Sensitivities of Subseasonal UFS Simulations to Changes in Parameterizations of Convection, Cloud Microphysics, and Planetary Boundary Layer

Benjamin W. Green[1,2], Eric Sinsky[3,4], Shan Sun[1], Vijay Tallapragada[3], Georg Grell[1]

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

Presentation to UFS Webinar, 11 May 2023

Funding for this work from: WPO Award NA20OAR4590314

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-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 **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 FY2024, all of NOAA's operational global Earth system prediction for lead times from **0-35+ days** will be consolidated into a **single UFS-based model** (GEFSv13)
  - Challenge: Ensure model has "reasonable" skill from daily to subseasonal time scales

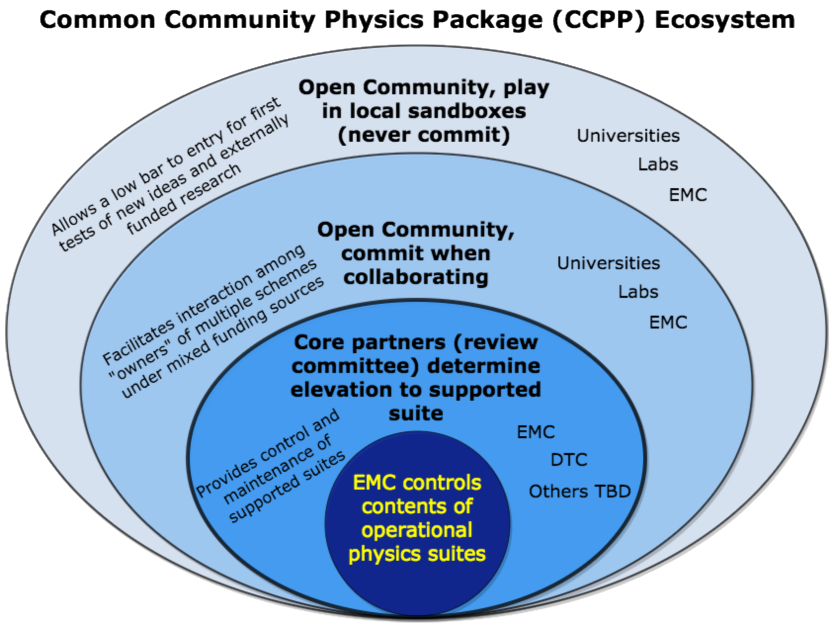
NPS Modeling System	Current Version	Q4 FY 22	Q1 FY 23	Q2 FY 23	Q3 FY 23	Q4 FY 23	Q1 FY 24	Q2 FY 24	Q3 FY 24	Q4 FY 24	Q1 FY 25	Q2 FY25	Q3 FY25	Q4 FY 25	Q1 FY 26	Q2 FY26	Q3 FY26	UFS Application
Global Weather, Waves & Global Analysis	GFS/ GDASv16.1																	UFS Medium Range & Sub-Seasonal
Global Weather and Wave Ensembles, Aerosols	GEFSv12	Coupled Reanalysis and SubX Reforecast Production							GFSv17/ GEFSv13	Seasonal Reforecast Production							GFSv18/ GEFSv14/ SFSv1	
Short-Range Regional Ensembles	SREFv7																	
Global Ocean & Sea-Ice	RTOFSv2				RTOFSv3													UFS Marine & Cryosphere
Global Ocean Analysis	GODASv2				GODASv3													
Seasonal Climate	CDAS/ CFSv2																	UFS Seasonal

Subseasonal moves from CFSv2 to GEFSv13

UNOFFICIAL!!! UNOFFICIAL!!!

Paradigm shift at NOAA: UFS and CCPP

- CCPP: Common Community 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 UFS runs



Credit: DTC website

Project setup (**in 2019**, when proposal was written)

- 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**

Project setup (**by 2022**, as project ended)

- Still using EMC's coupled UFS prototypes (they run "Xa" where X = 5, 7, 8) to do one-at-a-time physics testing:

Experiment #	Experiment Name	Convection	Boundary Layer	Microphysics
5a	UFS_P5	SASAS	EDMF	GFDL
5b	GF_5	**GF**	EDMF	GFDL
5c	MYNN_5	SASAS	**MYNN**	GFDL
5d	THOM_5	SASAS	EDMF	**Thompson**
7a	UFS_P**7**	SASAS	EDMF	GFDL
7b	GF_**7**	**GF**	EDMF	GFDL
8a	UFS_P8	SASAS	EDMF	**Thompson**
8b	GFDL_8	SASAS	EDMF	**GFDL**

- EMC switched their control microphysics from GFDL to Thompson for P8 based on medium-range results
- Compare (5a,b,c,d), (7a,b), (8a,b) to see impacts of physics changes (both isolated – within one prototype, and cumulative – across prototypes)

Experimental design: more details

- Run 5 additional sets of experiments (see previous slide). Follow protocol for EMC's coupled UFS prototypes for each experiment:
  - Initialize 1st and 15th of every month from 1 April 2011 through 15 March 2018 (**168 cases**)
  - C384 (~25 km) resolution: daily 1°x1° 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:
  - In our P5 tests, 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
  - P5: 64 vertical layers, dt=450s. P7 and P8: 127 vertical layers, dt=300s
    - P7, P8 need much more CPU than P5: Longer wait time for results; can't test as many schemes
  - Growing pains of UFS: MYNN unable to run in P7; Thompson had unrealistic precip in P7. Continued efforts throughout NOAA should prevent such issues going forward
  - Unique opportunity to isolate microphysics impact for P7 (GFDL) → P8 (Thompson)

What do we want to investigate?

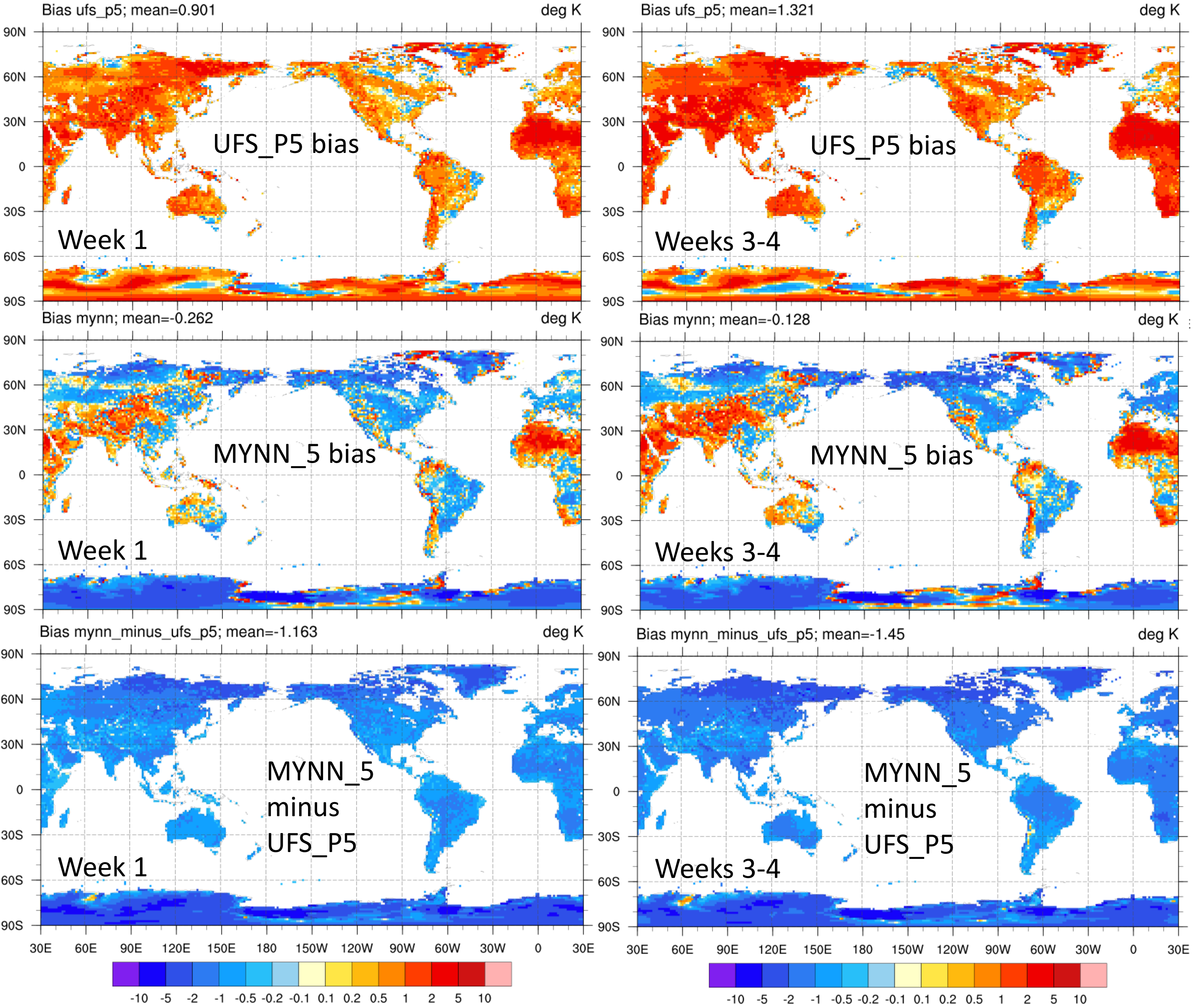
- Biases
  - T2m and precipitation are meaningful for end-users of S2S products
  - Global circulation biases are important too, due to teleconnections on S2S timescales
  - Maps (spatial patterns)
  - Vertical profiles (horizontally aggregated): Easy illustration of temporal evolution of bias
- Deterministic skill scores: ACs for Z500 & RMM index (MJO); HSS for precip

- Multi-physics ensembles: Research question **only** (untenable in ops), but expect forecast skill could be improved

- Note: More advanced (“process-based”) diagnostics left to others (data is on tape, happy to share!). Important to document/mitigate biases

Results: Prototype 5

- Prototype 5 run in 2020 (a lifetime ago in UFS development)
- 4 experiments:
  - “UFS_P5”: EMC control
  - “GF_5”: Use GF convection instead of SASAS
  - “MYNN_5”: Use MYNN PBL instead of default GFS PBL
  - “THOM_5”: Use Thompson MP instead of default GFDL MP
- P5 was the only round in which all 3 additional experiments were run
- Reminder for later: Careful not to generalize beyond P5 results from any scheme!

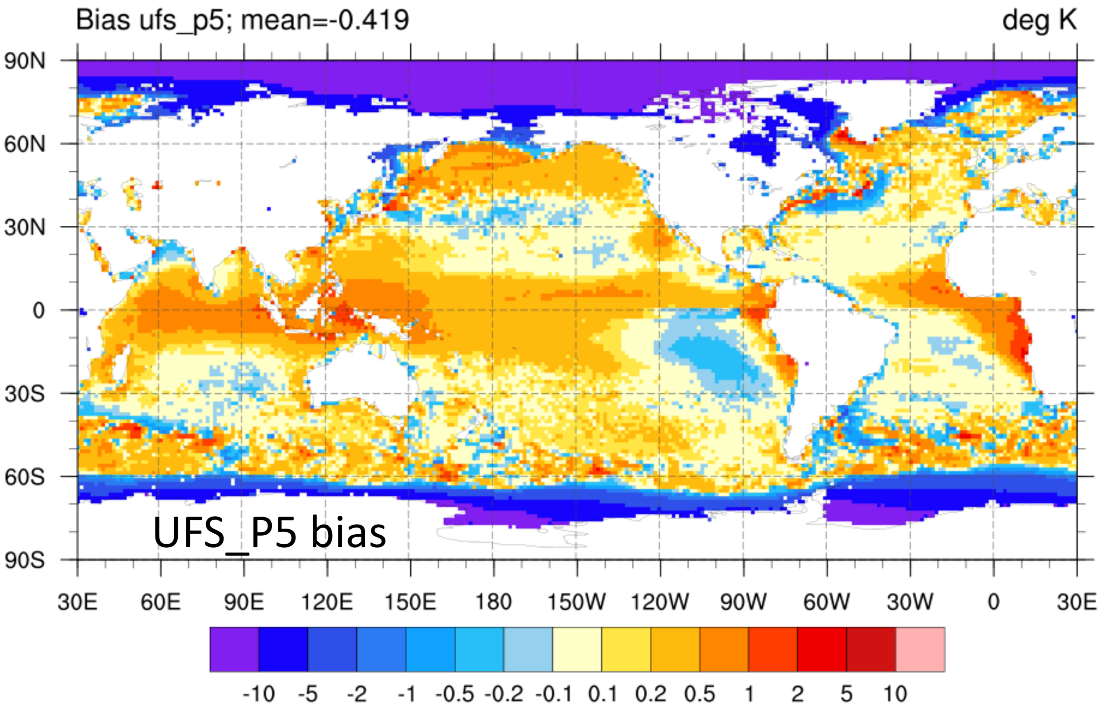


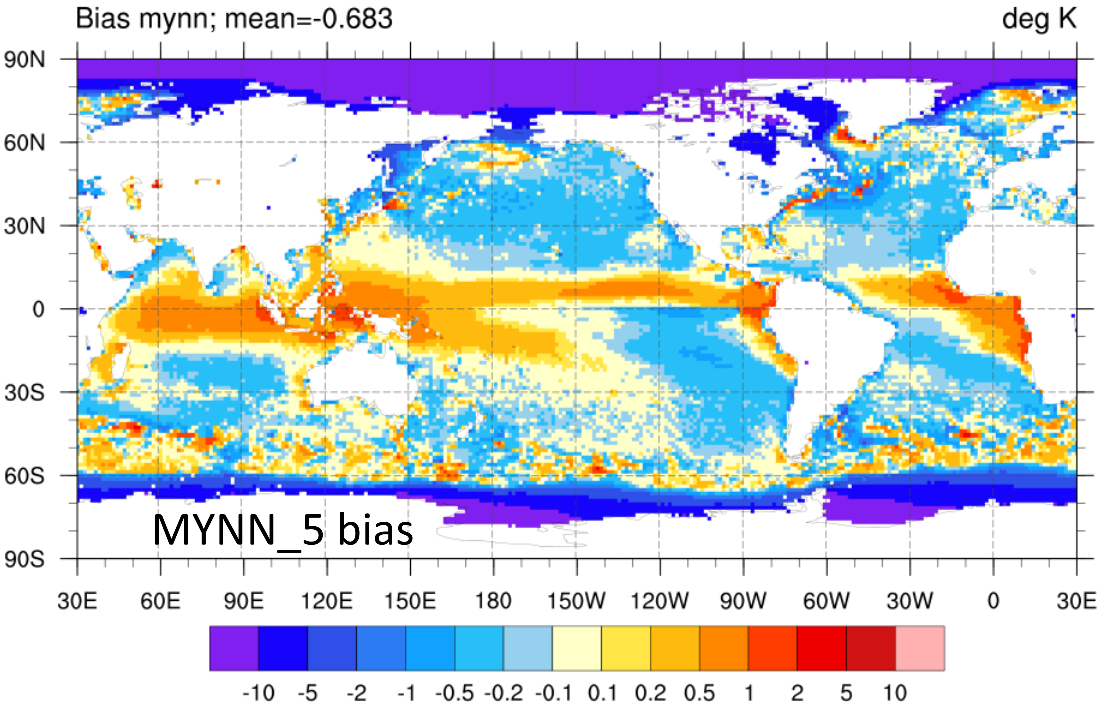
P5 results: T2m bias

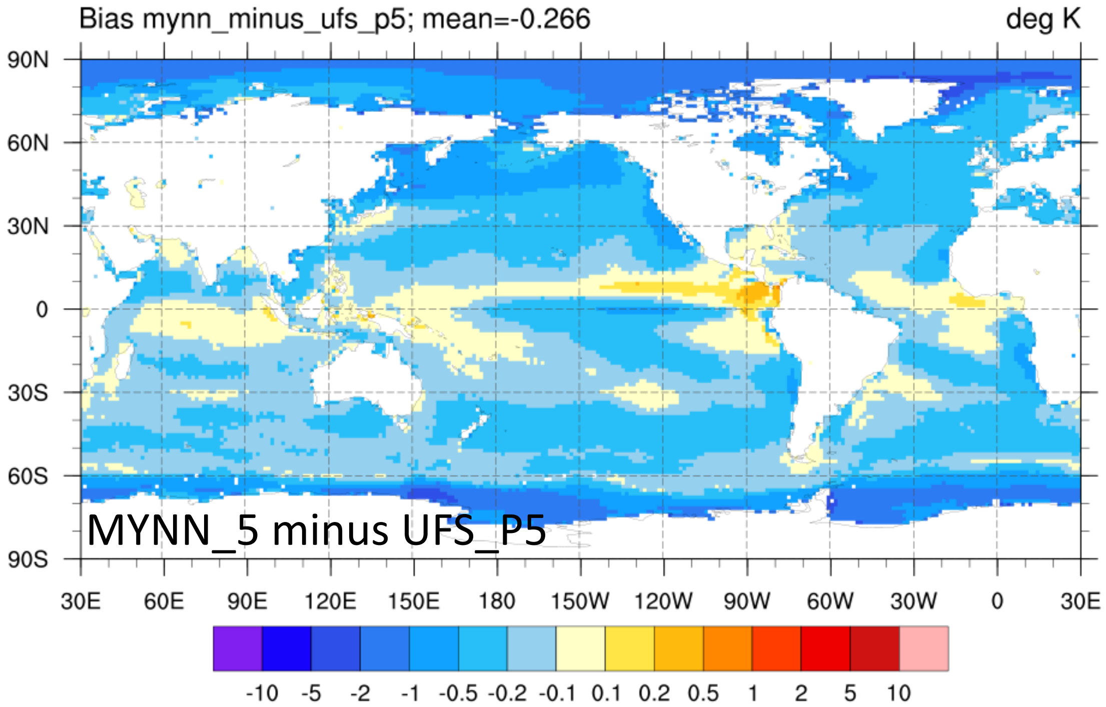
- Compare UFS_P5 (top) with MYNN_5 (middle) land biases (CFSR as truth)
- Difference plotted on bottom row
- Systematic cooling in MYNN_5 compared with UFS_P5
- Weeks 3-4: Patterns extremely similar to week 1!

P5 results: SST bias

- Using OSTIA as truth
- Weeks **3-4**, UFS_P5 vs. MYNN_5
- MYNN_5 cooler than UFS_P5 almost everywhere (exception: ITCZ); recall MYNN_5 has cooler T2m over all land



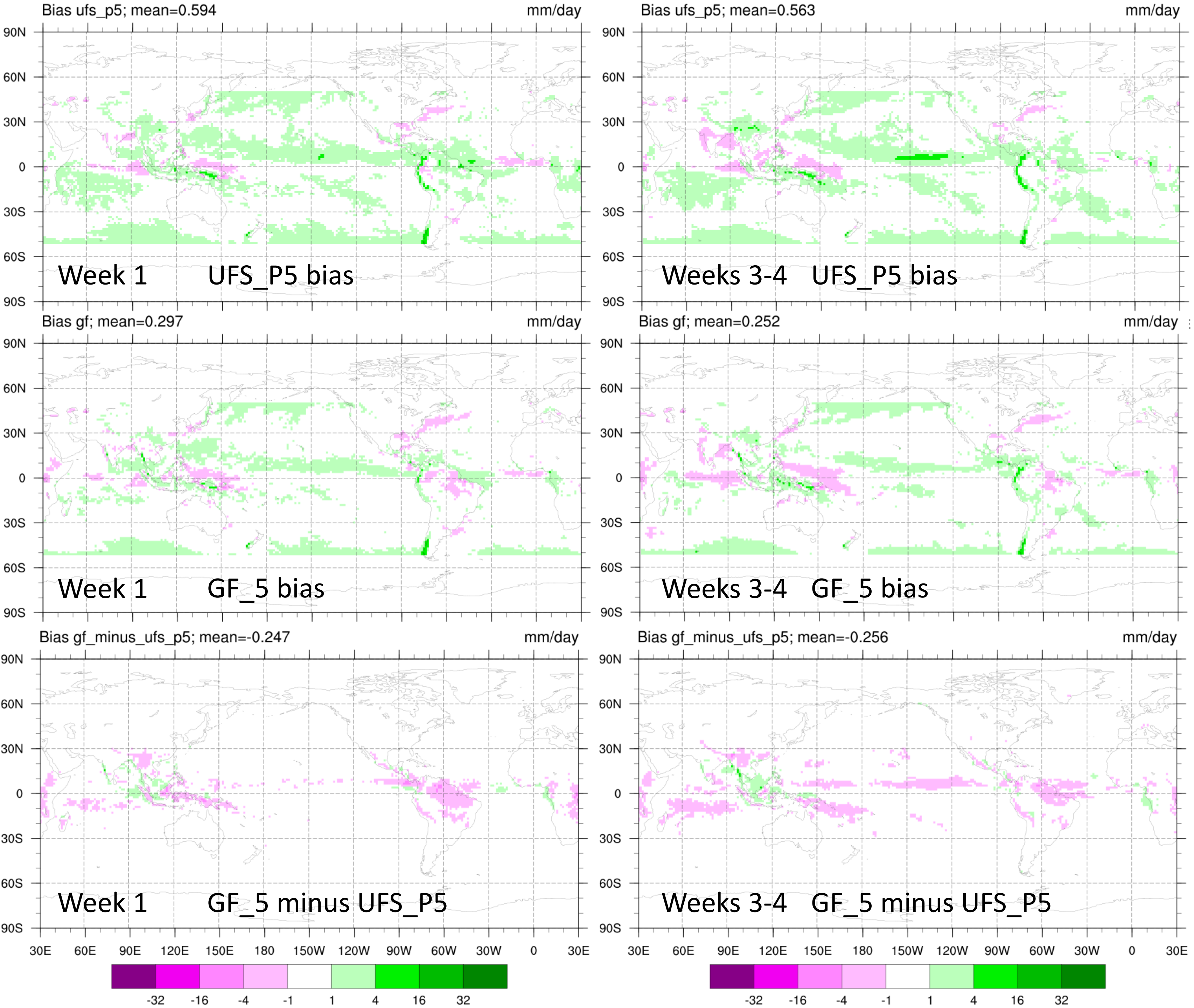




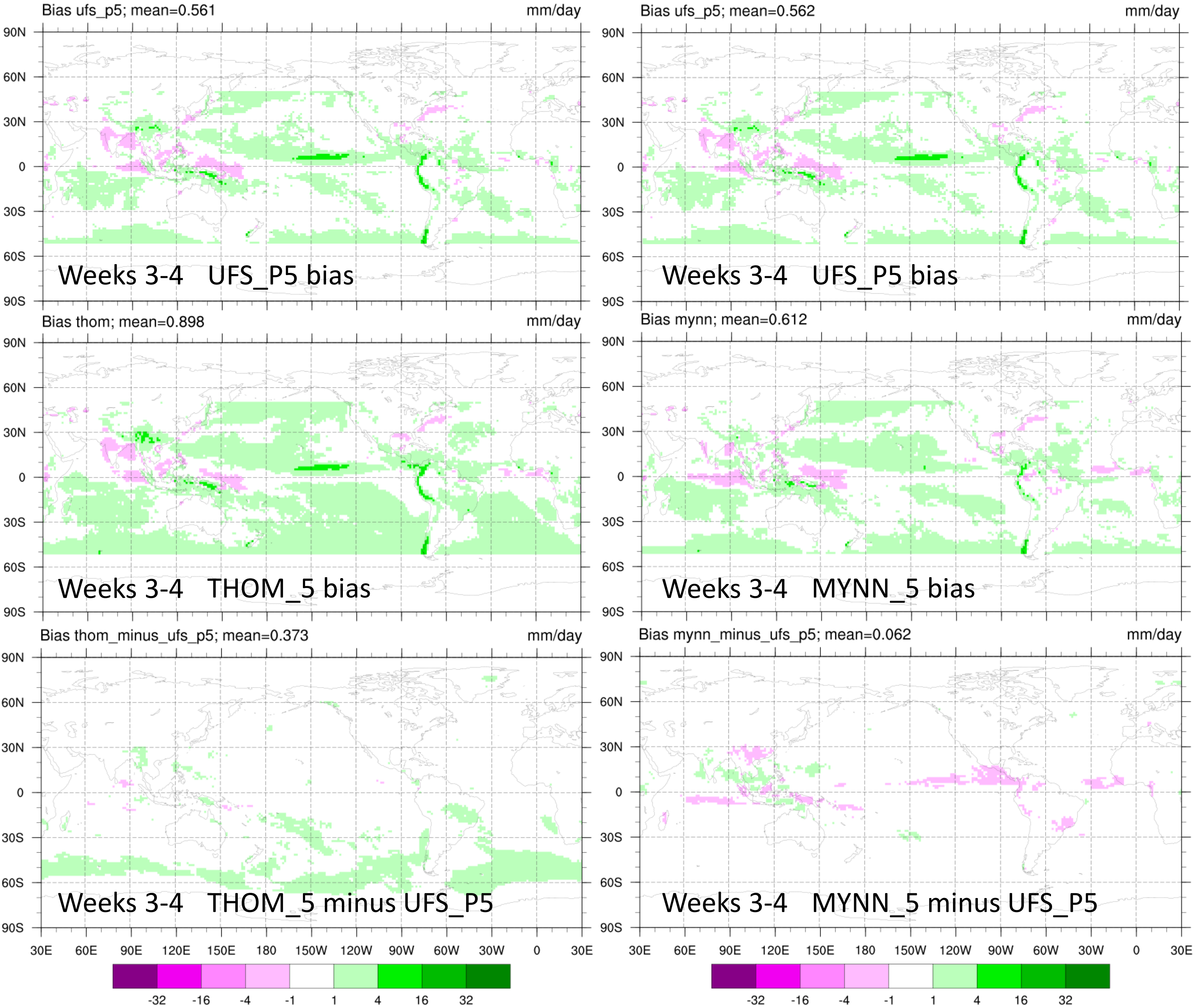
P5 results: T2m bias **over land**, and SST

- Comparing UFS_P5 (control) vs. MYNN_5: MYNN_5 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 grow more slowly than T2m, but MYNN_5 colder than UFS_P5 almost everywhere (exception: 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 – biases may be “baked in” after first ~1 week. More on this later…
  - But this potential shortcut likely won’t work for “second-order processes” (e.g., convection influencing T2m)

P5 results: QPF bias



- Compare UFS_P5 (top) with GF_5 (middle); TRMM as truth
- Difference plotted on bottom row
- GF_5 has smaller global mean bias than UFS_P5
- Systematic drying in GF_5 relative to UFS_P5
- Week 1 plots look very similar to weeks 3-4 plots



P5 results: QPF bias

- Compare UFS_P5 (top) with THOM_5 (middle left) and MYNN_5 (middle right); TRMM as truth
- Difference plotted on bottom row
- THOM_5 differs most from UFS_P5 in extratropics
- MYNN_5 differs most from UFS_P5 in tropics (moisture fluxes from ocean?)

