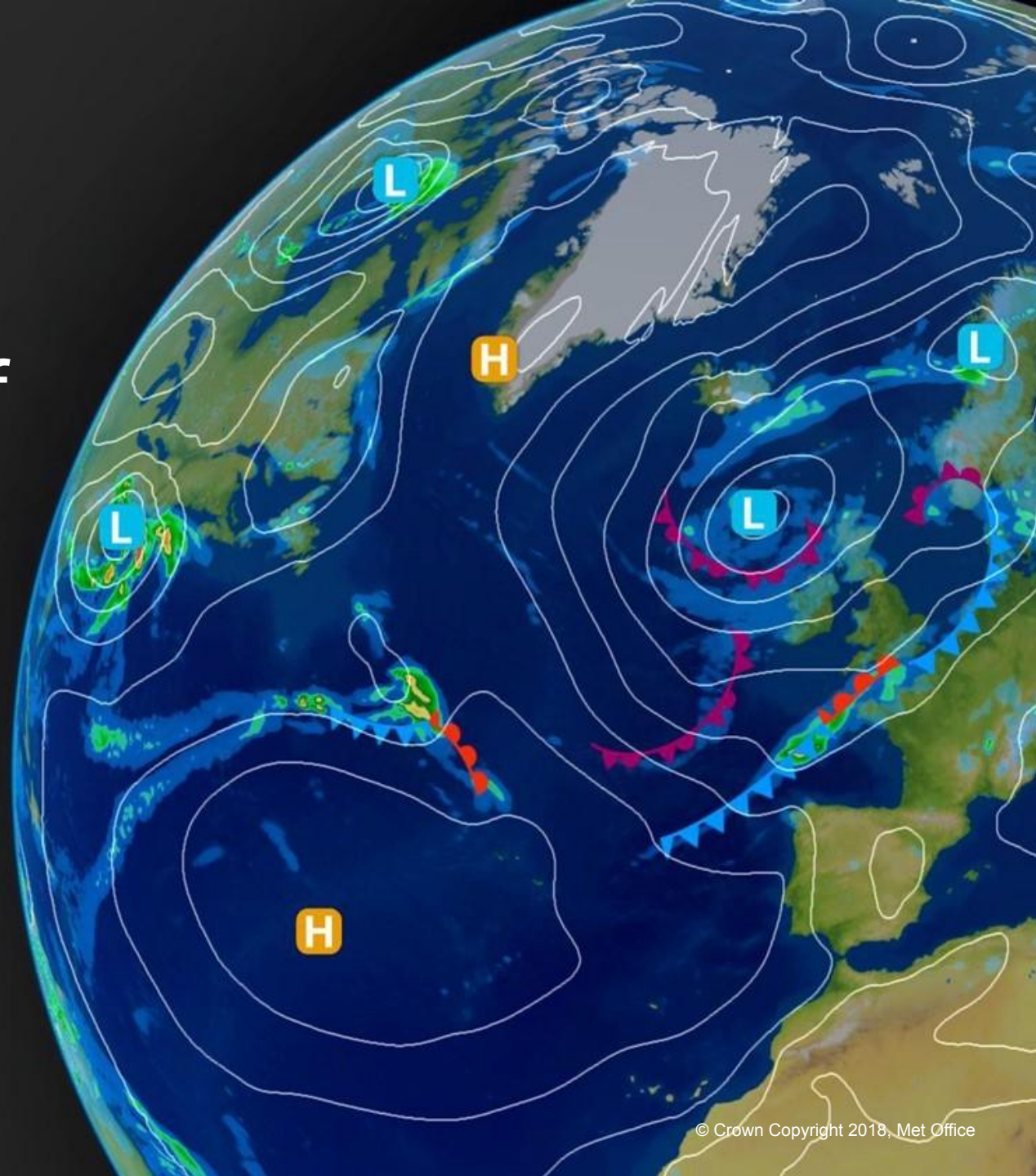
Met Office
Hadley Centre

A potential new approach for assessing suitability of new model versions for seasonal forecasting

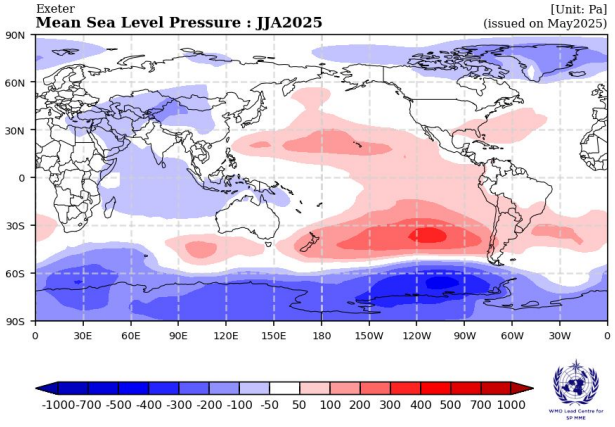


Jeff Knight, Met Office

Seamless Global Modelling Workshop
Bristol

6th June 2025

Met Office Seasonal to Decadal Group



- Major customer for Met Office models:
  - 4x GloSea6 forecast members per day => 540 years of simulation per year
  - 28x 24-year hindcast members per month => 4600 years of simulation per year
- GloSea6 phase 2 upgrade in 2025/6:
  - Roughly an increase of factor 3 or 4 => ~15,000 years of simulation each year
- DePreSys
  - 10x forecast => 100 years of simulation per year
  - 10x 70-year hindcast => 7000 years of simulation every few years
  - Will also go 4x larger in 2025/6
- Research simulations, MIPs etc additional
- Customers: UK Government, Copernicus, WMO (including being LC-ADCP) …

Met Office
Hadley Centre

Development cycle

- 2-year cycle:
  - Year 1: package testing, modification
  - End of year 1: freeze
  - Year 2: produce assessment runs
  - Year 2: conduct assessment
  - Hold assessment workshop
  - Release version

Development cycle

- 2-year cycle:
  - Year 1: package testing, modification
  - End of year 1: freeze
  - Year 2: produce assessment runs
  - Year 2: conduct assessment
  - Hold assessment workshop
  - Release version

- Schedule resource to work on model assessment
- Implement GloSea6 system changes for new version
- Rerun ocean assimilation if necessary
- Run selection of hindcasts
- Assess results

Met Office
Hadley Centre

Development cycle

- 2-year cycle:
  - Year 1: package testing, modification
  - End of year 1: freeze
  - Year 2: produce assessment runs
  - Year 2: conduct assessment
  - Hold assessment workshop
  - Release version

- Schedule resource to work on model assessment
- Implement GloSea6 system changes for new version
- Rerun ocean assimilation if necessary
- Run selection of hindcasts
- Assess results

