Transformers for Fast Emulation of Atmospheric Chemistry Box Models

Herbie Bradley
University of Cambridge, UK
hb574@cam.ac.uk

Nathan Luke Abraham
National Centre for Atmospheric Science, UK
Department of Chemistry, University of Cambridge, UK
nla27@cam.ac.uk

Peer Nowack
Climatic Research Unit
School of Environmental Sciences
University of East Anglia, UK
p.nowack@uea.ac.uk

Doug McNeall
Met Office Hadley Centre, UK
University of Exeter, UK
doug.mcneall@metoffice.gov.uk

Abstract

In computational models of atmospheric chemistry, chemical concentrations are determined by numerically solving large systems of ordinary differential equations which represent a set of chemical reactions. These solvers can be very computationally intensive, particularly those with the thousands or tens of thousands of chemical species and reactions that make up the most accurate models. We demonstrate the application of a deep learning transformer architecture to emulate an atmospheric chemistry box model, and show that this attention-based model outperforms LSTM and autoencoder baselines while providing interpretable predictions that are more than 2 orders of magnitude faster than a numerical solver.

1 Introduction

Atmospheric chemistry models are important scientific tools to understand the role of aerosols and ozone in climate change and their effect on human health. Simulations of atmospheric chemistry involve tens to hundreds of interacting chemical species, coupled by chemical reactions and integrated over time using implicit numerical solvers. These models can be very computationally expensive, due to the large number of coupled chemical species and stiffness of the underlying system of differential equations [4]. However, this expense is crucial to accurately model chemicals such as ozone which are not directly emitted into the atmosphere but have widespread effects on climate [15]. For the UKESM Earth system model, [18], the numerical integration component for atmospheric chemistry consists of roughly 10% of the total computational cost.

A growing body of recent work [7, 8, 9, 10, 17] has investigated the use of machine learning emulators to replace numerical solvers in atmospheric chemistry and achieve speedups of up to an order of magnitude. In this work, we develop a neural network emulator for an atmospheric chemistry box model using Temporal Fusion Transformers [12]. Our model is capable of interpretable, stable predictions of chemical concentrations over a multi-day time period consisting of hundreds of time steps, and achieves a speed up of several orders of magnitude over the numerical solver. This provides a potential path forward to develop an emulator for a global atmospheric chemistry model such as the UKCA model [1], which in turn fees into policy-informing global climate change projections.

Tackling Climate Change with Machine Learning: workshop at NeurIPS 2022.

2 Methodology

2.1 MOZART-4—The Chemical Model

The MOZART-4 mechanism [3] is an intermediate complexity atmospheric chemistry model for the troposphere. The mechanism consists of 83 chemical species, whose evolution is governed by 38 photolysis reactions and 158 gas phase reactions. We study MOZART-4 as a box model—a model representing the evolution of chemicals in a single grid cell, or box. We chose this mechanism since MOZART-4 is of similar chemical complexity to the UKCA model mentioned above and is available as a box model. A box model facilitates the generation of training data and as a result, enables comparisons to a variety of possible machine learning emulator setups.

The majority of prior work in emulation of atmospheric chemistry models [8, 9, 13, 10, 16] chose to use a deep learning approach. Kelp et al. [9], use an encoder-decoder neural network with a recurrent 'operator' block in the middle to emulate each time step, and found that this architecture successfully sped up the mechanism by several orders of magnitude, but predictions started to diverge from the mechanism results when a time horizon of more than 24 hours was used. We reproduce this architecture for comparison in our experiments. More recent work [10] built on these results using a similar model, but devised an 'online' training scheme in which the emulator is trained in parallel with the atmospheric chemistry model run and receives data sequentially. This approach showed a promising improvement in long-term stability of emulator predictions.

