Model validation using GOES-16 Brightness Temperatures

Sarah Griffin*, *Jason Otkin*, Gregory Thompson^, Sharon Nebuda*, Maria Frediani^, Tara Jensen^#%, Patrick Skinner~, Judith Berner^, Eric Gilleland^%, Timothy Supinie@, Fanyou Kong@, and Ming Xue@*

** Cooperative Institute for Meteorological Satellite Studies, University of Wisconsin-Madison*

^ National Center for Atmospheric Research

 # Research Applications Laboratory

% Developmental Testbed Center

~ Cooperative Institute for Mesoscale Meteorological Studies

@ Center for Analysis and Prediction of Storms

This work is supported by the Joint Technology Transfer Initiative

Outline

This presentation is a combination of two different approaches:

- Ensemble based
- Model configuration based

1. Methodology:
    - Method for Object-Based Diagnostic Evaluation (MODE)
    - Validation Statistics
2. Ensemble-based validation
3. Model configuration-based validation

Outline

This presentation is a combination of two different approaches:

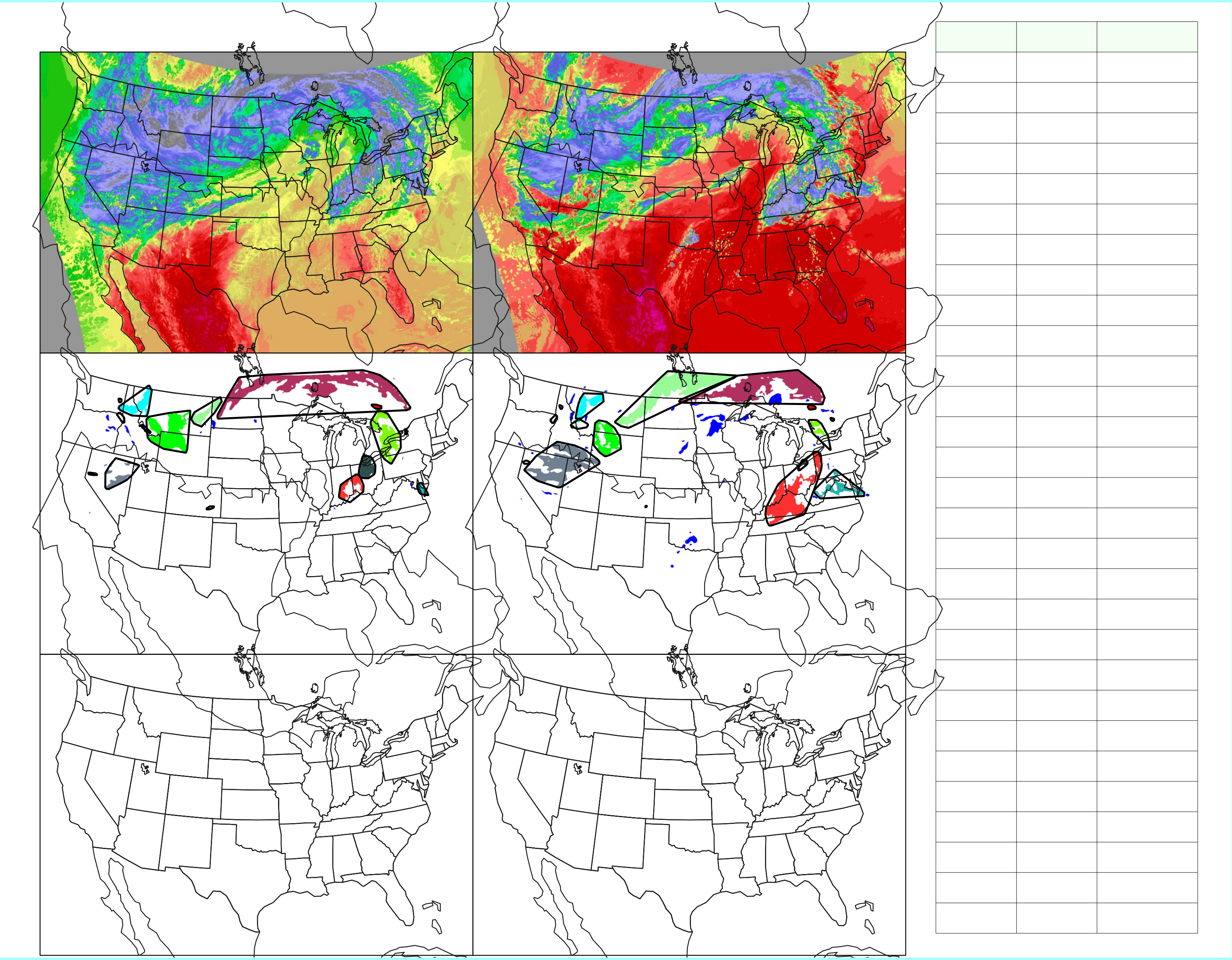
- Ensemble based
- Model configuration based

1. **Methodology:**
    - **Method for Object-Based Diagnostic Evaluation (MODE)**
    - **Validation Statistics**

2. Ensemble-based validation

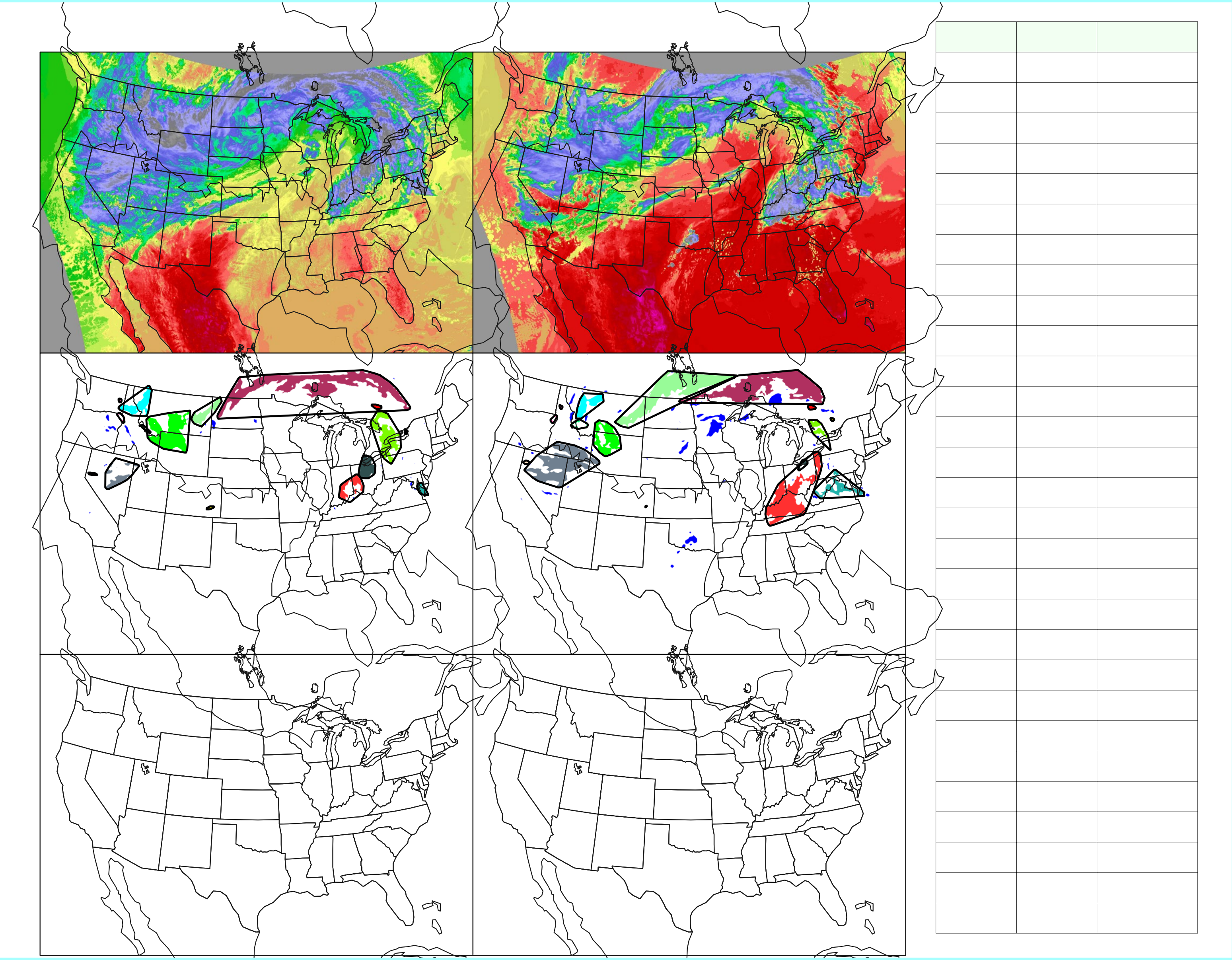
3. Model configuration-based validation

Method for Object-Based Diagnostic Evaluation (MODE)



1. Identify objects in forecast and observation fields.
2. Identify various object attributes for each object, such as location and size.
3. Match the forecast and observation cloud objects.
4. Output attributes for individual objects, such as location and size, and matched object pairs, such as the distance between object centers, ratio of object sizes, and overall interest score describing the "goodness" of the match for assessment.

Method for Object-Based Diagnostic Evaluation (MODE)



Clusters: one or more observation objects matched with one or more forecast objects

- Must have an interest score > 0.65
- Useful when analyzing matched object pairs, as otherwise smaller objects might not have a match and skew statistics
- Examples:
  - Gray objects over Nevada
  - Green objects over Ontario, Canada

Method for Object-Based Diagnostic Evaluation (MODE)

Interest Scores: similarity between matching forecast and observation MODE objects

Object Pair Attribute	User-Defined Weight (%)	Description
centroid_dist	4 (25.0)	Distance between objects' "center of mass"
boundary_dist	3 (18.75)	Minimum distance between the objects
convex_hull_dist	1 (6.25)	Minimum distance between the polygons surrounding the objects
angle_diff	1 (6.25)	Orientation angle difference
area_ratio	4 (25.0)	Ratio of the forecast and observation objects' areas (or its reciprocal, whichever yields a lower value)
int_area_ratio	3 (18.75)	Ratio of the objects' intersection area to the lesser of the observation or forecast area (whichever yields a lower value)

Validation Statistics

Mean Absolute Error (MAE): $\text{MAE}=\frac{1}{N}\sum_{i=1}^{N}\left|F_i - O_i\right|$

Mean Bias Error (MBE): $\text{MBE} = \frac{1}{N}\sum_{i=1}^{N}(F_i - O_i)$

F and O : forecast and observation BTs

Validation Statistics

Mean Absolute Error (MAE): MAE=$\frac{1}{N}\sum_{i=1}^{N}|F_i - O_i|$

Mean Bias Error (MBE): MBE = $\frac{1}{N}\sum_{i=1}^{N}(F_i - O_i)$

F and O : forecast and observation BTs

Two different approaches:

1. Over the full domain

Validation Statistics

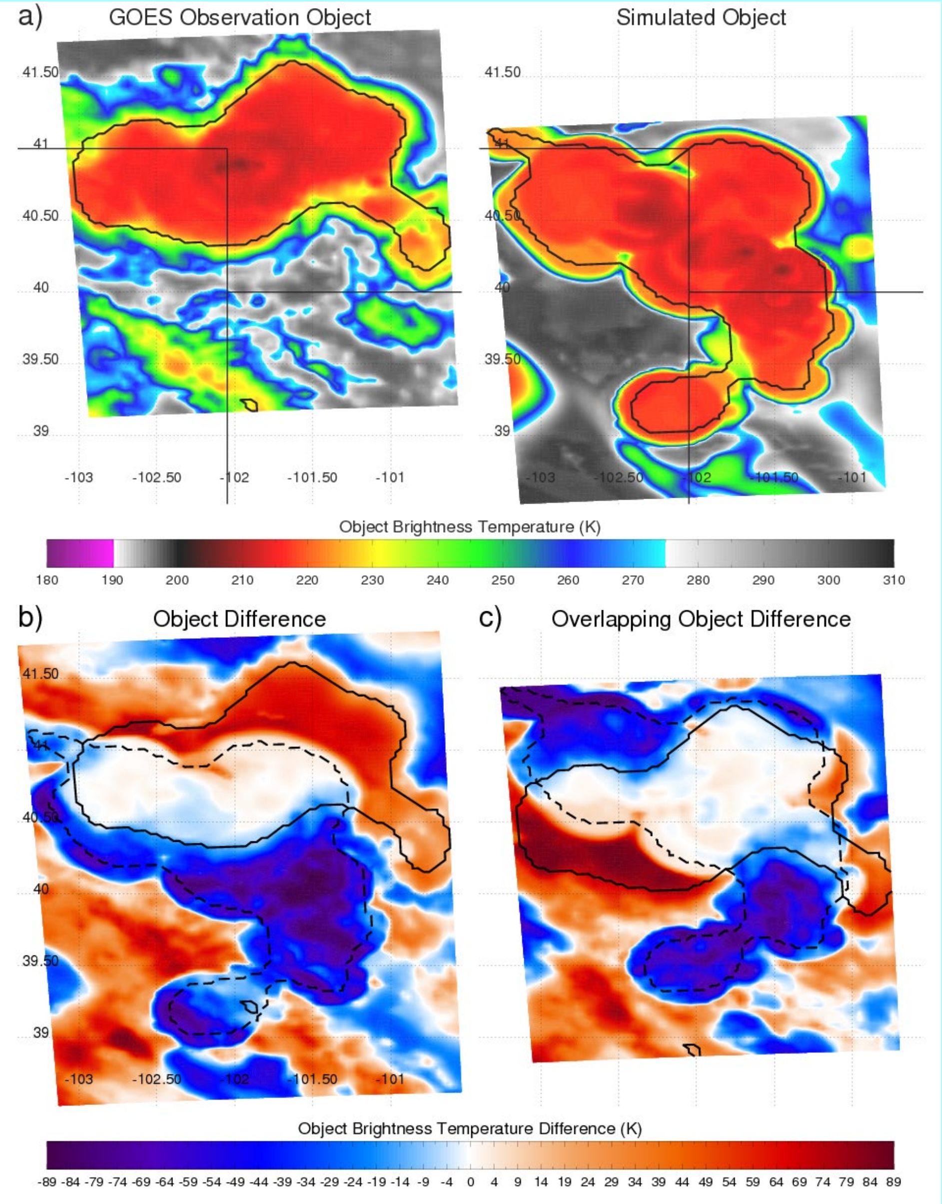
Mean Absolute Error (MAE): MAE= $\frac{1}{N}\sum_{i=1}^{N}|F_i - O_i|$

Mean Bias Error (MBE): MBE = $\frac{1}{N}\sum_{i=1}^{N}(F_i - O_i)$

F and O : forecast and observation BTs

Two different approaches:

1. Over the full domain
2. Over individual object/cluster matches where the displacement between objects has been removed.



Outline

This presentation is a combination of two different approaches:

- Ensemble based
- Model configuration based

1. Methodology:
   - Method for Object-Based Diagnostic Evaluation (MODE)
   - Validation Statistics

2. **Ensemble-based validation**

3. Model configuration-based validation

Model Configurations: WRF

1. SPP-MP ensemble: 5 members

- Time- and spatially-varying SPP perturbations were added to the graupel spectra Y-intercept parameter.
- Uncertainly is introduced in the cloud water gamma distribution at a scale of +/- 3.
- Vertical velocity was perturbed, which impacts cloud condensation and ice nucleation.

