Recreating observed convection-generated gravity waves from weather radar observations via a neural network and a dynamical atmospheric model

C. G. Kruse[1*], M. J. Alexander[1], M. Bramberger[1], A. Chattopadhyay[2], P. Hassanzadeh[2], B. Green[3], A. Grimsdell[1], L. Hoffmann[4]

[1]NorthWest Research Associates, Boulder, Colorado, USA
[2]Rice University, Houston, Texas, USA
[3]Stanford University, Palo Alto, California, USA
[4]Forschungszentrum Jülich, Jülich, Germany

Key Points:

- If realistic convective diabatic heating is supplied at the correct places and times, GW-resolving models can reasonably reproduce CGWs.
- While location and type of convective storm do influence latent heating, these difference are of 2nd order importance for CGW forcing.
- Drag due to convection-generated GWs from a compact source region is spread over O(1000) km due to lateral propagation

*3380 Mitchell Lane, Boulder, Colorado

Corresponding author: Christopher G. Kruse, `ckruse@nwra.com`

Abstract

Convection-generated gravity waves (CGWs) transport momentum and energy, and this momentum is a dominant driver of global features of Earth's atmosphere's general circulation (e.g. the quasi-biennial oscillation, the pole-to-pole mesospheric circulation). As CGWs are not generally resolved by global weather and climate models, their effects on the circulation need to be parameterized. However, quality observations of GWs are spatiotemporally sparse, limiting understanding and preventing constraints on parameterizations. Convection-permitting or -resolving simulations do generate CGWs, but validation is not possible as these simulations cannot reproduce the CGW-forcing convection at correct times, locations, and intensities.

Here, realistic convective diabatic heating, learned from full-physics convection-permitting Weather Research and Forecasting (WRF) simulations, is predicted from weather radar observations using neural networks and a previously developed look-up table. These heating rates are then used to force an idealized GW-resolving dynamical model. Simulated CGWs forced in this way did closely resemble those observed by the Atmospheric InfraRed Sounder in the upper stratosphere. CGW drag in these validated simulations extends 100s of kilometers away from the convective sources, highlighting errors in current gravity wave drag parameterizations due to the use of the ubiquitous single-column approximation. Such validatable simulations have significant potential to be used to further basic understanding of CGWs, improve their parameterizations physically, and provide more restrictive constraints on tuning *with confidence*.

Plain Language Summary

Thunderstorms generate waves in the atmosphere that can generate turbulence at commercial aircraft cruising altitudes and further aloft. At these higher altitudes, they eventually break, not only generating turbulence, but also exerting forces that affect the large-scale flows in the middle atmosphere. While these waves have been known to be important since at least the 1980s, they are difficult to observe. They can be simulated, but weather models do not simulate thunderstorms in the correct locations at the right times, meaning the simulated waves cannot be directly compared against observations. Here, weather radar observations are used as input to a look-up table and a neural network to force realistic thunderstorm motions and waves within a simplified weather model. This method was able to reproduce a satellite-observed case with notable skill. In one of the first simulations of thunderstorm-generated waves comparable to satellite observations, these waves travel 100s of kilometers away from the thunderstorms, conflicting with assumptions made in weather and climate models.

1 Introduction

Atmospheric gravity waves (GWs), or buoyancy waves, are mesoscale phenomena ($\approx 10 - 1000$ km wavelength), that transport momentum from lower to upper atmosphere layers and drive features in large-scale atmospheric circulation (Alexander et al., 2010). Convection is a primary source of atmospheric GWs, particularly in the tropics (C. C. Stephan et al., 2019a; Corcos et al., 2021; Liu et al., 2022) and summer extratropics (Hoffmann et al., 2013; Plougonven et al., 2015; C. Stephan et al., 2016; C. C. Stephan et al., 2019b), but also in winter hemisphere subtropical regions (Holt et al., 2017). In particular, convection-generated GWs (CGWs) are primary drivers of the stratospheric quasi-biennial oscillation (QBO) (Baldwin et al., 2001; Holt et al., 2016; Bushell et al., 2022), which influences tropospheric predictability in the tropics (Yoo & Son, 2016; Marshall et al., 2017; Abhik & Hendon, 2019; Martin et al., 2021; Anstey et al., 2021) and extra-tropics (Gray et al., 2018; Garfinkel et al., 2018). CGWs also play a role in the equator-to-pole Brewer-Dobson Circulation (Alexander & Rosenlof, 2003; C. Stephan et al., 2016), which is a primary driver of ozone and water vapor concentrations in the stratosphere (Hegglin & Shepherd, 2009).

Despite the importance of CGWs in climate and seasonal prediction, they remain largely unresolved in global prediction models, and their forcings on large-scale circulations must be parameterized (Richter et al., 2020; Bushell et al., 2022). The sparsity of quality observations of CGWs has prevented development of quantitative constraints on parameterizations (Alexander et al., 2021; Lee et al., 2022). As a result, these parameterizations are highly simplified using numerous idealizations and typically tuned to minimize a handful of global error metrics depending on the application (Richter et al., 2022). Instead of using observations to further fundamental understanding of CGWs and improve parameterizations, convection-permitting and -resolving simulations do internally generate CGWs and could be used. However, such simulations cannot reproduce the timings, locations, and intensities of actual convective sources, preventing validation of such simulations against the few CGW observations that exist. Without validation of such simulations, it is difficult to make progress in CGW research *with confidence*.

Here, a recently developed method is used to force an idealized GW-resolving model with reasonably-realistic diabatic heating at the correct locations and times in order to have a chance at simulating CGWs in a way that can be directly compared with observations following Grimsdell et al. (2010); C. Stephan and Alexander (2015); C. C. Stephan et al. (2016); Bramberger et al. (2020). This diabatic heating is predicted from weather radar observations of actual cases. Two methods are used to predict diabatic heating: the previously-developed look-up table (LT) method of Bramberger et al. (2020) and a new simple neural network (NN) model. This radar-derived heating is then provided to a GW-resolving idealized configuration of the Weather Research and Forecasting (WRF) model, which responds dynamically to the diabatic forcing in all ways the non-linear dynamical core and resolution allow. This method is tested against Atmospheric InfraRed Sounder (AIRS) and Project Loon super-pressure balloon observations in two cases. These two cases highlight the methods' abilities to reproduce observed CGWs. Previous work suggested the gravity wave spectrum above convection in WRF simulations was only modestly sensitive to the choice of microphysics parameterization (C. Stephan & Alexander, 2015), while the depth and strength of the convective latent heating are key determinants of the gravity wave spectrum (Bramberger et al., 2020). Our study also addresses how sensitive the CGW are to LT and NN methods based on specific locations/conditions.

The overall method to simulate actual cases of CGWs, the two tools used to predict convective diabatic heating, and the training data sets used for both tools are described in Section 2. The skill of the look-up table and NN models in predicting WRF-simulated diabatic heating profiles is presented in Section 3. Idealized model runs forced with the different diabatic heatings are then performed and compared to two cases of observed CGW: One observed by AIRS and one with Loon super-pressure balloon data in Section 4. Finally, Section 5 is a discussion of the results and conclusions. Details on the accessibility of data, NNs, WRF source codes, and analysis codes are given in Section 6.

2 Methods and Models

2.1 Overall Summary of the Method

CGWs are simulated within an idealized WRF configuration solely forced by convective diabatic heating. This diabatic heating, Q, is derived from the Multi-Radar, Multi-Sensor (MRMS) dataset, which merges numerous radar-derived quantities from all weather radars in the contiguous United States onto a single 0.01^o latitude, longitude ($\approx$ 1-km resolution) grid every two minutes (Zhang et al., 2016). Similar methods have been previously used to force CGWs from other weather radar data sets over the mid-latitude, Midwestern US (C. Stephan & Alexander, 2015; C. C. Stephan et al., 2016) and near Darwin, Australia (Grimsdell et al., 2010; Bramberger et al., 2020).

2.2 Predicting Convective Diabatic Heating from Weather Radar

2.2.1 Training Data

Two methods are used to predict profiles of Q given radar-observed quantities: the look-up table method of Bramberger et al. (2020) and a neural network (NN) method developed here. While radar reflectivities provide observations of falling convective precipitation, there are no observations of Q for training the methods. To work around this issue, the two methods are trained on full-physics, realistic convection-permitting ($\Delta x =$ 2-km, $\Delta z <$ 500-m resolution) WRF simulations of observed convective events. Within these simulations, the two methods are trained to predict simulated diabatic heating given simulated radar-observable quantities.

Two sets of full-physics, realistic WRF simulations were used for training: simulations of a case of significant deep tropical convection used by Bramberger et al. (2020) over Darwin, Australia (hereafter the Darwin run) and a simulation of typical diurnal convection over Florida (hereafter the Florida run).

The Darwin run simulated a 48-hour period, beginning 11 Jan 2003 at 12 UTC. The inner-most domain used a $\Delta x =$ 2-km resolution, was 408 km by 408 km wide, and was run three times with three slightly different model tops, effectively producing three ensemble members of the same case. A 10-km-deep upper sponge layer was used to prevent GW reflection off the top of the domain. The tropical physics suite was used (`https://www2.mmm.ucar.edu/wrf/users/physics/ncar_tropical_suite.php`). Initial and boundary conditions were provided by the ERA-Interim reanalysis. All three “ensemble members” of this case were included in the training and are together referred to as the Darwin run. The outer 20 km of the 2-km resolution domain were excluded from training, as were the first 12 hours of the simulations while initial imbalances dissipate and convection becomes well-developed. For complete details, see Bramberger et al. (2020).

The Florida run was completed as part of this work using WRFv4.4. A single $\Delta x =$ 2-km resolution domain was set up, with initial and boundary conditions from the ERA5 reanalysis (Hersbach et al., 2020). The domain was 1200 km by 1200 km wide, had a top at 1 hPa ($z \approx$ 45 km), and 110 vertical levels resulting in a nearly constant resolution of $\Delta z \approx$ 500 m above the tropopause. A 10-km-deep upper sponge layer was again specified. The tropical physics suite was again used. The period simulated was 72 hours, beginning 14 June 2018 at 12 UTC. Given large difference in resolution between the forcing reanalysis used for boundary conditions ($\Delta x \approx$ 31 km) and WRF ($\Delta x =$ 2 km), the outermost 200-km of the domain were excluded from training. The first 12 hours of the simulation were also excluded.

