Assimilation of High-Resolution Hurricane Inner-Core Data with the HWRF Hurricane Ensemble Data Assimilation System (HEDAS): Evaluation of the 2008-2011 vortex-scale analyses

Altuğ Aksoy[1,2]

[1]NOAA/AOML Hurricane Research Division
[2]U. Miami RSMAS Coop. Inst. Marine Atmos. Studies

Collaborators:
Sim Aberson, Tomislava Vukicevic,
Kathryn Sellwood, Sylvie Lorsolo, Thiago Quirino,
Bachir Annane, Shirley Murillo, Neal Dorst, Mike Page

The National Hurricane Center Seminar – 15 February 2012

Outline

- Brief introduction to data assimilation
- Advantages of ensemble-based data assimilation
- Description of HEDAS
- Overview of real-data cases
- Performance of analyses
- Conclusions

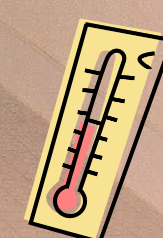
How Does *Scalar* Data Assimilation Work? Simplest Case

Let's consider the simplest possible scenario:

- **We are concerned about the temperature of a room – this is the quantity we would like to estimate (analysis),** T_a
- **We have two sources of information:**
  - » **A computer model that predicts the temperature (forecast, background)**
  - » **A thermometer placed somewhere in the room (observation)**

$T_f = 20^\circ\ \text{C}$ $T_a = 22^\circ\ \text{C}$ $T_o = 24^\circ\ \text{C}$

In the absence of any other information, the simplest approach would be to just take the mean of the two estimates!



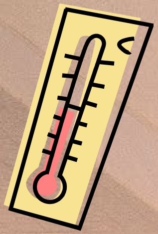
How Does *Scalar* Data Assimilation Work?
Better Approach: Estimates of Uncertainty

If we are a little more systematic about our approach, we would also try to obtain estimates of uncertainties associated with these measurements:

$T_f = 20^\circ\ \mathrm{C}$ $\quad 22^\circ\ \mathrm{C} < T_a < 24^\circ\ \mathrm{C}$ $\quad T_o = 24^\circ\ \mathrm{C}$

$\sigma_f = 2^\circ\ \mathrm{C}$ $\quad \sigma_o = 1^\circ\ \mathrm{C}$

Since our confidence in our observation is greater than in our model prediction, our final estimate should be closer to the observation, T_o*!*



How Does *Scalar* Data Assimilation Work? Statistical Estimation: Minimum Variance

With certain statistical assumptions, it is quite straightforward to obtain a mathematical relationship for the analysis temperature, T_a, that represents the *minimum-variance estimate*:

$$T_a = \frac{\sigma_o^2}{\sigma_f^2+\sigma_o^2} T_f + \frac{\sigma_f^2}{\sigma_f^2+\sigma_o^2} T_o$$

$T_f = 20^\circ \ \mathrm{C}$

$T_a = 23.2^\circ \ \mathrm{C}$

$T_o = 24^\circ \ \mathrm{C}$

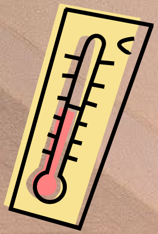
$\sigma_f = 2^\circ \ \mathrm{C}$

$\sigma_o = 1^\circ \ \mathrm{C}$

This approach also allows us to obtain an estimate for the analysis uncertainty, σ_a:

$$\sigma_a^2 = \left(\sigma_f^{-2}+\sigma_o^{-2}\right)^{-1}$$

$$\rightarrow \sigma_a \approx 0.45^\circ \ \mathrm{C}$$



How Does *Scalar* Data Assimilation Work?
Statistical Estimation: Incremental Form

With simple algebra, we can write the update (analysis) equation in *incremental* form:

$$T_a = T_f + \underbrace{\frac{\sigma_f^2}{\sigma_f^2 + \sigma_o^2}}_{K} \underbrace{(T_o - T_f)}$$

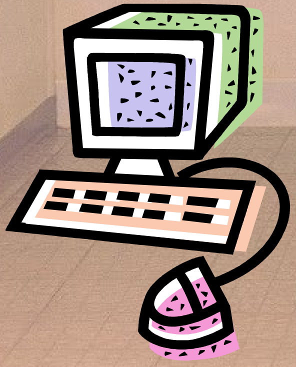
$T_f = 20^\circ$ C

$\sigma_f = 2^\circ$ C

Observation innovation or ***observation increment*** : Additional information introduced by the observation

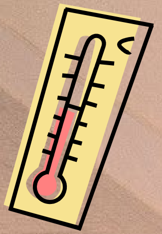
$T_o = 24^\circ$ C

$\sigma_o = 1^\circ$ C



Similarly, analysis uncertainty *σ_a can be expressed as:*

$$\sigma_a^2 = (1-K)\,\sigma_f^2$$



How Does *Scalar* Data Assimilation Work? Most Complex Case: Indirect Observations

A somewhat more complicated problem arises when we can only observe temperature indirectly, say through the measurement of infrared radiation:

$$\underbrace{P_a - P_f} = K'(P_o - \mathrm{H}(T_f))$$

$$T_f = 20^\circ \ \mathrm{C}$$

$$\sigma_f = 2^\circ \ \mathrm{C}$$

Now it is easiest to perform the update in *observation space* first!

Analysis increment is in observation space, but the observation and model increments are related through their correlation:

$$\Delta T \sim \rho_{PT} \Delta P$$

$$\textit{Obs. power} = P_o$$

$$P = \mathrm{H}(T) \approx \nu T^4$$

$$\sigma_P \approx \mathrm{H}(\sigma_T)$$

How Does *Scalar* Data Assimilation Work? Most Complex Case: Indirect Observations

A somewhat more complicated problem arises when we can only observe temperature indirectly, say through the measurement of infrared radiation:

$$P_a - P_f = K'(P_o - \mathrm{H}(T_f))$$

$T_f = 20^\circ$ C

$\sigma_f = 2^\circ$ C

Notice that this is the slope of the linear regression line between ΔT and ΔP !!

$Obs.\ power = P_o$

$P = \mathrm{H}(T) \approx \nu T^4$

$\sigma_P \approx \mathrm{H}(\sigma_T)$

$$\Delta T_a = \rho_{PT} \frac{\sigma_f}{H(\sigma_f)} \Delta P_a$$

How Does *Scalar* Data Assimilation Work?
Final Product: Full Update Equation

We have finally reached a full statistical solution for the scalar problem:

$$T_a - T_f = \rho_{PT}\,\frac{\sigma_f}{H(\sigma_f)}\,K'\left(P_o - \mathrm{H}(T_f)\right)$$

$T_f = 20^\circ$ C

$\sigma_f = 2^\circ$ C

To obtain the final form, let's expand K':

$$K' = \frac{H^2(\sigma_f)}{H^2(\sigma_f) + \sigma_o^{\,2}}$$

Obs. power $= P_o$

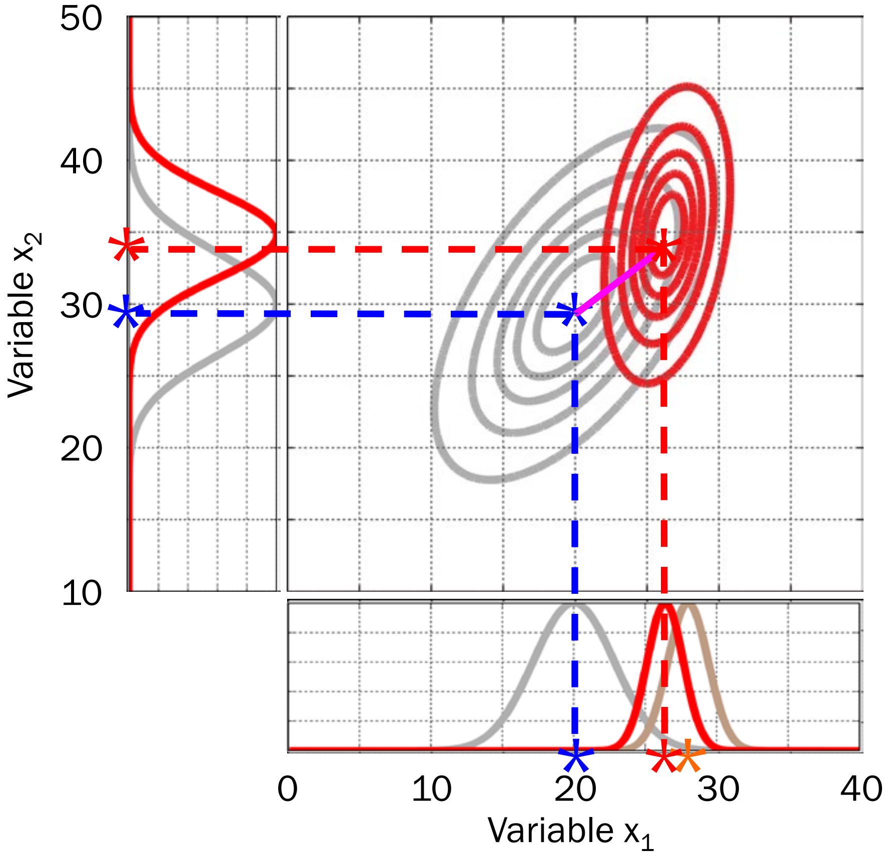
$P = \mathrm{H}(T) \approx \nu T^4$

$\sigma_P \approx \mathrm{H}(\sigma_T)$

$$K = \frac{\mathrm{cov}[\sigma_f, H(\sigma_f)]}{H^2(\sigma_f) + \sigma_o^{\,2}}$$

The DA Update Step with 2 Variables
The Full Bayesian Picture

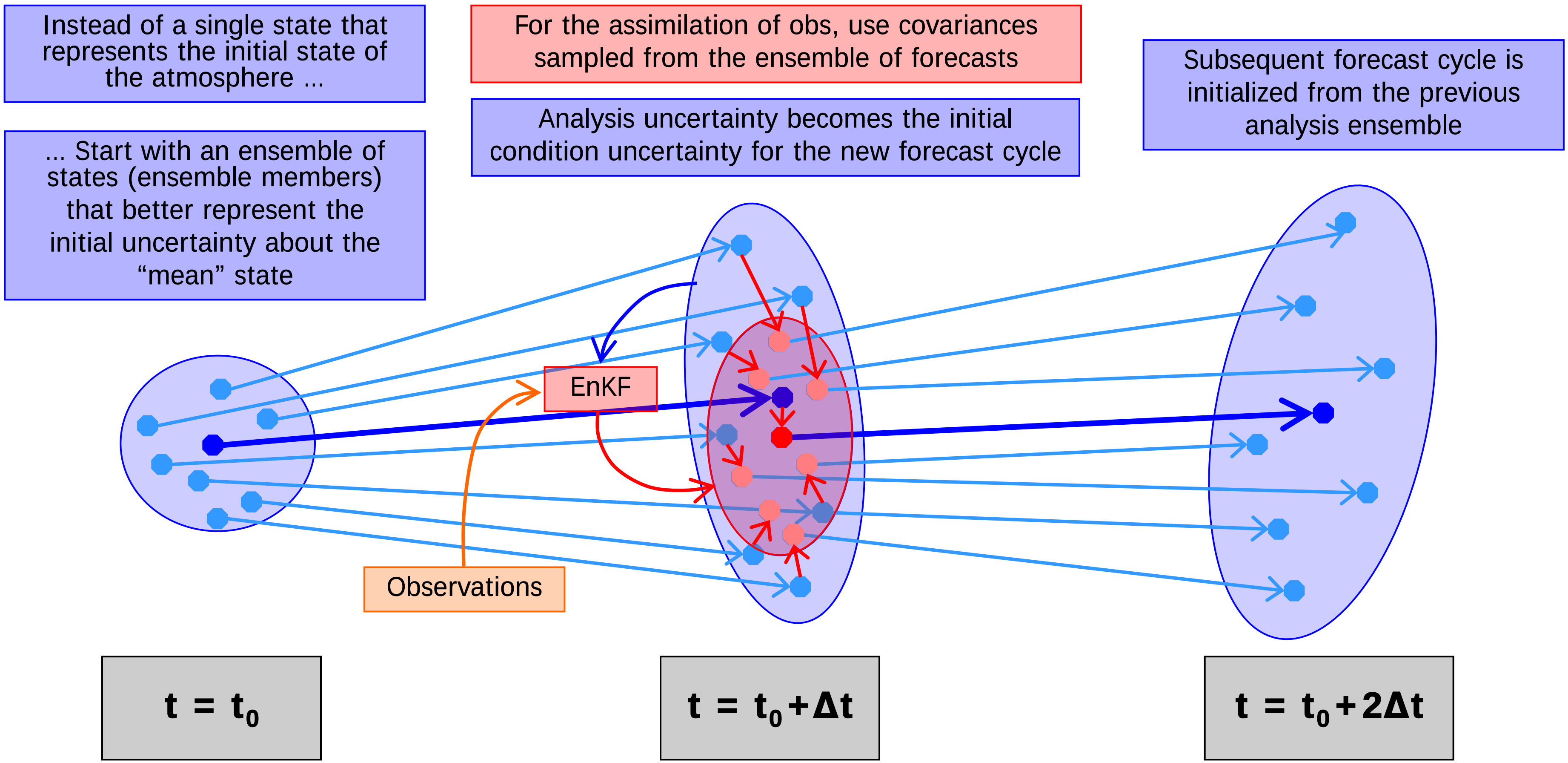
- When there are unobserved variables, information is propagated through the covariances among variables
- Let's see how an update would be carried out for two variables, one observed:



