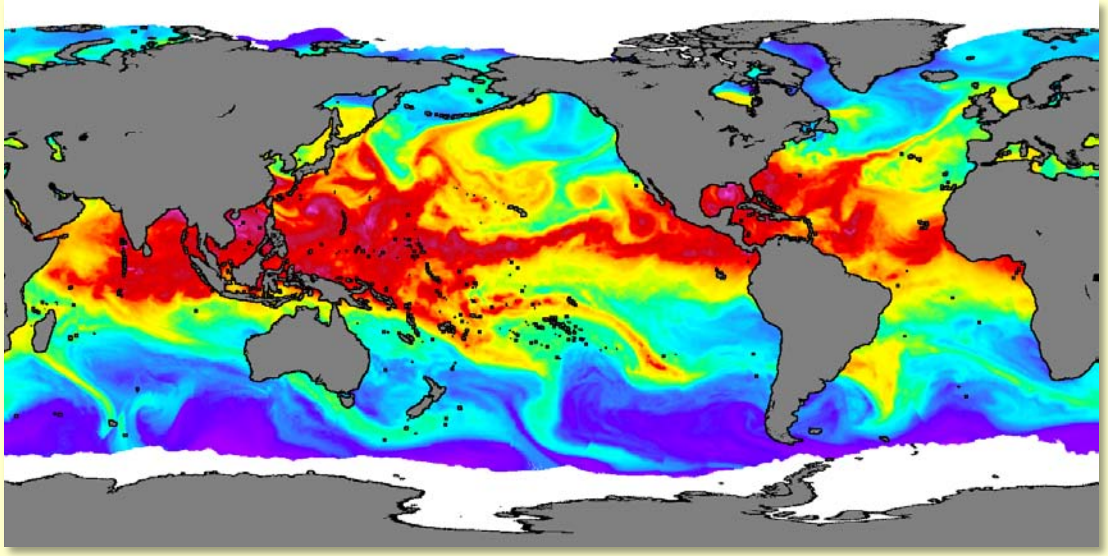
Including Model Quality Information in Detection and Attribution Studies: One Model, One Vote?

B.D. Santer, C. Bonfils, K.E. Taylor, P.J. Gleckler, T.P. Barnett, D.W. Pierce, T.M.L. Wigley, C. Mears, F.J. Wentz, W. Brüggemann, N.P. Gillett, S.A. Klein, S. Solomon, P.A. Stott, and M.F. Wehner

Program for Climate Model Diagnosis and Intercomparison
Lawrence Livermore National Laboratory, Livermore, CA 94550
Email: santer1@llnl.gov

14th Annual CCSM Workshop, Breckenridge, Colorado

June 18th, 2009



Detecting human effects on climate: Is it one model, one vote?

- Fingerprint studies rely on models to estimate both:
  - The pattern of response to human-caused changes in greenhouse gases (and/or other forcings). This pattern is called the "fingerprint"
  - Natural internal climate variability, which constitutes the background "noise" against which the fingerprint must be detected
- In estimating climate fingerprints and noise, should information from different models be given the same weight?
- Or should we exclude models that perform poorly in simulating aspects of observed climate likely to be important for a fingerprint study?

What did we learn in our previous fingerprint study?

PNAS

Identification of human-induced changes in atmospheric moisture content

B. D. Santer[a,b], C. Mears[c], F. J. Wentz[c], K. E. Taylor[a], P. J. Gleckler[a], T. M. L. Wigley[d], T. P. Barnett[e], J. S. Boyle[a], W. Brüggemann[f], N. P. Gillett[g], S. A. Klein[a], G. A. Meehl[d], T. Nozawa[h], D. W. Pierce[e], P. A. Stott[i], W. M. Washington[d], and M. F. Wehner[j]

[a]Program for Climate Model Diagnosis and Intercomparison, Lawrence Livermore National Laboratory, Livermore, CA 94550; [c]Remote Sensing Systems, Santa Rosa, CA 95401; [d]National Center for Atmospheric Research, Boulder, CO 80307; [e]Scripps Institution of Oceanography, La Jolla, CA 92037; [f]Institut für Unternehmensforschung, Universität Hamburg, 20146 Hamburg, Germany; [g]Climatic Research Unit, School of Environmental Sciences, University of East Anglia, Norwich NR4 7TJ, United Kingdom; [h]National Institute for Environmental Studies, Tsukuba 305-8506, Japan; [i]Hadley Centre for Climate Prediction and Research, United Kingdom Meteorological Office, Exeter EX1 3PB, United Kingdom; and [j]Lawrence Berkeley National Laboratory, Berkeley, CA 94720

We found that:

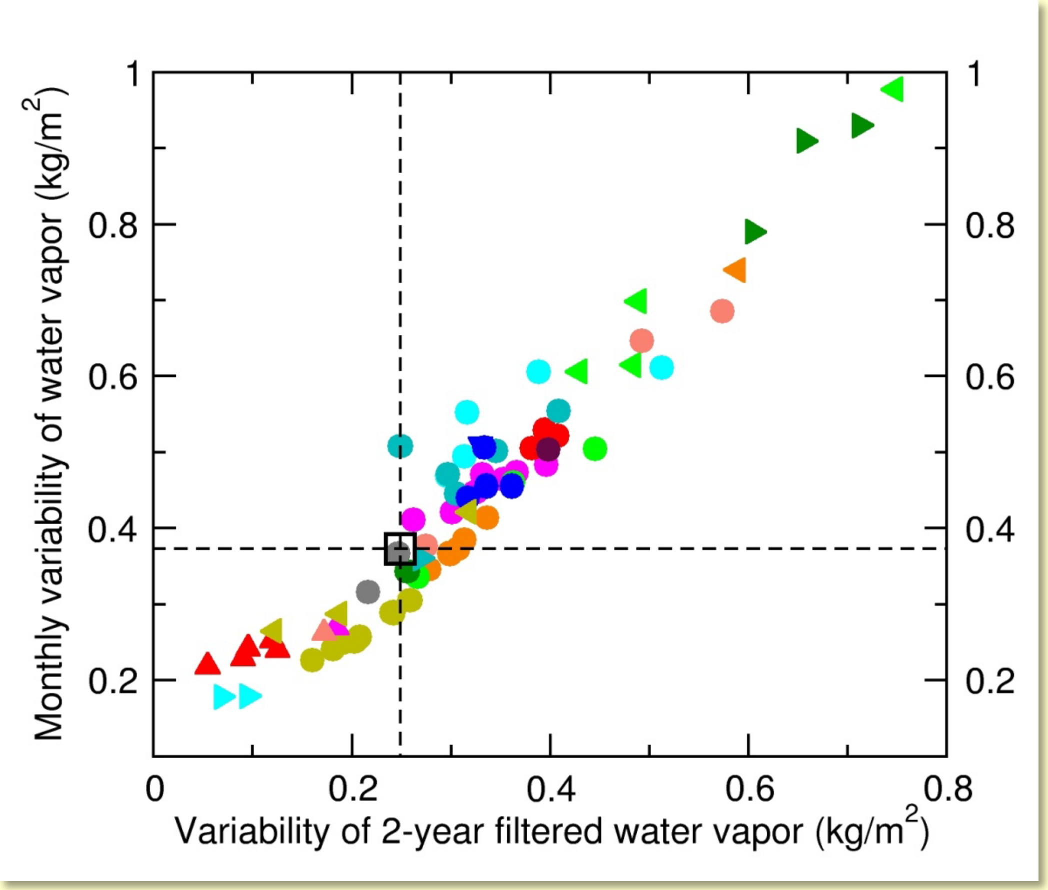
- There is an emerging human-caused signal in the increasing moisture content of Earth's atmosphere
- This signal is primarily due to human-caused increases in well-mixed greenhouse gases