To train the two methods described below, simulated radar quantities (i.e. inputs) and diabatic heating profiles (outputs) were paired at each grid point and time, but only for *convective* grid points. Grid points were deemed convective if the simulated rain rate exceeded 1 mm (10 min)$^{-1}$. In the Darwin and Florida runs, 1558031 and 180247 convective grid points were extracted, respectively.

2.2.2 Look-Up Table

The look-up table (LT) used here was the same as used by Bramberger et al. (2020). Briefly, to create their LT, convective grid points were binned by rain rate (RR) and echo top height (ET). Simulated diabatic heating profiles were then averaged within the simulated RR and ET bins. Then, given a RR and ET, a diabatic heating profile, $Q(z)$, is predicted via 2-D linear interpolation. The LT used here was trained only on the Darwin run, referred to as "DALT" in the figures.

2.2.3 Neural Networks

The LT method likely introduces errors due to the averaging applied within RR and ET bins, the dimensions of which are imposed. Additionally, it is not straightforward to expand the look-up table to take advantage of additional radar-observable quantities. Neural network architectures, and machine-learning methods in general, can provide a few advantages over a LT method. For example, NN training provides a flexible framework to increase the number of input quantities and more fully make use of available data. Additionally, averaging or compositing of heating profiles over RR and ET is not imposed, which may allow NNs to be more sensitive to input variables and distinguish between different diabatic heating regimes. Finally, the inherently non-linear nature of using an NN for prediction has potential in to increase skill by being better able to represent the complex structures of heating profiles. Here, five radar-observable quantities were used to predict diabatic heating profiles at a given point: radar reflectivities at 0 C, -10 C, and -20 C isotherms in addition to RR and ET used by the LT method. Prior to use with the NNs, all input variables and diabatic heatings were demeaned and then normalized by their standard deviations.

Here, a 40-neuron-wide, 6-layer-deep fully-connected NN with a hyperbolic tangent activation function was used to predict diabatic heating profiles gridpoint by gridpoint. Given the two sets of simulations to train on, three NNs were trained to predict diabatic heating: one trained on the Darwin run only, one trained on the Florida run only, and one trained on both, represented by "DANN", "FLNN", and "DAFLNN", respectively. The DANN was trained on all Darwin run convective grid points. The FLNN was trained on 90% of the Florida run convective grid points. The DAFLNN was trained on convective grid points from both simulations. Given the much smaller number of convective grid points in the Florida run, the Florida run profiles were duplicated until the number of Florida profiles was equal to the number of Darwin profiles to avoid data imbalance. A mean-squared error (MSE) loss function and a learning rate of 0.005 were used for training. Weights were updated after every batch of 10000 input-output pairs. Training continued until the epoch-accumulated MSE reduced by less 0.01%. These three NNs trained on the two training sets allow some inference of how generally applicable a NN trained on a single case of deep, tropical convection (e.g. the Darwin run) might be when used, for example, on a case of subtropical convection over the southeast US.

Limited hyperparameter optimization was performed in this problem. An NN with double the neurons (80 neurons, 6 layers) and an NN with an extra two layers (40 neurons, 8 layers) were trained on Darwin run profiles to predict a subset of convective profiles, also from the Darwin run. Changes in validation profiles (similar to Fig. 2, not shown) were minute, so the 40 neuron wide, 6 layer deep NN architecture was chosen. Further hyperparameter optimization is left to future work.

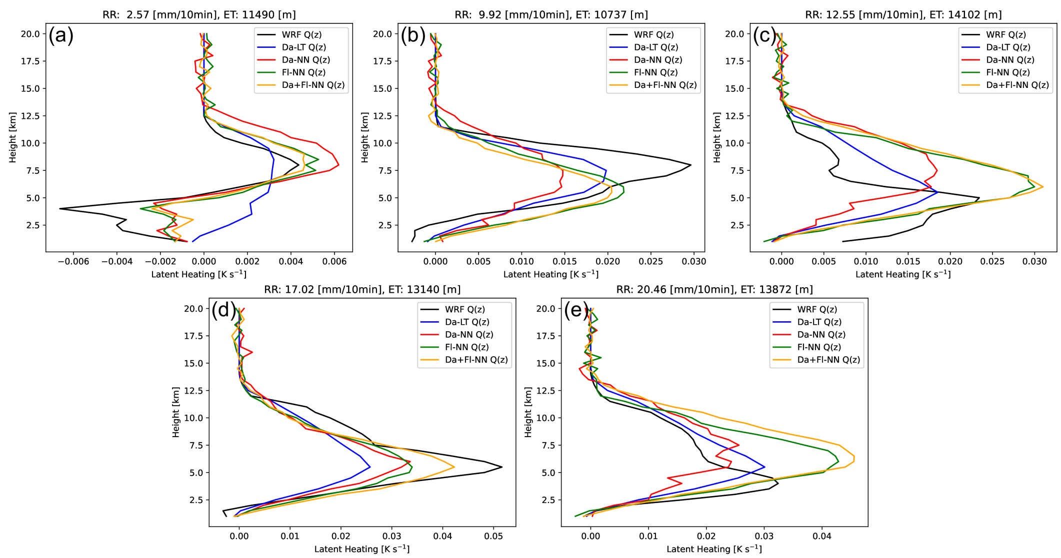

Figure 1. Individual profiles of WRF-simulated (black) and predicted (colors) latent heating. Profiles were randomly chosen from within the five, 5-mm rain rate bins from the Florida run. The tropopause was near $z = 15$ km for this case.

3 Evaluations of Diabatic Heating Predictions

The four methods of predicting diabatic heating are tested against the 10% of the Florida run profiles withheld from training. These withheld profiles were compiled by first binning all of the Florida-run convective grid points into RR bins of 5 mm (10 min)$^{-1}$ and then withholding a randomly chosen 10% of the profiles in each RR bin for testing. This process ensures the RR probability density function of the testing data is the same as in the training data and also ensures that the rarest, but most important profiles with the highest rain rates do not all end up being withheld from training. Rain rate is a good proxy for the magnitude of the diabatic heating above, which forces CGWs. Note that the two NNs that include profiles from the Florida run in training are being evaluated against Florida run profiles withheld from the same simulation.

WRF-simulated diabatic heating profiles and predictions from the four methods are shown for five randomly chosen profiles within the five RR bins in Fig. 1. By eye, the NNs predict WRF-simulated Q similarly. The DALT predictions are somewhat distinct, being more smooth in the vertical, which might be expected given the averaging inherent in the LT method. Encouragingly, all of the NNs represented the negative heatings near the surface due to evaporative cooling in the smallest RR profiles (Fig. 1a), whereas the DALT did not. While the DALT did not represent this feature here, look-up tables can be constructed to represent it (Lang & Tao, 2018; Tao et al., 2019).

Profiles of mean absolute error (MAE) and bias (MAE $= N^{-1} \sum_{i=1}^{N} |E_i|$, bias $= N^{-1} \sum_{i=1}^{N} E_i$, where $E_i = Q_{i,pred} - Q_{i,WRF}$ and Q is diabatic heating) validation statistics are presented in Fig. 2. Here, the bin-mean WRF-simulated diabatic heating profile is shown in black for reference, averaged over the number of profiles given in each panel title. In the smallest RR bin, all methods perform the worst, with MAE significantly larger than the bin-mean diabatic heating. This lack of predictive skill may be due insufficient information within the input quantities. Also, at these low RRs, not all of the profiles might be convective in nature, leading to errors when trying to predict a non-convective diabatic heating profile. At larger rain-rates (Fig. 2b-e), diabatic heat-

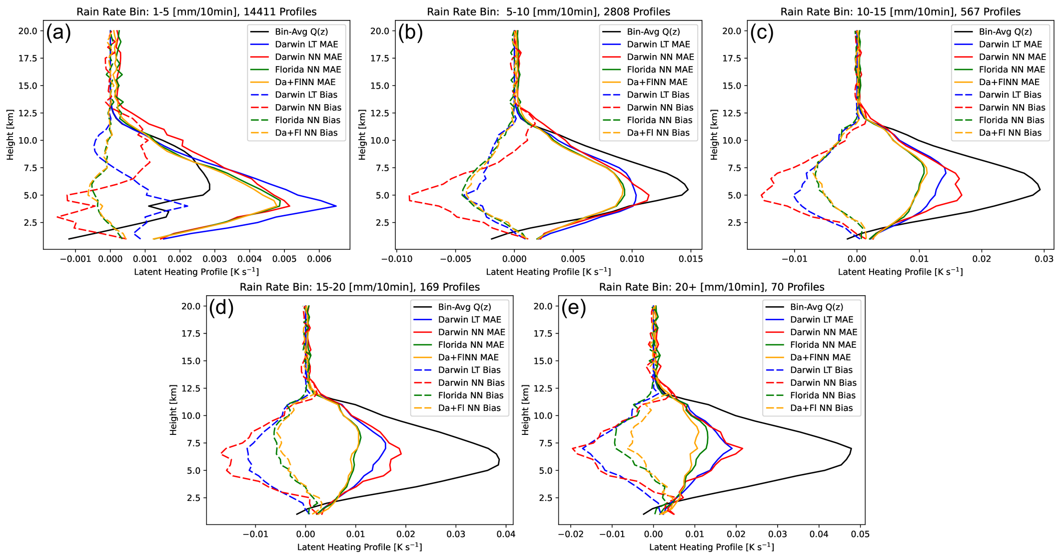

Figure 2. Validation statistics plotted as a function of height for the three NNs and the LT. All methods are tested against Florida run convective profiles from WRF (e.g. Fig. 1) that were withheld from training. Mean-absolute errors (MAE) are plotted as solid, colored lines. Mean errors (i.e. biases) are dashed. The mean latent heating profiles within the 5 mm (10 min)$^{-1}$ bins are plotted in solid black.

ings are much larger and all methods perform much better, with MAEs smaller than the mean heating rates.

Fig. 2 allows the predictive skill of the Darwin-trained LT and the Darwin-trained NN to be compared. Across all larger RR bins with more of a signal to predict, the two methods have very similar performance. Perhaps the DALT has slightly better skill than the DANN, with incrementally higher MAE by the DANN near the diabatic heating maxima apparently due to a weak bias in heating. However, the NNs perform notably better than the DALT for the smallest RRs, with smaller MAEs and biases in the lower half of the troposphere. Perhaps this is a reflection of the NNs' abilities to better represent more complex profiles of heatings, due to less averaging or compositing of the majority of profiles at these smaller RRs used in training, or a result of more information about the profile being used as input (i.e. reflectivities at 0C, -10C, and -20C used by the NNs and not the DALT).

