

Machine learning-driven characterization and prescription of aerosol optical properties for atmospheric models

Nilton Évora do Rosário¹, Karla M. Longo², Pedro H. Toso¹, Saulo R. Freitas², Marcia A. Yamasoe³, Luiz Flávio Rodrigues², Otavio Medeiros², Haroldo Campos Velho², Isilda da Cunha Menezes⁴, Ana Isabel Miranda⁴

¹ Departamento de Ciências Ambientais, Universidade Federal de São Paulo, Diadema, SP Brazil

² Instituto Nacional de Pesquisas Espaciais (INPE), São José dos Campos, SP, Brazil

³ Departamento de Ciências Atmosféricas, Instituto de Astronomia, Geofísica e Ciências Atmosféricas, Universidade de São Paulo, Cidade Universitária, São Paulo, SP, Brazil

⁴ Center for Environmental and Marine Studies (CESAM), Department of Environment and Planning, University of Aveiro, Campus Universitário de Santiago, 3810-193 Aveiro, Portugal

Correspondence to: Nilton do Rosário (nrosario@unifesp.br)

Abstract

Accurate modeling of aerosol optical properties is critical to simulate aerosol radiative effects. However, uncertainties regarding the simulation of aerosol-intensive optical properties are still significant. Therefore, the use of observations to constrain aerosol optical properties in models has been indicated as an option. Also, explicit computations of optical properties are still too costly for operational models, which makes observation-based prescriptions a convenient solution. We developed an observation-based prescription of aerosol optical properties driven by machine-learning techniques that can be applied in models. The Iberian Peninsula (IP) was taken as the reference domain, and the aerosol products from the AERONET sites across the IP were the main dataset. First, clustering was applied to define the typical aerosol optical regimes affecting the IP atmosphere. Five typical regimes were identified. Two of them were dominated by coarse mode, which was associated with Saharan dust. One was found to be close to pure dust, while the other indicated a mixed scenario of dust and pollution. Two of the non-dust regimes, strongly and moderately absorbing, were found to be associated with smoke. The remaining non-dust regime, with no clear association, occurs mostly in the eastern portion of the IP. Afterward, using aerosol-type columnar mass density from MERRA-2, a model was trained as a predictor of the optical regimes using the Random Forest method. The model was tested under distinct aerosol scenarios. Predictions' accuracy ranged from 60 to 75%, depending on the regime, while presenting an average accuracy of 70%.

Keywords: Aerosol Optical Properties, AERONET, MERRA-2, Machine-Learning, Random Forest

37 1. Introduction

38 The importance of aerosols in the Earth's climate system is undisputed. Aerosol particles
39 participate directly in the planetary energy budgets via the scattering and absorption of
40 terrestrial and solar radiation (Kim and Ramanathan 2008; IPCC, 2021; Li et al., 2022).
41 However, this participation is permeated by high complexity due to the variety of aerosol
42 particles sizes and composition, which cause significant uncertainty (Spencer et al. 2019;
43 IPCC, 2021; Li et al., 2022). The uncertainties and challenges in accurately representing
44 aerosol particles' processes in climate, weather, and environmental models arise from
45 various limitations. For instance, when it comes to aerosol direct interaction with radiation,
46 the current global aerosol monitoring system does not provide a comprehensive spatial-
47 temporal characterization of spectral complex refractive index and size distribution of the
48 aerosol particles, two critical information to characterize the particle absorption and
49 scattering (Samset et al. 2018; Li et al., 2022). This lack of observational data contributes
50 significantly to uncertainty in aerosol modeling and, therefore, to the uncertainty of the
51 aerosol radiative forcing.

52 The difficulty of the traditional libraries of aerosol optical and microphysical properties
53 (Shettle and Fenn, 1979; Koepke et al., 1997; Hess et al., 1998) to describe aerosol properties'
54 geographical variation, for instance, soil dust mineralogy (Adebisi et al., 2023), has been
55 central in the aerosol optical properties uncertainty debate. Another critical aspect is the
56 characterization of the state of the mixture of the aerosol particles in the model's aerosol
57 modules (Samset et al. 2018; Sand et al., 2021). Given the complex dynamic of aerosol particle
58 emission, transport, and removal in the atmosphere, numerical modelling of the state of the
59 mixture and the resultant complex refractive index and size distribution is widely recognized
60 as one of the most important sources of uncertainty in addressing aerosol particles' radiative
61 forcing (Sand et al., 2021). According to Sand et al. (2021), aerosol absorption is poorly
62 constrained, and the current climate models present a large range in the quantification of the
63 main absorbing aerosol species (black carbon (BC), organic aerosols (OA), and mineral dust).
64 Brown et al. (2021) findings indicate that biomass-burning aerosols in most climate models
65 are too absorbing, mainly due to treatments of aerosol mixing state, which idealized internal
66 mixing assumptions used in climate models overestimate Black Carbon(BC) absorption
67 enhancement when compared to observations. Saharan dust, a critical component of the
68 global aerosol system, has been found to absorb less solar radiation than models estimate
69 (Adebisi et al., 2023), and the primary cause pointed out is the models' overestimation of the
70 dust imaginary refractive index. Absorption is not the only issue facing aerosol particle
71 representation in climate models; the relative contribution of fine and coarse mode particles
72 is also a challenge. For instance, Adebisi et al. (2023) also found models underestimating
73 large dust particles when representing North African dust plumes.

74 Observation-constrained models have been recommended to mitigate models' current
75 difficulty in fully simulating aerosol properties and processes accurately (Samset et al. 2018;
76 Proske et al., 2024). In addition to the uncertainty aspects, explicit simulation of aerosol
77 composition and microphysical properties, followed by explicit computation of intensive
78 optical properties, is still too expensive computationally for operational models, which also
79 makes observational-based prescriptions a convenient solution. Zhong et al. (2022) used

80 relationships from an ensemble of aerosol models and satellite observations to identify the
81 primary source of uncertainty in aerosol modelling results in biomass burning regions. Their
82 study pointed out the incorrect simulations of lifetimes and the underestimation of mass
83 extinction coefficients as the main reasons for their difficulty in matching observed aerosol
84 optical depth (AOD). As the largest, time and device-consistent observational network,
85 capable of constraining multiple aerosol intensive microphysical and optical properties, the
86 Aerosol RObotic NETwork (AERONET) has been used worldwide to constrain models and
87 satellite algorithms (Omar et al., 2005; Li et al., 2010; Levy et al., 2010; Rosario et al., 2013;
88 Russel et al., 2014; Chen et al., 2023). Chen et al. (2023) developed an aerosol optical module
89 with observation-constrained Black Carbon properties to improve aerosol absorption
90 simulation. Their sensitivity simulations show a reduction of 18%–69% in the biases of
91 aerosol single-scattering co-albedo when compared with global observations from
92 AERONET. Li et al. (2010) used AERONET retrievals to evaluate and improve the
93 performance of a GCM aerosol optical module. They found their GCM to simulate flatter
94 Aerosol Optical Depth (AOD) spectral dependence, indicating an Angstrom Exponent (AE)
95 biased to low values, which suggests that the aerosol sizes simulated were too large. After
96 adjusting the aerosol's size based on AERONET retrievals, the agreement between simulated
97 and observed AOD improved for all aerosol regimes, but especially for smoke and dust
98 scenarios. Rosario et al. (2013) used a set of spectral optical models developed from
99 AERONET sky retrievals over distinct biomes combined with the concept of anisotropic areas
100 of influence of the AERONET sites (Hoelzemann et al., 2009) to constrain smoke aerosol
101 radiative effect modelling during South American biomass burning. By doing so, they were
102 able to capture the effect of the regional variability of smoke optical properties (absorption
103 and size-related) on the surface solar irradiance related to the biomes' distinct nature of
104 smoke.

105 Global and regional cluster analysis of AERONET long-term retrievals of aerosol properties
106 has proved valuable to classify observations in terms of aerosol optical regimes, providing
107 means to qualitative constraints on aerosol properties (Omar et al, 2005; Levy et al., 2007;
108 Russell et al., 2014; Li et al., 2019; Fan et al., 2020; Zhou et al., 2023). In these studies, the
109 number of identified typical aerosol optical regimes varied from 4 to 10, numbers that were
110 expected to likely represent either global or regional major aerosol scenarios, according to
111 each study's focus. In their study, Zhou et al. (2023) found that regional aerosol regime
112 classifications performed better than global classifications when applied to simulate AOD
113 during pollution episodes and in different seasons in Beijing, China. They found larger
114 differences between the strong and moderately absorbing aerosol regimes, namely dust and
115 smoke regimes, when comparing global and regional clustering results. This is a consequence
116 of the differences between China's regional dust and smoke aerosol particles' physical and
117 chemical characteristics and those of global dust and smoke mean features. Another aspect
118 highlighted by Zhou et al. (2023) is that smoke and dust-dominated optical regimes are more
119 frequent globally than in China. Their result suggests that regional classification better
120 captures typical aerosol optical regimes influencing a specific domain and, therefore, with
121 the potential to improve observation-constrained simulations of aerosol radiative forcing.

122 Focusing on the Iberian Peninsula (IP), this study sought to characterize the typical aerosol
123 optical regimes driving the variability of aerosol-intensive properties over the peninsula,

124 aiming to constrain aerosol optical properties prescription in atmospheric models using a
125 novel machine-learning approach. IP is a region affected by a highly dynamic and complex
126 set of aerosol mixing, including natural and anthropogenic particles (Cachorro et al., 2016;
127 Gomez-Amo et al., 2017). Natural sources include marine aerosols from the Atlantic Ocean
128 and Mediterranean Sea, mineral dust from North Africa, and, eventually, wildfire emissions.
129 Major anthropogenic sources are urban-industrial, particularly in more densely populated
130 regions, and biomass burning driven by human activities, especially in the north and central
131 Portugal and eastern and northern Spain. Regional column-integrated optical properties are
132 highly sensitive to the mixing of this diversity of aerosol types, in particular to dust and
133 smoke mixing (Gomez-Amo et al., 2017).

134 The manuscript is organized as follows: Section 2 includes a brief overview of the Iberian
135 Peninsula, focusing on the main atmospheric circulation features and major aerosol particle
136 sources affecting the region, followed by the description of the dataset and methods adopted
137 to identify, characterize, and prescribe the identified aerosol typical regimes. Results and
138 discussions are presented in Section 3. First, the identified aerosol optical regimes and their
139 major features are described and contextualized. Subsequently, the results of the novel
140 machine-learning approach to prescribing the optical regimes are discussed and evaluated.
141 Finally, the main findings of our study are highlighted in the conclusion section.

142

143 **2. Study Region, Data and Methods**

144

145 **2.1 Study region**

146 The Iberian Peninsula (**Figure 1**), comprising Spain and Portugal, exhibits diverse climate
147 conditions due to its complex topography and proximity to the Atlantic Ocean, the
148 Mediterranean Sea, and North Africa. The wind circulation over the peninsula is shaped by
149 its location between the Atlantic Ocean and the Mediterranean Sea, diverse topography, and
150 interactions between regional and global atmospheric patterns, leading to complex wind
151 circulations that significantly influence the region's climate. This results in distinct climate
152 zones, from arid deserts to lush green forests. The Mediterranean climate spans most of
153 Spain, including the eastern and southern coastal regions and central Portugal, featuring hot
154 and dry summers, especially inland. Winters are mild, rarely dropping below 10°C in coastal
155 areas. Most precipitation, often rain, occurs in autumn and winter, leading to dry summers
156 that increase wildfire risks. Wildfires regularly occur in the IP region, fueled by extreme
157 weather conditions, abnormal high temperature records combined with strong, dry winds
158 (Asfaw et al., 2022; Ermitão et al., 2023). Under these scenarios, the entire region can be
159 affected by smoke plumes that often shape the entire region's optical properties (Elias et al,
160 2004; Gomez-Amo et al., 2017). But wildfires are more frequent in the north and central
161 region of Portugal and the north and eastern portion of Spain (Ermitão et al., 2023; Alvares
162 et al., 2024). Oceanic climate is typical in northern coastal regions of Spain, such as Galicia,
163 Asturias, and the Basque Country, and parts of northern Portugal. The Atlantic Ocean
164 influences mild temperatures year-round, with minimal seasonal variation and abundant,
165 evenly distributed rainfall. Annual precipitation can exceed 1,000 mm, with frequent cloud

166 cover and high humidity, especially in winter. The Continental climate of the central plateau
167 (Meseta Central) and the Ebro Valley features extreme temperature variations, with hot
168 summers, highs often above 35°C, and winters below freezing. The central regions have less
169 precipitation than the coastal areas, with a semi-arid climate in some parts. Most rainfall
170 occurs in spring and autumn. Arid and Semi-Arid Climates are found in Southeastern Spain,
171 especially in Murcia and Almería, and parts of the Ebro Valley. These areas receive very low
172 rainfall, often less than 300 mm annually, leading to desert-like conditions like those in the
173 Tabernas Desert. Summers are extremely hot, while winters are mild. Southern Spain,
174 especially the Andalusian region, can be affected by hot and dry winds from the Sahara,
175 causing heat waves and dust storms.

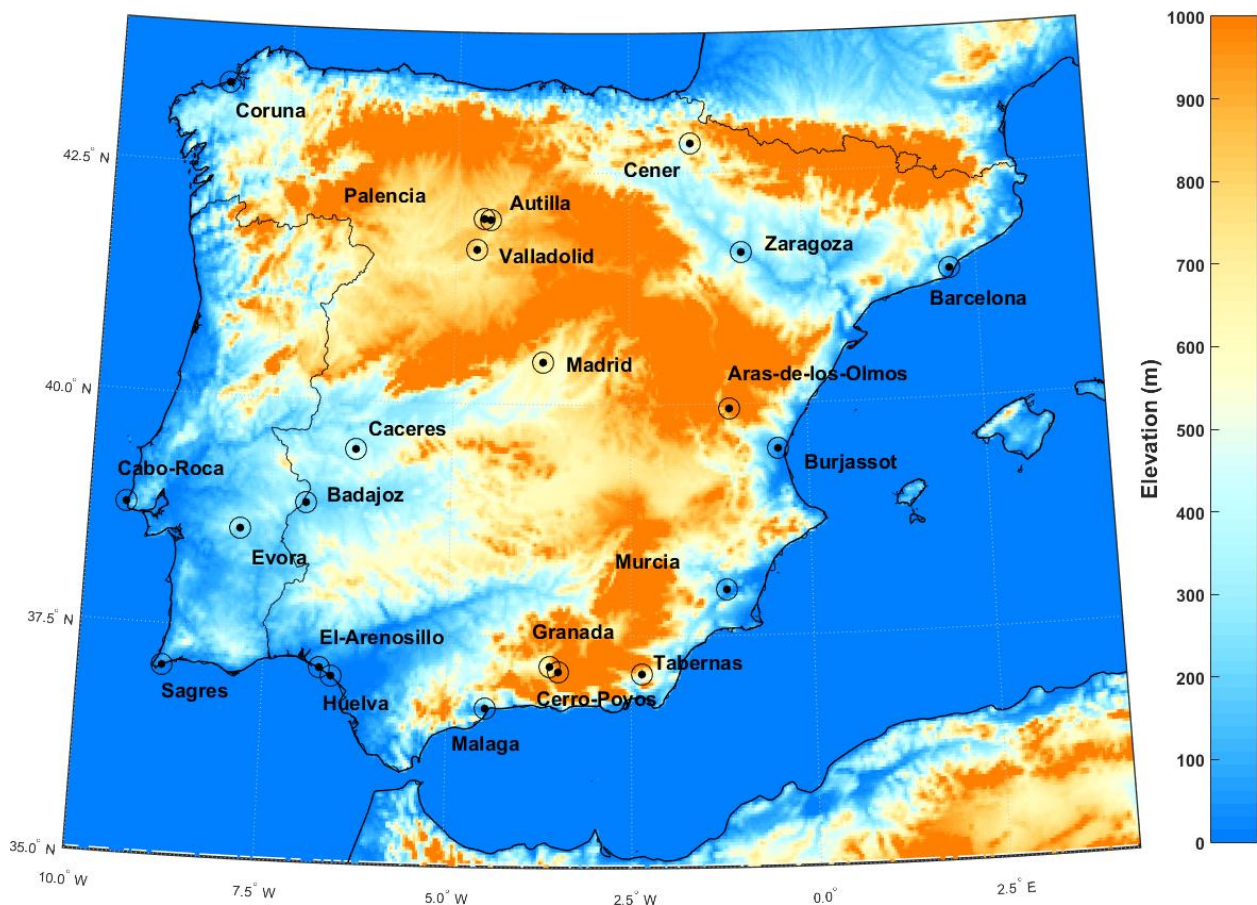
176 The occurrence of Saharan dust events on the Iberian Peninsula usually peaks in March and
177 June, with a marked minimum in April and lowest occurrence in winter according to Cachorro
178 et al. (2016). Depending on the synoptic conditions and circulation patterns, dust transport
179 can affect the entire peninsula (Toledano et al., 2007). The prevailing westerlies, blowing
180 from west to east, are the dominant wind pattern over the Iberian Peninsula. These winds
181 are most prominent in the mid-latitudes, including the Iberian Peninsula. More pronounced
182 in the northern region during autumn and winter, these winds bring moist air from the
183 Atlantic, increasing precipitation in Galicia, the Basque Country, and northern Portugal.
184 While they also affect central and southern areas, their impact is moderated by the
185 peninsula's topography and other wind systems. The northeast trade winds affect the
186 southern and western coasts of Portugal and southwestern Spain, creating a mild and dry
187 climate, especially in summer. In contrast, Mediterranean winds affect the eastern and
188 southeastern coasts. Additionally, the Iberian Thermal Low, resulting from intense heating
189 of the Iberian interior, creates a low-pressure area that draws air from the Atlantic and
190 Mediterranean shores, leading to converging wind patterns. This circulation pattern
191 enhances sea breeze penetration and moderates coastal temperatures. Southern Spain is
192 influenced by the Sahara winds, as said, these dry winds often carry dust, increasing the
193 temperature and reducing air quality. Calima is a type of wind that occurs when Saharan dust
194 reaches the peninsula, especially in summer, causing hazy skies, a reddish tint, and low
195 visibility. These winds are linked to high-pressure systems over North Africa and low-
196 pressure systems over the western Mediterranean.

