Advanced Mode Analysis
for Crash Simulation Results

Clemens-August Thole, Lialia Nikitina, Igor Nikitin, Tanja Clees

Fraunhofer Institute for Algorithms and Scientific Computing SCAI, Sankt Augustin, Germany

{Clemens-August.Thole,Lialia.Nikitina,Igor.Nikitin,Tanja.Clees}@scai.fraunhofer.de

Abstract:

Potential scatter of simulation results caused, for example, by buckling, is still a challenging issue for predictability. Principle component analysis (PCA) and correlation clustering are well-known mathematical methods for data analysis. In order to characterize scatter, methods of these types were applied to the ensemble of simulation results resulting from a number of runs using all node positions at all time steps. For industrially relevant problems, the size of the resulting data base is larger than 100 GBytes (even if compressed by FEMzip[1][7]) . As a result of applying the methods, the major components and influences dominating the differences between the simulation results are available.

PCA is a mathematical method which treats data bases globally, without built-in a-priori, physical knowledge. The selected modes do not separate different physical effects like buckling at different parts of the model. PCA rather tries to maximize the variations by combining several physical effects into one mode.

Difference PCA (DPCA) applies PCA analysis to the results for each part and time step. By analysis of the related covariance matrices, the local dimension of the scatter subspace can be identified and correlation between the scatter at different places can be analyzed. Using DPCA, different origins of scatter can be identified and physically meaningful components can be determined. DPCA, however, should be combined with clustering methods, for instance, in order to preprocess the data base.

Correlation clustering is an efficient method for the identification of strongly correlated components in bulky data. This method can be applied for causal analysis of large ensembles of even highly resolved crash simulation results. It groups all strongly correlated data items together, separates the model to a few such clusters and represents a structure of correlators in the model at a glance. The resulting so-called partition diagram and a series of appropriate visualizations can be used for analysis directly. Computed clusters can be used for DPCA, in addition.

The paper introduces the approaches and shows first results for an industrial model.

Keywords:

crash analysis, statistical analysis, correlation, clustering, mode analysis, principal component analysis

---

[1] FEMzip is a registered trademark of Fraunhofer Gesellschaft, Munich

1   Introduction

In [8] and [4] scatter of the Chrysler Neon model and a model of BMW were analyzed in detail. Buckling and contact heuristics were identified as major sources for scatter. For the design and optimization of car models it is very helpful to deal with a simulation model, which generates similar results even if slight changes of the model are performed. In order to investigate reasons for scatter several optimization tools like LS-OPT$^{TM\,2}$ [6] support the generation of variants as well as the analysis of scatter of key results (like intrusion). In addition, LS-OPT also supports correlation analysis between changes of model parameters and field results (like displacements at various nodal points) and various functions to visualize correlations and scatter of simulation results. In [4], new methods using correlation between displacements of nodes were introduced, which allow to find the origin of scatter in crash simulation. All these methods work if there is a single dominating source of scatter, but fail to separate the impact of several sources. In addition, handling of data bases consisting of several hundreds of GBytes is at least not efficiently realized by these methods, if possible at all.

2   PCA Analysis for Crash Simulation Results

According to [3], principle component analysis (PCA) was introduced by Pearson in the context of biological phenomena [5] and by Karhunen in the context of stochastic processes [2].

In [1], PCA was applied to full crash simulation results. Let $X_i(p,t)$ be the displacement of simulation run i out of n simulation runs at node p and time t. If $\overline{X}(p,t)$ is the mean of all simulation runs, the covariance matrix C can be defined as

$$C := \left[c_{i,j}\right]_{1 \leq i,j \leq n} \text{ and } c_{i,j} := \,< X_i - \bar{X}, X_j - \bar{X} >_2$$

The eigenvectors v_i of C form a new basis (principle components) and the λ_i (squareroots of the eigenvalues of C) provide a measure for the importance of each component.

If this method is applied to crash simulation results, n^2 scalar products between the simulations runs of length $3 * \#P * \#T$ have to be computed ($\#P$ number of points, $\#T$ number of time steps.)

From $\hat{X}(a) := \sum_{i=1}^{n} a_i X_i$,

follows that $\lambda_i = \left\|\hat{X}(v)\right\|_2$.

