A Deep Learning Parameterization for Ozone Dry Deposition

Velocities

S. J. Silva[1], C. L. Heald[1], S. Ravela[2], I. Mammarella[3], J. William Munger[4]

[1]Department of Civil and Environmental Engineering, Massachusetts Institute of Technology, Cambridge, MA

[2]Earth Signals and Systems Group, Department of Earth, Atmospheric, and Planetary Sciences, Massachusetts Institute of Technology, Cambridge, MA

[3]Institute for Atmospheric and Earth System Research/Physics, Faculty of Science, University of Helsinki, Finland

[4]School of Engineering and Applied Sciences, Harvard University, Cambridge, MA

Abstract

The loss of ozone to terrestrial and aquatic systems, known as dry deposition, is a highly uncertain process governed by turbulent transport, interfacial chemistry, and plant physiology. We demonstrate the value of using Deep Neural Networks (DNN) in predicting ozone dry deposition velocities. We find that a feedforward DNN trained on observations from a coniferous forest site (Hyytiälä, Finland) can predict hourly ozone dry deposition velocities at a mixed forest site (Harvard Forest, Massachusetts) more accurately than modern theoretical models, with a reduction in the Normalized Mean Bias (0.05 vs. ~0.1). The same DNN model, when driven by assimilated meteorology at 2°x2.5° spatial resolution, outperforms the Wesely scheme as implemented in the GEOS-Chem model. With more available training data from other climate and ecological zones, this methodology could yield a generalizable DNN suitable for global models.

Plain Language Summary

Ozone in the lower atmosphere is a toxic pollutant and greenhouse gas. In this work, we use a machine learning technique known as "deep learning", to simulate the loss of ozone to earth's surface. We show that our deep learning simulation of this loss process outperforms existing traditional models, and demonstrate the opportunity for using machine learning to improve our understanding of the chemical composition of the atmosphere.

1. Introduction

The uptake of atmospheric ozone (O_3) at the Earth's surface, commonly referred to as dry deposition, is a key process controlling the concentration of ozone in the lower atmosphere. On the global scale, dry deposition is responsible for 20-25% of all ozone loss in the troposphere (Lelieveld & Dentener, 2000). As exposure to surface ozone is globally responsible for 250,000 premature deaths per year (Cohen et al., 2017), and tropospheric ozone is a short-lived greenhouse gas which contributes ~20% of the radiative forcing of CO_2 (IPCC, 2013), it is important to understand the processes contributing to the lifetime and distribution of ozone in the troposphere.

In nature, dry deposition varies with changes in turbulent transport, surface chemistry, and plant physiology, though this complexity is challenging to describe on relevant spatial scales with current generation models (Wesely & Hicks, 2000; Wu et al., 2018). Instead, models typically calculate dry deposition as a linear function of concentration and a transfer velocity:

$$F = -V_d[O_3]$$

Where F is the total flux out of the atmosphere (negative indicating a downward flux), V_d is the "dry deposition velocity" and $[O_3]$ is the ozone concentration in the lowest layer of the atmosphere.

Modern parameterizations for calculating V_d are all largely based on a "resistor-in-series" framework as in Wesely (1989) (Hardacre et al., 2015). This resistor analogy attempts to parameterize the complex nonlinear processes governing dry deposition based on observed relationships, turbulent transport theory, heuristic assumptions regarding plant physiology, and a set of a priori estimated look up table values (e.g.

Zhang et al., 2002). The continuous development of these models has led to them becoming highly parameterized, and the large number of lookup table values and approximations within them limits their physical interpretability. Despite these elaborate parameterizations, recent work has shown that global models which use these schemes struggle to reproduce observations to within a factor of 1.4 (Clifton et al., 2016; Hardacre et al., 2015; Silva & Heald, 2018).

These current generation parameterizations of ozone dry deposition are conceptual and empirical schemes that do not directly represent the physics of the process. Thus, a machine learning solution potentially offers an effective alternative to improve dry deposition velocity predictions, with little loss of process-based knowledge. Machine learning methods have gained popularity in recent years as high quality methods to make rapid and accurate predictions from data in the earth and environmental sciences. Keller and Evans (2018) demonstrate the potential of using a Random Forest algorithm to replace gas-phase chemistry in a global chemical transport model. Nowack et al. (2018) use a Ridge regression technique to linearly parameterize ozone-temperature relationships for climate models, and find that their machine learning model can predict global ozone fields quite well at a fraction of the computational cost of traditional non-parameterized methods. Rasp et al. (2018) demonstrate the value of deep learning in predicting highly nonlinear subgrid processes in climate models, wherein their DNN produces consistent results 10 times faster than traditional models. In this work, we explore how a machine learning approach can be applied to the simulation and prediction of V_d, specifically focusing on the use and application of DNNs due to successful applications of the method in a computationally efficient manner. We further justify our machine learning model choice in section 4.3.