197 The wind circulation over the Iberian Peninsula is a dynamic and complex system shaped by
198 global atmospheric patterns, regional geography, and local topography. The interaction of
199 prevailing westerlies, trade winds, Mediterranean breezes, and local wind systems creates a
200 diverse wind regime that affects the peninsula's climate. Understanding these patterns is
201 essential for weather prediction, agriculture management, and tackling environmental
202 challenges. According to Cachorro et al. (2016), these complex and contrasting influences of
203 air masses from the Atlantic Ocean, Mediterranean Sea, European continent, and North Africa
204 lead to a large spatio-temporal variability in aerosol properties, types, and mixing processes
205 over the Iberian Peninsula. This makes the peninsula a challenging region for online
206 modeling of aerosol microphysical properties and mixing state, therefore an interesting
207 region to evaluate **observation-based approaches, such as those based on climatological
208 aerosol intensive optical properties from AERONET(Li et al., 2019; Fan et al., 2020; Zhou et
209 al., 2023).** Online modeling of optical properties involved in radiative interactions is usually

210 calculated as a mass-weighted average of individual species at each grid point. This
211 assumption of external mixing may not always be accurate, leading to significant
212 uncertainties, such as excessive absorption by smoke aerosols and inaccuracies in dust size
213 fractions. Observation-based approaches, such as those provided by AERONET retrieval
214 climatology, attribute intensive optical properties to an effective aerosol based on actual
215 observations. This method aims to reduce the uncertainties arising from the explicit
216 simulation of these properties in climate models. The Iberian Peninsula, influenced by a
217 variety of aerosol types—including dust, smoke, urban-industrial emissions, and marine
218 aerosols—presents an interesting region to test this hypothesis.

219

220



221

222 *Fig. 1: AERONET sites locations displayed on top of the Iberian Peninsula topography.*

223

224 2.2 AERONET aerosol inversion product

225 AERONET is a global ground-based network of sun photometers mainly aimed at
226 characterizing columnar aerosol particle properties (Holben et al., 1998). From the direct
227 Sun attenuation measurements, AERONET algorithms derive spectral Aerosol Optical Depth

228 (AOD_λ) at the wavelengths 0.34, 0.38, 0.44, 0.50, 0.67, 0.87, 0.94, and 1.02 μm. The interval
229 between direct sun measurements is typically 15 minutes, but only cloud-free conditions are
230 considered for aerosol retrievals. From the spectral dependency of AOD at these
231 wavelengths, AERONET provides Angstrom Exponent (AE), a parameter sensitive to the
232 aerosol particle size distribution (Eck et al., 1999). AERONET also provides several other
233 intensive properties that depend not on the amount but on the nature of the aerosol, related
234 to particle size, shape, and composition, from sky radiance measurements up to nine times a
235 day at the wavelengths 0.44, 0.67, 0.87, and 1.02 μm (Sinyuk et al., 2020). These intensive
236 properties include microphysical parameters, such as refractive indices ($n+ik$) and volume
237 size distribution, and also optical parameters like Single Scattering Albedo (SSA), asymmetry
238 parameter (ASY), Lidar Ratio (LR), Linear Depolarization Ratio (LDR), Angstrom Exponent,
239 among others (Holben et al., 1998; Dubovik et al. 2002). This set of aerosol intensive
240 properties is expected to capture most of the important aspects that differentiate the distinct
241 aerosols' optical regimes that affect the study region. For instance, the imaginary part of the
242 complex refractive index (k) and single scattering albedo (SSA) are properties indicated to
243 separate highly absorbing aerosol regimes from moderate and low absorbing regimes.
244 Angstrom Exponent (AE) and Asymmetry Parameter (ASY) are properties that help separate
245 aerosol regimes characterized by distinct size distributions. LR is highly sensitive to size and
246 composition-related information, for instance, the real part of the complex refractive index.
247 Meanwhile, LDR has high sensitivity to particle morphology, and it is widely used to separate
248 dust particles from other aerosol types. Given the dependency of these intensive properties
249 on the aerosol type (size and composition) and mixed state, it is possible to characterize the
250 aerosol scenarios over a specific AERONET site in terms of their nature and sources (Eck et
251 al., 1999; Dubovik et al., 2002). For instance, the LR is the ratio of the extinction coefficient to
252 the backscatter coefficient and is crucial for identifying different aerosol types. It reflects how
253 light scattering varies with particle size relative to the light wavelength. Small particles, like
254 smoke, have a high LR, while large particles, like sea salt, have a low LR. Therefore, with a
255 well-distributed regional network of AERONET's sun photometers, as that covering the
256 Iberian Peninsula, one can characterize the spatial dynamics of aerosol types and mixture
257 state influencing the regional aerosol regimes. Regarding the time period for the current
258 study, it extends from 2003 to 2023. However, due to calibration and other operational
259 aspects, some AERONET sites present different time ranges within this period.

260 Three key aspects of aerosol nature have been widely used to link aerosol regimes with
261 particle emission sources. These aspects are absorption efficiency, size distribution, and
262 shape (Dubovik et al., 2002). For instance, combustion-based sources, including biomass and
263 fossil fuel burning, produce aerosol dominated by fine mode particles, and absorption ranges
264 from moderate to strong, depending on the nature of biomass burning, fossil fuel, and ageing
265 processes. In contrast, natural sources, such as deserts and marine environments, produce
266 aerosols dominated by coarse-mode particles. Marine aerosol particles are characterized by
267 very low absorption, while dust aerosol can exhibit high absorption, mainly in the UV and VIS
268 bands (Smirnov et al., 2002; Dubovik et al., 2002). Furthermore, the irregular shape of dust
269 particles is a key factor that differentiates them from other aerosol types. This distinctive
270 feature is captured by AERONET retrievals of the LDR (Shin et al., 2018). Source attribution
271 provides valuable insights into the typical intensive optical properties affecting the
272 atmospheric column of a site resulting from complex aerosol state mixtures. This

273 understanding is crucial as it addresses a major challenge that current aerosol modules in
 274 **CMIP6 climate models face (Zhao et al., 2022)**. Reproducing climatological aerosol-intensive
 275 properties scenarios over specific regions has been a major goal of atmospheric models. In
 276 addition to evaluating aerosol modules in atmospheric models, AERONET's optical
 277 properties in typical regimes, which can be expressed as spectral aerosol optical models
 278 (Omar et al., 2005; Levy et al., 2007; Rosario et al., 2013; Zhou et al., 2023), are valuable for
 279 simulating aerosol direct radiative effects in environmental models (Rosario et al., 2013; Li
 280 et al., 2019). This approach is especially beneficial when high computational capacity is
 281 unavailable and explicit aerosol modules are not feasible.

282 **With more than 25 years of operating a vast network of Cimel Electronique Sun-sky**
 283 **radiometers across the world, AERONET has provided highly accurate, ground-truth**
 284 **measurements of aerosol optical depth and other properties (Giles et al., 2019). It has been**
 285 **widely used as the main reference to evaluate and validate satellites (Gupta et al, 2018) and**
 286 **model products (Gloß et al., 2021). The two most critical intensive optical properties to**
 287 **estimating aerosol radiative forcing retrieved by AERONET, single scattering albedo (SSA)**
 288 **and asymmetry parameter (ASY), are related, respectively, to absorption and size of the**
 289 **aerosol. Their accuracies are aerosol loading dependent (Dubovik et al., 2002). For AOD > 0.4**
 290 **at 440 nm (or > 0.2 at longer λ), SSA uncertainty $\approx \pm 0.03$, for lower AOD, uncertainty can be**
 291 **± 0.05 – 0.07 or larger. Regarding ASY, uncertainty is about ± 0.02 – 0.05 when AOD is high (\geq**
 292 **0.4 at 440 nm, ≥ 0.2 at longer wavelengths) but can be significantly larger at low AOD.**

293
 294
 295 Aiming to identify a representative set of typical aerosol regimes that affect the Iberian
 296 Peninsula, we applied cluster analysis methods (described in Sec. 2.4) to the AERONET **sky**
 297 **radiance retrievals dataset from 2003 to 2023**, taking advantage of the extensive coverage of
 298 AERONET sites across the region. **Table 1** presents a set of intensive properties provided by
 299 AERONET that was used to identify typical aerosol scenarios in the Iberian Peninsula
 300 atmospheric column. The variables displayed cover all three previously mentioned aspects,
 301 absorption efficiency, size distribution, and shape, which are expected to characterize the
 302 distinct nature of aerosol types and mixture anticipated in the study region.

303
 304 **Table 1:** List of AERONET inversions products (variables) used in clustering process
 305 **followed by their abbreviation as defined by AERONET.**

Variables	Abbreviation
Refractive Index - Real Part	RI _{Real} (440), RI _{Real} (670), RI _{Real} (870), RI _{Real} (1020)
Refractive Index - Imaginary part	RI _{Imag} (440), RI _{Imag} (670), RI _{Imag} (870), RI _{Imag} (1020)
Single Scattering Albedo	SSA(440), SSA(670), SSA(870), SSA(1020)
Asymmetry Parameter	ASY(440), SSA(670), SSA(870), SSA(1020)

Linear Depolarization ratio	LDR(440), LDR(670), LDR(870), LDR(1020)
Lidar Ratio	LR(440), LR(670), LR(870), LR(1020)
Fine and Coarse modes Volume median radius	VMR-F,VMR-C
Standard deviation from volume median radius, for Fine and Coarse modes	STD-F, STD-C
Fine and Coarse modes Effective radius	Reff-F, Reff-C

306

307 We selected only AERONET sites that operated for at least two years and that have sky
308 radiance inversion available with the highest quality level 2.0. Some selected sites are still
309 operational, while others have been discontinued. **Figure 1** illustrates the geographical
310 distribution of the chosen sites. Our selection encompasses various landscapes of the Iberian
311 Peninsula, from coastal plains regions (Coruña, Sagres, Burjassot) to highland plateaus in the
312 interior (Madrid, Valladolid, Aras-de-los-Olmos) and lowland valleys (Zaragoza, Murcia).
313 Regarding external air mass influence, sites in the southern border of IP are typically the first
314 to experience the transport of dusty air masses from North Africa, with locations such as El-
315 Arenosillo, Huelva, Malaga, and Sagres affected. The eastern sites (Barcelona, Burjassot, and
316 Murcia) are expected to be strongly influenced by the Mediterranean air masses. Western
317 and northern sites (Cabo da Roca, Coruna, Sagres) are directly under the influence of air mass
318 from the Atlantic Ocean. Additionally, the Portuguese countryside (Evora) and Spain's
319 eastern sites (Badajoz, Caceres) are located in regions that very often experience biomass
320 burning during the dry season (Ermitão et al., 2023; Silva et al., 2023; Hammed e tal., 2024;
321 Alvares et al., 2024).

322

323 2.3 Merra-2 Aerosol Diagnostic Product

324 The MERRA-2 (Modern-Era Retrospective Analysis for Research and Applications, Version 2)
325 Aerosol Diagnostic Product (ADP) is a comprehensive dataset provided by NASA that offers
326 global information about atmospheric aerosols (Gelaro et al., 2017; Buchart et al., 2017).
327 MERRA-2 combines observational data with numerical models(reanalysis project) to create
328 a detailed long-term record of atmospheric dynamics and composition from 1980 to the
329 present. Among other variables, the MERRA-2 ADP product offers a long-term view of
330 aerosol mass distribution by types and the related optical properties (Buchart et al., 2017).
331 Its extended temporal coverage allows analysis of aerosol trends, such as those related to
332 changes in atmospheric composition due to human activity and the impact on climate. Key
333 features of the MERRA-2 ADP include aerosol microphysical and optical properties such as
334 optical depth, mass concentration, and size distribution. These properties are crucial for
335 understanding aerosol loading and composition in the atmosphere and their role in the
336 Earth's radiation budget and climate system. A key aspect of MERRA-2 APD for this study is
337 that it provides aerosol-type column mass density, our target variable as a predictor of

338 aerosol optical model regime. The MERRA-2 APD includes diagnostics for the aerosol types
339 considered in most chemistry transport models: Dust (DT), Black-Carbon (BC), Organic
340 Carbon (OC), Sea-Salt (SS), and Sulfate (SF). The aerosol-type diagnostics variables cover
341 mass concentration at specific levels and are integrated in the entire atmospheric column,
342 which are applied to estimate columnar optical properties, such as extinction, scattering, and
343 absorption optical depths, at multiple wavelengths. For this study, the 550 nm wavelength
344 was used as a reference. Optical properties are a function of aerosol species, particle size, and
345 relative humidity. To convert from the simulated aerosol masses to optical quantities such as
346 aerosol optical depth, MERRA-2 uses Optics look-up tables (LUTs) derived from Mie
347 calculations using parameters from the Optical Properties of Aerosols and Clouds (OPAC;
348 Hess et al., 1998) , as described in Chin et al. (2002) and Colarco et al. (2010), except for dust-
349 type aerosol, which is based on Colarco et al. (2014). Therefore, these optical properties are
350 by-products of running the MERRA-2 reanalysis system and made available to the
351 community via MERRA-2 ADP. Further details on this can be found in Buchard et al. (2017).
352 From these extensive aerosol-driven optical properties, it is possible to derive several
353 MERRA-2 ADP intensive optical properties, such as Single Scattering Albedo (SSA).

354
355 Given that the aerosol optical properties retrieved from each AERONET site are influenced
356 by mixtures of different aerosol types, it is reasonable to assume that the impact of each
357 aerosol type on the column's intensive optical properties is primarily determined by its
358 concentration. Based on this premise, we propose a machine-learning approach that utilizes
359 the aerosol-type column mass density predicted by chemistry transport models to prescribe
360 the most accurate possible spatial distribution of the aerosol spectral optical model
361 developed through cluster analysis of AERONET data. A description of the method presented
362 in this study, exploring MERRA-2 products, can be found in subsection 2.5.

363

364 2.4 Optical models development: Cluster Analysis

365 Cluster analysis has been extensively used to develop aerosol optical models based on
366 AERONET sky inversion products (Omar et al., 2005; Levy et al., 2007; Russel et al., 2014). The
367 underlying principle is that AERONET instantaneous retrievals can be grouped into a certain
368 number of clusters, each representing different categories of aerosol regimes. These studies
369 have explored mainly the K-means clustering method, one of the most popular unsupervised
370 machine learning algorithms for partitioning a dataset into a pre-defined number of clusters.
371 However, specifying the number of clusters in advance poses a significant challenge for the
372 K-means method. Fortunately, there are techniques available that minimize the subjectivity
373 involved in this pre-definition. In our study, we adopted the Elbow method (Shi et al., 2021),
374 a widely used method for determining the optimal number of clusters (k) in a K-Means
375 clustering algorithm. It examines the relationship between the number of clusters and the
376 within-cluster sum of squares (WCSS), which measures the variance within each cluster (Eq.
377 1)

$$378 \quad \text{WCSS} = \text{minimize}(\sum_{k=1}^k W(C_k)) \quad (1)$$

379 where C_k is the k th cluster and $W(C_k)$ is the within-cluster variation. The total within-cluster
380 sum of squares (WCSS) measures the compactness of the clustering, and one wants it to be
381 as small as possible. If clusters are tight and well-separated, WCSS will be small, because
382 points are close to their centroids. If clusters are loose or overlapping, WCSS will be large.
383 The Elbow method is based on a plot of WCSS against the number of clusters. As the number
384 of clusters increases, WCSS always decreases; more clusters mean tighter groups. At some
385 point, the decrease becomes less significant, displaying a feature like an “elbow”. We ran our
386 clustering algorithm with k varying from 2 to 10 clusters. For each k , we calculated the total
387 WCSS. The k results against WCSS were displayed in a plot, and the optimal number of
388 clusters was defined based on the location (k) of the bend (elbow) in the plot.

389 The Elbow method has been widely used because of its straightforward approach to
390 estimating the most appropriate number of clusters. However, we recognize that it still
391 carries a certain degree of subjectivity, as it relies on visual interpretation. To reduce this
392 subjectivity, we combined the Elbow and Stability methods to evaluate the optimal number
393 of clusters that best represent the major aerosol regimes affecting the study region. Although
394 more rigorous methods are available in the literature, defining the number of clusters
395 remains a challenge, and different approaches often lead to distinct solutions (Krishnaveni et
396 al., 2023). Nevertheless, despite the limitations of the Elbow method, the number of clusters
397 identified in our study seems to provide a coherent characterization of optical regimes
398 affecting the Iberian Peninsula.

399

400

401 2.5 Optical models spacial prescription: Random Forest Technique

402 Once the optimal number of clusters is defined, which corresponds to the expected number
403 of major optical properties regimes to influence the study region, and the clustering process
404 is performed, each cluster is characterized by a set of AERONET instantaneous retrievals of
405 optical and microphysical properties that are expected to express an optical property regime.
406 Also, each AERONET instantaneous retrieval is tagged with the cluster number that it
407 belongs to. By averaging the instantaneous properties of each cluster, we set the reference
408 values that represent the mentioned major aerosol optical properties regimes.

409 We propose a machine-learning approach that utilizes the well-known random forests
410 supervised algorithm (Breiman, 2001) to spatially represent the aerosol optical models
411 defined by the cluster analysis for each AERONET site (described in section 2.4). The
412 implemented method was tested using aerosol column mass density data from MERRA-2
413 (Table 2) to establish the spatial distribution of the optical regime defined by the cluster's
414 average. This approach is also suitable for chemistry transport models.

415 MERRA-2 time series of column mass density for each aerosol type (DT, BC, OC, SS, SF) over
416 each AERONET site were collocated with the network inversion products used to derive the
417 clusters representing the distinct aerosol regimes over the Iberian Peninsula (described in
418 section 2.4). Merra-2 column mass densities are available with a frequency of 1 hour, while
419 AERONET optical properties instantaneous retrievals are provided at irregular times, due to

420 its dependence on cloud cover and AOD criteria (AOD at 440 nm > 0.4). So, for each AERONET
421 retrieval, our script searches for the MERRA-2 closest hour to synchronize the two datasets.
422 Each AERONET instantaneous aerosol microphysical and optical properties inversion
423 retrieval (Sinyuk et al., 2020) was connected to the corresponding cluster to which it
424 belonged. Likewise, each instantaneous aerosol microphysical and optical properties
425 inversion retrieval was also connected to the closest in-time combination of MERRA-2 data
426 of aerosol-type column mass density (DT, BC, OC, SS, SF). With this, we built a time series of
427 collocated MERRA-2 aerosol types of column mass density with the developed clusters
428 occurrences over each AERONET site, which was used in a training process aiming to predict
429 the suitable cluster given a specific combination of aerosol types of column mass density
430 predicted.

