Forecast Verification

Kathryn Newman

National Center for Atmospheric Research

Developmental Testbed Center

Outline

- Introduction to Forecast Verification
  - Introduction
  - Observations
  - Basic verification metrics
  - Uncertainty & Confidence intervals
- MET & MET-TC
  - MET Overview
  - Verification tools using model output
  - TC-specific tools

Introduction to Forecast Verification

Introduction

- What is Verification?
  - The process of comparing forecasts to relevant observations
  - Measures quality of forecasts
  - Evaluation of a particular model or condition

- Why Verify?
  - Help understand model biases and performance of models under certain conditions
  - Help users interpret forecasts
  - Identify forecast weakness, strengths, differences

Introduction

- Verification goals depend on the questions we want to answer
  - Determines which attribute(s) to measure
  - Drives choices in which statistics to compute, how to stratify the data, and what graphics to produce

- Before starting any verification study:
  1. **Identify multiple verification attributes** that provide answers to the questions of interest
     - ✓ *Position, wind, QPF, RI, landfall …*
  2. **Select measures and graphics** to appropriately measure and represent the attributes of interest
     - ✓ *Track (along/cross) error, Intensity error, Contingency tables …*
  3. **Identify a standard of comparison** that provides a reference level of skill
     - ✓ *CLIPER, SHIFOR, Baseline model …*

Observations

- Observations are an important consideration for TC verification
  - Quality and quantity of observations available
    - Typically sparse or intermittent
    - May infer characteristics from indirect measures (satellite)

Variable	Suggested observations	Suggested analyses
Position of storm center	Reconnaissance flights, visible & IR satellite imagery, passive microwave imagery	Best track, IBTrACS
Intensity – maximum sustained wind	Dropwinsonde, microwave radiometer	Best track, IBTrACS, Dvorak analysis
Intensity – central pressure	Ship, buoy, synop, AWS	IBTrACS, Dvorak analysis
Storm structure	Reconnaissance flights, Doppler radar, visible & IR satellite imagery, passive microwave	H*Wind, MTCSWA, ARCHER
Storm life cycle		NWP model analysis
Precipitation	Rain gauge, radar, passive microwave, spaceborne radar	Blended gauge-radar, blended satellite
Wind speed over land	Synop, AWS, Doppler radar	
Wind speed over sea	Buoy, ship reports, dropwinsondes, scatterometer, passive microwave imagers and sounders	H*Wind, MTCSWA
Storm surge	Tide gauge, GPS buoy	
Waves – significant wave height	Buoy, ship reports, altimeter	Blended analyses
Waves – spectra	Altimeter	

Suggested observations and analyses for verifying forecasts of TC variables and associated hazards. *(WMO report on TC verification)*

https://www.wmo.int/pages/prog/arep/wwrp/new/documents/TC_verification_Final_11Nov13.pdf

Observations

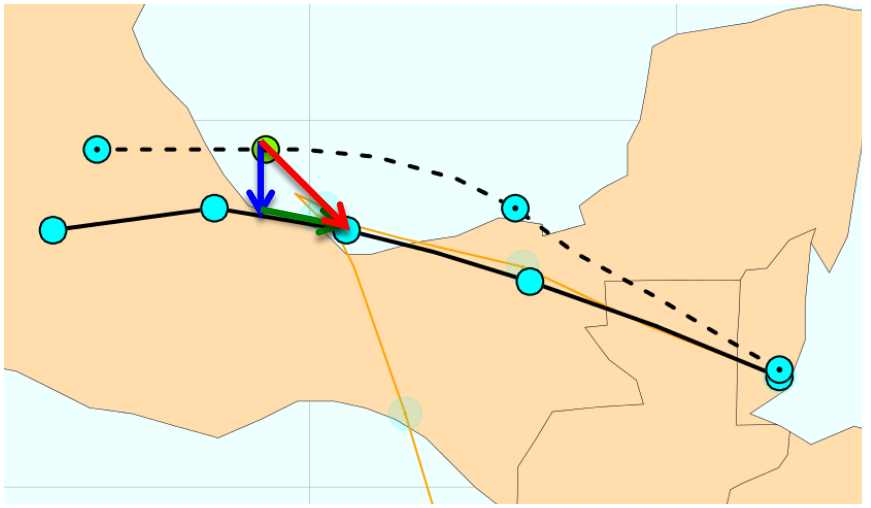
- Best track analysis
  - Subjective assessment of TC's center location and intensity (6 hr) using all observations available
  - Includes center position, maximum sfc winds, minimum center pressure, quadrant radii of 34/50/64 kt winds
  - Subjectively smoothed