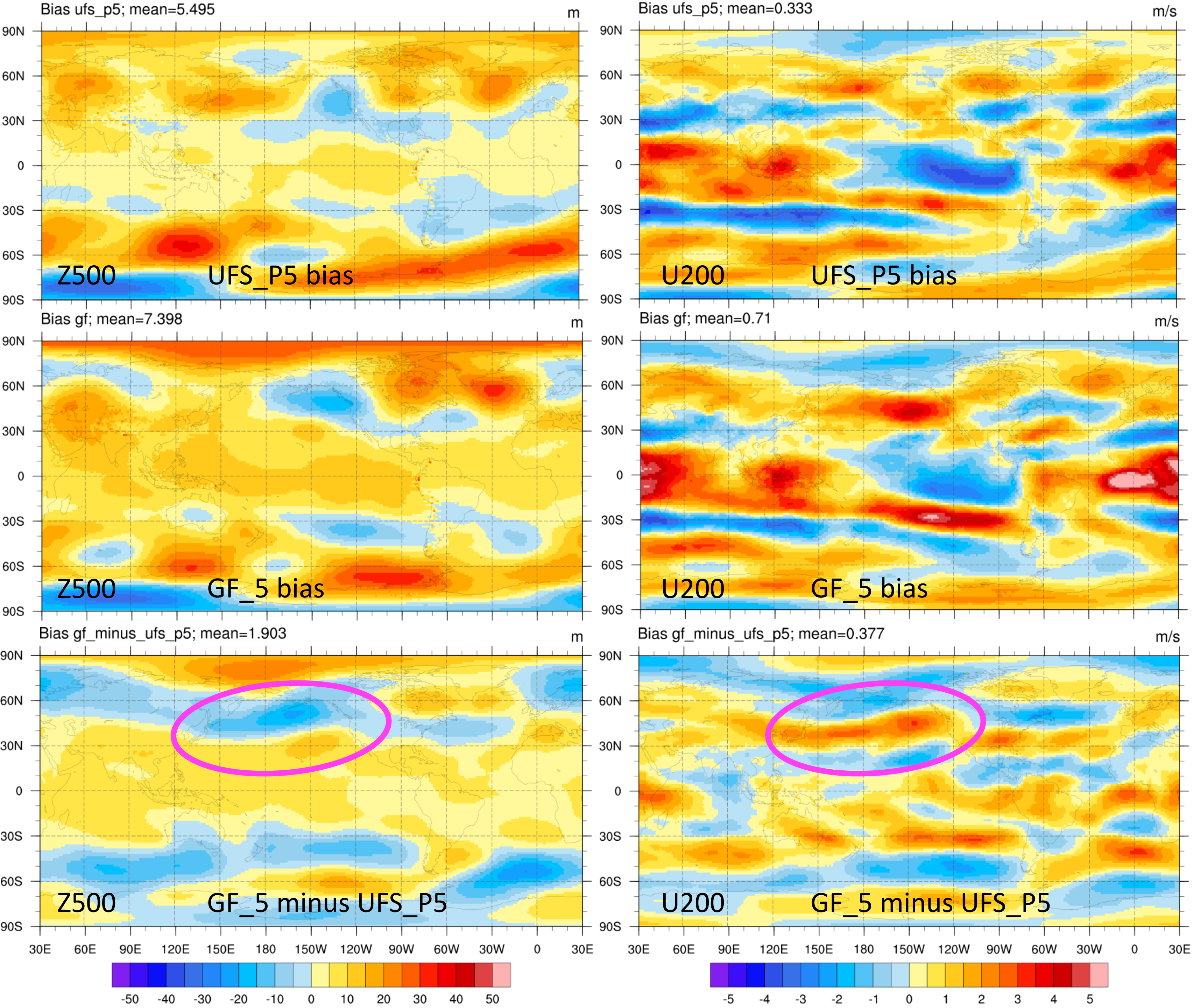
P5 results: Precipitation biases

- Biases become quite established within the first week (only shown for GF_5 & UFS_P5)
- GF_5 most different from UFS_P5 in tropics (convective precip)
- THOM_5 most different from UFS_P5 in extratropics (stratiform precip)
- MYNN_5 most different from UFS_P5 in tropics (moisture fluxes from ocean?)

- Precipitation biases useful for end-users, but model tuning for QPF notoriously difficult – may be useful to look at circulation biases first…

P5 results: Circulation biases

- Weeks 3-4 predictability for T2m (and especially QPF) is very low
- Mid-latitude subseasonal predictions rely heavily on teleconnections with tropics
- Mass/momentum biases might give some insight into global teleconnections
- Next, look at Z500 and U200



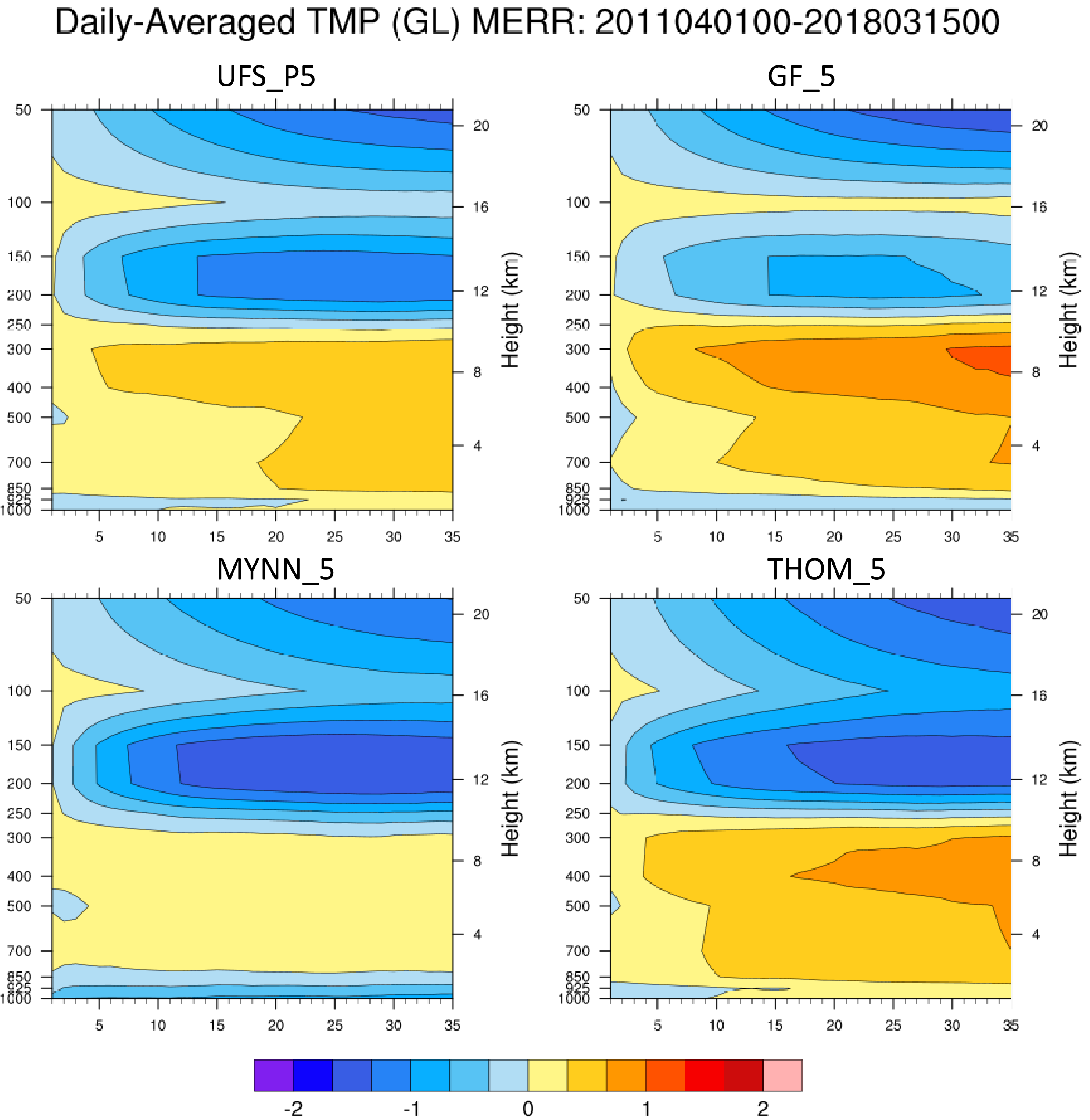
P5: Z500, U200

- Bias evolution slower; only show weeks 3-4
- Seasonal breakdown even noisier
- Both GF_5 and UFS_P5 have high Z500 bias, especially in SH
- Diff field shows GF_5 has higher heights in tropics and lower heights in NH extratropics: more baroclinicity
- U200 extremely noisy

Temporal evolution of biases

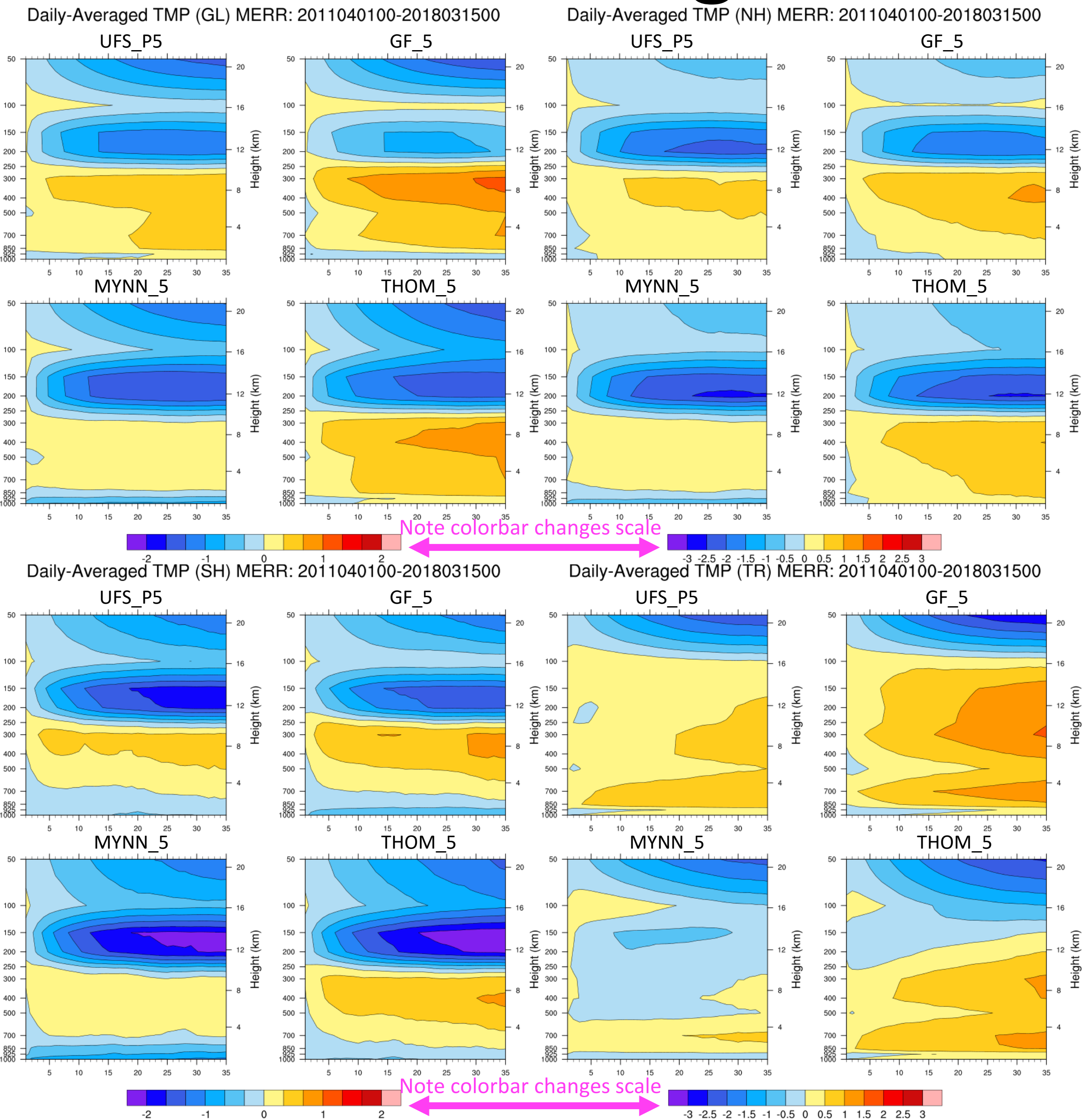
- Maps are useful for identifying geographic regions of interest, but can’t easily show temporal evolution
- Next set of plots: Average biases over a large spatial area (globe, hemisphere, tropical band) and show temporal evolution as a function of height
- These height (y-axis) vs. time (x-axis) plots will give a general sense of how quickly mean-state biases become established

P5 results: Global temperature bias (vs. CFSR)



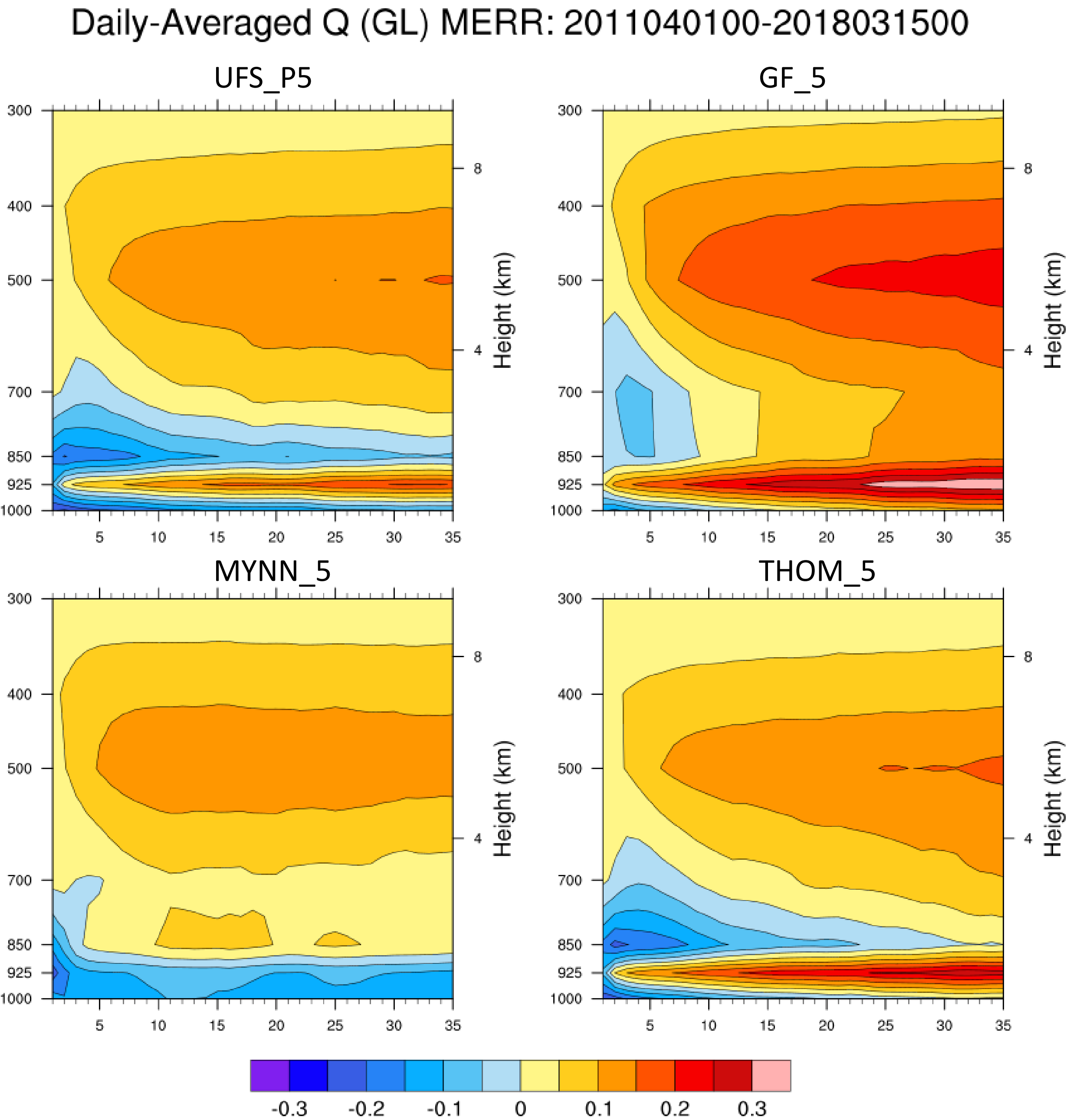
- Troposphere has a nearly monotonic temperature increase with increasing lead time; stratosphere has opposite pattern
  - Consistent with earlier: Could potentially look at biases in first ~1-2 weeks to get a sense of biases in weeks 3-4
  - May allow for **shorter runs to guide some subseasonal physics development**
- Switch from warm troposphere to cold stratosphere: reduced static stability near tropopause
- MYNN_5 (recall: 2020 version) has largest biases below 850mb, but smallest biases in rest of troposphere

P5 results: Regional temperature bias (vs. CFSR)



- NH & SH (20°-80°) more similar to each other (and to global) than tropics (±20°)
  - Biggest differences between NH and SH are below 700mb: makes sense as NH has much more land than SH
- Tropics do reveal some interesting features:
  - 500mb is locally cooler in all experiments
  - GF_5 is remarkably similar to UFS_P5

P5 results: Global Q bias (CFSR)

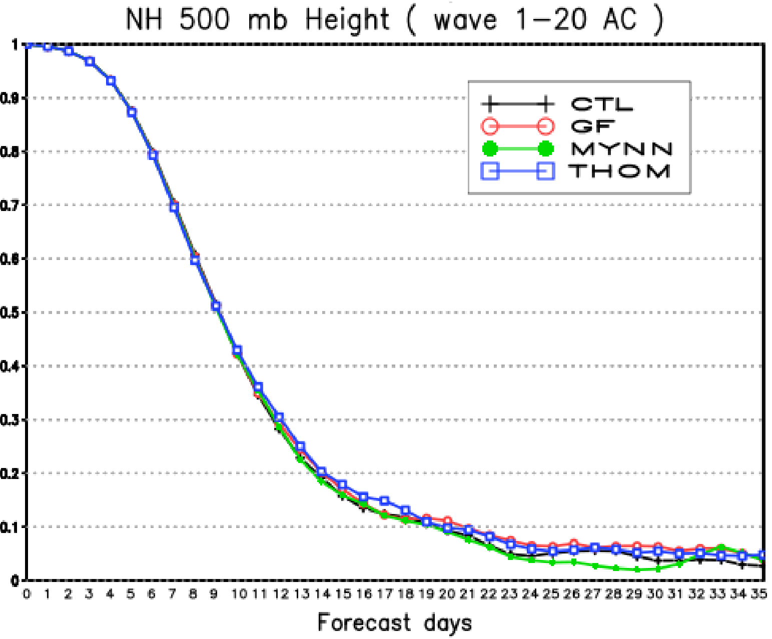


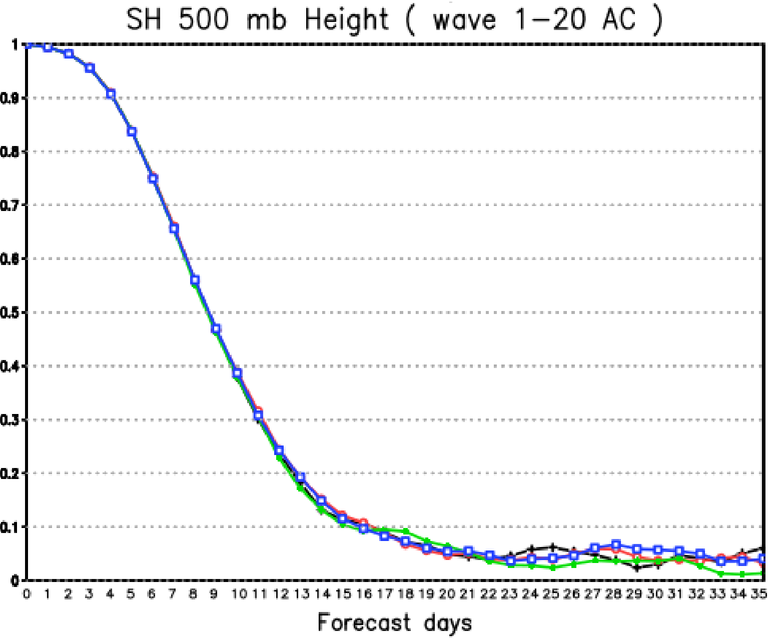
- Troposphere has nearly monotonic moisture increase w/ increasing lead time
  - Exception: MYNN_5 has nearly constant Q bias
  - In general, could **leverage shorter (1-2 week) runs to get insight into subseasonal biases**
- MYNN_5 (2020 version) most different from UFS_P5 below ~700mb
- THOM_5 most similar to UFS_P5 throughout
- GF_5 has more moisture throughout (recall GF_5 has less precipitation than UFS_P5)

