

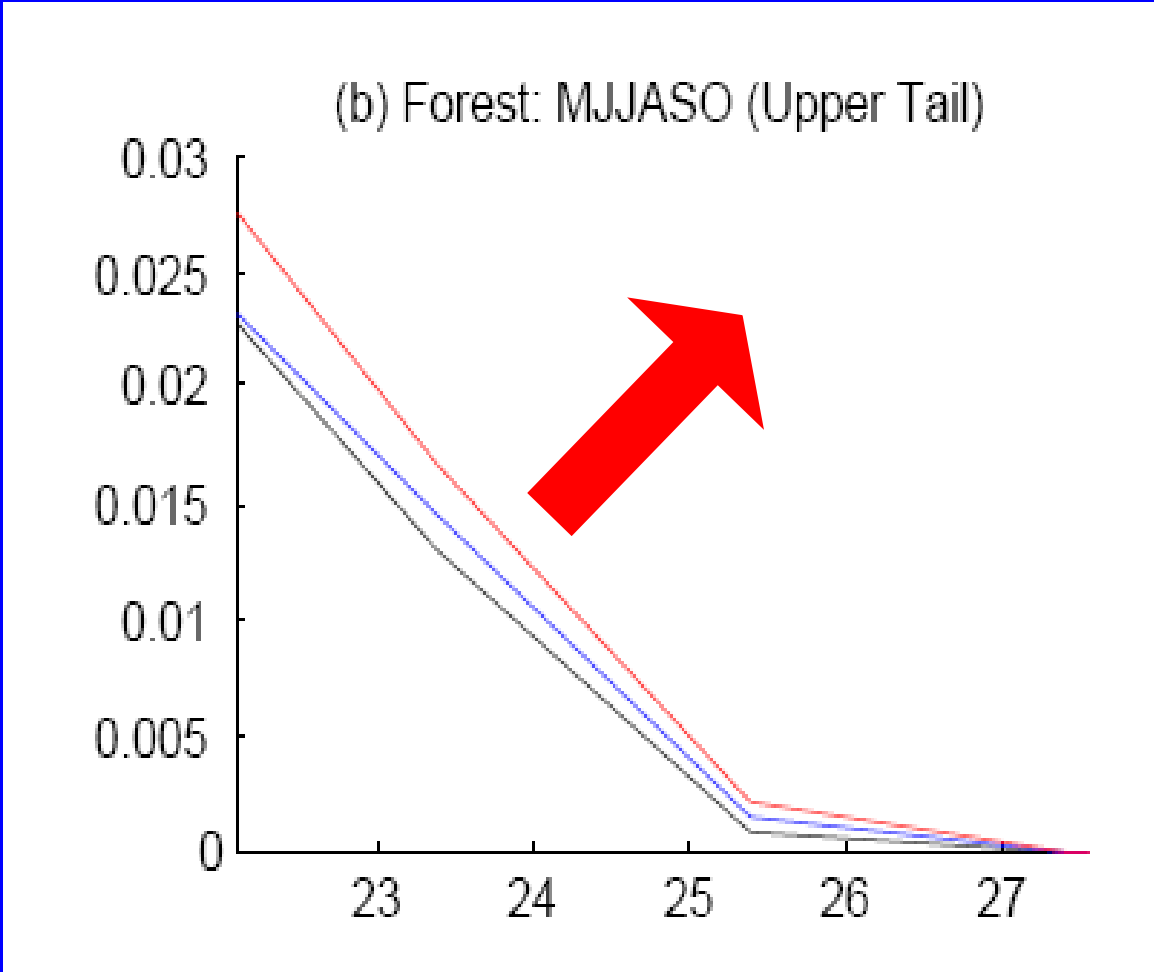
GCM Assessment using PDFs [and other thoughts]

Andy Pitman
Climate Change Research Centre

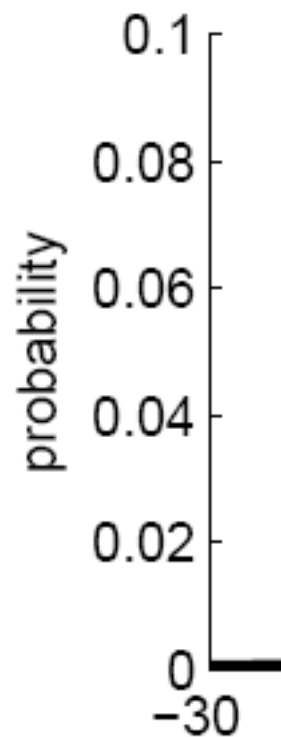
Why a “unified approach”