The $\hat{X}(v_i)$ show the major trends of the differences between the simulation results. The coefficients of the eigenvectors v_i correspond to the contribution of $\hat{X}(v_i)$ to $X_i - \bar{X}$ and can be used for cluster analysis and correlation with input parameters. If input parameters have been changed between the different simulation runs, the correlation analysis will indicate how certain trends can be avoided or increased by changing these inputs (e.g. thicknesses of parts) (c.f.[3], Chapter 2.4] for the properties of PCA analysis in general).

Principle Component Analysis is a mathematical method which determines mathematical trends in contrast to physical trends. To be more specific: λ, the square of the maximal eigenvalue of C, can be determined by

$$\lambda = \left(\max_{v}\left\|\hat{X}(v)\right\| \,\middle|\, \|v\| = 1\right)$$

and therefore will be in general a mixture out of several physical effects, like buckling.

---

^{2} LS-OPT is a registered trademark of Livermore Software Technology Corporation.

3   Difference PCA

Instead of considering the whole simulation results, correlation matrices can also be defined for the simulation results at parts of the model and for specific time steps. If P is a part of the model and T subset of the time steps, then $C_{P,T}$ can be defined as follows:

$$C_{P,T} := \left[c_{i,j}^{P,T}\right]_{1 \leq i,j \leq n} \text{ and } c_{i,j}^{P,T} := \frac{1}{N_{P,T}} \sum_{p \in P, t \in T} \left(X_i(p,t) - \bar{X}(p,t)\right) * \left(X_j(p,t) - \bar{X}(p,t)\right).$$

($N_{P,T}$ denotes the size of P times the size of T.)

The intrinsic dimension of the set of simulation results can be defined as the number of major components in its differences (for more formal definitions see [3], Chapter 3]). Buckling or any other local instability in the model or numerical procedures increase the intrinsic dimension of simulation results at parts which are affected compared to those, which are not affected. Therefore in the context of stability of crash simulation, those parts and time steps for which the intrinsic dimension increases are of particular interest.

Numerically this can be evaluated by determining eigenvectors and eigenvalues of

$$C_{P_1,T_1} - \tau \, C_{P_2,T_2}$$

for the covariance matrices of the simulation results at two different parts P_1 and P_2 and two different sets of time steps T_1 and T_2. If there are positive eigenvalues for a certain choice of τ (which separates noise from real signals), the simulation results at (P_1, T_1) show additional effects compared to those at (P_2, T_2). If v_{P_1,T_1} is the corresponding eigenvector, $\hat{X}(v_{P_1,T_1})$ shows the effect on (P_1, T_1) and also the impact on the other parts of the model. Similar methods can be used to remove those effects from this result, which do not affect (P_1, T_1) directly.

This approach has been filed for application of a Patent at the German Patent office (DPMA number 10 2009 057 295.3) by Fraunhofer Gesellschaft, Munich.

4   Correlation clustering

Causal analysis is generally performed by means of statistical methods, particularly by estimation of correlation of events. It is commonly known that correlation does not imply causation (this logical error is often referred as "cum hoc ergo propter hoc": "with this, therefore because of this"). Instead, strong correlation of two events does mean that they belong to the same causal chain. Two strongly correlated events either have direct causal relation or they have a common cause, i.e. a third event in the past, triggering these two ones. This common cause will be revealed if the whole causal chain, i.e. a complete sequence of causally related events, will be reconstructed.

For this purpose, one can use the following standard statistical description:

Input: experimental data matrix x_{ij}, i=1..N_{data}, j=1..N_{exp}. Every column in this matrix forms one experiment, every row forms a data item varied in experiments.

Then every data item is transformed to a z-score vector [9]:

$$\langle x_i \rangle = \Sigma_j \, x_{ij} \, / \, N_{exp}$$
with $dx_{ij} = x_{ij} - \langle x_i \rangle$ and $|dx_i| = sqrt(\Sigma_j \, dx_{ij}^2)$ and $z_{ij} = dx_{ij}/|dx_i|$

or by means of the equivalent alternative formula

$$z_{ij} = dx_{ij} / (rms(x_i) \, sqrt(N_{exp})),$$

with $rms(x_i) = sqrt(\Sigma_j \, dx_{ij}^2 / N_{exp})$. Here $\langle x_i \rangle$ is the mean of the i-th data item, dx_{ij} is the deviation from the mean of the i-th data item, and rms(xi) is the root mean square deviation of the i-th data item.