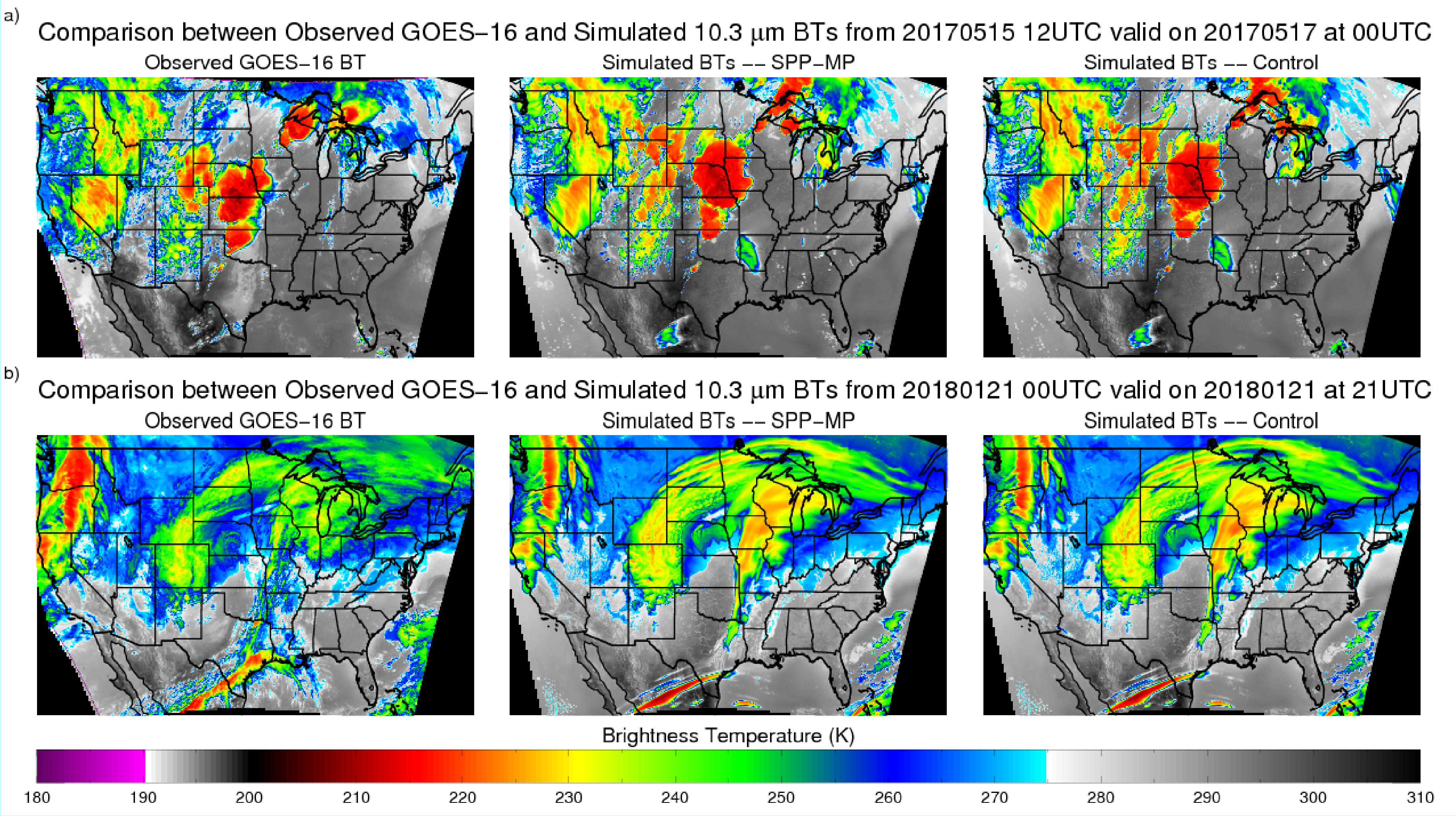
2. Control ensemble: 5 members

- White noise perturbations are introduced at the initialization time to four ensemble members.
- Fifth member is the unperturbed control initialization.

For more information, please see:

Thompson, G., J. Berner, M. Frediani, J. A. Otkin, and S. M. Griffin, 2020: A Stochastic Parameter Perturbation Method to Represent Uncertainty in a Microphysics Scheme. Submitted to *Mon. Wea. Rev*

Model Configurations: WRF



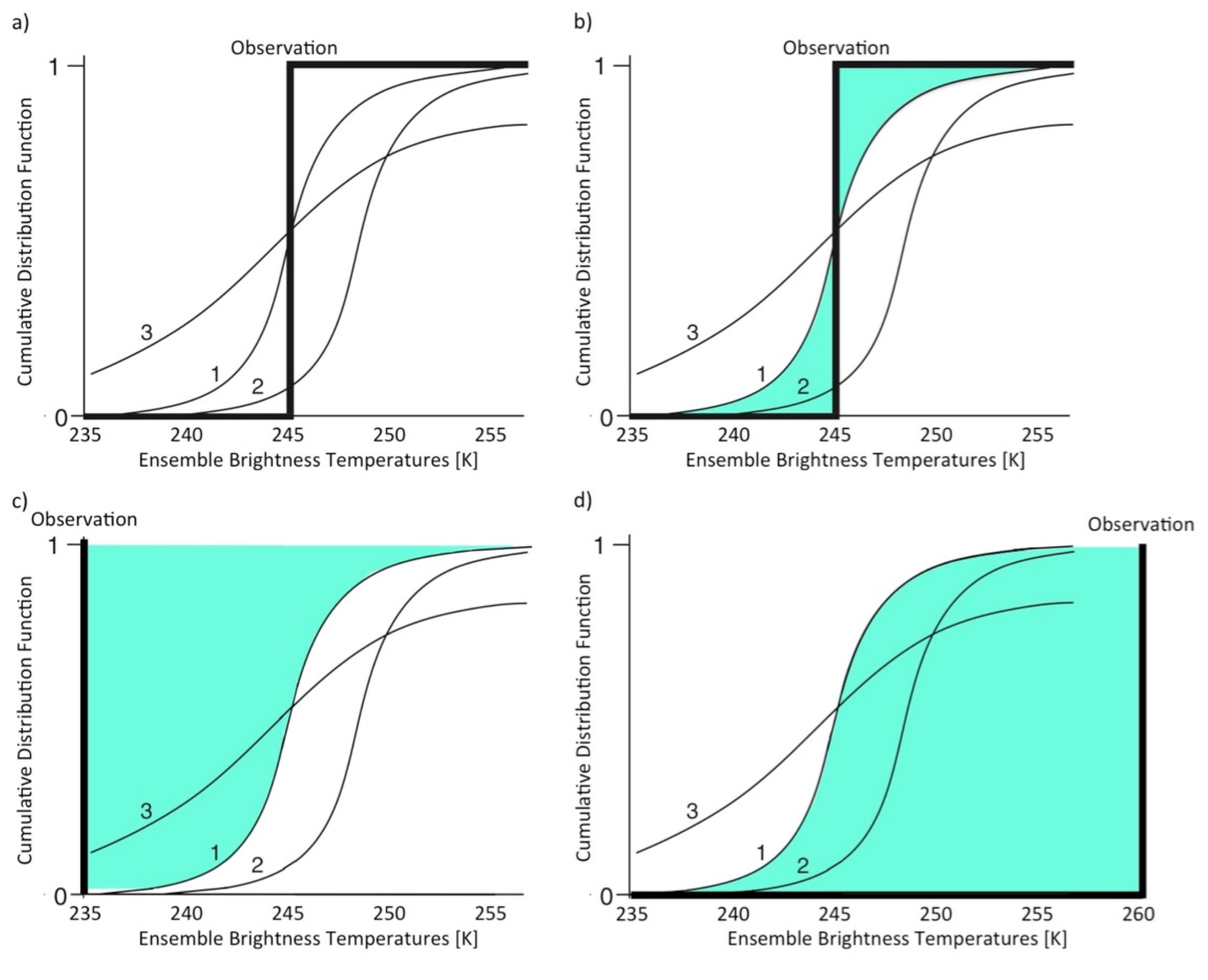
Ten 48-h forecasts, with forecast hours 0-5 are not used due to model spin-up.

- forecasts initialized at 12 UTC in May 2017
- forecasts initialized at 00 UTC in January 2018

Validation Statistics

Continuous Ranked Probability Score:

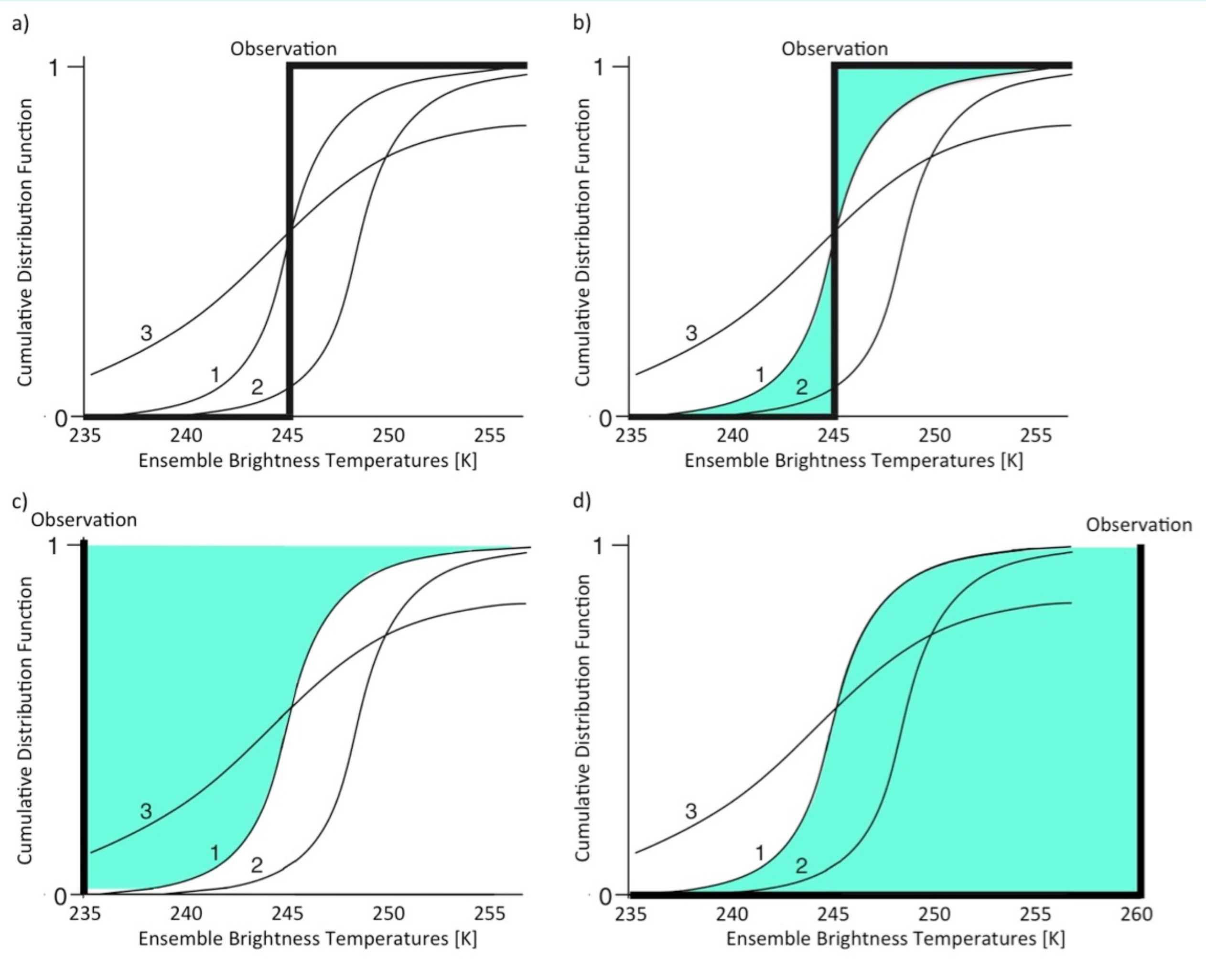
- compares the cumulative distribution function (CDF) of the simulated ensemble BTs to the observed BT at a given pixel
- Green indicates the CRPS for ensemble 1 for cases where the observation BT is within (b) and outside (c and d) the ensemble BT CDF



Validation Statistics

Continuous Ranked Probability Score:

- compares the cumulative distribution function (CDF) of the simulated ensemble BTs to the observed BT at a given pixel
- Green indicates the CRPS for ensemble 1 for cases where the observation BT is within (b) and outside (c and d) the ensemble BT CDF.

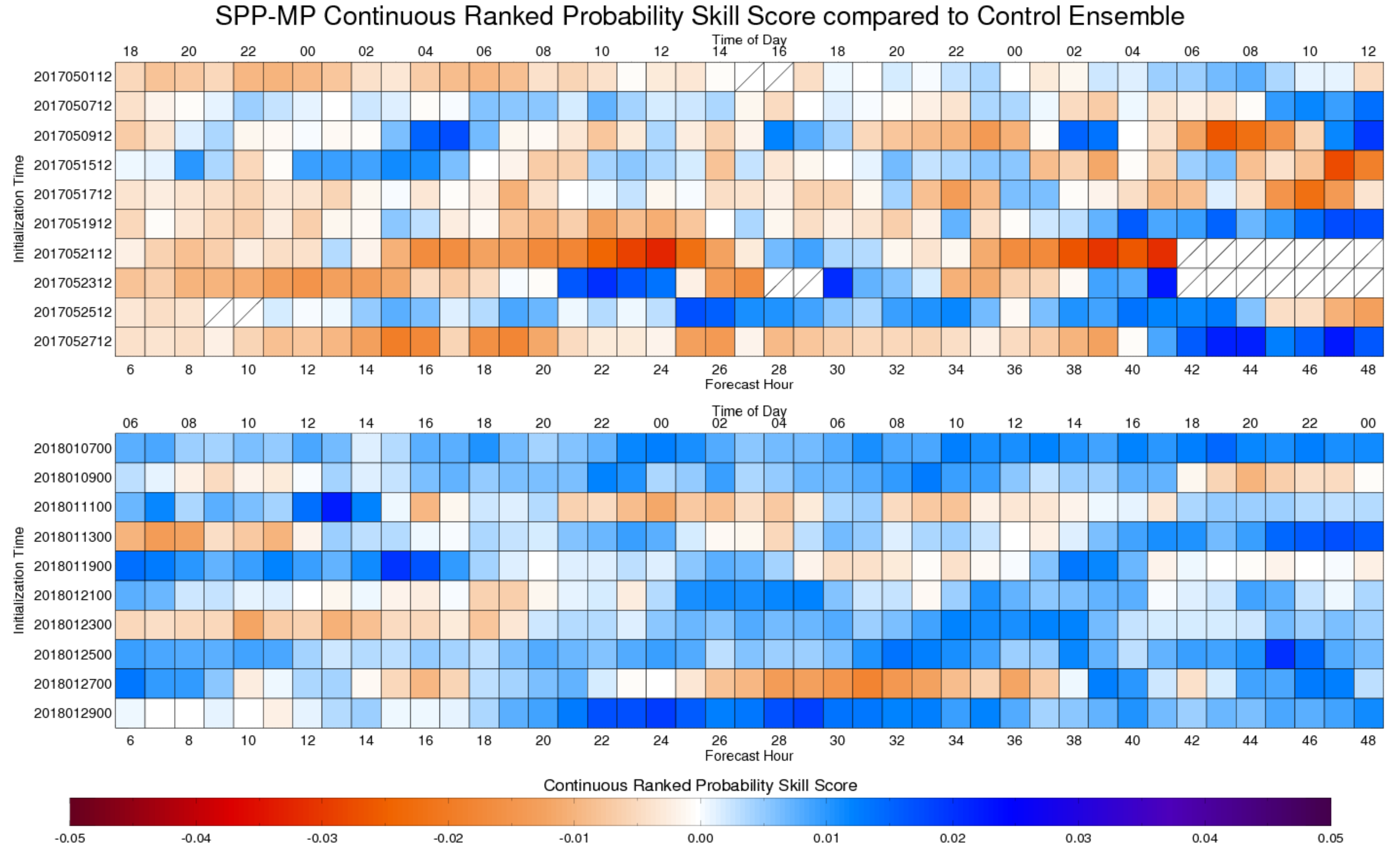


Continuous Ranked Probability Skill Score (CRPSS):

$$\mathrm{CRPSS} = 1 - \frac{\mathrm{CRPS}_{\mathrm{SPP-MP}}}{\mathrm{CRPS}_{\mathrm{Control}}}$$

