



HAL
open science

Unravelling the molecular significance of Rock-Eval® S2 pyrograms from soil and geological samples

Jérémy Jacob, Frédéric Delarue, Yoann Copard, Claude Le Milbeau, Laurent Grasset, Patrick Brockmann

► To cite this version:

Jérémy Jacob, Frédéric Delarue, Yoann Copard, Claude Le Milbeau, Laurent Grasset, et al.. Unravelling the molecular significance of Rock-Eval® S2 pyrograms from soil and geological samples. *Journal of Analytical and Applied Pyrolysis*, 2025, 191, pp.107211. <10.1016/j.jaap.2025.107211>. <insu-05101532>

HAL Id: insu-05101532

<https://insu.hal.science/insu-05101532v1>

Submitted on 6 Jun 2025

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.



Distributed under a Creative Commons CC BY 4.0 - Attribution - International License

1 **Unravelling the molecular significance of Rock-Eval® S2 pyrograms from soil and geological samples**

2

3 Jérémy Jacob¹, Frédéric Delarue², Yoann Copard³, Claude Le Milbeau⁴, Laurent Grasset⁵, Patrick

4 Brockmann¹

5

6 ¹ Laboratoire des Sciences du Climat et de l'Environnement, CEA-CNRS-UVSQ, Université

7 Paris-Saclay, 91198, Gif-sur-Yvette, France

8 ² Sorbonne Université, UPMC, CNRS, EPHE, PSL, UMR 7619 METIS, 4 place Jussieu, F-75005 Paris,

9 France

10 ³ Normandie Univ, UNIROUEN, UNICAEN, CNRS, M2C, France, 76000 Rouen, France

11 ⁴ Université d'Orléans, CNRS/INSU, BRGM, Institut des Sciences de la Terre d'Orléans (ISTO), UMR

12 7327, 45071 Orléans, France

13 ⁵ Institut de Chimie des Milieux et Matériaux de Poitiers, UMR 7285 CNRS, Université de Poitiers,

14 Poitiers, France

15

16 **Abstract**

17 Rock-Eval® pyrolysis is dedicated to rapidly determine the quality and quantity of organic matter

18 (OM) in environmental and geological samples. It has been proposed that detailed examination of S2

19 pyrograms acquired from a Flame Ionization Detector during programmed pyrolysis of samples can

20 provide complementary information to classical Rock-Eval® parameters. Previous mathematical

21 deconvolution of the S2 pyrogram has been suggested to semi-quantify distinct thermal fractions

22 that are supposed distinct in terms of chemical and/or biological lability. To date, there is no direct
23 molecular support for such an assumption. This work proposes a methodological approach based on
24 coupling a temperature-programmed pyrolyser to a standard mass spectrometer (Py-MS). A set of
25 reference soil samples was analysed. The set was completed by dissolved OM, source rock and coal
26 samples in order to test the insights of this methodological approach to various OM types. Data
27 analysis was based on T_{peak} values defined at temperatures of maximum *m/z* fragment production
28 during the temperature ramp. These T_{peak} values were grouped into clusters of distinct *m/z* values
29 for which a chemical attribution is proposed owing to an extensive literature survey. For the soil
30 samples, all the pyrolytic fractions described in the literature from mathematical deconvolution of
31 the S2 pyrogram were identified except for the most thermolabile fraction (280-320 °C). Fragments
32 with T_{peak} values between 330 and 390 °C were mainly attributed to proteins, lignin and
33 carbohydrates. T_{peak} values found in the 391-499 °C range corresponded to *m/z* fragments
34 attributed to aliphatics, lignin and aromatics while fragments with T_{peak} values higher than 480 °C
35 were only assigned to aromatics. In all samples, these fragments displayed a singular pattern of
36 decreasing *m/z* with increasing temperature, the significance of which remains to be fully elucidated.
37 This preliminary study provides key methodological guidelines for re-exploring Py-MS applications to
38 disentangle the chemical nature of OM.

39

40 **Keywords:** Rock-Eval[®] pyrolysis, S2 deconvolution, organic matter, soil, environmental matrices, Py-
41 MS

42

43 1. Introduction

44 Rock-Eval[®] pyrolysis is a common technique providing information on organic matter (OM)
45 quality and quantity in various environmental and geological samples [1-5]. During the last decades
46 there has been continuous efforts to generate more information than the classically used parameters
47 (Tmax, Total Organic Carbon, Hydrogen and Oxygen Index...). These efforts were mainly focused on
48 the multi-lobed S2 pyrolysis curves observed during soil, sediment and rock samples analyses, on the
49 argument that Tmax values of natural samples can be divided into several main temperature
50 fractions. Morphometric analyses and mathematical deconvolution of S2 curves into simple
51 Gaussians (considering the shape of the S2 curve of purified products; Sebag et al., 2006) were
52 proposed to define OM types within which a coherent thermal stability and biogeochemistry is
53 assumed [6-9]. Sebag et al. [7] proposed four elementary Gaussian components or pyrolytic families
54 (named F1 to F4) to resolve multi-lobed S2 curves obtained from soil samples. F1 (mean Tpeak = 307
55 °C) corresponds to labile bio-macromolecules found abundantly in fresh plants and soil litter. F2
56 (mean Tpeak = 371 °C) corresponds to resistant biopolymers such as lignin and cellulose. F3 (mean
57 Tpeak = 436 °C) corresponds to immature geo-macromolecules also (somewhat debatably) called
58 "humic substances". Finally, F4 (mean Tpeak = 481 °C) corresponds to "mature geo-
59 macromolecules". This approach has been applied to a large set of samples and scientific questions,
60 with some distinctions on the range and significance of the pyrolytic families [10-16]. An alternative
61 to mathematical deconvolution of the S2 peak consists in integrating the S2 peak area over fixed
62 temperature ranges [17-20].

63 These methods, and the numerous indexes built upon the pyrolytic families (R330 [21]; R400
64 [22]; S2I₃₅₀ and S2I₃₈₀ [17]; I-index [7, 18, 20]; and R- index [20]) were then applied by several authors
65 to assess the proportions of thermolabile and thermo-recalcitrant OM in soils [8, 11, 13, 17, 23, 24],
66 peats [25], riverine sediments [8], compost associated with pure minerals [26] and OM in aquifer
67 sediments [27]. Comparison of the results and indexes obtained by mathematical deconvolution (or

68 integration between temperature ranges) of S2 peaks with other techniques (Fourier Transform
69 Infrared Spectroscopy [21], ¹³C NMR [19, 26]) provided indirect molecular evidence by improving our
70 knowledge on the functional groups associated with the defined pyrolytic families. Similarly,
71 statistical analyses on large sets of samples were used to better discriminate the factors controlling
72 the distribution of the pyrolytic families [20].

73 Despite these laudable efforts, the molecular rationale for mathematically deconvoluting S2
74 pyrolytic curves obtained on soil samples into four to five Gaussian curves of homogeneous thermal
75 and chemical stability remains to be proven. As stated by Derenne and Quen ea [28], there is
76 currently no molecular information directly comparable to the Rock-Eval[®] pyrolysis signals from soils.
77 This prevents us from filling the gap between bulk and molecular analyses and thus achieving the
78 complete molecular elucidation of natural soil OM composition [29].

79 For instance, there is a large consensus that the first, thermolabile fraction (F1) comprises
80 carbohydrates whereas the late, thermo-recalcitrant fraction (F4) is mostly aromatic [25].
81 Conversely, the definition of the F2 and F3 fractions appears more elusive, with mentions of cellulose
82 and lignin, but also the so-called "humic substances". In addition, the number of Gaussian curves
83 required to satisfactorily explain the S2 signal can vary significantly between studies (from 4 to 11) in
84 agreement with the high variability of Tmax values recorded for soil samples, and the boundaries of
85 each defined fraction can vary amongst authors. Nevertheless, some common features are found
86 throughout these studies. There is a clear distinction between the OM that is responsible for the
87 signal below 330 °C ("thermally labile biological compounds" [6] or "little resistant bio-
88 macromolecules" [7]), between 330 and 400 °C ("cellulose, lignin and polypeptides" [6] or "resistant
89 biological constituents" [7]), and OM that is responsible for the signal above 400 °C ("immature
90 humic substances" [6] or "immature non-biotic constituents"/"immature geo-macromolecules" [7]).
91 Higher temperature fractions are defined either as "mature/reworked OM and charcoal" [6], or as
92 "mature refractory fraction" [7]. This distinction into several fractions of distinct thermal stability is

93 the foundation for the definition and use of different ratios that reflect the proportion of each
94 fraction to the S2 signal, as discussed above.

95 As pointed out by Sebag et al. [7], four or five Gaussian components cannot account for the
96 large diversity of biological constituents and their derivatives in soils and geological samples, and
97 there is currently no consensus on the adequate number of Gaussian curves that are necessary to
98 explain the S2 signal. In petroleum studies, the molecular composition of the S1 and S2 peaks has
99 long been addressed by means of offline, preparative, or online pyrolysis-GC-MS with cryo-trapping
100 [30, 31]. Cadeau et al. [32] recently proposed a new system that recovers hydrocarbons released
101 during Rock-Eval® analysis of standards, oils and source rocks to determine the molecular
102 composition of the Sh0, Sh1 and Sh2 peaks, those corresponding to free and absorbed hydrocarbons
103 in liquid-rich rock samples [33, 34, 35]. These approaches provided molecular clues about the
104 chemical composition of rather simple S1 and S2 peaks for oil shales and other petroleum-related
105 samples. They are not appropriate, however, to reveal the more complex chemical nature of soil
106 samples.

107 In the 90s, in-source pyrolysis mass spectrometry (Py-MS) and direct temperature-resolved
108 Field Ionization Mass spectrometry (FIMS) or Thermogravimetry/mass spectrometry were developed
109 for the molecular analysis of soil, dissolved and particulate marine and estuarine OM, dissolved
110 aquatic humic substances, plants [36-49] and sewage sludge and biowaste [50, 51]. These devices
111 make it possible to determine the chemical nature of materials by monitoring the evolution of
112 fragment emitted by pyrolysis under increasing temperatures directly in the source of the mass
113 spectrometer. An alternative to pyrolysis upstream the mass spectrometer is thermogravimetry. For
114 example, Lopez Capel et al. [52, 53] determined by this means three pools of soil OM: a thermally
115 labile fraction between 300 and 350 °C; a recalcitrant fraction between 400 and 500 °C and a
116 refractory fraction between 430 and 530 °C.

117 The molecular significance of Rock-Eval® S2 curves is here explored by applying a rather simple,
118 rapid, and cost-effective method inspired by the above-mentioned work to a set of standard soil and
119 geological samples. For this purpose, a temperature programmable pyrolyser was connected to a
120 standard single quadrupole mass spectrometer. Although this approach involves stronger
121 fragmentation and precludes the screening for molecular ions - unlike field ionization mass
122 spectrometry (FIMS) or in-source Py-MS - this allow to obtain molecular information in conditions
123 comparable to that of RE6 pyrolysis. In a previous work [54], we compared Rock-Eval® and Py-MS
124 results acquired on this set of samples by focusing on Total Ion Current (TIC) traces of Py-MS. This
125 allowed us evidencing time lags and temperature corrections, quantitative relationships and
126 differences in response between the two methods. Here, we discuss the evolution of the abundance
127 of m/z fragments during temperature-programmed pyrolysis to S2 curves obtained by Rock-Eval®
128 pyrolysis, which is discussed in terms molecular significance of the mathematical deconvolution of S2
129 peaks.

