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

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

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

PRESENTED AT:

INTRODUCTION

This project presents a surrogate model to estimate the 24h-ozone-gradient (ΔO_x) and uses the data of the Chemical and Transport Model (CTM) ATLAS [2]. This surrogate model performs much faster (minutes compared to days) than a CTM and enables the application in climate models. This represents the necessary step from understanding the chemistry and building sophisticated CTMs towards the usage of this knowledge in climate models, which is only feasibel by a much lower computation time.

Earth System Modelling (ESM) is a complex task that can be subdivided into several sub- and sub-sub-models. It is a common way to divide the underlying physics into domains of atmosphere, hydrosphere, solid-earth and ice-sheet. The atmospheric part of such models calculates the dynamics, the radiative transport, some amount of micro-physical processes and the chemistry of the atmosphere. The atmosphere contains hundreds of chemical species relevant for simulating this chemistry.

To account for such a highly dynamical and dimensional system the principle of the information technology algorithm divide and conquer is applied. This project solely concentrates on the estimation of the ozone gradient 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 chemistry and transport model (CTM) for stratospheric chemistry. It solves a coupled differential equation system using a stiff solver and a variable time-step. It includes a stratospheric chemistry scheme with 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 requires a high demand in computation performance. 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. This is a drawback, because 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 powerful surrogate models, that are tailored to the coupling into earth climate models.

[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. Machine Learning is based on a solid choice of data and it is therefore very important to understand what is fed into the NN.

Data is crucial for this project, since it is a fundamental necessity of Machine Learning. Everything the NN will eventually learn is derived from it. With millions of data samples, it can become a challenge to preserve an overview and to analyze it in a way that makes the structures of the data visible. Since the input variables have been determined wisely for the previous polynomial approach of SWIFT [1], a correlation between the input and the rate of change of ozone has already been shown. Nevertheless, further visualization methods from Data Science are discussed and used to get a deeper understanding of the data.