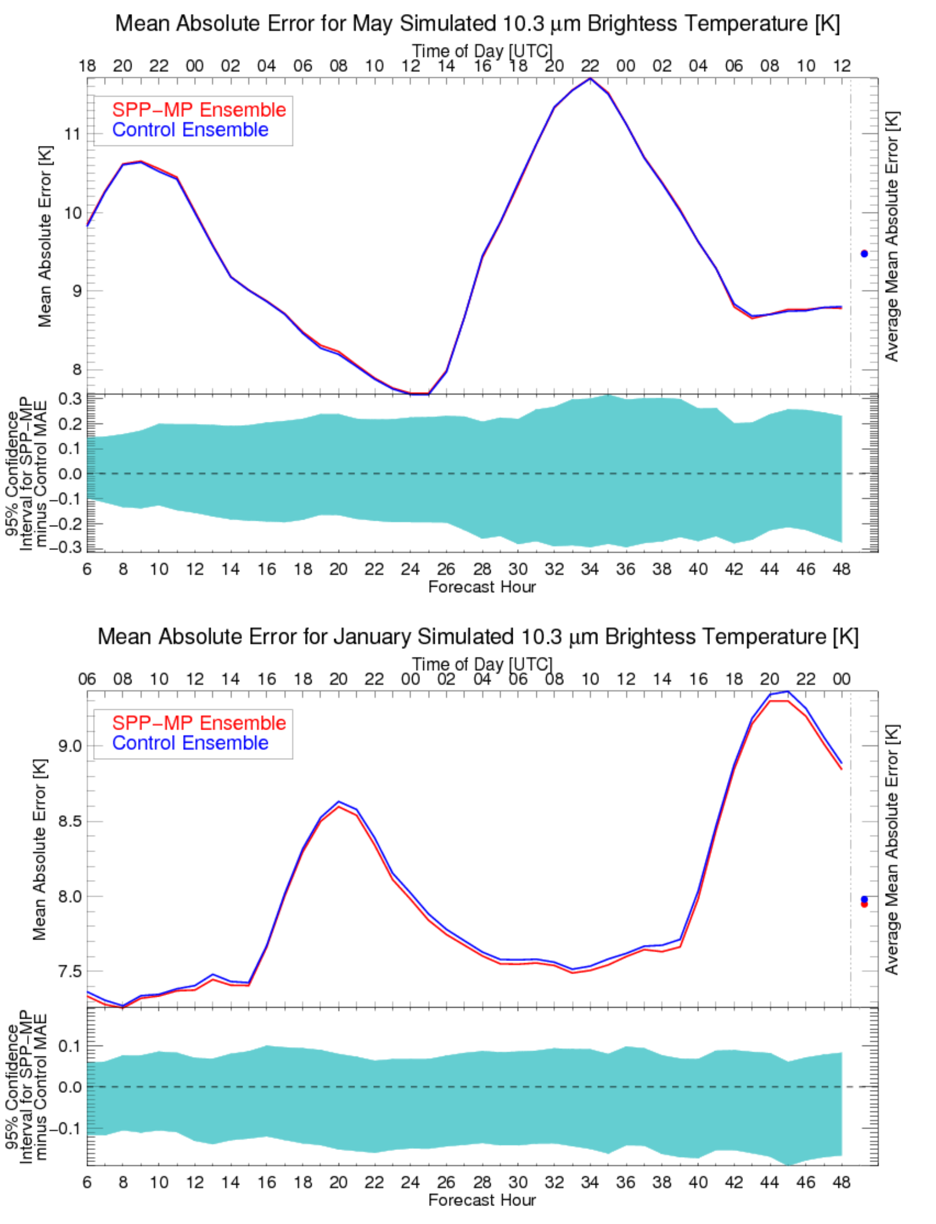
Positive CRPSS indicates the SPP-MP ensemble BTs more closely represent the observed GOES BT than the Control ensemble BTs.

Results: CRPSS



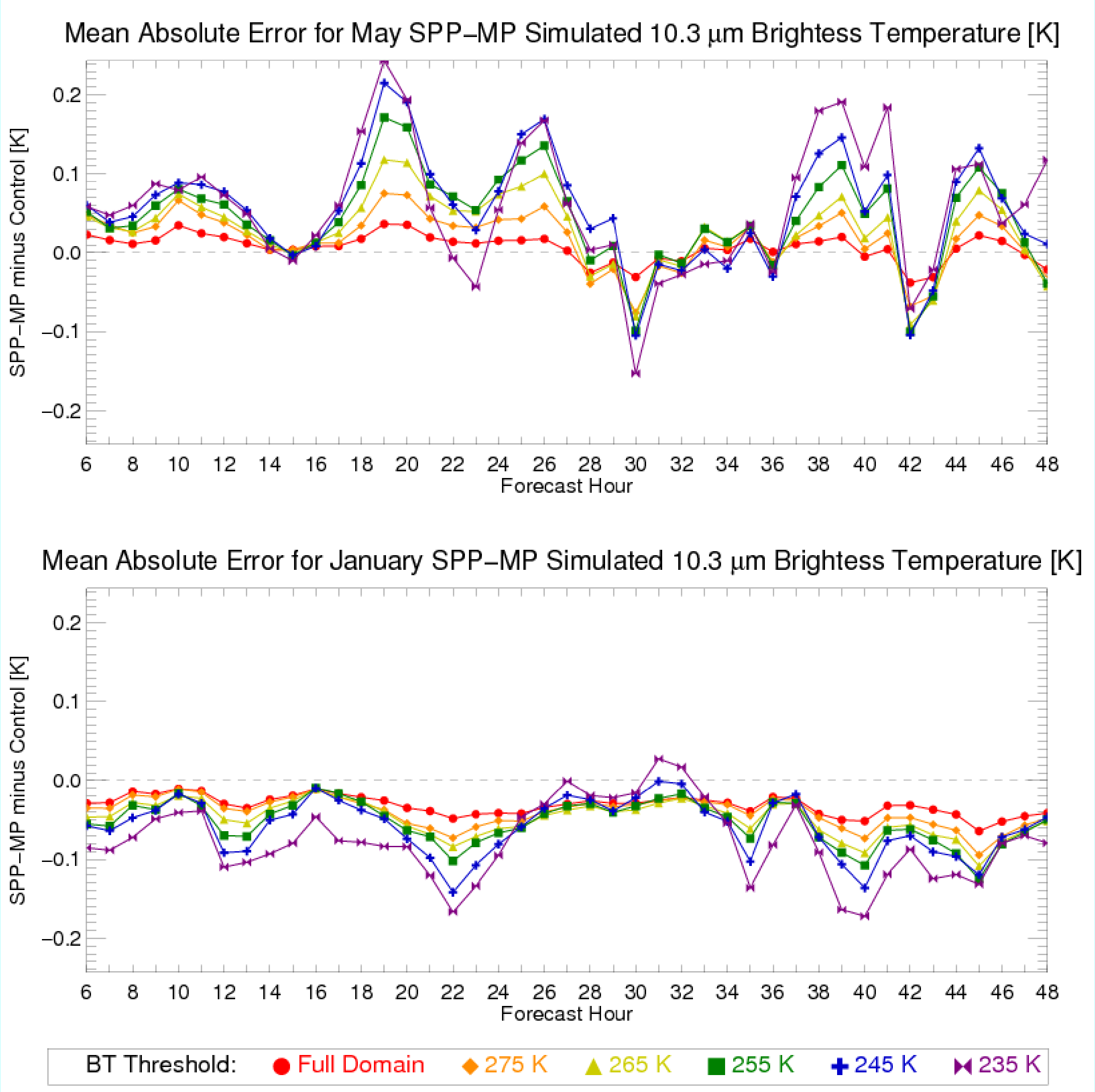
- CRPSS is negative for 61% of forecast hours in May.
    - ensemble not enclosing the observed BTs
        - 70.4% of May forecasts
        - 49.8% of January
- CRPSS positive for 75% of forecast hours in January.
    - smaller spread in SPP-MP ensemble than Control
        - 75.6% of May forecasts
        - 32.1% of January

Results: Domain MAE



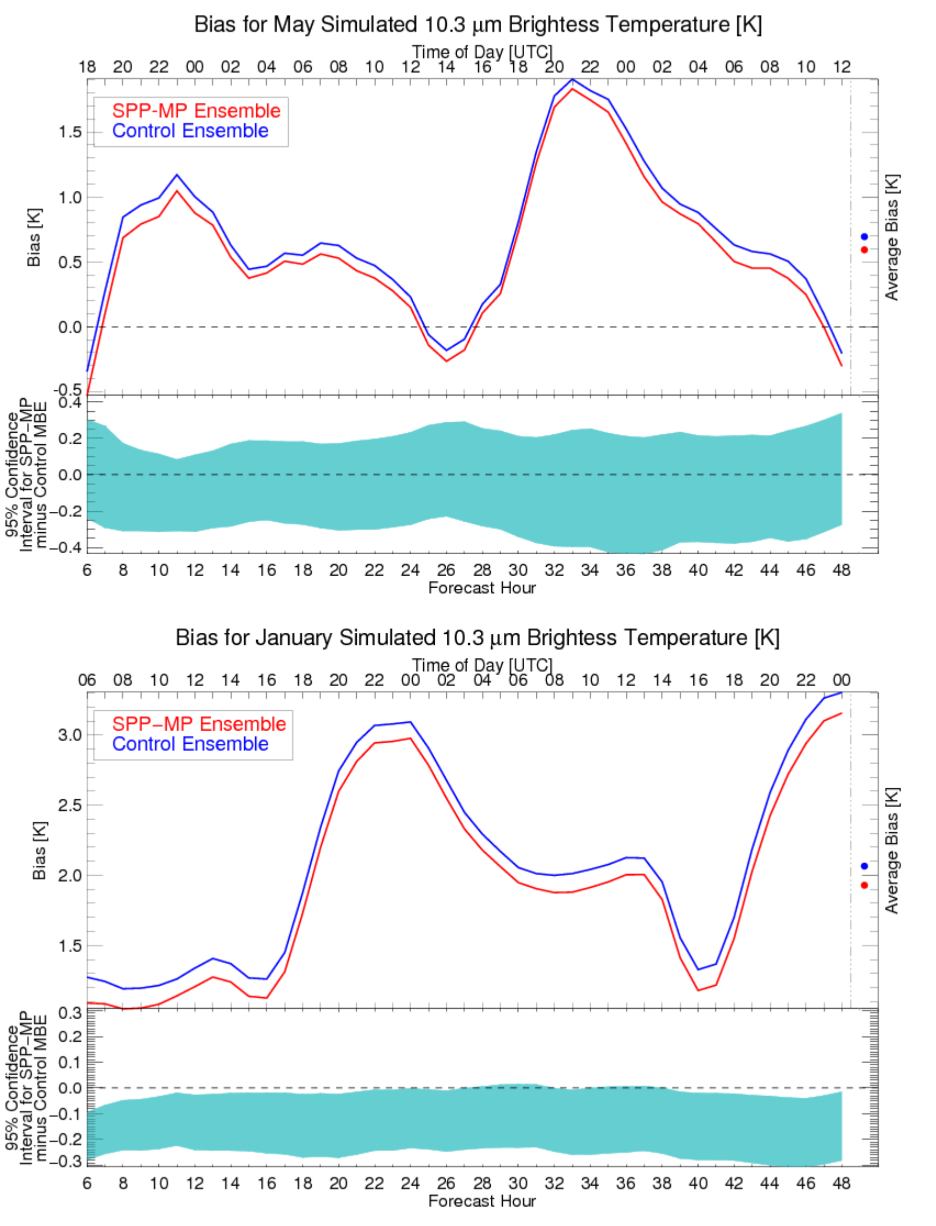
- Little difference between the MAE in May.
- Slightly larger difference in January, with SPP-MP slightly more accurate.
- Differences between SPP-MP and Control not statically significant.

Results: Domain MAE



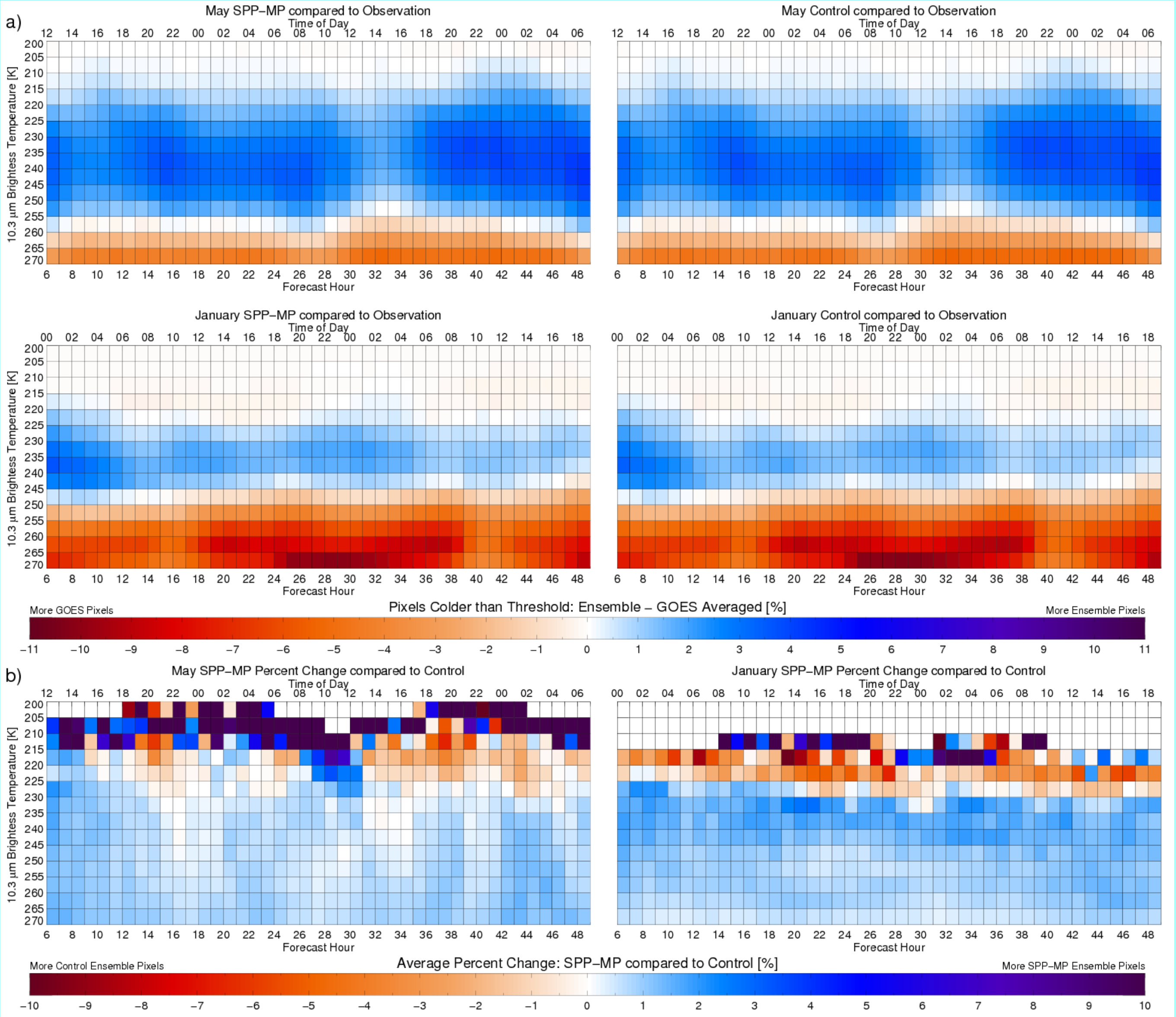
- Little difference between the MAE in May.
- Slightly larger difference in January, with SPP-MP slightly more accurate.
- Differences between SPP-MP and Control not statically significant.
- When analyzing pixels with observation or ensemble BT < threshold:
  - SPP-MP less accurate for colder thresholds in May (SPP-MP minus Control > 0)
  - SPP-MP more accurate for colder thresholds in Jan (SPP-MP minus Control < 0)

