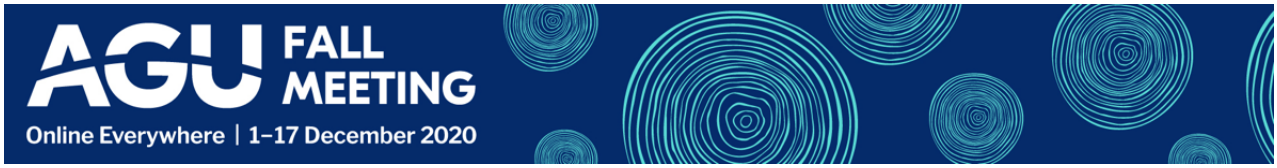
Estimating the Rate of Change of Stratospheric Ozone using Deep Neural Networks

Helge Mohn, Daniel Kreyling, Ingo Wohltmann, Merlin Barschke*, Markus Rex

Alfred Wegener Institute Helmholtz-Center for Polar and Marine Research Potsdam, Climate Sciences | Atmospheric Physics, Potsdam, Germany
*Technische Universität Berlin, Chair of Space Technology, Berlin, Germany

PRESENTED AT:



INTRODUCTION

We present a fast model for stratospheric ozone chemistry based on a neural network approach. The model is intended to replace the detailed chemistry schemes of chemistry and transport models (CTMs), general circulation models (GCMs) or Earth system models (ESMs), which are computationally very expensive.

The neural network (NN) model estimates the rate of change of ozone in 24 hours at a grid point and is trained on data of the detailed full chemistry model of the ATLAS chemistry and transport model (CTM) [2]. The benefit of this surrogate models is a much lower computation time (minutes instead of days) while the same level of accuracy is achieved. This represents a necessary step from understanding the chemistry and building sophisticated CTMs towards the usage of this knowledge in climate models, which is only feasible if much lower computation times can be achieved.

Modelling of the Earth system is a complex task and models usually contain a large number of sub-modules and parameterizations. This applies for example to the atmosphere, hydrosphere, solid earth and the ice sheets. Atmospheric chemistry is complex and usually involves dozens of species and hundreds of reactions with a wide range of concentrations and lifetimes.

This project concentrates on the estimation of the rate of change of ozone in the extrapolar stratosphere. The dynamics from the polar regions and from other layers of the atmosphere regarding the ozone change are not treated within this work.

The ATLAS model [2] is a Lagrangian CTM for stratospheric chemistry. It solves a coupled differential equation system using a stiff solver and a variable time-step. The stratospheric chemistry scheme of ATLAS has 46 active species, 171 reactions and heterogeneous chemistry on polar stratospheric clouds. It is not using the concept of chemical families. The application of the ATLAS CTM has high requirements on computational power. This is the reason why the coupling of full chemical models to climate models is generally not feasible with respect to the computation time of a global climate model. However, the incorporation of detailed chemistry is often desirable, in order to account for various feed-backs between chemistry, atmosphere and ocean. These complex chemical models motivate the formulation of faster but still accurate surrogate models, that are tailored to the coupling into earth climate models.

This project builds on the SWIFT project, which has a polar and an extrapolar surrogate model for the stratospheric ozone chemistry. We investigate an alternative approach to the polynomial approach used by extrapolar SWIFT by exploiting the improved approximation capability of NN with respect to nonlinear contexts.

[1] D. Kreyling, I. Wohltmann, R. Lehmann, and M. Rex. The extrapolar swift model (version 1.0): fast stratospheric ozone chemistry for global climate models. Geosci- entific Model Development, 11(2):753–769, 2018.

[2] I. Wohltmann and M. Rex. The lagrangian chemistry and transport model atlas: validation of advective transport and mixing. Geoscientific Model Development, 2(2):153–173, 2009.

DATA SCIENCE

Within this project a careful visual examination of the data was performed to analyze the big data set and to engineer new features.

Data is crucial for this project, since it is key for Machine Learning. All the NN will learn is based on the data. With millions of data samples, it can become a challenge to keep the overview and to make the structure in the data visible. Some effort has already been put into selecting appropriate input variables in the context of the development of the polynomial approach of extrapolar SWIFT [1]. Here, we use visualization methods from data science to get a deeper understanding of the data.

