Supplementary Figures

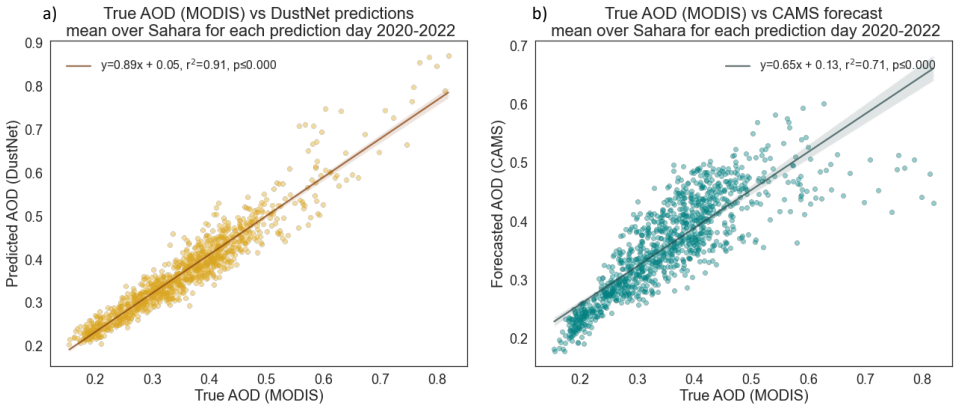

Fig. 5 Fig. A1: Spatially averaged daily AOD (2020-2022, n=1095) regressed between model predictions and MODIS data. Linear regression with corresponding y equation, Pearson's r^2 and p values were calculated for daily spatial mean AOD over the Sahara for 2020 - 2022. Shown in **a)** AOD prediction results from DustNet correspond to MODIS data well with high $r^2 = 0.91$, and only a slight tendency to overestimate higher AOD. In **b)** the mean AOD forecasts from CAMS are shown to correspond with MODIS data well, $r^2 = 0.71$ though, with more frequent tendency to underestimate both low and high AOD values. Results from both predictions are highly significant with $p < 0.0001$.

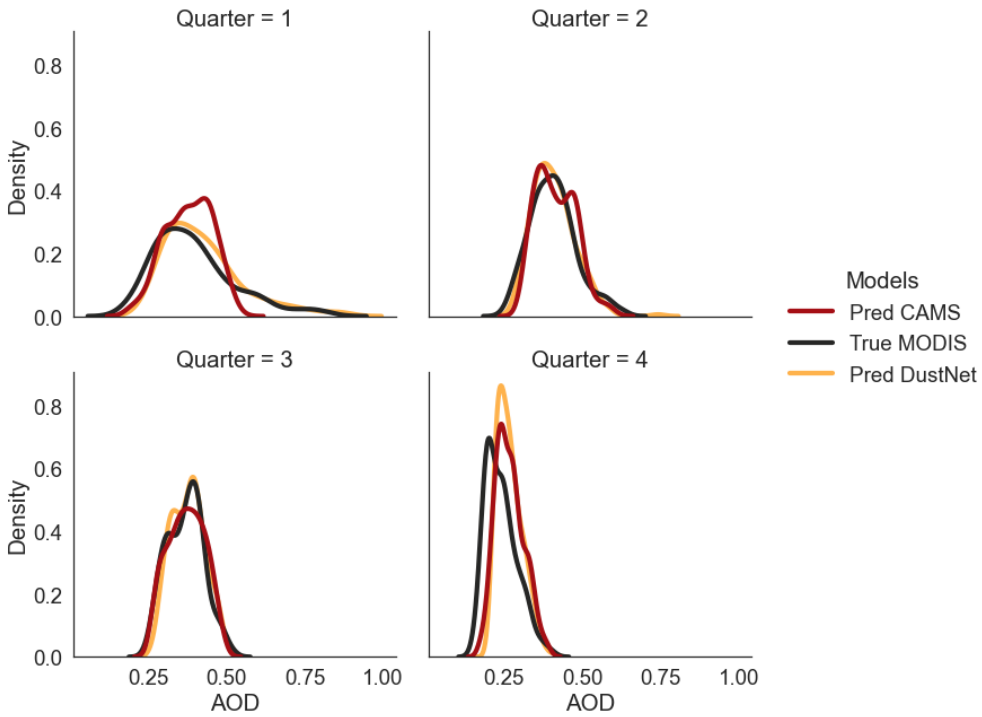

Fig. 6 Fig. A2: Seasonal mean distribution of daily AOD values. The data was averaged over the study region for the testing period of 2020-2022, and shows CAMS forecasts (red) DustNet predictions (yellow) and ground-true MODIS (black). The long tail, indicative of higher AOD values, is clearly missing in the CAMS distribution (red) for Quarter 1: January - March, while the lower AOD values are overestimated. The opposite is true for Quarter 4: October - December, where lower AOD values tend to be underestimated by both CAMS and DustNet in comparison to MODIS. Both models forecast fairly well during Quarter 2 and 3, although DustNet captures the bimodal distribution of AOD in Quarter 3 more skillfully than CAMS.