Results: Domain MBE



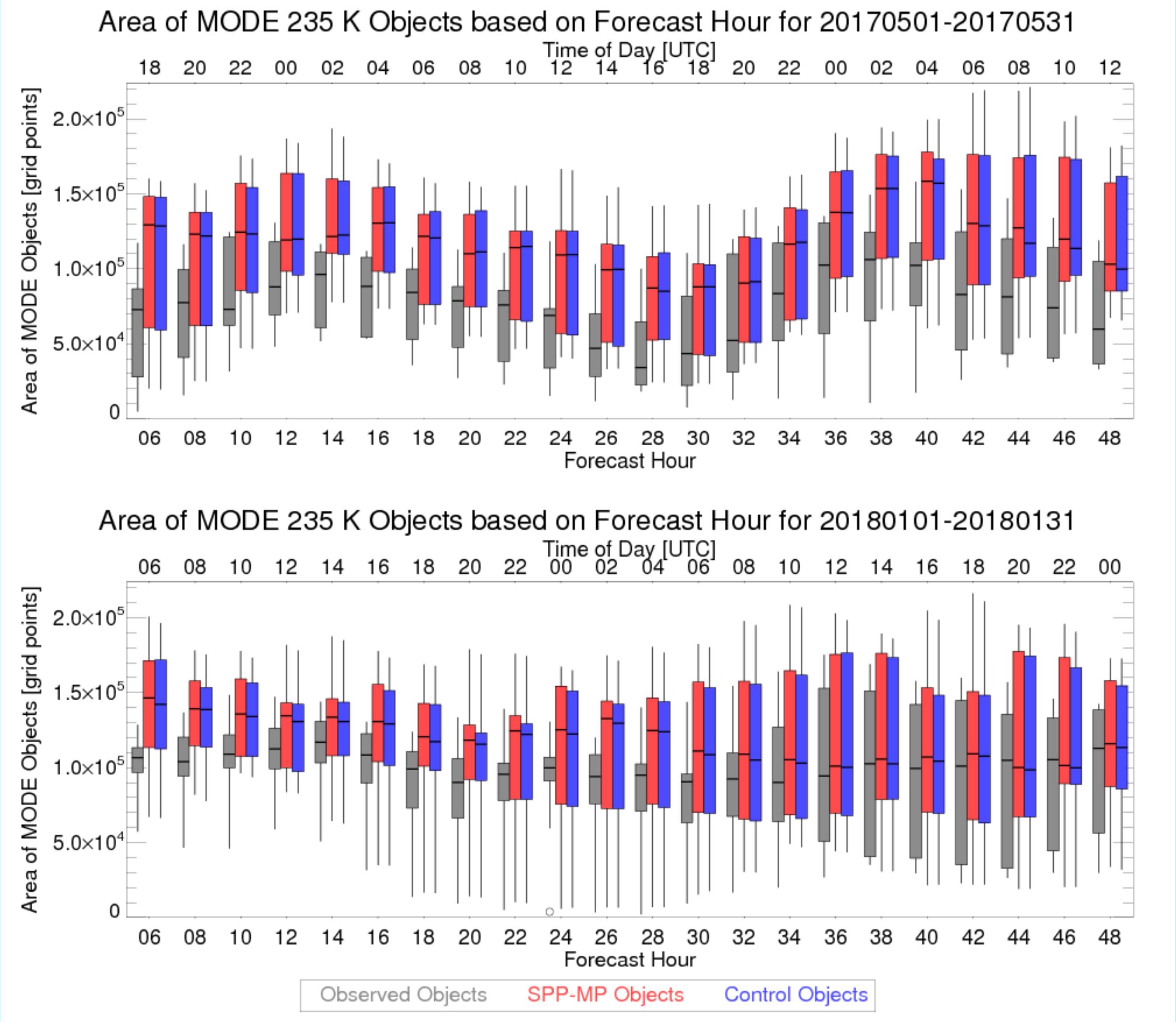
- Positive Bias = Domain BTs too high.
  - SPP-MP BTs are slightly lower than Control.
  - Difference statistically significant for January

Results: Domain MBE



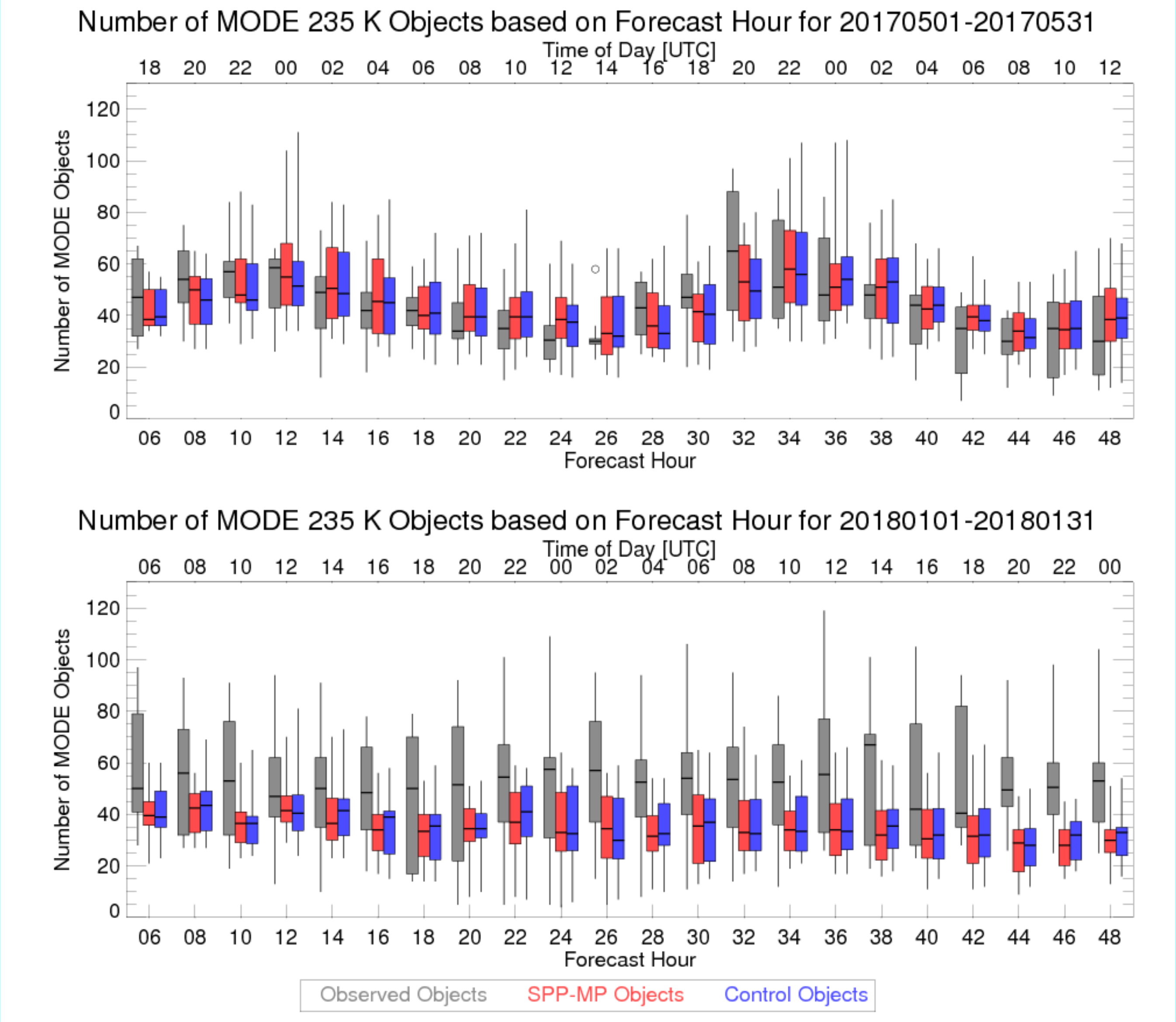
- Positive Bias = Domain BTs too high.
  - SPP-MP BTs are slightly lower than Control.
  - Difference statistically significant for January
- Positive bias due to not enough grid points with 270 K < BT < 255 K.
  - More pixels for SPP-MP compared to Control
- Control has more grid points with BT < 225 K than SPP-MP
  - SPP-MP reduces negative MBE at this BT threshold.

Results: Area of MODE Objects



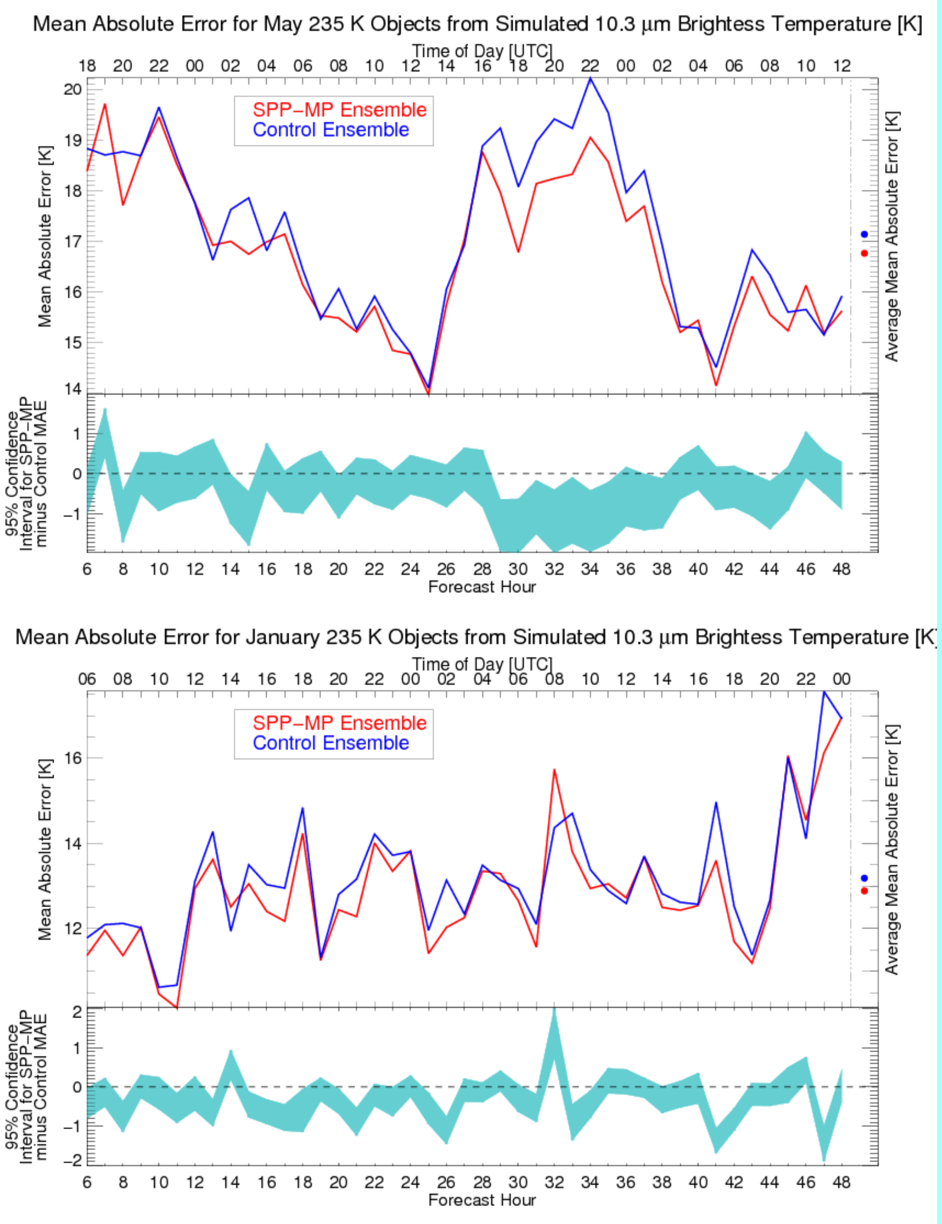
- Area encompassed by the simulated objects is much larger than the area of observed objects.
  - Average area encompassed by SPP-MP larger than Control for both months.

Results: Number of MODE Objects



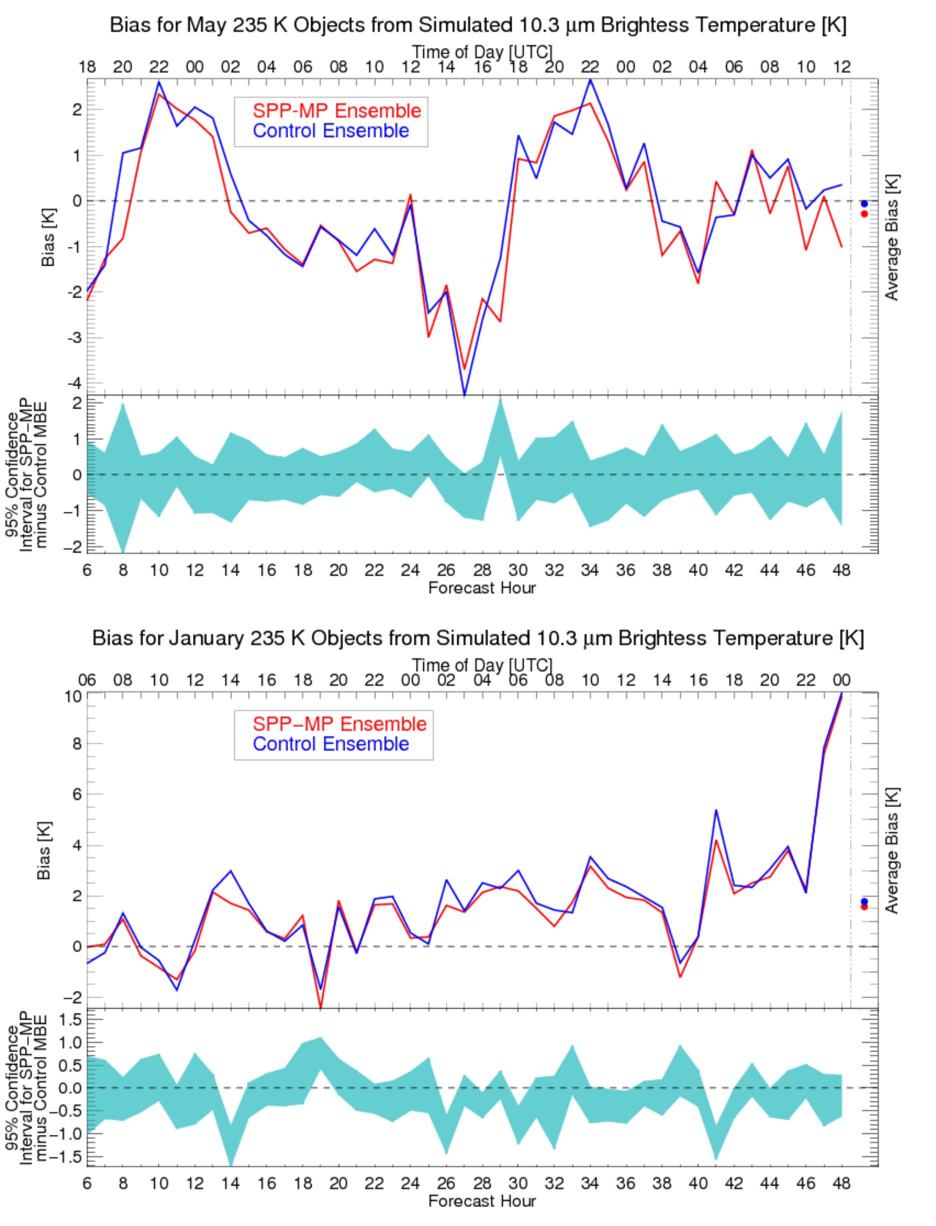
- Area encompassed by the simulated objects is much larger than the area of observed objects.
  - Average area encompassed by SPP-MP larger than Control for both months.
- Slightly more cloud objects in the SPP-MP ensemble than in the Control during May.
- SPP-MP has fewer objects compared to the Control in January.
- Average object size smaller in SPP-MP for 44% (14%) of May (January) forecasts.

Results: Object MAE and MBE



- MAE for SPP-MP is lower than the Control for both months.
  - More accurate at representing object BTs
  - Occasionally difference is statically significant.

Results: Object MAE and MBE



- MAE for SPP-MP is lower than the Control for both months.
  - More accurate at representing object BTs
  - Occasionally difference is statically significant.
- Bias is lower compared to the full domain
  - Too low for May, Control neutral
  - January object BTs still too high, but SPP-MP are slightly lower.
- Bias highly correlated with area ratio.
  - Larger forecast objects compared to observations has lower bias.

Conclusions

1. Model accuracy (MAE) can be analyzed different ways using the same metric.
    - SPP-MP and Control similar MAEs over full domain.
    - SPP-MP has lower accuracy in May when only using grid points with a BT lower than a given threshold in either observations or ensemble member.
        - Defining objects with that threshold and removing displacement results in higher accuracy for the SPP-MP for matched object pairs.
2. Bias also differs depending on analysis domain.
    - Positive over the full domain but negative for matched object pairs (May).
3. MODE allows for analyzing object number and sizes
    - SPP-MP produces more cloud objects in May 2017 compared to the Control
    - SPP-MP produces less cloud objects in January 2018, and both ensembles produce less than the observations.
    - Total area encompassed by objects for both ensembles is larger than the observations.

