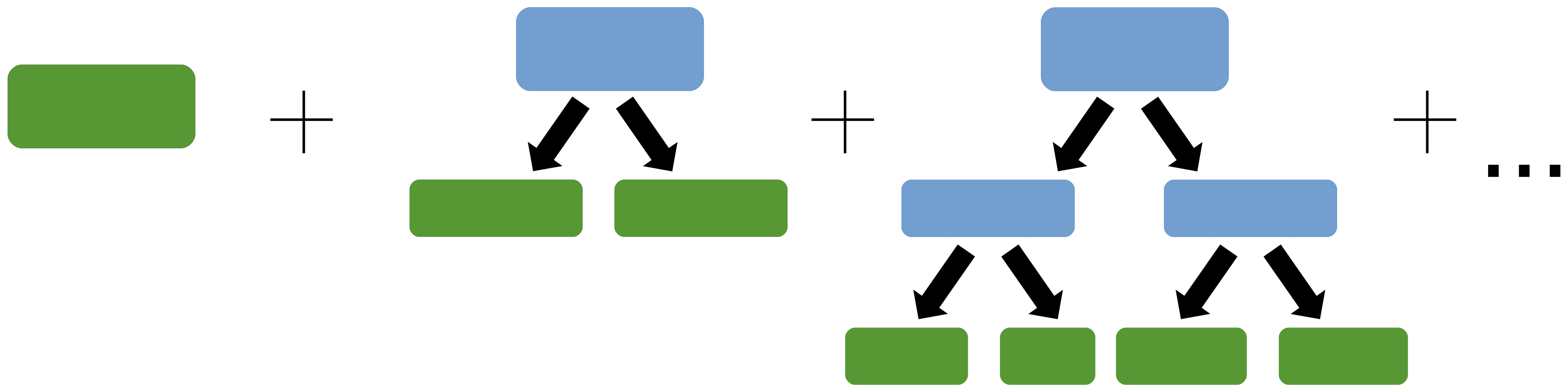
Investigation of the relative importance of meteorological and aerosol-related parameters in determining cloud microphysical properties using Machine Learning methods

Tobias Lord and Frida Bender, Department of Meteorology, Stockholm University, Sweden

Gradient Boosting Regression models are trained to predict Cloud Effective Radius over 10 regions.