What model data did we use in our PNAS paper?

- We used water vapor data from 22 different climate models (CMIP-3 archive)
- We used model 20th century ("20CEN") simulations to define the fingerprint that we searched for in observations
- We used water vapor data from model control runs (with no forcing changes) to estimate the noise of natural climate variability
- Water vapor observations from satellite-borne Special Sensor Microwave Imager (SSM/I).

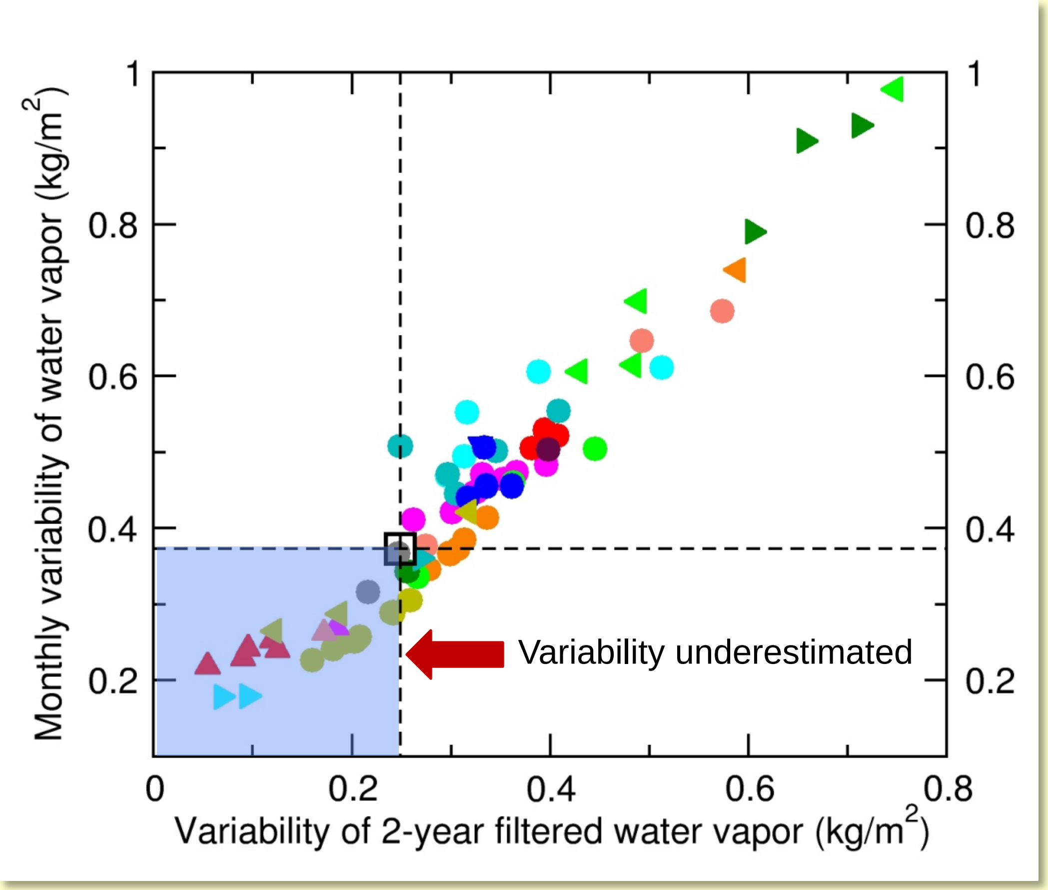
Although the models showed important differences in their performance, they had equal weight in the D&A study

The simulated variability ranges from 1/3 to 2.5 times the amplitude of observed variability.



Santer *et al.*, *Proceedings of U.S. National Academy of Sciences* (2007)

Although the models showed important differences in their performance, they had equal weight in the D&A study



Santer *et al*., *Proceedings of U.S. National Academy of Sciences* (2007)

If we use only the “top ten” models, can we still identify a human fingerprint in observed water vapor changes?

