Physically and Causally-Informed ML for Climate Modeling

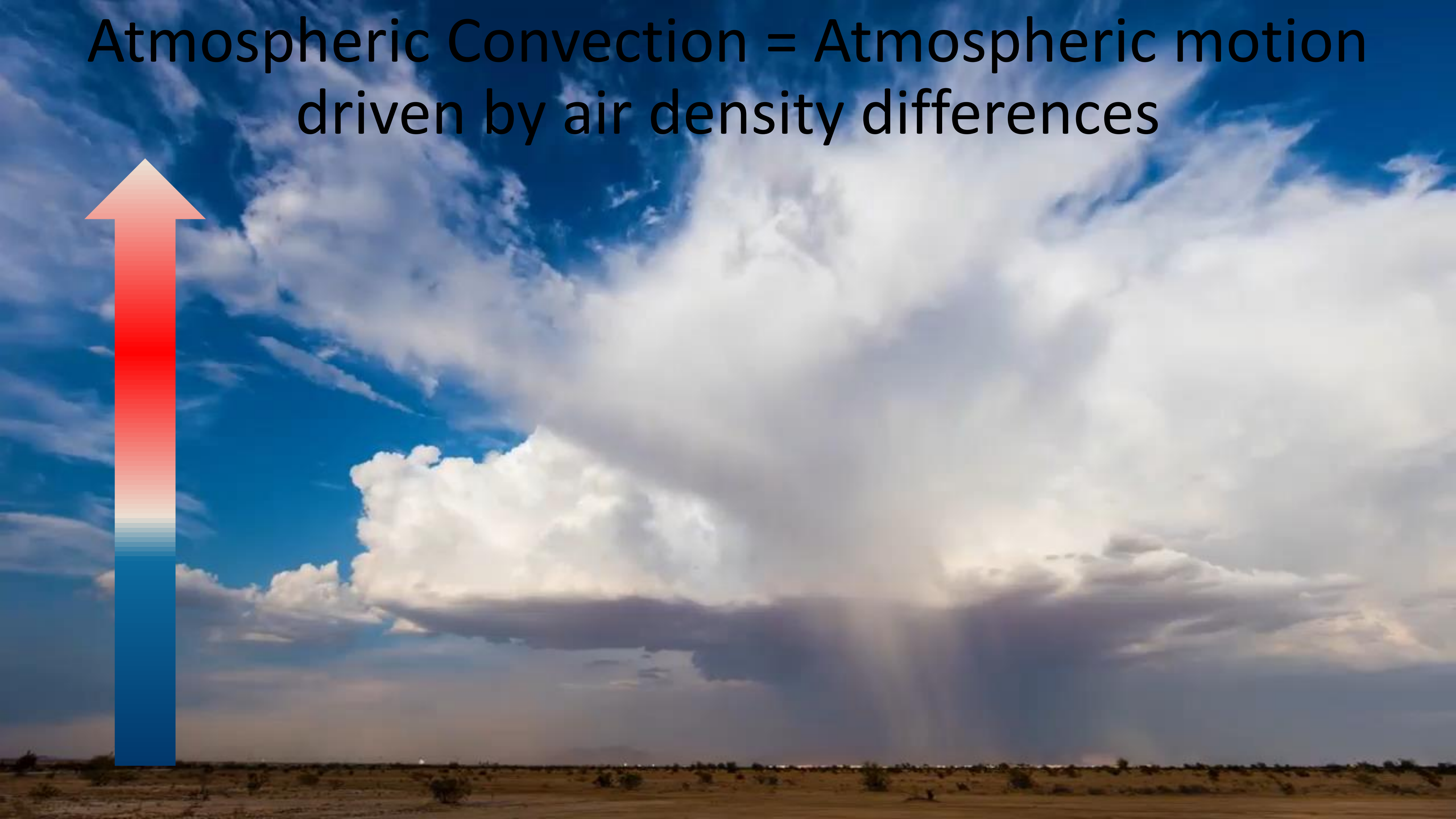
ai4oac2023 : Workshop on AI for Ocean, Atmosphere and Climate Dynamics

11-13 Apr 2023 Brest (France)

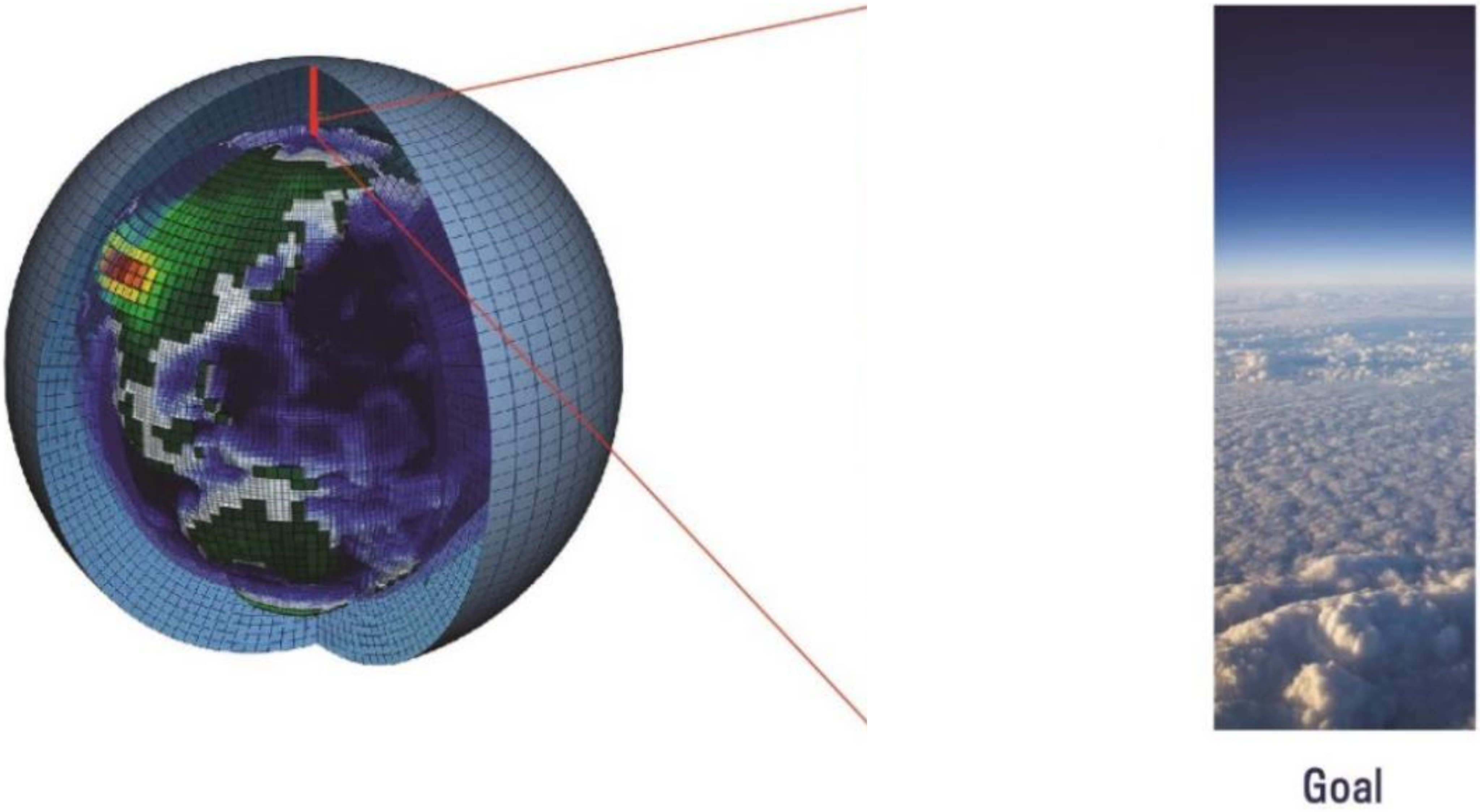
J. Yuval, P. O'Gorman (MIT)
A. Gupta, L. Peng (UCI)
S. Rasp (Climate AI)
F. Ahmed, D. Neelin (UCLA)
N. Lutsko (UCSD)

Tom Beucler (UNIL)
M. Pritchard (UCI),
P. Gentine (Columbia)

F. Iglesias-Suarez
V. Eyring, A. Gerhadus **(DLR)**
J. Runge (DLR, TUB)
S. Ganesh S., F. Tam,
M. Gomez (UNIL)

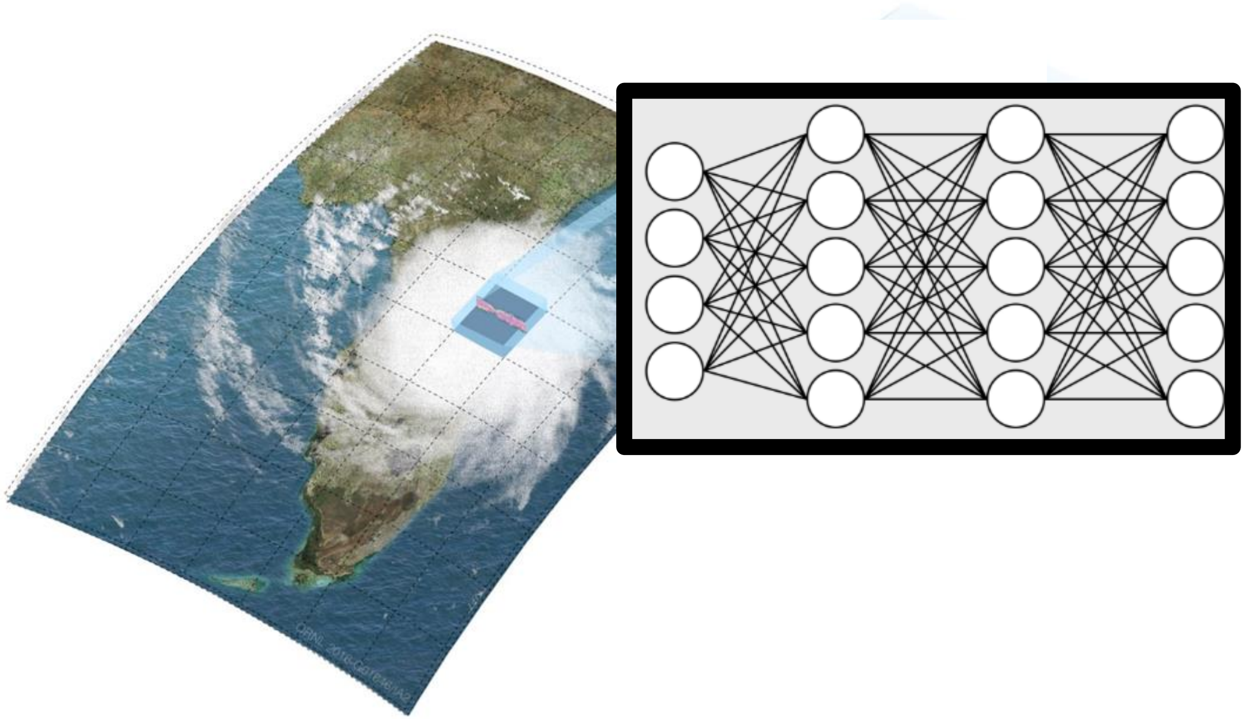

Atmospheric Convection = Atmospheric motion driven by air density differences

Motivation: Largest uncertainties in climate projections from clouds



Source: Zelinka et al. (2020), Meehl et al. (In Review), Gentine, Eyring & Beucler (2020)

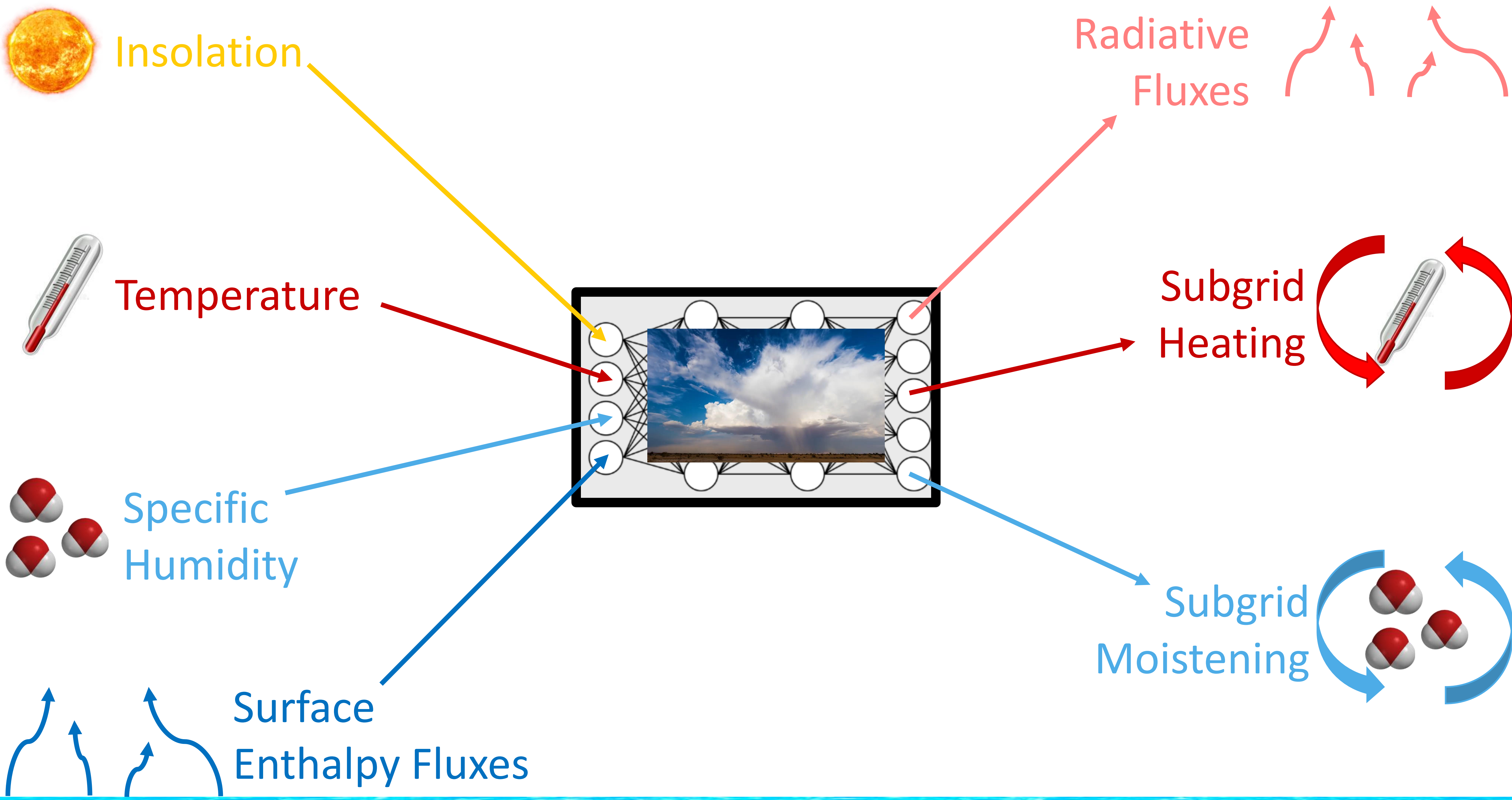
Goal: Machine Learning Subgrid-Scale Thermodynamics



Neural Network:
20 times faster

<u>Setup</u> : Super-Parameterized climate model, Aquaplanet & Earth-like
3 months of data (≈50M samples) for training, Year 2 for validation/test

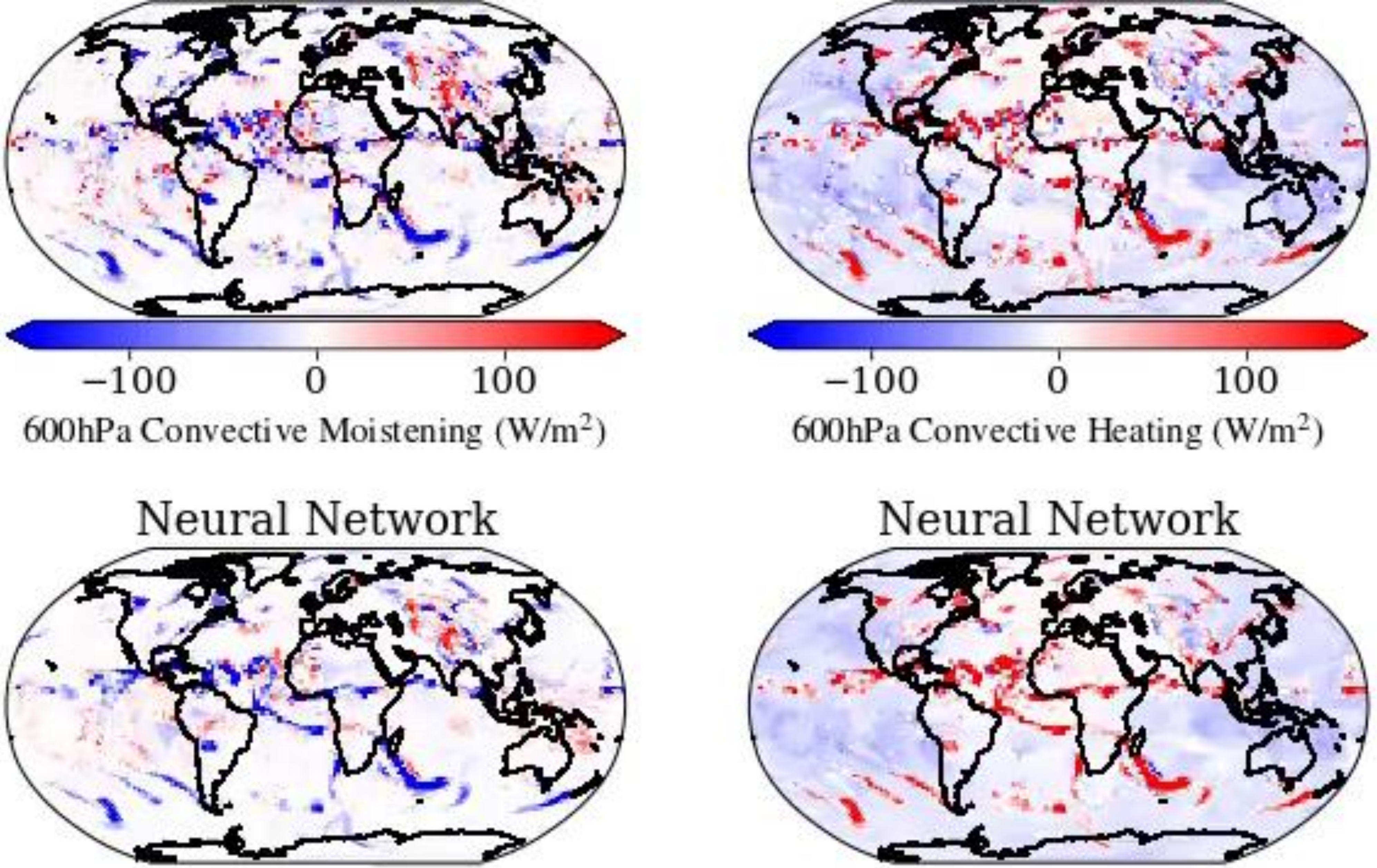
<u>Image source</u>: e3sm.org, <u>Model source</u>: Khairoutdinov et al. (2004)



<u>See</u>: Rasp et al. (2018), Brenowitz et al. (2018,2019), Gentine et al. (2018), Yuval et al. (2020), Krasnopolsky et al. (2013)

Truth
Super-param.
simulation

Prediction
NN
(offline)

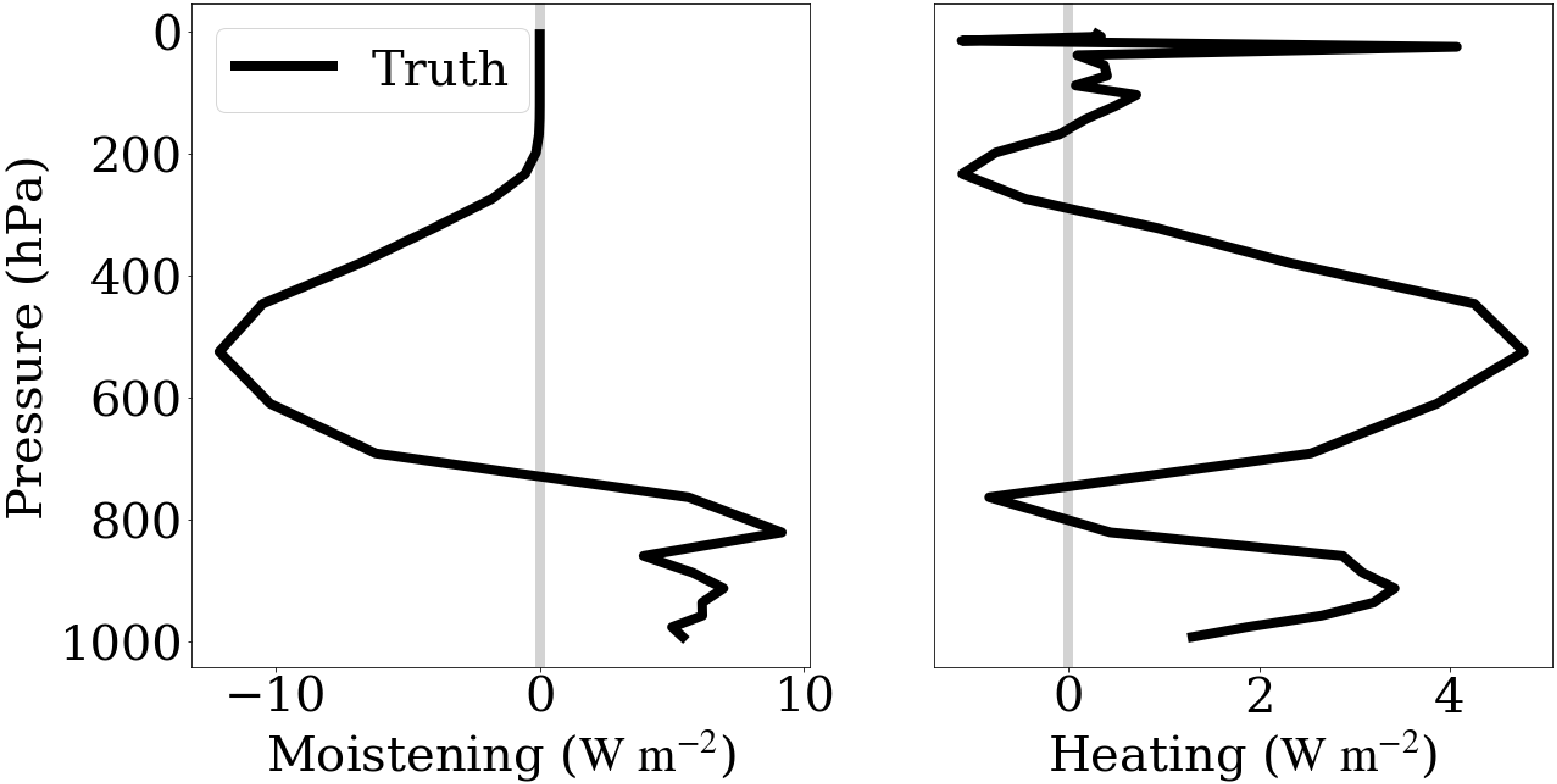


<u>Source</u>: *Mooers, Pritchard, Beucler et al. (2021)*

<u>See</u>: *Rasp et al. (2018), Brenowitz et al. (2018,2019), Gentine et al. (2018), Yuval et al. (2020), Krasnopolsky et al. (2013)*

