Supplement of Geosci. Model Dev., 19, 2691–2716, 2026
https://doi.org/10.5194/gmd-19-2691-2026-supplement

Geoscientific Model Development
Open Access

Supplement of

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

Nilton Évora do Rosário et al.

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

Supplement

Table S1 - Mean, median and standard deviation of optical and microphysical properties of each cluster.

Cluster		0	1	2	3	4
vmr_f	**mean**	0,1371	0,1434	0,1558	0,2045	0,1802
vmr_f	**median**	0,1320	0,1380	0,1520	0,2035	0,1720
vmr_f	**std**	0,0247	0,0315	0,0208	0,0419	0,0369
vmr_c	**mean**	2,0184	1,8822	3,1005	2,8167	2,8173
vmr_c	**median**	1,9950	1,8770	3,0310	2,7995	2,7740
vmr_c	**std**	0,2273	0,1700	0,4527	0,4189	0,5696
reff_f	**mean**	0,1161	0,1166	0,1431	0,1826	0,1653
reff_f	**median**	0,1130	0,1120	0,1400	0,1810	0,1560
reff_f	**std**	0,0202	0,0244	0,0171	0,0345	0,0332
reff_c	**mean**	1.6814	1,6115	2,4348	2,3110	2,2478
reff_c	**median**	1,6665	1,6130	2,3620	2,2790	2,1800
reff_c	**std**	0,1594	0,1319	0,4340	0,3781	0,4933
std_f	**mean**	0,6056	0,6737	0,4189	0,4687	0,4133
std_f	**median**	0,6170	0,6810	0,4100	0,4640	0,4090
std_f	**std**	0,0893	0,0737	0,0612	0,0641	0,0506
std_c	**mean**	0,5960	0,5402	0,6772	0,6301	0,6666
std_c	**median**	0,5960	0,5390	0,6870	0,6310	0,6740
std_c	**std**	0,0516	0,0430	0,0617	0,0496	0,0496
ssa440	**mean**	0,9095	0,9105	0,8594	0,9481	0,9365
ssa440	**median**	0,9104	0,9102	0,8599	0,9523	0,9406
ssa440	**std**	0,0267	0,0210	0,0316	0,0286	0,0311
ssa670	**mean**	0,9722	0,9793	0,8418	0,9457	0,9386
ssa670	**median**	0,9792	0,9831	0,8481	0,9481	0,9392
ssa670	**std**	0,0186	0,0117	0,0376	0,0290	0,0291
ssa870	**mean**	0,9733	0,9805	0,8072	0,9350	0,9281
ssa870	**median**	0,9806	0,9857	0,8139	0,9389	0,9284
ssa870	**std**	0,0203	0,0132	0,0491	0,0363	0,0344
ssa1020	**mean**	0,9752	0,9829	0,7873	0,9301	0,9225
ssa1020	**median**	0,9801	0,9900	0,8003	0,9360	0,9221
ssa1020	**std**	0,0197	0,0128	0,0568	0,0409	0,0374
asy440	**mean**	0,7550	0,7874	0,6732	0,7343	0,6896
asy440	**median**	0,7564	0,7877	0,6695	0,7362	0,6899
asy440	**std**	0,0249	0,0189	0,0272	0,0253	0,0189
asy670	**mean**	0,7190	0,7516	0,5845	0,6761	0,6299
asy670	**median**	0,7233	0,7517	0,5786	0,6851	0,6349
asy670	**std**	0,0210	0,0132	0,0443	0,0462	0,0358
asy870	**mean**	0,7215	0,7519	0,5473	0,6440	0,6006
asy870	**median**	0,7260	0,7524	0,5368	0,6505	0,6024
asy870	**std**	0,0196	0,0124	0,0540	0,0523	0,0448
asy1020	**mean**	0,7263	0,7554	0,5387	0,6320	0,5912
asy1020	**median**	0,7300	0,7554	0,5270	0,6359	0,5878
asy1020	**std**	0,0173	0,0123	0,0586	0,0546	0,0552
rfi_real440	**mean**	1,4715	1,4436	1,5087	1,4259	1,5244
rfi_real440	**median**	1,4712	1,4458	1,5091	1,4322	1,5318
rfi_real440	**std**	0,0376	0,0274	0,0685	0,0546	0,0521
rfi_real670	**mean**	1,4919	1,4581	1,5157	1,4194	1,5174

