living planet symposium | BONN 23–27 May 2022

TAKING THE PULSE OF OUR PLANET FROM SPACE

Machine learning-based constraint of the aerosol optical depth maps simulated by the CHIMERE chemistry-transport model

Farouk Lemmouchi[1], Juan Cuesta[1], Mathieu Lachatre[2], Julien Brajard[3], Adriana Coman[1], Claude derognat[2], and Matthias Beekmann[1]

[1] LISA, UMR 7583 CNRS, Université Paris-Est Créteil, France

[2] ARIA Technologies, Boulogne-Billancourt, France

[3] Nansen Environmental and Remote Sensing Center (NERSC), Bergen, Norway

2022-05-27

Aerosol optical depth (AOD) estimation

- Aerosol are suspended particles in the air like salts, dust, soots, etc with a radius size usually < 10µm. They are **harmful** for health and they can be transported several thousands of km.
- The abundance of aerosols vertically integrated in the atmosphere can be estimated by the **Aerosol Optical Depth** (AOD).
- We can estimate the AOD by **remote aerosol retrieval** algorithms e.g. Dark Target/Deep Blue, or by **chemistry-transport modeling (CTM)** e.g. CHIMERE.

Aerosol optical depth (AOD) estimation

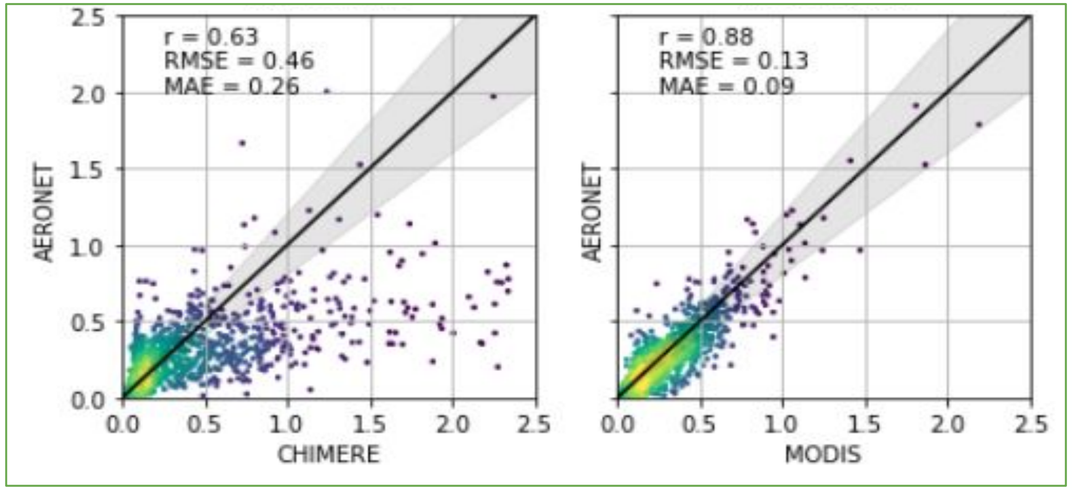
- Aerosol are suspended particles in the air like salts, dust, soots, etc with a radius size usually < 10µm. They are **harmful** for health and they can be transported several thousands of km.
- The abundance of aerosols vertically integrated in the atmosphere can be estimated by the **Aerosol Optical Depth** (AOD).
- We can estimate the AOD by **remote aerosol retrieval** algorithms e.g. Dark Target/Deep Blue, or by **chemistry-transport modeling (CTM)** e.g. CHIMERE.

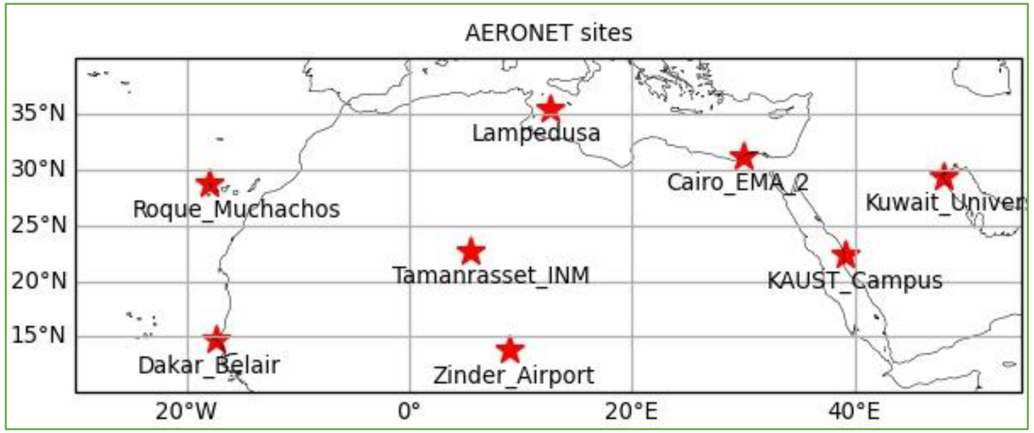
	CHIMERE AOD	MODIS AOD
Temporal resolution	1h	24h
spatial coverage	Full coverage	Cloud-free only
Accuracy	good	Better