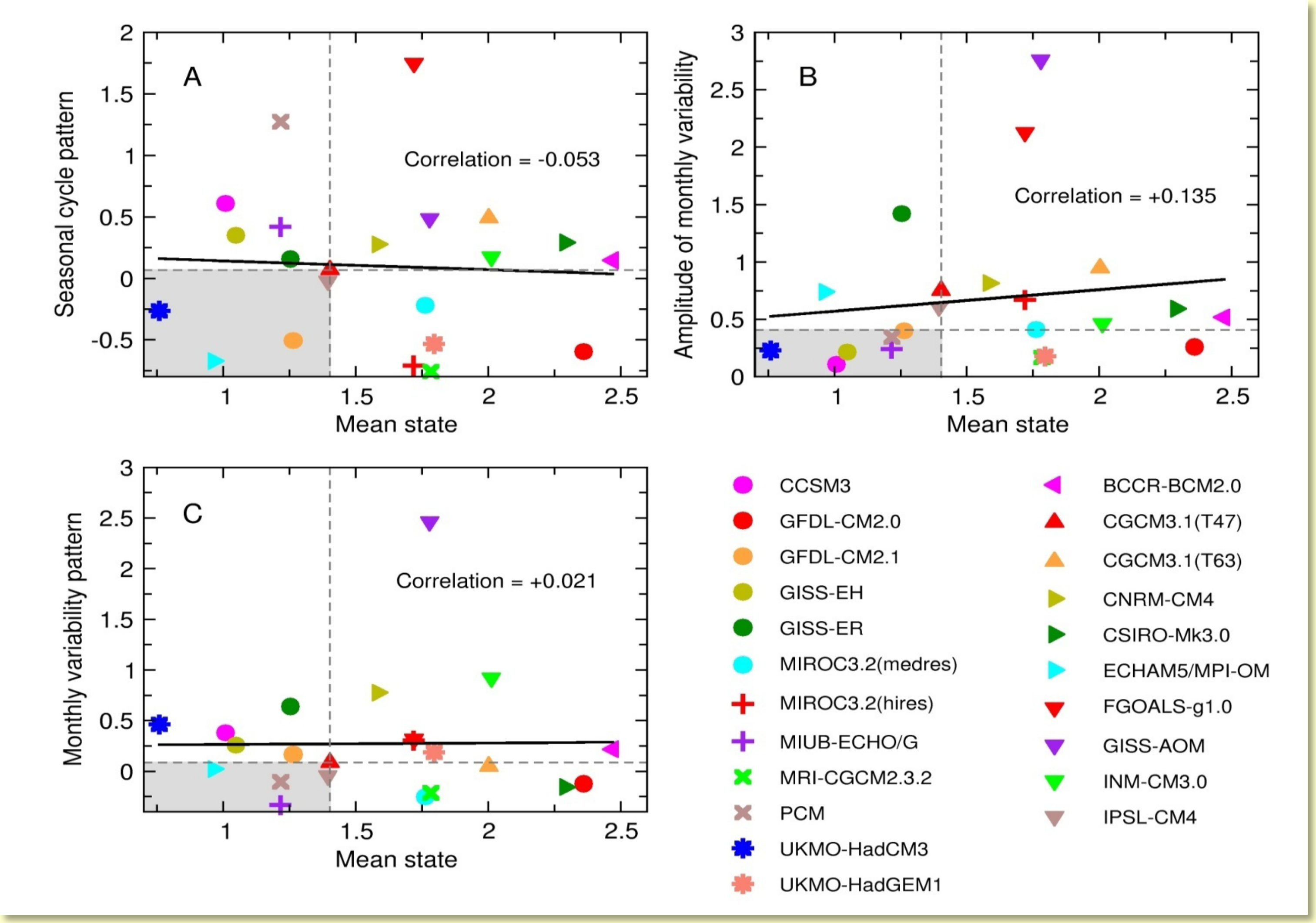
- We identified the “top ten” models (out of 22 in the CMIP-3 archive) in three sets of model performance metrics:
  - The climatological mean state and seasonal cycle pattern (“M+SC”)
  - The amplitude and pattern of variability on different timescales (monthly, 2-year, 10-year; “VA+VP”)
  - Mean state, seasonal cycle, and variability (“ALL”)
- This was done for:
  - Two different variables: Water vapor and sea-surface temperature (SST)
  - Five different geographical regions: AMO, PDO, Niño 3.4, tropical oceans (30°N-30°S), and near-global oceans (50°N-50°S)

How did we do the model ranking?

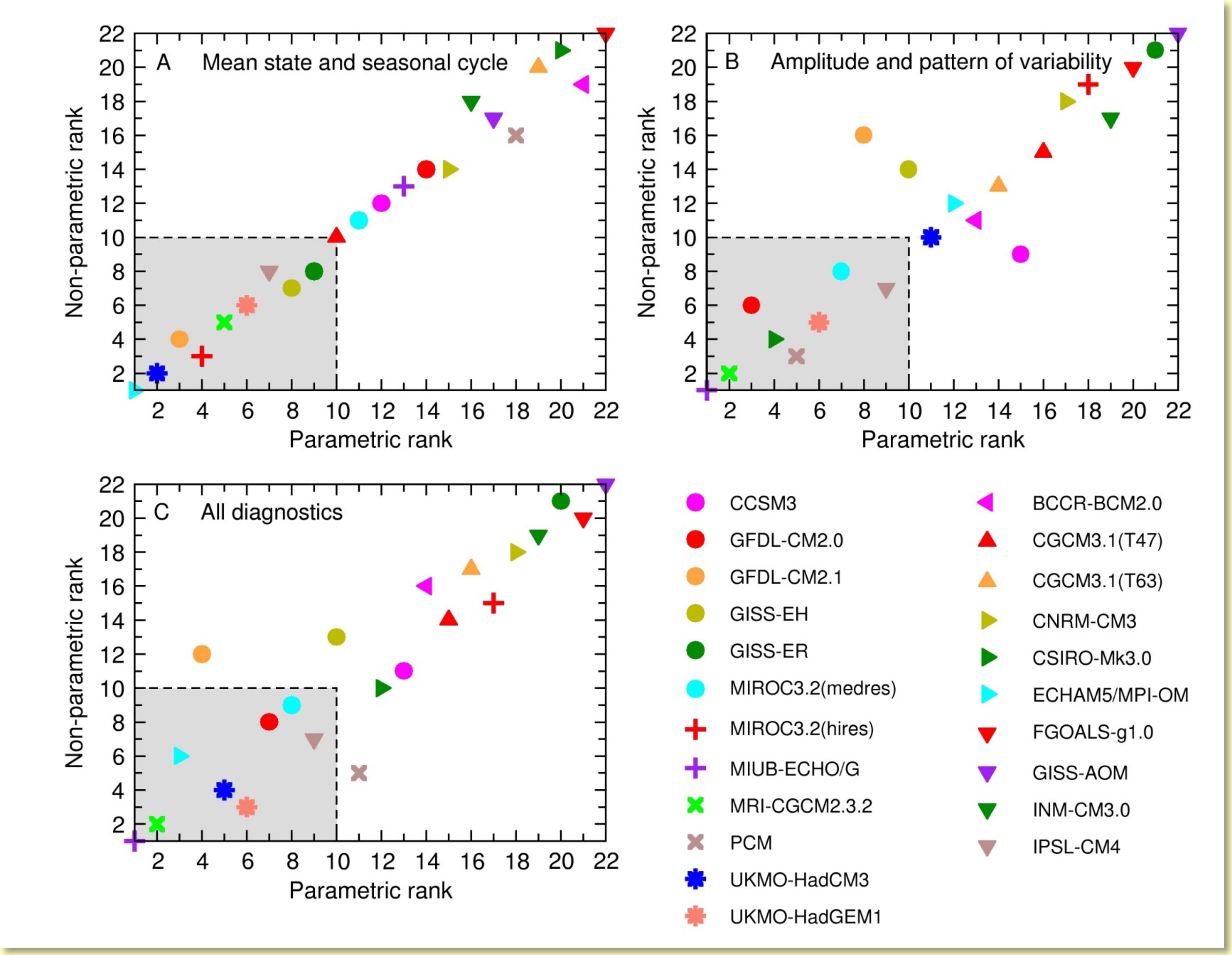
- M+SC: 20 model performance metrics
- VA+VP: 50 model performance metrics
- ALL: 70 model performance metrics
- For each set of metrics, model ranking was done in two different ways:
  - Parametrically: Rank is average of normalized values of individual metrics (“P”)
  - Non-parametrically: Average of the ranks for each individual metric (“NP”)
- In each case, identified “top ten” and “bottom ten” models
  - 12 cases: 3 groups of metrics (M+SC, VA+VP, ALL) × 2 ranking schemes (P, NP) × 2 groups of models (Top ten, Bottom ten)

