

Significant trend in vertical wind velocity variability revealed by machine learning

**Donifan Barahona¹, Katherine Breen^{1,2}, Minghui Diao³,
Derek Ngo³, Flor Vanessa Maciel⁴, Ryan Patnaude⁵**

1- Global Modeling and Assimilation Office, NASA Goddard Space Flight Center, 8800 Greenbelt rd, Greenbelt, MD, 20770, MD, United States.

2- Morgan State University, Baltimore, MD, United States.

3- Department of Meteorology and Climate Science San Jose State University, San Jose, CA, United States.

4- Department of Atmospheric and Oceanic Sciences, University of California, Los Angeles, CA, United States.

5- Department of Atmospheric Science, Colorado State University, Fort Collins, CO, United States

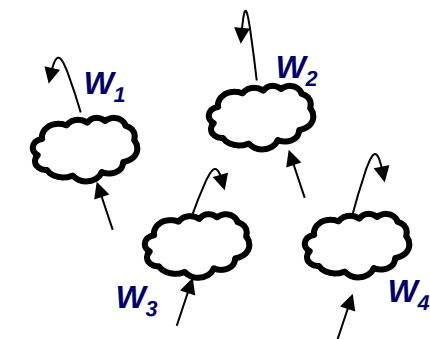
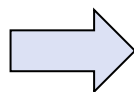
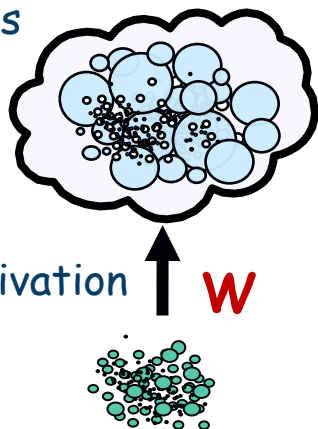
Vertical Wind and Cloud Formation

Vertical air motion drives the formation of cloud droplets and ice crystals.

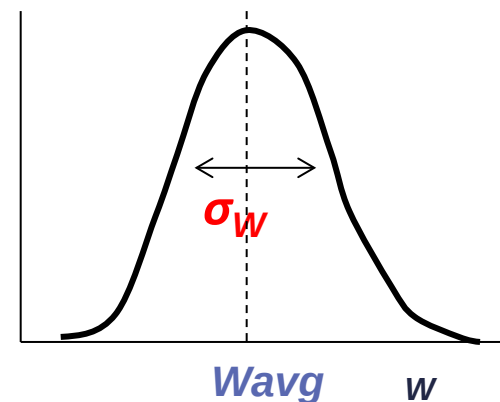
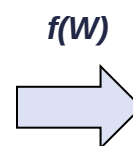
Cloud droplets, ice crystals

Activation

W



Grid cell (25-100 km)



RH, T, CCN, INP

$\Delta X < 1 \text{ km}$

$$\overline{N_a} = \frac{\int_0^\infty N_a(a_1, \dots, a_n, T, p, w) f(w) dw}{\int_0^\infty f(w) dw} \approx N_a|_{W_{\text{sub}}}$$

$W_{\text{sub}} \approx 0.8\sigma_W \equiv \text{Subgrid Updraft, Characteristic } W, \dots$

Parameterizing σ_w

Uncertainty in σ_w impacts cloud formation rates.

Different models for liquid and ice clouds.

CLUBB:
Consistent representation based on joint T, Q and W distributions.

- Complex
- PBL oriented

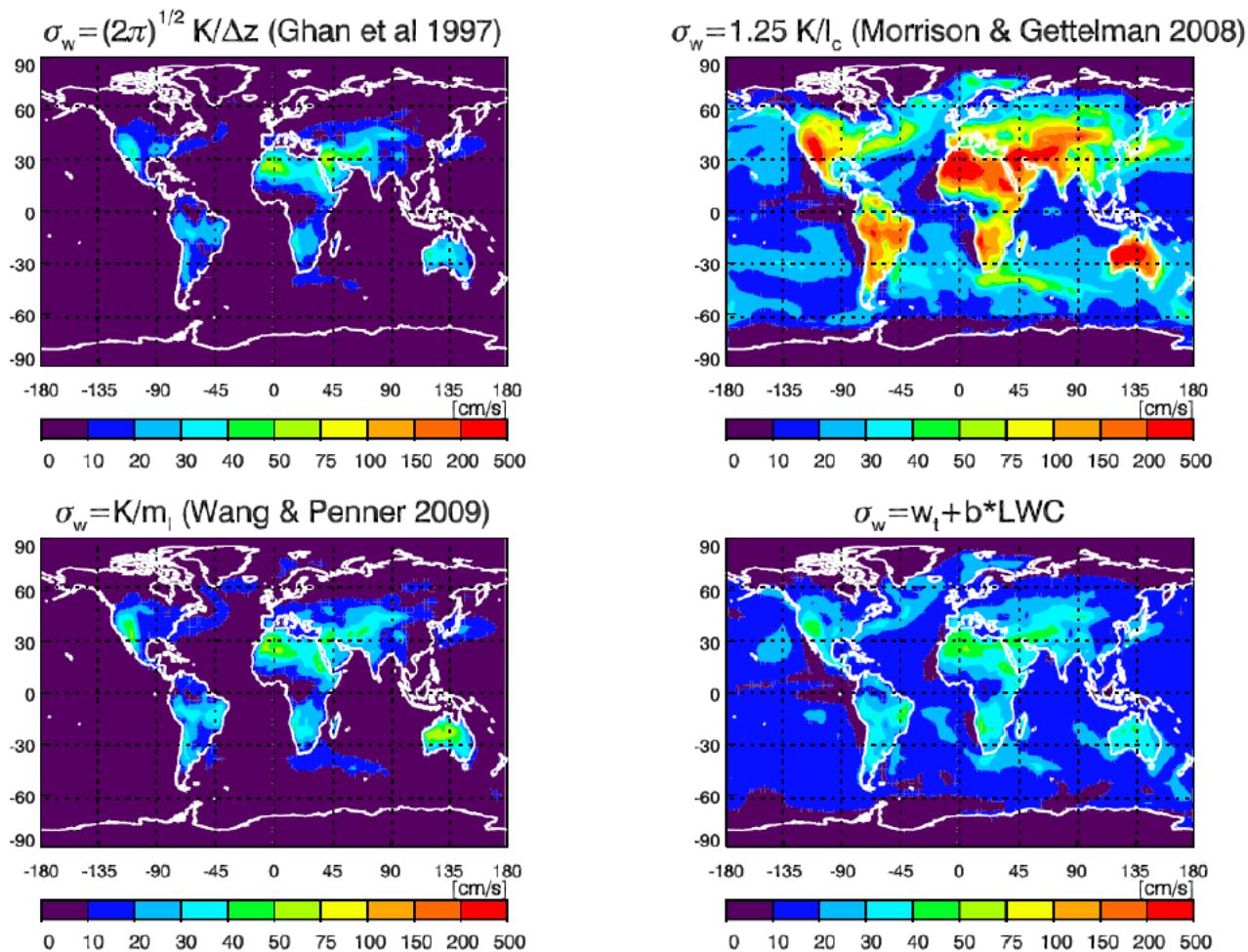


FIGURE 6: Annual averages of the grid-mean σ_w for the lowest five model levels ($\approx 2\text{km}$), calculated with four different parameterizations implemented in CAM-Oslo.

