Clustering of Hybrid Rocket Combustion Data

WAW Machine Learning 3

19.-20. November 2018

Alexander Rüttgers (SC-HPC)
Anna Petrarolo (RA-TRS)

Project ATEK (Antriebstechnologien und Komponenten für Trägersysteme)

- Project aim: Cost reduction of spacecraft systems by using *reusable* or *less complex* propulsion technologies

- Project inspired by private American space transport companies

- Tasks:
  - Numerical simulation and data analysis
  - Experiments / technical construction
  - Flight operation

- Participants: 8 different DLR institutes

Hybrid rockets in ATEK

- RA-TRS investigates new hybrid rocket fuels on a paraffin basis
- **Aim:** better theoretical understanding and optimization of combustion process
  - analysis of hydrodynamic instabilities
  - understand effect of the fuel‘s viscosity and surface tension on satellite droplet breakup (due to Kelvin-Helmholtz instability)
  - achieve higher regression rates / burn rates of the solid propellant
- combustion is captured with high-speed video camera
- Data analysis is performed by SC-HPC

DLR

Experimental measurements by RA-TRS

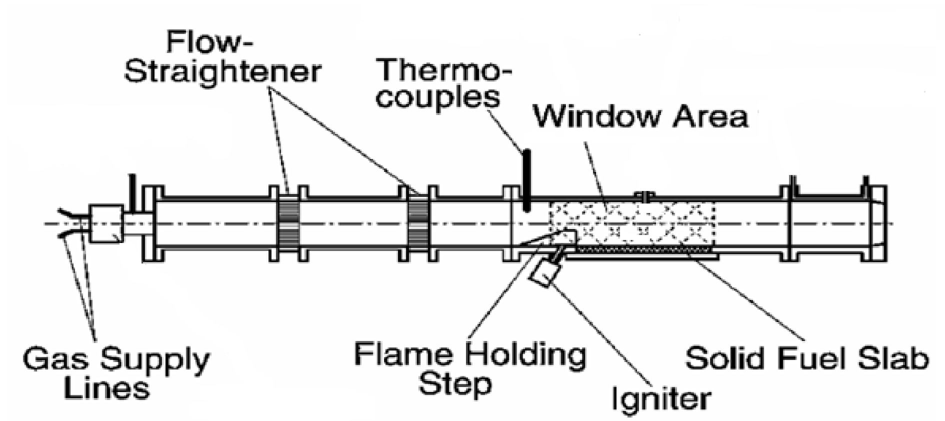


Table 2: Test matrix

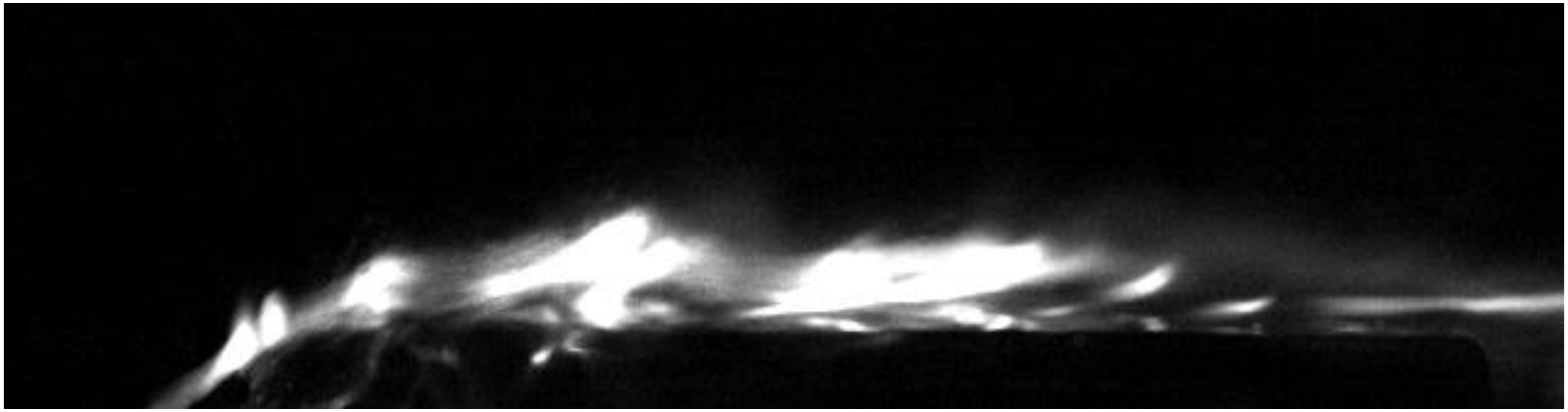
Test no.	Fuel	$\dot{m}_{Ox}[g/s]$	CH* filter
203	6805+5% polymer	50	-
234	6805+5% polymer	100	-
243	6805+5% polymer	10	-
253	6805	50	-
284	6805	50	✓

- combustion tests were performed with 4 different paraffin-based fuels in combination with gaseous oxygen