431 Therefore, the first step was to split the data into training (70%) and test (30%). The
432 algorithm uses training data to learn the relationship between the combination of aerosol-
433 types columnar mass density and the target, which are the developed clusters from
434 AERONET aerosol-intensive properties. The training was done using the Random Forest
435 Classification algorithm (RandomForestClassifier) from the Python package Scikit-Learn
436 (Abraham et al. 2014). The Random Forest classifier's hyperparameters were optimized
437 using RandomizedSearchCV, a stochastic method of parameter space exploration. The
438 parameter space included the number of decision trees (n_estimators: 50–500) and the
439 maximum depth of trees (max_depth: 1–20). The process used stratified k-fold cross-
440 validation to ensure representative sampling across aerosol regime classes. This
441 optimization method addressed the issues of class imbalance and aerosol regime
442 classification in atmospheric measurements. The imbalance refers to the unequal
443 representation of aerosol regimes in the dataset, where some clusters occur much less
444 frequently than others. This imbalance is typical in atmospheric aerosol measurements,
445 where extreme but radiatively important aerosol events like intense smoke episodes are rare
446 compared to more common background conditions.

447 The random search methodology was used to find parameter combinations inside the
448 parameter space without the processing demands of grid search. Cross-validated
449 performance indicators were used to select the final configuration in order to reduce
450 overfitting and ensure consistent performance across aerosol regimes. The confusion Matrix
451 was used to visualize the performance of the models, and we also calculated the following
452 indicators: Accuracy, Precision and Recall, and F1 score. Accuracy represents the number of
453 correctly classified data instances over the total; it checks the predictions against the actual
454 values in the test set and returns the percentage of times the model got right.

455 Precision and recall are two critical metrics for evaluating the performance of a classification
456 model. Precision is the proportion of true positives among all the predicted positive cases
457 (true and false), meaning it measures the accuracy of positive predictions (Eq. 2). Recall is
458 the proportion of true positives among all actual positive cases (true and false), meaning it
459 measures the model's ability to identify positive cases (Eq. 3). The F1 score, the harmonic
460 mean of a model's precision and recall, takes both precision and recall and provides a more
461 balanced measure of a model's performance (Eq. 4). The F1 score is set to be a value between
462 0 and 1, indicating, respectively, poor precision and recall and high precision and recall,
463 which is ideal.

464

465 Precision = True positive/(True positive + False positive) - (2)

466 Recall = True positive/(True positive + False negative) - (3)

467 F1 = 2 × (Precision × Recall)/(Recall + Precision) - (4)

468

469 **Table 2:** Predictor variables from Merra-2 (aerosol-type column mass density) used in the
470 machine learning process to prescribe the aerosol optical regime (optical model).

Variables	Abbreviation	Unity	Spatial resolution
Dust column mass density	DUCMASS	kg/m ²	0.5° × 0.625°
Black carbon column mass density	BCCMASS	kg/m ²	0.5° × 0.625°
Organic carbon column mass density	OCCMASS	kg/m ²	0.5° × 0.625°
SO ₂ column mass density	SO2CMASS	kg/m ²	0.5° × 0.625°
SO ₄ column mass density	SO4CMASS	kg/m ²	0.5° × 0.625°
Sea salt column mass density	SSCMASS	kg/m ²	0.5° × 0.625°

471 3. Results

472 The results section is divided into three subsections. The first one presents the results of
473 identifying the typical aerosol optical regimes affecting the Iberian Peninsula using cluster
474 analysis. The second subsection discusses the results and the performance of spatial
475 prescription of these typical aerosol regimes by applying machine learning (Random Forest)
476 to the columnar density of MERRA-2 aerosol components. Finally, case studies applying the
477 method developed are presented and discussed.

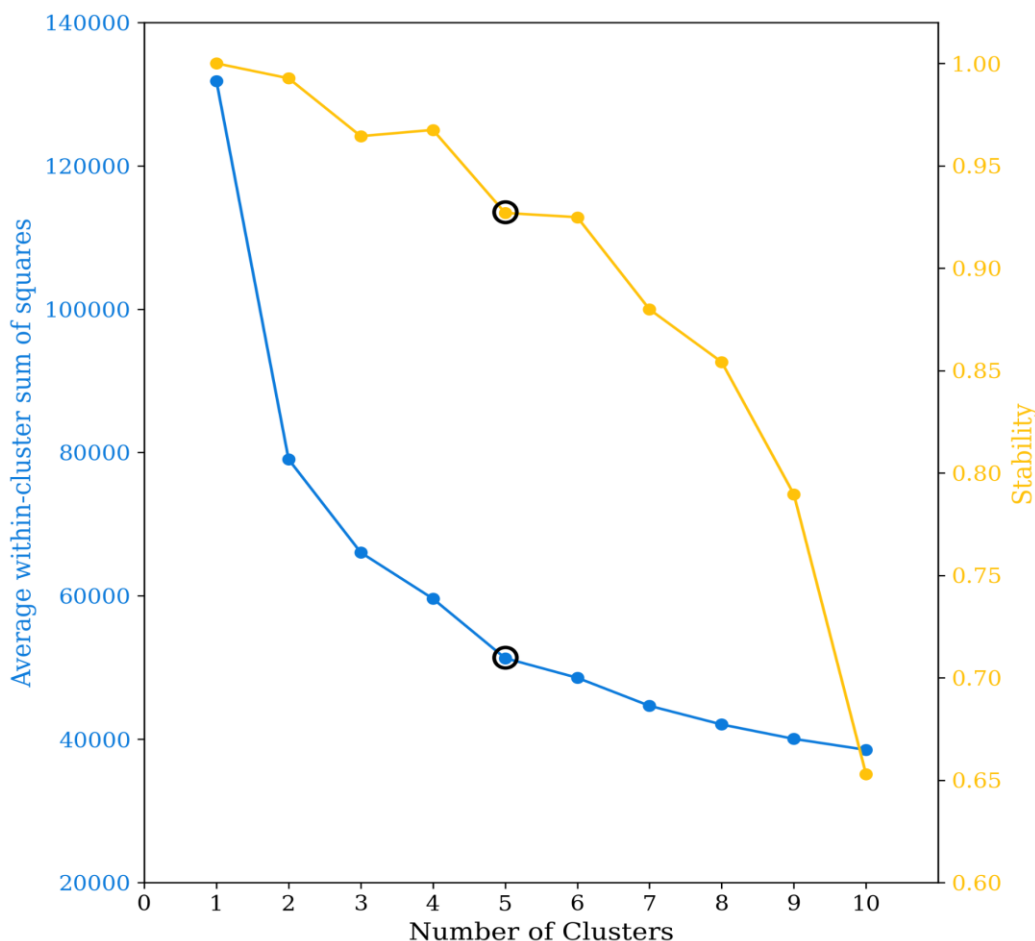
478

479 3.1 Cluster Analysis: Optical models development

480 The number of clusters (k) selected to characterize the typical optical aerosol regimes over
481 the Iberian Peninsula was defined based on the Elbow method (**Figure 2**), which indicated
482 five clusters were the optimal number to capture the aerosol regime variability. We also
483 evaluated from the Elbow method that there is a sharp bending at $k=2$, which we associated
484 with a clustering separation between aerosol regimes strongly dominated by coarse mode,
485 dust regimes, and regimes dominated by fine mode, non-dust regimes. However, to cover
486 more specific regimes within these two macro-regimes (dust regimes vs non-dust regimes),
487 a higher k is required, and $k=5$ is revealed to be the second sharpest bending. Cluster stability
488 as a function of the number of clusters was also evaluated as a way of evaluating whether the
489 clusters obtained are meaningful and not just artifacts of randomness or noise. High stability

490 suggests clusters represent real structure in the data, not just random fluctuations. The
491 stability for k=5 is above the 90% threshold, similar to k=6, a number after which stability
492 sharply decreases. Therefore, combining the Elbow method and stability reinforced k=5 as
493 an optimal cluster number to capture the typical aerosol scenarios over the Iberian
494 Peninsula, reducing the subjectivity usually associated with the K-means clustering method.

495
496



497

498 **Figure 2:** Average of sum of squares within-cluster and cluster stability as function of the
499 number of clusters.

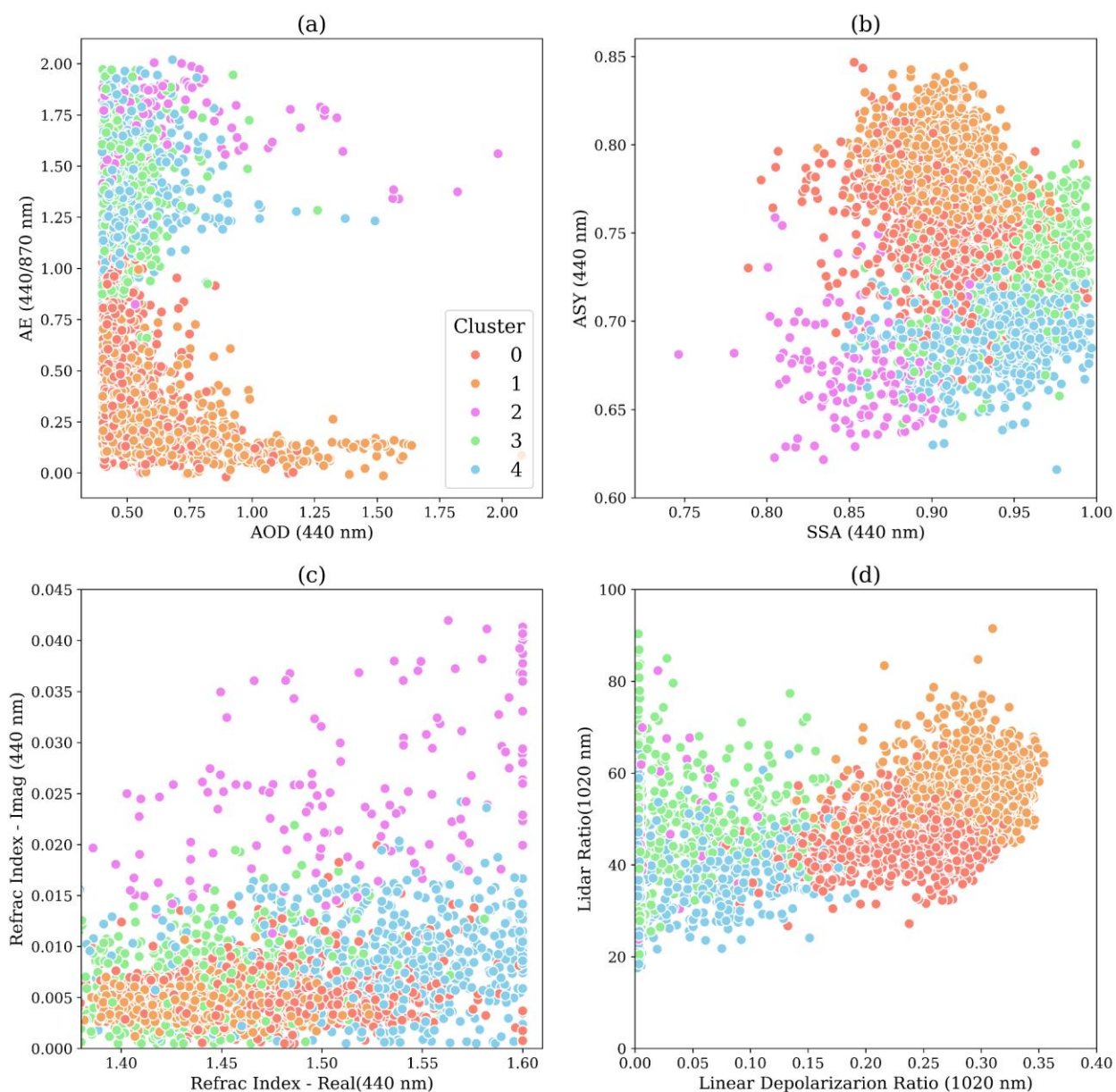
500

501 We applied the cluster analysis once we defined the optimal number of clusters. **Figure 3**
502 presents a combination of graphics used for aerosol properties analysis, highlighting the
503 obtained clusters' behavior and distinction. The first graphic (Fig. 3a) represents the Aerosol
504 Optical Depth (AOD) as a function of Angstrom Exponent (AE), which allows us to relate
505 aerosol loading variability with aerosol regimes dominated either by coarse or fine mode
506 (Eck et al., 1999). This analysis shows that two of the clusters (C0 and C1) are regimes

507 dominated by coarse mode particles ($AE < 1.0$), while the remaining three (C2, C3, and C4)
508 are regimes under the stronger influence of fine mode particles ($AE > 1.0$). The second plot
509 displays the asymmetric parameter against the single scattering albedo at 440 nm. This plot
510 aims to elucidate the clusters' distinctions related to particle absorption efficiency and the
511 asymmetry between hemispherical forward and backward scattering. Aerosol regimes
512 dominated by coarse particles tend to exhibit more significant forward scattering and,
513 consequently, higher asymmetry parameter values. In contrast, lower asymmetry parameter
514 values are expected in fine mode regimes (Eck et al., 1999; Dubovik et al., 2002). This pattern
515 is evident in the graphic; clusters C0 and C1 present higher asymmetry parameter values. It
516 is also possible to identify the distinction between the non-dust regimes C2, C3, and C4. C2
517 presents the lowest asymmetry parameter values, while it is the most absorbing of the
518 clusters, according to its single scattering albedo values. Small and highly absorbing particles
519 are commonly associated with urban pollution or fresh smoke plumes from biomass burning
520 (Dubovik et al., 2002; Omar et al., 2005; Levy et al. 2010; Martins et al., 2009). The C3 cluster
521 differs significantly from C2 by presenting higher asymmetry parameter values, an indication
522 of a shift to larger particle sizes. C3 has higher single-scattering albedo values, indicating a
523 less absorbing aerosol regime. SSA alone did not help to differentiate the two clusters
524 dominated by coarse mode particles (C0 and C1). C0 asymmetry parameter values tend to be
525 lower than those of C1, suggesting that the former could be a dusty mixture not as close to a
526 pure dust scenario as C1. The traditional plot of Lidar Ratio (LR) against Linear
527 Depolarization Ratio (LDR) (Kanitz et al. 2013, Illingworth et al., 2015) confirms this
528 hypothesis (Fig. 3d). Pure dust regimes of aerosol, due to their high level of non-spherical
529 particles, produce higher LDR (Groß et al., 2011). The C1 cluster presents higher values of
530 LDR than C0, indicating that C1 is closer to a pure dust regime. The C0, while a dust regime,
531 is likely to represent a mixed scenario given its LDR values consistent with dust and smoke
532 mixing (Kanitz et al. 2013). LDR values below 15%, which is the case of the clusters C2, C3,
533 and C4, are typically associated with fresh/aged smoke, urban-industrial pollution, and
534 marine particles scenarios. The analysis of the real part versus the imaginary part of the
535 complex refractive index (Fig. 3c) emphasizes C2 as the aerosol regime with the largest
536 absorption and highlights that the real part of the complex refractive index is the main aspect
537 differentiating C3 and C4.

538

539



541

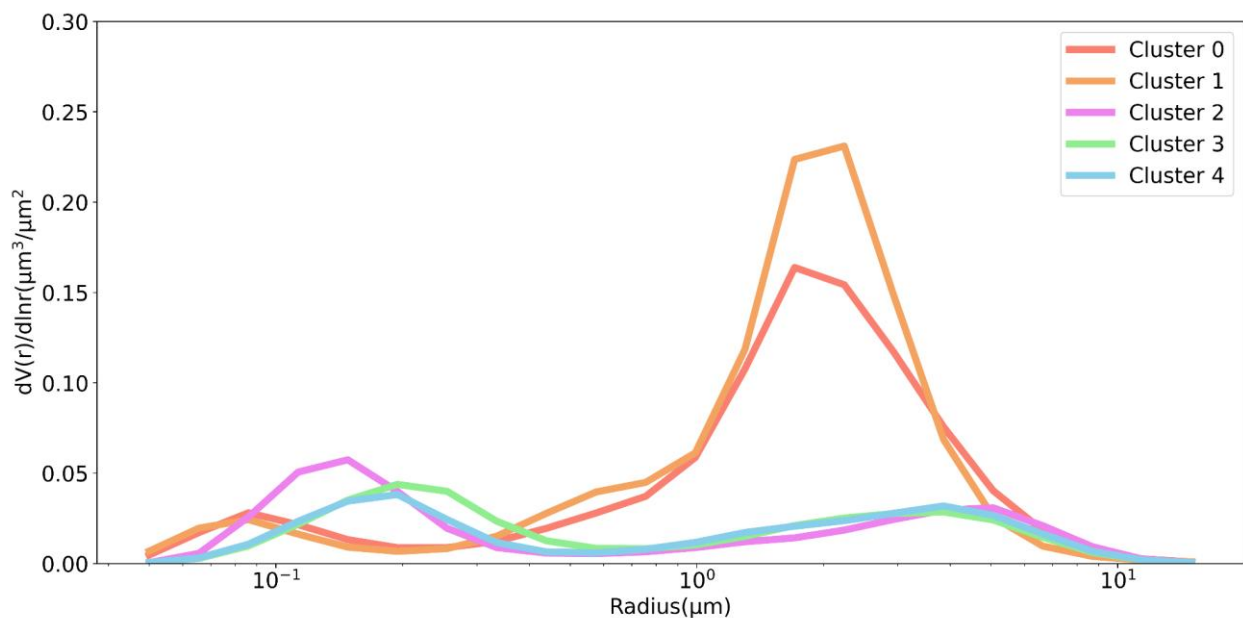
542 **Figure 3:** Scatterplot of the clusters elements as function of different parameters: (a) Extinction
 543 Angstrom Exponent (AE) as function of Aerosol Optical Depth (AOD) at 440 nm; (b) Asymmetry
 544 Parameter (ASY) as function of Single Scattering Albedo (SSA) at 440 nm; (c) Lidar Ratio as a function
 545 of Linear Depolarization Ratio at 1020 nm; (d) Refractive index at 440 nm: Imaginary part as function
 546 of Real part.