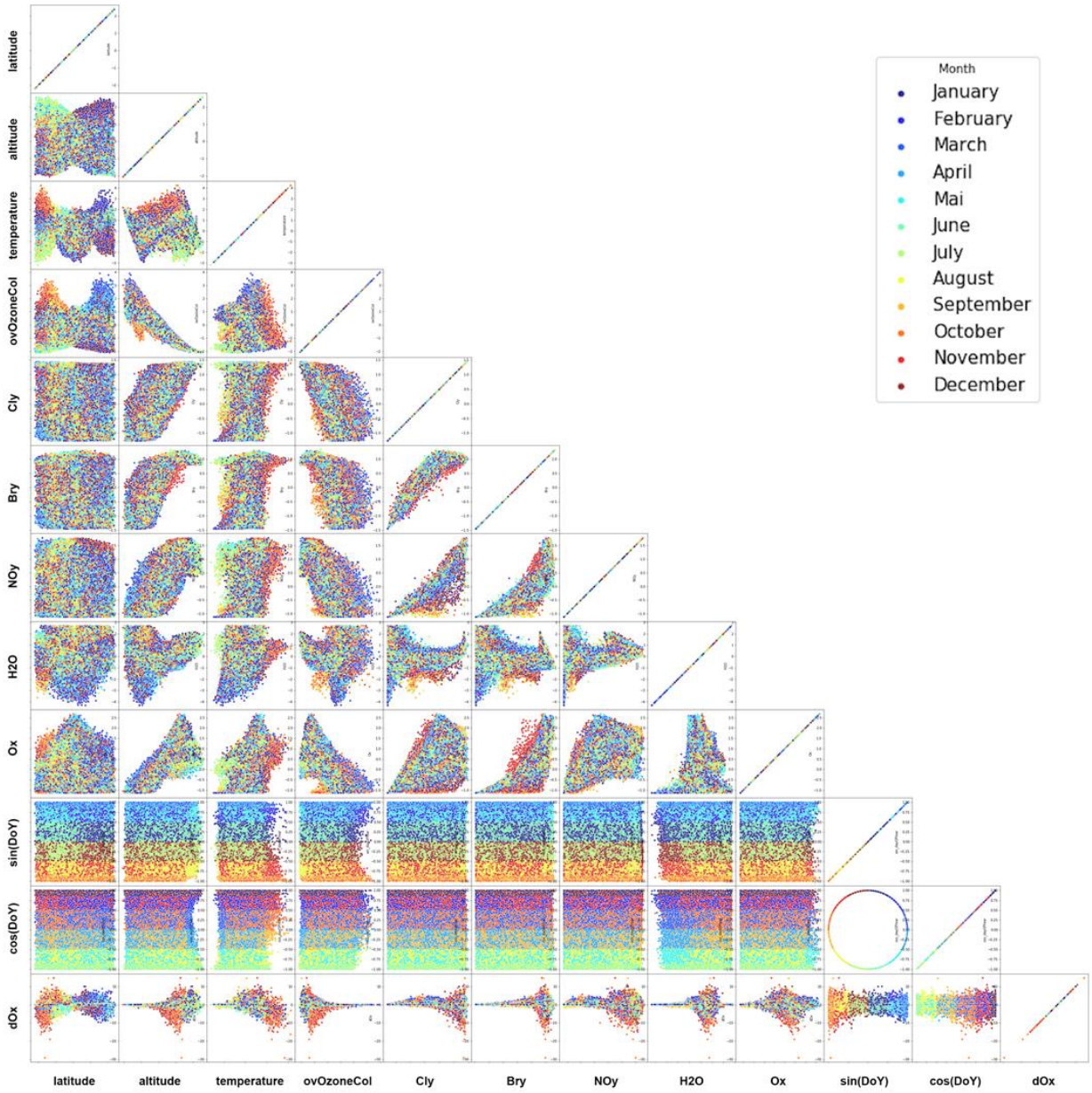


Fig. 1: Scatter-pair plot of input and output variables. Each diagram represents a bivariate scatter plot. Each point represents a data sample, and its color indicates the month of the data sample.

Fig. 1 gives an insight into the distribution among each variable. In contrast to the previous method [1] we added the day-of-year as a variable, to better exploit the machine learning method. By taking the sine and cosine of this value, its cyclical representation can be learned by the neural network model. The respective diagram shows the achieved cyclical representation of the new feature. By comparing the colors of the dots, we can learn that some variables vary over time quite significantly. In most cases the distributions are pretty dense and certain shapes become visible. In general only a few scatter dots are separated from the main distribution.

Fig. 2 further more explores the linear correlation of all variables among each other by showing the Pearson's correlation score. It can be seen that the time of year has an impact on the linear correlation. The monthly data has been sampled globally and put into a variable-wise comparison. The linear correlation with respect to the output variable dOx is rather low. Since the input variables have been determined carefully for the previous polynomial approach of SWIFT [1], a correlation between the input and the rate of change of ozone has already been shown. The intuition supports that a non-linear approach is required to use the already known correlation by a machine learning method. A neural network should therefore be well suited.