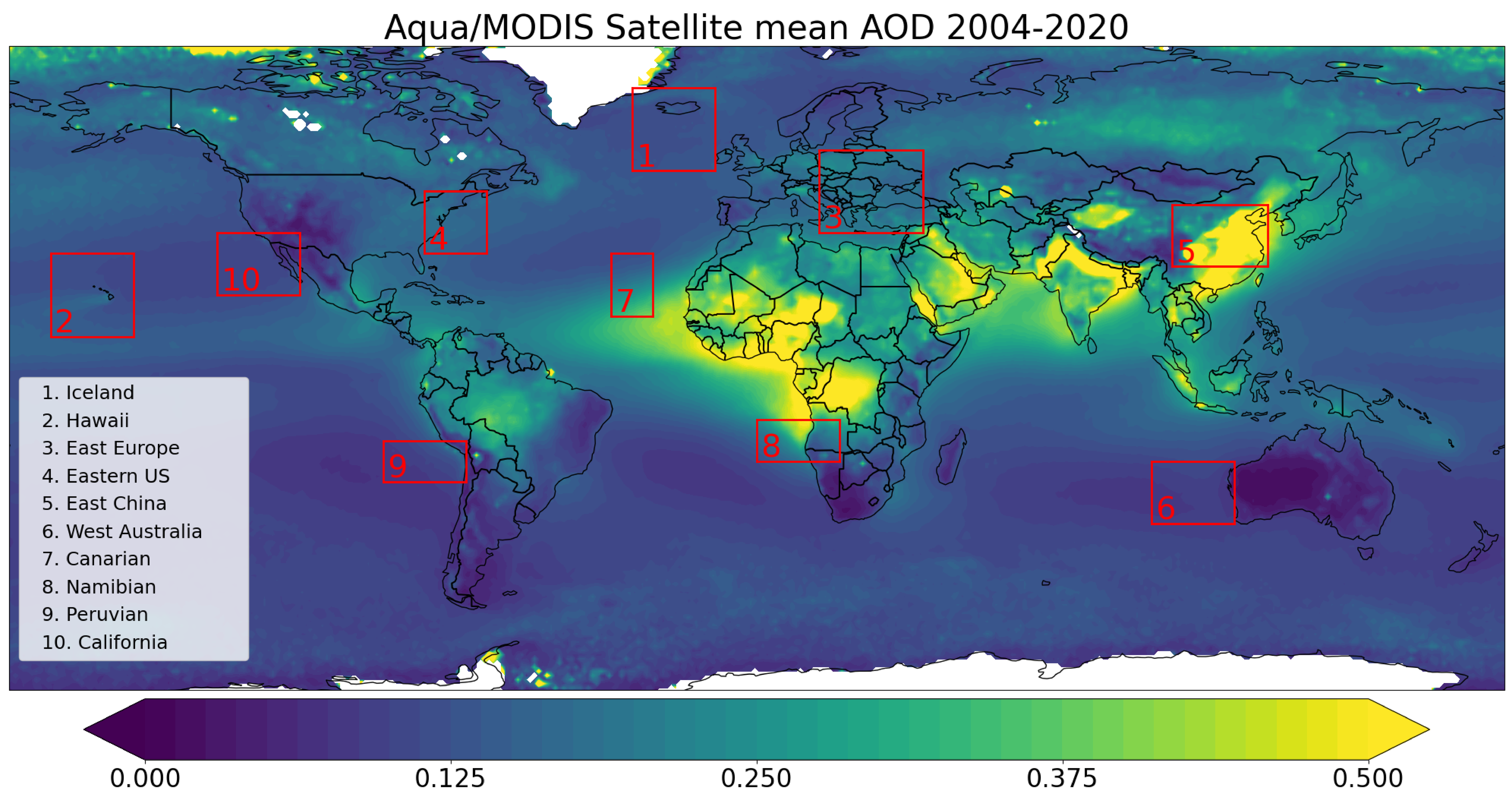


Input parameters:
-Meteorological variables. ERA-5 Reanalysis
-AOD and SO2 are used as Aerosol information. MODIS (Aqua) and OMI (Aura) respectively

Parameters are ranked depending on their importance in determining model's output of Cloud Effective Radius. Essentially how often they occur in the decision trees.

Funded by:
Swedish e-Science Research Centre

The computations were performed on resources provided by the Swedish National Infrastructure for Computing (SNIC) at NSC.



Results:
-Using large scale meteorological reanalysis data we show it is possible to predict a microphysical cloud parameter (cloud effective radius) through a simple Machine Learning Regression Model.

-Additionally, when also given aerosol information (AOD from satellite) the model score is slightly improved.

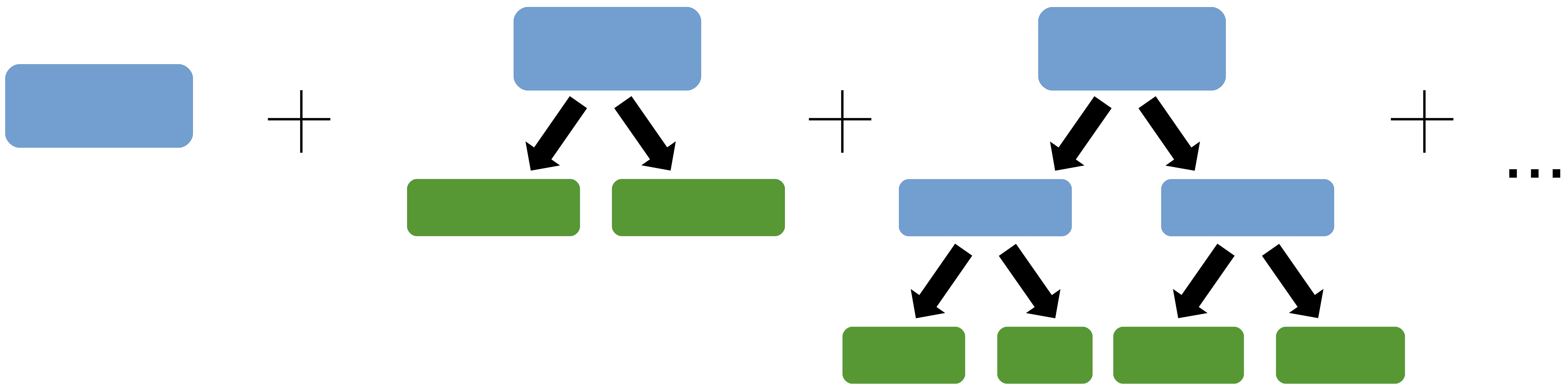
See Breakout session 3 for full results!

tobiaslord8@gmail.com

Gradient Boosting Regression

Stockholm University

A Gradient Boosting regression model is trained on meteorological reanalysis data and Aerosol satellite data for 10 regions. The model works as an ensemble of smaller decision trees.