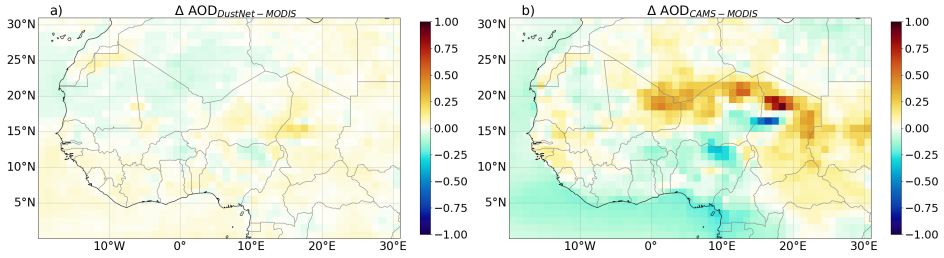

Fig. 7 Fig. A3: Bias of predictions **a)** DustNet and **b)** CAMS with respect to MODIS data. Note that the maximum bias produced by DustNet is 0.21 while the maximum bias for CAMS is 0.93.

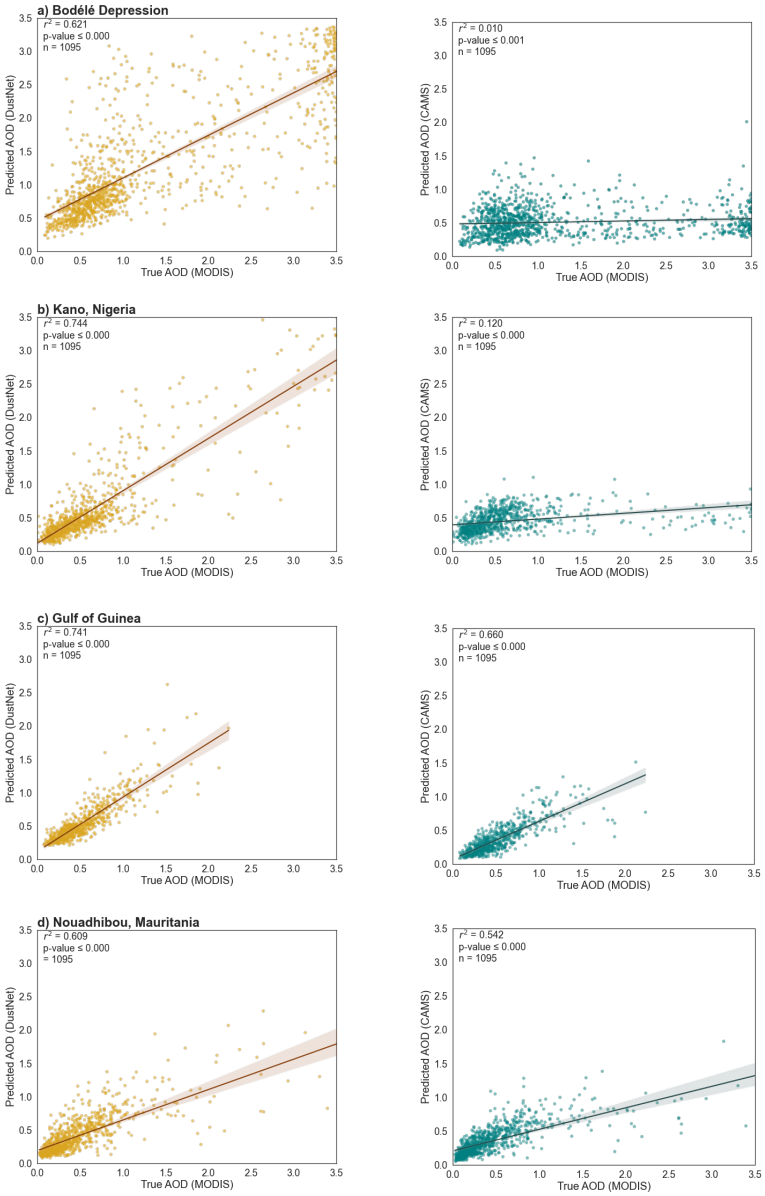

Fig. 8 Fig. A4: Scatter plot relationship between predicted mean AOD values (2020-2022) and MODIS data. Results for DustNet (**left panel**) and forecasts from CAMS (**right panel**) at four chosen locations show better agreement of DustNet predictions with MODIS data at each location. In **a)** the Bodélé Depression, Chad - highest source of dust in the Sahara - DustNet is significantly better than CAMS; **b)** Kano, Nigeria - the second most populous province **c)** Gulf of Guinea - over the ocean; and **d)** Nouadhibou, Mauritania - coastal location.