- **CHIMERE** AOD is less accurate in the regions lacking *in situ* monitoring stations for constraining the simulation e.g **Africa**.

Aerosol optical depth (AOD) estimation

Colocated AOD 550 nm of 12 months (2021)

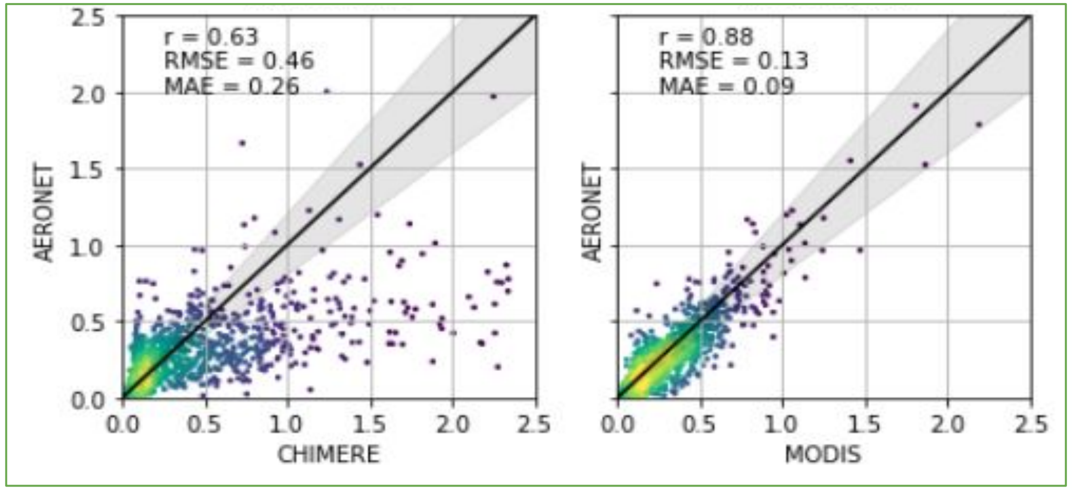


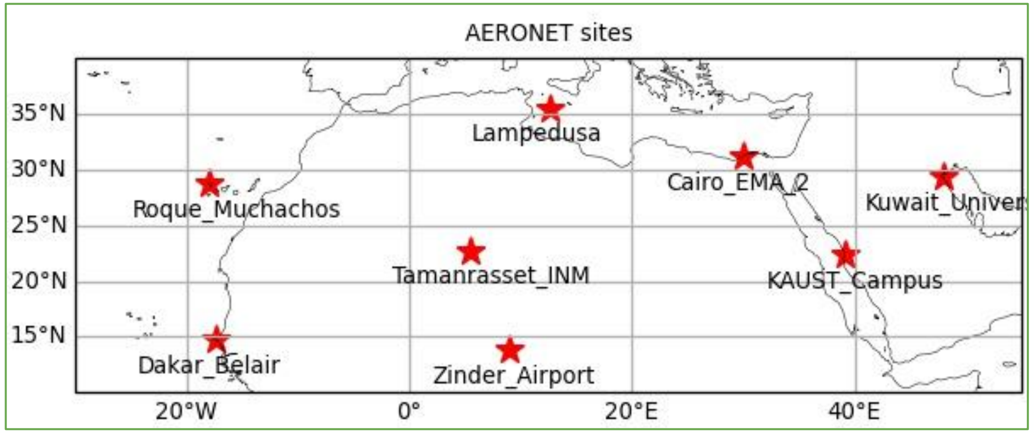


MODIS AOD compares better than CHIMERE against the 8 AERONET stations measurements.

