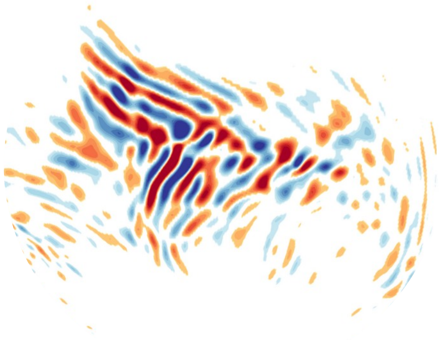
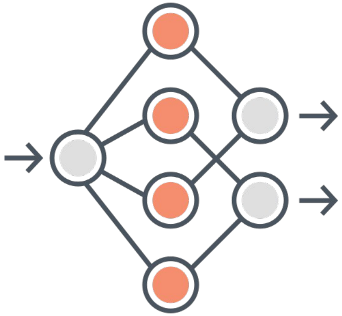
Nonlocal Deep Learning Parameterization for Climate Model Representation of Atmospheric Gravity Waves

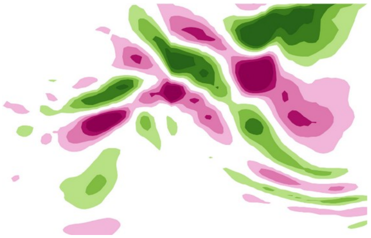
Aman Gupta, Aditi Sheshadri, Tom Meltzer, Sujit Roy, Valentine Anantharaj

Busan IAMAS-IACS-IAPSO Joint Assembly 2025

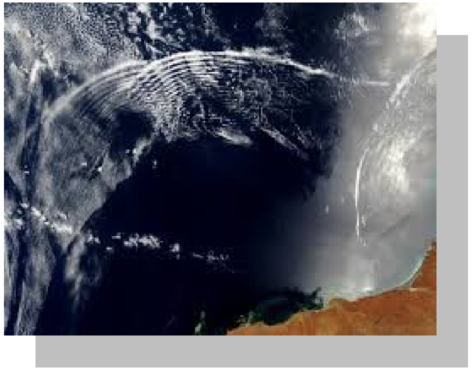
21th *July 2025*



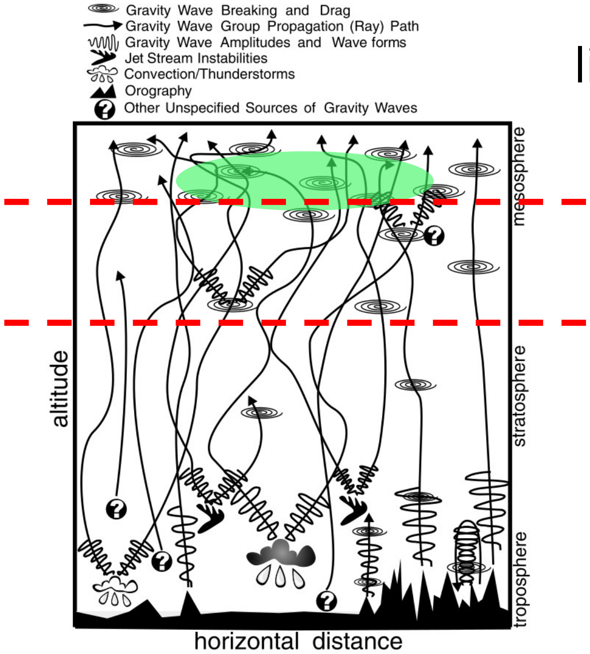




Atmospheric Gravity Waves (GWs)



Typical GW lifecycle



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

c_z ~ 0-15 m/s

c_H ~ 0-150 m/s

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

Critical Importance of Atmospheric Gravity Waves

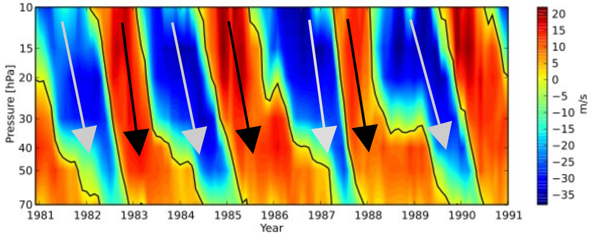
Atmospheric GWs induce clear air turbulence (CAT) and influence upper tropospheric predictability.

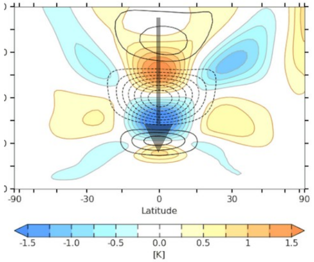
Severe Convectively Induced Turbulence Hitting a Passenger ...

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.

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).

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

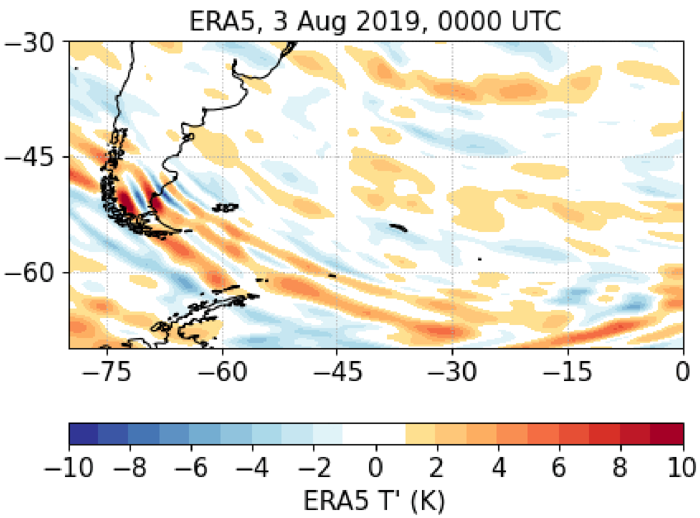




~ 28 month period

QBO animation credits: Hamid Pahlavan

Current GW Parameterizations have Notable Biases

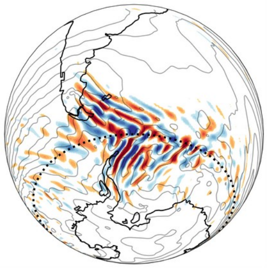


Key observed properties:

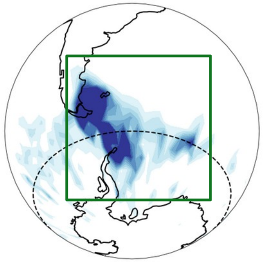
1) **Lateral propagation:** of wave fluxes away from source

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

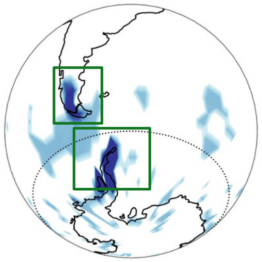
3) **Transience:** temporal coherence of wave packets



Resolved waves



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

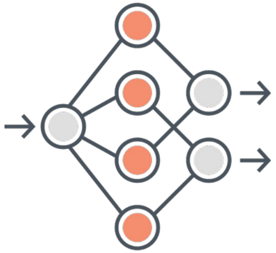
Objective: Use ML to Predict Subgrid-scale Gravity Wave Fluxes

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)

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



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

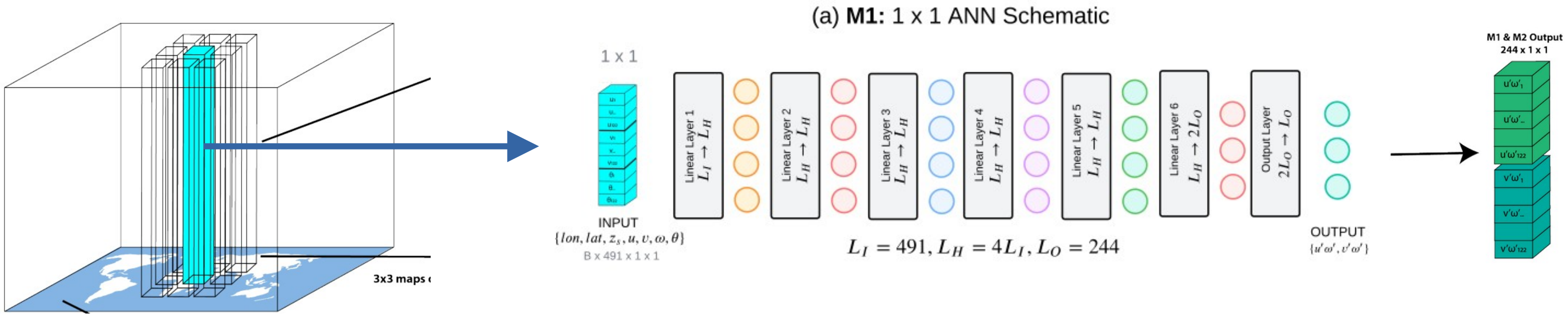
Gravity wave **momentum fluxes** from high-resolution reanalysis/obs
(unresolved by climate models)

Part 1: Nonlocal Emulation

Developing ML architectures to learn GW lateral propagation

We train three ML models with varying degrees of nonlocality

M1: Single Column

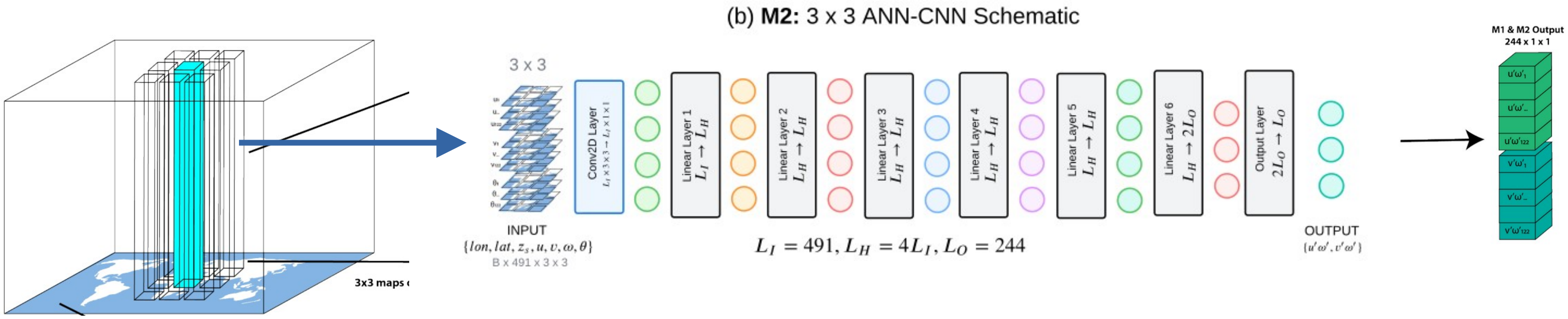


Model M1: inspired from traditional parameterizations

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

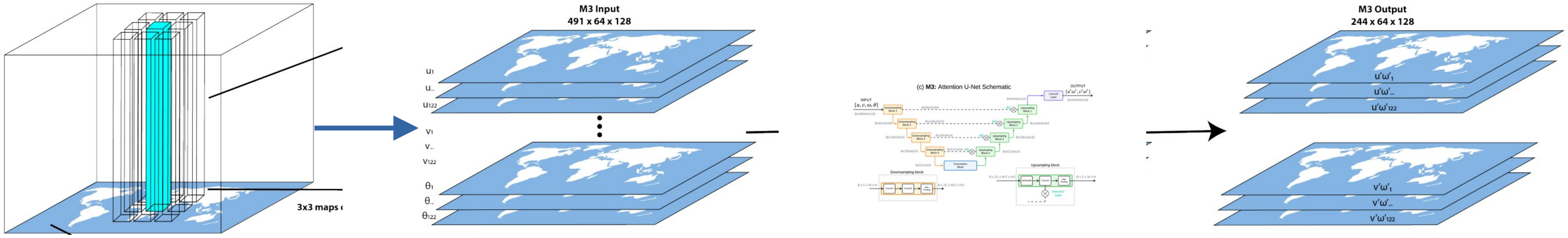


Model M2: Introducing slight nonlocality in space

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 nonlocal Attention UNet (Oktay et al. 2018)

Global input of dynamical variables to predict fluxes globally.

