Tropical Cyclone Forecast Assessment

Zachary Weller

Colorado State University

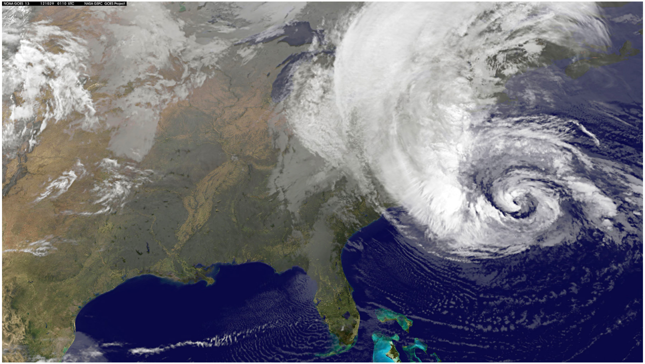
wellerz@stat.colostate.edu

Joint work with Dr. Jennifer Hoeting (CSU) and Dr. Ligia Bernardet (NOAA)

June 24, 2014

Motivation

- National Hurricane Center - tropical cyclone (TC) predictions since 1954
- Forecasts have typically improved each year, but are always seeking improvement due to potentially large economic and societal impacts



TC Models

- Forecast models are typically initialized every 6 hours producing updated forecasts.
- Typical output for an individual model run (forecast metrics):
  1. Lead time for prediction, given in 6 hour increments
  2. Track (location) given in lat/lon
  3. Minimum Sea Level Pressure (MSLP) given in millibars
  4. Intensity (1 minute max. sustained wind) given in knots
  5. Radii of sustained winds (34, 50, and 64 knots) for each storm quadrant

TC Models

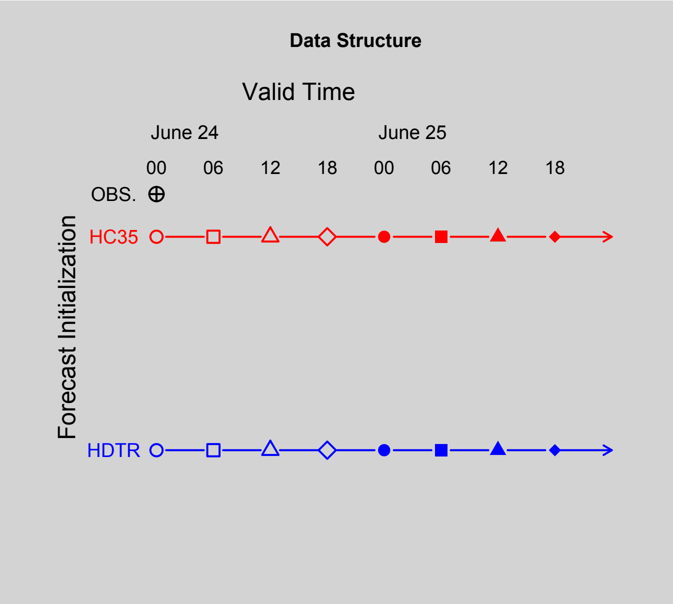
- Forecast models are typically initialized every 6 hours producing updated forecasts.
- Typical output for an individual model run (forecast metrics):
  1. Lead time for prediction, given in 6 hour increments
  2. Track (location) given in lat/lon
  3. Minimum Sea Level Pressure (MSLP) given in millibars
  4. Intensity (1 minute max. sustained wind) given in knots
  5. Radii of sustained winds (34, 50, and 64 knots) for each storm quadrant
- Current models of interest: dynamical models HC35 (control) and HDTR (experimental)
- Working Data:
  - Data from 2012 Atlantic Hurricane Season
  - Forecasts for 17 of 19 total storms
  - Variable number of forecasts from each model for each storm

