Honda Research Institute EU

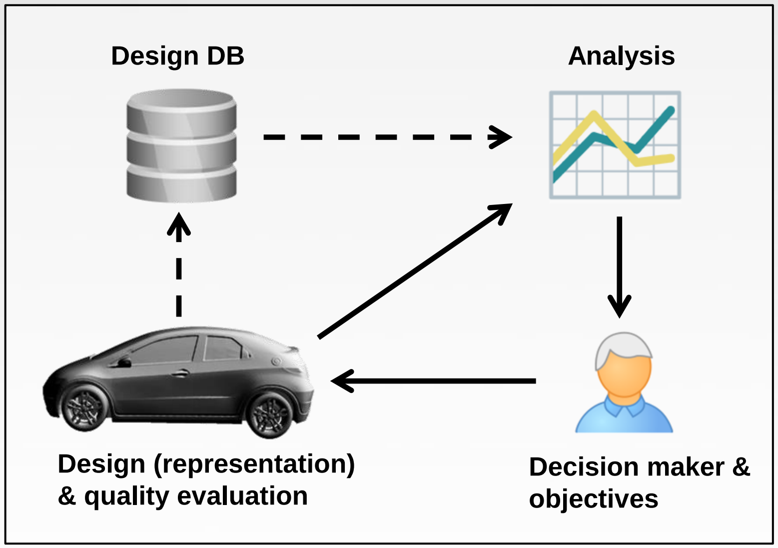
Engineering Data Analytics

innovation through science

B. Sendhoff and L. Gräning
Honda Research Institute Europe

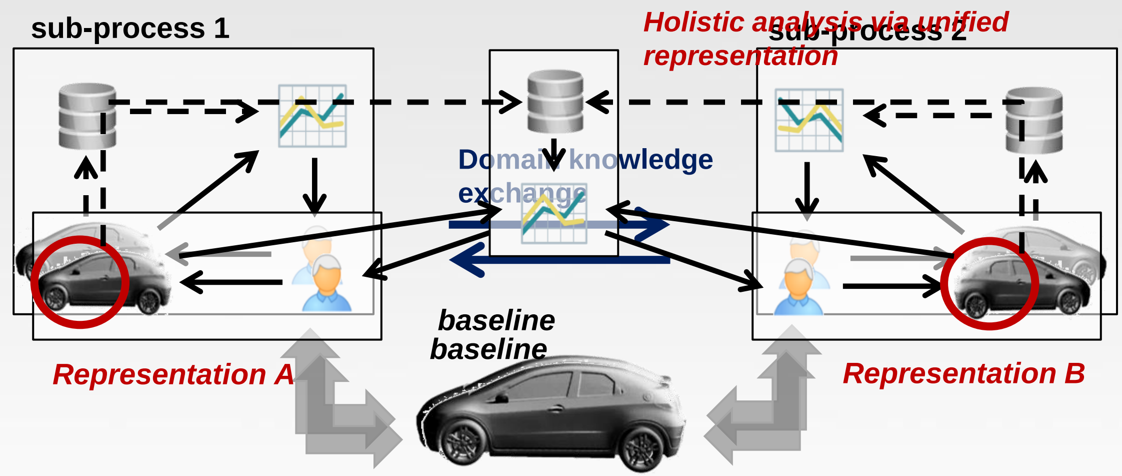
Engineering Data Analytics

From engineering design process …



- focus on one CAE discipline
- data analytics restricted to one discipline
- data analytics frequently restricted to one team and/or design step

… to the (holistic) design of complex systems



- optimal spatio-temporal problem decomposition
- multi-disciplinary design process
- multi team and multi tool integration
- cross-disciplinary data analytics
- knowledge extraction

Challenges: Size

- size of engineering CAE data scales with the demand for precision
- with the replacement of experiments by simulations data size increases dramatically
  - unsteady compressible LES simulations of mutiple engine cycles generated data in the **peta byte range**
  - high fidelity crash simulation
    ~100GB per analysis
    tera byte range
- test processes, e.g. for advanced driver assistant systems
  - sensory information can generate 5TB per hour
  - 10.000km estimated as 125 hours resulting in data in the **peta byte range**

Engine Simulation – Data Compression

- Compact feature based representation
  - store only data along traced particle
  - select & calculate features related to the design targets
  - preserve only streamlines with high information content (entropy based)

100% (1587)

50% (269)