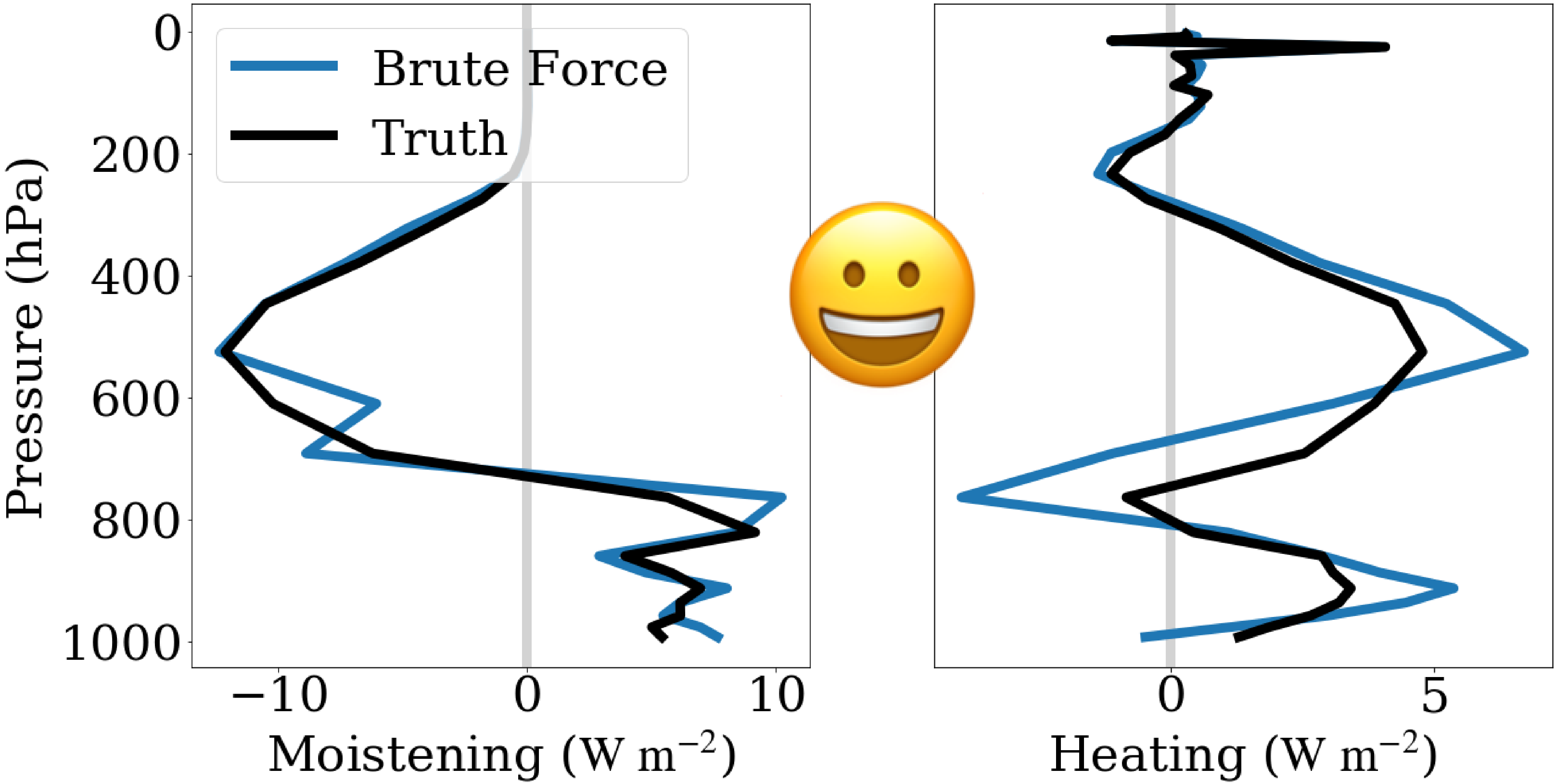
NNs perform well in the climate they were trained on…

Daily-mean Tropical prediction in cold climate



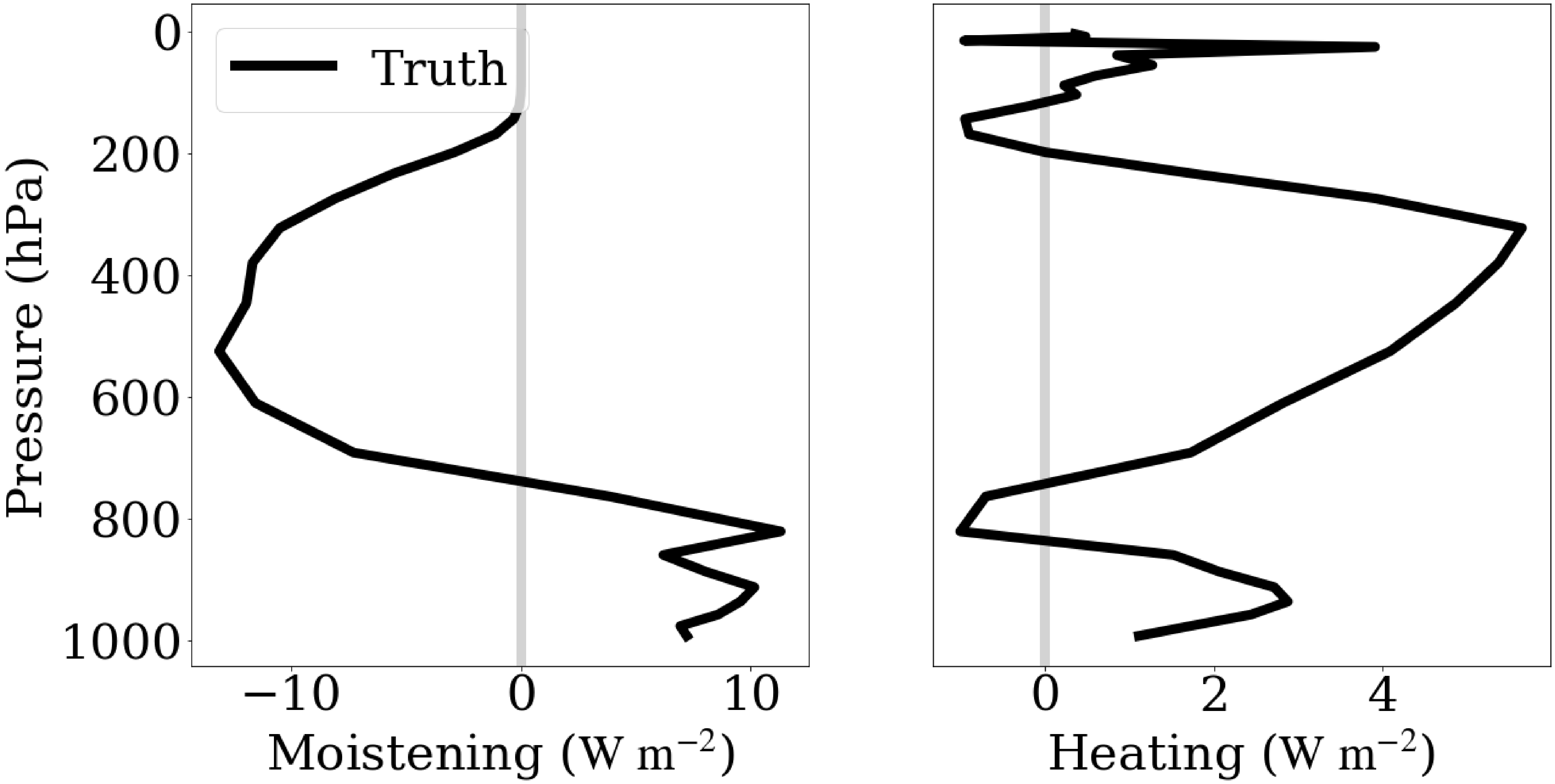
NNs perform well in the climate they were trained on…

Daily-mean Tropical prediction in cold climate



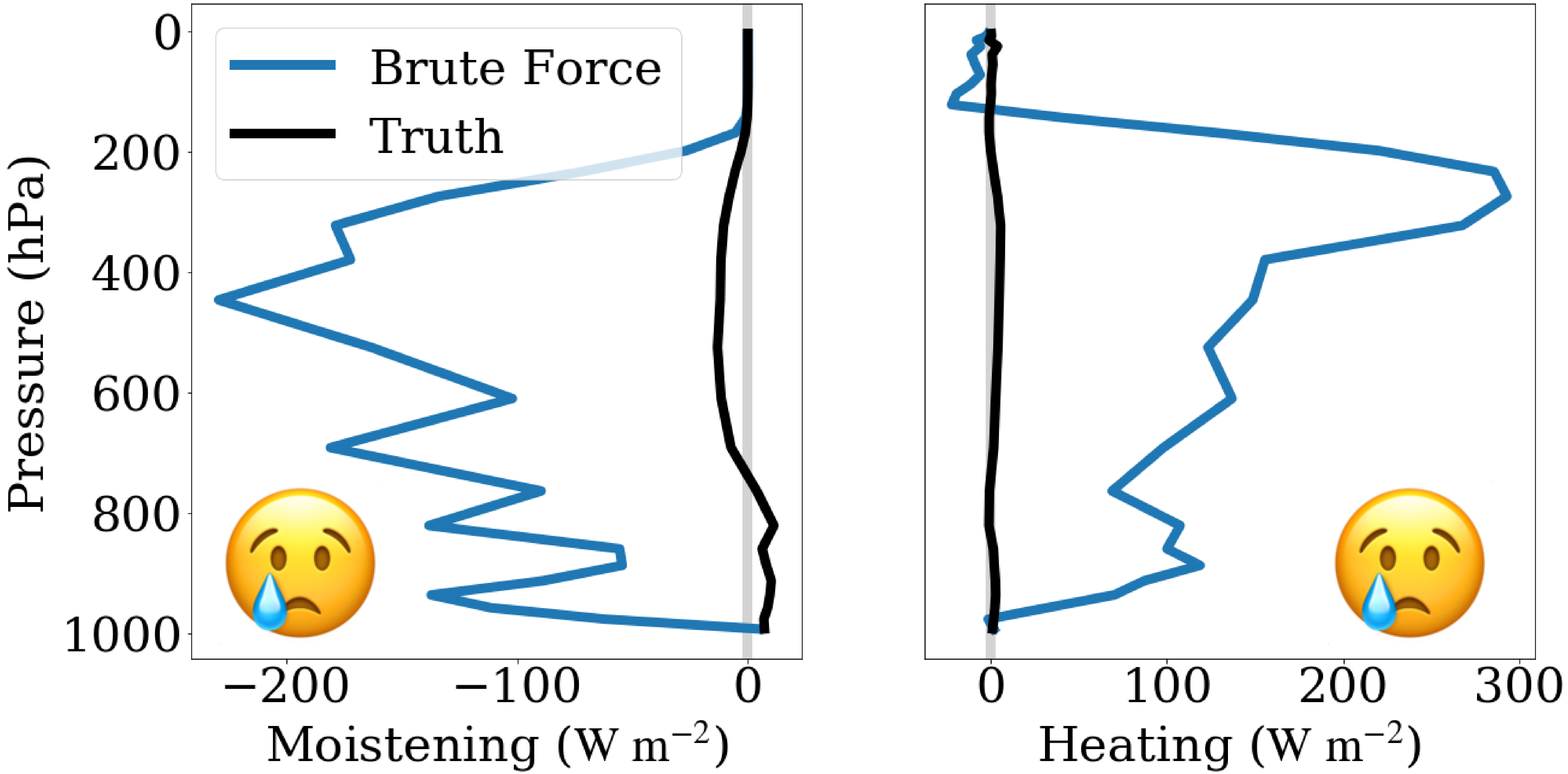
NNs perform well in the climate they were trained on…

Daily-mean Tropical prediction in warm climate



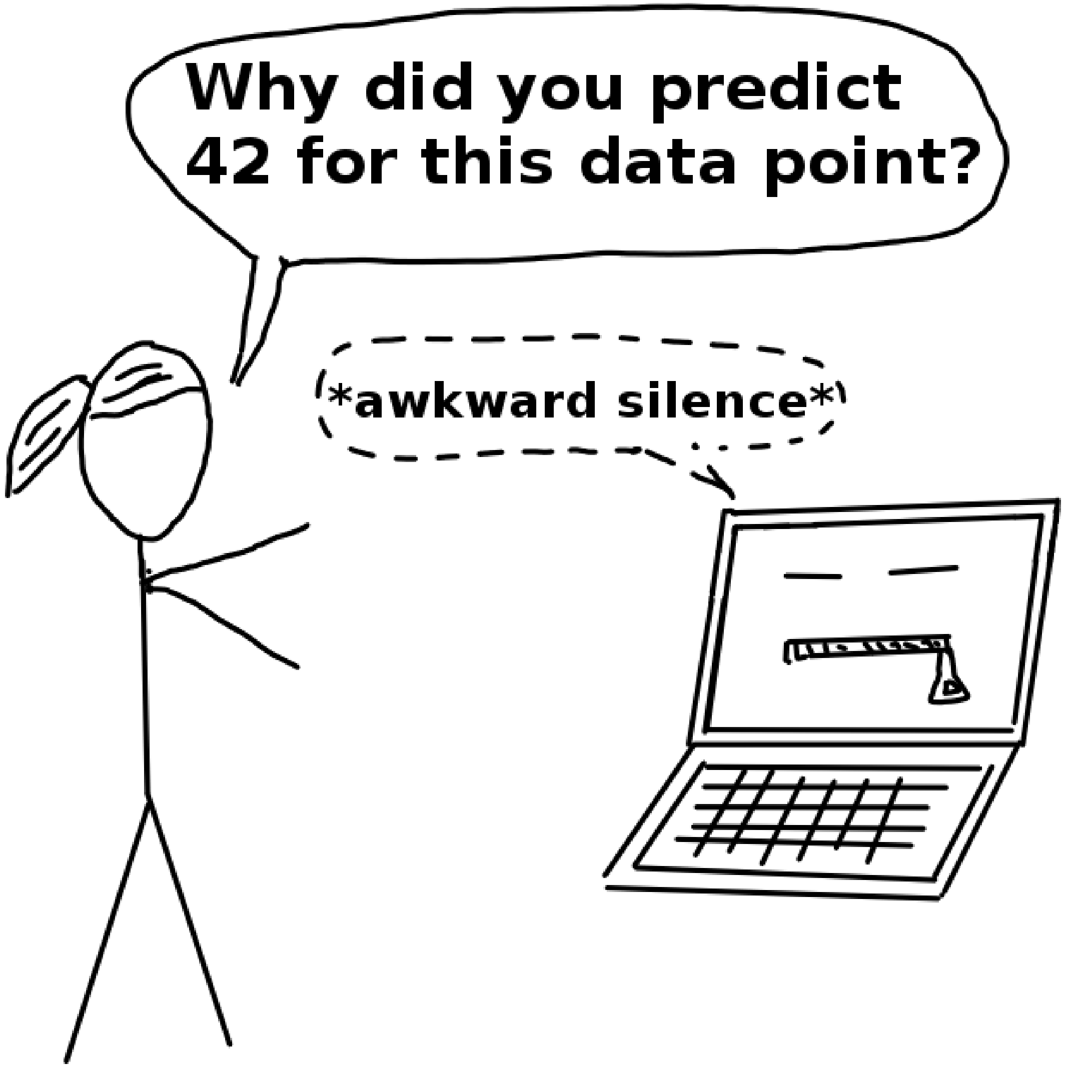
…but extrapolate poorly to a climate 8K warmer than training

Daily-mean Tropical prediction in warm climate



Problem 1: ML algorithms fail to generalize

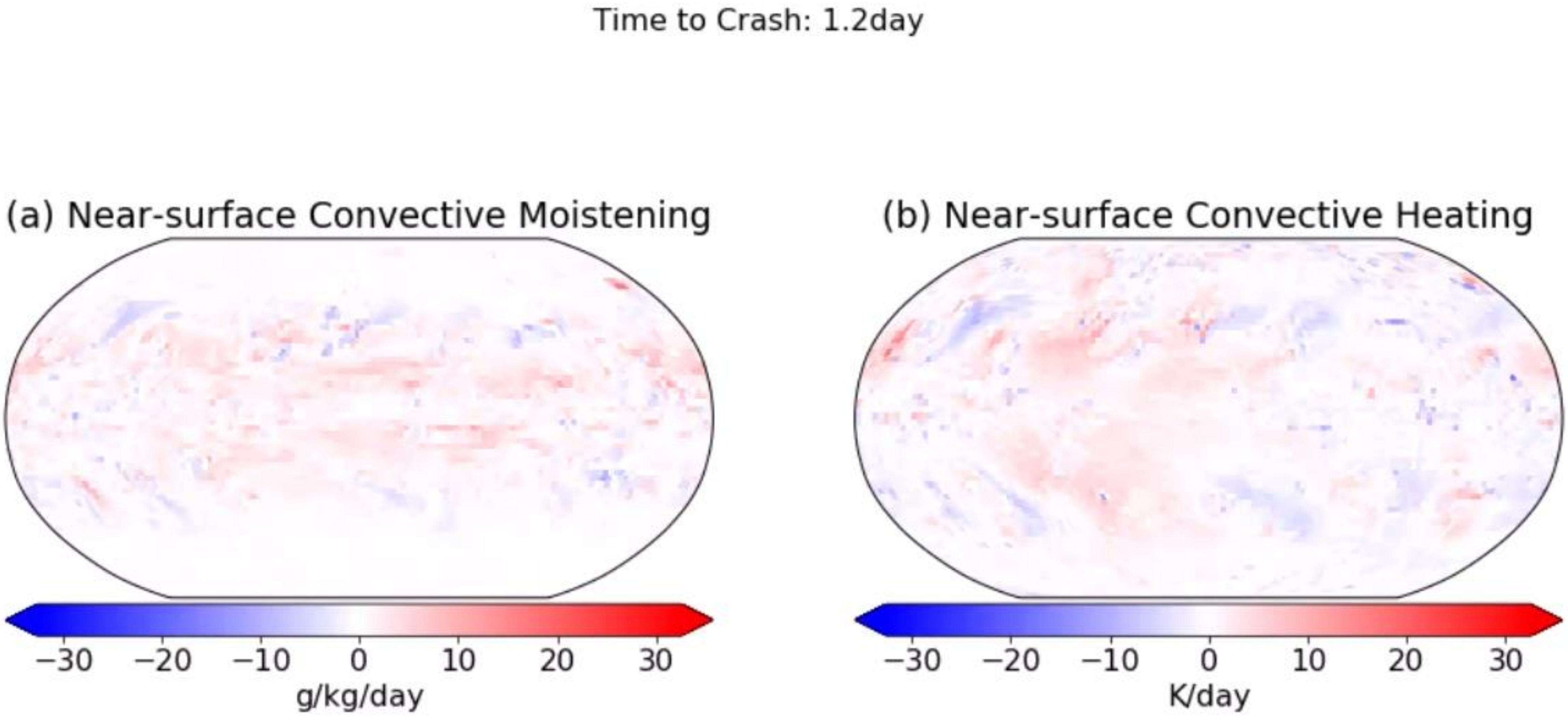
Problem 2: ML parametrizations hard to interpret/trust



Source: Interpretable Machine Learning, C. Molnar (2019)

Problem 1: ML algorithms fail to generalize

Problem 2: ML parametrizations hard to interpret/trust



See: Brenowitz, Beucler et al. (2020)

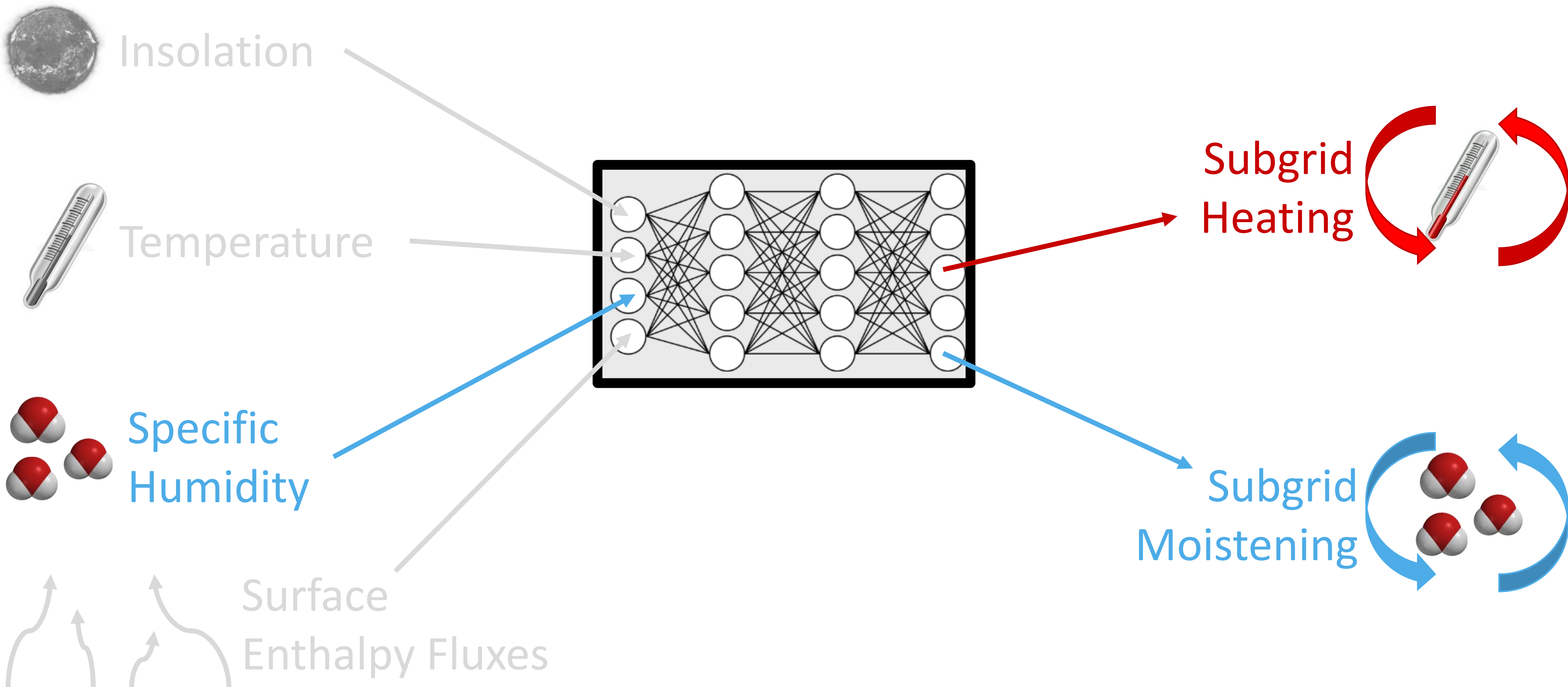
Problem 1: ML algorithms fail to generalize

Problem 2: ML parametrizations hard to interpret/trust

How can we design
data-driven models of convection
that are interpretable & generalize well?

1) How to combine ML & physical knowledge?
2) How to combine ML & causal inference?

1) **Physically-Informed ML**: Rescale inputs to convert extrapolation into ≈interpolation

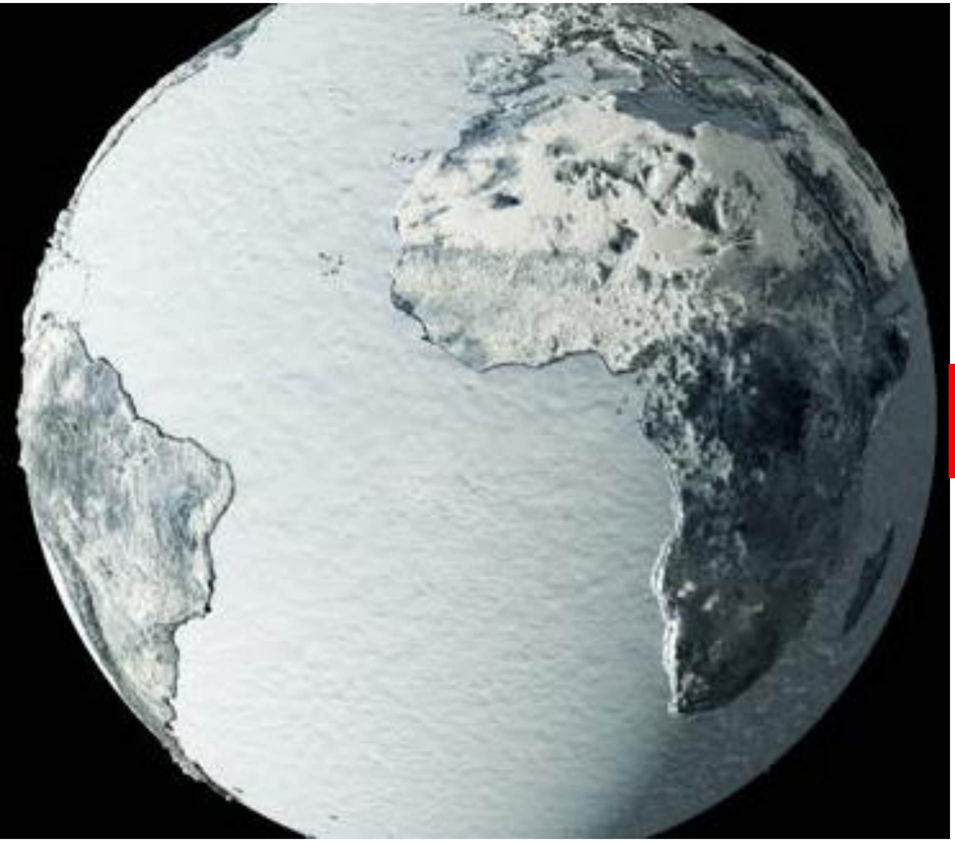


Idea: Break the model to uncover its weaknesses!

