A Deep Learning Parameterization of Gravity Wave Drag Coupled to an Atmospheric Global Climate Model

Zachary I. Espinosa[1], Aditi Sheshadri[1], Gerald R. Cain [2], Edwin P. Gerber[3], Kevin J. DallaSanta[4,5]

[1]Department of Earth System Science, Stanford University, Stanford, CA, USA
[2]Department of Computer Science, Stanford University, Stanford, CA, USA
[3]Courant Institute of Mathematical Sciences, New York University, New York, NY, USA
[4]NASA Goddard Institute for Space Studies, New York, NY, USA
[5]Universities Space Research Association, Columbia, MD, USA

Key Points:

- A neural network can accurately and stably emulate a physics-based parameterization of gravity wave drag when coupled to a climate model.
- When trained on one phase of the quasi-biennial oscillation, the emulator can generate an entire cycle of the quasi-biennial oscillation.
- The emulator captures the response of the original gravity wave parameterization to enhanced CO2.

Corresponding author: Zachary Espinosa, `zespinos@stanford.edu`

Abstract

We present a novel, single-column gravity wave parameterization (GWP) that uses machine learning to emulate a physics-based GWP. An artificial neural network (ANN) is trained with output from an idealized global climate model (GCM). We show that an ANN can learn the distribution of gravity wave drag directly from resolved flow variables. When coupled with a GCM, the ANN generates a quasi-biennial oscillation (QBO) with a realistic amplitude and period and is stable for multidecadal timescales. When forced by increasing concentrations of CO2, the ANN's climatological response is similar to that generated by the physics-based GWP. Finally, we show that the local horizontal wind components are the only essential training features for reasonable emulation and online stability. State-of-the-art GWPs differ in functional form, and are limited in their ability to incorporate observations. This work constitutes a significant step towards obtaining observationally validated, computationally efficient GWPs in GCMs.

Plain Language Summary

Atmospheric gravity waves (GWs) or "buoyancy waves" are generated by perturbations in a stably-stratified environment. They mediate momentum transport between the lower and middle atmospheres and play a leading-order role in driving middle atmospheric circulation. Due to computational constraints and a lack of observations, global climate models "parameterize" or estimate the effect of GWs on the large-scale flow. Current climate predictions are sensitive to uncertainties in these representations, particularly at the regional scale. Here, we present a novel approach to parameterizing GWs by training a neural network to emulate an existing gravity wave parameterization in a global climate model. This approach represents an appealing technique to build data-driven gravity wave schemes that can reduce existing uncertainties.

1 Introduction

Atmospheric gravity waves (GWs) play an important role in surface climate (Palmer et al., 1986) and a leading-order role in driving middle atmospheric circulation, structure, and variability (Fritts & Alexander, 2003). By transporting momentum, GWs impact the mean climatology (Bretherton, 1969; Sato & Hirano, 2019). They are critical for variability, impacting the jet stream and storm tracks (Fritts & Nastrom, 1992) and stratospheric dynamics (Antonita et al., 2007; Kang et al., 2018; Limpasuvan et al., 2012). Since much of the gravity wave spectrum (10^1 to 10^5 kilometers) is too fine to be captured at current model resolutions, models typically mimic its effects on the resolved circulation via explicit Gravity Wave Parameterizations (GWPs) (Fritts & Alexander, 2003; Richter et al., 2010).

Incorporating a realistic representation of these effects in numerical models, however, remains challenging for several reasons: i) The absolute magnitude of GW momentum flux is not well-constrained by observational or intermodel studies (Geller et al., 2013). ii) GWs are generated by a variety of sources, including orography, convection, and frontogenesis (Fritts & Alexander, 2003), but the representation of their sources is not uniform among models. iii) For a given source, the details of the GWP can vary greatly between models (Butchart et al., 2018), and even small changes within the same model can lead to diverging regional climate projections (Schirber, 2015). iv) The horizontal propagation of GWs is usually neglected in parameterizations, which is nonphysical and has an impact on the middle atmosphere (Xu et al., 2017). v) There tends to be a compensation between resolved and unresolved waves (Cohen et al., 2013), complicating observational and intermodel comparisons.

Despite these limitations, GWPs can be optimized towards maximizing global forecast skill scores (Alexander et al., 2019) or "tuned" to reduce climatological biases (Garcia et al., 2017). However, these limitations become apparent in simulations of future climate. Projections of the tropospheric and stratospheric circulations' response to anthropogenic forcing are sensitive to uncertainties in GWPs (Sigmond & Scinocca, 2010; Polichtchouk et al., 2018). Such limitations indicate the need for the continued development of GWPs, preferably with observational validation, computational efficiency, and minimal bias stemming from underlying physics-based assumptions.

An alternative approach to physics-based parameterization has emerged in the form of machine learning (ML). ML has been employed to parameterize processes such as convection (Rasp et al., 2018; Gentine et al., 2018) and radiation (Brenowitz & Bretherton, 2018; Roh & Song, 2020). Relatively little work has been done to investigate GWPs with ML technqiues. Matsuoka et al. (2020) demonstrated that a convolutional neural network can estimate the GW structure over Hokkaido, Japan when trained on high-resolution reanalysis data and Chantry et al. (2020) used an artificial neural network (ANN) to emulate a non-orographic GWP in numerical weather forecasting.

We perform a novel investigation into the efficacy of ML as a GWP in a GCM. We demonstrate that the drag due to breaking GWs and resultant circulation can be faithfully generated using an ANN that is trained using data from a global atmospheric model. The ML model, which we call WaveNet, differs from previous work in that it is trained using output from a GCM, provides global coverage, is stable for decadal integrations when coupled to the GCM, and accurately reproduces the large-scale atmospheric circulation. The potential value of a data-driven GWP lies in its computational efficiency after training, its ability to be trained and evaluated using an arbitrary amount of input data, which may include observations, reanalysis, and targeted integrations, and its potential to be appropriated into models containing GWPs thought to be less accurate.

2 Data

We train the ANN to emulate the M. Alexander and Dunkerton (1999) GWP as incorporated into an atmospheric model of intermediate complexity, the Model of an Idealized Moist Atmosphere (MiMA) (Jucker & Gerber, 2017; DallaSanta et al., 2019). MiMA captures key dynamical features of the stratosphere-troposphere system that depend critically on GWs at a resolution comparable to state-of-the-art stratosphere resolving GCMs. For additional details refer to Jucker and Gerber (2017) and DallaSanta et al. (2019). MiMA is integrated with T42 spectral resolution (triangular truncation at wavenumber 42, roughly equivalent to a 2.8-degree grid) in the horizontal and 40 vertical levels, with a model lid at 0.18 hPa and 23 levels above 100 hPa.

The implementation of the M. Alexander and Dunkerton (1999) GWP, hereafter referred to as the AD99 scheme, is based on its formulation within the GFDL Flexible Modeling System. The AD99 scheme aims to capture the effect of non-orographic gravity wave drag. It is the same scheme employed by GFDL Atmospheric Model 3 (Donner et al., 2011), except modified by Cohen et al. (2013) to ensure that all momentum that reaches 0.85 hPa is uniformly distributed to layers above the stratopause. The scheme assumes a Gaussian spectrum of GWs, discretized as a function of phase speed and launched from the upper troposphere (315 hPa).

