Aerosol classification using machine learning on sun photometer and lidar data with focus on United Kingdom (climatology)

Joelle Buxmann, Martin Osborne, Mike Protts, Debbie O'Sullivan








Panuganti Devara, Pawan Gupta, Govindan Pandithurai
Elena Lind

Thanks to all the engineers for their hard work on site!

This calibration of Met Office sun photometer data is part of a project that is supported by the European Commission under the Horizon 2020 – Research and Innovation Framework Programme, H2020-INFRADEV-2019-2, Grant Agreement number: 871115.

Thanks to the PHOTONS team in Lille!

We thank the (PI(s) and Co-I(s)) and their staff for establishing and maintaining the all sites used in this investigation.

www.metoffice.gov.uk

Operational lidar/sun photometer (&ceilometer) network

Main Function:

To detect volcanic ash layers over the UK for VAAC-London

Comprised of:

10 Lidar (Raymetric):

9 fixed operational sites

+ 1 research lidar

10 Sun-photometer (CIMEL) co-located with the Lidars

Ceilometers:

(could be used in future)

2 CL61 (depol.,Vaisala)

10 CHM15k (Lufft, OTT H.)

32 CL31 (Vaisala)

Aim here:

Develop a machine learning algorithm for

Aerosol classification based on sun photometer data

- using lidar data as validation

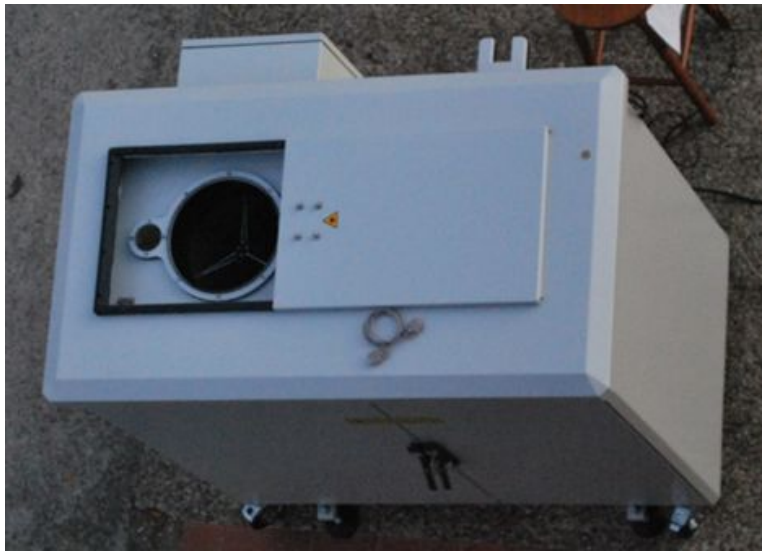
Future use for, e.g.:

- Nowcasting
- air quality alerts
- support volcanic ash identification

Met Office

Depol. Raman Lidar (Raymetrics)

Main features	Raymetrics Lidar (10)
Laser source	355nm (NdYAG, flash lamp-pumped)
Energy/pulse	~50mJ/pulse
Frequency	20 Hz
Power	~1 W
Detector	Both Analog and Photon Counting at 355nm (depolarization) and 387nm
Range recorded	120 km
Practical range	15km
Range resolution	15m
Time resolution	1min



Met Office

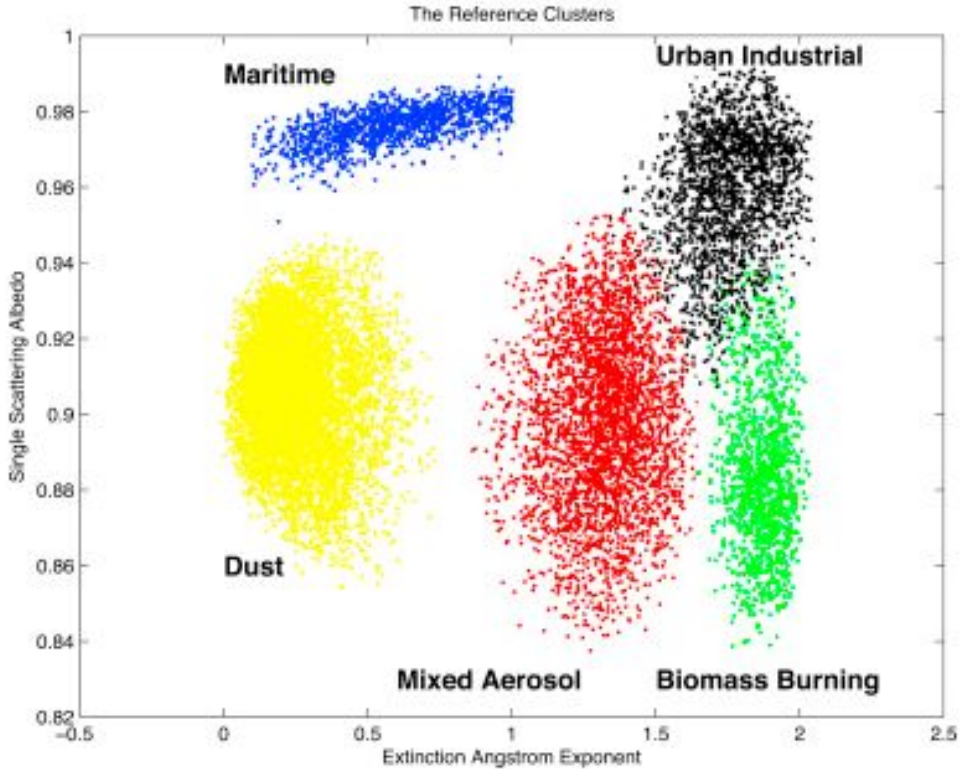
Our sun photometer (polar. version)

- Two high sensitivity detectors(UV-Vis and IR)
- 9 filter bands [nm] 340, 380, 440, 500, 675, 870, 936, 1020, 1640
- Polarization in 3 directions
- Direct sun measurements for Aerosol Optical Depth & Angstrom Exponent
- Indirect sun measurements for inversion products, e.g. Size distribution

Reference aerosol clusters

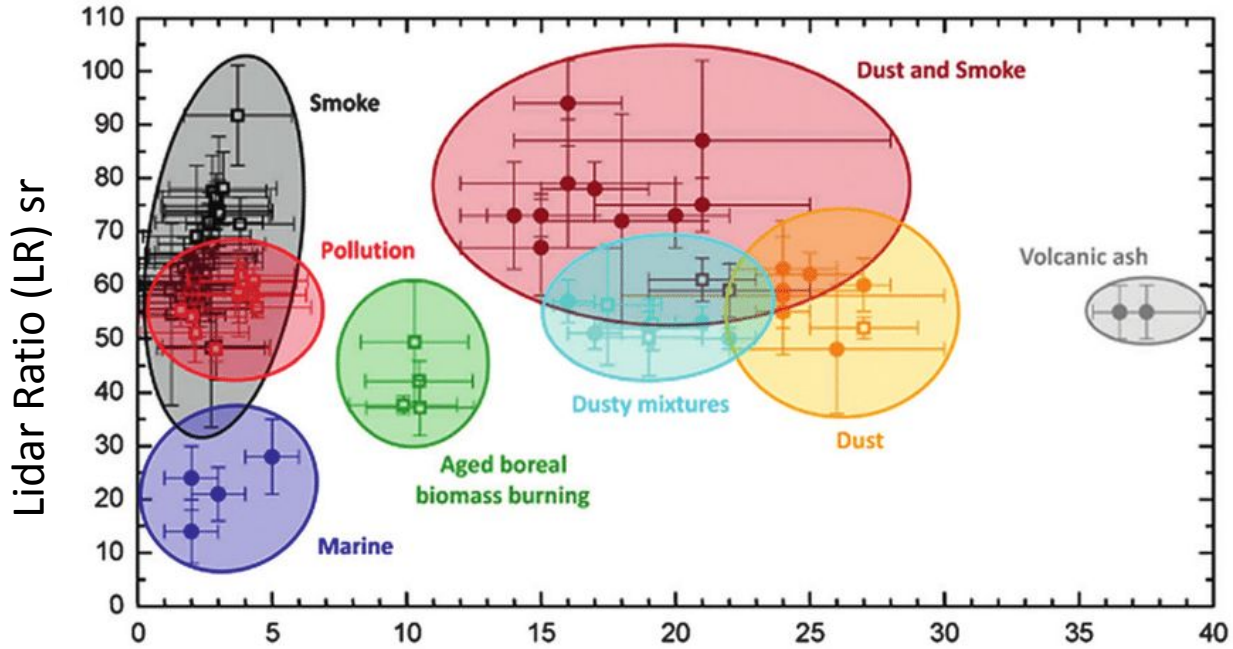
'in an ideal world'