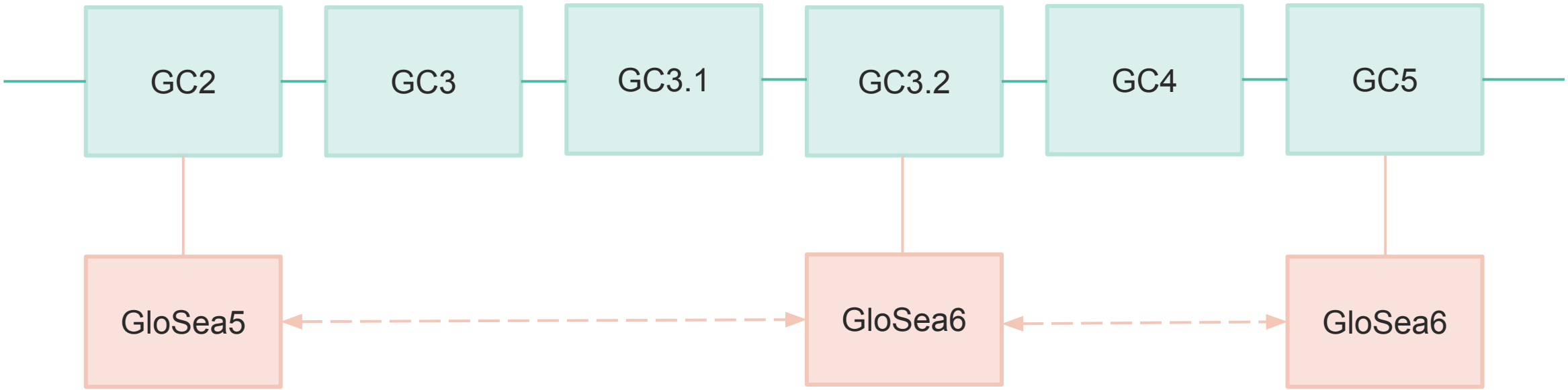
Met Office
Hadley Centre

Can we use AMIP simulations from the assessment

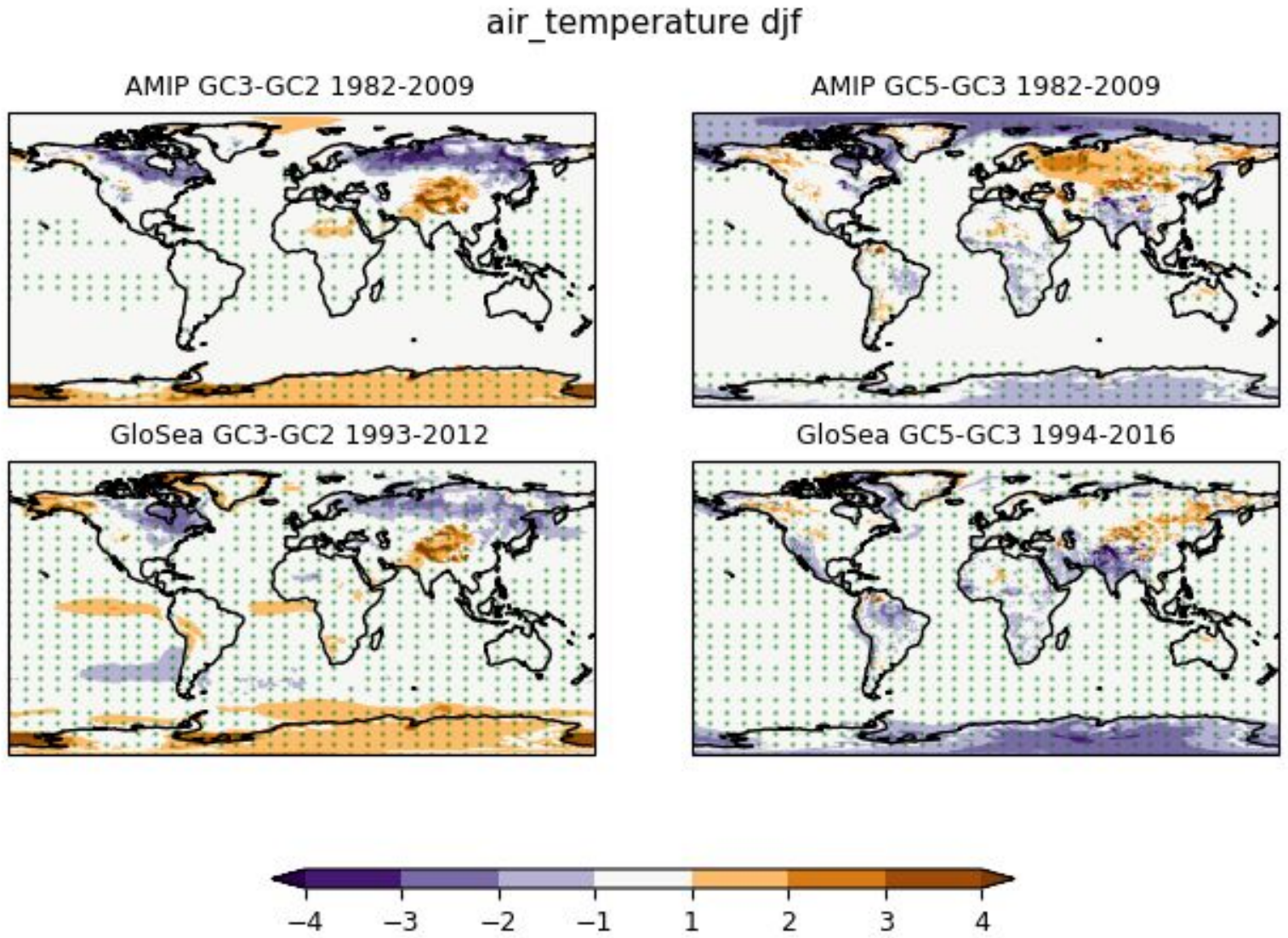
- Produced in the assessment – typically a single ~30-year simulation with observed SST
- Available about a year into the cycle and well in advance of the assessment workshop
- Somewhat of an analogue of seasonal forecasts – early seasonal systems were atmosphere only forced with persisted SSTs
- Seasonal forecasts are initialised with observations so the degree to which they have drifted is less than in a coupled control run

Met Office Hadley Centre

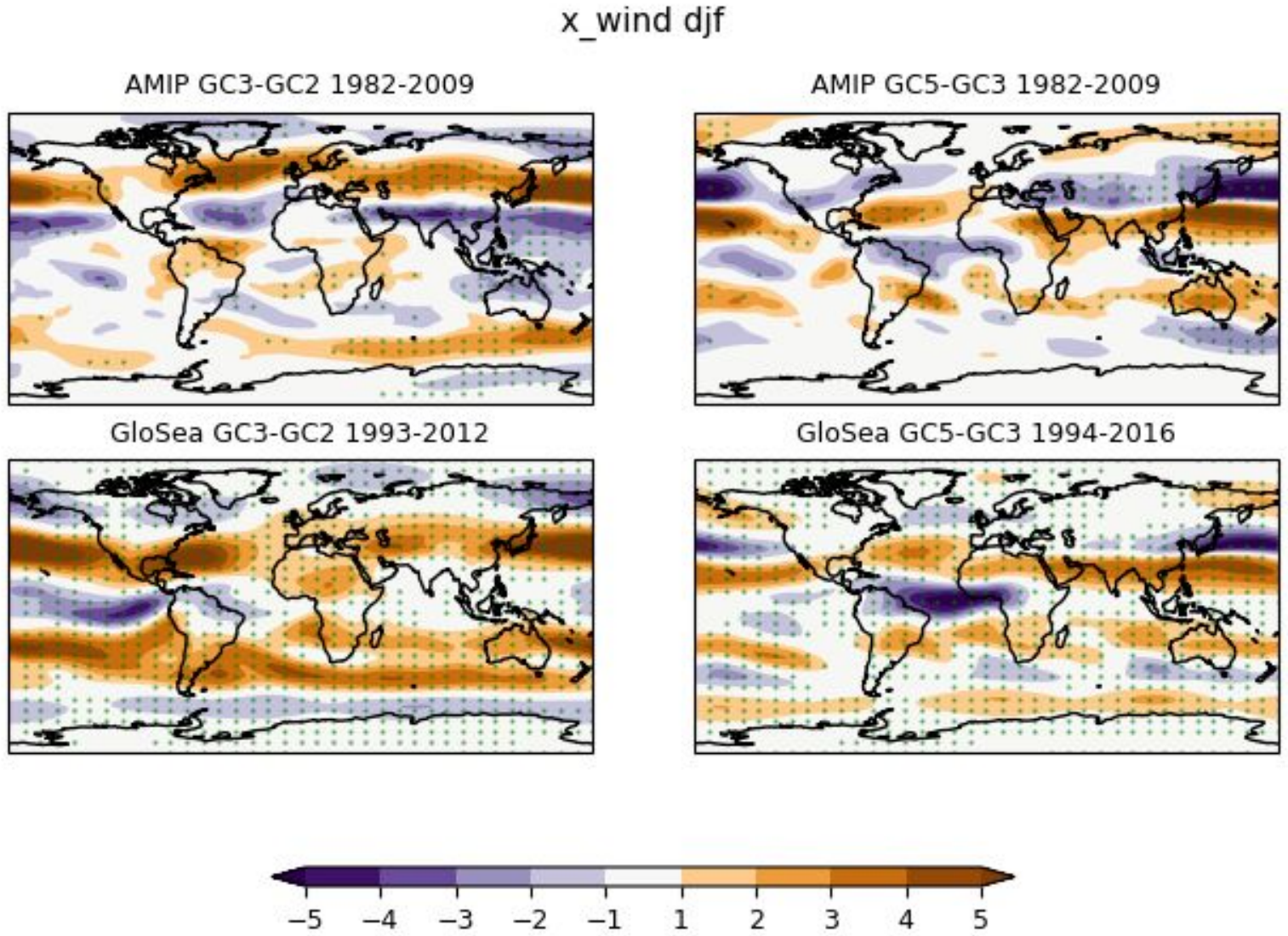
AMIP simulations with matching resolution (N216 O0.25) in same cycles as GloSea hindcast assessments



