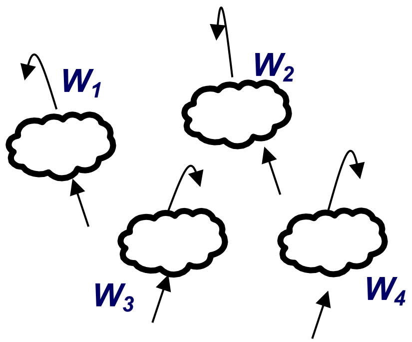
Deep learning parameterization of small-scale vertical velocity variability for atmospheric models

Donifan Barahona[1], Katherine Breen[1,2], Heike Kalesse-Los[3] and Johannes Röttenbacher[3]

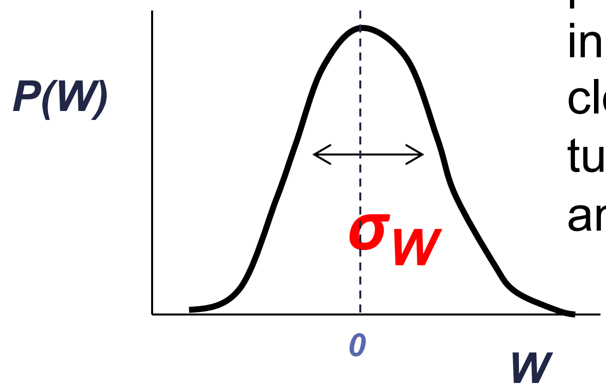
1. Global Modelling and Assimilation office, NASA Goddard Space Flight Center, Greenbelt, MD, USA.
2. Morgan State University, Baltimore, MD, USA.
3. Leipzig Institute for Meteorology, University of Leipzig, Leipzig, Germany.

Correspondence: donifan.o.barahona@nasa.gov

Subgrid Scale Dynamics is Highly Uncertain in Models

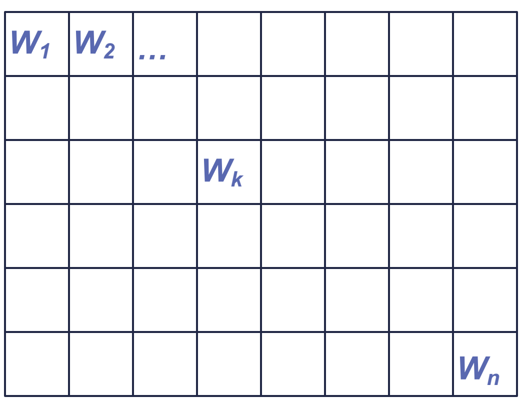


Grid cell (25-100 km)



Vertical wind motion plays a significant role in the representation of cloud formation, turbulence, PBL height and tracer transport.

P(W) can be explicitly simulated using ultra-high resolution models (km-scale and higher):



Many values for each 0.5° cell

σ_W

Downsampling

- ***Very Expensive***
- **Can't resolve all scales**
- **Biases still exist**

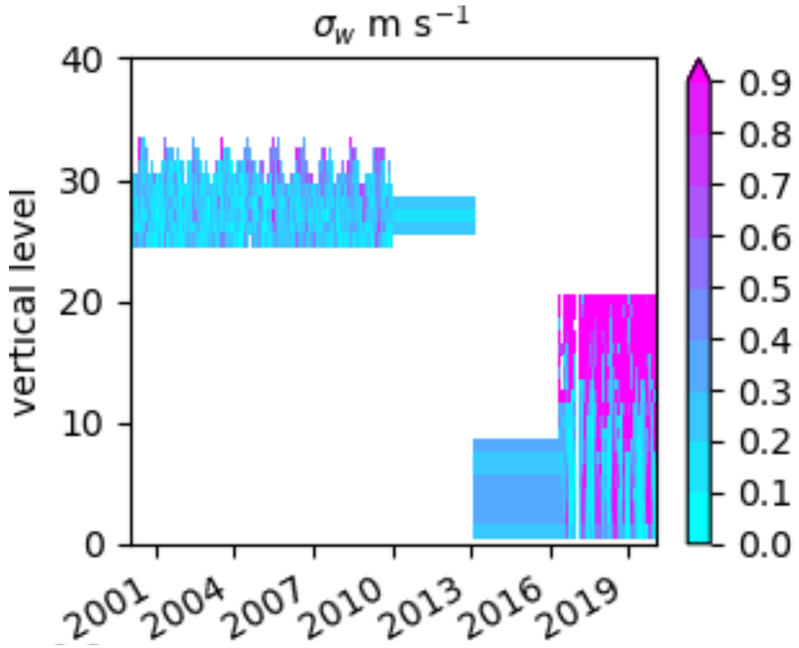
Wealth of data for training

Barahona et al. (2017)

Observational Constraints

W and σ_w can be retrieved using Doppler radar and lidar.