130

131 **2. Materials and methods**

132 2.1. Samples

133 The set of samples was presented in Jacob et al. [54]. It is constituted by a series of international
134 standards of soils ("Elliott", "Vertisol" and "Chernozem"), peat soil ("Pahokee"), source rocks
135 ("160000" and "Green"), coal ("Pocahontas") and dissolved OM ("Suwannee") that display significant
136 differences in RE6 parameters (TOC, HI, OI, Tmax) and in the shape of the S2 signals due to distinct
137 sources, diagenesis and maturity (see references in [54]).

138 2.2. Rock-Eval® pyrolysis and pyrolysis-mass spectrometry

139 Both methods were extensively described in Jacob et al. [54]. Briefly, for Rock-Eval® pyrolysis,
140 samples were analysed using "RecentSM" program: isothermal at 200 °C during 5 min, then an

141 increase to 650 °C at 30 °C/min and a final cessation of heating during 3 min down to ~540°C.
142 Samples were analysed at least twice. In the following, we will use TpS2 values as the temperature
143 effectively determined at the maximum of the S2 peak. Tpeak values stand for the temperature
144 determined either at the maximum of the Gaussian curves resulting from mathematical
145 deconvolution or at the maximum of the m/z curves acquired during programmed pyrolysis. TpS2
146 and Tpeak values are hence directly comparable.

147 Py-MS was realised with a Pyroprobe 5150 (CDS Analytical Inc.) pyrolyser directly coupled to a
148 single quadrupole ISQ7000 mass spectrometer (ThermoScientific), the Trace1310 GC gas
149 chromatograph (ThermoScientific) being only used as a bypass (deactivated capillary column, id 0.1,
150 60 cm length, oven maintained at 300°C mm). The temperature program set for the coil was similar
151 to that of Rock-Eval® pyrolysis. Samples were analysed at least three times.

152

153 2.3. Data management

154 In order to evaluate the extent to which there were clusters of m/z fragments issued from
155 organic compounds released at specific temperatures, we determined the Tpeak values of each m/z
156 Py-MS curve for several samples and replicates, at each nominal mass between m/z 50 and 150

157 Py-MS curves were extracted from Xcalibur .raw files after having converted .raw files into .cdf
158 files using the "Thermo File Converter" software. An in-house Python notebook was developed for
159 data management and data visualization. The script accomplishes two operations: first, extracting
160 and reorganizing 3d (m/z , scan/time and intensity) data from the converted .cdf file; then, smoothing
161 the curve (over 100 scans) and visualizing data as the evolution of intensity for selected m/z along
162 scans. Scan numbers were converted into time of analysis and then to temperature of pyrolysis by
163 using the temperature program equation over the temperature ramp (200 to 650 °C at 30 °C/min).

164 The package "peakutils" was then used to detect the temperature of maximum abundance of
165 each m/z (Tpeak), for each sample. The list of Tpeak values together with their corresponding m/z
166 was then exported to a .csv file and compiled for all samples and replicates.

167 The Python notebook can be tested with Binder on selected result files acquired on standard
168 samples at the following address: <https://github.com/PBrockmann/Pyrograms>.

169 The result files can also be downloaded here:

170 https://thredds-su.ipsl.fr/thredds/catalog/ipsl_thredds/brocksce/Pyrograms/catalog.html

171 For data management reasons, we only considered Tpeak values of fragments in the m/z 50-150
172 range. These fragments constitute by far the largest proportion of the TIC signal (estimated from
173 absolute abundances in the total spectrum, integrated over the whole TIC peak) in Green ($91.4 \pm 1\%$),
174 Elliott ($93.7 \pm 0.7\%$), Vertisol ($85.2 \pm 4.8\%$), Chernozem ($91.8 \pm 1.5\%$), Suwannee ($90.1 \pm 1.2\%$),
175 Pahokee ($92.9 \pm 1.9\%$) and Pocahontas ($78.9 \pm 7.2\%$) samples. These values are higher than those
176 reported, for example, for Time-Resolved Pyrolysis-Field Ionization Mass Spectrometry [1, 2] due to
177 stronger ionization and fragmentation with the device used. Tpeaks were individually verified for all
178 samples and exported to an Excel sheet. This operation led to at least 200 Tpeak values (observable;
179 obs) per sample over at least two analytical replicates (n) for every standard sample (Elliott: n=4,
180 obs=443; Vertisol: n=3, obs=278; Chernozem: n=3, obs=251; Pahokee: n=3, obs=365; Suwannee: n=3,
181 obs=309; Pocahontas: n=3, obs=318; Green: n=2, obs=200; 160000: n=2, obs=202). Finally, Tpeak
182 values were converted into temperatures equivalent to TpS2 values as measured by Rock-Eval®
183 ($T_{peak_{conv}} = (T_{peak\ Py-MS} - 114.69)/0.8134$; see Jacob et al. [54]). The inventory of Tpeak values in
184 the m/z 50-150 range was not exhaustive since some Tpeaks of double peaks or shoulders could not
185 be automatically determined. Although this work refers to the mass spectrometric literature for
186 molecular attributions of m/z values (cf. section 4.7.), m/z attribution to a specific class of
187 compounds is not unequivocal. In addition to multiple peaks and shoulders in some m/z curves that

188 indicate a distinct thermal behaviour of the compounds they derive from, the same m/z fragments
189 can be produced from compound families that are chemically distinct.

190 Clustering of $T_{\text{peak}_{\text{conv}}}$ values was done with the k-means classification function in the XLSTAT®
191 software. Although 3 to 5 clusters, depending on the sample, were defined with hierarchical
192 clustering (realized with the HCA function in XLSTAT®), it was decided to set the k-means
193 classification with 4 clusters for homogeneity, and in order to be in line with the number of Gaussian
194 curves into which S2 pyrolysis curves are deconvoluted in most cases.

195

196 **3. Results**

197 S2 and TIC signals have been compared previously [54]. Here we focus on the analysis of m/z
198 curves and compare the distribution of their $T_{\text{peak}_{\text{conv}}}$ values to literature data on mathematical
199 deconvolution of S2 signals.

200 3.1. Total mass spectra

201 Total mass spectra illustrate the relative abundance of m/z fragments within the time range of
202 the TIC peak (Fig. 1 for Elliott and Green and Fig. S1 for all samples). Some common features were
203 observed in the total mass spectra of all standard samples: a decreasing trend of abundance towards
204 higher m/z and strong abundance of the m/z 91 fragment in nearly all samples, although with
205 different relative abundances.

206 In Green and 160000, m/z 91 emerges only slightly from a decreasing trend of m/z relative
207 abundance with increasing m/z . In Pahokee, Pocahontas and Suwannee, m/z 91 strongly emerges or
208 dominates. In addition to m/z 91, m/z 94 is also abundant in Pahokee and Suwannee samples (Fig.
209 S1). In soil samples Chernozem, Elliott and Vertisol, the relative abundance of m/z 91 compared to
210 other fragments is similar. In addition, clusters of m/z can be identified in all mass spectra at m/z 51-

211 58, m/z 63-71, m/z 76-85, m/z 89-97 and 102-113. Clusters of heavier fragments (m/z 115-123, 127-
212 135, 141-147) are less evident. Fragments larger than m/z 200 are barely discernible.

213

214 3.2. Evolution of selected m/z with time/temperature

215 A convenient manner of depicting the distinct evolution of m/z fragment intensity with
216 time/temperature consists in plotting the intensity of selected fragments (here m/z 57, 78, 91, 92, 94
217 and 107) with temperature. In Fig. 2 (for Elliott and Green samples, in Fig. S2 for all samples), the
218 intensity of m/z fragments is plotted against temperature by converting Py-MS acquisition time to
219 temperature, using the temperature program, and then converting to $T_{\text{peak}_{\text{conv}}}$ for comparison with
220 T_{pS2} (see Methods). Only data for the 5-20 min time range, corresponding to the linear increase in
221 temperature with time, are shown, excluding data acquired during the isothermal steps. From the
222 selected m/z fragments of which the intensity is plotted over temperature, it becomes evident that
223 fragments are recorded over distinct time ranges, with differing maximum release temperatures. This
224 illustrates the sequential release and cracking of compounds during programmed pyrolysis.

225 The curve of Py-MS TIC evolution with temperature of Green is a simple Gaussian with few
226 variations in the $T_{\text{peak}_{\text{conv}}}$ values. This is in agreement with the observation by Sebag et al. [7] that
227 the S2 pyrolysis curves of source rocks, in contrast to soils and recent sediments, only require a single
228 Gaussian curve to be mathematically deconvoluted. This is indeed true for those with limited organic
229 sources such as Green that is Type I OM mainly composed of algal OM. The Py-MS TIC curve is more
230 complex for the Elliott sample. It is built upon m/z curves whose T_{peaks} are distinct. In addition,
231 some of the m/z Py-MS curves display shoulders or are bimodal, for example, the curves of m/z 57
232 and 78 in Elliott (Fig. 2). This suggests that (1) distinct compounds affording the same fragments are
233 released at distinct times or that (2) the same compounds (thus displaying the same fragments) are
234 released at two distinct times and/or have distinct thermal behaviours. From a quantitative point of
235 view, the main features identified from total mass spectra (Fig. 1) are also revealed from m/z

236 evolution with temperature. Upon the selected fragments displayed in Fig. 2, m/z 57 is dominant in
237 Green whereas m/z 91 is the most abundant fragment in Elliott, associated with abundant m/z 94
238 and 107.

239 3.3. Clusters of $T_{\text{peak}_{\text{conv}}}$ values

240 $T_{\text{peak}_{\text{conv}}}$ values determined in the m/z 50 to 150 range for replicate analyses of standard
241 samples were grouped into four clusters whose characteristics are shown in Table 1. In soil samples,
242 C1 was centered around 365 °C, C2 around 430 °C, C3 around 460 °C and C4 around 490 °C, with
243 significant intra-cluster variance. In geological samples of higher thermal maturity, the clusters were
244 shifted towards higher temperature values, specifically for Pocahontas. Note that the clusters
245 defined for the Green and 160000 samples were defined on a narrower range of temperatures and
246 displayed lower variance.

