Knowledge Discovery in Manufacturing Simulations

Steffen Strassburger

Ilmenau University of Technology
Germany

Contents

- **Motivation and Methodology**
- Case Study 1: Gold Mining Facility
- Case Study 2: Manufacturing Line
- Summary & Future Work

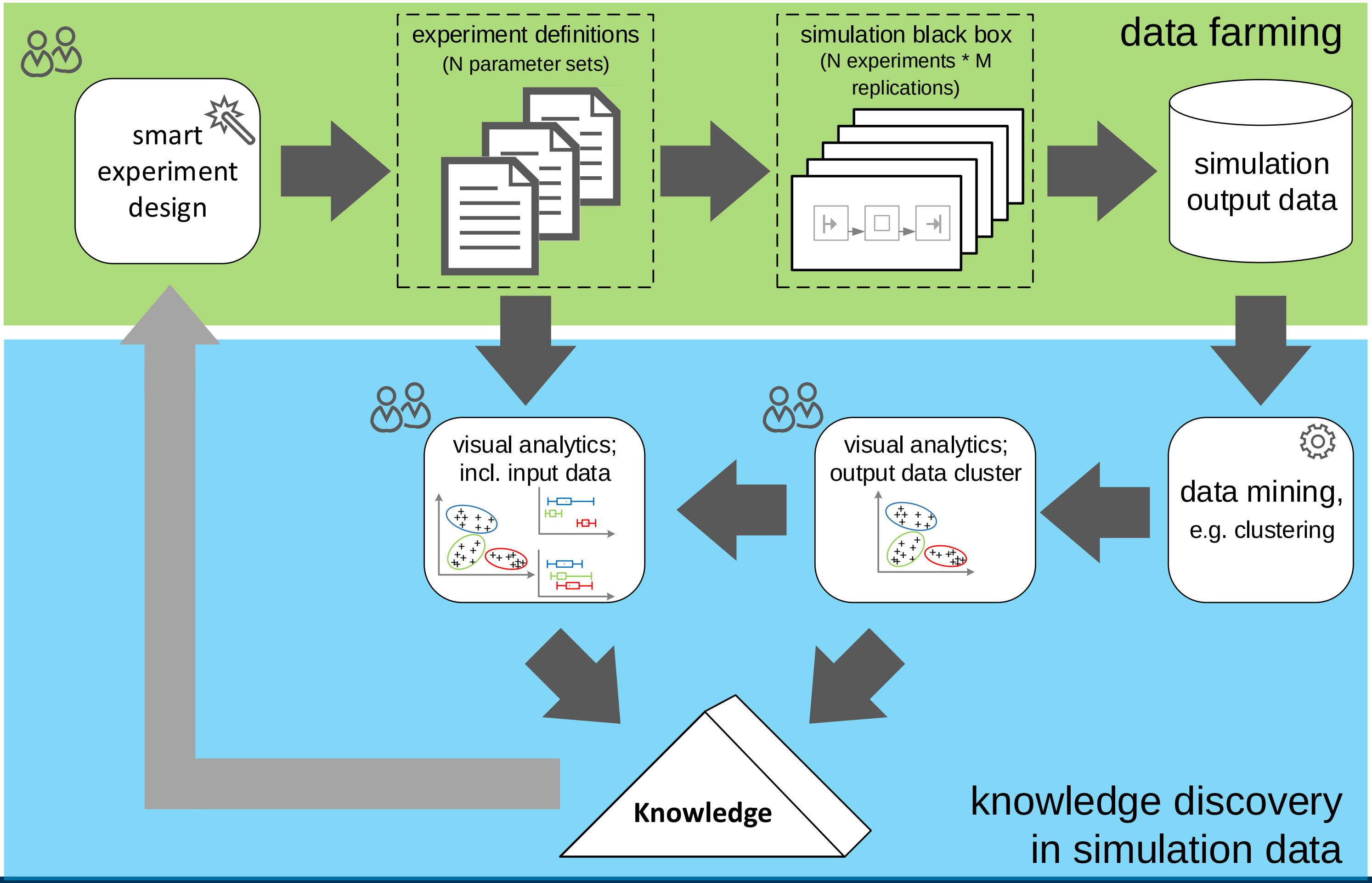
Traditional Simulation Studies Limit Knowledge Discovery

- Common way of conducting simulation studies:
  - Define project scope and goal
  - Build and validate simulation model
  - Manually experiment with model
  - Implement improved parameters in the real system

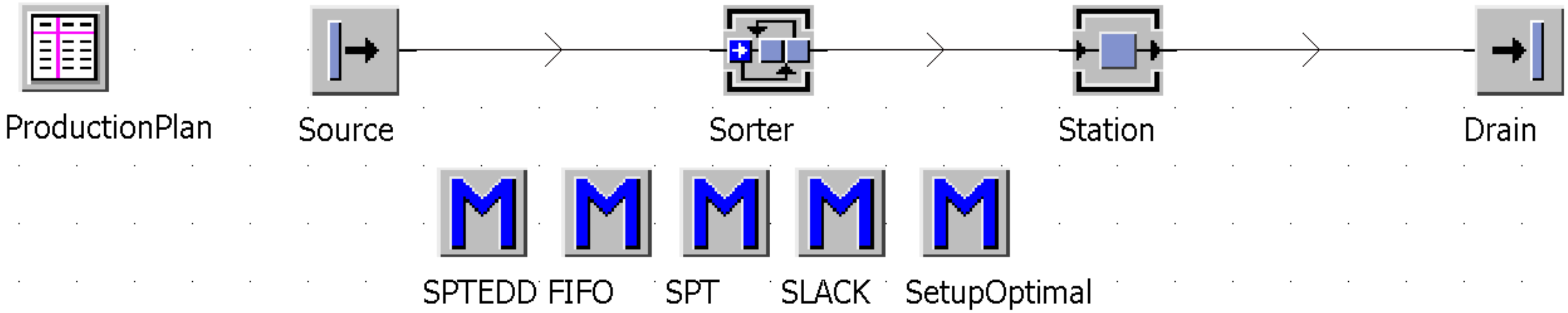
Objective: To derive a new way for Knowledge Discovery in Simulations

- Investigate ways for broad scale simulation experiments (“data farming”)
- Investigate data mining algorithms to uncover interesting patterns and relations
- Investigate ways for gaining knowledge (useful information) about the system through visually aided analysis (“visual analytics”)

Combining Simulation and Visual Analytics: Knowledge Discovery in Simulations



A basic example*



*Example from [FBS2015a] and [FBS2015b]

Data Creation, basic example

Input Factor	Margins	Levels
Inter arrival time	60s-240s	18 (10s interval)
Sorter capacity	10-1000	10 (100 Slots interval)
Sorter strategy	5 strategies	5
Product mixture (Seven product types)	0-100% per product	47 Experiments (NOLH-Sampling*)
Random number stream	1-10	10

