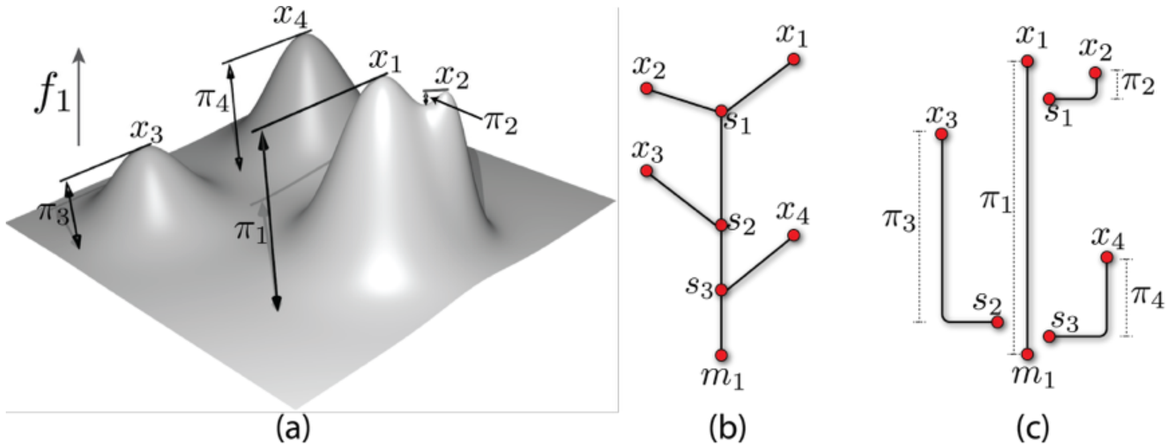
Using Maximum Topology Matching to Explore Differences in Species Distribution Models



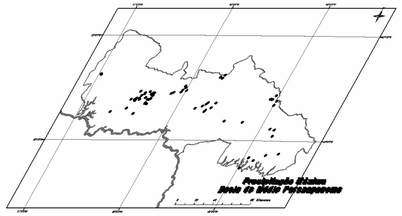
Jorge Poco
Harish Doraiswamy
Marian Talbert
Jeffrey Morisette
Cláudio Silva

NYU TANDON SCHOOL OF ENGINEERING

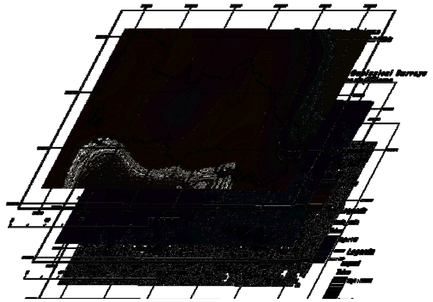
USGS science for a changing world

Species Distribution Models (SDM)

Observations



Environmental layers



Algorithm

Geographic position

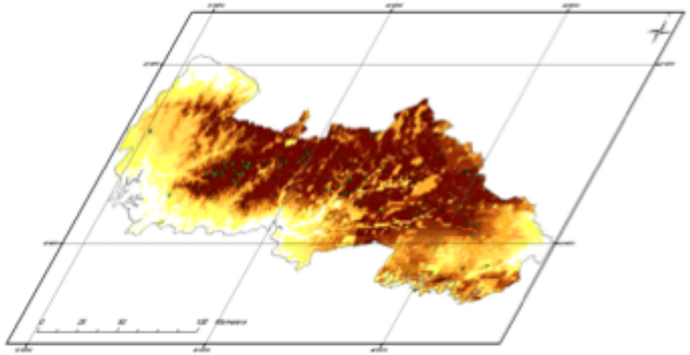
Temperature

Precipitation

Topography

NDVI

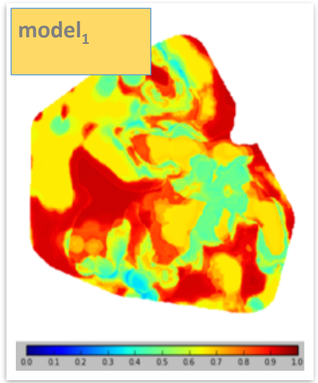
Predictive distribution

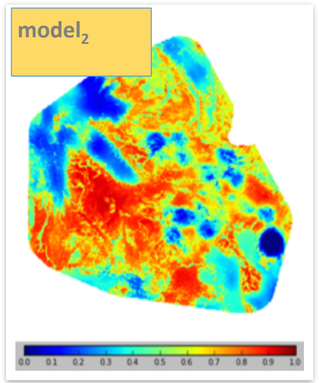


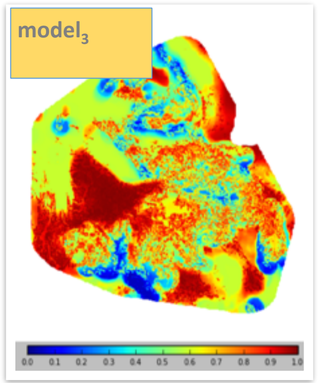
$$f : \mathbb{R}^d \rightarrow [0, 1]$$

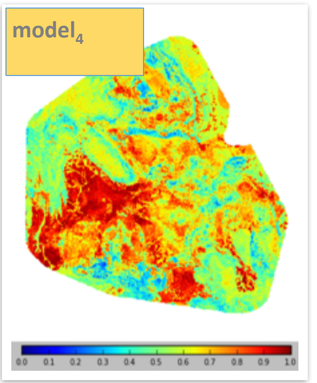
Model Inter-comparison

Predict Distribution









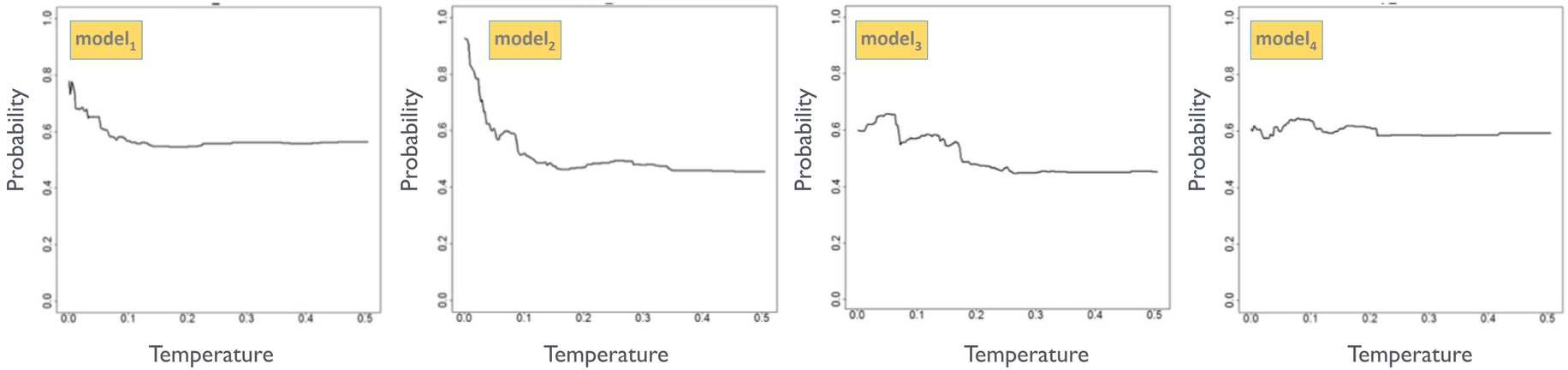
Why are they different?
How do we compare them?