+ different fuel configurations

- video camera (1024 x 1024 pixels) captures 10 000 frames / second

Experiment 284



Video: fuel = pure paraffin
oxygen mass flow = 50 *g*/*s*,
CH*-filter (i.e. wavelengths emitted from CH* are filmed)
experiment 3 *s* = 30 000 frames / 8GB data per experiment

Clustering of data matrix S

- Cluster analysis is used to group a set of objects in such a way that the objects in the same group are similar (in some sense) to each other.
- Our aim: clustering of combustion data = identify different phases of the flow

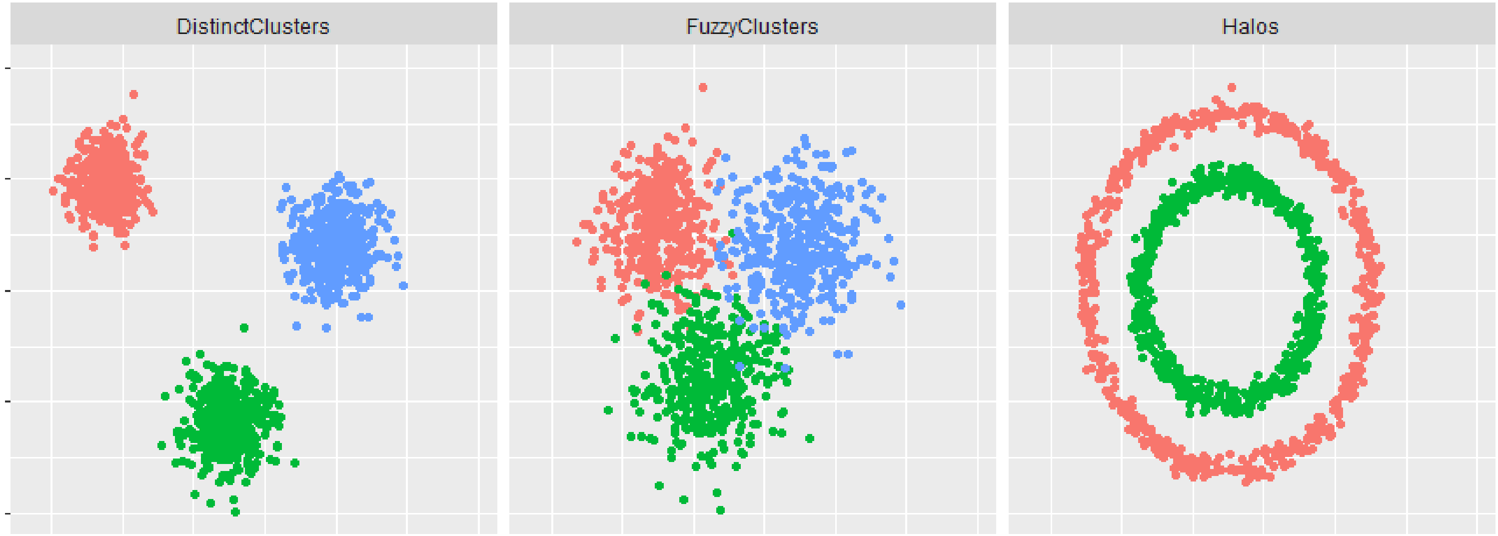


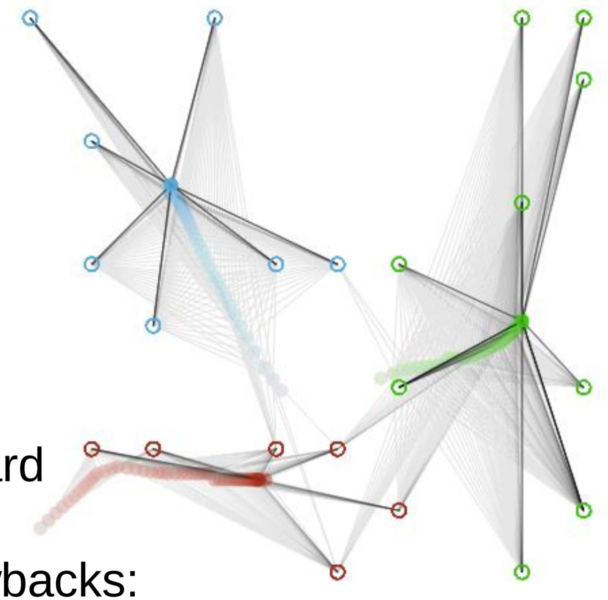
Illustration of different clusters

Centroid-based clustering (K-means++)

- Input: data matrix *S*, number of centroids *K*

- K-means iteratively minimizes an objective function

$$J = \sum_{i=1}^{K} \sum_{\boldsymbol{x}_j \in C_i} \left\| \boldsymbol{x}_j - \boldsymbol{\mu}_i \right\|^2$$



with $\boldsymbol{x}_j$ as set of observations and with $\boldsymbol{\mu}_i$ as centroids of the *K* clusters C_i