247

248 4. Discussion

249 4.1. $T_{\text{peak}_{\text{conv}}}$ distribution in soil samples

250 "Bulk" quantitative differences in the eight standard samples were in Jacob et al. [54]. Major
251 distinctions in m/z abundance are also noted in the total mass spectra of samples (Fig. 1 and Fig. S1).
252 This denotes disparities in the molecular composition of the studied OM samples in accordance with
253 (1) their origin and (2) the degree of degradation or thermal alteration they were subjected to. This is
254 also highlighted by the Py-MS curves of selected m/z fragments (Fig. 2) illustrating various T_{peaks} for
255 distinct m/z . Although Fig. 2 does not provide a full picture of $T_{\text{peak}_{\text{conv}}}$ variability within or across
256 samples, it suggests that the chemical composition of OM that emits those m/z fragments is distinct
257 at distinct temperatures. This is in agreement with the mathematical deconvolution of RE6 S2 curves
258 that makes the assumption of the sequential cracking of chemically and thermally distinct OM during
259 pyrolysis. These types of OM are classically grouped into four to five families, each characterized by
260 rather homogeneous temperatures of maximum release and thought to be chemically or physically
261 homogeneous. In order to test the hypothesis that there are such pyrolytic families of homogeneous
262 temperatures of maximum release, the $T_{\text{peak}_{\text{conv}}}$ values determined for each m/z fragment in the 50-
263 150 range were used as a key parameter. The distribution of $T_{\text{peak}_{\text{conv}}}$ values against m/z fragments
264 is displayed for four replicates the Elliott sample in Fig. 3.

265 For the Elliott sample, $T_{\text{peak}_{\text{conv}}}$ values of individual m/z curves range from 332 to 581 °C. There
266 is a rather good reproducibility of $T_{\text{peak}_{\text{conv}}}$ values between the four replicates of the Elliott sample,
267 with a standard deviation from 0 to 13 °C (average 2.7 °C on 443 $T_{\text{peak}_{\text{conv}}}$ values over the four
268 replicates). On examining the distribution of $T_{\text{peak}_{\text{conv}}}$ values by increasing temperatures (Fig. 3a),
269 several slope breaks can be distinguished. This suggests that $T_{\text{peak}_{\text{conv}}}$ values are not randomly
270 distributed and that several groups can be distinguished. Distribution of the $T_{\text{peak}_{\text{conv}}}$ values against
271 m/z and individualised by replicate analyses (Fig. 3b) confirms that the $T_{\text{peak}_{\text{conv}}}$ values are not
272 randomly distributed. Four clusters can be distinguished (Fig. 3c; Table 1 and Table 2). Within the

273 third and fourth groups, a singular pattern can be observed with several series of decreasing
274 $T_{peak_{conv}}$ values with increasing m/z , a pattern that is more evident between m/z 114 and 124, m/z
275 128 and 134 and m/z 142 and 148 (Fig. 3b).

276 The distribution of $T_{peak_{conv}}$ values (averaged over replicates) against m/z in Chernozem,
277 Vertisol and Pahokee is similar to that observed for Elliott (Fig. 4). As observed for Elliott, and with
278 few exceptions, the $T_{peak_{conv}}$ values determined for each m/z are consistent within replicates (mean
279 standard deviation=2.4 °C for Elliott, 120 observables, n=4; 4.6 °C for Vertisol, 98 observables, n=3;
280 3.6 °C for Chernozem, 96 observables, n=3; 3.5 °C for Pahokee, 124 observables, n=3). In addition,
281 the $T_{peak_{conv}}$ values determined for each m/z fragment are relatively consistent across samples,
282 except for Pahokee, which displays, on average, higher $T_{peak_{conv}}$ values for the same m/z fragments
283 in the 440-520 °C range (clusters 3 and 4). Despite these rather good agreements of $T_{peak_{conv}}$ values,
284 some discrepancies can be evidenced between samples. In cluster 1 (340-400 °C), as an example,
285 more m/z fragments were found in Elliott and Pahokee than in Vertisol and Chernozem. This denotes
286 chemical differences between soil samples that are additional to differences in relative abundance of
287 m/z fragments, not discussed here. Some minor differences were observed in $T_{peak_{conv}}$ values in the
288 440-520 °C temperature range (clusters 3 and 4) between soil samples. In most cases, and for a given
289 m/z fragment, $T_{peak_{conv}}$ values are higher for Pahokee, then Chernozem, then Vertisol and Elliott.
290 Although this must be confirmed by further studies, these differences may illustrate mineral matrix
291 effects [55] on the cracking temperature of these polymers. For instance, Dembicki [56] showed that,
292 due to catalytic effects, clay minerals decrease activation energies to lower values. This is likely
293 explaining why $T_{peak_{conv}}$ values are lower in Elliott (silt loamy soil) than in Vertisol and Chernozem
294 (sandy clay), and in Pahokee (peaty soil). Finally, the pattern of decreasing $T_{peak_{conv}}$ values with
295 increasing m/z in the 450-520 °C range for several series of m/z , already observed for Elliott, is also
296 noted for Chernozem, Vertisol and Pahokee.

297

298 4.2. Comparison of S2 and of Py-MS TIC deconvolution and $T_{peak_{conv}}$ clustering for soils

299 Left panel of Figure 5 depicts the clusters of $T_{peak_{conv}}$ values and T_{peak} and $T_{peak_{conv}}$ values of
300 Gaussian curves obtained after mathematical deconvolution of the S2 and TIC peaks determined on
301 standard soil samples (from [54]). These results are compared to the temperature ranges of fractions
302 defined in the literature (central panel, Fig. 5), and to the same parameters as in the left panel but
303 for non-soil samples (right panel, data from [54]). The parameters of each cluster are given in Table
304 1. Although the deconvolution of the S2 pyrolysis curve resulted in Gaussian curves centred around
305 316 °C in Pahokee and 321 °C in Elliott (black hyphens, Fig. 5), no fragment was found for which a
306 $T_{peak_{conv}}$ value was determined in the 290-330 °C range that corresponds to the F1 [7], F2 [8] or A1
307 [19, 20] pyrolytic fractions in the samples. This range corresponds to the pyrolytic family defined as
308 "thermally labile biological compounds" [6-8, 19, 20]. Because these compounds are essentially
309 found in soil litters, they are not expected to represent a large fraction of the signal in the standard
310 soil samples since these were taken from deeper horizons and can thus be devoid of fresh material.
311 Nonetheless, this is in contradiction with the appearance of a Gaussian curve at low temperatures
312 from the deconvolution of the S2 pyrolysis curves [54]. An alternative explanation would imply that
313 compounds released at this temperature are light/small compounds with minimal production of
314 fragments higher than m/z 50, as discussed in [54].

315 The first (by order of increasing temperatures) series of clusters of $T_{peak_{conv}}$ values that could be
316 defined falls within the 330-390 °C range (332-385 °C for Elliott, 329-395 °C for Chernozem, 343-393
317 °C for Vertisol and 332-387 °C for Pahokee; Table 1 and Fig. 5). These ranges are fairly homogeneous
318 and in agreement with ranges defined for F2 [7], F3 [8], A2 [19, 20], and the higher temperature
319 range of the S2a [18] pyrolytic fractions that are attributed to cellulose, lignin and, tentatively,
320 polypeptides by Disnar et al. [6]. $T_{peak_{conv}}$ values of the second Gaussian obtained by deconvolution
321 of the TIC peak fall within the temperature ranges of this cluster (grey hyphens in Fig. 5), but not

322 Tpeak values defined on the second Gaussian curve from deconvolution of the S2 peak, which are
323 always higher.

324 The second series of clusters of Tpeak_{conv} values ranges between 391 and 455 °C (391-433 °C for
325 Elliott, 416-455 °C for Chernozem, 398-449 °C for Vertisol, and 396-448 °C for Pahokee (Table 1 and
326 Fig. 5). These ranges fall within the range of temperatures proposed in the literature for the third
327 pyrolytic fraction (defined as F3, F4, A3, and the lower temperature range of S2b, depending on the
328 authors; Fig. 5) considered to consist, for soils, of "immature humic substances" [6].

329 The third series of clusters range from 434 to 499 °C (434-464 °C for Elliott, 456-486 °C for
330 Chernozem, 450-486 °C for Vertisol and 451-499 °C for Pahokee, Table 1 and Fig. 5). They thus
331 overlap by ~20 °C with the range of the second series of clusters. These ranges are consistent with
332 the F5 pyrolytic fraction of Copard et al. [8], with the lower values of the F4 and A4 fractions, and
333 with the upper range of the S2b fraction (Fig. 5). These pyrolytic fractions are defined in the
334 literature as characteristic of "mature" or "reworked" OM. Within this series of clusters, there is a
335 better agreement with Tpeak_{conv} values defined on the third Gaussian resulting from deconvolution
336 of the S2 peak than for the other series of clusters.

337 The fourth series of Tpeak_{conv} clusters ranges between 465 and 588 °C (465-581 °C for Elliott,
338 487-588 °C for Chernozem, 487-588 °C for Vertisol and 500-581 °C for Pahokee, Table 1 and Fig. 5).
339 Again, this series of clusters overlaps with the previous one by ~30 °C. It could correspond to the F6
340 fraction of Copard et al. [8] considered as reflective of recalcitrant, fossil OM, but also to the S2c of
341 Gillespie et al. [18] and to the upper temperature range of the A4 fraction of Albrecht et al. [19] and
342 Sebag et al. [20].

343 Although both S2 mathematical deconvolution and Tpeak clustering resulted in four fractions,
344 these are thus not systematically comparable. The first fraction defined from S2 deconvolution (F1,
345 F2, A1 and lower range of S2a) could not be identified. Only the first and last Tpeak clusters are in
346 general good agreement with F2, [7], F3 [8], the upper part of S2a [18] and A2 [19, 20] on the one

347 hand and F6 [8], S2c [18] and the upper part of A4 [19, 20] on the other hand. The second and third
348 Tpeak clusters more or less correspond to F3 and F4 [7], F4 and F5 [8], A3 and the lower part of A4
349 [19, 10], and S2b [18], with some overlaps between Tpeak clusters. Thus, clusters 2 and 3 (and to a
350 lesser extent cluster 4) could be considered as a continuous series, calling into question a possible
351 artificial clustering of a homogeneous set of Tpeak_{conv} values.

352 Overall, the examination of Tpeak_{conv} values for soil samples appears to validate a cluster of
353 temperatures in the 330-390 °C range that could correspond to the second pyrolytic fraction
354 commonly reported in the literature. The same applies to the high temperature cluster. The picture is
355 more complex for intermediate Tpeak_{conv} values with no clear clustering and variable correspondence
356 with pyrolytic families defined from Rock-Eval® S2 deconvolution.

357 4.3. Comparison of S2 and of Py-MS TIC deconvolution and Tpeak_{conv} clustering for other samples

358 The series of clusters in Suwannee is not very distinct from those of the soils, except that the four
359 clusters appear to be systematically set at lower temperatures than their equivalents in soils.
360 Conversely, the series of clusters in Pocahontas is distinct, with no Tpeak_{conv} value below 399 C°
361 (except for *m/z* 64 at 352-360 °C not presented, cf. below) and maximum temperatures recorded at
362 631 °C, due obviously to the higher thermal maturity recorded by this sample. Except for a cluster
363 that was defined between 349 and 394 C° in Green (but with only 4 Tpeak_{conv} values), Green and
364 160000 samples are characterized by concentrated clusters in a narrower range of temperatures
365 (444-495 °C for Green and 416-503 °C for 160000) than those observed for soils. For non-soil samples
366 such as Pocahontas, Green and 160000 there is hence no or little correspondence with the
367 temperature ranges defined in the literature for soil samples.

