Simulation of the aerosol size distribution using a neural network surrogate for the Modal Aerosol Module (MAM7)

Katherine H. Breen[1,2], Donifan Barahona[1], Anton Darmenov[1], Matthew Thompson[3]
[1]NASA Goddard Space Flight Center, Greenbelt, MD, United States, [2]Universities Space Research Association, Columbia, MD, USA, [3]Science Systems and Applications, Inc., Lanham, MD, USA

GMAO
Global Modeling & Assimilation Office

Abstract

One objective of atmospheric simulations is to quantify the distribution of aerosols and their properties. Accurate parameterizations of the processes governing aerosol mass, particle number, and particle size distribution are important for predicting the Earth's net radiative balance and aerosol-cloud interactions. The Modal Aerosol Module (MAM7) is a two-moment aerosol model that simulates mass, number, and size distribution of seven modes comprised of internally mixed aerosol species. The two-moment scheme adds significant computational expense but allows for the prediction of varying particle size distribution relative to the bulk method which predicts only total mass. In this work, we developed a neural network surrogate model for MAM7 (MAMnet) to predict the aerosol number concentration in NASA's Global Earth Observing System (GEOS) without adding prohibitive computational expense. MAMnet, can be driven by output from a single moment, mass-based, aerosol scheme (Goddard Chemistry Aerosol and Radiation model (GOCART)) or from reanalysis products (Modern-Era Retrospective analysis for Research and Applications, Version 2 (MERRA-2)). MAMnet was trained using number concentrations from a 5-year GEOS/MAM7 simulation at 1-degree horizontal resolution and using the total mass calculated across modes as inputs, as well as temperature and air density. The model architecture for MAMnet was based on AlexNet, the 2012 winner of the ImageNet Large Scale Visual Recognition Challenge. While some modifications were necessary to accommodate our problem, important aspects of the network were preserved. MAMnet was able to reproduce zonal dynamics and spatial distributions of the aerosol number concentration however predictability in the upper troposphere was poor.

How are aerosols represented in GEOS?

The Goddard Earth Observing System (GEOS) model consists of a set of components that numerically represent different aspects of the Earth system. The modal aerosol module (MAM7) was developed to simulate aerosol size distribution and number concentration, but at a high computational cost.

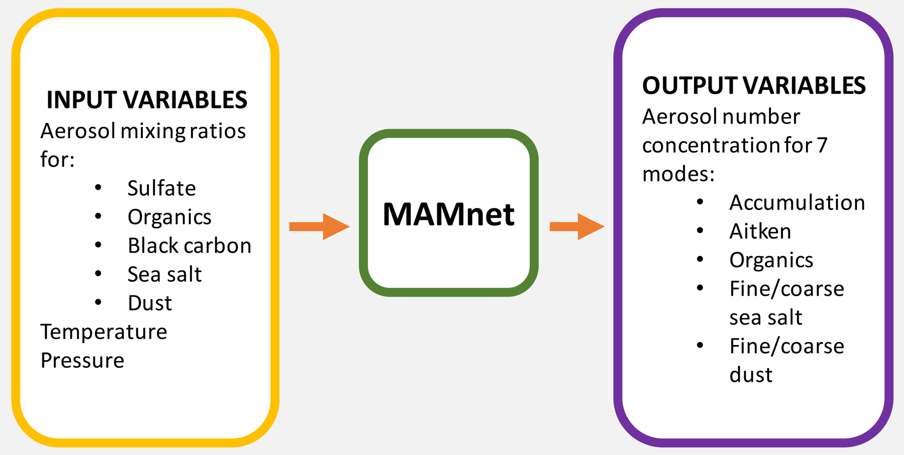
Table 1: Comparison of the Goddard Chemistry Aerosol and Radiation model (GOCART) and the Modal Aersol Module (MAM)

	GOCART	MAM
Approach	Bulk – mass-based, single moment aerosol model, uses only total mass. All particles of the same size have the same composition.	Modal - two-moment aerosol model (predicts mass and number, infer size distribution per species)
Aerosol representation	Externally mixed (all species separate, all particles of the same size have same composition)	Internally mixed (multiple species in the same mode, more realistic)
Computational expense	Relatively low - **operational**	**High**
Notes	Comparable with assimilated aerosols (consistent with MERRA2)	Transport of additional tracers adds computational expense BUT **predicts particle size distribution – important for predicting radiation and aerosol-cloud interactions**

The AGCM configuration of GEOS+MAM was used to develop a robust training dataset for a neural network (NN) surrogate model. We ran a 5-year GEOS+MAM simulation at 1-degree horizontal resolution with 72 vertical levels, generating ~1M samples per grid cell at 3-hourly temporal resolution.

