



**HAL**  
open science

## **BOOM! Tephrochronological dataset and exploration tool of the Southern (33–46° S) and Austral (49–55° S) volcanic zones of the Andes**

Consuelo Martínez Fontaine, Vanessa Peña-Araya, Chiara Marmo, Marine Le Morvan, Guillaume Delpech, Karen Fontijn, Giuseppe Siani, Lucile Cosyn-Wexsteen

### ► To cite this version:

Consuelo Martínez Fontaine, Vanessa Peña-Araya, Chiara Marmo, Marine Le Morvan, Guillaume Delpech, et al.. BOOM! Tephrochronological dataset and exploration tool of the Southern (33–46° S) and Austral (49–55° S) volcanic zones of the Andes. *Quaternary Science Reviews*, 2023, 316, pp.108254. 10.1016/j.quascirev.2023.108254. hal-04219211

**HAL Id: hal-04219211**

**<https://hal.science/hal-04219211>**

Submitted on 26 Sep 2023

**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 Attribution 4.0 International License

1 **BOOM! Tephrochronological dataset and exploration tool of the**  
2 **Southern (33–46° S) and Austral (49–55° S) Volcanic Zones of the**  
3 **Andes**

4 Consuelo Martínez Fontaine<sup>a, b, 1</sup>, Vanessa Peña-Araya<sup>c</sup>, Chiara Marmo<sup>a</sup>, Marine Le Morvan<sup>e</sup>, Guillaume  
5 Delpech<sup>a</sup>, Karen Fontijn<sup>d</sup>, Giuseppe Siani<sup>a</sup>, Lucile Cosyn-Wexsteen<sup>a</sup>

6 <sup>a</sup>Géosciences Paris-Saclay (GEOPS), Université Paris-Saclay, CNRS UMR 8148, 91405 Orsay, France

7 <sup>b</sup>Laboratoire des Sciences du Climat et de l'Environnement (LSCE), Laboratoire mixte CNRS-CEA, Avenue  
8 de la Terrasse, 91198 Gif-sur-Yvette Cedex, France

9 <sup>1</sup>Present address: Centro de investigación en Tecnologías para la Sociedad C+, Universidad del  
10 Desarrollo. Av. Plaza 680, Edificio S, 7610658 Las Condes, Chile. E-mail: [cdpmartinez@udd.cl](mailto:cdpmartinez@udd.cl)

11 <sup>c</sup>Inria, Laboratoire CNRS, Université Paris-Saclay - Bât. 660 91405 Orsay Cedex, France

12 <sup>d</sup>Department of Geosciences Environment and Society, Université Libre de Bruxelles, Av. F Roosevelt 50,  
13 1050 Brussels, Belgium

14 <sup>e</sup>Inria Saclay – Île-de-France, Parietal team, 1 Rue Honoré d'Estienne d'Orves, 91120 Palaiseau, France

15 Corresponding author: Consuelo Martínez Fontaine

16

## 17 **Abstract**

18 Tephrochronology studies the deposits of explosive volcanic eruptions in the stratigraphic record. The  
19 Southern (SVZ, 33–46° S) and Austral (AVZ, 49–55° S) Volcanic Zones of the Andes are two very active  
20 volcanic zones where tephrochronology is of great use. There, it can be used to improve chronologies of  
21 paleoenvironmental records in Patagonia, an area providing valuable records at global scale; as well as  
22 to identify areas likely to be affected by volcanic eruptions in the future, essential for producing volcanic  
23 hazard maps. The close proximity of many volcanic centers with recurrent explosive activity, which have  
24 very similar geochemical compositions, and very often poor age constraints, represent a challenge for the  
25 study of tephrochronology in the region. In addition to this, the ever-growing amount of  
26 tephrochronological information in the area, dispersed in different types of publications which vary  
27 greatly in format, makes the integration of the data produced by different actors, and consecutively its  
28 interpretation, increasingly difficult. Here we address this issue by compiling the BOOM! dataset, which  
29 integrates ~30 years of research on 32 active volcanic centers and 132 different eruptions, which took  
30 place during the last 20,000 years. To help users find and reuse data in the large dataset, we developed  
31 an online platform which provides user-friendly tools for exploring it, and helps users download subsets  
32 of it. To integrate this very heterogeneous information, special attention was given to include information  
33 which allows users to evaluate data quality and comparability, as well as to provide tools in the explorer  
34 for users to filter data by different criteria. The integration of this dataset opens new perspectives for the  
35 development of novel visualizations of tephrochronological data, for example, to better understand the  
36 multidimensional uncertainties associated with it. For example, uncertainties associated with analytical  
37 precision, with age estimates of both tephra deposits and volcanic eruptions, and of tephra classification.  
38 Additionally, it allows for the use of robust statistical tools to correlate tephra deposits, including those  
39 based on machine learning algorithms, which are here explored.

40 **Keywords:** Tephrochronology, Southern Volcanic Zone of the Andes, Austral Volcanic Zone of the Andes,  
41 Geochemistry, Radiocarbon, Machine learning

42

## 43 **1 Introduction**

44 Tephrochronology is a discipline of geosciences which uses the deposits of explosive volcanic eruptions  
45 as stratigraphic and chronological markers. These deposits are called tephtras, and by studying them,  
46 researchers can reconstruct the eruptive history of volcanic centers: how often volcanic activity occurs,  
47 the magnitude of the eruptions, and the dispersion of the volcanic products. The Southern (SVZ, 33–46°  
48 S) and Austral (AVZ, 49–55° S) Volcanic Zones of the Andes are two very active volcanic zones, where  
49 ~65 volcanic centers have been identified. Many of these centers have had recurrent explosive activity in  
50 historical times, e.g., Llama (Naranjo & Moreno, 2005), Puyehue-Cordón Caulle (Naranjo et al., 2017),  
51 Calbuco (Sellés & Moreno, 2011), Hudson (Naranjo et al., 1993) and Lautaro volcanic centers (Mayr et al.,  
52 2019). While for others, evidence of recurrent explosive activity during the last ~20,000 years has also  
53 been identified, e.g., Mocho-Choshuenco (Rawson et al., 2015), Michinmahuida (Amigo et al., 2013) and  
54 Mount Burney volcanic centers (Smith et al., 2019). More so, some of these volcanic centers are very  
55 explosive, with distal dispersion of tephtras being registered hundreds of kilometers from their source, as  
56 the H1 eruption from Hudson volcanic center (Naranjo & Stern, 1998; Kilian et al., 2003; Stern, 2008; Del  
57 Carlo et al., 2018; Smith et al., 2019). Because of the latter, one important concern in these areas is  
58 protecting surrounding villages and other important infrastructure (e.g., power plants or dams) from  
59 potentially dangerous eruptions. Part of this work is done by the Chilean Geology and Mining agency  
60 SERNAGEOMIN, which produces volcanic hazard maps of the different volcanic centers and regions in  
61 the country and distributes them to the population (e.g., Bertin et al., 2018). In order to produce these  
62 maps, SERNAGEOMIN relies on tephrochronological and volcanological information from their own  
63 surveys as well as from peer-reviewed scientific publications (e.g., Sellés & Moreno, 2011; Bertin et al.,  
64 2018), which helps them identify areas likely to be affected by future eruptions. Besides this very  
65 important application, and because of the recurrent explosive activity of the SVZ and AVZ, there is great  
66 potential of using tephrochronology to improve the chronologies of paleoenvironmental records in  
67 Patagonia (Fontijn et al., 2014). If tephtras deposited during the same eruptive event are identified in  
68 different sedimentary archives (e.g., lake sediment cores, marine sediment cores, peat cores, or

BOOM! Tephrochronology dataset and exploration tool of the Southern (33–46° S) and Austral (49–55° S)  
Volcanic Zones of the Andes

69 archaeological sites), the chronologies of different records (e.g., paleoclimatological,  
70 paleoceanographical, or archaeological records) can be synchronized. Paleoenvironmental records in  
71 Patagonia provide valuable climatic records, such as the variations in the Southerly Westerly Winds (e.g.,  
72 Moreno et al., 2018); valuable oceanographic records, such as changes in oceanic circulation in the  
73 Southern Ocean (e.g., Siani et al., 2013); as well as unique records describing interactions between human  
74 populations, climate change and volcanic activity at high latitudes (e.g., Villarosa et al., 2006; Prieto et al.,  
75 2013). In order to correctly integrate these records, having good chronologies is essential, thus  
76 highlighting the potential of using tephrochronology in the area.

77 In the last four decades, tephrochronology together with volcanology have increased our understanding  
78 of the eruptive history of the SVZ and AVZ, revealing higher than previously thought recurrence rates and  
79 explosivity of many of the volcanic centers in the area (e.g., Chaitén and Michinmahuida volcanoes (Amigo  
80 et al., 2013; Watt et al., 2013; Moreno et al., 2015; Alloway et al., 2017a, b; Martínez Fontaine et al., 2021)).  
81 This work, and especially the study of sedimentary archives which favor the preservation of tephras  
82 (such as lake cores (e.g., Bertrand et al., 2014)), has revealed a very intricate tephrostratigraphic record  
83 in the area, given by the close proximity of the volcanic centers together with their high eruptive  
84 recurrence rates. In order to disentangle this record, tephrochronologists try to fingerprint tephras using  
85 their physical characteristics, geochemical composition, stratigraphic position, and age (e.g., Lowe, 2011).  
86 By doing this, they can correlate tephras deposited in different locations and ideally identify their volcanic  
87 source and the specific eruption during which they were deposited. In the SVZ and AVZ, doing this is not  
88 trivial. For example, many volcanic centers and eruptions have very similar geochemical compositions,  
89 or in other cases, the age estimates of tephra deposits have high uncertainties, or they do not exist at all,  
90 making it hard to distinguish tephras based on these criteria. This issue is enhanced by the *unFAIRness* of  
91 the tephrochronological data (e.g., physical characteristics, geochemical composition, stratigraphic  
92 position, and age of tephras) in the area, i.e., data is not readily Findable, Accessible, Interoperable, or  
93 Reusable (Wilkinson et al., 2016; Abbott et al., 2022). At present, the available tephrochronological and  
94 volcanological information of the SVZ and AVZ can be found dispersed in journals, undergraduate or  
95 doctoral theses, publications from government institutions (SERNAGEOMIN), and personal collections,

96 which vary greatly in format, and with some information sometimes very difficult to find. This, together  
97 with the increased data stream associated with the development of analytical techniques (such as Electro  
98 Probe microanalysis (EMPA, also referred to as EMP), typically used for analyzing glass (or mineral)  
99 major element composition of volcanic ash shards, e.g., Lowe (2011)), makes the integration of the data  
100 produced by different actors (researchers and practitioners from government institutions), and  
101 consecutively its interpretation, more and more difficult. The increasing need for organization of the vast  
102 and growing amount of tephrochronological and volcanological data in the region, is an issue that has  
103 already been addressed in previous publications (Fontijn et al., 2014, 2016; Alloway et al., 2017a, 2017b).  
104 Here we address it by making two main contributions. First, by compiling a dataset which standardizes  
105 and integrates ~30 years of research mainly in tephrochronology (complemented with volcanological  
106 information), on 32 active volcanic centers and 132 different eruptions of the SVZ and AVZ during the  
107 last 20,000 years. Additionally, we developed an online platform<sup>1</sup> which provides user-friendly tools for  
108 the exploration of the large dataset, and helps users download subsets of it. Both contributions aim to be  
109 a step towards making these data FAIR (Wilkinson et al., 2016).

## 110 **2 Methods**

### 111 **2.1 Understanding the problem: Interviews with domain experts**

112 In order to characterize the work and needs of people using tephrochronology in the SVZ and AVZ, we  
113 conducted a series of interviews with researchers and practitioners who are part of this community,  
114 including volcanologists, tephrochronologists, palaeoclimatologists, among others. This information was  
115 then used as input to design the dataset structure and the exploration tool. Between June and August  
116 2020, we interviewed 19 people: two researchers in volcanology, one researcher in paleoceanography  
117 and a team of 16 people from the Chilean National Volcanic Vigilance Network, part of SERNAGEOMIN.  
118 However small, the interviewed sample represents to some extent the diversity of the  
119 tephrochronological community in the area. Each interview lasted between one and three hours and was

---

<sup>1</sup> <https://boom-vis.lisn.upsaclay.fr/>

BOOM! Tephrochronology dataset and exploration tool of the Southern (33–46° S) and Austral (49–55° S)  
Volcanic Zones of the Andes

120 divided in two main phases. In the first phase, we asked them about their demographic information (age,  
121 research area, position) and their daily work using tephrochronological data. More specifically, we  
122 focused on understanding what scientific questions they try to answer, which data they use to do so, how  
123 they acquire, organize and analyze these data, and what problems they commonly encounter in this  
124 process. In the second phase, we asked them to show us a concrete example from their work in order to  
125 confirm or complement the answers given during phase one, and particularly to understand how they  
126 use visualizations to organize data and answer their scientific questions. Finally, we showed them an  
127 early prototype of the exploration tool in order to get initial feedback to iterate its design. All interviews  
128 were audio recorded with the participants' consent (available in section S1 of the supplementary  
129 material). A summary of the answers from the interviews can be found in Table S1 in the supplementary  
130 material.

131 The interviews had two main results. On the one hand, we confirmed that the unFAIRness of the data was  
132 an important issue for the participants. All of them expressed how difficult and time-costly it was for  
133 them to collect, organize and analyze tephrochronological data given the dispersion of data sources and  
134 the lack of common publication standards. On the other hand, the interviews allowed us to list the  
135 information more routinely used by tephrochronologists working in the area, as well as pinpointing the  
136 problems they encounter when using these data (Table S1). Then, we designed a structure that contains  
137 this information in a way that, together with the explorer, addresses those issues. In particular, the main  
138 problems we addressed were: difficulties organizing the bulk of data, ignoring the existence of data, not  
139 understanding a priori if data from different sources are comparable or not, and problems visualizing  
140 data in its multidimensionality (e.g., visualizing at once the geographical position of the tephra deposits,  
141 their geochemical composition, stratigraphy and chronology). More detailed descriptions on the design  
142 of the BOOM! dataset and explorer, and how they address these issues, are presented in sections 3 and 4,  
143 respectively. The information from the interviews substantially aided the development of the structure  
144 of the dataset and the exploration tool.

## 145 2.2 Data collection

146 We collected tephrochronological and volcanological information on 32 active volcanic centers belonging  
147 to the SVZ and AVZ, between volcanoes Llaima (~38.7° S) and Mount Burney (~52.3° S). Here, volcanic  
148 centers are considered active if evidence of volcanic activity occurring during the last ~20,000 years has  
149 been identified. This time frame is relevant for volcanology as it captures the volcanic centers that are  
150 likely to still erupt in the future. However, it is also informed by practical limitations due to the presence  
151 of the Patagonian Ice Sheet (PIS) in the area during the last glacial period (~20,000–60,000 years Before  
152 Present; (Davies et al., 2020)). During the last glacial period, the PIS covered most of the Andes south of  
153 ~38° S, and thus, unconsolidated tephra deposits older than ~20,000 years have been largely eroded.

154 Three types of data were collected:

- 155 i. Physical characteristics and geochemical analyses (e.g., EMPA) of **pyroclastic material** (tephras  
156 themselves). A full list of the geochemical analyses included in the dataset is detailed in Table S2.
- 157 ii. Radiocarbon (<sup>14</sup>C) ages of **organic matter** *associated* with the tephras, typically recovered from  
158 within the deposits or from palaeosols immediately underlying them (used to estimate when  
159 tephras were deposited).
- 160 iii. Geochemical composition and <sup>40</sup>Ar/<sup>39</sup>Ar ages of **effusive material** (e.g., lava, breccia,  
161 prismatically joint blocks).

162 Even though this dataset is aimed to help the tephrochronological community, volcanological data (i.e.,  
163 data of effusive material) was also collected to complement the tephrochronological information.  
164 Because effusive material builds the volcanic edifice itself, there is little uncertainty regarding its volcanic  
165 source. On the contrary, tephra deposits can be identified thousands of kilometers away from any  
166 volcanic center, and often their source has not been robustly identified yet. Thus, the geochemical  
167 composition of effusive materials was included in the dataset as a geochemical “ground truth”, to which  
168 to compare the geochemical composition of tephra deposits for which the volcanic source has not been  
169 yet identified. This information is particularly useful for volcanic centers for which little  
170 tephrochronological information is available. On the other hand, <sup>40</sup>Ar/<sup>39</sup>Ar ages provide age constraints

BOOM! Tephrochronology dataset and exploration tool of the Southern (33–46° S) and Austral (49–55° S)  
Volcanic Zones of the Andes

171 when no  $^{14}\text{C}$  ages or other stratigraphic constraints are available. Errors associated to  $^{40}\text{Ar}/^{39}\text{Ar}$  ages are  
172 usually one order of magnitude larger than those of  $^{14}\text{C}$  ages and generally do not allow to differentiate  
173 between eruptions in the SVZ and AVZ. However, in recent years  $^{40}\text{Ar}/^{39}\text{Ar}$  errors are getting smaller  
174 thanks to the development of more precise analytical techniques.

175 From the publications in the literature, we prioritized those which included analyses of the geochemical  
176 composition of tephtras, together with age or stratigraphic constraints, as long as the composition  
177 reported was magmatic (in this particular geologic setting, this corresponds to  $\sim 40\text{--}80$  wt.%  $\text{SiO}_2$ ).  
178 Otherwise, and because of the data scarcity for many volcanic centers, we included all publications we  
179 could find in the literature which met these criteria. However, we always tried to include information  
180 necessary for users to evaluate the quality and comparability of the data (sections 4 and S5) and flagged  
181 the data when some issue was identified (section S5.4, Table S4). In total, we collected data from 72  
182 scientific publications published in peer-reviewed journals, five publications from SERNAGEOMIN, and  
183 two doctoral theses (the whole list of publications can be found below). These publications included  
184 works on tephrochronology (31), volcanology (28), geological mapping (6), palynology (6), glaciology  
185 (2), volcanic hazard mapping (1), archaeology (1), limnology(1), pedology (1) and paleoceanography (1).  
186 The information included in the dataset was extracted either directly from the publication tables, text  
187 and supplementary material, or alternatively through private requests to the authors when the data  
188 discussed in the publication was not readily available. The authors were informed of the purpose of the  
189 data request beforehand. From this process,  $\sim 16,800$  sample analyses were included in the dataset,  
190 identified as sourced in 132 different eruptive events, from 32 different volcanic centers. Tephra deposits  
191 described in the literature for which the volcanic source has not been identified yet were also included in  
192 the dataset, which correspond to  $\sim 11\%$  of the sample analyses.

193 Publications in peer reviewed journals:

194 Abarzúa & Moreno, 2008; Abarzúa et al., 2004; Alloway et al., 2015; Alloway et al., 2017a; Alloway et al.,  
195 2017b; Amigo et al., 2013; Bertrand et al., 2008a; Bitschene et al., 1993; Bouvet de Maisonneuve et al.,  
196 2012; Brahm et al., 2018; Bucchi et al., 2015; Carel et al., 2011; Casati et al., 2019; Clapperton et al., 1995;

197 Constantini et al., 2011; Del Carlo et al., 2018; D'Orazio et al., 2003; Fontijn et al., 2016; Futa & Stern,  
198 1988; Geoffroy et al., 2018; Gerlach et al., 1988; Haberle & Lumley, 1998; Harambour, 1988; Heusser et  
199 al., 1989 Heusser et al., 2000; Hickey-Vargas et al., 1989; Jacques et al., 2014; Kilian et al., 2003;  
200 Kratzmann et al., 2008; Lara et al., 2004; Lara et al., 2006; Lohmar et al., 2012; López-Escobar et al., 1992;  
201 López-Escobar et al., 1993; López-Escobar et al., 1995; Martínez Fontaine et al., 2021; Mayr et al., 2019;  
202 Miranda et al., 2013; Weller et al., 2014; Moreno & León, 2003; Moreno et al., 2015; Moreno, 2004;  
203 Morgado et al., 2015; Morgado et al., 2019; Motoki et al., 2006; Naranjo & Moreno, 1991; Naranjo et al.,  
204 1993; Naranjo & Stern, 2004; Naranjo & Moreno, 2005; Naranjo & Stern, 1998; Naranjo et al., 2017; Pesce  
205 & Moreno, 2014; Rawson et al., 2015; Rawson et al., 2016; Reubi et al., 2011; Schindlbeck et al., 2014;  
206 Simmons et al., 2020; Singer et al., 2008; Smith et al., 2019; Stern et al., 2011; Stern et al., 2015; Stern,  
207 1992; Stern, 2008; Tagiri et al., 1993; Villarosa et al., 2006; Wastegård et al., 2013; Watt et al., 2011a;  
208 Watt et al., 2011b; Watt et al., 2013; Weller et al., 2014; Weller et al., 2015; Weller et al., 2017; Weller et  
209 al., 2019.

210 Publications from SERNAGEOMIN :

211 Naranjo & Stern, 1998; Naranjo & Moreno, 2005; Lara et al., 2006; Moreno et al., 2010; Sellés & Moreno,  
212 2011; Bertin et al., 2018.

213 Doctoral theses: Lohmar, 2008; Mella, 2008.

### 214 **2.3 Data Visualization: Development of the BOOM! explorer**

215 The BOOM! explorer is complementary to the dataset itself and aims to provide tools for users to make  
216 the most of it. As mentioned before, we designed the explorer to address the main problems participants  
217 expressed during the interviews (section 2.1, Table S1). In turn, this helped us better understand the  
218 dataset we were collecting and modify its structure in order to better answer to community requests. For  
219 example, when looking for ways to visualize if data from different sources were comparable or not, we  
220 realized that important information to evaluate this was missing in the dataset, such as the type of  
221 register the sample corresponds to (pyroclastic material, organic matter, or effusive material) or the type  
222 of analysis performed in the sample (bulk versus micro analytical). Taking this into consideration, this

223 information was included in the dataset and users can choose which type of register and analysis to  
224 visualize in the explorer. This process led an iterative redesign of the explorer, which included: modifying  
225 the visual representation of the data, including additional User Interface interactions (e.g., filtering data  
226 by different criteria), and reorganizing the visual elements in the explorer. This iterative redesign was  
227 also supported by informal meetings with practitioners, some of them from the same group we had  
228 previously interviewed, where they told us their impressions of the tool, expressed their need for  
229 additional features, and gave us general feedback. The development of the explorer additionally helped  
230 us identify gaps in the dataset; include additional information, for example, a particular type of data of a  
231 particular volcanic center or eruptive event in order to better characterize them; as well as identify  
232 outliers and flag data. A detailed description of how the explorer can be used to explore the dataset is  
233 exemplified in section S6 by a use case.