Relationship between different measures of model skill

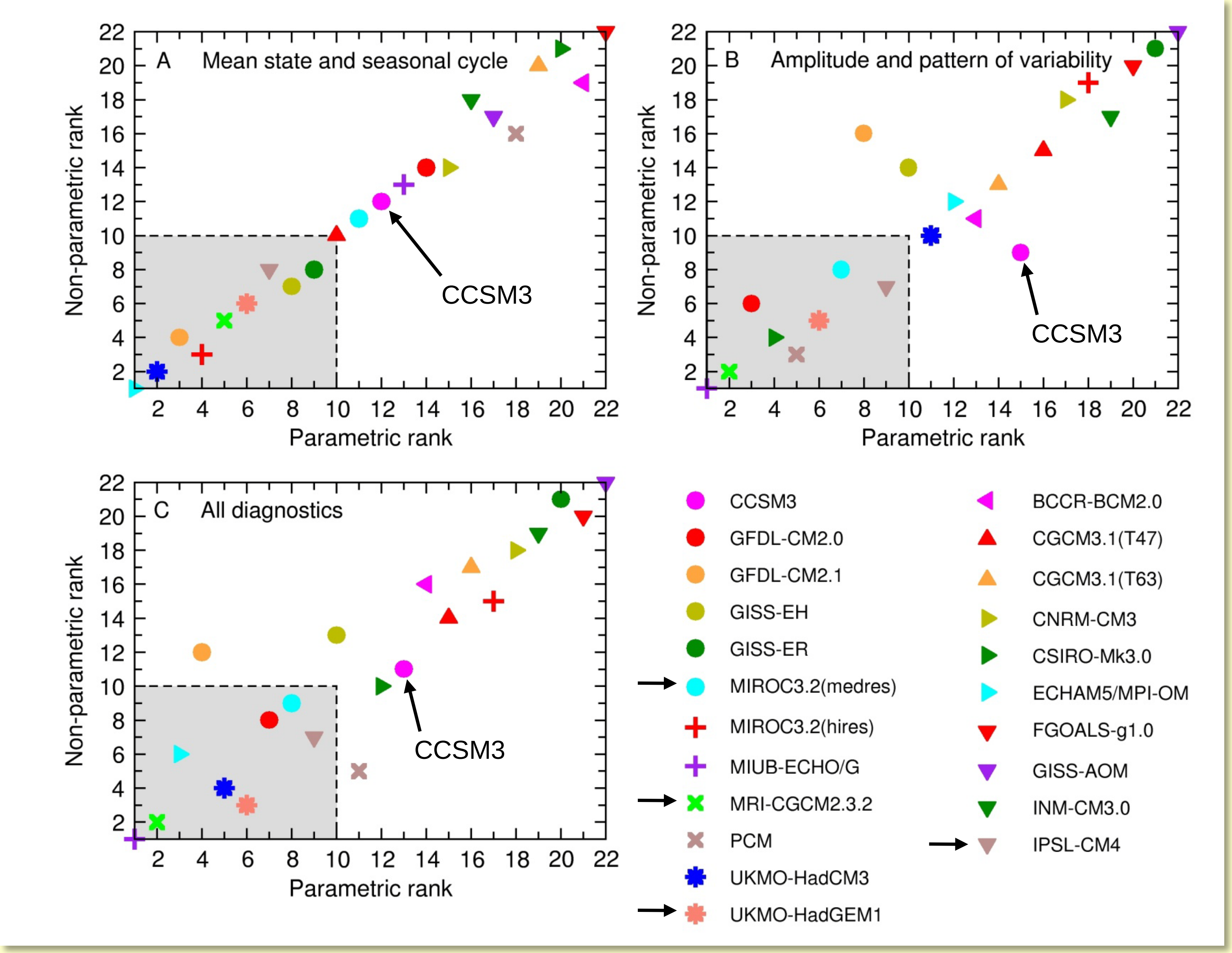
There is no relationship between skills to simulate the mean state and the variability.