- Optimization problem itself is known to be NP-hard

- K-means is known to be robust but has two drawbacks:
  - Algorithm can converge to local optimum solutions
  - Number of clusters *K* is not directly clear

Strategies to avoid drawbacks of K-means

- **Avoid local optimum solutions**
  - Algorithm is run multiple times (here: 10-times)
  - Take solution with smallest J (not a big difference in our case)
  - Implementation of K-Means++*
    - Choose the initial centers less randomly
- **Selection of K in K-means?**
  - Detailed analysis of J depending on K (here: algorithm is used for K= 2, …, 10)
  - Runtime of algorithm at least linearly in K
  - Note that an optimal K is often problem dependent

*Arthur and Vassilvitskii. K-means++: The advantages of careful seeding. SODA ’07, 2007

Analysis of objective function *J* depending on *K*

- Evaluation function *f(K)** with *J(K)* as minimum of *J* for *K* clusters

$$f(K) = \begin{cases} 1, & \text{if } K = 1 \\ \frac{J(K)}{\alpha_K J(K-1)}, & \text{if } J(K-1) \neq 0, \forall K > 1 \\ 1, & \text{if } J(K-1) = 0, \forall K > 1 \end{cases}$$

$$\alpha_K = \begin{cases} 1 - \frac{3}{4d}, & \text{if } K = 2 \text{ and } d > 1, \\ \alpha_{K-1} + \frac{1-\alpha_{K-1}}{6} & \text{if } K > 2 \text{ and } d > 1. \end{cases}$$

where *d* is number of dimensions and α_K is a weight factor.

- For high problem dimensions $f(K) \approx J(K)/J(K-1)$ for all *K* > 1 and *f*(*K*)=1 otherwise.

*Pham, Dimov, and Nguyen. Selection of k in k-means clustering. *Journal of Mechanical Engineering Science*, 2005.

Estimate number of clusters *K*

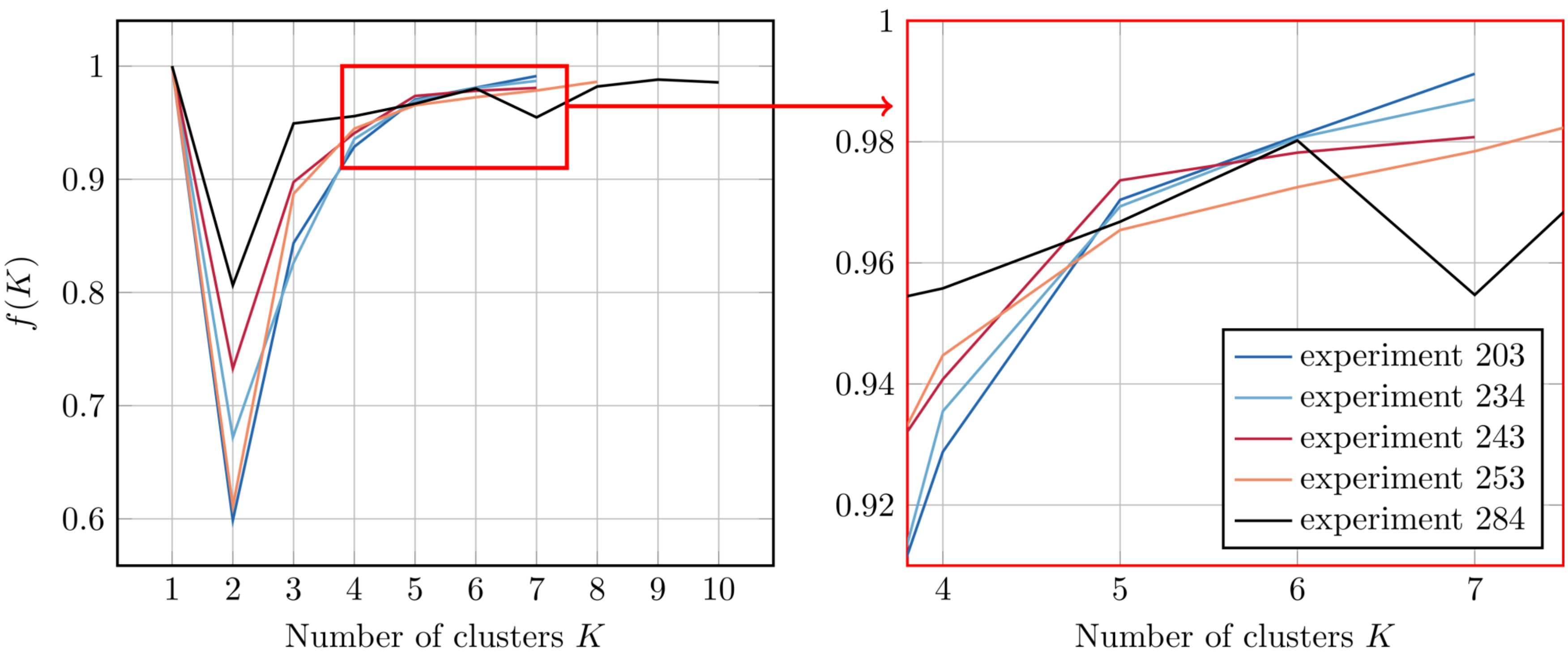


Figure: Evaluation function *f*(*K*) to determine the number of clusters *K*.