In this work we present the first parameterization of ozone dry deposition velocities using a DNN. We train a model using data from a coniferous forest site in Hyytiälä, Finland, and evaluate the model fidelity against observations at the largely deciduous Harvard Forest in the northeastern United States.

2. Data Sources

Two datasets of V_d were used in this work, observations at Hyytiälä Forest in southern Finland, and at Harvard Forest in the northeastern United States.

The Hyytiälä Forest site is part of the SMEAR II research station (Keronen et al., 2003; Zhou et al., 2017; https://www.atm.helsinki.fi/SMEAR/). The site is located in a coniferous boreal forest with a climate characterized by cold winters and mild humid summers. The ozone dry deposition velocity data record (Junninen et al., 2009; https://avaa.tdata.fi/web/smart/smear) used in this work extends from 2001-2013, with observations averaged to hourly intervals (Rannik et al., 2012). This results in 55995 total observations of V_d and associated meteorological and physical parameters for model development. This is the primary dataset used for model training, validation, and testing in this study.

The Harvard Forest data were collected at the Harvard Forest Environmental Measurements Site is a component of the Harvard Forest LTER and part of the AmeriFlux network. It is surrounded by a mixed deciduous/coniferous forest (Munger et al., 1996; Munger & Wofsy, 2018; http://harvardforest.fas.harvard.edu). The climate is similar to Hyytiälä, except summers are warmer, and winters more mild. The data record used here is from 1992-2005, with observations at an hourly time resolution.

This dataset contains 38363 observations that we use as a test for model generalizability in Section 4.2.

We trained the models in this work using surface observations common to both sites: wind speed, sensible heat flux, air temperature, relative humidity, photosynthetically active radiation, and time (month and hour).

3. Deep Learning

The deep learning model was developed and trained with open source software in the R programming language, using the Keras API (keras.io, Chollet & others, 2015) and the TensorFlow libraries (Martín Abadi et al., 2015; http://tensorflow.org), all facilitated by the RStudio Interface to Tensorflow (Chollet & Allaire, 2018), https://tensorflow.rstudio.com). The model we used in this work was a so-called "Multilayer Perceptron" or "Feedforward Deep Neural Network" category of machine learning models, as they are the most basic form of DNNs.

Deep Neural Networks consist of interconnected layers that transform the input data (here, environmental parameters) into the desired output (ozone dry deposition velocities). Each layer consists of a number of nodes, which is a weighted sum of the output values of the previous layer plus a bias term, all within a nonlinear function known as an activation function. The weights and the bias within each node are the values that are optimized by model training, using a method known as stochastic gradient descent, by iterating on small samples (batches) of data for a given number of iterations (epochs). The activation function is used in part to improve the ability of the DNN to capture nonlinearities within the data. For a DNN such as the one used here, the total number of trainable parameters per layer is calculated as $(n+1)m$, with n weights

(equal to the number of inputs to the layer) plus one bias term repeated *m* times (equal to the number of outputs from the layer). We describe our selection process for the hyperparameters (number of nodes and layers, activation function, etc) in the following section. For more details on deep learning, see Goodfellow et al. (2016).

3.1 Hyperparameter Selection and Dataset Preprocessing

In designing DNNs, there are many preprocessing decisions that need to be made prior to training the model, and no common set of rules or heuristics governing the decision making process (Goodfellow et al., 2016; Islam & Murase, 2001). In general, we focus on keeping the number of free and tunable parameters in our model as small as possible, as opposed to optimizing every detail.

We use the RMSprop stochastic gradient descent optimizer (Tieleman & Hinton, 2012) in this work, with the Keras default parameters (Chollet & others, 2015). For these default parameters, the initial learning rate was set to 0.001, the decay factor used was 0.9, and no gradient clipping was applied to the norm or the absolute gradient value. This optimizer improves performance over a basic stochastic gradient descent method, without adding substantial computational overhead (Goodfellow et al., 2016; Tieleman & Hinton, 2012). The batch size used for training was 64, which was an empirically chosen as a tradeoff between computational speed and model convergence. We used a rectified linear unit (ReLU) nonlinear activation function (Glorot et al., 2011) due to both its accelerated convergence over traditional (e.g. sigmoidal) activation functions, appropriateness for neural network regression of a real valued number (real valued output), and its popularity in the literature (Chollet & Allaire, 2018). We additionally add 20% dropout to each layer for each epoch to reduce model overfitting (Chollet &

Allaire, 2018). The loss metric to be minimized is the Mean Squared Error (MSE), which is typical for DNN continuous real number regressions.