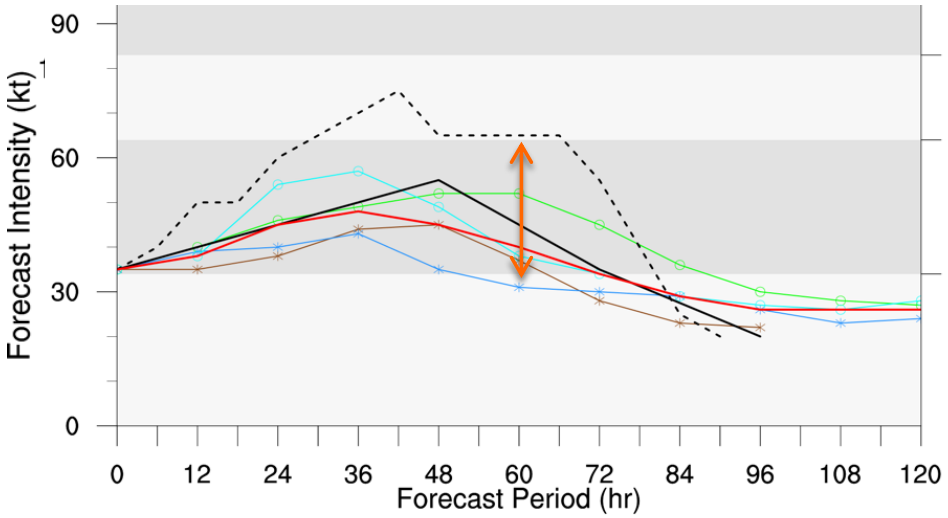
```
AL, 09, 2011082200,   , BEST,   0, 179N,  650W,  60,  993, TS,  34, NEQ,  130,   30,   30,   90, 1010,  125,  30,  60,   0,  L,   0,   ,   0,   0,     IRENE, M, 12, NEQ,  360,   60,   15,  120
AL, 09, 2011082200,   , BEST,   0, 179N,  650W,  60,  993, TS,  50, NEQ,   30,    0,    0,   30, 1010,  125,  30,  60,   0,  L,   0,   ,   0,   0,     IRENE, M, 12, NEQ,  360,   60,   15,  120
AL, 09, 2011082206,   , BEST,   0, 182N,  659W,  65,  990, HU,  34, NEQ,  130,   60,   60,   90, 1010,  150,  15,  75,   0,  L,   0,   ,   0,   0,     IRENE, D, 12, NEQ,  360,   60,   15,  120
AL, 09, 2011082206,   , BEST,   0, 182N,  659W,  65,  990, HU,  50, NEQ,   40,   25,   20,   35, 1010,  150,  15,  75,   0,  L,   0,   ,   0,   0,     IRENE, D, 12, NEQ,  360,   60,   15,  120
AL, 09, 2011082206,   , BEST,   0, 182N,  659W,  65,  990, HU,  64, NEQ,   25,    0,    0,    0, 1010,  150,  15,  75,   0,  L,   0,   ,   0,   0,     IRENE, D, 12, NEQ,  360,   60,   15,  120
AL, 09, 2011082212,   , BEST,   0, 189N,  670W,  70,  989, HU,  34, NEQ,  160,   60,   60,   90, 1010,  200,  15,  85,   0,  L,   0,   ,   0,   0,     IRENE, D, 12, NEQ,  180,   60,    0,  120
AL, 09, 2011082212,   , BEST,   0, 189N,  670W,  70,  989, HU,  50, NEQ,   40,   25,   20,   35, 1010,  200,  15,  85,   0,  L,   0,   ,   0,   0,     IRENE, D, 12, NEQ,  180,   60,    0,  120
AL, 09, 2011082212,   , BEST,   0, 189N,  670W,  70,  989, HU,  64, NEQ,   25,    0,    0,    0, 1010,  200,  15,  85,   0,  L,   0,   ,   0,   0,     IRENE, D, 12, NEQ,  180,   60,    0,  120
AL, 09, 2011082218,   , BEST,   0, 193N,  680W,  75,  988, HU,  34, NEQ,  160,   60,   40,   90, 1010,  250,  15,  85,   0,  L,   0,   ,   0,   0,     IRENE, D, 12, NEQ,  180,   60,    0,  120
AL, 09, 2011082218,   , BEST,   0, 193N,  680W,  75,  988, HU,  50, NEQ,   40,   30,   20,   35, 1010,  250,  15,  85,   0,  L,   0,   ,   0,   0,     IRENE, D, 12, NEQ,  180,   60,    0,  120
AL, 09, 2011082218,   , BEST,   0, 193N,  680W,  75,  988, HU,  64, NEQ,   25,    0,    0,    0, 1010,  250,  15,  85,   0,  L,   0,   ,   0,   0,     IRENE, D, 12, NEQ,  180,   60,    0,  120
AL, 09, 2011082300,   , BEST,   0, 197N,  688W,  80,  981, HU,  34, NEQ,  160,   70,   50,  100, 1010,  300,  15, 105,  30,  L,   0,   ,   0,   0,     IRENE, D, 12, NEQ,  180,   60,    0,  120
AL, 09, 2011082300,   , BEST,   0, 197N,  688W,  80,  981, HU,  50, NEQ,   70,   30,   30,   70, 1010,  300,  15, 105,  30,  L,   0,   ,   0,   0,     IRENE, D, 12, NEQ,  180,   60,    0,  120
AL, 09, 2011082300,   , BEST,   0, 197N,  688W,  80,  981, HU,  64, NEQ,   25,    0,    0,   35, 1010,  300,  15, 105,  30,  L,   0,   ,   0,   0,     IRENE, D, 12, NEQ,  180,   60,    0,  120
AL, 09, 2011082306,   , BEST,   0, 201N,  697W,  80,  978, HU,  34, NEQ,  180,  120,   90,  130, 1010,  300,  15, 105,   0,  L,   0,   ,   0,   0,     IRENE, D, 12, NEQ,  180,  120,   35,  150
AL, 09, 2011082306,   , BEST,   0, 201N,  697W,  80,  978, HU,  50, NEQ,   90,   60,   40,   70, 1010,  300,  15, 105,   0,  L,   0,   ,   0,   0,     IRENE, D, 12, NEQ,  180,  120,   35,  150
AL, 09, 2011082306,   , BEST,   0, 201N,  697W,  80,  978, HU,  64, NEQ,   45,   30,   20,   35, 1010,  300,  15, 105,   0,  L,   0,   ,   0,   0,     IRENE, D, 12, NEQ,  180,  120,   35,  150
AL, 09, 2011082312,   , BEST,   0, 204N,  706W,  80,  978, HU,  34, NEQ,  180,  120,   90,  130, 1008,  300,  15, 105,   0,  L,   0,   ,   0,   0,     IRENE, D, 12, NEQ,  180,  120,   35,  150
AL, 09, 2011082312,   , BEST,   0, 204N,  706W,  80,  978, HU,  50, NEQ,   90,   60,   40,   70, 1008,  300,  15, 105,   0,  L,   0,   ,   0,   0,     IRENE, D, 12, NEQ,  180,  120,   35,  150
```

Follows ATCF format – more on this later!

TC Metrics

- **Track Error**: great-circle distance between the forecast location and the actual location of the storm center (nmi)
- **Along-track Error**: indicator of whether a forecasting system is moving a storm too slowly/quickly
- **Cross-track Error**: indicates displacement to the right/left of the observed track
- **Intensity Error**: Difference between forecast and actual intensity (kts)
  - Raw intensity errors (bias) vs. absolute intensity errors (magnitude of error)