DLR

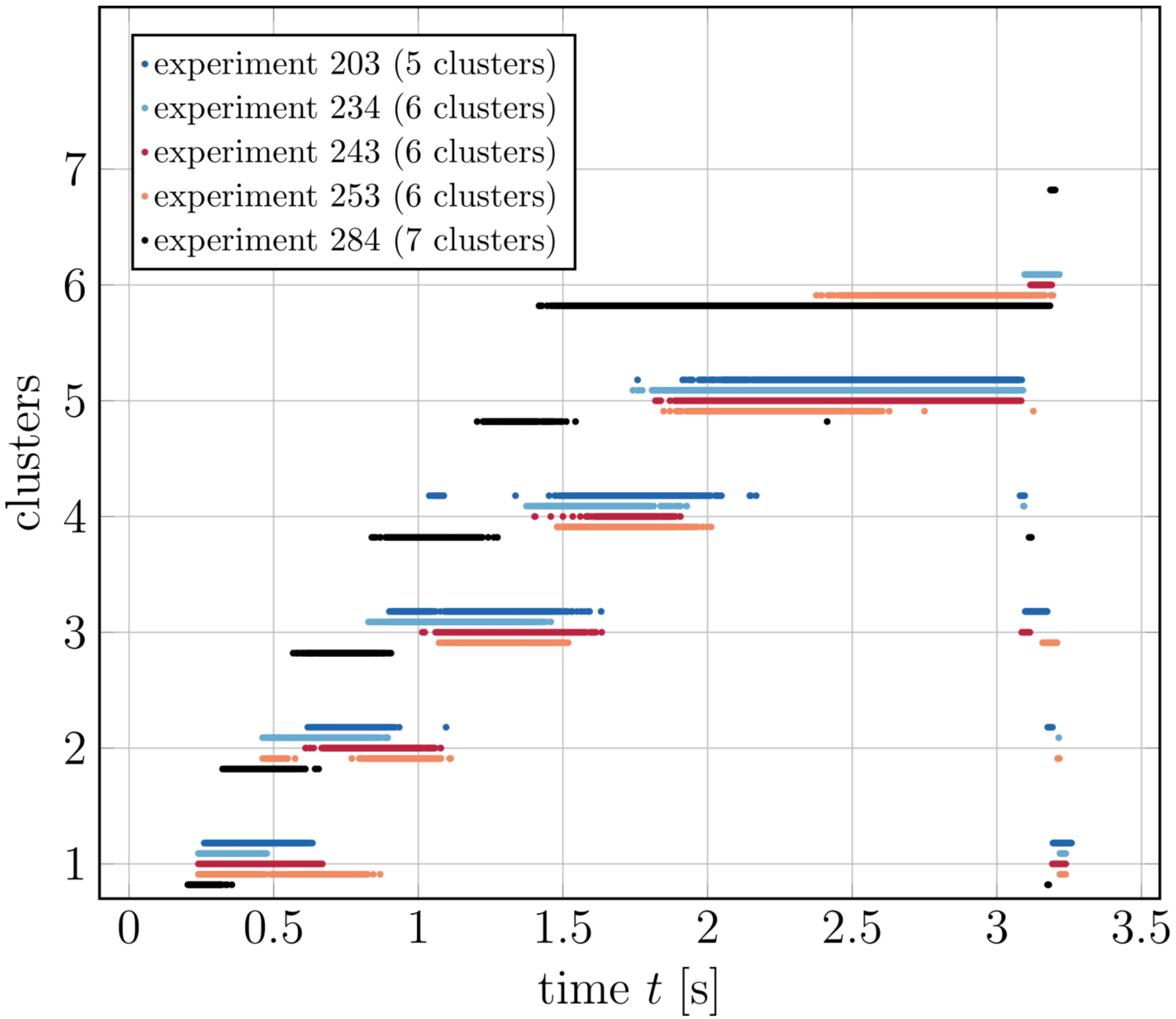


Figure: Distribution of the frames in 5 experiments to their corresponding clusters.