- Simulations over a few days identify [some] systematic errors in the models – errors that are apparent in longer term means
- This suggests [to me] that looking at finer spatial scales gives you some useful insights
- Clearly cannot look at $\Delta t \cdot N_p \cdot N_l \cdot N_r \cdot N_m$ data
- So, we [typically] average over some dimensions – time, area, levels, realizations, models ...
 - And hide potentially important errors and phenomenon

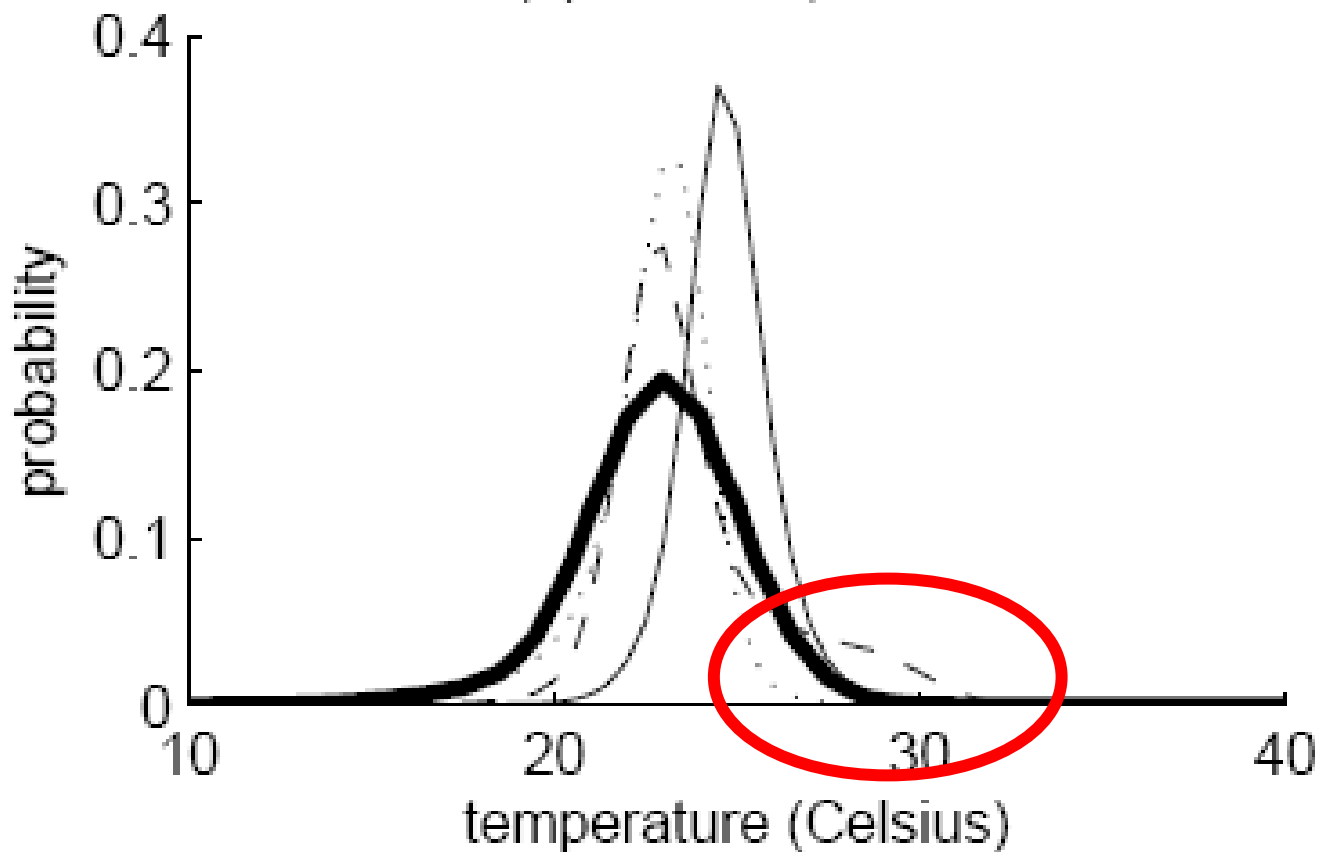
Impact of increase of leaf-level CO₂ from 280 to 350 to 500



