Assessing the simulation of clouds, radiation, and precipitation in climate simulations

Robert Pincus, Patrick Hofmann, and Crispian Batstone
Climate Diagnostics Center, Physical Sciences Division, NOAA/ESRL

Karl Taylor and Peter Gleckler
Program for Climate Model Diagnosis and Intercomparison, LLNL

Definitions

Global numerical weather prediction (NWP):

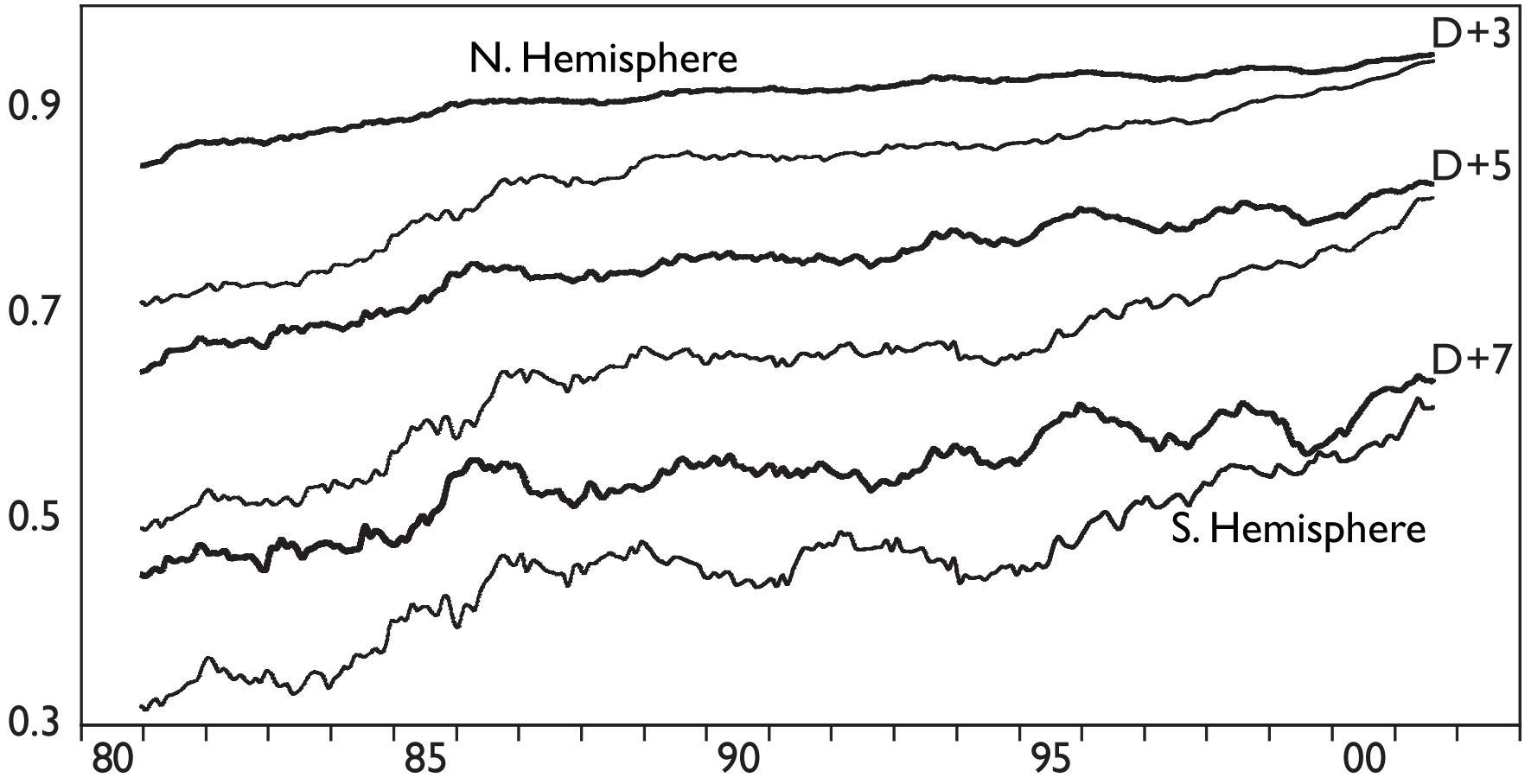
- days-to-months forecast duration
- assimilation system
- initial value problem: forecasts for specific times

Climate modeling (GCMs):

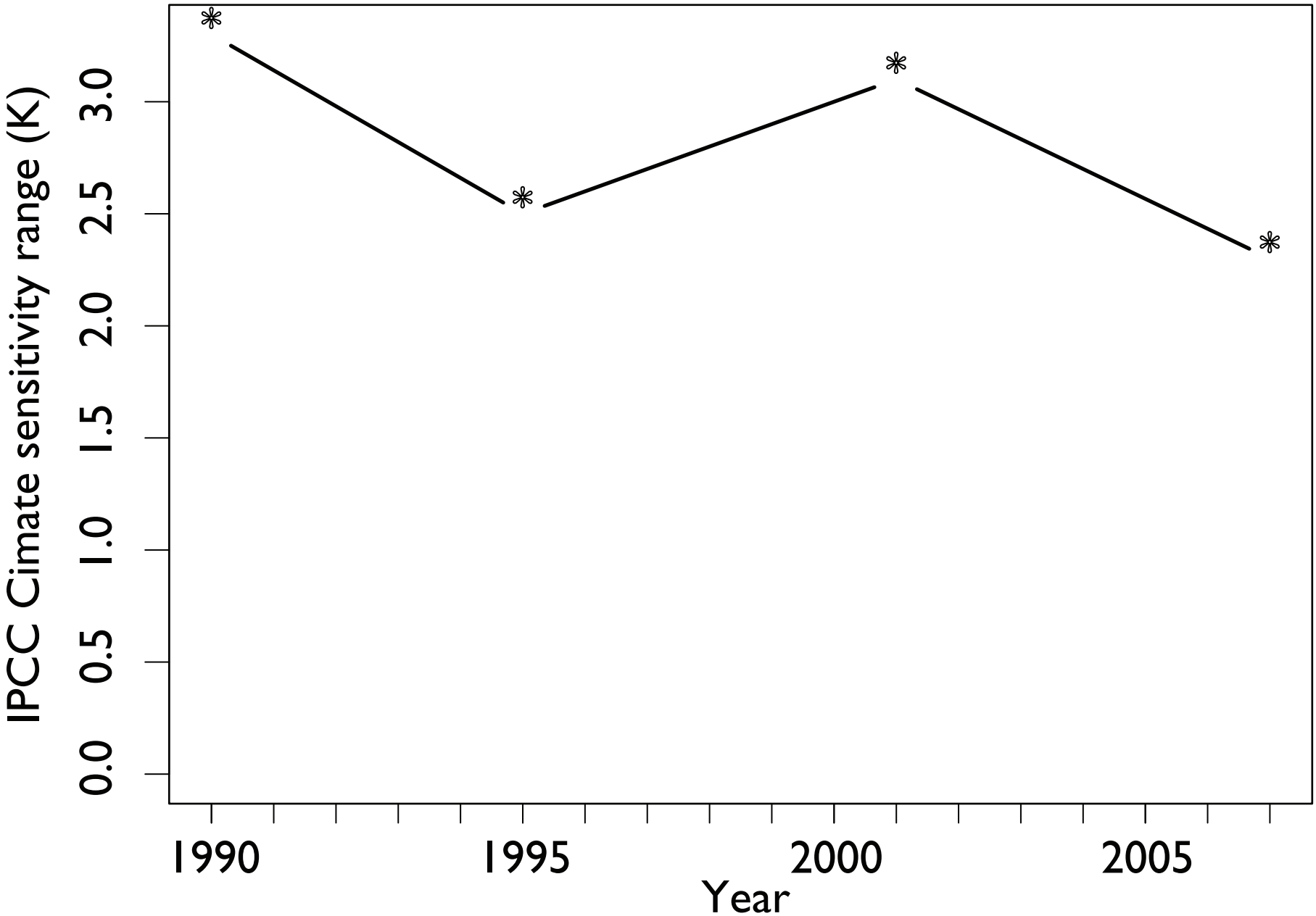
- decade-to-century projections
- no assimilation system
- boundary value problem: statistical evaluation (mean, modes of variability, relationships)

How have these models improved over time?