- The background consists of x_1 and x_2
- Both variables are normally distributed
- The two variables are correlated
- x_1 is directly observed as y^0
- The observation is also normally distributed; but has smaller error than x_1
- Since observation error is smaller than x_1 background error, the analysis x_1^a is closer to y^0
- The analysis error s_1^a is smaller than both s_1^f and s_y
- The covariance between x_1 and x_2 relates changes in x_1 to changes in x_2 -> slope of regression
- The joint analysis probability distribution has become narrower, reflecting improved estimates in both x_1 and x_2

Ensemble-based Data Assimilation: The General Process

In ensemble-based data assimilation, we compute sample covariances from an ensemble of forecasts



Advantages of Ensemble-Based Data Assimilation

- Background covariances are sampled from the forecast ensemble
  → Flow-dependent covariances that are independent of any assumptions for the nature of flow (e.g., geostrophy)
- Provides a natural basis for probabilistic forecasts
- Easy to implement and maintain
    - No adjoints are needed to be developed which is especially complicated for highly nonlinear and discontinuous sub-grid parameterization schemes and/or nonlinear observation operators
- Straightforward application to domains with multiple nests
- Easily lends itself to parallelization
- Performance (so far) comparable to variational schemes

NOAA/AOML/HRD's HWRF Ensemble Data Assimilation System (HEDAS)

- **Forecast model:**
  - Exp. HWRF with 2 nested domains (9/3 km hor. resolution, 42 vert. levels)
  - Static inner nest to accommodate covariance computations
  - Ferrier microphysics, explicit convection on inner nest
- **Ensemble system:**
  - Initialized (cold start) from *GFS-EnKF (NOAA/ESRL)* ensemble member analyses
  - 30 ensemble members
- **Data assimilation:**
  - Square-root EnKF filter (Whitaker and Hamill 2002)
  - Assimilates data only on the inner nest
  - Covariance localization (Gaspari and Cohn 1999)
  - No explicit covariance treatment
  - Filter solver parallelized using OpenMP

Aircraft Data of Interest

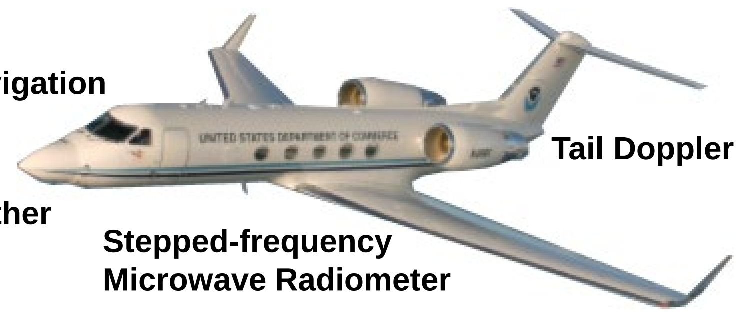
P-3 Aircraft (eye penetrations)

Chemistry inlet
GPS
Humidity
Radar altimeter
Gust probe
C-Band Weather Radar
Tail Doppler Radar
Radiometer
Pylons
Lower-fuselage Weather Radar
Launch tubes and chutes: Dropwindsonde, AXBT, AXCP, AXCTD
(Stepped-frequency Microwave Radiometer, cloud physics, aerosol)

Observation	Error
Doppler wind speed	2 ms^{-1}
FL/Dropsonde Temperature	0.5 K
FL/Dropsonde zonal/merid. wind speed	2 ms^{-1}
SFMR	Variable, mean ~5 ms^{-1}

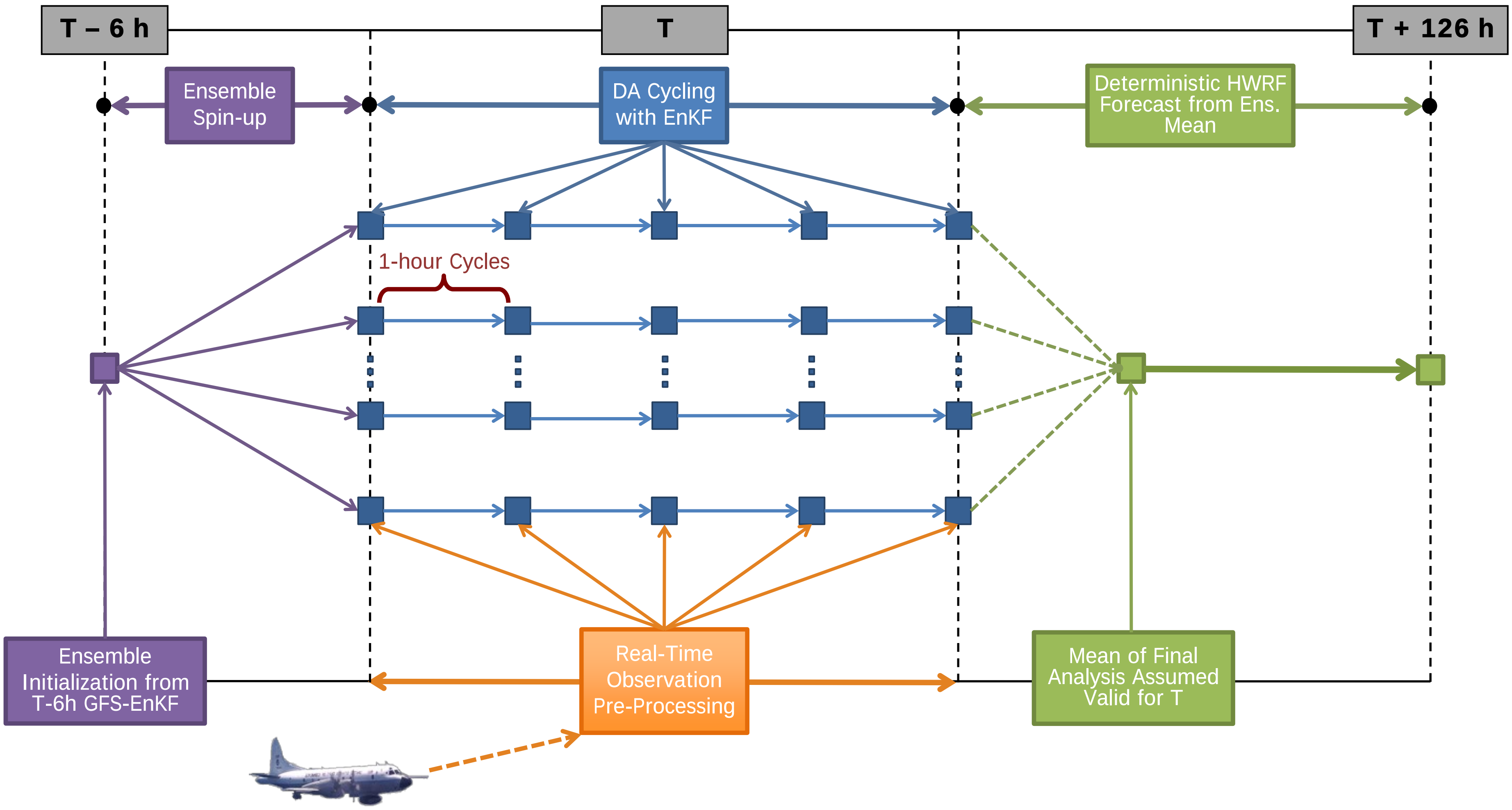
HRD EnKF Effort Primarily Focusing on
Dropsonde, Doppler Radar, Flight-Level, and *SFMR* Obs

G-IV Aircraft (environmental)

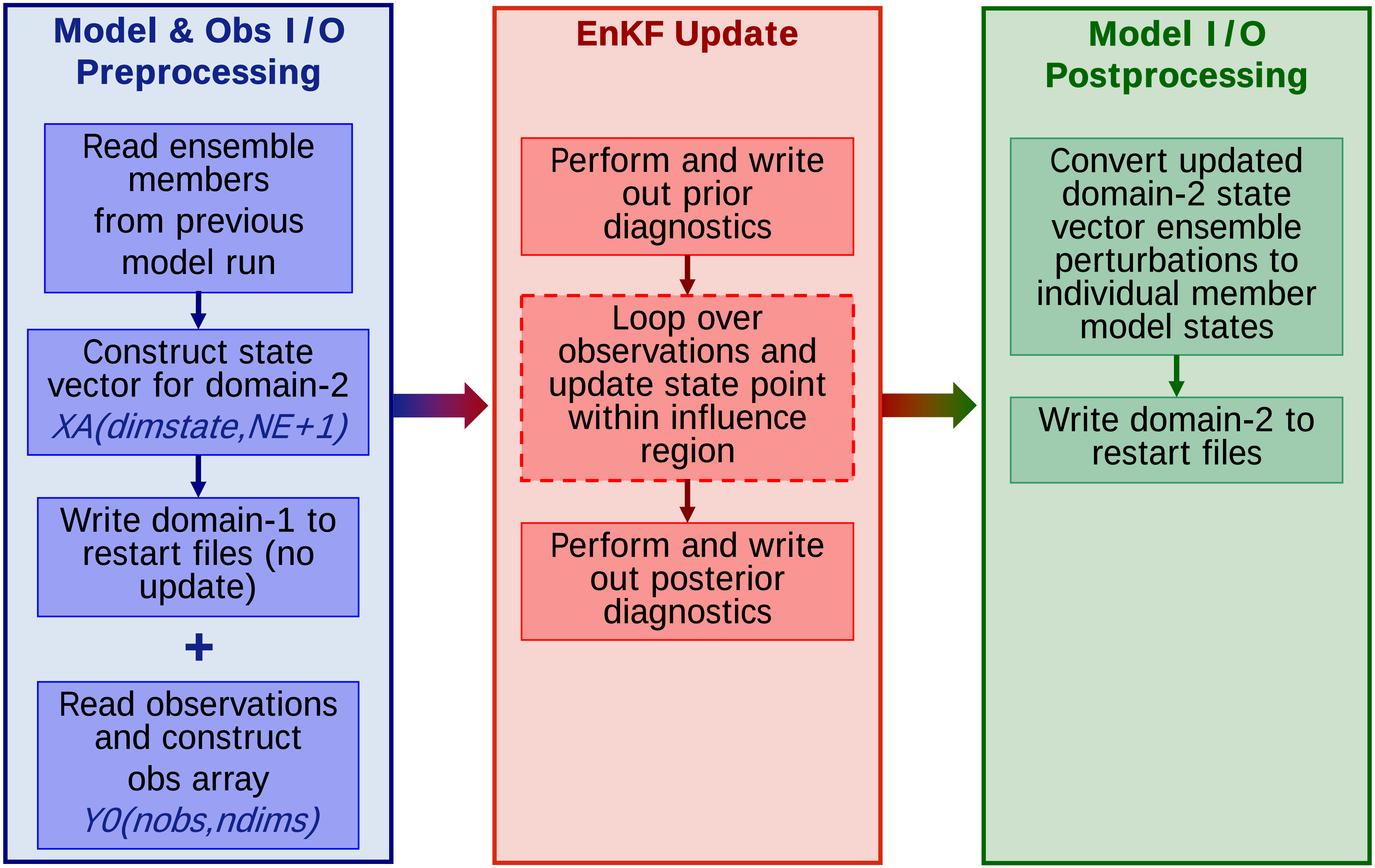


GPS navigation
Tail Doppler Radar (new)
C-Band Weather Radar
Stepped-frequency Microwave Radiometer
Dropwindsonde chute