(d) Baltic, Tmax

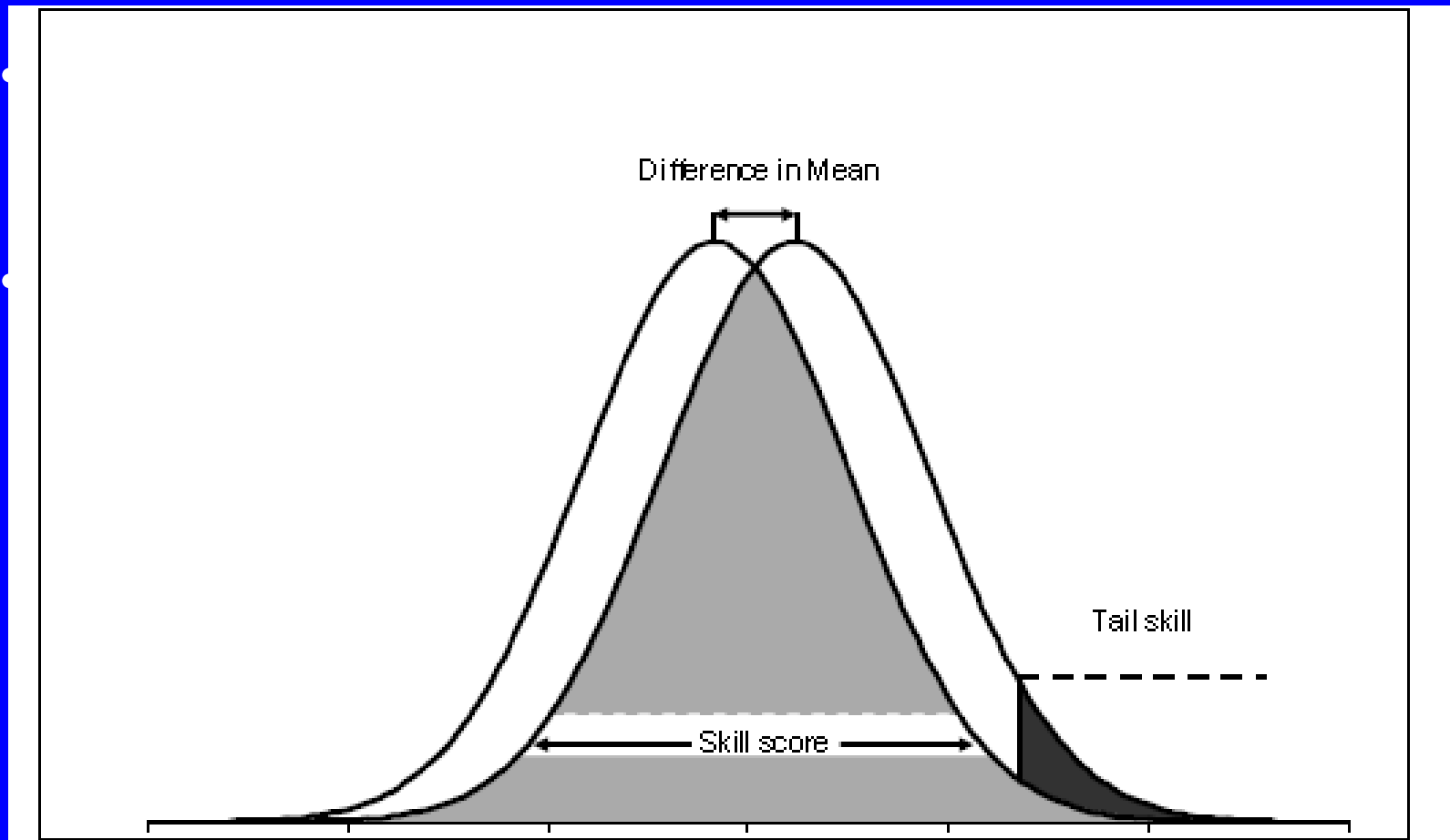


(b) Amazon, Tmin

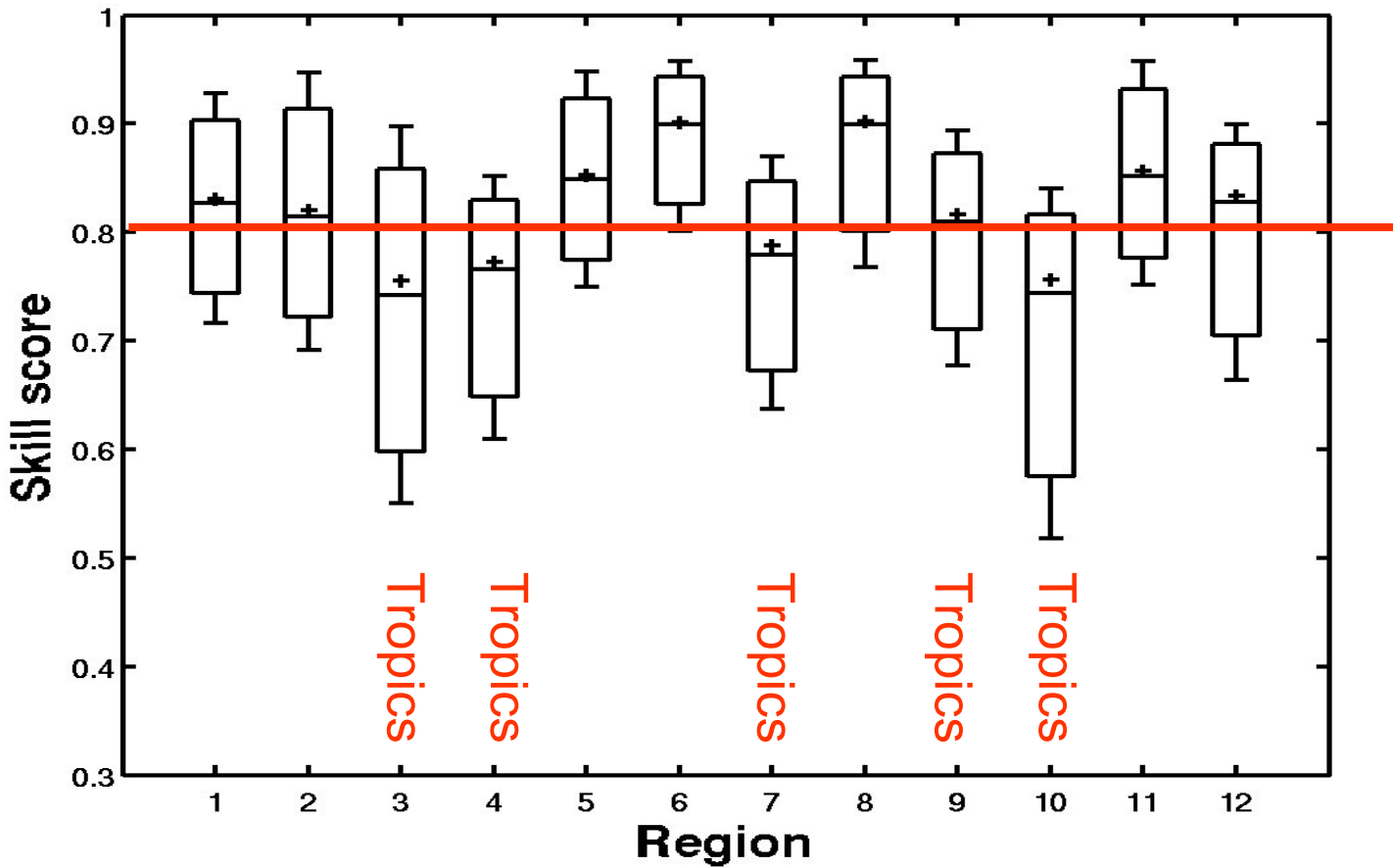


UNSW efforts

- used AR4 IPCC models, and used daily data



Regional skill - precipitation



Climate models are generally tropically challenged

Advantages and disadvantages

- Advantages
 - All data can be used
 - Transferable and comparable metric
 - Easy and intuitive
 - Preserves extremes and anomalies
 - Quite robust to data gaps and errors
 - Discriminates between models well

Advantages and disadvantages

- Disadvantages
 - No preservation of time dimension
 - Need highest resolution data possible
 - Need to sample the PDF to plot surfaces – or plot a skill surface

Thoughts on evaluation

- Evaluation is application specific
- there is no “right way” - many useful ways
- AR4 methods are *fine* at global scales

Thoughts on evaluation

- A “pool” of metrics is difficult to design to avoid hidden duplication and bias or non-sampling of important things
- Means based evaluation are:
 - *fine* if you are focused on mean responses
 - invalid if you then focus on extremes

Conclusions

- Our pdf-based methods are not *the* solution
 - they could be part of a “collection of metrics”
- Any collection should include many non-overlapping measures
- Need to exist as a tool box
- Metrics are of no use without a *benchmark*.
 - The best model may be useless
 - The worse model may be good enough
 - What are the standards we are striving to meet?

Conclusions

- How good does the ACCESS model have to be to be good enough ?
- What does the ACCESS model have to be able to be able to do to be useful ?
- If we do not know the answer to these questions, then what is the basis for the decisions on what needs to be in the model ?