Image source: IT Biz Advisor

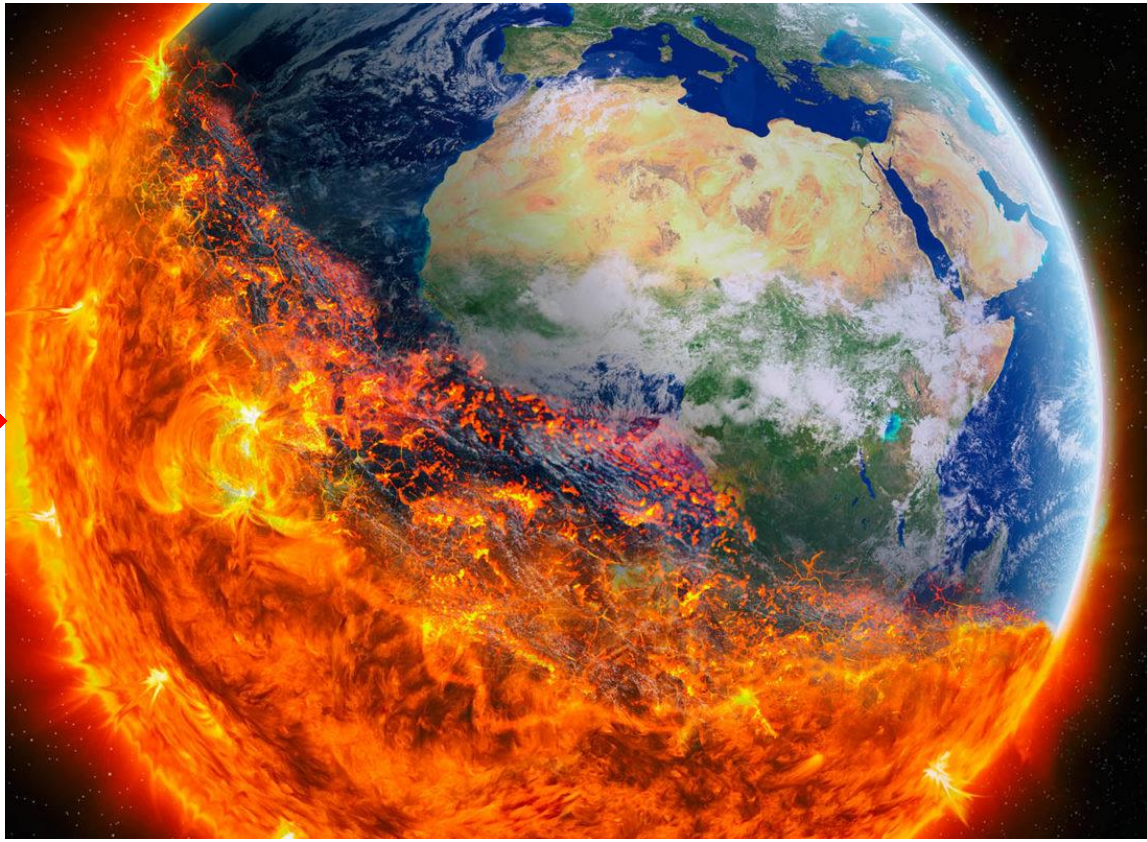
Generalization Experiment: Uniform +8K warming

Training and Validation on
cold aquaplanet simulation



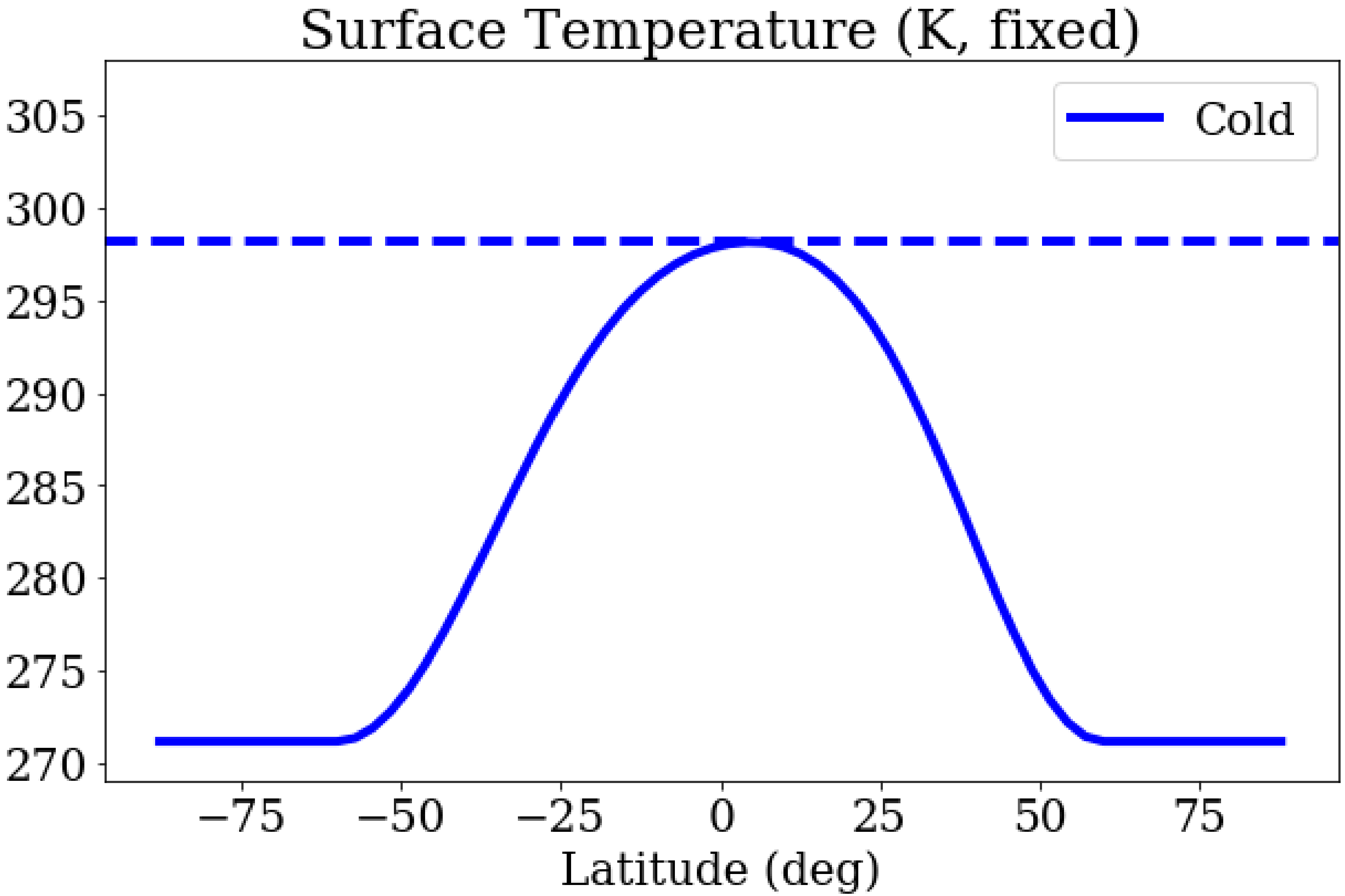
+8K

Test on warm aquaplanet simulation

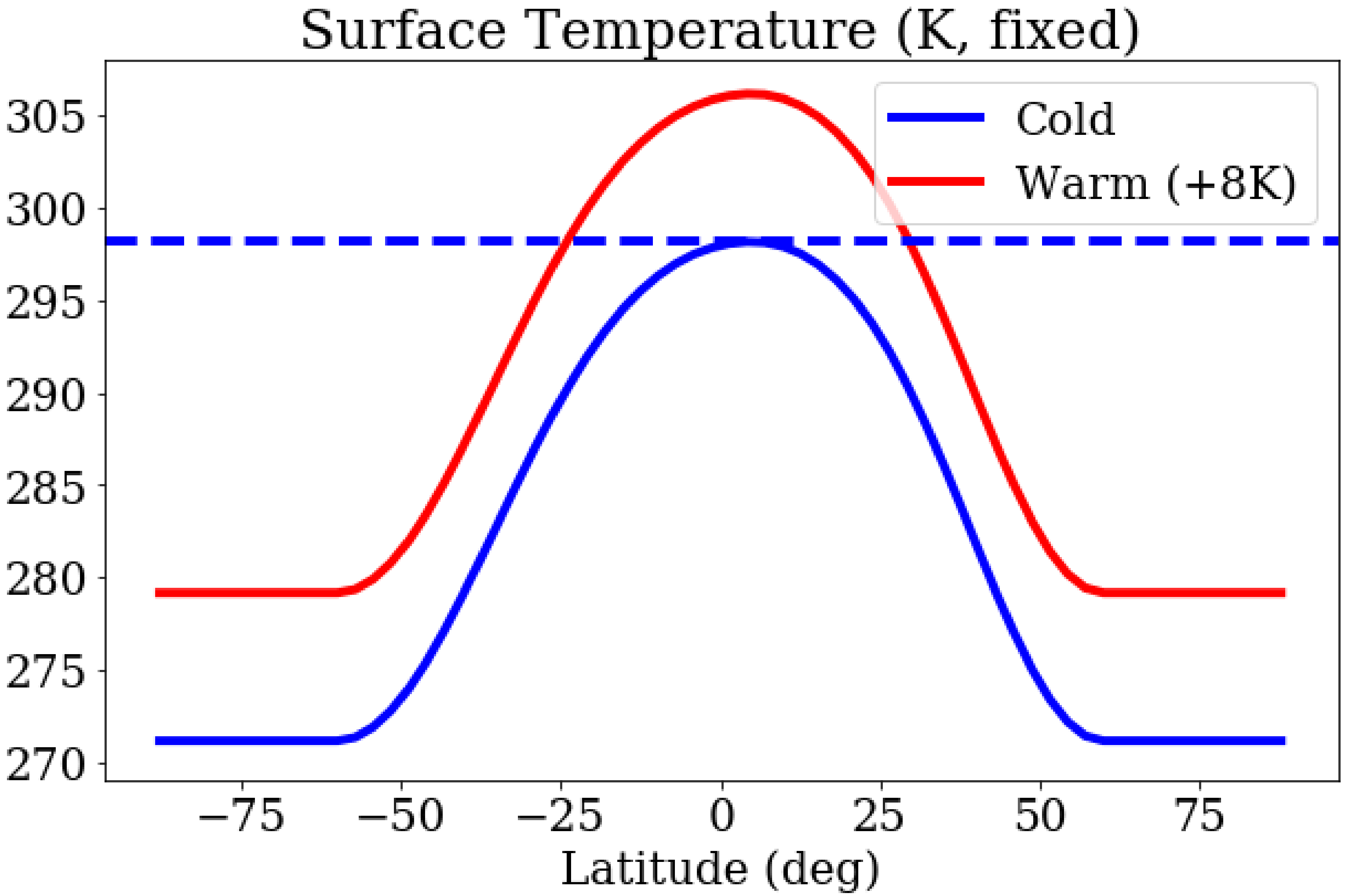


<u>Images</u>: Rashevskyi Viacheslav, Sebastien Decoret

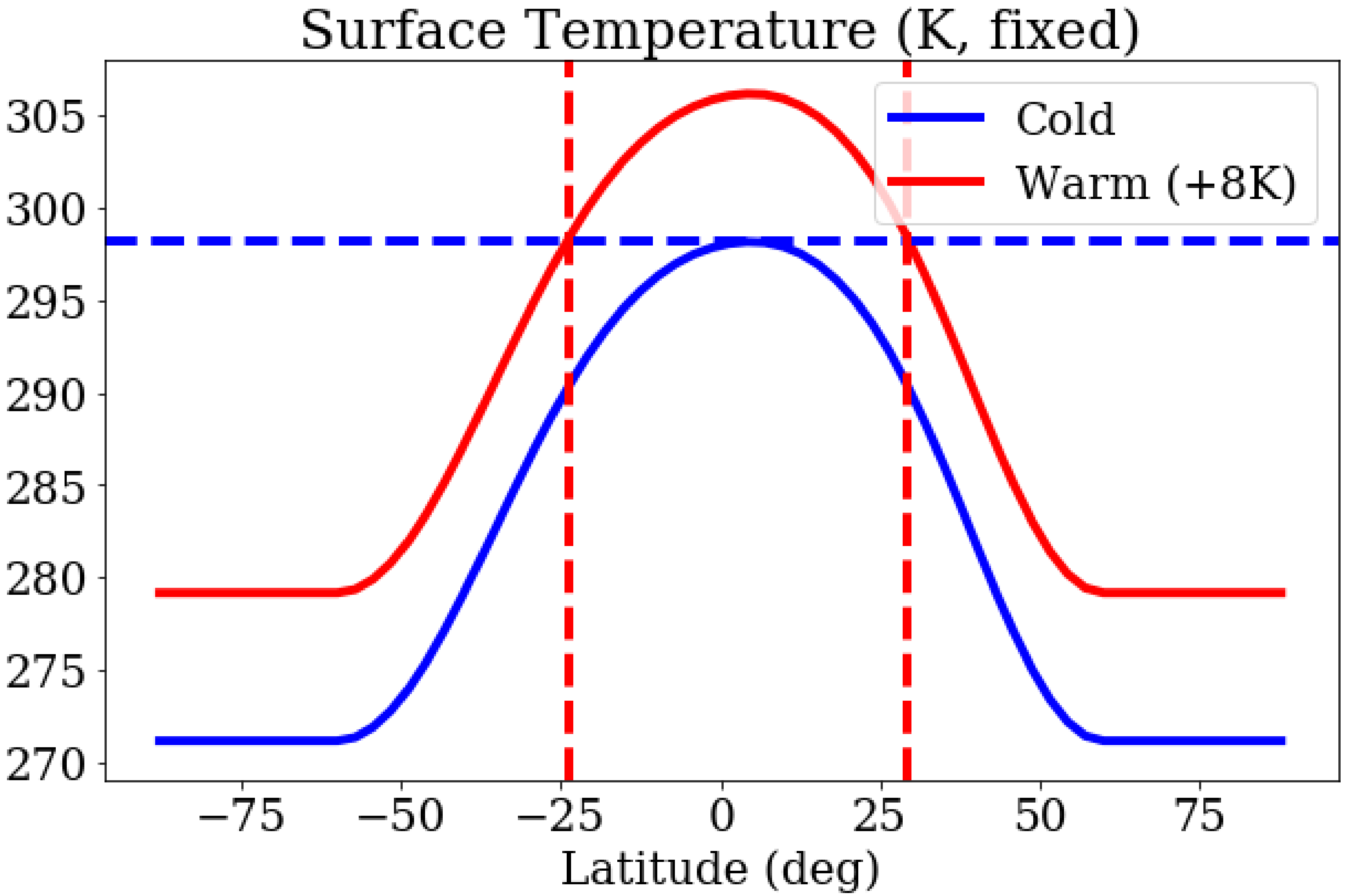
Generalization Experiment: Uniform +8K warming



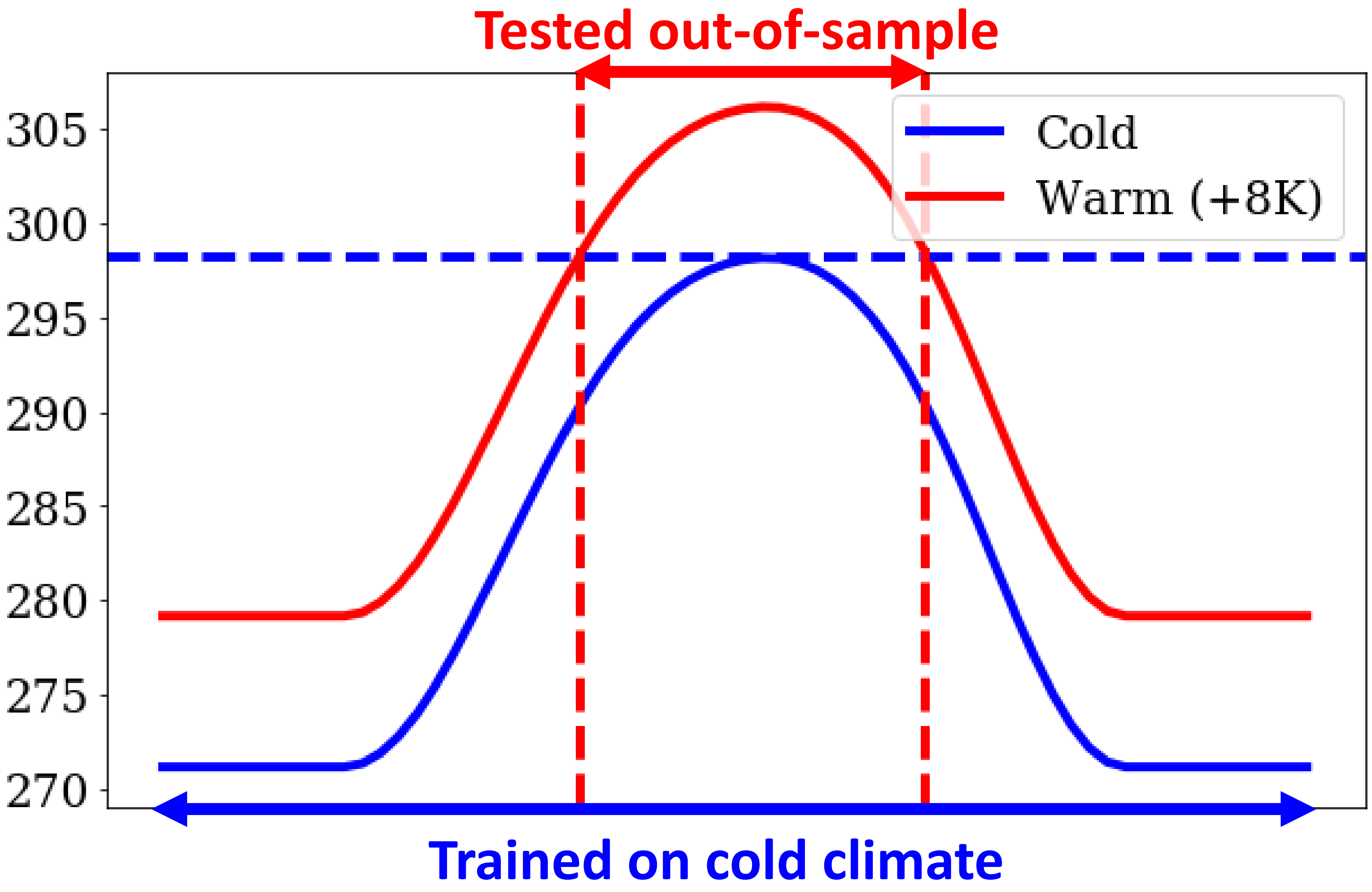
Generalization Experiment: Uniform +8K warming



Generalization Experiment: Uniform +8K warming

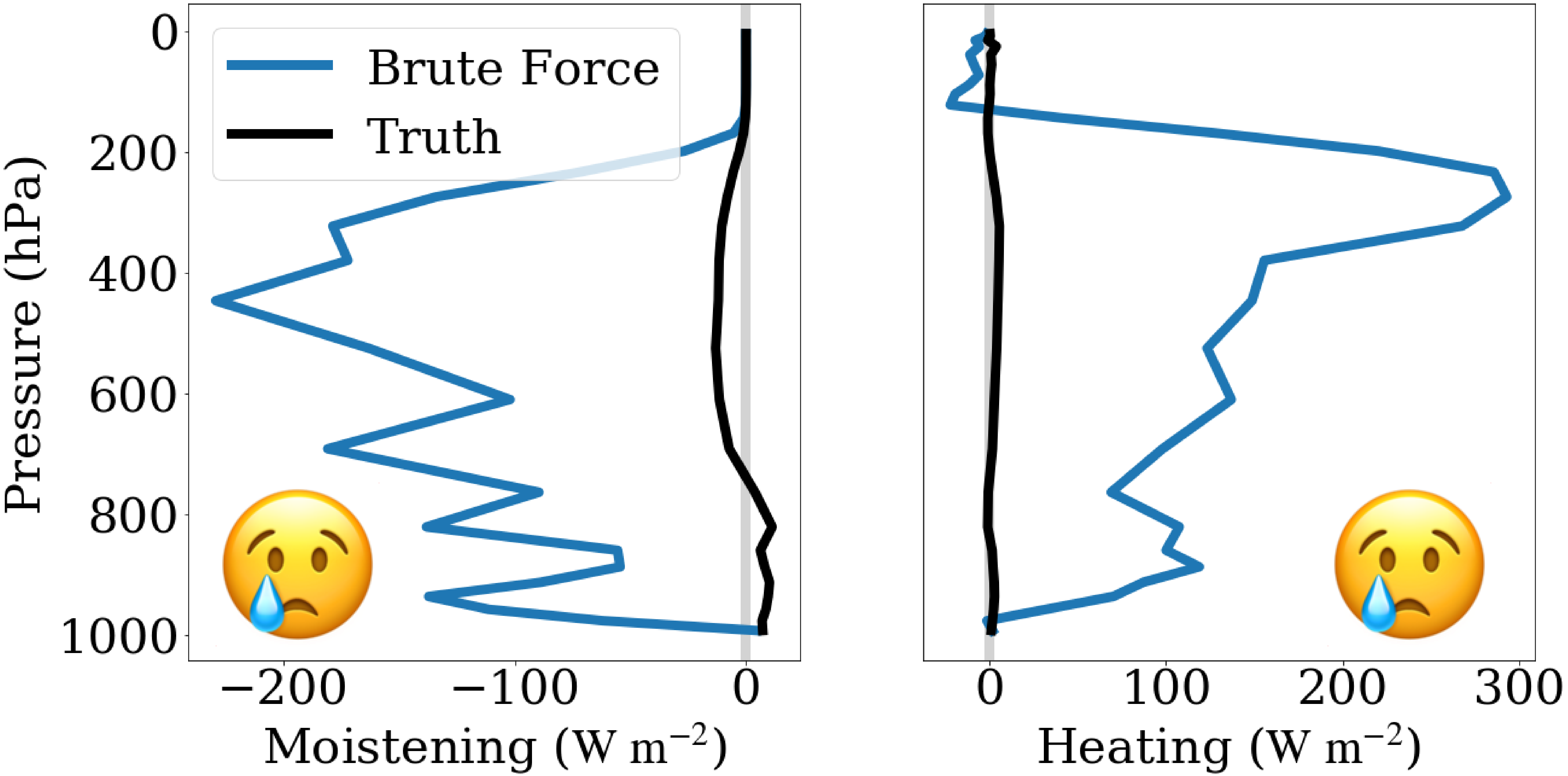


Generalization Experiment: Uniform +8K warming



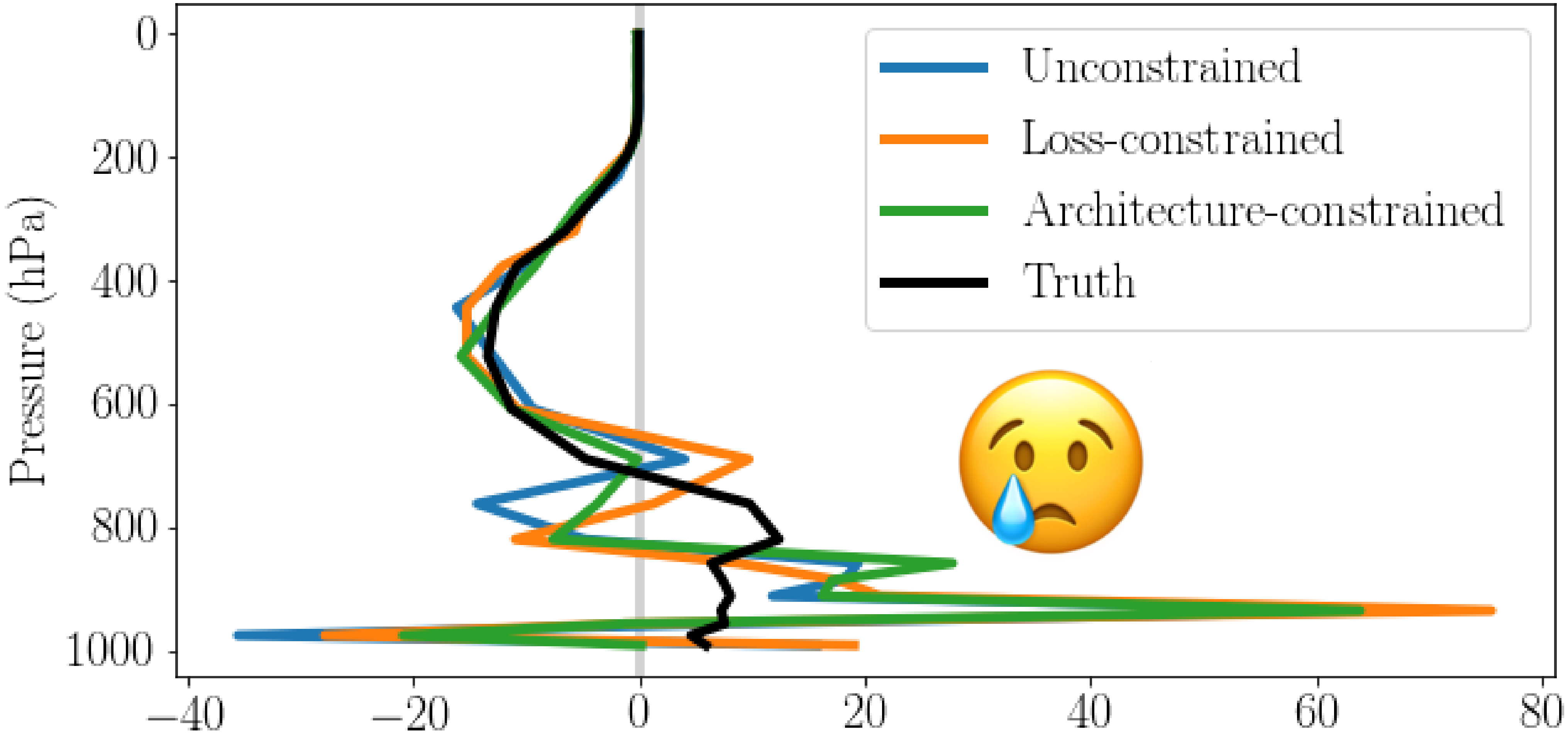
NNs extrapolate poorly to a climate 8K warmer than training…

