Deep Learning Ensembles for Improved Atmospheric Composition Modeling

Jennifer Sleeman[1], Christoph A. Keller[2,3], Christopher Ribaudo[1], David Chung[1], Mimi Szeto[1]
[1] The Johns Hopkins University Applied Physics Laboratory
11100 Johns Hopkins Road
Laurel, MD 20723
[2] Morgan State University, Baltimore, MD 21251
[3] NASA Goddard Space Flight Center,
Global Modeling and Assimilation Office, Code 610.1,
Greenbelt, MD 20771

Abstract

Better forecasting of atmospheric composition is a critical aspect of environmental and climate monitoring. Among climate and weather numeric modeling, often ensembles are used to improve the forecasting power and to quantify the uncertainty of the model. However, the numerical simulation of atmospheric chemistry, critical for composition simulations, is computationally too expensive to generate numerical composition ensembles. One way to address this problem is to use deep learning to emulate the slow physical model. In this work we study the feasibility of two different deep learning methods and show how an emulator could be used to realistically estimate uncertainties of atmospheric composition forecasts, bypassing the need to run costly numerical ensemble simulations. One of the methods builds upon Fourier neural operators and the NVIDIA FourCastNet architecture and the second method builds on conditional Generative Adversarial Networks. We design the models to respond to perturbations to the most important drivers of air pollution, including meteorology and pollutant emissions. We apply this framework to the NASA GEOS Composition Forecast System (GEOS-CF), which produces daily global composition forecasts at approximately 25 km^2 horizontal resolution. Due to computational constraints, GEOS-CF currently has limited capability to produce probabilistic estimates or to optimally assimilate trace gas observations. We show how a deep learning emulator has the potential to improve composition forecasts produced by GEOS-CF or other, similar types of applications. These methods could be applied to other types of ensemble-based models, potentially providing a large speed-up in overall modeling time.

Introduction

Poor air quality has strong implications on human health and can exacerbate the impact of a changing climate (Fuller et al. 2022). In turn, warming of the atmosphere could increase extreme air quality events (Shukla et al. 2019). Understanding the role of anthropogenic emissions and its effects on the atmosphere and the climate involves understanding the atmospheric chemistry. Atmospheric chemistry models are a central tool to understand, predict, and mitigate environmental problems such as air quality degradation, stratospheric ozone loss, and ecosystem damage. Chemistry models are computationally very expensive as they need to capture the distribution of hundreds of stiffly coupled chemical species, along with the physical dynamics of the atmosphere. Because of this, chemistry models are orders of magnitude slower than weather models that do not include chemistry. The high computational cost of atmospheric chemistry makes it challenging to simulate atmospheric composition in near real-time. Due to this high computational cost, ensembles are difficult to run and therefore uncertainty estimation is in many cases unobtainable.

Deep learning could potentially overcome this computational burden by providing a way to emulate atmospheric chemistry models. In this study, we asked the question, can we build a deep learning model that can learn to forecast atmospheric chemistry concentrations up to ten days into the future, using just a few timesteps of information from a set of initialized ensemble members. If we could achieve this goal, without a significant loss in forecast skill, the performance savings could be immense.

However, there are still many unanswered questions related to building a deep learning model that could learn to forecast days into the future for a set of ensemble members (given only a few timesteps of input). Specifically, we need to better understand the role of emissions and meteorology on the forecast skill of atmospheric chemical concentrations.

In this paper, we begin to answer these questions using two different deep learning models that are based on well-researched methods for weather forecasting. We emulate the NASA GEOS Composition Forecast System (GEOS-CF) as a use case atmospheric chemistry modeling system.

