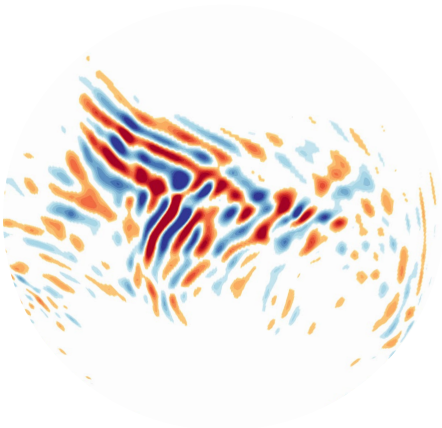
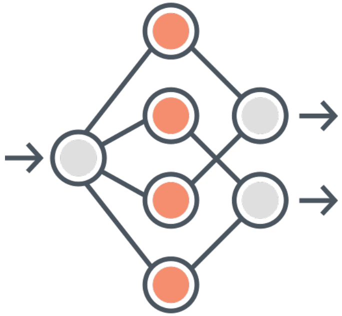
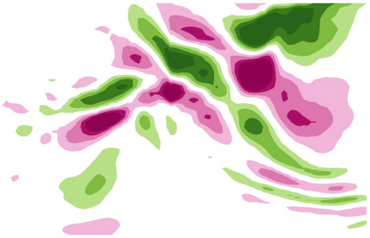
Nonlocal Deep Learning Parameterization for Process Representation in Climate Models

Aman Gupta, Aditi Sheshadri, Tom Meltzer, Sujit Roy, Valentine Anantharaj

30th CESM Workshop, Boulder, CO
11th *June 2025*



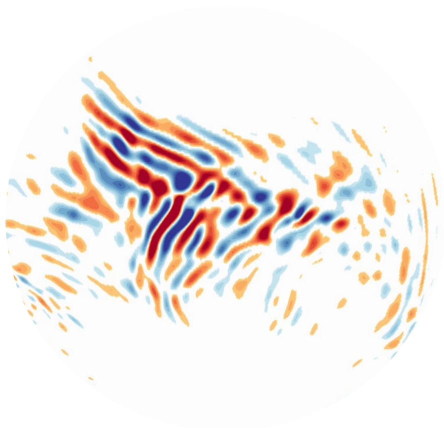
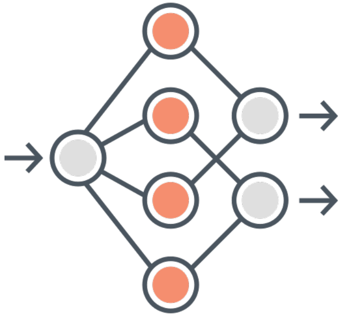
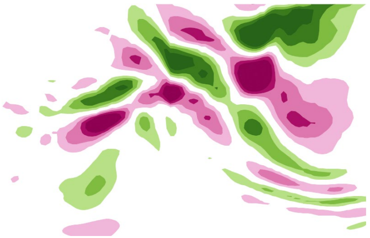




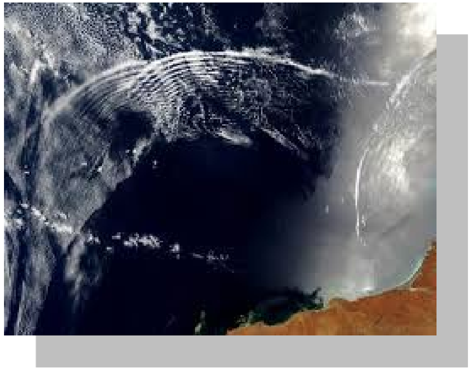
Nonlocal Deep Learning Parameterization for Atmospheric Gravity Waves Representation in Climate Models

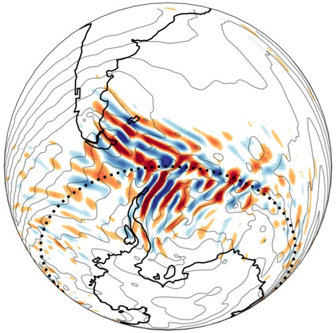
Aman Gupta, Aditi Sheshadri, Tom Meltzer, Sujit Roy, Valentine Anantharaj

30th CESM Workshop, Boulder, CO
11th *June 2025*

Atmospheric Gravity Waves (GWs)

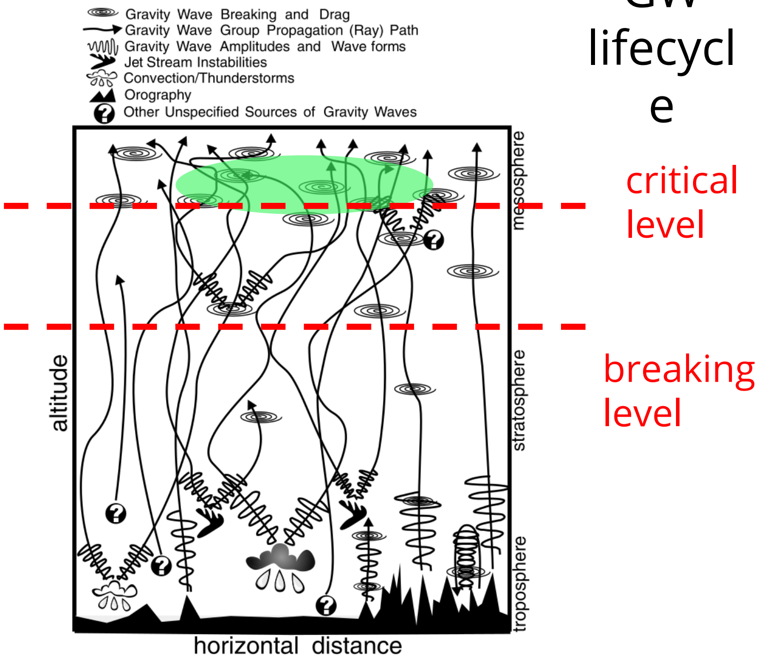




- Sources: jets, convection, mountains etc.
- Multiple scales: 100 m to 1000s km

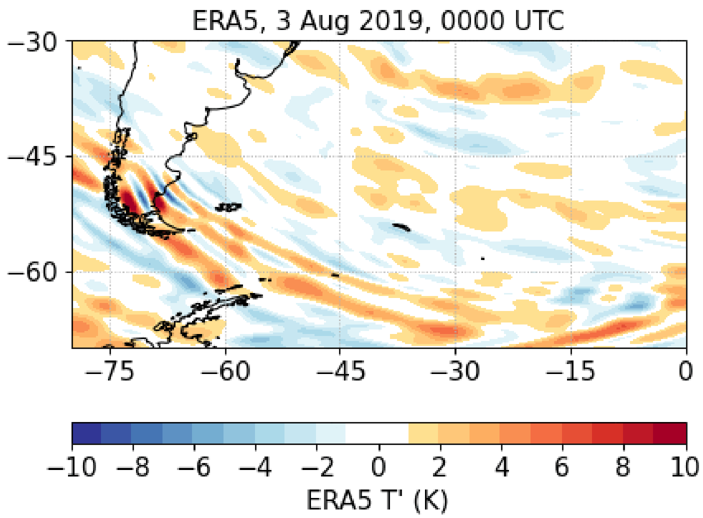
c_z ~ 0-15 m/s
c_H ~ 0-150 m/s

Typical GW lifecycle



- Vertical coupling: carry near surface momentum to upper atmosphere within hours. 10x faster propagation in the horizontal.

Current GW Parameterizations have Notable Biases

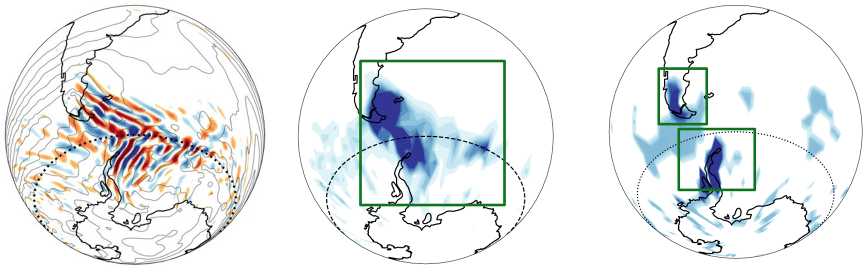


Key properties:

1) **Lateral propagation:** of wave fluxes away from source

2) **Refraction:** changes in wavenumber as they propagate

3) **Transience:** temporal coherence of wave packets



phase structure   resolved flux   parameterized flux

Biases in:

a) QBO representation

b) “cold-pole” bias in Austral summer stratosphere

c) Midlatitude jet strength and mesospheric overturning circulation

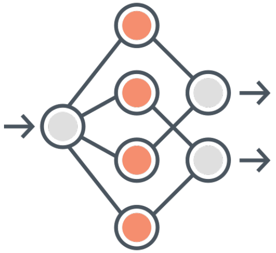
ML to Learn Subgrid-scale Gravity Wave Fluxes

Learn momentum fluxes from high-resolution, GW-resolving data

Couple the ML flux predictor to a coarse-resolution climate model

Background atmospheric conditions
(resolved by climate models)

$$\begin{bmatrix} u \\ v \\ \theta \\ \omega \end{bmatrix}$$

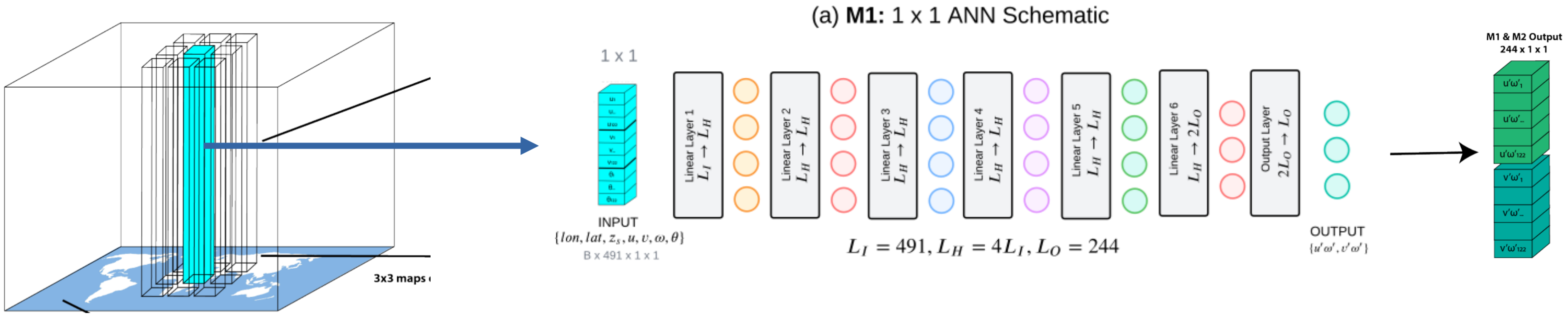


$$\begin{bmatrix} u'\omega' \\ v'\omega' \end{bmatrix}$$

Gravity wave **momentum fluxes** from high-resolution reanalysis/obs
(unresolved by climate models)

