



**HAL**  
open science

## Puy de Dôme Station (France): A Stoichiometric Approach to Compound Classification in Clouds

Pascal Renard, Angelica Bianco, Janne Jänis, Timo Kekäläinen, Maxime Bridoux, Laurent Deguillaume

### ► To cite this version:

Pascal Renard, Angelica Bianco, Janne Jänis, Timo Kekäläinen, Maxime Bridoux, et al.. Puy de Dôme Station (France): A Stoichiometric Approach to Compound Classification in Clouds. *Journal of Geophysical Research: Atmospheres*, 2022, 127 (16), pp.e2022JD036635. 10.1029/2022JD036635 . insu-03777320

**HAL Id: insu-03777320**

**<https://insu.hal.science/insu-03777320>**

Submitted on 6 Oct 2022

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.



L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# JGR Atmospheres

## RESEARCH ARTICLE

10.1029/2022JD036635

## Puy de Dôme Station (France): A Stoichiometric Approach to Compound Classification in Clouds

Pascal Renard<sup>1</sup>, Angelica Bianco<sup>1</sup> , Janne Jänis<sup>2</sup>, Timo Kekäläinen<sup>2</sup>, Maxime Bridoux<sup>3</sup>, and Laurent Deguillaume<sup>1,4</sup> 

<sup>1</sup>Laboratoire de Météorologie Physique, CNRS, Université Clermont Auvergne, Aubière, France, <sup>2</sup>Department of Chemistry, University of Eastern Finland, Joensuu, Finland, <sup>3</sup>CEA, DAM, DIF, Arpajon, France, <sup>4</sup>Observatoire de Physique du Globe de Clermont-Ferrand, CNRS, Université Clermont Auvergne, Aubière, France

### Key Points:

- Analyzing cloud water by high resolution-mass spectrometry shows the presence of non-polar, semi-polar and phosphorus-containing compounds
- A multidimensional stoichiometric constraint classification was applied to assign molecular formulas according to six main categories
- A statistical study highlights correlations between these categories and air mass history, such as proteins with continental boundary layer

### Supporting Information:

Supporting Information may be found in the online version of this article.

### Correspondence to:

L. Deguillaume and M. Bridoux,  
laurent.deguillaume@uca.fr;  
maxime.bridoux@cea.fr

### Citation:

Renard, P., Bianco, A., Jänis, J., Kekäläinen, T., Bridoux, M., & Deguillaume, L. (2022). Puy de Dôme station (France): A stoichiometric approach to compound classification in clouds. *Journal of Geophysical Research: Atmospheres*, 127, e2022JD036635. <https://doi.org/10.1029/2022JD036635>

Received 13 FEB 2022

Accepted 15 JUL 2022

### Author Contributions:

**Formal analysis:** Pascal Renard  
**Methodology:** Angelica Bianco, Janne Jänis, Timo Kekäläinen, Maxime Bridoux  
**Project Administration:** Maxime Bridoux, Laurent Deguillaume  
**Writing – original draft:** Pascal Renard, Angelica Bianco, Maxime Bridoux, Laurent Deguillaume

**Abstract** Seven cloud water samples were collected from May to October 2018 at the Puy de Dôme station (PUY) in France and analyzed by positive-ion atmospheric pressure photoionization [(+)APPI] Fourier transform ion cyclotron resonance mass spectrometry. The assigned formulas (ranging from 3,865 to 6,380) were attributed using the multidimensional stoichiometric constraint classification of Rivas-Ubach et al. (2018, <https://doi.org/10.1021/acs.analchem.8b00529>) to six main categories (RUCs): LipidC, ProteinC, Amino-sugarC, CarbohydrateC, NucleotideC, and OxyaromaticC. Back trajectories were calculated by the computing atmospheric trajectory tool (CAT) model to obtain information on the air mass history. Partial least square regressions were performed using chemical data, CAT back-trajectory calculations and FT-ICR MS data to analyze the environmental variability of the organic sample composition. ProteinC is correlated with the continental surface for air masses transported within the boundary layer, and Amino-sugarC is strongly correlated with acetate, NO<sub>3</sub><sup>-</sup> and NH<sub>4</sub><sup>+</sup>, suggesting Anthropogenic sources for amino sugars and proteins. LipidC is correlated with the sea surface for air masses transported within the free troposphere, confirming the long-range transport of marine biogenic sources. Concerning Oxy-aromaticC, given the correlations with oxidants and pollutants, as well as anti-correlations with local influence, we proposed a mechanism of oxidation from remote anthropogenic sources.

**Plain Language Summary** Clouds were sampled on top of the Puy de Dôme mountain in France to study their chemical composition. Cloud droplets were collected by impaction with samplers specifically designed for that. The seven samples, collected from May to October 2018, were characterized by a high-resolution mass spectrometry method, revealing thousands of organic compounds carrying carbon, hydrogen, nitrogen, sulfur, and phosphorus atoms, belonging to both natural and anthropogenic sources. Using a recently developed classification method, organic compounds were shared into classes based on their respective numbers of elements to lipids, proteins, amino sugars, carbohydrates, and aromatic compounds. To better understand the variability of the molecular fingerprints of each collected sample, we applied statistical analysis which enabled us to link the history of the air-masses calculated with a three-dimensional kinematic trajectory code and the chemical composition of the clouds. For example, proteins were related to the time spent by the air mass above continental surfaces at low altitude (in the boundary layer). Amino sugars and aromatic compounds were strongly correlated with anthropogenic sources. Finally, lipids were correlated with the time spent over the ocean in the free troposphere, confirming potential long-range transport from marine source.

## 1. Introduction

Cloud droplets contain a complex mixture of water-soluble organic matter (WSOM) originating from the scavenging of soluble gases and the dissolution of cloud condensation nuclei (Ervens et al., 2015; Herrmann et al., 2015). During the cloud lifetime, aqueous-phase reactions lead to (photo-)oxidative transformations that potentially produce small organic compounds, such as diacids and dicarbonyls (Li et al., 2020; Tomaz et al., 2018; Zhang et al., 2017) and high-molecular-weight organic matter (Herrmann et al., 2015; Li et al., 2017; Renard et al., 2014). In addition, cloud WSOM is a substrate for microorganisms (Bianco, Deguillaume, et al., 2019; Renard et al., 2016; Vaïtilingom et al., 2013; Wei et al., 2017). Most of the current analyses of cloud WSOM with classical analytical targeted technologies have contributed to a better understanding of cloud composition and chemistry (Bianco et al., 2016; Li et al., 2020; Löflund et al., 2002; Triesch et al., 2021; van Pinxteren et al., 2005). However, a significant fraction of this complex matrix remains uncharacterized.

Non-targeted approaches relying on the ultrahigh resolving power of Fourier transform ion cyclotron resonance mass spectrometry have shown great potential to unravel the extremely high molecular diversity of cloud WSOM. Indeed, FT-ICR MS provides unmatched sensitivity and mass accuracy, generating a highly resolved description of the cloud WSOM composition (Altieri et al., 2008; Bianco, Deguillaume, et al., 2019; Boone et al., 2015; Cook et al., 2017; LeClair et al., 2012; Mazzoleni et al., 2010; Zhao et al., 2013).

Recent studies have shown the presence of relatively less polar compounds (e.g., terpenoids), as well as anthropogenic aromatics, in cloud WSOM (Lebedev et al., 2018; van Pinxteren et al., 2005; Wang et al., 2015, 2020). Compared to electrospray ionization (ESI), atmospheric pressure photoionization (APPI) has seldom been used to characterize atmospheric WSOM (Nizkorodov et al., 2011). In this study, we used APPI to extend the analytical window of WSOM characterization to include its nonpolar components. Indeed, APPI provides more uniform ionization efficiency than ESI across a broader range of compound classes, such as nonpolar compounds (Marshall & Rodgers, 2008) and a wider linear dynamic range (Kauppila & Syage, 2021). Moreover, APPI minimizes ion suppression and matrix effects (Hanold et al., 2004). We used positive-ion (+)APPI for selective ionization of nitrogen-containing species (Podgorski et al., 2012).

The molecular formulas, derived from FT-ICR MS analysis, are difficult to be associated with specific molecular structures, as each represents many possible isomeric arrangements but allows instead to define elemental compound categories, such as CHO or CHNO. Elemental compound categories are usually represented graphically using van Krevelen (vK) (H/C vs. O/C) diagrams, a widely used graphical tool that allows for the analyst to plot crucial information related to the chemical composition of a sample, visualize the structural classes of organic compounds (e.g., lipid-like, protein-like) (Kew et al., 2017; Kim et al., 2003) and describe the chemical reactions in terms of additions and losses within elemental compound categories (Heald et al., 2010). Nevertheless, a large overlap exists between the vK-categories leading to incorrect classification (Brockman et al., 2018; Rivas-Ubach et al., 2018). For example, the protein-like category overlaps with other categories mainly because nitrogen is not accounted for in the classical two-dimensional 2D-vK diagram, losing information about the oxidation state (An et al., 2019). To address this issue, Rivas-Ubach et al. proposed a multidimensional stoichiometric constraint classification (MSCC) based on C, H, N, O, P and S stoichiometric ratios (Table S1 in Supporting Information S1) in six main categories (hereafter RUCs): LipidC, ProteinC, Amino-sugarC, CarbohydrateC, NucleotideC, and Oxy-aromaticC (Rivas-Ubach et al., 2018). Initially applied to plant metabolites, the MSCC has been tested toward generic databases. Rivas-Ubach et al. found that lipids (97%), oxy-aromatics (97.3%) and proteins (99.9%) from various databases matched respectively within the stoichiometric constraints of LipidC, Oxy-aromaticC and ProteinC.

The aim of this study was to characterize the molecular composition of 7 cloud water sampled at the Puy de Dôme station (PUY) in France. Previously, we used (−)ESI/FT-ICR MS (Bianco, Riva, et al., 2019); in this study, we used (+)APPI, a complementary ionization technique, allowing us to tend toward a greater exhaustiveness. Data were interpreted using the MSCC for the first time to determine the composition of atmospheric WSOM, investigating in particular the presence of sulfur and phosphorus-containing compounds. The cloud water samples were further compared using multivariate statistical tools, evaluating the influence of inorganic ions and air mass history calculated by the computing atmospheric trajectory tool (CAT) model (Baray et al., 2020; Renard et al., 2020). This approach helps to evaluate possible relationships between air mass history and WSOM variability in the cloud water samples collected at PUY.