547

548 **Figures 4 and 5** present the clusters averages for selected features: size distribution,
 549 complex refractive index, single scattering albedo, and asymmetry parameter. A more
 550 detailed summary of the mean behavior of the clusters is presented in **Table 3**. The average
 551 size distribution of the clusters confirms that aerosol regimes affecting the Iberian Peninsula

552 vary between two scenarios dominated by coarse mode (C0, C1), named here as dust regimes,
 553 and three scenarios when coarse mode is not dominant, here considered as non-dust
 554 regimes. There are differences between the dust regimes: C1 is associated with a higher
 555 coarse particle loading than C0. Among the non-dust regimes (C2, C3, and C4), the main
 556 difference is seen between C2 and the other two. C2 is characterized by a larger fine particle
 557 loading. Between C3 and C4, one can observe a larger radius spread for C3 regarding the
 558 contribution of the fine mode, which can be inferred from a potential growth of particles via
 559 processes such as water uptake, aging, and coagulation, or that the aerosol regime mixture
 560 includes sources that naturally produce slightly larger fine particles. These features usually
 561 indicate more aged, more hygroscopic, or more humidified aerosol compared to freshly
 562 emitted, dry fine particles.

563



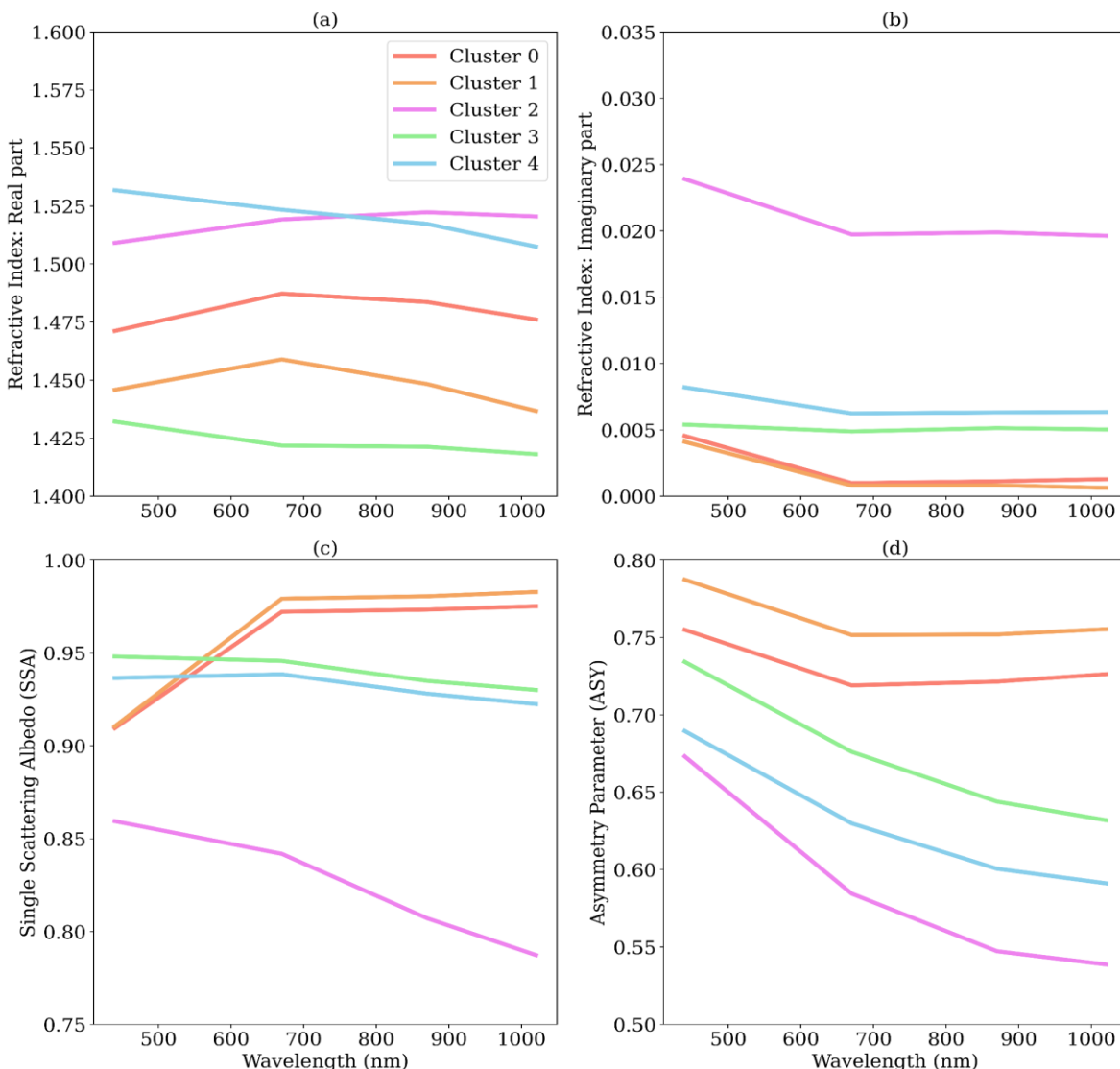
564 **Figure 4:** Clusters mean volume particle size distribution as a function of radius. These size
 565 distributions correspond to the average of the instantaneous size distributions retrieved by
 566 AERONET from each identified cluster. The numeric values of each cluster size distribution can
 567 be found in Table S2 in the supplement.

569

570 Clusters C2 and C4 have close values for the real part of the refractive index, but cluster C2
 571 has a much larger imaginary part, justifying its lowest SSA (Figure 5). The C2 strong
 572 absorption combined with its smaller particles suggests that it is likely associated with fresh
 573 smoke (Reid et al., 1998; Reid et al., 2005). The average of the real part of the complex
 574 refractive index corroborates the difference between the C3 and C4 aerosol regimes.
 575 According to Moise et al. (2015), a variation as such observed between C3 and C4 (1.4 to 1.5)
 576 could produce an increment of 12 % in estimating the direct aerosol radiative forcing over
 577 the solar spectrum wavelength range. Zhao et al. (2019) also showed that the direct aerosol
 578 radiative forcing is estimated to vary by 40 % when the real part of the complex index values
 579 varies between 1.36 and 1.56. The reasons for the differences observed between the real

580 parts of C3 and C4 remain unclear. However, the spatial distribution of the clusters (see Fig.
 581 6) indicates that C3 is more prevalent in the eastern region of the Iberian Peninsula, which is
 582 the wettest area and more exposed to air masses from the Mediterranean and Eastern
 583 Europe. Additionally, the low values of the real part of the complex refractive index for C3
 584 align with aerosol regimes that have a strong contribution from sulfate particles. The spectral
 585 dependency of the single scattering albedo corroborates our attribution of the C0 and C1 to
 586 a dust regime. Dust particles are characterized by strong absorption in the UV spectrum
 587 (Dubovik et al., 2002), which decreases as the wavelength increases, a feature present in both
 588 C0 and C1. Also, consistent with dust-dominated regimes, C0 and C1 have the largest mean
 589 asymmetry parameter at all wavelengths.

590



591

592 **Figure 5:** Clusters average of complex refractive index, (a) Real and (b) Imaginary parts, (c)
 593 single scattering albedo and (d) asymmetry parameter.

594

595 The analysis above and the summary provided by **Table 3** provide several specific
596 characteristics that help us to contextualize the clusters. To enhance this understanding, we
597 add the spatial (**Figure 6**) and seasonal (**Figure 7**) distribution of the clusters into our
598 analysis. C0 and C1 aerosol regimes are dominated by dust, where C1 is the closest regime
599 to what we could call a pure dust scenario. Both aerosol regimes, C0 and C1, affect practically
600 the entire Peninsula (**Figure 6**) and all year round, but it is more frequent in the southern
601 part of the Peninsula, an expected feature considering that the dust particles are mainly
602 transported from North Africa (Cachorro et al., 2016; Gómez-Amo et al., 2017). The C2
603 cluster is the most absorbing regime, and is characterized by the smallest fine mode particles
604 (**Table 3**). Its spatial distribution (**Figure 6**), with more frequent occurrence along the belt
605 spanning from Evora, in Portugal, to Caceres, in Spain, a region known for high recurrence of
606 biomass burning, reinforces our hypothesis. Additionally, the seasonal distribution of C2 in
607 this region coincides with the peak of the biomass burning season. The C3 aerosol regime
608 also occurs over **all AERONET sites throughout the year**, but it is dominant in the eastern and
609 northeastern portions of the Iberian Peninsula. Among non-dust regimes, its unique feature
610 is its very low real part of the refractive index. C4, as C3, is weakly absorbing according to its
611 single scattering albedo. **C4** is present across the entire Peninsula, but its occurrence
612 increases in the central and northern portions, which are more prone to biomass burning. An
613 important feature of C4 is that its occurrence increases during the summer and the beginning
614 of autumn (**Figure 7**) in the central region of the Iberian Peninsula, from Évora (Portugal) to
615 Madrid (Spain), when the region's biomass burning season is underway. These aspects led
616 us to hypothesize that C4 is an aerosol regime under the strong influence of smoke aerosol
617 particles.

618
619 **Table 3: Summary of the clusters based on the average of optical and microphysical properties.**
620 *A detailed description of the clusters can be found in Tables S1 and S2 in the supplement. The*
621 *values in the brackets correspond to standard deviation.*

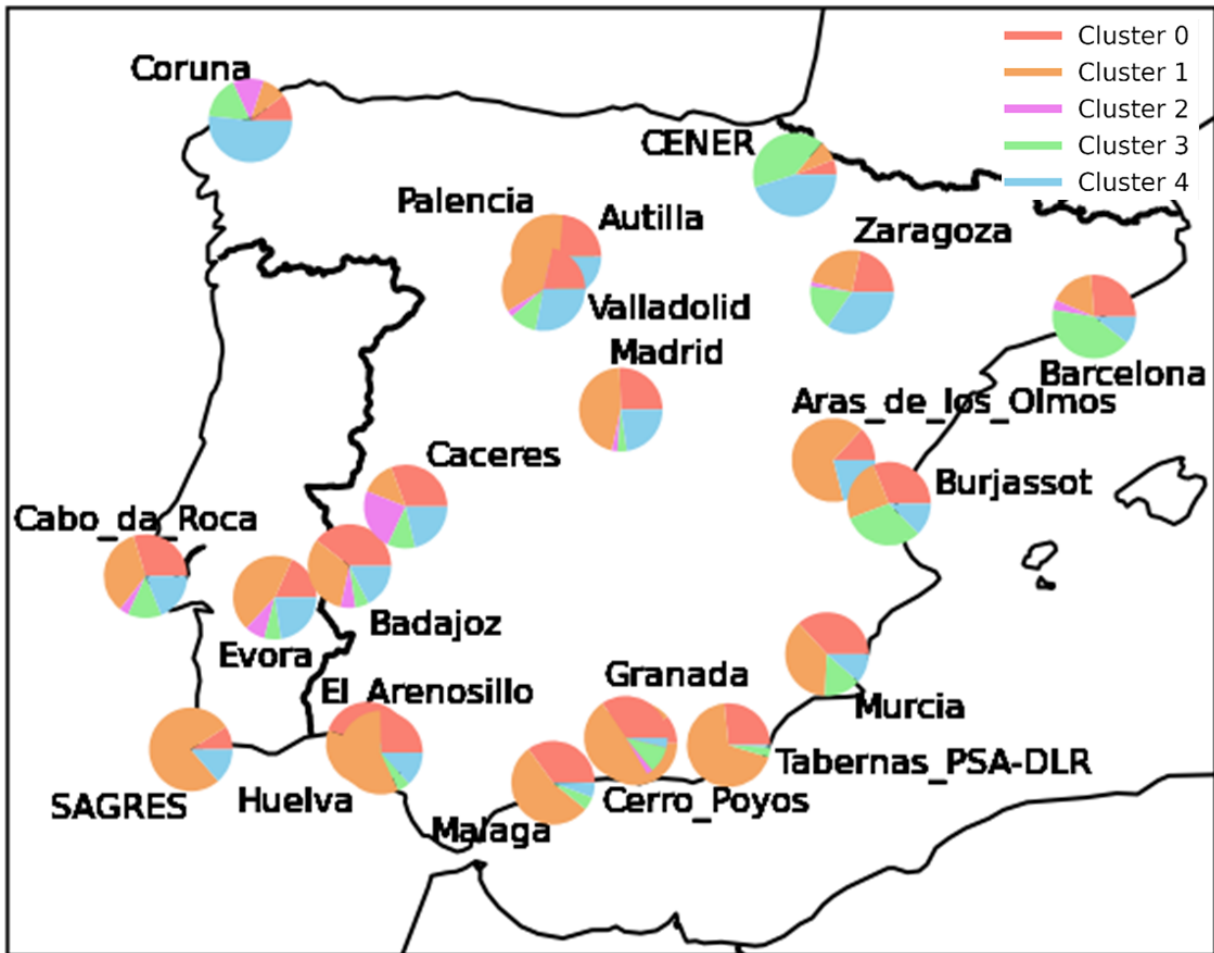
Properties	Cluster0 (Polluted dust)	Cluster1 (Pure dust)	Cluster2 (Strongly absorbing smoke)	Cluster3 (Urban- Industrial Pollution)	Cluster4 (Moderately absorbing smoke)
Number of records	1308	1665	153	660	604
Percentage (%)	29.76	37.88	3.48	15.01	13.74
Ref_idx_Real (440 nm)	1.47(0.04)	1.44(0.03)	1.51(0.07)	1.43(0.06)	1.52(0.05)
Ref_idx_Img (440 nm)	0.005(0.002)	0.004(0.001)	0.025(0.009)	0.006(0.004)	0.009(0.004)
VMR-F	0.14(0.03)	0.14(0.03)	0.16(0.02)	0.21(0.04)	0.18(0.04)
STD - F	0.61(0.09)	0.67(0.07)	0.42(0.06)	0.47(0.06)	0.41(0.05)
REff-F	0.12(0.02)	0.12(0.02)	0.14(0.02)	0.18(0.03)	0.17(0.03)
REff-C	1.68(0.16)	1.61(0.13)	2.44(0.43)	2.31(0.38)	2.25(0.49)
VMR-C	2.02(0.23)	1.88(0.17)	3.10(0.45)	2.82(0.42)	2.82(0.57)
STD-C	0.60(0.52)	0.54(0.04)	0.68(0.06)	0.63(0.05)	0.67(0.05)
AOD (440 nm)	0.50(0.11)	0.58(0.21)	0.64(0.29)	0.48(0.09)	0.51(0.13)
SSA (440 nm)	0.91(0.03)	0.91(0.02)	0.86(0.03)	0.95(0.03)	0.94(0.03)

ASY (440 nm)	0.76(0.02)	0.79(0.19)	0.67(0.03)	0.73(0.03)	0.69(0.02)
AE(440/870 nm)	0.40(0.25)	0.24(0.14)	1.67(0.20)	1.43(0.26)	1.47(0.25)
LR(1020 nm)	64(9)	70(8)	89(16)	77(17)	61(15)
LDPR(440 nm)	0.17(0.04)	0.21(0.04)	0.01(0.03)	0.03(0.04)	0.03(0.05)

622

623

624



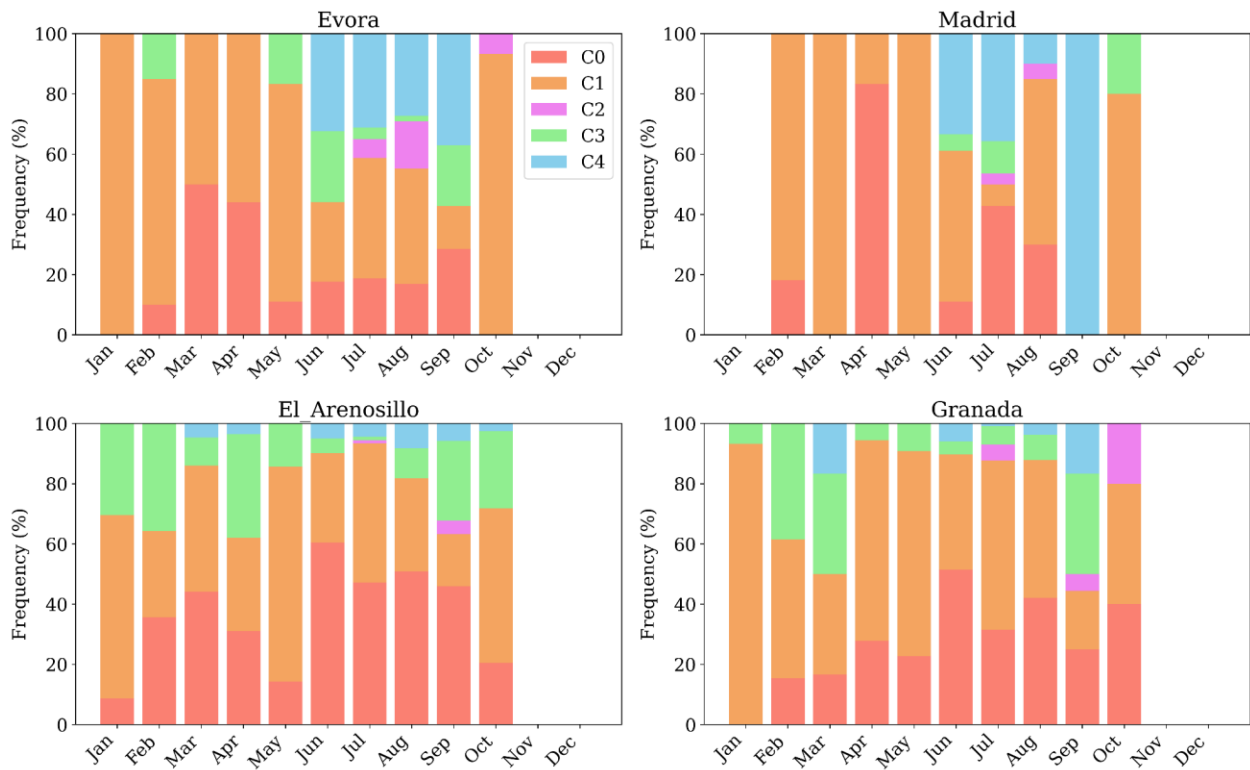
625

626 **Figure 6: Proportions** of the occurrence of the clusters of aerosol regimes at the AERONET
 627 sites across the Iberian Peninsula.