We use a simple heuristic to bound the number of layers in the neural network by the number of iterations that an independent regression may need to obtain convergence (less than 10% error). For this purpose, we use an iterative "boosted" linear model from the XGBoost library (Chen & Guestrin, 2016). A boosted model is a machine learning algorithm that iteratively reduces bias in a model estimate. We investigated the improvement of the boosted model with each additional iteration, using the observations at Hyytiälä Forest. Once the MSE of the boosted model improved by less than 10%, we selected that iteration number as the number of layers in our model. For the Hyytiälä Forest dataset, that number was 3.

We posit that the number of nodes in the each layer of the network (neurons) must be at most as many as the number of independent and orthogonal dimensions within the dataset (Angus, 1991; Seuret et al., 2017). While it is possible to estimate independence in a variety of ways, we simply choose the number of significant principal components as the number of independent dimensions (Seuret et al., 2017). This is because the dimensionality of the optimum estimate for a linear model is its principal components, and a non-linear model for the same estimation problem must admit fewer dimensions (Jain et al., 2000). In this case, more than half of the variance was explained with the first 16 eigenvalues, and more than 90% explained with the first 32. We chose 16 in this work to keep the total number of trainable parameters a full order of magnitude lower than the size of the training dataset.

We normalize all the data to have mean of zero and standard deviation of one using the mean and standard deviation of the test (Hyytiälä Forest) data, as is typical to ensure

that model parameters are well behaved during the training and to improve convergence (Goodfellow et al., 2016). Once the data are processed by the DNN, the output is rescaled for interpretation.

4. Results and Discussion

4.1 Model Training

The model was trained for 500 epochs (iterations of a batch of data) using the specifications in section 3 on observations from the Hyytiälä Forest dataset. Following typical DNN training and evaluation techniques (Chollet & Allaire, 2018; Hastie et al., 2001), we randomly split the dataset into training (70%), validation (10%), and test (20%) subsets. The model parameters were trained using the training dataset and both hyperparameter selection and overfitting is assessed by evaluating the trained DNN on the independent validation subset. If the DNN performs substantially better on the training data as compared to the validation data, it is likely that the model is overfit. Using this method, we tuned the hyperparameters to what is summarized in section 3. Once satisfactory performance was achieved on the training and validation sets, a final evaluation was completed using the test subset of data. Model summary statistics are shown in Table 1, with the MSE, correlation, and the normalized mean bias (NMB). Broadly, the DNN model is trained to capture the majority of the variability in the training dataset, and reduces the mean disagreement between the model and observations to near zero. From Table 1, it is apparent that the validation dataset evaluation statistics are very similar (within 0.01) to those of the training dataset, consistent with little overfitting. As expected, the model errors on the training and validation datasets are more optimistic than the test dataset (Hastie et al., 2001), however the results are still quite good compared to other models of ozone dry

deposition velocities. Over similar mixed forests, current models of dry deposition have NMB of ~0.1 to 0.4 (Wu et al., 2018), higher than the test data performance of this model.

	MSE	R	NMB
Training	0.02	0.82	0.04
Validation	0.02	0.81	0.05
Test	0.07	0.56	0.05

Table 1. Summary Statistics for the DNN model of Hyytiälä Forest ozone dry deposition velocities, including the mean squared error (MSE), Pearson correlation coefficient (R), and normalized mean bias (NMB).

4.2 Model Generalizability

The DNN model applied to the Hyytiälä test data has predictive skill, with performance as good or better than many theoretical parameterizations (Silva & Heald, 2018; Wu et al., 2018). However, the overall generalization error (error on the test set) of the DNN model as it applies to all mixed forests similar to Hyytiälä is likely underestimated in this situation because all of the data are not i.i.d. (independent and identically distributed random variables). In time series data from natural dynamical systems like ozone dry deposition velocities, the test data are not fully independent from the training data, even when randomly sampled. In this case, an improved test for the generalization of the DNN model is to use a source of data observed at a different location. We use observations at the Harvard Forest site for this purpose.

We evaluated the model trained in section 4.1 on the observations at the Harvard Forest site (1992-2005). Monthly average deposition velocities are shown in Figure 1 for the observations and the DNN. Broadly, the DNN reproduces the observations well,