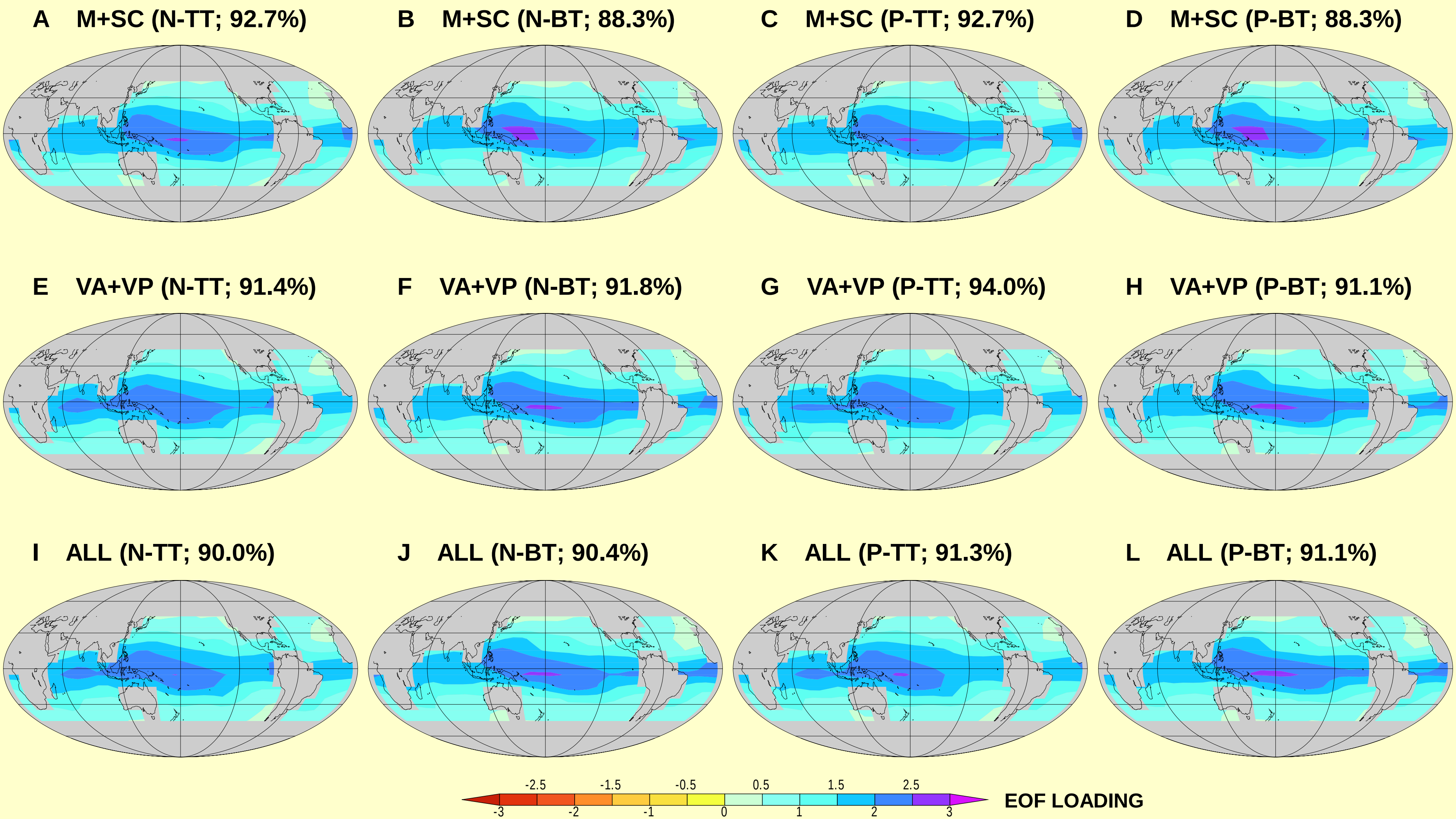
Overall ranking of model performance



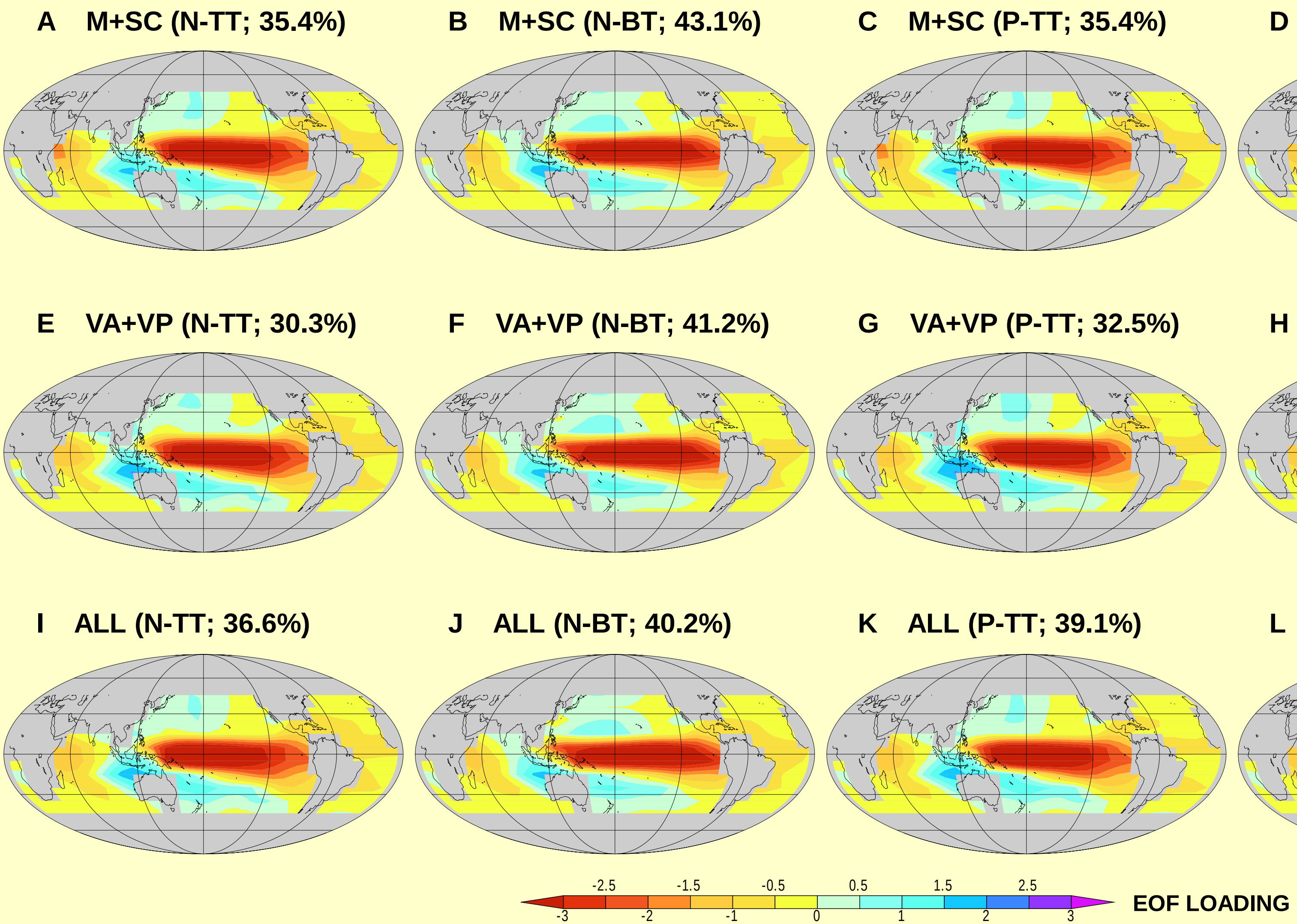
Overall ranking of model performance



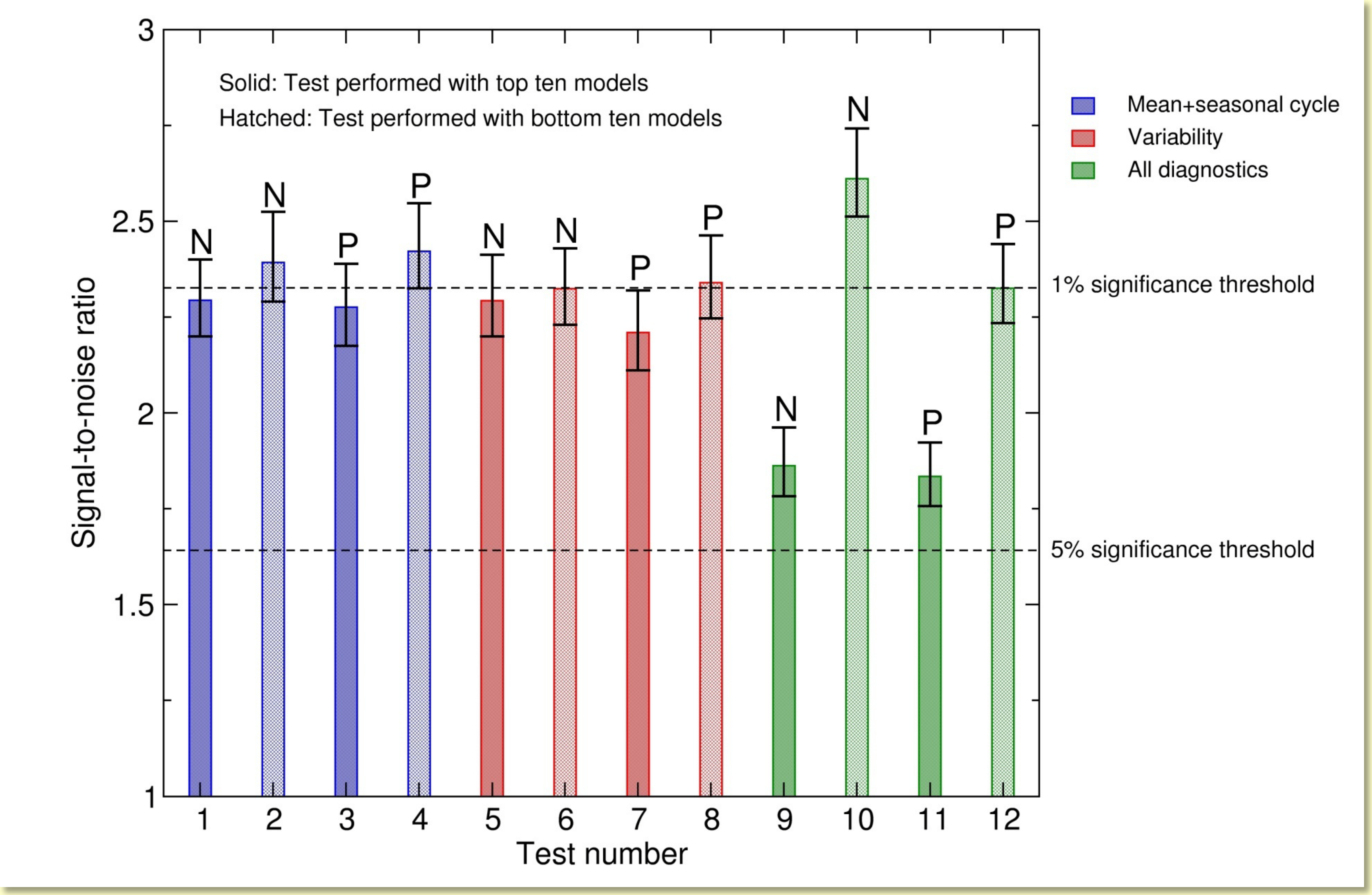