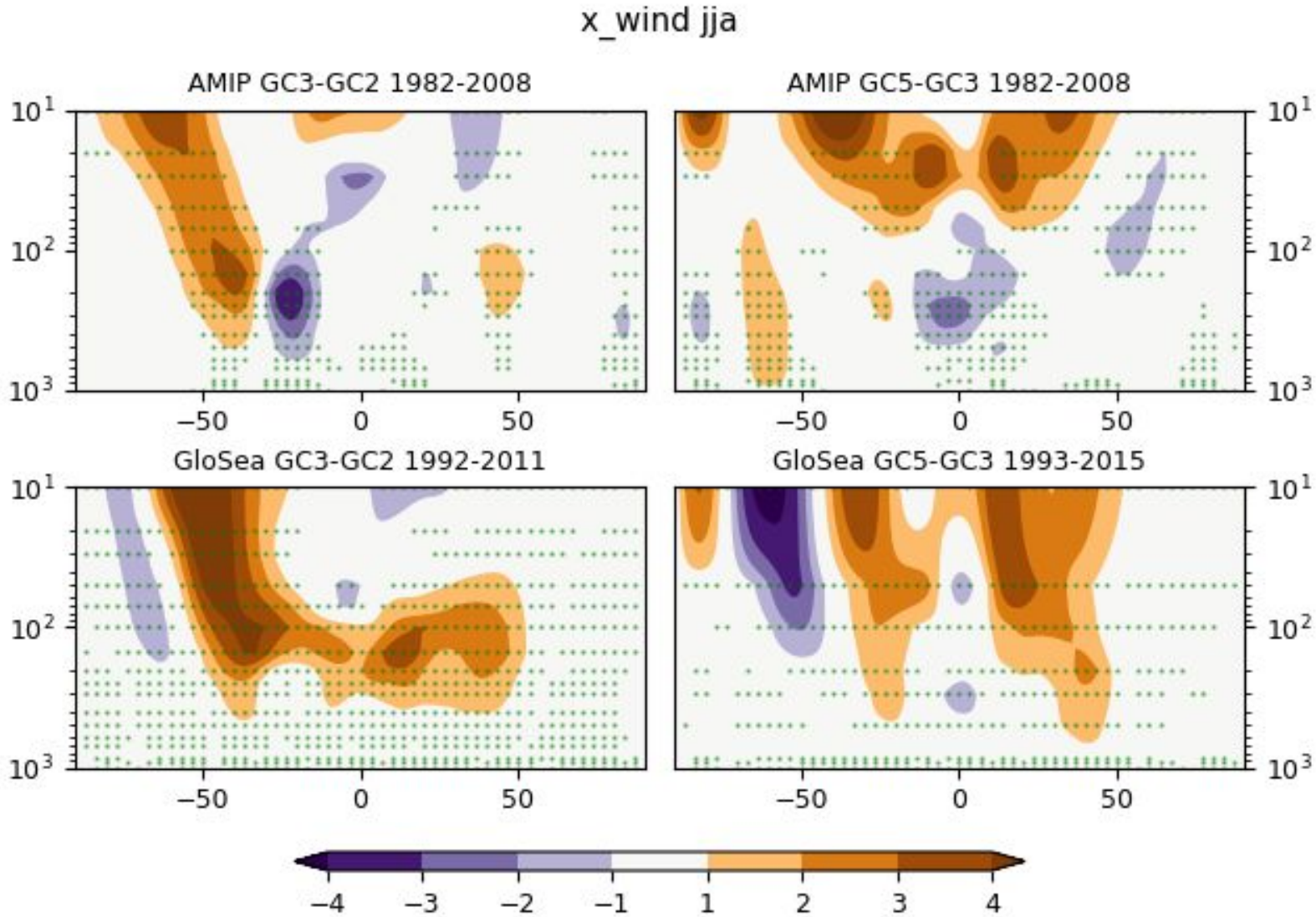
Comparison of change in mean bias (1.5m T DJF mean)



Comparison of change in mean bias (u_{200} DJF mean)

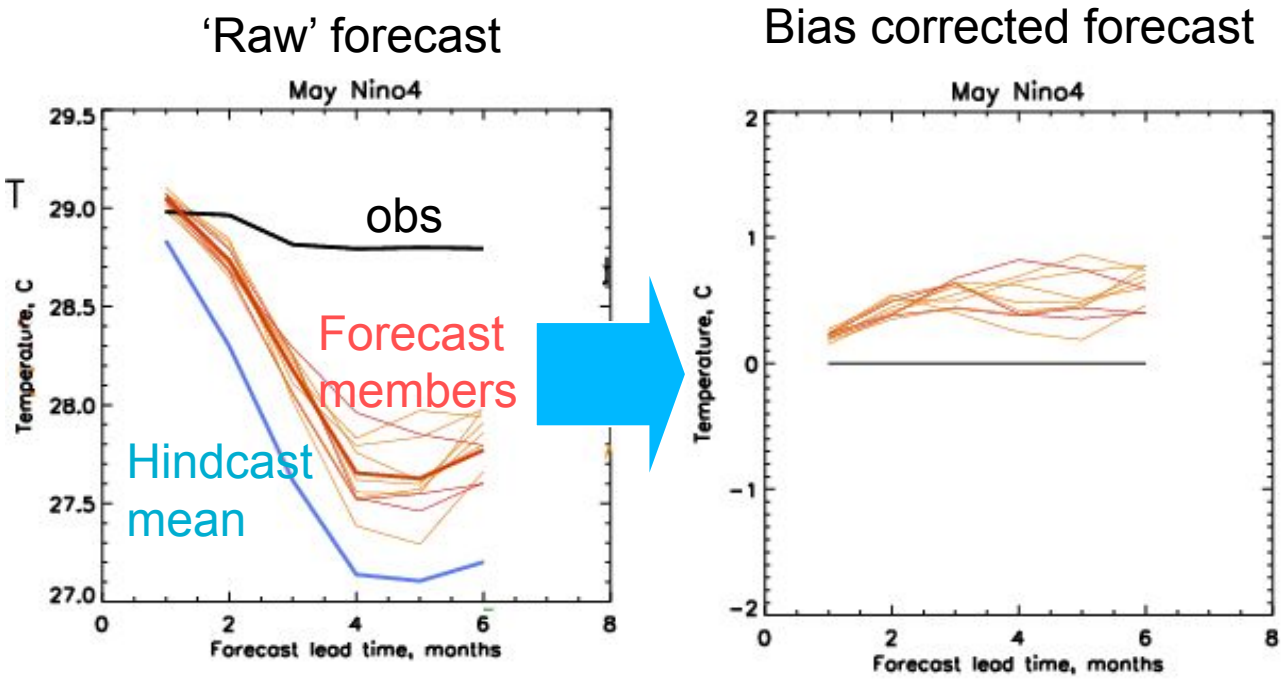


Comparison of change in mean bias ($\bar{u}$ JJA mean)