In this way, data items are transformed to N_{data} vectors in N_{exp}-dimensional space. All these z-vectors belong to an $(N_{exp}-2)$-dimensional unit-norm sphere, formed by intersection of a sphere $|z|=1$ with a hyperplane $\Sigma_j\, z_{ij}=0$.

An important role of this representation is the following. All z-vectors which are located close one to the other correspond to strongly correlated data items. Two such data items being displayed on 2D plot will show a linear dependence between the items. This plot is coincident up to a scaling with the plot of one z-vector versus the other, which is concentrated near $z_1=z_2$ due to approximate coincidence of z-vectors. Also in the case when two z-vectors are approximately opposite, they correspond to strongly correlated data items and linear data plot, approximately coincident with $z_1=-z_2$, up to a scaling.

The scalar product of two z-vectors is equal to Pearson's correlator since

$$(z_1,z_2) = \Sigma_j\, z_{1j}\, z_{2j} \quad \text{and} \quad (z_1,z_2) = corr(x_1,x_2)$$

It is also related with the angular distance of the z-vectors on the sphere:

$$(z_1,z_2)=\cos\theta\,(z_1,z_2),$$

Pearson's correlator, often used as a measure of correlation, indicates strong correlation in the limit of $corr\rightarrow1$, corresponding to $z_1\rightarrow z_2$, $\theta\rightarrow0$ or in the limit $corr\rightarrow-1$, corresponding to $z_1\rightarrow-z_2$, $\theta\rightarrow\pi$. The sign of Pearson's correlator shows the direction of dependence of data items: in the case $corr=1$, z_1 increases with increase of z_2, in the case $corr=-1$, z_1 decreases with increase of z_2. If not the direction of dependence is of interest, but only the fact of such dependence, one can formally consider a sphere of z-vectors as projective space, where opposite points are glued together and are considered as identical points. In this space approximately opposite vectors become approximately coincident.

The task of correlation clustering is to group all strongly correlated data items in one cluster and subdivide all data items to a few such clusters. This allows to identify the parts of data which behave similarly and to reduce the number of independent degrees of freedom of the system to few cluster centers. Clustering provides a compact representation of experimental data, suitable for a direct analysis by the user.

Correlation clustering becomes straightforward in z-space representation: one only needs to group together closely located z-vectors with any suitable general clustering technique. Many efficient clustering techniques are known to be applicable to this problem [10]:

- density-based clustering
- k-means clustering
- dendrogram clustering
- spectral clustering

Details of these techniques in application to correlation clustering are presented in [11]. Most of them possess as a control parameter a cluster diameter diam related to Pearson's correlator by the formula

$$diam=arccos(corr).$$

While spectral clustering is equivalent to global PCA, the other clustering techniques, e.g. k-means, do not force orthogonality of the components and allow to identify generic clusters with non-orthogonal center positions.

This approach is subject of an international patent application (PCT/EP2010/061439) by Fraunhofer Gesellschaft, Munich.

5   Example

The new methods were tested on the publically available Ford Taurus test case (version with 854692 elements) as provided by [12]. Simulations were performed by LS-DYNA[TM,3]. The thicknesses of all parts were changed by means of an experimental design (design-of-experiment, DoE) consisting of 100 simulation runs.

The corresponding ensemble was used for all three PCA-type analysis steps, namely global, local and difference PCA. In particular, DPCA profits from larger ensembles, together with a preprocessing (by means of correlation clustering) yielding an appropriate selection of target clusters.

