Burning Phases Separation by Application of Clustering Algorithm to Hybrid Rocket Combustion Data

By Anna PETRAROLO,[1)] Alexander RUETTGERS,[2)] and Mario KOBALD[1)]

[1)]*German Aerospace Center (DLR), Institute of Space Propulsion, Propellants Department, 74239 Hardthausen, Germany*

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

K-means is a clustering algorithm that allows the classification of data into specific groups or clusters. Data points belonging to the same cluster are more similar (same properties and/or features) to each other than those in other groups. In this study, K-means++ clustering, an improved variant of the K-means clustering, was applied to hybrid rocket combustion data, in order to get insights into the complex flow phenomena. Hybrid rocket propulsion is a promising technology for many applications, because it allows for cost reductions, still being able to deliver similar performance as solid and liquid propulsion. However, the complex fluid dynamic and combustion phenomena taking place in the hybrid rocket combustion chamber are not fully understood yet and still a matter of ongoing research. In the last years, many combustion tests with different paraffin-based fuels were performed at the German Aerospace Center (DLR) with an optically accessible combustion chamber. For gaining a better insight into the combustion process, the tests were captured with a high-speed video camera. This led to a huge amount of data images for each test, which needed to be analysed in detail. In the framework of this study, the combustion data set was clustered with a K-means++ algorithm. From the results it is possible to observe that different clusters, corresponding to different combustion phases, were identified by the algorithm. A dependency of the combustion flame behaviour on the oxidizer mass flow and fuel composition was also found.

Key Words: Hybrid Rocket Combustion, K-means algorithm, Clustering, Optical Investigations

Nomenclature

$C = c_k$	:	cluster, [-]
J	:	squared error, [-]
K	:	number of clusters, [-]
$X = x_i$	:	set of data points, [-]
d	:	problem dimension, [-]
n	:	dimension of the set of data points, [-]
α_k	:	weight factor, [-]
μ_k	:	mean of cluster c_k (centroid), [-]

1. Introduction

Hybrid rocket engines have several advantages compared to classical solid or liquid rockets. Due to the fact that the propellants are stored in two different states of matter, hybrid motors are safer than solid motors. This also contributes to reduce the total costs of the engine. Moreover, they are characterized by controllable thrust, including shut off and restart capability. With respect to liquid engines, they are mechanically simpler and, consequently, cheaper.[9)] Finally, their performance are in between those of solid and liquid engines. However, due to the diffusion limited combustion process typical for this kind of engines (the propellants are not pre-mixed, but they need to gasify and mix with each other before being able to react), hybrid systems using conventional polymeric fuels are characterized by poor regression rate performance (resulting in low thrust level). In order to overcome this problem, the so-called liquefying hybrid rocket fuels, such as paraffin-based ones, can be used. These fast burning fuels are characterized by low viscosity and surface tension and they experience a different combustion mechanism with respect to conventional polymeric fuels.[8)] During the combustion, instead of pyrolysing, they form a thin liquid layer on the fuel surface, which becomes unstable due to the high-speed gas flow in the fuel port.[7)] The liquid layer instabilities produce, in turn, droplets that entrain in the gas flow, thus working like a spray injection along the length of the motor. This causes an increase in the fuel burning area and, consequently, an increase in the regression rate of the fuel grain. Unfortunately, the entrainment phenomenon in hybrid rocket combustion process is still a matter of ongoing research and not yet fully understood.

For a better understanding of the experiments, the combustion process has been captured with a high-speed video camera that is able to capture 10 000 frames per second. This produces a huge amount of data, which needs to be analyzed in detail. Therefore, it is important to group the images into subsets such that the essential flow structures and their length in time can be identified. In each combustion test, at least three different flow phases are expected, which correspond to the ignition phase, the steady combustion state and the extinction of the flame. Furthermore, turbulent or other irregular structures might exist in the dataset. Even if these turbulent structures only exist within a short period of time, they might strongly affect the overall combustion behaviour. Therefore, a clustering of the dataset is required, which is able to group the data into separate flow phases and furthermore to detect strongly irregular combustion phenomena.

In this work, K-means++, an improved variant of K-means clustering,[11,12)] using the Euclidean metric is applied to the combustion data. Since the number of clusters K is not known in advance, a detailed analysis has been performed to estimate K in each experiment. As a result, it is shown that there is not a single optimal choice for K in our dataset but that for each

experiment several values should be considered. This results can be generalized to the clustering of experimental datasets in other areas of fluid dynamics. The analysis gives detailed insights into the combustion process and allows to identify optimal experimental configurations for future experiments.

2. Data Clustering

The goal of data clustering is to discover the natural grouping of a set of objects. Practically, given a representation of n objects, the clustering algorithm finds K groups based on a measure of similarity such that the similarities between objects in the same group are high, while the similarities between objects in different groups are low.[5] Of course, there is not a unique notion of similarity. The clusters can differ in terms of their shape, size and density. The presence of noise in the data makes the detection of clusters even more difficult. In fact, an ideal cluster is composed of a set of points that is compact and isolated. In reality, a cluster is a subjective entity that is in the eye of the beholder and whose significance and interpretation requires domain knowledge. It is also important to underline there is no best clustering algorithm. Each algorithm imposes a structure on the data either explicitly or implicitly. When there is a good match between the model and the data, good partitions are obtained. Since the structure of the data is not known a priori, it is important to try different approaches to determine an appropriate algorithm for the data to analyze.[5] Clustering algorithms can be generally divided into two groups:

- *hierarchical*, which recursively find nested clusters either in agglomerative mode or in divisive mode;
- *partitional*, which find all the clusters simultaneously as a partition of the data and do not impose a hierarchical structure.

The most well-known and simplest partitional algorithm is K-means. Let $X = x_i, i = 1, ..., n$, be the set of n dimensional points to be clustered into a set of K clusters, $C = c_k, k = 1, ..., K$. K-means algorithm finds a partition such that the squared error between the empirical mean of a cluster and the points in the cluster is minimized. Let μ_k be the mean of cluster c_k. The squared error between μ_k and the points in cluster c_k is defined as:

$$J(c_k) = \sum_{x_i \in c_k} \|x_i - \mu_k\|^2 \tag{1}$$

The goal of K-means is to minimize the sum of the squared error over all K clusters,

$$J(C) = \sum_{k=1}^{K} \sum_{x_i \in c_k} \|x_i - \mu_k\|^2 \tag{2}$$

The minimization of this objective function leads only to a local minimum, even though there is a large probability that K-means could converge to the global optimum when clusters are well separated.[11] K-means starts with an initial partition with K clusters and assigns patterns to clusters so as to reduce the squared error. Since the squared error always decreases with an increase in the number of clusters K (with $J(C) = 0$ when $K = n$), it can be minimized only for a fixed number of clusters.[11]

