On the chemical composition and structure of incipient soot in a laminar diffusion flame

Jessy Elias, Alessandro Faccinetto, Cornelia Irimiea, Nicolas Nuns, Claire Pirim, Cristian Focsa, Hervé Vezin, Xavier Mercier

On the chemical composition and structure of incipient soot in a laminar diffusion flame

J. Elias[1,2], A. Faccinetto[1,], C. Irimiea[3], N. Nuns[4], C. Pirim[5], C. Focsa[5], H. Vezin[6], X. Mercier[1]*

[1]*Univ. Lille, CNRS, UMR 8522, PC2A, F-59000 Lille, France*

[2]*ADEME, F-49004, Angers, France*

[3]*DMPE, ONERA, Univ. Paris Saclay, F-91123 Palaiseau, France*

[4]*Univ. Lille, M. E. Chevreul Institute, F-59000 Lille, France*

[5]*Univ. Lille, CNRS, UMR 8523, PhLAM, F-59000 Lille, France*

[6]*Univ. Lille, CNRS, UMR 8516, LASIRe, F-59000 Lille, France*

*Corresponding author

Abstract

The transformation of molecular precursors in the gas phase into nanoparticles in the condensed phase (*soot inception*) during combustion has not yet been fully understood. While several hypotheses on soot inception are currently being examined, their validation requires detailed knowledge of the physical-chemical properties of the matter first formed in the condensed phase (*incipient soot*). However, characterizing short-lived species in a reactive environment at high temperature is inherently difficult. In this work, we propose a multi-analytical approach to indirectly characterize chemical composition and structure of incipient soot obtained in controlled laboratory conditions. The results are explained in light of recent developments on the characterization of the optical and magnetic properties of incipient soot, and used to build a phenomenological model of soot inception. The acquired information is key to support the hypothesis selection and to understand the reactivity of combustion byproducts and thus their impact on health and the environment.

Keywords

Soot inception, incipient soot, multi-diagnostics characterization, structure, chemical composition, diffusion flames, ToF-SIMS, Raman spectroscopy.

1. Introduction

Soot is made of fractal-like aggregates of carbonaceous nanoparticles formed during the incomplete combustion of hydrocarbons. A vast body of literature has been devoted to understanding the fundamental mechanisms governing the transition from the gas to the condensed phase (*soot inception*) in hydrocarbon flames. As a result, a number of hypotheses on the main mechanisms of soot inception are currently being under validation by the scientific community [1–6]. Incipient soot is defined as the matter first formed in the condensed phase from molecular precursors like polycyclic aromatic hydrocarbons (PAHs) and their derivatives [7]. Over the years, it became clear that detailed knowledge of the physical-chemical properties of incipient soot is key to understand soot inception. However, it also became evident that many physical-chemical properties of incipient soot significantly differ from mature soot. In addition to being fundamental for the hypothesis selection, these differences affect the interpretation of the output of several diagnostics, and therefore must be considered for the accurate interpretation of the experimental data.

During the last two decades, a remarkable number of investigations on the formation and characterization of incipient soot have been published, most focusing on the soot nanoparticles size distribution and structure. Unlike primary particles that can reach diameters up to tens of nanometers and are characterized by a classic core-shell internal nanostructure [8,9], incipient particles can be as small as 1-3 nm [10,11] and appear to be constituted of 2-3 atomic layers [12] of either amorphous or randomly oriented carbonaceous clusters [13–15]. The core of primary particles has been shown to be compatible with the coagulation of 2-3 incipient particles [8,16], suggesting that incipient particles may grow into primary particles. However, the coagulation rate of incipient particles in flames is generally sufficiently low [17–19] that they can coexist and be detected along with larger particles [20–24].

The specific heat capacity and density [25] and the complex refractive index function [26–29] of incipient soot are significantly lower than mature soot. Incipient soot is characterized by weak absorption

in the visible light spectrum, and lack the strong thermal radiation emission that can be fitted with Planck's law typical of mature soot [30,31].

In contrast, little and mostly indirect information on the chemical composition and structure of incipient soot is currently known. Implementing the diagnostics required to detect and analyze short-lived species or nanoparticles in a complex reactive environment at high temperature and interpreting the resulting data is a challenging task, especially since sampling is required for online and ex situ analyses. Incipient soot is richer in small PAHs [32] and aliphatics [33,34] than mature soot, although significant differences exist between diffusion and premixed flames [35]. Direct evidence on the structure of isolated molecules thermally desorbed from incipient soot has been recently produced [36,37]. The higher H/C ratio [9,38,39] and smaller particle size [14,40] of incipient soot result in different chemical reactivity [41–43] and hygroscopic properties [44,45] than mature soot.

To address this lack of fundamental knowledge, in this work we present the fine characterization of the chemical composition and structure of incipient soot generated in a co-flow laminar diffusion flame of methane. This flame has already been extensively characterized in our previous works using different in situ and ex situ diagnostics (*Methods* 2.1-2.2). The soot and PAH regions have already been mapped in situ, and used to select the sampling points that are investigated in this work by time of flight secondary ions mass spectrometry (ToF-SIMS, *Methods* 2.3) and Raman spectroscopy (*Methods* 2.4-2.5). The data analysis and interpretation rely on principal component analysis (PCA, *Methods* 2.6) to infer information on the properties of incipient soot: the principal components representative of the chemical composition and of the structure are first identified (*Results* 3.1-3.4), then cross-compared (*Results* 3.5), and finally correlated to the data obtained in situ or already available in the literature in a thorough multi-diagnostic approach (*Discussion* 4.1). In particular, the atomic fraction of hydrogen obtained from the analysis of the mass spectra [35] has been suggested to be linked to the formation rate of covalent C-C bonds [46]. The structure of incipient soot is derived from the detailed analysis of the Raman spectra and is shown to evolve as a function of the appearance of curved graphene layers, the presence of finite-size carbon crystallites, and the changing distance between graphene-like layers. These results are discussed in the

framework of the recent advancements in the characterization of the concentration and structure of persistent radicals in incipient soot [47], and more specifically of our recent electron paramagnetic resonance (EPR) measurements [48]. The cross-examination of these properties allows the identification of several properties of the chemical species involved in the phase transition that are used to propose a phenomenological model of soot inception (*Discussion* 4.2).

2. Materials and methods

2.1 Flame and sampling

The investigated flame is a 120 mm high axisymmetric co-flow laminar methane/air diffusion flame stabilized on a burner at atmospheric pressure. The burner consists of a central injector supplied with 0.52 L min^{-1} of methane and surrounded by an 86.6 L min^{-1} air shield (standard T and p). Samples are extracted from the flame axis using a dilutive quartz microprobe inserted radially into the flame and operated with nitrogen to reduce the collision probability in the sampling line and quench the reactivity in the sampled flow. The sampling scheme is shown in the supporting information (SI 8.1) where it is compare to the mapping of the soot volume fraction obtained by LII as detailed below. The collected species are deposited on Ti wafers with the same 1/8″ diameter impactor used in our previous works [35,46,49]. High-velocity impaction (~34 m s^{-1}) results in carbonaceous particles aggregating at the impaction site at the center of the wafer and in the condensable gas scattering all over the wafer surface [46]. ToF-SIMS and Raman measurements are performed on two selected ROIs: the materials deposited directly at the impaction site (impaction ROI), and the materials scattered on the wafer surface surrounding it (halo ROI).

2.2 In situ characterization of the flame

Soot particles have already been mapped using in situ laser induced incandescence (LII) [50]. LII measurements are performed at 1064 nm using the fundamental of a Nd:YAG laser (Quantel Brilliant B). The LII is recorded at 90° with respect to the laser beam using a fast ICCD camera (Princeton Instruments

PI-MAX2, Roper Scientific). The LII signal is summed over the 580-800 nm spectral range. Soot particles are heated below the sublimation threshold at 130 mJ cm^{-2}.

PAHs [50] and OH radicals [51] have already been mapped in this flame using in situ laser induced fluorescence (LIF). Briefly, LIF measurements are performed at several excitation wavelengths in the UV-Vis spectral range. The excitation source is a Nd:YAG laser (Quanta-ray, Spectra Physics) used to pump an optical parametric oscillator (premiScan-ULD/240, GWU-Lasertechnik) to provide tunable wavelengths at 213-532 nm. The PAH fluorescence spectra, induced at 532 nm, are recorded using a fast ICCD camera (PI-MAX2, Roper Scientific) coupled to a spectrometer (IHR320, Jobin Yvon). To avoid any interference with the LII, the laser fluence is set to 13 mJ cm^{-2}, which is below the activation threshold of the LII. The PAH fluorescence is summed over the 550-700 nm spectral range. OH-LIF spectra are obtained by exciting the A-X (0-0) band (309-312 nm) and by collecting the fluorescence signal in the R bands spectral range (305-308 nm). The emitted fluorescence is collected at 90° and imaged with a spectrometer (IHR320, Jobin Yvon) coupled to a photomultiplier tube (XP2020Q, Photonics). The anti-Stokes collection scheme minimizes potential spectral interferences with the intense PAHs fluorescence above 300 nm. The results obtained in the previous works are used to establish a sampling scheme and are critical for the interpretation of the results.

2.3 Secondary ions mass spectrometry (ToF-SIMS)

Bi_3^+ primary ions are used to sputter and ionize the species on the samples surface. The secondary ions are accelerated and analyzed on a time-of-flight mass spectrometer with a maximum resolving power $m/\Delta m \approx 10^4$ at 200 m/z. This technique is sensitive to high mass organic compounds ($m/z > 200$) due to the use of primary Bi_3^+ ions having 25 keV energy with very low current intensity of 0.3 pA (static mode). Mass spectra in positive polarity are recorded at 50 scans per acquisition on a 500 × 500 μm^2 surface with an image resolution of 128 × 128 pixels. Acquisitions are performed on three different zones on each ROI, halo and impaction. The information extracted from the mass spectra includes: the time-of-flight of 520 peaks selected after background and fragment ions removal that is calibrated into

m/z and used to identify unknown species, and the peak normalized ion counts I_n defined as $I_n = I/I_{TIC}$ where I is the peak ion count and I_{TIC} the total relevant ion count (total ion count after background removal). Peaks are identified using mass defect analysis, allowing the assignment of a molecular formula to the selected m/z following the protocol established in our previous works [35,46,49]. The mass spectra (SI 8.1), the mass defect plot with the complete peak list (SI 8.3) is available in the supporting information.

The atomic fraction of hydrogen [H] is calculated from the mass spectra using the peaks identified by mass defect analysis, where the normalized signal intensity w_i is used to weight the individual contributions of the identified ions $N_{X,i}$ [46,52]:

$$[\mathrm{H}] = \frac{\mathrm{N_H}}{\mathrm{N_H} + \mathrm{N_C} + \mathrm{N_O}} \qquad \text{Equation 2.1.}$$

With $$\mathrm{N_X} = \sum_i \mathrm{N_{X,i}} w_i \ , \ \mathrm{X} = \mathrm{H, C, O} \ , \ \sum_i w_i = 1 \qquad \text{Equation 2.2.}$$

2.4 Raman spectroscopy

The Raman microscope (InVia Reflex, Renishaw) is equipped with a $\lambda = 514.5$ nm diode-pumped solid state laser. The laser power is reduced to 0.05%, corresponding to a low laser irradiance of 1.63 kW cm^{-2}, to avoid structural changes in the sample. The laser beam is focused on the samples with a 20× magnification objective. The Raman spectral intensity and wavelength are calibrated with a Si wafer by using the pure Si peak at 520 cm^{-1}. The same instrument settings (600 lines mm^{-1} grating, acquisition time 6 s times 80 accumulations per spectrum), chosen after a set of test measurements to provide the best signal-to-noise ratio for the analysis, are used for all samples. In particular, for samples collected at low flame HAB, long acquisition time and a preliminary analysis performed on all samples have been deemed necessary to find the optimal laser irradiance to record high-quality Raman spectra, the existence of which is acknowledged in our previous work but not explored because of the insufficient signal-to-noise ratio [46]. About 30 Raman spectra per ROIs are recorded between 700 and 2500 cm^{-1} for each sample.

2.5 Analysis of the Raman spectra

The first-order Raman spectra of carbonaceous materials are generally characterized by two main spectral features, labeled D and G bands, at around 1350 and 1600 cm^{-1}. As detailed in Table 1, an extensive body of literature has been dedicated to analyzing correlations between chemical and structural properties of graphene, graphite, highly oriented pyrolytic graphite, carbon films, and soot to some of the band parameters. [H] is often calculated from the ratio between the slope of the photoluminescence baseline and the height of the G band [53,54]. Likewise, the degree of graphitization, quantified through the size of the aromatic domain, has been estimated from the D-to-G band height ratio [11,36,55,56]. The degrees of disorder and chemical maturity have also been estimated from the shift and broadening of the G band that seems to depend on the crystallite size [57]. These approaches are relatively simple to implement but not always optimal for flame-sampled materials. To extract more detailed structural information, fittings with four [58], five [55,59–61] and up to six [62] different peaks have been discussed in the literature. However, the peak fitting is a challenging task since the D and G bands are broad, partially overlap, fold in several contributions, and therefore the exact number and origin of the active Raman modes are difficult to assess. A common approach to the spectral analysis uses the complete set or subsets of the peaks listed in Table 1.

Following this approach, the Raman spectra obtained in this work are analyzed by curve fitting using different combinations of the peaks listed in Table 1. The curve fitting is performed with the peak analyzer tool of OriginPRO 2021 (Levenberg-Marquardt algorithm, χ^2 tolerance value of 10^{-6} to reach convergence). The best results ($R^2 > 0.995$) are obtained with six peak fits: D4, D1, D3, G, D2 [59] expanded to include the peak at 1222 cm^{-1} [62] here labeled D5 for naming consistency. An example of the peak fitting is shown in Figure 1. Peaks D4 and D5 are both required to optimally fit the left wing of the D band, peak D3 is necessary for the otherwise underrepresented local minimum between bands, and peaks G and D2 are required to reproduce the asymmetrical shape of the G band.

In the analysis of Raman spectra of carbonaceous materials, the variability range of the peak centers and widths is sometimes fixed or limited for the fitting to reach convergence [55,63]. However, as shown in Table 1, some peak assignments are still under debate, therefore limiting the variability range of the spectral parameters can ultimately mislead the data interpretation. In particular, the choice of the photoluminescence baseline impacts the determination of both the peak centers and widths. In the literature, the photoluminescence background is either absent [62], not discussed [46,55,59], or subtracted [58,64,65] before fitting. In this work, third-degree polynomial functions are found to be the lowest degree polynomials that approximate with high confidence ($R^2 > 0.995$) the rising shape of the photoluminescence background in the spectral region 800-2100 cm^{-1}. The polynomial coefficients and constant term are set as free parameters and fitted with the peaks: this approach has the important practical advantage that during the fitting it is not necessary to limit peak centers and widths to reach convergence.

Table 1. Nomenclature and known or proposed assignments of the peaks used for the spectral analysis of the first-order Raman spectra of carbonaceous materials.

Peak	Observed Raman shift / cm^{-1}				Line shape	Bibliographic references
	Sadezky et al. [59]	Minutolo et al. [62]	Parent et al. [55]	Carpentier et al. [64]		
D4	~1200	1176(4)	1180(5)	1175-1190	Lorentz	Assignment under debate. Characteristic of highly disordered materials: polyenes [66], ionic impurities [67], peripheric sp^2 or sp^3 carbon atoms [68].
D5	/	1222(6)	/	1269	Lorentz	Assignment under debate. Finite crystal size or defects [69,70]. Poorly linked polyaromatic units [64]. Already observed in flame-sampled materials [62].
D1	~1350	1340(4)	1350(5)	1337-1372	Lorentz	Certain assignment. A_{1g} ("breathing") mode of graphene only active in case of lattice disturbance (edges due to finite layers, heteroatoms) [71–74].
D3	~1500	1529(4)	1550(5)	1505-1544	Gauss	Assignment under debate. Possibly a convolution of several other modes [55]. Amorphous carbon [66,67,75], finite-size

						graphite crystals [76], defects outside graphene layers [77]. A detailed discussion is available [62].
G	~1580	1598(4)	1580(5)	1600	Lorentz	Certain assignment. E_{2g} mode of ideal graphite lattice [71,73].
D2	~1620	1618(4)	1620(5)	1616-1618	Lorentz	Certain assignment. Layers of graphene not directly sandwiched between other layers [55,62,73].

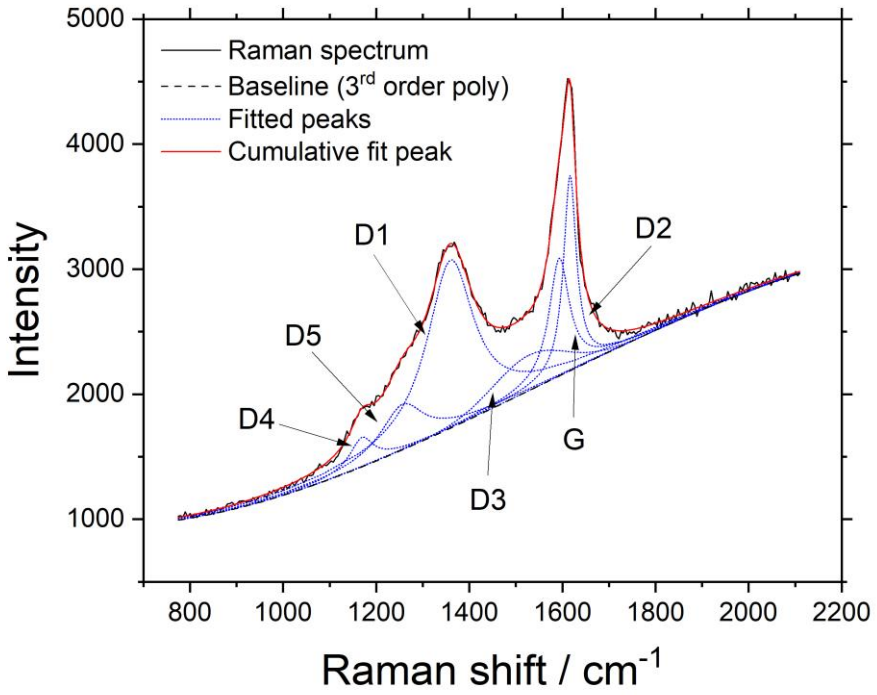


Figure 1. Raman spectra analysis. Raman spectrum obtained from a sample collected at 55 mm HAB (514.5 nm excitation wavelength, 1.63 kW cm^{-2} laser irradiance) and six-peak fitting. The figure shows the acquired spectrum (solid black line) and the spectral analysis including the photoluminescence baseline (dashed black line), the peaks D4, D5, D1, D3, G and D2 (dotted blue lines) and the cumulative fit curve (solid red line).

A detailed discussion on the calculation of [H] from the Raman spectra and comparison with other works (SI 8.5) and on the photoluminescence background (SI 8.6) are available in the supporting information.

2.6 Principal components analysis

Nowadays, the level of complexity of the output of many diagnostics makes it difficult to rely on intuition or visual analysis to extract trends from raw data. Therefore, statistical tools like multivariate analysis become increasingly useful to extract any meaningful chemical information from complex databases.