Ecologist's Approach

Visualize one-dimension at a time (default response curve)

Predictors

Temperature	[0, 0.5]
Precipitation	mean
Topography	mean
NDVI	mean
…	…



SHORTCOMINGS:

- Restricts the analysis to one dimension at time
- Interaction between dimensions are lost because of dimensionality reduction

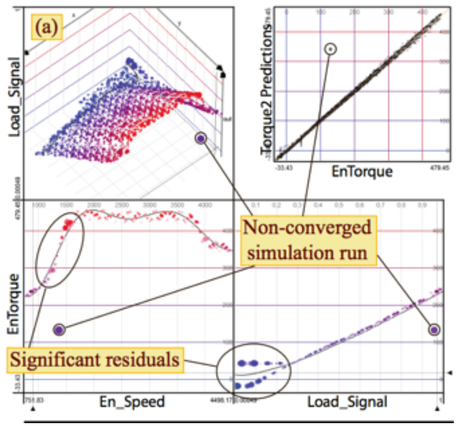
Goals

- **Explore** SDMs in the high dimensional domain
- **Compare** different SDMs

Our approach is to use the **topology of SDMs** for exploring and comparing them

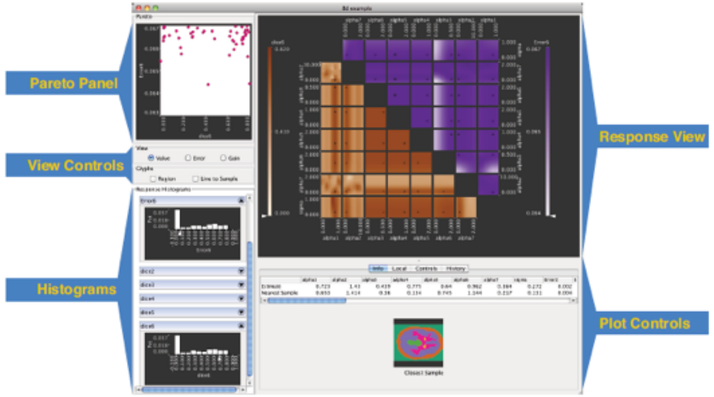
Challenges: Exploring High Dimensional Functions

[HyperMoVal]



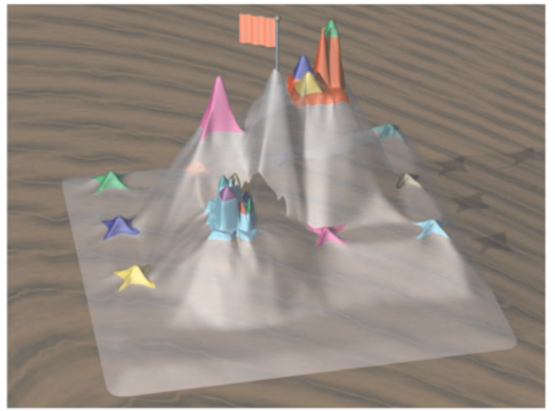
[Piringer et al. 2010]

[Tuner]



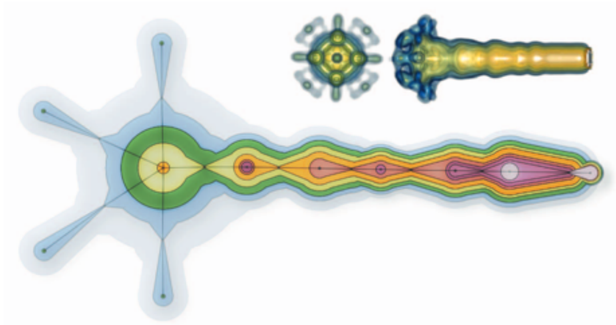
[Torsney-Weir et al. 2011]

[Topological Landscape]



[Weber et al. 2007]

[Topological Spines]

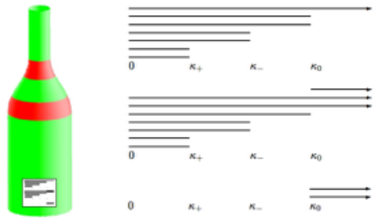


[Correa et al. 2011]

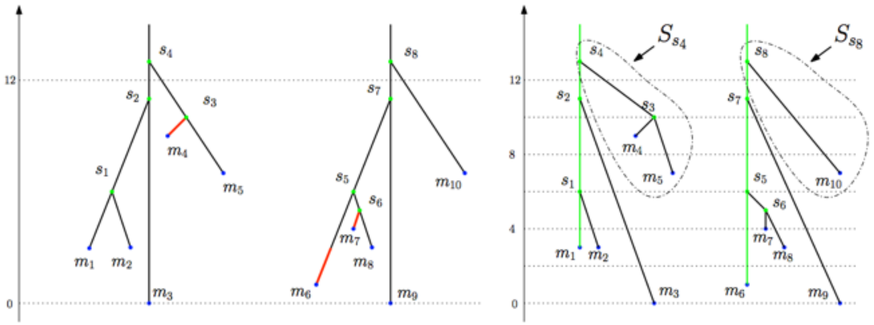
- Shortcoming:
  - Require to manually explore and compare two high dimensional spaces

Challenges: Comparing Scalar Functions

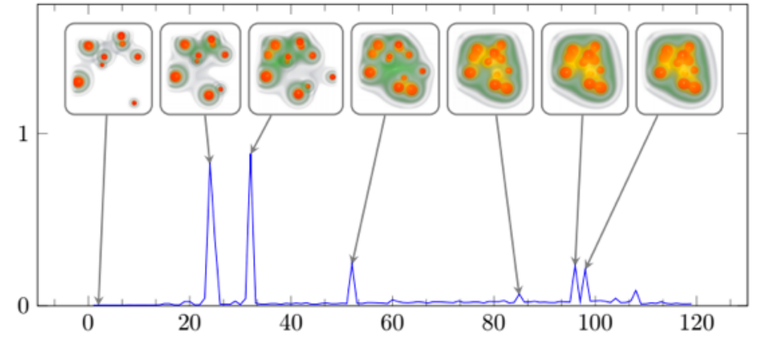
- Existing work:
  - Distance between visual representations of topological persistence
    [Carlsson et al. 2004][Cohen-Steiner et al. 2007]
  - Distance between topological data structures
    - Merge trees / Reeb graphs
      [Morozov et al. 2013][Beketayev et al. 2014][Bauer et al . 2014]
    - Extremum graphs
      [Narayanan et al. 2015]
- Shortcomings:
  - Location of critical points not considered
  - Similar features can be far away



[Carlsson et al. 2004]

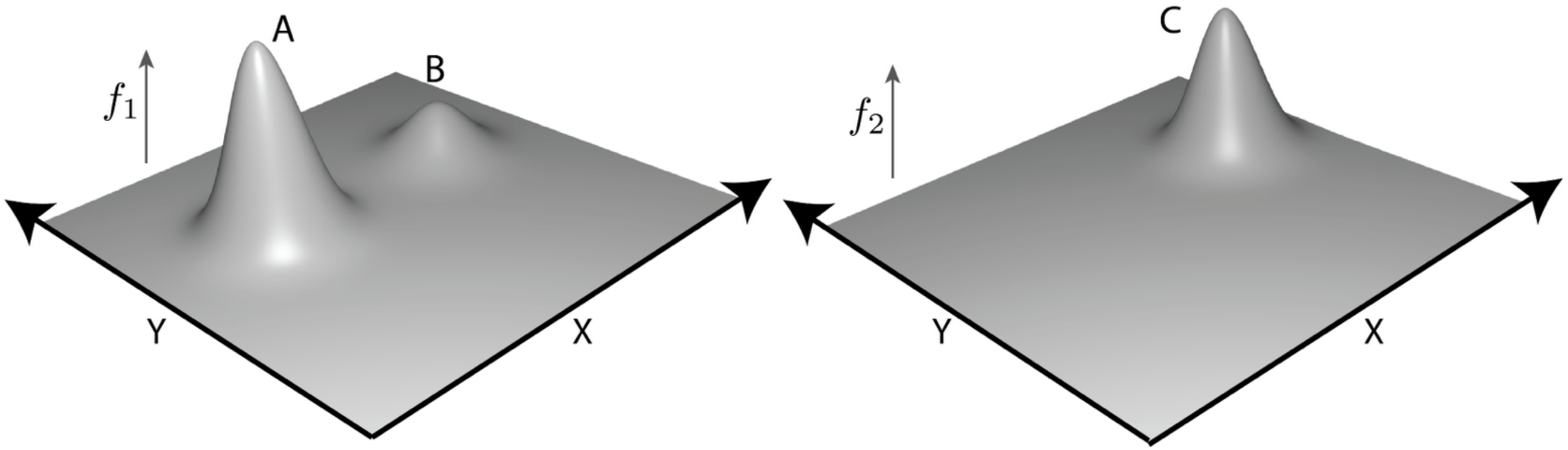


[Beketayev et al. 2014]