The combustion dataset is clustered with a K-means++ algorithm[2] that bases on the classical K-means algorithm.[11,12] K-means++ and K-means only differ in the initialization of the cluster centroids. In contrast to K-means, K-means++ reduces the probability that the algorithm terminates with local optimal findings for $\min(J)$ as final output. This is achieved by using an improved seeding algorithm. K-means++ takes only the first centroid from a uniform random distribution. Then, all further cluster centroids are taken randomly from a weighted probability distribution that scales with the squared distance to the closest centroids that have already been chosen (see Arthur and Vassilvitskii[2] for a more detailed description). In average, this leads to a more homogeneous distribution of the initial centroids. Furthermore, it can be shown that the objective function J of the initial centroids chosen by K-means++ differs from the objective function of the optimal clustering only by a factor of $O(\log K)$ in expectation. Nevertheless, all experiments in Section **4.** are clustered with ten independent runs of K-means++ to avoid obtaining local optimal solutions

Next, we motivate the choice of a K-means type clustering for our specific problem. Since each observation $x_i \in \mathbb{N}^d$ corresponds to a d-pixel grayscale image represented by integer values, the dimensionality d of the problem is high. Nevertheless, the Euclidean metric is widely used to quantify the difference of two images. It is known from the literature that the Euclidean metric is prone to the so-called "Curse of dimensionality"for high-dimensional data, cf. Steinbach et al.[18] This can be a severe problem for density-based clustering such as DBSCAN[4] or OPTICS[1] in which an ϵ-neighbourhood has to be specified, since a close to optimal choice for ϵ can be difficult in practice. To avoid this problem, a K-means type clustering has been chosen, in which the number of clusters K instead of the parameter ϵ has to be predefined. For our specific application, this input quantity can be roughly estimated from background knowledge in combustion theory, whereas a parameter such as ϵ cannot. Nevertheless, since the number of clusters in the combustion dataset is not known in advance, a detailed analysis to find K is necessary. This article follows a twofold approach to determine K.

First, different values for K are clustered in a range from one up to ten. Using the different results for the minimized objective function, a function $f(K)$ is built to estimate an optimal choice of K. In the following, an evaluation function $f(K)$ according to Pham et al.[16] is used. Using this approach, the objective function is

$$f(K) = \begin{cases} 1 & \text{if } K = 1 \\ \frac{J(K)}{\alpha_K J(K-1)} & \text{if } J(K-1) \neq 0, \forall K > 1 \\ 1 & \text{if } J(K-1) = 0 \forall K > 1 \end{cases} \tag{3}$$

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

where $J(K)$ is the optimal solution of the objective function for K clusters, d is the number of dimensions and α_K is a weight factor. The weight factor α_K is used to reduce the effect of the problem dimension. It converges to 1 in the limit $d \to \infty$. Therefore, in the high-dimensional application that is considered here $f(K) \approx J(K)/J(K-1)$ for all $K > 1$ and $f(K) = 1$

otherwise. According to Pham et al.,[16] $f(K)$ represents the ratio between real distortion and estimated distortion. The output of $f(K)$ is close to 1 when the data distribution is uniform. Values of $f(K)$ smaller than 1 reflect that the data distribution is more concentrated. As a result, a small value of $f(K)$ is an indicator for a choice of K that gives well-defined clusters.

Second, even though the problem dimension is high, the centroids μ_k can still be visualized and give information on the mean combustion in the corresponding clusters c_k. Therefore, it is essential to give a physical interpretation of the phenomena that are shown in each cluster center. Then, the optimal number of clusters depends on the leading question. If only the main flow phases of the combustion are relevant, K can be chosen comparatively low. On the other hand, if short-time turbulent structures are relevant, K has to be chosen larger. The second approach therefore leads to an ambiguity for K but ensures that the results are relevant in practice.

3. Combustion Tests

The combustion tests were performed at the German Aerospace Center (DLR), Institute of Space Propulsion in Lampoldshausen, at the test complex M11. An already existing modular combustion chamber, used in the past to investigate the combustion behavior of solid fuel ramjets,[3] was adjusted and used for the test campaigns at atmospheric pressure. A side view of the whole combustion chamber set-up is shown in Fig. 1. The oxidizer main flow is entering the combustion chamber from the left, after having passed two flow straighteners. The mass flow rate is adjusted by a flow control valve and it is measured with a Coriolis flow meter. A high frequency static pressure sensor is mounted in the combustion chamber. Ignition is done via an oxygen/hydrogen torch igniter from the bottom of the chamber. A test sequence, reported in Table 1, is programmed before the test and is run automatically by the test bench control system. More details about the test bench and test settings are given in Kobald et al. and Petrarolo et al.[10,13,15]

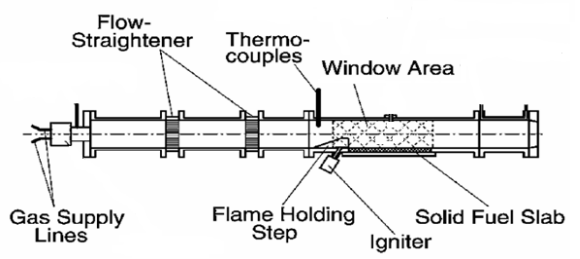


Fig. 1. Side view of the atmospheric combustion chamber set-up, adapted from[19]

In the framework of this research, all tests were done at atmospheric pressure and with an oxidizer mass flow ranging from 10 to 120 g/s. Combustion tests were performed using a single-slab paraffin-based fuel with a 20° forward facing ramp angle (see Fig. 2), in combination with gaseous oxygen. Two different fuel compositions were analyzed in this study: pure paraffin 6805 from the manufacturer Sasol Wax and the same paraffin with 5% mass addition of a commonly available polymer. Burning time was 3 seconds for each test. For video data acquisition a Photron Fastcam SA 1.1 high speed video camera was used with a maximum resolution of 1024x1024 pixel. The frame rate, resolution and shutter time of the camera were adjusted for each test, according to the test conditions and position of the camera. A test was also done using a CH* chemiluminescence imaging technique, with a band-pass filter centered around 431 nm placed in front of the camera. The excited CH* molecules emit photons around this wavelength, when they relax back to a lower energy state. Since high CH* concentration exists only in the main reaction zone, the resulting images provide a good indication of the instantaneous flame sheet location and topology.

Table 1. Automatic test sequence

Time [s]	Action
T-0.2	Open ignition valves
T0	Open oxidizer main valve, start high-speed camera
T+0.3	Close ignition valves
T+3	Close oxidizer main valve, start nitrogen purge
T+5	End of sequence

Fig. 2. Fuel slab configuration used in this research, before (top) and after (bottom) combustion test

In this study, five combustion tests have been analyzed. The test matrix is presented in Table 2.

Table 2. Test matrix

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

4. Results and Discussion

This section is subdivided into two parts. First, the output of the clustering algorithm is analyzed to determine the number of relevant clusters in the dataset. Second, a physical interpretation of the different clusters and their centroids is given and the results are connected to the experiments.

4.1. Analysis of the number of clusters

The experiments listed in Table 2 have been clustered with a K-means++ algorithm and different input values for K up to ten. Furthermore, all runs of the algorithm with all values for K have been repeated ten times to avoid obtaining local optimal minimum solutions for J. In all cases, the relative differences in $min(J)$ in the different runs was below 0.5% and the obtained clusters were almost identical. Therefore, this indicates that the clustering results are optimal with respect to the Euclidean distance norm.

