Accelerating Progress in Subseasonal to Seasonal Prediction Capabilities by Improving Subgrid-Scale Parameterizations in the Unified Forecast System

Lead PI: Benjamin W. Green[1,2]

Co-PI: Vijay Tallapragada[3]

Collaborators: Shan Sun[1], Eric Sinsky[3,4], Georg Grell[1], Fanglin Yang[3], Jian-Wen Bao[5]

[1]NOAA/OAR/GSL; [2]CU/CIRES; [3]NOAA/NWS/EMC; [4]IMSG; [5]NOAA/OAR/PSL

WPO Award NA20OAR4590314

Presentation to Weeks 3-4/S2S Webinar, 5 April 2021

Background

- Most model physics development takes place for systems run on timescales of global weather (< 2 weeks), or even shorter: not as much at S2S+ timescales
  - Why? Mostly practical: shorter timescales = more/faster runs

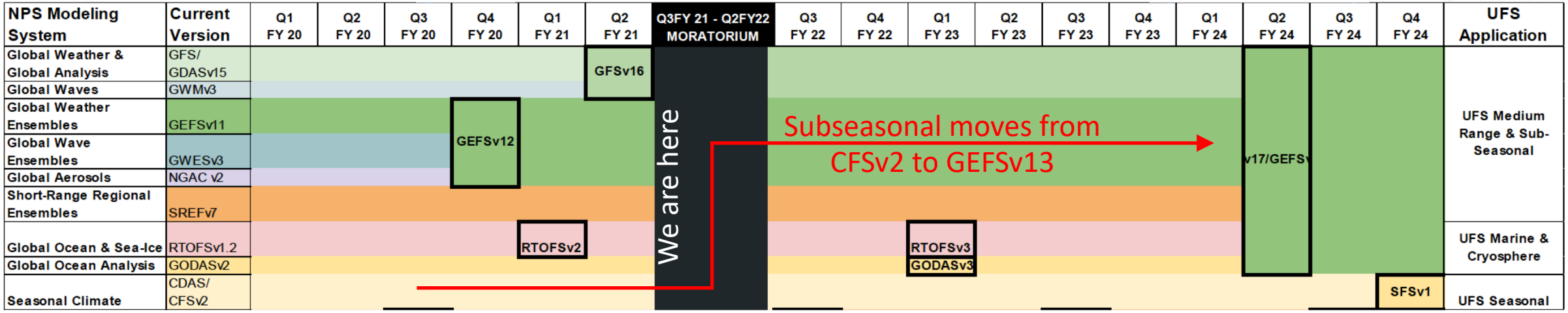
- Physics development for S2S out to climate scales evolves much more slowly, and separately from NWP (weather) timescale

- Paradigm shift (at least at operational centers): consolidate modeling systems (dynamical cores, and subgrid-scale physics) to run across many timescales – “minutes-to-months” (next presenter uses phrase “minutes-to-millenia”)

- Leverage paradigm shift to look at model physics at subseasonal timescales

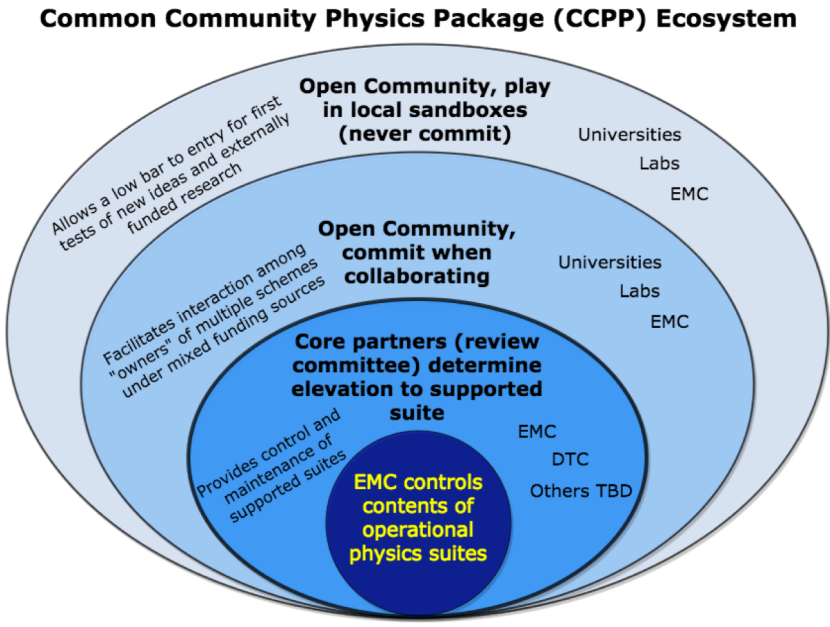
Paradigm shift at NOAA: UFS and CCPP

- NOAA is consolidating its modeling systems into Unified Forecast System (UFS)
- Current operations:
  - Runs out to 16 days done by **atmosphere+wave GFS/GEFS**
  - Runs out to 45 days (and beyond) done by **fully-coupled CFSv2**
  - GFS/GEFS have very different physics schemes than CFSv2
- By 2024, all of NOAA's operational **global** Earth system prediction for lead times from **1 to ~45 days** will be consolidated into a **single UFS-based model** (GEFS)
  - New challenge: Ensure model has "reasonable" skill from daily to subseasonal time scales



Paradigm shift at NOAA: UFS and CCPP

- CCPP: Community Common Physics Package
- Strips (atmospheric) physics out of dynamical core; allows for easy replacement of physics *schemes* and entire physics *suites*
- Common framework used by operations (NOAA/EMC) and research partners
- This project: Leverage CCPP to swap in alternative parameterizations for convection, cloud microphysics, and planetary boundary layer for subseasonal runs



Credit: DTC website

Project goals

- Use “one-at-a-time” tests that swap parameterizations of convection, microphysics, and PBL to examine impact of these schemes on coupled UFS subseasonal runs
- Leverage ongoing coupled UFS development at EMC (they run “Experiment 1”):

Experiment #	Experiment Name	Convection	Boundary Layer	Microphysics
1	UFS_P5	SASAS	EDMF	GFDL
2	GF	**GF**	EDMF	GFDL
3	MYNN	SASAS	**MYNN**	GFDL
4	Thompson	SASAS	EDMF	**Thompson**

- GF, MYNN, and Thompson schemes are developed by NOAA/GSL, NCAR, and other partners primarily for use in high-resolution short-range NWP
- Comparing Experiment 1 to 2, 3, or 4 gives insight into impacts of convection/PBL/microphysics, **accelerating S2S physics development**

Experimental design

- Run 3 additional sets of experiments (see previous slide). Follow EMC's "Prototype 5" protocol for each experiment:
  - Initialize 1st and 15th of every month from 1 April 2011 through 15 March 2018 (**168 cases**)
  - C384 (~25 km) resolution, 64 vertical layers: daily 1x1 output on isobaric & surface levels
  - 35-day runs
  - CMEPS mediator used to couple the following models:
    - FV3 atmosphere
    - MOM6 ocean
    - CICE6 sea ice
    - WW3 wave
- Some notes:
  - There were a total of 3 crashed cases (1/168 in MYNN, 2/168 in Thompson) due to ice model issues. We can get statistically meaningful results without these cases
  - Future iteration of Experiments 2-4 will follow protocol for a later EMC test of coupled UFS

What do we want to investigate?

- Biases
  - T2m and precipitation are meaningful for end-users of S2S products
  - Global circulation biases are important too, due to teleconnections on S2S timescales

- Deterministic skill scores: Bivariate RMM index for MJO skill

- Vertical profiles (case studies and aggregate): Ongoing work

- Multi-physics ensembles: More of a research question, but expect forecast **skill** could be improved by combining Experiments 1-4 into an ensemble

Preliminary results

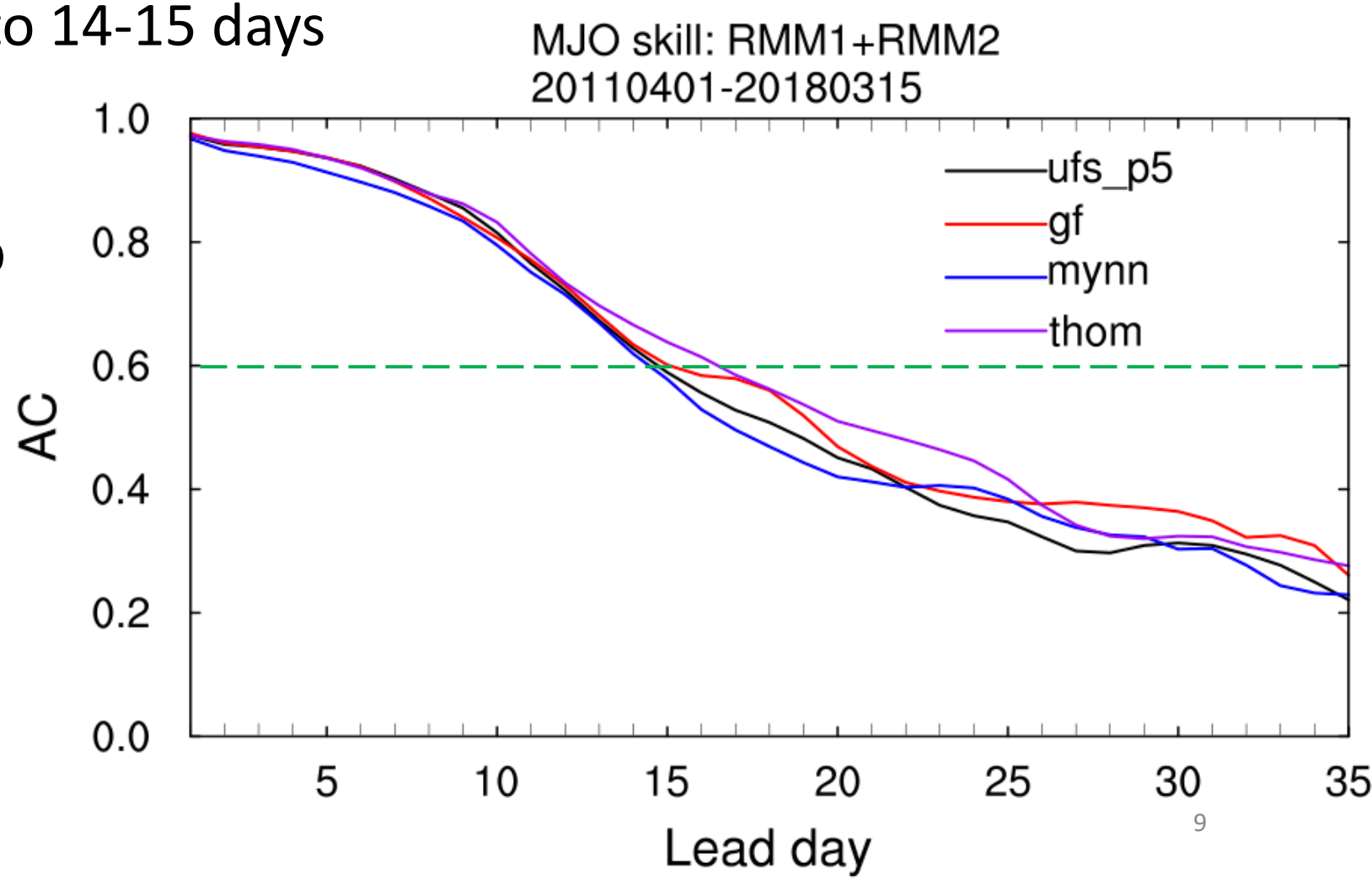
- Note: **Project is in early stages, so results still coming in**

- Skill scores: Bivariate RMM

- Biases: T2m, SST, precipitation, OLR (circulation is still a work in progress)

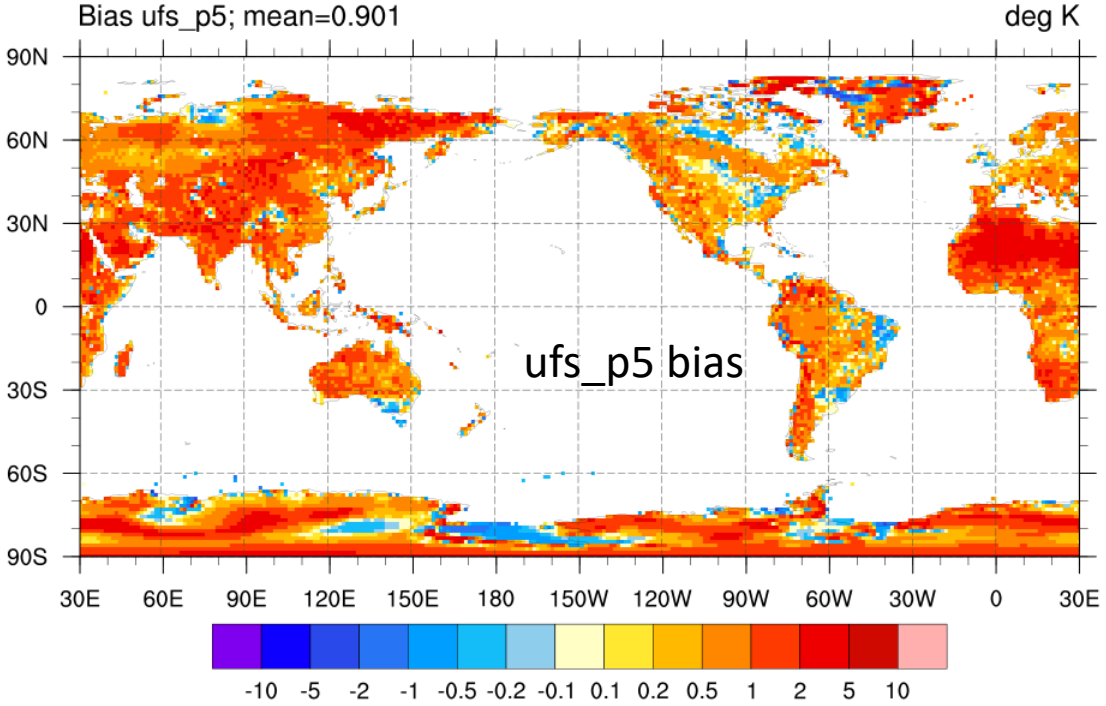
Preliminary results: RMM skill score

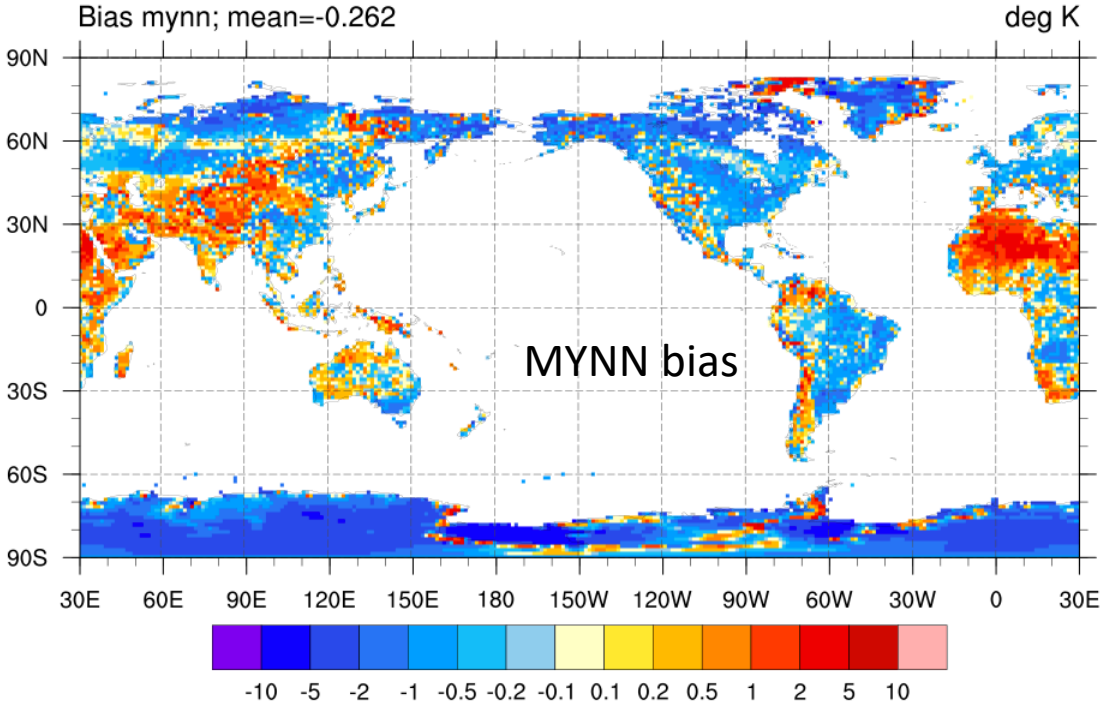
- Control ("ufs_p5") never has the highest RMM skill score
- Using score of 0.6 as a threshold, Thompson experiment is skillful out to 16 days (**year round**); others skillful to 14-15 days
- Still need to look at various teleconnections: RMM skill is meaningless if relationship between tropics and mid-latitudes is wrong