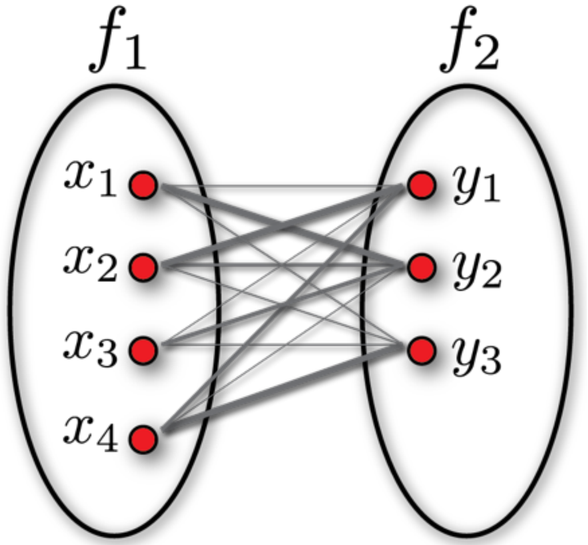
[Narayanan et al. 2015]

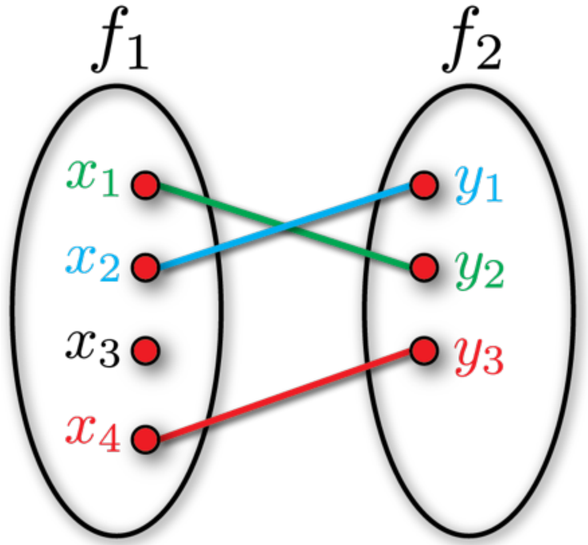
Motivation



Is there a way to compute a locality-aware similarity measure?

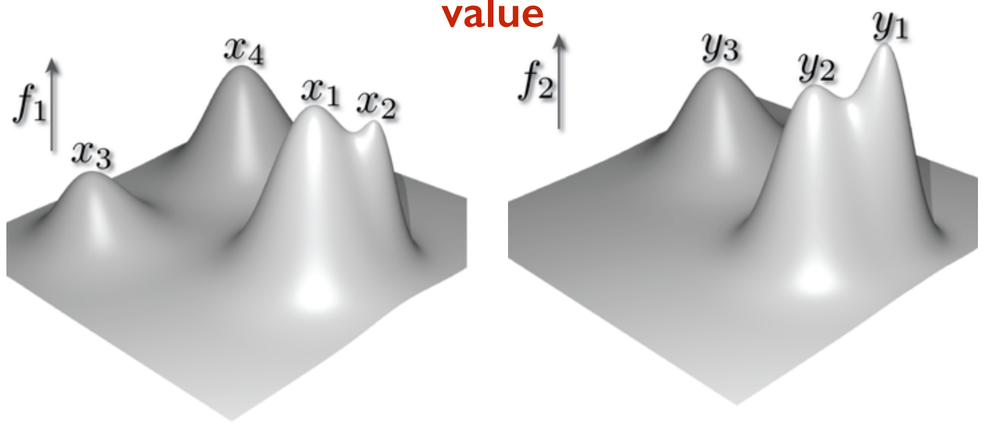
Maximum Topology Matching

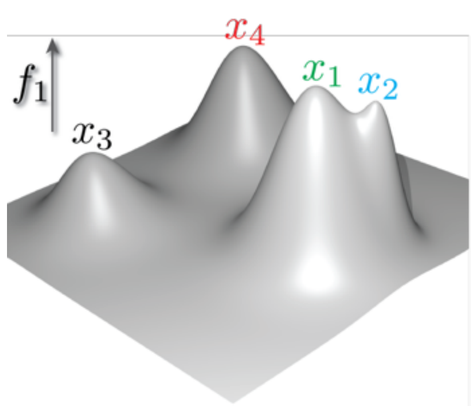


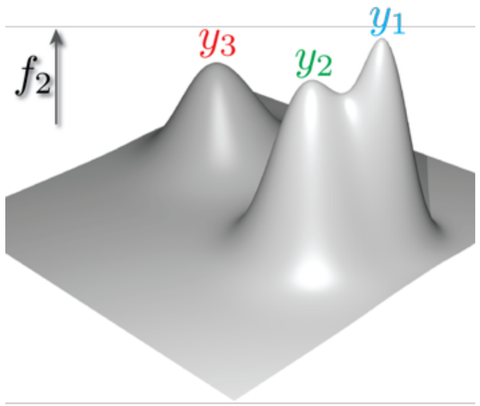


$$w_{a,b} = (1 - \delta_{a,b}) e^{\frac{d_g(a,b)}{r^2}}$$

function value — locality

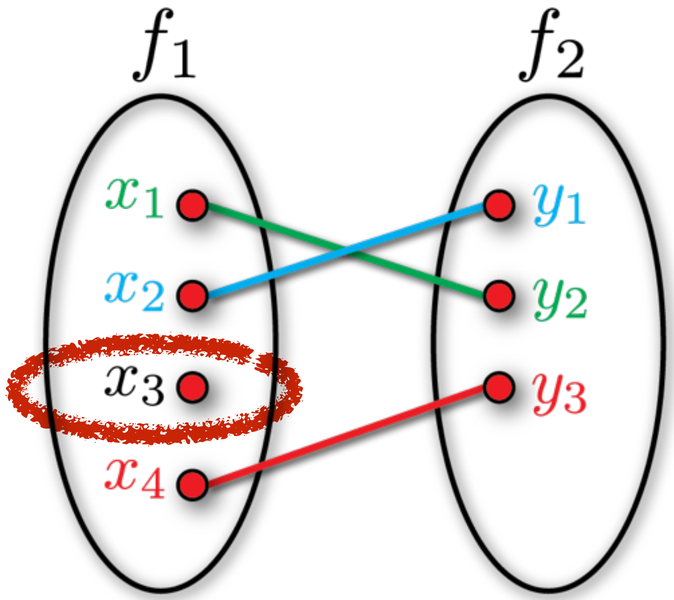


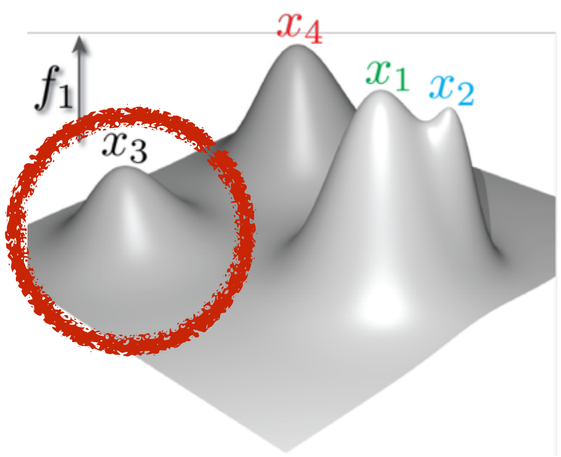


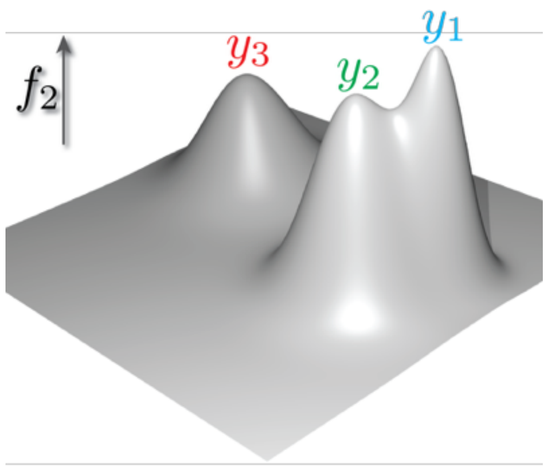


Topological Similarity

Intuition: It is the minimum simplification required to obtain a perfect matching between two functions