The algorithm was implemented vectorized in Matlab and was executed sequentially on a workstation with 128 GB main memory. A complete clustering of a single experiment with an image data size of about 8 GB integer values required about 1.5 days on the workstation. Since the combustion data was converted from integer to floating point numbers to allow a computation of the cluster centroids (arithmetic mean of all images in the cluster), the main memory requirement of the program was in the order of 60 GB. Furthermore, it is noted that the complexity of K-means type clustering increases linearly with K, which restricts the maximum value for K in practice.

As explained in Section **2.**, we first determine K for each experiment. For this purpose, Fig. 3 visualizes $f(K)$ as defined in Equations (3) and (4) for all experiments. In the region of K that is considered, $f(K)$ always has its global minimum at $K = 2$. This indicates that there are two main structures in the combustion on the coarse scale: an ignition phase and a much longer steady combustion process. But, as known from the combustion theory, more structures can be found on a short-time scale. Furthermore, we note that $f(K = 2) \approx 0.8$ for experiment 284 but $f(K = 2) \approx 0.6$ for the experiments 203 and 253. This already indicates that more short-time phenomena can be found in experiment 284 compared to the remaining experiments. Moreover, $f(K)$ of test 284 has a local minimum for $K = 7$. Later on, it is shown that a short-time turbulence is resolved with $K = 7$ and the results will be interpreted in Subsection **4.2.**. The outcome of $f(K)$ in Fig. 3 for the experiments 203, 234, 243 and 253 are not conclusive in the first place. In all cases, values for K in the range $2, \ldots, 6$ can be justified. For larger values $K > 6$, further short-time structures might exist even if $f(K) \approx 1$ but the computational effort increases strongly and results cannot be achieved within a reasonable amount of time.

Since the dataset consists of images from a high-speed video camera, it is interesting to analyze the correlation between the clusters and the points in time of the corresponding images. This correlation is shown in Fig. 4 for all five combustion experiments, each with individual choices for K. As a first result, the K-means++ algorithm has identified structures that are connected in time and therefore can be interpreted as a separate flow phase. For a better illustration of the output, the clusters have been ordered in time such that the first cluster represents

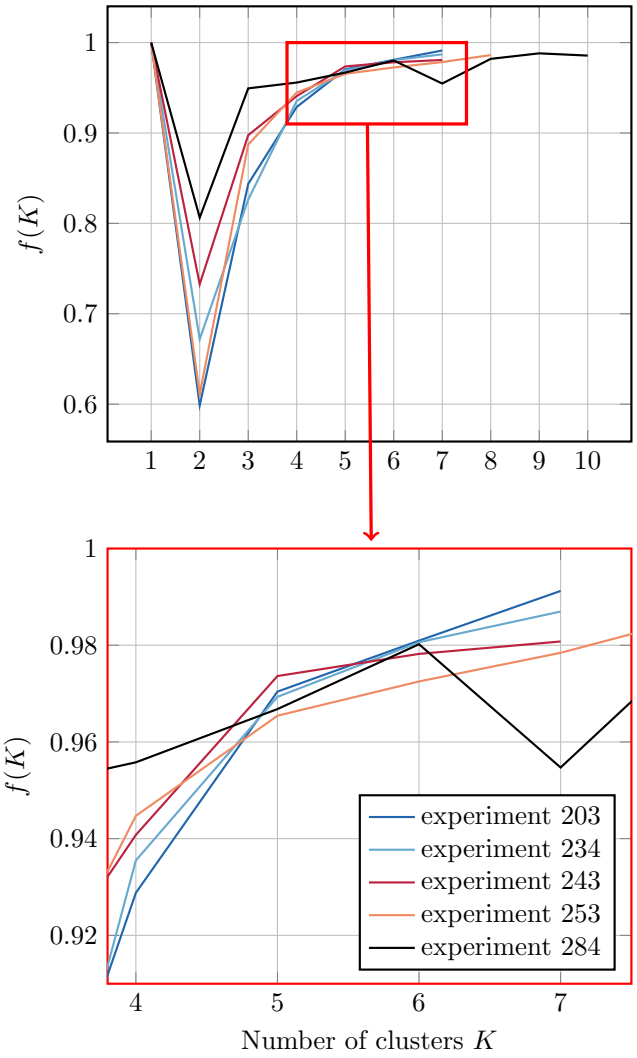


Fig. 3. Evaluation function $f(K)$ for each combustion experiment to determine the number of clusters K.

images from the first flow phase and analogously for all other clusters. Furthermore, the different scatter plots in Fig. 4 have been slightly shifted in vertical position. The reason for this shift in y-direction is to avoid an overlap of the different experiments. Again, it is noted that the clusters for experiment 284 differ from the remaining experiments. More precisely, cluster 6 in experiment 284 is a long-time flow phase that consists of 16980 of the 30000 images or, equivalently, exists for about 1.7 seconds in time. In the other experiments, the long-time phase has a length of about 1.2 seconds in time. Moreover, an extinction phase for about $t > 3.2$ seconds is resolved in all experiments. Since this phase only exists for a short period in time, it is allocated to the initial clusters C_1, C_2 and C_3. It is expected that the extinction phase can be separated into its own cluster for much larger choices of K.

Figure 4 lists 6 separate clusters for the experiments 234, 243, 253 and 5 clusters for experiment 203. These choices will be justified in the following. According to Fig. 3, it was concluded that $K \leq 6$. In practice, a value of K that is too small has the effect that relevant flow phases are not resolved. On the other hand, if K is too large the only potential drawback is that identical flow phases are separated in two clusters. But, as the separated structures are connected in time, see Fig. 4, this situation

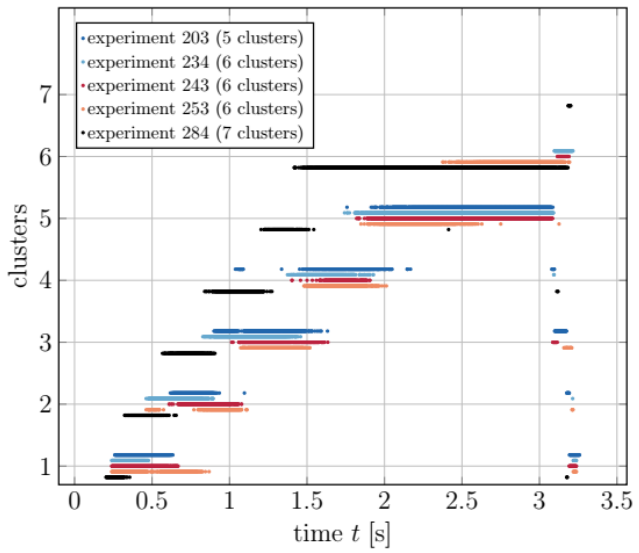


Fig. 4. Distribution of the frames in 5 experiments to their corresponding clusters. For a better visualization and to avoid an overlap, the points have been slightly shifted in vertical direction.

can be identified. As a result, a comparatively large value $K = 6$ has been chosen for the experiments 234, 243 and 253 except for $K = 5$ for experiment 203. In experiment 203, the effect that an identical flow phase is separated into two clusters already occurs for $K = 6$. As an example, this effect is visualized in Fig. 5 for the clusters with $K = 5$ and $K = 7$ in experiment 203. It is obvious that the clusters C_6 and C_7 for $K = 7$ belong to the same period in time and therefore represent the same flow phase. This effect is also visible for $K = 6$ but less pronounced. As a result, experiment 203 is analyzed with $K = 5$ clusters in Section **4.2.**.

