Aeolus Wind Retrieval Assimilation at the GMAO

Will McCarty

Global Modeling and Assimilation Office
NASA Goddard Space Flight Center

UMD Seminar
28 October 2019

Introduction

From a physics perspective, it is relatively straight forward to measure the mass field from space

- The atmosphere is made of molecules
- Those molecules have a Temperature
- Those molecules emit photons at said Temperature
- Some of those photons reach a detector strapped to a satellite up in space

It’s less straightforward to measure the wind in space

Introduction

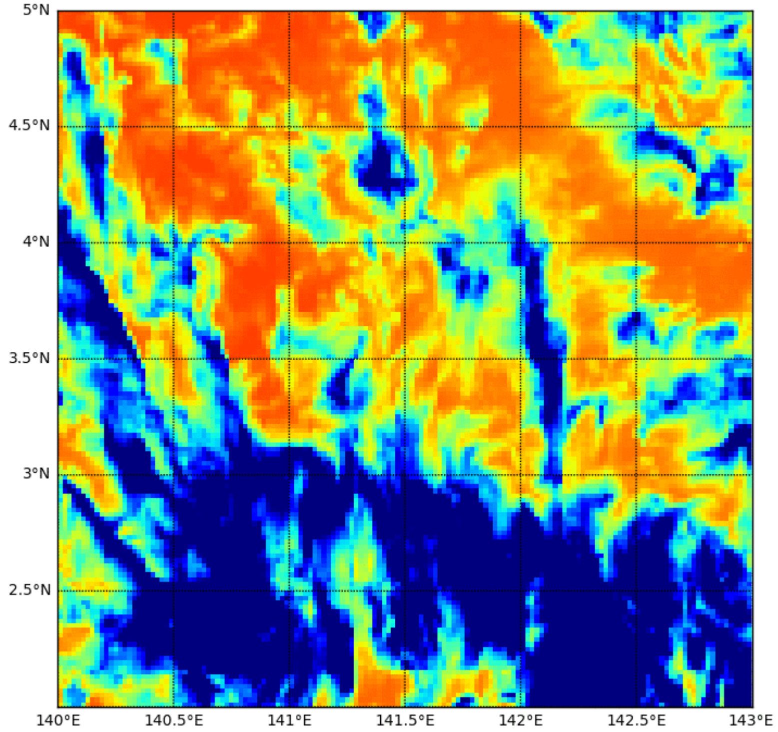
From a physics perspective, it is relatively straight forward to measure the mass field from space

- The atmosphere is made of molecules
- Those molecules have a Temperature
- Those molecules emit photons at said Temperature
- Some of those photons reach a detector strapped to a satellite up in space

It’s less straightforward to measure the wind in space

- Historically, this has been performed using Atmospheric Motion Vectors (AMVs)
- AMVs track satellite-observed features (e.g. clouds, water vapor gradients) in space ($\delta \mathbf{x}$) and time (δt)
- For this, spatial resolution is most important (features need to be resolved), so these data have been historically derived from satellite imagery

But fundamentally, AMVs *infer* the wind



From Indirect to Direct…Enter Aeolus

ESA has launched the Aeolus satellite to measure wind directly from space

- Measuring the doppler shift in the backscattered light from a 355 nm lidar 30° off-nadir
- Measurements from Clouds and Aerosols
  - Sharp backscatter signal (Mie)
- Measurements in clear sky
  - Broad backscatter signal (Rayleigh)
- Lidars are big and expensive and difficult to point
  - Only one wind component is measured horizontally

The GMAO has been involved in Aeolus for over a decade

- In 2009, I arrived and started preparing for Aeolus as a post-doc (under the auspices of the JCSDA and Aeolus preparadness)
- Then the satellite was delayed indefinitely, and I focused elsewhere
- Accepted as a member of the ESA Aeolus Cal/Val Team to help validate the observations before wide-release
- This has been coordinated across collaborations with NOAA (NWS/EMC: I. Genkova, NESDIS/STAR: Aeolus Tech. Maturation Program)

Assimilation Perspective

The Global Modeling and Assimilation Office (GMAO) was prepared to monitor and assimilate the Aeolus L2B horizontal line of sight (HLOS) winds

- Pre-launch test data allowed for ingest on Day 1 of L2B BUFR release via ECMWF
- A statement to pre-launch preparedness both in terms of JCSDA and ECMWF/ESA delivery of useful proxy data

A control experiment (without Aeolus) using standard GMAO test configuration has been run for the initial assessment of data record through 20 Jan 2019

- The Goddard Earth Observing System (GEOS) atmospheric data assimilation system
- Model: GEOS Model, ~1/4° (C360 cubed sphere grid)
- Analysis: GSI Hybrid 4D-EnVar, full global observing system, 1/2° analysis, 1° (C90) ensembles
  - Note, the GSI assimilation procedure is co-developed with NOAA partners, thus the implementation methods are similar

Monitoring and departure statistics have been calculated comparing Aeolus with the control ‘background’ (short-term forecast) fields for this period

- No active assimilation experimentation yet

Initial Departure Assessment

Each wind retrieval is classified by retrieval type (Rayleigh or Mie) and as Cloudy or Clear

- Four subtypes are considered: Rayleigh/Clear, Rayleigh/Cloudy, Mie/Clear, and Mie/Cloudy

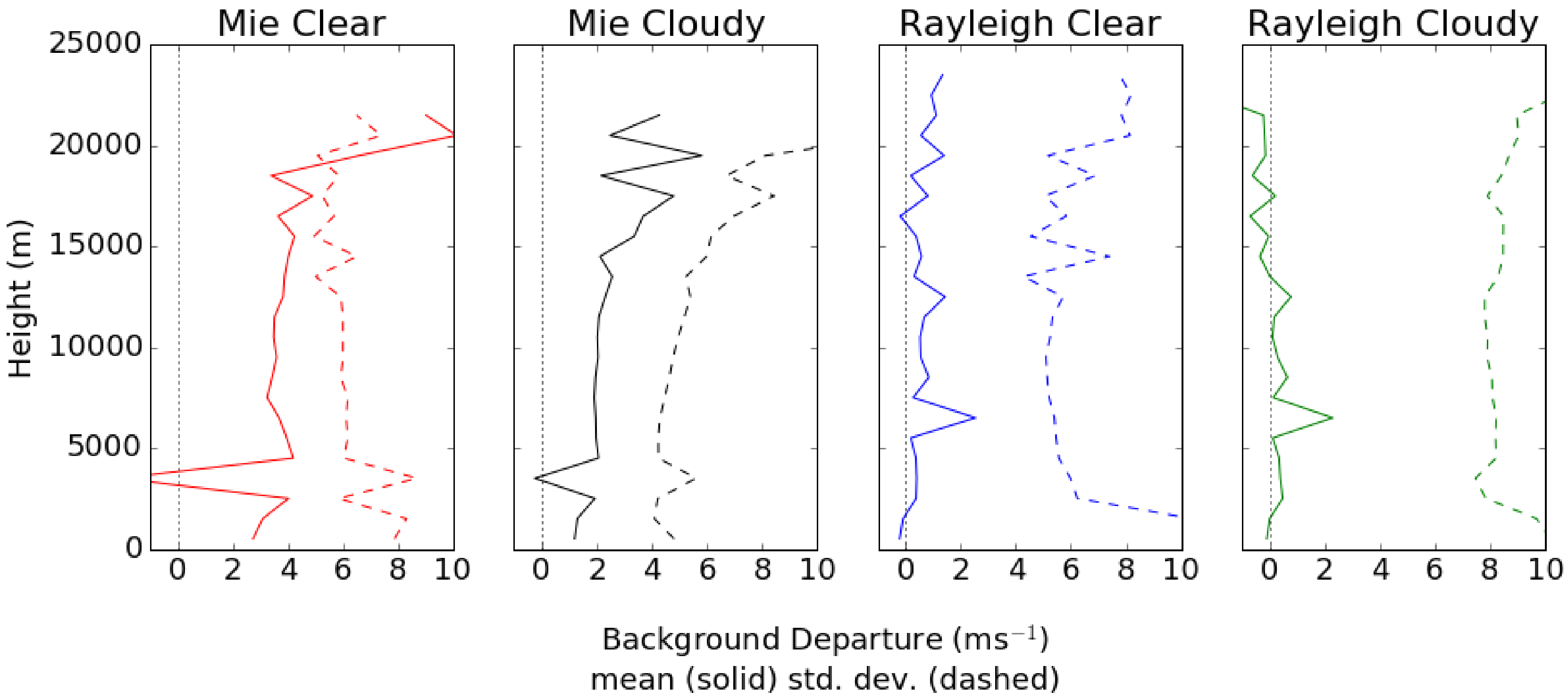
All observations are assimilated as single-point HLOS observations at the centroid

- No effort is made to account for along-track observation resolution

Initial QC procedures:

- Discard if |HLOS| > 1000 ms^{-1} (outright discard unrealistic values)
- Reject if Confidence Flag indicates invalid retrieval
- For all results hereafter, a ±20 ms^{-1} gross check is applied to the departures for initial results

Global Statistics: 3 Sept – 30 Nov 2018



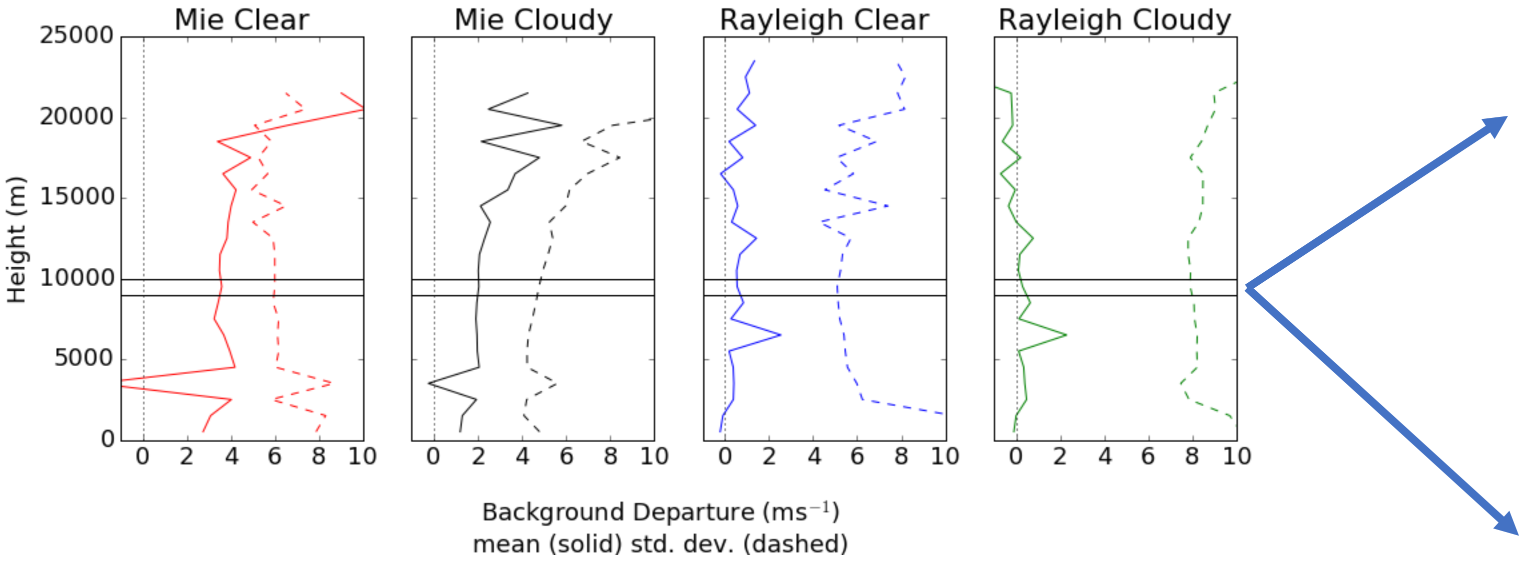
Initial comparisons to our model showed known issues