Summary: P5 biases

- For many fields, regardless of which UFS experiment, bias magnitude increases with increasing lead time
- Potential to discern subseasonal-length biases through shorter (1-2 week) runs
  - Fits well with UFS-based GEFSv13 paradigm: 0-35 day runs all from same model
- Not going to get into "which is better": all model components (including physics) have changed substantially since 2020 (when P5 was run)
  - Impact of changing baseline UFS prototype will be examined later
- Biases only tell part of the story: want to look now at skill scores

P5 results: Z500 anomaly correlations

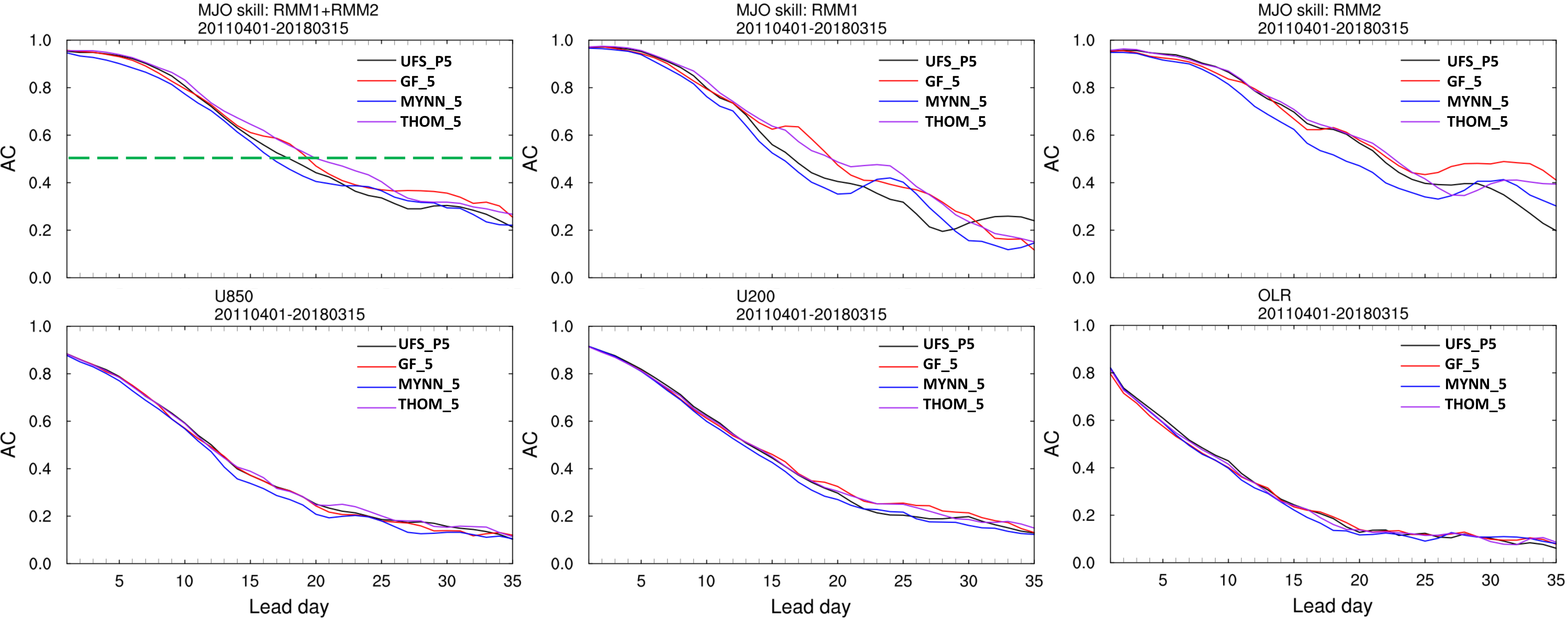




- Not a common subseasonal metric, but important for medium-range weather (UFS-based GEFSv13 will cover days 0-35+)

- Differences **not** statistically significant:
  - Exception: THOM_5 worse than UFS_P5 in NH for 120-192h

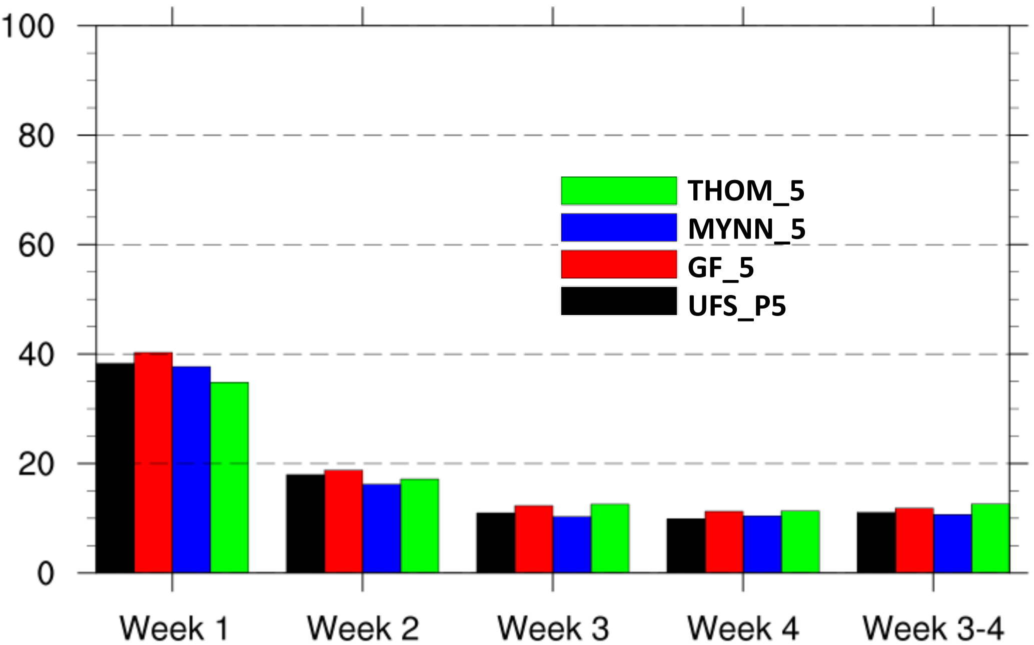
P5 results: MJO skill

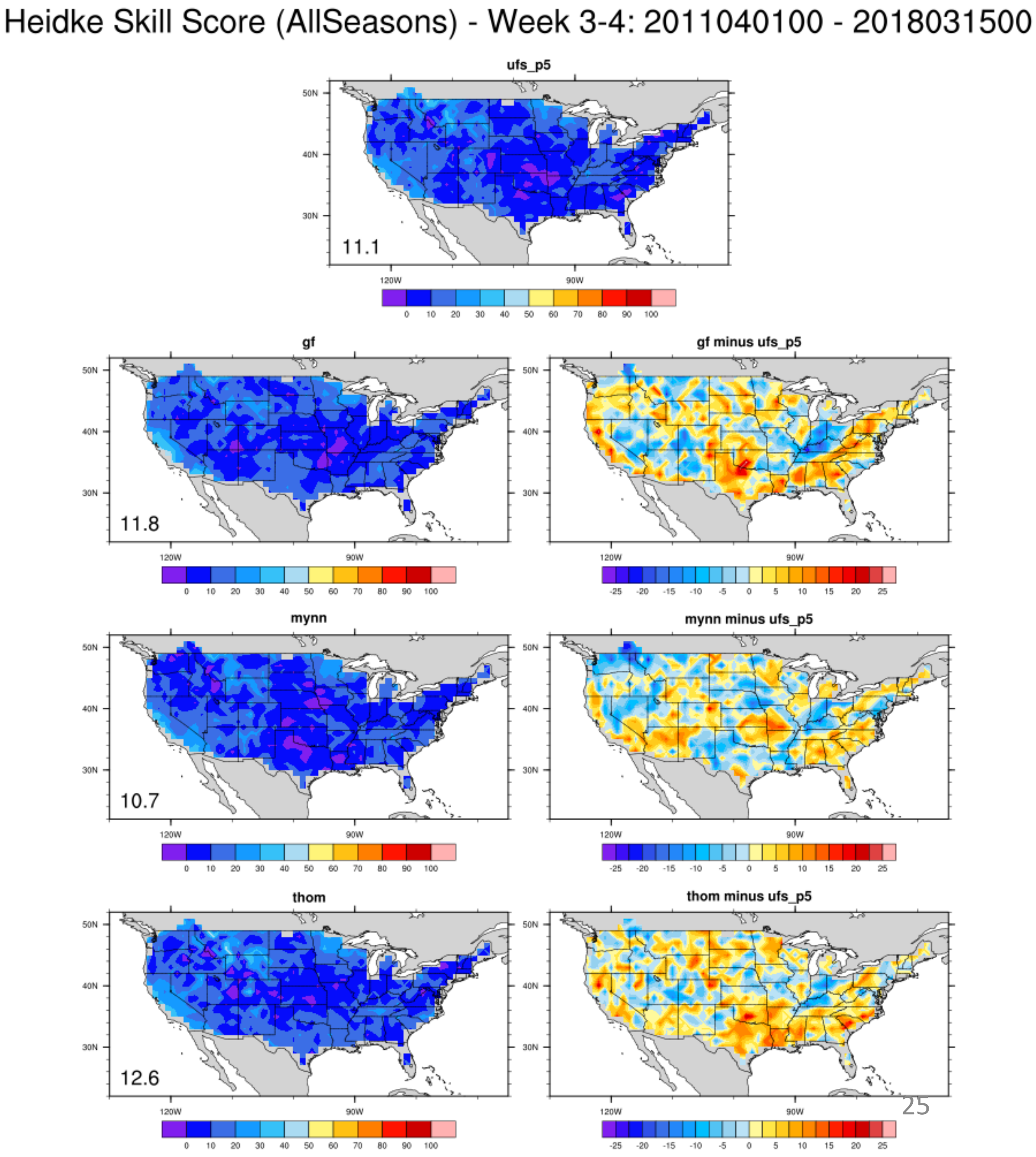


- Differences not statistically significant
- MYNN_5 appears less skillful due to:
  - Worse RMM2 skill (also somewhat RMM1)
  - Worse U850 skill: 850mb is closer to PBL, so not surprising this experiment differs there

P5 results: QPF skill (HSS) over CONUS

- Aggregated over whole year
- No one scheme is consistently better/worse than any other
- By weeks 3+4, all experiments have areas of HSS < 0 (worse than random)



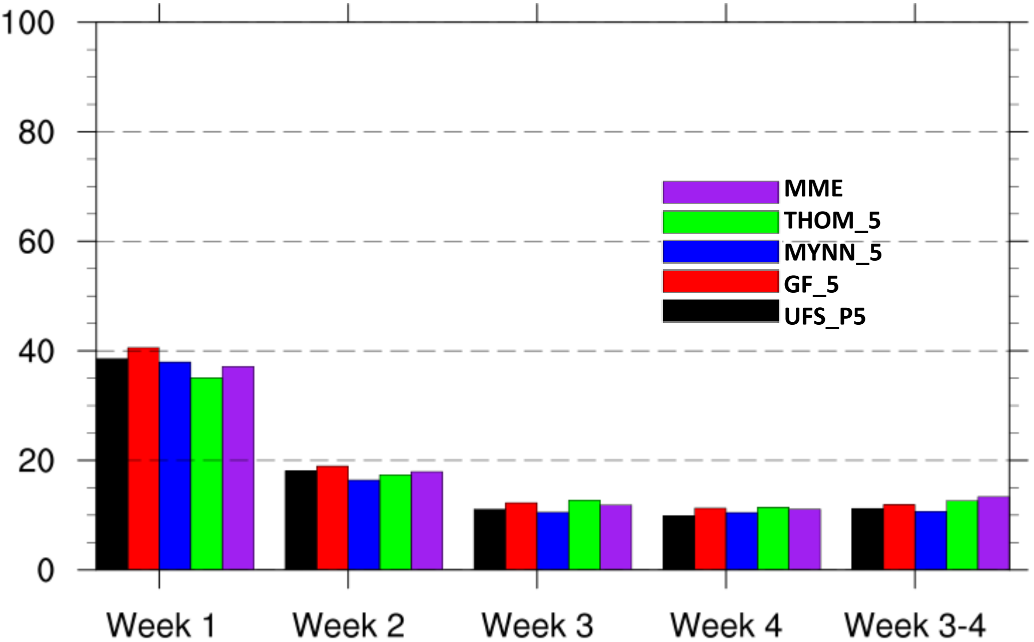


P5 results: QPF skill over CONUS, **adding MME**

- Previous studies have shown benefits of multi-model ensembles (MMEs) including multi-physics ensembles (MPEs) for subseasonal prediction

- Here, only a research curiosity: **not operationally feasible**

- Interestingly, MME (MPE) only does better – slightly – in combined weeks 3+4



P5 results: summary

- For several different fields, mean state biases become established in the first 1-2 weeks, then grow monotonically into the subseasonal
  - Good news for UFS-based GEFSv13: all global operational forecasts for days 0-35+ will be done by one single model (can infer subseasonal biases from shorter-term biases)
- UFS-based MPE not operationally feasible, didn't do better than any individual configuration in the one metric considered (QPF HSS over CONUS)
- No one experiment is uniformly better (or worse) than any other: dependencies on variable/phenomenon of interest, lead time, geographic location…
  - Are some of the more specific findings within P5 framework applicable to other UFS coupled subseasonal prototypes?

Impact of changing baseline: P5 → P7

- UFS development from partners throughout R2O funnel: good, but rapid pace!
- Example: Want to test 2 radiation schemes: Scheme R1 against Scheme R2
  - During testing, others may be changing radiation schemes (R1 and/or R2), land surface schemes, atmospheric dynamics, air/ocean coupling, etc… all are interconnected!
  - How applicable is your R1 vs. R2 test when various changes have been made?

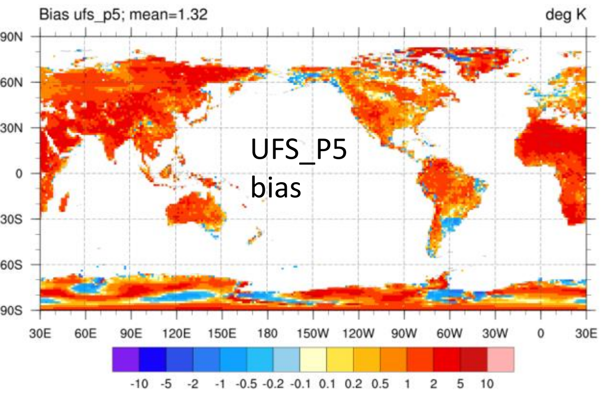
- P5 was based on UFS code from 2020; P5 runs and evaluation took almost 1 year
  - Impossible to make recommendations from “ancient” code
  - What to do? **Pivot!**

- New strategy: Repeat physics tests from P5 in P7 using “top of trunk”
  - Growing pains: Issues with MYNN PBL and Thompson MP in P7, so only GF vs. SASAS tested

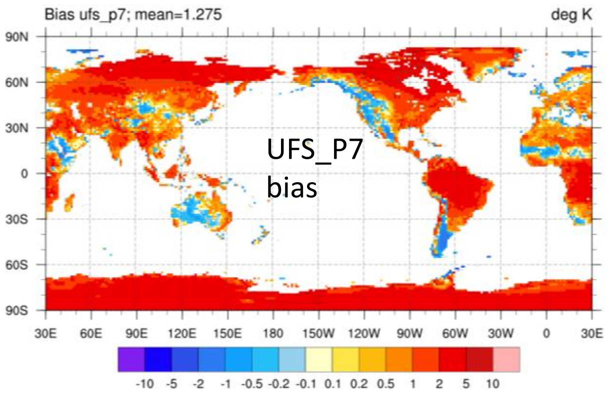
An incomplete list of changes between P5 and P7

- Timestep: 450s → 300s (increased CPU cost)
- Vertical layers: 64 → 127 (also increased CPU cost)
- Atmospheric initial conditions: CFSR → GEFSv12
- Land mask: Non-fractional → fractional
- Land model: Noah → Noah-MP
- NSST off → NSST on (NSST simulates SST diurnal cycle)

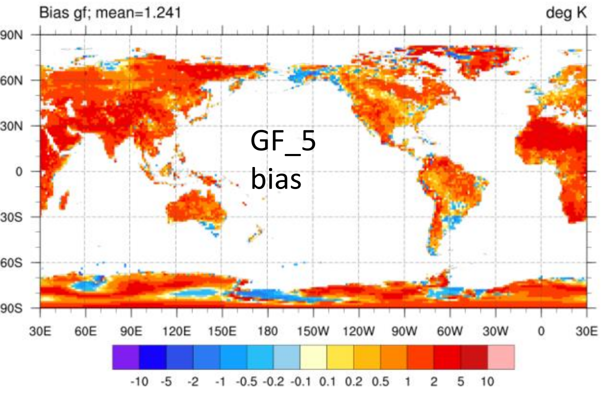
Weeks 3-4 T2m bias, SASAS vs. GF in P5 and P7

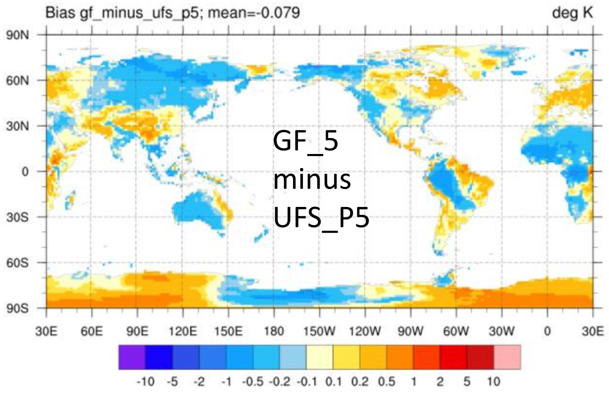


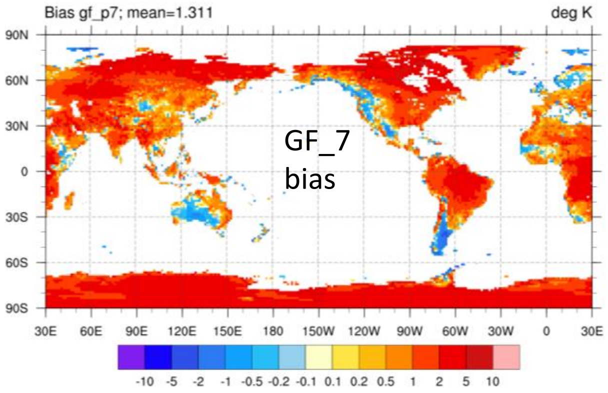
P5, verified against **CFSR**
All 168 cases

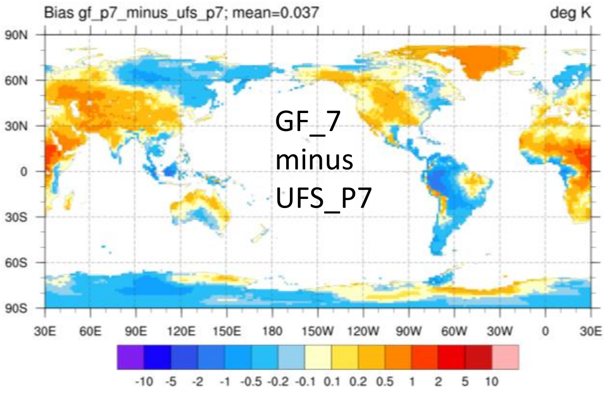


P7, verified against **GEFSv12**
All 168 cases



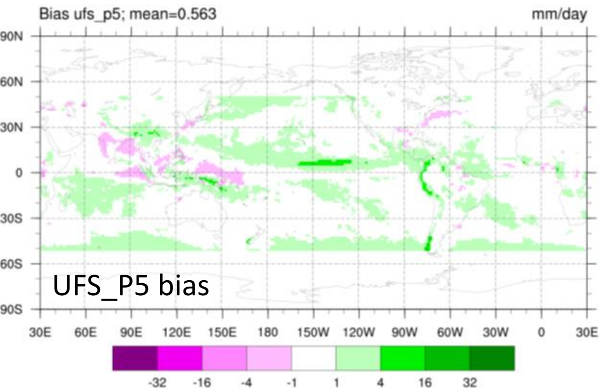




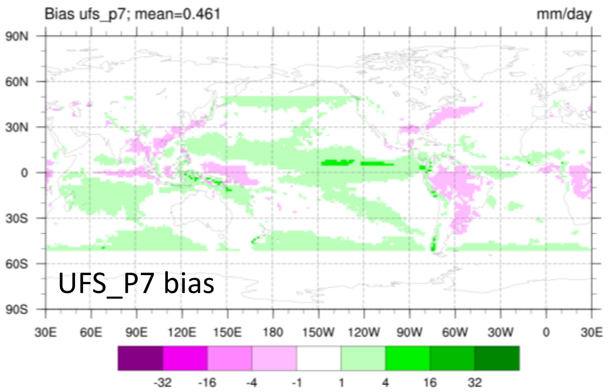


- Note: Verifying analysis changed to be consistent with updated ICs
- Impact of going from SASAS to GF (bottom right panels):
  - Consistencies (both P5 and P7): Cooling over Siberia; warming over China
  - Inconsistencies: Everywhere else, including global mean (GF cooler in P5, warmer in P7)

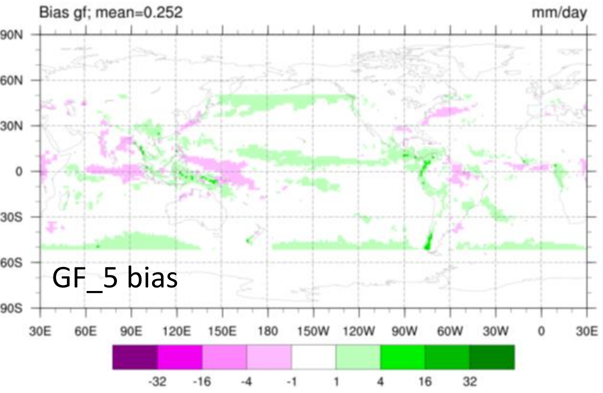
Weeks 3-4 QPF bias, SASAS vs. GF in P5 and P7