2012 Atlantic Hurricanes

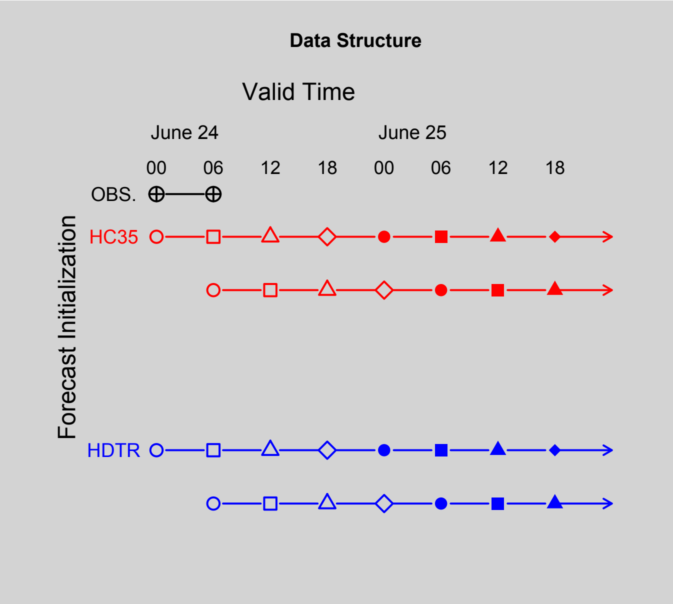
Verification Techniques

- After a storm, best track data are compiled (i.e. the "truth")
  - contains information on the "observed" values: track, MSLP, and intensity, etc.
  - data are compiled through a reanalysis
  - some uncertainty in these data
- Forecast model output is compared to best track for each forecast metric separately, creating prediction errors
- Traditionally, univariate analysis is done on the prediction errors for each forecast metric for a given lead time
- Models are often compared via a homogeneous comparison: prediction errors from the two models are matched on storm, forecast initialization time, and lead time
- For example: compare intensity errors at the 24 hour lead time from the two models via a paired t-test

Data Format

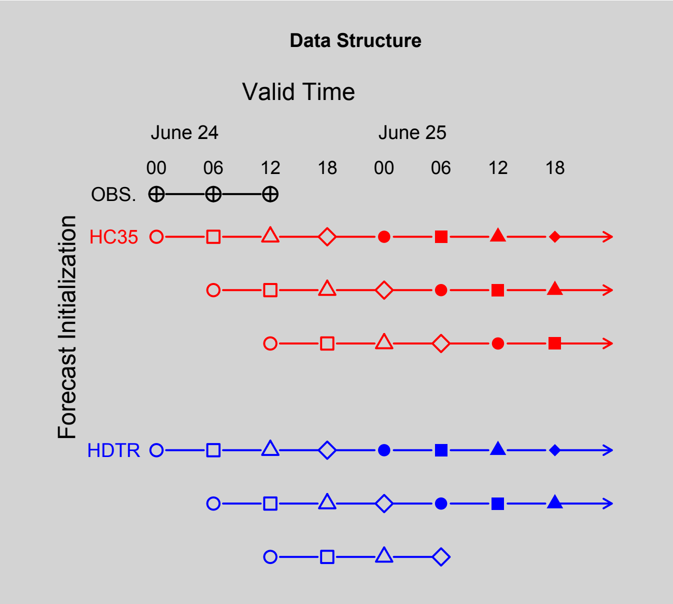

Data Structure
Valid Time
June 24
June 25
00
06
12
18
00
06
12
18
OBS.
HC35
HDTR
Forecast Initialization

Data Format



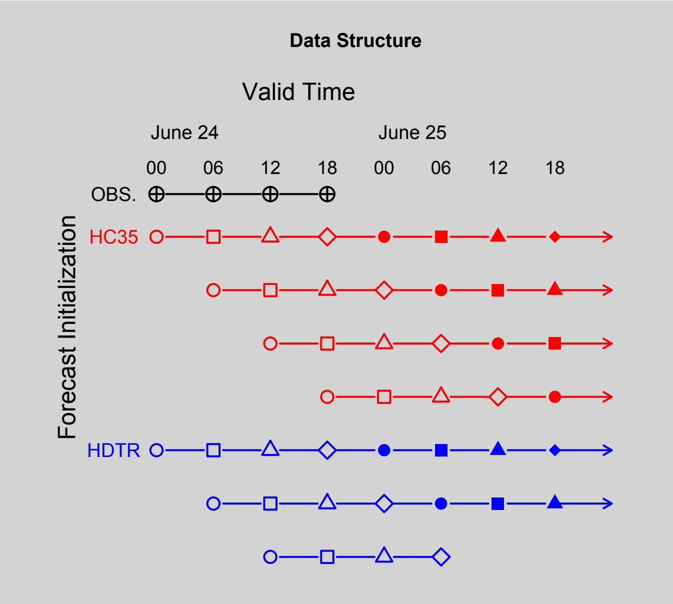
Data Structure
Valid Time
June 24
June 25
00
06
12
18
00
06
12
18
OBS.
HC35
HDTR
Forecast Initialization