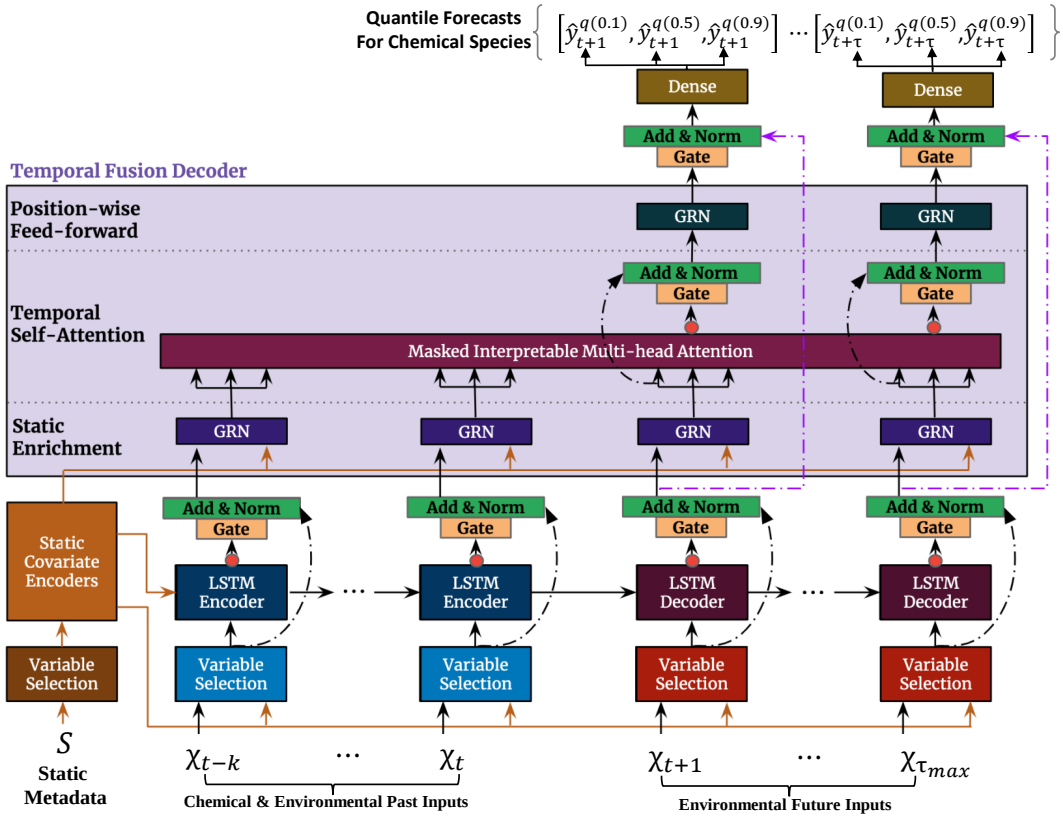

Figure 1: Temporal Fusion Transformer (TFT) architecture. The TFT inputs static metadata, time-varying past inputs and time-varying known future inputs. The GRN is a Gated Residual Network [12], and the Variable Selection networks are a weighted sum of GRNs, one for each input variable. The output is a set of quantile forecasts for each chemical species τ time steps into the future. Figure partially reproduced from [12].

The mechanism is implemented in Fortran 90 using the Kinetic PreProcessor (KPP) and BOXMOX. KPP [2] is a tool which converts a defined set of chemical reactions and rate constants into Fortran

files which will solve the ODE system representing the chemical mechanism. BOXMOX [11] extends KPP to box models and provides a convenient interface to specify initial concentrations, boundary conditions, and environmental parameters.

Using a set of chemically-plausible initial condition ranges [9] and Latin Hypercube Sampling [20], we use BOXMOX to generate a training dataset of 5000 runs of the MOZART-4 box model and a test dataset of 100 runs, each of which ran for 3 days of simulation time using a 15 minute time step. The training dataset consists of 1.435 million data points. We standardize the data to zero mean and unit standard deviation, ignore the first time step of each run as a warm-up to reduce variance, and mix midday & midnight starts of the simulation to reduce the possibility of the emulator overfitting to characteristics of the start time. The input to the emulator at each time step consists of the concentrations of the 83 chemical species (over one or more past time steps), concatenated with environmental variables (temperature, pressure, and solar zenith angle) which affect the speed of chemical reactions. The solar zenith angle acts as a time representation for the model, since the time of day can be calculated from it given a history of at least 2 time points.

2.2 Temporal Fusion Transformers

Our model, a Temporal Fusion Transformer (TFT) [12], is an attention-based architecture capable of interpretable multi-horizon time series modeling. The TFT has the ability to output prediction intervals of various percentiles for each time step, such as the 10^{th}, 50^{th}, and 90^{th} percentiles, via a dense layer on top of the decoder output. The model is trained with a quantile loss [21] which drives the prediction intervals to be well calibrated. We use early stopping with a patience of 10 epochs to halt training when the quantile loss on the validation set plateaus, dropout [19] with $p = 0.1$, the Rectified Adam gradient descent optimizer [14] with a learning rate of 0.01, and clip gradient norms to 1.0.