rfi_real670	**median**	1,4873	1,4589	1,5192	1,4218	1,5234
rfi_real670	**std**	0,0278	0,0184	0,0538	0,0316	0,0366
rfi_real870	**mean**	1,4877	1,4478	1,5180	1,4209	1,5144
rfi_real870	**median**	1,4837	1,4483	1,5223	1,4213	1,5173
rfi_real870	**std**	0,0239	0,0160	0,0500	0,0278	0,0341
rfi_real102	**mean**	1,4804	1,4363	1,5144	1,4184	1,5070
rfi_real102	**median**	1,4761	1,4367	1,5205	1,4181	1,5075
rfi_real102	**std**	0,0247	0,0169	0,0477	0,0277	0,0343
rfi_imag44	**mean**	0,0049	0,0043	0,0253	0,0060	0,0085
rfi_imag44	**median**	0,0045	0,0041	0,0239	0,0054	0,0082
rfi_imag44	**std**	0,0021	0,0013	0,0086	0,0037	0,0043
rfi_imag670	**mean**	0,0015	0,0011	0,0215	0,0053	0,0066
rfi_imag670	**median**	0,0010	0,0008	0,0197	0,0049	0,0062
rfi_imag670	**std**	0,0012	0,0007	0,0079	0,0032	0,0034
rfi_imag87	**mean**	0,0016	0,0012	0,0222	0,0056	0,0068
rfi_imag87	**median**	0,0011	0,0008	0,0199	0,0051	0,0063
rfi_imag87	**std**	0,0014	0,0009	0,0092	0,0037	0,0035
rfi_imag102	**mean**	0,0017	0,0012	0,0222	0,0057	0,0068
rfi_imag102	**median**	0,0013	0,0006	0,0196	0,0050	0,0063
rfi_imag102	**std**	0,0015	0,0010	0,0102	0,0039	0,0036
LDR440	**mean**	0,1732	0,2114	0,0131	0,0313	0,0344
LDR440	**median**	0,1773	0,2094	0,0036	0,0038	0,0036
LDR440	**std**	0,0420	0,0362	0,0253	0,0427	0,0455
LDR670	**mean**	0,2214	0,2647	0,0186	0,0358	0,0380
LDR670	**median**	0,2346	0,2705	0,0118	0,0115	0,0107
LDR670	**std**	0,0454	0,0303	0,0182	0,0387	0,0399
LDR870	**mean**	0,2426	0,2874	0,0194	0,0371	0,0401
LDR870	**median**	0,2577	0,2937	0,0127	0,0120	0,0114
LDR870	**std**	0,0467	0,0300	0,0184	0,0410	0,0431
LDR1020	**mean**	0,2535	0,2993	0,0122	0,0317	0,0362
LDR1020	**median**	0,2689	0,3038	0,0047	0,0044	0,0041
LDR1020	**std**	0,0474	0,0307	0,0210	0,0445	0,0482
LR440	**mean**	64,3056	70,2883	88,7544	77,1293	60,8497
LR440	**median**	63,9365	69,5020	86,8230	75,7680	61,3060
LR440	**std**	9,0500	7,9190	16,3755	16,5475	14,7009
LR670	**mean**	43,1043	50,1851	54,1279	60,6682	47,8367
LR670	**median**	43,1125	49,6430	50,9810	60,6010	45,9210
LR670	**std**	4,3556	4,3933	12,2167	12,2187	10,6693
LR870	**mean**	43,8387	52,6138	42,2470	51,4978	39,3875
LR870	**median**	44,0395	52,1030	38,8990	50,7710	38,2370
LR870	**std**	4,3887	4,7388	11,3011	12,5723	10,4292
LR1020	**mean**	46,0858	56,3432	37,7698	47,0174	35,4478
LR1020	**median**	46,1805	55,7800	34,3160	46,3820	34,7830
LR1020	**std**	4,9359	5,5016	10,6425	11,8851	9,2532

Table S2 - Mean volume particle size distribution as function of radius for each cluster.

