Towards Downscaling Global AOD with Machine Learning

Josh Millar, Paula Harder, Lilli Freischem, Philipp Weiss, Philip Stier

Motivation

- Poor air quality is a significant health threat, especially in urban areas.
- High-resolution (HR) **Aerosol Optical Depth (AOD)** forecasts are needed as proxy for improve air pollutant mass concentrations.
- Current GCM forecasts of AOD lack spatial resolution = difficult to accurately represent local-scale variability

AOD = column-integrated light extinction over an atmospheric column

Here, we evaluate GCM/reanalysis-scale to local-scale downscaling of global AOD using ML and assess potential for ML bias correction with in-situ AOD observations.

Data

For training, MODIS MOD04_L2 AOD data is used.

Processing

- reprojection: sinusoidal → regular
- gap-filling of missing values
  - mean imputation used
  - post-coarsening

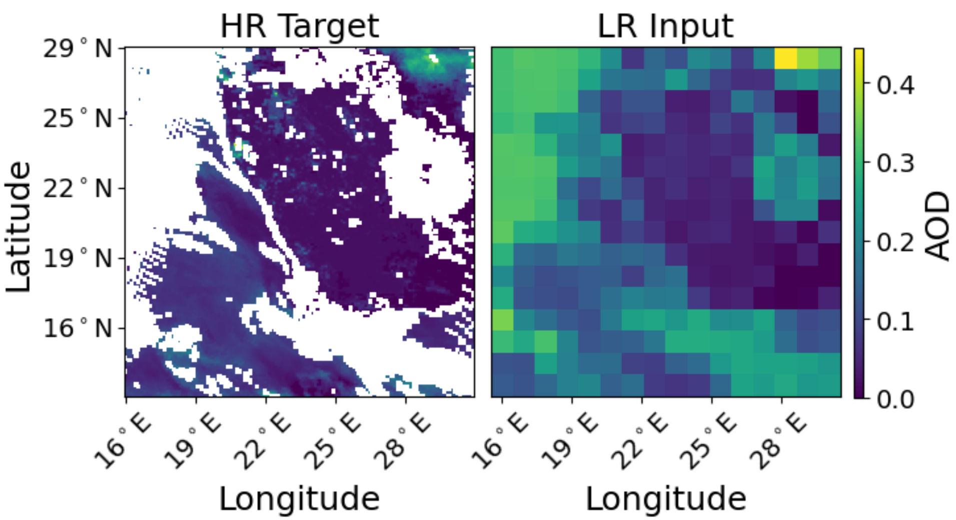

Fig 1: HR reprojected MOD04_L2 AOD image crop alongside its coarsened gap-filled LR equivalent, forming a LR-HR input-target pair.

Data (cont.)

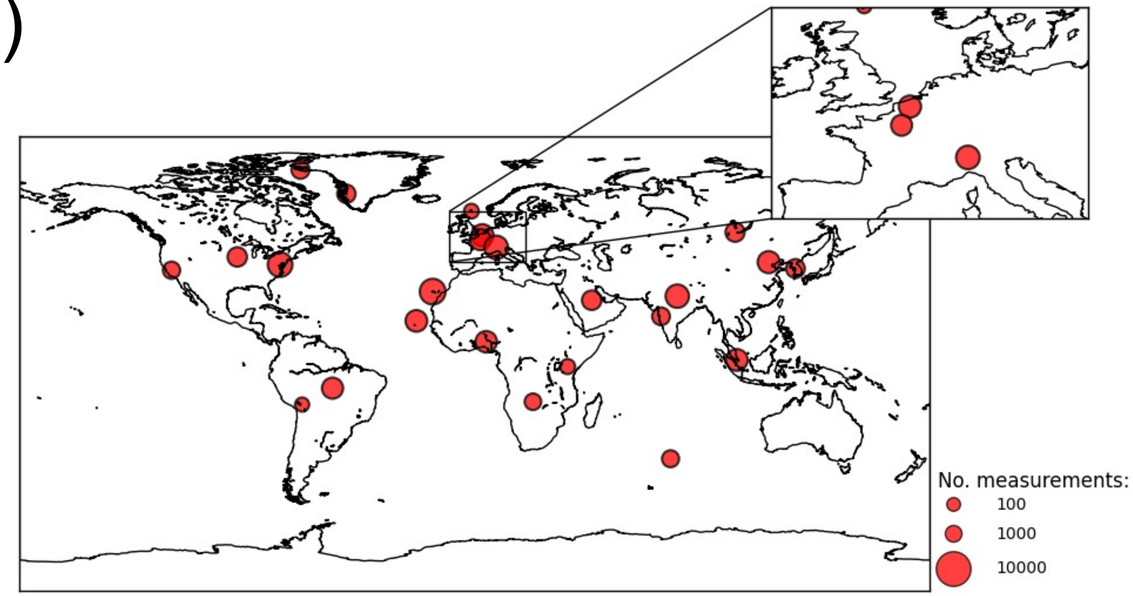

Fig 2: The AERONET stations selected for evaluation.