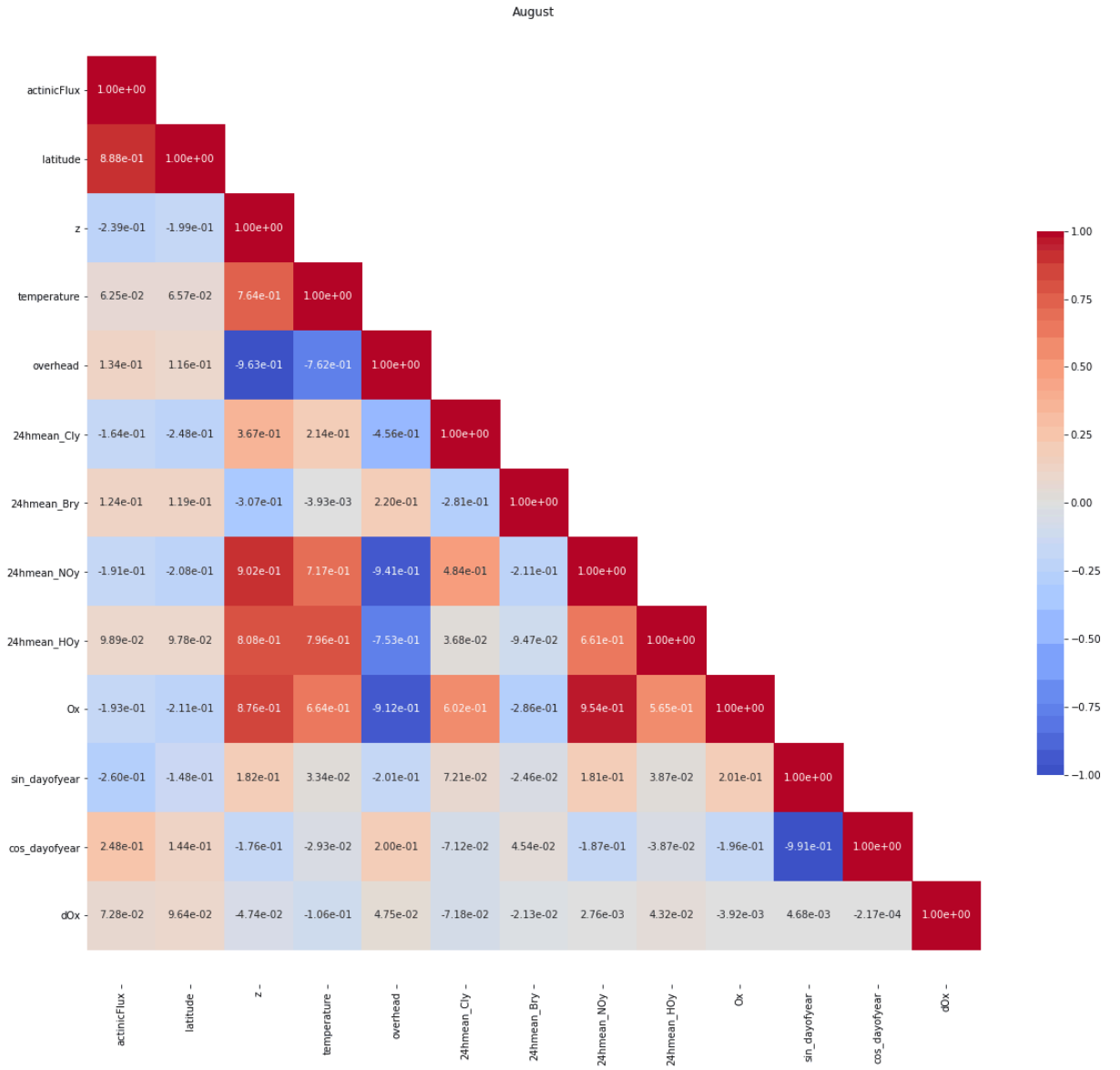


Fig. 2: Pearson Correlation Heatmap: A characteristic monthly correlation is shown. Each square represents the result of a pair of variables. The data set has been divided into monthly groups. The animation shows the results one month at a time and shows the current month in the title (above). Bluish colors (negative values) show a negative linear correlation and reddish colors (positive values) show a positive linear correlation. Greyish colors show no or a weak linear correlation.

[1] D. Kreyling, I. Wohltmann, R. Lehmann, and M. Rex. The extrapolar swift model (version 1.0): fast stratospheric ozone chemistry for global climate models. Geosci- entific Model Development, 11(2):753–769, 2018.

METHODOLOGY

This approach uses a deep feed-forward Neural Network (NN) with fully-connected layers as a regression model. It was implemented using the open source machine learning framework PyTorch [3]. In general such a NN has one input-layer, several hidden-layers and one output layer (see fig. 3).

