# A Process-Informed Determination of Credibility Across Different Downscaling Methods

**Melissa S. Bukovsky[1],**

**Seth A. McGinnis[2], Rachel R. McCrary[2], Linda O. Mearns[2]**

[1]University of Wyoming, Laramie, WY, USA

[2]NSF National Center for Atmospheric Research, Boulder, CO, USA

DOE EESM PI Meeting

U.S. Department of Energy | Office of Science
HYPERFACETS
A Framework for Improving Analysis and Modeling of Earth System and Intersectoral Dynamics at Regional Scales

UW | Haub School of Environment and Natural Resources

NSF

NCAR
Operated by UCAR

# Models/Methods

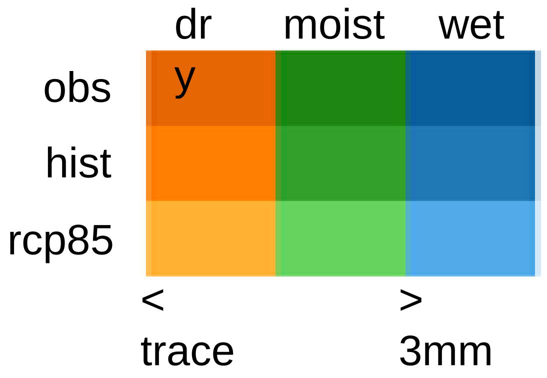
- No downscaling (just GCMs)
  - **Raw**
  - MPI-ESM-LR, GFDL-ESM2M, HadGEM2-ES
- Dynamical (NA-CORDEX)
  - **RegCM4**
  - **WRF** (nudged)
- Statistical – spatial
  - **CNN** (basic U-Net)*****
  - **LOCA***

- Statistical – point
  - **SDSM** (multivariate regression)*****
  - **qdm** (quantile delta mapping)

# Focus on wet/dry day occurrence & the ‘ingredients’ for precipitation, in May in one location.

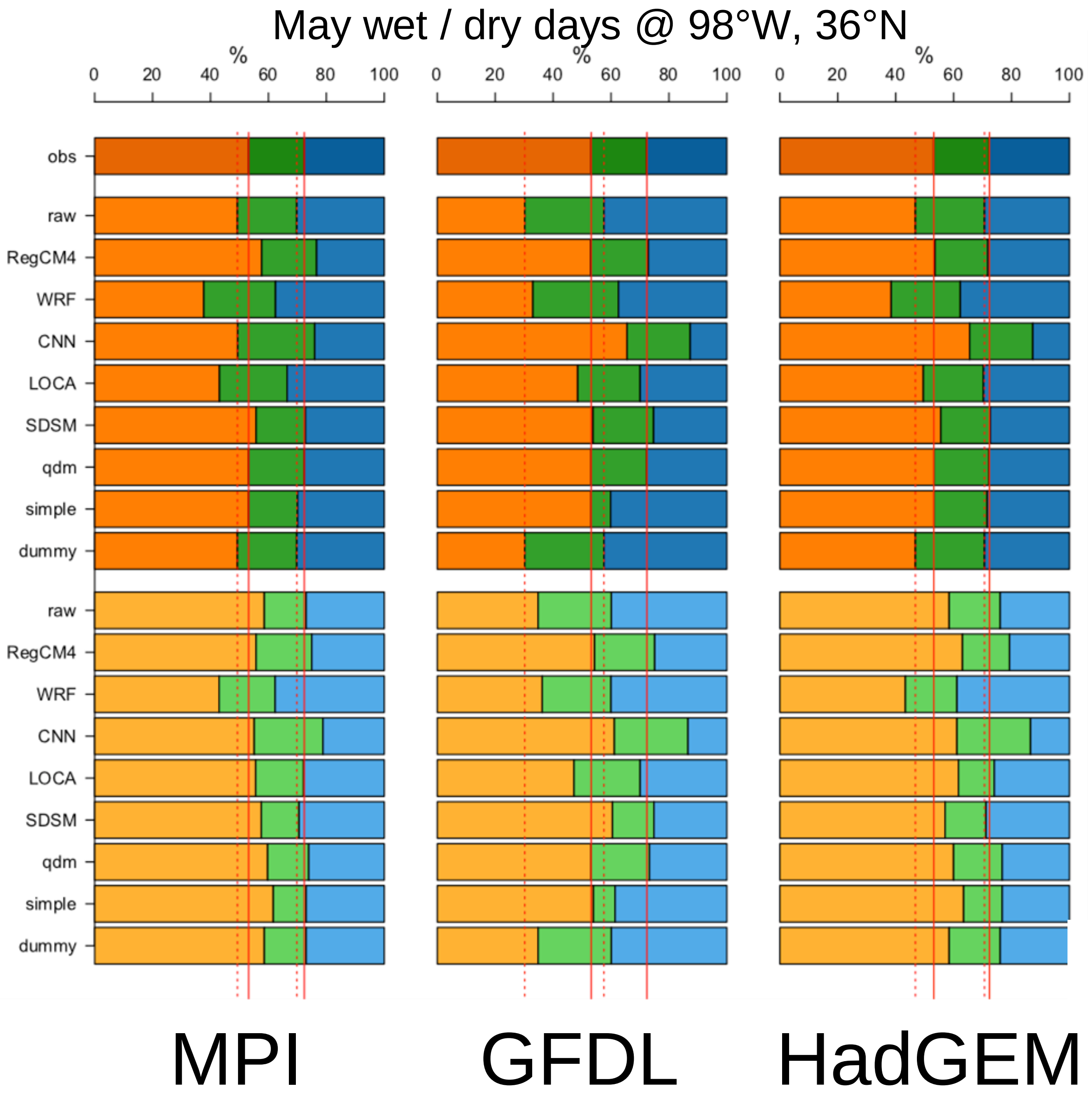
Thresholds:

- Wet > 3mm/day
- Dry < 0.254mm




This focus:

- Simplifies the problem
- Reduces influence of model intensity bias
- Focuses on atmospheric setup for precipitation




**3 CMIP5-era GCMs**

## We evaluated precipitation occurrence in May in the Southern Great Plains (SGP)

Precipitation at this test point (◆) results from a combination of processes that provide moisture & lift:

a) Moisture advection by SGP low-level jet

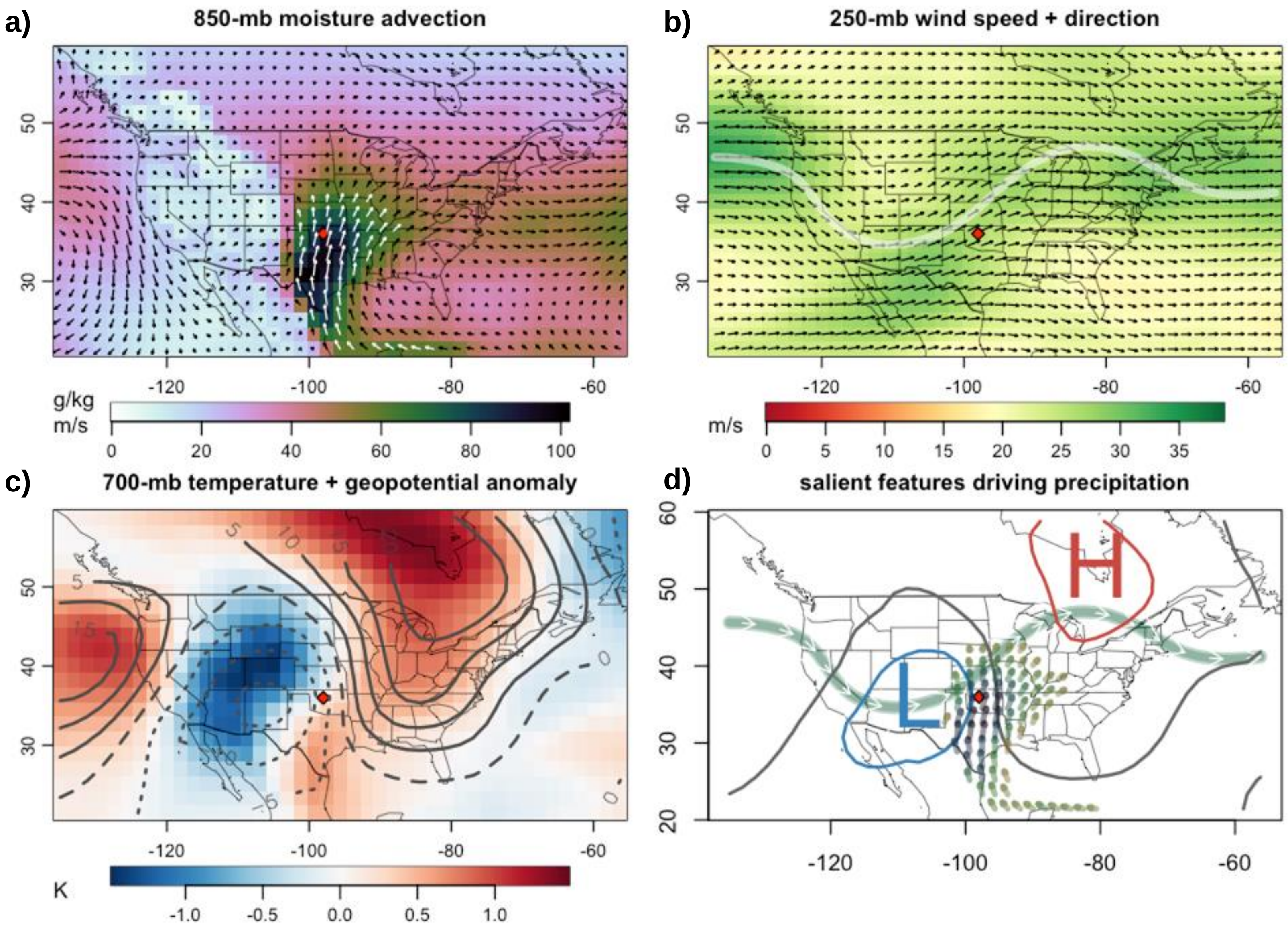
b) Upstream of an upper-level trough in the jetstream, ahead of a jet max.

c) Upstream of mid-level trough.