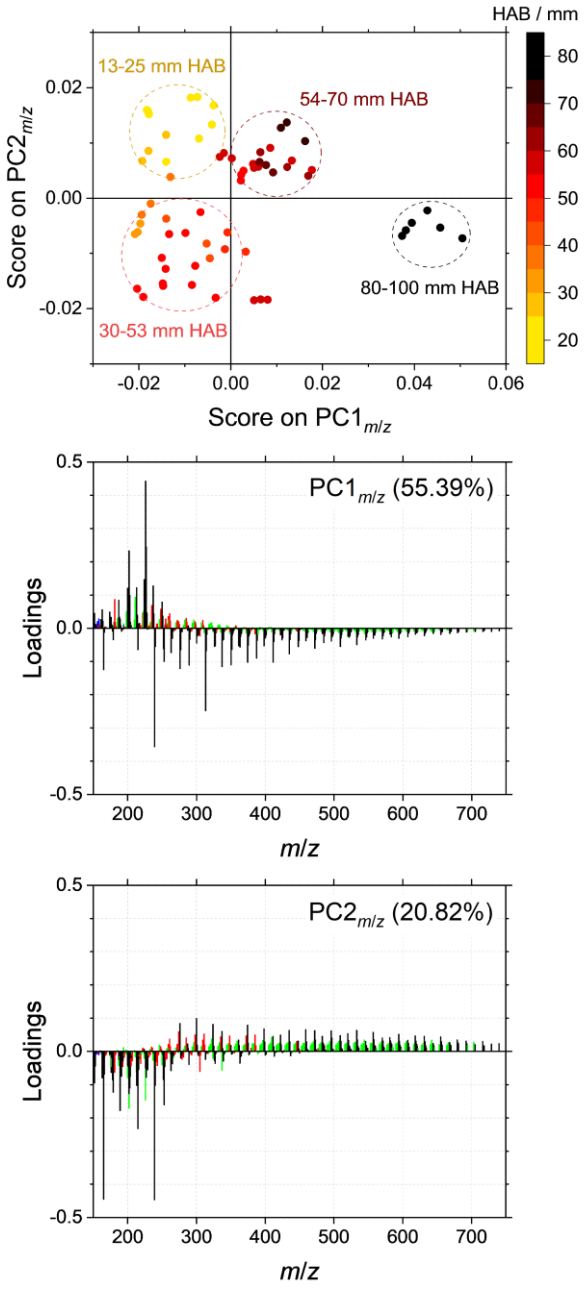
Principal component analysis (PCA) is used to reduce the dimensionality of the mass spectra and Raman spectra databases and to assist the interpretation of the variance [78,79]. The PCA shows the changes of the variance of a set of observables calculated from a set of variables. The loading plots find the variables that have the largest effect on each principal component. Therefore, if a principal component can be linked to a specific physical phenomenon, the PCA indicates the variables upon which this phenomenon depends. In this work, the m/z from the ToF-SIMS mass spectra or the spectral parameters from the peak fitting of the Raman spectra are used as variables, while the data acquisitions on the sampling zones and various ROIs are used as observables. The additional information on the physical phenomena is provided by the complementary diagnostics (LII and LIF that map the formation regions of soot and molecular precursors, and EPR that provides the spin concentration), and by the presence of visible soot deposits on the analyzed substrates. PCA on non-similar quantities (positions and widths of the Raman peaks) is performed on the correlation matrix. PCA on similar quantities (ToF-SIMS ion counts normalized to the total relevant ion count and Raman peak areas normalized to the total area of the spectrum after removing the photoluminescence background) is performed on the covariance matrix. The results are then discussed in terms of scores and loadings [35,80,81], with the primary goal of finding correlations between the chemical composition (ToF-SIMS mass spectra) and the structure (Raman spectra) of the samples.

The four principal components of the PCA on mass spectrometry data (SI 8.4) and on Raman spectroscopy data (SI 8.7) are shown in the supporting information.

3. Results

3.1 The chemical composition of soot

A total of 132 ToF-SIMS mass spectra collected at 21 different HABs are analyzed. The data recorded on impaction ROIs show the contributions of both soot (only present above the soot inception HAB) and condensable gas, while the data recorded on the surrounding halo ROIs only show the contribution of condensable gas. The information extracted from the mass spectra includes the m/z and the normalized ion count I_n of 520 selected peaks. Rather than working with the complete set of detected peaks, scalar variables are built to represent trends in the mass spectra. To this end, PCA is a particularly useful tool since the principal components are already defined as linear combinations of the original variables (I_n), and can be related to changes of the variance through the score plots [79]. The results of the PCA on the complete database of mass spectra are discussed in the exploratory data analysis (SI 8.8). The results of the PCA on the impaction ROIs are shown in Figure 2. The $PC2_{m/z}$ against $PC1_{m/z}$ score plot lays the foundation for the discussion since $PC1_{m/z}$ and $PC2_{m/z}$ represent a significant fraction of the total variance (76.21%) of the chemical composition. The score plot reveals several data clusters representing samples having sufficiently different chemical composition to be classified on either $PC1_{m/z}$ or $PC2_{m/z}$. It should be stressed that the interpretation proposed herein is based on the analysis of $PC1_{m/z}$ and $PC2_{m/z}$, which, while responsible for the highest contributions, are by no means the only contributors to the variance of the database.



240

241 **Figure 2. Results of the PCA (covariance matrix) performed on the peak normalized areas extracted from the**
242 **impaction ROIs mass spectra.** The figure shows the $PC2_{m/z}$ against $PC1_{m/z}$ score plot with data cluster attributions (top
243 panel) and the corresponding loading plots (middle and bottom panels). The loadings are color-mapped on the assigned
244 formula: $C_mH_n^+$ (black), $C_mH_nO_p^+$ (red) and unassigned (green).

245

3.2 High m/z hydrocarbons ($PC1_{m/z}$)

On $PC1_{m/z}$ (55.39% of the total variance explained) three data clusters can be identified as shown in Figure 2. All mass spectra at 13-53 mm HAB (yellow-orange datapoints) exhibit negative scores on $PC1_{m/z}$. The high dispersion of the datapoints on $PC1_{m/z}$ indicates that the PCA cannot find statistically significant differences among the mass spectra of the samples collected in this HAB range. The mass spectra at 54-70 mm HAB (red-brown datapoints) are closely clustered with low positive scores in $PC1_{m/z}$, except the mass spectra at 56 mm HAB with close-to-zero negative scores. Finally, the mass spectra at 80-100 mm HAB (black datapoints) exhibit high positive scores in $PC1_{m/z}$. The *m*/*z* contributing to $PC1_{m/z}$ are shown in the corresponding loading plot in Figure 2. A sign inversion of the loadings occurs in the 240-300 *m*/*z* range (corresponding to C19-C24 hydrocarbons), which essentially divides the loadings into two groups: high *m*/*z* ions having high negative loadings, and low *m*/*z* ions having high positive loadings. The former are almost invariably identified as $C_mH_n^+$ by mass defect analysis and their molecular formulae are consistent with PAHs or their derivatives, whereas the latter are identified as $C_mH_n^+$, $C_mH_nO_p^+$ and other unassigned ions by mass defect analysis and their molecular formulae are consistent with PAHs, their derivatives or oxygen-containing hydrocarbons. In the score plot, a sign inversion of the loadings of $PC1_{m/z}$ occurs between 53 and 56 mm HAB that calls attention to a change in the chemical composition of the samples. This change, consisting of the depletion of high *m*/*z* PAHs and their derivatives in favor of lower *m*/*z* PAHs, their derivatives, and oxygen-containing hydrocarbons, is the first important information extracted from the analysis of the chemical composition of the samples.

3.3 The atomic fraction of hydrogen ($PC2_{m/z}$)

On $PC2_{m/z}$ (20.82% of the variance explained) two data clusters can be identified as shown in Figure 2. Mass spectra at 30-53 mm HAB (orange-red datapoints) and 80-100 mm HAB (black datapoints) exhibit negative scores, whereas mass spectra at 13-25 mm HAB (yellow datapoints) and 54-70 mm HAB (brown datapoints) exhibit positive scores. The *m*/*z* contributing to $PC2_{m/z}$ are also shown in the corresponding loading plot in Figure 2. Again, a sign inversion of the loadings occurs approximately in the 240-300 *m*/*z*

range. Here, high *m*/*z* ions have loadings close to zero and therefore only marginally contribute to the phenomenon represented by $PC2_{m/z}$, whereas low *m*/*z* ions having high negative loadings are predominantly identified as $C_mH_n^+$.

[H] is linearly correlated to $PC2_{m/z}$ ($R^2 = 0.9223$) for samples collected up to 70 mm HAB, as shown in Figure 3. High [H] corresponds to high negative scores on $PC2_{m/z}$, i.e. to samples collected at 30-53 mm and 80-100 mm HAB. As further discussed below, this finding is particularly interesting as it shows that samples at 30-53 mm HAB have the highest concentration of hydrogen potentially available for chemical reactivity. Furthermore, the highest negative loadings on $PC2_{m/z}$ correspond to hydrocarbons in the range 150-260 *m*/*z*. Higher *m*/*z* species, being either hydrocarbons, oxygen-containing hydrocarbons or other unknown species, tend to show significantly lower loadings (see Figure 2) and thus give smaller contributions to $PC2_{m/z}$. Above 80 mm HAB the linear correlation between [H] and $PC2_{m/z}$ is lost, possibly because of the oxidation processes becoming dominant as further discussed below.

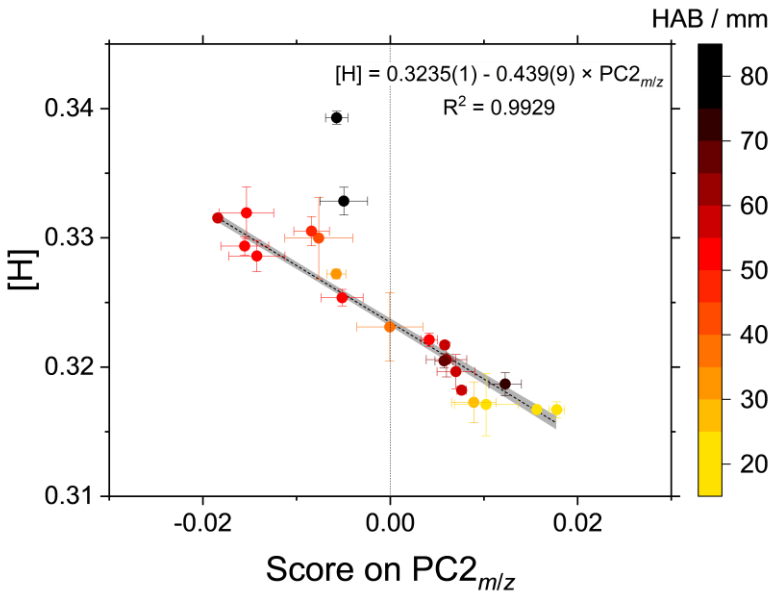


Figure 3. [H] against $PC2_{m/z}$. The linear regression (black dashed line) and the 95% confidence interval (gray area) do not include the data collected at 80-100 mm HAB (two isolated black datapoints), for which the correlation is lost.

3.4 Raman spectroscopy: a challenging interpretation

The first-order feature of the Raman spectra, consisting of the D and G bands in the 800-2100 cm^{-1} range, is analyzed by peak fitting. Extracting information from the Raman spectra of carbonaceous materials is no trivial task: the D and G bands are broad and partially overlap, and therefore the exact number and origin of the active Raman modes are still debated. The approach proposed herein is based on a six-peak fitting (D4, D5, D1, D3, G, D2) and focuses on finding correlations between the spectral parameters but avoiding existing empirical relationships that might not be valid for the entire range of the investigated soot properties, especially for incipient soot.

A total of 790 Raman spectra of impaction ROIs collected at 22 different HABs are acquired. The extracted spectral parameters include: the peak normalized area A_n defined as $A_n(X) = A(X)/\Sigma_i A(X_i)$, where $A(X)$ is the area of peak X after subtracting the photoluminescence baseline, the peak position x_c and the width w_{FWHM} taken as the full width at half maximum (FWHM). Once again, PCA is used to reduce the dimensionality and to find statistically significant correlations; the discussion is based on the $PC2_{An}$ against $PC1_{An}$ score plot.

Figure 4 shows the results of the PCA performed on the A_n for all impaction ROIs. The scores on $PC1_{An}$ (66.48% of the variance explained) classify the datapoints in three data clusters based on their sampling HAB: positive scores at 15-53 mm HAB, low negative scores at 54-65 mm HAB, and high negative scores at 70-100 mm HAB. These distinct data clusters represent the tendency of A_n to assume constant values at 15-53 mm HAB (yellow-red datapoints), then to transition through 54-65 mm HAB (brown datapoints) to eventually reach different constant values at 70-100 mm HAB (black datapoints). The loading plot of $PC1_{An}$ shows that different peaks provide opposing contributions to $PC1_{An}$. Peaks D4 and D3 have negative loadings: in particular, peak D3 has high negative loading, i.e. it contributes extensively to the phenomenon represented by $PC1_{An}$, and A_n(D3) shows the largest relative difference of all peaks between 15-53 mm HAB and 70-100 mm HAB. Peaks D5, G and D2 have low positive loadings, hence

their contribution to $PC1_{An}$ opposes that of peaks D4 and D3. Finally, peak D1 has loading close to zero, and therefore contributes only marginally to the phenomenon represented by $PC1_{An}$.

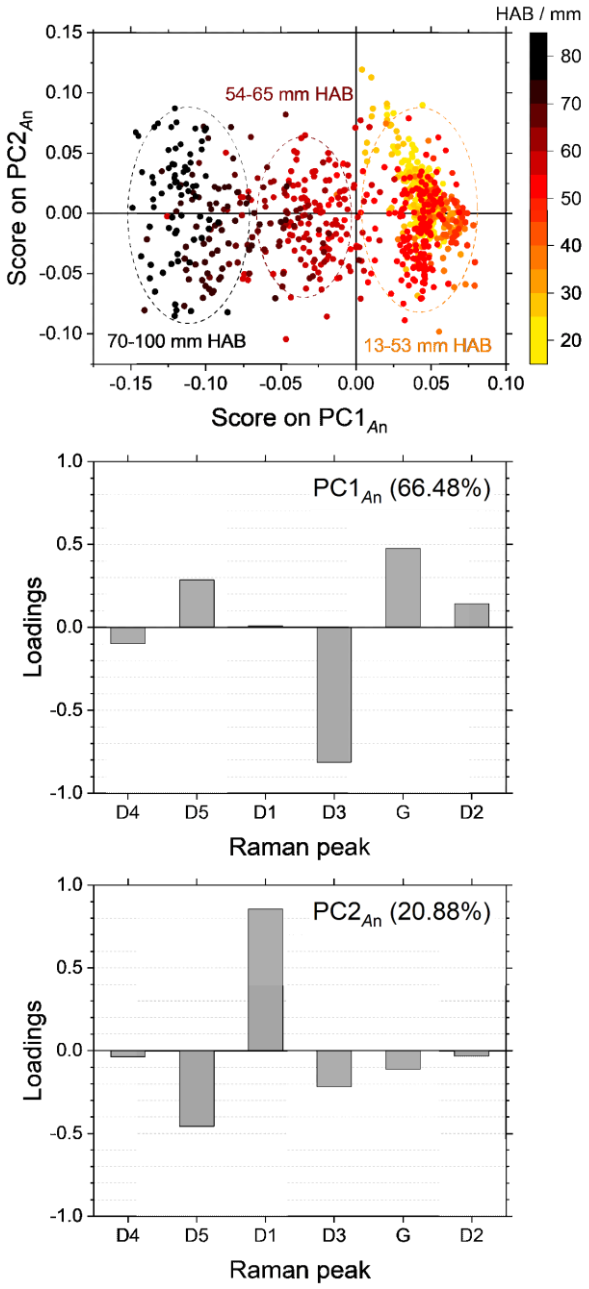
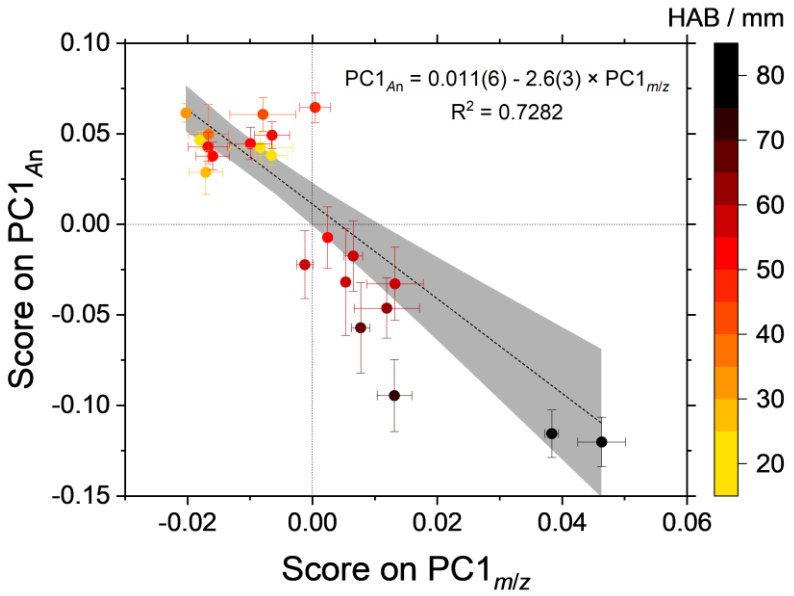


Figure 4. Results of the PCA (covariance matrix) performed on the peak normalized areas A_n extracted from the impaction ROIs Raman spectra. The figure shows the $PC2_{An}$ against $PC1_{An}$ score plot with data cluster attributions (top panel) and the corresponding loading plots (middle and bottom panels).

318 The higher-order principal components do not show any appreciable dependence on the HAB or other
319 identified properties of the samples.
320 The second parameter derived from the fittings, x_c, is independent of both the flame and sampling
321 conditions. The data scores overlap in the same region of the score plots and cannot be clearly clustered
322 (SI 8.7), with the only exception of a small data cluster having high positive scores in the third principal
323 component that is consequence of a ~10 cm^{-1} shift to a higher wavenumber of the D band occurring in
324 samples collected at 40-45 mm HAB. A similar behavior has already been observed in materials sampled
325 from premixed flames [11]: there, the authors remark that strains in the lattice or doping from electrical
326 charges could be at the origin of band shifts in both the D and G bands, and suggest that it could be
327 evidence of a transition from a planar lattice to the bent crystallites typical of soot nanoparticles.
328 Akin to what has been previously observed for A_n, the last parameter w_{FWHM}, also shows a separation in
329 three data clusters that represent the tendency of all peaks except D5 to broaden in samples collected at
330 54-70 mm HAB (SI 8.7). This peak broadening is consistent with the global broadening of the Raman
331 spectra, with D and G bands spanning over 1000-1800 cm^{-1} in samples collected at 15-53 mm HAB and
332 over 900-1900 cm^{-1} in those collected at 70-100 mm HAB.

333 **3.5 Linking structure and chemical composition**

334 As emerged from the data analysis so far, $PC1_{m/z}$ (high m/z hydrocarbons), $PC2_{m/z}$ ([H]) and $PC1_{An}$
335 (Raman peaks D5, D3 and G) are convenient choices to represent changes of the chemical composition
336 and structure of the samples. As shown in Figure 5, $PC1_{m/z}$ and $PC1_{An}$ are correlated: high negative scores
337 on $PC1_{m/z}$ correspond to high positive scores on $PC1_{An}$ (13-53 mm HAB), low positive scores on $PC1_{m/z}$
338 correspond to low negative scores on $PC1_{An}$ (54-70 mm HAB), and high positive scores on $PC1_{m/z}$
339 correspond to high negative scores on $PC1_{An}$ (80-100 mm HAB).



340

341 **Figure 5. Linking structure and chemical composition.** $PC1_{An}$ against $PC1_{m/z}$ plot. The linear fitting (black dashed line)
342 and the 95% confidence interval (gray area) are shown. In this and all following figures, the averages calculated on all the
343 available ROIs are compared since it was not possible to perform ToF-SIMS and Raman analyses on the same locations on
344 the sample surface.