Parameterizing σ_w

“Input updraft velocity fluctuations can explain as much as 48% of temporal variability in output ice crystal number and 61% in droplet number in GEOS-5 and up to 89% of temporal variability in output ice crystal number in CAM5.1.” (Sullivan et al. 2016)

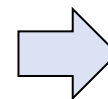
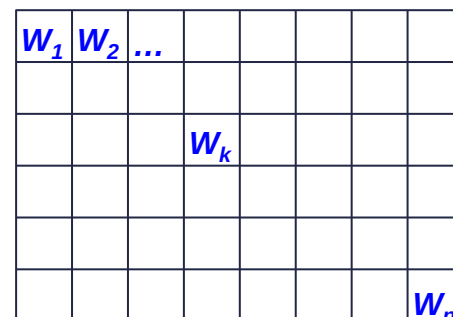
In radiative terms “the breadth of the range [of the effect of σ_w on the AIE] is 0.4 W m^{-2} , ...comparable to a substantial fraction of the total diversity of current aerosol forcing estimates.” (West et al. 2014).

There is currently not a global, independent estimate of σ_w that allows us to reduce such an uncertainty

Estimating σ_w

Explicit: Kilometer-scale models

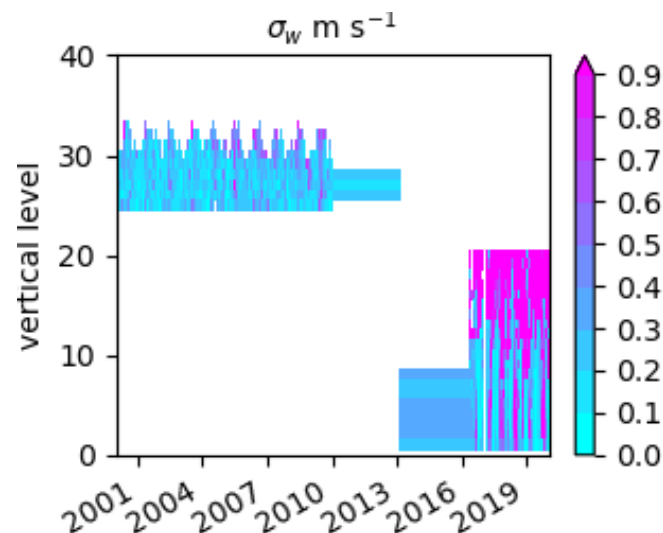
- Physically-based.
- Very expensive. Unfeasible for operational forecasts or long-term climate prediction but **provide excellent data for training.**
- **Biases still exist. Can't resolve all scales.**

 σ_w

High-res downsampling

Observational Data: W and σ_w can be retrieved using Doppler radar and lidar.

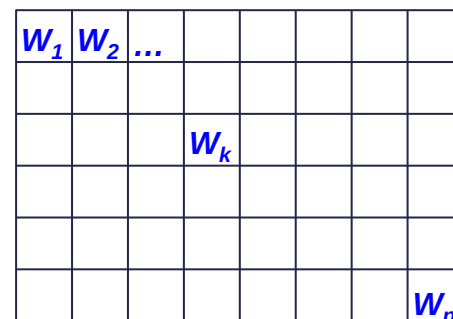
- Constrained by observations
- Very Limited domain. Satellite-based estimates are crude.
- Experimental error could be significant.



Estimating σ_w

Explicit: Kilometer-scale models

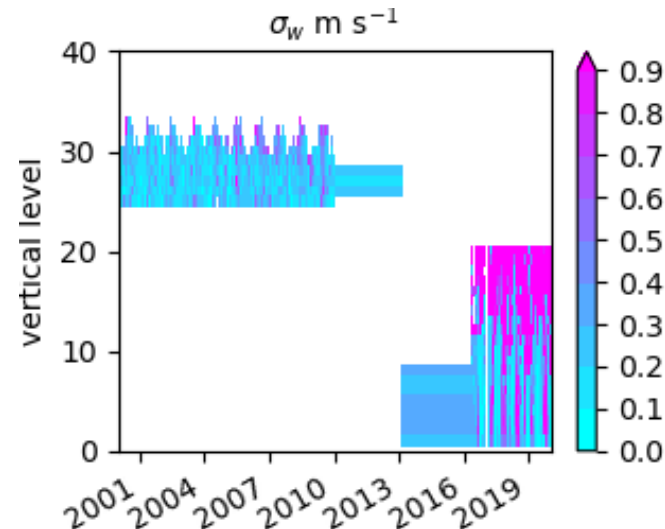
- Physically-based.
- Very expensive. Unfeasible for operational forecasts or long-term climate prediction but **provide excellent data for training.**
- **Biases still exist. Can't resolve all scales.**



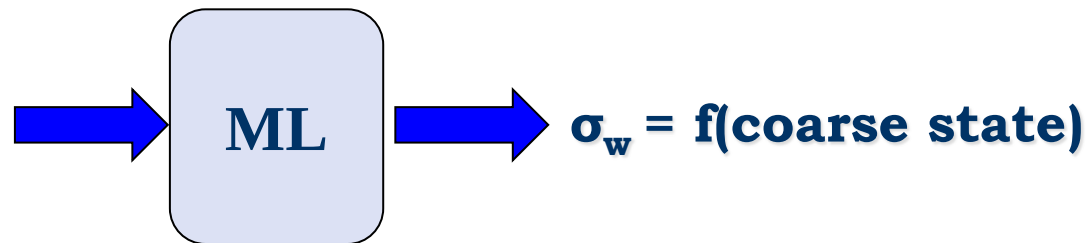
High-res downsampling

Observational Data: W and σ_w can be retrieved using Doppler radar and lidar.

- Constrained by observations
- Very Limited domain. Satellite-based estimates are crude.
- Experimental error could be significant.



**Experimental Observations,
Kilometer-Scale simulations
Climate Reanalysis**

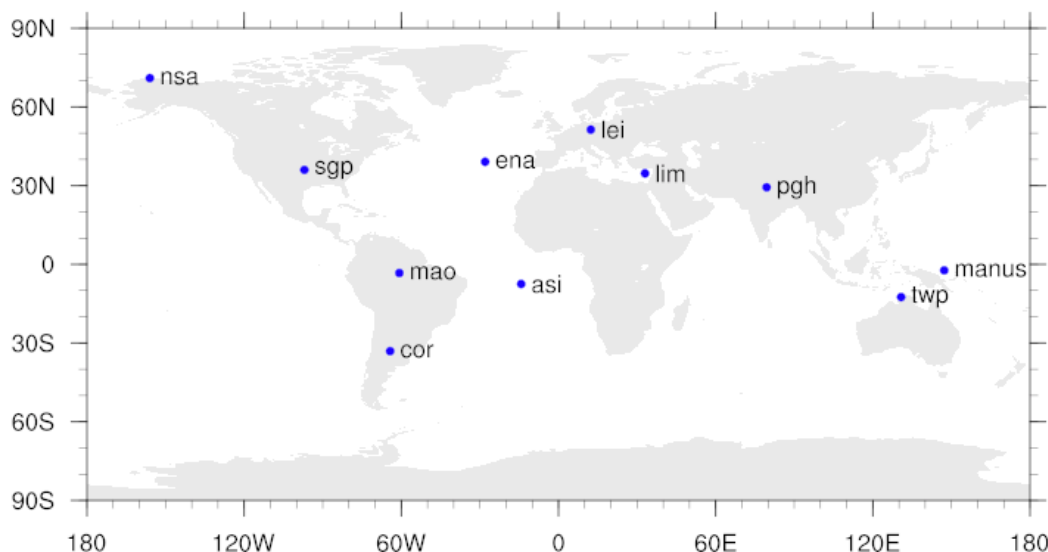


Data and Methods

GEOS Nature run (G5NR): 2 years of free-running, non-hydrostatic 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:



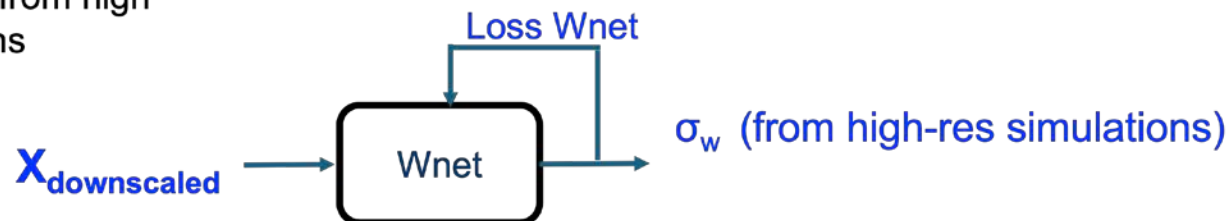
W retrieved using Doppler radar and lidar, at 11 different sites around the world. Mostly PBL sites, 4 cirrus, 1 convective

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

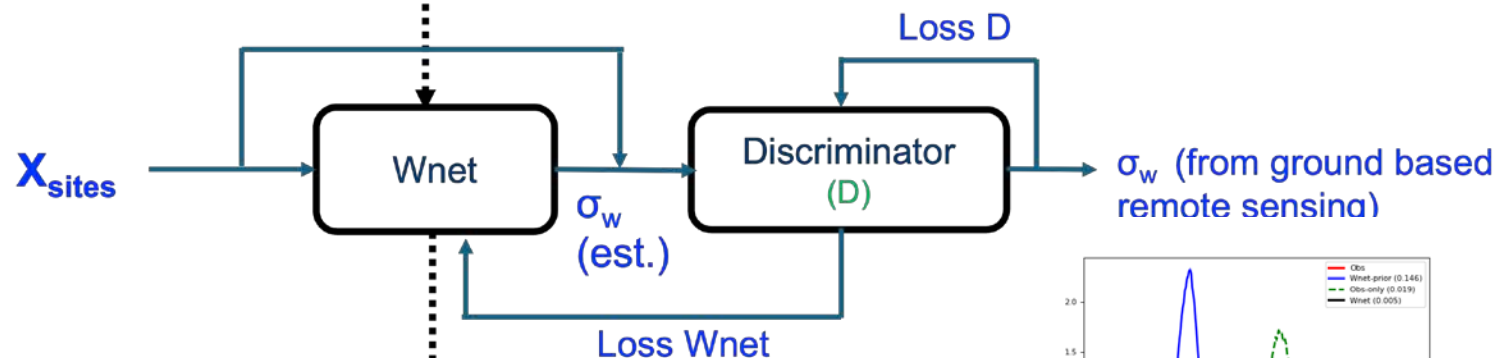
About 100 years of data on **W** at temporal resolution between 20 s and 5 min.

Developing a Global Estimate of σ_w

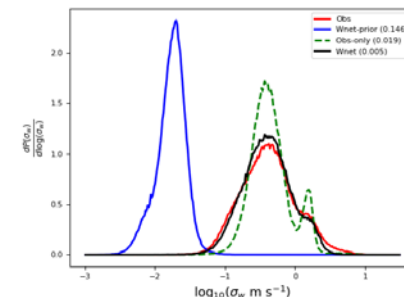
1- Surrogate model from high resolution simulations



2- Probabilistic refinement using observational data at ground sites

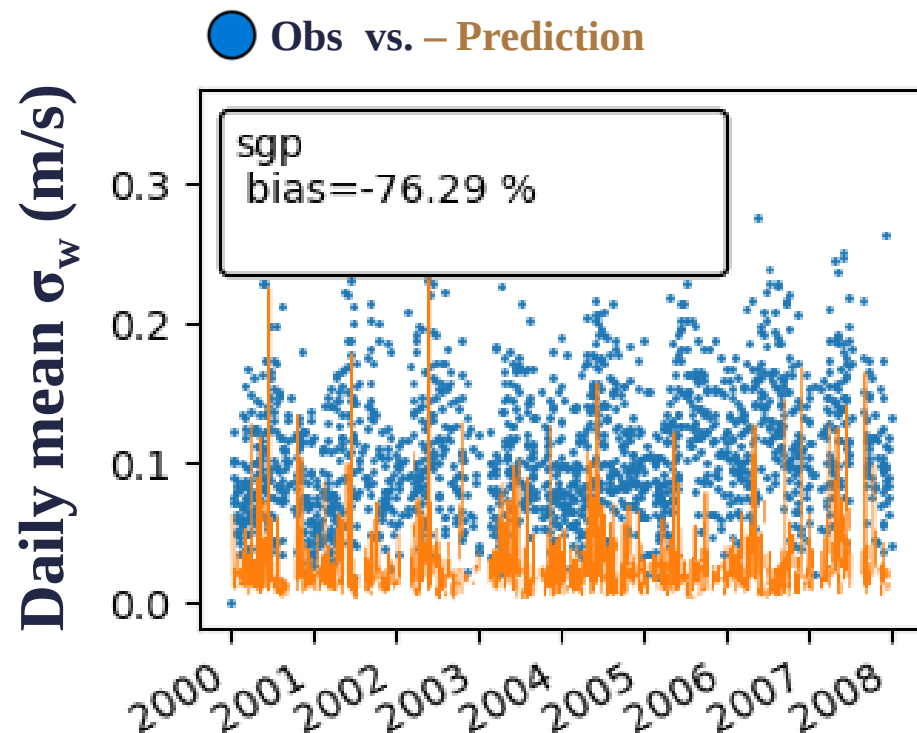
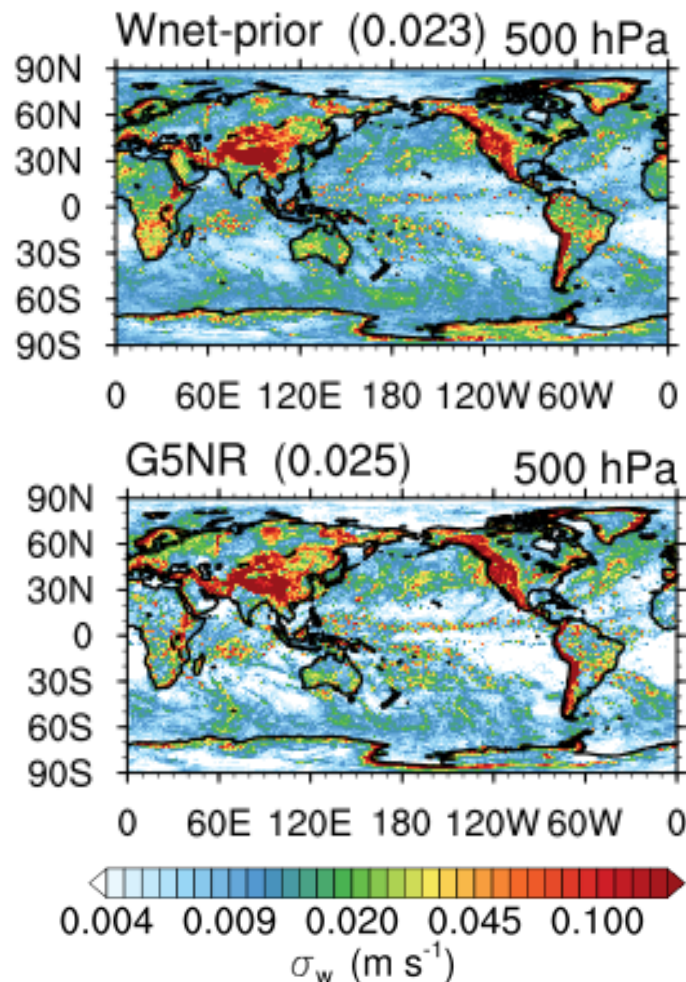