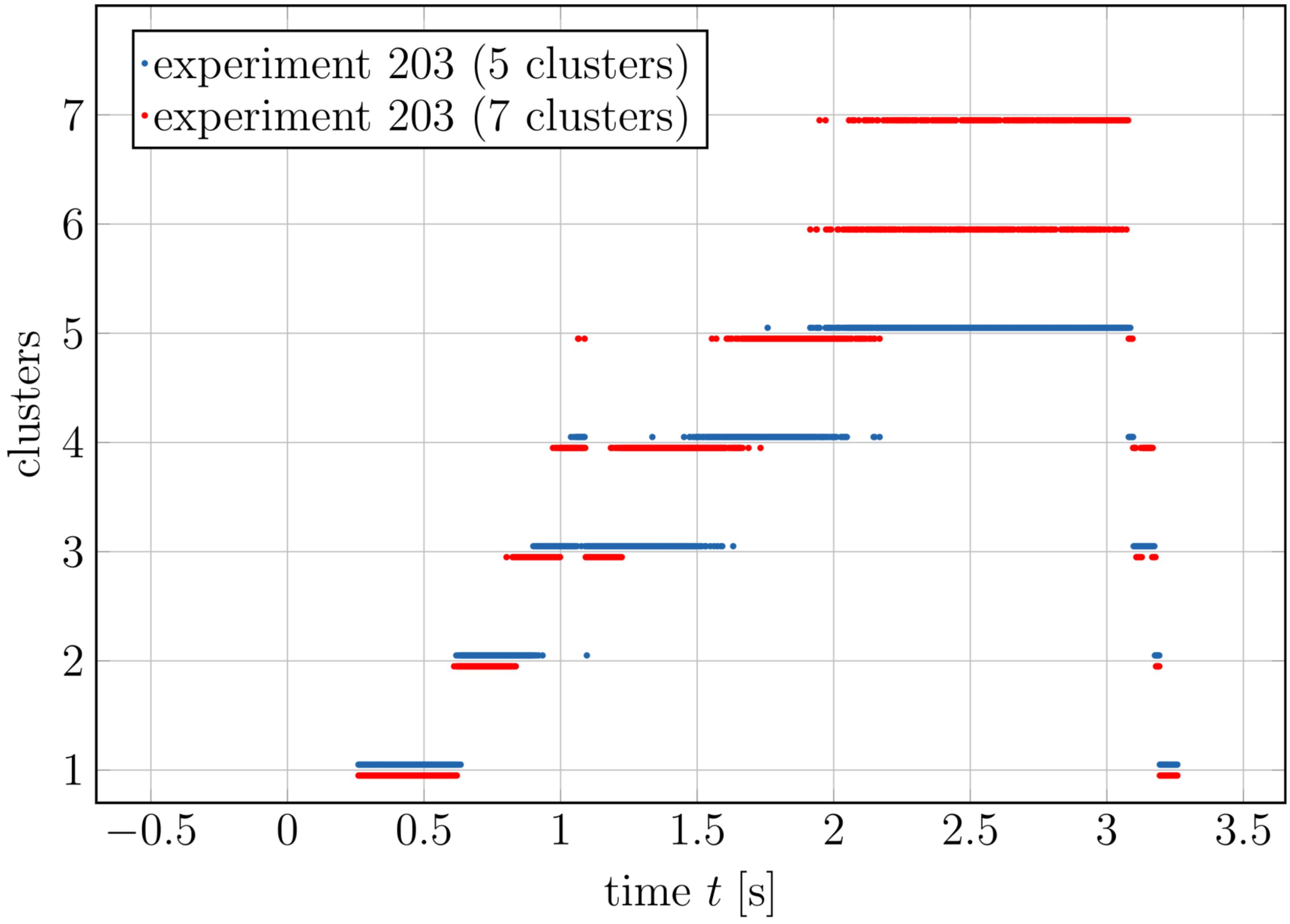
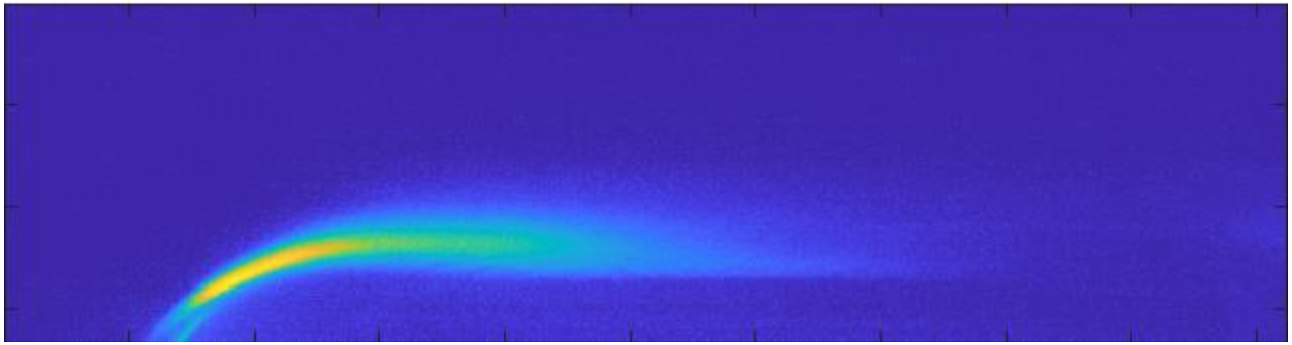
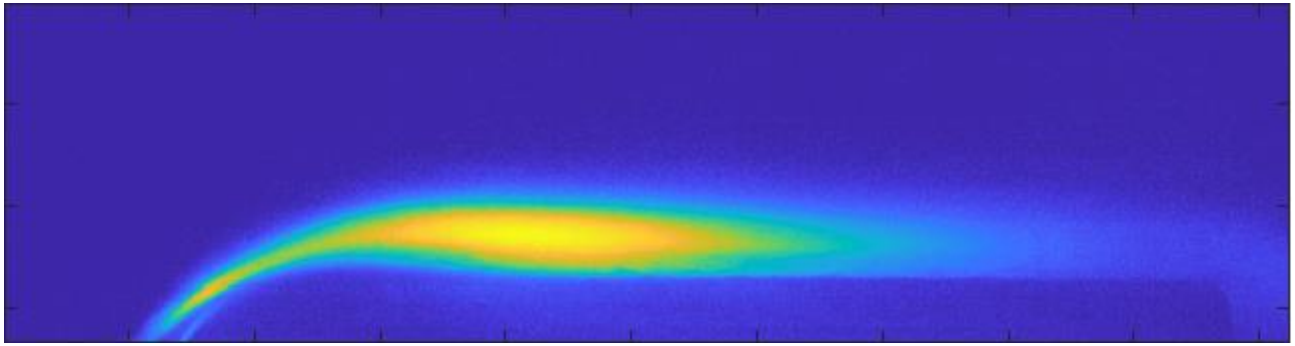