indicating that the model is generalizable. The general seasonal variability is captured by the DNN, with an overall hourly correlation of 0.35 (0.5 in the summer months, 0.2 in the winter). This correlation is higher than many global models (Silva & Heald, 2018) but 5-10% lower than current generation point models (Wu et al., 2018). The bias of the DNN is relatively very small, with the absolute NMB ~0.1. This is as good or better than most current generation models and much higher than global models (Hardacre et al., 2015; Silva & Heald, 2018; Wu et al., 2018).
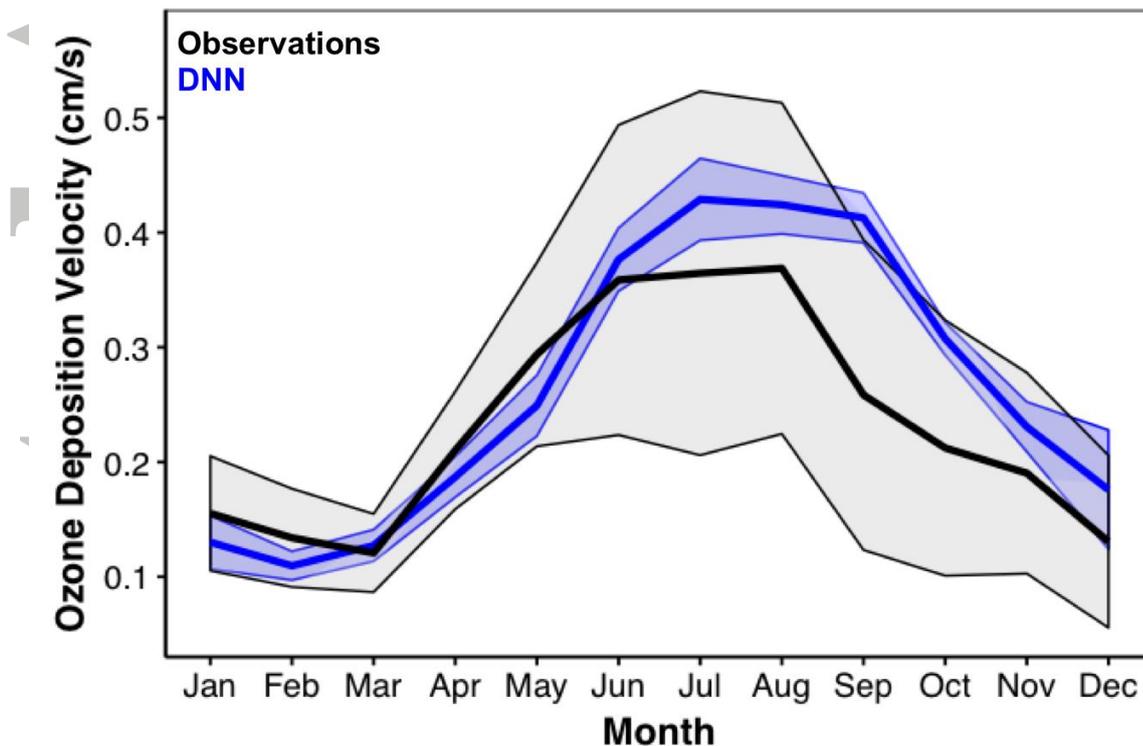
Figure 1. Monthly average V_d over Harvard Forest from 1992-2005. Observations are in black, and the DNN is in blue. The shaded regions represent ±1 standard deviation across years.

4.3 Comparison with Other Models

While machine learning parameterizations are currently interesting and novel methodological approaches, that alone does not warrant their use. Since DNNs are fundamentally "black box" models that lack interpretability, their use needs to be justified further than purely a demonstration of predictive skill. We justify the added value in using a DNN by comparison with several more interpretable approaches including: a simple linear model, a Random Forest, and a Ridge regression.

Using the same training, validation, and test split set as for the DNN parameterization, we trained a simple linear model, a Random Forest (Liaw & Wiener, 2002), and a Ridge regression (Friedman et al., 2010). Table 2 shows summary statistics over Harvard Forest. We include the correlation, the MSE, and the relative computational time of model prediction.

Overall, the Random Forest and DNN outperform the linear and Ridge regression, likely due in part to how V_d responds nonlinearly to the input variables used here. While the correlation and MSE summary statistics are very similar for both the DNN and Random Forest regression, the Random Forest prediction time is 27% slower than the DNN. Though the model prediction time is not perfectly optimized, these results are consistent with previous work demonstrating rapid computation from DNNs (Rasp et al., 2018), and slower implementations of Random Forests (Keller & Evans, 2018). Given the computational challenges many models already face, the lack of process-based information currently available in models of V_d, and the great potential for DNN model portability and retraining (Chollet & Allaire, 2018) we believe that the application of a DNN for this purpose is well justified.

Model	R	MSE	Time
DNN	0.35	0.20	1.00
Linear	0.22	0.52	< 0.01
Random Forest	0.36	0.19	1.27
Ridge	0.33	0.43	0.07
Model	R	MSE	Time
DNN	0.35	0.20	1.00
Linear	0.22	0.52	< 0.01
Random Forest	0.36	0.19	1.27
Ridge	0.33	0.43	0.07

Table 2. Summary Statistics (Pearson correlation coefficient, R, Mean Squared Error, MSE, and Computational time of model prediction relative to the DNN) for the DNN parameterizations, the linear model, a Random Forest Regression, and a Ridge regression of ozone dry deposition velocities at Harvard Forest.