368 4.4. Chemical attribution of *m/z* Tpeak_{conv} values

369 Graphical and statistical treatments only provided a partially satisfactory distinction of classes of
370 fragments depending on the temperature at which their maximum was released. As a matter of fact,

371 some $T_{\text{peak}_{\text{conv}}}$ values of replicate analyses of one standard sample were found in two distinct
372 clusters, which shows the limit of clustering. In order to better understand the molecular significance
373 of pyrolytic fractions, fragments were organised according to both their chemical attribution and
374 their $T_{\text{peak}_{\text{conv}}}$ values. m/z fragments were tentatively attributed to specific chemical families
375 following an extensive literature survey (see Supplementary Table S1 and related references).

376 Although the attribution of m/z fragments to specific chemical families is non-unique and thus
377 debatable, it provided a first step to assess the nature of the compounds that are progressively
378 produced during the course of pyrolysis. m/z fragments were then tentatively attributed to distinct
379 chemical families for the Elliott sample. The result of this attribution is displayed in the m/z 50-150
380 range in Table 3. When two $T_{\text{peak}_{\text{conv}}}$ values were defined for a single m/z fragment, the m/z
381 fragment was attributed to the most probable chemical family, depending on the $T_{\text{peak}_{\text{conv}}}$ value and
382 chemical attribution of neighboring (in terms of temperature) m/z fragments. Sulfur was identified in
383 all samples through the m/z 64 fragment.

384 Five chemical families were distinguished. "Aromatics" are composed of m/z fragments identified
385 in the literature as originating from benzene, "alkylaromatics", "alkylbenzenes", or "naphthalenes".
386 m/z fragments interpreted in the literature as representative of "aliphatic polymers",
387 "hydrocarbons", "alkane fragments", "low-mass olefins" or "lipids" were grouped into "Aliphatics".
388 The "Carbohydrates" family comprises fragments attributed to "carbohydrates with pentose and
389 hexose subunits", "mono- and polysaccharides" or "cellulose". The "Lignin" family is composed of
390 fragments attributed to "phenols and lignin monomers", "dihydroxybenzenes", "alkyl-substituted
391 dihydroxybenzenes" and "alkylphenols". Finally, fragments relative to "proteins", "N-compounds or
392 N-containing compounds", "peptides", "amino-acid species and amides", "alkyl pyrroles and indoles"
393 were grouped in the "Proteins" chemical family.

394 Despite being produced with different temperature programs, the result of m/z attribution is in
395 general agreement with results acquired on soils by Py-FIMS [44, 57, 58] and with results from the

396 Rock-Eval[®] analysis of pure organic materials [10, 22]. In other words, the $T_{\text{peak}_{\text{conv}}}$ values found for
397 chemical families are in line with the temperatures of maximum release defined in the literature.

398 Fig. 6 reports the ranges of $T_{\text{peak}_{\text{conv}}}$ values for the 5 chemical families of m/z fragments in the
399 50-150 range, in the Elliott sample. As already discussed, and considering the nature of the samples,
400 the most labile fraction described in the literature between 280 and 320 °C could not be identified by
401 significant $T_{\text{peak}_{\text{conv}}}$ values. The cluster of $T_{\text{peak}_{\text{conv}}}$ values that occurs between 340 and 400 °C is
402 constituted by m/z fragments attributed to Sulphur, Carbohydrates, Phenols and Proteins (Fig. 6).
403 This result is in line with the chemical attribution of the fraction of soil OM that cracks between 330
404 and 400 °C to "cellulose, lignin and polypeptides" [6] or "resistant biological constituents" [7]. A
405 meticulous examination of differences in T_{peak} occurrences between soil samples in this 340-400 °C
406 cluster could give valuable information on the chemical differences in thermolabile OM between soil
407 samples, although quantitative data could be more efficient. For example, and contrary to Elliott, a
408 $T_{\text{peak}_{\text{conv}}}$ value could not be defined in Chernozem for m/z 98 (Carbohydrates), m/z 61 and 124
409 (Phenols) and m/z 74 (Protein).

410 Most fragments of the 415 to 480 °C $T_{\text{peak}_{\text{conv}}}$ cluster relate to Aliphatics, Lignin and Aromatics,
411 and to a lesser extent to Carbohydrates (m/z 50, 82, 114, 144, 146), although some of these
412 fragments appear isolated in the 415-440 °C range. Due to the difficulties in defining distinct T_{peak}
413 clusters in this temperature range (cf. 4.2.), it is considered that all fragments in the 415-480 °C range
414 represented a homogeneous, although chemically complex, polymeric OM.

415

416 4.5. A singular pattern in the 415-480°C range

417 The distribution of m/z fragments against $T_{\text{peak}_{\text{conv}}}$ values in the 415-480°C range is
418 characterized by the already mentioned singular pattern with $T_{\text{peak}_{\text{conv}}}$ values decreasing with
419 increasing m/z (Fig. 3, 4 and 7). This pattern is more evident in the m/z 110-150 range. For lower m/z ,

420 abundant fragments attributed to Lignin might mask this pattern. Three series of m/z ($\sim 115-123$,
421 $\sim 128-135$ and $\sim 140-149$), separated by ~ 14 units, are each characterized by a ~ 50 °C drop in
422 temperature (Fig. 7) in most samples, except for Green and 160000, for which the difference is lower.
423 This pattern was identified in all soil samples but also in Suwannee, Green and 160000. In Pocahontas
424 this pattern was also observed but occurs at higher temperatures (500-560 °C). This singular and
425 similar pattern regardless of the samples, but recorded at distinct $T_{\text{peak}_{\text{conv}}}$ values, suggests that the
426 considered m/z fragments are representative of a coherent pool of OM mostly composed, for soils,
427 of Aliphatics and Lignin, with some Aromatics and Carbohydrates.

428 The considered pattern suggests a continuum of cracking in this cluster. This could reflect the
429 progressive cracking of components of the polymers based on thermochemical considerations
430 (degree of unsaturation, aromaticity, double bond location on the alkyl chains, branching...). For
431 example, as the unsaturation increases (leading to a series of m/z with 2 mass unit differences),
432 lower activation energies may be involved [60]. The lower temperature range recorded for Green
433 and 160000 samples for this cluster could be related to their more aliphatic nature compared to
434 other samples, as already illustrated by the low relative abundance of m/z 91 in the total mass
435 spectra of Green and 160000 (Fig. 1). For soil samples, these results thus provide new insight into this
436 yet poorly resolved or understood RE pyrolytic fraction considered in the literature as "immature
437 humic substances" [6] or "immature non-biotic constituents"/"immature geo-macromolecules" [7].

438 Surprisingly, the pattern is also valid for the Pocahontas sample, but with $T_{\text{peak}_{\text{conv}}}$ values of the
439 corresponding m/z fragments are about ~ 70 °C higher than for soils (Fig. 7). Mineral matrix effects
440 cannot be invoked because Pocahontas is devoid of a significant mineral fraction (as Suwannee). The
441 higher $T_{\text{peak}_{\text{conv}}}$ values recorded for Pocahontas is likely due to the higher degree of
442 reticulation/condensation of the polymers and/or higher aromaticity since Pocahontas is mainly
443 composed of a network of alkylbenzenes and alkylnaphthalenes [60]. This probably explains the
444 scarcity of $T_{\text{peak}_{\text{conv}}}$ values lower than 500 °C for this sample: (1) m/z 64 $T_{\text{peak}_{\text{conv}}}$ value was found at

445 356 °C and attributed to Sulphur; (2) m/z 83, 97, 98, 99, 111, 112, 113 and 125 $T_{peak_{conv}}$ values were
446 found in the 417-470 °C range. These were attributed to Aliphatics in soil samples, as well as in
447 Pocahontas. All other $T_{peak_{conv}}$ values defined for Pocahontas (500-600 °C) encompass the range
448 defined for the fourth cluster of $T_{peak_{conv}}$ values (470-590 °C) defined in soil samples. Very consistent
449 $T_{peak_{conv}}$ values were found for m/z 78, 128 and 138 in both Pocahontas (Fig. 7) and soil samples
450 (Figs. 3, 4 and 7). Thus, these fragments and their associated $T_{peak_{conv}}$ values could attest to a
451 contribution of fossil organic carbon in soil samples. Results are thus consistent with interpretations
452 of the high temperature fraction (~480 °C) defined from the mathematical deconvolution of the S2
453 pyrogram as "mature/reworked OM and charcoal" [6] or "mature refractory fraction" [7]. It appears
454 peculiarly relevant for Chernozem in which Schmidt et al. [61] noted a significant contribution of
455 charred organic carbon. The monitoring of such fragments in environmental samples should be
456 examined in order to determine the contribution of charcoal/fossil carbon/black carbon.

457

458 **Conclusion**

459 Temperature-resolved pyrolysis directly coupled to a mass spectrometer applied to various
460 environmental samples provides direct evidence for the molecular composition of pyrolytic fractions
461 defined from the mathematical deconvolution of Rock-Eval® S2 pyrograms. Only the most
462 thermolabile fraction (280-320 °C) described in the literature for soil samples was not detected.
463 Then, and according to their thermal lability, four main clusters of T_{peak} values were described. The
464 first cluster of T_{peak} values (340-400 °C) is formed by fragments assigned to proteins, lignin and
465 carbohydrates. The second and third clusters (415-480 °C) correspond to fragments attributed to
466 aliphatics, lignin and aromatics. In all samples, this cluster displays a singular pattern of decreasing
467 m/z with increasing temperature, the significance of which remains to be elucidated. The fourth
468 cluster (T_{peak} values > 480 °C) is constituted by fragments exclusively attributed to aromatic
469 moieties.

470 These results show that the mathematical deconvolution of S2 curves of soil samples obtained from
471 Rock-Eval® pyrolysis can be partly validated by direct molecular information, although the complex
472 415-480 °C cluster should probably not be considered as a simple Gaussian and requires further
473 attention. For other types of samples such as mature geological samples, the strategy must be
474 adapted, and the attribution of *m/z* fragments could differ. This approach opens new windows for
475 the elucidation of the molecular nature of complex organic matrices. Further work will require a
476 closer scrutiny of the significance of *m/z* achieved by using higher mass resolution and lower energy.
477 Extensive data mining of the results could afford more relevant variables than the Tpeak values
478 alone.

479

480 **Acknowledgments**

481 The authors wish to thank David Curry, Thomas Gentzis and Joseph Curiale for their constructive
482 comments on an earlier version of the manuscript. Rachel Boscardin at ISTO laboratory in Orléans is
483 acknowledged for Rock-Eval® analyses. This work received financial support from the OTARIE project
484 (Région Centre Val de Loire, European Council through the FEDER fund) and from the Labex
485 VOLTAIRE (ANR-10-LABX-100-01). The authors appreciated interactions with Benoit Auguin (Quad
486 Service, France) from the initiation to the finalization of the project. J. Jacob, F. Delarue and Y.
487 Copard are indebted to J.R. Disnar for raising their awareness of the potential of Rock-Eval® pyrolysis
488 and for long-standing inspiring discussions on this project. M.F. Dignac and V. Grossi are thanked for
489 their advice during data interpretation. E. Rowley-Jolivet is acknowledged for English language
490 editing of the manuscript.