Daily-mean Tropical prediction in warm climate



…even when constrained to conserve mass and energy

Daily-mean Tropical prediction in (+4K) warming experiment



Convective moistening (W m^{-2})

<u>See:</u> Beucler et al. (2019)

Physically rescale the data to convert extrapolation into interpolation

$$\begin{bmatrix} \text{Specific humidity}\,(p) \\ \text{Temperature}\,(p) \\ \text{Surface Pressure} \\ \text{Solar Insolation} \\ \text{Latent Heat Flux} \\ \text{Sensible Heat Flux} \end{bmatrix} \overset{\text{NN}}{\mapsto} \begin{bmatrix} \text{Subgrid moistening}\,(p) \\ \text{Subgrid heating}\,(p) \\ \text{Radiative fluxes} \end{bmatrix}$$

Brute Force: Not Climate-Invariant

Physically rescale the data to convert extrapolation into interpolation

<u>Goal</u>: Uncover **climate-invariant** mapping from climate to convection

$$\begin{bmatrix} \widetilde{\text{Specific humidity}}\,(\widetilde{p}) \\ \widetilde{\text{Temperature}}\,(\widetilde{p}) \\ \widetilde{\text{Surface Pressure}} \\ \widetilde{\text{Solar Insolation}} \\ \widetilde{\text{Latent Heat Flux}} \\ \widetilde{\text{Sensible Heat Flux}} \end{bmatrix} \overset{\widetilde{\text{NN}}}{\longmapsto} \begin{bmatrix} \widetilde{\text{Subgrid moistening}}\,(\widetilde{p}) \\ \widetilde{\text{Subgrid heating}}\,(\widetilde{p}) \\ \widetilde{\text{Radiative fluxes}} \end{bmatrix}$$

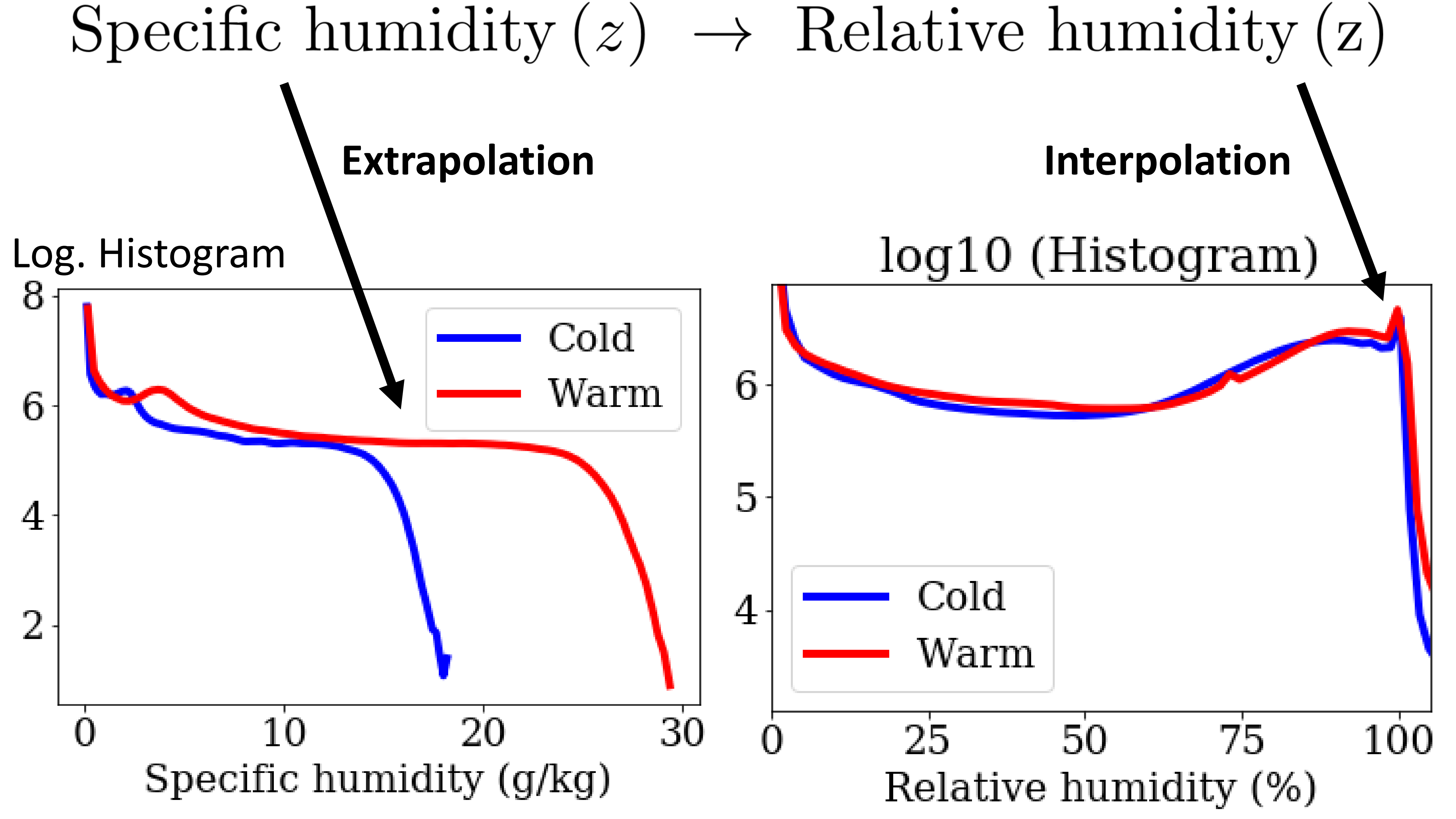
<u>Goal</u>: **Climate-Invariant**

Physically rescale the data
to convert extrapolation into interpolation

<u>Goal</u>: Uncover **climate-invariant** mapping from climate to convection

$$\begin{bmatrix} \widetilde{\text{Specific humidity}}\,(\widetilde{p}) \\ \widetilde{\text{Temperature}}\,(\widetilde{p}) \\ \widetilde{\text{Surface Pressure}} \\ \widetilde{\text{Solar Insolation}} \\ \widetilde{\text{Latent Heat Flux}} \\ \widetilde{\text{Sensible Heat Flux}} \end{bmatrix} \overset{\widetilde{\text{NN}}}{\longmapsto} \begin{bmatrix} \widetilde{\text{Subgrid moistening}}\,(\widetilde{p}) \\ \widetilde{\text{Subgrid heating}}\,(\widetilde{p}) \\ \widetilde{\text{Radiative fluxes}} \end{bmatrix}$$

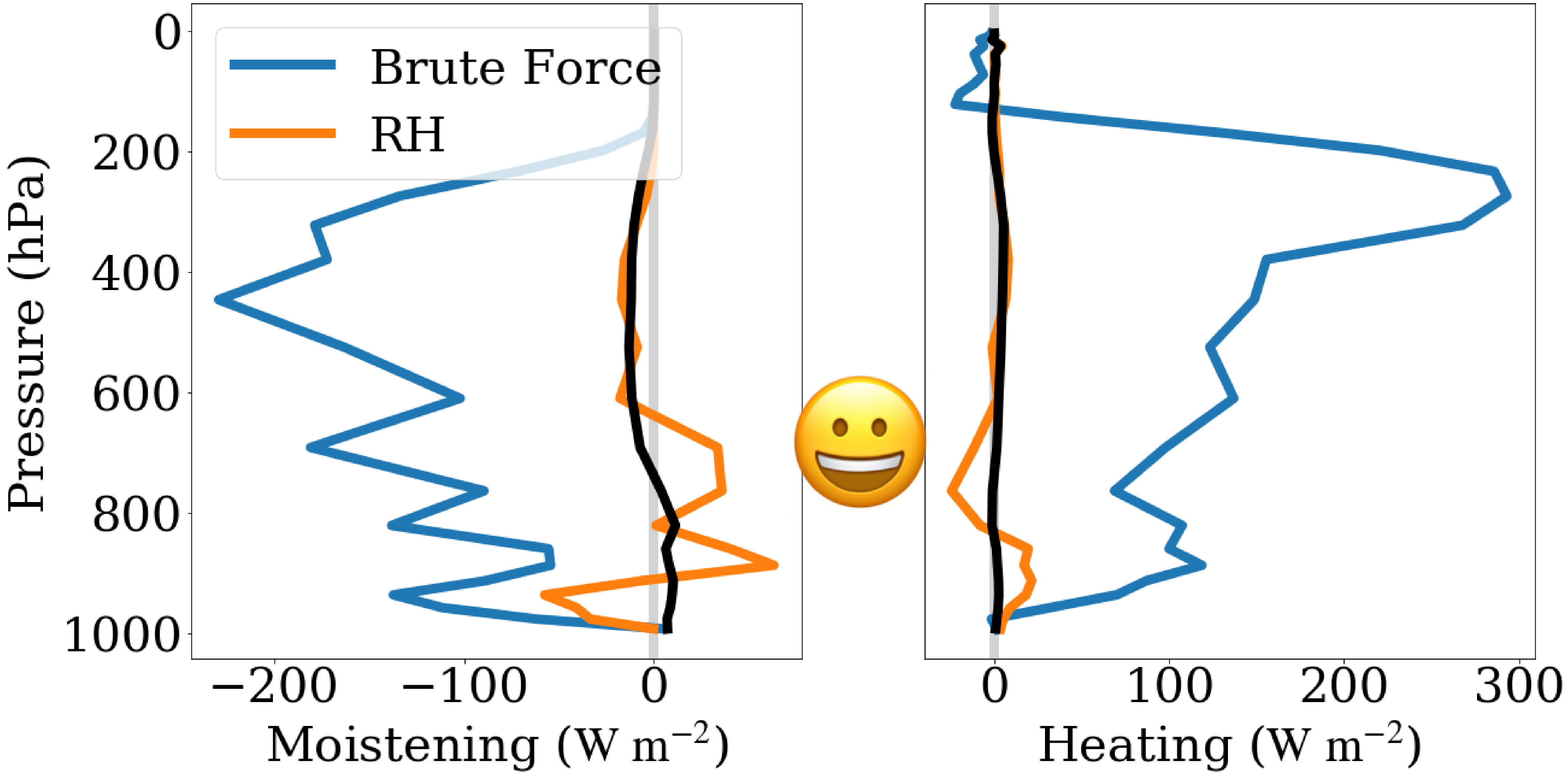
How to choose the physical rescaling?

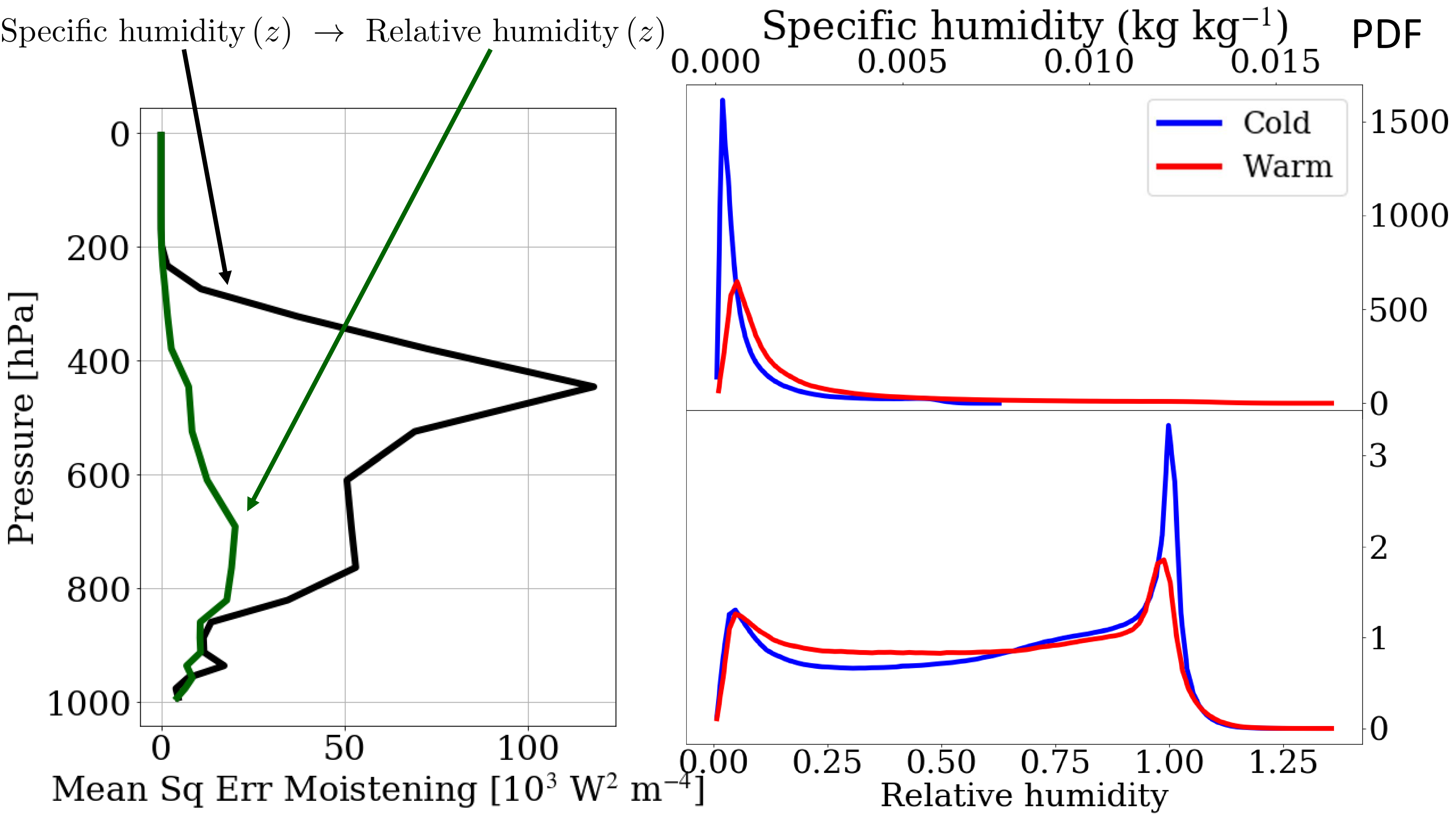


Specific humidity (z) $\rightarrow$ Relative humidity (z)
Extrapolation
Interpolation
Log. Histogram
8
6
4
2
Cold
Warm
0
10
20
30
Specific humidity (g/kg)
log10 (Histogram)
6
5
4
Cold
Warm
0
25
50
75
100
Relative humidity (%)

Specific humidity (z) $\rightarrow$ Relative humidity (z)

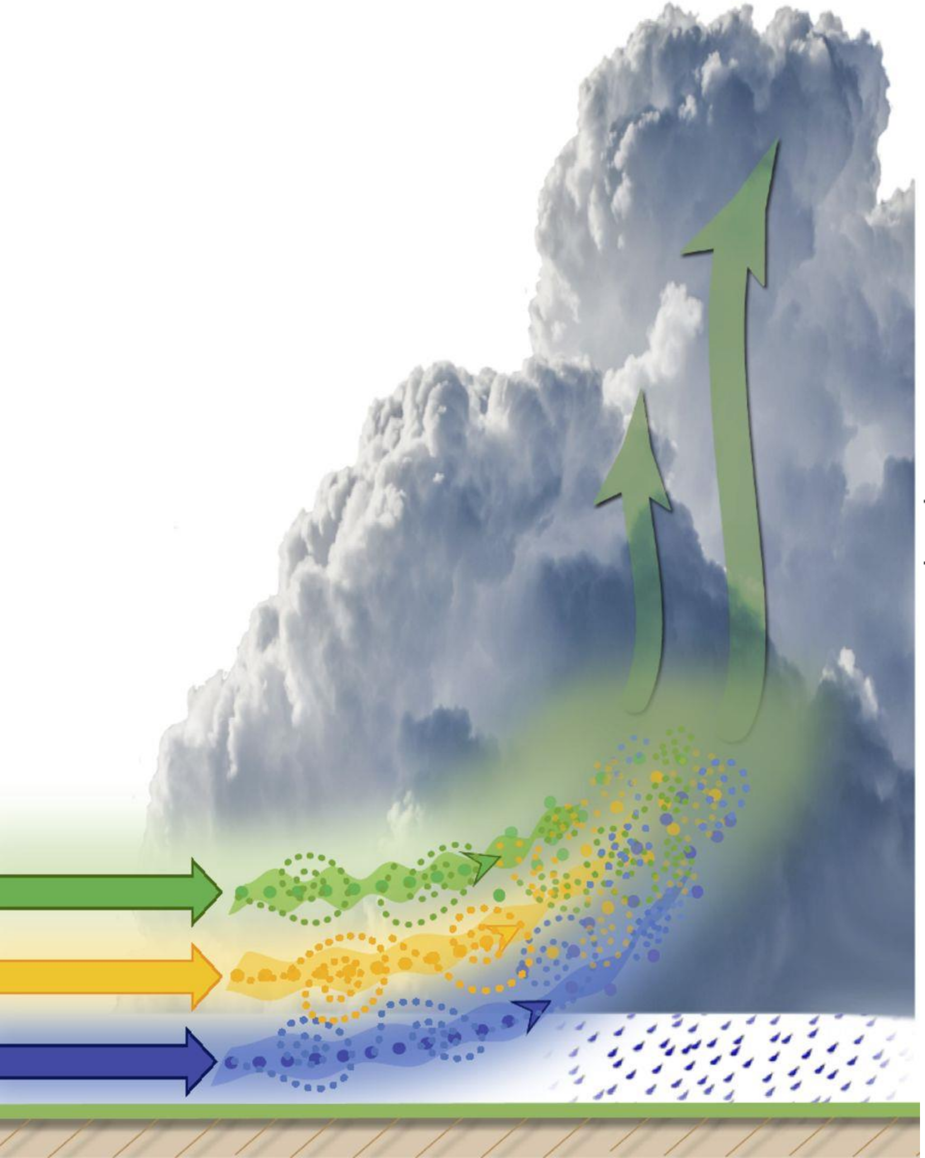
Generalization improves dramatically!





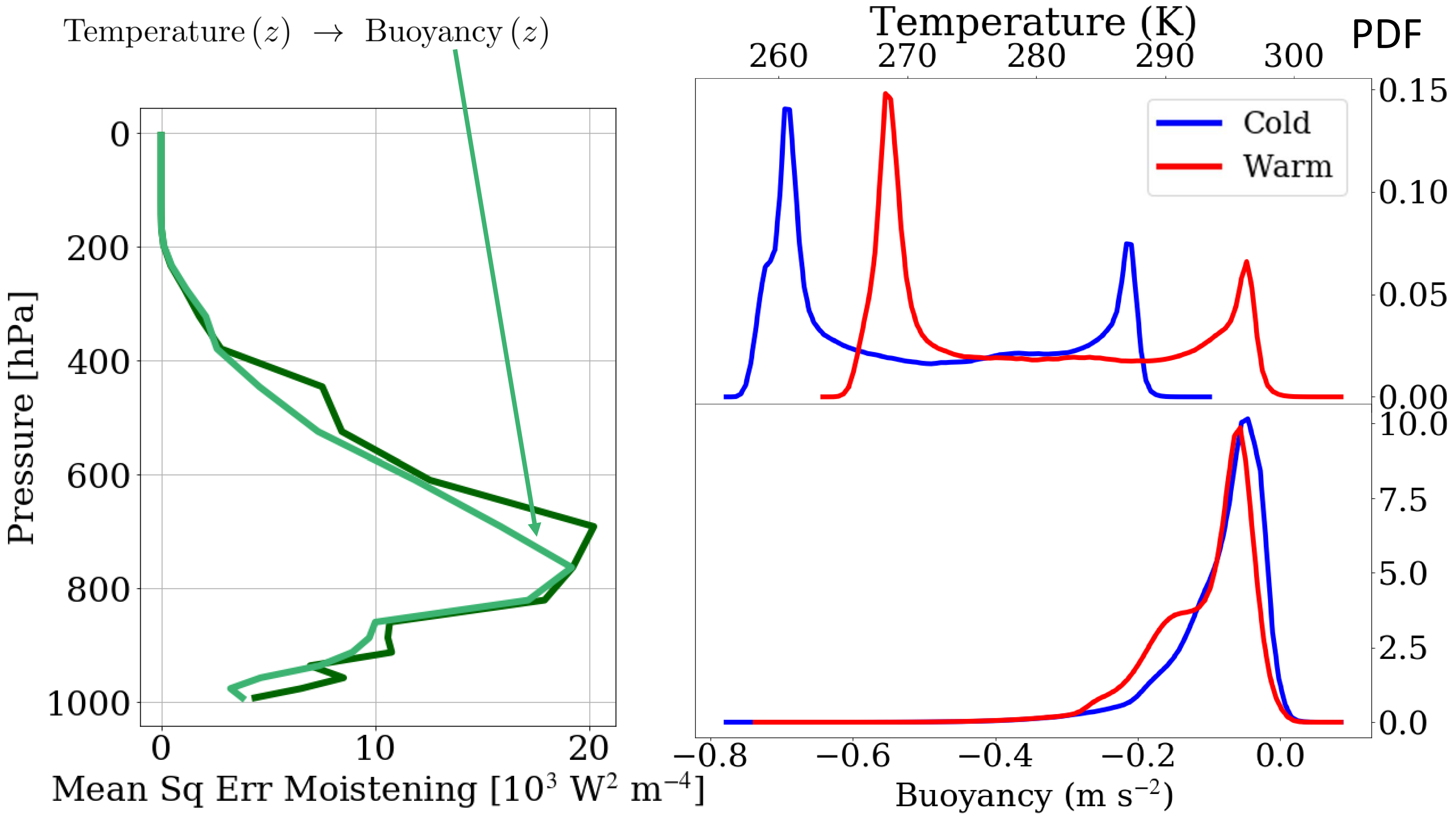
Specific humidity (z) $\rightarrow$ Relative humidity (z)
Pressure [hPa]
0
200
400
600
800
1000
0
50
100
Mean Sq Err Moistening [10^3 W^2 m^{-4}]
Specific humidity (kg kg^{-1})
0.000
0.005
0.010
0.015
PDF
Cold
Warm
1500
1000
500
0
3
2
1
0
0.00
0.25
0.50
0.75
1.00
1.25
Relative humidity

Observations suggest a strong relationship between buoyancy & moist convection across scales