234 Given the above process, we decided to build a light visualization tool, which can be easily installed in a  
235 local environment. In this way, other researchers can make modifications/improvements to the explorer.  
236 The BOOM! explorer is published under an MIT License<sup>2</sup>, which allows to reproduce it and build from it,  
237 under the same license as the original one. The tool was developed in Javascript, using node.js<sup>3</sup> as the  
238 back-end environment. For the visualization components, we used the Leaflet library<sup>4</sup> for the map, D3.js<sup>5</sup>  
239 for the timelines, and dc.js<sup>6</sup> with Crossfilter<sup>7</sup> for the geochemical composition scatterplots. The dataset is  
240 delivered as a CSV file (see section 3 for more details), but in order to allow fast queries in the  
241 visualization, we exported it to a SQLite<sup>8</sup> dataset that is accessed from node.js. This transformation also  
242 includes precomputed queries that help with the performance of the tool. The BOOM! explorer  
243 (<https://boom-vis.lisn.upsaclay.fr/>) is hosted by the Laboratoire Interdisciplinaire des Sciences du

---

<sup>2</sup> <https://mit-license.org/>

<sup>3</sup> <https://nodejs.org/>

<sup>4</sup> <https://leafletjs.com/>

<sup>5</sup> <https://d3js.org/>

<sup>6</sup> <https://dc-js.github.io/dc.js/>

<sup>7</sup> <https://crossfilter.github.io/crossfilter/>

<sup>8</sup> <https://www.sqlite.org>

244 Numérique, LISN (Université Paris-Saclay, CNRS & Inria), France, which guarantees its hosting and access  
245 as long as the server exists.

### 246 **3 BOOM! Dataset structure**

247 Tephrochronology identifies deposits from explosive volcanic eruptions in the stratigraphic record,  
248 samples, and characterizes them based on their physical characteristics, geochemical composition,  
249 stratigraphy and age. Then, based on this characterization and the comparison with available  
250 tephrochronological (and volcanological) information in the area, tephrochronologists ideally identify  
251 the volcanic source and the specific eruption during which the tephra was deposited, i.e., they *classify* it.  
252 The BOOM! dataset follows this structure. It is composed of a collection of descriptions and laboratory  
253 analyses of tephra deposits identified in the SVZ and AVZ (complemented with data on some lava flows).  
254 In the BOOM! dataset, each of these descriptions or analyses is called a “sample observation”, and one  
255 sample of a tephra deposit may be described by one or many sample observations. For example, in order  
256 to characterize a tephra deposit, researchers can describe the physical characteristics of the deposit as a  
257 whole (e.g., its color, grain size, and thickness) and at the same time analyze the geochemical composition  
258 of dozens of individual volcanic glass shards obtained in just one sample. Then, one sample observation  
259 corresponds to the analyzed geochemical composition of one individual glass shard (as exemplified in  
260 Figure 1). The sample as a whole, on the other hand, is described by the physical characteristics of the  
261 tephra deposit and the geochemical composition of all the analyzed glass shards. Additionally, tephra  
262 deposits are ideally not described isolated, but as part of a stratigraphic column, and their relative  
263 stratigraphic position is used to identify their volcanic source. In the BOOM! dataset, each stratigraphic  
264 column is called a “section”, and each interval of that column is called a “sub section”. In total, 1,303  
265 sections are contained in the BOOM! dataset. However, it is important to note that for 876 sections, only  
266 one tephra deposit is described, and no name is given for the section by the authors. The remaining 427  
267 sections are given a name in the original publication and are described by between one and 46 sub  
268 sections, characterized by between one and ten samples, each described by between one and 105 sample  
269 observations, corresponding to a total of 2,899 samples and 16,768 sample observations.

270 These descriptions of tephra deposits in the BOOM! dataset have been, for the most part, already  
271 classified, i.e., their volcanic source and the specific eruptive event during which they were deposited  
272 have been identified by the authors. Thus, they can be used to characterize different volcanic centers and  
273 individual eruptive events in the SVZ and AVZ and serve as a reference to compare unclassified samples  
274 to. In the BOOM! dataset, an eruptive event is characterized by all the samples (and respective sample  
275 observations) which have been correlated with that eruptive event (Figure 1). Since these samples have  
276 been collected in different locations, in different types of sedimentary archives, and provide different  
277 types of information (physical characteristics, geochemical composition, age), together they can be used  
278 to provide a more complete characterization of an eruption. For example, when did the eruption occur,  
279 what was its explosivity, the dispersion of its products, and the type of the eruption it corresponded to.  
280 In the same way, all the samples attributed to a particular volcanic center can be used to describe it. These  
281 samples may or may not be attributed to a particular eruptive event yet, in which case they are labeled  
282 as *unknown* eruptions. Nevertheless, all the samples attributed to a volcanic center can be used to  
283 describe it, though grouping them by eruption provides more information to understand their eruptive  
284 history: the recurrence of volcanic activity, its general explosivity, if temporal changes in volcanic activity  
285 have been registered, etc.

286 In practice, the BOOM! dataset is distributed as two .CSV files: a main file named BOOMDataset, and a  
287 secondary file named MeasurementRuns. In the BOOMDataset file, each sample observation corresponds  
288 to a row, which can be characterized by a maximum of 80 attributes (columns), which in turn can be  
289 grouped by the type of information they provide (Figure 1, Figure 2, Table S3):

- 290 1. **Identification attributes:** these attributes correspond to information used to identify the  
291 sample observation, but which do not further describe the tephra deposit itself. They can be  
292 subdivided into ID attributes, Position attributes, Reference attributes, and Analysis attributes.
- 293 2. **Characterization attributes:** they correspond to the actual observations (data) of the samples,  
294 which are used to characterize –and ideally fingerprint– tephra deposits. These attributes can be  
295 further subdivided into: Stratigraphy attributes (the stratigraphic position of the tephra deposit

296 and age estimate(s)); Physical characteristics attributes (color, thickness, and grain size of the  
297 tephra deposit); and Geochemical composition attributes (major elements (both *raw* and  
298 normalized), trace elements and isotope ratios compositions of either pyroclastic or effusive  
299 material).

300 3. **Interpretation attributes:** They correspond to information that is inferred based on the  
301 characterization of the tephra deposit (together with previous tephrochronological information).  
302 In particular: the volcanic source of the deposit (Volcano); the specific eruptive event during  
303 which the volcanic products (tephra or lava) were deposited (Event); and the Magnitude and  
304 Volcanic Explosivity Index (VEI) of the eruptive event. The fact that these are *interpretation*  
305 attributes is very relevant in tephrochronology. On the one hand, because tephra can be deposited  
306 far away from their volcanic source (e.g., H1 eruption from Hudson volcano has been identified  
307 ~700–900 km away from its source (Naranjo & Stern, 1998; Kilian et al., 2003; Stern, 2008; Del  
308 Carlo et al., 2018; Smith et al., 2019)), identifying which volcanic center and during which  
309 eruption a tephra was deposited is not always trivial. Additionally, the tephrochronological  
310 record is incomplete and biased towards the stratigraphic sections that are best preserved or  
311 more easily accesible. Thus, the interpretation of the eruptive event, volcanic center, magnitude  
312 and VEI may be updated as new information becomes available. In fact, the volcanic source of  
313 some of the tephra samples in the BOOM! dataset have been questioned and re-interpreted over  
314 time. For example, the volcanic source of MIC2 and COR1 tephra deposits identified by Naranjo &  
315 Stern (2004) as sourced from the Michinmahuida and Corcovado volcanic centers, have been  
316 questioned in later publications, such as Watt et al., (2009, 2013), Amigo et al. (2013) and Alloway  
317 et al. (2017a, b). In these cases, the reinterpretation of the samples is indicated by a flag (see  
318 below).

319 4. **Metadata:** A final group of attributes corresponds to Flags and Comments, which give additional  
320 information to interpret the data. On the one hand, the Flag attribute is used to communicate  
321 issues identified when including the data in the BOOM! dataset. Sample observations can be  
322 flagged for seven types of reasons (detailed in Table S4): problems with the sample ID

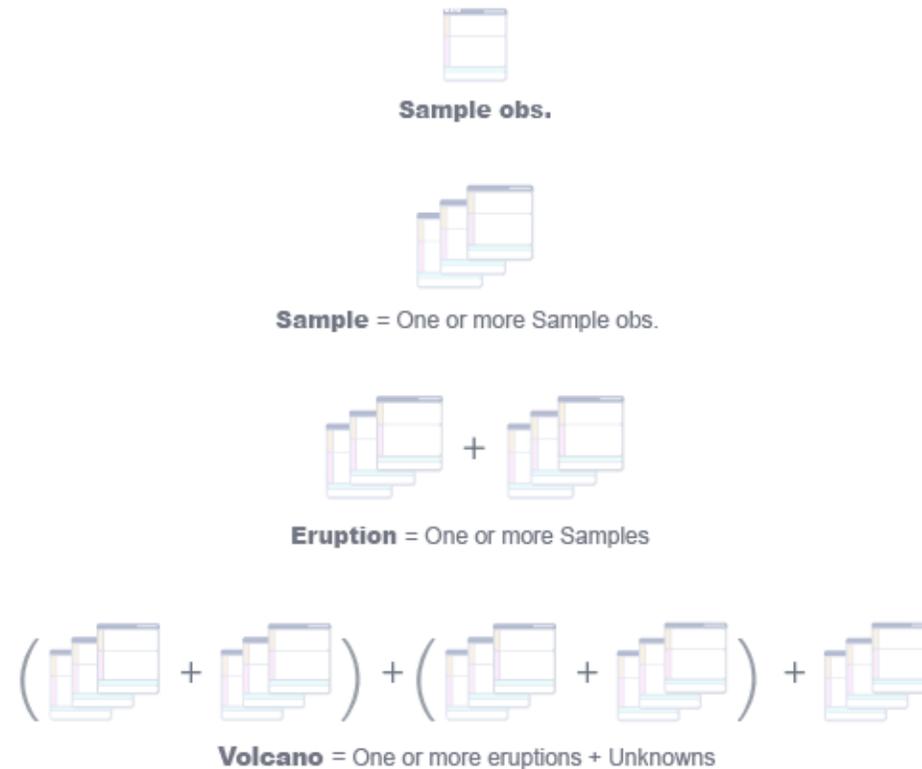
BOOM! Tephrochronology dataset and exploration tool of the Southern (33–46° S) and Austral (49–55° S)  
Volcanic Zones of the Andes

323 (SampleID\_Issue), its position (Position\_Issue), the Digital Object Identifier (DOI) of the  
324 publication where the data was obtained from (DOI\_Issue), its age (Age\_Issue), its geochemical  
325 composition (Geochemistry\_Issue), the volcanic source it is attributed to (VolcanicSource\_Issue),  
326 or the name given to the eruptive event (EventName\_Issue). In every case, the type of flag is  
327 indicated under the Flag attribute, and the particular reason why the sample observation is  
328 flagged is detailed under the Flag Description attribute. On the other hand, the Comments  
329 attribute corresponds to additional information that can help users interpret the data, which is  
330 not described by any other attribute.

331 As mentioned before, in total ~16,800 sample observations are included in the BOOM! dataset,  
332 interpreted as sourced from 32 volcanic centers and 132 different eruptive events. The amount of  
333 information related to each volcanic center varies from a total of less than 20 sample observations for  
334 volcanic centers such as Huequi, Corcovado, Yanteles, or Cay, to thousands of sample observations for  
335 volcanic centers such as Llaima, Quetrupillán, Mocho-Choshuenco, Puyehue-Cordón Caulle and Chaitén.  
336 For each volcanic center, sample observations attributed to between one (Aguilera volcano) and twenty-  
337 six (Mocho-Choshuenco) different eruptive events are included in the dataset. Because of the  
338 heterogeneity of the information included in the BOOM! dataset (the different types of sedimentary  
339 archives, sources of the information, dates of publication, and of uses of the tephrochronological  
340 information), not every sample observation is described by all of the 80 attributes previously described.  
341 In fact, in the BOOMDataset file, 58.3% of the attribute information is “missing” (Figure 2).

# BOOM! DATASET STRUCTURE

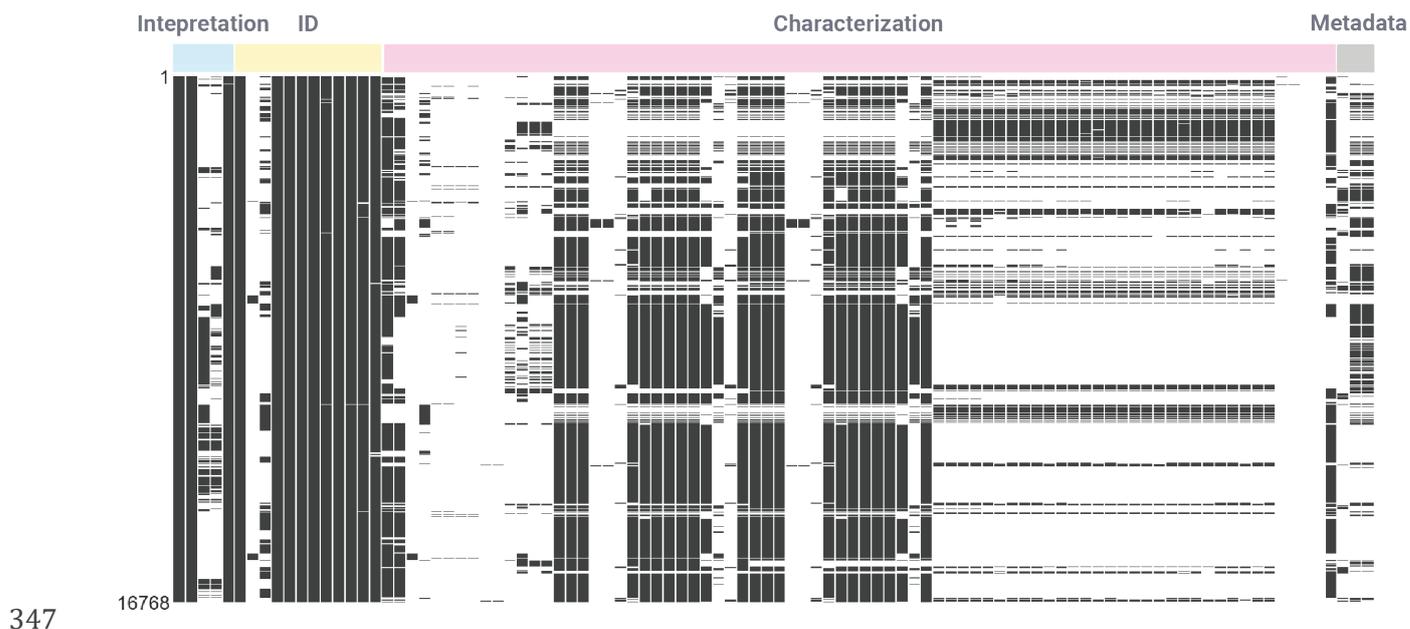
SAMPLE OBSERVATION		
IDENTIFICATION	<ul style="list-style-type: none"> <li><b>ID</b> <ul style="list-style-type: none"> <li>Sample ID: LAZ-T7A</li> <li>Sample Observation ID: 12-05-4-23</li> <li>IGSN:</li> </ul> </li> <li><b>POSITION</b> <ul style="list-style-type: none"> <li>Location: La Zeta, Esquel</li> <li>Latitude: -42.89</li> <li>Longitude: -71.36</li> </ul> </li> </ul>	<ul style="list-style-type: none"> <li><b>REFERENCE</b> <ul style="list-style-type: none"> <li>Authors: Alloway et al., 2017b</li> <li>DOI: <a href="https://doi.org/10.1002/jqs.2976">https://doi.org/10.1002/jqs.2976</a></li> </ul> </li> <li><b>ANALYSIS</b> <ul style="list-style-type: none"> <li>Type of register: Pyroclastic material</li> <li>Type of analysis: Micro analytical</li> <li>Analyzed material: Glass shards</li> <li>Analytical technique: LA-ICP-MS</li> </ul> </li> </ul>
	CHARACTERIZATION	<ul style="list-style-type: none"> <li><b>STRATIGRAPHY</b> <ul style="list-style-type: none"> <li><b>Sedimentary Archive</b> <ul style="list-style-type: none"> <li>Type of Section: Outcrop</li> <li>SectionID: La Zeta, Esquel</li> <li>SubSection ID: LAZ-T7A</li> <li>SubSection distance from top:</li> </ul> </li> <li><b>Age</b> <ul style="list-style-type: none"> <li>Historical age:</li> <li><sup>14</sup>C age:</li> <li><sup>14</sup>C laboratory code:</li> <li>Stratigraphic position:</li> <li><sup>40</sup>Ar/<sup>39</sup>Ar age:</li> </ul> </li> </ul> </li> <li><b>PHYSICAL CHARACTERISTICS</b> <ul style="list-style-type: none"> <li>Color: pale to dark grey</li> <li>Thickness: 8 cm</li> <li>Grain size min: fine ash</li> <li>Grain size max: medium ash</li> </ul> </li> </ul>
<b>INTERPRETATION</b>		
<ul style="list-style-type: none"> <li>Volcano: Michinmahuida</li> <li>Event: Lepué</li> <li>Magnitude:</li> <li>VEI:</li> </ul>		



342

343 **Figure 1** BOOM! dataset structure. To the left are shown the attributes in the BOOMDataset and MeasurementRuns files, grouped by the type of  
 344 information they provide, as described in the text. Attributes from the MeasurementRuns file correspond to the information below “Quality of analysis”.  
 345 The metadata attributes are not illustrated. The information shown corresponds to sample LAZ-T7A (Alloway et al. 2017b). The full description of the  
 346 attributes and units is described in Table S3.

BOOM! Tephrochronology dataset and exploration tool of the Southern (33–46° S) and Austral (49–55° S) Volcanic Zones of the Andes

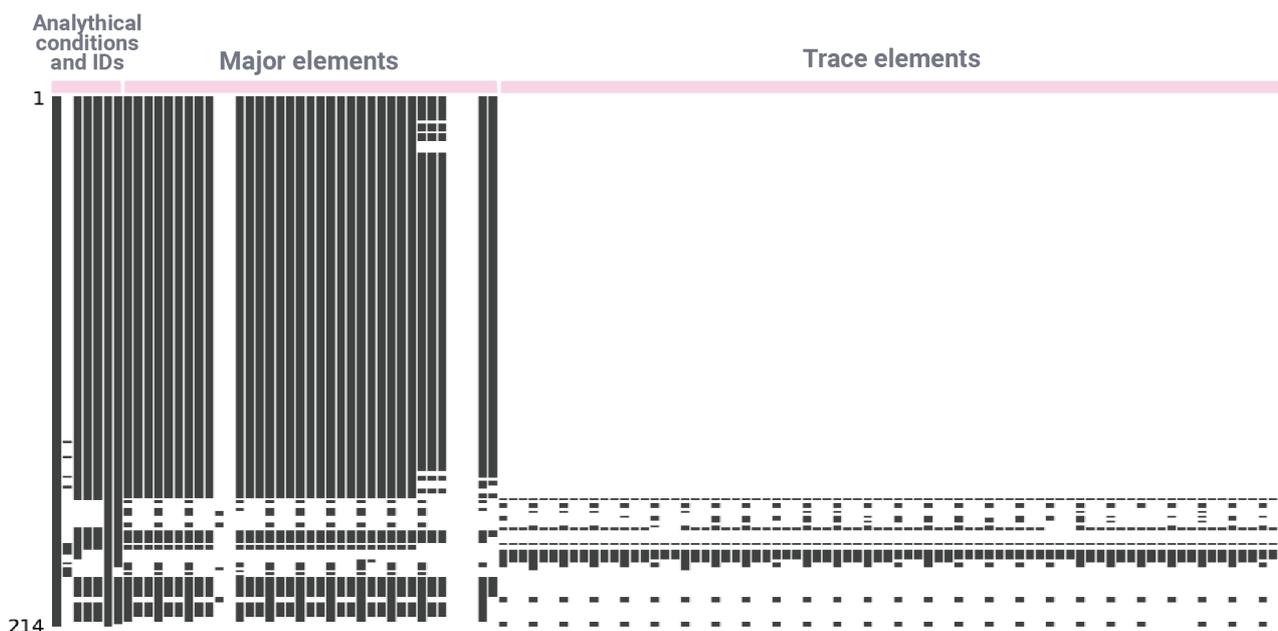


347

348 **Figure 2** Representation of the “missing” values in the BOOMDataset file. Each column represents an  
349 attribute and each row a sample observation. The attributes are grouped according to the type of  
350 information they provide: Interpretation attributes, ID attributes, Characterization attributes, and  
351 Metadata. Dark grey cells represent filled values and white cells represent “missing” values. The figure  
352 was created using the Python library Missigno<sup>9</sup>.

353 In addition to the main BOOMDataset file, a second CSV file named MeasurementRuns is also distributed  
354 (Figure 3). This file contains information which is part of the Geochemical composition attributes  
355 previously described (attributes under “Quality of analysis” in Figure 1). It corresponds to information  
356 used by geochemists to evaluate the quality and comparability of the geochemical analyses performed by  
357 different laboratories. We decided to store this information in a different file because of the way the  
358 geochemical composition of most of the samples in the BOOM! dataset were analyzed. Most of the data in  
359 the dataset corresponds to analyses of the geochemical composition of individual volcanic glass shards  
360 (~88% of the sample observations in the BOOMDataset file), which were analyzed either by EMPA (for  
361 major elements) or by Laser Ablation - Inductively Coupled Plasma - Mass Spectrometry LA-ICP-MS, for  
362 trace elements. Individual glass shards are analyzed in “batches”, and during one batch (called a  
363 “measurement run” in the BOOM! dataset), dozens or hundreds of individual glass shards from different  
364 samples can be analyzed. To evaluate the accuracy and precision of the analyses performed during one

<sup>9</sup> <https://github.com/ResidentMario/missingno>



365

366 **Figure 3** Representation of the “missing” values in the MeasurementRuns file. Dark grey cells represent  
367 filled values and white cells represent missing values. The figure was created using the Python library  
368 Missigno.