Aerosol optical depth (AOD) estimation

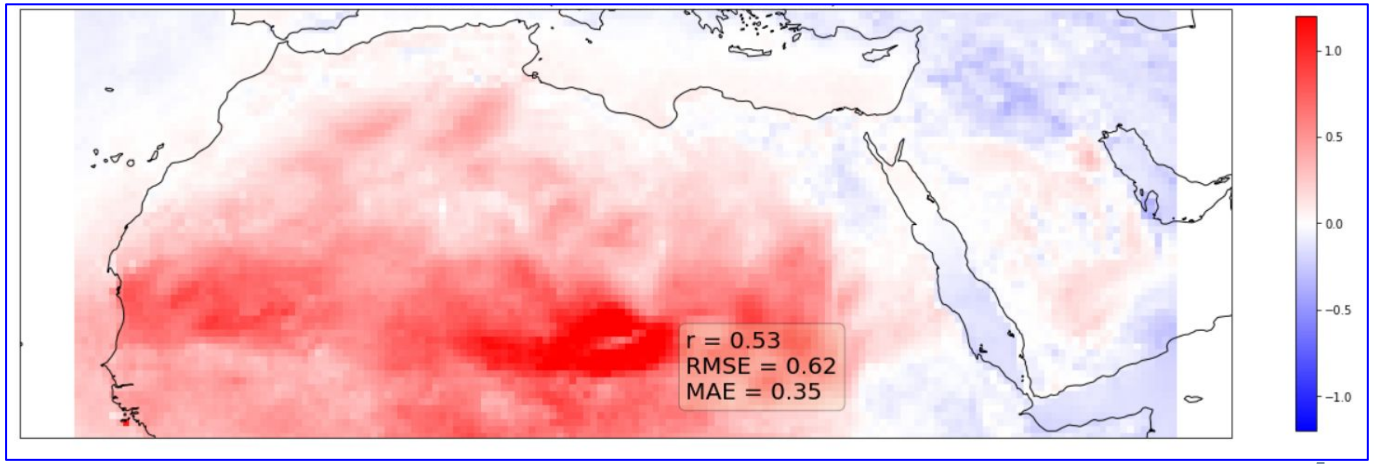
Colocated AOD 550 nm of 12 months (2021)





MODIS AOD compares better than CHIMERE against the 8 AERONET stations measurements.

AOD year 2021 average bias CHIMERE-MODIS AQUA



CHIMERE overestimate AOD in the deserts and underestimate it elsewhere.

Principle of AOD simulation post-processor

One year (2021) of data over north Africa with 0.45° horizontal resolution.

CHIMERE AOD + other variables → Trained AI Model → AOD_{550nm} corrected

Inputs **Output**

Inputs

- CHIMERE simulation at 1:00 PM : AOD_{550} ,*PM_{10}, PM_{25}, pDUST, pOCAR, pWATER, pSALT, ROOH, HCNM, $pH2SO_4$, $pHNO_3$, pNH_3, sphu, O_3, SO_2, CO, OH, NH_3, TOL, NO_X, NO_Y*
- ERA5 at 1:00 PM: *pres, relh, soil moisture, boundary layer height, albedo*

Reference

MODIS/aqua $AOD_{550\ nm}$

- Random forest
- Neural networks
- Multiple Linear Regression

The **goal** is to improve the AOD estimation of CHIMERE chemistry-transport model using MODIS AOD observations.

Principle of AOD simulation post-processor

One year (2021) of data over north Africa with 0.45° horizontal resolution.

Training + testing dataset :

The first 20 days of each month = 240 days

=> Table of ~1.4M simulation ground pixel and 97 features

Validation data :