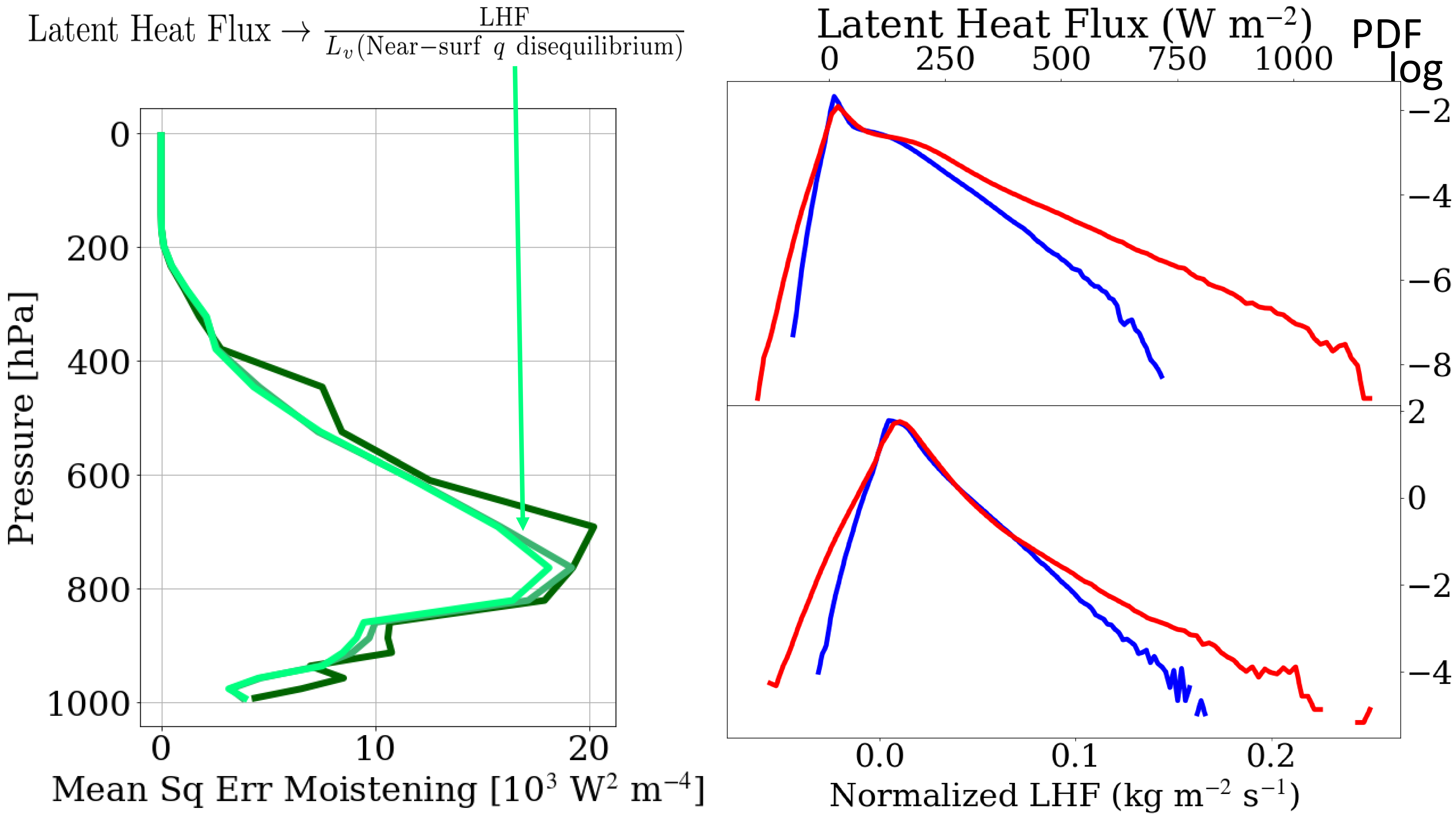


$$\text{Buoyancy}\,(z) \stackrel{\text{def}}{=} g \times \frac{\text{Temp parcel} - \text{Temp}(z)}{\text{Temp}(z)}$$

<u>See:</u> Schiro et al. (2018), Ahmed & Neelin (2018), Ahmed et al. (2020)

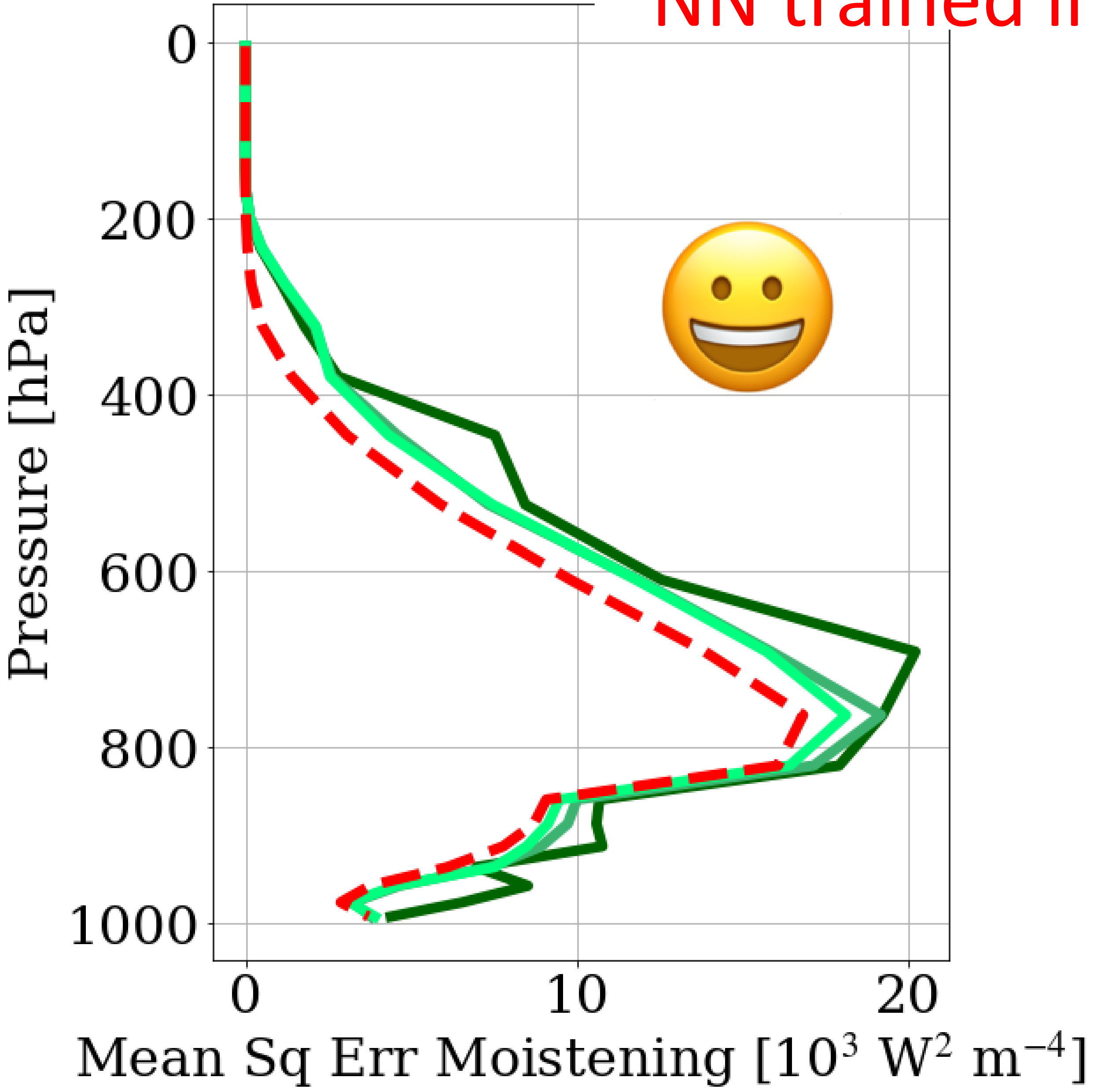


Temperature (z) → Buoyancy (z)
Pressure [hPa]
0
200
400
600
800
1000
0
10
20
Mean Sq Err Moistening [10^3 W^2 m^{-4}]
Temperature (K)
260
270
280
290
300
PDF
0.15
0.10
0.05
0.00
Cold
Warm
10.0
7.5
5.0
2.5
0.0
−0.8
−0.6
−0.4
−0.2
0.0
Buoyancy (m s^{-2})

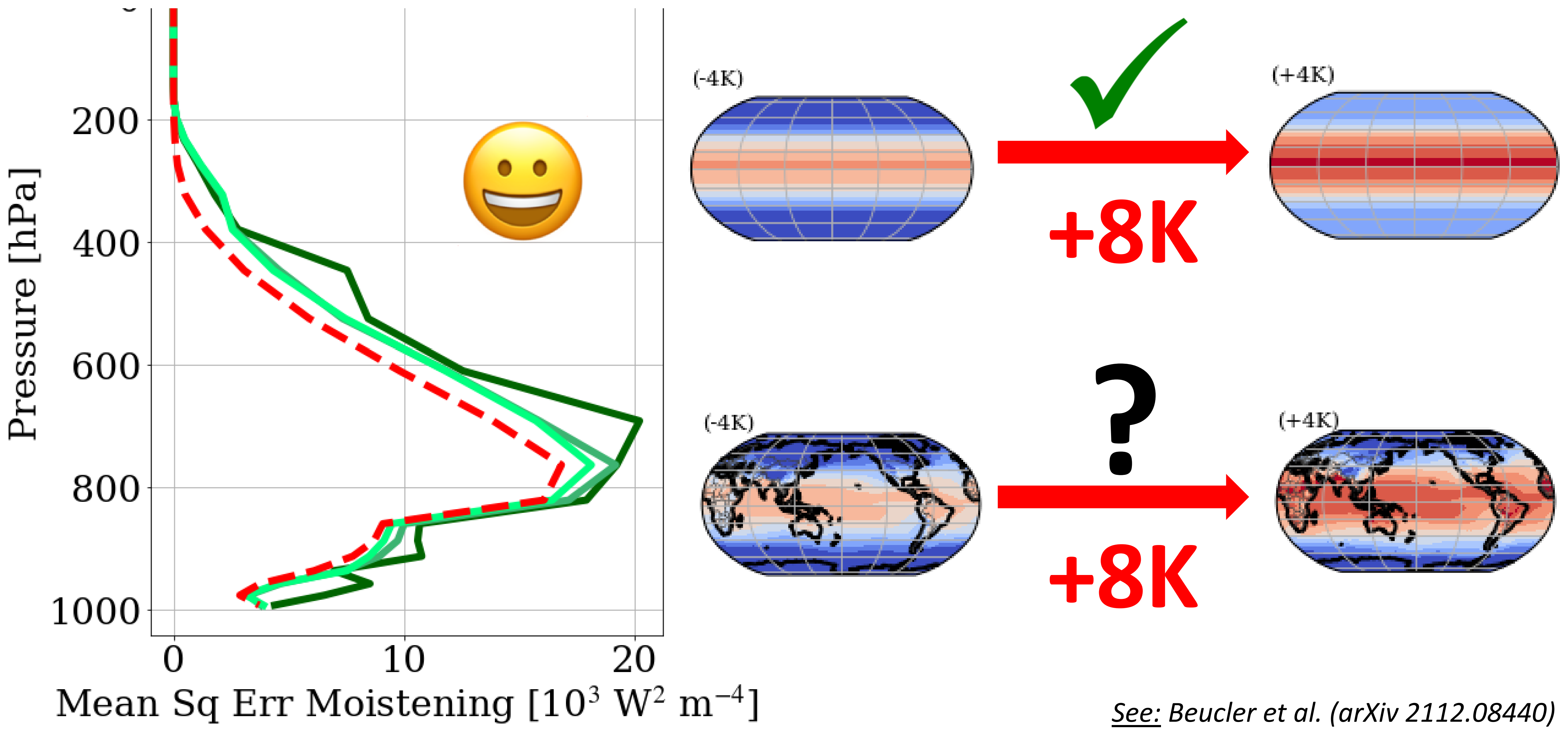


Latent Heat Flux → $\frac{\mathrm{LHF}}{L_v(\mathrm{Near{-}surf}\ q\ \mathrm{disequilibrium})}$
Pressure [hPa]
0
200
400
600
800
1000
0
10
20
Mean Sq Err Moistening [10^3 W^2 m^{-4}]
Latent Heat Flux (W m^{-2})
0
250
500
750
1000
PDF
log
−2
−4
−6
−8
2
0
−2
−4
0.0
0.1
0.2
Normalized LHF (kg m^{-2} s^{-1})

Climate-Invariant NNs generalization error close to NN trained in warm climate

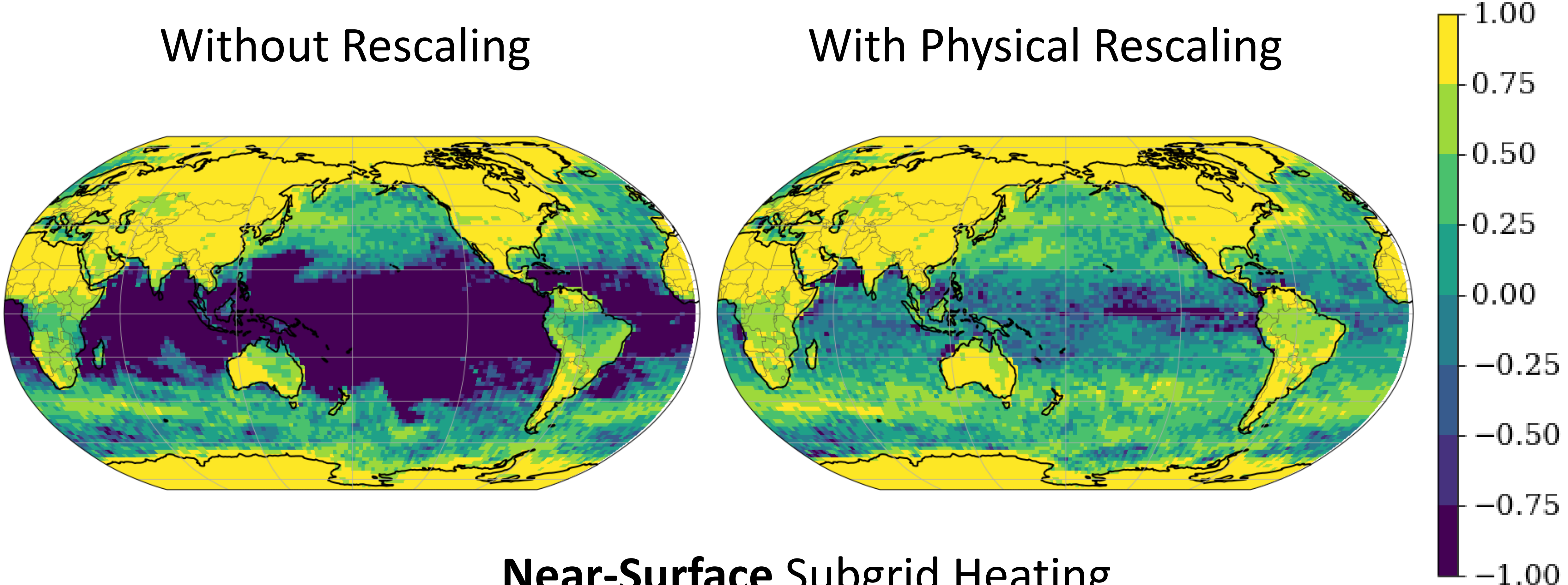


Problem 3: Physically Rescaling Inputs allows NNs to generalize from cold to warm climate



See: *Beucler et al. (arXiv 2112.08440)*

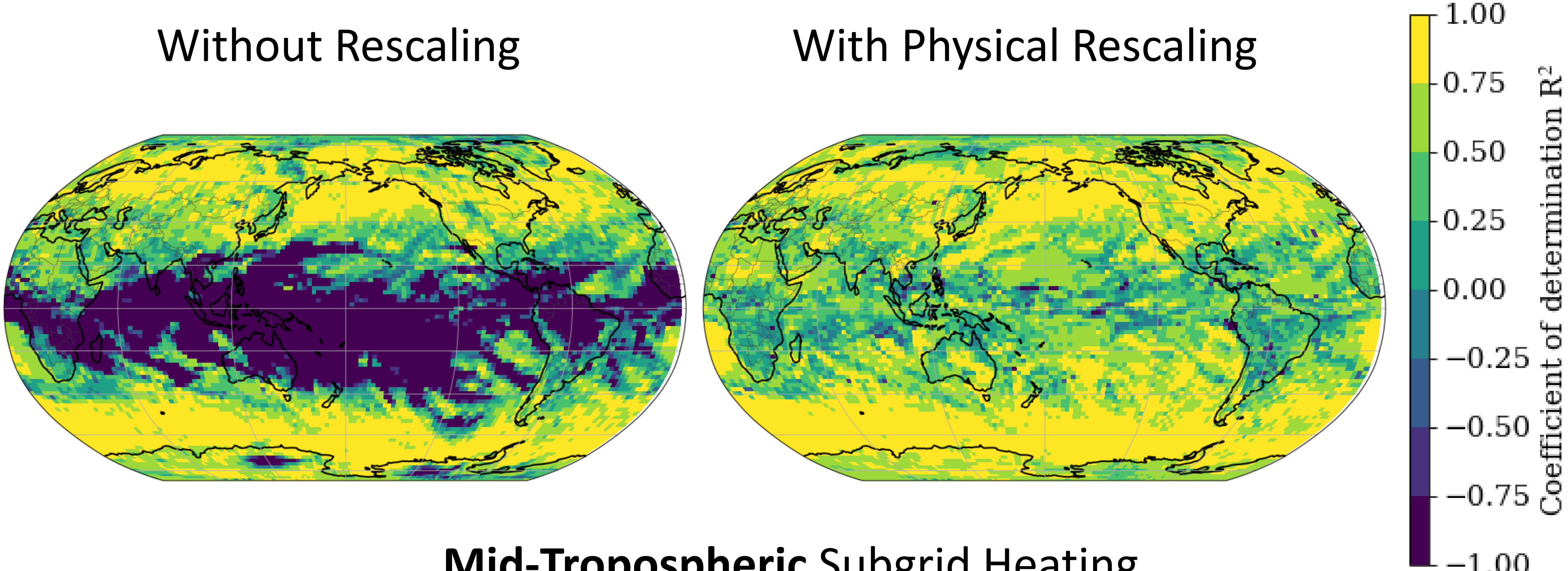
Physically-Rescaled Neural Networks Generalize Better Across Climates **in Earth-like configurations**



Near-Surface Subgrid Heating

<u>See:</u> Beucler et al. (arXiv 2112.08440)

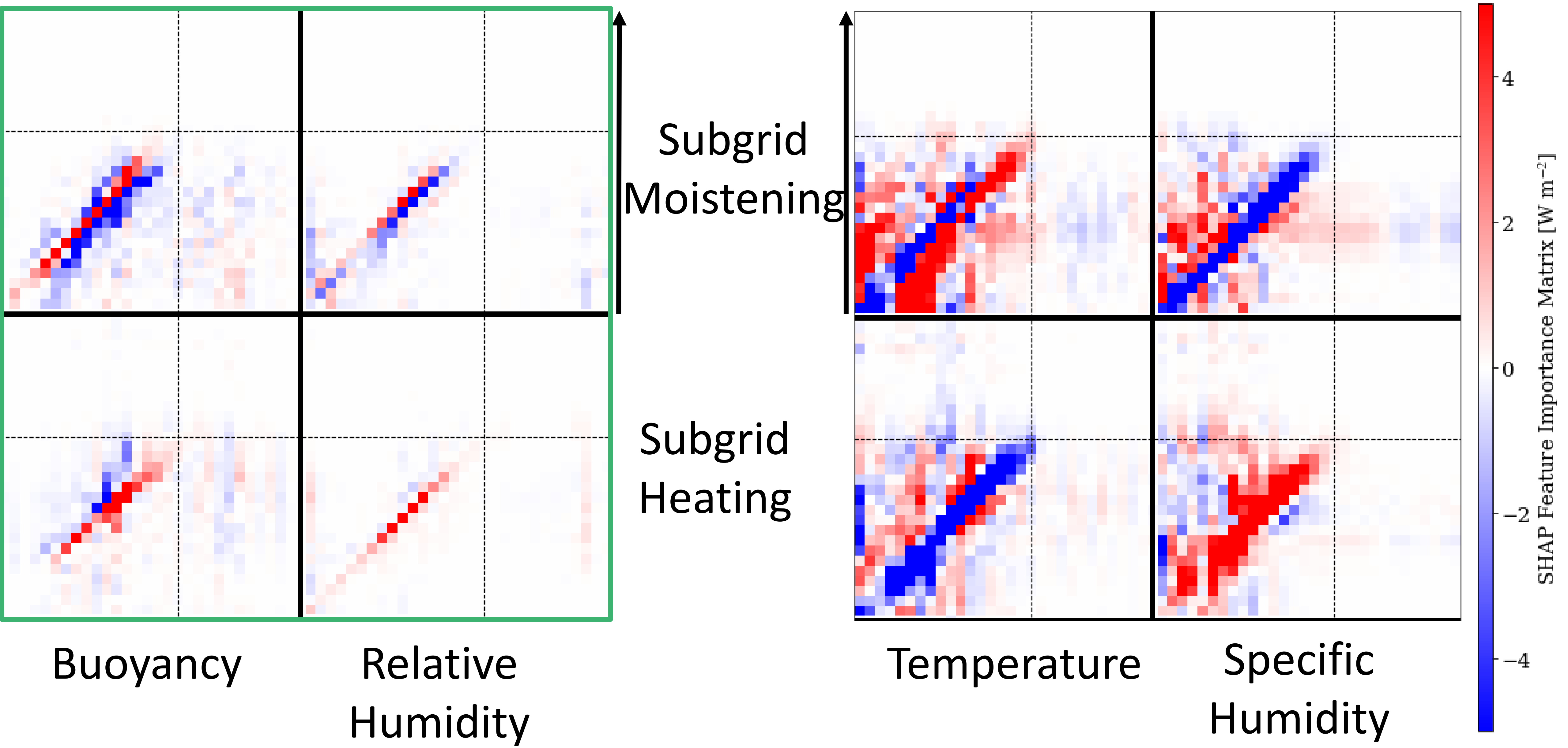
Physically-Rescaled Neural Networks Generalize Better Across Climates **in Earth-like configurations**



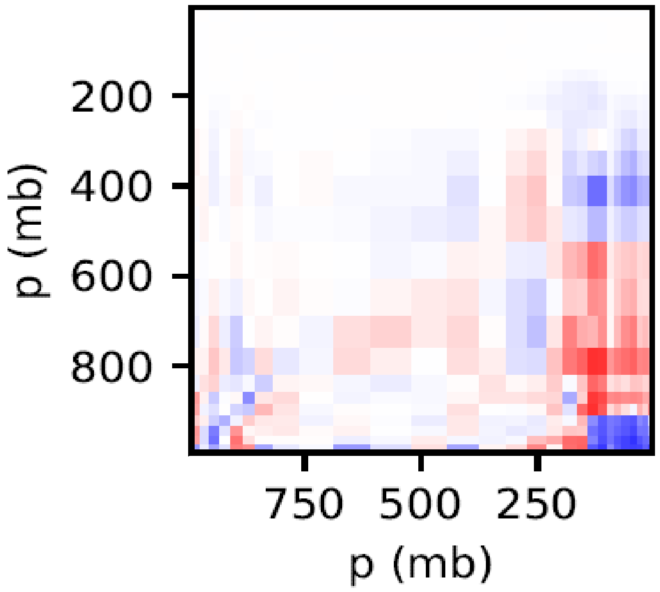
Mid-Tropospheric Subgrid Heating

<u>*See:*</u> *Beucler et al. (arXiv 2112.08440)*

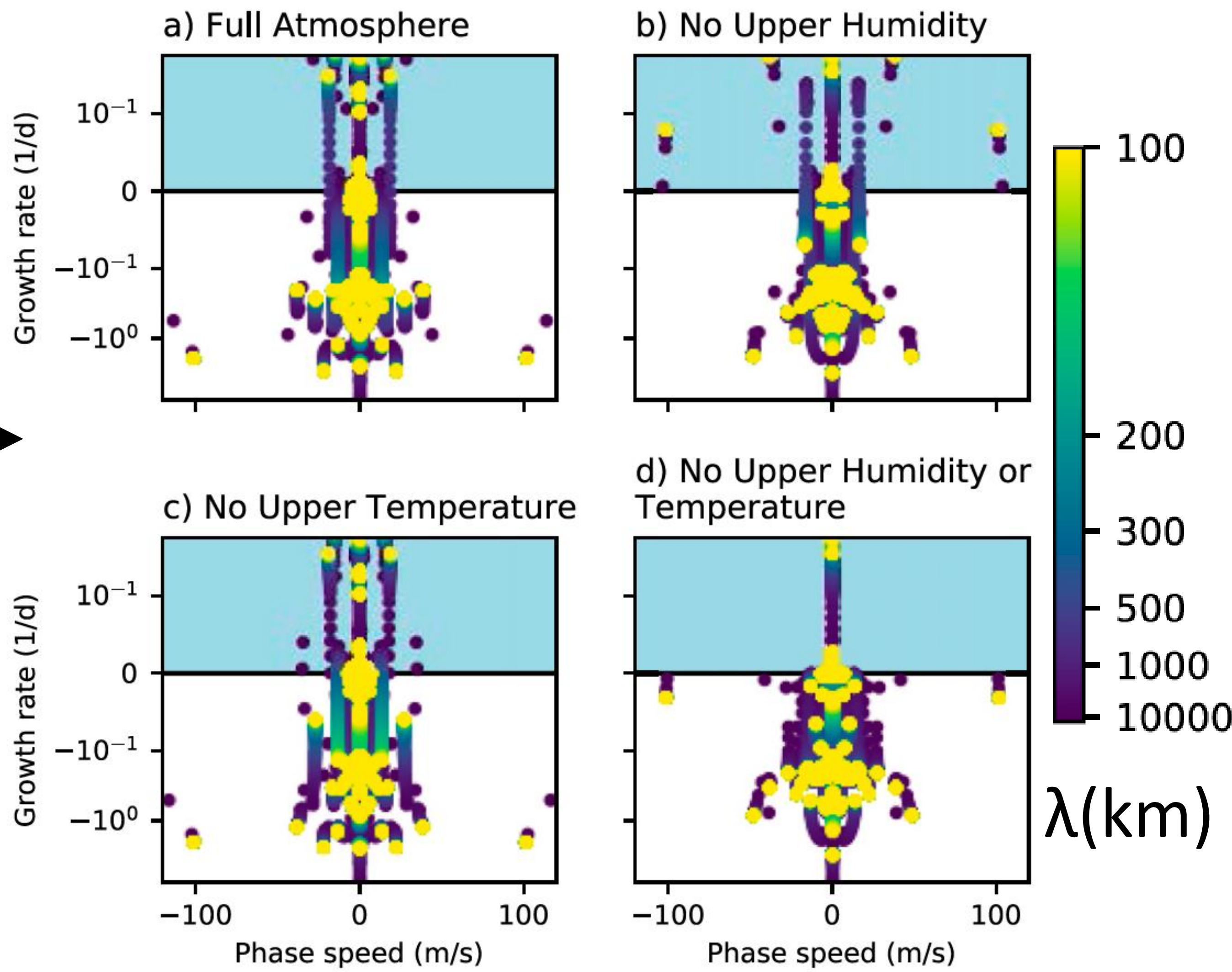
Climate-invariant NNs more local than Brute-Force NNs