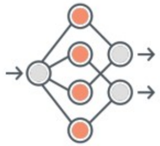
Training Configuration for Models M1-M3

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)

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



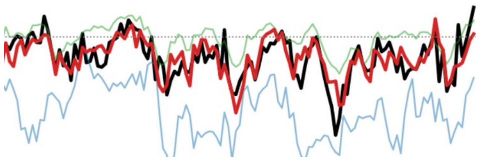
$$\begin{bmatrix} u'\omega' \\ v'\omega' \end{bmatrix}$$

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

- M1-M3 first trained on 4 years of ERA5 reanalysis (3 years training + 1 year validation). Identical hyperparameters and similar model sizes.
- Later, re-trained on 4 months (NDJF) of 1.4 km global ECMWF-IFS model.
- Trained on different feature sets, both globaly and exclusively in the stratosphere.

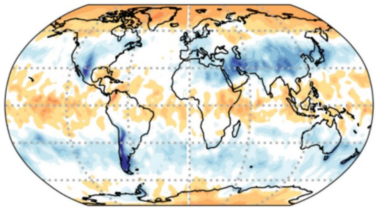
Evaluate performance beyond RMSE

Test 1. Temporal Evolution



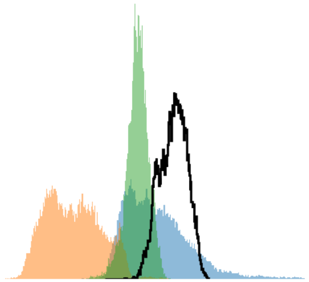
Does the model correctly learn the temporal wave evolution

Test 2. Seasonal Averages



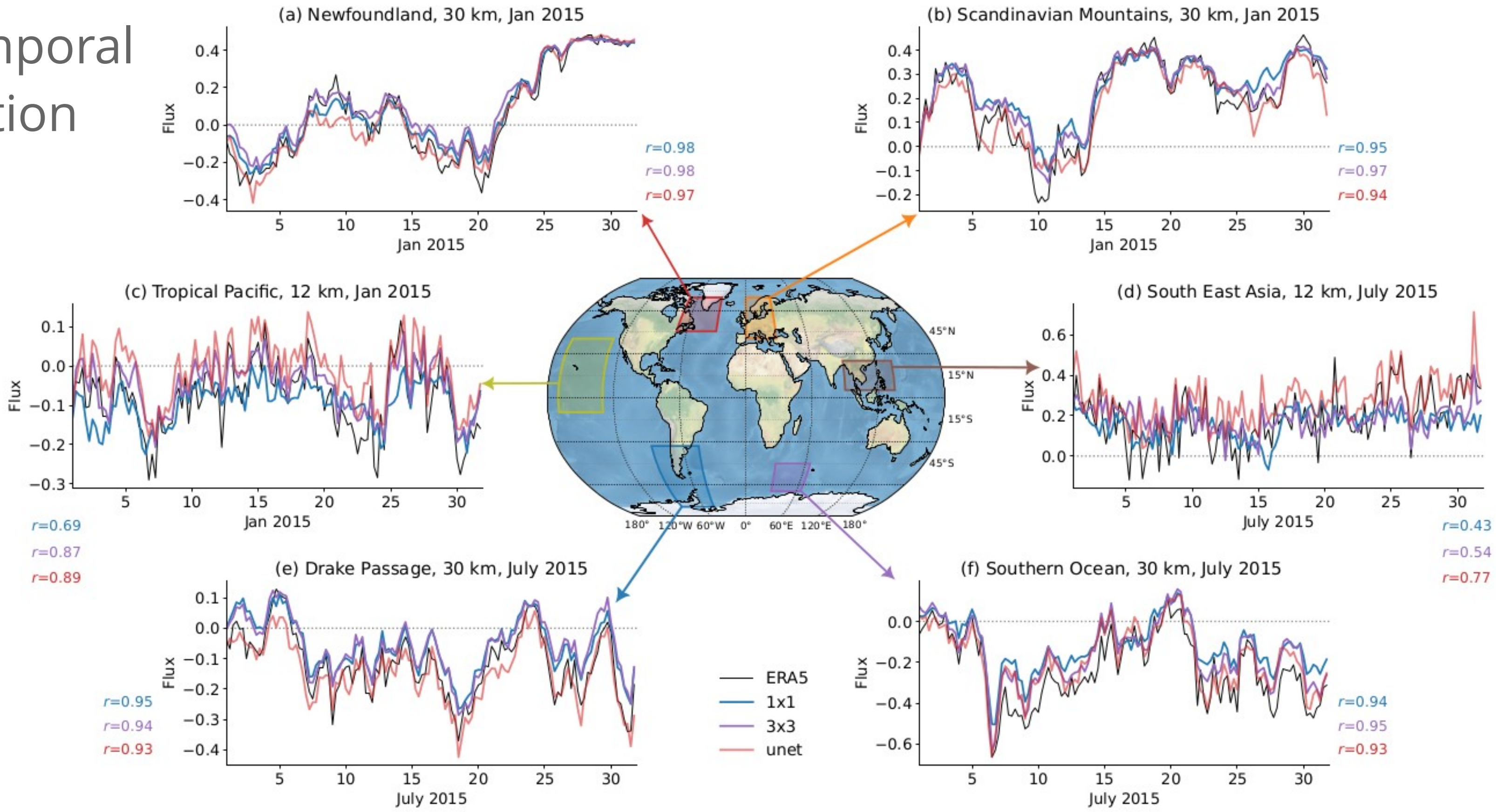
Does the model generate accurate global flux distribution?

Test 3. Flux distribution



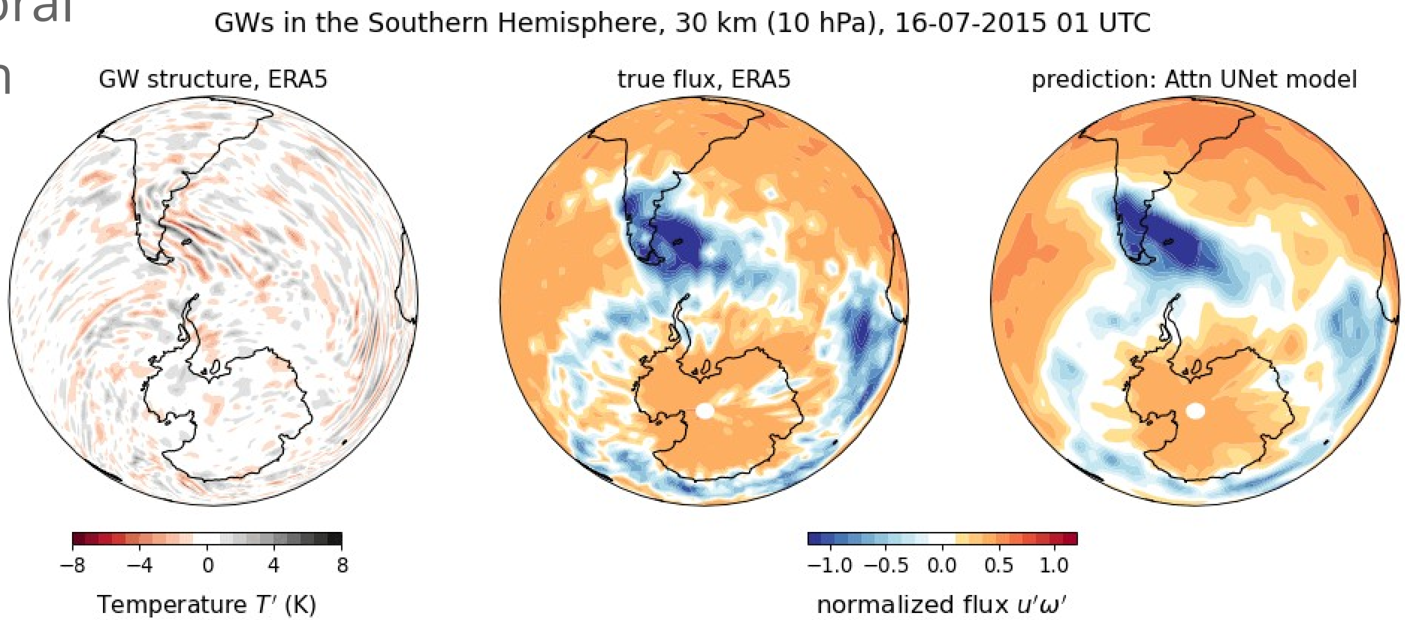
Does the model generate desired statistics?