ECMWF 500hPa Z forecast anomaly correlation (Simmons and Hollingsworth, 2001)



Range of climate sensitivity in IPCC reports



Measuring model skill

Forecast skill scores are computed by all modeling centers

- Explicitly defined by WMO in 1980s
- Computed monthly and shared among centers

There's no agreement on what constitutes climate model skill

This is both a social and a conceptual problem:

- We desire projections on long time scales
- We can evaluate against present-day observation
- We don't know what relationships (if any) link the two

We'll assume that present-day skill is necessary and seek

- climate model metrics analogous to NWP skill scores

What is a metric?

Metrics evaluate accuracy (how good), not why

- Low-order measure of skill
- Applied to forced modes, not internal variability

It's not a diagnostic measure

- Those are aimed at finding out what's wrong so the model can be improved; usually at the process level

NWP metrics (skill scores) are very simple

- Bias, RMS error, anomaly correlation
- Winds, temperature, pressure/geopotential height
- Large domains (e.g. globe, tropics, extra-tropics)

Metrics for clouds, radiation, and precipitation

NWP models don't evaluate these quantities

- Believed not to affect forecast skill strongly
- Point observations aren't representative
- Variables are not assimilated, so can't use analysis

Climate models need to evaluate these quantities

- "Clouds play a central role in climate" (Mom, apple pie)
- Observations *are* available on relevant time/space scales

We have computed a set of metrics for clouds, etc.

- across a range of climate models

Guinea pigs

We scored the GCM runs used for IPCC AR4

- "Climatology" is typically 1979-1999
- Ensemble average where relevant

Two classes of models, four ringers:

- Atmosphere-only, specified sea surface temperature (12)
- Coupled ocean/atmosphere, specified GHGs (21)
- "Super-CAM" (MMF/SP-CAM)
- "IPCC mean model" separately for AMIP, coupled runs
- Clouds from 40 year ECMWF reanalysis (ERA-40)
- Current-ish version of ECMWF forecast model

Decisions, decisions

What defines a metric?

- Physical parameter
- Verifying observations
- Time/space domain
- Statistical quantity

What to evaluate

Evaluation quantities have to be *observable* and *relevant*

- Global evaluations lead to using satellite data
- (but not CERES SRBAVG-nonGEO)

A suite of parameters and verifying observations:

- Total cloud fraction: ISCCP (and MODIS)
- Surface precipitation rate: GPCP (and Xie-Arkin)
- Cloud radiative effect (CRE): CERES-ES4 (and ERBE)
- For comparison: TOA radiation fluxes

When and where to evaluate

We score the global composite seasonal cycle

- (Global domain, averaged month-by-month)
- “Contains” errors in mean annual cycle, mean climate
- Practical (i.e. people will give you data at this resolution)

How to evaluate (1)

Mean differences (bias) and pattern errors are both relevant

How about this:

Mean bias $\bar{E} = \bar{m} - \bar{o}$

Root mean square error $E = \sqrt{\overline{(m - o)^2}}$

Centered RMS error $E' = \sqrt{\overline{[(m - \bar{m}) - (o - \bar{o})]^2}}$

Correlation (anomaly or pattern correlation)

$$R = \frac{\overline{(m - \bar{m}) - (o - \bar{o})}}{\sigma_m \sigma_o}$$

Ratio of standard deviations σ_m / σ_o

How to evaluate (2)
Taylor diagrams

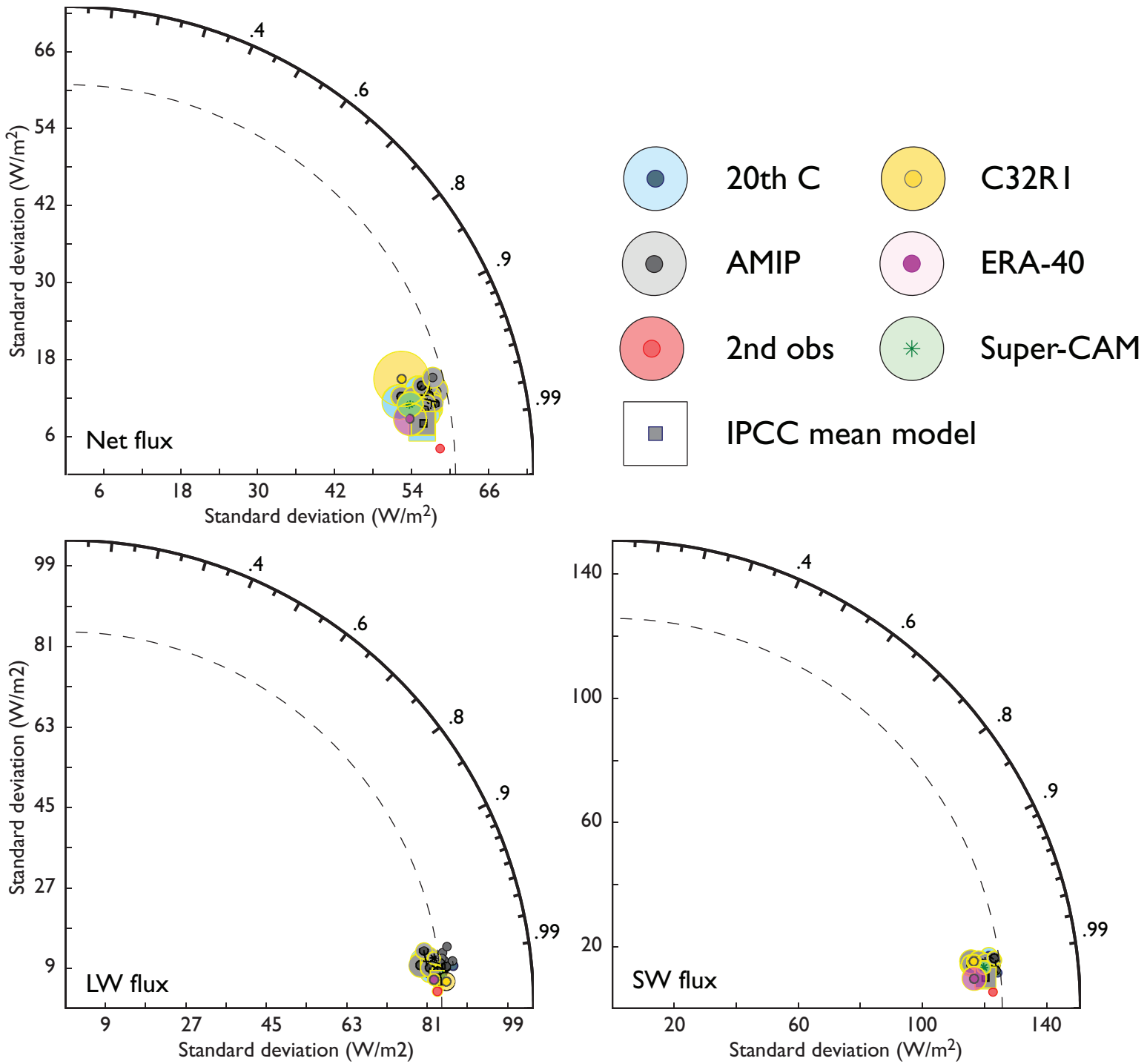
Note 1: RMS error is the quadratic sum of bias and centered RMS error: $E^2 = \bar{E}^2 + E'^2$