- Limited domain.
- Radar retrievals need clouds, lidar is confined to the boundary layer.
- **Experimental error could be significant.**

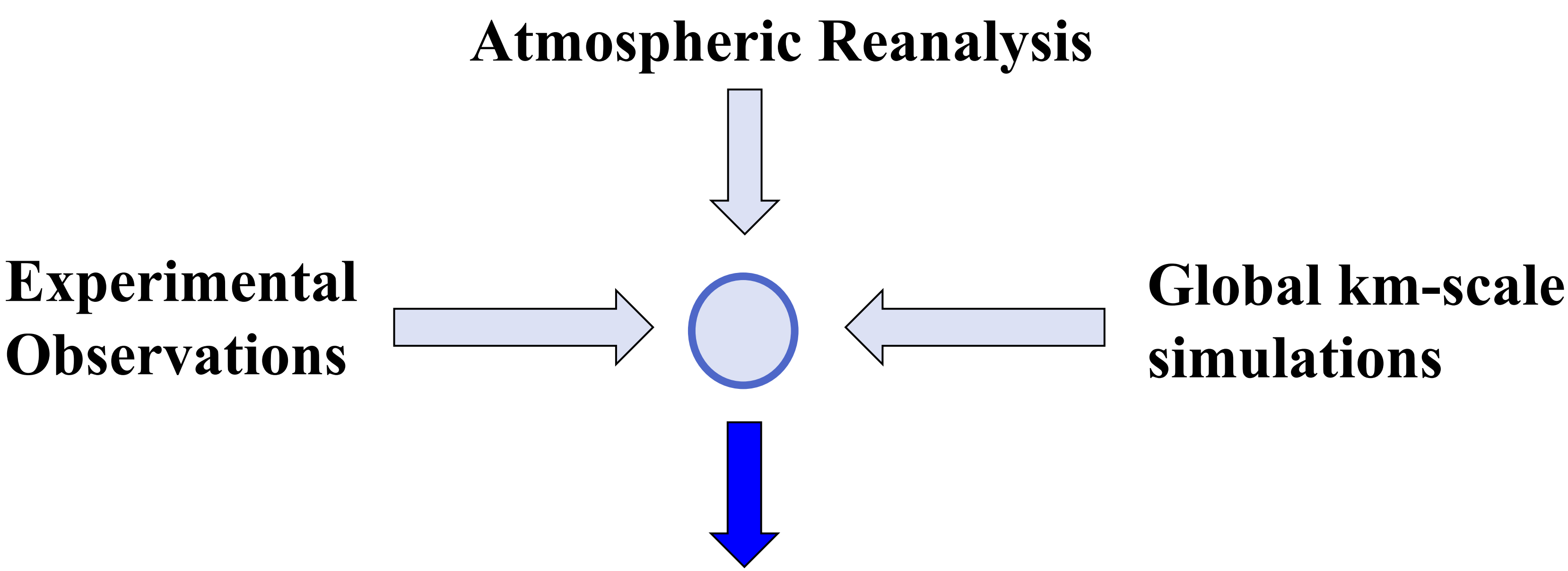


The impact of uncertainty can be mitigated by developing a model that aims to match the statistics of σ_w.

Generative adversarial networks:

"The goal of a generative modeling algorithm is to learn a Pmodel(x) that approximates Pdata(x) as closely as possible." *Goodfellow et al. 2020*

Parameterizing Subgrid variability in W



$$F(X) = \sigma_W \;;\; P(\sigma_W) \sim P_{observed}(\sigma_W)$$

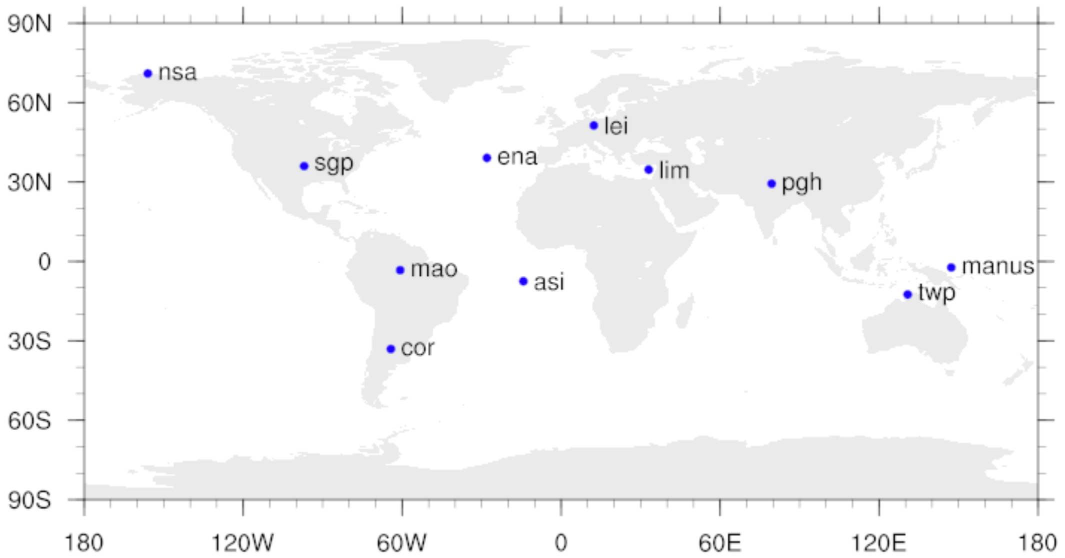
X = Atmospheric State (i.e., T, P, Winds, …) **at coarse resolution**
F = Artificial Neural Network
P = Probability distribution function

Data and Methods

GEOS Nature run (G5NR): 2 years of free-running simulation of the NASA GEOS-5 model with prescribed SST at 7 km spatial resolution. ~ 5Pb of data.

MERRA-2: NASA 1980-present atmospheric reanalysis. Provides the atmospheric state for each of the sites.

Observational Data:



W retrieved using Doppler radar and lidar, at 11 different sites around the world.

Mostly from DOE-ASR archive. New data at the "lei" and "lim" sites (CloudNet).

About 100 years of data on *W* at temporal resolution between 20 s and 5 min.