- ca. 500.000 simulation runs total

*NOHL: Nearly Orthogonal Latin Hybercube

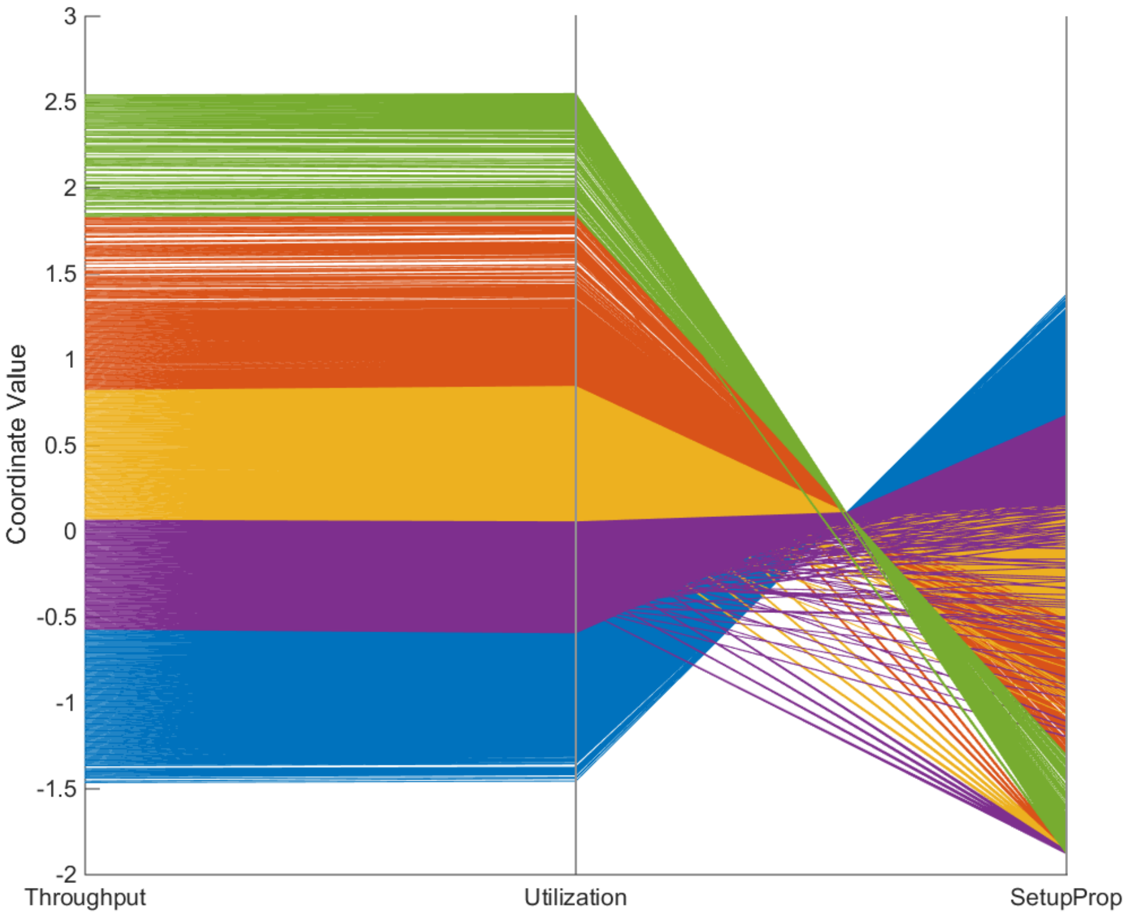
Data Creation

- 50.000 experiments (10 replications) = 500.000 sets of data
- Each set represents one simulation run
- Each simulation run has a set of input values (defined by experimental design) and output values
- Output values are generated by the simulation model; represent system performance measures, e.g. throughput of jobs, station utilization,..

Data Mining / Clustering

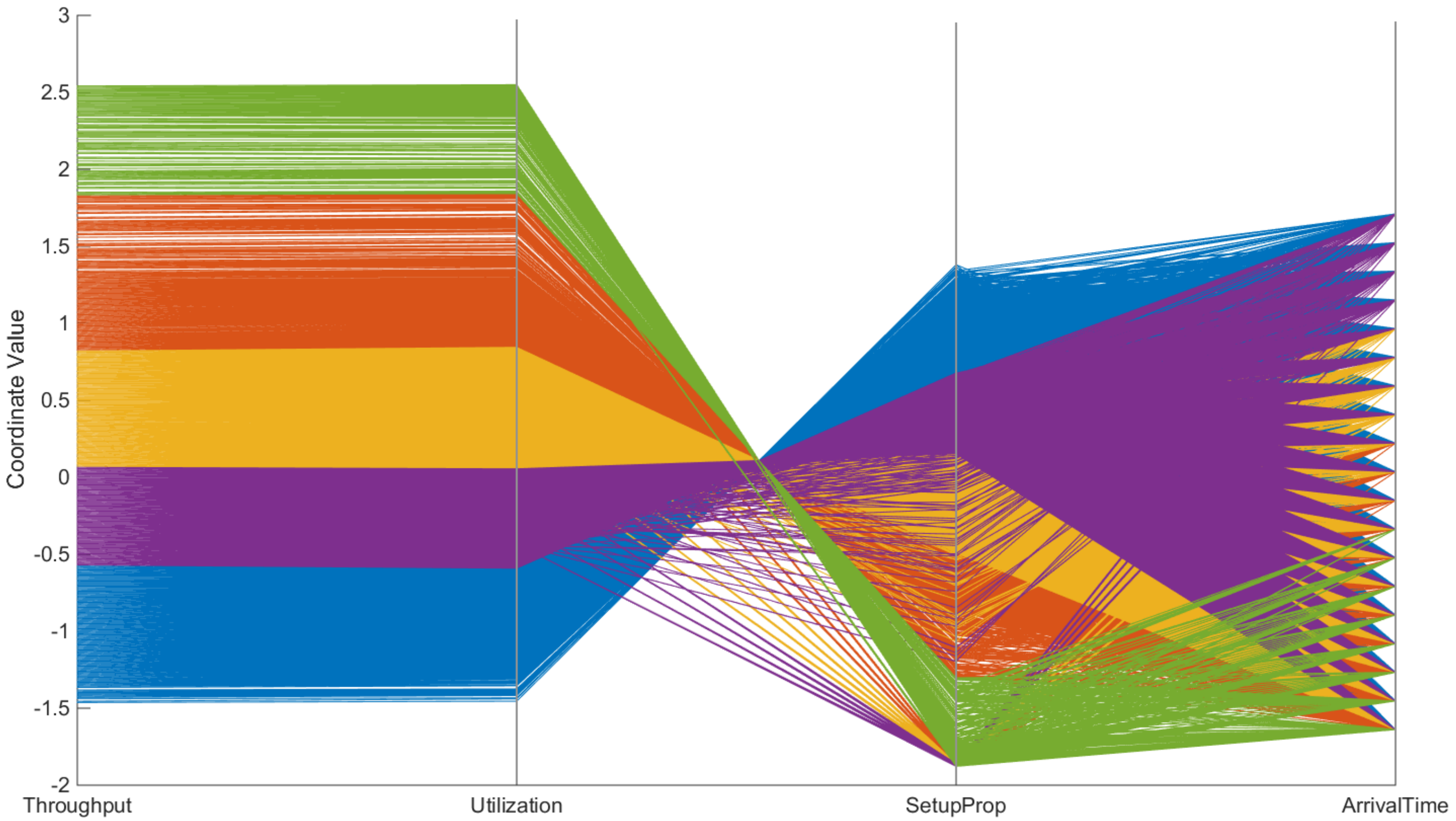
- Data mining : apply algorithms over data that are able to extract patterns
- Here: Application of a clustering algorithm
  - Group simulation runs into clusters based on output performance values
  - Simulation runs in the same cluster have similar performance values
  - Investigate clusters
  - Create knowledge by finding interesting relations to corresponding input parameter values