Motivation for causal ML: Eliminating spurious link with stratospheric q & T eliminates instability

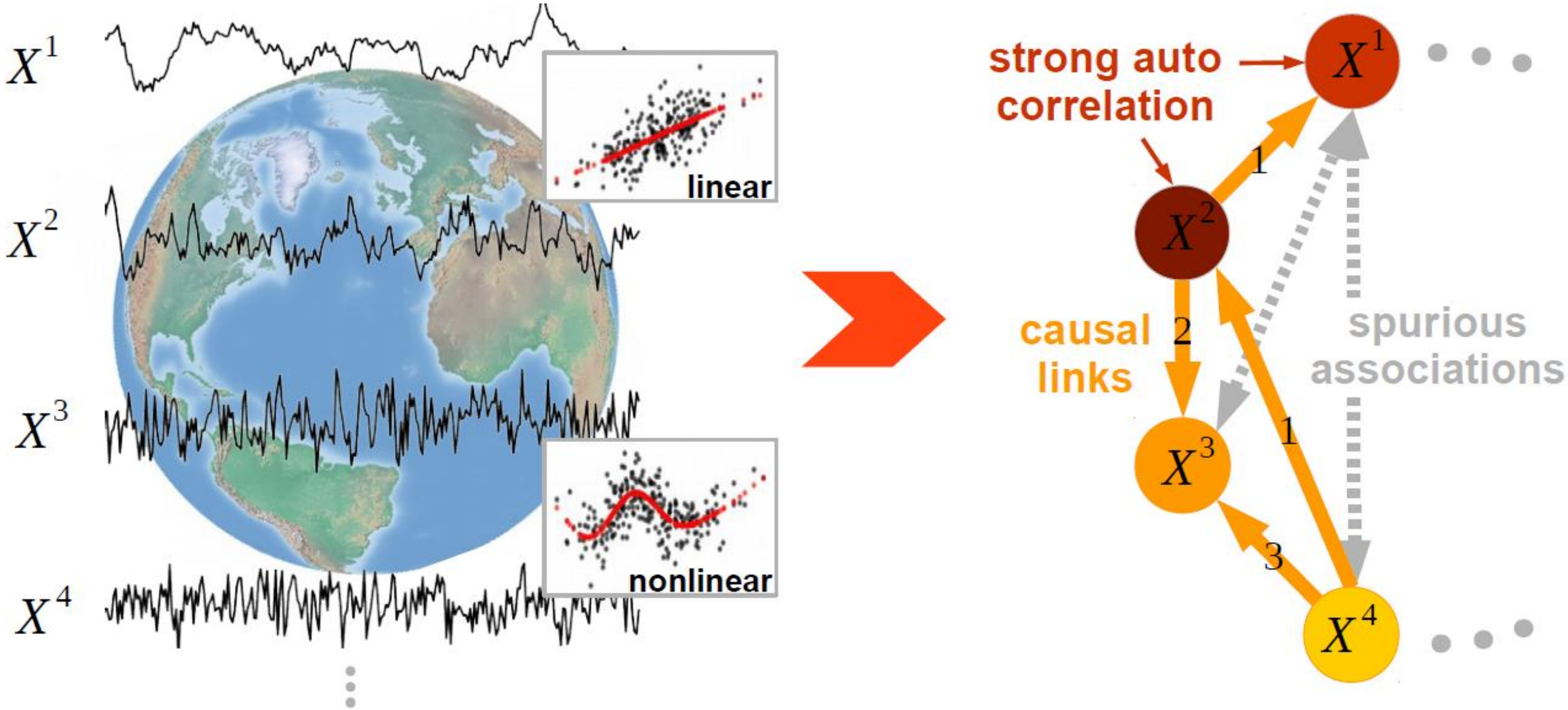


Couple to linearized gravity wave dynamics



<u>See</u>: ***Brenowitz, Beucler et al. (2020),*** *Kuang (2018, 2007), Herman and Kuang (2013)*

2) Causally-Informed ML: Phrase regression as causal discovery problem to select inputs *a priori* without degrading performance



Source: Runge et al. (2019), See: Kretschmer et al. (2016), Runge et al. (2019), Spirtes & Glymour (1991)

Causal discovery: Assumptions for PC1/PCMCI

Causal Markov Condition:

dependence ⇒ connectedness

Faithfulness assumption:

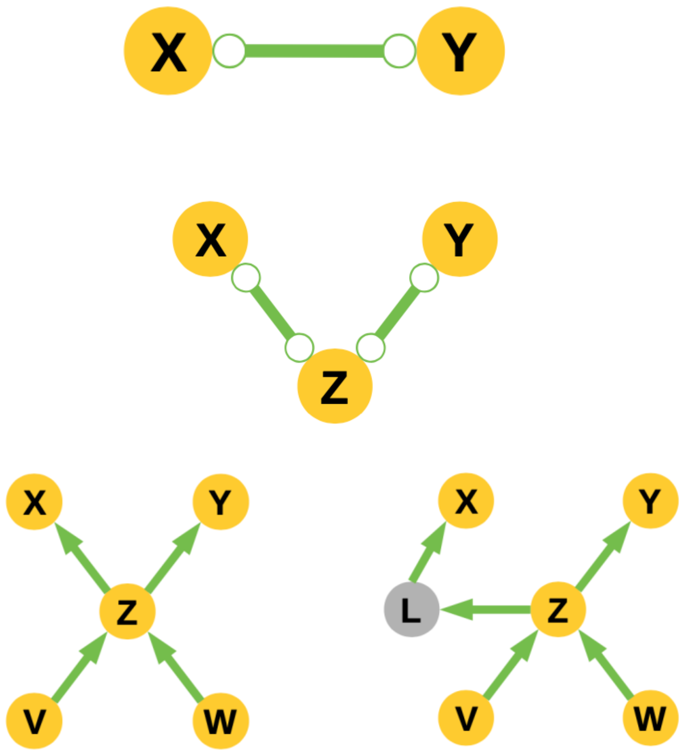
independence ⇒ no causal link

Causal sufficiency:

All common causes are “observed”

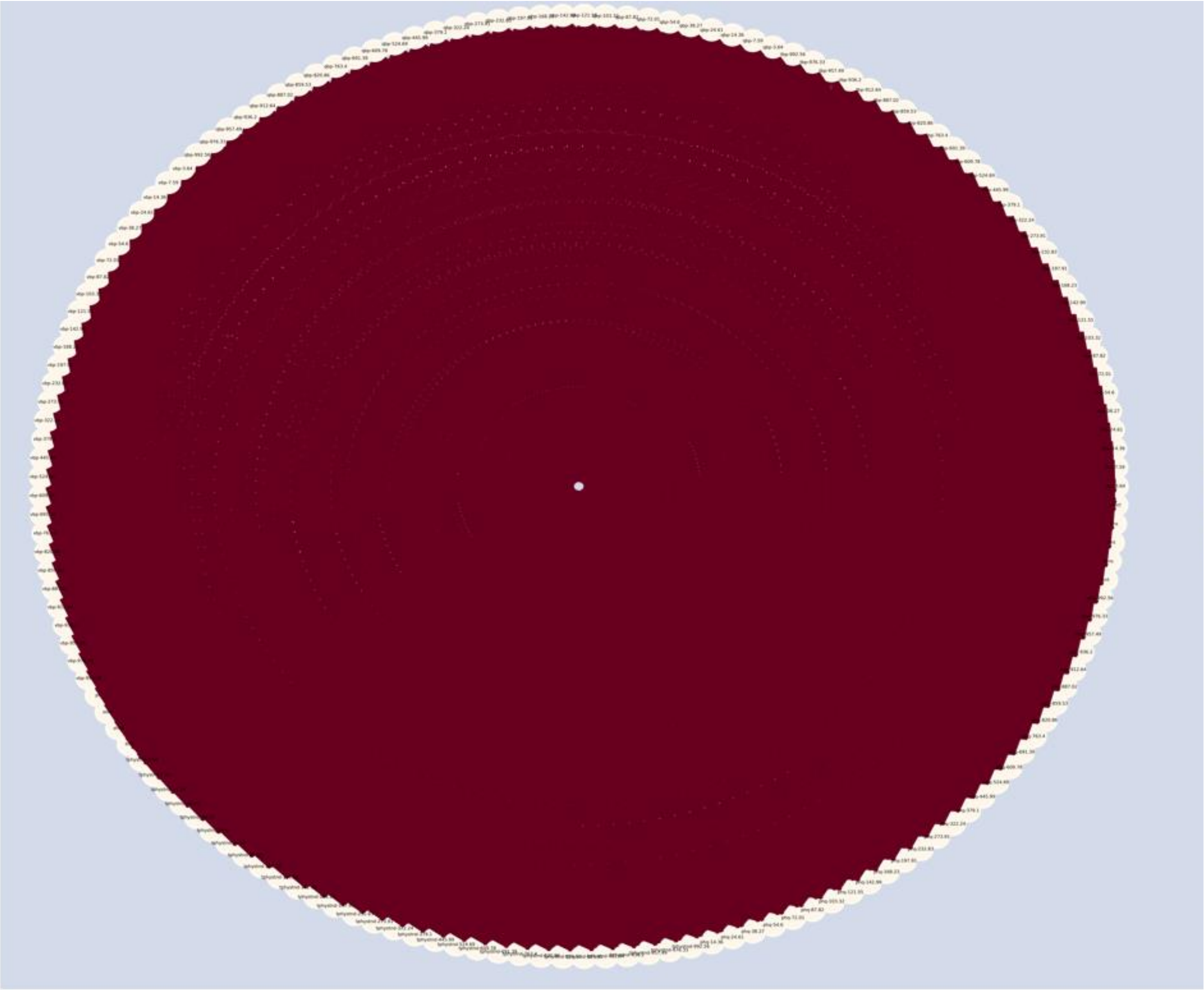
Causal stationarity:

Relationships C^t throughout time series



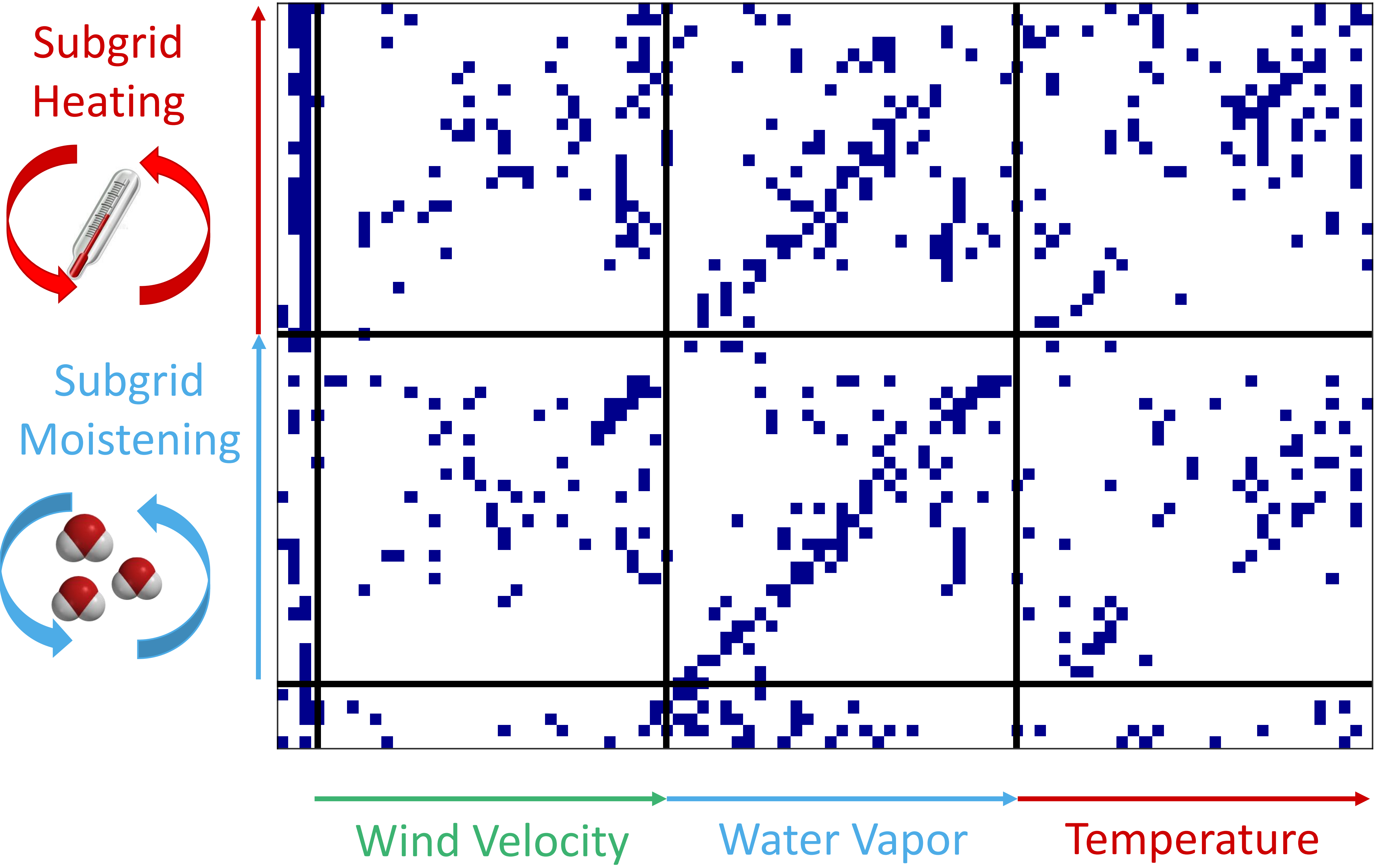
Adapted from: *Fernando Iglesias-Suarez (DLR)*

Before PC1: Fully-connected

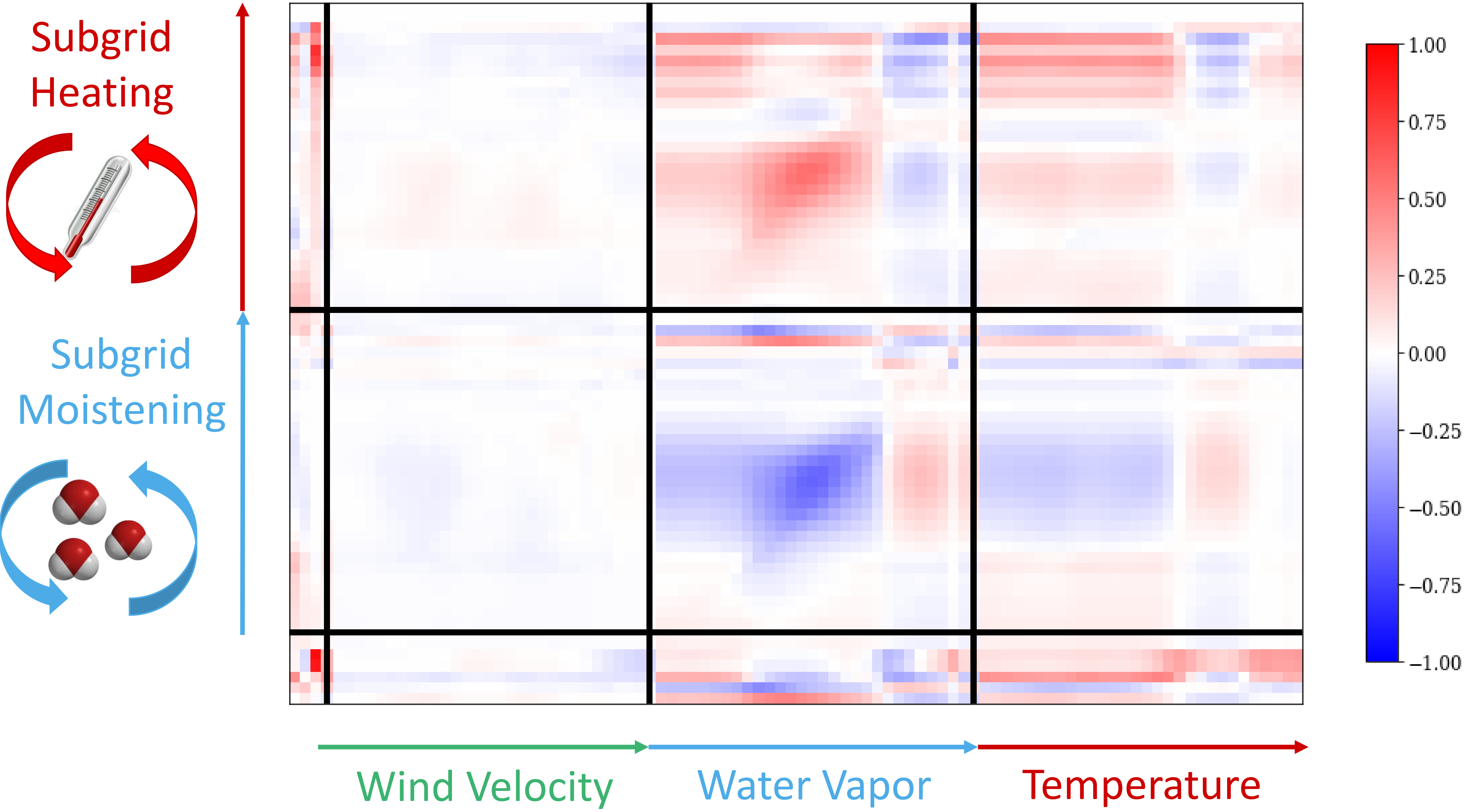


Fully-connected Inputs-to-Outputs

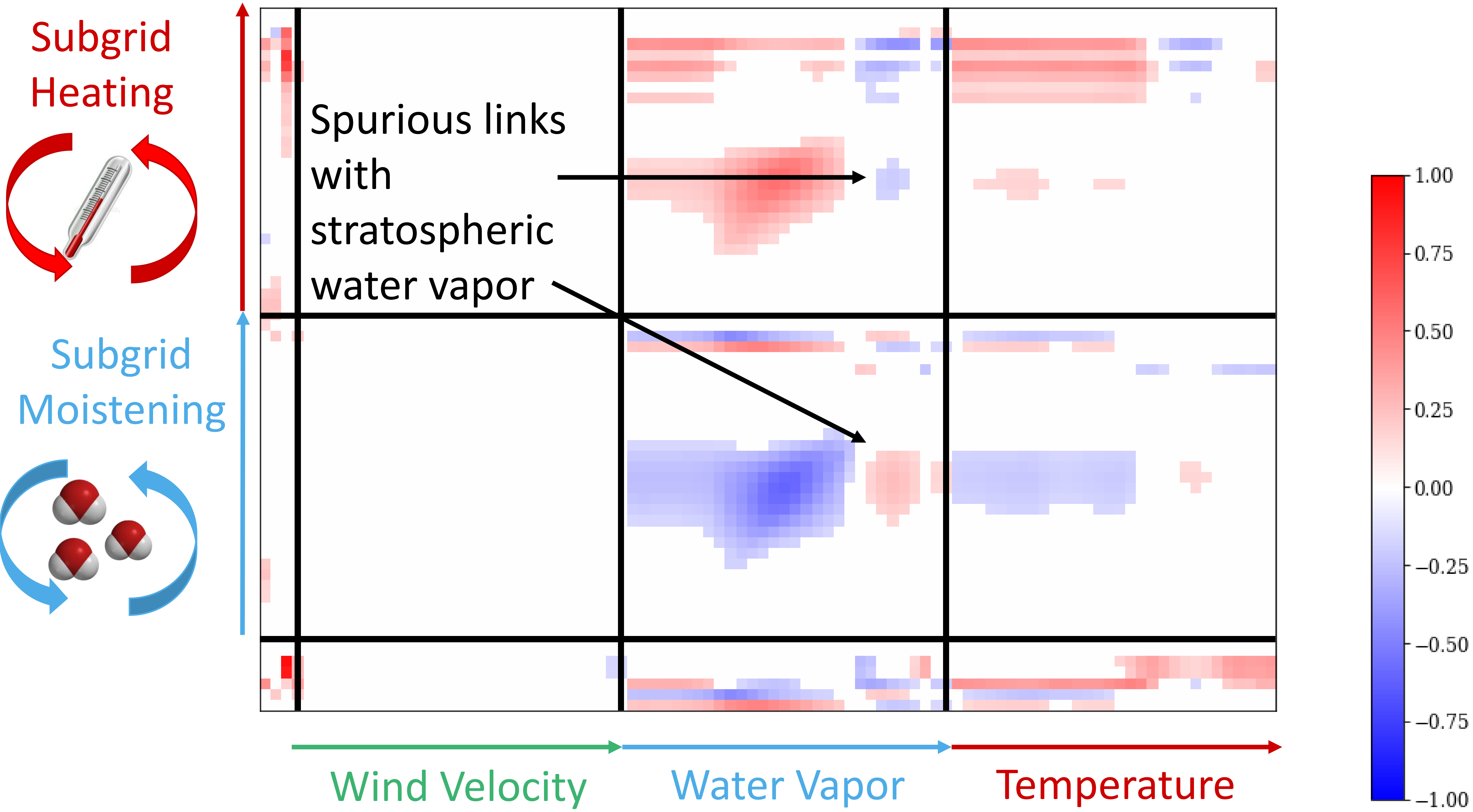
Source: Fernando Iglesias-Suarez (DLR), See: Spirtes and Glymour (1991)