1. Temporal Evolution



M1-M3 models skillfully learn the intermittent and coherent evolution of GW fluxes in the atmosphere over both orographic and nonorographic hotspots. **Nonlocal models uniformly perform better.**

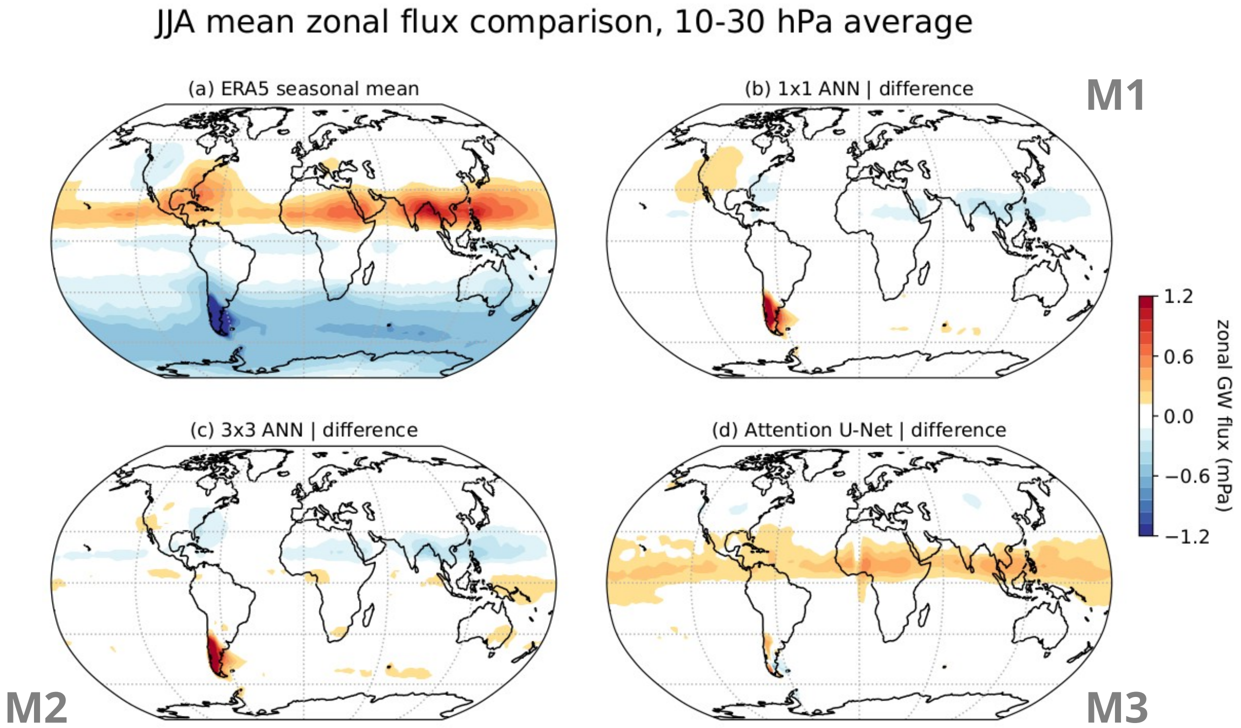
1. Temporal Evolution



Attention Unet (M3) correctly predicts 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 Averages



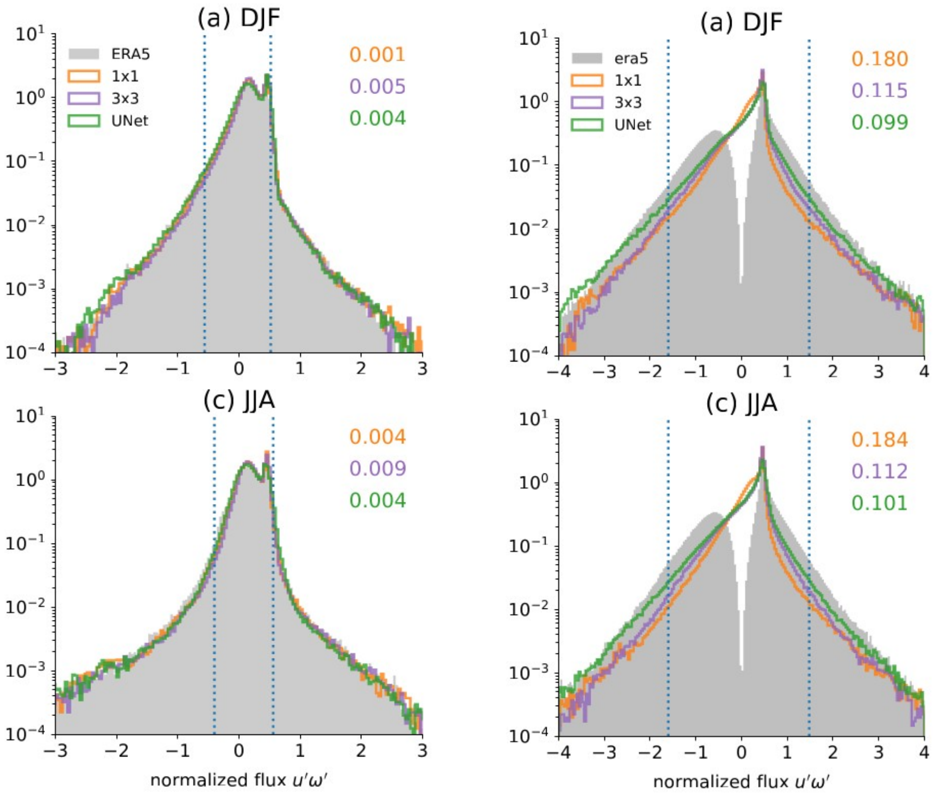
All of M1, M2, M3 generate commendable predictions.

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

3. Global Flux Distribution

Hellinger distance between two distributions:

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

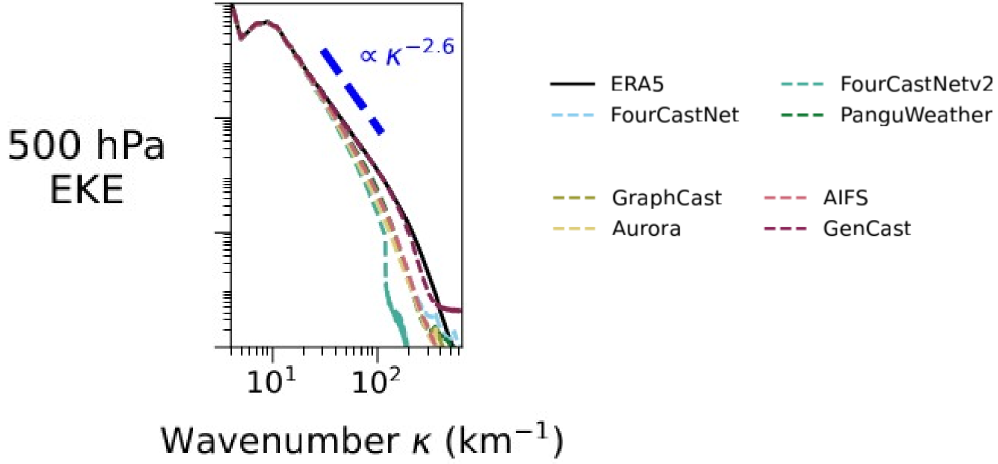


Seasonal averages

daily averages

The seasonally averaged distributions are reproduced quite well.

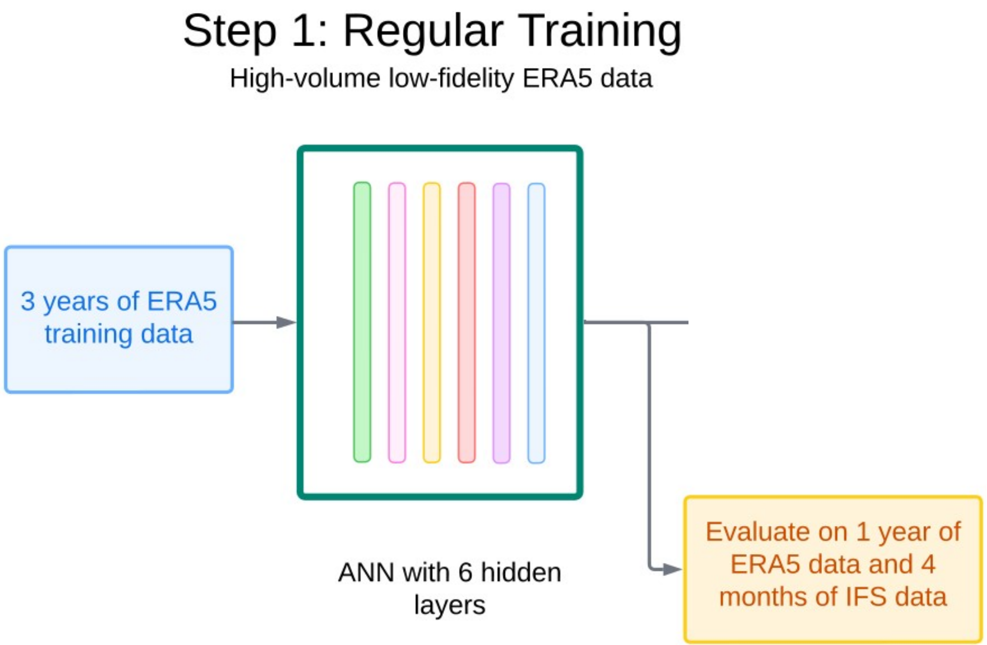
... but the neural nets struggle with small values – predict zeros instead. Similar to AIWP models underestimating small scales?



Part 2: Transfer Learning (TL)

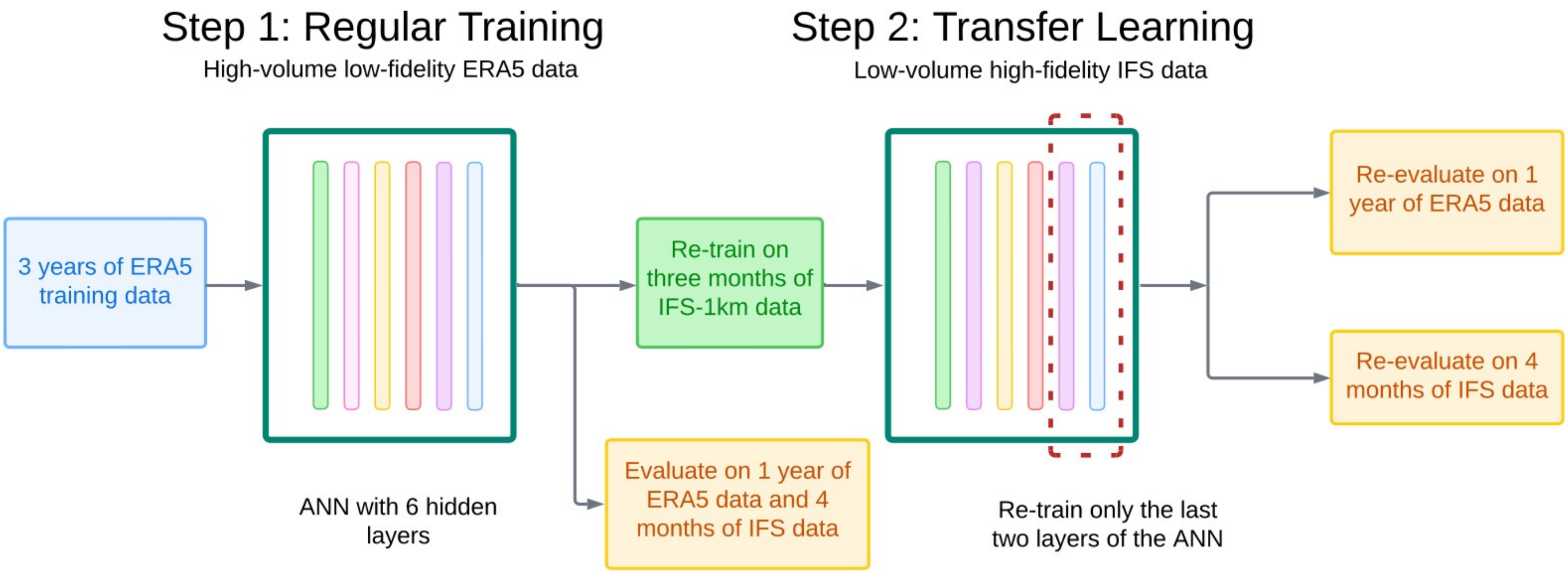
Blending low-fidelity datasets with high-fidelity datasets

Improving predictions using transfer learning (TL) on high-res datasets



ERA5 under-resolves mesoscale GWs.