Correlation clustering can handle the whole ensemble efficiently. However, if parameter dependencies are not to be analyzed, as is the case here, a quarter of the original ensemble - in general, approximately 25 runs only - is sufficient to obtain statistically meaningful results.
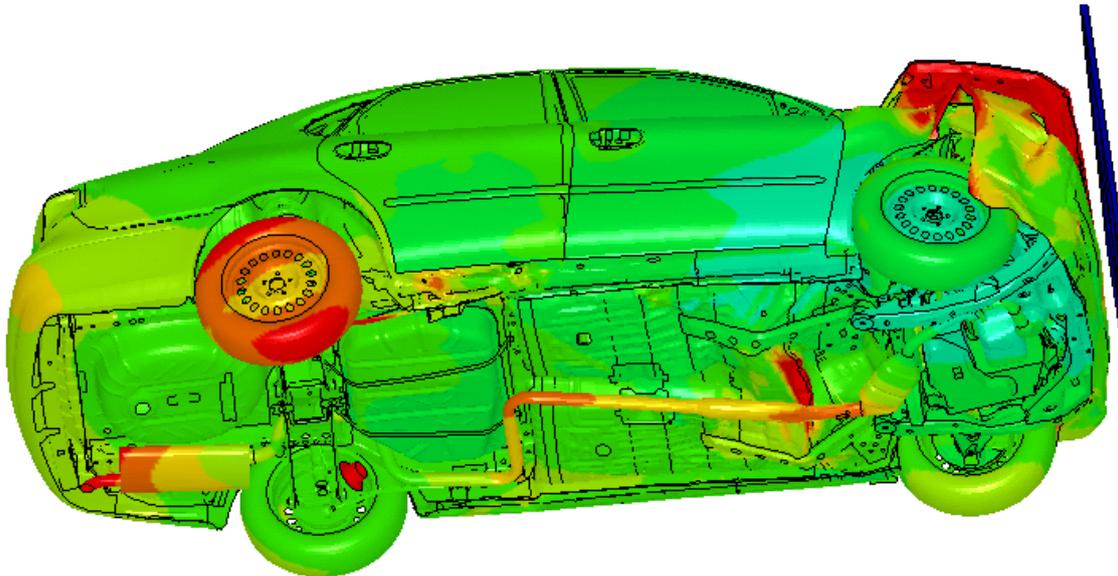
Figure 1: Maximal local difference between simulation results as color coding computed by DIFF-CRASH [4] (range 0-100 mm)

Figure 1 shows the maximal scatter between the simulation results computed by DIFF-CRASH[4], a tool for the statistical analysis of crash-simulation results. It shows substantial scatter for example in the floor section.

5.1   Results with global PCA

Figure 2 shows the results of the principle component analysis. The outliers of $log(\frac{\lambda_i}{\lambda_{i-1}})$ show, that there are a few dominating components starting at component 80. Most interesting are the two last components, which clearly separate from the others.

The fringe plots on Figure 3 show the two most important components at the final state. Both cover almost the whole car. The contributions of each simulation run to the most important component are directly correlated with the thicknesses of the group of remaining parts, which have been changed simultaneously. The coefficients of the second component are not correlated (even if non-linear basis functions are used) with any of the thickness variation.

---

[3] LS-DYNA is a registered trademark of Livermore Software Technology Corporation.
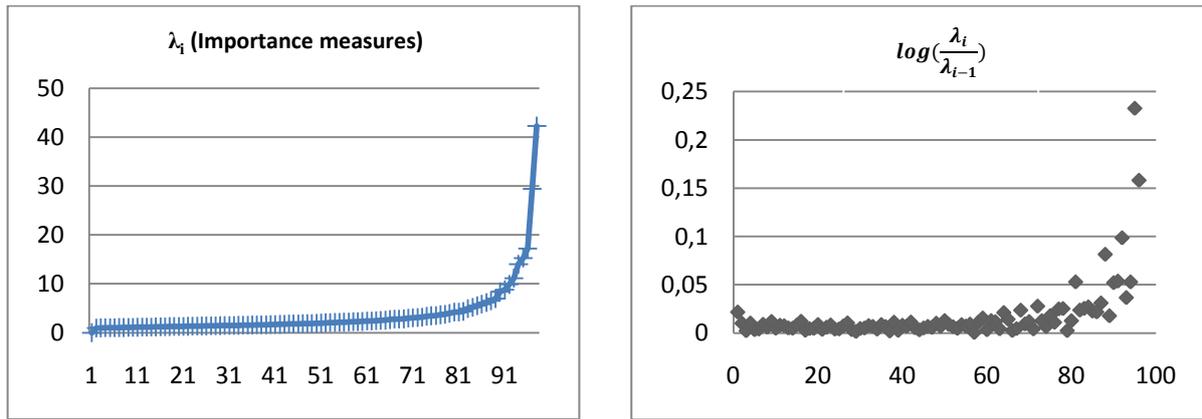[4] DIFF-CRASH is a registered trademark of Fraunhofer Gesellschaft, Munich

