Evaluation of climate and air quality models based on observations

aerosol extreme events, aerosol trends, …

A. Mortier, MET Norway

PROGRAMME OF THE EUROPEAN UNION

Outline

- introduction (models, products, challenges, …)
- some results
  - aerosol extreme events
  - aerosol trends
- an interactive evaluation interface: pyaerocom / AeroVal

Augustin Mortier, PhD

MET Norway, Climate and Air Quality department

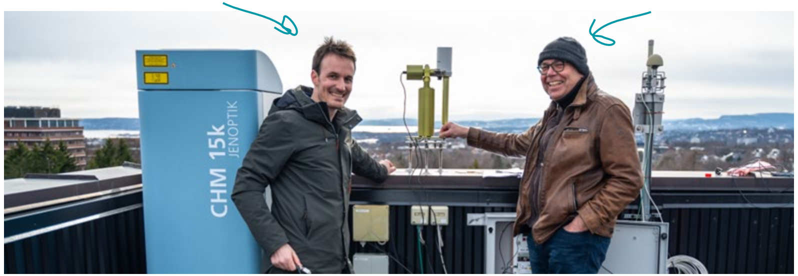
augustinm@met.no

Experience & Activities

- lidars/sunphotometers (PhD thesis)
- aerosol trends, model evaluation
- data visualization & web development
    - ceilometers (vprofiles.met.no)
    - policy support (policy.atmosphere.copernicus.eu)
    - model evaluation (aeroval.met.no)
    - aerosol alerts (aerosol-alerts.atmosphere.copernicus.eu)

me

M. Schulz



Introduction

Introduction

Motivations for model evaluation

- *How does my model perform? (strengths, weaknesses)*
- *How can I improve my model?*

Nobody likes a wrong model prediction

→ *Are there biases in my model's predictions?*

→ *How well does my model capture the variability?*

time/space patterns or trends

→ *How sensitive is my model to parameters?*

- ***Which parameter?***
- ***Where?***
- ***When?***

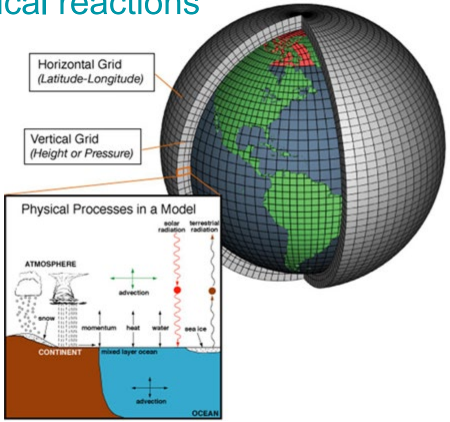
Introduction

Models

"**Air quality models** use mathematical and numerical techniques to simulate the physical and chemical processes that affect **air pollutants** as they disperse and react in the atmosphere."
→ *CHIMERE, EMEP*, LOTOS-EUROS, MOCAGE, …, + ENSEMBLE*

regional, high resolution, good representation of emissions, transport and chemical reactions

"**Climate models** use quantitative methods to simulate the interactions of the important **drivers of climate**, including **atmosphere**, **oceans**, **land surface** and **ice**. They are used for a variety of purposes from study of the dynamics of the climate system to projections of future climate."
→ *Earth System Models: NorESM*, CanESM, CESM, GFDL, EC-Earth (IFS)…*

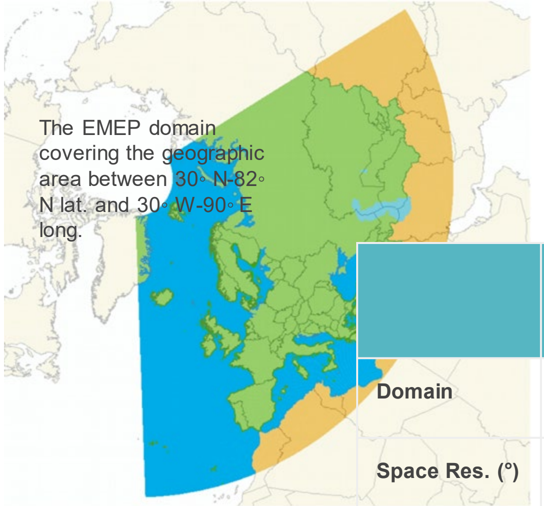


global models, simplified chemistry, but several components

NorESM2 Vs EMEP MSC-W

NorESM2 components		
Atmosphere	CAM6-Nor	2x2 / 1x1, 32 levels (up to 3 hPa, 40 km)
Land	CLM5	2x2 / 1x2, +/- 20 layers
River routing	MOSART	
Ocean	BLOM	1x1, 53 levels
Sea ice	CICE5	1x1, 8+3 levels

CAM = Community Atmosphere Model / CLM = Community Land Model / MOSART = Model for Scale Adaptive River Transport / BLOM = Bergen Layered Ocean Model / CICE = Los Alamos Sea Ice Modela



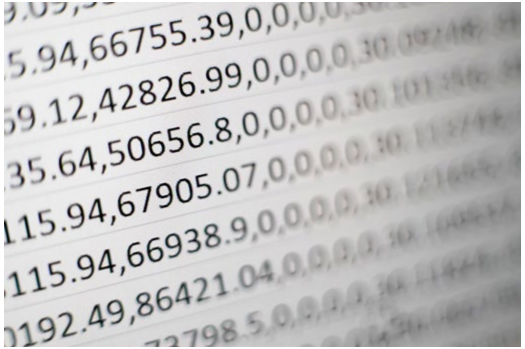
The EMEP domain covering the geographic area between 30◦ N-82◦ N lat. and 30◦ W-90◦ E long.

	NorESM (CAM6-Nor)	EMEP
Domain	Global	Regional (Europe)
Space Res. (°)	2x2 / 1x1	0.1 x 0.1
Levels	32	20 (10 in PBL)
Tracers	≈30	≈200

Introduction

Standard atm. model products: aerosol, gases, meteorological parameters

- 3D: time, lat, lon
- 4D: time, lat, lon, elevation



- Air quality models
  - Concentrations
    - PM_{10}, $PM_{2.5}$
    - Chemical species: O_3, SO_2, SO_4, NO, BC, …
  - Meteorological parameters: T, RH, P, precip., u/v, …
- Climate models
  - Radiative fluxes
  - Optical and microphysical properties: AOD, scat/abs, ext(z), Size Distribution
  - Meteorological parameters: T, RH, P, precip., u/v, …

Model evaluation

→ comparisons

- *model VS ground-based obs.*
- *model VS satellite*

but methods also relevant for

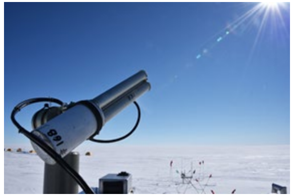
- *satellite VS ground-based obs → "validation"*
- *model v1 VS model v2 → sensitivity tests*

Introduction

Observations

from the **ground**

atmospheric Lidar



photometer

weather station

aethalometer

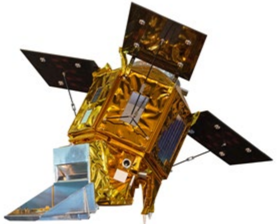
nephelometer

pyranometer

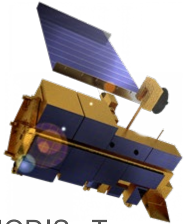
PM monitor

@obs. stations + mobile systems (boats, cars, drones)

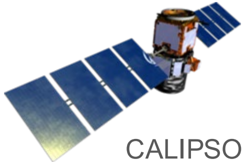
from **space**



Sentinel-5P
NO_2, O_3, …



MODIS - Terra
Spectroradiometer



CALIPSO
Lidar (CALIOP)