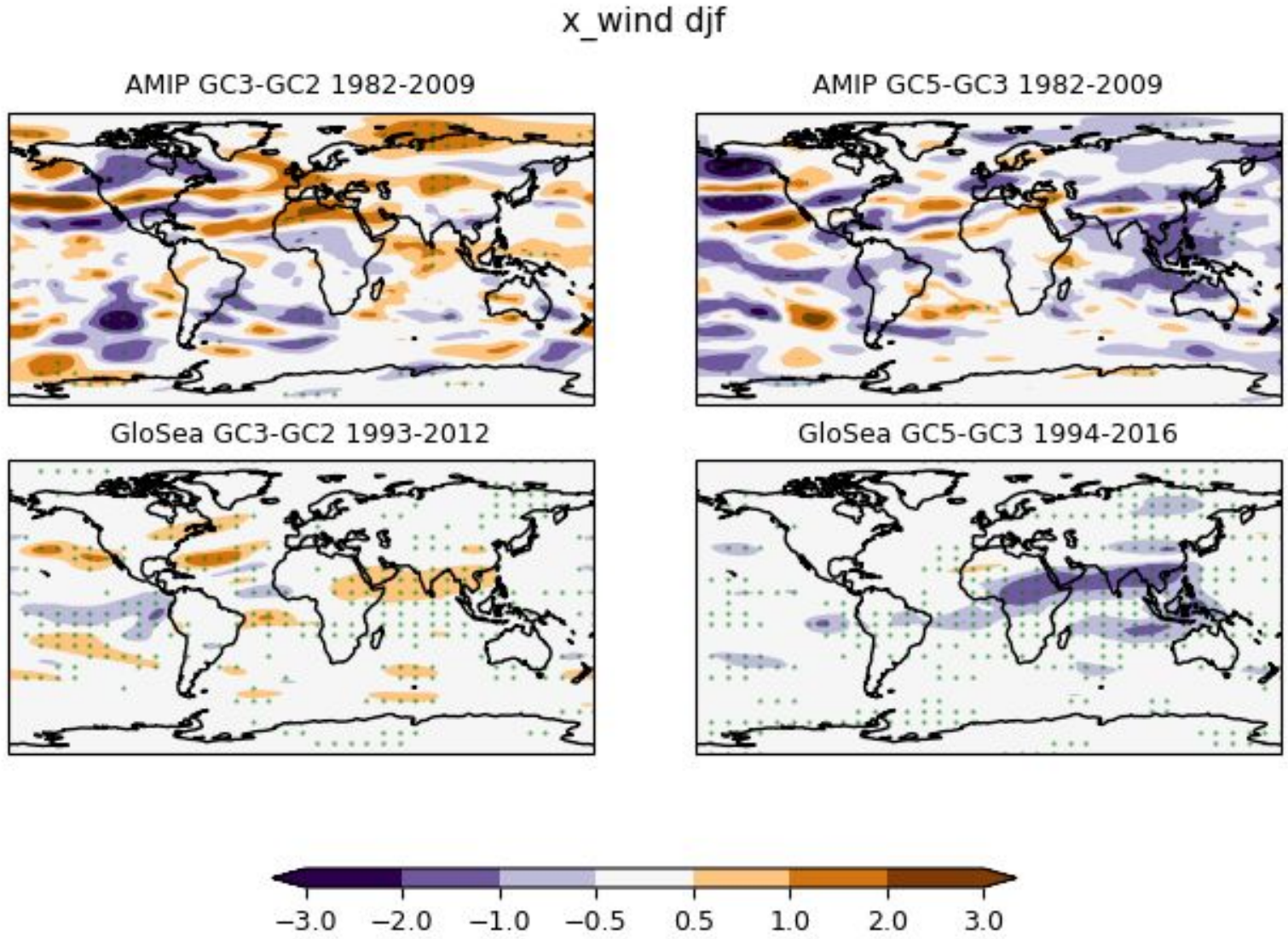


Bias correction in seasonal systems

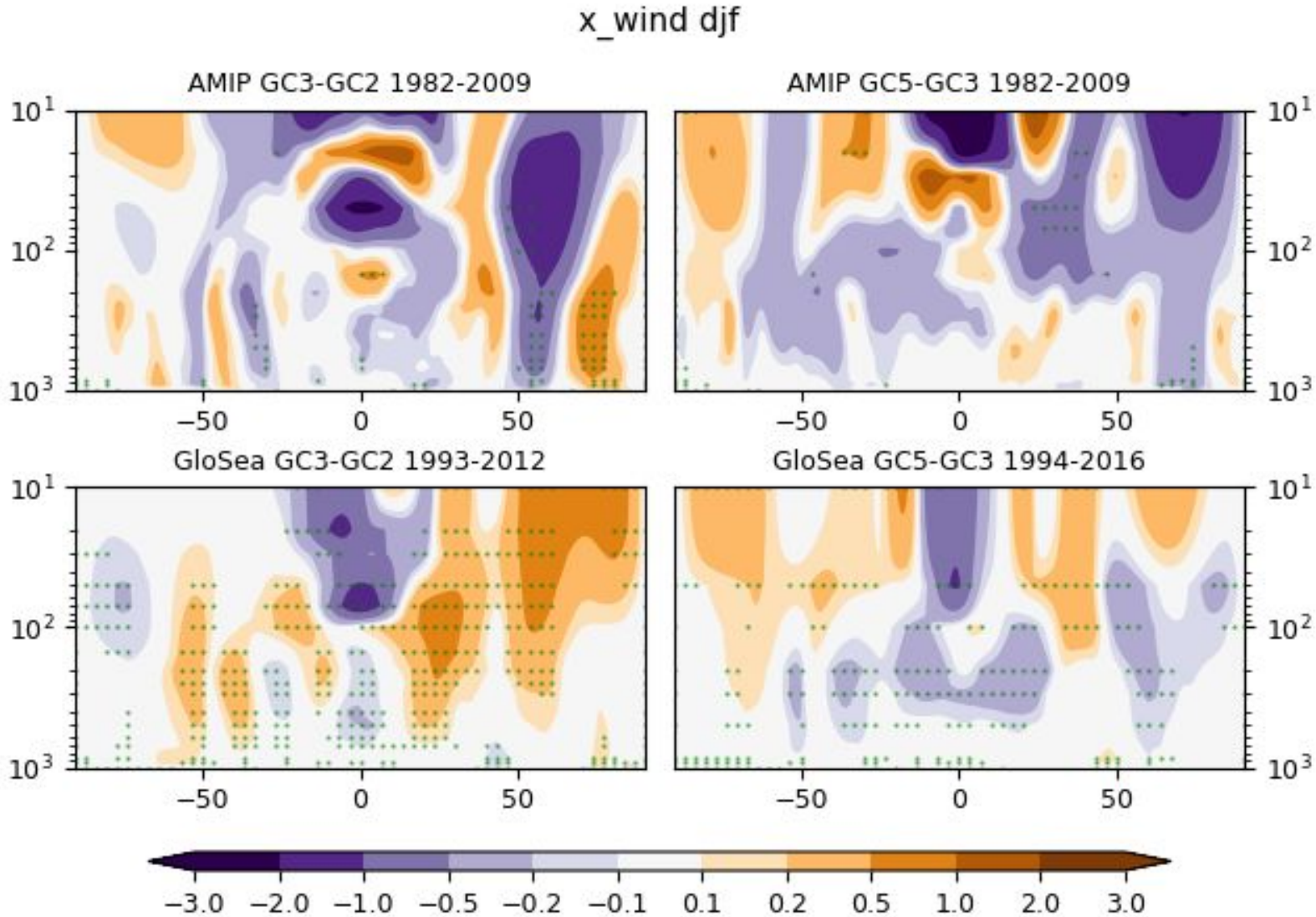
- Hindcasts over past years (GloSea case 1993-2016) provide time-dependent bias correction to forecasts
- While reducing model bias is generally very important (not least because of non-linearities), in seasonal forecasting bias correction reduces the importance of biases
- Mostly, we are concerned with other aspects of performance, such as the scale of year-to-year variability or the strength of teleconnections



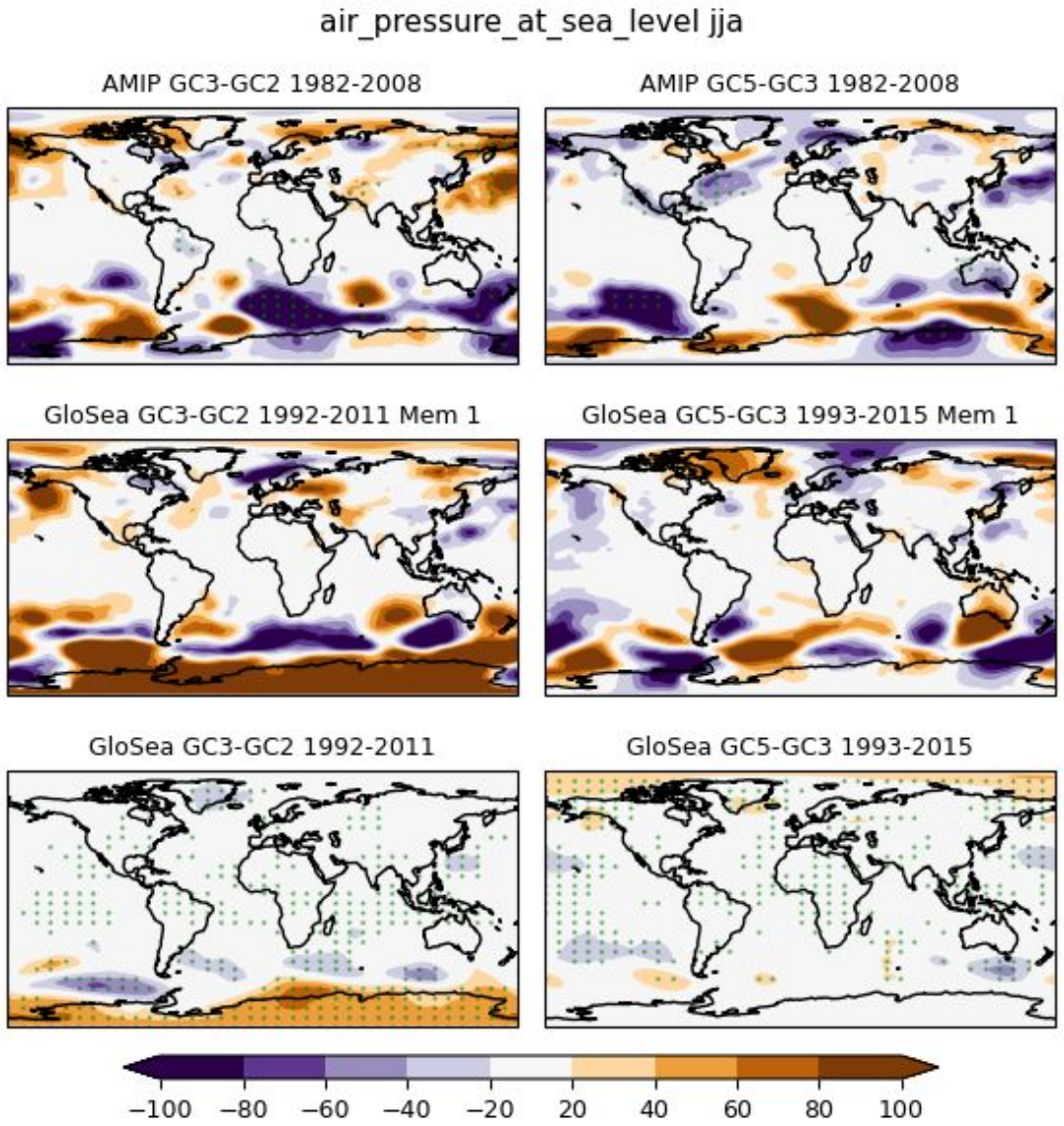
Changes in standard deviation (u_{200} DJF mean)