Radius(μm)	Clusters mean volume particle size distribution as function of radius ($\mu m^3/\mu m^2$)	
[0.05,	**Cluster 0** ([0.00746956, 0.01921329, 0.03028816, 0.02589049, 0.01613817,	
0.065604	0.01084328, 0.0100023 , 0.01328514, 0.02074599, 0.02914561,	
0.086077	0.03857136, 0.06109866, 0.11383102, 0.17465919, 0.1699613 ,	
0.112939	0.12803852, 0.08359337, 0.0475923 , 0.02261995, 0.0087932 ,	
0.148184	0.00289074, 0.00091922])	
0.194429		
0.255105	**Cluster 1** ([0.01246812, 0.02194062, 0.02617934, 0.01872784, 0.01098944,	
0.334716	0.00810001, 0.00956317, 0.01684171, 0.03133539, 0.04367875,	
0.439173	0.04915988, 0.06776062, 0.13481862, 0.26612858, 0.28392134,	
0.576227	0.17208717, 0.07599688, 0.03015943, 0.01177114, 0.00472025,	
0.756052	0.0020729 , 0.00112952])	
0.991996		
1.301571	**Cluster 2** ([0.00053764, 0.0060057 , 0.02952382, 0.06068518, 0.0691188 ,	
1.707757	0.05030102, 0.02522124, 0.0117204 , 0.00697745, 0.00614911,	
2.240702	0.00750429, 0.01076075, 0.01521814, 0.01884122, 0.02239693,	
2.939966	0.02783376, 0.03364816, 0.03310523, 0.02281578, 0.01021044,	
3.857452	0.00288488, 0.00052916])	
5.06126		
6.640745	**Cluster 3** ([0.00054038, 0.0035513 , 0.0161313 , 0.03539967, 0.04535381,	
8.713145	0.04615952, 0.04072243, 0.02634627, 0.01467314, 0.0099306,	
11.4323	0.00974848, 0.01311973, 0.01957372, 0.02657935, 0.03080725,	
15.0]	0.03367282, 0.03391425, 0.02744734, 0.01601158, 0.00667934,	
	0.00204447, 0.00047592])	
	Cluster 4 ([0.00057772, 0.00390664, 0.01567725, 0.03059 , 0.03936049,	
	0.04083636, 0.03145199, 0.01545702, 0.00783089, 0.00688574,	
	0.00928641, 0.01456289, 0.02156659, 0.02550996, 0.02730787,	
	0.03069614, 0.03361232, 0.02923737, 0.01757699, 0.00721282,	
	0.0021053 , 0.00047402])	