Improving predictions using transfer learning (TL) on high-res datasets

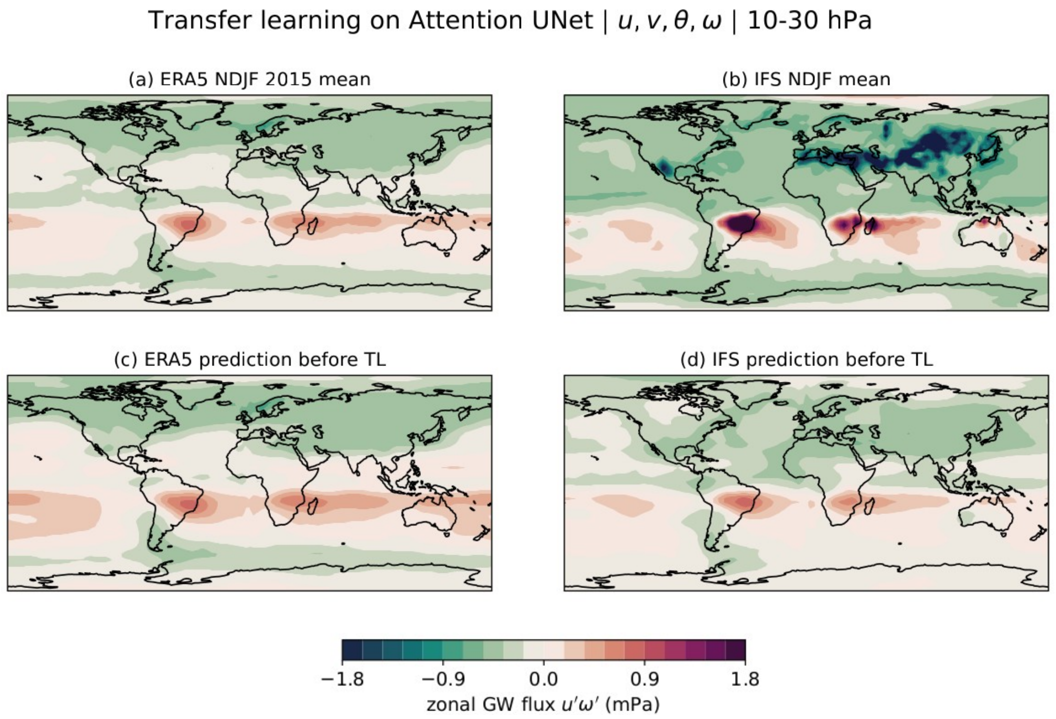


ERA5 under-resolves mesoscale GWs. This underestimation is corrected by transfer learning on limited-period-high-resolution fluxes from a kilometer-scale global models

Partly Retraining on 1km global ECMWF-IFS: Best of Both Worlds?

NDJF Fluxes
from ERA5

NDJF Fluxes
from IFS-1km



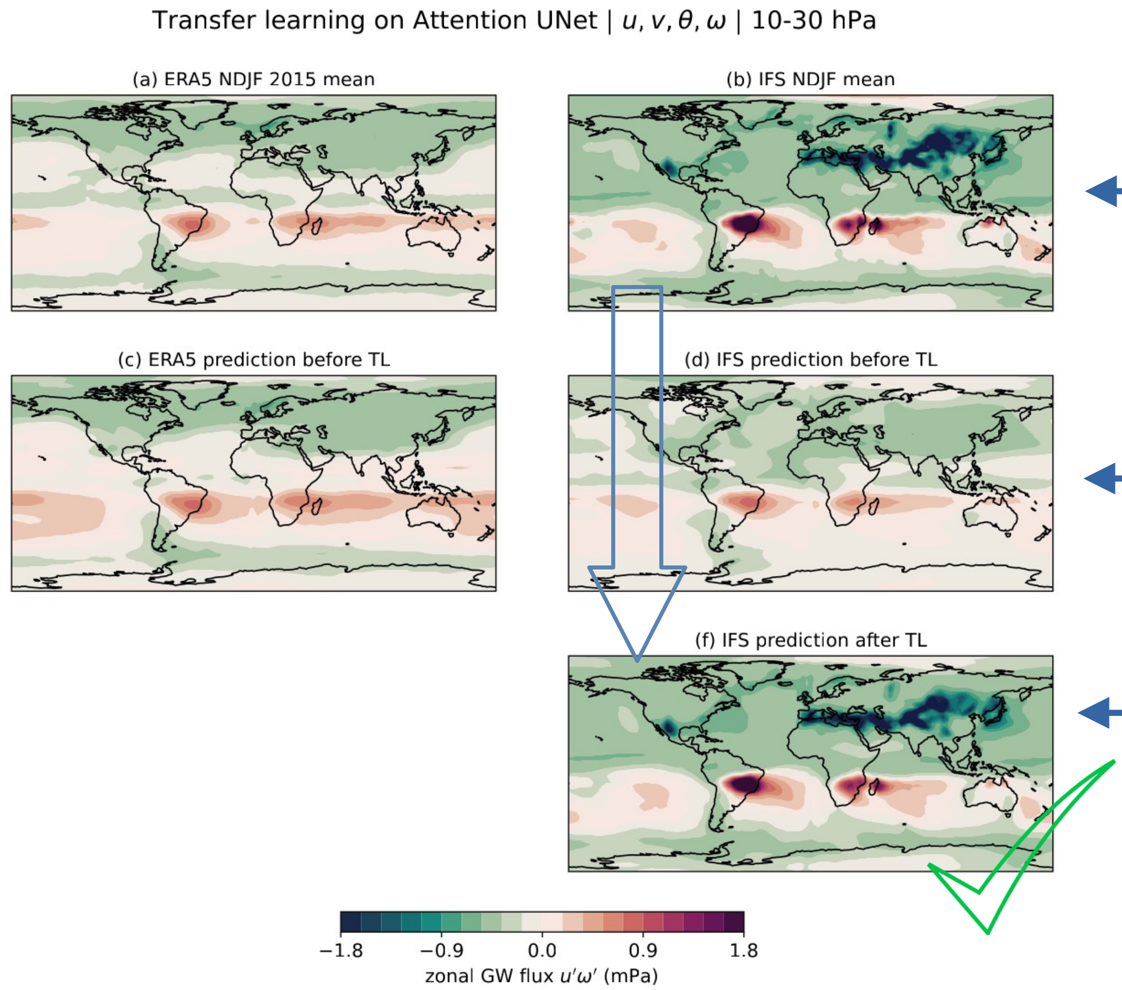
Trained only on ERA5
Tested on ERA5
Skillful prediction

Trained only on ERA5
Tested on IFS-1km
Under-resolves fluxes

Partly Retraining on 1km global ECMWF-IFS: Best of Both Worlds?

NDJF Fluxes from ERA5

Trained only on ERA5
Tested on ERA5
Skillful prediction



NDJF Fluxes from IFS-1km

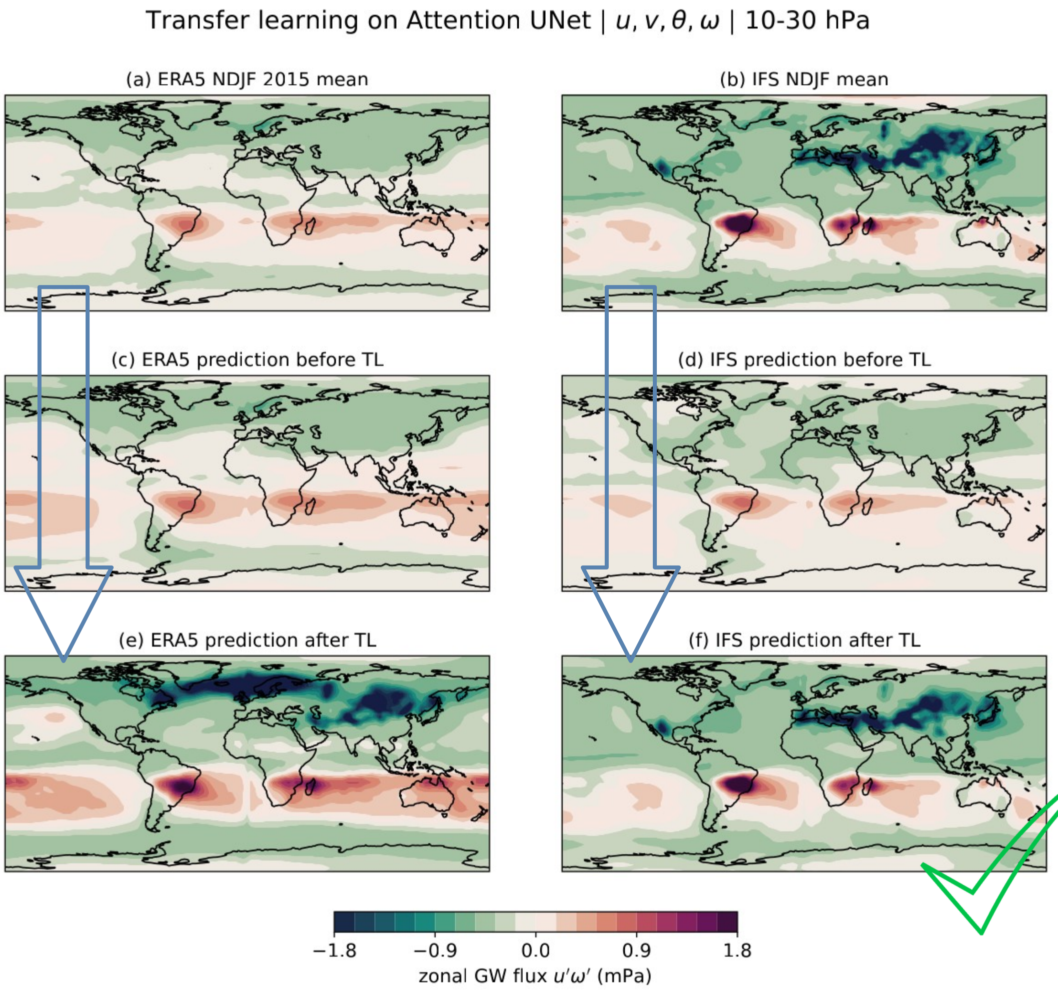
Trained only on ERA5
Tested on IFS-1km
Under-resolves fluxes

Correction of IFS flux prediction after transfer learning (as expected)

Partly Retraining on 1km global ECMWF-IFS: Best of Both Worlds?

