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

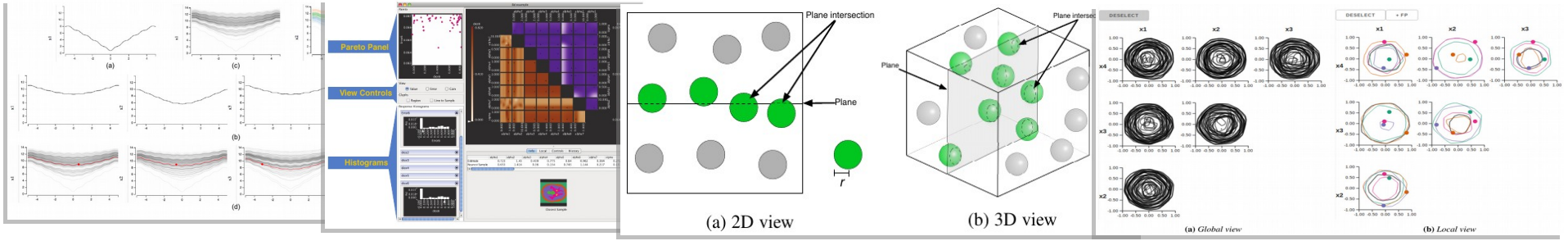
About me

- 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

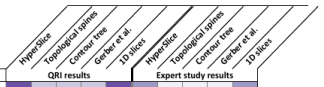


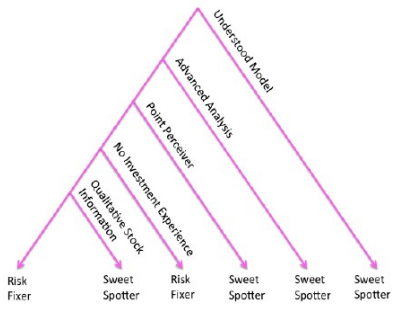
Multi-dimensional spaces

API design

Users/tasks

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 those data cases"	Summary statistics: variance, mean, SA
Find extremum	"Find data cases possessing an extreme value of an attribute over its range within the data set"	Find local/global min/max
Determine range	"Given a set of data cases and an attribute of interest, find the span of values within the set"	What is the range of possible outputs?
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"	What types of shapes do the manifolds have
Find anomalies	"Identify any anomalies within a given set of data cases with respect to a given relationship or expectation, e.g. statistical outliers"	Do areas of the manifold have shapes unlike any others
Cluster	"Given a set of data cases, find clusters of similar attribute values"	Areas of the manifold have similar shapes
Correlate	"Given a set of data cases and two attributes, determine useful relationships between the values of those attributes"	1D vs 2D relationships

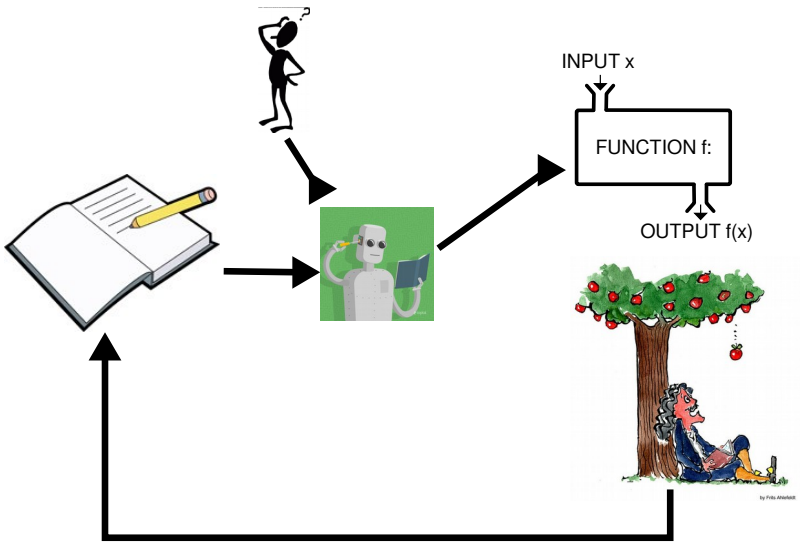




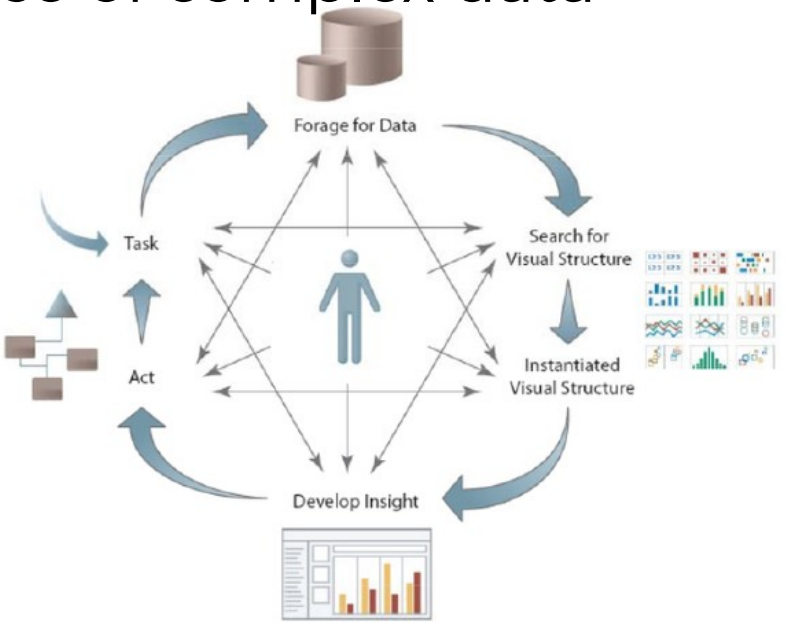
Visualization and machine learning

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

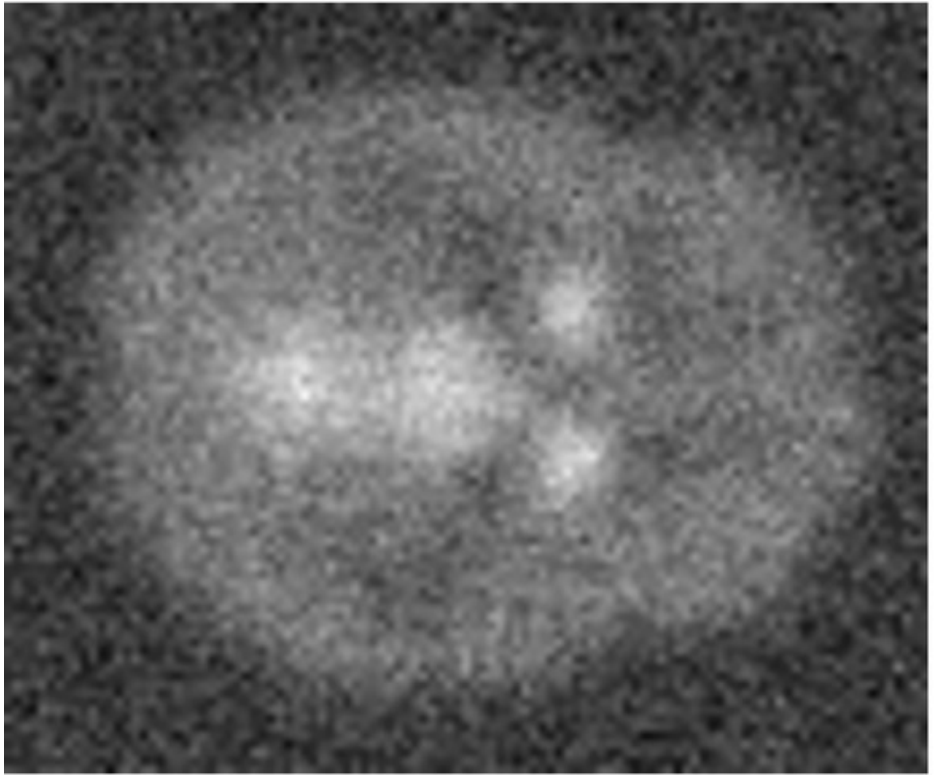
Meta parameters

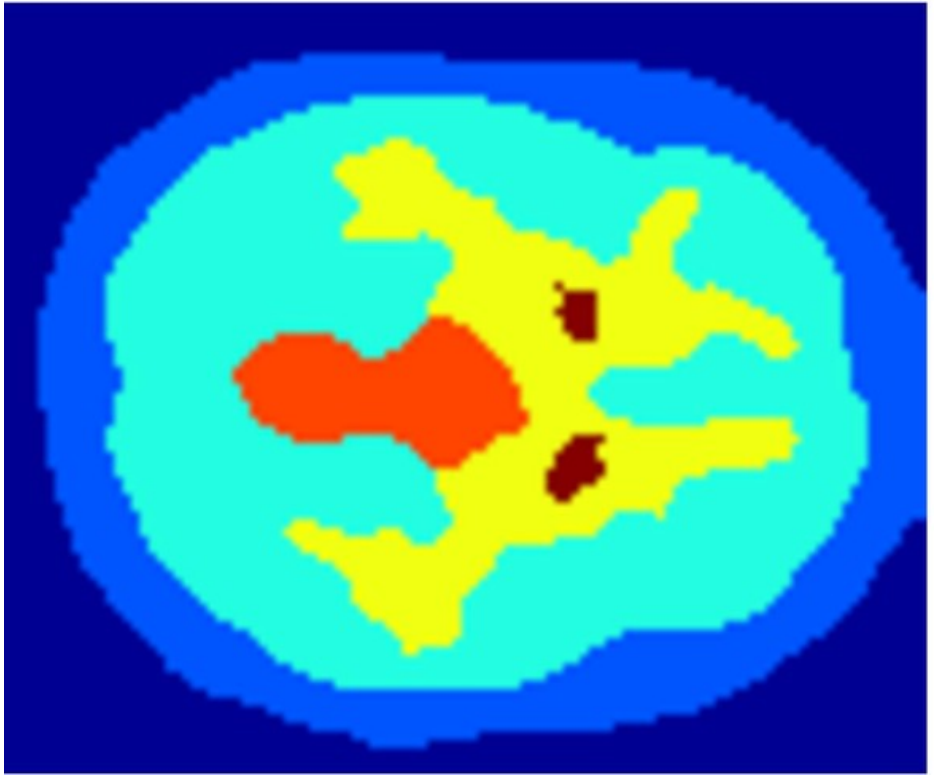
What are meta parameters?

Control how learning takes place

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

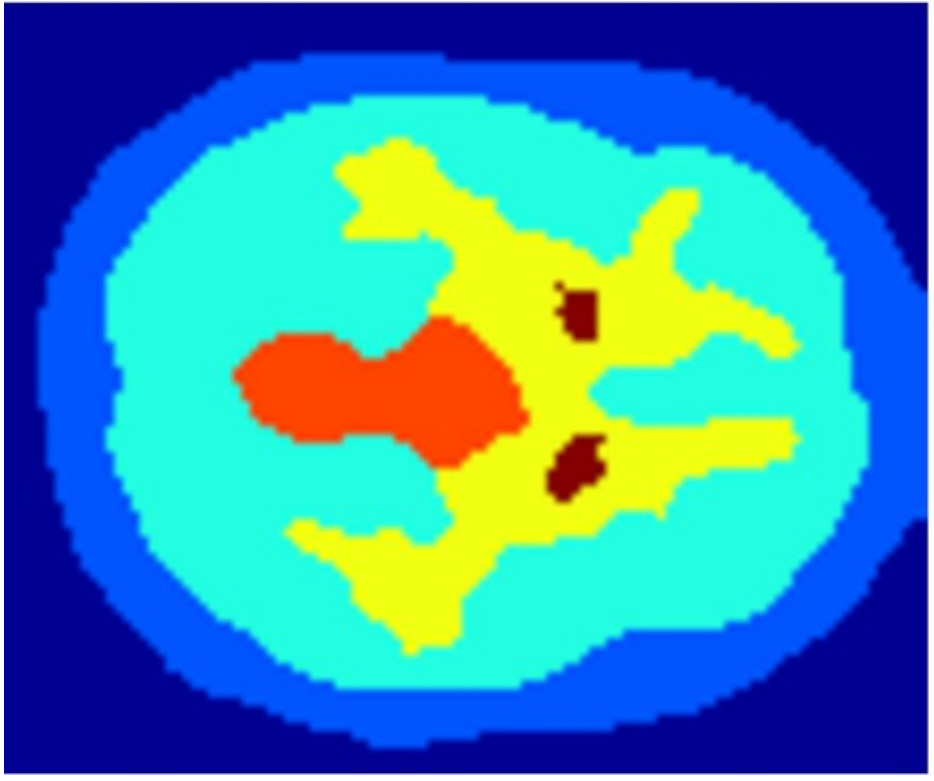
Why study meta parameters?