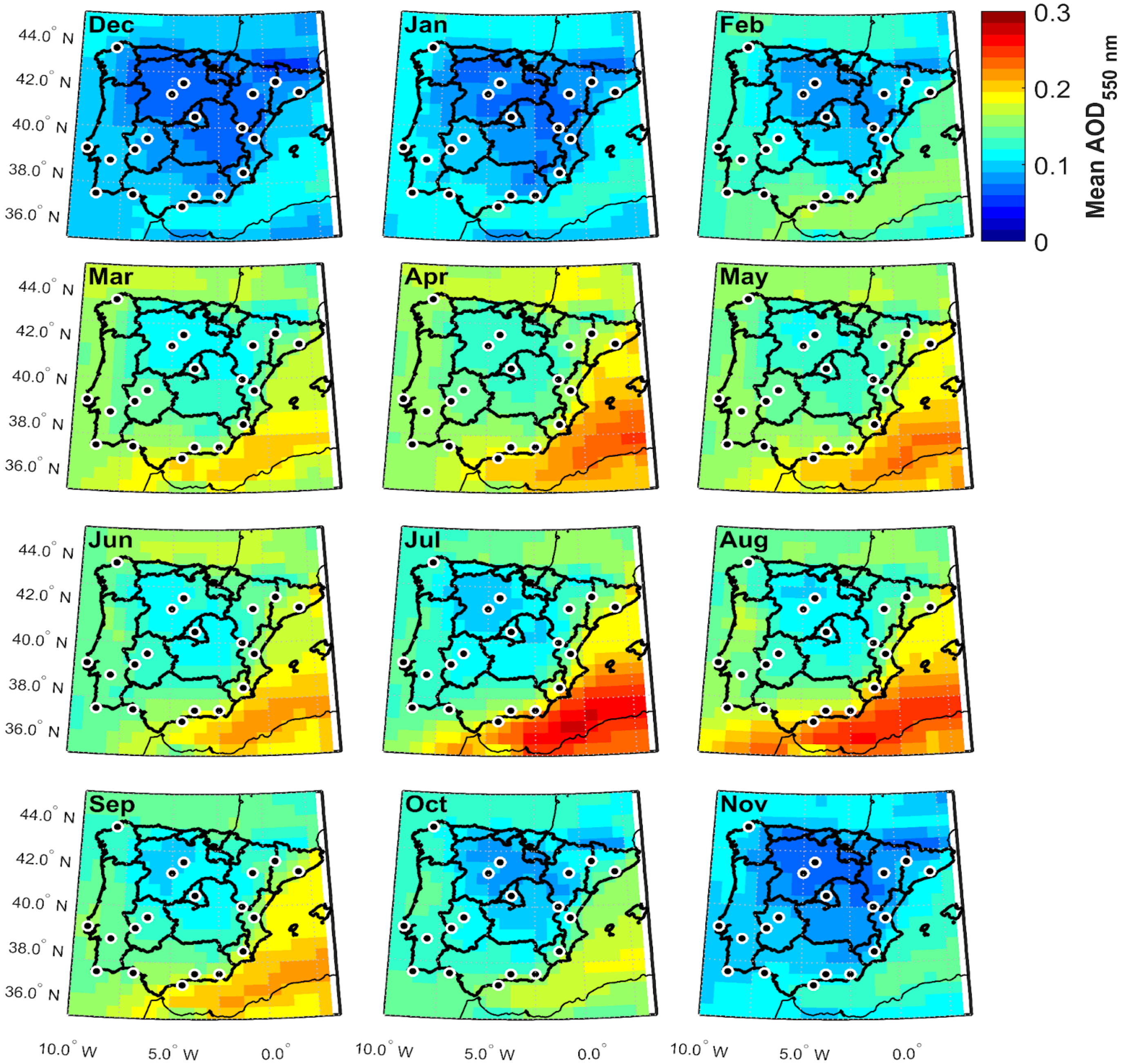

Figure S1 - Monthly climatology of Aerosol Optical Depth at 550 nm over the Iberian Peninsula from MERRA-2 aerosol products to provide a perspective on the spatial and monthly dynamic of aerosol loading in the region.