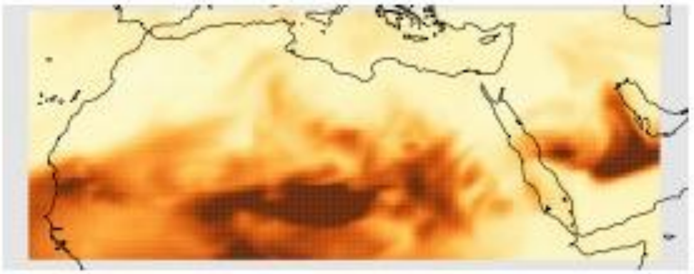
The rest of the days => 122 days

- Random forest
- Neural networks
- Multiple Linear Regression

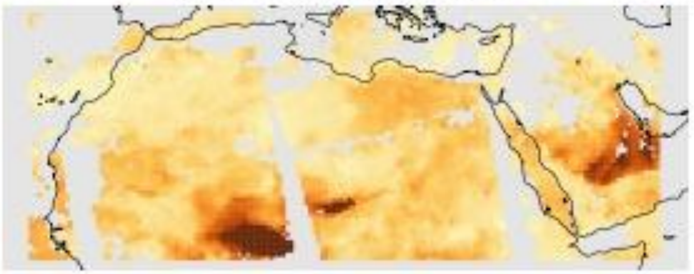
The **goal** is to improve the AOD estimation of CHIMERE chemistry-transport model using MODIS AOD observations.

Results of AOD simulation post-processor

Corrected AOD : case of 21 Jan 2021

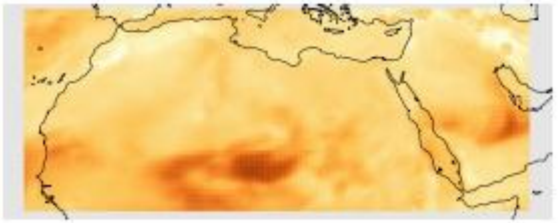


CHIMERE $AOD_{550\ nm}$

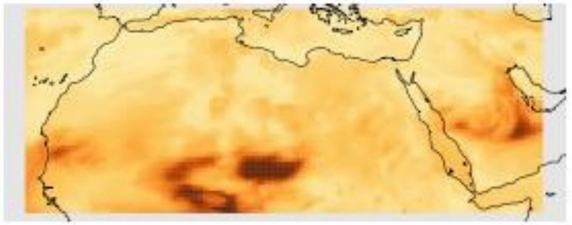


MODIS AQUA $AOD_{550\ nm}$

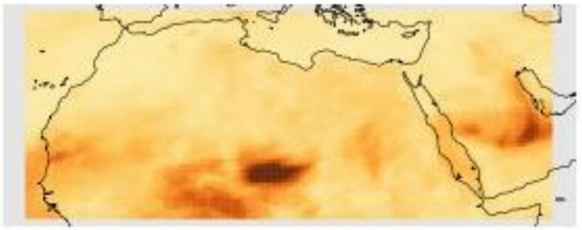
CHIMERE AOD_{550nm} corrected



Multiple linear model



Neural network

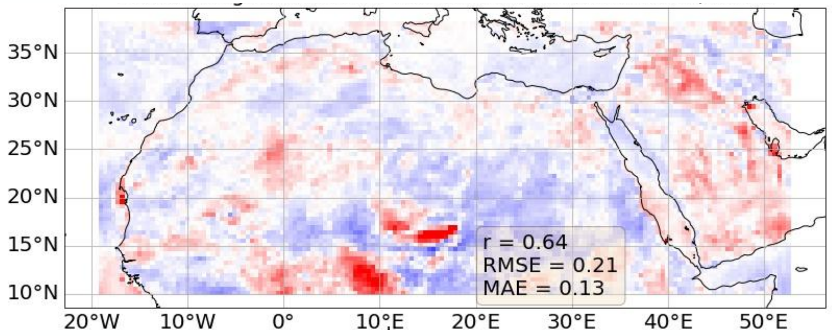


Random forests

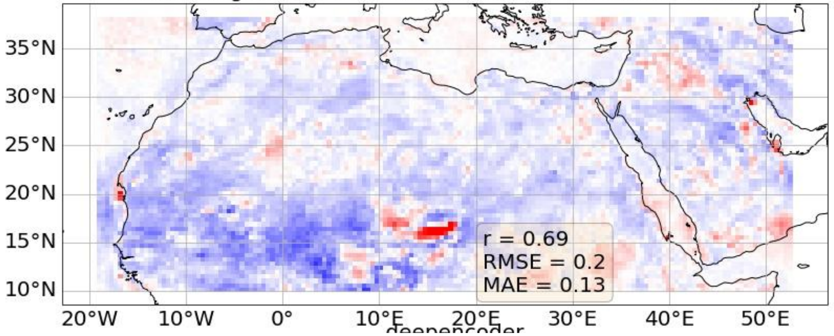
Good overall performance.