Outline

This presentation is a combination of two different approaches:

- Ensemble based
- Model configuration based

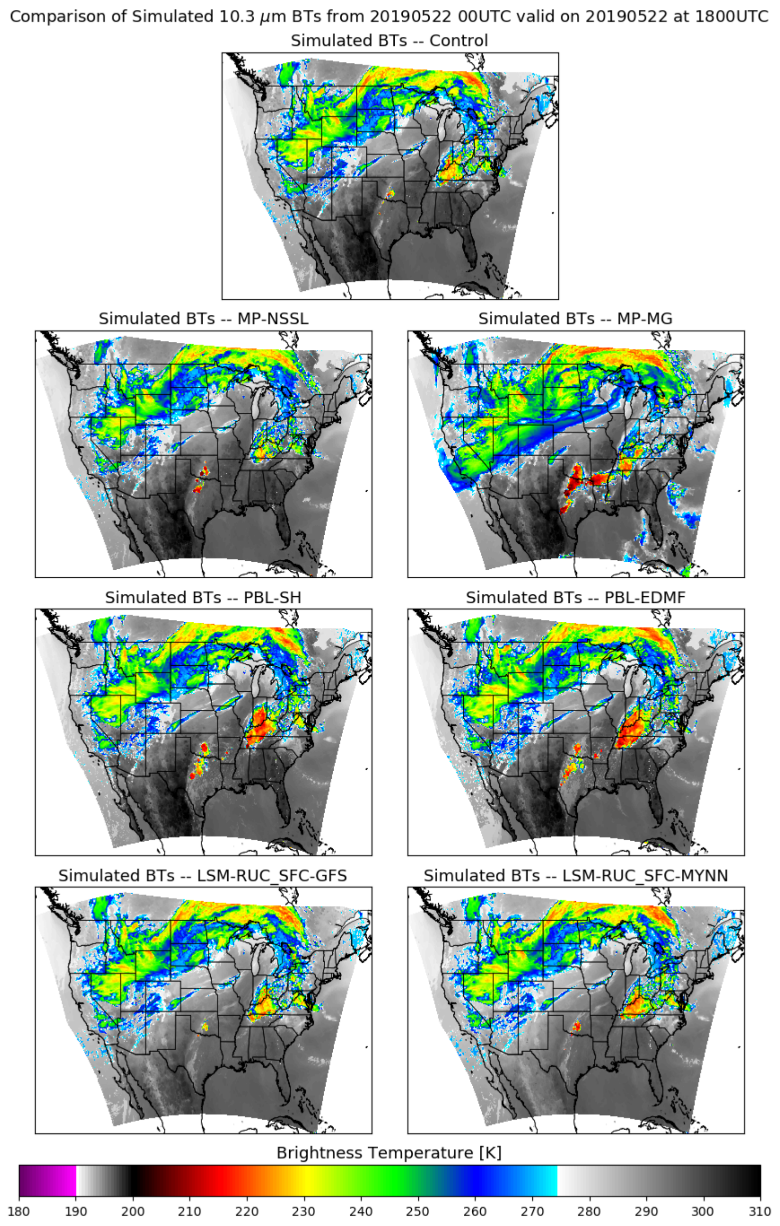
1. Methodology:
   - Method for Object-Based Diagnostic Evaluation (MODE)
   - Validation Statistics

2. Ensemble-based validation

3. **Model configuration-based validation**

Model Configurations: FV3-LAM

Name	Microphysics Scheme	Planetary Boundary Layer Scheme	Surface Layer	Land Surface Model
Control	Thompson	MYNN	GFS	Noah
MP-NSSL	National Severe Storms Laboratory	MYNN	GFS	Noah
MP-MG	Morrison-Gettelman	MYNN	GFS	Noah
PBL-SH	Thompson	Shin-Hong	GFS	Noah
PBL-EDMF	Thompson	EDMF	GFS	Noah
LSM-RUC_SFC-GFS	Thompson	MYNN	GFS	RUC
LSM-RUC_SFC-MYNN	Thompson	MYNN	MYNN	RUC



Methodology

1. Object-based Threat Score (OTS) :

$$OTS = \frac{1}{A_f + A_o}\left[\sum_{p=1}^{P} I^p (a_f^p + a_o^p)\right]$$

A_f and A_o : Area of all forecasted and observed objects.

P : number of matched forecast and observation object pairs

I^p : interest score between the matched forecast and observation object

a_f^p and a_o^p : areas of the forecast and observation objects in the matched pair

Methodology

1. Object-based Threat Score (OTS) :

$$OTS = \frac{1}{A_f + A_o}\left[\sum_{p=1}^{P} I^p \left(a_f^p + a_o^p\right)\right]$$

A_f and A_o : Area of all forecasted and observed objects.

P : number of matched forecast and observation object pairs

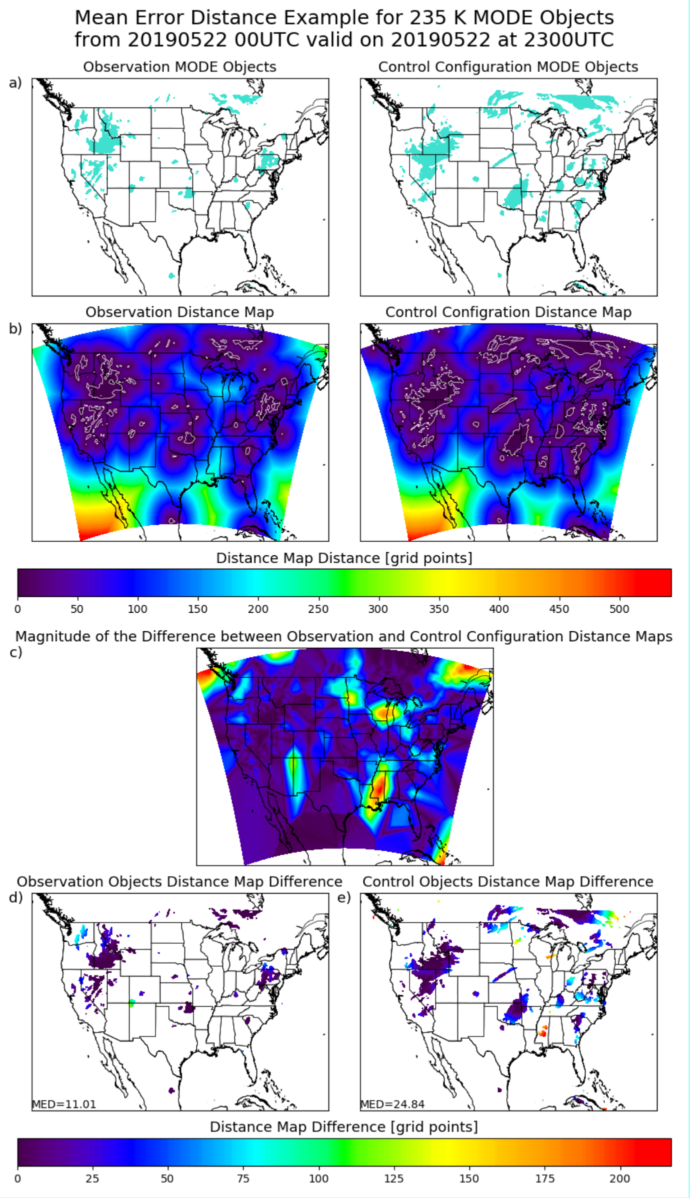
I^p : interest score between the matched forecast and observation object

a_f^p and a_o^p : areas of the forecast and observation objects in the matched pair

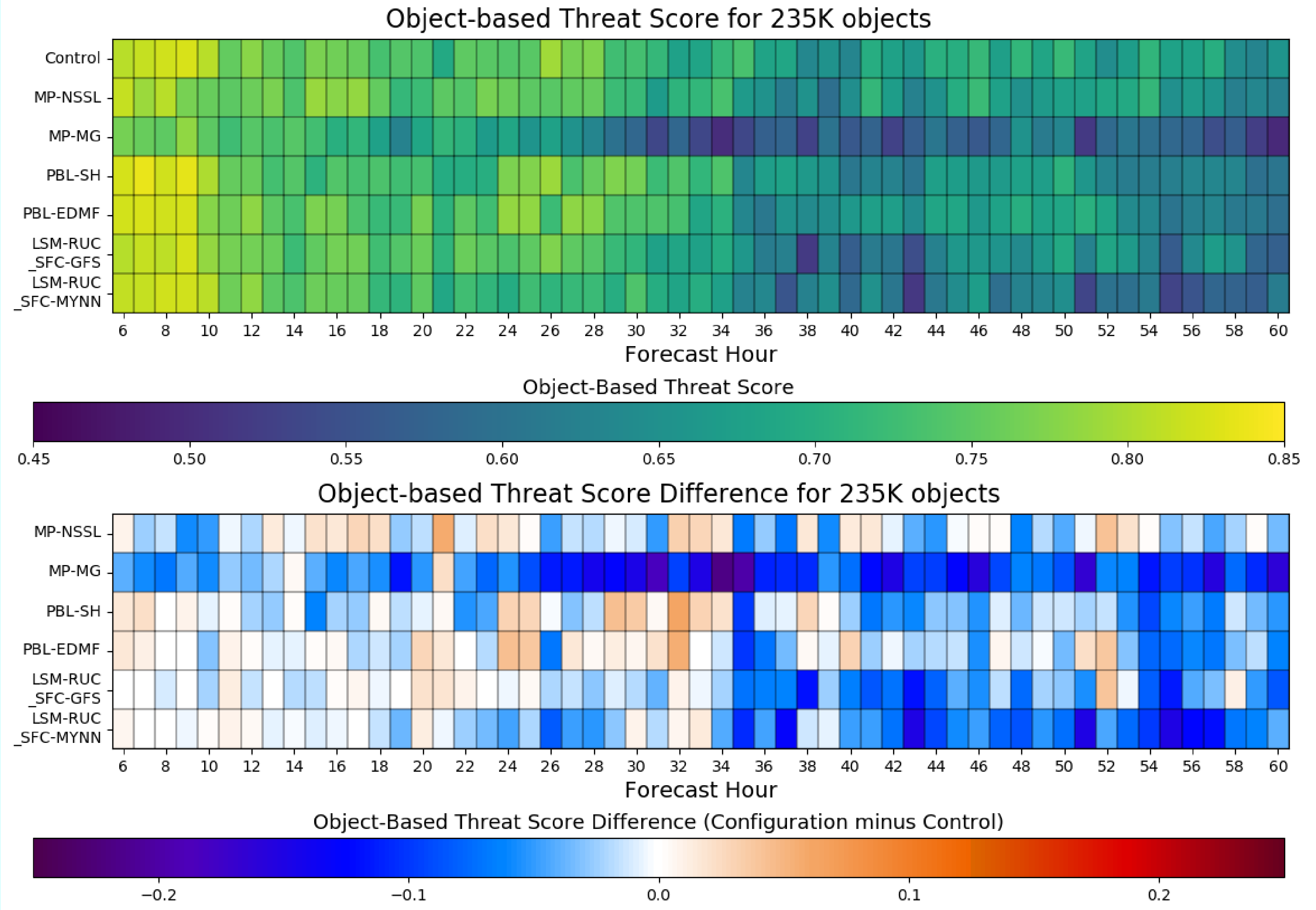
2. Mean Error Distance (MED):

Calculates distance between every grid point identified as a forecast (observation) object to the closest grid point identified as an observation (forecast) object.

- distance map: shortest distance between every grid point and the nearest grid point identified as an object
- **MED from forecast to observation ≠ MED from observation to forecast**

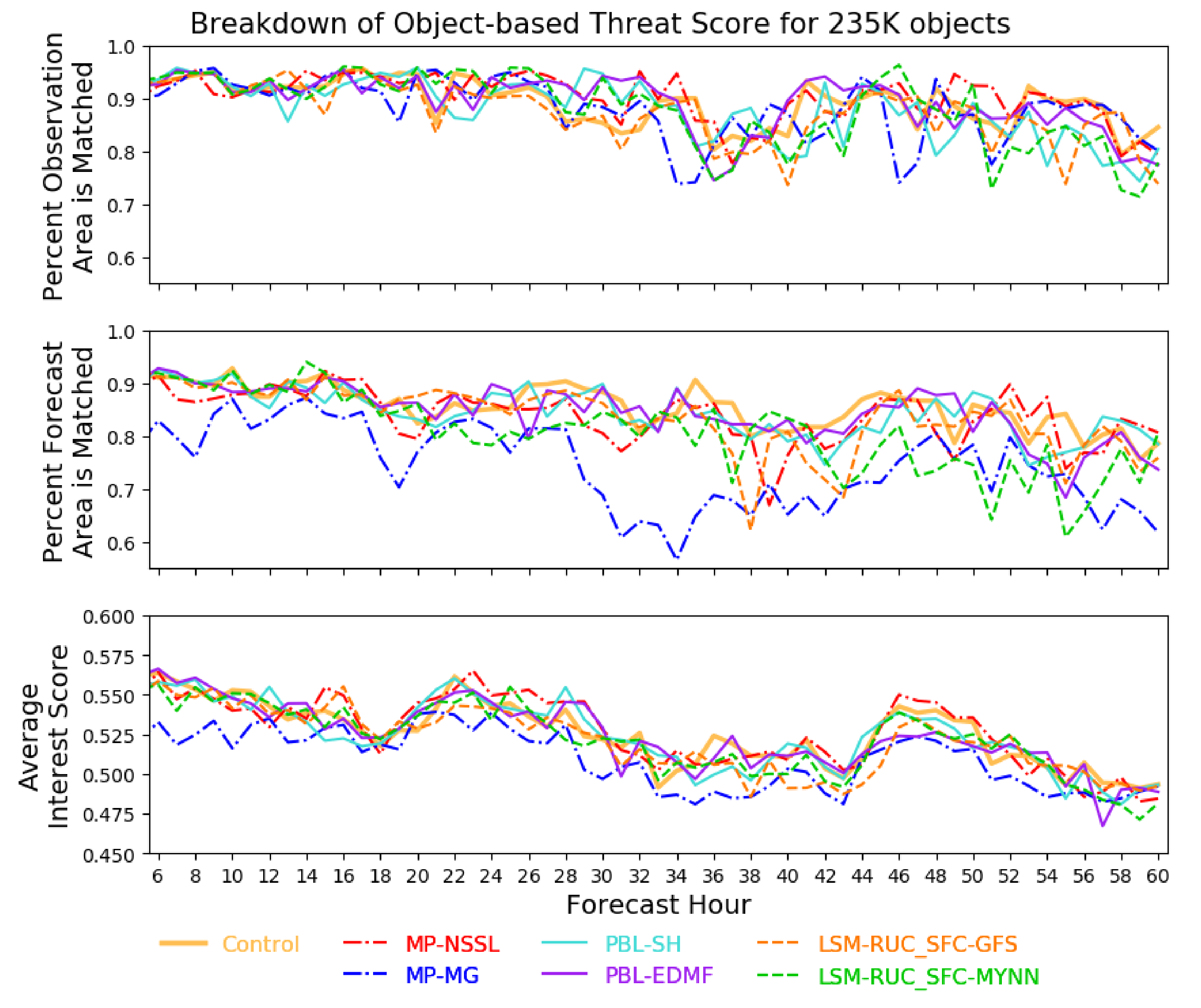


Results: Object-Based Threat Score



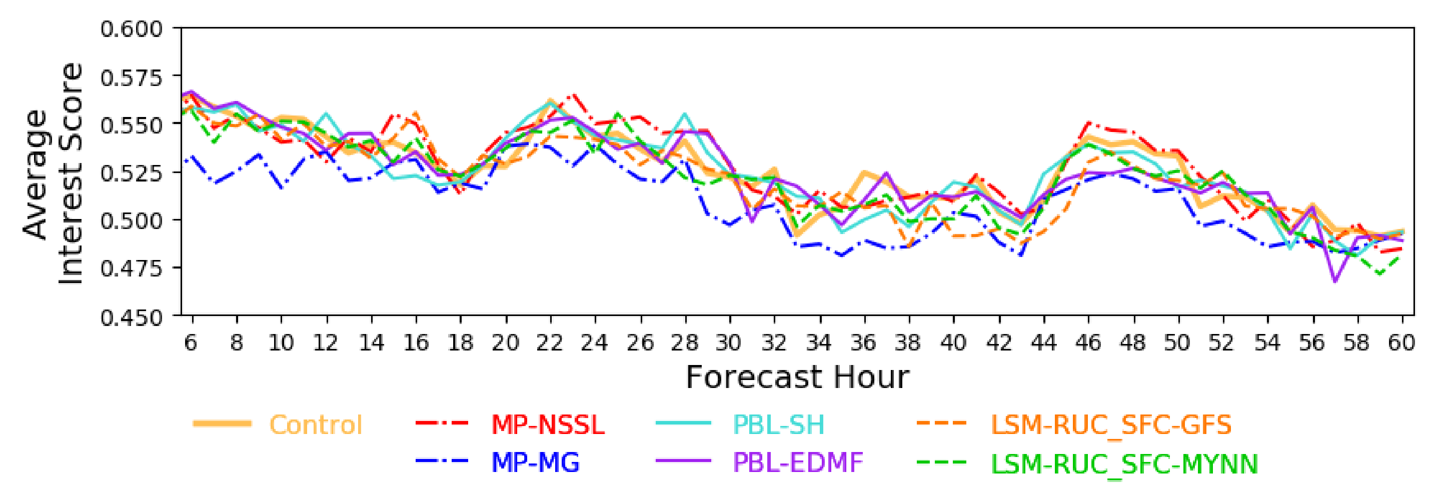
- Control has the highest average OTS.
- MP-MG has the lowest average OTS.
- LSM-RUC_SFC-MYNN has the steepest decline in OTS by forecast hour.
  - Correlated with an increased number of objects
- Parameter changes have a neutral to positive impact on OTS in early FHs compared to Control.