4.4 Applicability in 3-D Models

Due to coarser spatiotemporal resolution and potential errors in meteorological fields (Gelaro et al., 2017), it is not immediately clear how the DNN developed here would perform in a regional-global scale 3D model. To explore this we evaluate the value of the developed DNN at the Harvard Forest site using assimilated meteorology, and compare it with a global model (GEOS-Chem v9-02, www.geos-chem.org) parameterization of dry deposition (Silva & Heald, 2018) using that same meteorology.

We use the Modern Era Retrospective Meteorological ReAnalysis (MERRA) meteorology of the driving variables of the DNN regridded to match the GEOS-Chem spatial resolution of 2°x2.5°, with all three-hourly fields linearly interpolated to the hourly resolution to directly match the date and time of the observations.

Seasonal average hourly deposition velocities at Harvard Forest are shown in Figure 2 for the observations, the DNN driven by MERRA (DNN-MERRA), and the Wesely (1989) deposition scheme as implemented in the GEOS-Chem model. The DNN-MERRA parameterization reproduces the observations better than the Wesely (1989) parameterization, with substantially lower summertime bias, and the overall seasonal cycle well represented. The DNN-MERRA NMB (0.12) is lower than the Wesely parameterization NMB (0.3) while the hourly correlation is similar (~0.35). The DNN-MERRA NMB is in line with the test data evaluation from Hyytiälä Forest, while the correlation is lower, indicating that the model captures less of the observed variability when generalized to the Harvard Forest location and when meteorology is degraded to the 2°x2.5° spatial resolution. Overall, this test suggests that the DNN-MERRA model outperforms current parameterizations, and could potentially be applied effectively in regional and global scale models.
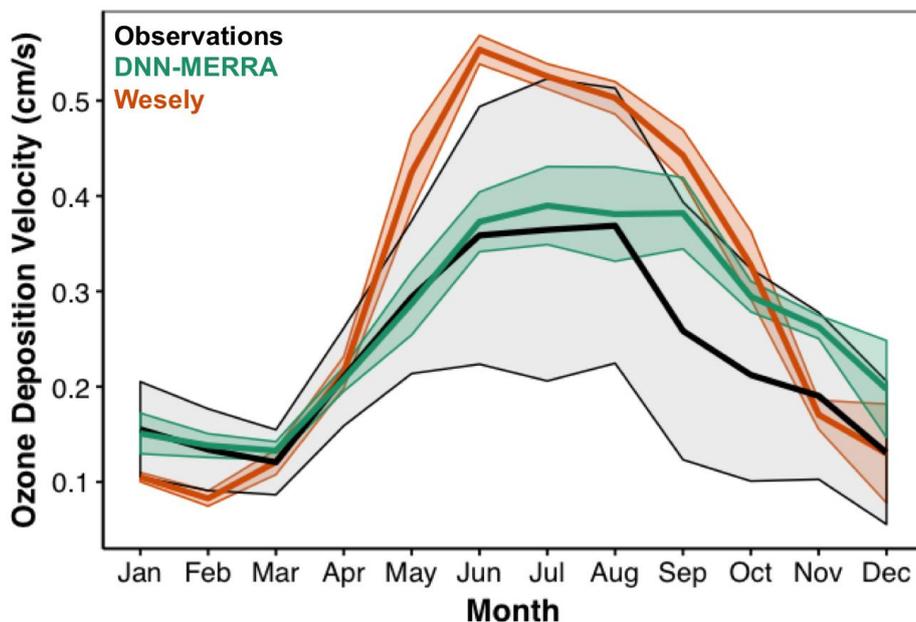
Figure 2. Monthly average V_d over Harvard Forest from 1992-2005. Observations are in black, the DNN driven by MERRA meteorology (DNN-MERRA) is in green, and the

Weseley parameterization as implemented in the GEOS-Chem model (Weseley) is in orange. The shaded regions represent ±1 standard deviation across years.

4.5 Data Requirements

As with all machine learning methods, the DNN in this work excels at reproducing observations that are within the bounds of the data it is trained on. There is no guarantee that the model will perform well over a vastly different vegetation type or meteorological scenario. Thus, more long-term observations over varied regions are necessary if this methodology were to be applied globally.

In defining the parameters of the DNN in section 3.1, we required the somewhat arbitrary condition that we have at least an order of magnitude more data than free parameters in our model. To achieve this with the DNN as structured here, we require at least 12,330 observations. At the Hyytiälä site, there are generally 3500 – 6500 observations taken each year, which would require ~5 years of observations to achieve the constraint on total observations used in this work.