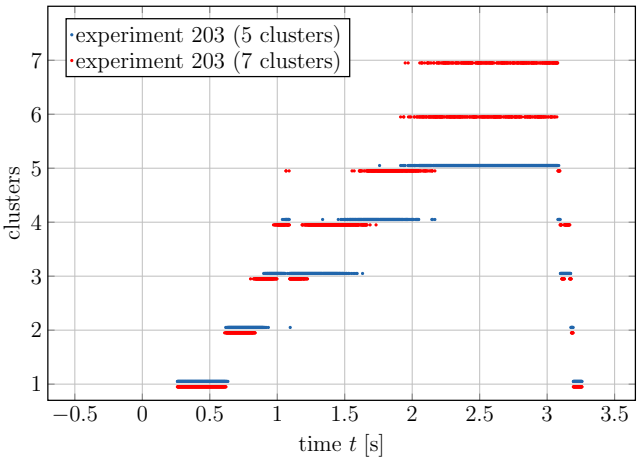


Fig. 5. Distribution of the frames in experiment 203 to $K = 5$ and $K = 7$ clusters.

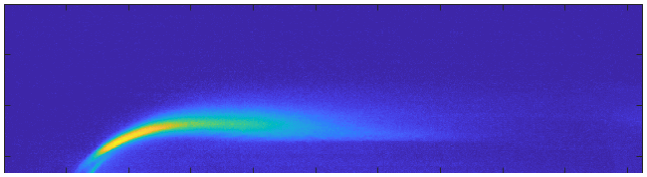
4.2. Analysis of the combustion

As already said in Section **3.**, five combustion tests have been analyzed in the framework of this study (see Table 2). The results obtained give many insights into the hybrid combustion process and allow to separate the different burning phases.

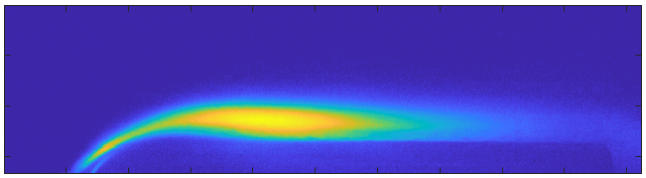
First of all, it is interesting to have a look at test 284. This combustion video was realized with a band-pass filter centered around 431 nm placed in front of the high-speed camera. It is generally recognized (see Devriendt et al.[6] and Schefer[17]) that the primary species contributing to flame luminescence are the electronically excited species CH*, C2* and OH*. All three species show a close correspondence across the main reaction zone and are thus equally suitable as markers for the flame zone location. In particular, the concentrations of CH* increase rapidly to a maximum within the flame and then decay rapidly downstream of the reaction zone.[17] Therefore, the CH* images of test 284 give a good representation of the main flame location. Fig. 6 shows the centroids belonging to the 7 clusters separated by the algorithm. A centroid of a cluster is an average image coming from all the frames belonging to that cluster. The first cluster goes from t=+0.2 until t=+0.33 seconds and it represents the initial ignition phase. In fact, until t=+0.3 seconds the valve of the torch igniter is open (see Table 1) and the gases coming from the hydrogen/oxygen combustion enter the combustion chamber from the bottom head of the fuel slab. They start gasify the fuel, which, in turn, initiates burning with the oxygen entering the chamber. Since the ignition is coming from the bottom of the combustion chamber, the first area that starts to gasify and to burn is exactly the fuel step, as it can be seen in the image of centroid 1 in Fig. 6. Clusters 2, 3 and 4 are representing the different phases of the fuel ignition transient and are going from t=+0.33 until around t=+1.23 seconds. Cluster 2 is starting directly after the closing of the ignition valves, thus no combustion gases are coming from the igniter any more and the fuel slab starts to burn without any energy addition from the outside. This is the start of a self-sustained combustion. At this point, the oxygen mass flow is not yet settled down ($\dot{m}_{Ox} < 10\ g/s$), but it is still increasing; therefore, there is still no steady-state. As the fuel slab starts to burn, different phases can be observed in these 3 clusters: at the beginning (cluster 2) the fuel slab is burning just in the front, then, as the time goes on (and the oxygen mass flow increases $\dot{m}_{Ox} < 25\ g/s$), it starts to burn in the middle (cluster 3) and, finally, in cluster 4 ($\dot{m}_{Ox}$ reaches 40 g/s), almost the whole surface is burning. It is also possible to notice that, in cluster 4, the brightness of the flame already starts going down. This means that the temperature in the combustion chamber is already quite high and a quasi steady-state flame is appearing. This could seem strange, but it is important to remember that the CH* radicals are more likely to appear at the beginning of the combustion process, when the temperature is not yet too high. Later, they are further reacting (the most likely reaction is: $CH^* + O_2 = CO + OH^*$) and, therefore, their concentration decreases while the OH* concentration increases. So, it is more likely that the brighter burning regions are seen at the beginning of the combustion (clusters 2,3,4), while during the steady-state phase (clusters 5 and 6), the concentration of the CH* decreases and so does also the flame brightness (see Fig. 7). Clusters 5 and 6 also show a more or less constant flame thickness on the whole fuel slab length. In the time window of these two clusters (between t=+1.23 and t=+3.18 seconds), the oxidizer mass flow is more or less steady (it increases from 45 to 50 g/s), the temperature reaches the maximum flame temperature and the flame is well settled down on the fuel slab surface. This situation does not change until the closing of the oxygen main valve and the opening of the nitrogen purge at t=+3.18 seconds, when cluster 7 starts. This last phase represents the combustion extinction phase. According to the test sequence, at t=+3 seconds, the oxygen main valve

closes and the nitrogen purge starts. In the video, it is possible to see that the nitrogen purge comes at around t=+3.08 until t=+3.1 seconds. At t=+3.18 seconds the flame is completely extinguished. This causes, in turn, a drop in the chamber pressure, which practically causes a suction of still-burning paraffin gases from the rear part of the fuel slab. In the video it is clearly visible that these reacting gases are coming from the back and they are travelling until the fuel slab head (see Fig. 7). So, cluster 7 represents the flame extinguish phase, where the remaining paraffin gases are still burning with the rest of the oxygen available in the combustion chamber. As it can be noticed from Fig. 4, only clusters 1 (first ignition transient), 6 (steady state) and 7 (extinction) are well separated. Clusters 2,3,4 and 5 contain elements that could also belong to other clusters (they are slightly overlapped on each other). This happens because the transition between the ignition transient and the steady state phase is not well defined. The flow dynamics and, consequently, the flame brightness change slowly and gradually and it is not really possible to define when one phase is finished and when a new one starts. Thus, these four clusters represent the transition between the ignition and the steady state phases. On the other hand, the transition between the steady state and the extinction phase is well defined because it begins exactly when the nitrogen purge starts (at t=+3 seconds).