Data Format



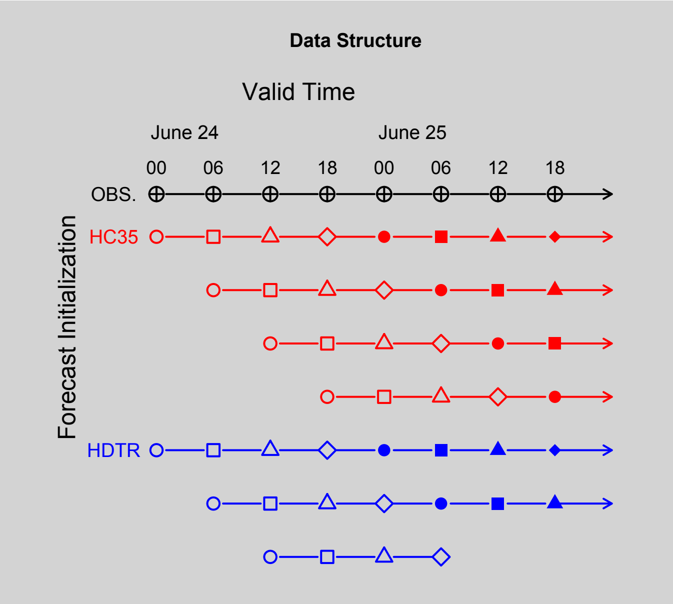
Data Structure
Valid Time
June 24
June 25
00
06
12
18
00
06
12
18
OBS.
Forecast Initialization
HC35
HDTR

Data Format



Data Structure
Valid Time
June 24
June 25
00
06
12
18
00
06
12
18
OBS.
HC35
HDTR
Forecast Initialization

Data Format



Data Structure
Valid Time
June 24
June 25
00
06
12
18
00
06
12
18
OBS.
HC35
HDTR
Forecast Initialization

Goals

1. Help hurricane modelers evaluate model performance for various metrics
   - Construct statistical tests to compare TC forecasting model performance. Ideally, could compare more than two models simultaneously and tests would be multivariate.
   - Graphical methods to evaluate model performance.

Goals

1. Help hurricane modelers evaluate model performance for various metrics
   - Construct statistical tests to compare TC forecasting model performance. Ideally, could compare more than two models simultaneously and tests would be multivariate.
   - Graphical methods to evaluate model performance.
2. Understand effects of (spatial and temporal) correlation in the data on the results of above tests and account for this correlation.
   - Adjust p-values and CI's used in model comparison.
   - Reduce computation time needed for retrospective model runs?

Goals

1. Help hurricane modelers evaluate model performance for various metrics
   - Construct statistical tests to compare TC forecasting model performance. Ideally, could compare more than two models simultaneously and tests would be multivariate.
   - Graphical methods to evaluate model performance.
2. Understand effects of (spatial and temporal) correlation in the data on the results of above tests and account for this correlation.
   - Adjust p-values and CI's used in model comparison.
   - Reduce computation time needed for retrospective model runs?
3. Understand the relationship between the dynamical model forecast errors and errors in the dynamical model's 3-D environment.
   - For forecasting intensity (max winds), statistical forecasting models still out-perform dynamical forecasting models
   - Statistical Hurricane Intensity Prediction System (SHIPS)
   - Example: errors in the sea surface temperature may be related to errors in the intensity forecast.