Challenges

- By nature
  - **observations** are **not perfect**: continuity, uncertainty, noise, outliers
  - **models** are **not perfect**: simplification of processes, uncertainty, outliers (and bugs)
  - **representativeness** (time, space):
    - deserts, oceans, forests / urban vs rural: local emissions
    - sunphotometer: wintertime?
  - **homogeneity** of measurements: observation networks (calibration, protocols)
  - **wet VS dry**: some in-situ instruments measure dried particles
- Technical
  - **data access:** ACTRIS RI, EBAS, …
  - **data reading**
    - model: NetCDF
    - observations: anything (NetCDF, csv, excel, …)
  - make sure things are **comparable**: units, point of view (attenuated backscatter profile)
  - **colocation**: nearest neighbor, inter/extrapolation, time resampling
  - concentration / deposition → average / accumulate
  - **vertical** coordinates: altitude VS layers / pressure

Introduction

Some standard scores

unitless

Statistics		Range	Perfect Score
R	linear?	[-1,1]	1
R Spearman	monotonic?	[-1,1]	1
NMB (%)	deviation?	[-∞,+∞]	0
MNMB (%)	deviation? ++ for small values	[-200,200]	0
NRMSE (%)	absolute error	[0,200]	0
FGE	absolute error ++ when outliers	[0,2]	0

$$R = \frac{\sum_{i=1}^{n}(o_i - \bar{o})(m_i - \bar{m})}{\sqrt{\sum_{i=1}^{n}(o_i - \bar{o_i})^2}\sqrt{\sum_{i=1}^{n}(m_i - \bar{m_i})^2}}$$

$$\rho = 1 - \frac{6\sum_{i=1}^{n}(m_i - o_i)^2}{n(n^2-1)}$$

$$NMB = \frac{\sum_{i=1}^{n}(m_i - o_i)}{\sum_{i=1}^{n} o_i}$$

$$MNMB = \frac{2}{n}\sum_{i=1}^{n}\left(\frac{m_i - o_i}{m_i + o_i}\right)$$

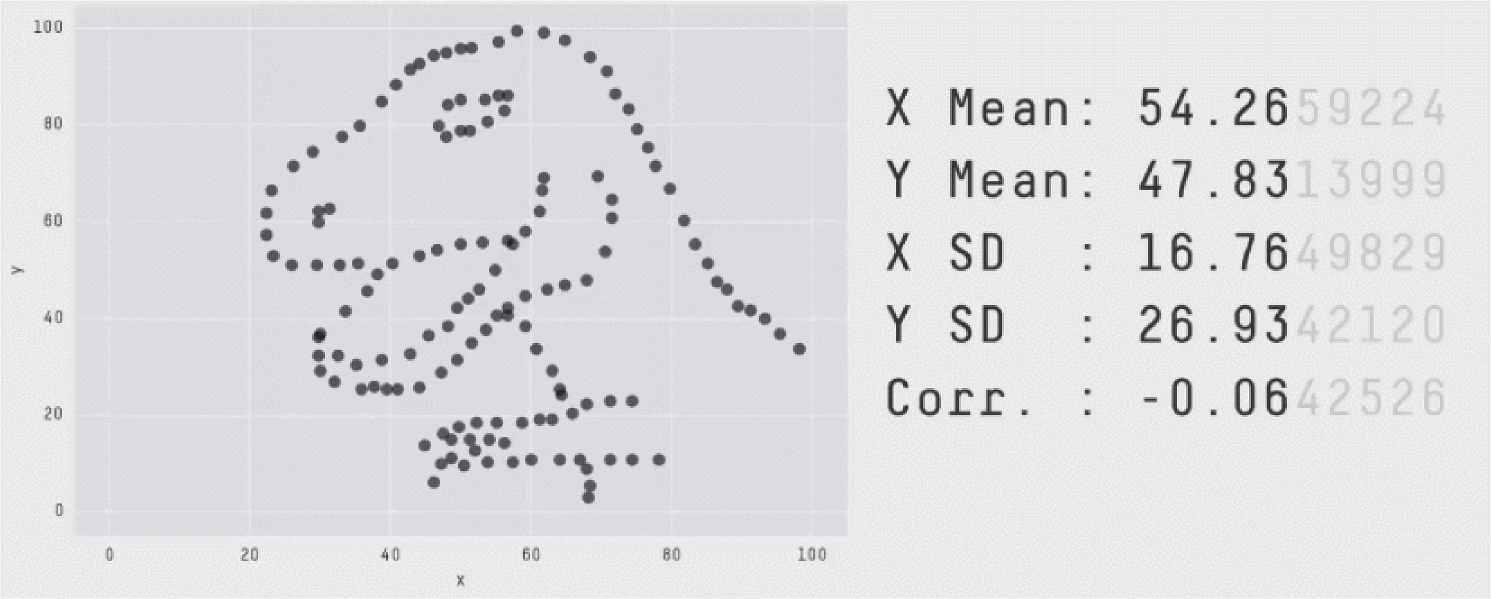
$$NRMSE = \frac{\sqrt{\frac{1}{n}\sum_{i=1}^{n}(o_i - m_i)^2}}{\bar{o}}$$