We train three ML models with varying degrees of nonlocality

M1: Single Column

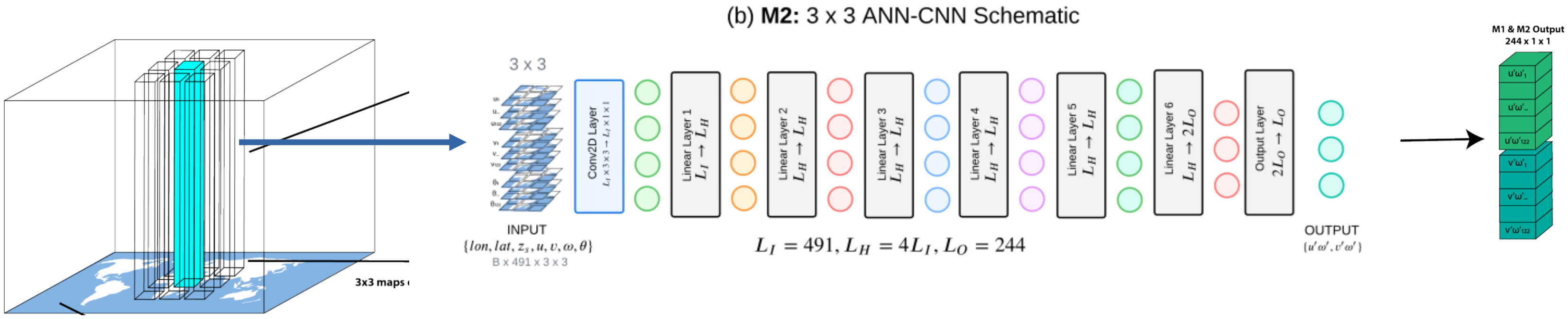


Model M1: inspired from traditional parameterizations

Dynamical variables in a column used to predict flux in the column

We train three ML models with varying degrees of nonlocality

M2: Multiple Columns

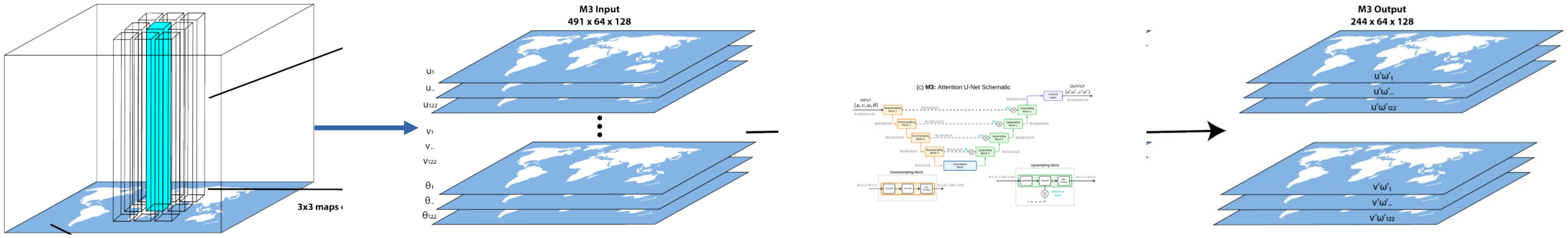


Model M2: Introducing slight nonlocality in space

Dynamical variables in 1 + 8 neighboring columns to predict fluxes in the central column

We train three ML models with varying degrees of nonlocality

M3: Global Attention U-Net



Model M3: Globally <u>nonlocal</u> Attention UNet (Oktay et al. 2018)
Global input of dynamical variables to predict fluxes globally.

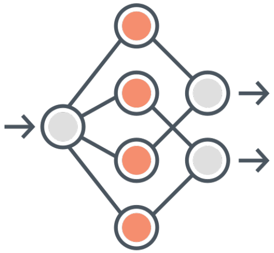
ML to Learn Subgrid-scale Gravity Wave Fluxes

Learn momentum fluxes from high-resolution, GW-resolving data

Couple the ML flux predictor to a coarse-resolution climate model

Background atmospheric conditions
(resolved by climate models)

$$\begin{bmatrix} u \\ v \\ \theta \\ \omega \end{bmatrix}$$



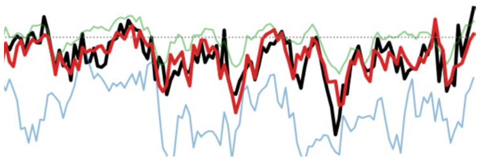
$$\begin{bmatrix} u'\omega' \\ v'\omega' \end{bmatrix}$$

Gravity wave **momentum fluxes** retrieved from high-resolution reanalysis
(unresolved by climate models)

Trained on 4 years of ERA5 and 4 months (NDJF) of 1.4 km ECMWF-IFS

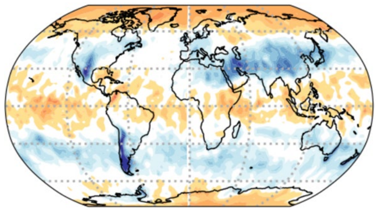
Evaluate performance beyond RMSE

Test 1. Temporal Evolution



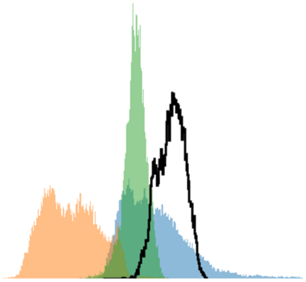
Does the model correctly learn the temporal wave evolution

Test 2. Seasonal Averages



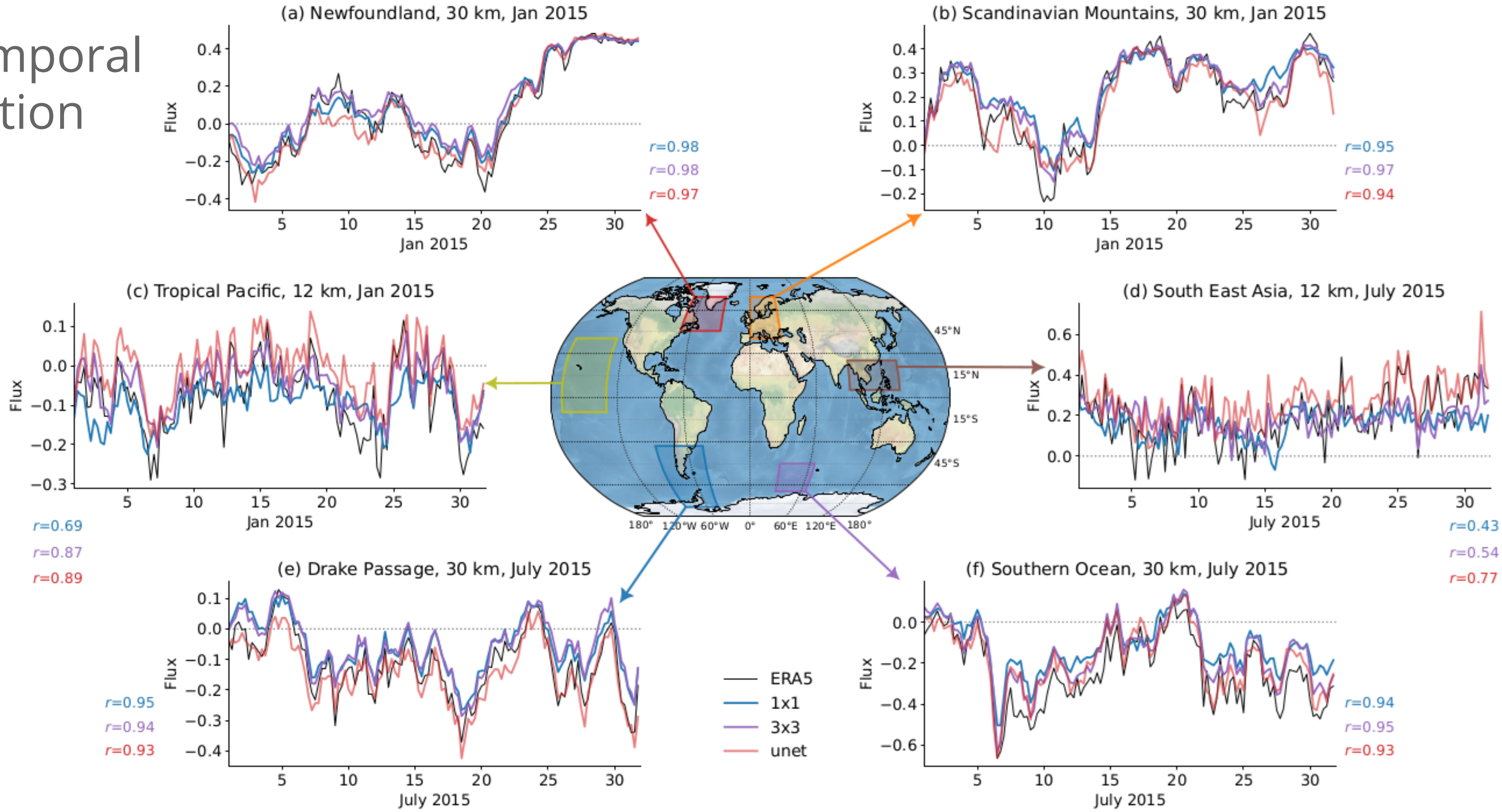
Does the model generate accurate global flux distribution?

Test 3. Flux distribution

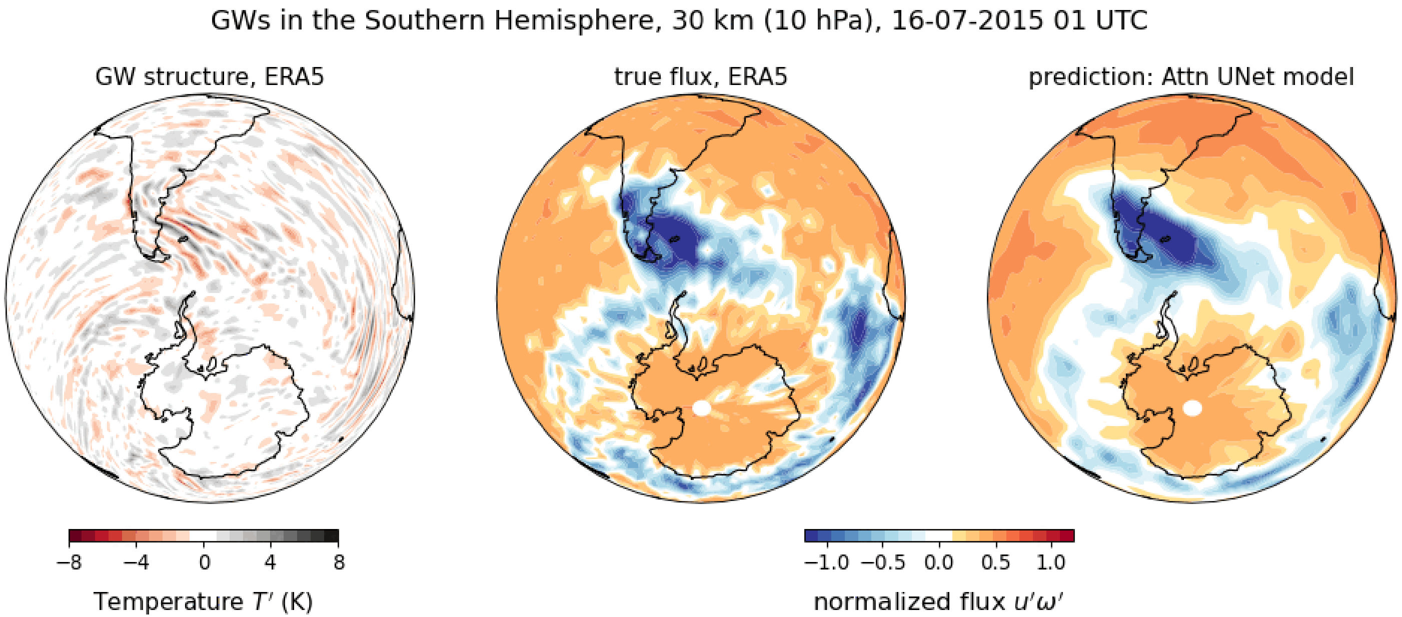


Does the model generate desired statistics?