A broad spectrum of phase speeds is chosen, with a half width of 35 m/s, and centered around the speed of the zonal wind at the launch level. The total amplitude of the momentum stress is 0.0043 Pa. The broad spectrum ensures that there is a rich source of GWs at MiMA's lower boundary. The momentum associated with each wave is deposited at its linear breaking level. The intermittency of observed GWs is taken into account via a scaling parameter: the breaking level is based on the behavior of a large wave, indicative of the median amplitude of the distribution, but the momentum flux is rescaled to provide the momentum deposition associated with an average wave. This better captures the true breaking level of GWs, but smooths out the momentum deposition in time.

The chief idealization of the scheme is in our choice of source spectrum. We assume a uniform spectrum that varies only with respect to the zonal winds at the source level. This crudely accounts for filtering of the wave spectrum by the troposphere and was critical for the simulation of the QBO.

We utilized five years of six-hourly output from one integration of MiMA, yielding 11,796,480 training samples per year. The second year of output is used for training, while all others are reserved for testing. Not all runs utilize the full year of training data, and discrepancies are specified per experiment. All training data are standardized by removing the mean and scaling to unit variance, calculated for each variable across each pressure level. Test data was similarly standardized using the mean and variance calculated from training data. Output variables were inverse transformed before presentation.

3 Neural Network Architecture

An artificial neural network (ANN) is a computing system of interconnected layers of computational nodes. For a detailed review of ANNs, we recommend Brenowitz and Bretherton (2018). We trained two ANNs to generate zonal and meridional drag separately. The ANN architecture is described in Figure S1.

The input layer of both ANNs accepts a single vertical column of resolved flow variables from MiMA (Table S1). The output layer produces a single vertical column of 33 gravity wave drag (GWD) values, corresponding to the upper 33 model levels (approximately 0.18hPa to 436hPa). This includes the upper three sponge layers and three layers below AD99's average launch level. During online simulations with the ANN, no drag is specified below level 33, i.e. zeros are appended to WaveNet's output to extend to the full length of the vertical column. The result of each node in the output layer does not pass through an activation function. For all other nodes, we use the Rectified Linear Unit (ReLU). In total, this ANN architecture contains 3,848,481 trainable parameters when using the full set of resolved flow variables and shifts slightly, with a lower limit of 3,806,753, when training with a subset of resolved flow variables.

The loss function - in our case, the logcosh error - is computed for a minibatch of 1,024 training samples that are drawn from a pseudo-shuffled training dataset. The logcosh error is defined as

$$L(y, y^p) = \sum_{i=1}^{n} log(cosh(y_i^p - y_i))$$

where n is the size of the dataset and y_i and y_i^p are the i^{th} truth and prediction, respectively. The logcosh error is found to consistently outperform mean-squared and mean-absolute loss functions. Parameter updates are performed according to Adam optimization (Kingma & Ba, 2014). We started each training session with a learning rate of 10^{-3} and reduced it when improvement plateaued for more than 5 epochs, with an epoch defined as 1,500 batches. We stopped training when performance plateaued for more than 10 epochs, which occurred at 200 epochs on average. We did not perform any regularization. All weights are initialized using Xavier initialization (Glorot & Bengio, 2010).

4 Offline Performance

4.1 Large-Scale Circulation

We start by training WaveNet on one year of MiMA output and testing it offline on the three years proceeding and one year preceding the training period. In order to assess how well WaveNet captures GWD associated with large-scale circulation, we an-

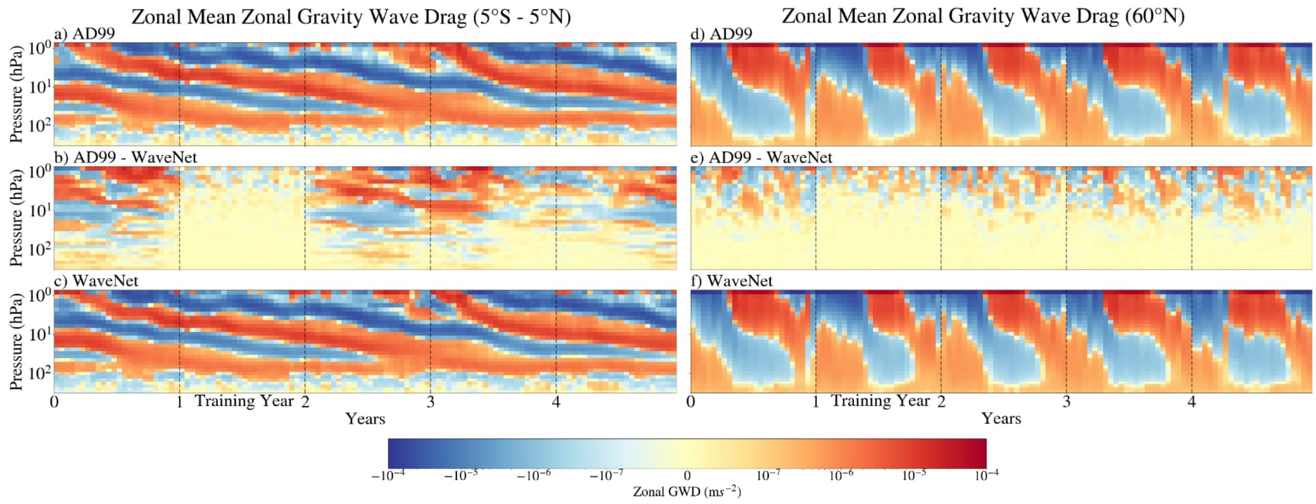


Figure 1. Pressure-Time profiles of 15-day averaged zonal drag at 5°S - 5°N and 60°N for AD99 (a,d), WaveNet (c,f), and their difference (b,e) during offline, uncoupled testing. Vertical dashed lines separate years. The ANN is trained on the second year and tested on all other years. The westward (blue) and eastward (red) bands in the left column correspond to drag associated with the QBO and in the right column correspond to drag associated with the seasonal cycle of the polar vortex.

alyze the vertical profile of drag tendencies at 5°S - 5°N and 60N (Figure 1). WaveNet is trained on one year of global data (year 2), containing the westward phase of the QBO and a full seasonal cycle of the polar vortex. The ANN captures the changes in GW driving associated with the eastward phases preceding and proceeding the training period. The result suggests that WaveNet can transfer learning from regions (horizontally or vertically) containing eastward wind samples to a region where it has not experienced similar samples during the training period.

While comparison between Figure 1a and Figure 1c shows that WaveNet can capture the gross features of the "out-of-sample" periods, differences between AD99 and WaveNet are larger at these times, as seen in the middle panel. Notably, the errors decrease around month 48, when the winds are more similar in structure to the training period. This suggests more varied training data or regularization techniques may improve WaveNet's performance. Movies S1-S4 show strong similarities on a global scale between the horizontal GWD generated by WaveNet and AD99 at 10 hPa and 100 hPa for a six month time series of test data. These results are reassuring with respect to WaveNet's potential to globally generalize from regional observations.

