Forecast Verification: Part III

Probabilistic Forecasts of Discrete Predictands: Brier Skill Score

- Brier Score: the mean squared error of the probability forecasts,

$$\mathrm{BS} = \frac{1}{n}\sum_{k=1}^{n}(y_k - o_k)^2,$$

where the observation O is equal to 1 occurrence and 0 for nonoccurrence.

- *How would you formulate the Brier Skill Score?*
- Brier Skill Score: note that BS is equal to 0 for perfect forecasts

$$\mathrm{BSS} = \frac{\mathrm{BS} - \mathrm{BS}_{\mathrm{ref}}}{0 - \mathrm{BS}_{\mathrm{ref}}} = 1 - \frac{\mathrm{BS}}{\mathrm{BS}_{\mathrm{ref}}},$$

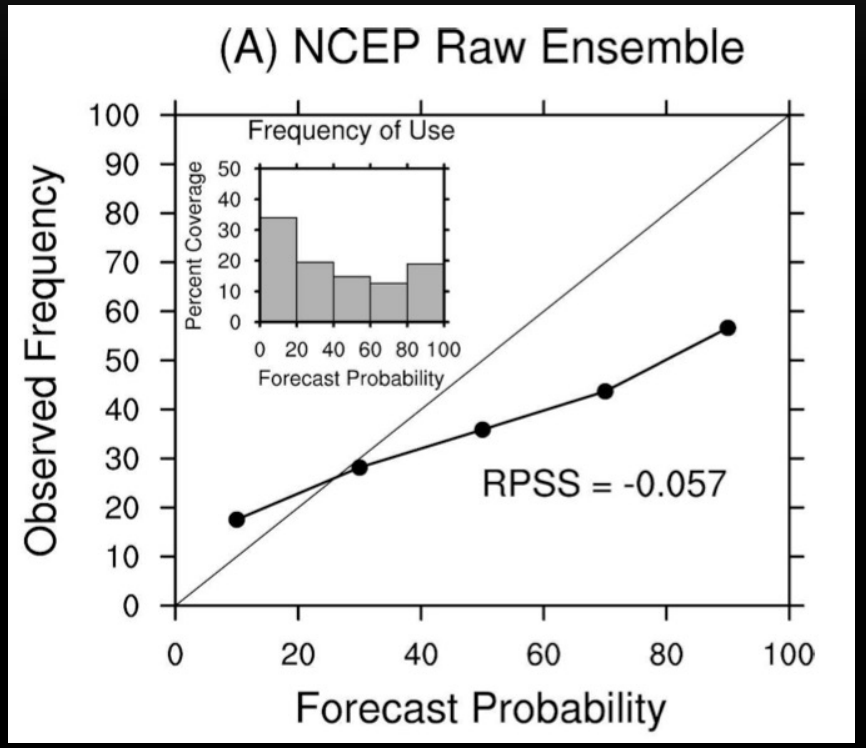
$$\mathrm{SS}_{\mathrm{ref}} = \frac{\mathrm{A} - \mathrm{A}_{\mathrm{ref}}}{\mathrm{A}_{\mathrm{perf}} - \mathrm{A}_{\mathrm{ref}}} \times 100\%,$$

- Usually the reference forecasts are the relevant climatological relative frequencies that may vary with location and/or season. Positive BSS indicates better skill than the reference forecasts.