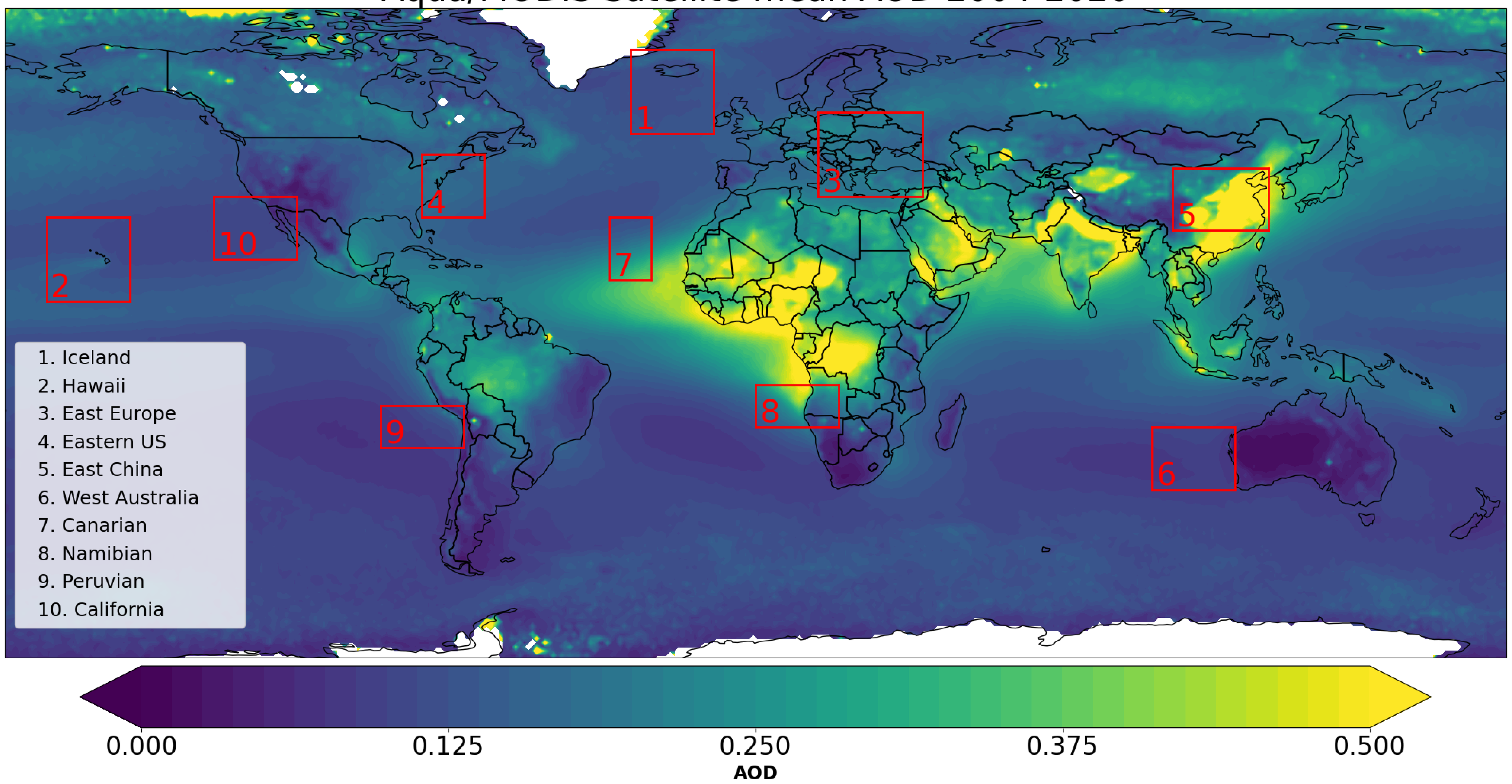
Input Parameters:
-Daily meteorological data: ERA-5 Reanalysis
T, RH, Gph, u, v, w, divergence.
All on 700,850 and 1000hPa

-AOD: L3 Dark Target Deep Blue
MODIS (Aqua)

-SO2: L3 Total Column Density in PBL
OMI (Aura)

Output: Cloud Effective Radius (r_e) which is compared to observations from MODIS (Aqua)