1. Temporal Evolution



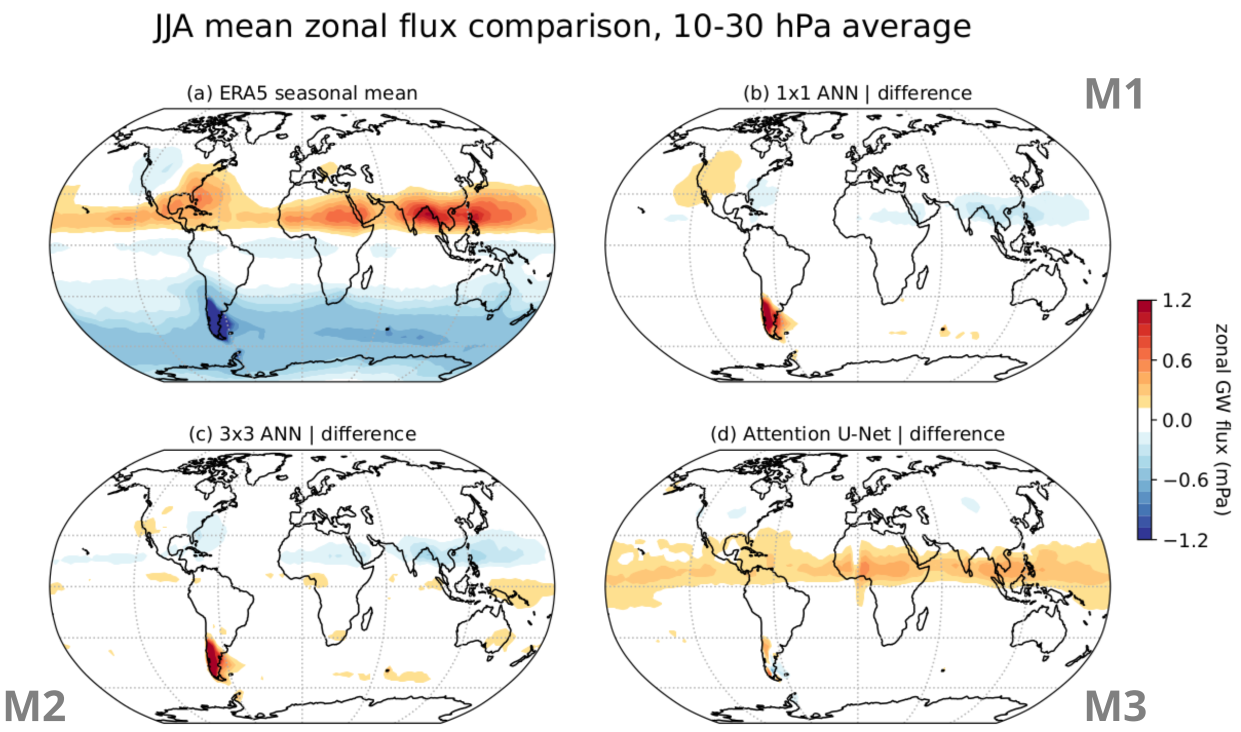
ML models skillfully learn the intermittent and coherent evolution of GW fluxes in the atmosphere over both orographic and nonorographic hotspots. **Nonlocal models perform better.**



Attention UNet correctly identifies wave excitation and lateral propagation over multiple hotspots over the Southern Ocean (Andes, small islands, storm tracks, Antarctic Peninsula, etc.)

Successful simulation of belts of midlatitude GW activity in both hemispheres without special provisions for recurrence.

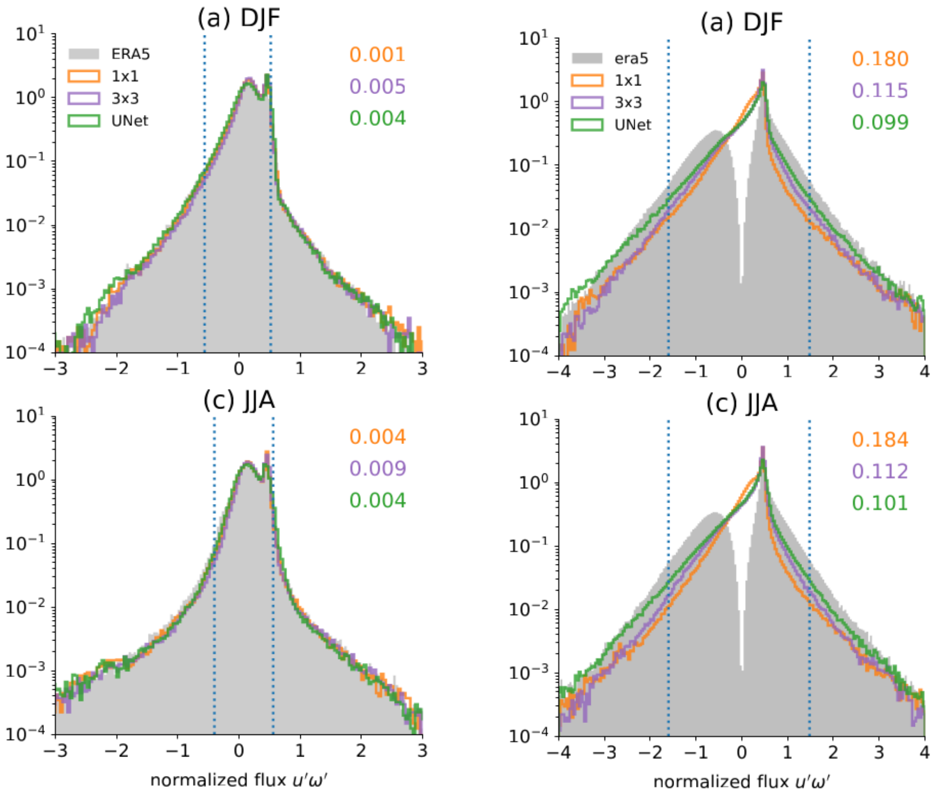
2. Seasonal Average



All of M1, M2, M3 generate reasonable predictions.

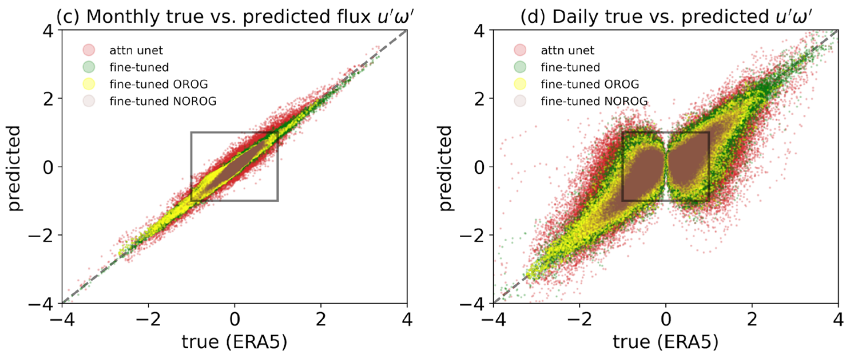
Attention UNets generate the most accurate predictions in the midlatitudes (where horizontal propagation is most prominent).

3. Global Flux Distribution



Seasonal averages

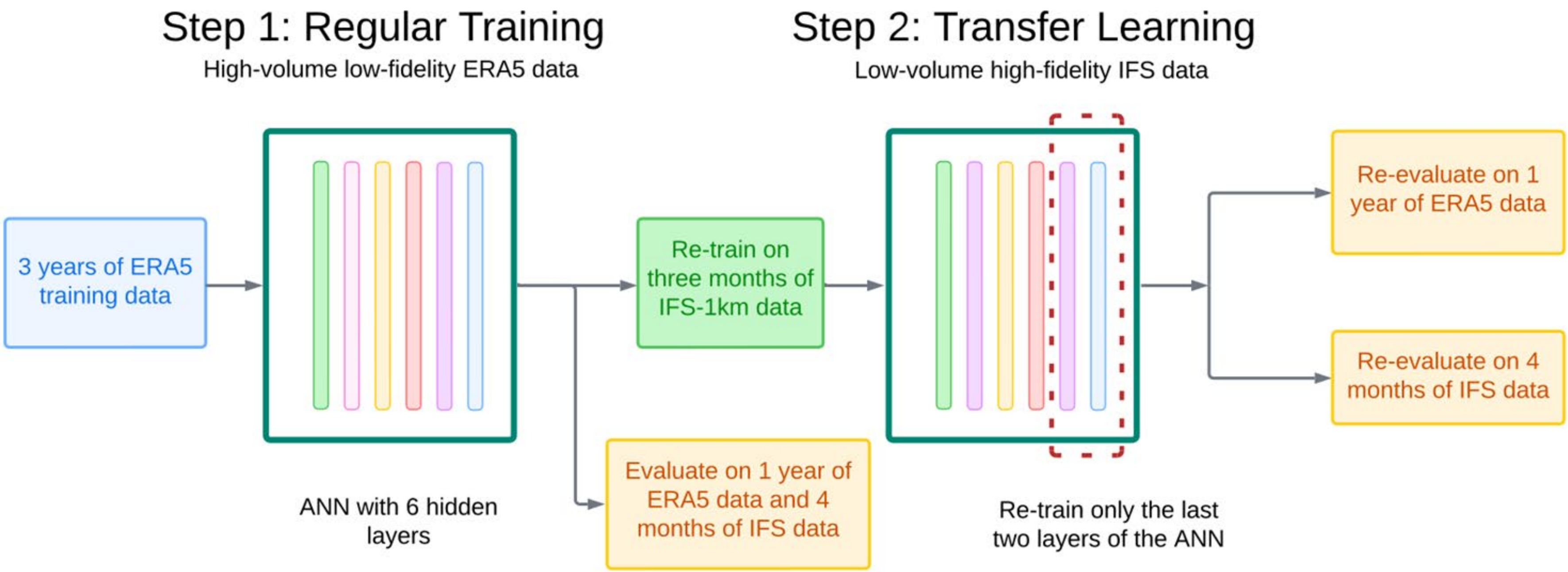
daily averages



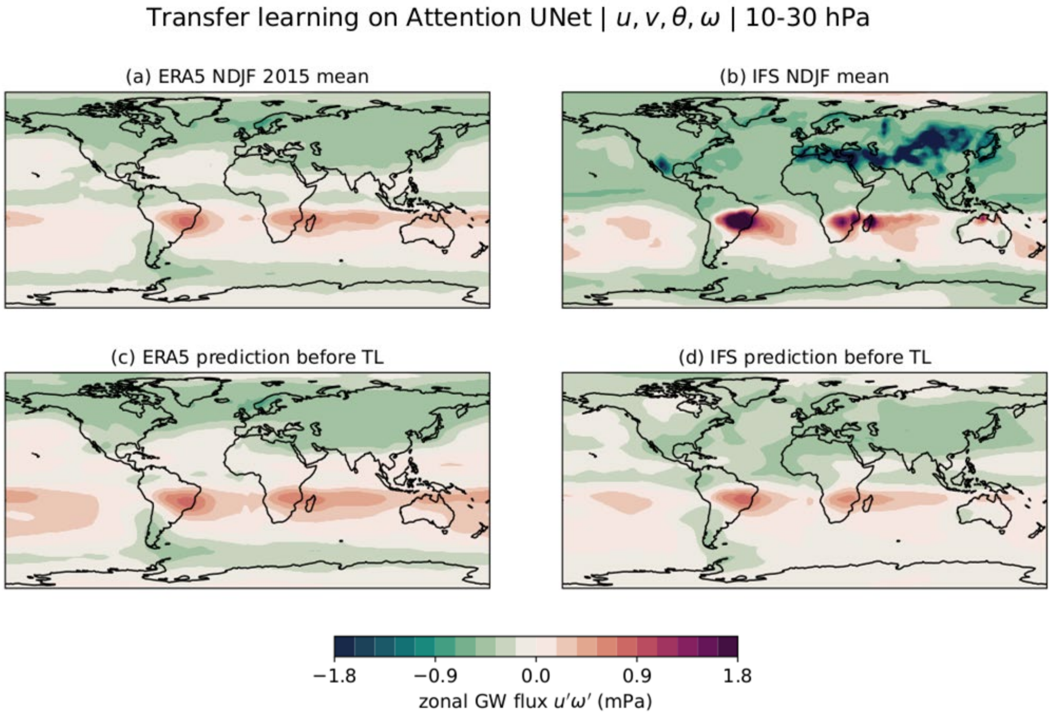
The seasonally averaged distributions are reproduced quite well.

… but the neural nets struggle with small values – predict zeros instead.