Figure 2: PCA analysis for full car model: Plots of importance measures for the components as well as their relative growth
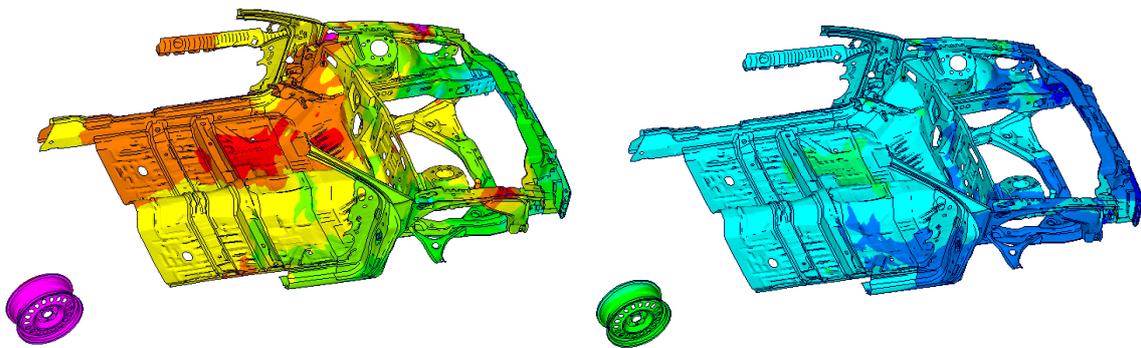

Figure 3: Fringe plots of the two most important components from global PCA on a selected number of parts
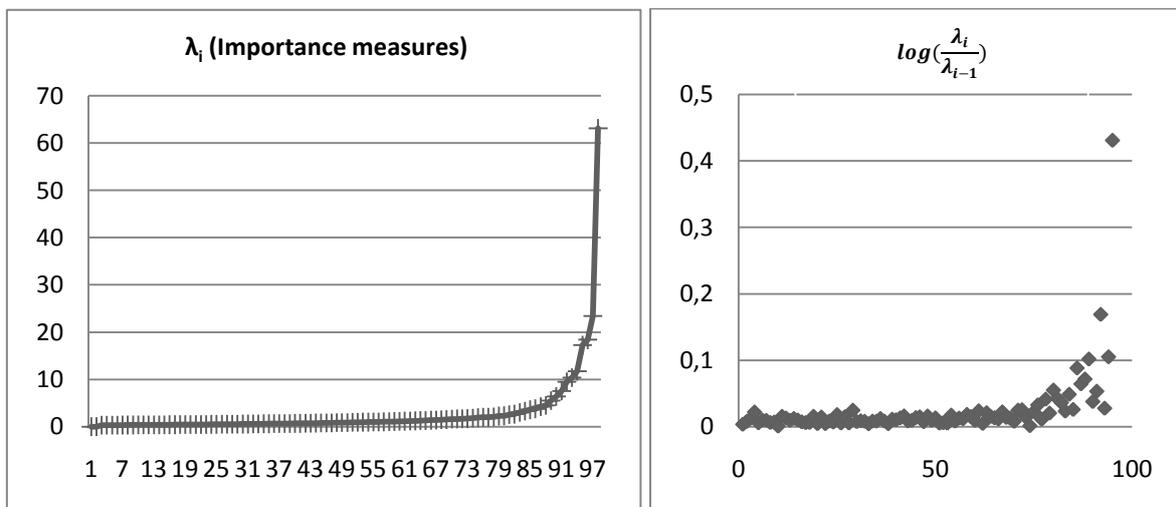

Figure 4: PCA analysis for part 2000085 at time step 150 ms: Plots of importance measures for the components as well as their relative growth

5.2   Results with local PCA for spitwall (Part 2000085)

In a second step the PCA analysis has been applied only to the spitwall at time step 150 ms. The intrusion of the spitwall is one of the design criteria for the safety of the car and therefore of special interest for the design engineer.

Figure 4 shows that 4 components are dominating the differences of the simulation runs at the spitwall.