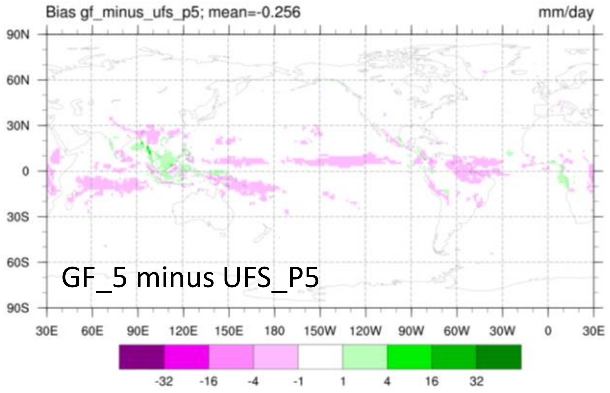


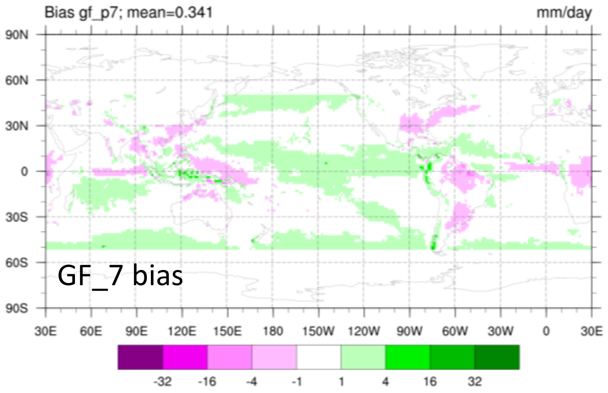
P5, verified against TRMM
All 168 cases

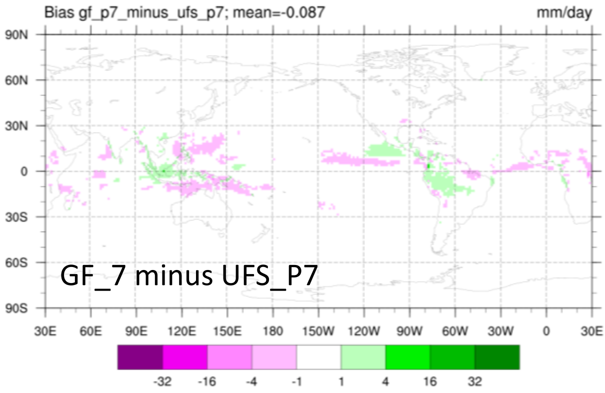


P7, verified against TRMM
All 168 cases



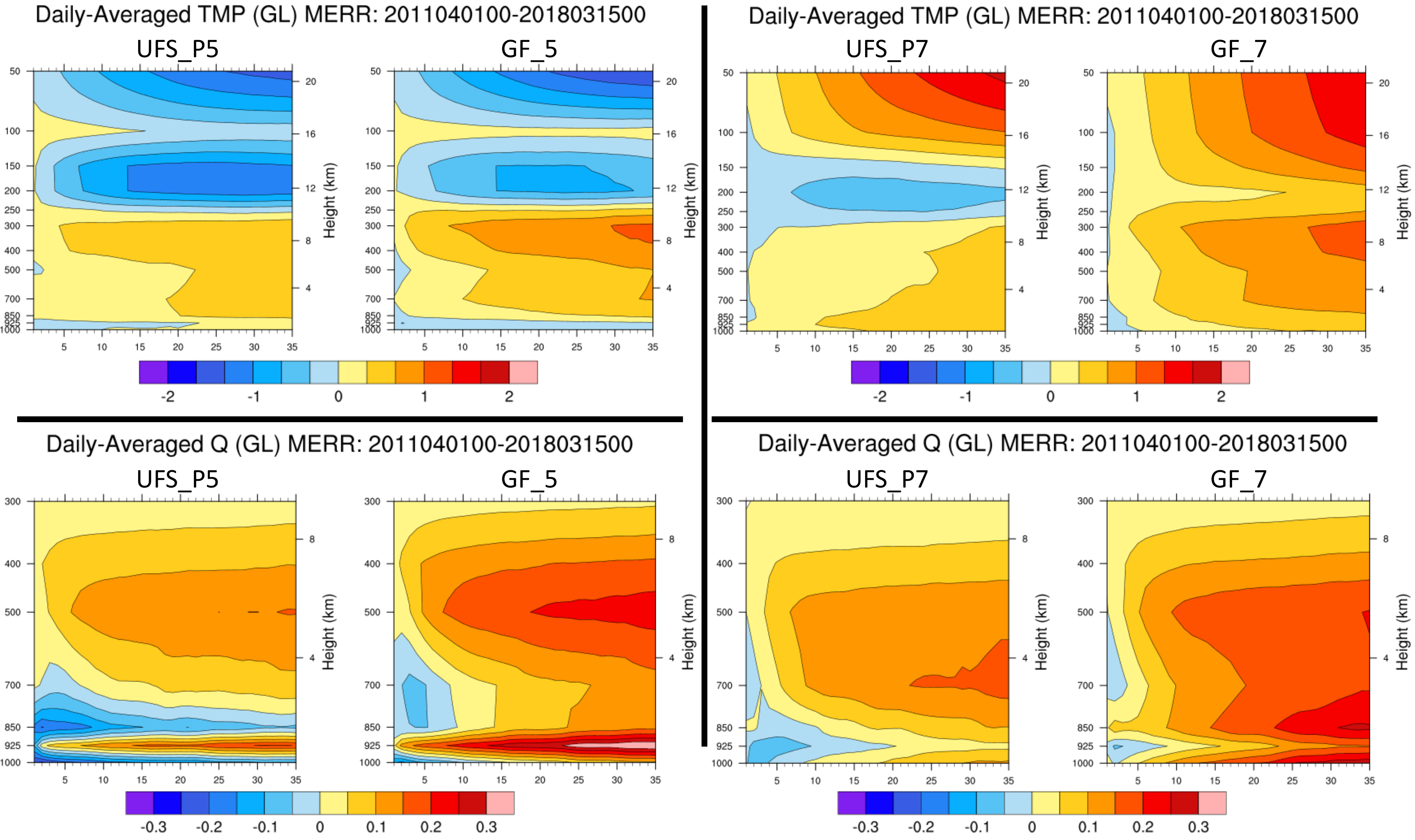






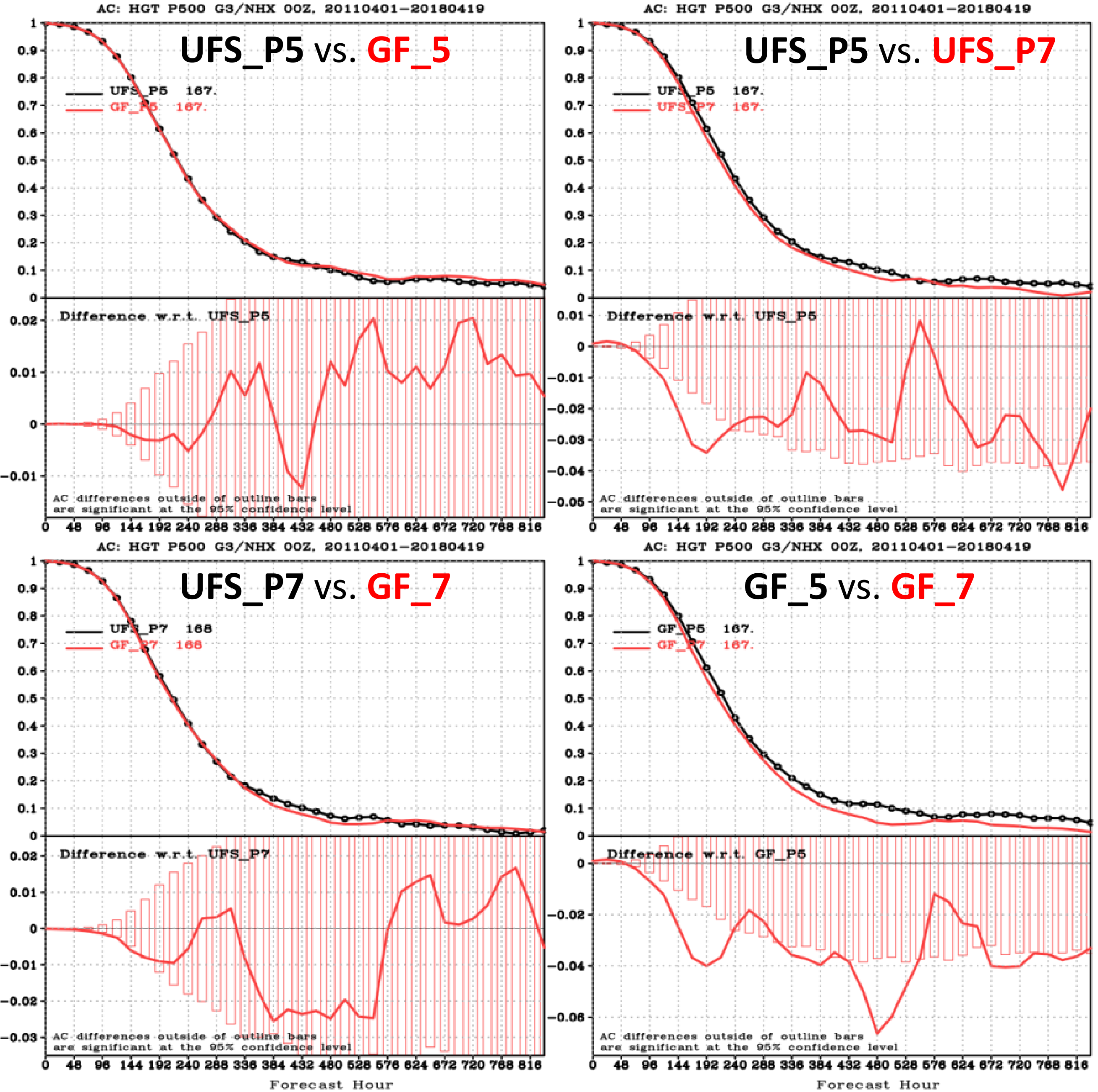
- Impact of going from SASAS to GF (bottom right panels):
  - Consistencies: Global drying trend; locally wetter near Maritime Continent
  - Inconsistencies: GF drying relative to SASAS is weakened in P7

Global bias evolution: T & Q, SASAS vs. GF in P5 & P7



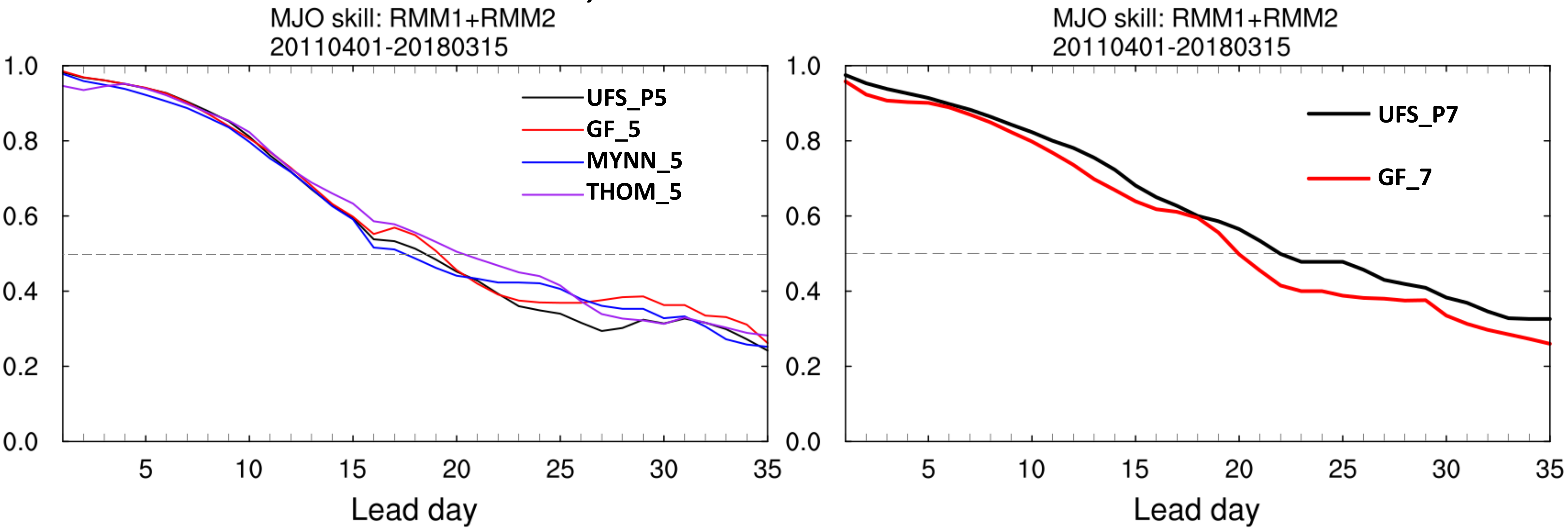
- Note: verifying analysis is CFSR for P5; GEFSv12 for P7
- Good news: in general, magnitude of biases generally increase monotonically with increasing lead time for **both P5 and P7**
- GF slightly warmer (and more moist) than SASAS in P5; substantially warmer (and more moist) than SASAS in P7
- **But:** In P5, bias structures are similar between SASAS and GF; in P7, bias structures between 2 schemes are quite different

NHX Z500 ACC: P5 vs. P7, GF vs. SASAS



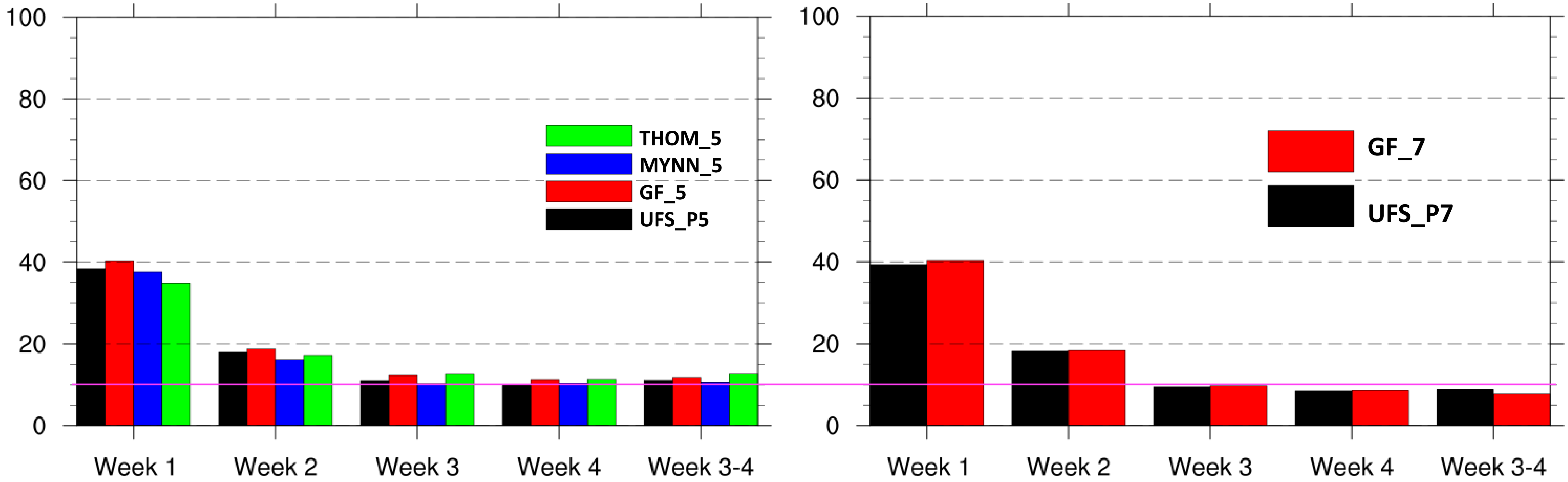
- P5 (top left): no significant difference between UFS_P5 and GF_5
- P7 (bottom left): GF_7 statistically significantly worse than UFS_P7 for first ~144 hours
- For both SASAS and GF (right panels), P7 has statistically significantly worse Z500 skill for days 3-10 compared to P5
- For NH Z500 ACC, changing convection scheme (left panels) not as important as going from P5 → P7 (right panels)

MJO RMM1+2 skill, P5 vs. P7



- P5 vs. P7 comparison: GF essentially unchanged, but SASAS improves by 3-4 days (using 0.5 AC threshold) from P5 to P7; both schemes improve in first ~15 days

Precip HSS, CONUS, year-round, P5 vs. P7



- P5 vs. P7 comparison: weeks 1-2 very little change, but both SASAS and GF are worse in P7 for week 3 and beyond
- GF typically better than SASAS, but *barely*

Summary: Impact of changing baseline, P5 → P7

- In general, bias magnitudes increase monotonically with increasing lead time for **both P5 and P7**

- However: specific details on relative scheme performance (SASAS vs. GF) are not necessarily consistent between P5 and P7
  - Some combination of changes to schemes themselves, plus changes to other components of coupled UFS between P5 and P7
  - Key message: Use caution when trying to generalize performance of a physics scheme

- Next: a highly controlled experiment that isolates the impact of switching from GFDL MP to Thompson MP in the context of moving from P7 to P8

P7 → P8: Isolating impact of microphysics change

- GFDL MP used as default for coupled UFS prototypes 1-7

- P8: Switch to Thompson MP **based on medium-range weather results** (NOT subseasonal results)

- To isolate impact of MP change from all other changes between P7 and P8, we ran a parallel P8, but with GFDL MP ("P8_GM")

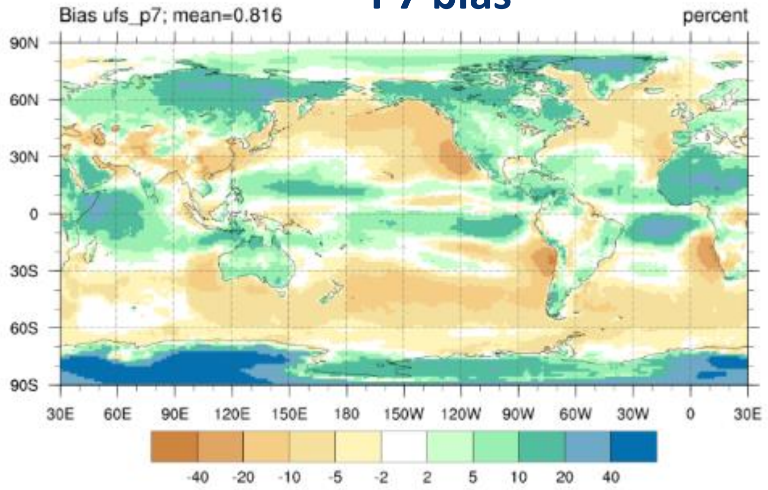
- Next slide: Show one example (EMC did extensive comparison of control P8 with P8_GM) of how P8_GM can be leveraged to discern whether changed results between P7 and P8 are (or are not) attributable to changed microphysics

Annual Weeks 3&4

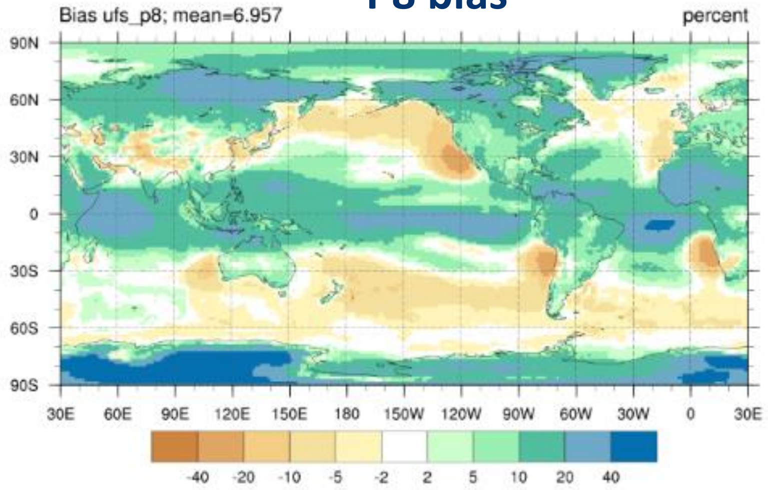
Total Cloud Cover wrt CERES

Credit: Lydia Stefanova, EMC

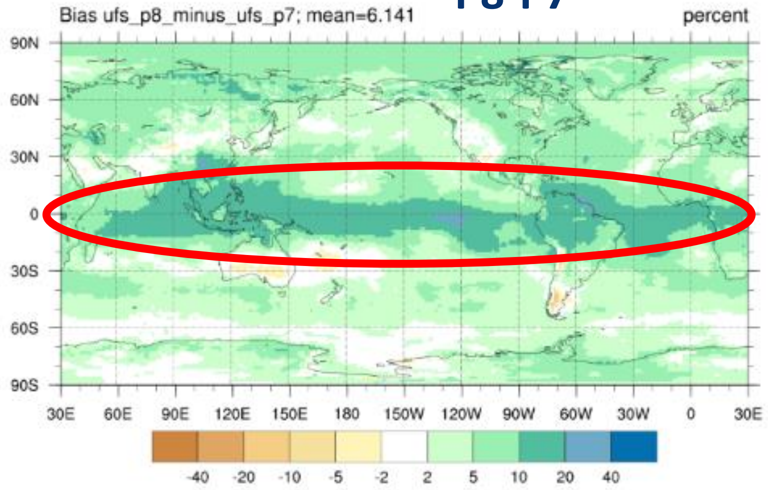
P7 bias



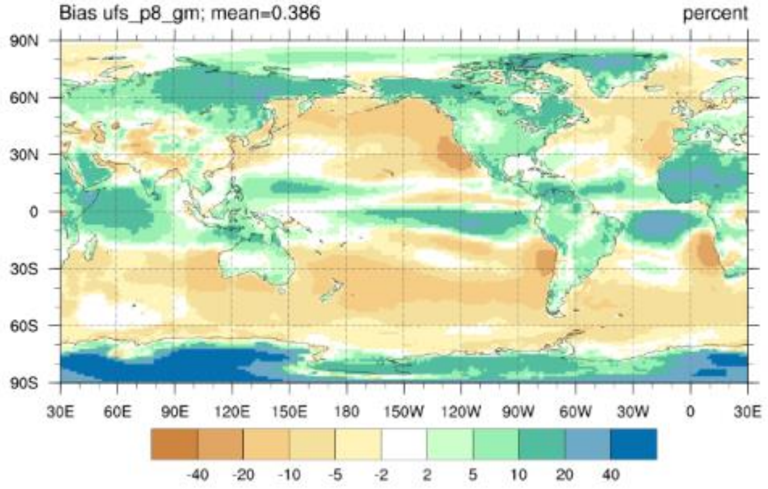
P8 bias



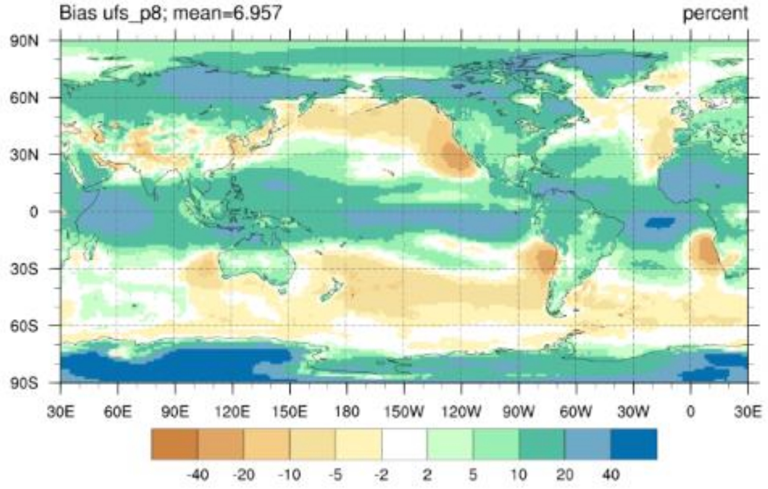
P8-P7



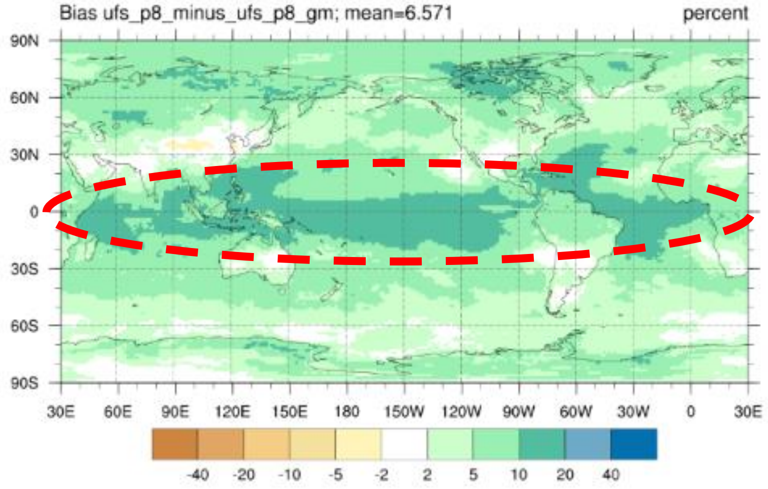
P8_GM bias



P8 bias



P8-P8_GM



P8’s global increase of cloud cover, **particularly along the equator**, is ATTRIBUTABLE to Thompson MP