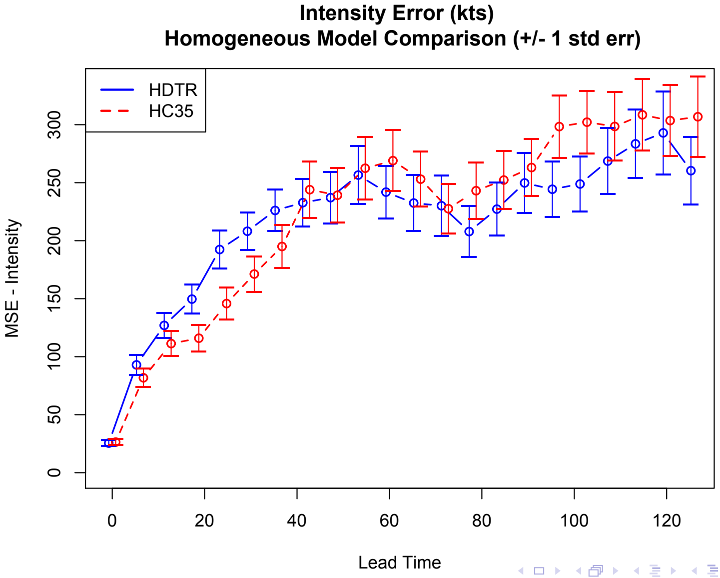
Some Verification Literature

1. Statistical focus:
    - Probabilistic forecasts (Murphy and Winkler, 1987)
    - Spatial methods (Gilleland, et. al. 2009)
    - Ensembles and proper scoring (Gneiting and Raftery, 2007)
    - Assessing multivariate forecasts (Gneiting, et. al. 2008)
2. Atmospheric Science focus:
    - Progress and Challenges in Verification (Ebert, et. al. 2013)
    - Forecast Verification: A Practitioner's Guide in Atmospheric Science (Jolliffe and Stephenson 2012)
    - New Techniques to Assess Wind Radii Forecasts and Storm Asymmetry (Davis et. al. 2010)
    - Developmental Testbed Center (DTC): yearly summaries of model performance (Bernardet, et. al. 2012)

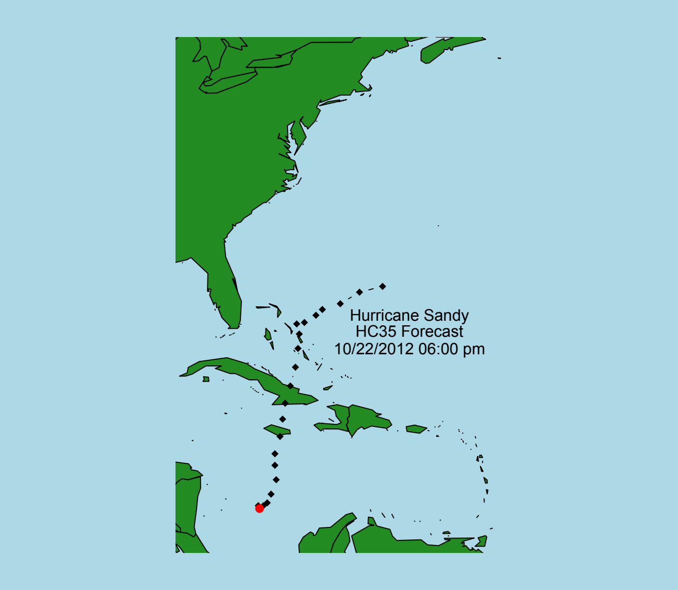
Paired t-tests

1. Paired t-tests are traditionally used to compare the performance of two forecasting models at a given lead time.
   - For the number of storms, let $i \in \{1, 2, \ldots, I\}$
   - For the number of forecasts for storm i, let $j \in \{1, 2, \ldots, n_i\}$
   - For the total sample size, let $N = \sum_{i=1}^{I} n_i$
   - Let $y_{ij} = |e_{ij}^{HC35}| - |e_{ij}^{HDTR}|$, where $e_{ij}^{M} = Forecast - Observed$ values for a forecasting metric (e.g. intensity) from forecasting model M for a fixed lead time (e.g. 24 hours)
2. Consider the model:
$$y_{ij} = \mu + \epsilon_{ij}. \tag{1}$$
3. We are interested in inference for μ (e.g., $\mu = 0$).
4. Challenge: multiple types of correlation