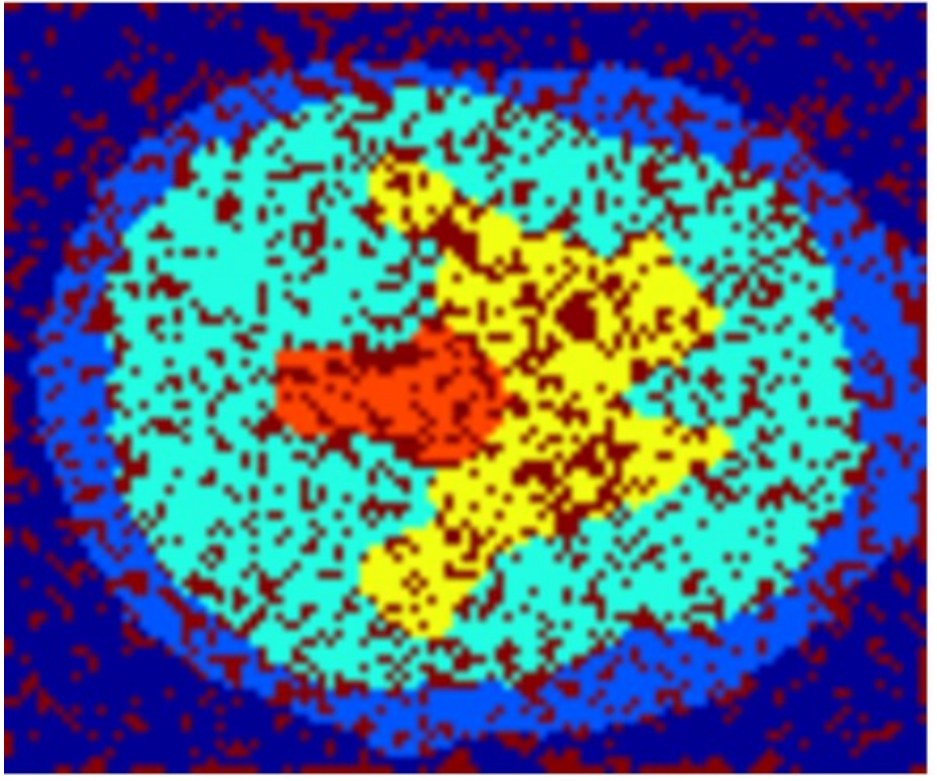




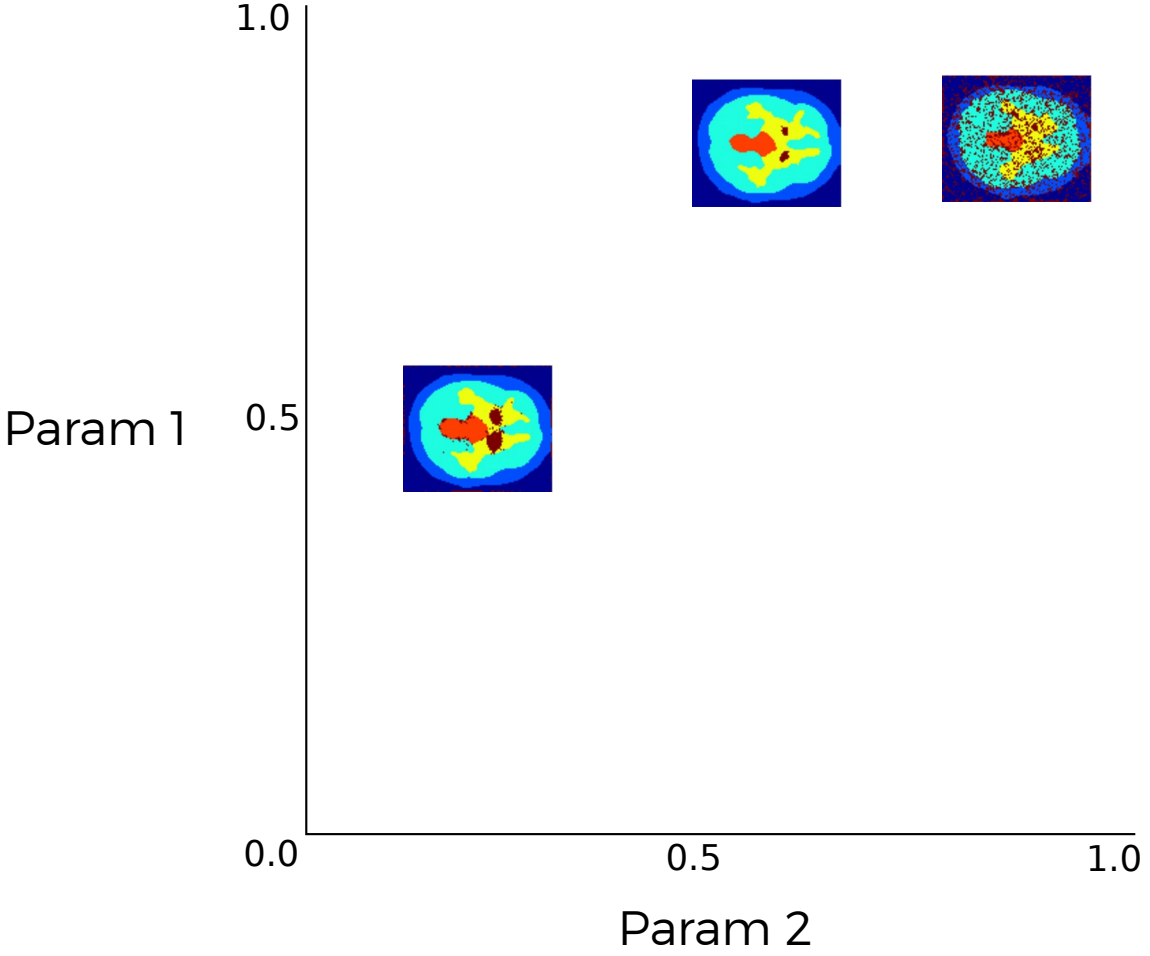
Why study meta parameters?