Visualizing Clustering Results

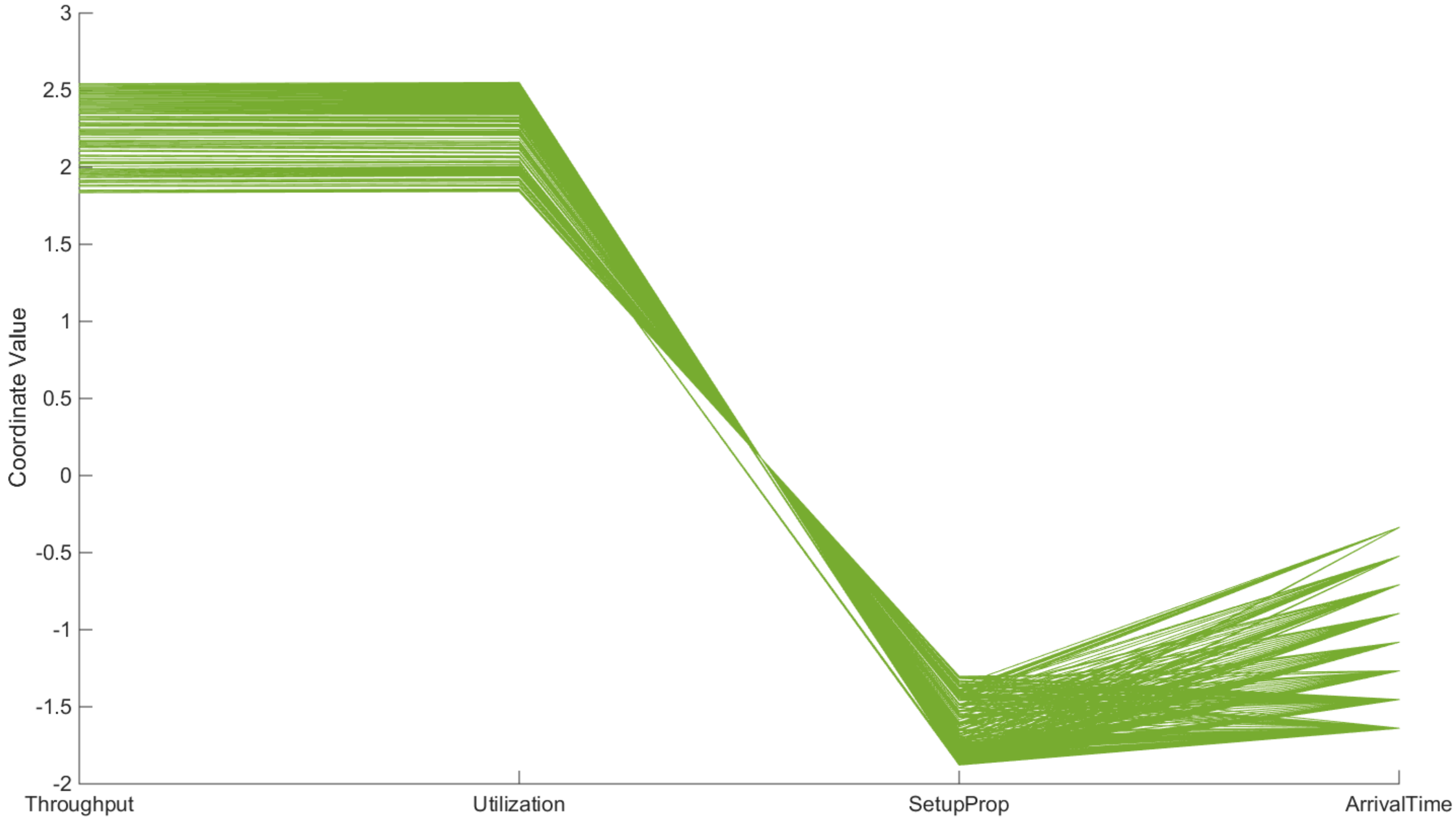


- Each line represents one simulation run
- Clusters are indicated through colour
- Hierarchy of system performance through clusters
- Next step : investigation of input parameters