$$FGE = \frac{2}{n}\sum_{i=1}^{n}\left|\frac{m_i - o_i}{m_i + o_i}\right|$$

m: model, o: observation, n: number of points

Introduction

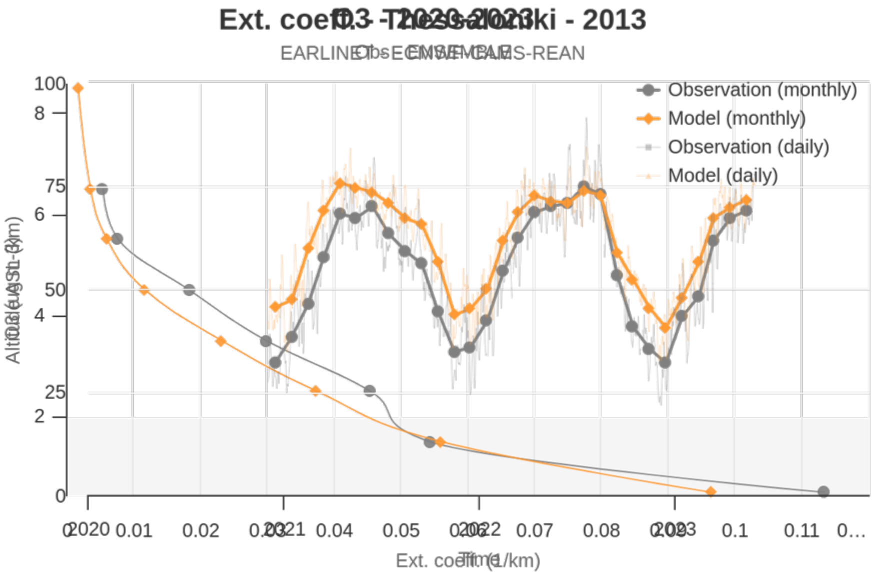
Always plot your data



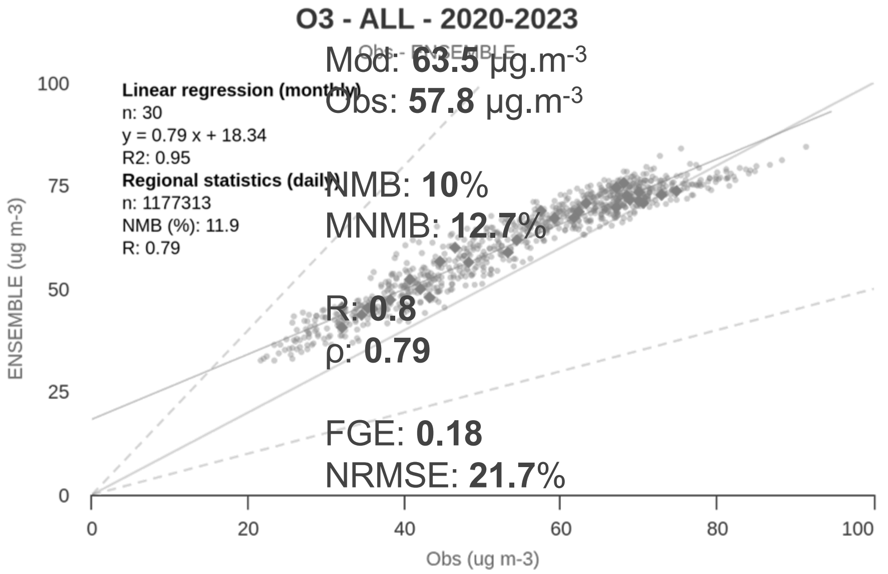
The Datasaurus dozen

Some charts

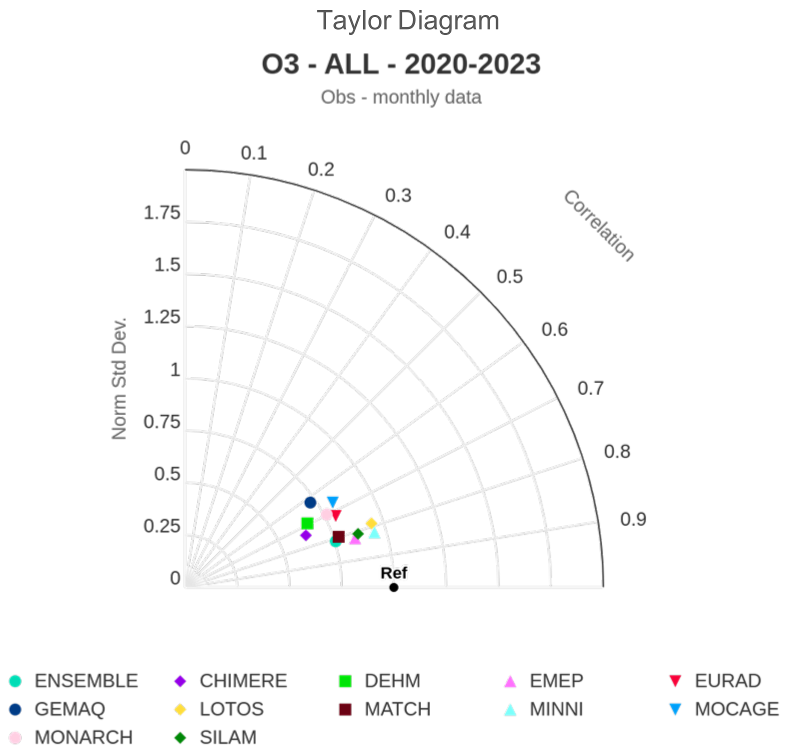
Time Series / plotting things on the same axis



Scatterplot



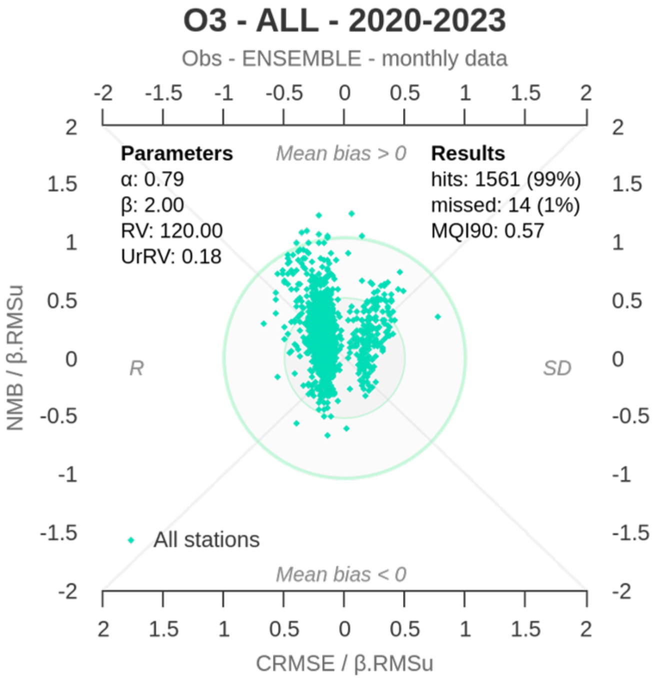
Some charts



Taylor Diagram
O3 - ALL - 2020-2023
Obs - monthly data
Correlation
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
Norm Std Dev.
1.75
1.5
1.25
1
0.75
0.5
0.25
0
Ref
ENSEMBLE
CHIMERE
DEHM
EMEP
EURAD
GEMAQ
LOTOS
MATCH
MINNI
MOCAGE
MONARCH
SILAM

Some charts

FAIRMODE assessment target plot



- RMS_u: measurement uncertainty
- CRMSE: Centered Root Mean Squared Error

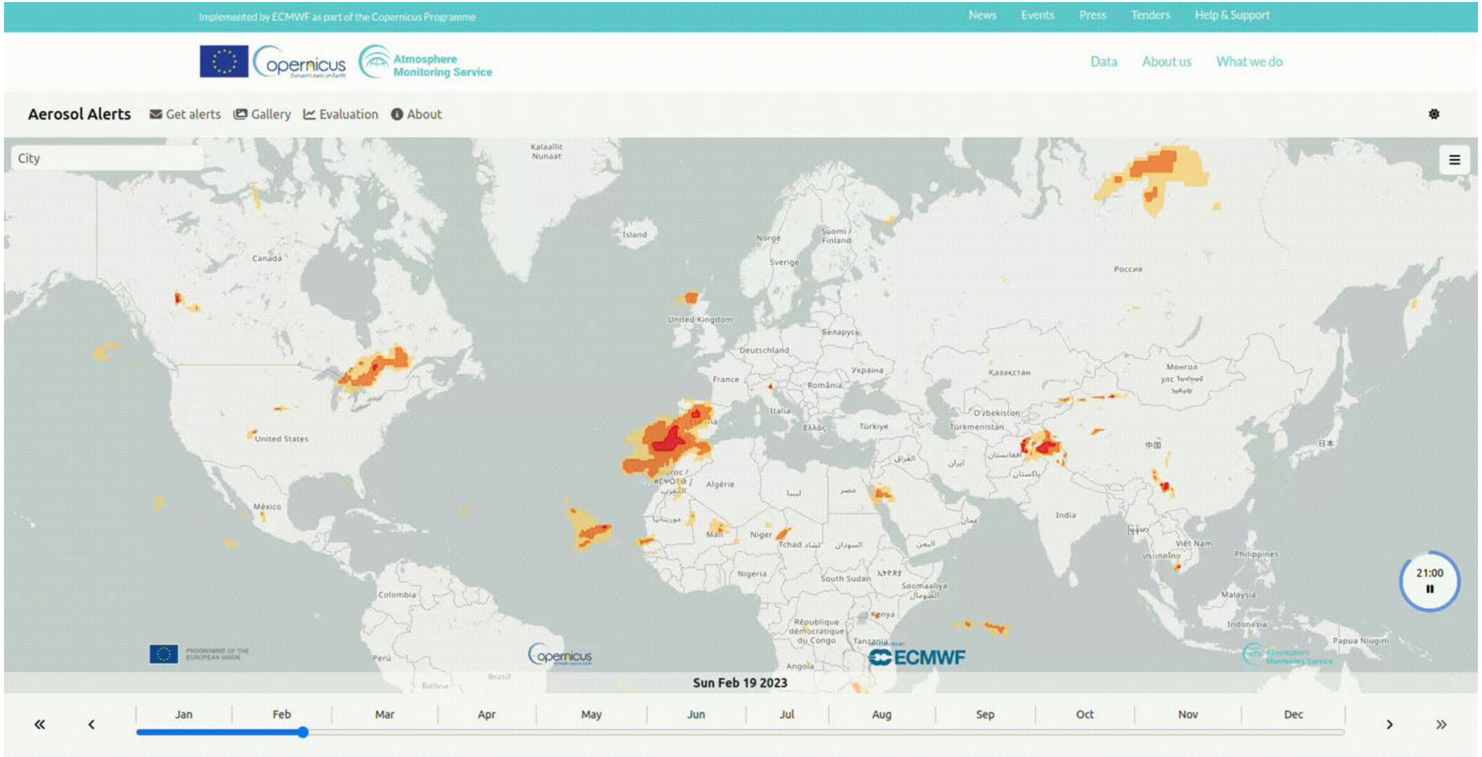
$$CRMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}[(M_i - \bar{M}) - (O_i - \bar{O})]^2}$$