Results of AOD simulation post-processor

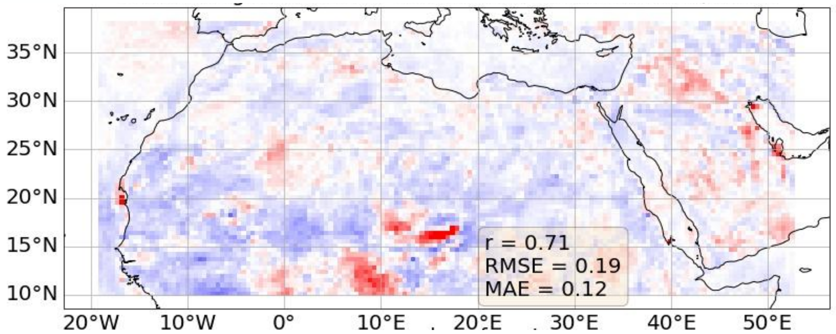
Multiple linear regression

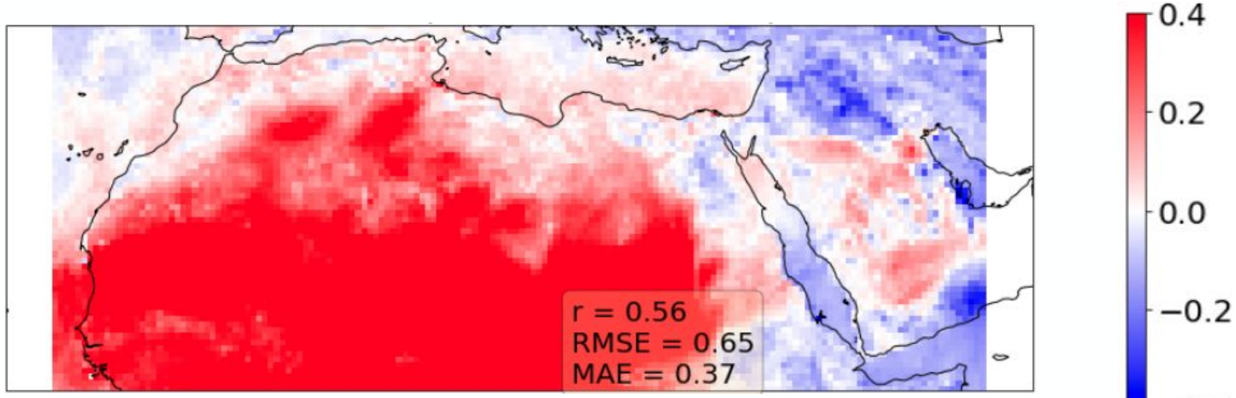


Neural network



Random forests



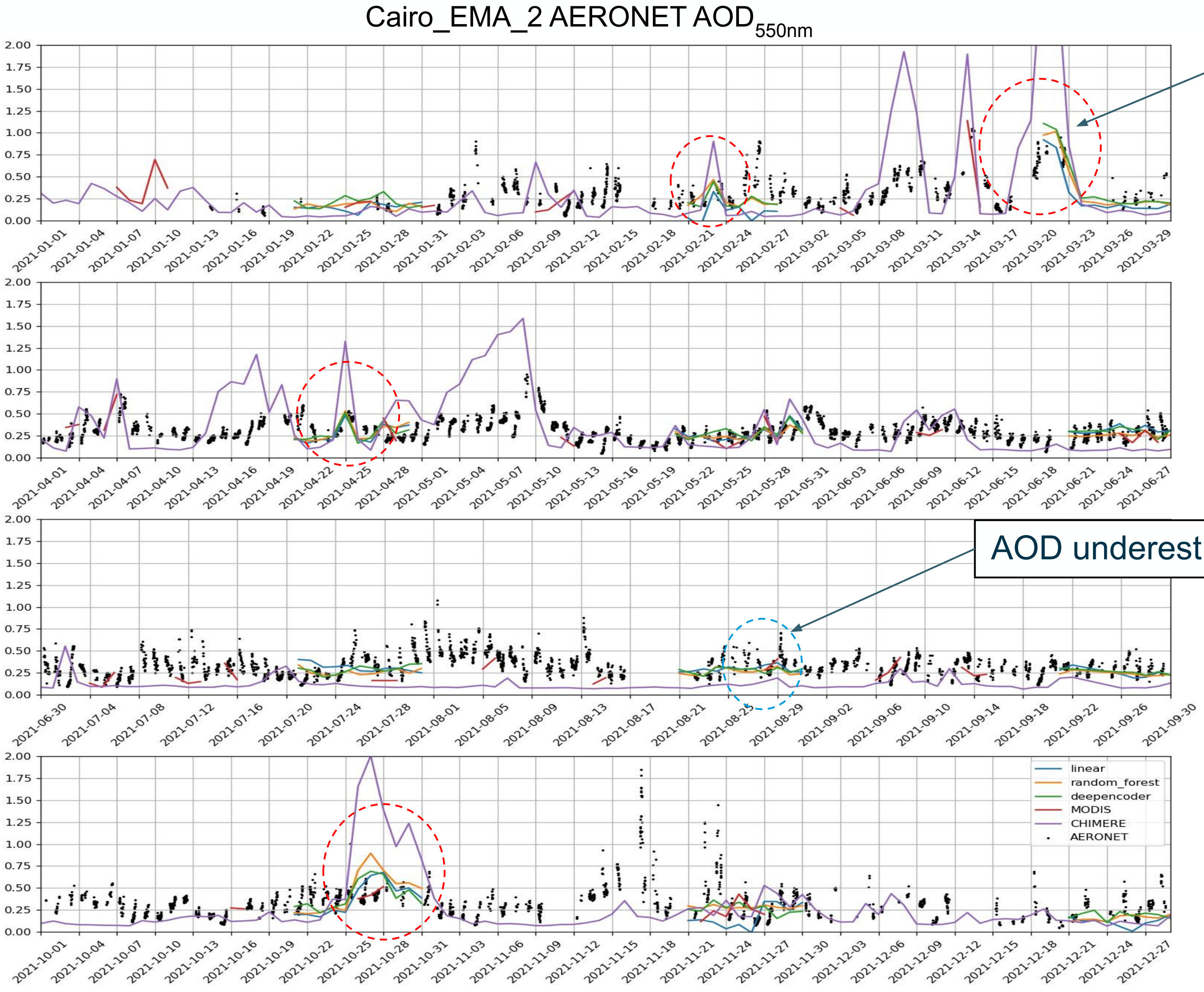


CHIMERE AOD average bias wrt MODIS on validation dataset

	Improvement in r	Improvement in RMSE
Multiple linear regression	14 %	68 %
Random forests	27 %	71 %
Neural network	23 %	69 %

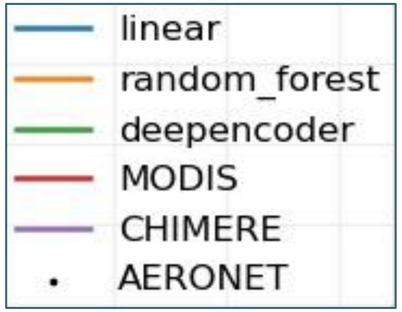
AOD comparison: CHIMERE corrected vs MODIS Aqua. N = 1.2M

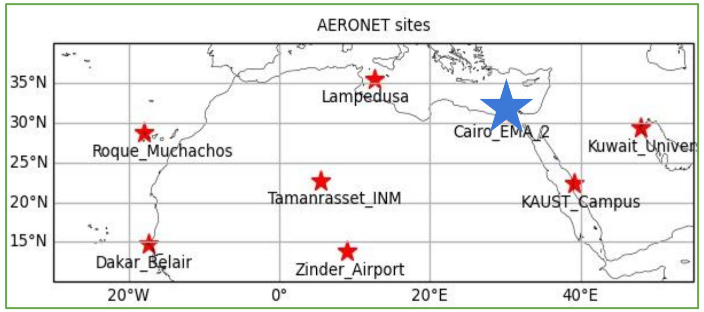
Results of AOD simulation post-processor



AOD Overestimation corrected

AOD underestimation corrected





	Improvement in r	Improvement in RMSE
Multiple linear regression	16 %	67 %
Random forests	19 %	67 %
Neural network	11 %	65 %

AOD comparison: CHIMERE corrected vs 8 AERONET Stations. N = 330

Conclusion and perspectives

Conclusions

- We have corrected CHIMERE overestimation of AOD over north Africa using MODIS aqua observations with a post processing approach.
- The best model for the correction is random forests model.