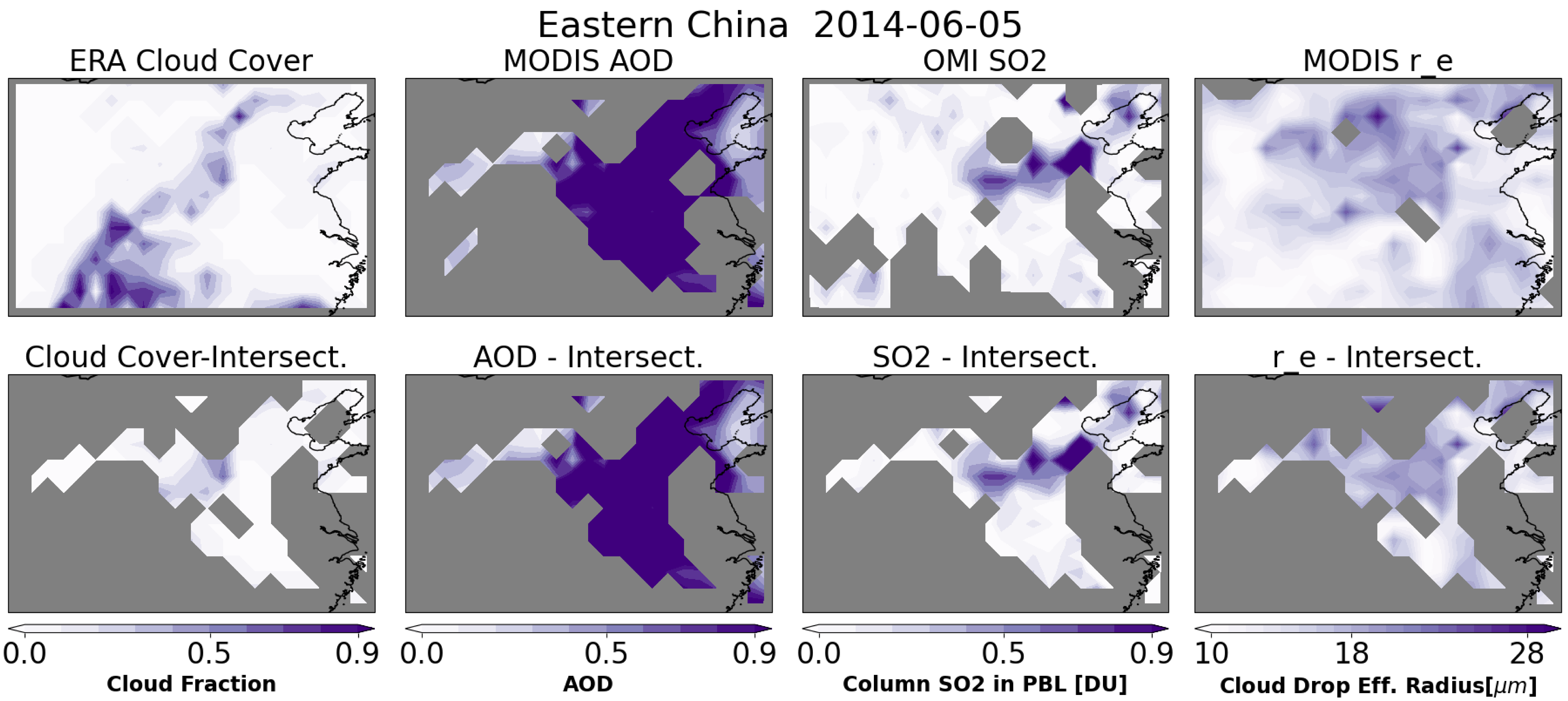
Aqua/MODIS Satellite mean AOD 2004-2020



Map shows the regions included in the study. Both Volcanic, Anthropogenic and Stratocumulus regions were included.

Reanalysis and Satellite Data coverage

Stockholm University



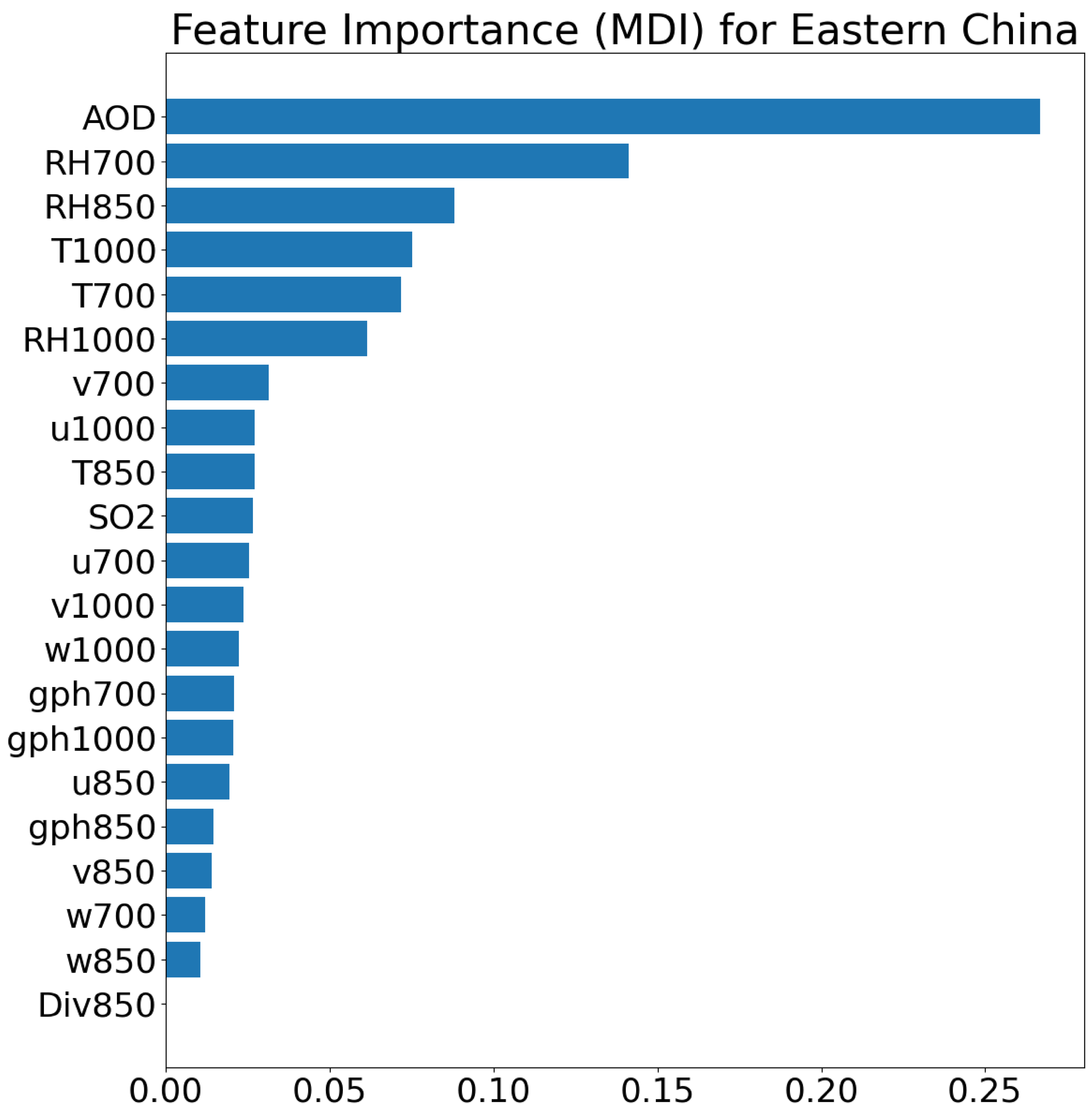
-Data coverage varies between ERA-5 and satellite retrievals.