3- Inference driven by reanalysis data



X = Atmospheric State (i.e., T, P, Winds, ...) **at coarse resolution**

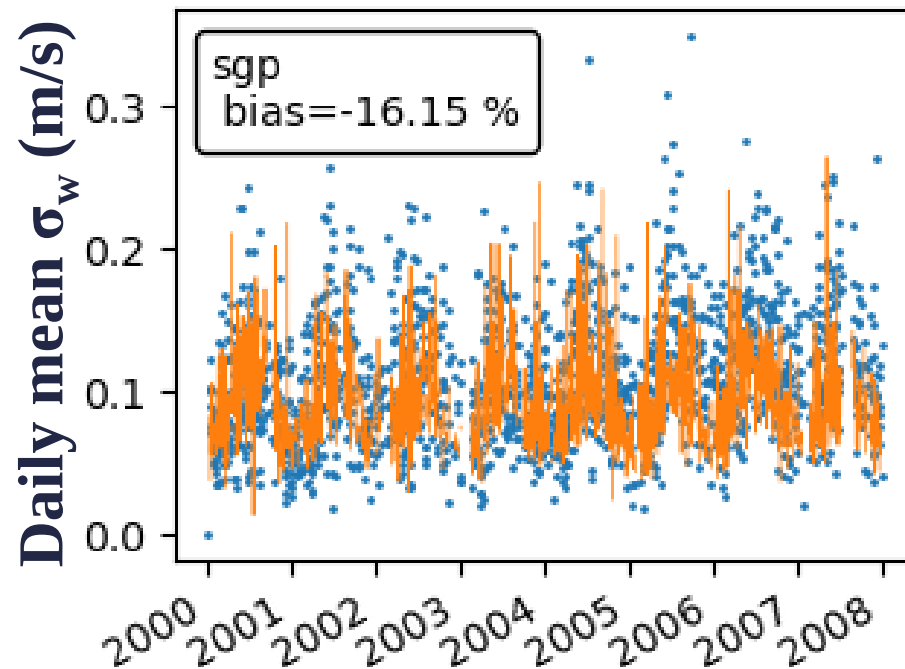
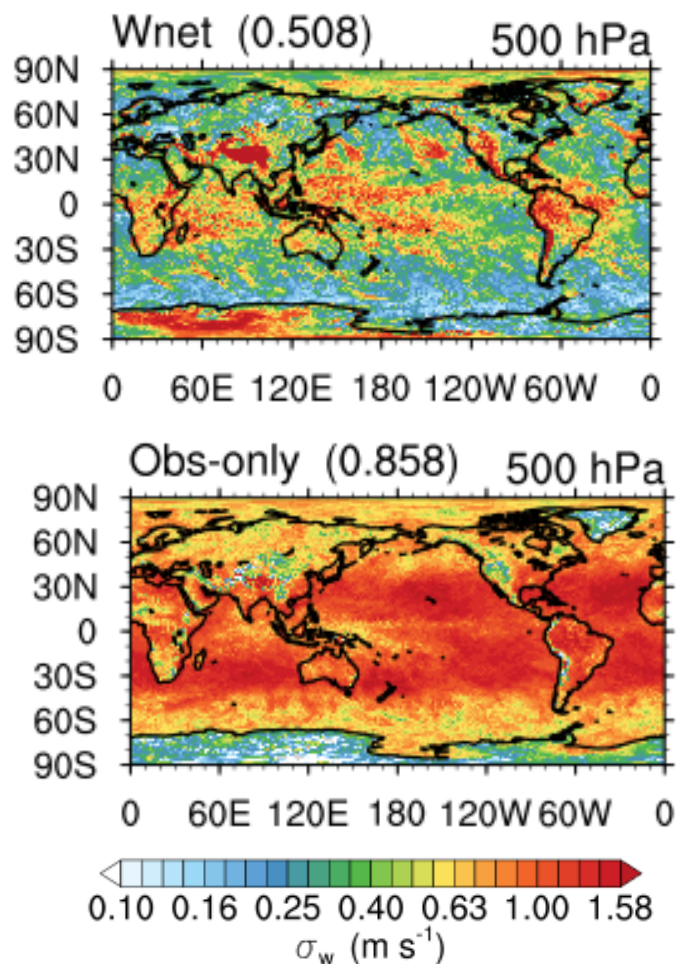
Results: Wnet-prior reproduces the GSRM



The NN reproduces the GSRM output within 0.005 m s⁻¹.

However, it tends to underestimate variability, reflecting the bias of the high-resolution simulations.

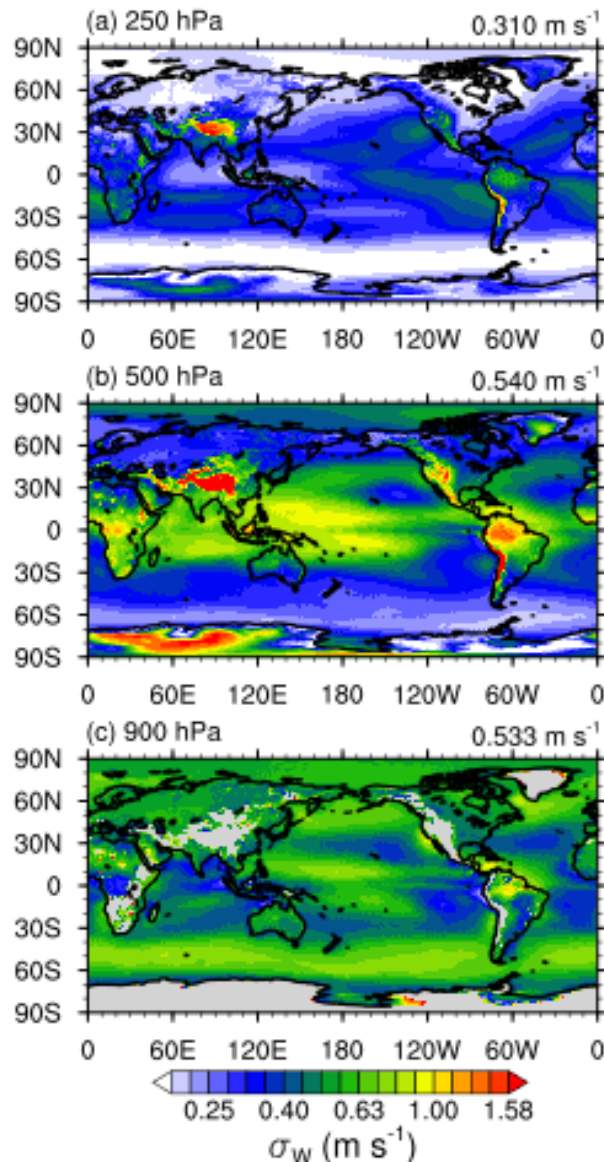
Results: GAN-Trained NN Parameterization Achieves High Accuracy



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”).