HEDAS Cycling Flow



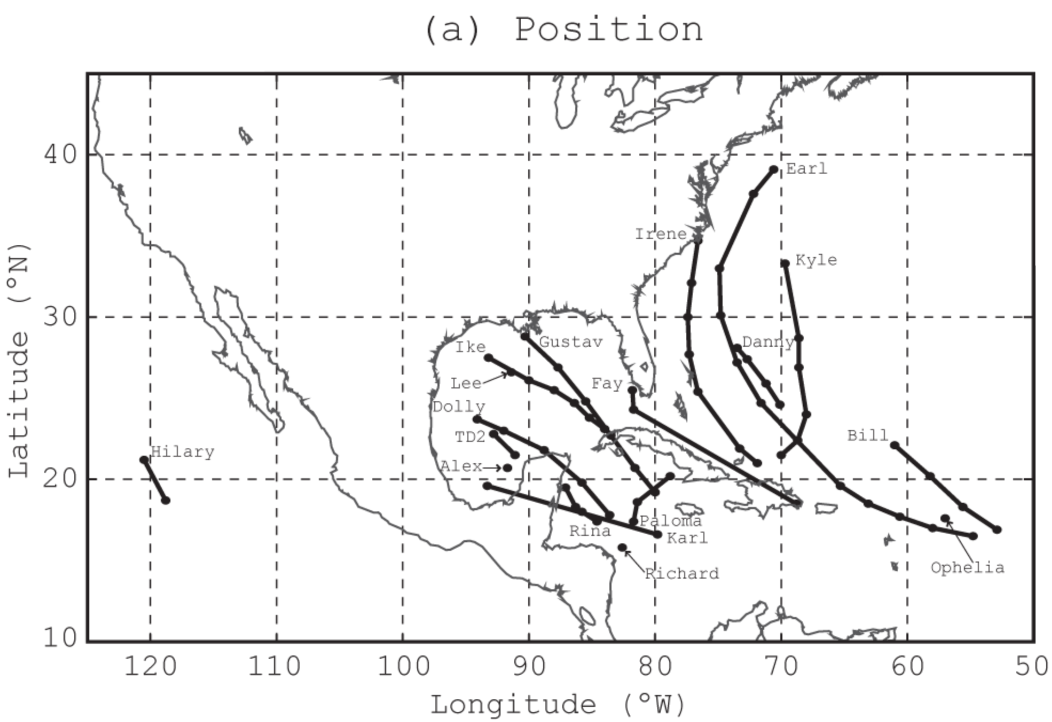
HEDAS EnKF Workflow

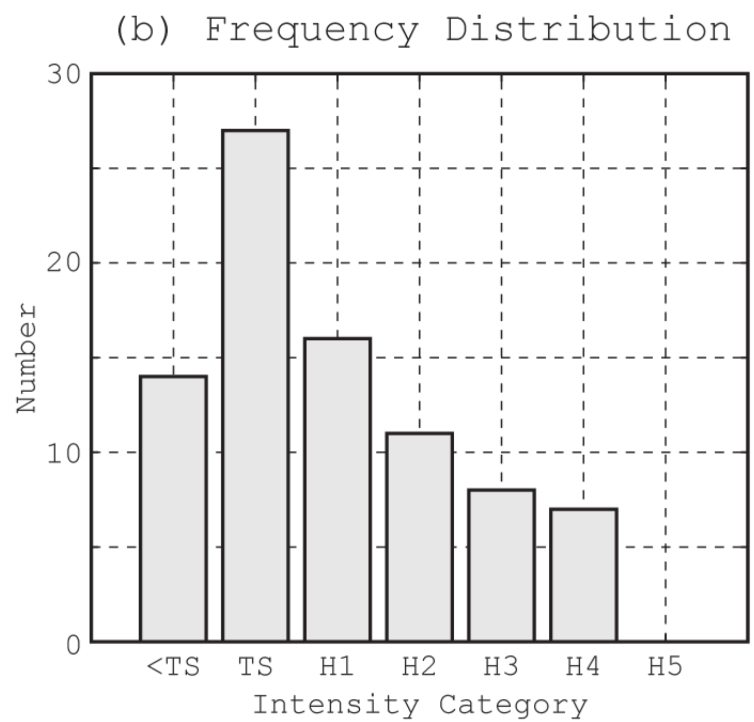


2008-2011 Real-Data Cases Considered

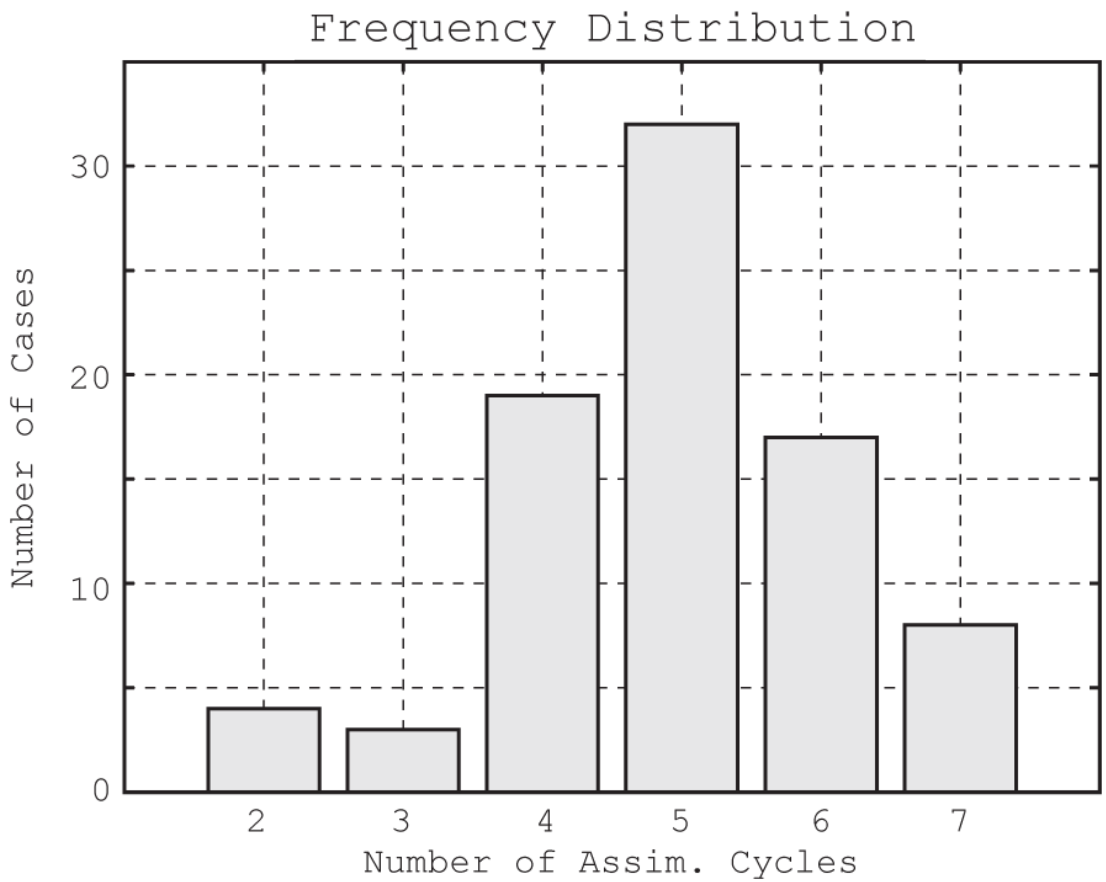
2008:		Ike	09-10-00Z	Danny	08-26-12Z	Karl	09-16-18Z
Dolly	07-20-12Z	Ike	09-10-12Z	Danny	08-27-00Z	Richard	10-23-06Z
Dolly	07-21-00Z	Ike	09-11-00Z	Danny	08-27-12Z	Tomas	11-04-00Z
Dolly	07-21-12Z	Ike	09-11-12Z	Danny	08-28-00Z	Tomas	11-04-12Z
Dolly	07-22-00Z	Ike	09-12-00Z	**2010:**		Tomas	11-15-00Z
Dolly	07-22-12Z	Ike	09-12-18Z	Alex	06-29-00Z	Tomas	11-06-12Z
Fay	08-14-12Z	Kyle	09-23-00Z	TD2	07-07-00Z	Tomas	11-07-00Z
Fay	08-15-00Z	Kyle	09-24-12Z	TD2	07-07-12Z	**2011:**	
Fay	08-15-06Z	Kyle	09-25-00Z	TD2	07-08-00Z	Irene	08-24-00Z
Fay	08-15-18Z	Kyle	09-25-12Z	Earl	08-29-00Z	Irene	08-24-12Z
Fay	08-18-18Z	Kyle	09-26-00Z	Earl	08-29-12Z	Irene	08-25-12Z
Fay	08-19-06Z	Kyle	09-26-18Z	Earl	08-30-00Z	Irene	08-26-00Z
Gustav	08-30-00Z	Kyle	09-27-00Z	Earl	08-30-12Z	Irene	08-26-12Z
Gustav	08-30-12Z	Kyle	09-27-18Z	Earl	08-31-00Z	Irene	08-27-00Z
Gustav	08-31-00Z	Paloma	11-07-06Z	Earl	09-01-12Z	Irene	08-27-12Z
Gustav	08-31-12Z	Paloma	11-07-18Z	Earl	09-02-00Z	Lee	09-02-00Z
Gustav	09-01-00Z	Paloma	11-08-18Z	Earl	09-02-12Z	Ophelia	09-24-18Z
Gustav	09-01-12Z	**2009:**		Earl	09-03-00Z	Hilary	09-28-18Z
Dolly	07-20-12Z	Ana	08-17-00Z	Earl	09-03-18Z	Hilary	09-29-18Z
Dolly	07-21-00Z	Bill	08-19-00Z	Earl	09-04-00Z	Rina	10-26-00Z
Dolly	07-21-00Z	Bill	08-19-12Z	Karl	09-13-00Z	Rina	10-26-18Z
Dolly	07-20-12Z	Bill	08-20-00Z	Karl	09-13-12Z	Rina	10-27-00Z
Dolly	07-21-00Z	Bill	08-20-12Z	Karl	09-14-00Z	Rina	10-27-18Z