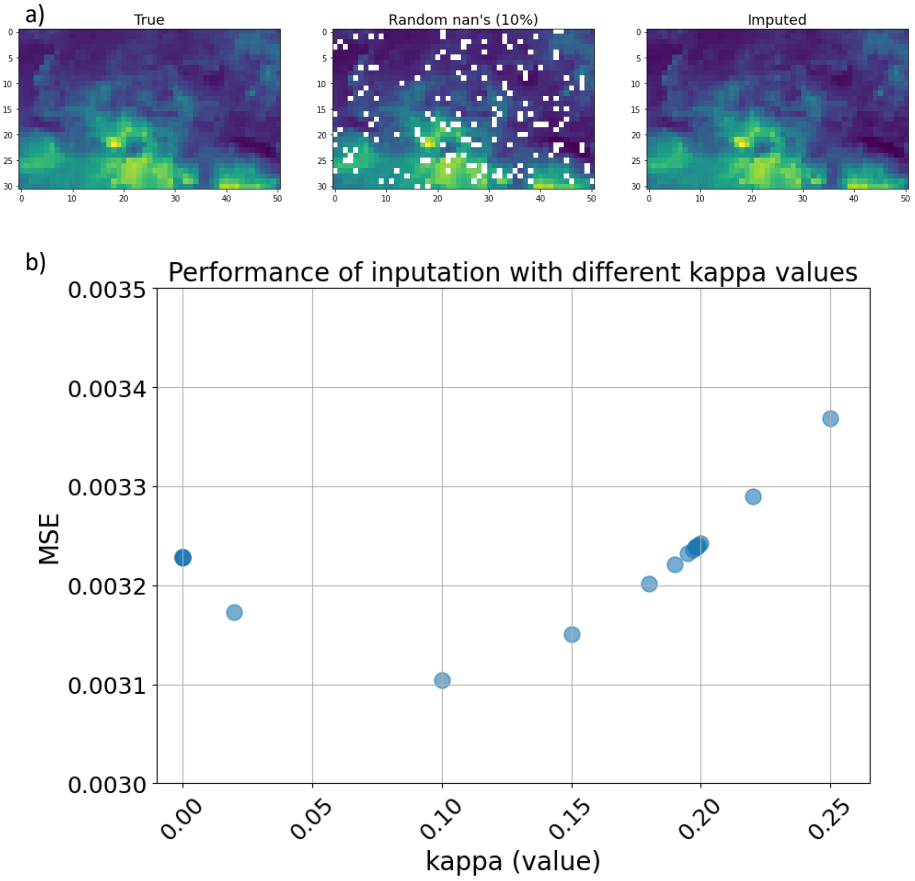

Fig. 9 Fig. A5: Performance validation of imputation code. Three visual inspections of imputation in **a)** show the "true" values on the left, randomly assigned 10% of missing values in the middle, and imputed with Lattice Kriging method values in the right panel. In **b)** the results of MSEs between "true" and imputed values are displayed for different kappa values (Kriging hyperparameter) used during imputation. We tested 24 different kappa values ranging between 0.000001 and 0.1998) The MSE appears to be insensitive to changes in kappa values producing only marginal improvements in the overall MSE.

