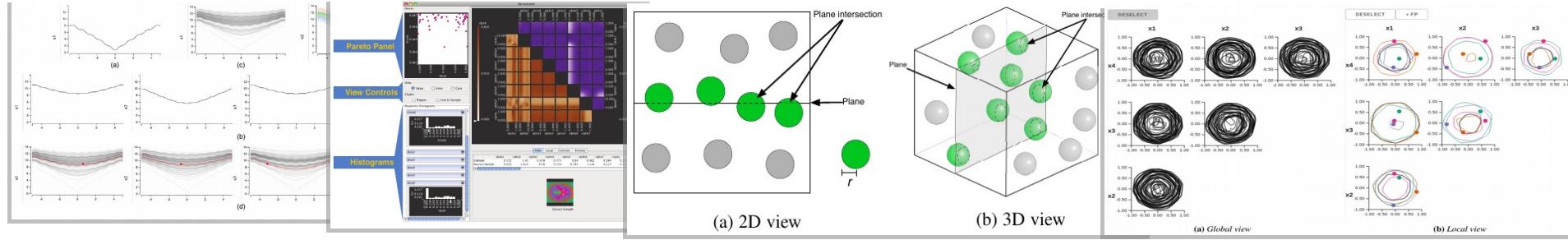


About me

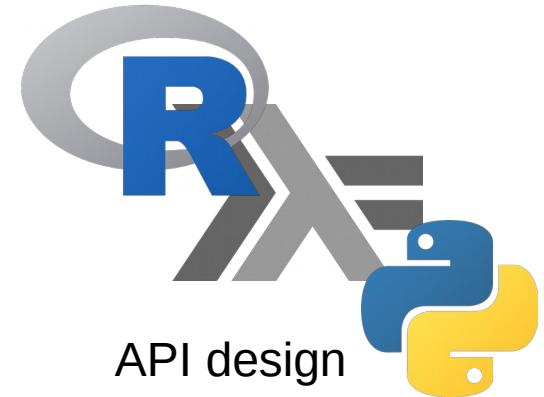


- Born: Allentown, PA
- Undergrad: Georgetown
- Finance: NYC
- Master's: Simon Fraser
- PhD/Postdoc: University of Vienna
- Next: Swansea University!

My research

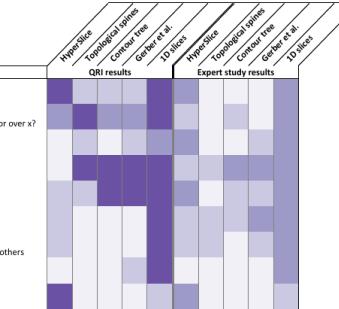


Multi-dimensional spaces

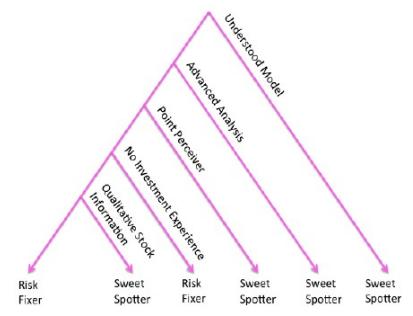


API design

Task	Task description for discrete data items from [AES05]	Our adaption to continuous scalar functions
Retrieve value	"Given a set of specific cases, find attributes of those cases"	Given an x , what is the function value?
Filter	"Given some concrete conditions attribute values, find data cases satisfying those conditions."	For what parameter values is the function equal or over x ?
Compute derived value	"Given a set of data cases, compute an aggregate numeric representation of the data cases"	Summary statistics: variance, mean, SA Find local/global min/max
Find extremum	"Given a set of data cases possessing an extreme value of an attribute over its range within the data set"	What is the range of possible outputs?
Determine range	"Given a set of data cases and an attribute of interest, find the span of values within the set"	What types of shapes do the manifolds have
Characterize distribution	"Given a set of data cases and a quantitative attribute of interest, characterize the distribution of that attribute's values over the set"	Do areas of the manifold have shapes unlike any others
Find anomalies	"Identify any anomalies within a given set of data cases with respect to a given relationship or expectation, e.g. statistical outliers"	Areas of the manifold have similar shapes
Cluster	"Given a set of data cases, find clusters of similar attribute values"	1D vs 2D relationships
Correlate	"Given a set of data cases and two attributes, determine useful relationships between the values of those attributes"	



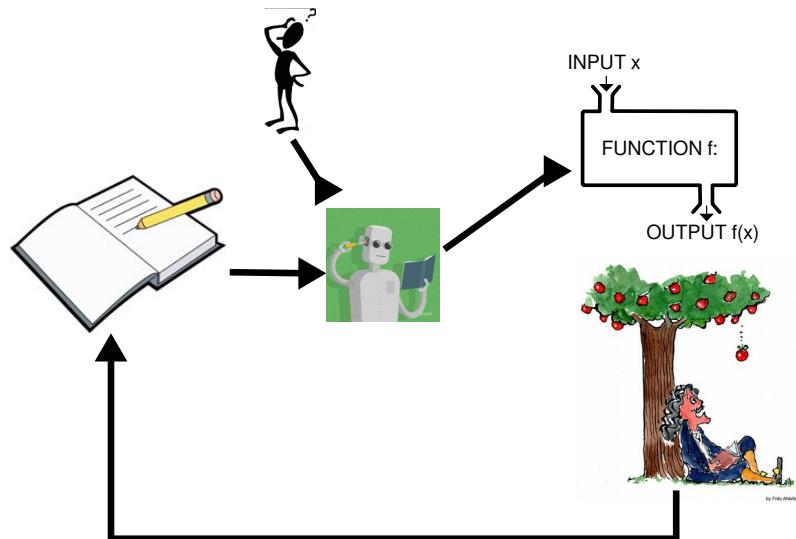
Users/tasks



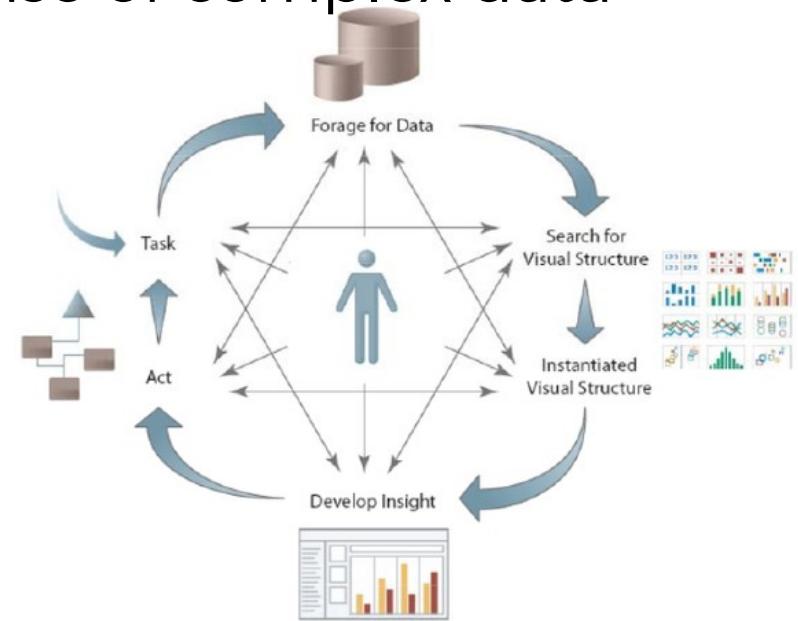
Visualization and machine learning

Visualization and machine learning

similar goals: make sense of complex data



Machine learning



Visualization

Morton, Kristi, Ross Bunker, Jock Mackinlay, Robert Morton, and Chris Stolte. "Dynamic workload driven data integration in tableau," Proceedings of the 2012 acm sigmod international conference on management of data. 2012.

Who helps whom?

both!

- Visualization helps machine learning:
evaluating models
- Machine learning helps visualization:
machine learning for embedded analysis

Outline

- Visualization helping machine learning
- Machine learning helping visualization
- What does the future hold?

Visualization helping machine learning

How do they work together?

- Building models
 - Meta parameters
 - Parameter selection
- Validating models
- Understanding models

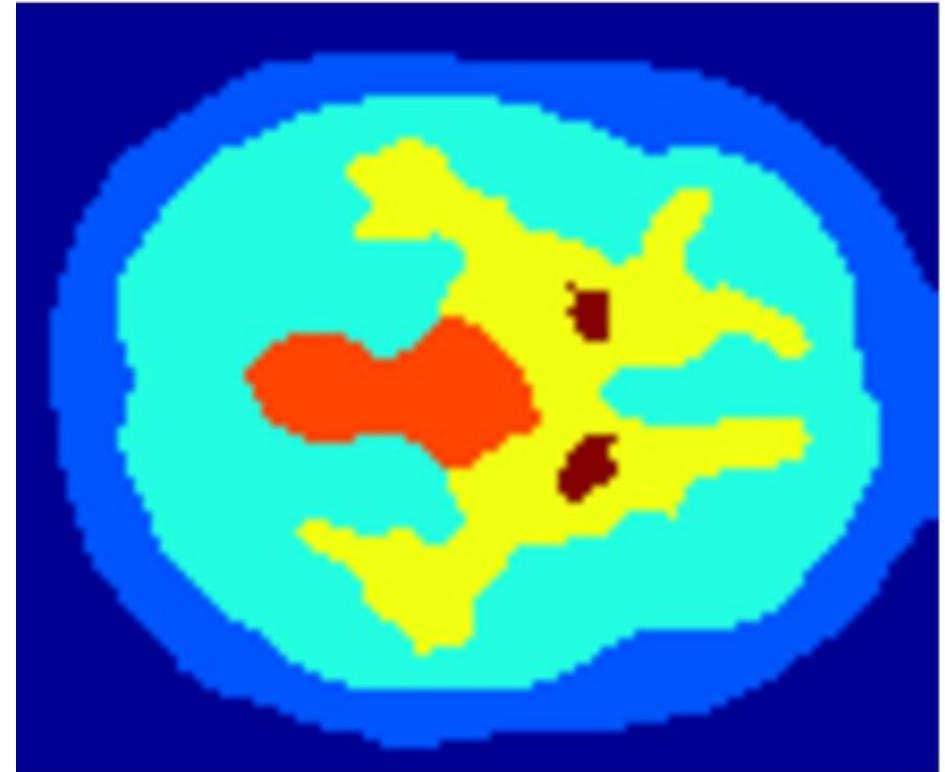
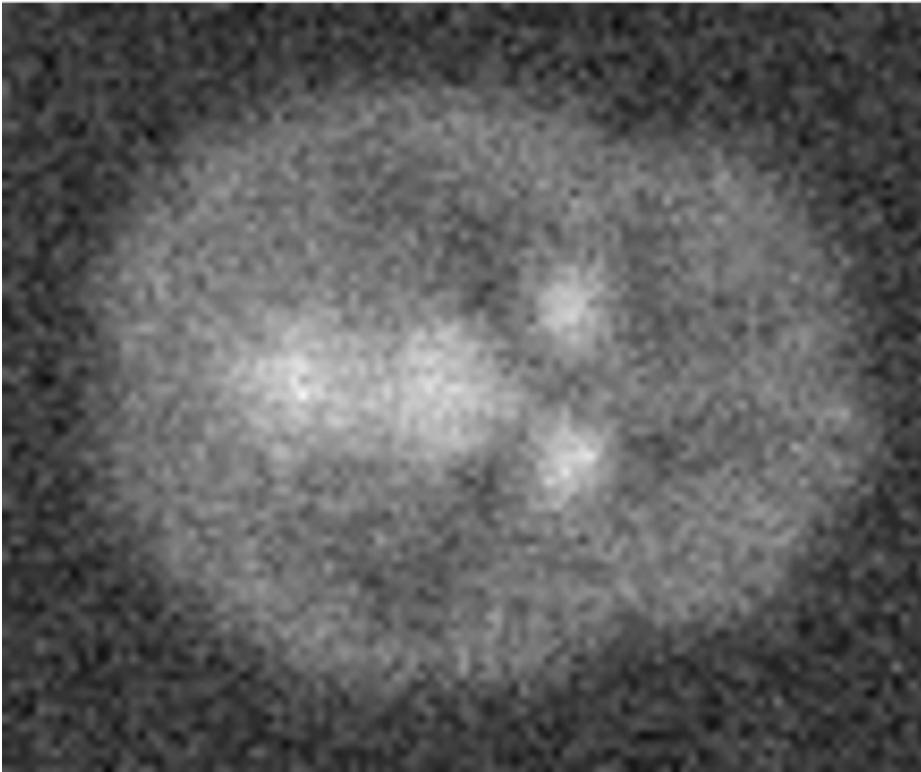
Meta parameters

What are meta parameters?