Distribution of Cases

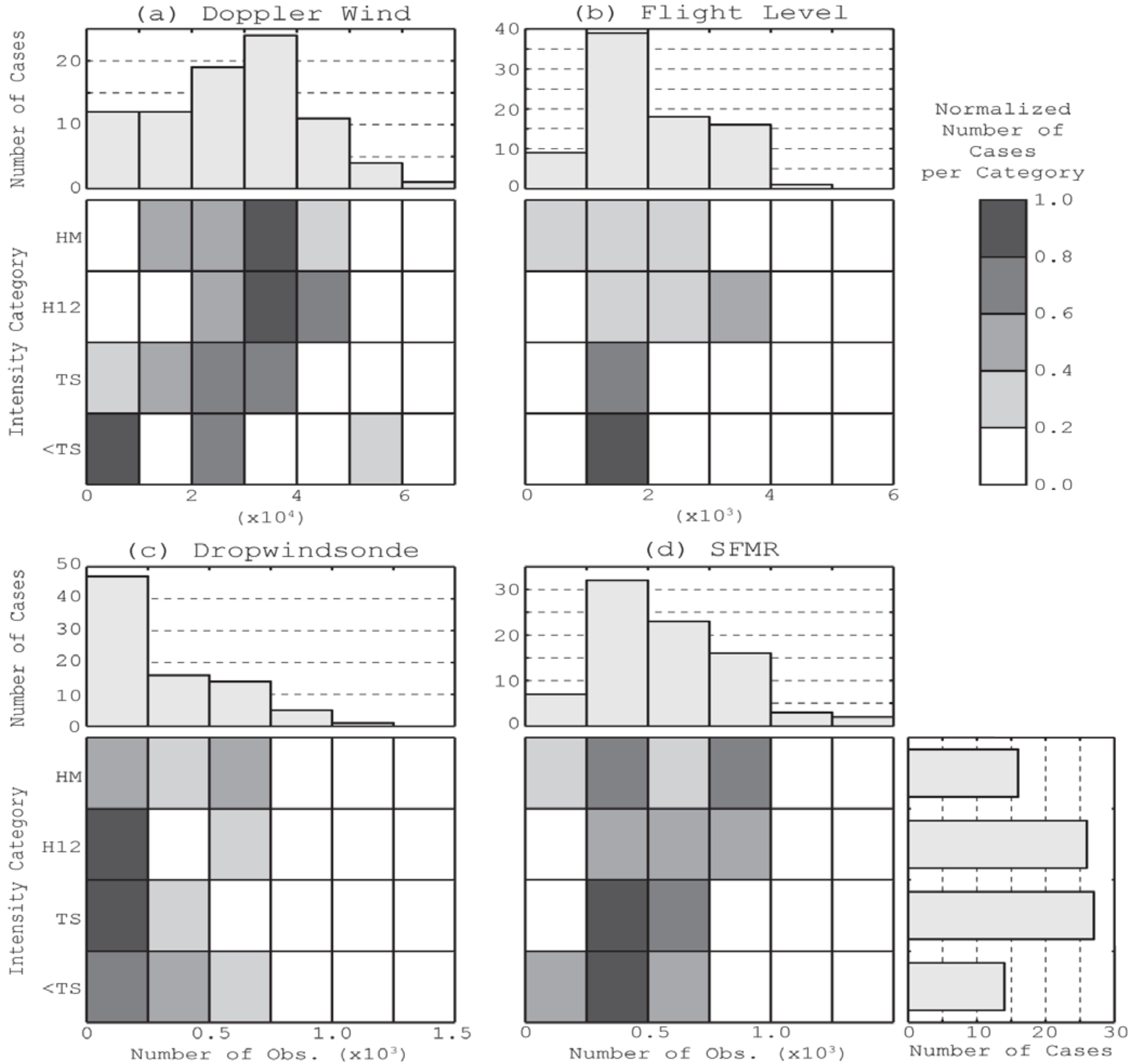




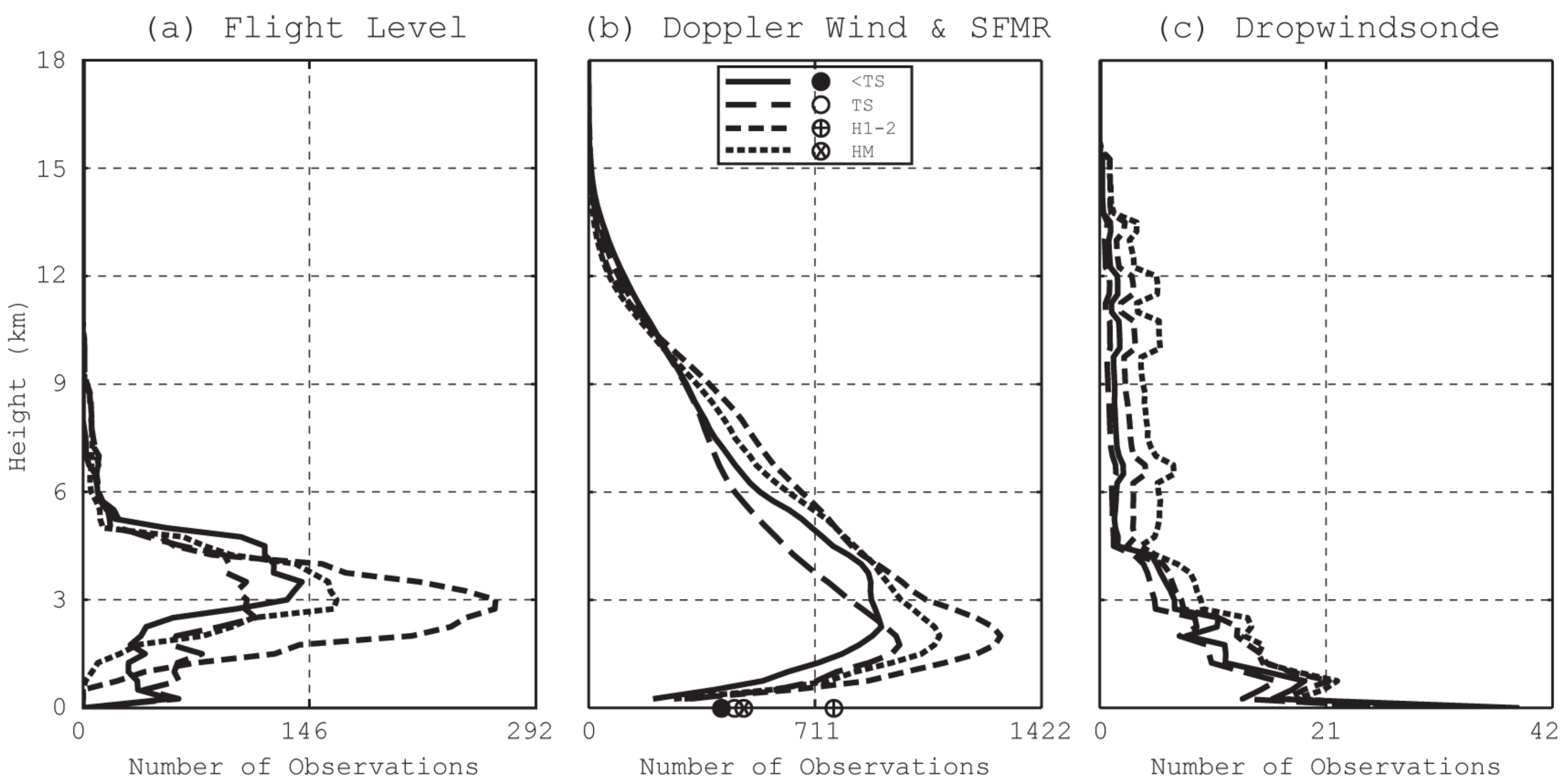
Number of Assimilation Cycles



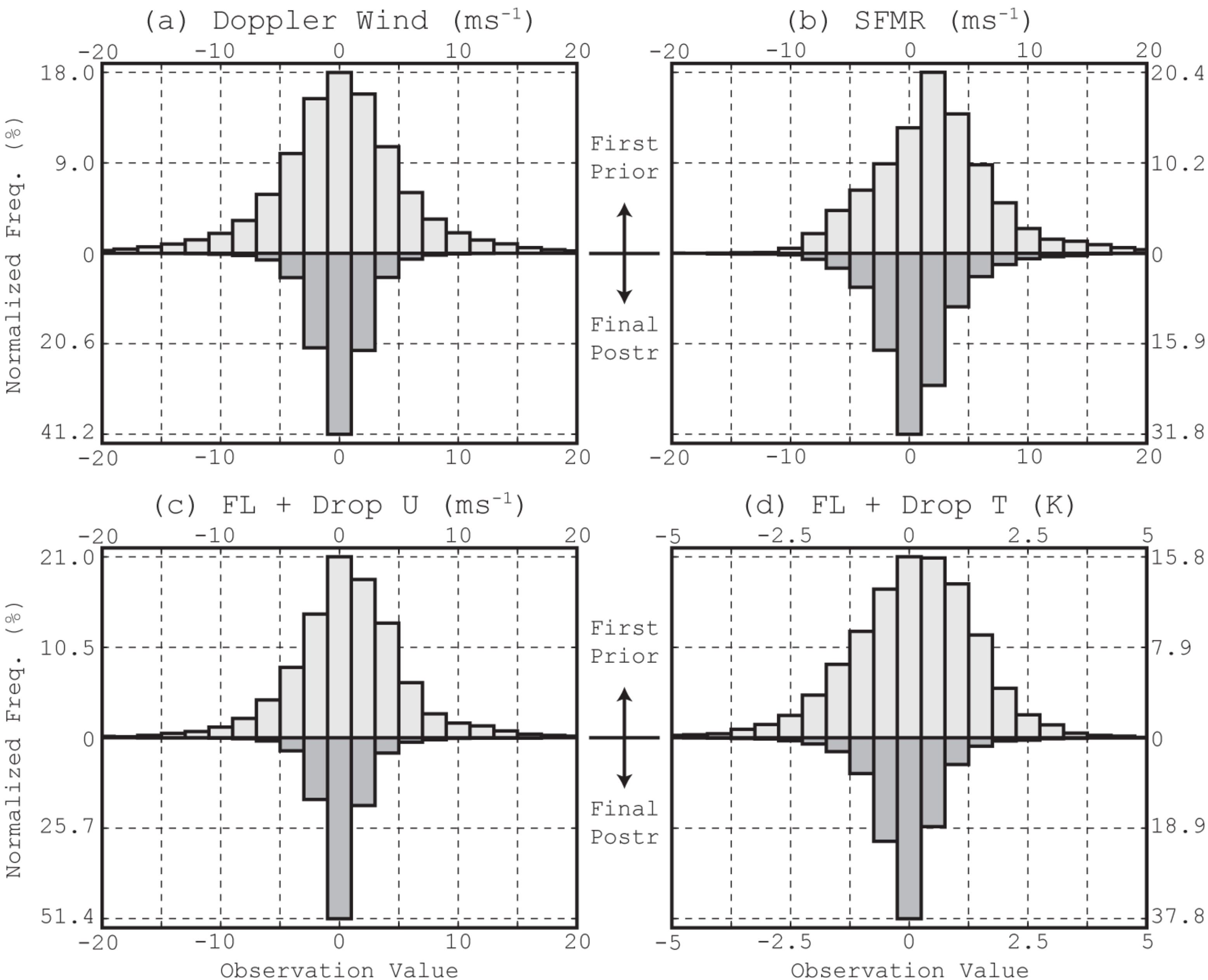
Number of Observations Assimilated