Perspectives

- Develop a post processor that uses MODIS Terra for the training.

living planet symposium | BONN 23–27 May 2022

TAKING THE PULSE OF OUR PLANET FROM SPACE

Thank you for your attention

Machine learning-based constraint of the aerosol optical depth maps simulated by the CHIMERE chemistry-transport model

Farouk Lemmouchi[1], Juan Cuesta[1], Mathieu Lachatre[2], Julien Brajard[3], Adriana Coman[1], Claude derognat[2], and Matthias Beekmann[1]

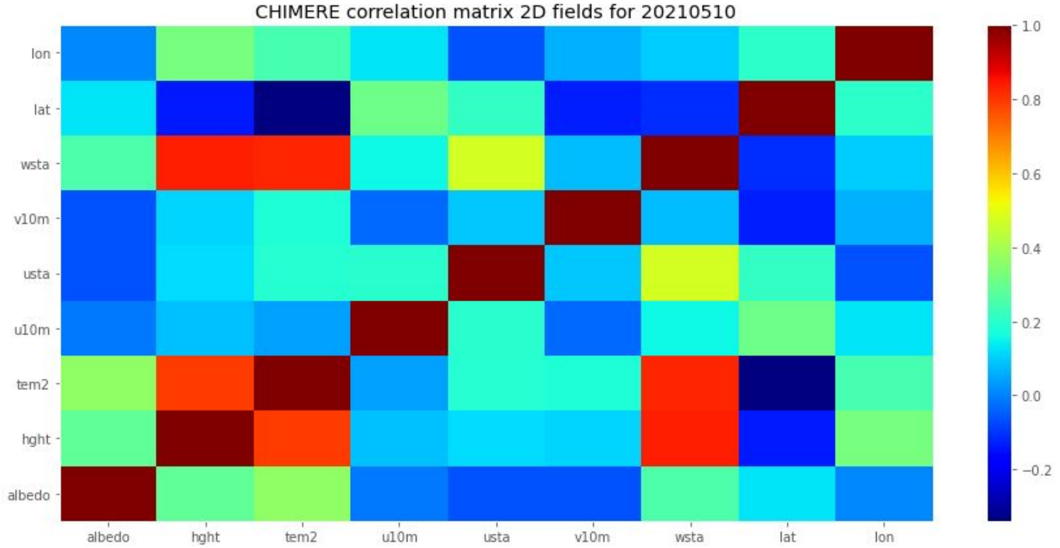
[1] LISA, UMR 7583 CNRS, Université Paris-Est Créteil, France

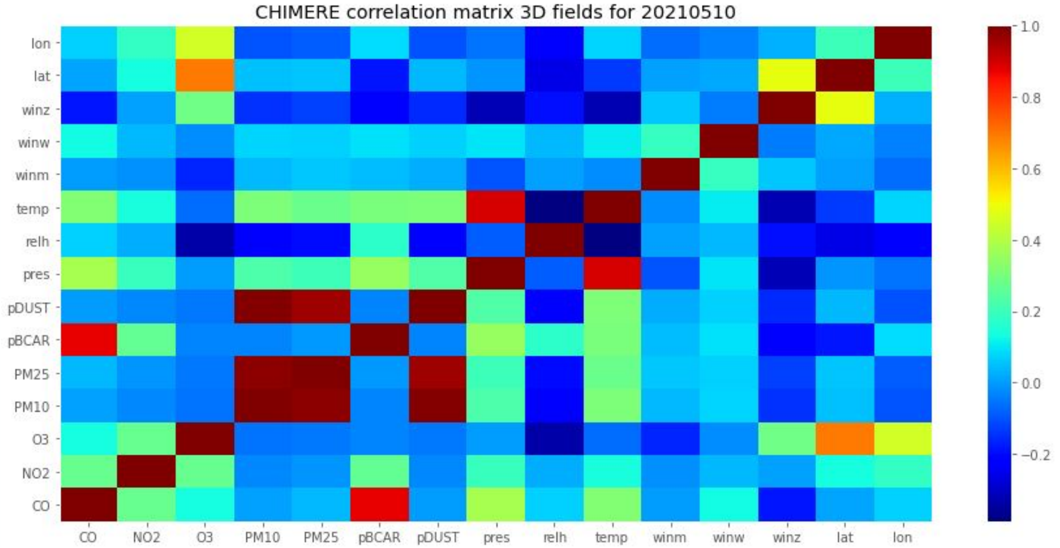
[2] ARIA Technologies, Boulogne-Billancourt, France

