VISUALIZATION FOR UNDERSTANDING REGRESSION MODELS

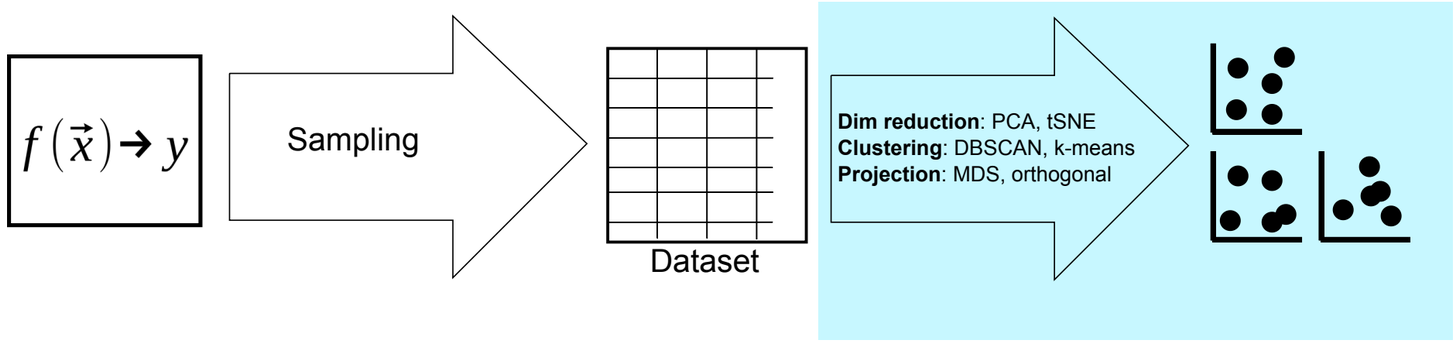
Thomas Torsney-Weir

8 September, 2020

Swansea University
Prifysgol Abertawe

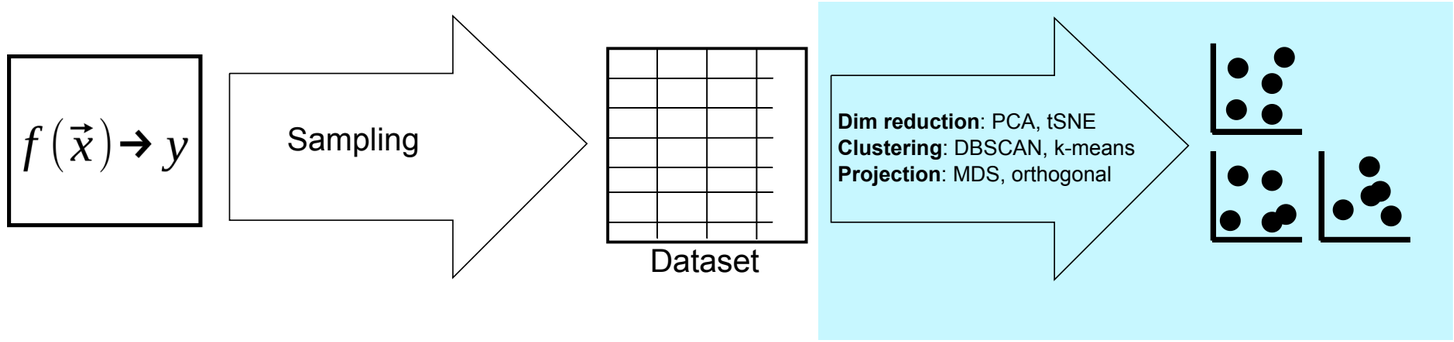
VISUALIZATION PIPELINES

Discrete

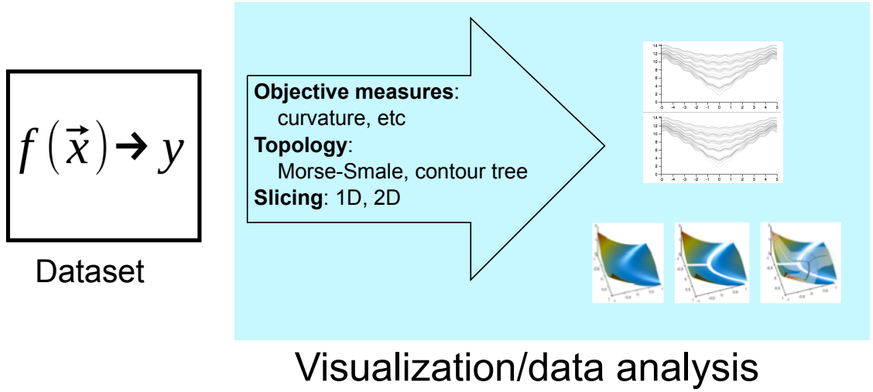


VISUALIZATION PIPELINES

Discrete



Continuous



AGENDA

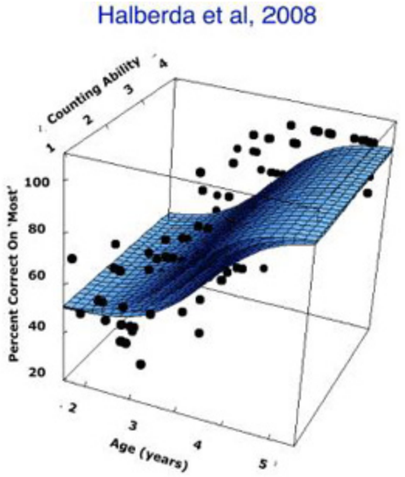
Benefits of treating a regression model itself as the “dataset” for visual data analysis

- What are regression algorithms?
- Overview of slicing
- Advantages of regression model as dataset

WHAT ARE REGRESSION MODELS?

"predict the value of one or more *continuous* target variables given the value of a D-dimensional vector of input variables" (Bishop 2006)

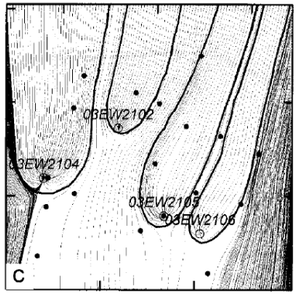
- Important bits:
  - Take a number of factors as input (often continuous)
  - Output is a scalar
  - Inputs are often meaningful
  - Conceptually a multi-dimensional surface (manifold)



key issue: how do we understand this multi-dimensional surface?