Instead of constraining the total number of observations, one could instead propose to require the generalization error to reach a certain threshold value. We use a limited set of the Hyytiälä Forest data (the year 2008) to train a DNN with the same parameters as in section 3.1. We explore the generalization loss over Harvard Forest as a function of the number of months observed, with the goal is to evaluate at what point the generalization loss approaches the loss of the fully trained DNN (loss = 4.1). We explore all possible months as a start date, and range from 1 to 12 months of continuous data throughout the year, with the DNN retrained on each combination of months. Figure 3 shows that after approximately 6 months of continuous data collection (~2800

observations) a generalization loss of within 50% of the fully trained DNN is reached, a value that should outperform the Weseley parameterization as implemented in GEOS-Chem (Silva & Heald, 2018). In this specific case, training on months containing data from summers is more effective at reducing the generalization loss than winter months. As more continuous months of data are used the loss begins to approach the fully trained DNN. The final value using just one full year is still far from the overall loss, indicating that more than a year of data is valuable for model accuracy and generalizability. This points to the high value of long-term observations in varied regions, many of which currently have limited data availability (Silva & Heald, 2018).
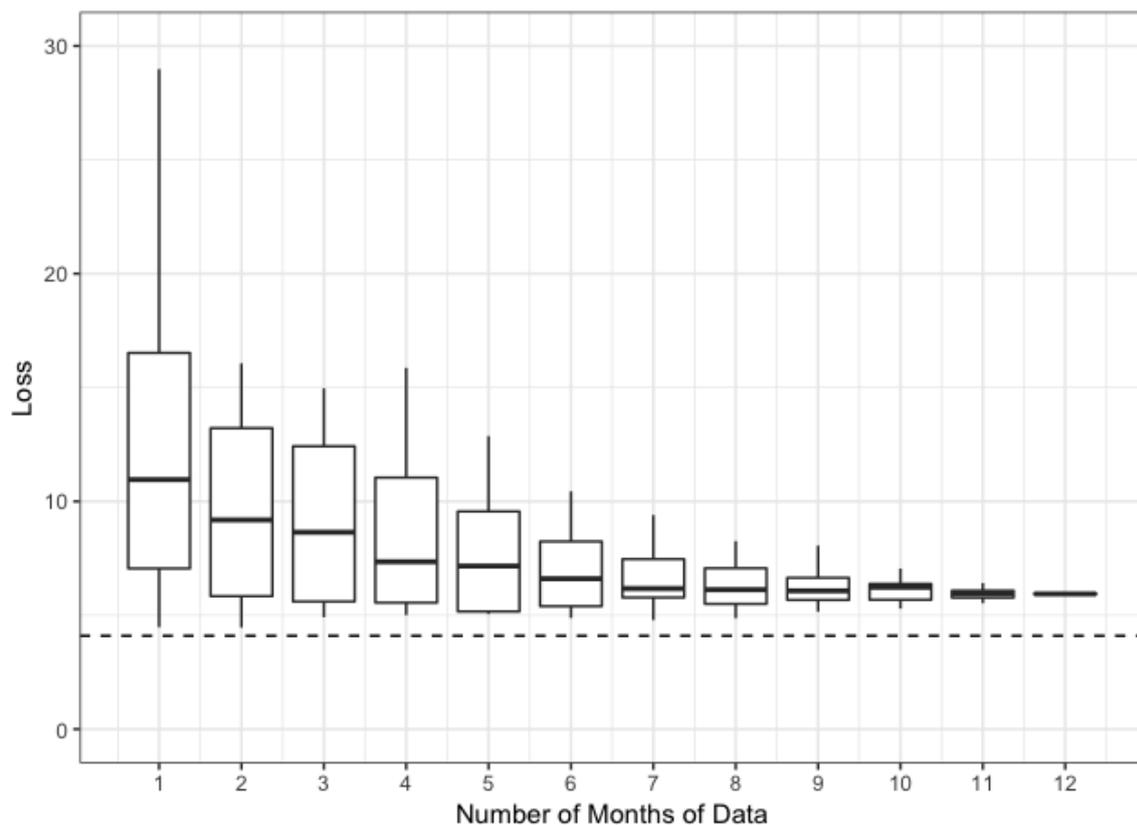
Figure 3. The generalization loss at Harvard Forest as a function of continuous months of training data from Hyytiälä Forest in the year 2008. The horizontal line is the generalization loss when all data from all years are used. Boxplots are plotted in the style of Tukey (see McGill et al., 1978), with the middle bars corresponding to medians,

with the lower and upper hinges corresponding to the 25th and 75th percentile, respectively. The vertical whiskers extend to the hinge to the most extreme value within 1.5*the IQR defined by the hinges.

5. Conclusions

We present the first parameterization of ozone dry deposition velocities using a deep learning framework. The DNN parameterization in this work outperforms other statistical models in terms of speed and accuracy, and it is also less biased compared to observations relative to the commonly used physical parameterization of Wesely (1989). With continued data collection and improved data availability, particularly from long-term measurement sites, machine learning methods like deep learning are likely to continue to grow in their applicability and performance, and present valuable opportunities for reducing model bias and uncertainty for highly parameterized model processes.