628

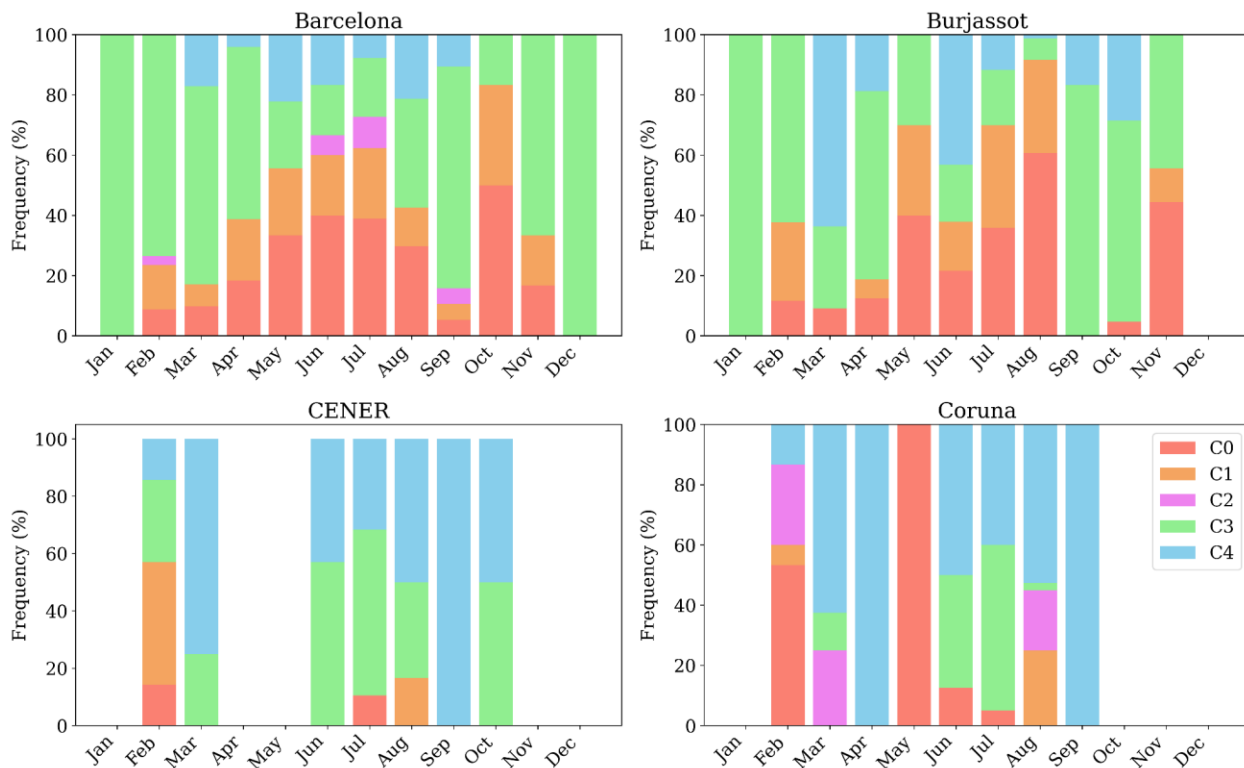
629 **Figures 7 and 8** provide a perspective view on the seasonal occurrence of each cluster based
 630 on sites that represent different regions of the Iberian Peninsula.

631



632

633 **Figure 7:** Clusters relative monthly occurrence over the AERONET sites representative of the
634 Iberian Peninsula western lowlands (Evora), highlands plateau (Madrid) and southeast
635 lowlands (El Arenosillo, Granada).



636
 637 **Figure 8:** Clusters relative monthly occurrence over the AERONET sites representatives of the
 638 following Iberian Peninsula regions: Eastern Coast (Barcelona, Burjassot) and Northern
 639 (Coruna, CENER).

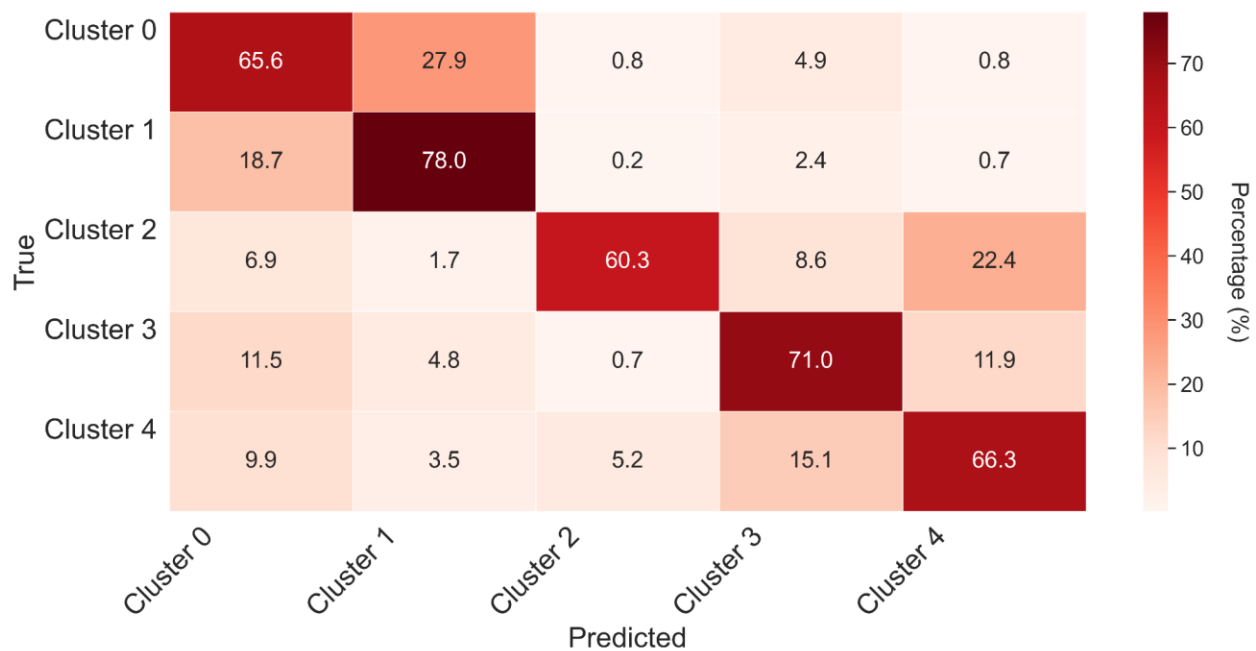
640 **3.2 Random Forest Classifier: Performance and Optical models spatial dynamic**

641 The Random Forest training of MERRA-2 aerosol-type column mass density as predictors of
 642 aerosol optical regime covered 70% of the AERONET sky inversions used in this study, combining
 643 datasets from all sites. The testing dataset, constituted by the remaining 30%, was used to evaluate
 644 the model's performance. The best parameters obtained from the optimization using
 645 RandomizedSearchCV were the number of decision trees of 477 ($n_{estimators} = 477$) and the
 646 maximum depth of trees of 19 ($max_depth=19$). There are several metrics for assessing machine
 647 learning performance. **Figure 9** presents the one used in this study, the Normalized Confusion
 648 Matrix (NCM), which expresses the percentage of correct and incorrect predictions (where
 649 the classifier got confused). In the matrix, the rows represent the true labels, and the columns
 650 represent the predicted ones. The values along the diagonal indicate the percentage of times
 651 where the predicted matches the true label. The other cells reflect instances where the
 652 classifier mislabeled an observation; the column tells us what the classifier predicted, and the
 653 row tells us the correct label.

654 For all clusters, the classifier's correct predictions surpassed the incorrect predictions, with
 655 a maximum frequency of correct prediction close to 80% obtained for C1. The minimum
 656 percentage of correct prediction, about 60%, was obtained for C2, the highest absorbing
 657 cluster. Regarding dust regime clusters, despite the struggle to predict C0, it is possible to see
 658 that, in this case, the classifier's main confusion is with the C1, which is also a cluster related

659 to an aerosol scenario dominated by coarse mode particles (dust regime), as with C0. The classifier's confusion in this case is between the two dust-regime models; therefore, the induced error in radiative transfer calculations would be lower than that if the confusion was between a dust and a non-dust regime, especially like C2, which is substantially different from any of the dust regimes. Rarely does the classifier take either C0 or C1 as C2, C3, and C4, a case where substantial error in the radiative effect would be expected. By combining C0 and C1 results in the NCM, the percentage of correct prediction achieved by the classifier indicating dust regime is higher than 95%. Similarly, the classifier rarely takes C3 and C4 as C0, C1, and C2. Given that C3 and C4 are also close in terms of their optical properties, especially concerning absorption, some degree of confusion among them is expected. Nevertheless, these aspects of the confusion matrix among close clusters are important to identify where the model needs extra training, for instance, considering longer time series when available and adding new and relevant predictors, such as Brown Carbon, an important aerosol component not available in the current MERRA-2 aerosol reanalysis products. C2, the least frequent and the one representing the most absorbing aerosol regime over the Iberian Peninsula, is rarely mislabeled as C0 or C1, but often mislabeled as C3 or C4. Still, the score percentage is around 60%.

676



677

678 **Figure 9:** Normalized confusion matrix of the Random Forest classifier applied to the prediction
 679 of the clusters that describe the typical aerosol optical regime based on MERRA-2 aerosol
 680 components column mass density.

681 To provide further insight into the model performance, we also examined other metrics
 682 commonly used to evaluate Random Forest training. Precision, Recall, and F1 score were
 683 calculated for both scenarios, the trained model applied to the test and to the train dataset
 684 (Table 4). The results indicate that the model generalizes well, without significant

685 overfitting. Even for Cluster 2, which has a small number of occurrences, the model was able
 686 to maintain high precision and score. The general accuracy did not drop critically for the test
 687 data (0.70) when compared with the train dataset (0.88), another indicator of the model's
 688 ability to generalize. The trained model applied to the test dataset achieved a general
 689 accuracy of 70 %, meaning it correctly predicted the aerosol regime in three out of four cases.
 690 For all clusters, all metrics adopted were higher than 0.60, with precision and recall values
 691 exceeding 0.75 in some cases. The precision metric indicates how often the positive
 692 predictions are correct. The model precision varied within the specific optical regimes (ex.,
 693 non-dust) and among optical regimes (dust, non-dust). It showed higher precision in
 694 identifying C1 than C0, the two dust regimes. Among the non-dust regime clusters, the lowest
 695 precision obtained was 0.62 for the prediction of C2; nevertheless, this precision is still a
 696 promising outcome considering the limited number of samples of this cluster available for
 697 the training process. Given its strong absorption nature, mislabeling the C2 aerosol regime
 698 would translate into high radiative error; therefore, as mentioned, extra training is required
 699 to improve the model prediction for C2 occurrence.

700

701 **Table 4.** Performance metrics values of the trained model applied to the test and the
 702 train(within parenthesis) dataset to predict aerosol optical regimes based on aerosol-type
 703 column mass density.

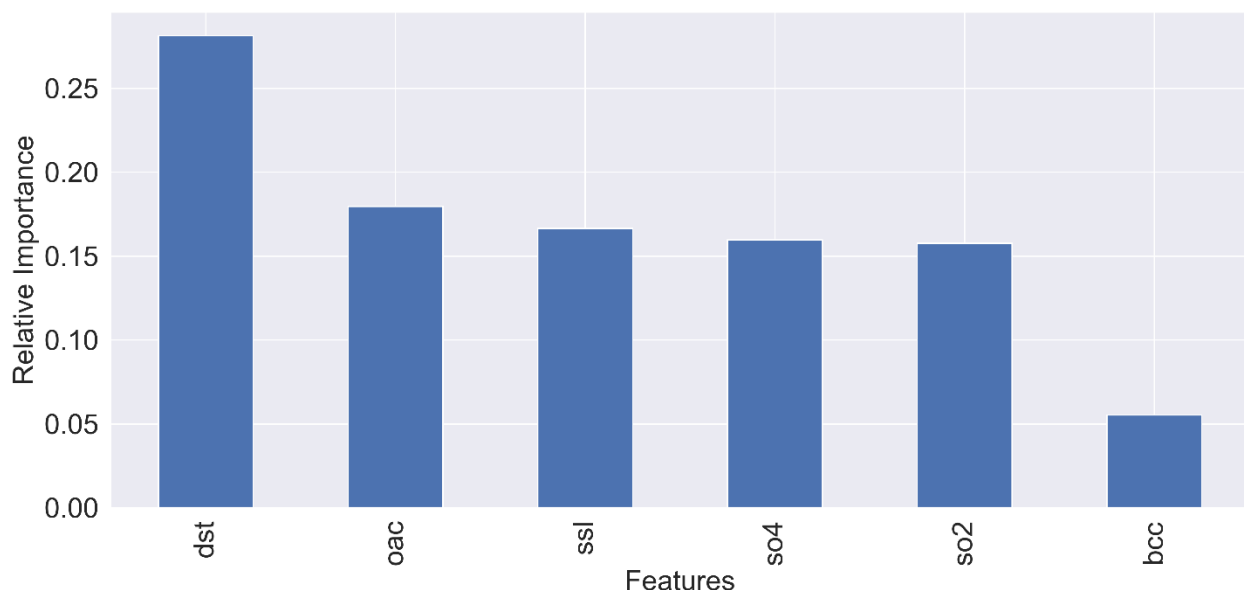
Clusters	Precision	Recall	F1-Score	Support (N)
0	0.62(0.89)	0.62(0.78)	0.62(0.83)	394(848)
1	0.68(0.86)	0.70(0.94)	0.69(0.90)	452(1140)
2	0.62(0.92)	0.60(0.76)	0.61(0.83)	62(111)
3	0.76(0.93)	0.73(0.94)	0.74(0.94)	251(439)
4	0.68(0.91)	0.69(0.93)	0.69(0.92)	185(397)

704

705 **Figure 10** illustrates the relative importance of the predictor variables for the grids
 706 consisting of the AERONET sites, highlighting the influence of each aerosol-type column mass
 707 density on the model's decision-making. The results indicate that the presence of dust over
 708 the Iberian Peninsula is the primary factor affecting the aerosol optical properties in this
 709 region. This finding aligns with actual conditions, as the transport of Saharan dust to the
 710 peninsula is the main driver of aerosol optical properties variability in the area. Dust is
 711 followed by organic carbon, sea salt, and sulfate aerosol types. Organic carbon relevance is
 712 associated with biomass burning, a critical aerosol source during the dry season.
 713 Interestingly, black carbon column mass density did not rank among the top predictors.
 714 Despite the expectation that black carbon might serve as a significant indicator of the aerosol
 715 optical regime due to its association with smoke-influenced aerosols. There is considerable
 716 uncertainty in black carbon simulations in atmospheric chemistry models, including
 717 reanalyses such as MERRA-2, which may hinder its effectiveness in predicting the aerosol
 718 regime observed at AERONET monitoring sites.

719 We also managed to calculate the relative importance of the predictors from Table 1 in the
 720 cluster prediction; the result is presented in the Supplement (Figure SS1). Consistency can
 721 be observed between the score scale from SS1 and that derived from MERRA-2 with respect
 722 to aerosol types (Figure 10). In Figure 10, dust (dst) mass variability emerges as the most
 723 influential factor in determining which cluster should be applied. In the SS1 figure, which
 724 presents the importance of the optical parameters from Table 1 for clustering, the scores
 725 appear well distributed, with a maximum value close to 0.1. Nevertheless, it is evident that
 726 higher wavelengths (near-infrared at 870 and 1020 nm) and specific optical parameters—
 727 namely the Asymmetry Parameter and the Linear Depolarization Ratio—exhibit the greatest
 728 importance, as they are most effective in distinguishing dust from other aerosol types.

729



730

731 **Figure 10:** Relative importance of the predictor variables, i. e. the degree of influence of each
 732 aerosol-type column mass density on the model decision-making. *dst* - Dust, *oac* - Organic
 733 Carbon, *ssl* - Sea-Salt, *so4* - Sulfate *so2* - Sulfur dioxide(precursor of *so4*), *bcc* - Black Carbon.

734

735 3.3 Application: Case studies

736 From the testing dataset, we selected some case studies that significantly impacted local
 737 populations, garnered media attention, and represented different aerosol scenarios in the
 738 Iberian Peninsula. This selection provides a visual (qualitative) demonstration of the model's
 739 predicting capability (**Table 5**).

740

741 **Table 5:** List of case studies of aerosols high loading events over Iberian Peninsula selected
 742 to highlight as examples of the classifier trained model application.

Case study	Date	Nature (Reference link)
------------	------	-------------------------

#01	June 27, 2023	Smoke ¹
#02	October 16, 2017	Dust and Smoke ²
#03	August 11, 2016	Smoke ³
#04	March 17, 2022	Dust ⁴

743 1-<https://earthobservatory.nasa.gov/images/151507/canadian-smoke-reaches-europe>

744 2-<https://atmosphere.copernicus.eu/saharan-dust-and-smoke-over-france-and-uk>

745 3-<https://earthobservatory.nasa.gov/images/88552/fires-rage-in-portugal>

746 4- <https://earthobservatory.nasa.gov/images/149645/dusty-storm-clouds-over-europe>

747

748 We set our trained model to prescribe the spatial distribution of aerosol optical regimes
749 (clusters) that best fit various scenarios based on MERRA-2 aerosol-type column mass
750 density. The results for all cases studied are presented in **Figure 11**. To minimize
751 uncertainties associated with the estimates of aerosol absorptivity, AERONET SSA retrievals
752 are limited to cases where the AOD at 440 nm exceeds 0.4 (Dubovik and King, 2000; Holben
753 et al., 2006). Therefore, the discussion of the optical regime prescriptions was focused on
754 areas where AOD was above this threshold. The uncertainty in retrieved SSA is ~0.03 at AOD
755 at 440 nm = 0.4 and decreases at higher AOD levels (Sinyuk et al., 2020).

756 For our regional analysis, we used the MERRA-2 AOD field as a reference, since AERONET
757 provides local AOD at specific sites. Given that the prescription is done based on a map of the
758 combination of aerosol types column density from models, in this case MERRA-2, the only
759 way to filter areas across the Iberian Peninsula where AOD at 440 nm > 0.4 is to use the AOD
760 field from MERRA-2.

761 Case#01 occurred from June 1 to 25, 2023, coinciding with large-scale wildfire events in
762 Quebec, Canada. A substantial portion of smoke from these wildfires crossed the Atlantic
763 Ocean and reached Western Europe, especially the Iberian Peninsula, resulting in darkened
764 skies in the affected countries. Our trained model predicted that the most suitable aerosol
765 optical regime for the areas impacted by the smoke (Portugal, Western, and Northern Spain)
766 is C4, which corroborates our previous discussion associating the C4 optical regime with
767 regional smoke.