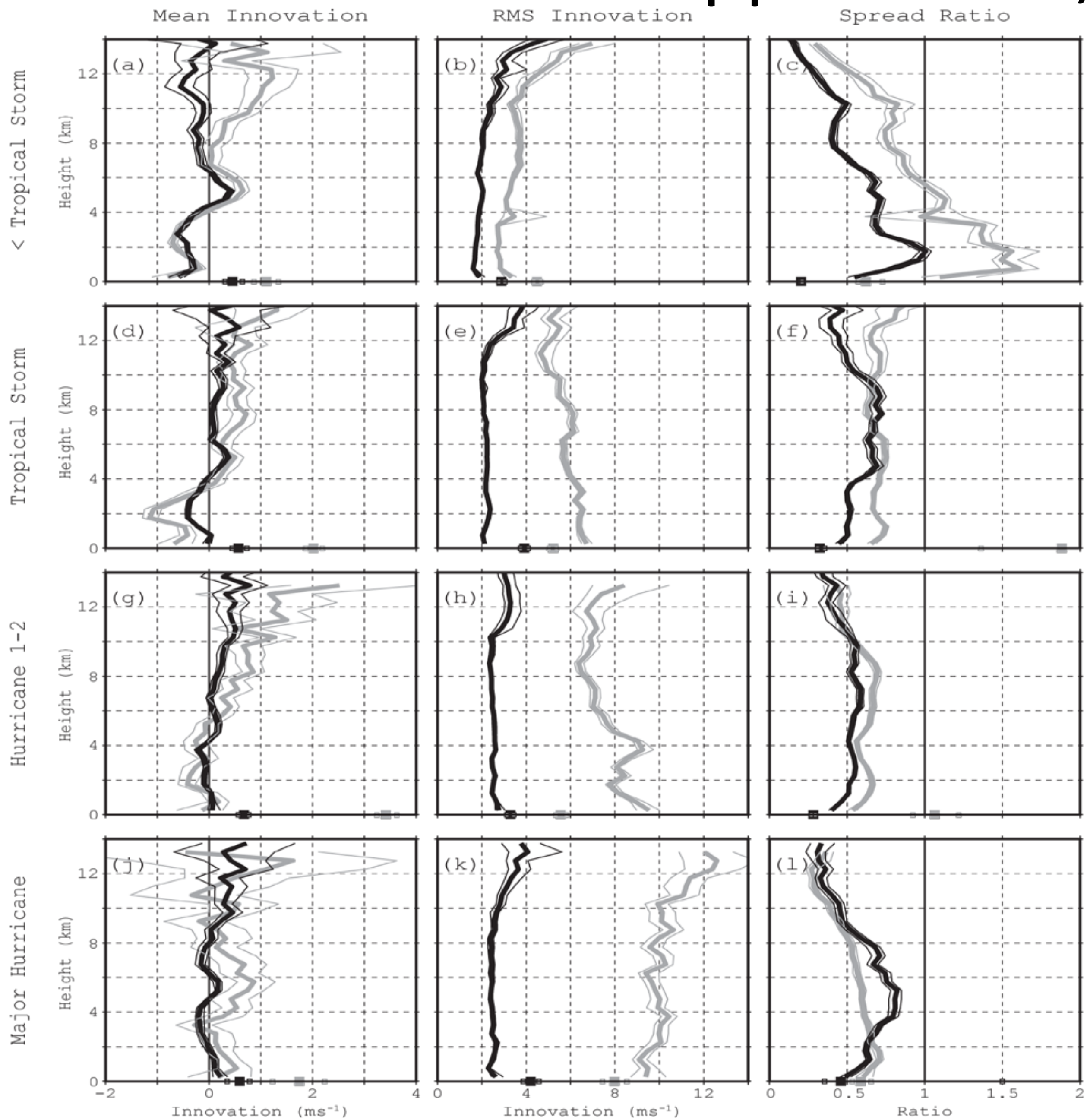
Number of Observations Assimilated



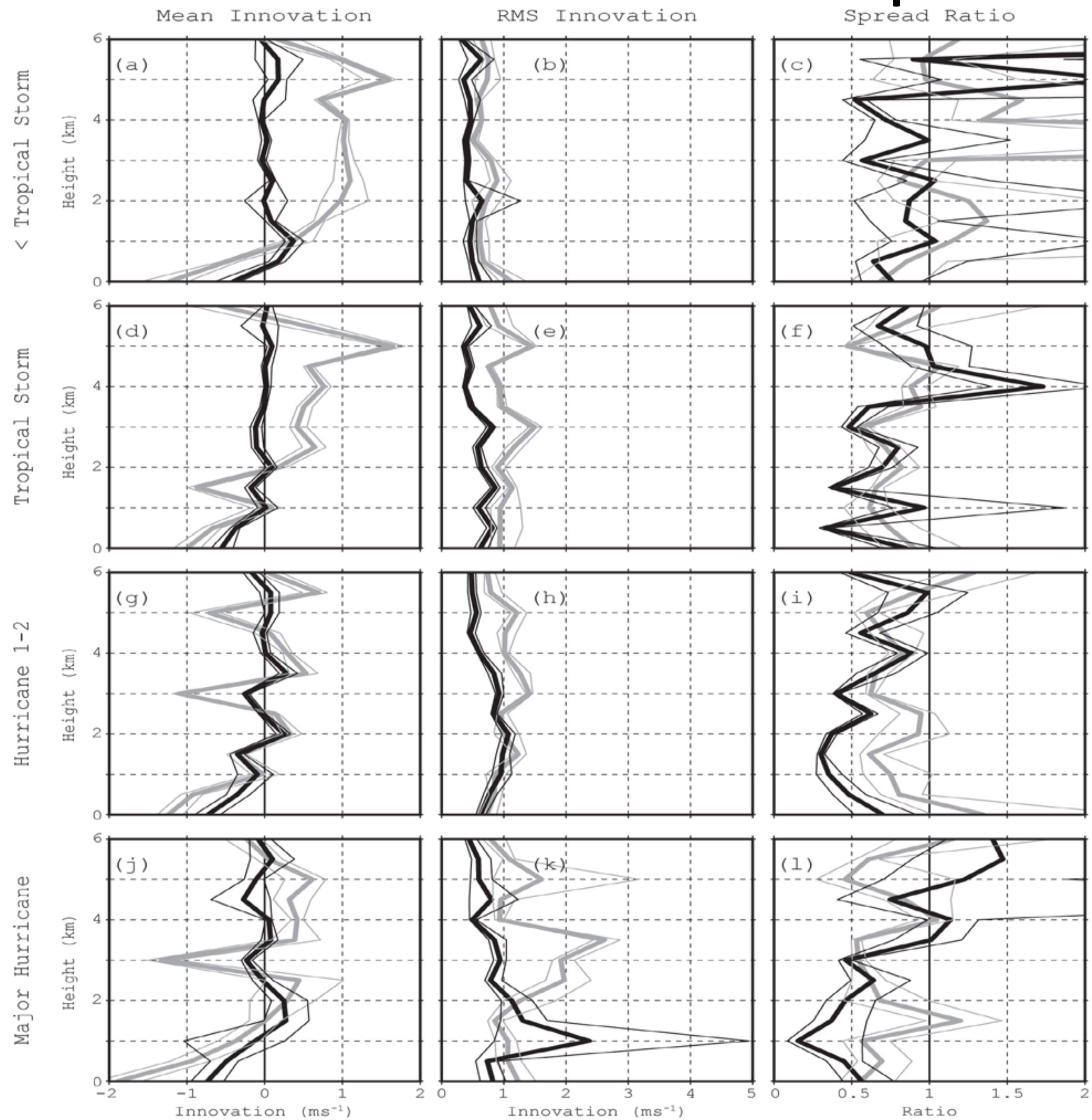
Distribution of Innovations (O-F & O-A)



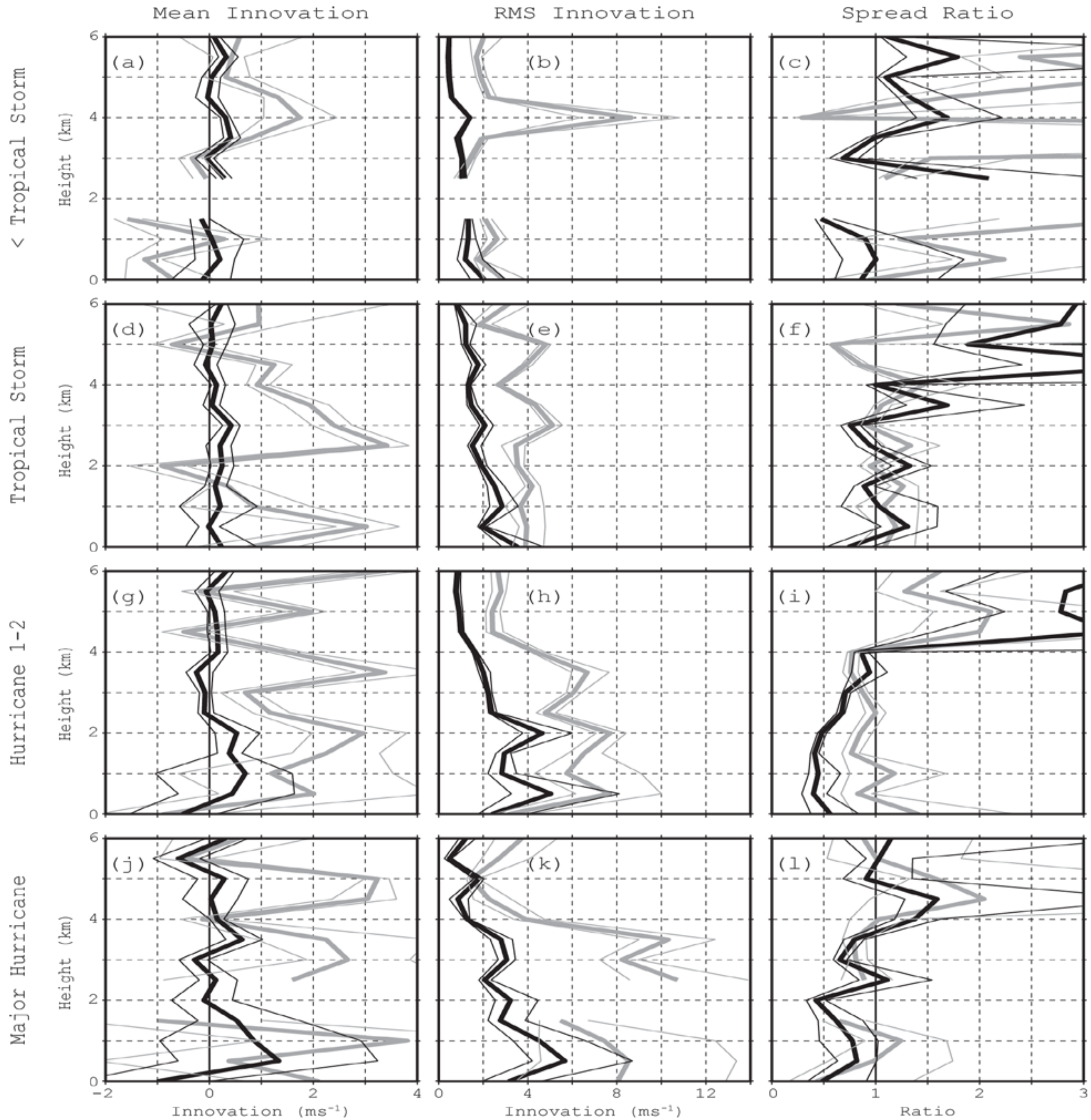
Innovation Statistics – Doppler Wind, SFMR