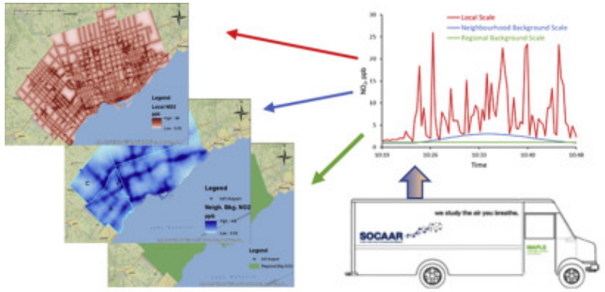
APPLICATION AREAS

Geostatistics



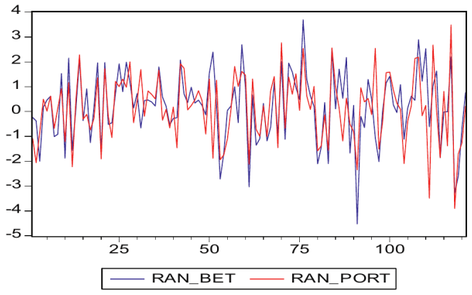
(Tonkin and Larson 2002)

Urban studies



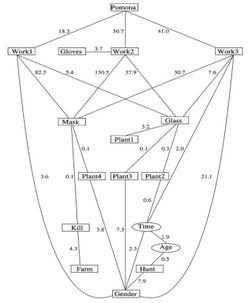
(Shairsingh et al. 2019)

Finance



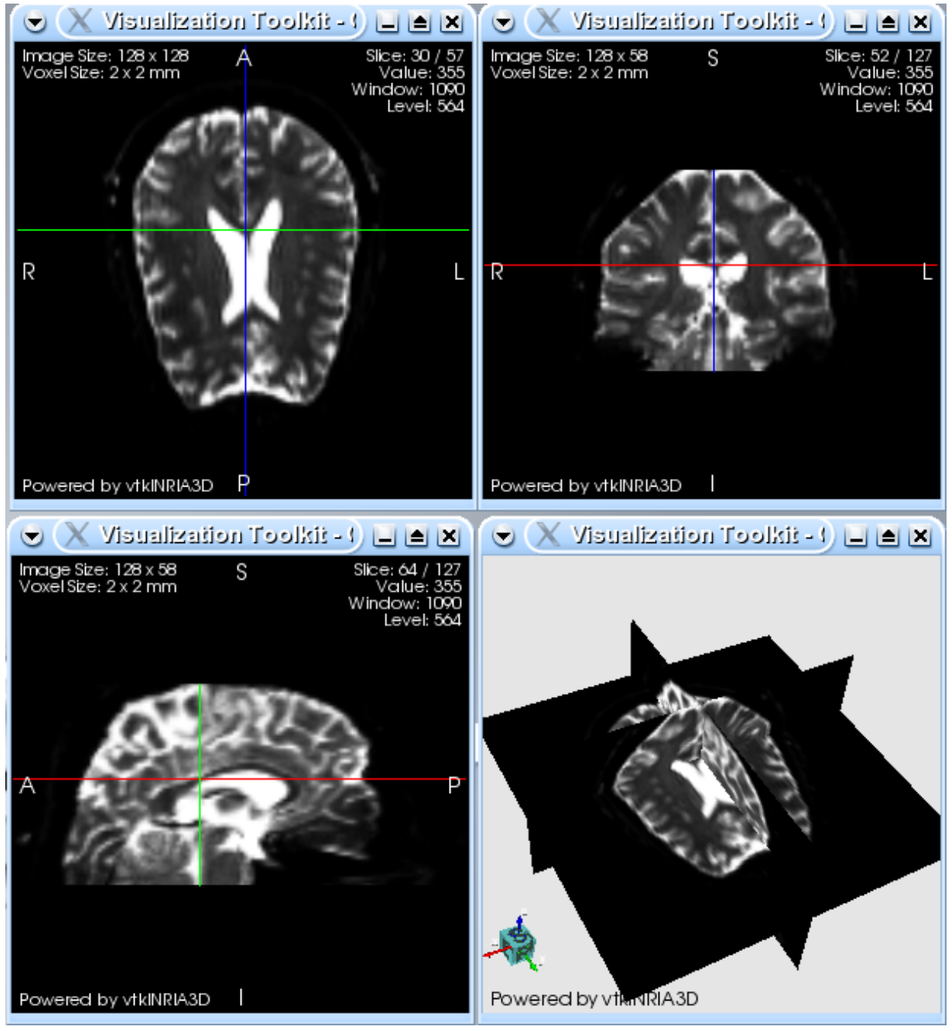
(Anghelache and Anghel 2014)

Epidemiology



(Pittavino et al. 2017)

