ESA-ECMWF Workshop 2021

INTEGRATING REANALYSIS AND SATELLITE CLOUD INFORMATION TO ESTIMATE DOWNWARD LONG-WAVE RADIATION FLUXES USING MULTIVARIATE ADAPTIVE REGRESSION SPLINES

APPLICATION TO EUMETSAT LSA-SAF

Francisco Lopes,[1] Emanuel Dutra,[1,2] and Isabel Trigo,[1,2]

fmtlopes@fc.ul.pt, endutra@fc.ul.pt, isabel.trigo@ipma.pt

[1]Instituto Dom Luiz (IDL), Faculty of Sciences, University of Lisbon, Campo Grande, 1749-016, Lisbon, Portugal.
[2]Portuguese Institute for Sea and Atmosphere (IPMA), Rua C do Aeroporto, 1749-077, Lisbon, Portugal.

Motivation

- Downward long-wave radiation (DLR) is essential to: Earth's surface energy balance, heat exchange fluxes, climate variability and global warming calculations;
- DLR is highly dependent on the vertical profiles of atmospheric temperature and, most noticeably, water vapour. Clouds significantly influence DLR variability by increasing the total effective emissivity of the sky;
- Semi-empirical, physical and hybrid (physical + remote sensed data) models have some degree of dependency with factors that hinder their accuracy (particular calibration conditions, quality and availability of atmospheric profiles database and satellite information accuracy);

Objectives

- Review of LSA-SAF DLR algorithm (Trigo et al. 2010[1]): reanalysis integration with new atmospheric profiles database (ERA5) for the same conditions (periods, locations) used in the original calibration, i.e. TIGR-like database (Chevallier et al. 2000[2]) heavily based on ERA-40 and MODTRAN4 fluxes;
- A new formulation is proposed: combination of reanalysis (ERA5), ground (BSRN+ARM) and remote sensed (MSG) data with a machine learning model that uses Multivariate Adaptive Regression Splines (MARS);
- Benchmarking and validation of proposed methodology against other models estimates and several ground stations within MSG-disk;

Multivariate Adaptive Regression Splines - MARS

MARS main reference: JH Friedman. Multivariate adaptive regression splines. The annals of statistics, 19(1):1–67, 1991.

- A non-parametric technique that automatically builds a model having into account nonlinearities;
- Belongs to a group of regression algorithms used to predict continuous (numerical) variables;
- Makes use of **hinge functions*** with the form max(0, x-cte) or max(0, cte-x) that can model interaction between two or more variables, where cte are the "knots" of the hinge functions;
- MARS 2-stage building process: forward and backward pass;

(Training set)

Forward pass

- Adds repeatedly basis functions in pairs;
- Finds pair of basis functions with maximum reduction in **sum-of-squares residual error**;
- Searches all combinations;
- Calculates coefficient of each term, a linear regression over terms is applied;
- Process stops when residual error is very small or when max. number of terms is reached;

Backward pass

- **Compensates** the forward pass (usually builds an **overfit** model that is good to the data but not to the new data);
- Removes terms one by one (**pruning**), deleting the least effective term at each step;
- Subsets are compared using the **general cross-validation** (GCV) criteria;
- Has the advantage to choose any term to delete, while forward pass only sees next pair of terms at each step;

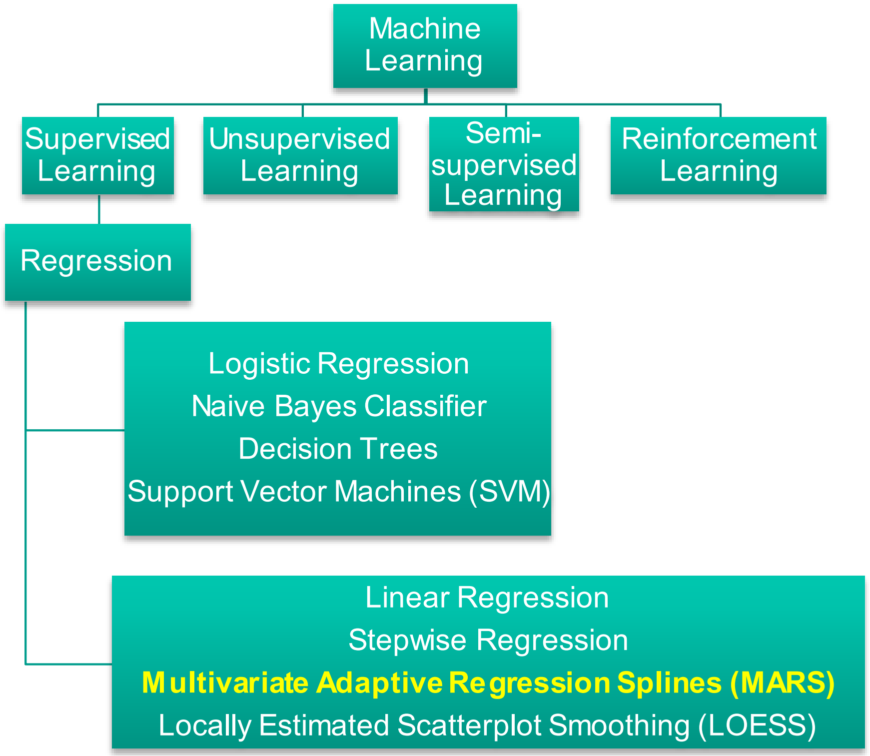

Fig. 2 – MARS belongs to the family of machine learning algorithms.

***Hinge function:**
h(x-c) = max(0, x-c) = {x−c, if x>0; and 0, if x≤c},
where c is a constant also known as a knot

MARS – ERA5 Integration

- ERA5 data extraction (https://cds.climate.copernicus.eu);
- ERA5 hourly outputs: total column water vapour (tcwv), 2-metre air temperature (t2m), 2-metre dewpoint temperature (d2m), surface thermal radiation downwards (strd), and total cloud cover (tcc);
- MARS calibration with ERA5 database:
  - Predictand: strd fluxes;
  - Predictors: tcwv, t2m and d2m;
  - Sky conditions definition: tcc

    Clear: tcc = 0
    Cloudy: tcc > 0.9

Table 1 – Predictors assessment for MARS calibration. Bias (μ) and root mean square error (RMSE) are depicted ($W.m^{-2}$). Analysis performed against ERA5 fluxes during a period of ~4.6 years (1994-1995, 2004-2005 and 2014-2015).