Comparison of the validation profiles for the DANN, FLNN, and DAFLNN allow some inferences to be made on how generally applicable a NN trained on a single case of deep, tropical convection might be. In all RR bins except the lowest, the FLNN outperforms the DANN, with MAE reduced by about 33% relative to DANN. This might not be too surprising as the FLNN was trained on the same run from which these testing data were withheld. For all but the highest RR bin, including the Darwin-run profiles in the NN training did not change the predictive skill much. However, at the highest RRs, inclusion of the Darwin profiles in training did notably increase the predictive skill of the NN. This is likely due to the fact that the Darwin run included much stronger (RRs 65+ mm (10 min)$^{-1}$) and deeper (tropopause at $z = 18$ km near Darwin vs $z =$ 15 km over Florida) convection, having more convective grid points at these higher rain rates from which to learn.

To summarize, convective diabatic heating exhibits significant point-to-point variability and is a challenge to predict skillfully given only a handful of radar-observable quantities. Both the LT and NN methods have similar predictive skill at larger RRs. The NNs appear to be better able to predict the complex heating profiles at the smallest RRs. More representative training data (e.g. from the Florida run) increases predictive skill. Finally, as largely expected, more training data (i.e. including both runs in the training) can further increase skill incrementally.

4 Evaluations of Simulated CGWs

4.1 Idealized WRF Configuration

The four tools described above were used to predict convective diabatic heating from MRMS data. Then, these heatings were supplied to the same idealized configuration of WRF used by C. Stephan and Alexander (2015) and Bramberger et al. (2020). Briefly, the 3-D super cell idealized case within WRFv3.7 was the starting point. The initialization code was modified to remove the default initial warm bubble. All physical parameterizations were disabled. WRF's "open" boundary conditions were used, designed to allow small-amplitude GWs to propagate out of the domain without affecting the interior solution. The Coriolis parameters were constant across the domain and set using a latitude of 28.5 degrees north. The namelist parameter "pert_coriolis" was set to true to only allow the Coriolis forces to be applied to the wind speed deviations from the initial profiles. Initial profiles were taken from MERRA2 (Gelaro et al., 2017), averaged between 25 and 34 degrees latitude, -77 and -84 degrees longitude at times closest to the measurements of interest (see cases below). A key modification was made to the WRF variable registry, which allowed WRF to read the internal diabatic heating variable, "h_diabatic", from a file via an auxiliary input stream. The modified WRF source code, along with a diff relative to the original source code, are provided. See the Open Research section below for details.

The four tools were used to create 3-D diabatic heating files readable by WRF on the 2-km resolution WRF grid every two minutes. Heatings were only provided within the dashed box in Fig. 3e, tapered from zero to the full amounts between the dashed and solid boxes. Additionally, the small heatings produced by the NNs above the echo top heights (e.g. Fig. 1) were set to zero. These heating files were read by WRF, updating the constant diabatic heating used to force changes in temperature every two minutes. As WRF integrates forward in time (a $\Delta t = 10$ s was used), WRF's dynamical core responds to this heating in every way the governing equations and resolution allow. Convective updrafts and compensating subsidence are forced. All mechanisms that generate CGWs (i.e. diabatic heating, obstacle, mechanical oscillator) act to the extent possible, as forced by the provided diabatic heating.

4.2 Evaluation against AIRS

4.2.1 The Case of Interest

In order to evaluate the idealized WRF simulations, an attempt was made to reproduce CGWs observed by the Atmospheric InfraRed Sounder (AIRS) in the stratosphere. Brightness temperature perturbations from AIRS radiance measurements averaged over 42 channels with wavelengths near 4 μm and 2 channels with wavelengths near 15 μm are shown in panel (a) of Figs. 3 and 4. For details on the brightness temperature products, see Hoffmann et al. (2013, 2014, 2017). Vertical observational filter kernels, averaged over all channels included in each product, are shown in Fig. 5, which depict the relative importance of different altitudes in emitting radiation at the selected wavelengths to the AIRS sensor. The 4 μm channel set is most sensitive to stratospheric temperature perturbations at about 30 − 40 km of altitude. The 15 μm channel set is most

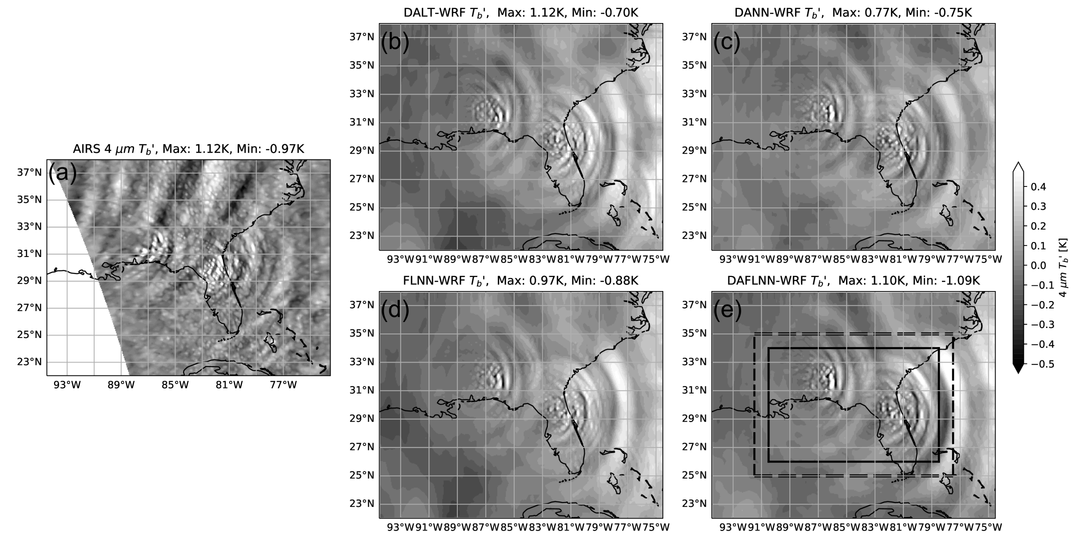

Figure 3. Maps of observed (a) and WRF-simulated (b-e) T_b'. AIRS observations shaded in (a) were collected over 18:41 to 18:45 UTC on 22 July 2018. The WRF-simulated T_b' were computed using output at 18:50 UTC. Approximate vertical and horizontal AIRS observational filters were applied to WRF in (b-e). Diabatic heating, Q, supplied to WRF was limited to within the boxes in (e), with a cosine ramp transitioning predicted Q from zero to its full amount between the dashed and solid lines.

sensitive at about 40 − 45 km. Note the different vertical width and sensitivity of the two kernel functions.

In both of these products, small-scale perturbations within eastward-directed semi-circular GWs are apparent just north of the gulf coast and over northern Florida. These observations are consistent with localized convective sources below, which was the case as seen in the MRMS lowest reflectivity mosaic at 18 UTC (2 pm local) on 22 July 2018 in Fig. 6, valid about 40 minutes prior to the AIRS data being collected overhead. Earlier analyses of reflectivities indicate these two convective features initiated approximately six hours earlier (8 am local) and so were rapidly developing up to the time of the AIRS overpass.

To simulate this case, the idealized WRF model was configured with 110 evenly-spaced vertical levels extending up to $z = 80$ km ($\Delta z \approx 727$ m), with a 10-km deep GW-absorbing sponge at the top. This depth was chosen in order to cover as much of the AIRS observational kernels within a physically-interpretable portion of the domain as possible. The idealized model was initialized 6 UTC, 22 July 2022 with the wind (Fig. 7) and stability (not shown) profile from MERRA2 and integrated forward 30 hours in time. Four simulations were completed, forced by diabatic heatings produced by the four tools described above updated every two minutes. Variables were output every 10 minutes .

4.2.2 Application of AIRS Observational Filters to WRF Output

In order to validate the four runs against AIRS data, both vertical and horizontal observational filters were applied to the WRF output to approximate the brightness temperature perturbations that would be seen by the AIRS sensor viewing through the simulated atmosphere. The vertical observational filter was applied first by taking the

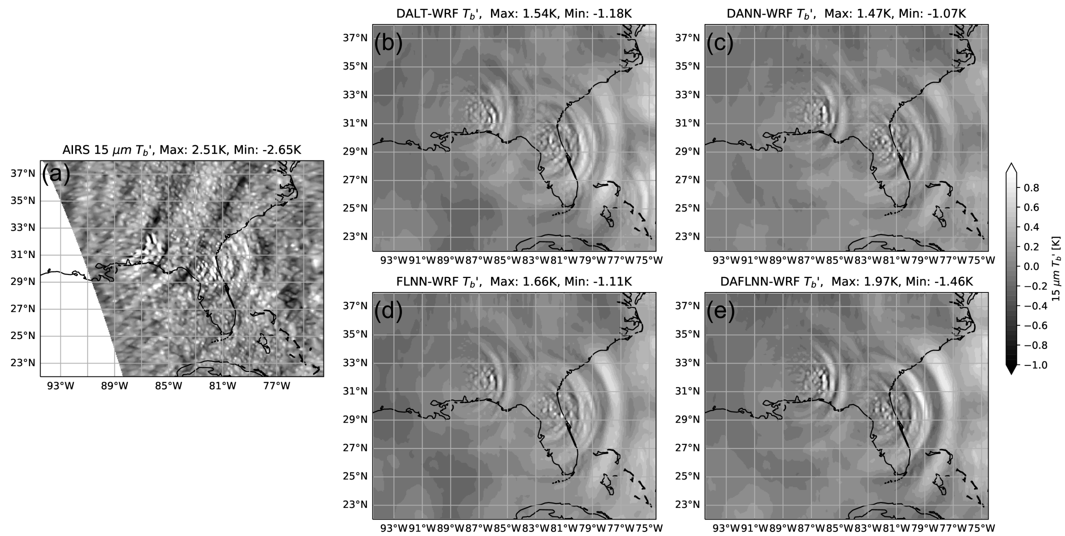

Figure 4. As in Fig. 3, but for the 15 μm product. Note the gray-shading range is twice that in Fig. 3.

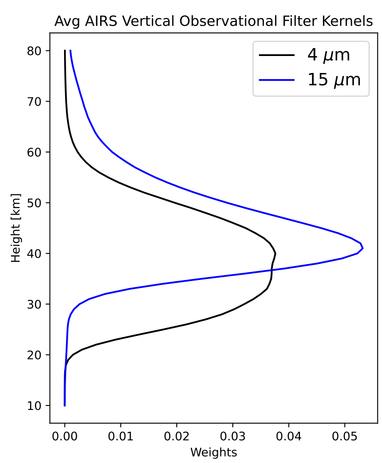

Figure 5. Average vertical observational filters for the 4 μm and 15 μm brightness temperature perturbation products. The 4 (15) μm kernel plotted here is the average of kernels of 42 (2) individual channels (Hoffmann et al., 2013, 2014, 2017) to reduce noise. These kernels were computed assuming climatological midlatitude atmospheric conditions.