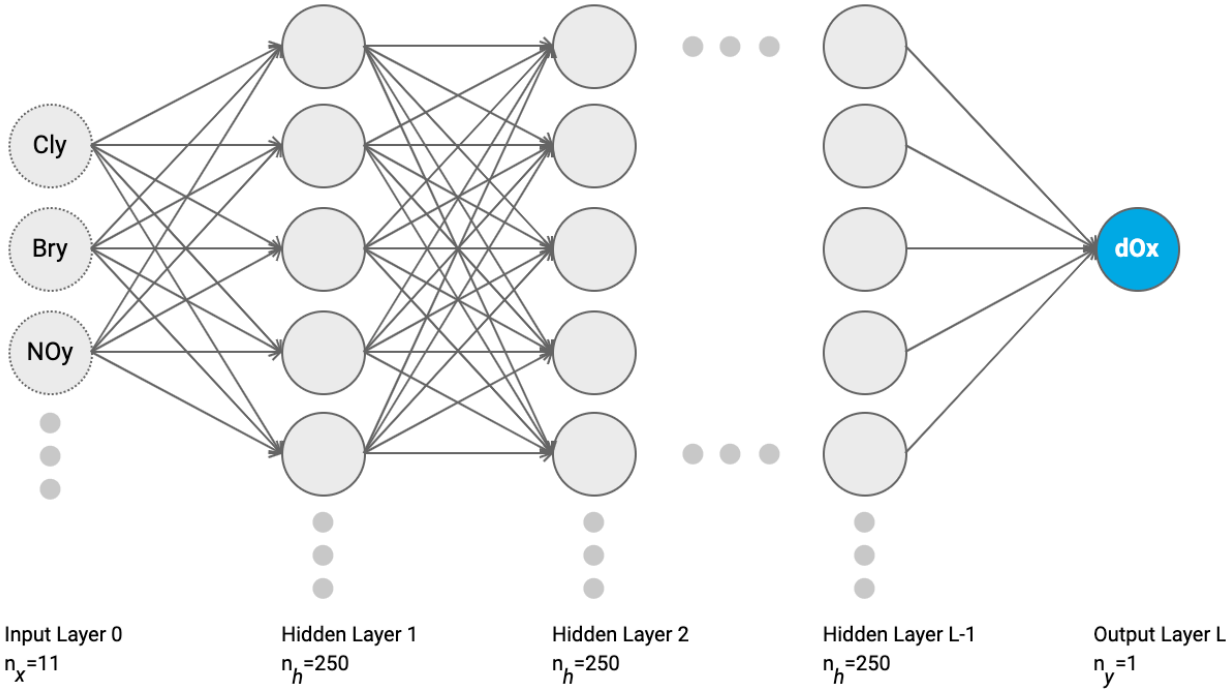


Fig. 3: Sketch of the NN-architecture. A fully connected NN is shown. On the left side the input neurons are shown, followed by the hidden layers (middle) and the regression output neuron to the right. The number of layers and the number of neurons per layer need to be optimized by a hyperparameter search.

Our input layer uses eleven input variables that are normalized by zero-mean and unit-variance. In addition to the previous polynomial approach SWIFT [1], which uses nine input features, this approach extends this selection by two more features that represent the day-of-year as a vector.

In our case the output layer consists of only one neuron that outputs our regression estimate. During training this estimate is also normalized by zero-mean and unit-variance. This output of the NN-model is the 24h-ozone-gradient (dO_x), which is used to estimate the volume mixing ratio of ozone of the next day:

$$O_x^t = O_x^{t-1} + \Delta O_x$$

The NN-models of this project use a fix time step for the ozone gradient of 24h.

Hyperparameter Search

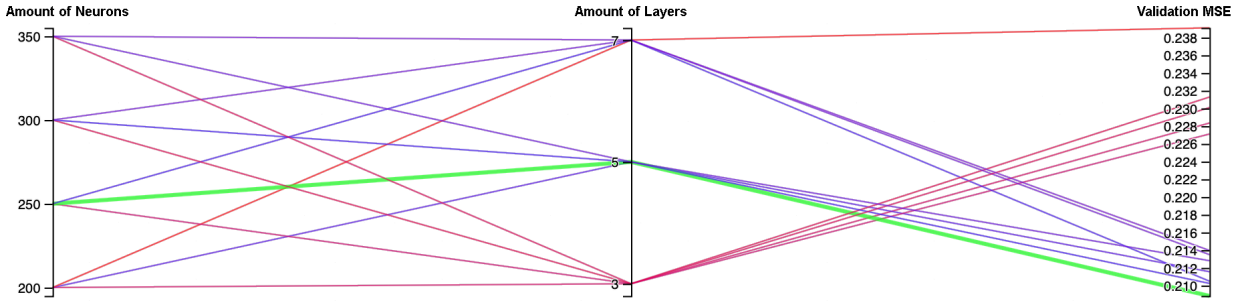


Fig. 4: HP search of the architecture parameters. The results with the lowest mean-square-error (MSE) on the right hand side shows the best combination.

HP are like adjustable screws to the NN model. These parameters can be set or adjusted by the scientist prior to or during the training. In addition to the HP, there are the NN-Parameter (NN-P), which group all internal model parameters like weights, biases and gradients.

To find a well-suited combination of HP a grid-seach was performed as can be seen in fig. 4.

The architecture of the hidden-layers need to be determined to be able to reproduce the context to be learned. In general the number of hidden-layers (L_h) and the number of neurons per layer (n_h) can be adjusted. To do so a hyperparameter (HP) seach was performed that included the number of layers and neurons per layer in addition to other adjustable parameters. The outcome of the grid-search was that 5 layers with 250 neurons per layer enable a good performance of the NN.

Training & Validation

Different sub-sets of the data-set have been separated by a random choice to enable an unbiased evaluation (see fig. 5). The data-set has a size of 23e6 data samples.