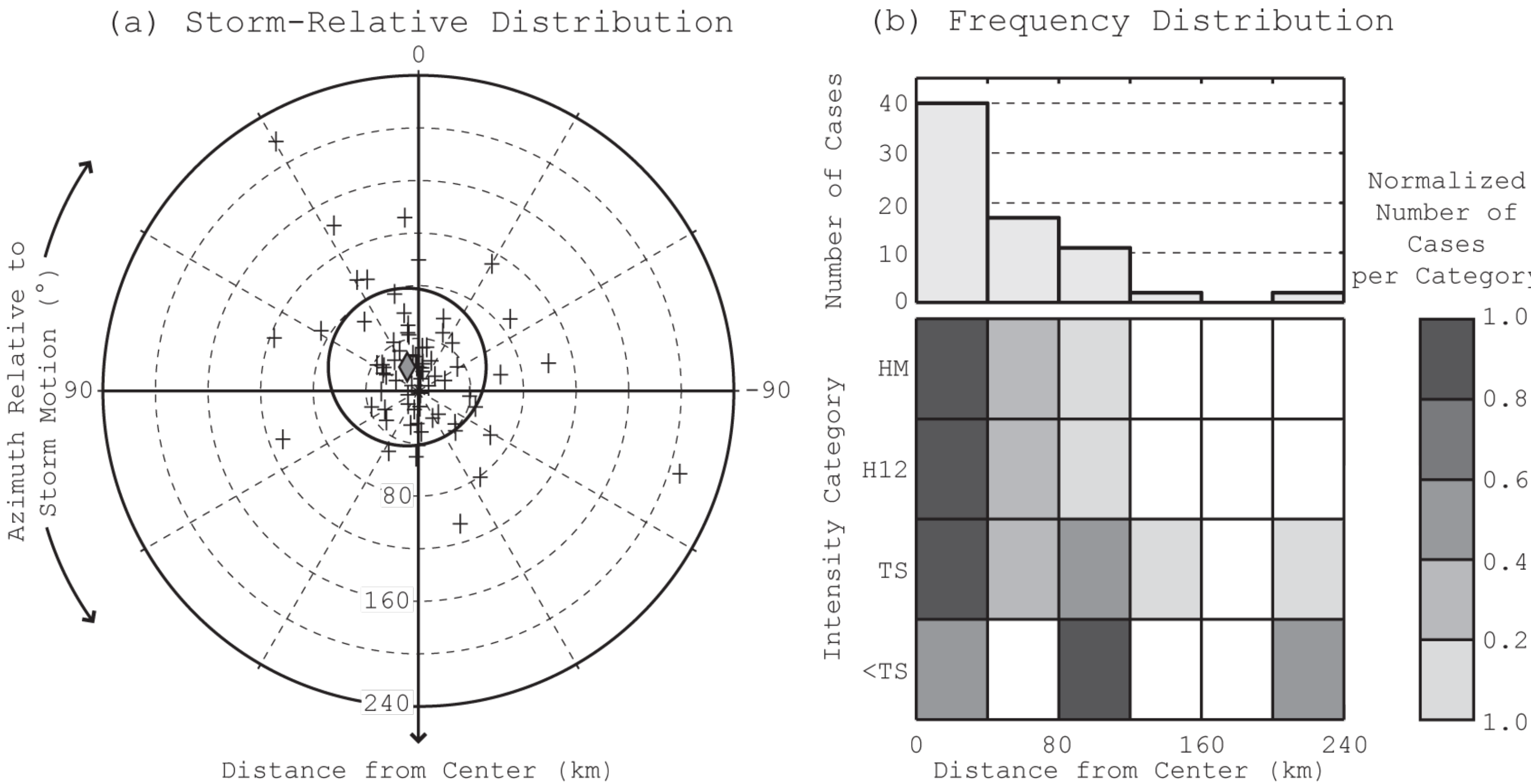
Innovation Statistics – Temperature



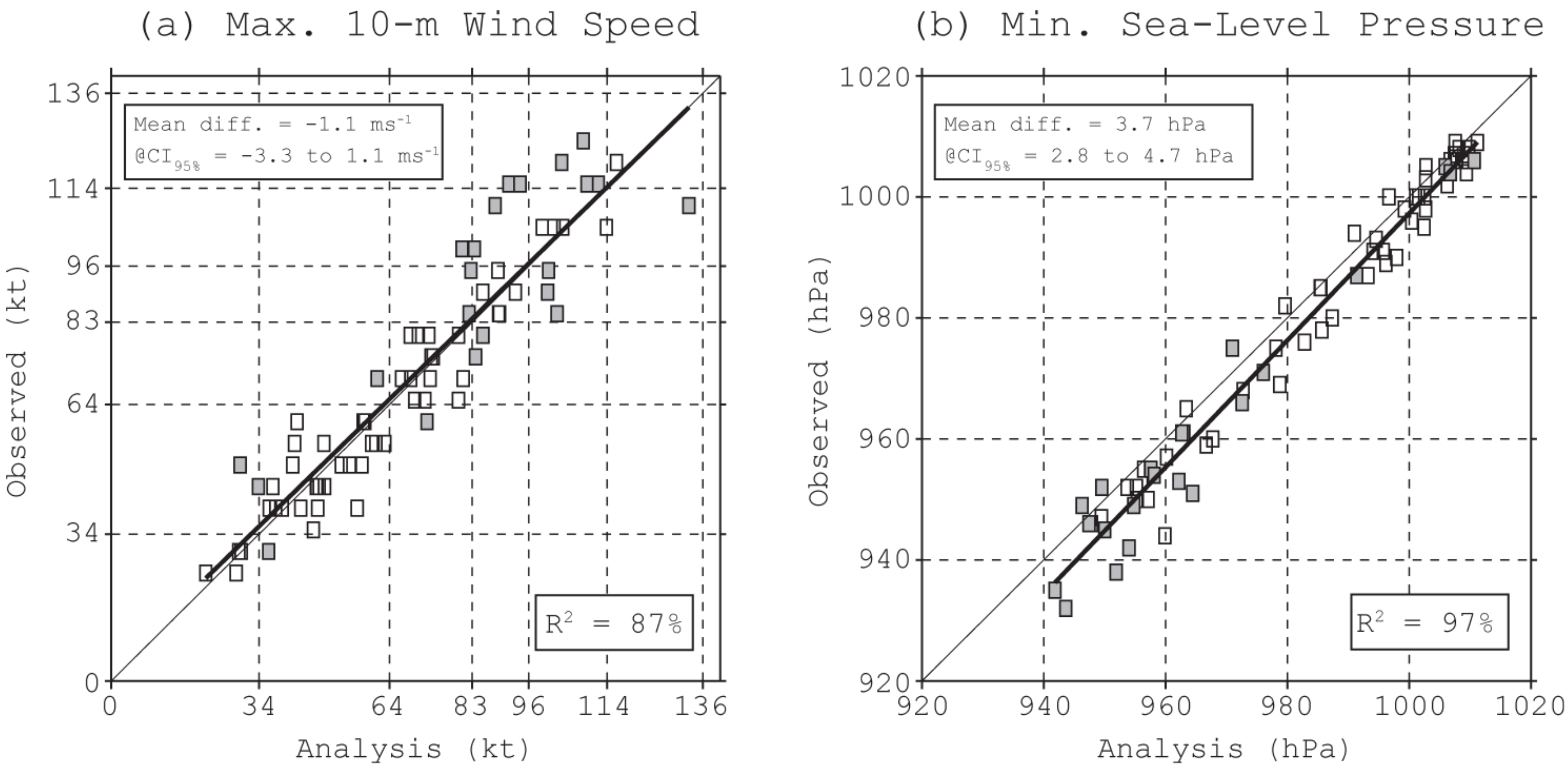
Innovation Statistics – Zonal Wind Speed