Datasets

Train/validate MAMnet – GEOS+MAM data (~90%)



Why are aerosols important?

Accurate simulations of atmospheric chemistry require knowing the concentration of chemical species (aerosols) in the atmosphere

For a given species X:

- The **mixing ratio** (Cx) of X is the mass fraction of X (mass of X per mass of air)
- The **bulk number concentration** (Nx) of X is the number of molecules per unit volume of air.

$Nx = Cx * Na \Rightarrow Nx = g(Cx)$

We want to know g(Cx) for a size distribution!

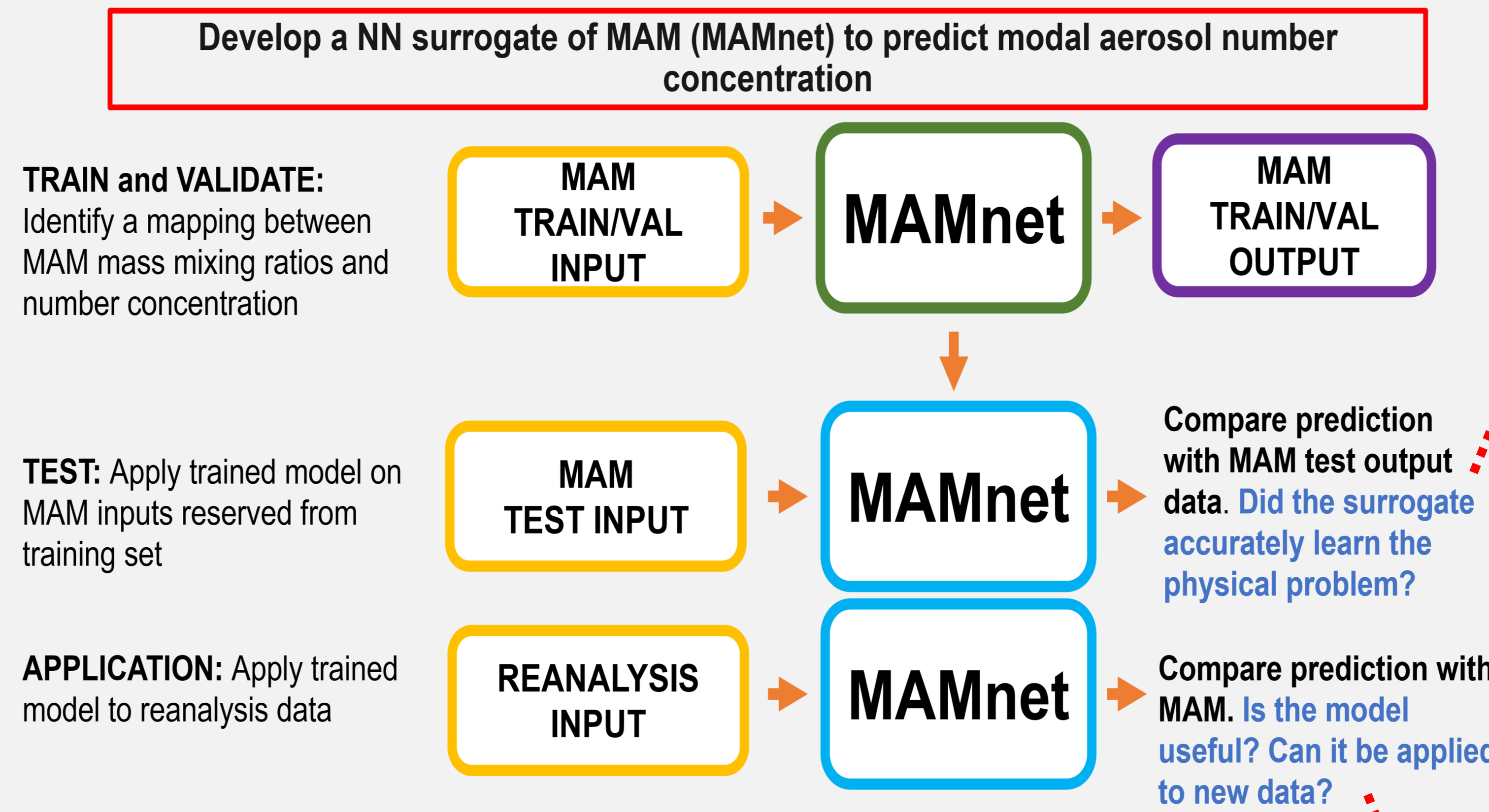
Where Na is the total number of molecules per unit volume air