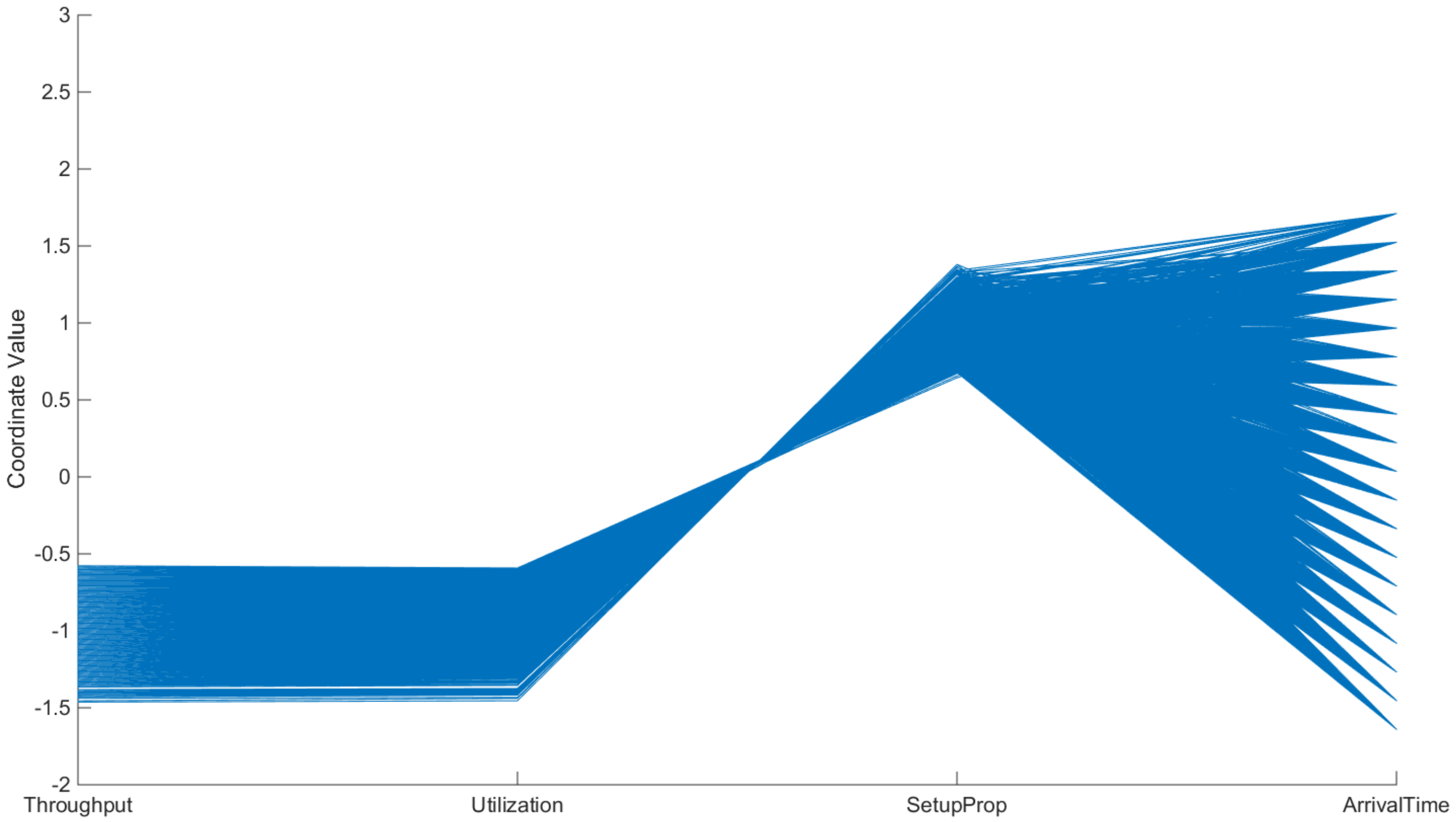
Visualizing results, relations to input parameters (arrival time)



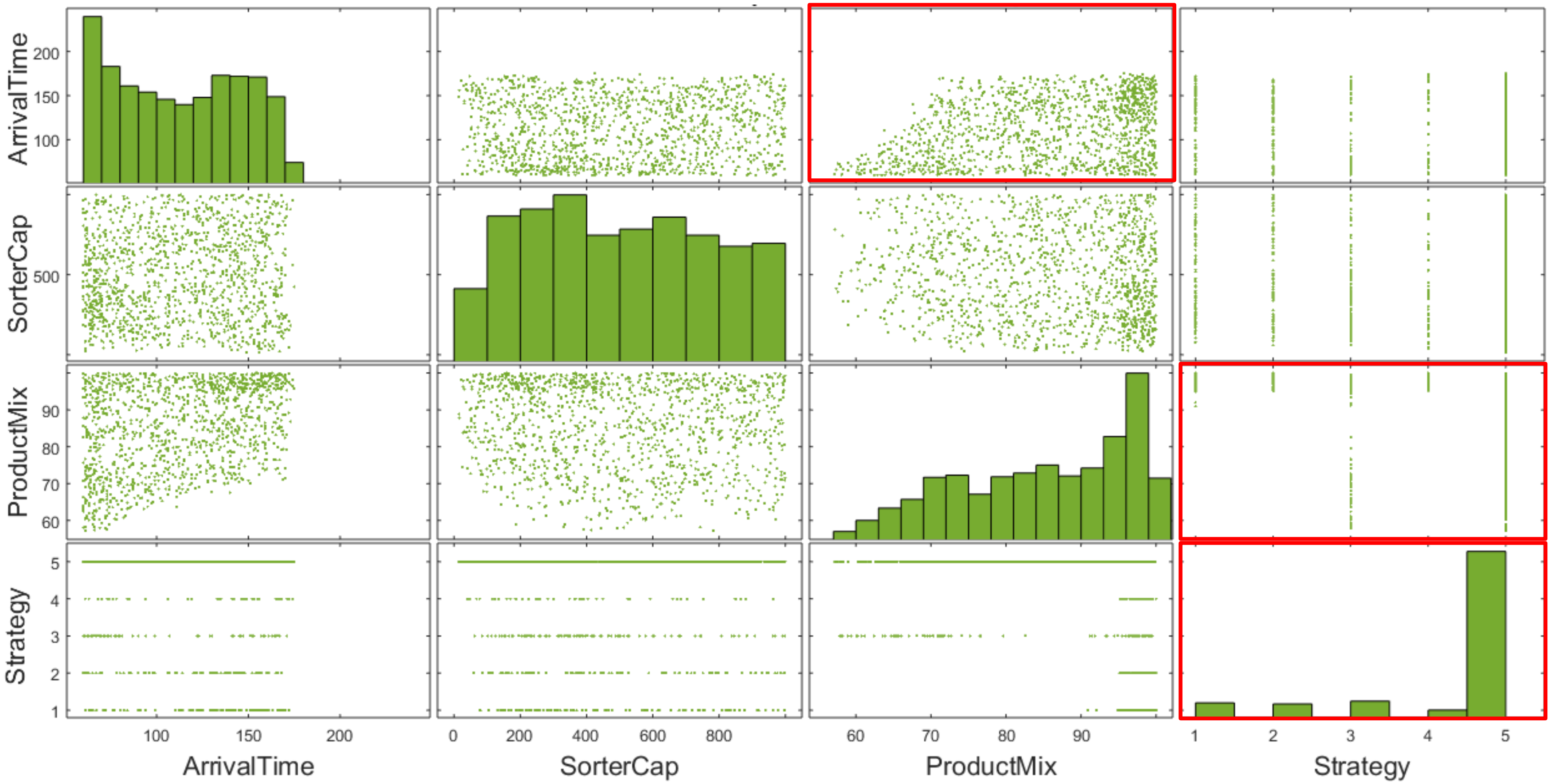
Interaction (Filtering), Relations to Input Parameters