491 **Tables**

492 Table 1: Barycentre, intra-cluster variance, minimum (Min) and maximum (Max) values and standard
 493 deviation (Stdev) for the 4 clusters of Tpeakconv values defined in the *m/z* 50-150 range of the eight
 494 standard samples.

Sample	Cluster	Barycentres (°C)	Intra-cluster variance	Min (°C)	Max (°C)	Stdev (°C)	Observables
Elliott	C1	354	101	332	385	8,3	71
	C2	418	51	391	433	5,5	78
	C3	449	66	434	464	7,6	194
	C4	480	630	465	581	16,7	100
Chernozem	C1	352	308	329	395	14,4	15
	C2	443	79	416	455	7,2	95
	C3	469	46	456	486	5,5	100
	C4	504	551	487	588	15,2	41
Vertisol	C1	361	128	343	393	8,7	24
	C2	433	139	398	449	10,3	94
	C3	466	74	450	486	7,3	131
	C4	508	767	487	588	21,2	29
Pahokey	C1	355	185	332	387	12,3	70
	C2	424	267	396	448	14,5	19
	C3	475	92	451	499	7,9	254
	C4	524	769	500	581	23,9	22
Suwanee	C1	360	133	334	379	9,7	47
	C2	400	80	381	419	7,5	74
	C3	439	91	422	455	8,6	74
	C4	471	181	455	514	10,6	114
Pocahontas	C1	430	224	399	458	13,1	23
	C2	514	92	492	535	7,9	121
	C3	555	92	535	572	8,4	113
	C4	589	159	572	631	10,7	61
Green	C1	372	484	349	394	19,8	4
	C2	454	26	444	462	4,3	26
	C3	469	8	462	475	2,3	118
	C4	482	23	476	495	4,2	52
160000	C1	429	43	416	438	5,9	36
	C2	448	20	440	457	4,0	102
	C3	467	33	459	476	5,5	54
	C4	492	67	482	503	7,4	10

495

496

497

498

499 Table 2: Affection of m/z fragments to distinct $T_{peak_{conv}}$ clusters in the Elliott sample. When two
 500 $T_{peak_{conv}}$ values were found for the same fragment, the $T_{peak_{conv}}$ values are indicated.

$T_{peak_{conv}}$ cluster	m/z fragment
340 - 400 °C	58, 59, 60, 61, 64, 68, 73, 74, 86, 87, 96, 98, 109, 110, 111, 112, 119, 121, 124, 125, 137, 138
400 – 440 °C	53, 54, 58, 61, 67, 72, 80, 81, 82, 87, 95, 96, 109, 110, 111, 114, 123, 124, 125, 133, 135, 136, 137, 138, 148, 149, 150
440 - 470 °C	50, 51, 55, 56, 57, 61, 62, 64, 65, 66, 67, 69, 70, 71, 77, 79, 83, 84, 85, 86, 91, 92, 93, 94, 97, 98, 99, 100, 104, 105, 106, 108, 111, 112, 113, 117, 118, 119, 120, 121, 122, 125, 126, 132, 133, 134, 137, 140, 145, 146, 147, 148
470 - 591 °C	63, 64, 74, 75, 76, 78 (470 °C), 78 (570 °C), 89, 90, 101, 102, 103, 107, 114, 115, 116, 127, 128 (475 °C), 128 (565 °C), 129, 130, 131, 139, 141, 142, 143, 144, 145, 146

501

502 Table 3: Attribution of m/z fragments to chemical families in the Elliott sample according to literature
 503 data (Table S1).

Chemical family	m/z
Aromatics	78, 92, 106, 120, 128, 134, 138, 142, 148
Aliphatics	55, 56, 57, 58, 65, 67, 70, 74, 76, 81, 83, 84, 87, 89, 91, 92, 93, 96, 97, 98, 99, 100, 101, 102, 103, 105, 106, 107, 108, 109, 111, 112, 113, 114, 115, 116, 117, 118, 119, 120, 122, 123, 124, 125, 126, 127, 128, 129, 130, 131, 132, 133, 134, 135, 148
Carbohydrates	50, 58, 60, 68, 82, 86, 96, 98, 110, 112, 114, 144, 146
Lignin	51, 54, 61, 69, 71, 72, 75, 77, 79, 80, 85, 86, 90, 94, 95, 110, 124, 136, 140, 150
Proteins	59, 73, 74, 87, 109, 111, 125, 137
Sulfur	64
Non-attributed	145, 53, 62, 63, 64, 66, 78, 104, 119, 121, 137, 138, 139, 141, 143, 147, 149

504

505

506 **Figure captions**

507 Figure 1: Total mass spectra of Elliott and Green standard samples analysed by Py-MS in the m/z 50-
508 300 range. Total mass spectra illustrate the relative abundance of m/z fragments recorded during the
509 time interval of the TIC peak. Background subtraction is performed before and after the TIC peak.

510 Figure 2: Evolution of TIC (red curve) and relative abundance of selected m/z (57, 78, 91, 92, 94 and
511 107) with temperature (converted to be comparable to TPs2) for Elliott and Green samples. The
512 same curves for the eight standard samples are displayed in Fig. S2. $T_{peak_{conv}}$ values are indicated for
513 both TIC and m/z curves. The amount of material used for each analysis (in mg) is indicated after the
514 sample name.

515 Figure 3: $T_{peak_{conv}}$ values determined for individual m/z values (obs=443) in the 50-150 range for four
516 replicate analyses of the Elliott standard sample (11.1, 14, 14.4 and 16.7 mg). (a) Ordered by
517 increasing temperature; (b) Ordered against m/z values and organized by replicate analysis. Arrows
518 represent a singular pattern of decreasing m/z with increasing $T_{peak_{conv}}$ values discussed later; (c)
519 Grouped after cluster analysis. Horizontal dotted line corresponds to the limits of defined clusters.

520 Figure 4: Comparison of $T_{peak_{conv}}$ values (averaged per sample) organized versus m/z values in the
521 50-150 range for soil samples. Elliot (green dots, $n=4$), Chernozem (red dots, $n=3$), Vertisol (circled in
522 black, $n=3$) and Pahokee (black dots, $n=3$). Horizontal dotted line corresponds to the limits of defined
523 clusters.

524 Figure 5: Comparison of cluster analysis of $T_{peak_{conv}}$ values for soil standard sample with ranges of
525 temperatures defined in the literature for T_{peak} values of Gaussian curves resulting from
526 mathematical deconvolution of the S2 pyrolytic curve (boxes and whisker plots; see Jacob et al. [54]).
527 In order to allow direct comparison of soil $T_{peak_{conv}}$ clusters with literature data, these are presented
528 next to each other. Clusters of $T_{peak_{conv}}$ values for non-soil samples (Suwannee, Pocahontas, Green
529 and 160000) are presented in the four panels on the right. Parameters of the $T_{peak_{conv}}$ clusters are

530 given for each standard sample in Table 1. T_{peak} values defined from the deconvolution of the S2
531 pyrolysis curve are indicated, for each sample, as black hyphens (horizontal bars), and T_{peak} values
532 defined on each Gaussian curve obtained from the mathematical deconvolution of the TIC peak are
533 indicated as grey and thick hyphens (from Jacob et al. [54]). Note that m/z 64 was isolated before
534 clustering in the Pocahontas sample, and appears as a single point.

535 Figure 6: $T_{peak_{conv}}$ values for each m/z in the 50-150 range, organized according to their chemical
536 attribution. $T_{peak_{conv}}$ values represent average values for 4 replicate analyses of the Elliott sample.

537 Figure 7: Distribution on $T_{peak_{conv}}$ values of m/z fragments in the 100-150 range for replicate
538 analyses of Elliott, Vertisol, Chernozem, Pahokee, Green, Pocahontas and Suwannee standard
539 samples.

540

541

542 **References**

- 543 [1] Espitalié, J., Laporte, J.L., Madec, M., Marquis, F., Leplat, P., Paulet, J., Boutefeu, A., 1977.
544 Méthode rapide de caractérisation des roches mères, de leur potentiel pétrolier et de leur degré
545 d'évolution. Revue de l'Institut Français du Pétrole 32, 23-42.
- 546 [2] Espitalié, J., Deroo, G., Marquis, F., 1985a. La pyrolyse Rock-Eval® et ses applications; première
547 partie. Revue de l'Institut Français du Pétrole 40, 563-579.
- 548 [3] Espitalié, J., Deroo, G., Marquis, F., 1985b. La pyrolyse Rock-Eval® et ses applications; deuxième
549 partie. Revue de l'Institut Français du Pétrole 40, 755-784.
- 550 [4] Lafargue, E., Marquis, F., Pillot, D., 1998. Rock-Eval® 6 applications in hydrocarbon exploration,
551 production, and soil contamination studies. Revue de l'Institut Français du Pétrole 53, 421-437.
- 552 [5] Behar, F., Beaumont, V., Penteado, H.D.B., 2001. Rock-Eval® 6 technology: performances and
553 developments. Oil Gas Science and Technology – Revue de l'Institut Français du Pétrole 56, 111-134.
- 554 [6] Disnar, J.R., Guillet, B., Keravis, D., Di-Giovanni, C., Sebag, D., 2003. Soil organic matter (SOM)
555 characterization by Rock-Eval® pyrolysis: scope and limitations. Organic Geochemistry 34, 327-343.
- 556 [7] Sebag, D., Disnar, J.R., Guillet, B., Di Giovanni, C., Verrechia, E.P., Durand, A., 2006. Monitoring
557 organic matter dynamics in soil profiles by 'Rock-Eval® pyrolysis': bulk characterization and
558 quantification of degradation. European Journal of Soil Science 57, 344-355.
- 559 [8] Copard, Y., Di-Giovanni, C., Martaud, T., Albéric, P., Olivier, J.E., 2006. Using Rock-Eval® 6 pyrolysis
560 for tracking fossil organic carbon in modern environments: implications for the roles of erosion and
561 weathering. Earth Surface Processes and Landforms 31, 135-153.
- 562 [9] Jacob, J., 2003. Enregistrement des variations paléoenvironnementales depuis 20000 ans dans le
563 Nord Est du Brésil (Lac Caço) par les triterpènes et autres marqueurs organiques. Ph.D. thesis,
564 Université d'Orléans, Orléans, France, 296 pp.