Global Estimate of σ_w



We developed a long-term data set on σ_w spanning 42 years, on 3-hourly intervals, at half-degree, from 1 hPa to 1000 hPa.

<https://portal.nccs.nasa.gov/datashare/Wstd-MERRA2/>

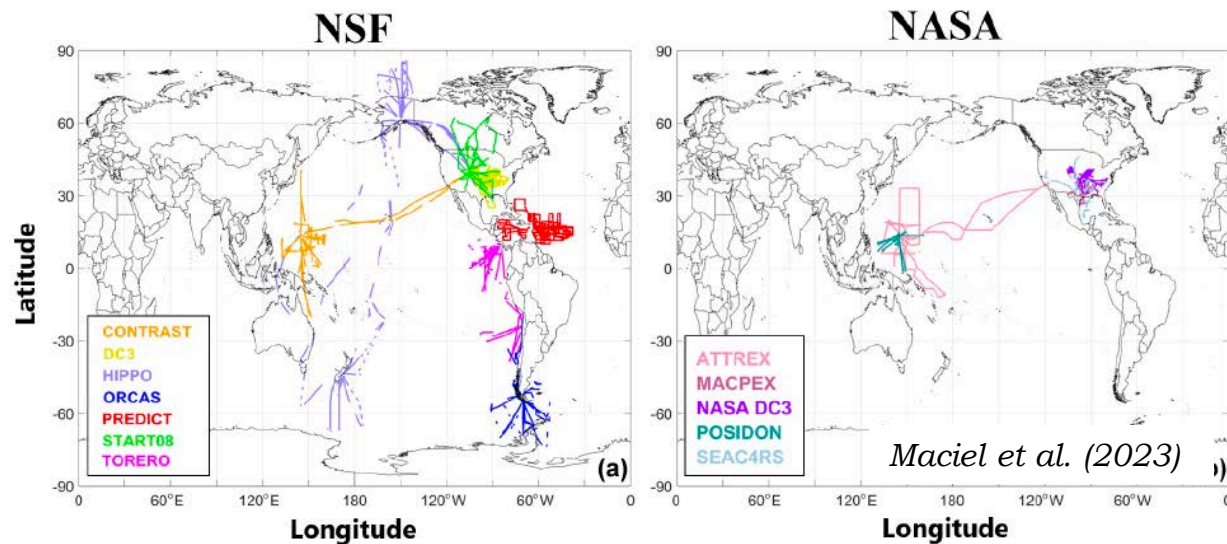
Familiar patterns:

- High variability in convective regions and storm tracks.
- Turbulence induced by orography and the jet stream.
- In general σ_w decreases with altitude.

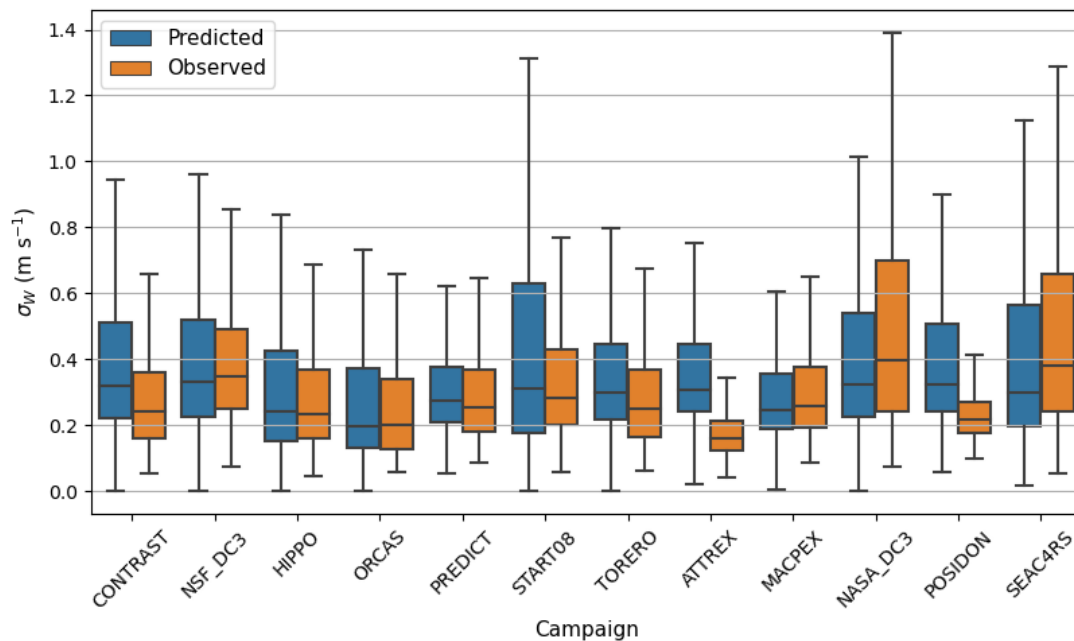
Global mean σ_w derived from MERRA-2 for 1980-2023.

Barahona et al. *Under review*.