Test MAMnet

Input: Reserved GEOS+MAM test set (~10%), GOCART, NASA reanalysis datasets (MERRA-2)
Compare MAMnet output with: Reserved GEOS+MAM test set, observational aerosol number concentration

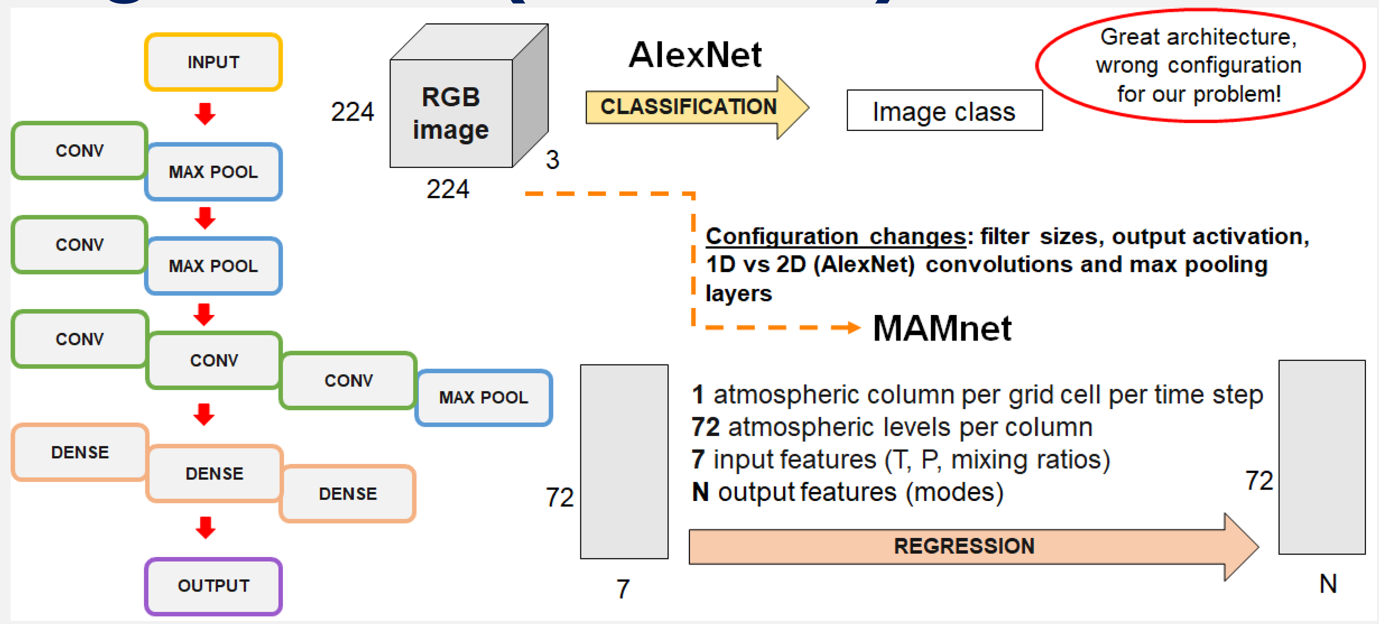
Observational datasets for aerosol number concentration were developed by:
Asmi, A., et al. "Number size distributions and seasonality of submicron particles in Europe 2008–2009." *Atmospheric Chemistry and Physics* 11.11 (2011): 5505-5538.