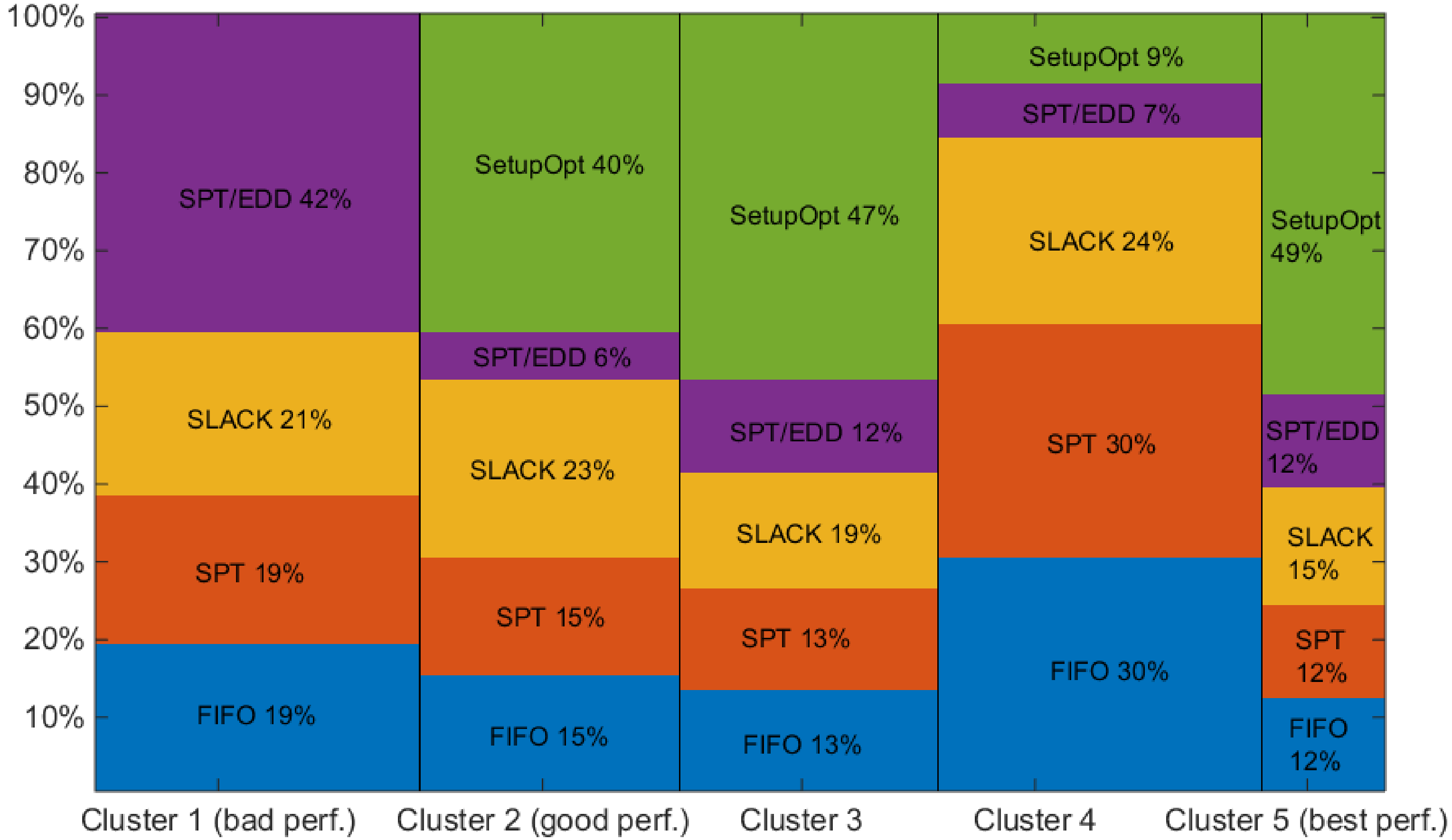
Interaction (Filtering), Relations to Input Parameters



Scatter matrix plots of input parameters can further help to identify relationships



Visualizing results, relations to input parameters (sorting strategy)



TECHNISCHE UNIVERSITÄT ILMENAU

Interpretation and Knowledge Creation

- Knowledge creation in our basic example system:
  - Short inter arrival times are prerequisite for good system performance
  - Break even point where frequency of jobs leads to blocking through setup processes
  - Setup optimal scheduling helps to reduce this effect
- We successfully extracted useful information about our system through data mining (clustering) and suitable visualizations and interaction (filtering)