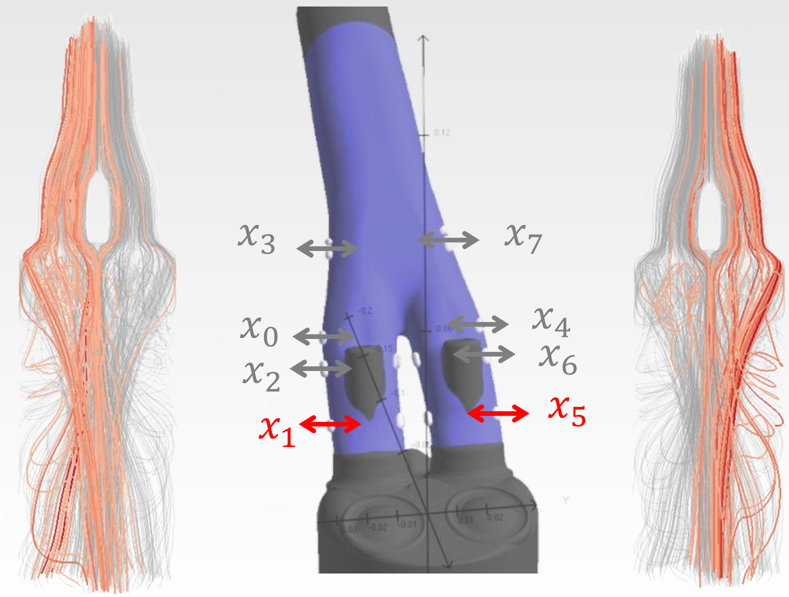
simulation with standard CFD mesh

turbulence mixing in cylinder

$$\boxed{l(u) : u \mapsto p, v, (M_v, M_t, M_d, M_s, M_\omega)^T}$$

1/5 of the streamlines preserve 50% of information

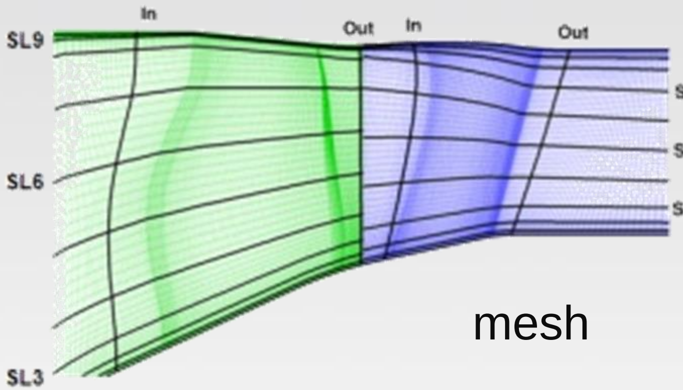
- Design objective (e.g. correlation to engine efficiency) based data compression



- reduction of uncorrelated flow data
- feature similarity estimated based on feature distributions along individual streamlines
- correlation between design changes and stream line features has been derived using mutual information

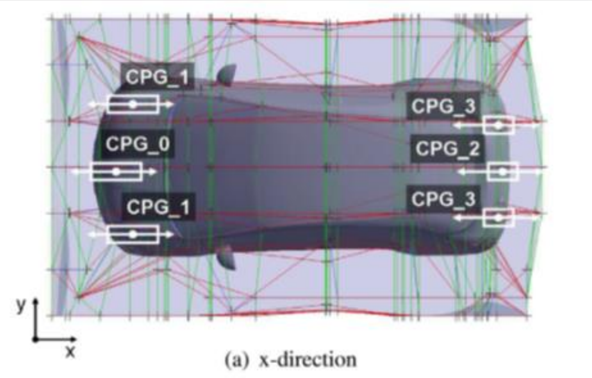
CAE Design Process – Three Spaces

(geometric)
constraints



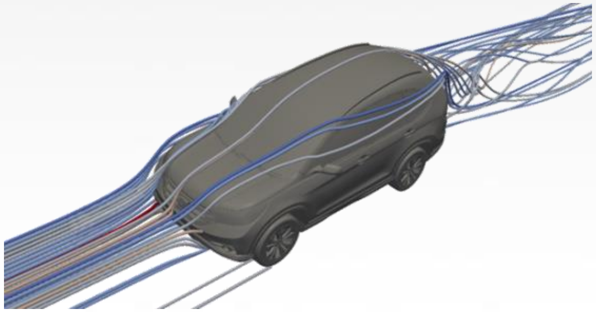
CAE

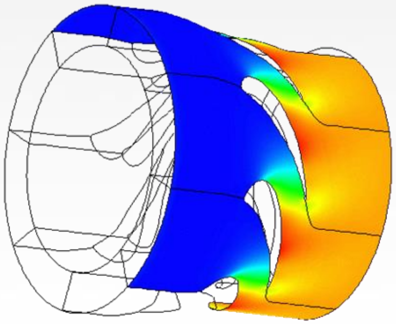
mesh

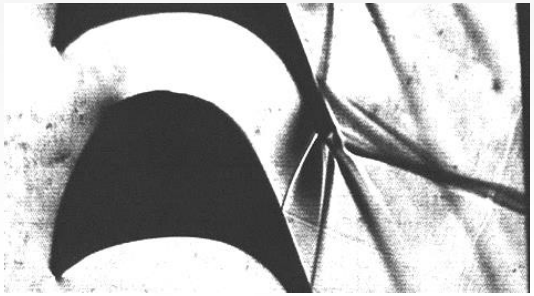


Representation

CAD or other

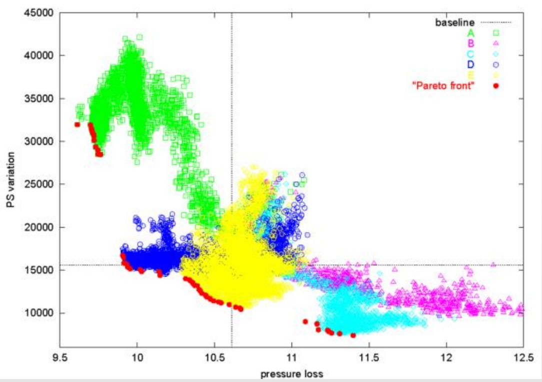






Flow Field

Constraints



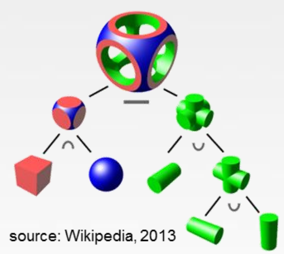
Objective Values

Challenges: Heterogeneity

- different model qualities (LES/RANS), different mesh sizes and mesh geometries due to different objectives and different stages in the design process requiring different local precision

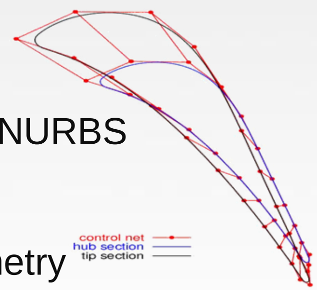
- different classes of design representations and design evaluation (CAE tools)