Meta parameters can have a large influence on performance



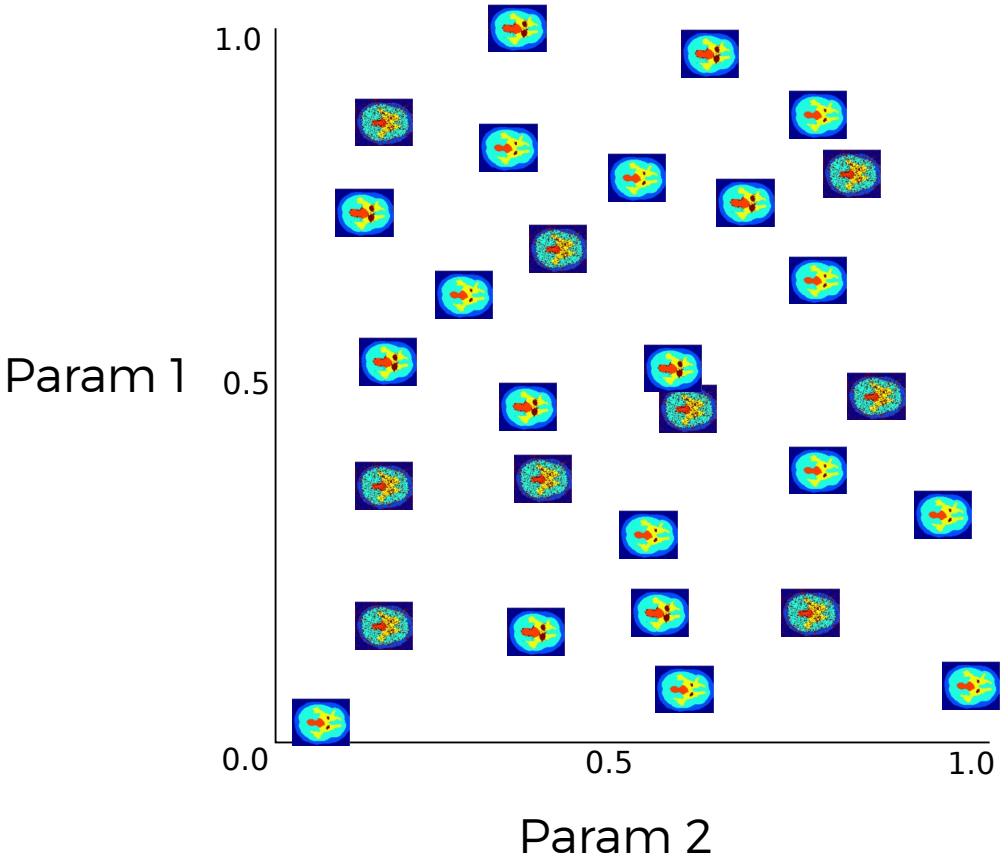


Manual method

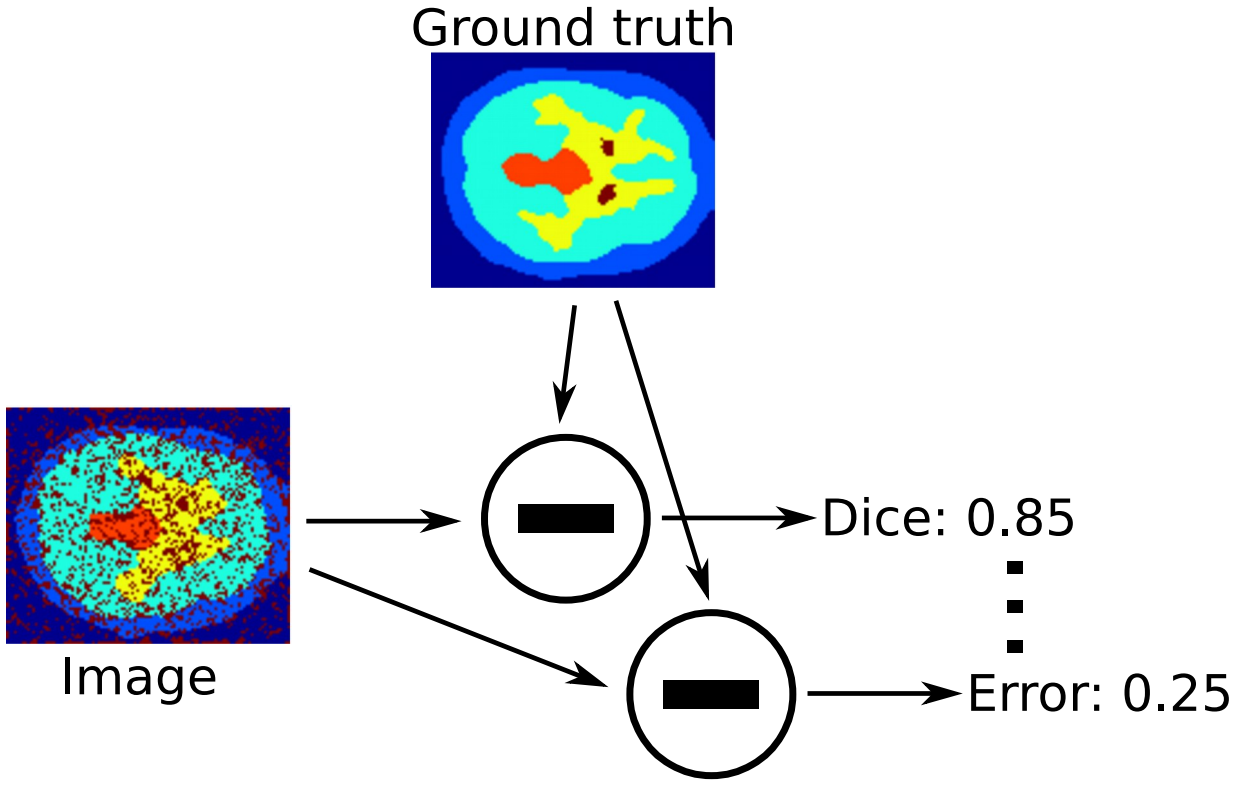


Manual method

Manual tuning is time consuming and error prone

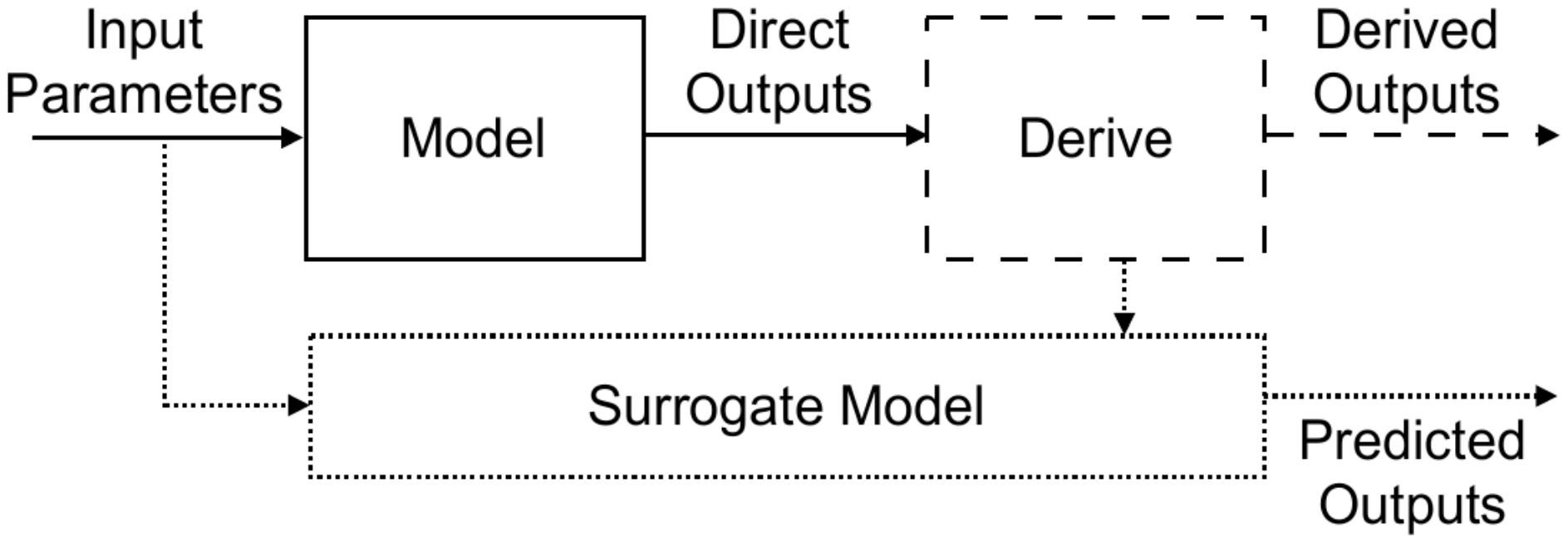


Objective measures



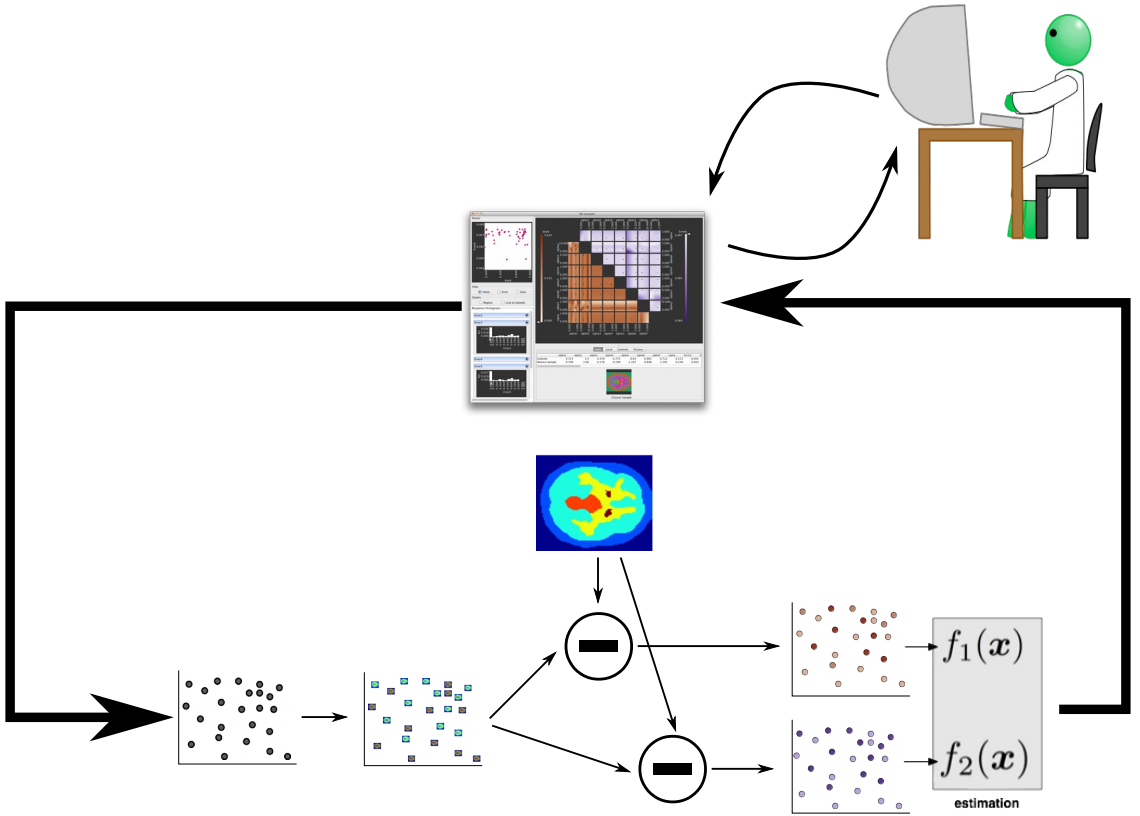
Visual parameter space exploration

Use a more principled approach

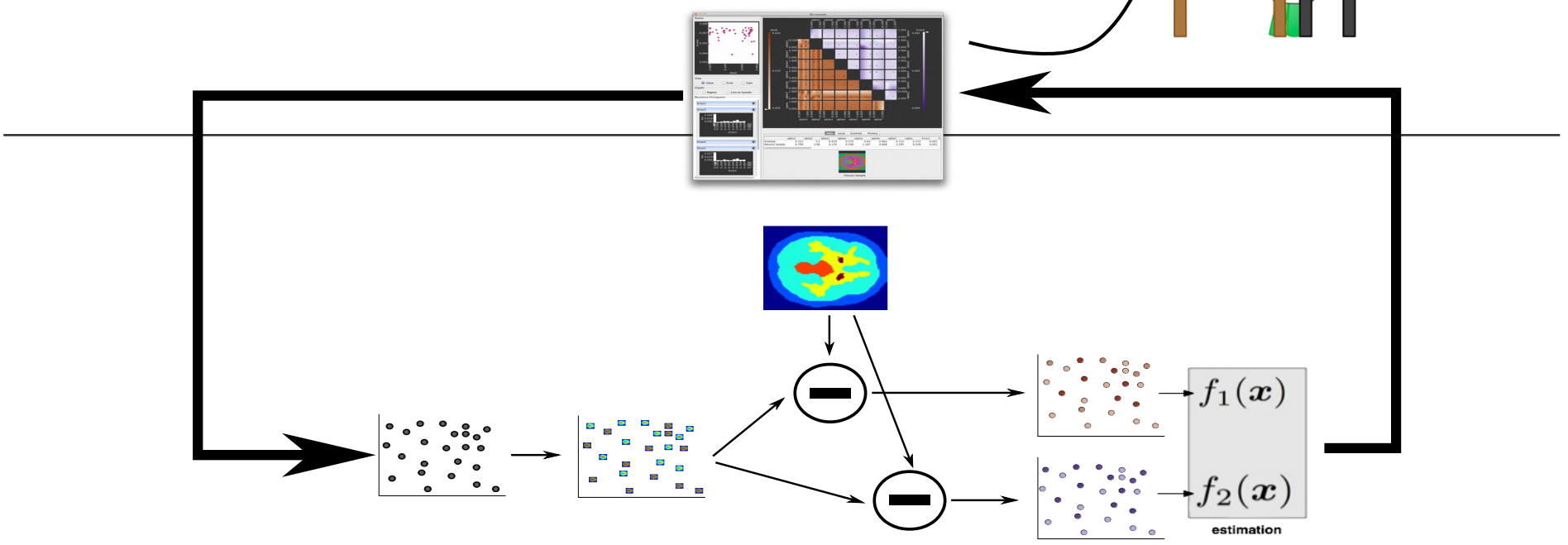


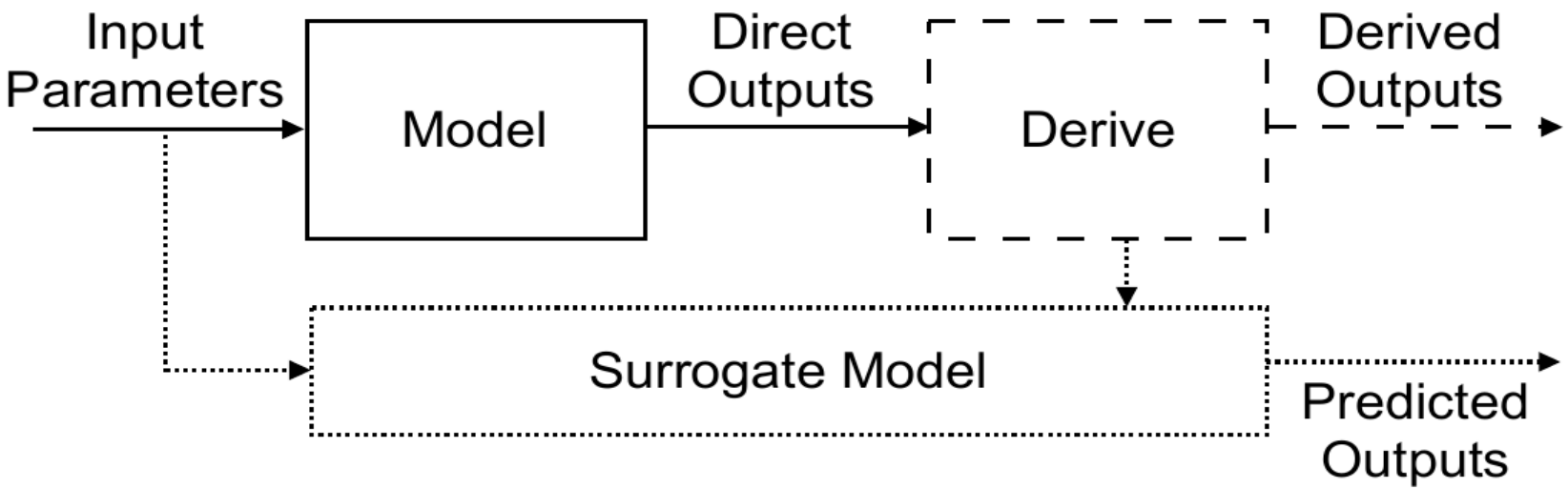
Sedlmair, Michael, Christoph Heinzl, Stefan Bruckner, Harald Piringer, and Torsten Möller. "Visual parameter space analysis: A conceptual framework," IEEE Transactions on Visualization and Computer Graphics. 2014.