Extreme Aerosol Events

CAMS aerosol alerts service

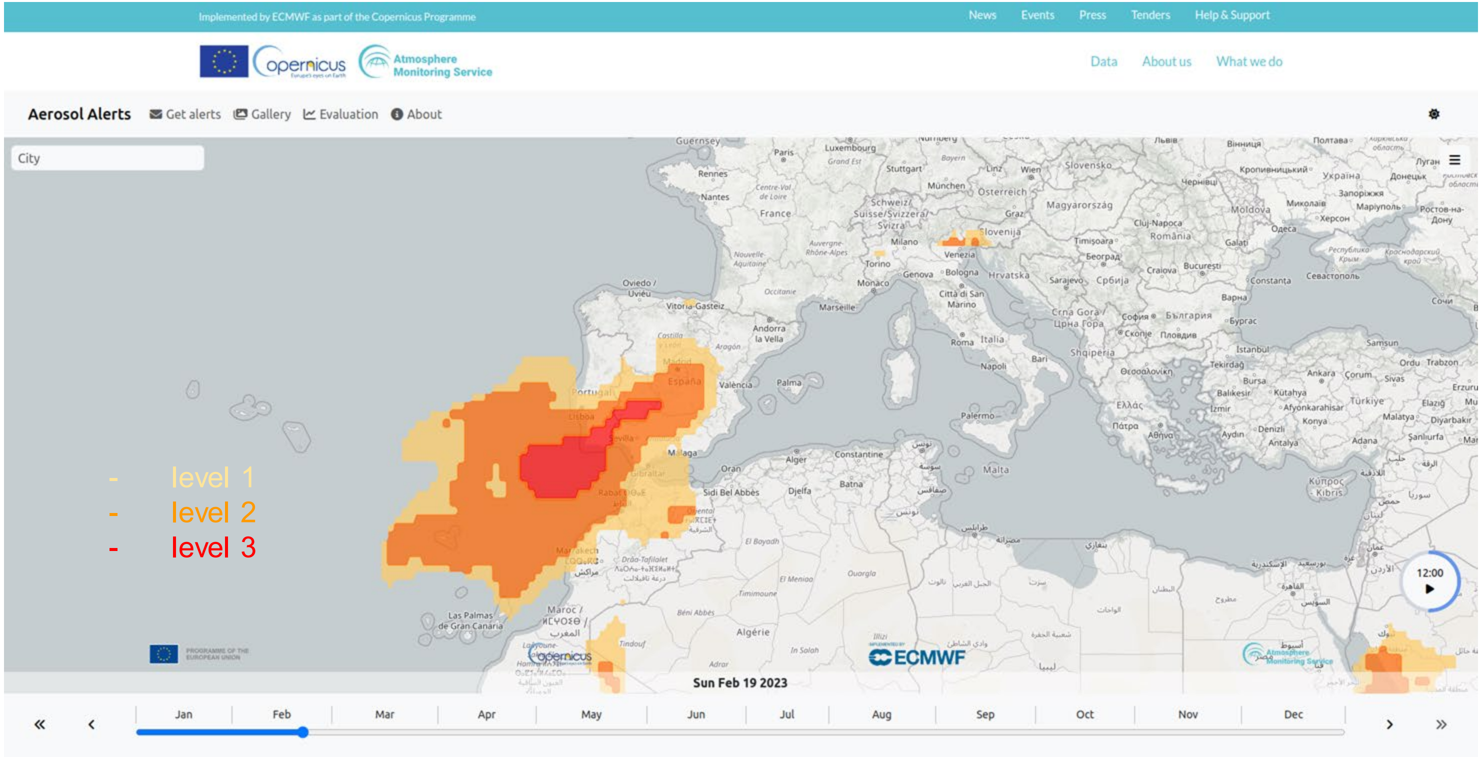
Evaluation of extreme aerosol events

CAMS Aerosol Alerts Service



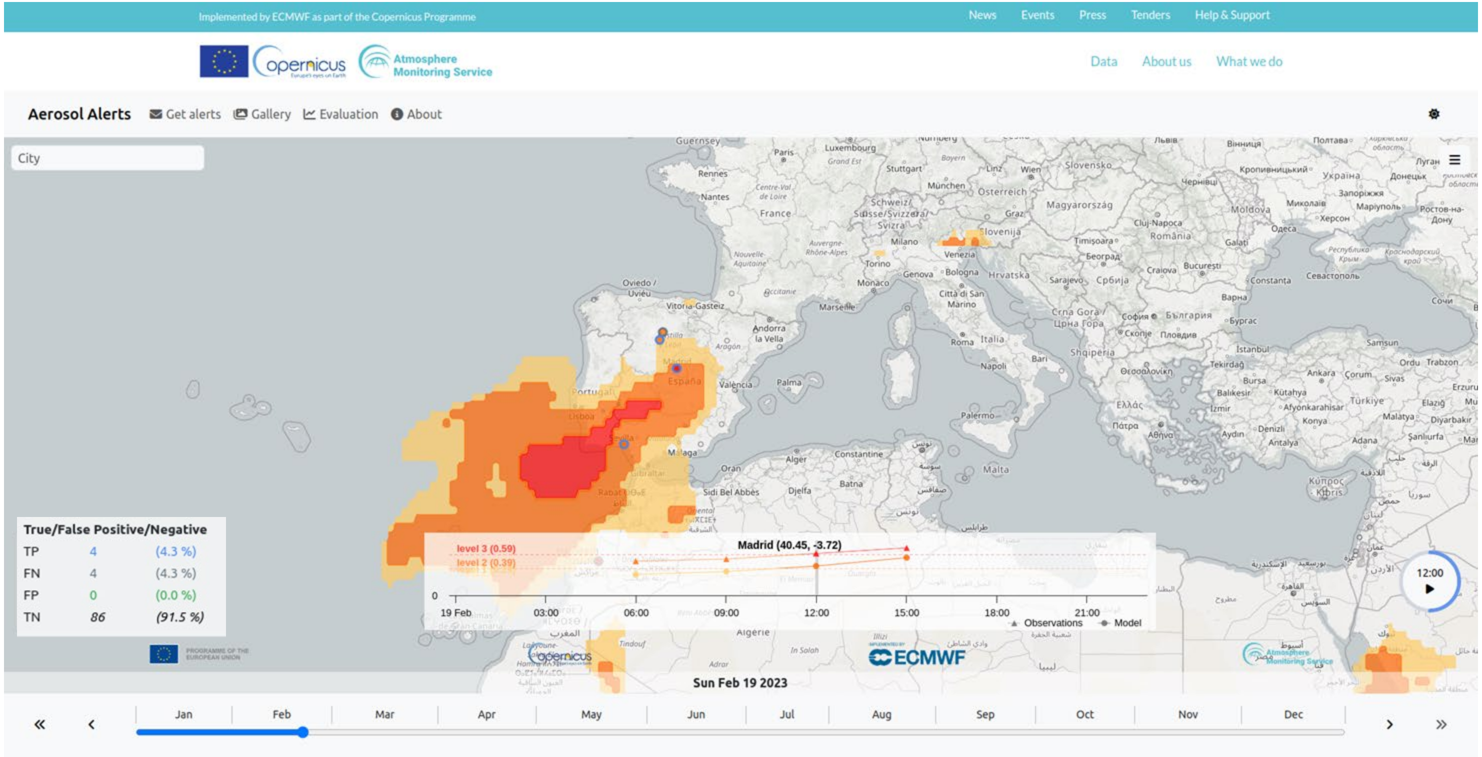
Evaluation of extreme aerosol events

CAMS Aerosol Alerts Service



Evaluation of extreme aerosol events

Does the model get it or not?



Evaluation of extreme aerosol events

Does the model get it or not?

Forecasts Evaluation

This online evaluation, based on colocated observations, allows to explore the model performances in forecasting the aerosol alerts, by region and season.