Summary

- Unique opportunity to test impacts of different atmospheric physics schemes in fully-coupled, subseasonal-length UFS runs

- Biases become established within first 1-2 weeks, then grow in magnitude with lead time: **holds true for multiple coupled UFS prototype frameworks**
  - Great result for UFS-based GEFSv13 (responsible for all global forecasts days 0-35+): can discern subseasonal biases through shorter-length runs

- However, **specifics** of scheme comparisons changed between prototypes: inappropriate to say if one scheme is better (or worse) than another

Future work

- Coupled prototypes are done: EMC is moving ahead to implement UFS-based GEFSv13 to have operational global ensemble predictions from days 0-35+

- NOAA is beginning UFS-based **Seasonal Forecast System** (SFS) development
  - SFS (1-year forecasts) planned to replace operational CFSv2 (9-month forecasts) in FY26
  - GSL **already working to make substantial contributions UFS-based SFS** with other (NOAA) partners

- All subseasonal experiments shown here are available upon request (ben.green@noaa.gov): These datasets are quite valuable for physics comparisons using advanced diagnostics

Future work

- Coupled prototypes are done: EMC is moving ahead to implement UFS-based GEFSv13 to have operational global ensemble predictions from days 0-35+

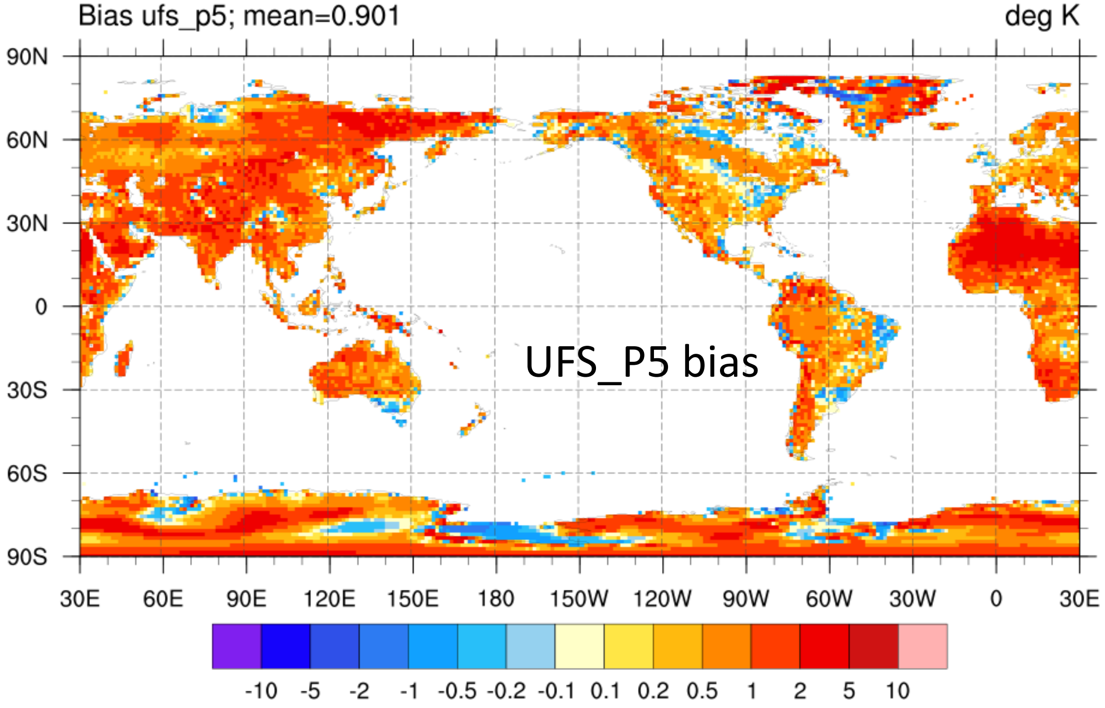
- NOAA is beginning UFS-based **Seasonal Forecast System** (SFS) development
  - SFS (1-year forecasts) planned to replace operational CFSv2 (9-month forecasts) in FY26
  - GSL **already working to make substantial contributions UFS-based SFS** with other (NOAA) partners

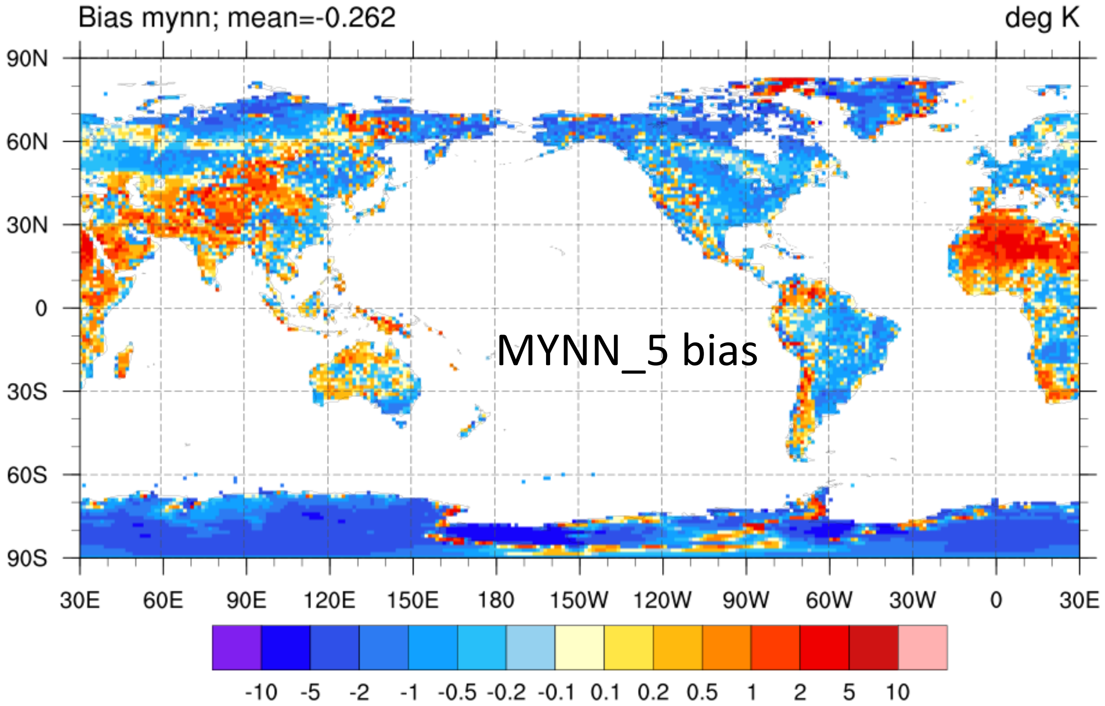
- All subseasonal experiments shown here are available upon request (ben.green@noaa.gov): These datasets are quite valuable for physics comparisons using advanced diagnostics

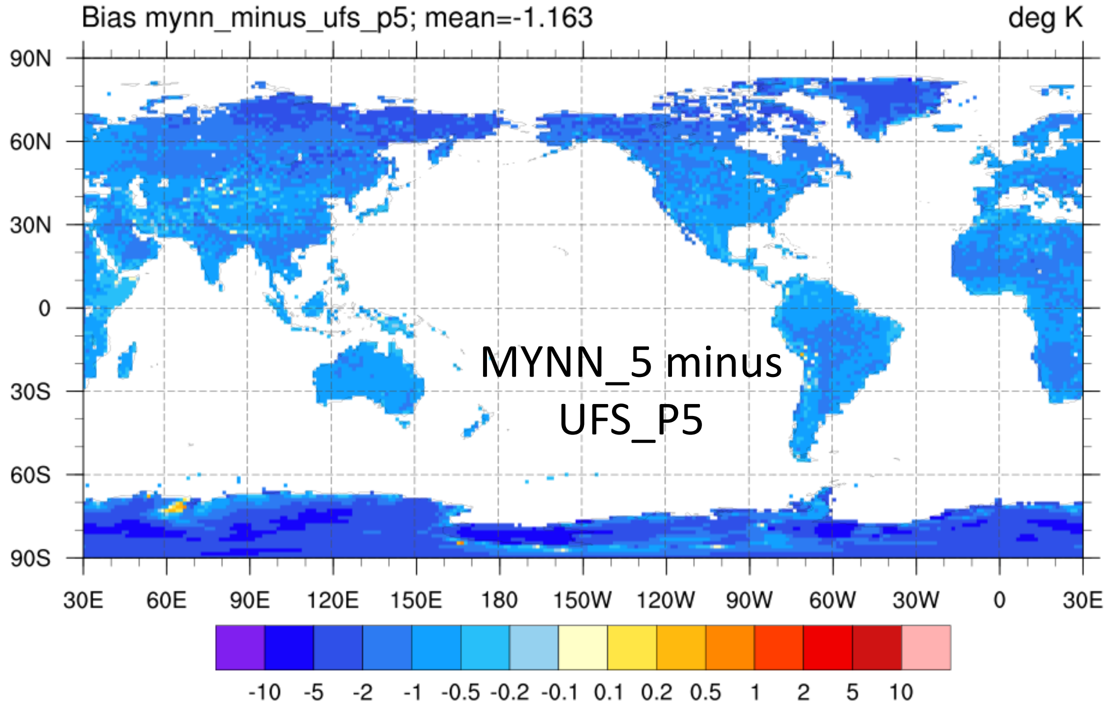
Extra slides

P5 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)

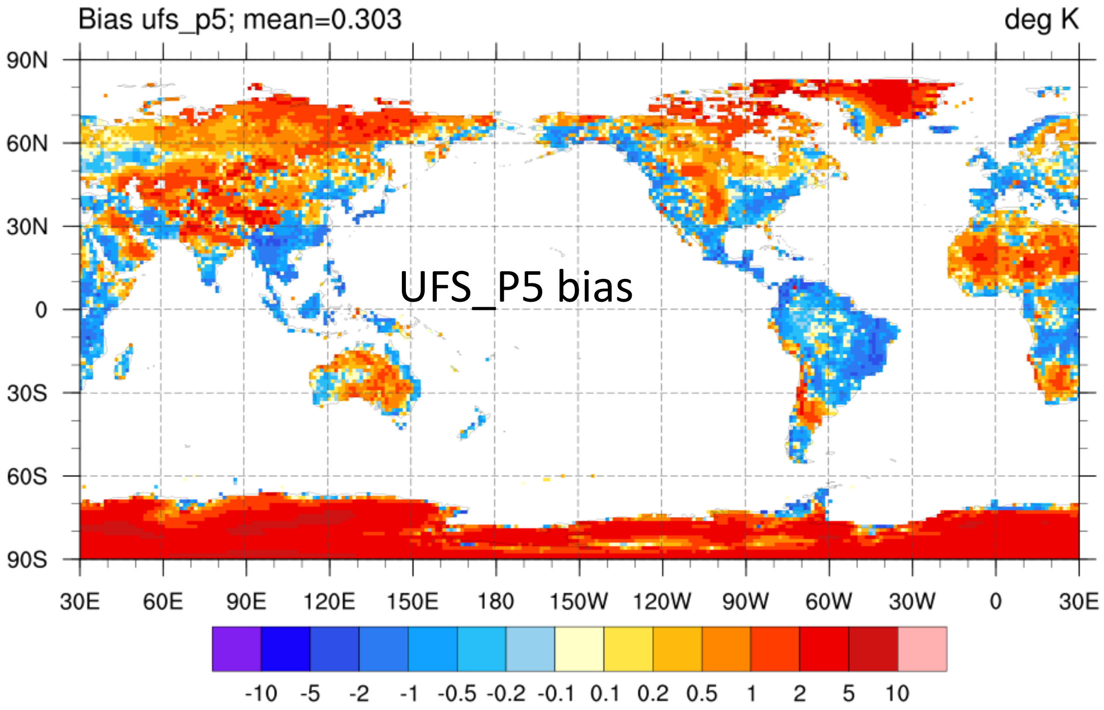


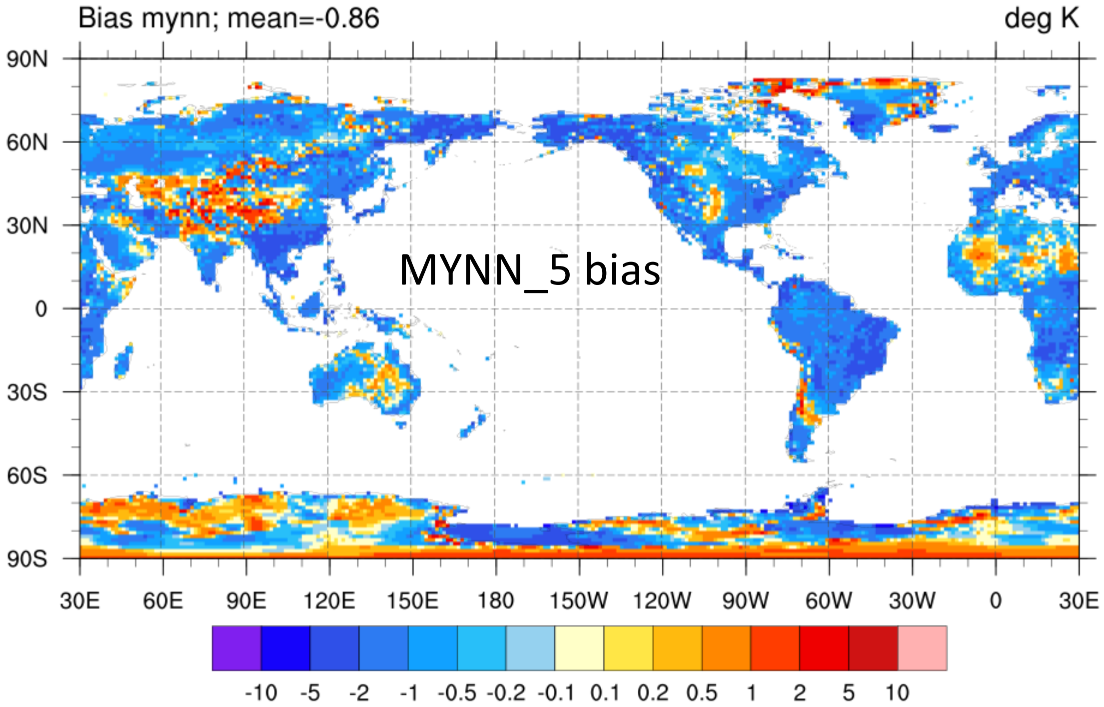


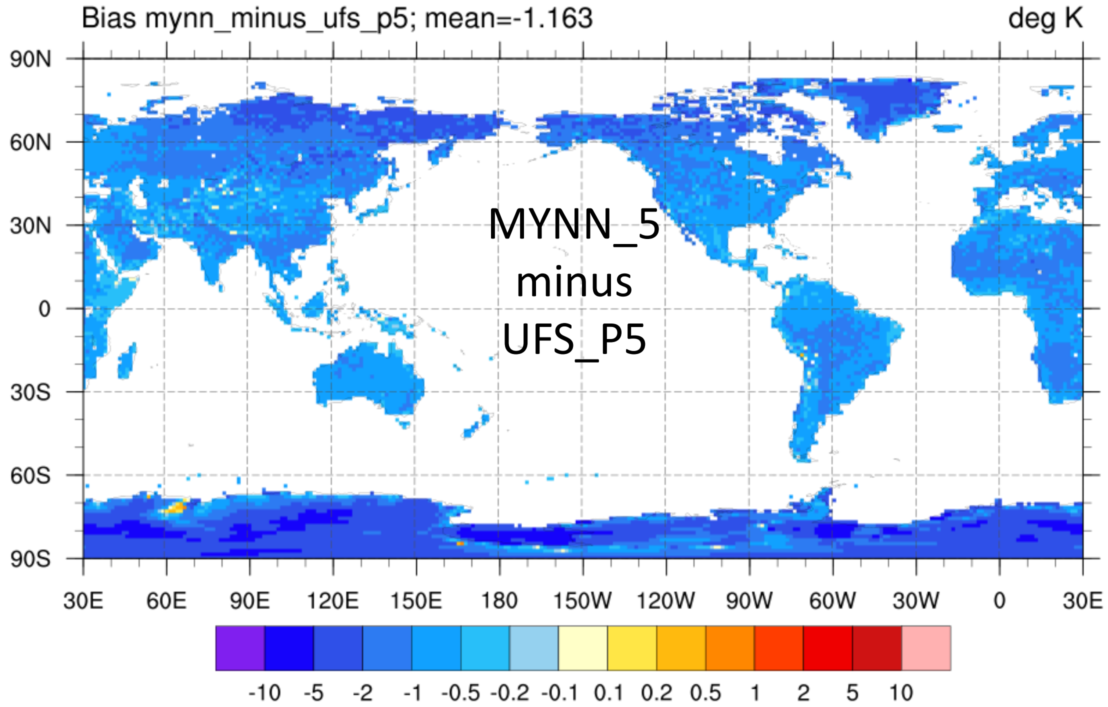


P5 results: T2m bias

- Land biases (left) using **ERA5** as truth
- Week **1**, UFS_P5 vs. MYNN_5: 167 cases
- Systematic cooling in MYNN_5 relative to UFS_P5 (bottom right)

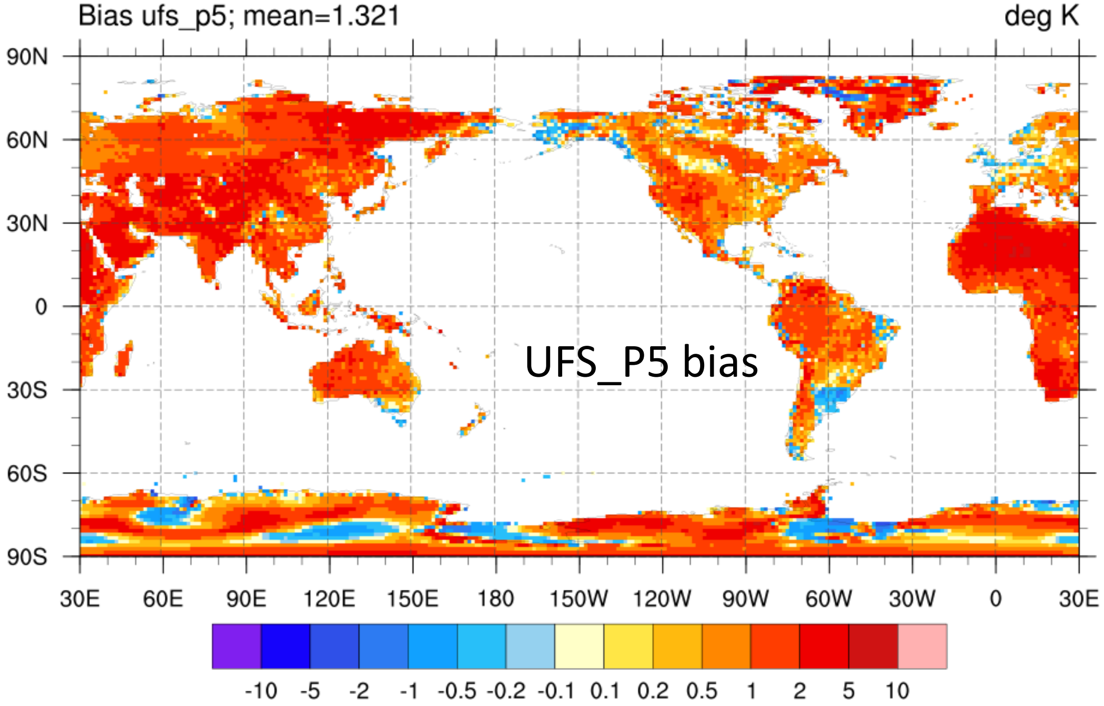


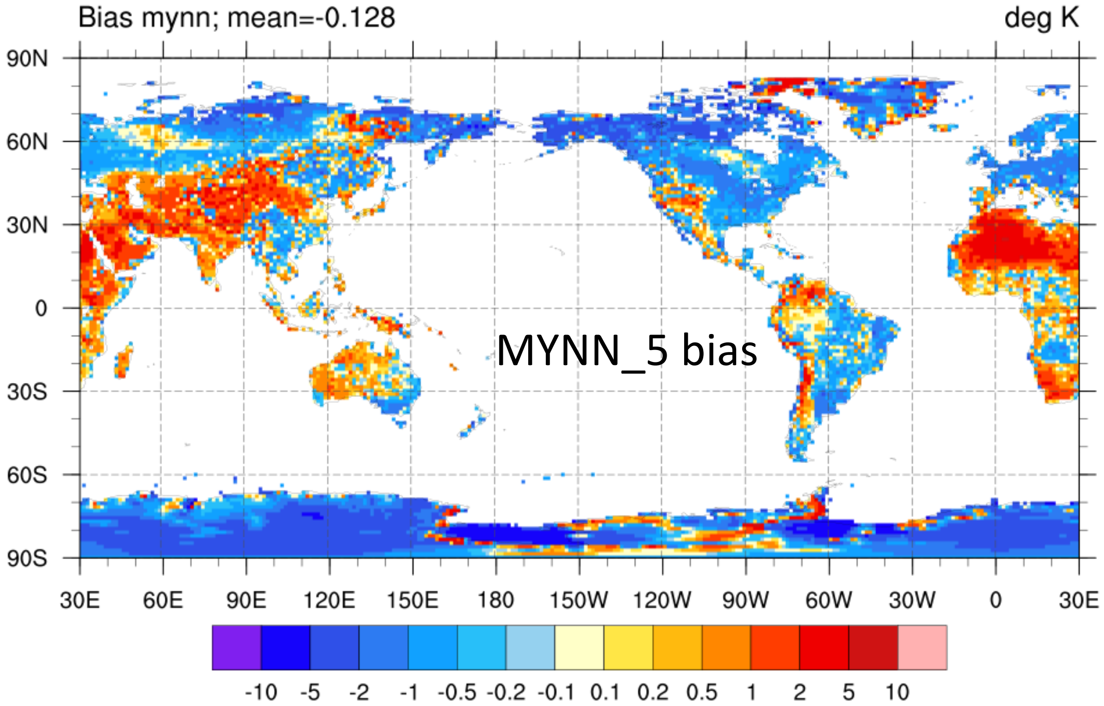


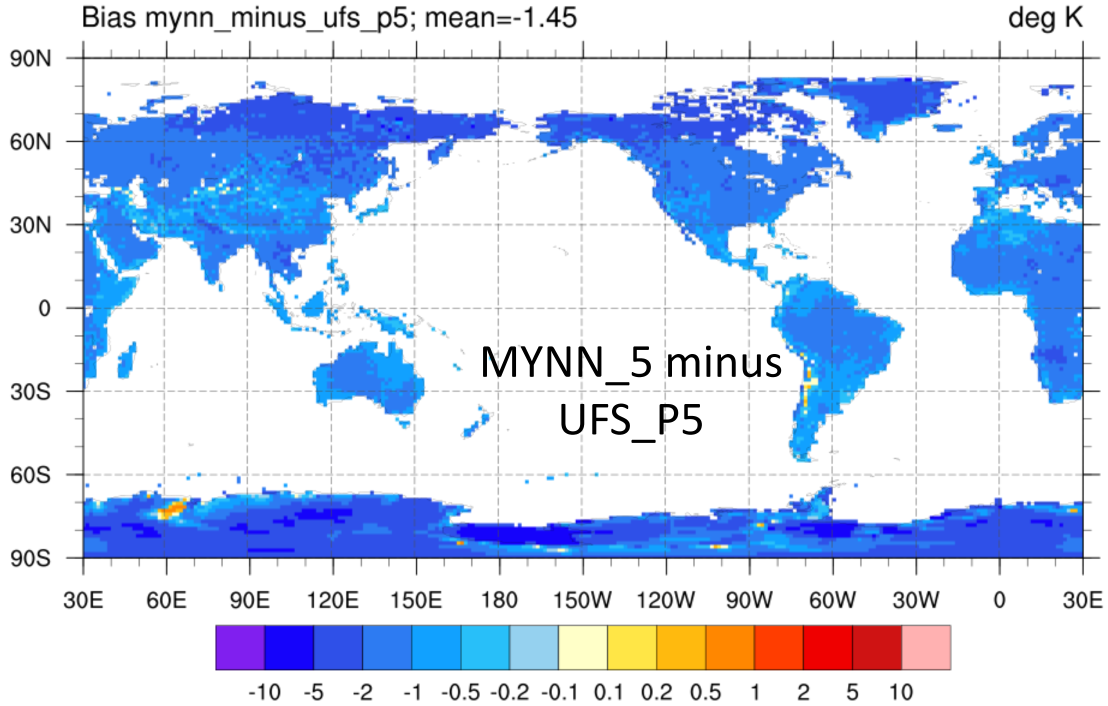


P5 results: T2m bias

- 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)