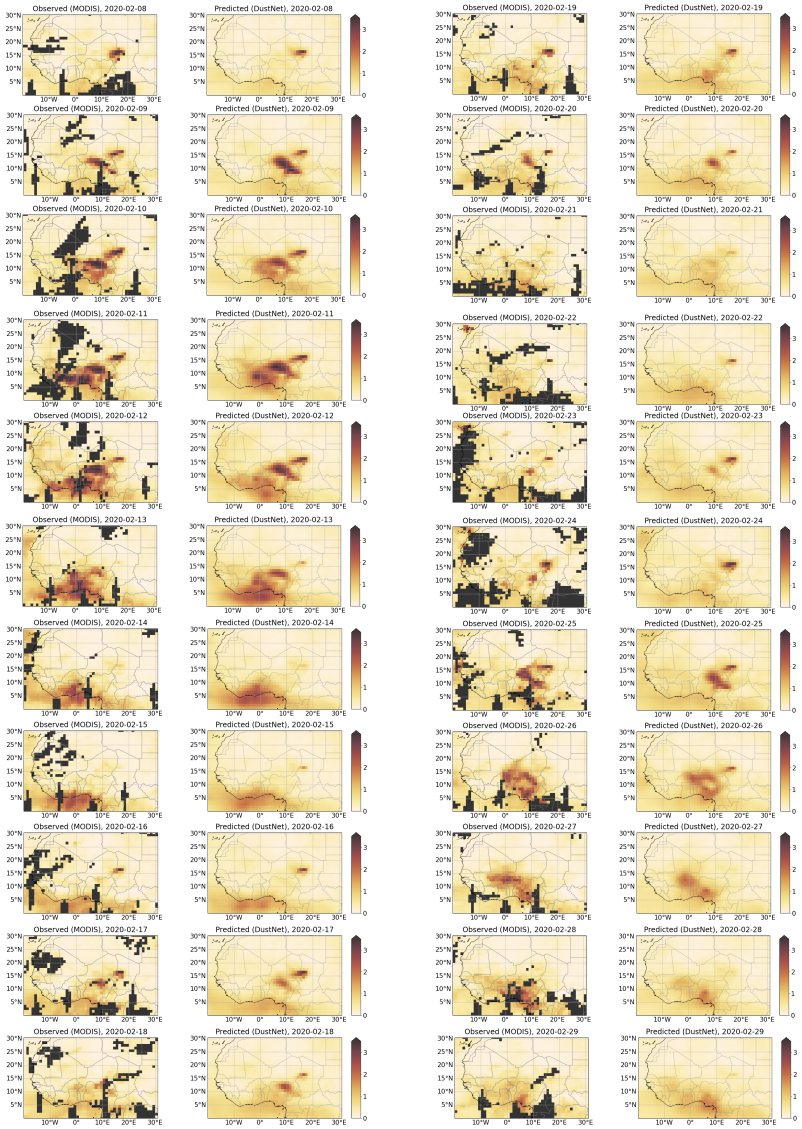

Fig. 10 Fig. A6: Comparison of daily AOD values as observed by MODIS (mean of Aqua and Terra) (**left panel** in both columns) and corresponding DustNet predictions (**right panel**) for selected continuous 3 weeks (22 days), from 8th - 29th February 2020. The dark grey colour in the MODIS maps represents missing values. Despite an initial assumption of heavy reliance on the past 5 days of AOD during training, DustNet presents a skillful ability to predict the next time-step (24-hr) which visibly differs from the last 5 days. This is evident on 13th -14th Feb and 21st Feb, where the AOD values start to decrease despite an increasing past trend. Similarly, prediction of an increasing AOD from 22nd - 26th Feb was captured, despite the previous 5 days of decreasing AOD. The south-western direction of aerosol transport during boreal winter is also skillfully captured (10th - 14th Feb), as is the position of the Bodélé Depression during dust generation episodes (22nd - 26th Feb), but without overly relying on this location as a constant dust source (27th - 29th Feb).

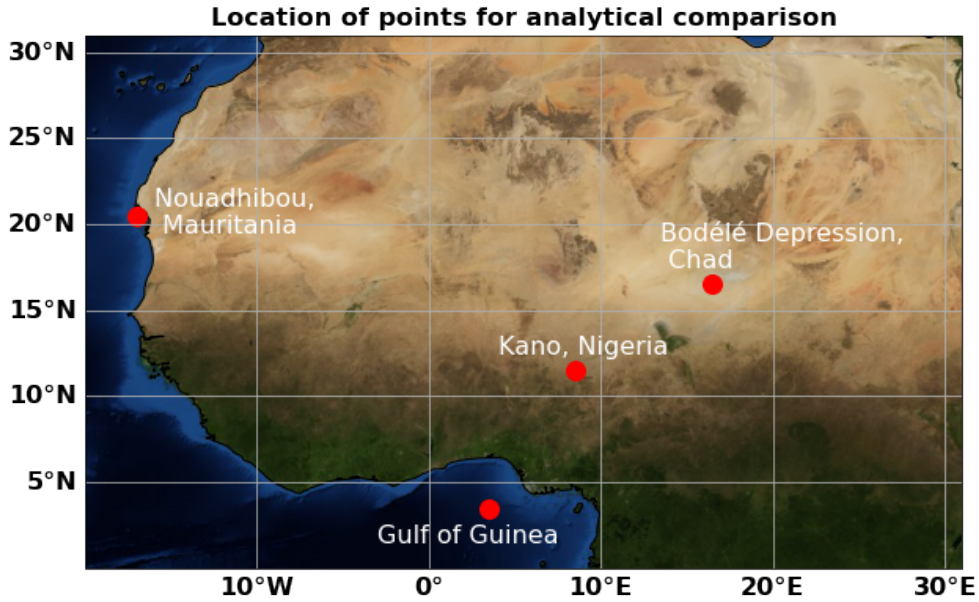

Fig. 11 Fig. A7: Locations of selected grid points used to asses the model's predictive accuracy on a local scale ($1^\circ \times 1^\circ$ grid size). The background image for the December view of Blue Marble is available from NASA https://visibleearth.nasa.gov/collection/1484/blue-marble?page=4