Control how learning takes place

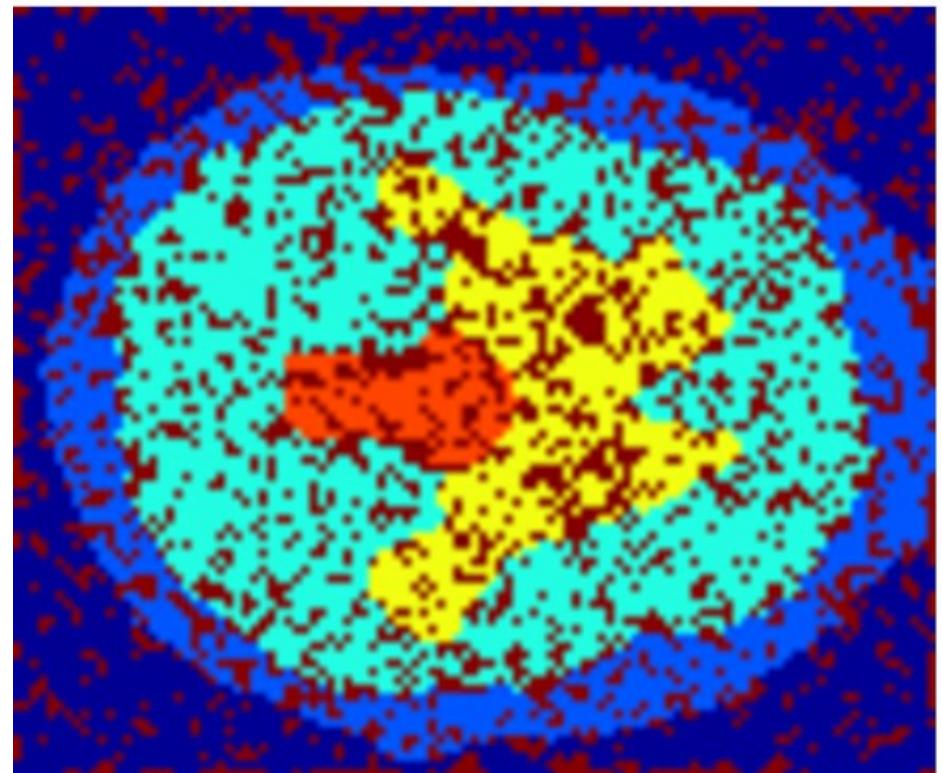
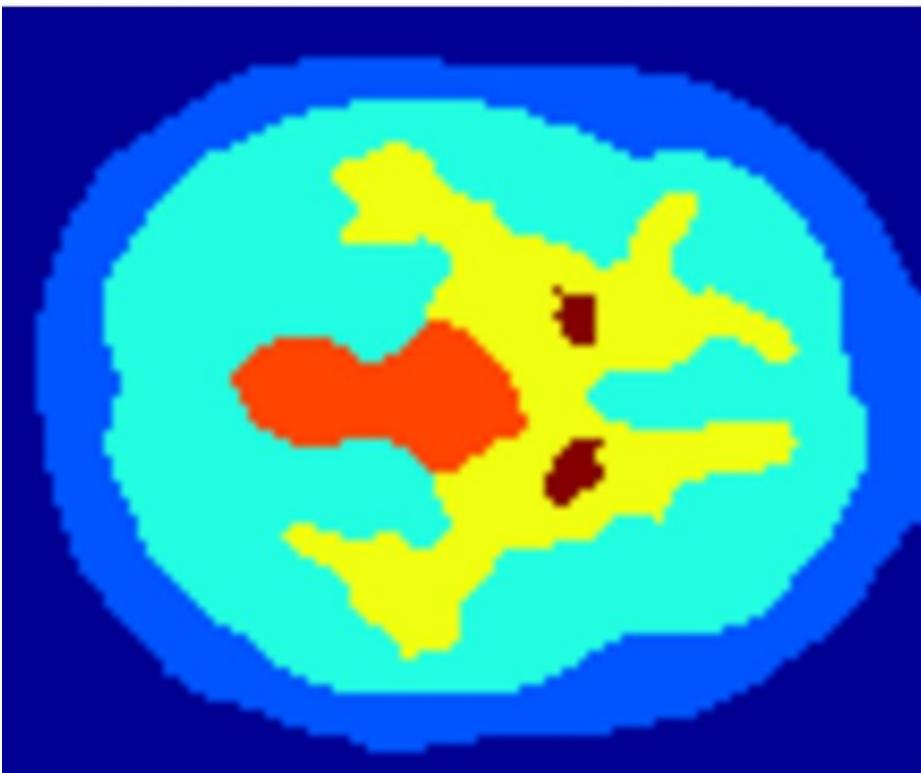
- Learning rate
- Number and size of network layers
- Slack variables
- Stopping conditions

Why study meta parameters?

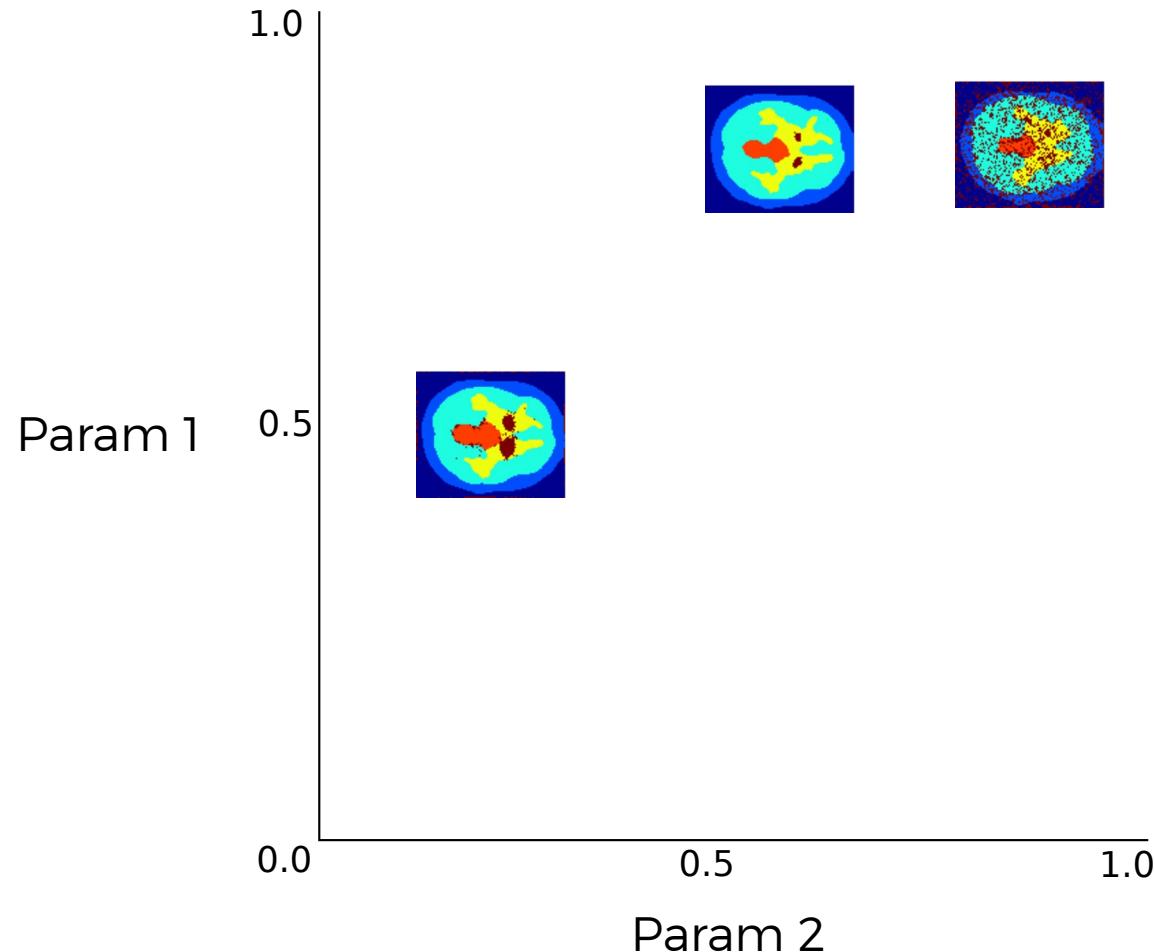


Why study meta parameters?