	Clear Sky		Cloudy Sky	
Predictors	μ	RMSD	μ	RMSD
tcwv, t2m, d2m	0.3	7.9	-0.1	16.1
tcwv, t2m	0.4	8.9	-0.3	16.4
tcwv	-0.8	24.7	-1.0	21.2
tcwv, d2m	-0.7	23.5	-0.2	16.9
t2m, d2m	0.3	13.1	0.9	18.5
t2m	0.6	20.5	1.0	21.5
d2m	-1.0	32.2	0.7	19.2

Observational data

Station	Network	Location	Lat. and Lon.	Elev.	Years	Annual DLR
Brasilia	BSRN	Brazil	15.60˚S; 47.71˚W	1023	44.47	364.45
Budapest	BSRN	Hungary	47.43˚N; 19.18˚E	139	0.51	373.82
Cabauw	BSRN	Netherlands	51.97˚N; 4.93˚E	0	91.76	323.69
Camborne	BSRN	U.K.	50.22˚N; 5.32˚W	88	72.70	324.57
Carpentras	BSRN	France	44.08˚N; 5.06˚E	100	88.39	321.74
Cener	BSRN	Spain	42.82˚N; 1.60˚W	471	64.17	321.71
De Aar	BSRN	South Africa	30.67˚S; 23.99˚E	1287	39.05	303.88
Eastern North Atlantic	BSRN	Azores	39.09˚N; 28.03˚W	15.2	6.24	359.34
Florianopolis	BSRN	Brazil	27.61˚S; 48.52˚W	11	35.62	386.40
Gandhinagar	BSRN	India	23.11˚N; 72.63˚E	65	9.85	401.45
Gobabeb	BSRN	Namibia	23.56˚S; 15.04˚E	407	47.12	338.67
Neumayer	BSRN	Antarctica	70.65˚S; 8.25˚W	42	92.28	216.87
Niamey	ARM	Africa	13.48˚N; 2.18˚E	223	6.36	392.11
Lindenberg	BSRN	Germany	52.21˚N; 14.12˚E	125	87.36	315.06
Palaiseau	BSRN	France	48.71˚N; 2.21˚E	156	97.61	322.61
Paramaribo	BSRN	Suriname	5.81˚N; 55.22˚W	4	3.63	421.16
Payerne	BSRN	Switzerland	46.82˚N; 6.94˚E	491	98.04	315.05
Petrolina	BSRN	Brazil	9.07˚S; 40.32˚W	387	47.25	386.86
Sede Boqer	BSRN	Israel	30.86˚N; 34.78˚E	500	46.80	332.86
São Martinho da Serra	BSRN	Brazil	29.44˚S; 53.82˚W	489	37.74	327.19
Sonnblick	BSRN	Austria	47.05˚N; 12.96˚E	3109	39.23	249.07
Tamanrasset	BSRN	Algeria	22.79˚N; 5.53˚E	1385	99.17	330.70
Toravere	BSRN	Estonia	58.25˚N; 26.46˚E	70	98.05	308.71

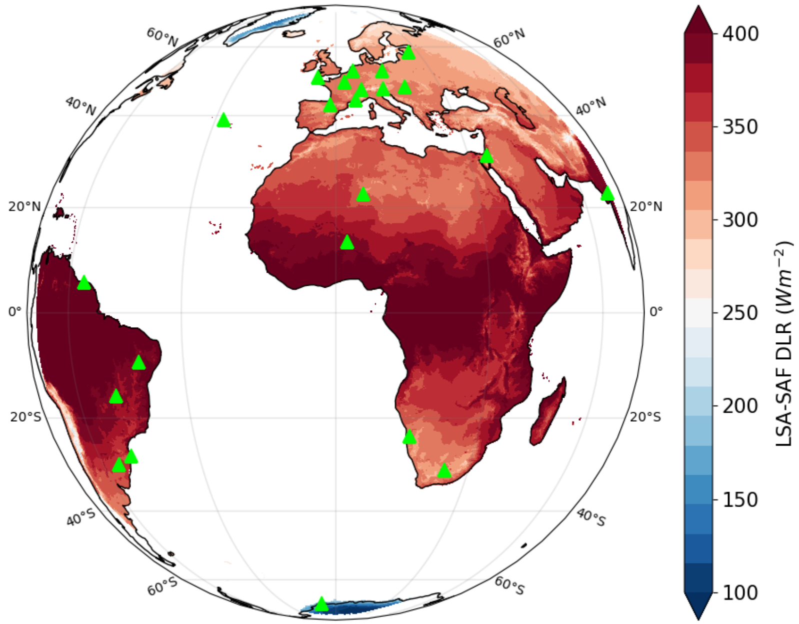

Fig. 3 – Annual mean (2020) DLR at surface estimated with the LSA-SAF operational algorithm within the MSG-disk. Depiction of all 23 measuring stations (green triangles) used for the validation of MARS estimates comprising a 16-year period of study (from 2004 to 2019).

Table 2 – Ground stations used. Name, network of origin, geographical coordinates (˚), elevation (m), years available (i.e. % of available data, in years, between 2004 and 2019), and annual mean DLR ($W.m^{-2}.year^{-1}$).