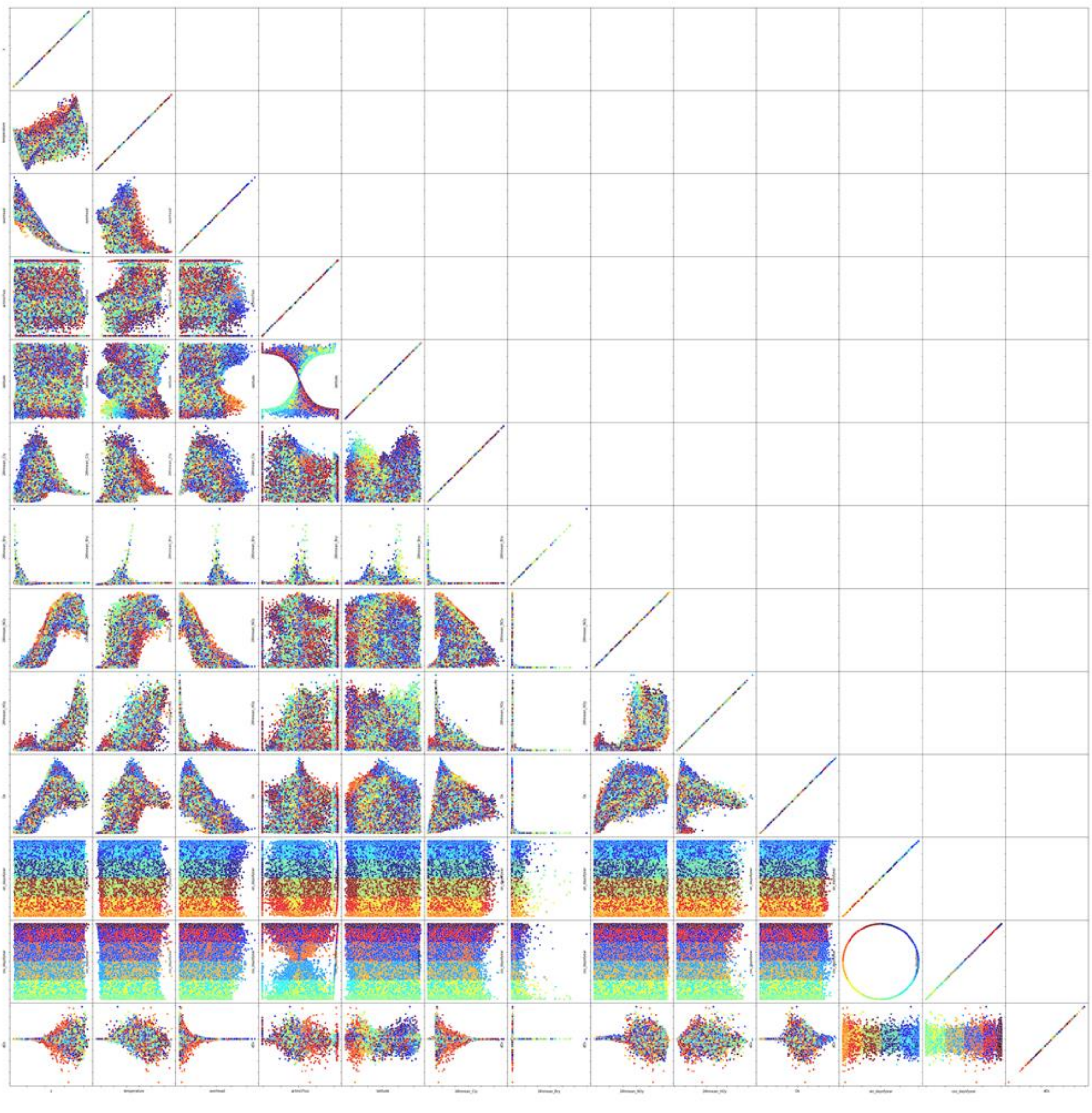


Fig. 7: Scatter-pair plot of the input and output features

Fig. 7 gives an insight into the distribution among each features. This approach involves also the day-of-year as a feature, which needed to be engineered to become able for the usage with a machine learning method. Instead of using the day-of-year as a scalar value (0..365), a vector has been created. By taking the sine and cosine of this value, this cyclical representation can be learned by the neural network model. This makes sense, because the last day and first day of a year have a distance of one day.

Fig. 8 further more explores the linear correlation of all features among each other. A negative value (blue) depicts a negative linear correlation and the positive (red) vice-versa. The greyish color depicts a very low or no linear correlation. The time of year has an impact on the linear correlation due to a change in daylength and total actinic flux per calender day.