## 2. Methods

### 2.1. Cloud Sampling and Sample Treatment

Sampling was performed at the PUY (45.77°N, 2.96°E, 1,465 m a.s.l.) in the Massif Central region (France). PUY is part of the French national platform Cézeaux-Aulnat-Opme-puy de Dôme (CO-PDD) (Baray et al., 2020) and belongs to the following international networks: the European Monitoring and Evaluation Program, Global Atmosphere Watch, and Aerosols, Clouds, and Trace Gases Research Infrastructure.

For the last 25 years, CO-PDD has evolved to become a full instrumented platform for atmospheric research. PUY is completed by additional sites located at lower altitudes and adding the vertical dimension to the atmospheric observations: Opme (660 m), Cézeaux (410 m), and Aulnat (330 m). The integration of different sites offers a unique combination of in situ and remote sensing measurements capturing and documenting the variability of

particulate and gaseous atmospheric composition, but also the optical, biochemical, and physical properties of aerosol particles, clouds, and precipitations.

Air masses reaching the PUY station were transported in the atmospheric boundary layer (ABL) and/or in the free troposphere, depending on the seasons and time of the day, but with prevailing westerly winds. The PUY summit is frequently under cloudy conditions, on average, 30% of the year, with higher occurrences during winter and autumn than during spring and summer (Baray et al., 2019). This makes PUY a reference site to study and sample clouds (Bianco et al., 2018; Lallement et al., 2018; Lebedev et al., 2018; Renard et al., 2016; Wang et al., 2020; Wirgot et al., 2017).

Seven cloud water samples from May to October 2018 were collected at PUY with the cloud impactor previously described (average sampling time of 3 hr 45 min) (Deguillaume et al., 2014).

Sampling was performed using aluminum cloud water collectors under nonprecipitating and nonfreezing conditions as described in Deguillaume et al. (2014). Such active collectors allow to collect specifically cloud droplets. They consist of a suction device and a sampling device allowing to select by inertia a lower limit size of cloud droplets (cut-off diameter around 7  $\mu\text{m}$ ). Cloud droplets were collected by impaction onto a rectangular plate and then run into a sterilized bottle.

Before cloud collection, impactors were cleaned using Milli-Q water and sterilized by autoclaving. Immediately after sampling, a fraction of the aqueous volume was filtered using a 0.2  $\mu\text{m}$  nylon filter (Fisherbrand™) to eliminate particles and microorganisms. The samples were then stored at  $-20^{\circ}\text{C}$ .

Solid-phase extraction (SPE) was used to concentrate cloud WSOM and remove the inorganic salts from the cloud water samples before FT-ICR MS analysis. The Strata-X (Phenomenex) cartridges (recommended for neutral compounds) were used for SPE. The analytical procedure is fully described in Bianco et al. (2018).

The microphysical, chemical and microbiological characterization of cloud water samples is performed in the frame of the observation service PUYCLOUD and data of more than 300 cloud water samples are available online at <https://www.opgc.fr/data-center/public/data/puycloud>. More details about the physicochemical analysis are reported in the SI. The chemical and microphysical properties of the seven cloud water samples analyzed in this study are reported in Table S2 of Supporting Information S1.

## 2.2. Dynamical Analysis

The CAT model is a three-dimensional kinematic trajectory code using initialization wind fields from the recent reanalysis European Centre for Medium-Range Weather Forecasts ERA-5 (Hoffmann et al., 2019) used and described the CAT model in the classification of clouds sampled at PUY, detailing the same procedure adopted in this study. The air mass history was modeled by counting the number of trajectory points over the sea and continental surfaces. Then, the continental and sea surfaces were vertically subdivided using the altitude of the ABL height (ABLH) interpolated on the trajectory points (data summarized in Table S2 of Supporting Information S1).

All of these characteristics were then compiled for each cloud water sample, providing a so-called “CAT matrix.” Thus, the matrices indicated, for each cloud water sample, the distribution of the zones crossed by their 72-hr backward trajectory. The relationship between the air mass history and the cloud chemical composition was the subject of a statistical analysis, as described in Renard et al. (2020). The back-trajectory plot with the CAT model displays correlation between cloud composition and air mass history (Renard et al., 2020) and the calculation of the boundary layer height indicates that the puy de Dôme summit was in free troposphere during sampling. This is also supported by the low temperature measured during sampling and the low particle concentration.

## 2.3. (+)APPI FT-ICR MS Analysis

(+)APPI analysis was performed using a 12-T Bruker solariX XR FT-ICR instrument (Bruker Daltonik GmbH, Bremen, Germany) equipped with a dynamically harmonized ICR cell (Paracell) and an APPI-II ion source fitted with a 10.6-eV krypton discharge VUV lamp (Syagen Corp.). For the (+)APPI measurements, toluene (HPLC-grade) was used as a dopant. The sample and the solution of toluene/methanol (1:1, v/v) were infused into the ion source at the same time (via the main and auxiliary sample sprayers) at flow rates of 10 and 2  $\mu\text{L min}^{-1}$ , respectively. Dry nitrogen was used as the nebulizing (4.0  $\text{L min}^{-1}$ ) and drying gas (2.5 bar,  $220^{\circ}\text{C}$ ). The mass

spectra were first calibrated externally by using APCI-L Tuning Mix (part no. G1969-85010; Agilent Technologies, Santa Clara, CA). SolarixControl 2.2 software was used for the data acquisition, and the mass spectra were further processed and analyzed with Bruker DataAnalysis 4.4 software. A total of 100 coadded time-domain transients (8 Mword each) were summed, zero-filled once and full-sine apodized to provide the final 16 Mword magnitude-mode data at an  $m/z$  range of 100–2,000. For peak picking, the signal-to-noise (S/N) ratio was set at seven, and the relative intensity threshold was  $\geq 0.01\%$ . Internal recalibration was applied using a custom-made calibration list for WSOM samples.

#### 2.4. Formula Assignments

Assigned molecular formulas were calculated from the DataAnalysis® matrix using Composer software (Sierra Analytics, Modesto, CA) as previously described in Bianco, Deguillaume, et al. (2019).

After calibration with the formula list, detected peaks in the filtered and calibrated peak lists are assigned to the closest molecular formulas from the full theoretical mass list within a mass tolerance of  $\pm 3$  ppm. To avoid any unreasonable formula assignments, we proceeded by steps, increasing the elements ranges. Step 1, the software attributed the formula with  $C_{1-10}H_{2-20}N_{0-1}O_{0-4}P_{0-0}S_{0-1}$ , step 2,  $C_{2-35}H_{4-70}N_{0-2}O_{0-5}P_{0-1}S_{0-1}$ , and step 3  $C_{4-70}H_{8-140}N_{0-4}O_{0-20}P_{0-3}S_{0-1}$ . More, rules adapted from Kind and Fiehn were applied to exclude formulas that did not occur abundantly in WSOM (e.g.,  $H/C \leq 3$ ,  $N/C \leq 1$ ,  $O/C \leq 1.5$ ,  $P/C \leq 0.33$  and  $S/C \leq 0.3$ ) (Kind & Fiehn, 2007). Mass signals found in the blank (323 low intensity peaks) were excluded (not subtracted) in Composer software. The mean carbon number (C) and mean elemental ratios (e: H/C, N/C, O/C, P/C and S/C) were converted into relative abundance weighted ( $RA_w$ ) values ( $C_w$  and  $e_w$ ) using Equations 1 and 2, respectively:

$$C_w = \frac{\sum (RA_i \times C)}{\sum RA_i} \quad (1)$$

$$e_w = \frac{\sum (RA_i \times e)}{\sum RA_i} \quad (2)$$

where  $RA_i$  is the relative abundance (RA) for each individual formula,  $i$ ,  $C$ , the number of carbon atoms and  $e$ , an elemental ratio (H/C, N/C, O/C, P/C or S/C).

#### 2.5. Statistical Analysis

Partial least squares (PLS) regressions were performed to establish correlations between the chemical concentrations ( $H_2O_2$ , inorganic carbon (IC),  $Fe^{2+}$ ,  $Na^+$ ,  $NH_4^+$ ,  $NO_3^-$ ,  $SO_4^{2-}$ , acetate and formate), CAT matrix (air mass history), and FT-ICR MS data (i.e., carbon numbers (C), elemental ratios (H/C, N/C, O/C, P/C and S/C) and aromaticity index (AI)). This panel of inorganic ions is representative of the diversity of atmospheric sources, such as  $Na^+$  for marine source,  $NH_4^+$  and  $NO_3^-$  for anthropogenic influence,  $SO_4^{2-}$  for both marine and anthropogenic influence;  $Fe^{2+}$  and  $H_2O_2$  are related to the cloud oxidative capacity; acetate and formate are proxy of organic chemistry. Mann–Whitney and Kruskal–Wallis nonparametric tests were conducted to validate significant differences between two and among several data groups, respectively (not shown). Two air mass categories were stated to be different when the probability for the groups to have identical data distributions was lower than 5% ( $p$  value  $< 0.05$ ). These statistical tests were selected because of the nonnormal distribution of the data, according to the Shapiro–Wilk normality test. A least absolute shrinkage and selection operator regression was also performed to consolidate the results (data not shown). Statistical analysis was performed using Excel XLSTAT software (Addinsoft, 2021).

### 3. Results

Agglomerative hierarchical clustering (AHC) was used to categorize cloud water samples based on the long-term monitoring of their chemical composition following the work from Renard et al. (2020). The AHC determined four categories from the six main inorganic ions ( $Cl^-$ ,  $Mg^{2+}$ ,  $Na^+$ ,  $NH_4^+$ ,  $NO_3^-$ , and  $SO_4^{2-}$ ): Highly marine, Marine, Continental, Polluted. The seven samples studied in this work belonged to the so-called “Marine” category (i.e., the ionic concentrations were low compared to those of the hundreds of clouds sampled at PUY).