Wilks 2011

Probabilistic Forecasts of Discrete Predictands (cont'd): Reliability Diagram

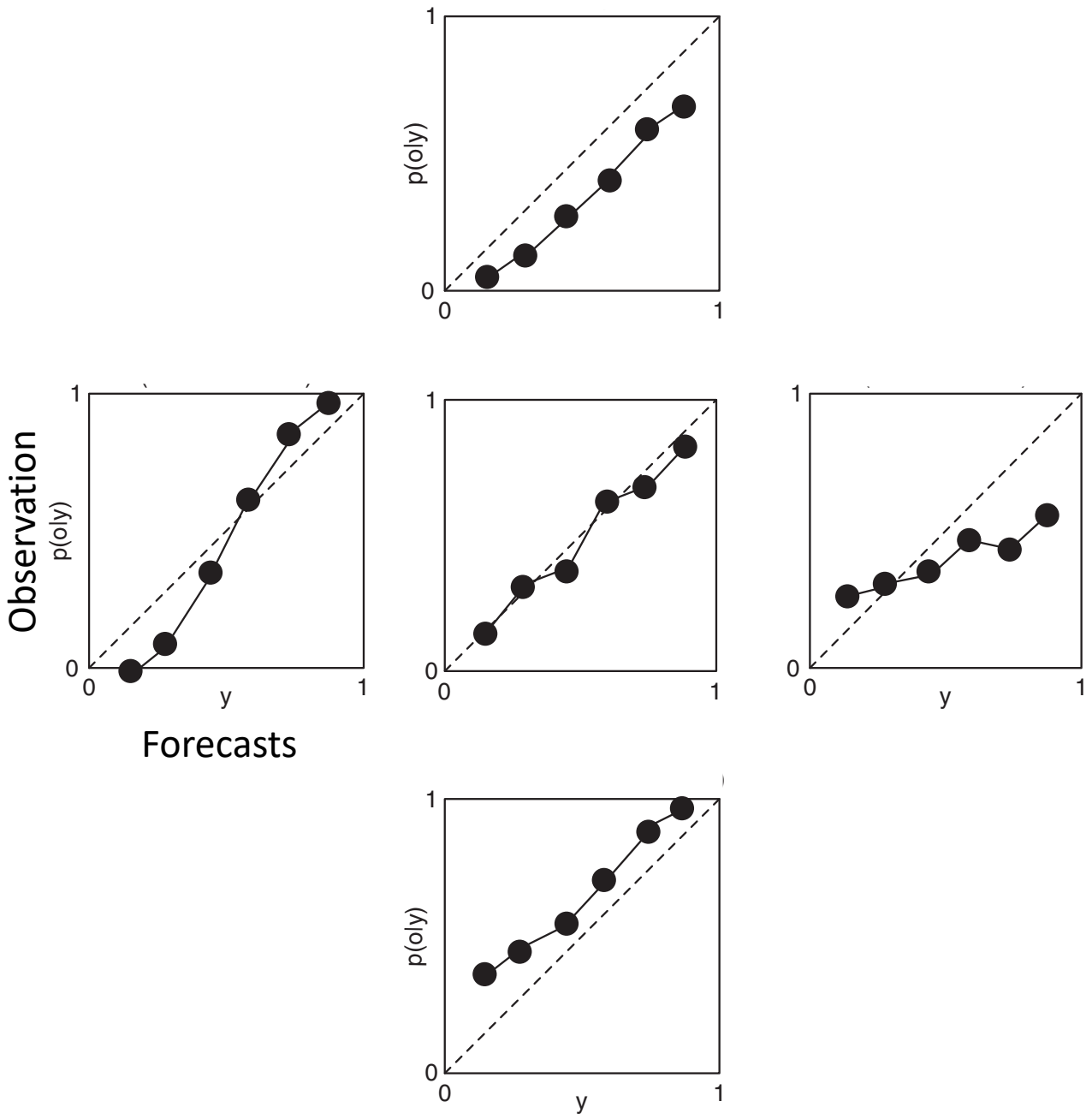
- The reliability diagram is a graph that shows the full joint distribution of forecasts and observations for probability forecasts of a binary predictand (1-occurrence vs. 0-nonoccurrance).
- The reliability diagram has two elements:
  - the calibration distributions: conditional distributions of the observation given the forecast
  - the refinement distribution: the frequency of use of each of the possible forecasts.
- Perfect forecasts: The conditional event relative frequency is essentially equal to the forecast probability, so that the dots fall along the 1:1 line.



Reliability diagrams for week-2 tercile probability forecasts of 850-hPa temperature.
(Whitake and Wei 2006, MWR)