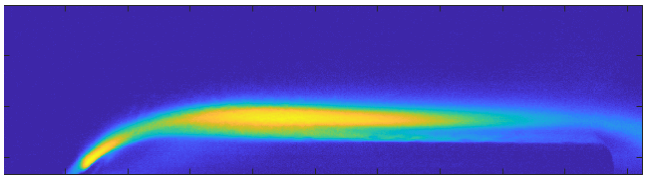
For what concerns the test without the filter, they present more or less the same phases. The first 3 clusters (between t=+0.5–0.7 and t=+1.45–1.65 seconds, depending on the test) represent the ignition transient of the fuel slab (note that the first visible flame, which corresponds to the first self-sustained flame, is only visible starting at t=+0.3–0.4 seconds). The first cluster represents the initial flame, which is not yet developed over the whole fuel slab length. Actually, only in test 243, there is a brighter visible area in front of the fuel slab (see Fig. 8). It seems like a recirculation zone where most likely the paraffin gases are entrained in a kind of vortex (probably caused by the step) and start burning with the oxidizer. Then, the vortex becomes bigger (cluster 2) and more paraffin gases get entrained in the recirculation area, until the whole fuel slab burns (cluster 3). It is possible to note that in test 243 there is always a small recirculation area directly after the step, which is not burning. The oxidizer is not able to reach that area because it burns completely before and so no combustion is occurring there. It is also observed that the higher the oxidizer mass flow, the faster the ignition phase. In tests 203, 234 and 253 (see, respectively, Fig. 9, 10, 11), the flame is already developed on the whole fuel surface in cluster 2, while, in test 243 ($\dot{m}_{Ox} = 10\ g/s$), it appears only in cluster 3 (see Fig. 8). This is due to the faster dynamics of the convection and diffusion processes, which allow a faster burning process (more turbulence, faster heat exchange, higher regression rate). Cluster 4 (between t=+1.45–1.6 and t=+1.86–2.2 seconds) represents the transition between the ignition and the steady-state, while cluster 5 (between t=+1.86–2.2 and t=+3.18 seconds) identifies the steady-state. This is easily recognizable as the longer phase, where the oxidizer mass flow has settled down to its steady state value. Only in test 253, the steady-state is divided into two smaller clusters (cluster 5 and 6). This is due to the presence of a reflection on the window (visible on the left side of the image), which changes the average brightness of the frames that



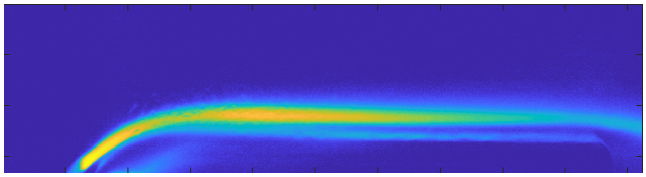
(a) centroid 1 (1320/30000 frames)



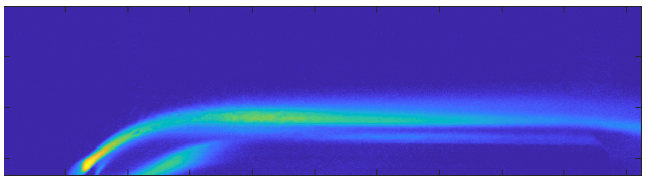
(b) centroid 2 (2619/30000 frames)



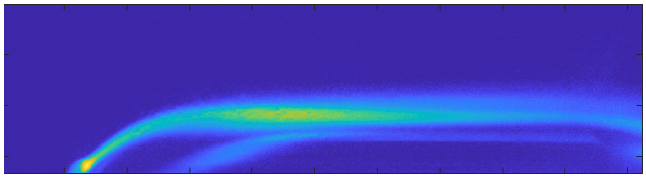
(c) centroid 3 (2942/30000 frames)



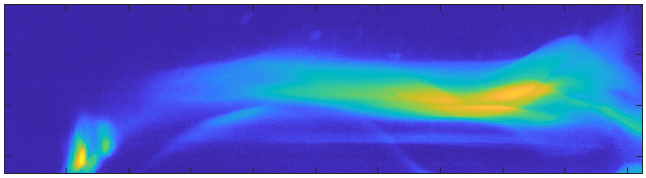
(d) centroid 4 (3493/30000 frames)



(e) centroid 5 (2452/30000 frames)



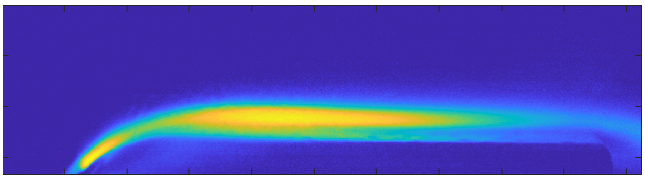
(f) centroid 6 (16980/30000 frames)



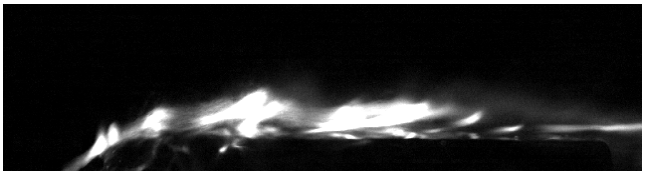
(g) centroid 7 (194/30000 frames)

Fig. 6. Visualization of the seven centroids in experiment 284.

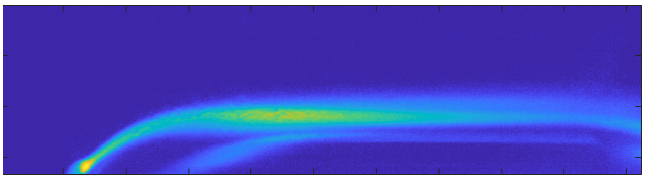
are then recognized as a new cluster by the algorithm. Since the flame in these two clusters is identical, it is possible to merge them and consider them as one. As it can be noticed from Figures 8, 9, 10, 11, in all the tests there is not a big difference between clusters 4 and 5. This is again due to the fact that there is no clear distinction between the end of the ignition phase and the beginning of the steady-state. Moreover, in test 234 ($\dot{m}_{Ox} = 100\,g/s$) and 203 ($\dot{m}_{Ox} = 50\,g/s$), the steady-state flame is not so flat like in test 243 ($\dot{m}_{Ox} = 10\ g/s$). This trend was