Evaluation against Aircraft Data

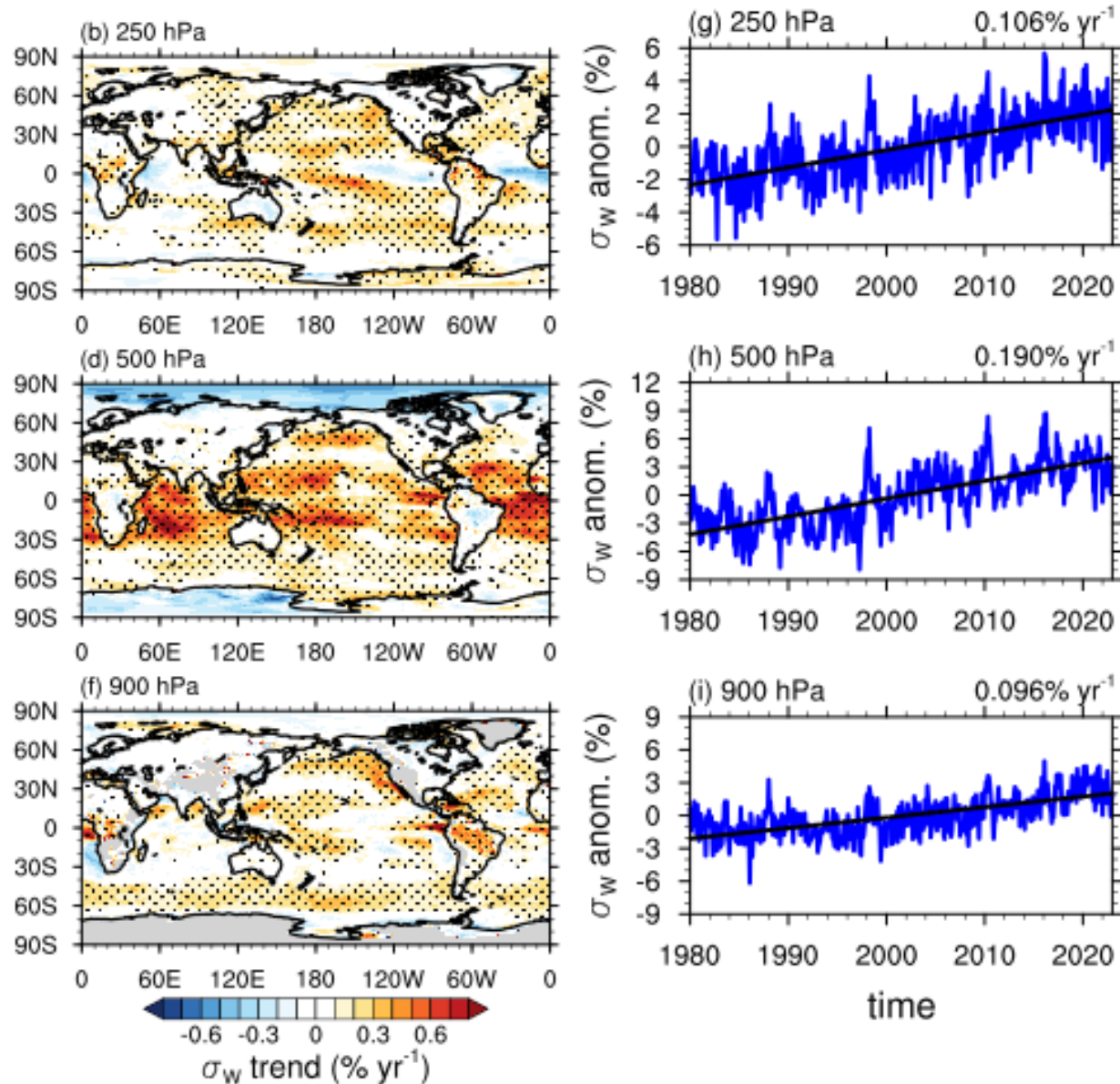


MERRA2-derived σ_w was interpolated along flight tracks for different field campaigns.



Rigorous test:

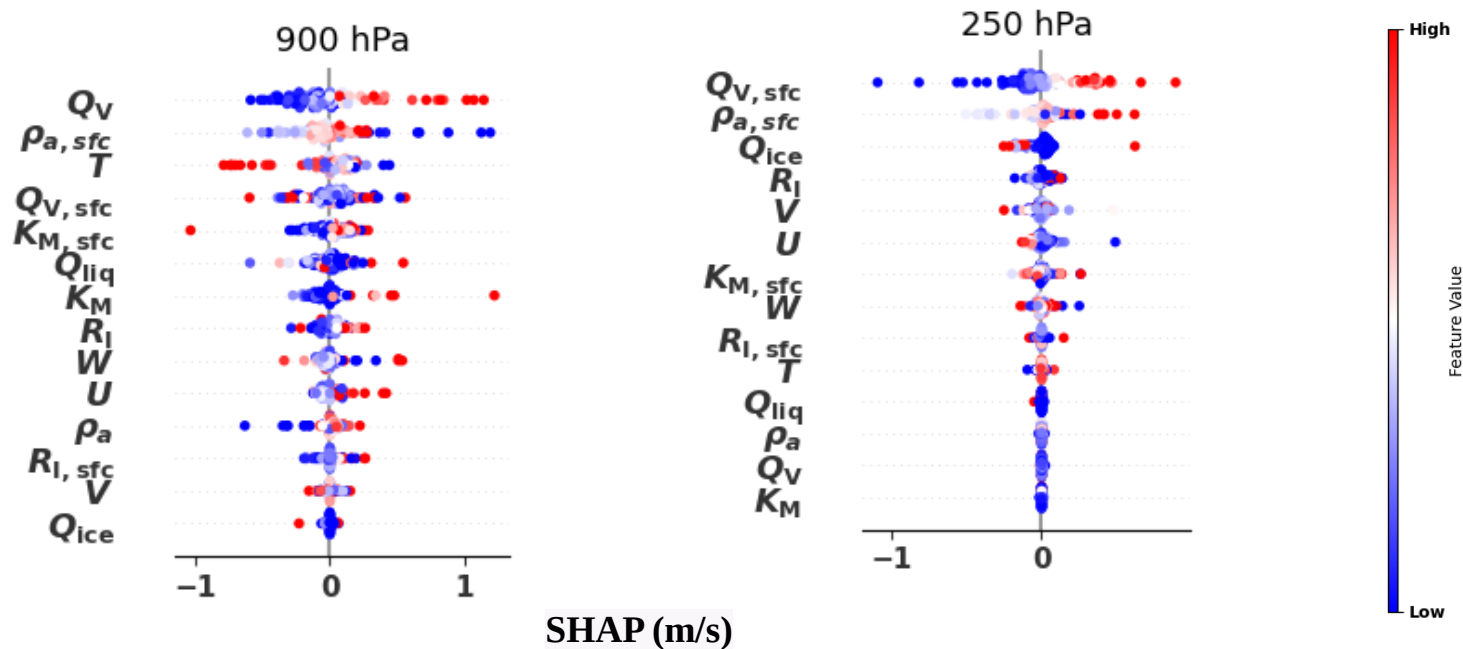
Data not used during training, and σ_w is measured directly instead of retrieved by remote sensing.

Trends in σ_w 

Significant trends in σ_w emerged during the last 40 years, with a general increase over the tropics/subtropics and decrease towards high latitudes.

Implies enhanced turbulence in the lower latitudes.

Explainable Machine Learning Analysis



The SHAP value indicates the sensitivity of σ_w to a given input. Red on the right means positive correlation, inputs are organized by importance.

Surface water vapor and air density are the main factors at all levels. Horizontal winds become important at higher altitude and in the Arctic. These are linked to convective activity and jet-stream strength.

Evolving meteorological patterns, like jet-activity and convection, have systematically impacted the intensity and frequency of small scale processes, such as turbulence, hence cloud microphysics.

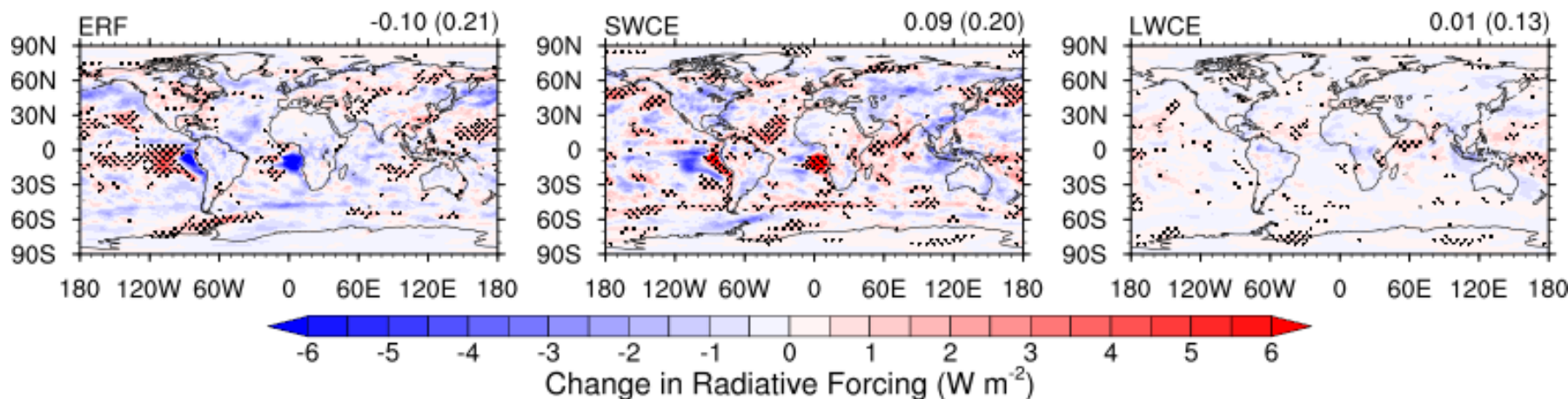
Evolving σ_w as a forcing

Due to its impact on cloud microphysics σ_w affects the radiative properties of clouds.