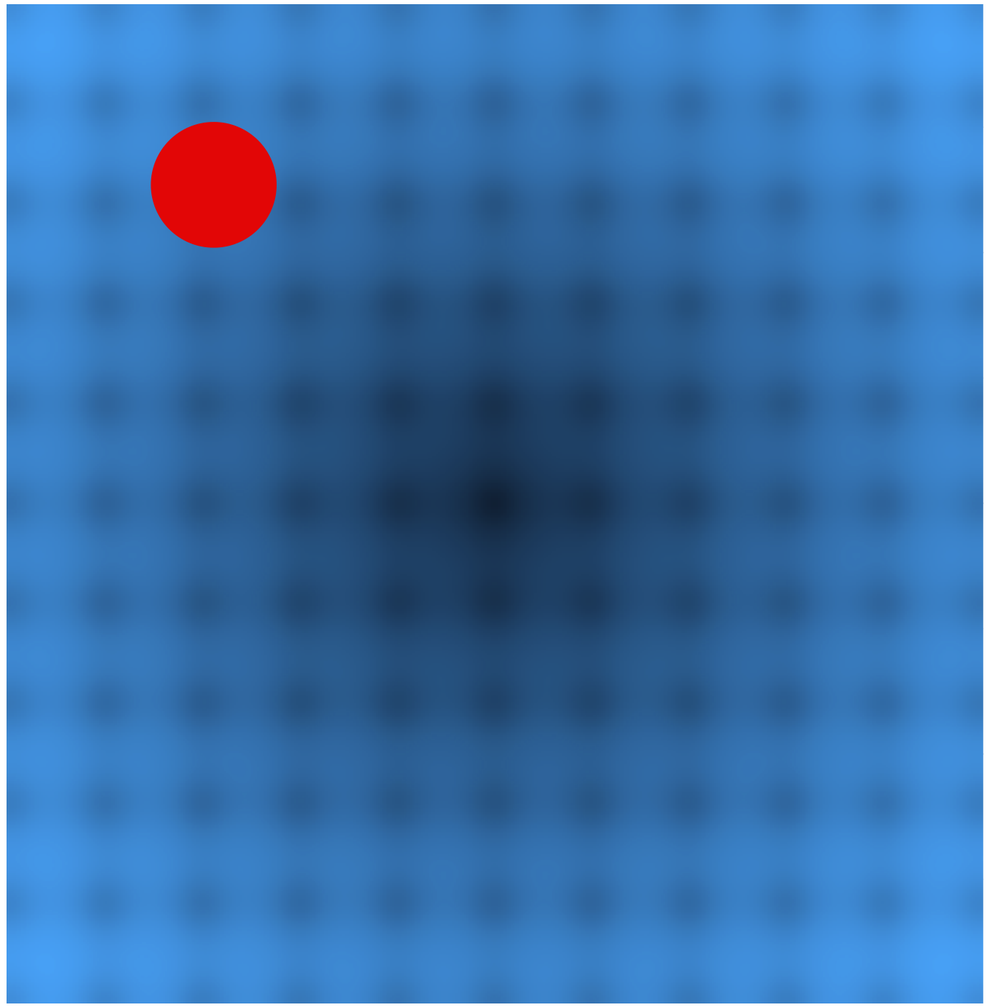
SLICING



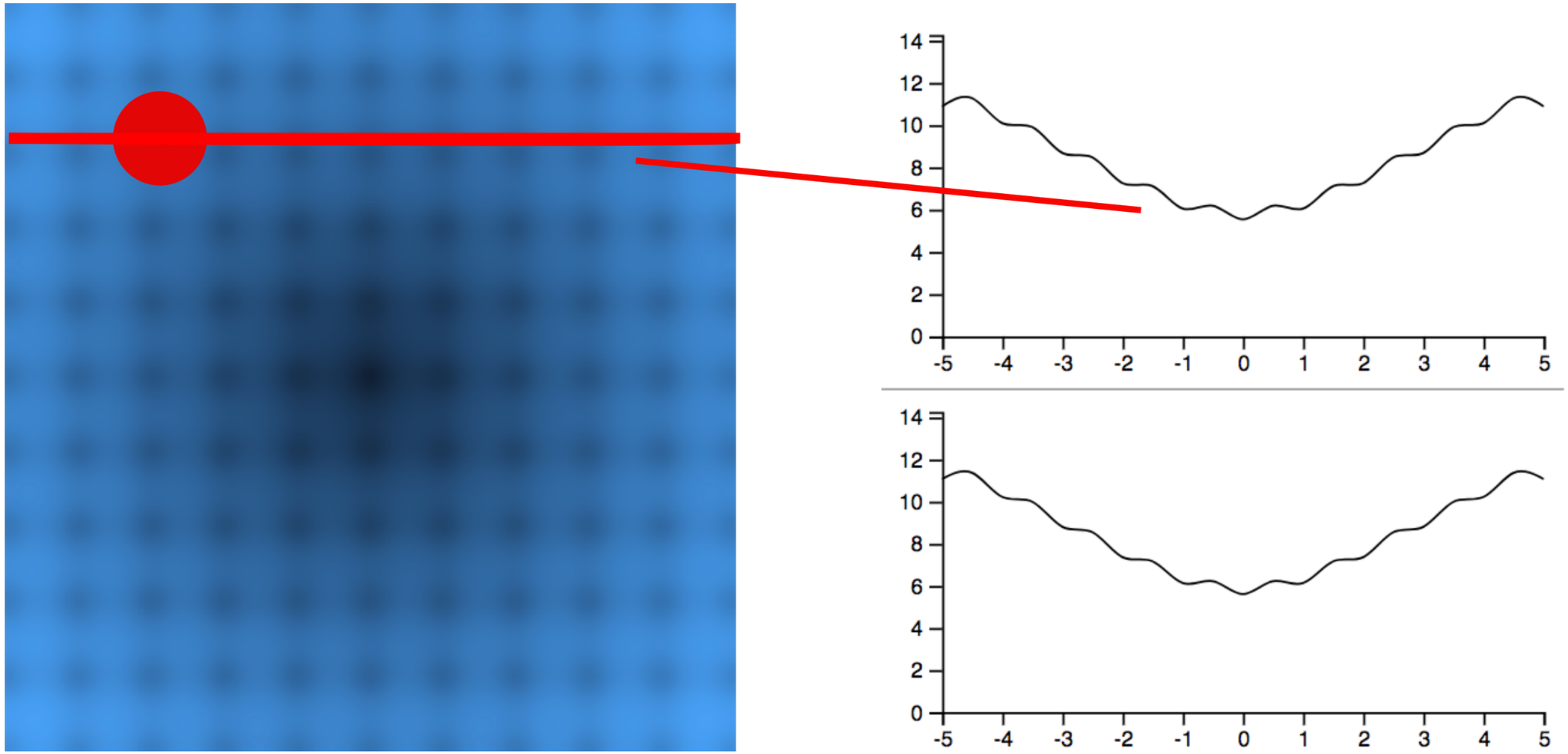
- **Pros**
  - Reduces dimensionality
  - Easy to understand metaphor
- **Cons**
  - Focus point selection important

