ATCF: LESSONS LEARNED ON TC CONSENSUS FORECASTING

John Knaff, NOAA/NESDIS and
Buck Sampson, NRL, Monterey

Definitions

- TC – tropical cyclone
- Aid – individual model forecast
- Guidance – any method that provides a forecast
- Early guidance – guidance available at the forecast time
- Late guidance – guidance that is available after the forecast has been issued
- Interpolation – a method that blends initial conditions and late guidance to form early guidance (add the deltas, applied at all times, fazed out)
- Track- positions (lat, lon) associated with a TC
- Intensity – the maximum (1-minute) wind speed associated with a TC

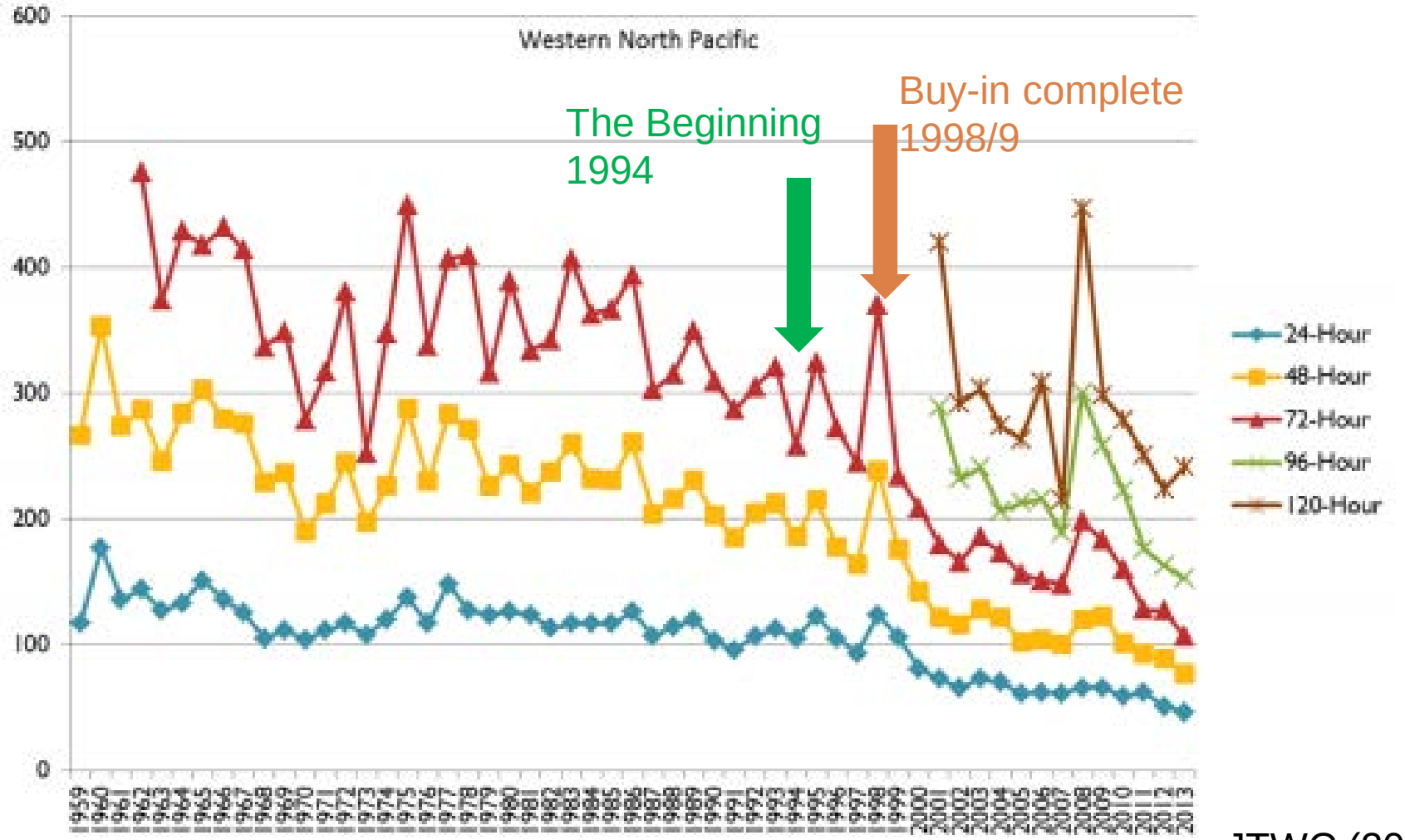
History

- Consensus forecasting in economics in the late 1960's
- Applied to Meteorology in the early 1970's
- Applied to track forecasting in the 1990's
- Applied to intensity forecasting in 2000's