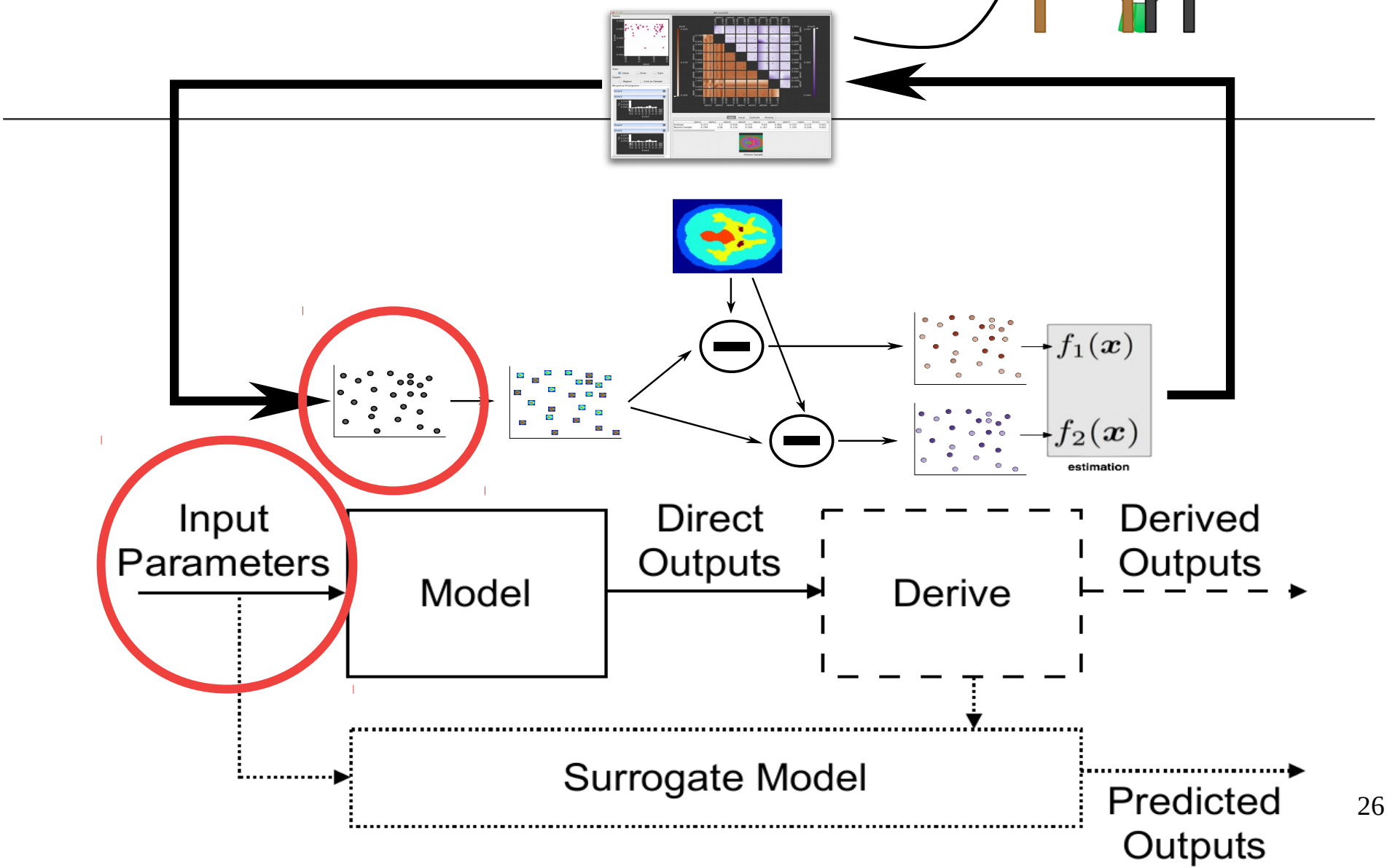
Tuner

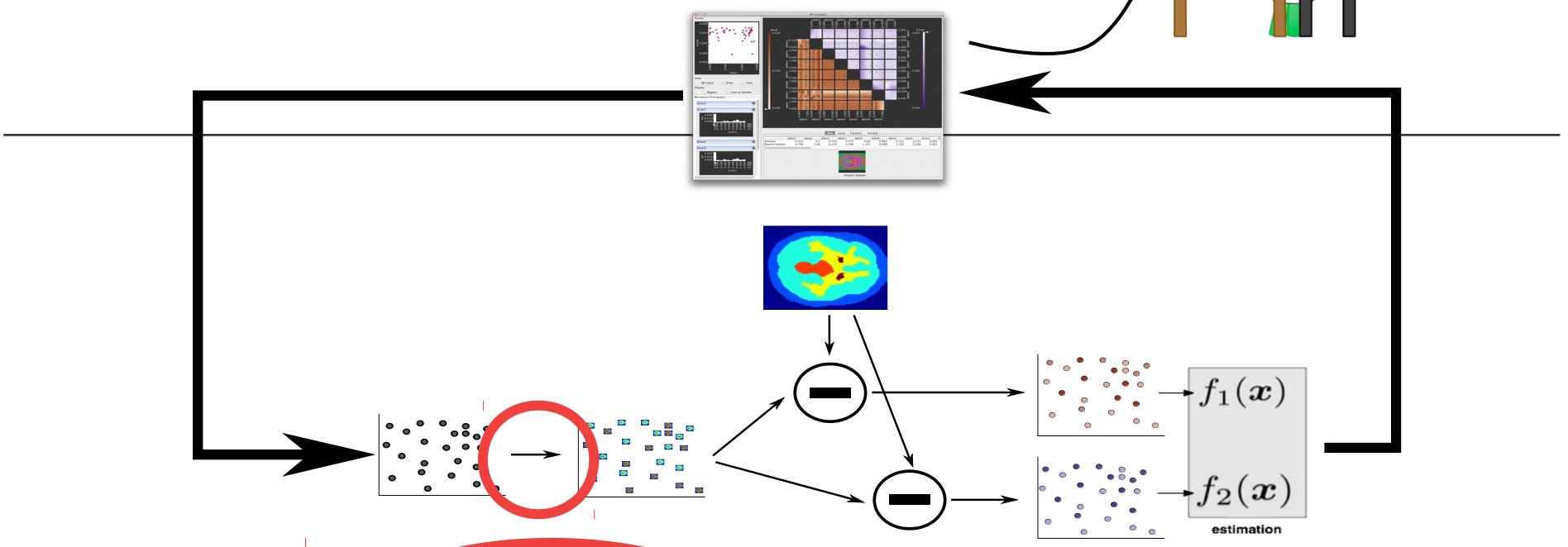


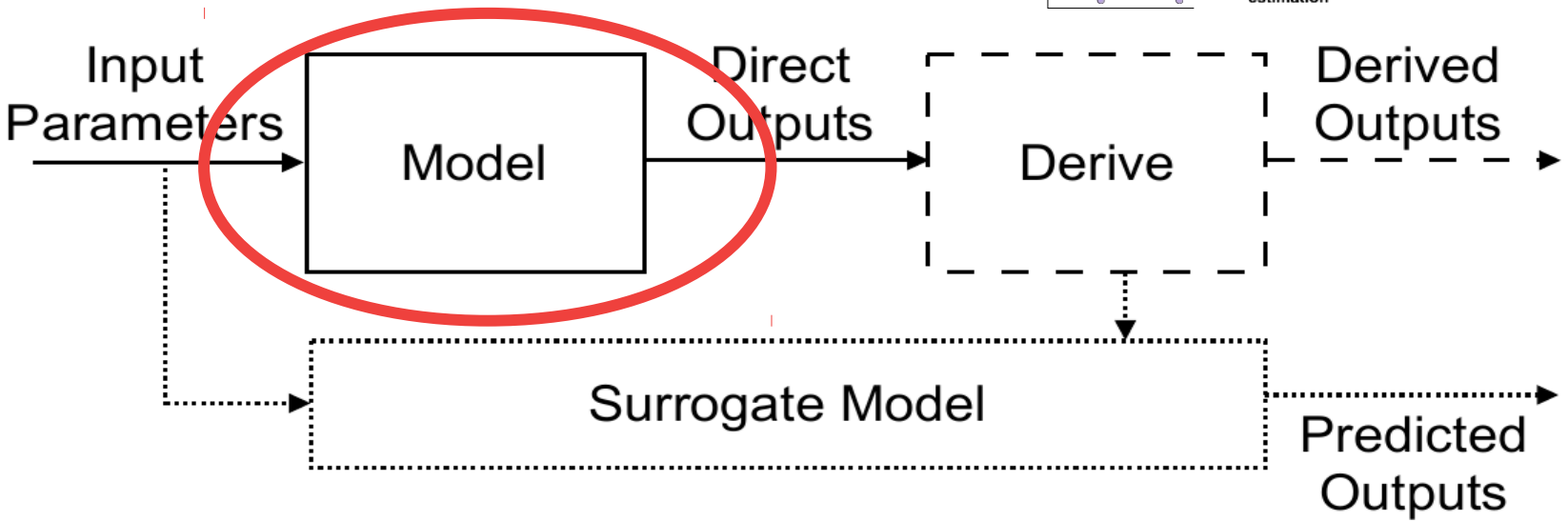
Torsney-Weir, Thomas, Ahmed Saad, Torsten Möller, Britta Weber, Hans-Christian Hege, Jean-Marc Verbavatz, and Steven Bergner. "Tuner: Principled parameter finding for image segmentation algorithms using visual response surface exploration," IEEE Transactions on Visualization and Computer Graphics. 2011.

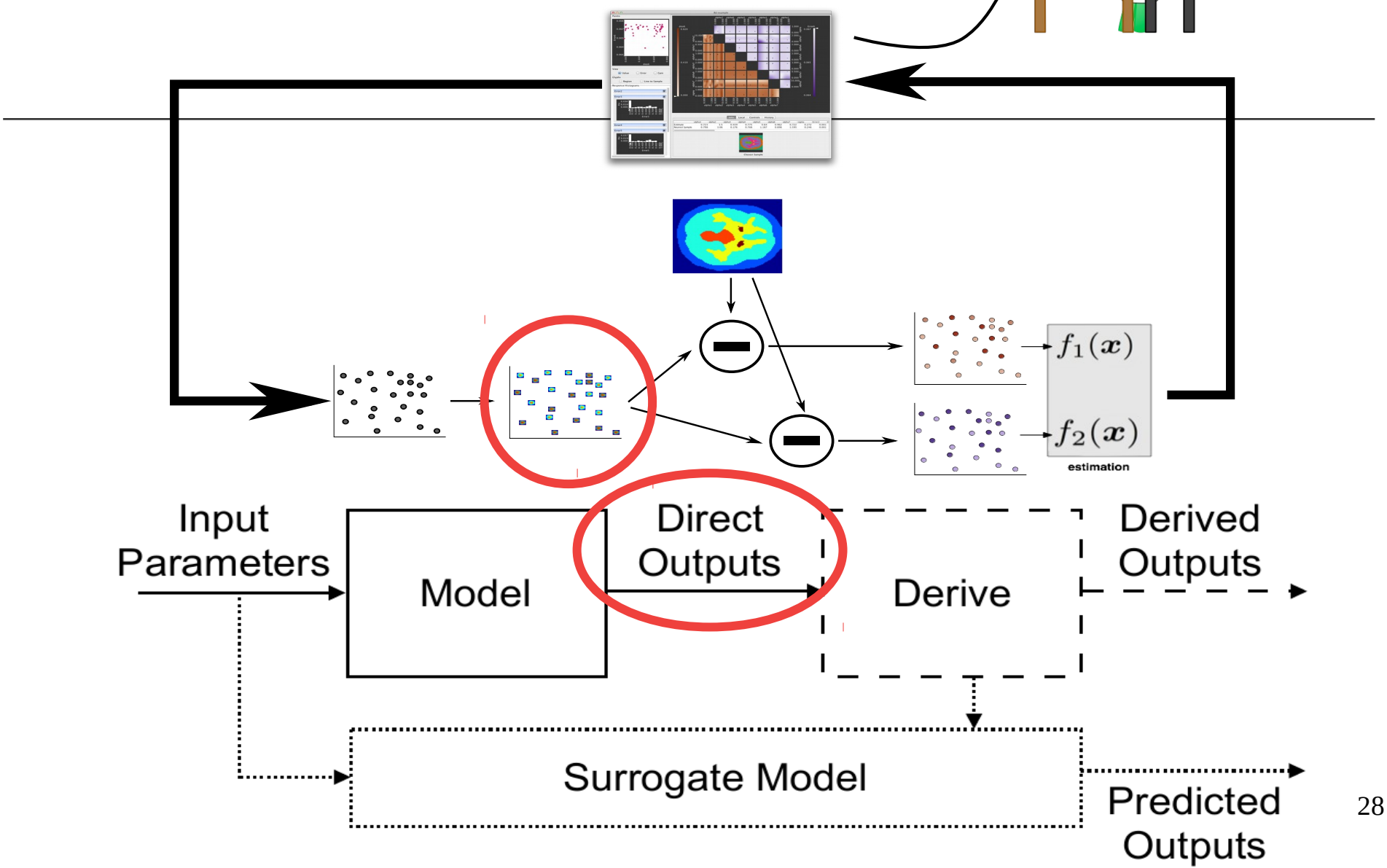

$f_1(\boldsymbol{x})$
$f_2(\boldsymbol{x})$
estimation

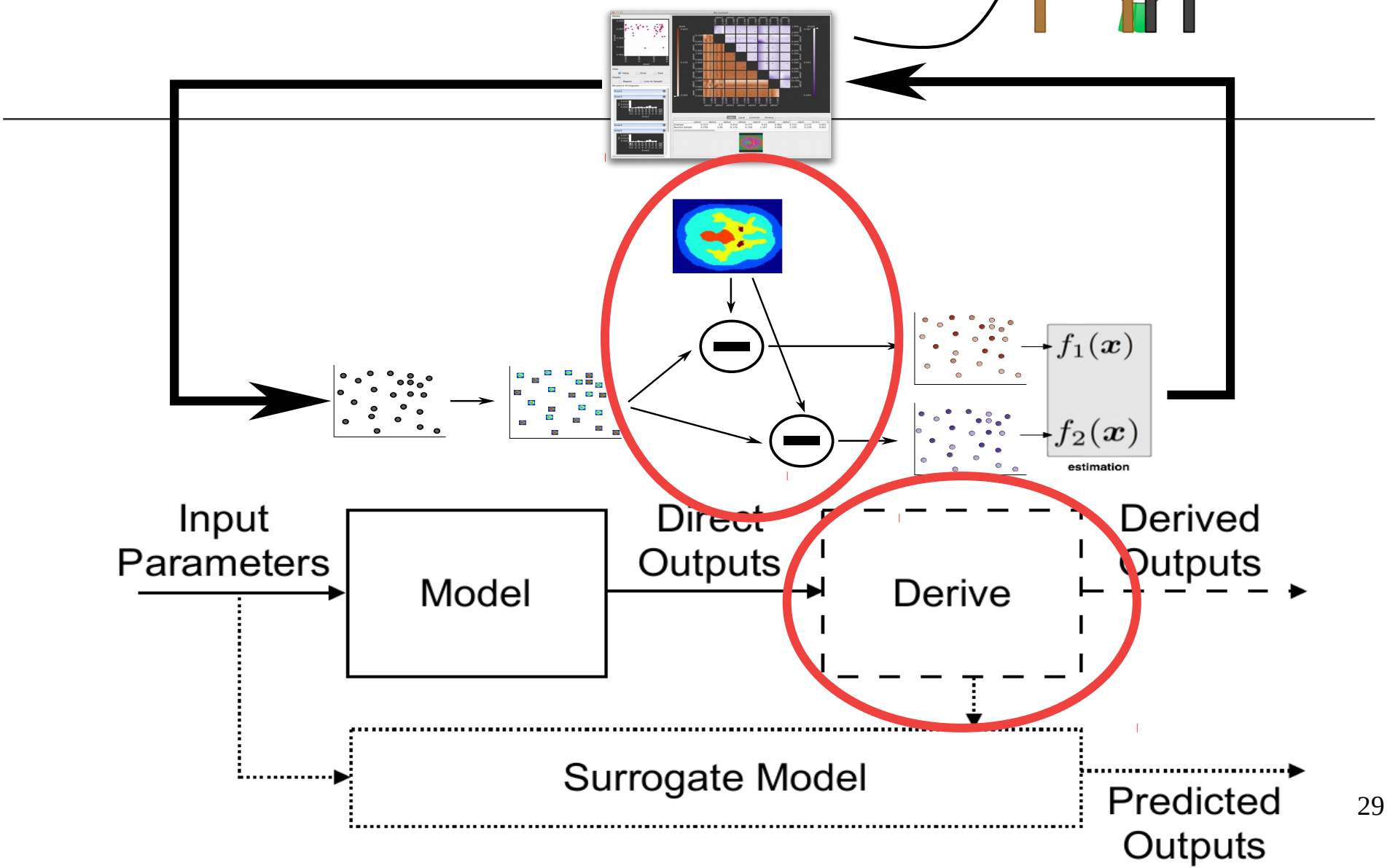

Input Parameters
Model
Direct Outputs
Derive
Derived Outputs
Surrogate Model
Predicted Outputs