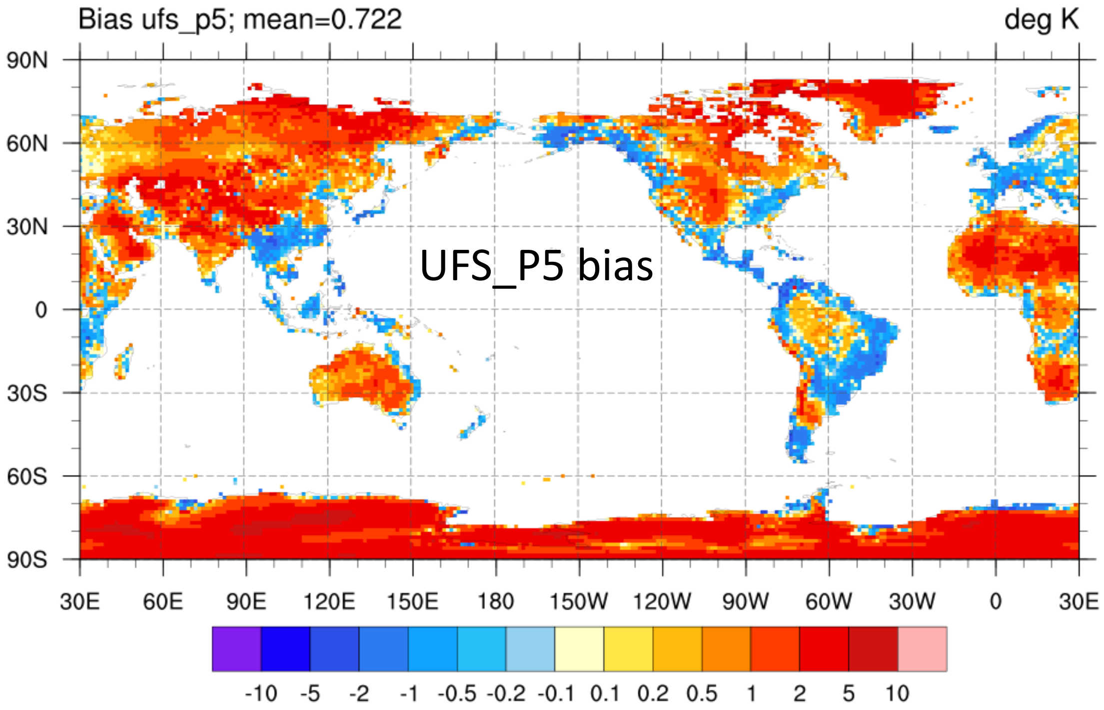


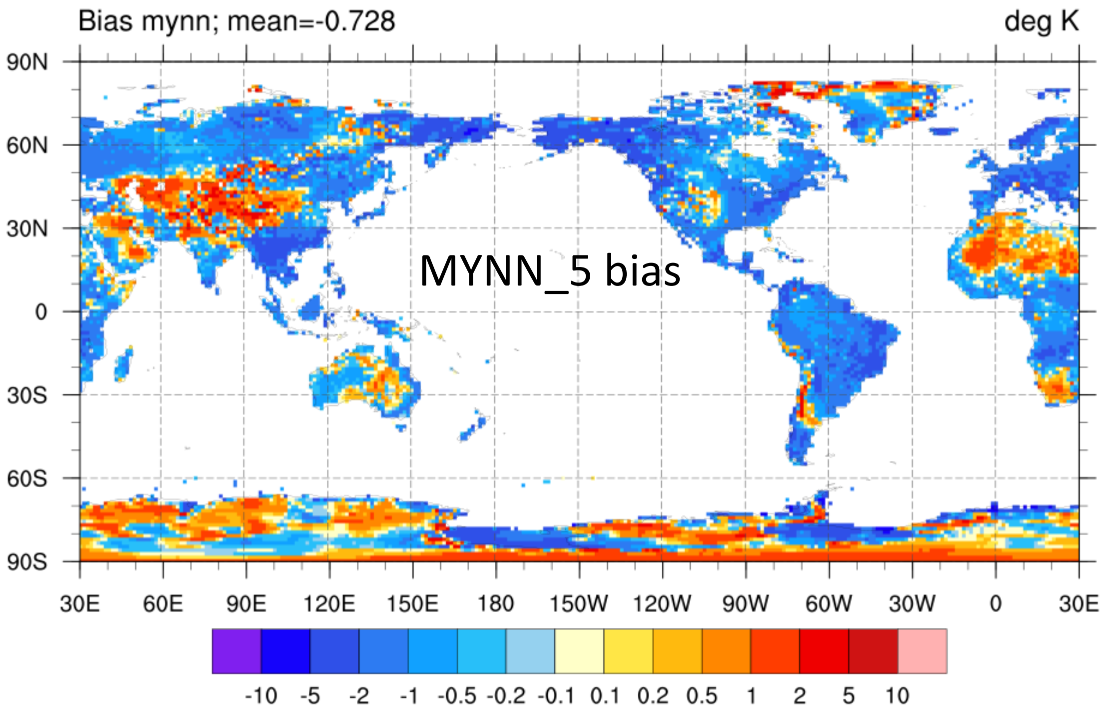


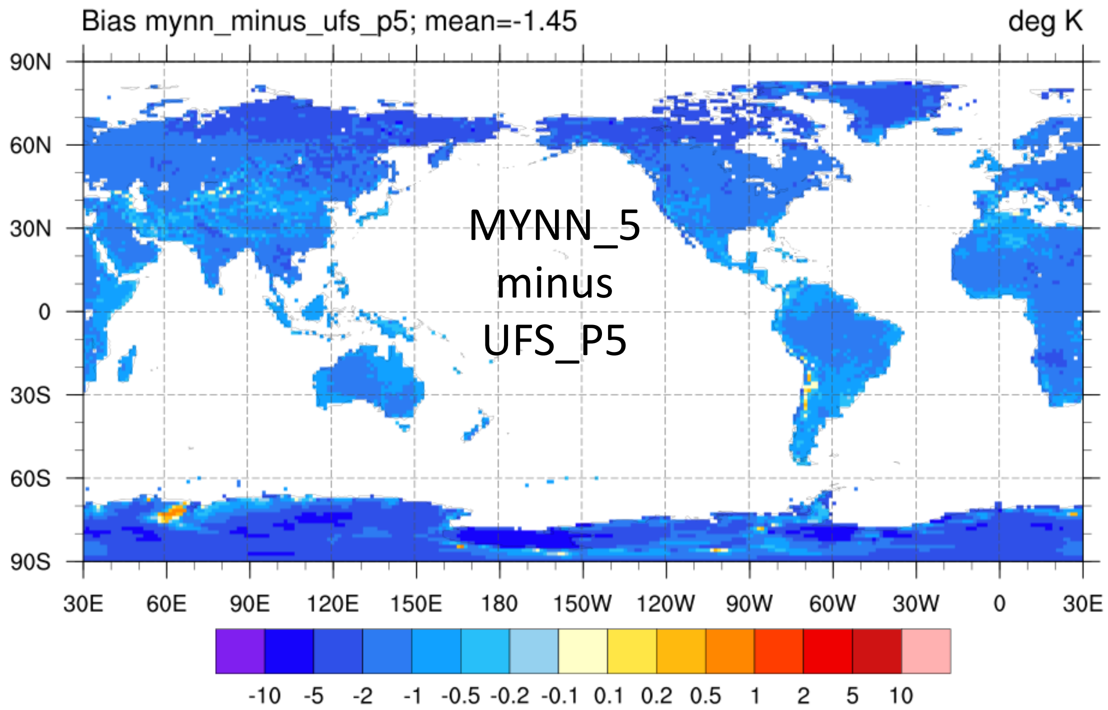


P5 results: T2m bias

- Land biases (left) using **ERA5** as truth
- Weeks **3-4**, UFS_P5 vs. MYNN_5: 167 cases
- Patterns extremely similar to week 1!
- Systematic cooling in MYNN_5 relative to UFS_P5 (bottom right)

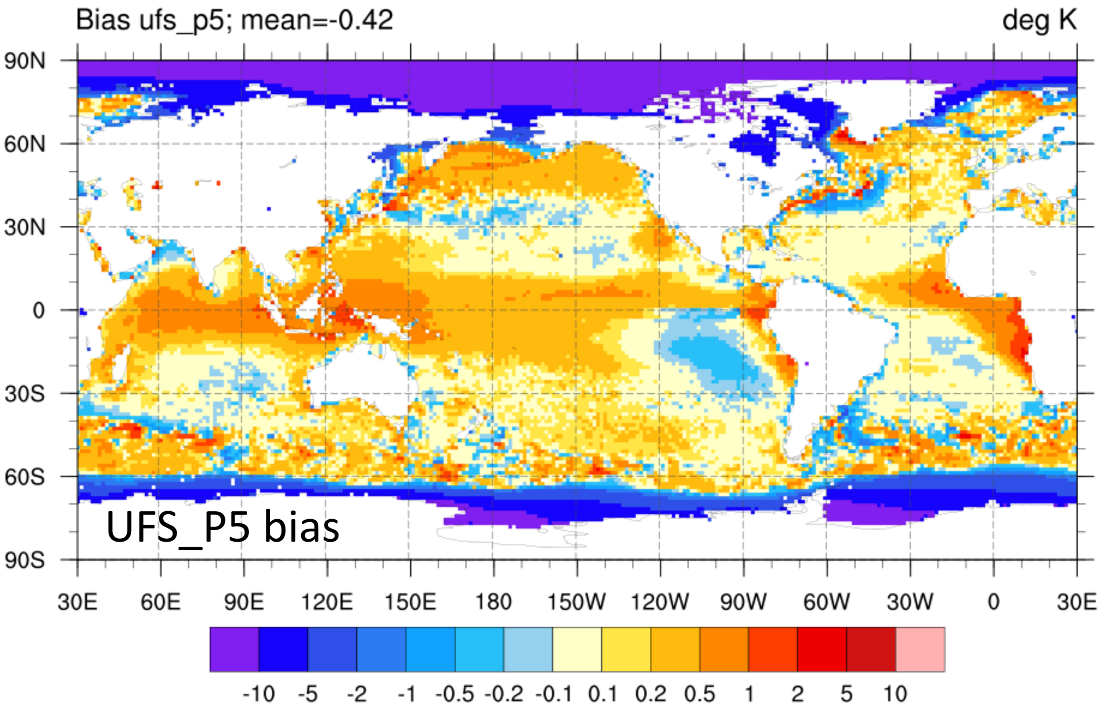


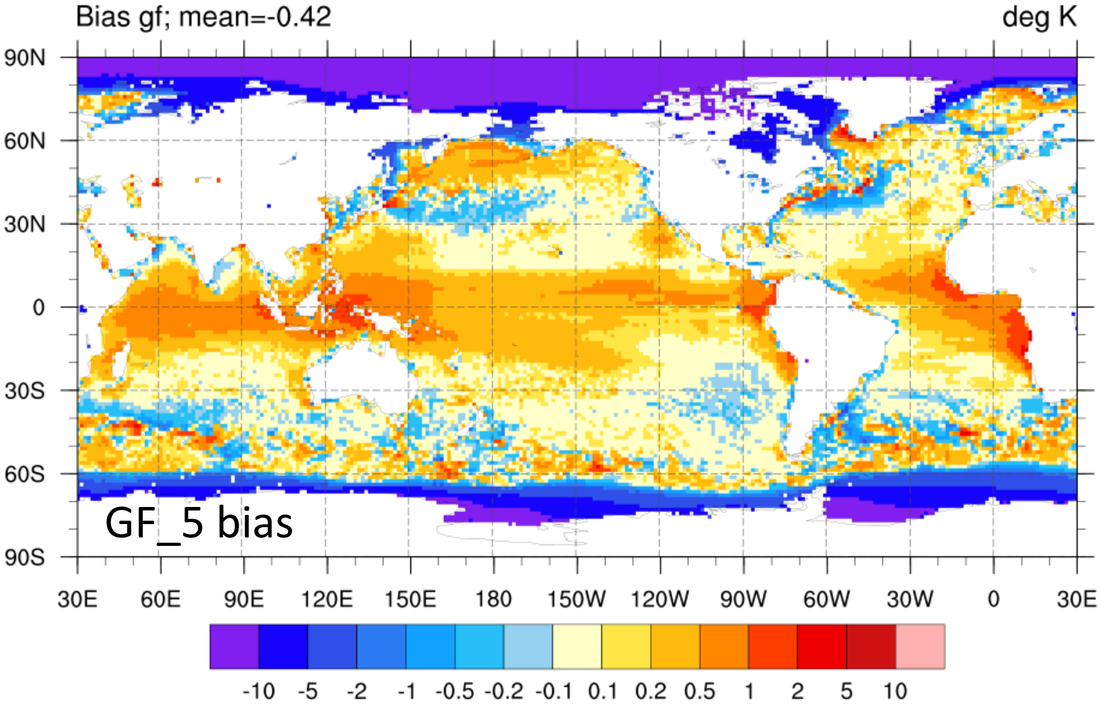


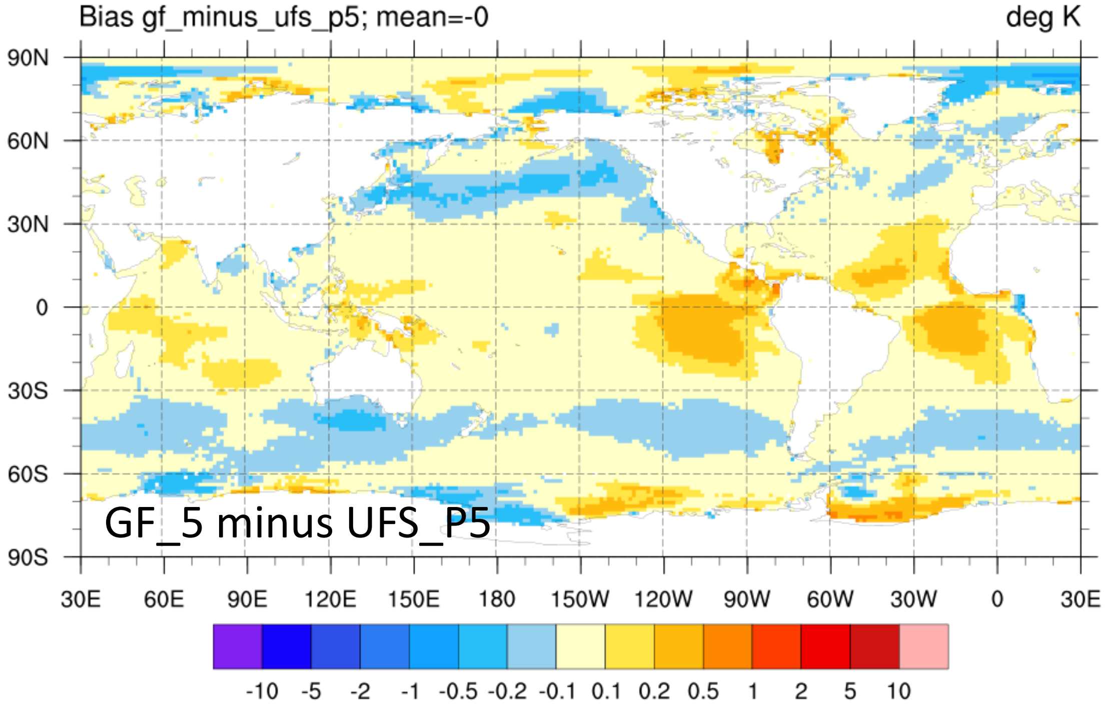


P5 results: SST bias

- Using OSTIA as truth
- Weeks **3-4**, UFS_P5 vs. GF_5: 168 cases
- GF_5 slightly cooler in mid-latitudes, slightly warmer off western coasts of Africa and S. America

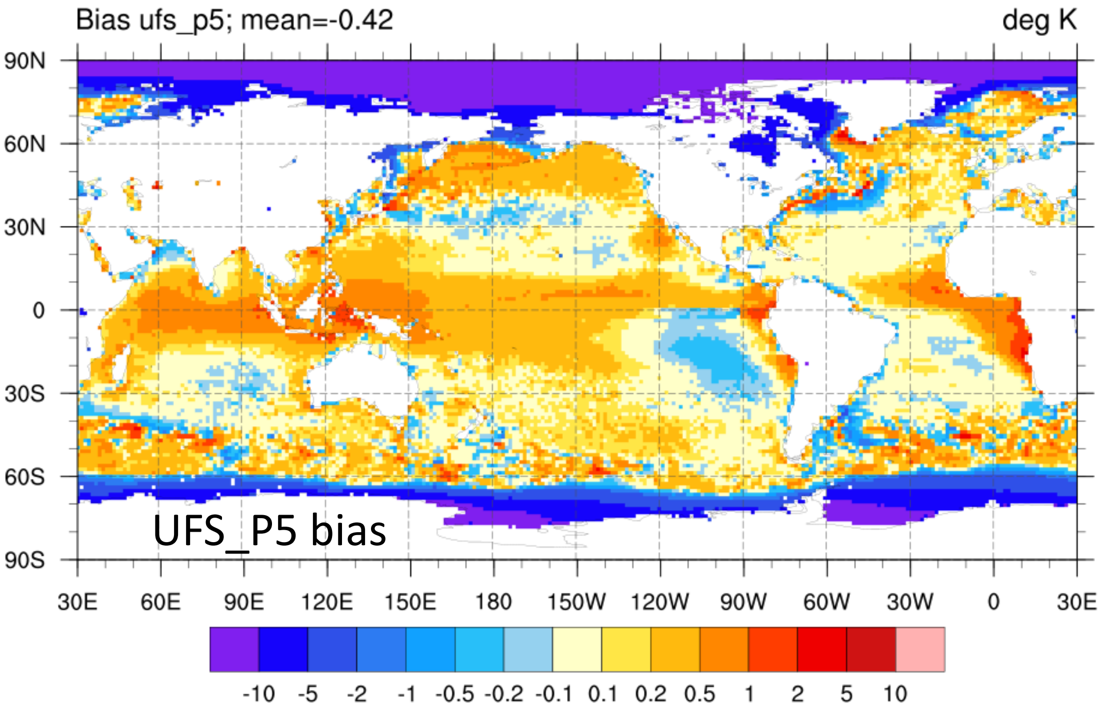


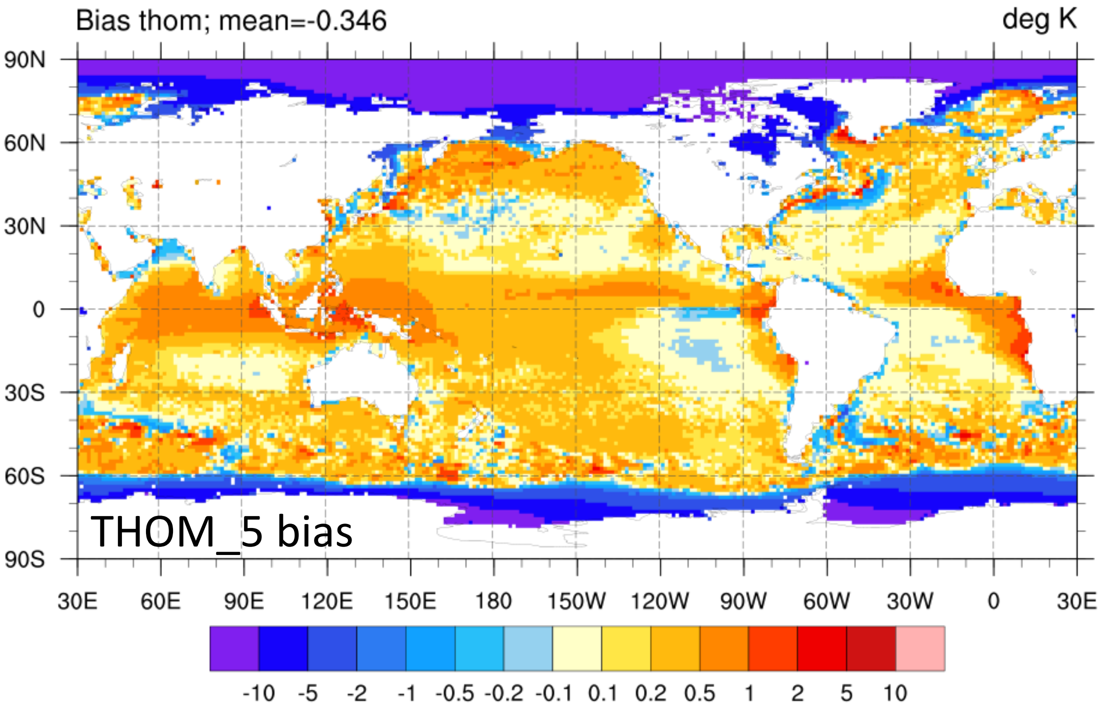


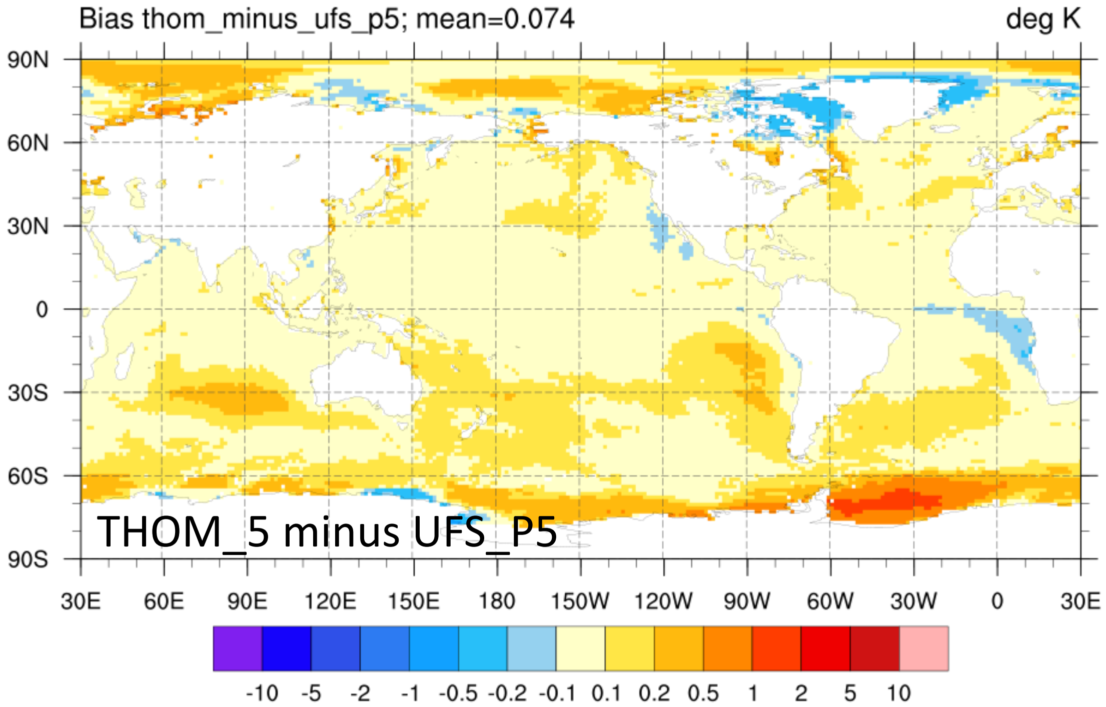


P5 results: SST bias

- Using OSTIA as truth
- Weeks **3-4**, UFS_P5 vs. THOM_5: 166 cases
- THOM_5 slightly warmer off west coasts of S. America and Australia