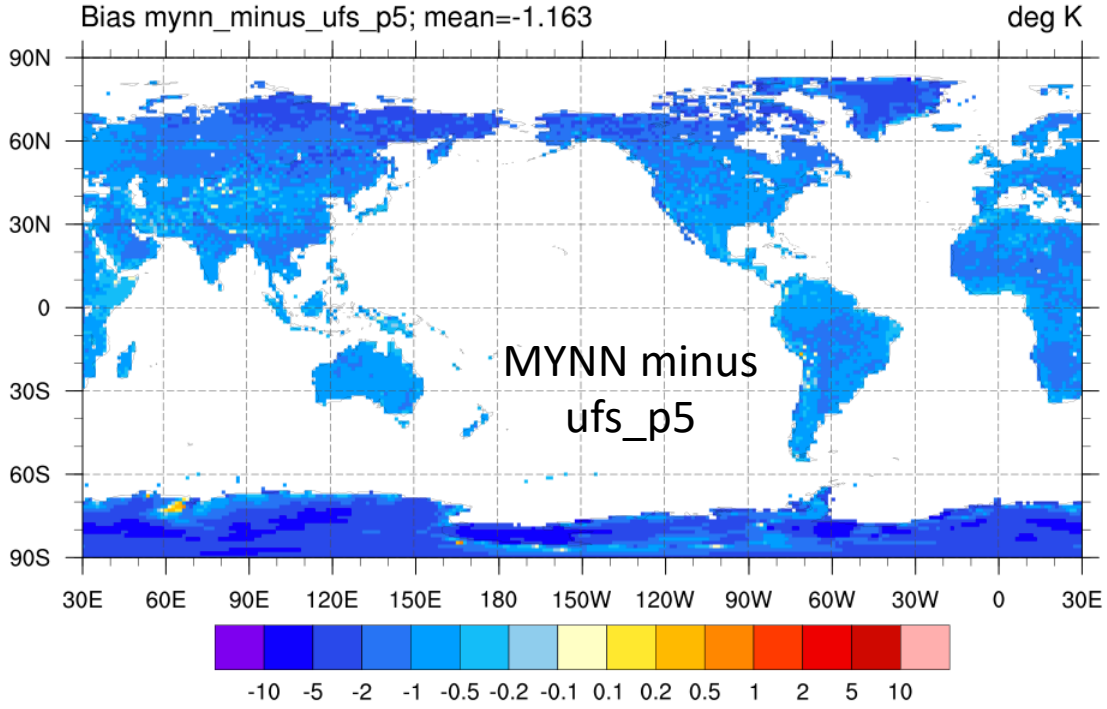


Preliminary results: T2m bias

- Land biases (left) using CFSR as truth
- Week **1**, ufs_p5 vs. MYNN: 167 cases
- Systematic cooling in MYNN relative to ufs_p5 (bottom right)

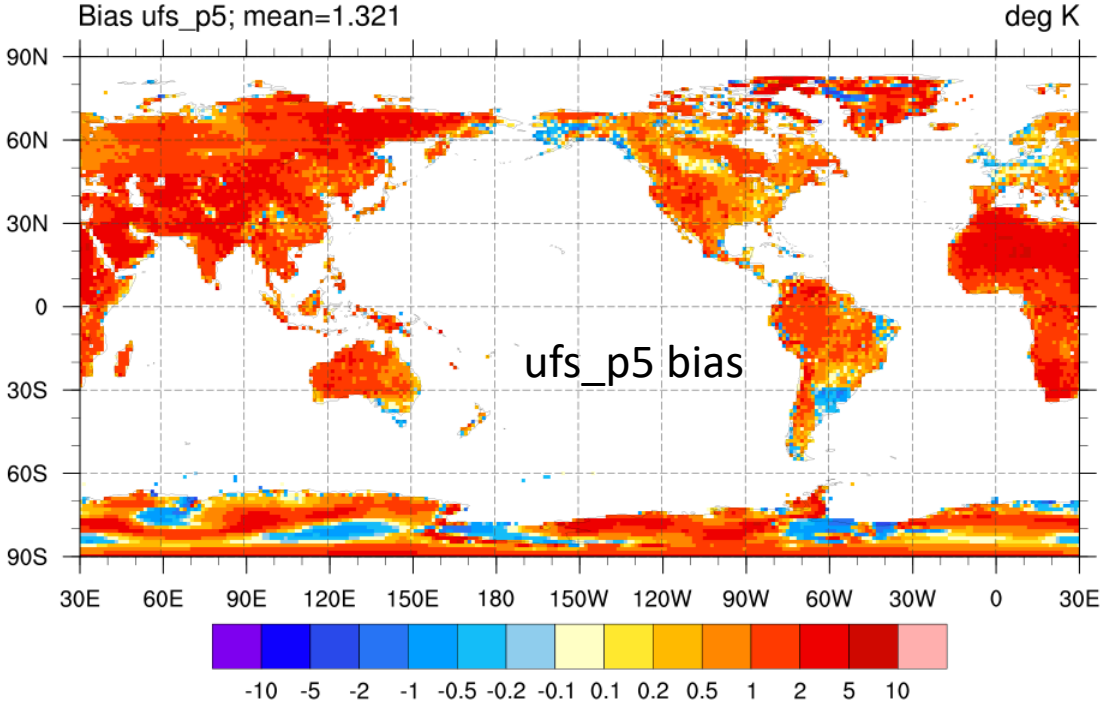


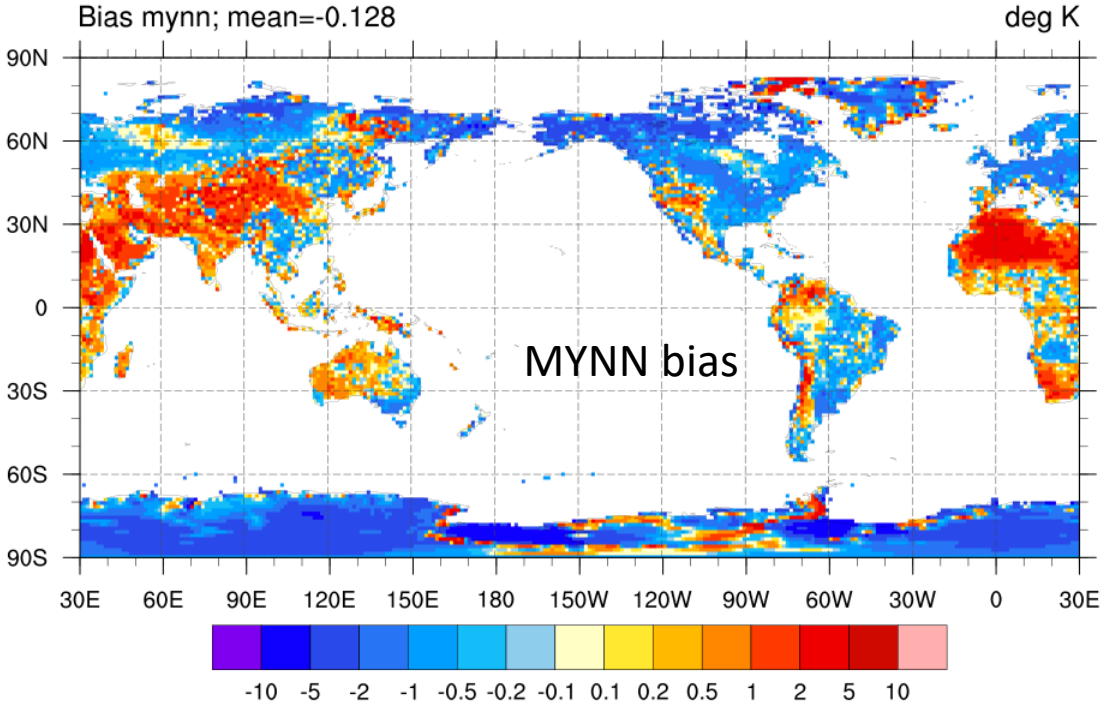


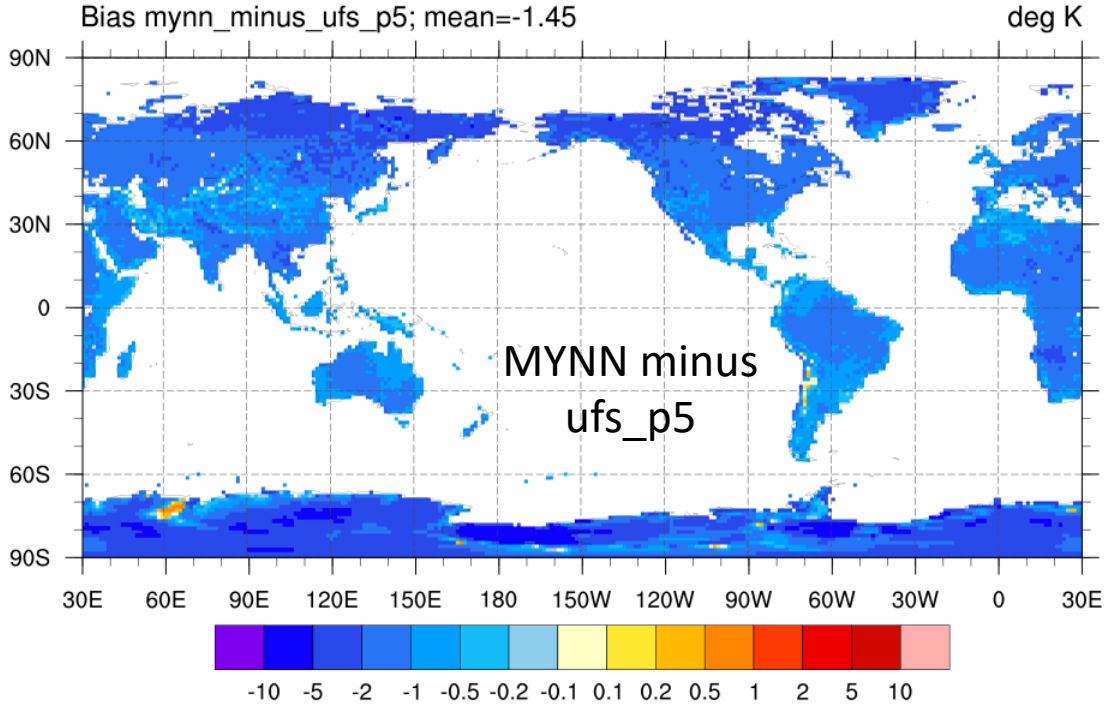


Preliminary results: T2m bias

- Land biases (left) using CFSR as truth
- Weeks **3-4**, ufs_p5 vs. MYNN: 167 cases
- Patterns extremely similar to week 1!
- Systematic cooling in MYNN relative to ufs_p5 (bottom right)

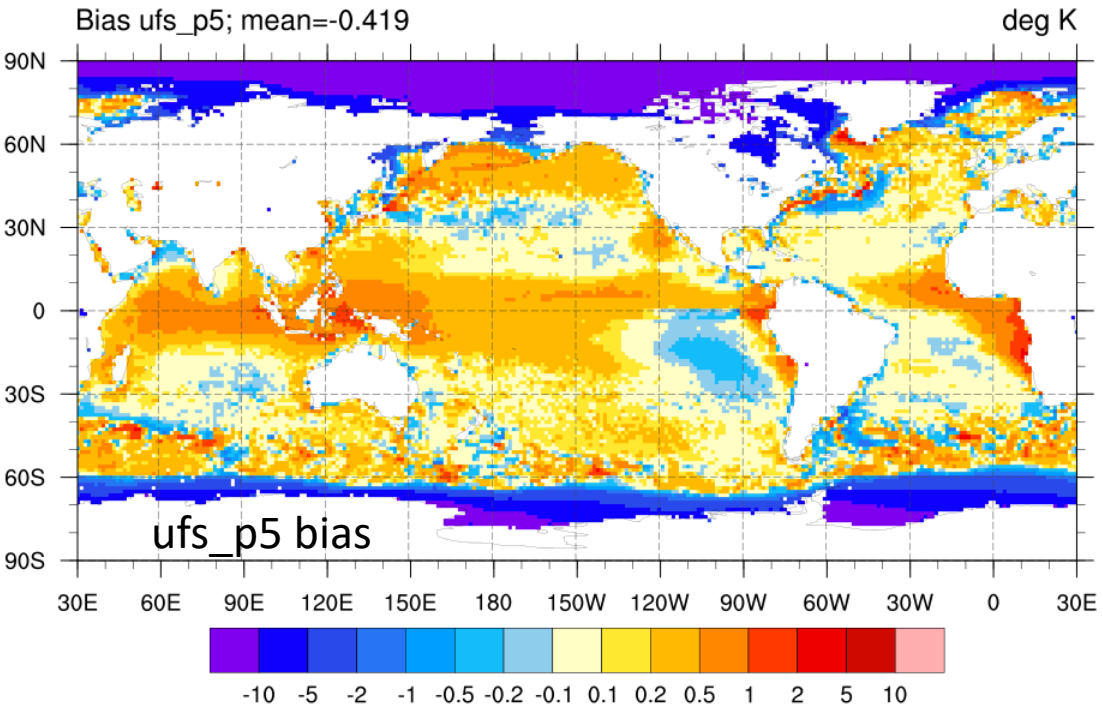


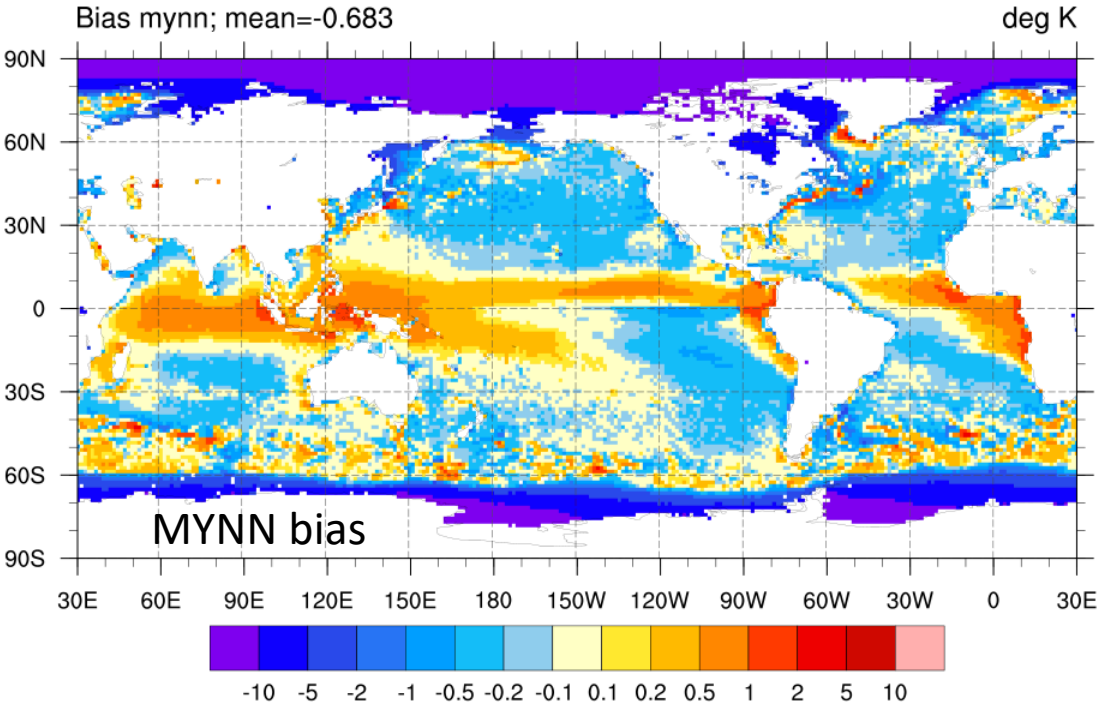


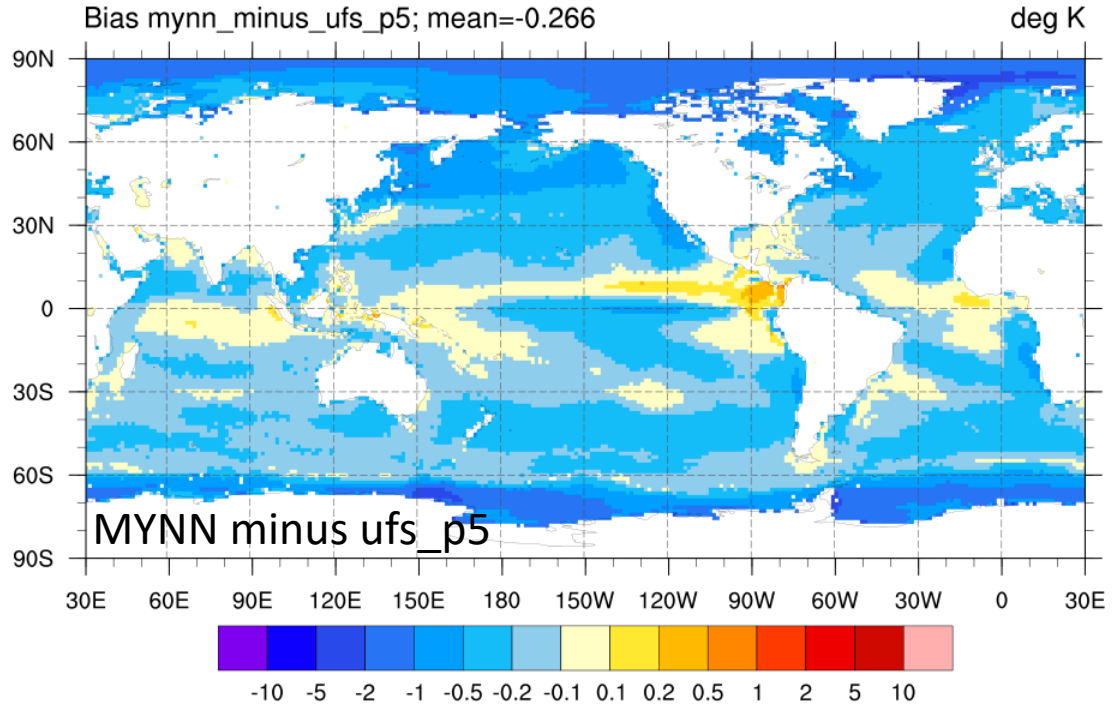


Preliminary results: SST bias

- Using OSTIA as truth
- Weeks **3-4**, ufs_p5 vs. MYNN: 167 cases
- MYNN warmer than ufs_p5 along ITCZ; cooler everywhere else (recall T2m cooler over all land)