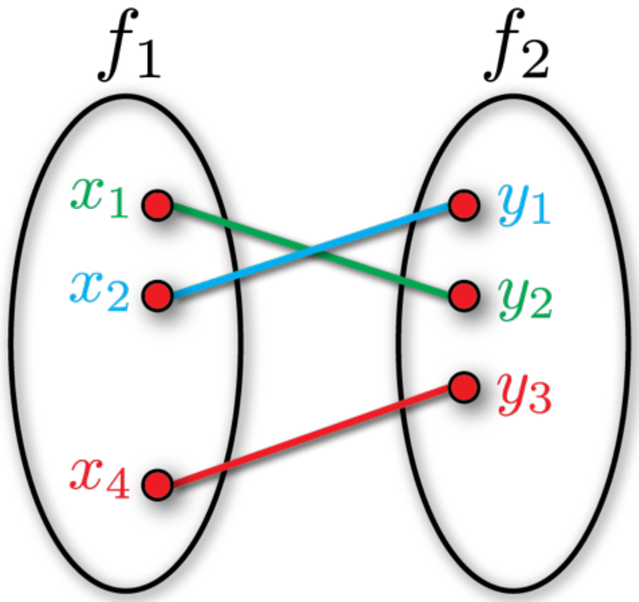


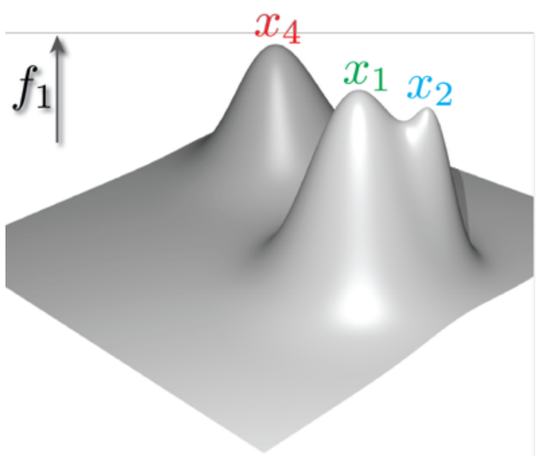


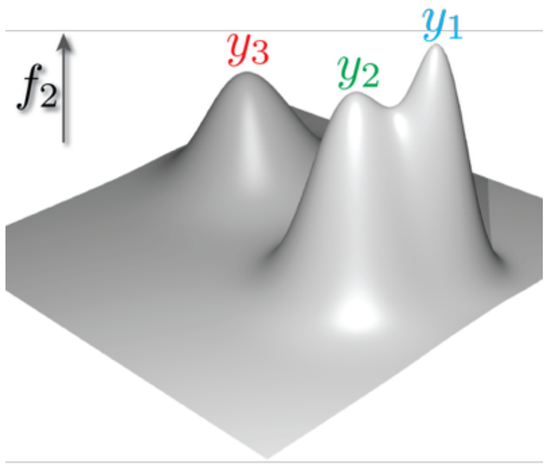


Topological Similarity

Intuition: It is the minimum simplification required to obtain a perfect matching between two functions