Parameterization Development

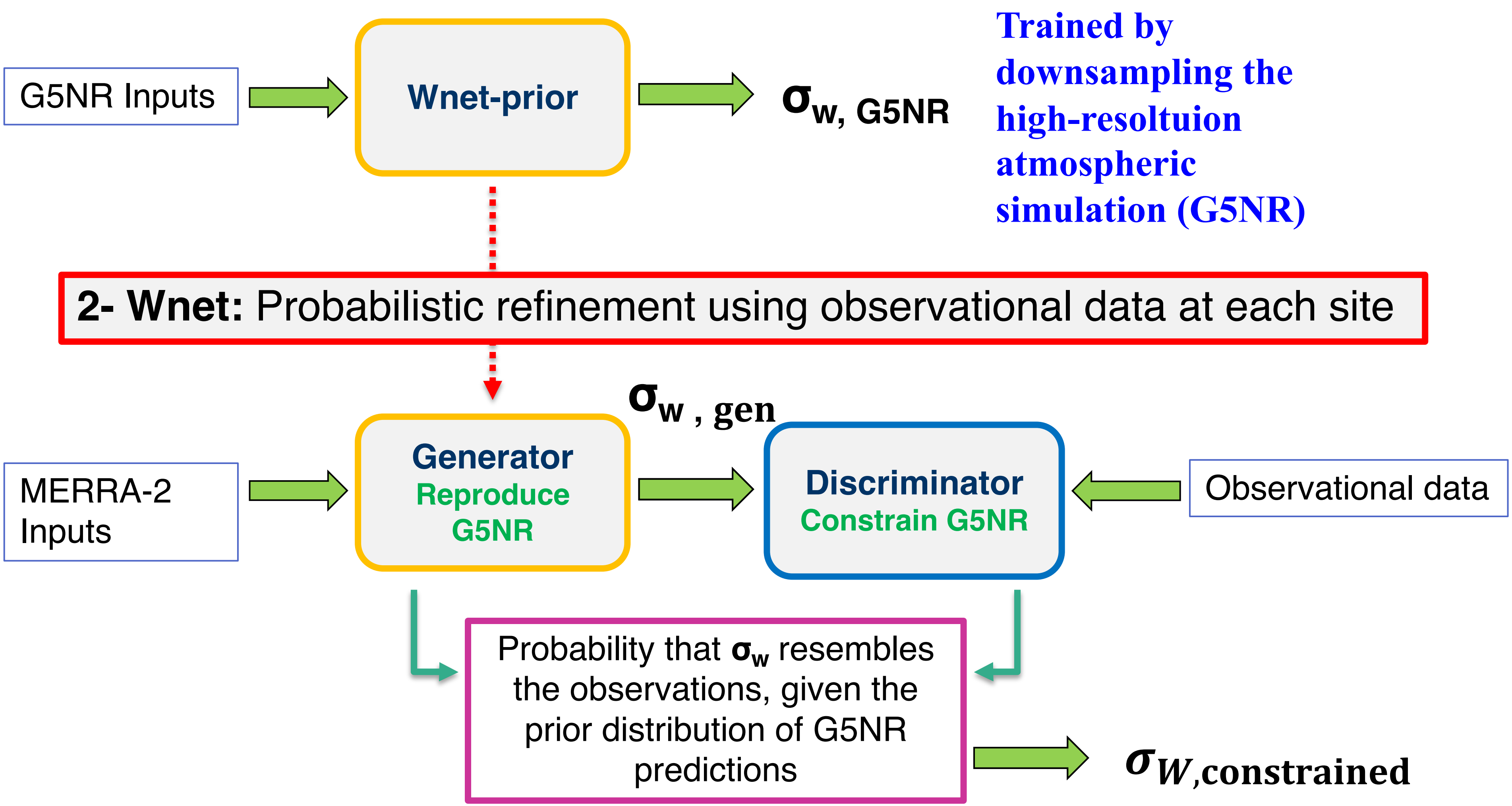
1- Wnet-prior: Surrogate model for G5NR vertical wind velocity

G5NR Inputs ~ 50 km → Wnet-prior → $\sigma_{w,\ \mathrm{G5NR}}$

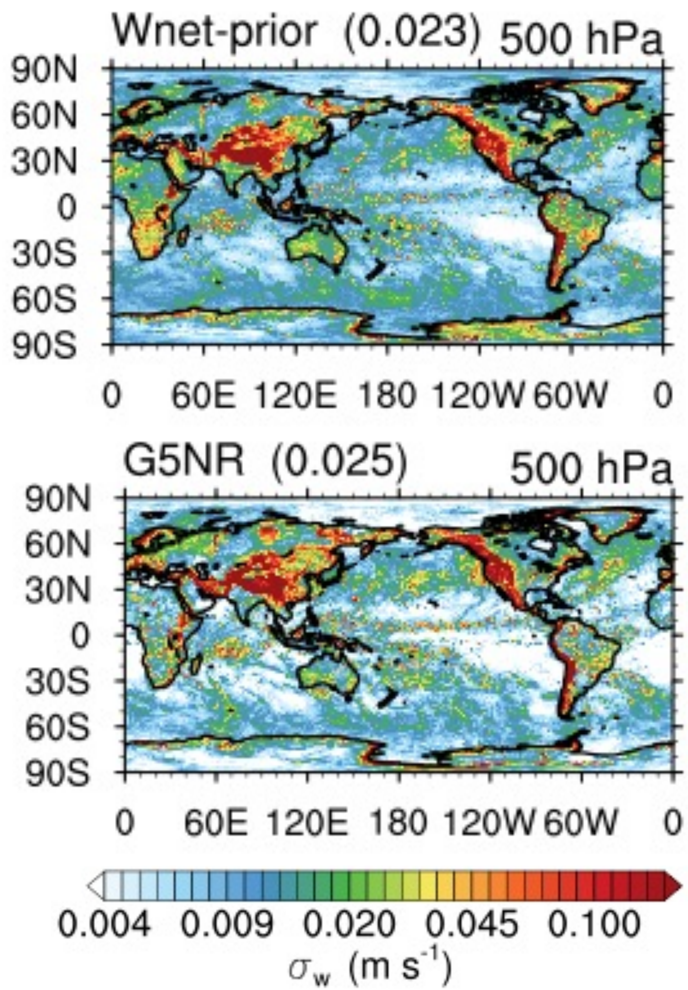
Trained by downsampling the high-resoltuion atmospheric simulation (G5NR)