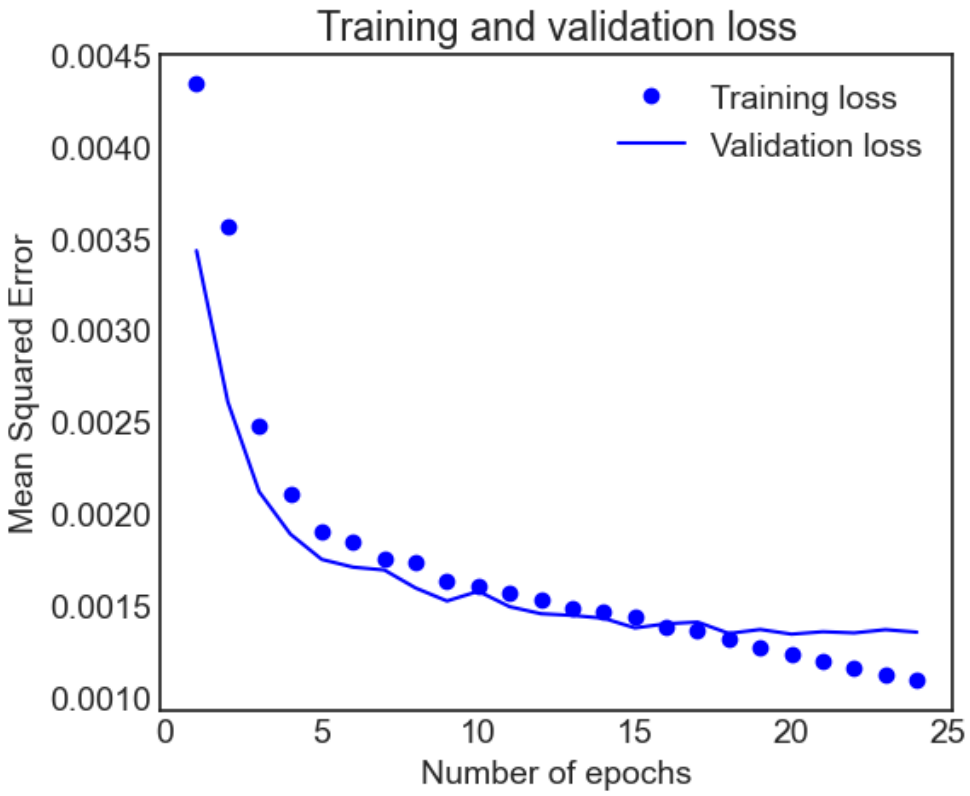

Fig. 12 Fig. A8: Training and validation loss for the optimal model - DustNet. The model's architecture ensures Early Stopping is performed following the 4th iteration without any improvement in validation loss. Here, stopping occurred after 24 epochs and the model with the lowest ratio of training to validation loss was saved and used for predictions.