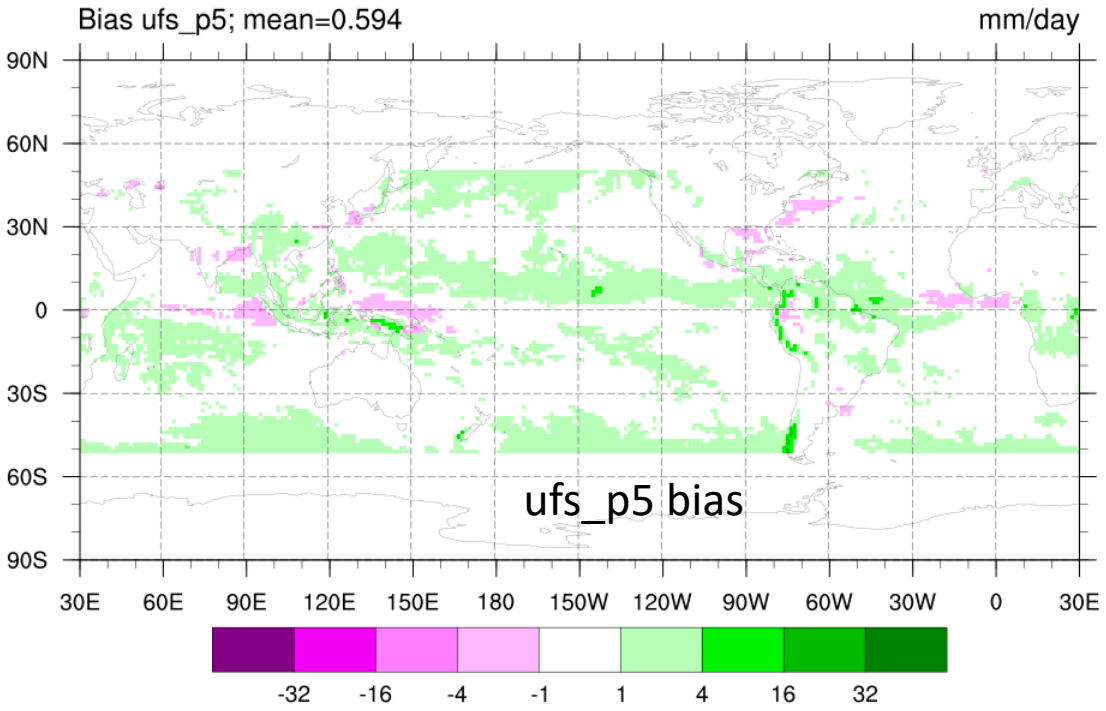


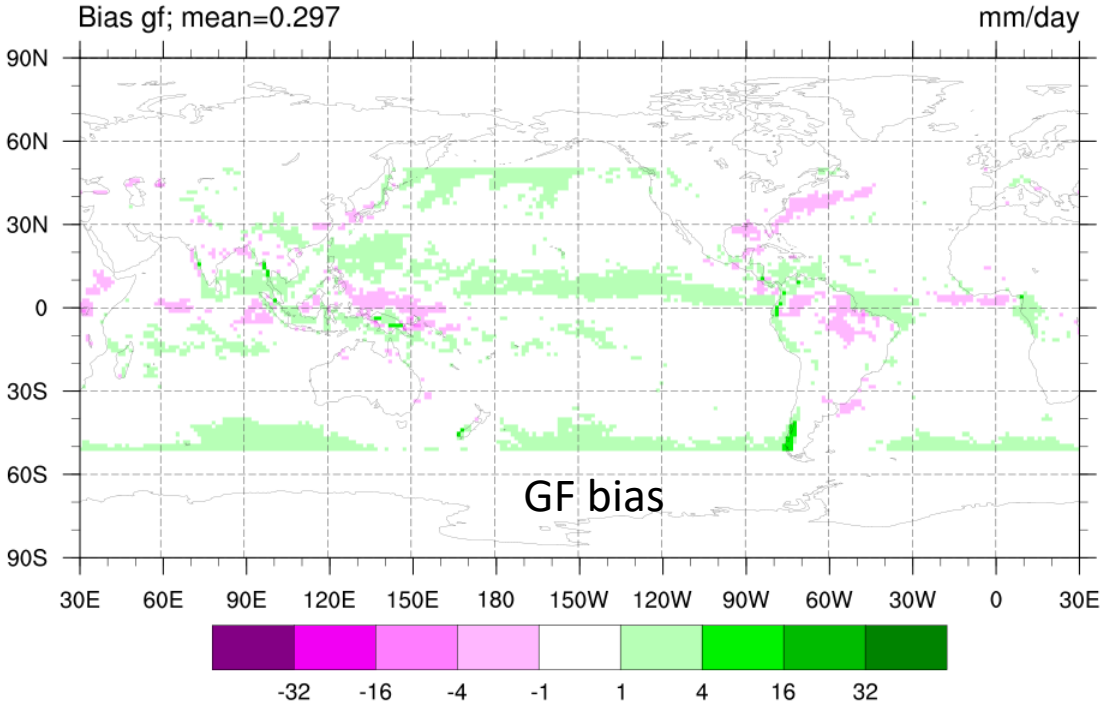
Preliminary results: T2m bias **over land**, and SST

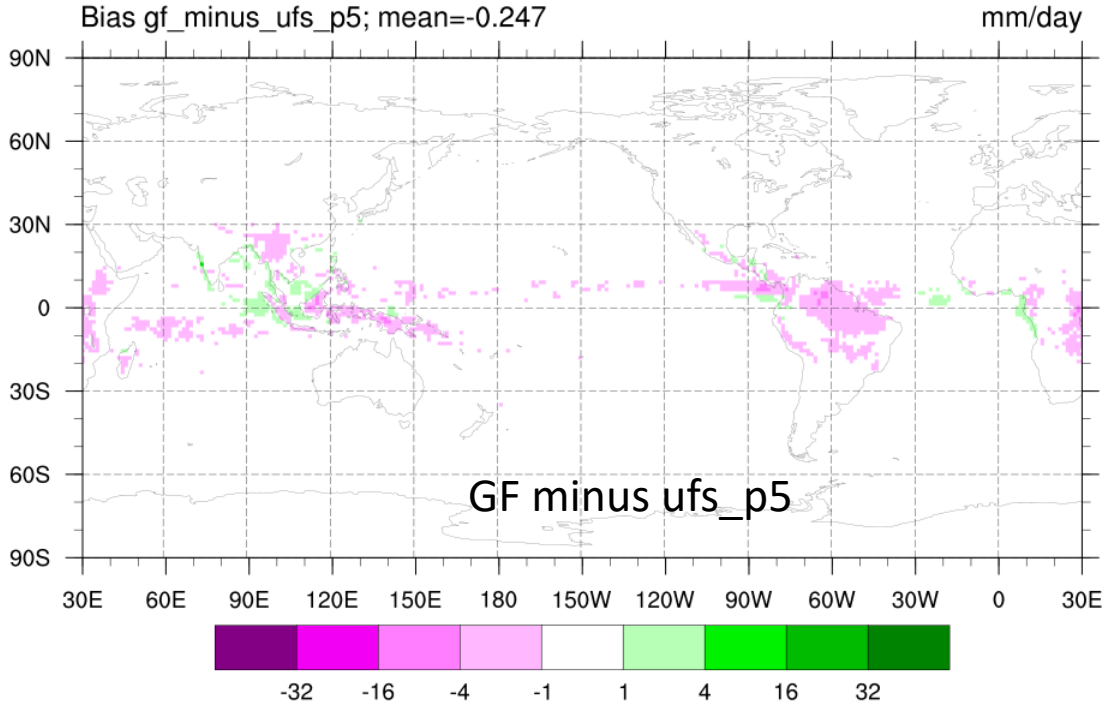
- Comparing ufs_p5 (control) vs. MYNNPBL: MYNN consistently colder
  - True across almost all land masses
  - True across seasons (not shown)
  - True over all lead times (only week 1 vs. weeks 3-4 shown)
- Bias patterns from week 1 to weeks 3-4 are quite similar, just growing in magnitude with increasing lead time
- SST biases for MYNN grow more slowly than T2m, but end up colder than ufs_p5 everywhere (except for ITCZ)
- **A thought:** Biases from "first-order processes" (e.g., PBL influencing T2m) at S2S timescales may be reduced by tuning based on shorter runs. In other words, if we want to reduce T2m biases for S2S, we may be able to tune/test modifications to PBL scheme with 7-day (or shorter) runs
  - But this potential shortcut won't work for "second-order processes" (e.g., convection influencing T2m; PBL influencing OLR…)

Preliminary results: QPF bias

- Using TRMM as truth
- Week **1**, ufs_p5 vs. GF: 168 cases
- GF has smaller mean bias than ufs_p5
- Systematic drying in GF relative to ufs_p5 (bottom right)

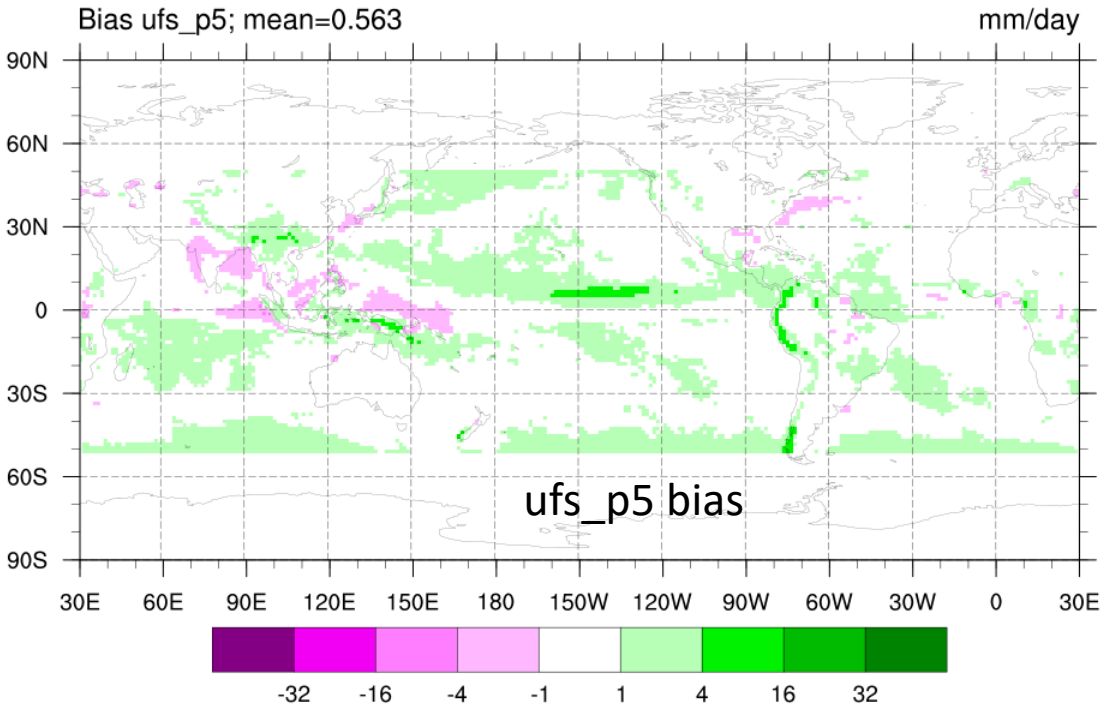


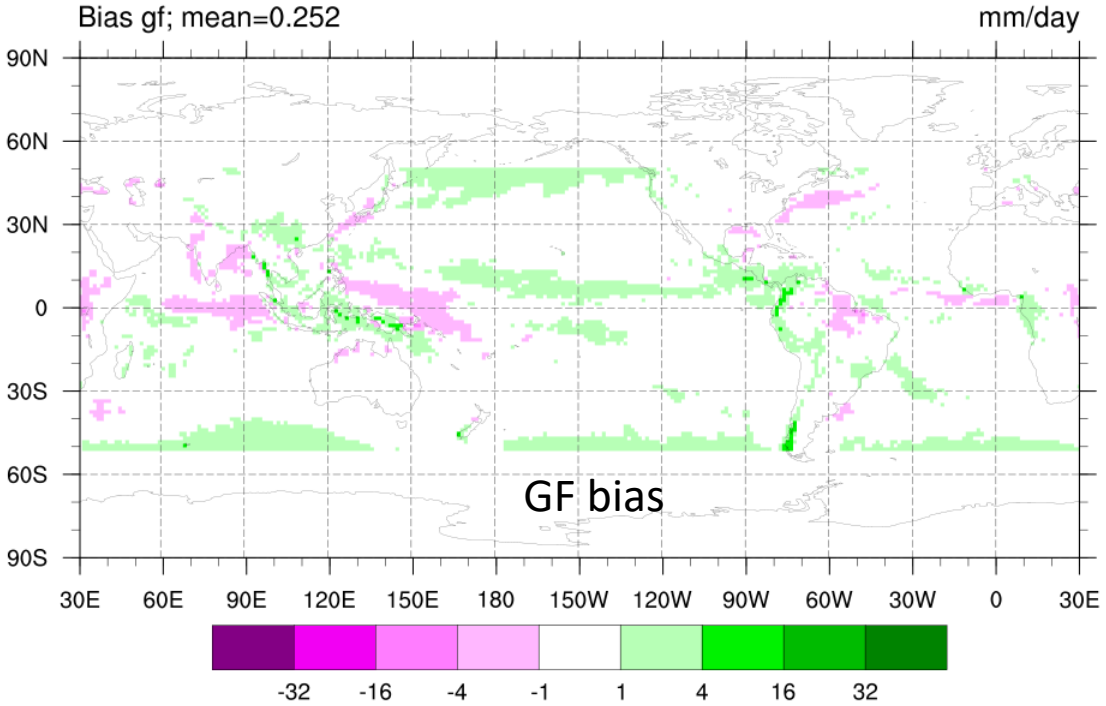


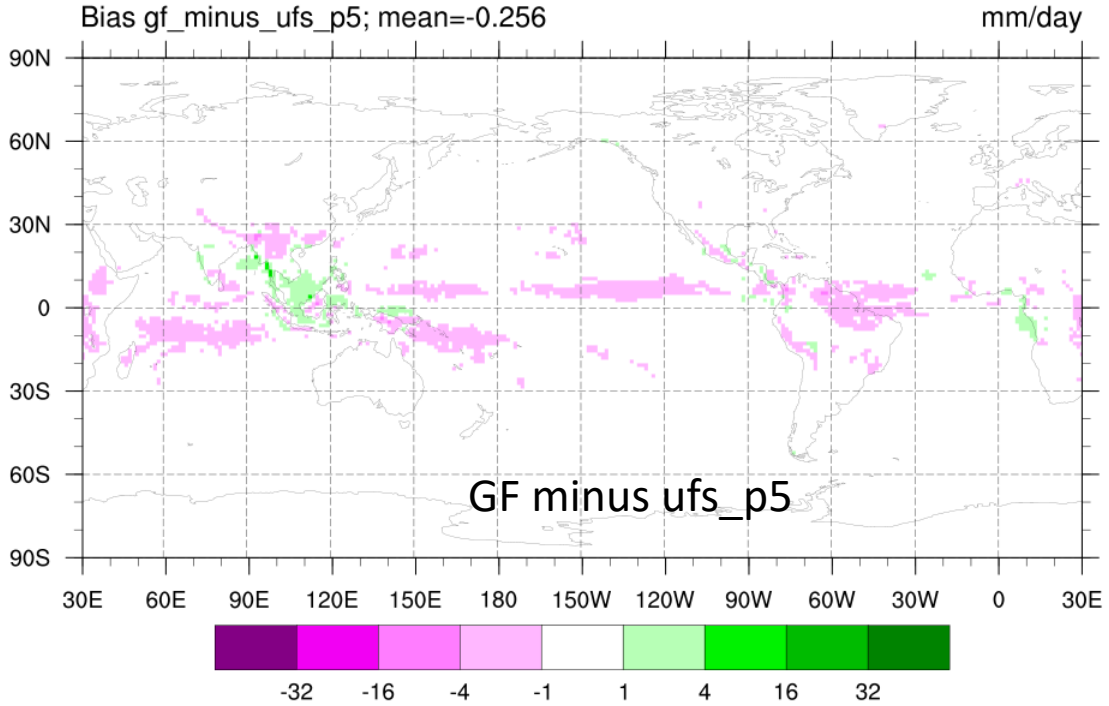


Preliminary results: QPF bias

- Using TRMM as truth
- Weeks **3-4**, ufs_p5 vs. GF: 168 cases
- GF has smaller mean bias than ufs_p5
- Systematic drying in GF relative to ufs_p5 (bottom right)