369 measurement run, geochemists repeatedly analyze the composition of different secondary standards  
370 along with the glass shards to be studied. These secondary standards ideally have certified reference  
371 values (e.g., Jochum et al., 2006, 2011, 2016) which can be compared with the analyzed values during  
372 each measurement run to estimate their accuracy and precision. In particular, by comparing the mean  
373 value of each analyzed element of each secondary standard to the certified values, the accuracy of the  
374 analysis of each element during that particular measurement run can be estimated (more details on how  
375 to do this with the BOOM! dataset is explained in section 4.2.2). In a similar way, the precision can be  
376 estimated as the standard deviation of the analyzed composition of each element on each secondary  
377 standard analyzed during a particular measurement run. Taking this into consideration, in order to  
378 provide information for other researchers to evaluate the comparability of their geochemical analyses,  
379 researchers generally publish: the label given by them to a particular measurement run; which samples  
380 and sample observations were analyzed during that run; which secondary standards were analyzed; the  
381 number of analyses of each secondary standard in that run; the mean value and the standard deviation  
382 of each analyzed element of each secondary standard; as well as the analytical totals if major elements

383 were analyzed. In the MeasurementRuns file, each row corresponds to the information of one secondary  
384 standard during a particular run, thus one run can correspond to several rows. The reason for analyzing  
385 different secondary standards during one measurement run is that they have different geochemical  
386 compositions (e.g., GOR132-G (~46 wt.% SiO<sub>2</sub>), StHs6/80-G (~63.7 wt.% SiO<sub>2</sub>), ATHO-G (~75.6 wt.%  
387 SiO<sub>2</sub>) (Jochum et al., 2006)), and the performance of an analytical technique is associated with that  
388 composition. Taking the latter into consideration, to assess the quality of the analyses, geochemists  
389 usually consider the accuracy and precision of the secondary standard or multiple secondary standards  
390 with a geochemical composition relatively similar to that of the analyzed sample. In addition, in order to  
391 evaluate the comparability of EMPA analyses, the analytical conditions during each measurement run are  
392 also considered (e.g., beam size, accelerating voltage, and beam current). Thus, this information is also  
393 stored in the MeasurementRuns file. In both the MeasurementRuns and BoomDataset files, each batch  
394 analysis label is indicated under the measurement run attribute, which links both files. The  
395 MeasurementRuns file contains 81 different measurement runs where between one and seven secondary  
396 standards were analyzed. The measurement runs described in the MeasurementRuns file corresponds to  
397 10,891 of the sample observations in the BOOMDataset, attributed to 26 volcanic centers and 61 eruptive  
398 events. In the same way as with the BOOMDataset file, because of the heterogeneity of the data, ~69% of  
399 the cells in the MeasurementRuns file are “missing” (Figure 3).

400 The two files of the dataset are hosted on the ESPRI server of the IPSL, France, which guarantees its  
401 hosting and access as long as the server exists. Both sets can be downloaded directly from the IPSL  
402 catalog<sup>10</sup> as a Web Map service (WMS), Web Feature service (WFS) or as a CSV file; or using the BOOM!  
403 explorer to download subsets of it (see section S6 for more details).

---

<sup>10</sup><https://doi.org/10.14768/47b4525f-ff39-4940-a963-4d2673f2362e>

## 404 **4 Data comparability**

405 As already mentioned, the BOOM! dataset integrates tephrochronological (and volcanological)  
406 information extracted from 79 different publications. Even though all these publications provide relevant  
407 information for tephrochronology, they can be very heterogeneous: they correspond to different types of  
408 publications (scientific publications, government reports, and doctoral theses); from different disciplines  
409 (tephrochronology, volcanology, archaeology, limnology, geological mapping, palynology, glaciology,  
410 pedology, and paleoceanography); which were published through 30 years of research (1988-2021 C.E.);  
411 and which additionally can have very different research goals (e.g., produce volcanic hazard maps, do  
412 research on volcanology and tephrochronology, and improve chronologies of sedimentary archives for  
413 paleoclimate, palynology or paleoceanography studies). The idea behind collecting the BOOM! dataset, is  
414 that all this information together can be used to better describe eruptive events and volcanic centers in  
415 the SVZ and AVZ and help disentangle the tephrochronological record there. However, because of the  
416 heterogeneities in the source of the information, different publications might study the same  
417 phenomenon in a slightly different way. This does not necessarily mean that the information cannot be  
418 integrated, but it is necessary to understand what the information actually represents in order to  
419 correctly interpret it. Taking this into account, several of the attributes included in the dataset,  
420 correspond to information for the user to evaluate the comparability of the data. In the following we focus  
421 on attributes associated with age estimates and geochemical composition comparability. For further  
422 description of other relevant attributes, such as type of register, type of section, and flags, the reader is  
423 referred to the supplementary material (section S5).

### 424 **4.1 Age estimate comparability**

425 A very important part of fingerprinting tephra deposited during different eruptive events in a very active  
426 volcanic region, such as the SVZ and AVZ, is estimating the age of the eruption. In the timeframe  
427 comprehended in the dataset (last 20,000 years), the age of a tephra is generally estimated by <sup>14</sup>C dating  
428 organic matter which has been identified *associated* to the tephra deposits. This is done because the  
429 pyroclastic material itself generally cannot be directly dated with enough precision to differentiate

BOOM! Tephrochronology dataset and exploration tool of the Southern (33–46° S) and Austral (49–55° S)  
Volcanic Zones of the Andes

430 between eruptions (for example, by the  $^{40}\text{Ar}/^{39}\text{Ar}$  method, for which errors are generally on the order of  
431 thousands of years, whereas eruptions can occur every as little as tens of years (e.g., Singer et al., 2008)).  
432 Because of this, researchers try to identify organic matter which is ideally embedded *within* the tephra  
433 deposit itself, and alternatively immediately above or below it. This organic matter can be  $^{14}\text{C}$  dated with  
434 a precision of tens of years, and in that way, provide a more precise reference for when the eruption  
435 occurred. Both the stratigraphic position of the organic matter with respect to the tephra (above, within,  
436 or below), and the specific material that has been dated, are important aspects to consider when  
437 evaluating the comparability of the different age estimates. For instance, if a piece of charcoal is found  
438 within a tephra deposit, it is assumed that the charcoal is the result of hot pyroclastic material burning  
439 living trees during an explosive eruption. Thus, by dating the charcoal, a more or less precise age of when  
440 the eruptive event happened, can be obtained. Alternatively, if no organic matter is found within the  
441 tephra deposit itself, soil that has formed either above or below the deposit, might be dated.  
442 Unfortunately, the  $^{14}\text{C}$  age of the soil might not represent the *true* age of the tephra deposit. Soil is formed,  
443 among others, by the degradation of organic matter, and during its formation it incorporates organic  
444 matter that is contemporary to the soil formation process, but it can also incorporate older organic matter  
445 that was already present when soil formed; as well as younger organic matter which can infiltrate from  
446 above, for example through a permeable tephra layer. Thus, soil  $^{14}\text{C}$  ages or bulk organic sediment  $^{14}\text{C}$   
447 ages, correspond to a mean value of organic matter of different ages. Additionally, soil can be found either  
448 below or above the tephra deposit, and thus, it might have formed long before the eruption, or it might  
449 have taken years to develop after the tephra deposition. In these cases, the  $^{14}\text{C}$  ages from the soil are  
450 generally interpreted as maximum and minimum ages of the tephra deposit, respectively. In a similar  
451 way, if organic macro remnants, such as charcoal or wood, are found below or above the tephra deposit  
452 and dated, those ages are also considered as maximum and minimum ages. Because of the latter, a good  
453 practice, which can help reduce the age estimate uncertainties is replicates, i.e., dating more than once a  
454 determined sample. In the BOOM! dataset, however, most  $^{14}\text{C}$  ages have not been replicated, which can  
455 be associated with increased expenses, but also with material availability.

456 Another important aspect to consider when estimating the age of eruptive events, is the type of  
457 sedimentary archive where the tephra -and associated organic matter- were identified. Depending on the  
458 depositional environment, different types of organic matter for  $^{14}\text{C}$  dating will be available. For example,  
459 in the ocean, generally wood or charcoal are not identified and planktic foraminifera are dated instead to  
460 provide an age estimate of tephtras. However, these ages cannot be directly compared with on land ages,  
461 as the concentration of  $^{14}\text{C}$  in the ocean and the atmosphere are not in equilibrium. Because of this, a  
462 correction must be applied to marine ages in order to compare them with atmospheric ages. This  
463 correction changes in time and space, and is a subject of ongoing research (e.g., Siani et al., 2013; Merino-  
464 Campos et al., 2019). A similar situation occurs with  $^{14}\text{C}$  ages in lake sediment cores. In this case, ideally  
465 terrestrial macrofossils found within the lake sediment cores are dated, which represent the atmospheric  
466  $^{14}\text{C}$  age, as long as the sediment has not been reworked. In many cases, however, these are not available  
467 and bulk sediments are date instead. Because of different processes occurring in lakes, bulk sediment  $^{14}\text{C}$   
468 content is generally not in equilibrium with the atmosphere. Among these processes: terrestrial input of  
469 *old* organic matter by river inflow, groundwater  $^{14}\text{C}$  content, dissolution of aged carbonate or variations  
470 in biological activity in the lake (e.g., Geyh et al., 1998; Yu et al., 2007). Because of this, the difference  
471 between the contemporary atmosphere and lacustrine sediment  $^{14}\text{C}$  age can change both in time and  
472 space and is characteristic of each lake. Thus, marine  $^{14}\text{C}$  ages and bulk sediment  $^{14}\text{C}$  ages in lake sediment  
473 cores, provide a loose reference of the time of deposition of tephra deposits, but should in no case be  
474 directly compared with on land ages without a proper correction.

475 Taking the latter into consideration, in order to correctly interpret the  $^{14}\text{C}$  ages in the BOOM! dataset in  
476 terms of time of deposition of tephtras, in addition to the  $^{14}\text{C}$  age and respective analytical error, users  
477 should also consider the analyzed material (e.g., charcoal, wood, soil, organic macro remnants, planktic  
478 foraminifera), the type of analysis (micro analytical (e.g., charcoal, wood) or bulk (e.g., soil, bulk  
479 sediment)), the stratigraphic position of the analyzed material (e.g., X cm above or below the tephra), the  
480 type of section (e.g., marine core, lake core, outcrop), and whether or not replicates for the sample have  
481 been published.

## 482 **4.2 Geochemical composition comparability**

483 An important aspect when fingerprinting tephras is assessing their geochemical composition. Because  
484 many volcanic centers and eruptive events have distinct compositions, by analyzing the geochemical  
485 composition of tephras, researchers can discriminate between potential sources. To do this,  
486 tephrochronologists most often analyze the major element composition of volcanic products (~82% of  
487 the sample observations in the BOOM! dataset). However, especially in the SVZ, many volcanic centers  
488 and different eruptive events have overlapping major element compositions, thus additional information  
489 is needed to discriminate the volcanic source of tephras. Because of the latter, in some cases trace  
490 elements are also analyzed (~23% of the sample observations in the BOOM! dataset), and less frequently,  
491 isotopic ratios (for example  $^{87}\text{Sr}/^{86}\text{Sr}$  and  $^{143}\text{Nd}/^{144}\text{Nd}$ ), which provide a good discrimination tool,  
492 however are less routinely analyzed (0.1% of the sample observations in the BOOM! dataset). When  
493 comparing the geochemical composition of different volcanic products, there are several aspects to  
494 consider to understand if data from different publications are comparable or not. In the first place, it is  
495 important to understand what kind of volcanic product (pyroclastic or effusive material), and particular  
496 material (e.g., individual glass shards, bulk tephra, melt inclusions, etc.) has been analyzed. Different  
497 materials can represent different processes associated with eruptive activity, and as such, can have  
498 different geochemical compositions, even if they correspond to the same eruptive event (for more details,  
499 see section 4.2.1). Secondly, it is important to check if the data interpreted meet community quality  
500 criteria and if not, filter the data accordingly (section 4.2.2). Finally, when comparing the major element  
501 composition of different sample observations, it is important that the major element compositions are  
502 normalized to a volatile-free composition in order to compare them. Additionally, as different major  
503 elements are analyzed in different publications (for example  $\text{FeO}_T$  versus  $\text{Fe}_2\text{O}_{3T}$ ), it is also important to  
504 check that the normalizations are comparable (section 4.2.3). In the following sections we describe these  
505 three aspects in detail.

### 506 **4.2.1 Analyzed volcanic products**

507 Pyroclastic and effusive material represent different expressions of eruptive activity, and because of that,  
508 even if they correspond to the same eruptive event, they will not necessarily have the same geochemical

509 composition. In addition, different materials can be analyzed within the same type of register, which  
510 should also be taken into consideration when interpreting the data. For example, in a pyroclastic deposit,  
511 generally individual glass shards or crystals are analyzed, which represent different magmatic processes  
512 and thus they will have very different geochemical compositions, even though they correspond to the  
513 same tephra deposit (e.g., Morgado et al., 2015). In the BOOM! dataset, the analyzed materials were  
514 entered as they were presented in the publications not to lose the raw information and as a way to make  
515 explicit the heterogeneity on how data is published. Analyzed materials within pyroclastic deposits  
516 include: tephra, bulk tephra, pumice, bulk pumice, scoria, bulk scoria, lapilli, accretionary lapilli,  
517 individual accretionary lapilli, bulk ash, bulk glass, glass shards, matrix glass, melt inclusions, closed melt  
518 inclusions, open melt inclusions, and juvenile lithics; whereas for effusive material, analyzed material  
519 comprehend: lava, prismatically jointed blocks (PJB), and breccia.

520 Another important aspect to note when using the geochemical composition of volcanic products for  
521 fingerprinting tephtras, is what in the BOOM! dataset is called the “type of analysis”. Pyroclastic deposits  
522 and effusive material can be analyzed either micro-analytically (e.g., individual glass shards or minerals)  
523 or instead, the sample can be crushed and analyzed as a whole. The latter case is called a “bulk” analysis,  
524 and it represents a mean value of minerals, glass shards and even lithics, and most often they display  
525 narrow trends in bivariate plots. When available, micro-analytical analyses are preferred to bulk analyses  
526 because the glass (melt) composition might be more sensitive to small amounts of fractional  
527 crystallization and thus may reflect a more variable geochemical composition. Thus, providing more  
528 options for fingerprinting tephtras, for example, of different eruptions from the same volcanic source  
529 (Lowe, 2011). When bulk analyses are performed, on the other hand, the composition of different  
530 eruptive events and even different volcanic centers can be very similar and does not always allow to  
531 discriminate the source of volcanic products.

532 In order to correctly interpret the geochemical composition of eruptive events and volcanic centers in  
533 the BOOM! dataset, the user should take into consideration: the type of register (effusive material,  
534 pyroclastic material), type of analysis (micro-analytical, bulk) and the analyzed material (e.g., lava,

535 pumice, wood). These qualitative attributes can help users understand if the geochemical compositions  
536 they are interpreting represent similar processes or not, and thus if they can be directly compared.

#### 537 **4.2.2 Quality control of analyses**

538 Generally, when researchers obtain the results from the geochemical analyses they perform (but also  
539 when collecting information to compare their samples with), they check if the data meets some  
540 community criteria, to understand if the analyses are *good* analyses. Among these: if the analysis was  
541 correctly performed, if what is said to be analyzed was in fact analyzed, and if the accuracy and precision  
542 of the analysis is within an acceptable range. In most cases, researchers will publish only the data that  
543 meet these criteria. However, different researchers, disciplines, or fields of application, can follow slightly  
544 different criteria. Thus, it is important for the users to have access to the information necessary for them  
545 to evaluate the latter, so they can filter data according to their own criteria. In the BOOM! dataset, we  
546 have included several attributes for users to evaluate this, which we describe in the following.

547 A first order quality control of analyses of geochemical composition is a check of the analytical conditions  
548 in which they were performed. Depending on the technique employed, this information is more or less  
549 routinely communicated. Because most of the BOOM! dataset corresponds to EMPA analyses of glass  
550 shards (~76% of sample observations), we have included only the analytical conditions for analyses  
551 performed by EMPA, which correspond to beam size, accelerating voltage, beam current of a particular  
552 measurement run, stored in the MeasurementRuns file. Common values are described in Table S3.

553 An additional important aspect to evaluate when assessing the quality of a geochemical analysis,  
554 particularly of major element analyses, is that the analytical total is within accepted values. Major (and  
555 minor) elements correspond to the geochemical elements found more abundantly in igneous rocks, and  
556 their abundances are typically expressed in weight percentage (wt.%) of oxides: SiO<sub>2</sub>, TiO<sub>2</sub>, Al<sub>2</sub>O<sub>3</sub>, FeO  
557 and/or Fe<sub>2</sub>O<sub>3</sub>, MnO, MgO, CaO, Na<sub>2</sub>O, K<sub>2</sub>O, P<sub>2</sub>O<sub>5</sub>. Additionally, in some cases Cl and F are also analyzed and  
558 considered with minor elements. The sum of the analyzed major (and minor) elements composition is  
559 called the “analytical total” (“Total” in the BOOM! dataset), and typically is between 90–100 wt.%. The  
560 difference between the analytical total and a 100 wt.% can derive from different situations, among them:

561 analytical error (including due to surface roughness of the samples, poor sample positioning, and loss of  
562 alkalis during the analysis), from post-depositional hydration (e.g., alteration of the volcanic products),  
563 or from the presence of non-degassed volatiles in the groundmass of the volcanic products. For a more  
564 detailed discussion on *acceptable* analytical totals depending on the volatile content of the volcanic  
565 products, the reader is referred to Pearce et al. (2008). The value of the analytical total that is considered  
566 as a *good* analysis also depends on the analytical technique employed, which is associated with the type  
567 of analysis performed (bulk or micro analytical). On the one hand, when bulk samples are analyzed, for  
568 example, by X-ray fluorescence (XRF), Atomic Absorption Spectroscopy (AAS), or Inductively Coupled  
569 Plasma Emission Spectrometry (ICP-AES), in addition to the major element composition, a value called  
570 Loss on ignition (LOI) is sometimes estimated. The LOI corresponds to the volatile content of the sample,  
571 for example H<sub>2</sub>O, CO<sub>2</sub>, SO<sub>2</sub>, which can derive from different degrees of post-depositional hydration or syn-  
572 eruptive characteristics of the volcanic products. In this case, the analytical total plus the LOI should  
573 approach 100 ± 2 wt.%, otherwise it is considered a *bad* analysis. In the dataset, ~1.9% of the sample  
574 observations corresponding to bulk analyses of major elements of pyroclastic or effusive material, are  
575 outside this range, which corresponds to ~0.2 % of the sample observations in the whole BOOM! dataset.  
576 On the other hand, when glass shards are analyzed by EMPA, the volatile content is usually not directly  
577 analyzed. In this case, analytical totals lower than 100 ± 2 wt.% do not necessarily reflect poor quality  
578 analyses or alteration of the samples, but more often reflect characteristics of the volcanic deposit and  
579 environment, which are not accounted for by the major elements alone. Because of the latter, there is no  
580 consensus regarding acceptable totals for EMPA glass data, and thus, in the dataset this value is stored  
581 when presented in the literature (~86% of the sample observations analyzed for major elements in the  
582 BOOM! dataset have been published along with their analytical totals) for the user to judge for  
583 themselves. For ~95% of the sample observations analyzed for major elements with EMPA, the analytical  
584 total is between 95 and 102 wt%. Samples with analytical totals lower than 90 wt% were not included in  
585 the dataset.

586 Another important quality control researchers perform is making sure that what was intended to be  
587 analyzed, actually was. To do this, they generally check that the geochemical composition is within

588 plausible ranges. In the case of bulk analyses, this means that the composition is magmatic, i.e., that SiO<sub>2</sub>  
589 is between ~40–80 wt.%. In the case of micro-analytical analyses, it depends on what it is analyzed,  
590 generally glass or mineral crystals. Sometimes, the groundmass in pumice, scoria, individual glass shards,  
591 etc., can be very microlite-rich, particularly for relatively mafic samples (e.g., of basaltic andesite  
592 composition) and so it can be difficult to analyze the glass phase using a defocused beam by EMPA or LA-  
593 ICP-MS; on the other hand, when crystals are analyzed, if they are very small (<20 µm) it might be difficult  
594 to analyze their composition at specific locations (e.g., core vs. rim). In order to check if the analyses were  
595 performed on the desired material, researchers check the composition of major elements and evaluate  
596 whether it is feasible for the expected phase (e.g., glass vs. a mineral phase). In the BOOM! dataset, data  
597 was included as long as it was magmatic and indicated as either bulk tephra or individual glass shards in  
598 the original publication. The user can choose to filter the information based on their own criteria.

599 One of the most robust ways of reducing the uncertainties in correlations based on the geochemical  
600 composition of tephras is side-by-side analysis. This means, analyzing in the same laboratory and with  
601 the same methodology unknown and reference tephras (Lowe et al., 2017). In the BOOM! dataset, this  
602 has strictly been done in only publication (Smith et al., 2019). Smith et al. (2019) reanalyzed reference  
603 tephras previously identified on land by Stern (1992, 2008) and by Weller et al. (2015), to check for  
604 potential correlations with the tephras identified by them in a lake sediment core. On the contrary, most  
605 publications a side-by-side analysis is not done. However, in many publications unknown tephras  
606 identified in several sections are analyzed which allow authors to robustly correlate them and  
607 characterize eruptions methodologies (e.g., Rawson et al., 2015; Fontijn et al., 2016; Alloway et al., 2017a,  
608 b). These analyses are done in the same laboratory and following the same methodologies, reducing the  
609 uncertainty in their correlation. In general though, samples which have been analyzed in previous  
610 publications are not reanalyzed, probably because of the increase in expenses or complications obtaining  
611 the samples. Alternatively, one important way of evaluating if analyses performed in different  
612 laboratories are comparable, as well as of evaluating their accuracy and precision (see below), is  
613 analyzing secondary standards along with the unknown samples and reporting the results.