For assessing bias correction potential, CAMS data is used with AERONET as ‘unbiased’ reference.

- Direct comparison with in-situ AERONET measurements, interpolated to 550nm.
- CAMS data is scale-height adjusted using a fixed 2.1km scale-height.
- Collocation as follows:
    - Spatial: Nearest single point.
    - Temporal: Nearest reading within ±12 hours.

Method

Architecture + Experimental Setup

- progressive upsampling: 5x + 2x → 10x
- maintained ReLU (Rectified Linear Unit) outside ResBlocks
- removed batch norm inside ResBlocks
- custom Mean Squared Error (MSE) loss
- early stopping on validation loss

Baseline: Lanczos interpolation (non-trainable)

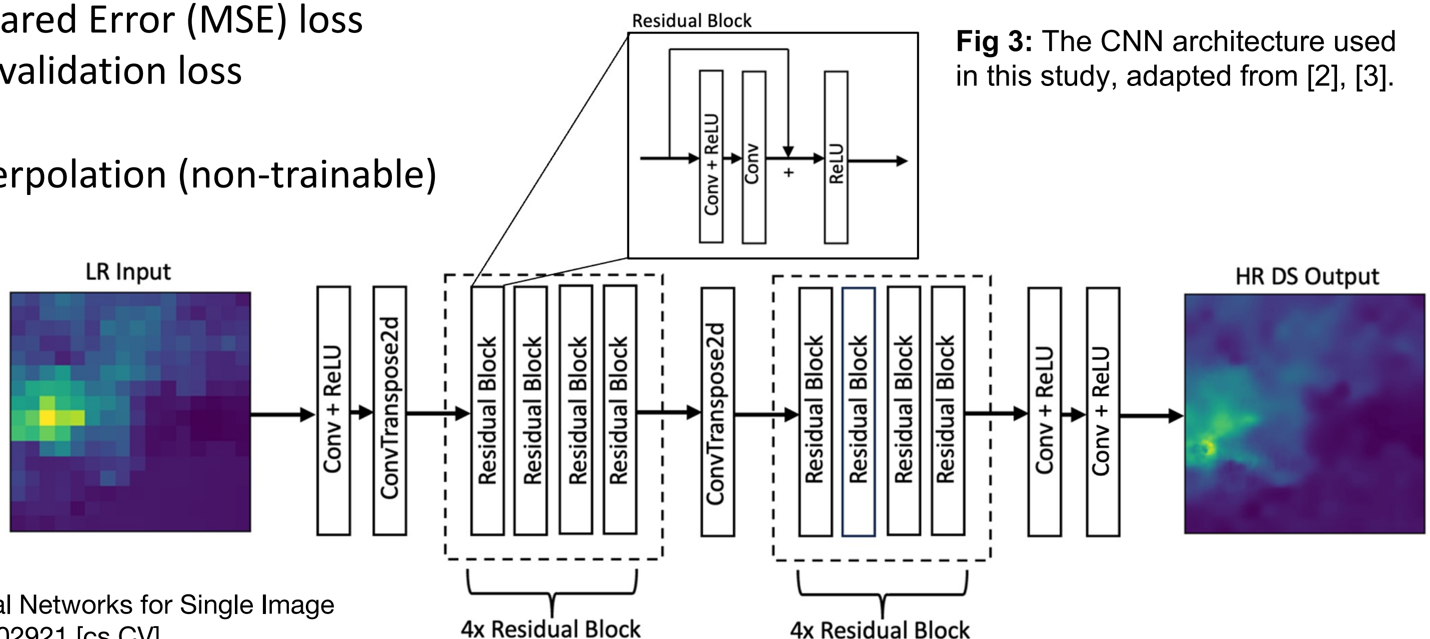

Fig 3: The CNN architecture used in this study, adapted from [2], [3].

[1] Lim et al. Enhanced Deep Residual Networks for Single Image Super-Resolution. 2017. arXiv: 1707.02921 [cs.CV].

[2] Paula Harder et al. Physics-Constrained Deep Learning for Climate Downscaling. 2023. arXiv: 2208.05424 [physics.ao-ph].