Results: Object-Based Threat Score



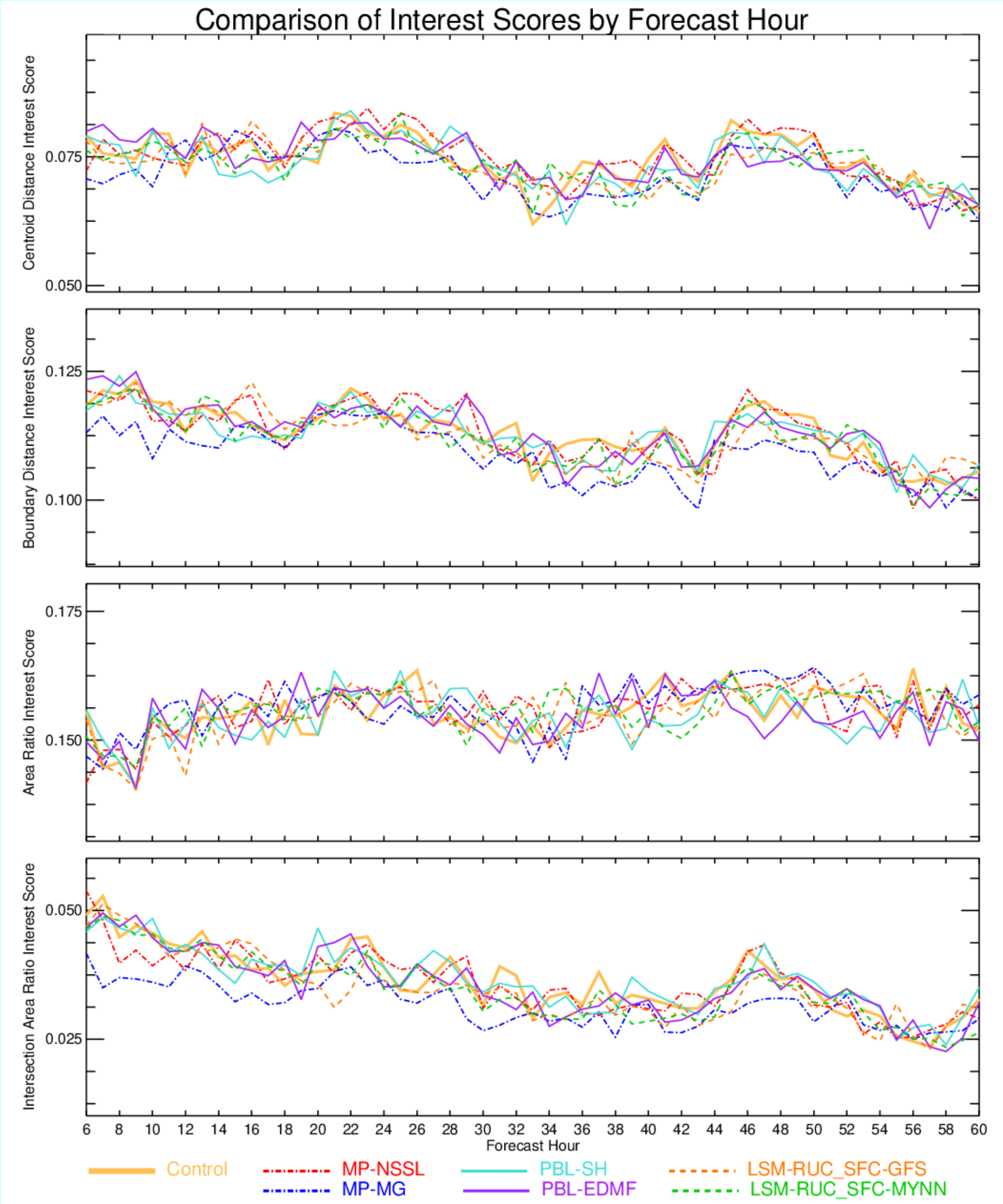
- Similar Percent of Observation Objects matched ($\frac{a_o}{A_o}$)
- MP-MG much lower Percent Forecast Objects matched ($\frac{a_f}{A_f}$)
  - MP-MG has highest number of objects.
- Local maximum in interest scores due to lower distance between matched objects ($\frac{1}{P}\sum_{p=1}^{P} I^p$)

Results: Object-Based Threat Score



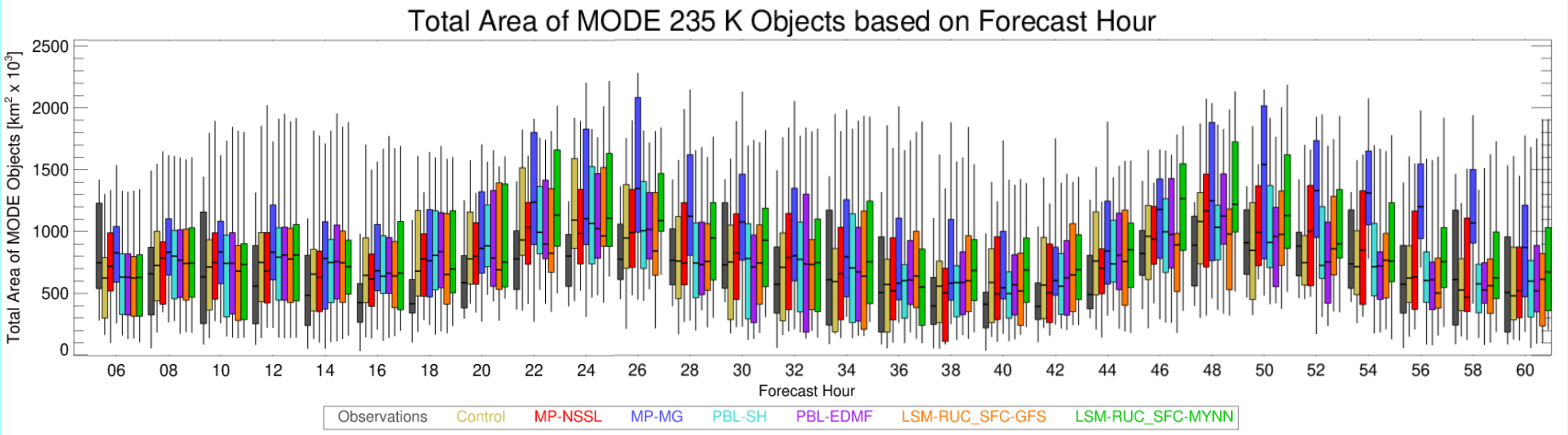
- Diurnal Cycle in Average Interest Scores corresponding to 5-9 pm Central Daylight Time.

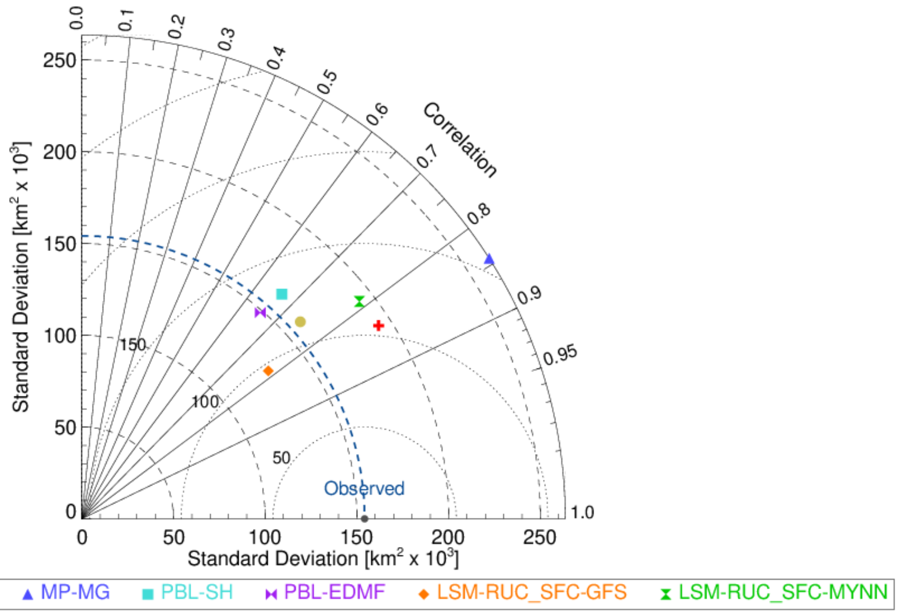
Results: Object-Based Threat Score



- Diurnal Cycle in Average Interest Scores corresponding to 5-9 pm Central Daylight Time.
- Break down Interest Scores into 4 main components.
  - Diurnal cycle in Centroid Distance and Boundary Distance
  - Remove this attributes, diurnal cycle much weaker.

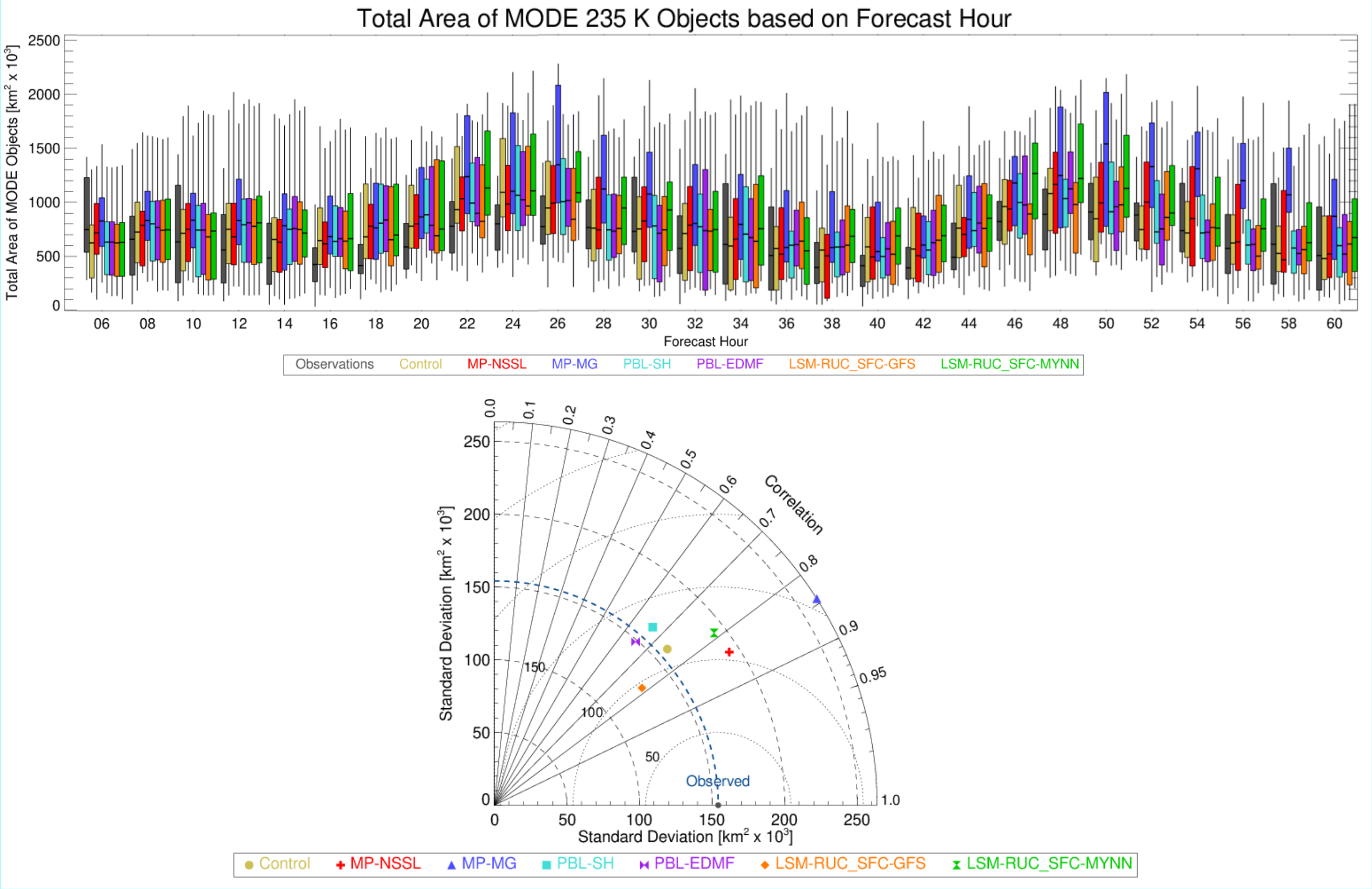
Results: Area of MODE objects





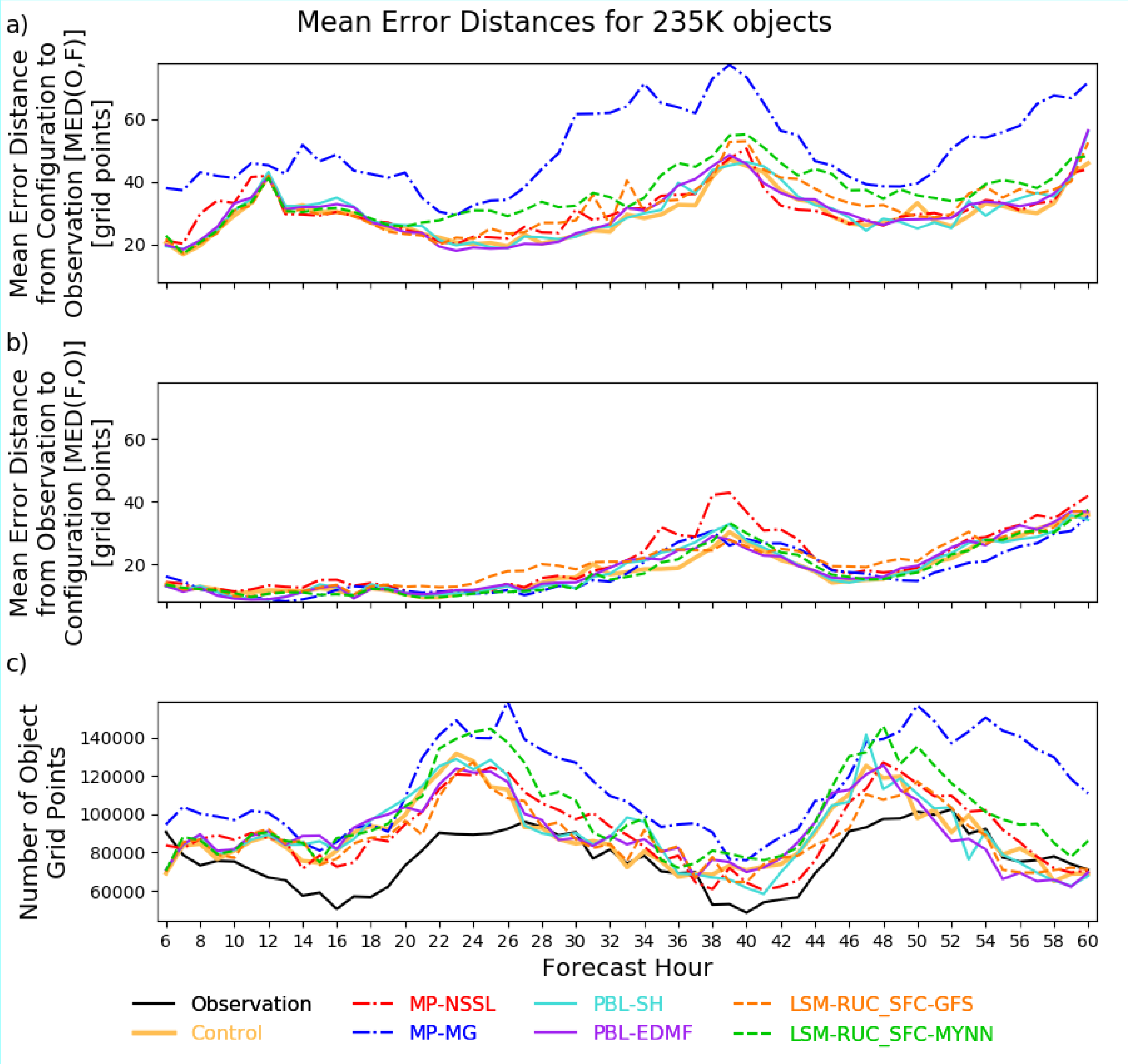
- Taylor diagram: bottom image
  - Pearson correlation coefficient (solid lines)
  - standard deviation (dashed lines)
  - root-mean-square difference along the dashed semi-circles in the plot.
    - mean squared error after accounting for biases