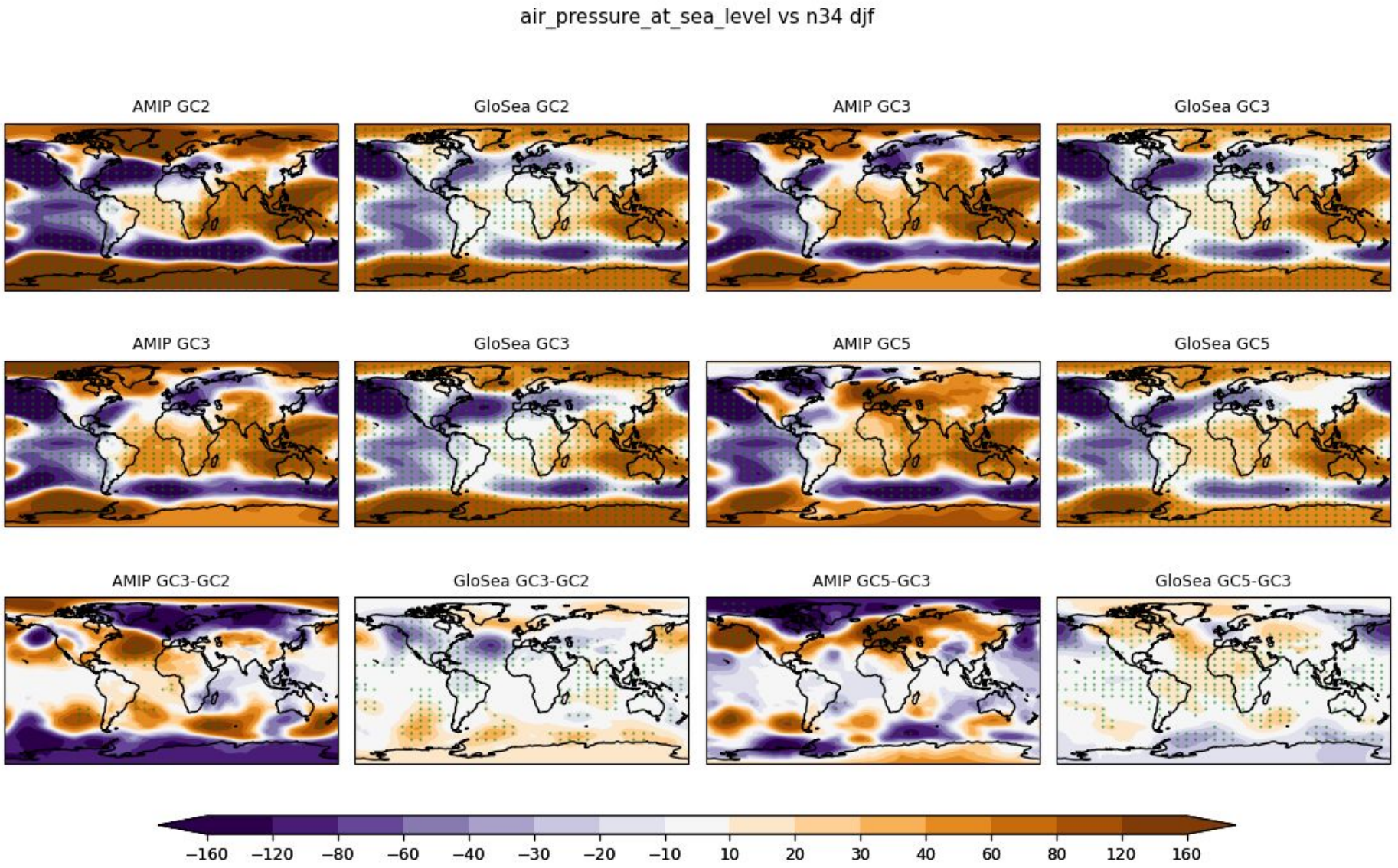
Changes in standard deviation ($\bar{u}$ DJF mean)



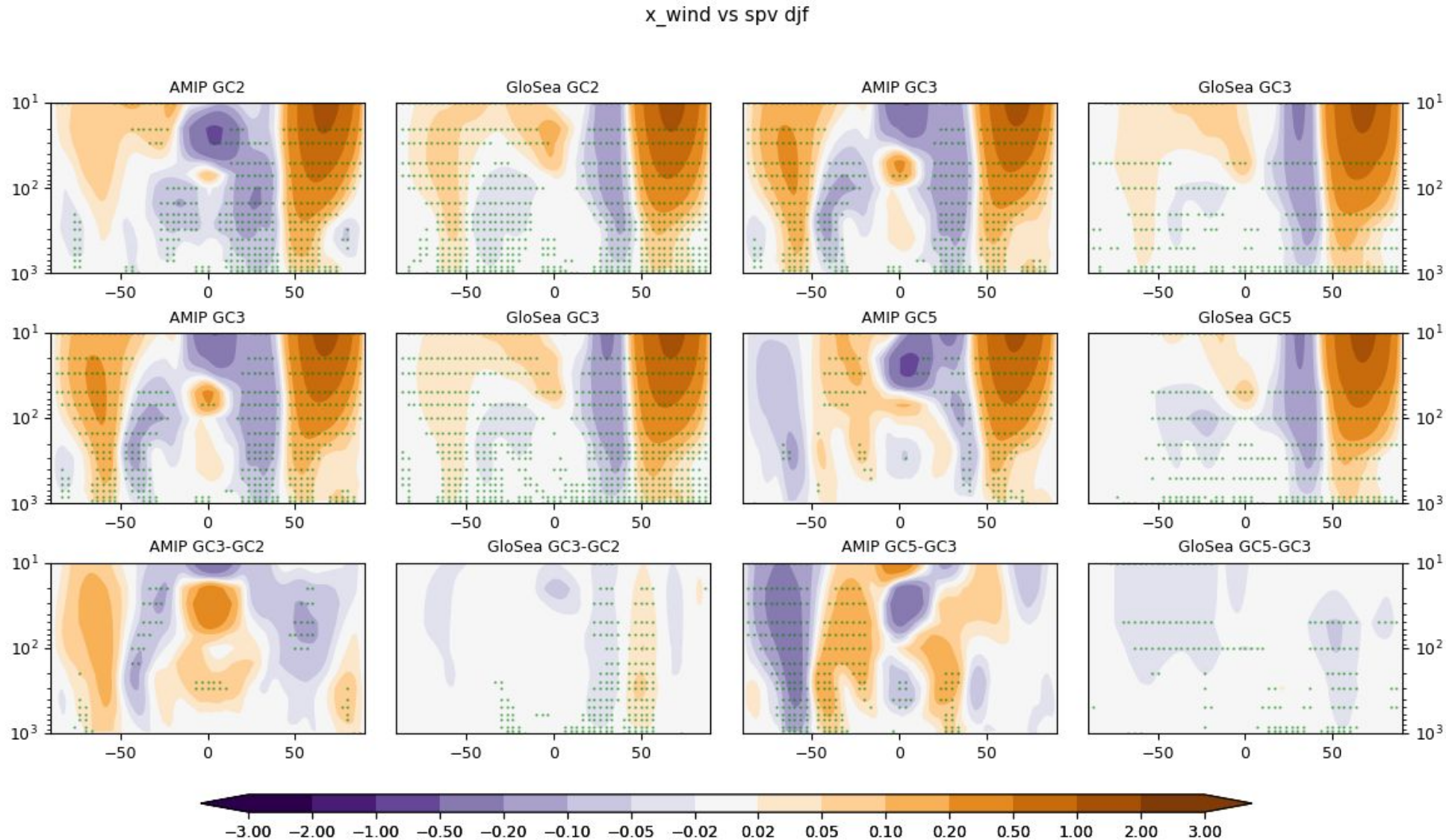
Comparison with 1 member (MSLP JJA mean)