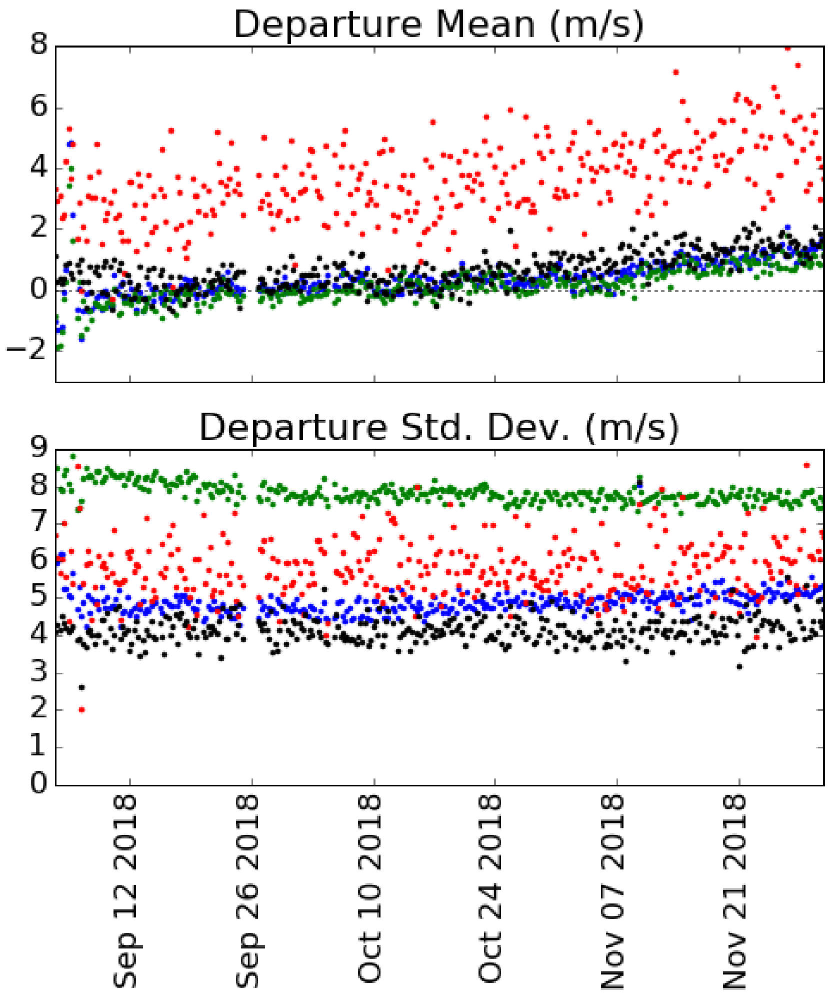
- Mie retrievals are heavily biased: approaching (cloudy), exceeding (clear) 2 ms^{-1}
- Rayleigh retrievals less-so
- Bias > 1 m/s for cloudy; > 2 m/s for clear above ~3500 m

’Hot Pixel’ issues seen in Mie and Rayleigh

- Bias spike between 6000-6500 m

Global Statistics: 3 Sept – 30 Nov 2018

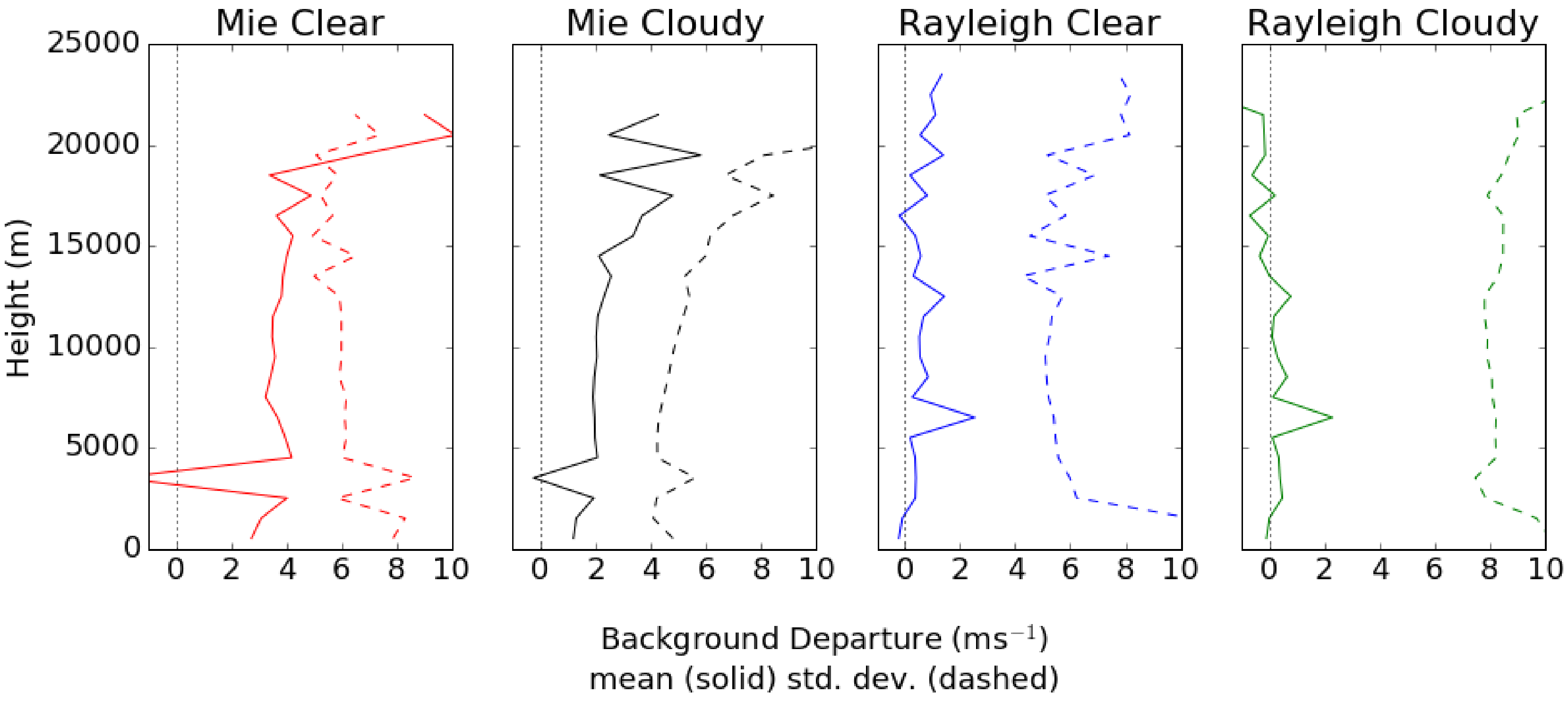




From 9000-10000 m, the larger bias is readily apparent in time series

- All subtypes show some hint of an increase in bias w.r.t. time

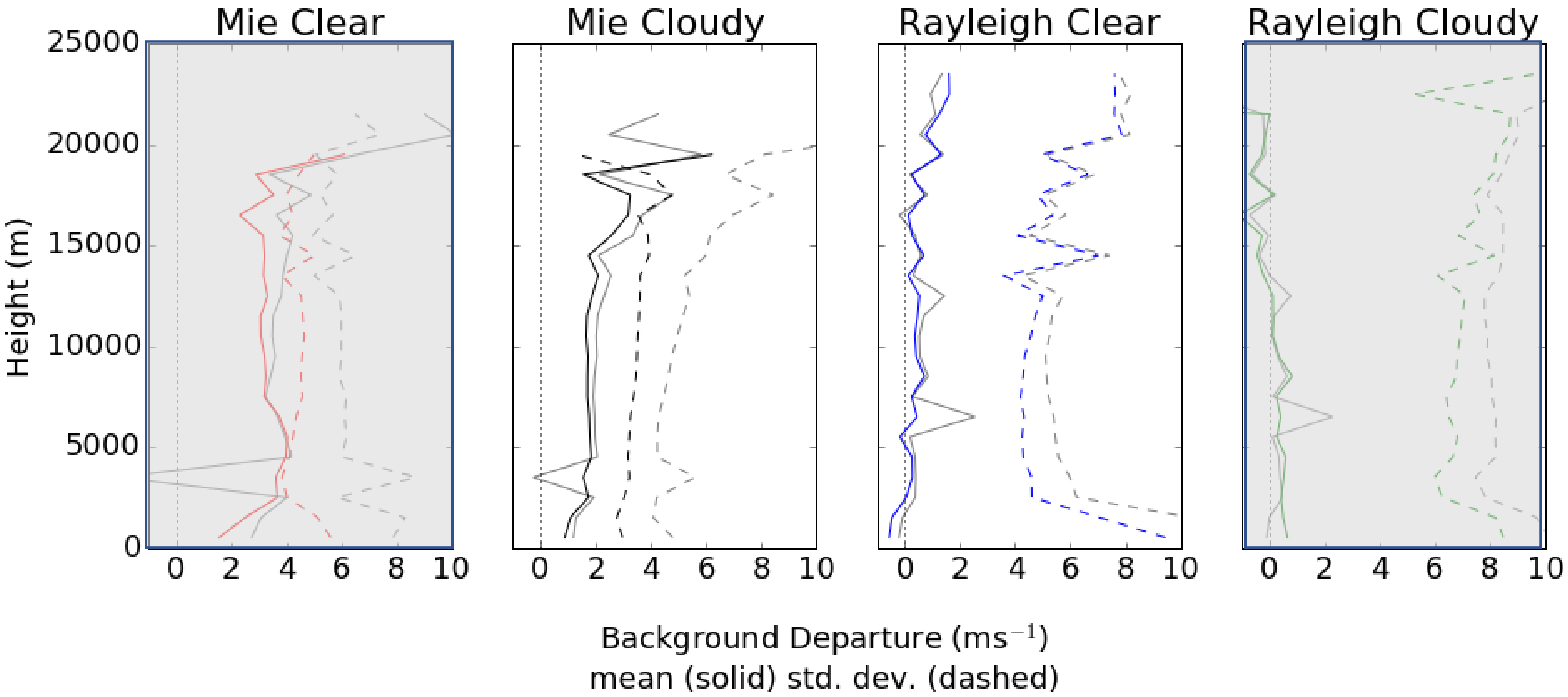
Quality Control



Michael Rennie (ECMWF) & the Aeolus Team provided a list of suggested quality control and bias corrections to target the best observations for assimilation

- Note: 12 Sept – 16 Oct 2018 is the targeted assimilation inter-comparison period