Meta parameters can have a large influence on performance

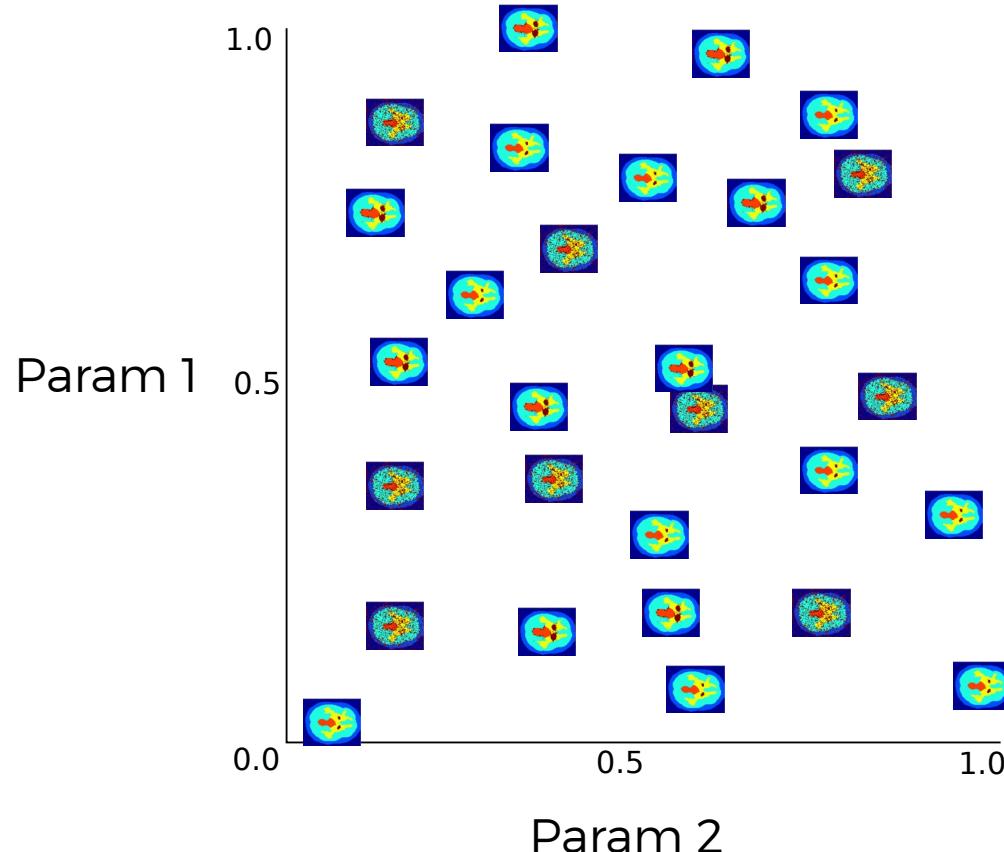


Manual method

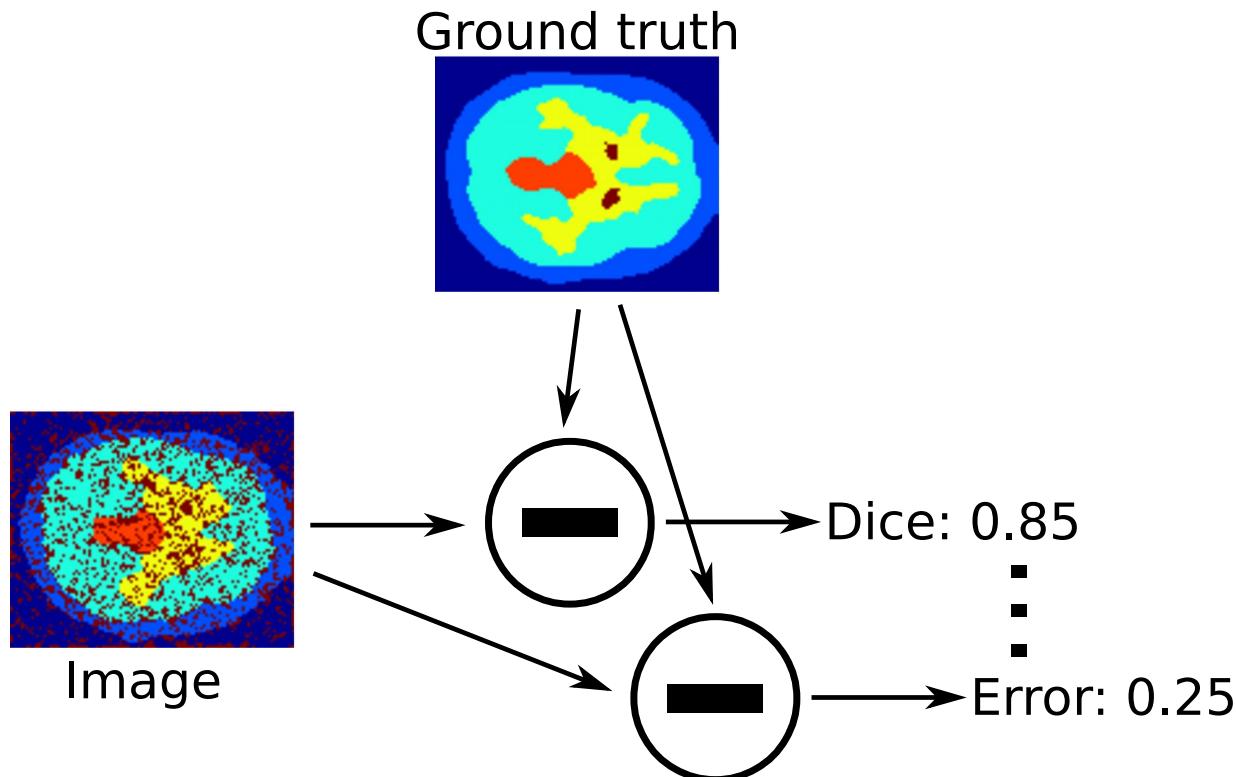


Manual method

Manual tuning is time consuming and error prone

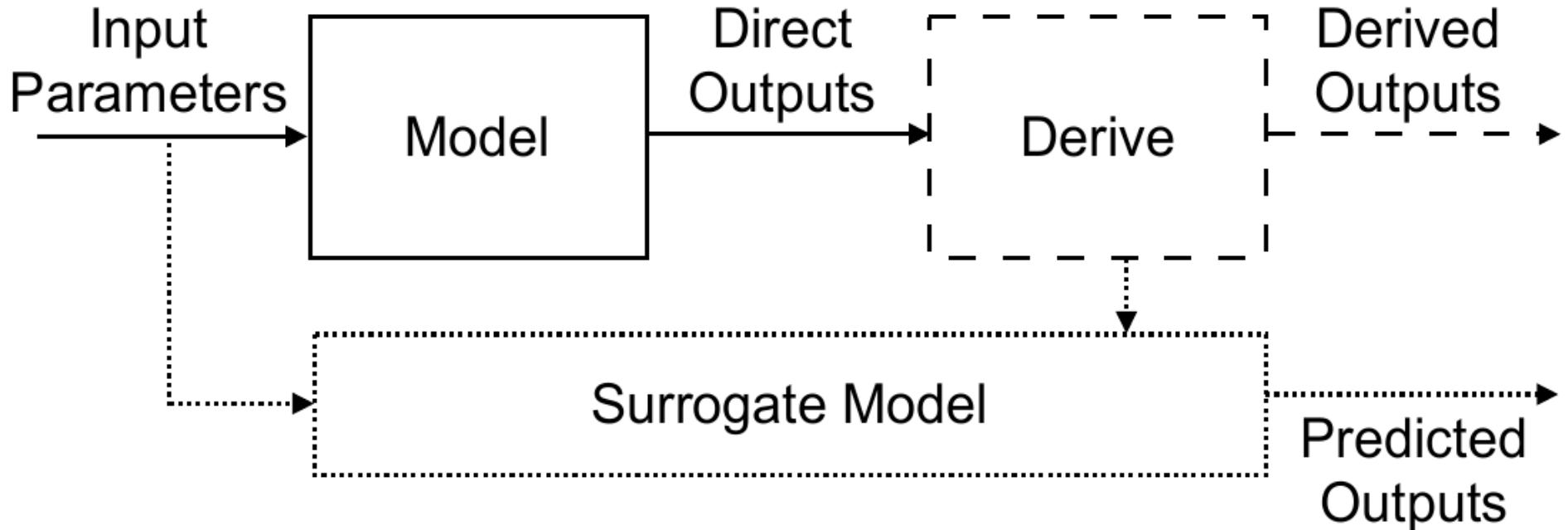


Objective measures



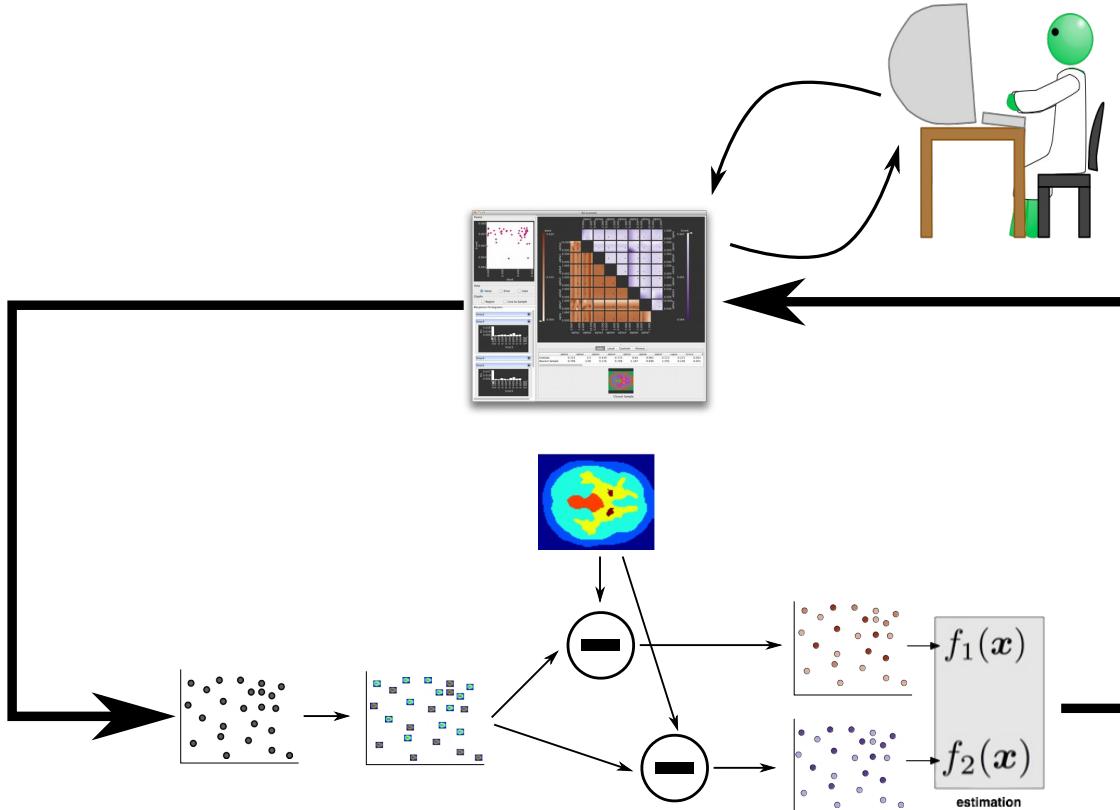
Visual parameter space exploration

Use a more principled approach

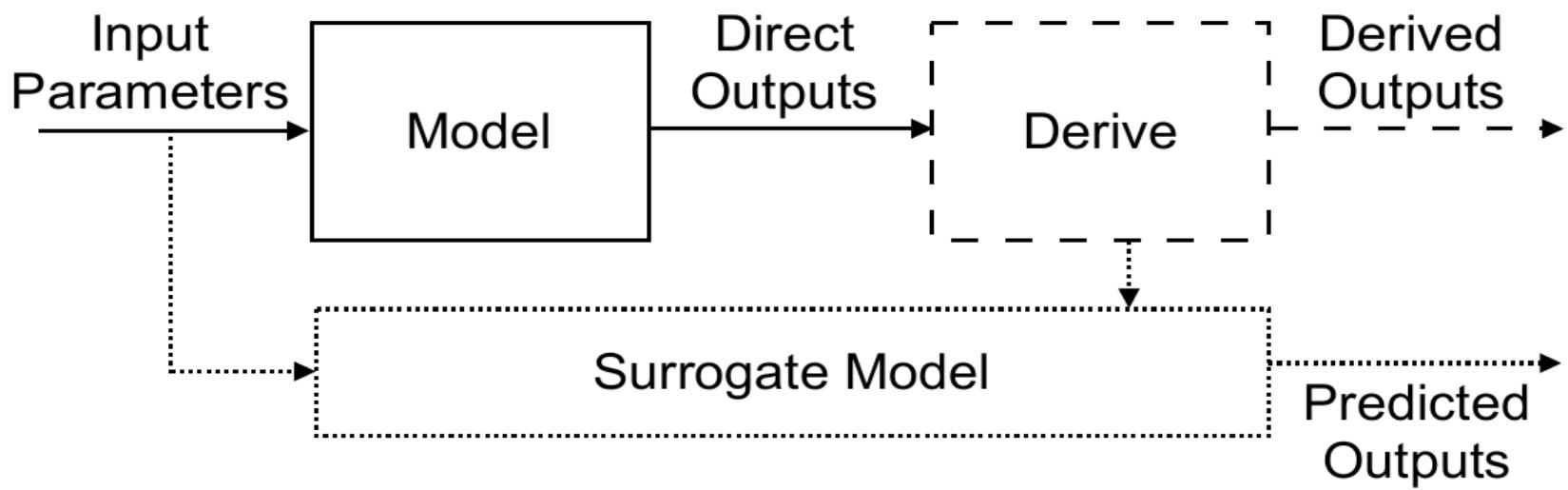
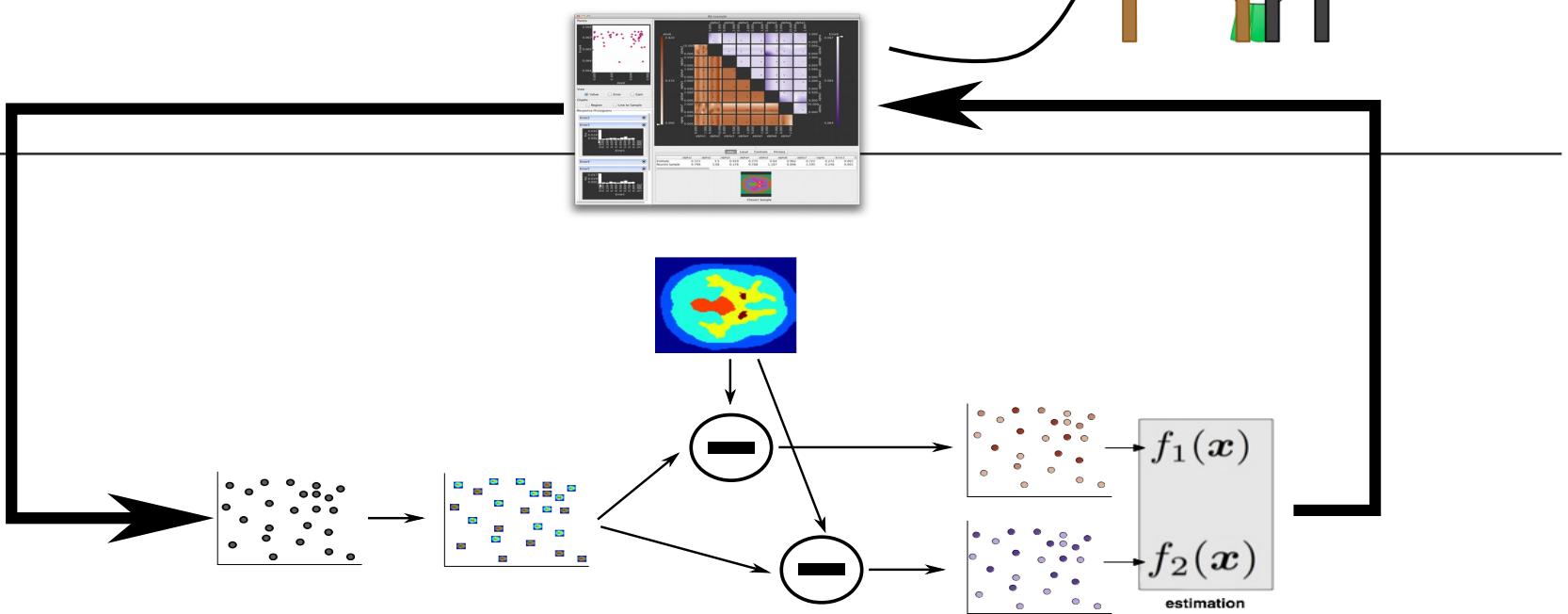


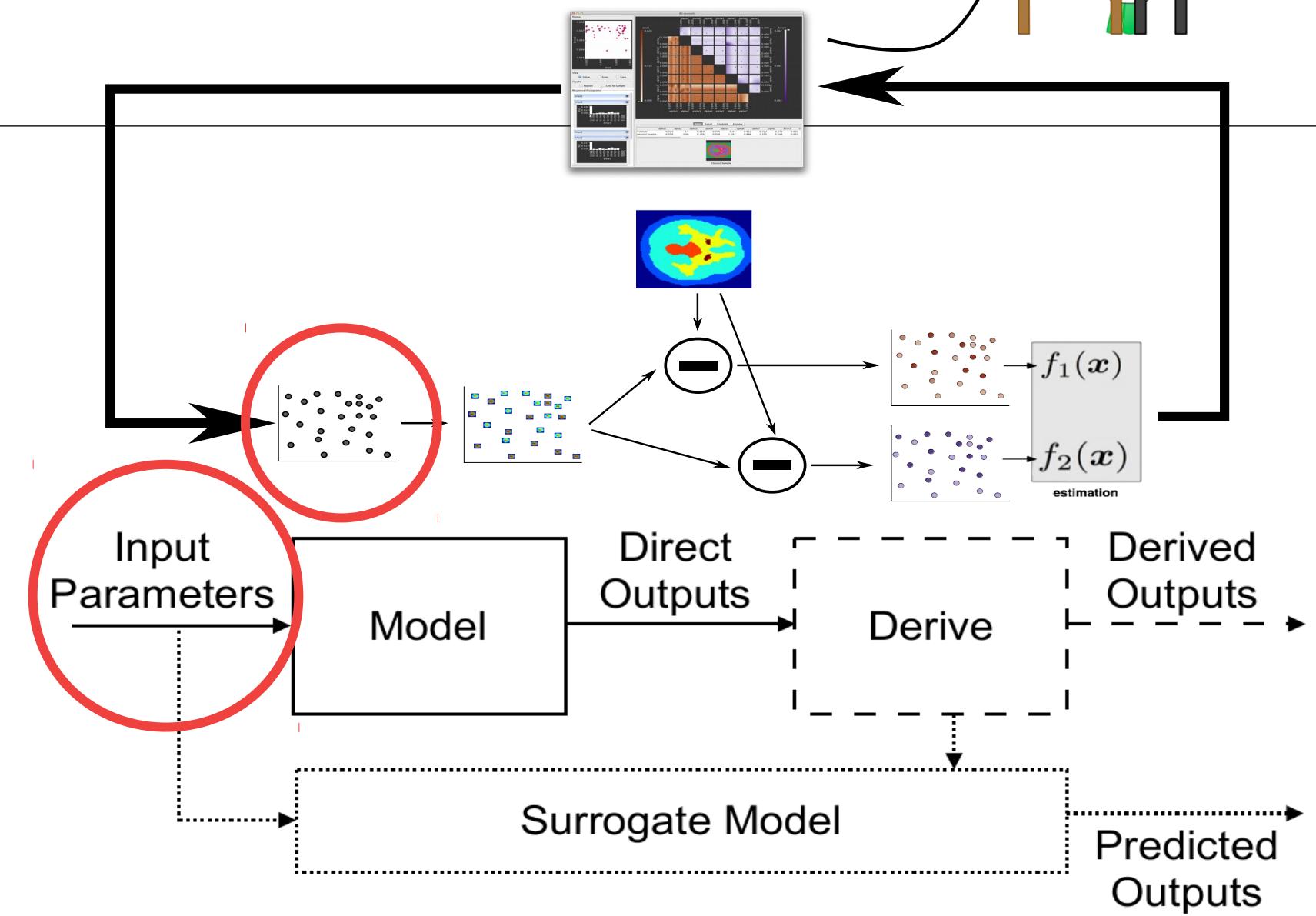
Sedlmair, Michael, Christoph Heinzl, Stefan Bruckner, Harald Piringer, and Torsten Möller. "Visual parameter space analysis: A conceptual framework," IEEE Transactions on Visualization and Computer Graphics. 2014.