Changes in teleconnections (DJF MSLP mean vs Niño 3.4)



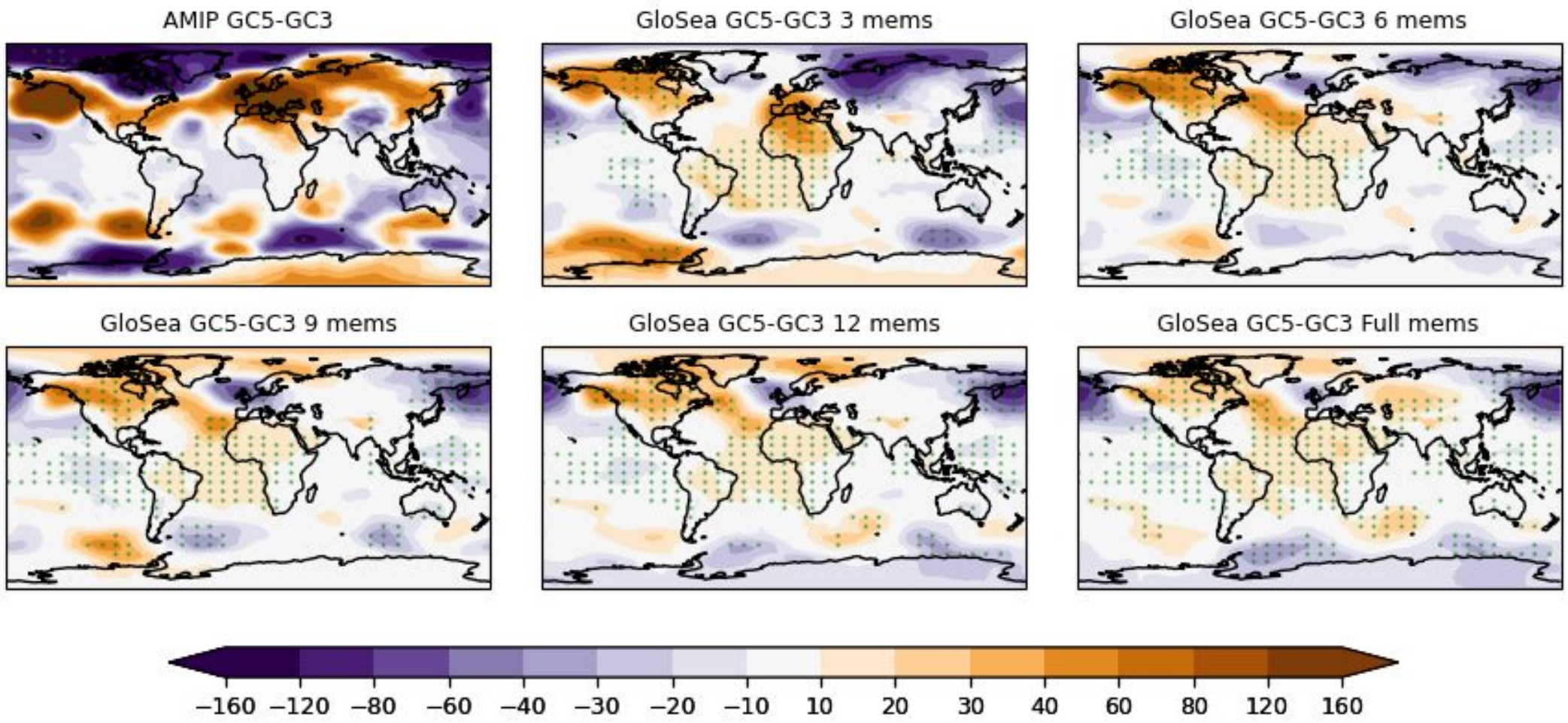
DJF $\bar{u}$ mean vs NH Strat Polar Vortex strength)



Conclusions

- Differences in mean bias fields are frequently similar in AMIP and GloSea and are widely statistically significant in each
- Broadly speaking, therefore, the change in AMIP bias gives an approximate indication of the patterns and amplitudes of the change in GloSea
- For variance and teleconnections, however, AMIP differences between versions generally bear little resemblance to those in GloSea, are much larger in magnitude, and are not significant
- The AMIP runs cannot tell us the expected changes in GloSea because the estimates of these quantities are under-sampled
- To go further, we need more data – we will try running an ensemble of AMIP runs
- If this works, it could provide better feedback for model developers and better models for seasonal forecast groups

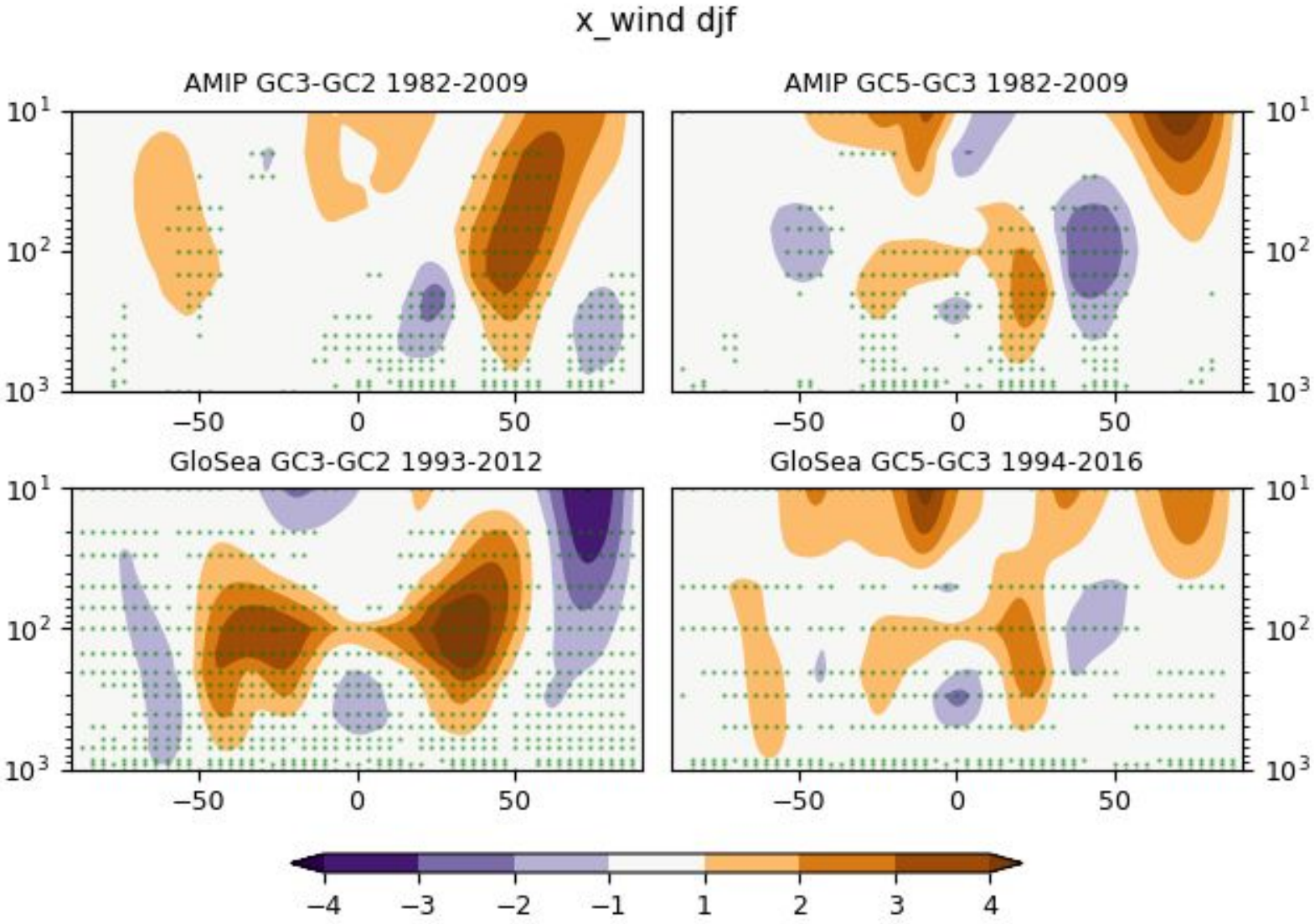
If we had more years by running an ensemble, how big should it be?