Results: Area of MODE objects



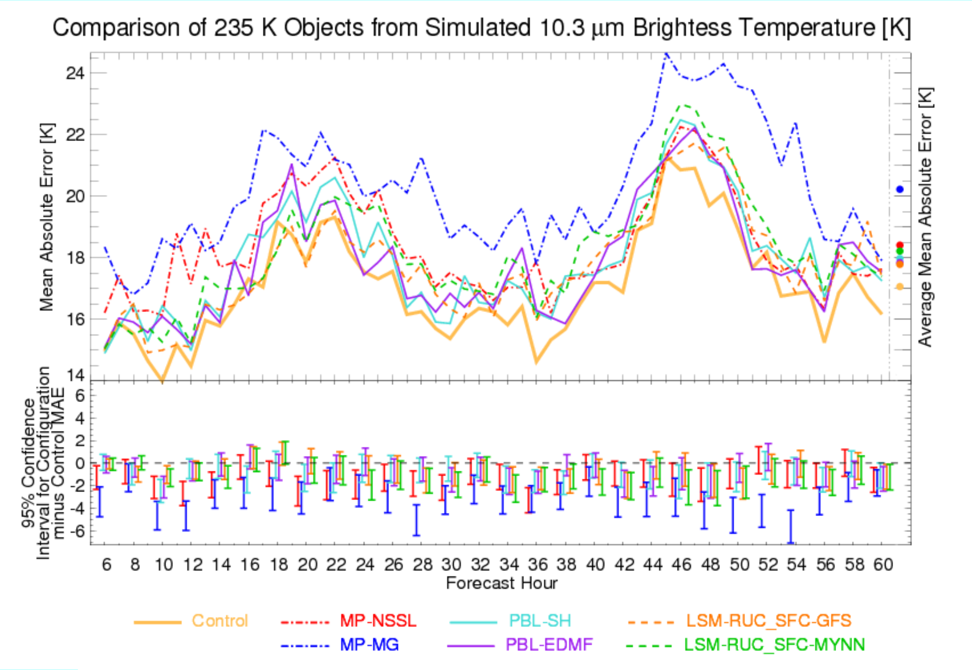
- Diurnal cycle like OTS.
  - higher OTS does not correspond to more overlapping area.
- MP-MG has the largest amount of area encompassed by objects
  - largest spread in median area
- Changes to the PBL results in less correlation between areas.
- LSM-RUC_SFC-GFS has lowest RMSD.

Results: Mean Error Distance



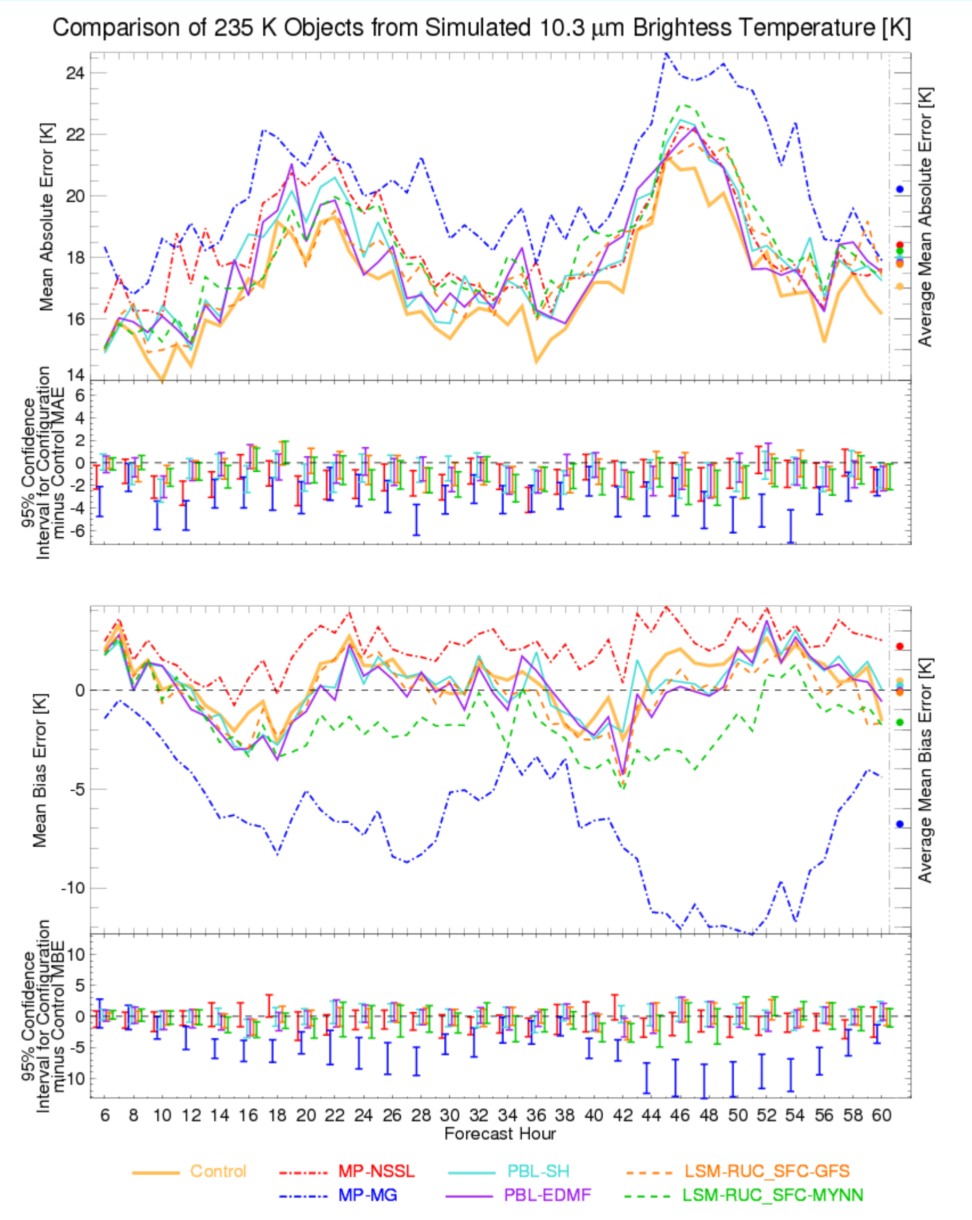
- MED from forecast to observation > MED from observation to forecast
  - Due to more forecast object grid points than observations.
- MP-MG has highest MED from forecast to observation.
- Diurnal cycle: More grid points and lower MED
  - Not indicate more overlapping grid points.

Results: Object-based MAE and MBE



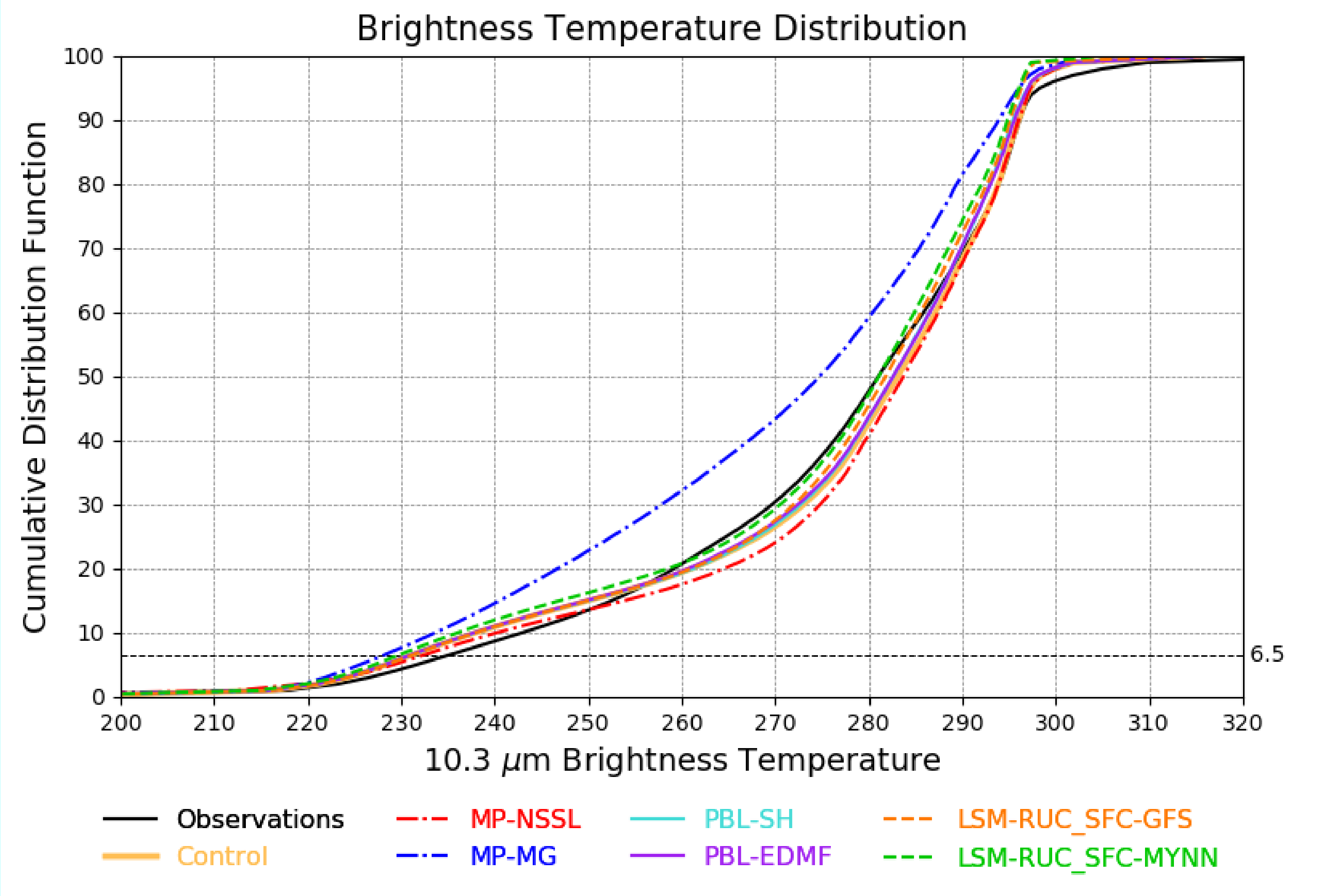
- Diurnal cycle in model accuracy
  - Opposite of OTS (MAE high when OTS low)
  - Centroid distance removed for MAE.
- MP-MG highest MAE, Control lowest.
  - MP-MG difference from Control statistically significant.
  - MP-NSSL next highest MAE.

Results: Object-based MAE and MBE



- Diurnal cycle in model accuracy
  - Opposite of OTS (MAE high when OTS low)
  - Centroid distance removed for MAE.
- MP-MG highest MAE, Control lowest.
  - MP-MG difference from Control statistically significant.
  - MP-NSSL next highest MAE.
- Changing microphysics scheme has largest impact on MBE (and MAE)
  - MP-MG low bias in object BTs
  - MP-NSSL has a high bias in object BTs.
  - MBE is correlated with an increased number of forecast object grid points compared to the observation object.

Brightness Temperature Bias



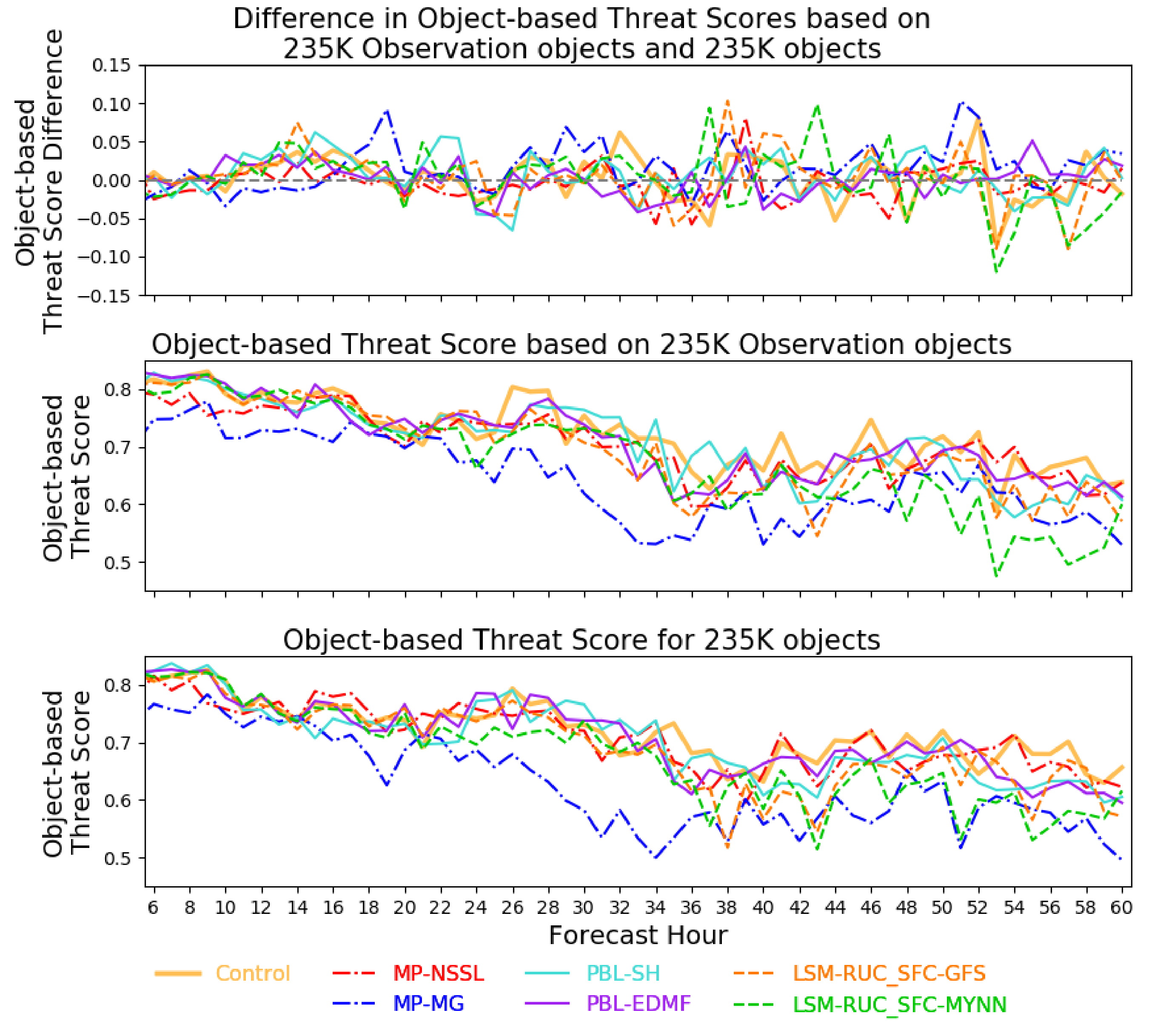
BT corresponding to the 6.5th percentile:

- Observations: 235.0 K
- Control : 231.0 K
- MP-NSSL: 232.3 K
- MG-MG: 228.1 K
- PBL-SH: 230.9 K
- PBL-EDMF: 230.9 K
- LSM-RUC_SFC-GFS: 231.1 K
- LSM-RUC_SFC-MYNN: 229.7 K