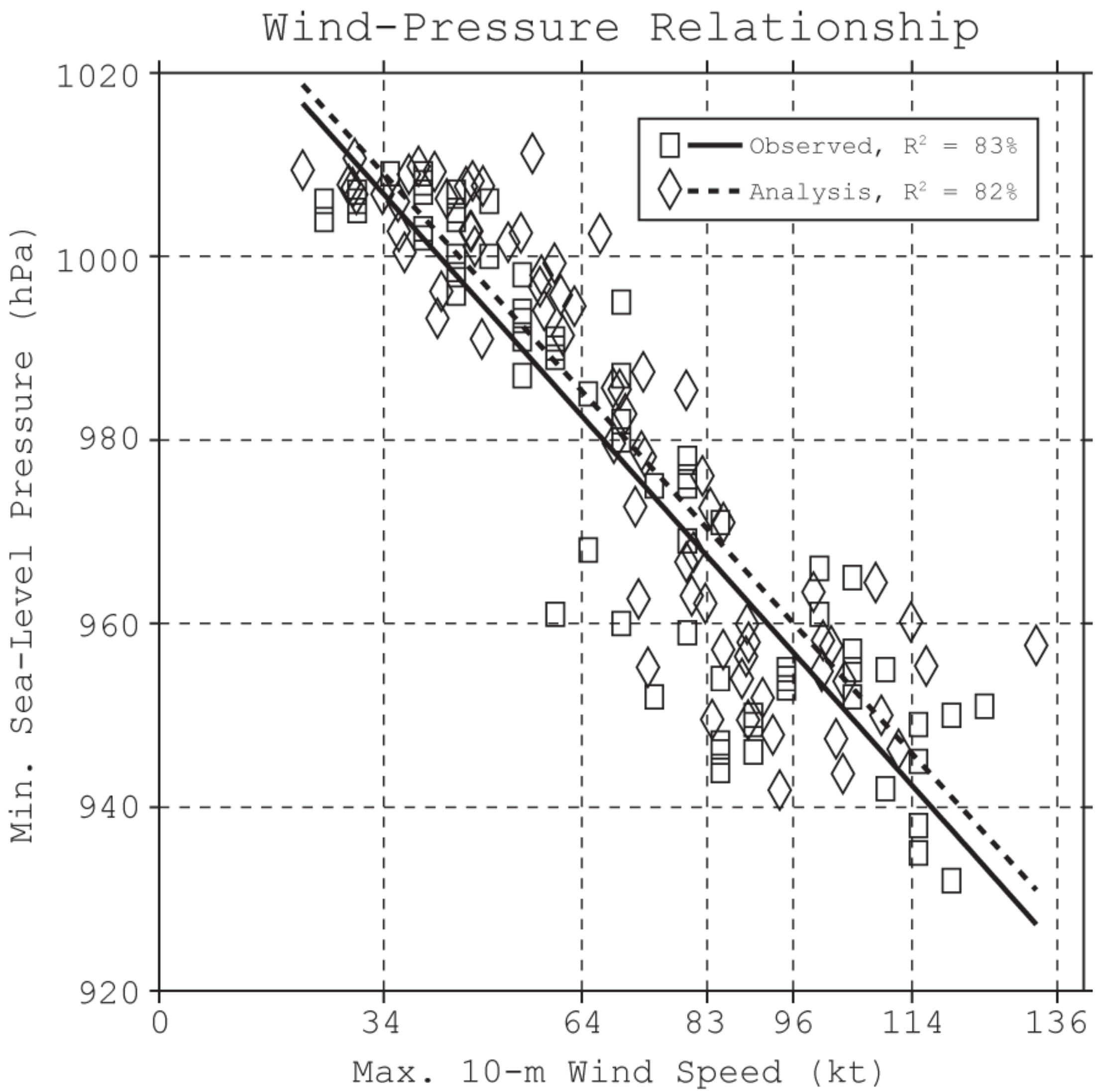
Analysis Position Error



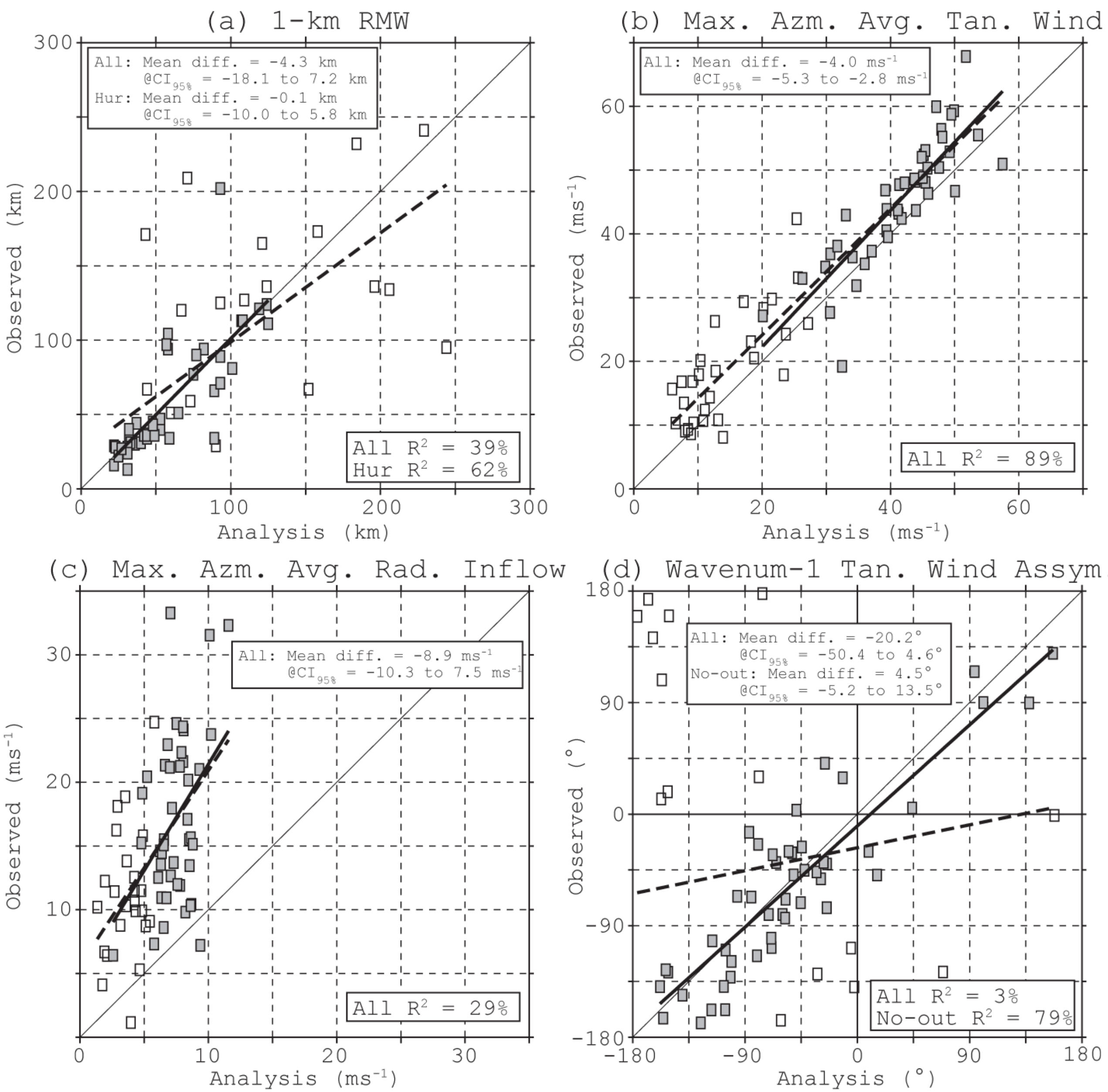
Analysis Intensity Error



Analysis Wind-Pressure Relationship

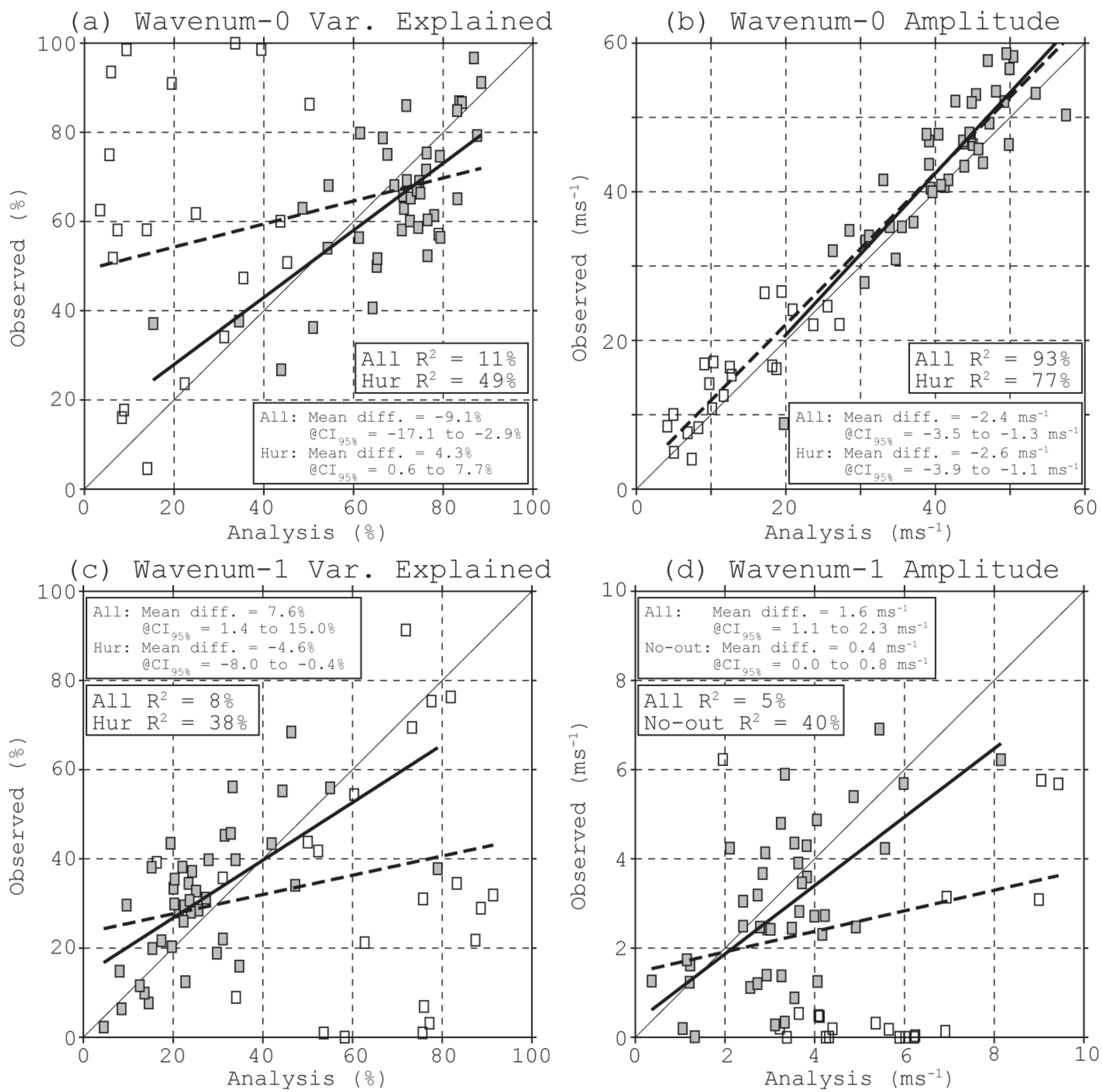


Analysis Doppler-Derived Structure



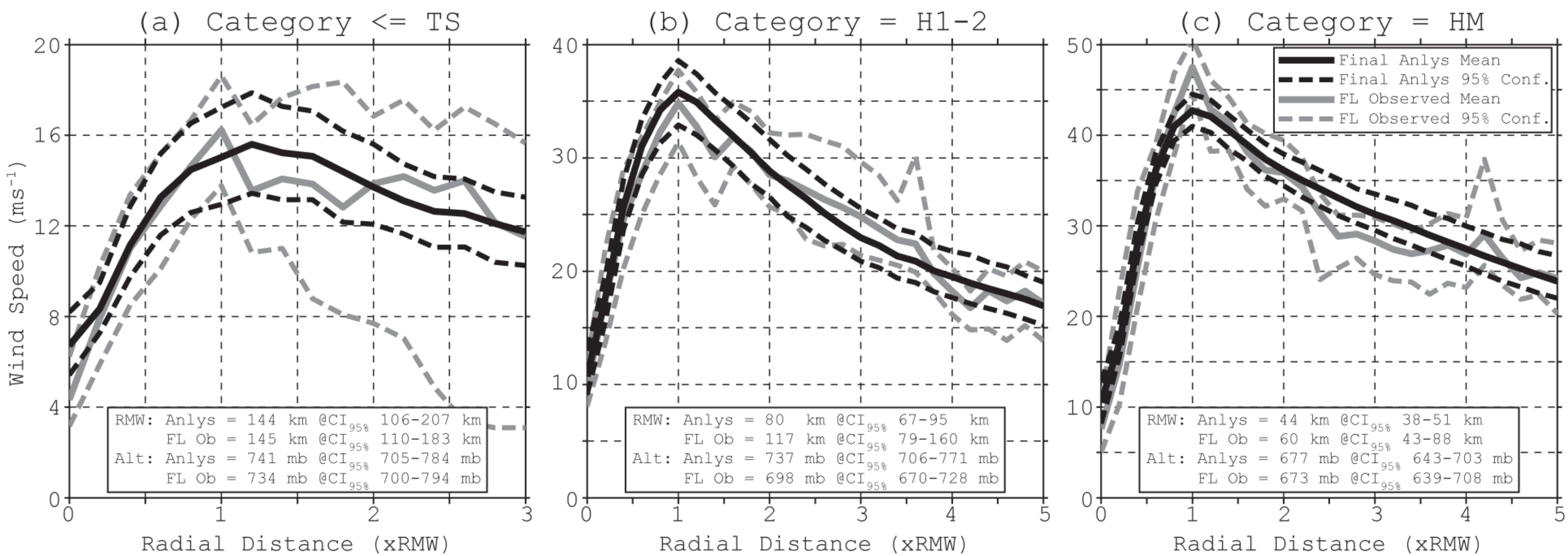
2008 Fay (1)
2008 Gustav (2)
2008 Ike (1)
2008 Kyle (4)
2008 Paloma (1)
2009 Danny (1)
2010 Earl (1)
2010 Karl (2)
2010 Tomas (1)
2011 Irene (2)