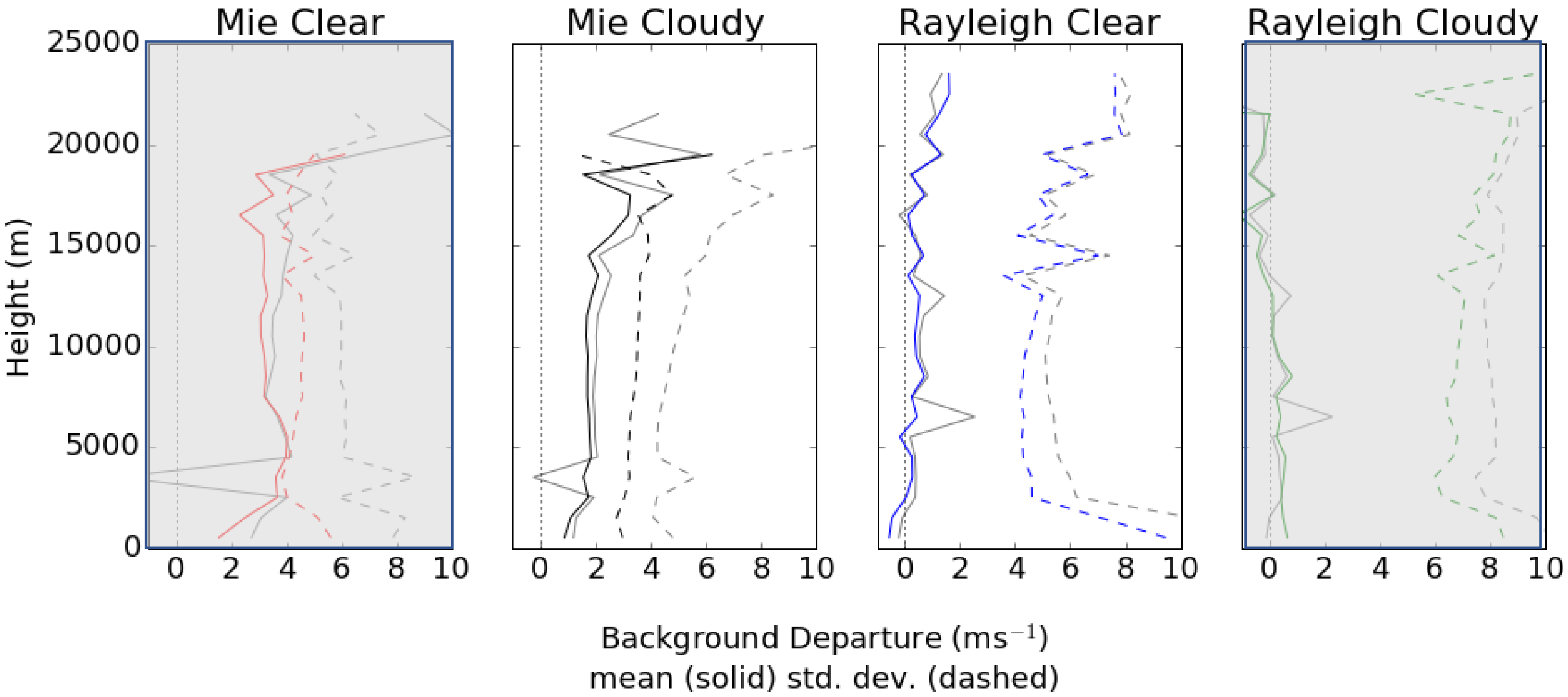
Quality Control



By applying the quality control procedures, it can be seen that for the two suggested use types (Mie Cloudy, Rayleigh Clear) that the standard deviations are dramatically reduced and the hot spot regions are avoided/resolved

- Good observations may be lost using the ECMWF-suggested pressure-based blacklisting for hot spot avoidance

Quality Control



At this point, two key issues remain prior to initial assimilation experiments:

- Large biases
- Observation error handling

Handling Bias

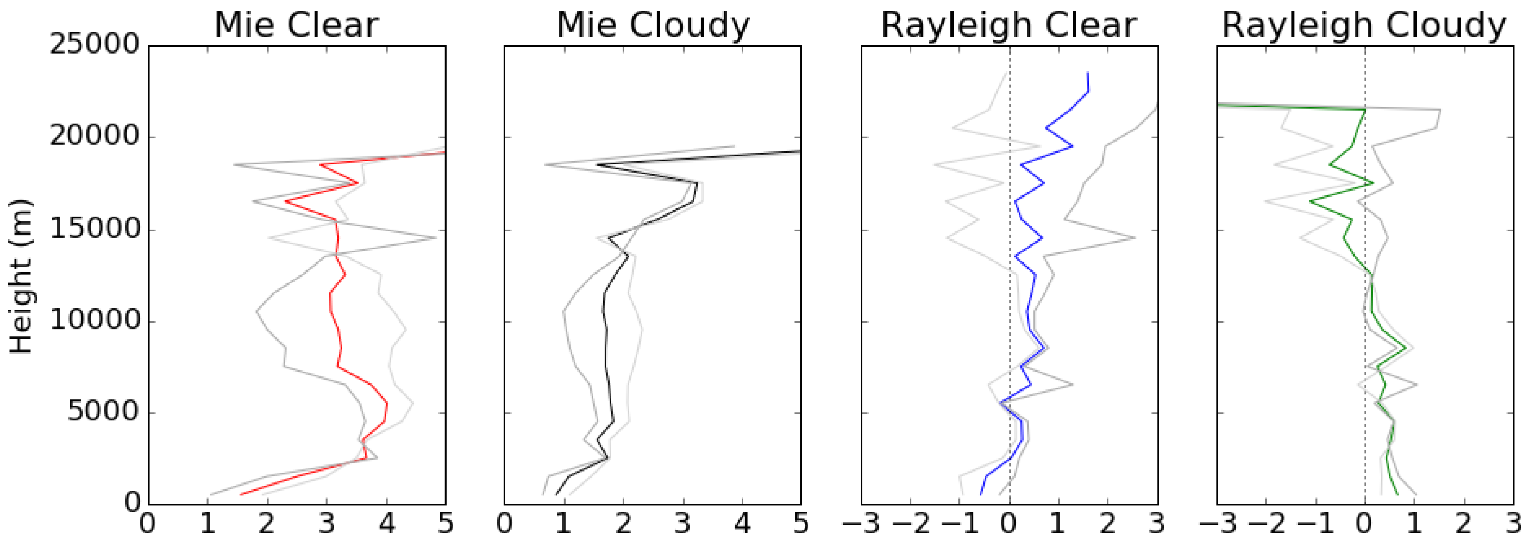
From both this study and ECMWF suggestion, two components of bias need to be removed prior to assimilation

- ECMWF suggested applying a -1.35 ms^{-1} bias correction to Mie Cloudy winds
- Additionally, ECMWF noted substantial biases as a function of orbit node, which cancel out in global averaging
  - Ascending/descending biases expected to be removed/fixed via improved calibration

Handling Bias

From both this study and ECMWF suggestion, two components of bias need to be removed prior to assimilation

- ECMWF suggested applying a bias correction of -1.35 ms^{-1} to Mie Cloudy winds
- Additionally, Gert-Jan Marseille (KNMI) and others have noted substantial biases as a function of orbit node, which cancel out in global averaging
  - Ascending/descending biases expected to be removed/fixed via improved calibration



Mean Background Departure (ms^{-1})
Ascending (Light Gray) & Descending (Dark Gray) Node

Bias Correction

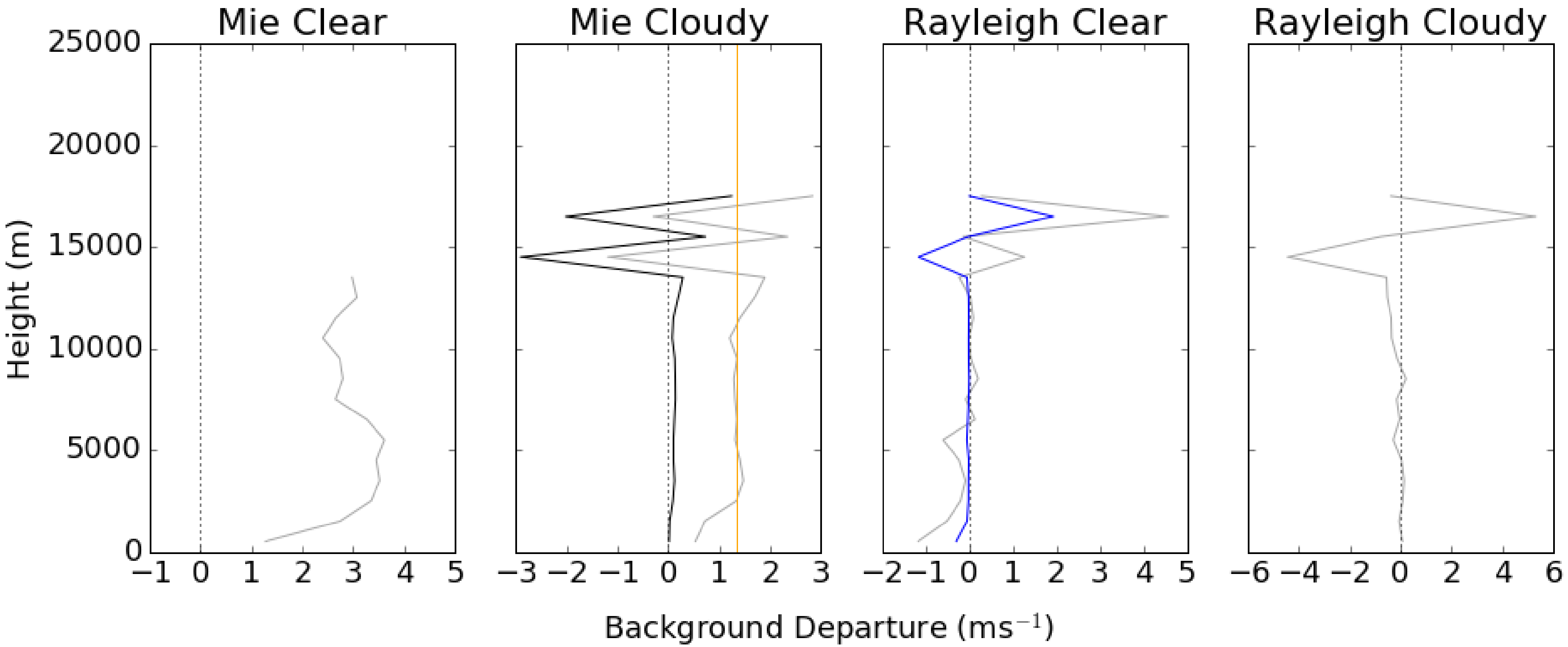
GMAO initial implementation allows for two bias corrections to be applied as a function of height

- Global correction: a constant value
- Node correction: Node*constant, where Node == 1 for ascending node; -1 for descending node

Values were determined from:

- GMAO-derived departures
  - Not ECMWF-suggested correction
- 12-30 Sept 2018
  - A period of high-quality observations
  - Temporal drifts in bias will not be captured with static, non-adaptive bias correction

Departures after Bias Correction



Before (gray), After (color), & ECMWF-Suggested (orange) Bias Correction

Bias correction successfully removes most of the tropospheric bias

- Some issue above ~ 14 km
- Disconnect between mean and nodal BC (sum of each component != zero)