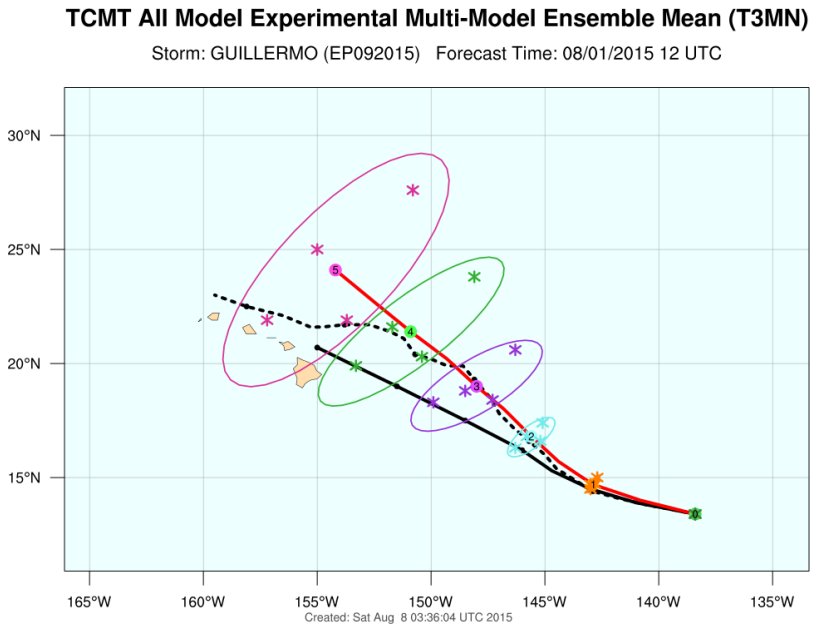


Graphics courtesy of NCAR TCMT

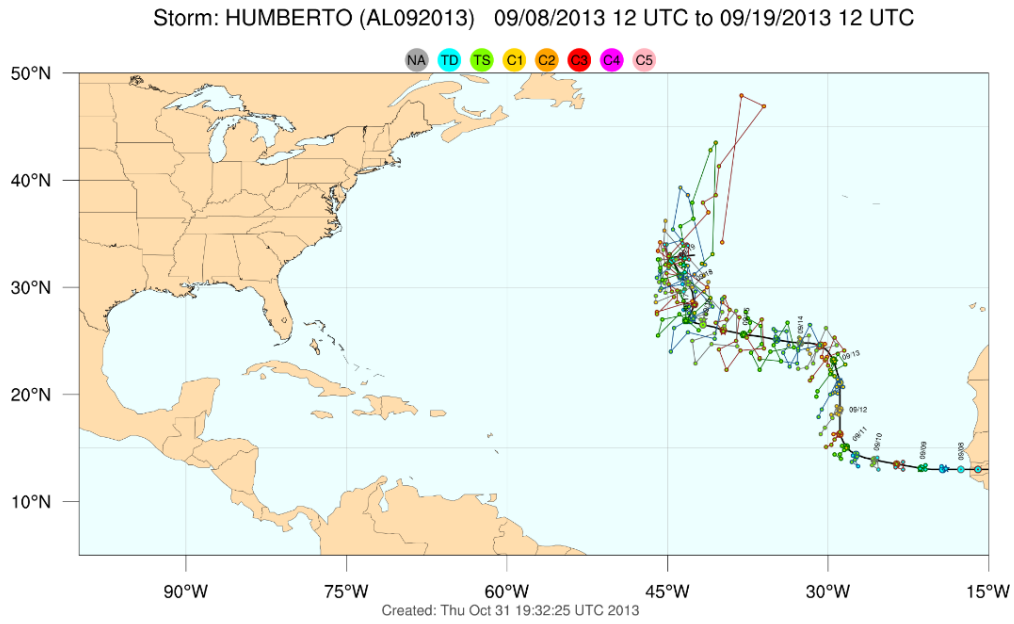
TC Metrics

- Storm structure, precipitation, wind speed, storm surge, waves, probabilistic forecasts and ensembles…
  - Going beyond basic track and intensity error
- New approaches for TC verification evolving

Probability ellipses derived from ensembles



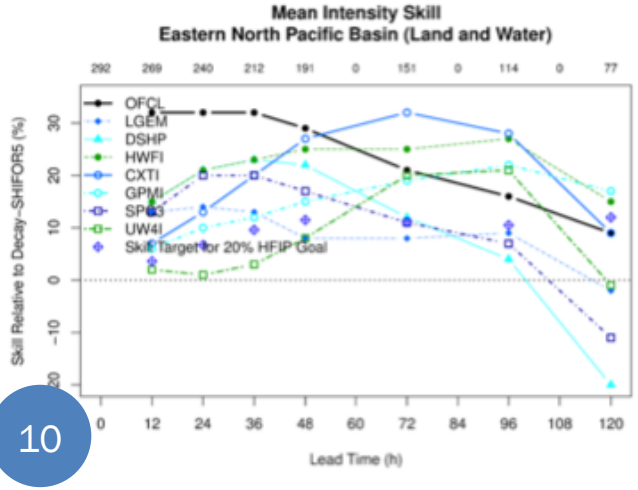
Evaluation of forecast consistency

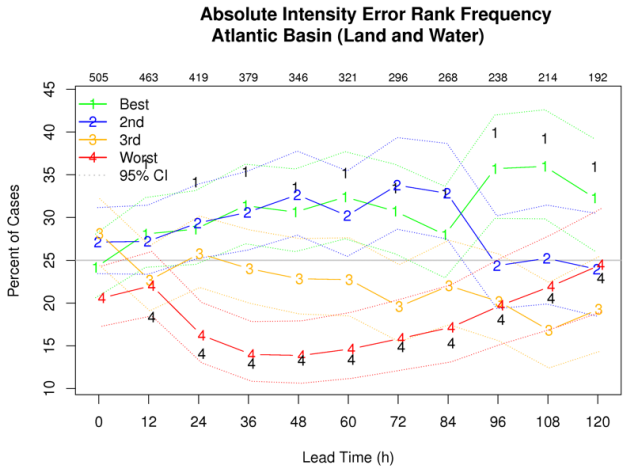


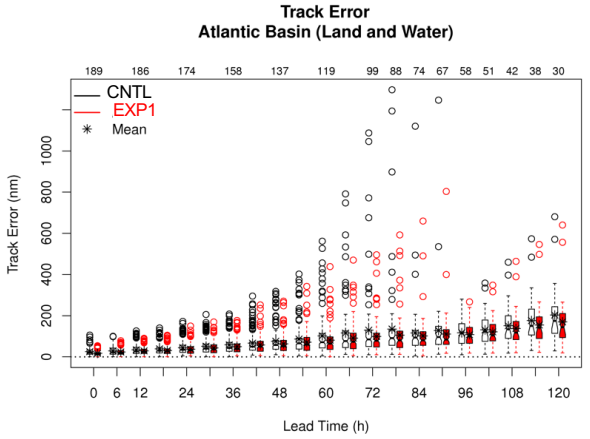
Images from NCAR. Methodology following Hamill et al 2011 (left), Fowler et al 2010 (right)

TC metrics

- Skill Scores: Used as a standard of comparison, skill diagrams are often used to compare model skill relative to CLIPER/SHIFOR
- Frequency of Superior Performance & Rank frequency: ranking a particular model forecast relative to the performance multiple model forecasts
- Distribution of errors: Box plots can be used to highlight the distributions of the errors in the forecasts