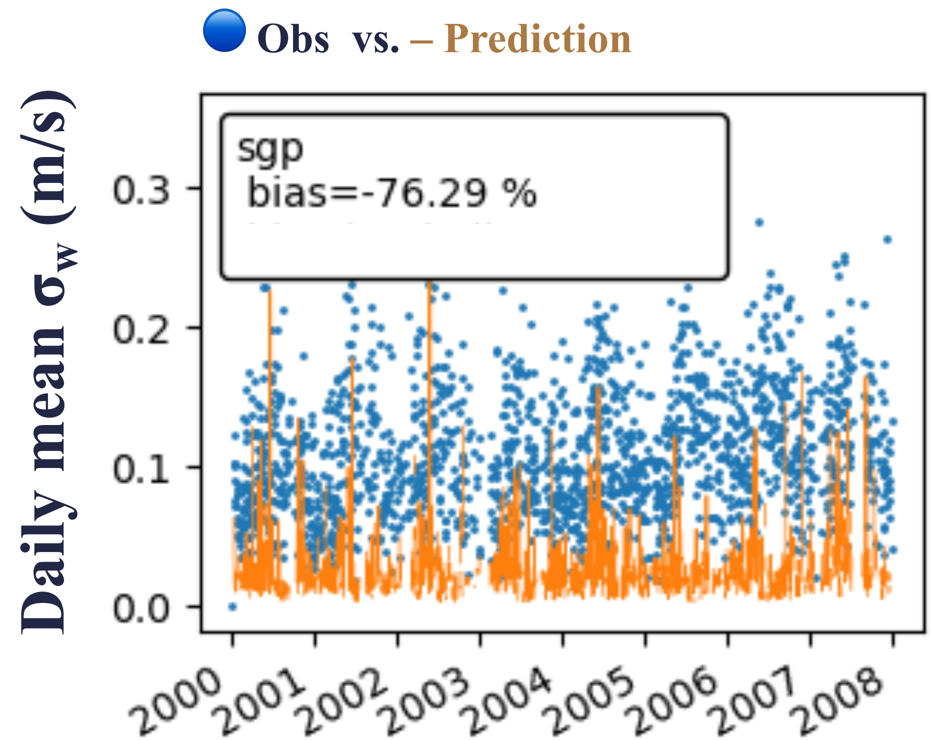
Parameterization Development

1- Wnet-prior: Surrogate model for G5NR vertical wind velocity



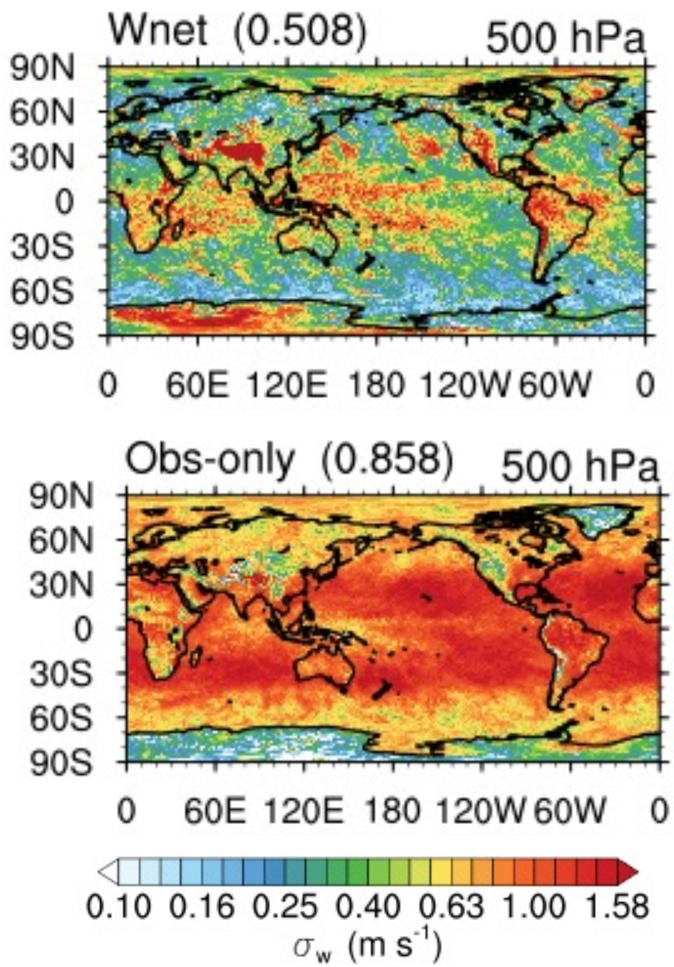
Results: Wnet-prior reproduces the GCRM

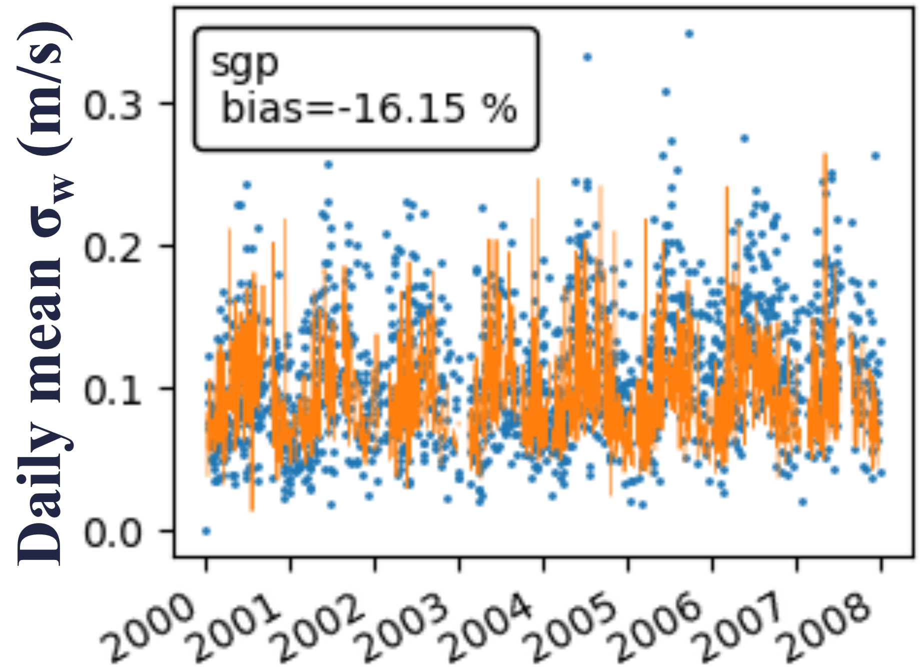




The ANN reproduces the GCRM output within 0.005 m s^{-1}.