Comparison against AMVs

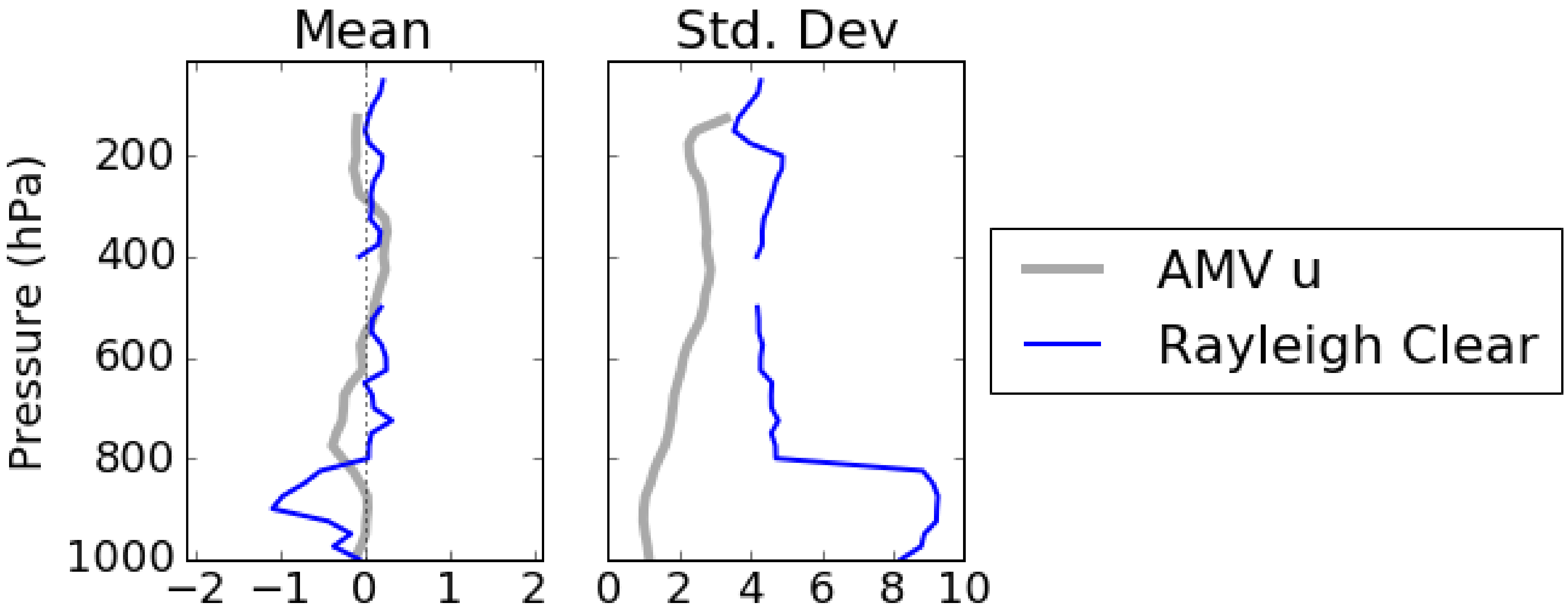
For comparison to AMVs, we are considering the bulk statistics for:

- 1-31 October 2018 (all cycles)
- 45°S to 45°N
- Aeolus HLOS; AMV u wind component
- AMVs from all GEO types: (all satellites, cloud & WV-derived winds)
- Aeolus winds are bias corrected based on previous plots

The latitudinal boundaries are chosen in an effort to keep LOS & AMV comparisons somewhat consistent

Note, no colocation is performed

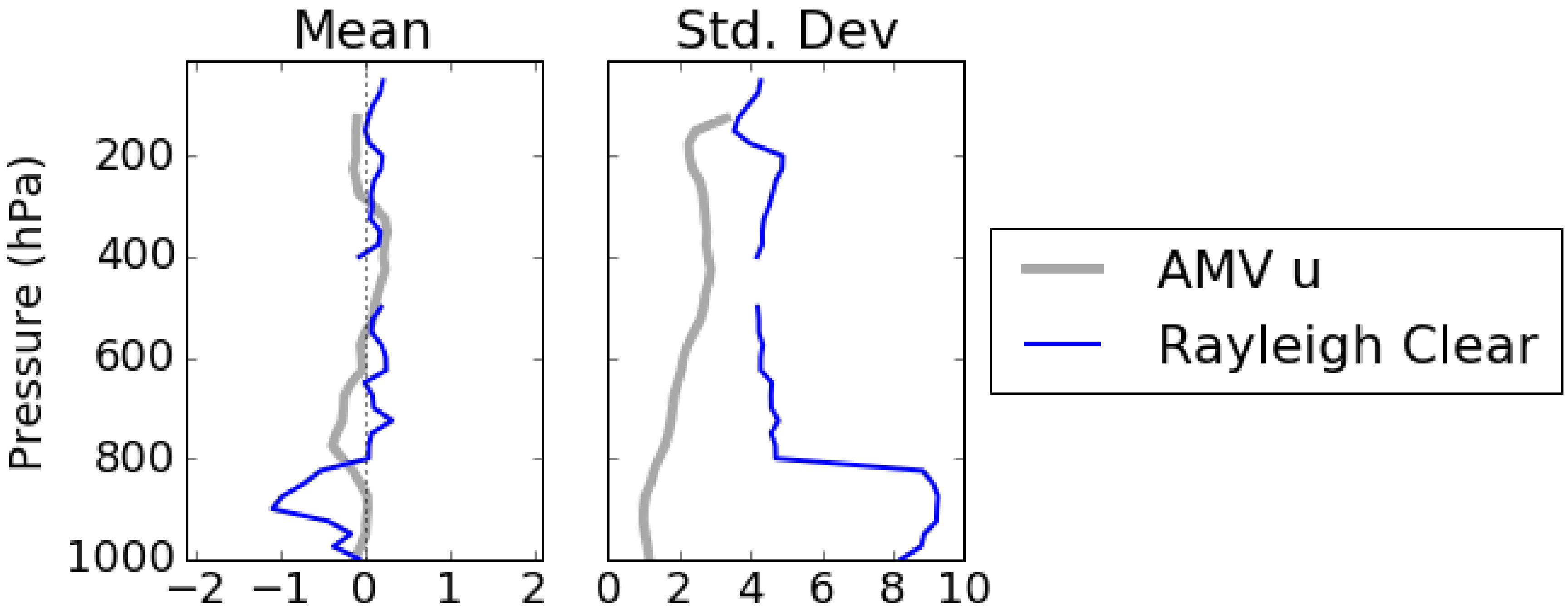
Comparison against AMVs



First, the AMVs show biases of their own

- These may reflect both model biases and AMV retrieval biases
- If there are model forecast biases, then it is possible that the bias is projected onto the Aeolus HLOS winds via BC procedure, since there is no anchoring performed
- Not shown: AMV v means are slightly different, but standard deviations are the same

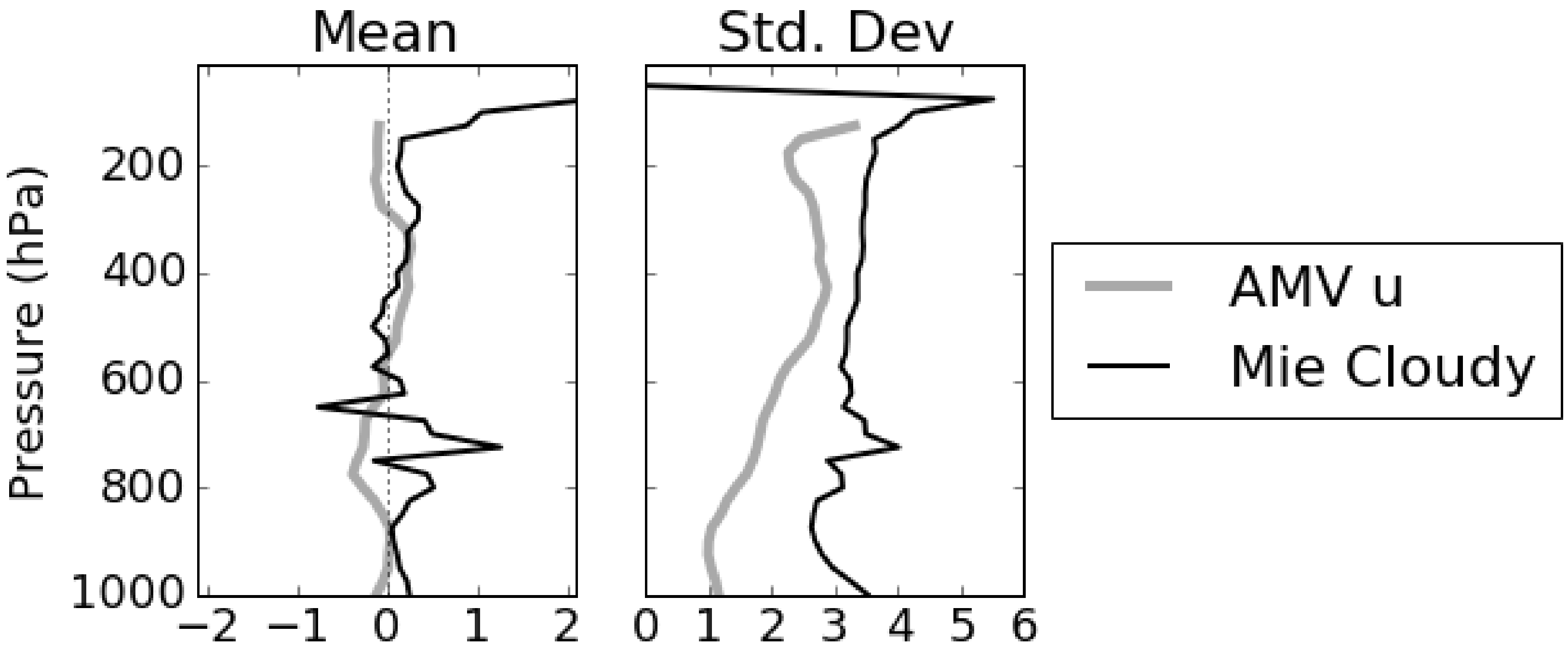
Comparison against AMVs



Compared to the AMVs, the Rayleigh Clear HLOS winds have:

- A fundamentally different bias structure
- Larger variance at all levels
- A very large variance in the lowest 200 hPa

Comparison against AMVs



Compared to AMVs, the Mie Cloudy HLOS winds show:

- More varying bias structures between 800 – 600 hPa
- Much higher quality measurements near-surface in terms of bias and variance
- Larger variance at all levels

Prescribing Observation Errors for Assimilation

The L2B product comes with estimates of observation error – hereafter EE

- The EE is based on the retrieval – (e.g. signal-to-noise)

The applicability of this retrieval-based EE in data assimilation is under evaluation

- In data assimilation, the observation error (hereafter σ_o) fundamentally provides the observation weight
- The σ_o is not, however, simply an exact definition of the error in instrument space
  - For example, in radiance assimilation, it is not an NEdT. In temperature assimilation, it is not simply the noise level of the thermistor
- Representative error is a key component of σ_o in data assimilation
  - Perhaps more important in HLOS wind assimilation – particularly for Rayleigh winds:
    - Scale issues (~100 km integration length) between model (handled as a single point) and observation
    - Sub-integration-scale variability translated to uncertainty
- Therefore, σ_o is larger than the true observation error, so even if EE is perfect, it must be inflated