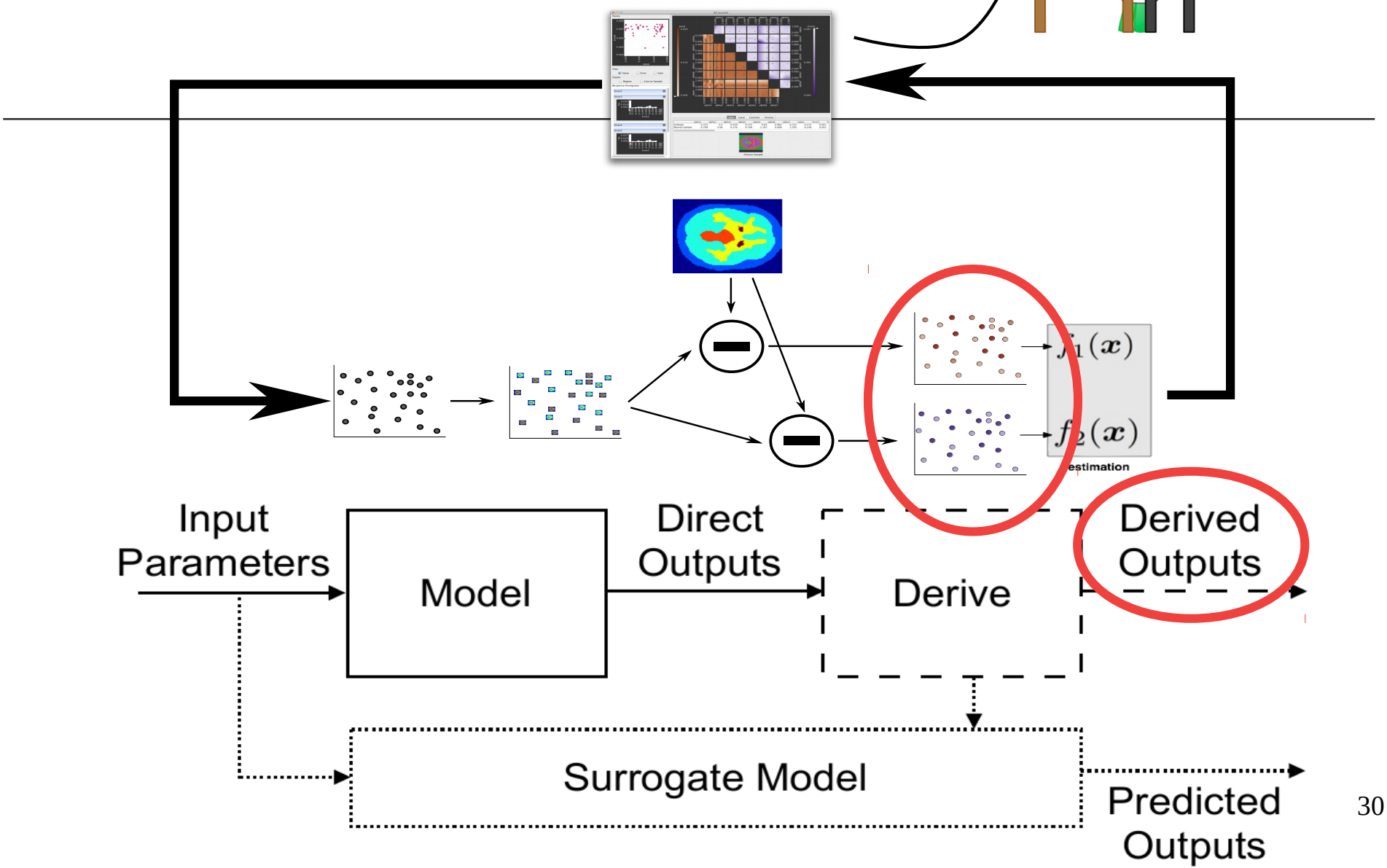

$f_1(\boldsymbol{x})$
$f_2(\boldsymbol{x})$
estimation
Input Parameters
Model
Direct Outputs
Derive
Derived Outputs
Surrogate Model
Predicted Outputs

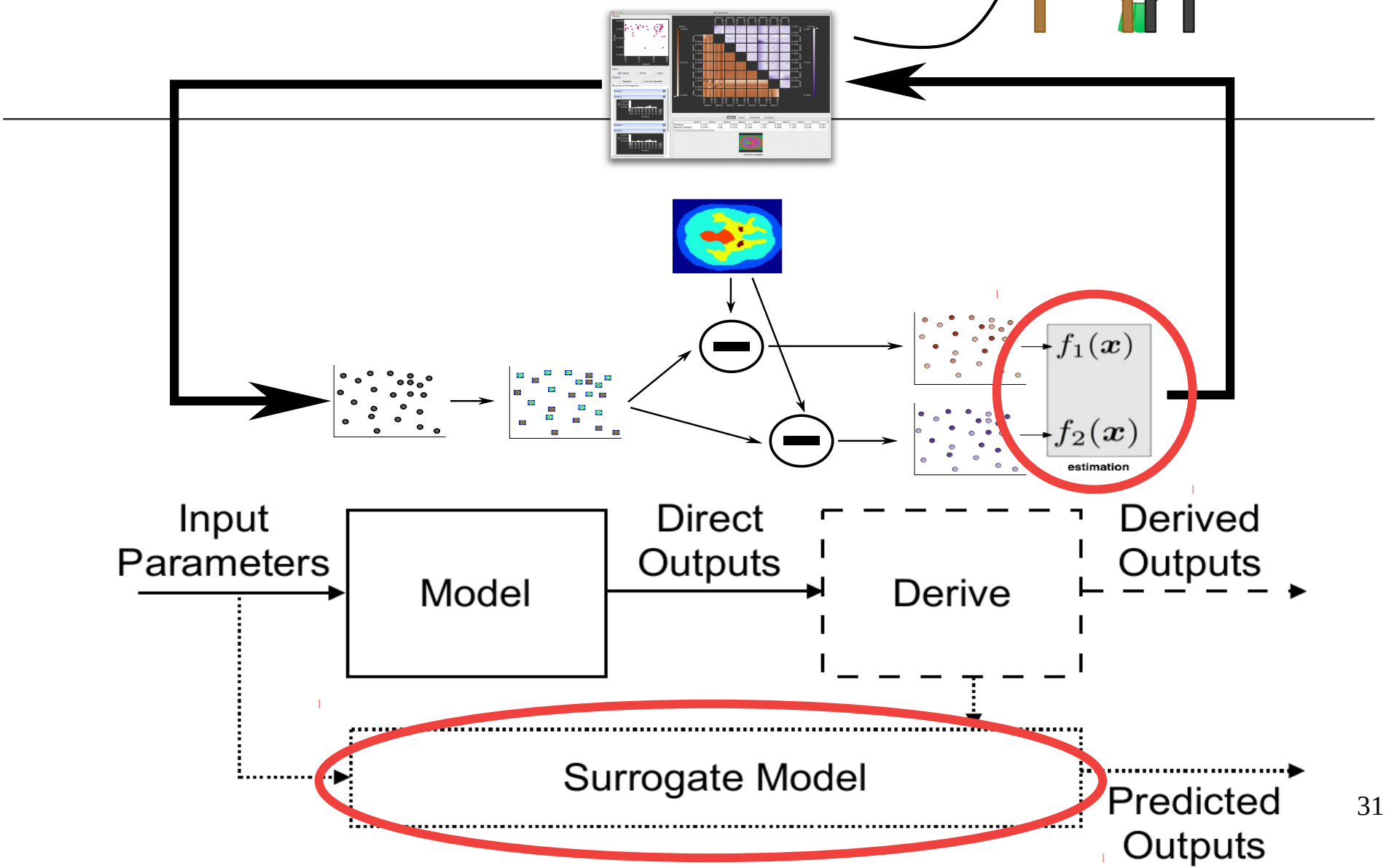

$f_1(\boldsymbol{x})$
$f_2(\boldsymbol{x})$
estimation


Input Parameters
Model
Direct Outputs
Derive
Derived Outputs
Surrogate Model
Predicted Outputs

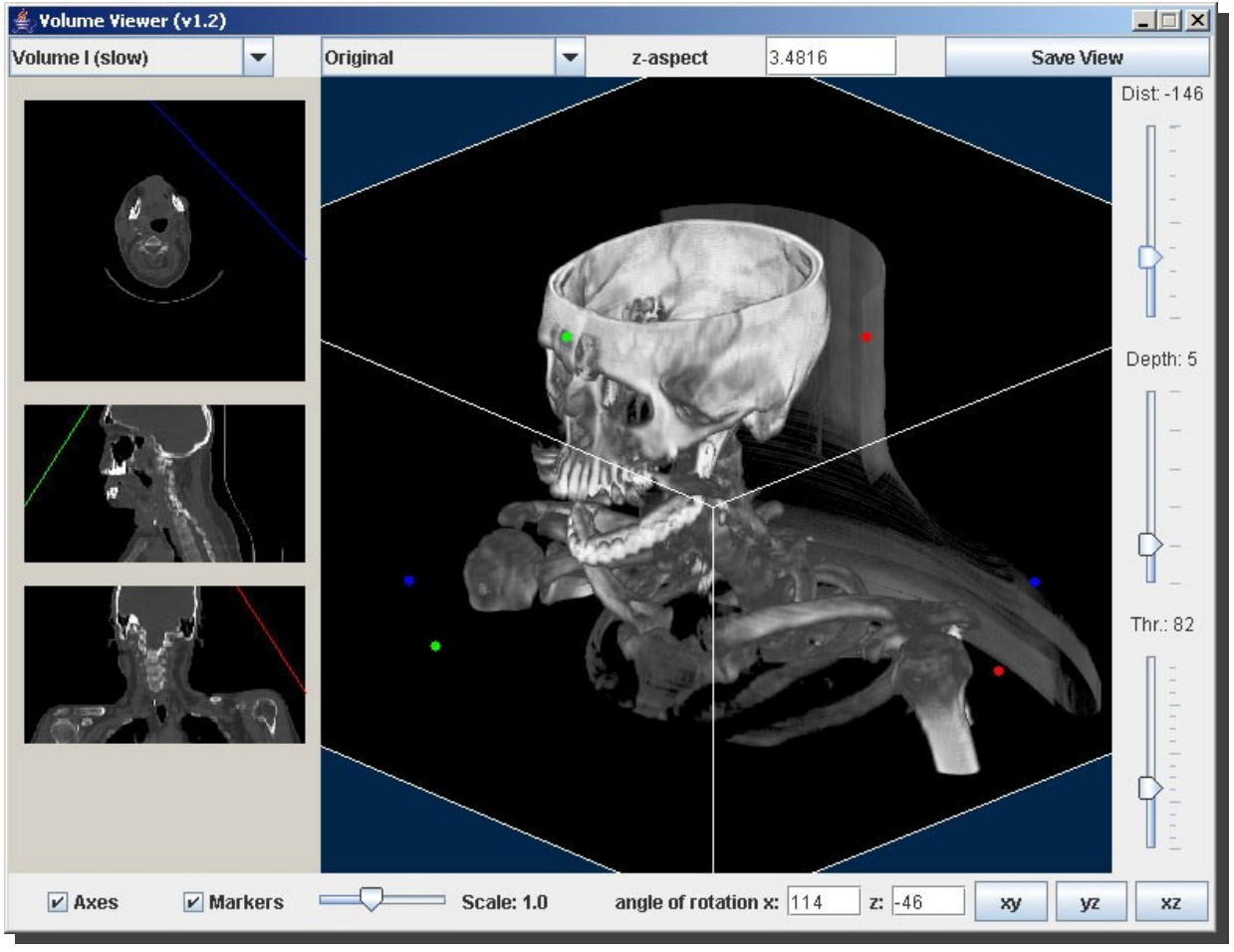

$f_1(\boldsymbol{x})$
$f_2(\boldsymbol{x})$
estimation
Input Parameters
Model
Direct Outputs
Derive
Derived Outputs
Surrogate Model
Predicted Outputs


$f_1(\boldsymbol{x})$
$f_2(\boldsymbol{x})$
estimation
Input Parameters
Model
Direct Outputs
Derive
Derived Outputs
Surrogate Model
Predicted Outputs


$f_1(\boldsymbol{x})$
$f_2(\boldsymbol{x})$
estimation
Input Parameters
Model
Direct Outputs
Derive
Derived Outputs
Surrogate Model
Predicted Outputs

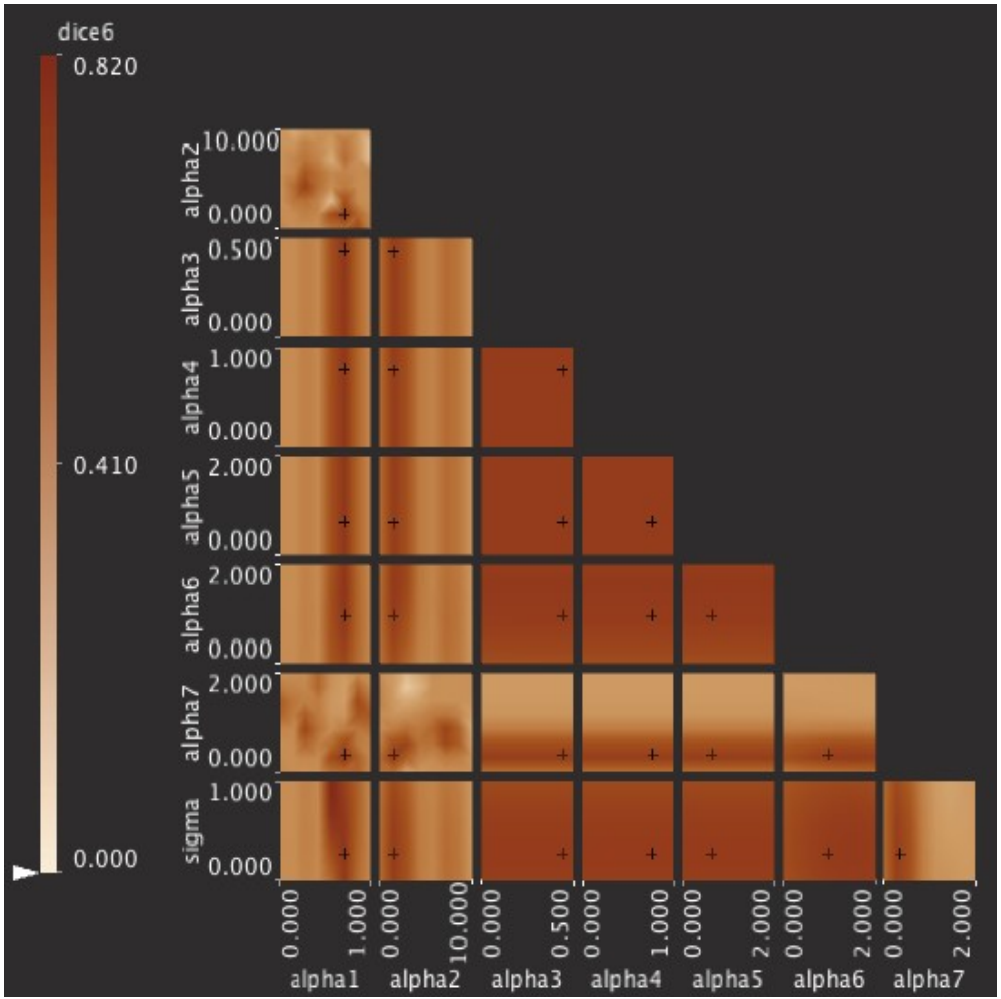

$f_1(\boldsymbol{x})$
$f_2(\boldsymbol{x})$
estimation
Input Parameters
Model
Direct Outputs
Derive
Derived Outputs
Surrogate Model
Predicted Outputs