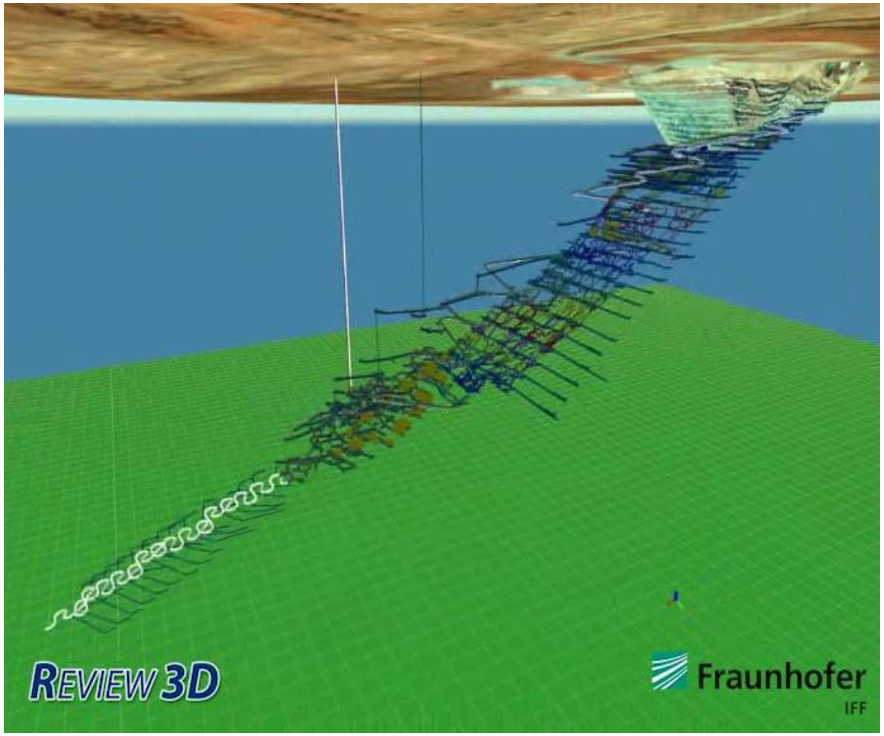
Short Summary

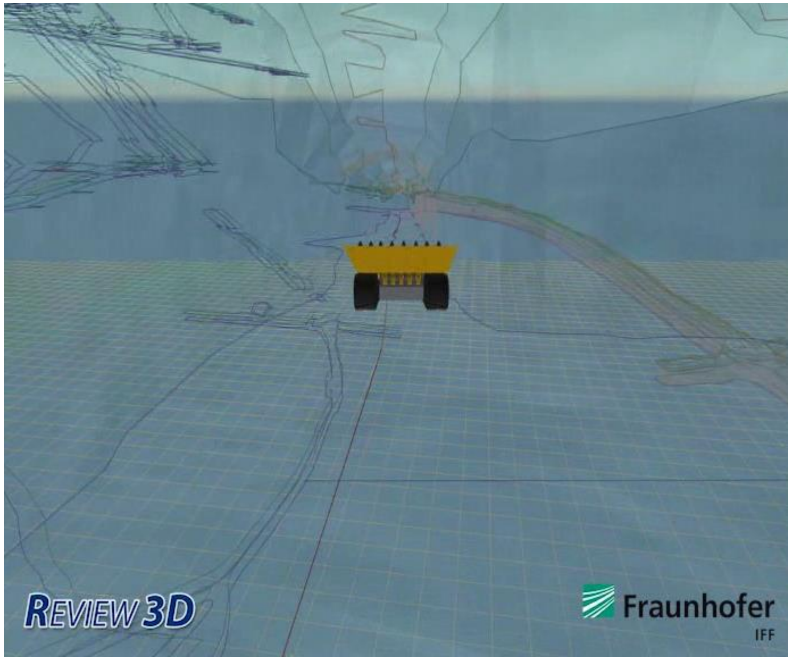
- Knowledge discovery process with simulation as a data generator and visualization as an analysis tool
- Potential improvements
  - different data mining tools and methods
  - investigation of other suitable visualizations
- Next: Application to more complex real life production systems

Contents

- Motivation and Methodology
- **Case Study 1: Gold Mining Facility**
- Case Study 2: Manufacturing Line
- Summary & Future Work

Case Study: A Gold Mining Facility in Australia*





*Case study from [FB+2016]

Key Questions

- Are there any interesting patterns and relations in the simulation data that may create additional knowledge for the mining engineers?
- Which configuration of haul trucks performs consistently well at base level 1100?
- How does it hold up against deeper loading levels and how do performance and cost parameters alter when depth of loading points increases?

Data Generation (1)

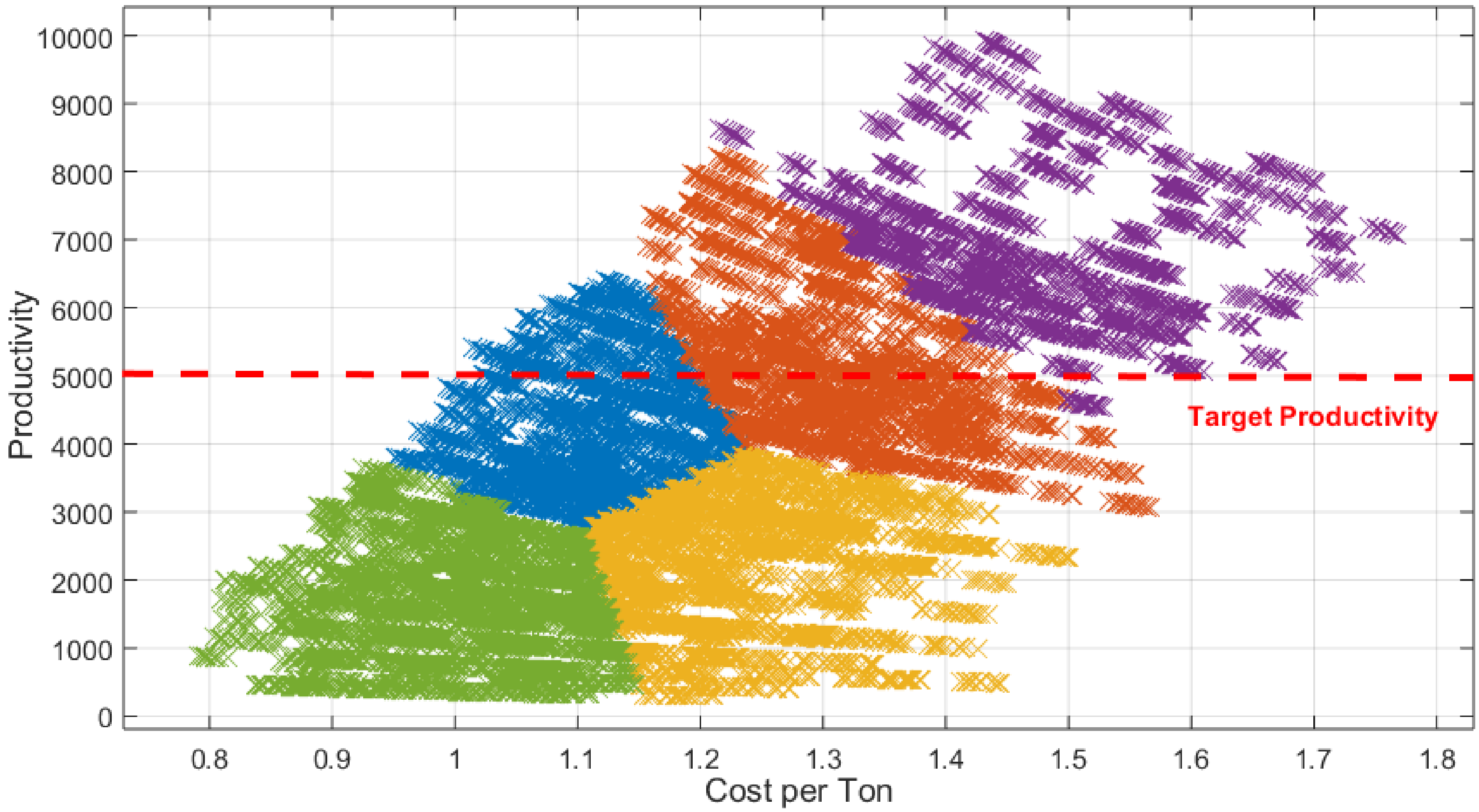
Factor	Margin	Scale	Description
Decision Factors			
#Trucks	1 – 20	discrete	number of haul trucks
Tonnage	[20-50] (increments of 10)	categorical	payload of a haul truck in ton
LoadingPort	[1100-2000] (increments of 100)	categorical	depth of loading port in meters below surface
Shift	9 - 11	discrete	shift regime with active shift duration in hours
Noise Factors			
SpeedDown	8 - 14	continuous	truck down driving speed in km/h
LoadingTime	2 - 5	continuous	loading time for a haul truck in minutes
WorkshopRate	10 - 20	continuous	probability for unscheduled maintenance

Data Generation (2)

- 262.141 experiments*
- Simulation period of 31 days (1 day for warm up and 30 days for result data collection)
- Parallel execution on computing cluster: runtime ~30h on 10 machines (300h)
- Implementation
  - Simulator: SLX
  - Data collection: MongoDB
  - Data computation (experiment design, data mining, visualization): Matlab