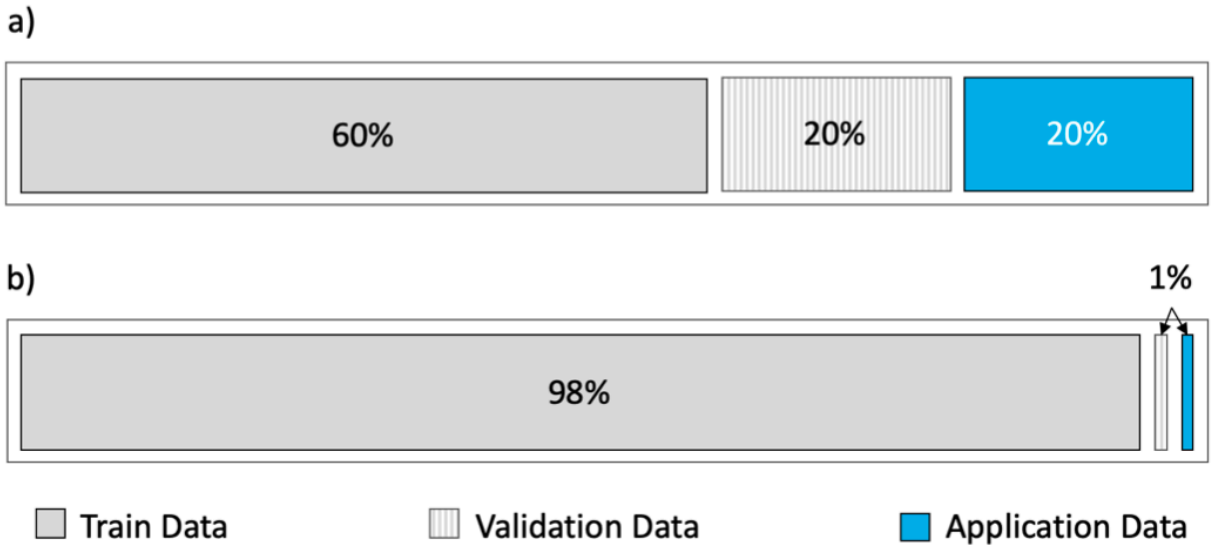


Fig. 5: Different subsets of the available data. a) traditional splitting
b) optional modified splitting that gained popularity in the era of Big Data

The supervised training has been performed for several epochs by using the training data until the validation MSE (validation data) showed no further increase for several training batches (see fig. 6). The resulting model was then stored for further evaluation. To compare several models of different architecture another set of data, called application data, was used.

Fig. 6 starts with a NN that is not trained (left: Backpropagation 0). By backpropagating the MSE and therefore adjusting the weights of the NN, the performance increases very quickly for the first backpropagations and then saturates. This lower boundary can be explained by the uncertainty of the data-set for the task of estimating the output with respect to the chosen time-step. The variance in MSE can be explained by local minimas of the hyperplane. These need to be overcome, which results in a decreased performance measurement. The bias is defined as the difference between training and validation set. For this well-tuned NN the bias is rather low.

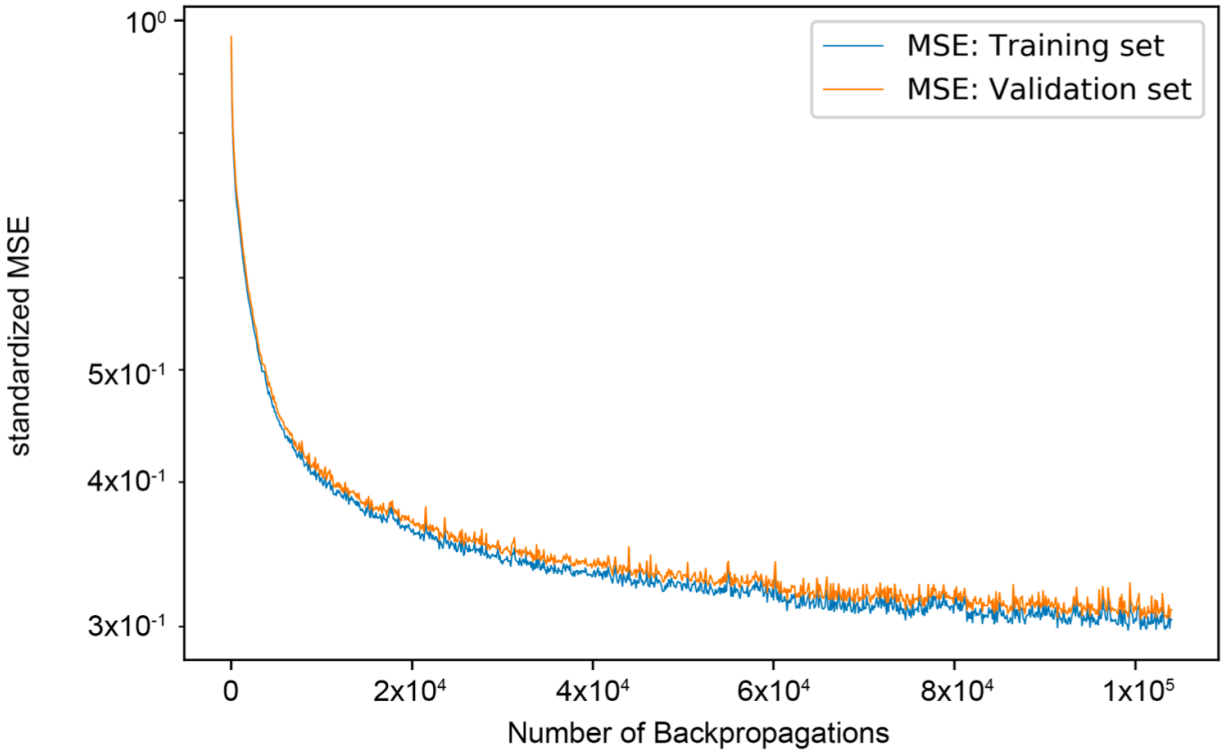


Fig. 6: Training process: the performance of each training step (backpropagation) is measured by the MSE of the training and validation set. The validation data is not used for training and offers an unbiased evaluation.