Total Aerosols (AOD) - IFS-OSUITE - D+1 - NAMERICA - 2023 (AMJ)

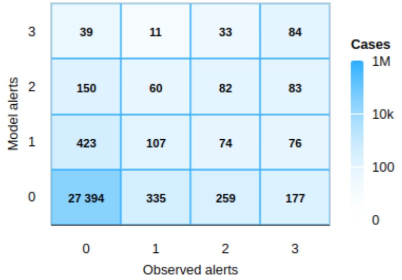
Stations in selected region

AERONET_V3_Level15

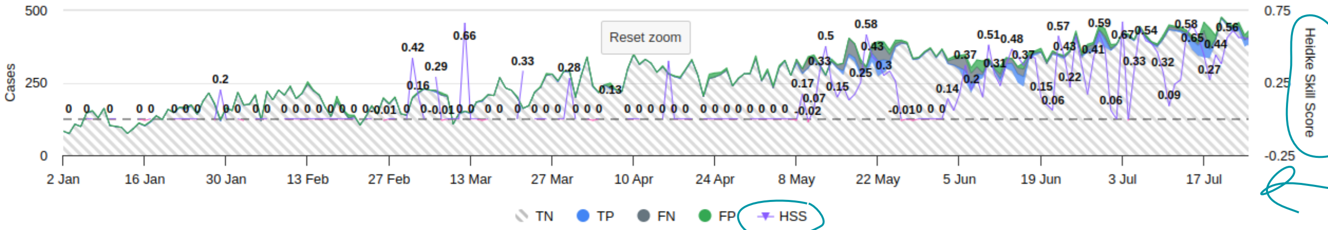
Hit/Miss ratios

	Absolute	Relative
Hit	28004	95.3 %
Miss	1383	4.7 %
TN	27394	93.2 %
TP	610	2.1 %
FN	771	2.6 %
FP	612	2.1 %

Contingency table



Skill scores time series



T/F P/N: True/False Positive/Negative; HSS: Heidke Skill Score

Aerosol Trends Evaluation

Mortier et al., 2020

Aerosol Trends Evaluation

Study Period

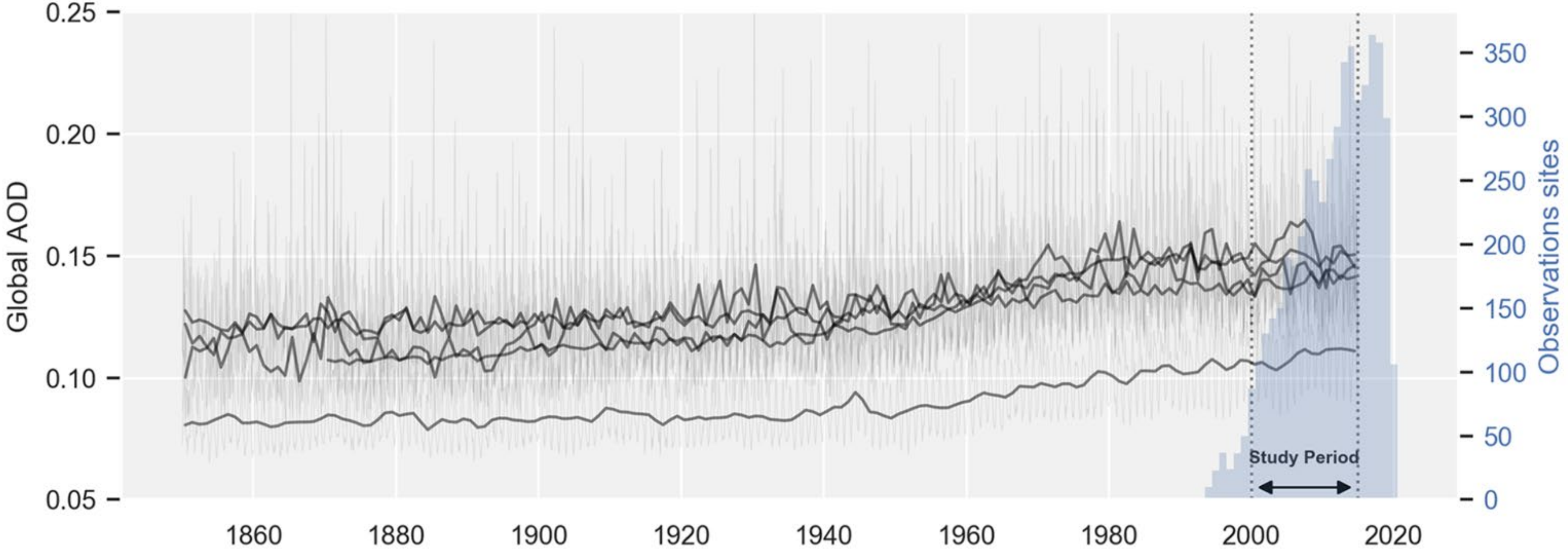


Figure 1: Global AOD computed from model historical runs (Oslo CTM3, GFDL-AM4, CanESM5, CESM2, IPSL-CM6A, ECHAM-HAM) at monthly (gray lines) and yearly (black lines) resolutions, overlaid with the number of active observation sites in the AERONET sun photometer network.

Aerosol Trends Evaluation

Datasets

Observations

Parameter	Type	Observation networks	Models
AOD	Column	AERONET[1]	ECMWF-Rean, NorESM2, SPRINTARS, ECHAM-HAM, GEOS, Oslo CTM3, GFDL-AM4, BCC-CUACE, CanESM5, CESM2, IPSL-CM6A
AOD_f	Column	AERONET	NorESM2, SPRINTARS, ECHAM-HAM, GEOS, Oslo CTM3, GFDL-AM4
AOD_c	Column	AERONET	ECMWF-Rean, NorESM2, SPRINTARS, ECHAM-HAM, Oslo CTM3, GFDL-AM4
AE	Column	AERONET	ECMWF-Rean, NorESM2, SPRINTARS, ECHAM-HAM, GEOS, Oslo CTM3, GFDL-AM4
$PM_{2.5}$	Surface	EMEP[2], IMPROVE[3]	ECMWF-Rean, NorESM2, SPRINTARS, ECHAM-HAM, GEOS
PM_{10}	Surface	EMEP, IMPROVE	ECMWF-Rean, NorESM2, SPRINTARS, ECHAM-HAM, GEOS
SO_4	Surface	EMEP, IMPROVE, CASTNET[4], CAPMoN[5], EANET[6]	ECMWF-Rean, NorESM2, SPRINTARS, ECHAM-HAM, GEOS, Oslo CTM3, BCC-CUACE
σ_{sp}	Surface	GAW-WDCA[7] (incl. IMPROVE, NOAA FAN[8], ACTRIS[9], EMEP)	NorESM2, Oslo CTM3
σ_{ap}	Surface	GAW-WDCA (incl. NOAA FAN, ACTRIS, EMEP)	NorESM2, Oslo CTM3

Table 1: List of observations and model datasets used in this study (see text for explanation).

Models

Model	Group	Natural interactive emissions	Anthropogenic emissions	Meteorology	Resolution (degrees)	References
ECMWF-Rean	CAMS-Rean	D, SS	MACCity	RA	0.7 × 0.7	Inness et al. (2019); Zhang et al. (2009)
SPRINTARS	AP3	D, SS, DMS, Oce VOCs	C: SO_2, BC, OC	N	0.56 × 0.56	Takemura et al. (2000, 2002, 2005)
ECHAM-HAM	AP3	D, SS, DMS	C: SO_2, BC, OC	fSST	1.875 × 1.875	Tegen et al. (2019); Neubauer et al. (2019)
GEOS	AP3	D, SS, DMS, Oce VOCs	O: SO_2, SO_4, BC, OC, NH_3	N	1.00 × 1.00	Bian et al. (2017); Chin et al. (2002); Colarco et al. (2010)
Oslo CTM3	AP3	D, SS	C: SO_2, SO_4, BC, OC, NH_3	S	2.25 × 2.25	Lund et al. (2018); Myhre et al. (2009)
GFDL-AM4	AP3	D, SS, DMS, Oce & Veg OC	C: SO_2, SO_4, BC, OC	fSST&N	1. × 1.25	Zhao et al. (2018a,b)
BCC-CUACE	AP3	D, SS, DMS	C: SO_2, BC, OC	F	2.8 × 2.8	Zhang et al. (2012, 2014); Wang et al. (2014)
NorESM2	CMIP6	D, SS, DMS, MSA, VOCs	C: SO_2, SO_4, OC, BC	F	1.89 × 2.50	Seland et al. (2020); Kirkevåg et al. (2018)
CanESM5	CMIP6	D, SS, DMS	C: SO_2, OC, BC	F	2.77 × 2.81	Swart et al. (2019)
CESM2	CMIP6	D, SS, DMS_{clim}	C: SO_2, OC, BC	F	0.94 × 1.25	Danabasoglu et al. (submitted); Tilmes et al. (2019)
IPSL-CM6A	CMIP6	D, SS, DMS_{clim}	*C: SO_2, BC, OC, NH_3	fSST	2.50 × 1.27	Lurton et al. (2019)

Table 2: Information on models used in this study (CAMS-Rean: CAMS reanalysis, AP3: AeroCom phase III, CMIP6: historical experiments from CMIP6).