Models description

- **MARS_MOD**: ~4.6 years of ERA5 (tcwv, t2m, d2m as predictors and strd as predictand) to calibrate the model;
- **LSA_MOD**: ~4.6 years of ERA5, same inputs as for MARS_MOD but using operational algorithm;
- **MARS_OBS**: ~10.4 years of measured data (random selection of 6 months of data from each of the 23 stations) + ERA5; stations with less than 6 months provide 40% of data instead;
- **LSA_OBS**: ~10.4 years of measured data + ERA5, same inputs as for MARS_OBS but using operational algorithm;
- **LSA_OPER**: ~1.5 years of TIGR-like database (ERA40) + MODTRAN4 fluxes;
- **ERA5**: direct output - surface thermal radiation downwards (strd);

Acronym	Model	Predictors	Predictand	Cloud info.	Calibration period
MARS_MOD **LSA_MOD**	MARS LSA	tcwv, t2m, d2m (ERA5)	strd (ERA5)	tcc (ERA5)	1992-1993, 2002-2003, 2012-2013
MARS_OBS **LSA_OBS**	MARS LSA	tcwv, t2m, d2m (ERA5)	BSRN + ARM	cma (MSG)	2004-2019
LSA_OPER	LSA	tcwv, t2m, d2m (ERA40)	MODTRAN4	tcc (ERA-40)	1992-1993
ERA5	Reanalysis	-	-	-	-

Table 3 – List of models and respective calibration inputs used in the analysis.

Results – ERA5 integration + MARS

- Better adjustments to ERA5 fluxes are obtained with MARS_MOD under clear and cloudy sky conditions;
- "S" shape curve from the operational product is mostly related to TIGR-like database and MODTRAN-4 fluxes used in original calibration. LSA model with ERA5 (LSA_MOD) eliminates this feature;
- MARS deviations under cloudy periods: related to the misrepresentation of clouds in ERA5, i.e. total cloud cover (used for sky classification);
- More accurate cloud information can be used from satellites to define sky conditions: e.g. MSG cloud mask (cma);
- Further validation of MARS estimates using ground measurements;

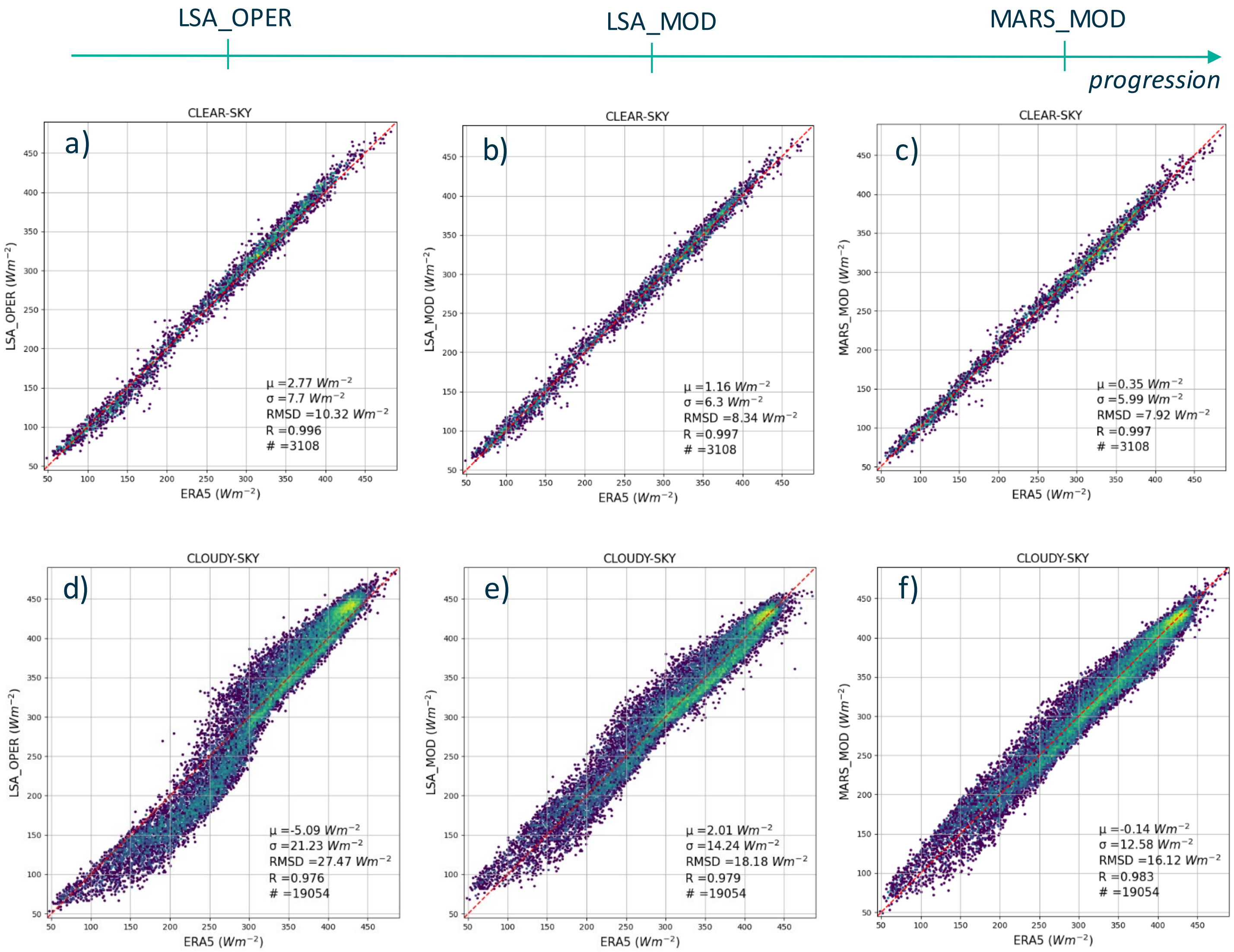

Fig. 4 - Benchmarking (1994-1995, 2004-2005 and 2014-2015) for: **(a, d)** LSA_OPER; **(b, e)** LSA_MOD; (c, f) MARS_MOD.