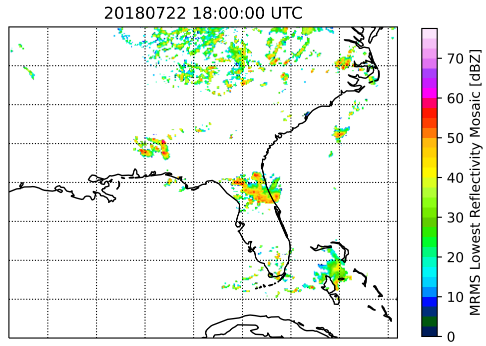

Figure 6. Multiple Radar, Multiple Sensor (MRMS) mosaic of lowest weather radar reflectivity valid 18 UTC on 22 July 2018, approximately 40 minutes prior to the AIRS observations in Figs. 3 and 4.

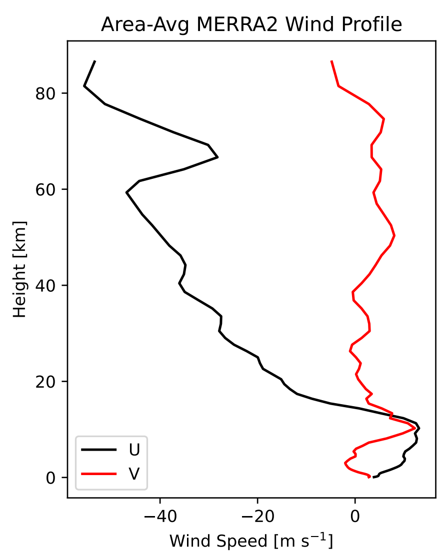

Figure 7. MERRA2 wind components area-averaged between 25N and 34N, 77W and 84W, valid 18 UTC on 22 July 2018. This wind (and stability, not shown) profile was used to initialize all idealized WRF simulations.

vertically-weighted average of WRF temperature perturbations (T') using the kernels in Fig. 5 as weights. Temperature perturbations were computed by first applying spatial high-pass filtering following Kruse and Smith (2015) to retain scales smaller than 500 km, similar to the high-pass filtering applied when removing background brightness temperature (T_b) from AIRS swaths (Hoffmann et al., 2013, 2014).

After application of the vertical observational filter, the simulated T_b' field is still at 2-km horizontal resolution, containing small-scale, large amplitude T_b'. However, the field of view of individual AIRS footprints is $\approx$ 13.5 km $\times$ 13.5 km at nadir, increasing to 41 km $\times$ 21.4 km at the edges of cross-track scans within an AIRS swath (Aumann et al., 2003; Hoffmann et al., 2013). Cross-track scans are $\approx$ 18 km apart, leading to a slight underlap of footprints in this direction. To roughly approximate the AIRS horizontal observational filters and scanning geometries, the 2-km resolution WRF-simulated T_b' were coarsened to 16-km resolution.

4.2.3 WRF Validation Against AIRS

The WRF-simulated T_b' approximately visible to the AIRS sensor are shown in panels (b-e) in Figs. 3 and 4. The model output time was 18:50 UTC, $\approx$ 7 minutes after the AIRS overpass over the region. Overall, the CGWs in WRF do resemble the CGWs emanating from the two regions of convection in the AIRS observations.

Small-scale T_b' features are apparent in both the observations and all WRF simulations. The larger-scale eastward propagating GW to the east of the convective sources also closely resembles those seen in the data. The minimum and maximum T_b' in WRF, due to the small-scale perturbations right above convection, are very comparable to those in the observations. Though, in the WRF output, T_b' minima and maxima were very sensitive to the degree to which WRF was coarsened. For example, coarsening to 20-km resolution reduced the simulated extrema by about half, due to significant small-scale CGW variability unresolved by AIRS. The amplitude of the CGW features to the east of the convection is quite comparable to that seen in the observations and not sensitive to the degree to which output was coarsened.

Several differences between the models and the observations can be noted as well, however. Phase-lines of the CGWs southeast of the convection appear slightly rotated clockwise relative to those in WRF. This might be due to latitudinal variations in the background flow (e.g. $\partial_y U$, $\partial_y\, \partial_z U$) in reality that were unrepresented by the horizontally-homogeneous profiles used to initialize WRF (i.e. Fig. 7). Additionally, observed large-scale GWs with northeast-southwest-oriented phase lines in the northern part of the domain are not present in the models. These GWs are likely due to sources outside of the spatiotemporal domain represented by WRF or outside of the region where convective forcing was supplied (Fig. 3e) and, hence, were not represented. Finally, the observations include significant noise, particularly in the 15 μm product (Fig. 4), where only two AIRS channels were averaged.

Brightness temperature perturbations along 29N are shown in Fig. 8. East of 79W, the CGW amplitudes *and phases* are very similar to the observations, at least in the 4 μm T_b' (panel a). The comparison east of 79W is not as good in the 15 μm product (panel b), though, the significant noise in the observations ($\sim$ 0.3K), potentially of similar amplitude to the CGWs according to WRF, obscures the comparison. While the CGWs do not obviously emerge from noise in such a transect, CGWs are visible through the noise when plotted spatially in Fig. 4a. (Note noise in the 4 μm channel is smaller $\sim$ 0.1 K.) The simulated CGWs (Fig. 4 (b-e)) do resemble those visible through the noise in the observations. West of 79W, the high-amplitude, small-scale perturbations in WRF do not match in phase with those observed (Fig. 8). Simulated perturbation amplitudes are similar to the observations, being similar in the 4 μm product and slightly smaller in the 15 μm product. Perhaps the simulated amplitudes could be made more comparable with

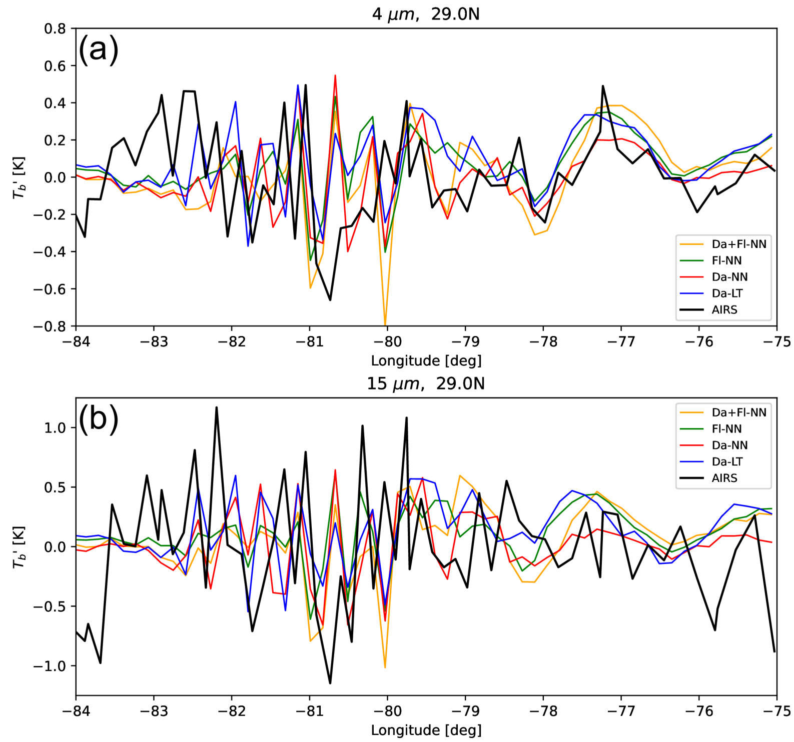

Figure 8. Brightness temperature comparison along 29N over Florida and to the east. WRF output was coarsened to 16-km to approximate an average horizontal observational filter of AIRS.

the observations with a more realistic treatment of AIRS footprint geometries and sizes and/or the addition of noise to the WRF output, however, this was not performed here. Still, the exact locations and phases of these small-scale CGWs right above the sources are likely inherently unpredictable, meaning matching simulated phases with observations may not be realistic.

While the evaluations of diabatic heating predictions by the four tools could suggest one tool is better than the other (e.g. comparing MAE from the DANN versus the DAFLNN in Fig. 2), the CGWs produced by all four diabatic heatings are quite similar between the four runs (Figs. 3, 4, 8). It is unclear if the small differences in AIRS-visible simulated CGWs between the four simulations are significant, being attributable to differences in the diabatic heatings, or if these differences are essentially within an ensemble spread where only diabatic heatings were purturbed (i.e. indistinguishable). As such, it is difficult to claim one tool is better than the other when validating the simulated CGWs against the AIRS observations. However, the similarity of CGWs between the four solutions, all resembling the observations quite well, allows the conclusion that if a reasonable diabatic heating, in this case learned from a microphysics parameterization within a covection-permitting and not convection-resolving simulation, is supplied to a GW-resolving model at correct locations and times, the CGWs resulting from this forcing can be quite realistic.

4.3 WRF Validation Against Loon Super-Pressure Balloons

For further evaluation, another case was simulated using the modified idealized WRF configuration. Here, a case of typical diurnal convection over Florida was simulated that

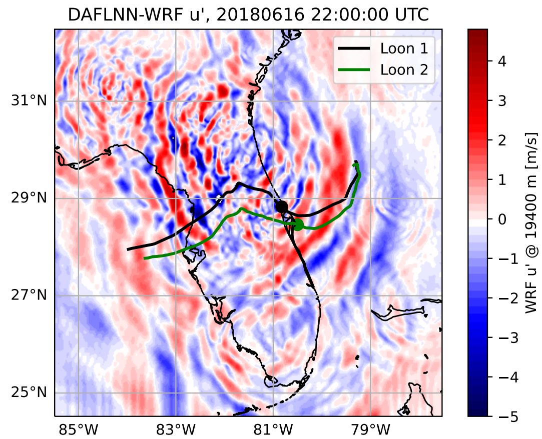

Figure 9. Horizontal cross-section of u' at $z = 19400$ m at 22 UTC on 16 June 2018. Here, the entire WRF domain is shown. The idealized WRF model was initialized 10 hours prior to the valid time. The two Loon super-pressure balloon tracks are shown during the 24 hour period beginning at 12 UTC, 16 June 2018. The circles indicate the positions of the super-pressure balloons at the valid time. The height was chosen to be an approximate average height of the balloons (c.f. Fig. 10).

happened to have two super-pressure balloons, flown by Project Loon (hereafter Loon), advecting from east-to-west overhead near $z = 19.4$ km. Loon was a Google project, and later an Alphabet subsidiary, that flew 2131 super-pressure balloons nearly globally in order to provide wireless internet access to rural areas (Rhodes & Candido, 2021). Loon balloons carried a payload with instruments measuring pressure, temperature, and horizontal velocities from GPS (Friedrich et al., 2017) at 1 Hz. These balloons also had the capability of changing their density, allowing some altitude control and steering by catching winds at different altitudes. A flag recorded when vertical maneuvering occurred. These data have been used in a handful of studies up to this point (Friedrich et al., 2017; Schoeberl et al., 2017; Conway et al., 2019; Lindgren et al., 2020). Only the 1 Hz location, height, and horizontal wind observations are used here.