Aerosol Trends Evaluation

Region definition & Observation coverage

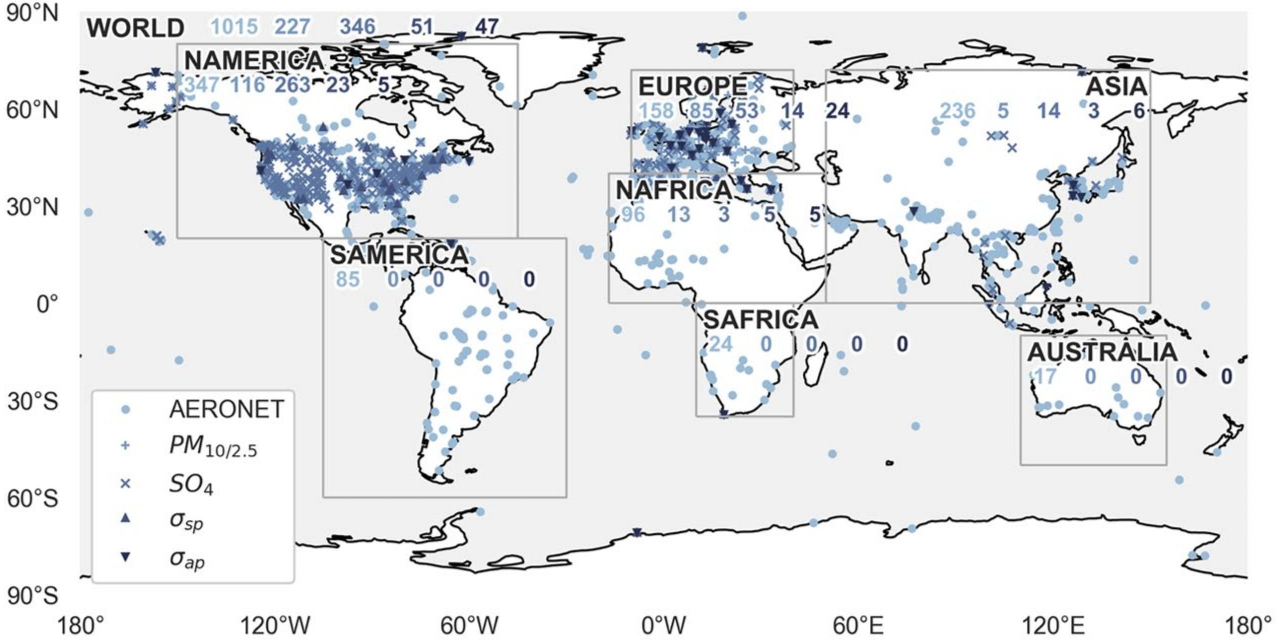


Figure 2: Distribution of the observations within the different regions considered in this study. The numbers reported within each region correspond to the maximum number of stations given for the observation networks, corresponding to the five observation types found in the legend.

Aerosol Trends Evaluation

Regional Time Series

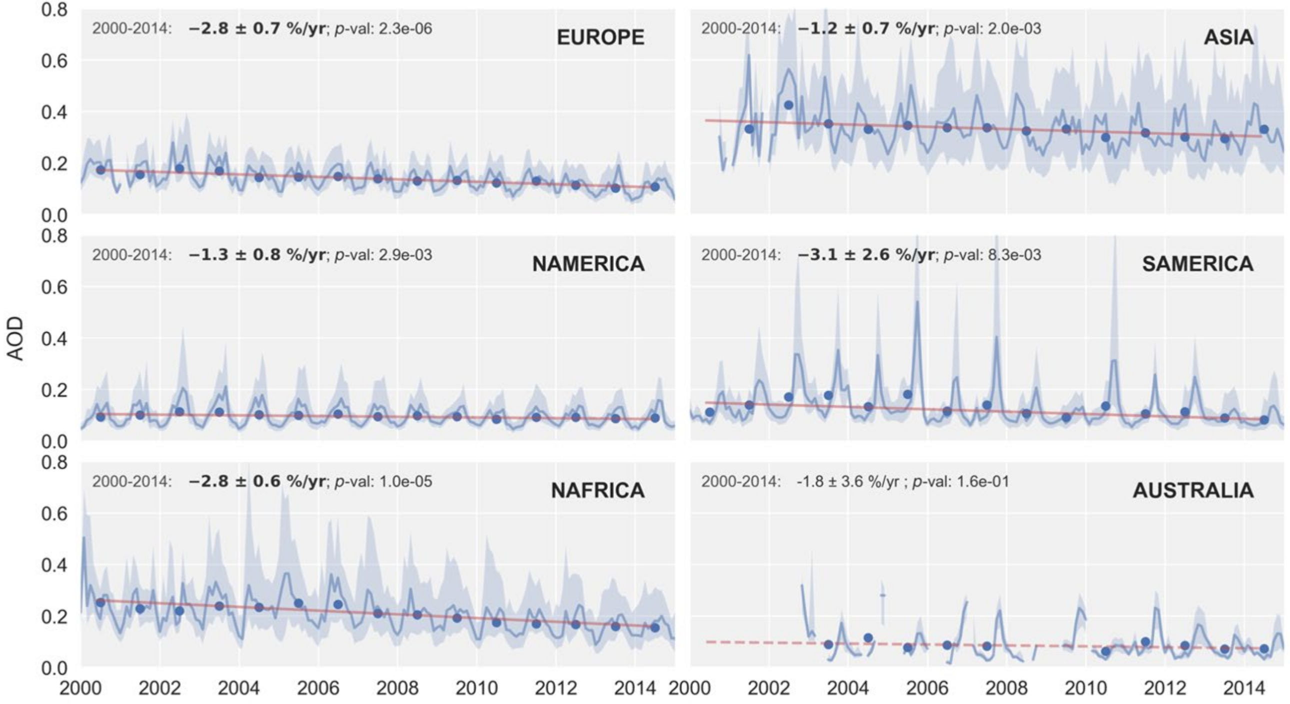


Figure 3: Regional time series of AOD. The dark blue line corresponds to the median, and the light blue envelope is bound by the first and third quartiles of all valid points at the corresponding month, respectively. The blue dots correspond to the yearly averages which are used to compute the linear trend. The latter is displayed as a continuous line when the trend is significant and as a dashed line when it is not. Trend values, an error estimate and a significance value are given in each panel.