[1] D. Kreyling, I. Wohltmann, R. Lehmann, and M. Rex. The extrapolar swift model (version 1.0): fast stratospheric ozone chemistry for global climate models. Geosci- entific Model Development, 11(2):753–769, 2018.

[3] A. Paszke et al., Automatic differentiation in PyTorch, NIPS-W, 2017

PERFORMANCE COMPARED TO ATLAS

The resulting Neural Network (NN) models are not only capable of learning the context of an 11-dimensional hyperplane, but also achieve a better accuracy than the previous polynomial approach of SWIFT [1].

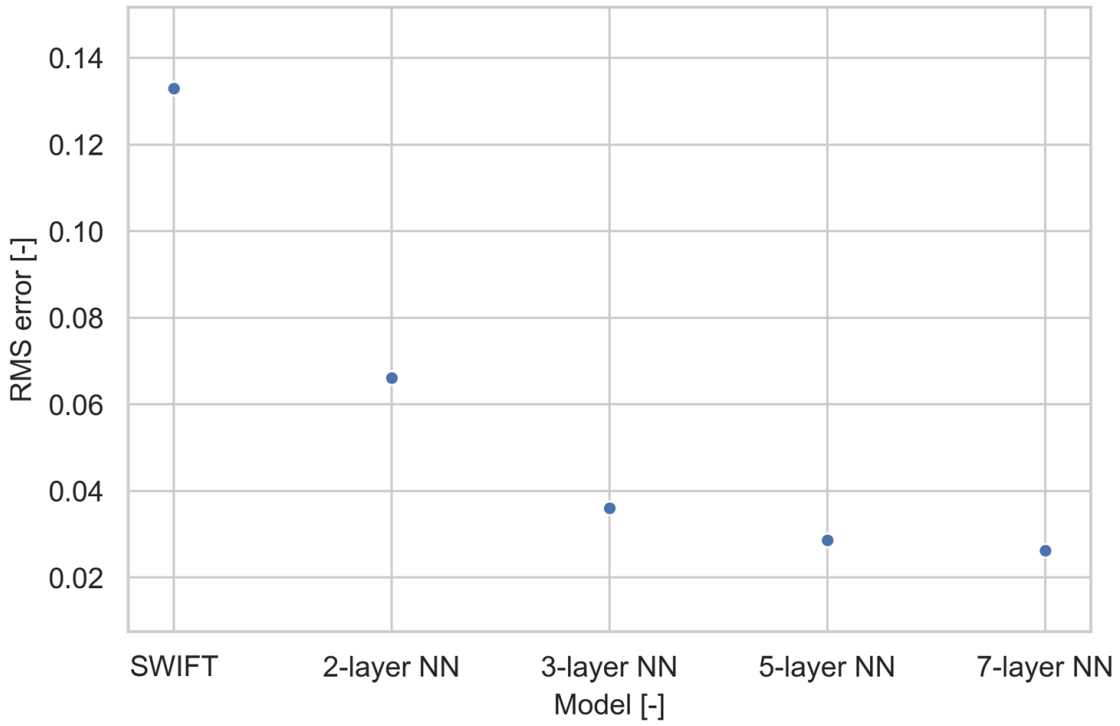


Fig. 7: Comparison of the Root-Mean-Square-Error (RMSE) of different regression models. Left the previous polynomial approach SWIFT [1] *and to the right different NN models (this project) having a different number of layers.*

Fig. 7 shows the comparison to the results of SWIFT [1]. Different NN-models having a different number of layers perform each very well. Starting with a three-layer NN the performance starts to converge.

The performance of this machine learning approach can be assessed by visualizing and comparing to the corresponding results of the ATLAS model (see fig. 8). The images on the left side depict the monthly mean dOx plotted with respect to the altitude and equivalent latitude. Reddish colors show an increase in ozone and blueish colors a decrease.
The images on the right show the difference of the results on the left side compared with the ATLAS result (upper left). The more white (zero difference) can be seen, the closer the model outcome is to the ATLAS result. The ATLAS minus NN-model plot (bottom right) shows significantly less deviations in comparison to the previous polynomial approach SWIFT (plot middle right). This is especially true for the bins at the edges. With the application of this new NN-model these areas perform similar well to other areas.