For this case, an 800 km $\times$ 800 km $\times$ 55 km domain was used at $\Delta x = 2$-km horizontal resolution and $\Delta z = 500$ m vertical resolution via 110 evenly spaced vertical levels. The idealized configuration was initialized at 12 UTC on 16 June 2018 and integrated 24 hours in time. Diabatic heatings were again computed from MRMS data via the same LT and three NNs and supplied to WRF every two minutes. Figure 9 shows the zonal wind perturbations relative to the initial profile at $z = 19.4$ km, near the altitude of two Loon super-pressure balloons (Fig. 10), valid at 22 UTC on 16 June 2018. The tracks of the two super-pressure balloons, along with their locations at the output valid time, are also depicted in Fig. 9.

The WRF output was then 4-D linearly interpolated to the time, altitude, latitude, and longitude of the observations taken by both Loon flights during the 24 hours of the four simulations. The Loon height, zonal wind perturbation, and meridional wind perturbation time series for both flights are shown in Fig. 10. Initially, there are no perturbations occuring at the Loon locations, as the convective forcing did not begin immediately and when it did occur, it was some distance northwest. When the CGWs do reach the Loon locations, as noted in the previous case, the differences in heatings provided by the four tools do not seem to result in significant differences in the simulated CGWs they force.

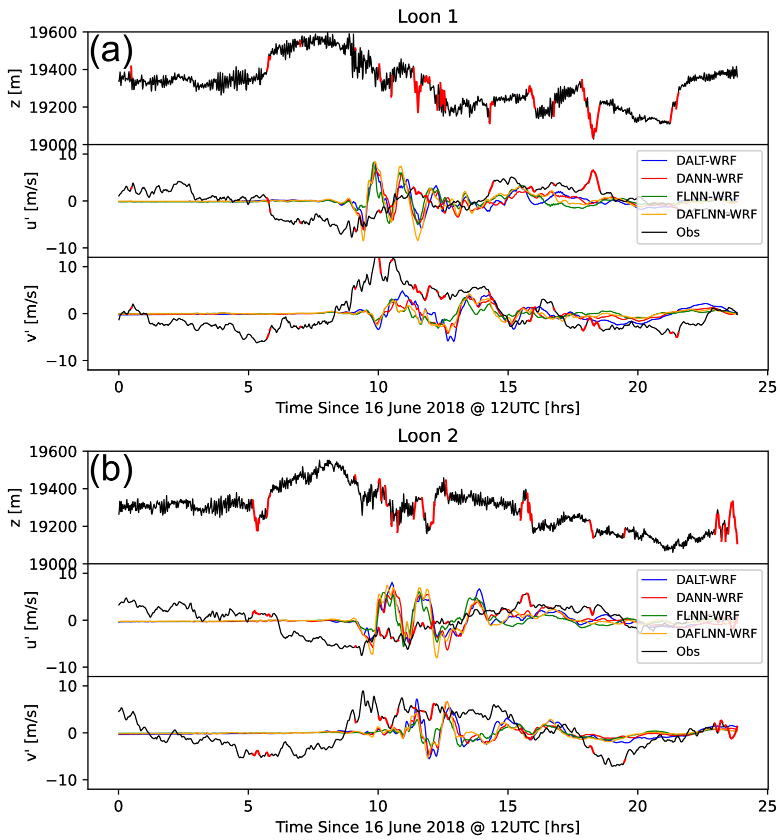

Figure 10. Time series of Loon super-pressure balloon GPS altitude, zonal wind perturbations, and meridional wind perturbations (black) for two Loon flights that happened to drift over Florida 16 June 2018. The four WRF runs were 4-D linearly interpolated (colors) to the latitudes, longitudes, heights, and times for comparison with the observations (obs).The portions of the observed time series (black) highlighted in red indicate periods where the super-pressure balloon was maneuvering vertically.

In this case, none of the idealized WRF simulations were able to well-reproduce the observations. Here, simulated wind speed perturbations are relative to the initial wind at the altitude of interest. The Loon perturbations are relative to the mean over the 24 hour period presented. The simulated $u' = u(t) - u(t_{init})$ amplitudes were generally notably higher than in the Loon observations. The simulated v' compared, perhaps, slightly better to the observations. Likely the best point of comparison was in the arrival times of the CGWs to the Loon locations. For example, about 8 hours after initialization, the appearance of significant simulated CGW perturbations appear. This timing roughly corresponds to when higher-frequency variability appears in Loon as well.

It is difficult to say whether or not the overall method of recreating CGWs did not work in this case. While the time series comparisons are poor (Fig. 10) and wind speed uncertainty is reported to be much smaller (0.23 m s^{-1}, Friedrich et al. (2017)) than the observed variations, data in this case are limited to only two transects. Comparisons of GWs along individual transects can be misleading, as small differences in the location of interest relative to the GWs can lead to significant differences of the apparent GW field sampled on a transect when, spatially, the GW fields are similar (c.f. Fig. 8, 4). Additionally, the data quality is somewhat questionable in this particular case. The por-

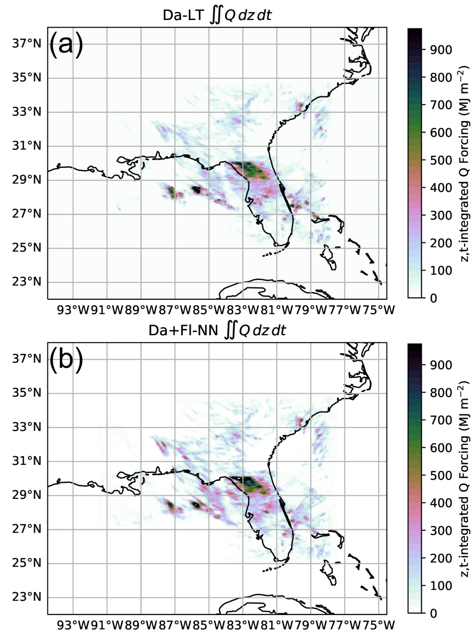

Figure 11. Height- and time-integrated latent heating (Q) predicted by (a) the look-up table method and (b) the DAFLNN on the WRF domain for the AIRS case. Latent heating was zeroed outside of the dashed line in (d).

tions of the Loon time series highlighted in red indicate times when the super-pressure balloon was vertically maneuvering by changing its density. It is unknown if this maneuvering was performed to steer the balloons or an automated response to oppose the influences of CGWs.

5 GW Analysis of the AIRS Case

A primary motivation for the overall method of forcing an idealized model with weather-radar-derived diabatic heating was to produce validatable simulations of CGWs and then use these validated simulations to study CGWs. Here, the CGWs within the AIRS-validated case above are briefly analyzed. The objectives are to see how far laterally CGWs can propagate in this case, to see where they dissipate, and how strong the drag decelerations are. All of these objectives are currently relevant to the development and improvement of GW parameterization in weather and climate models, which has not been well constrained by observations or constrained by directly validated CGW-resolving simulations such as these.

Over the entire 30-hour AIRS-validated WRF simulation, the convective diabatic forcings were fairly compact. The height- and time-integrated diabatic heating over the entire simulation is shown from the DALT and DAFLNN predictions in Fig. 11. The corresponding maps from the DANN and FLNN predictions were largely similar and so are not shown. The most intense, prolonged heating resulted from the convective region over northern Florida, with more localized and weaker forcings scattered within the domain elsewhere. This localization of CGW forcing simplifies interpretation of GW analyses somewhat, as the CGWs can largely be interpreted as being generated by a single localized source.

GW amplitudes are illustrated in Fig. 12. Amplitudes were computed using the discrete Hilbert Transform following Eckermann et al. (2015) and Mercier et al. (2008), allowing phase-averaged quantities to be produced in physical (e.g. x,y) space. These amplitudes were then averaged over all output times during the 30-hour simulation, from output every two minutes. At $z = 40$ km, CGWs are most apparent over and to the east of the diabatic forcing (c.f. Figs. 12a, b, c, g and 11b). The prevalence of CGW activity to the east is largely expected, considering the strong easterly wind shear in the ambient winds below this altitude (Fig. 7) forcing critical-level dissipation of the westward-propagating CGWs. The CGW activity is most spread out according to u' amplitudes (Fig. 12a) and most localized according to w' amplitudes (Fig. 12b), with the spread of vertical fluxes of horizontal momentum ($MF_x = \overline{\rho}\,\widehat{u'w'}$, $MF_y = \overline{\rho}\,\widehat{v'w'}$ with hats here indicating phase averaging via Hilbert transform) in between. In terms of momentum flux, CGWs can clearly propagate O(1000) km away from their source, consistent with the modeling study of Sun et al. (2023), observational study of Corcos et al. (2021) and inconsistent with the conventional column-approximation in parameterizations.

Vertical fluxes of zonal (b-d) and meridional (f-h) momentum are shown at $z =$ 20 km, 40 km, and 60 km in Fig. 12 to give a sense for how CGWs both dissipate and spread with height. The color shading scales are reduced with height, implying CGW dissipation and momentum deposition. Alternatively, lateral spreading can result in spreading and reduction of fluxes, too (Eckermann et al., 2015). However, the spatial extents do not appear to change significantly with height, suggesting GW dissipation.