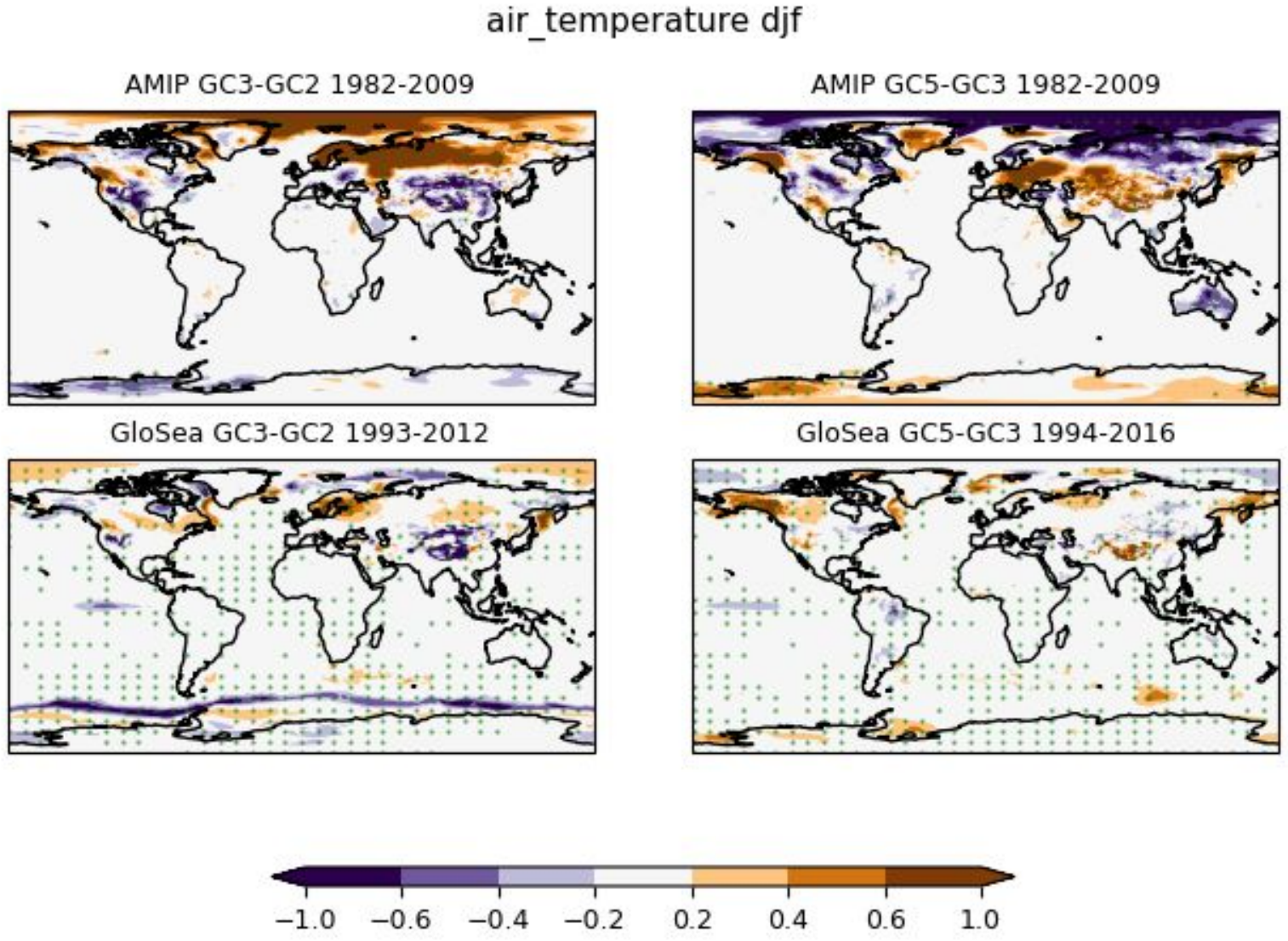
Change in DJF MSLP telecon to Nino3.4 (Pa/K)

Questions and Answers

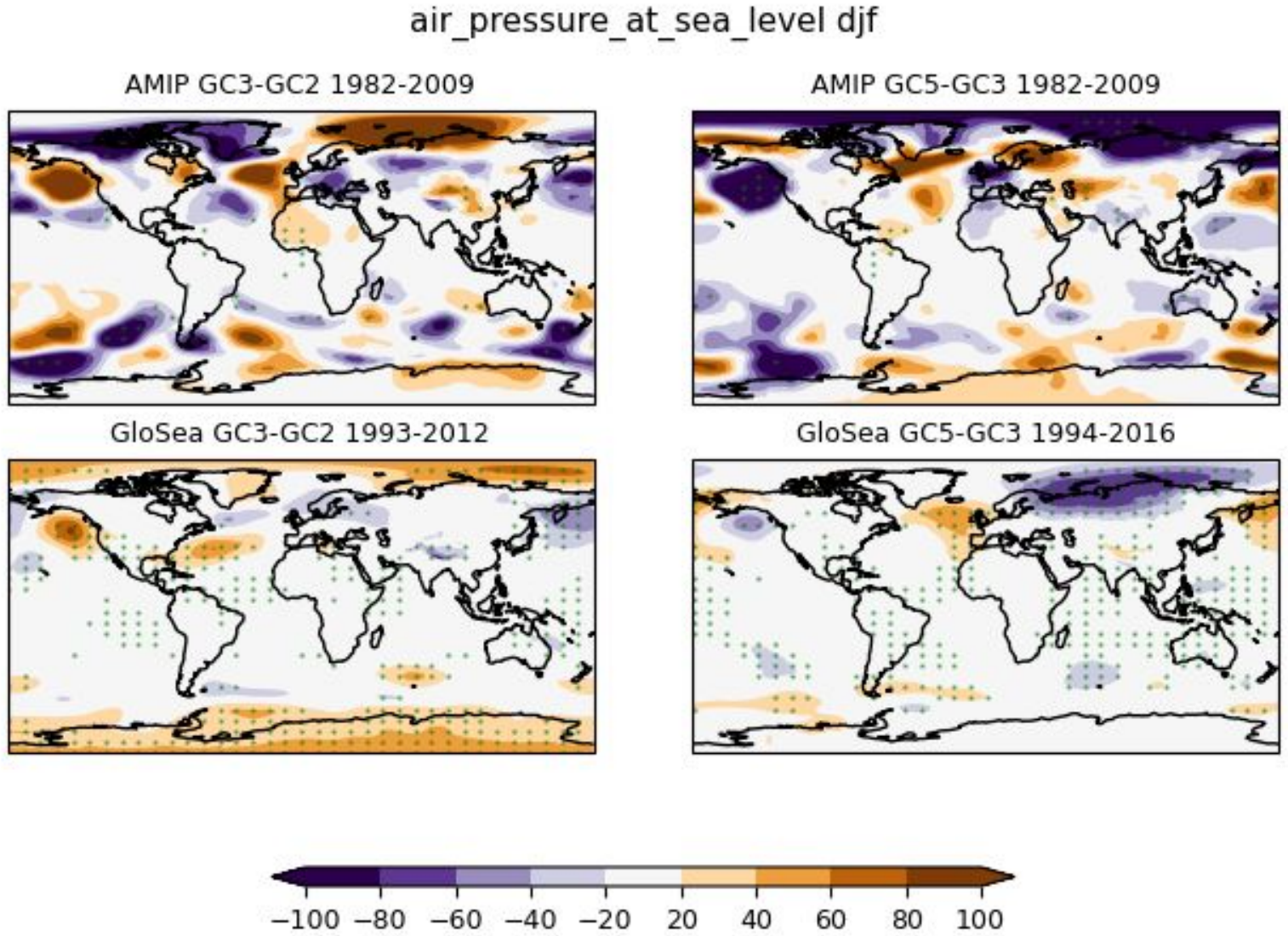
GC Assessment vs GloSea clim (ubar DJF)