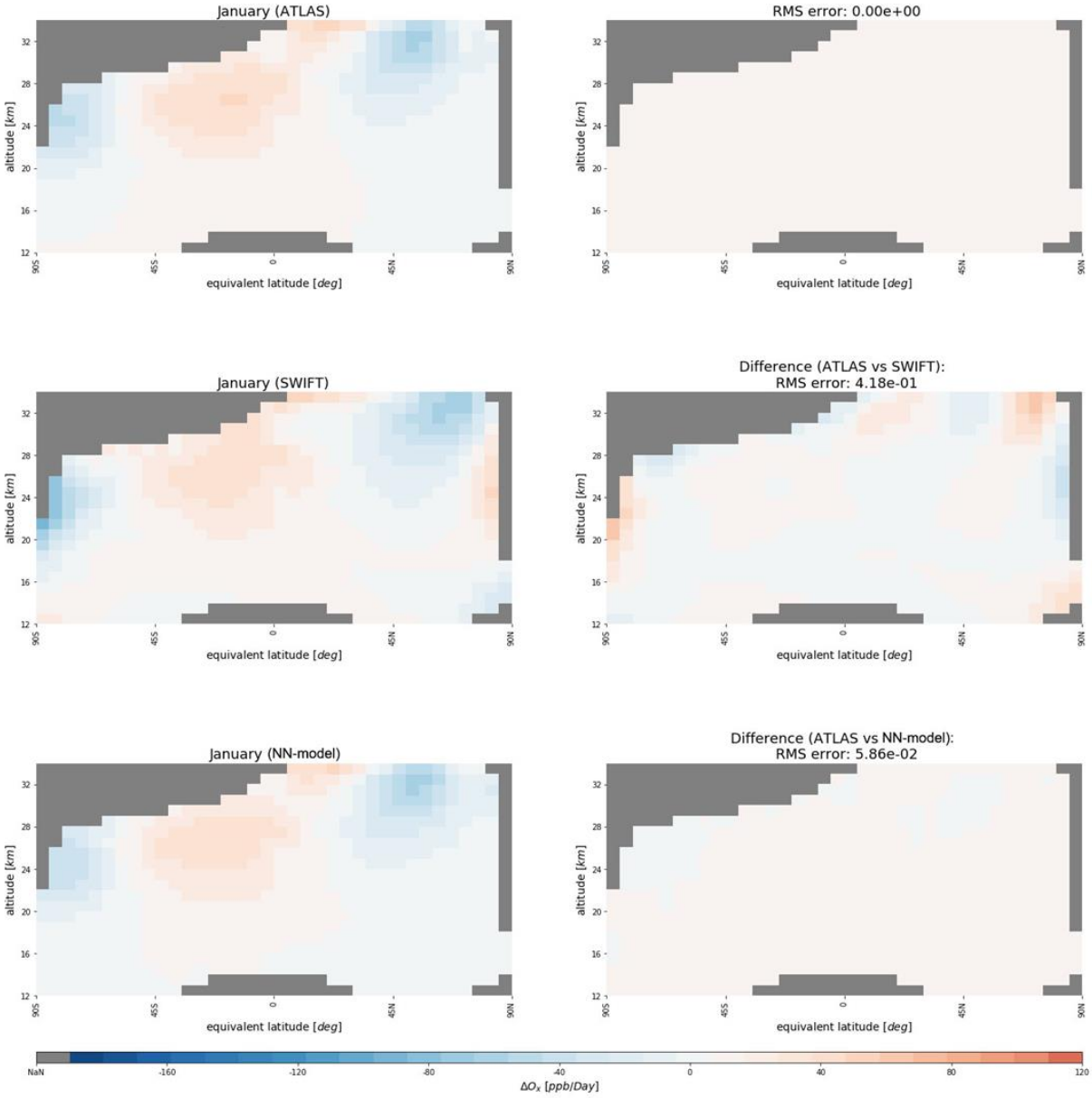


Fig. 8: Comparison for January of the raw-data source ATLAS (top row), previous polynomial based approach SWIFT *(middle row) and Neural Network approach (bottom row).*
Left column: Bin-wise (altitude vs equivalent latitude) monthly mean of the rate of change of ozone ΔOx
*Right column: Difference images (*ATLAS minus surrogate model)

[1] D. Kreyling, I. Wohltmann, R. Lehmann, and M. Rex. The extrapolar swift model (version 1.0): fast stratospheric ozone chemistry for global climate models. Geosci- entific Model Development, 11(2):753–769, 2018.

SUMMARY AND OUTLOOK

We demonstrate that a neural network (NN) approach can be used successfully to develop a fast model for stratospheric ozone chemistry. The NN model is based on a set of 11 input parameters and calculates the ozone tendency at a given grid point. The NN model is trained on data of the detailed full chemistry model of the ATLAS Chemistry and Transport Model [2]. In addition, we show that the NN model achieves a better accuracy than the existing fast stratospheric chemistry model SWIFT [1], which is based on a polynomial approach. The results of this master's thesis are used as the starting point of the PhD project with the working title 'Robust Machine Learning applied to Stratospheric Chemistry Models'.

By now only a subset of the variables available at each Lagrangian air parcel has been used. A feature importance ranking algorithm that ranks the variables by their importance will be used to further optimize the choice of data.

The application of a chemical model in a GCM requires that all input variables, which are needed by the neural network to estimate dOx, are available to the model. This requirement can be fulfilled by variables already present in the GCM, new climatologies or even further models.

A change in environmental conditions could lead to input data that is outside of the trained parameter space. Futher studies shall enable the NN to know what it has been trained on. A model that calculates the model-uncertainty beside the dOx estimate would enable a robust application of the NN. If the NN is used outside the trained parameter space, an alternative source (e.g. a climatology or another regression model) should be used.