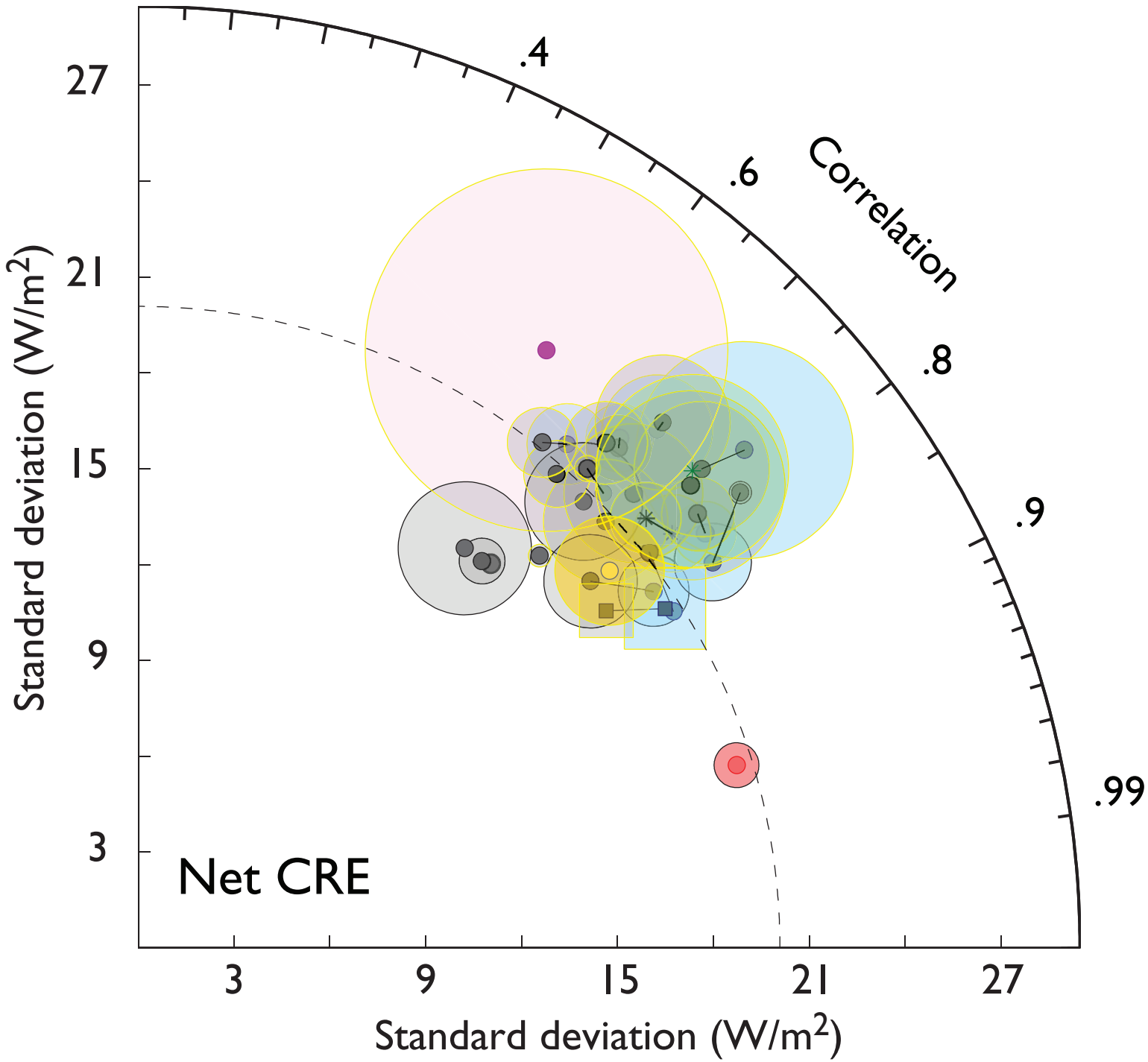
Note 2: The ratio of standard deviations, the correlation coefficient, and the centered RMS error are geometrically related

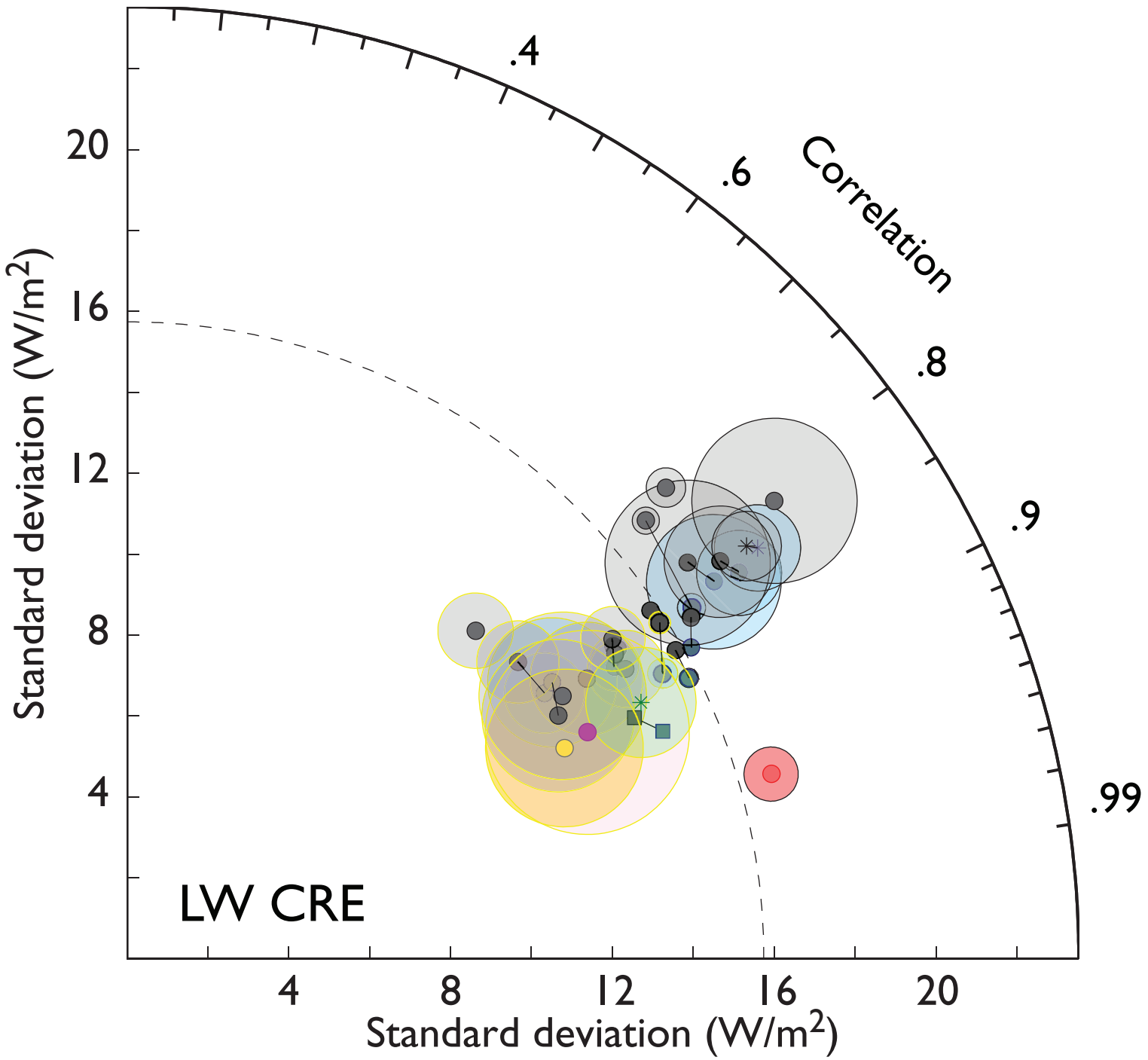
$E' = \sigma_m + \sigma_o + 2\sigma_m\sigma_o \cos R$