ICLR

Imperial College London

UNIVERSITY OF OXFORD

Results

Table 1. Metrics for the baseline and CNN

	MODIS (MOD04_L2)		CAMS		
	HR DS (baseline)	HR DS (CNN)	LR	HR DS (baseline)	HR DS (CNN)
KGE	0.571	**0.939**	0.646	**0.758**	0.738
MSE	0.008	**0.005**	0.075	**0.066**	0.069
NMBE	**0.017**	0.039	0.203	**0.092**	0.098

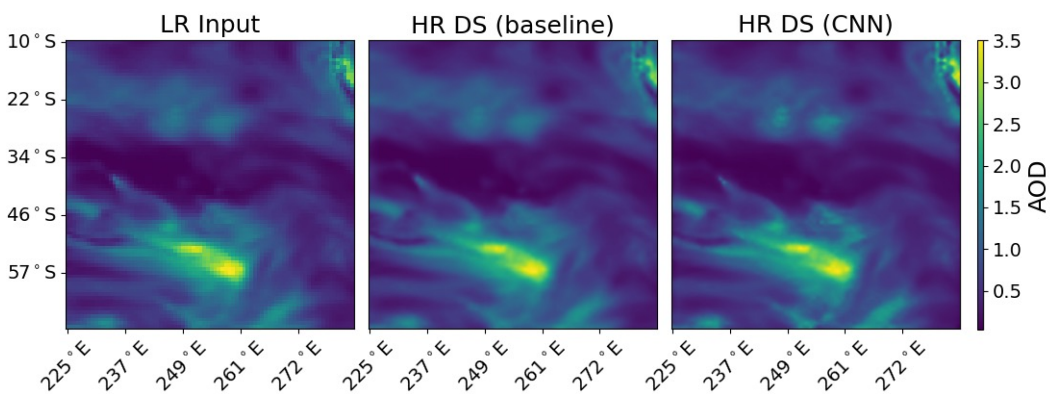

Fig 4. Example outputs, from baseline and ResNet (CNN), on a CAMS image.

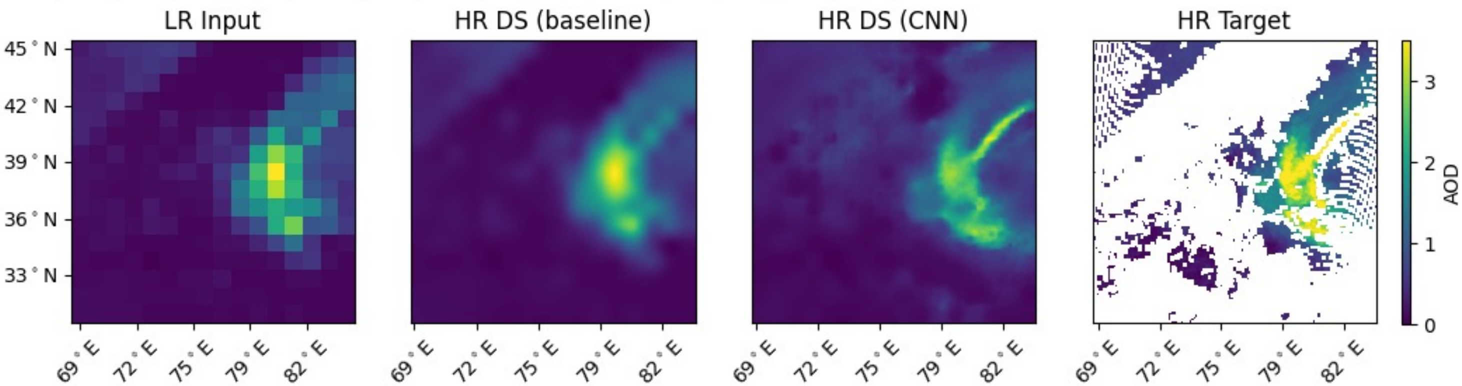

Fig 5. Example outputs, from baseline and ResNet (CNN), on a MOD04_L2 image.

Conclusions

Results

- outperforms interpolation baseline on MODIS AOD
- struggles with bias correction on CAMS reanalysis

Limitations

- gap-filling introduces potential bias, or bias inherent in MODIS data
- wavelength shift from AERONET 550nm to nearest-available

Future Work

- incorporate physical constraints / enforce conservation laws
- temporal regularization via recurrence mechanisms
- unsupervised fine-tuning on reanalysis/GCM data
    - could help mitigate distribution shift issues

Thank you!