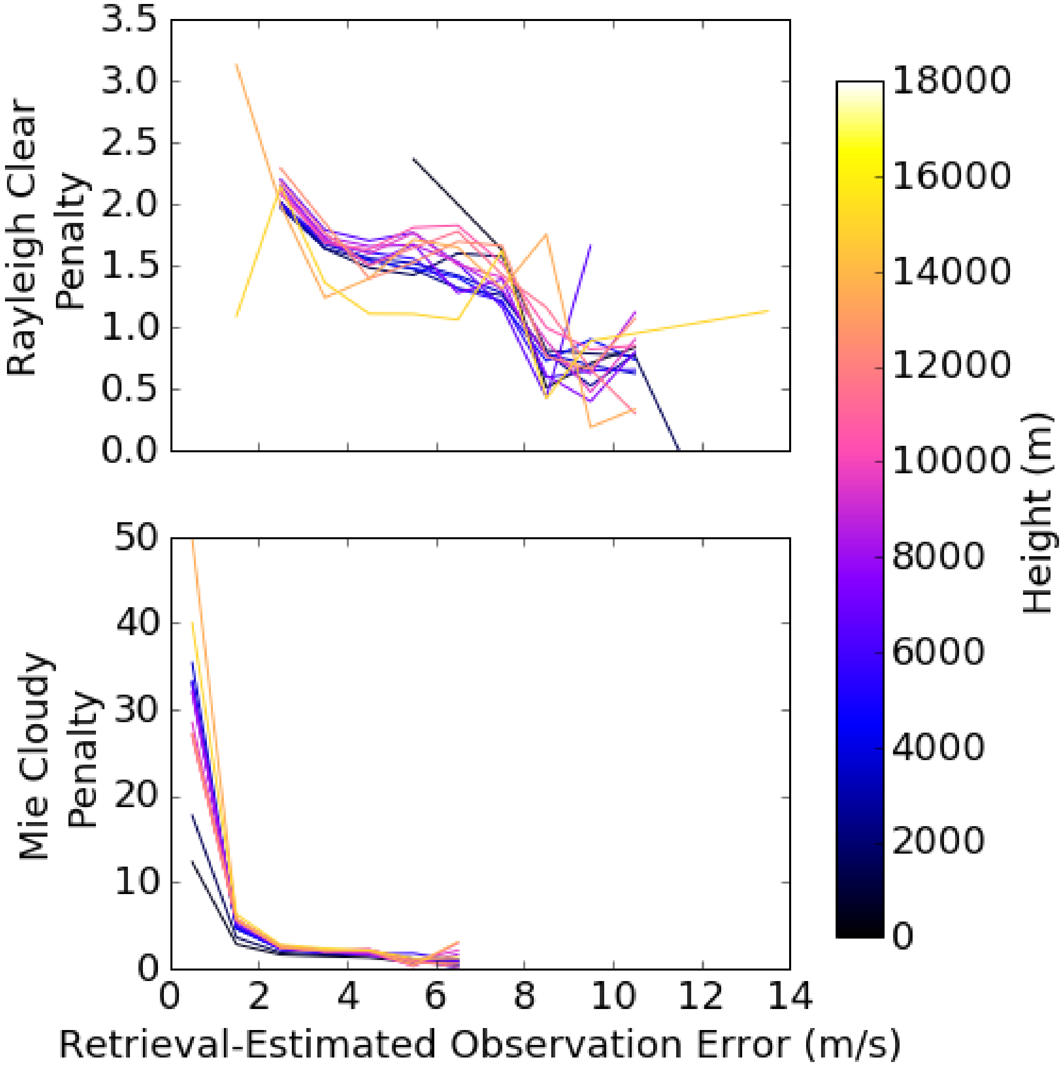
To understand the character of the EE relative to the background departures, the *penalty* is considered

- $Penalty = \frac{1}{n}\sum\frac{(O-B)^2}{\sigma_O^2}$
- For these plots, we calculate the penalty assuming σ_o = EE

Penalty vs. Estimated Error

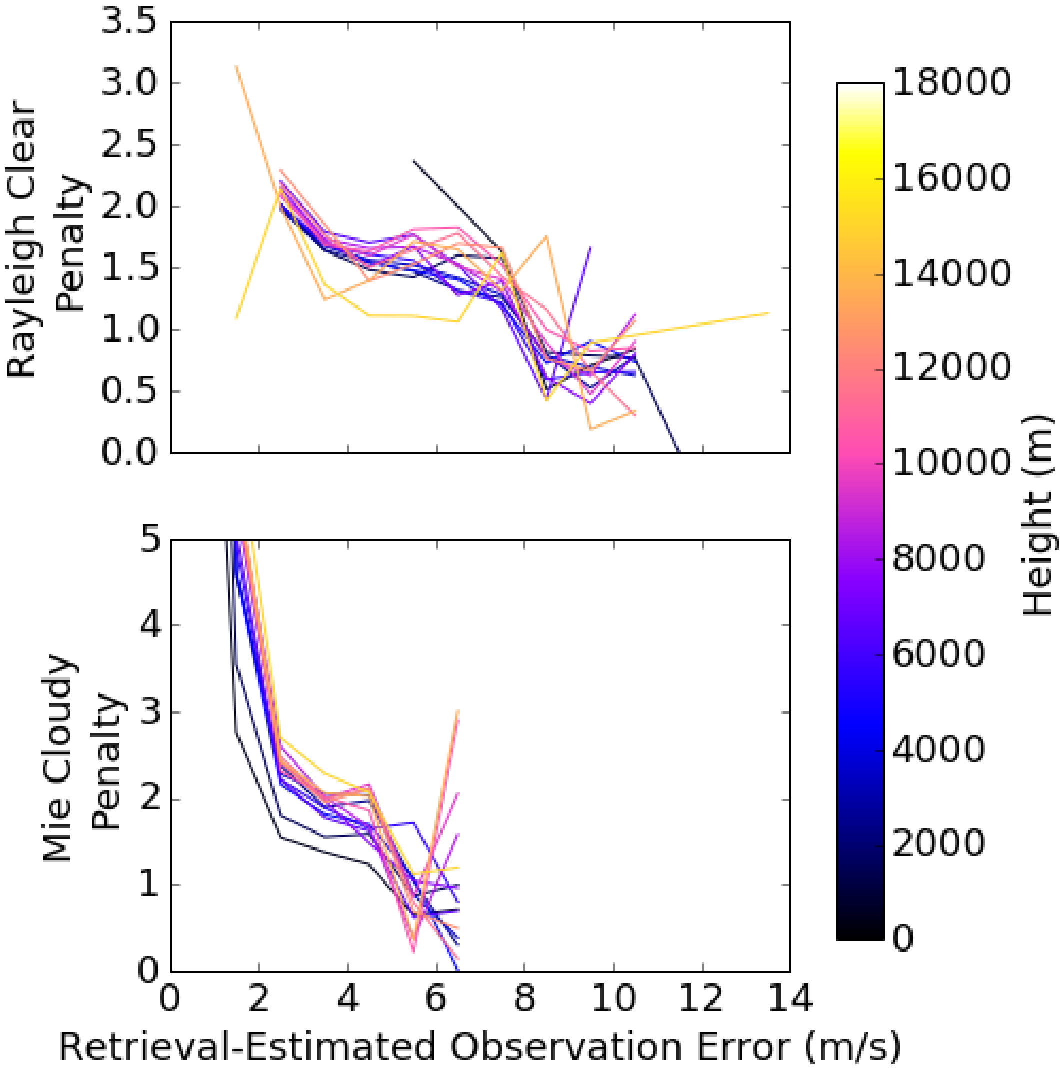
Penalty shows that the estimated errors for both Rayleigh Clear and Mie Cloudy HLOS winds are smaller than the variance of the background departure for most values of EE

- Roughly a factor of 1.5-2x for Rayleigh Clear, as a function of height
- Ratio < 1 for large error values
- Mie penalties large in cases where EE is small (small denominator)



Penalty vs. Estimated Error

Zooming in, Mie Cloudy departure variance 1-2x as large for wind EE greater than 2 m/s



Prescribing Observation Errors for Assimilation

To account for EE in the definition of σ_o a simple error model has been implemented as a function of height

- $\sigma_o(z) = m(z) * EE + b(z)$

A target *penalty* is used to initially tune the vertical profiles of m & b

- In a perfect world, Penalty == 1:
  - Background error is non-zero, thus making penalty larger
  - Representativeness error makes σ_o larger than ‘pure’ observation error
  - Representativeness can also have horizontally and vertically correlated error structures, which are not well-handled in DA and are thus handled by inflation of error variance
- A value of 0.45, which is similar to values used for AMVs, is initially used for both Mie Cloudy and Rayleigh Clear

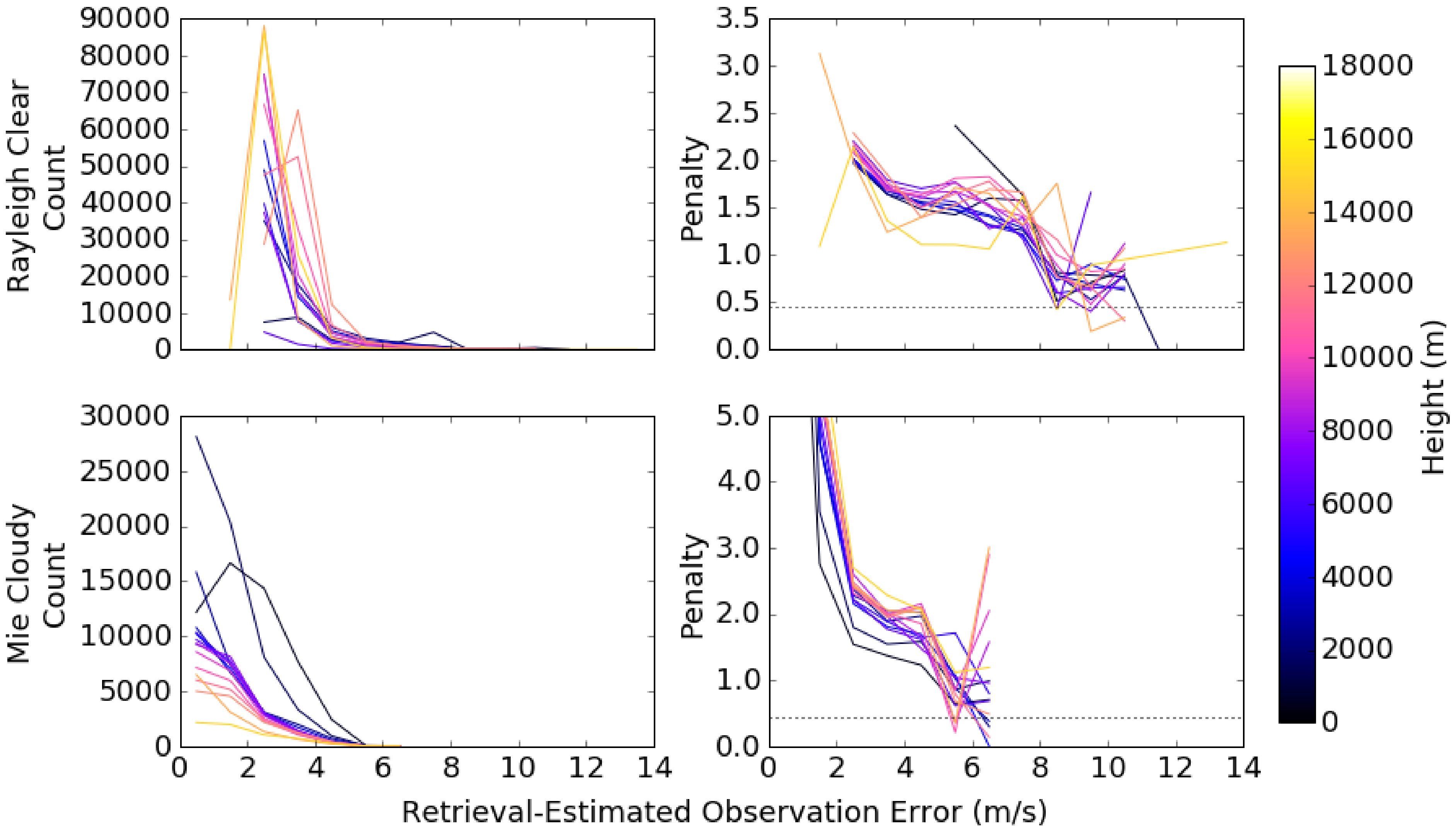
$$Penalty = \frac{1}{n}\sum\frac{(O-B)^2}{\sigma_o^2}$$