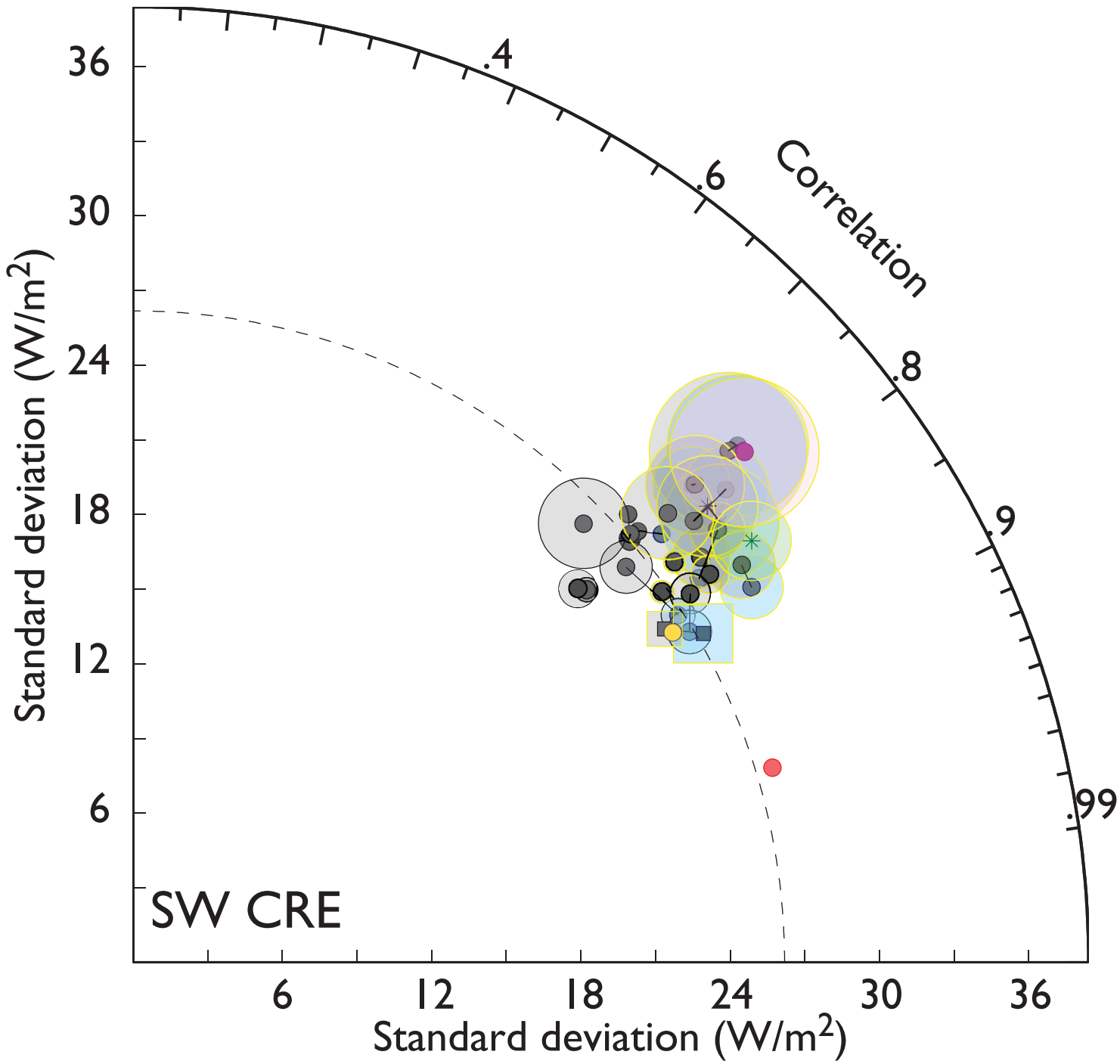
The "Taylor diagram" uses these relationships

We've modified the Taylor diagram to show the bias


Net flux
Standard deviation (W/m2)
Standard deviation (W/m2)
6
18
30
42
54
66
.4
.6
.8
.9
.99
LW flux
Standard deviation (W/m2)
Standard deviation (W/m2)
9
27
45
63
81
99
SW flux
Standard deviation (W/m2)
Standard deviation (W/m2)
20
60
100
140
20th C
C32R1
AMIP
ERA-40
2nd obs
Super-CAM
IPCC mean model


Net CRE
Correlation
.4
.6
.8
.9
.99
Standard deviation (W/m^2)
3
9
15
21
27
Standard deviation (W/m^2)
3
9
15
21
27


LW CRE
Correlation
.4
.6
.8
.9
.99
Standard deviation (W/m2)
4
8
12
16
20
Standard deviation (W/m2)
4
8
12
16
20


SW CRE
Correlation
.4
.6
.8
.9
.99
36
30
24
18
12
6
Standard deviation (W/m2)
6
12
18
24
30
36
Standard deviation (W/m2)

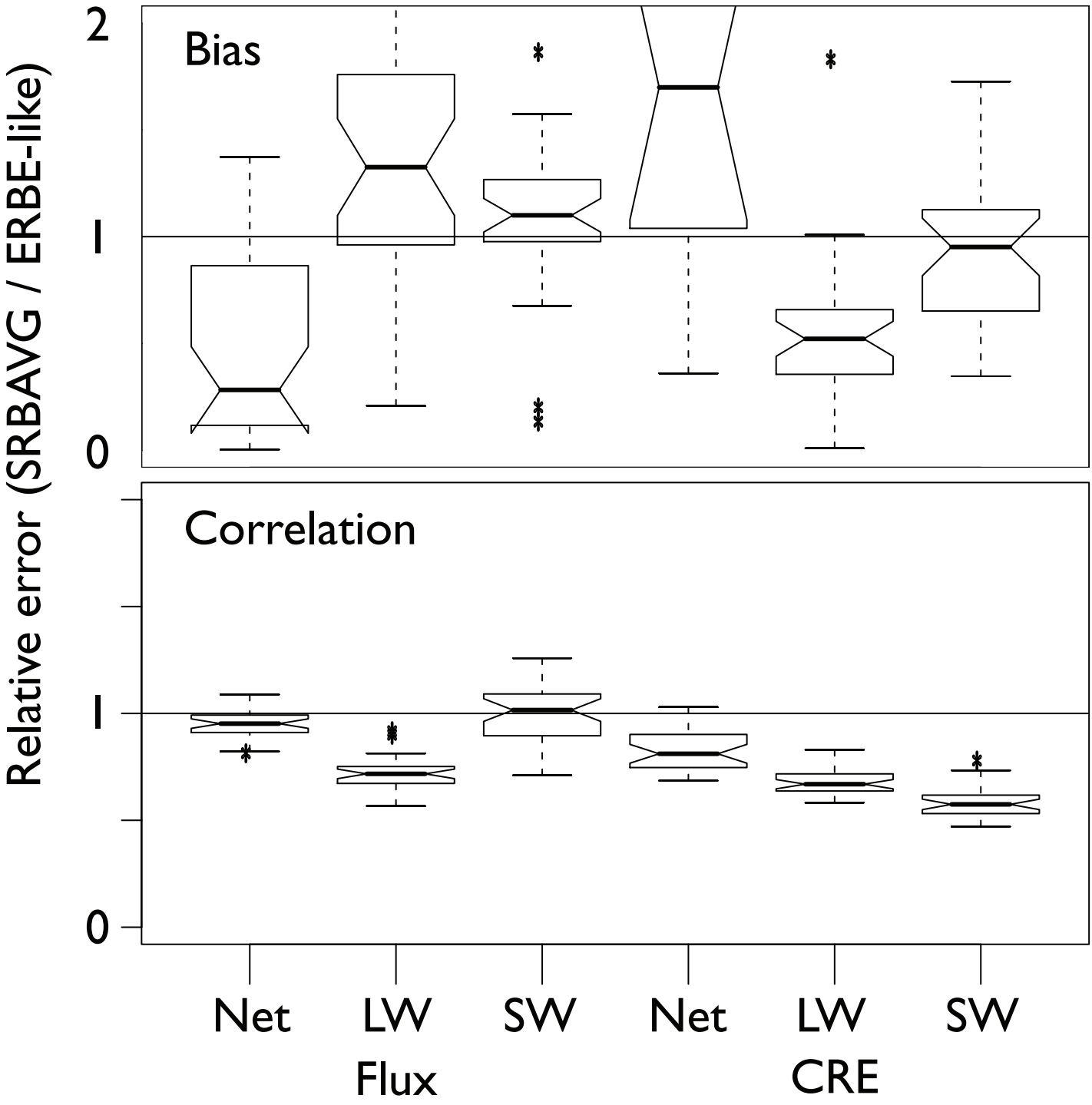
What if we'd had modern flux measurements?