(a) centroid 3

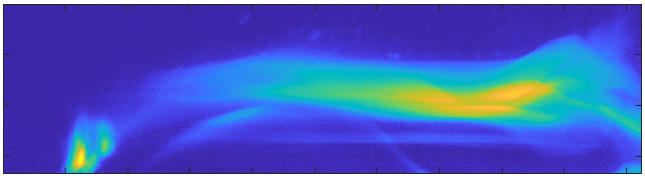


(b) experiment at $t = 0.7006\ s$



(c) centroid 6

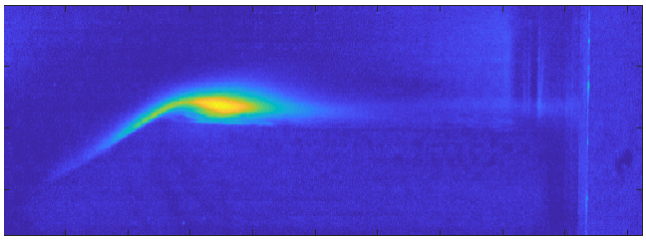
(d) experiment at $t = 2.3576\ s$



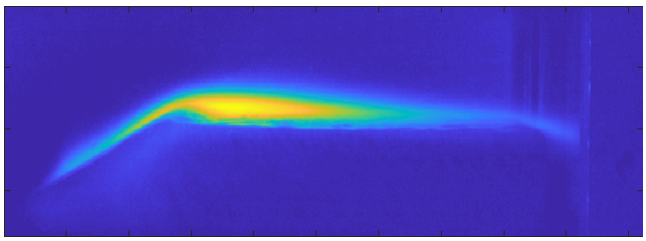
(e) centroid 7

(f) experiment at $t = 3.1927\ s$

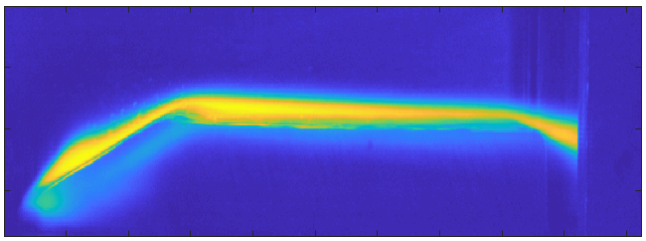
Fig. 7. Comparison between individual frames and the corresponding centroids for experiment 284.



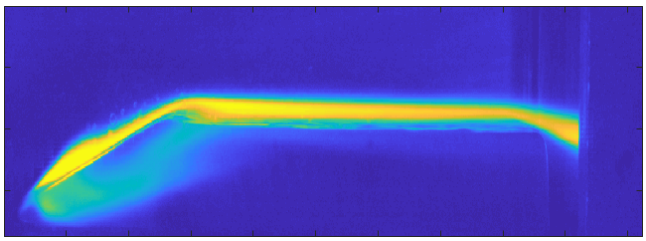
(a) centroid 1 (4677/30000 frames)



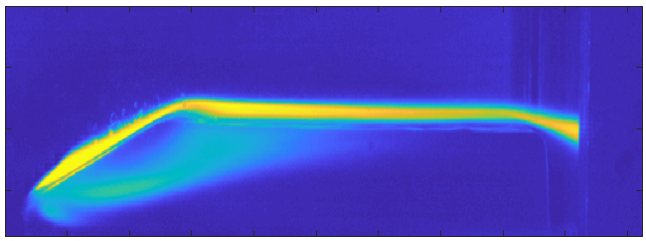
(b) centroid 2 (3995/30000 frames)



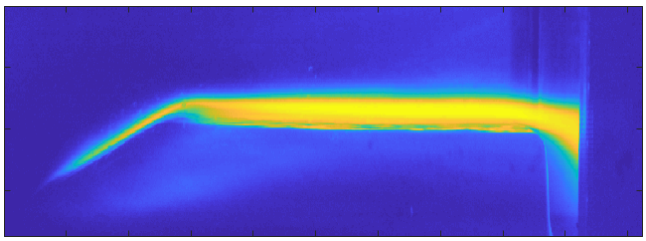
(c) centroid 3 (5521/30000 frames)



(d) centroid 4 (2950/30000 frames)



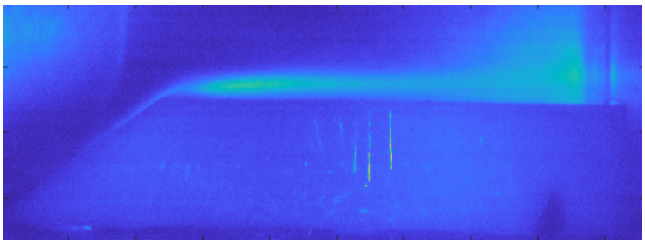
(e) centroid 5 (12093/30000 frames)



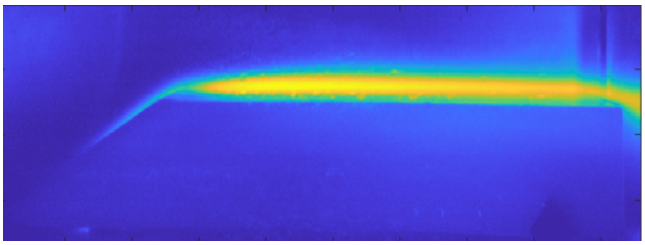
(f) centroid 6 (764/30000 frames)

Fig. 8. Visualization of the six centroids in experiment 243.

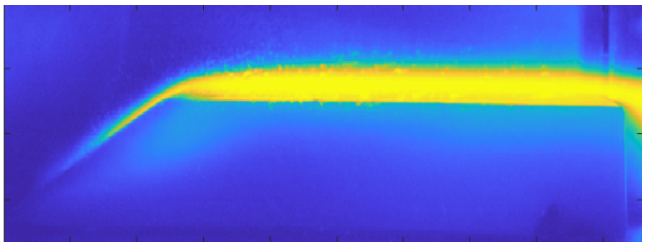
already observed during other test campaigns (see Petrarolo et al.[14]), where it was showed that the flame height depends on the fuel formulation (fuel viscosity) and oxidizer mass flow. In particular, for the lower viscosity fuel (pure paraffin 6805), the flame height decreases with increasing oxidizer mass flow. On the other hand, the flame height of the fuel with the higher viscosity (6805+5%polymer) increases with increasing oxidizer mass flow. The explanation for this trend could be found in the balance between the vaporization and the entrainment regression rate, which contribute to the final total regression rate in liquefying hybrid rocket fuels. The combination of fuel viscosity and oxidizer mass flow decides which of the two regression rates plays a more important role. For the lower viscosity fuel, the droplets entrainment at high oxidizer mass flows dominates direct gasification. This brings to an important increase in the regression rate and a decrease in the liquid layer thickness (see Karabeyoglu et al.[7]). Consequently, also the flame height tends to decrease. On the other hand, for the higher viscos-



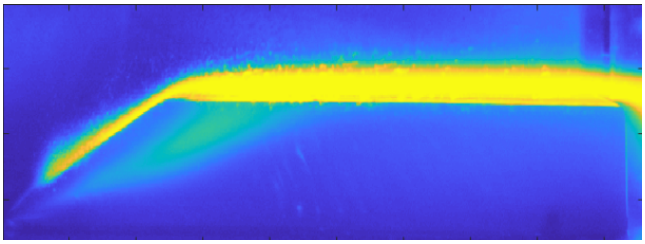
(a) centroid 1 (4361/30000 frames)



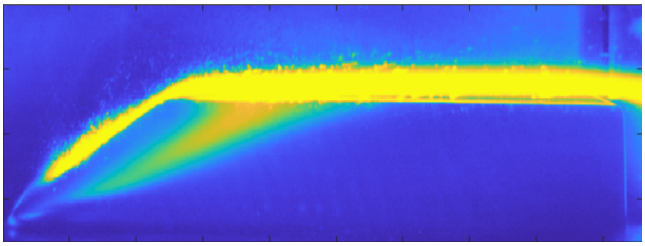
(b) centroid 2 (2953/30000 frames)