- 565 [10] Carrie, J., Sanei, H., Stern, G., 2012. Standardisation of Rock–Eval pyrolysis for the analysis of
566 recent sediments and soils. *Organic Geochemistry* 46, 38-53.
- 567 [11] Saenger, A., Cécillon, L., Sebag, D., Brun, J.-J., 2013. Soil organic carbon quantity, chemistry and
568 thermal stability in a mountainous landscape: a Rock–Eval pyrolysis survey. *Organic Geochemistry* 54,
569 101-114.
- 570 [12] Saenger, A., Cécillon, L., Poulénard, J., Bureau, F., De Daniéli, S., Gonzalez, J.-M., Brun, J.-J., 2015.
571 Surveying the carbon pools of mountain soils: a comparison of physical fractionation and Rock-Eval®
572 pyrolysis. *Geoderma* 241, 279-288.
- 573 [13] Cécillon, L., Baudin, F., Chenu, C., Houot, S., Jolivet, R., Kätterer, T., Lutfalla, S., MacDonald, A.,
574 Van Oort, F., Plante, A.F., 2018. A model based on Rock-Eval® thermal analysis to quantify the size of
575 the centennially persistent organic carbon pool in temperate soils. *Biogeosciences* 15, 2835.
- 576 [14] Soucémariadin, L., Cécillon, L., Chenu, C., Baudin, F., Nicolas, M., Girardin, C., Barré, P., 2018.
577 Is Rock-Eval® 6 thermal analysis a good indicator of soil organic carbon lability?—a method-
578 comparison study in forest soils. *Soil Biology and Biochemistry* 117, 108-116.
- 579 [15] Soucémariadin, L., Cécillon, L., Chenu, C., Baudin, F., Nicolas, M., Girardin, C., Delahaie, A.,
580 Barré, P., 2019. Heterogeneity of the chemical composition and thermal stability of particulate
581 organic matter in French forest soils. *Geoderma* 342, 65-74.
- 582 [16] Poeplau, C., Barré, P., Cécillon, L., Baudin, F., Sigurdsson, B.D., 2019. Changes in the Rock-Eval®
583 signature of soil organic carbon upon extreme soil warming and chemical oxidation - A comparison.
584 *Geoderma* 337, 181-190.
- 585 [17] Hetényi, M., Nyilas, T., Tóth, T.M., 2005. Stepwise Rock-Eval® pyrolysis as a tool for typing
586 heterogeneous organic matter in soils. *Journal of Analytical and Applied Pyrolysis* 74, 45-54.

- 587 [18] Gillespie, A.W., Sanei, H., Diocion, A., Ellert, B.H., Regier, T.Z., Chevrier, D., Dynes, J.J., Tarnocai,
588 C., Gregorich, E.G., 2014. Perennially and annually frozen soil carbon differ in their susceptibility to
589 decomposition: analysis of subarctic earth hummocks by bioassay, XANES and pyrolysis. *Soil Biology*
590 *and Biochemistry* 68, 106-116.
- 591 [19] Albrecht, R., Sebag, D., Verrecchia, E., 2015. Organic matter decomposition: bridging the gap
592 between Rock-Eval pyrolysis and chemical characterization (CPMAS ¹³C NMR). *Biogeochemistry* 112,
593 101-111.
- 594 [20] Sebag, D., Verrecchia, E.P., Cécillon, L., Adatte, T., Albrecht, R., Aubert, M., Bureau, F., Cailleau,
595 G., Copard, Y., Decaens, T., Disnar, J.R., Hetényi, M., Nyilas, T., Trombino, L., 2016. Dynamics of soil
596 organic matter based on new Rock-Eval® indices. *Geoderma* 284, 185-203.
- 597 [21] Delarue, F., Laggoun-Defarge, F., Disnar, J.R., Lottier, N., Gogo, S., 2011. Organic matter sources
598 and decay assessment in a Sphagnum-dominated peatland (Le Forbonnet, Jura Mountains, France):
599 impact of moisture conditions. *Biogeochemistry* 106, 39-52.
- 600 [22] Disnar, J.R., Guillet, B., Keravis, D., Di-Giovanni, C., Sebag, D., 2003. Soil organic matter (SOM)
601 characterization by Rock-Eval® pyrolysis: scope and limitations. *Organic Geochemistry* 34, 327-343.
- 602 [23] Hetényi, M., Nyilas, T., Csanád, S., Brukner-Wein, A., 2006. Heterogeneous organic matter from
603 the surface horizon of a temperate zone marsh. *Organic Geochemistry* 37, 1931-1942.
- 604 [24] Czirbus, N., Nyilas, T., Raucsik, B., Hetényi, M., 2016. Investigation of the effect of soil mineral
605 composition on soil organic matter stability. *Soil and Water Research* 11, 147-154.
- 606 [25] Delarue, F., Disnar, J.R., Copard, Y., Gogo, S., Jacob, J. and Laggoun-Défarge, F., 2013. Can Rock-
607 Eval® pyrolysis assess the biogeochemical composition of organic matter during peatification?
608 *Organic Geochemistry* 61, 66-72.

609 [26] Le Mer, G., Barthod, J., Dignac, M.F., Barré, P., Baudin, F., Rumpel, C., 2020. Inferring the impact
610 of earthworms on the stability of organo-mineral associations, by Rock-Eval® thermal analysis and
611 ¹³C NMR spectroscopy. *Organic Geochemistry* 144, 104106.

612 [27] Tambach, T.J., Veld, H., Griffioen, J., 2009. Influence of HCl/HF treatment on organic matter in
613 aquifer sediments: A Rock-Eval® pyrolysis study. *Applied Geochemistry* 24, 2144-2151.

614 [28] Derenne, S., Quénéa, K., 2015. Analytical pyrolysis as a tool to probe soil organic matter. *Journal*
615 *of Analytical and Applied Pyrolysis* 111, 108-120.

616 [29] Hedges, J.I., Eglinton, G., Hatcher, P.G., Kirchman, D.L., Arnosti, C., Derenne, S., Evershed, R.P.,
617 Kögel-Knabner, I., de Leeuw, J.W., Littke, R., Michaelis, W., Rullkötter, J. 2000. The molecularly-
618 uncharacterized component of non living organic matter in natural environments. *Organic*
619 *Geochemistry* 31, 945-958.

620 [30] BJOROY, M., HALL, K., HALL, P.B., 1992. Detailed hydrocarbon analyser for well site and laboratory
621 use. *Marine and Petroleum Geology* 9, 648-665.

622 [31] Behar, F., Vandenbroucke, M., Tang, Y., Marquis, F., Espitalié, J., 1997. Thermal cracking of
623 kerogen in open and closed systems. *Organic Geochemistry* 26, 321-339.

624 [32] Cadeau, P., Romero-Sarmiento, M-F., Sissmann, O., Beaumont, V., 2020. On-line recovery system
625 coupled to a Rock-Eval® device: An analytical methodology to characterize liquid and solid samples.
626 *Organic Geochemistry* 144, 104014.

627 [33] Romero-Sarmiento, M.F., Pillot, D., Letort, G., Lamoureux-Var, V., Beaumont, V., Huc, A.Y.,
628 Garcia, B., 2016a. New Rock-Eval® method for characterization of unconventional shale resource
629 systems. *Oil & Gas Science and Technology – Revue de l'Institut Français du Pétrole Energies*
630 *Nouvelles* 71, 37.

631 [34] Romero-Sarmiento, M.-F., Euzen, T., Rohais, S., Jiang, C., Littke, R., 2016b. Artificial thermal
632 maturation of source rocks at different thermal maturity levels: application to the Triassic Montney
633 and Doig formations in the Western Canada Sedimentary Basin. *Organic Geochemistry* 97, 148-162.

634 [35] Romero-Sarmiento, M.-F., Ramiro-Ramirez, S., Berthe, G., Fleury, M., Littke, R., 2017.
635 Geochemical and petrophysical source rock characterization of the Vaca Muerta Formation,
636 Argentina: implications for unconventional petroleum resource estimations. *International Journal of*
637 *Coal Geology* 184, 27-41.

638 [36] Schulten, H.-R., Abbt-Braun, G., Frimmel, F.H., 1987. Time-resolved pyrolysis field ionisation
639 mass spectrometry of humic material isolated from freshwater. *Environ. Sci. Technol.* 21, 349-357.

640 [37] Hempfling, R., Schulten, H.-R., 1990. Chemical characterization of the organic matter in forest
641 soils by Curie point pyrolysis-GC/MS and pyrolysis-field ionization mass spectrometry. *Organic*
642 *Geochemistry* 15, 131-145.

643 [38] Schnitzer, M., Schulten, H.-R., 1992. The analysis of soil organic matter by pyrolysis-field
644 ionization mass spectrometry. *Soil Science Society of America* 56, 1811-1817.

645 [39] Benner, R., Pakulski, J.D., McCarthy, M., Hedges, J.I., Hatcher, P.G., 1992. Bulk chemical
646 characteristics of dissolved organic matter in the ocean. *Science* 255, 1561-1564.

647 [40] Boon, J.J., 1992. Analytical pyrolysis mass spectrometry: new vistas opened by temperature-
648 resolved in-source PYMS. *International Journal of Mass Spectrometry and Ion Processes* 118-119,
649 755-787.

650 [41] Boon, J.J., Klap, V.A., Eglinton, T.I., 1998. Molecular characterization of microgram amounts of
651 oceanic colloidal organic matter by direct temperature-resolved ammonia chemical ionization mass
652 spectrometry. *Organic Geochemistry* 29, 1051-1061.

653 [42] Gibson, E.K., 1973. Thermal analysis-mass spectrometer computer system and its application to
654 the evolved gas analysis of Green River Shale and Lunar soil samples. *Thermochimica Acta* 5, 243-
655 255.

656 [43] Schulten, H.-R., Schnitzer, M., 1993. Temperature-resolved in-source pyrolysis-soft ionization
657 mass spectrometry of soil humic acids. *Organic Geochemistry* 20, 17-25.

658 [44] Leinweber, P., Schulten, H.-R., 1995. Composition, stability and turnover of soil organic matter:
659 investigations by off-line pyrolysis and direct pyrolysis-mass spectrometry. *Journal of Analytical and*
660 *Applied Pyrolysis* 32, 91-110.

661 [45] Leinweber, P., Schulten, H.-R., 1999. Advances in analytical pyrolysis of soil organic matter.
662 *Journal of Analytical and Applied Pyrolysis* 49, 359-383.

663 [46] Eglinton, T.I., Boon, J.J., Minor, E.C., Oison, R.J., 1996. Microscale characterization of algal and
664 related particulate organic matter by direct temperature-resolved mass spectrometry. *Marine*
665 *Chemistry* 52, 27-54.

666 [47] Minor, E.C., Eglinton, T.I., Boon, J.J., 2000. Direct temperature-resolved mass spectrometry as a
667 technique for the semi-quantitative analysis of marine particulate organic matter. *Journal of*
668 *Analytical and Applied Pyrolysis* 53, 19-34.

669 [48] Minor, E.C., Boon, J.J., Harvey, H.R., Mannino, A., 2001. Estuarine organic matter composition as
670 probed by direct temperature-resolved mass spectrometry and traditional geochemical techniques.
671 *Geochimica et Cosmochimica Acta* 65, 2819-2834.

672 [49] Braadbaart, F., Boon, J.J., van der Horsta, J., van Bergen, P.F., 2004. Laboratory simulations of
673 the transformation of peas as a result of heating: the change of the molecular composition by DTMS.
674 *Journal of Analytical and Applied Pyrolysis* 71, 997-1026.