4.2 Interpretability of the ANN

To further evaluate the quality of predictions, we calculate the R^2 coefficient of determination for a one year time series of test data at each spatial location and average for each vertical level. R^2 is defined as one minus the proportion of the sum of squares

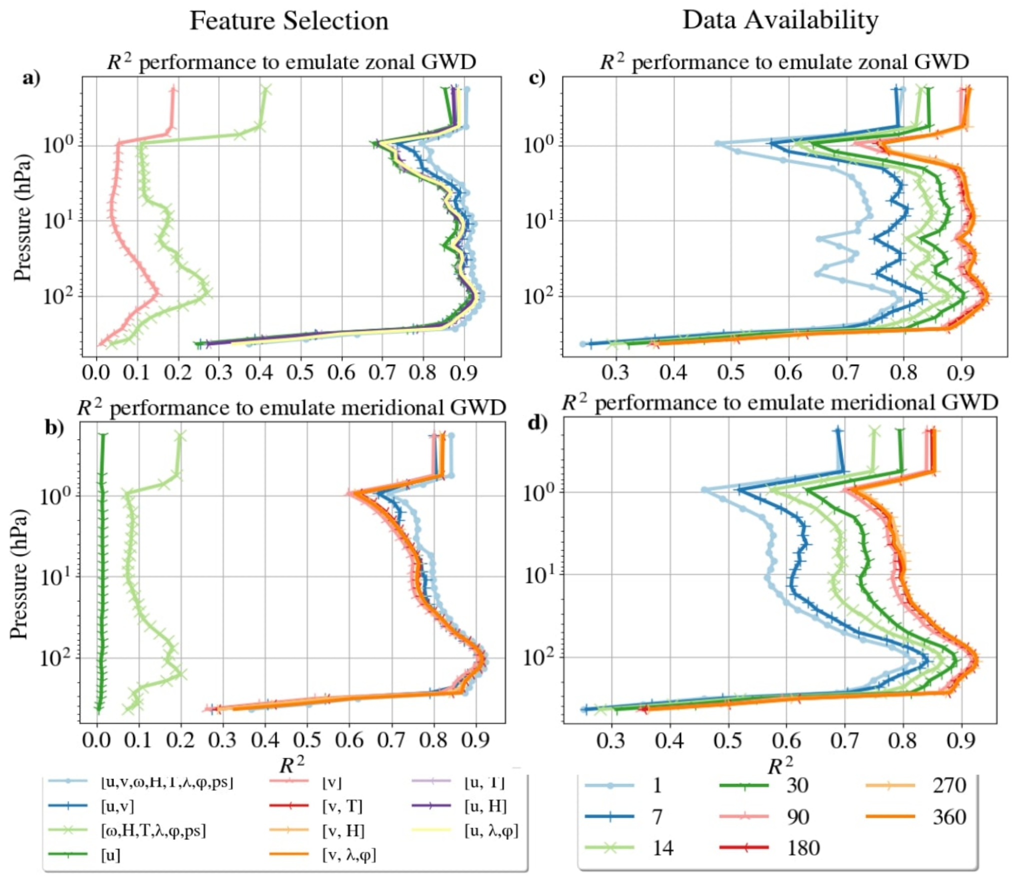


Figure 2. R^2 values are computed for one year of test data at each spatial location and averaged for each model pressure level. Presented are the R^2 values for the feature experiments (a and b), where ten subsets of the full input variable set are provided during training, and data availability experiments (c and d), where the size of the training set is restricted to the number of days specified in the legend.

of residuals to the total sum of squares. We also train WaveNet using ten distinct subsets of the full resolved flow variable set to determine which training features are most critical (Figure 2ab). When trained with the full feature set (light blue, circle inscribed lines), the average pressure weighted R^2 value above the source level is .91 for zonal and .88 for meridional GWD predictions. Below this level, performance significantly degrades as AD99 overwhelmingly outputs trivial, nonphysical GWD. The zonal (meridional) wind component is the only flow variable necessary to retain 94.4% (96.1%) of the ANN's performance for zonal (meridional) GWD predictions.

To further investigate the role of horizontal wind features, we calculate Shapley Additive Explanations (SHAP) (Lundberg & Lee, 2017a). The SHAP value for a feature is the weighted sum of the conditional expectation of the marginal contribution of including that feature in the prediction. The SHAP values for this study are calculated using Deep SHAP, an approximation technique that combines DeepLIFT with Shapley values from collinear cooperative game theory (Shrikumar et al., 2017; Lundberg & Lee, 2017b). Figure S2 shows that on average the horizontal wind components on levels directly above and below the level of prediction have the largest contribution to the model's prediction. This is consistent with our physical understanding of GWs, where dissipation is linked to critical levels (i.e., where the phase speed of a wave is equal to the speed of the background flow). This spatially local dependence suggests that our approach may generalize well to observational datasets that contain measurements at a small range of pressure levels, e.g., observations from superpressure balloons (Podglajen et al., 2016; Lindgren et al., 2020).

4.3 Sensitivity to Amount of Training Data

To understand how performance degrades as less training data is made available to the ANN, we incrementally decrease the number of training samples from one year to one day (Figure 2cd). To account for the effects of seasonality, we sample uniformly in space and time from one year of training data to generate a subset with coverage of the entire seasonal cycle. We find that a quarter of a year of data (equivalent to three months or 2.9 million samples) is sufficient to retain 98% of WaveNet's performance compared to training with one year of data. This suggests the promise of a data-driven GWP to be trained and validated on observational datasets of similar resolution and size, as well as to generalize to simulations that more realistically model the physics of GWs.

5 Coupling WaveNet with MiMA

Successful offline emulation of key atmospheric structures does not necessarily engender good online performance. It has been shown that small offline errors can trigger instabilities or accumulate to generate large errors (Brenowitz & Bretherton, 2018). Here, we study the online stability and performance of the emulator by replacing AD99 with WaveNet in MiMA. We couple the ANN, written in Python, with MiMA, written in Fortran, by using an interoperability package, forpy. Different approaches can be taken to couple ML algorithms with Fortran (Chantry et al., 2020; Ott et al., 2020). The main advantage of forpy is that it supports complex network structures and requires nominal effort to recouple with alternate ML schemes. The primary disadvantage of this interface is that it is slow compared with alternate approaches. When coupled, WaveNet slows the entire GCM by roughly 7.5x, a result of the forpy interface and WaveNet's size. The runtime may be further optimized by switching to an alternate coupling interface, performing quantization and pruning, reducing the number of trainable weights, and migrating to a GPU compatible GCM. Optimizing and analyzing WaveNet's run-time is a subject of on-going study and beyond the scope of this work. For all subsequent online analysis, we use the versions of WaveNet trained with one year of data that accept as input (u,T; zonal ANN) and (v,T; meridional ANN).