NDJF Fluxes from ERA5

NDJF Fluxes from IFS-1km



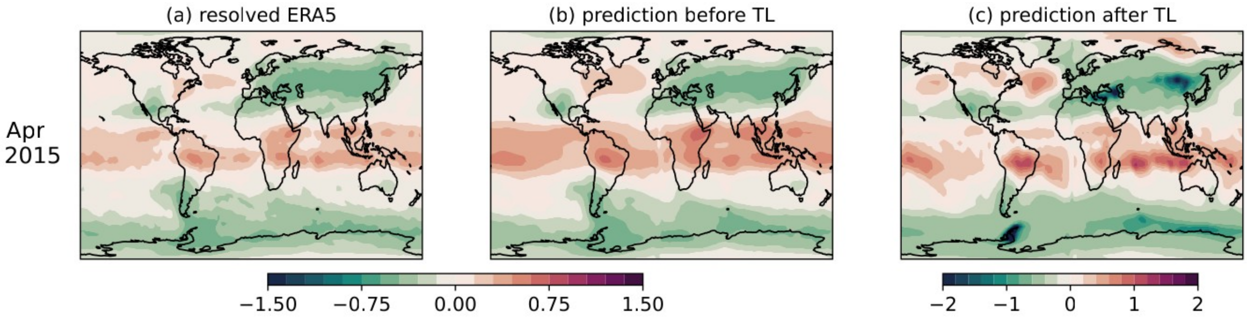
Trained only on ERA5
Tested on ERA5
Skillful prediction

Trained only on ERA5
Tested on IFS-1km
Under-resolves fluxes

Improved magnitudes in ERA5 after transfer learning

Correction of IFS flux prediction after transfer learning (as expected)

TL Yields Skillful Predictions on Out-of-set Months

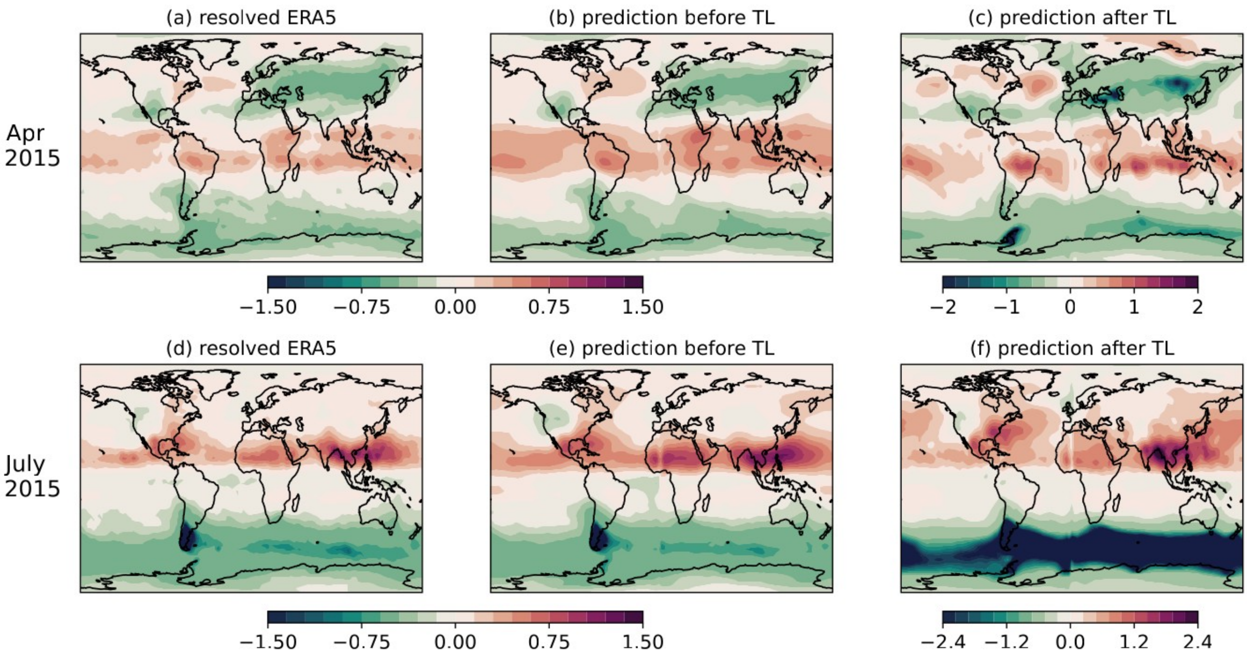


Following TL, models predict stronger fluxes, while identifying the correct hotspots.

The models blend learnings from both l*ow-fidelity high-volume* and *high-fidelity low-volume* datasets.

Models provide effective ‘scaling’ of fluxes even on out-of-set months.

TL Yields Skillful Predictions on Out-of-set Months

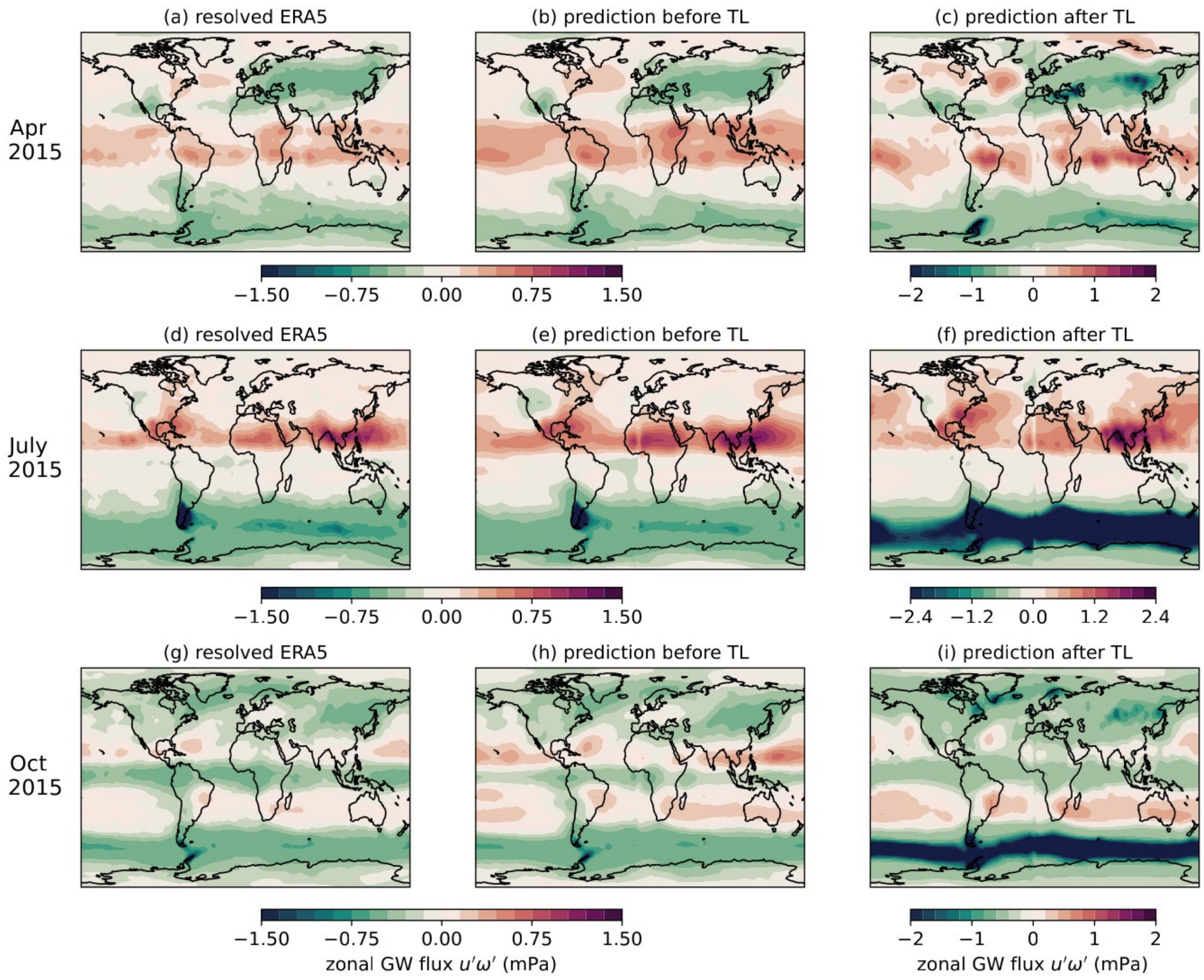


Following TL, models predict stronger fluxes, while identifying the correct hotspots.

The models blend learnings from both l*ow-fidelity high-volume* and *high-fidelity low-volume* datasets.

Models provide effective 'scaling' of fluxes even on out-of-set months.

TL Yields Skillful Predictions on Out-of-set Months

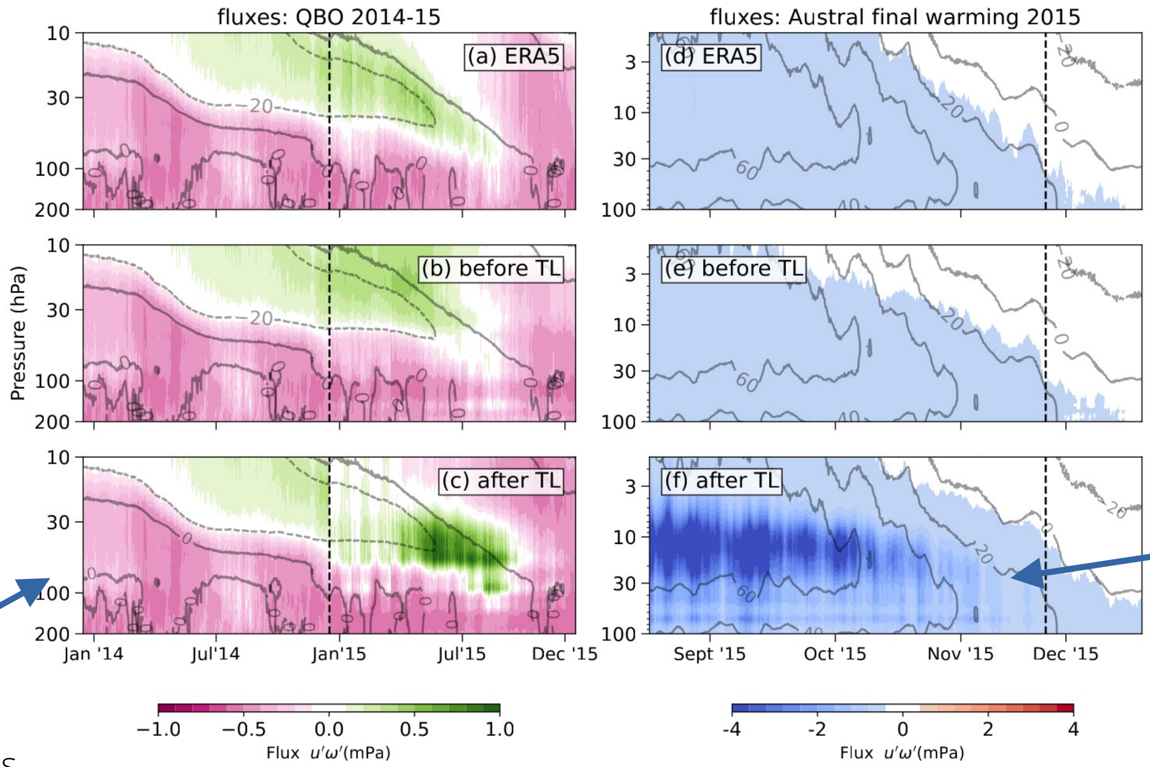


Following TL, models predict stronger fluxes, while identifying the correct hotspots.