Climatologies of σ_w were calculated using the output of two pre-run simulations representing preindustrial (1850) and present day (mean 2010's) conditions.

Two identical simulations with constrained winds, sst, and aerosol forcing were run differing only in the prescribed σ_w climatologies.

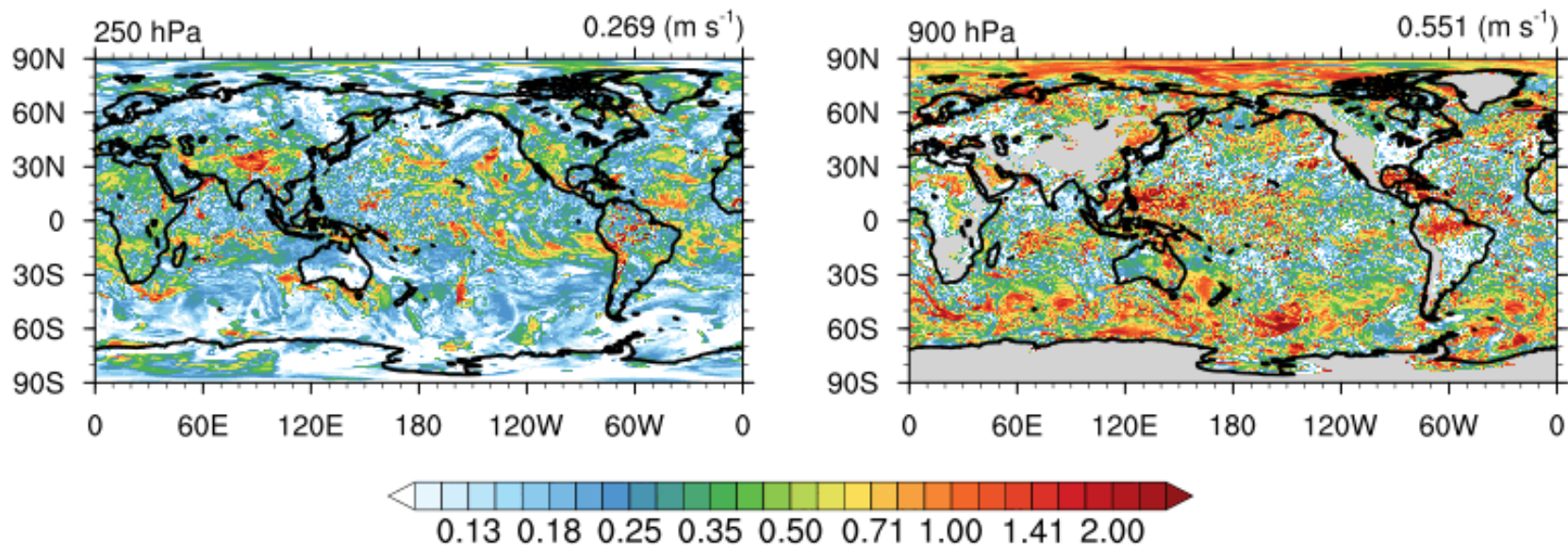
Evolving σ_w represents a radiative forcing of about -0.1 W m^{-2} , mostly through the shortwave.



Conclusions

- An observation-constrained neural network, Wnet, was developed to estimate subgrid scale variability in W using the coarse meteorological state.
- A long term data set was developed, revealing significant trends in σ_w over the last four decades, likely driven by the evolution of water vapor, convection and jet stream strength
- The connection between large scale climate patterns and cloud microphysics through σ_w may represent an indirect radiative effect on climate, slightly counterbalancing greenhouse warming.

sigmaW (m/s) 01-Jun 2019 (01H)

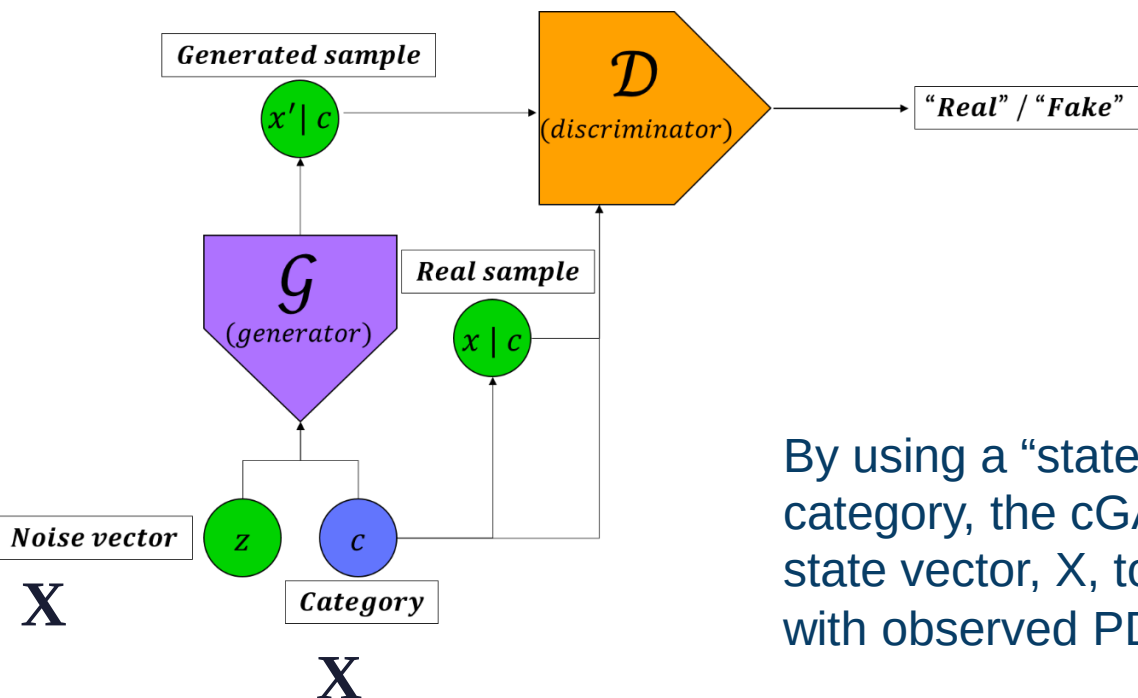


Accounting for Error through Adversarial Training

The PDF of σ_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)