Results – MARS validation

- Models that use measured data seem to have better performance (MARS_OBS, LSA_OBS);
- Use of MARS to estimate DLR values at surface provides better adjustments to observations;
- Higher errors: ERA5;
- ERA5 overall underestimation is transferred into MARS and LSA models (MARS_MOD, LSA_MOD), being reduced in MARS_OBS;
- LSA_OPER overall underestimation due to original calibration (TIGR-like, MODTRAN-4);

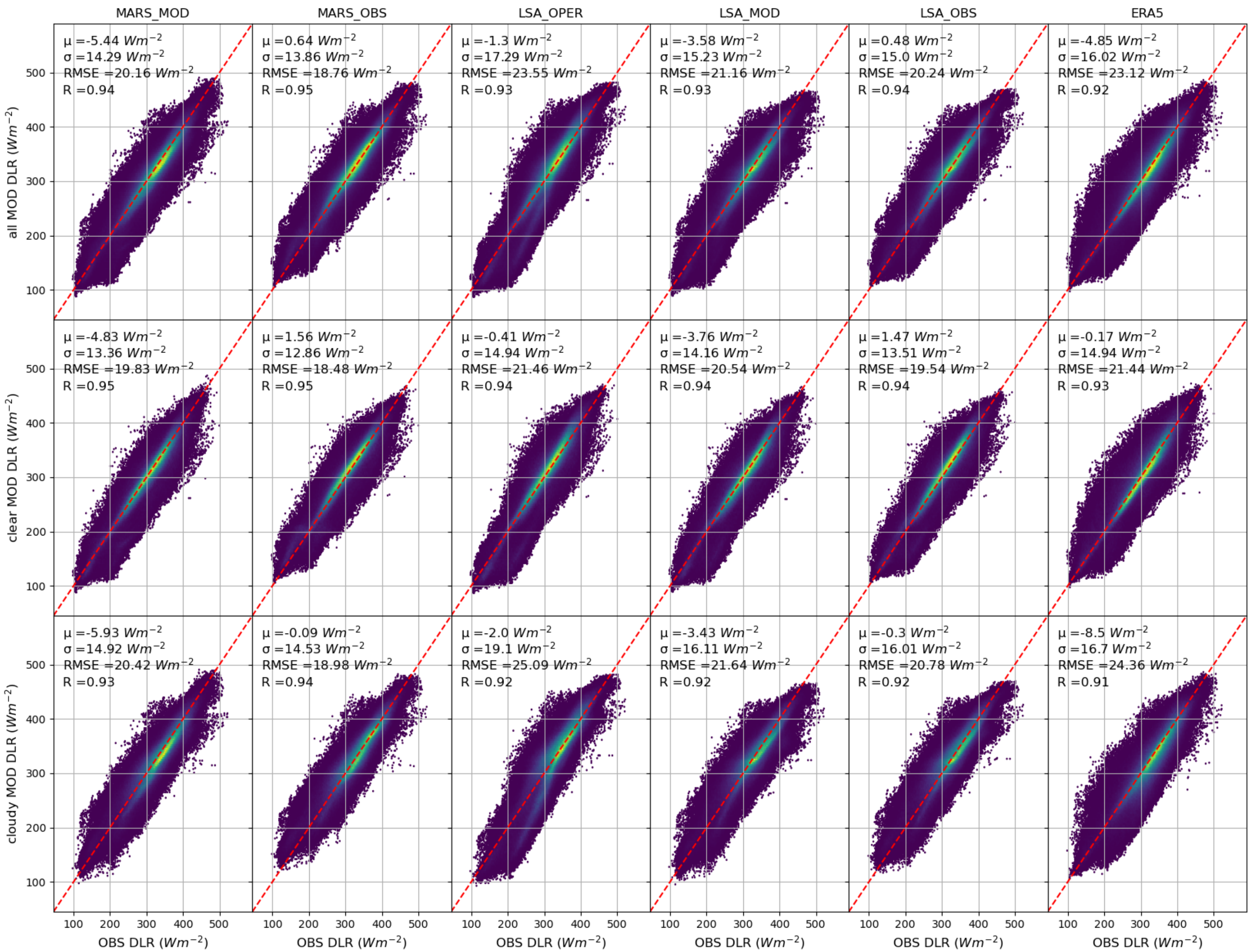

Fig. 5 - DLR fluxes ($W.m^{-2}$) for all, clear and cloudy sky conditions from all models vs. hourly observations from 23 stations.

Results – MARS validation

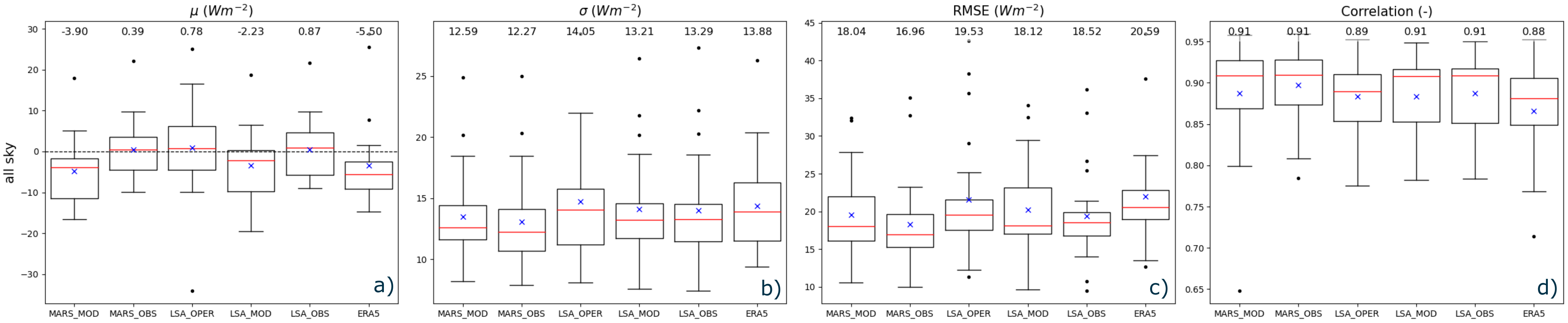

Fig. 6 - Boxplots for all sky conditions considering different error metrics, i.e. bias (μ), deviation (σ), RMSE, and Pearson's correlation (R), found in all measuring stations for each model. Validation period: between 2004-2019.