BOOM! Tephrochronology dataset and exploration tool of the Southern (33–46° S) and Austral (49–55° S)  
Volcanic Zones of the Andes

614 An important evaluation to make to correctly interpret the available geochemical information, is  
615 assessing the accuracy and precision of the analyses. In the case of isotope ratios included in the BOOM!  
616 dataset ( $^{87}\text{Sr}/^{86}\text{Sr}$  and  $^{143}\text{Nd}/^{144}\text{Nd}$ ), an analytical error (generally  $2\sigma$ ) is directly provided by the  
617 laboratories and is included in the dataset, which allows users to understand the analytical precision of  
618 that data. In the case of major and trace elements, however, assessing this is less straightforward. The  
619 most appropriate way to assess the accuracy and precision of these analyses, is to consider the  
620 composition of the secondary standards analyzed along with the samples, as described in section 3. From  
621 the total of sample observations analyzed for major and trace elements, for ~64% (10,438 of the sample  
622 observations), it was possible to obtain the full secondary standard information (i.e., name of the  
623 standards analyzed in each measurement run, number of analyses performed during each measurement  
624 run, analyzed mean value and standard deviation for each analyzed element for each standard). In ~2.8%  
625 of the cases (453 sample observations) only the mean analyzed values were published. Thus, for ~33.2%  
626 of the sample observations analyzed it is not possible to evaluate the accuracy and precision of the  
627 analyses. Additionally, from the 33 secondary standards analyzed in the publications included in the  
628 BOOM! dataset, certified values are available for only 19 of them, which corresponds to ~55% of the  
629 sample observations, from which ~53% has the full information (mean, standard deviation and n). Eight  
630 of the certified secondary standards correspond to glass: GOR128-G, GOR132-G, KL2-G, ML3B-G, T1-G,  
631 StHs6/80-G, ATHO-G (Jochum et al., 2006), NIST SRM 610 (Jochum et al., 2011), and eleven to whole rock  
632 powders AGV-2, BCR-2, BHVO-1, BHVO-2, BIR-1, JA-1, JA-2, W-2 (Jochum et al., 2016), S-Y<sup>11</sup>, OREAS184<sup>12</sup>,  
633 OREAS700<sup>13</sup>. For the whole rock analyses included in the dataset, only the mean values were published.

634 In order to provide a broad idea of the accuracy and precision of the analyses included in the BOOM!  
635 dataset, we have estimated the accuracy and precision of the analyses for which the authors provided the  
636 full information, and the certified values are available online. The accuracy is estimated as the mean  
637 analyzed value versus the certified value of each element analyzed on each secondary standard during

---

<sup>11</sup> Certified values available at: <https://www.nrcan.gc.ca/our-natural-resources/minerals-mining/mining-resources/sy-4-diorite-gneiss/8025>

<sup>12</sup> Certified values available at: <https://www.oreas.com/crm/oreas-700/>

<sup>13</sup> Certified values available here: <https://www.oreas.com/crm/oreas-184/>

638 each measurement run, and the precision as the relative standard deviation (RSD) of the  $n$  analyzed  
639 values of each element on each secondary standard during each measurement run (RSD =  
640  $(1SD/\text{mean}) \cdot 100$ ). Figure 4a and b show the results of these estimations, grouped by secondary standard.  
641 Jupyter notebooks written in Python are available in the github repository<sup>14</sup> of the project for users to  
642 explore the overall dataset accuracy and precision, as well as for each measurement run or sample.  
643 Overall, the highest accuracies and precisions are observed for  $\text{SiO}_2$ , which range between  $\sim 0.96$  and  
644  $\sim 1.02$ , and  $\sim 0.09\%$  and  $1.54\%$ , respectively; and the lowest accuracies and precisions are observed for  
645  $\text{P}_2\text{O}_5$ , which range between  $\sim 0.16$  and  $\sim 2.00$ , and between  $3\%$  and  $206\%$ , respectively.

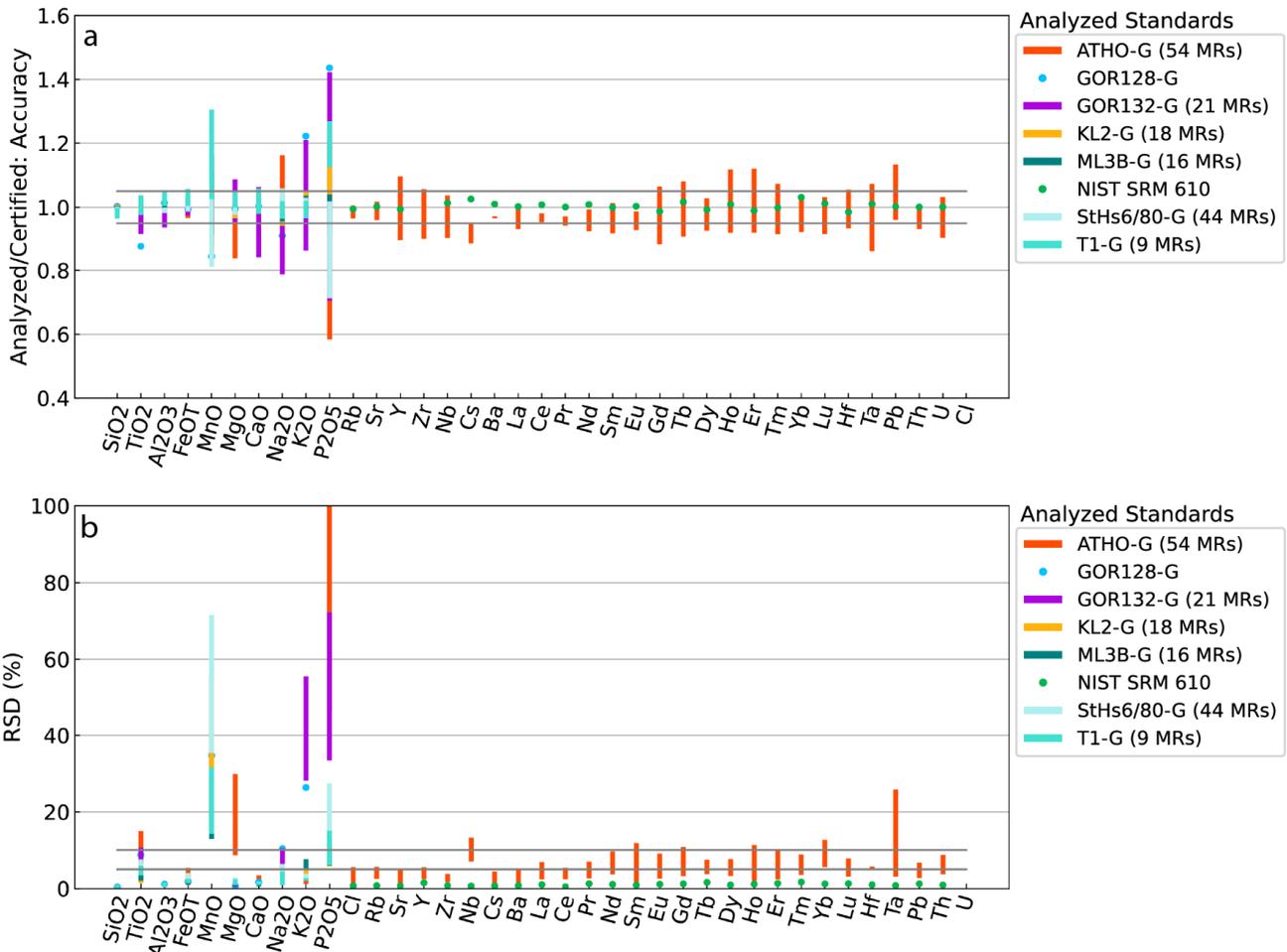
646 As mentioned in section 3, more than one secondary standard is generally analyzed by run, which are  
647 intended to represent the geochemical composition of the studied samples. Because of that, more than  
648 one value for accuracy and precision for each analyzed element is associated to each batch (for example  
649 as in Figure 5a and b). In this case, researchers will generally evaluate the accuracy and precision of a  
650 sample observation by comparing it with the secondary standard(s) with the geochemical composition  
651 closest to that of the unknown sample. Thus, the precise analytical error is not calculated, rather it is  
652 generally estimated as the highest value of precision and accuracy for each element. For example, along  
653 with sample CLD025A ( $\sim 55\text{--}57$  wt.%  $\text{SiO}_2$ ), analyzed during measurement run 11112 by Fontijn et al.  
654 (2016), secondary standards GOR132-G (45.5 wt.%  $\text{SiO}_2$ ), KL2-G (50.3 wt.%  $\text{SiO}_2$ ), T1-G (58.6 wt.%  $\text{SiO}_2$ ),  
655 StHs6/80-G (63.7 wt.%  $\text{SiO}_2$ ) and ATHO-G (75.6 wt.%  $\text{SiO}_2$ ) were analyzed. In this case, standards KL2-G  
656 and T1-G have the composition closest to the sample. During that run, the analysis of  $\text{TiO}_2$ , for example,  
657 is estimated as  $\sim 5\%$  for standard KL2-G, and  $\sim 3\%$  for standard T1-G, whereas precisions are  $2\%$  and  $5\%$ ,  
658 respectively. Thus, the estimated accuracy and precision for the analysis of  $\text{TiO}_2$  are both  $\sim 5\%$ .

659 Users should take into consideration all of the above when interpreting the data and the dataset, and filter  
660 regarding their own criteria.

---

<sup>14</sup> <https://github.com/consuelola/BOOM/Notebooks>

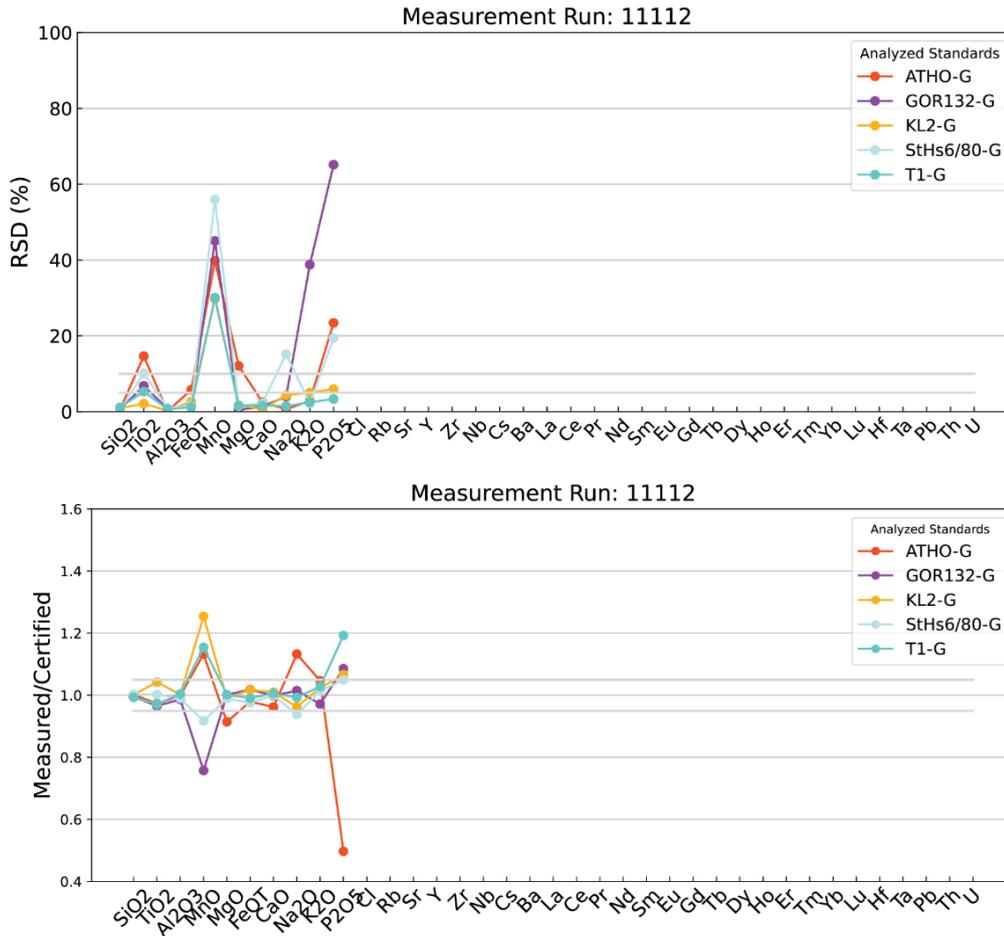
BOOM! Tephrochronology dataset and exploration tool of the Southern (33–46° S) and Austral (49–55° S) Volcanic Zones of the Andes



661

662 **Figure 4.** Accuracies (a) and precisions (b) of the geochemical composition analyses included in the  
 663 BOOM! dataset, estimated as described in the text. When the same secondary standard was analyzed in  
 664 more than three measurement runs, the 1σ range are plotted as a vertical line and the number of  
 665 measurement runs considered is indicated in the legend as #MRs. When a secondary standard was  
 666 analyzed during three or less measurement runs, accuracy and precision are plotted as a dot. Bold grey  
 667 horizontal lines correspond to 0.95 and 1.05 values for the accuracy and 5 and 10% for the precision.  
 668 Only secondary standards with certified values are shown, analyzed during measurement runs for which  
 669 the name of the secondary standard for which the full information is published (as described in the text):  
 670 ATHO-G, GOR128-G, GOR132-G, KL2-G, ML3B-G, T1-G, StHs6/80-G, (Jochum et al., 2006), NIST SRM 610  
 671 (Jochum et al., 2011).

BOOM! Tephrochronology dataset and exploration tool of the Southern (33–46° S) and Austral (49–55° S) Volcanic Zones of the Andes



672 **Figure 5** Accuracies and precisions of the geochemical composition analyses of each secondary standard  
 673 analyzed during measurement run 11112 (Fontijn et al., 2016), as an example. a: Accuracy for major  
 674 elements, calculated as mean analyzed value of each element of each secondary standard analyzed during  
 675 measurement run 11112, versus the certified values (Jochum et al., 2006). b: Precision of the geochemical  
 676 analyses, calculated as mean analyzed value of each element of each secondary standard analyzed during  
 677 measurement run 11112, divided by the standard deviation of the n analyses during that measurement  
 678 run. Bold grey horizontal lines correspond to 0.95 and 1.05 values for the accuracy and 5 and 10% for the  
 679 precision.

680 **4.2.3 Major element normalization**

681 As already mentioned, the major element composition of volcanic products is expressed in wt.%, and  
 682 ideally the sum of all the major oxides and LOI (when analyzed) should approach 100 wt.%. However,  
 683 because of post-depositional alteration, variable volatile contents in magmas, and analytical errors, the  
 684 analytical totals are generally not exactly 100 wt.%. In order to eliminate the effects of these processes in  
 685 the composition of oxides, researchers normalize the major element composition to a 100% volatile-free  
 686 (also named anhydrous) basis, i.e., by the sum of the major oxides without LOI (when analyzed). In order  
 687 for users to be able to directly compare the major element compositions of sample observations in the

BOOM! Tephrochronology dataset and exploration tool of the Southern (33–46° S) and Austral (49–55° S)  
Volcanic Zones of the Andes

688 BOOM! dataset, both the *raw* and normalized major elements are stored. Because of the heterogeneities  
689 in the data sources in the dataset, when normalizing major elements two aspects should be taken into  
690 consideration, which we describe in the following.

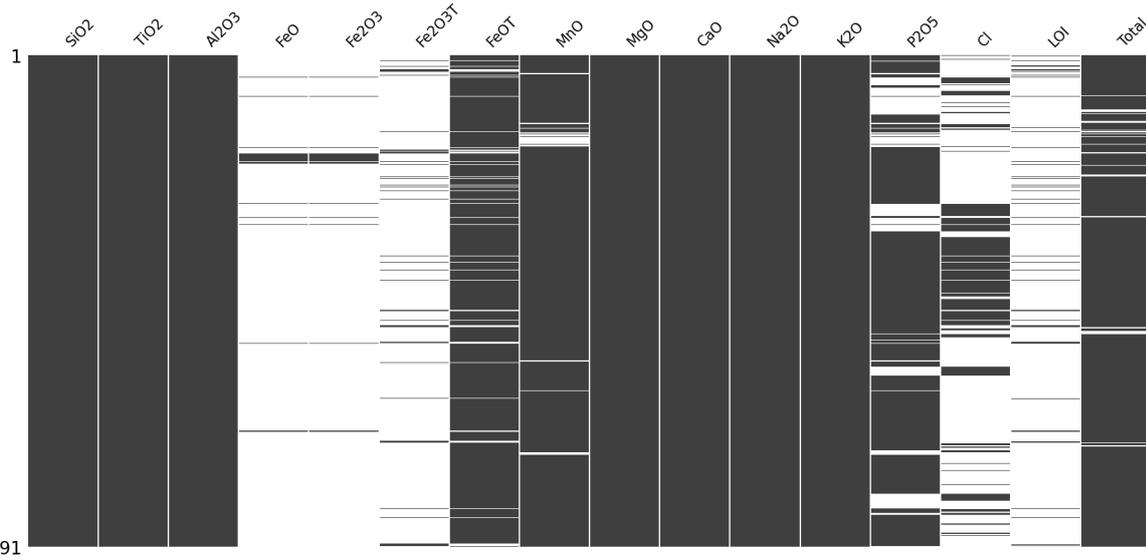
691 In the first place, in order to normalize the major element compositions of samples by a comparable  
692 amount, special attention must be paid to how iron (Fe) was reported. Fe can exist in volcanic rocks in  
693 two states of oxidation ( $\text{Fe}^{2+}$ , as FeO; and  $\text{Fe}^{3+}$ , as  $\text{Fe}_2\text{O}_3$ ), which exist in different proportions in different  
694 types of rocks. Most analytical techniques employed to analyze volcanic rocks are not able to distinguish  
695 between these two states of Fe, and thus it is presented in most publications as a *total* Fe content. For  
696 example, when individual volcanic glass shards are analyzed by EMPA, Fe is usually reported as total Fe  
697 as FeO, which might be communicated by authors as FeO,  $\text{FeO}_T$  or  $\text{FeO}^*$ . In contrast, when bulk samples  
698 are analyzed, for example by XRF or Inductively Coupled Plasma - Optical Emission Spectrometry (ICP-  
699 OES), Fe is typically reported as total Fe as  $\text{Fe}_2\text{O}_3$ , and it might be presented in the publication either as  
700  $\text{Fe}_2\text{O}_3$  or  $\text{Fe}_2\text{O}_{3T}$ . Additionally, other techniques such as AAS and wet chemical analysis (titration) can  
701 discriminate between  $\text{Fe}^{2+}$  and  $\text{Fe}^{3+}$ , however they are not commonly used. In this case, publications  
702 calculate the total Fe as  $\text{FeO}_T = \text{FeO} + 0.899 \cdot \text{Fe}_2\text{O}_3$ , or as  $\text{Fe}_2\text{O}_{3T} = 1.1 \cdot \text{FeO} + \text{Fe}_2\text{O}_3$ . Thus, when normalizing  
703 major element compositions of heterogeneous sources, it is important to calculate the analytical total  
704 using the same expression for Fe, as using one or the other will result in different normalized  
705 compositions. In the BOOM! dataset, ~90% of the major element compositions were analyzed by EMPA.  
706 Because of this, the normalization in the dataset is done considering Fe as  $\text{FeO}_T$  for calculating the  
707 analytical total. When  $\text{FeO}_T$  was not directly presented in the original publication, it was calculated either  
708 as:  $\text{FeO} + 0.899 \cdot \text{Fe}_2\text{O}_3$ , when both FeO and  $\text{Fe}_2\text{O}_3$  were analyzed (~3% of major element observations);  
709 or as  $0.899 \cdot \text{Fe}_2\text{O}_{3T}$ , when Fe was analyzed as  $\text{Fe}_2\text{O}_{3T}$  (~5% of major element observations). To save the  
710 original data and its heterogeneity, Fe is described in the dataset by four attributes: FeO,  $\text{Fe}_2\text{O}_3$ ,  $\text{FeO}_T$  and  
711  $\text{Fe}_2\text{O}_{3T}$ , and only the original data is filled in the non-normalized (*raw*) version of the dataset.

712 Second, different publications might analyze different sets of major elements and thus, samples might not  
713 be normalized by a comparable analytical total. In particular, elements found in lower concentrations

BOOM! Tephrochronology dataset and exploration tool of the Southern (33–46° S) and Austral (49–55° S)  
Volcanic Zones of the Andes

714 (such as  $P_2O_5$ , MnO or Cl) are not always analyzed (Figure 6), because of the increased amount of time  
715 required to analyze them with reasonable precision, versus the information they provide (when using  
716 EMPA). However, because of their low concentrations, whether these elements are analyzed or not, will  
717 have a small influence on normalizations. We have performed a small test in order to check if the  
718 normalized values of samples where different major elements have been analyzed are still comparable.  
719 We re-normalized the major element compositions of sample observations where all major elements  
720 included in the BOOM! dataset were analyzed by a new analytical total, calculated as the sum of all major  
721 elements analyzed, excluding Cl,  $P_2O_5$  and MnO. Then, we calculated the variation between the  
722 normalized and the re-normalized compositions of each element (e.g.,  $SiO_2$  <sub>renormalized</sub> –  $SiO_2$  <sub>normalized</sub>) for  
723 each sample observation and compared it to the maximum achievable precision for that element. For the  
724 latter, we considered the 95% CL of the certified values of the secondary standards analyzed along with  
725 the samples (Figure 7). We only tested sample observations for which both the analytical total are  
726 included in the BOOM! dataset and the secondary standards analyzed have certified values (~61% of  
727 sample observations analyzed for major elements). Because the secondary standards used for whole-  
728 rock and glass analyses are different, we compare the sample observations with their respective  
729 secondary standards. In all cases, except for  $SiO_2$ ,  $Al_2O_3$ , and FeO, the difference of the analyzed  
730 composition between the normalizations is lower than the 95%CL of the geostandards analyzed along  
731 with the samples (Figure 7, S1, S2). Nevertheless, in the case of  $SiO_2$ ,  $Al_2O_3$ , and FeO, ~99% of the sample  
732 observations the difference is within or lower than the 95% CL of the standards for the range of  
733 compositions of the sample observations. For  $Al_2O_3$ , the difference between the two normalizations is  
734 lower than 0.13 wt.% in~ 99% of the cases, which is lower than the 0.2 wt.% analytical error of  $Al_2O_3$  for  
735 most secondary standards analyzed along with the samples, except for BHVO-2G, for which it is 0.1 wt.%.  
736 Thus, even though in the BOOM! dataset sample observations are included for which different major  
737 elements were analyzed, their normalized compositions are still comparable.