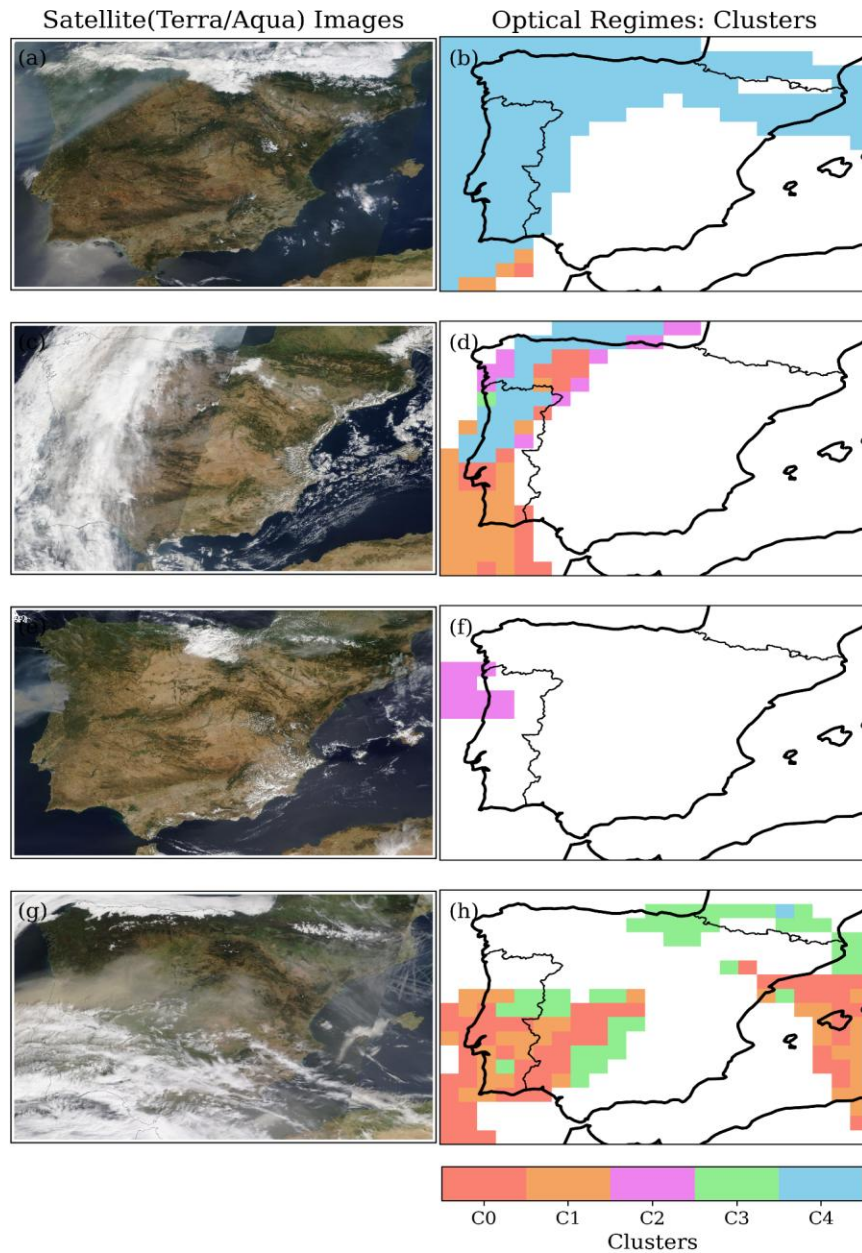
768 Case#02 features an emblematic event on October 16, 2017, marked by a simultaneous
769 massive wildfire in central and northern Portugal and a strong dust transport from North
770 Africa via the south of Portugal. The path connecting the smoke and dust produced a strong
771 northward transport affecting the United Kingdom, influenced by the synoptic conditions
772 associated with the ex-hurricane Ophelia, located just north of the Iberian Peninsula
773 (Osborne et al., 2019). The optical regime prescription identified the C4 cluster as the
774 appropriate regime from central Portugal northward to the UK. Meanwhile, the area affected
775 by dust, spanning from North Africa to southern and central Portugal, was characterized by
776 a mix of C0 and C1, the clusters associated with dust regimes. As the dust plume arrived in

777 Portugal, the model indicated a gradual transition from C1, indicative of pure dust, to C0,
778 which represents conditions of dust mixed with smoke (Gómez-Amo et al., 2017). The
779 random distribution of C2 within the larger C4 regions likely reflects the model's response
780 to the specific conditions dictated by the aerosol-type column mass densities. This could
781 suggest patches of high-absorbing aerosol-type within a less-absorbing large-scale smoke
782 plume, although there is insufficient evidence to draw definitive conclusions.

783 Case#03, dated August 16, 2016, involved strong wildfire emissions in northern Portugal.
784 Most of the smoke was transported toward the Atlantic Ocean, while the remainder of the
785 peninsula experienced low aerosol loading conditions. Consistent with **fresh** smoke aerosol
786 scenarios, the model prescribed the C2 optical regime, **the highest absorbing cluster. In**
787 **strong biomass burning events, especially at the early stage of the emission process, the ratio**
788 **of elemental carbon to organic carbon is usually high, which has been shown to explain the**
789 **high absorption features of fresh smoke plume (Schwink et al, 2024). Additionally, previous**
790 **studies have also shown that Brown Carbon(BrC) absorption is strongest in fresh smoke**
791 **plumes and decreases with atmospheric processing (Saleh, R., et al., 2014; Pokhrel, et al.,**
792 **2017).**

793 Case#04 pertains to an extreme Saharan dust transport that affected most of the Iberian
794 Peninsula on March 15-17, 2022. During this event, the 24-hour average concentration of
795 PM_{2.5} reached as high as 700 $\mu\text{g m}^{-3}$ in parts of Spain (Rodriguez and López-Darias, 2024).
796 The pollution episode was dominated by dust, and indeed, the model prescribed the optical
797 regimes **C0 and C1, which indicate pure dust and dusty conditions** for most of the Iberian
798 Peninsula. This demonstrates our approach's capability to differentiate specific scenarios
799 within dust regimes. For non-dust regimes **such as** C2, a highly absorbing regime, we would
800 not expect to see widespread prescriptions, as we hypothesize that it is associated with fresh,
801 high-absorbing pollution plumes.

802 **Figure 6**, depicting the occurrence of each cluster across the Iberian Peninsula, corroborates
803 our hypothesis by indicating that the C2 regime is mainly present in specific areas where
804 aerosol loading increases are primarily attributed to biomass burning, such as the western
805 lowlands of the Iberian Peninsula (Evora, Badajoz, and Caceres) and in the Galicia region
806 (Coruna). The C3 optical regime was not linked to large-scale dust transport or smoke plumes
807 across the Iberian Peninsula, suggesting it might be associated with high levels of local or
808 regional pollution. **Figure 6** shows that the C3 regime is commonly observed throughout the
809 year in the eastern portion of the Iberian Peninsula. The results of these case studies,
810 combined with performance evaluations, highlight the capability and potential of this
811 machine-learning approach, which uses clustering and random forest classification to
812 prescribe optical models from aerosol-type columnar mass density to calculate aerosol
813 particles' direct radiative effect in atmospheric models. By constraining modelling with
814 observational data, we can help mitigate the known uncertainties related to aerosol direct
815 radiative forcing. Additionally, our method's straightforwardness and lower computational
816 cost, **when compared to the 3D modeling of optical properties based on online Mie**
817 **calculations,** favor operational modeling when infrastructure is limited.



818

819 **Figure 11:** Case studies of distinct aerosol scenarios over the Iberian Peninsula selected to test
 820 our machine-learning based approach to predict the best optical property regime: (a, b)
 821 Case#01 on June 27, 2023; (c, d) Case#02 on October 16, 2017; (e, f) Case#03 on August 11,
 822 2016; (g, h) Case#04 on March 17, 2022. On the left side, MODIS Terra and Aqua/NASA True
 823 color satellite images (<https://wvs.earthdata.nasa.gov/>); and on the right the cluster spatial
 824 distribution prescribed by the model.

825 **Figure 12** shows the single scattering albedo at 550 nm, comparing the current approach
 826 and MERRA-2 reanalysis results. The MERRA-2 columnar total SSA was calculated based on
 827 the ratio of total scattering aerosol optical depth to total extinction aerosol optical depth,
 828 both provided in MERRA-2 aerosol products. For smoke scenarios on June 27, 2023, MERRA-
 829 2 indicated a more absorbing optical regime (SSA at 550 nm ~ 0.86 - 0.90) compared to the

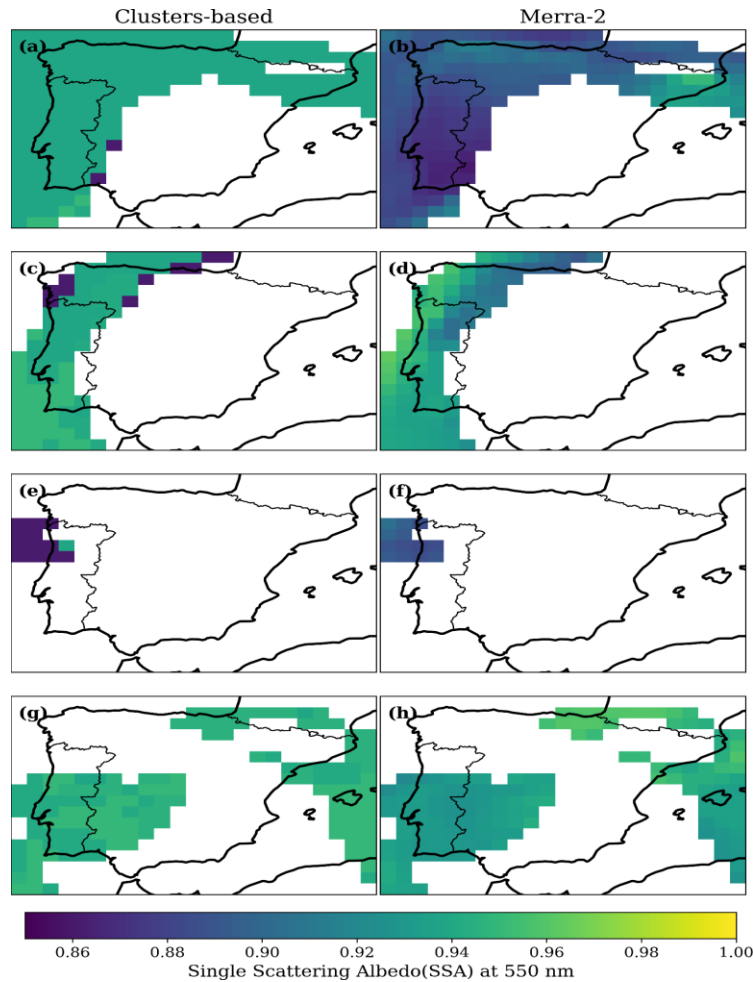
830 current approach (SSA at 550 nm \sim 0.95). On this day, the average SSA at 550 nm over the
831 AERONET site in Coruna City, which was directly affected by Canadian smoke, exceeded 0.95.
832 **The opposite was observed for the strong smoke emission event that occurred over northern**
833 **Portugal on August 11, 2016. Current strategy prescribed a much lower SSA, therefore the**
834 **strongest absorbing regime, when compared with the MERRA-2 calculation. Due to the**
835 **absence of a site in the northern part of Portugal, we were not able to compare the prescribed**
836 **and simulated values with AERONET data.** For the dust scenario, on March 17, 2022, the
837 current approach prescribed a less absorbing optical regime (SSA at 550 nm \sim 0.94 - 0.95)
838 compared to MERRA-2, which reported a SSA at 550 nm of roughly 0.92 - 0.94. The analysis
839 of SSA at 550 nm over AERONET sites affected by the dust event surpassed 0.94. While these
840 cases highlight differences between the prescriptions based on the clusters and MERRA-2
841 results, they are only sufficient to warrant further investigation. To gain a statistical
842 perspective on whether the findings from these case studies are isolated incidents or
843 indicative of a trend, we compare a much larger sample of MERRA-2 SSA at 550 nm across
844 various AERONET sites in the Iberian Peninsula using the clusters approach. We focused only
845 on MERRA-2 aerosol scenarios for AOD at 550 nm larger than 0.3, **which correspond to AOD**
846 **higher than 0.4 at 440 nm previously mentioned, and c**onducted the comparison segmented
847 by the optical regimes defined by the clusters.

848 **Figure 13** shows the count distribution of MERRA-2 SSA at 550 nm for the aerosol regimes
849 represented by the clusters C0, C1, C3, and C4, as classified by the random forest classifier **we**
850 **developed. Histograms of clusters of SSA at 550 nm presented in Figure 13 were generated**
851 **following a Gaussian distribution, considering the cluster average as the central value of each**
852 **optical regime cluster and standard deviation as the typical spread. A similar analysis was**
853 **conducted for the Angstrom Exponent (Figure 14) to evaluate aspects related to particle size**
854 **distribution.** Based on **Figure 13**, we found that, on average, our aerosol optical regime
855 prescription based on the clusters (AERONET) is less absorbing than MERRA-2 for aerosol
856 regimes C0, C1, C3, and C4. More significant differences are observed for C1, C3, and C4.
857 Cluster C1 corresponds to a dust scenario closer to pure dust, while C4 is dominated by
858 smoke. Regarding the particle size indicator (AE), it was observed that MERRA-2 has a lower
859 contribution of coarse particles in the dust regimes compared to the cluster-based
860 prescriptions (**Figure 14a, b**). This finding aligns with Adebisi et al. (2023), who noted that
861 climate models tend to underestimate large dust particles, mainly when representing North
862 African dust plumes. Conversely, for the non-dust regimes (C3, C4), MERRA-2 shows a larger
863 relative contribution of coarse particles than the clusters-based prescription (**Figure 14c, d**).
864 **Figure 15** shows the results for C2. For this specific regime, on average, prescriptions based
865 on the cluster (AERONET) are more absorbing than MERRA-2, opposite to the findings of the
866 other clusters. Regarding AE, under the C2 regime, MERRA-2's mean AE is lower than that
867 prescribed from the cluster, suggesting a lower relative contribution of fine mode in the
868 reanalysis simulations. This is similar to the findings related to the two other fine-mode
869 dominant regimes (C3 and C4).

870 As demonstrated by the SSA and AE distributions (Figures 13, 14, 15) and evaluated from
871 Tables 1 and 4, the model can also predict the occurrence of the minority cluster C2 (3–4
872 percent of samples). **The model preserves the distribution of optical properties of less**
873 **frequent aerosol regimes while capturing MERRA-2 features without the need for explicit**
874 **class imbalance treatment, with C2's highly absorbing and dominant fine mode conditions**

875 reflected in both SSA and AE predictions, with the distributions of values across clusters
 876 showing coherence with MERRA-2 values. With C2's highly absorbing and dominant fine
 877 mode conditions reflected in both SSA and AE predictions, the distributions across clusters
 878 demonstrate agreement between expected and observed distribution values.