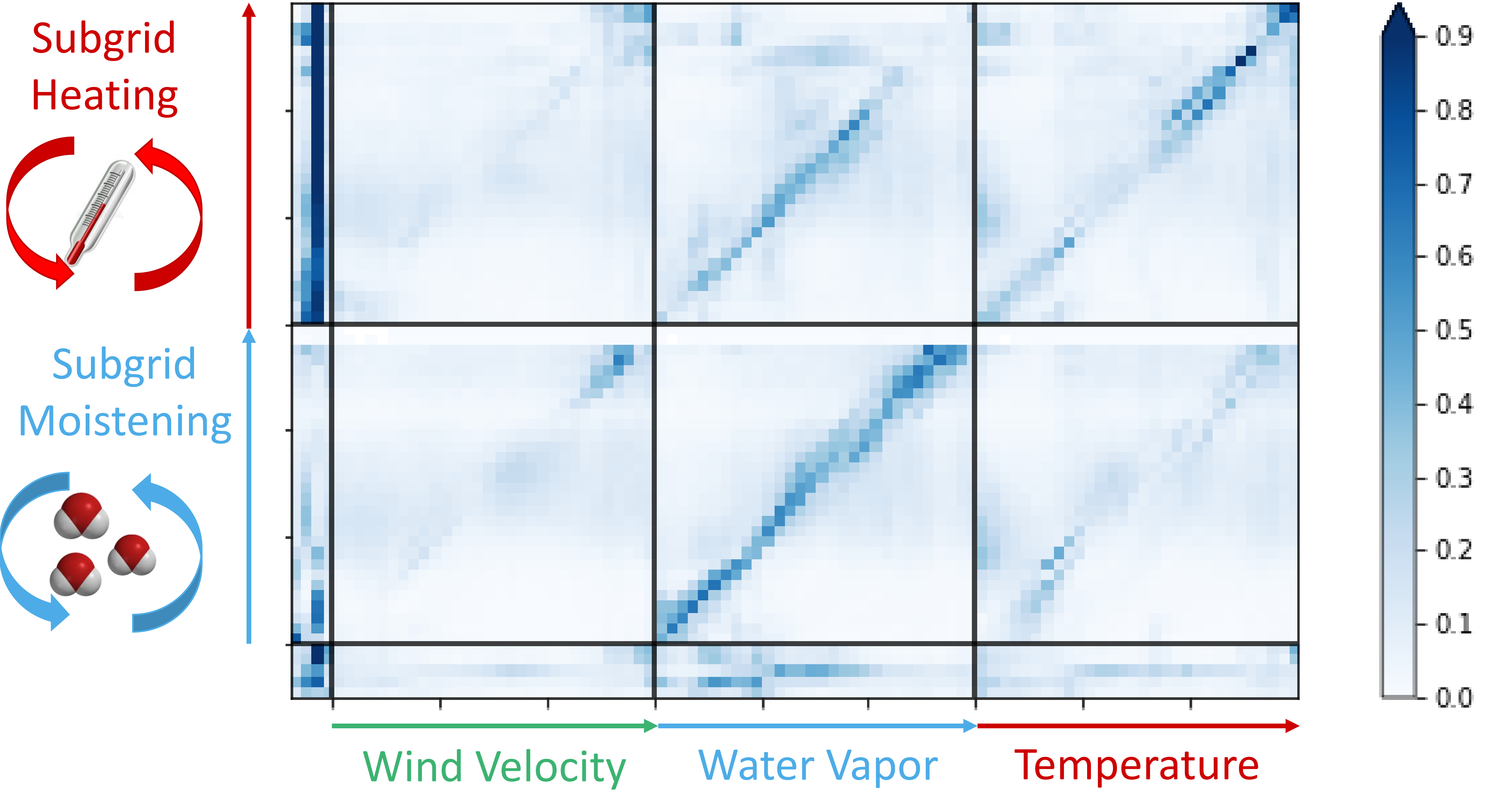
Subgrid Heating
Subgrid Moistening
Wind Velocity
Water Vapor
Temperature



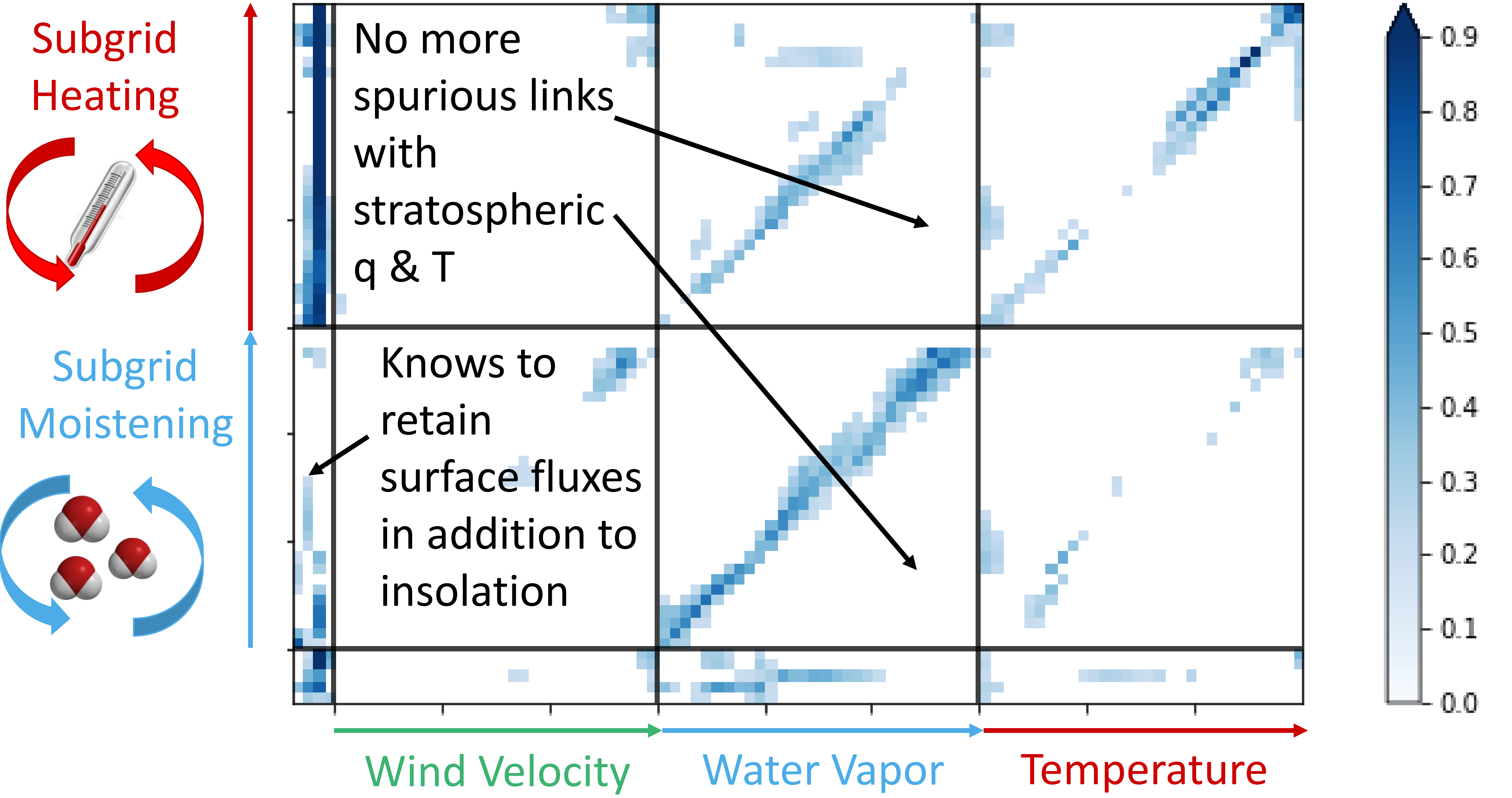
Subgrid
Heating
Subgrid
Moistening
Wind Velocity
Water Vapor
Temperature
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75
−1.00



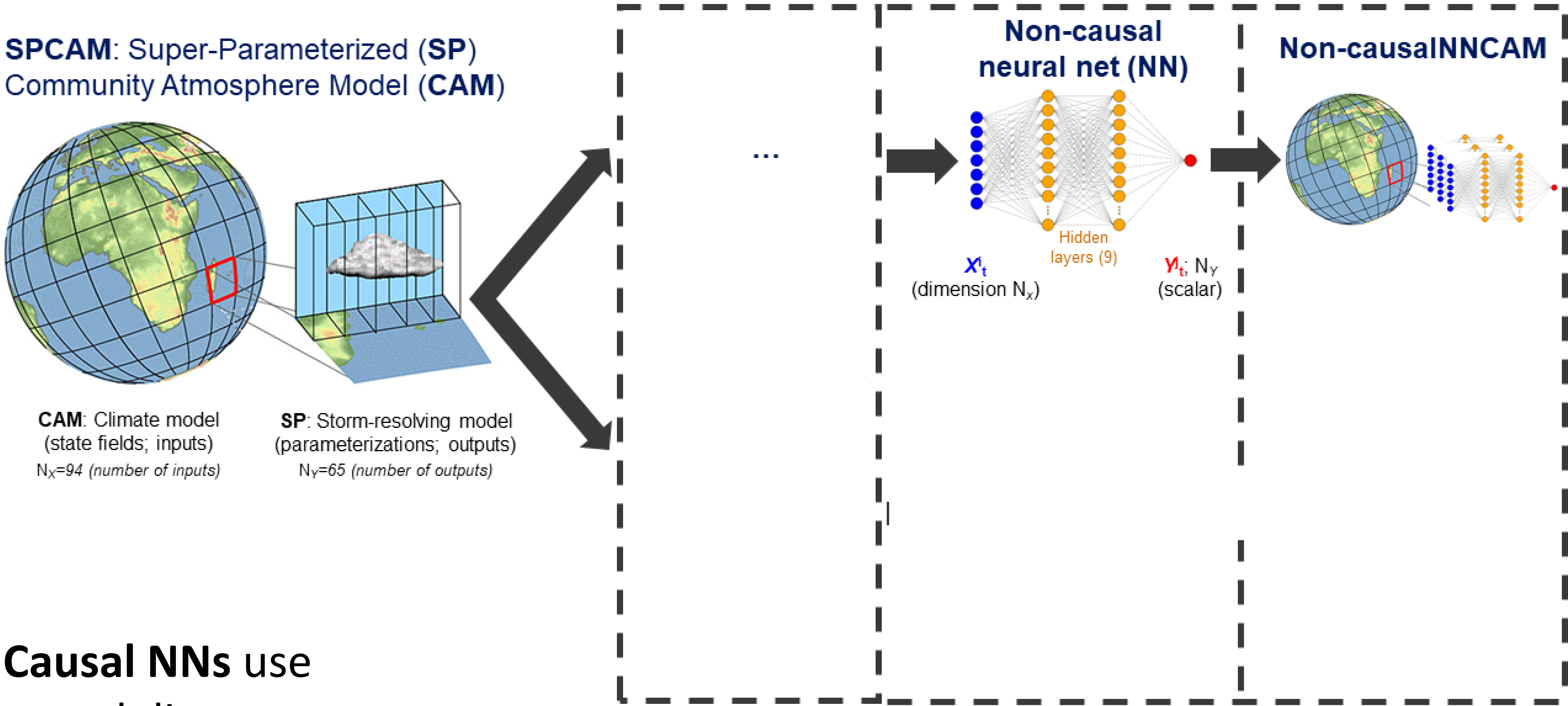
Subgrid Heating
Subgrid Moistening
Spurious links with stratospheric water vapor
Wind Velocity
Water Vapor
Temperature
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75
−1.00



Source: Fernando Iglesias-Suarez (DLR)



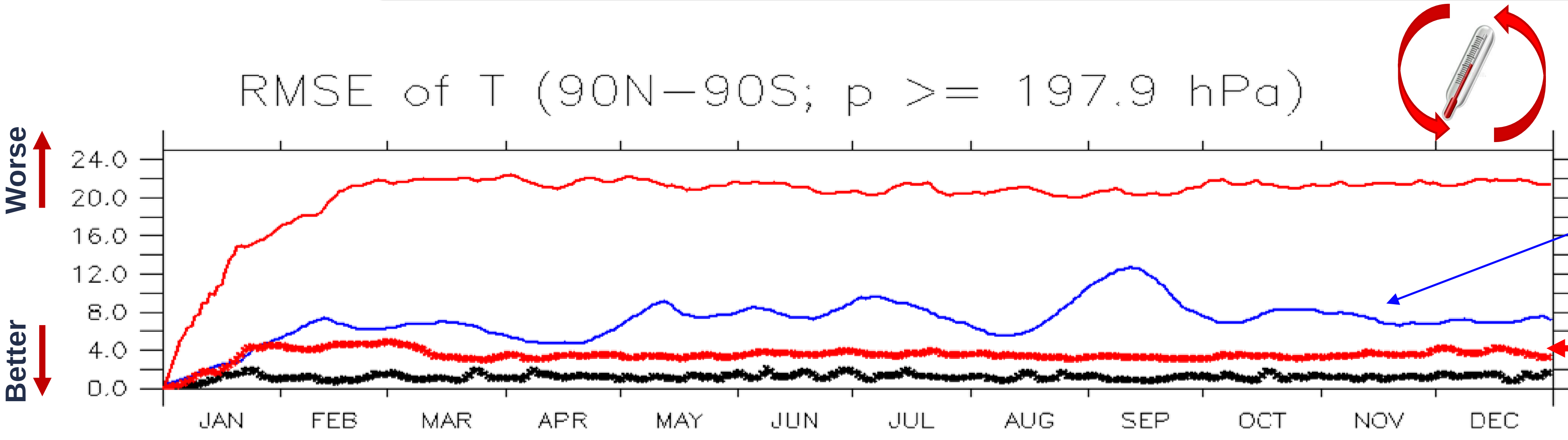
Source: Fernando Iglesias-Suarez (DLR)



Causal NNs use causal discovery to eliminate spurious inputs for each output

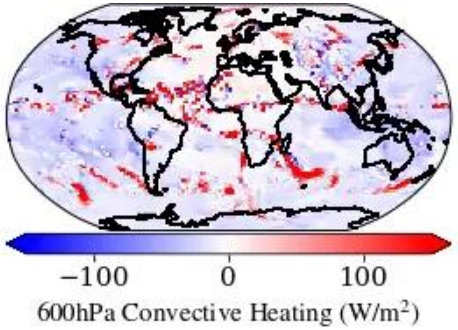
Source: Fernando Iglesias-Suarez (DLR)

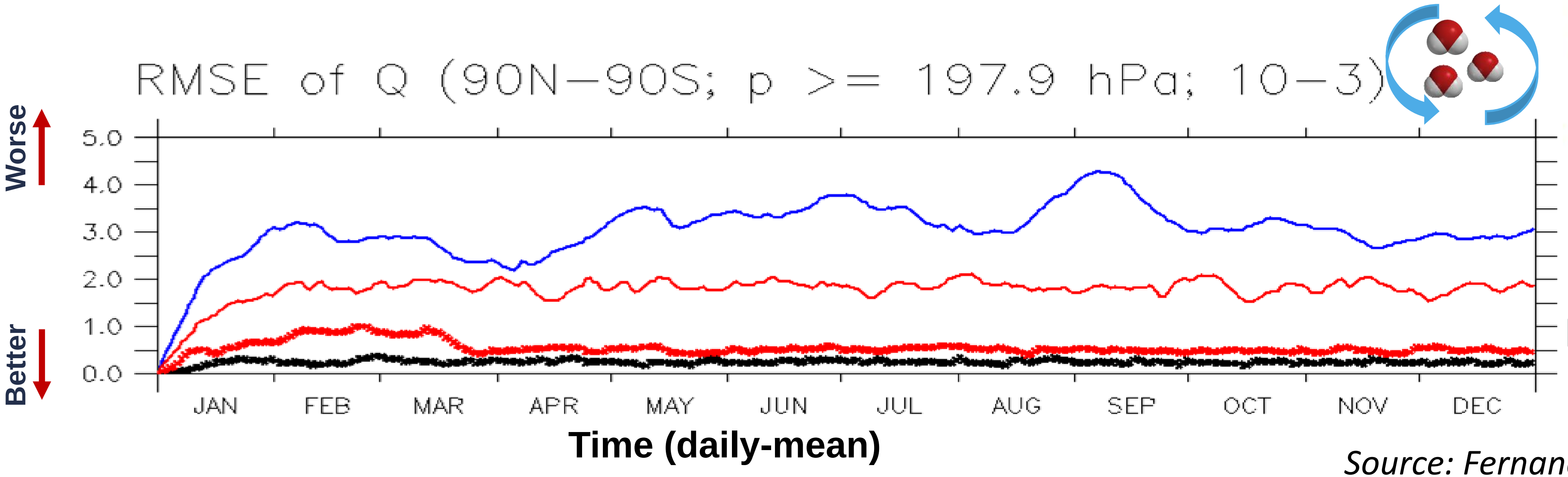
Causally-Informed Neural Nets are Stable Online with Low Error



Brute-force

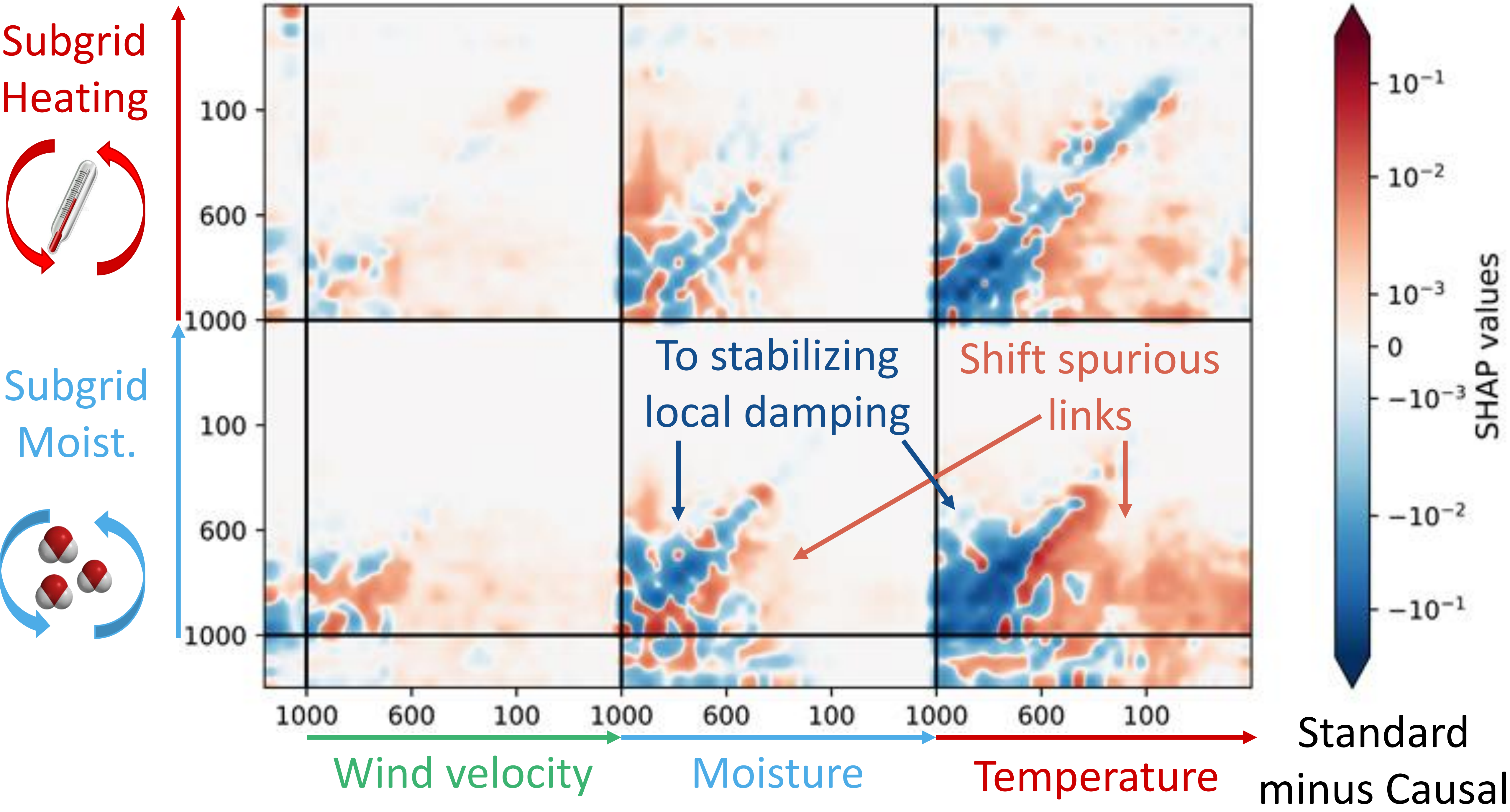
Causally-informed





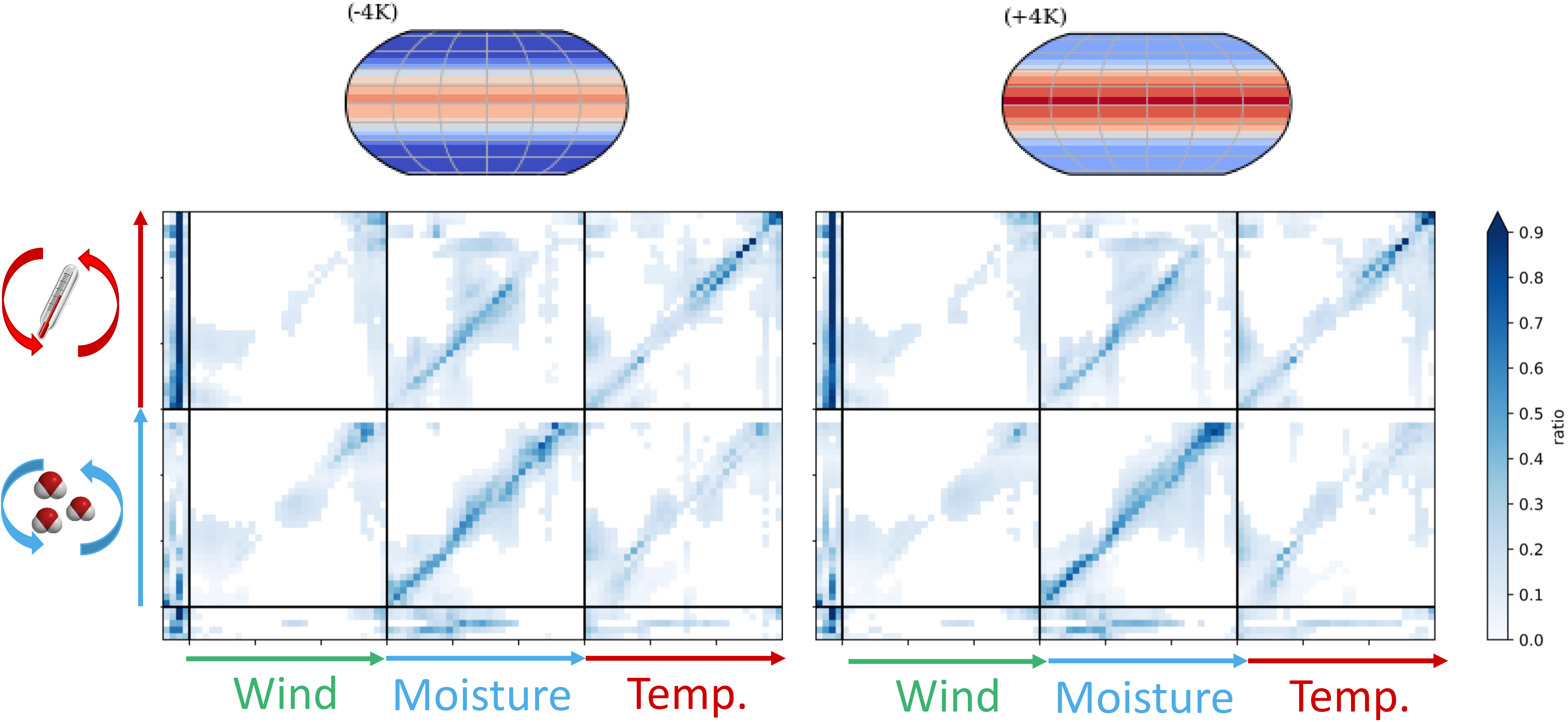
The model runs smoothly

Source: Fernando Iglesias-Suarez (DLR)



Subgrid Heating
Subgrid Moist.
100
600
1000
100
600
1000
To stabilizing local damping
Shift spurious links
1000
600
100
1000
600
100
1000
600
100
Wind velocity
Moisture
Temperature
10^{-1}
10^{-2}
10^{-3}
0
-10^{-3}
-10^{-2}
-10^{-1}
SHAP values
Standard minus Causal

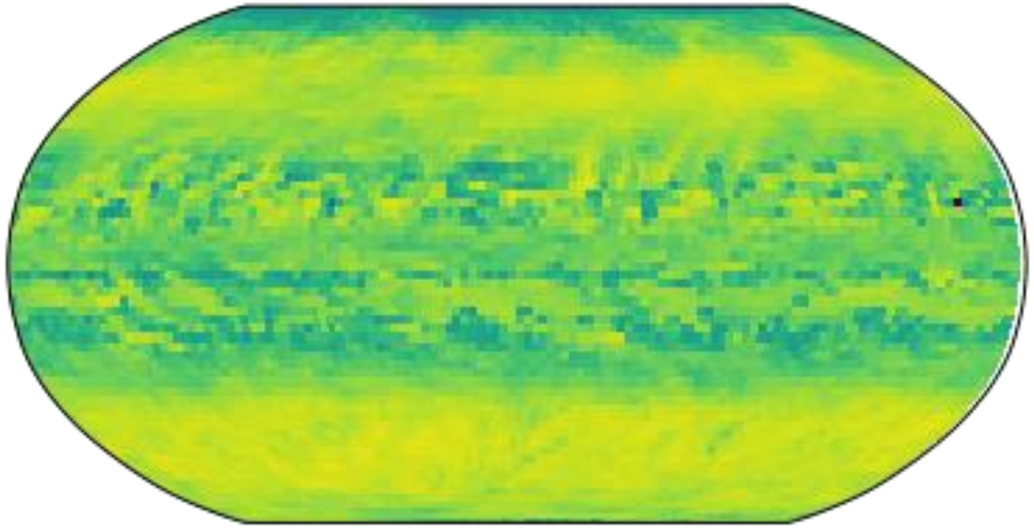
Possible Explanation: Causal Links are Climate-Invariant