The meridional spread of these validated CGWs are shown in Fig. 13, where zonally- and temporally-averaged ($\overline{(\,.\,)}^{xt} = L^{-1}T^{-1}\int\int(\,.\,)\,dt\,dx$, where L and T are the lengths and periods over which the quantity is averaged) wave and convective quantities are shown as a function of latitude and height. The largest CGW amplitudes occur directly over the highest diabatic heating, but extend north and south of the peak heating with height (Fig. 13a, b). This spread is also seen in the contours of vertical flux of zonal and meridional momentum (Fig. 13c, d), though, this spread with height is more subtle in this variable.

The zonal and meridional CGW drag was quantified via

$$(GWD_x, GWD_y) = -\frac{1}{\overline{\rho}^{xt}}\frac{\partial}{\partial z}\left(\overline{\rho}^{xt}\overline{u'w'}^{xt}, \overline{\rho}^{xt}\overline{v'w'}^{xt}\right) \tag{1}$$

and shown in panels (e-f). The influences of lateral divergences of lateral fluxes of horizontal momentum can be important (Sun et al., 2023), but were not investigated here. The vertical profiles of zonal drag are largely consistent with linear GW theory. Westward-propagating GWs producing negative zonal momentum flux encountered critical levels and dissipated in the region of strong negative zonal wind shear between $z = 15$ km and 20 km (Fig. 7). This results in negative drags of ≈ 1 m s^{-1} day^{-1}, though, these values of drag are somewhat subjective as they depend on the choices made in areas over which fluxes were averaged. The eastward-propagating waves do not encounter critical levels, but do grow with altitude and gradually reach overturning amplitudes and dissipate, indicated by the general increase in positive drag with height (Fig. 13e). However, zonal drags rise sharply in the layers of positive shear above $z = 35$ km, as CGWs propagating into these layers encounter shear that brings the environment a bit closer to their phase speeds, forces GWs toward steepening and saturating (see Kruse et al. (2016) for further discussion on this effect, but for orogaphic GWs). The growth in amplitudes due to these local zonal wind maxima can be seen in Fig. 13a and b. Interestingly, the zonal and meridional drags are fairly invariant in latitude despite localized forcing, except at the highest altitudes, highlighting the effects of lateral propagation on drag.

It should be noted that the idealized WRF configuration used no physical parameterizations and so did not use a turbulence parameterization. Also, the vertical resolution of $\Delta z = 727$ m may be coarse relative to the scales of motions involved in CGW breakdown. Both simulation characteristics will likely affect some details of how and where these simulated CGWs break and deposit momentum. Testing how turbulence param-

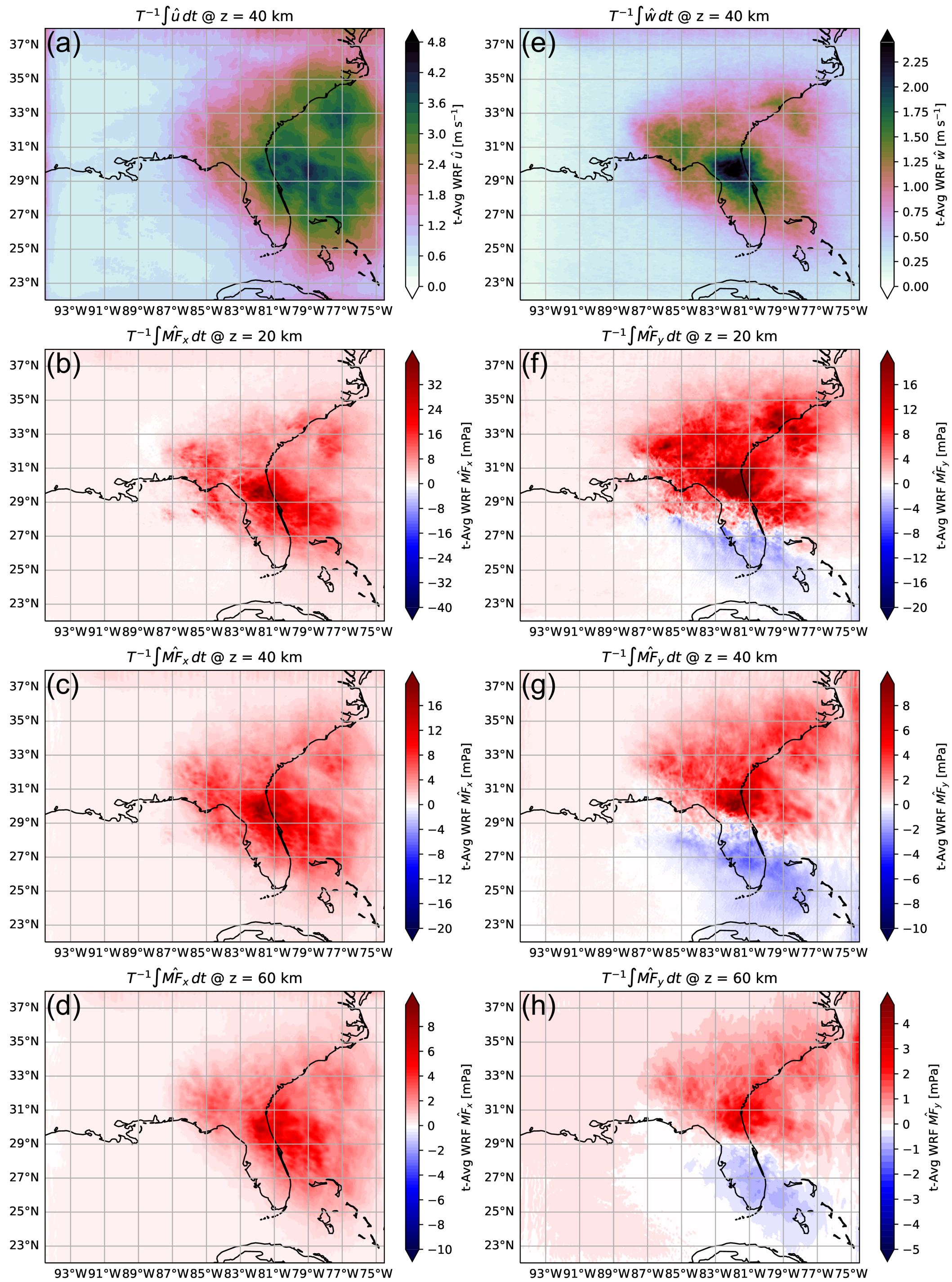

Figure 12. Phase-averaged (via Hilbert transform, $(\hat{.})$), time-averaged GW amplitudes of (a) u', (e) w', (b-d) vertical flux of zonal momentum, (f-h) and vertical flux of meridional momentum at selected levels indicated in the panel titles. These analyses are of the DAFLNN-forced WRF run. Comparison with Fig. 11d gives an indication of how different variables tend to spread laterally and how this spread varies with height. Note every panel has an individual color shading range.

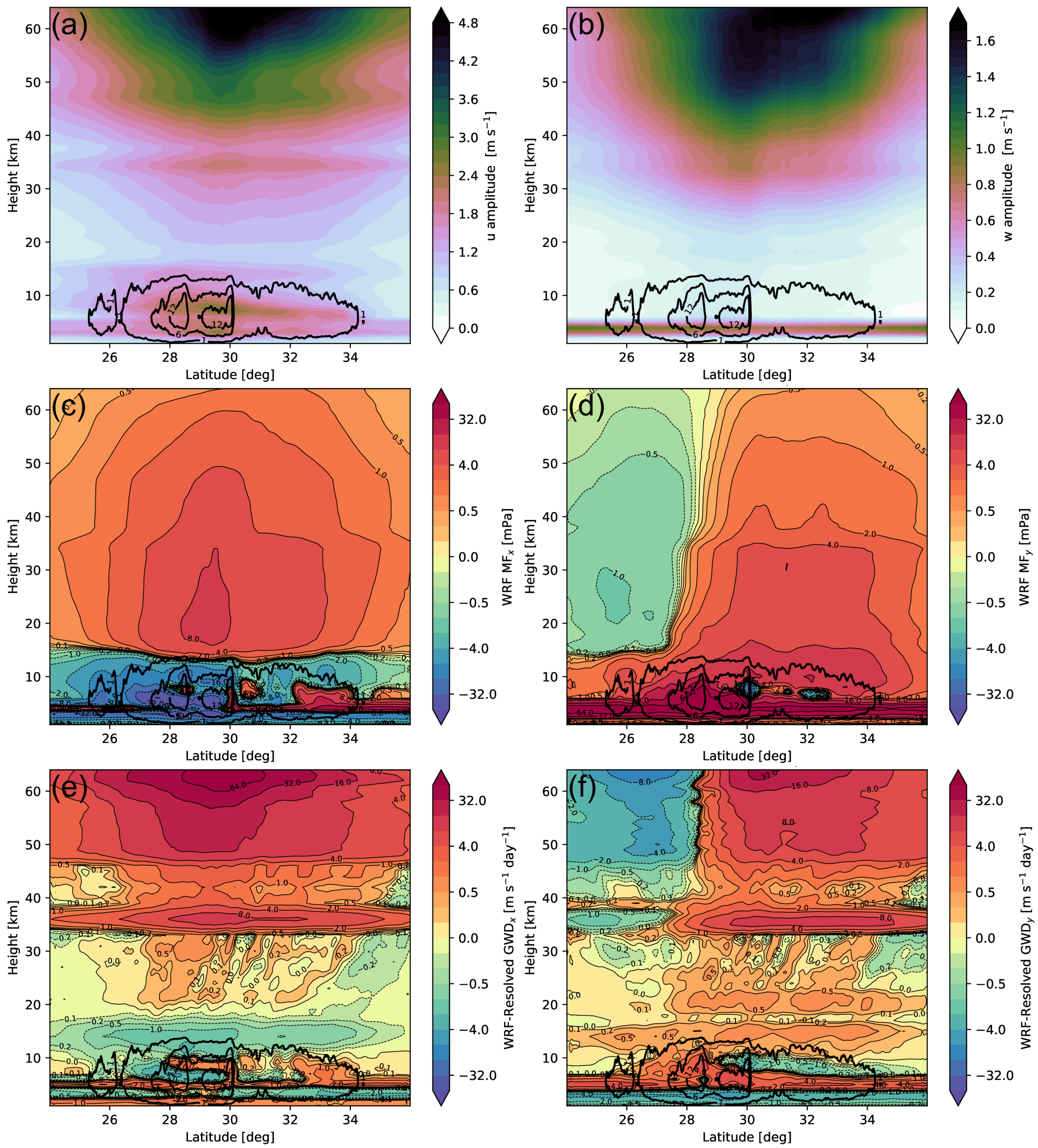

Figure 13. Time- and zonal-mean (a) u' amplitude, (b) w' amplitude, (c) vertical flux of zonal momentum, (d) vertical flux of meridional momentum, (e) zonal GWD, and (f) meridional GWD. The entire 30-hour simulation was included in the time averaging. The outer 200 km of the domain were excluded. The vertical fluxes of horizontal momentum and zonal drags were smoothed along latitude with a 42-km moving average smoother. The thick black contours depict the time-, zonal-mean latent heating at 1, 6, and 12 K day^{-1}.

eterizations and vertical resolution affect drag on the mesoscales is certainly warranted, but is left to future work.