Figure: Distribution of the frames in experiment 203 to K=5 and K=7 clusters.

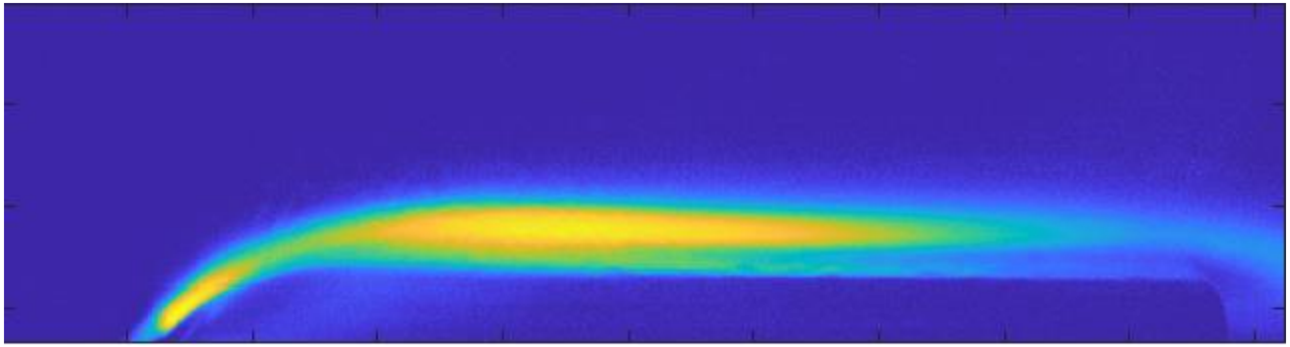
Experiment 284 with *K*=7



cluster 1
(1320/30000 frames)
ignition phase
(ignition comes from bottom of the chamber)

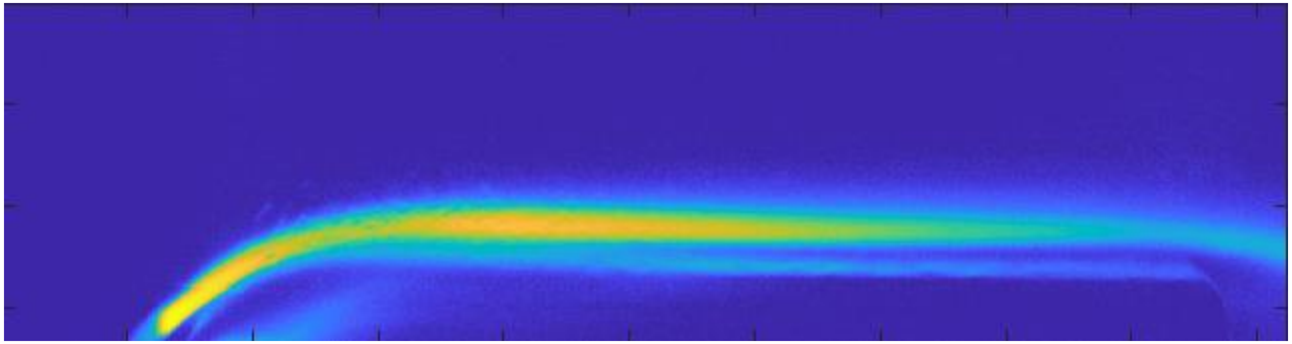


cluster 2
(2942/30000 frames)
burn phase without energy from outside (ignition valves closed)