- Statistical summary (all sky): hourly data from all measuring stations in all model variants;
- Lower errors: MARS calibrate with measured data;
- Higher errors: ERA5 Reanalysis (related to cloud representation);
- Outliers occurrence in all models: GVN, SON, SMS stations (high latitudes, high altitudes, possible measuring equipment malfunction);

Results – Spatial distribution

<u>MSG-disk mapping:</u>

- **Annual mean differences (2020);**

➢ Spatial variation is consistent between all models;

➢ Dominant gradients: higher values in tropics and lower values in high latitudes;

➢ Systematic differences are observed: e.g. higher altitudes (ERA5) and higher latitudes (LSA_OPER);

➢ MARS_OBS smaller differences towards LSA_OPER with slight underestimation within the tropics;

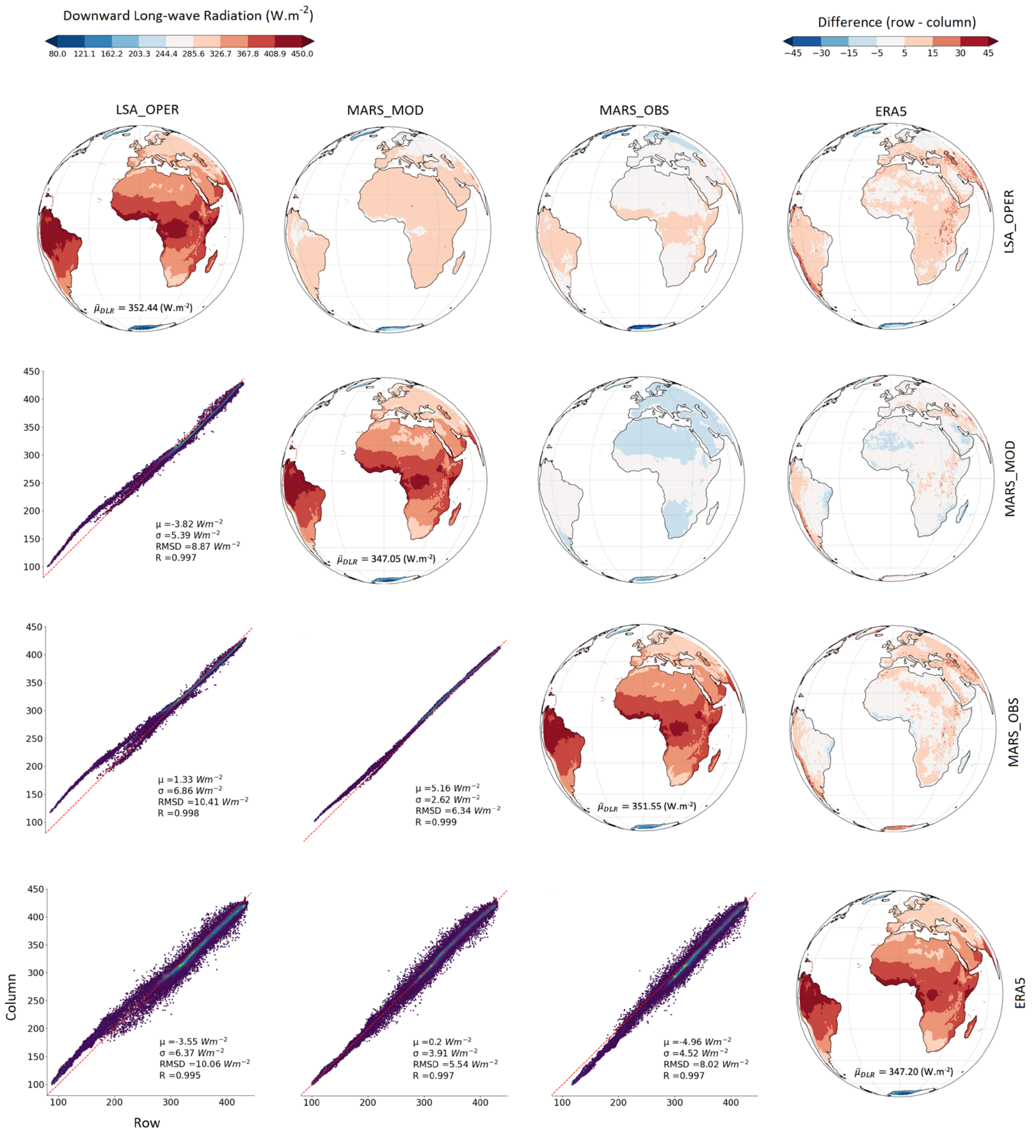

Fig. 7 - DLR (W.m^{-2}) mapping: annual means 2020.

Results – Spatial distribution

<u>MSG-disk mapping:</u>

- **Seasonal mean differences 2020 (DJF);**

➢ High agreement among all models with respect to seasonality;

➢ LSA_OPER increases underestimation towards other models estimates where lower values of DLR fluxes occur (e.g. eastern Europe);