345

346 The sign inversion of $PC1_{m/z}$, which occurs approximately in the 240-300 m/z range and corresponds to
347 C19-C24 hydrocarbons, represents the depletion of high m/z hydrocarbons ($PC1_{m/z} < 0$). The Raman
348 peaks affected by this change of the chemical composition are identified by comparing the scores on
349 $PC1_{m/z}$ to the A_n of the individual peaks obtained from the curve fitting of the Raman spectra. As shown in
350 Figure 6, several behaviors of A_n against $PC1_{m/z}$ are identified.

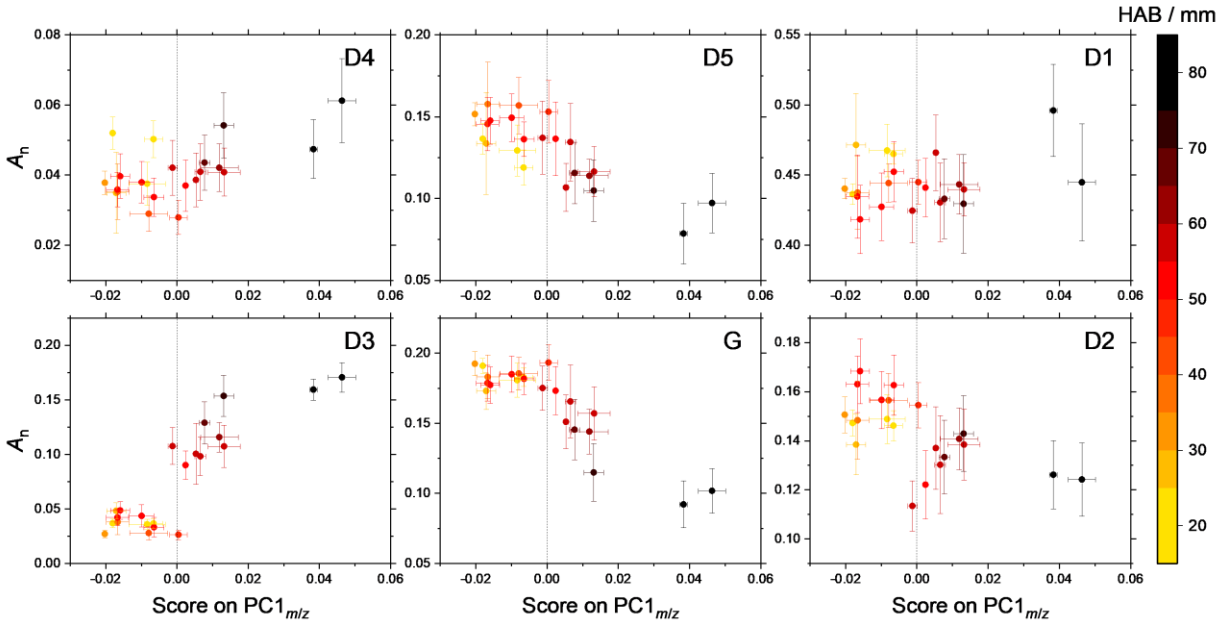


Figure 6. A_n (Raman spectral analysis) against scores on PC1$_{m/z}$ (ToF-SIMS analysis).

A_n(D4) is weakly correlated to PC1$_{m/z}$. It is not clear whether a threshold at the sign inversion of PC1$_{m/z}$ occurs. While the origin of peak D4 is not yet understood, it has been suggested to be related to the presence of polyenes [66], peripheric carbon atoms [59,66,68] or ionic impurities [67]. In any case, peak D4 seems to be only weakly affected by the presence of high *m*/*z* hydrocarbons.

A_n(D5) is correlated to PC1$_{m/z}$, suggesting a dependence on the presence of high *m*/*z* hydrocarbons. Although peak D5 has already been observed in flame-sampled materials [62], only little information is available on its origin. Peak D5 has been proposed to be activated by the relaxation of the selection rules from finite crystal size effects or defects [69,70] or the merging of Raman active modes of poorly linked polyaromatic units [64,82].

A_n(D1) is uncorrelated to PC1$_{m/z}$. Peak D1 has been assigned to the A_{1g} ("breathing") mode of graphene that is only active in case of lattice disturbances like edges or the presence of heteroatoms [71–74,83], and has already been established in the literature as a useful marker for mature soot [55,62,84,85]. In this work, peak D1 is not found to be an effective marker for the presence of high *m*/*z* hydrocarbons.

A_n(D3) is correlated to $PC1_{m/z}$, and furthermore the data clustering already observed in $PC1_{An}$ is preserved. In particular, the sharp transition from 13-53 mm HAB (yellow-orange datapoints) to 54-65 mm HAB (red-brown datapoints) corresponds to the sign inversion of $PC1_{m/z}$, i.e. to the depletion of high *m*/*z* hydrocarbons. The origin of peak D3 is not yet fully understood: in previous works, it has been attributed to the presence of amorphous carbon [66,67,75], finite-size graphite crystals [76], or defects outside the graphene layers [77]. Density functional calculations suggest that it might be the convolution of several Raman active modes [55]. New insights on the origin of peak D3 come from EPR measurements. Narrow EPR linewidths are characteristic of amorphous carbon or dehydrogenation during maturation [86]. Figure 7a shows the remarkable correlation between the EPR linewidth of soot samples collected from the same flame analyzed in this work [48] and the corresponding A_n(D3) that supports the attribution of peak D3 to vibrations of amorphous carbon.

A_n(G) is correlated to $PC1_{m/z}$ with a well-defined threshold behavior. A_n(G) is constant for negative $PC1_{m/z}$ at 13-53 mm HAB, then suddenly decreases at the sign inversion of $PC1_{m/z}$. Peak G has been unequivocally assigned to the vibration mode E_{2g} of the ideal graphitic lattice [71,73]. The correlation between the EPR linewidth and A_n(G) shown in Figure 7b suggests that the transition to amorphous carbon is associated with the reduction of the size of the graphitic lattice.

A_n(D2) is weakly correlated to $PC1_{m/z}$. Peak D2 has been assigned to isolated graphene layers not directly sandwiched between other layers [55,62,73].

No clear trends emerge when comparing the individual A_n to $PC2_{m/z}$ (SI 8.9). A_n(D4) shows a weak correlation to $PC2_{m/z}$, while A_n(D5), A_n(D1), A_n(D3), A_n(G) and A_n(D2) are found to be uncorrelated to $PC2_{m/z}$.

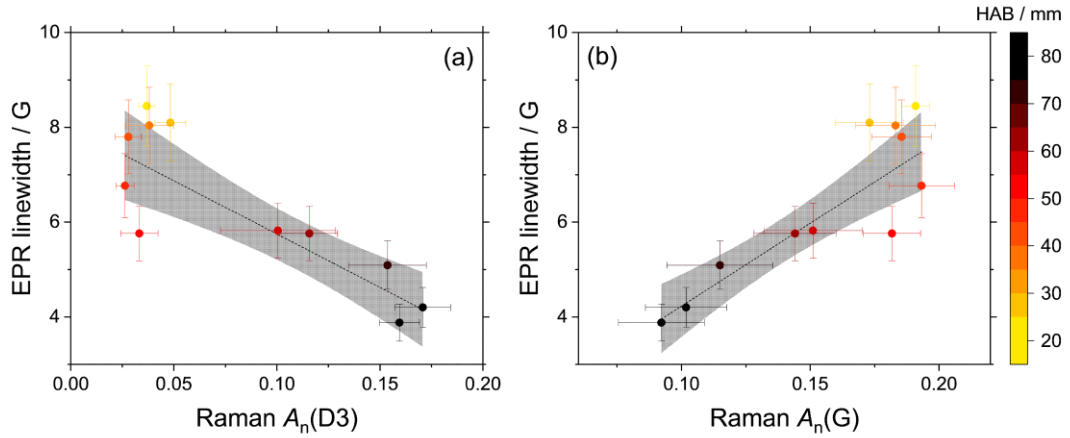
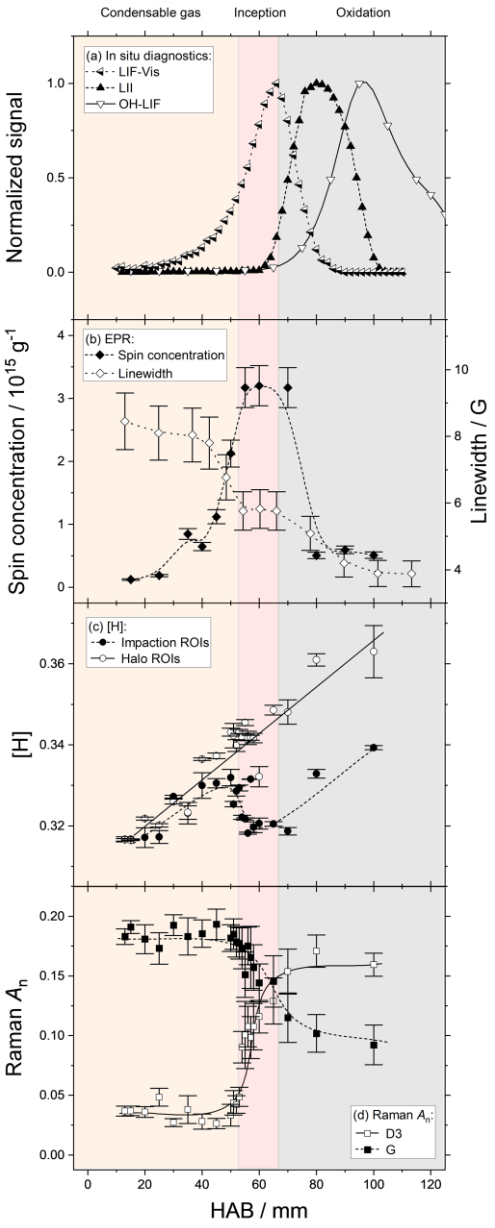


Figure 7. EPR linewidth and interpretation of Raman spectra. EPR linewidth against A_n(D3) (a) and against A_n(G) (b). The linear fitting (black dashed line) and the 95% confidence interval (gray area) are shown. Both correlations are statistically significant ($R^2 > 0.7$).

4. Discussion

4.1 Connecting the dots: toward a better understanding of soot inception

Figure 8 shows the available data that is found to be dependent on the flame HAB and that includes: (a) the normalized in situ profiles of the fluorophores (obtained by laser induced fluorescence, LIF) and the laser induced incandescence (LII) signal [50] that is directly proportional to the soot volume fraction; (b) the spin concentration of the persistent radicals (obtained by electron paramagnetic resonance, EPR) [48]; (c) the [H] of the impaction and halo ROIs (obtained by ToF-SIMS, this work); (d) the intensity of the vibrations of carbon crystals of finite size and flat graphene-like structures (obtained by Raman spectroscopy, this work). This multi-diagnostic approach allows the identification of distinct regimes in the flame axis based on differences in the chemical composition, structure and optical properties of the condensable gas and of the condensed phase. Information on incipient soot and its molecular precursors is then inferred from the comparative analysis.



406

407 **Figure 8. Dependence of the investigated quantities on the flame HAB.** The in-situ characterization of the flame
408 includes (a) the normalized profiles of the fluorescence induced with 532 nm excitation wavelength (half down left
409 triangles [35,50]), the LII signal (solid up triangles [50]), and the OH radicals fluorescence (hollow down triangles [51]).
410 The ex-situ characterization of the samples includes: (b) the spin concentration of persistent radicals (solid diamonds
411 [48]) (c) the [H] of the impaction (solid rounds) and halo (hollow rounds) ROIs, (d) A_n(D3) (hollow squares) and A_n(G)
412 (solid squares) of the impaction ROIs. All connecting lines are guidelines.

413

At 13-30 mm HAB the chemical composition of the condensable gas does not change significantly with the HAB or the ROI. In this HAB range, high *m*/*z* hydrocarbons are already available (Figure 2). The mass defect of the identified ions corresponds to or is slightly higher than pericondensed PAHs, but much lower than the aliphatic and benzene oligomers limits [35,87]. The vast majority of the molecular formulae assigned by mass defect analysis are consistent with PAHs and PAHs substituted by short aliphatic chains, and are in good agreement with works that identified in young soot aliphatic molecules [33,34] and low *m*/*z* aliphatic fragment ions [35].

Evidence of change in the flame properties and in the chemistry of the extracted samples begins around 30 mm HAB. First, the induced fluorescence emission progressively shifts toward visible wavelengths [50]. As previously discussed, evidence has been found that this emission, which was first observed in the 1980s [88,89] and initially attributed to large aromatic molecules [90], might be associated with long-lived PAH excimers [50,91] or radicals [48] instead. In parallel, the spin concentration of the persistent radicals increases (Figure 8b). Around 40 mm HAB, the EPR bandwidth (Figure 8b) begins to decrease and the [H] of the impaction and halo ROIs begins to diverge (Figure 8c).

In the impaction ROIs below 53 mm HAB, the intensity of the vibrations attributed to amorphous carbon and to the ideal graphitic lattice (peaks D3 and G, respectively, Figure 8d) is constant. This lack of evolution conveys important information: first, it supports the assignment of the ions detected by mass spectrometry to flat and isolated or weakly bound PAHs. Second, it clearly shows that the structure of the matter in the condensable gas does not change against the HAB. In this HAB range, no black body-like emission typical of the matter in the condensed phase is detected in situ in the flame (absence of LII signal, Figure 8a).

At 53-56 mm HAB, several drastic changes occur within a few mm in the ex situ samples that acquire the character of a discontinuity point. The chemical composition of the impaction ROIs (condensable gas and condensed phase) and of the halo ROIs (condensable gas only) diverge (Figure 8c). In particular, the decreasing [H] of the impaction ROIs, not observed at lower HAB or in the halo ROIs, is consistent with an emerging chemical reactivity compatible with the increasing formation rate of covalent C-C bonds that

supports the occurrence of cross-linking or dimerization as discussed in our previous work [46]. In the impaction ROIs, the relative contribution of high *m*/*z* hydrocarbons decreases in favor of lower *m*/*z* hydrocarbons and oxygen-containing hydrocarbons ($PC1_{m/z}$ loadings sign inversion, Figure 2). This depletion is consistent with the recently observed accumulation of PAHs on soot particles [92], and strongly suggests that incipient soot acts as sinks for high *m*/*z* hydrocarbons. The structure of the species in the impaction ROIs undergoes significant changes within the narrow 53-54 mm HAB range. First, the intensity of the vibrations of the graphitic lattice, that are constant up to 53 mm HAB, begin to decrease (Raman peak G, Figure 8c): a similar phenomenon has already been observed and suggested to be an indicator of the formation of curved structures as opposed to the flat aromatic species more abundant before soot inception [37,56]. At the same time, the intensity of the vibrations attributed to amorphous carbon suddenly rises (Raman peak D3, Figure 8c). The spin concentration and the fluorescence emission reach their maximum at 55-70 mm HAB (Figure 8b) and 65-70 mm HAB (LIF induced with 532 nm excitation wavelength, Figure 8a), respectively. The EPR bandwidth (Figure 8b) remains approximately constant, while black body-like emission of matter in the condensed phase is initially detected at 60 mm HAB (in situ LII signals with SNR $\geq$ 3, Figure 8a).

At 70-100 mm HAB the soot volume fraction reaches its maximum while the concentration of fluorophores detected in situ decreases against HAB (LII and LIF profiles, respectively, Figure 8a). High *m*/*z* hydrocarbons are depleted (Figure 2). The intensity of the vibrations of the graphitic lattice further decreases, while the intensity of the vibrations attributed to amorphous carbon reaches its maximum (Raman peaks G and D3, respectively, Figure 8b). The spin concentration reaches a minimum, and the EPR bandwidth further decreases (Figure 8b). All these changes are already well documented in the literature [1,3,4] and represent the transition from young to mature and (partially) oxidized soot. This second transition occurs in a broader HAB range than the former, and involves more clearly the structure rather than the chemical composition of the sampled materials (no sign inversion of $PC1_{m/z}$, Figure 2). The oxidation processes becoming dominant (rise of the concentration of OH radicals that begins around 65 mm HAB) reasonably explains the depletion of the high *m*/*z* hydrocarbons (high scores in $PC1_{m/z}$,

Figure 2), but also the disappearance of the mature soot particles that are almost entirely oxidized higher in the flame (the diffusion flame is non-smoking). The transition to a regime where the dominating mechanism is oxidation could also explain the loss of linearity in the correlation between [H] and $PC2_{m/z}$ previously discussed (see section 3.3), as the hydrocarbons found in this region likely have a more diverse nature than the PAHs generated only by molecular growth processes like the HACA.

4.2 A simple phenomenological model of soot inception

An overview of the experimental findings is provided in Figure 9. Based on purely phenomenological considerations, the hypothesis is made that in this diffusion flame the reactive species responsible for soot inception are formed from the pool of the PAHs and substituted PAHs already available at low HAB. As the flame temperature increases, the concentration of reactive species also increases until it can sustain the formation of adducts. In the flame axis, a critical point is reached at 53-54 mm HAB (1600-1650 K [51]), where the increased formation rate of C-C bonds results in a global decrease of [H], immediately followed by the detection of graphite-like crystals in the condensed phase ex situ, and slightly higher at 60 mm HAB (~1700 K) by in situ black body-like emission.

A possibility that should be considered is that the chain of events leading to soot inception begins before the black body-like emission of soot is detected. In this context, the lower limit of detection of the LII diagnostic becomes a key point for the interpretation of all experimental data. In this work, the first changes in the chemical composition and structure of the samples consistent with a phase transition occur 6-7 mm lower than the first black body-like emission attributed to particles in the condensed phase. This discrepancy cannot be accounted for even when including the uncertainty attributed to the sampling process (±2 mm). Based on the literature reviewed so far, there is consensus that incepting soot particles have different physical-chemical properties than fully formed primary soot particles, and in particular their absorption function E(m) is significantly lower [28,29]. Therefore, the black body-like emission is detectable only when the particle concentration in the flame becomes sufficiently high, or alternatively when coagulation and surface growth have already resulted in an increased E(m). These two scenarios are

not mutually exclusive and are both consistent with a shifted particle detection by LII on the axial profile of the flame.