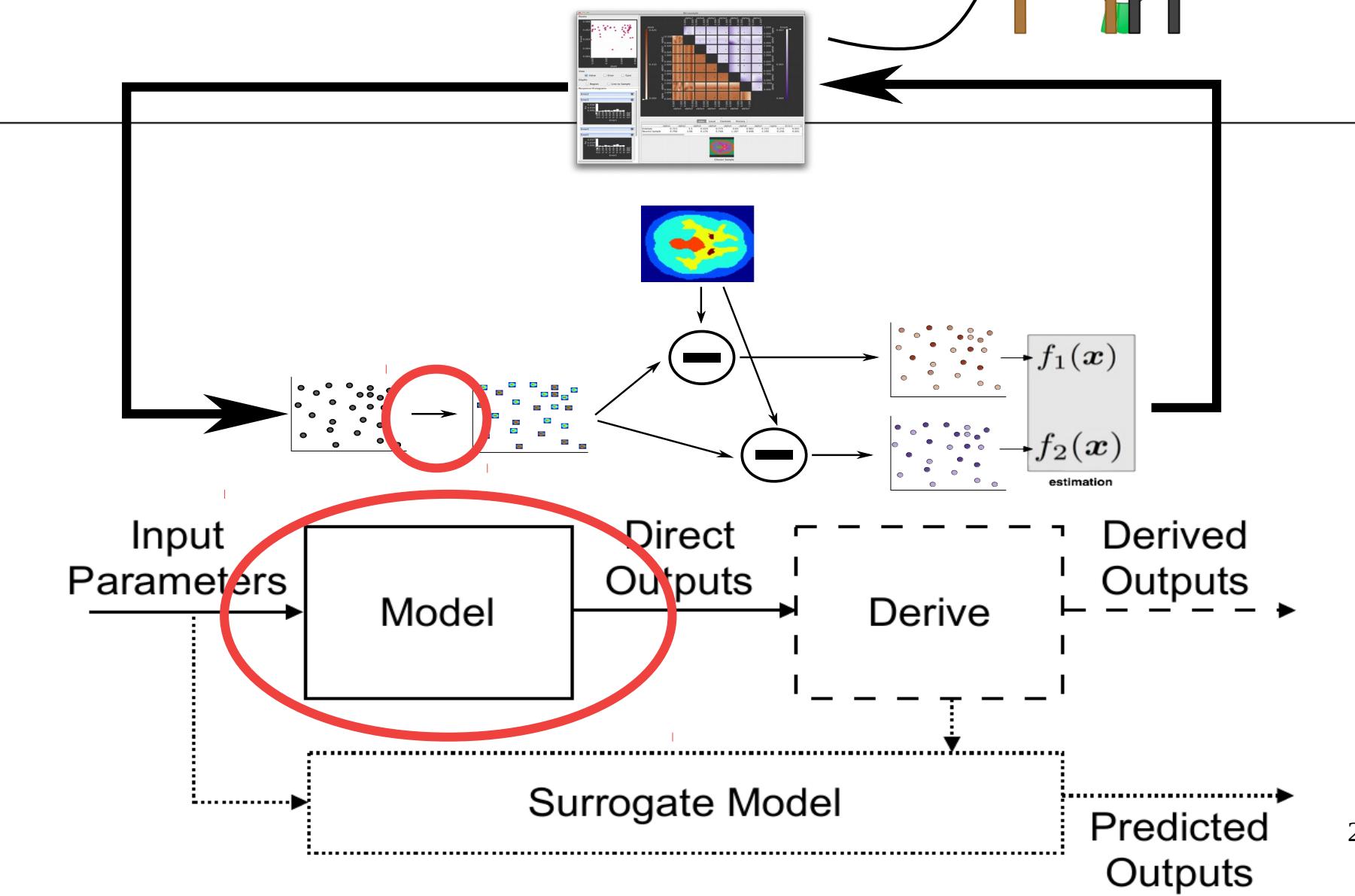
Tuner

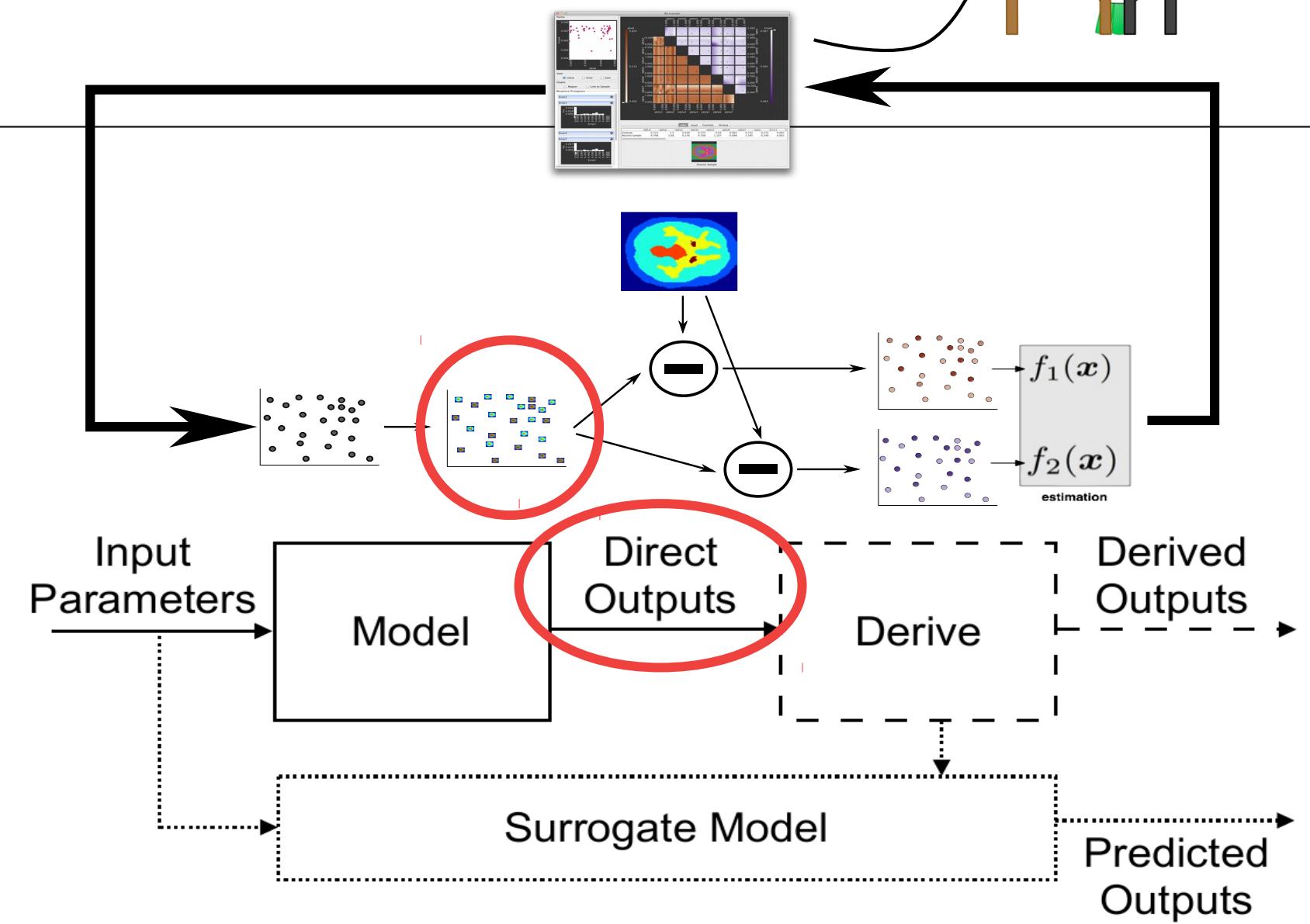


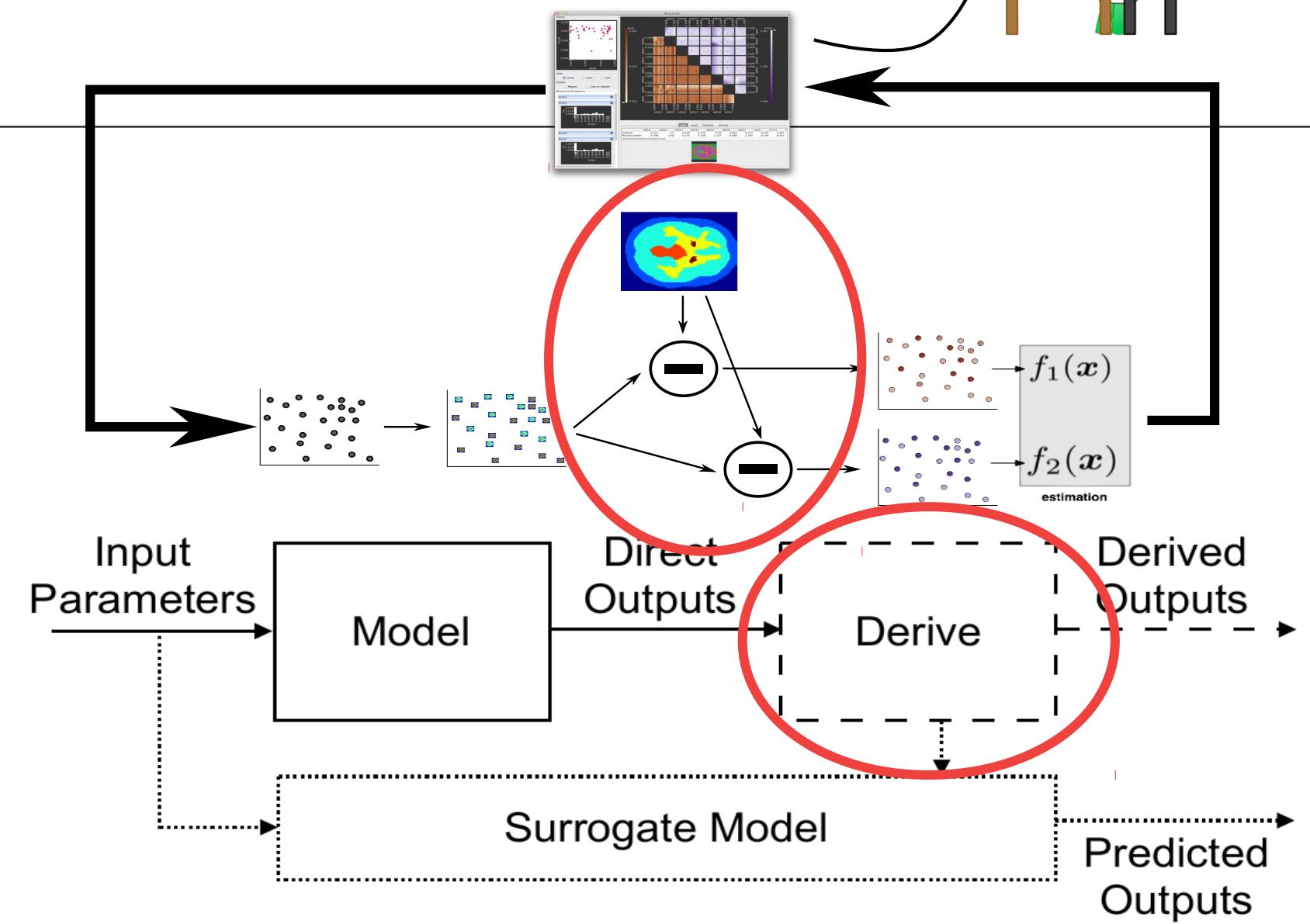
Torsney-Weir, Thomas, Ahmed Saad, Torsten Möller, Britta Weber, Hans-Christian Hege, Jean-Marc Verbaatz, and Steven Bergner. "Tuner: Principled parameter finding for image segmentation algorithms using visual response surface exploration," IEEE Transactions on Visualization and Computer Graphics. 2011.