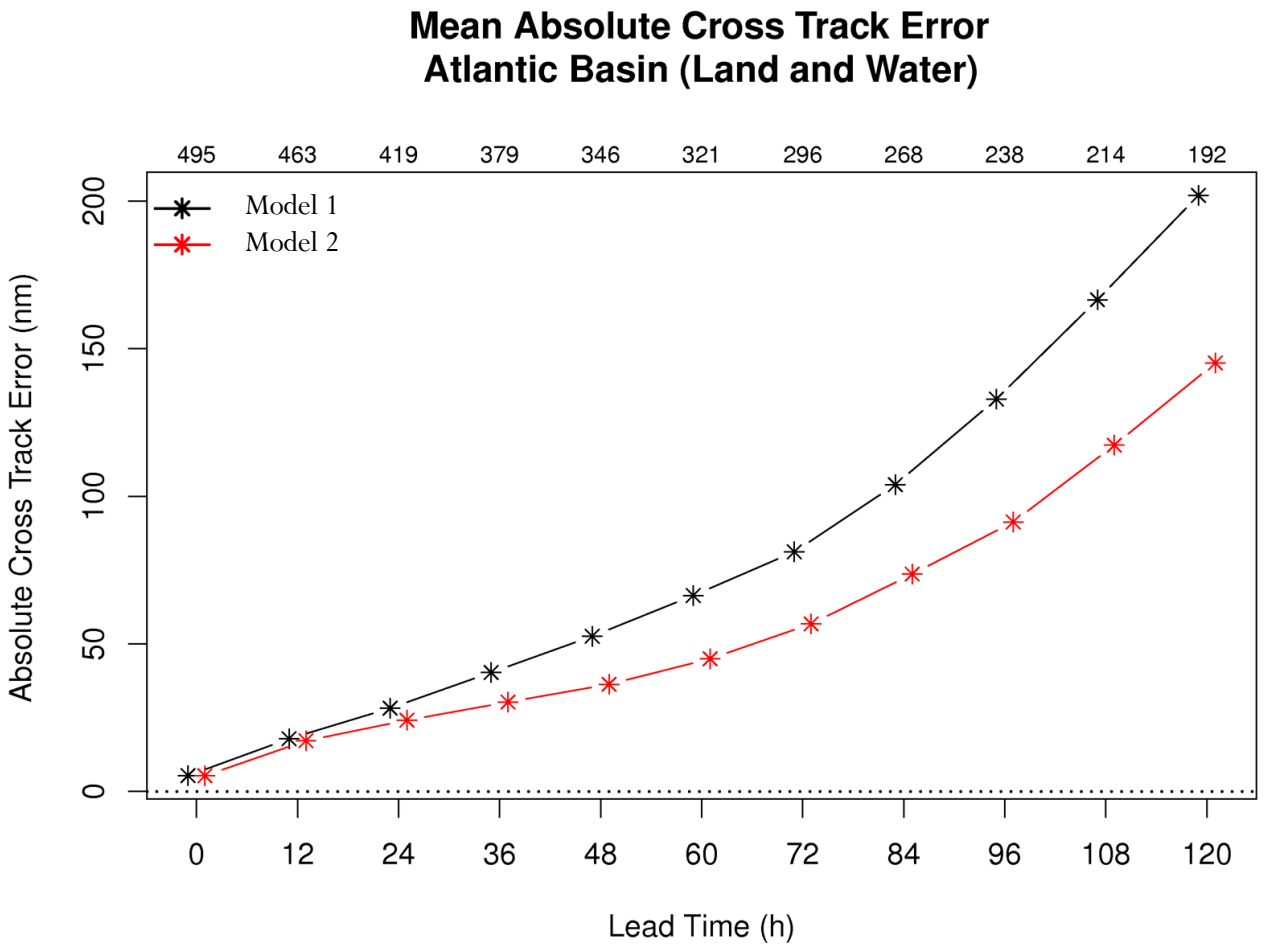




Uncertainty

- Observations and analysis products as well as models themselves are subject to uncertainty
- Need to be aware of sample size!
  - TCs typically have smaller samples due to lower frequency of occurrence relative to other weather phenomena
- Accounting for sampling uncertainty:
  - Verification statistic is a realization of a random process
  - What if the experiment were re-run under identical conditions? Would you get the same answer?

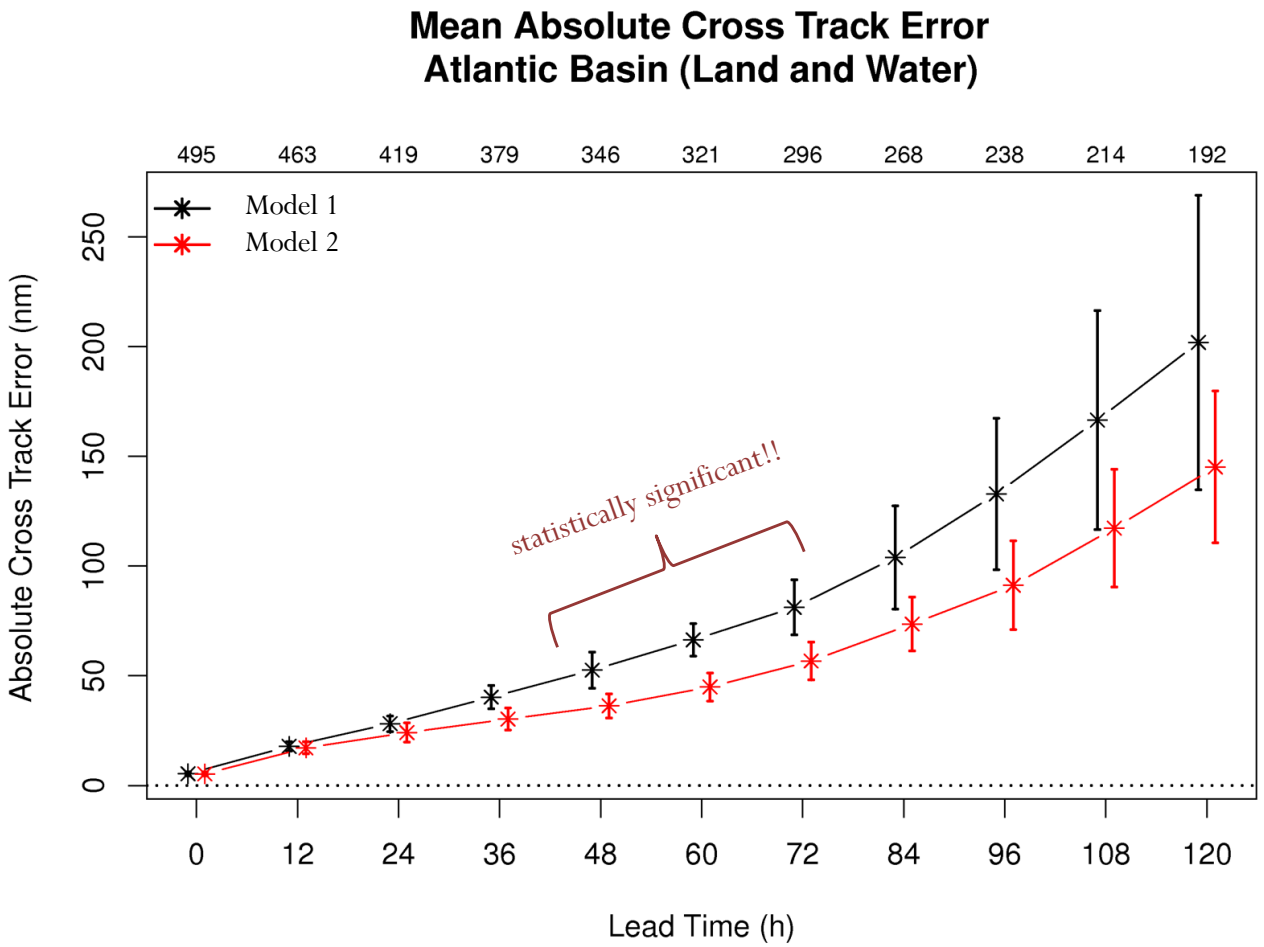
Confidence intervals



Mean absolute cross-track errors for two models.

Scores are very similar at short lead times, but seem to diverge at longer lead times

Confidence intervals



Confidence Intervals (CIs) indicate no significant difference between 0-36 h, after 84 h

Statistical significance indicated where CIs don't overlap

Multiple methods for computing CIs:

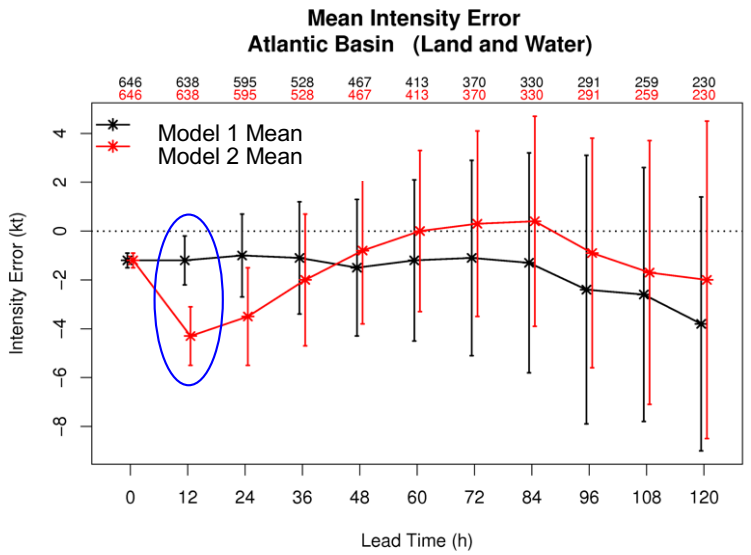
- Standard error about the mean or median
- Bootstrapping