CAD

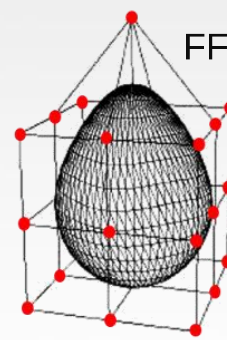


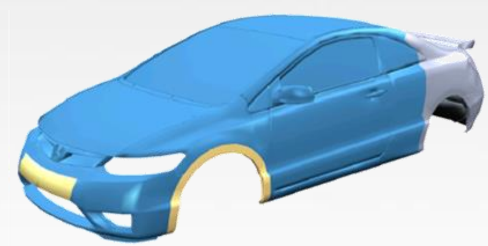
Constructive Solid Geometry

NURBS



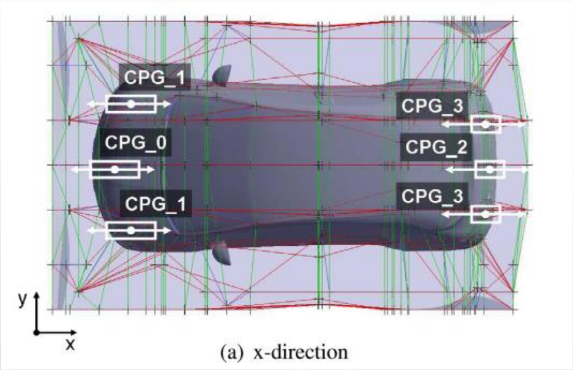
FFD

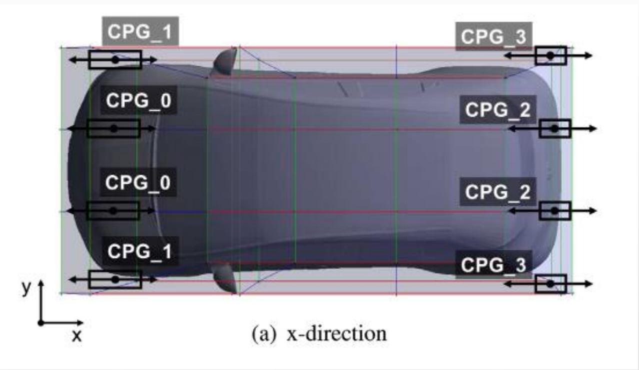




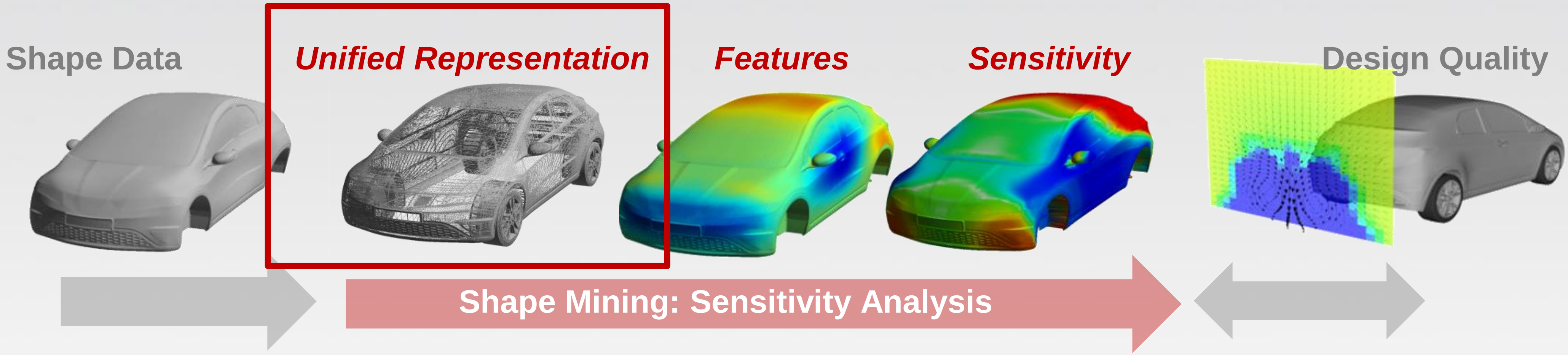
RBF based Space Deformation

- variations of design representations in one class due to different design focus, e.g. by different design teams or different stages of development





Unified Representation and Shape Mining



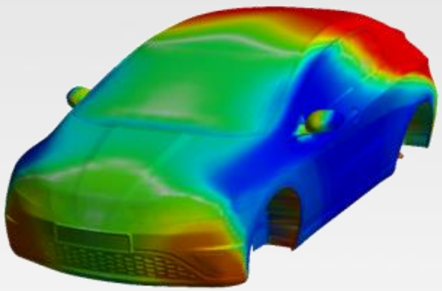
- unstructured polygonal surface mesh (simple and high-dimensional) $\mathcal{M} : (\mathcal{V}, \mathcal{K})$
  - list of vertices $\mathcal{V} = (\vec{v}_1, \ldots, \vec{v}_n)$
  - p simplices $\{i_1, i_2, i_3, \ldots, i_\mu\}$, with $i_l, l \in [1 \ldots n]$
  - enclosing a polygonal face made up of μ segments
  - the surface patches define a locally linear approximation of the surface