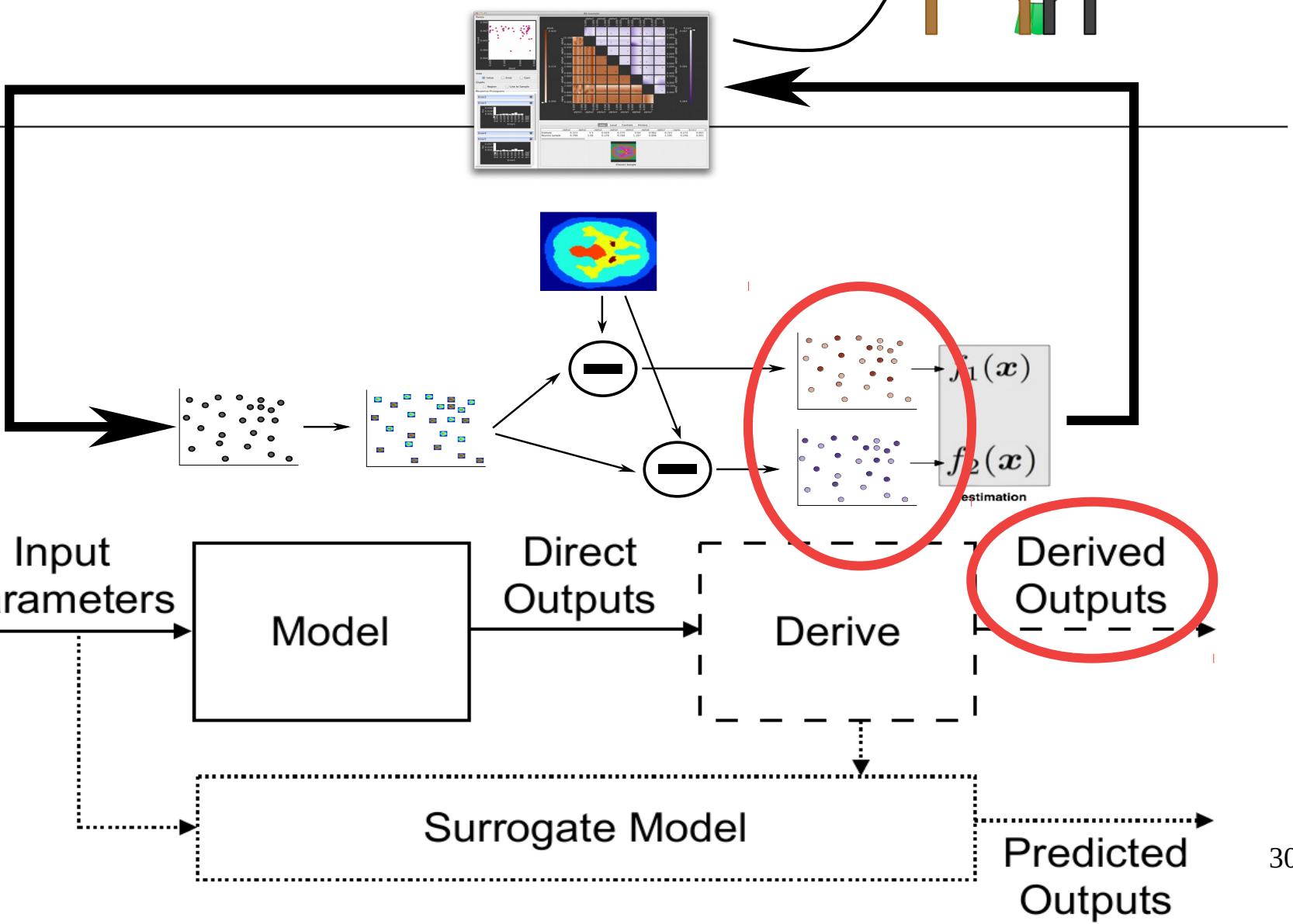


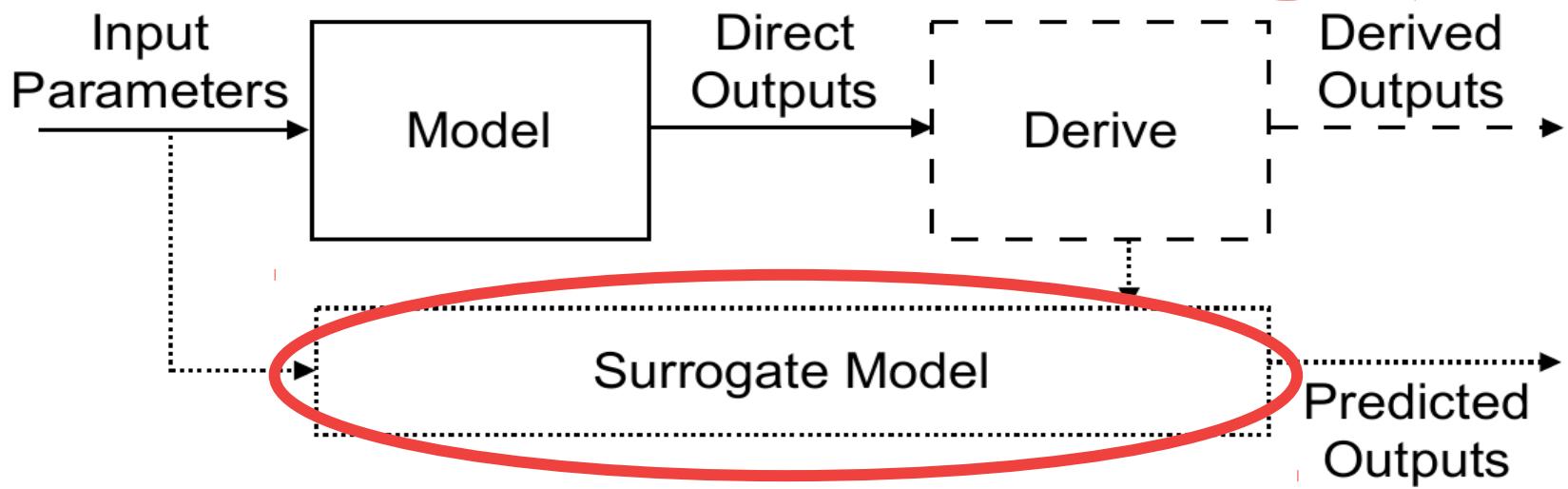
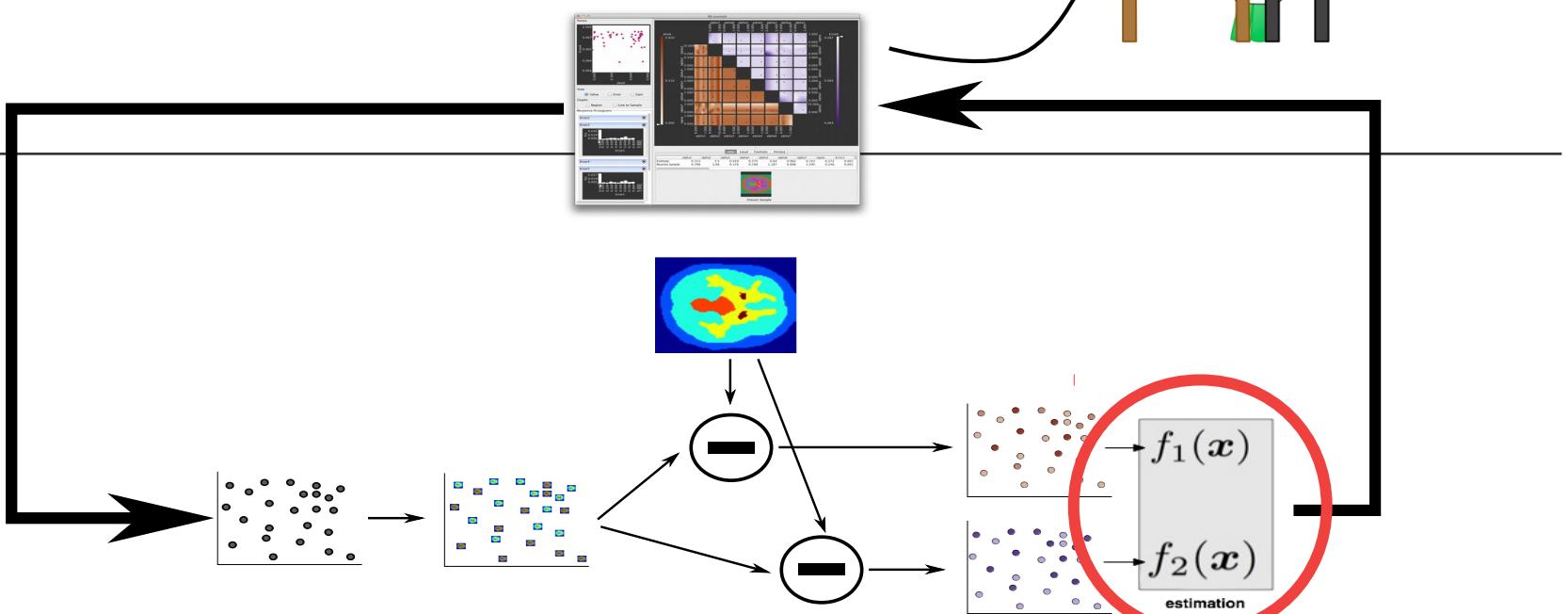