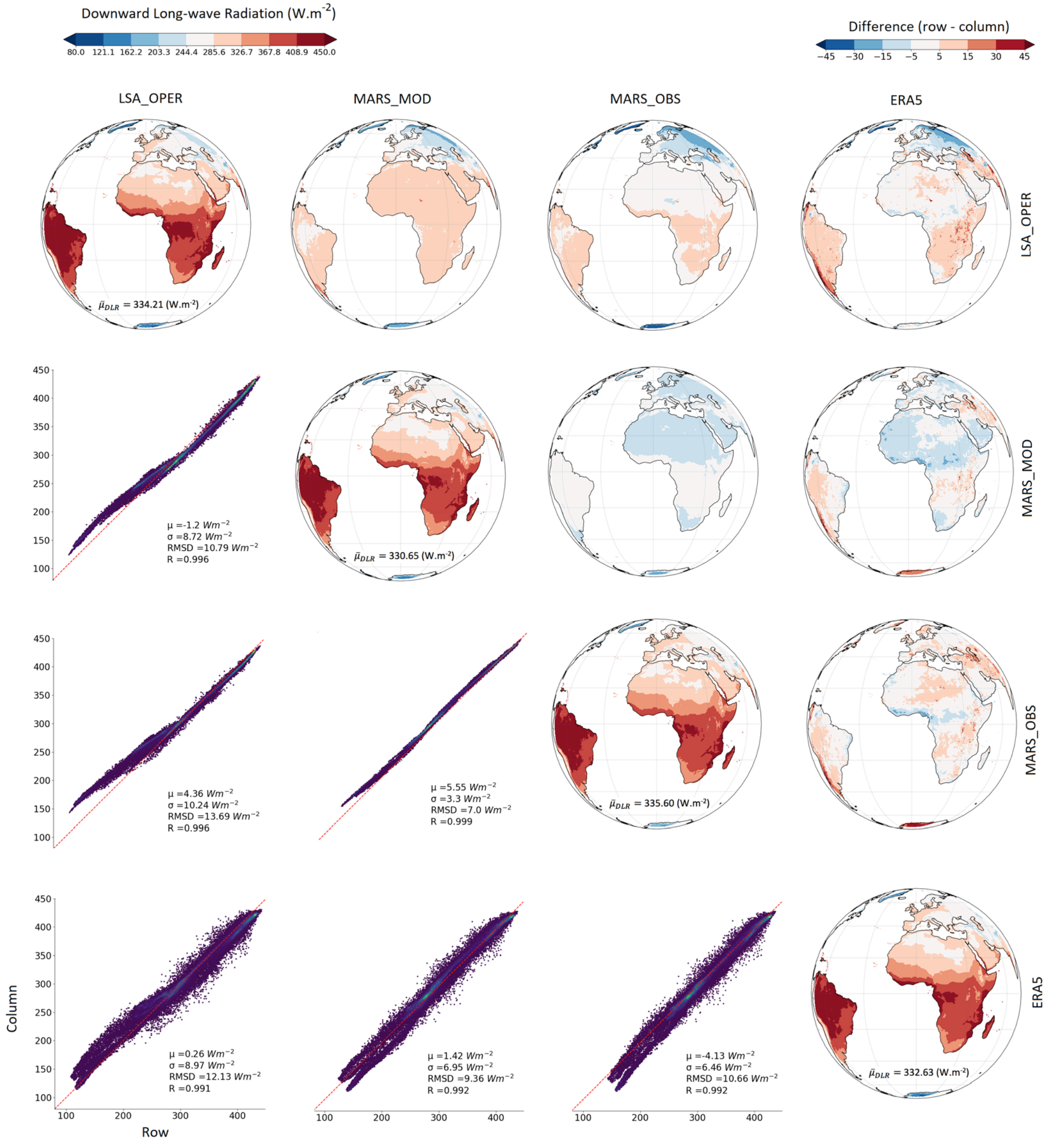

Fig. 8 - DLR (W.m^{-2}) mapping: seasonal (DJF) means 2020.

Results – Spatial distribution

MSG-disk mapping:

- **Seasonal mean differences 2020 (MAM);**

➢ High agreement among all models with respect to seasonality;

➢ Increase of DLR fluxes, particularly in northern Africa and Europe; in comparison with LSA_OPER, the error decrease in these regions, except ERA5 (general underestimation);