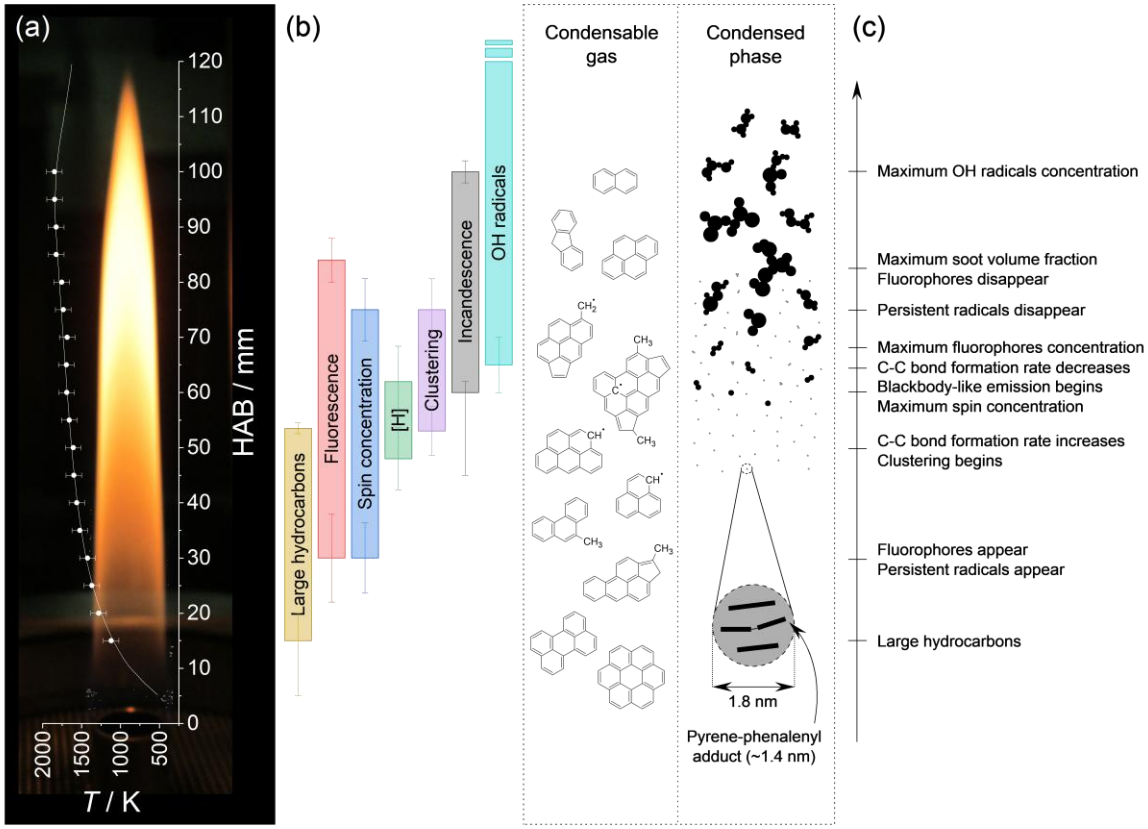


Figure 9. Phenomenological scenario for soot inception. (a) flame picture with overlaid HAB scale and temperature profile. (b) Significant HAB ranges: large hydrocarbons (ToF-SIMS on the impaction ROIs, yellow), fluorophores (LIF at 532 nm excitation wavelength, red), spin concentration (EPR, blue), slope inversion of [H] (green), clustering (changes of the D3 and G Raman peaks, purple), black body-like emission (LII, gray), and OH radicals (LIF-OH, cyan). The error bars of the quantities obtained from ex-situ diagnostics (EPR, ToF-SIMS and Raman) include the shift introduced with the sampling uncertainty (estimated to ±2 mm from the change in the flame luminosity and propagated with a Taylor expansion). (c) summary of the experimental observation. The depicted soot particles are to scale (2 nm diameter incipient soot particles, 5-15 nm diameter primary particles, ~50 nm gyration diameter aggregates). The size of the depicted adducts is calculated from aromatic C-C and C-H bond distances.

With this uncertainty in mind, it is interesting to notice that the spin concentration of the persistent radicals reaches its maximum immediately before the black body-like emission attributed to particles in the condensed phase is first detected, and begins to decrease when the OH radicals appear and the soot volume fraction reaches its maximum (Figure 8). These correspondences strongly suggest that the detected persistent radicals are closely related to the reactive species responsible for soot inception: when the reactive medium is extracted from the flame and deposited on a solid substrate during sampling, the surviving reactive species likely transform into the stable products detected by the ex situ diagnostics, including the persistent radicals detected by EPR. However, a trace is clearly left in the samples as the spin concentration peaks shortly before the black body-like emission attributed to soot particles is first detected. In addition, the spin concentration is linearly correlated to the induced fluorescence at 532 nm [48], i.e. the reactive species that generate the persistent radicals and the fluorophores may share a common origin. The induced fluorescence [35,50,91] and recent works on the characterization of the persistent radicals [47,48,93] suggest important contributions from resonance-stabilized radicals of PAHs substituted by short aliphatic chains. The availability of radical PAHs substituted with side chains might contribute to further stabilize any multi-layered structure formed via interplanar covalent bonds in addition to the weaker staking interactions that are known to be insufficient to hold together clusters of small PAHs at the flame temperature [94]. This hypothesis is additionally supported by the correlation between the EPR linewidth and the intensity of the vibration of mode E_{2g} of the ideal graphitic lattice found in this work. The decreasing intensity of the vibration of mode E_{2g} (Figure 7b) and the disappearance of large PAHs (Figure 2) suggest that the initial formation of amorphous carbon is associated with the reduction of the size of the aromatic domains. This decreasing intensity of the vibration cannot be a consequence of the increasing number of lattice defects as the intensity of the vibration of mode A_{1g} of graphene is not correlated to the soot inception process (peak D1, see Figure 6). Therefore, it must be concluded that the observed reduction of the size of the aromatic domains is caused by the formation of curved and/or multi-layer structures.

Higher in the flame, around 70-75 mm HAB, this equilibrium is upset. The slope of [H] changes its sign again, indicating that the formation rate of C-C bonds decreases, the emission of the fluorophores and the spin concentration of the persistent radicals are reduced to roughly one tenth of their peak value. All this information indicates that soot inception reaches an abrupt halt shortly before the maximum soot volume fraction is detected. As already noticed above, this second series of changes corresponds remarkably well to the rise of the concentration of OH radicals. Therefore, it can be inferred that the OH radicals first inactivate the most reactive species in the flame, which subsequently slows down the formation rate of C-C bonds, and finally stabilizes the structure of the particles to that of mature soot. In particular, soot sampled above the maximum soot volume fraction show no visual trace of the brown organic matter that is prominent in the samples collected at lower HAB [48].

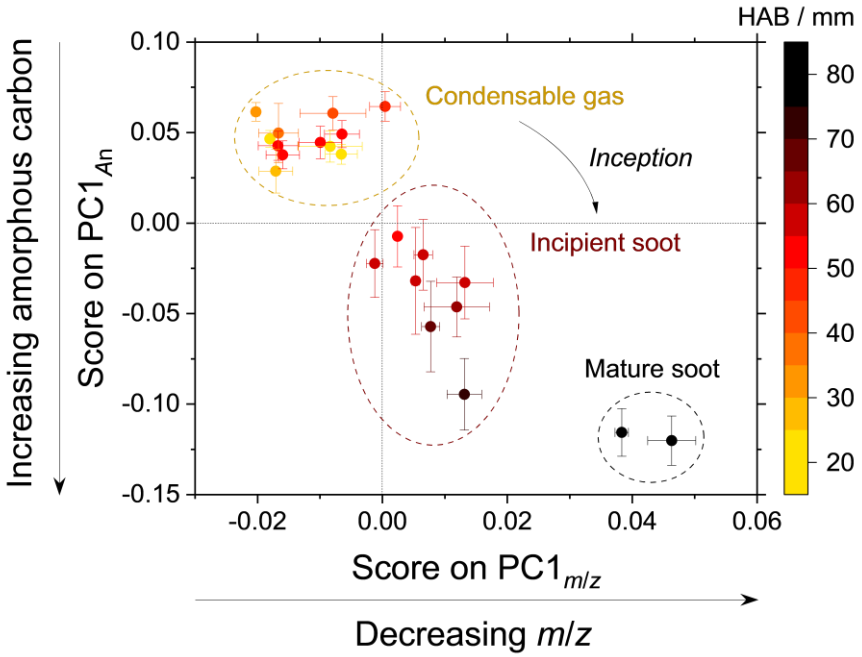
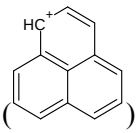
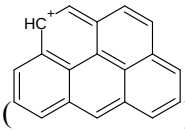


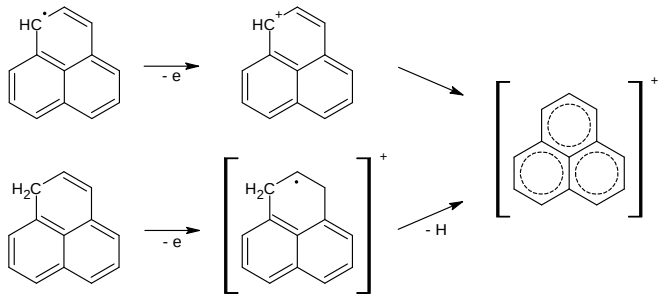
Figure 10. Soot inception identified by the PCA (Figure 5 revisited). $PC1_{An}$ against $PC1_{m/z}$ plot emphasizing the existence of three data clusters. Increasing scores on $PC1_{m/z}$ represent the increasing relative contribution of low m/z PAHs and their derivatives (section 3.1), while decreasing scores on $PC1_{An}$ represent the increasing relative contribution of amorphous carbon (section 3.4). Notice how the data cluster assigned to the incipient soot stands out from the other two data clusters.

In light of this discussion, the cross-comparison of the analysis of the chemical composition and structure of soot becomes an effective way to represent soot inception. Figure 10 shows the $PC1_{An}$ against $PC1_{m/z}$ plot, already discussed in section 3.5, with three highlighted data clusters representing different regimes separated by discontinuities and assigned to condensable gas, incipient soot and mature soot. This representation well summarizes the main information extracted from this study, and in particular highlights the existence of incipient soot as a distinct regime with chemical composition and structure significantly different than either the condensable gas and the mature soot.

In our previous work, the potential link between the formation of dimers of small PAHs and soot inception has been discussed at length [46]. As a final remark, we would like to bring attention to the cations that mark the beginning and the maximum of the negative $PC1_{m/z}$ loadings distribution in the analysis of mass spectrometry data, and specifically to the first cation of the distribution $C_{13}H_9^+$ (165.070 u) and to the cation with the largest variability in the sequence of high *m*/*z* hydrocarbons $C_{19}H_{11}^+$ (239.086 u). Among the possible isomers, the cations of 1H-phenalene (

) and 6H-benzo[*cd*]pyrene (

) are particularly interesting as they have attracted considerable interest in theoretical chemistry for their ability to form relatively stable π dimers [95,96]. On the one hand, care must be taken when interpreting ex situ mass spectrometry data. Benzenoid PAHs with an odd number of carbon atoms contain a CH_2 group that can undergo post-ionization hydrogen elimination to form the same cations as resonance-stabilized radicals as shown below for the case of phenalene/phenalenyl:



Therefore, it can be argued that the secondary ions detected ex situ mostly derive from stable PAHs rather than reactive radicals since the latter are less likely to survive the sampling process. On the other hand, the very existence of a maximum of the spin concentration before the inception zone is clear evidence that at least part of the information on the reactive radicals is preserved for a long time after the sampling. Within the limits imposed by this interpretation, the intermediate species detected at the very beginning of the soot inception zone in our previous work [46] could be the stable products deriving from the reactive radicals in the flame. For instance, reactions reminiscent of the formation of Frenklach's E-bridges [2] or Selvakumar's π-radicals localization [97] like:

pyrene ($C_{16}H_{10}$) + phenalene* ($C_{13}H_9$) → pyrene-phenalene* ($C_{29}H_{15}$) + 4H

would result in adducts capable of forming cations consistent with the observed *m*/*z*. The characteristic size of the adducts, calculated from the aromatic bond length of PAHs in the considered mass range and shown in scale in Figure 9, is 0.8-1.5 nm, in reasonably good agreement with the estimated size of the aromatic domains in this flame [46]. Additionally, similar reactions would explain the observed decrease of [H], and being resonance-stabilized, the adducts could be sufficiently stable to survive the sampling and be detected ex situ.

5. Conclusions

In this work, the chemical composition and the structure of incipient soot generated in a co-flow laminar diffusion flame of methane are investigated with a multi-diagnostic approach. Principal component analysis is used to determine the physical variables involved in soot inception, and to link them to soot inception, growth and oxidation.

The main finding of this work relies on the comparative analysis of the chemical composition (by secondary ion mass spectrometry) and the structure (by Raman spectroscopy) of incipient soot of samples extracted from the flame. The results are discussed in light of the information obtained from complementary diagnostics that include the spin concentration of the extracted samples (by electron paramagnetic resonance), and the in-flame distribution of the soot volume fraction (by laser induced incandescence) and of different classes of fluorophores (by laser induced fluorescence). The comparative analysis shows compelling evidence of the existence of an intermediate state of the matter extracted from the flame, attributed to incipient soot, having significantly different chemical composition and structure than the condensable gas and the mature soot. The appearance of this intermediate state occurs suddenly shortly before the in situ detection of soot particles by laser induced incandescence in the axial direction of the flame.

In terms of chemical composition, the atomic percentage of hydrogen [H] of the intermediate state shows a slope inversion that suggests the existence of an emerging reactivity absent in the condensable gas. Based on our past investigation [46], this reactivity is attributed to an increased formation rate of C-C bonds associated with the formation of incipient soot. The candidate reactive species are identified as C12–C20 CmHn+ hydrocarbons. In addition, the change of slope of [H] corresponds remarkably well to the maximum of the spin concentration detected by electron paramagnetic resonance, which suggests that some of the reactive species are radicals.

In terms of structure, the physical phenomena linked to the appearance of the intermediate state are the appearance of amorphous carbon and the progressive disappearance of large and flat aromatic structures, while the overall number of lattice defects is unaffected.

6. Acknowledgments

The authors would like to thank Dr. Edouard Capoen for his invaluable support in the preparation of the deposition substrates. The authors would also like to acknowledge the contribution of the *Centre d' Etudes et de Recherches Lasers et Applications* (CERLA) platform for the materials and equipment.

Financial support for this work was provided by the French *Agence de la Transition Ecologique* (ADEME).

7. Bibliography

[1] Martin JW, Salamanca M, Kraft M. Soot inception: carbonaceous nanoparticle formation in flames. Prog Energy Combust Sci 2022;88:100956. https://doi.org/10.1016/j.pecs.2021.100956.
[2] Frenklach M. Reaction mechanism of soot formation in flames. Phys Chem Chem Phys 2002;4:2028–37. https://doi.org/10.1039/B110045A.
[3] D'Anna A. Combustion-formed nanoparticles. Proc Combust Inst 2009;32:593–613. https://doi.org/10.1016/j.proci.2008.09.005.
[4] Wang H. Formation of nascent soot and other condensed-phase materials in flames. Proc Combust Inst 2011;33:41–67. https://doi.org/10.1016/j.proci.2010.09.009.
[5] Desgroux P, Mercier X, Thomson KA. Study of the formation of soot and its precursors in flames using optical diagnostics. Proc Combust Inst 2013;34:1713–38. https://doi.org/10.1016/j.proci.2012.09.004.
[6] Frenklach M, Mebel AM. On the mechanism of soot nucleation. Phys Chem Chem Phys 2020;22:5314–31. https://doi.org/10.1039/D0CP00116C.
[7] Michelsen HA, Colket MB, Bengtsson P-E, D'Anna A, Desgroux P, Haynes BS, et al. A review of terminology used to describe soot formation and evolution under combustion and pyrolytic conditions. ACS Nano 2020;14:12470–90. https://doi.org/10.1021/acsnano.0c06226.
[8] Alfè M, Apicella B, Barbella R, Rouzaud J-N, Tregrossi A, Ciajolo A. Structure–property relationship in nanostructures of young and mature soot in premixed flames. Proc Combust Inst 2009;32:697–704. https://doi.org/10.1016/j.proci.2008.06.193.
[9] Kholghy MR, Veshkini A, Thomson MJ. The core–shell internal nanostructure of soot – A criterion to model soot maturity. Carbon 2016;100:508–36. https://doi.org/10.1016/j.carbon.2016.01.022.
[10] Sgro LA, Barone AC, Commodo M, D'Alessio A, De Filippo A, Lanzuolo G, et al. Measurement of nanoparticles of organic carbon in non-sooting flame conditions. Proc Combust Inst 2009;32:689–96. https://doi.org/10.1016/j.proci.2008.06.216.
[11] Commodo M, De Falco G, Bruno A, Borriello C, Minutolo P, D'Anna A. Physicochemical evolution of nascent soot particles in a laminar premixed flame: from nucleation to early growth. Combust Flame 2015;162:3854–63. https://doi.org/10.1016/j.combustflame.2015.07.022.
[12] Moallemi A, Kazemimanesh M, Kostiuk LW, Olfert JS. The effect of sodium chloride on the nanoparticles observed in a laminar methane diffusion flame. Combust Flame 2018;188:273–83. https://doi.org/10.1016/j.combustflame.2017.10.009.
[13] Totton TS, Chakrabarti D, Misquitta AJ, Sander M, Wales DJ, Kraft M. Modelling the internal structure of nascent soot particles. Combust Flame 2010;157:909–14. https://doi.org/10.1016/j.combustflame.2009.11.013.
[14] Teini PD, Karwat DMA, Atreya A. Observations of nascent soot: molecular deposition and particle morphology. Combust Flame 2011;158:2045–55. https://doi.org/10.1016/j.combustflame.2011.03.005.
[15] Botero ML, Sheng Y, Akroyd J, Martin J, Dreyer JAH, Yang W, et al. Internal structure of soot particles in a diffusion flame. Carbon 2019;141:635–42. https://doi.org/10.1016/j.carbon.2018.09.063.
[16] Ishiguro T, Takatori Y, Akihama K. Microstructure of diesel soot particles probed by electron microscopy: first observation of inner core and outer shell. Combust Flame 1997;108:231–4. https://doi.org/10.1016/S0010-2180(96)00206-4.

[17] D'Alessio A, Barone AC, Cau R, D'Anna A, Minutolo P. Surface deposition and coagulation efficiency of combustion generated nanoparticles in the size range from 1 to 10nm. Proc Combust Inst 2005;30:2595–603. https://doi.org/10.1016/j.proci.2004.08.267.
[18] Hou D, Zong D, Lindberg CS, Kraft M, You X. On the coagulation efficiency of carbonaceous nanoparticles. J Aerosol Sci 2020;140:105478. https://doi.org/10.1016/j.jaerosci.2019.105478.
[19] Huo Z, Cleary MJ, Sirignano M, Masri AR. A sectional soot formation kinetics scheme with a new model for coagulation efficiency. Combust Flame 2021;230:111444. https://doi.org/10.1016/j.combustflame.2021.111444.
[20] De Falco G, Picca F, Commodo M, Minutolo P. Probing soot structure and electronic properties by optical spectroscopy. Fuel 2020;259:116244. https://doi.org/10.1016/j.fuel.2019.116244.
[21] Zhao B, Yang Z, Johnston MV, Wang H, Wexler AS, Balthasar M, et al. Measurement and numerical simulation of soot particle size distribution functions in a laminar premixed ethylene-oxygen-argon flame. Combust Flame 2003;133:173–88. https://doi.org/10.1016/S0010-2180(02)00574-6.
[22] Abid AD, Tolmachoff ED, Phares DJ, Wang H, Liu Y, Laskin A. Size distribution and morphology of nascent soot in premixed ethylene flames with and without benzene doping. Proc Combust Inst 2009;32:681–8. https://doi.org/10.1016/j.proci.2008.07.023.
[23] Carbone F, Moslih S, Gomez A. Probing gas-to-particle transition in a moderately sooting atmospheric pressure ethylene/air laminar premixed flame. Part II: Molecular clusters and nascent soot particle size distributions. Combust Flame 2017;181:329–41. https://doi.org/10.1016/j.combustflame.2017.02.021.
[24] Shariatmadar H, Aleiferis PG, Lindstedt RP. Particle size distributions in turbulent premixed ethylene flames crossing the soot inception limit. Combust Flame 2022;243:111978. https://doi.org/10.1016/j.combustflame.2021.111978.
[25] Michelsen HA. Effects of maturity and temperature on soot density and specific heat. Proc Combust Inst 2021;38:1197–205. https://doi.org/10.1016/j.proci.2020.06.383.
[26] Betrancourt C, Mercier X, Liu F, Desgroux P. Quantitative measurement of volume fraction profiles of soot of different maturities in premixed flames by extinction-calibrated laser-induced incandescence. Appl Phys B 2019;125:16. https://doi.org/10.1007/s00340-018-7127-2.
[27] Kelesidis GA, Pratsinis SE. Soot light absorption and refractive index during agglomeration and surface growth. Proc Combust Inst 2019;37:1177–84. https://doi.org/10.1016/j.proci.2018.08.025.
[28] Yon J, Cruz JJ, Escudero F, Morán J, Liu F, Fuentes A. Revealing soot maturity based on multi-wavelength absorption/emission measurements in laminar axisymmetric coflow ethylene diffusion flames. Combust Flame 2021;227:147–61. https://doi.org/10.1016/j.combustflame.2020.12.049.
[29] Minutolo P, Commodo M, D'Anna A. Optical properties of incipient soot. Proc Combust Inst 2023;39:1129–38. https://doi.org/10.1016/j.proci.2022.09.019.
[30] Mouton T, Mercier X, Wartel M, Lamoureux N, Desgroux P. Laser-induced incandescence technique to identify soot nucleation and very small particles in low-pressure methane flames. Appl Phys B 2013;112:369–79. https://doi.org/10.1007/s00340-013-5446-x.
[31] Betrancourt C, Liu F, Desgroux P, Mercier X, Faccinetto A, Salamanca M, et al. Investigation of the size of the incandescent incipient soot particles in premixed sooting and nucleation flames of n-butane using LII, HIM, and 1 nm-SMPS. Aerosol Sci Technol 2017;51:916–35. https://doi.org/10.1080/02786826.2017.1325440.
[32] Carbone F, Canagaratna MR, Lambe AT, Jayne JT, Worsnop DR, Gomez A. Detection of weakly bound clusters in incipiently sooting flames via ion seeded dilution and collision charging for (APi-TOF) mass spectrometry analysis. Fuel 2021;289:119820. https://doi.org/10.1016/j.fuel.2020.119820.
[33] Cain JP, Camacho J, Phares DJ, Wang H, Laskin A. Evidence of aliphatics in nascent soot particles in premixed ethylene flames. Proc Combust Inst 2011;33:533–40. https://doi.org/10.1016/j.proci.2010.06.164.