Sun photometer (column integrated properties)
- used for machine learning



[P. Hamill et al. 2016, Atmospheric environment]

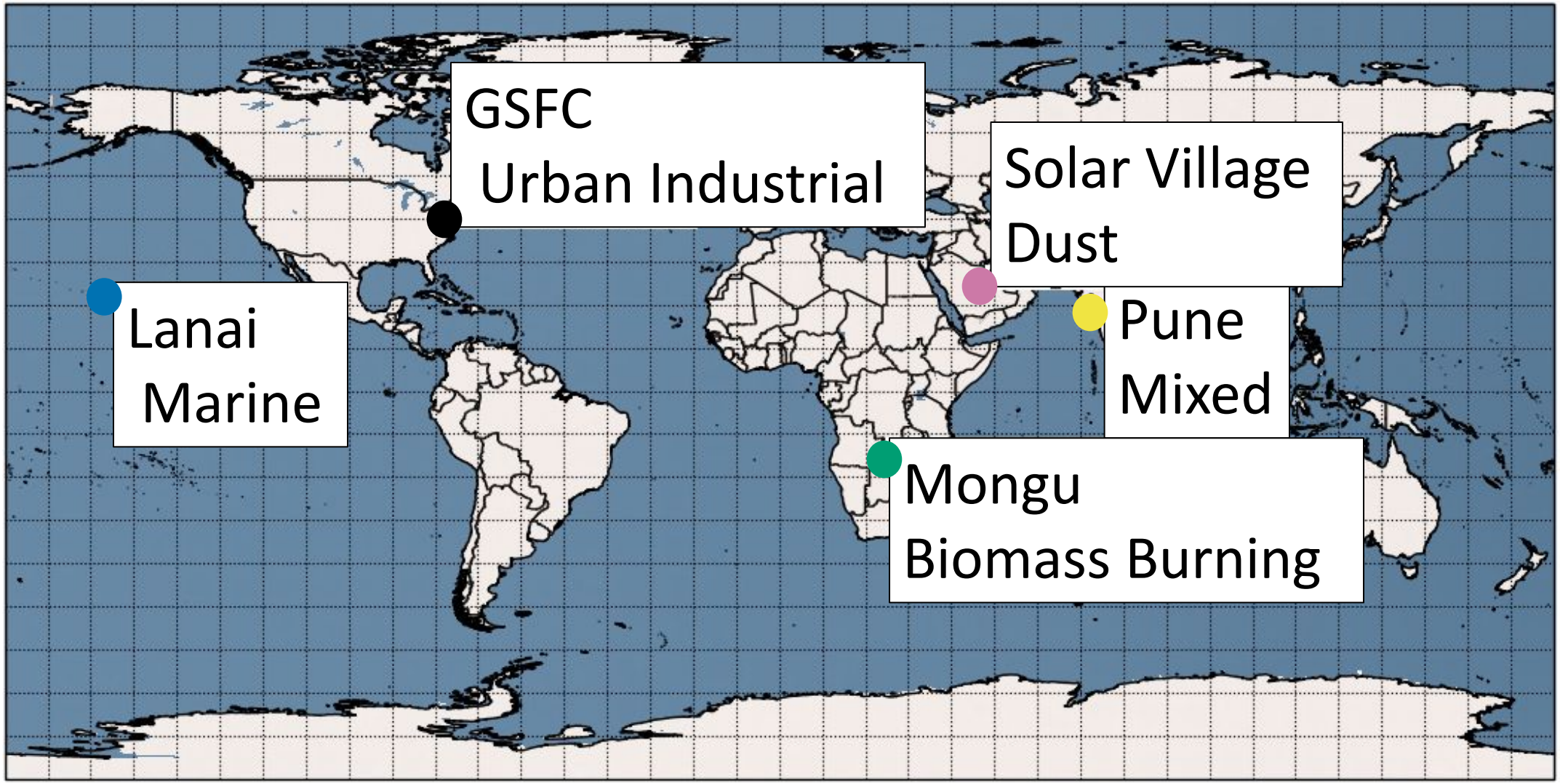
Lidar (height resolved properties)
- used for validation



adapted from [A.J. Illingworth et al. 2015]

Reference aerosol clusters –

'in real world-select specific sites based on dominant aerosol from literature'



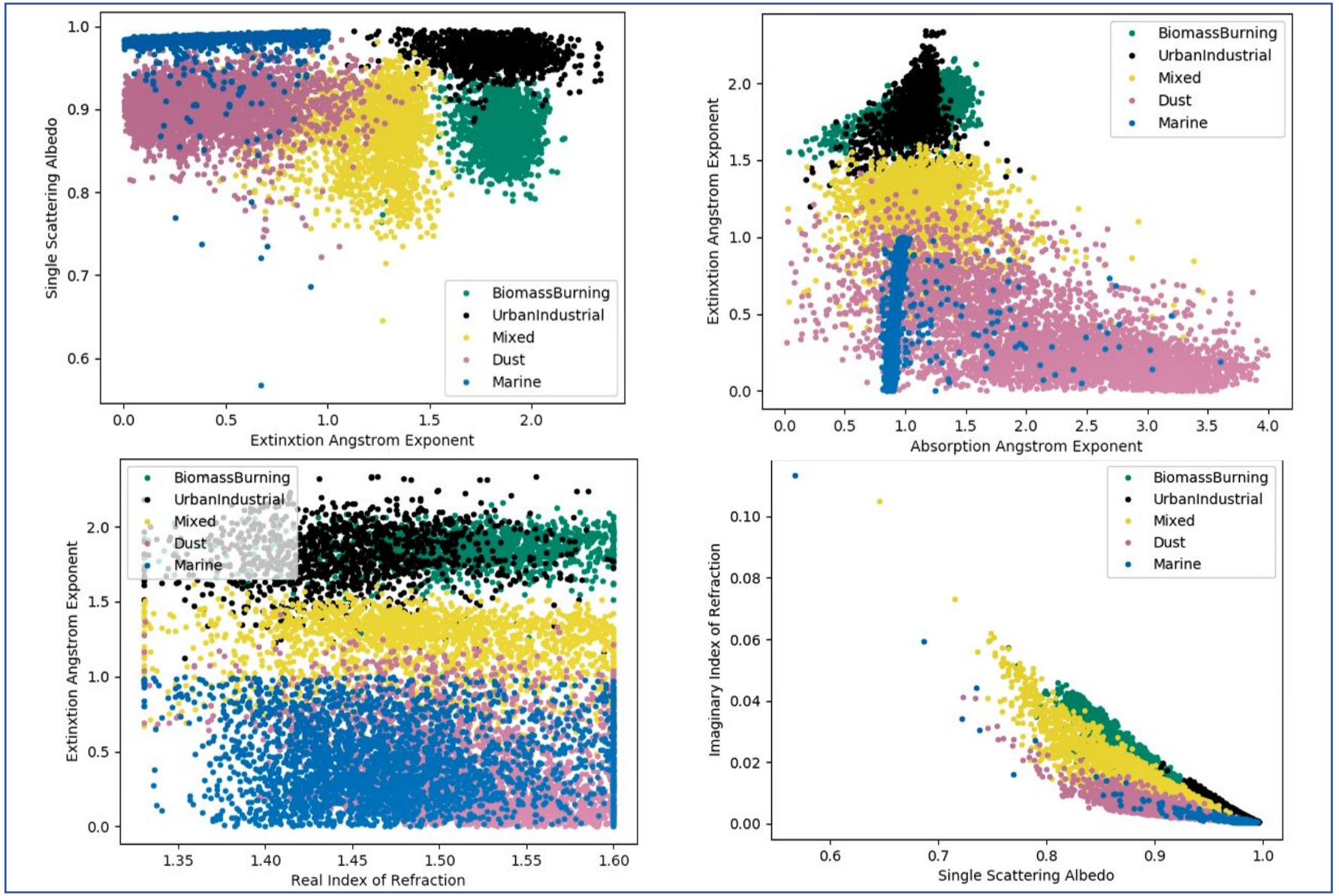
[Site selection from: P. Hamill et al. 2016, Atmospheric environment]