The models blend learnings from both l*ow-fidelity high-volume* and *high-fidelity low-volume* datasets.

Models provide effective ‘scaling’ of fluxes even on out-of-set months.

Physically Consistent Performance on Key Stratospheric Features



Improved prediction
around QBO transitions
in July '15

Improved prediction
around vortex
breakdown
in Sept-Oct '15

Despite, transfer learning
only on NDJF data from
IFS-1km

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

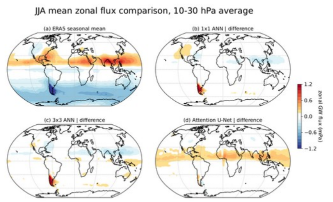
Aman Gupta[1], Aditi Sheshadri[1], Sujit Roy[2,3], Valentine Anantharaj[4]

[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

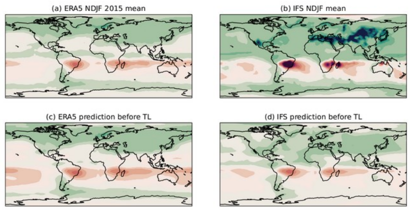
Submitted to JAMES (**preprint: tinyurl.com/gwbaco25**)

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

Key Conclusions

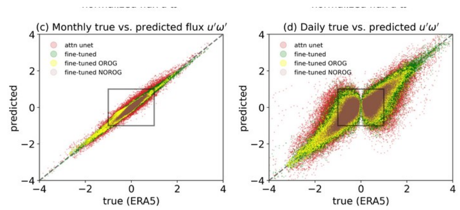


1. **Skillful performance:** The three ML schemes learn nonlocal propagation, temporal coherence, and seasonal distributions of GW fluxes from high-resolution data.



2. **Importance of nonlocality:** 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

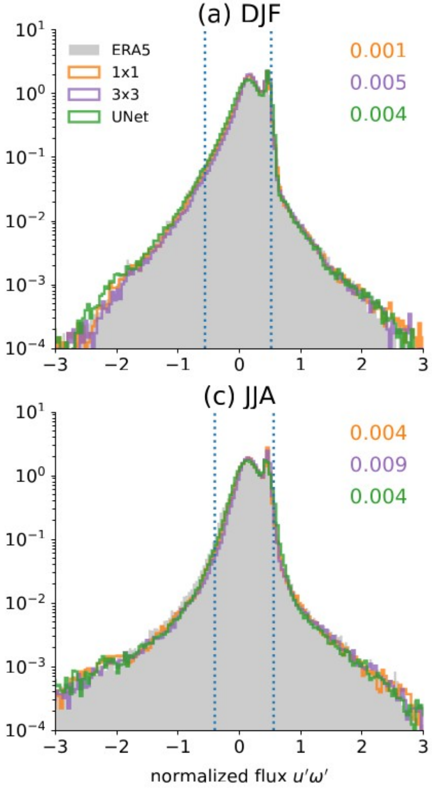
1. Testing performance on dissimilar model outputs: high-resolution CAM and ICON runs.

2. Coupling the ML scheme to a climate model (CAM7) to test "online" performance and stratospheric variability: a software engineering challenge

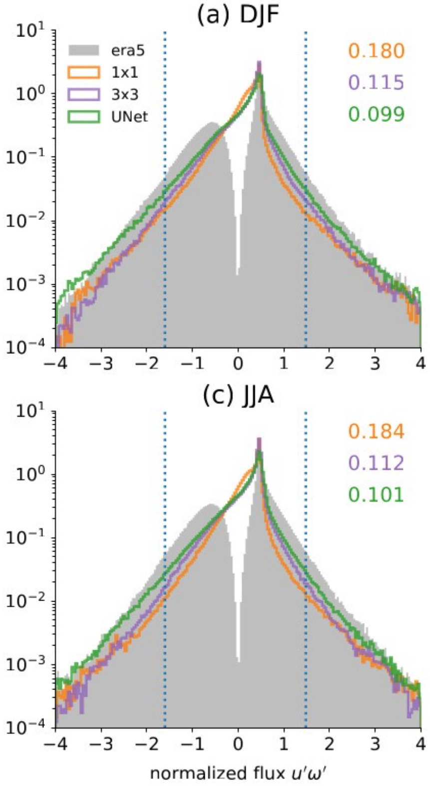
Supplement

3. Global Flux Distribution

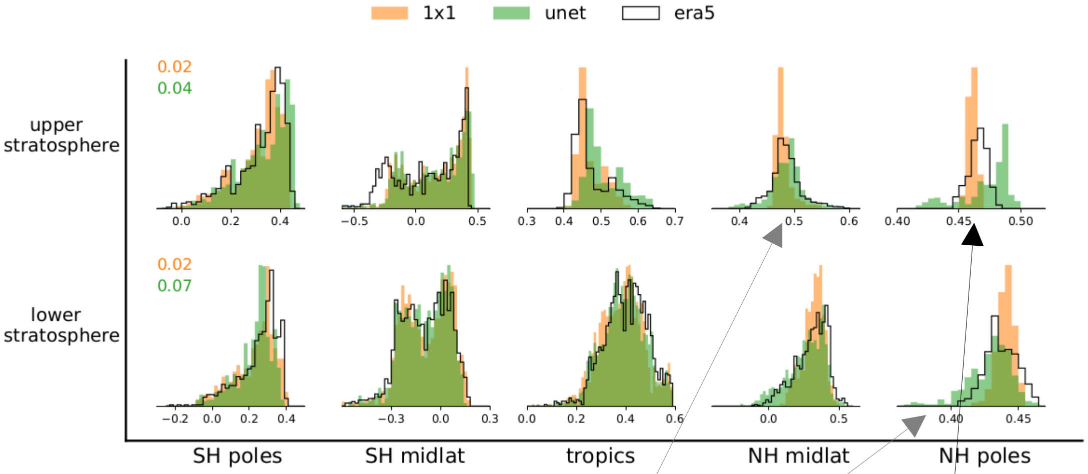
$$\mathcal{H}(p,q) = \frac{1}{2}\int_{x\in X}\left(\sqrt{p(x)}-\sqrt{q(x)}\right)^2 dx = 1-\int_{x\in X}\sqrt{p(x)q(x)}dx.$$



