Comparison of clustering techniques for hybrid rocket fuel combustion data

A. Rüttgers[a], C. Debus[a], M. Siggel[a], A. Petrarolo[b] and M. Kobald[b]

[a] *German Aerospace Center (DLR), Simulation and Software Technology, High-Performance Computing Department, Linder Hoehe, 51147 Cologne, Germany*

[b] *German Aerospace Center (DLR), Institute of Space Propulsion, Propellants Department, Im Langen Grund, 74239 Hardthausen, Germany*

Clustering techniques were applied to hybrid rocket combustion tests to better understand the complex flow phenomena. Novel techniques such as hybrid rockets that allow for cost reductions of space transport vehicles are of high importance in space flight. However, the combustion process in hybrid rocket engines is still a matter of ongoing research and not fully understood yet. Recently, combustion tests with different paraffin-based fuels have been performed at the German Aerospace Center (DLR). For a better understanding of the experiments, the combustion process has been captured with a high-speed video camera, which leads to a huge amount of images for each test. In order to catch the essential flow structures, the combustion dataset has been analyzed with unsupervised machine learning techniques.

In this talk, we compare the outcome of different clustering techniques, e.g. centroid models, density models, and spectral clustering, on the test data. Depending on the approach that is considered, the algorithms either tend to resolve long-running flow phases [1] (e.g. centroid models) or reveal short-term turbulence effects [2] (e.g. spectral clustering). Since both information are essential to characterize the fluid flow, it is therefore advantageous to combine different approaches. As a result, valuable insights into the different combustion phases were obtained and a comparison of the quality of the combustion flame in the different tests could be made.

Clustering of thousands of images is very demanding with respect to computing time and memory requirement. We therefore employ HeAT [3], the Helmholtz Analytics Toolkit, for the high-performance data analysis on a parallel cluster at DLR. HeAT provides a basic infrastructure for tensor linear algebra, has an automatic differentiation algorithm, supports multi-hardware-GPU/CPU-systems and allows the implementation of the algorithms in a computationally efficient manner. Using HeAT, the computing time in the order of several days (sequential run) could be reduced to roughly one hour (parallel run).

References

[1] Rüttgers, A., Petrarolo, A., and Kobald, M., "Clustering of paraffin-based hybrid rocket fuels combustion data", *Experiments in Fluids*, Vol. 61, No. 4, 2020, https://doi.org/10.1007/s00348-019-2837-8

[2] Debus, C., Rüttgers, A., Petrarolo, A., Kobald, M., and Siggel, M., "High-performance data analytics of hybrid rocket fuel combustion data using different machine learning approaches", *AIAA SciTech Forum*, 2020, in press

[3] Krajsek, K., Comito, C., Götz, M., Hagemeier, B., Knechtges, P., and Siggel, M., "The Helmholtz Analytics Toolkit (HeAT) - A Scientific Big Data Library for HPC", *Extreme Data: Demands, Technologies, and Services. Workshop Proceedings, IAS series*, 2019, pp. 57–60.

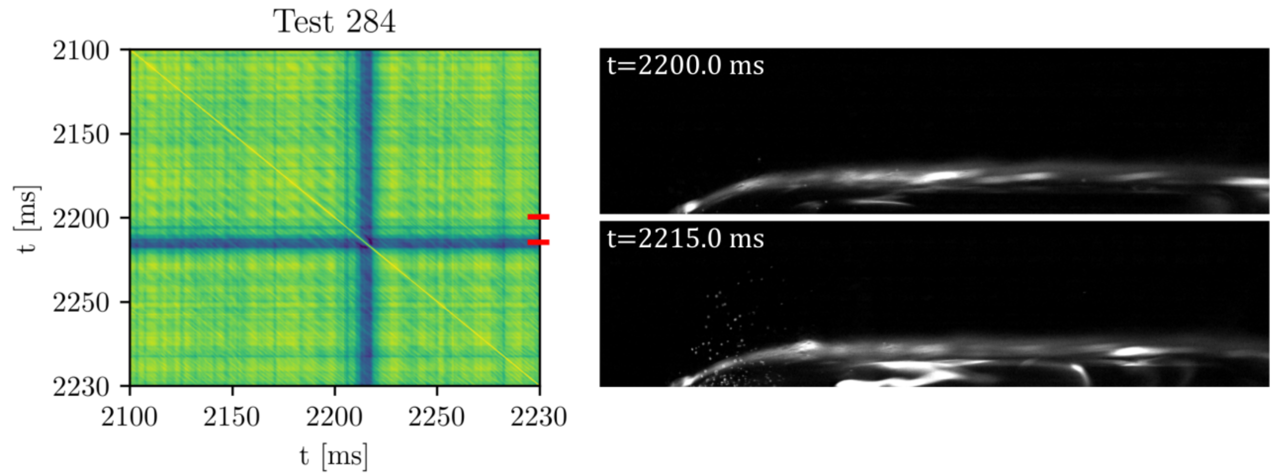

Figure 1: Spectral clustering result of test 284 with a Gaussian kernel. The enlarged view of irregularities in the similarity matrix (left) corresponds with differences in the flame appearance. The points in time of the sample images are indicated with red ticks in the similarity matrix.