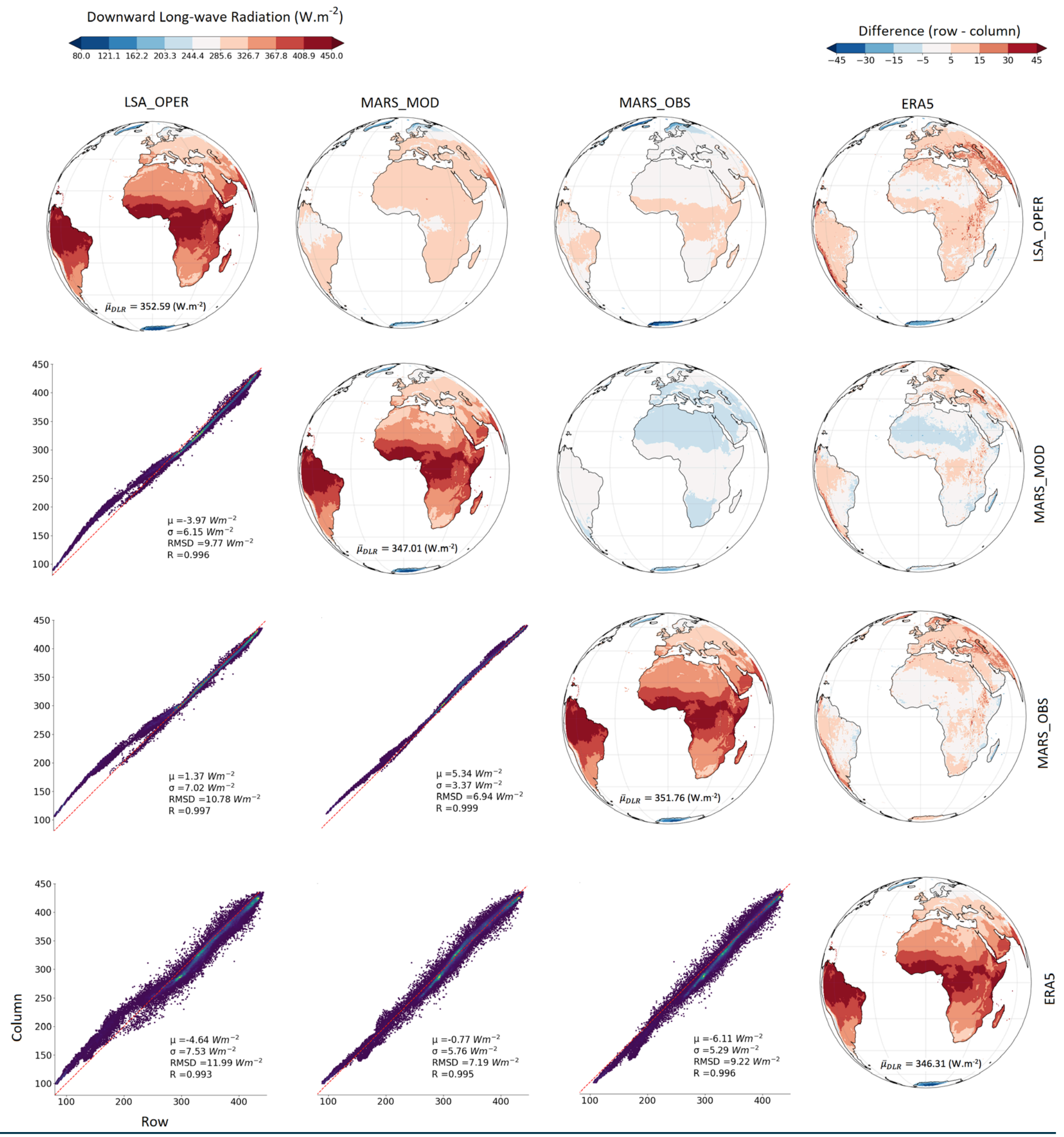

Fig. 9 - DLR (W.m^{-2}) mapping: seasonal (MAM) means 2020.

Results – Spatial distribution

MSG-disk mapping:

- **Seasonal mean differences 2020 (JJA);**

➢ High agreement among all models with respect to seasonality;

➢ Higher increase of DLR fluxes in northern Africa and Europe, and higher decrease in southern Africa (very small differences between all models);

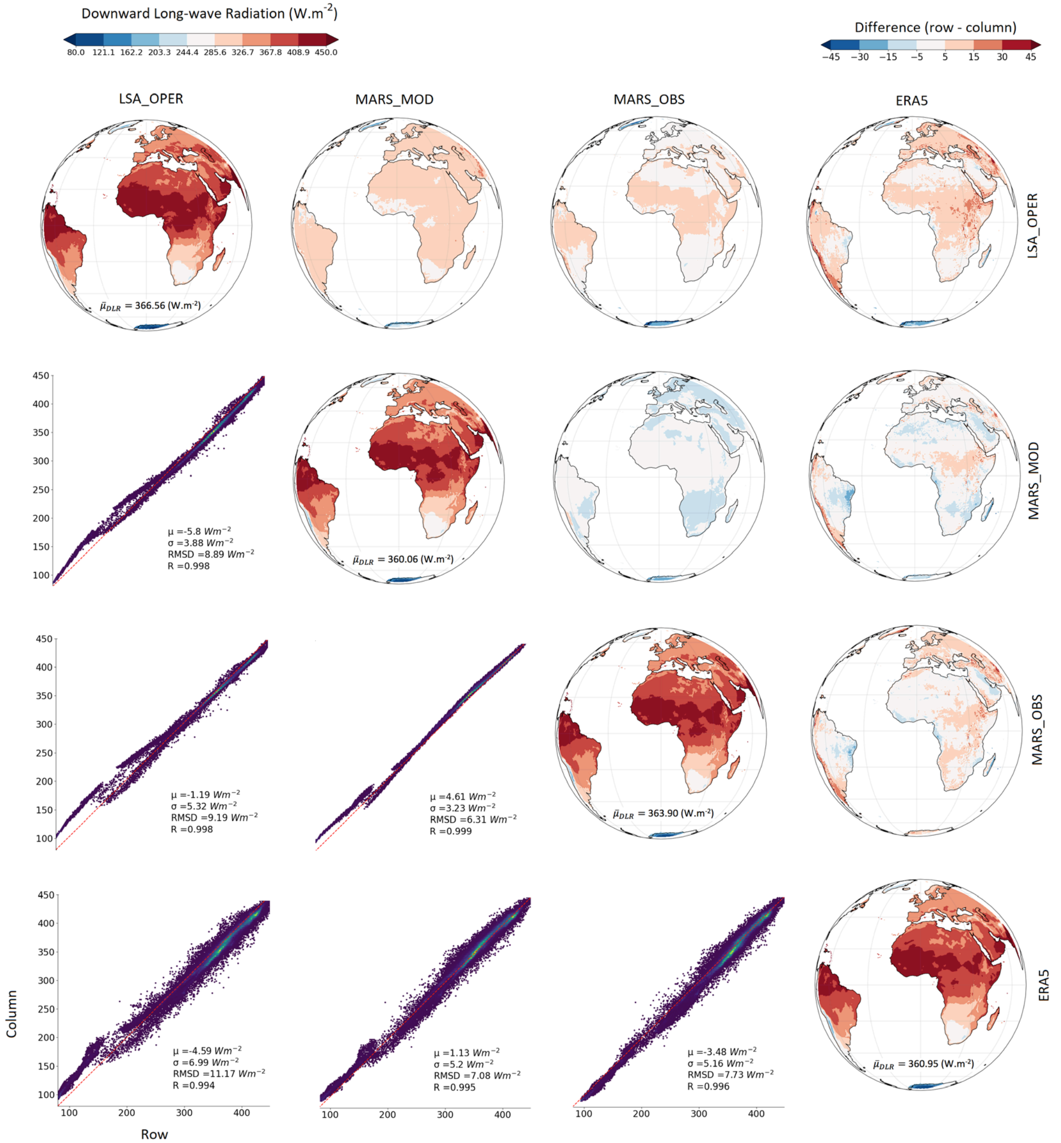

Fig. 10 - DLR (W.m^{-2}) mapping: seasonal (JJA) means 2020.