Seasonal averages



daily averages

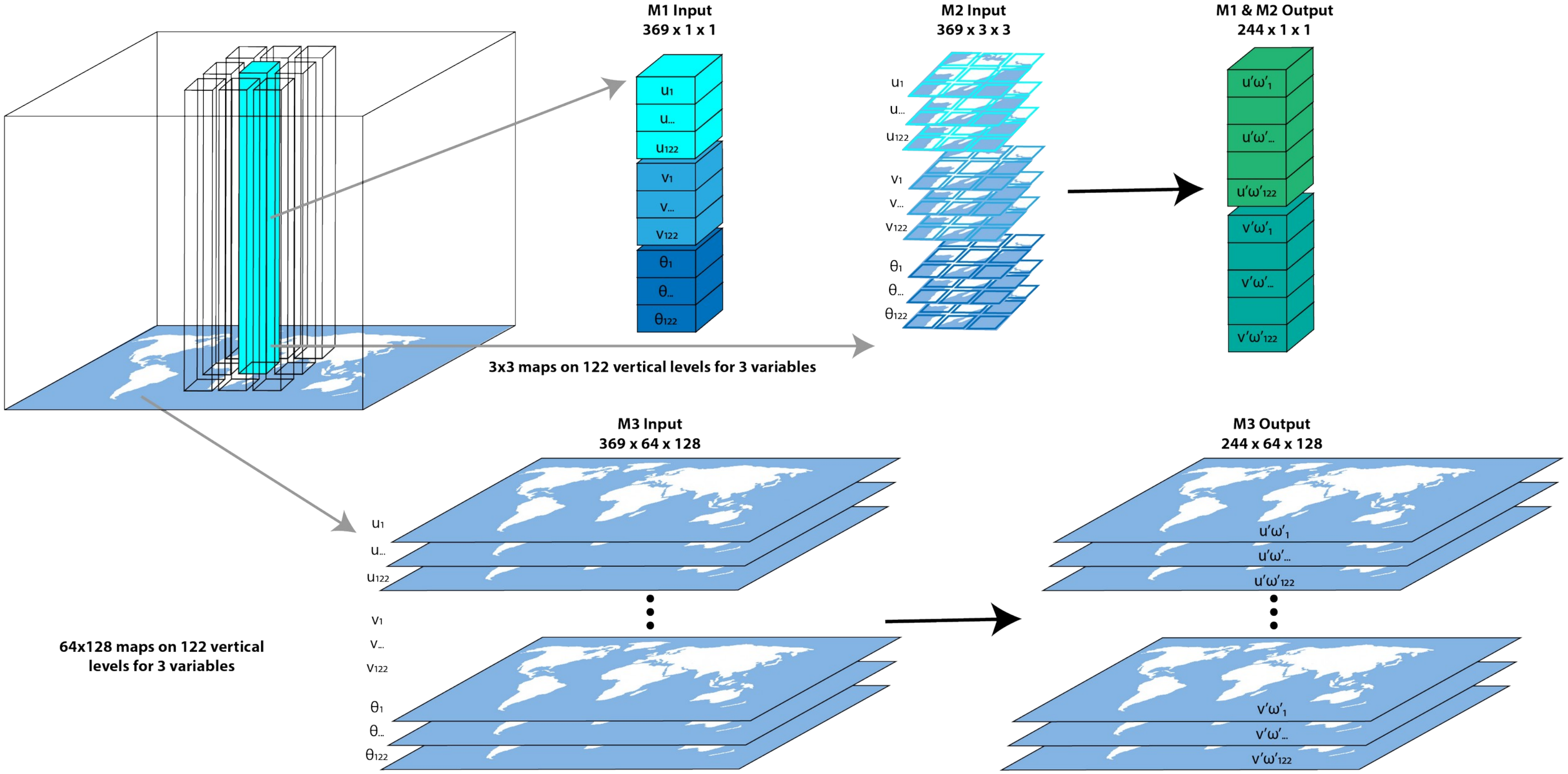


The three models generate comparable distribution tails for all seasons

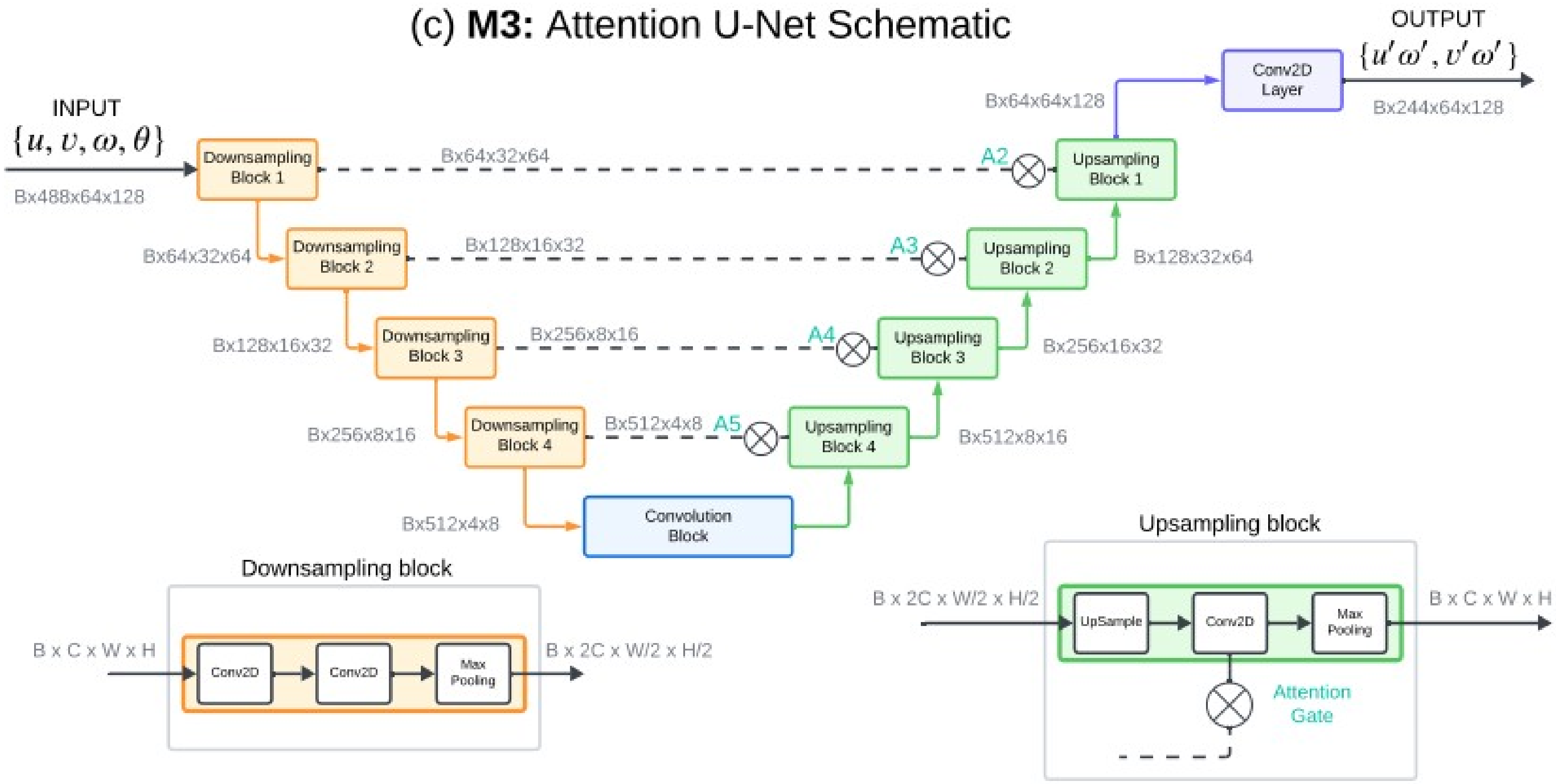
Prominent narrow bias in flux predictions by ANNs

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

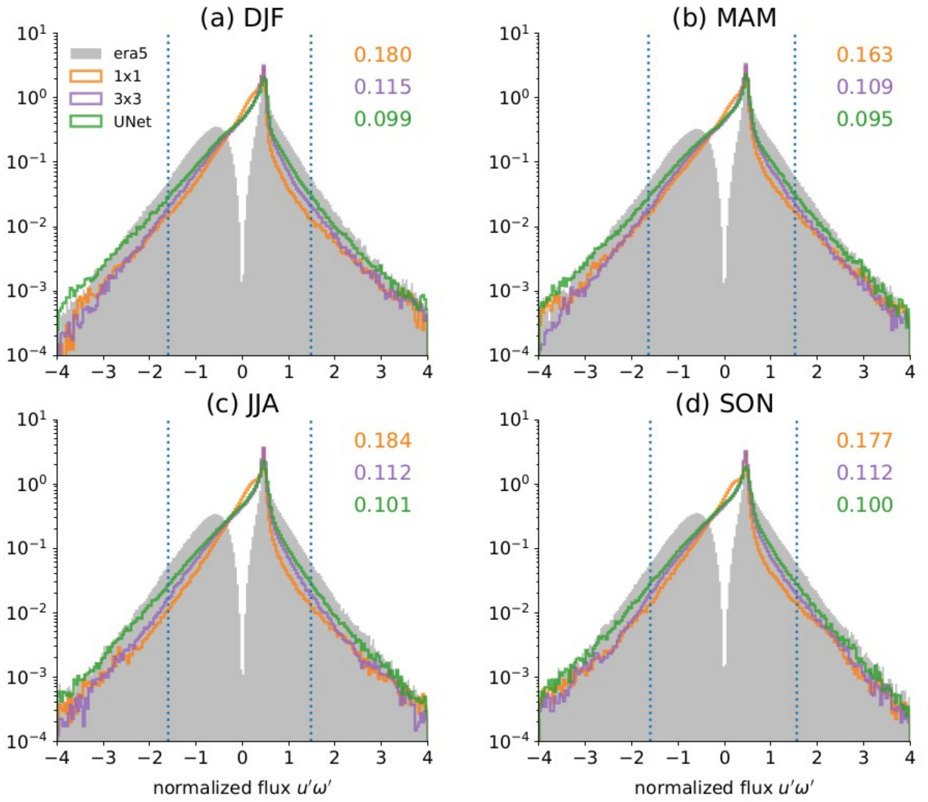
Learning nonlocality through nonlocal architectures