Reference aerosol clusters –'in real world'

Selected Parameters from quality controlled AERONET data level 2.0: https://aeronet.gsfc.nasa.gov/

- Single Scattering Albedo (SSA) (@440nm)
- Extinction Angstrom Exponent (EAE)
- Absorption Angstrom Exponent (AAE) (@440nm-870nm)
- Imaginary Index of Refraction (IRI) (@440nm)
- Real Index of Refraction (RIR) (@440nm)

(Marine lev 1.5, filtered EAE<1, as in Sayer et al.2012)

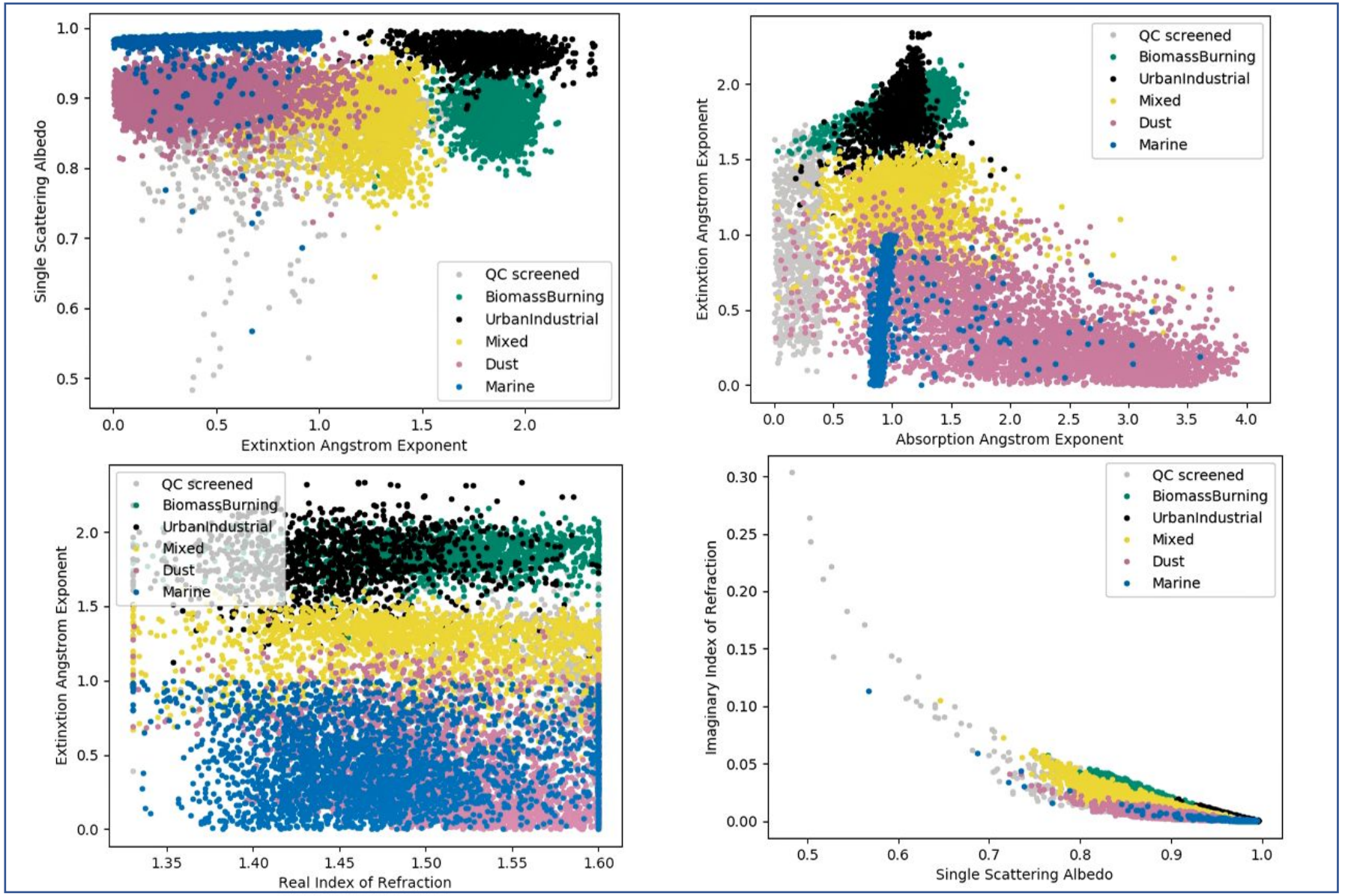


Reference aerosol clusters –'in real world'

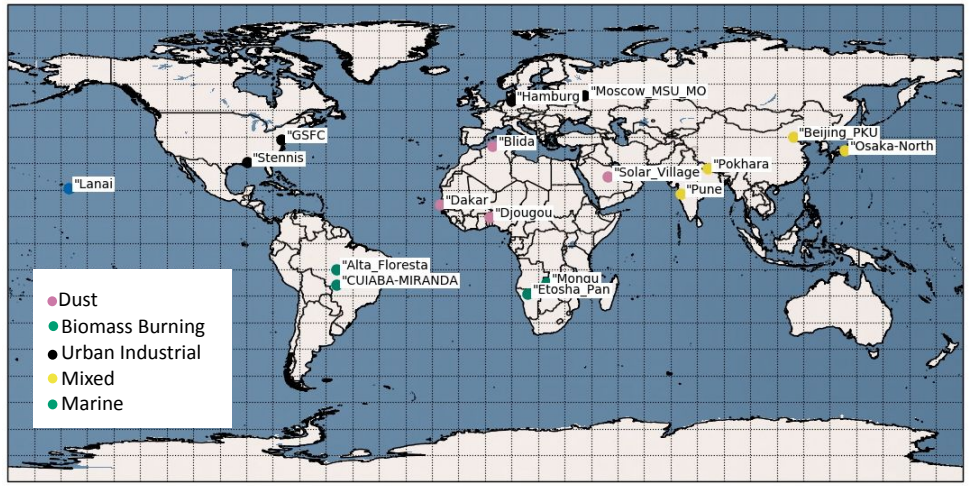
Selected Parameters from quality controlled AERONET data level 2.0:

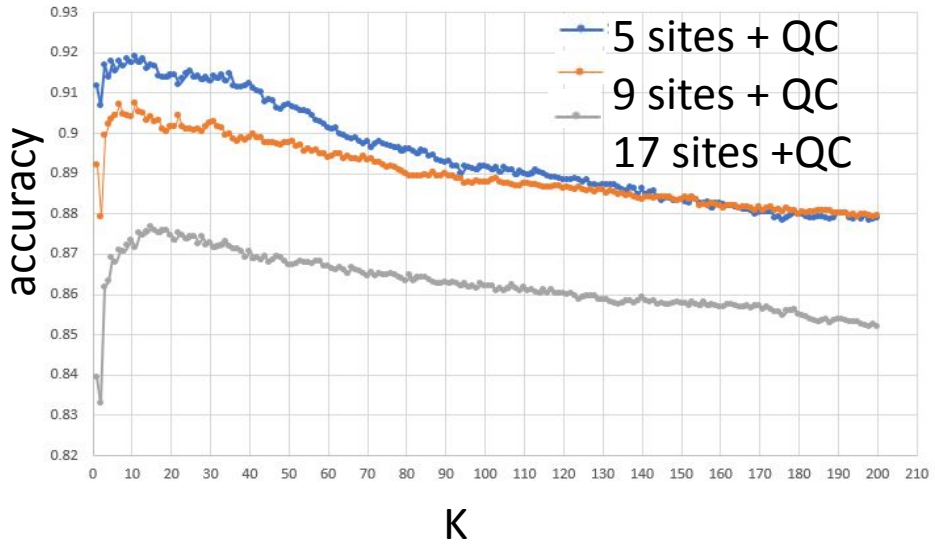
https://aeronet.gsfc.nasa.gov/

Additional cluster: 'QC screened' from level 1.5 data with:
SSA<0.9
and
AAE<0.4
those were not/rarely found in the chosen reference clusters level 2.0 data