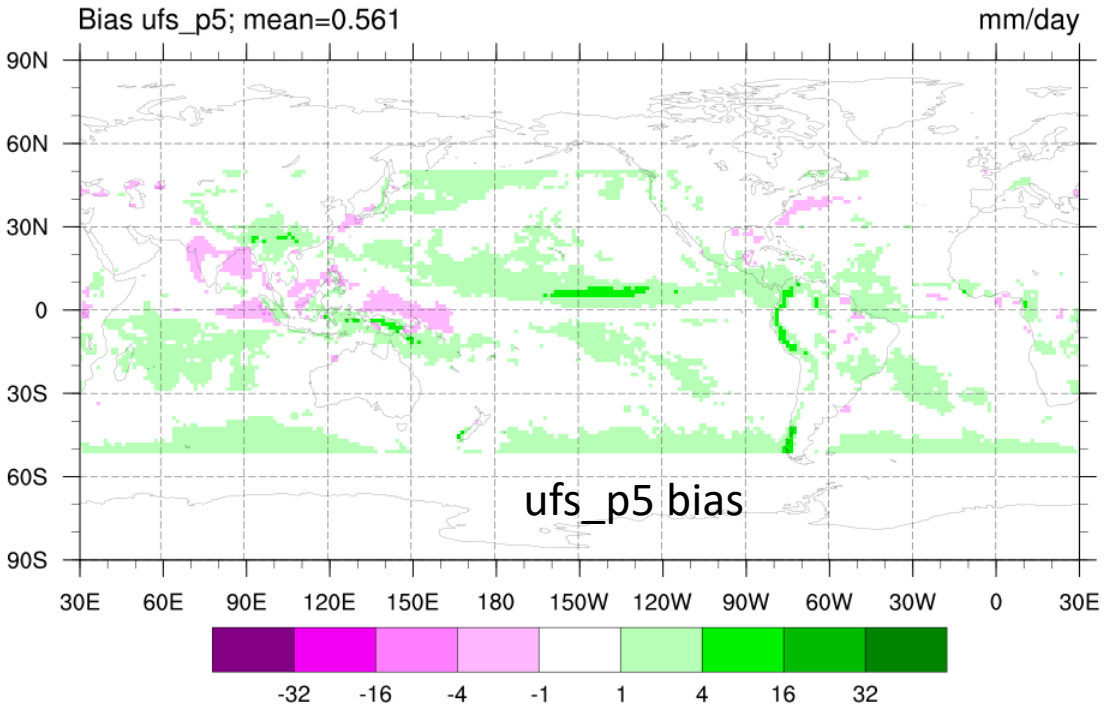


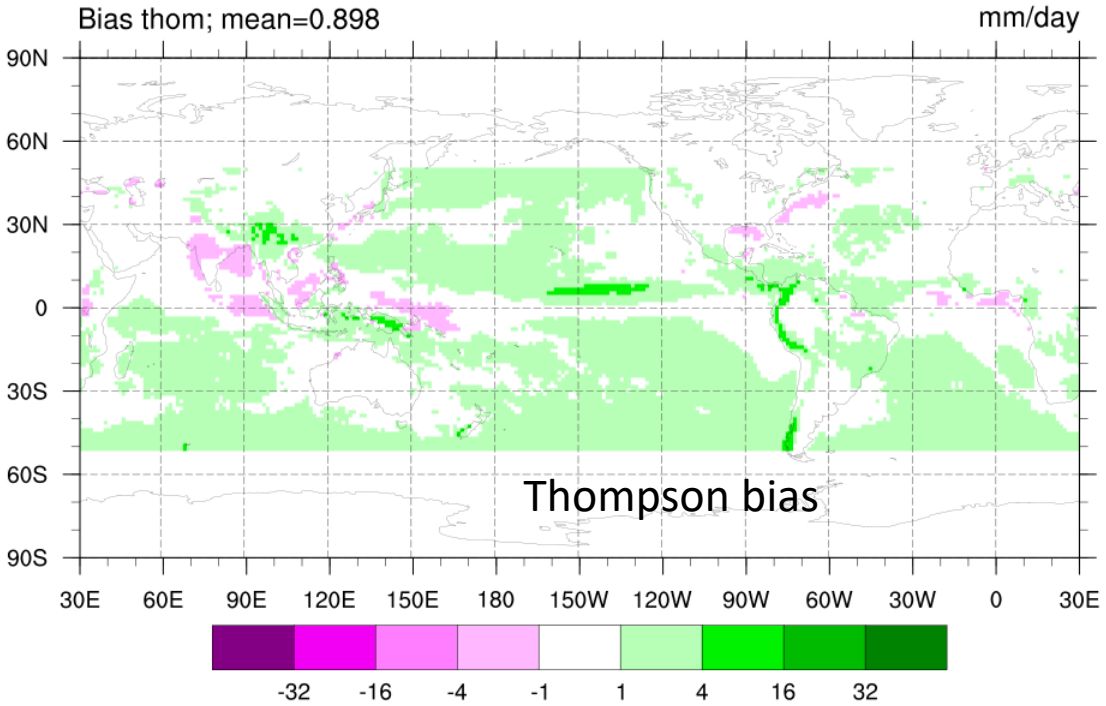


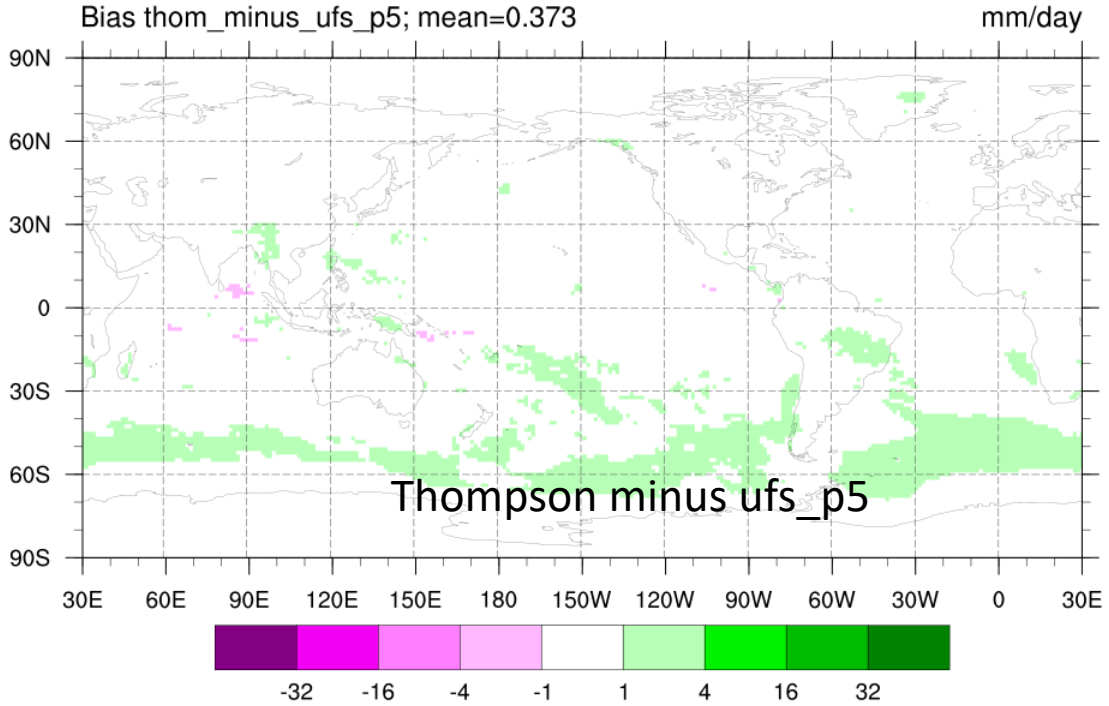


Preliminary results: QPF bias

- Using TRMM as truth
- Weeks **3-4**, ufs_p5 vs. **Thompson**: 166 cases
- Thompson has biggest impact in extratropics: **tropics looks almost identical**
  - Convective scheme much more important here

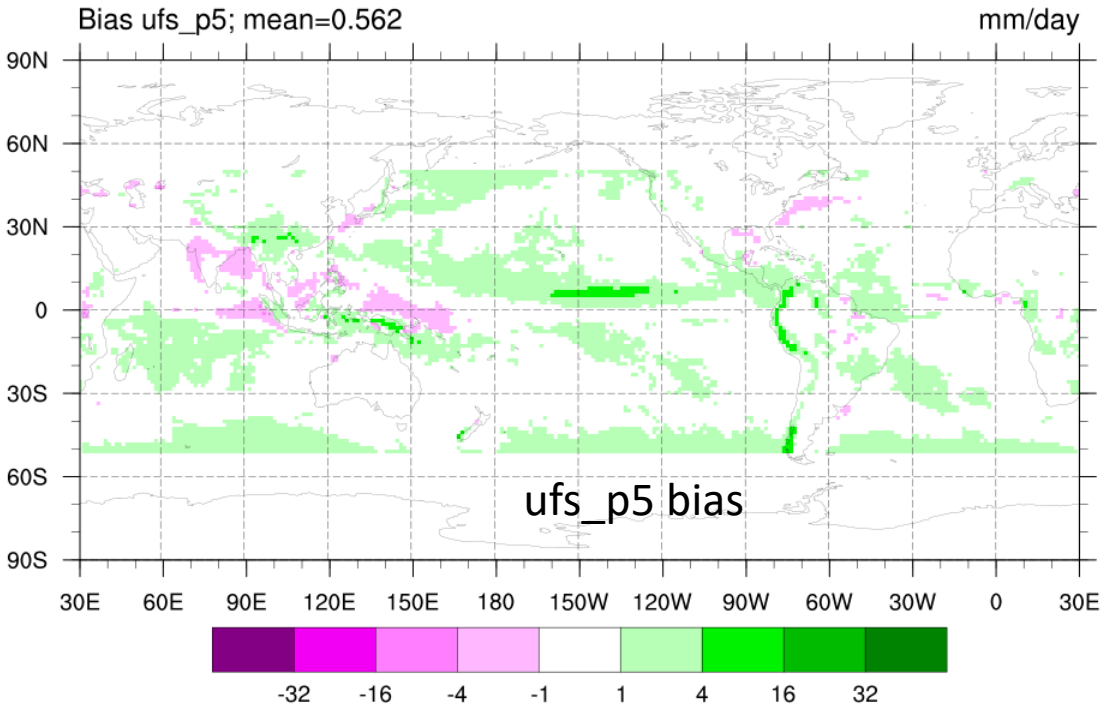


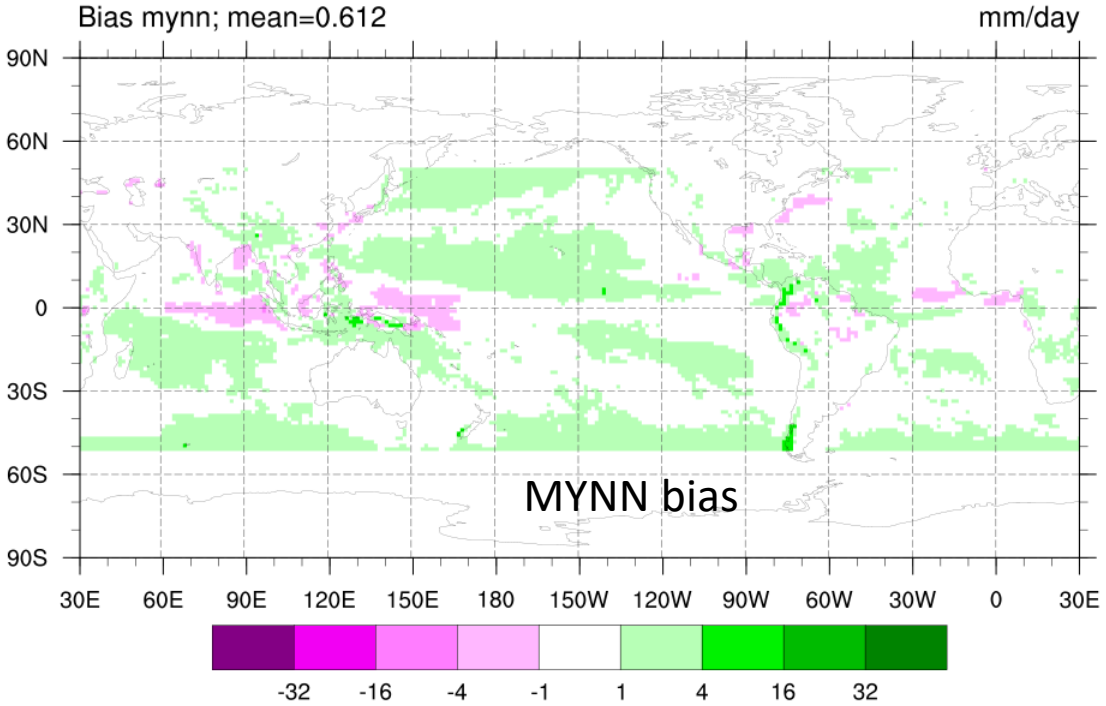


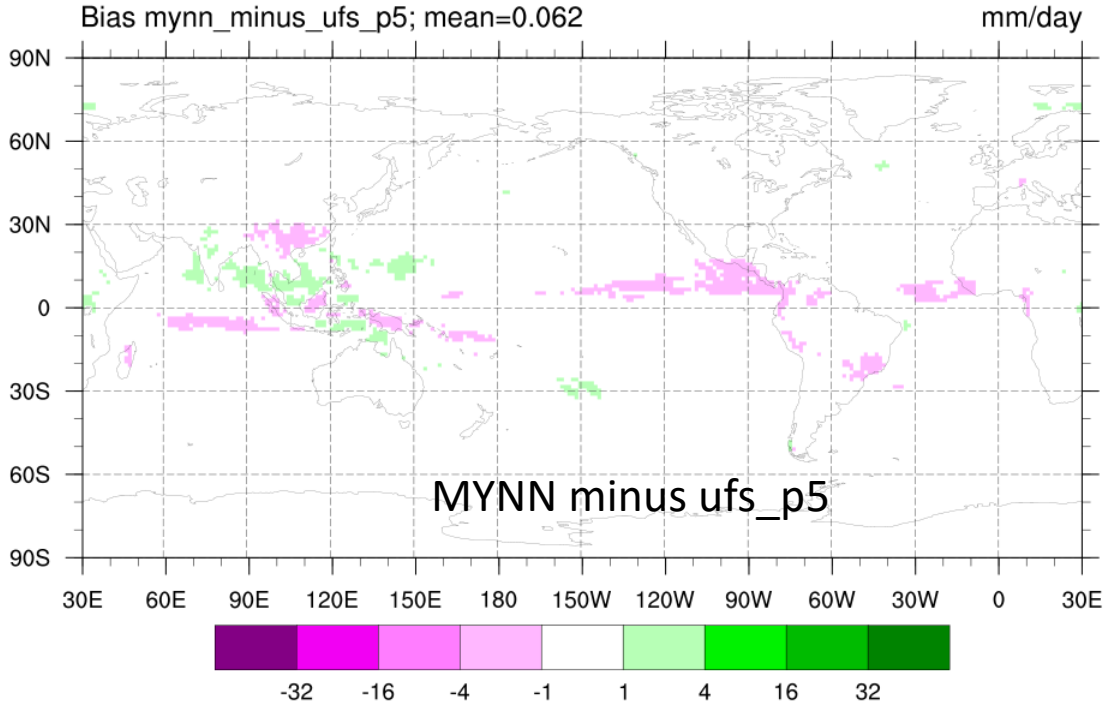


Preliminary results: QPF bias

- Using TRMM as truth
- Weeks **3-4**, ufs_p5 vs. **MYNN**: 167 cases
- MYNN has biggest impact in **tropics**: likely due to PBL transport of moisture from ocean (would require more investigation)