Example Plot



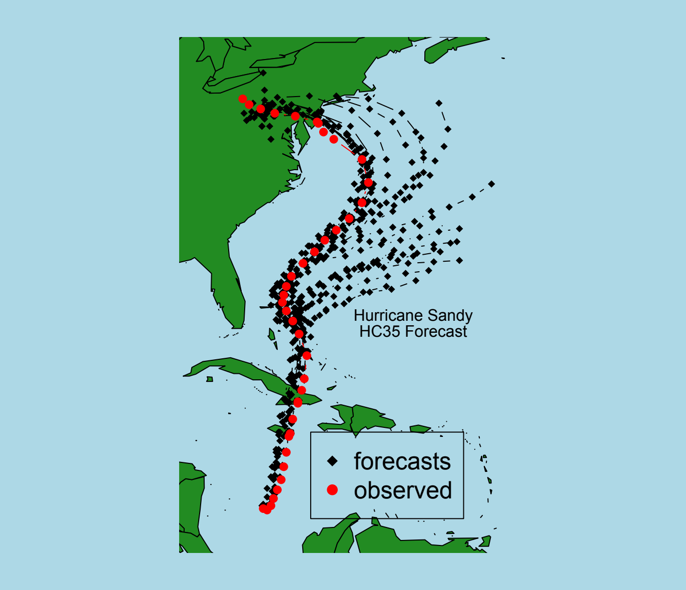
Correlation: Hurricane Sandy



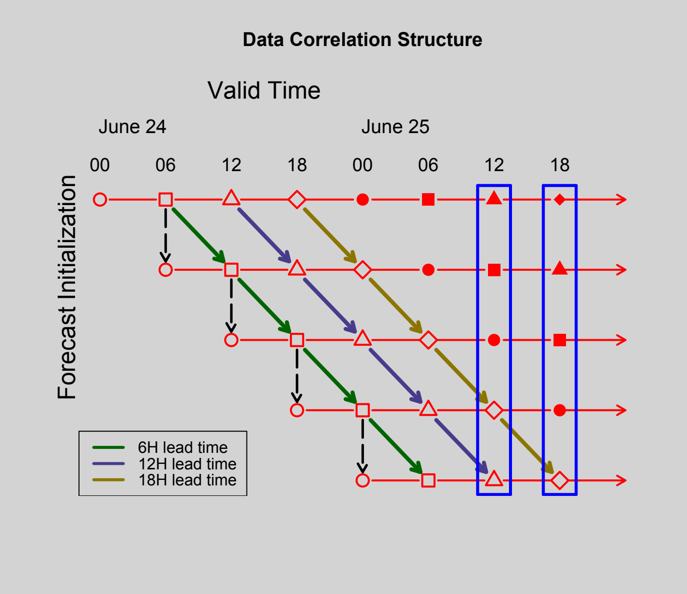
Correlation: Hurricane Sandy

Correlation: Hurricane Sandy

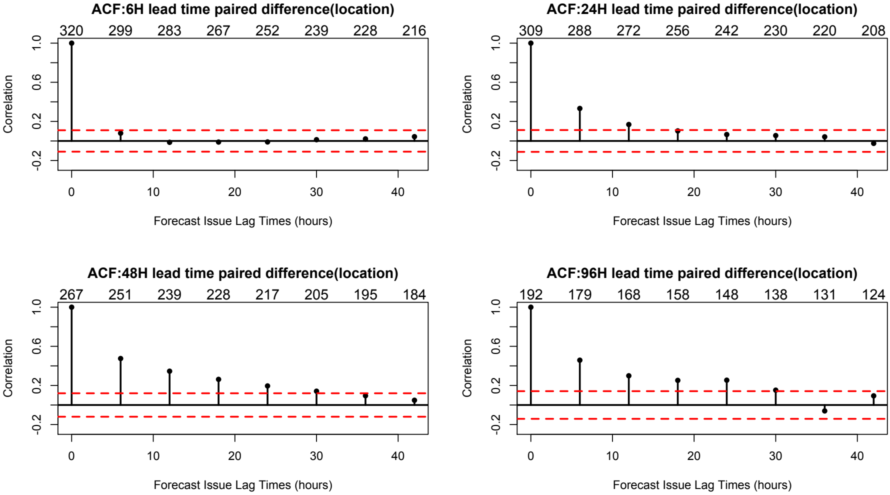
Correlation: Hurricane Sandy



Correlation



Correlation



Paired t-tests: correlation adjustment

Recall the model: $y_{ij} = \mu + \epsilon_{ij}$.

Treating the data as coming from a repeated measures experiment (subject = storm), and modeling the ϵ_{ij} as an AR(1) process gives adjusted CI's for μ.