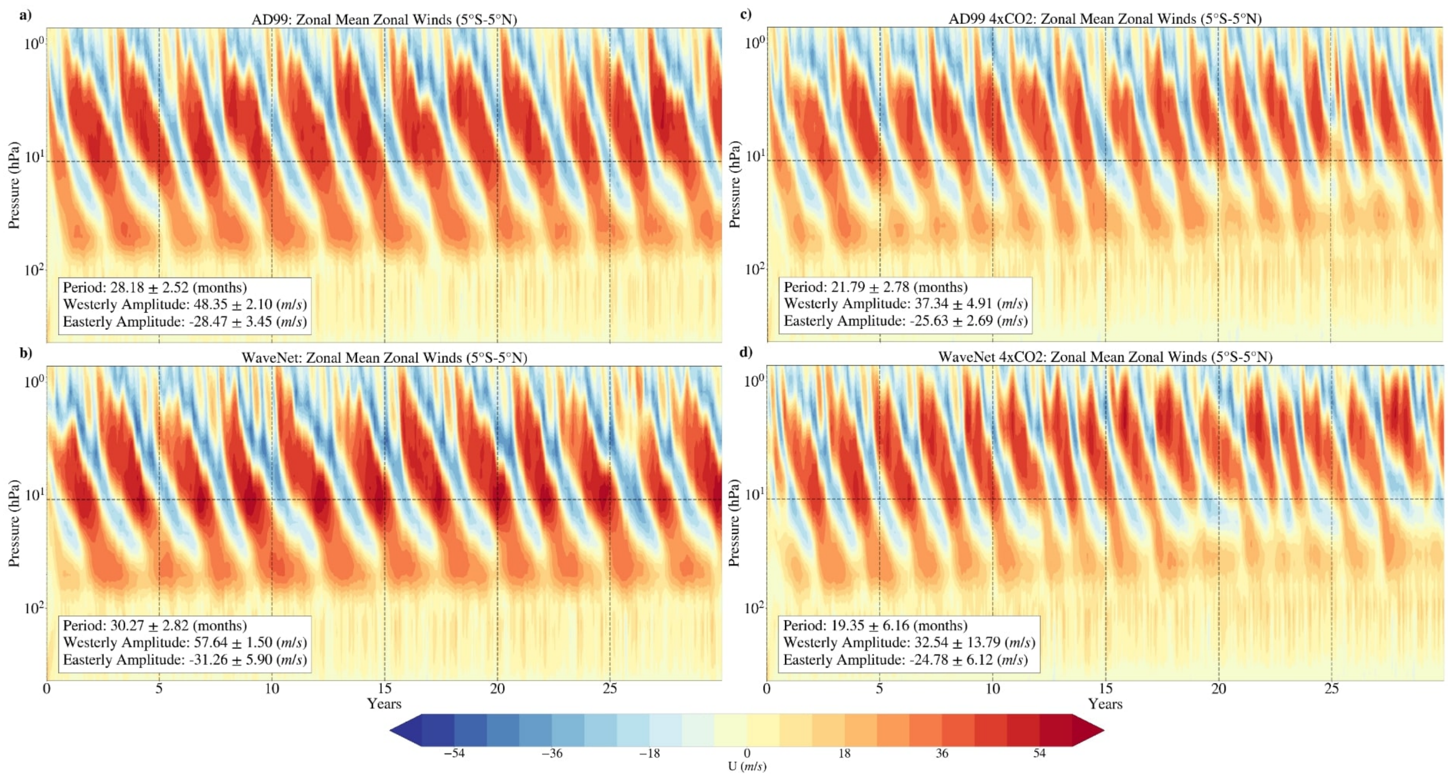


Figure 3. Pressure-Time profiles of the zonal mean zonal wind, averaged from between 5° S and 5° N and smoothed with a 15-day low pass filter show the behavior of the QBO in 30 year integrations of (a) the control version of MiMA with the AD99 parameterization, (b) the model coupled with WaveNet, (c) a 4xCO2 integration with the AD99 parameterization, and (d) a 4xCO2 integration coupled with WaveNet. Vertical dashed lines separate 5 year segments. The eastward (red) and westward (blue) bands correspond to winds associated with breaking GWs in opposite phases of the QBO. The QBO period and amplitudes are calculated using the transition time (TT) method. The dashed-horizontal line in each panel delineates the model level (≈10 hPa) where the TT method is used.

6 Online Performance

To demonstrate the ability of WaveNet to act as a faithful emulator online, we complete a 30-year integration of MiMA coupled to WaveNet using two configurations. Figure 3 shows pressure-time profiles of 15-day averaged zonal mean zonal winds between 5°S and 5°N for the two configurations for AD99 (a,c) and WaveNet (b,d). Following the transition time (TT) method described in Richter, Anstey, et al. (2020), we calculate the period of each QBO cycle as the difference in time between every other phase change for the 5°S to 5°N averaged zonal mean zonal wind time series at roughly 10 hPa. The westerly (easterly) amplitude is taken as the maximum (minimum) value of the time series for each QBO cycle. The numbers in the lower left corner of each panel show these statistics calculated for years 1 through 30 with the spin-up time omitted. From the same experiments, we plot the average zonal winds and temperature as a function of pressure and latitude in Figure 4.

6.1 Baseline Emulation

The first experiment utilizes the same model configuration as was used to generate training data and serves to evaluate WaveNet's ability to emulate AD99 while online. For this scenario, WaveNet produces a QBO with an average period of 30.27 ± 2.82 months, slightly longer than the 28.18 ± 2.51 month period generated by AD99. This difference can easily be observed in Figure S3 which shows that the error magnitude oscillates with out-of-phase samples. WaveNet produces remarkable consistency in the overall height and fine-scale features of the QBO. The asymmetry between easterly and westerly amplitudes, which is well appreciated in models and observations (DallaSanta et al., 2021), is fully captured by WaveNet. The amplitude fidelity is limited by the training period. The simulated westerlies are too strong compared to those generated by AD99, as the ANN was trained over an easterly dominated year; however the difference between easterly amplitudes is not statistically significant. This suggests that training an ANN on limited observations (less than one QBO cycle) may provide spectral insight even if the entire amplitude cycle is not well-sampled. WaveNet produces a climatology similar to AD99 (Figure 4). WaveNet generates a stronger polar vortex (Figure S4) and cooler temperatures (Figure 4j) in the upper atmosphere in both poles. The general agreement between WaveNet and AD99 observed in these results is a strong indicator that WaveNet can act as a faithful emulator of AD99.

6.2 Forcing Emulation

To be useful, a data-driven GWP must accurately emulate GWD across a range of model configurations and scenarios not fully captured in training data. The second experiment presented here serves as a first step towards understanding WaveNet's sensitivity to model configuration. Here, we completed a coupled run using four times preindustrial concentrations of CO2 (1200ppm) or roughly triple that of the baseline run (390ppm), which was optimized for early 21st century scenarios. Figures 3c and 3d show that when using either AD99 or WaveNet the average period and amplitude of the QBO decrease and the variance increases. The amplitude decrease is common among state-of-the-art models (Richter, Butchart, et al., 2020) and is attributed to increased tropical upwelling, which is separate from GWD. However, we observe that the QBO generated by WaveNet has a larger response to forcing and enhanced variance.

Models differ regarding the period projection; and a mechanism is not yet clear, as the intermodel correlation between period decrease and increased gravity wave flux vanishes when only fixed-source GWPs are considered (Richter, Butchart, et al., 2020). Thus, it is significant that AD99 and WaveNet project similar changes to QBO period, highlighting the role of changes in the background state. Our idealized model results point to further investigation with more complex models, e.g. within the GFDL Flexible Modeling System. Figure 4 and Figure S4 show that WaveNet and AD99 have a very similar climatological response to external forcing (e.g. both produce stronger polar jets and similar tropical changes). Although the regions that are statistically distinguishable vary, likely due to changes in regional variability, the fraction of total distinguishable points remains roughly the same for the 4xCO2 (36% u, 31% temp) and baseline scenarios (39% u, 28% temp).

7 Discussion and Conclusion

We have demonstrated that an ANN can skillfully learn the salient features of GW momentum transport directly from resolved flow variables. We have produced an emulator that can stably run when coupled with a simplified global atmospheric model for multidecadal time-periods and reproduce the large-scale circulation. We have shown that the most important input features for WaveNet's predictions are the horizontal wind components local to the vertical level of prediction, and we have demonstrated that WaveNet can generalize using limited training data, online and offline. Finally, WaveNet produces a climatological response to CO2 forcing similar to that generated by AD99. The success of these experiments implies that an approach like WaveNet may open a new avenue by which the advantages of high-resolution GW simulations (Remmler et al., 2015) or observational datasets (Lindgren et al., 2020) can be incorporated into current GCMs.