**Table 1**  
Assigned Formulas, Elemental Ratios and Aromaticity Index of the 7 Cloud Water Samples Collected at PUY

Cloud water sample	3-May	9-May	24-Aug	14-Sep	24-Sep	1-Oct	8-Oct	Mean	STD <sup>a</sup>	RSD <sup>b</sup>	Min	Max
Assigned formulas <sup>c</sup>	4,026	4,372	5,788	3,764	3,713	3,699	6,625	4,570	1,168	0	3,699	6,625
C <sup>d</sup>	17.2	17.3	19.1	16.3	17.3	17.1	17.7	17.4	0.9	0.0	16.3	19.1
H/C <sup>e</sup>	1.45	1.49	1.46	1.41	1.46	1.48	1.42	1.45	0.03	0.02	1.41	1.49
N/C <sup>e</sup>	0.093	0.095	0.089	0.086	0.092	0.108	0.084	0.093	0.008	0.087	0.084	0.108
O/C <sup>e</sup>	0.199	0.175	0.199	0.195	0.188	0.176	0.183	0.188	0.010	0.055	0.175	0.199
P/C <sup>e</sup>	0.000	0.007	0.000	0.004	0.000	0.000	0.012	0.003	0.005	1.373	0.000	0.012
S/C <sup>e</sup>	0.002	0.010	0.007	0.011	0.001	0.002	0.016	0.007	0.006	0.801	0.001	0.016
Aliphatic (AI = 0) <sup>f</sup>	18.4%	20.3%	22.8%	16.9%	17.0%	19.2%	17.8%	18.9%	2.1%	11.1%	16.9%	22.8%
Olefinic (0 < AI ≤ 0.5) <sup>f</sup>	74.2%	75.7%	71.8%	74.8%	77.5%	76.5%	75.2%	75.1%	1.8%	2.4%	71.8%	77.5%
Aromatic (0.5 < AI < 0.67) <sup>f</sup>	6.9%	3.6%	4.9%	8.0%	5.1%	4.2%	6.5%	5.6%	1.6%	28.2%	3.6%	8.0%
Condensed aromatic (AI ≥ 0.67) <sup>f</sup>	0.5%	0.4%	0.4%	0.3%	0.3%	0.1%	0.6%	0.4%	0.2%	41.4%	0.1%	0.6%

<sup>a</sup>Standard deviation. <sup>b</sup>Relative standard deviation (RSD = STD/mean). <sup>c</sup>Number of distinct assigned formulas. <sup>d</sup>Relative abundance weighted (RA<sub>w</sub>) mean carbon number (C). <sup>e</sup>RA<sub>w</sub> mean elemental ratios <sup>f</sup>aromaticity index (AI: aliphatic, olefinic, aromatic and condensed aromatic).

However, each sample was derived from a unique air mass history, and significant differences in the WSOM content were detected in our study. The molecular characterization of cloud water samples is first presented in a traditional way, with elemental ratios (e.g., H/C, O/C), elemental compound categories (e.g., CHONPS are compounds containing C, H, O, N, P and S), and vK diagrams (O/C vs. H/C). We then introduce the application of MSCC (Rivas-Ubach et al., 2018) and the benefits of using this method in the complex matrix of the cloud WSOM.

### 3.1. Molecular Characterization of WSOM in Cloud Water Samples

#### 3.1.1. Elemental Ratios

Molecular formulas of the form C<sub>c</sub>H<sub>h</sub>N<sub>n</sub>O<sub>o</sub>S<sub>s</sub>P<sub>p</sub> were assigned to the (+)APPI FT-ICR mass spectra of WSOM. The total assigned formulas for each sample ranged from 3,699 to 6,625 (Table 1).

Although the APPI source minimizes ion suppression and matrix effects, the RA of compounds is affected by the molecular ionization efficiencies and is therefore not a direct measure of the concentration of a compound in the mixture. Nevertheless, assuming similar ionization efficiencies, RA-weighted values may improve the comparison between samples or between the WSOM subcategories described below. The RA-weighted mean carbon number (C) and mean elemental ratios (H/C, N/C, O/C, P/C and S/C) are displayed in Table 1. The C values (Table 1) ranged from 16.3 to 19.1, with a mean H/C of 1.45, a mean O/C of 0.19 and a mean N/C of 0.093, confirming the well-known selectivity of (+)APPI toward (a) low-polarity molecules and (b) nitrogen-containing compounds (Bianco et al., 2018; Cook et al., 2017; Zhao et al., 2013). Furthermore, the O/C ratio was relatively lower than those observed in (-)ESI (≈0.5) (Brege et al., 2018; Cook et al., 2017; Mazzoleni et al., 2010).

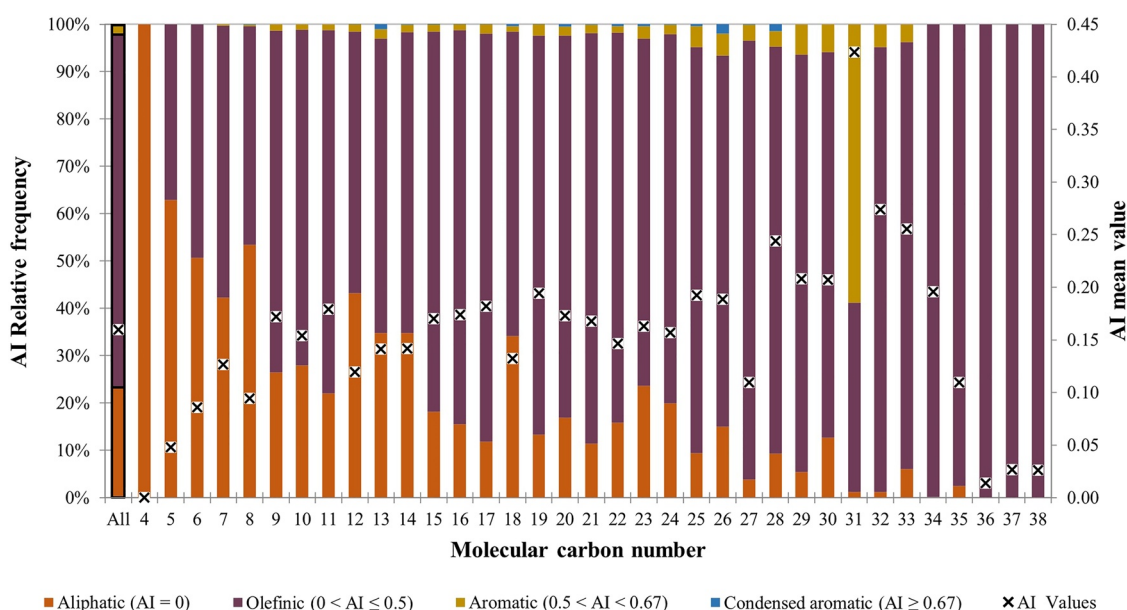
#### 3.1.2. Aromaticity Index (AI)

The weighted AI was computed according to the Melendez-Perez et al. (2016) Equation 3:

$$AI = \frac{1 + 0.5(2C - H - O - 2S - N - P)}{C} \quad (3)$$

With an average of 75.1%, olefinic (0 < AI ≤ 0.5) structures were homogeneously predominant in (+)APPI (Table 1 and Figure S1 in Supporting Information S1), followed by aliphatics (AI = 0; 18.9%), aromatics (0.5 < AI < 0.67) and condensed aromatics (AI ≥ 0.67). Compounds falling in the condensed aromatic classification were more heterogeneous (highest relative standard deviations) than aliphatic, olefinic and even aromatic compounds, reflecting the diversity of sources and the complexity of atmospheric reactivity.

Molecular trends across the seven cloud events were examined with respect to the number of C atoms in the molecular formulas (Figure 1). The AI value increased with the number of C atoms (from 4 to 31 C), following the same pattern observed in other studies focusing on cloud WSOM with ESI ionization (Mazzoleni et al., 2012; Zhao et al., 2013). The aliphaticity of compounds tends to decrease with the number of carbons (from 4 to 38 C), while the proportion of olefinic compounds increased. With an average of 3.5% of the total molecular formulas detected across all samples (from 4 to 38 C), the relative frequency of aromatic and condensed aromatic structures was generally low (Figure 1). Moreover, the mean AI was similar to that observed in (+)ESI (Brege et al., 2018).



**Figure 1.** Aromaticity index (AI) relative frequency (histogram, left axis) and AI mean value (black cross, right axis) versus molecular carbon number (seven pooled cloud water samples, collected in 2018 at PUY).

### 3.2. A Stoichiometric Approach to Compound Categorization

The vK diagrams did not reflect the complexity of thousands of compounds containing C, H, N, O, P and S, as observed in this work. However, the MSCC based on the stoichiometric ratios of these elements was recently introduced and initially applied to plant metabolites (Rivas-Ubach et al., 2018). Here, we applied the MSCC algorithm to the seven cloud water samples (Table S1 in Supporting Information S1). Using this model, we assigned cloud WSOM molecular formulas according to the six main RUCs: Amino-sugarC, CarbohydrateC, LipidC, NucleotideC, Oxy-aromaticC, and ProteinC.