Use Supervised Machine learning approach





- K Nearest Neighbour (KNN) technique
- Divide data set in training (70%) and test (30%) set
- Use Mahalanobis distance [P.Hamill et al. 2016]:

Then the Mahalanobis distance from the test point to the cluster is defined by

$$D_M = \left[(\mathbf{x}-\mathbf{m})^{\mathbf{T}}\mathbf{S}^{-1}(\mathbf{x}-\mathbf{m})\right]^{1/2},$$

where $\mathbf{S} = \mathbf{cov}(\mathbf{x}_i,\mathbf{x}_j)$ is the covariance matrix whose elements are defined by

$$\mathbf{S} = E[(\mathbf{x}-\mathbf{m})(\mathbf{x}-\mathbf{m})]^{\mathbf{T}}.$$

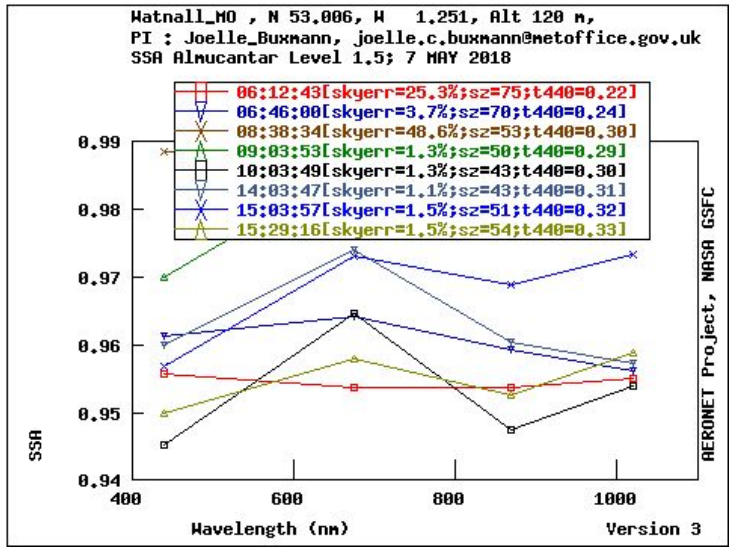
- KNN with K=13
- 5 sites + lev15 QC class

- □92% accuracy on test data set

(predicted versus assigned class)

Example Case Study for validation on lidar data

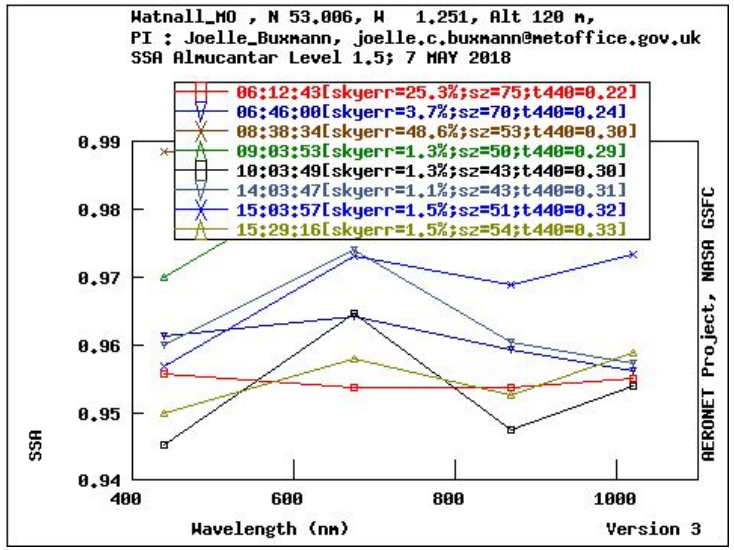
ML- Classification on
Sun photometer data:



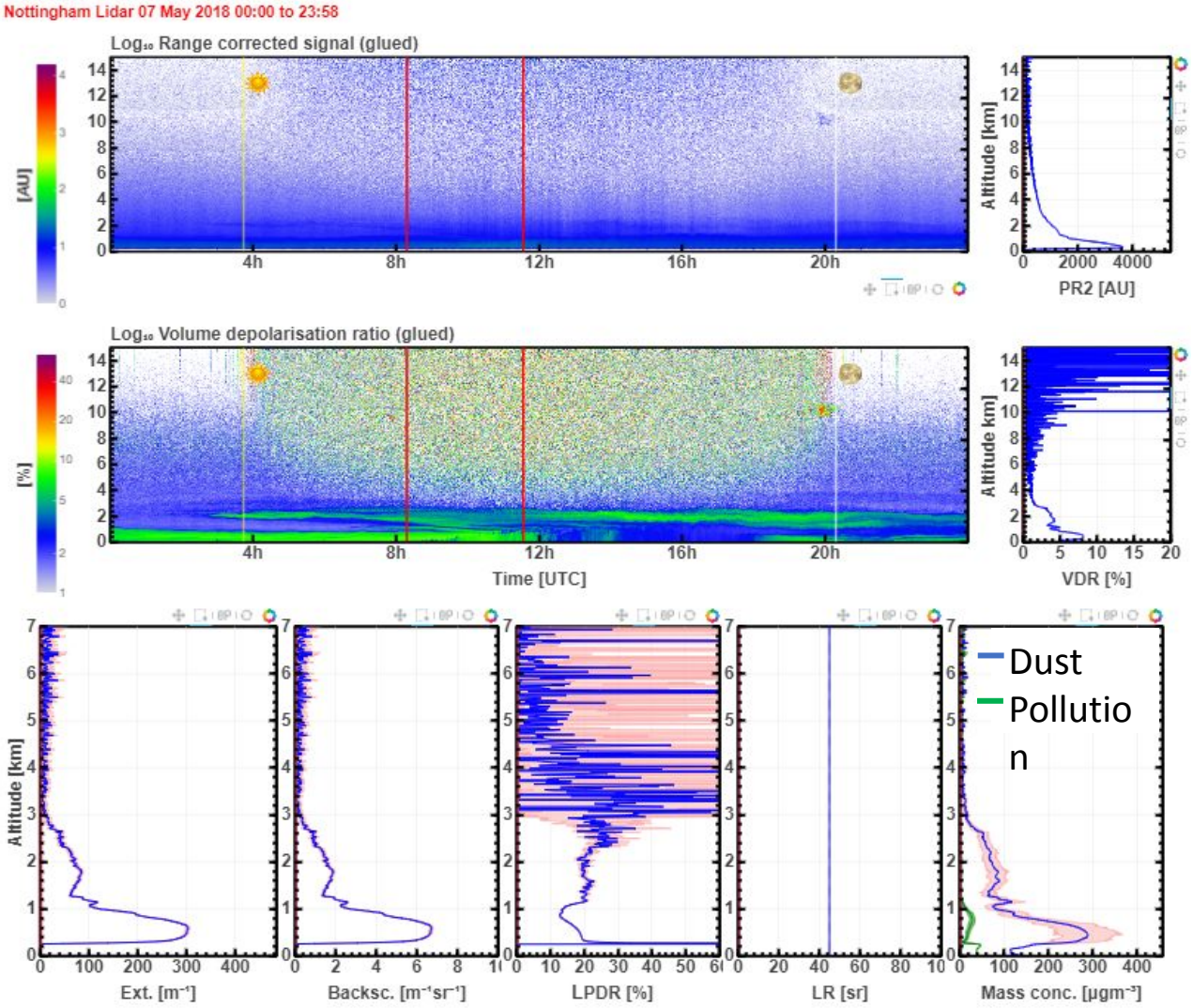
SSA	EAE	AAE	RRI	IRI
0.9558	1.1553	1.0855	1.4626	0.0044
0.9613	1.1576	1.1149	1.4458	0.0036
0.9886	1.0749	1.0224	1.6	0.001
0.97	1.0076	1.7434	1.5965	0.0027
0.9452	0.9658	1.1285	1.5825	0.0052
0.96	1.0808	1.2023	1.4898	0.0031
0.9569	1.21	1.7522	1.4913	0.0036
0.95	1.2376	1.3507	1.4792	0.0044

Example Case Study for validation on lidar data

ML- Classification on
Sun photometer data:

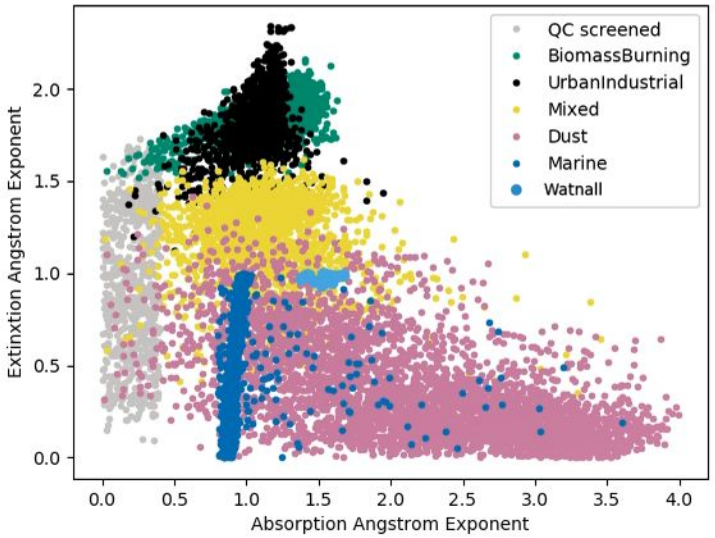


SSA	EAE	AAE	RRI	IRI
0.9558	1.1553	1.0855	1.4626	0.0044
0.9613	1.1576	1.1149	1.4458	0.0036
0.9886	1.0749	1.0224	1.6	0.001
0.97	1.0076	1.7434	1.5965	0.0027
0.9452	0.9658	1.1285	1.5825	0.0052
0.96	1.0808	1.2023	1.4898	0.0031
0.9569	1.21	1.7522	1.4913	0.0036
0.95	1.2376	1.3507	1.4792	0.0044



Example Case Study for validation on lidar data

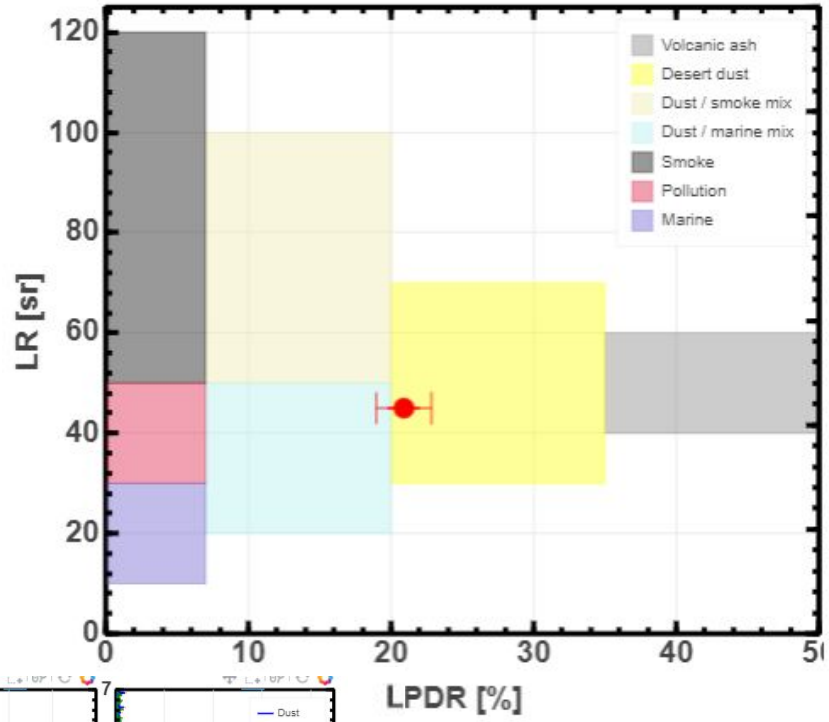
ML- Classification on
Sun photometer data:

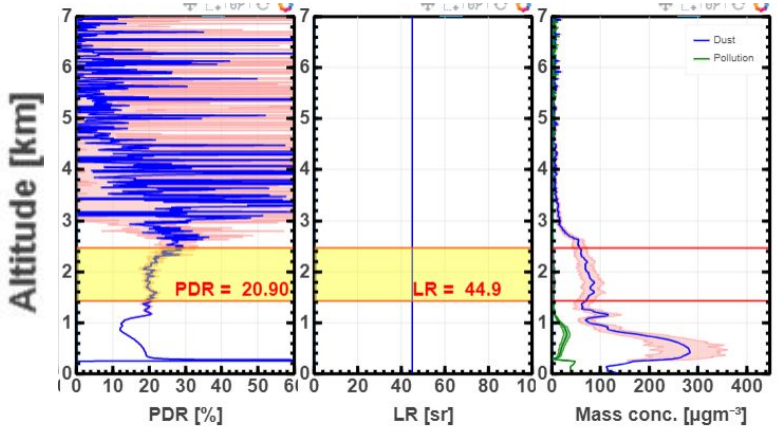


SSA	EAE	AAE	RRI	IRI
0.9558	1.1553	1.0855	1.4626	0.0044
0.9613	1.1576	1.1149	1.4458	0.0036
0.9886	1.0749	1.0224	1.6	0.001
0.97	1.0076	1.7434	1.5965	0.0027
0.9452	0.9658	1.1285	1.5825	0.0052
0.96	1.0808	1.2023	1.4898	0.0031
0.9569	1.21	1.7522	1.4913	0.0036
0.95	1.2376	1.3507	1.4792	0.0044