(c) centroid 3 (6525/30000 frames)

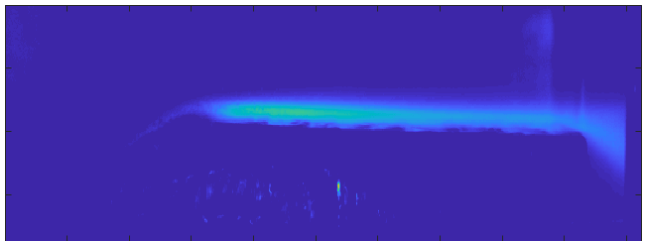


(d) centroid 4 (5429/30000 frames)

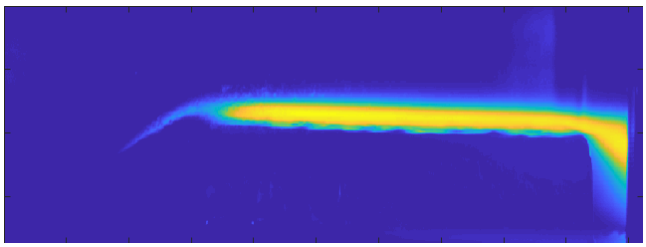


(e) centroid 5 (10732/30000 frames)

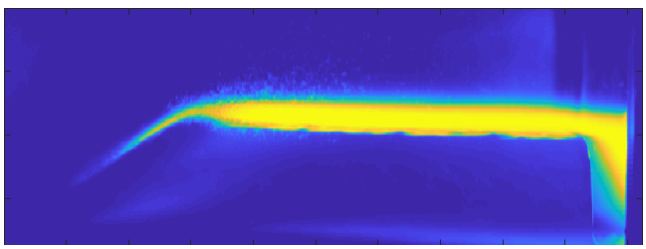
Fig. 9. Visualization of the five centroids in experiment 203.



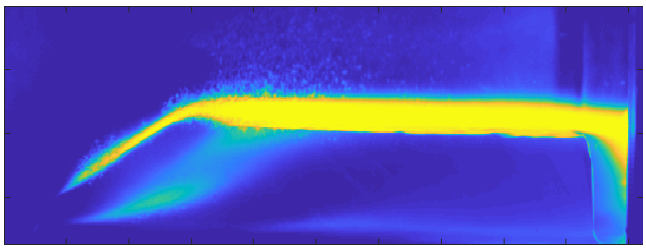
(a) centroid 1 (2533/30000 frames)



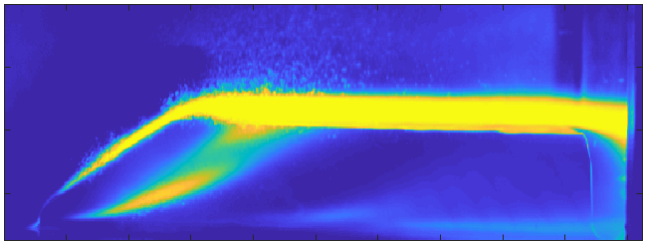
(b) centroid 2 (4093/30000 frames)



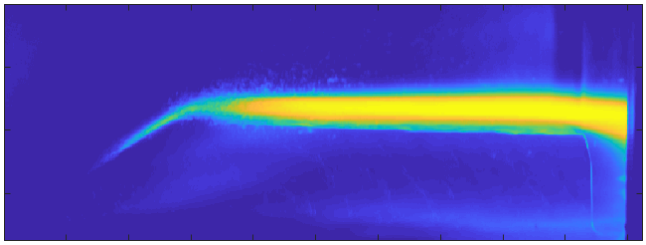
(c) centroid 3 (5273/30000 frames)



(d) centroid 4 (4326/30000 frames)



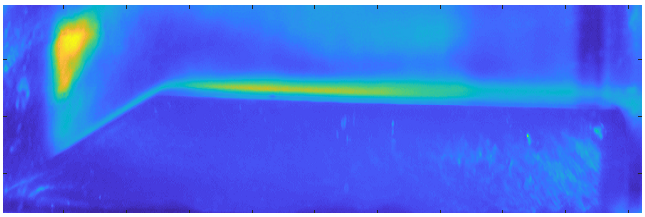
(e) centroid 5 (12598/30000 frames)



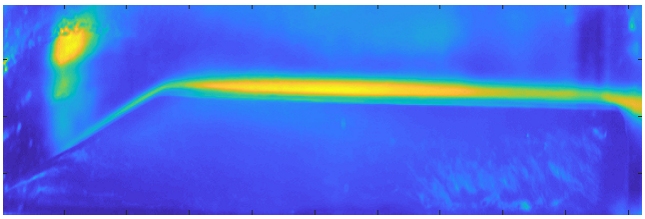
(f) centroid 6 (1177/30000 frames)

Fig. 10. Visualization of the six centroids in experiment 234.

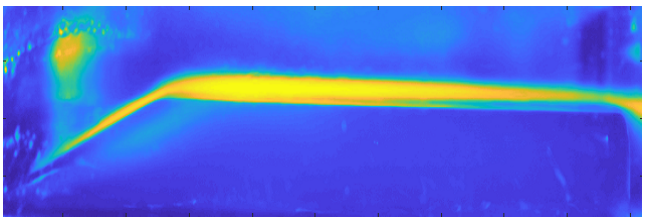
ity fuel (which is the case of the tests analyzed in this study), even at high oxidizer mass flows, the vaporization regression rate plays still an important role with respect to the entrainment mass transfer. This means that the evaporation blowing of the gaseous phase mass transfer from the fuel surface is still pretty high. This pushes the flame sheet further away from the liquid layer and, consequently, increases the flame height. Additionally, the steady-state flame seems to be a bit more detached from the fuel surface in test 243. This is probably also due to the different oxidizer mass flow rate. At higher mass flow rates (tests 234, 203, 253), the dynamic pressure exerted by the gas flow on the flame is higher and so the flame gets pushed towards the fuel surface. Moreover, the blowing effect from the



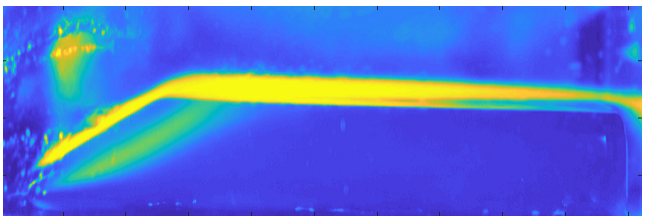
(a) centroid 1 (5329/30000 frames)



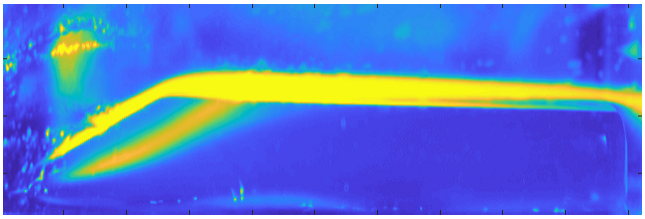
(b) centroid 2 (3404/30000 frames)