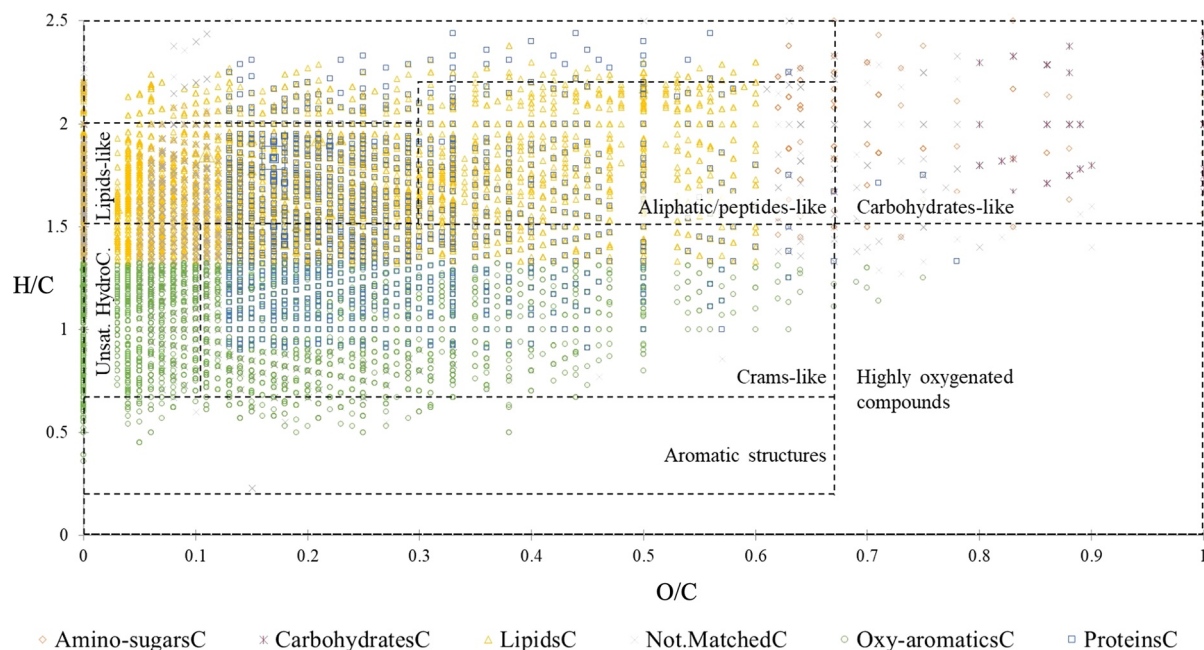
#### 3.2.1. RUCs Versus vK-Diagram

Figure 2 displays the vK diagram of the seven pooled cloud water samples, where each compound is colored according to its RUC. Using the third N-dimension, the MSCC led to a better discrimination of the classes of compounds that overlap in the 2D-vK diagram (Figure S2 in Supporting Information S1). This result is particularly obvious for ProteinC, overlapping most vK categories (Figure 2 and Figure S2 in Supporting Information S1), or for CarbohydrateC and Amino-sugarC, also overlapping in the 2D-vK diagram but distinguishable in the 3D-vK diagram (Figure S2 in Supporting Information S1). No assigned formulas fall into NucleotideC, mainly because of the P/C ratio ( $\geq 0.1$ ) in the MSCC, which is quite high for a cloud matrix. Approximately 5% of the total assigned formulas do not match any group of the MSCC (hereafter Not.MatchedC). Most of these compounds display an O/C ratio too low to be classed in ProteinC or an N/C ratio too high to be classed in LipidC (Figure 2, Table S1 in Supporting Information S1). Overall, the MSCC represents a substantial improvement over the classic 2D-vK approach (i.e., O/C and H/C ratios), as shown in Figure 2, where the distinctions are not adequate, particularly between Lipid-like and Peptide-like categories.

Elemental compound categories (e.g., CHO, CHNO, etc.) are commonly used to describe vK diagrams. Indeed, they facilitate the comparison of samples and allow for the categories to be highlighted. Here, the elemental compound categories are discussed based on the RUCs obtained for the seven pooled cloud water samples.

#### 3.2.2. RUC Relative Distribution

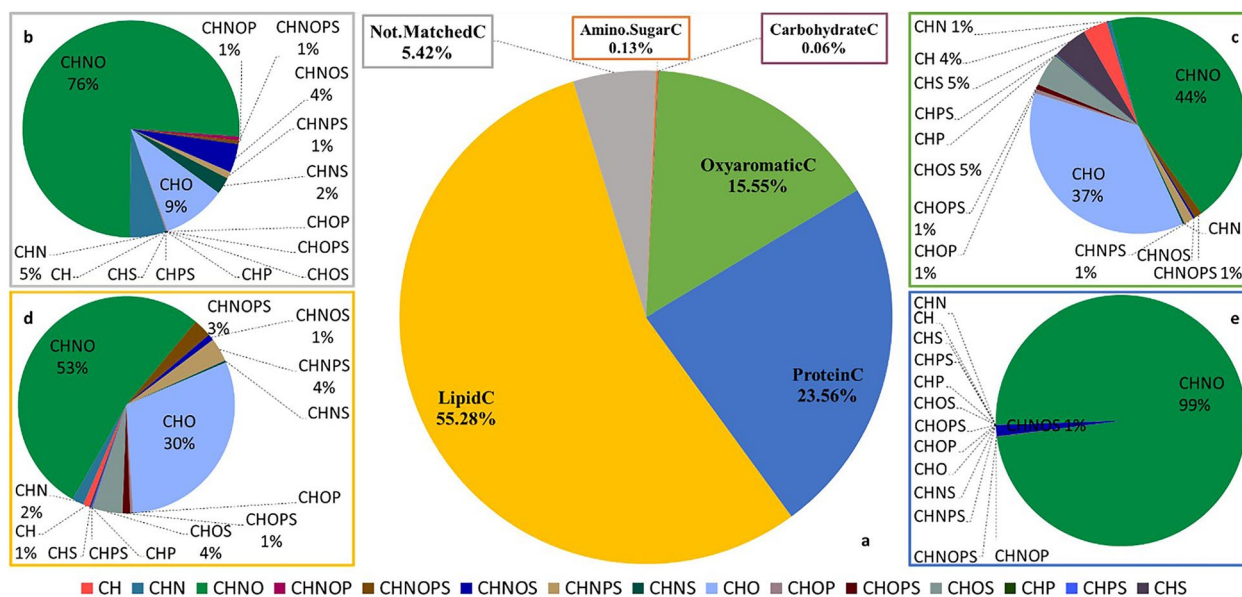
Figure 3 displays the relative distribution of RUCs for all samples and the proportions of elemental compound categories for each RUC. Figure S4 in Supporting Information S1 details the relative distribution of RUCs for each cloud water sample. LipidC, ProteinC and Oxy-aromaticC represent 94.4% of the assigned formulas, while



**Figure 2.** Van Krevelen (vK) diagrams of the seven pooled cloud water samples collected in 2018 at PUY. Rivas-Ubach et al. categories (RUCs) are reported using different colors (Amino-sugarC in orange, CarbohydrateC in purple, LipidC in yellow, Oxy-aromaticC in green, ProteinC in blue and Not.MatchedC in gray). Dashed black lines delimit standard vK classes (aliphatic/peptides-like, carboxyl-rich alicyclic molecules-like (CRAM-like), aromatic structures...) (Rivas-Ubach et al., 2018).

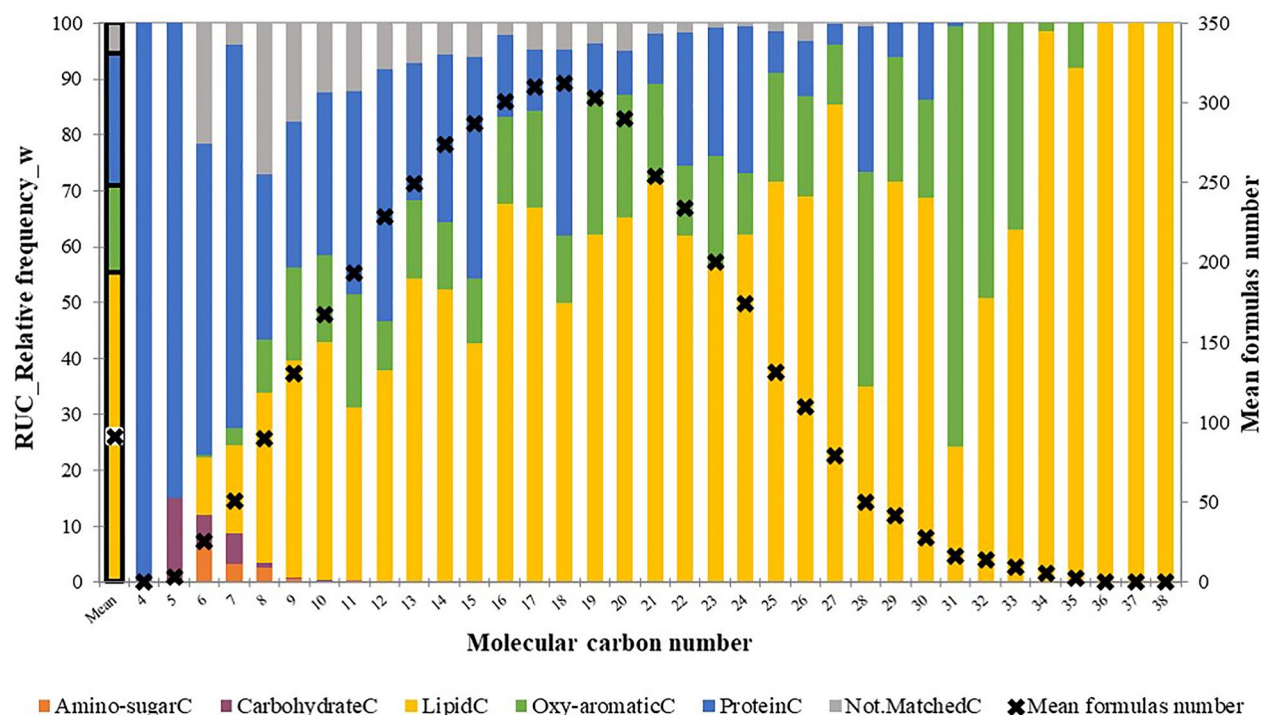
Amino-sugarC and CarbohydrateC are seldom detected, and 5.4% of formulas do not belong to any RUC (Not.MatchedC).

LipidC represents over half (55.3%) of all assigned molecular formulas. Lipids are a complex family of compounds. Fatty acids and fatty alcohols are, for example, relatively unreactive, while surfactants are highly surface-active. Elemental compound categories of LipidC reflect this complexity (Rapf et al., 2018; Triesch



**Figure 3.** (a) RUCs of the seven pooled cloud water samples collected in 2018 at PUY (Rivas-Ubach et al., 2018). RUCs are reported using different colors (Amino-sugarC in orange, CarbohydrateC in purple, LipidC in yellow, Oxy-aromaticC in green, ProteinC in blue and Not.MatchedC in gray). b–e. Elemental compound categories of each RUC, displayed with different colors (e.g., CHNO: green, CHO: light blue).





**Figure 4.** RUCs relative frequency (histogram, left axis) reported using different colors (Amino-sugarC in orange, CarbohydrateC in purple, LipidC in yellow, Oxy-aromaticC in green, ProteinC in blue and Not.MatchedC in gray), and mean formula number (black cross, right axis) versus molecular carbon number (seven pooled cloud water samples, collected in 2018 at PUY).

et al., 2021; van Pinxteren et al., 2020). CHNO and CHO in LipidC constitute 52.8% and 30.5% of the group, respectively. Phospholipids and sulfolipids compounds account for 14.0% of LipidC.