We compare the TFT with the encoder-decoder architecture described above [9]), with the addition of a pseudo-online training scheme in which the emulator is trained on data points from the simulation in order, which has been shown to improve stability [10]. We also compare with LSTM [6] and random forest baselines. We plan to release our codebase as open-source, to encourage further development of emulators for atmospheric chemistry.

3 Results

Table 1: Evaluation metrics for our models, using a history of length 10 and averaged across all chemical species over a 3 day simulation window. We show the mean squared error (MSE), mean absolute error (MAE), $R^2 \in [0, 1]$, and ozone absolute error in ppm.

Model	MSE	MAE	R^2	O_3 error (ppm)
TFT	**0.0274**	**0.1672**	0.965	**0.1238**
Autoencoder [10]	0.0306	0.1980	0.958	0.1429
LSTM	0.0289	0.2884	**0.971**	0.2087
Random Forest	0.0297	0.2721	0.916	0.2925

Table 1 shows our results across several evaluation metrics, for a training scheme using the past 10 time steps of chemical concentrations for each time step of prediction. We include the ozone error in parts per million to indicate the chemical scale of the errors, since this species is one of the most important to accurately model. These results show that the TFT emulator outperforms the closest baseline by an average of **8.4%**, in addition to providing more easily interpretable predictions. Figure 1 shows kernel density estimates from the TFT emulated chemistry compared to the ground truth, and demonstrates that our model is a good approximation to the simulated chemistry for two important chemical species.

To compare the running times of the emulator and the chemical simulation, we timed the creation of the training dataset consisting of 5000 runs of the MOZART-4 box model: this took $\sim$35 minutes on a single CPU, although the efficiency could likely be improved further. Meanwhile, the batched inference using the TFT model for predictions of 5000 concentration time series of the same length took $\sim$19 seconds on a V100 GPU, averaged over multiple predictions—a speed difference of

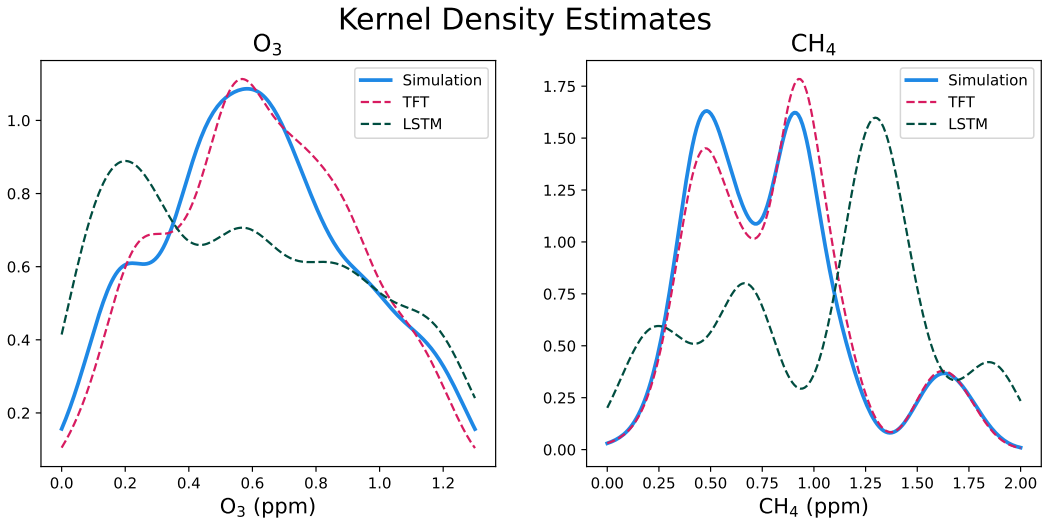

Figure 2: Un-normalized kernel density estimates for Ozone and Methane (in ppm), using the TFT and LSTM baseline for emulation predictions.

$\sim$**2** orders of magnitude. The LSTM and autoencoder baselines have even faster inference, since transformer inference is inherently slightly slower due to architectural limitations.