- vertex displacement – measuring differences to a chosen reference mesh

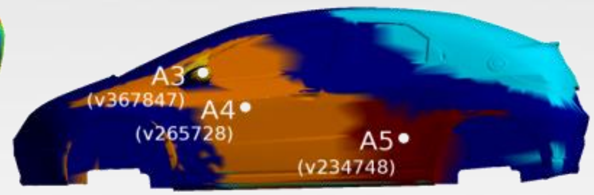
$$\Delta_i^r \triangleq \delta_{i,j}^{r,m} = \delta(\vec{v}_i^r, \vec{v}_j^m) = (\vec{v}_j^m - \vec{v}_i^r) \circ \vec{n}_i^r, \delta \in (-\infty, +\infty)$$

Shape Mining for Design Processes

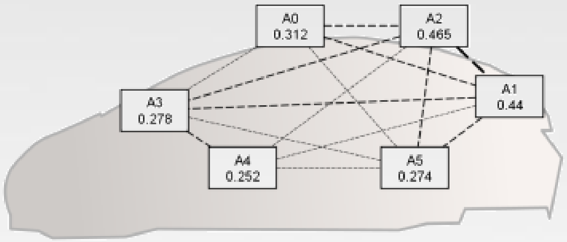
Surface Sensitivity



Sensitive Areas
(DimensionalityReduction)

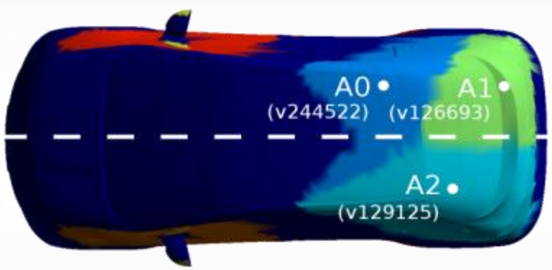


Interaction Analysis

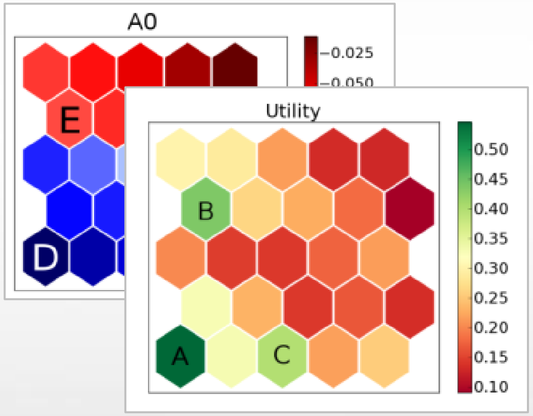


Shape Mining: Modeling Surface Interrelations

Geodesic distance

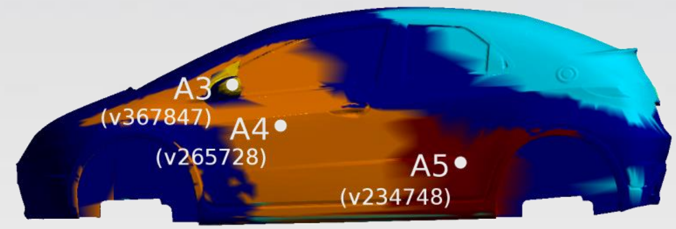


Concept Learning

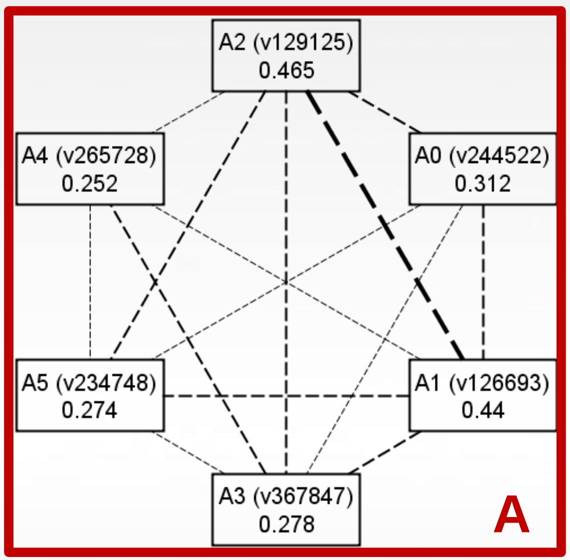


Interaction Analysis

- Interaction analysis provides an algorithmic basis to support and revise the decomposition of complex problems



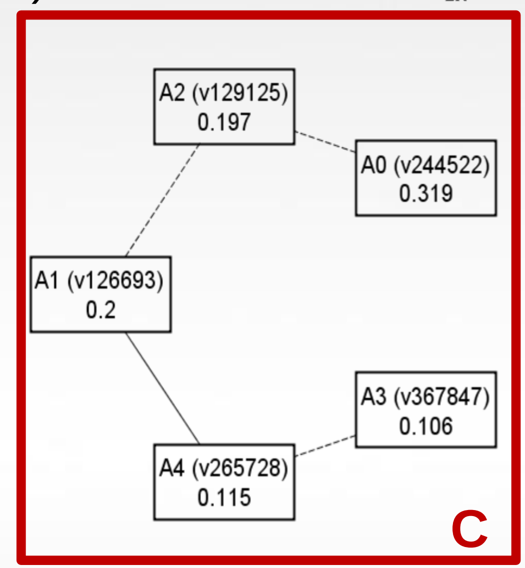
- Interaction calculation & visualization
  - Interaction values are normalized by the entropy of the design quality
  - Line thickness codes strength of the three-way interactions