ProteinC represents 23.6% of assigned molecular formulas in (+)APPI and consists of CHNO (98.5%), CHNOP, CHNOPS and CHNOS. The assignment of CHNOS (1.44%) could suggest the presence of proteinogenic amino acids such as cysteine or methionine.

Oxy-aromaticC represents 15.5% of the assigned molecular formulas. This category cannot be described, like LipidC or ProteinC, as a chemical group per se. Oxy-aromaticC consists of compounds with low H/C ratios (<1.3), usually represented by aromatic or unsaturated compounds. These compounds contain one or more aromatic rings with diverse elemental compositions that typically participate in specific biological functions, making up a large variety of compounds along the O/C axis, such as polyphenolics (Rivas-Ubach et al., 2018). Like LipidC, Oxy-aromaticC is highly diversified in terms of elemental compound categories. CHNO is 44.1%, but CHO attains 36.6%, while P and S compounds represent 15.0% of Oxy-aromaticC.

Both LipidC and Oxy-aromaticC contain compounds without oxygen (CH, CHS, CHNP...), 7.03% and 11.5%, respectively. Oxygen not-containing compounds are mostly detected in (+)APPI (Lin et al., 2018).

Amino-sugarC represents only 0.13% of the assigned molecular formulas. Similar to ProteinC, Amino-sugarC contains only CHNO (99%) and CHNOS. Microbial amino sugars, compared to lipids, have been shown to be a relatively stable fraction of microbial biomass and could be used as biomarkers to study the balance between carbon sequestered in soils versus that released into the atmosphere (Liang et al., 2015).

### 3.2.3. RUCs Versus Carbon Number Trends

Figure 4 (right axis) displays the mean formula number per sample versus the carbon number (i.e., the number of carbon atoms, from 4 to 38 C, per formula) of all assigned molecular formulas across all 7 samples. A peak is observed at 18 C, representing 312 formulas (i.e., the average number of assigned formulas with 18 carbon atoms is 321 per cloud water sample). 99% of the assigned formulas contain 6–30 C and are normally distributed ( $18.0 \pm 7.2$  C).

Figure 4 (left axis) depicts, for each molecular carbon number, the relative frequency of RUC. ProteinC consists mainly of rather small molecules ( $15.5 \pm 4.4$  C), and amino-sugarC ( $9.4 \pm 2.4$  C) is detected from 6 to 14 C. The relative frequency of ProteinC decreases from 4 to 31 C, and among this category, CHNO is the main elemental compound category. CHNOP, CHNOPS, and CHNOS are only present from 7 to 15 C (Figure S4d in Supporting Information S1).

The relative frequency of LipidC roughly increases with increasing numbers of C atoms (from 6 to 38 C), and from 8 to 27 C, LipidC is the foremost category ( $17.7 \pm 4.8$  C). Among LipidC, P and S compounds are dominant in the C range of 24–34; otherwise, CHNO and CHO are the main elemental compound categories (Figure S4b in Supporting Information S1).

The relative frequency of Oxy-aromaticC roughly increases from 6 to 35 C ( $17.6 \pm 5.1$  C). Few small molecules ( $C < 8$ ) match in this category. Conversely, Oxy-aromaticC is predominant at 31 C. Among Oxy-aromaticC, P and S compounds are dominant above 22 C; otherwise, CHNO and CHO are the main elemental compound categories (Figure S4c in Supporting Information S1).

Overall, P and S compounds are dominant among the largest molecules (from 24 to 34 C, Figure S4a in Supporting Information S1). Furthermore, by adding heteroatoms, the representation and interpretation of elemental compound categories become increasingly complicated, contrary to the accessible RUC graphical representation. In addition, RUC could be used to perform more relevant statistical analysis on cloud water samples.

## 4. Discussion

To better understand how environmental parameters affect cloud molecular composition, multivariate statistical analysis was performed on air mass history and ionic chromatography data together with the FT-ICR MS data using PLS regressions.

### 4.1. RUC Versus FT-ICR MS Data

To perform PLS analysis (Figure S5 in Supporting Information S1), the matrix of the explanatory variables ( $X_s$ ) was composed of the FT-ICR MS data, that is, the AI, the mean carbon number (C) and mean elemental ratios (H/C, N/C, O/C, P/C and S/C). The matrix of dependent variables ( $Y_s$ ) used consisted of the RUC matrix. The PLS regression provides a predictive model, with a positive quality index ( $Q^2 = 0.33$ ) from four orthogonal components.

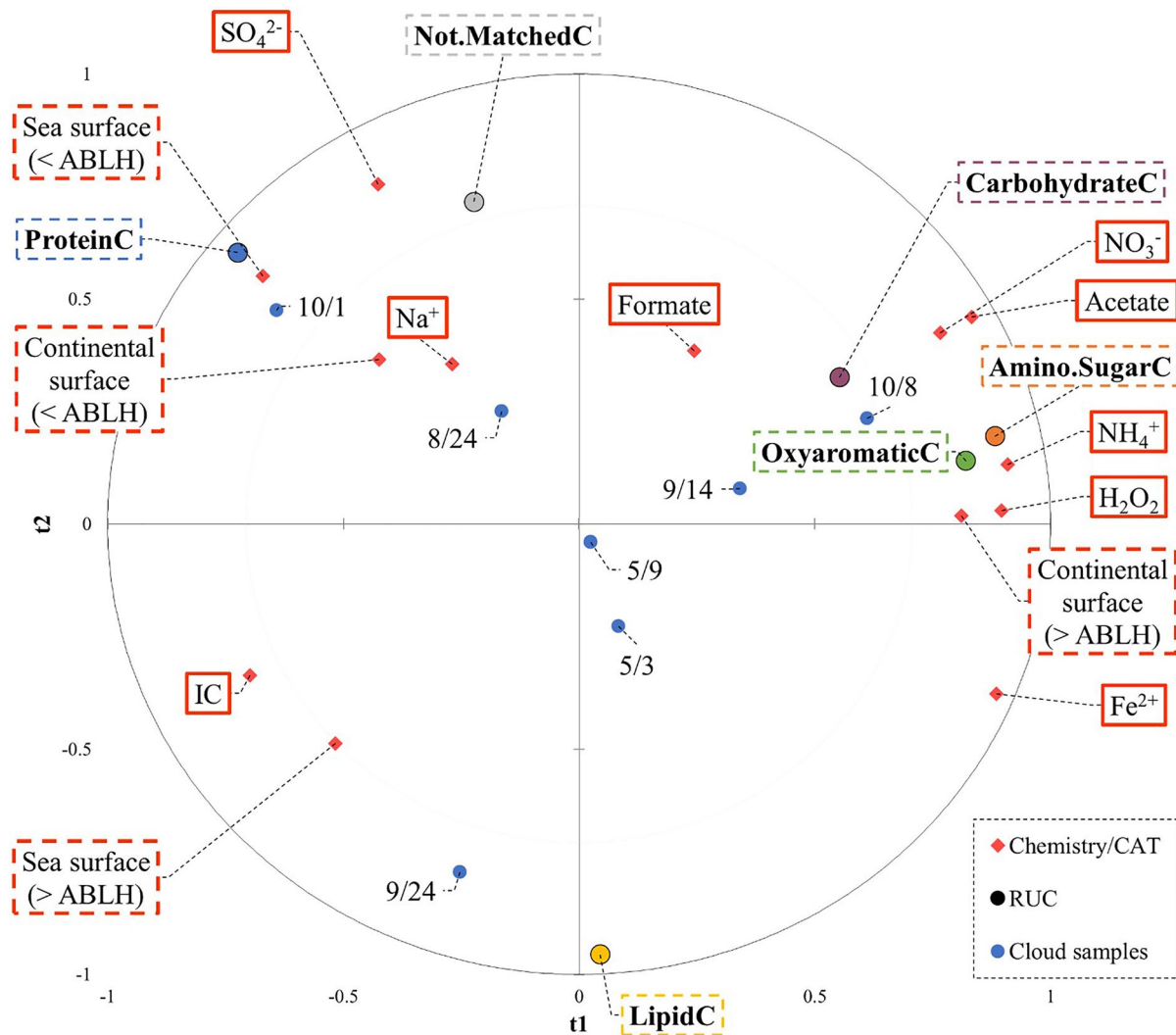
Table S3 in Supporting Information S1 details the correlations between the RUCs and the FT-ICR MS data. ProteinC was correlated (unlike other RUCs) with N/C ( $R = 0.74$ ,  $p$  value = 0.20) and with aliphatic compounds (AI = 0) ( $R = 0.67$ ,  $p$  value = 0.05). ProteinC roughly anti-correlates with all other RUCs, in particular LipidC and Oxy-aromaticC, which illustrates the importance of the N/C ratio in the classification of formulas.

LipidC did not show any noticeable (anti-)correlation, except with ProteinC ( $R = -0.60$ ,  $p$  value = 0.27). LipidC includes approximately half of the formulas assigned in FT-ICR MS. This is the main category, likely very heterogeneous. This category should be subdivided according to the same Rivas-Ubach method.

Oxy-aromaticC is consistently anti-correlated with H/C ( $R = -0.94$ ,  $p$  value = 0.003) and correlates with aromatics. Amino-sugarC is correlated with condensed aromatics ( $R = 0.69$ ,  $p$  value = 0.07). The proportion of CarbohydrateC, low among these seven cloud water samples, is anti-correlated with the N/C ratio ( $R = -0.72$ ,  $p$  value = 0.01).

P and S compounds significantly correlate with Amino-sugarC and, to a lesser extent, with CarbohydrateC and Oxy-aromaticC, suggesting P- and S-substitution.