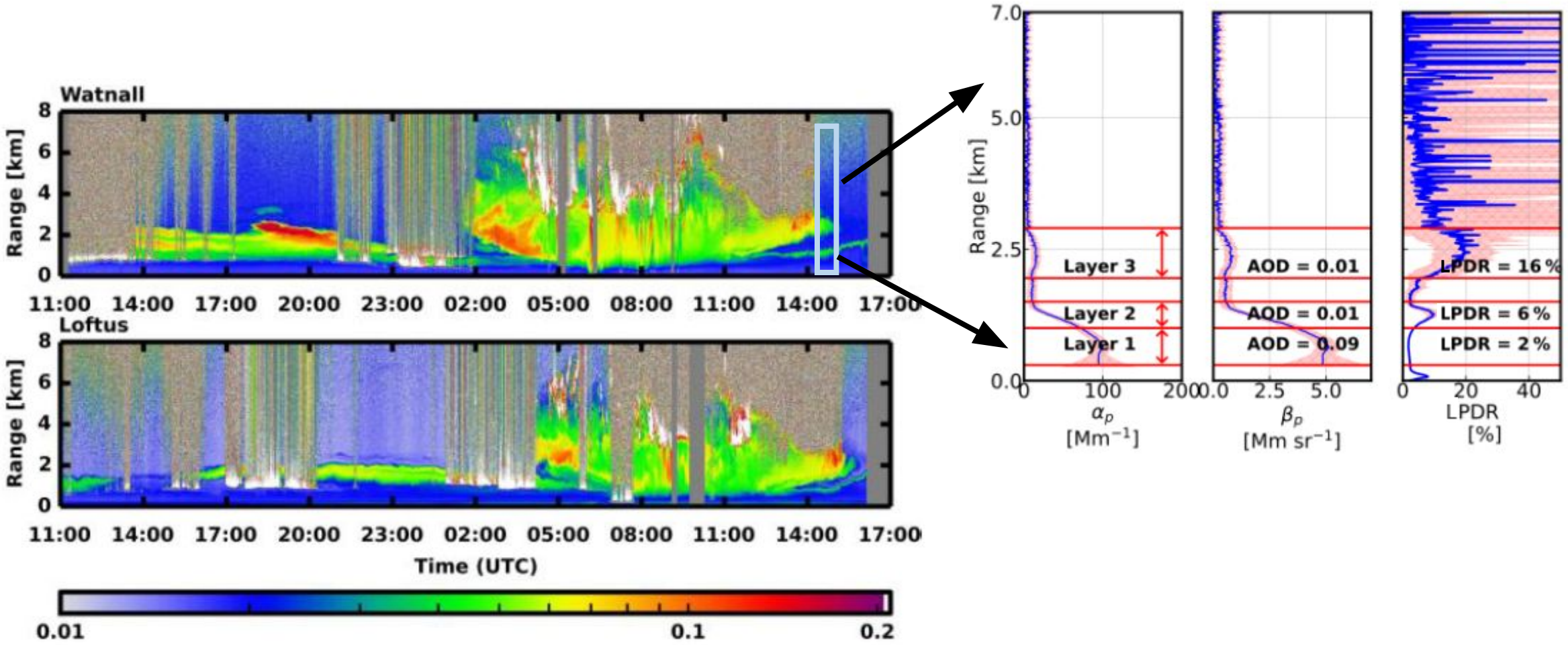
Dust ←

lidar





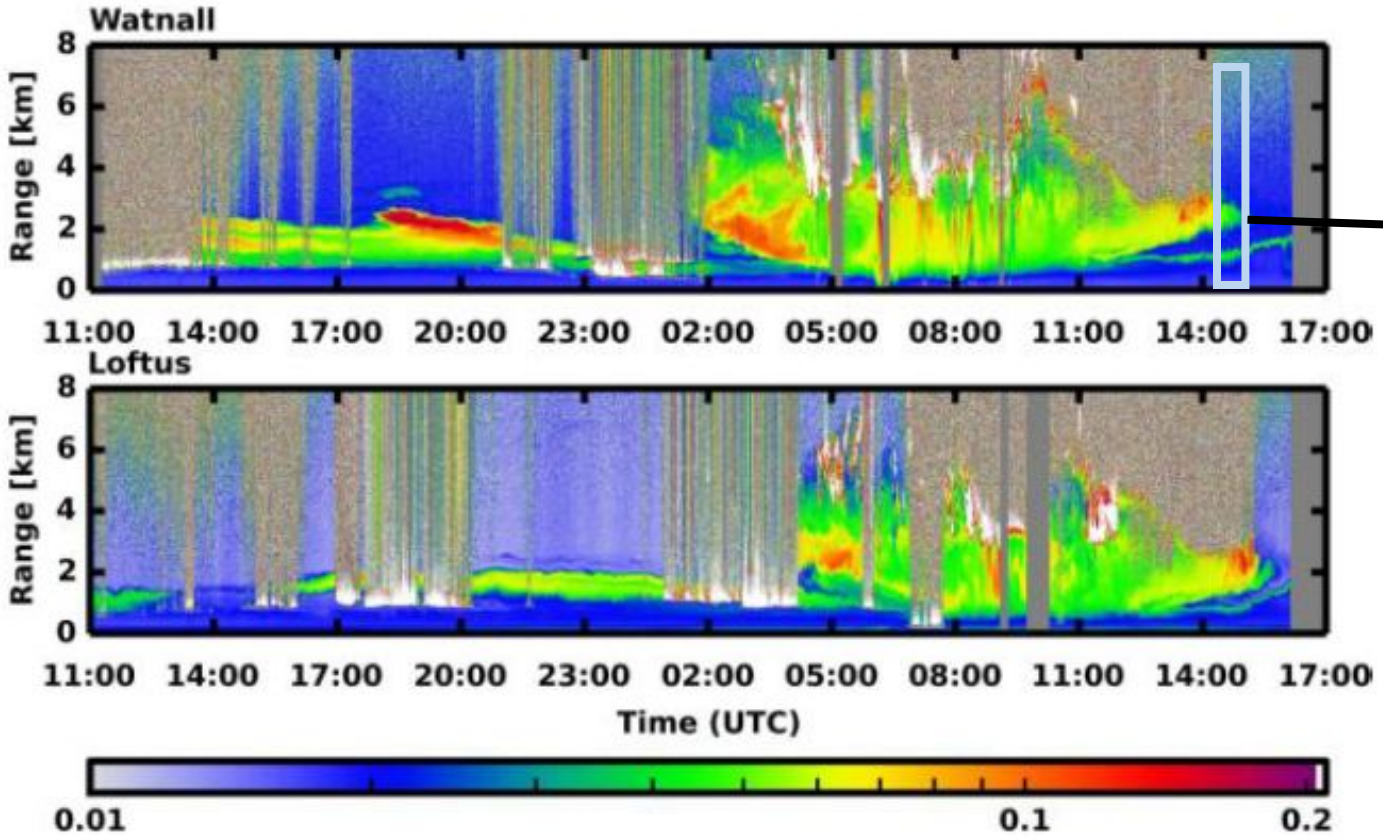
Example Case Study for validation on lidar data: October 2017

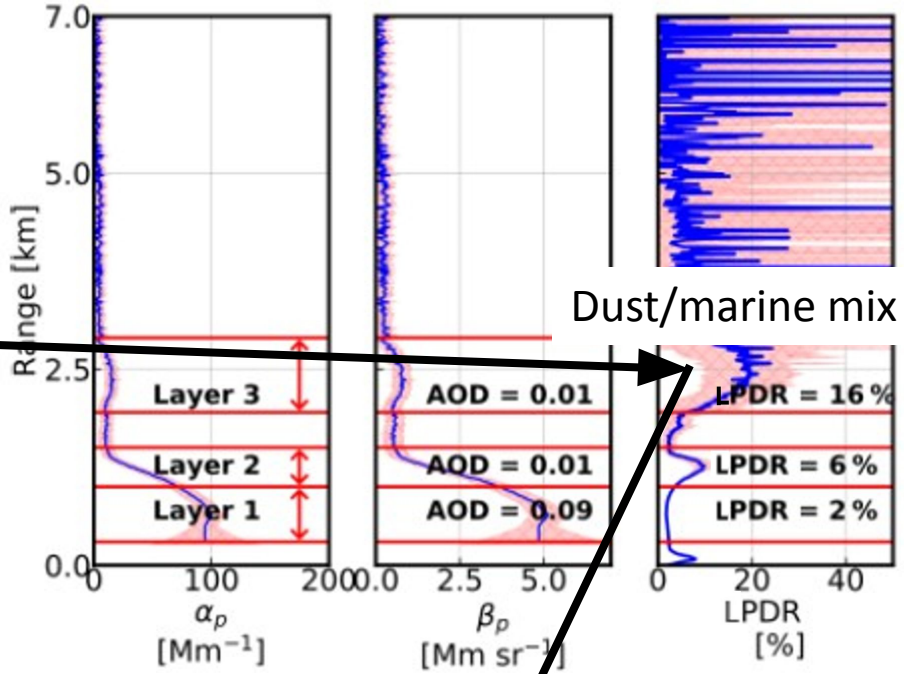


linear volume depolarisation ratios for 15 and 16 October 2017

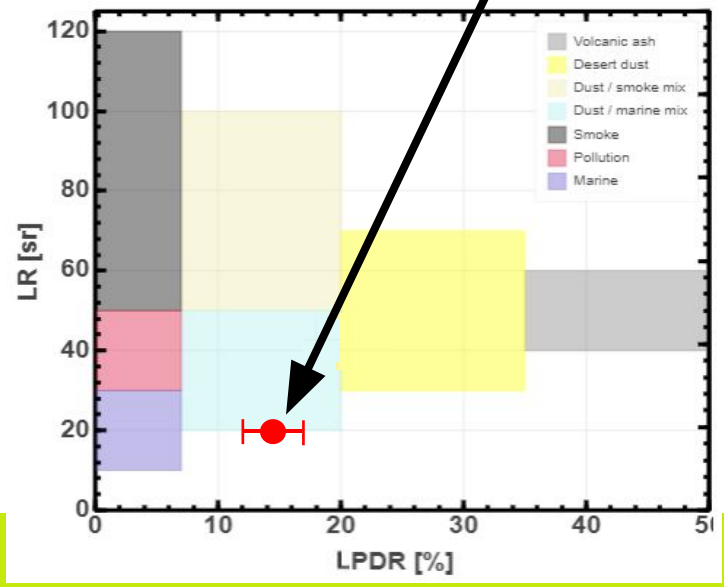
[Storm Ophelia: Osborne et al. 2019, ACP]