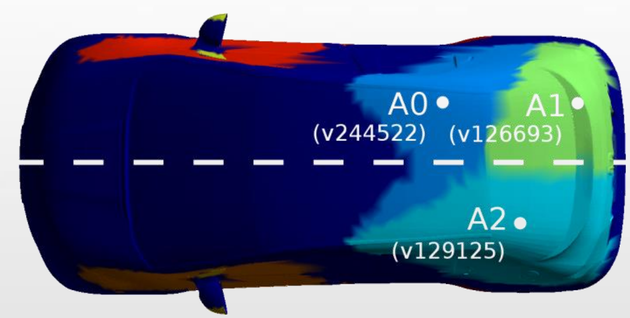


a) drag force

b) rear lift force

c) combined $\phi_F = \hat{\phi}_{F_{LR}} \cdot \hat{\phi}_{F_D}$





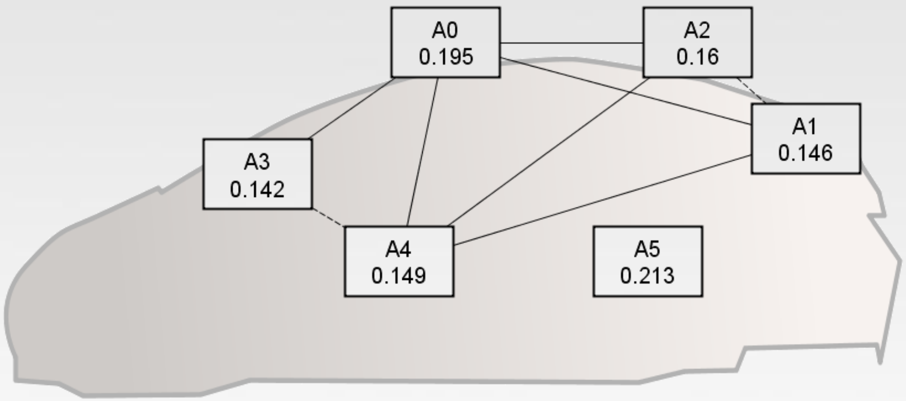
A. Redundancy among all parameters

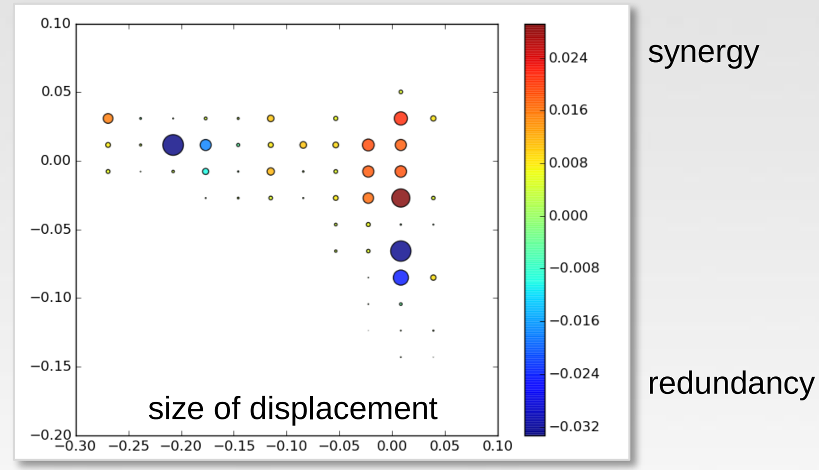
B. Interaction between parameters at roof and side