There are, however, a number of challenges that may emerge before the advantages of an approach like WaveNet can be fully realized in a GCM. First, ANNs do not inherently conserve energy or momentum. Second, the lack of interpretability of ANNs may serve as a substantial barrier to their widespread adoption. Additional effort is necessary to consider how WaveNet's behavior may relate to the GW dispersion relations. Third, WaveNet in its current setup is too slow. In order to make coupling WaveNet with a GCM computationally feasible, a cost-performance analysis is planned to reduce WaveNet's

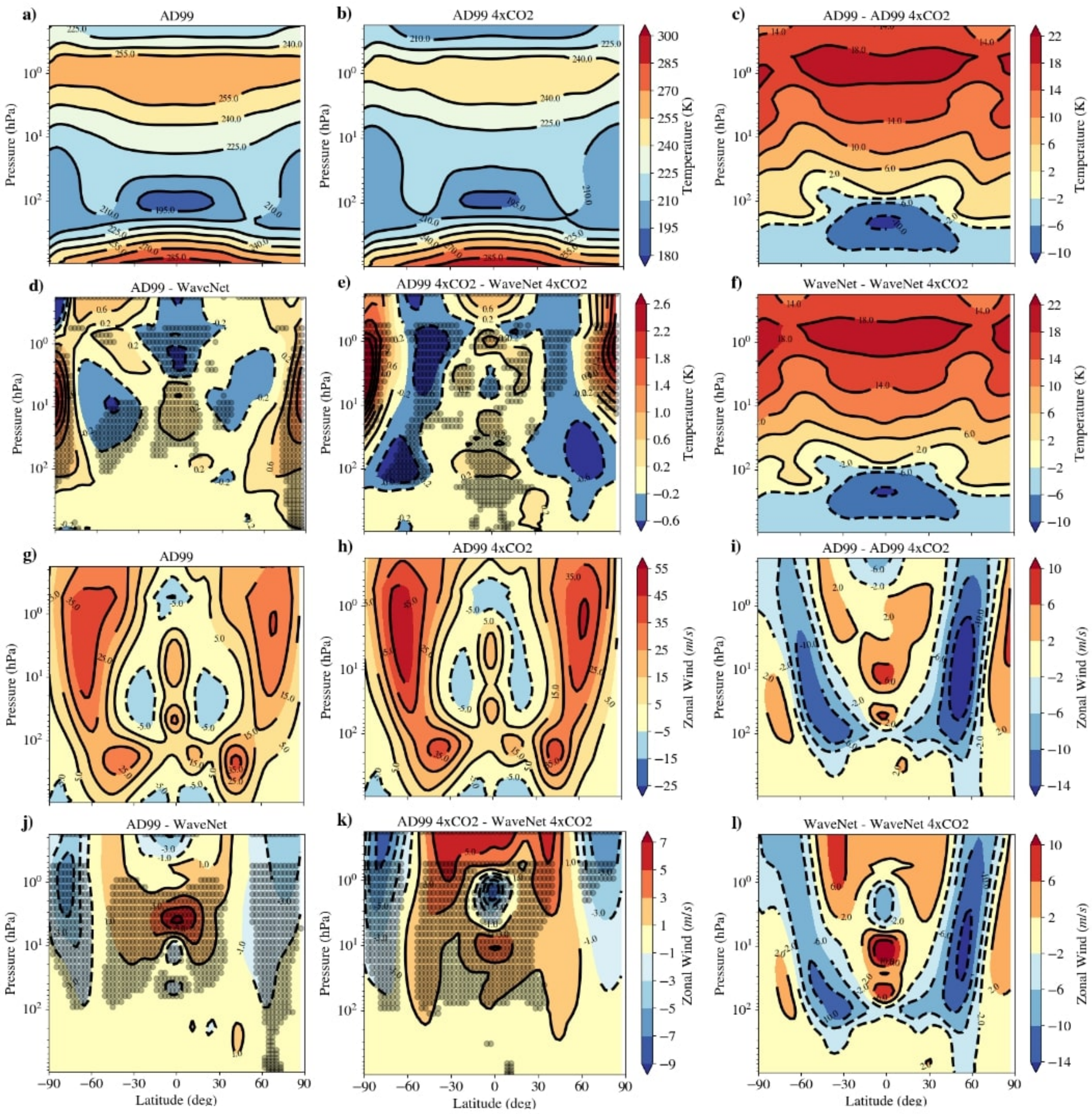


Figure 4. Pressure-Latitude profiles of average temperature (a,b) and zonal winds (g,h) in the baseline (a,g) and 4xCO2 (b,h) simulations for AD99. Panels d, e, j, and k present the climatological difference between AD99 and WaveNet for average temperature and zonal winds for the baseline and 4xCO2 simulations. Regions that are statistically distinguishable ($p<0.05$) via the Student's t-test are dotted. The climate change signal for average temperature and zonal winds for WaveNet and AD99 are presented in the rightmost column.

complexity, optimize coupling, and utilize GPU hardware. A next test is to examine how WaveNet generalizes when run using various model configurations and trained on regional datasets and orographic and nonorographic GWPs.

Nevertheless, our results suggest that machine learning may represent a powerful alternative to existing GWPs. The approach presented here constitutes a first step toward obtaining such GWPs for global climate prediction.

Acknowledgments

We thank Joan Alexander and Pedram Hassanzadeh for useful discussions. ZIE and AS acknowledge support from the NSF through grant OAC-2004492. EPG acknowledges support from the NSF through grant OAC-2004572. KD was funded by the NASA Postdoctoral Program at the Goddard Institute for Space Studies. This research received support by the generosity of Eric and Wendy Schmidt by recommendation of the Schmidt Futures program. We thank two anonymous reviewers for their constructive feedback, which greatly improved the manuscript.

Data Availability

The ANN is built using Keras, a deep learning framework that wraps Tensorflow. All training source code is available at `https://doi.org/10.5281/zenodo.4428931` (Espinosa, 2021). Training took on the order of 12 hours on a graphical processing unit and varied according to data size and trainable parameters. The implementation of SHAP values is available at `https://github.com/slundberg/shap` and is not maintained or owned by this project group (Lundberg & Lee, 2017a). MiMA is documented by Jucker and Gerber (2017) and Garfinkel et al. (2020), maintained at `https://github.com/mjucker/MiMA` and available at `https://zenodo.org/record/3984605#.YVYXtzHMKHs`. The model code, forpy coupling code, trained ANNs, run parameters, and modified configuration for MiMA are available at `https://zenodo.org/record/5533166`. The coupling library, forpy, developed and maintained by Elias Rabel is well documented and available at `https://github.com/ylikx/forpy` (Rabel et al., 2018).

References

Alexander, Bacmeister, J., Ern, M., Gisinger, S., Hoffmann, L., Holt, L., ... Wright, C. (2019). *Seeking new quantitative constraints on orographic gravity wave stress and drag to satisfy emerging needs in seasonal-to-subseasonal and climate prediction.* Retrieved from `http://www.issibern.ch/teams/consonorogravity/#`

Alexander, M., & Dunkerton, T. (1999). A spectral parameterization of mean-flow forcing due to breaking gravity waves. *Journal of the Atmospheric Sciences*, *56*(24), 4167–4182.

Antonita, T. M., Ramkumar, G., Kumar, K. K., Appu, K., & Nambhoodiri, K. (2007). A quantitative study on the role of gravity waves in driving the tropical stratospheric semiannual oscillation. *Journal of Geophysical Research: Atmospheres*, *112*(D12).

Brenowitz, N. D., & Bretherton, C. S. (2018). Prognostic validation of a neural network unified physics parameterization. *Geophysical Research Letters*, *45*(12), 6289–6298.