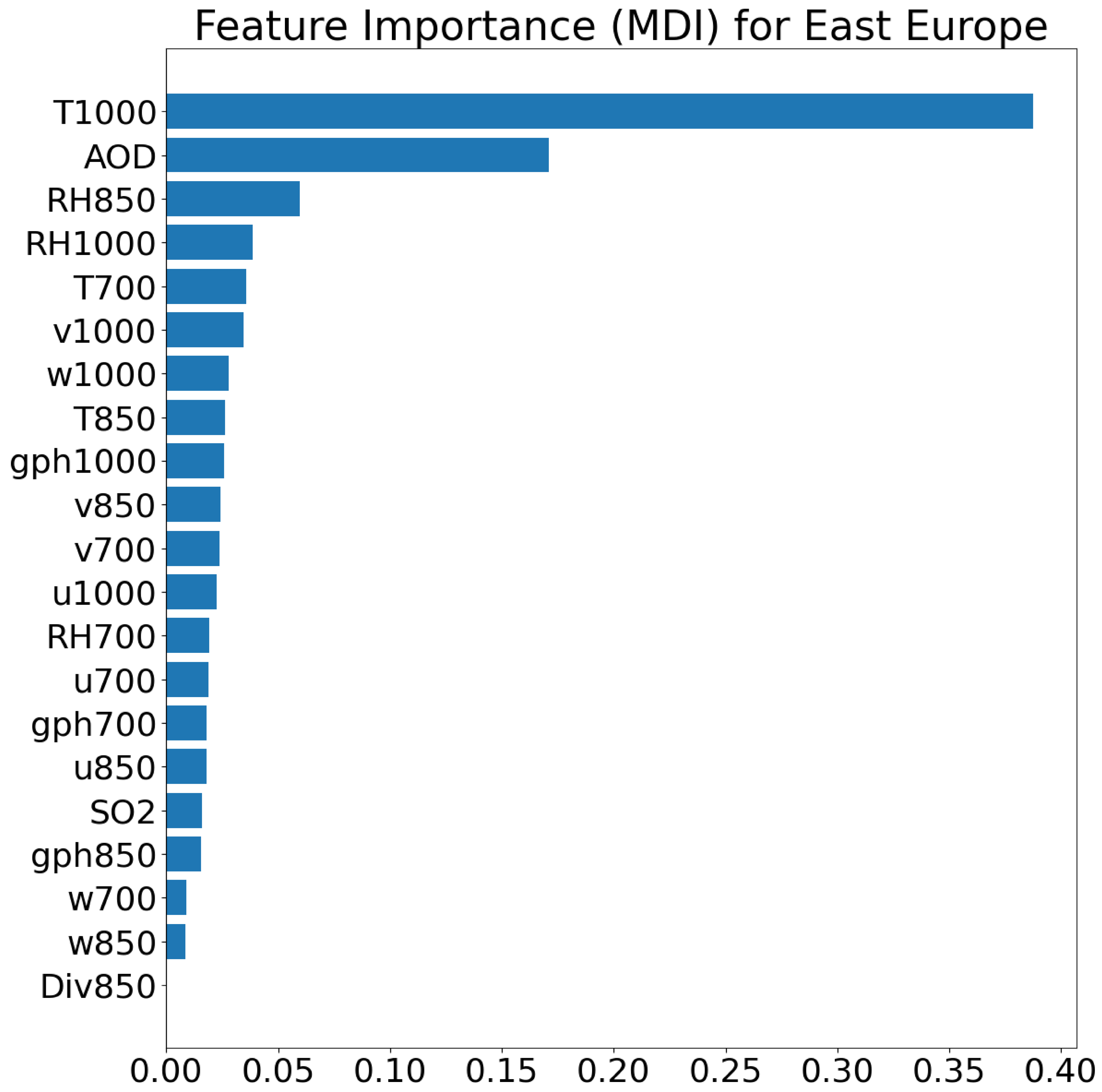
-On each given day, the data points where all data are available are given to the model.