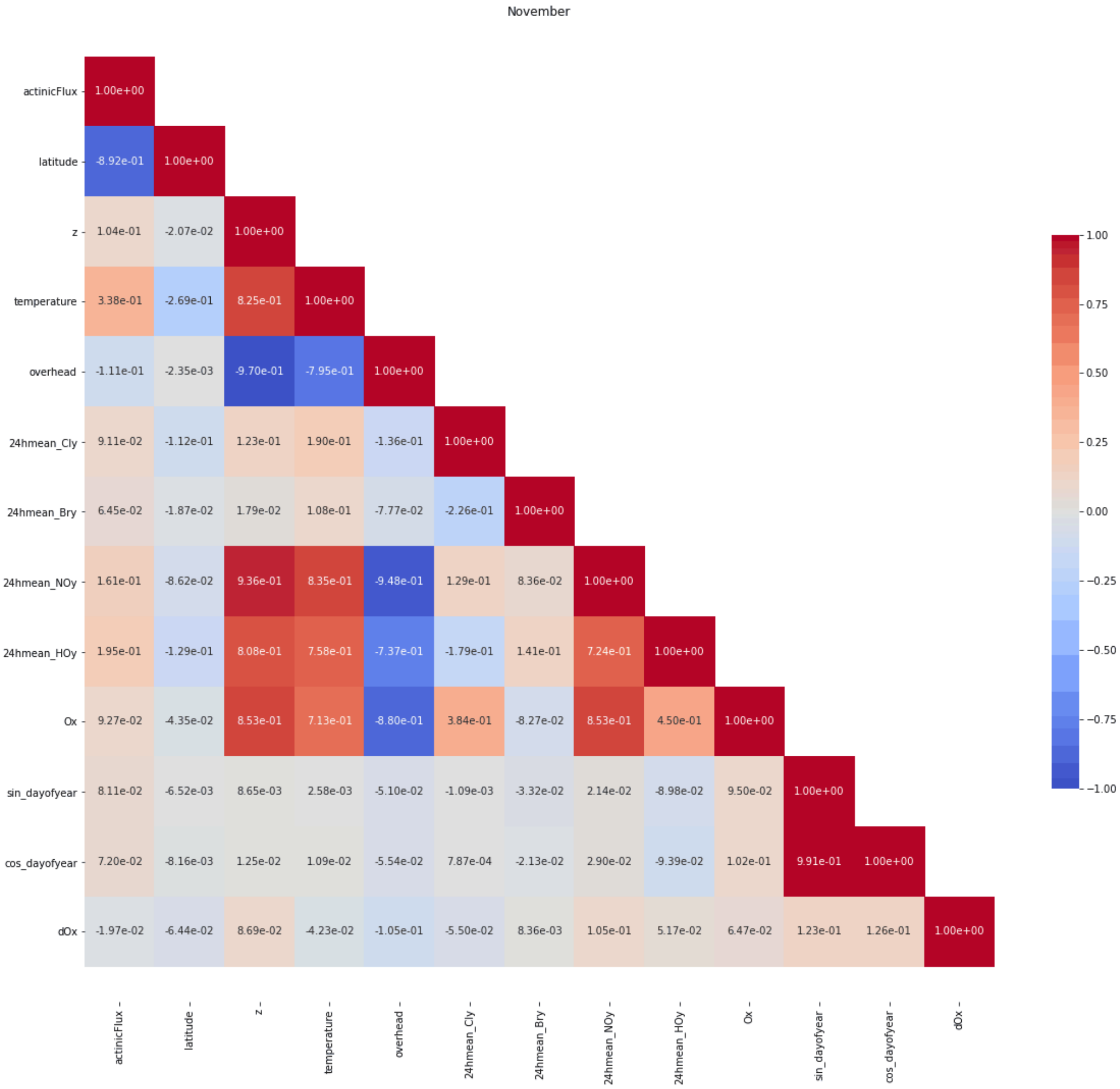


Fig. 8: Pearson Correlation Heatmap

[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.

METHODOLOGY

This approach uses a deep feed-forward Neural Network (NN) with fully-connected layers as a regression model. In general such a NN has one input-layer, several hidden-layers and one output layer (see fig. 1).