Example Case Study for validation on lidar data: October 2017





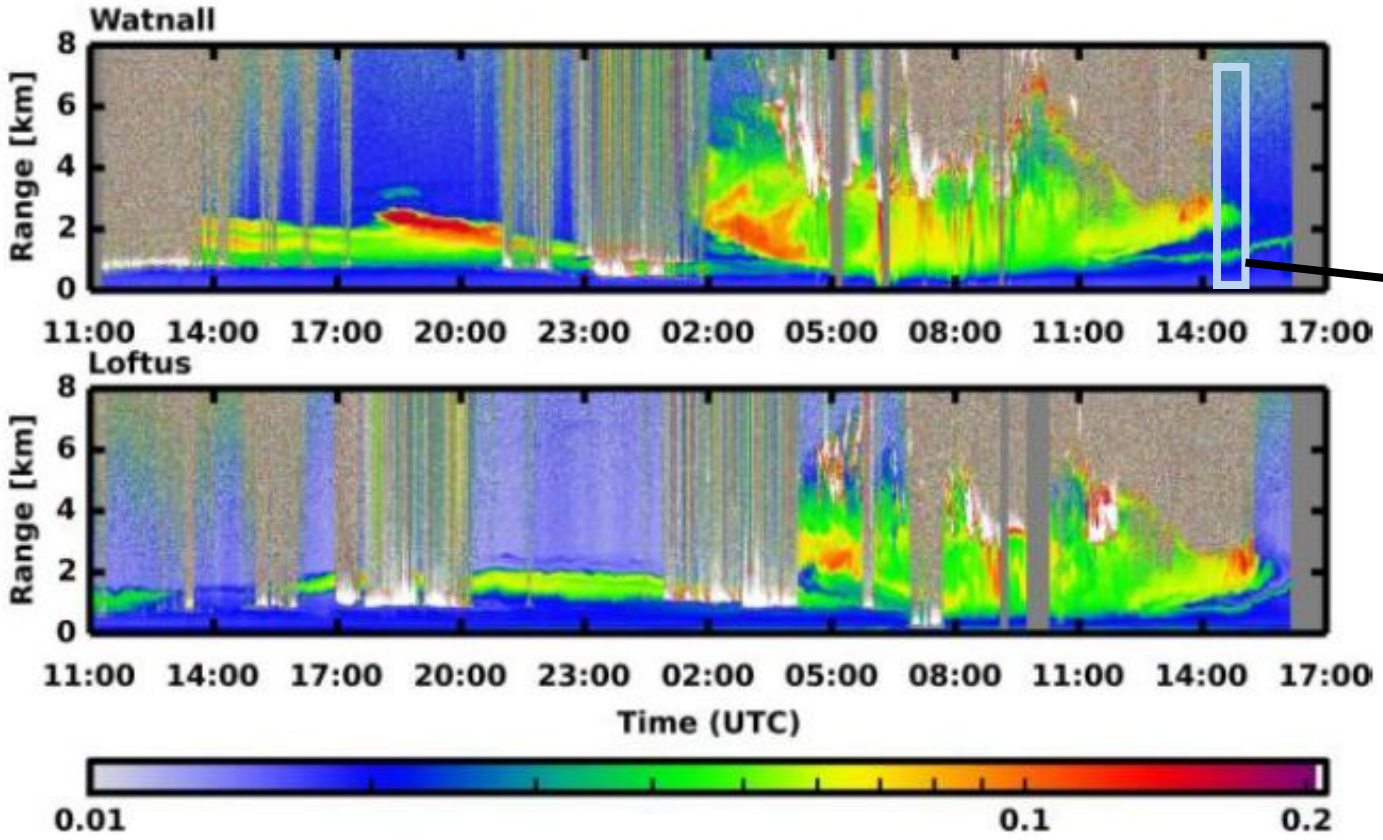
Dust/marine mix

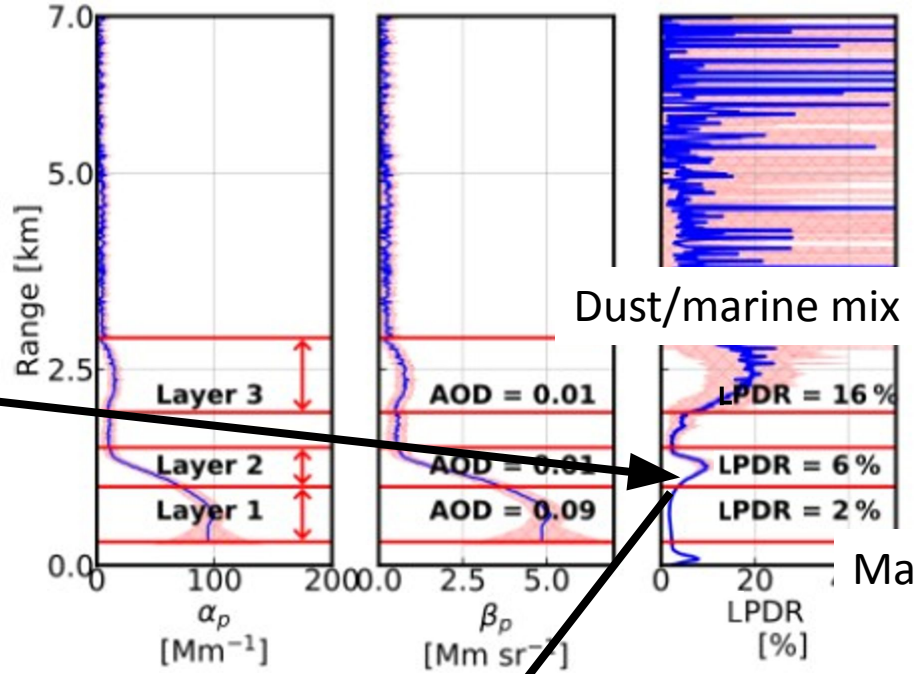


Lidar volume linear depolarisation ratios for 15 and 16 October 2017

[Storm Ophelia: Osborne et al. 2019, ACP]

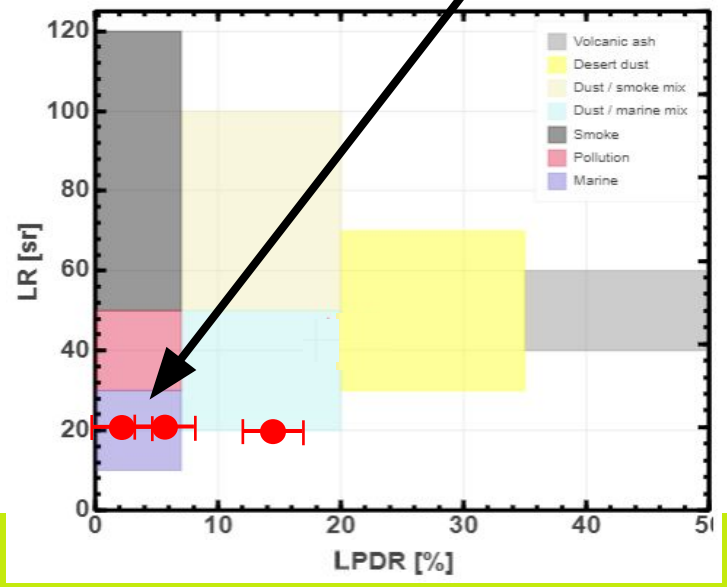
Example Case Study for validation on lidar data: October 2017





Dust/marine mix

Marine



Lidar volume linear depolarisation ratios for 15 and 16 October 2017

[Storm Ophelia: Osborne et al. 2019, ACP]