To further reduce the calculation time, a multiple GPU application of the NN model will be integrated into a benchmarking with the previous polynomial approach SWIFT [1].

[1] D. Kreyling, I. Wohltmann, R. Lehmann, and M. Rex. The extrapolar swift model (version 1.0): fast stratospheric ozone chemistry for global climate models. Geosci- entific Model Development, 11(2):753–769, 2018.

[2] I. Wohltmann and M. Rex. The lagrangian chemistry and transport model atlas: validation of advective transport and mixing. Geoscientific Model Development, 2(2):153–173, 2009.

ABSTRACT

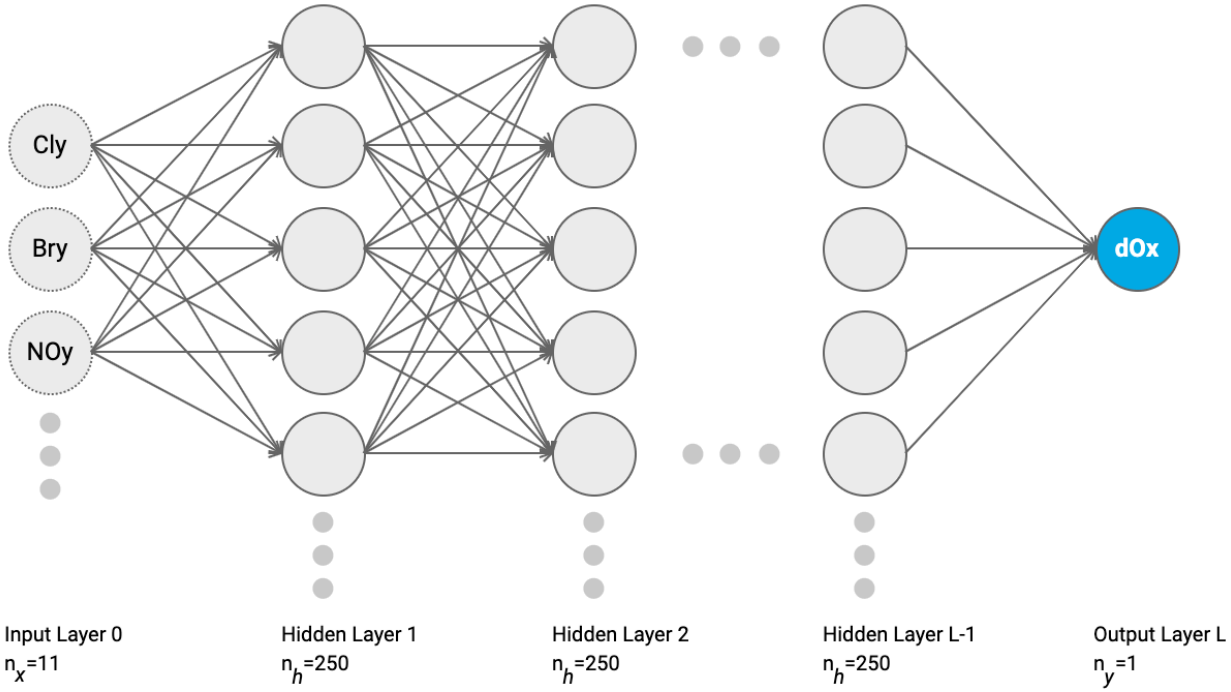
Due to the intensive ozone research in recent decades, the processes that influence stratospheric ozone are well understood. The chemistry and transport model ATLAS was developed to simulate the chemistry and transport of stratospheric ozone globally. The chemical rate of change of ozone is calculated at each model point and time step of the model by solving a system of differential equations that requires 55 input parameters (chemical species, temperatures, ...). But the computational effort to solve this complex system of differential equations is very high, and with respect to the overall limited computation time, this prevents the inclusion of ozone chemistry into ESMs.

This project proposes a data-driven machine learning approach to predict the rate of change of stratospheric ozone. To derive a data set from modelled data, ATLAS was run for several short model runs. The rate of change of ozone and 55 parameters were stored at each model point and time step. By observing the co-variances of the high-dimensional feature-space, a large data set with reduced dimensionality has been created. A supervised learning algorithm used this data set of input and output pairs to train a deep feed-forward neural network (NN). This involved the identification and optimisation of several hyperparameters and to find a well-functioning combination of depth (number of layers) and width (number of neurons per layer). In this way, the NN model capacity is optimised with respect to the data itself.

To evaluate this approach, the results were compared with another data-driven approach called SWIFT. The SWIFT model employs a repro-modelling approach that uses polynomials to approximate the rate of change of ozone.

The resulting NN model is not only capable of learning the context of an eleven-dimensional hyperplane, but also improves the RMSE by about one order of magnitude compared to SWIFT's previous polynomial approach. In addition, the deviations of the predictions at the boundaries (altitude and latitude) are significantly lower, which is a challenge for the polynomial approach.

Only fully coupled ozone climate set ups are able to consider the complex interactions of the stratospheric ozone layer and climate. This is a step towards a computationally very fast but accurate application of an interactive ozone scheme in climate models.



(https://agu.confex.com/data/abstract/agu/fm20/5/2/Paper_737325_abstract_704476_0.png)