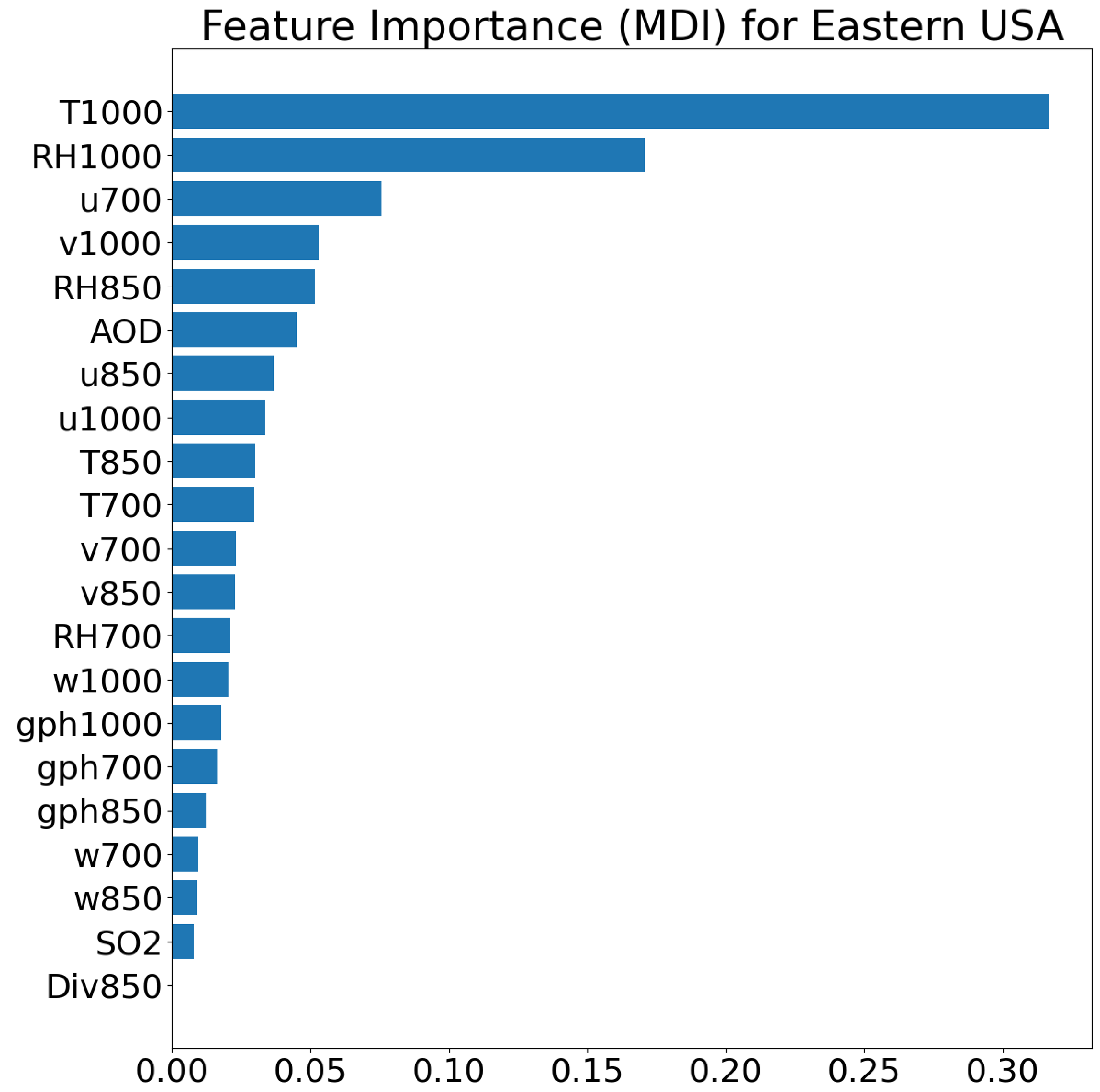
-Meteorological data are approximately synchronised with A-train overpass time.

Parameter Importance

Parameters are ranked depending on their importance in determining model’s output of cloud effective radius.







Over Eastern China (shown above) AOD is the most important.

Over Eastern Europe (shown above) AOD is the 2nd most important.