Aerosol Trends Evaluation

Trends calculations

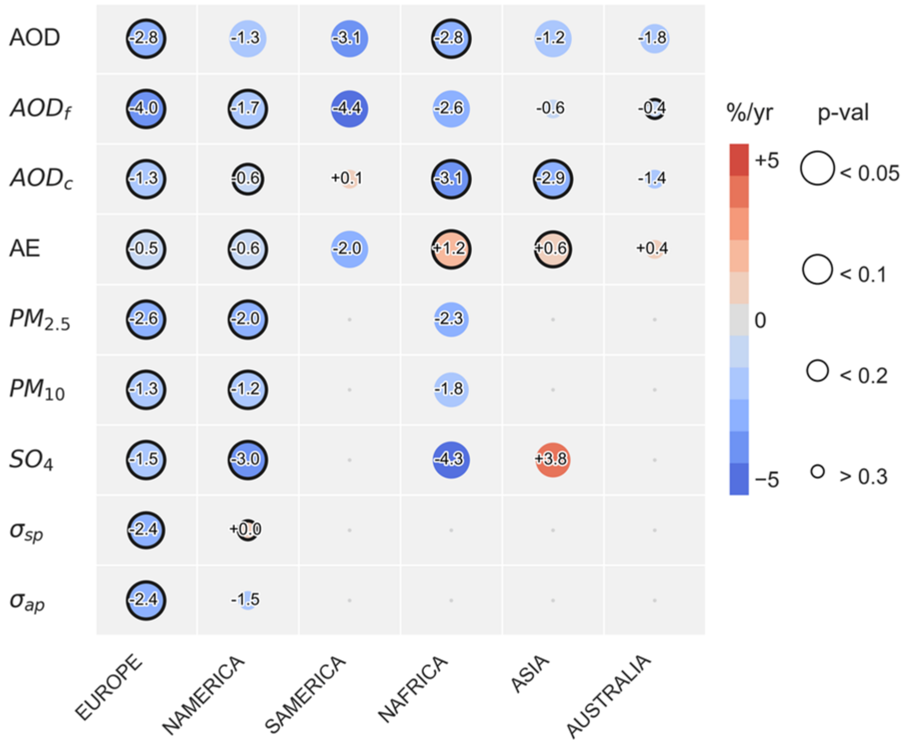


Figure 5: Regional trends in the aerosol properties computed with the observation datasets. The color of the circles corresponds to the slope, while the radius indicates the p value. The largest circles represent the trends that are significant with a confidence level of 95 %. The circles outlined in black indicate the trends associated with a representativity greater than 50 %.

Aerosol Trends Evaluation

Trends evaluation

Figure 6: Regional trends in the aerosol properties computed with observations and models co-located in space and time with the observations. The error bars correspond to the uncertainty in the trend as calculated using both the uncertainty in the Theil–Sen slope and the residuals. The bold font indicates that the trends are significant at a confidence level of 95 % (p value < 0.05).

Aerosol Trends Evaluation

Trends evaluation

Aerosol Trends Evaluation

Representativeness of observed trends → model subset selection

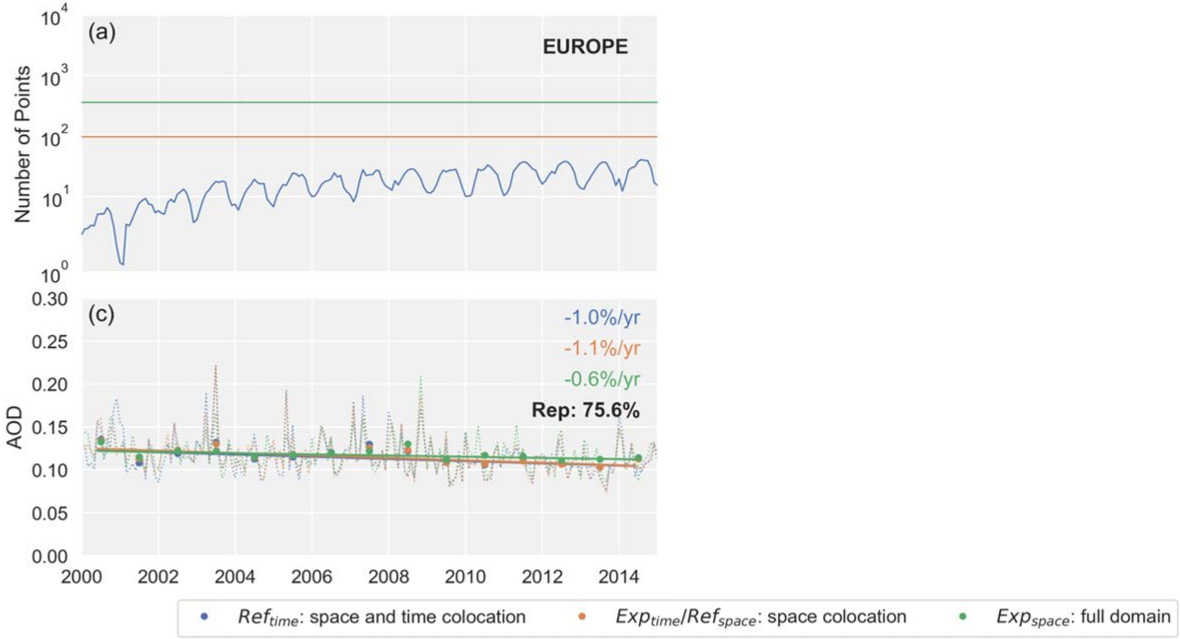


Figure 4: Three regional AOD time series and respective trends, constructed from model data (NorESM2) for the investigation of the representativity of the observational data. Panels (a) and (b) correspond to the number of points used to compute the regional time series for the three different datasets. Panels (c) and (d) show the time series, the trends and the resulting representativity value (black, bold). The blue color (Reftime) corresponds to the model output co-located in space and time with the available observations. Panels (a) and (b) show an overall increase in the number of available observations (more stations) combined with a seasonal cycle (less AOD available in wintertime). The orange color (Exptime⁄Refspace) corresponds to the model output co-located in space with the stations providing measurements, using the complete time series from 2000 to 2014. The green color (Expspace) corresponds to the model output in the whole geographic region (see Fig. 2), using all of the grid boxes without any co-location with the observations.

Aerosol Trends Evaluation

Modelled trends

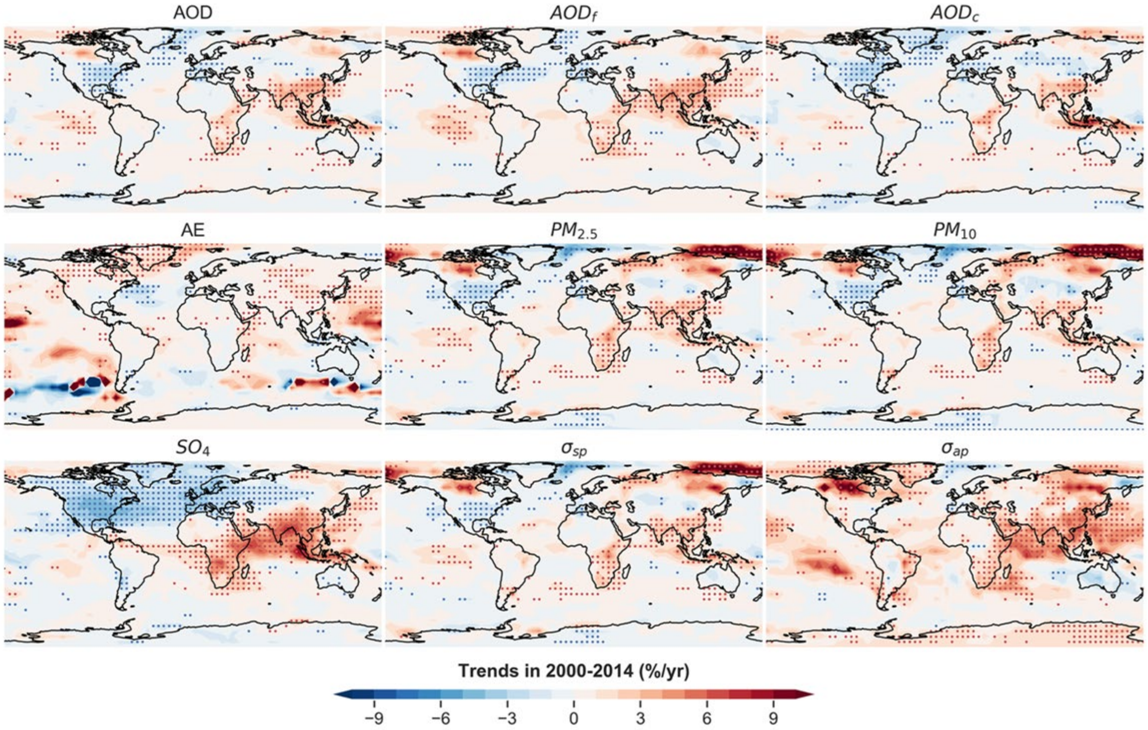


Figure 7: Global trends in aerosol properties using NorESM2 data regridded at a 5∘ × 5∘ resolution. The blue and red dots indicate significant negative and positive trends, respectively.