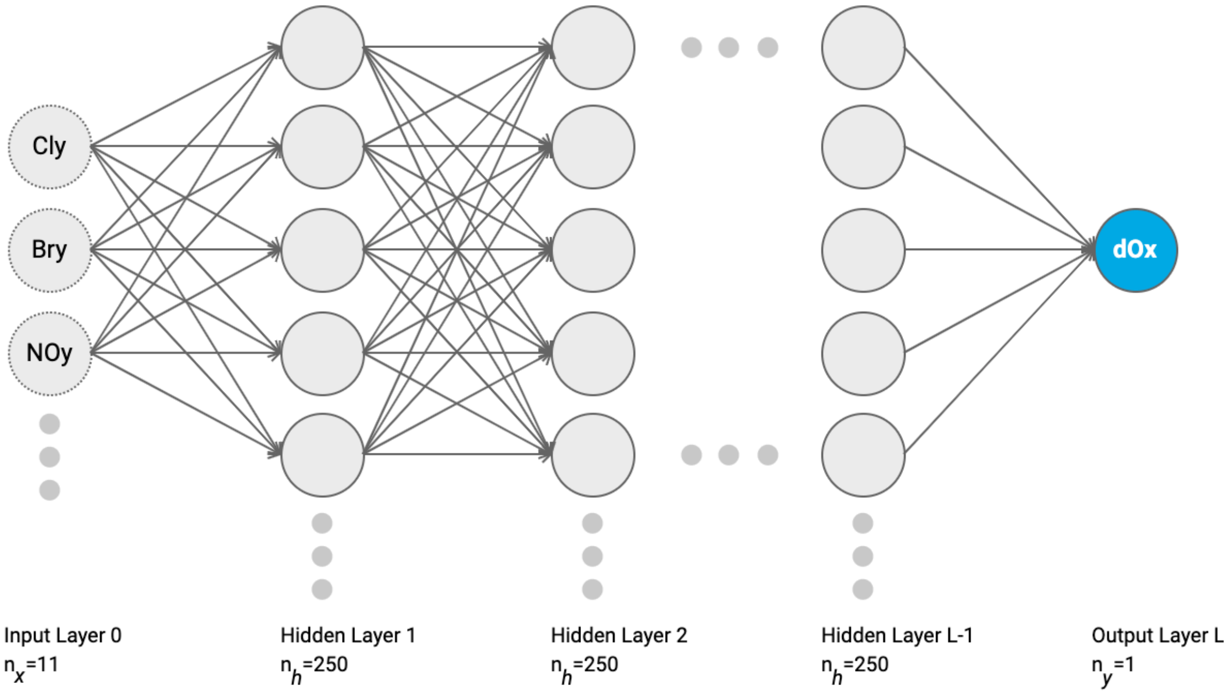


Fig. 1: Sketch of the NN-architecture

Our input layer uses eleven input features 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 appends 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 time step of the ozone gradient is not limited by this approach, but has been chosen as a daily gradient.

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 the number of layers and neurons per layer in addition to other adjustable parameters.

Hyperparameter Search

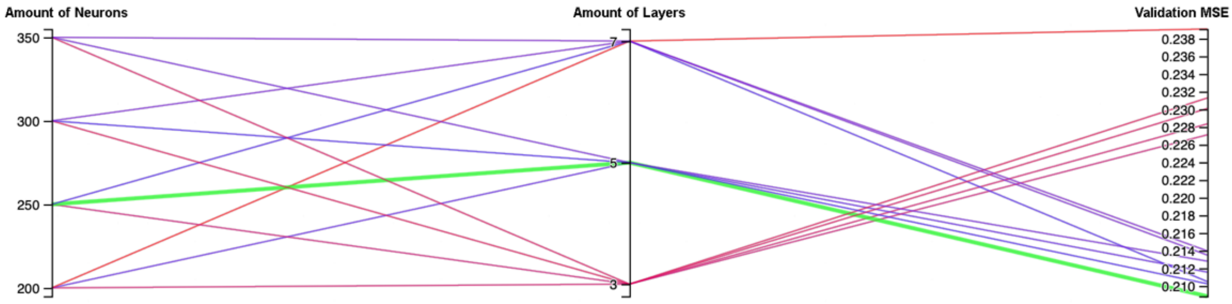


Fig. 2: HP search of the architecture parameters. The results with the lowest mean-square-error (MSE) on the right hand side shows the best combination.

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 of the nth order.

To find a well-suited combination of HP a grid-seach was performed.

Training & Validation

Different sets of the data have been separated to enable an unbiased evaluation (see fig. 3). The training process involved data that was separated from the validation data.