d) Diagram view of a-c

*Do the downscaling methods rain on days when it makes sense for it to rain?*

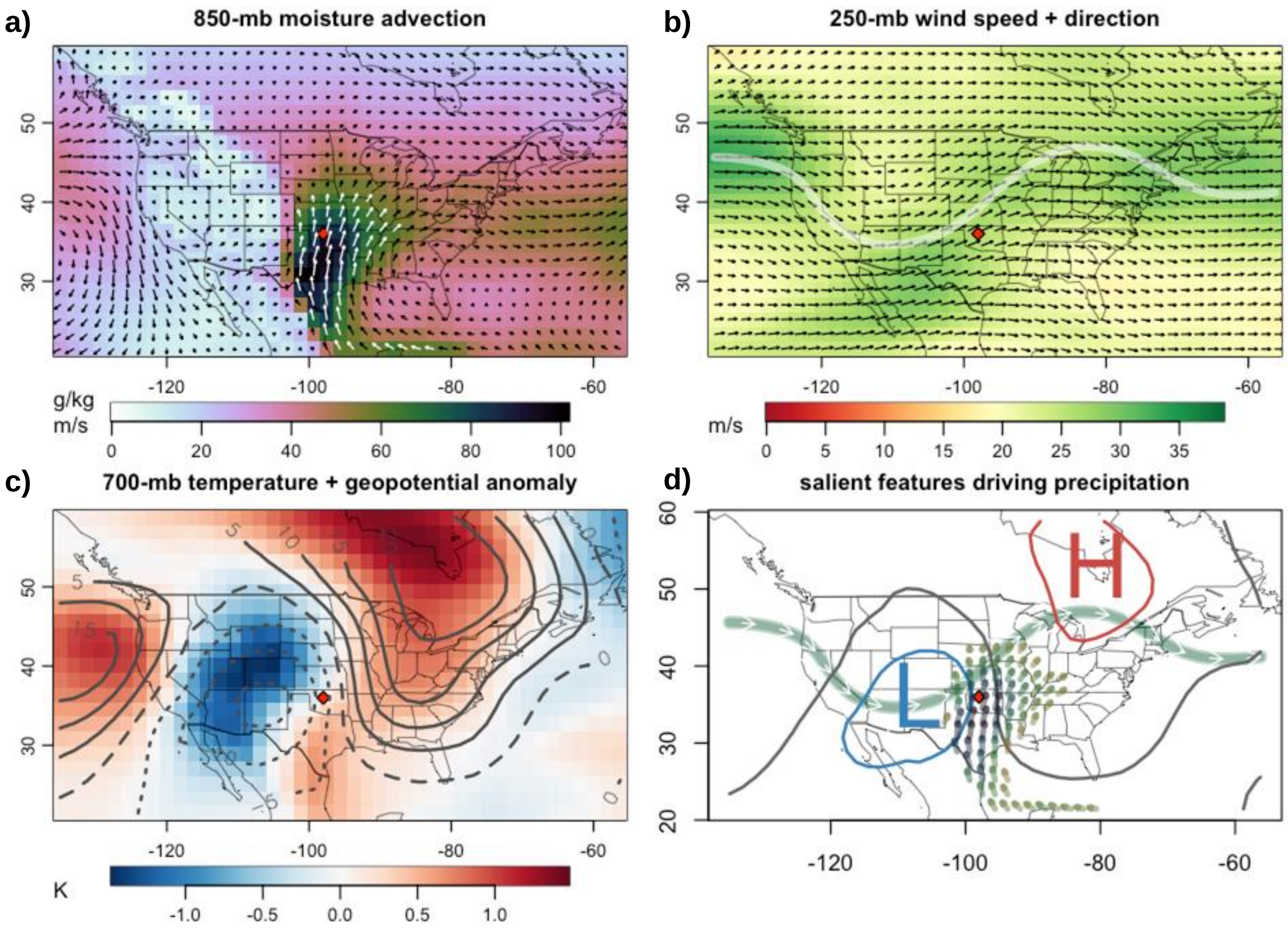
## Observed May Wet-Day Climatology, 1980-2005



Comparison methods:

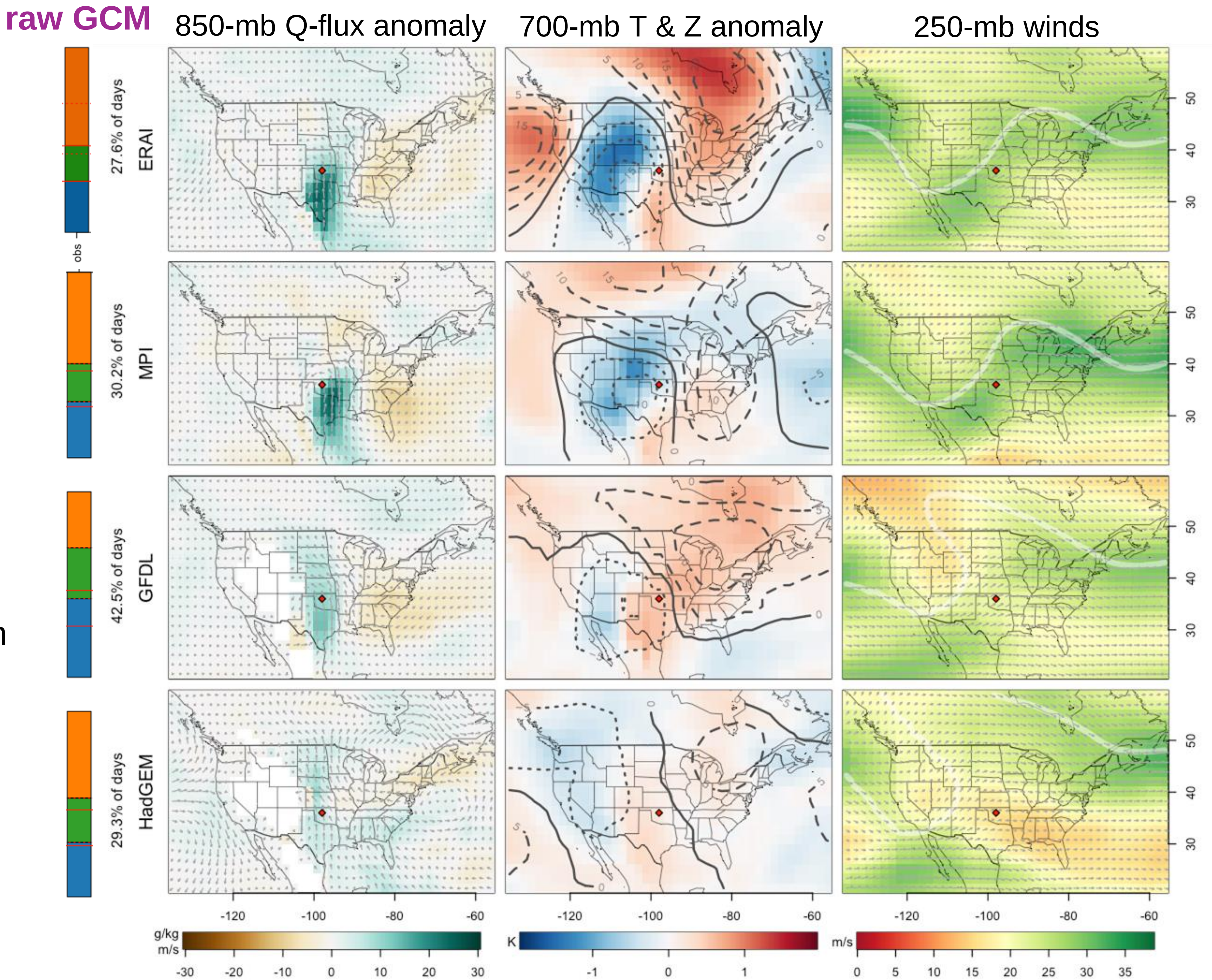
- GCM large-scale forcing + precipitation in GCMs
- GCM large-scale forcing + statistical method wet days
- RCM large-scale forcing + RCM wet days

## Observed May Wet-Day Climatology, 1980-2005



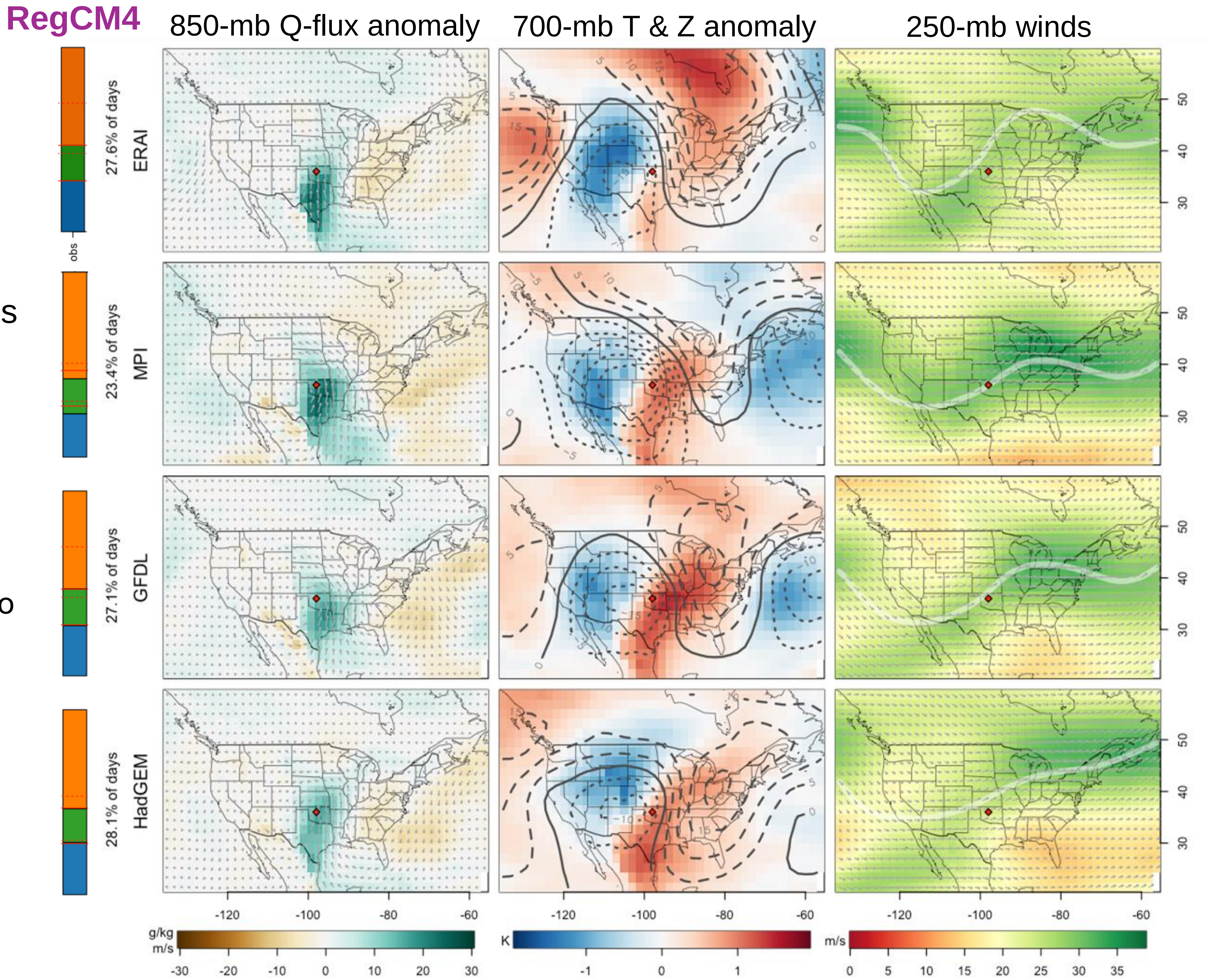
To evaluate credibility, we average upper-level conditions on test-point wet-days.