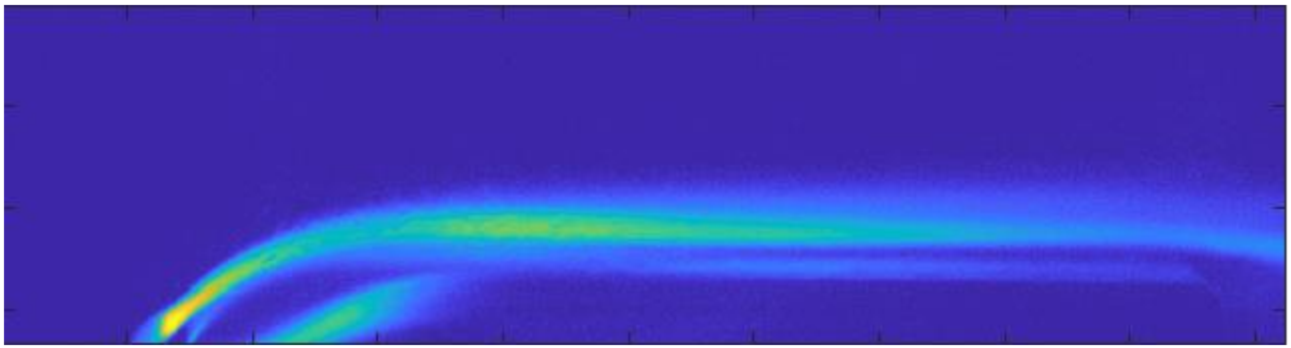


cluster 3
(3493/30000 frames)
fuel slap burns in the middle (oxygen mass flow increases)

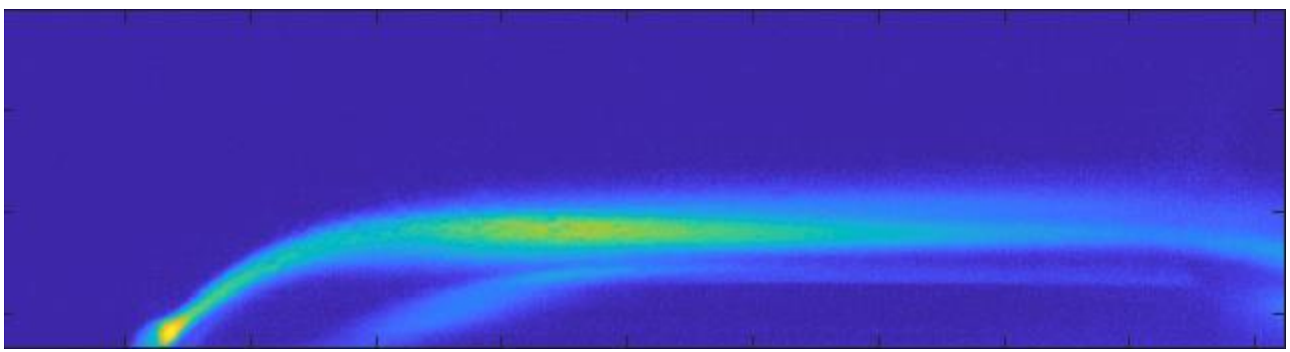
Experiment 284 with *K*=7



cluster 4
(3493/30000 frames)
whole surface is burning (brightness decreases due to $CH^* + O2 = CO + OH^*$)

