# Simulation Data Mining for Supporting Bridge Design

Ninth Australasian Data Mining Conference

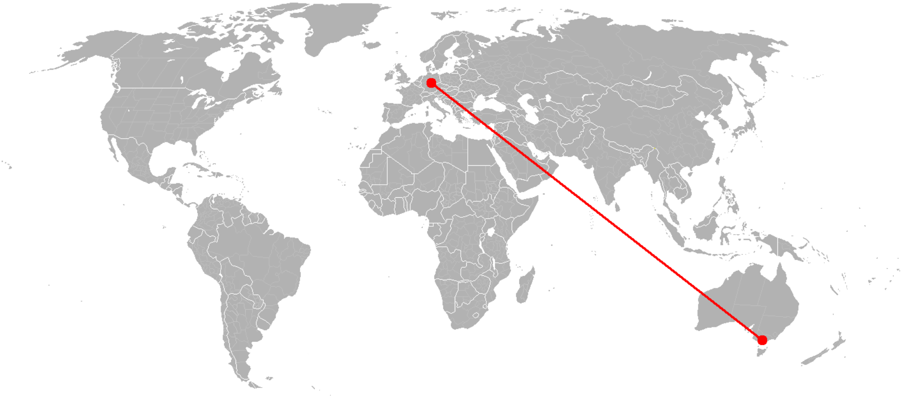
Steven Burrows[†] Benno Stein[†] Jörg Frochte[‡]
David Wiesner[†] Katja Müller[†]

[†] Bauhaus University Weimar
[‡] Bochum University of Applied Science

1–2 December 2011

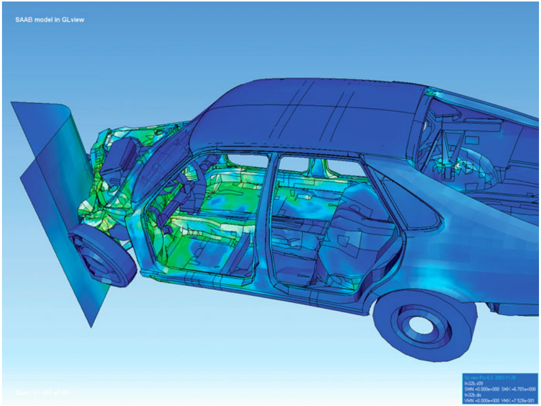
# About Me

- Undergrad, Honours, PhD, RMIT University, up to 2010.
- PostDoc, Bauhaus University Weimar, 2011 to current.
- "Strategies for Robust Design of Structures" project.
  - Includes simulation data mining and civil engineering sub-projects.



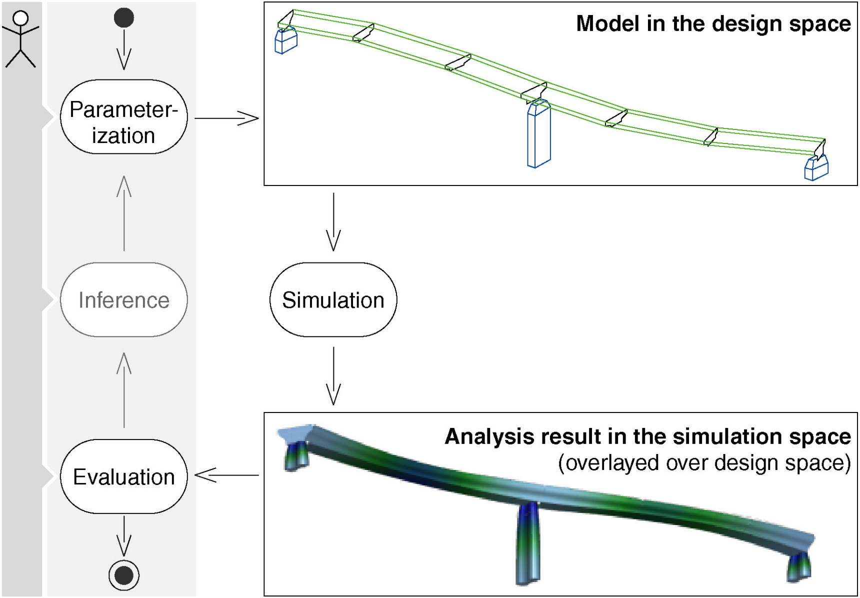
# Applications of Simulation Data Mining