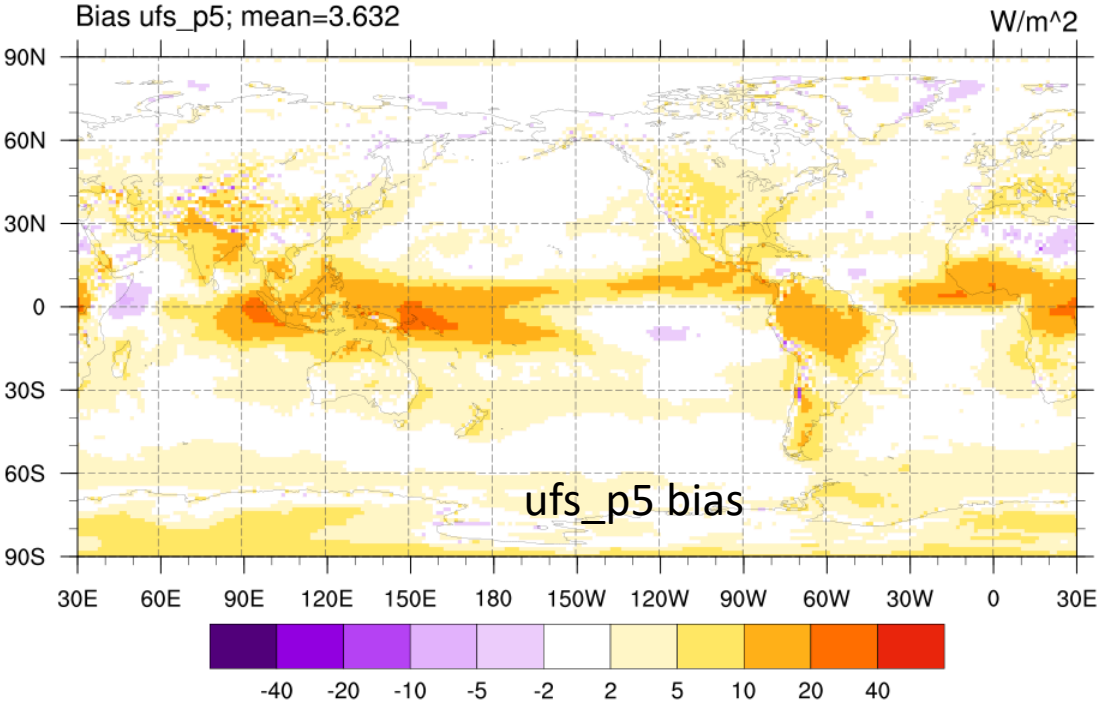
Preliminary results: Precipitation bias

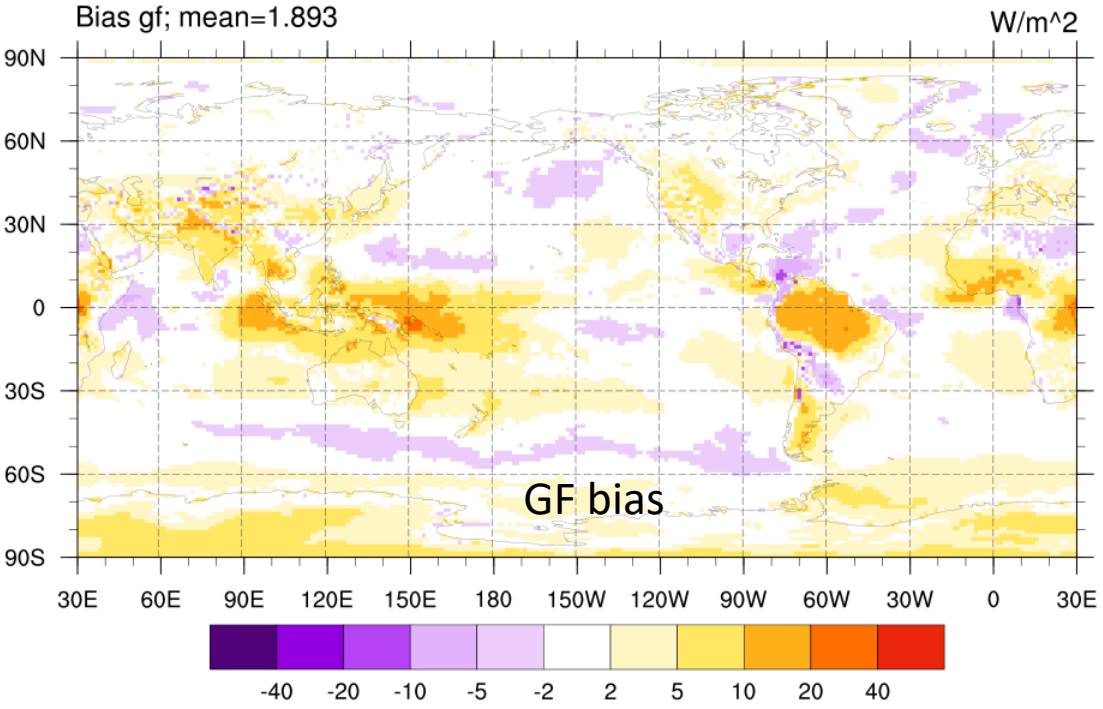
- Comparing ufs_p5 (control) vs. GF: GF has smaller bias than ufs_p5
- GF consistently has less precip than ufs_p5
  - True across seasons (not shown)
  - True over all lead times (only week 1 vs. weeks 3-4 shown)
  - Not necessarily true uniformly in space, especially when looking by season (not shown)
  - Bias patterns from week 1 to weeks 3-4 are somewhat similar, including in magnitude: precipitation errors saturate very quickly (not surprising)
  - Moreso because bias magnitudes don't increase from week 1 to weeks 3-4, tuning convective scheme for precip may not require 35-d runs (could get by with shorter runs)

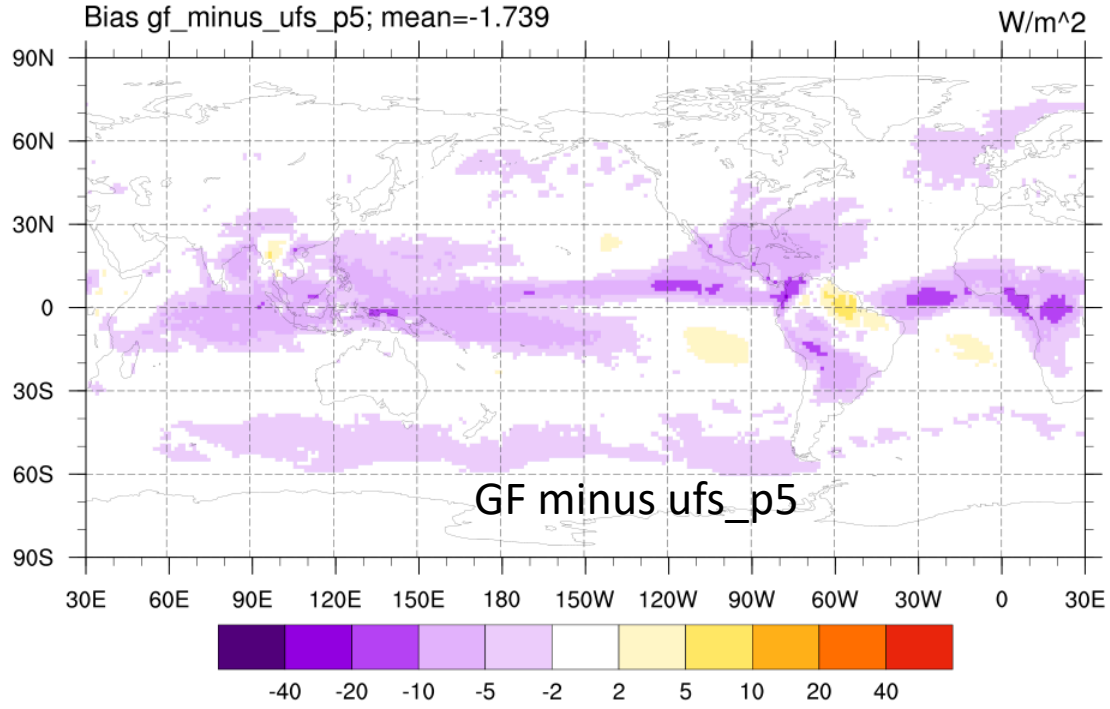
- For microphysics: Impact noted in extratropics; tropics has only minor change because convection scheme is the same
- For PBL: Impact noted in tropics (not extratropics), likely because of different moisture fluxes through PBL

Preliminary results: OLR bias

- Using HRIS as truth
- Week **1**, ufs_p5 vs. GF: 168 cases
- GF has lower mean bias than ufs_p5
- Deeper tropical convection (lower OLR) in GF relative to ufs_p5 (bottom right)

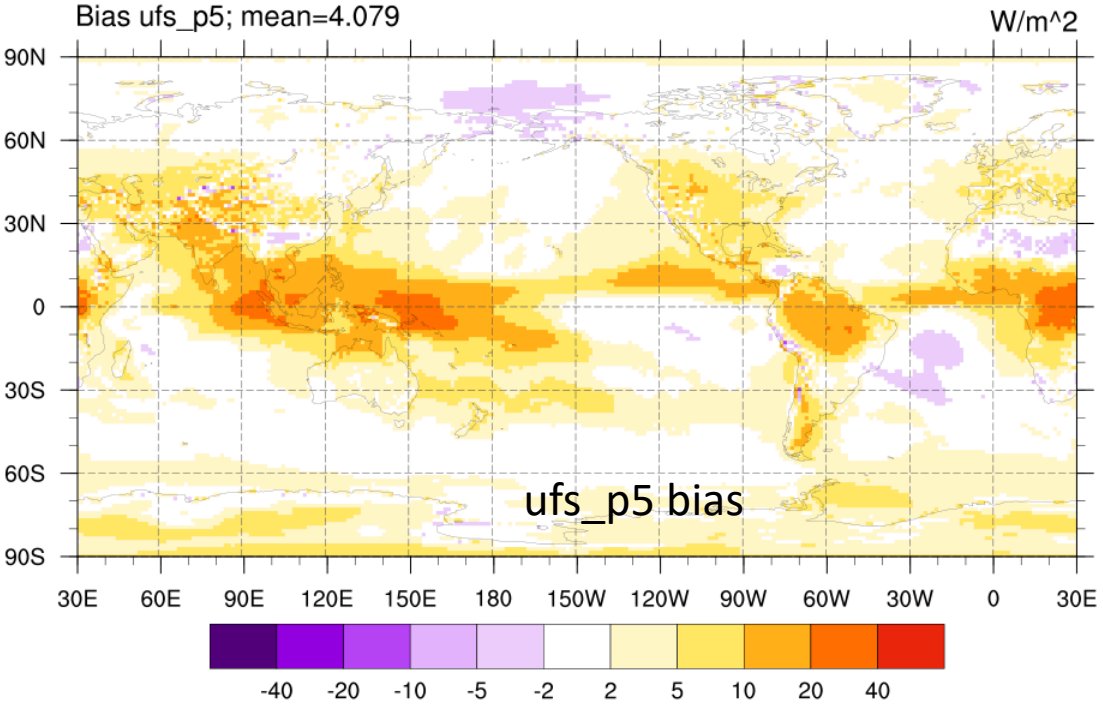


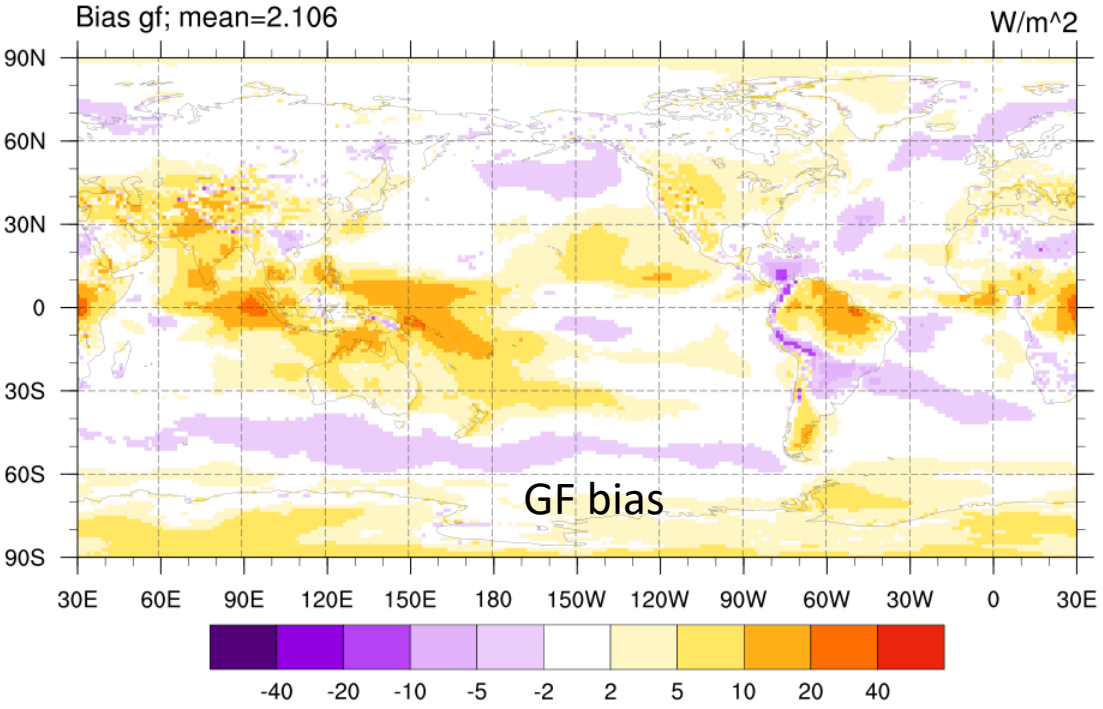


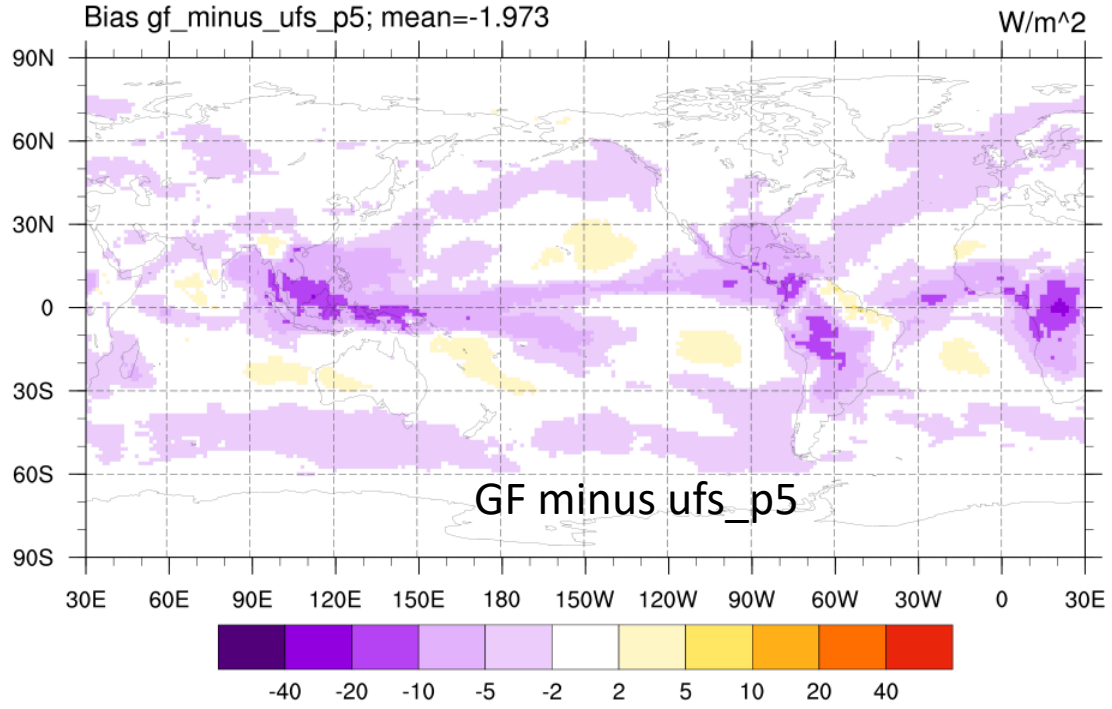


Preliminary results: OLR bias

- Using HRIS as truth
- Weeks **3-4**, ufs_p5 vs. GF: 168 cases
- GF has lower mean bias than ufs_p5
- Bias patterns similar to week 1, but larger
- Still deeper tropical convection in GF

