

Key scientific advances since the AR5?

- Beginning to recognise the limits of performance based weighting
- The emergence of interdependence based weighting schemes and recognition of their importance
- The understanding that PPEs and MMEs are not transitive
- The use of perfect models to perform out of sample testing
- Applying data assimilation approaches to paleo time series has advanced and has applications related to understanding decadal variability

Shortcomings, gaps and opportunities?

- Shortcomings
- Metrics are not transitive, optimising one generally doesn't provide benefits for others, not possible to optimise all metrics
- The importance of tuning choices in confounding weighting
- The recognition mean state weighting has pitfalls (e.g. tuning, arbitrary model assumptions etc)
- The danger of performance weighting based on the right answer for the wrong reason-ensuring you understand the mechanism
- Currently no concrete recommendations for how to move beyond model democracy

- Gaps
- Identifying the role of structural vs internal variability across MMEs

- Opportunities
- Moving beyond mean state weighting, incorporating temporal variability into weighting schemes
- The combination of weighting schemes and emergent constraint approaches
- Knowledge transfer between the weighting scheme community and the DandA community
- Use long time series records both historical and paleo to understand internal variability
- Recommended weighting schemes for specific purposes

Long-term perspective?

- Moving beyond model democracy is a likely imperative
- Identifying the role of structural vs internal variability across MMEs
- Moving beyond mean state weighting, incorporating temporal/spatial variability into weighting schemes
- The development of a statistical methodology/ strategy to account for multiple chains of influence, orthogonal uncertainties
- Better documentation that tracks model interdependence
- What does it mean for models to be interdependent, moving beyond error covariance and model lineage
- Assessing model interdependence by examining simpler idealised problems, toy models, idealised experiments etc