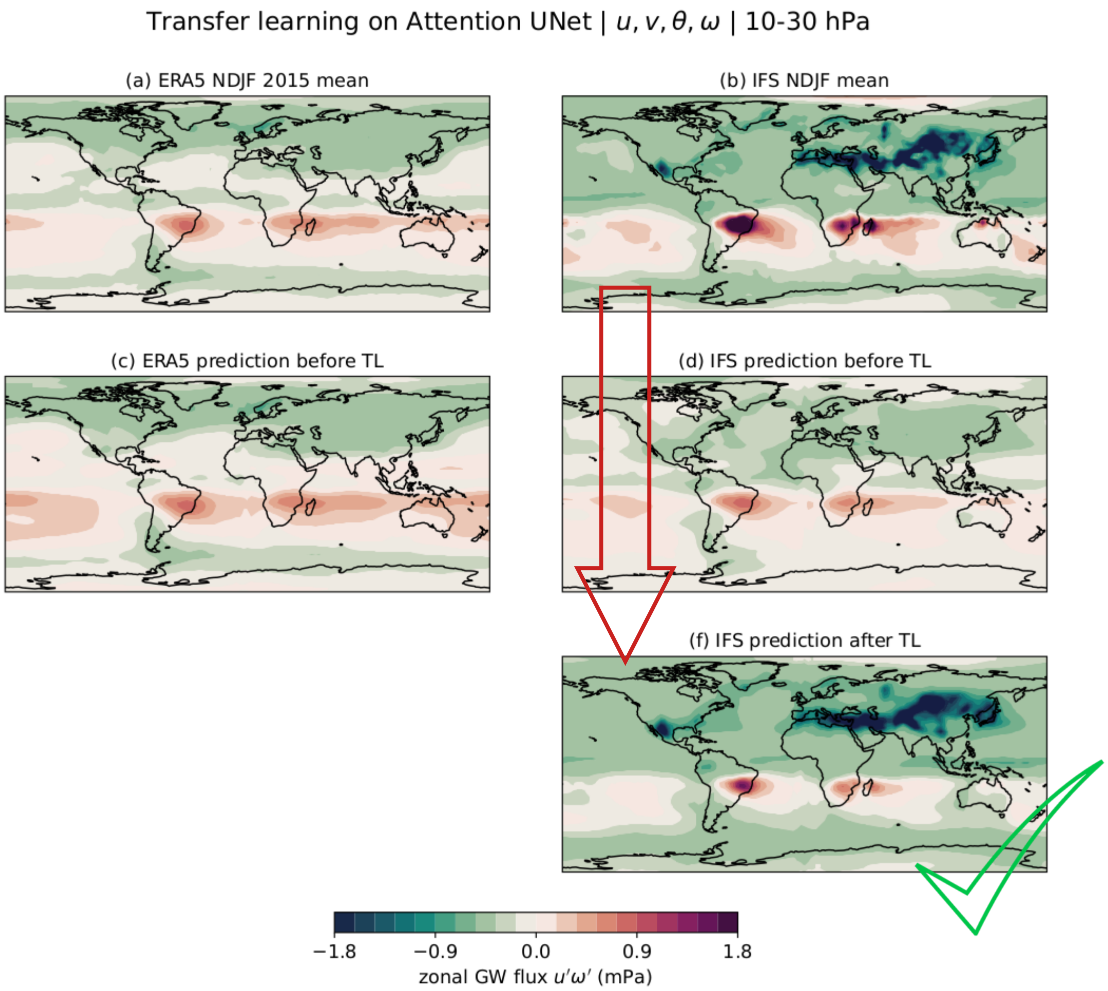
Improving predictions using transfer learning



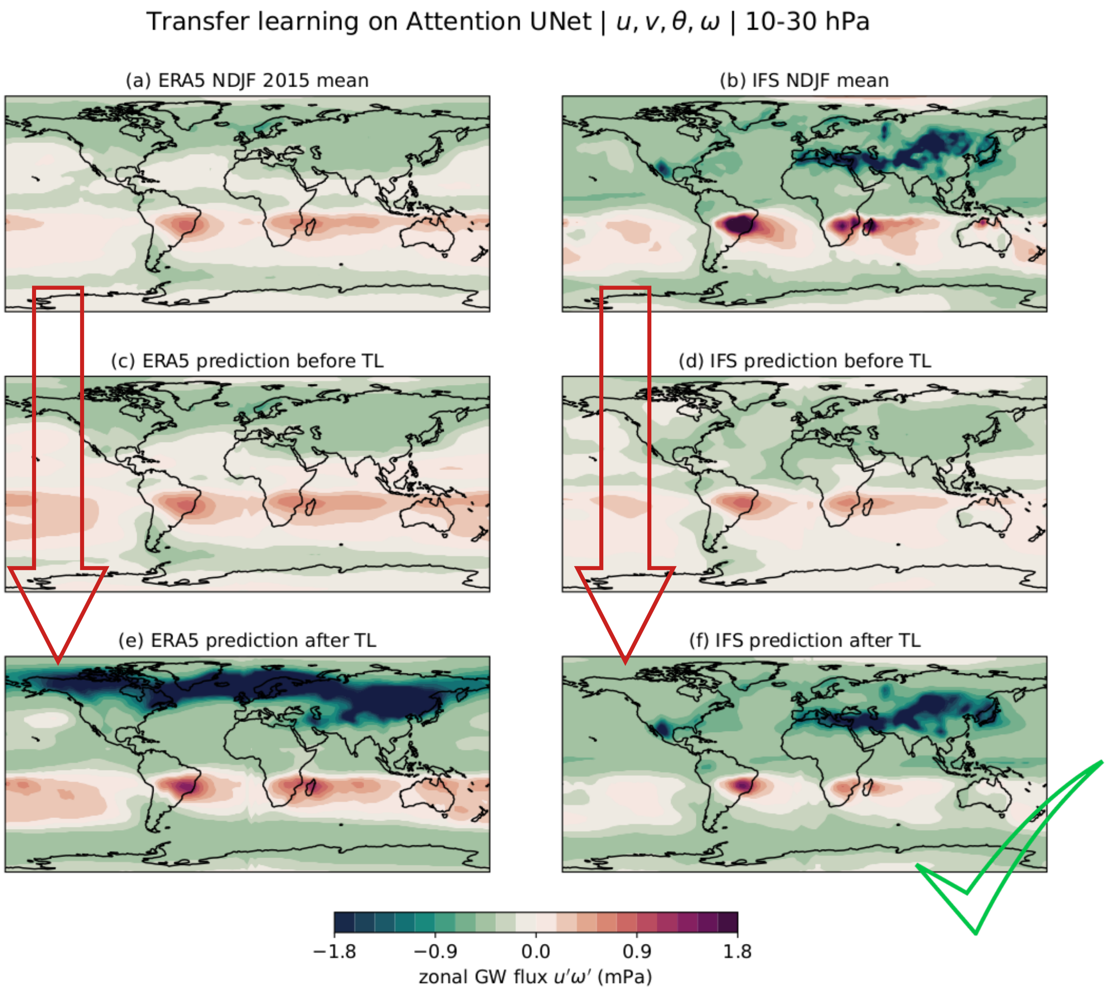
Next, the learning is augmented by using limited-but-high-resolution data which fully resolved GWs



ML models trained only on ERA5 underestimate fluxes in 1km-IFS



Following Transfer Learning, models predict stronger fluxes in ERA5, while identifying the correct hotspots.



Following Transfer Learning, models predict stronger fluxes in ERA5, while identifying the correct hotspots.

Thus, the models effectively blend learnings from both *low-fidelity high-volume* and *high-fidelity low-volume* datasets.

Potential to learn from multiple high-resolution climate datasets.

A Nonlocal Deep Learning Parameterization for Climate Model Representation of Atmospheric Gravity Waves: Offline Performance

Aman Gupta[1], Aditi Sheshadri[1], Sujit Roy[2,3], Valentine Anantharaj[4]

[1]Department of Earth System Science, Stanford University, Stanford, USA
[2]Earth System Science Center, The University of Alabama in Huntsville, Huntsville, AL, USA
[3]NASA Marshall Space Flight Center, Huntsville, AL, USA
[4]Oak Ridge National Laboratory, Oak Ridge, Tennessee, USA

Submitted to JAMES (preprint: authorea.com//1263806

Code:
github.com/DataWaveProject/nonlocal_gwfluxes
HiRes IFS data: *https://osf.io/gx32s/*

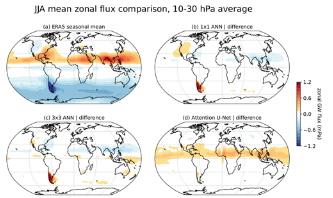
Key Conclusions

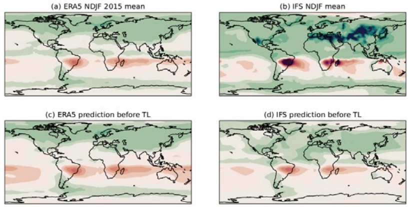
1. The three ML schemes learn nonlocal propagation, temporal coherence, and seasonal distributions of GW fluxes from high-resolution data.

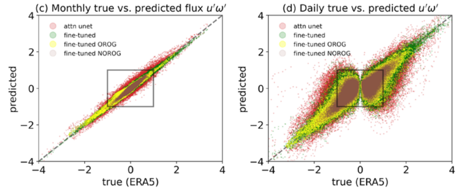
2. The model with the highest embedded nonlocality generates the best predictions.

3. Transfer learning allows blending multiple datasets to improve performance

4. Limitation: the schemes proficiently predict large-amplitude GW packets, but predicting small values is still a challenge