Wilks 2011

(a) Example Calibration Functions



Reliability Diagram

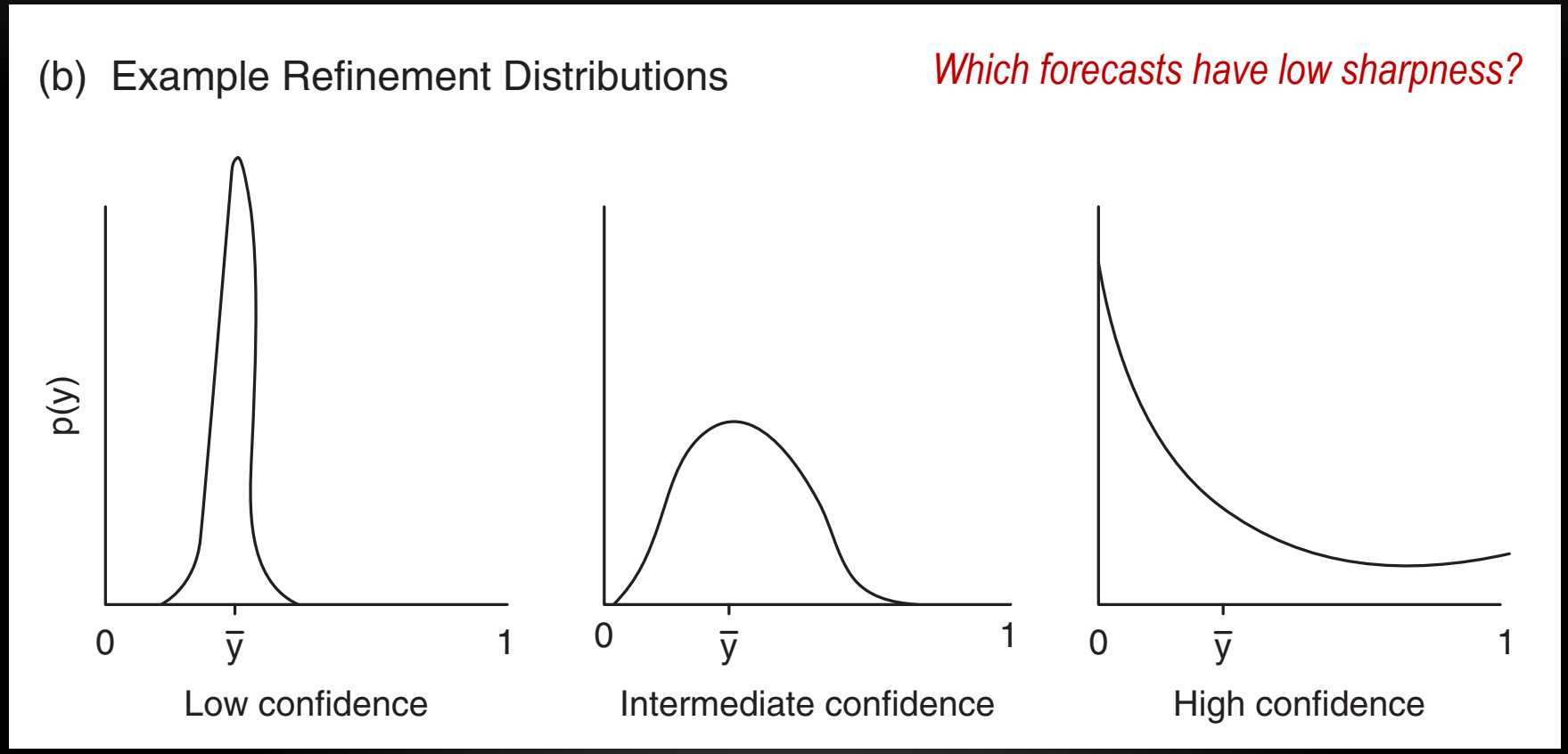
Which one represents a good forecast?
Which one represents over-forecasting?
Which one represents under-forecasting?
Which one represents good resolution?
Which one represents poor resolution?

Pause and Think

- Center: good forecasts, the conditional event relative frequency is essentially equal to the forecast probability.
- Top: the forecasts are consistently too large relative to the conditional event relative frequencies (positive biases)
- Bottom: the forecast probabilities are consistently too small relative to the corresponding conditional event relative frequencies (negative biases)
- Right: poor resolution, the conditional outcome relative frequencies depend only weakly on the forecasts
- Left: good resolution, the forecasts are able to identify subsets of forecast occasions for which the outcomes are quite different from each other.

Wilks 2011

Unconditional Distributions of Forecasts



Left: low confidence, Forecasts that deviate little from their average value exhibit little confidence.
Right: Forecasts that are frequently extreme—that is, specifying probabilities close to the certainty values y = 0 and y = 1 (right panel)—exhibit high confidence.