Tuner

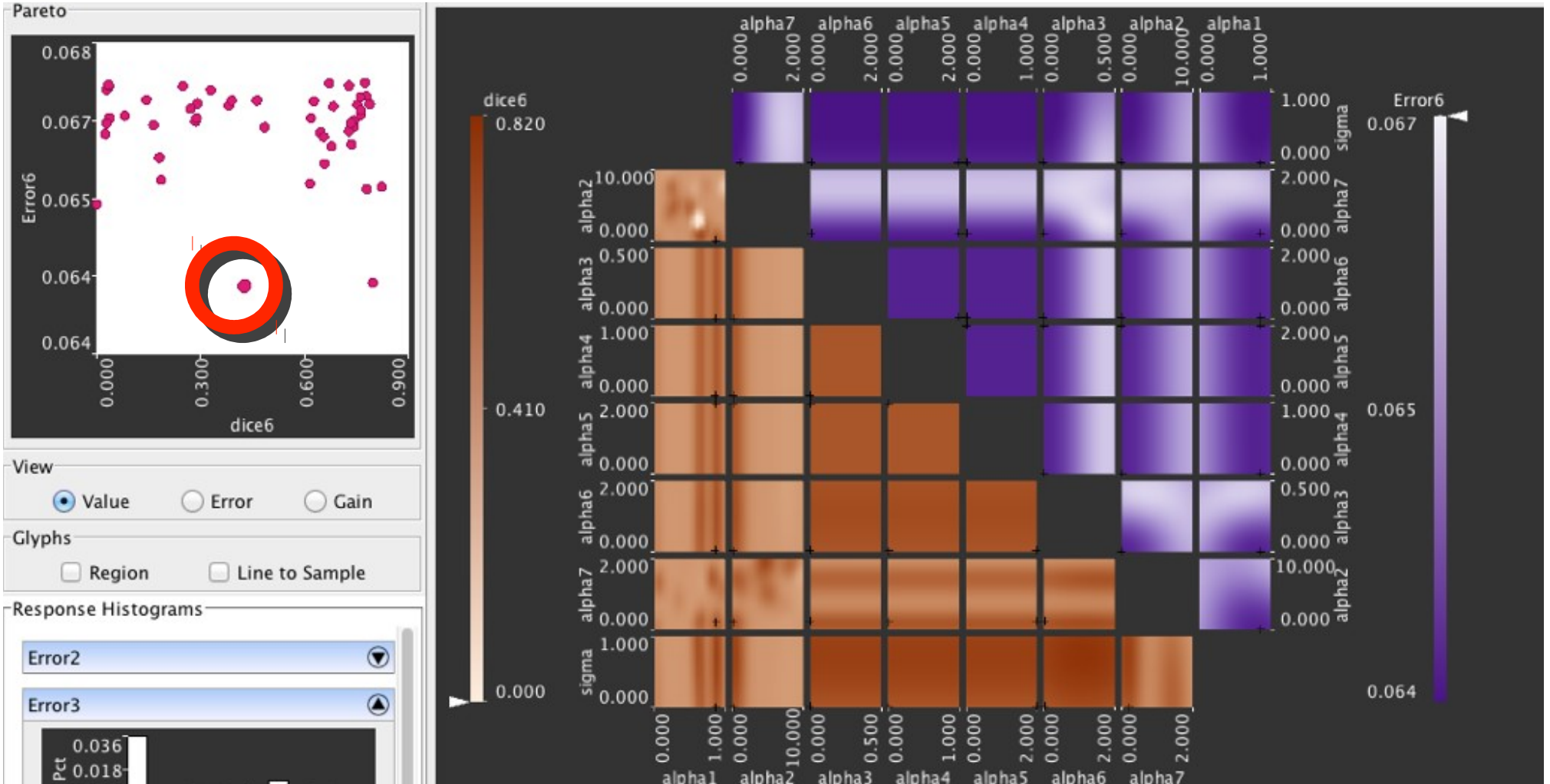


https://imagej.nih.gov/ij/plugins/volume-viewer.html

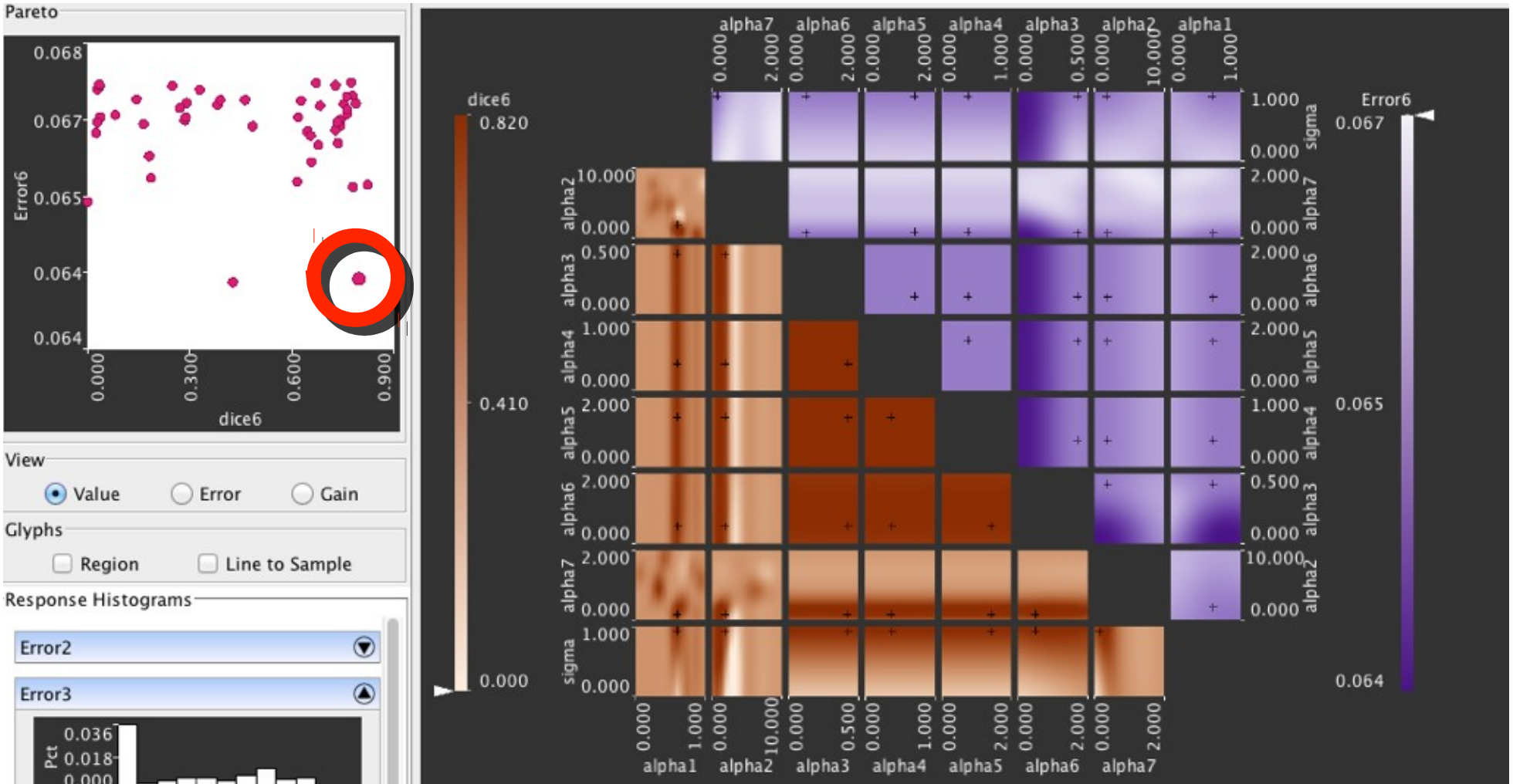
Tuner


dice6
0.820
0.410
0.000
alpha2
10.000
0.000
alpha3
0.500
0.000
alpha4
1.000
0.000
alpha5
2.000
0.000
alpha6
2.000
0.000
alpha7
2.000
0.000
sigma
1.000
0.000
0.000
1.000
alpha1
0.000
10.000
alpha2
0.000
0.500
alpha3
0.000
1.000
alpha4
0.000
2.000
alpha5
0.000
2.000
alpha6
0.000
2.000
alpha7

Tuner


Pareto
0.068
0.067
0.065
0.064
Error6
0.000
0.300
0.600
0.900
dice6
View
Value
Error
Gain
Glyphs
Region
Line to Sample
Response Histograms
Error2
Error3
0.036
0.018
Pct
dice6
0.820
0.410
0.000
alpha7
alpha6
alpha5
alpha4
alpha3
alpha2
alpha1
sigma
Error6
0.067
0.065
0.064

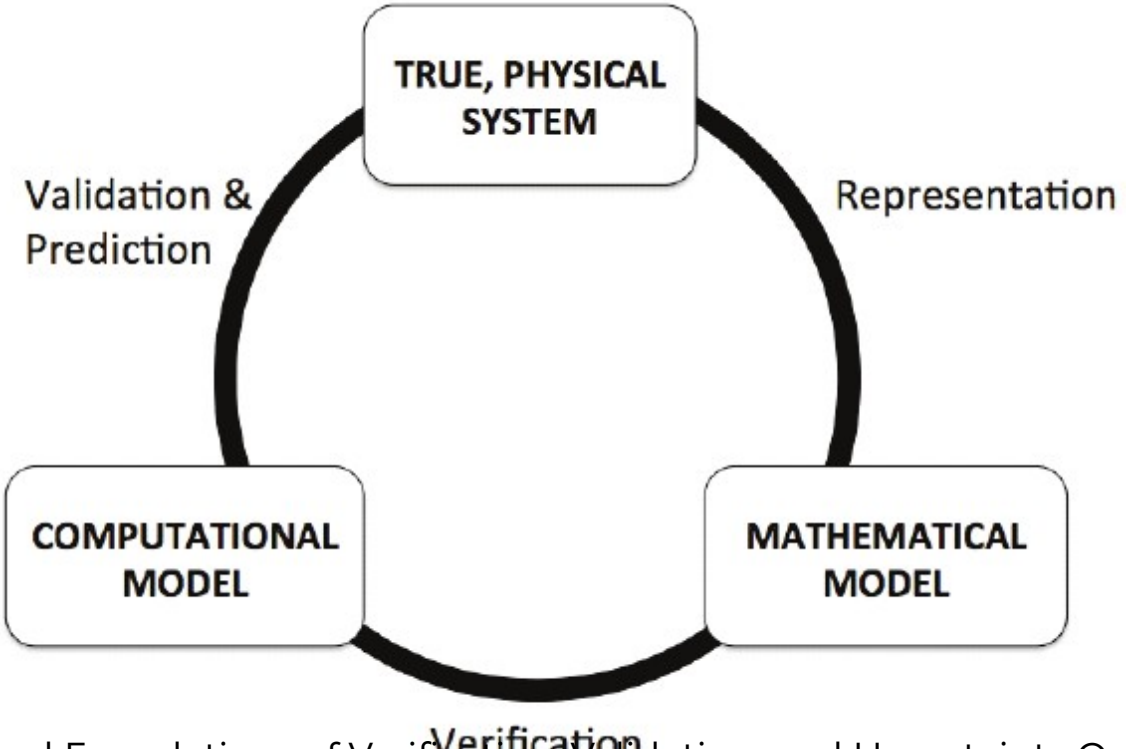
Tuner


Pareto
0.068
0.067
0.065
0.064
0.064
Error6
0.000
0.300
0.600
0.900
dice6
View
Value
Error
Gain
Glyphs
Region
Line to Sample
Response Histograms
Error2
Error3
0.036
0.018
0.000
Pct
dice6
0.820
0.410
0.000
alpha7
alpha6
alpha5
alpha4
alpha3
alpha2
alpha1
sigma
Error6
0.067
0.065
0.064

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

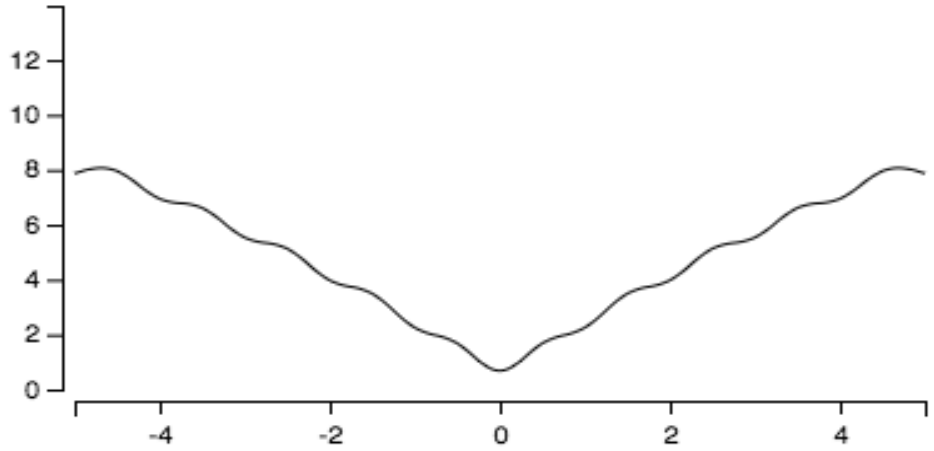
- What is the best model for my data?
- How well will these models generalize?
- Summary statistics are not always enough
- Balancing multiple objectives is difficult
- Certain training points might be very important