Is the “fingerprint” pattern of water vapor changes in response to external forcing sensitive to model quality information?



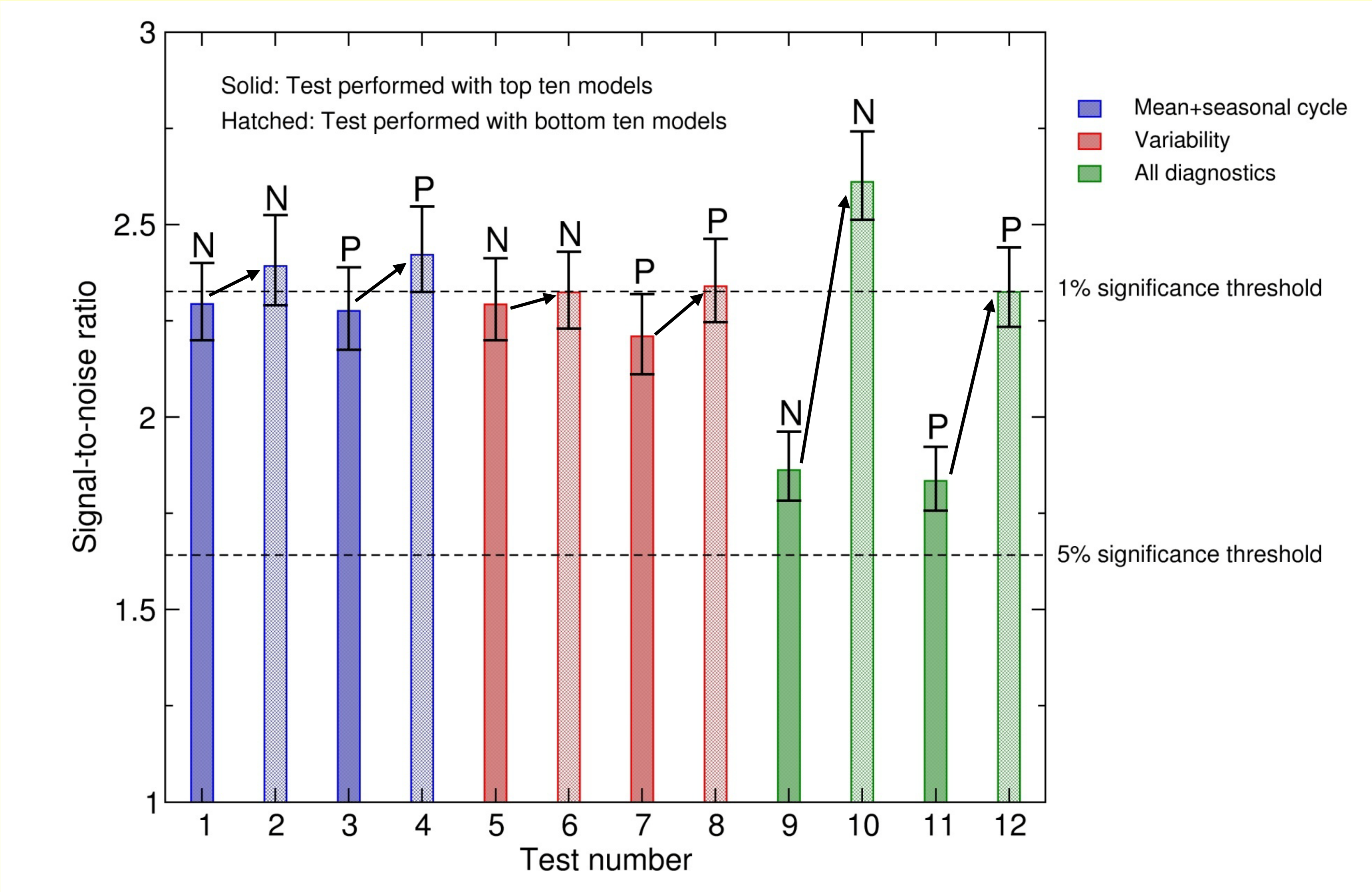
Is the pattern of internally-generated variability sensitive to model quality information?