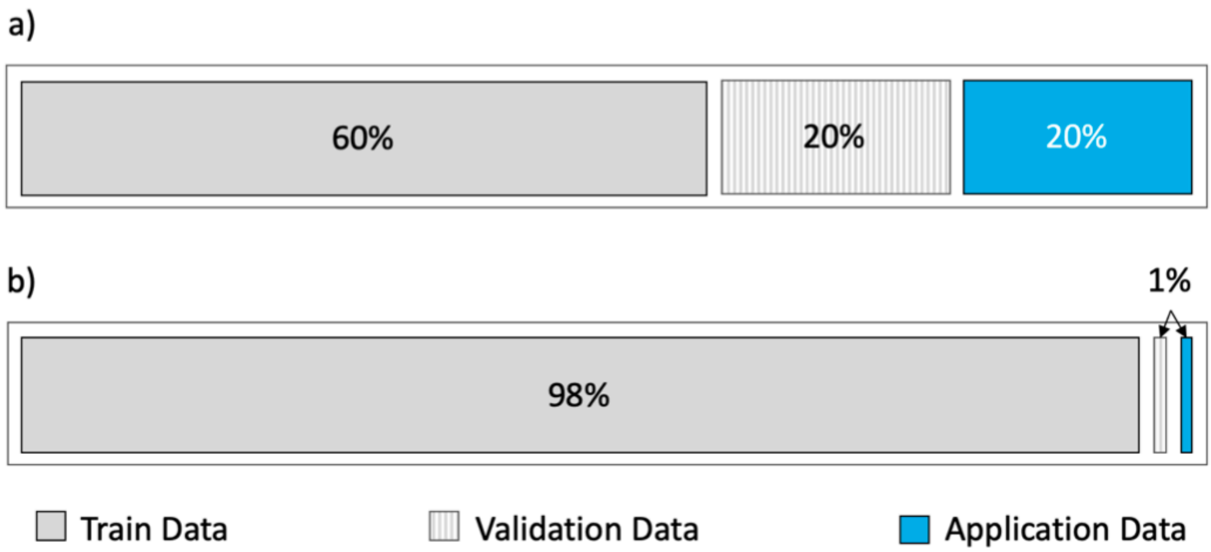


Fig. 3: Different subsets of the available data. a) traditional splitting
b) optional modified splitting due to the era of Big Data

The supervised training has been performed for several epochs until the validation MSE showed no further increase for several training batches (see fig. 4). The resulting model was then stored for further evaluation. To compare several models of different architecture a further set of data called application-data was used.

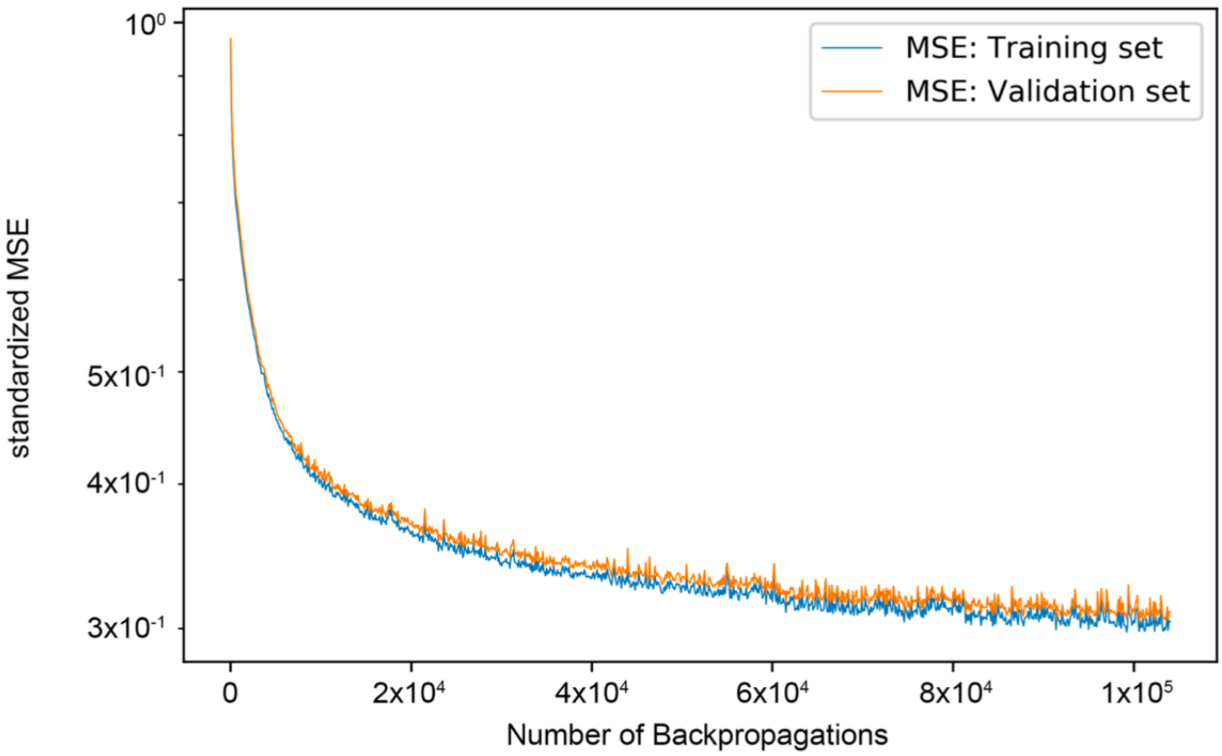


Fig. 4: Training process

[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.