MULTI-D SLICING

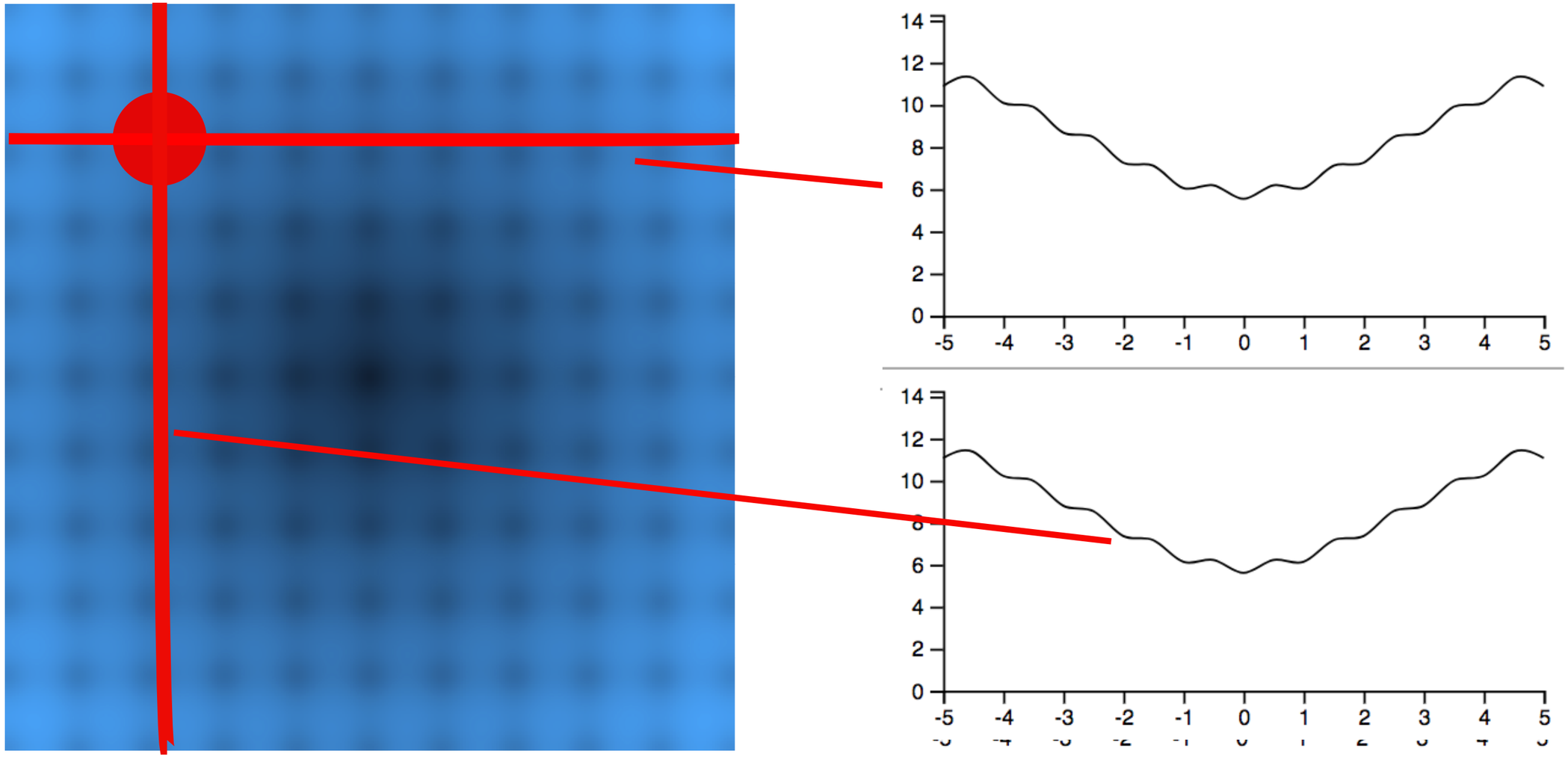
MULTI-D SLICING



MULTI-D SLICING

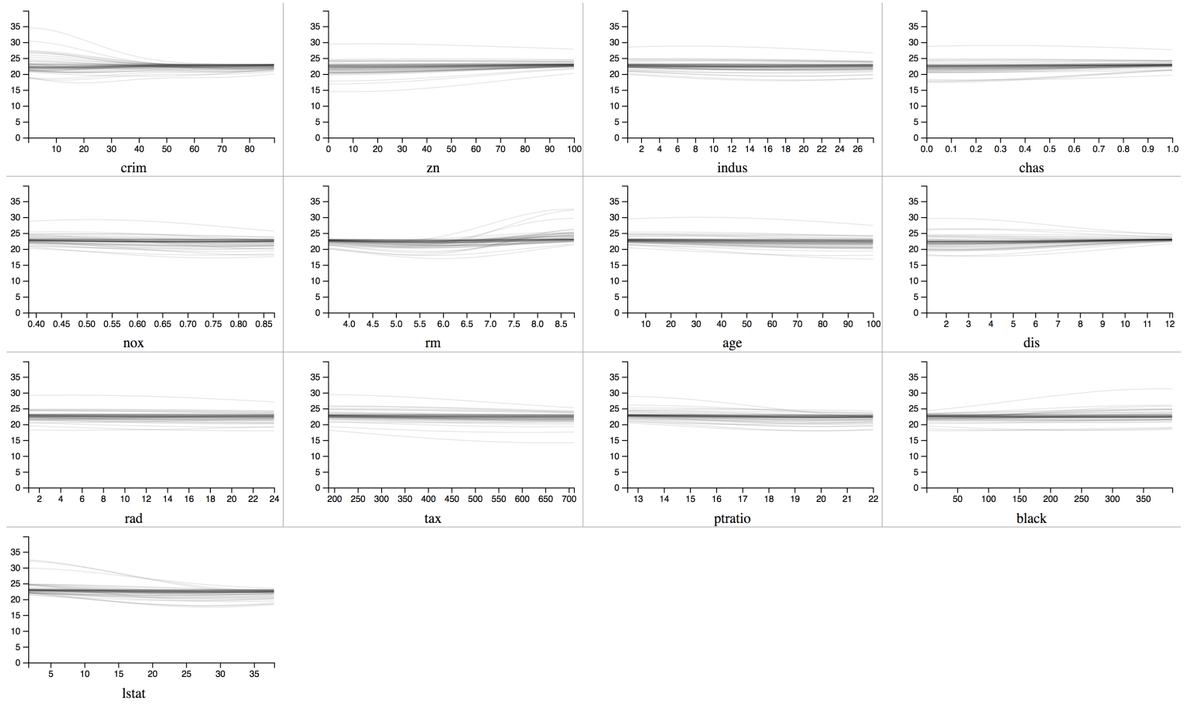


MULTI-D SLICING



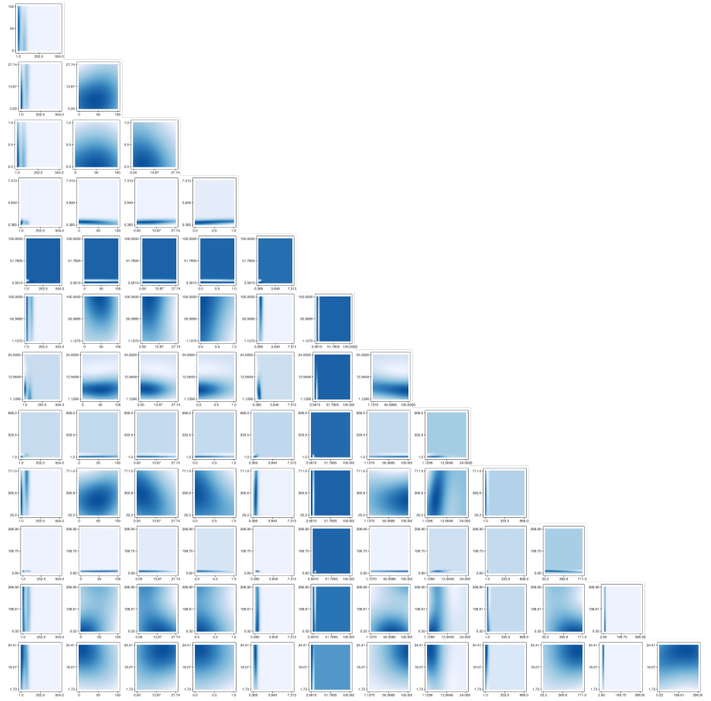
SLICING

1D



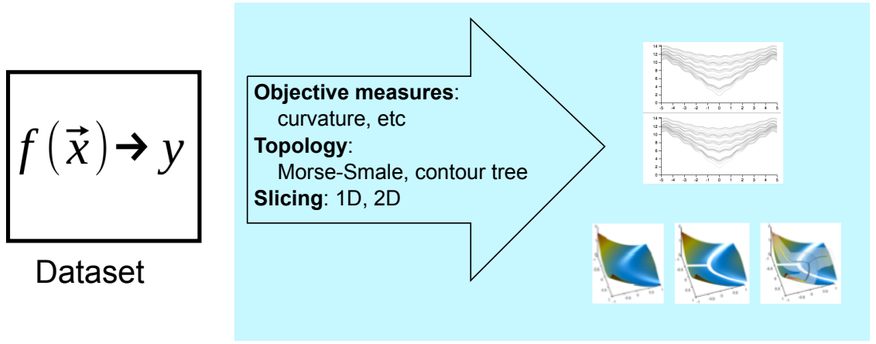
Sliceplorer (Torsney-Weir, Sedlmair, and Möller 2017)