- **MPI** (top) ★★★★★
- **GFDL** (middle) ★★★☆☆
  - Okay but weak, messy
  - Note: more wet days
- **HadGEM** (bottom) ★☆☆☆☆
  - Weak moisture flux
  - Poor mid-level spatial pattern
  - Jet stream shifted N
- This baseline credibility is inherited by all downscaling methods.



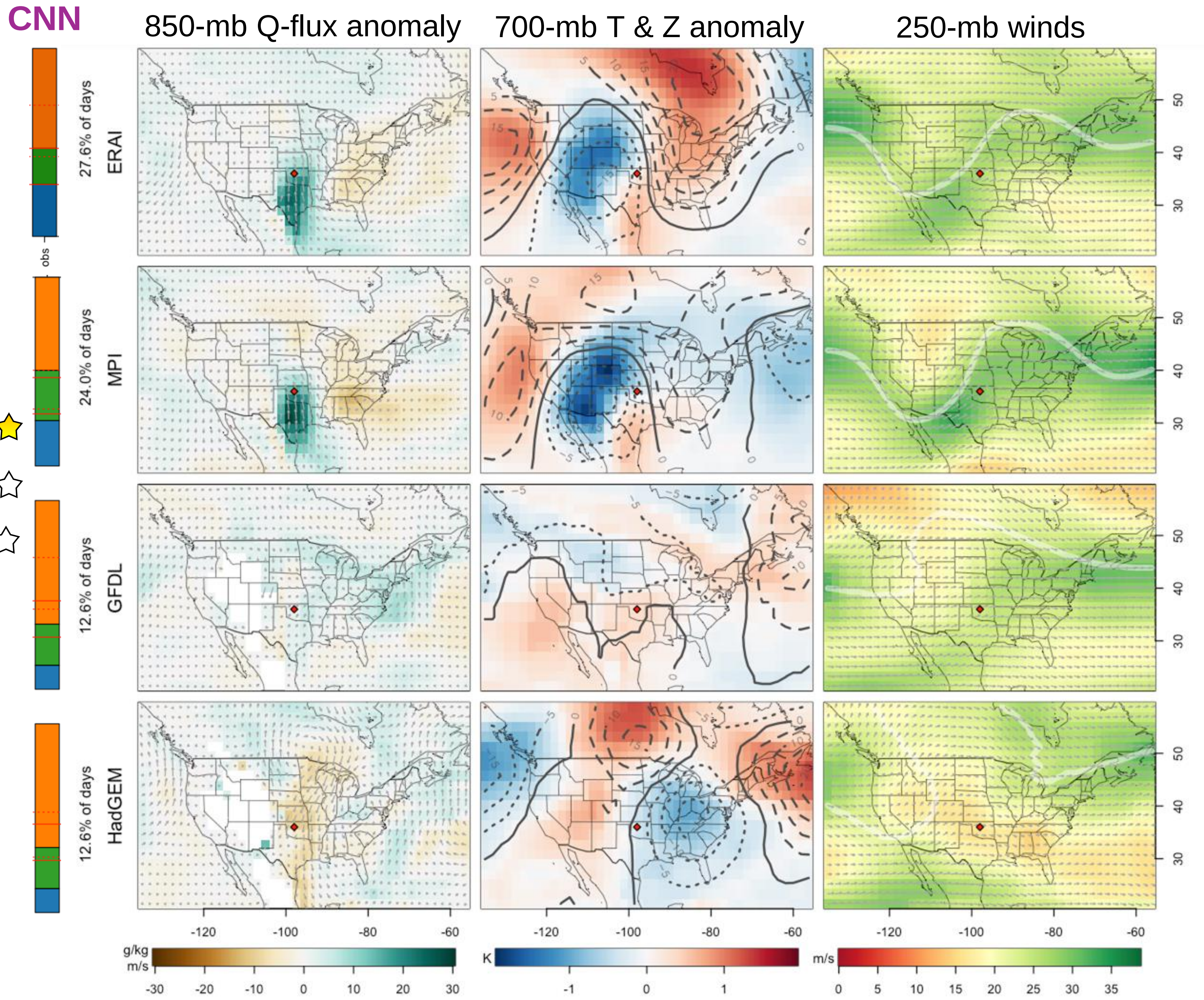
# Dynamical downscaling can improve on bad GCM inputs.