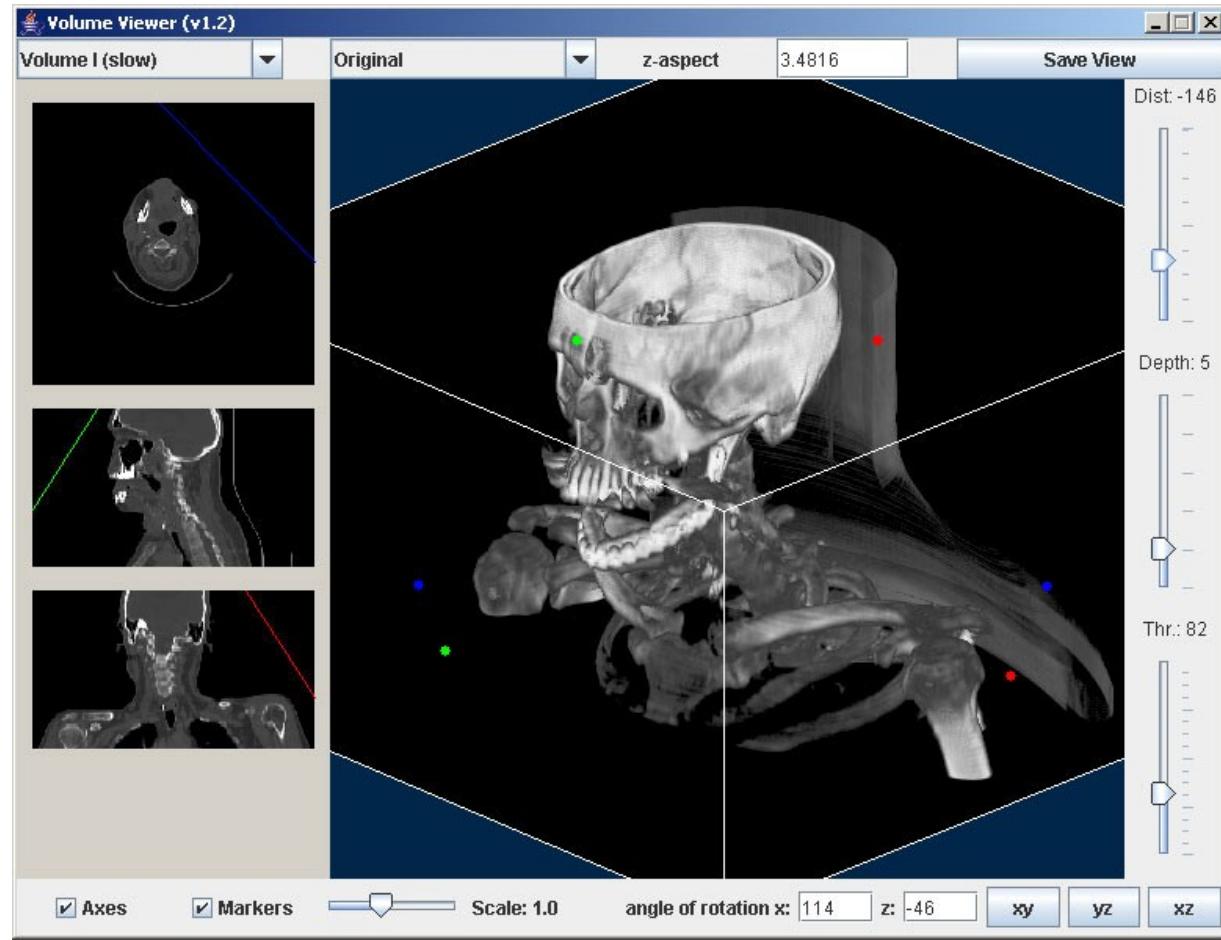




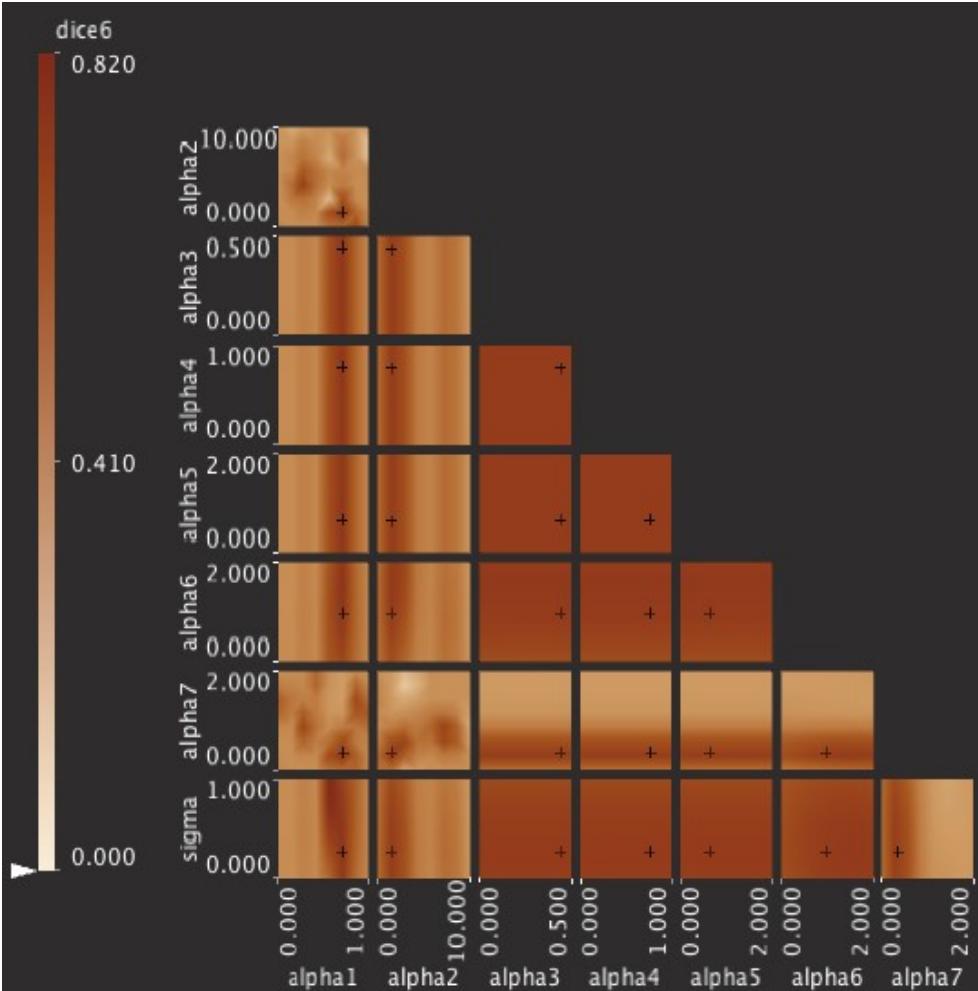




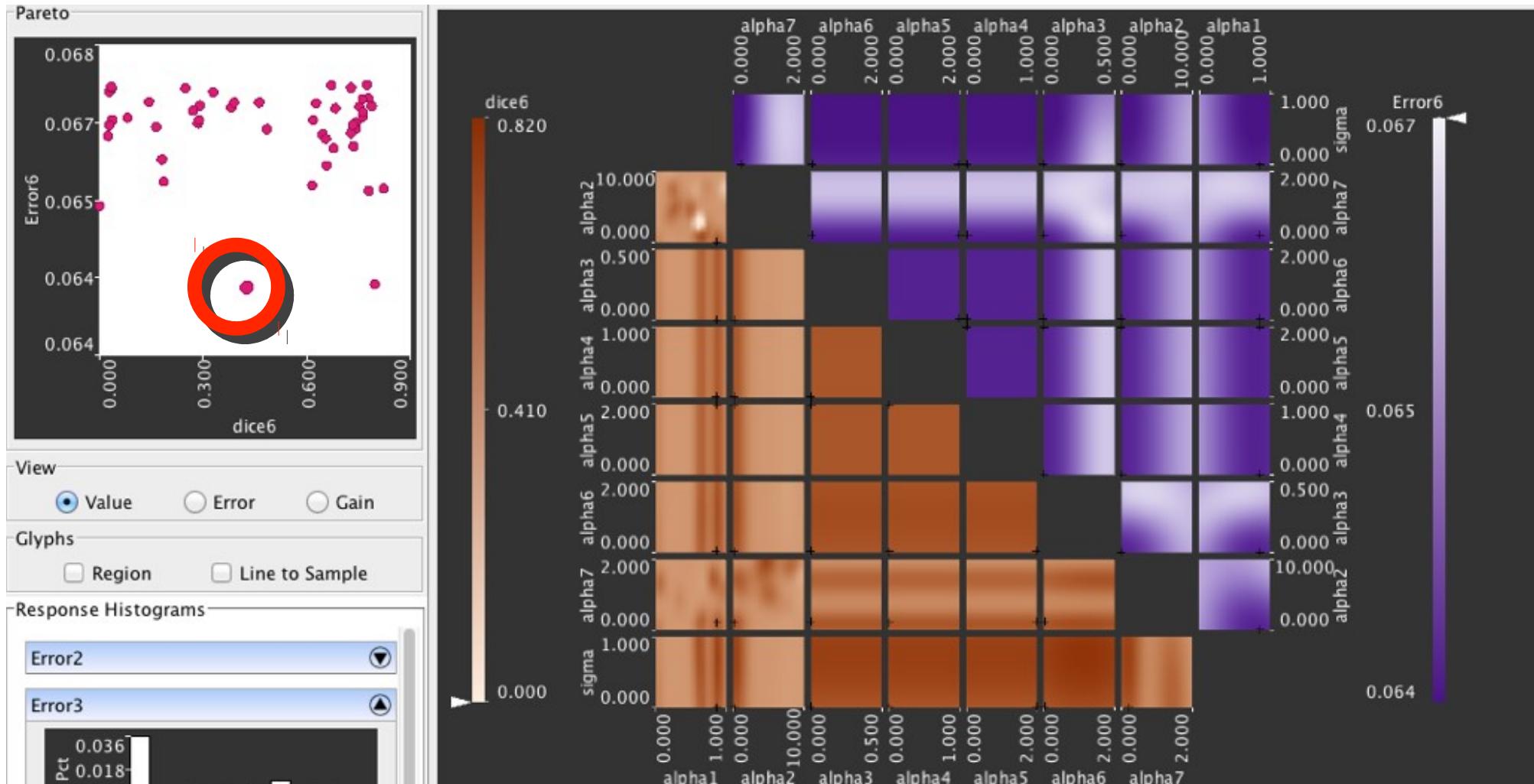
Tuner



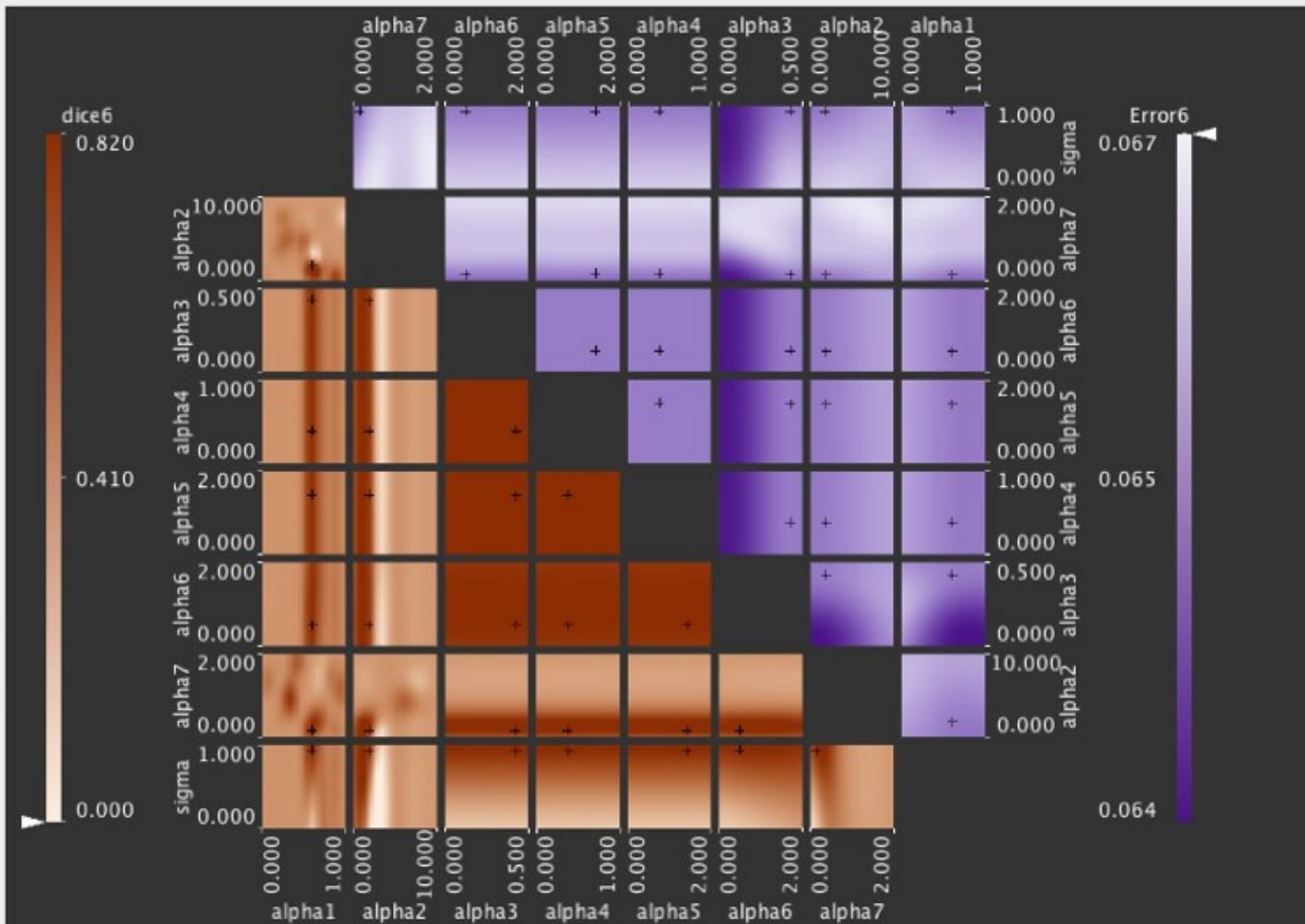
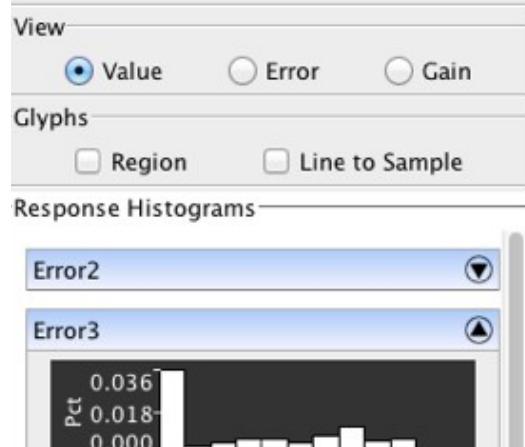
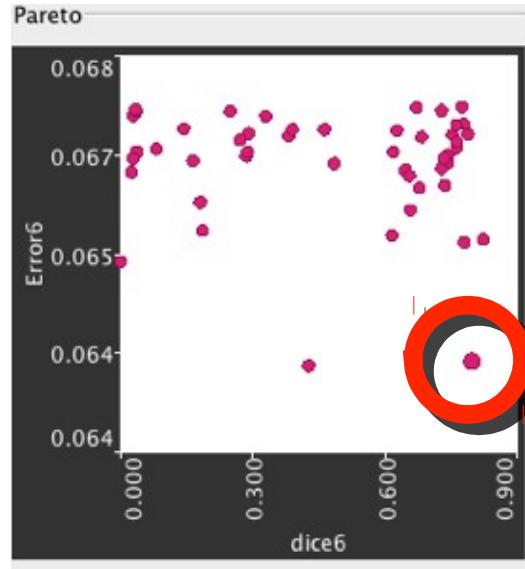
Tuner



Tuner



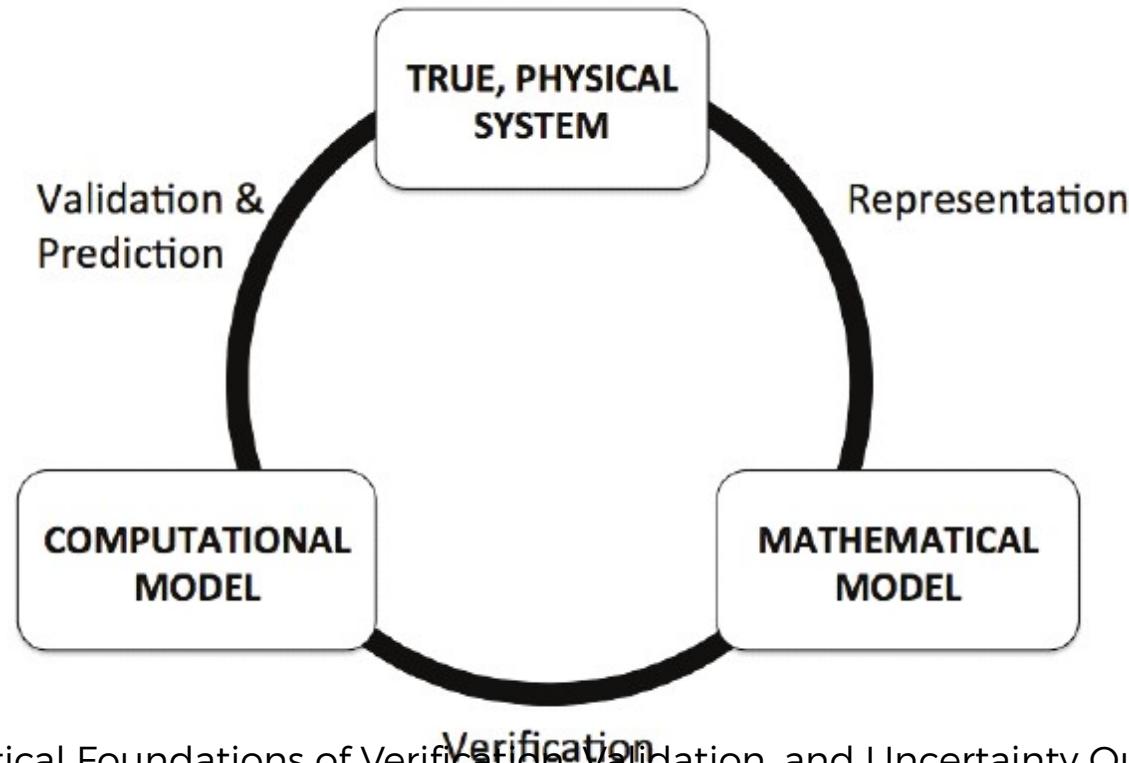
Tuner



Validating and verifying models

What do we mean?

How do we know when our models are working?



Committee on Mathematical Foundations of Verification, Validation, and Uncertainty Quantification; Board on Mathematical Sciences and Their Applications, Division on Engineering and Physical Sciences, National Research Council. *Assessing the reliability of complex models: Mathematical and statistical foundations of verification, validation, and uncertainty quantification* 2012.

Model selection

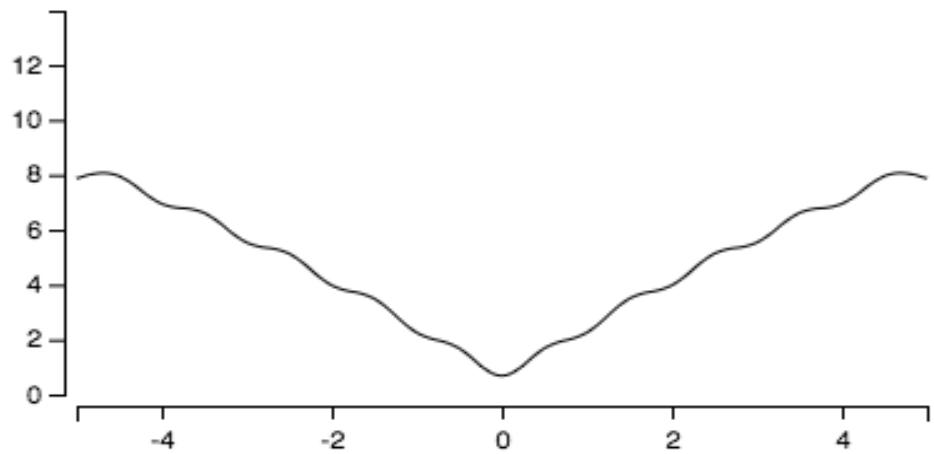
- What is the best model for my data?
- How well will these models generalize?
- Summary statistics are not always enough
- Balancing multiple objectives is difficult
- Certain training points might be very important

Example

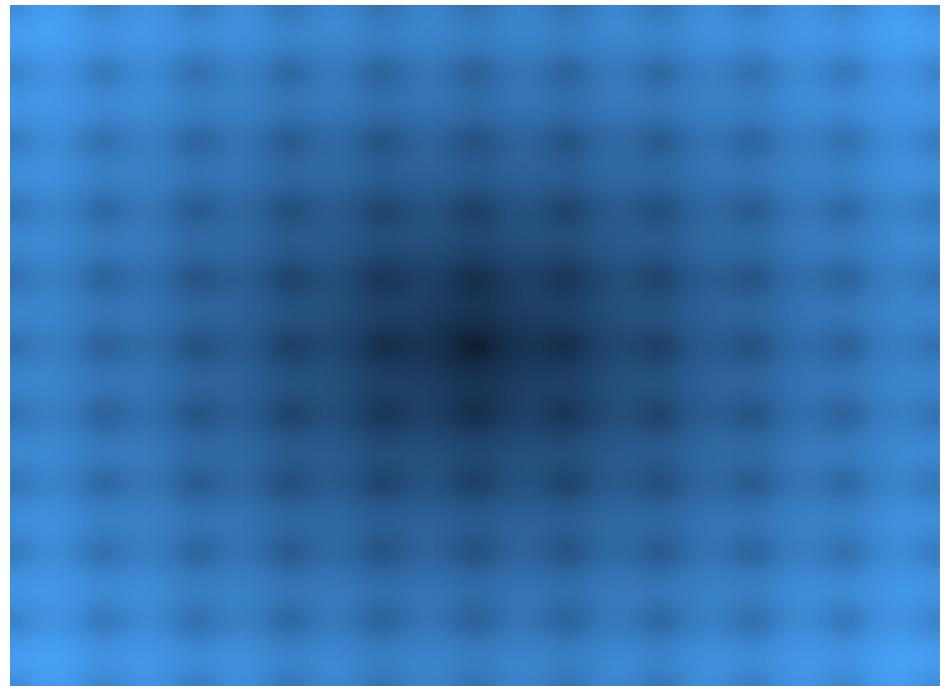
Boston housing dataset

- 13 factors
 - crime rate
 - number of rooms
 - etc
- want to predict home price
- what regression model to pick?