However, it tends to underestimate variability, reflecting the bias of the high-resolution simulations.

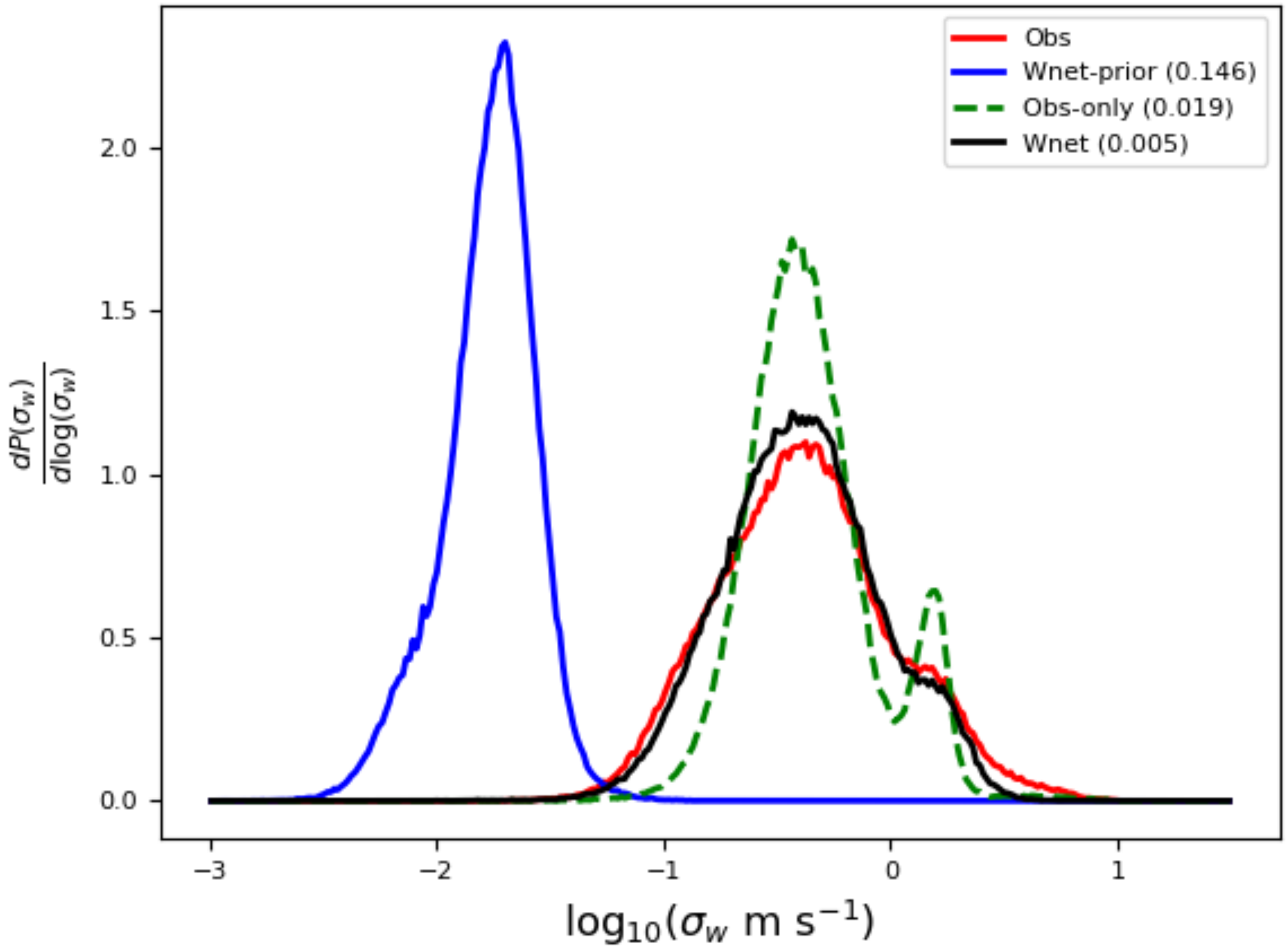
Results: GAN-Trained NN Parameterization Achieves High Accuracy





Using observational data is critical for the performance.
Wnet exhibits spatial structure that would be missing in a model trained using only observational data ("Obs-Only").