6 Discussion and Conclusions

If reasonably realistic diabatic heating is supplied at the correct locations and times in a GW-resolving model, the CGWs generated within that model can resemble observed CGWs quite well. This overall method (i.e. forcing CGW-resolving simulations with observations of convection) shows significant promise in furthering CGW research and parameterization development *with confidence*, as it allows full 4-D fields of realistic CGWs to be generated and analyzed rigorously.

Here, diabatic heating was learned from full-physics, $\Delta x = 2$-km, $\Delta z < 500$-m resolution WRF simulations. These simulations were convection-permitting, but not convection-resolving (Jeevanjee, 2017; Jeevanjee & Zhou, 2022), and diabatic heatings are predicted by the WRF Single-Moment 6-class (WSM6) microphysics scheme (Hong & Lim, 2006). The good agreement between simulated and observed CGWs (Figs. 3, 4) suggests the convection permitted by these resolutions and the heatings predicted by this microphysics scheme are reasonably realistic, at least when it comes to CGW forcing.

The look-up table method and NNs had similar skill at predicting the WRF-simulated diabatic heating profiles at larger rain rates, while the NNs showed promise at being better able to represent complexities in heating profiles (e.g. evaporative cooling layers) at smaller rain rates. The vast majority of gridpoints deemed "convective" (i.e. having a rain rate exceeding 1 mm $(10\ \text{min})^{-1}$) had these smaller rain rates. This increased performance by NNs at smaller rain rates could be attributable to the inherent ability of such an architecture to represent such profiles, potentially the increased information contained by the additional radar reflectivities used as input, or just a reflection the NNs being trained mostly small-rain-rate profiles. Perhaps a proper hyperparameter optimization, a loss function used to emphasize skill of the larger-amplitude heating profiles, or an architecture more appropriate for this application (e.g. one that uses spatial input to account for the 3-D tilting of convection observations due to wind shear) could enhance skill in this application over all rain rates.

While machine learning methods still have significant potential to further improve skill in predicting convective diabatic heating beyond conventional methods (e.g. look-up tables), variations in CGWs generated by the different heatings predicted here were small. It is unclear if better heatings will be significant when it comes to CGW forcing.

In the $\Delta x = 2$-km, $\Delta z = 727$-m resolution idealized WRF configuration used here, larger-scale CGWs that apparently propagate more laterally validated the best against AIRS observations, with both phases and amplitudes reproduced reasonably well *quantitatively*. The WRF configuration was also able to reproduce the smaller-scale, more vertically-propagating CGWs above convective sources as well, at least in amplitudes. Still, these small-scale CGWs were highly sensitive to the details sampling a simulation as if AIRS were viewing through it. A more accurate treatment of how AIRS might sample these simulated CGWs that takes into account viewing geometries of individual footprints, variations in horizontal observational filtering with viewing zenith angle, and perhaps even radiative transfer would likely alter how a hypothetical AIRS sensor would see these CGWs. This is particularly relevant as these small-scale CGWs right over the convection are responsible for much of the momentum flux (Fig. 12).

Finally, CGWs are inherently non-stationary and propagate away from the convection. A spectrum of horizontal and vertical group velocities is generated. In the validated simulation presented here, it is clear CGWs propagate 100s of kilometers away from the convective sources. The most momentum fluxed and drag deposited does occur above the convective sources, but significant drags still occur 100s of kilometers away.

These results provide more evidence for relaxing the commonly employed single-column approximation in GW parameterizations, which assumes GWs propagate only vertically.

7 Open Research

The MRMS reflectivity data were retrieved from an archive at Iowa State University, accessible here: `https://mtarchive.geol.iastate.edu/2018/06/16/mrms/ncep/SeamlessHSR/`. Numerous other variables are derived from the MRMS dataset in real time, but are not publicly archived. Chuntao Liu, at Texas A&M University - Corpus Christi, has personally archived a handful of these additional variables, and the precipitation rates, echo top heights, and reflecitivies at 0^oC, -10^oC and -20^oC isotherms were provided by him. The AIRS brightness temperature data products applied in this study are available open access (Hoffmann, 2021). DAFLNN-forced WRF output at 10-minute output frequency, the trained NNs, parsed training data, the Bramberger et al. (2020) lookup-table, time-averaged DAFLNN-forced WRF output, and all Python scripts and WRF source codes used are archived at the Stanford Digital Repository (https://doi.org/10.25740/kq456hs1417).

Acknowledgments

This work was supported by grants from the NSF OAC CSSI program (NSF Grants 2005123 and 2004512) and by the generosity of Eric and Wendy Schmidt by recommendation of the Schmidt Futures program. This work was largely performed with high-performance computing support from Cheyenne (Computational Information Systems Laboratory, 2019) provided by NCAR's Computational and Information Systems Laboratory, sponsored by the National Science Foundation. Finally, we would like to acknowledge physics suite recommendations by Jimy Dudhia at NCAR and very helpful discussions on machine learning with Minah Yang at NYU.

References

Abhik, S., & Hendon, H. H. (2019). Influence of the qbo on the mjo during coupled model multiweek forecasts. *Geophysical Research Letters*, *46*(15), 9213-9221. doi: https://doi.org/10.1029/2019GL083152

Alexander, M. J., Geller, M., McLandress, C., Polavarapu, S., Preusse, P., Sassi, F., ... Watanabe, S. (2010). Recent developments in gravity-wave effects in climate models and the global distribution of gravity-wave momentum flux from observations and models. *Quarterly Journal of the Royal Meteorological Society*, *136*(650), 1103-1124. doi: https://doi.org/10.1002/qj.637

Alexander, M. J., Liu, C. C., Bacmeister, J., Bramberger, M., Hertzog, A., & Richter, J. H. (2021). Observational validation of parameterized gravity waves from tropical convection in the whole atmosphere community climate model. *Journal of Geophysical Research: Atmospheres*, *126*(7), e2020JD033954. (e2020JD033954 2020JD033954) doi: https://doi.org/10.1029/2020JD033954

Alexander, M. J., & Rosenlof, K. H. (2003). Gravity-wave forcing in the stratosphere: Observational constraints from the upper atmosphere research satellite and implications for parameterization in global models. *Journal of Geophysical Research: Atmospheres*, *108*(D19). doi: https://doi.org/10.1029/2003JD003373

Anstey, J. A., Simpson, I. R., Richter, J. H., Naoe, H., Taguchi, M., Serva, F., ... Yukimoto, S. (2021). Teleconnections of the quasi-biennial oscillation in a multi-model ensemble of QBO-resolving models. *Quarterly Journal of the Royal Meteorological Society*, 1–25. doi: 10.1002/qj.4048

Aumann, H., Chahine, M., Gautier, C., Goldberg, M., Kalnay, E., McMillin, L., ... Susskind, J. (2003). Airs/amsu/hsb on the aqua mission: design, science objectives, data products, and processing systems. *IEEE*

Transactions on Geoscience and Remote Sensing, *41*(2), 253-264. doi: 10.1109/TGRS.2002.808356

Baldwin, M. P., Gray, L. J., Dunkerton, T. J., Hamilton, K., Haynes, P. H., Randel, W. J., … Takahashi, M. (2001). The quasi-biennial oscillation. *Reviews of Geophysics*, *39*(2), 179-229. doi: https://doi.org/10.1029/1999RG000073

Bramberger, M., Alexander, M. J., & Grimsdell, A. W. (2020). Realistic simulation of tropical atmospheric gravity waves using radar-observed precipitation rate and echo top height. *Journal of Advances in Modeling Earth Systems*, *12*(8), e2019MS001949. doi: https://doi.org/10.1029/2019MS001949

Bushell, A. C., Anstey, J. A., Butchart, N., Kawatani, Y., Osprey, S. M., Richter, J. H., … Yukimoto, S. (2022). Evaluation of the quasi-biennial oscillation in global climate models for the sparc qbo-initiative. *Quarterly Journal of the Royal Meteorological Society*, *148*(744), 1459-1489. doi: https://doi.org/10.1002/qj.3765

Computational Information Systems Laboratory. (2019). *Cheyenne: HPE/SGI ICE XA System (University Community Computing)* [Supercomputer]. Boulder, CO: National Center for Atmospheric Research. doi:10.5065/D6RX99HX.

Conway, J. P., Bodeker, G. E., Waugh, D. W., Murphy, D. J., Cameron, C., & Lewis, J. (2019). Using project loon superpressure balloon observations to investigate the inertial peak in the intrinsic wind spectrum in the midlatitude stratosphere. *Journal of Geophysical Research: Atmospheres*, *124*(15), 8594-8604. doi: https://doi.org/10.1029/2018JD030195

Corcos, M., Hertzog, A., Plougonven, R., & Podglajen, A. (2021). Observation of gravity waves at the tropical tropopause using superpressure balloons. *Journal of Geophysical Research: Atmospheres*, *126*(15), e2021JD035165. (e2021JD035165 2021JD035165) doi: https://doi.org/10.1029/2021JD035165

Eckermann, S. D., Ma, J., & Broutman, D. (2015). Effects of horizontal geometrical spreading on the parameterization of orographic gravity wave drag. part i: Numerical transform solutions. *Journal of the Atmospheric Sciences*, *72*(6), 2330 - 2347. doi: 10.1175/JAS-D-14-0147.1

Friedrich, L. S., McDonald, A. J., Bodeker, G. E., Cooper, K. E., Lewis, J., & Paterson, A. J. (2017). A comparison of loon balloon observations and stratospheric reanalysis products. *Atmospheric Chemistry and Physics*, *17*(2), 855–866. doi: 10.5194/acp-17-855-2017

Garfinkel, C. I., Schwartz, C., Domeisen, D. I. V., Son, S.-W., Butler, A. H., & White, I. P. (2018). Extratropical atmospheric predictability from the quasi-biennial oscillation in subseasonal forecast models. *Journal of Geophysical Research: Atmospheres*, *123*(15), 7855-7866. doi: https://doi.org/10.1029/2018JD028724

Gelaro, R., McCarty, W., Suárez, M. J., Todling, R., Molod, A., Takacs, L., … Zhao, B. (2017). The modern-era retrospective analysis for research and applications, version 2 (merra-2). *Journal of Climate*, *30*(14), 5419 - 5454. doi: 10.1175/JCLI-D-16-0758.1

Gray, L. J., Anstey, J. A., Kawatani, Y., Lu, H., Osprey, S., & Schenzinger, V. (2018). Surface impacts of the Quasi Biennial Oscillation. *Atmos. Chem. Phys.*, *18*(11), 8227–8247. doi: 10.5194/acp-18-8227-2018

Grimsdell, A. W., Alexander, M. J., May, P. T., & Hoffmann, L. (2010). Model study of waves generated by convection with direct validation via satellite. *Journal of the Atmospheric Sciences*, *67*(5), 1617 - 1631. doi: 10.1175/2009JAS3197.1

Hegglin, M. I., & Shepherd, T. G. (2009). Large climate-induced changes in ultraviolet index and stratosphere-to-troposphere ozone flux. *Nature Geoscience*, *2*(10), 687–691.