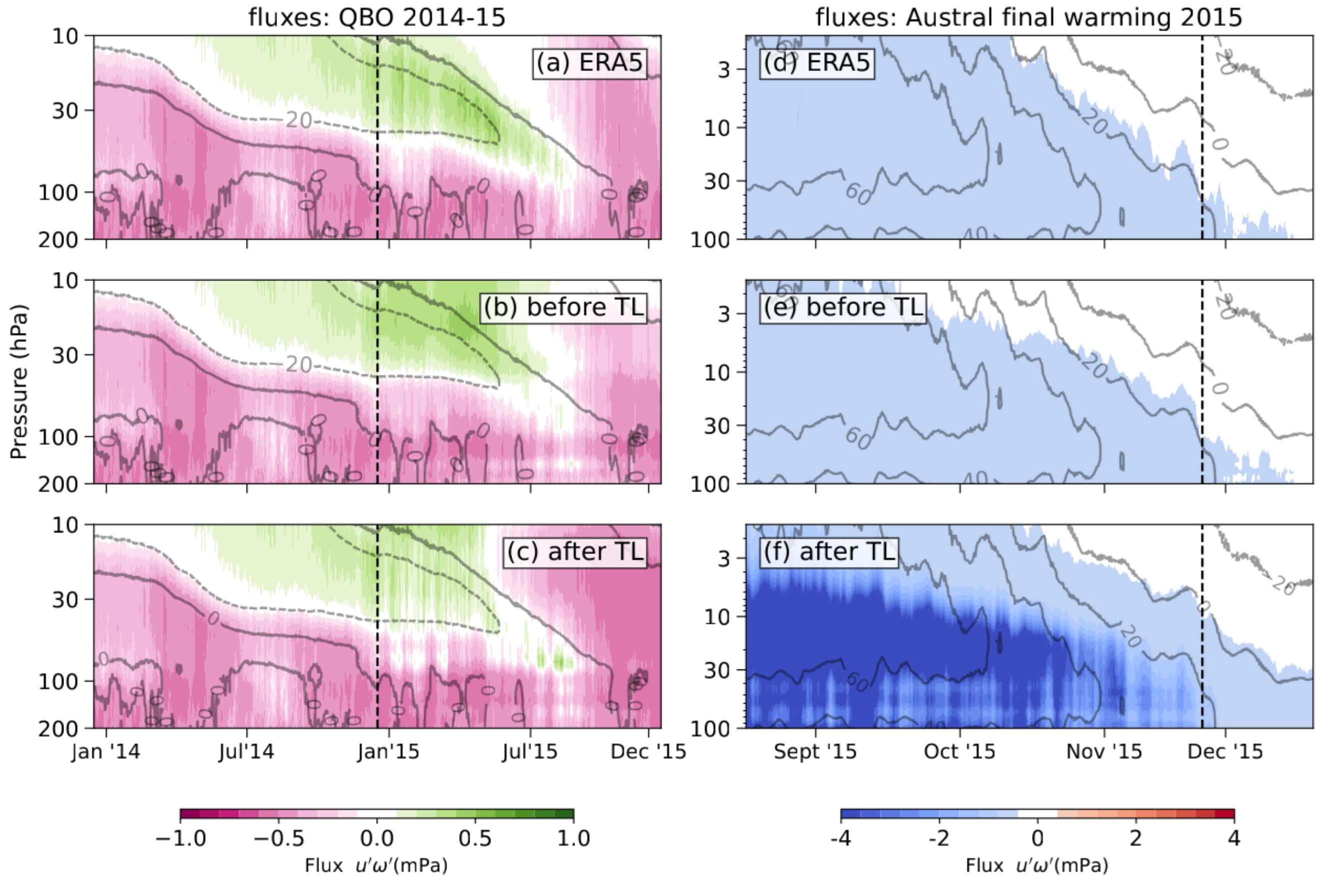


In Progress

1. Testing performance on dissimilar model outputs: high-resolution CAM and ICON runs.

2. Coupling to test online performance in CAM7: challenges

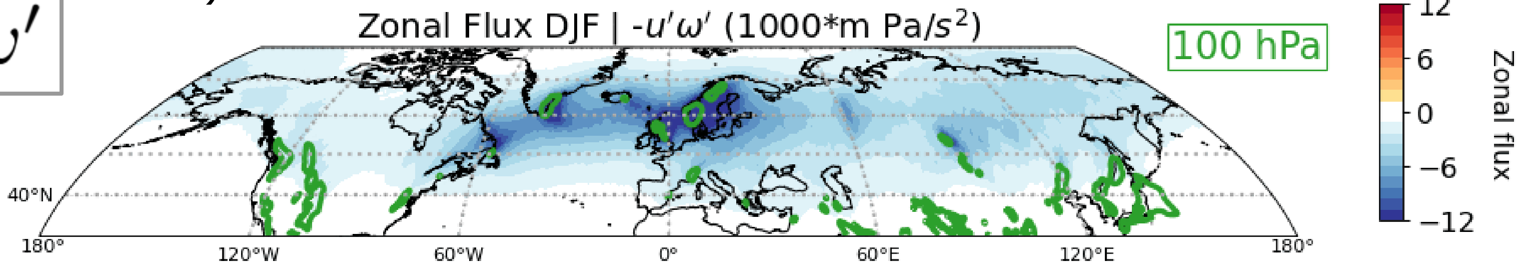
Supplement


fluxes: QBO 2014-15
fluxes: Austral final warming 2015
(a) ERA5
(b) before TL
(c) after TL
(d) ERA5
(e) before TL
(f) after TL
Pressure (hPa)
Jan '14
Jul'14
Jan'15
Jul'15
Dec '15
Sept '15
Oct '15
Nov '15
Dec '15
Flux $u'\omega'$(mPa)
Flux $u'\omega'$(mPa)

GWs form a belt of wave activity in the middle atmosphere

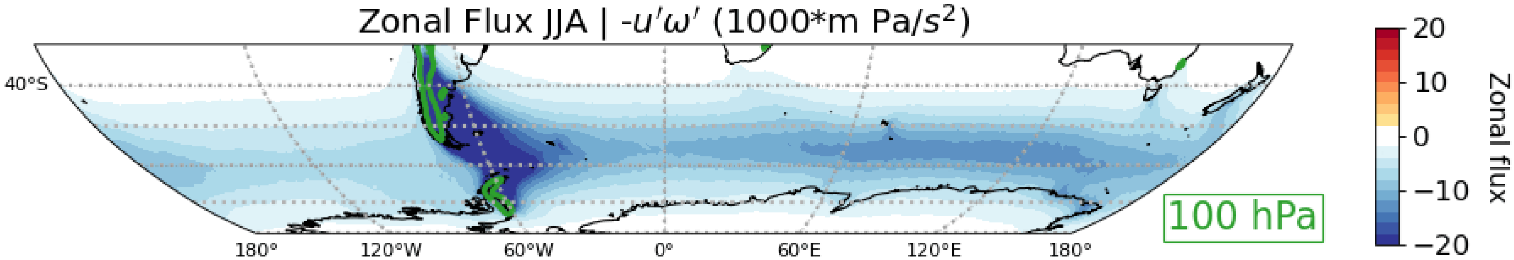
Green: Flux envelope, Color: Flux at 2 hPa (~45 km)

$$-u'\omega'$$



1979-2022 December January February mean From reanalysis

local GW generation ⇨ propagation through strong shear ⇨ global spreading



1979-2022 June July August mean from reanalysis

Gupta, Sheshadri, Alexander, Birner (2024), GRL | Insights on Lateral Gravity Wave Propagation in the Extratropical Stratosphere from 44 Years of ERA5 Data

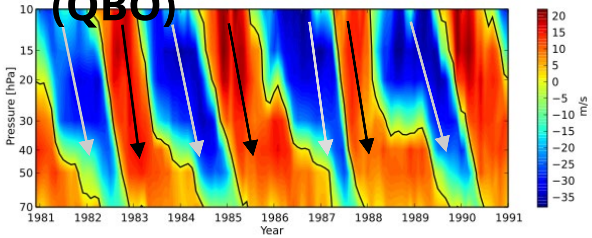
Critical Impacts of Gravity Waves

Atmospheric GWs induce clear air turbulence (CAT)
and influence upper tropospheric predictability.

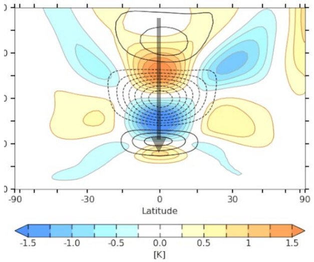
Severe Convectively Induced Turbulence Hitting a Passenger ...
by S Gisinger · 2024 · Cited by 2 — The **Singapore Airlines flight** SQ321 was on its way from London to Singapore when severe **turbulence** was encountered over Myanmar on 21 May 2024.

Tropical **Quasi-Bienniel Oscillation (QBO)**

Key drivers of global circulation and periodic
wind patterns, in the middle atmosphere.
Indirectly influencing Antarctic summer heat extremes via polar vortex variability (Choi et al., 2024).



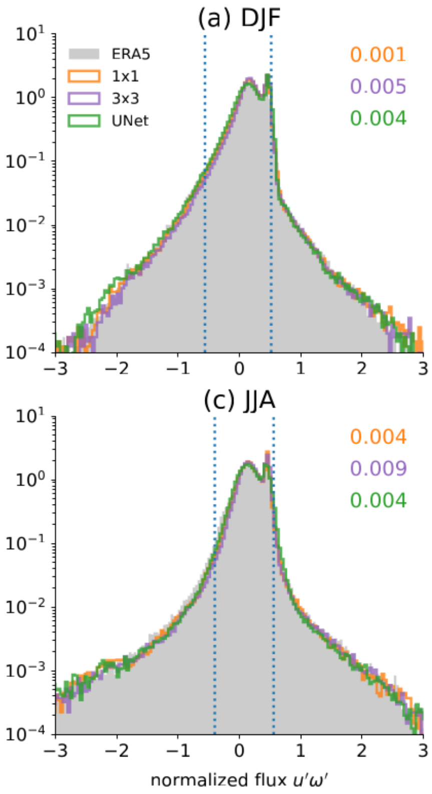
~ 28 month
period



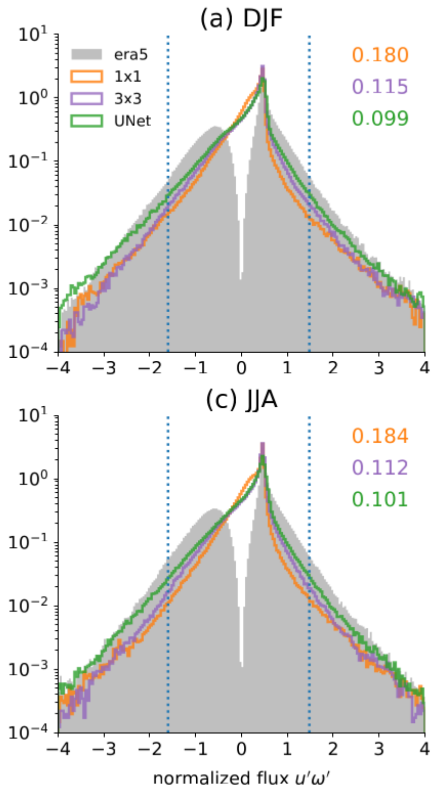
QBO animation credits: Hamid Pahlavan

3. Global Flux Distribution

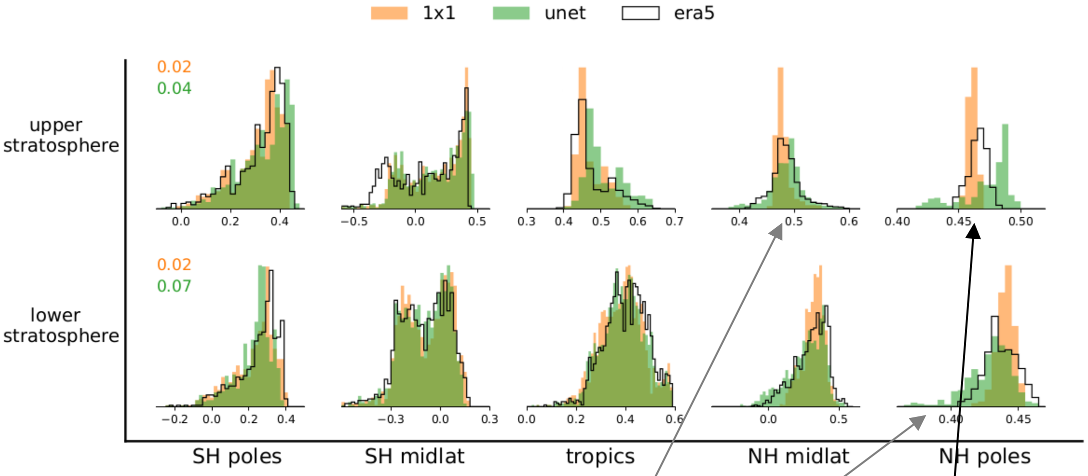
$$\mathcal{H}(p,q) = \frac{1}{2}\int_{x\in X}\left(\sqrt{p(x)} - \sqrt{q(x)}\right)^2 dx = 1 - \int_{x\in X}\sqrt{p(x)q(x)}dx.$$



Seasonal averages



daily averages



The three models generate comparable distribution tails for all seasons

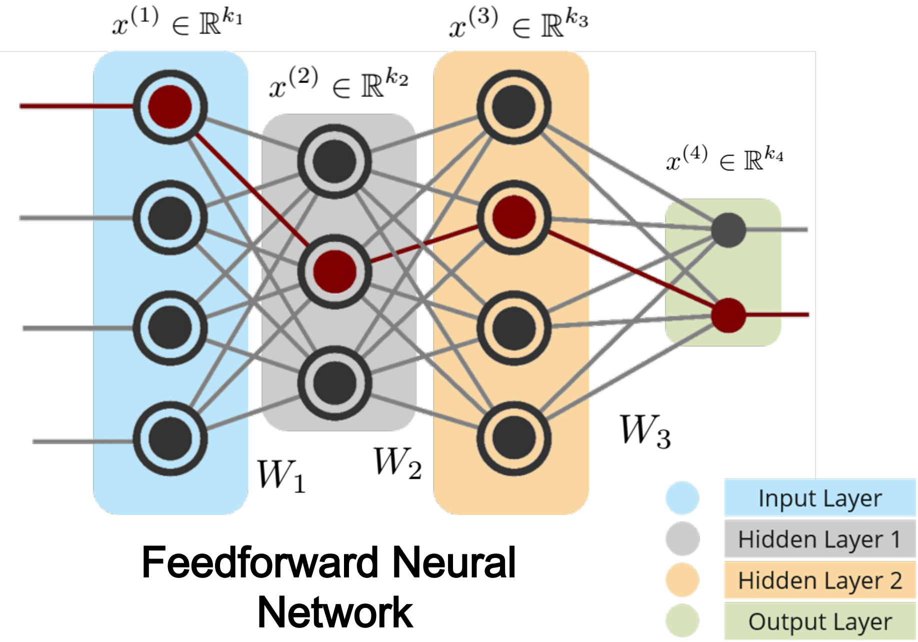
Prominent narrow bias in flux predictions by ANNs

Areas of weak GW activity (in summer stratosphere) most challenging to simulate.