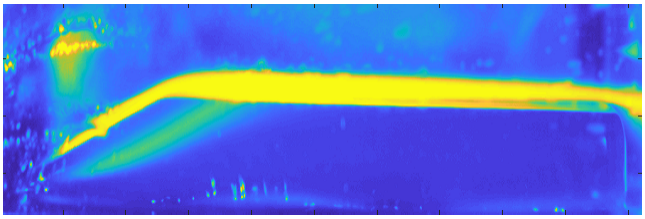
(c) centroid 3 (4582/30000 frames)



(d) centroid 4 (4268/30000 frames)



(e) centroid 5 (5912/30000 frames)



(f) centroid 6 (6505/30000 frames)

Fig. 11. Visualization of the six centroids in experiment 253.

fuel surface to the flame (due to the vaporization of the fuel) is higher in the lower oxidizer mass flow case (test 243). Thus the flame gets pushed far from the fuel surface. So this behaviour of the flame is due to the balance of these two forces (dynamic pressure and blowing). Finally, cluster 6 (between t=+3.18 and t=+3.2–3.24 seconds) represents the extinction phase. It is the shorter phase and it starts just after the closing of the oxidizer main valve and opening of the nitrogen purge (t=+3 seconds, according to the test sequence; in the video the purge comes at t=+3.08 seconds).

As it can be observed, the big differences between the tests with and without the filter are the transients. In fact, the initial ignition phase, when the ignition valve is still open (between t=+0.2 and t=+0.33–0.35 seconds), is not visible in the test without the filter (nothing is visible in the video of tests 243 before t=+0.35 seconds). The brightness of the combustion gases of the torch igniter is too weak to be visible without the filter. Thus, the first visible flame is that of the fuel slab itself when it starts burning. Also the extinction phase in the tests without the filter is different from that of test 284. This is, of course, also due to the filter that allows us to see only the CH* emissions.

5. Conclusion

K-means++ clustering was applied to the hybrid combustion dataset. To the best of our knowledge, this is the first application of image clustering in the specific application of hybrid rocket combustion. The analysis revealed several interesting short-time phenomena in the dataset and clearly indicate the potential of unsupervised learning techniques for the structuring of large datasets. This results could be also generalized to the clustering of experimental datasets in other areas of fluid dynamics.

In the future, a parallel implementation of the algorithm on the HPDA-cluster operated at DLR Simulation and Software Technology might remove the restrictions that were necessary to obtain results for such a large dataset within a reasonable amount of time, e.g. the limited amount of up to $K = 10$ clusters that was investigated. Larger values for K might resolve further short-time structures and give further insights into the complex combustion process. Different clustering algorithms will be also investigated and compared, in order to identify the appropriate algorithm to analyze the data at hand.

Acknowledgments

This research was carried out under the project *Antriebstechnologien und Komponenten für Trägersysteme* (ATEK) by the German Aerospace Center (DLR). The support of the M11 and the propellants department is greatly acknowledged.

References

1) M. Ankerst, M. Breunig, H.-P. Kriegel, and J. Sander. Optics: Ordering points to identify the clustering structure. *SIGMOD Rec.*, 28(2):49–60, June 1999.
2) D. Arthur and S. Vassilvitskii. K-means++: The advantages of careful seeding. In *Proceedings of the Eighteenth Annual ACM-SIAM Symposium on Discrete Algorithms*, SODA '07, pages 1027–1035, Philadelphia, PA, USA, 2007. Society for Industrial and Applied Mathematics.
3) H. K. Ciezki, J. Sender, W. Clau-oslash, A. Feinauer, and A. Thumann. Combustion of solid-fuel slabs containing boron particles in step combustor. *Journal of Propulsion and Power*, 19(6):1180–1191, November 2003.
4) M. Ester, H.-P. Kriegel, J. Sander, and X. Xu. A density-based algorithm for discovering clusters a density-based algorithm for discovering clusters in large spatial databases with noise. In *Proceedings*

of the Second International Conference on Knowledge Discovery and Data Mining, KDD'96, pages 226–231. AAAI Press, 1996.

5) Anil K. Jain. Data clustering: 50 years beyond k-means. *Pattern Recognition Letters*, 31:651–666, 2010.
6) B. Ceursters K. Devriendt, H. Van Hook and J. Petters. *Chem. Phys. Lett.*, 261:450–456, 1996.
7) A. Karabeyoglu, D. Altman, and B. J. Cantwell. Combustion of liquefying hybrid propellants: Part 1, general theory. *Journal of Propulsion and Power*, Vol.18(No. 3):610–620, May 2002.
8) A. Karabeyoglu, B.J. Cantwell, and D. Altman. Development and testing of paraffin-based hybrid rocket fuels. In *37th AIAA/ASME/SAE/ASEE Joint Propulsion Conference and Exhibit*, Salt Lake City, Utah, July 2001. American Institute of Aeronautics and Astronautics.
9) A. Karabeyoglu, J. Stevens, D. Geyzel, B. Cantwell, and D. Micheletti. High performance hybrid upper stage motor. In *47th AIAA/ASME/SAE/ASEE Joint Propulsion Conference and Exhibit*. American Institute of Aeronautics and Astronautics, July 2011.
10) M. Kobald, A. Petrarolo, and S. Schlechtriem. Combustion visualization and characterization of liquefying hybrid rocket fuels. In *51st AIAA/SAE/ASEE Joint Propulsion Conference*. American Institute of Aeronautics and Astronautics, July 2015.
11) S. Lloyd. Least squares quantization in pcm. *IEE Transactions on Information Theory*, 28(2):129–137, 1982.
12) J. MacQueen. Some methods for classification and analysis of multivariate observations. In *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Statistics*, pages 281–297, Berkeley, Calif., 1967. University of California Press.
13) A. Petrarolo and M. Kobald. Evaluation techniques for optical analysis of hybrid rocket propulsion. *Journal of Fluid Science and Technology*, 11(4):JFST0028–JFST0028, 2016.
14) A. Petrarolo, M. Kobald, H. K. Ciezki, and S. Schlechtriem. Principal and independent component analysis of hybrid combustion flame. *International Journal of Energetic Materials and Chemical Propulsion*, 18(1):pp. 9–29, 2019.
15) A. Petrarolo, M. Kobald, and S. Schlechtriem. Understanding Kelvin-Helmholtz instability in paraffin-based hybrid rocket fuels. *Experiments in Fluids*, page 59:62, 2018.
16) D. Pham, S. Dimov, and C. Nguyen. Selection of k in k-means clustering. *Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science*, 219(1):103–119, 2005.
17) R. W. Schefer. Flame sheet imaging using CH chemiluminescence. *Combustion Science and Technology*, 126: 1-6:255–279, 2010.
18) M. Steinbach, L. Ertöz, and V. Kumar. The challenges of clustering high dimensional data. In *New directions in statistical physics*, pages 273–309. Springer, 2004.
19) A. Thumann and H. K. Ciezki. *Combustion of Energetic Materials, chap. Comparison of PIV and Colour-Schlieren Measurements of the Combusiton Process of Boron Particle Containing Soild Fuel Slabs in a Rearward Facing Step Combustor*, volume Vol. 5, pp. 742-752. Begell House Inc., 2002.