EVALUATION

This project presents a whole machine learning project from the very beginning of dealing with the raw-data of a CTM-model (ATLAS [2]) to the final assessment of the results. The resulting Neural Network (NN) models are not only capable of learning the context of an eleven-dimensional hy- perplane, but also achieve a better accuracy than the previous polynomial approach of SWIFT [1].

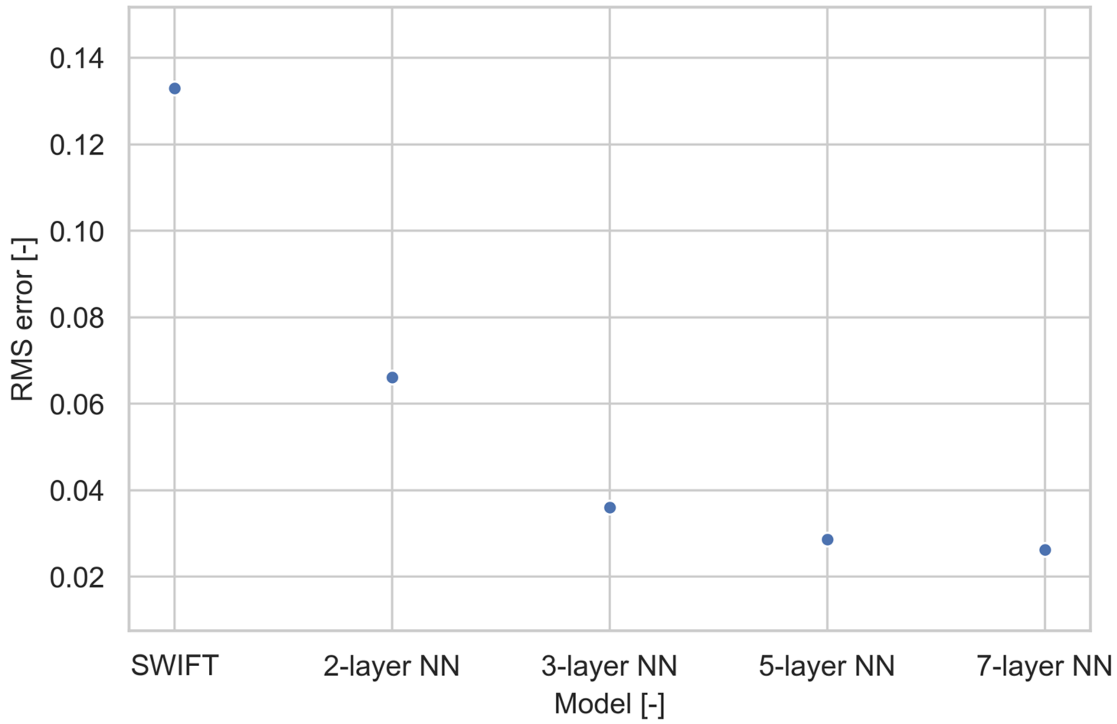


Fig. 5: Comparisom of the RMSE of different regression models. Left the previous polynomial approach SWIFT and to the right different NN models having a different number of layers of this approach.

Fig. 5 shows the comparisom to the results of SWIFT. 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 further more highlighted by visdualizing and comparing it to the raw-data of the ATLAS CTM. The difference images on the right side support the above mentioned performance of this approach (bottom). Especially the bins at the outer areas have been problematic for the approach of SWIFT. By the use of this new NN-model these areas perform similar well to the other areas..