C. Redundancies and Synergies cancel each other out

Local Interaction – Flow Field Interaction

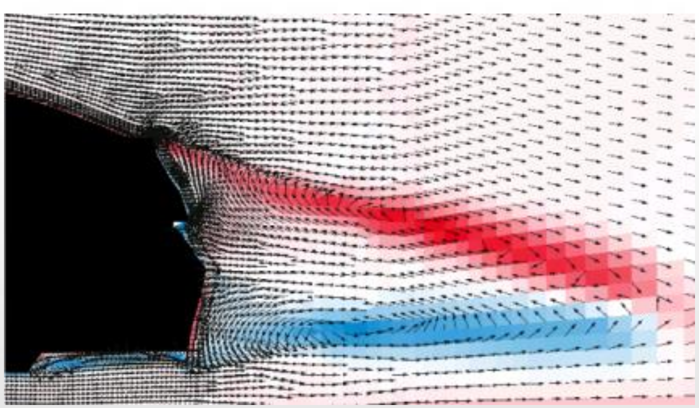
- local interaction analysis

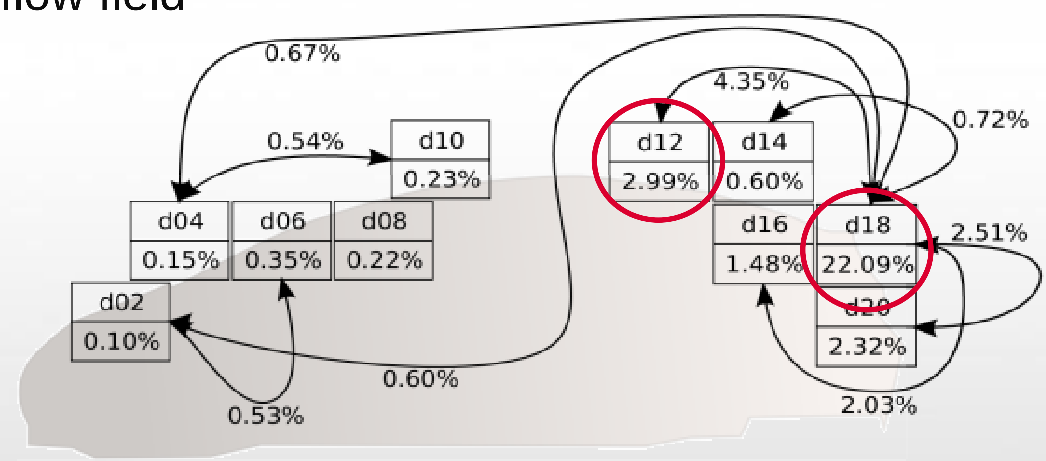




Interaction map visualizing local interaction between A1 and A4

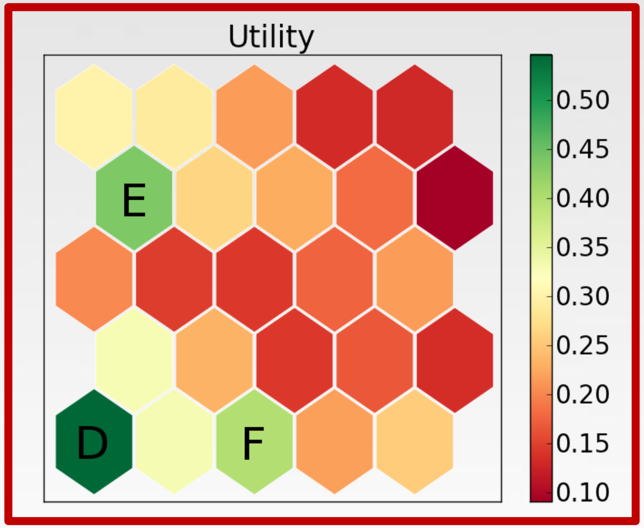
- interaction analysis based on the flow field

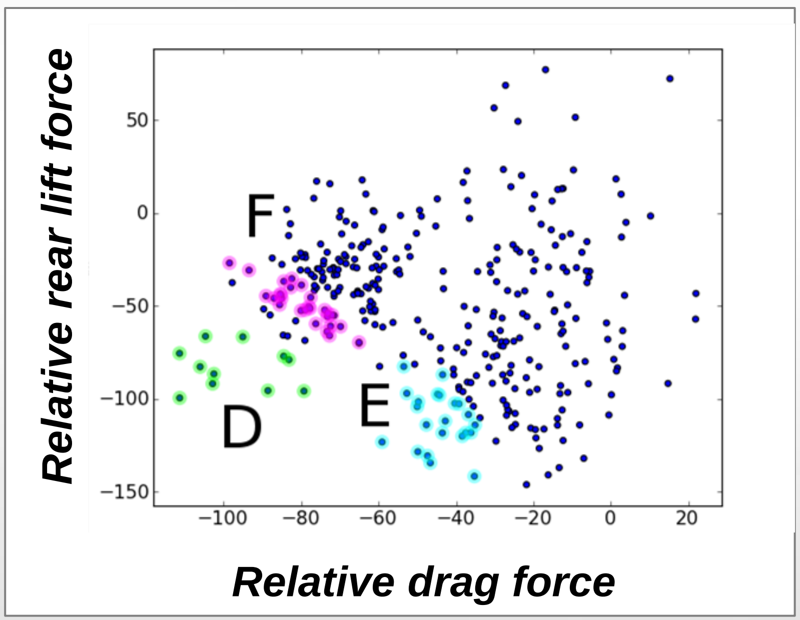




Concept Learning

- Estimating the relevance of the design concepts





Utility:

$$\mathrm{util}(A \to C, \boldsymbol{D}) = \underbrace{\mathrm{vol}(O, \boldsymbol{D})}_{} \cdot \underbrace{\mathrm{IS}(A \to C, \boldsymbol{D})}_{}$$

Hypervolume of non-dominated solution set

Interestingness measure: correlation between A and C

$$\mathrm{IS}(A \to C, \boldsymbol{D}) = \sqrt{\mathrm{conf}(A \to C, \boldsymbol{D}) \times \mathrm{conf}(C \to A, \boldsymbol{D})}.$$

$$\mathrm{conf}(A \to C, \boldsymbol{D}) = \frac{\mathrm{supp}(A \to C, \boldsymbol{D})}{\mathrm{cov}(A \to C, \boldsymbol{D})}$$

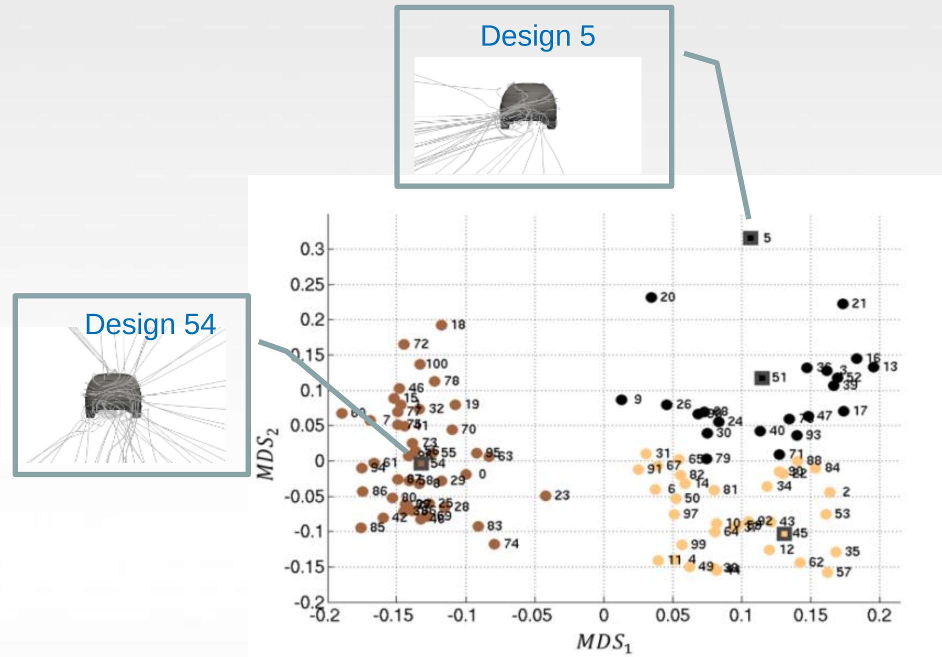
- hyper volume measures the quality of the concept in an multi-objective sense
- coverage measures how many designs in my data relate to the design concept
- support measures how many of the designs that belong to my concept relate to the similar objectives