Example

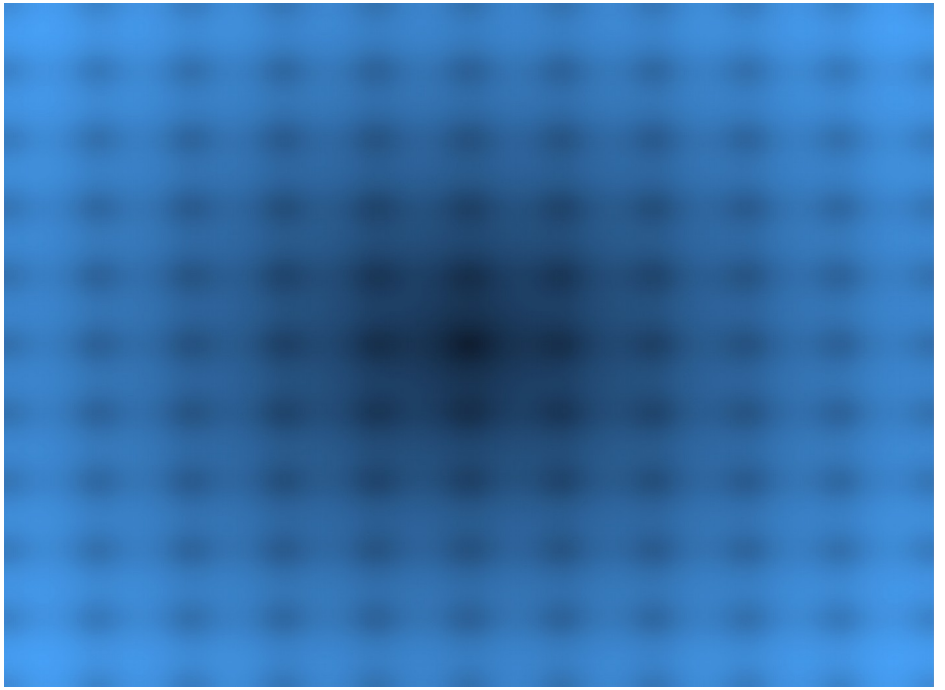
Boston housing dataset

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

Function inspection



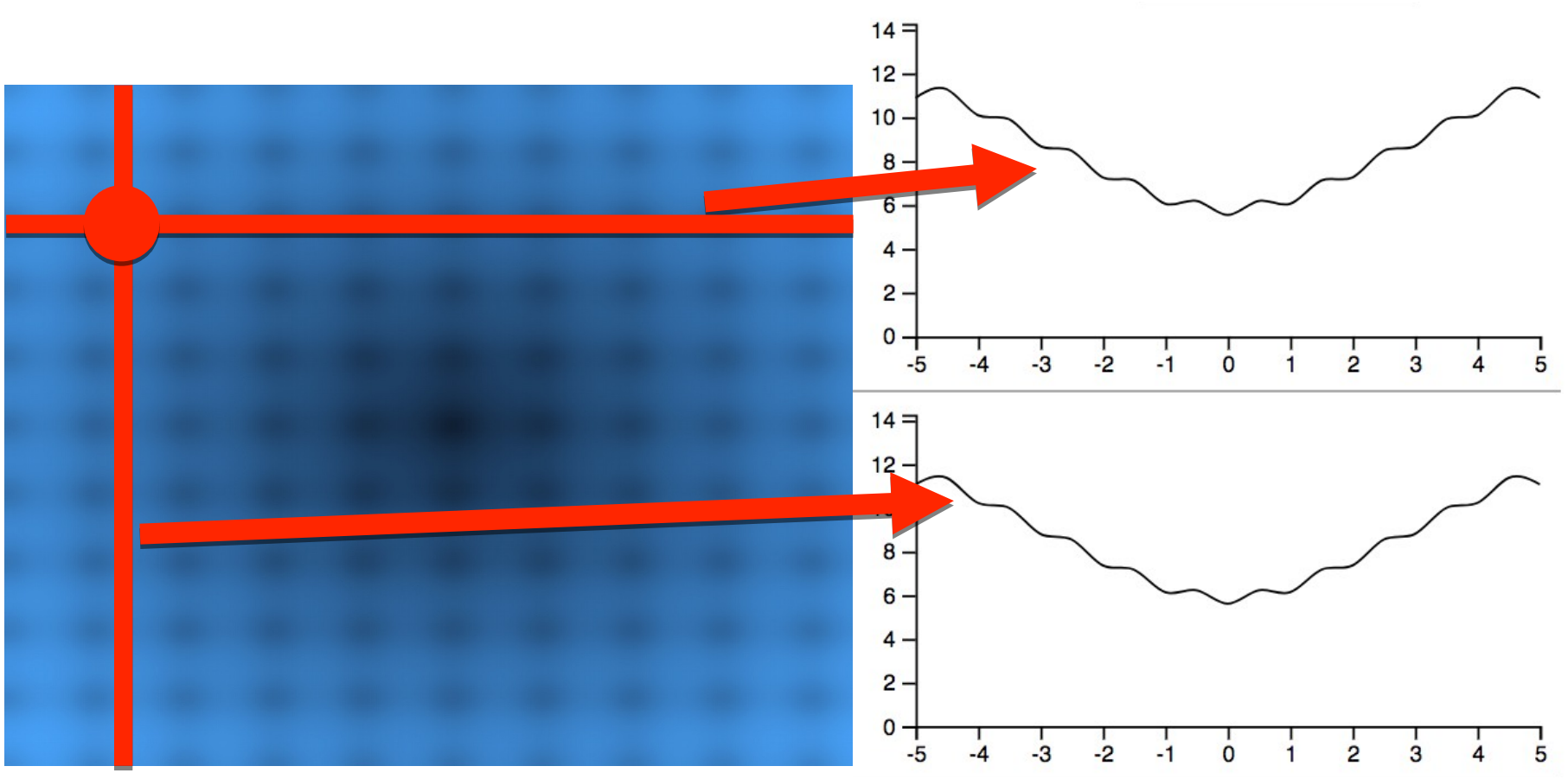
1D function



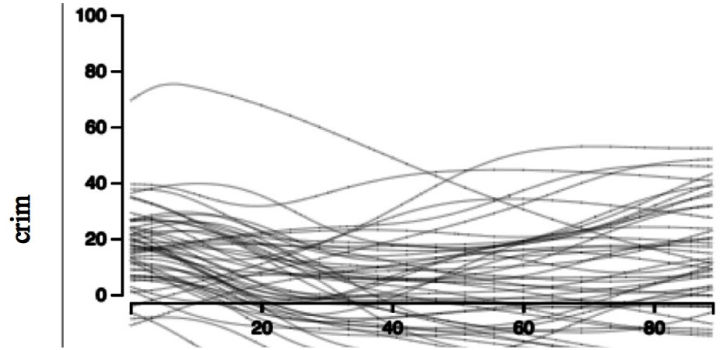
2D function

More than this gets hard!

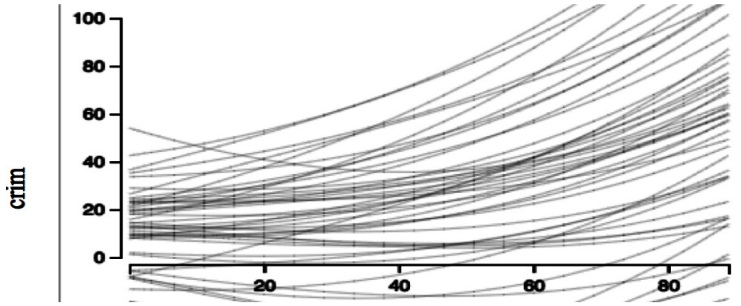
Slicing



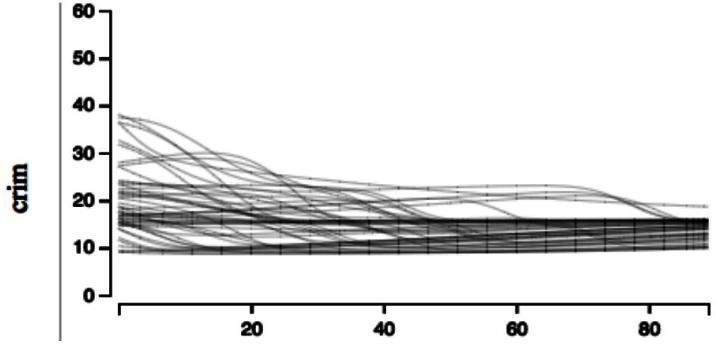
Sliceplorer views



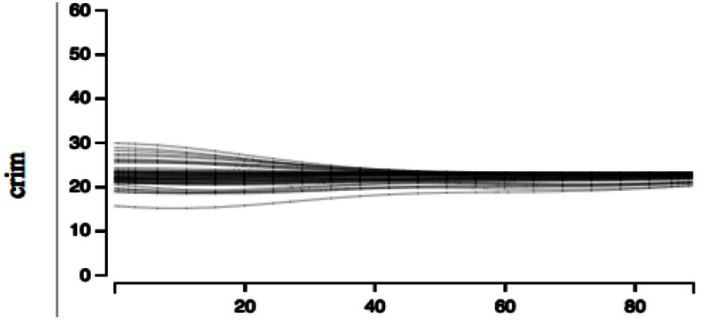
NN (single layer)



SVM (polynomial)



NN (dual layer)



SVM (RBF)

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

Summary

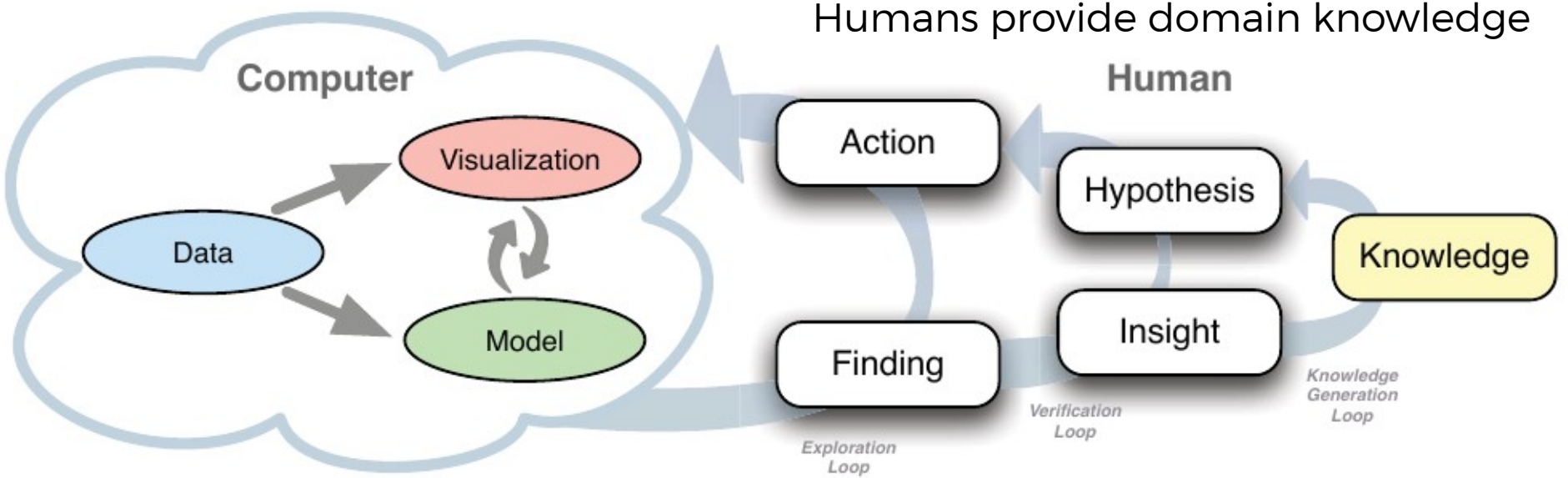
- Just an answer is not enough (context)
- Humans have trouble understanding complex models
- Interactivity can bring people into the model

Machine learning helping visualization

How?

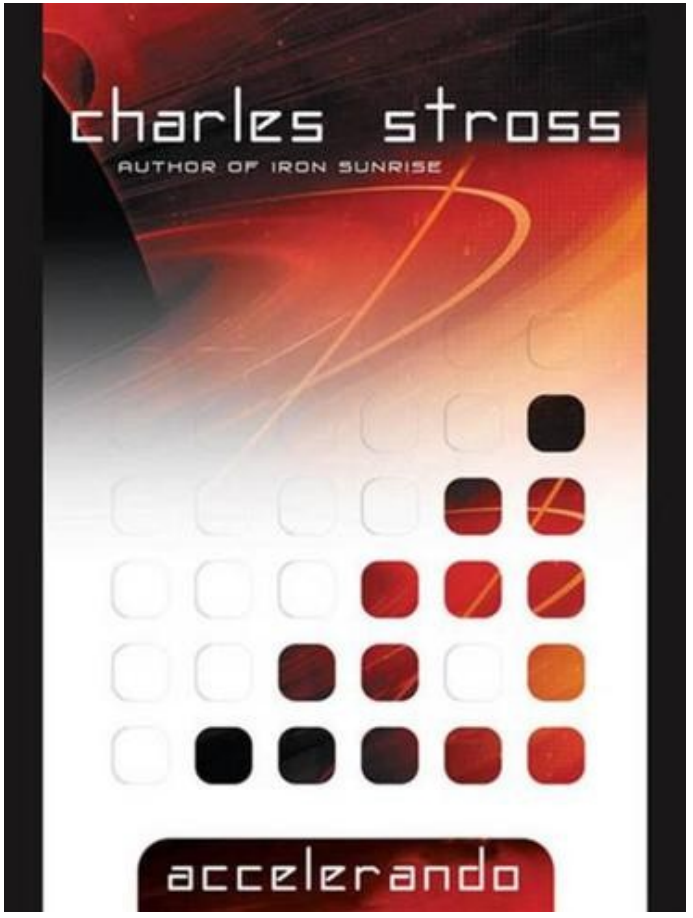
Computers are good at calculating

Humans provide domain knowledge



Sacha, D., A. Stoffel, F. Stoffel, Bum Chul Kwon, G. Ellis, and Daniel A Keim. "Knowledge generation model for visual analytics," Visualization and Computer Graphics, IEEE Transactions on. 2014.