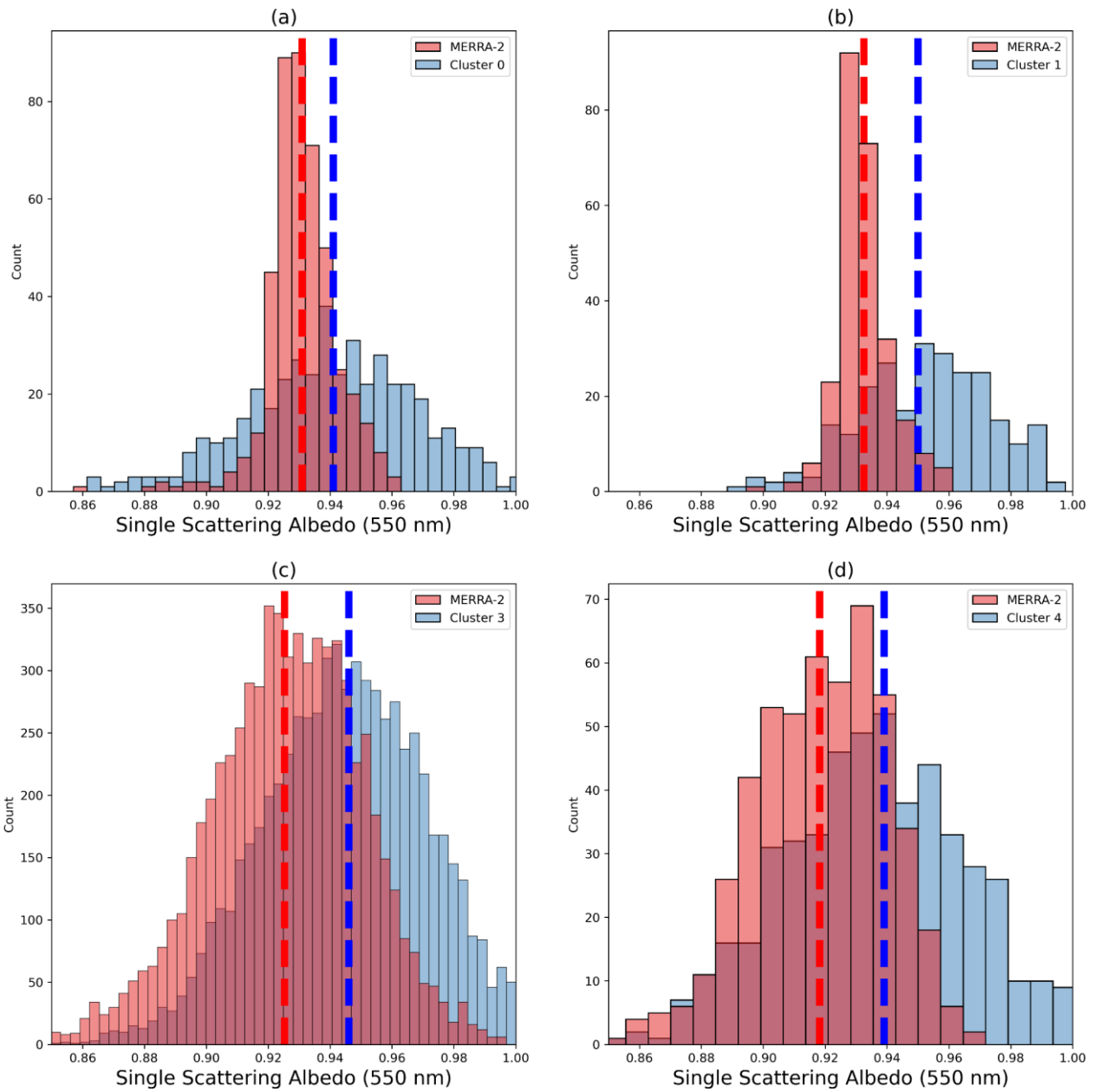
879



880

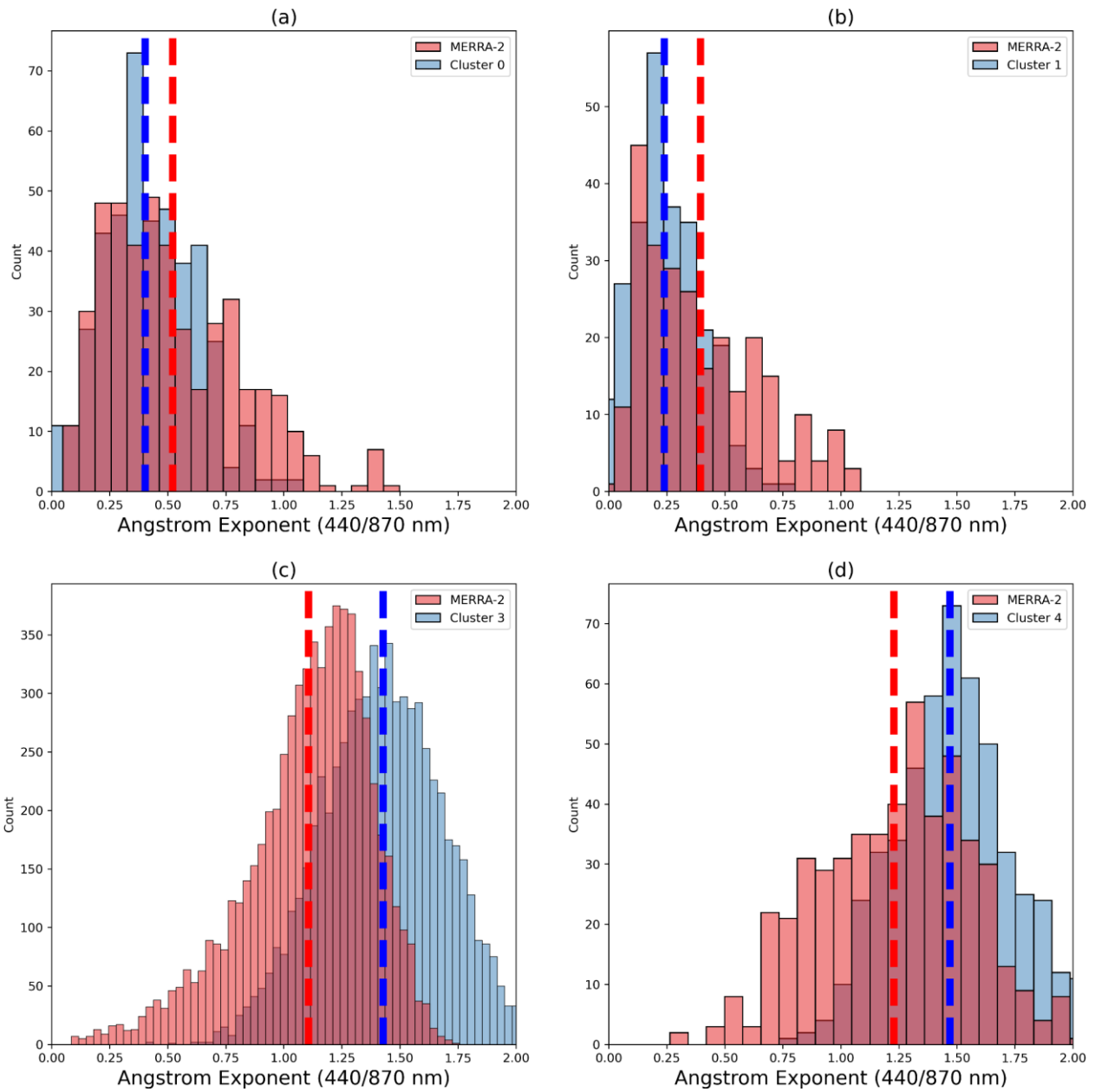
881 **Figure 12:** Single Scattering Albedo (SSA) prescription based on the current study approach
 882 (left) and that simulated by MERRA-2 (right) for the selected case studies of Table 2: (a, b)
 883 Case#01 on June 27, 2023; (c, d) Case#02 on October 16, 2017; (e, f) Case#03 on August 11,
 884 2016; (g, h) Case#04 on March 17, 2022.

885



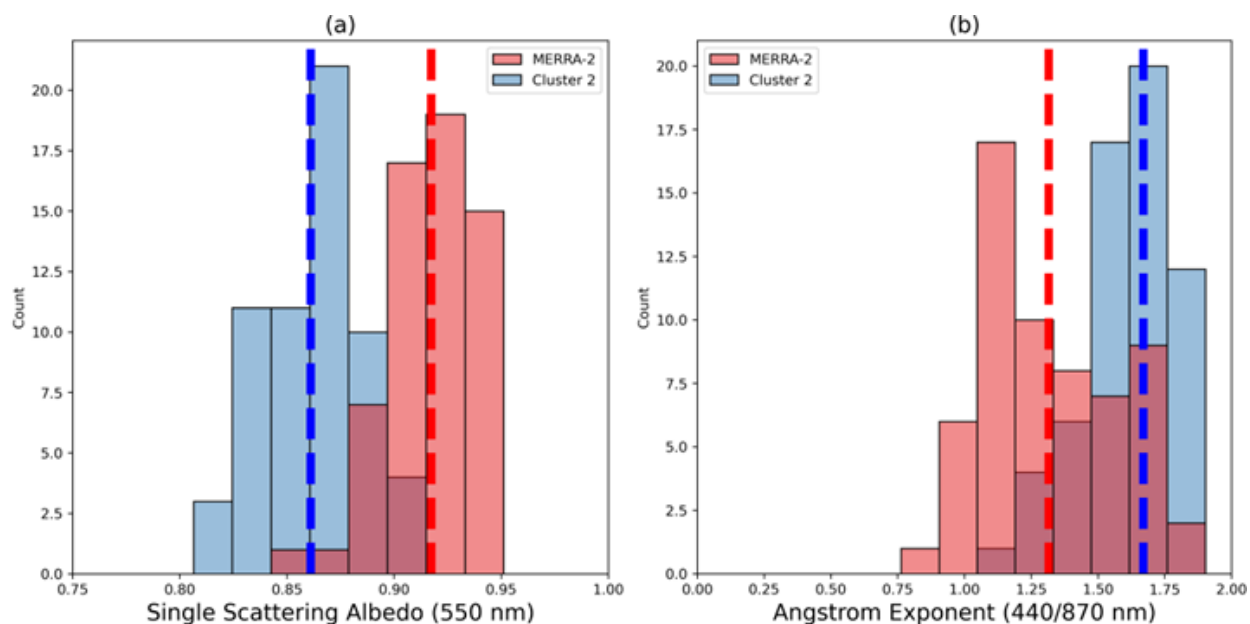
886

887 **Figure 13:** Current study prescription and MERRA-2 simulation of Single Scattering Albedo
 888 (SSA) frequency distribution as function of the optical regime (clusters): a) Cluster 0; b)
 889 Cluster 1; c) Cluster 3; e) Cluster 4. **The dashed lines represent the mean values.**



890
 891 **Figure 14:** Current study prescription and MERRA-2 simulation of Angstrom Exponent (AE)
 892 frequency distribution as function of the optical regime (clusters): a) Cluster 0; b) Cluster 1;
 893 c) Cluster 3; e) Cluster 4. **The dashed lines represent the mean values.**

894



895
 896
 897 **Figure 15:** Current study prescription and MERRA-2 simulation of (a) Single Scattering
 898 Albedo and (b) Angstrom Exponent (AE) frequency distribution for the Cluster 2 scenario.
 899 **The dashed lines represent the mean values.**

900

901 4. Conclusions

902 This study emphasizes the importance of observational-based research to constrain the
 903 prescription of aerosol-intensive properties in atmospheric models. We aimed to
 904 characterize the typical aerosol intensive optical properties affecting the Iberian Peninsula
 905 (IP) using data from the atmospheric column AERONET sky inversion products. We
 906 employed K-means clustering to analyze historical aerosol intensive properties across all
 907 AERONET sites that operated for at least two years and had the highest quality dataset level
 908 (2.0) available. We identified five distinct clusters (C0, C1, C2, C3, and C4) representing
 909 different optical regimes, illustrating the predominant aerosol scenarios in the IP. The key
 910 difference among these clusters lies in the contribution of coarse-mode particles and their
 911 absorption efficiency. Clusters C0 and C1 are dominated by coarse-mode particles and
 912 classified as dust regimes due to their association with Saharan dust transport. In particular,
 913 the optical properties of C1 closely resemble a pure dust scenario, while C0 indicates a more
 914 mixed situation, which we refer to as dusty. On the other hand, clusters C2 and C4 are
 915 identified as non-dust regimes, linked to strong and moderate absorption related to smoke
 916 plumes. Cluster C3, also a non-dust regime, is more frequently observed in the eastern part
 917 of the IP and differs from C4 mainly by having a much lower real part of the refractive index.
 918 After identifying the typical aerosol regimes affecting the IP, we utilized aerosol-type
 919 columnar mass density data (dust, organic carbon, black carbon, sea salt, and sulfates) from

920 MERRA-2 to predict the aerosol optical regime at each grid point using the supervised
921 learning methodology Random Forest. We tested the performance of the trained model
922 under various aerosol scenarios. The accuracy of the predictions for the aerosol optical
923 regimes ranged from 60% to 75%, depending on the regime, with an average accuracy of
924 70%. Notably, the accuracy exceeded 90% when predicting solely dust or non-dust optical
925 regimes.

926 An analysis of MERRA-2 simulations alongside this study's AERONET cluster-based
927 prescriptions of optical regime indicators, such as absorption (SSA) and size (AE), reveals
928 that MERRA-2 is generally more absorbing for the aerosol optical regimes (C0, C1, C3, and
929 C4) impacting the atmosphere of the Iberian Peninsula, except for the most absorbing
930 regime(C2). Specifically, the reanalysis simulations indicate higher absorption under the
931 non-dust regimes C3 and C4. When examining the relative contributions of fine and coarse
932 modes, the cluster-based prescription indicates a larger average contribution of coarse
933 particles than the MERRA-2 under dust regimes (C0, C1). Conversely, for the non-dust
934 regimes (C2, C3, C4), MERRA-2 shows a lower relative contribution from the fine mode
935 compared to the clusters-based prescription.

936 Our findings contribute to enhancing the understanding of the dynamic aerosol optical
937 properties over the Iberian Peninsula and highlight the potential of machine-learning
938 approaches to improve the representation of aerosol radiative forcing in atmospheric
939 models. Many atmospheric modelling systems are not designed to simulate aerosol-intensive
940 microphysical and optical properties in real time. Additionally, computational cost remains
941 a common limitation worldwide. Our approach integrates AERONET-derived intensive
942 properties based on climatological optical regimes to refine the model, coupled with
943 predicted aerosol-type columnar mass density. This integration can help reduce regional
944 uncertainty in the simulation of aerosol radiative forcing.

945 **Nevertheless, we acknowledge that additional research and analysis are necessary to build**
946 **on the developments and findings presented here. Among the potential limitations and**
947 **directions for future work, we emphasize the importance of better understanding the impact**
948 **of AERONET parameter uncertainties on the clustering process, as well as conducting an**
949 **intercomparison between basic and more advanced clustering approaches. A natural**
950 **extension of this study would be the development of a comprehensive investigation focused**
951 **on radiative transfer calculations, within which the proposed method could be thoroughly**
952 **evaluated.**

953

954

955

956

957

958

959 **Competing interests**

960 The authors declare that they have no conflict of interest.

961 **Acknowledgements and financial support**

962 The authors acknowledge the financial support of FCT—Science and Technology Portuguese
963 Foundation, which funded the project FIRESMOKE ([http://doi.org/10.54499/PTDC/CTA-
964 MET/3392/2020](http://doi.org/10.54499/PTDC/CTA-
964 MET/3392/2020)) through national funds. Thanks are also owed to the financial support
965 given to CESAM by FCT (UID Centro de Estudos do Ambiente e Mar (CESAM) +
966 LA/P/0094/2020) through national funds. We also acknowledge the financial support of
967 CNPq - National Council for Scientific and Technological Development (CNPq) through the
968 funding processes CNPq N^o 441851/2023-1 and CNPq N^o 172486/2023-8. Author HFCV also
969 thanks to the CNPq grant No 315349/2023-9. We thank AERONET and MERRA-2 PIs and
970 teams for their effort in establishing and maintaining the sites and the reanalysis
971 development used in this study. We acknowledge the use of imagery from the Worldview
972 Snapshots application (<https://wvs.earthdata.nasa.gov>), part of the Earth Observing System
973 Data and Information System (EOSDIS).

974

975 **Author contributions**

976 NR, KL and PT designed and performed the research, analyzed the data, and wrote the first
977 version of the paper. MY, SF, LF, OM, HFCV contributed to writing, discussion, review and
978 editing. ICM and AIM conceptualization and coordination of the Project FIRESMOKE,
979 discussion, review and editing.

980 **Code and data availability.**

981 All the datasets (AERONET and MERRA-2) used in this study are publicly available and were
982 downloaded from their respective websites (<https://aeronet.gsfc.nasa.gov/>; and
983 <https://disc.gsfc.nasa.gov/datasets?project=MERRA-2>). Code and dataset required to
984 conduct the analyses herein is available at <https://doi.org/10.5281/zenodo.15178347>
985 (Rosario, 2025).

986 **References**

987 Abraham, A, F Pedregosa, M Eickenberg, P Gervais, A Mueller, J Kossaifi, A Gramfort, B
988 Thirion, and G Varoquaux. 2014. “Machine Learning for Neuroimaging with Scikit-Learn.”
989 *Front Neuroinform* 8: 14.

- 990 Adebisi, A.A., Huang, Y., Samset, B.H. et al. Observations suggest that North African dust
991 absorbs less solar radiation than models estimate. *Commun Earth Environ* 4, 168 (2023).
992 <https://doi.org/10.1038/s43247-023-00825-2>.
- 993 Alvarez, Albert, Judit Lecina-Diaz, Enric Batllori, Andrea Duane, Lluís Brotons, Javier Retana,
994 Spatiotemporal patterns and drivers of extreme fire severity in Spain for the period 1985–
995 2018, *Agricultural and Forest Meteorology*, Volume 358, 2024, 110185, ISSN 0168-1923,
996 <https://doi.org/10.1016/j.agrformet.2024.110185>
- 997 Asfaw, H. W., McGee, T. K., & Correia, F. J. (2022). Wildfire preparedness and response during
998 the 2016 Arouca wildfires in rural Portugal. *International Journal of Disaster Risk Reduction*,
999 73, 102-895. <https://doi.org/10.1016/j.ijdr.2022.102895>
- 1000 Breiman, Leo. 2001. "Random Forests". *Machine Learning* 45 (1): 5–32.
1001 <https://doi.org/10.1023/a:1010933404324>.
- 1002 Brown H, Liu X, Pokhrel R, Murphy S, Lu Z, Saleh R, Mielonen T, Kokkola H, Bergman T, Myhre
1003 G, Skeie RB, Watson-Paris D, Stier P, Johnson B, Bellouin N, Schulz M, Vakkari V, Beukes JP,
1004 van Zyl PG, Liu S, Chand D. Biomass burning aerosols in most climate models are too
1005 absorbing. *Nat Commun*. 2021 Jan 12;12(1):277. doi: 10.1038/s41467-020-20482-9. PMID:
1006 33436592; PMCID: PMC7804930.
- 1007 Buchard-Marchant, V.J., C.A. Randles, A.M. da Silva, A. Darmenov, P.R. Colarco, R. Govindaraju,
1008 R.A. Ferrare, J. Hair, A. Beyersdorf, L.D. Ziemba, and H. Yu (2017), The MERRA-2 Aerosol
1009 Reanalysis, 1980 Onward. Part II: Evaluation and Case Studies, *J. Climate*, 30, 6851-6872,
1010 doi:10.1175/JCLI-D-16-0613.1.
- 1011 Cachorro, V. E., Burgos, M. A., Mateos, D., Toledano, C., Bennouna, Y., Torres, B., de Frutos, Á.
1012 M., and Herguedas, Á.: Inventory of African desert dust events in the north-central Iberian
1013 Peninsula in 2003–2014 based on sun-photometer–AERONET and particulate-mass–EMEP
1014 data, *Atmos. Chem. Phys.*, 16, 8227–8248, <https://doi.org/10.5194/acp-16-8227-2016>,
1015 2016.
- 1016 Chen, G., Wang, J., Wang, Y., Wang, J., Jin, Y., Cheng, Y., et al. (2023). An aerosol optical module
1017 with observation-constrained black carbon properties for global climate models. *Journal of*
1018 *Advances in Modeling Earth Systems*, 15, e2022MS003501.
1019 <https://doi.org/10.1029/2022MS003501>
- 1020 Chin, M., Ginoux, P., Kinne, S., Torres, O., Holben, B. N., Duncan, B. N., Martin, R. V., Logan, J. A.,
1021 Higurashi, A., and Nakajima, T.: Tropospheric aerosol optical thickness from the GOCART
1022 model and comparisons with satellite and sun photometer measurements, *J. Atmos. Sci.*, 59,
1023 461–483, [https://doi.org/10.1175/1520-0469\(2002\)059<0461:taotft>2.0.co;2](https://doi.org/10.1175/1520-0469(2002)059<0461:taotft>2.0.co;2), 2002.
- 1024 Colarco, P., Da Silva, A., Chin, M., and Diehl, T.: Online simulations of global aerosol
1025 distributions in the NASA GEOS-4 model and comparisons to satellite and ground-based
1026 aerosol optical depth, *J. Geophys. Res.-Atmos.*, 115, D14207,620
1027 <https://doi.org/10.1029/2009JD012820>, 2010.