Objectives



AlexNet adaptation for regression (MAMnet)

AlexNet citation:
Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, *25*, 1097-1105.



Results and Summary

Did the MAMnet accurately learn the physical problem? YES

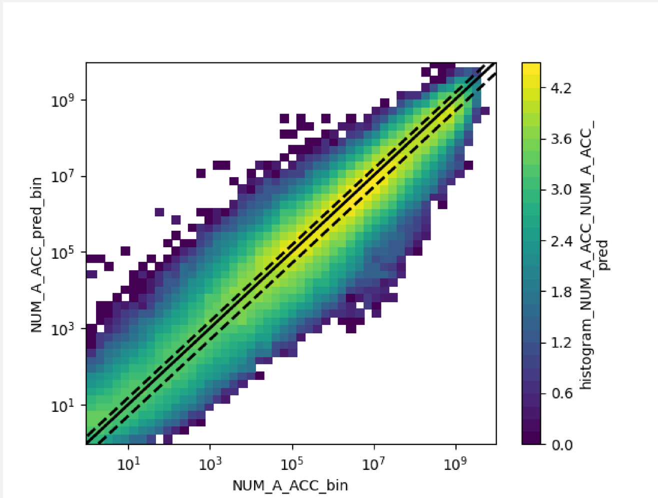


Figure 1. Histogram of GEOS+MAM test data ("true", x axis) vs. MAMnet prediction for aerosol number concentration in accumulation mode.

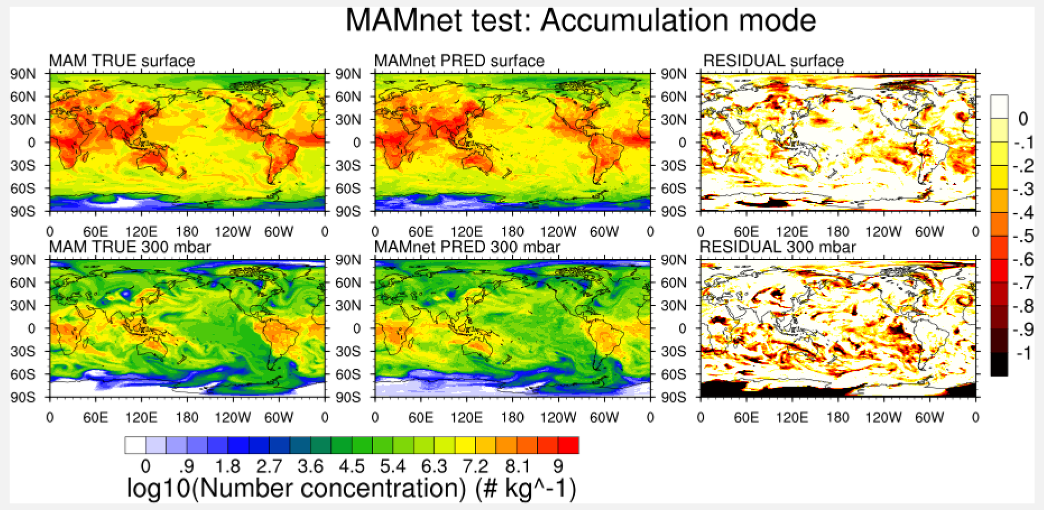


Figure 2. True and predicted data are aerosol number concentration in accumulation mode. Left: GEOS+MAM reserved test data ("true"). Middle: MAMnet prediction on GEOS+MAM test data. Right: Residual (true-pred). Top: surface. Bottom: Upper troposphere.

Prior to MAMnet training, ~10% of the data were reserved for testing. Figures 1 and 2 show predictions made by the trained MAMnet model applied to the reserved test data.

Overall, MAMnet was able to reproduce the test data (not used during training/validation), indicating that MAMnet was trained on a dataset that encapsulated the statistical variability of the system and "learned" patterns from data as opposed to overfitting. Additionally, this test suggests that the AlexNet architecture was successfully adapted for a regression task on atmospheric data.