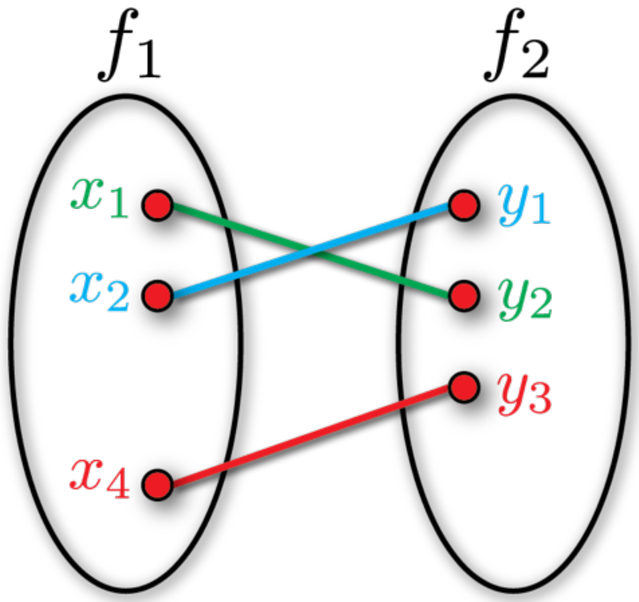


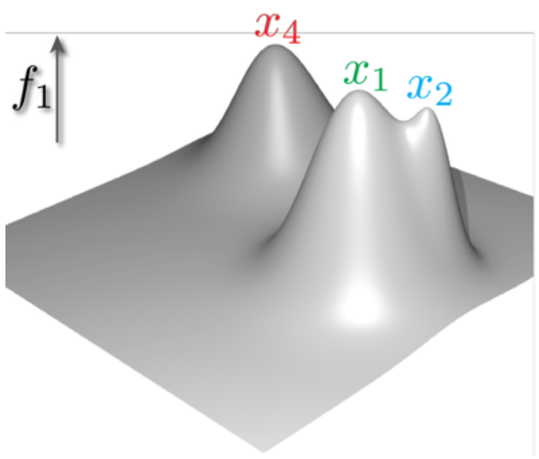


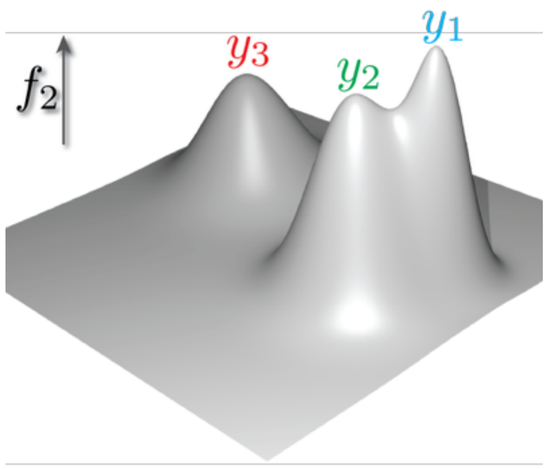


Functional Similarity

Intuition: Measures the amount of change required to get **identical functions**

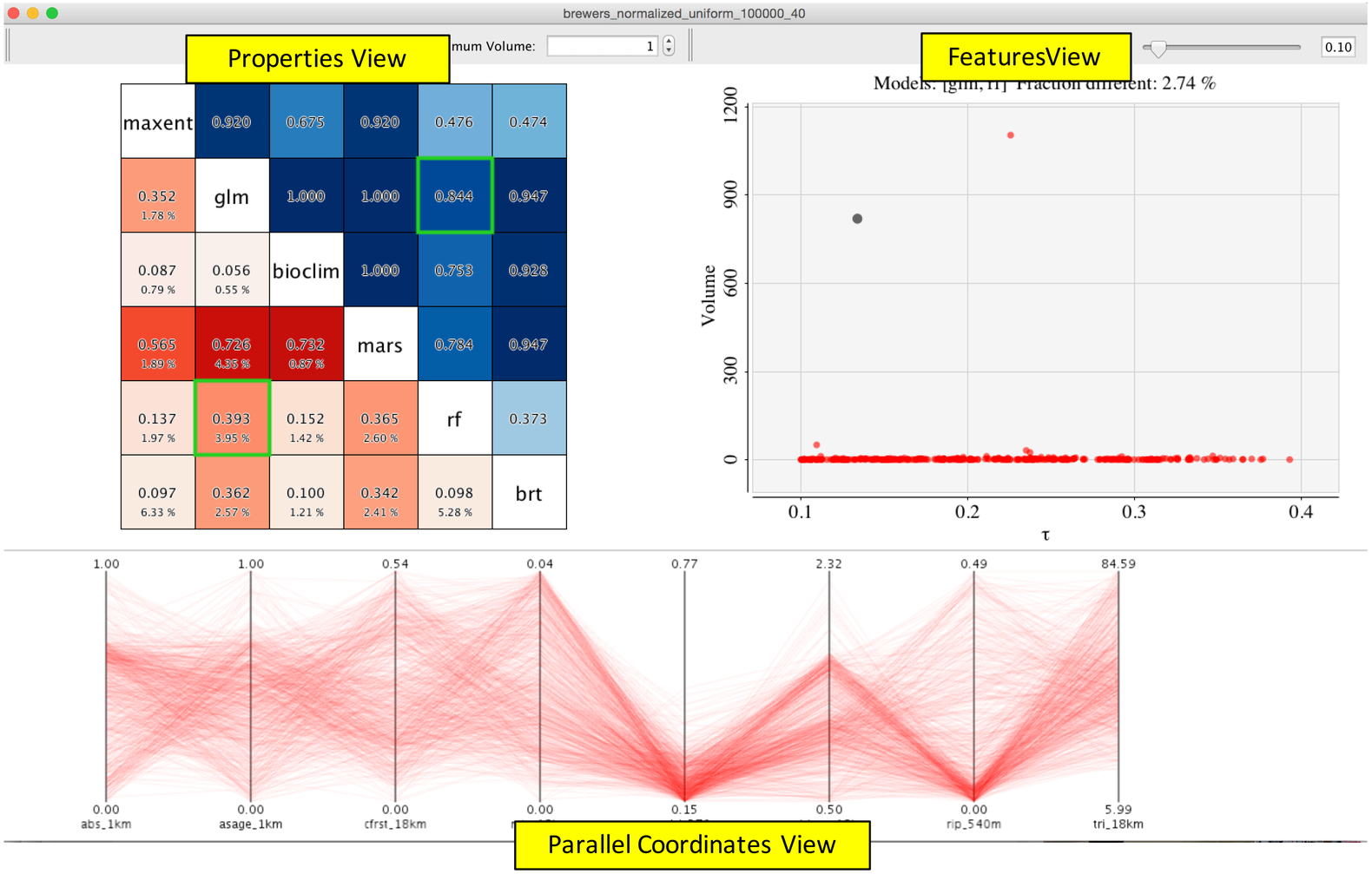




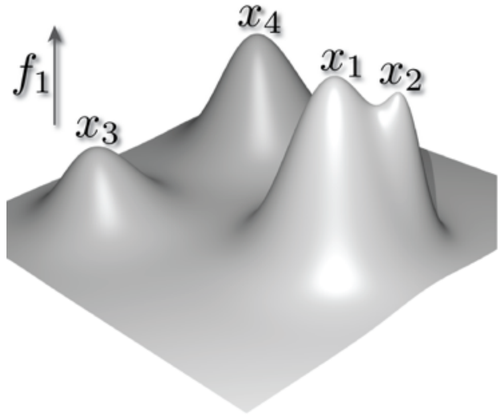


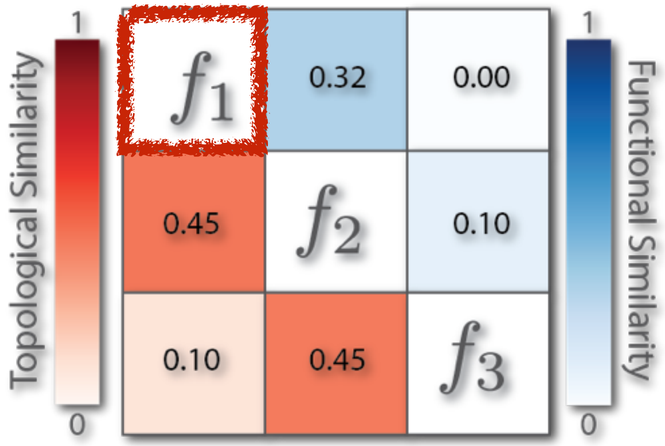
Exploration Framework

Exploration Framework

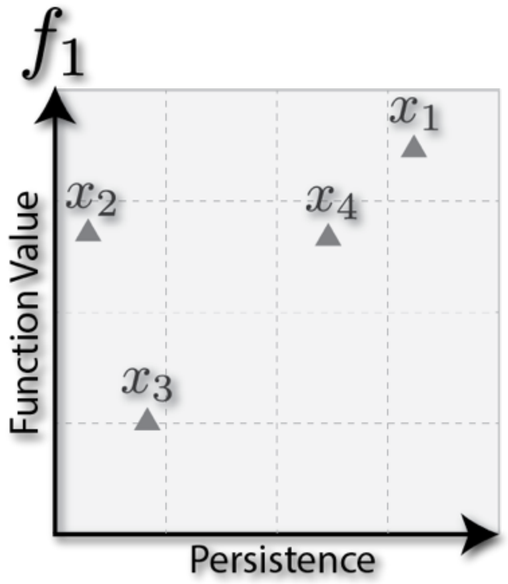


Exploring a SDM

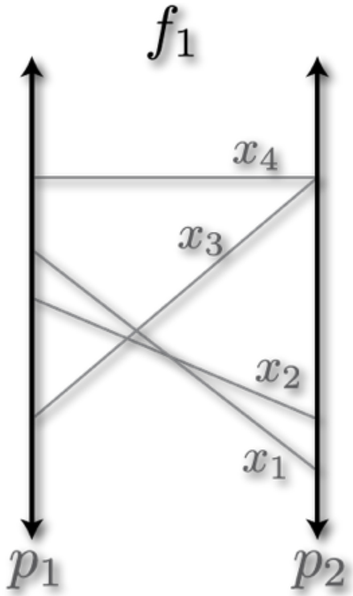




Properties View

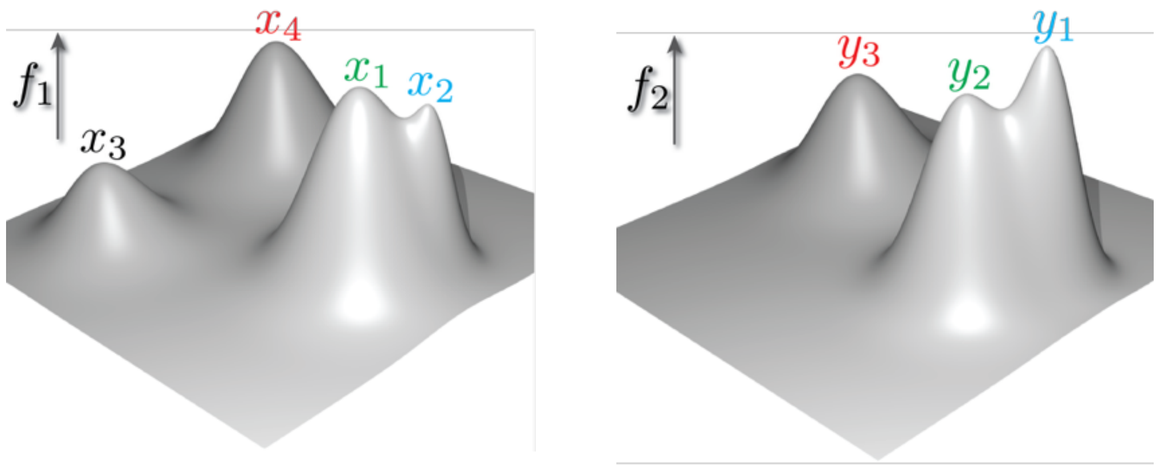


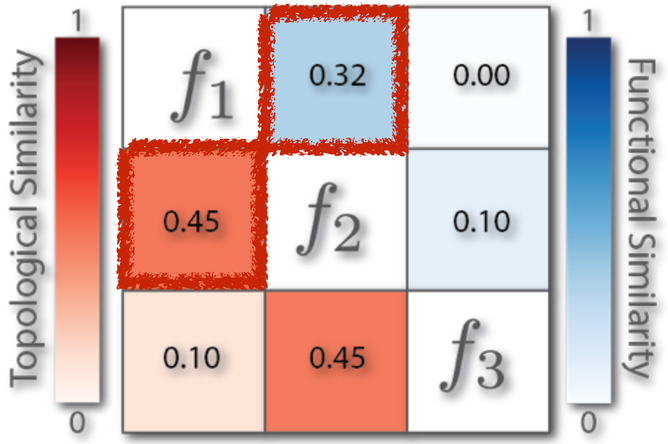
Features View