Background

The NASA GEOS Composition Forecast System (GEOS-CF) combines the NASA GEOS Earth System Model with the GEOS-Chem chemistry module to produce daily analyses and 5-day forecast of atmospheric composition at approximately $25x25\,km^2$ (Keller, Knowland et al. 2021). GEOS-CF predicts the spatiotemporal evolution of more than 200 chemical species, including all major air pollutants such as ozone, nitrogen dioxide, and fine particulate matter (Keller, Knowland et al. 2021). The GEOS-CF system is much slower than a comparable weather forecasting model and takes $6-8$ hours to complete one full compute cycle, consisting of a one-day analysis (constrained by me-

teorological and composition observations) and a 5-day prediction initialized from the one-day analysis. Because of the high compute requirements, the GEOS-CF system currently does not include ensemble predictions or state-of-the art data assimilation methodologies, as commonly used for weather prediction.

Related Work

Early work by Debry et al. (Debry and Mallet 2014) used machine learning for learning an ensemble for the Prev'Air operational platform using weighting and a ridge regression approach. This was geared towards an operational environment which includes the incorporation of observational information.

In work by Bihlo et al. (Bihlo 2021), a conditional Generative Adversarial Network (cGAN) architecture was used, described as a vid2vid model which was built based on the pix2pix architecture (Isola, Zhu et al. 2017). The Generator was described as a U-Net architecture with a bottleneck of Long Short-Term Memory (LSTM) networks. ERA5 Reanalysis data was used to train the model and a Monte Carlo dropout approach was used after the cGAN was trained to simulate an ensemble. The ensemble was used to forecast different weather variables and showed promising results on variable prediction tasks except for total precipitation. They ran an ensemble of 100 realizations of the trained model for atmospheric parameters that corresponded to the ERA5 variables - 500 hPa geopotential height, two-meter temperature and total precipitation. They computed the ensemble mean and standard deviation for the ensemble for each variable.

The cGAN method was extended by Brecht and Bihlo (Brecht and Bihlo 2022) using a similar architecture from Bihlo's previous work, except in this work they explored training the model using ECMWF IFS operational ensemble data with a spatial resolution 0.5 x 0.5. They trained the deep learning method to learn the statistical properties of the ensemble and the spread given the control forecast.

FourCastNet (Pathak, Subramanian et al. 2022) is a deep neural network that learns to forecast 2D atmospheric variables globally. It uses a Fourier Neural Operator (FNO) combined with a vision transformer architecture and claims strong performance for modeling complex PDE systems. FourCastNet was trained to forecast weather variables using the ERA5 Reanalysis data. Ensembling was done by randomly initializing conditions.

In these later methods, ensembling was performed by a random process. However, the initial conditions for the ensemble members of many models are defined based a significant amount of scientific rigor, and domain knowledge. Randomness may not sufficiently represent this knowledge and rigor.

Emulating Ensembles Using Deep Learning

We define an ensemble as M and an ensemble member as m where $m_{0..n} \in M$ and n represents the number of total members in the ensemble. Each m is an atmospheric chemistry model and initialized according to domain knowledge, with both emissions and meteorology present in the initial conditions. A subset of m, the training members is defined by tm and a non-overlapping subset, the unseen held-out members, is defined by um. When emulating M, a deep learning method will be trained on $tm_{0..n-j}$, where j represents the total number of unseen held-out members. After the model is trained it will be applied to $um_{n-j+1..n}$. Given t timesteps of information, the trained model will be used to forecast d days of concentrations for a set of defined species. In this work we used the following species due to the relevance to air quality: Carbon Monoxide (CO), Nitric Oxide (NO), Nitrogen Dioxide (NO2), and Tropospheric Ozone (O3).

Creating the GEOS-CF Ensemble Dataset

We produced a comprehensive set of retrospective model output, including ensemble simulations, to serve as the training data for the deep learning models. To make the problem computationally feasible, these simulations were conducted at a horizontal resolution of approximately $100x100\,\mathrm{km}^2$. The training data consists of a one-year (2021) simulation of a GEOS-CF like system, plus 32 ensemble simulations over the same time period that were designed to capture the model sensitivity to the main drivers of atmospheric composition, namely pollutant emissions and meteorology. The 32 meteorological ensembles were produced by constraining the meteorological fields to the 32 ensembles produced by the GEOS Forward Processing (GEOS-FP) system. For the 32 emission ensembles, we randomly perturbed the anthropogenic emissions of nitrogen oxide (NO) and carbon monoxide (CO) using Gaussian kernels with an approximate length scale of 250 km.