Results – Spatial distribution

<u>MSG-disk mapping:</u>

- **Seasonal mean differences 2020 (SON);**

➢ High agreement among all models with respect to seasonality;

➢ Decrease of DLR fluxes in northern hemisphere and increase in southern hemisphere;

➢ Similar spatial variation between all models in comparison to spring spatial variation;

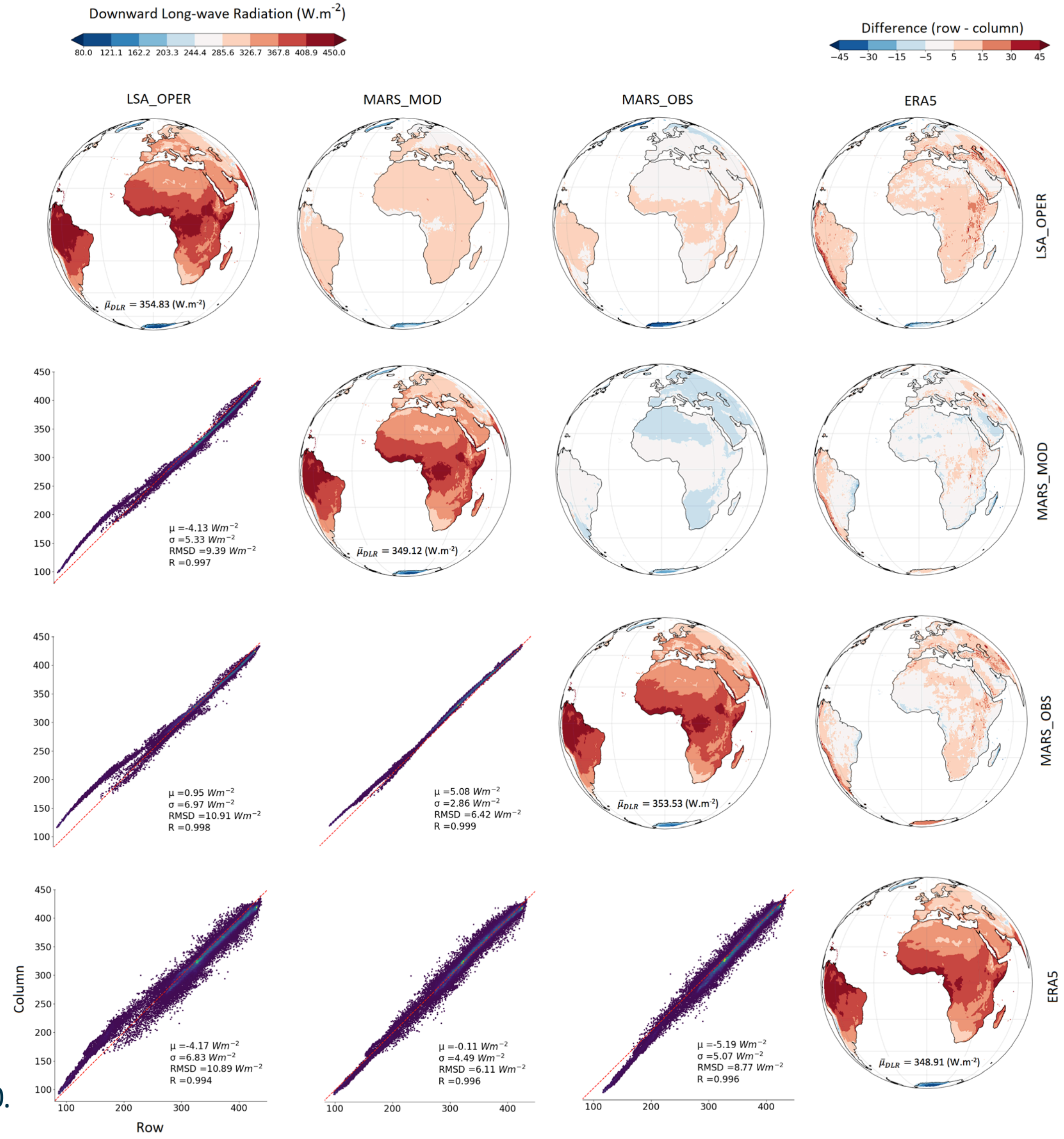

Fig. 11 - DLR (W.m^{-2}) mapping: seasonal (SON) means 2020.

Conclusions

Summary

- LSA-SAF operational product calibrated with ERA5 atmospheric profiles database seems to provide better adjustments for DLR fluxes instead of using the original calibration (ERA-40, MODTRAN-4); Better description of near surface variables and radiation fluxes;
- Overall, MARS performs well, particularly when measured data is used for calibration;
- DLR fluxes maps (MSG-disk) show a general consistency between all model estimates, in particular MARS_OBS shows a higher approximation to the LSA-SAF operational product than the remaining models;

Future work

- Improvements should be expected in the future with the replacement of McRad radiation scheme (currently operational in the ERA5 reanalysis) by the ecRad scheme;
- Other MARS variants can be tested with a new configuration of predictors, e.g. use of satellite measured thermal infrared bands top of atmosphere radiances for calibration, similarly to (Zhou et al., 2018[3]);

References & acknowledgements

[1]Trigo, I.F., C. Barroso, P. Viterbo, S.C. Freitas, and I.T. Monteiro. Estimation of downward long-wave radiation at the surface combining remotely sensed data and NWP data, J. Geophys. Res., 115, D24118, 2010. https://doi.org/10.1029/2010JD013888

[2]Chevallier, F. Chédin, A., Chéruy, F., and Morcrette, J.-J. TIGR-like atmospheric-profile databases for accurate radiative-flux computation, Q. J. R. Meteorol. Soc., 126, pp. 777–785, 2000. https://doi.org/10.1002/qj.49712656319

[3]Zhou W., Wang, T., Shi, J., Peng, B., Zhao, R., and Yu, Y. Remote Sensed Clear-Sky Surface Longwave Downward Radiation by Using Multivariate Adaptive Regression Splines Method. IEEE International Geoscience and Remote Sensing Symposium, 22-27 July 2018, Valencia, Spain. https://doi.org/10.1109/IGARSS.2018.8519297

This research was co-funded by the *Fundação para a Ciência e a Tecnologia* (FCT) and *Associação para a Investigação e Desenvolvimento de Ciências* (FCiências.ID), research grant PTDC/CTA-MET/28946/2017

https://www.fciencias-id.pt/node/1 http://idl.campus.ciencias.ulisboa.pt/ https://www.ipma.pt/en/index.html https://www.eumetsat.int/lsa-saf

fmtlopes@fc.ul.pt

Thank you.