Figure *5* shows the most important two components as fringe plots for their displacement on the geometry of the car. The most important component is very similar to that of the global analysis. The second component shows buckling at the floor section which seems to influence the spit wall at the 150 ms.
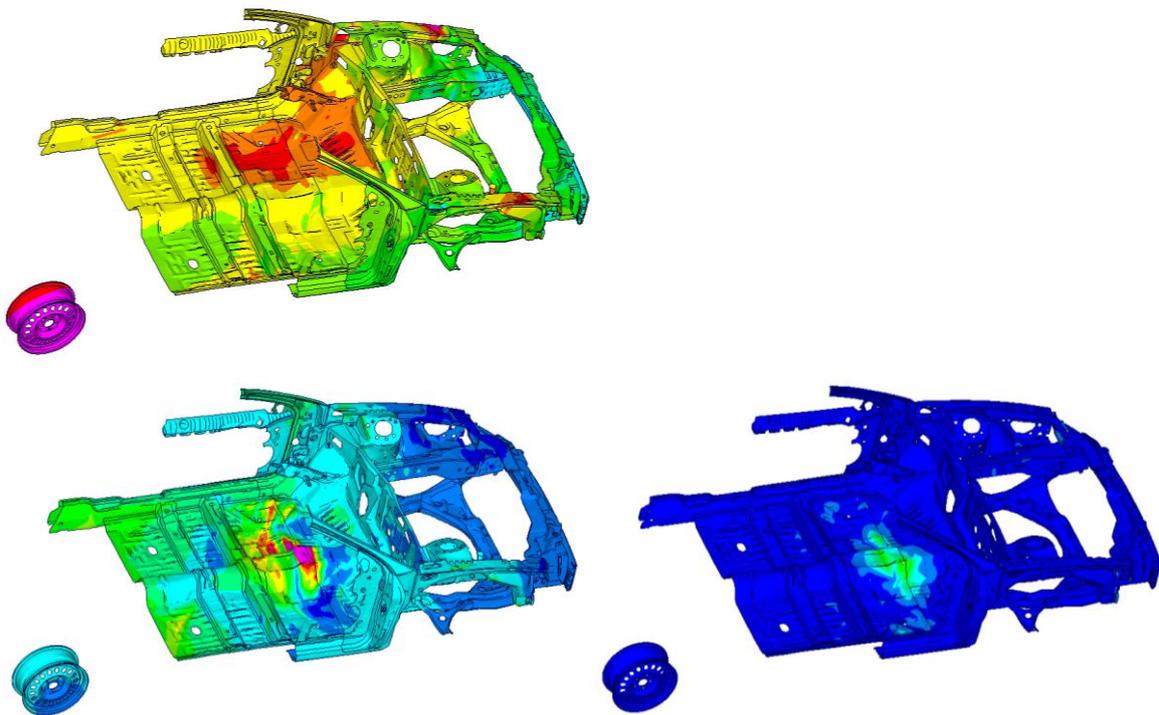

Figure 5: Fringe plot of the displacements of the two most important components from the local PCA analysis of the spitwall (Part 2000085) at 150 ms and for the second component also at 40 ms (Range 0-30mm)

5.3   Results with DPCA for spitwall (Part 2000085)

In a final step, DPCA was applied to the spitwall and the floor section. Using this analysis it turned out, that the buckling at the floor section has almost no impact on the scatter at the spitwall (see Figure 6). Performing a DPCA analysis of the spitwall with all other Parts and time steps buckling of the Part 2000014 (longitudinal rail) between time step 6 and 7 is very important for the behavior of the spitwall.