- **GOOD: ALL** ☆☆☆☆☆
- All 3 downscaled GCM simulations now show strong moisture flux anomaly & sensible spatial patterns and magnitudes
- Jetstream placement has also been improved
  - Note: WRF can't fix the jet stream to the same extent because those simulations are nudged for wavelengths > 2000km



# Complex statistical downscaling can make things worse.

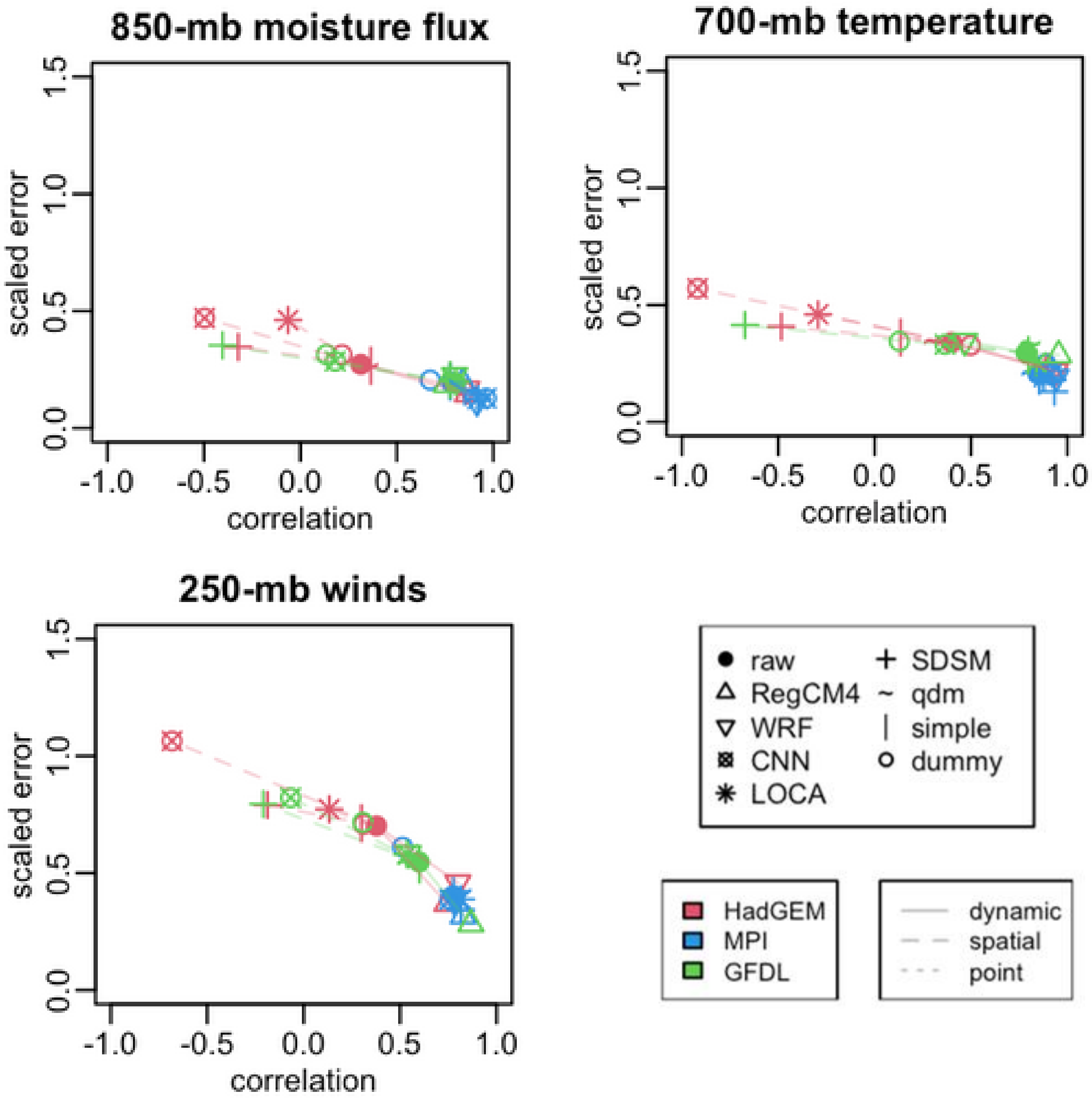
- CNN exaggerates the patterns & inherited credibility.
  - MPI still **GOOD** ★★★★★
  - GFDL now **BAD** ★☆☆☆☆
  - HadGEM **VERY BAD** ☆☆☆☆☆
- Statistical methods often do poorly when asked to extrapolate beyond the data they were trained on.