We passed on SRBAVG-nonGEO because of missing data (clear-sky and CRE)

Then Norm Loeb gave us access to SRBAVG-GEO

Differences are interesting

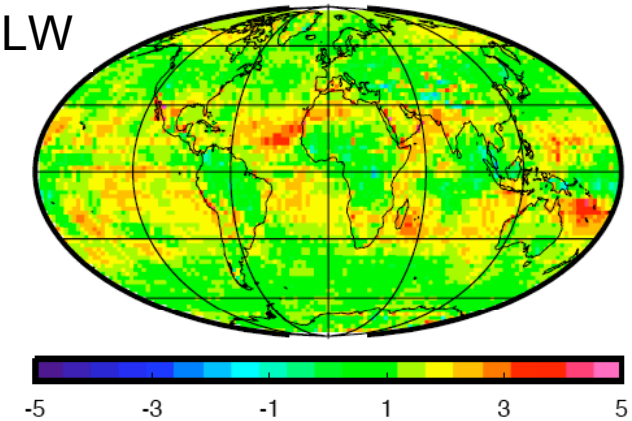
- TOA flux balance reduces bias (esp net flux)
- Models are generally better correlated with SRBAVG than with ERBE-like
- Differences with ES-4 and ERBE can be large relative to model errors (that's a first)


Bias
Correlation
Relative error (SRBAVG / ERBE-like)
2
1
0
1
0
Net
LW
SW
Net
LW
SW
Flux
CRE

CERES ERBE-Like minus nonGEO All-Sky TOA Flux Difference (Wm^{-2})

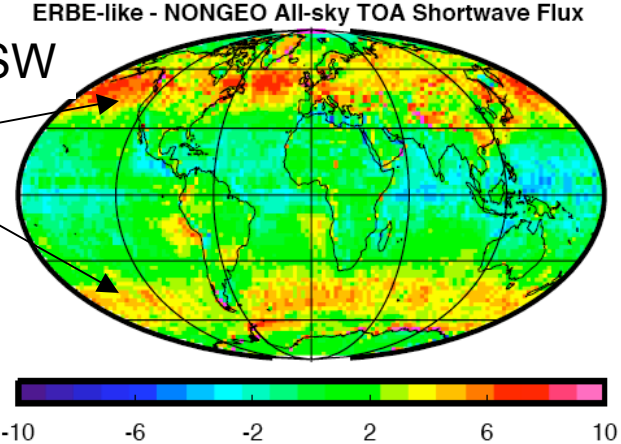
- Differences due to Scene iD + ADMs
- ERBE albedo increase with viewing geometry more pronounced at high latitudes.

LW



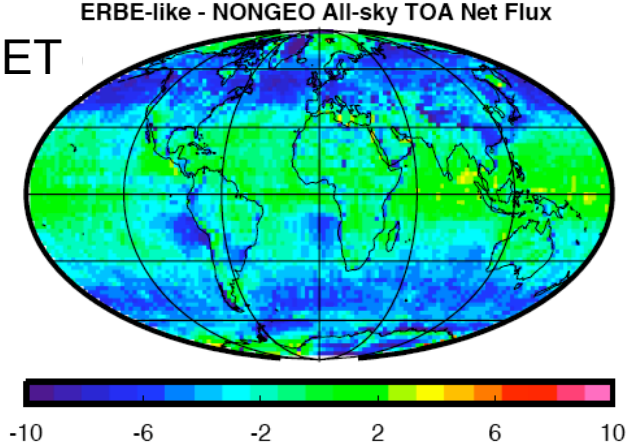
ERBE-like - NONGEO All-sky TOA Shortwave Flux