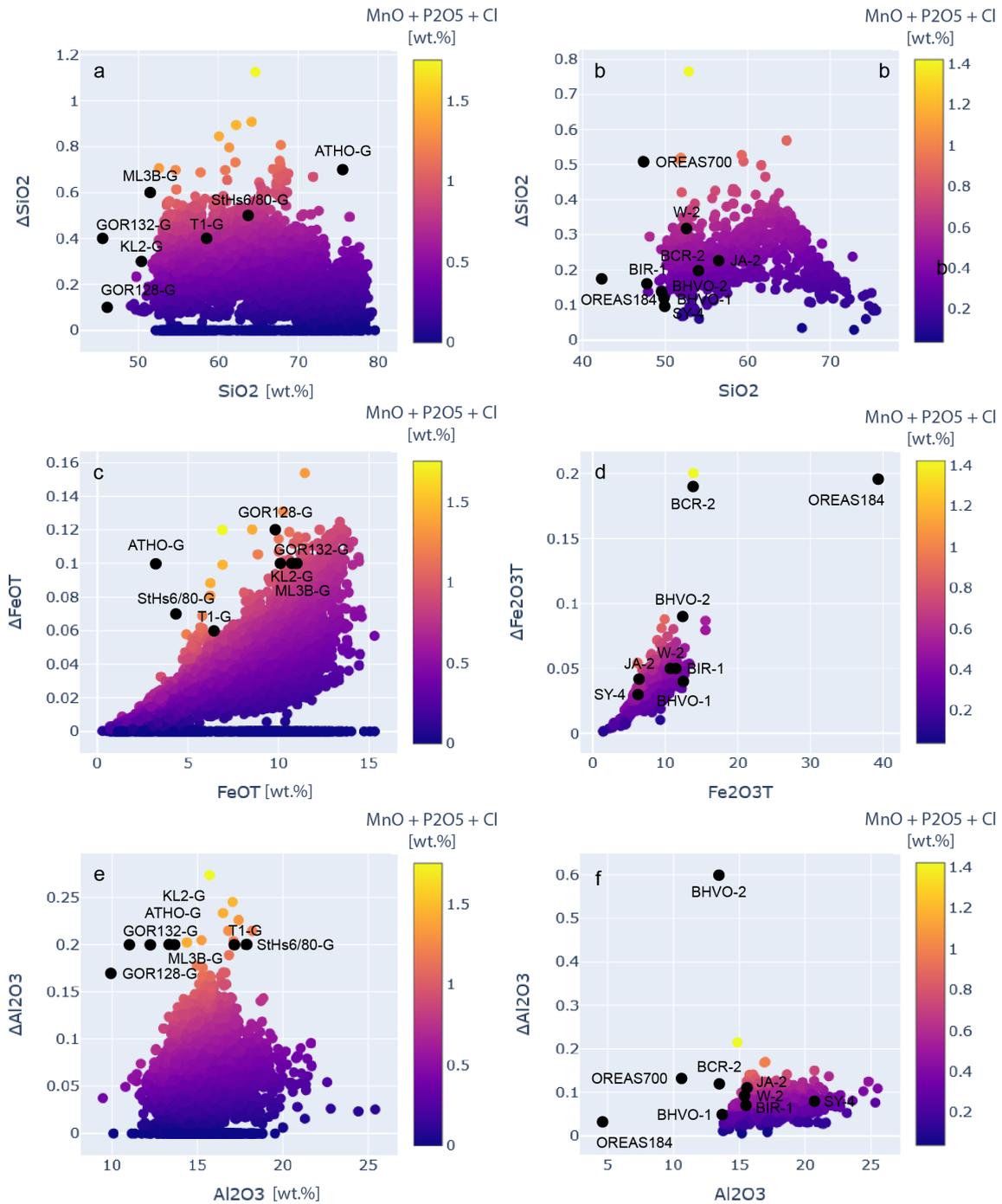
BOOM! Tephrochronology dataset and exploration tool of the Southern (33–46° S) and Austral (49–55° S) Volcanic Zones of the Andes



738

739 **Figure 6** Heterogeneity of the analyzed major elements contained in the BOOM! dataset. The rows  
740 correspond to the sample observations. Dark grey cells represent “filled” values and white cells represent  
741 “missing” values. In this case, 13,791 sample observations have major element data, but not all elements  
742 have been analyzed for each sample. The figure was created using the Python library Missigno.

BOOM! Tephrochronology dataset and exploration tool of the Southern (33–46° S) and Austral (49–55° S) Volcanic Zones of the Andes



743

744 **Figure 7.** Impact of normalizing major element compositions of sample observations in the BOOM!  
 745 dataset by different analytical totals.  $\Delta$ : Normalized value considering the analytical total as the sum of  
 746 all analyzed major elements - Normalized value considering the analytical total without MnO, P<sub>2</sub>O<sub>5</sub> and  
 747 Cl. a, c, e: glass shard analyses; b, d, f: whole-rock analyses. Only samples for which the analytical total is  
 748 published in the original publication are plotted. The 95% CL of the certified secondary standards  
 749 analyzed along with the samples are plotted as a reference of the highest achievable precision of the  
 750 analyses. GOR128-G, GOR132-G, KL2-G, ML3B-G, T1-G, StHs6/80-G, ATHO-G and T1-G (Jochum et al.,  
 751 2006), BCR-2, BHVO-1, BHVO-2, BIR-1, JA-2, W-2 (Jochum et al., 2016), SY-4  
 752 (<https://www.nrcan.gc.ca/our-natural-resources/minerals-mining/mining-resources/sy-4-diorite-gneiss/8025>),  
 753 OREAS184 (<https://www.oreas.com/crm/oreas-184/>), OREAS700  
 754 (<https://www.oreas.com/crm/oreas-700/>).

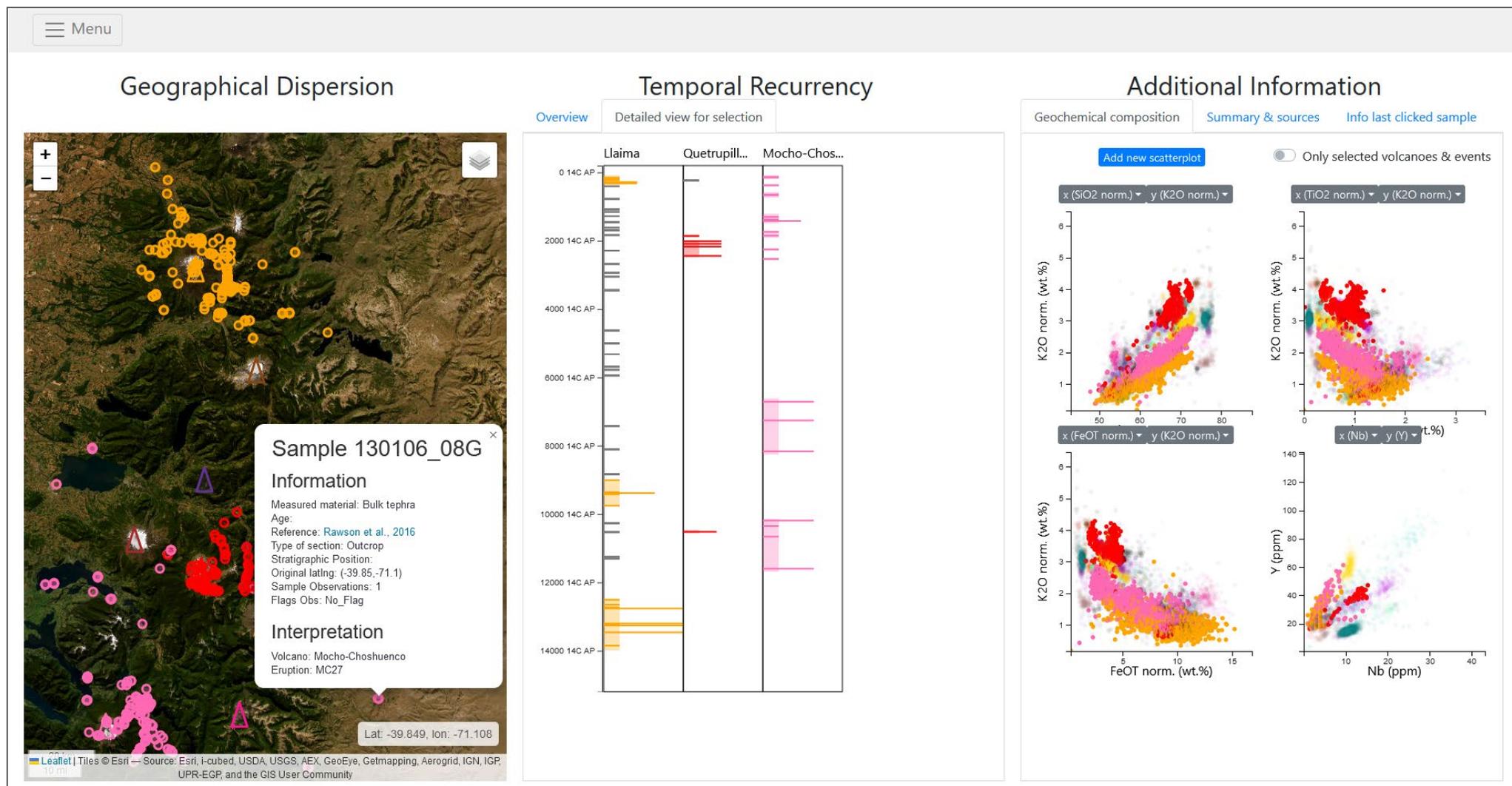
## 755 **5 Using the BOOM! Dataset to explore potential correlations**

756 The BOOM! dataset and explorer were developed to help in the difficult task of disentangling the  
757 tephrochronological record in the SVZ and AVZ, which in practice means robustly correlating tephra  
758 deposited during the same eruptive event. The latter is generally done by fingerprinting them, i.e.,  
759 identifying unique characteristics that distinguish tephra deposits. Doing this is often quite difficult,  
760 because of the multidimensionality of the tephrochronological information, as communicated during the  
761 interviews (Table S1). Because of this, researchers often use different tools to explore and visualize  
762 different type of data (GIS software, Excel, R, Adobe Illustrator, etc.), and so, characterizing each volcanic  
763 center and eruptive event can become very tedious, especially when dealing with large amounts of data,  
764 such as in the BOOM! dataset. With this in mind, the BOOM! explorer (<https://boom-vis.lisn.upsaclay.fr/>)  
765 was created as a complementary tool to the dataset. By using it, users can visualize at the same time the  
766 geographical distribution of volcanic centers and the volcanic products originated during specific  
767 eruptive events, as well as their geochemical composition, and the available information to assess the  
768 chronologies of the eruptive events (Figure 8). The BOOM! explorer can at the same time be used to  
769 compare the available data for each volcanic center and known eruptive event and evaluate the “finger-  
770 printability” of the tephra. In section S6 of the supplementary material we exemplify how the BOOM!  
771 explorer can help users in correlations with an example. It is important to note that the BOOM! explorer  
772 is not meant to be a comprehensive tool, and users can always download the dataset, add supplementary  
773 information, and produce different visualizations fitted to their specific needs; as well as download the  
774 source code of the explorer and make changes fitted to their needs. In this regard, the explorer can also  
775 be used as a catalog of the data in the BOOM! dataset. By using it, users can perform a first inspection of  
776 the dataset, understand the amount and type of information available for each volcanic center and  
777 eruptive event, and chose which information to download. Additionally, we have produced a Jupyter  
778 notebook called Correlations, available in the github<sup>15</sup> repository, which users can use to plot their own  
779 unknowns and compare them with the BOOM! dataset.

---

<sup>15</sup> <https://github.com/consuelola/BOOM/assets/Notebooks>

780



781

782

783

**Figure 8** Screenshots of the sections of the exploration tool (<https://boom-vis.lisn.upsaclay.fr>), showing different visualizations of the data in the dataset. In the figure, volcanic centers Llaima, Quetrupillán, Mocho-Choshuenco have been selected. For more details on how to use the explorer, the reader is referred to section S6.

784 Even though being able to visually compare tephrochronological data in its multidimensionality is very  
785 helpful, sometimes it is not enough to robustly distinguish tephras. The latter is especially true in areas  
786 where the geochemical composition of many volcanic centers and eruptive events is very similar, as is  
787 often the case in the SVZ and AVZ. This is worsened by having large amounts of data from different  
788 eruptive events. The latter is not uncommon, and because of it, statistical tools are sometimes used by  
789 tephrochronologists to find ways to more robustly fingerprint tephras (Lowe et al., 2011; 2017). In  
790 section 5.1, we briefly evaluate the use of geochemical composition together with machine learning  
791 algorithms to classify samples in the dataset labeled as *unknown* volcano. This exploration is not aimed  
792 to be exhaustive either, but rather to provide a starting point and encourage the use of the BOOM! dataset  
793 to explore the application of machine learning for tephra classification.

## 794 **5.1 Machine learning application**

795 The BOOM! dataset offers a great opportunity for the application of machine learning in the discipline of  
796 tephrochronology. To demonstrate this potential, we trained models on the BOOM! dataset to perform  
797 automatic tephra correlation. In machine learning terminology, this corresponds to a classification task,  
798 where the input are sample observations describing the geochemical composition of tephras, and the  
799 outcome to predict is the volcanic system which originated the tephra deposit. For simplicity, in this  
800 experiment we classified each sample observation rather than the sample as a whole, which can be done,  
801 for example, by a majority vote (e.g., Bolton et al., 2020). In practice, when the geochemical composition  
802 of more than one sample observation is available for one tephra deposit, researchers consider the  
803 geochemical trend described by all the sample observations to classify the sample. Thus, future  
804 applications should take this into consideration.

### 805 **5.1.1 Preprocessing**

806 Prior to training the models, a few preprocessing steps are necessary. First of all, only sample  
807 observations for which the volcanic source has been identified were considered, which corresponds to  
808 ~89% of the dataset. This is necessary because the volcanic source “labels” are needed both for training  
809 the models and for evaluating their performances. From the latter, we discarded sample observations for

810 which the volcanic source is debated in different publications (sample observations with a  
811 VolcanicSource\_Issue flag, corresponding to 86 sample observations). Additionally, we only considered  
812 sample observations for which the major or trace elements composition of the volcanic products were  
813 analyzed, i.e., sample observations which correspond to  $^{14}\text{C}$  ages of organic matter were ignored. From  
814 these, we discarded sample observation which are geochemical composition outliers (31 sample  
815 observations), those which have analytical totals lower than 95 wt.%, and those with LOI higher than 5  
816 wt.%, as they might correspond to altered samples (3,242 sample observations, for more details the  
817 reader is referred to section 4.2.2). After this treatment, we only kept sample observations from volcanic  
818 centers left with more than 10 sample observations, from at least two different samples. This is required  
819 in order to have sufficient data for training and testing the models. The latter left a total of 13,925 sample  
820 observations from 2,167 samples from 27 different volcanic centers. As a result of the preprocessing, the  
821 volcanic centers Cay, Macá, Yanteles, Corcovado, Cordón Cabrera and Subsidiary Vcha dome were not  
822 included in the dataset. The whole preprocessing is done in the preprocessing function in `utils.py`,  
823 available in the github repository<sup>16</sup>. We call the resulting dataset the majors *or* traces dataset, as it  
824 includes all sample observations for which either major elements, trace elements or both have been  
825 measured.

### 826 5.1.2 Methods

827 We trained four machine learning classifiers or estimators: Logistic Regression, k-Nearest Neighbors  
828 (kNN), Random Forest (RF) and Gradient Boosting (GB). These classifiers were trained with volcanic  
829 centers as target values and major and trace element compositions as features. Because of the high  
830 percentage of missing data (~62 %), imputation of the missing values was performed prior to model  
831 fitting for all models, except for Gradient Boosting which natively handles missing values. Four different  
832 imputation methods were considered: mean imputation (with `sklearn's SimpleImputer`) which imputes  
833 a constant value, in this case the mean value of each feature; k-Nearest Neighbors imputation (with  
834 `sklearn's KNNImputer`) which imputes by the mean value of the nearest neighbors, in this case 15

---

<sup>16</sup> [https://github.com/consuelola/BOOM/Machine\\_learning/src](https://github.com/consuelola/BOOM/Machine_learning/src)

835 neighbors; iterative conditional imputation (with sklearn’s IterativeImputer) relying on either a Bayesian  
836 Ridge (BR) regressor or a Random Forest regressor. This method is related to the well-known imputation  
837 method MICE (Multiple Imputation by Chained Equations, van Buuren, 2018). It consists in imputing  
838 features iteratively, in a round-robin fashion, using a model (in our case one of BR or RF) that takes as  
839 target the feature to impute, and as input data of all the other features. This results in a total of 13  
840 imputer-estimator pairs tested. For all imputation algorithms and models, we used the scikit-learn<sup>17</sup>  
841 python library (Pedregosa et al., 2011). The code is available on the run\_experiments.py function on  
842 github.

843 The BOOM dataset is subject to class imbalance, i.e., the number of sample observations for each volcanic  
844 center varies greatly. For example, six of the 27 volcanic centers are represented by less than 50 sample  
845 observations in the dataset (Yate, Apagado, Hornopirén, Huequi, Lautaro and Aguilera), while five  
846 volcanic centers are represented by more than 1,000 sample observations (Llaima, Quetrupillán, Mocho-  
847 Choshuenco, Puyehue-Cordón Caulle, and Chaitén). To account for this class imbalance, the performances  
848 of the thirteen models were evaluated by comparing both their accuracies and balanced accuracies. The  
849 accuracy indicates the proportion of correctly classified sample observations over the entire test set,  
850 whereas the balanced accuracy weights sample observations according to the inverse prevalence of their  
851 class (in this case, the volcanic center). Balanced accuracy therefore gives more weight to rarer classes  
852 than accuracy, and allows a better understanding of whether these rarer classes tend to be well-classified  
853 or not. It may, however, be affected by a higher variance: because of the little data available for some of  
854 the volcanic centers (for example, 11 sample observations for Huequi) it can be more sensitive to changes  
855 in the hyperparameters of the models chosen in each fold, to the randomness of the imputing mechanism,  
856 and to the train-test divisions of the sample observations.

857 For model evaluation, we performed a 10-fold cross validation, which gives us 10 performance values  
858 (here accuracy and balanced accuracy) for each imputer-estimator pair tested. On each fold, a grid search  
859 is performed on the training set with an inner 5-fold cross-validation to select the best hyperparameters

---

<sup>17</sup> <https://scikit-learn.org>

860 for the estimator. The model is then retrained with the hyperparameters identified on the whole train set  
861 before being evaluated on the test set. Note that we optimized the hyperparameters of the estimators but  
862 used default hyperparameters for the imputation algorithms to alleviate the computational burden.

863 As many samples are composed of several sample observations, special care was taken not to separate  
864 sample observations belonging to the same sample in the train and test sets. This is very important, as in  
865 practice, either the volcanic source is known for all sample observations in a sample, or unknown for all  
866 of them. Moreover, sample observations coming from the same sample are most likely to be more similar  
867 than sample observations coming from the same volcanic source in general. The failure to keep together  
868 all sample observations from a given sample in the same set would thus not reflect a practical use case  
869 and artificially inflate performances. To ensure that we do not split sample observations inappropriately,  
870 we used a custom `GridSearchCV_with_groups` object as the usual `GridSearchCV` object from scikit-learn  
871 does not allow yet to specify groups.

872 Model training was expensive in terms of memory (~228GB) and processing time (~24 hours on 30  
873 cores), mainly due to the iterative conditional imputation with RF regressor which used a lot of memory  
874 and took two orders of magnitude longer than all other imputation schemes.

### 875 **5.1.3 Results**

876 The mean, standard deviation (SD), and maximum error of the ten accuracies and balanced accuracies  
877 for each model are shown in Table 1 and Figure 9, and the detailed results for each fold can be accessed  
878 in the github repository in the `results/major_or_traces` folder. Additionally, the confusion matrix, feature  
879 importances, and bivariate plots to visualize the classifications were produced for each of the thirteen  
880 models and are also available in the github repository, in the `figures/major_or_traces` folder.