675 [50] Conesa, J.A., Marcilla, A., Moral, R., Moreno-Caselles, J., Perez-Espinosa, A., 1998. Evolution of
676 gases in the primary pyrolysis of different sewage sludges. *Thermochimica Acta* 313, 63-73.

677 [51] Smidt, E., Eckhardt, K.-U., Lechner, P., Schulten, H.-R., Leinweber, P., 2005. Characterization of
678 different decomposition stages of biowaste using FT-IR spectroscopy and pyrolysis-field ionisation
679 mass spectrometry. *Biodegradation* 16, 67-79.

680 [52] Lopez Capel, E., Abbott, G.D., Thomas, K.M., Manning, D.A.C., 2006a. Coupling of thermal
681 analysis with quadrupole mass spectrometry and isotope ratio mass spectrometry for simultaneous
682 determination of evolved gases and their carbon isotopic composition. *Journal of Analytical and*
683 *Applied Pyrolysis* 75, 82-89.

684 [53] Lopez Capel, E., de la Rosa Arranz, J.M., González-Vila, F.J., González-Perez, J.A., Manning, D.A.C.,
685 2006b. Elucidation of different forms of organic carbon in marine sediments from the Atlantic coast
686 of Spain using thermal analysis coupled to isotope ratio and quadrupole mass spectrometry. *Organic*
687 *Geochemistry* 37, 1983-1994.

688 [54] Jacob, J., Delarue, F., Copard, Y., Le Milbeau, C., Grasset, L., Brockmann, P., 2023. Comparison
689 between Rock-Eval® and temperature-programmed pyrolysis/mass spectrometry for the analysis of
690 environmental and geological samples. *Journal of Analytical and Applied Pyrolysis* 173, 106078.

691 [55] Langford, F.F., Blanc-Valleron, M.M., 1990. Interpreting Rock- Eval pyrolysis data using graphs of
692 pyrolysable hydrocarbons vs. total organic carbon. *AAPG Bulletin* 74, 799-804.

693 [56] Dembicki, H. Jr., 1992. The effects of the mineral matrix on the determination of kinetic
694 parameters using modified Rock Eval pyrolysis. *Organic Geochemistry* 18, 531-539.

695 [57] Monreal, C.M., Schulten, H.R., Kodama, H., 1997. Age, turnover and molecular diversity of soil
696 organic matter in aggregates of a Gleysol. *Canadian Journal of Soil Science* 77, 379-388.

- 697 [58] Schulten, H.-R., Leinweber, P., 1999. Thermal stability and composition of mineral-bound organic
698 matter in density fractions of soil. *European Journal of Soil Science* 50, 237-248.
- 699 [59] Schulten, H.R., Lattimer, R.P., 1984. Application of mass spectrometry to polymers. *Mass*
700 *Spectrometry Reviews* 3, 231-315.
- 701 [60] Nali, M., Corana, F., Montanari, L., Pellegrini, L., 1994. A pyrolysis-gas chromatography/mass
702 spectrometry study on coals. *Journal of Analytical and Applied Pyrolysis* 29, 15-23.
- 703 [61] Schmidt, M.W.I., Skjemstad, J.O., Gehrt, E., Kögel-Knabner, I. 1999. Charred organic carbon in
704 German chernozemic soils. *European Journal of Soil Science* 50, 351-365.
- 705

707 Table S1: List of *m/z* fragments together with the chemical families they were attributed to, from the literature.

<i>m/z</i>	Chemical family	Ref.	<i>m/z</i>	Chemical family	Ref.	<i>m/z</i>	Chemical family	Ref.	<i>m/z</i>	Chemical family	Ref.
50	Polysaccharides	[1]	81	Aliphatic polymers	[1]	103	N-containing compounds	[11]	124	Alk-dihydroxybenz	[7]
51	Lignin	[1]	81	Proteins	[1]	103	N-compounds	[10]	124	Dihydroxybenzenes	[9]
51	Polyphenols	[1]	81	N-containing compounds	[11]	105	Lipids	[1]	124	Phenols and lignin monomers	[11]
52	Lignin	[1]	81	N-compounds	[10]	106	Alkylaromatics (alkylbenzenes)	[2]	124	Lignin monomers and phenols	[10]
54	Lignin	[1]	82	Mono- and polysaccharides	[2]	106	M,D,T,P-benzene	[6]	125	Lipids	[1]
54	Proteins	[1]	82	Carbo-pent/hexose	[6]	106	Ethyl- to docosyl-benzene	[6]	125	N-containing compounds	[11]
54	Polyphenols	[1]	82	Cellulose	[4]	106	Benzene	[4]	125	N-compounds	[10]
55	Aliphatic polymers	[1]	82	Methyl furan	[4]	106	Lipids	[1]	126	Mono- and polysaccharides	[2]
55	Hydrocarbons	[3]	82	Aliphatic polymers	[1]	106	Benzenes	[7]	126	Carbohydrates	[12]
56	Low-mass olefins	[6]	82	Polysaccharides	[1]	106	Alkylbenzenes	[9]	126	Carbo-pent/hexose	[6]
56	Proteins	[1]	82	Polysaccharides	[5]	106	Alkylaromatics	[11]	126	Phenols	[6]
56	Alkanes	[7]	82	Carbohydrates	[11]	106	Alkylaromatics	[10]	126	Cellulose	[4]
57	Aliphatic polymers	[1]	82	Pent/hexose-carbo	[10]	107	Nitrogen compounds	[2]	126	Lipids	[1]
57	Hydrocarbons	[3]	84	Mono- and polysaccharides	[2]	107	Aliphatic polymers	[1]	126	Polysaccharides	[1]
57	Peptides	[11]	84	Carbo-pent/hexose	[6]	108	Phenols and lignin monomers	[2]	126	Polysaccharides	[5]
57	Peptides	[10]	84	Low-mass olefins	[6]	108	Lignin monomers	[12]	126	Polysaccharides	[8]
58	Mono- and polysaccharides	[2]	84	Cellulose	[4]	108	Phenols	[6]	126	Levoglucosenone / Carbohydrates	[9]
58	Acetone	[6]	84	Hydroxyfuran	[4]	108	Phenol	[4]	126	Carbohydrates	[11]
58	Polysaccharides	[1]	84	Polysaccharides	[1]	108	Aliphatic polymers	[1]	126	Pent/hexose-carbo	[10]
58	Alkane fragments	[7]	84	Alkanes	[7]	108	Phenols	[7]	126	Lignin monomers and phenols	[10]
59	N-containing compounds	[12]	84	Carbohydrates	[11]	108	Alkylphenols	[9]	127	Lipids	[1]
59	Acetamide	[6]	84	Peptides	[11]	108	Phenols and lignin monomers	[11]	128	Mono- and polysaccharides	[2]
59	N-compounds	[6]	84	Pent/hexose-carbo	[10]	108	Lignin monomers and phenols	[10]	128	Polysaccharide	[4]
59	Acetamide	[9]	85	N-containing compounds	[12]	109	Trimethyl pyrrole	[6]	128	Lipids	[1]
59	N-containing compounds	[11]	85	Amino-acid species and amides	[12]	109	Aliphatic polymers	[1]	128	Polysaccharides	[1]
59	N-compounds	[10]	85	Lignin	[1]	109	N-containing compounds	[11]	128	Naphtalenes	[7]
60	Mono- and polysaccharides	[2]	86	Mono- and polysaccharides	[2]	109	N-compounds	[10]	129	Lipids	[1]

60	Amino-acid species and amides	[12]	86	Lignin	[1]	110	Phenols and lignin monomers	[2]	129	Peptides	[11]
60	Carbohydrates	[12]	87	Amino-acid species and amides	[12]	110	Mono- and polysaccharides	[2]	129	Peptides	[10]
60	Carbo-pent/hexose	[6]	87	Aliphatic polymers	[1]	110	Carbohydrates	[12]	130	Lipids	[1]
60	Acetic acid	[6]	87	Peptides	[11]	110	Carbo-pent/hexose	[6]	131	Nitrogen compounds	[2]
60	Cellulose	[4]	87	Peptides	[10]	110	Phenols	[6]	131	N-compounds	[6]
60	Acetic acid	[4]	88	Lignin	[1]	110	Cellulose	[4]	131	Prot-alk-pyrroles/indoles	[4]
60	Lignin	[1]	89	Aliphatic polymers	[1]	110	Aliphatic polymers	[1]	131	Lipids	[1]
60	Polyphenols	[1]	90	Lignin	[1]	110	Polysaccharides	[1]	131	Proteins	[1]
60	Polysaccharides	[5]	91	Lipids	[1]	110	Polysaccharides	[5]	132	Mono- and polysaccharides	[2]
60	Carbohydrates	[11]	91	Peptides	[11]	110	Alk-dihydroxybenz	[7]	132	Carbohydrates	[12]
60	Pent/hexose-carbo	[10]	91	Peptides	[10]	110	Dihydroxybenzenes	[9]	132	Carbo-pent/hexose	[6]
61	Lignin	[1]	92	Alkylaromatics (alkylbenzenes)	[2]	110	Carbohydrates	[11]	132	Lipids	[1]
61	Polyphenols	[1]	92	M,D,T,T,P-benzene	[6]	110	Phenols and lignin monomers	[11]	132	Polysaccharides	[1]
63	Lignin	[1]	92	Benzene	[4]	110	Pent/hexose-carbo	[10]	132	Alk-tetralins/benzofurans	[7]
64	Lignin	[1]	92	Benzenes	[7]	110	Lignin monomers and phenols	[10]	132	Carbohydrates	[11]
65	Aliphatic polymers	[1]	92	Alkylbenzenes	[9]	111	Aliphatic polymers	[1]	132	Pent/hexose-carbo	[10]
67	Nitrogen compounds	[2]	92	Alkylaromatics	[11]	111	N-containing compounds	[11]	133	Lipids	[1]
67	Pyrrole	[6]	92	Alkylaromatics	[10]	111	N-compounds	[10]	134	Alkylaromatics (alkylbenzenes)	[2]
67	Prot-alk-pyrroles/indoles	[4]	93	N-compounds	[6]	112	Mono- and polysaccharides	[2]	134	M,D,T,T,P-benzene	[6]
67	Aliphatic polymers	[1]	93	Aliphatic polymers	[1]	112	Carbohydrates	[12]	134	Ethyl- to docosyl-benzene	[6]
67	Proteins	[1]	93	Proteins	[1]	112	Carbo-pent/hexose	[6]	134	Lipids	[1]
67	N-containing compounds	[11]	93	N-containing ions	[5]	112	Cellulose	[4]	134	Alkylbenzenes	[9]
67	N-compounds	[10]	94	Phenols and lignin monomers	[2]	112	Aliphatic polymers	[1]	134	Alkylaromatics	[11]
68	Cellulose	[4]	94	Lignin monomers	[12]	112	Polysaccharides	[1]	134	Alkylaromatics	[10]
68	Furan	[4]	94	Phenols	[6]	112	Carbohydrates	[11]	135	Amino-acid species and amides	[12]
68	Lignin	[1]	94	Phenol	[4]	112	Pent/hexose-carbo	[10]	135	Lipids	[1]
69	Lignin	[1]	94	Lignin	[1]	113	Lipids	[1]	135	Peptides	[11]
69	Hydrocarbons	[3]	94	Phenols	[7]	114	Mono- and polysaccharides	[2]	135	Peptides	[10]
70	Amino-acid species and amides	[12]	94	Alkylphenols	[9]	114	Carbohydrates	[12]	136	Lipids	[1]
70	Low-mass olefins	[6]	94	Phenols and lignin monomers	[11]	114	Carbo-pent/hexose	[6]	136	Phenols	[7]
70	Polysaccharides	[1]	94	Lignin monomers and phenols	[10]	114	Polysaccharide	[4]	137	N-containing compounds	[11]