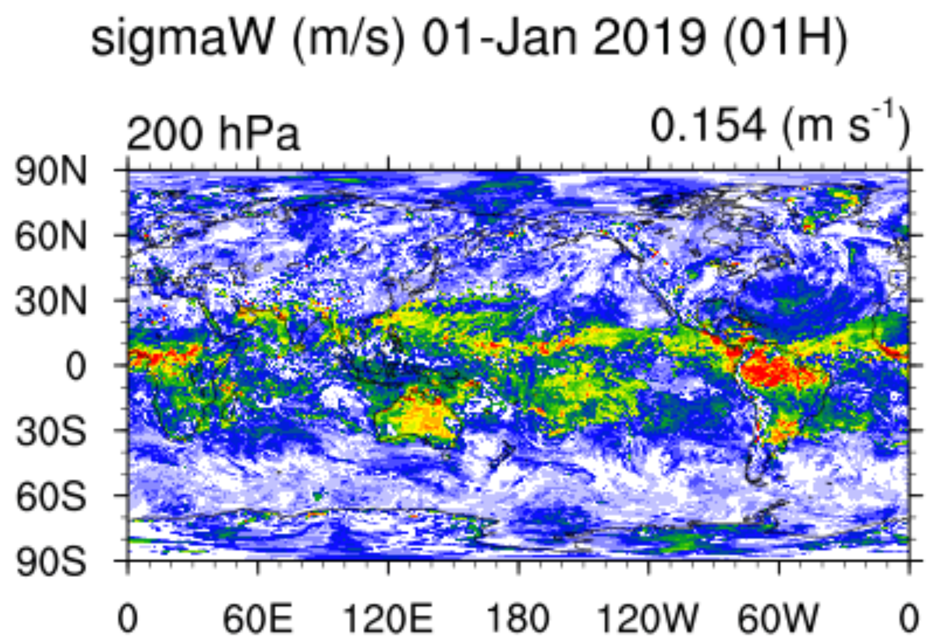
What about the PDF?



The statistics generated by the Wnet parameterization approximate well the observations, minimizing the effect of uncertainty on σ_W.