95% CI's for μ: location error (km)

i.i.d. errors

Lead Time	Lower	Upper
24	-2.16	5.76
48	2.84	18.44
96	21.16	72.51

AR-1 errors

Lead Time	Lower	Upper
24	-4.54	8.55
48	-3.30	25.67
96	3.15	94.77

Paired t-tests: correlation adjustment

95% CI's for μ: intensity error (knots)

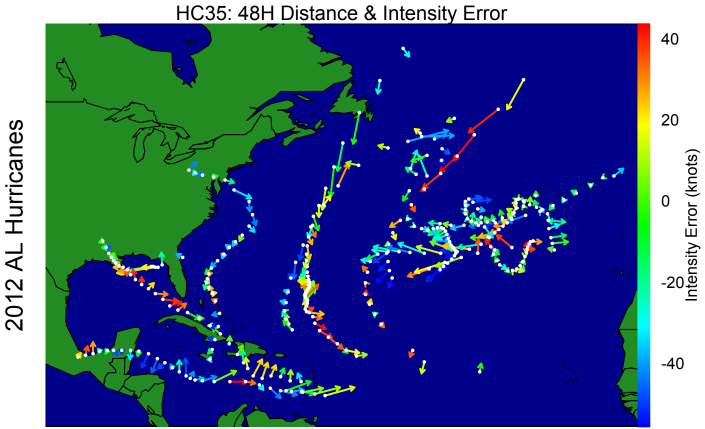
i.i.d. errors

Lead Time	Lower	Upper
24	-2.59	-0.90
48	-1.25	0.56
96	0.19	2.91

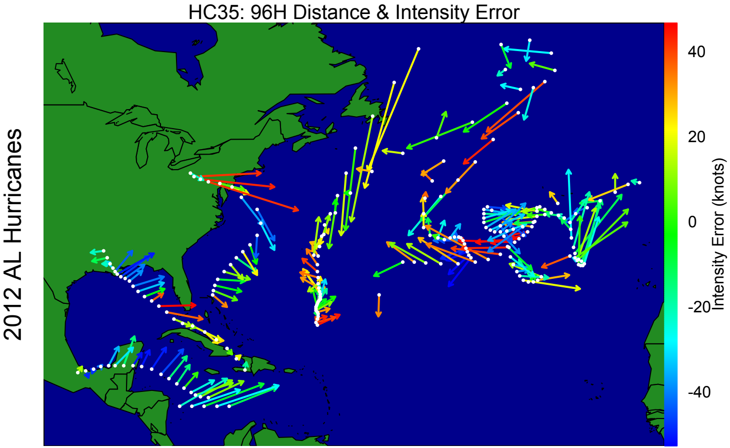
AR-1 errors

Lead Time	Lower	Upper
24	-2.88	-0.36
48	-1.70	0.92
96	-0.84	3.73

Graphical Techniques: Evaluating a Single Model



Graphical Techniques: Evaluating a Single Model



Future Work: 3-D Enviroment

1. Dynamical models attempt to emulate the 3-dimensional physical environment.
   - Sea surface temperature (SST)
   - Air temperature
   - Relative humidity (RH)
   - Wind speed and shear at various pressure gradients
2. Our data contains 100+ variables describing 3-D physical environment

Future Work: 3-D Enviroment

1. Dynamical models attempt to emulate the 3-dimensional physical environment.
   - Sea surface temperature (SST)
   - Air temperature
   - Relative humidity (RH)
   - Wind speed and shear at various pressure gradients
2. Our data contains 100+ variables describing 3-D physical environment
3. Question: are errors in any of the 3-D environment variables related to forecast errors? (in particular, intensity errors)
4. Errors in 3-D environment can be computed by using GFS (global forecast system) data, but data are not on the same scale
5. Errors in the 3-D environment are highly correlated for some variables

Future Work: 3-D Enviroment

1. Dynamical models attempt to emulate the 3-dimensional physical environment.
   - Sea surface temperature (SST)
   - Air temperature
   - Relative humidity (RH)
   - Wind speed and shear at various pressure gradients
2. Our data contains 100+ variables describing 3-D physical environment
3. Question: are errors in any of the 3-D environment variables related to forecast errors? (in particular, intensity errors)
4. Errors in 3-D environment can be computed by using GFS (global forecast system) data, but data are not on the same scale
5. Errors in the 3-D environment are highly correlated for some variables

Goal: Use statistics to improve dynamical forecasting models, which are being outperformed by statistical forecasting models!

Future Work

1. Multivariate spatio-temporal model of forecast errors
2. Compare several forecasting models at once
3. Examine other forecasting metrics (e.g. radii of max winds: a measure of storm structure)

The End

Thank you to:

1. Ligia Bernardet & Christina Holt (NOAA)
2. Mrinal Biswas (NOAA)
3. STATMOS research network
4. Dr. Peter Guttorp (University of Washington)
5. Justin Wagner (University of Washington)
6. Dr. Alexandra Schmidt (UFRJ) and all the PASI instructors

Questions? Perguntas?