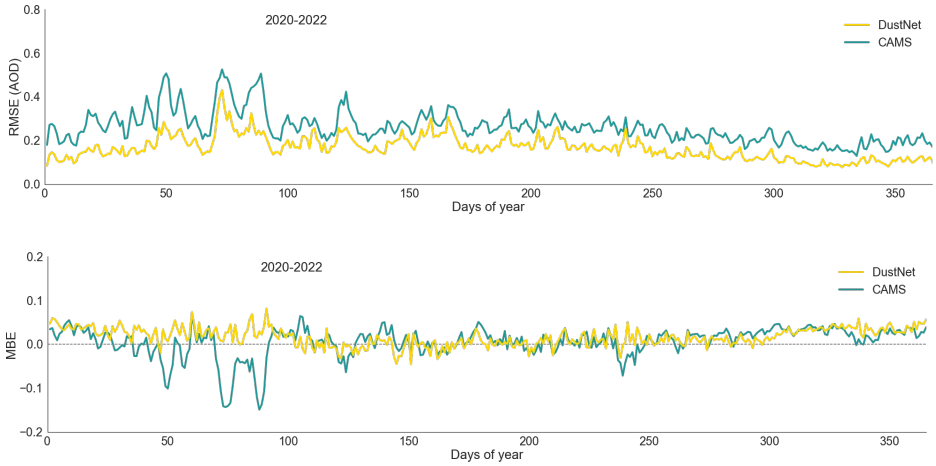

Fig. 13 Fig. A9: In the **top panel** the temporal mean RMSE values for AOD predicted by DustNet (yellow) and CAMS (cyan) are shown. The dashed line represents the climatological mean from MODIS. At all time-steps the DustNet model predictions show smaller (better) errors than those produced by CAMS. The **lower panel** shows the temporal mean bias errors (MBE) from the DustNet predictions (yellow) and CAMS (cyan). Here, the DustNet bias fluctuates close to zero more often than the bias produced by CAMS.

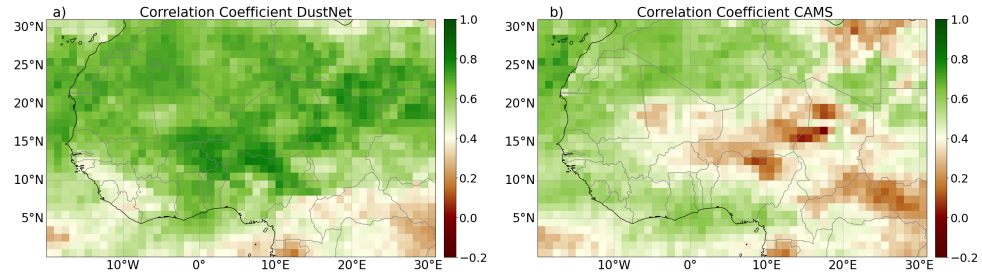

Fig. 14 Fig. A10: Daily MODIS correlation coefficient with predictions from **a)** DustNet and **b)** CAMS. The maximum correlation for DustNet is 0.82 and minimum is 0.16, while the maximum correlation for CAMS is 0.75 and minimum is -0.04. Values with weaker (≤ 0.4) correlation are represented in white to brown, while values with stronger >0.4 correlation are shown in green shades. With predominantly green shades, especially over the Sahara region, the advantage of DustNet predictions is clearly visible.

Represented in Figure 14 is the daily correlation coefficient between the ground true values from MODIS and DustNet predictions in panel a), and between MODIS and CAMS in panel b). The area of the Saharan Desert, where mineral dust is the main contributor to the AOD values, is the stronghold of DustNet-produced predictions. In comparison, CAMS displays a lower (weaker) correlation with MODIS over areas in the Saharan Desert, which were highlighted in previous figures (Fig. 3 and Fig. 10) as dust generation regions. The regions south of 8°N and east of 10°E display weaker correlation for both models. This is not surprising, since DustNet's training regime did not include any data indicative of black carbon or secondary organic aerosols, which seasonally dominate the AOD over the equatorial region of Central African forests ([73]).

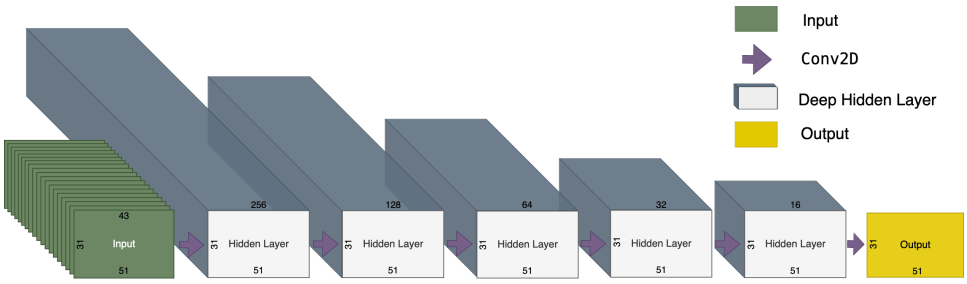

Fig. 15 Fig. A11: Schematic representation of simple Conv2D model. **From left:** the Input Layer of shape (31,51,43) is represented in green. Following, are the 5 Hidden Layers of this same width and height as Inputs, but with different depths. The depths (number of hidden connections) were set in decreasing order to 256, 128, 64, 32 and 16. The last 2D Convolution with depth 1 created the output which shape was matching our target AOD.