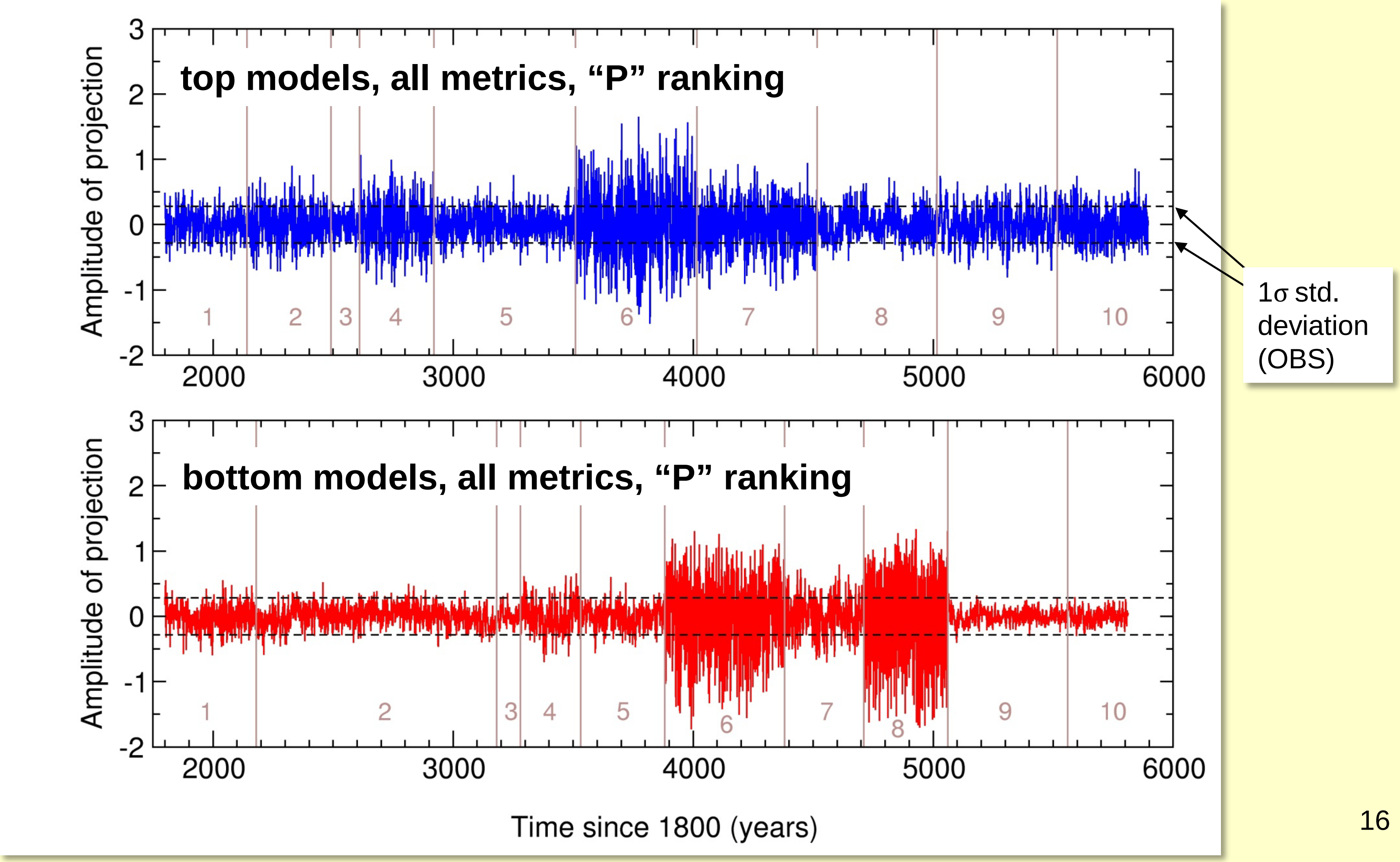
We can identified a human “fingerprint” in the observed water vapor changes in each of the 12 cases



We can identified a human "fingerprint" in the observed water vapor changes in each of the 12 cases



Why do D&A results based on the “top ten” and “bottom ten” models have different S/N ratios?



Conclusions (I)

- Model errors are complex in space and time

- Even for a straightforward application (identifying a human fingerprint in observed water vapor changes), it is not easy to make an unambiguous identification of the "top ten" models

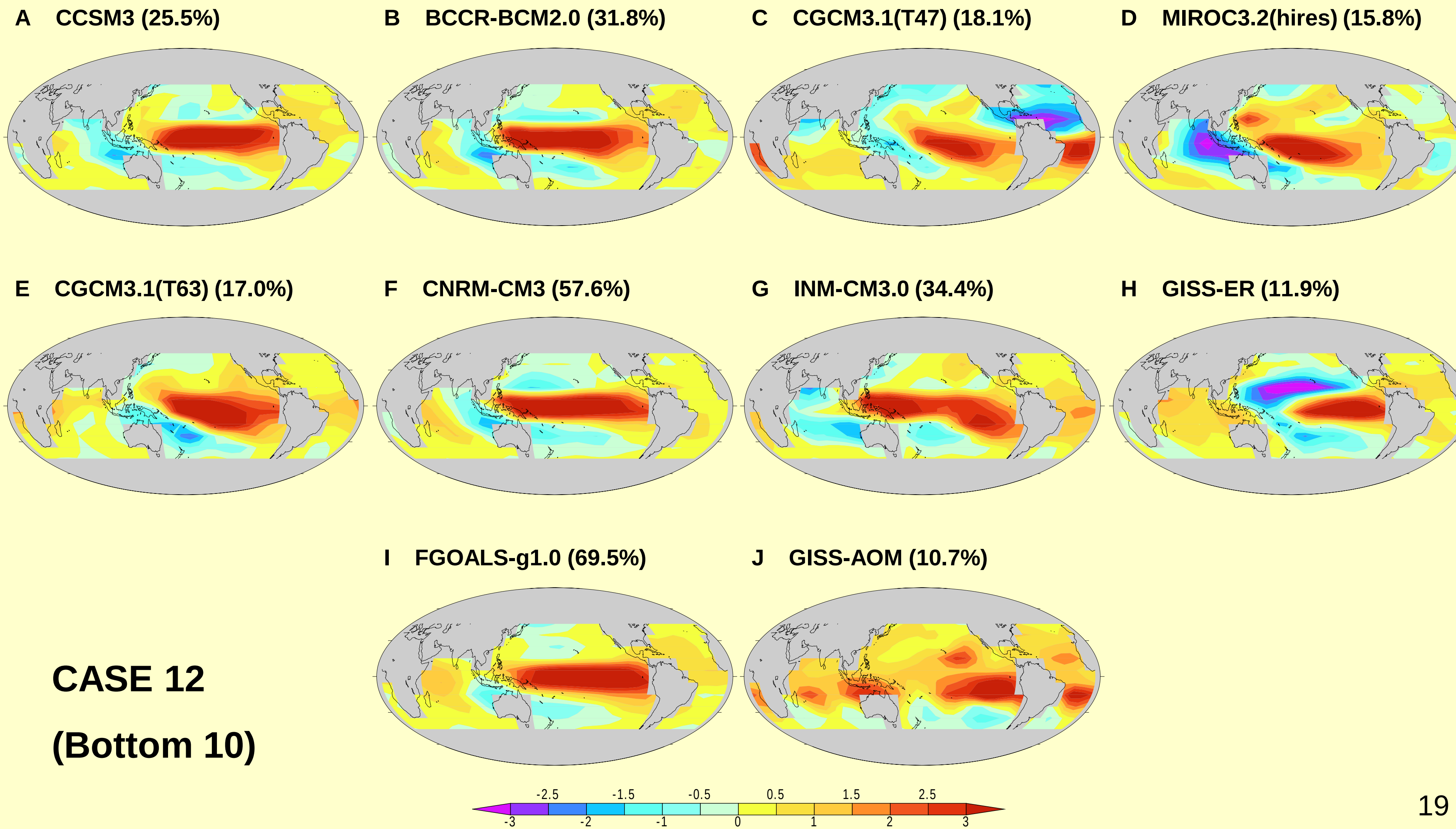
- In the water vapor example, there is not a clear relationship between model errors in simulating the mean state and the temporal variability
  - ➡ These results imply that it may be difficult to come up with objective, scientifically-defensible schemes for weighting projections of future climate change