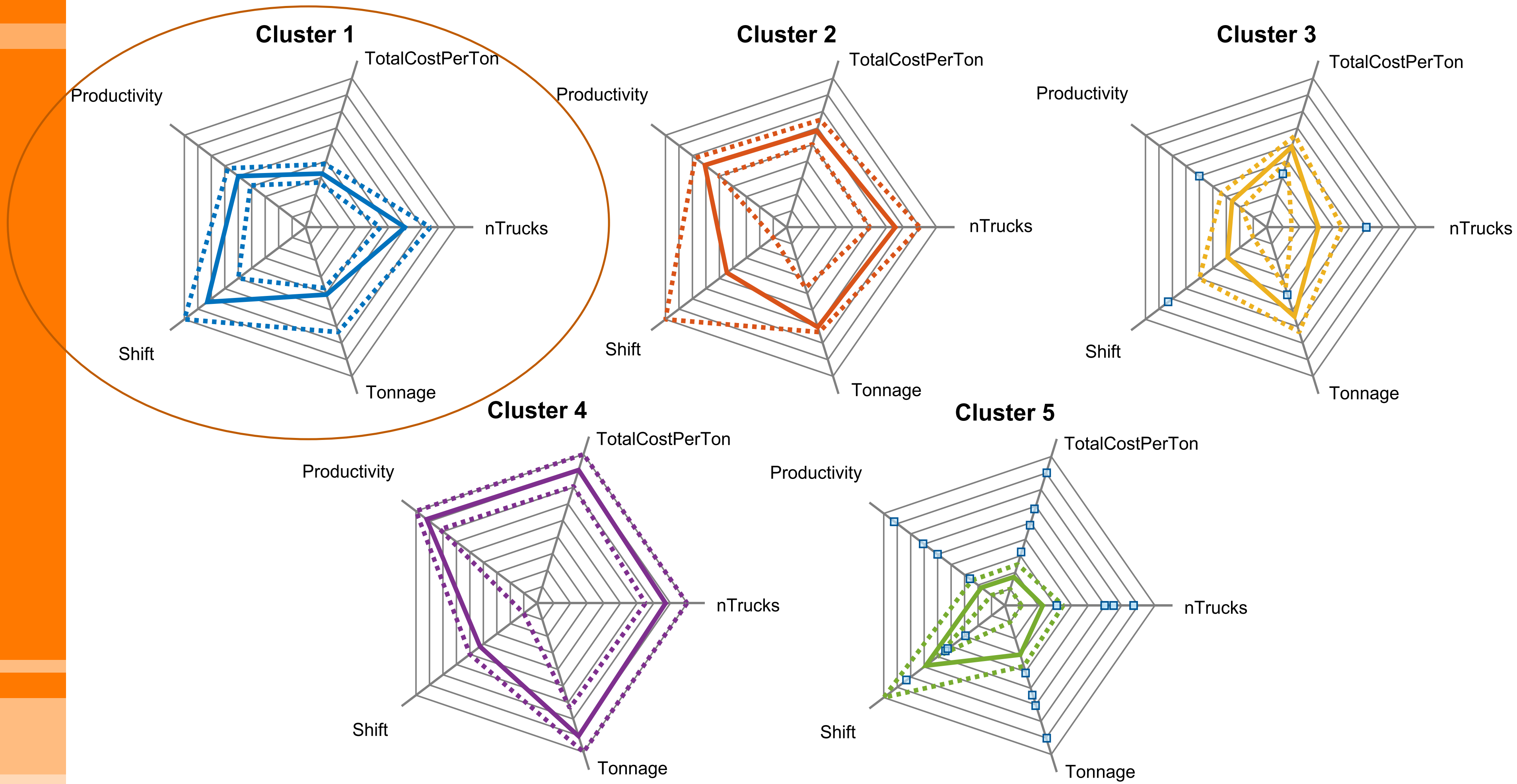
(*based on a nearly balanced nearly orthogonal hypercube design [VS+2011])