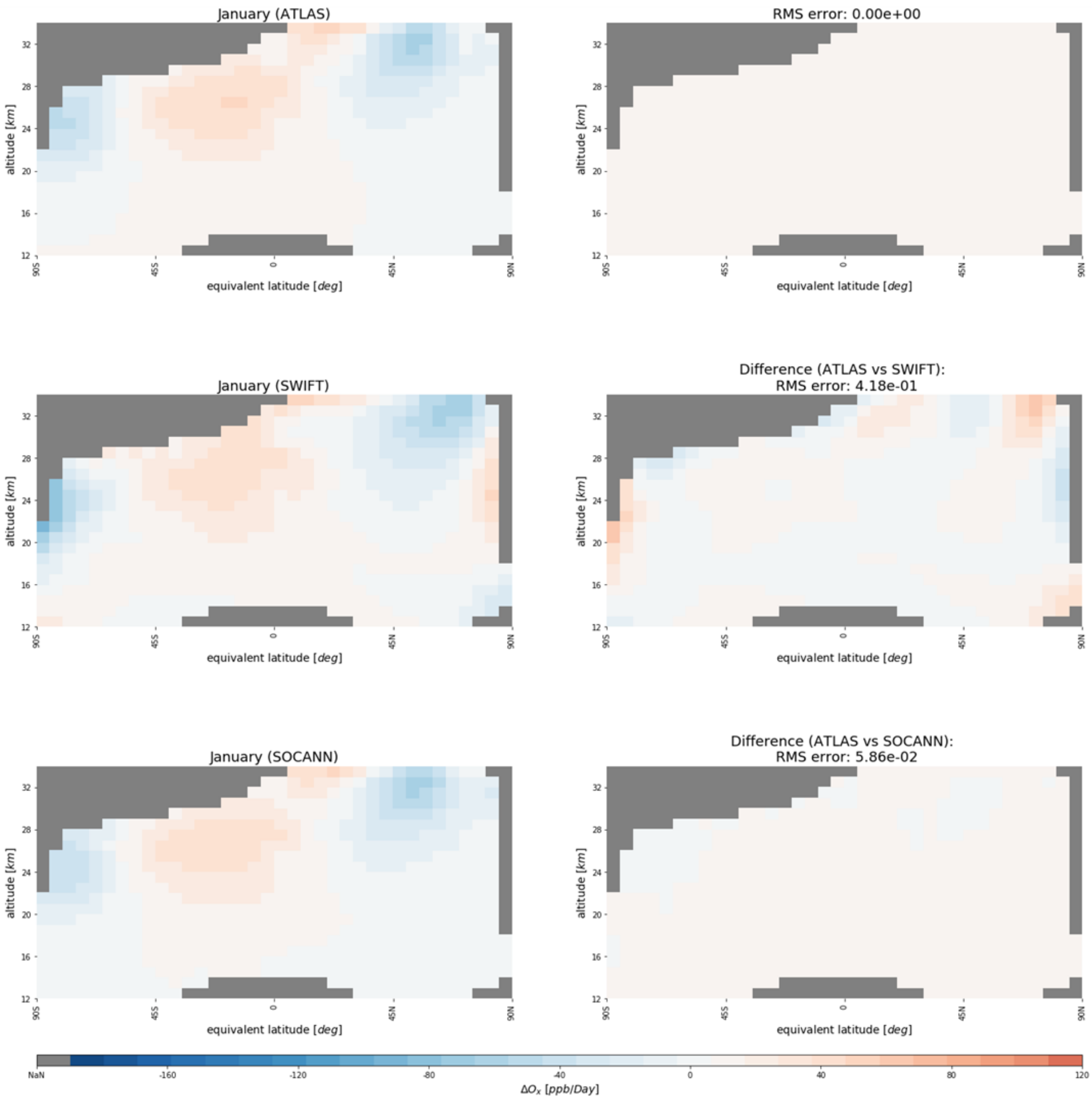


Fig. 6: Visual comparisom of the raw-data source (top), SWIFT (middle) and this method (bottom)
Left: Bin-wise (altitude vs equivalent latitude) monthly mean ΔO_x
Right: Difference images with respect to the raw-data

[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.

OUTLOOK

This Neural Network (NN) approach could successfuly learn the context of an eleven-dimensional hyperplane. This can be understood as a feasibility confirmation and has become a start of a new PHD project having the working title 'Robust Machine Learning applied to Stratospheric Chemistry Models' (Helge Mohn).