Acknowledgements

References

Angus, J. E. (1991). *Criteria for Choosing the Best Neural Network: Part 1* (No. NHRC-91-

16). NAVAL HEALTH RESEARCH CENTER SAN DIEGO CA, NAVAL HEALTH

RESEARCH CENTER SAN DIEGO CA. Retrieved from

http://www.dtic.mil/docs/citations/ADA247725

Chen, T., & Guestrin, C. (2016). XGBoost: A Scalable Tree Boosting System. In

Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge

Discovery and Data Mining (pp. 785–794). New York, NY, USA: ACM.

https://doi.org/10.1145/2939672.2939785

Chollet, F., & Allaire, J. J. (2018). *Deep learning with R*. Shelter Island, NY: Manning

Publications Co.

Chollet, F., & others. (2015). *Keras*. Retrieved from https://keras.io

Clifton, O. E., Fiore, A. M., Munger, J. W., Malyshev, S., Horowitz, L. W., Shevliakova, E., et

al. (2016). Interannual variability in ozone removal by a temperate deciduous

forest. *Geophysical Research Letters*, 2016GL070923.

https://doi.org/10.1002/2016GL070923

Cohen, A. J., Brauer, M., Burnett, R., Anderson, H. R., Frostad, J., Estep, K., et al. (2017).

Estimates and 25-year trends of the global burden of disease attributable to

ambient air pollution: an analysis of data from the Global Burden of Diseases

Study 2015. *The Lancet*, *0*(0). https://doi.org/10.1016/S0140-6736(17)30505-6

Friedman, J., Hastie, T., & Tibshirani, R. (2010). Regularization Paths for Generalized

Linear Models via Coordinate Descent. *Journal of Statistical Software*, *33*(1), 1–

22.

Gelaro, R., McCarty, W., Suárez, M. J., Todling, R., Molod, A., Takacs, L., et al. (2017). The

    Modern-Era Retrospective Analysis for Research and Applications, Version 2

    (MERRA-2). *Journal of Climate*, *30*(14), 5419–5454.

    https://doi.org/10.1175/JCLI-D-16-0758.1

Glorot, X., Bordes, A., & Bengio, Y. (2011). Deep Sparse Rectifier Neural Networks. In G.

    Gordon, D. Dunson, & M. Dudík (Eds.), *Proceedings of the Fourteenth International*

    Conference on Artificial Intelligence and Statistics (Vol. 15, pp. 315–323). Fort

    Lauderdale, FL, USA: PMLR. Retrieved from

    http://proceedings.mlr.press/v15/glorot11a.html

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning*. MIT Press.

Hardacre, C., Wild, O., & Emberson, L. (2015). An evaluation of ozone dry deposition in

    global scale chemistry climate models. *Atmos. Chem. Phys.*, *15*(11), 6419–6436.

    https://doi.org/10.5194/acp-15-6419-2015

Hastie, T., Tibshirani, R., & Friedman, J. (2001). *The Elements of Statistical Learning*. New

    York, NY, USA: Springer New York Inc.

IPCC. (2013). Index. In T. F. Stocker, D. Qin, G.-K. Plattner, M. Tignor, S. K. Allen, J.

    Boschung, et al. (Eds.), *Climate Change 2013: The Physical Science Basis.*

    Contribution of Working Group I to the Fifth Assessment Report of the

    Intergovernmental Panel on Climate Change (pp. 1523–1535). Cambridge, United

    Kingdom and New York, NY, USA: Cambridge University Press.

    https://doi.org/10.1017/CBO9781107415324

Islam, M., & Murase, K. (2001). A new algorithm to design compact two-hidden-layer

    arti®cial neural networks. *Neural Networks*, 14.

Jain, A. K., Duin, R. P. W., & Mao, J. (2000). Statistical Pattern Recognition: A Review. *IEEE Trans. Pattern Anal. Mach. Intell.*, *22*(1), 4–37. https://doi.org/10.1109/34.824819

Junninen, H., Lauri, A., Keronen, P., Aalto, P., Hiltunen, V., Hari, P., & Kulmala, M. (2009). Smart-SMEAR: on-line data exploration and visualization tool for SMEAR stations, *14*, 11.

Keller, C. A., & Evans, M. J. (2018). Application of random forest regression to the calculation of gas-phase chemistry within the GEOS-Chem chemistry model v10. *Geoscientific Model Development Discussions*, 1–25. https://doi.org/10.5194/gmd-2018-229

Keronen, P., Reissell, A., Rannik, Ü., Pohja, T., Siivola, E., Hiltunen, V., et al. (2003). Ozone flux measurements over a Scots pine forest using eddy covariance method: performance evaluation and comparison with flux-profile method, *8*, 19.