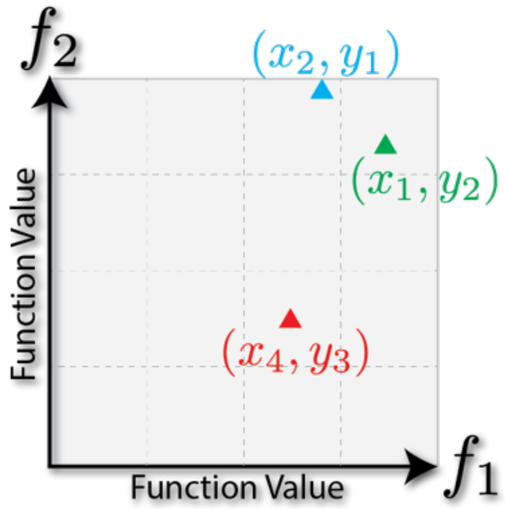
Parallel Coordinates View

Similarities Between Models

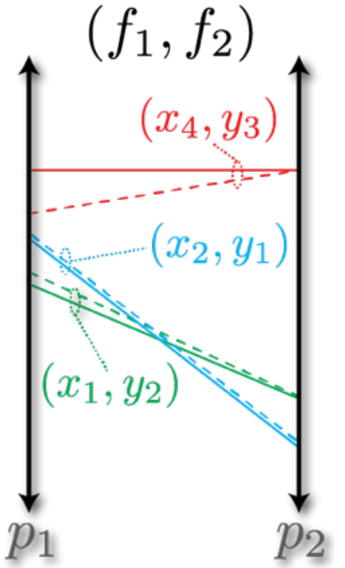




Properties View

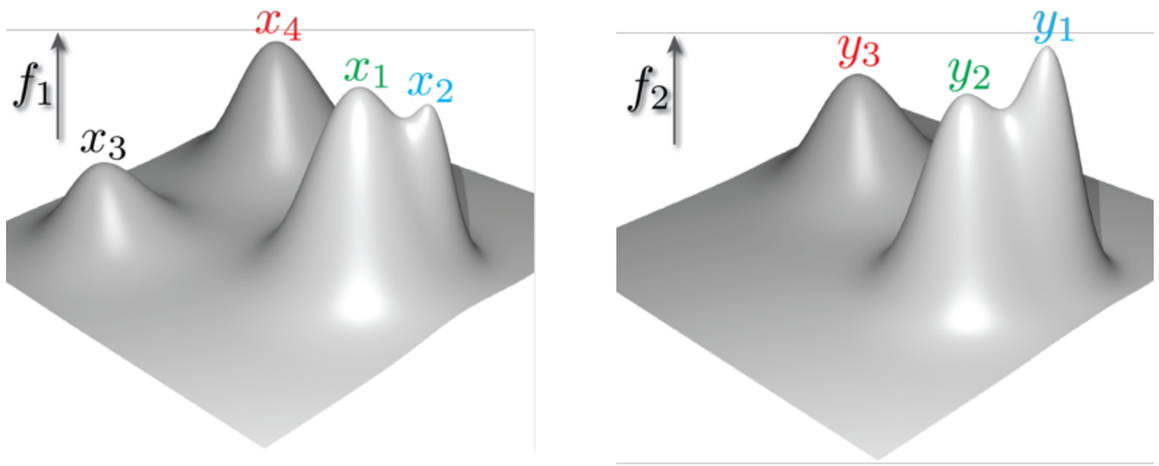


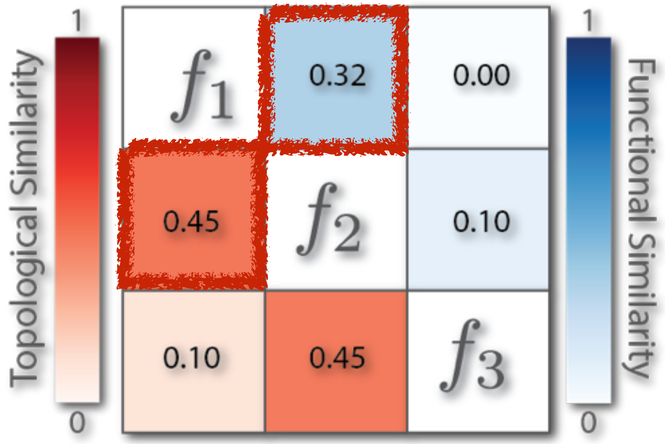
Features View



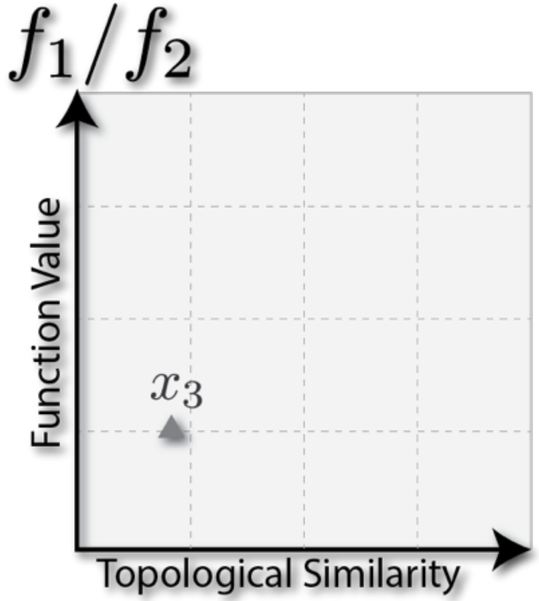
Parallel Coordinates View

Differences Between Models

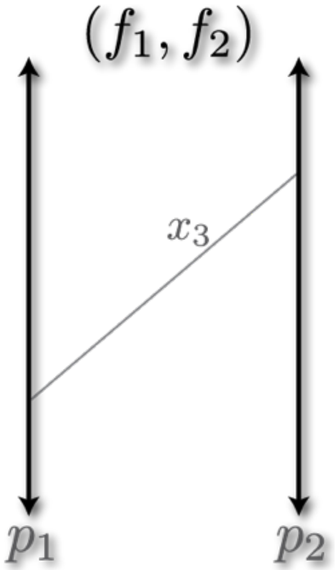




Properties View



Features View



Parallel Coordinates View

Case Studies

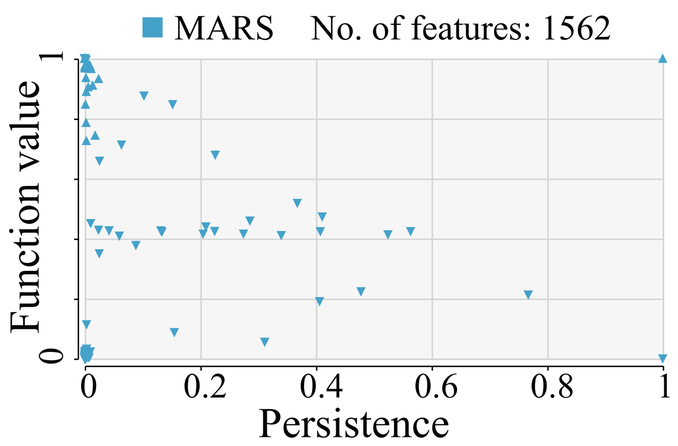
Data Sets

- Brewer's Sparrow
  - 8 predictors
  - 5 models
- Sagebrush
  - 8 predictors
  - 5 models
- Spruce Fir
  - 9 predictors
  - 5 models