2D



Hyperslice (Wijk and Liere 1993)

BENEFITS OF PIPELINE

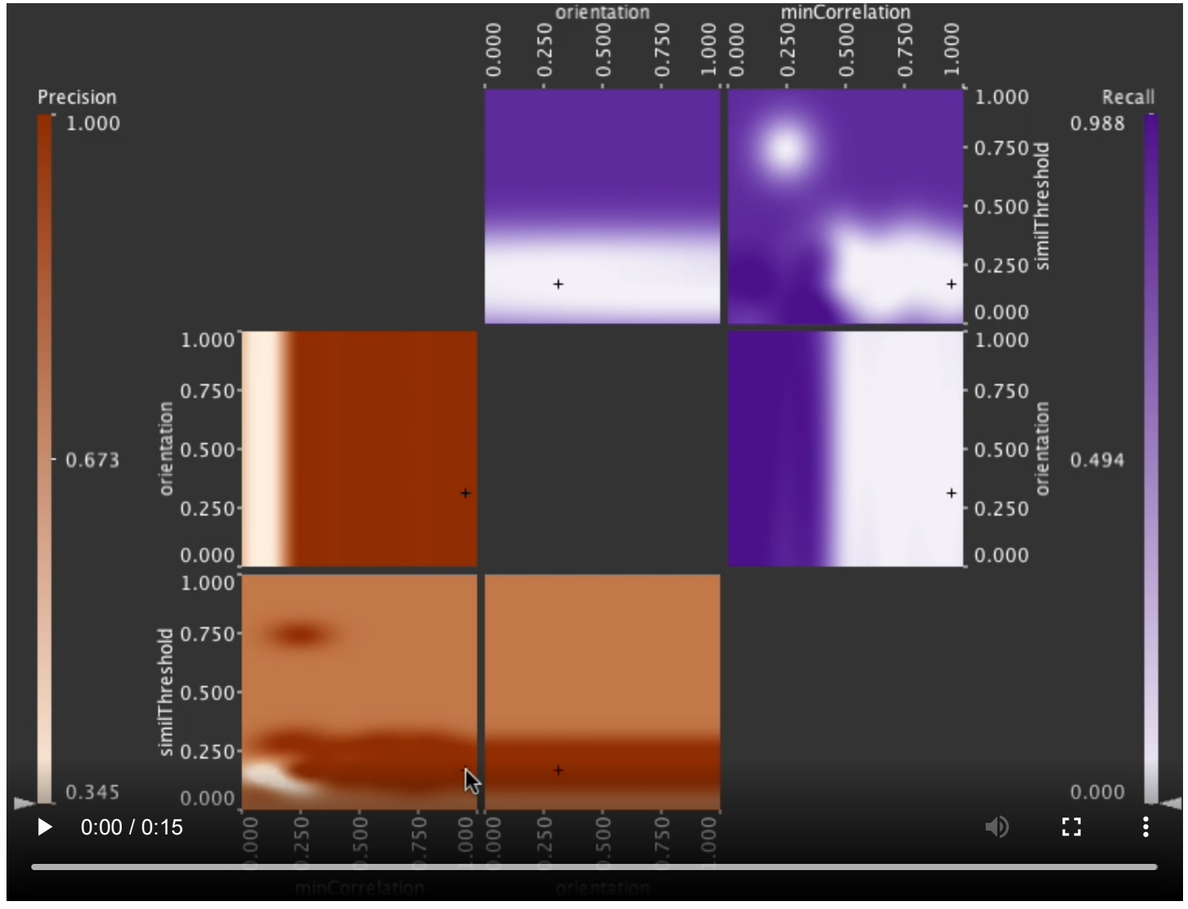


- Fast rendering
- Focus point selection

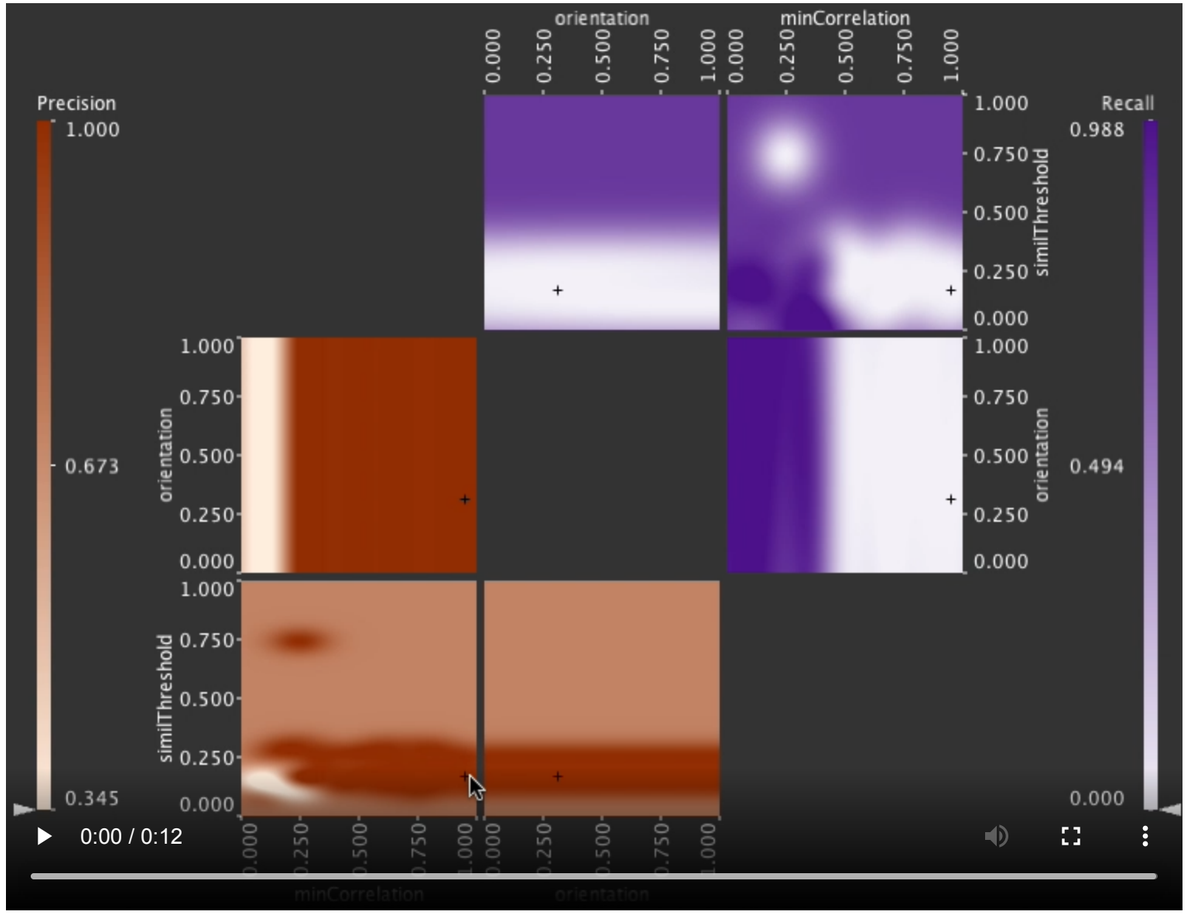
FAST RENDERING

Because we know the details of the model, the visualization system can execute the regression model at any point