Lelieveld, J., & Dentener, F. J. (2000). What controls tropospheric ozone? *Journal of Geophysical Research: Atmospheres*, *105*(D3), 3531–3551. https://doi.org/10.1029/1999JD901011

Liaw, A., & Wiener, M. (2002). Classification and Regression by randomForest. *R News*, *2*(3), 18–22.

Martín Abadi, Ashish Agarwal, Paul Barham, Eugene Brevdo, Zhifeng Chen, Craig Citro, et al. (2015). *TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems*. Retrieved from https://www.tensorflow.org/

McGill, R., Tukey, J. W., & Larsen, W. A. (1978). Variations of Box Plots. *The American Statistician*, *32*(1), 12–16. https://doi.org/10.2307/2683468

Munger, J. W., Wofsy, S. C., Bakwin, P. S., Fan, S.-M., Goulden, M. L., Daube, B. C., et al. (1996). Atmospheric deposition of reactive nitrogen oxides and ozone in a

temperate deciduous forest and a subarctic woodland: 1. Measurements and

mechanisms. *Journal of Geophysical Research: Atmospheres*, *101*(D7), 12639–

12657. https://doi.org/10.1029/96JD00230

Munger, W., & Wofsy, S. (2018). Canopy-Atmosphere Exchange of Carbon, Water and

Energy at Harvard Forest EMS Tower since 1991 [Data set]. Environmental Data

Initiative.

https://doi.org/10.6073/pasta/dd9351a3ab5316c844848c3505a8149d

Nowack, P., Braesicke, P., Haigh, J., Abraham, N. L., Pyle, J., & Voulgarakis, A. (2018).

Using machine learning to build temperature-based ozone parameterizations for

climate sensitivity simulations. *Environmental Research Letters*, *13*(10), 104016.

https://doi.org/10.1088/1748-9326/aae2be

Rannik, Ü., Altimir, N., Mammarella, I., Bäck, J., Rinne, J., Ruuskanen, T. M., et al. (2012).

Ozone deposition into a boreal forest over a decade of observations: evaluating

deposition partitioning and driving variables. *Atmos. Chem. Phys.*, *12*(24), 12165–

12182. https://doi.org/10.5194/acp-12-12165-2012

Rasp, S., Pritchard, M. S., & Gentine, P. (2018). Deep learning to represent subgrid

processes in climate models. *Proceedings of the National Academy of Sciences*,

201810286. https://doi.org/10.1073/pnas.1810286115

Seuret, M., Alberti, M., Liwicki, M., & Ingold, R. (2017). PCA-Initialized Deep Neural

Networks Applied to Document Image Analysis. In *2017 14th IAPR International

Conference on Document Analysis and Recognition (ICDAR)* (Vol. 01, pp. 877–882).

https://doi.org/10.1109/ICDAR.2017.148

Silva, S. J., & Heald, C. L. (2018). Investigating Dry Deposition of Ozone to Vegetation.

Journal of Geophysical Research: Atmospheres, 2017JD027278.

https://doi.org/10.1002/2017JD027278

Tieleman, T., & Hinton, G. (2012). *Lecture 6.5—RmsProp: Divide the gradient by a running average of its recent magnitude*.

Wesely, M. L. (1989). Parameterization of surface resistances to gaseous dry deposition in regional-scale numerical models. *Atmospheric Environment (1967)*, *23*(6), 1293–1304. https://doi.org/10.1016/0004-6981(89)90153-4

Wesely, M. L., & Hicks, B. B. (2000). A review of the current status of knowledge on dry deposition. *Atmospheric Environment*, *34*(12–14), 2261–2282. https://doi.org/10.1016/S1352-2310(99)00467-7

Wu, Z., Schwede, D. B., Vet, R., Walker, J. T., Shaw, M., Staebler, R., & Zhang, L. (2018). Evaluation and Intercomparison of Five North American Dry Deposition Algorithms at a Mixed Forest Site. *Journal of Advances in Modeling Earth Systems*, *10*(7), 1571–1586. https://doi.org/10.1029/2017MS001231

Zhang, L., Moran, M. D., Makar, P. A., Brook, J. R., & Gong, S. (2002). Modelling gaseous dry deposition in AURAMS: a unified regional air-quality modelling system. *Atmospheric Environment*, *36*(3), 537–560. https://doi.org/10.1016/S1352-2310(01)00447-2

Zhou, P., Ganzeveld, L., Rannik, Ü., Zhou, L., Gierens, R., Taipale, D., et al. (2017). Simulating ozone dry deposition at a boreal forest with a multi-layer canopy deposition model. *Atmos. Chem. Phys.*, *17*(2), 1361–1379. https://doi.org/10.5194/acp-17-1361-2017