Figure 7 shows the dominating component at 30 ms and 150 ms. At 30 ms it is an almost isolated effect at the longitudinal rail which affects later almost all parts.
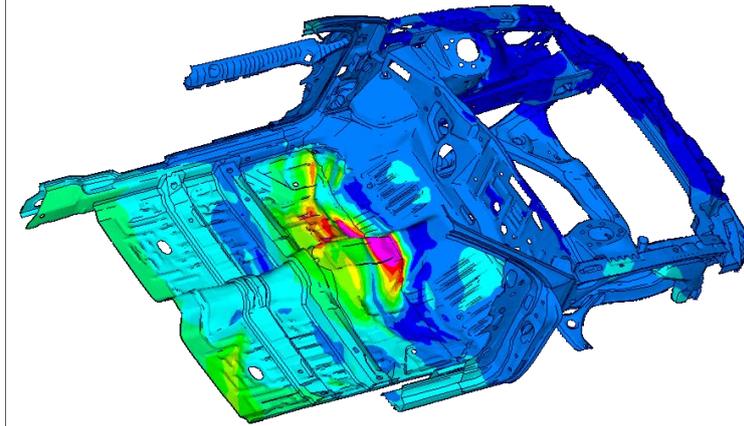
Figure 6: Fringe plot of the most important component of the DPCA analysis of the floor section and the spitwall.
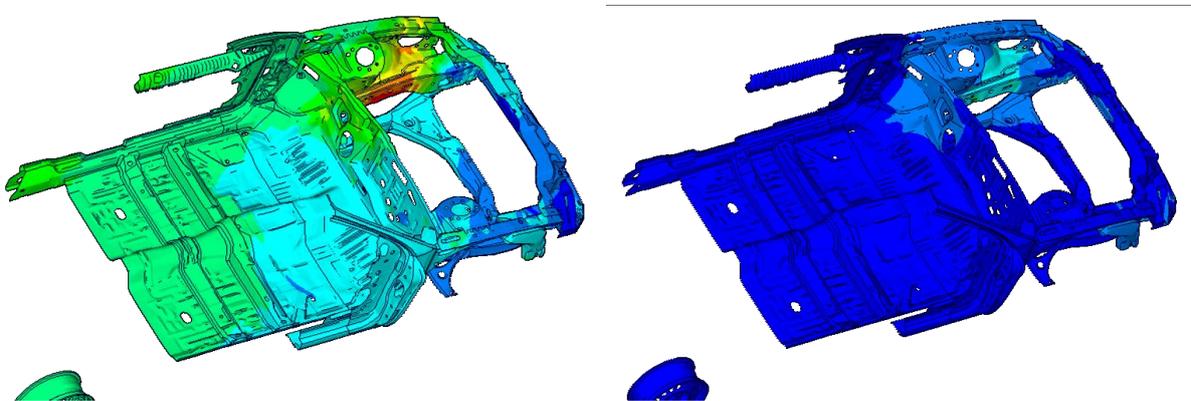

Figure 7: Fringe plot of the most important component of the DPCA analysis of spitwall and longitudinal rail after a clean-up with other scatter candidates.

5.4   Results with z-score k-means correlation clustering

The partition diagram, depicted in Figure 8, shows at a glance the distribution of correlators in the model. The following correlation clusters have been detected:

- The red cluster appears at 25ms on the driver's door and then propagates to the right side of the car.
- The yellow cluster appears at 35ms in the place of connection of a part 2000610 and the longitudinal rail 2000014, and then propagates to the spitwall 2000085.
- The green cluster starts on the bumper at the moment of the crash contact with the barrier and extends over the motor hood.
- The cyan cluster corresponds to a buckling of the bottom longitudinal structure 2000096 at 70ms; the influence first propagates through the driver's seat and then extends to the whole left-rear side of the car.
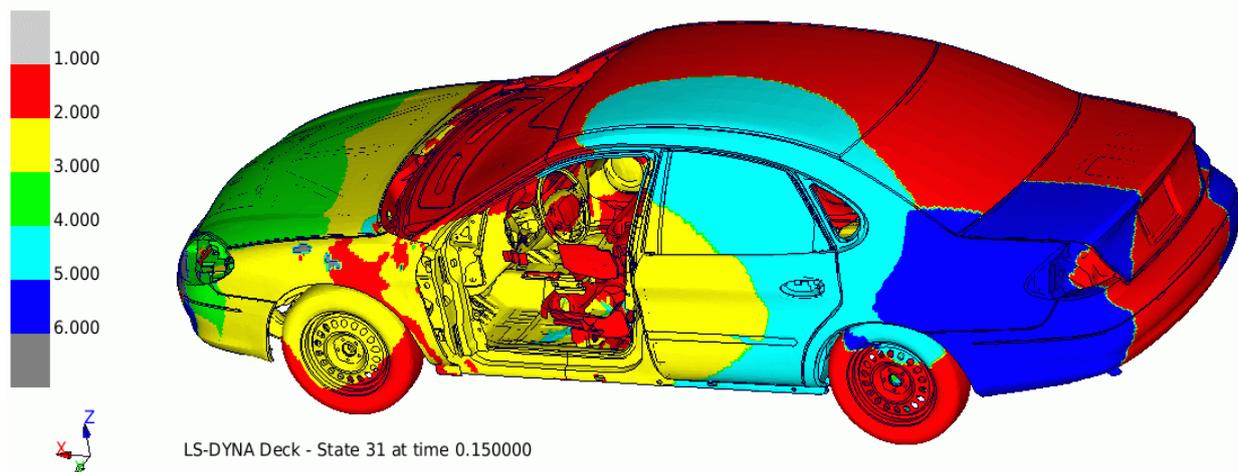- The blue cluster appears at 55ms on the left rear wheel and occupies the rear part of the car.