Wilks 2011

An Example

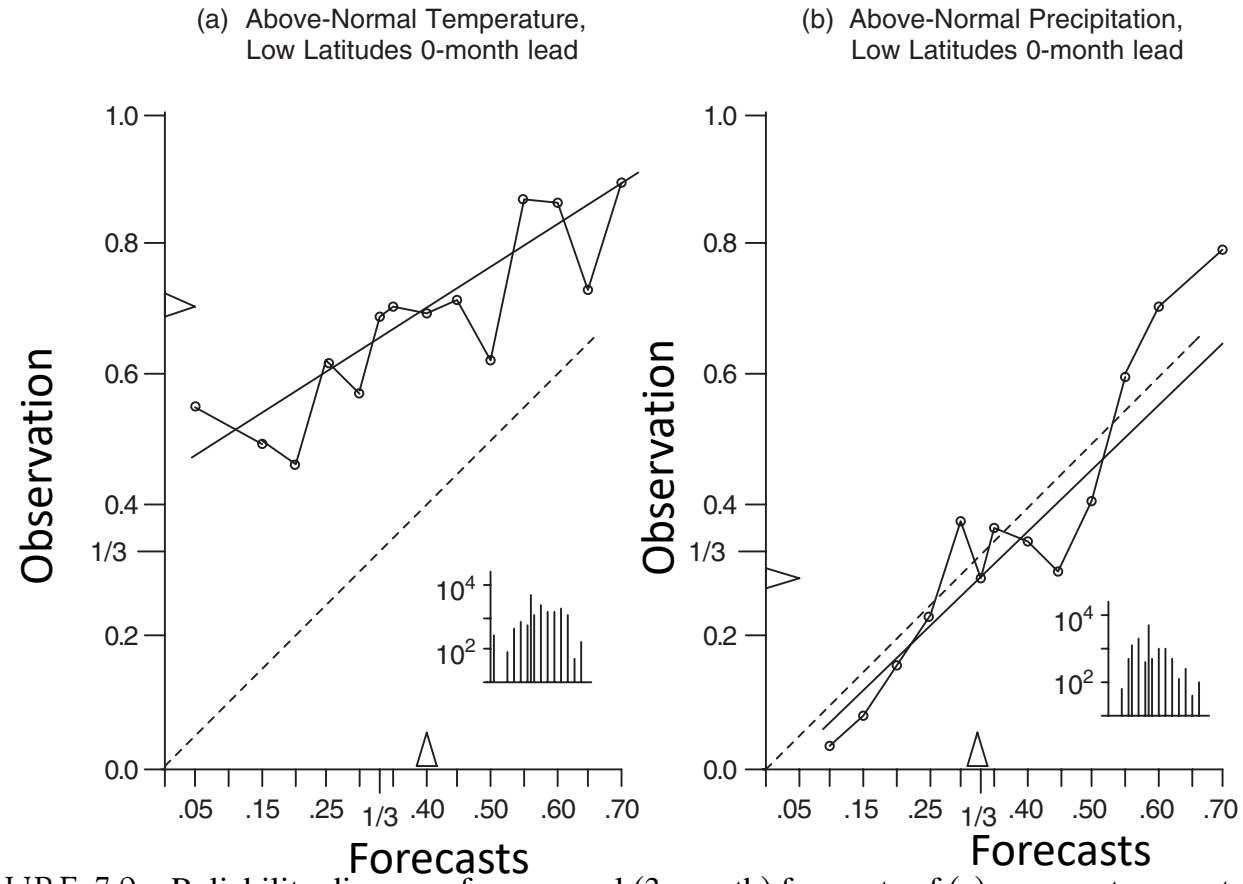


FIGURE 7.9 Reliability diagrams for seasonal (3-month) forecasts of (a) average temperature warmer than the climatological upper tercile, and (b) total precipitation wetter than the climatological upper tercile, for global land areas equatorward of 30°, during the period 1997–2000. From Wilks and Godfrey (2002).

Pause and Think

How would you characterize the two forecasts based on the reliability diagrams?

- Under-forecasting or cold biases for the temperature forecasts.
- The refinement distributions (insets, with logarithmic vertical scales) show much more confidence (more frequent use of more extreme probabilities) for the temperature forecasts.

Wilks 2011

Nonprobabilistic Forecasts of Fields

- Mean Squared Error (MSE): the spatial average of the squared difference between the observation and forecast at each of the M grid points. The root-mean squared error (RMSE) is also widely used.

$$\mathrm{MSE} = \frac{1}{M}\sum_{m=1}^{M}(y_m - o_m)^2.$$

- Anomaly correlation or pattern correlation: measures the similarity between the observed and forecast spatial patterns.