Waveno. Decomposition – Doppler vs. Analysis

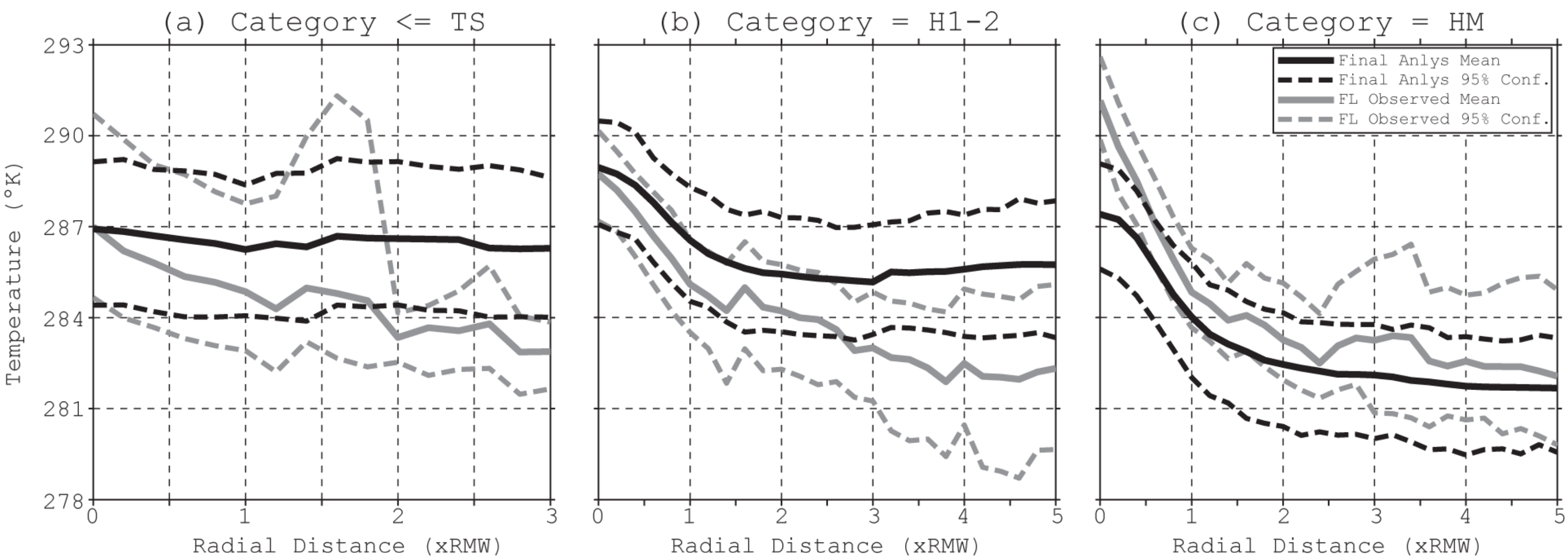


2008 Fay (1)
2008 Gustav (2)
2008 Ike (4)
2008 Kyle (7)
2008 Paloma (1)
2009 Bill (1)
2009 Danny (1)
2011 Irene (2)
2011 Hilary (1)

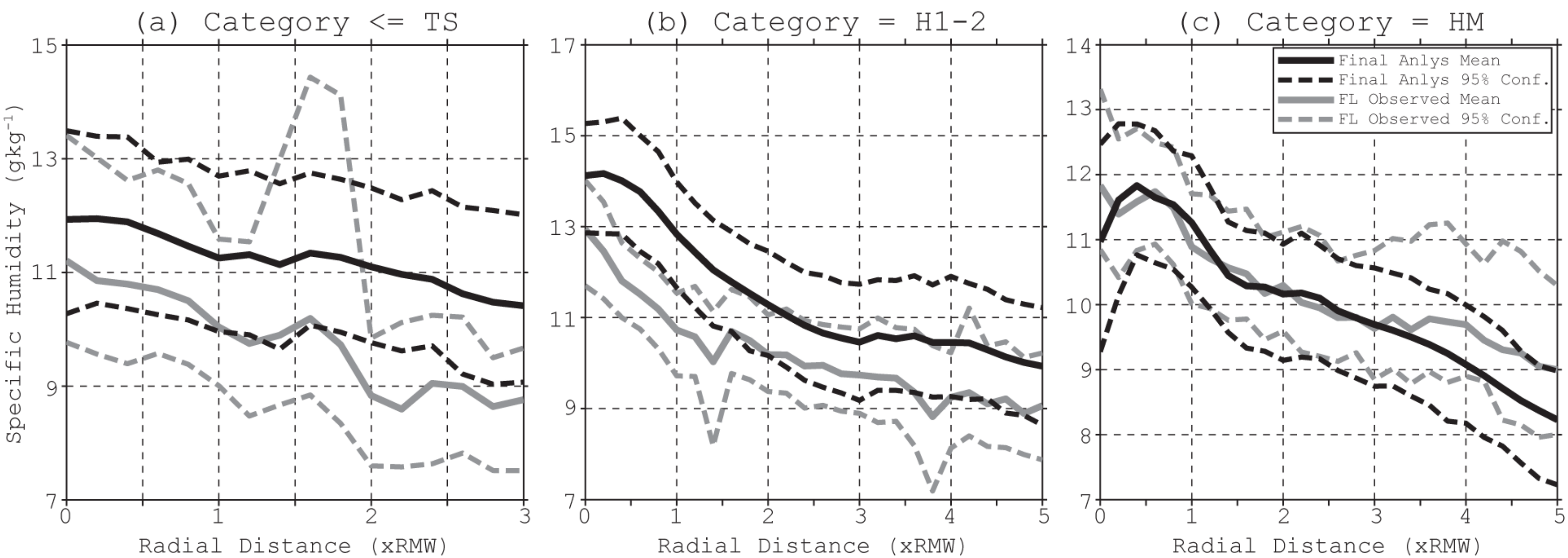
Radial Profiles – FL Wind Speed



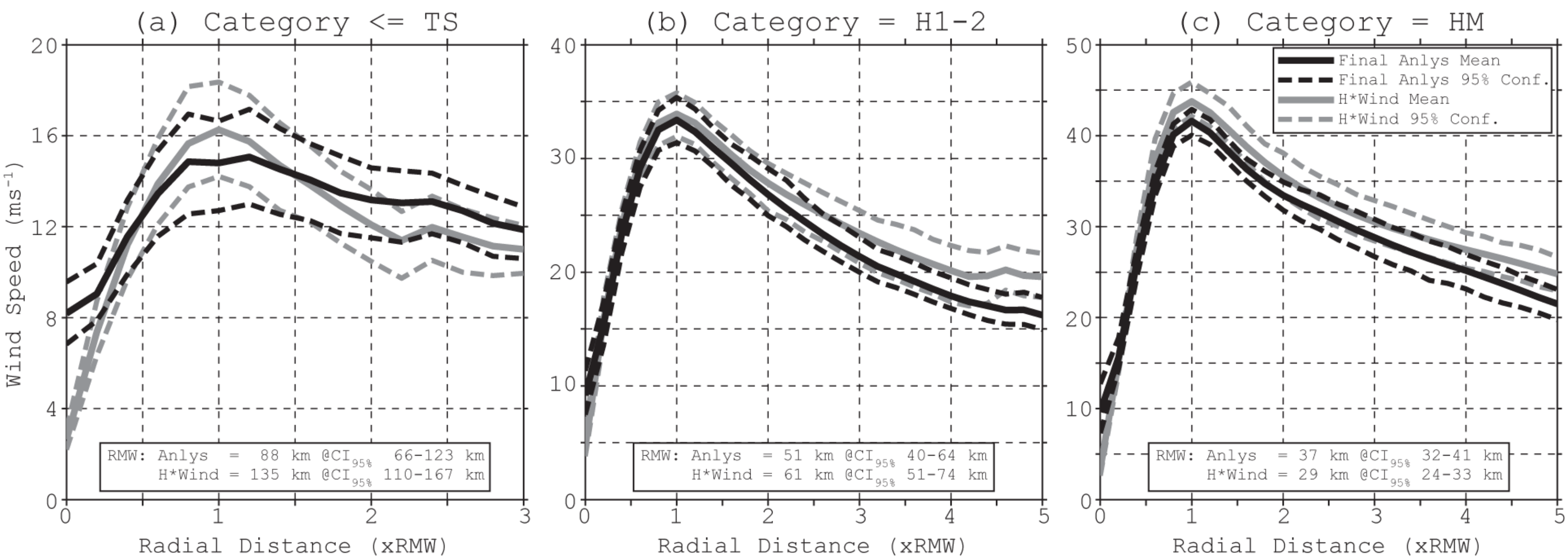
Radial Profiles – FL Temperature



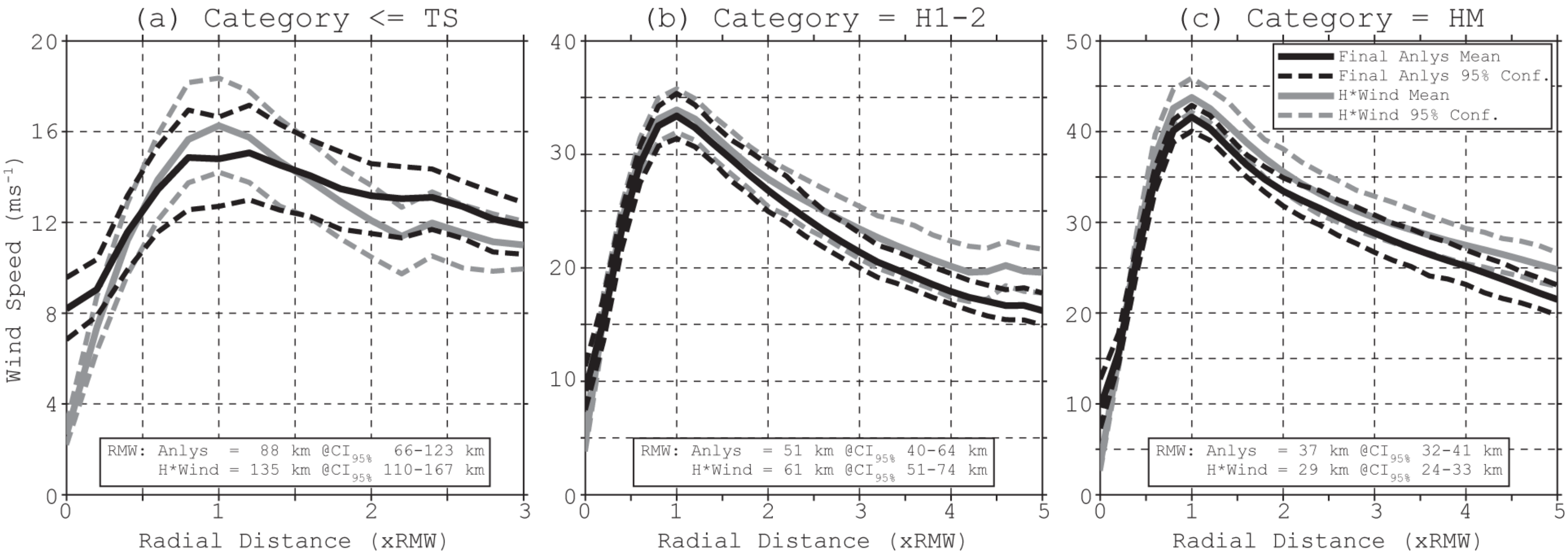
Radial Profiles – FL Specific Humidity



Radial Profiles – H*Wind Surface Wind Speed



Radial Profiles – H*Wind Surface Wind Speed



Summary - 1

- HEDAS runs with 83 cases from 2008-2011 seasons are analyzed
- Realistic distribution of cases by intensity (most are tropical storm to cat-1 hurricanes)
- In a typical run:
  - 4-5 cycles of data are assimilated
  - 30-40K observations are assimilated, but Doppler wind observations dominate
  - Best sampling is achieved below ~9 km
- Improvements in observation space are seen that can be directly associated with data assimilation

Summary - 2

- Position errors in analyses are on average ~20 km compared to the best track, tend to be "ahead" of the best track position as assim. cycles extend beyond the synoptic time
- Good fit of analysis intensity to best track is seen, but analyses are systematically weaker in MSLP by ~3 mb
- Structurally,
  - RMW fit much better for hurricanes
  - Azimuthal wavenumber-1 asymmetry is impacted by few outliers, especially from the 2008 season
  - Wavenumber 0 and 1 structures well captured