Function inspection



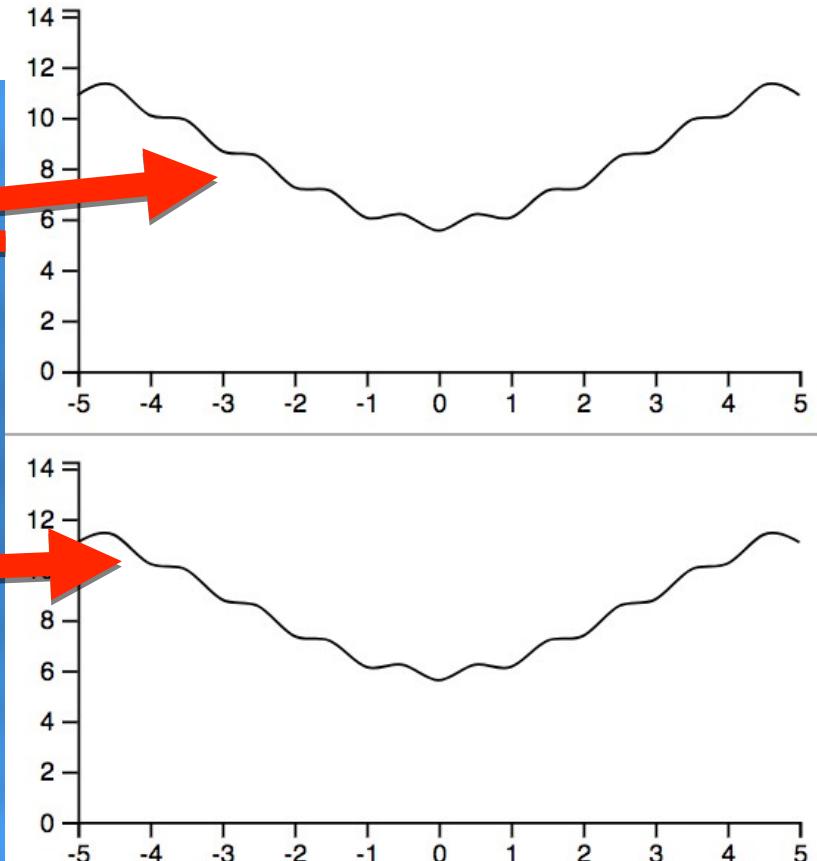
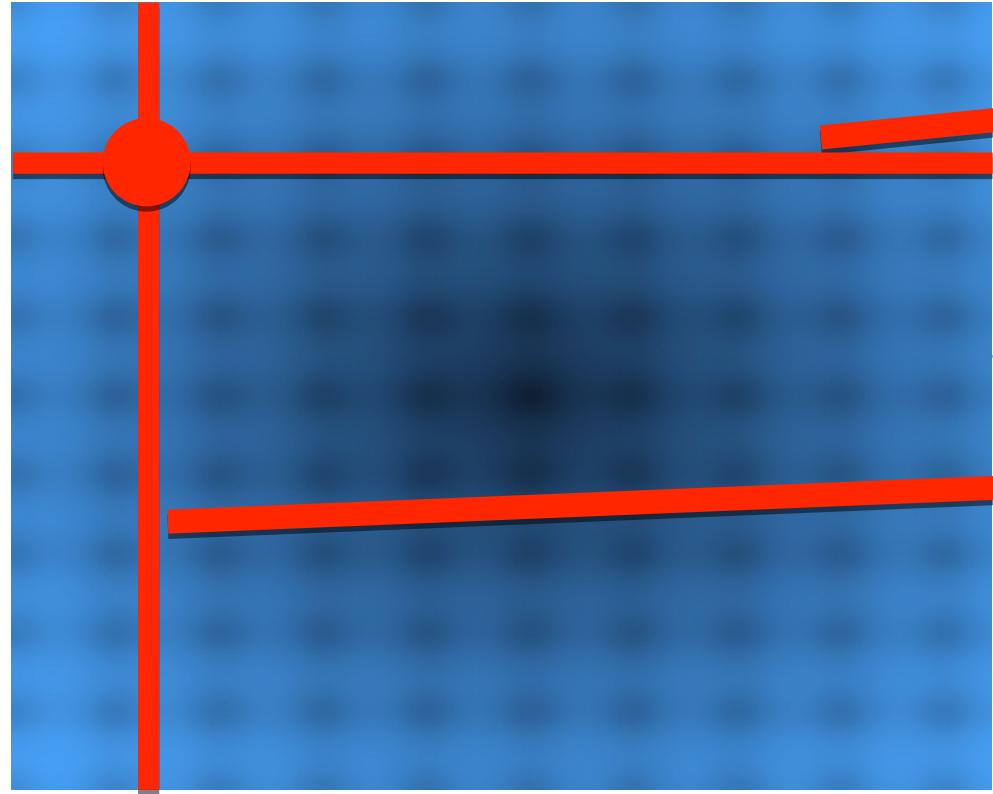
1D function



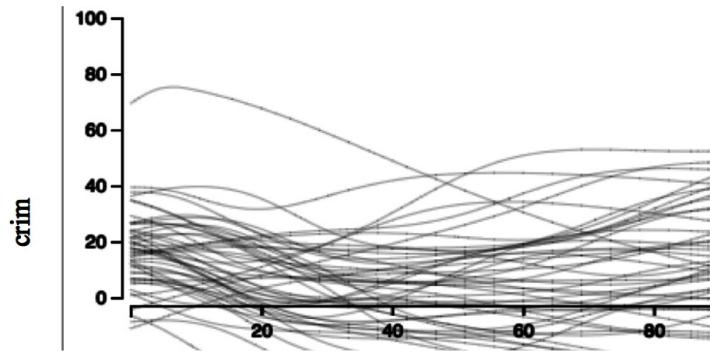
2D function

More than this gets hard!

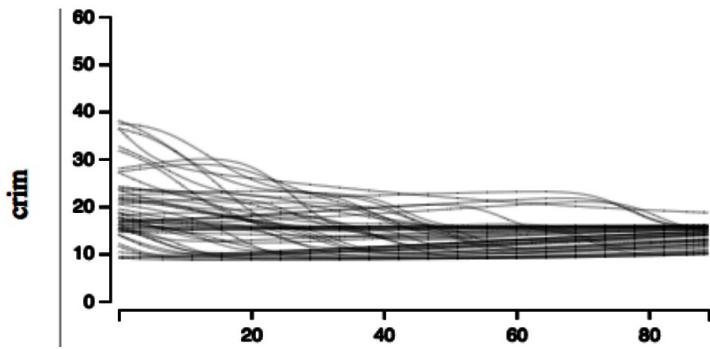
Slicing



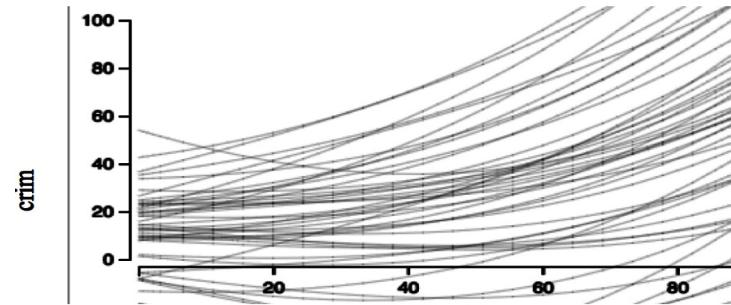
Sliceplorer views



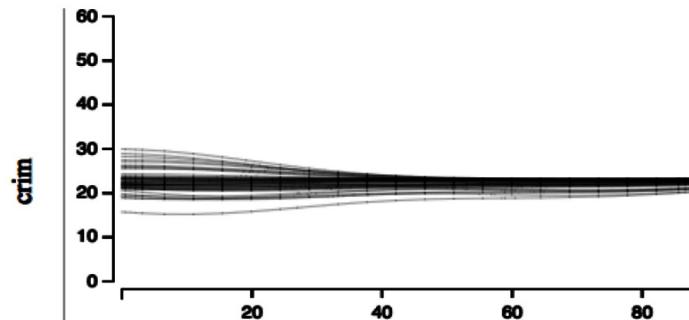
NN (single layer)



NN (dual layer)



SVM (polynomial)



SVM (RBF)

Torsney-Weir, Thomas, Michael Sedlmair, and Torsten Möller.
"Sliceplorer: 1D slices for multi-dimensional continuous functions,"
Computer Graphics Forum. 2017.

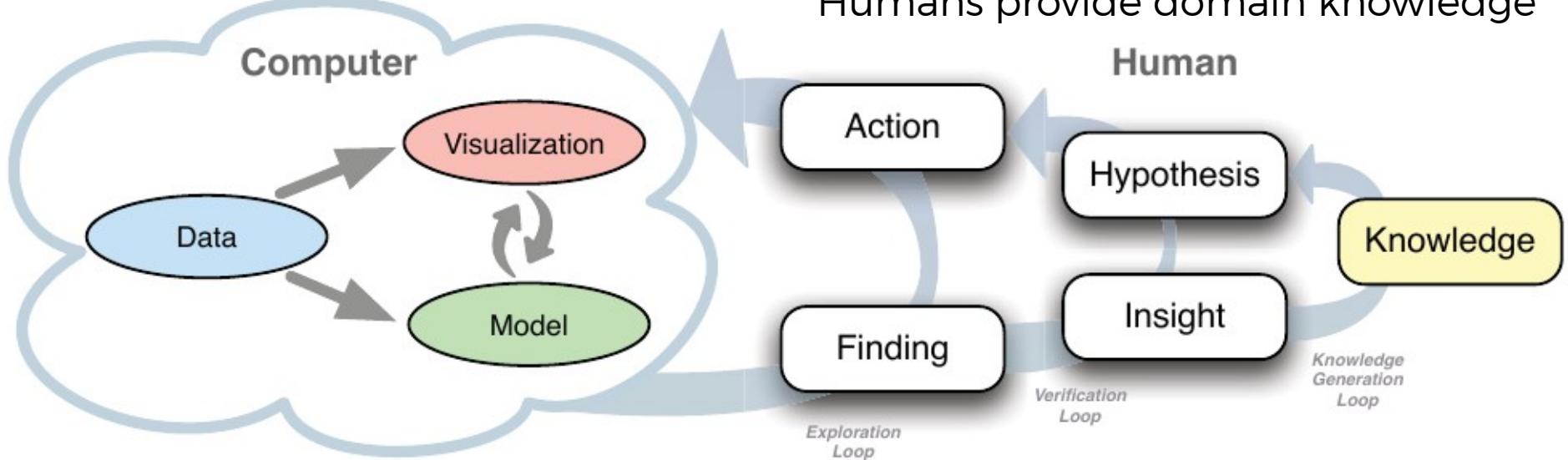
Summary

- Just an answer is not enough (context)
- Humans have trouble understanding complex models
- Interactivity can bring people into the model

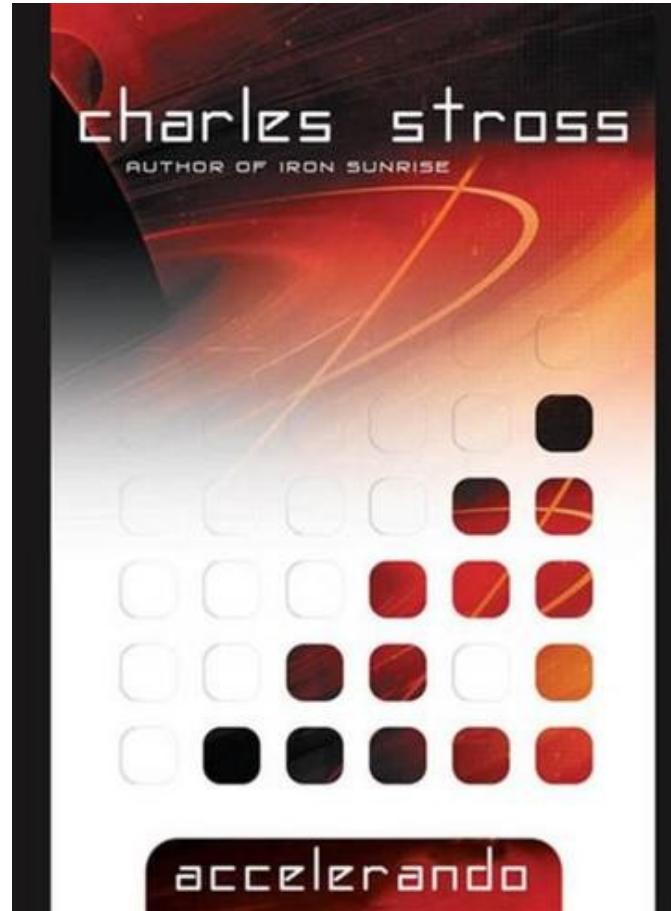
Machine learning helping visualization

How?

Computers are good at calculating



Book ad!