- Our positive detection of a human fingerprint in satellite-based estimates of water vapor change is a robust result
  - ➡ It is relatively unaffected by incorporating "model quality" information in the detection study

Conclusions (II)

- Use of the “bottom ten” models for detecting anthropogenic effects on water vapor leads to an overestimate of S/N ratios
  - Introduces biases in D&A results
- The “fingerprint” of water vapor changes in response to external forcing is relatively insensitive to model quality information
  - Fingerprint structure is dictated by zero-order physics
- The structure of the dominant mode of water vapor variability is also remarkably insensitive to model quality information
  - Very similar noise modes are estimated from “top ten” and “bottom ten” models (despite large differences in noise structure in individual models)

Patterns of internally-generated variability in individual model control runs



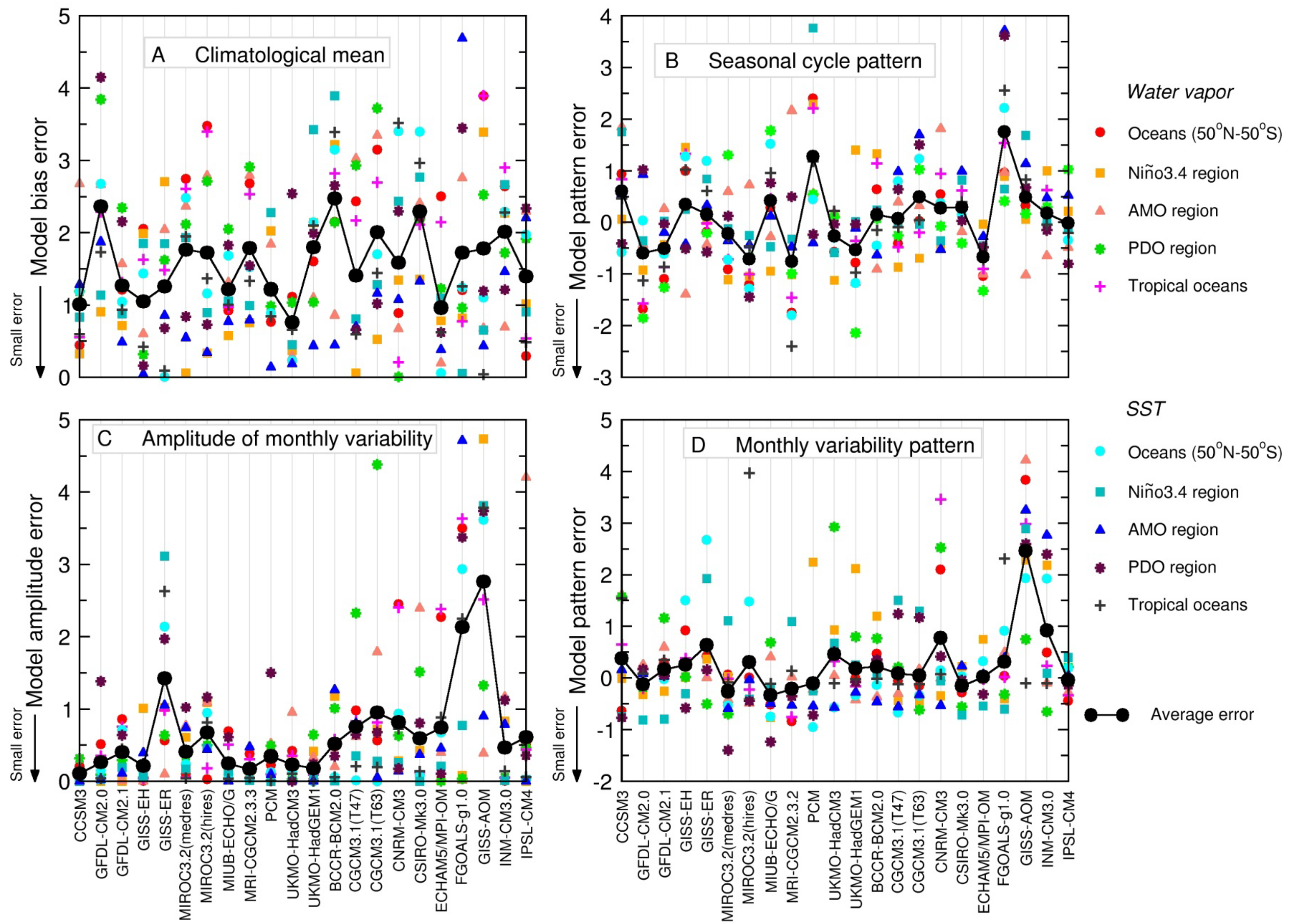
CASE 12

(Bottom 10)

What observational data did we use in our PNAS paper?

- Water vapor retrievals were available since Sept. 1987 from the satellite-borne Special Sensor Microwave Imager (SSM/I)
- Based on measurements of microwave emissions from 22 GHz water vapor absorption line
- SSM/I retrievals unavailable over highly emissive land surface and ice
- We used data for 19-year period Jan. 1988 to Dec. 2006

Results for individual regions, variables, and metrics



Results for individual regions, variables, and metrics