This first PLS allows for establishing links between cloud water samples and RUCs. The clouds are too few to draw conclusions on our panel; however, the high  $Q^2$  and the correlations between the clouds and the components demonstrate that the model correctly describes all the clouds and could allow for classifying the clouds based on the FT-ICR MS data. In other words, designed to classify molecules, the MSCC method could help to classify a cloud water sample. In Renard et al. (2020) clouds sampled at PUY are classified according to their inorganic ion composition ( $\text{Cl}^-$ ,  $\text{Mg}^{2+}$ ,  $\text{Na}^+$ ,  $\text{NH}_4^+$ ,  $\text{NO}_3^-$  and  $\text{SO}_4^{2-}$ ) as marine, highly marine, continental or polluted. In



**Figure 5.** Partial least squares (PLS) chart with  $t$  components on axes  $t_1$  and  $t_2$ . The correlation map superimposes the cloud water samples and the dependent ( $Y_s$ ) and explanatory ( $X_s$ ) variables.  $Y_s$ , from the RUC matrix, are symbolized by black circles (shape fill: Amino-sugarC in orange, CarbohydrateC in purple, LipidC in yellow, Oxy-aromaticC in green, ProteinC in blue and Not.MatchedC in gray);  $X_s$ , from the chemical and computing atmospheric trajectory tool (CAT) matrices by red diamonds. The seven cloud water samples are represented by blue circles.

the same way, clouds could be classified according to their organic composition, as ProteinC (e.g., sample 10/1 in Figure S5 of Supporting Information S1), LipidC, Oxy-aromaticC.

Additional dimensions in the model, such as ionic content and information on air mass history, could link clouds with potential sources and improve the classification.

#### 4.2. Influence of Air Mass History and Chemistry at PUY

To perform the PLS analysis (Figure 5), the matrix of explanatory variables ( $X_s$ ) is composed of a CAT matrix (sea surfaces (</> ABLH) and continental surfaces (</> ABLH)) and chemical data ( $Fe^{2+}$ ,  $H_2O_2$ , IC,  $SO_4^{2-}$ ,  $NO_3^-$ , acetate, formate,  $Na^+$  and  $NH_4^+$ ). The matrix of the dependent variables ( $Y_s$ ) is an RUC matrix as well.

Table S4 in Supporting Information S1 details the correlations between the RUCs, chemistry, and CAT model. ProteinC is correlated with sea surface (<ABLH) ( $R = 0.94$ ,  $p$  value = 0.09) and, to a lesser extent, with continental surface (<ABLH). Similar correlations between amino acids and the time spent by the air mass over the sea and continental surfaces within the ABL have been observed by (Renard et al., 2022). Even though amino acids (directly measured in cloud water samples) and ProteinC (a statistical category of protein) should be compared

**Table 2**  
From the Partial Least Square (PLS) (Figure 5, With Three Components, t1, t2 and t3)

a.	RUCs	t1	t2	t3	b.	Cloud water sample	t1	t2	t3
	Amino-sugarC	<b>0.88</b>	0.20	0.03		5/3	0.08	-0.23	-0.31
	CarbohydrateC	0.55	0.33	-0.50		5/9	0.02	-0.04	<b>0.77</b>
	LipidC	0.04	<b>-0.96</b>	0.14		8/24	-0.16	0.25	-0.54
	Oxy-aromaticC	<b>0.82</b>	0.14	-0.20		9/14	0.34	0.08	-0.04
	ProteinC	<b>-0.72</b>	<b>0.60</b>	-0.14		9/24	-0.25	<b>-0.77</b>	-0.04
	Not.Matched	-0.22	<b>0.72</b>	<b>0.64</b>		10/1	<b>-0.64</b>	0.47	0.14
						10/8	<b>0.61</b>	0.23	0.02

Note. a. Correlation matrix of the variables with t (component) for each RUC, and b. Standardized scores on t for each cloud water sample. In red (in blue), the highest (anti-)correlation and standardized scores. The bold value denoted  $R > 0.6$  (or  $< -0.6$ ).

with caution, sea salt may drive amino acids to the aerosol–air interface, where they can more readily undergo chemical reactions, including peptide formation under certain conditions (Angle et al., 2022). These consistent correlations may confirm direct influences from the boundary layer in terms of nitrogen supply. ProteinC is also anti-correlated with  $\text{Fe}^{2+}$  and  $\text{H}_2\text{O}_2$  concentrations, suggesting a potential influence of photochemistry on protein (amino acid) concentrations. Sulfur-containing proteinogenic amino acids (cysteine and methionine) might be degraded directly by  $\text{H}_2\text{O}_2$  (Lundeen et al., 2014); and  $\text{Fe}^{2+}$  and  $\text{H}_2\text{O}_2$  produce, via Fenton chemistry,  $\text{HO}\cdot$  radical, known to oxidize proteins (Liu et al., 2017).

The correlations with continental surface ( $< \text{ABLH}$ ) as well as  $\text{SO}_4^{2-}$  suggest potential anthropogenic sources for ProteinC.  $\text{SO}_4^{2-}$  is known to be ubiquitous, but at PUY sulfate, nitrate and ammonium are correlated (Renard et al., 2020), consistent with an anthropogenic source of organic S capable of being transported long distances, as proposed in Altieri et al. (2009).

The origin of LipidC is heterogeneous and could explain why this category is not significantly correlated with the air mass history. Some subcategories could be related to the sea surface ( $> \text{ABLH}$ ); van Pinxteren et al. (2020) observed an accumulation of the total dissolved lipids in the sea surface microlayer, and among them terpenoids and fatty acids, lipid composition which is also found on aerosols (Triesch et al., 2021). Sunlight might initiate complex radical chemistry increasing lipids complexity in the environment (Rapf et al., 2018). LipidC is also robustly anti-correlated with  $\text{SO}_4^{2-}$ , confirming a potential marine biogenic source for LipidC. A possible explanation could be the formation of sulfolipids (Triesch et al., 2021).

Oxy-aromaticC, Amino-sugarC and, to a lesser extent, CarbohydrateC are correlated with each other and with continental surface ( $> \text{ABLH}$ ), pollutants ( $\text{NH}_4^+$  and  $\text{NO}_3^-$ ) and oxidants ( $\text{Fe}^{2+}$  and  $\text{H}_2\text{O}_2$ ), implying anthropogenic sources, according to Renard et al. (2020). Amino-sugarC could also be the result of the reactivity between Oxy-aromaticC and ProteinC.

The proportion of CarbohydrateC is low among these seven cloud water samples, but this category seems coherently anti-correlated with IC ( $R = -0.78$ ,  $p$  value = 0.02) and correlated with acetate.

The model's predictive quality index is still positive (predictive) but slightly lower than that of the previous PLS ( $Q^2 = 0.21$  with 3 components).  $Q^2$  presents significant disparities according to the RUC (detailed in Table S5 of Supporting Information S1); correlations are quite high concerning Amino-sugarC, ProteinC and Oxy-aromaticC ( $Q^2 = 0.41$ , 0.52 and 0.57), which means the PLS is predictive. Conversely, LipidC is inhomogeneous and the  $Q^2$  is negative. With only seven samples, the PLS is non-predictive for this category. The results are preliminary and need further investigation. Nevertheless, PLS regression is a powerful statistical tool when analyzing small sample sizes or data with a nonnormal distribution (Chin & Newsted, 1999).

As described above, one of the aims of this PLS regression is to classify the cloud water samples according to their organic composition. Amino-sugarC and Oxy-aromaticC are correlated to component t1 (Table 2), and then 10/8 and 9/14 samples would be in these anthropogenic RUCs (to distinguish these 2 RUCs, an additional component would be necessary, and therefore, more samples are needed). Biogenic LipidC is robustly anti-correlated

with component t2 as a 9/24 sample. ProteinC is anti-correlated with t1 and correlated with t2, as 10/1 sample with local influences.

In addition to this preliminary classification, the PLS algorithm provides a complex equation:  $Y = X W_h C_h + E_h$ , where  $Y$  is the matrix of the dependent variables and  $X$  is the matrix of the explanatory variables.  $W_h$  and  $C_h$  are the matrices generated by the PLS algorithm, and  $E_h$  is the matrix of the residuals.

The predictive model provided by this PLS needs further investigation but could improve the cloud chemistry model we developed in parallel. The objective is to have an overview of the organic sample composition from the CAT model in the absence of suitable analysis, such as FT-ICR MS data.

Essentially, MSCC allows for cloud water samples to be compared more efficiently and accurately than a vK diagram would. The RUCs provided by the MSCC could be preferentially linked to marine or continental sources in the boundary layer or free troposphere. The RUCs could thus highlight the impact of transport on the composition of the cloud water sample. Finally, in the event of an industrial accident and the release of potentially toxic compounds into the atmosphere (work in progress), the MSCC would indicate the predominant categories of molecules (e.g., proteins, lipids, etc.) present in the cloud and therefore the primary type of reactivity one would expect.

## Conflict of Interest

The authors declare no conflicts of interest relevant to this study.

## Data Availability Statement

The list of elemental formula for each sample is reported in the Reference list. The table reports in column A the label, in column B the day of sampling, in column C the elemental formula, in column D the formula type, in column E the experimental mass and in column F the total abundance. Columns G, H, I, J, K and L report the number of C, H, N, O, P and S, respectively (<https://drive.uca.fr/d/28fb3e7c0a034793891e/>). Column M reports the classification following Rivas-Ubach categories. The physico-chemical and microbiological composition of cloud water samples is reported in the database PUYCLOUD and available free of charge (<https://www.opgc.fr/data-center/public/data/puyccloud>). Rivas-Ubach software used for the classification is available free of charge (Rivas-Ubach et al., 2018).

## Acknowledgments

The authors acknowledge Sierra Analytics that provided the software Composer for data treatment. The authors also acknowledge the I-Site CAP20-25. This work was supported by CEA/CNRS through contract CEA CAJ\_16-193/C22259-Av1. The PUY station is an instrumented site of the Observatoire de Physique du Globe de Clermont-Ferrand (OPGC) observatory and LaMP laboratory, supported by the Université Clermont Auvergne, by the Center National de la Recherche Scientifique, and by the Center National d'Etudes Spatiales (CNES). The authors acknowledge the financial support from the OPGC, the Federation de Recherche en Environnement through the Contrat Plan Etat Région founded by Region Auvergne – Rhône-Alpes, the French ministry, and the Fond Européen de Développement Régional from the European community. This project was also supported by the European Union's Horizon 2020 Research and Innovation Programme (EU FT-ICR MS project; grant agreement 731077). The FT-ICR MS facility is supported by Biocenter Finland (Instruct-FI), Biocenter Kuopio and the European Regional Development Fund (grant A70135).