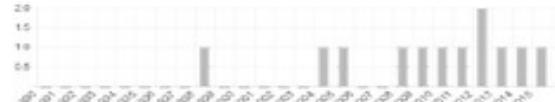
Methods

- Clustering
- Classification
- Regression

New keyword search:

framework

found in 12 papers

cooccurred with 40 author keywords
in a topic cluster with 41 author keywords**framework** found in 12 VIS Papers

Conf.	Year	Title
VAST	2015	Characterizing Provenance in Visualization and Data Analysis: An Organizational Framework of Provenance Types and Purposes
InfoVis	2014	Design Activity Framework for Visualization Design
VAST	2013	An Extensible Framework for Provenance in Human Terrain Visual Analytics
InfoVis	2012	Design Study Methodology: Reflections from the Trenches and the Stacks
VAST	2012	An AffordanceBased Framework for Human Computation and HumanComputer Collaboration
VAST	2011	Visual analytic roadblocks for novice investigators
InfoVis	2010	behaviorism: a framework for dynamic data visualization
InfoVis	2009	A Nested Model for Visualization Design and Validation
InfoVis	2008	A Framework of Interaction Costs in Information Visualization
VIS	2005	Framework for visualizing higherorder basis functions
InfoVis	2004	A Knowledge TaskBased Framework for Design and Evaluation of Information Visualizations
InfoVis	1998	An operator interaction framework for visualization systems

framework cooccurs with 40 other keywords:

information visualization 4x, design 3x, evaluation 3x, visualization 3x, model 2x, provenance 2x, theory 2x, visual analytics 2x, analytic gap 1x, analytic provenance 1x, basis function 1x, bookmarks 1x, cognitive model 1x, conceptual model 1x, cybersecurity 1x, decisions 1x, design study 1x, dynamic data 1x, extensibility 1x, finite elements 1x, human complexity 1x, human computation 1x, human terrain analysis 1x, information art 1x, interaction 1x, interface evaluation 1x, investigative analysis 1x, knowledge tasks 1x, methodology 1x, narrative 1x, nested model 1x, operators 1x, process 1x, qualitative experiment 1x, roadblock 1x, spreadsheets 1x, tessellation 1x, user interaction 1x, view/value 1x, visualization systems 1x

framework is part of the topic cluster Visualization Systems, Toolkits, and Environments containing 41 keywords

visualization systems 15x, framework 12x, toolkits 6x, visualization system and toolkit design 3x, radviz 2x, social software 2x, system 2x, visualization environment 2x, visualization framework 2x, analytic environments 1x, authoring environment 1x, bard 1x, decision support systems 1x, ecco 1x, enzo 1x, fundexplore 1x, geovista studio 1x, information visualization architecture 1x, information workspace 1x, integrated visualization system 1x, many eyes 1x, mobjects 1x, modular visualization environments 1x, objectorientated visualization toolkit 1x, pad++ 1x, physicallybased systems 1x, piccolo 1x, renderman 1x, responsive workbench 1x, similan 1x, socialnetsense 1x, software infrastructure 1x, software tools 1x, system architecture 1x, toolkit design 1x, treenetviz 1x, user interface toolkits 1x, view space exploration framework 1x, visad 1x, visual analytics infrastructures 1x, volume visualization framework 1x

Isenberg, Petra, Tobias Isenberg, Michael Sedlmair, Jian Chen, and Torsten Möller. "Visualization as seen through its research paper keywords," IEEE Transactions on Visualization and Computer Graphics. 2017.

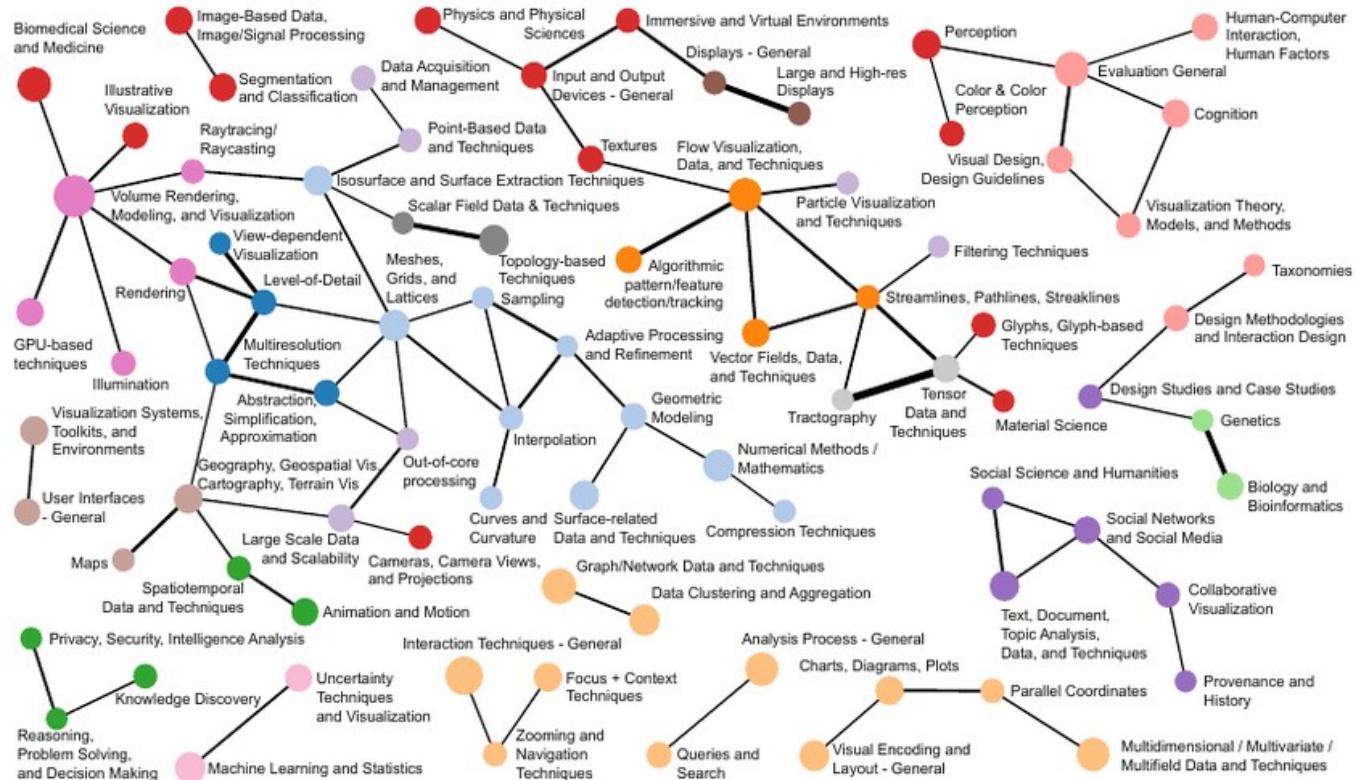
KeyVis

goal: find relevant papers for a research project

- What are "key" papers in subject?
- How helpful are keywords?
- Do keywords relate to each other?

KeyVis

step 1: cluster the papers based on keywords



KeyVis

step 2: build an interface for the clustering

Search for VIS paper keywords

New keyword search:

Getting started [Search](#) Topics About

framework
found in 12 papers
cooccurred with 40 author keywords
in a topic cluster with 41 author keywords

The chart displays the frequency of the keyword 'framework' in 12 VIS papers. The y-axis represents frequency from 0.0 to 2.0. The x-axis lists 12 papers, each with a bar indicating its frequency. The frequencies range from approximately 0.2 to 1.8.

framework found in 12 VIS Papers

Conf.	Year	Title
VAST	2015	Characterizing Provenance in Visualization and Data Analysis: An Organizational Framework of Provenance Types and Purposes
InfoVis	2014	Design Activity Framework for Visualization Design
VAST	2013	An Extensible Framework for Provenance in Human Terrain Visual Analytics
InfoVis	2012	Design Study Methodology: Reflections from the Trenches and the Stacks
VAST	2012	An AffordanceBased Framework for Human Computation and HumanComputer Collaboration
VAST	2011	Visual analytic roadblocks for novice investigators
InfoVis	2010	behaviorism: a framework for dynamic data visualization
InfoVis	2009	A Nested Model for Visualization Design and Validation
InfoVis	2008	A Framework of Interaction Costs in Information Visualization
Vis	2005	Framework for visualizing higherorder basis functions
InfoVis	2004	A Knowledge TaskBased Framework for Design and Evaluation of Information Visualizations
InfoVis	1998	An operator interaction framework for visualization systems

framework cooccurs with 40 other keywords:

information visualization 4x design 3x evaluation 3x visualization 3x model 2x provenance 2x theory 2x visual analytics 2x analytic gap 1x analytic provenance 1x basis function 1x bookmarks 1x cognitive model 1x conceptual model 1x cybersecurity 1x decisions 1x design study 1x dynamic data 1x extensibility 1x finite elements 1x human complexity 1x human computation 1x human terrain analysis 1x information art 1x interaction 1x interface evaluation 1x investigative analysis 1x knowledge tasks 1x methodology 1x narrative 1x nested model 1x operators 1x process 1x qualitative experiment 1x roadblock 1x spreadsheets 1x tessellation 1x user interaction 1x viewvalue 1x visualization systems 1x

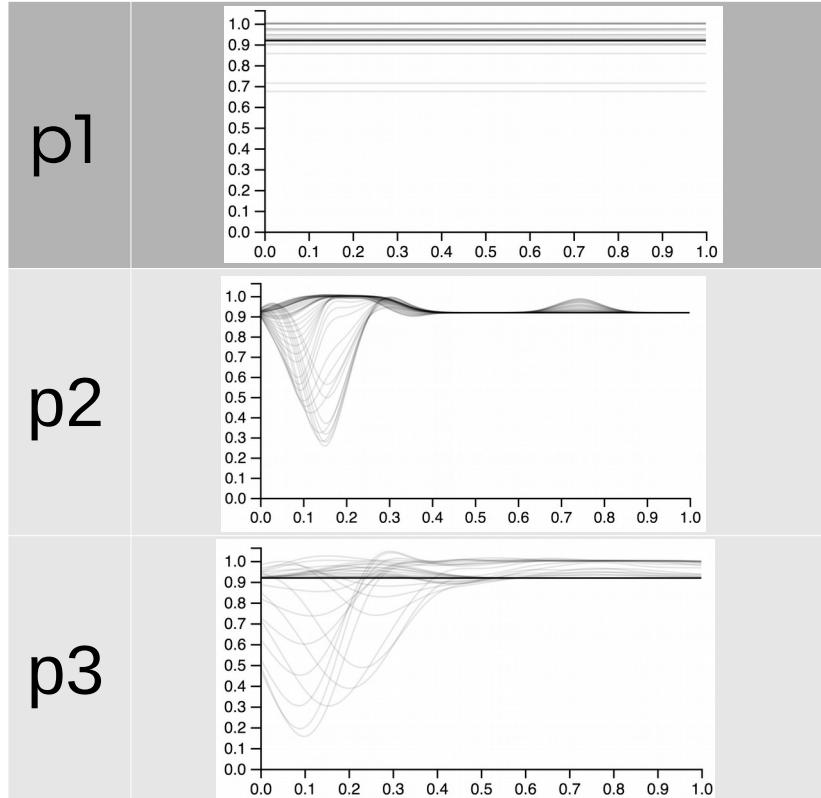
framework is part of the topic cluster Visualization Systems, Toolkits, and Environments containing 41 keywords

visualization systems 15x framework 12x toolkits 6x visualization system and toolkit design 3x radviz 2x social software 2x system 2x visualization environment 2x visualization framework 2x analytic environments 1x authoring environment 1x bard 1x decision support systems 1x echo 1x enzo 1x fundexplore 1x geovista studio 1x information visualization architecture 1x information workspace 1x integrated visualization system 1x many eyes 1x mobjects 1x modular visualization environments 1x objectoriented visualization toolkit 1x pad++ 1x physicallybased systems 1x piccolo 1x renderman 1x responsive workbench 1x similar 1x socialnseisense 1x software infrastructure 1x software tools 1x system architecture 1x toolkit design 1x treemetyliz 1x user interface toolkits 1x view space exploration framework 1x visad 1x visual analytics infrastructures 1x volume visualization framework 1x

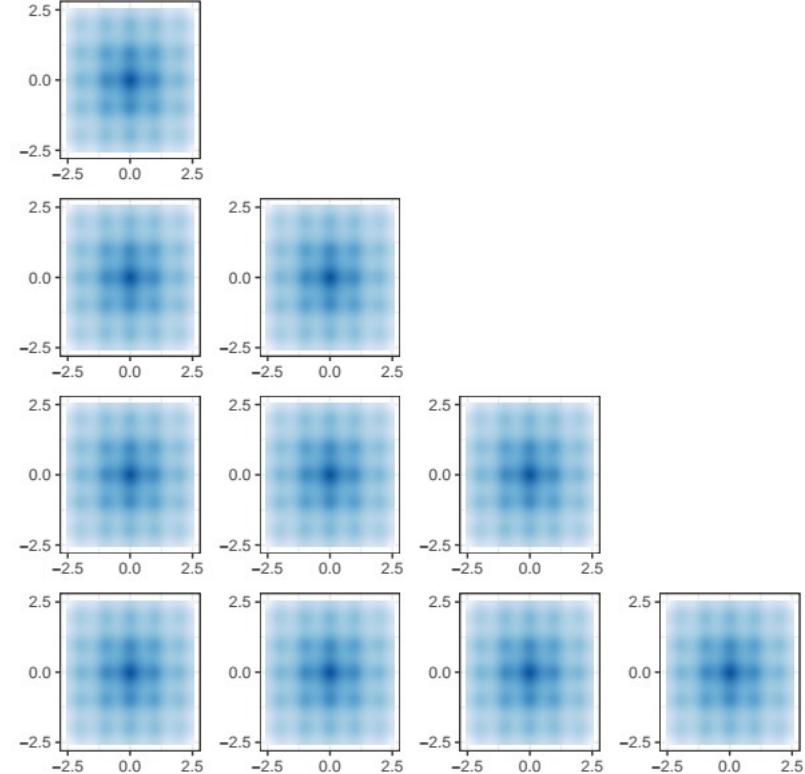
Regression

Slicing

1D slices



2D slices



The future!

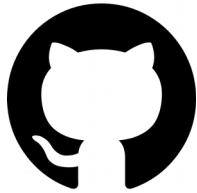
Interesting projects

- How can humans and machines work together?
- Visualizing regression algorithms
- Understand what "understandability" means

Thanks!

www.tomtorsneyweir.com

thomas.torsney-weir@univie.ac.at



@gabysbrain