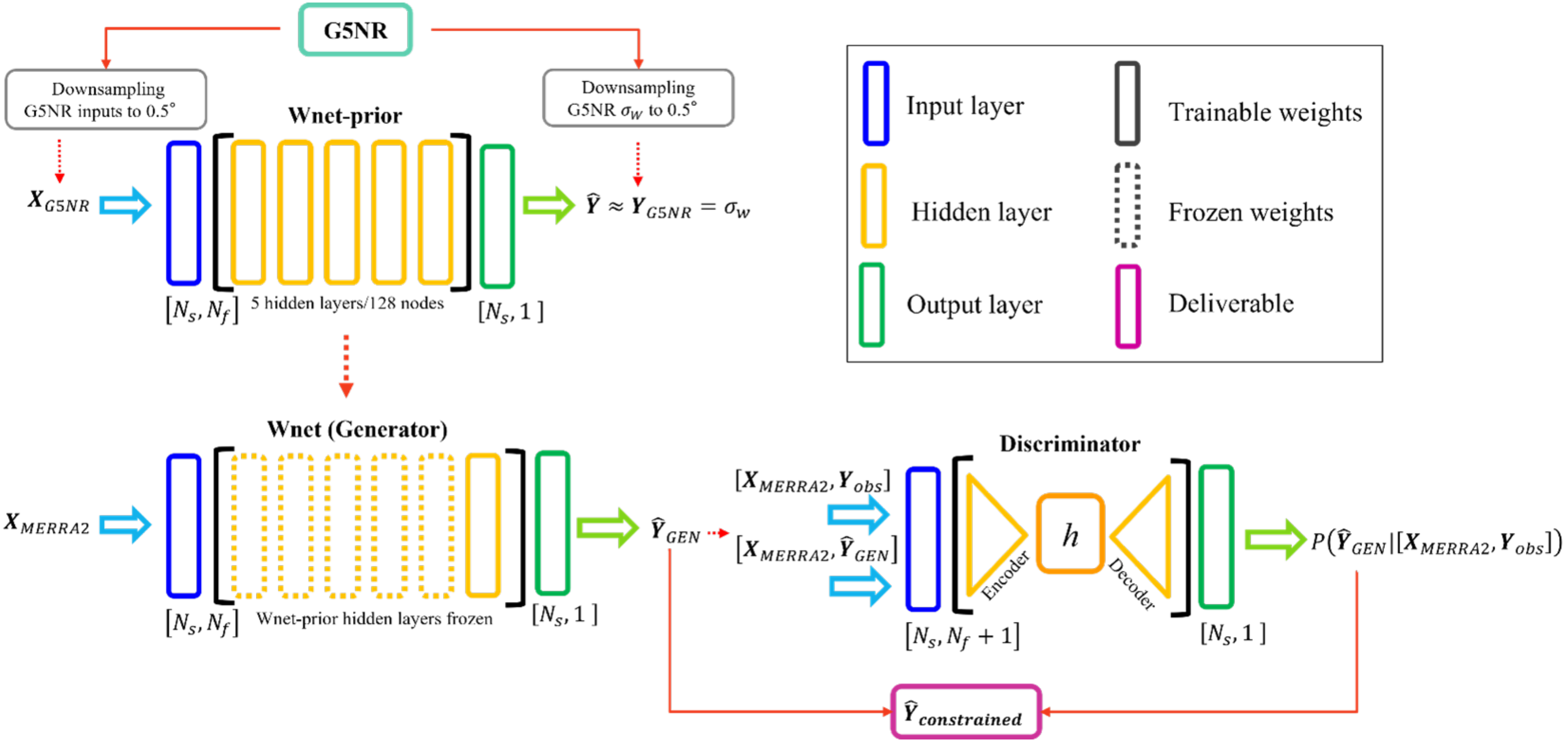
Conclusions

- A neural network was developed to parameterize subgrid scale variability in W in climate models.
- Using observational data during training was critical for the parameterization performance.
- Training using adversarial algorithms resulted in a parameterization able to reproduce the observed statistics of σ_W.
- The ANN inherits spatial structure from GCRM simulations not obvious in the observational data.



Barahona et al. Submitted.

Artificial Neural Network Development



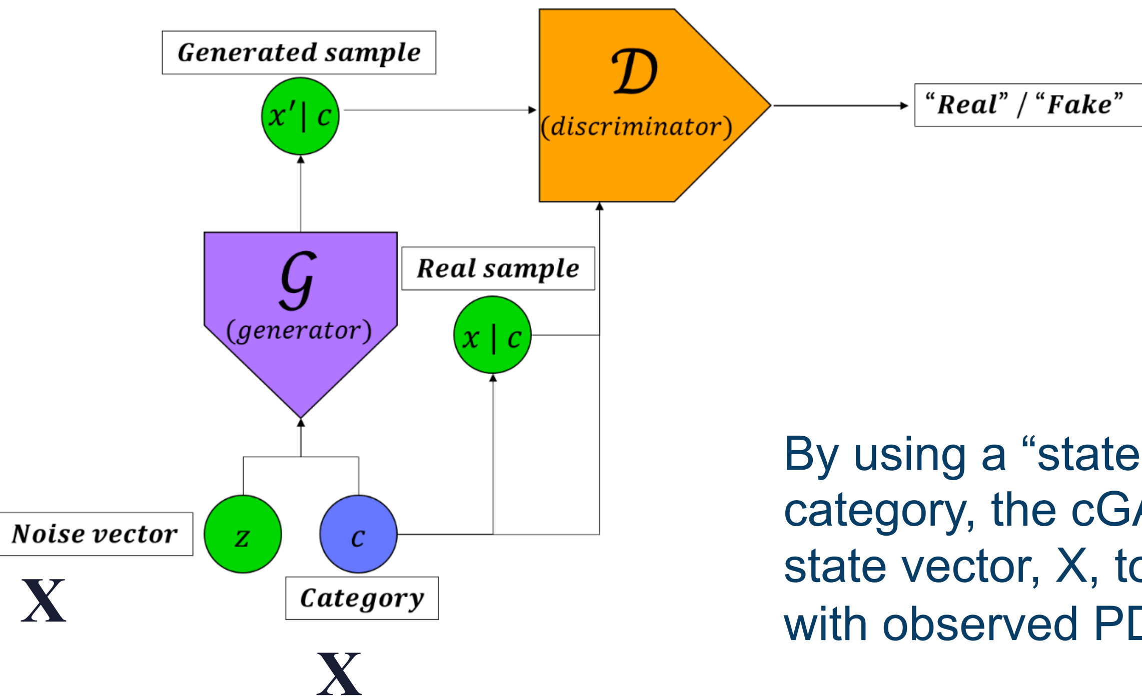
Barahona et al. *AIES* (2023). *Submitted.*

Accounting for Error through Adversarial Training

The PDF of $\boldsymbol{\sigma_w}$ is more robust to experimental and sampling error than individual measurements:

- Even a few observations may represent the PDF
- Non-systematic error cancels out in the PDF

Parameterization must aim to reproduce the PDF-> Generative Adversarial Networks (GANs)