- Car crashworthiness (Kuhlmann et al., 2005; Mei and Thole, 2007).
- Occupant restraint systems (Zhao et al., 2010).
- Aviation (Fayyad et al., 1996; Painter et al., 2006).
- Semiconductor manufacturing (Brady and Yellig, 2005).




http://en.wikipedia.org/wiki/Finite_element_method

# Interactive Bridge Design in Civil Engineering



# Supporting Bridge Design

Key idea:

- Mine patterns in pre-computed bridge simulation results.

Why simulation data mining?:

- Faster simulations, provide diagnosis, automated design, etc.

Consider models $\{m_i \in M\}$ and simulation results $\{y_i \in Y\}$:

- $\|\mathbf{y}_1 \ominus \mathbf{y}_2\| < \varepsilon \quad \Leftrightarrow \quad \varphi_{\mathrm{Design}}(\mathbf{m}_1, \mathbf{m}_2) \approx 1.$
- Develop $\varphi_{\mathrm{Design}}$ to predict the similarity of two designs with regards to learned behavior.

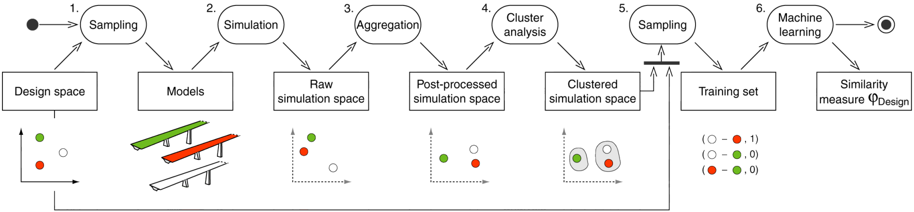
Questions we can answer:

1. Identify good '?' in $\varphi_{\mathrm{Design}}(\mathbf{m}_1, ?) \approx 1.$
2. Predict the behavior of a new model $m$.
3. Learn cost optimization rules for any equivalence class $M' \subseteq M$.

# Methodology

Computation of the similarity measure in six steps:

1. Sample candidate designs.
2. Simulate the models.
3. Aggregate the simulation results.
4. Cluster the simulation results.
5. Sample the simulation results.
6. Learn a mapping from $\{m_i \in M\}$ to $\{y_i \in Y\}$.




Future work disclaimer:

- There are still competing alternatives in many steps to be explored.

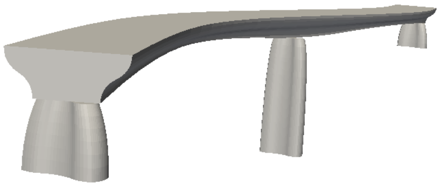
# Step 1: Sample Candidate Designs

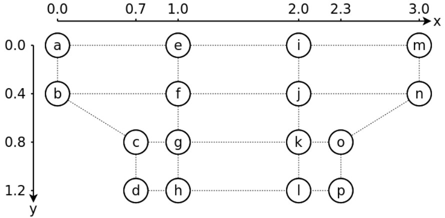
Data format:

- IFC (Industry Foundation Classes): An object-oriented data model for describing entities in the construction and building industries.
- IFC-Bridge: An extension to IFC for bridges.
- NURBS (Non-Uniform Rational Basis Spline): Novel extension to IFC-Bridge in the project.

Data set:

- 14 641 geometry and material permutations of the model below.