SW



ERBE-like - NONGEO All-sky TOA Net Flux

NET

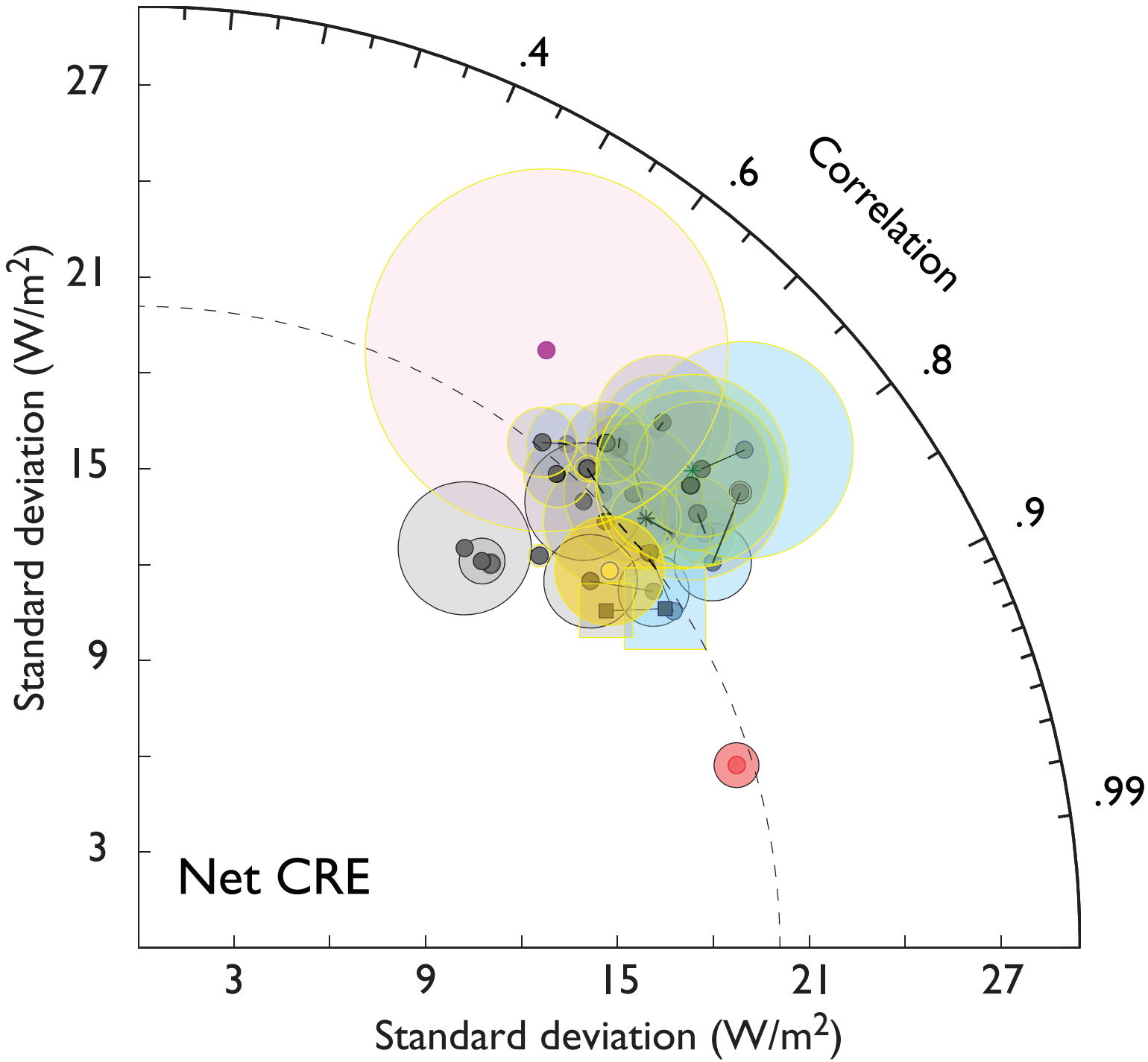


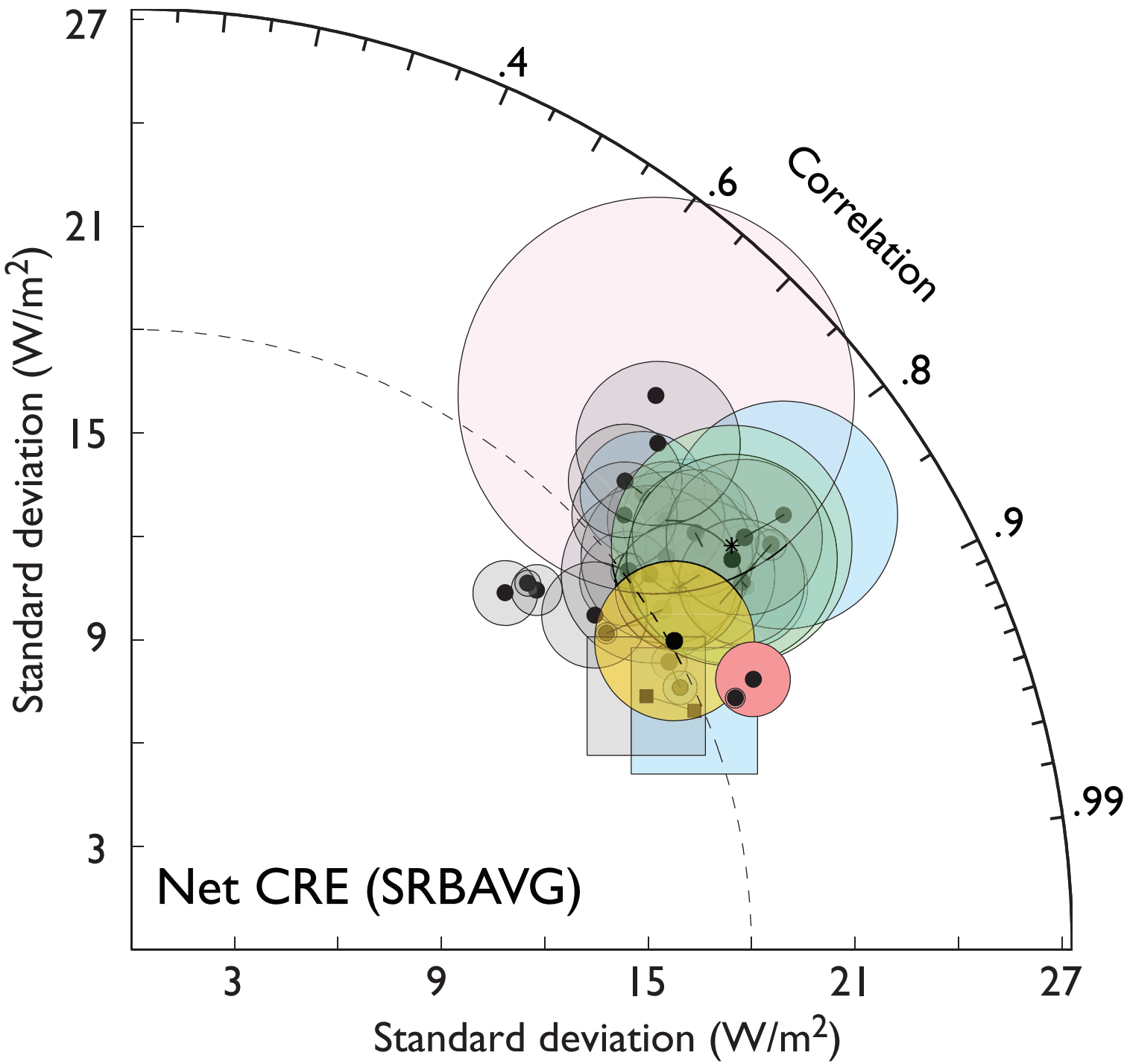
Global Mean Difference

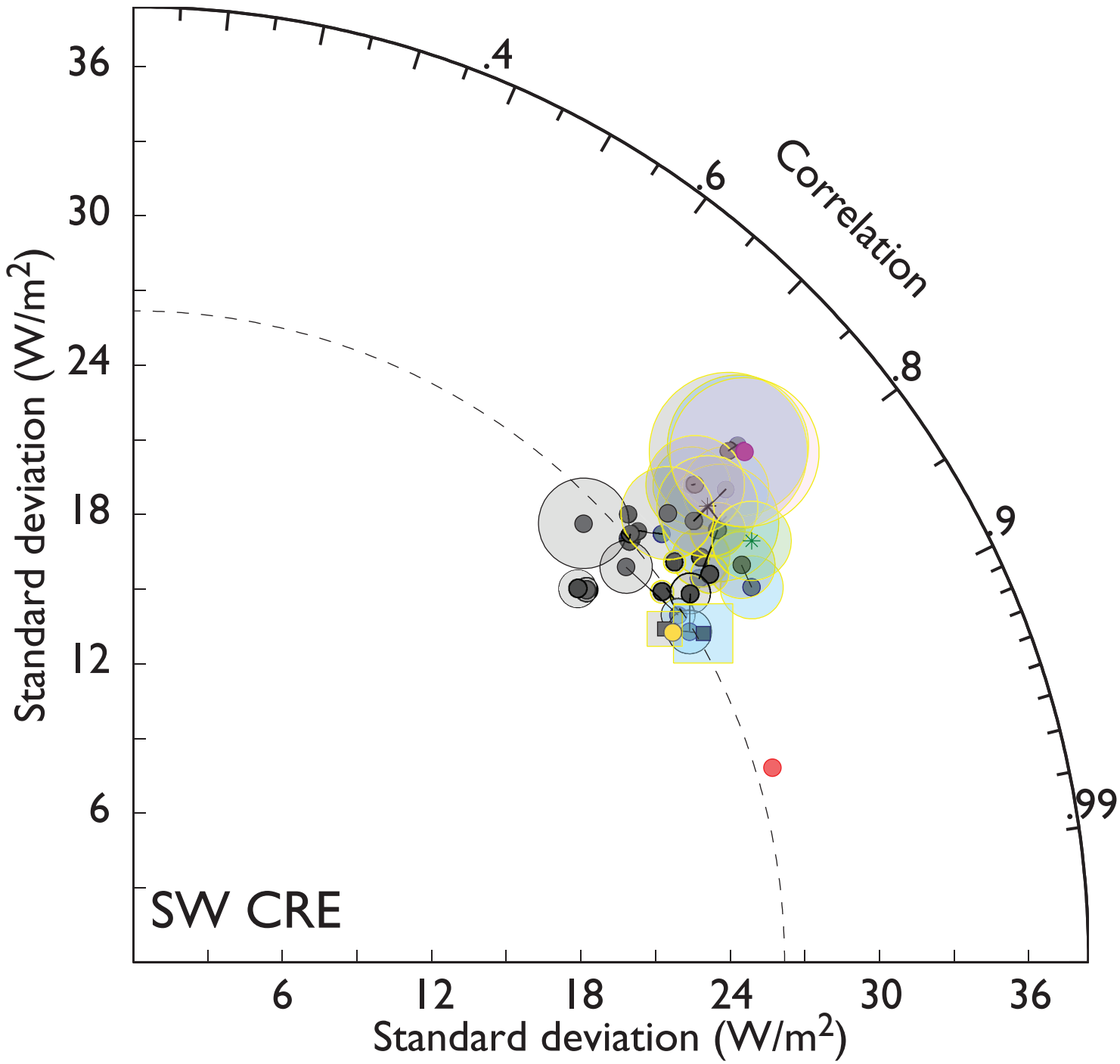
1.3 Wm^{-2}

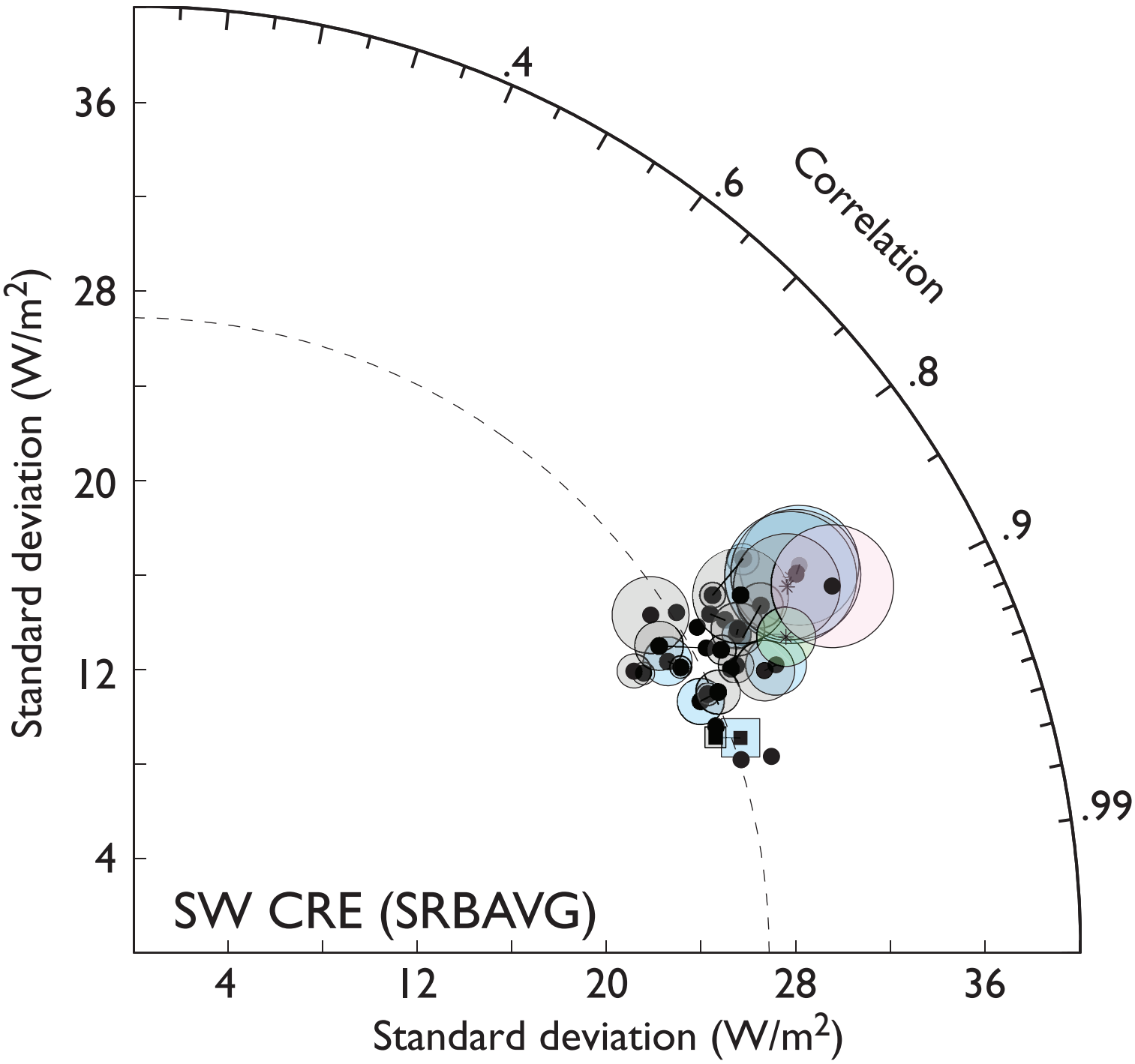
1.7Wm^{-2}

-3.0 Wm^{-2}


Net CRE
Correlation
.4
.6
.8
.9
.99
Standard deviation (W/m^2)
3
9
15
21
27
Standard deviation (W/m^2)
3
9
15
21
27


Net CRE (SRBAVG)
Correlation
.4
.6
.8
.9
.99
27
21
15
9
3
Standard deviation (W/m2)
3
9
15
21
27
Standard deviation (W/m2)


SW CRE
Correlation
.4
.6
.8
.9
.99
36
30
24
18
12
6
Standard deviation (W/m2)