By now only a subset of the avbailable data of each lagrangian air parcel has been used. An important feature ranking involving several methods shall now be performed to further optimise the choice of data.

The application of a chemical model in a GCM requires the availability of all input features to perform the estimation of ΔO_x. Thereby the choice of input features need to be accessibel through the GCM itself or by other means. This can involve climatologies or even further models, that estimate the gradients of the necessary features.

Upcoming works can build on the results of this NN approach and extend it by investigating the sensitivity of a NN model with respect to a more or less dense sampled parameter space. Furthermore a change in environmental parameters could lead to a covariate shift. The trained hyper area of the eleven-dimensional hyperplane may not reflect the changed context. The degree of adaptability of a NN model should therefore be assessed. Data-driven models perform very well when it comes to interpolation and when they are used within the trained distribution of data. To detect whether an input of the NN is known or new data (outside the trained distribution), this PHD projects puts a focus on uncertainty estimation. Finally a model estimate will be accompanied by a uncertainty estimate that enables it to determine if the estimate can be used or not.

To reduce the computation time further more a multiple GPU application of the NN-model will be put into benchmark 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

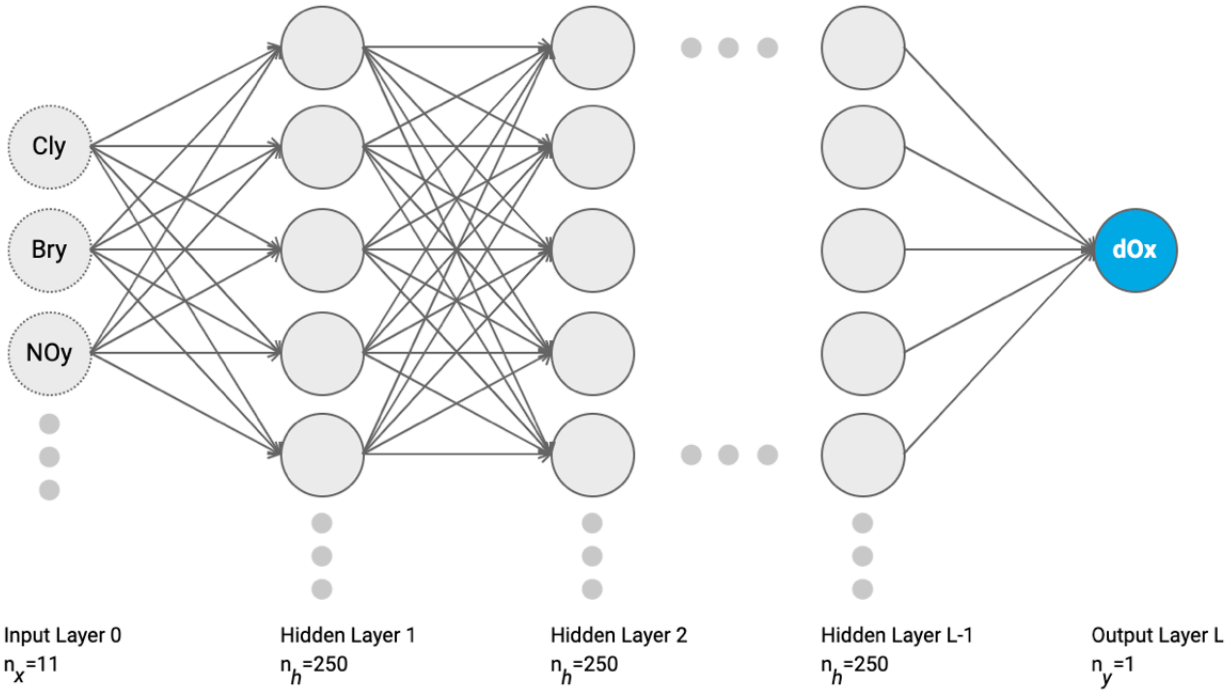
Due to the intensive ozone research in recent decades, the processes that influence stratospheric ozone are well understood. The chemistry and transport model ATLAS was developed to simulate the chemistry and transport of stratospheric ozone globally. The chemical rate of change of ozone is calculated 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)