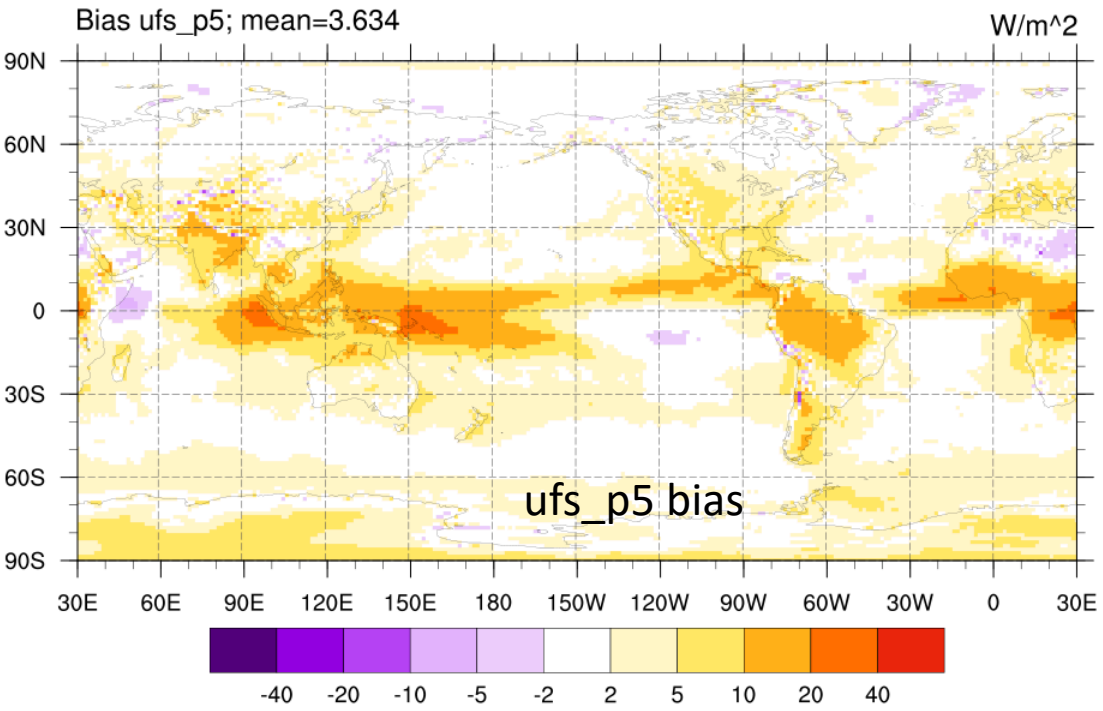


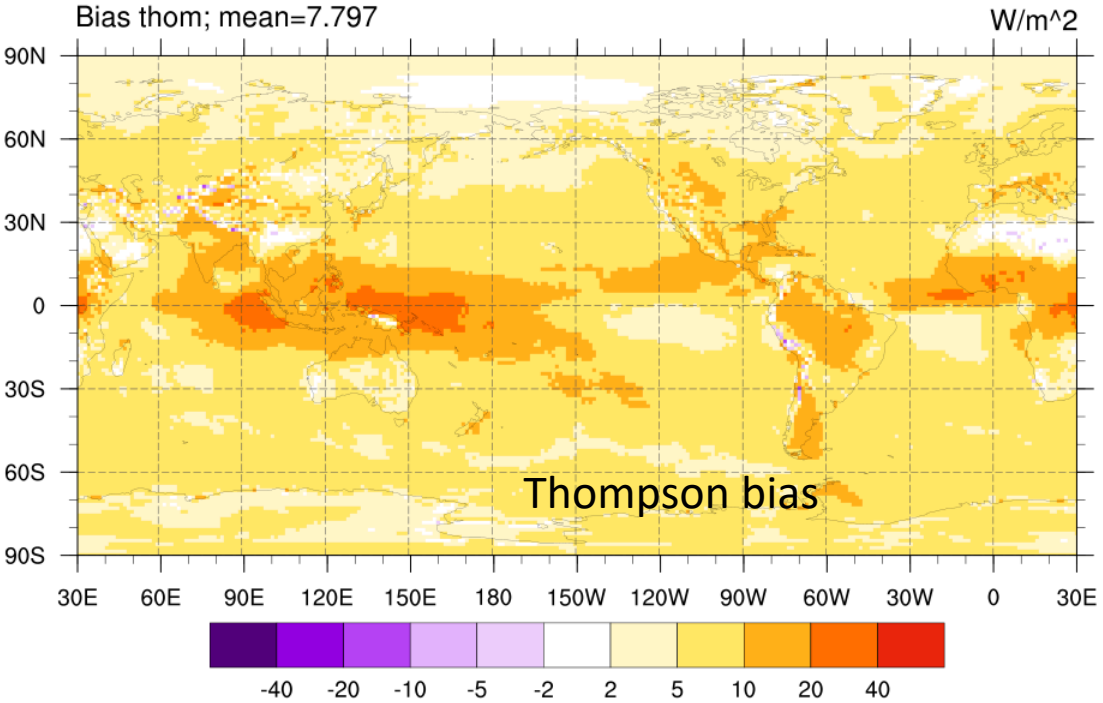


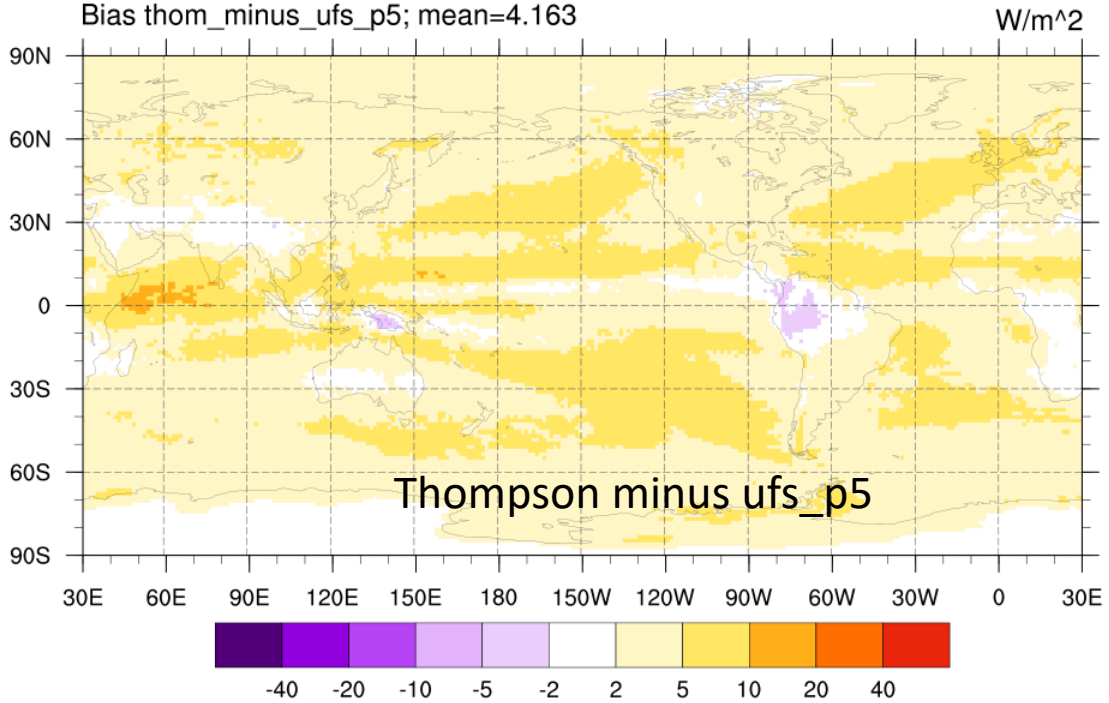


Preliminary results: OLR bias

- Using HRIS as truth
- Week **1**, ufs_p5 vs. Thompson: 166 cases
- Nearly uniformly higher OLR in Thompson: bug or feature? Requires further investigation

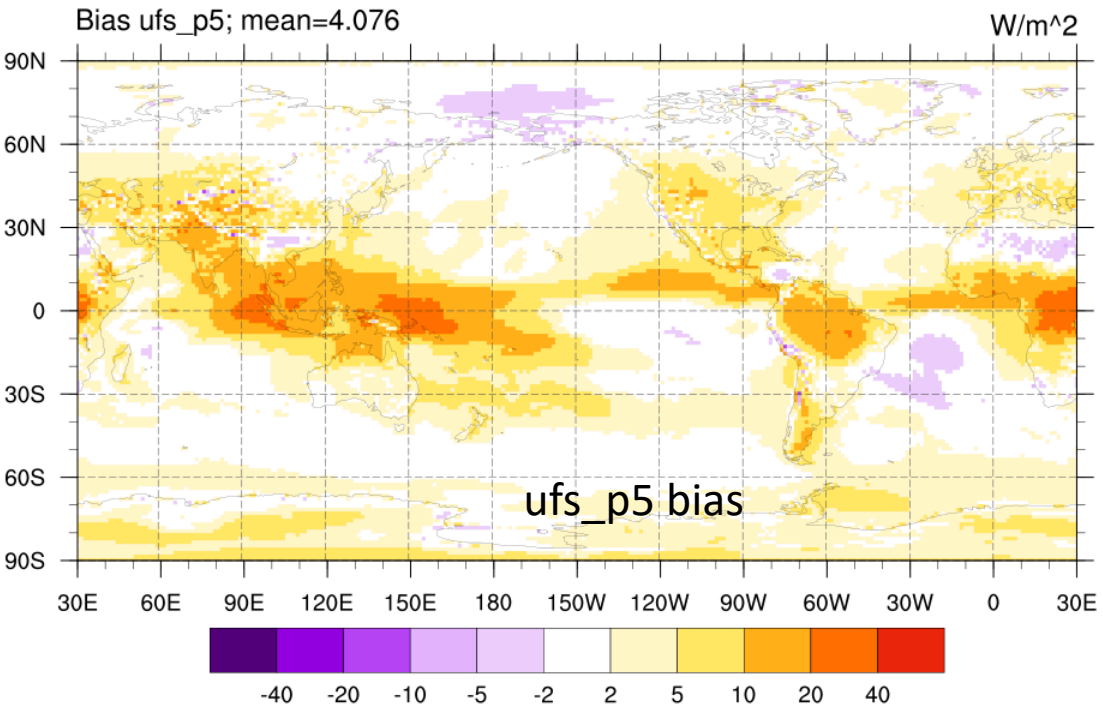


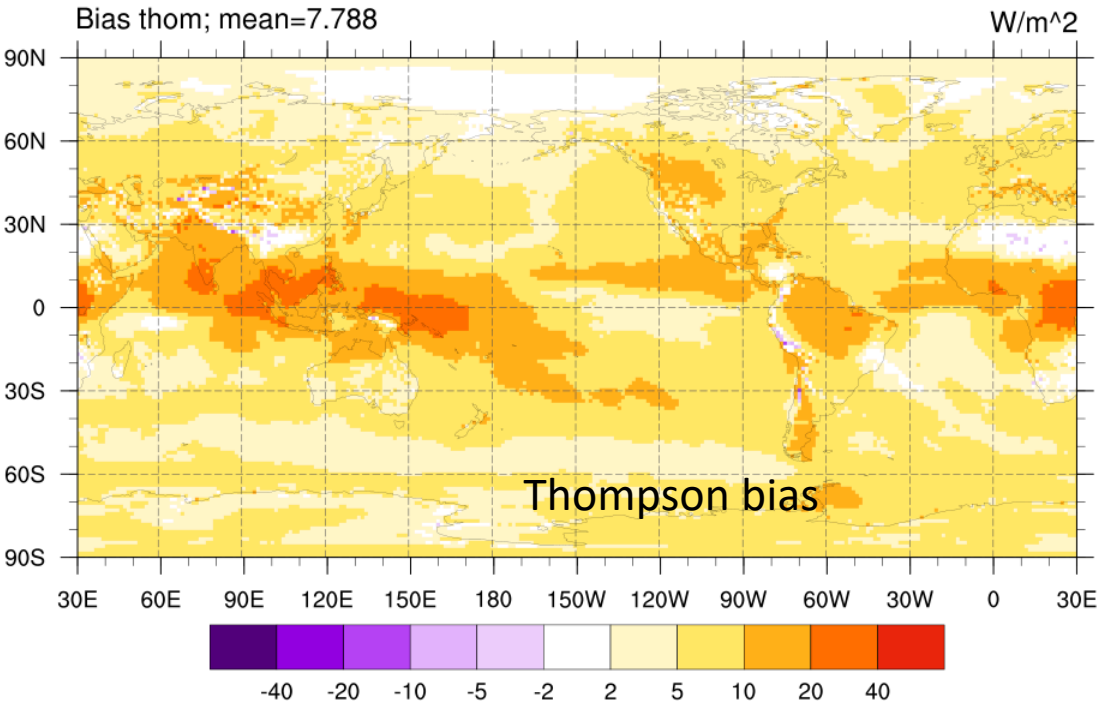


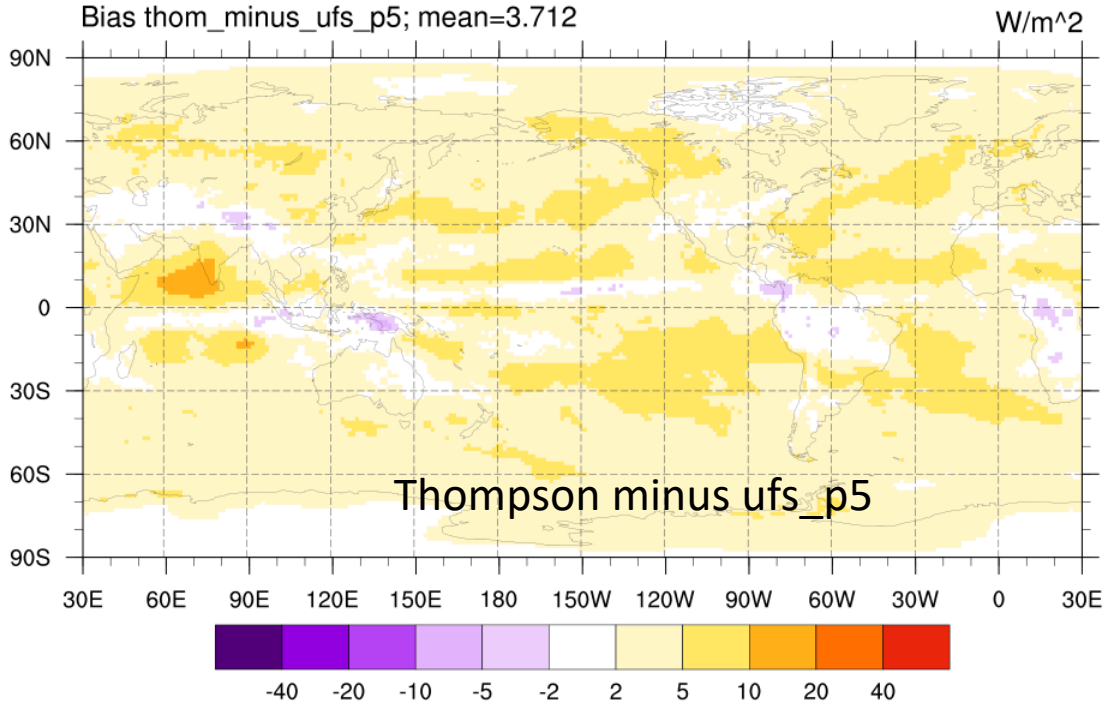


Preliminary results: OLR bias

- Using HRIS as truth
- Weeks **3-4**, ufs_p5 vs. Thompson: 166 cases
- Results quite similar to Week 1: Needs more investigation







Preliminary results: OLR bias

- Comparing ufs_p5 (control) vs. GF: GF has lower mean bias than ufs_p5
- GF consistently has lower OLR (deeper convection) in tropics
  - True across seasons (not shown)
  - True over all lead times (only week 1 vs. weeks 3-4 shown)
  - But… precip was slightly lower in GF over tropics: counterintuitive at first, but OLR does not necessarily have a one-to-one relationship with precip. This OLR/precip relationship is something for physics developers to look at more carefully

- For microphysics: Thompson had global increase in OLR right away
  - Bug or feature? Requires further investigation. Fortunately, this can be accomplished with much shorter runs
  - Note: Thompson scheme has had substantial modifications since these runs were performed, so results could look quite different now

Summary of preliminary results

- Swapping in alternative physics shows promise in subseasonal forecast performance

- Skill scores for MJO (and for Z500, not shown) are quite similar across all 4 experiments: why?
  - Coincidence or compensating errors?
  - Do **multiple** physics schemes need to be changed at once to see a bigger impact?