Methods

We explored adapting two methods for learning to forecast ensemble mean and to estimate the ensemble uncertainty. We modified both the Bihlo cGAN model and the FourCastNet model to provide flexibility in choice of model architecture and input features. Both models are considered state of the art in terms of their forecast skill for weather variables.

FourCastNet (Pathak, Subramanian et al. 2022) uses an autoregressive approach, meaning it predicts one timestep out and feeds that prediction back into the model for the next timestep. We continued using the autoregressive approach. We however did not use its built-in ensembling for this work, as our goal was to work with the ensemble data itself from GEOS-CF.

The cGAN model based on Bihlo (Bihlo 2021) is trained using eight timesteps of input and forecasts eight timesteps into the future. We kept this time stepping and used the vid2vid architecture, modified for our dataset.

The novelty we introduce in this work (to both networks) is adapting them for forecasting atmospheric chemistry variables, which is essential for air quality forecasting. In order to better understand features that could be useful in this type of forecast, we performed a set of sensitivity experiments. We then setup experiments where we train each model to learn to forecast the ensemble mean and standard deviation.

Experimental Results

It is important to understand the effects of emissions and meteorology on concentrations. Therefore, this study includes experiments that address understanding how including emissions information with concentrations will affect the overall forecast skill. In addition, we explore how including emissions and meteorology for future timesteps influences the forecast skill. Finally, the third experiment examines the forecast skill of both cGAN and FourCastNet for a held-out set of ensemble members using a small set of timesteps of initialization. In this experiment we forecast up to 10 days with both methods.

Metrics

We measure performance in terms of Mean Absolute Error (MAE) and Root Mean Squared Error(RMSE).

$$\text{MAE} = \frac{1}{n}\sum_{j=1}^{n}|y_j - \hat{y}_j| \tag{1}$$

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{j=1}^{n}(y_j - \hat{y}_j)^2} \tag{2}$$

Including Emissions Information

In the first set of experiments, we compared the forecast skill when emissions is included vs. not included. Emission data is readily available in the GEOS model. We measured performance in terms of the average RMSE. Early experiments show a smaller average RMSE when emissions are included with concentrations as input for FourCastNet. We saw a similar trend in forecast skill using the cGAN model. For the cGAN model we plotted the set of ensemble members, represented as different trends. In Figure 1, we show, consistent with the FourCastNet model, the cGAN forecast skill does improve with emissions information included. Overall, it appears that including emissions as a feature in addition to the concentration values, improves the forecast skill.

Including Emissions and Meteorological Information

In the following set of sensitivity experiments we tested how well cGAN performs when meteorological information is included with emissions for a given forecast and we compared this with also having some knowledge of future emissions and meteorological information (as the GEOS model does have this information available). In Figure 2, we show the results of CO and in Figure 3, we show the results of NO2. For each chemical species, the MAE is lower when knowledge of future emissions and meteorological information is included.

From these results, having insight into future meteorology and emissions did improve the forecast skill of concentrations for all forecasted species. We did not conduct this experiment using FourCastNet due to the complexities of the network.

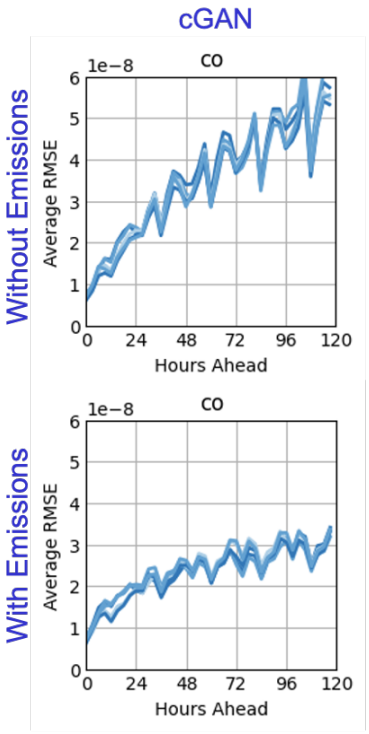

Figure 1: Early experiments show a smaller average RMSE when emissions are included with concentrations as input for cGAN.