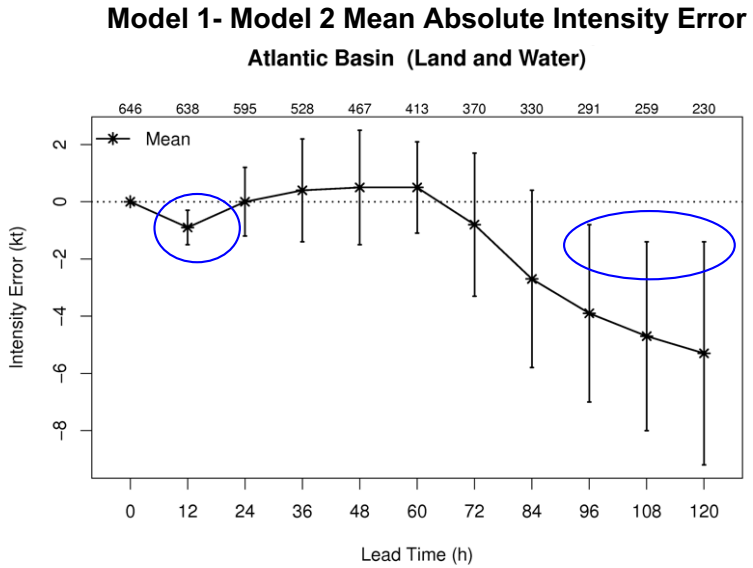
Choice of alpha value for CIs

- e.g. 95%

Confidence intervals

- Two ways to examine scores:
  - CI about absolute scores
    - May be difficult to differentiate model performance differences
    - SS where two model CIs do not intersect
  - CI about Pairwise Differences
    - May allow for differentiation of model performance.
    - SS where CIs do not encompass 0
    - Stronger test – removes common forecast challenges



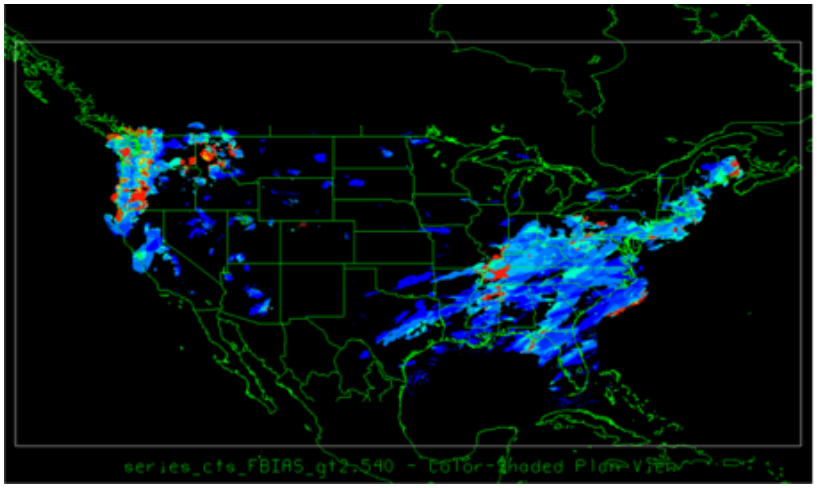


Model Evaluation Tools

MET & MET- Tropical Cycle

Model Evaluation Tools

- What is MET?
  - MET is a set of tools for evaluating model forecasts
- A modular set of forecast evaluation tools
  - Freely available, highly configurable, fully documented, supported
- MET includes:
  - Reformatting tools
  - Statistical tools
  - Analysis tools
- MET works directly with post-processed model output to perform a large variety of statistical analyses



Precipitation frequency bias generated from MET output

Model Evaluation Tools

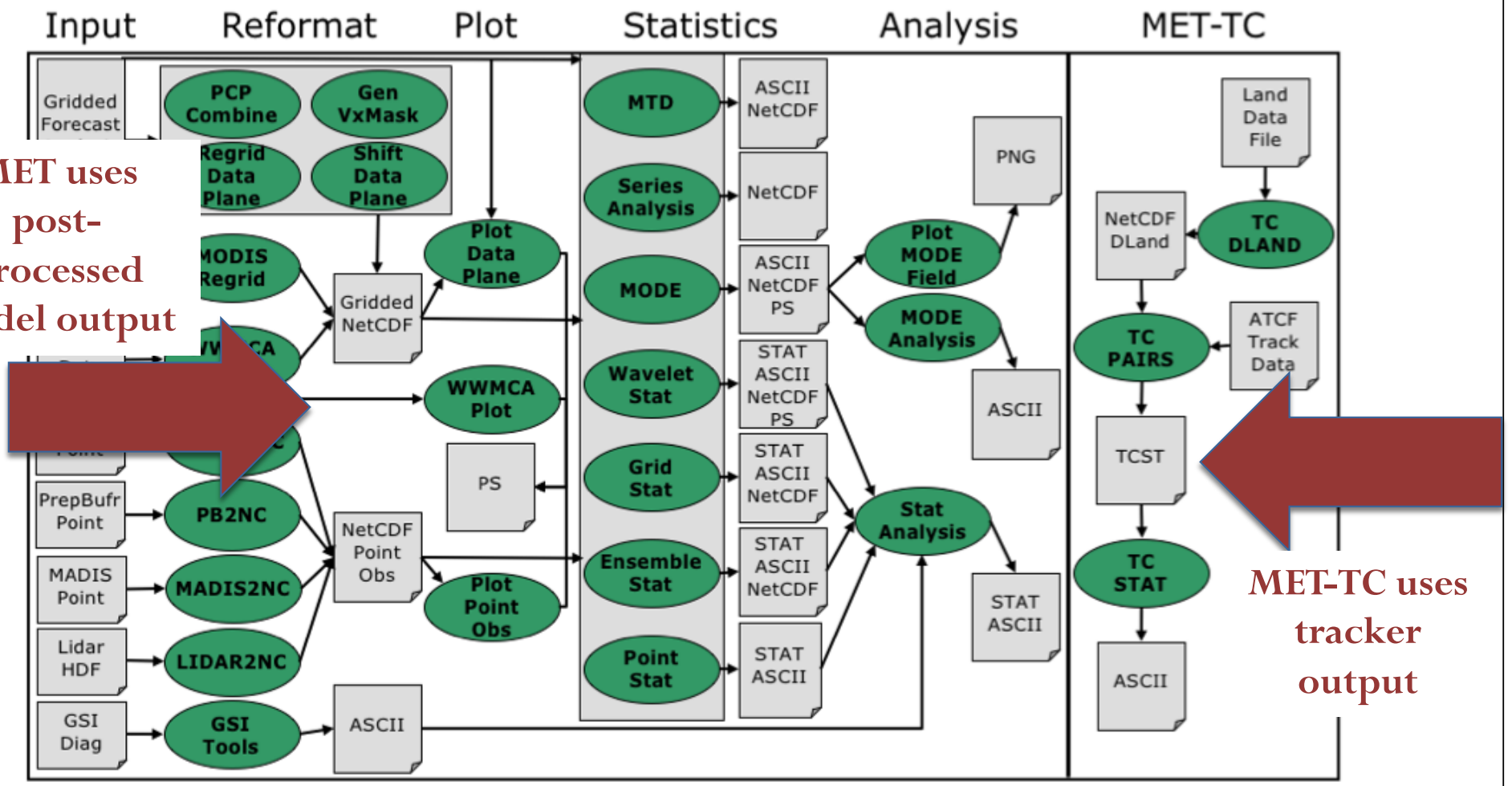
- Overview of tools
- MET provides a variety of verification techniques:
  - Gridded model data to point-based observations
  - Gridded model data to gridded observations
  - Ensemble and probabilistic verification methods
  - Aggregating output through time and space
  - Object-based verification
  - Tropical cyclone verification
    - Tropical cyclone evaluation through MET-TC