Bretherton, F. P. (1969). Momentum transport by gravity waves. *Quarterly Journal of the Royal Meteorological Society*, *95*(404), 213–243.

Butchart, N., Anstey, J., Hamilton, K., Osprey, S., McLandress, C., Bushell, A., . . . others (2018). Overview of experiment design and comparison of models participating in phase 1 of the sparc quasi-biennial oscillation initiative (qboi). *Geoscientific Model Development*, *11*(3).

Chantry, M., Hatfield, S., Dueben, P., Polichtchouk, I., & Palmer, T. (2020). Machine learning emulation of gravity wave drag in numerical weather forecasting. *Journal of Advances in Modeling Earth Systems*, e2021MS002477.

Cohen, N. Y., Gerber, & Oliver Bühler, E. P. (2013). Compensation between resolved and unresolved wave driving in the stratosphere: Implications for downward control. *Journal of the atmospheric sciences*, *70*(12), 3780–3798.

DallaSanta, K., Gerber, E. P., & Toohey, M. (2019). The circulation response to volcanic eruptions: The key roles of stratospheric warming and eddy interactions. *Journal of Climate*, *32*(4), 1101–1120.

DallaSanta, K., Orbe, C., Rind, D., Nazarenko, L., & Jonas, J. (2021). Dynamical and trace gas responses of the quasi-biennial oscillation to increased co2. *Journal of Geophysical Research: Atmospheres*, *126*(6), e2020JD034151.

Donner, L. J., Wyman, B. L., Hemler, R. S., Horowitz, L. W., Ming, Y., Zhao, M., . . . others (2011). The dynamical core, physical parameterizations, and basic simulation characteristics of the atmospheric component am3 of the gfdl global coupled model cm3. *Journal of Climate*, *24*(13), 3484–3519.

Espinosa, Z. I. (2021). *Learning-gwd-with-mima* [Computer Software]. Zenodo. Retrieved from `https://doi.org/10.5281/zenodo.4428931`

Fritts, D. C., & Alexander, M. J. (2003). Gravity wave dynamics and effects in the middle atmosphere. *Reviews of geophysics*, *41*(1).

Fritts, D. C., & Nastrom, G. D. (1992). Sources of mesoscale variability of gravity waves. part ii: Frontal, convective, and jet stream excitation. *Journal of the Atmospheric Sciences*, *49*(2), 111–127.

Garcia, R. R., Smith, A. K., Kinnison, D. E., Cámara, Á. d. l., & Murphy, D. J. (2017). Modification of the gravity wave parameterization in the whole atmosphere community climate model: Motivation and results. *Journal of the Atmospheric Sciences*, *74*(1), 275–291.

Garfinkel, C. I., White, I., Gerber, E. P., Jucker, M., & Erez, M. (2020). The building blocks of northern hemisphere wintertime stationary waves. *Journal of Climate*, *33*(13), 5611–5633.

Geller, M. A., Alexander, M. J., Love, P. T., Bacmeister, J., Ern, M., Hertzog, A., ... others (2013). A comparison between gravity wave momentum fluxes in observations and climate models. *Journal of Climate*, *26*(17), 6383–6405.

Gentine, P., Pritchard, M., Rasp, S., Reinaudi, G., & Yacalis, G. (2018). Could machine learning break the convection parameterization deadlock? *Geophysical Research Letters*, *45*(11), 5742–5751.

Glorot, X., & Bengio, Y. (2010). Understanding the difficulty of training deep feedforward neural networks. In *Proceedings of the thirteenth international conference on artificial intelligence and statistics* (pp. 249–256).

Jucker, M., & Gerber, E. (2017). Untangling the annual cycle of the tropical tropopause layer with an idealized moist model. *Journal of Climate*, *30*(18), 7339–7358.

Kang, M.-J., Chun, H.-Y., Kim, Y.-H., Preusse, P., & Ern, M. (2018). Momentum flux of convective gravity waves derived from an offline gravity wave parameterization. part ii: Impacts on the quasi-biennial oscillation. *Journal of the Atmospheric Sciences*, *75*(11), 3753–3775.

Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.

Limpasuvan, V., Richter, J. H., Orsolini, Y. J., Stordal, F., & Kvissel, O.-K. (2012). The roles of planetary and gravity waves during a major stratospheric sudden warming as characterized in waccm. *Journal of atmospheric and solar-terrestrial physics*, *78*, 84–98.

Lindgren, E. A., Sheshadri, A., Podglajen, A., & Carver, R. W. (2020). Seasonal and latitudinal variability of the gravity wave spectrum in the lower stratosphere. *Journal of Geophysical Research: Atmospheres*, *125*(18), e2020JD032850.

Lundberg, S. M., & Lee, S.-I. (2017a). A unified approach to interpreting model predictions. In I. Guyon et al. (Eds.), *Advances in neural information processing systems 30* (pp. 4765–4774). Curran Associates, Inc. Retrieved from `http://papers.nips.cc/paper/7062-a-unified-approach-to-interpreting-model-predictions.pdf`

Lundberg, S. M., & Lee, S.-I. (2017b). A unified approach to interpreting model predictions. In *Advances in neural information processing systems* (pp. 4765–4774).

Matsuoka, D., Watanabe, S., Sato, K., Kawazoe, S., Yu, W., & Easterbrook, S. (2020). Application of deep learning to estimate atmospheric gravity wave parameters in reanalysis data sets. *Geophysical Research Letters*, *47*(19), e2020GL089436.

Ott, J., Pritchard, M., Best, N., Linstead, E., Curcic, M., & Baldi, P. (2020). A fortran-keras deep learning bridge for scientific computing. *Scientific Programming*, *2020*.

Palmer, T., Shutts, G., & Swinbank, R. (1986). Alleviation of a systematic westerly bias in general circulation and numerical weather prediction models through an orographic gravity wave drag parametrization. *Quarterly Journal of the Royal Meteorological Society*, *112*(474), 1001–1039.

Podglajen, A., Hertzog, A., Plougonven, R., & Legras, B. (2016). Lagrangian temperature and vertical velocity fluctuations due to gravity waves in the lower stratosphere. *Geophysical Research Letters*, *43*(7), 3543–3553.

Polichtchouk, I., Shepherd, T. G., & Byrne, N. J. (2018). Impact of parametrized nonorographic gravity wave drag on stratosphere-troposphere coupling in the northern and southern hemispheres. *Geophysical Research Letters*, *45*(16), 8612–8618.

Rabel, E., Ruger, R., Govoni, M., & Ehlert, S. (2018). *Forpy: A library for fortran-python interoperability.* Retrieved from `https://github.com/ylikx/forpy`

Rasp, S., Pritchard, M. S., & Gentine, P. (2018). Deep learning to represent subgrid processes in climate models. *Proceedings of the National Academy of Sciences*, *115*(39), 9684–9689.

Remmler, S., Hickel, S., Fruman, M. D., & Achatz, U. (2015). Validation of large-eddy simulation methods for gravity wave breaking. *Journal of the Atmospheric Sciences*, *72*(9), 3537–3562.

Richter, J. H., Anstey, J. A., Butchart, N., Kawatani, Y., Meehl, G. A., Osprey, S., & Simpson, I. R. (2020). Progress in simulating the quasi-biennial oscillation in cmip models. *Journal of Geophysical Research: Atmospheres*, *125*(8), e2019JD032362.