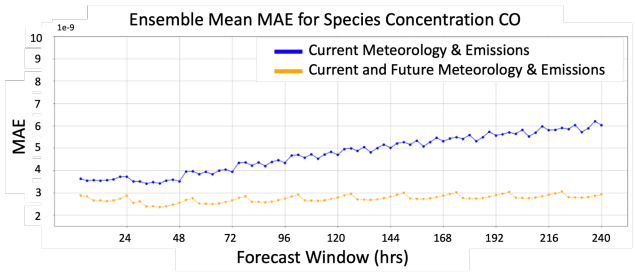

Figure 2: Comparing cGAN forecast skill for CO when current meteorology and emissions is known with current and future projected meteorology and emissions.

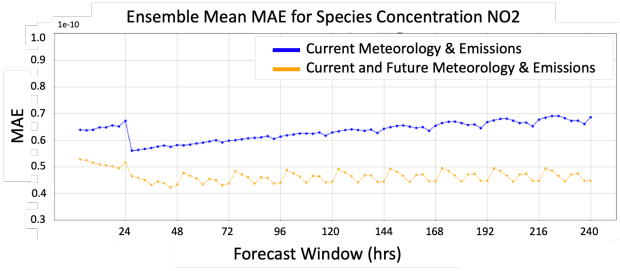

Figure 3: Comparing cGAN forecast skill for NO2 when current meteorology and emissions is known with current and future projected meteorology and emissions.

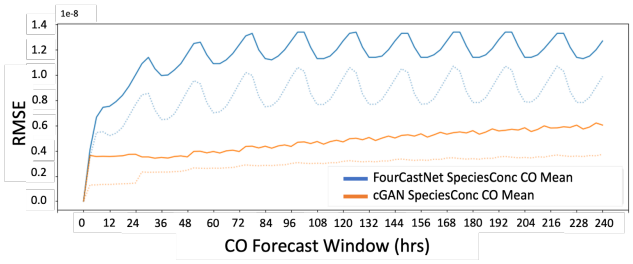

Figure 4: Comparing FourCastNet and cGAN ten day forecast for CO by comparing RMSE scores for mean (solid) and spread (dotted).

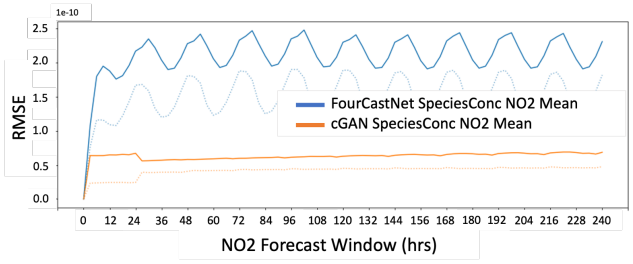

Figure 5: Comparing FourCastNet and cGAN ten day forecast for NO2 by comparing RMSE scores for mean (solid) and spread (dotted).

Learning Ensemble Mean and Estimating Uncertainty

In this experiment we study how FourCastNet and cGAN could be used to forecast ten days of the ensemble with a few timesteps of information from a set of ensemble members.

For this setup, both FourCastNet and cGAN were trained on two months of GEOS-CF data for 12 of the ensemble members and the test set contained two months of data for five of the ensemble members. Results were based on meteorological and emissions data present and an evaluation was setup to measure performance of concentrations for CO, NO, NO2, O3.

We compare the forecast of up to ten days for FourCastNet and cGAN trained with emissions and meteorological information. We show the results for CO in Figure 4 and NO2 in Figure 5. Included in each plot is the RMSE for both the mean and the spread.