# Step 2: Simulate the Models

Input:

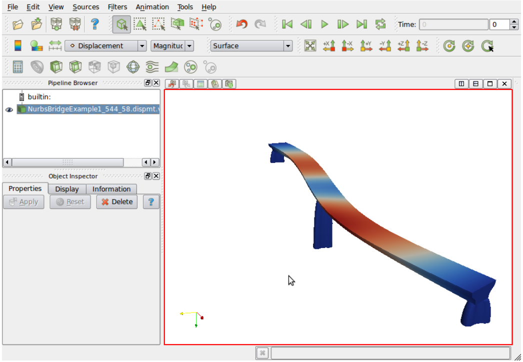
- IFC-Bridge data models (Lebegue et al., 2007) with NURBS.

Simulation Engine:

- Finite Element Method implementation (Gerold, 2010). Our “oracle”.

Output:

- VTK (Visualization Toolkit) format (Schroeder et al., 1996).



# Step 3: Aggregate the Simulation Results

Original data:

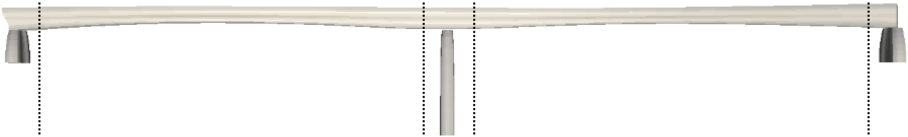
- 12 064 points and measurements from the FEM mesh.

Process:

- Consultation with a Numerics professor.

Aggregated data (45 measurements):

- Five regions (below).
- Maximum displacement, strain, and stress.
- X, Y, and Z co-ordinates.



# Step 4: Cluster the Simulation Results

Goal:

- Learn similar groupings of simulated models.

Clustering algorithms:

- K-means (Hartigan and Wong, 1979).
- Hierarchical Agglomerative Clustering (Gowda and Krishna, 1978).
- AiTools implementation (http://webis.de/research/projects/aitools).

Evaluation:

- Expected Density measure (Stein et al., 2003).
- Higher quality clusterings give have higher expected density score.

# Step 5: Sample the Simulation Results

Example (350 items):

| | |
|---|---|
| Cluster A: 100 items. | $= 4\,950$ $\left(\frac{100(100-1)}{2}\right)$ positive pairs. |
| Cluster B: 120 items. | $= 7\,140$ $\left(\frac{120(120-1)}{2}\right)$ positive pairs. |
| Cluster C: 130 items. | $= 8\,385$ $\left(\frac{130(130-1)}{2}\right)$ positive pairs. |
| | $= 20\,475$ positive pairs. |
| | $+ 40\,600$ negative pairs. |
| | $= 61\,075$ total pairs. |

Sampling strategy (from approximately $10^8$ pairs):

- Class balance.
- Equal sampling from each cluster of positive pairs.
- Random sampling for negative pairs.

# Step 6: Machine Learning

Training data:

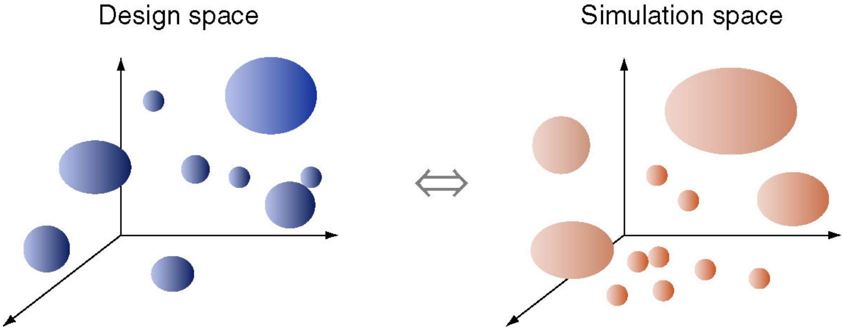
- Duples in the form $\langle \mathbf{m}_k \ominus \mathbf{m}_l,\, c_j \rangle$.