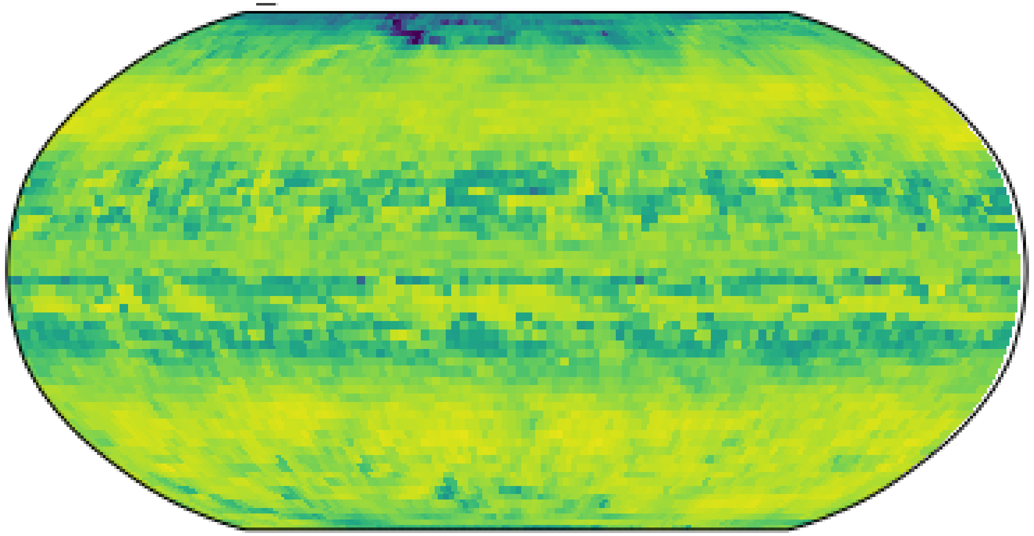
Source: *Fernando Iglesias-Suarez (DLR)*

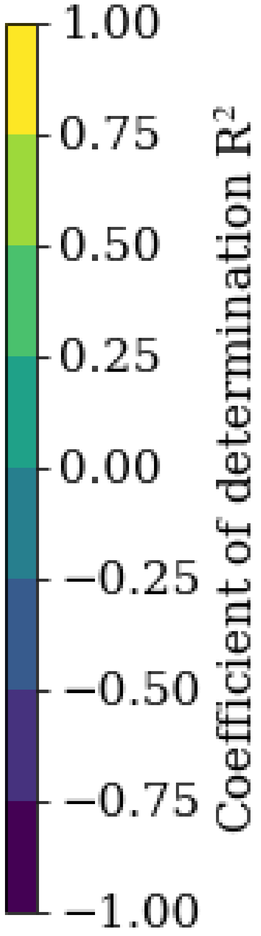
Physically and Causally-Informed Neural Nets generalize to warmer climates with ≈10% of inputs

Climate-Invariant
(94 inputs)



Climate-Invariant
+ Causally-Informed
(11 inputs)





Mid-Tropospheric Subgrid Heating

Outlook: **Multidata** causal feature selection: Causal ML for tropical cyclone prediction

Causal Markov Condition:

dependence ⇒ connectedness

Faithfulness assumption:

independence ⇒ no causal link

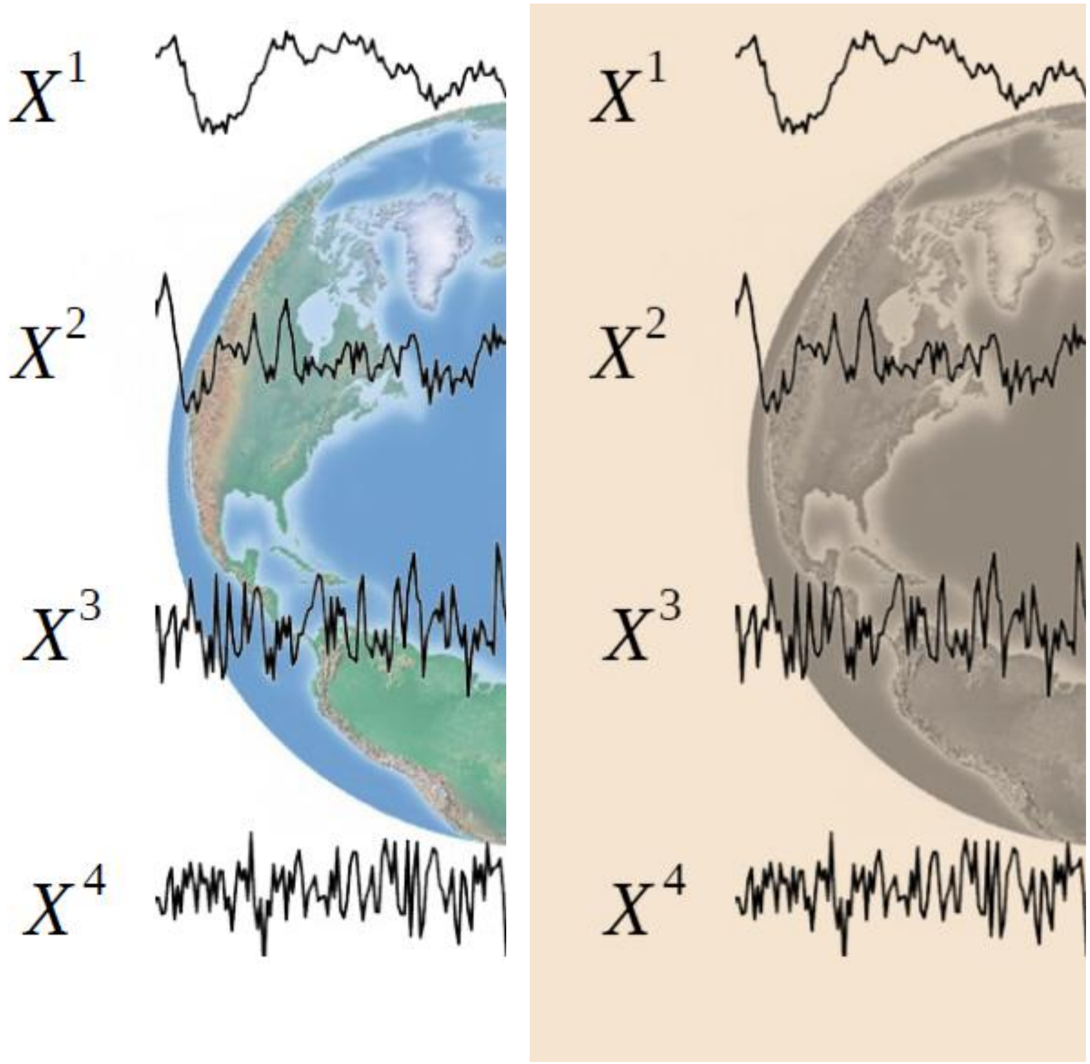
Causal sufficiency:

All common causes are "observed"

Causal stationarity:

Relationships C^t throughout time series

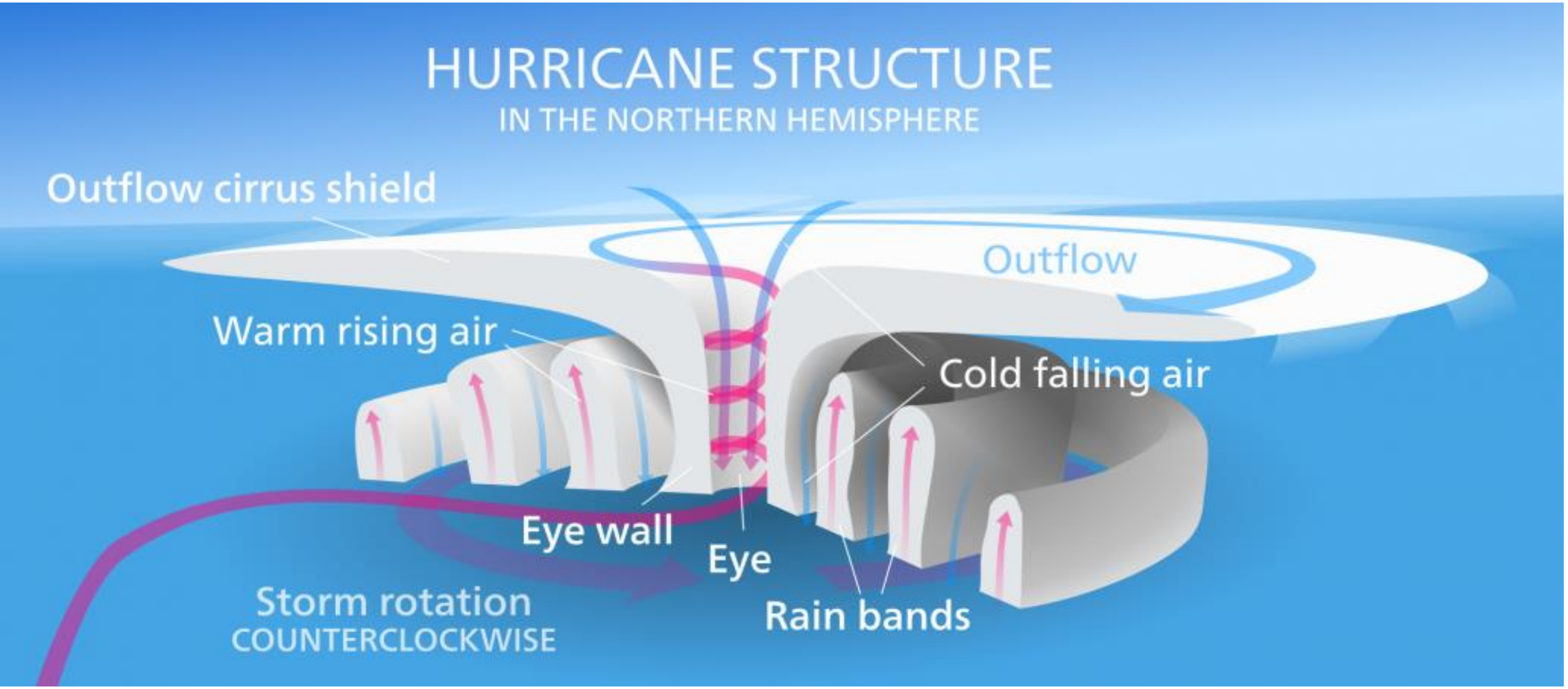
Common causal structure



Adapted from: *Fernando Iglesias-Suarez (DLR)*

<u>Outlook:</u> **Multidata** causal feature selection: Causal ML for tropical cyclone prediction

Feature = Meteorological variable + vertical lvl + horizontal sector + Time lag



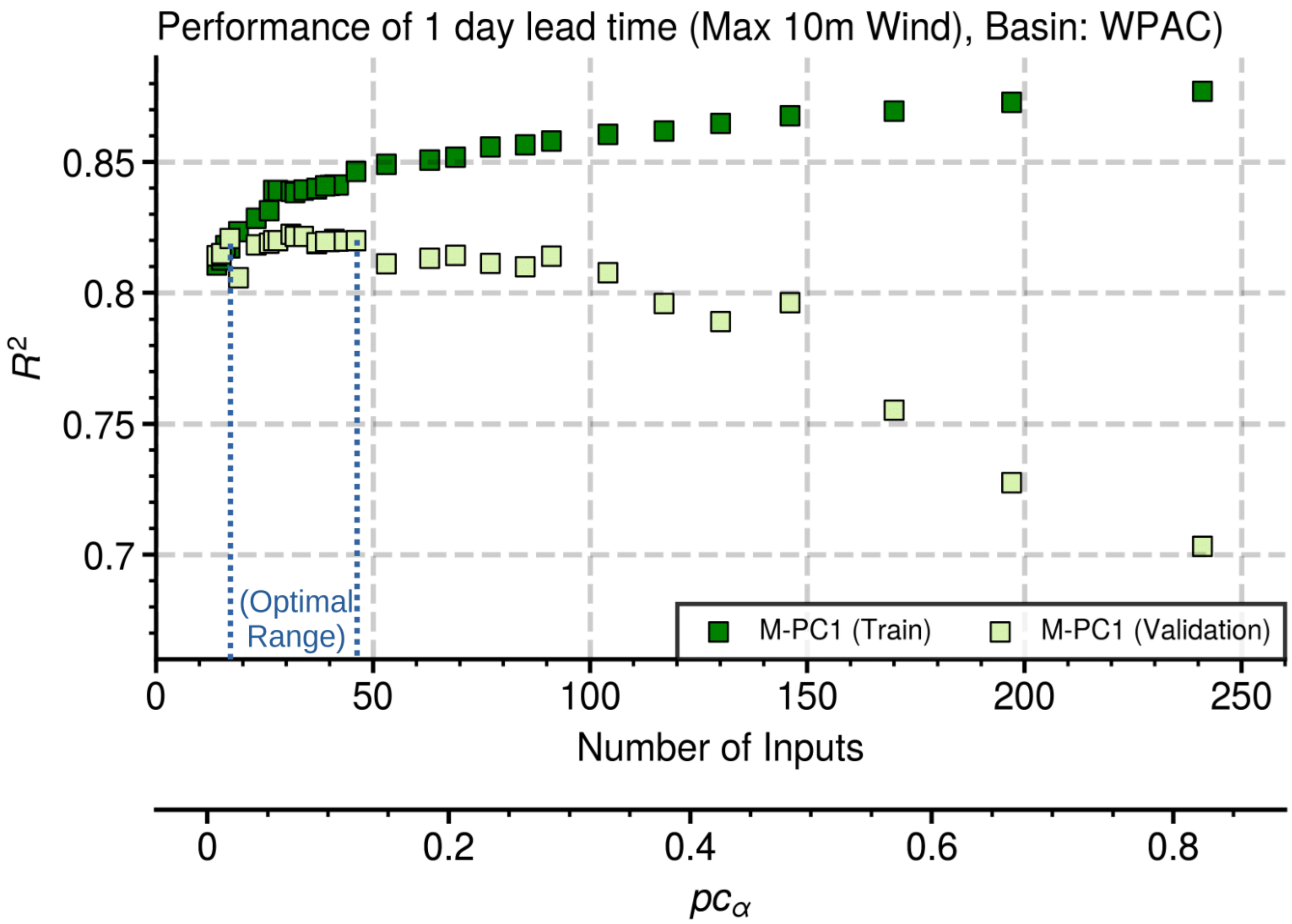
- Hor. Divergence
- Vertical velocity
- Relative vorticity
- Relative humidity
- Geopotential z
- Eq. pot. Temp.
- Wind shear
- Column integrals

<u>Image Source</u>: internetgeography.net

The optimal predictor set helps ML generalize to an independent validation set (50 separate TCs)

Multidata causal feature selection outperforms:

1. Random feature selection
2. XAI-based feature selection (Random forest)
3. Lagged correlation-based feature selection

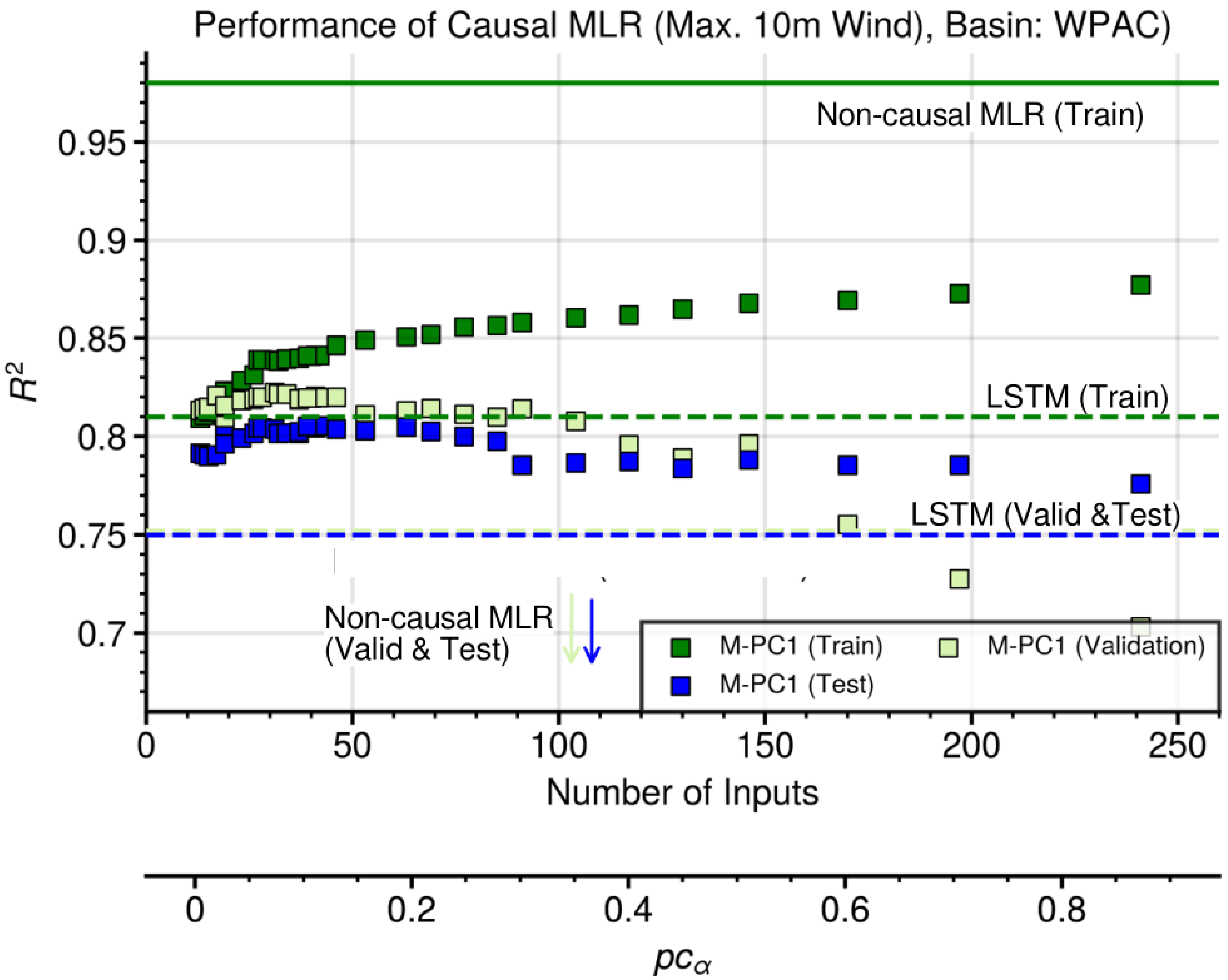


<u>See</u>: *Ganesh S. et al. (2023, arXiv:2304.05294)*

The optimal predictor set helps ML generalize to an independent validation set & improves performance

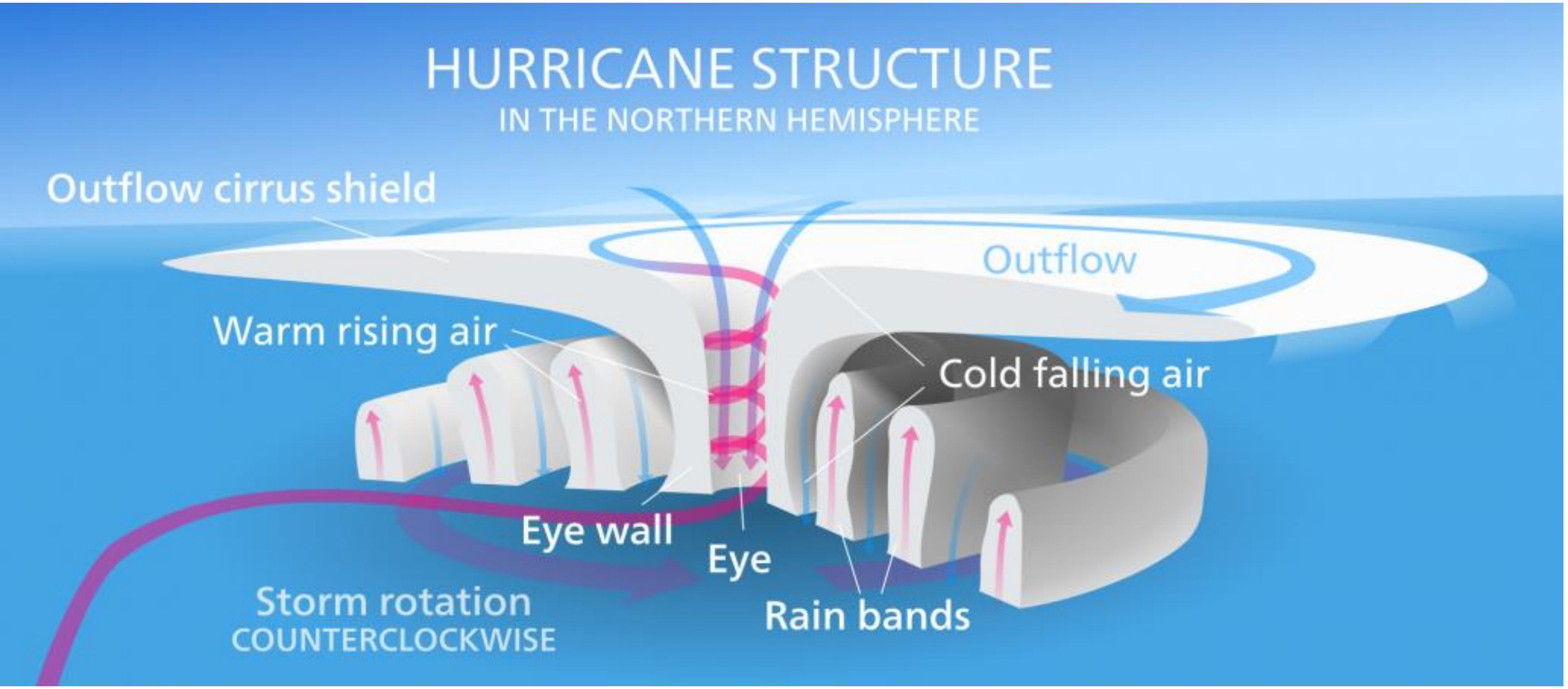
Multidata causal feature selection outperforms:

1. Random feature selection
2. XAI-based feature selection (Random forest)
3. Lagged correlation-based feature selection

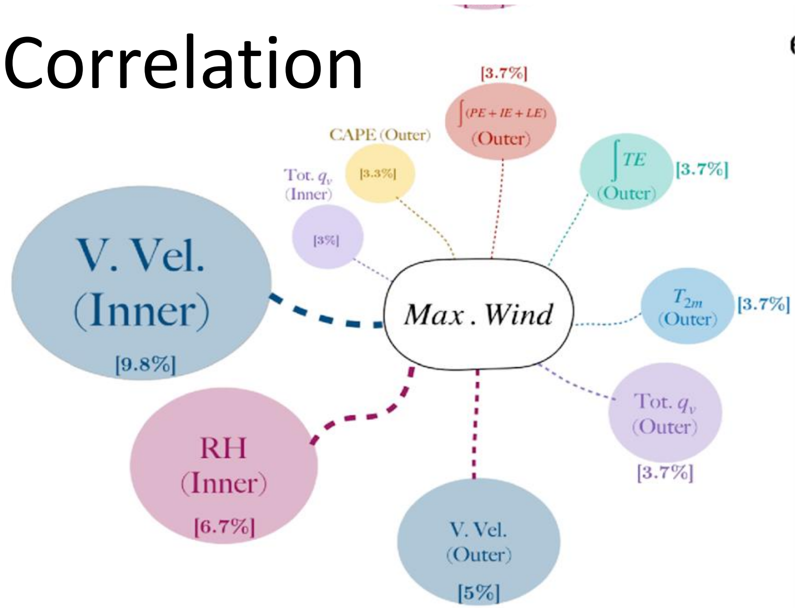


See: Ganesh S. et al. (2023, arXiv:2304.05294)

Causal feature selection not only removes spurious links, but also suggests new predictors for TC intensity



Correlation



<u>See</u>: Ganesh S. et al. (2023, arXiv:2304.05294)

Conclusion

1. Generalization: Physically rescaling the inputs (and outputs) of neural networks helps them generalize to unseen climates and geographies

→ Test climate-invariant neural nets online, directly train on observations

2. In absence of physical knowledge, causal discovery helps objectively & a priori select inputs to train more parsimonious models, which lead to stable coupled simulations with reduced drifts

3. Physically-informed + (multidata) causally-informed ML helps create Physically-consistent + generalizable + causal models

→ Applicable to different settings (tropical cyclones, heatwaves)

When do we need physical knowledge and causal consistency?

Thank you

www.unil.ch/dawn

tom.beucler@unil.ch

lms.ecmwf.int