[3] Nansen Environmental and Remote Sensing Center (NERSC), Bergen, Norway

2022-05-27

Annex


CHIMERE correlation matrix 2D fields for 20210510
lon
lat
wsta
v10m
usta
u10m
tem2
hght
albedo
albedo hght tem2 u10m usta v10m wsta lat lon
1.0 0.8 0.6 0.4 0.2 0.0 −0.2


CHIMERE correlation matrix 3D fields for 20210510
lon
lat
winz
winw
winm
temp
relh
pres
pDUST
pBCAR
PM25
PM10
O3
NO2
CO
CO NO2 O3 PM10 PM25 pBCAR pDUST pres relh temp winm winw winz lat lon
1.0 0.8 0.6 0.4 0.2 0.0 −0.2

Annex

	aod_550	albedo	swrd	soim	sreh	hght	slhf	sshf	PM100	PM101	...	TOL3	NOX0	NOX1	NOX2	NOX3	NOY0	NOY1	NOY2	NOY3	MODIS
1275	0.173605	0.069919	898.295843	1.000000	0.738572	279.161852	98.242683	4.697709	70.857063	56.149590	...	6.172077e-05	0.536248	0.081407	0.010487	0.050271	2.003414	0.541531	0.327809	0.571008	0.230000
1276	0.170628	0.069919	899.238946	1.000000	0.718886	289.801069	100.163923	4.644310	63.921501	56.359470	...	3.352798e-05	0.343165	0.072713	0.010419	0.051613	1.568657	0.528776	0.314933	0.558541	0.231400
1277	0.168853	0.069919	899.605610	1.000000	0.714466	317.049167	100.317717	4.828505	58.366013	56.333328	...	1.717326e-05	0.237690	0.070321	0.010072	0.052869	1.234403	0.541795	0.298145	0.542610	0.250571
1278	0.168419	0.069919	898.825471	1.000000	0.735737	376.624689	102.076261	6.822394	55.027561	56.193081	...	8.639563e-06	0.189995	0.070526	0.009580	0.053663	1.044674	0.590920	0.284075	0.539785	0.264333
1279	0.170032	0.069919	897.689597	1.000000	0.803382	343.023533	89.508130	8.997177	53.121582	55.523144	...	4.709343e-06	0.132864	0.072429	0.009187	0.053752	0.797566	0.654688	0.277905	0.547298	0.272000
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
3406	0.041233	0.069650	404.070170	0.894535	0.731780	712.829796	69.509489	24.042798	8.867656	9.818521	...	1.833118e-09	0.502794	0.418913	0.141939	0.047425	2.704281	2.311237	0.833513	0.332473	0.035000
3407	0.154296	0.069645	429.228642	1.000000	0.713132	583.670152	43.704343	6.797994	9.263952	11.305456	...	2.864839e-07	0.850047	0.705863	0.047614	0.053254	3.921051	3.632363	0.469975	0.331330	0.087000
3408	0.093284	0.069641	454.739576	1.000000	0.808057	483.371571	162.203435	9.918905	26.724852	12.300276	...	5.034299e-10	0.289551	0.247068	0.088981	0.041052	0.548024	0.544248	0.348961	0.231820	0.195333
3409	0.109697	0.069641	460.826986	1.000000	0.823266	651.673693	132.336687	0.571692	29.810478	15.199094	...	5.385578e-09	0.284930	0.240420	0.086628	0.039719	0.538856	0.532429	0.342192	0.226627	0.223833
3410	0.086402	0.069641	463.430507	1.000000	0.804189	1018.642767	123.626614	-8.397679	29.961962	12.743091	...	3.667450e-08	0.293133	0.213712	0.081407	0.037910	0.535727	0.498445	0.328404	0.222273	0.236250

1445129 rows × 97 columns

Annex

Multiple linear regression

Built with ⅓ of the available dataset
fit_intercept=False, positive=False
Important coefficients: pOH, specific humidity

Random forest model

2 fold cross validation using 10k data points for forests exploration, then using ⅔ of the available dataset for training.

max_features=[20, 25, 30]
n_bins=[64, 128]
min_samples_leaf=[2, 4, 6],
min_samples_split=[2, 4, 6],
n_estimators=[100]
scoring='neg_mean_squared_error

Neural network model

Training dataset : ⅓ of the available dataset
Testing dataset : ⅓ of the available dataset
N_layers : 4 to 10
N neurons : 10 to 50 per layer
Adam optimizer with respect to mse.
'batch_size', 5000, 20000, step=1000
100 random trials
50 epocs