A: design concept C: objective space D: simulated data

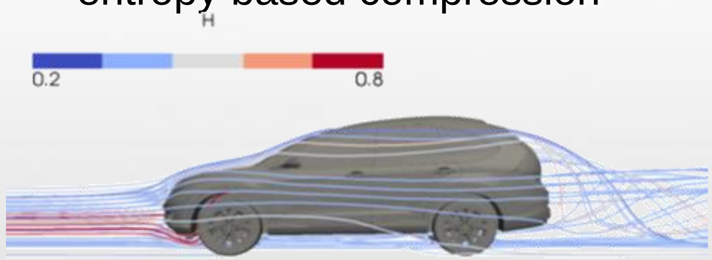
Flow Field based Analysis with Compressed Data

- Flow field analysis based on the quantification of streamline similarities

  - Visualization & analysis of flow field similarities



  - entropy based compression



  - Flow field retrieval

λ	$\lvert L_Q \rvert$	Best matches id	Time fraction	Correct matches
1.0	*2450*	*77,73,95*	*100.0%*	*100.0%*
0.9	423	77,73,95	≈22.0%	90.0%
0.7	264	77,73,95	≈ 17.8%	90.0%
0.5	161	77,95,73	≈12.5%	50.0%
0.2	46	95,77,73	≈2.5%	10.0%

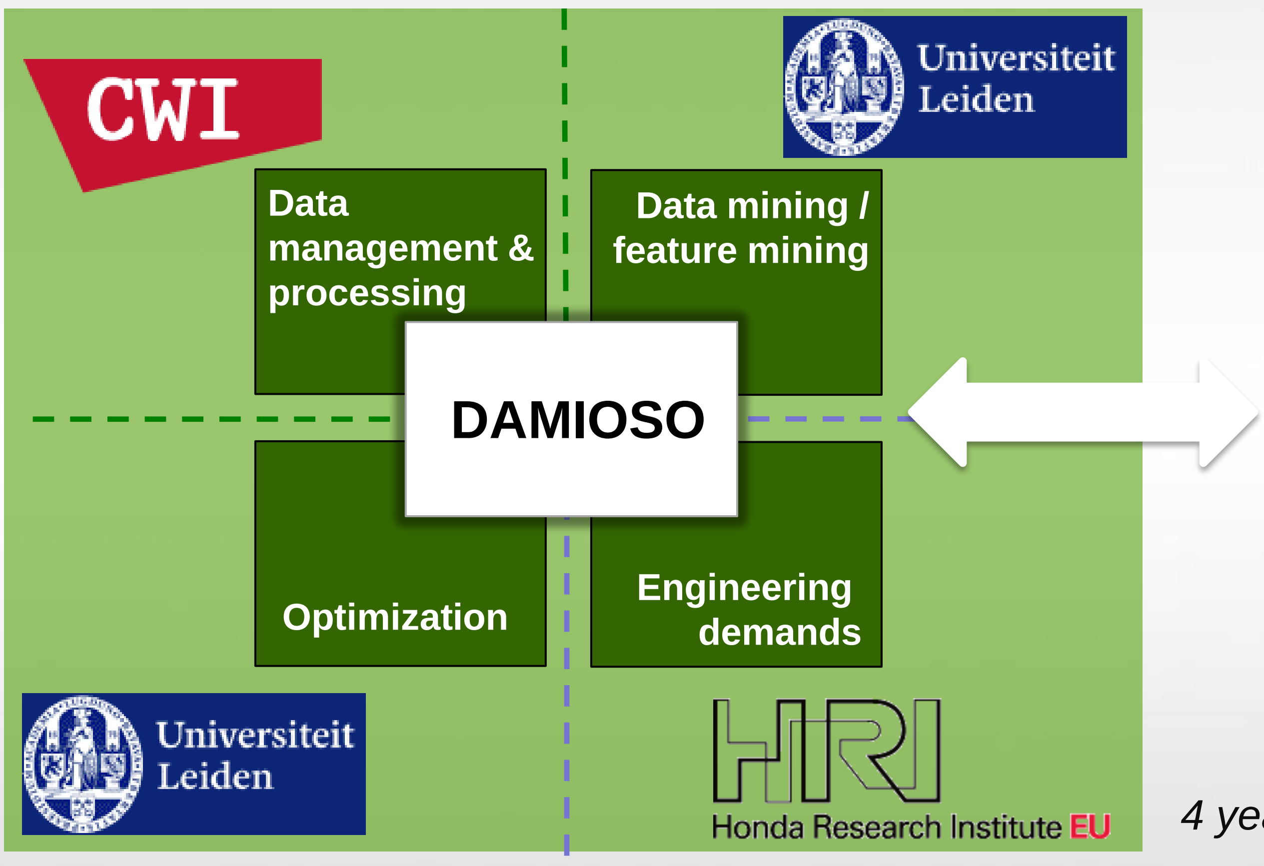
same best matches for reduction to ~10% of stream lines, with only 18% of the computation time compared to the raw stream line set

Summary

- Engineering Data Analytics (EDA) targets the analysis of data from different design processes, dfferent disciplines and representations
- challenges are the data size and the heterogeneity of the data
- intelligent data compression and a good choice of a unified representation are prerequisites for a successful data analysis process that augments engineering knowledge
- interacton analysis and concept learning can provide new design insight and measure higher order correlations between design variations
- flow field analysis provide important and more comprehensive information on the design process
- interaction and utility can equally be used for combining different domains
- EDA is not restrcited to the design process, it also includes e.g. test scenarios with variations to the sensor set-up,environmental variations, different drivers, scenarios