Is the model useful – can it be applied to new data? YES

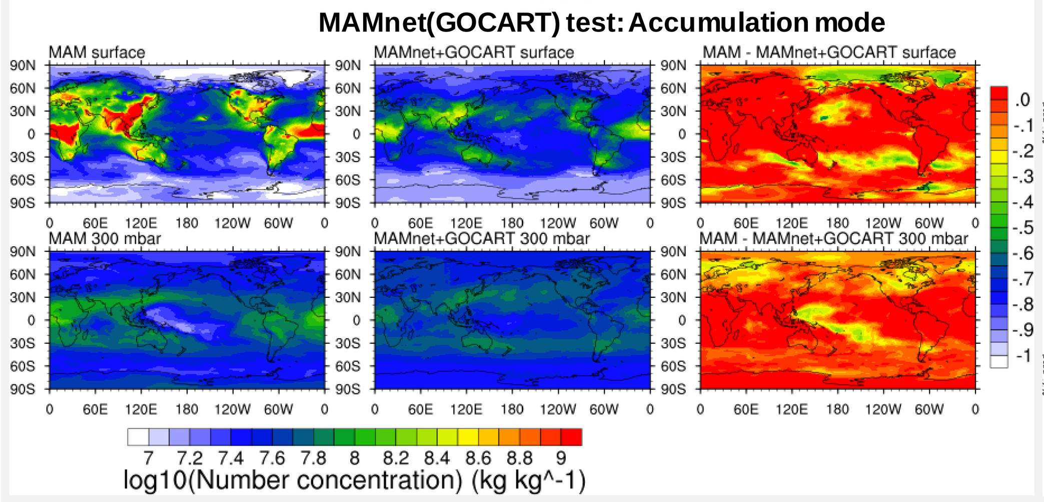


Figure 3. Left: GEOS+MAM reserved test data ("true"). Middle: MAMnet prediction on GOCART inputs. Right: Residual (true-pred). GOCART is the operational aerosol number concentration module in GEOS.

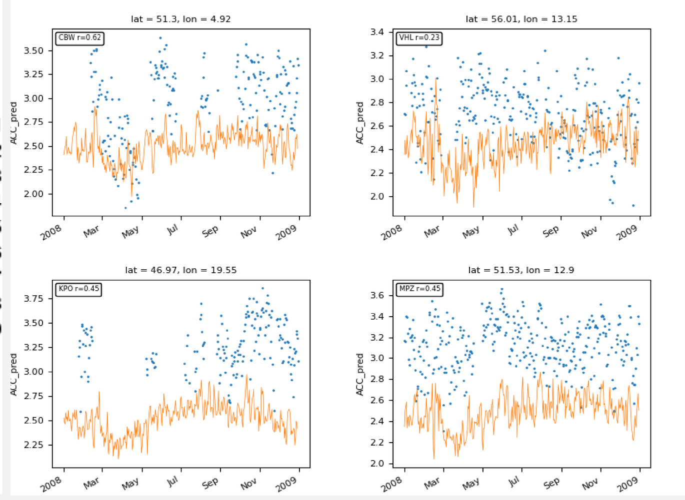


Figure 4. Trained MAMnet model run on reanalysis data (MERRA2; orange) compared with near-surface observational datasets (blue). See Asmi et. Al. (2011) for site descriptions.

After testing MAMnet on reserved GEOS+MAM data, MAMnet was applied to GOCART inputs. GOCART is the operational aerosol module in GEOS but neglects spatiotemporal variability in the aerosol size distribution (see Table 1). Figure 3 shows a comparison between the GEOS+MAM test set (left) and the MAMnet(GOCART) prediction (middle). As a final test, we applied MAMnet to MERRA-2 reanalysis data, which are used to drive GEOS, and compared the output with observational data (Figure 4). In some cases, correlation was high (r>0.6), but further improvements are necessary. We are currently working with an additional probabilistic framework to assimilate observations during training.

These tests indicate that inputs from a low fidelity model (GOCART) were successfully used to replicate the output of a high fidelity model (MAM7) using a surrogate neural network (Fig. 3).

Lastly, while MAMnet is able to reproduce observational data with some degree of accuracy, more development is necessary (Fig. 4).

E-mail: Katherine.h.breen@nasa.gov | *Web:* gmao.gsfc.nasa.gov

National Aeronautics and Space Administration