881 **The BOOM! dataset allows the learning of successful models for tephra correlation** - The best model  
882 obtained is the iterative conditional imputation with BR regressor and RF as prediction model. It achieves  
883 92% (SD: 3) accuracy on average over the cross-validation folds, and 79% (SD: 6) balanced accuracy. This  
884 performance highlights a clear signal in the geochemical data. While the successful application of machine  
885 learning has been demonstrated in previous studies (Bolton et al., 2020; Pignatelli & Piochi, 2021; Uslular

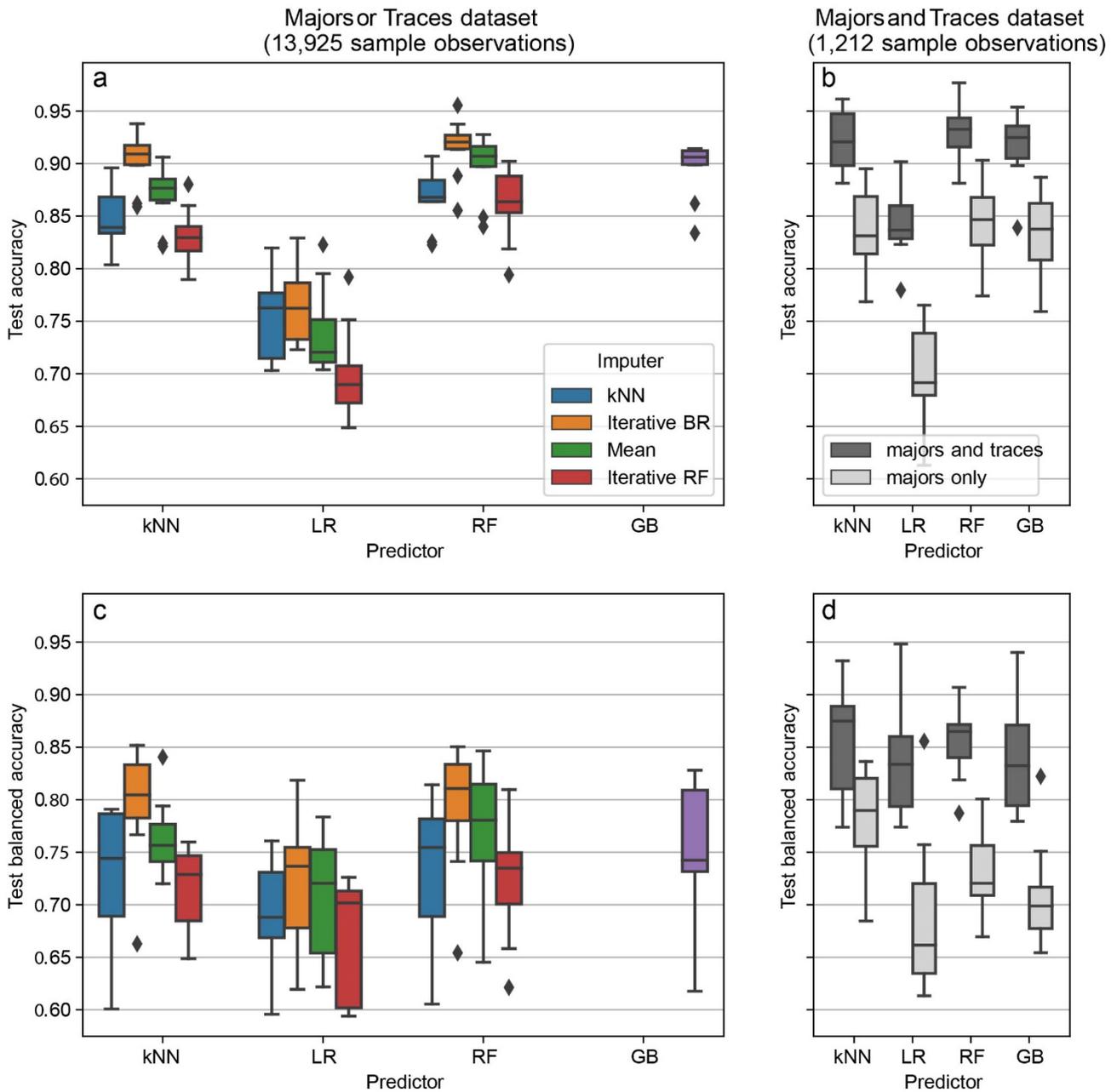
886 et al., 2022), it is the first time that it is demonstrated at this scale (13,925 sample observations from  
 887 2,167 samples and 27 different volcanic centers) on heterogeneous data. Indeed, the geochemical  
 888 analyses were performed by many different laboratories, and many sample observations have been  
 889 analyzed for either major or trace elements, which can affect the performance of machine learning  
 890 models. It is thus positive and encouraging that models good enough to be of practical interest (>90%  
 891 accuracy) can be learned from this data set. Note that here we focused on a few classical and effective  
 892 models, without trying to obtain the best possible performances. It is thus very likely that improvements  
 893 are possible, for example by using ensembling techniques, or by improving the handling of missing values.

894 **Table 1.** Performances of the thirteen models tested in the major *or* traces set, i.e., the dataset containing  
 895 sample observations on which either major elements, trace elements or both have been measured. For  
 896 each of the imputer-estimator pairs, the mean accuracy and balanced accuracy of the 10-folds of the  
 897 cross-validation are presented. Imputers are indicated in the columns and estimators in the rows LR:  
 898 Logistic Regression; kNN: k-Nearest Neighbors; RF: Random Forest; GB: Gradient Boosting; BR: Bayesian  
 899 Ridge.

	Iterative conditional imputation with			
	Simple imputer with mean	kNN Imputer	BR regressor	RF regressor
LR	0.74/0.71	0.75/0.69	0.76/0.72	0.70/0.67
kNN	0.87/0.76	0.85/0.73	0.90/0.80	0.83/0.71
RF	0.90/0.77	0.87/0.73	0.92/0.79	0.86/0.73
GB				0.90/0.75

900 **The imputation algorithm matters** – Our results (Figure 9a and 9c) show that the choice of imputation  
 901 algorithm plays an important role in the performance of the models, both with regards to accuracy and  
 902 balanced accuracy. Interestingly, the ranking of imputation methods remains similar across models. The  
 903 kNN Imputer and iterative conditional imputation with RF led to the worst performances, while the mean  
 904 imputation and iterative conditional imputation with BR had the best performances. This is surprising as

905 iterative conditional imputation with RF is a state-of-the-art imputation method. This situation might be  
906 related to the structure of the “missingness” patterns (Figure S5). The dataset has a block structure, with  
907 10,305 sample observations for which only the major elements have been analyzed, 2,408 sample  
908 observations for which only trace elements have been analyzed, and “only” 1,212 sample observations  
909 for which both major and trace elements have been analyzed. In order to impute the major elements in  
910 the “traces only” block and conversely, imputation algorithms need to learn the relationships between  
911 the majors and traces from the sample observations for which both have been analyzed, which represent  
912 here a relatively small fraction of the data. This is a rather difficult setting, which may explain that mean  
913 imputation outperforms iterative conditional imputation with RF or the kNN imputer. In contrast, the  
914 iterative conditional imputation with BR regression outperforms mean imputation for all models.  
915 Previous work using machine learning for identifying the volcanic source of volcanic products have either  
916 not had any missing values in their dataset (Bolton et al., 2020) or have solely used a simple imputation  
917 with either the mean (Pignatelli et al., 2021) or with zero (Uslular et al., 2020). Our results highlight the  
918 importance of having good imputation schemes to improve overall performances, especially with  
919 datasets compiled from very heterogeneous data, such as the BOOM! dataset. In what follows, we  
920 consider the accuracies and balanced accuracies obtained with the iterative conditional imputation with  
921 BR regression.



922

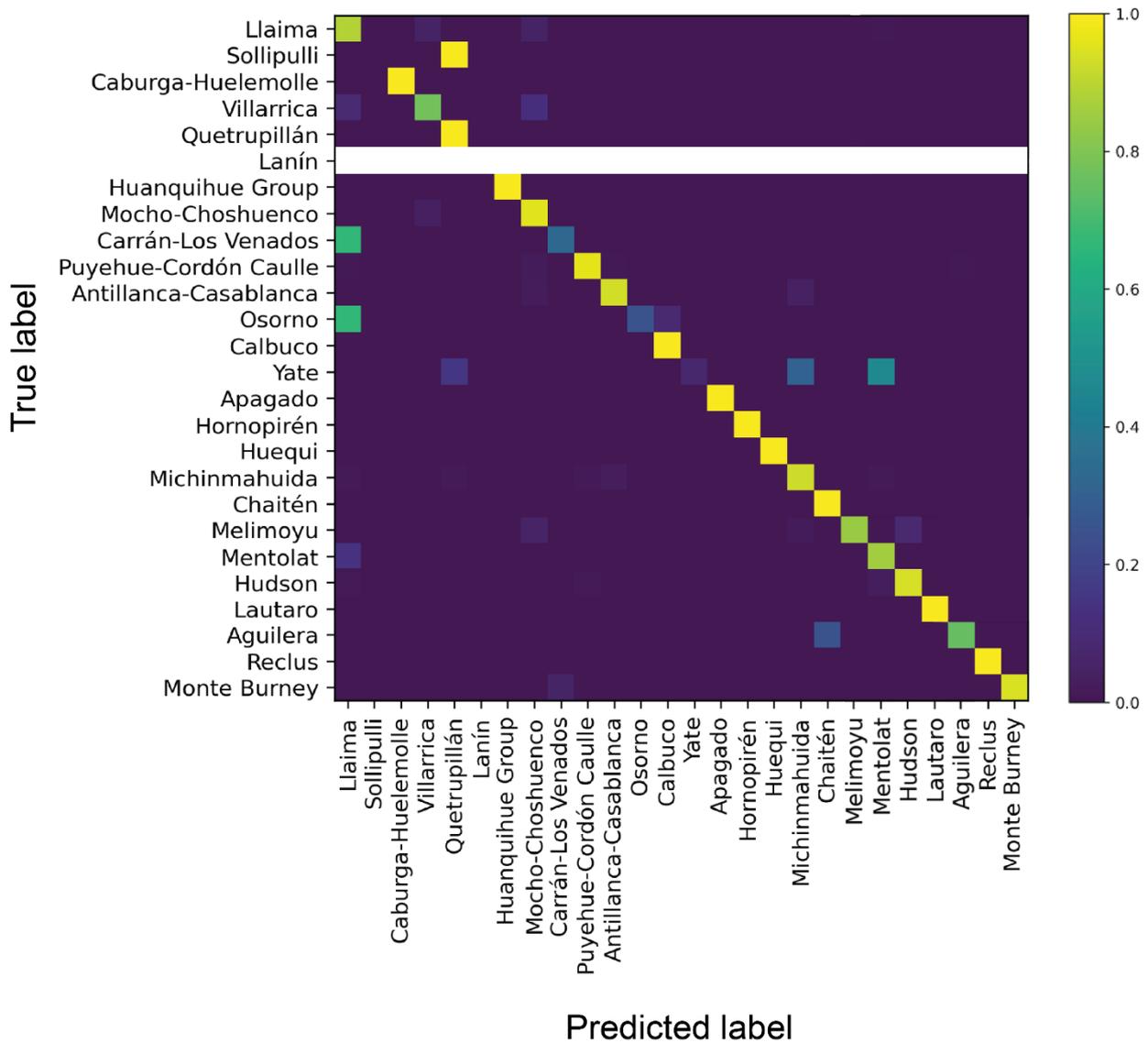
923 **Figure 9.** Boxplots showing the performances of the thirteen imputation-estimator models tested.  
 924 Boxplots were obtained considering the performances on each of the 10 folds of the cross validation. a  
 925 and c correspond to the accuracy and balanced accuracy on the “major *or* traces” set, i.e., the dataset  
 926 containing sample observations on which either major elements, trace elements or both (13,925 sample  
 927 observations) have been measured. In the legend the different imputation mechanisms are indicated:  
 928 kNN: k-Nearest Neighbors; Iterative BR: iterative conditional imputation with Bayesian Ridge regressor;  
 929 Mean: Simple Imputer with the mean value of each element; Iterative RF: iterative conditional imputation  
 930 with Random Forest regressor. On the x axis, the different predictors tested are indicated: kNN: k-Nearest  
 931 Neighbors; LR: Logistic Regression; RF: Random Forest; GB: Gradient Boosting. b and d correspond to the  
 932 performances obtained on the “major *and* traces set”, i.e., the dataset considering only sample  
 933 observations for which *both* major elements *and* at least one trace element have been measured. In b and  
 934 d, the comparison of the performances considering only the major elements of the major and traces set,  
 935 versus the performances considering major and trace elements of the major and traces set, are plotted.

936 **Random Forest performs best, closely followed by KNN and Gradient Boosting** – Overall, RF had the  
937 best performances (acc. 92% / balanced acc. 79%), followed by kNN (90% / 80%), GB (90% / 75%), and  
938 finally LR, whose performance was significantly lower (76% / 72%). These results are not surprising as  
939 RF and GB have been shown to work well with a wide range of datasets (Fernández-Delgado et al., 2014).  
940 In particular, RF is among the best performing algorithms in previous work assessing the use of machine  
941 learning for identifying the volcanic source of volcanic products, (Bolton et al., 2020; Pignatelli & Piochi,  
942 2021; Uslular et al., 2022). In contrast, GB was tested exclusively in the South Aegean Active Volcanic Arc,  
943 with similar performances to that of RF (Uslular et al., 2022). kNN displays more variable results in the  
944 literature. Good performances were observed in a study with Alaska tephras (Bolton et al., 2020) and  
945 kNN had the best performance in the Neapolitan region (Pignatelli & Piochi, 2021), yet it obtained  
946 relatively bad performances in the South Aegean Active Volcanic Arc (Uslular et al., 2022). Finally, LR  
947 displayed the worst performances considering both balanced and unbalanced accuracies, being lower  
948 than all other models outside of 1SD. This reflects the fact that a linear model seems inappropriate on  
949 such a dataset.

950 **Performances on rarer classes are satisfying** – Balanced accuracies are systematically lower than  
951 accuracies, reflecting the fact that the learned classifiers are better on classes that have more sample  
952 observations. This behavior is expected as the more samples we have of a class, the more accurately we  
953 can characterize it. As can be observed in the confusion matrix for the model with the highest accuracy  
954 (Figure 10), the five volcanic centers with more than 1000 sample observations are well classified by the  
955 model. Only Llaima is sometimes classified as Villarrica or Mocho-Choshuenco, however it is relatively  
956 well classified as well. Rarer classes are relatively well classified as well. In fact, from the five less  
957 represented classes, Apagado, Hornopirén, and Huequi are very well classified. The few poorly classified  
958 classes over the entire dataset (Sollipulli, Villarrica, Carrán-Los Venados, Osorno, Yate, and Aguilera), are  
959 represented by a number of sample observations ranging from ~40 to 120, except for Villarrica with 858  
960 sample observations. The latter indicates that classes that are less represented are more likely to be badly  
961 classified. Thus, including more information in the future from the poorly classified samples could  
962 increase the performances of the models.

963 **Assessing the usefulness of traces in addition to major elements** – In tephrochronology trace  
964 elements are generally analyzed to increase the chances of fingerprinting tephtras. However it is unclear  
965 to which point measuring trace elements in addition to major elements can help better identify the  
966 volcanic source of a deposit. To answer this question, we restricted our attention to sample observations  
967 for which *both major and* trace elements are available, as both are needed to evaluate the effect of adding  
968 traces in a model. As most samples have only a subset of trace elements measured, we included all  
969 samples that have at least one trace element measured. It resulted in a dataset comprising 1,212 sample  
970 observations, belonging to 908 samples and representing 25 volcanoes. On this dataset, we trained the  
971 same models as above to predict the volcanic source, however, in this case, we first considered only the  
972 major elements (in the github repository this experiment corresponds to `major_and_traces_restricted`),  
973 and then, both the major and trace elements (`major_and_traces` in the github repository). All training  
974 details are similar to above. When only major elements are considered, the dataset has few missing values  
975 (0.1%) so we just imputed them by the mean. However, there are many missing values in the trace  
976 elements (30%), therefore we compared the various imputation methods on the majors and traces  
977 dataset. The results (Figure S6) show that mean imputation is one of the best methods for this dataset.  
978 Thus, we chose to compare the results with majors only and with majors and traces when both datasets  
979 are imputed by the mean (Figure 9b and 9d). Note that these results are affected by a higher variance  
980 than previously as we are using a dataset with roughly 1,000 sample observations, compared to more  
981 than 10,000 previously.

BOOM! Tephrochronology dataset and exploration tool of the Southern (33–46° S) and Austral (49–55° S) Volcanic Zones of the Andes



982

983 **Figure 10.** Confusion matrix showing the proportion of predicted sample observations by each volcanic center. Rows with no data correspond to classes in which no sample observations were left in the train set. The colored bar to the right of the figure indicates the proportion of classified sample observations, yellow being 1, thus all sample observations were classified as that volcanic center; and purple being 0, thus no sample observations were classified as that volcanic center.

988 Figure 9 (b, d) shows that using traces in addition to majors improves performances substantially across  
 989 all models. The accuracy of kNN, RF and GB rises by 10 percent from 80-85% to 90-95%, while that of LR  
 990 increases by 15 percent. The effect of adding traces as predictors is even stronger when looking at the  
 991 balanced accuracy, except for kNN. While the balanced accuracy greatly varies across models when using  
 992 majors elements only, it is roughly on par across models when adding traces. This may indicate that some  
 993 rarer classes may be better classified when traces are used in addition to majors. Surprisingly, in terms

BOOM! Tephrochronology dataset and exploration tool of the Southern (33–46° S) and Austral (49–55° S)  
Volcanic Zones of the Andes

994 of balanced accuracy, even LR performs almost as well as RF or kNN when traces are considered, reaching  
995 83.4%. This is an interesting observation as LR is a simple and easy to inspect model which  
996 tephrochronology experts might prefer over models that are more difficult to interpret. Overall, these  
997 results highlight the potential of traces for better (automated) fingerprinting of tephras.

998 The results discussed here constitute a first exploration showing encouraging performances. With more  
999 than 90% accuracy and around 80% balanced accuracy, the tests here performed highlight that machine  
1000 learning with the BOOM! dataset has a great potential to improve correlations in the SVZ and AVZ.  
1001 Nevertheless, models would benefit from an improvement in their balanced accuracy to help disentangle  
1002 all tephrochronological records in the region, including the under-represented ones. As mentioned  
1003 before, imputation algorithms have a noteworthy effect on performances when considering the whole  
1004 dataset. Future work should thus focus on understanding the best imputation strategies, given the  
1005 missigness structure of the dataset. In addition, it is important to include in the future a method to make  
1006 pooled predictions, i.e., to predict the source not only for each sample observation independently, but for  
1007 the whole sample at once. This question has been explored by Bolton et al. (2020) and performances of  
1008 all models improved when considering the predictions at the sample level rather than at the sample  
1009 observation level. Finally, using the BOOM! dataset, a more in-depth study can be developed to further  
1010 the use of machine learning not only to identify the volcanic source of tephras, but to identify the exact  
1011 eruption originating the tephra. In this sense, additional information included in the dataset could be  
1012 included in the models, such as the ages, the type of volcanic product, and the location where they were  
1013 identified. The code for the experiment here performed is available on github and can be used to improve  
1014 our baselines.

## 1015 **6 Conclusions and perspectives**

1016 The BOOM! dataset here presented integrates an unprecedented amount of tephrochronological (and  
1017 volcanological) information in the SVZ and AVZ in a machine-readable way. It contains information on 32  
1018 different volcanic centers and 132 eruptive events that occurred during the last 20,000 years, extracted  
1019 from 79 different scientific publications, in ten different disciplines, produced through ~30 years of  
1020 research. The machine-readable dataset allowed for the development of the BOOM! explorer, which  
1021 provides novel visualizations of the tephrochronological information in the area. In particular, the BOOM!  
1022 explorer allows users to visualize data in its multidimensionality (geographical distribution, chronology,  
1023 and geochemical composition), filter data according to different criteria of comparability, and easily  
1024 compare the available information for a given volcanic center or eruption, as well as directly linking the  
1025 information to the source publication. The development of the explorer in parallel to the dataset greatly  
1026 aided the development of its structure, by providing clear ways to organize the bulk of data and identify  
1027 information that was missing in order to meet community needs.

1028 In order to effectively integrate this heterogeneous data, it was key to incorporate information for users  
1029 to evaluate its quality and comparability, which is a repeated request from the tephrochronological  
1030 community. However, much of this information is not always communicated explicitly in publications, or  
1031 in a way that is easily integrated with data from other publications, which impedes the reuse of data and  
1032 can lead to misinterpretation of it. For example, in many publications it is not communicated in which  
1033 state of oxidation of Fe is analyzed, rather it is deduced from the analytical technique. The latter can lead  
1034 to confusion, especially when researchers from different disciplines reuse the information. In other cases,  
1035 the position, type of register, analytical totals, or secondary standards analyzed along with the samples  
1036 are not indicated in publications and could not be obtained when contacting the authors. Other types of  
1037 information are not always communicated in a machine-readable way, making it very time consuming to  
1038 integrate it with other data. In particular, including the stratigraphy of sections was a common request  
1039 by the community, however in many publications this is communicated only in figures showing  
1040 stratigraphic columns without the exact stratigraphic position and thickness of tephtras. Future work in

BOOM! Tephrochronology dataset and exploration tool of the Southern (33–46° S) and Austral (49–55° S)  
Volcanic Zones of the Andes

1041 tephrochronology would ideally publish these data in a machine-readable way to increase its re usability  
1042 and the efficiency of tephrochronological work in the area.

1043 The dataset and explorer are not meant to be comprehensive, and the work here presented is expected  
1044 to be regarded as a basis from which to build upon. Regarding the dataset, more information could be  
1045 incorporated to better describe tephra deposits. For example, more information to describe the physical  
1046 characteristics of the tephra deposits, such as macroscopic description of it and photos; a more thorough  
1047 description of the stratigraphy of sedimentary archives and their isotopic stratigraphy, when existing; as  
1048 well as schematic representations of stratigraphic columns; supplementary dating methods, e.g., from  
1049 dendrochronology, and more information to describe  $^{14}\text{C}$  ages, for example reservoir ages and the choice  
1050 of macrofossil dated, when suited, as well as the calibrated ages by the last published calibration curve;  
1051 regarding the geochemical composition, analyses of mineral compositions or additional isotopic ratios  
1052 such as  $\delta^{18}\text{O}$ . Regarding the explorer, additional or alternative visualizations could be developed to better  
1053 visualize the uncertainty of the geochemical composition analyses and  $^{14}\text{C}$  ages, the uncertainties related  
1054 to age estimates of eruptive events. Additionally, an important request of the community to pursue in the  
1055 future is finding ways to better communicate the uncertainty associated with the interpretation of the  
1056 volcanic source of tephtras, and how this is susceptible to change in the future, as new tephrochronological  
1057 information becomes available. Finally, an important request which could not be met in this version of  
1058 the dataset and the explorer, is that people can upload their own data and compare it to the dataset, which  
1059 would be of great use for the community.