Data Analysis (1): Clustering of simulation runs based on output performance*

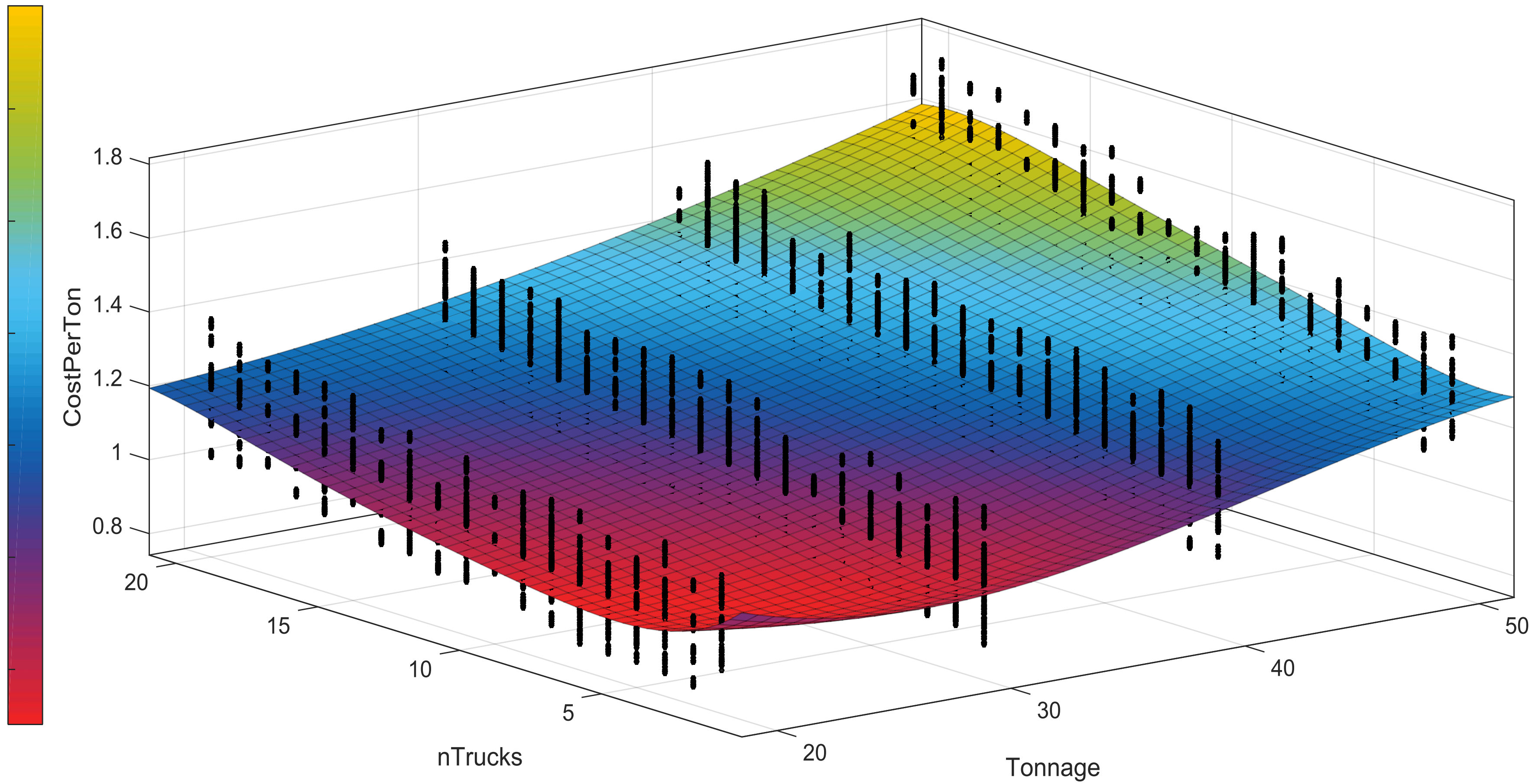


* using productivity (tons per day) and cost per ton

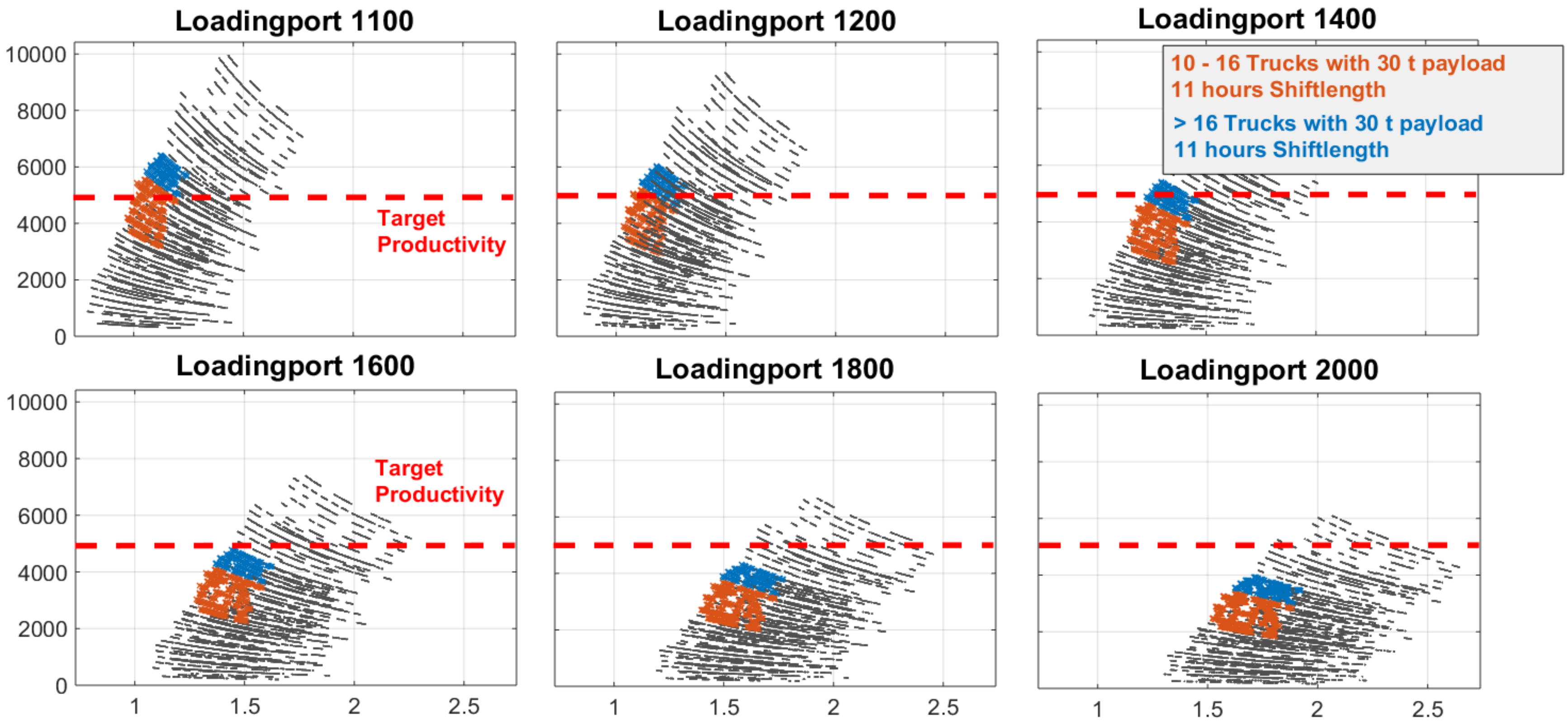
Data Analysis (2): Linear regressing models help identify most influential input parameters



Data Analysis (3): Different Levels of Cost per Ton for each Truck Portfolio



How does productivity and cost hold up on deeper loading levels?



Short Summary

- First real world simulation model used to validate our methodology
- Generated knowledge by using data mining algorithms and visual data exploration

Contents

- Motivation and Methodology
- Case Study 1: Gold Mining Facility
- **Case Study 2: Manufacturing Line**
- Summary & Future Work

Case Study 2

- Slides removed due to existing confidentiality agreement.

Contents

Summary

- Knowledge Discovery in (Manufacturing) Simulation Data is possible
- Differences from related work
  - Integrated process of data farming, data mining, and visual analytics
  - Challenges intrinsic to application domain, i.e., experiment design for product mixes
- Real world case studies confirm benefits of method

Future and Ongoing Work

- Extensions to methodology
  - Further data mining methods
  - Finding suitable visualizations
  - Making visualizations more interactive
- Software framework
- Online Mining of simulation data
  - Make results available while experiments are still running
  - Define new experiments "on the fly"

Literature

[FBS2015a] Feldkamp N., Bergmann S., Strassburger S. “Visual Analytics of Manufacturing Simulation Data”. In: *Proceedings of the 2015 Winter Simulation Conference*, eds. L. Yilmaz, W. K. V. Chan, I. Moon, T. M. K. Roeder, C. Macal, and M. D. Rossetti, pp. 779-790. Huntington Beach, CA, USA, December 6-9, 2015.

[FBS2015b] Feldkamp N., Bergmann S., Strassburger S. “Knowledge Discovery in Manufacturing Simulations”. In: *Proceedings of the 2015 ACM SIGSIM PADS Conference*, pp. 3-12. June 9-12, 2015, London, UK.

[FB+2016] Feldkamp, N., S. Bergmann, S. Strassburger, and T. Schulze. 2016. "Knowledge Discovery in Simulation Data: A Case Study of a Gold Mining Facility". In *Proceedings of the 2016 Winter Simulation Conference,* edited by T. M. K. Roeder, P. I. Frazier, R. Szechtman, E. Zhou, T. Huschka, and S. E. Chick, 1607–1618.

[VS+2011] Vieira, H., S. M. Sanchez, K. H. Kienitz, and M. C. N. Belderrain. 2011. "Improved Efficient, Nearly Orthogonal, Nearly Balanced Mixed Designs." In *Proceedings of the 2011 Winter Simulation Conference (WSC 2011),* edited by S. Jain, R. Creasey, J. Himmelspach, K. P. White, and M. C. Fu, 3600–3611, Piscataway, N.J.: IEEE.