MET-Tropical Cyclone

- WHAT is MET-TC?
  - A set of tools to aid in TC forecast evaluation and verification
  - Developed to replicate (and add to) the functionality of the NHC verification software
  - Modular set of tools which utilize the MET software framework
    - Allows for additional capabilities and features to be added to future releases
- WHY use MET-TC?
  - Provides a standard set of verification metrics and comprehensive output statistics
  - Available to all users
    - Enables consistent forecast evaluation studies to be undertaken across the community

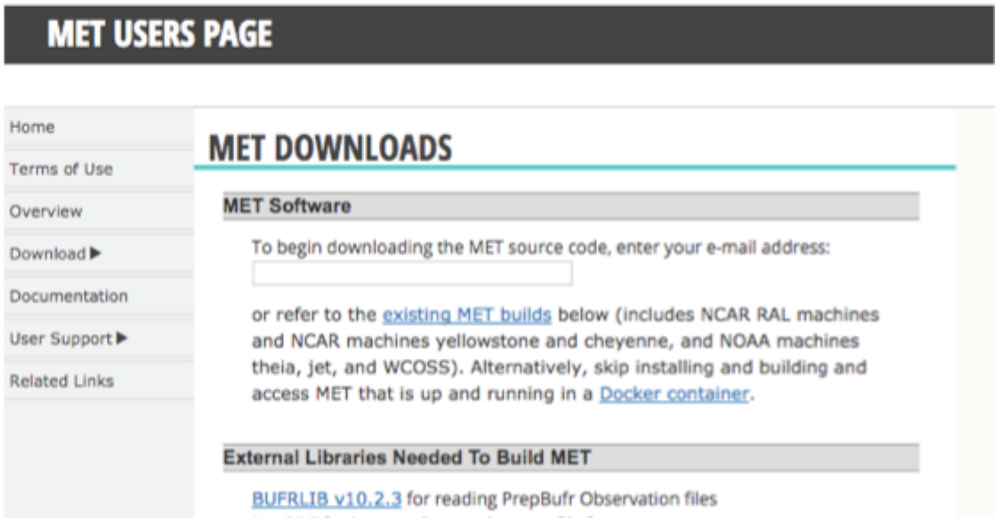
Model Evaluation Tools

MET Overview v6.1



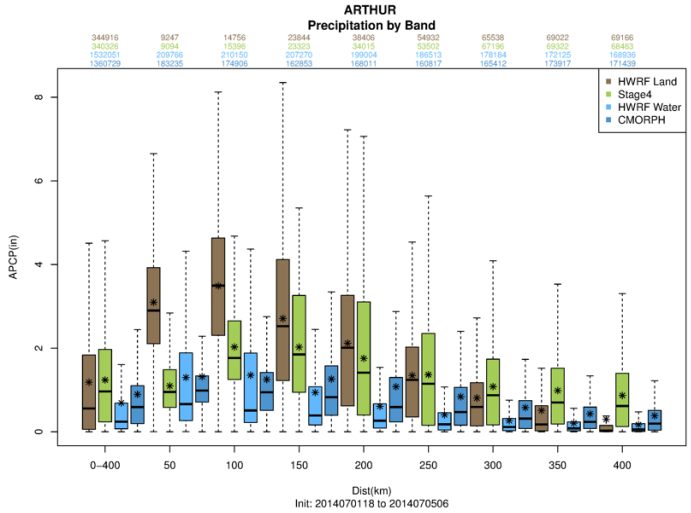
Compile & Build

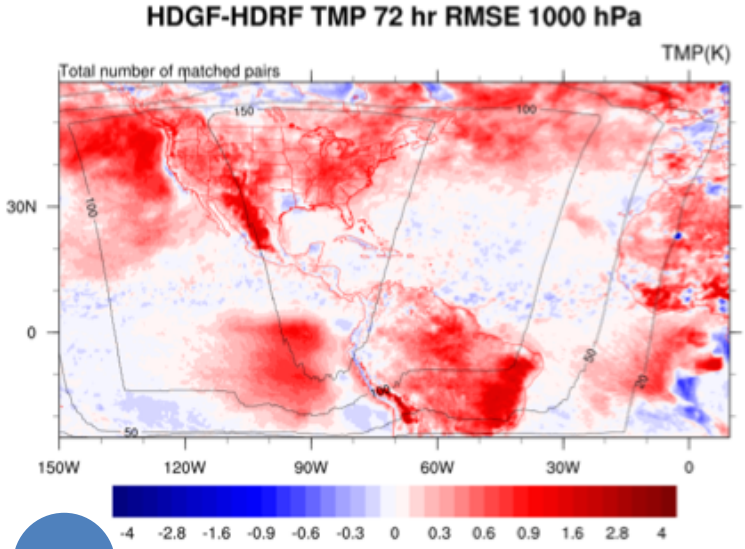
- Download MET (must be v4.1 or newer for MET-TC capabilities) release and compile locally
  - Need to register to download: www.dtcenter.org/met/users
- Supported platforms and compilers
  1. Linux with GNU compilers
  2. Linux with Portland Group (PGI) compilers
  3. Linux with Intel compilers

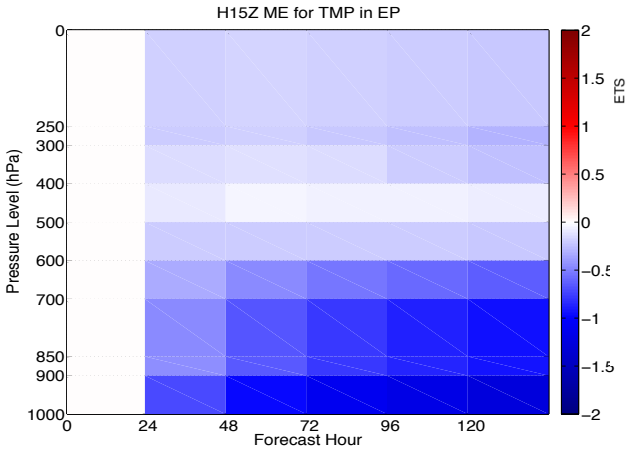


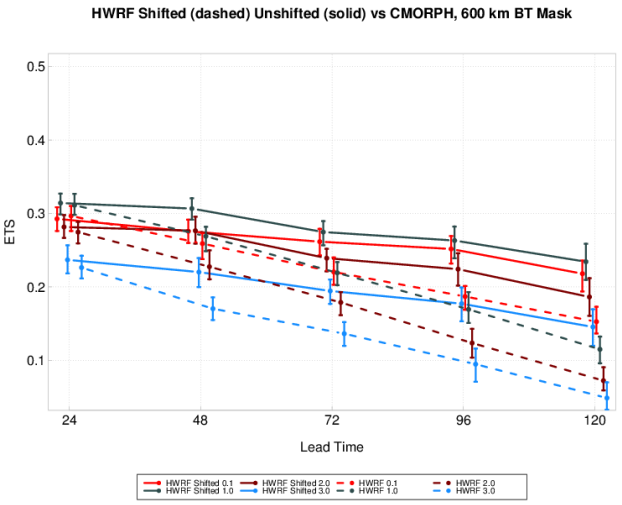
HWRF verification using MET

- MET verification tools using HWRF model output
  - Large scale: verified against GFS, other configurations
    - TMP, SPFH, HGT …
  - Storm scale QPF verification