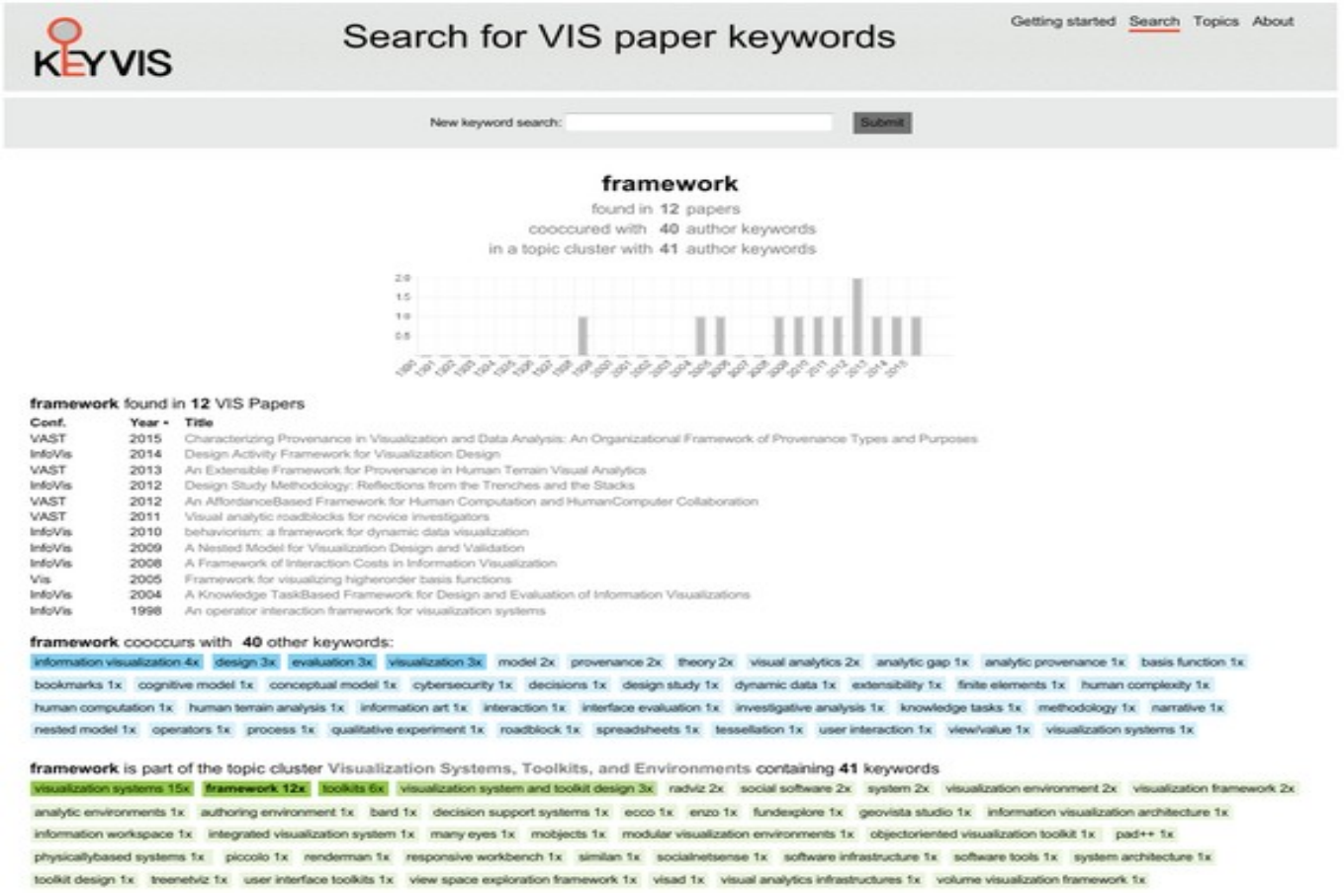
Book ad!



Methods

- Clustering
- Classification
- Regression

KeyVis



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.

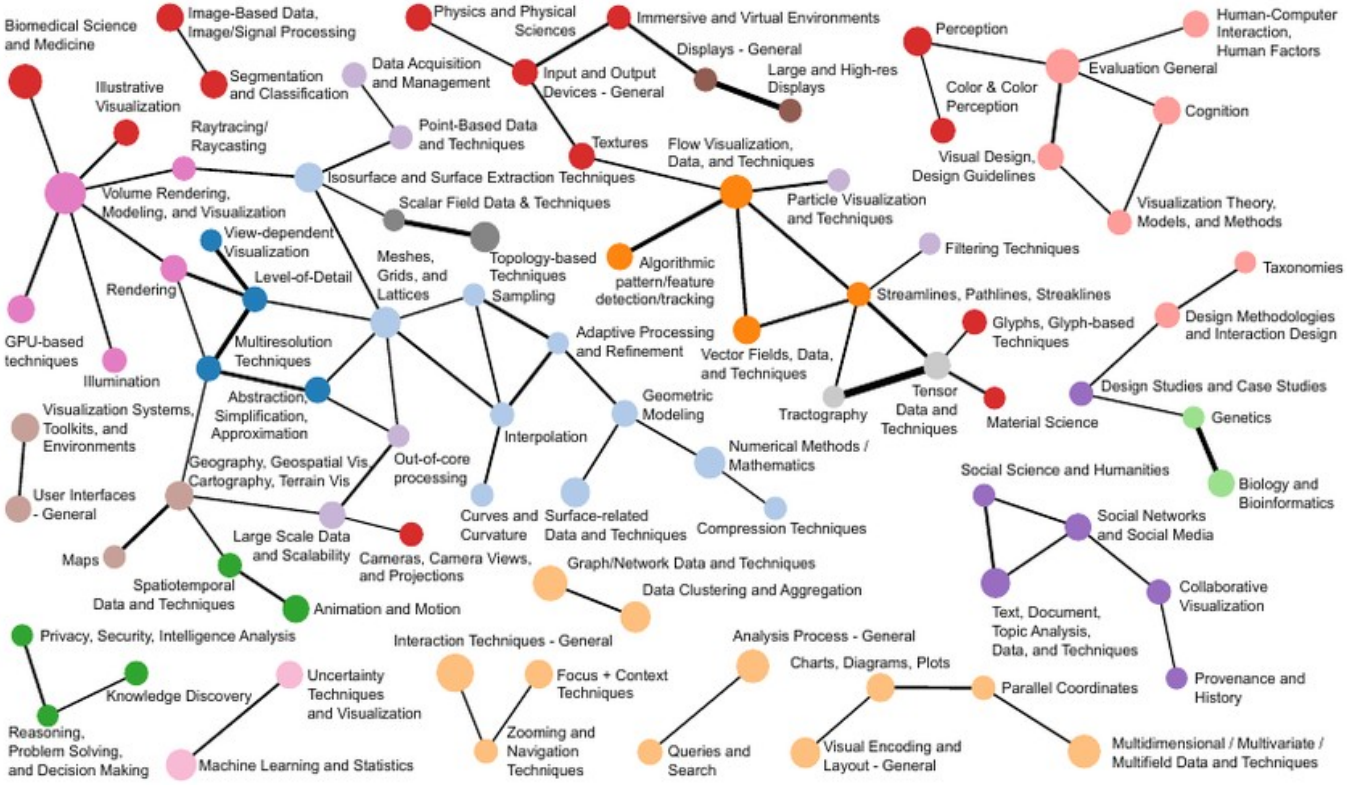
KeyVis

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



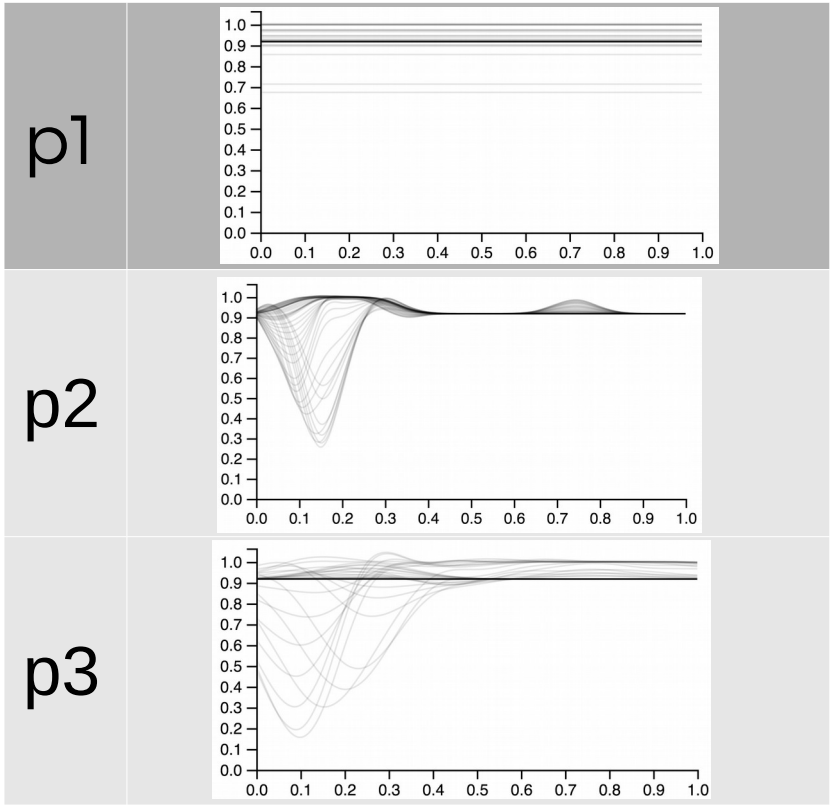
KeyVis

step 2: build an interface for the clustering

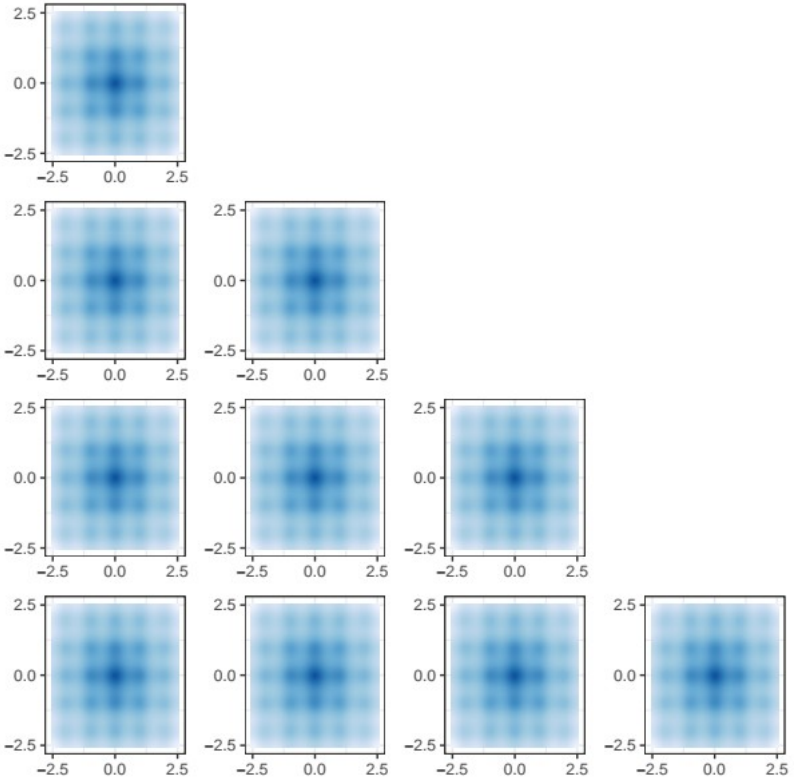
Regression

Slicing

1D slices



2D slices



The future!

Interesting projects

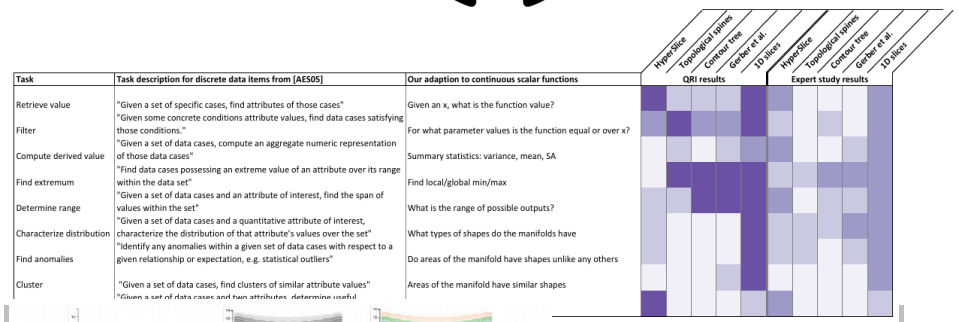
- How can humans and machines work together?
- Visualizing regression algorithms
- Understand what "understandability" means

Thanks!

www.tomtorsneyweir.com

thomas.torsney-weir@univie.ac.at

@gabysbrain



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 those data cases"	Summary statistics: variance, mean, SA
Find extremum	"Find data cases possessing an extreme value of an attribute over its range within the data set"	Find local/global min/max
Determine range	"Given a set of data cases and an attribute of interest, find the span of values within the set"	What is the range of possible outputs?
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"	What types of shapes do the manifolds have
Find anomalies	"Identify any anomalies within a given set of data cases with respect to a given relationship or expectation, e.g. statistical outliers"	Do areas of the manifold have shapes unlike any others
Cluster	"Given a set of data cases, find clusters of similar attribute values"	Areas of the manifold have similar shapes
	"Given a set of data cases and two attributes, determine useful