[34] Cain J, Laskin A, Kholghy MR, Thomson MJ, Wang H. Molecular characterization of organic content of soot along the centerline of a coflow diffusion flame. Phys Chem Chem Phys 2014;16:25862–75. https://doi.org/10.1039/C4CP03330B.
[35] Irimiea C, Faccinetto A, Mercier X, Ortega I-K, Nuns N, Therssen E, et al. Unveiling trends in soot nucleation and growth: When secondary ion mass spectrometry meets statistical analysis. Carbon 2019;144:815–30. https://doi.org/10.1016/j.carbon.2018.12.015.
[36] Schulz F, Commodo M, Kaiser K, De Falco G, Minutolo P, Meyer G, et al. Insights into incipient soot formation by atomic force microscopy. Proc Combust Inst 2018;37:885–92. https://doi.org/10.1016/j.proci.2018.06.100.
[37] Commodo M, Kaiser K, De Falco G, Minutolo P, Schulz F, D'Anna A, et al. On the early stages of soot formation. Molecular structure elucidation by high-resolution atomic force microscopy. Combust Flame 2019;205:154–64. https://doi.org/10.1016/j.combustflame.2019.03.042.
[38] Vander Wal RL, Tomasek AJ. Soot nanostructure: dependence upon synthesis conditions. Combust Flame 2004;136:129–40. https://doi.org/10.1016/j.combustflame.2003.09.008.
[39] Vander Wal RL, Yezerets A, Currier NW, Kim DH, Wang CM. HRTEM Study of diesel soot collected from diesel particulate filters. Carbon 2007;45:70–7. https://doi.org/10.1016/j.carbon.2006.08.005.
[40] Eaves NA, Dworkin SB, Thomson MJ. The importance of reversibility in modeling soot nucleation and condensation processes. Proc Combust Inst 2015;35:1787–94. https://doi.org/10.1016/j.proci.2014.05.036.
[41] Camacho J, Tao Y, Wang H. Kinetics of nascent soot oxidation by molecular oxygen in a flow reactor. Proc Combust Inst 2015;35:1887–94. https://doi.org/10.1016/j.proci.2014.05.095.
[42] De Falco G, Bocchicchio S, Commodo M, Minutolo P, D'Anna A. Raman spectroscopy of nascent soot oxidation: structural analysis during heating. Front Energy Res 2022;10:878171.
[43] Russo C, Ciajolo A, Cimino S, La Matta V, La Rocca A, Apicella B. Reactivity of soot emitted from different hydrocarbon fuels. Effect of nanostructure on oxidation kinetics. Fuel Process Technol 2022;236:107401. https://doi.org/10.1016/j.fuproc.2022.107401.
[44] Grimonprez S, Faccinetto A, Batut S, Wu J, Desgroux P, Petitprez D. Cloud condensation nuclei from the activation with ozone of soot particles sampled from a kerosene diffusion flame. Aerosol Sci Technol 2018;52:814–27. https://doi.org/10.1080/02786826.2018.1472367.
[45] Wu J, Faccinetto A, Grimonprez S, Batut S, Yon J, Desgroux P, et al. Influence of the dry aerosol particle size distribution and morphology on the cloud condensation nuclei activation. An experimental and theoretical investigation. Atmos Chem Phys 2020;20:4209–25. https://doi.org/10.5194/acp-20-4209-2020.
[46] Faccinetto A, Irimiea C, Minutolo P, Commodo M, D'Anna A, Nuns N, et al. Evidence on the formation of dimers of polycyclic aromatic hydrocarbons in a laminar diffusion flame. Commun Chem 2020;3:112. https://doi.org/10.1038/s42004-020-00357-2.
[47] Vitiello G, De Falco G, Picca F, Commodo M, D'Errico G, Minutolo P, et al. Role of radicals in carbon clustering and soot inception: a combined EPR and Raman spectroscopic study. Combust Flame 2019;205:286–94. https://doi.org/10.1016/j.combustflame.2019.04.028.
[48] Elias J, Faccinetto A, Vezin H, Mercier X. Investigation of resonance-stabilized radicals associated with soot particle inception using advanced electron paramagnetic resonance techniques. Commun Chem 2023;6:99. https://doi.org/10.1038/s42004-023-00896-4.
[49] Irimiea C, Faccinetto A, Carpentier Y, Ortega I-K, Nuns N, Therssen E, et al. A comprehensive protocol for chemical analysis of flame combustion emissions by secondary ion mass spectrometry. Rapid Commun Mass Spectrom 2018;32:1015–25. https://doi.org/10.1002/rcm.8133.
[50] Mercier X, Carrivain O, Irimiea C, Faccinetto A, Therssen E. Dimers of polycyclic aromatic hydrocarbons: the missing pieces in the soot formation process. Phys Chem Chem Phys 2019;21:8282–94. https://doi.org/10.1039/C9CP00394K.

[51] Elias J, Faccinetto A, Batut S, Carrivain O, Sirignano M, D'Anna A, et al. Thermocouple-based thermometry for laminar sooting flames: Implementation of a fast and simple methodology. Int J Therm Sci 2023;184:107973. https://doi.org/10.1016/j.ijthermalsci.2022.107973.
[52] Dobbins RA, Fletcher RA, Chang H-C. The evolution of soot precursor particles in a diffusion flame. Combust Flame 1998;115:285–98. https://doi.org/10.1016/S0010-2180(98)00010-8.
[53] Casiraghi C, Piazza F, Ferrari AC, Grambole D, Robertson J. Bonding in hydrogenated diamond-like carbon by Raman spectroscopy. Diam Relat Mater 2005;14:1098–102. https://doi.org/10.1016/j.diamond.2004.10.030.
[54] Buijnsters JG, Gago R, Jiménez I, Camero M, Agulló-Rueda F, Gómez-Aleixandre C. Hydrogen quantification in hydrogenated amorphous carbon films by infrared, Raman, and x-ray absorption near edge spectroscopies. J Appl Phys 2009;105:093510. https://doi.org/10.1063/1.3103326.
[55] Parent P, Laffon C, Marhaba I, Ferry D, Regier TZ, Ortega I-K, et al. Nanoscale characterization of aircraft soot: a high-resolution transmission electron microscopy, Raman spectroscopy, X-ray photoelectron and near-edge X-ray absorption spectroscopy study. Carbon 2016;101:86–100. https://doi.org/10.1016/j.carbon.2016.01.040.
[56] Commodo M, D'Anna A, De Falco G, Larciprete R, Minutolo P. Illuminating the earliest stages of the soot formation by photoemission and Raman spectroscopy. Combust Flame 2017;181:188–97. https://doi.org/10.1016/j.combustflame.2017.03.020.
[57] Ferrari AC, Robertson J. Interpretation of Raman spectra of disordered and amorphous carbon. Phys Rev B 2000;61:14095–107. https://doi.org/10.1103/PhysRevB.61.14095.
[58] Le KC, Lefumeux C, Bengtsson P-E, Pino T. Direct observation of aliphatic structures in soot particles produced in low-pressure premixed ethylene flames via online Raman spectroscopy. Proc Combust Inst 2019;37:869–76. https://doi.org/10.1016/j.proci.2018.08.003.
[59] Sadezky A, Muckenhuber H, Grothe H, Niessner R, Pöschl U. Raman microspectroscopy of soot and related carbonaceous materials: spectral analysis and structural information. Carbon 2005;43:1731–42. https://doi.org/10.1016/j.carbon.2005.02.018.
[60] Russo C, Apicella B, Lighty JS, Ciajolo A, Tregrossi A. Optical properties of organic carbon and soot produced in an inverse diffusion flame. Carbon 2017;124:372–9. https://doi.org/10.1016/j.carbon.2017.08.073.
[61] Haller T, Rentenberger C, Meyer JC, Felgitsch L, Grothe H, Hitzenberger R. Structural changes of CAST soot during a thermal–optical measurement protocol. Atmos Meas Tech 2019;12:3503–19. https://doi.org/10.5194/amt-12-3503-2019.
[62] Minutolo P, Commodo M, Santamaria A, De Falco G, D'Anna A. Characterization of flame-generated 2-D carbon nano-disks. Carbon 2014;68:138–48. https://doi.org/10.1016/j.carbon.2013.10.073.
[63] Catelani T, Pratesi G, Zoppi M. Raman characterization of ambient airborne soot and associated mineral phases. Aerosol Sci Technol 2014;48:13–21. https://doi.org/10.1080/02786826.2013.847270.
[64] Carpentier Y, Féraud G, Dartois E, Brunetto R, Charon E, Cao A-T, et al. Nanostructuration of carbonaceous dust as seen through the positions of the 6.2 and 7.7 μm AIBs. Astron Astrophys 2012;548:A40. https://doi.org/10.1051/0004-6361/201118700.
[65] Dasappa S, Camacho J. Evolution in size and structural order for incipient soot formed at flame temperatures greater than 2100 K. Fuel 2021;291:120196. https://doi.org/10.1016/j.fuel.2021.120196.
[66] Dippel B, Jander H, Heintzenberg J. NIR FT Raman spectroscopic study of flame soot. Phys Chem Chem Phys 1999;1:4707–12. https://doi.org/10.1039/A904529E.
[67] Cuesta A, Dhamelincourt P, Laureyns J, Martínez-Alonso A, Tascón JMD. Raman microprobe studies on carbon materials. Carbon 1994;32:1523–32. https://doi.org/10.1016/0008-6223(94)90148-1.

[68] Livneh T, Bar-Ziv E, Senneca O, Salatino P. Evolution of reactivity of highly porous chars from Raman microscopy. Combust Sci Technol 2000;153:65–82. https://doi.org/10.1080/00102200008947251.
[69] Tan PH, Hu CY, Dong J, Shen WC, Zhang BF. Polarization properties, high-order Raman spectra, and frequency asymmetry between Stokes and anti-Stokes scattering of Raman modes in a graphite whisker. Phys Rev B 2001;64:214301. https://doi.org/10.1103/PhysRevB.64.214301.
[70] Mafra DL, Samsonidze G, Malard LM, Elias DC, Brant JC, Plentz F, et al. Determination of LA and TO phonon dispersion relations of graphene near the Dirac point by double resonance Raman scattering. Phys Rev B 2007;76:233407. https://doi.org/10.1103/PhysRevB.76.233407.
[71] Tuinstra F, Koenig JL. Raman spectrum of graphite. J Chem Phys 1970;53:1126–30. https://doi.org/10.1063/1.1674108.
[72] Katagiri G, Ishida H, Ishitani A. Raman spectra of graphite edge planes. Carbon 1988;26:565–71. https://doi.org/10.1016/0008-6223(88)90157-1.
[73] Wang Y, Alsmeyer DC, McCreery RL. Raman spectroscopy of carbon materials: structural basis of observed spectra. Chem Mater 1990;2:557–63. https://doi.org/10.1021/cm00011a018.
[74] Di Donato E, Tommasini M, Fustella G, Brambilla L, Castiglioni C, Zerbi G, et al. Wavelength-dependent Raman activity of D2h symmetry polycyclic aromatic hydrocarbons in the D-band and acoustic phonon regions. Chem Phys 2004;301:81–93. https://doi.org/10.1016/j.chemphys.2004.02.018.
[75] Jawhari T, Roid A, Casado J. Raman spectroscopic characterization of some commercially available carbon black materials. Carbon 1995;33:1561–5. https://doi.org/10.1016/0008-6223(95)00117-V.
[76] Nemanich RJ, Solin SA. First- and second-order Raman scattering from finite-size crystals of graphite. Phys Rev B 1979;20:392–401. https://doi.org/10.1103/PhysRevB.20.392.
[77] Beyssac O, Goffé B, Petitet J-P, Froigneux E, Moreau M, Rouzaud J-N. On the characterization of disordered and heterogeneous carbonaceous materials by Raman spectroscopy. Spectrochim Acta A 2003;59:2267–76. https://doi.org/10.1016/S1386-1425(03)00070-2.
[78] Abdi H, Williams LJ. Principal component analysis. Wiley Interdiscip Rev Comput Stat 2010;2:433–59. https://doi.org/10.1002/wics.101.
[79] Bro R, Smilde AK. Principal component analysis. Anal Methods 2014;6:2812–31. https://doi.org/10.1039/C3AY41907J.
[80] Pei L, Jiang G, Tyler BJ, Baxter LL, Linford MR. Time-of-flight secondary ion mass spectrometry of a range of coal samples: a chemometrics (PCA, cluster, and PLS) analysis. Energ Fuels 2008;22:1059–72. https://doi.org/10.1021/ef7003199.
[81] Duca D, Irimiea C, Faccinetto A, Noble JA, Vojkovic M, Carpentier Y, et al. On the benefits of using multivariate analysis in mass spectrometric studies of combustion-generated aerosols. Faraday Discuss 2019;218:115–37. https://doi.org/10.1039/C8FD00238J.
[82] Negri F, Castiglioni C, Tommasini M, Zerbi G. A computational study of the Raman spectra of large polycyclic aromatic hydrocarbons: toward molecularly defined subunits of graphite. J Phys Chem A 2002;106:3306–17. https://doi.org/10.1021/jp0128473.
[83] Pimenta MA, Dresselhaus G, Dresselhaus MS, Cançado LG, Jorio A, Saito R. Studying disorder in graphite-based systems by Raman spectroscopy. Phys Chem Chem Phys 2007;9:1276–90. https://doi.org/10.1039/B613962K.
[84] Ess MN, Ferry D, Kireeva ED, Niessner R, Ouf F-X, Ivleva NP. In situ Raman microspectroscopic analysis of soot samples with different organic carbon content: structural changes during heating. Carbon 2016;105:572–85. https://doi.org/10.1016/j.carbon.2016.04.056.
[85] Baldelli A, Rogak SN. Morphology and Raman spectra of aerodynamically classified soot samples. Atmos Meas Tech 2019;12:4339–46. https://doi.org/10.5194/amt-12-4339-2019.
[86] Skrzypczak-Bonduelle A, Binet L, Delpoux O, Vezin H, Derenne S, Robert F, et al. EPR of radicals in primitive organic matter: a tool for the search of biosignatures of the most ancient traces of life. Appl Magn Reson 2008;33:371–97. https://doi.org/10.1007/s00723-008-0083-y.

[87] D'Anna A, Sirignano M, Kent J. A model of particle nucleation in premixed ethylene flames. Combust Flame 2010;157:2106–15. https://doi.org/10.1016/j.combustflame.2010.04.019.
[88] Beretta F, Cincotti V, D'Alessio A, Menna P. Ultraviolet and visible fluorescence in the fuel pyrolysis regions of gaseous diffusion flames. Combust Flame 1985;61:211–8. https://doi.org/10.1016/0010-2180(85)90102-6.
[89] Coe DS, Haynes BS, Steinfeld JI. Identification of a source of argon-ion-laser excited fluorescence in sooting flames. Combust Flame 1981;43:211–4. https://doi.org/10.1016/0010-2180(81)90018-3.
[90] Smyth KC, Shaddix CR, Everest DA. Aspects of soot dynamics as revealed by measurements of broadband fluorescence and flame luminosity in flickering diffusion flames. Combust Flame 1997;111:185–207. https://doi.org/10.1016/S0010-2180(97)00017-5.
[91] Sirignano M, Collina A, Commodo M, Minutolo P, D'Anna A. Detection of aromatic hydrocarbons and incipient particles in an opposed-flow flame of ethylene by spectral and time-resolved laser induced emission spectroscopy. Combust Flame 2012;159:1663–9. https://doi.org/10.1016/j.combustflame.2011.11.005.
[92] Shariatmadar H, Hampp F, Lindstedt RP. Quantification of PAH concentrations in premixed turbulent flames crossing the soot inception limit. Proc Combust Inst 2021;38:1163–72. https://doi.org/10.1016/j.proci.2020.06.359.
[93] Martin JW, Pascazio L, Menon A, Akroyd J, Kaiser K, Schulz F, et al. π-diradical aromatic soot precursors in flames. J Am Chem Soc 2021;143:12212–9. https://doi.org/10.1021/jacs.1c05030.
[94] Sabbah H, Biennier L, Klippenstein SJ, Sims IR, Rowe BR. Exploring the role of PAHs in the formation of soot: pyrene dimerization. J Phys Chem Lett 2010;1:2962–7. https://doi.org/10.1021/jz101033t.
[95] Small D, Zaitsev V, Jung Y, Rosokha SV, Head-Gordon M, Kochi JK. Intermolecular π-to-π bonding between stacked aromatic dyads. experimental and theoretical binding energies and near-IR optical transitions for phenalenyl radical/radical versus radical/cation dimerizations. J Am Chem Soc 2004;126:13850–8. https://doi.org/10.1021/ja046770i.
[96] Xiang Q, Guo J, Xu J, Ding S, Li Z, Li G, et al. Stable olympicenyl radicals and their π-dimers. J Am Chem Soc 2020;142:11022–31. https://doi.org/10.1021/jacs.0c02287.
[97] Selvakumar PK, Martin JW, Lorenzo MD, Paskevicius M, Buckley CE. Role of π-radical localization on thermally stable cross-links between polycyclic aromatic hydrocarbons. J Phys Chem A 2023;127:6945–52. https://doi.org/10.1021/acs.jpca.3c03769.
[98] Faccinetto A, Focsa C, Desgroux P, Ziskind M. Progress toward the quantitative analysis of PAHs adsorbed on soot by laser desorption/laser ionization/time-of-flight mass spectrometry. Environ Sci Technol 2015;49:10510–20. https://doi.org/10.1021/acs.est.5b02703.

8. Supporting Information

8.1 Flame sampling scheme

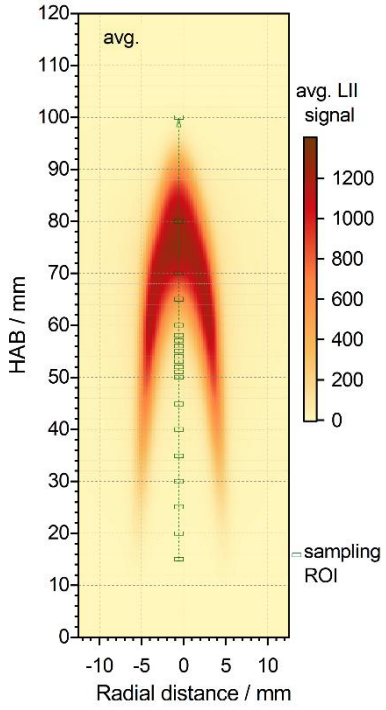


Figure SI 1. LII mapping of soot in the diffusion flame and sampling scheme (green hollow rectangles).