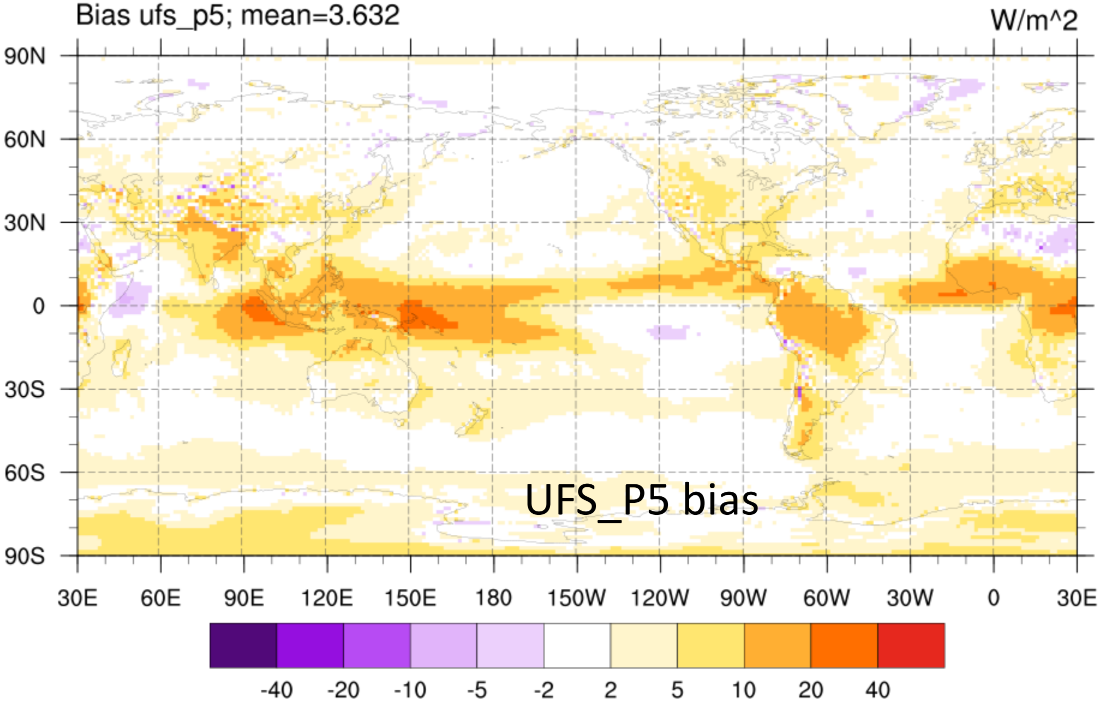


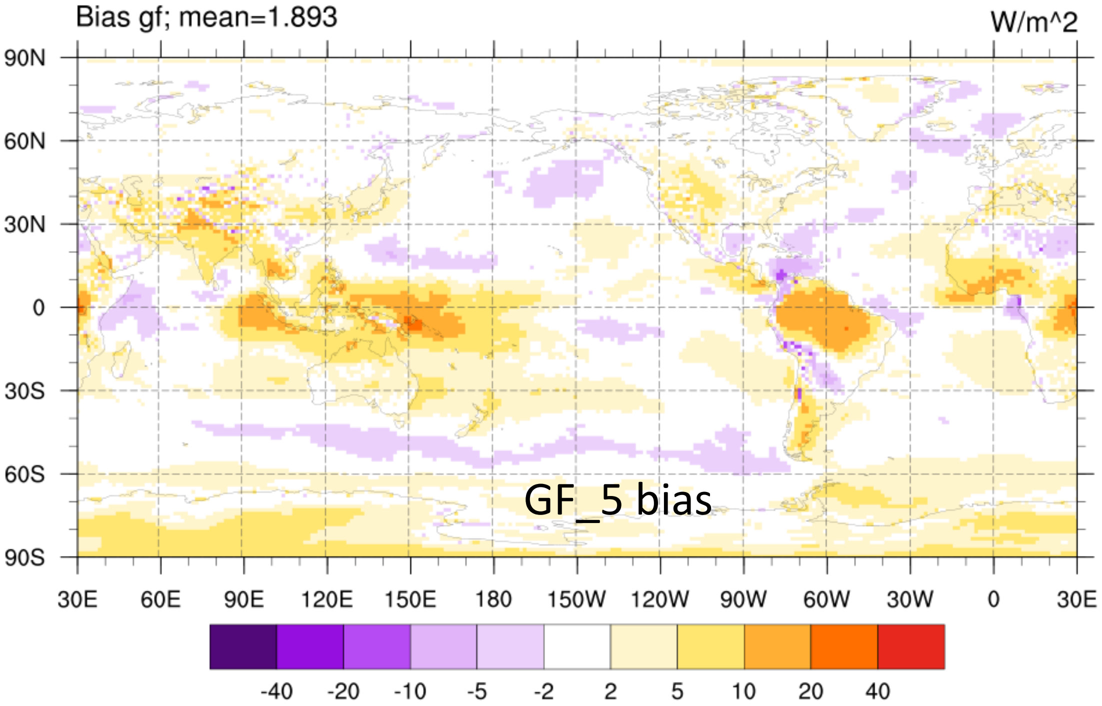


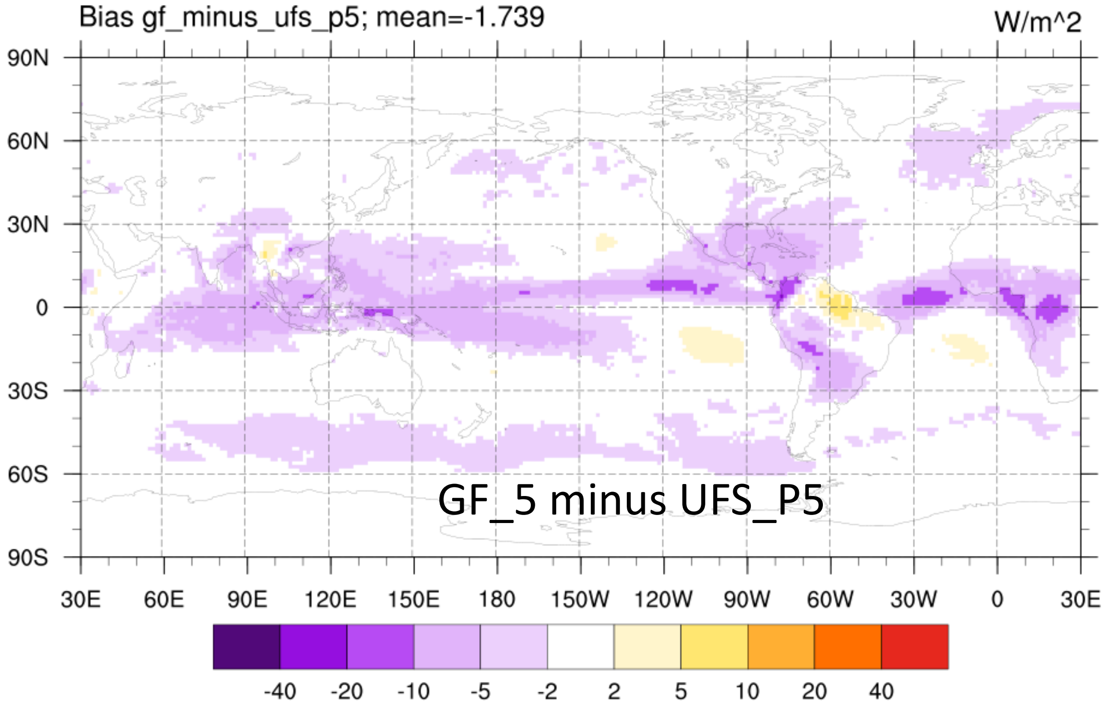


P5 results: OLR bias

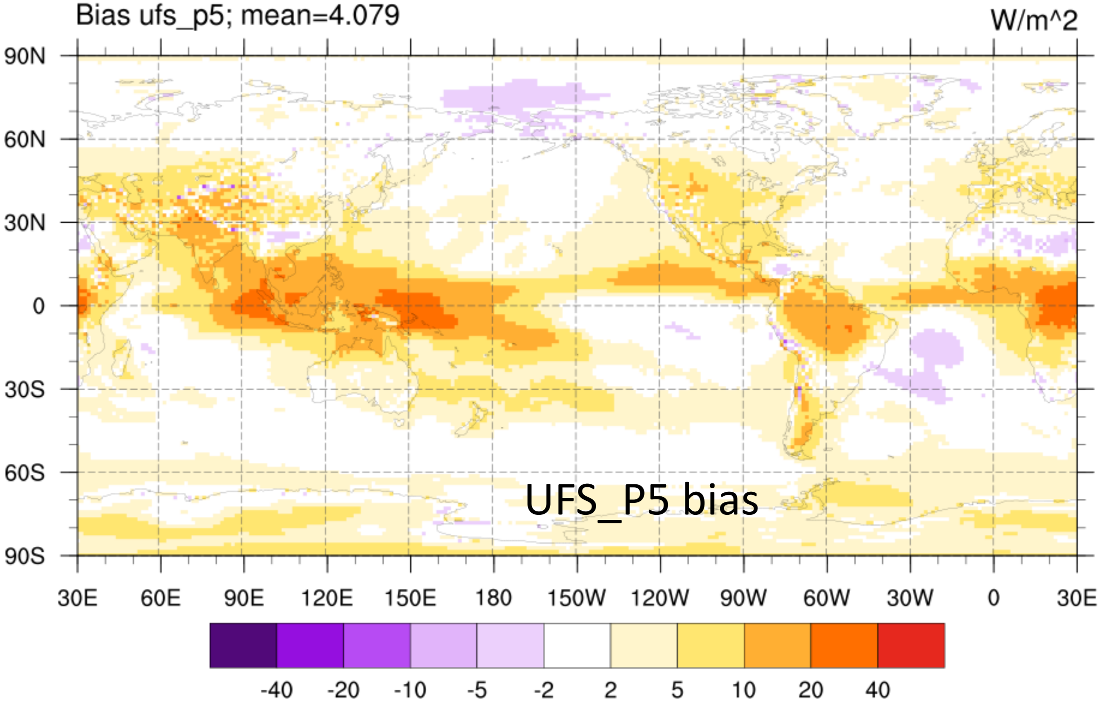
- Using HRIS as truth
- Week **1**, UFS_P5 vs. GF_5: 168 cases
- GF_5 has lower mean bias than UFS_P5
- Deeper tropical convection (lower OLR) in GF_5 relative to UFS_P5 (bottom right)



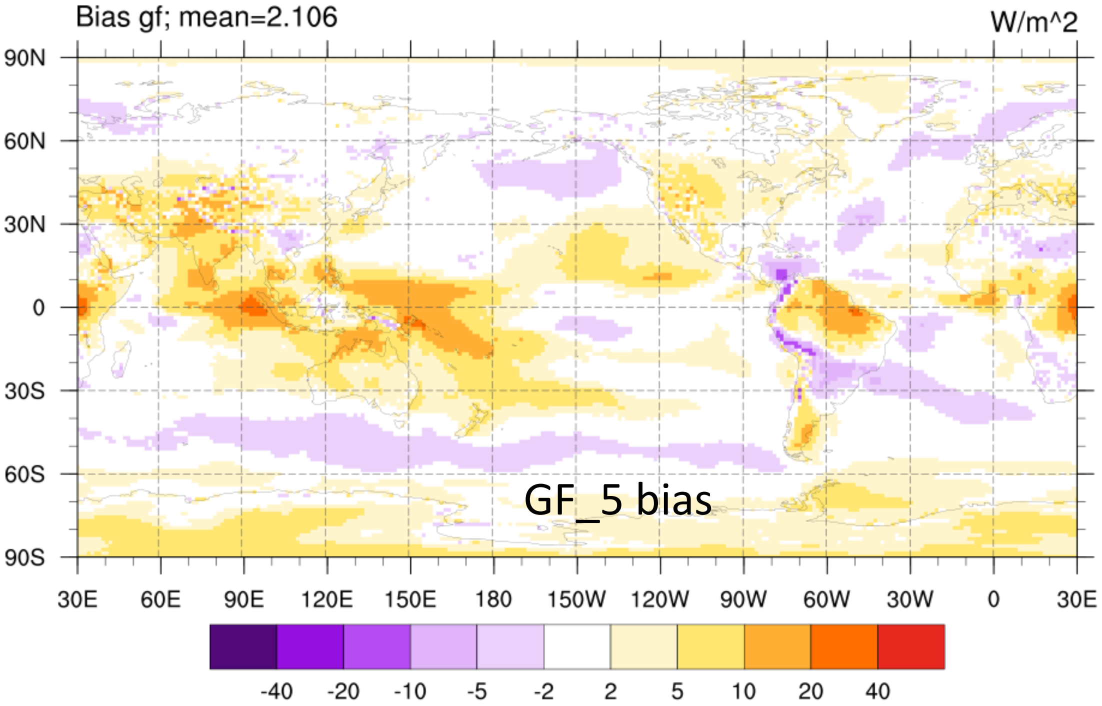


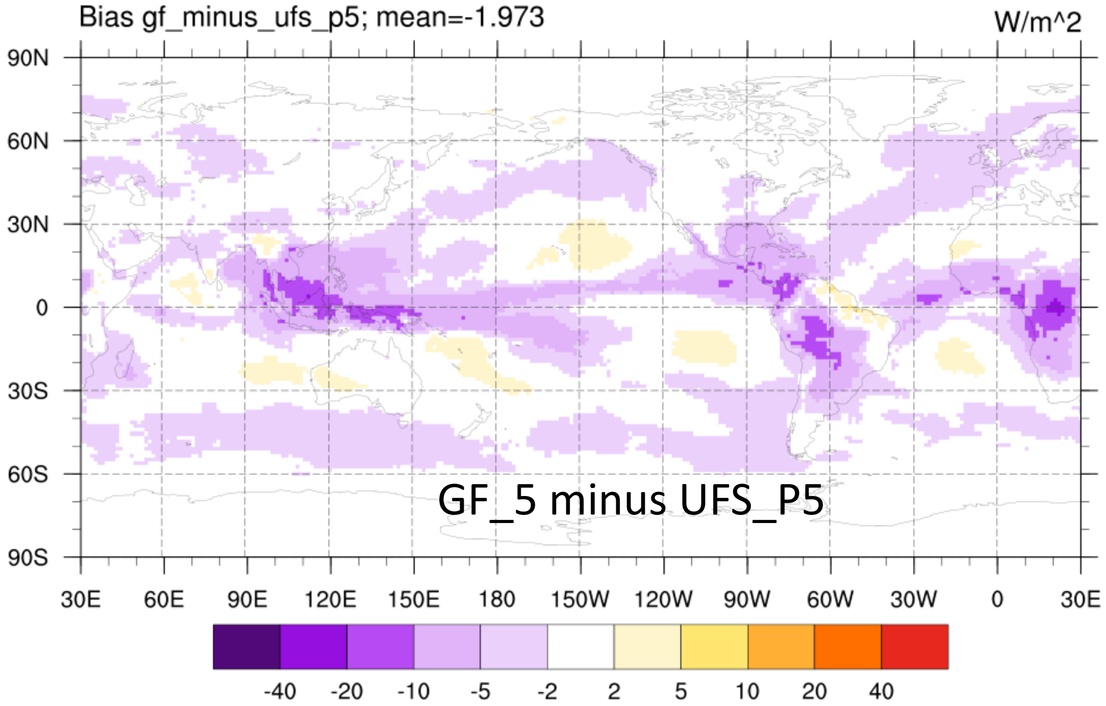


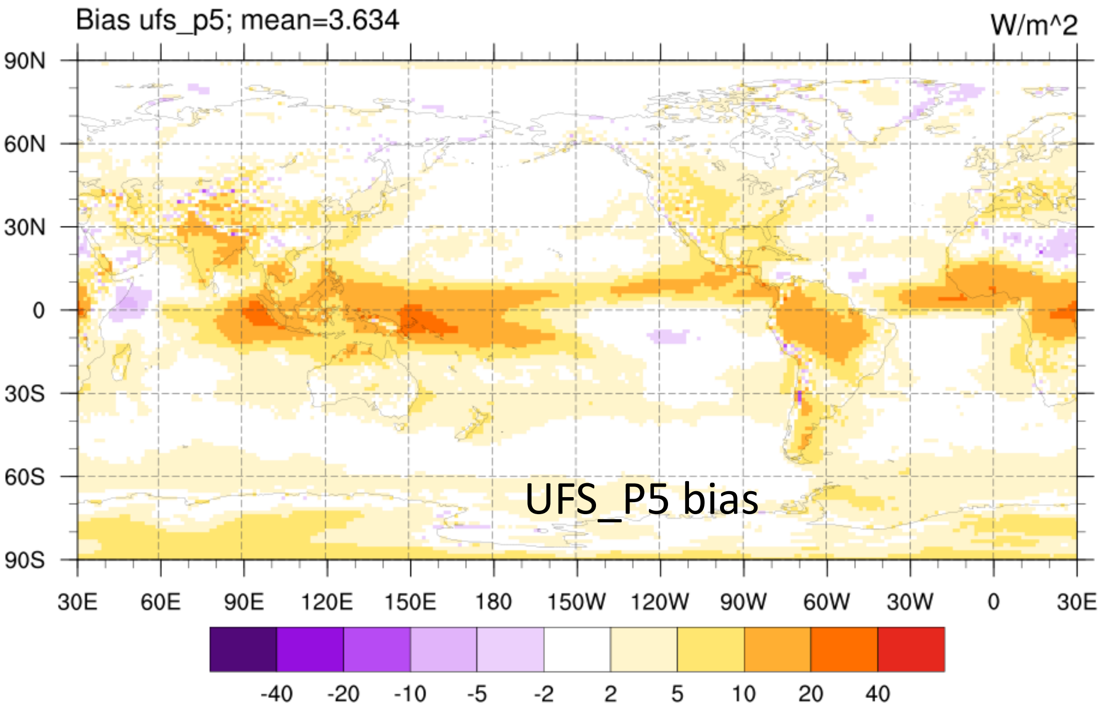
P5 results: OLR bias



- Using HRIS as truth
- Weeks **3-4**, UFS_P5 vs. GF_5: 168 cases
- GF_5 has lower mean bias than UFS_P5
- Bias patterns similar to week 1, but larger
- Still deeper tropical convection in GF_5