## References

- Addinsoft (2021). *XLSTAT statistical and data analysis solution*. XLSTAT Your Data Anal. Solut.
- Altieri, K. E., Seitzinger, S. P., Carlton, A. G., Turpin, B. J., Klein, G. C., & Marshall, A. G. (2008). Oligomers formed through in-cloud methylglyoxal reactions: Chemical composition, properties, and mechanisms investigated by ultra-high resolution FT-ICR mass spectrometry. *Atmospheric Environment*, 42(7), 1476–1490. <https://doi.org/10.1016/j.atmosenv.2007.11.015>
- Altieri, K. E., Turpin, B. J., & Seitzinger, S. P. (2009). Oligomers, organosulfates, and nitroxy organosulfates in rainwater identified by ultra-high resolution electrospray ionization FT-ICR mass spectrometry. *Atmospheric Chemistry and Physics*, 9(7), 2533–2542. <https://doi.org/10.5194/acp-9-2533-2009>
- An, Y., Xu, J., Feng, L., Zhang, X., Liu, Y., Kang, S., et al. (2019). Molecular characterization of organic aerosol in the Himalayas: Insight from ultra-high-resolution mass spectrometry. *Atmospheric Chemistry and Physics*, 19(2), 1115–1128. <https://doi.org/10.5194/acp-19-1115-2019>
- Angle, K. J., Nowak, C. M., Davasam, A., Dommer, A. C., Wauer, N. A., Amaro, R. E., & Grassian, V. H. (2022). Amino acids are driven to the interface by salts and acidic environments. *Journal of Physical Chemistry Letters*, 13(12), 2824–2829. <https://doi.org/10.1021/acs.jpcclett.2c00231>
- Baray, J.-L., Bah, A., Cacaault, P., Sellegri, K., Pichon, J.-M., Deguillaume, L., et al. (2019). Cloud occurrence frequency at puy de Dôme (France) deduced from an automatic camera image analysis: Method, validation, and comparisons with larger scale parameters. *Atmosphere*, 10(12), 808. <https://doi.org/10.3390/atmos10120808>
- Baray, J.-L., Deguillaume, L., Colomb, A., Sellegri, K., Freney, E., Rose, C., et al. (2020). Cézéaux-Aulnat-Opme-Puy De Dôme: A multi-site for the long-term survey of the tropospheric composition and climate change. *Atmospheric Measurement Techniques*, 13(6), 3413–3445. <https://doi.org/10.5194/amt-13-3413-2020>
- Bianco, A., Deguillaume, L., Chaumerliac, N., Vaïtilingom, M., Wang, M., Delort, A.-M., & Bridoux, M. C. (2019). Effect of endogenous microbiota on the molecular composition of cloud water: A study by Fourier-transform ion cyclotron resonance mass spectrometry (FT-ICR MS). *Scientific Reports*, 9(1), 7663. <https://doi.org/10.1038/s41598-019-44149-8>
- Bianco, A., Deguillaume, L., Vaïtilingom, M., Nicol, E., Baray, J.-L., Chaumerliac, N., & Bridoux, M. (2018). Molecular characterization of cloud water samples collected at the Puy de Dôme (France) by Fourier transform ion cyclotron resonance mass spectrometry. *Environmental Science & Technology*, 52(18), 10275–10285. <https://doi.org/10.1021/acs.est.8b01964>

- Bianco, A., Riva, M., Baray, J.-L., Ribeiro, M., Chaumerliac, N., George, C., et al. (2019). Chemical characterization of cloudwater collected at Puy de Dôme by FT-ICR MS reveals the presence of SOA components. *ACS Earth Space Chem*, 3(10), 2076–2087. <https://doi.org/10.1021/acsearthspacechem.9b00153>
- Bianco, A., Voyard, G., Deguillaume, L., Mailhot, G., & Brigante, M. (2016). Improving the characterization of dissolved organic carbon in cloud water: Amino acids and their impact on the oxidant capacity. *Scientific Reports*, 6(1), 37420. <https://doi.org/10.1038/srep37420>
- Boone, E. J., Laskin, A., Laskin, J., Wirth, C., Shepson, P. B., Stirm, B. H., & Pratt, K. A. (2015). Aqueous processing of atmospheric organic particles in cloud water collected via aircraft sampling. *Environmental Science & Technology*, 49(14), 8523–8530. <https://doi.org/10.1021/acs.est.5b01639>
- Brege, M., Paglione, M., Gilardoni, S., Decesari, S., Facchini, M. C., & Mazzoleni, L. R. (2018). Molecular insights on aging and aqueous-phase processing from ambient biomass burning emissions-influenced Po Valley fog and aerosol. *Atmospheric Chemistry and Physics*, 18(17), 13197–13214. <https://doi.org/10.5194/acp-18-13197-2018>
- Brockman, S. A., Roden, E. V., & Hegeman, A. D. (2018). Van Krevelen diagram visualization of high resolution-mass spectrometry metabolomics data with OpenVanKrevelen. *Metabolomics*, 14(4), 48. <https://doi.org/10.1007/s11306-018-1343-y>
- Chin, W. W., & Newsted, P. R. (1999). Structural equation modeling analysis with small samples using partial least squares. In R. Hoyle (Ed.), *Statistical strategies for small sample research* (pp. 307–341). Sage Publications. (last access: 10 January 2022) Retrieved from <https://www.researchgate.net/publication/242370645>
- Cook, R. D., Lin, Y.-H., Peng, Z., Boone, E., Chu, R. K., Dukett, J. E., et al. (2017). Biogenic, urban, and wildfire influences on the molecular composition of dissolved organic compounds in cloud water. *Atmospheric Chemistry and Physics*, 17(24), 15167–15180. <https://doi.org/10.5194/acp-17-15167-2017>
- Deguillaume, L., Charbouillot, T., Joly, M., Vaütilingom, M., Parazols, M., Marinoni, A., et al. (2014). Classification of clouds sampled at the puy de Dôme (France) based on 10 yr of monitoring of their physicochemical properties. *Atmospheric Chemistry and Physics*, 14(3), 1485–1506. <https://doi.org/10.5194/acp-14-1485-2014>
- Ervens, B., Renard, P., Tlili, S., Ravier, S., Clément, J.-L., & Monod, A. (2015). Aqueous-phase oligomerization of methyl vinyl ketone through photooxidation – Part 2: Development of the chemical mechanism and atmospheric implications. *Atmospheric Chemistry and Physics*, 15(16), 9109–9127. <https://doi.org/10.5194/acp-15-9109-2015>
- Hanold, K. A., Fischer, S. M., Cormia, P. H., Miller, C. E., & Syage, J. A. (2004). Atmospheric pressure photoionization. 1. General properties for LC/MS. *Analytical Chemistry*, 76(10), 2842–2851. <https://doi.org/10.1021/ac035442i>
- Heald, C. L., Kroll, J. H., Jimenez, J. L., Docherty, K. S., DeCarlo, P. F., Aiken, A. C., et al. (2010). A simplified description of the evolution of organic aerosol composition in the atmosphere: Van Krevelen diagram of organic aerosol. *Geophysical Research Letters*, 37(8), L08803. <https://doi.org/10.1029/2010GL042737>
- Herrmann, H., Schaefer, T., Tilgner, A., Styler, S. A., Weller, C., Teich, M., & Otto, T. (2015). Tropospheric aqueous-phase chemistry: Kinetics, mechanisms, and its coupling to a changing gas phase. *Chemistry Review*, 115(10), 4259–4334. <https://doi.org/10.1021/cr500447k>
- Hoffmann, L., Günther, G., Li, D., Stein, O., Wu, X., Griessbach, S., et al. (2019). From ERA-Interim to ERA5: The considerable impact of ECMWF's next-generation reanalysis on Lagrangian transport simulations. *Atmospheric Chemistry and Physics*, 19(5), 3097–3124. <https://doi.org/10.5194/acp-19-3097-2019>
- Kauppila, T., & Syage, J. (2021). Photoionization at elevated or atmospheric pressure: Applications of APPI and LPPI (pp. 267–303). <https://doi.org/10.1002/9783527682201.ch8>
- Kew, W., Blackburn, J. W. T., Clarke, D. J., & Uhrin, D. (2017). Interactive van Krevelen diagrams - Advanced visualisation of mass spectrometry data of complex mixtures: Interactive van Krevelen Diagrams. *Rapid Communications in Mass Spectrometry*, 31(7), 658–662. <https://doi.org/10.1002/rcm.7823>
- Kim, S., Kramer, R. W., & Hatcher, P. G. (2003). Graphical method for analysis of ultrahigh-resolution broadband mass spectra of natural organic matter, the van Krevelen diagram. *Analytical Chemistry*, 75(20), 5336–5344. <https://doi.org/10.1021/ac034415p>
- Kind, T., & Fiehn, O. (2007). Seven Golden Rules for heuristic filtering of molecular formulas obtained by accurate mass spectrometry. *BMC Bioinformatics*, 8(1), 105. <https://doi.org/10.1186/1471-2105-8-105>
- Lallement, A., Besaury, L., Tixier, E., Sancelme, M., Amato, P., Vinatier, V., et al. (2018). Potential for phenol biodegradation in cloud waters. *Biogeosciences*, 15(18), 5733–5744. <https://doi.org/10.5194/bg-15-5733-2018>
- Lebedev, A., Polyakova, O. V., Mazur, M., Artaev, V. B., Canet, I., Lallement, A., et al. (2018). Detection of semi-volatile compounds in cloud waters by GCxGC-TOF-MS. Evidence of phenols and phthalates as priority pollutants. *Environmental Pollution*, 241, 616–625. <https://doi.org/10.1016/j.envpol.2018.05.089>
- LeClair, J. P., Collett, J. L., & Mazzoleni, L. R. (2012). Fragmentation analysis of water-soluble atmospheric organic matter using ultrahigh-resolution FT-ICR Mass Spectrometry. *Environmental Science & Technology*, 46(8), 4312–4322. <https://doi.org/10.1021/es203509b>
- Li, J., Wang, X., Chen, J., Zhu, C., Li, W., Li, C., et al. (2017). Chemical composition and droplet size distribution of cloud at the summit of Mount Tai, China. *Atmospheric Chemistry and Physics*, 17(16), 9885–9896. <https://doi.org/10.5194/acp-17-9885-2017>
- Li, T., Wang, Z., Wang, Y., Wu, C., Liang, Y., Xia, M., et al. (2020). Chemical characteristics of cloud water and the impacts on aerosol properties at a subtropical mountain site in Hong Kong SAR. *Atmospheric Chemistry and Physics*, 20(1), 391–407. <https://doi.org/10.5194/acp-20-391-2020>
- Liang, C., Gutknecht, J., & Balsler, T. (2015). Microbial lipid and amino sugar responses to long-term simulated global environmental changes in a California annual grassland. *Frontiers in Microbiology*, 6. <https://doi.org/10.3389/fmicb.2015.00385>
- Lin, P., Fleming, L. T., Nizkorodov, S. A., Laskin, J., & Laskin, A. (2018). Comprehensive molecular characterization of atmospheric Brown carbon by high resolution mass spectrometry with electrospray and atmospheric pressure photoionization. *Analytical Chemistry*, 90(21), 12493–12502. <https://doi.org/10.1021/acs.analchem.8b02177>
- Liu, F., Lai, S., Tong, H., Lakey, P. S. J., Shiraiwa, M., Weller, M. G., et al. (2017). Release of free amino acids upon oxidation of peptides and proteins by hydroxyl radicals. *Analytical and Bioanalytical Chemistry*, 409(9), 2411–2420. <https://doi.org/10.1007/s00216-017-0188-y>
- Löflund, M., Kasper-Giebl, A., Schuster, B., Giebl, H., Hitznerberger, R., & Puxbaum, H. (2002). Formic, acetic, oxalic, malonic and succinic acid concentrations and their contribution to organic carbon in cloud water. *Atmospheric Environment*, 36(9), 1553–1558. [https://doi.org/10.1016/S1352-2310\(01\)00573-8](https://doi.org/10.1016/S1352-2310(01)00573-8)
- Lundeen, R. A., Janssen, E. M.-L., Chu, C., & McNeill, K. (2014). Environmental photochemistry of amino acids, peptides and proteins. *CHIMIA International Journal for Chemistry*, 68(11), 812–817. <https://doi.org/10.2533/chimia.2014.812>
- Marshall, A. G., & Rodgers, R. P. (2008). Petroleomics: Chemistry of the underworld. *Proceedings of the National Academy of Sciences*, 105(47), 18090–18095. <https://doi.org/10.1073/pnas.0805069105>