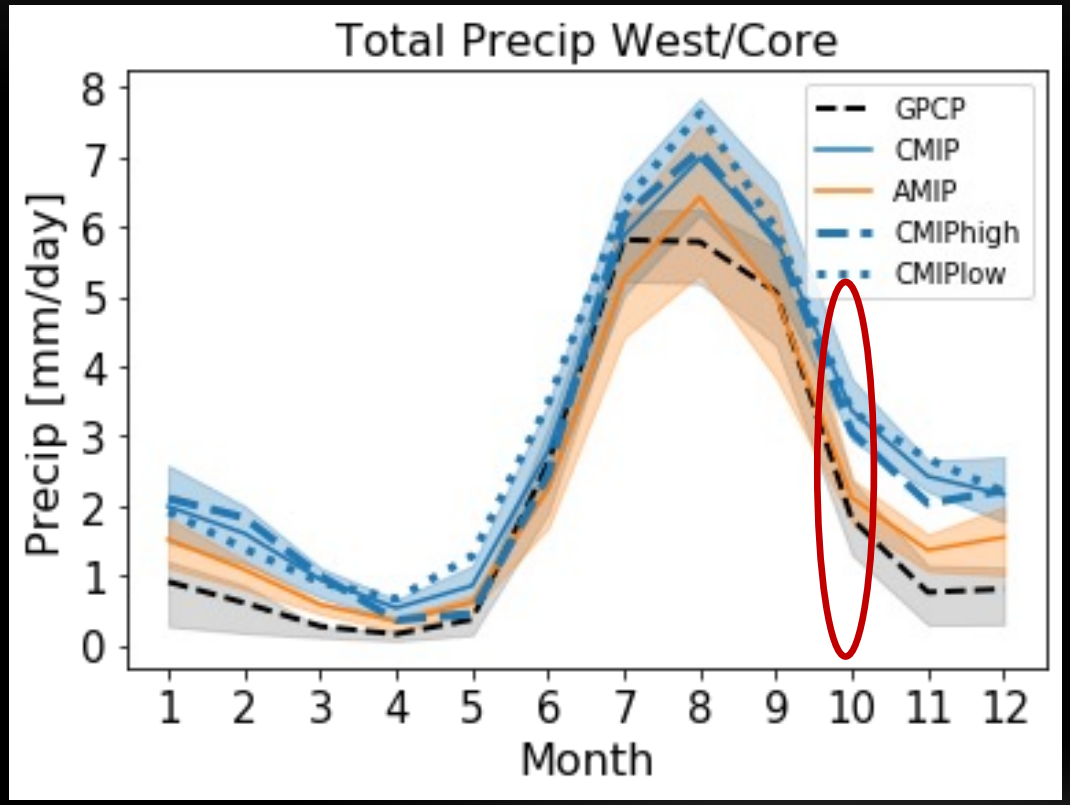
$$\mathrm{AC_C} = \frac{\sum_{m=1}^{M}(y'_m - \bar{y}')(o'_m - \bar{o}')}{\left[\sum_{m=1}^{M}(y'_m - \bar{y}')^2 \sum_{m=1}^{M}(o'_m - \bar{o}')^2\right]^{1/2}}.$$

Wilks 2011

Performance-oriented and Physics-oriented Model Evaluation

- Evaluation of model forecasts is an indispensible component of the model development effort.
- Standard performance-oriented metrics, such as ACC and RMSE, focus on forecast performance and are routinely used in operational centers for forecast verification.
- Physics-oriented diagnostics focus on the critical processes or phenomena and, through evaluation of key processes/variables, shed light on the deficiencies of the physical parameterization or other errors in a model.
- In brevity, physics-oriented evaluation not only provides information on how well a model performs but also on why a model may fail in a certain aspect.

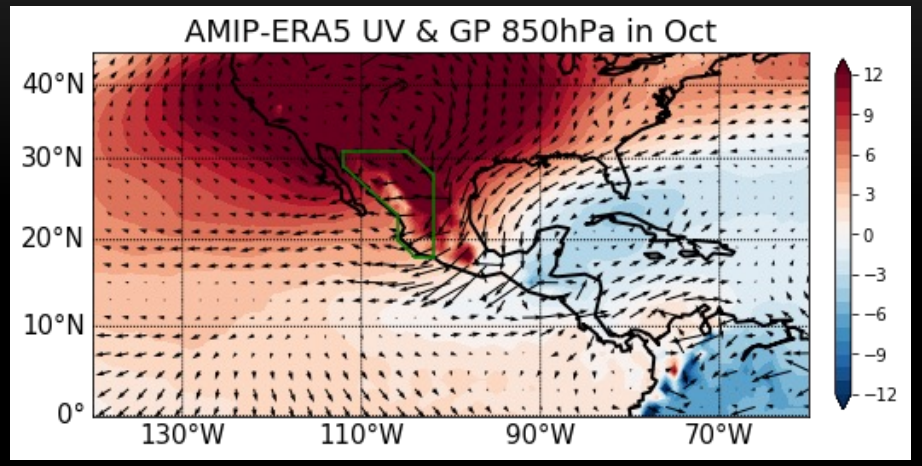
Physics-oriented Model Evaluation: an example