Example Case Study for validation on lidar data: October 2017

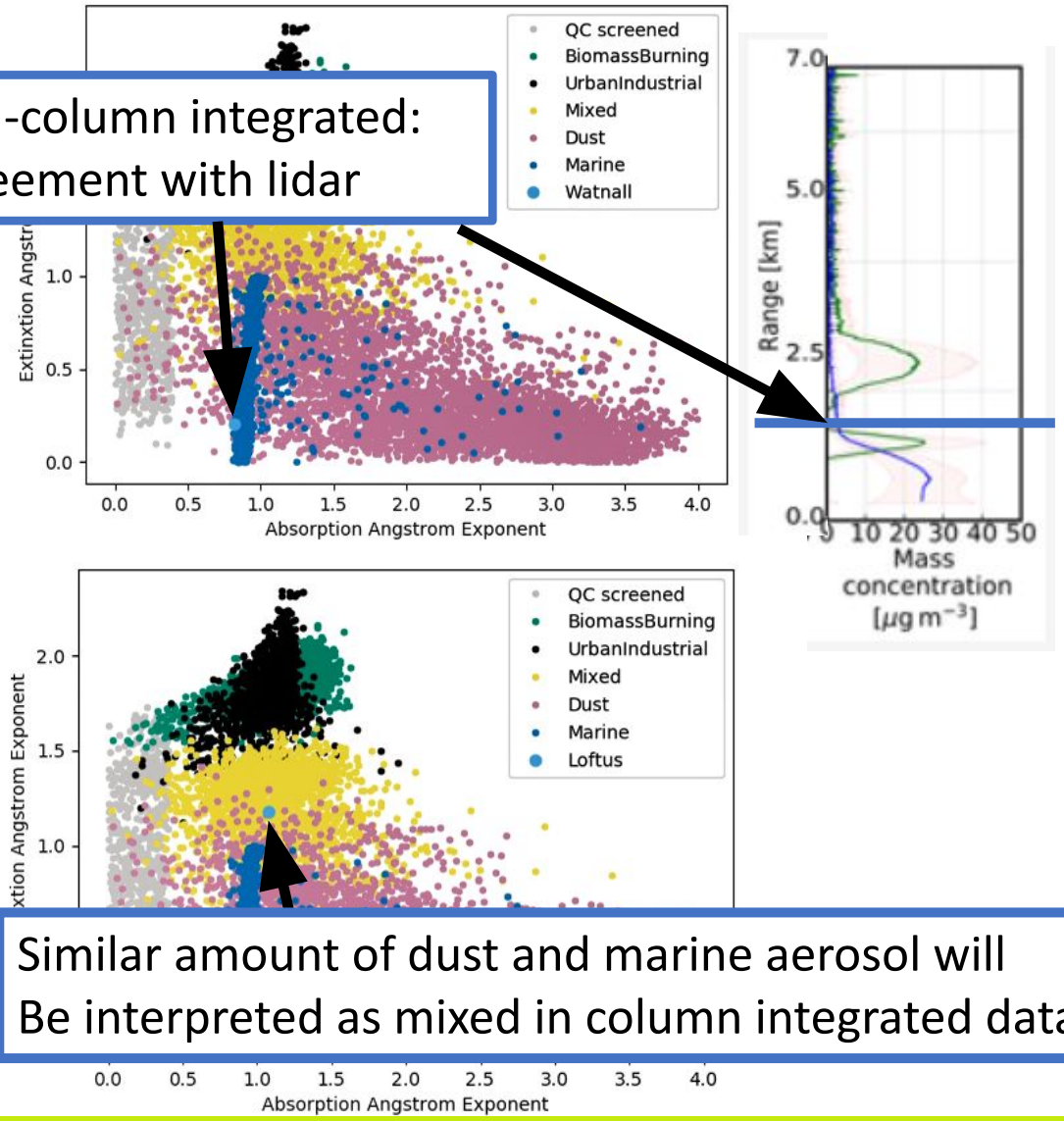
Location	Date	Time	Layer height(km)	PLDR (%)	LR (sr)	Max concentration (μg m−3)	Aerosol type lidar	Aerosol type SP-M
Watnall	16/10/17	14:30 to 15:00	0.3 to 1 km	2±0.3	19.5	25.2±10	Marine	Marine
			1 to 1.5 km	6±2.1	19.5	30±10	Marine	
			1.95 to 2.9 km	16±3.7	19.5	25±9	Dust mix	
Loftus	15/10/17	15:00 to 15:30	0.3 to 1.1 km	2±0.4	27	37±14	Marine	Mixed
			1.4 to 2.45 km	26±7.2	27	58±30	Dust	

[Storm Ophelia: Osborne et al. 2019, ACPD]

Main aerosol type -column integrated:
Marine SP in agreement with lidar

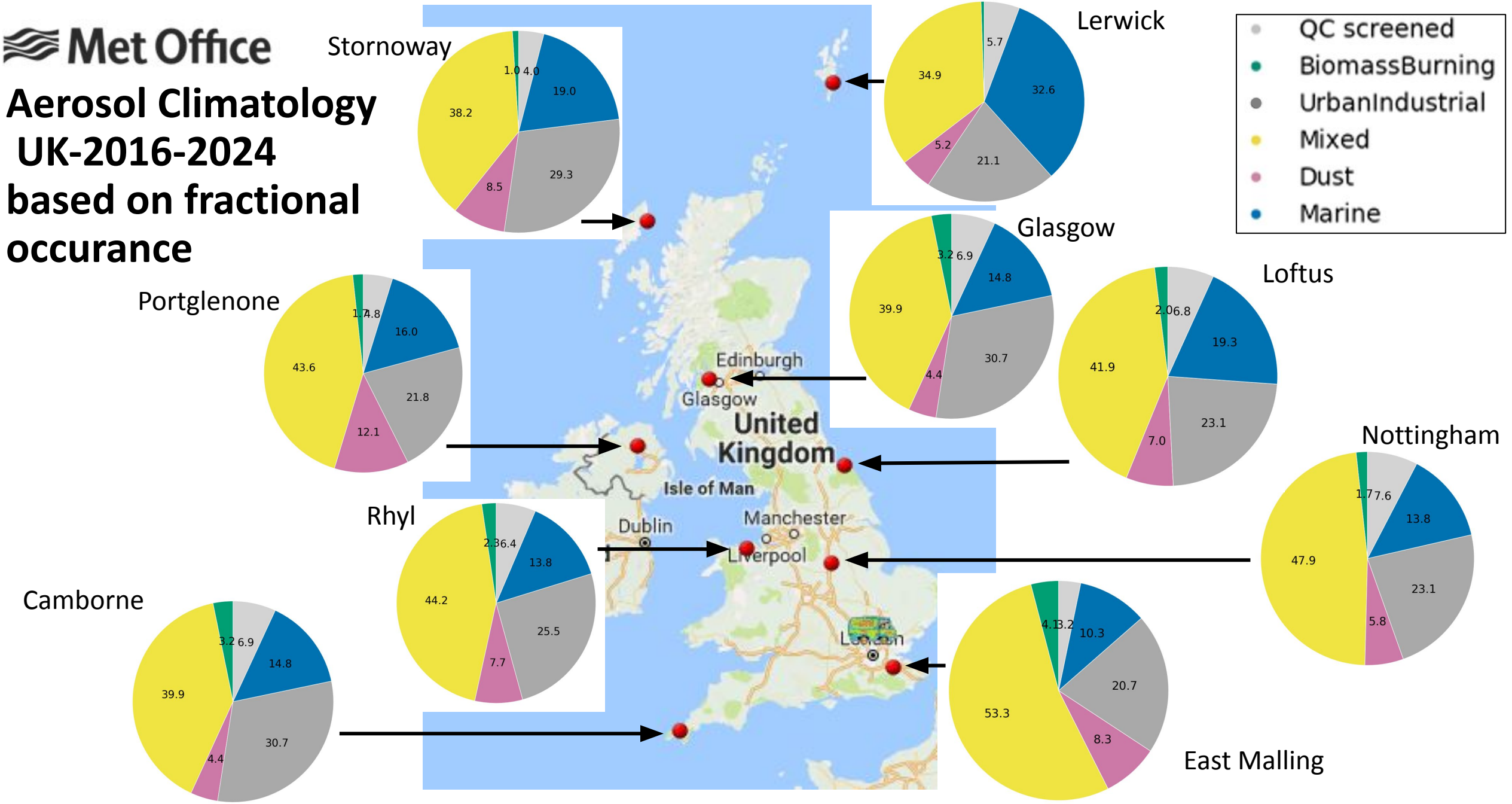
Sun photometer (SP)

Lidar



Similar amount of dust and marine aerosol will
Be interpreted as mixed in column integrated data

Aerosol Climatology UK-2016-2024 based on fractional occurance



Questions?
Please contact me:
Joelle.c.buxmann@metoffice.gov.uk

Summary

- Supervised Machine Learning (K Nearest Neighbour) has been applied to built an aerosol Classification Scheme
- Preliminary results show high accuracy 92% on test/train data set including level 1.5 data suitable for Nowcasting
- Case studies so far show good agreement with lidar prediction of aerosol types
- Climatology: coastal sites show higher marine aerosols, while generally the UK seems to be dominated by mixed aerosols

Outlook

- Add Volcanic Ash to the scheme
- Validation for more case studies
- Climatology: Look at seasonal variations; account for aerosol amount
- Satellite data + CAMS model comparison
- Write Nowcasting application