3 fps

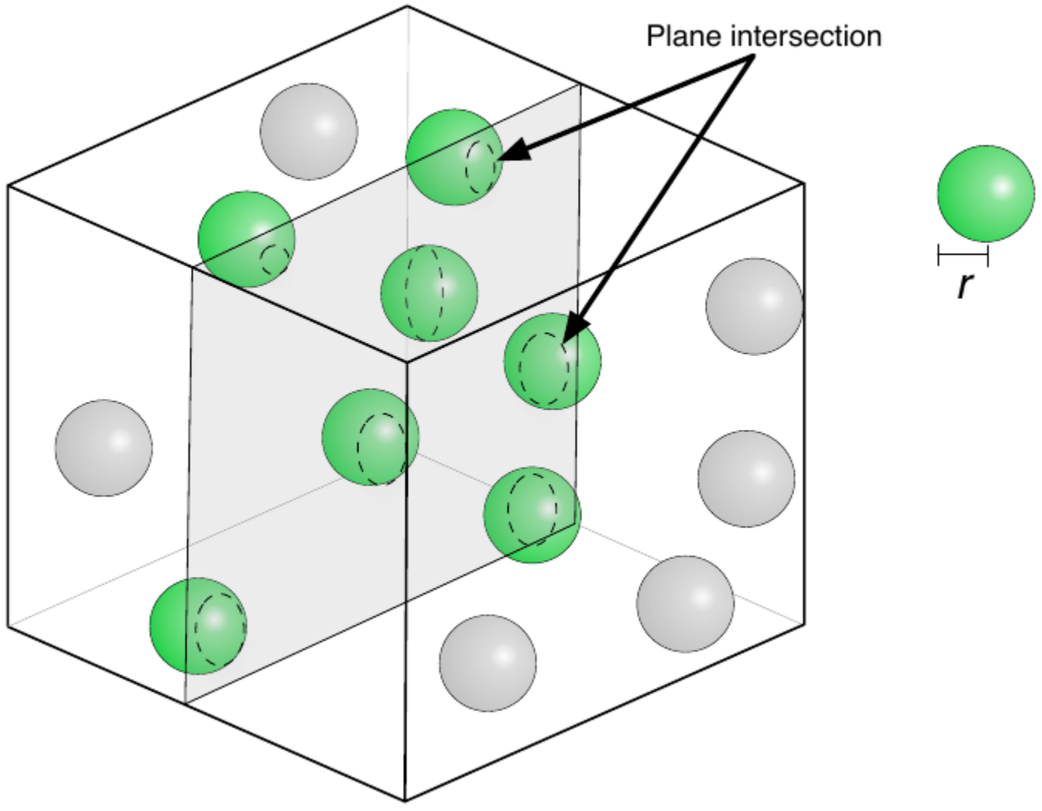


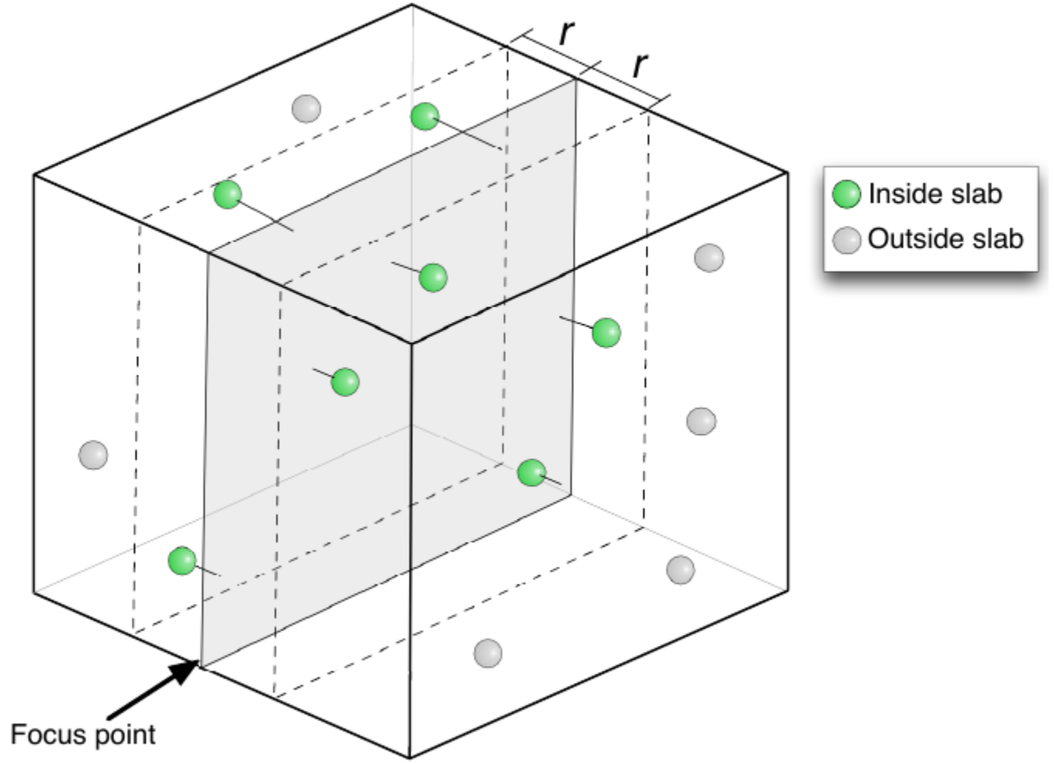
30 fps



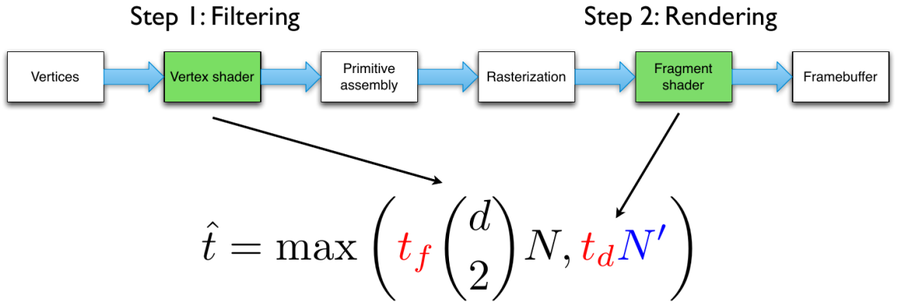
FAST RENDERING

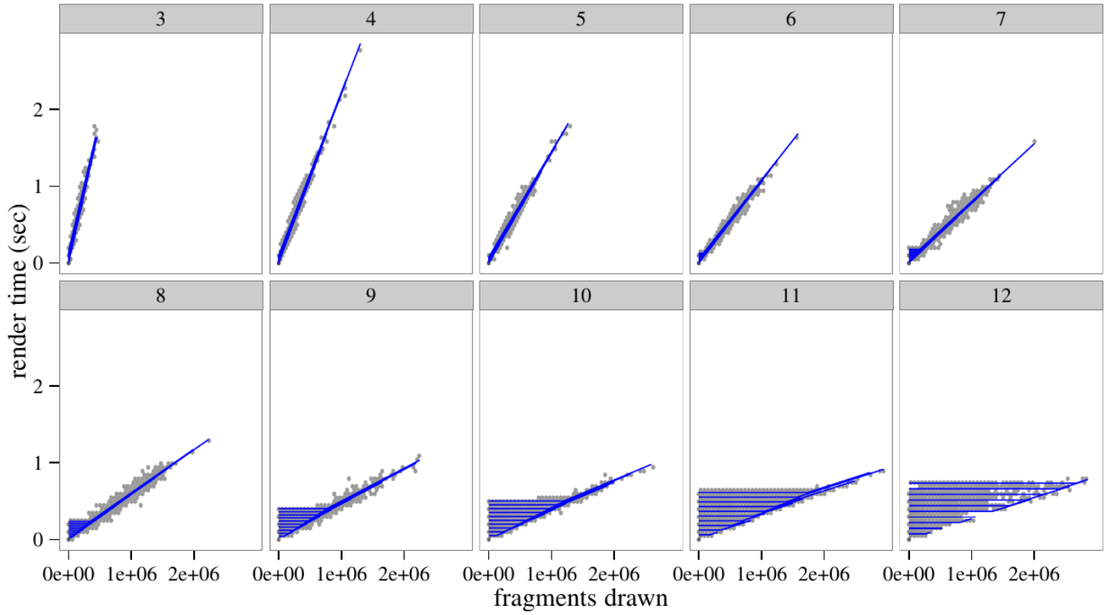
Knowing we were using Gaussian process models, we could analyze the geometry of the scene to figure out how to make it run faster