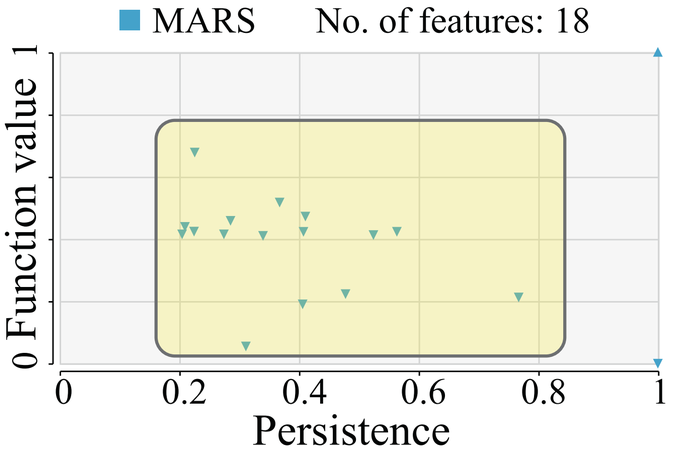
Implementation

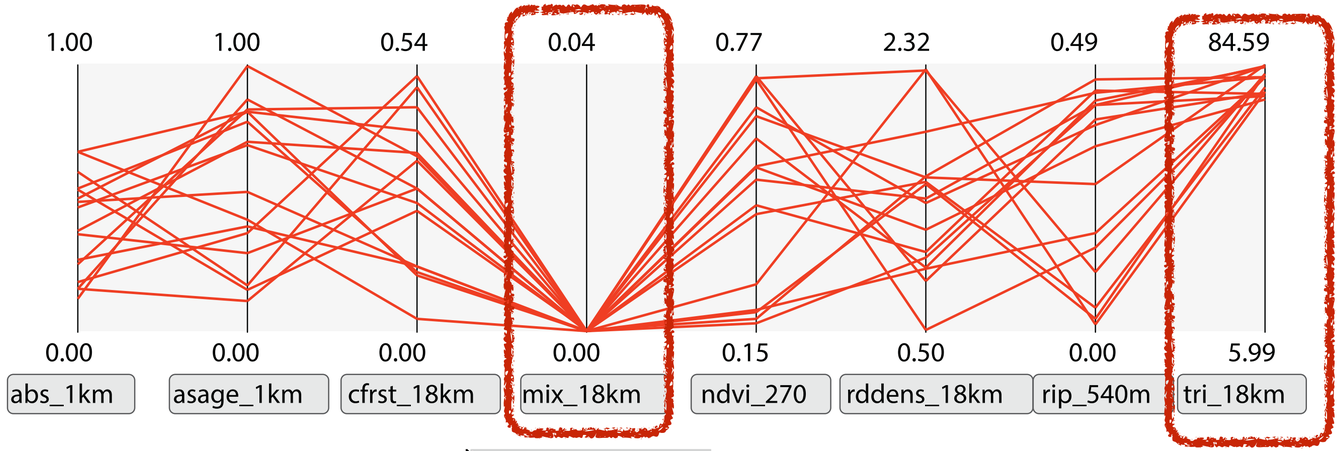
- Domain of SDM
  - Discretize as a graph
  - Sample n=10^5 points in the high-dimensional space
    - Latin Hypercube Sampling
  - Compute the k-nearest-neighbor graph (k=40) for each SDM
- Scalar function
  - Function values are computed using the SDM on the vertices of the Graph
  - Linearly interpolated within each edge
- Cut-off radius for computing edge weights r = 0.1

Exploring a SDM

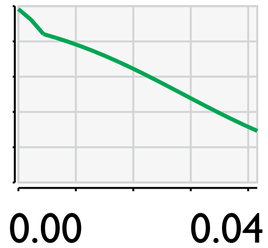


Simplification

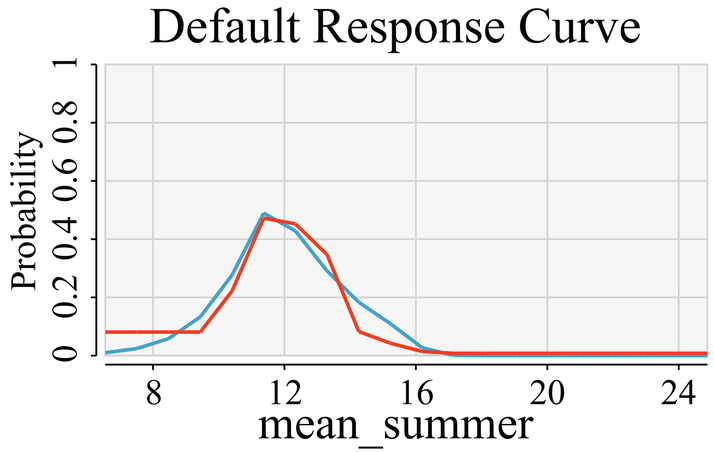


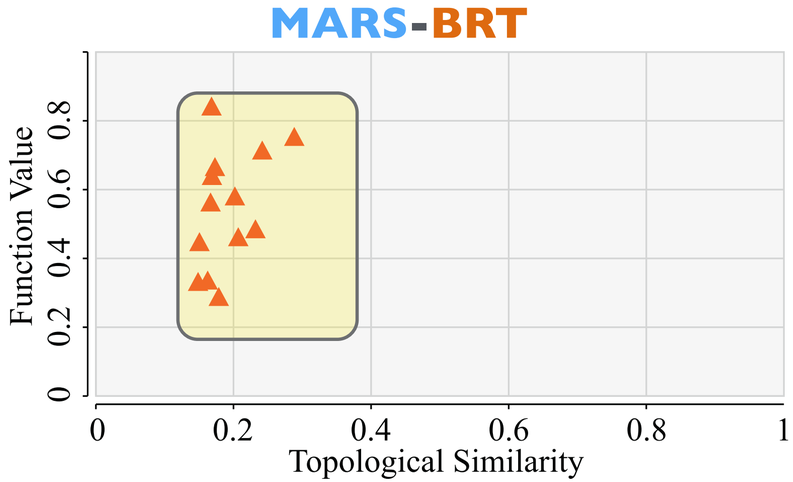


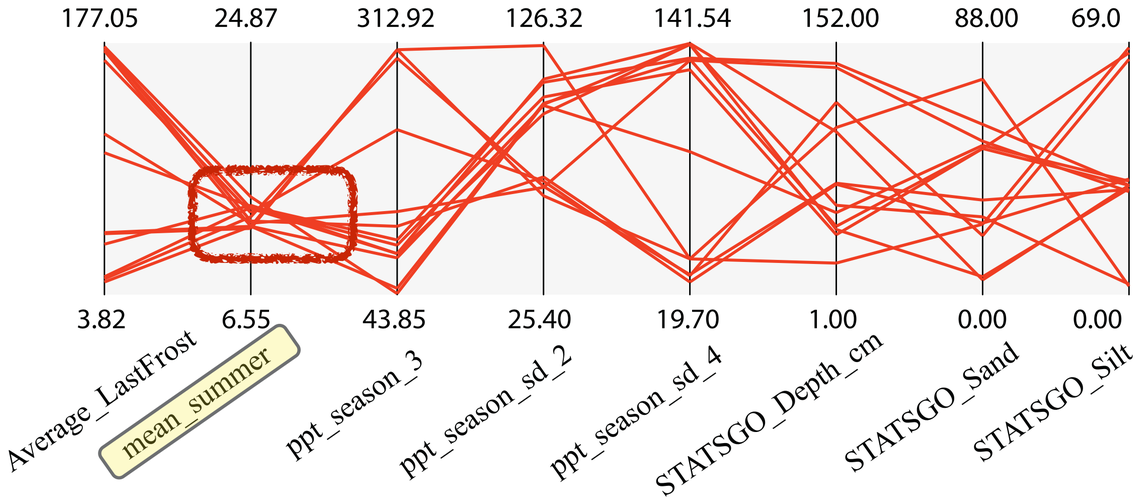
default response curve

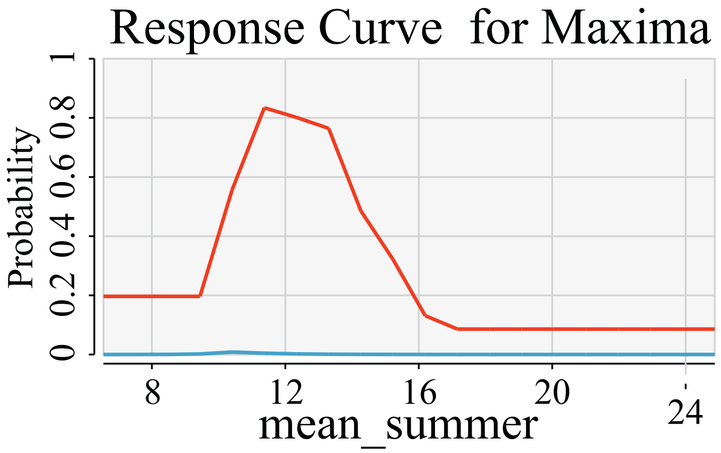


Exploring Differences: MARS vs BRT









Conclusions

- Positive feedback from Ecologists
  - Surprised by results
  - Integrating into SAHM package for VisTrails
- Other Contributions
  - Robustness to noise
    - perturb function values
    - perturb extrema locations
  - Experimentally evaluate effect of parameters to the similarity measures
    - sample size
    - # neighbors
    - neighborhood radius r