6
12
18
24
30
36
Standard deviation (W/m2)


SW CRE (SRBAVG)
Correlation
.4
.6
.8
.9
.99
36
28
20
12
4
Standard deviation (W/m2)
4
12
20
28
36
Standard deviation (W/m2)

Other things we learned

No model excels in all metrics

Averaging improves scores

- Ensemble means score better than individual members
- All-model mean is much better than individual models
- Model results are distributed around observations

Models have characteristic errors

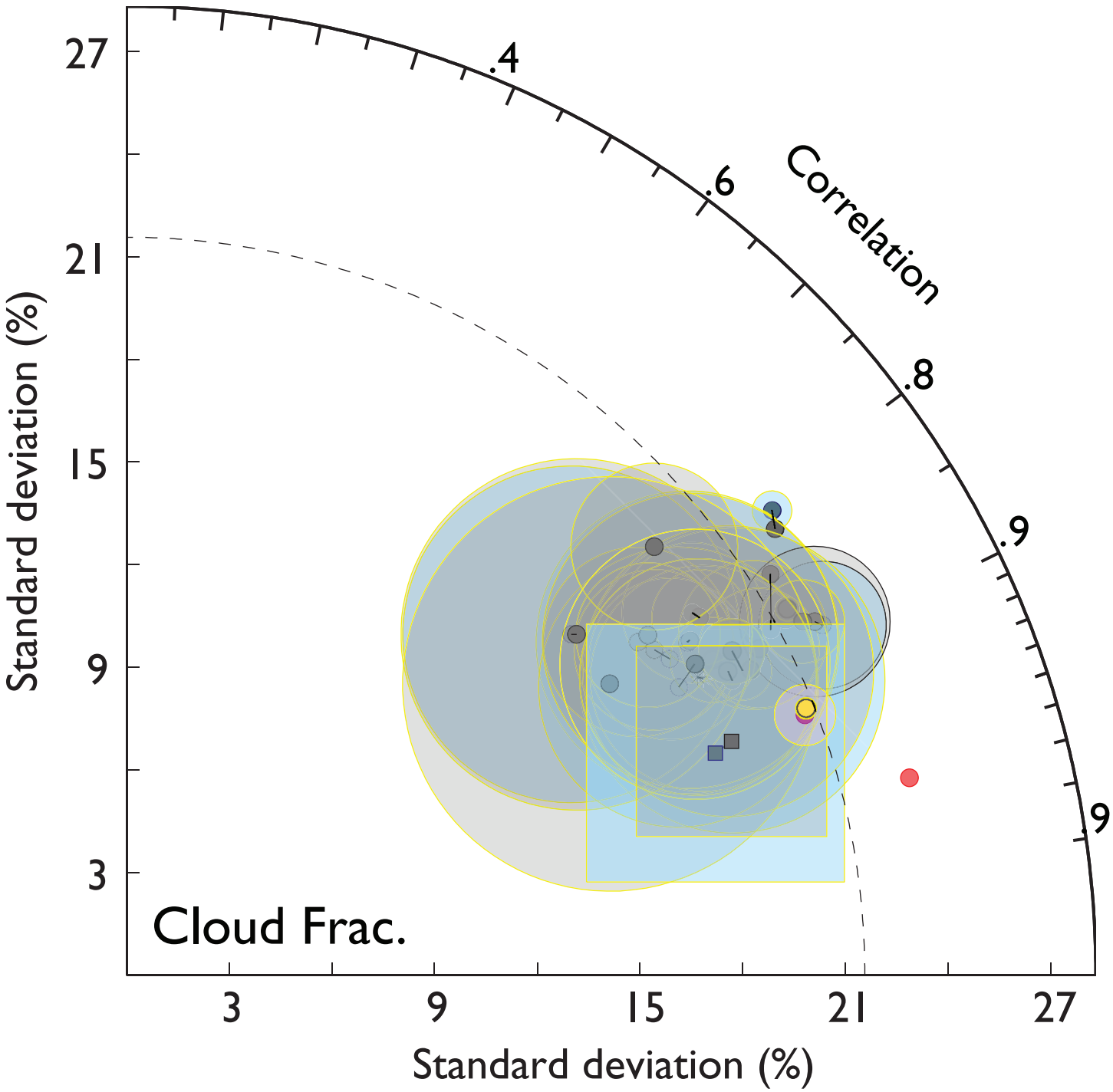
- AMIP and 20th century scores are similar per model
- Super-CAM MMF isn't tons better than CAM