Aerosol Trends Evaluation

Modelled trends

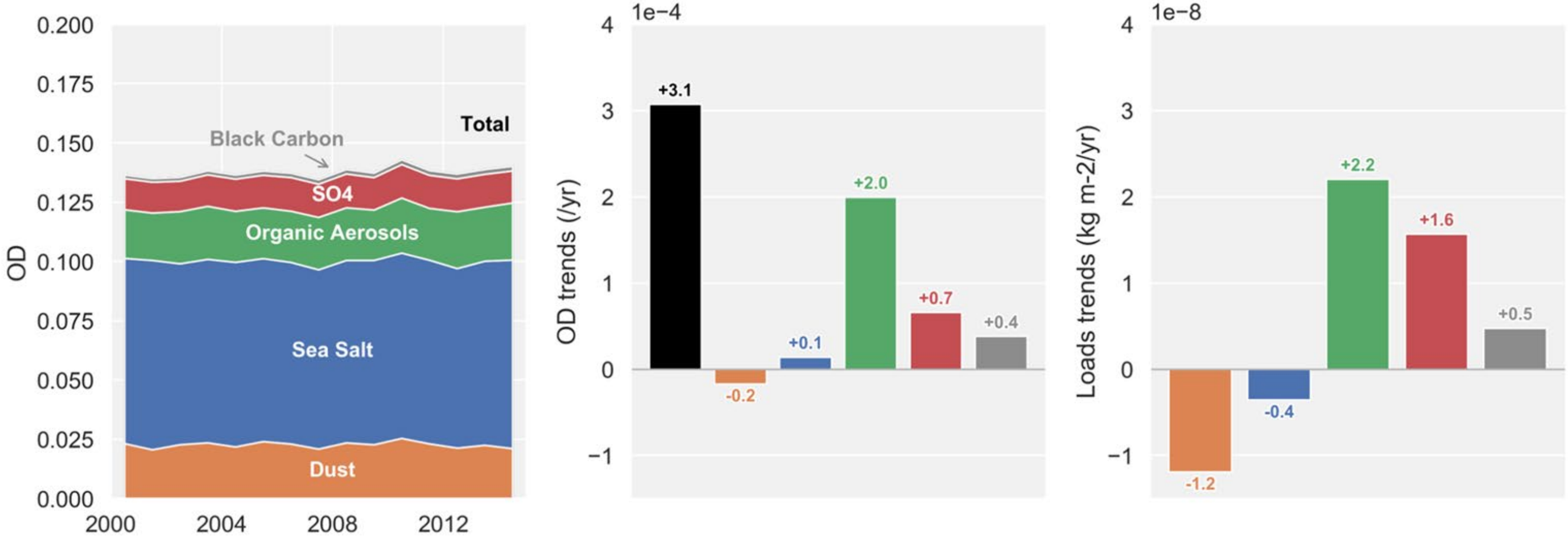


Figure 8: Absolute trends in OD and emissions of the main aerosol species computed with NorESM2. The y axis of the trends in OD and the emissions is given according to the power of 10 indicated at the top left corner of each of the subplots.

pyaerocom / AeroVal

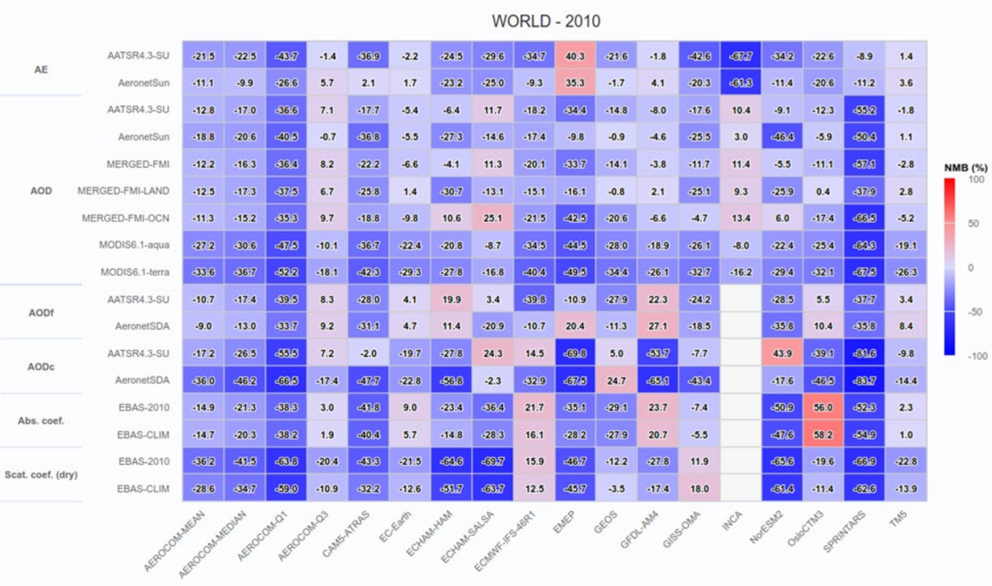
an interactive evaluation interface

pyaerocom / AeroVal

Motivations

~2 years ago - Gliß et al., 2021: AeroCom phase III **multi-model** evaluation of the aerosol life cycle and optical properties using **ground- and space-based remote sensing** as well as **surface in situ observations** (*https://acp.copernicus.org/articles/21/87/2021*)

- *models: 13 + ENSEMBLE (Mean/Median)*
- *observations:*
  - *remote sensing: AERONET, satellites (MODIS, AATSR, Merged-FMI dataset)*
  - *in situ: EBAS (Scat./Abs. Coef.)*



AeroVal overall evaluation heatmap, adapted from Gliß et al., 2021

How to interpret all of this?

→ exploratory/evaluation tool

About

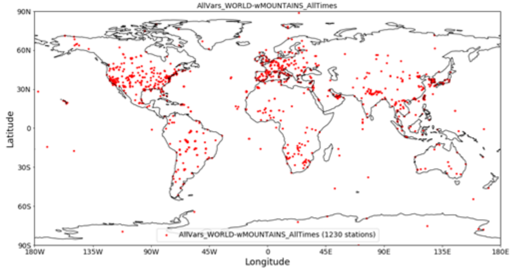
"*pyaerocom*: python library (≥3.6) for **processing** and **plotting** of data [related to the **AeroCom** project].

This includes support for reading and processing of **model data** (e.g. AeroCom, EMEP, …), **satellite data** (e.g. MODIS, AATSR, …) and **ground based observation** datasets (e.g. AERONET, EBAS, EARLINET, Ghost, AirNow, MarcoPolo, EEA, …).

In addition, pyaerocom provides tools for **colocation** and **cross evaluation** of different datasets using **commonly used statistical metrics** such as several biases, gross-errors, or correlation coefficients."

https://github.com/metno/pyaerocom

```
import pyaerocom as pya
obs_reader = pya.io.ReadUngridded('AeronetSunV3Lev2.daily')
od550aer = obs_reader.read(vars_to_retrieve='od550aer')
# plot station coordinates
od550aer.plot_station_coordinates();
```



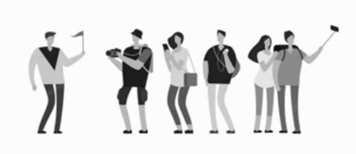
pyaerocom
Data reading (obs, mod) + colocation
→ **colocated files**

pyaeroval
Read pyaerocom colocated files + computation of statistics +
→ **JSON files**

AeroVal
aeroval.met.no
visualization of **JSON files**

used in

- CAMS2_83 (regional mo
- CAMS2_82 (global mode



Conclusion

Conclusion

Some take home messages

- Important to evaluate model (in the present/past) to determine **forecasts reliability**
- Model evaluation highlights strengths and weaknesses of models → **model improvement**
- Challenges
  - Measurements **representativeness**
  - **Colocation** (time resampling, space)
- **Scores**: R, NMB, RMSE, …
- **Charts**: visualization of data / scores

Look at you data from different angles