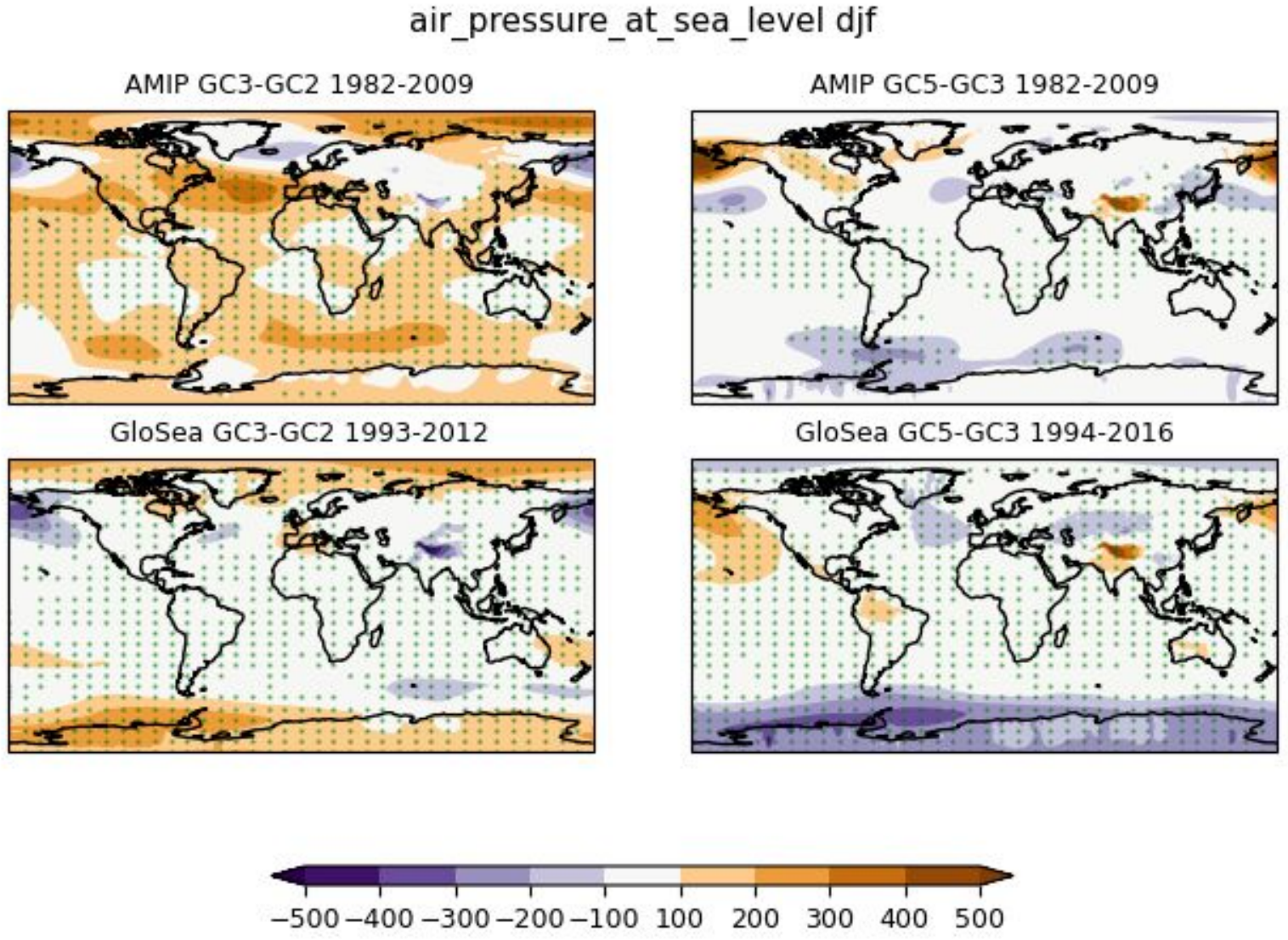
Changes in standard deviation (1.5mT DJF)



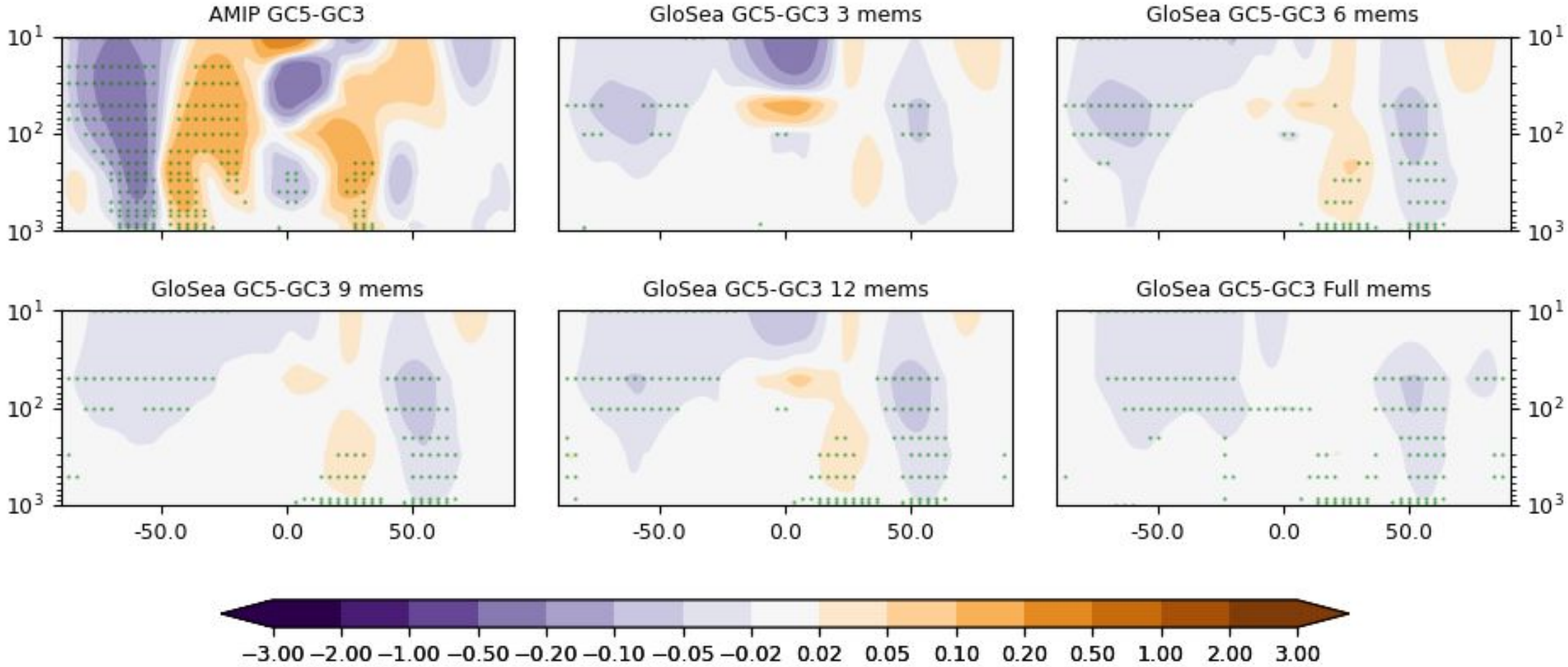
Changes in standard deviation (MSLP DJF)



Comparison of change in mean bias (MSLP DJF mean)



If we had more years by running an ensemble, how big should it be?



Change in DJF zm u telecon to SPV (m/s per m/s)

Conclusions from variance/teleconnection plots

- AMIP differences between versions generally bear little resemblance to those in GloSea
- Features in the AMIP differences tend to be much larger
- AMIP differences generally not significant, whereas GloSea differences are
- Suggests AMIP has too much noise due to undersampling, i.e. an insufficient number of years
- Confirmed by single members from GloSea, which have similar characteristics
- Higher requirement for data for these quantities than for means

Recommendations

- Use AMIP to estimate the changes in GloSea hindcasts, thereby allowing quicker input to development
- Implement an ensemble to widen the range of variables to include more that are of interest to S2D
- The amount of data necessary appears to be ~250 years, or around 8-10 parallel members
- Extend AMIP runs to ~2020 to better match the GloSea hindcast period and provide more data
- Perform statistical testing routinely, especially for these additional metrics
- Keep doing final GloSea hindcast and assessment, as we still need to exactly know the ultimate level of performance