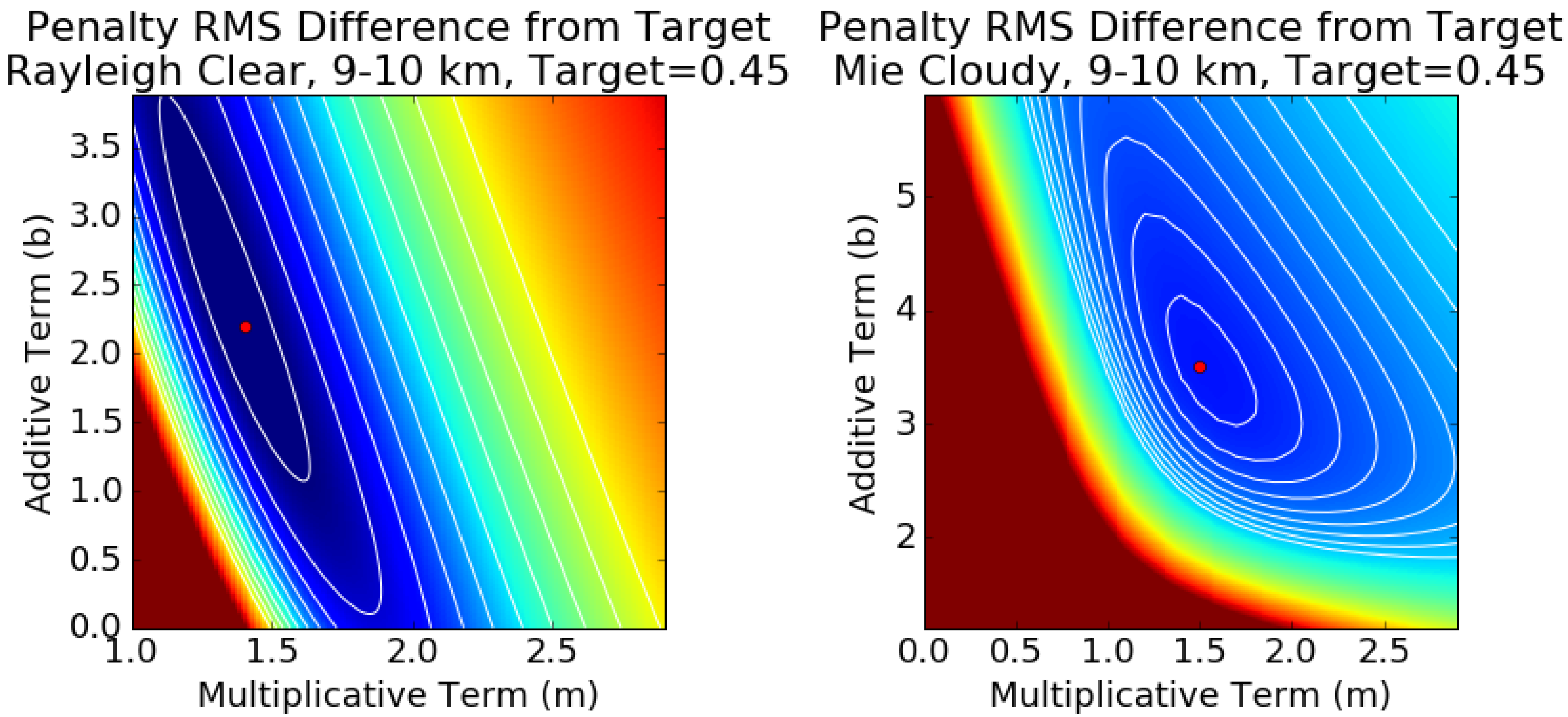
This model was designed to provide a (overly?) simple flexibility to tune the error, either as a multiplicative inflation of EE or as an additive constant.

Penalty vs. Estimated Error

An important point to this tuning exercise is that all obs are tuned to a single penalty

  - The tuned m(z) & b(z) values will be largely weighted by EE values with the most observations





The target penalty was used to determine values of m(z) & b(z) as a function of vertical height bins

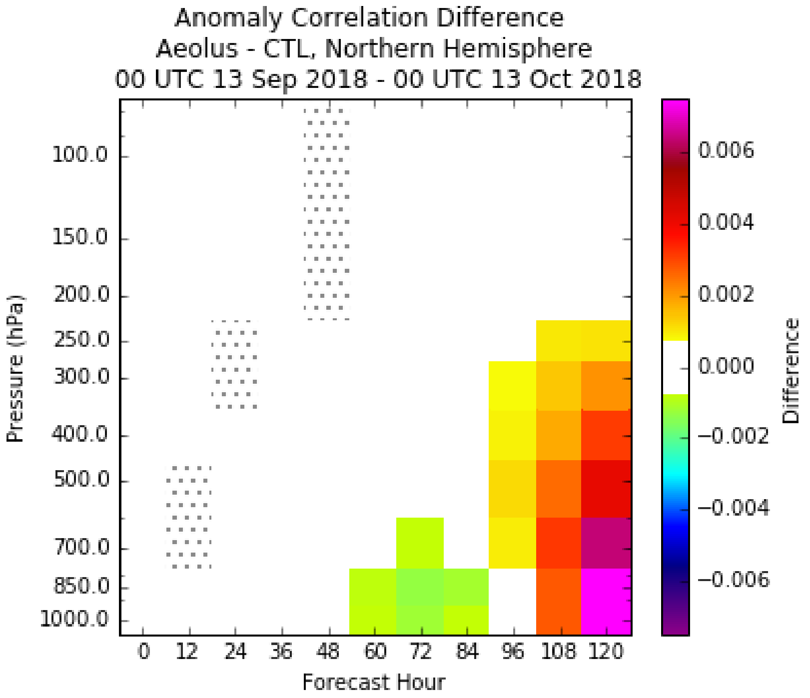
- The m & b that minimize the difference between the new & target penalty (shaded/contoured) was selected at each level
- Plots shown are for 9-10 km

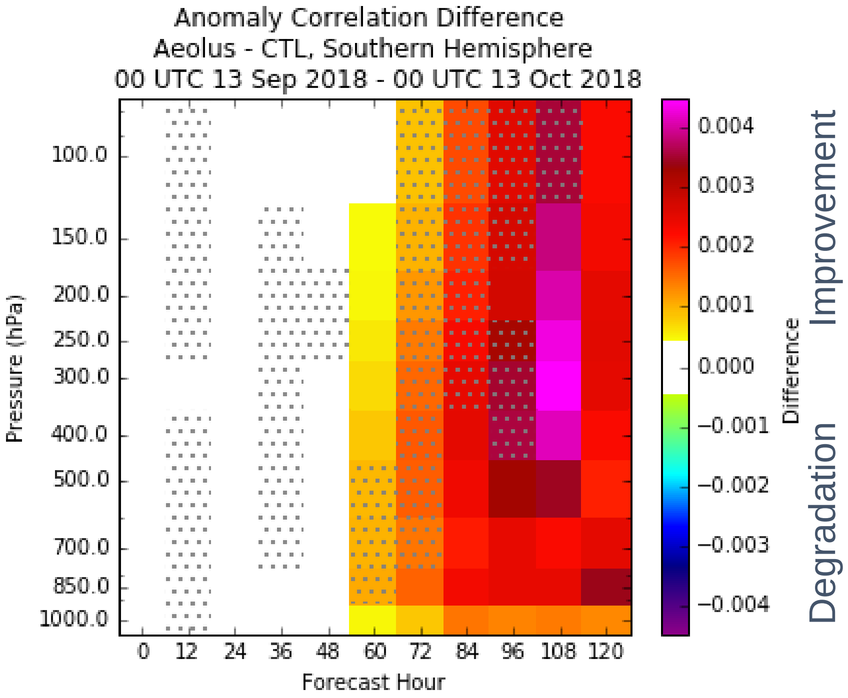
At This Point…

The Aeolus data was ready for a standard Observing System Experiment (OSE)

- CTL was run
- Experiment: CTL + Aeolus
    - Using methods shown thusfar

Forecast Verification – Height Anomaly Plots


Anomaly Correlation Difference
Aeolus - CTL, Northern Hemisphere
00 UTC 13 Sep 2018 - 00 UTC 13 Oct 2018


Anomaly Correlation Difference
Aeolus - CTL, Southern Hemisphere
00 UTC 13 Sep 2018 - 00 UTC 13 Oct 2018

Height anomaly forecast verification show neutral impacts in the Northern Hemisphere.

- Insignificant improvement at day 5, degradation near surface at day 3

General improvement to day 4 in the Southern Hemisphere, including statistical significance in the upper troposphere

- Hopefully the second laser allows for longer sample periods to gain more confidence