- "First-order" tuning may not require 35-d runs (potential to tune based on shorter runs)
  - But "second-order" impacts (e.g., impact of convection on T2m) can't take this shortcut

- CCPP provides a nice framework in which to do physics testing

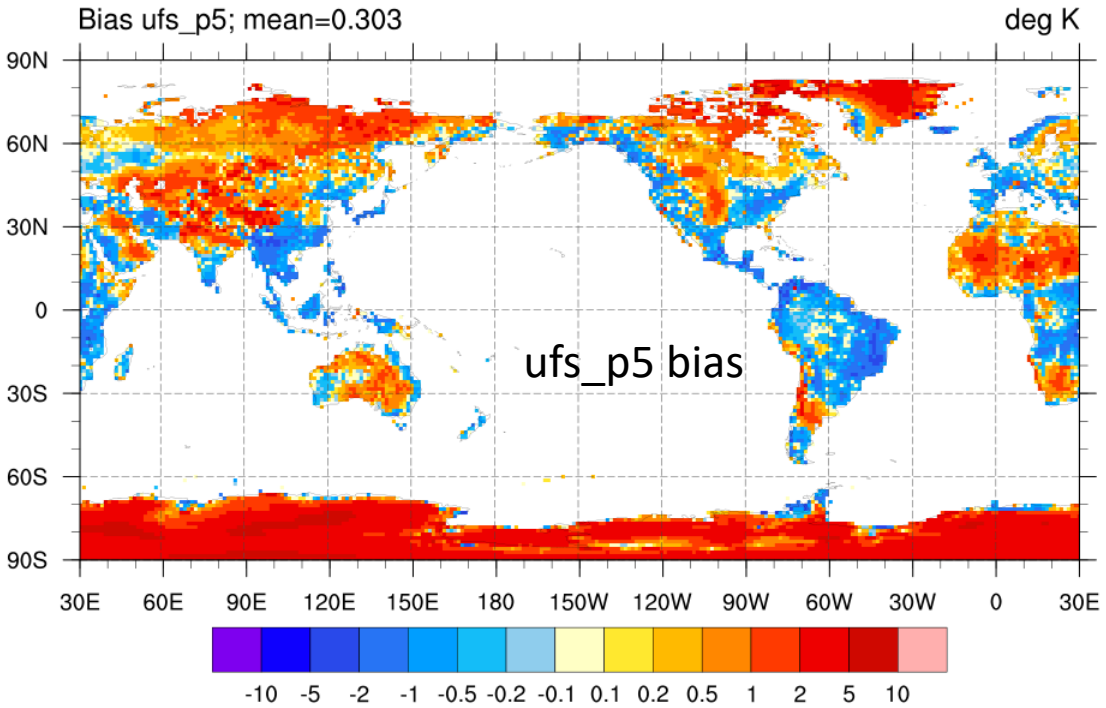
Future work

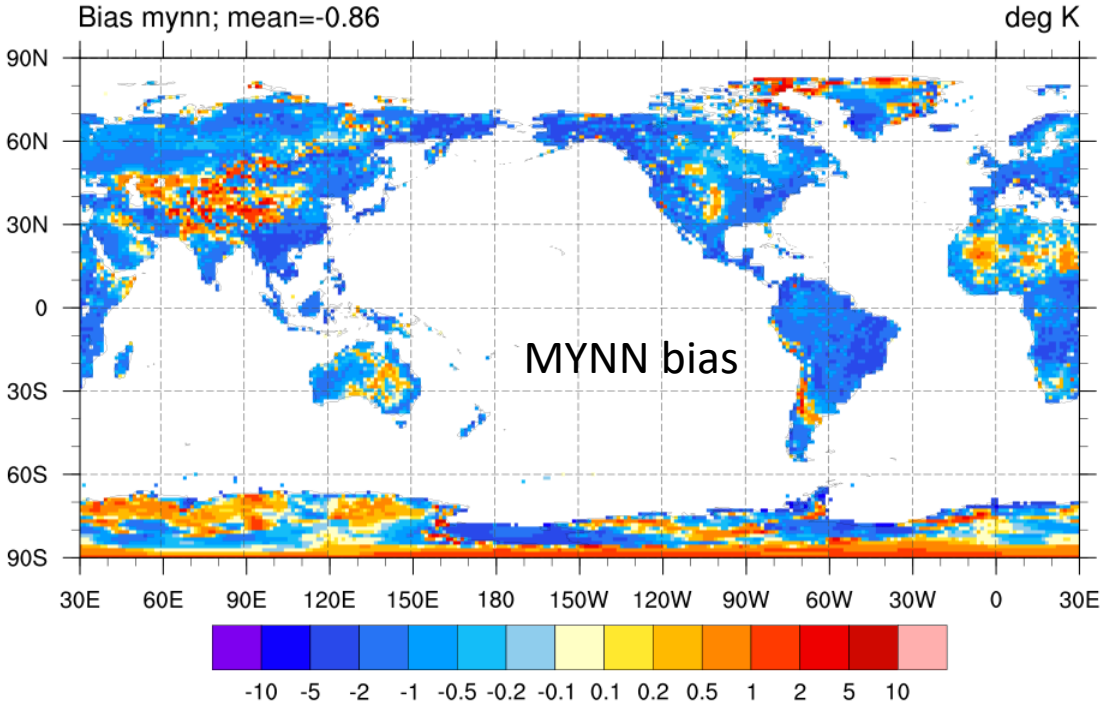
- Bias fields: Upper-level circulation, including MJO-related fields
- Vertical profiles: case studies, aggregate results (including moisture in PBL)
- Collaborate with physics developers: **why** are the biases the way they are? Can we adjust the schemes based on these results?
- As coupled UFS continues to develop (at EMC), conduct another set of experiments with newest model updates: are conclusions the same, or do (relative) results differ?

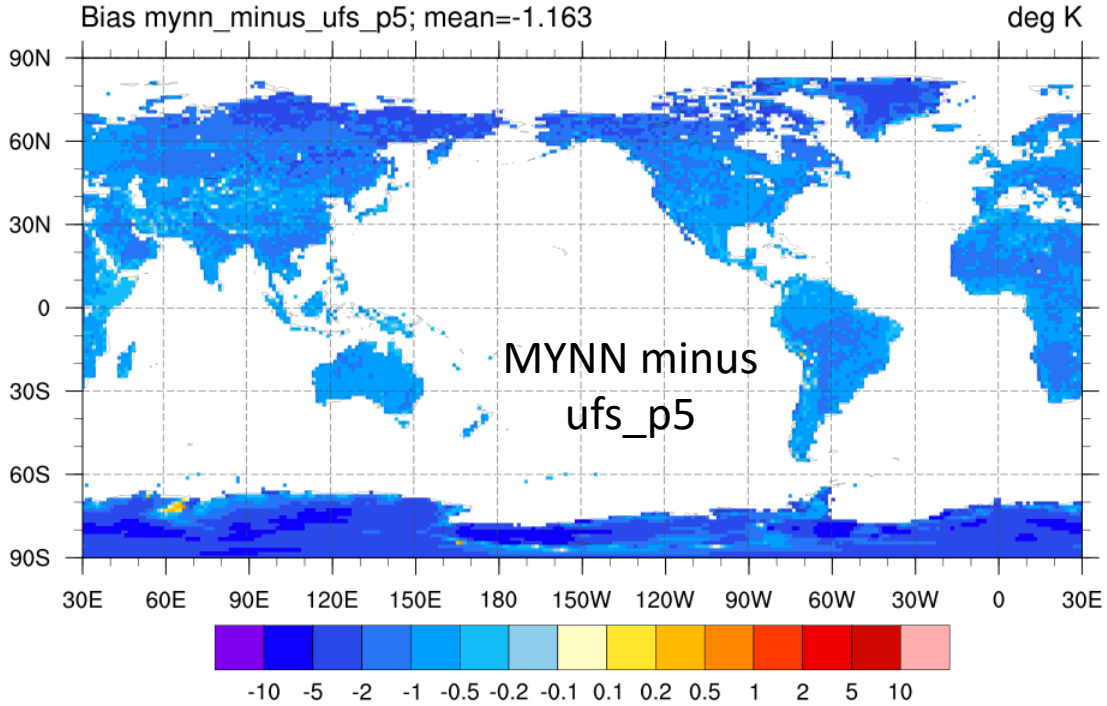
Extra slides

Preliminary results: T2m bias

- Land biases (left) using ERA5 as truth
- Week **1**, ufs_p5 vs. MYNN: 167 cases
- Systematic cooling in MYNN relative to ufs_p5 (bottom right)

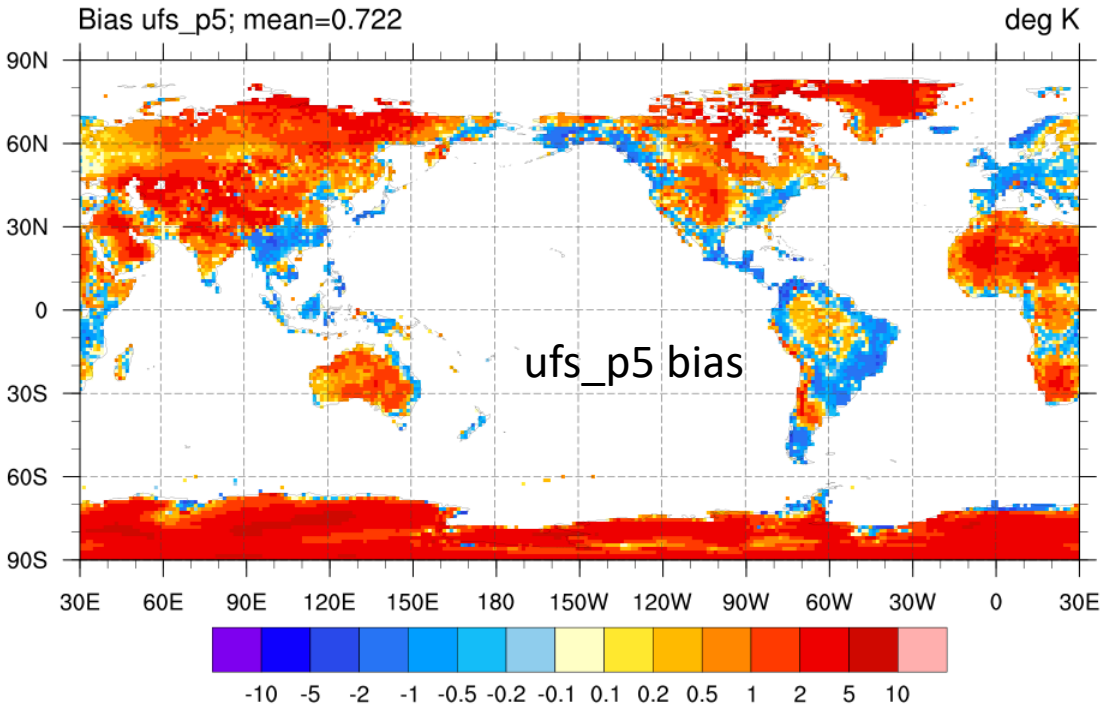


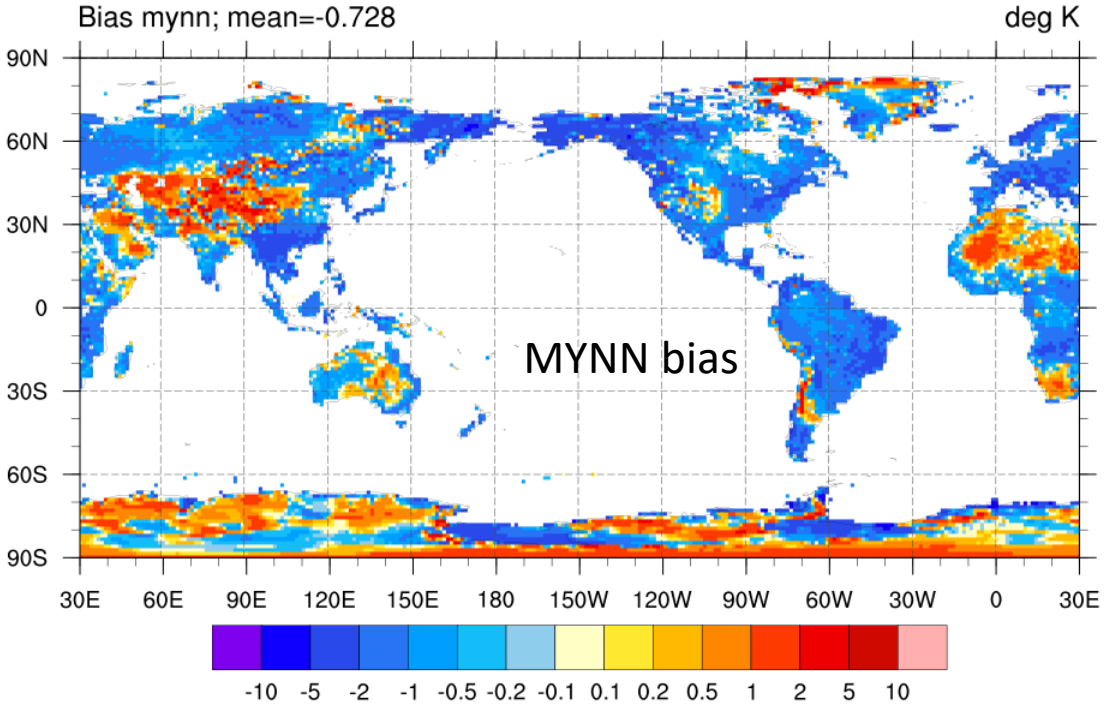


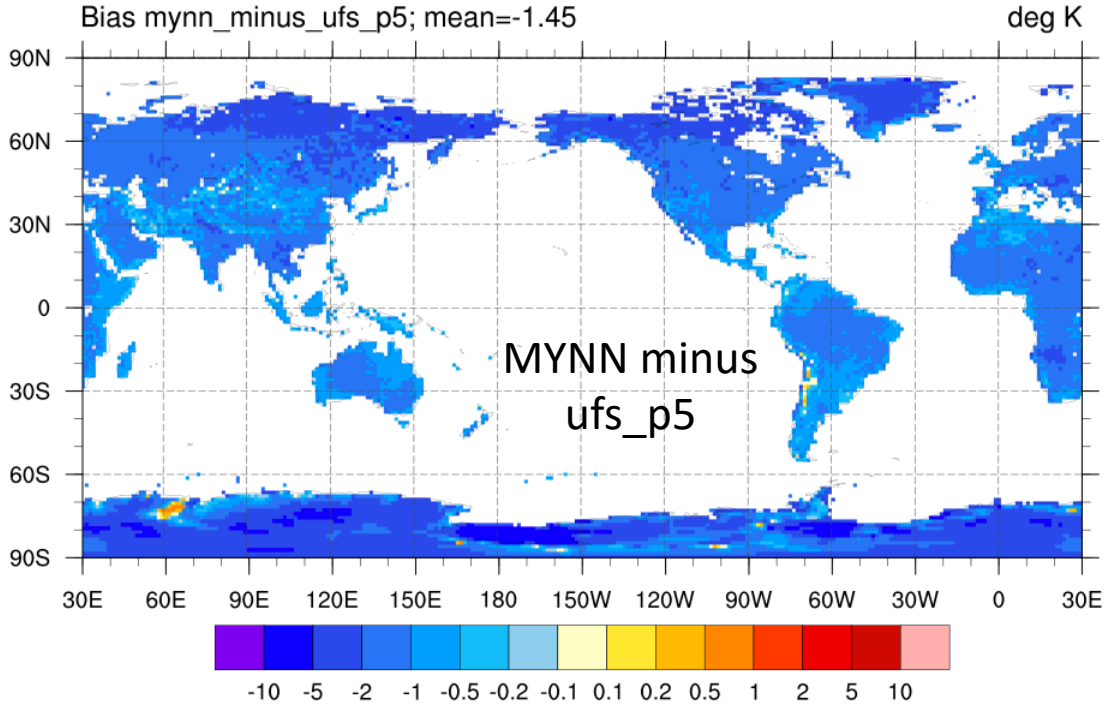


Preliminary results: T2m bias

- Land biases (left) using ERA5 as truth
- Weeks **3-4**, ufs_p5 vs. MYNN: 167 cases
- Patterns extremely similar to week 1!
- Systematic cooling in MYNN relative to ufs_p5 (bottom right)