Figure 8: Results of correlation clustering in form of partition diagram.

6   Summary and Outlook

Local principle component analysis and DPCA were introduced to the analysis of optimization and stability of crash simulation results. For the most important component, a clear correlation with the simultaneous change of the thickness of a group of parts could be identified. DPCA turned out to be useful for the identification of the most important buckling component influencing the spitwall as a secondary effect.

We have also introduced correlation clustering, an efficient method for identification of strongly correlated components in bulky data, and applied it for causal analysis of crash simulation results. This method allows to group together all strongly correlated data items, to separate the model to a few such clusters and to represent a structure of correlators in the model at a glance by means of the partition diagram, for instance.

First efficient software implementations of the methods as well as necessary data management procedures are available. Future work shall be concerned with several aspects such as automatic workflows involving the methods presented for stability analysis, combinations with developments described in [13] and automatic workflows for parameter studies for processes or process chains as well as on enhancements of the methods for non-linear data analysis.

7   Acknowledgments

8   References

[1]   S. Ackermann, L. Gaul, M. Hanss, T. Hambrecht. Principal component analysis for detection of globally important input parameters in nonlinear finite element analysis. In Optimisation and Stochastic Days 5.0. dynardo - dynamic software&engineering, Weimar, 2008. http://www.dynardo.de/bibliothek/wost/wost-50/

[2]   K. Karhunen. Zur Spektraltheorie stochastischer Prozesse. Ann. Acad. Sci. Fennicae, 34, 1946.

[3]   M. Lee, J. Verleysen. Nonlinear Dimension Reduction. Information Science and Statistics. Springer Science+Business Media, 2007.

[4]   L. Mei, C.-A. Thole. Data analysis for parallel car-crash simulation results and model optimization. Simulation modelling practice and theory, 16(3):329–337, 2008.

[5]   K. Pearson. On lines and planes of closest fit to systems of points in space. Philosophical Magazine, 2:559–572, 1901.

[6]   W. Stander, N. Rouw. New features in ls-opt version 3. In Wayne L. Mindle, editor, 9th International LS-DYNA Users Conference 2006. Livermore Software Technology Corporation, Livermore California, 2006. http://www.dynalook.com/international-conf-2006/2Optimization.pdf.

[7]   C.-A. Thole. Compression of ls-dyna simulation results. In Procs. 5th European LS-DYNA user conference, Birmingham., 2005. http://www.dynalook.com/european-conf-2005/Modelling_II/Thole.pdf.

[8]   C.-A. Thole, L. Mei. Reason for scatter in simulation results. In Procs. 4th European LS-Dyna User Conference. Dynamore GmbH, Germany, 2003. http://www.dynalook.com/european-conf-2003/crash-automotive-applications/reasons-for-scatter-in-crash-simulation-results.

[9]   R.J. Larsen, M.L. Marx. An Introduction to Mathematical Statistics and Its Applications, Prentice Hall 2001.

[10] P.-N. Tan, M. Steinbach, V. Kumar. Introduction to Data Mining, Addison-Wesley, 2005.

[11] I. Nikitin, L. Nikitina, T. Clees. Stochastic analysis and nonlinear metamodeling of crash test simulations and their application in automotive design, to appear in: Computational Engineering: Design, Development and Applications, Ed.: F.Columbus, Pub.: Nova Science, New York 2010.

[12] NCAC, http://www.ncac.gwu.edu/vml/models.html

[13] T. Clees, D. Steffes-lai. PRO-CHAIN - Efficient Statistical Analysis of Process Chains Applied to a Forming-to-Crash Example. In Procs 9th LS-DYNA Forum, Bamberg, Germany, Oct 12-13, 2010. Dynamore GmbH.