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., … Thépaut, J.-N. (2020). The era5 global reanalysis. *Quarterly Jour-*

nal of the Royal Meteorological Society, *146*(730), 1999-2049. Retrieved from https://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/qj.3803 doi: https://doi.org/10.1002/qj.3803

Hoffmann, L. (2021). *AIRS/Aqua Observations of Gravity Waves* [Dataset]. (Jülich DATA, V1) doi: https://doi.org/10.26165/JUELICH-DATA/LQAAJA

Hoffmann, L., Alexander, M. J., Clerbaux, C., Grimsdell, A. W., Meyer, C. I., Rößler, T., & Tournier, B. (2014). Intercomparison of stratospheric gravity wave observations with airs and iasi. *Atmospheric Measurement Techniques*, *7*(12), 4517–4537. doi: 10.5194/amt-7-4517-2014

Hoffmann, L., Spang, R., Orr, A., Alexander, M. J., Holt, L. A., & Stein, O. (2017). A decadal satellite record of gravity wave activity in the lower stratosphere to study polar stratospheric cloud formation. *Atmospheric Chemistry and Physics*, *17*(4), 2901–2920. doi: 10.5194/acp-17-2901-2017

Hoffmann, L., Xue, X., & Alexander, M. J. (2013). A global view of stratospheric gravity wave hotspots located with atmospheric infrared sounder observations. *Journal of Geophysical Research: Atmospheres*, *118*(2), 416-434. doi: https://doi.org/10.1029/2012JD018658

Holt, L. A., Alexander, M. J., Coy, L., Liu, C., Molod, A., Putman, W., & Pawson, S. (2017). An evaluation of gravity waves and gravity wave sources in the southern hemisphere in a 7 km global climate simulation. *Quarterly Journal of the Royal Meteorological Society*, *143*(707), 2481-2495. doi: https://doi.org/10.1002/qj.3101

Holt, L. A., Alexander, M. J., Coy, L., Molod, A., Putman, W., & Pawson, S. (2016). Tropical waves and the quasi-biennial oscillation in a 7-km global climate simulation. *Journal of the Atmospheric Sciences*, *73*(9), 3771 - 3783. doi: 10.1175/JAS-D-15-0350.1

Hong, S.-Y., & Lim, J.-O. J. (2006). The wrf single-moment 6-class microphysics scheme (wsm6). *Asia-Pacific Journal of Atmospheric Sciences*, *42*(2), 129–151.

Jeevanjee, N. (2017). Vertical velocity in the gray zone. *Journal of Advances in Modeling Earth Systems*, *9*(6), 2304-2316. doi: https://doi.org/10.1002/2017MS001059

Jeevanjee, N., & Zhou, L. (2022). On the resolution-dependence of anvil cloud fraction and precipitation efficiency in radiative-convective equilibrium. *Journal of Advances in Modeling Earth Systems*, *14*(3), e2021MS002759. (e2021MS002759 2021MS002759) doi: https://doi.org/10.1029/2021MS002759

Kruse, C. G., & Smith, R. B. (2015). Gravity wave diagnostics and characteristics in mesoscale fields. *Journal of the Atmospheric Sciences*, *72*(11), 4372 - 4392. doi: 10.1175/JAS-D-15-0079.1

Kruse, C. G., Smith, R. B., & Eckermann, S. D. (2016). The midlatitude lower-stratospheric mountain wave "valve layer". *Journal of the Atmospheric Sciences*, *73*(12), 5081 - 5100. doi: 10.1175/JAS-D-16-0173.1

Lang, S. E., & Tao, W.-K. (2018). The next-generation goddard convective–stratiform heating algorithm: New tropical and warm-season retrievals for gpm. *Journal of Climate*, *31*(15), 5997 - 6026. doi: https://doi.org/10.1175/JCLI-D-17-0224.1

Lee, H.-K., Kang, M.-J., Chun, H.-Y., Kim, D., & Shin, D.-B. (2022). Characteristics of latent heating rate from gpm and convective gravity wave momentum flux calculated using the gpm data. *J. Geophys. Res.: Atmospheres*, *127*(17), e2022JD037003. doi: https://doi.org/10.1029/2022JD037003

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. (e2020JD032850 10.1029/2020JD032850) doi: https://doi.org/10.1029/2020JD032850

Liu, C., Alexander, J., Richter, J., & Bacmeister, J. (2022). Using trmm latent heat as a source to estimate convection induced gravity wave momentum flux in the lower stratosphere. *Journal of Geophysical Research: Atmospheres*, *127*(1), e2021JD035785. (e2021JD035785 2021JD035785) doi: https://doi.org/10.1029/2021JD035785

Marshall, A. G., Hendon, H. H., Son, S.-W., & Lim, Y. (2017). Impact of the quasi-biennial oscillation on predictability of the madden–julian oscillation. *Climate Dynamics*, *49*(4), 1365–1377.

Martin, Z., Son, S.-W., Butler, A., Hendon, H., Kim, H., Sobel, A., ... Zhang, C. (2021). The influence of the quasi-biennial oscillation on the madden–julian oscillation. *Nature Reviews Earth & Environment*, *2*(7), 477–489.

Mercier, M. J., Garnier, N. B., & Dauxois, T. (2008). Reflection and diffraction of internal waves analyzed with the hilbert transform. *Physics of Fluids*, *20*(8), 086601. doi: 10.1063/1.2963136

Plougonven, R., Hertzog, A., & Alexander, M. J. (2015). Case studies of nonorographic gravity waves over the southern ocean emphasize the role of moisture. *Journal of Geophysical Research: Atmospheres*, *120*(4), 1278-1299. doi: https://doi.org/10.1002/2014JD022332

Rhodes, B., & Candido, S. (2021). *Loon stratospheric sensor data.* Zenodo. Retrieved from `https://doi.org/10.5281/zenodo.5119968` doi: 10.5281/zenodo.5119968

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. (e2019JD032362 10.1029/2019JD032362) doi: https://doi.org/10.1029/2019JD032362

Richter, J. H., Butchart, N., Kawatani, Y., Bushell, A. C., Holt, L., Serva, F., ... Yukimoto, S. (2022). Response of the quasi-biennial oscillation to a warming climate in global climate models. *Quarterly Journal of the Royal Meteorological Society*, *148*(744), 1490-1518. doi: https://doi.org/10.1002/qj.3749

Schoeberl, M. R., Jensen, E., Podglajen, A., Coy, L., Lodha, C., Candido, S., & Carver, R. (2017). Gravity wave spectra in the lower stratosphere diagnosed from project loon balloon trajectories. *Journal of Geophysical Research: Atmospheres*, *122*(16), 8517-8524. doi: https://doi.org/10.1002/2017JD026471

Stephan, C., & Alexander, M. J. (2015). Realistic simulations of atmospheric gravity waves over the continental u.s. using precipitation radar data. *Journal of Advances in Modeling Earth Systems*, *7*(2), 823-835. doi: https://doi.org/10.1002/2014MS000396

Stephan, C., Alexander, M. J., & Richter, J. H. (2016). Characteristics of gravity waves from convection and implications for their parameterization in global circulation models. *Journal of the Atmospheric Sciences*, *73*(7), 2729 - 2742. doi: 10.1175/JAS-D-15-0303.1

Stephan, C. C., Alexander, M. J., Hedlin, M., de Groot-Hedlin, C. D., & Hoffmann, L. (2016). A case study on the far-field properties of propagating tropospheric gravity waves. *Monthly Weather Review*, *144*(8), 2947 - 2961. doi: 10.1175/MWR-D-16-0054.1

Stephan, C. C., Strube, C., Klocke, D., Ern, M., Hoffmann, L., Preusse, P., & Schmidt, H. (2019a). Gravity waves in global high-resolution simulations with explicit and parameterized convection. *Journal of Geophysical Research: Atmospheres*, *124*(8), 4446-4459. doi: https://doi.org/10.1029/2018JD030073

Stephan, C. C., Strube, C., Klocke, D., Ern, M., Hoffmann, L., Preusse, P., & Schmidt, H. (2019b). Intercomparison of gravity waves in global convection-permitting models. *Journal of the Atmospheric Sciences*, *76*(9), 2739 - 2759. doi: 10.1175/JAS-D-19-0040.1

Sun, Y. Q., Hassanzadeh, P., Alexander, M. J., & Kruse, C. G. (2023). Quantifying

3d gravity wave drag in a library of tropical convection-permitting simulations for data-driven parameterizations. *Authorea Preprints*.

Tao, W.-K., Iguchi, T., & Lang, S. (2019). Expanding the goddard csh algorithm for gpm: New extratropical retrievals. *Journal of Applied Meteorology and Climatology*, *58*(5), 921 - 946. doi: https://doi.org/10.1175/JAMC-D-18-0215.1

Yoo, C., & Son, S.-W. (2016). Modulation of the boreal wintertime Madden-Julian oscillation by the stratospheric quasi-biennial oscillation. *Geophysical Research Letters*, *43*(3), 1392-1398. doi: 10.1002/2016GL067762

Zhang, J., Howard, K., Langston, C., Kaney, B., Qi, Y., Tang, L., … Kitzmiller, D. (2016). Multi-radar multi-sensor (mrms) quantitative precipitation estimation: Initial operating capabilities. *Bulletin of the American Meteorological Society*, *97*(4), 621 - 638. doi: 10.1175/BAMS-D-14-00174.1