Interesting results across all species shows cGAN tends to out-perform FourCastNet. We also see that FourCastNet has oscillatory behavior. These results could be due to known instabilities in FourCastNet, or could be due to the smaller data set size. FourCastNet could be overfit in this experiment. Loss plots do indicate validation loss is slightly worse than training loss. However when we examine the concentration distributions we also see similar cyclical patterns in concentrations. It could be that FourCastNet is more sensitive to these patterns.

Conclusions and Future Work

We explored the potential of two deep learning models, cGAN and FourCastNet, to produce ensemble predictions of atmospheric composition. Both methods show promise predicting the spatiotemporal evolution of key air pollutants up to 10 days into the future.

We demonstrated that the inclusion of emissions appears to improve the forecast skill of species concentrations when using both a cGAN and FourCastNet.

We demonstrated using a cGAN trained on a set of ensemble member data, where emission data and meteorological data was given for the current timestep and future projections, the cGAN was able to make significantly more accurate forecasts of ensemble mean and spread when future projections were included (predictions were made for a 24-hour period).

We also demonstrated that FourCastNet, for the first time, could be applied to forecasting trace gases. This model was also trained on ensemble member data. We showed that we could make a 10-day forecast given unseen ensemble member data using just seven timesteps of information (21 hours) for FourCastNet and eight timesteps for the cGAN. We are working towards three timesteps (nine hours). The error that accumulated over those 10 days was minimal.

Surprisingly, cGAN out-performed FourCastNet for all species. Given that the cGAN was trained using a single GPU and for approximately two days and FourCastNet was trained on four GPUs for more than two days, these results are compelling. However, given the size of the dataset, more experiments with larger datasets would be required to confirm these findings. Future work will include experiments with a larger number of ensemble members.

The cGAN and the FourCastNet model could be combined with a numerical composition forecasting system to produce invaluable ensemble information at a computationally feasible cost. Future work includes testing such a hybrid application with the GEOS-CF model, with the goal to improve its data assimilation architecture and to provide uncertainty estimates along with the main composition forecast.

Our methodology could be applied to other ensemble-based modeling systems and provides a way to significantly reduce the time to run the ensemble, with minimal loss in overall performance. This could enable larger ensembles, providing an opportunity to conduct experiments that would otherwise be computationally infeasible to run. Future work will include exploring this idea of applying our methodology to other types of modeling systems.

Acknowledgements

This work has been funded by NASA grant NNH21ZDA001N-AIST21-0024.

References

Bihlo, A. 2021. A generative adversarial network approach to (ensemble) weather prediction. *Neural Networks*, 139: 1–16.

Brecht, R.; and Bihlo, A. 2022. Deep learning for ensemble forecasting. In *EGU General Assembly Conference Abstracts*, EGU22–2058.

Debry, E.; and Mallet, V. 2014. Ensemble forecasting with machine learning algorithms for ozone, nitrogen dioxide and PM10 on the Prev'Air platform. *Atmospheric environment*, 91: 71–84.

Fuller, R.; Landrigan, P. J.; Balakrishnan, K.; et al. 2022. Pollution and health: a progress update. *The Lancet Planetary Health*, 6(6): e535–e547.

Isola, P.; Zhu, J.-Y.; et al. 2017. Image-to-image translation with conditional adversarial networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 1125–1134.

Keller, C. A.; Knowland, K. E.; et al. 2021. Description of the NASA GEOS composition forecast modeling system GEOS-CF v1. 0. *Journal of Advances in Modeling Earth Systems*, 13(4): e2020MS002413.

Pathak, J.; Subramanian, S.; et al. 2022. Fourcastnet: A global data-driven high-resolution weather model using adaptive fourier neural operators. *arXiv preprint arXiv:2202.11214*.

Shukla, P. R.; Skea, J.; Calvo Buendia, E.; et al. 2019. IPCC, 2019: Climate Change and Land: An IPCC Special Report on Climate Change, Desertification, Land Degradation, Sustainable Land Management, Food Security, and Greenhouse Gas Fluxes in Terrestrial Ecosystems. In press.