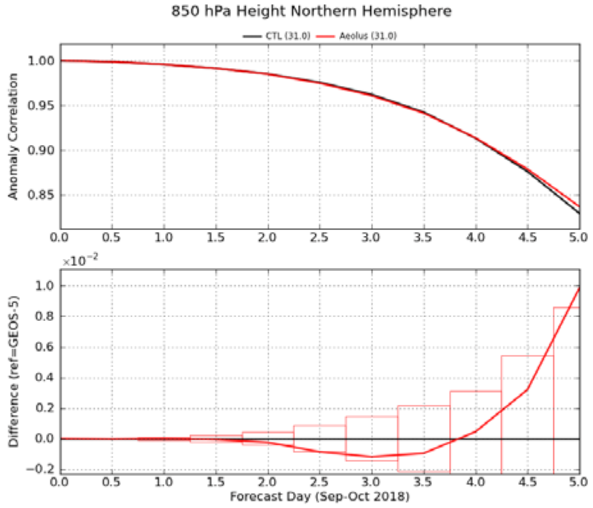
Height Anomaly Plots – by level

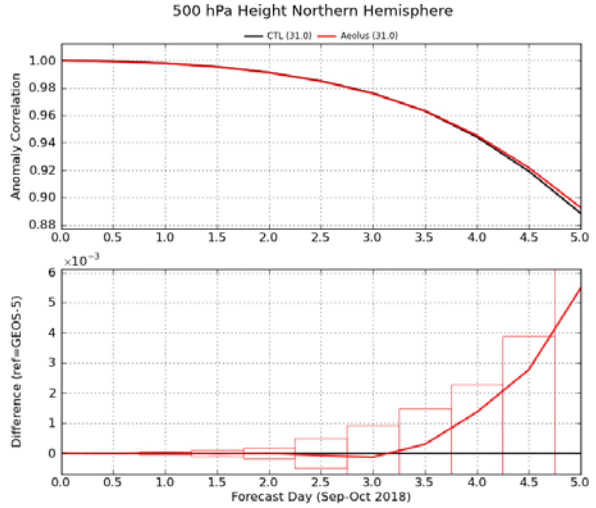
850 hPa

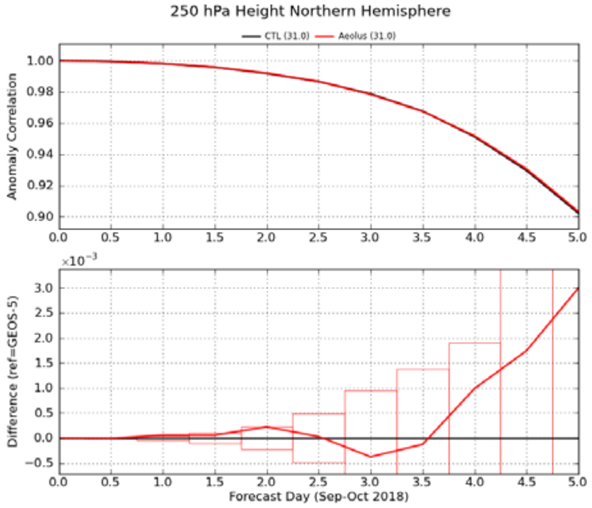
500 hPa

250 hPa

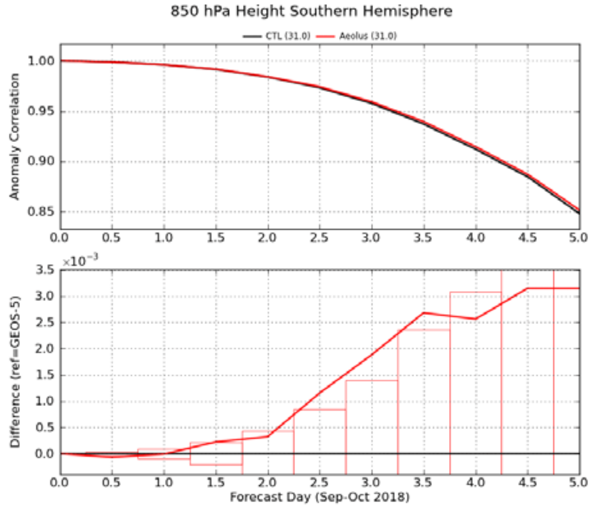
Northern Hemisphere

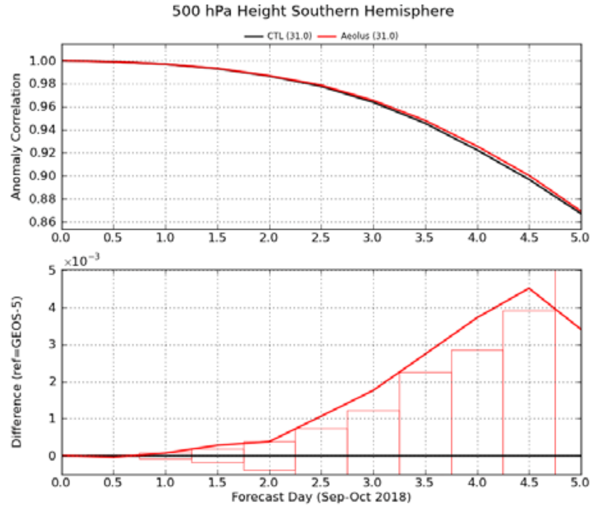

850 hPa Height Northern Hemisphere

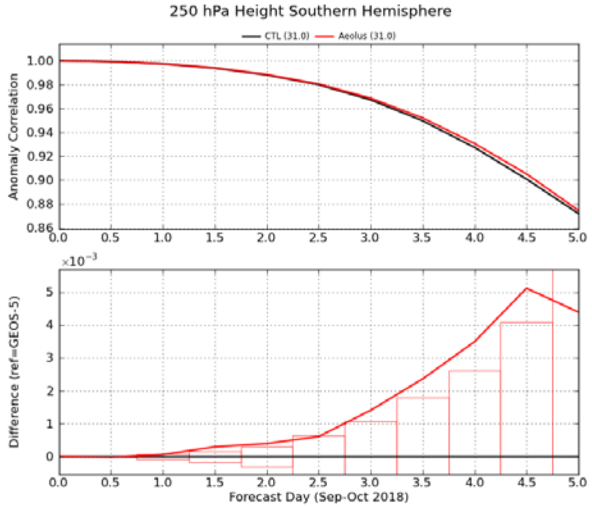

500 hPa Height Northern Hemisphere


250 hPa Height Northern Hemisphere

Southern Hemisphere


850 hPa Height Southern Hemisphere

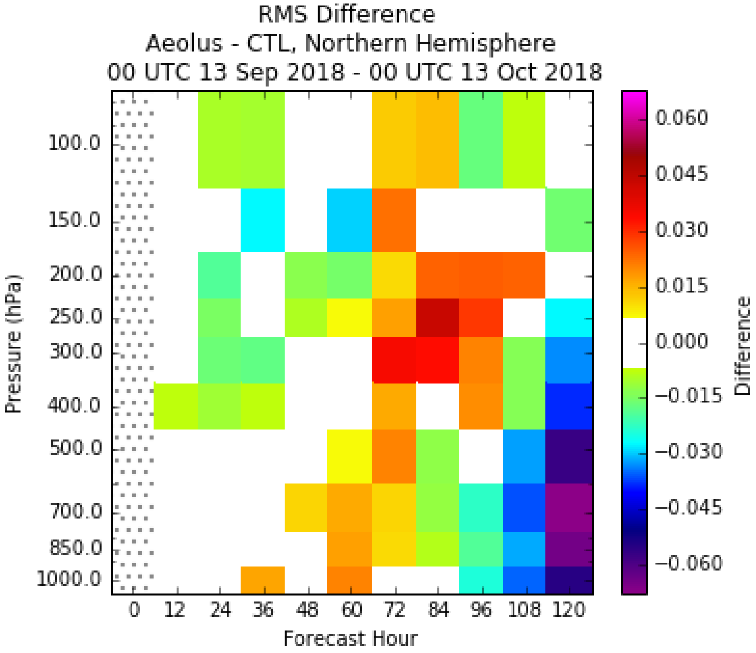

500 hPa Height Southern Hemisphere

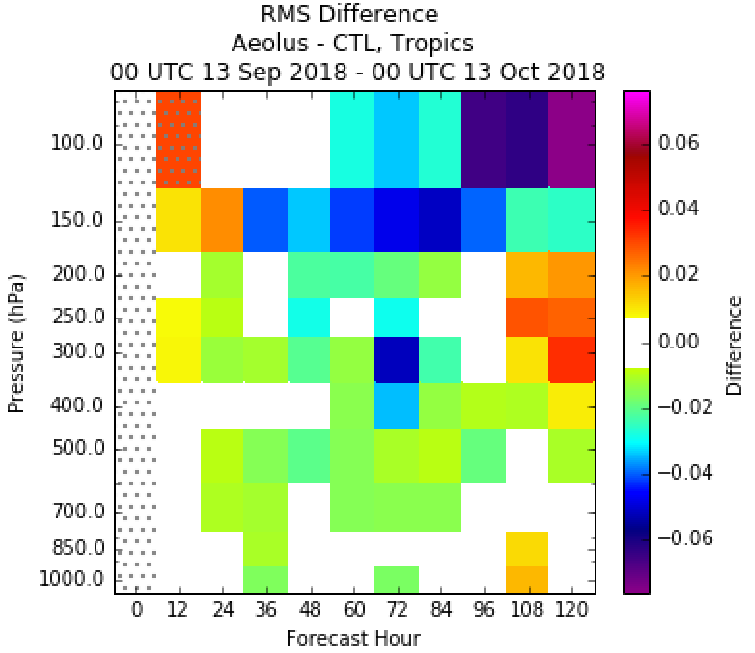

250 hPa Height Southern Hemisphere

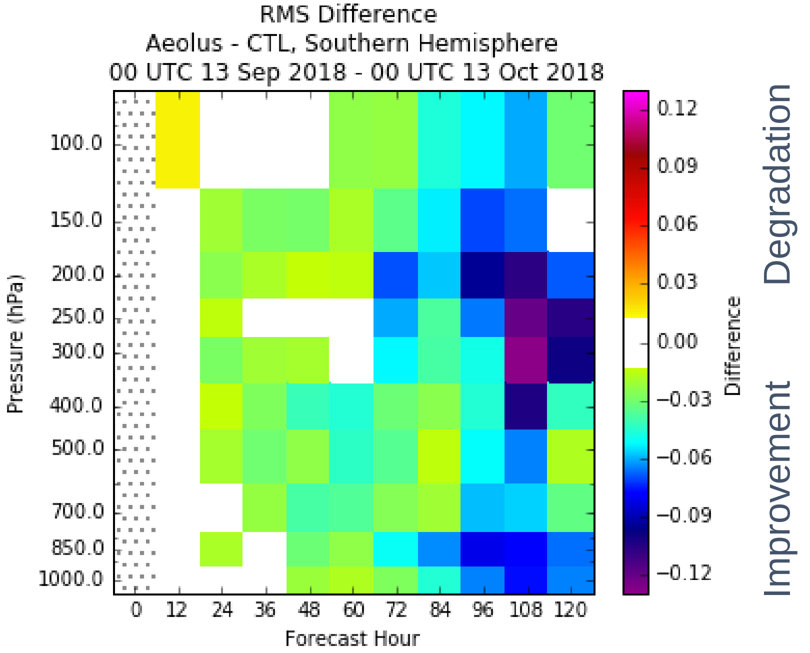
Mostly consistent with previous plot, these are the absolute magnitudes, differences, and error bars* of the AC at various levels

* These plots are generated using a GMAO tool that assumes a hardwired 0.9 confidence interval; the contoured plot on the previous slide assumes a 0.95 confidence interval for the hatching

Forecast Verification – RMS Plots







Zonal Wind RMS forecast verification show neutral impacts globally

- Patterns inconsistent throughout the forecast in the Northern Hemisphere
- Generally of the correct sign in the tropics until Day 5, particularly near UT/LS
- Consistently of the correct sign through column to Day 5 in Southern hemisphere

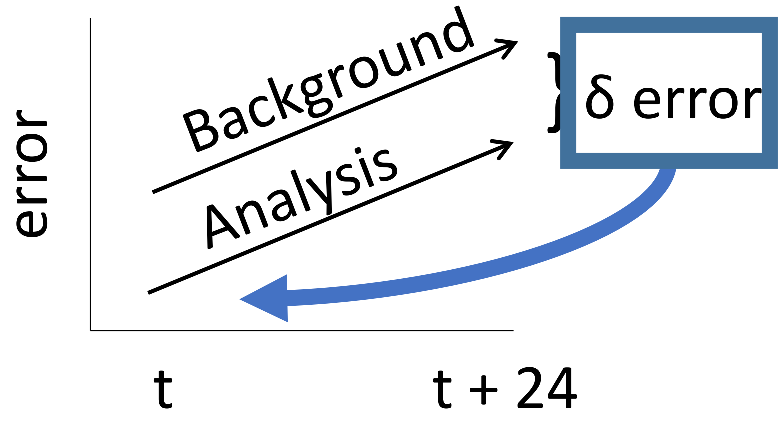
One month is a very small sample size for this metric

Forecast Sensitivity - Observation Impact (FSOI)

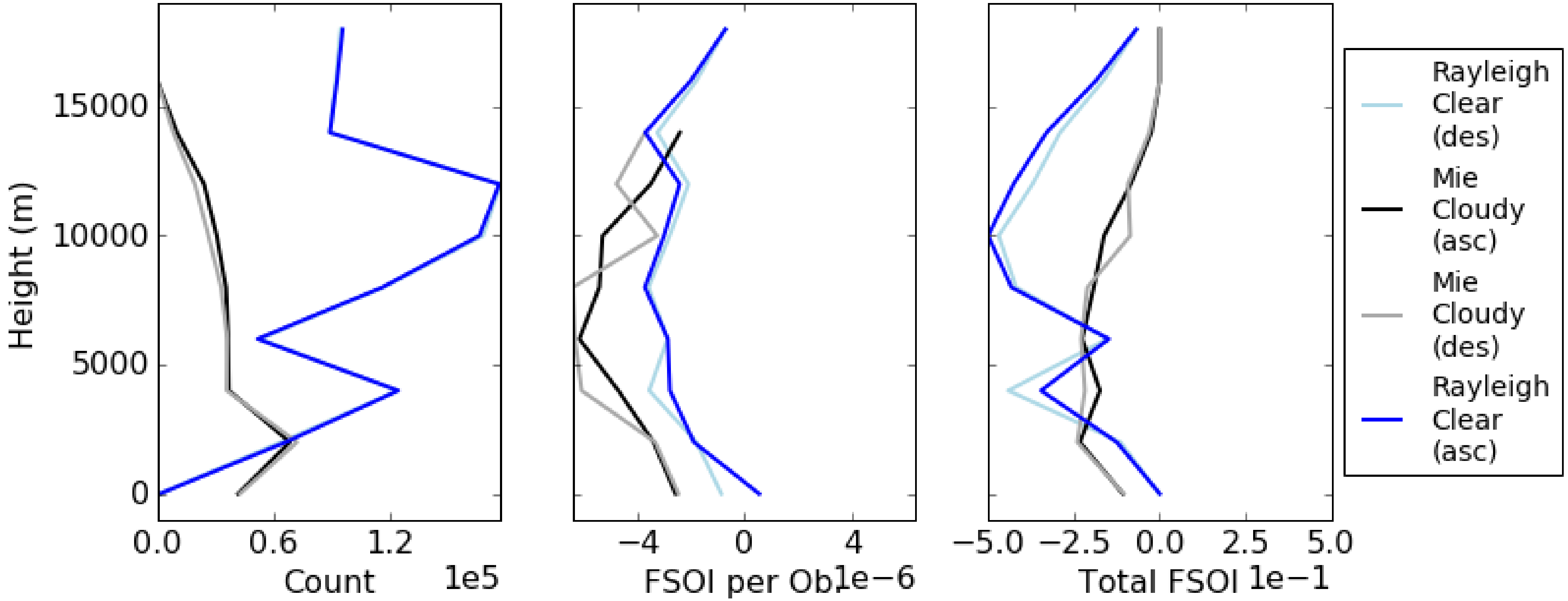
FSOI is a measure of 24 hour forecast error reduction projected into observation space