1060 Even though compiling the BOOM! dataset and providing the explorer is a great contribution to the  
1061 tephrochronological community in the area, an important challenge for the future is finding ways for the  
1062 dataset to incorporate more information. In its current form as described here, the version of the dataset  
1063 is static. Taking the latter into consideration, a collaboration with the National Volcanic Network (RNVV)  
1064 of SERNAGEOMIN is already in place to look for ways of transferring the dataset and explorer to secure  
1065 its maintenance and evolution in the future.

## 1066 **7 Data Availability**

1067 The BOOM! dataset is hosted on the ESPRI server of the IPSL, France, which guarantees its hosting and  
1068 access as long as the server exists. Both files of the dataset can be downloaded directly from the IPSL  
1069 catalog (<https://doi.org/10.14768/47b4525f-ff39-4940-a963-4d2673f2362e>) as a Web Map service  
1070 (WMS), Web Feature service (WFS) or as a CSV file. Additionally, either the entire dataset or subsets of it  
1071 can be downloaded as a CSV file through the BOOM! explorer (<https://boom-vis.lisn.upsaclay.fr>).

## 1072 **8 Acknowledgements**

1073 This study benefited from the Chilean scholarship for PhD *Becas Chile, convocatoria* 2017. For data  
1074 hosting and processing, it benefited from the IPSL mesocenter ESPRI facility, which is supported by CNRS,  
1075 UPMC, Labex L-IPSL, CNES and *Ecole Polytechnique*. We especially thank Karim Ramage and Vincent  
1076 Douet from ESPRI for their kind support. Additionally, we would like to especially thank the team from  
1077 the RNVV from SERNAGEOMIN and Sebastian Watt, who kindly participated in the interviews and gave  
1078 continuous feedback for the development of the platform. We also thank every researcher who kindly  
1079 shared details on the information they discuss in their papers in order to enrich the dataset: Brent  
1080 Alloway, Charles Stern, Daniel Sellés, Derek Weller, Eduardo Morgado, Francisco Bucchi, Harriet Rawson,  
1081 Hugo Moreno, José Naranjo, Luis Lara, Maarten Van Daele, Rebecca Smith, Sebastien Bertrand, Simon  
1082 Haberle, and Silke Lohmar. Finally, we thank Tamara Basualto for helping with the graphic design in the  
1083 paper, and Sebastián Miranda, Camila Miranda and Taylor Quilty for helping us come up with BOOM!

## 1084 **9 References**

- 1085 Abarzúa A.M. & Moreno P.I. 2008. Changing fire regimes in the temperate rainforest region of southern  
1086 Chile over the last 16,000 yr. *Quaternary Research*, 69 (1), 62–71.  
1087 <https://doi.org/10.1016/j.yqres.2007.09.004>
- 1088 Abarzúa A.M., Villagrán C. & Moreno P.I. 2004. Deglacial and postglacial climate history in east-central  
1089 Isla Grande de Chiloé, southern Chile (43°S). *Quaternary Research*, 62 (1), 49–59.  
1090 <https://doi.org/10.1016/j.yqres.2004.04.005>

BOOM! Tephrochronology dataset and exploration tool of the Southern (33–46° S) and Austral (49–55° S)  
Volcanic Zones of the Andes

- 1091 Abbott, P., Bonadonna, C., Bursik, M., Cashman, K., Davies, S., Jensen, B., Kuehn, S., Kurbatov, A., Lane, C.,  
1092 Plunkett, G., Smith, V., Thomlinson, E., Thordarsson, T., Walker, J.D., & Wallace, K. 2022.  
1093 Community Established Best Practice Recommendations for Tephra Studies-from Collection  
1094 through Analysis (4.0.2) [Data set]. *Zenodo*. <https://doi.org/10.5281/zenodo.6568306>
- 1095 Alloway, B.V., Pearce, N.J.G., Moreno, P.I., Villarosa, G., Jara, I., De Pol-Holz, R. & Outes, V. 2017a. An 18,000  
1096 year-long eruptive record from Volcán Chaitén, northwestern Patagonia: Paleoenvironmental  
1097 and hazard-assessment implications. *Quaternary Science Reviews*. 168, 151–181.  
1098 <https://doi.org/10.1016/j.quascirev.2017.05.011>
- 1099 Alloway, B.V., Moreno, P.I., Pearce, N.J.G., De Pol-Holz, R., Henríquez, W.I., Pesce, O.H., Sagredo, E.,  
1100 Villarosa, G. & Outes, V. 2017b. Stratigraphy, age and correlation of Lepué Tephra: a widespread  
1101 c. 11 000 cal a BP marker horizon sourced from the Chaitén Sector of southern Chile. *Journal of*  
1102 *Quaternary Science*. 32 (6), 795–829. <https://doi.org/10.1002/jqs.2976>
- 1103 Alloway, B.V., Pearce, N.J.G., Villarosa, G., Outes, V. & Moreno, P.I. 2015. Multiple melt bodies fed the AD  
1104 2011 eruption of Puyehue-Cordón Caulle, Chile. *Scientific Reports*, 5: 17589.  
1105 <https://doi.org/10.1038/srep17589>
- 1106 Amigo, Á., Lara, L.E. & Smith, V.C. 2013. Holocene record of large explosive eruptions from Chaitén and  
1107 Michinmahuida Volcanoes, Chile. *Andean Geology*, 40 (2), 227–248.  
1108 <https://doi.org/10.5027/andgeov40n2-a03>
- 1109 Bertin, L., Moreno, H. & Becerril, L., 2018. *Peligros del Campo Volcánico Carrán-Los Venados, Región de los*  
1110 *Ríos, Carta Geológica de Chile, Serie Geológica Ambiental, N° 33. Servicio Nacional de Geología y*  
1111 *Minería – Chile (SERNAGEOMIN), Subdirección Nacional de Geología.*
- 1112 Bertrand, S., Castiaux, J. & Juvigné, E. 2008a. Tephrostratigraphy of the late glacial and Holocene  
1113 sediments of Puyehue Lake (Southern Volcanic Zone, Chile, 40°S). *Quaternary Research*. 70 (3),  
1114 343–357. <https://doi.org/10.1016/j.yqres.2008.06.001>
- 1115 Bertrand, S., Charlet, F., Charlier, B., Renson, V. & Fagel, N. 2008b. Climate variability of southern Chile  
1116 since the Last Glacial Maximum: a continuous sedimentological record from Lago Puyehue (40°S).  
1117 *Journal of Paleolimnology*, 39, 179–195. <https://doi.org/10.1007/s10933-007-9117-y>

BOOM! Tephrochronology dataset and exploration tool of the Southern (33–46° S) and Austral (49–55° S)  
Volcanic Zones of the Andes

- 1118 Bertrand, S., R. Daga, R. Bedert, & K. Fontijn. 2014. Deposition of the 2011–2012 Cordón Caulle tephra  
1119 (Chile, 40°S) in lake sediments: Implications for tephrochronology and volcanic analogies. *Journal of*  
1120 *Geophysical Research: Earth Surface*, 119, 2555–2573, <https://doi.org/10.1002/2014JF003321>
- 1121 Bitschene, P., Arias, N., Arizmendia, A., Giacosa, R., Griznik, M., Fernández, M., Márquez, M. & Nillni, A.  
1122 1993. Volcanology and environmental impact of the August 1991 eruption of the Hudson Volcano  
1123 (Patagonian Andes; Chile). *Zentralblatt Geologie u. Paläontologie*. Stuttgart. H 1-2: 493-500.
- 1124 Bolton, M.S.M., Jensen, B.J.L., Wallace, K., Praet, N., Fortin, D., Kaufman, D., De Batist, M. 2020. Machine  
1125 learning classifiers for attributing tephra to source volcanoes: an evaluation of methods for Alaska  
1126 tephra. *Journal of Quaternary Science*, 35 (1–2), 81–92. <https://doi.org/10.1002/jqs.3170>
- 1127 Bouvet de Maisonneuve, C., Dungan, M.A., Bachmann, O. & Burgisser, A. 2012. Insights into shallow  
1128 magma storage and crystallization at Volcán Llaima (Andean Southern Volcanic Zone, Chile).  
1129 *Journal of Volcanology and Geothermal Research*. 211–212, 76–91.  
1130 <https://doi.org/10.1016/j.jvolgeores.2011.09.010>
- 1131 Brahm, R., Parada, M. A., Morgado, E., Contreras, C. & McGee, L. E. 2018. Origin of Holocene trachyte lavas  
1132 of the Quetupillán volcanic complex, Chile: Examples of residual melts in a rejuvenated  
1133 crystalline mush reservoir. *Journal of Volcanology and Geothermal Research*, 357, 163–176.  
1134 <https://doi.org/10.1016/j.jvolgeores.2018.04.020>
- 1135 Bucchi, F., Lara, L. E. & Gutiérrez, F. 2015. The Carrán–Los Venados volcanic field and its relationship with  
1136 coeval and nearby polygenetic volcanism in an intra-arc setting. *Journal of Volcanology and*  
1137 *Geothermal Research*, 308, 70-81. <https://doi.org/10.1016/j.jvolgeores.2015.10.013>
- 1138 Carel, M., Siani, G. & Delpech, G. 2011. Tephrostratigraphy of a deep-sea sediment sequence off the south  
1139 Chilean margin: New insight into the Hudson volcanic activity since the last glacial period. *Journal*  
1140 *of Volcanology and Geothermal Research*, 208 (3–4), 99–111.  
1141 <https://doi.org/10.1016/j.jvolgeores.2011.09.011>
- 1142 Casati, E., D’Amico, M., Šefrna, L., Trombino, L., Tunesi, A. & Previtali, F. 2019. Geo-pedological  
1143 contribution to the reconstruction of Holocene activity of Chaitén volcano (Patagonia, Chile).  
1144 *Journal of South American Earth Sciences*, 94: 102222.  
1145 <https://doi.org/10.1016/j.jsames.2019.102222>
- 1146 Clapperton C.M., Sugden D.E., Kaufman D.S. & McCulloch R.D. 1995. The last glaciation in central Magellan  
1147 Strait, southernmost Chile. *Quaternary Research*, 44 (2), 133–148.  
1148 <https://doi.org/10.1006/qres.1995.1058>

BOOM! Tephrochronology dataset and exploration tool of the Southern (33–46° S) and Austral (49–55° S)  
Volcanic Zones of the Andes

- 1149 Costantini, L., Pioli, L., Bonadonna, C., Clavero, J. & Longchamp, C. 2011. A Late Holocene explosive mafic  
1150 eruption of Villarrica volcano, Southern Andes: The Chaimilla deposit, *Journal of Volcanology and*  
1151 *Geothermal Research*, 200 (3–4), 143–158. <https://doi.org/10.1016/j.jvolgeores.2010.12.010>
- 1152 Davies, B.J., Darvill, C.M., Lovell, H., Bendle, J.M., Dowdeswell, J.A., Fabel, D., García, J.L., Geiger, A., Glasser,  
1153 N.F., Gheorghiu, D.M., Harrison, S., Hein, A.S., Kaplan, M.R., Martin, J.R.V., Mendelova, M., Palmer,  
1154 A., Pelto, M., Rodés, Á., Sagredo, E.A., Smedley, R.K., Smellie, J.L., Thorndycraft, V.R., 2020. The  
1155 evolution of the Patagonian Ice Sheet from 35 ka to the present day (PATICE). *Earth-Science*  
1156 *Reviews* 204: 103152. <https://doi.org/10.1016/j.earscirev.2020.103152>
- 1157 Del Carlo, P., Di Roberto, A., D’Orazio, M., Petrelli, M., Angioletti, A., Zanchetta, G., Maggi, V., Daga, R.,  
1158 Nazzari, M. & Rocchi, S., 2018. Late Glacial-Holocene tephra from southern Patagonia and Tierra  
1159 del Fuego (Argentina, Chile): A complete textural and geochemical fingerprinting for distal  
1160 correlations in the Southern Hemisphere. *Quaternary Science Reviews*, 195 (1), 153–170.  
1161 <https://doi.org/10.1016/j.quascirev.2018.07.028>
- 1162 D’Orazio, M., Innocenti, F., Manetti, P., Tamponi, M., Tonarini, S., González-Ferrán, O., Lahsen, A. & Omarini,  
1163 R. 2003. The Quaternary calc-alkaline volcanism of the Patagonian Andes close to the Chile triple  
1164 junction: geochemistry and petrogenesis of volcanic rocks from the Cay and Maca volcanoes  
1165 (~45°S, Chile). *Journal of South American Earth Sciences*, 16 (4), 219–242.  
1166 [https://doi.org/10.1016/S0895-9811\(03\)00063-4](https://doi.org/10.1016/S0895-9811(03)00063-4)
- 1167 Fernández-Delgado M., Cernadas E., Barro S., & Amorim D. 2014. Do we need hundreds of classifiers to  
1168 solve real world classification problems? *Journal of machine learning research*, 15 (1): 3133–3181.
- 1169 Fontijn, K., Lachowycz, S.F., Rawson, H., Pyle, D.M., Mather, T.A., Naranjo, J.A. & Moreno-Roa, H. 2014. Late  
1170 Quaternary tephrostratigraphy of southern Chile and Argentina. *Quaternary Science Reviews*, 89,  
1171 70–84. <https://doi.org/10.1016/j.quascirev.2014.02.007>
- 1172 Fontijn, K., Rawson, H., Van Daele, M., Moernaut, J., Abarzúa, A.M., Heirman, K., Bertrand, S., Pyle, D.M.,  
1173 Mather, T.A., De Batist, M., Naranjo J.A. & Moreno, H., 2016. Synchronisation of sedimentary  
1174 records using tephra: A postglacial tephrochronological model for the Chilean Lake District.  
1175 *Quaternary Science Reviews*. 137, 234–254. <https://doi.org/10.1016/j.quascirev.2016.02.015>

BOOM! Tephrochronology dataset and exploration tool of the Southern (33–46° S) and Austral (49–55° S)  
Volcanic Zones of the Andes

- 1176 Futa K. & Stern C. R. 1988. Sr and Nd isotopic and trace element compositions of Quaternary volcanic  
1177 centers of the southern Andes. *Earth and Planetary Science Letters*, 88 (3–4), 253–262.  
1178 [https://doi.org/10.1016/0012-821X\(88\)90082-9](https://doi.org/10.1016/0012-821X(88)90082-9)
- 1179 Geoffroy, C. A., Alloway, B. V., Amigo, Á., Parada, M. A., Gutierrez, F., Castruccio, A., Pearce, N. J. G., Morgado,  
1180 E. & Moreno, P.I. 2018. A widespread compositionally bimodal tephra sourced from Volcán  
1181 Melimoyu (44°S, Northern Patagonian Andes): Insights into magmatic reservoir processes and  
1182 opportunities for regional correlation. *Quaternary Science Reviews*, 200, 141–159.  
1183 <https://doi.org/10.1016/j.quascirev.2018.09.034>
- 1184 Gerlach, D.C., Frey, F.A., Moreno-Roa, H. & López-Escobar, L. 1988. Recent Volcanism in the Puyehue-  
1185 Cordon Caulle Region, Southern Andes, Chile (40.5°S): Petrogenesis of Evolved Lavas. *Journal of*  
1186 *Petrology*, 29 (2), 333–382. <https://doi.org/10.1093/petrology/29.2.333>
- 1187 Geyh, M.A., Schotterer, U., Grosjean, M. 1998. Temporal changes of the <sup>14</sup>C reservoir effect in lakes.  
1188 *Radiocarbon*, 40 (2), 921–931. <https://doi.org/10.1017/S0033822200018890>
- 1189 Haberle, S.G. & Lumley, S.H. 1998. Age and origin of tephtras recorded in postglacial lake sediments to the  
1190 west of the southern Andes, 44°S to 47°S. *Journal of Volcanology and Geothermal Research*, 84 (3–  
1191 4), 239–256. [https://doi.org/10.1016/S0377-0273\(98\)00037-7](https://doi.org/10.1016/S0377-0273(98)00037-7)
- 1192 Harambour, S.V. 1988. Sobre et hallazgo del mítico volcán Reclus, ex Mano del diablo, Hielo Patagónico  
1193 Sur, Magallanes, Chile. *Revista Geológica de Chile*, 15 (2), 173–179.
- 1194 Heusser, C.J., Heusser, L.E. & Hausser A. 1989. A 12,000 yr B.P. tephra layer at Bahía Inutil (Tierra del  
1195 Fuego, Chile). *Anales del Instituto de la Patagonia*, 19 (1), 39–49.
- 1196 Heusser, C. J., Heusser, L. E., Lowell, T.V., Moreira, A. & Moreira, S. 2000. Deglacial palaeoclimate at Puerto  
1197 del Hambre, subantarctic Patagonia, Chile. *Journal of Quaternary Science*, 15 (2), 101–114.  
1198 [https://doi.org/10.1002/\(SICI\)1099-1417\(200002\)15:2<101::AID-JQS500>3.0.CO;2-Y](https://doi.org/10.1002/(SICI)1099-1417(200002)15:2<101::AID-JQS500>3.0.CO;2-Y)
- 1199 Hickey-Vargas, R.L., Moreno Roa, H., López-Escobar, L. & Frey, F.A. 1989. Geochemical variations in  
1200 Andean basaltic and silicic lavas from the Villarrica-Lanin volcanic chain (39.5° S): an evaluation  
1201 of source heterogeneity, fractional crystallization and crustal assimilation. *Contributions to*  
1202 *Mineralogy and Petrology*. 103, 361–386. <https://doi.org/10.1007/BF00402922>
- 1203 Jacques, G., Hoernle, K., Gill, J., Wehrmann, H., Bindeman, I. & Lara, L.E. 2014. Geochemical variations in  
1204 the Central Southern Volcanic Zone, Chile (38–43°S): The role of fluids in generating arc magmas.  
1205 *Chemical Geology*, 371, 27–45. <https://doi.org/10.1016/j.chemgeo.2014.01.015>

BOOM! Tephrochronology dataset and exploration tool of the Southern (33–46° S) and Austral (49–55° S)  
Volcanic Zones of the Andes

- 1206 Jochum, K.P., Stoll, B., Herwig, K., Willbold, M., Hofmann, A.W., Amini, M., Aarburg, S., Abouchami, W.,  
1207 Hellebrand, E., Mocek, B., Raczek, I., Stracke, A., Alard, O., Bouman, C., Becker, S., Dücking, M., Brätz,  
1208 H., Klemd, R., de Bruin, D., et al. 2006. MPI-DING reference glasses for in situ microanalysis: New  
1209 reference values for element concentrations and isotope ratios. *Geochemistry, Geophysics,*  
1210 *Geosystems*, 7 (2): Q02008. <https://doi.org/10.1029/2005GC001060>
- 1211 Jochum, K.P., Weis, U., Stoll, B., Kuzmin, D., Yang, Q., Raczek, I., Jacob, D.E., Stracke, A., Birbaum, K., Frick,  
1212 D.A., Günther, D. & Enzweiler, J. 2011. Determination of Reference Values for NIST SRM 610—617  
1213 Glasses Following ISO Guidelines. *Geostandards and Geoanalytical Research*, 34 (4), 397–429.  
1214 <https://doi.org/10.1111/j.1751-908X.2011.00120.x>
- 1215 Jochum, K.P., Weis, U., Schwager, B., Stoll, B., Wilson, S.A., Haug, G.H., Andreae, M.O. & Enzweiler, J. 2016.  
1216 Reference Values Following ISO Guidelines for Frequently Requested Rock Reference Materials.  
1217 *Geostandards and Geoanalytical Research*, 40 (3), 333–350. [https://doi.org/10.1111/j.1751-](https://doi.org/10.1111/j.1751-908X.2015.00392.x)  
1218 [908X.2015.00392.x](https://doi.org/10.1111/j.1751-908X.2015.00392.x)
- 1219 Kilian, R., Hohner, M., Biester, H., Wallrabe-Adams, H.J., Stern, C.R. 2003. Holocene peat and lake sediment  
1220 tephra record from the southernmost Chilean Andes (53-55°S). *Revista Geológica de Chile* 30 (1),  
1221 23–37. <http://dx.doi.org/10.4067/S0716-02082003000100002>
- 1222 Kratzmann, D.J., Carey, S., Scasso, R. & Naranjo, J.A. 2008. Compositional variations and magma mixing in  
1223 the 1991 eruptions of Hudson volcano, Chile. *Bulletin of Volcanology*. 71, 419–439.  
1224 <https://doi.org/10.1007/s00445-008-0234-x>
- 1225 Lara, L. E., Naranjo, J. A. & Moreno, H. 2004. Lanín volcano (39.5°S), Southern Andes: geology and  
1226 morphostructural evolution. *Revista Geológica de Chile*, 31 (2), 241–257.  
1227 <https://doi.org/10.4067/S0716-02082004000200004>
- 1228 Lara, L.E., Moreno, H., Naranjo, J.A., Matthews, S. & Pérez de Arce, C. 2006. Magmatic evolution of the  
1229 Puyehue-Cordón Caulle Volcanic Complex (40° S), Southern Andean Volcanic Zone: From shield  
1230 to unusual rhyolitic fissure volcanism. *Journal of Volcanology and Geothermal Research*, 157 (4),  
1231 343–366. <https://doi.org/10.1016/j.jvolgeores.2006.04.010>

BOOM! Tephrochronology dataset and exploration tool of the Southern (33–46° S) and Austral (49–55° S)  
Volcanic Zones of the Andes