Neural Network as a Collection of Perceptrons

Brain is a network of interconnected neurons. For any input/actions, only selected neurons fire at a given time. A **multi-layer perceptron (MLP)** is a collection of neurons with equisized, fully-connected hidden layers. Similarly, a size-varying MLP without loops is called a **feedforward neural network**.

Consider a feedforward neural network arranged as an input layer, 2 hidden layers, and an output layer:



Feedforward Neural Network

Forward Propagation

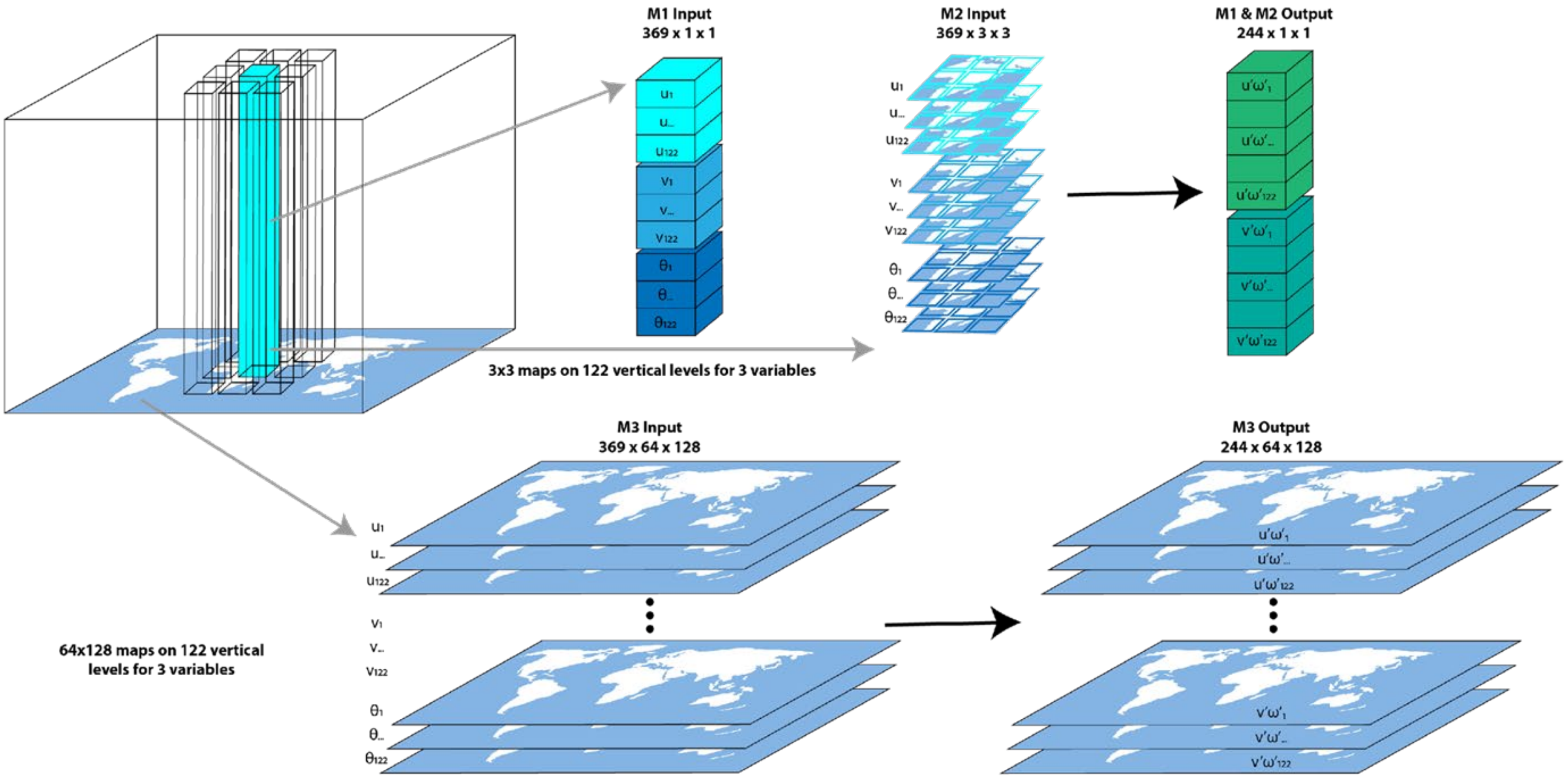
(1) Each layer maps to the next using a set of weights

(2) The linear transformation is followed by a non-linear activation σ(.)

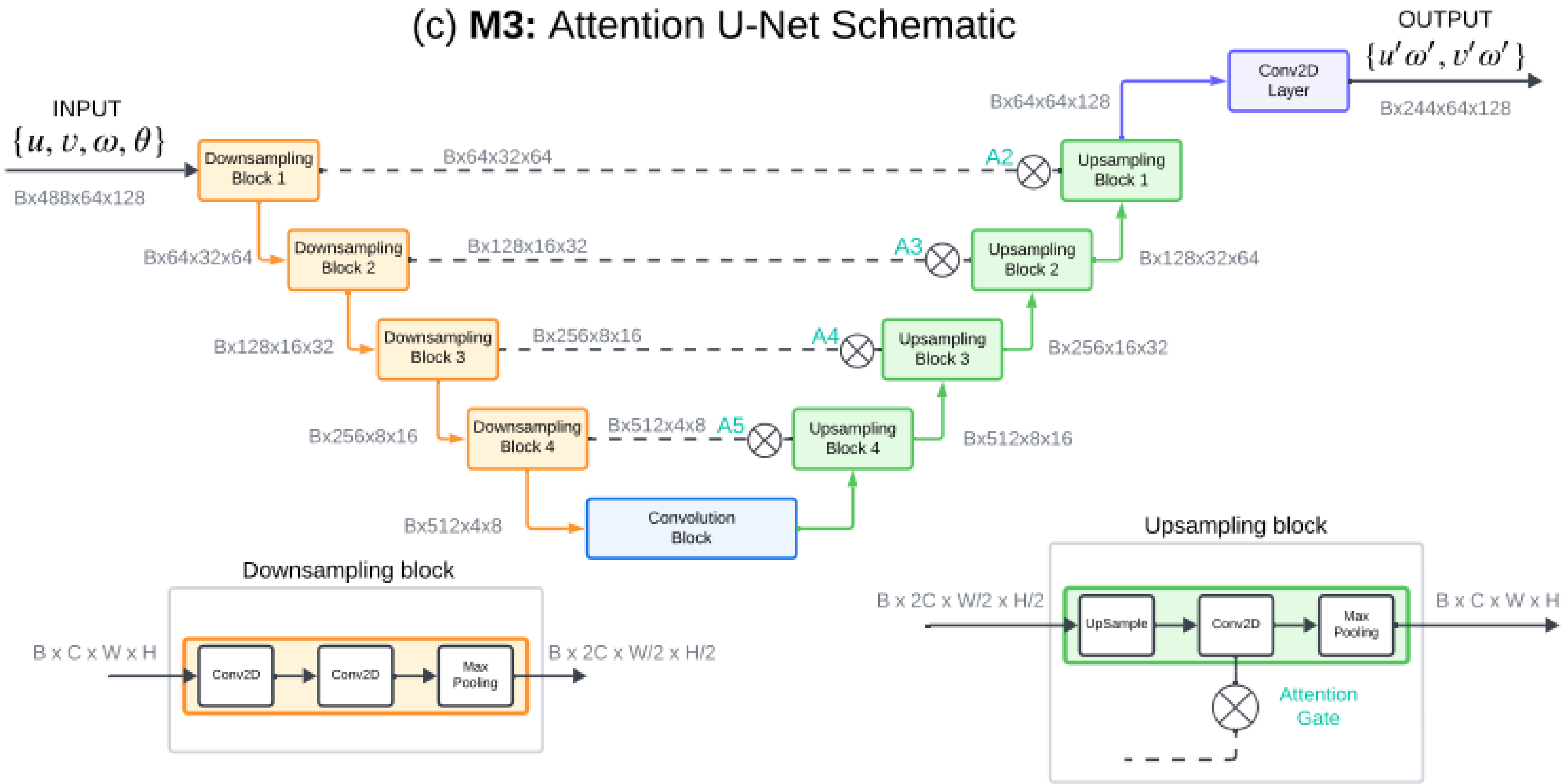
$$\boxed{x^{(i+1)} = \sigma\left(W_i^T x^{(i)}\right)}$$

$$W_i \in \mathbb{R}^{k_i \times k_{i+1}}, \sigma_i : \mathbb{R}^{k_{i+1}} \to \mathbb{R}^{k_{i+1}}$$

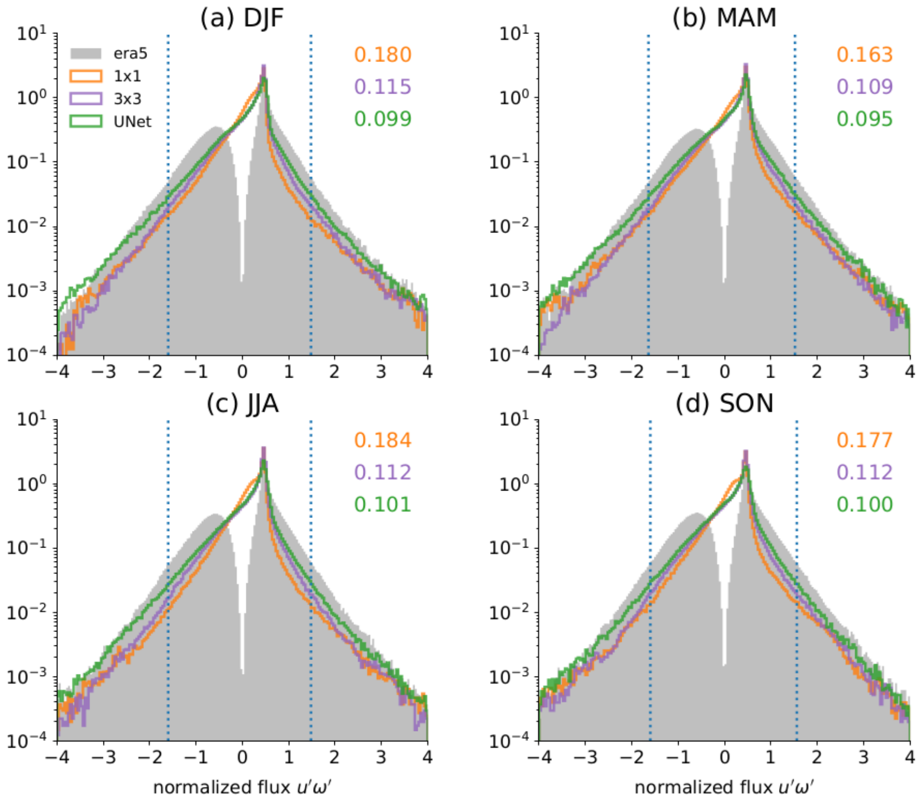
Learning nonlocality through nonlocal architectures