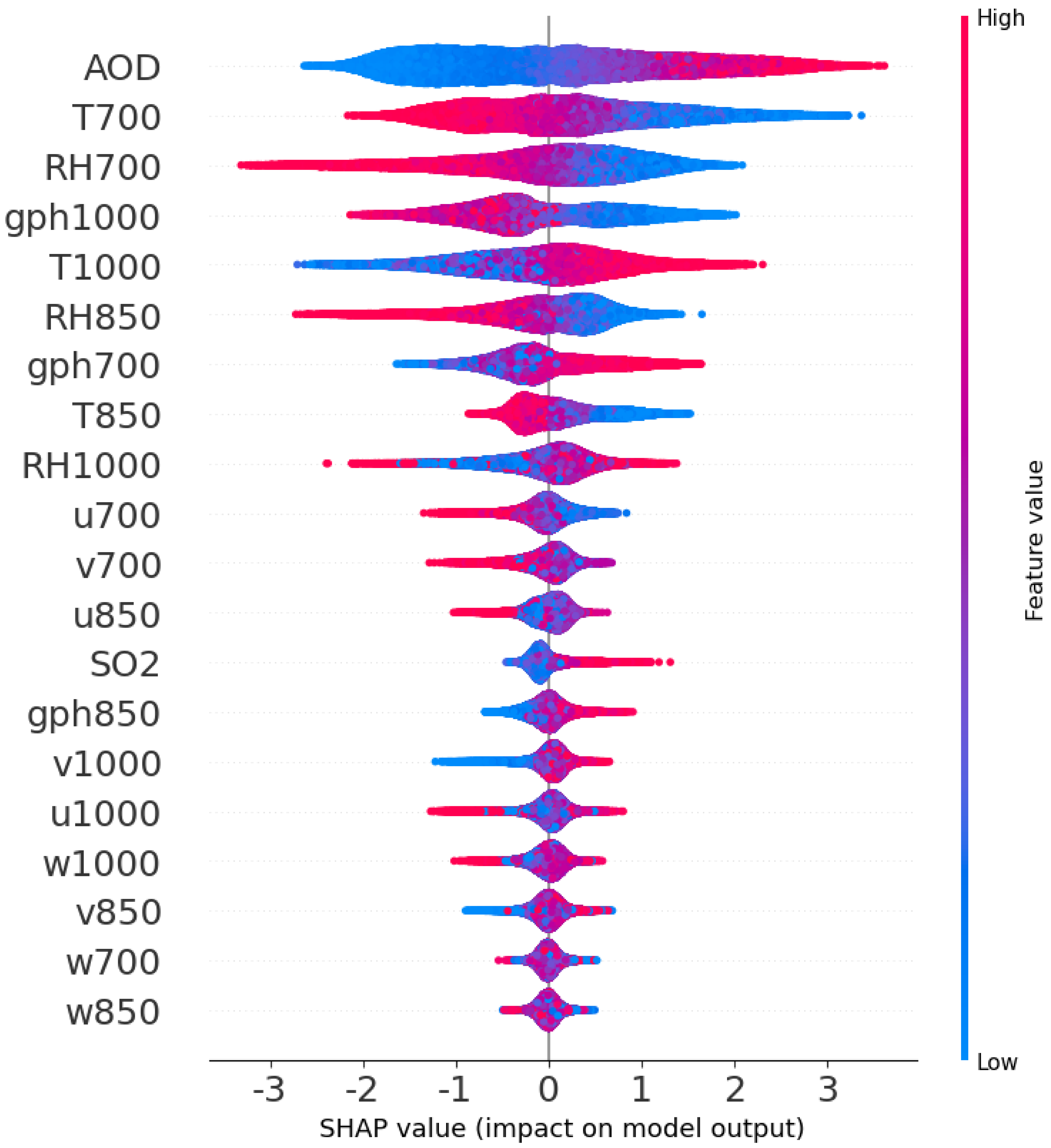
SO2 is not important for any region.

Over Eastern USA (shown above) AOD is the 6th most important.