8.2 ToF-SIMS mass spectra

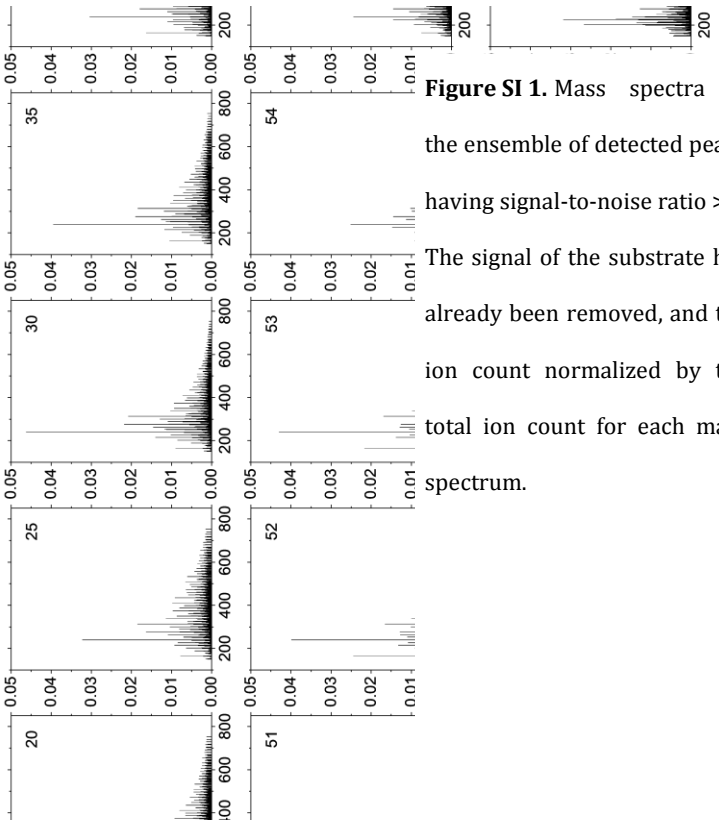


Figure SI 1. Mass spectra of the ensemble of detected peaks having signal-to-noise ratio > 3. The signal of the substrate has already been removed, and the ion count normalized by the total ion count for each mass spectrum.

8.3 ToF-SIMS mass defect analysis

Figure SI 2 shows the mass defect plot obtained from the ensemble of used to build the peak list in Table SI 1 and for the multivariate analysis.

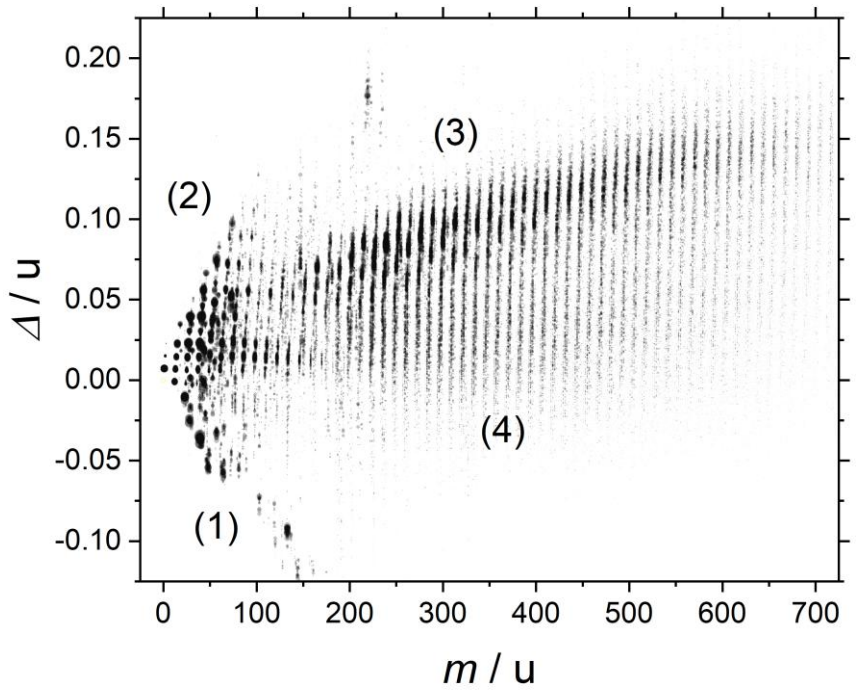


Figure SI 2. Mass defect plot showing the ensemble of detected peaks having signal-to-noise ratio > 3. The size of the datapoints is proportional to the logarithm of the peak intensity. The low mass defect series (1), also very prominent in the blank, is assigned to Ti oxides from the deposition substrates. The low m/z peaks (2) bear contributions from the blank and fragment ions. The high m/z and high mass defect peaks (3) are typical of soot and contain many identified hydrocarbons. The high m/z and low mass defect peaks (4), also typical of soot, are consistent with oxygen-containing hydrocarbons or low-hydrogen carbon cluster ions, however, certain assignments are not possible with the limited available resolving power (m/⍰m ≈ 104 at 200 m/z) [46].

Table SI 1. Peak list used in the data reduction. The 95% confidence level on the m/z is shown between brackets.

m/z	Formula	m/z	Formula	m/z	Formula	m/z	Formula	m/z	Formula
150.02(2)	?	224.06(2)	$C_{18}H_8^+$	303.12(3)	$[^{13}C]C_{24}H_{14}^+$	394.02(4)	?	511.14(5)	$C_{41}H_{19}^+$
150.04(2)	$C_{12}H_6^+$	225.03(2)	?	304.12(3)	$C_{24}H_{16}^+$	395.03(4)	?	512.15(5)	$[^{13}C]C_{41}H_{19}^+$
151.03(2)	?	225.06(2)	$[^{13}C]C_{18}H_8^+$	305.10(3)	$C_{23}H_{13}O^+$	396.04(4)	?	513.15(5)	?
151.05(2)	$[^{13}C]C_{12}H_6^+$	226.05(2)	?	306.10(3)	$[^{13}C]C_{23}H_{13}O^+$	397.06(4)	?	516.03(5)	?
152.03(2)	?	226.08(2)	$C_{18}H_{10}^+$	307.01(3)	?	398.10(4)	$C_{32}H_{14}^+$	518.05(5)	?
152.06(2)	$C_{12}H_8^+$	226.82(2)	?	309.02(3)	?	399.11(4)	$[^{13}C]C_{32}H_{14}^+$	519.06(5)	?
153.04(2)	?	227.08(2)	$[^{13}C]C_{18}H_{10}^+$	309.05(3)	$C_{21}H_9O_3^+$	400.12(4)	$C_{32}H_{16}^+$	520.12(5)	$C_{42}H_{16}^+$
153.07(2)	$[^{13}C]C_{12}H_8^+$	228.09(2)	$C_{18}H_{12}^+$	310.04(3)	$[^{13}C]C_{21}H_9O_3^+$	401.13(4)	$[^{13}C]C_{32}H_{16}^+$	521.13(5)	$[^{13}C]C_{42}H_{16}^+$
154.04(2)	?	229.07(2)	$C_{17}H_9O^+$	311.03(3)	?	402.13(4)	?	522.14(5)	$C_{42}H_{18}^+$
154.07(2)	$C_{11}H_8N^+$	229.10(2)	$[^{13}C]C_{18}H_{12}^+$	311.07(3)	$C_{21}H_{11}O_3^+$	403.12(4)	?	523.15(5)	$[^{13}C]C_{42}H_{18}^+$
155.05(2)	$C_{11}H_7O^+$	230.07(2)	$[^{13}C]C_{17}H_9O^+$	312.08(3)	$[^{13}C]C_{21}H_{11}O_3^+$	404.12(4)	$C_{31}H_{16}O^+$	524.15(5)	$C_{42}H_{20}^+$
155.09(2)	$C_{12}H_{11}^+$	230.11(2)	?	313.10(3)	$C_{25}H_{13}^+$	405.01(4)	?	525.15(5)	$[^{13}C]C_{42}H_{20}^+$
156.09(2)	$C_{11}H_{10}N^+$	231.08(2)	$C_{17}H_{11}O^+$	314.10(3)	$[^{13}C]C_{25}H_{13}^+$	405.12(4)	$[^{13}C]C_{31}H_{16}O^+$	529.04(5)	?
157.07(2)	$C_{11}H_9O^+$	232.09(2)	$[^{13}C]C_{17}H_{11}O^+$	315.11(3)	$C_{25}H_{15}^+$	407.03(4)	?	531.07(5)	?
158.10(2)	$C_{11}H_{12}N^+$	233.00(2)	?	316.10(3)	$C_{24}H_{12}O^+$	408.03(4)	?	533.13(5)	$C_{43}H_{17}^+$
159.08(2)	$C_{11}H_{11}O^+$	233.06(2)	$C_{16}H_9O_2^+$	317.10(3)	$[^{13}C]C_{24}H_{12}O^+$	409.05(4)	?	534.14(5)	$[^{13}C]C_{43}H_{17}^+$
161.00(2)	?	234.01(2)	?	318.00(3)	?	409.08(4)	?	535.15(5)	$C_{43}H_{19}^+$
161.04(2)	$C_{13}H_5^+$	235.01(2)	?	318.11(3)	?	410.08(4)	?	536.15(5)	$[^{13}C]C_{43}H_{19}^+$
162.02(2)	?	235.04(2)	$C_{15}H_7O_3^+$	319.11(3)	?	411.12(4)	$C_{33}H_{15}^+$	537.15(5)	?
162.04(2)	$[^{13}C]C_{13}H_5^+$	236.02(2)	?	320.01(3)	?	412.12(4)	$[^{13}C]C_{33}H_{15}^+$	540.03(5)	?
163.03(2)	?	236.04(2)	$[^{13}C]C_{15}H_7O_3^+$	321.02(3)	?	413.13(4)	$C_{33}H_{17}^+$	542.05(5)	?
163.05(2)	$C_{13}H_7^+$	237.03(2)	?	322.02(3)	?	414.13(4)	$[^{13}C]C_{33}H_{17}^+$	544.10(5)	?
164.03(2)	?	237.07(2)	$C_{19}H_9^+$	322.05(3)	?	415.13(4)	?	545.11(5)	?
164.06(2)	$[^{13}C]C_{13}H_7^+$	238.04(2)	?	323.05(3)	?	416.00(4)	?	546.14(5)	$C_{44}H_{18}^+$
165.07(2)	$C_{13}H_9^+$	238.07(2)	$[^{13}C]C_{19}H_9^+$	324.04(3)	?	416.12(4)	$C_{32}H_{16}O^+$	547.15(5)	$[^{13}C]C_{44}H_{18}^+$
166.08(2)	$[^{13}C]C_{13}H_9^+$	239.09(2)	$C_{19}H_{11}^+$	324.08(3)	$C_{26}H_{12}^+$	417.13(4)	$[^{13}C]C_{32}H_{16}O^+$	548.15(5)	$C_{44}H_{20}^+$
167.09(2)	$C_{13}H_{11}^+$	240.09(2)	$[^{13}C]C_{19}H_{11}^+$	325.10(3)	$[^{13}C]C_{26}H_{12}^+$	418.02(4)	?	549.16(5)	$[^{13}C]C_{44}H_{20}^+$
168.06(2)	$C_{12}H_8O^+$	241.10(2)	$C_{19}H_{13}^+$	326.11(3)	$C_{26}H_{14}^+$	420.04(4)	?	550.15(6)	?
169.07(2)	$C_{12}H_9O^+$	242.08(2)	$C_{18}H_{10}O^+$	327.12(3)	$[^{13}C]C_{26}H_{14}^+$	421.04(4)	?	553.04(6)	?
174.01(2)	?	243.08(2)	$[^{13}C]C_{18}H_{10}O^+$	328.12(3)	$C_{26}H_{16}^+$	422.10(4)	$C_{34}H_{14}^+$	555.06(6)	?
174.05(2)	$C_{14}H_6^+$	244.09(2)	$C_{18}H_{12}O^+$	329.10(3)	$C_{25}H_{13}O^+$	423.11(4)	$[^{13}C]C_{34}H_{14}^+$	557.13(6)	$C_{45}H_{17}^+$
175.02(2)	?	245.10(2)	$[^{13}C]C_{18}H_{12}O^+$	330.10(3)	$[^{13}C]C_{25}H_{13}O^+$	424.12(4)	$C_{34}H_{16}^+$	558.14(6)	$[^{13}C]C_{45}H_{17}^+$
175.05(2)	$[^{13}C]C_{14}H_6^+$	246.00(2)	?	331.00(3)	?	425.13(4)	$[^{13}C]C_{34}H_{16}^+$	559.15(6)	$C_{45}H_{19}^+$
176.03(2)	?	246.03(2)	?	333.02(3)	?	426.14(4)	$C_{34}H_{18}^+$	560.15(6)	$[^{13}C]C_{45}H_{19}^+$
176.06(2)	$C_{14}H_8^+$	247.02(2)	?	334.03(3)	?	427.14(4)	$[^{13}C]C_{34}H_{18}^+$	561.15(6)	?
177.04(2)	?	248.02(2)	?	335.03(3)	?	428.14(4)	?	562.16(6)	?
177.07(2)	$[^{13}C]C_{14}H_8^+$	248.05(2)	$C_{16}H_8O_3^+$	335.07(3)	$C_{23}H_{11}O_3^+$	428.71(4)	?	566.04(6)	?
178.05(2)	?	249.03(2)	?	336.07(3)	$[^{13}C]C_{23}H_{11}O_3^+$	429.00(4)	?	568.09(6)	?
178.08(2)	$C_{14}H_{10}^+$	249.05(2)	$[^{13}C]C_{16}H_8O_3^+$	337.04(3)	?	429.13(4)	$C_{33}H_{17}O^+$	569.09(6)	?
179.05(2)	?	250.04(3)	?	337.10(3)	$C_{27}H_{13}^+$	431.02(4)	?	570.14(6)	$C_{46}H_{18}^+$
179.09(2)	$C_{14}H_{11}^+$	250.07(3)	$C_{20}H_{10}^+$	338.10(3)	$[^{13}C]C_{27}H_{13}^+$	433.05(4)	?	571.15(6)	$[^{13}C]C_{46}H_{18}^+$
180.09(2)	$[^{13}C]C_{14}H_{11}^+$	251.08(3)	$[^{13}C]C_{20}H_{10}^+$	339.12(3)	$C_{27}H_{15}^+$	434.07(4)	?	572.15(6)	$C_{46}H_{20}^+$
181.07(2)	$C_{13}H_9O^+$	252.09(3)	$C_{20}H_{12}^+$	340.12(3)	$[^{13}C]C_{27}H_{15}^+$	435.12(4)	$C_{35}H_{15}^+$	573.16(6)	$[^{13}C]C_{46}H_{20}^+$
182.07(2)	$[^{13}C]C_{13}H_9O^+$	253.10(3)	$[^{13}C]C_{20}H_{12}^+$	341.13(3)	$C_{27}H_{17}^+$	436.12(4)	$[^{13}C]C_{35}H_{15}^+$	574.16(6)	?
183.05(2)	$C_{12}H_7O_2^+$	254.11(3)	$C_{20}H_{14}^+$	342.13(3)	?	437.13(4)	$C_{35}H_{17}^+$	577.03(6)	?
185.00(2)	?	255.08(3)	$C_{19}H_{11}O^+$	343.11(3)	$C_{26}H_{15}O^+$	438.13(4)	$[^{13}C]C_{35}H_{17}^+$	579.06(6)	?
185.03(2)	?	256.09(3)	$[^{13}C]C_{19}H_{11}O^+$	344.01(3)	?	439.14(4)	?	581.13(6)	$C_{47}H_{17}^+$
186.01(2)	?	256.99(3)	?	345.02(3)	?	440.14(4)	?	582.14(6)	$[^{13}C]C_{47}H_{17}^+$
187.02(2)	?	257.06(3)	$C_{18}H_9O_2^+$	346.03(3)	?	442.01(4)	?	583.15(6)	$C_{47}H_{19}^+$
187.05(2)	$C_{15}H_7^+$	258.01(3)	?	347.04(3)	?	444.03(4)	?	584.15(6)	$[^{13}C]C_{47}H_{19}^+$
188.03(2)	?	258.07(3)	$[^{13}C]C_{18}H_9O_2^+$	348.04(3)	?	445.04(4)	?	585.16(6)	?
188.06(2)	$[^{13}C]C_{15}H_7^+$	259.01(3)	?	348.08(3)	$C_{24}H_{12}O_3^+$	446.09(4)	$C_{32}H_{14}O_3^+$	586.16(6)	?
189.04(2)	?	259.04(3)	$C_{17}H_7O_3^+$	349.08(3)	$[^{13}C]C_{24}H_{12}O_3^+$	447.10(4)	$[^{13}C]C_{32}H_{14}O_3^+$	590.04(6)	?
189.07(2)	$C_{15}H_9^+$	260.03(3)	?	350.10(4)	$C_{28}H_{14}^+$	448.12(4)	$C_{36}H_{16}^+$	592.07(6)	?
190.05(2)	?	261.03(3)	?	351.11(4)	$[^{13}C]C_{28}H_{14}^+$	449.13(4)	$[^{13}C]C_{36}H_{16}^+$	594.14(6)	$C_{48}H_{18}^+$
190.07(2)	$[^{13}C]C_{15}H_9^+$	261.06(3)	$C_{17}H_9O_3^+$	352.12(4)	$C_{28}H_{16}^+$	450.13(5)	$C_{36}H_{18}^+$	595.15(6)	$[^{13}C]C_{48}H_{18}^+$
191.09(2)	$C_{15}H_{11}^+$	262.03(3)	?	353.13(4)	$[^{13}C]C_{28}H_{16}^+$	451.14(5)	$[^{13}C]C_{36}H_{18}^+$	596.16(6)	$C_{48}H_{20}^+$
192.09(2)	$[^{13}C]C_{15}H_{11}^+$	262.06(3)	$[^{13}C]C_{17}H_9O_3^+$	354.14(4)	$C_{28}H_{18}^+$	452.14(5)	?	597.16(6)	$[^{13}C]C_{48}H_{20}^+$
193.07(2)	$C_{14}H_9O^+$	263.08(3)	$C_{21}H_{11}^+$	355.00(4)	?	453.00(5)	?	598.16(6)	?
193.10(2)	$C_{15}H_{13}^+$	264.09(3)	$[^{13}C]C_{21}H_{11}^+$	355.12(4)	$C_{27}H_{15}O^+$	453.13(5)	$C_{35}H_{17}O^+$	603.05(6)	?
194.01(2)	?	265.10(3)	$C_{21}H_{13}^+$	357.02(4)	?	455.02(5)	?	605.10(6)	?
194.08(2)	?	266.11(3)	$[^{13}C]C_{21}H_{13}^+$	357.09(4)	$C_{26}H_{13}O_2^+$	457.05(5)	?	607.15(6)	$C_{49}H_{19}^+$
194.10(2)	$C_{14}H_{12}N^+$	267.09(3)	?	358.02(4)	?	458.05(5)	?	608.16(6)	$[^{13}C]C_{49}H_{19}^+$
195.09(2)	$C_{14}H_{11}O^+$	267.12(3)	$C_{21}H_{15}^+$	358.10(4)	$[^{13}C]C_{26}H_{13}O_2^+$	459.12(5)	$C_{37}H_{15}^+$	609.16(6)	$C_{49}H_{21}^+$
197.06(2)	$C_{13}H_9O_2^+$	268.09(3)	$C_{20}H_{12}O^+$	359.04(4)	?	460.12(5)	$[^{13}C]C_{37}H_{15}^+$	616.06(6)	?
198.01(2)	?	269.10(3)	$C_{20}H_{13}O^+$	360.05(4)	?	461.13(5)	$C_{37}H_{17}^+$	618.13(6)	$C_{50}H_{18}^+$
198.04(2)	?	270.01(3)	?	361.09(4)	$C_{29}H_{13}^+$	462.14(5)	$[^{13}C]C_{37}H_{17}^+$	619.15(6)	$[^{13}C]C_{50}H_{18}^+$
199.02(2)	?	271.02(3)	?	362.10(4)	$[^{13}C]C_{29}H_{13}^+$	463.14(5)	?	620.15(6)	$C_{50}H_{20}^+$
199.04(2)	?	272.02(3)	?	363.11(4)	$C_{29}H_{15}^+$	464.14(5)	?	621.16(6)	$[^{13}C]C_{50}H_{20}^+$
200.03(2)	?	272.05(3)	$C_{18}H_8O_3^+$	364.12(4)	$[^{13}C]C_{29}H_{15}^+$	466.01(5)	?	622.16(6)	?
200.06(2)	$C_{16}H_8^+$	273.05(3)	$[^{13}C]C_{18}H_8O_3^+$	365.13(4)	$C_{29}H_{17}^+$	468.03(5)	?	627.05(6)	?
201.04(2)	?	274.03(3)	?	366.14(4)	$[^{13}C]C_{29}H_{17}^+$	470.06(5)	?	629.09(6)	?
201.07(2)	$[^{13}C]C_{16}H_8^+$	274.07(3)	$C_{18}H_{10}O_3^+$	367.13(4)	?	471.08(5)	?	631.15(6)	$C_{51}H_{19}^+$
202.05(2)	?	275.07(3)	$[^{13}C]C_{18}H_{10}O_3^+$	368.01(4)	?	472.12(5)	$C_{38}H_{16}^+$	632.15(6)	$[^{13}C]C_{51}H_{19}^+$
202.08(2)	$C_{16}H_{10}^+$	276.09(3)	$C_{22}H_{12}^+$	370.02(4)	?	473.14(5)	$[^{13}C]C_{38}H_{16}^+$	633.16(6)	$C_{51}H_{21}^+$
203.09(2)	$[^{13}C]C_{16}H_{10}^+$	277.10(3)	$[^{13}C]C_{22}H_{12}^+$	371.03(4)	?	474.13(5)	$C_{38}H_{18}^+$	634.16(6)	$[^{13}C]C_{51}H_{21}^+$
203.83(2)	?	278.11(3)	$C_{22}H_{14}^+$	371.11(4)	$C_{27}H_{15}O_2^+$	475.14(5)	$[^{13}C]C_{38}H_{18}^+$	640.06(6)	?
204.09(2)	$C_{16}H_{12}^+$	279.08(3)	$C_{21}H_{11}O^+$	372.04(4)	?	476.14(5)	?	642.13(6)	?
204.84(2)	?	279.12(3)	$[^{13}C]C_{22}H_{14}^+$	372.08(4)	$C_{26}H_{12}O_3^+$	477.14(5)	$C_{37}H_{17}O^+$	643.14(6)	?
205.07(2)	$C_{15}H_9O^+$	280.99(3)	?	372.75(4)	?	479.02(5)	?	644.15(6)	$C_{52}H_{20}^+$
205.10(2)	$[^{13}C]C_{16}H_{12}^+$	281.10(3)	$C_{21}H_{13}O^+$	373.04(4)	?	481.05(5)	?	645.16(6)	$[^{13}C]C_{52}H_{20}^+$
206.07(2)	$[^{13}C]C_{15}H_9O^+$	283.01(3)	?	373.08(4)	$[^{13}C]C_{26}H_{12}O_3^+$	483.11(5)	$C_{39}H_{15}^+$	646.16(6)	?