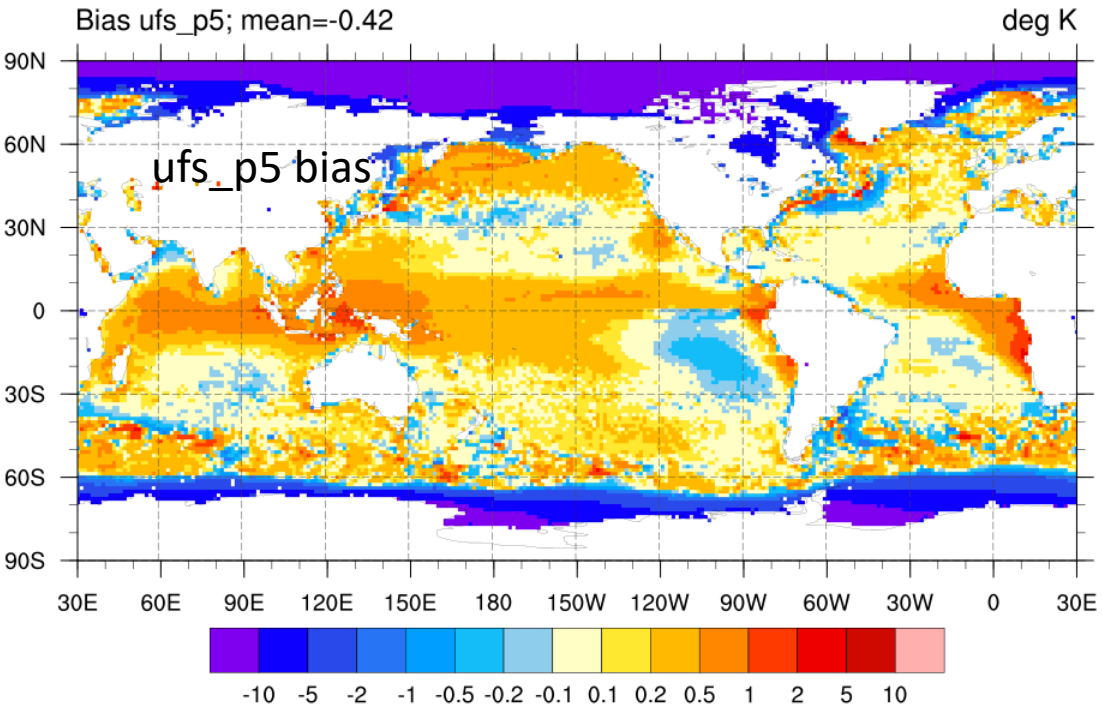


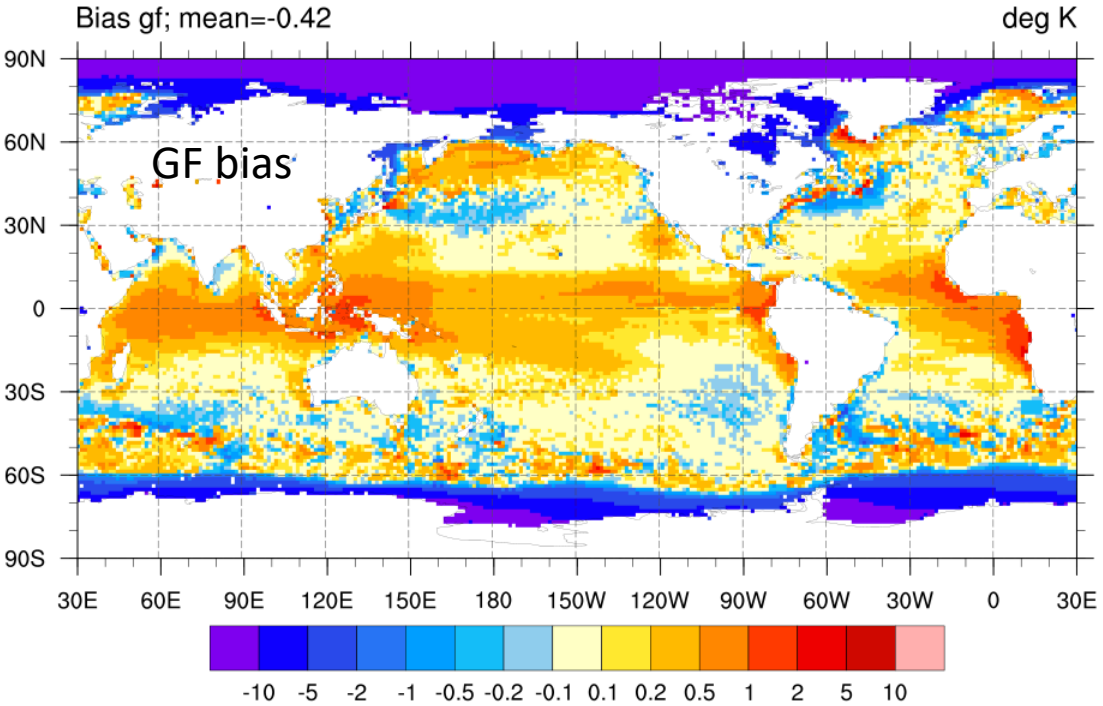


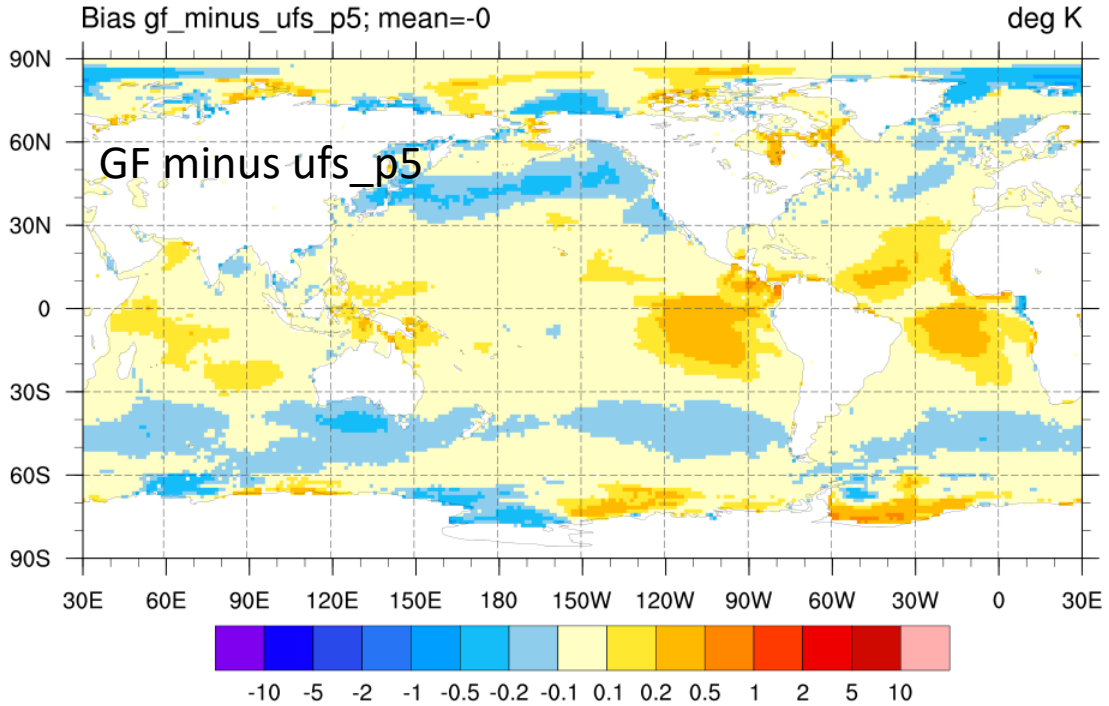


Preliminary results: SST bias

- Using OSTIA as truth
- Weeks **3-4**, ufs_p5 vs. GF: 168 cases
- GF slightly cooler in mid-latitudes, slightly warmer off western coasts of Africa and S. America

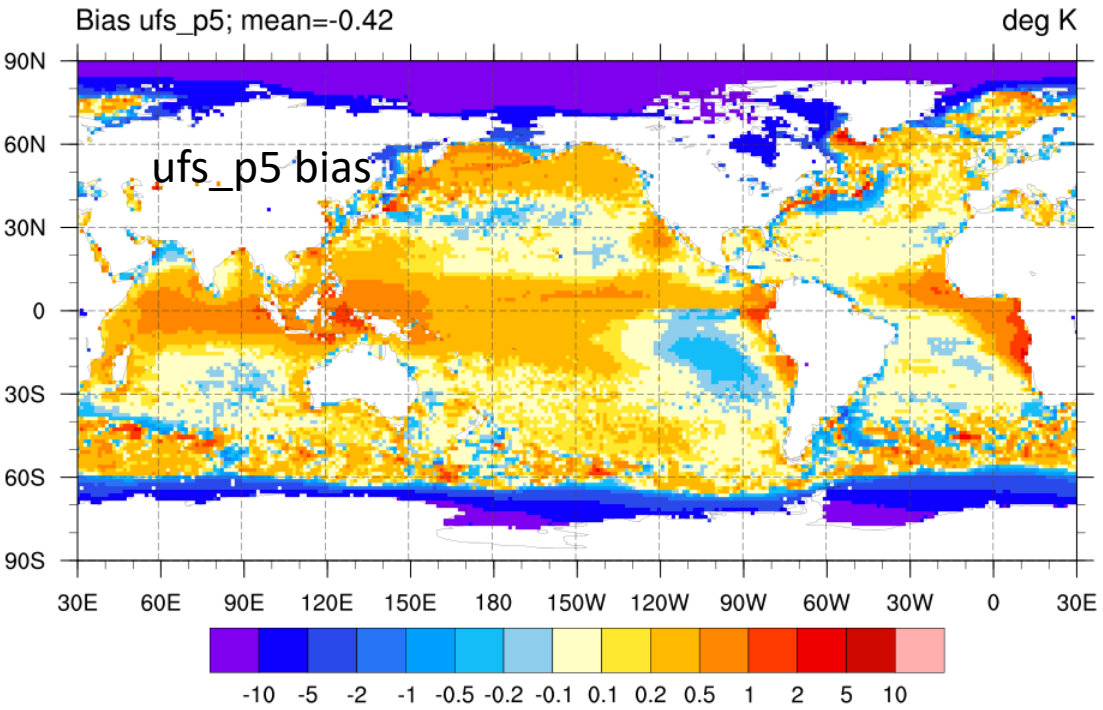


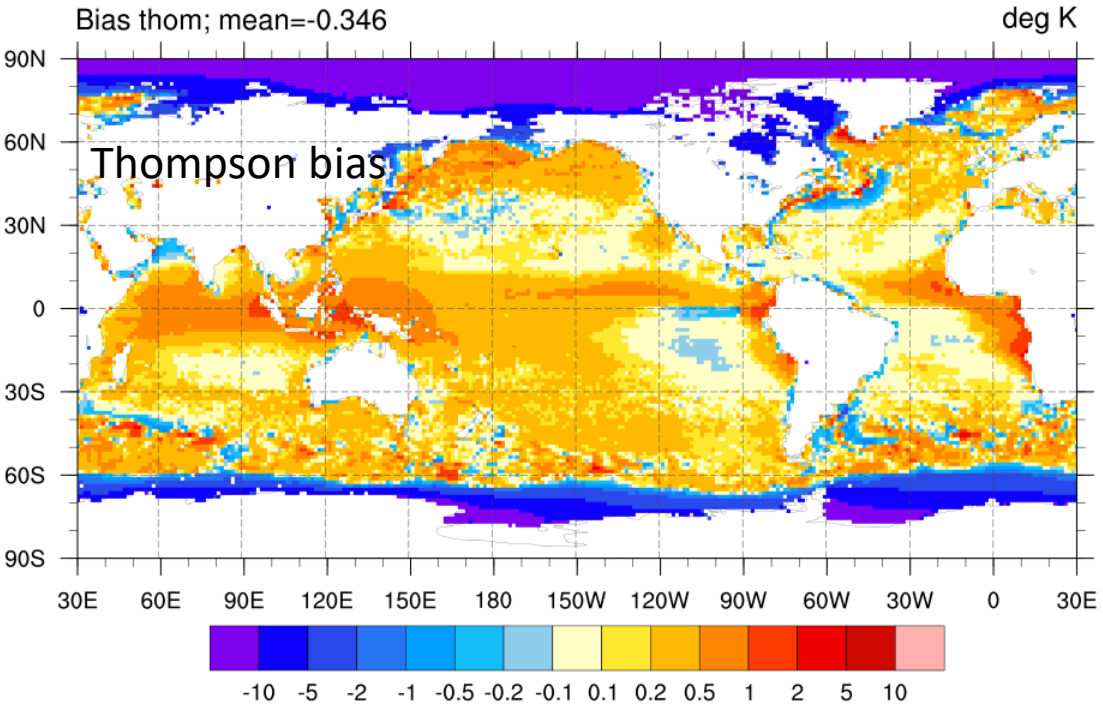


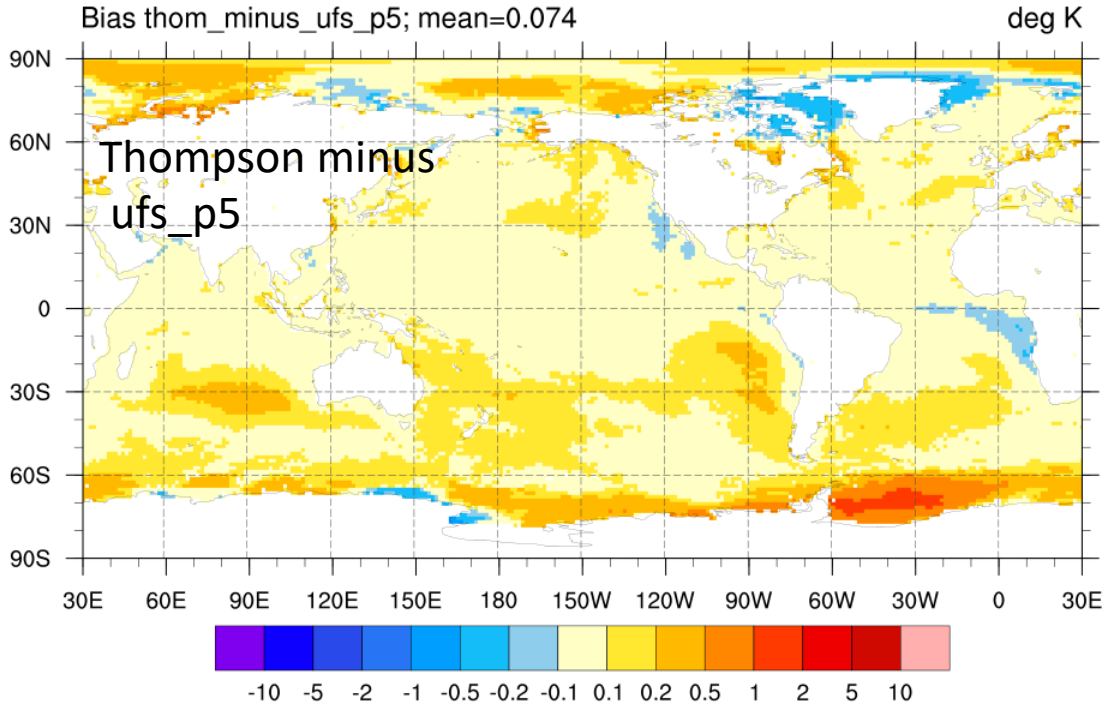


Preliminary results: SST bias

- Using OSTIA as truth
- Weeks **3-4**, ufs_p5 vs. Thompson: 166 cases
- Thompson slightly warmer off west coasts of S. America and Australia







Preliminary results: RMM skill score

- Control (“ufs_p5”) almost always has the lowest RMM skill score
- Using score of 0.6 as a threshold, Thompson experiment is skillful out to 14 days (**year round**); others skillful to 13 days
- Still need to look at various teleconnections: RMM skill is meaningless if relationship between tropics and mid-latitudes is wrong