# We summarize the wet-day composite plots with spatial correlation and scaled MAE vs obs

- All methods perform well (high cor, low MAE) when the GCM is good (blue)
- Dynamical methods (△▽) can improve on the parent GCMs
- Simple point-based statistical methods ( | ~ ) stick close to the parent GCM
- When the GCM is bad (red), complex statistical methods (crossed symbols) can reduce already-low credibility

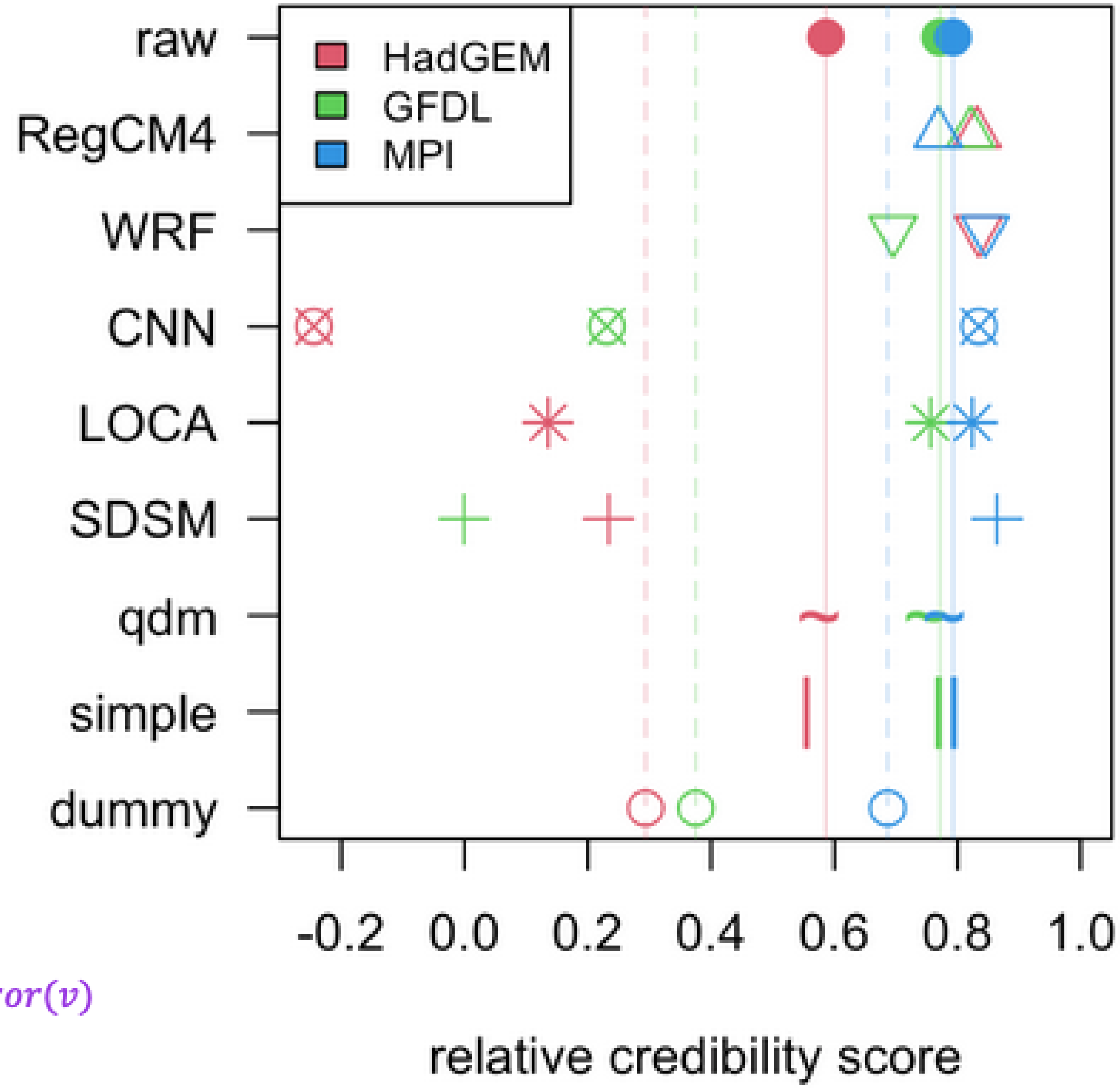
Subset for example purposes, all fields from previous slides completed □



# Then we metricized this.

- CNN, LOCA, and SDSM can perform worse than random noise (dummy method) when downscaling poorly performing “bad” GCMs (e.g., HadGEM), and often perform considerably worse when downscaling moderately performing “ugly” GCMs (e.g.,GFDL).
- However, for well performing GCMs (MPI), they also perform well and are as credible as the dynamical downscaling methods, but *only* for well performing ”good” GCMs.
- Simpler statistical methods, qdm and simple, inherit the credibility of the GCM.
- RCMs add value.