206.10(2)	?	284.03(3)	?	374.10(4)	$C_{30}H_{14}^+$	484.11(5)	$[^{13}C]C_{39}H_{15}^+$	653.07(7)	?
207.08(2)	$C_{15}H_{11}O^+$	285.02(3)	?	375.11(4)	$[^{13}C]C_{30}H_{14}^+$	485.13(5)	$C_{39}H_{17}^+$	655.15(7)	$C_{53}H_{19}^+$
210.01(2)	?	285.06(3)	$C_{19}H_9O_3^+$	376.12(4)	$C_{30}H_{16}^+$	486.14(5)	$[^{13}C]C_{39}H_{17}^+$	656.16(7)	$[^{13}C]C_{53}H_{19}^+$
211.02(2)	?	286.06(3)	$[^{13}C]C_{19}H_9O_3^+$	377.12(4)	$[^{13}C]C_{30}H_{16}^+$	487.14(5)	$C_{39}H_{19}^+$	657.16(7)	$C_{53}H_{21}^+$
211.05(2)	?	287.08(3)	$C_{23}H_{11}^+$	378.13(4)	?	488.14(5)	$[^{13}C]C_{39}H_{19}^+$	664.06(7)	?
212.02(2)	?	288.08(3)	$[^{13}C]C_{23}H_{11}^+$	379.11(4)	$C_{29}H_{15}O^+$	489.15(5)	?	666.10(7)	?
212.05(2)	?	289.10(3)	$C_{23}H_{13}^+$	380.12(4)	$[^{13}C]C_{29}H_{15}O^+$	490.01(5)	?	668.15(7)	$C_{54}H_{20}^+$
213.03(2)	?	290.11(3)	$[^{13}C]C_{23}H_{13}^+$	381.02(4)	?	492.03(5)	?	669.16(7)	$[^{13}C]C_{54}H_{20}^+$
213.07(2)	$C_{17}H_9^+$	291.12(3)	$C_{23}H_{15}^+$	381.13(4)	$C_{29}H_{17}O^+$	494.06(5)	?	670.17(7)	$C_{54}H_{22}^+$
214.07(2)	$[^{13}C]C_{17}H_9^+$	292.09(3)	$C_{22}H_{12}O^+$	383.03(4)	?	495.07(5)	?	679.14(7)	$C_{55}H_{19}^+$
215.09(2)	$C_{17}H_{11}^+$	293.10(3)	$[^{13}C]C_{22}H_{12}O^+$	384.04(4)	?	496.12(5)	$C_{40}H_{16}^+$	680.15(7)	$[^{13}C]C_{55}H_{19}^+$
216.09(2)	$[^{13}C]C_{17}H_{11}^+$	294.00(3)	?	385.04(4)	?	497.13(5)	$[^{13}C]C_{40}H_{16}^+$	681.16(7)	$C_{55}H_{21}^+$
217.07(2)	?	295.01(3)	?	385.08(4)	$C_{27}H_{13}O_3^+$	498.14(5)	$C_{40}H_{18}^+$	692.15(7)	$C_{56}H_{20}^+$
217.10(2)	$C_{17}H_{13}^+$	296.01(3)	?	386.09(4)	$[^{13}C]C_{27}H_{13}O_3^+$	499.15(5)	$[^{13}C]C_{40}H_{18}^+$	693.16(7)	$[^{13}C]C_{56}H_{20}^+$
218.08(2)	$C_{16}H_{10}O^+$	296.04(3)	?	387.11(4)	$C_{31}H_{15}^+$	500.14(5)	?	694.16(7)	?
219.08(2)	$[^{13}C]C_{16}H_{10}O^+$	297.04(3)	?	388.12(4)	$[^{13}C]C_{31}H_{15}^+$	501.15(5)	?	703.13(7)	?
220.09(2)	?	298.03(3)	?	389.12(4)	?	503.03(5)	?	705.16(7)	$C_{57}H_{21}^+$
220.83(2)	?	298.06(3)	$C_{20}H_{10}O_3^+$	390.13(4)	?	505.04(5)	?	716.15(7)	$C_{58}H_{20}^+$
222.01(2)	?	299.07(3)	$[^{13}C]C_{20}H_{10}O_3^+$	391.13(4)	?	507.08(5)	?	718.16(7)	$C_{58}H_{22}^+$
222.03(2)	$C_{14}H_6O_3^+$	300.09(3)	$C_{24}H_{12}^+$	392.00(4)	?	508.10(5)	?	729.16(7)	$C_{59}H_{21}^+$
223.02(2)	?	301.10(3)	$[^{13}C]C_{24}H_{12}^+$	392.12(4)	$C_{30}H_{16}O^+$	509.14(5)	$C_{41}H_{17}^+$	740.15(7)	$C_{60}H_{20}^+$
224.03(2)	?	302.11(3)	$C_{24}H_{14}^+$	393.13(4)	$[^{13}C]C_{30}H_{16}O^+$	510.14(5)	$[^{13}C]C_{41}H_{17}^+$	753.16(8)	$C_{61}H_{21}^+$

8.4 PCA on the ToF-SIMS mass spectra

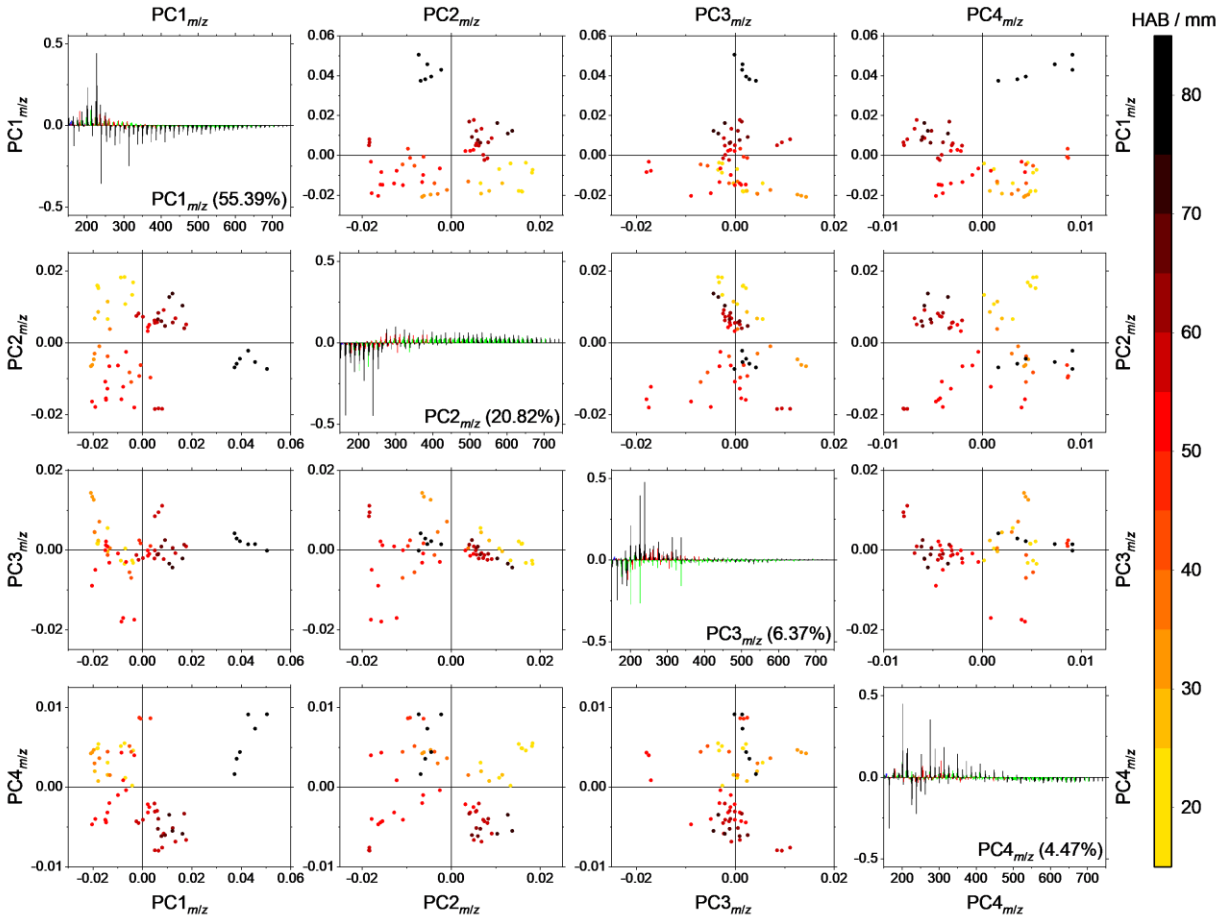


Figure SI 3. Results of the PCA (covariance matrix) performed on the mass spectra of the impaction ROIs. The figure shows the score plots of the first four principal components and the corresponding loading plots in the main diagonal. The scores are color-mapped on the sampling HAB (scale on the side of the figure). The loadings are color-mapped on the assigned formula: $C_mH_n^+$ (black), $C_mH_nO_p^+$ (red) and unassigned (green).

8.5 Calculating [H] from the Raman spectra

In Raman spectroscopy, [H] of organic films has been shown to be proportional to log $m/I(\mathrm{G})$, where m is the slope of the linear fit of the photoluminescence background and $I(\mathrm{G})$ is the intensity of the Raman G band [53,54]. In this work, for the data collected on the impaction ROIs (the data of the halo ROIs cannot be used due to the absence of the G band), log $m/I(\mathrm{G})$ shows a linear dependence ($R = 0.8706$) on [H] up to 70 mm HAB:

$$\log\frac{m}{I(\mathrm{G})} = -13(2) + 44(6) \times [\mathrm{H}] \qquad \text{Equation 8.3.}$$

At higher HAB, the linearity is lost. In Figure SI 4, the fitting function is compared to empirical equations obtained from the analysis of hydrogenated carbon films prepared by chemical vapor deposition [53,54]. The three empirical equations differ significantly in their slope and intercept. A likely explanation is in the different depth sensibility at the sample surface of the used diagnostic: ToF-SIMS in static mode (<1-2 nm, this work), nuclear reaction analysis (<2-5 nm, [53]) and elastic recoil detection analysis (<2-5 nm, [54]). On the other hand, even the shortest penetration depth of a Vis laser beam typically used for Raman spectroscopy on highly absorbing carbonaceous materials can be estimated in the range of a few μm [98]. Therefore, caution should be paid when using these equations since they compare desorption-related phenomena inherently linked to the material surface to interactions occurring nearly in the bulk.

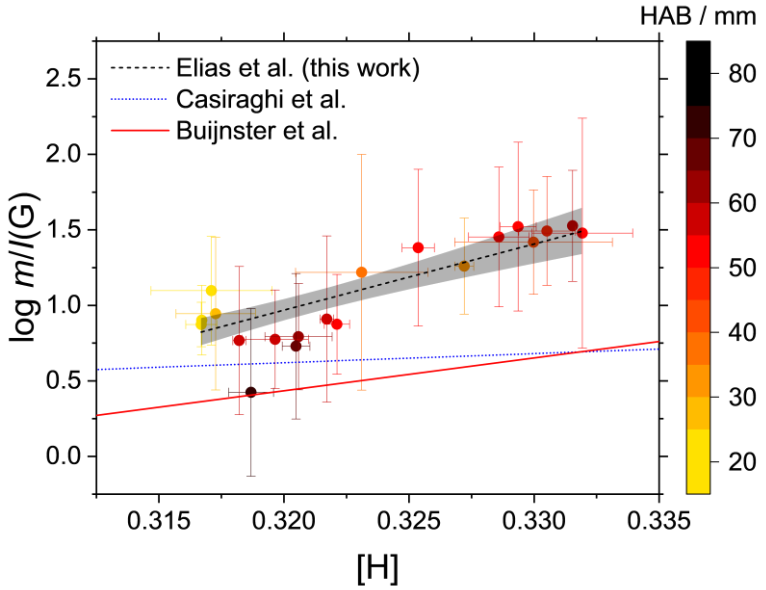


Figure SI 4. log m/I(G) against [H]. The linear fitting does not include 80-100 mm HAB datapoints, for which the linear correlation is lost. The 95% confidence band is shown (gray area). Two equations obtained from the analysis of amorphous hydrogenated carbon films prepared by chemical vapor deposition [53,54] are shown by comparison.

8.6 The Raman photoluminescence background

In the Raman spectra, the integrated photoluminescence area A_{PL} is considered to be proportional to the local concentration of condensable species deposited on the substrate and fluorescing when excited at 514.5 nm. As shown in Figure SI 5, log A_{PL} is linearly correlated to both $PC1_{An}$ and $PC1_{m/z}$. A_{PL} decreases by one order of magnitude at 53-54 mm HAB, which corresponds to the transition to the second cluster of $PC1_{An}$ in Figure 4. A_{PL} remains constant up to 60 mm HAB then decreases by another order of magnitude in 65-80 mm HAB samples, in accordance with the transition to the third cluster of $PC1_{An}$. Therefore, peaks D5, D3 and G and A_{PL} are linked to the presence of high m/z hydrocarbons. In particular, high m/z hydrocarbons in the mass spectra correspond to weak peak D3, intense peaks G and D5 and high A_{PL}. Conversely, the absence of high m/z hydrocarbons from the mass spectra corresponds to intense peak D3, weak peaks G and D5 and low A_{PL}. The A_{PL} spanning two orders of magnitude can be a consequence of

changes in the concentration in the gas phase or the deposition efficiency of the condensable hydrocarbons.