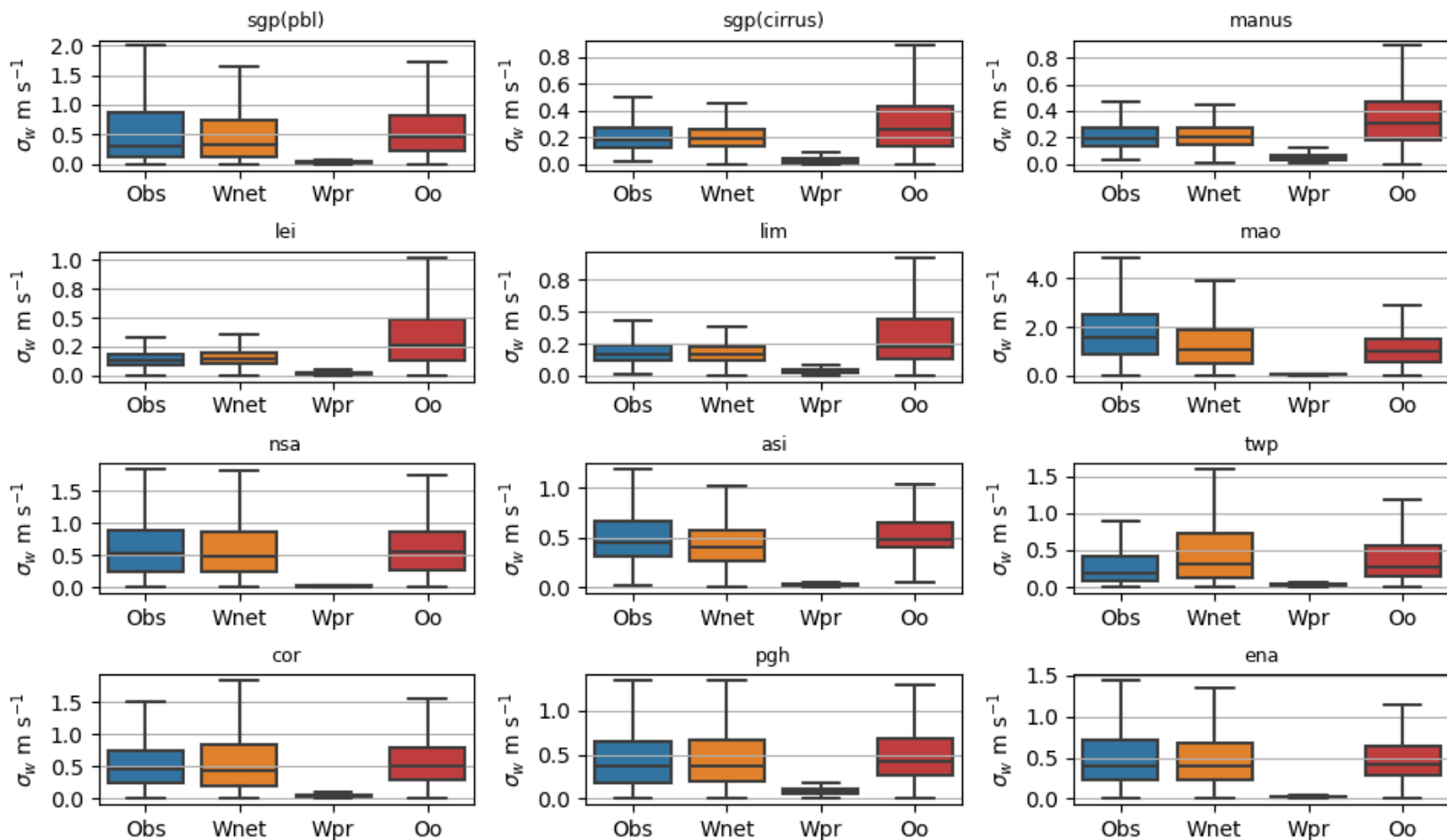


By concatenating labelled data to the target, it is possible to restrict the PDF within a particular reference class

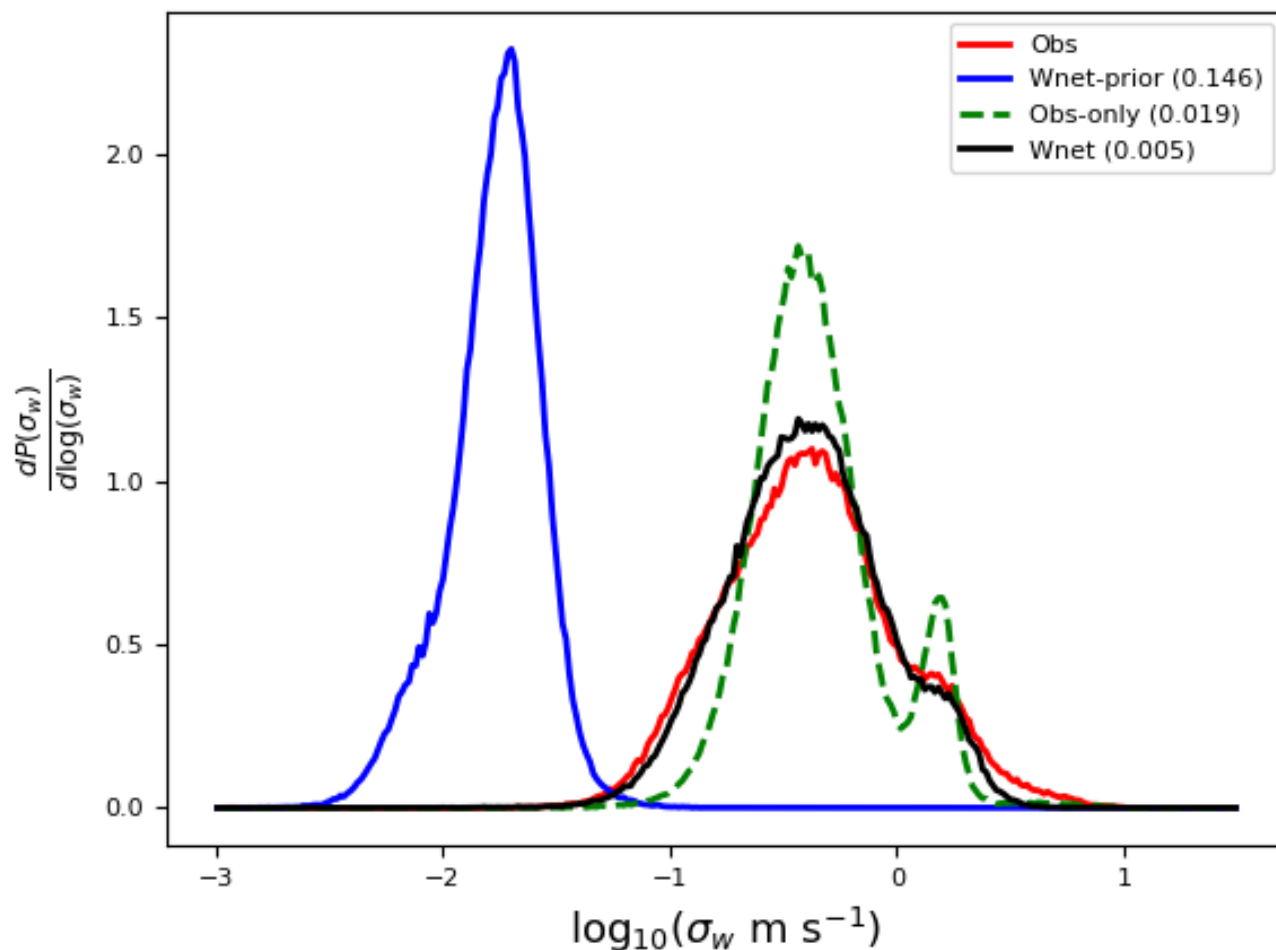
By using a "state vector" X , as both input and category, the cGAN yields a regression from the state vector, X , to the target (σ_w) while conforming with observed PDF.

Conditional GAN

Results: ANN models at all sites



What about the PDF?



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