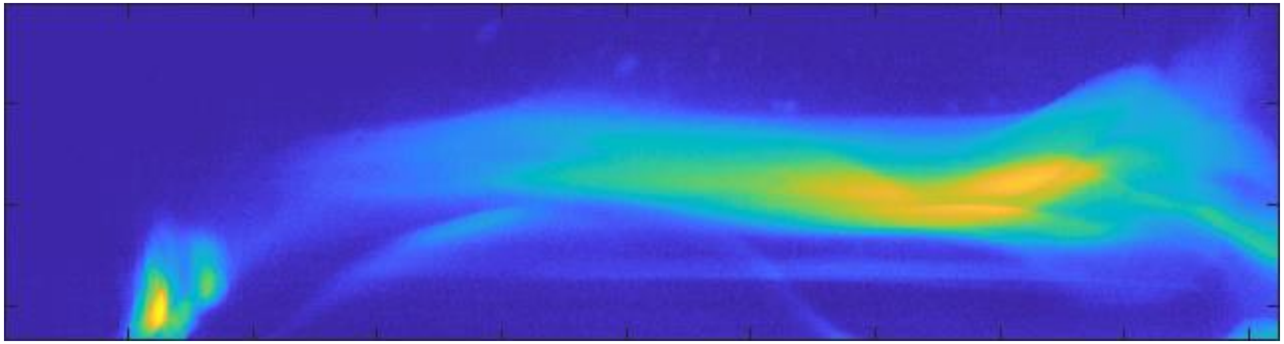


cluster 5
(2452/30000 frames)
large side flame close to the camera appears



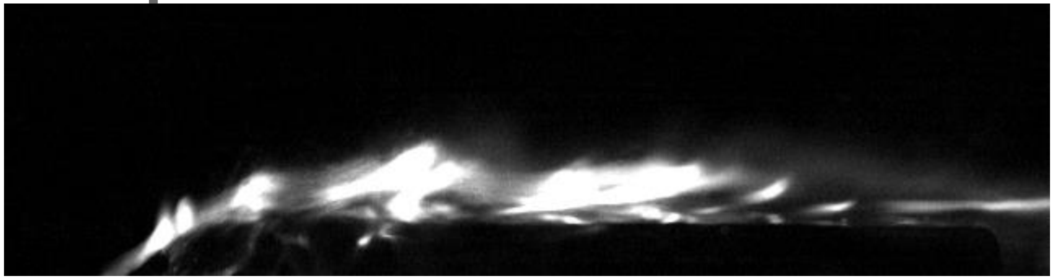
cluster 6
(16980/30000 frames)
constant combustion (with low CH* concentration, largest cluster)

Experiment 284 with *K*=7

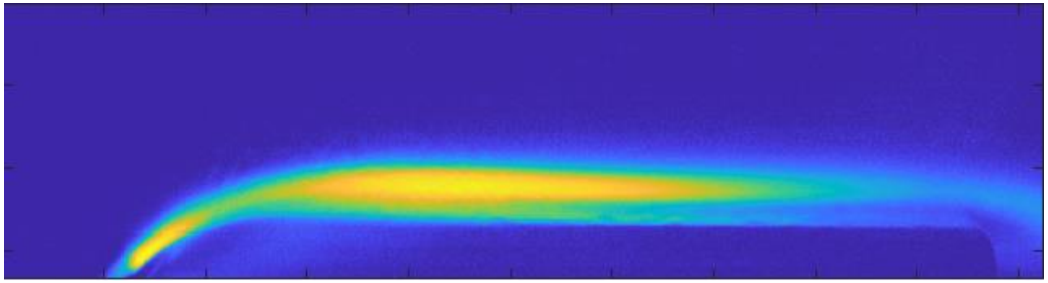


cluster 7
(194/30000 frames)
flame extinguishing phase (oxygen valve closes, nitrogen purge)

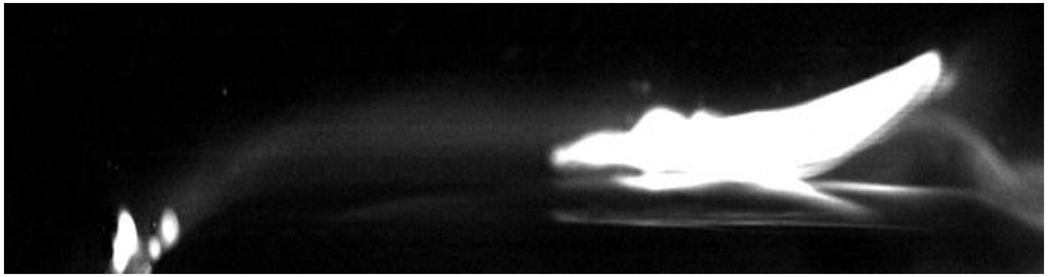
Comparison: centroids – individual frames



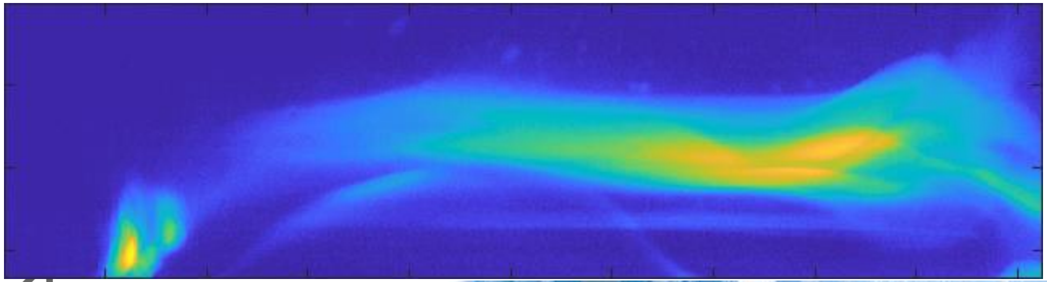
experiment at t = 0.7006 s



centroid 3



experiment at t = 2.3576 s



centroid 7

Conclusion and outlook

- Analysis of turbulent combustion experiments with a pressure of 15 bar – 30 bar in combustion chamber.

- Clustering is very time consuming at the moment (about 1.5 days per experiment for sequential code)
  - Usage of the HPDA cluster from SC for this problem
  - HeAT – Helmholtz Analytics Toolkit (see talk by Martin Siggel)
  - Mini batch K-means++ instead of K-means++?
  - Dimensionality reduction with an autoencoder + clustering of reduced dataset

- Comparison with different approach (e.g. density-based clustering, hierarchical clustering, …)

Thank you for your attention!