- 1028 Colarco, P. R., Nowottnick, E. P., Randles, C. A., Yi, B., Yang, P., Kim, K.-M., Smith, J. A., and
1029 Bardeen, C. G.: Impact of radiatively interactive dust aerosols in the NASA GEOS-5 climate
1030 model: Sensitivity to dust particle shape and refractive index, *J. Geophys. Res.-Atmos.*, **119**,
1031 753–786, <https://doi.org/10.1002/2013JD020046>, 2014
- 1032 Dubovik, O., B. Holben, T. F. Eck, A. Smirnov, Y. J. Kaufman, M. D. King, D. Tanré, and I. Slutsker,
1033 2002: Variability of Absorption and Optical Properties of Key Aerosol Types Observed in
1034 Worldwide Locations. *J. Atmos. Sci.*, **59**, 590–608, [https://doi.org/10.1175/1520-0469\(2002\)059<0590:VOAAOP>2.0.CO;2](https://doi.org/10.1175/1520-0469(2002)059<0590:VOAAOP>2.0.CO;2).
- 1036 Eck, T. F., Holben, B. N., Reid, J. S., Dubovik, O., Smirnov, A., O'Neill, N. T., Slutsker, I., and Kinne,
1037 S.: Wavelength dependence of the optical depth of biomass burning, urban, and desert dust
1038 aerosols, *J. Geophys. Res.*, **104**, 31333–31349, doi:10.1029/1999jd900923, 1999.
- 1039 Elias, Thierry Ghislain, Ana Maria Silva, Maria João Figueira, Nuno Belo, Sergio Pereira, Paola
1040 Formenti, Gunter Helas, "Aerosol extinction and absorption in Évora, Portugal, during the
1041 European 2003 summer heat wave," *Proc. SPIE 5571, Remote Sensing of Clouds and the*
1042 *Atmosphere IX*, (30 November 2004); <https://doi.org/10.1117/12.566579>
- 1043 Ermitão, T.; Páscoa, P.; Trigo, I.; Alonso, C.; Gouveia, C. Mapping the Most Susceptible Regions
1044 to Fire in Portugal. *Fire* **2023**, *6*, 254. <https://doi.org/10.3390/fire6070254>
- 1045 Fan, Y. , X. Sun, H. Huang, R. Ti, X. Liu The primary aerosol models and distribution
1046 characteristics over China based on the AERONET data *J. Quant. Spectrosc. Ra.*, **275** (2021),
1047 10.1016/j.jqsrt.2021.107888
- 1048 Gelaro, R., and Coauthors, 2017: The Modern-Era Retrospective Analysis for Research and
1049 Applications, Version 2 (MERRA-2). *J. Climate*, **30**, 5419–5454,
1050 <https://doi.org/10.1175/JCLI-D-16-0758.1>.
- 1051 Gómez-Amo, J. L., Estellés, V., Marcos, C., Segura, S., Esteve, A. R., Pedrós, R., Utrillas, M. P., and
1052 Martínez-Lozano, J. A.: Impact of dust and smoke mixing on column-integrated aerosol
1053 properties from observations during a severe wildfire episode over Valencia (Spain), *Science*
1054 *Total Environ.*, **599–600**, 2121–2134, <https://doi.org/10.1016/j.scitotenv.2017.05.041>,
1055 2017.
- 1056 Groß, S., Tesche, M., Freudenthaler, V., Toledano, C., Wiegner, M., Ansmann, A., Althausen, D.
1057 and Seefeldner, M. (2011) 'Characterization of Saharan dust, marine aerosols and mixtures
1058 of biomass-burning aerosols and dust by means of multi-wavelength depolarization and
1059 Raman lidar measurements during SAMUM 2', *Tellus B: Chemical and Physical Meteorology*,
1060 **63(4)**, p. 706-724. Available at: <https://doi.org/10.1111/j.1600-0889.2011.00556.x>.
- 1061 Hamed, R.A.; Alawode, G.L.; Montoya, L.E.; Krasovskiy, A.; Kraxner, F. Exploring Drivers of
1062 Wildfires in Spain. *Land* **2024**, *13*, 762. <https://doi.org/10.3390/land13060762>
- 1063 Henok Workeye Asfaw, Tara K. McGee, Fernando Jorge Correia, Wildfire preparedness and
1064 response during the 2016 Arouca wildfires in rural Portugal, *International Journal of Disaster*

- 1065 Risk Reduction, Volume 73, 2022, 102895, ISSN 2212-4209,
1066 <https://doi.org/10.1016/j.ijdr.2022.102895>.
- 1067 Hess, M., P. Koepke, and I. Schult, 1998: Optical properties of aerosols and clouds: The
1068 software package OPAC. *Bull. Amer. Meteor. Soc.*, 79, 831–844.
- 1069 Hoelzemann, J. J., Longo, K. M., Fonseca, R. M., do Rosario, N. M. E., Elbern, H., Freitas, S. R., and
1070 Pires, C.: Regional representativity of AERONET observation sites during the biomass
1071 burning season in South America determined by correlation studies with MODIS Aerosol
1072 Optical Depth, *J. Geophys. Res.*, 114, D13301, doi:10.1029/2008jd010369, 2009
- 1073 Holben, B. N., Eck, T. F., Slutsker, I., Tanre, D., Buis, J. P., Setzer, A., Vermote, E., Reagan, J. A.,
1074 Kaufman, Y. J., Nakajima, T., Lavenu, F., Jankowiak, I., and Smirnov, A.: AERONET – A Federated
1075 Instrument Network and Data Archive for Aerosol Characterization, *Remote Sens. Environ.*,
1076 66, 1–16, doi:10.1016/s0034-4257(98)00031-5, 1998.
- 1077 Illingworth, A. J., and Coauthors, 2015: The EarthCARE Satellite: The Next Step Forward in
1078 Global Measurements of Clouds, Aerosols, Precipitation, and Radiation. *Bull. Amer. Meteor.*
1079 *Soc.*, 96, 1311–1332, <https://doi.org/10.1175/BAMS-D-12-00227.1>.
- 1080 IPCC, 2021: Climate Change 2021 - the Physical Science Basis, Contribution of Working Group
1081 I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change [Masson-
1082 Delmotte, V., P. Zhai, A. Pirani, S.L. Connors, C. Péan, S. Berger, N. Caud, Y. Chen, L. Goldfarb,
1083 M.I. Gomis, M. Huang, K. Leitzell, E. Lonnoy, J.B.R. Matthews, T.K. Maycock, T. Waterfield, O.
1084 Yelekçi, R. Yu, and B. Zhou (eds.)]. Cambridge University Press, In Press, Published: 9 August
1085 2021.
- 1086 Kanitz, T., A. Ansmann, R. Engelmann, and D. Althausen, 2013: North-south cross sections of
1087 the vertical aerosol distribution over the Atlantic Ocean from multiwavelength
1088 Raman/polarization lidar during Polarstern cruises. *J. Geophys. Res. Atmos.*, 118, 2643–
1089 2655, doi:10.1002/jgrd.50273.
- 1090 Kim, D. and Ramanathan, V. (2008) Solar Radiation Budget and Radiative Forcing Due to
1091 Aerosols and Clouds. *Journal of Geophysical Research: Atmospheres*, 113, D02203.
1092 <https://doi.org/10.1029/2007JD008434>
- 1093 Koepke, P., M. Hess, I. Schult, and E. P. Shettle (1997), Global aerosol data set, *Rep. 243*, Max-
1094 Planck-Inst. für Meteorol., Hamburg, Germany.
- 1095 Levy, R. C., Remer, L. A., Kleidman, R. G., Mattoo, S., Ichoku, C., Kahn, R., and Eck, T. F.: Global
1096 evaluation of the Collection 5 MODIS dark-target aerosol products over land, *Atmos. Chem.*
1097 *Phys.*, 10, 10399–10420, doi:10.5194/acp-10-10399-2010, 2010
- 1098 Levy, R. C., Remer, L. A., and Dubovik, O.: Global aerosol optical properties and application to
1099 Moderate Resolution Imaging Spectroradiometer aerosol retrieval over land, *J. Geophys.*
1100 *Res.-Atmos.*, 112, D13210, <https://doi.org/10.1029/2006JD007815>, 2007.

1101 Li, J., Carlson, B.E., Yung, Y.L. et al. Scattering and absorbing aerosols in the climate system.
1102 Nat Rev Earth Environ 3, 363–379 (2022). <https://doi.org/10.1038/s43017-022-00296-7>

1103 Li, J., L. Liu, A. A. Lacis, and B. E. Carlson (2010), An optimal fitting approach to improve the
1104 GISS ModelE aerosol optical property parameterization using AERONET data, J. Geophys.
1105 Res., 115, D16211, doi:10.1029/2010JD013909.

1106 Li, Z.; Zhang, Y.; Xu, H.; Li, K.; Dubovik, O.; Goloub, P. The Fundamental Aerosol Models Over
1107 China Region: A Cluster Analysis of the Ground-Based Remote Sensing Measurements of
1108 Total Columnar Atmosphere. Geophys. Res. Lett. 2019, 46, 4924–4932

1109 Martins, J. V., Artaxo, P., Kaufman, Y. J., Castanho, A. D., and Remer, L. A.: Spectral absorption
1110 properties of aerosol particles from 350–2500 nm, Geophys. Res. Lett., 36, L13810,
1111 <https://doi.org/10.1029/2009GL037435>, 2009.

1112 Moise, T., Flores, J. M., and Rudich, Y.: Optical properties of secondary organic aerosols and
1113 their changes by chemical processes, Chem. Rev., 115, 4400–4439, 2015.

1114 Osborne, M., Malavelle, F. F., Adam, M., Buxmann, J., Sugier, J., Marenco, F., and Haywood, J.:
1115 Saharan dust and biomass burning aerosols during ex-hurricane Ophelia: observations from
1116 the new UK lidar and sun-photometer network, Atmos. Chem. Phys., 19, 3557–3578,
1117 <https://doi.org/10.5194/acp-19-3557-2019>, 2019.

1118 Proske, U., Ferrachat, S., and Lohmann, U.: Developing a climatological simplification of
1119 aerosols to enter the cloud microphysics of a global climate model, Atmos. Chem. Phys., 24,
1120 5907–5933, <https://doi.org/10.5194/acp-24-5907-2024>, 2024.

1121 Ramanathan, V., P. J. Crutzen, J. T. Kiehl, and D. Rosenfeld. 2001. “Aerosols, Climate, and the
1122 Hydrological Cycle”. *Science* 294 (5549). <https://doi.org/10.1126/science.1064034>.

1123 Reid, J. S. and Hobbs, P. V.: Physical and optical properties of smoke from individual biomass
1124 fires in Brazil, J. Geophys. Res., 103, 32 013–32 031, 1998

1125 Reid, J. S., Eck, T. F., Christopher, S. A., Koppmann, R., Dubovik, O., Eleuterio, D. P., Holben, B.
1126 N., Reid, E. A., and Zhang, J.: A review of biomass burning emissions part III: intensive optical
1127 properties of biomass burning particles, Atmos. Chem. Phys., 5, 827–849,
1128 <https://doi.org/10.5194/acp-5-827-2005>, 2005.

1129 Rodríguez, S. and López-Darias, J.: Extreme Saharan dust events expand northward over the
1130 Atlantic and Europe, prompting record-breaking PM₁₀ and PM_{2.5} episodes, Atmos. Chem.
1131 Phys., 24, 12031–12053, <https://doi.org/10.5194/acp-24-12031-2024>, 2024.

1132 Rosario, N. E.: Machine learning-driven characterization and prescription of aerosol optical
1133 properties for atmospheric models, Zenodo [code],
1134 <https://doi.org/10.5281/zenodo.14825197>, 2025.

- 1135 Rosário, N. E., Longo, K. M., Freitas, S. R., Yamasoe, M. A., and Fonseca, R. M.: Modeling the
1136 South American regional smoke plume: aerosol optical depth variability and surface
1137 shortwave flux perturbation, *Atmos. Chem. Phys.*, 13, 2923–2938,
1138 <https://doi.org/10.5194/acp-13-2923-2013>, 2013.
- 1139 Russell, P. B., Kacenelenbogen, M., Livingston, J. M., Hasekamp, O. P., Burton, S. P., Schuster, G.
1140 L., Johnson, M. S., Knobelspiesse, K. D., Redemann, J., Ramachandran, S., and Holben, B.: A
1141 multiparameter aerosol classification method and its application to retrievals from
1142 spaceborne polarimetry, *J. Geophys. Res.-Atmos.*, 119, 9838–9863,
1143 <https://doi.org/10.1002/2013JD021411>, 2014
- 1144 Saleh, R., Robinson, E. S., Tkacik, D. S., Ahern, A. T., Liu, S., Aiken, A. C., Sullivan, R. C., Presto,
1145 A. A., Dubey, M. K., Yokelson, R. J., Donahue, N. M., & Robinson, A. L. (2014). Brownness of
1146 organics in aerosols from biomass burning linked to their black carbon content. *Nature*
1147 *Geoscience*, 7(9), 647-650. <https://doi.org/10.1038/ngeo2220>
- 1148 Samset, B.H., Stjern, C.W., Andrews, E. et al. Aerosol Absorption: Progress Towards Global and
1149 Regional Constraints. *Curr Clim Change Rep* 4, 65–83 (2018).
1150 <https://doi.org/10.1007/s40641-018-0091-4>
- 1151 Sand, M., Samset, B. H., Myhre, G., Gliß, J., Bauer, S. E., Bian, H., Chin, M., Checa-Garcia, R.,
1152 Ginoux, P., Kipling, Z., Kirkevåg, A., Kokkola, H., Le Sager, P., Lund, M. T., Matsui, H., van Noije,
1153 T., Oliví, D. J. L., Remy, S., Schulz, M., Stier, P., Stjern, C. W., Takemura, T., Tsigaridis, K., Tsyro,
1154 S. G., and Watson-Parris, D.: Aerosol absorption in global models from AeroCom phase III,
1155 *Atmos. Chem. Phys.*, 21, 15929–15947, <https://doi.org/10.5194/acp-21-15929-2021>, 2021.
- 1156 Schwink SK, Mael LE, Dunnington TH, Schmid MJ, Silberstein JM, Heck A, Gotlib N, Hannigan
1157 MP, Vance ME. Impacts of Aging and Relative Humidity on Properties of Biomass Burning
1158 Smoke Particles. *ACS EST Air*. 2024 Dec 6;2(1):109-118. doi: 10.1021/acsestair.4c00224.
1159 PMID: 39817254; PMCID: PMC11730893.
- 1160 Shettle, E. P. and Fenn, R. W.: Models for the Aerosols of the Lower Atmosphere and the
1161 Effects of Humidity Variations on Their Optical Properties, AFGL-TR-79-0214, 94, 1979
- 1162 Shi, C., Wei, B., Wei, S. et al. A quantitative discriminant method of elbow point for the optimal
1163 number of clusters in clustering algorithm. *J Wireless Com Network* 2021, 31 (2021).
1164 <https://doi.org/10.1186/s13638-021-01910-w>
- 1165 Shin, S.-K., Tesche, M., Kim, K., Kezoudi, M., Tatarov, B., Müller, D., and Noh, Y.: On the spectral
1166 depolarisation and lidar ratio of mineral dust provided in the AERONET version 3 inversion
1167 product, *Atmos. Chem. Phys.*, 18, 12735–12746, [https://doi.org/10.5194/acp-18-12735-](https://doi.org/10.5194/acp-18-12735-2018)
1168 2018, 2018.
- 1169 Silva, P.; Carmo, M.; Rio, J.; Novo, I. Changes in the Seasonality of Fire Activity and Fire
1170 Weather in Portugal: Is the Wildfire Season Really Longer? *Meteorology* 2023, 2, 74-86.
1171 <https://doi.org/10.3390/meteorology2010006>

- 1172 Sinyuk, A., Holben, B. N., Eck, T. F., Giles, D. M., Slutsker, I., Korkin, S., Schafer, J. S., Smirnov, A.,
1173 Sorokin, M., and Lyapustin, A.: The AERONET Version 3 aerosol retrieval algorithm,
1174 associated uncertainties and comparisons to Version 2, *Atmos. Meas. Tech.*, 13, 3375–3411,
1175 <https://doi.org/10.5194/amt-13-3375-2020>, 2020.
- 1176 Smirnov, A., B. N. Holben, Y. J. Kaufman, O. Dubovik, T. F. Eck, I. Slutsker, C. Pietras, and R. N.
1177 Halthore, 2002: Optical Properties of Atmospheric Aerosol in Maritime Environments. *J.*
1178 *Atmos. Sci.*, 59, 501–523, [https://doi.org/10.1175/1520-
1179 0469\(2002\)059<0501:OPOAAI>2.0.CO;2](https://doi.org/10.1175/1520-0469(2002)059<0501:OPOAAI>2.0.CO;2).
- 1180 Spencer, RS, RC Levy, LA Remer, S Mattoo, GT Arnold, DL Hlavka, KG Meyer, A Marshak, EM
1181 Wilcox, and SE Platnick. 2019. “Exploring Aerosols near Clouds with High-Spatial-Resolution
1182 Aircraft Remote Sensing during SEAC(4)RS.” *J Geophys Res Atmos* 124: 2148–73.
- 1183 Toledano, C., Cachorro, V. E., de Frutos, A. M., Sorribas, M., and Prats, N.: Inventory of African
1184 Desert Dust Events Over the Southwestern Iberian Peninsula in 2000–2005 with an
1185 AERONET Cimel Sun Photometer, *J. Geophys. Res.*, 112, D21201,
1186 doi:10.1029/2006JD008307, 2007
- 1187 Zhao, G., Tan, T., Zhao, W., Guo, S., Tian, P., and Zhao, C.: A new parameterization scheme for
1188 the real part of the ambient urban aerosol refractive index, *Atmos. Chem. Phys.*, 19, 12875–
1189 12885, <https://doi.org/10.5194/acp-19-12875-2019>,
1190 <https://acp.copernicus.org/articles/19/12875/2019/> 2019.
- 1191 Zhong Q, Schutgens N, van der Werf GR, van Noije T, Bauer SE, Tsigaridis K, Mielonen T,
1192 Checa-Garcia R, Neubauer D, Kipling Z, Kirkevåg A, Olivie DJL, Kokkola H, Matsui H, Ginoux P,
1193 Takemura T, Le Sager P, Rémy S, Bian H, Chin M. Using modelled relationships and satellite
1194 observations to attribute modelled aerosol biases over biomass burning regions. *Nat*
1195 *Commun.* 2022 Oct 7;13(1):5914. doi: 10.1038/s41467-022-33680-4. PMID: 36207322;
1196 PMCID: PMC9547058.
- 1197 Zhou, P.; Wang, Y.; Liu, J.; Xu, L.; Chen, X.; Zhang, L. Difference between global and regional
1198 aerosol model classifications and associated implications for spaceborne aerosol optical
1199 depth retrieval. *Atmos. Environ.* **2023**, *300*, 119674.