Richter, J. H., Butchart, N., Kawatani, Y., Bushell, A. C., Holt, L., Serva, F., … others (2020). Response of the quasi-biennial oscillation to a warming climate in global climate models. *Quarterly Journal of the Royal Meteorological Society*.

Richter, J. H., Sassi, F., & Garcia, R. R. (2010). Toward a physically based gravity wave source parameterization in a general circulation model. *Journal of the Atmospheric Sciences*, *67*(1), 136–156.

Roh, S., & Song, H.-J. (2020). Evaluation of neural network emulations for radiation parameterization in cloud resolving model. *Geophysical Research Letters*, *47*(21), e2020GL089444.

Sato, K., & Hirano, S. (2019). The climatology of the brewer–dobson circulation and the contribution of gravity waves. *Atmos. Chem. Phys*, *19*, 4517–4539.

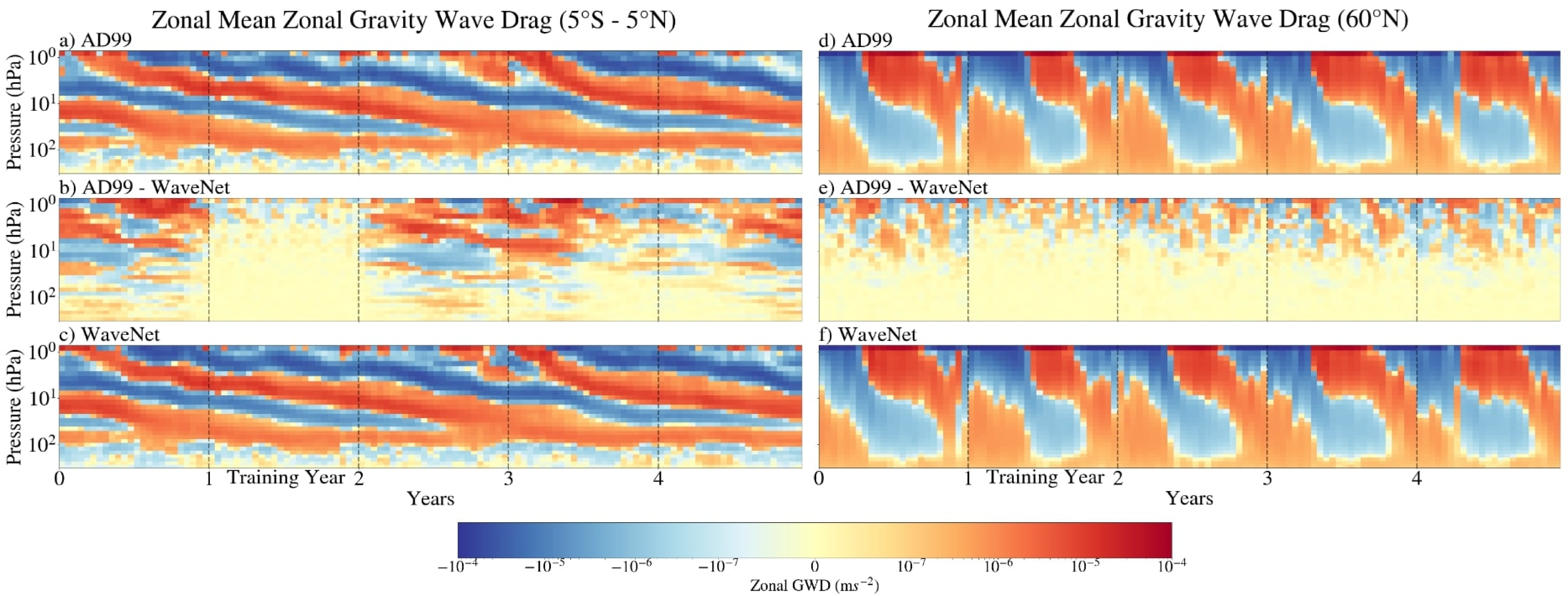
Schirber, S. (2015). Influence of enso on the qbo: Results from an ensemble of idealized simulations. *Journal of Geophysical Research: Atmospheres*, *120*(3), 1109–1122.

Shrikumar, A., Greenside, P., & Kundaje, A. (2017). Learning important features through propagating activation differences. *arXiv preprint arXiv:1704.02685*.

Sigmond, M., & Scinocca, J. F. (2010). The influence of the basic state on the northern hemisphere circulation response to climate change. *Journal of climate*, *23*(6), 1434–1446.

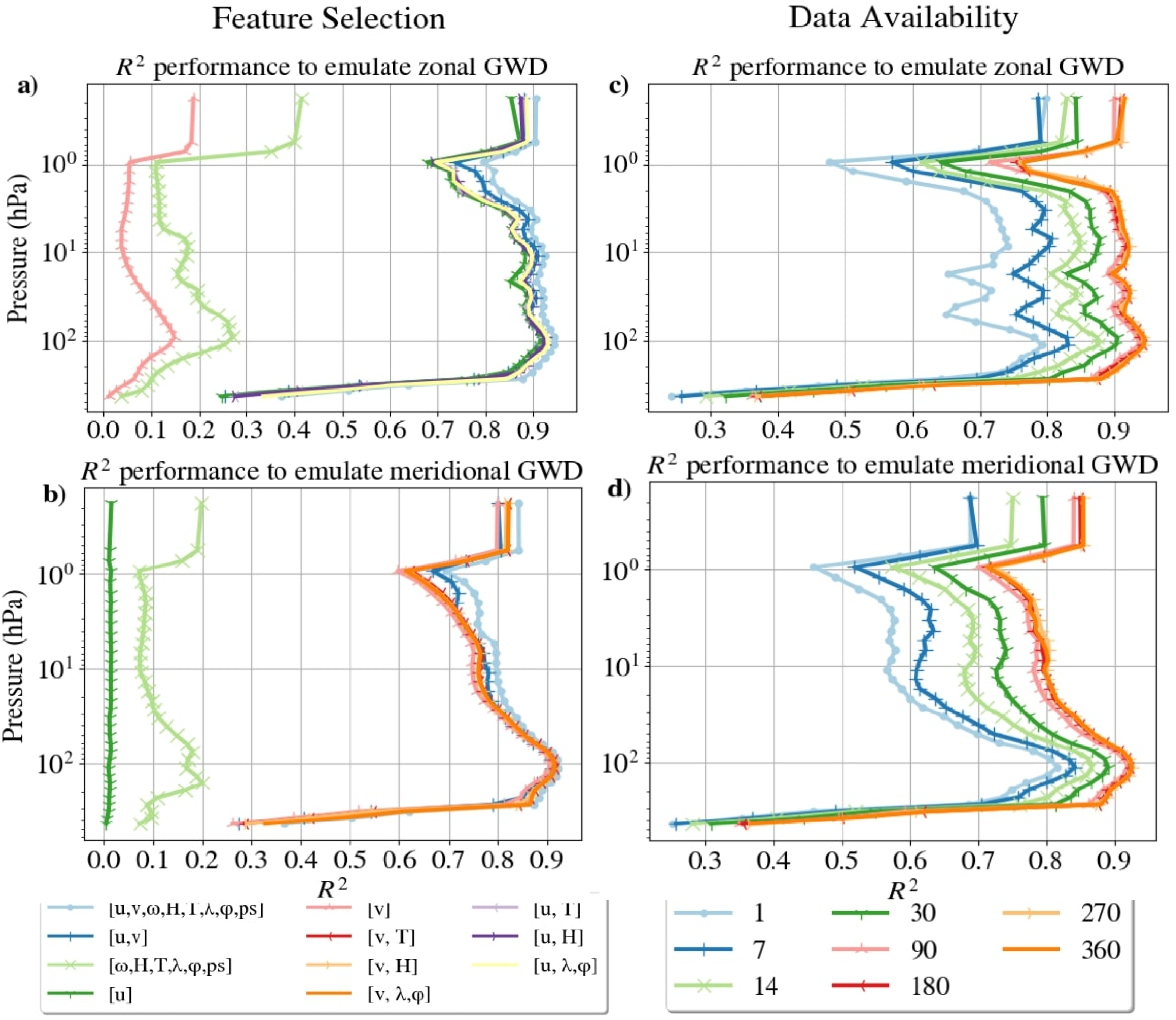
Xu, X., Wang, Y., Xue, M., & Zhu, K. (2017). Impacts of horizontal propagation of orographic gravity waves on the wave drag in the stratosphere and lower mesosphere. *Journal of Geophysical Research: Atmospheres*, *122*(21), 11–301.