4 Discussion & Conclusion

Our results show that a transformer specialised for time-series modeling can supply significantly better predictions for a chemical box model, while still providing interpretable output that quantifies the relative importance of features and time steps. We also demonstrate the scale of the speed-ups that can be obtained from replacing an expensive numerical solver for atmospheric chemistry with a transformer model.

In future work, we will extend our work to the UKCA [1] 3D atmospheric chemical model, with the goal of replacing the numerical solver with a stable, fast machine learned emulator, achieving significant speedups in global climate modeling. The MOZART-4 mechanism we emulate in this work is of comparable complexity to UKCA. The numerical solver for the UKCA component in the UKESM [18] Earth system model takes up 10% of the entire wall-clock computation time, and replacing the UKCA chemical solver with a trained neural network emulator could save up to $\sim$530 CPU core hours per UKESM model-year. Applied to a full Coupled Model Intercomparison Project (CMIP) [5] round, a 10% saving is the equivalent of over 1000 model-years of simulation, with a significant associated reduction in energy usage for the HPC system.

In conclusion, we demonstrated promising results from a deep learning attention-based architecture for emulation of a numerical solver for atmospheric chemistry box models, and show a path forward to emulation of a global chemical model. We hope that this work will contribute to a better understanding of emulation techniques for chemical numerical solvers, and potentially spur applications to other emulation problems.

References

[1] S. Archer-Nicholls, N. L. Abraham, Y. M. Shin, J. Weber, M. R. Russo, D. Lowe, S. R. Utembe, F. M. O'Connor, B. Kerridge, B. Latter, R. Siddans, M. Jenkin, O. Wild, and A. T. Archibald, "The common representative intermediates mechanism version 2 in the united kingdom chemistry and aerosols model," *Journal of Advances in Modeling Earth Systems*, vol. 13, no. 5, e2020MS002420, 2021. DOI: https://doi.org/10.1029/2020MS002420.

[2] V. Damian, A. Sandu, M. Damian, F. Potra, and G. R. Carmichael, "The kinetic preprocessor kpp-a software environment for solving chemical kinetics," *Computers & Chemical Engineering*, vol. 26, no. 11, pp. 1567–1579, 2002.

[3] L. K. Emmons, S. Walters, P. G. Hess, J.-F. Lamarque, G. G. Pfister, D. Fillmore, C. Granier, A. Guenther, D. Kinnison, T. Laepple, J. Orlando, X. Tie, G. Tyndall, C. Wiedinmyer, S. L. Baughcum, and S. Kloster, "Description and evaluation of the Model for Ozone and Related chemical Tracers, version 4 (MOZART-4)," *Geoscientific Model Development*, vol. 3, no. 1, pp. 43–67, 2010.

[4] E. Esentürk, N. L. Abraham, S. Archer-Nicholls, C. Mitsakou, P. Griffiths, A. Archibald, and J. Pyle, "Quasi-newton methods for atmospheric chemistry simulations: Implementation in ukca um vn10.8," *Geoscientific Model Development*, vol. 11, no. 8, pp. 3089–3108, 2018. DOI: 10.5194/gmd-11-3089-2018.

[5] V. Eyring, S. Bony, G. A. Meehl, C. A. Senior, B. Stevens, R. J. Stouffer, and K. E. Taylor, "Overview of the coupled model intercomparison project phase 6 (cmip6) experimental design and organization," *Geoscientific Model Development*, vol. 9, no. 5, pp. 1937–1958, 2016. DOI: 10.5194/gmd-9-1937-2016.

[6] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, 1997.

[7] C. A. Keller and M. J. Evans, "Application of random forest regression to the calculation of gas-phase chemistry within the geos-chem chemistry model v10," *Geoscientific Model Development*, vol. 12, no. 3, pp. 1209–1225, 2019. DOI: 10.5194/gmd-12-1209-2019.

[8] M. M. Kelp, C. W. Tessum, and J. D. Marshall, "Orders-of-magnitude speedup in atmospheric chemistry modeling through neural network-based emulation," *arXiv preprint arXiv:1808.03874*, 2018.

[9] M. M. Kelp, D. J. Jacob, J. N. Kutz, J. D. Marshall, and C. W. Tessum, "Toward Stable, General Machine-Learned Models of the Atmospheric Chemical System," *Journal of Geophysical Research: Atmospheres*, vol. 125, no. 23, 2020.