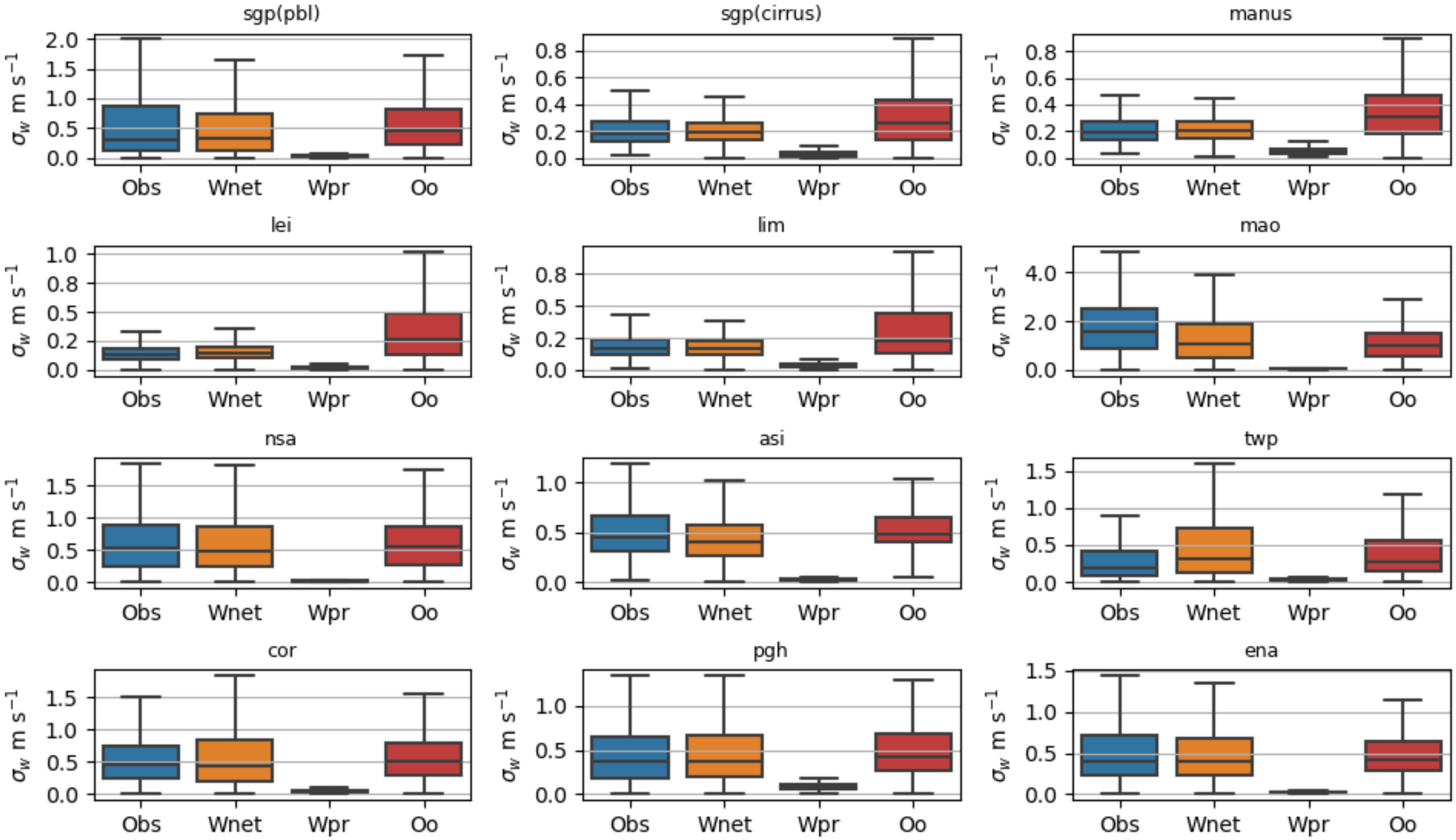
Conditional GAN

By concatenating labelled data to the target, it is possible restrict the PDF within a particular reference class

By using a "state vector" **X**, as both input and category, the cGAN yields a regression from the state vector, X, to the target ($\boldsymbol{\sigma_w}$) while conforming with observed PDF.

Park et al. Electronics 2020, 9, 688;
doi:10.3390/electronics9040688

Results: ANN models at all sites

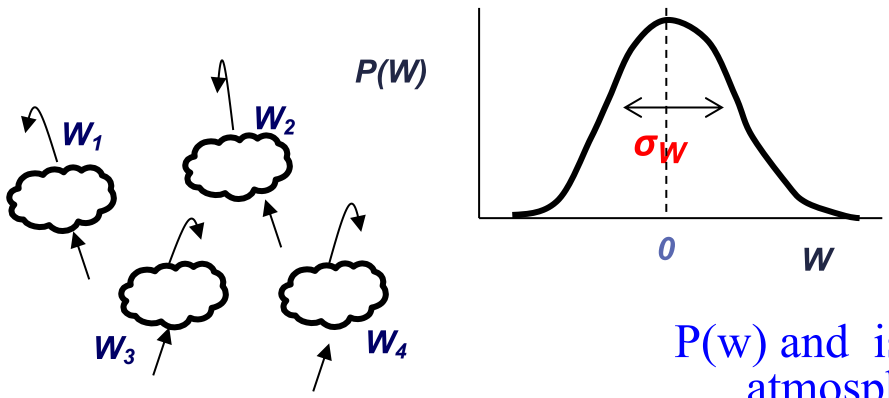


Vertical Wind

Vertical air motion drives important atmospheric processes, like the formation of cloud droplets and ice crystals, and atmospheric boundary layer mixing.

Atmospheric models typically cannot resolve the dynamics of air parcel ascent

Small-scale wind fluctuations are thus characterized by a subgrid distribution of vertical wind velocity, ***W*****,** with standard deviation σ_W .



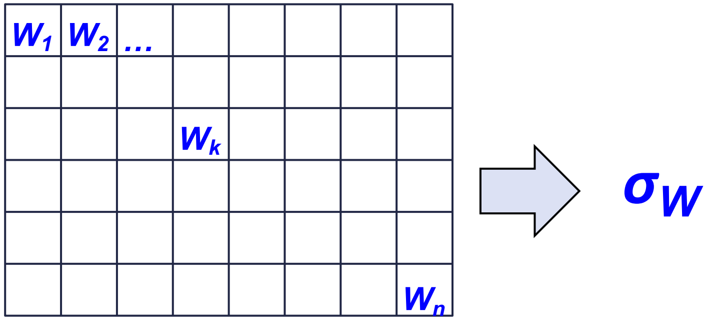
Grid cell (25-100 km)

P(w) and is very uncertain in atmospheric models

Parameterizing σ_w

Theory: Physically based and consistent. Slow to develop with many unconstrained parameters. Could be expensive.

Explicit: Sub-kilometer scale models
- Physically-based.
- Very expensive. Unfeasible for operational forecasts or long-term climate prediction but p**rovide excellent data for training.**
- **Biases still exist. Can't resolve all scales.**



High-res downsampling

Observational Data: W and σ_w can be retrieved using Doppler radar and lidar.
- Robust and fast.
- Constrained by observations
- Limited domain.
- **Experimental error could be significant**.