- Each assimilated observation has its own impact metric
  - Allows for the aggregation of the metric in different ways
    - per instrument, channel, footprint, etc.
- **A negative value equates a reduction in error, so NEGATIVE = GOOD**

FSOI was run for the entire time period (1 month, 4 cycles per day)



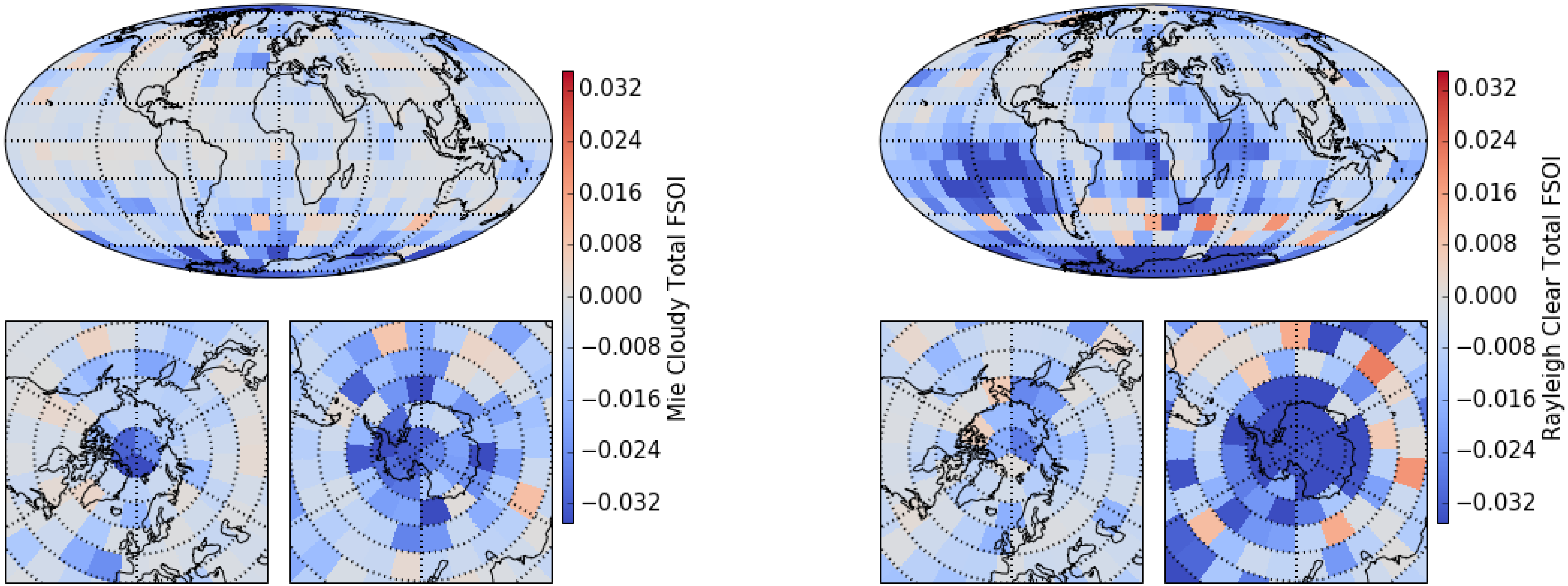
Forecast Sensitivity - Observation Impact



FSOI as a function of height

- No notable difference seen between ascending and descending nodes in this metric
- Total impact is the sum over all observations, so there is some expectation that the impact scales with observation count
  - Rayleigh FSOI minimum corresponds to the observation count maximum
  - With mie, the impact scales less obviously
- Per observation, both the Mie and Rayleigh show signs of most impact in the middle troposphere
  - Consistent in that AMVs are generally located in the upper and lower troposphere – the middle troposphere is a relative data void!

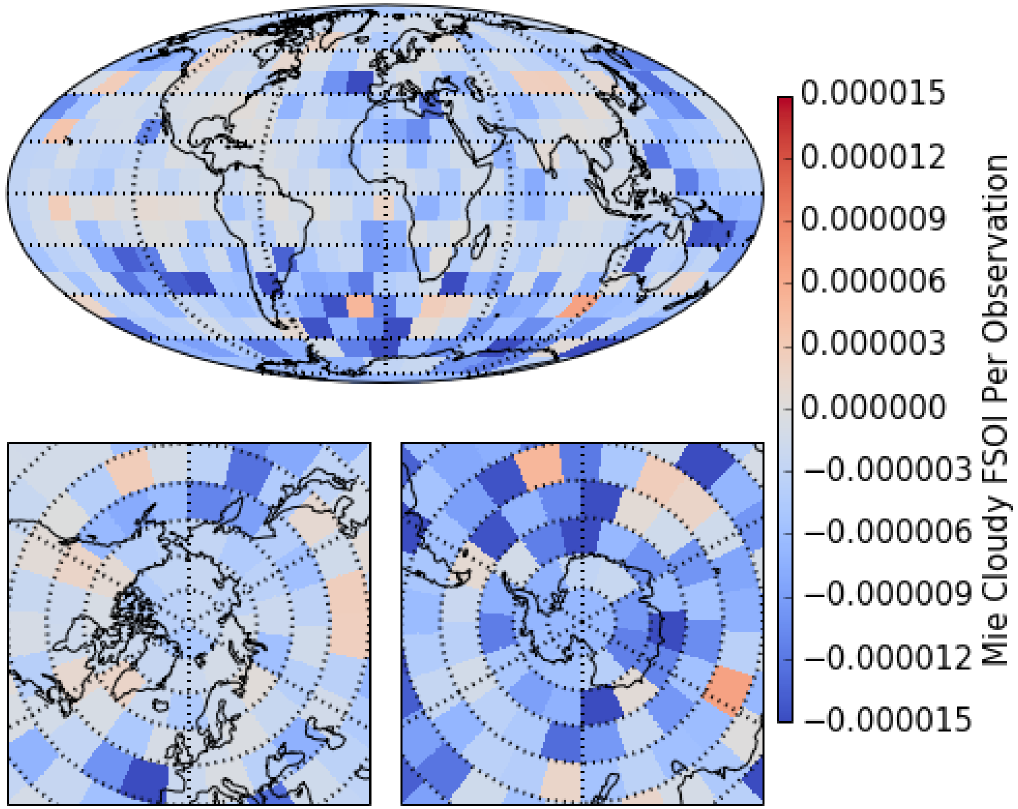
FSOI

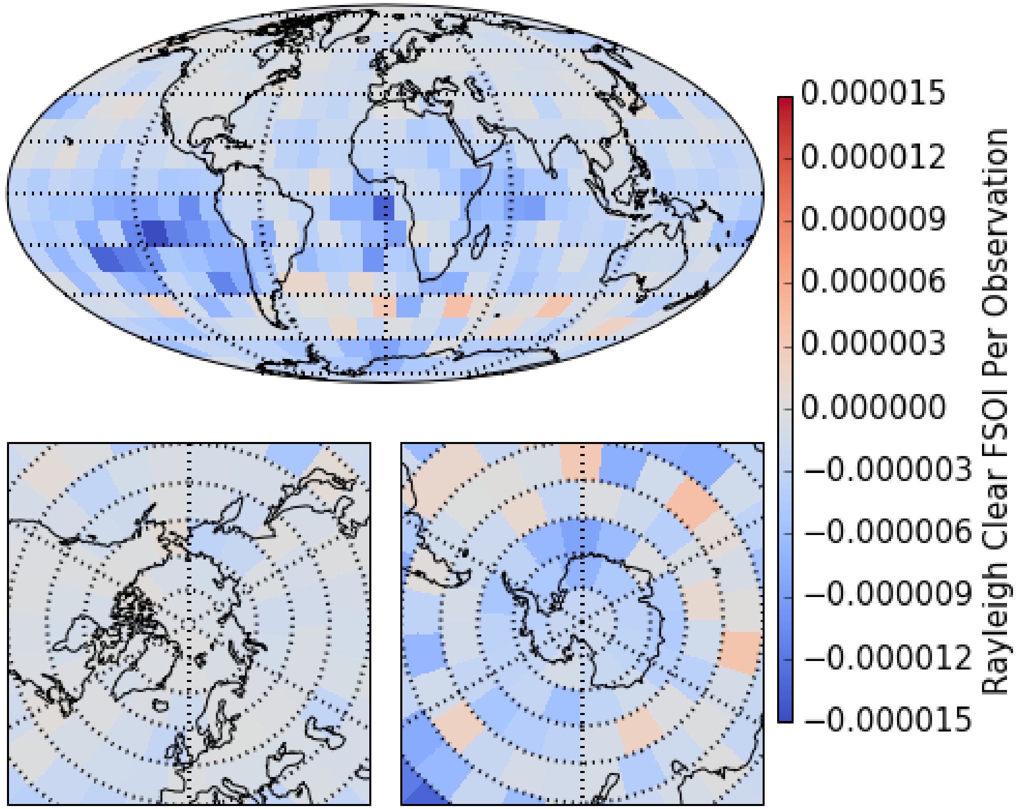


Spatial Distribution of FSOI (Total)

- Both show largest impacts in Southern Hemisphere – consistent with forecast skills
- Both show benefit over Antarctica, minimal impact in southern storm track
- Rayleigh shows benefit in subtropics that is not as apparent in Mie

FSOI





Spatial Distribution of FSOI (Per Observation)

- Generally similar patters to total
- Rayleigh maintains large subtropical signal
- Rayleigh Antarctic signal weaker per-observation than the Mie; Mie stronger signal in N. Hemisphere

Conclusions

The results seem promising

- Perhaps even better than expected considering the initial assessment of background departures
- It would be ideal to get a reprocessed dataset with as many issues resolves as can be upstream

Bias correction is a real issue

- I don't want the data delivered with an applied bias correction unless it can be removed
- Would the BUFR table allow this?
- Would there need to be additional sequences for different corrections
- Variational would be ideal, but:
    - Are there enough anchoring observations?
    - Implementing a VarBC isn't always trivial (in some systems, it requires wiring throughout the entire system)
- Offline but dynamic correction is probably the most reasonable
    - Perhaps a unified tool for accomplishing this consistently

Scientifically, I envisioned this effort as a potential anchor to the AMVs – which biases are still unaccounted for?