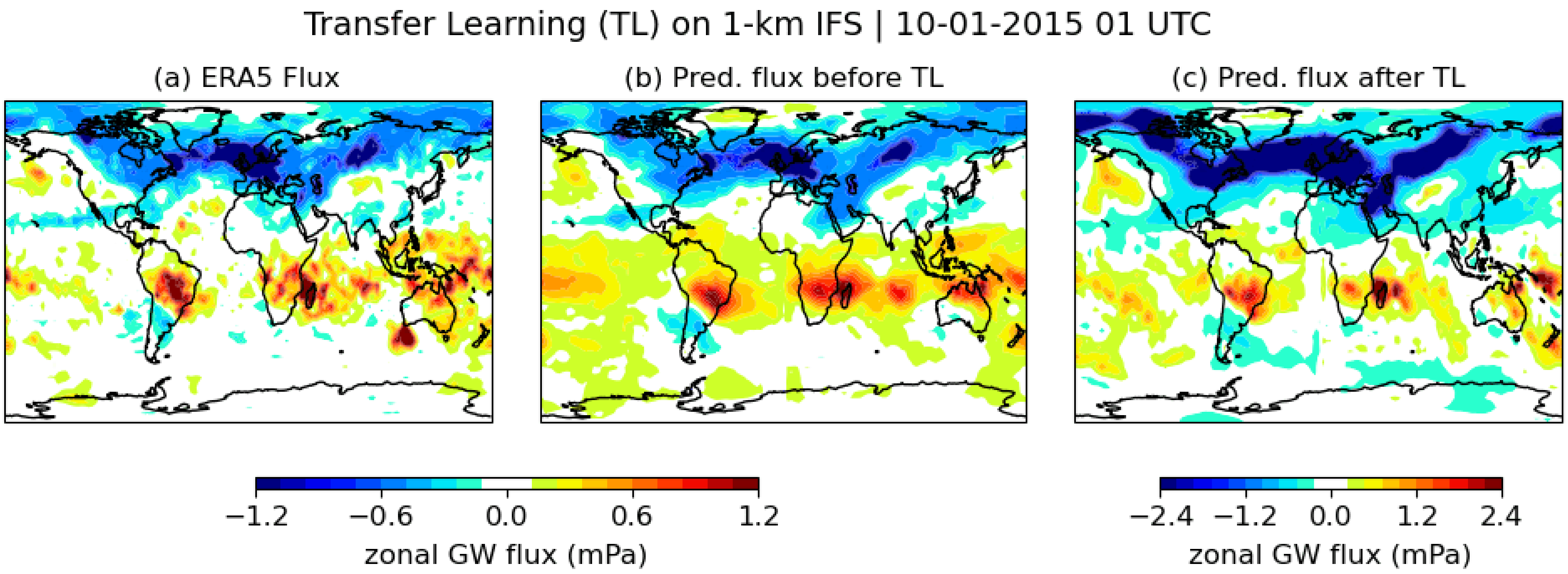
Attention UNet Schematic

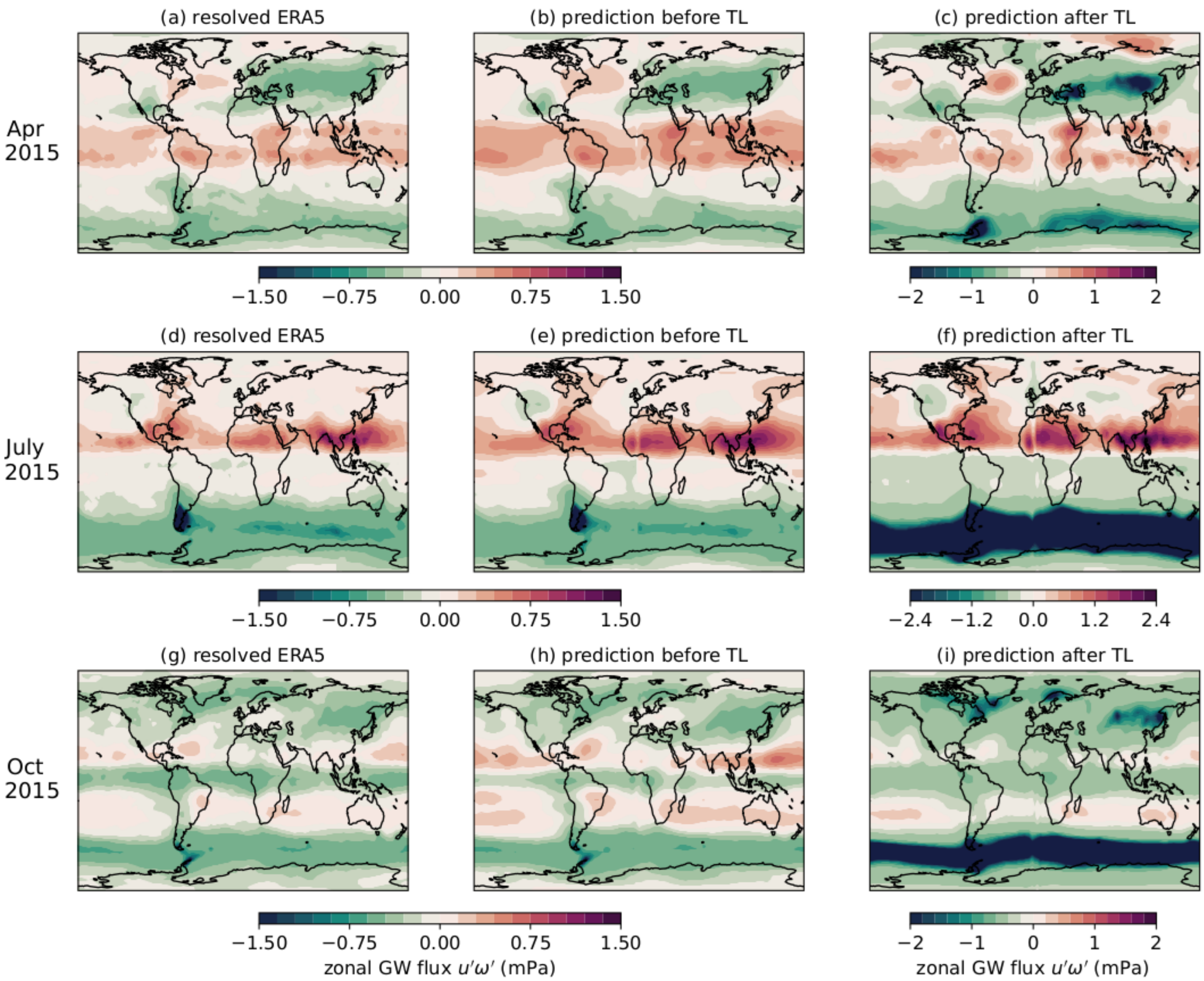


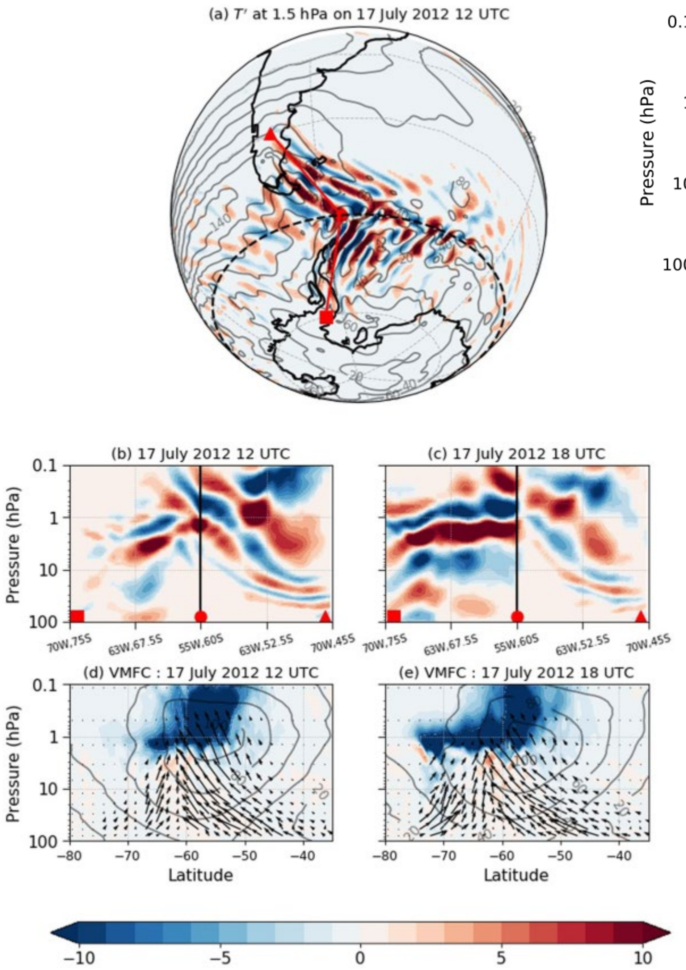
Daily Sampled Flux Distributions

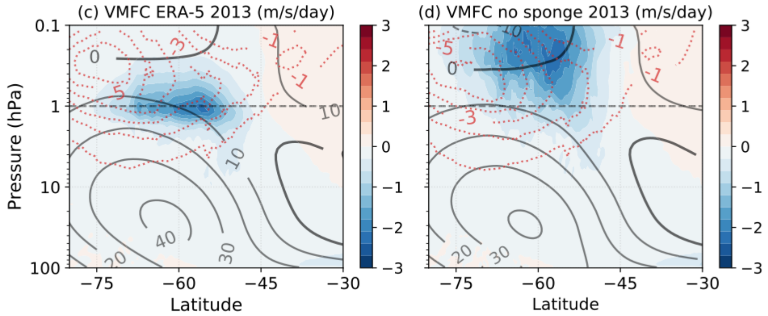


Transfer Learning on out-of-set months

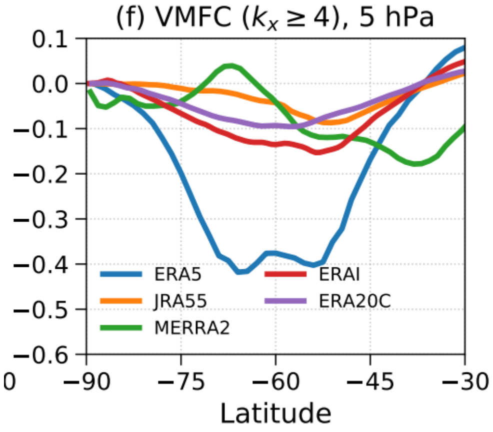


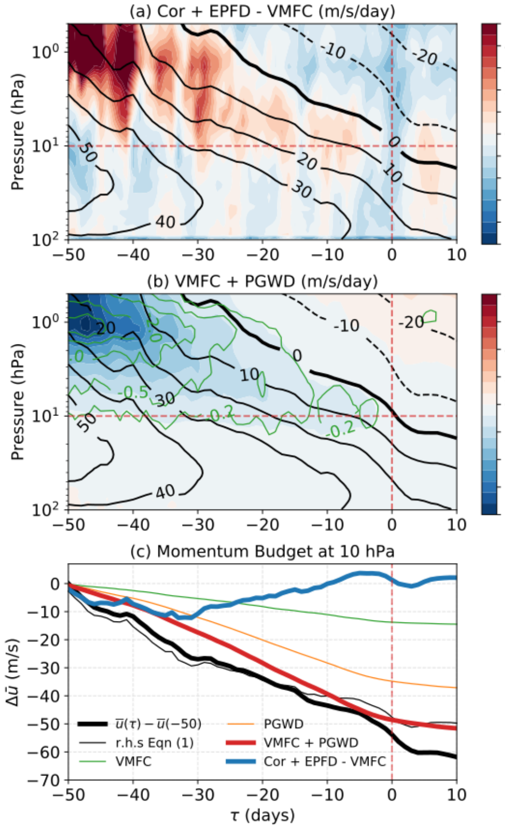

(a) resolved ERA5
(b) prediction before TL
(c) prediction after TL
Apr 2015
−1.50
−0.75
0.00
0.75
1.50
−2
−1
0
1
2
(d) resolved ERA5
(e) prediction before TL
(f) prediction after TL
July 2015
−1.50
−0.75
0.00
0.75
1.50
−2.4
−1.2
0.0
1.2
2.4
(g) resolved ERA5
(h) prediction before TL
(i) prediction after TL
Oct 2015
−1.50
−0.75
0.00
0.75
1.50
zonal GW flux $u'\omega'$ (mPa)
−2
−1
0
1
2
zonal GW flux $u'\omega'$ (mPa)