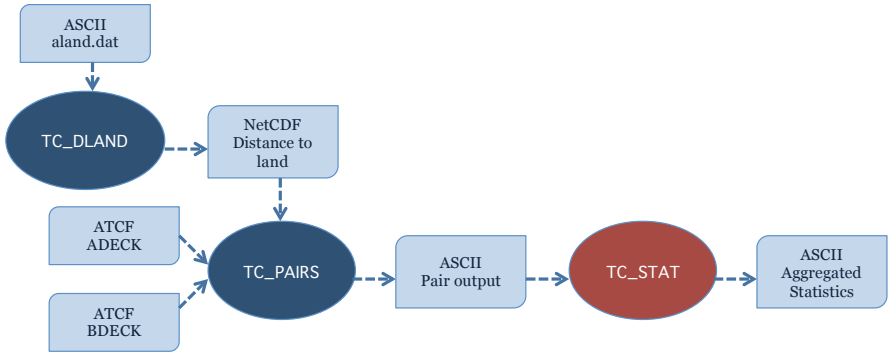








MET-TC Tools



- TC-dland
  - Pre-computes distance to land file for use to TC-pairs
    - ✓ More efficient than computing distances on the fly
- TC-pairs
  - Reads ATCF files to produce pair statistics (with reference TC dataset) on independent model input or user-specified consensus forecasts
    - ✓ Pair generation can be subset based on user-defined filtering criteria
    - ✓ Includes computation of consensus forecasts and baseline models
- TC-stat
  - Provides summary statistics and filtering jobs on TC-pairs output
    - ✓ Stratifies pair output by various conditions and thresholds
    - ✓ Produces summary statistics on specific column(s) of interest
    - ✓ Includes RIRW job type for rapid intensification studies
    - ✓ Include PROBRIRW job type to apply probabilistic methods edeck files (RI)

MET-TC: Getting Started…

- Model output must be run through an internal/external vortex tracking algorithm *(GFDL vortex tracker – previous lecture)*
- The input files must be in Automated Tropical Cyclone Forecasting System (ATCF) format.
  - Must adhere to for MET-TC tools to properly parse the input data (first 8 columns required)

For detailed information on ATCF format:
http://www.nrlmry.navy.mil/atcf_web/docs/database/new/abdeck.txt

- The best track analysis is used primarily used as the observational dataset in MET-TC.

All operational model aids and best track analysis can be found on the NHC ftp server: ftp://ftp.nhc.noaa.gov/atcf/archive/

MET-TC: easy filtering criteria

MODEL	WATCH/WARNING STATUS
STORM ID	THRESHOLD: Any value: initial time, valid time
BASIN	WATER ONLY
CYCLONE	RAPID INTENSITY
STORM NAME	LANDFALL
INITALIZATION TIME: Include, exclude, beginning, end	EVENT EQALIZATION
INITALIZATION/VALID HR	CONSENSUS FORECAST
VALID TIME: Include, exclude, beginning, end	LAG FORECAST
LEAD TIME	INTERPOLATED FORECASTS
MASKING	

MET-TC: TC_stat

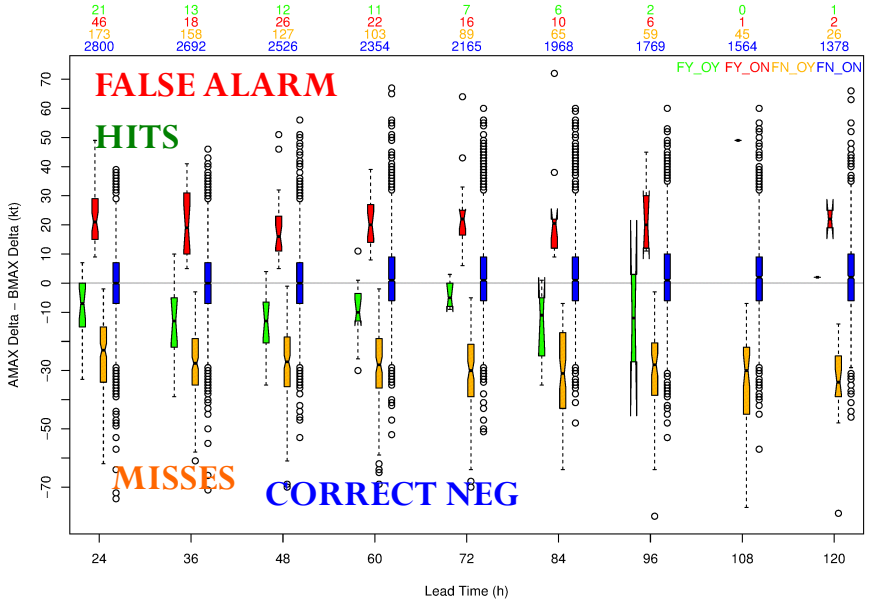
- The **filter** job applies a flexible set of filtering criteria to subset track data
- The **summary** job computes summary statistics for one or more columns of data
- The **rirw** job identifies rapid intensification or weakening in the forecast and analysis track and applies flexible criteria to derive event contingency tables and statistics
- The **probrirw** job derives probabilistic statistics summarizing performance of probabilistic forecasts for a specified intensity change (contained in edeck)

MET-TC: HWRF RIRW Verification

- MET-TC includes Rapid Intensity Change verification capabilities
  - Default 30kt change over 24hr. Also includes relaxation capabilities for further diagnosis of model behavior
  - Contingency table statistics, distributions corresponding to the 4 quadrants of the contingency table

		Observation		
		RI	***No RI***	***Total***
Model Forecast	***RI***	128 (0.3%)	253 (0.6%)	381 (0.9%)
	No RI	1623 (4.1%)	37654 (94.9%)	39277 (99%)
	Total	1751 (4.4%)	37907 (95.6%)	39658 (100%)