- Mazzoleni, L. R., Ehrmann, B. M., Shen, X., Marshall, A. G., & Collett, J. L. (2010). Water-soluble atmospheric organic matter in fog: Exact masses and chemical formula identification by ultrahigh-resolution fourier transform ion cyclotron resonance mass spectrometry. *Environmental Science & Technology*, *44*(10), 3690–3697. <https://doi.org/10.1021/es903409k>
- Mazzoleni, L. R., Saranjampour, P., Dalbec, M. M., Samburova, V., Hallar, A. G., Zielinska, B., et al. (2012). Identification of water-soluble organic carbon in non-urban aerosols using ultrahigh-resolution FT-ICR mass spectrometry: Organic anions. *Environmental Chemistry*, *9*(3), 285. <https://doi.org/10.1071/EN11167>
- Melendez-Perez, J. J., Martínez-Mejía, M. J., & Eberlin, M. N. (2016). A reformulated aromaticity index equation under consideration for non-aromatic and non-condensed aromatic cyclic carbonyl compounds. *Organic Geochemistry*, *95*, 29–33. <https://doi.org/10.1016/j.orggeochem.2016.02.002>
- Nizkorodov, S. A., Laskin, J., & Laskin, A. (2011). Molecular chemistry of organic aerosols through the application of high resolution mass spectrometry. *Solvent Extr.*, *19*.
- Podgorski, D. C., McKenna, A. M., Rodgers, R. P., Marshall, A. G., & Cooper, W. T. (2012). Selective ionization of dissolved organic nitrogen by positive ion atmospheric pressure photoionization coupled with fourier transform ion cyclotron resonance mass spectrometry. *Analytical Chemistry*, *6*(11), 5085–5090. <https://doi.org/10.1021/ac300800w>
- Rapf, R. J., Perkins, R. J., Dooley, M. R., Kroll, J. A., Carpenter, B. K., & Vaida, V. (2018). Environmental processing of lipids driven by aqueous photochemistry of  $\alpha$ -keto acids. *ACS Central Science*, *4*(5), 624–630. <https://doi.org/10.1021/acscentsci.8b00124>
- Renard, P., Bianco, A., Baray, J.-L., Bridoux, M., Delort, A.-M., & Deguillaume, L. (2020). Classification of clouds sampled at the puy de Dôme station (France) based on chemical measurements and air mass history matrices. *Atmosphere*, *11*(7), 732. <https://doi.org/10.3390/atmos11070732>
- Renard, P., Brissy, M., Rossi, F., Leremboure, M., Jaber, S., Baray, J.-L., et al. (2022). Free amino acid quantification in cloud water at the Puy de Dôme station (France). *Atmospheric Chemistry and Physics*, *22*(4), 2467–2486. <https://doi.org/10.5194/acp-22-2467-2022>
- Renard, P., Canet, I., Sancelme, M., Wirgot, N., Deguillaume, L., & Delort, A.-M. (2016). Screening of cloud microorganisms isolated at the Puy de Dôme (France) station for the production of biosurfactants. *Atmospheric Chemistry and Physics*, *16*(18), 12347–12358. <https://doi.org/10.5194/acp-16-12347-2016>
- Renard, P., Reed Harris, A. E., Rapf, R. J., Ravier, S., Demelas, C., Coulomb, B., et al. (2014). Aqueous phase oligomerization of methyl vinyl ketone by atmospheric radical reactions. *Journal of Physical Chemistry C*, *118*(50), 29421–29430. <https://doi.org/10.1021/jp5065598>
- Rivas-Ubach, A., Liu, Y., Bianchi, T. S., Tolić, N., Jansson, C., & Paša-Tolić, L. (2018). Moving beyond the van Krevelen diagram: A new stoichiometric approach for compound classification in organisms. *Analytical Chemistry*, *90*(10), 6152–6160. <https://doi.org/10.1021/acs.analchem.8b00529>
- Tomaz, S., Cui, T., Chen, Y., Sexton, K. G., Roberts, J. M., Warneke, C., et al. (2018). Photochemical cloud processing of primary wildfire emissions as a potential source of secondary organic aerosol. *Environmental Science & Technology*, *52*(19), 11027–11037. <https://doi.org/10.1021/acs.est.8b03293>
- Triesch, N., van Pinxteren, M., Engel, A., & Herrmann, H. (2021). Concerted measurements of free amino acids at the Cabo Verde islands: High enrichments in submicron sea spray aerosol particles and cloud droplets. *Atmospheric Chemistry and Physics*, *21*(1), 163–181. <https://doi.org/10.5194/acp-21-163-2021>
- Vaitilingom, M., Deguillaume, L., Vinatier, V., Sancelme, M., Amato, P., Chaumerliac, N., & Delort, A.-M. (2013). Potential impact of microbial activity on the oxidant capacity and organic carbon budget in clouds. *Proceedings of the National Academy of Sciences*, *110*(2), 559–564. <https://doi.org/10.1073/pnas.1205743110>
- van Pinxteren, D., Plewka, A., Hofmann, D., Müller, K., Kramberger, H., Svrčina, B., et al. (2005). Schmücke hill cap cloud and valley stations aerosol characterisation during FEBUKO (II): Organic compounds. *Atmospheric Environment*, *39*(23–24), 4305–4320. <https://doi.org/10.1016/j.atmosenv.2005.02.014>
- van Pinxteren, M., Fomba, K. W., Triesch, N., Stolle, C., Wurl, O., Bahlmann, E., et al. (2020). Marine organic matter in the remote environment of the Cape Verde islands – An introduction and overview to the MarParCloud campaign. *Atmospheric Chemistry and Physics*, *20*(11), 6921–6951. <https://doi.org/10.5194/acp-20-6921-2020>
- Wang, M., Perroux, H., Fleuret, J., Bianco, A., Bouvier, L., Colomb, A., et al. (2020). Anthropogenic and biogenic hydrophobic VOCs detected in clouds at the puy de Dôme station using Stir Bar Sorptive Extraction: Deviation from the Henry's law prediction. *Atmospheric Research*, *237*, 104844. <https://doi.org/10.1016/j.atmosres.2020.104844>
- Wang, R., Wang, Y., Li, H., Yang, M., Sun, L., Wang, T., & Wang, W. (2015). Cloud deposition of PAHs at Mount Lushan in southern China. *Science of the Total Environment*, *526*, 329–337. <https://doi.org/10.1016/j.scitotenv.2015.04.047>
- Wei, M., Xu, C., Chen, J., Zhu, C., Li, J., & Lv, G. (2017). Characteristics of bacterial community in cloud water at Mt Tai: Similarity and disparity under polluted and non-polluted cloud episodes. *Atmospheric Chemistry and Physics*, *17*(8), 5253–5270. <https://doi.org/10.5194/acp-17-5253-2017>
- Wirgot, N., Vinatier, V., Deguillaume, L., Sancelme, M., & Delort, A.-M. (2017). H<sub>2</sub>O<sub>2</sub> modulates the energetic metabolism of the cloud microbiome. *Atmospheric Chemistry and Physics*, *17*(24), 14841–14851. <https://doi.org/10.5194/acp-17-14841-2017>
- Zhang, G., Lin, Q., Peng, L., Yang, Y., Fu, Y., Bi, X., et al. (2017). Insight into the in-cloud formation of oxalate based on in situ measurement by single particle mass spectrometry. *Atmospheric Chemistry and Physics*, *17*(22), 13891–13901. <https://doi.org/10.5194/acp-17-13891-2017>
- Zhao, Y., Hallar, A. G., & Mazzoleni, L. R. (2013). Atmospheric organic matter in clouds: Exact masses and molecular formula identification using ultrahigh-resolution FT-ICR mass spectrometry. *Atmospheric Chemistry and Physics*, *13*(24), 12343–12362. <https://doi.org/10.5194/acp-13-12343-2013>