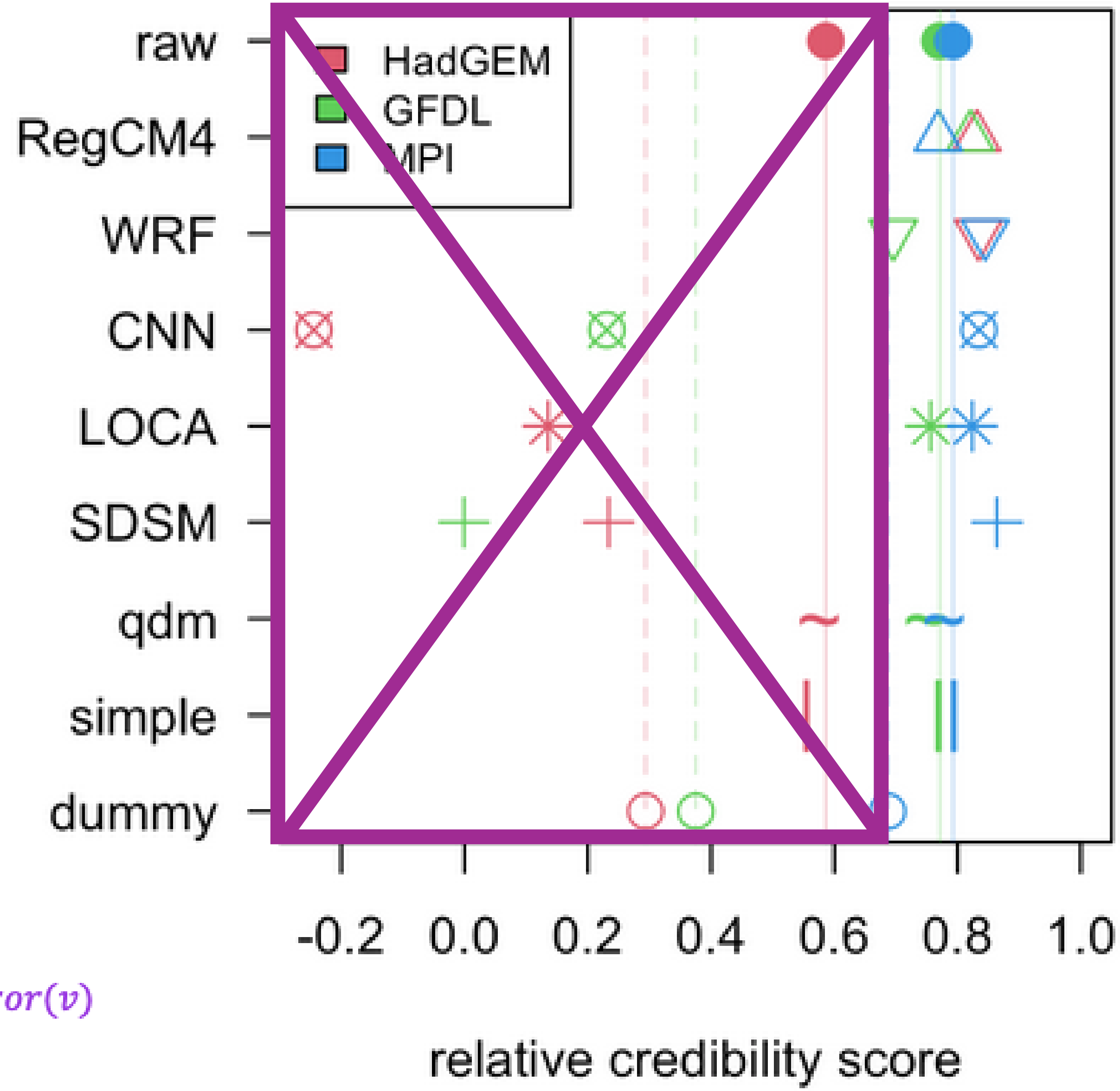
$$relative\ credibility = \frac{1}{2{*}n} \sum_{v=1}^{n} correlation(v) + 1 - scaled\ error(v)$$



# Then we metricized this.

- CNN, LOCA, and SDSM can perform worse than random noise (dummy method) when downscaling poorly performing “bad” GCMs (e.g., HadGEM), and often perform considerably worse when downscaling moderately performing “ugly” GCMs (e.g.,GFDL).
- However, for well performing GCMs (MPI), they also perform well and are as credible as the dynamical downscaling methods, but *only* for well performing ”good” GCMs.
- Simpler statistical methods, qdm and simple, inherit the credibility of the GCM.
- RCMs add value.

$$\text{relative credibility} = \frac{1}{2*n} \sum_{v=1}^{n} \text{correlation}(v) + 1 - \text{scaled error}(v)$$



# SUMMARY

- The output from a downscaling method can be more or less credible based on how it responds to the input errors.  Here:
  - All methods perform well when the GCM is good.
  - Dynamical methods can improve on the parent GCM’s credibility.
  - Simple point-based statistical methods stick close to the parent GCM.
  - When the GCM has low credibility, complex statistical methods can make it even worse.
- These results suggest the complex statistical methods warrant further scrutiny.
  - What do these results mean for the projections from these methods?  An environment undergoing climate change might not be that different from one that is shifted due to bias.
- Projections inherit historic credibility.
  - See paper for additional conclusions/discussion regarding projections of future climate.
- This is an example of a framework.
  - There’s room for expansion.
- These results may not hold everywhere.

# DISCUSSION

- To generalize this framework:
  - Pick a downscaled variable that is driven by important resolved processes in a region of interest
  - Stratify data based on values of that variable at a point or over a small sub-region
  - Create composites of the driver variables for the different strata
  - Evaluate composites vs observations & physical process understanding
  - Evaluate plausibility of changes in physical processes

- Future work:

- Generalize and automate to weather regimes.
- Try other regions.

# Acknowledgements

- This work was supported by DOE grant DE-SC0016438
- NCAR is sponsored by NSF under Cooperative Agreement #1852977
- DoD ESTCP for previous support of the NA-CORDEX data archive
- WCRP's WGRC & WGCM and the CORDEX modeling groups for data
- NCAR CISL for computing resources
- *Everyone who has contributed to useful discussions around this topic!*


*Submitted to Earth's Future!*

Melissa.Bukovsky@uwyo.edu

U.S. Department of Energy | Office of Science
HYPERFACETS
A Framework for Improving Analysis and Modeling of Earth System and Intersectoral Dynamics at Regional Scales

UW | Haub School of Environment and Natural Resources

NCAR Operated by UCAR