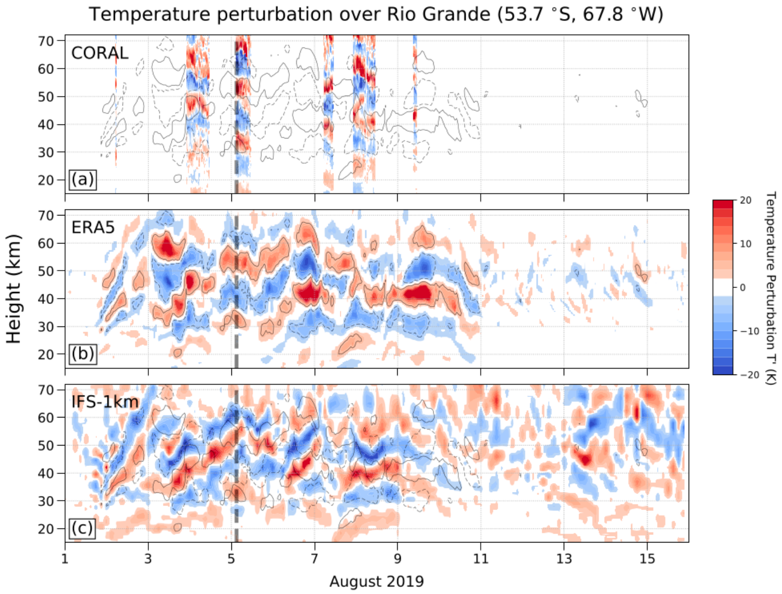
$$\overline{u}_t = \underbrace{\left(f - \frac{1}{R\cos\phi}(\overline{u}\cos\phi)_\phi\right)\overline{v}^*}_{\text{Cor}} - \underbrace{\overline{u}_p\overline{\omega}^*}_{\text{vAdv}} + \underbrace{\frac{1}{R\cos\phi}\vec{\nabla}\cdot\vec{F}}_{\text{EPFD}} + \underbrace{\overline{X}}_{\text{PGWD}}$$

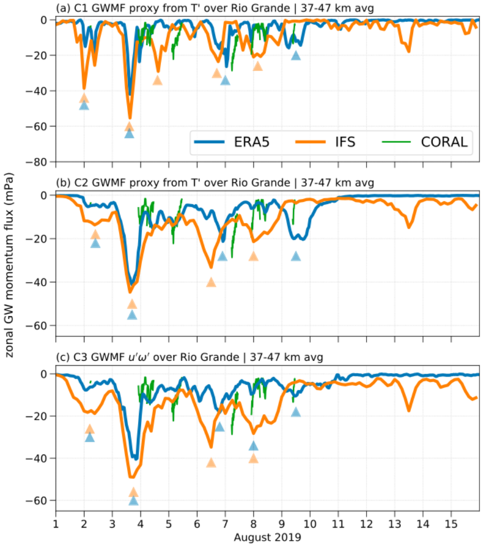


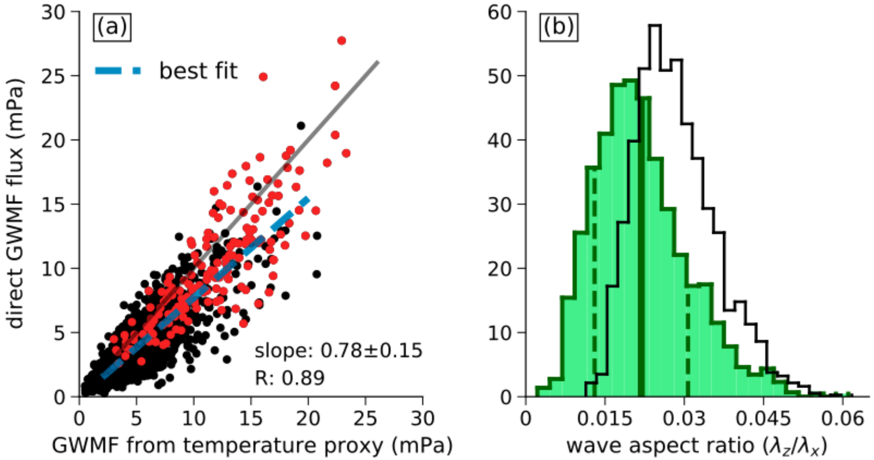


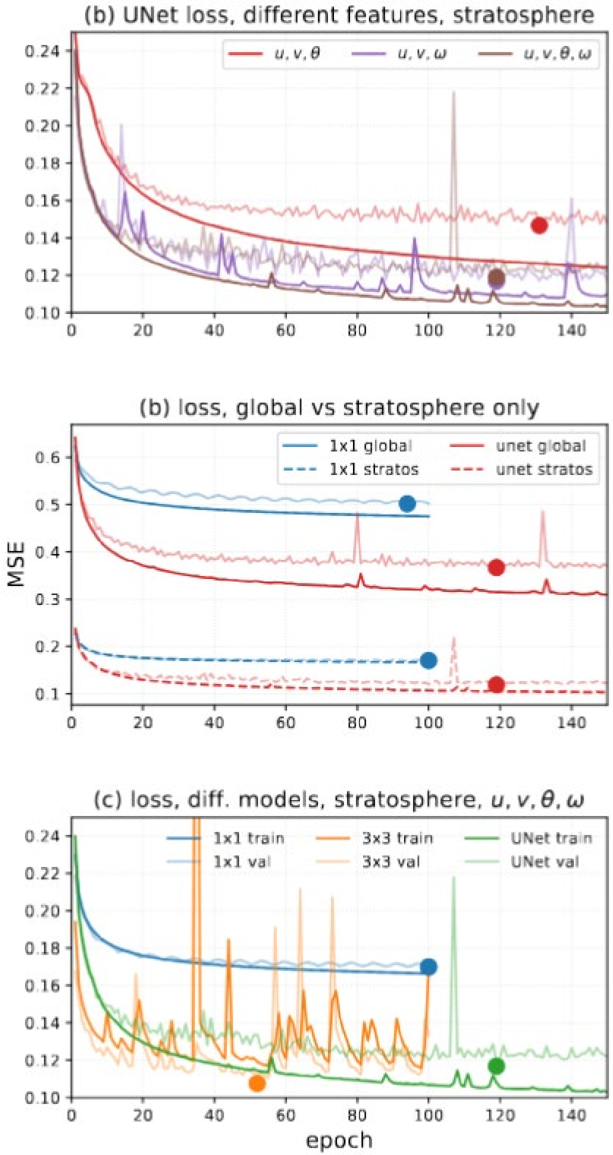
$$E_p = \frac{1}{2}\frac{g^2}{N^2}\overline{\left(\frac{T'}{T_0}\right)^2},$$

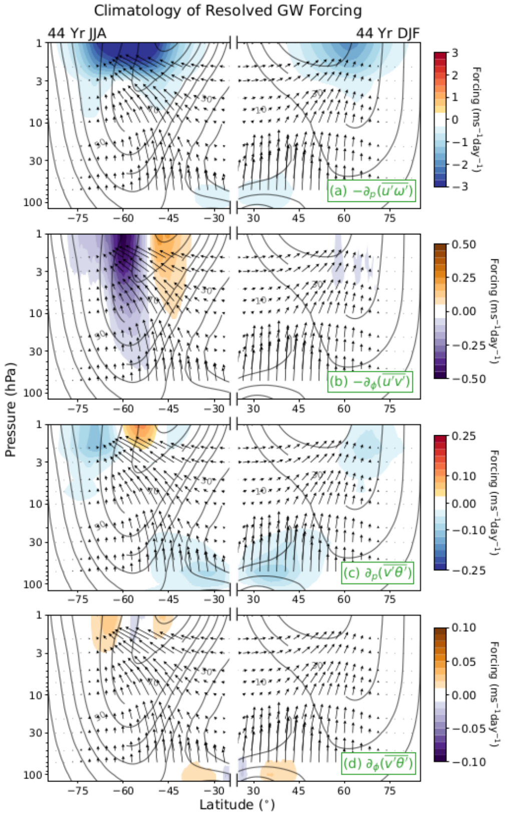
$$\mathbf{F} = g^{-1}(\overline{u'\omega'}, \overline{v'\omega'}).$$

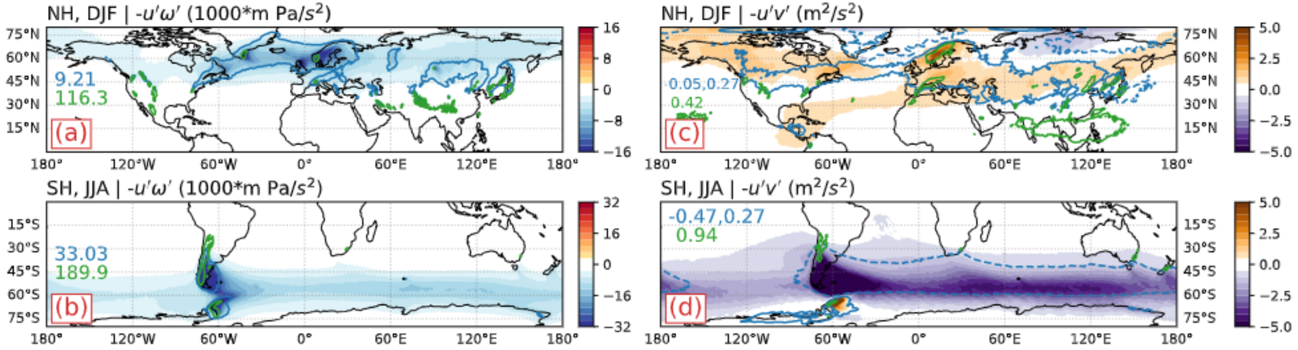


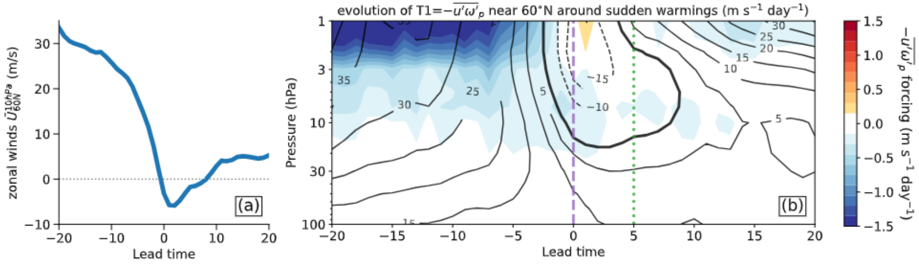





(b) UNet loss, different features, stratosphere
u, v, θ
u, v, ω
u, v, θ, ω
(b) loss, global vs stratosphere only
1x1 global
1x1 stratos
unet global
unet stratos
MSE
(c) loss, diff. models, stratosphere, u, v, θ, ω
1x1 train
1x1 val
3x3 train
3x3 val
UNet train
UNet val
epoch


Climatology of Resolved GW Forcing
44 Yr JJA
44 Yr DJF
Pressure (hPa)
Latitude (°)
Forcing (ms^{-1}day^{-1})
(a) $-\partial_p(\overline{u'\omega'})$
(b) $-\partial_\phi(\overline{u'v'})$
(c) $\partial_p(\overline{v'\theta'})$
(d) $\partial_\phi(\overline{v'\theta'})$


NH, DJF | -$u'\omega'$ (1000*m Pa/s^2)
NH, DJF | -$u'v'$ (m^2/s^2)
SH, JJA | -$u'\omega'$ (1000*m Pa/s^2)
SH, JJA | -$u'v'$ (m^2/s^2)
9.21
116.3
-0.05,0.27
0.42
33.03
189.9
-0.47,0.27
0.94
(a)
(b)
(c)
(d)


evolution of T1=$-\overline{u'\omega'}_p$ near 60°N around sudden warmings (m s^{-1} day^{-1})
zonal winds $\bar{U}^{10hPa}_{60N}$ (m/s)
Lead time
Pressure (hPa)
$-\overline{u'\omega'}_p$ forcing (m s^{-1} day^{-1})
(a)
(b)