- 1232 Lara, L. E. & Moreno, H. 2006. *Geología del complejo volcánico Puyehue-Cordón Caulle, Región de Los Lagos*.  
1233 Servicio Nacional de Geología y Minería, Carta Geológica de Chile, Serie Geológica Básica, 99,  
1234 Escalera 1:50.000, Santiago.
- 1235 Lohmar, 2008. *Pétrologie des grands dépôts d'ignimbrites des volcans Villarrica (Licán et Pucón) et*  
1236 *Llaima (Ignimbrite Curacautín), dans les Andes du Sud (Chili)*. PhD Thesis, Université Blaise  
1237 Pascal - Clermont Ferrand II, France.
- 1238 Lohmar, S., Parada, M., Gutiérrez, F., Robin, C. & Gerbe, M.C. 2012. Mineralogical and numerical  
1239 approaches to establish the pre-eruptive conditions of the mafic Licán Ignimbrite, Villarrica  
1240 Volcano (Chilean Southern Andes). *Journal of Volcanology and Geothermal Research*, 235–236, 55–  
1241 69, <https://doi.org/10.1016/j.jvolgeores.2012.05.006>
- 1242 López-Escobar, L., Parada, M.A., Moreno, H., Frey, F.A. & Hickey-Vargas, R.L. 1992. A contribution to the  
1243 petrogenesis of Osorno and Calbuco volcanoes, Southern Andes (41°00'–41°30'S): a comparative  
1244 study. *Revista Geológica de Chile*, 19 (2), 211–226.
- 1245 López-Escobar, L., Kilian, R., Kempton, P.D. & Tagiri, M. 1993. Petrography and geochemistry of  
1246 Quaternary rocks from the Southern Volcanic Zone of the Andes between 41°30' and 46°00'S,  
1247 Chile. *Revista Geológica de Chile*, 20, 33–55.
- 1248 López-Escobar, L., Parada, M.A., Hickey-Vargas, R.L., Frey, F.A., Kempton, P.D. & Moreno, H. 1995. Calbuco  
1249 Volcano and minor eruptive centers distributed along the Liquiñe-Ofqui Fault Zone, Chile (41°–  
1250 42° S): contrasting origin of andesitic and basaltic magma in the Southern Volcanic Zone of the  
1251 Andes. *Contributions to Mineralogy and Petrology*, 119, 345–361.  
1252 <https://doi.org/10.1007/BF00286934>
- 1253 Lowe, D.J. 2011. Tephrochronology and its application: A review. *Quaternary Geochronology*, 6, 107–153.  
1254 <https://doi.org/10.1016/j.quageo.2010.08.003>
- 1255 Lowe, D.J., Pearce, N.J.G., Jorgensen, M.A., Kuehn, S.C., Tryon, C.A. & Hayward, C.L. 2017. Correlating  
1256 tephtras and cryptotephtras using glass compositional analyses and numerical and statistical  
1257 methods: Review and evaluation. *Quaternary Science Reviews*, 175, 1–44.  
1258 <https://doi.org/10.1016/j.quascirev.2017.08.003>
- 1259 Martínez Fontaine, C., Siani, G., Delpech, G., Michel, E., Villarosa, G., Manssouri, F. & Nouet, J. 2021. Post-  
1260 glacial tephrochronology record off the Chilean continental margin (~41° S). *Quaternary Science*  
1261 *Reviews*, 261: 106928. <https://doi.org/10.1016/j.quascirev.2021.106928>

BOOM! Tephrochronology dataset and exploration tool of the Southern (33–46° S) and Austral (49–55° S)  
Volcanic Zones of the Andes

- 1262 Mayr, C., Smith, R.E., García, M.L., Massafferro, J., Lücke, A., Dubois, N., Maidana, N., I., Meier, W. J-H., Wissel,  
1263 H. & Zolitschka, B. 2019. Historical eruptions of Lautaro Volcano and their impacts on lacustrine  
1264 ecosystems in southern Argentina. *Journal of Paleolimnology*, 62, 205–221.  
1265 <https://doi.org/10.1007/s10933-019-00088-y>
- 1266 McNamara, K., Rus, A.C., Cashman, K.V., Castruccio, A. & Abarzúa, A.M. 2019. Comparison of lake and land  
1267 tephra records from the 2015 eruption of Calbuco volcano, Chile. *Bulletin of Volcanology*, 821: 10  
1268 (2019). <https://doi.org/10.1007/s00445-019-1270-4>
- 1269 Mella, M. 2008. Petrogêneses do complexo vulcânico Yate (42, 30°S), Andes do Sul, Chile. Tese de  
1270 Doutorado, Instituto de Geociências, Universidade de São Paulo, São Paulo.  
1271 <https://doi.org/10.11606/T.44.2009.tde-04032009-091537>
- 1272 Moreno, P.I. & León A.L. 2003. Abrupt vegetation changes during the last glacial to Holocene transition in  
1273 mid-latitude South America. *Journal of Quaternary Science*, 18 (18), 787–800.  
1274 <https://doi.org/10.1002/jqs.801>
- 1275 Moreno, P.I. 2004. Millennial-scale climate variability in northwest Patagonia over the last 15 000 yr.  
1276 *Journal of Quaternary Science*, 19 (1), 35–47. <https://doi.org/10.1002/jqs.813>
- 1277 Moreno, P.I., Alloway, B.V., Villarosa, G., Outes, V., Henríquez, W.I., De Pol-Holz, R. & Pearce, N.J.G., 2015. A  
1278 past-millennium maximum in postglacial activity from Volcán Chaitén, southern Chile. *Geology*,  
1279 43 (1), 47–50. <https://doi.org/10.1130/G36248.1>
- 1280 Moreno, P. I., Vilanova, I., Villa-Martínez, R., Dunbar, R.B., Mucciarone, D.A., Kaplan, M.R., Garreaud, R. D.M.,  
1281 Rojas, M., Moy, C. M., De Pol-Holz, C.M. & Lambert, F. 2018. Onset and Evolution of Southern  
1282 Annular Mode-Like Changes at Centennial Timescale. *Scientific Reports*, 8: 3458.  
1283 <https://doi.org/10.1038/s41598-018-21836-6>
- 1284 Moreno, H., Lara, L. & Orozco, G. 2010. *Geología del Volcán Osorno, Región de Los Lagos*. Servicio Nacional  
1285 de Geología y Minería, Carta Geológica de Chile, Serie Geológica Básica, N° 126. *Servicio Nacional*  
1286 *de Geología y Minería – Chile* (SERNAGEOMIN), *Subdirección Nacional de Geología*.
- 1287 Morgado, E., Parada, M.A., Contreras, C., Castruccio, A., Gutiérrez, F. & McGee, L.E. 2015. Contrasting  
1288 records from mantle to surface of Holocene lavas of two nearby arc volcanic complexes:  
1289 Caburgua-Huelemolle Small Eruptive Centers and Villarrica Volcano, Southern Chile. *Journal of*

BOOM! Tephrochronology dataset and exploration tool of the Southern (33–46° S) and Austral (49–55° S)  
Volcanic Zones of the Andes

- 1290 *Volcanology and Geothermal Research*, 306, 1–16.
- 1291 <https://doi.org/10.1016/j.jvolgeores.2015.09.023>
- 1292 Morgado, E., Morgan, D.J., Harvey, J., Parada, M.Á., Castruccio, A., Brahm, R., Gutiérrez, F., Georgiev, B. &  
1293 Hammond, S.J. 2019. Localised heating and intensive magmatic conditions prior to the 22–23  
1294 April 2015 Calbuco volcano eruption (Southern Chile). *Bulletin of Volcanology*. 81: 24.  
1295 <https://doi.org/10.1007/s00445-019-1280-2>
- 1296 Motoki, A., Orihashi, Y., Naranjo, J.A., Hirata, D., Skvarca, P. & Anma, R. 2006. Geologic reconnaissance of  
1297 Lautaro Volcano, Chilean Patagonia. *Revista Geológica de Chile*, 33 (1), 177–187.  
1298 <http://doi.org/10.4067/S0716-02082006000100008>
- 1299 Naranjo, J.A. & Moreno, H. 1991. Actividad explosiva postglacial en el volcan Llaima, Andes del Sur (38°  
1300 45'S). *Revista Geológica de Chile*, 18, 69–80.
- 1301 Naranjo, J.A., Moreno, H. & Banks, N.G. 1993. *La erupción del volcán Hudson en 1991 (46°S), Región XI,*  
1302 *Aisén, Chile*, Boletín 44. *Servicio Nacional Geología y Minería -Chile (SERNAGEOMIN), Subdirección*  
1303 *Nacional de Geología*.
- 1304 Naranjo, J.A. & Moreno, H. 2005. *Geología del Volcán Llaima, Región de la Araucanía. Carta Geológica de*  
1305 *Chile, Serie Geología Básica, N° 88. Servicio Nacional de Geología y Minería - Chile*  
1306 *(SERNAGEOMIN), Subdirección Nacional de Geología*.
- 1307 Naranjo, J.A., Singer, B.S., Jicha, B.R., Moreno, H. & Lara, L.E. 2017. Holocene tephra succession of Puyehue-  
1308 Cordón Caulle and Antillanca/Casablanca volcanic complexes, southern Andes (40–41°S). *Journal*  
1309 *of Volcanology and Geothermal Research*, 332, 109–128.  
1310 <https://doi.org/10.1016/j.jvolgeores.2016.11.017>
- 1311 Naranjo, J.A. & Stern, C.R. 2004. Holocene tephrochronology of the southernmost part (42°30'–45°S) of  
1312 the Andean Southern Volcanic Zone. *Revista Geológica de Chile*, 31 (2), 225–240.  
1313 <https://doi.org/10.4067/S0716-02082004000200003>
- 1314 Naranjo, J.A., Stern, C.R. 1998. Holocene explosive activity of Hudson Volcano, southern Andes. *Bulletin of*  
1315 *Volcanology* 59, 291–306. <https://doi.org/10.1007/s004450050193>
- 1316 [Scikit-learn: Machine Learning in Python](#), Pedregosa et al., JMLR 12, pp. 2825–2830, 2011.

BOOM! Tephrochronology dataset and exploration tool of the Southern (33–46° S) and Austral (49–55° S)  
Volcanic Zones of the Andes

- 1317 Pearce, N.J.G, Bendall, C.A., and Westgate, J.A. 2008. Comment on “Some numerical considerations in the  
1318 geochemical analysis of distal microtephra” by A.M. Pollard, S.P.E. Blockley and C.S. Lane. *Applied*  
1319 *Geochemistry*, 23 (5), 1353-1364. <https://doi.org/10.1016/j.apgeochem.2008.01.002>
- 1320 Pesce, O.H. & Moreno, P.I. 2014. Vegetation, fire and climate change in central-east Isla Grande de Chiloé  
1321 (43°S) since the Last Glacial Maximum, northwestern Patagonia. *Quaternary Science Reviews*, 90,  
1322 143–157. <https://doi.org/10.1016/j.quascirev.2014.02.021>
- 1323 Pignatelli, A. & Piochi, M. 2021. Machine learning applied to rock geochemistry for predictive outcomes:  
1324 The Neapolitan volcanic history case. *Journal of Volcanology and Geothermal Research*, 415:  
1325 107254. <https://doi.org/10.1016/j.jvolgeores.2021.107254>
- 1326 Prieto, A., Stern, C.R. & Estévez, J.E. 2013. The peopling of the Fuego-Patagonian fjords by littoral hunter-  
1327 gatherers after the mid-Holocene H1 eruption of Hudson Volcano. *Quaternary International*, 317,  
1328 3–13. <https://doi.org/10.1016/j.quaint.2013.06.024>
- 1329 Rawson, H., Naranjo, J. A., Smith, V. C., Fontijn, K., Pyle, D. M., Mather, T. A. & Moreno, H. 2015. The  
1330 frequency and magnitude of post-glacial explosive eruptions at Volcán Mocho-Choshuenco,  
1331 southern Chile. *Journal of Volcanology and Geothermal Research*, 299, 103–129.  
1332 <https://doi.org/10.1016/j.jvolgeores.2015.04.003>
- 1333 Rawson, H., Keller, T., Fontijn, K., Pyle, D. M., Mather, T.A., Smith, V.C. & Naranjo, J. A. 2016. Compositional  
1334 variability in mafic arc magmas over short spatial and temporal scales: Evidence for the signature  
1335 of mantle reactive melt channels. *Earth and Planetary Science Letters*. 456, 66–77.  
1336 <https://doi.org/10.1016/j.epsl.2016.09.056>
- 1337 Reubi, O., Bourdon, B., Dungan, M.A., Koornneef, J.M., Sellés, D., Langmuir, C.H. & Aciego, S. 2011.  
1338 Assimilation of the plutonic roots of the Andean arc controls variations in U-series disequilibria  
1339 at Volcán Llama, Chile. *Earth and Planetary Science Letters*, 303 (1–2), 37–47.  
1340 <https://doi.org/10.1016/j.epsl.2010.12.018>
- 1341 Romero, J.E., Morgavi, D., Arzilli, F., Daga, R., Caselli, A., Reckziegel, F., Viramonte, J., Díaz-Alvarado, J.,  
1342 Polacci, M., Burton, M. & Perugini, D. 2016. Eruption dynamics of the 22–23 April 2015 Calbuco  
1343 Volcano (Southern Chile): Analyses of tephra fall deposits. *Journal of Volcanology and Geothermal*  
1344 *Research*, 317, 15-29. <https://doi.org/10.1016/j.jvolgeores.2016.02.027>

BOOM! Tephrochronology dataset and exploration tool of the Southern (33–46° S) and Austral (49–55° S)  
Volcanic Zones of the Andes

- 1345 Schindlbeck, J.C., Freundt, A. & Kutterolf, S. 2014. Major changes in the post-glacial evolution of magmatic  
1346 compositions and pre-eruptive conditions of Llaima Volcano, Andean Southern Volcanic Zone,  
1347 Chile. *Bulletin of Volcanology*. 76: 830. <https://doi.org/10.1007/s00445-014-0830-x>
- 1348 Sellés, D. & Moreno, H. 2011. *Geología del Volcán Calbuco, Región de Los Lagos. Carta Geológica de Chile,*  
1349 *Serie Geología Básica, N° 130. Servicio Nacional de Geología y Minería – Chile (SERNAGEOMIN),*  
1350 *Subdirección Nacional de Geología.*
- 1351 Siani, G., Michel, E., De Pol-Holz, R., DeVries, T., Lamy, F., Carel, M., Isguder, G., Dewilde, F. & Laurantou, A.  
1352 2013. Carbon isotope records reveal precise timing of enhanced Southern Ocean upwelling during  
1353 the last deglaciation. *Nature Communications*, 4: 2758. <https://doi.org/10.1038/ncomms3758>
- 1354 Simmons, I.C., McGarvie, D., Cortés, J.A., Calder, E.S. & Pavez, A. 2020. Holocene volcanism at the  
1355 Quetrupillán Volcanic Complex (39°30' S, 71°43' W), southern Chile. *Volcanica*, 3 (1), 115–137.  
1356 <https://doi.org/10.30909/vol.03.01.115137>
- 1357 Singer, B.S., Jicha, B.R., Harper, M.A., Naranjo, J.A., Lara, L.E. & Moreno-Roa, H. 2008. Eruptive history,  
1358 geochronology, and magmatic evolution of the Puyehue-Cordón Caulle volcanic complex, Chile.  
1359 *GSA Bulletin*, 120 (5–6), 599–618. <https://doi.org/10.1130/B26276.1>
- 1360 Smith, R.E., Smith, V.C., Fontijn, K., Gebhardt, A.C., Wastegård, S., Zolitschka, B., Ohlendorf, C., Stern, C. &  
1361 Mayr, C. 2019. Refining the Late Quaternary tephrochronology for southern South America using  
1362 the Laguna Potrok Aike sedimentary record. *Quaternary Science Reviews*. 218, 137–156.  
1363 <https://doi.org/10.1016/j.quascirev.2019.06.001>
- 1364 Stern, C. R., Moreno, P. I., Villa-Martínez, R., Sagredo, E. A., Prieto, A. & Labarca, R. 2011. Evolution of ice-  
1365 dammed proglacial lakes in Última Esperanza, Chile: implications from the late-glacial R1  
1366 eruption of Reclús volcano, Andean Austral Volcanic Zone. *Andean Geology*, 38 (1), 82–97.
- 1367 Stern, C.R., de Porras, M.E. & Maldonado, A. 2015. Tephrochronology of the upper Río Cisnes valley (44°S),  
1368 southern Chile. *Andean Geology*, 42 (2), 173–189. <https://doi.org/10.5027/andgeoV42n2-a02>
- 1369 Stern, C.R. 1992. Tephrochronology of Magallanes: new data and implications. *Anales del Instituto de la*  
1370 *Patagonia* 21: 129-139.

BOOM! Tephrochronology dataset and exploration tool of the Southern (33–46° S) and Austral (49–55° S)  
Volcanic Zones of the Andes

- 1371 Stern, C.R., 2008. Holocene tephrochronology record of large explosive eruptions in the southernmost  
1372 Patagonian Andes. *Bulletin of Volcanology*, 70, 435–454. [https://doi.org/10.1007/s00445-007-](https://doi.org/10.1007/s00445-007-0148-z)  
1373 [0148-z](https://doi.org/10.1007/s00445-007-0148-z)
- 1374 Uslular, G., Kiyıkçı, F., Karaarslan, E. & Gençalioglu Kuşcu, G. 2022. Application of machine-learning  
1375 algorithms for tephrochronology: a case study of Plio-Quaternary volcanic fields in the South  
1376 Aegean Active Volcanic Arc. *Earth Science Informatics* 15, 1167–1182.  
1377 <https://doi.org/10.1007/s12145-022-00797-5>
- 1378 Tagiri, M., Moreno, H., López-Escobar, L. & Notsu, K. 1993. Two magma types of the high-alumina basalt  
1379 series of Osorno, Southern Andes (41°06'S)-plagioclase dilution effect. *Journal of Mineralogy,*  
1380 *Petrology and Economic Geology*, 88 (7), 359–371. <https://doi.org/10.2465/ganko.88.359>
- 1381 van Buuren, S. 2018. *Flexible Imputation of Missing Data* (Second edition). CRC/Chapman & Hall.  
1382 <https://doi.org/10.1201/9780429492259>
- 1383 Villarosa, G., Outes, V., Hajduk, A., Crivelli Montero, E., Sellés, D., Fernández, M. & Crivelli, E. 2006.  
1384 Explosive volcanism during the Holocene in the Upper Limay River Basin: The effects of ashfalls  
1385 on human societies, Northern Patagonia, Argentina. *Quaternary International*, 158 (1), 44–57.  
1386 <https://doi.org/10.1016/j.quaint.2006.05.016>
- 1387 Watt, S.F.L., Pyle, D.M., Mather, T.A., Martin, R.S. & Matthews, N.E. 2009. Fallout and distribution of  
1388 volcanic ash over Argentina following the May 2008 explosive eruption of Chaitén, Chile. *Journal*  
1389 *of Geophysical Research*, 114: B04207. <https://doi.org/10.1029/2008JB006219>
- 1390 Watt, S.F.L., Pyle, D.M. & Mather, T.A. 2011a. Geology, petrology and geochemistry of the dome complex  
1391 of Huequi volcano, southern Chile. *Andean Geology*, 38 (2), 335–348.  
1392 <https://doi.org/10.5027/andgeov38n2-a05>
- 1393 Watt, S.F.L., Pyle, D.M., Naranjo, J.A., Rosqvist, G., Mella, M., Mather, T.A. & Moreno, H., 2011b. Holocene  
1394 tephrochronology of the Hualaihue region (Andean southern volcanic zone, ~42° S), southern  
1395 Chile. *Quaternary International*, 246 (1–2), 324–343.  
1396 <https://doi.org/10.1016/j.quaint.2011.05.029>

BOOM! Tephrochronology dataset and exploration tool of the Southern (33–46° S) and Austral (49–55° S)  
Volcanic Zones of the Andes

- 1397 Watt, S.F.L., Pyle, D.M. & Mather, T.A. 2013. Evidence of mid-to late-Holocene explosive rhyolitic eruptions  
1398 from Chaitén Volcano, Chile. *Andean Geology*, 40 (2), 216–226.  
1399 <https://doi.org/10.5027/andgeoV40n2-a02>
- 1400 Weller, D., Miranda, C.G., Moreno, P.I., Villa-Martínez, R. & Stern, C.R. 2014. The large late-glacial Ho  
1401 eruption of the Hudson volcano, southern Chile. *Bulletin of Volcanology*, 76, Article: 831.  
1402 <https://doi.org/10.1007/s00445-014-0831-9>
- 1403 Weller, D.J., Miranda, C.G., Moreno, P.I., Villa-Martínez, R. & Stern, C.R. 2015. Tephrochronology of the  
1404 southernmost Andean Southern Volcanic Zone, Chile. *Bulletin of Volcanology*, 77, Article: 107.  
1405 <https://doi.org/10.1007/s00445-015-0991-2>
- 1406 Weller, D.J., de Porras, M.E., Maldonado, A., Méndez, C. & Stern, C.R. 2017. Holocene tephrochronology of  
1407 the lower Río Cisnes valley, southern Chile. *Andean Geology*, 44 (3), 229–248.  
1408 <https://doi.org/10.5027/andgeov44n3-a01>
- 1409 Weller, D.J., de Porras, M.E., Maldonado, A., Méndez C. & Stern, C.R. 2019. Petrology, geochemistry, and  
1410 correlation of tephra deposits from a large early-Holocene eruption of Mentolat volcano, southern  
1411 Chile. *Journal of South American Earth Sciences*, 90, 282–295.  
1412 <https://doi.org/10.1016/j.jsames.2018.12.020>
- 1413 Wilkinson, M., Dumontier, M., Aalbersberg, I., G., Appleton, M., Axton, A., Baak, N., Blomberg, J-W., Boiten,  
1414 L.B., da Silva Santos, P.E., Bourne, J., Bouwman, A.J., Brookes, T., Clark, M., Crosas, I., Dillo, O.,  
1415 Dumon, S., Edmunds, C.T., Evelo, R., Finkers, et al. 2016. The FAIR Guiding Principles for scientific  
1416 data management and stewardship. *Scientific Data*, 3: 160018 (2016).  
1417 <https://doi.org/10.1038/sdata.2016.18>
- 1418 Yu, S.-Y., Shen, J., Colman, S.M. 2007. Modeling the radiocarbon reservoir effect in lacustrine systems.  
1419 *Radiocarbon*, 49 (3), 1241–1254. <http://doi.org/10.1017/S0033822200043150>
- 1420