DAMIOSO – Data Mining on High Volume Simulation Output

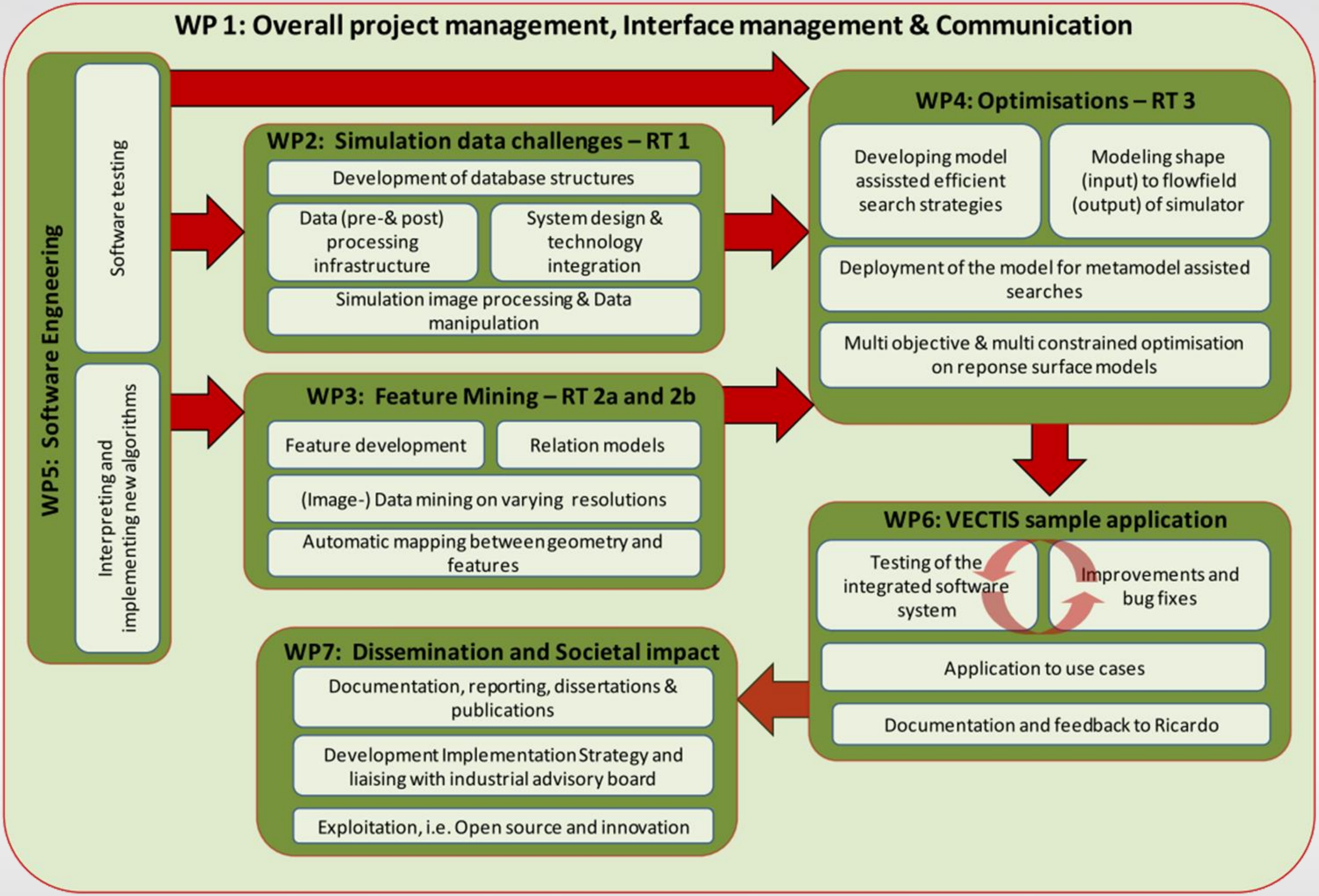
- Algorithms for managing, mining, learning and optimizing based on massive data generated by computational simulations
- Combining optimization and simulation technologies under the scope of massive data and long simulation time



Industry User Committee

4 year project funded by the NWO

DAMIOSO


WP 1: Overall project management, Interface management & Communication
WP5: Software Engineering
Software testing
Interpreting and implementing new algorithms
WP2: Simulation data challenges – RT 1
Development of database structures
Data (pre-& post) processing infrastructure
System design & technology integration
Simulation image processing & Data manipulation
WP3: Feature Mining – RT 2a and 2b
Feature development
Relation models
(Image-) Data mining on varying resolutions
Automatic mapping between geometry and features
WP4: Optimisations – RT 3
Developing model assissted efficient search strategies
Modeling shape (input) to flowfield (output) of simulator
Deployment of the model for metamodel assisted searches
Multi objective & multi constrained optimisation on reponse surface models
WP6: VECTIS sample application
Testing of the integrated software system
Improvements and bug fixes
Application to use cases
Documentation and feedback to Ricardo
WP7: Dissemination and Societal impact
Documentation, reporting, dissertations & publications
Development Implementation Strategy and liaising with industrial advisory board
Exploitation, i.e. Open source and innovation

Thanks to

S. Menzel,
M. Olhofer,
CSOA team @ HRI

colleagues from
Honda R&D Americas
Honda R&D Japan

partners at
University Leiden (T. Bäck)
University Birmingham (X. Yao)
University Surrey (Y. Jin)