70	Lignin	[1]	95	Nitrogen compounds	[2]	114	Aliphatic polymers	[1]	137	N-compounds	[10]
70	Polyphenols	[1]	95	Hydroxypyryl	[6]	114	Polysaccharides	[1]	138	Phenols and lignin monomers	[2]
70	Alkanes	[7]	95	Prot-alk-pyrroles/indoles	[4]	114	Polysaccharides	[8]	138	Alkylaromatics (alkylbenzenes)	[2]
70	Peptides	[11]	95	Lignin	[1]	114	Carbohydrates	[11]	138	Phenols	[6]
70	Peptides	[10]	95	Polyphenols	[1]	114	Pent/hexose-carbo	[10]	138	Monomeric lignins	[6]
71	N-containing compounds	[12]	95	N-containing compounds	[11]	115	Aliphatic polymers	[1]	138	Guaiacyl units from lignin residues	[4]
71	Lignin	[1]	95	N-compounds	[10]	115	Peptides	[11]	138	Phenol	[4]
71	Polyphenols	[1]	96	Mono- and polysaccharides	[2]	115	Peptides	[10]	138	Monomeric products from lignin	[1]
72	Carbohydrates	[12]	96	Carbohydrates	[12]	116	Lipids	[1]	138	Alk-dihydroxybenz	[7]
72	Carbo-pent/hexose	[6]	96	Carbo-pent/hexose	[6]	117	Nitrogen compounds	[2]	138	Dihydroxybenzenes	[9]
72	Cellulose	[4]	96	Cellulose	[4]	117	N-compounds	[6]	138	Phenols and lignin monomers	[11]
72	Lignin	[1]	96	Furfural and/or pyrone	[4]	117	Prot-alk-pyrroles/indoles	[4]	138	Lignin monomers and phenols	[10]
72	Polysaccharides	[1]	96	Aliphatic polymers	[1]	117	Aliphatic polymers	[1]	139	N-containing compounds	[11]
72	Alkane fragments	[7]	96	Polysaccharides	[1]	117	Proteins	[1]	139	N-compounds	[10]
72	Carbohydrates	[11]	96	Polysaccharides	[5]	117	N-containing ions	[5]	140	Phenols	[6]
72	Pent/hexose-carbo	[10]	96	Polysaccharides	[8]	118	Aliphatic polymers	[1]	140	Monomeric lignins	[6]
73	Amino-acid species and amides	[12]	96	Carbohydrates	[11]	119	Aliphatic polymers	[1]	140	Monomeric products from lignin	[1]
73	Peptides	[11]	96	Pent/hexose-carbo	[10]	120	Phenols and lignin monomers	[2]	140	Phenols and lignin monomers	[11]
73	Peptides	[10]	97	Amino-acid species and amides	[12]	120	Alkylaromatics (alkylbenzenes)	[2]	142	Alk-napth	[2]
74	Amino-acid species and amides	[12]	97	Aliphatic polymers	[1]	120	M,D,T,T,P-benzene	[6]	142	M,D,T,T,P-benzene	[6]
74	Aliphatic polymers	[1]	97	Peptides	[11]	120	Ethyl- to docosyl-benzene	[6]	142	Naphtalenes	[7]
74	Polysaccharides	[1]	97	Peptides	[10]	120	Phenol	[4]	142	Alkylaromatics	[11]
74	Peptides	[11]	98	Mono- and polysaccharides	[2]	120	Aliphatic polymers	[1]	142	Alkylaromatics	[10]
74	Peptides	[10]	98	Carbohydrates	[12]	120	Alkylbenzenes	[9]	144	Mono- and polysaccharides	[2]
75	Lignin	[1]	98	Carbo-pent/hexose	[6]	120	Alkylaromatics	[11]	144	Carbo-pent/hexose	[6]
75	Peptides	[11]	98	Low-mass olefins	[6]	120	Peptides	[11]	144	Polysaccharides	[1]
75	Peptides	[10]	98	Cellulose	[4]	120	Alkylaromatics	[10]	144	Polysaccharides	[8]
76	Aliphatic polymers	[1]	98	Furfuryl alcohol	[4]	122	Phenols and lignin monomers	[2]	144	Carbohydrates	[11]
77	Lignin	[1]	98	Lipids	[1]	122	Phenols	[6]	144	Pent/hexose-carbo	[10]
78	Benzene	[4]	98	Polysaccharides	[1]	122	Aliphatic polymers	[1]	145	Nitrogen compounds	[2]
78	Proteins	[1]	98	Polysaccharides	[8]	122	Phenols	[7]	146	Tetralins / benzofurans	[7]

78	Benzenes	[7]	98	Alkanes	[7]	122	Alkylphenols	[9]	148	Alkylaromatics (alkylbenzenes)	[2]
79	Nitrogen compounds	[2]	98	Carbohydrates	[11]	122	Phenols and lignin monomers	[11]	148	M,D,T,T,P-benzene	[6]
79	Pyridine	[6]	98	Pent/hexose-carbo	[10]	122	Lignin monomers and phenols	[10]	148	Ethyl- to docosyl-benzene	[6]
79	N-compounds	[6]	99	N-containing compounds	[12]	123	Tetramethyl pyrrole	[6]	148	Alkylbenzenes	[9]
79	Lignin	[1]	99	Amino-acid species and amides	[12]	123	Aliphatic polymers	[1]	148	Alkylaromatics	[11]
79	Proteins	[1]	99	Lipids	[1]	123	N-containing compounds	[11]	148	Alkylaromatics	[10]
79	N-containing ions	[5]	99	Peptides	[11]	123	N-compounds	[10]	150	Phenols and lignin monomers	[2]
79	N-containing compounds	[11]	99	Peptides	[10]	124	Phenols and lignin monomers	[2]	150	Monomeric lignins	[6]
79	N-compounds	[10]	100	Aliphatic polymers	[1]	124	Phenols	[6]	150	Guaiacyl units from lignin residues	[4]
80	Lignin	[1]	101	Aliphatic polymers	[1]	124	Monomeric lignins	[6]	150	Phenol	[4]
81	Nitrogen compounds	[2]	102	Cellulose	[4]	124	Guaiacyl units from lignin residues	[4]	150	Monomeric products from lignin	[1]
81	Methyl pyrrole	[6]	102	Aliphatic polymers	[1]	124	Phenol	[4]	150	Phenols	[7]
81	N-compounds	[6]	103	Benzonitrile	[6]	124	Lipids	[1]	150	Phenols and lignin monomers	[11]
81	Prot-alk-pyrroles/indoles	[4]	103	Aliphatic polymers	[1]	124	Monomeric products from lignin	[1]			

708

709 References

- 710 [1] Hempfling, R., Schulten, H.-R., 1990. Chemical characterization of the organic matter in forest soils by Curie point pyrolysis-GC/MS and pyrolysis-field ionization mass spectrometry. *Organic*
711 *Geochemistry* 15, 131-145.
- 712 [2] Leinweber, P., Schulten, H.-R., Horte, C., 1992. Differential thermal analysis, thermogravimetry and in-source pyrolysis-mass spectrometry studies on the formation of soil organic matter.
713 *Thermochimica Acta* 200, 151-167.
- 714 [3] Gibson, E.K., 1973. Thermal analysis-mass spectrometer computer system and its application to the evolved gas analysis of Green River Shale and Lunar soil samples. *Thermochimica Acta* 5,
715 243-255.
- 716 [4] Saiz-Jimenez, C., Haider, K., Meuzelaar, H.L.C., 1979. Comparisons of soil organic matter and its fractions by pyrolysis mass-spectrometry. *Geoderma* 22, 25-37.
- 717 [5] Schulten, H.-R., 1987. Pyrolysis and soft ionization mass spectrometry of aquatic/terrestrial humic substances and soils. *Journal of Analytical and Applied Pyrolysis* 12, 149-186.
- 718 [6] Schnitzer, M., Schulten, H.-R., 1992. The analysis of soil organic matter by pyrolysis-field ionization mass spectrometry. *Soil Science Society of America* 56, 1811-1817.
- 719 [7] Simmler, N., Schulten, H.-R., Yun, Y., Meuzelaar, H.L.C., 1992. Thermochemical Analysis of U.S. Argonne Premium Coal Samples by Time-Resolved Pyrolysis-Field Ionization Mass
720 Spectrometry. In: Meuzelaar H.L.C. (Ed.), *Advances in Coal Spectroscopy. Modern Analytical Chemistry*. Springer, Boston, MA, USA, pp. 295-337.
- 721 [8] Boon, J.J., 1992. Analytical pyrolysis mass spectrometry: new vistas opened by temperature-resolved in-source PYMS. *International Journal of Mass Spectrometry and Ion Processes* 118-
722 119, 755-787.
- 723 [9] Schulten, H.-R., Schnitzer, M., 1993. Temperature-resolved in-source pyrolysis-soft ionization mass spectrometry of soil humic acids. *Organic Geochemistry* 20, 17-25.
- 724 [10] Monreal, C.M., Schulten, H.R., Kodama, H., 1997. Age, turnover and molecular diversity of soil organic matter in aggregates of a Gleysol. *Canadian Journal of Soil Science* 77, 379-388.
- 725 [11] Schulten, H.-R., Leinweber, P., 1999. Thermal stability and composition of mineral-bound organic matter in density fractions of soil. *European Journal of Soil Science* 50, 237-248.
- 726 [12] Leinweber, P., Schulten, H.-R., 1995. Composition, stability and turnover of soil organic matter: investigations by off-line pyrolysis and direct pyrolysis-mass spectrometry. *Journal of*
727 *Analytical and Applied Pyrolysis* 32, 91-110.

728 Table S2: Tpeak values determined for each m/z fragment in replicate analyses of standard samples
729 and organized by clusters.

730 → See attached file

731 Figure S1: Total Mass spectra illustrate the relative abundance of m/z fragments recorded during the
732 time interval of the TIC peak. Background subtraction is performed before and after the TIC peak.
733 The amount of sample used is indicated after the sample name.

734 Figure S2: Evolution of TIC (red curve) and relative abundance of selected m/z (57, 78, 91, 92, 94 and
735 107) with temperature (converted from time of Py-MS analysis) for the eight standard samples.
736 Tpeak values are indicated for both TIC and m/z curves. The amount of material used for each
737 analysis (in mg) is indicated after the sample name.