Figure 1.



Zonal Mean Zonal Gravity Wave Drag (5°S - 5°N)
Zonal Mean Zonal Gravity Wave Drag (60°N)
a) AD99
b) AD99 - WaveNet
c) WaveNet
d) AD99
e) AD99 - WaveNet
f) WaveNet
Pressure (hPa)
Training Year
Years
Zonal GWD (ms^{-2})

Figure 2.

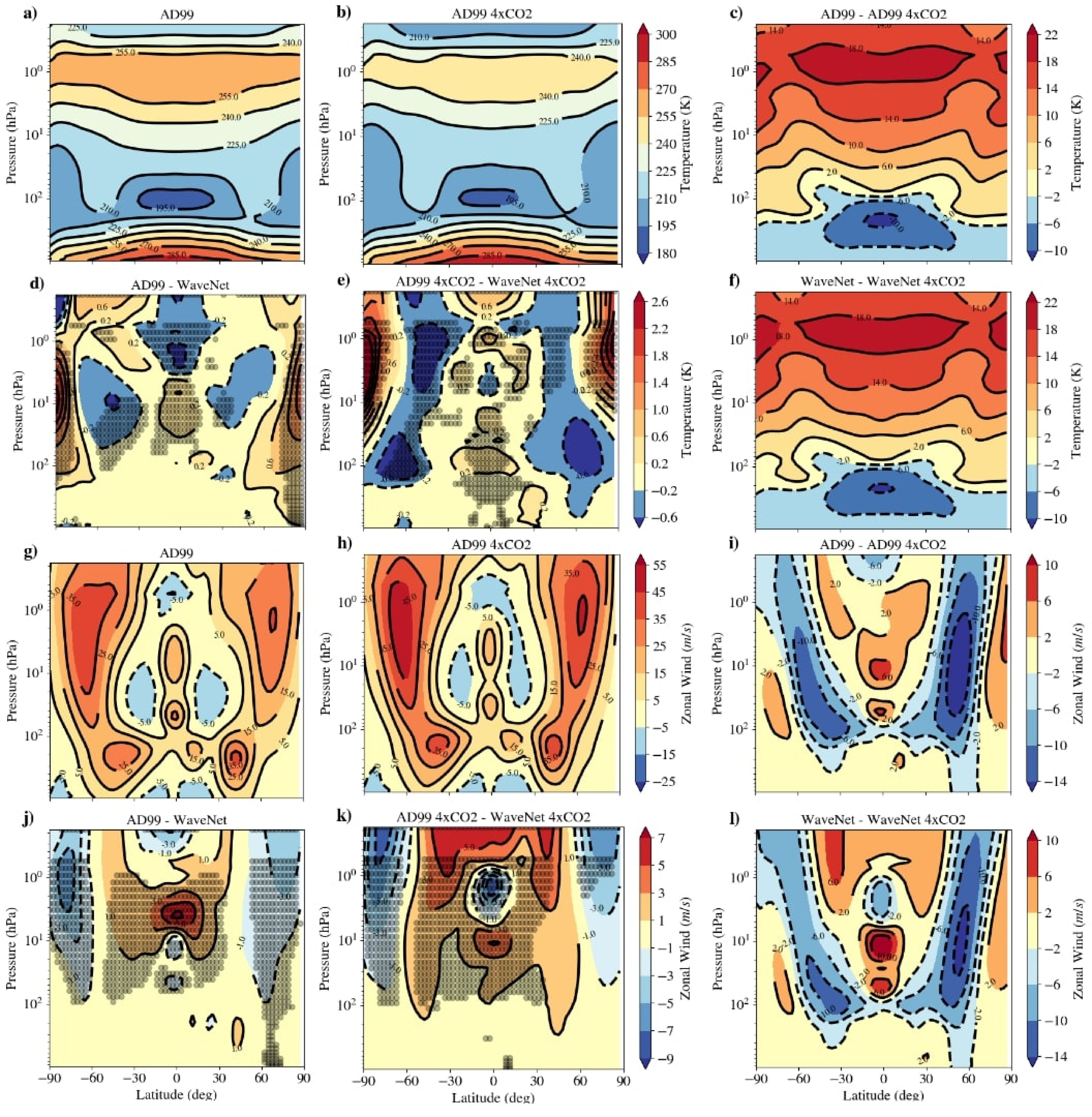


Feature Selection
Data Availability
R^2 performance to emulate zonal GWD
R^2 performance to emulate meridional GWD
a)
b)
c)
d)
Pressure (hPa)
R^2
[u,v,ω,H,T,λ,φ,ps]
[u,v]
[ω,H,T,λ,φ,ps]
[u]
[v]
[v, T]
[v, H]
[v, λ,φ]
[u, T]
[u, H]
[u, λ,φ]
1
7
14
30
90
180
270
360

Figure 3.

a)
AD99: Zonal Mean Zonal Winds (5°S-5°N)
Period: 28.18 ± 2.52 (months)
Westerly Amplitude: 48.35 ± 2.10 (m/s)
Easterly Amplitude: -28.47 ± 3.45 (m/s)
b)
WaveNet: Zonal Mean Zonal Winds (5°S-5°N)
Period: 30.27 ± 2.82 (months)
Westerly Amplitude: 57.64 ± 1.50 (m/s)
Easterly Amplitude: -31.26 ± 5.90 (m/s)
c)
AD99 4xCO2: Zonal Mean Zonal Winds (5°S-5°N)
Period: 21.79 ± 2.78 (months)
Westerly Amplitude: 37.34 ± 4.91 (m/s)
Easterly Amplitude: -25.63 ± 2.69 (m/s)
d)
WaveNet 4xCO2: Zonal Mean Zonal Winds (5°S-5°N)
Period: 19.35 ± 6.16 (months)
Westerly Amplitude: 32.54 ± 13.79 (m/s)
Easterly Amplitude: -24.78 ± 6.12 (m/s)
Pressure (hPa)
Years
U (m/s)

Figure 4.


a) AD99
b) AD99 4xCO2
c) AD99 - AD99 4xCO2
d) AD99 - WaveNet
e) AD99 4xCO2 - WaveNet 4xCO2
f) WaveNet - WaveNet 4xCO2
g) AD99
h) AD99 4xCO2
i) AD99 - AD99 4xCO2
j) AD99 - WaveNet
k) AD99 4xCO2 - WaveNet 4xCO2
l) WaveNet - WaveNet 4xCO2
Pressure (hPa)
Temperature (K)
Zonal Wind (m/s)
Latitude (deg)
-90 -60 -30 0 30 60 90