Results: Brightness Temperature Bias

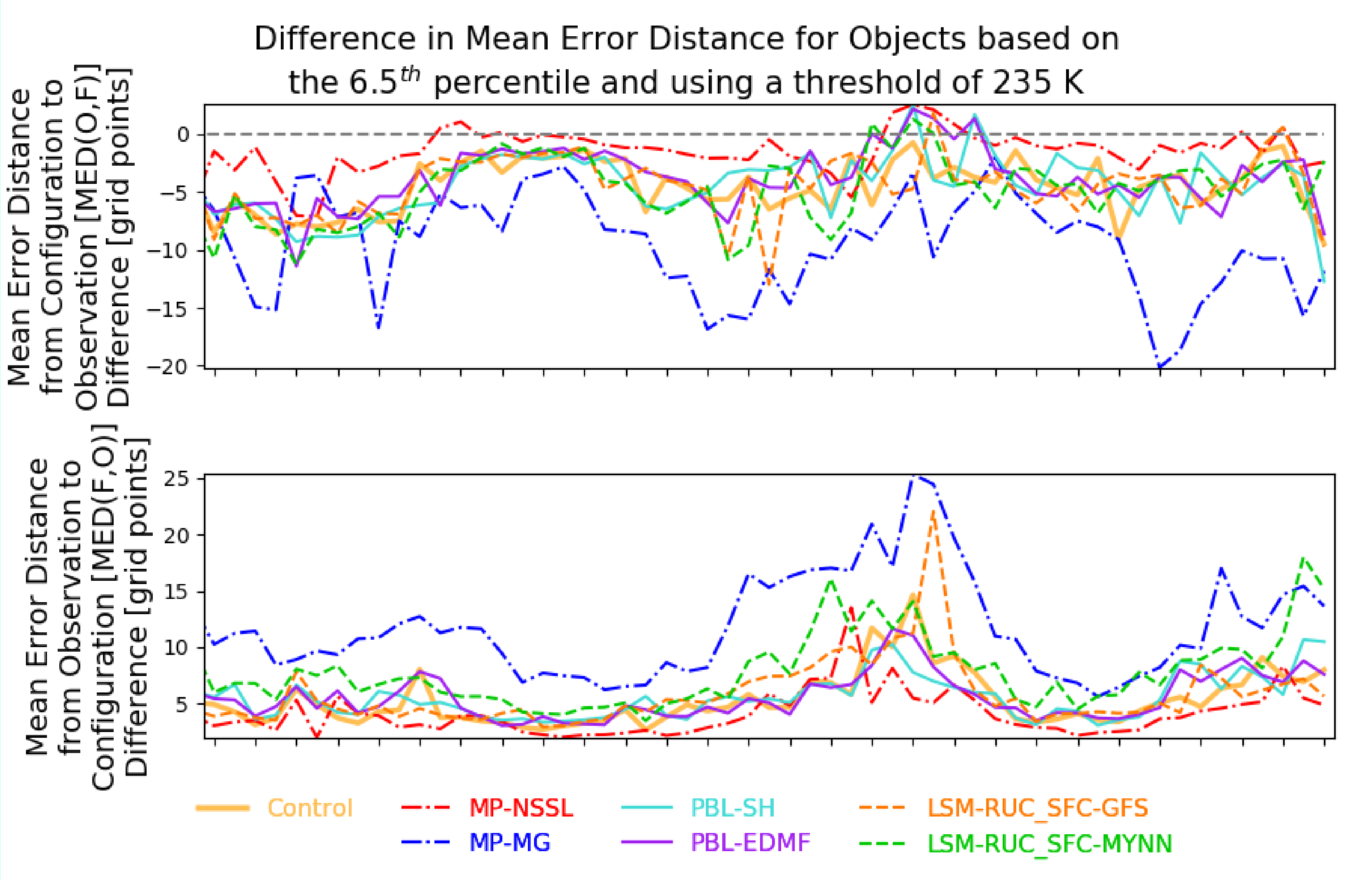
- Overall, little change from 235.0 K threshold.

Results: Brightness Temperature Bias



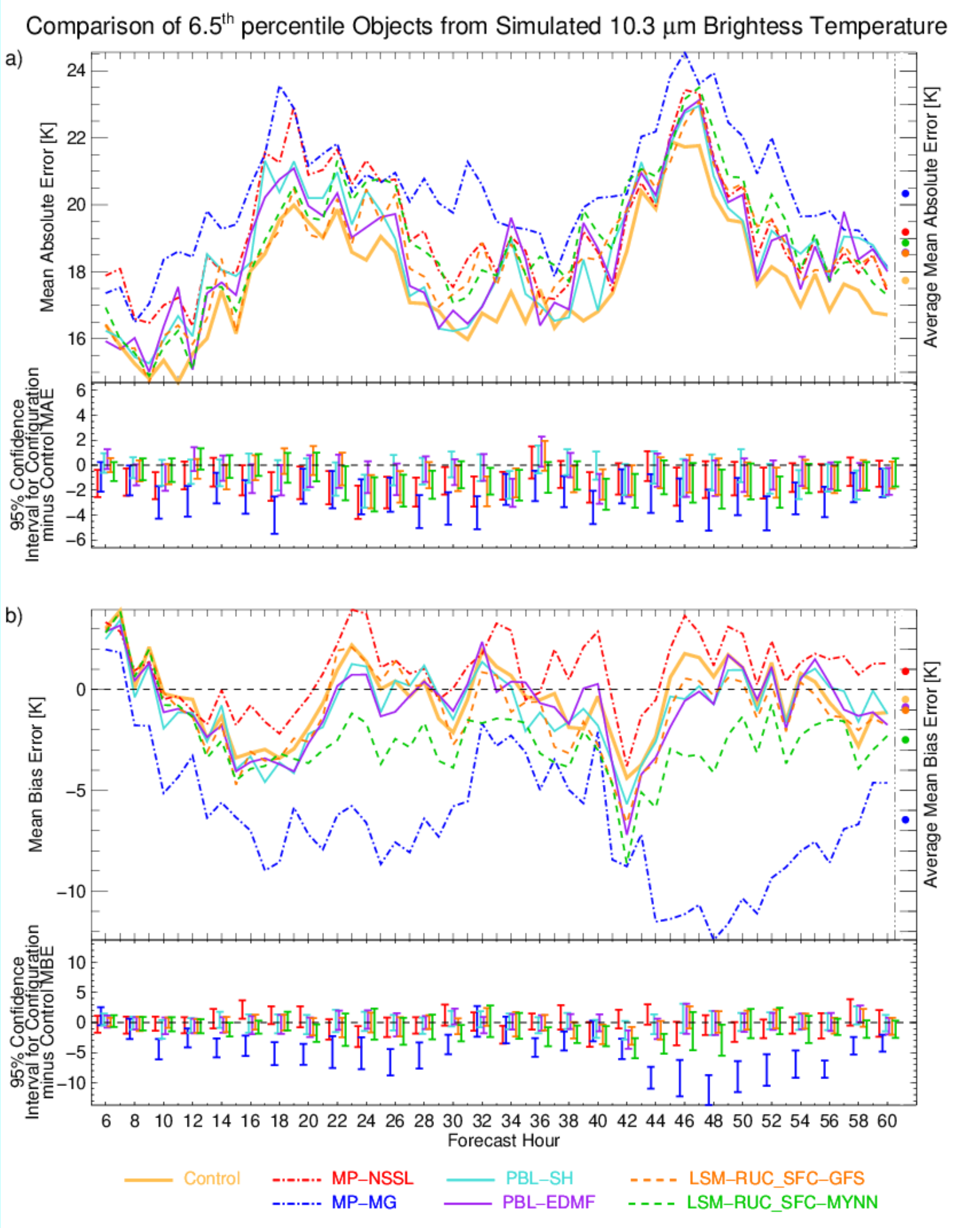
- Overall, little change from 235.0 K threshold.
- OTS:
  - Neutral changes
  - MP-MG still lowest.

Results: Brightness Temperature Bias



- Overall, little change from 235.0 K threshold.
- OTS:
  - Neutral changes
  - MP-MG still lowest.
- MED:
  - MED from forecast to observation decreases
  - MED from observation to forecast increases
    - MP-MG highest

Results: Brightness Temperature Bias



- Overall, little change from 235.0K threshold.
- OTS:
  - Neutral changes
  - MP-MG still lowest.
- MED:
  - MED from forecast to observation decreases
  - MED from observation to forecast increases
    - MP-MG highest
- MAE/MBE:
  - MP-MG highest MAE
  - MP-MG lowest MBE

Conclusions

1. Changing the microphysics scheme from Thompson:
- Morrison-Gettelman results in lower BTs, which are overall less accurate.
- NSSL results in higher BTs, which are also less accurate than Control.

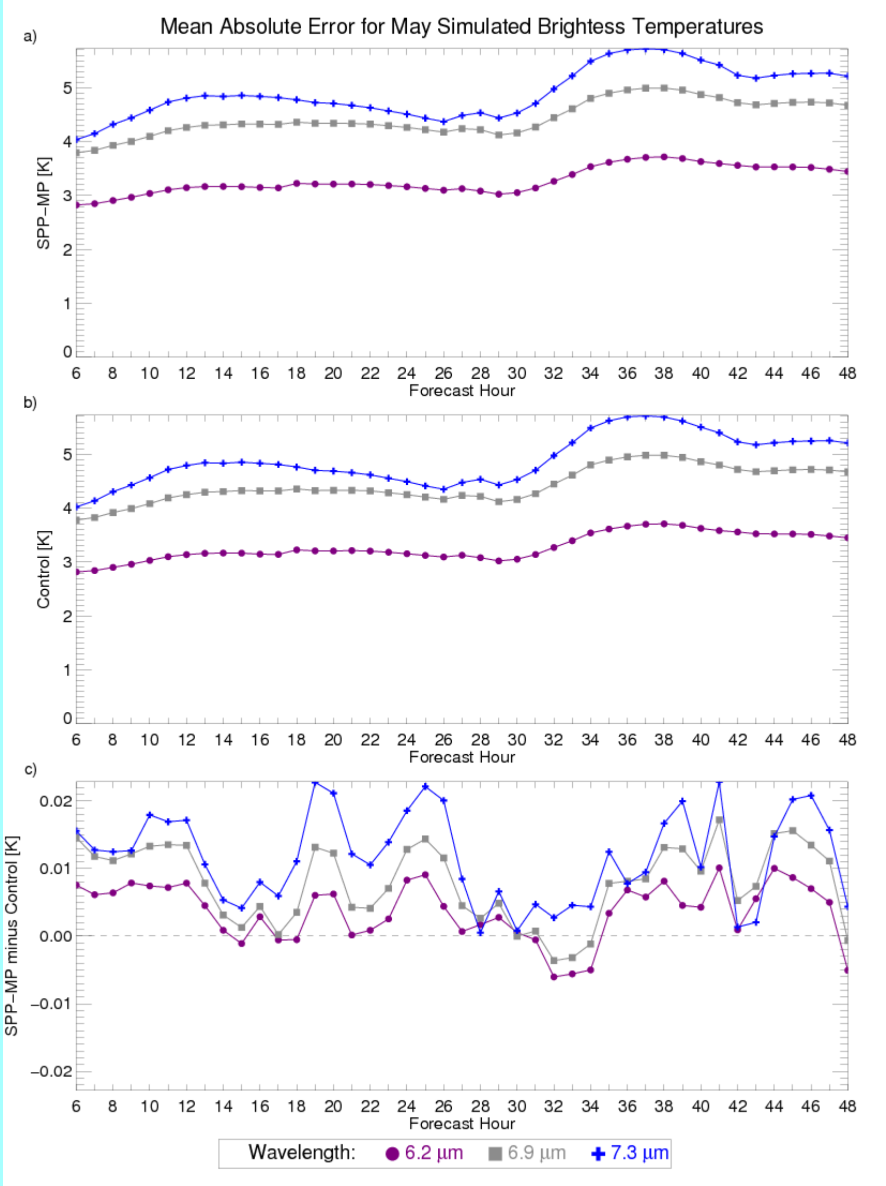
2. Changing the PBL scheme from MYNN:
- reduces the high BT bias, though the BTs are less accurate based on the OTS and MAE.

3. Updates to the surface also reduce the accuracy of simulated BTs.

4. Accounting for model bias when calculating the OTS does not impact the relative performance of each model configuration.

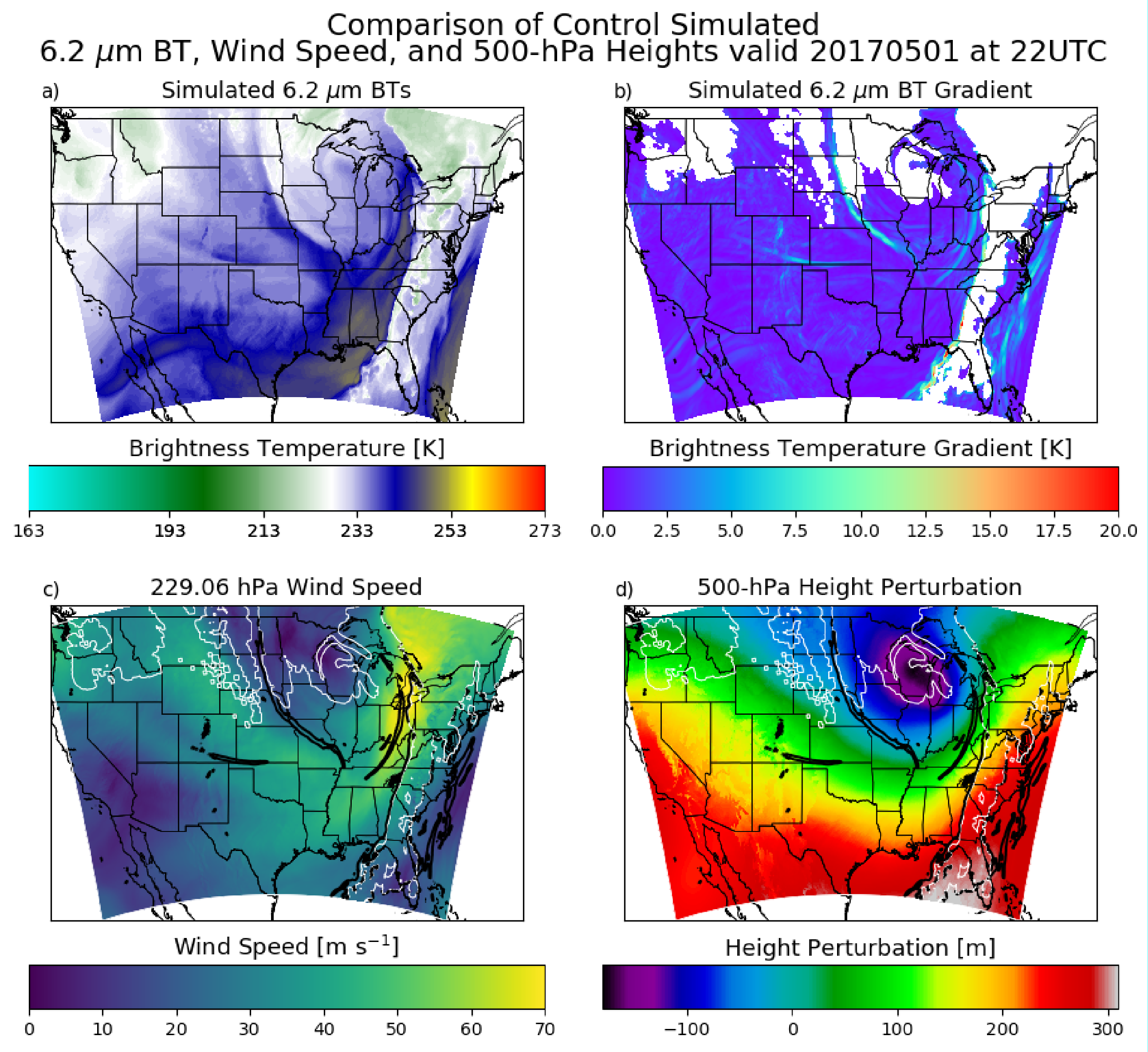
Future Work

1. Expand beyond the 10.3 µm brightness temperatures.
    - Water vapor BTs



Future Work

1. Expand beyond the 10.3 μm brightness temperatures.
   - Water vapor BTs
   - Can we correlate features in the WV BTs to synoptic features?



Future Work

1. Expand beyond the 10.3 µm brightness temperatures.
    - Water vapor BTs
    - Can we correlate features in the WV BTs to synoptic features?

2. Other model fields:
    - Radar reflectivity
    - Snow cover

Future Work

1. Expand beyond the 10.3 μm brightness temperatures.
   - Water vapor BTs
   - Can we correlate features in the WV BTs to synoptic features?

2. Other model fields:
   - Radar reflectivity
   - Snow cover

Questions??

Email:
sarah.griffin@ssec.wisc.edu