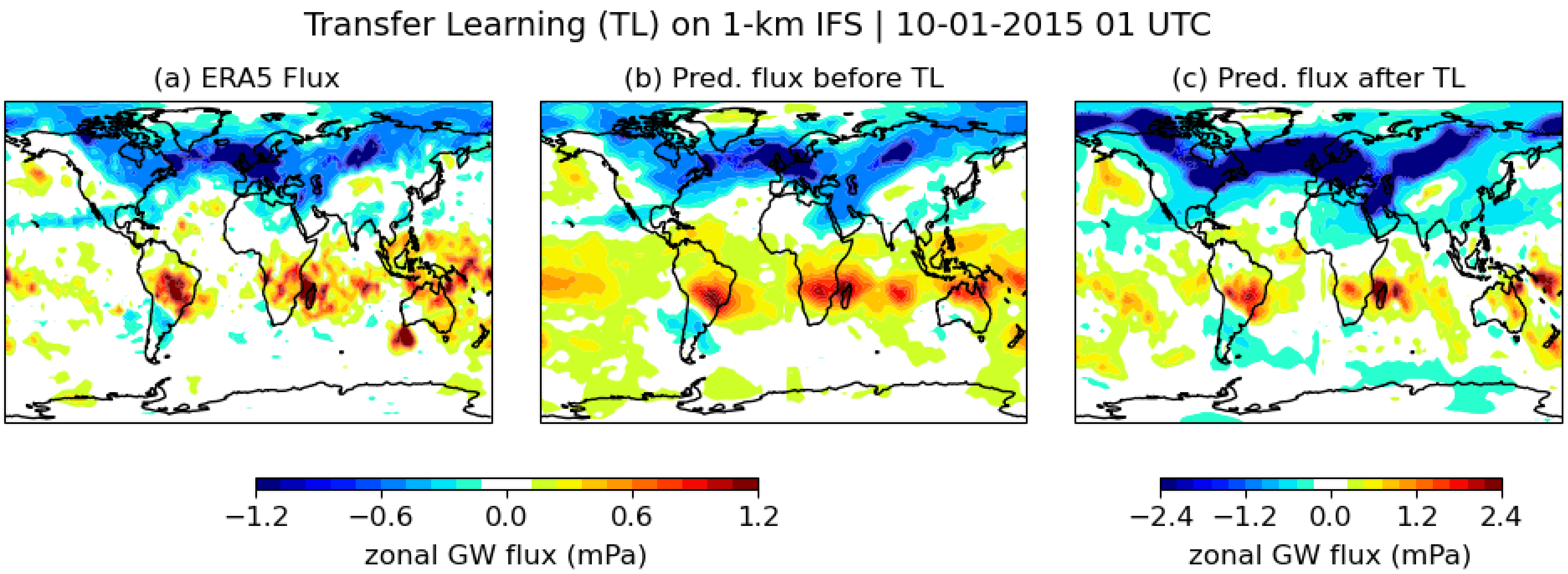
Attention UNet Schematic

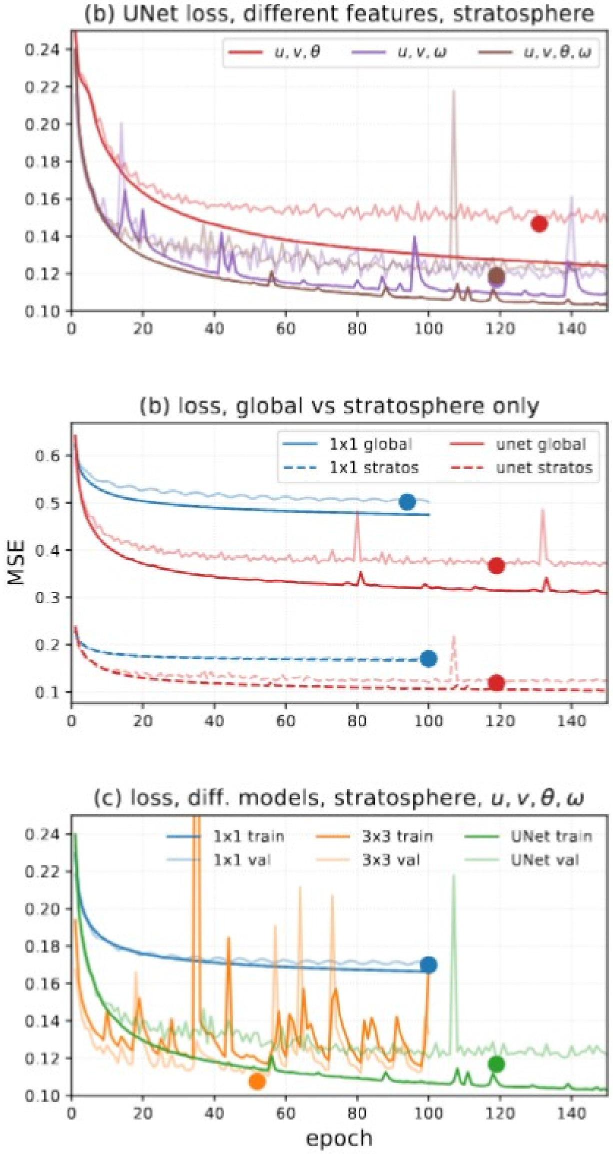


Daily Sampled Flux Distributions



Transfer Learning on out-of-set months

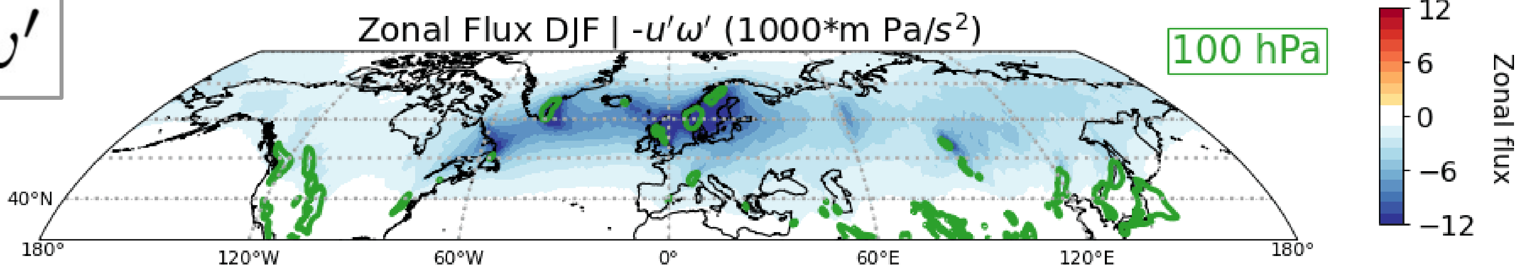



(b) UNet loss, different features, stratosphere
u, v, θ
u, v, ω
u, v, θ, ω
(b) loss, global vs stratosphere only
1x1 global
1x1 stratos
unet global
unet stratos
MSE
(c) loss, diff. models, stratosphere, u, v, θ, ω
1x1 train
1x1 val
3x3 train
3x3 val
UNet train
UNet val
epoch

GWs form a belt of wave activity in the middle atmosphere

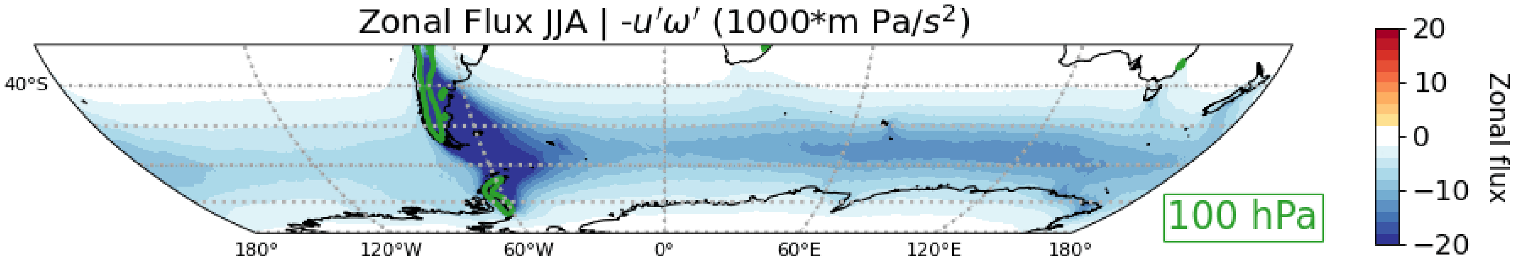
Green: Flux envelope, Color: Flux at 2 hPa (~45 km)

$$-u'\omega'$$



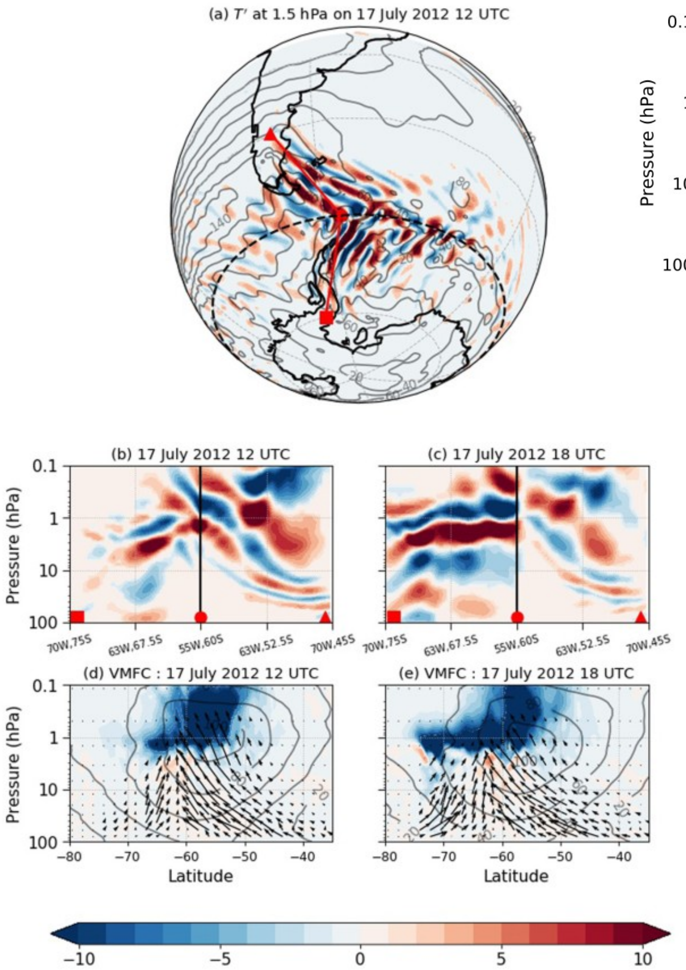
1979-2022
December
January
February
mean
From
reanalysis

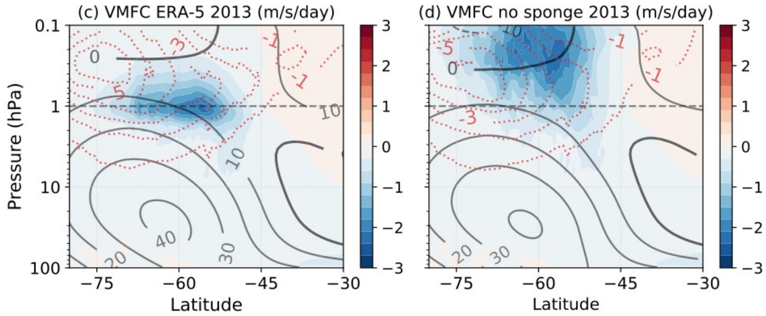
local GW generation ⇨ propagation through strong shear ⇨ global spreading



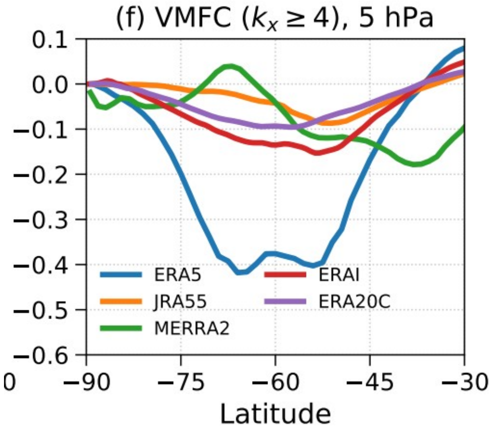
1979-2022
June
July
August
mean from
reanalysis

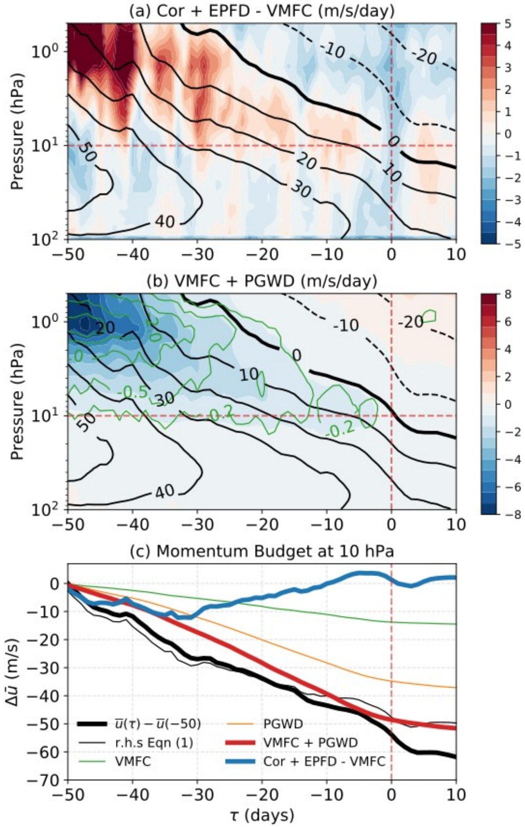
Gupta, Sheshadri, Alexander, Birner (2024), GRL | Insights on Lateral Gravity Wave Propagation in the Extratropical Stratosphere from 44 Years of ERA5 Data





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

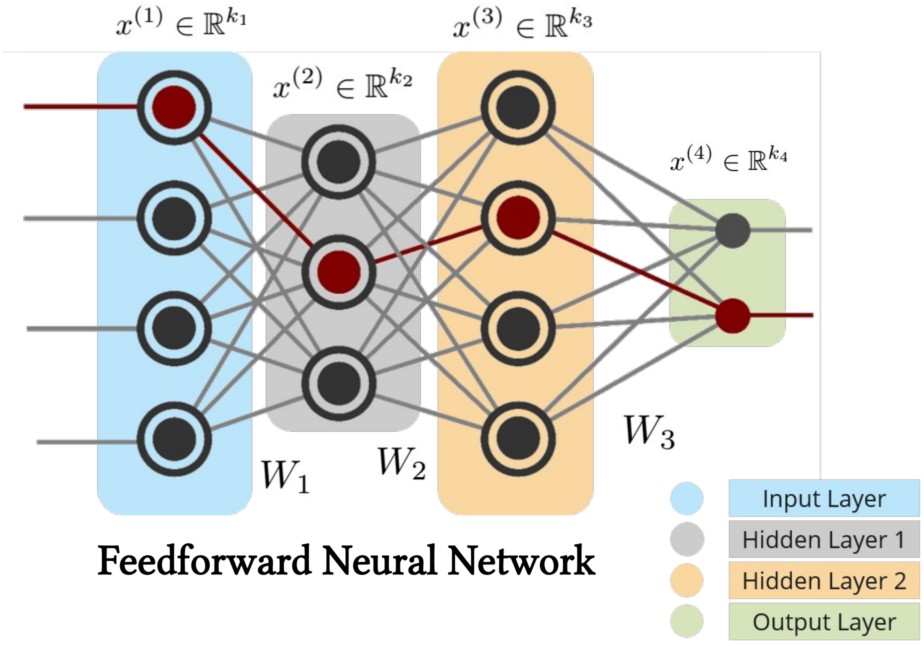




Neural Network as a Collection of Perceptrons

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:



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(.)

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

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