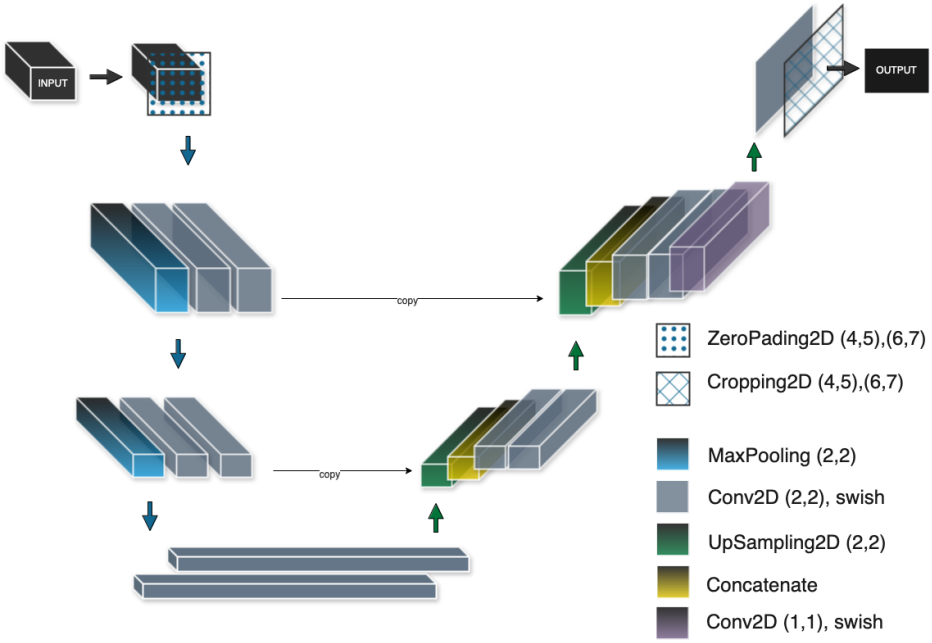

Fig. 16 Fig. A12: Illustrative sketch of U-NET model architecture with individual blocks representing model layers. The input layer (31,51,43) is first padded with 2D zeros layer which increases the height and width of input shape (40,64,43). The encoding pathway (blue arrows down) includes 2 successive layers of Conv2D which increase the depth of the input size ($40 \times 64 \times 64$). Following MaxPooling layer decreases the first two dimensions, while Conv2D increases the third ($20 \times 32 \times 128$). After the second MaxPooling and double Conv2D the input is reshaped to (10×128). The decoding pathway (green arrows up) includes 2D Upsampling and Concatenation which now increases the width, height and depth to ($20 \times 32 \times 384$). Following 2 layers of Conv2D decrease the depth while UpSampling and Concatenation increases the shape to ($40 \times 64 \times 192$). The last two layers of Conv2D decreased the depth to (40, 64, 64) while its final layer brought the depth down to ($40 \times 64 \times 1$). Last layer, Cropping2D ensured the output matched the target size of ($31 \times 51 \times 1$).