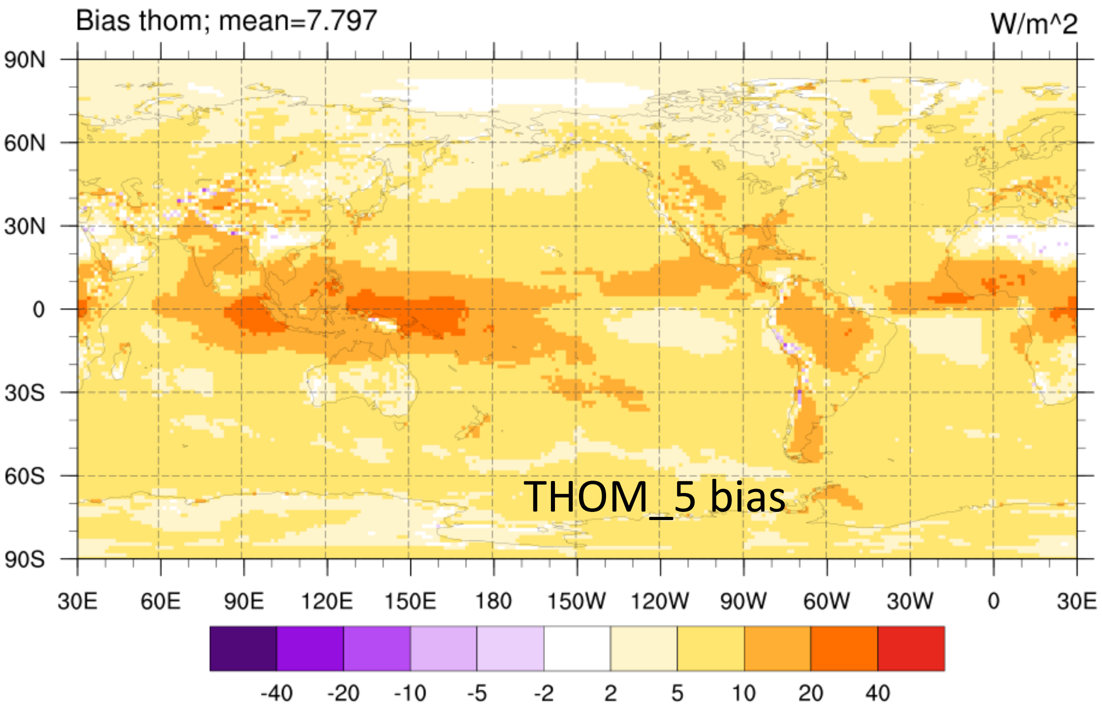


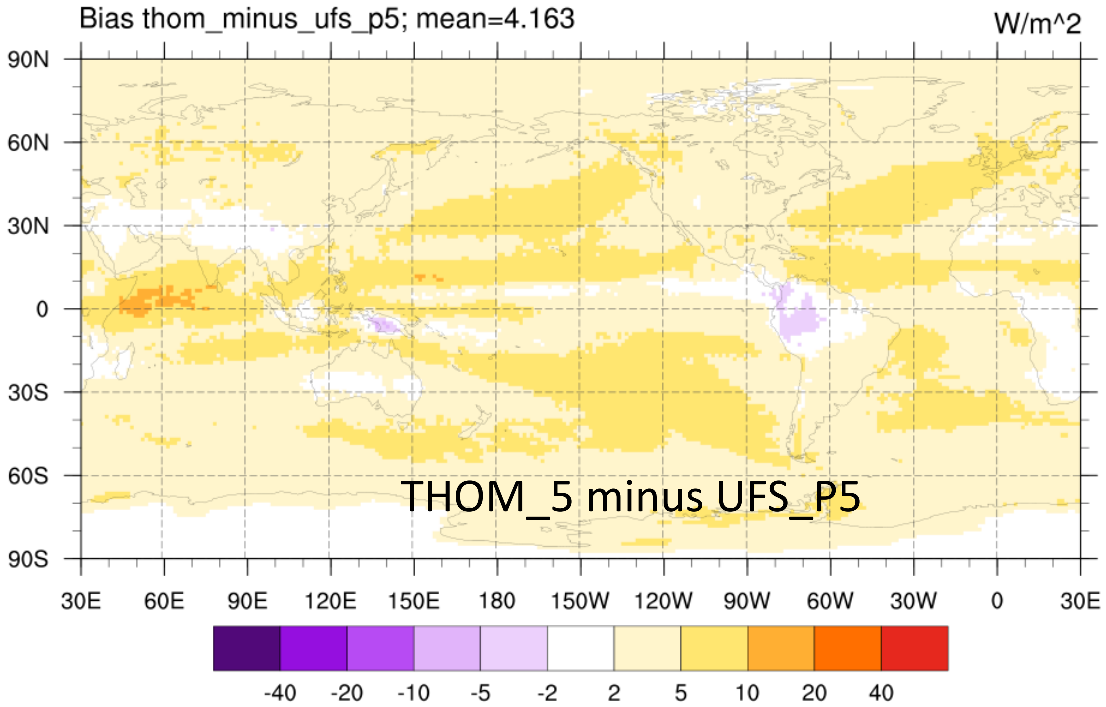




P5 results: OLR bias

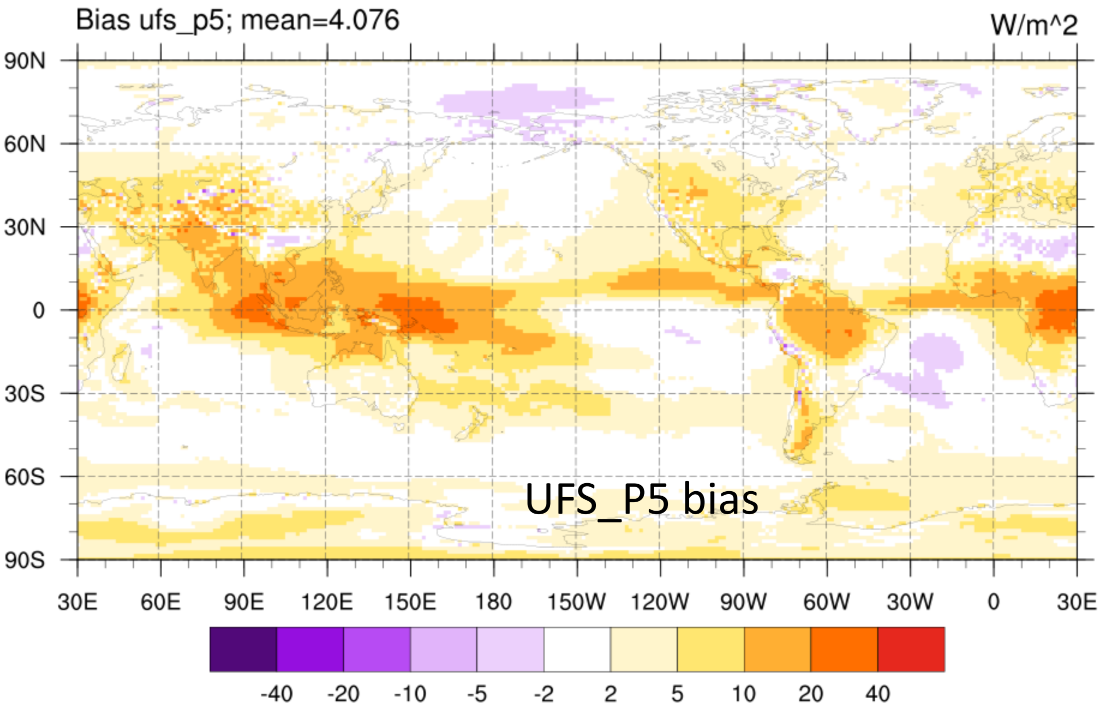
- Using HRIS as truth
- Week **1**, UFS_P5 vs. THOM_5: 166 cases
- Nearly uniformly higher OLR in THOM_5: bug or feature?

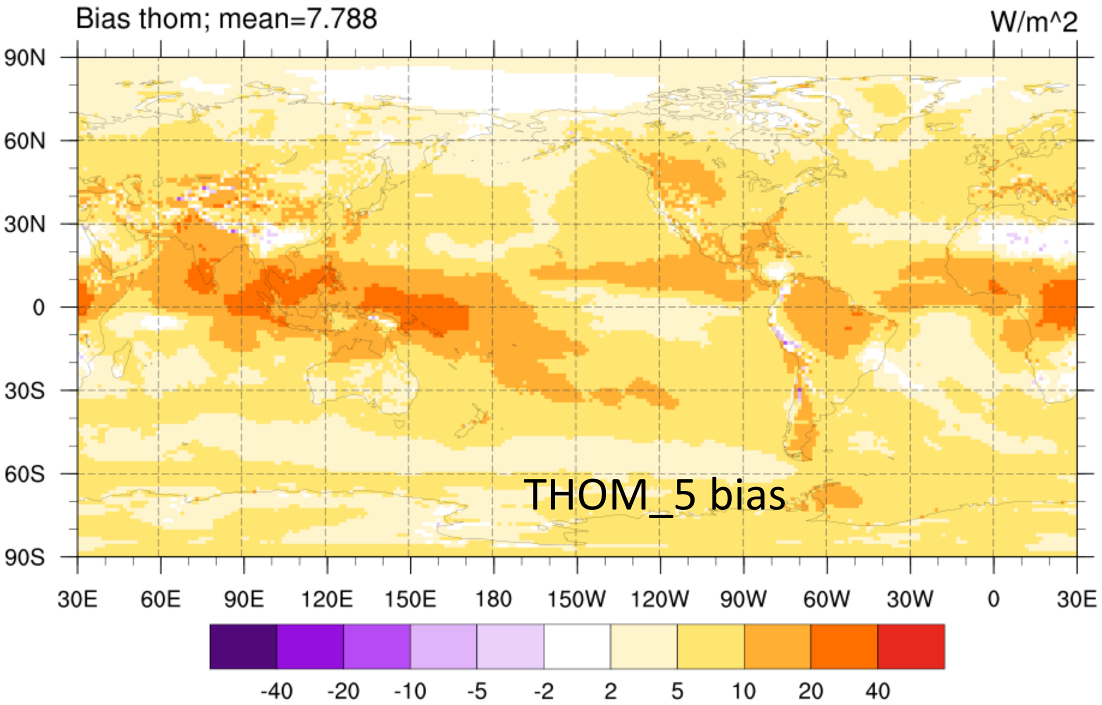


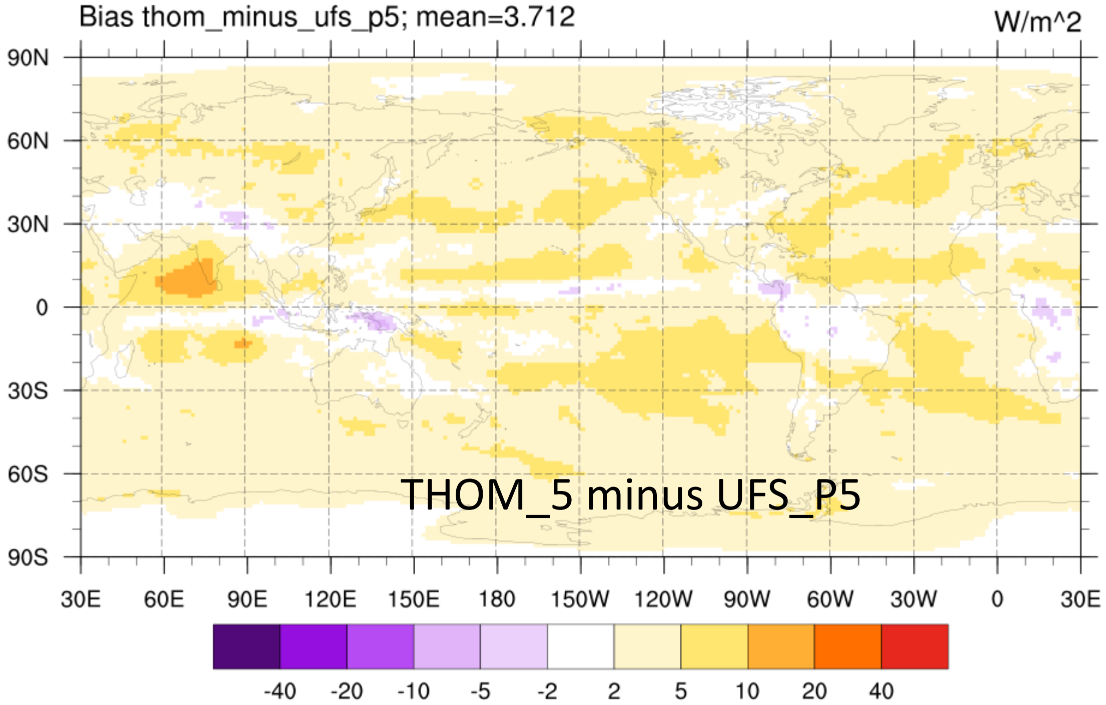


P5 results: OLR bias

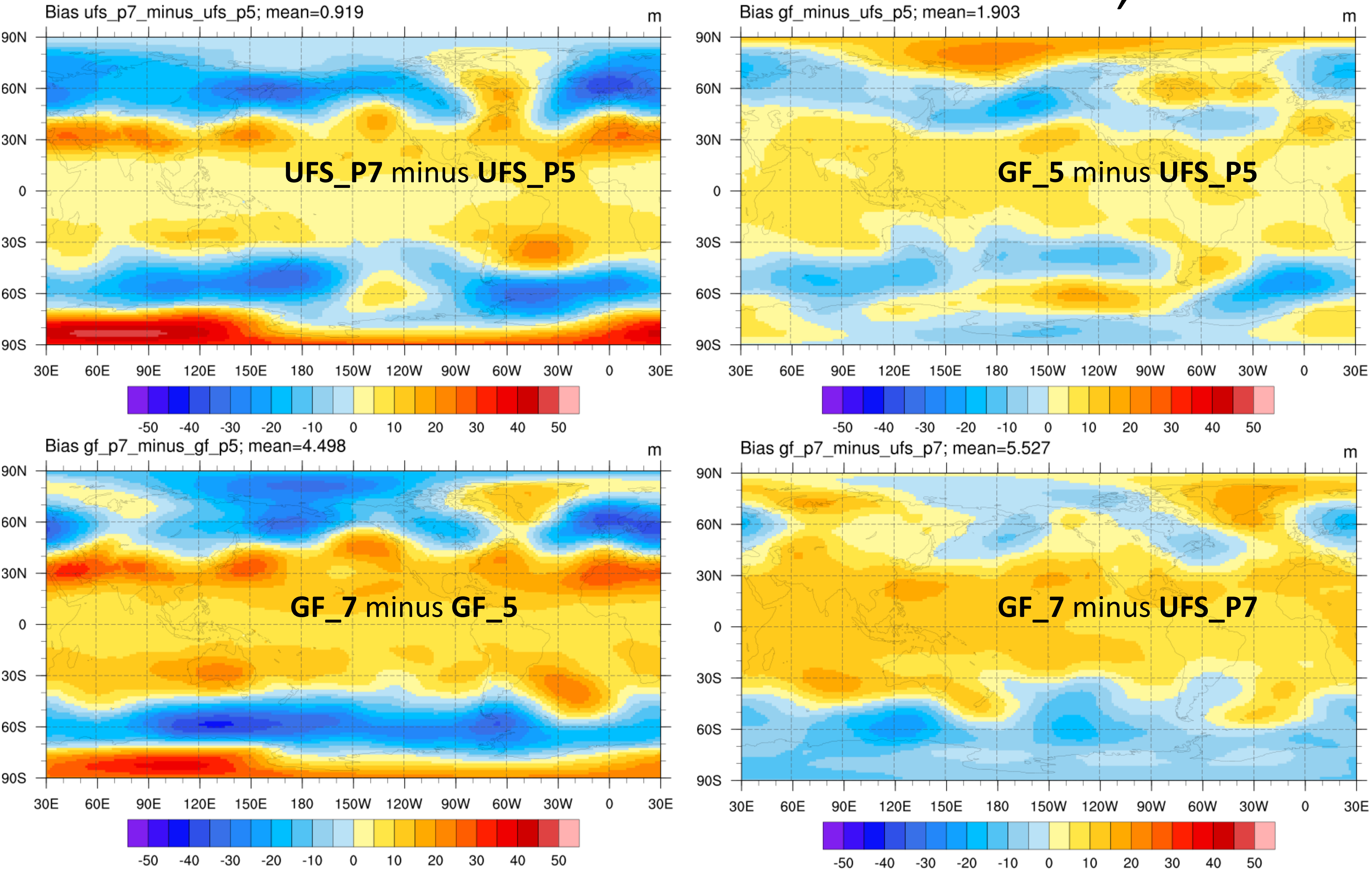
- Using HRIS as truth
- Weeks **3-4**, UFS_P5 vs. THOM_5: 166 cases
- Results quite similar to Week 1





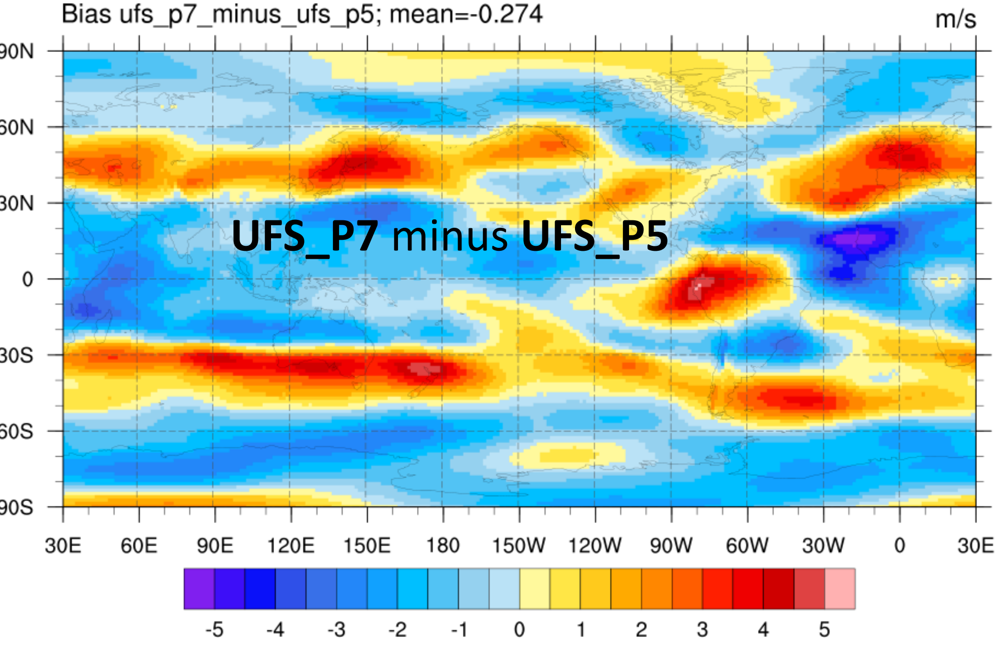


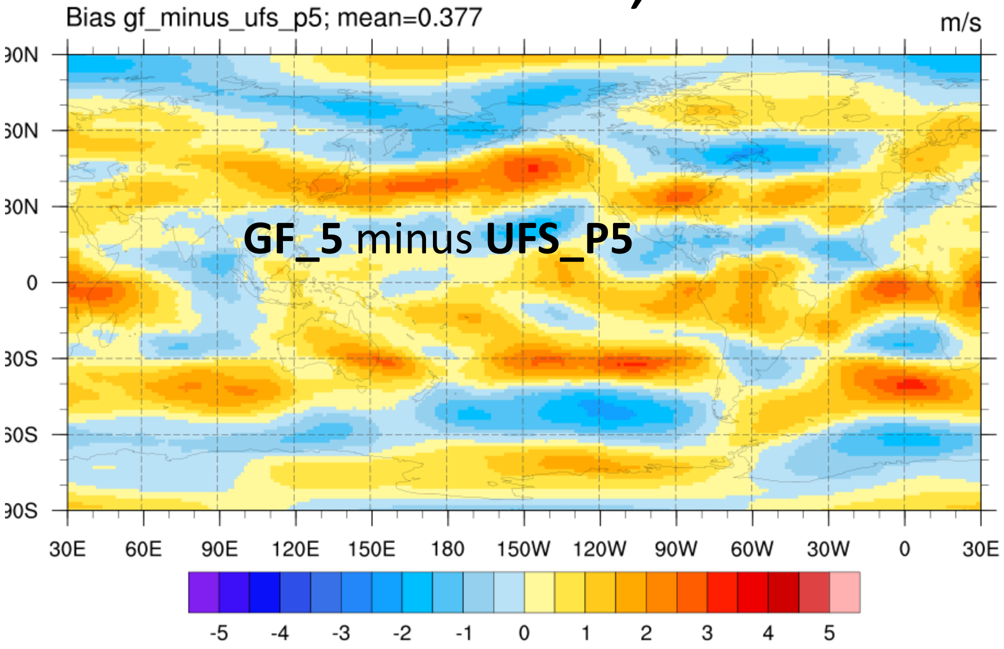
Weeks 3-4 Z500 bias: P5 vs. P7, GF vs. SASAS

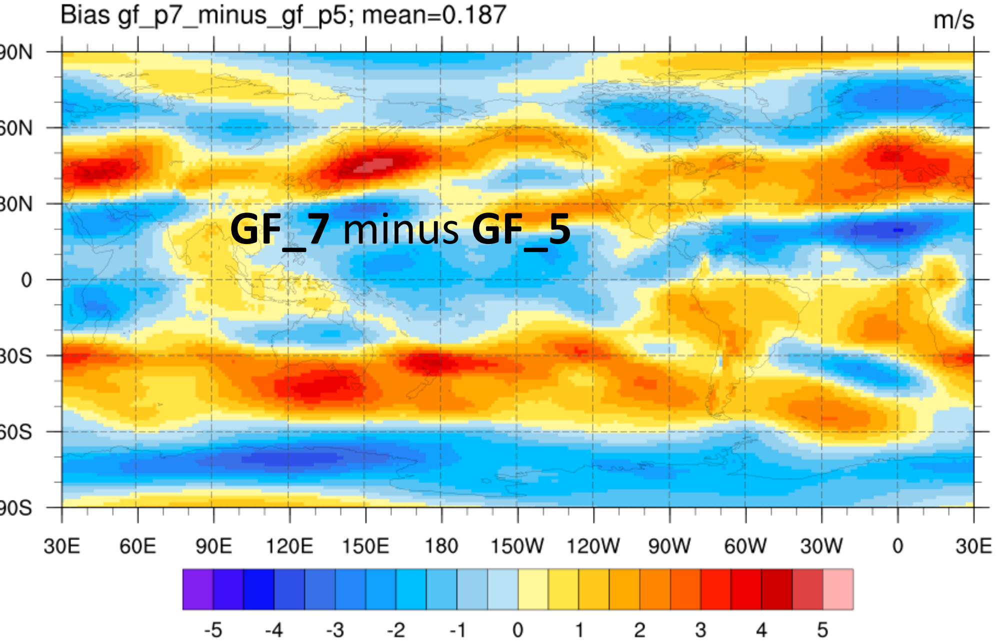


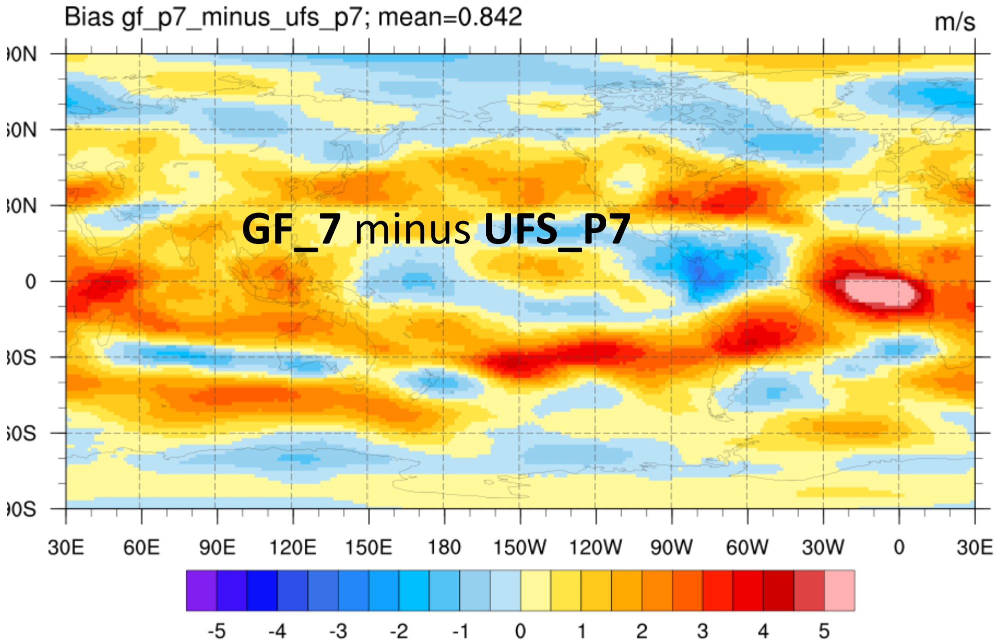
- P5 → P7 has large change in Antarctic Z500 bias (left)
- GF (right) and P7 (left) have higher heights 30S-30N compared to SAS (right) and P5 (left), respectively
- GF (right) and P7 (left) have lower heights poleward of ±60° compared to SAS (right) and P5 (left), respectively: stronger jets?

Weeks 3-4 U200 bias: P5 vs. P7, GF vs. SASAS









- P5 → P7 has stronger jets (consistent with Z500)
- SAS vs. GF: much higher frequency patterns in difference fields (right)