Learning:

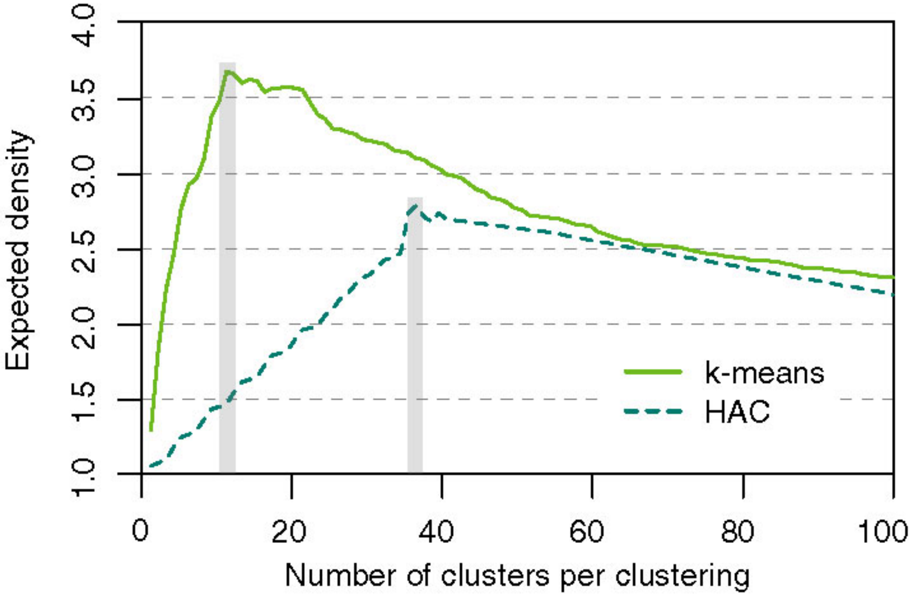
- Ten-fold cross validation.
- Naive Bayes and Maximum Entropy classifiers (Burrows et al., 2011).

Outcome:

- Class probability estimates [0, 1] for evaluating $\varphi_{\text{Design}}$.



# Clustering Results



# Accuracy Results

| | K-means (12) | | HAC (37) | |
|---|---|---|---|---|
| Data set size | Naive bayes | Entropy | Naive bayes | Entropy |
| 100 | 94.0 | 94.0 | 94.0 | 96.0 |
| 200 | 92.5 | 93.0 | 90.0 | 91.0 |
| 500 | 85.4 | 90.6 | 89.8 | 90.8 |
| 1 000 | 91.4 | 94.3 | 90.8 | 91.2 |
| 2 000 | 88.6 | 92.4 | 89.8 | 91.0 |
| 5 000 | 89.4 | 92.7 | 89.3 | 90.5 |
| 10 000 | 89.6 | 92.4 | 88.5 | 89.4 |
| 20 000 | 89.8 | 92.3 | 89.6 | 90.3 |
| 50 000 | 89.9 | 92.7 | 89.1 | 90.1 |
| 100 000 | 89.7 | 92.4 | 89.1 | 89.7 |
| 200 000 | 89.8 | 92.5 | 89.1 | 89.8 |
| all | 89.8 | 92.5 | 89.1 | 89.7 |

# Future Work

- Use of a rank correlation co-efficient such as Spearman's *rho*, Pearson's *r*, or Kendall's *tau* to compare the correlation of the ranks of $\varphi_{\text{Design}}$ with the ranks of the cosine similarity taken from the simulation space.
- Apply clustering instead of ranking for the evaluation, and compare the coverage of the clusterings (F-measure).
- Apply *domain decomposition* as a parallelization technique for solving partial differential equations in FEM analysis.

# Summary

- Mine patterns in pre-computed bridge simulation results for knowledge discovery.
- Six step methodology for computing $\varphi_{\mathrm{Design}}$ so that new questions can be answered.
- Initial results are promising, but more remains for future work.

Thankyou!

Steven Burrows
steven.burrows@uni-weimar.de
www.webis.de