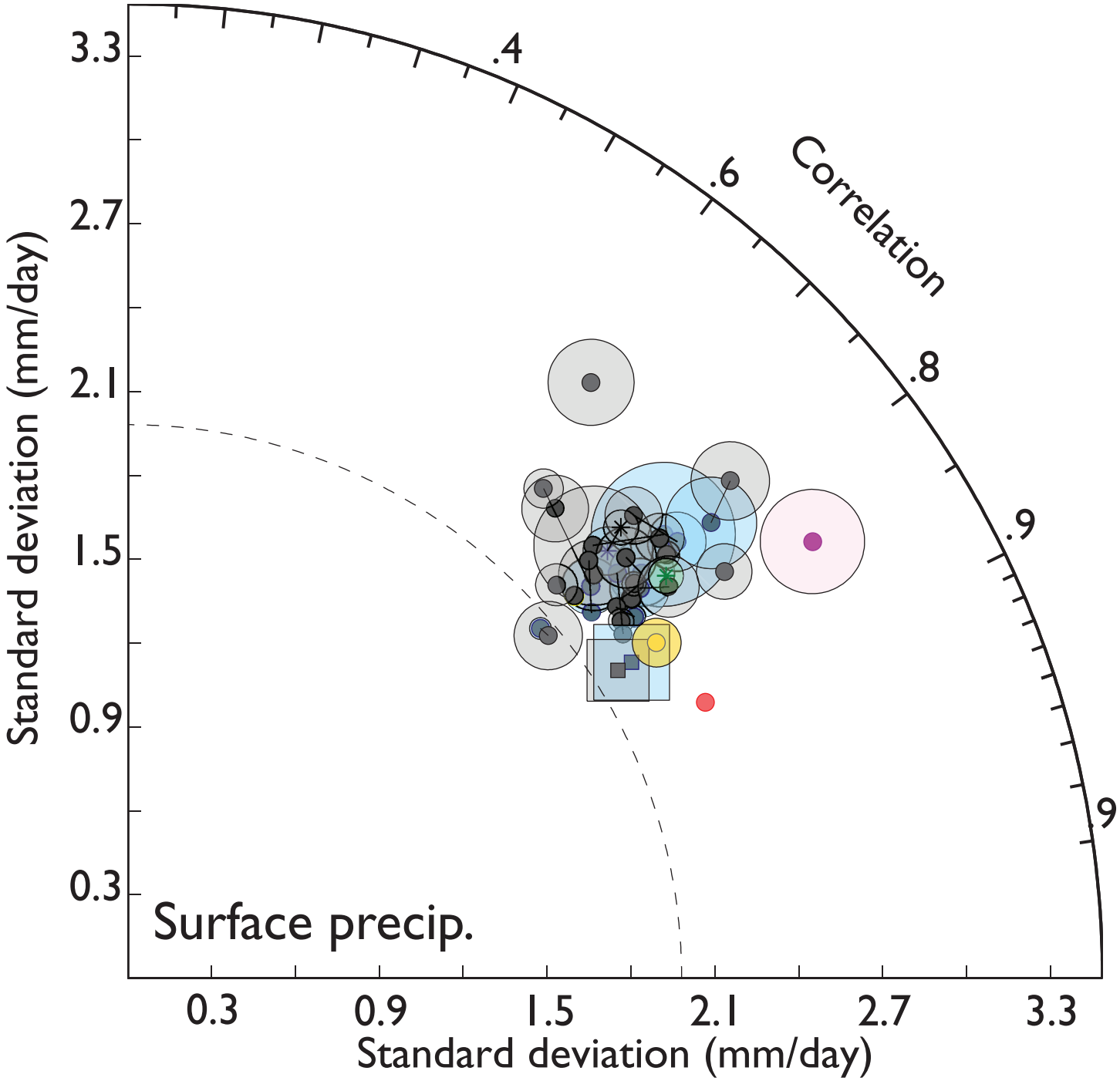
What to remember

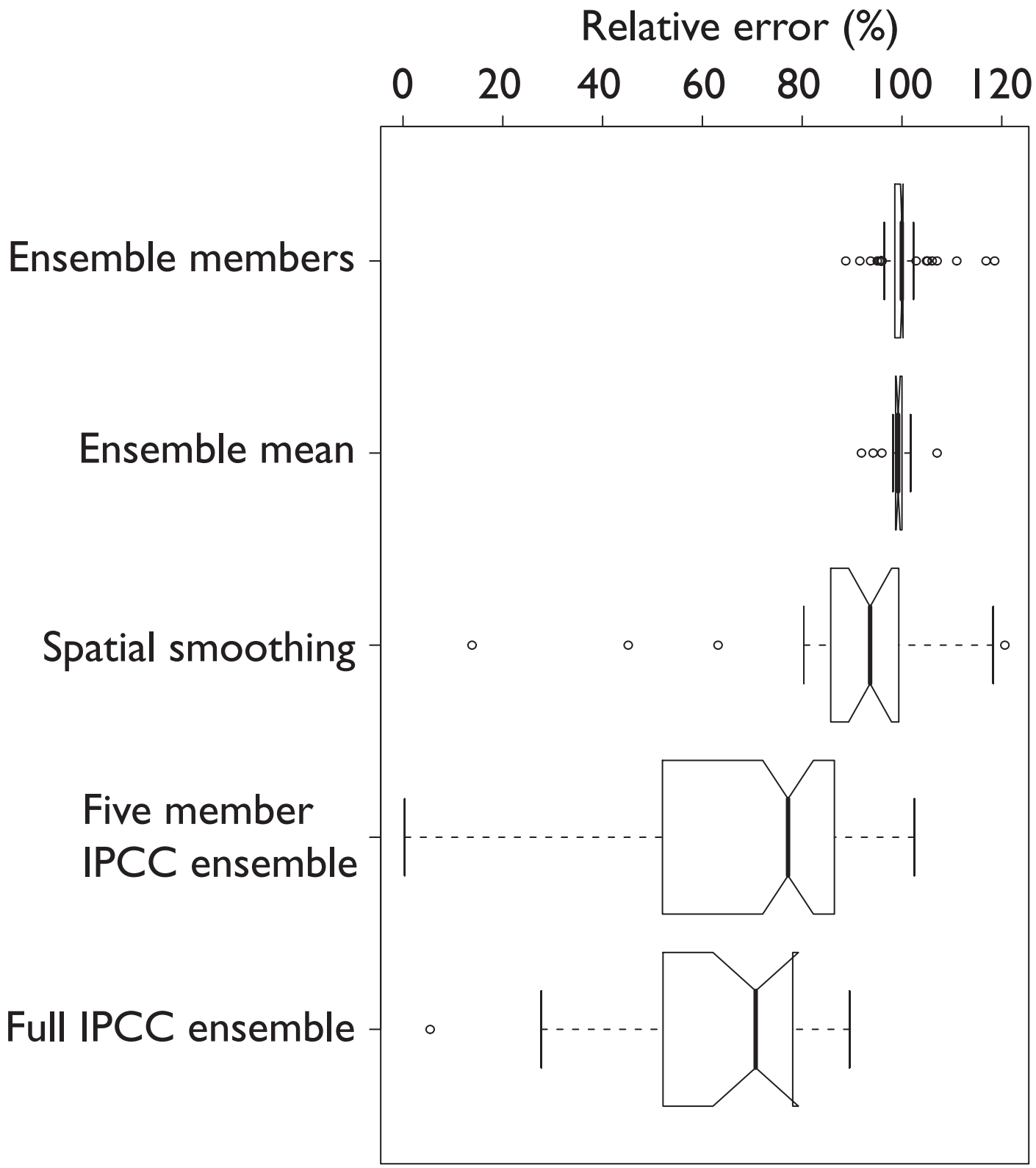
Model skill is still generally less than observational "uncertainty"

The biggest open question is how to relate measurable skill to confidence in climate change predictions

SRBAVG-GEO is precisely what the modeling community has been waiting for


Correlation
.4
.6
.8
.9
.9
Standard deviation (%)
27
21
15
9
3
Cloud Frac.
3
9
15
21
27
Standard deviation (%)


Surface precip.
Correlation
.4
.6
.8
.9
.9
Standard deviation (mm/day)
3.3
2.7
2.1
1.5
0.9
0.3
0.3
0.9
1.5
2.1
2.7
3.3
Standard deviation (mm/day)


Relative error (%)
0
20
40
60
80
100
120
Ensemble members
Ensemble mean
Spatial smoothing
Five member
IPCC ensemble
Full IPCC ensemble