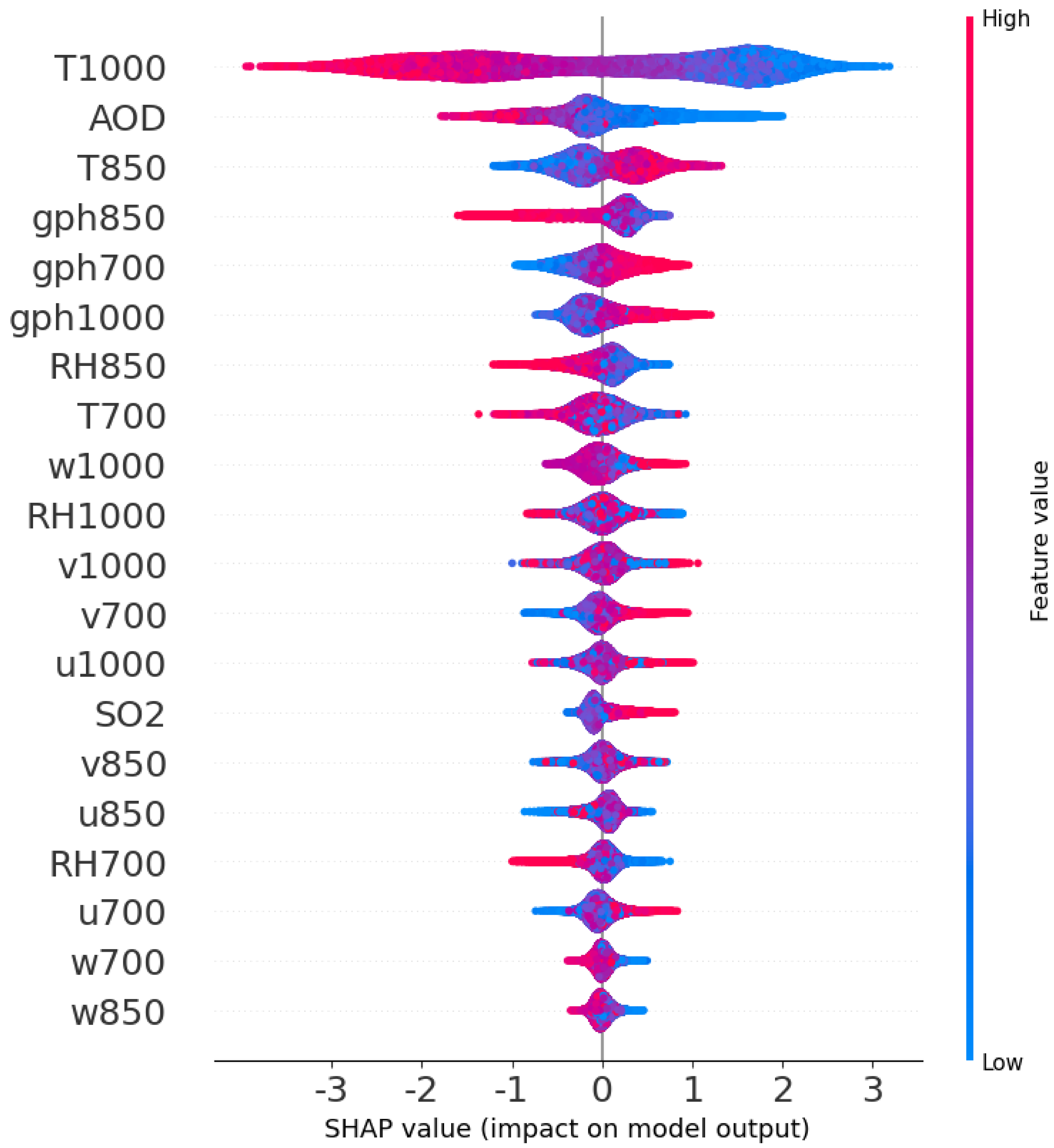
Parameters’ Impact on Model Output

SHAP-value plot: X-value shows the parameters impact on the model output. In color is the relative value of the specific parameter.

Eastern China



Eastern Europe



For Eastern China (shown above), a <u>large</u> AOD value will tend to make the model predict <u>larger</u> values of r_e. This represents a + sign in the table (see next slide).

For Eastern Europe (shown above), a <u>small</u> AOD value will tend to make the model predict <u>larger</u> values of r_e. This represents a - sign in table (see next slide).

Stockholm University

Main Results and Conclusions

Region	AOD rank out of 21	SO2 rank out of 21	RMSE with AOD, SO2 (relative error)	RMSE without AOD, SO2 (relative error)	RMSE Mean Reference model	AOD impact on output
Iceland	10	19	3.59 (22%)	3.65 (22%)	4.84	0
Hawaii	3	17	3.78 (19%)	3.88 (20%)	4.44	-
East China	1	10	4.57 (30%)	4.69 (31%)	5.07	+
East US	6	20	4.03 (28%)	4.08 (28%)	4.78	+
California	5	20	3.96 (29%)	4.03 (30%)	5.10	0
Peru	3	20	3.65 (25%)	3.85 (27%)	5.53	-
West Australia	4	20	3.76 (22%)	3.85 (23%)	5.18	-
Namibia	5	20	3.28 (26%)	3.33 (26%)	4.08	0
Canary	2	20	3.73 (22%)	3.88 (23%)	4.71	+
South East Europe	2	17	4.40 (26%)	4.46 (26%)	4.79	-

-A Gradient Boosting Regression model predicts r_e with scores (RMSE, R^2) better than a mean reference model for all tested regions.

-Model scores are slightly improved when AOD input is included, but SO2 does not improve score.

-RH and T are found to be the most important variables for r_e prediction

-Sign of AOD effect on r_e output varies between regions