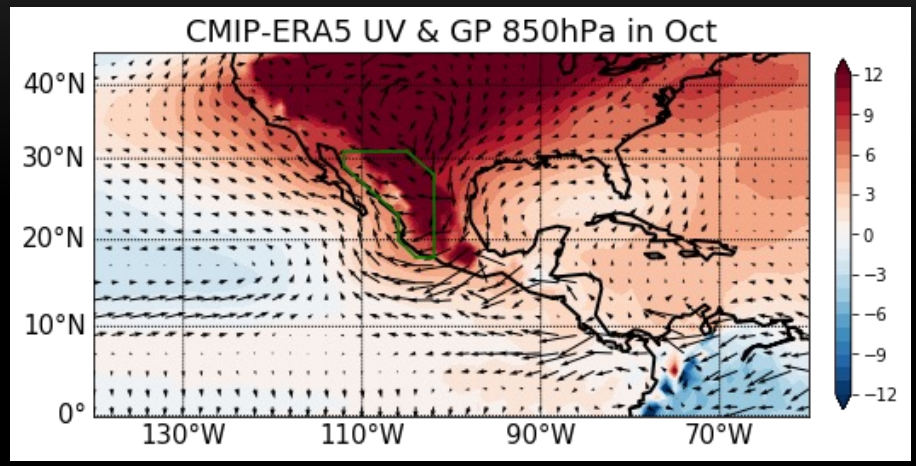


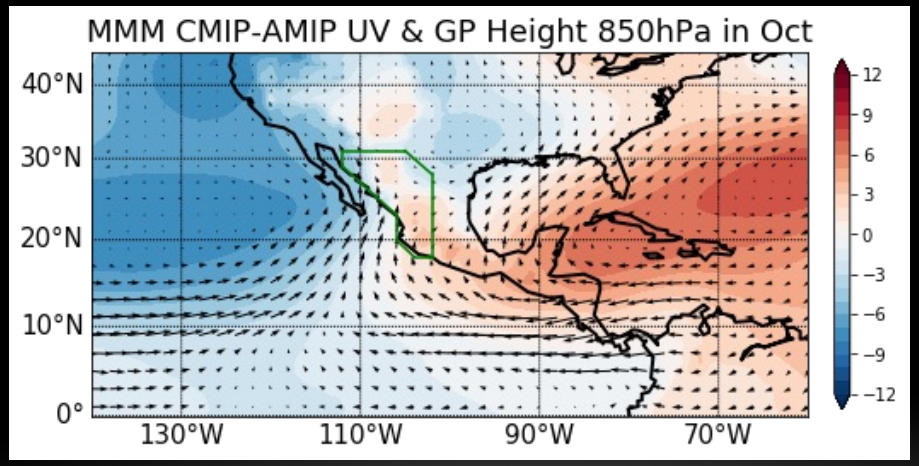
- Seasonal cycle of precipitation in the region of North American Monsoon from AMIP, CMIP multi-model ensemble means (MMM) and GPCP (observation).
- The AMIP MMM shows a better agreement with the observation.
- CMIP overpredicts the late season precipitation.

Ye and Wang, in prep

Physics-oriented Model Evaluation: an example (cont'd)







- Both AMIP and CMIP overestimate H850 over North American Continent.
- CMIP has smaller biases in H850 over the East Pacific. This, however, results in strong pressure gradient force along the coast and thus overpredicted in-land moisture transport, which contributes to the overpredicted precipitation.
- A model may get the right result for the wrong reason!

Ye and Wang, in prep

References

- Wilks, 2011, “Statistical Methods in the Atmospheric Sciences”, Chapter 7
- Wilks, 2011, “Statistical Methods in the Atmospheric Sciences”, Section 3.5
- Whitaker, J. S., Wei, X., & Vitart, F. (2006). Improving Week-2 Forecasts with Multimodel Reforecast Ensembles, *Monthly Weather Review*, *134*(8), 2279-2284.