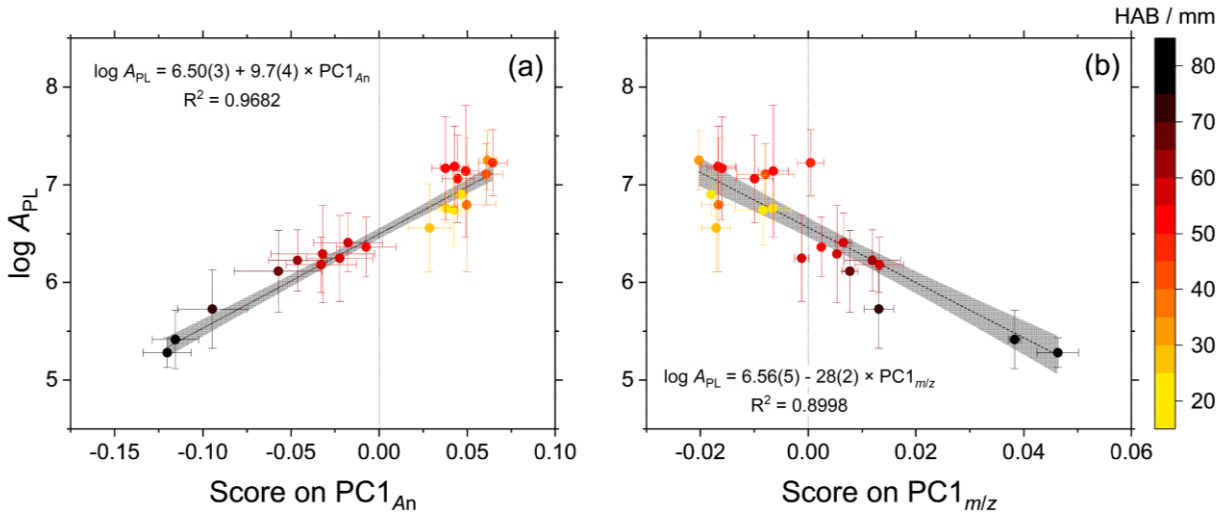
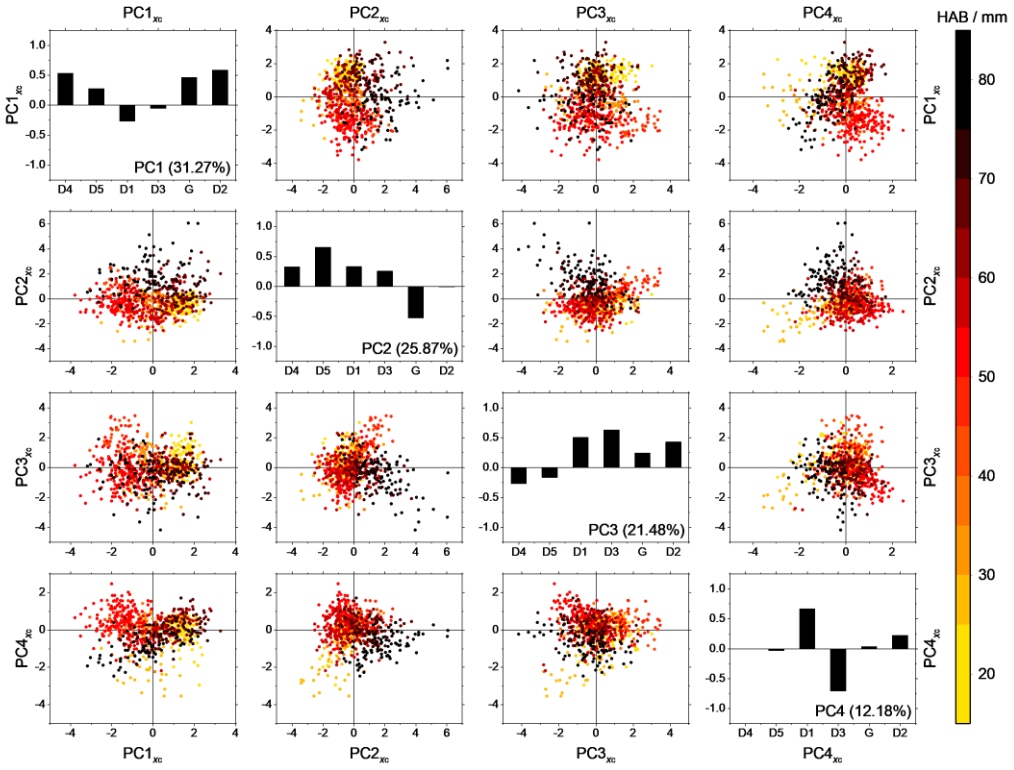
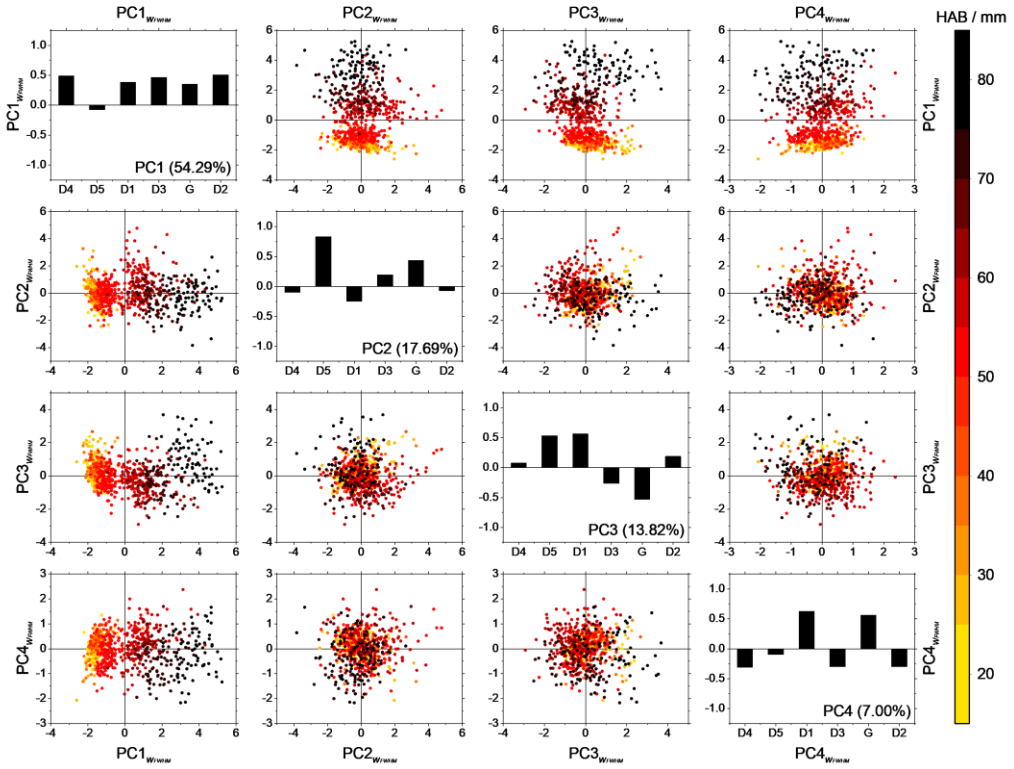


Figure SI 5. Dependences of the integrated photoluminescence area. A_{PL} against (a) scores on $PC1_{An}$, and (b) scores on $PC1_{m/z}$. The linear fittings (black dashed lines) and the 95% confidence interval (gray areas) are shown.

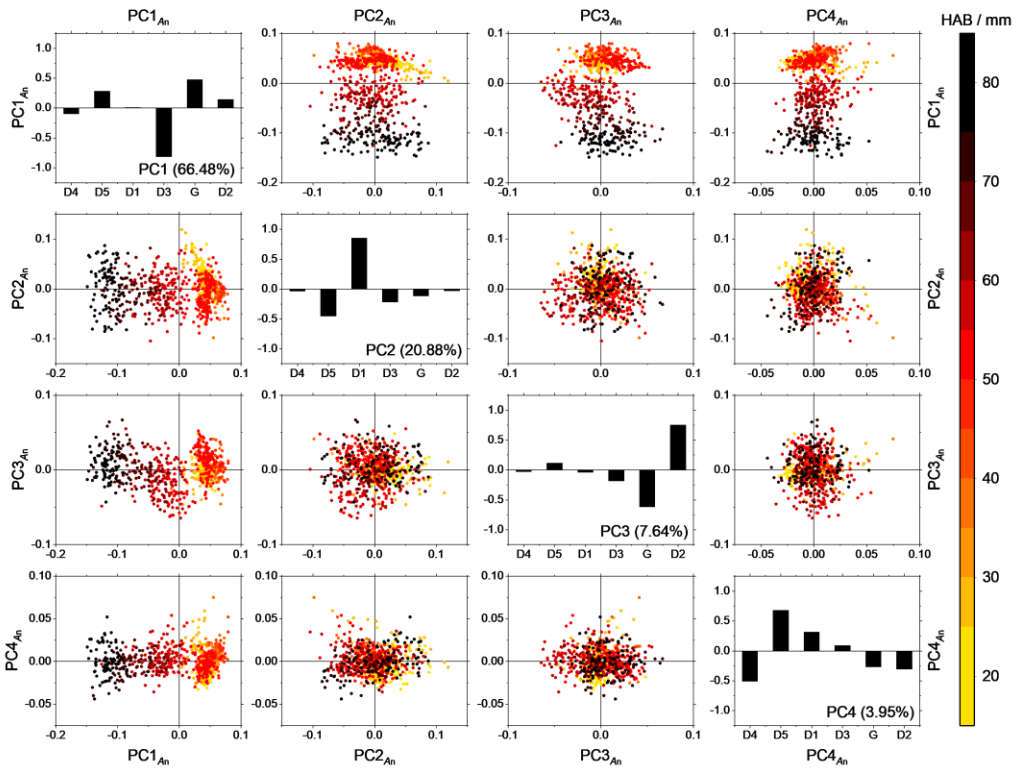
8.7 PCA on the Raman spectral parameters x_c, w_{FWHM} and A_n



(a)



(b)



(c)

959

960 **Figure SI 6. Results of the PCA performed on the Raman spectral parameters on the impaction ROIs.** (a) peak
961 positions x_c and (b) peak widths w_{FWHM} (different scales, correlation matrix is used). (c) normalized areas A_n (similar
962 scale, covariance matrix is used). The figures show the score plots of the first four principal components and the
963 corresponding loading plots in the main diagonal. The scores are color-mapped on the sampling HAB (the colormap scale
964 is on the side).

965

966 ## 8.8 Exploratory data analysis

967 Exploratory PCA is performed on the totality of the mass spectra after removing the background and
968 fragment ions (covariance matrix, 520 *m*/*z* values as variables, 150 mass spectra as observations). As a
969 rule of thumb, adding information to the score plots helps the principal components related to physical
970 phenomena, while the loading plots pinpoint the variables affecting the respective principal components.
971 Figure SI 7 shows the $PC2_{ex}$ against $PC1_{ex}$ score plot of the exploratory analysis including all available

information on the sampling HAB and analysis ROI. The corresponding Raman spectra are shown in the insets.

At 13-35 mm HAB (yellow datapoints), no statistically significant differences are found between the mass spectra of the impaction and halo ROIs that appear closely clustered together with high positive scores in $PC1_{ex}$. At 40-70 mm HAB, the mass spectra of the impaction and halo ROIs are split into different data clusters in $PC1_{ex}$: the impaction ROIs have positive scores on $PC1_{ex}$ (red-brown solid datapoints) while the halo ROIs have negative scores on $PC1_{ex}$ (red-brown hollow datapoints). In this HAB range, Raman spectra are detected only in the impaction ROIs but not in the halo ROIs. At 80-100 mm HAB, the mass spectra of the impaction and halo ROIs are found scattered with negative scores in $PC1_{ex}$. The impaction ROIs have high negative scores on $PC2_{ex}$ (black solid datapoints) while the halo ROIs have low negative or positive scores on $PC2_{ex}$ (black hollow datapoints). Again, Raman spectra are detected only in the impaction ROIs but not in the halo ROIs.

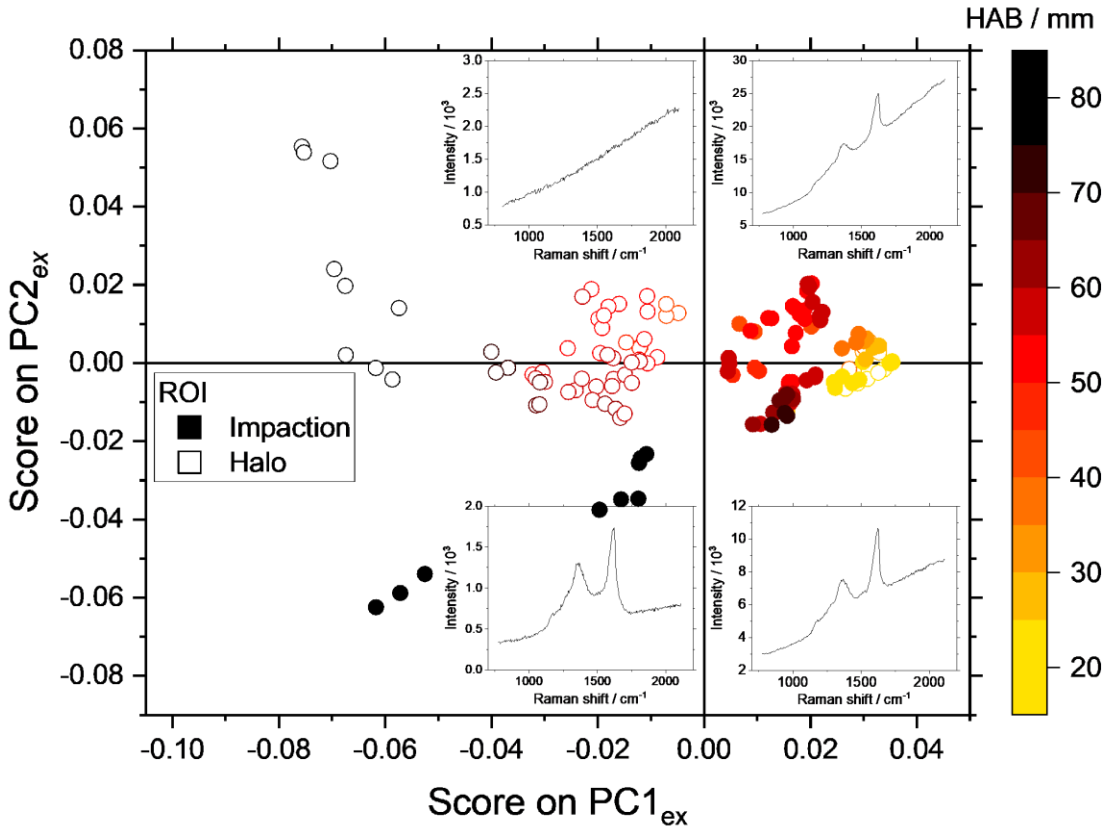


Figure SI 7. Exploratory data analysis. $PC2_{ex}$ against $PC1_{ex}$ score plot, all mass spectra in the database included. The datapoints color and interior represent the sampling HAB and the analysis location on the substrate (solid, impaction ROIs and hollow, halo ROIs), respectively. (b) The corresponding Raman spectra are shown in the insets.

From the analysis of the score plot in Figure SI 7, decreasing scores in $PC1_{ex}$ correspond to the progressively increasing sampling HAB, but also to a rather net data clustering of the impaction and halo ROIs. The phenomena responsible for this data clustering are associated with the HAB and ROI change, with the former likely related to the flame chemistry and the latter possibly related to a phase repartition during the sampling [35,49]. To focus on the flame chemistry only, the PCA is performed again on the impaction ROIs only (*Results* 3.1). However, the halo ROIs also contain information useful for understanding the formation of incipient soot particles.

8.9 Dependences of the Raman spectral parameters

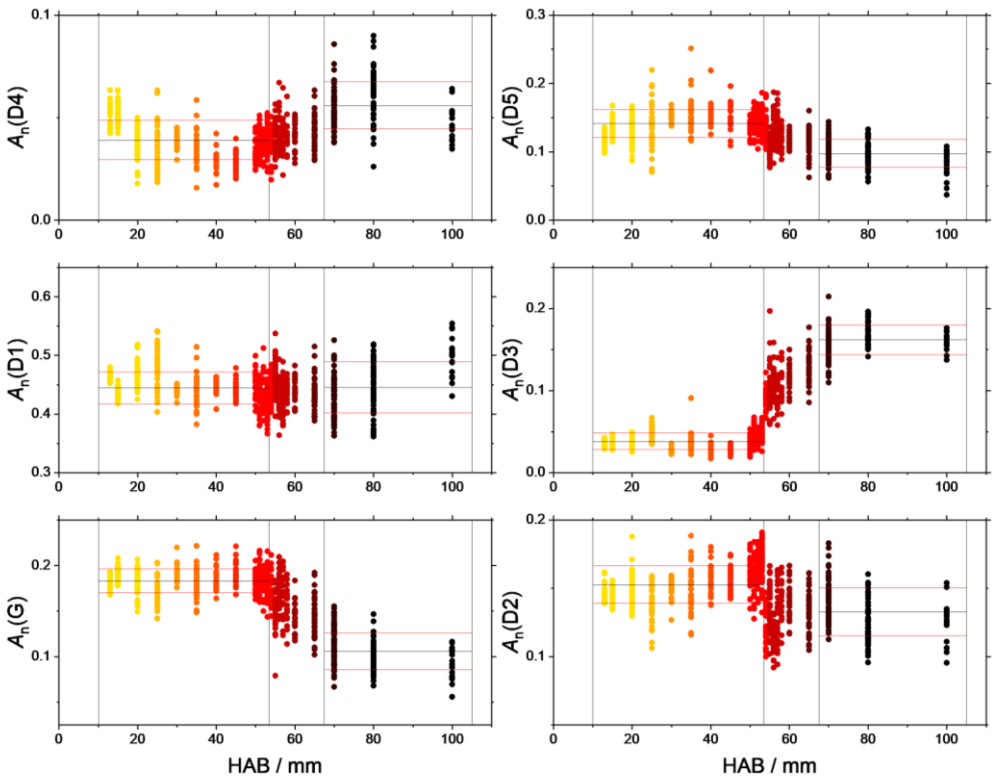


Figure SI 8. Dependence of the peak normalized areas A_n(X) on the HAB. The horizontal black and red lines represent the average and ±1 standard deviation, respectively, calculated upstream (15-53 mm HAB) and downstream (70-100 mm HAB) the flame transition region (54-60 mm HAB), and given in Table SI 2. The data are color-mapped on the sampling HAB.

Table SI 2. Average values of the fitted spectral parameters calculated upstream (15-53 mm HAB) and downstream (70-100 mm HAB) the flame transition region (54-60 mm HAB). The uncertainty represents one standard deviation.

	x_c / cm^{-1}			w_{FWHM} / cm^{-1}			A_n		
	15-53 mm	70-100 mm	Difference	15-53 mm	70-100 mm	Difference	15-53 mm	70-100 mm	Difference
D4	1170±3	1172±5	2±6	51±6	98±20	47±21	0.04±0.01	0.06±0.01	0.02±0.02
D5	1255±4	1265±7	10±8	118±10	113±12	5±16	0.14±0.02	0.10±0.02	0.04±0.03
D1	1363±5	1361±2	2±5	112±7	129±12	17±14	0.45±0.03	0.45±0.04	0.00±0.05
D3	1517±14	1522±13	5±19	99±14	162±12	63±18	0.04±0.01	0.16±0.02	0.12±0.02
G	1594±2	1590±3	4±4	50±2	55±4	5±4	0.18±0.01	0.10±0.02	0.08±0.02
D2	1618±1	1618±2	0±1	29±2	41±4	12±4	0.15±0.01	0.13±0.02	0.02±0.02

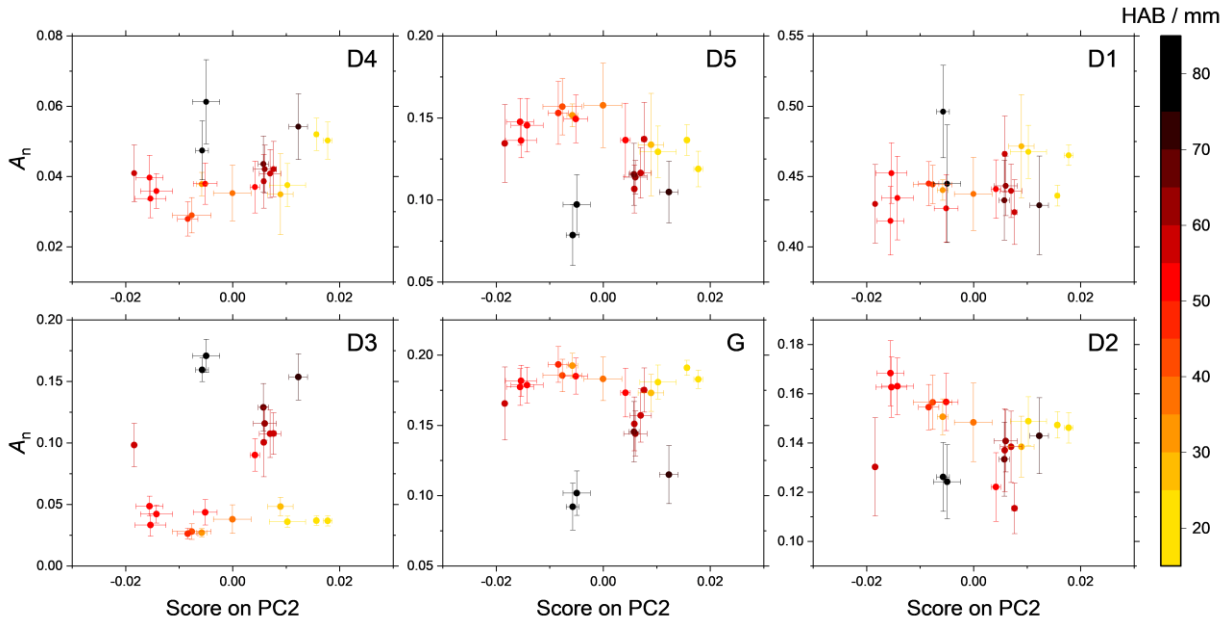


Figure SI 9. A_n against scores on PC2$_{m/z}$. Notice that, since it was not possible to perform the two analyses on the same locations on the sample surface, the average and standard deviation calculated on all available data are used instead.