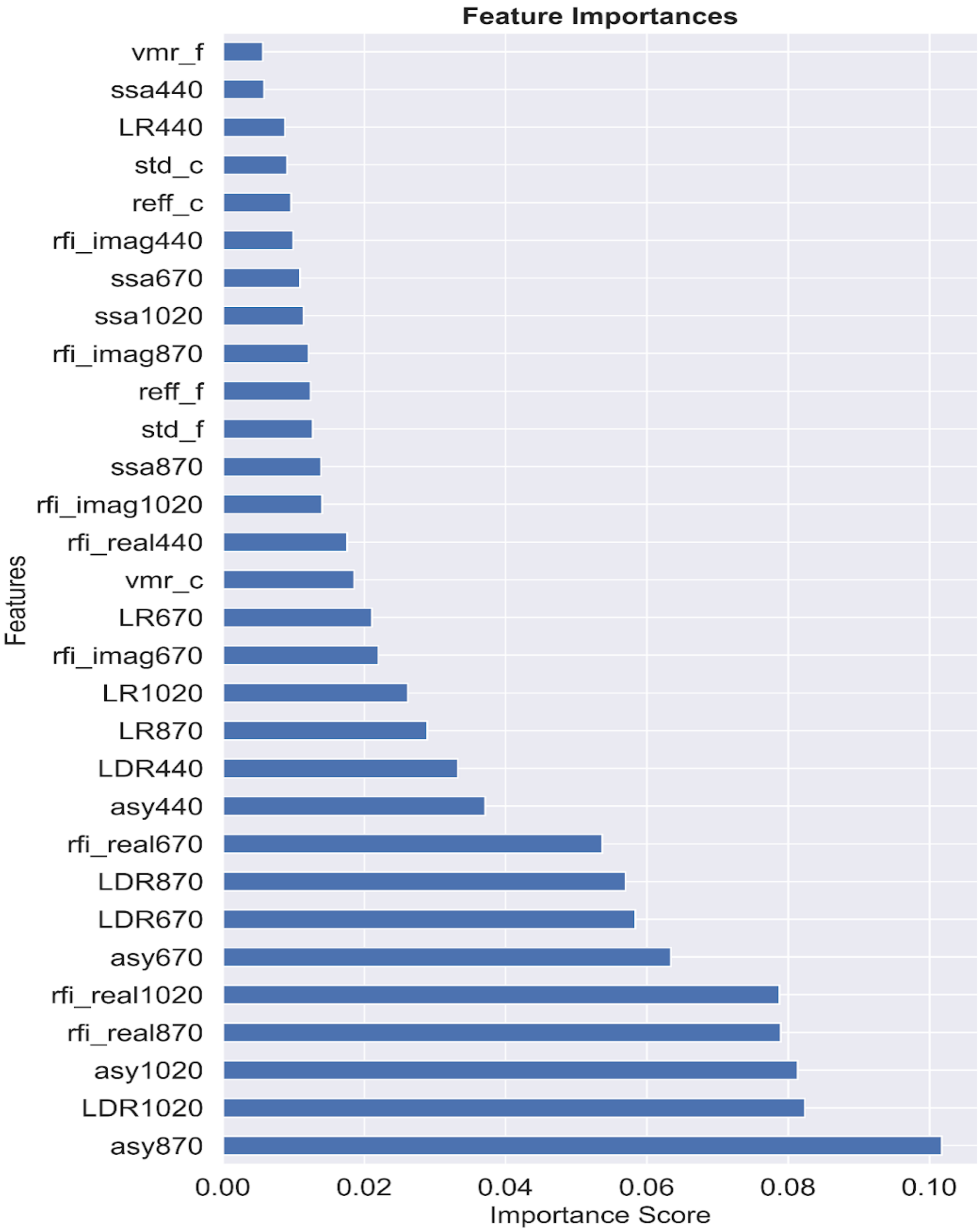

Figure S2 - The relative importance of predictors from Table 1(in the manuscript)) in cluster prediction. It is possible to observe consistency between this score scale and the importance score scale regarding aerosol types (Figure 10). In Figure 10, the dust (dst) mass variability emerges as the most relevant factor for determining which cluster should be applied. In the figure below, which shows the importance of the optical parameters from Table 1 for clustering, the scores are well distributed (maximum close to 0.1). However, it became clear that higher wavelengths (near infrared 870 and 1020 nm) and optical parameters (ASYmmetry parameter and Linear Depolarization Ratio), that best differentiate dust from other aerosol types, presented the highest importances

The following tables present a summary of the results obtained for the cluster robustness analysis.

Table S3 - Results for jitter and bootstrap robustness tests. The jitter robustness test was conducted by adding Gaussian noises testing different uncertainty possibilities (5%, 10%, 20%) to the standardized feature matrix.

Random Perturbation	Jitter Robustness (5%/ 10% / 20%)	Bootstrap
Mean ARI	0.8570 / 0.8511 / 0.8183	0.8405
Std ARI	0.2039 /0.1887 / 0.1698	0.1925
Min ARI	0.4823 / 0.4821/ 0.4801	0.4646
Max ARI	0.9813 / 0.9650 / 0.9215	0.9837

Table S4 - Results for the cluster retention score metric. The test was also conducted for three scenarios by adding Gaussian noises testing different uncertainty possibilities (5%, 10%, 20%) to the standardized feature matrix.

Cluster-wise Stability	5%	10%	20%
Cluster 0	0.9692	0.9638	0.9456
Cluster 1	0.9916	0.9844	0.9721
Cluster 2	0.9905	0.9907	0.9851
Cluster 3	0.8890	0.8922	0.8830
Cluster 4	0.9292	0.9377	0.9339

Table S5 - Results for the consensus (co-assignment) matrix metric. The test was also conducted for three scenarios by adding Gaussian noises testing different uncertainty possibilities (5%, 10%, 20%) to the standardized feature matrix.

Co-assignment matrix	5%	10%	20%
Mean Intra-cluster probability	0.9302	0.9276	0.9087
Mean Inter-cluster probability	0.0512	0.0500	0.0561