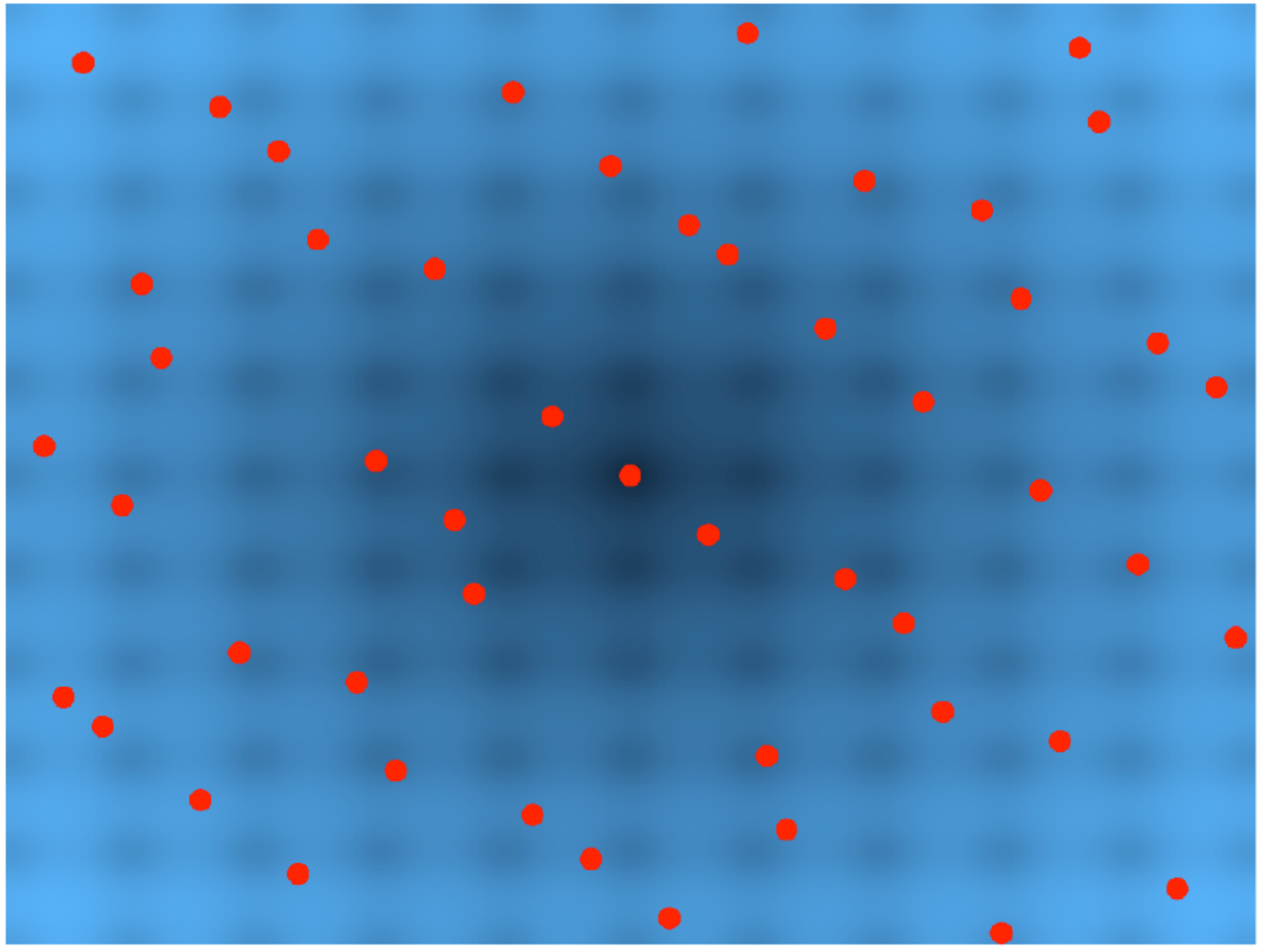
FAST RENDERING

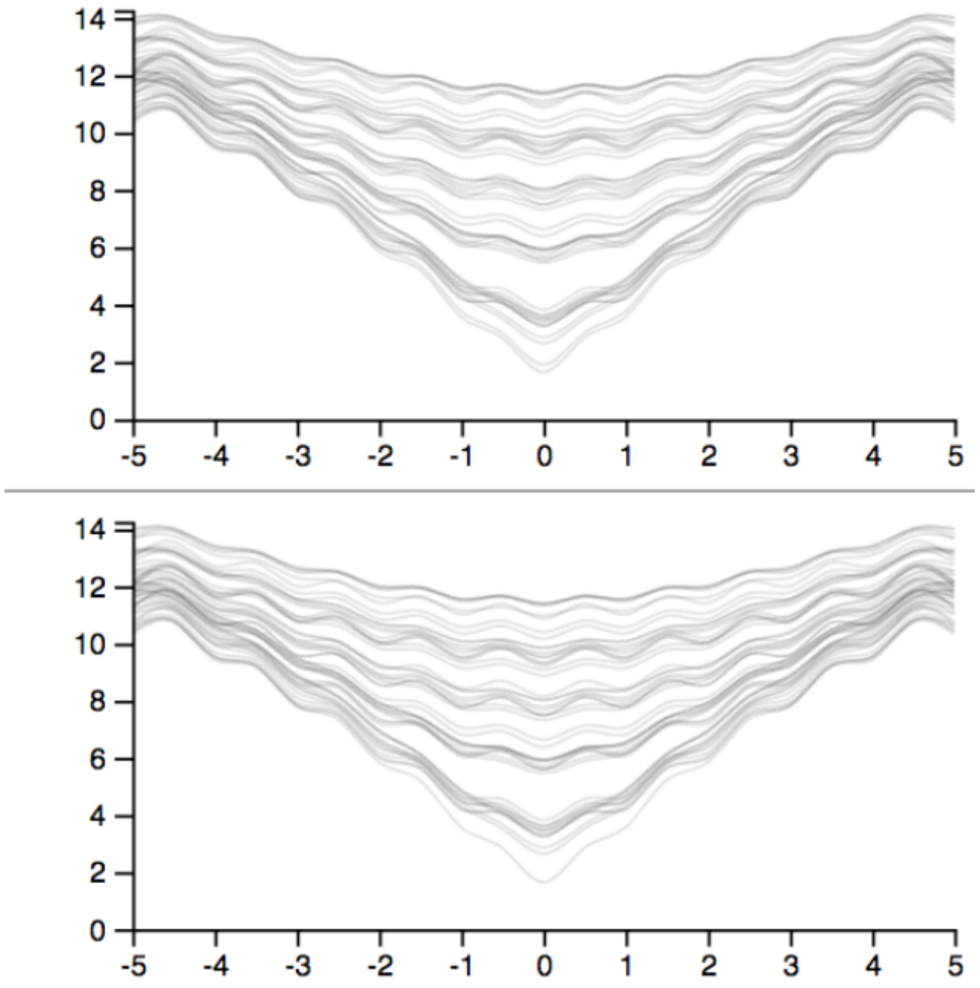




(Torsney-Weir et al. 2017)

FOCUS POINT SELECTION





Focus point sampling using space-filling design (Torsney-Weir, Sedlmair, and Möller 2017)

CONCLUSION

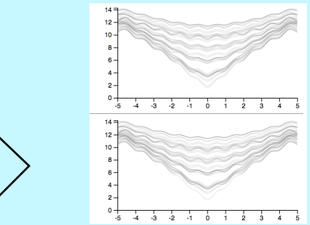
Treating the regression model itself as the datatype allows us to analyze the model more efficiently.

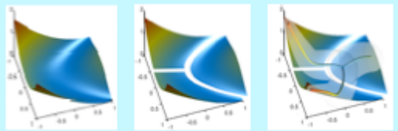
- Slice-based visualization
- Efficient rendering
- Control of sampling

$$f(\vec{x}) \rightarrow y$$

Dataset

Objective measures:
curvature, etc
Topology:
Morse-Smale, contour tree
Slicing: 1D, 2D





Visualization/data analysis

THANKS!

Questions?

t.d.torsney-weir@swansea.ac.uk

Swansea University
Prifysgol Abertawe

REFERENCES

Anghelache, Constantin, and Mădălina Gabriela Anghel. 2014. “Using the Regression Model in the Analysis of Financial Instruments Portfolios.” *Procedia Economics and Finance*, International Conference on Applied Statistics (ICAS) 2013, 10 (January): 324–29. https://doi.org/10.1016/S2212-5671(14)00308-6.

Bishop, Christopher M. 2006. *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Secaucus, NJ, USA: Springer-Verlag New York, Inc.

Pittavino, M., A. Dreyfus, C. Heuer, J. Benschop, P. Wilson, J. Collins-Emerson, P. R. Torgerson, and R. Furrer. 2017. “Comparison Between Generalized Linear Modelling and Additive Bayesian Network; Identification of Factors Associated with the Incidence of Antibodies Against Leptospira Interrogans Sv Pomona in Meat Workers in New Zealand.” *Acta Tropica* 173 (September): 191–99. https://doi.org/10.1016/j.actatropica.2017.04.034.