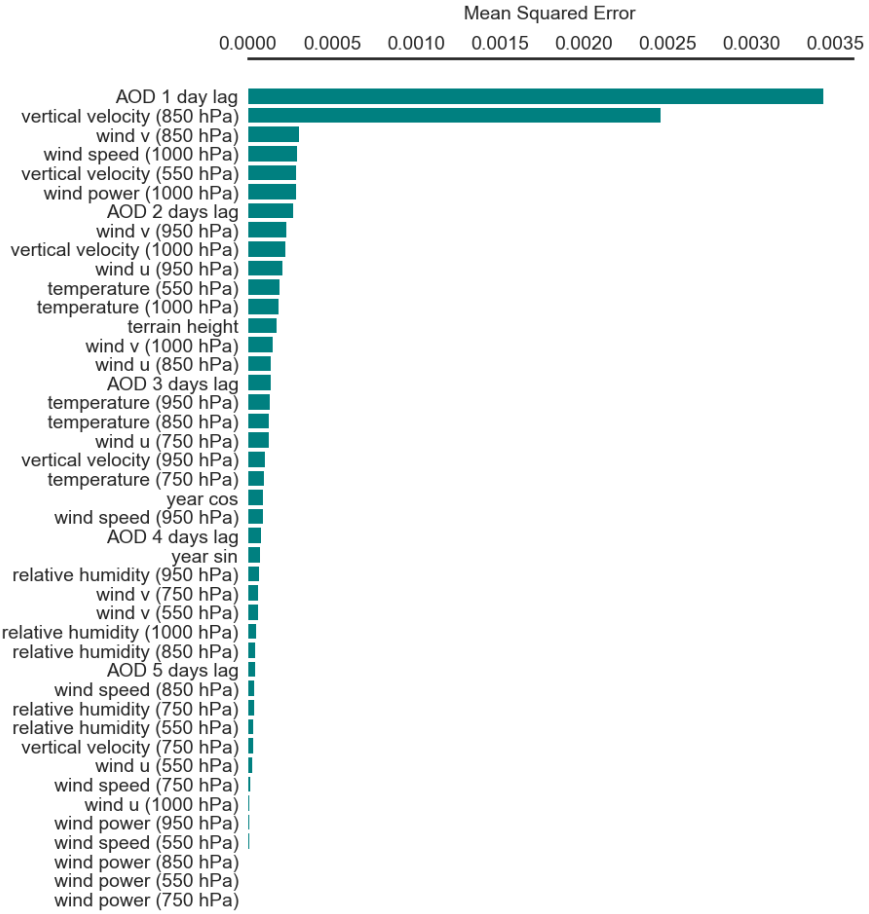

Fig. 17 Fig. A13: DustNet model's feature importance based on MSE. The features with the highest MSE indicates the highest importance for the pre-trained DustNet model in creating 24-hours (1-step ahead) predictions. The bar chart indicates that the model predictions result in highest MSE when the information on 'AOD 1 day lag' feature values are removed (zeroed-out). Similarly, when feature values for 'vertical velocity (850hPa) were removed from prediction run, the model resulted in higher MSE. This suggests that the pre-trained DustNet model relies on 'yesterdays' AOD values and vertical velocity at 850hPa the most during prediction run.

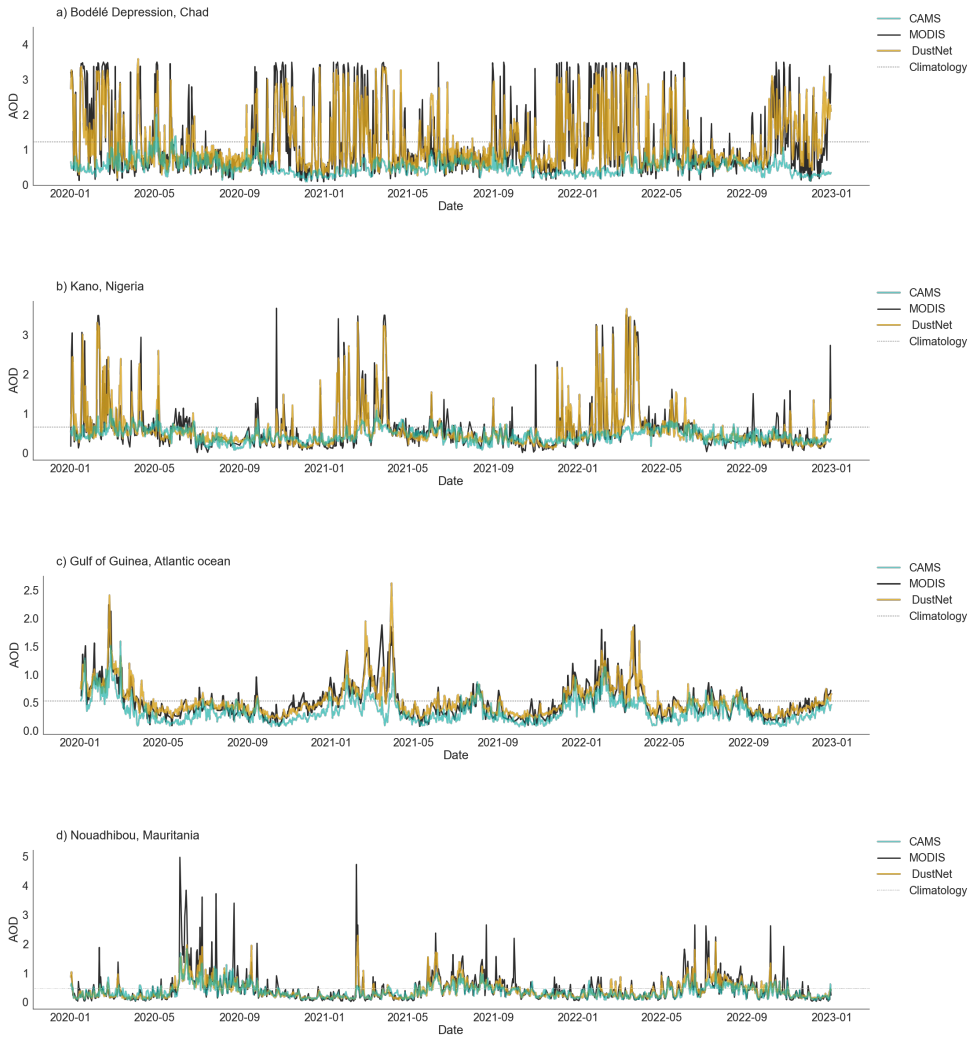

Fig. 18 Fig. A14: Same as Fig. 4, but for daily data.