Future Work

- Each dimension normalized between 0 and 1
  - Can they be standardized instead?
- Neighborhood radius fixed to 0.1 based on discussions with our collaborators
  - Can a different weighting scheme be used irrespective of the domain?
- Use other metaphors to visualize SDMs
  - Eg. topological spines

Acknowledgements

This work was supported in part by

- Google Faculty Award
- IBM Faculty Award
- Moore-Sloan Data Science Environment at NYU
- NYU School of Engineering
- NYU's Center for Urban Science and Progress (CUSP)
- AT&T
- NSF award CNS-1229185
- DOE
- NASA Biodiversity Program award NNH11AS091

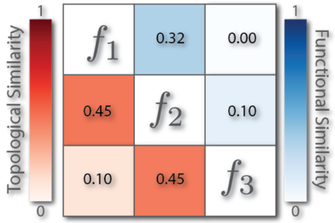
Using Maximum Topology to Explore Differences in Species Distribution Models

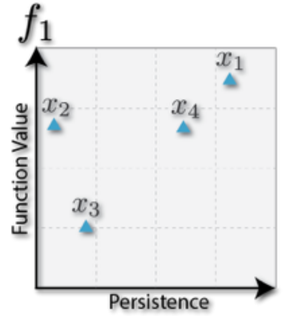
Thank You!

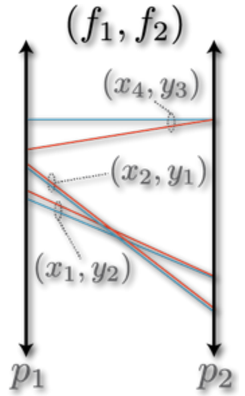
More Information:

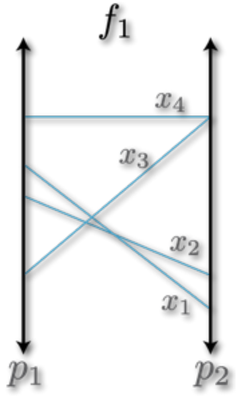
Jorge Poco
jpocom@uw.edu

Harish Doraiswamy
harishd@nyu.edu





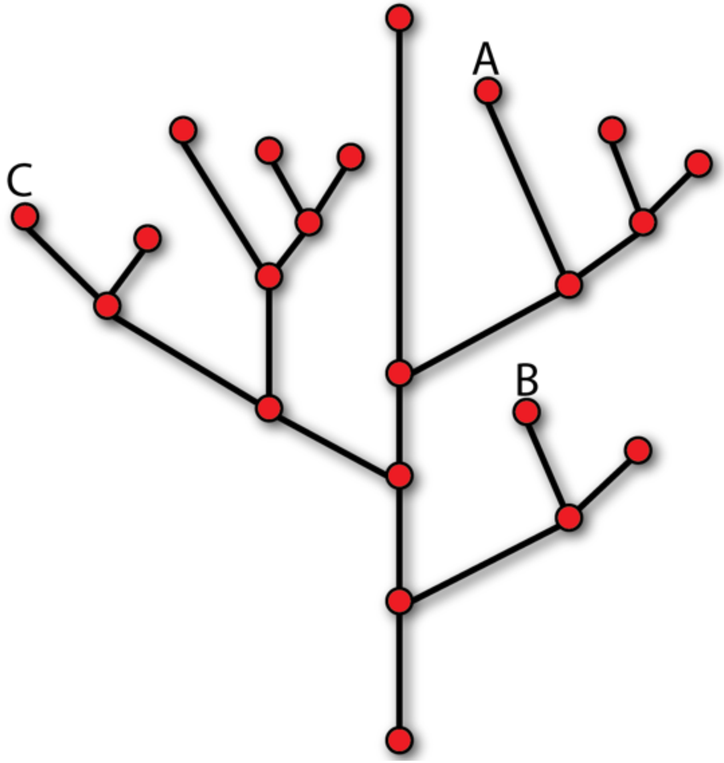


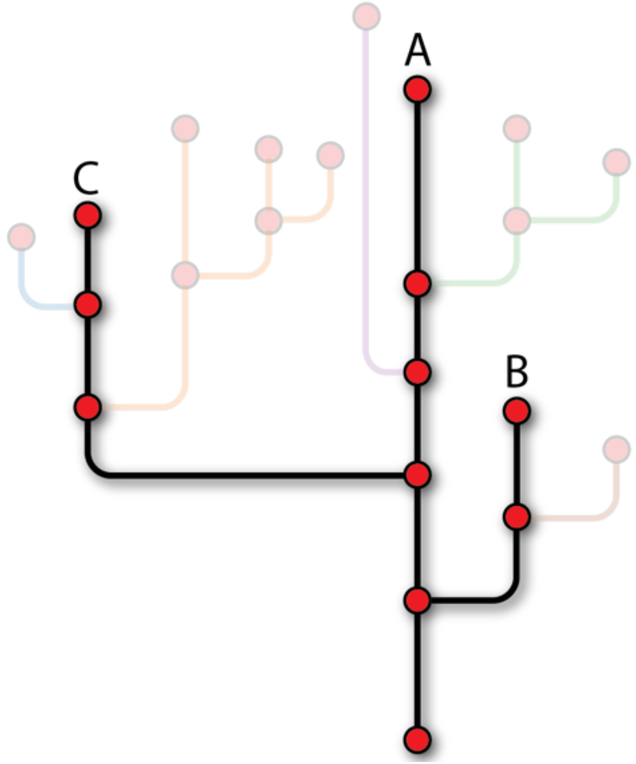


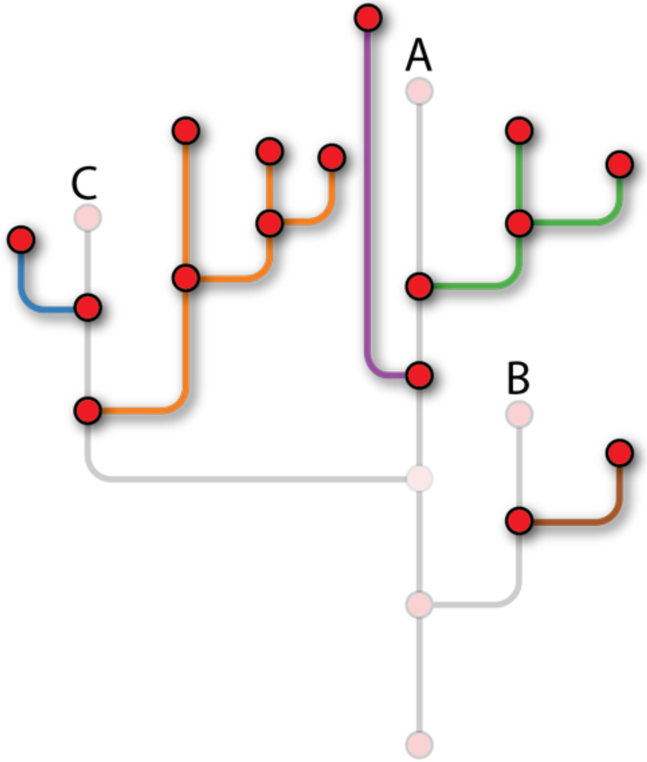
NYU TANDON SCHOOL OF ENGINEERING

USGS science for a changing world

Similarity Measures: Topological Similarity







$$\tau = \max(\tau_1, \tau_2)$$

- Intuition: It is the minimum simplification required to obtain a perfect matching between two functions

Effect to Noise