Shairsingh, Kerolyn K., Cheol-Heon Jeong, Jonathan M. Wang, Jeffrey R. Brook, and Greg J. Evans. 2019. “Urban Land Use Regression Models: Can Temporal Deconvolution of Traffic Pollution Measurements Extend the Urban LUR to Suburban Areas?” *Atmospheric Environment* 196 (January): 143–51. https://doi.org/10.1016/j.atmosenv.2018.10.013.

Tonkin, Matthew J., and Steven P. Larson. 2002. “Kriging Water Levels with a Regional-Linear and Point-Logarithmic Drift.” *Groundwater* 40 (2): 185–93. https://doi.org/10.1111/j.1745-6584.2002.tb02503.x.

Torsney-Weir, Thomas, Steven Bergner, Derek Bingham, and Torsten Möller. 2017. “Predicting the Interactive Rendering Time Threshold of Gaussian Process Models with HyperSlice.” *IEEE Transactions on Visualization and Computer Graphics* 23 (2): 1111–23. https://doi.org/10.1109/TVCG.2016.2532333.

Torsney-Weir, Thomas, Michael Sedlmair, and Torsten Möller. 2017. “Sliceplorer: 1D Slices for Multi-Dimensional Continuous Functions.” *Computer Graphics Forum* 36 (3): 167–77. https://doi.org/10.1111/cgf.13177.

Wijk, Jarke J. van, and Robert van Liere. 1993. “HyperSlice: Visualization of Scalar Functions of Many Variables.” In *Proceedings of the 4th Conference on Visualization*, 119–25. IEEE Computer Society. https://doi.org/10.1109/VISUAL.1993.398859.