[10] M. M. Kelp, D. J. Jacob, H. Lin, and M. P. Sulprizio, "An online-learned neural network chemical solver for stable long-term global simulations of atmospheric chemistry," *Journal of Advances in Modeling Earth Systems*, vol. 14, no. 6, 2022, e2021MS002926 2021MS002926. DOI: https://doi.org/10.1029/2021MS002926.

[11] C. Knote, J. Barré, and M. Eckl, "BEATBOX v1.0: Background Error Analysis Testbed with Box Models," *Geoscientific Model Development*, vol. 11, no. 2, pp. 561–573, 2018.

[12] B. Lim, S. Ö. Arık, N. Loeff, and T. Pfister, "Temporal fusion transformers for interpretable multi-horizon time series forecasting," *International Journal of Forecasting*, vol. 37, no. 4, pp. 1748–1764, 2021.

[13] C. Liu, H. Zhang, Z. Cheng, J. Shen, J. Zhao, Y. Wang, S. Wang, and Y. Cheng, "Emulation of an atmospheric gas-phase chemistry solver through deep learning: Case study of Chinese Mainland," *Atmospheric Pollution Research*, vol. 12, no. 6, p. 101 079, 2021.

[14] L. Liu, H. Jiang, P. He, W. Chen, X. Liu, J. Gao, and J. Han, "On the variance of the adaptive learning rate and beyond," in *International Conference on Learning Representations*, 2020.

[15] P. J. Nowack, N. Luke Abraham, A. C. Maycock, P. Braesicke, J. M. Gregory, M. M. Joshi, A. Osprey, and J. A. Pyle, "A large ozone-circulation feedback and its implications for global warming assessments," *Nature Climate Change*, vol. 5, no. 1, pp. 41–45, 2015. DOI: 10.1038/nclimate2451.

[16] J. Schreck, C. Becker, D. J. Gagne, K. Lawrence, S. Wang, C. Mouchel-Vallon, J. Choi, and A. Hodzic, "Neural network emulation of the formation of organic aerosols based on the explicit gecko-a chemistry model," *Earth and Space Science Open Archive*, p. 85, 2022. DOI: 10.1002/essoar.10510054.1.

[17] J. S. Schreck, C. Becker, D. J. Gagne, K. Lawrence, S. Wang, C. Mouchel-Vallon, J. Choi, and A. Hodzic, "Neural network emulation of the formation of organic aerosols based on the explicit gecko-a chemistry model," *Journal of Advances in Modeling Earth Systems*, 2022.

[18] A. A. Sellar, C. G. Jones, J. P. Mulcahy, Y. Tang, A. Yool, A. Wiltshire, F. M. O'Connor, M. Stringer, R. Hill, J. Palmieri, S. Woodward, L. de Mora, T. Kuhlbrodt, S. T. Rumbold, D. I. Kelley, R. Ellis, C. E. Johnson, J. Walton, N. L. Abraham, M. B. Andrews, T. Andrews, A. T. Archibald, S. Berthou, E. Burke, E. Blockley, K. Carslaw, M. Dalvi, J. Edwards, G. A. Folberth, N. Gedney, P. T. Griffiths, A. B. Harper, M. A. Hendry, A. J. Hewitt, B. Johnson, A. Jones, C. D. Jones, J. Keeble, S. Liddicoat, O. Morgenstern, R. J. Parker, V. Predoi, E. Robertson, A. Siahaan, R. S. Smith, R. Swaminathan, M. T. Woodhouse, G. Zeng, and M. Zerroukat, "Ukesm1: Description and evaluation of the u.k. earth system model," *Journal of Advances in Modeling Earth Systems*, vol. 11, no. 12, pp. 4513–4558, 2019. DOI: https://doi.org/10.1029/2019MS001739.

[19] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A simple way to prevent neural networks from overfitting," *J. Mach. Learn. Res.*, vol. 15, no. 1, pp. 1929–1958, 2014.

[20] M. Stein, "Large sample properties of simulations using latin hypercube sampling," *Technometrics*, vol. 29, no. 2, pp. 143–151, 1987.

[21] R. Wen, K. Torkkola, B. Narayanaswamy, and D. Madeka, "A multi-horizon quantile recurrent forecaster," in *NeurIPS 2017*, 2017.