

# Enhancing Comparative Model Analysis using Persistent Homology

Bastian Rieck   Heike Lette

Interdisciplinary Center for Scientific Computing  
Heidelberg University

November 9, 2014



UNIVERSITÄT  
HEIDELBERG  
ZUKUNFT  
SEIT 1386



**HGS**  
**MathComp**

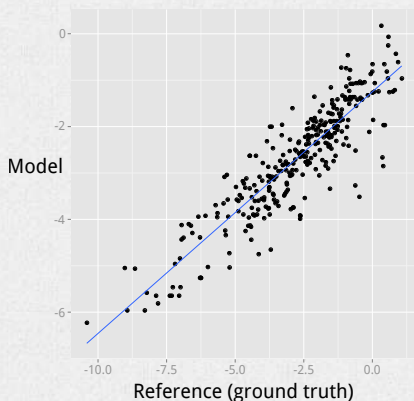


# Motivation

- ▶ Different predictive models for a process
- ▶ Which models are suitable?
  
- ▶ Example: Solubility of chemical compounds
- ▶ 19 different models
- ▶ 1267 chemical compounds, described by 228-dimensional feature vectors
- ▶ Measured ground truth



# Quality measures



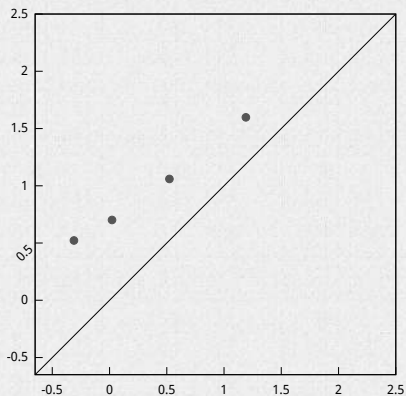
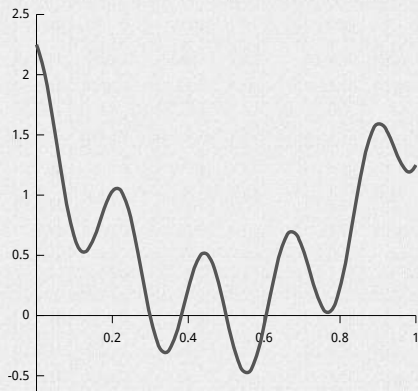
## State of the art

- ▶ Root-mean-square error (RMSE)
- ▶ Correlation coefficient ( $R^2$ )

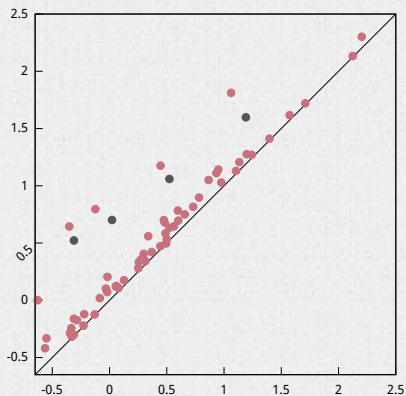
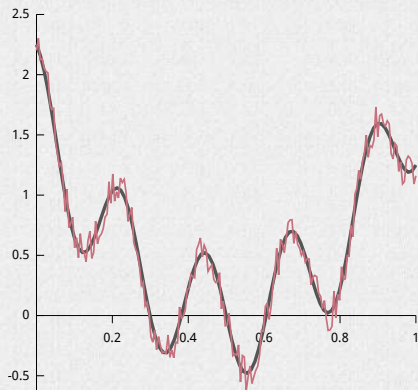
## Shortcomings

- ▶ Noise
- ▶ Reduced effects of outliers

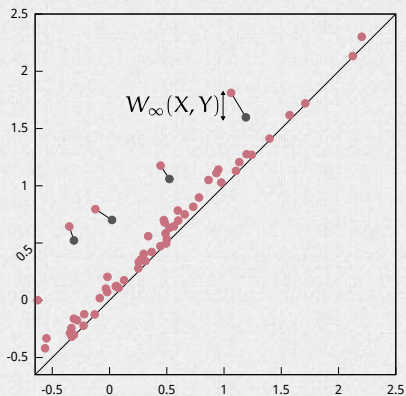
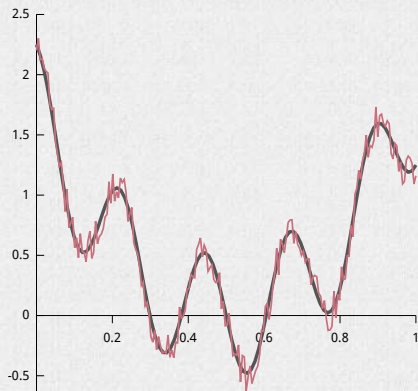
# Persistent homology



# Persistent homology

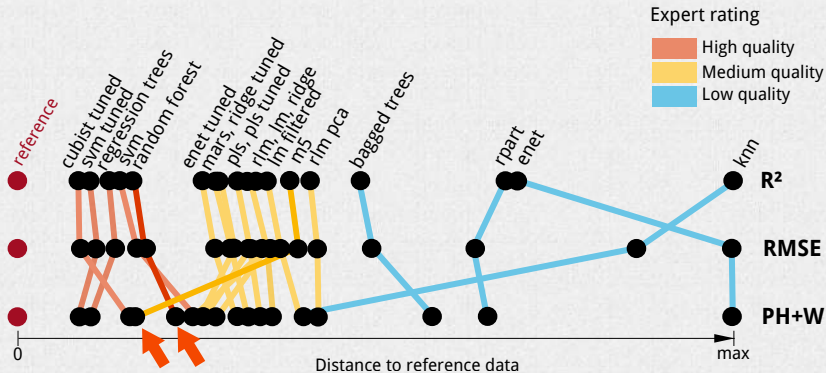


# Persistent homology

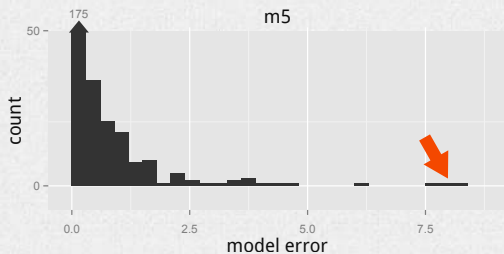
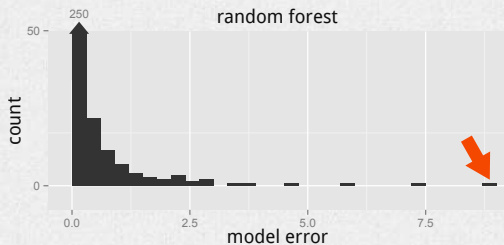


# Results

## Comparison to reference data



# Why the larger diversity?





# Results

## Model landscapes (model-to-model comparison)