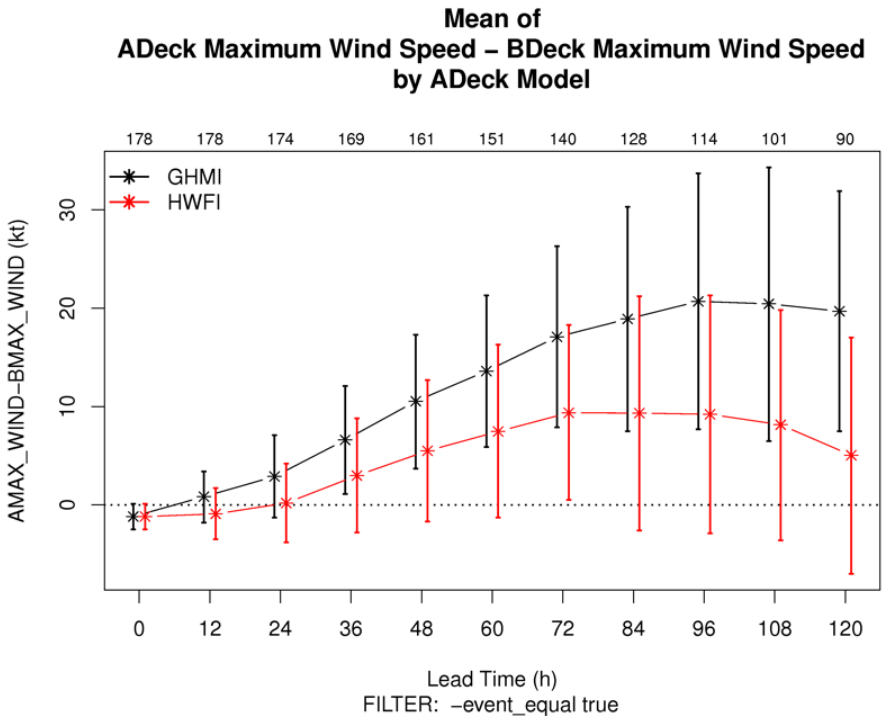
POD	7.3%
PODN	99.3%
FAR	66.4%
RI Event Rate	4.4%

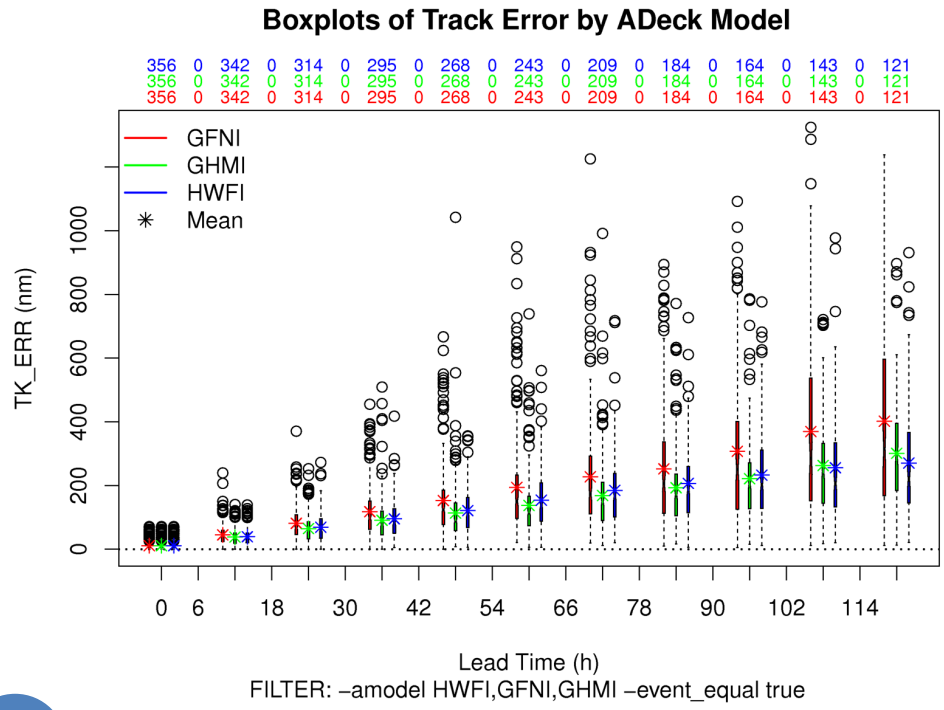


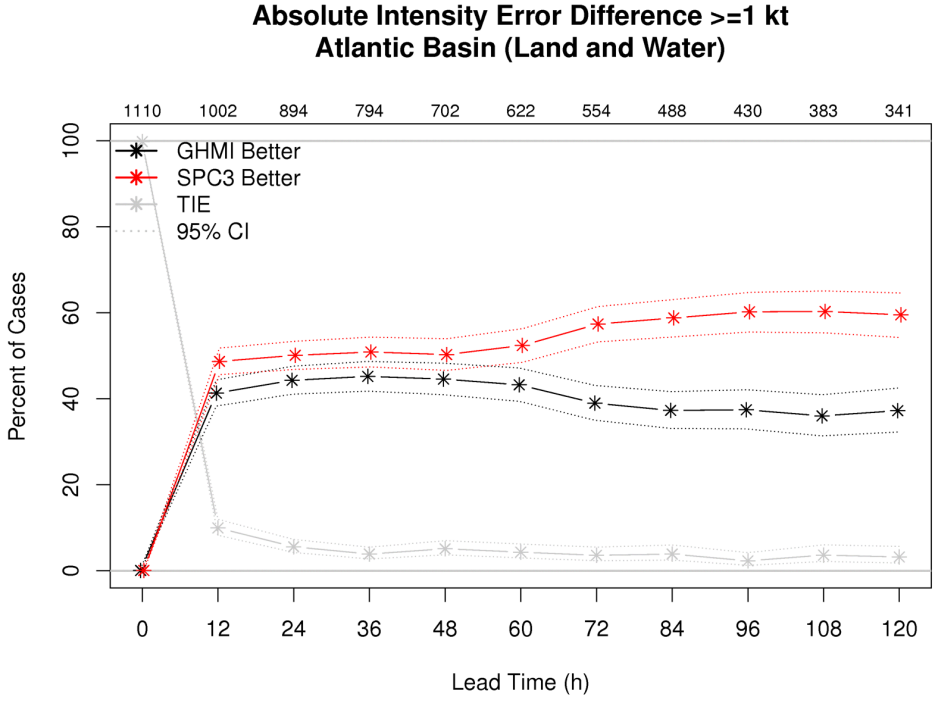
Graphics tools

Graphical capabilities are included in the MET-TC release

- `R statistical language`







References & Further Reading

- Verification methods for tropical cyclone forecasts: https://www.wmo.int/pages/prog/arep/wwrp/new/documents/TC_verification_Final_11Nov13.pdf
- Gilleland, E., 2010: Confidence intervals for forecast verification. NCAR Technical Note NCAR/TN-479+STR, 71pp. *Available at:*

  http://nldr.library.ucar.edu/collections/technotes/asset-000-000-000-846.pdf
- Jolliffe and Stephenson (2011): Forecast verification: A practitioner's guide, 2nd Edition, Wiley & sons
- JWGFVR (2009): Recommendation on verification of precipitation forecasts. WMO/TD report, no.1485 WWRP 2009-1
- Nurmi (2003): Recommendations on the verification of local weather forecasts. ECMWF Technical Memorandum, no. 430
- Wilks (2006): Statistical methods in the atmospheric sciences, ch. 7. Academic Press
- NHC forecast verification: http://www.nhc.noaa.gov/verification/index.shtml
- WWRP/WGNE Joint Working Group on Forecast Verification Research: http://www.cawcr.gov.au/projects/verification/

Appendix C of MET Documentation: http://www.dtcenter.org/met/users/docs/overview.php

- For MET code download and user's guide:

  www.dtcenter.org/met/users

- Contact for MET questions, help, comments:

  met_help@ucar.edu

- HWRF questions?

  hwrf-help@ucar.edu