Common Methods

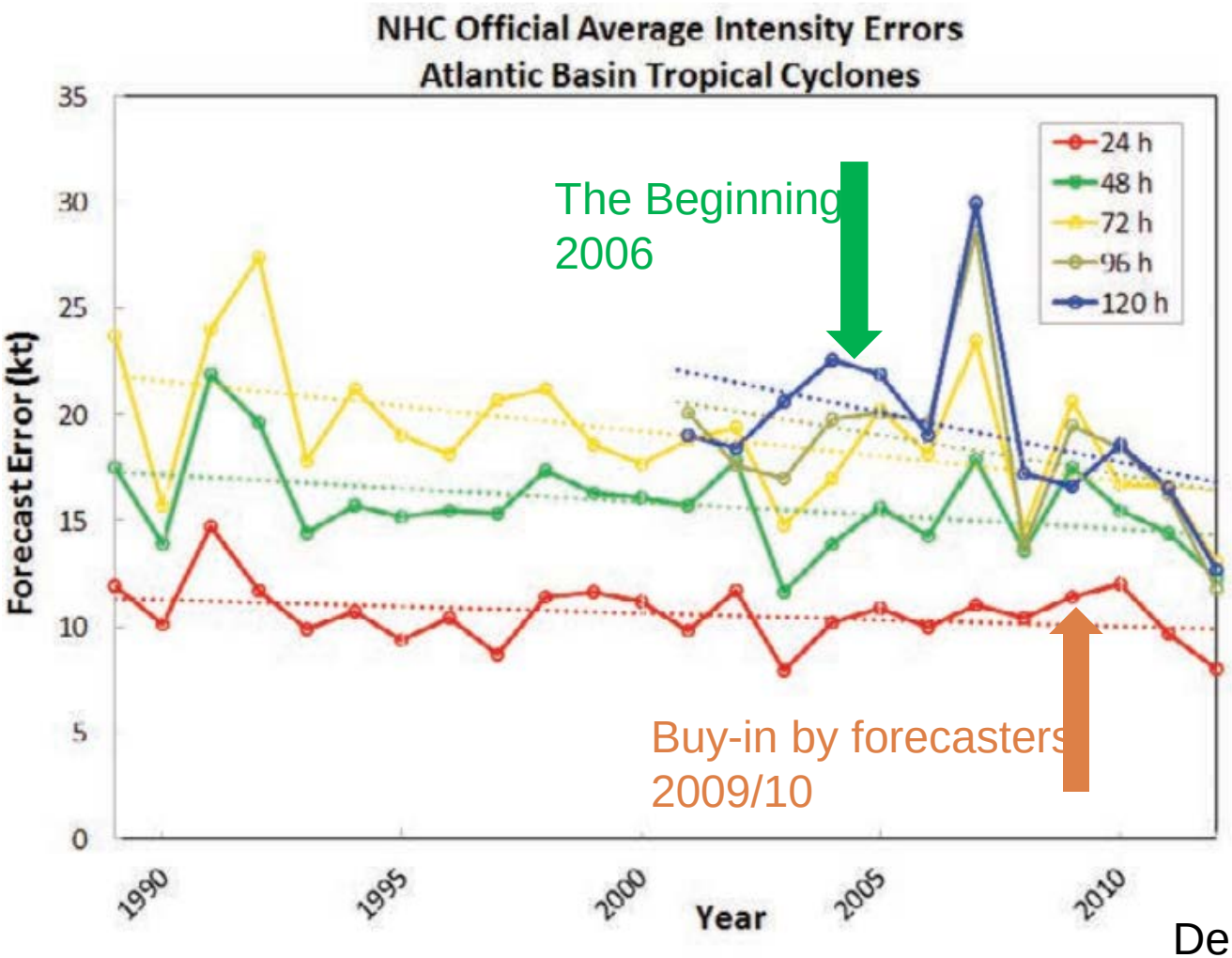
- Equally weight aids
- Weighted aids (typically based on RMS errors of recent performance)
- Regress on the aids (can provide negative weights)
- User (human) selected

Track Improvements



JTWC (2014)

Intensity Improvements

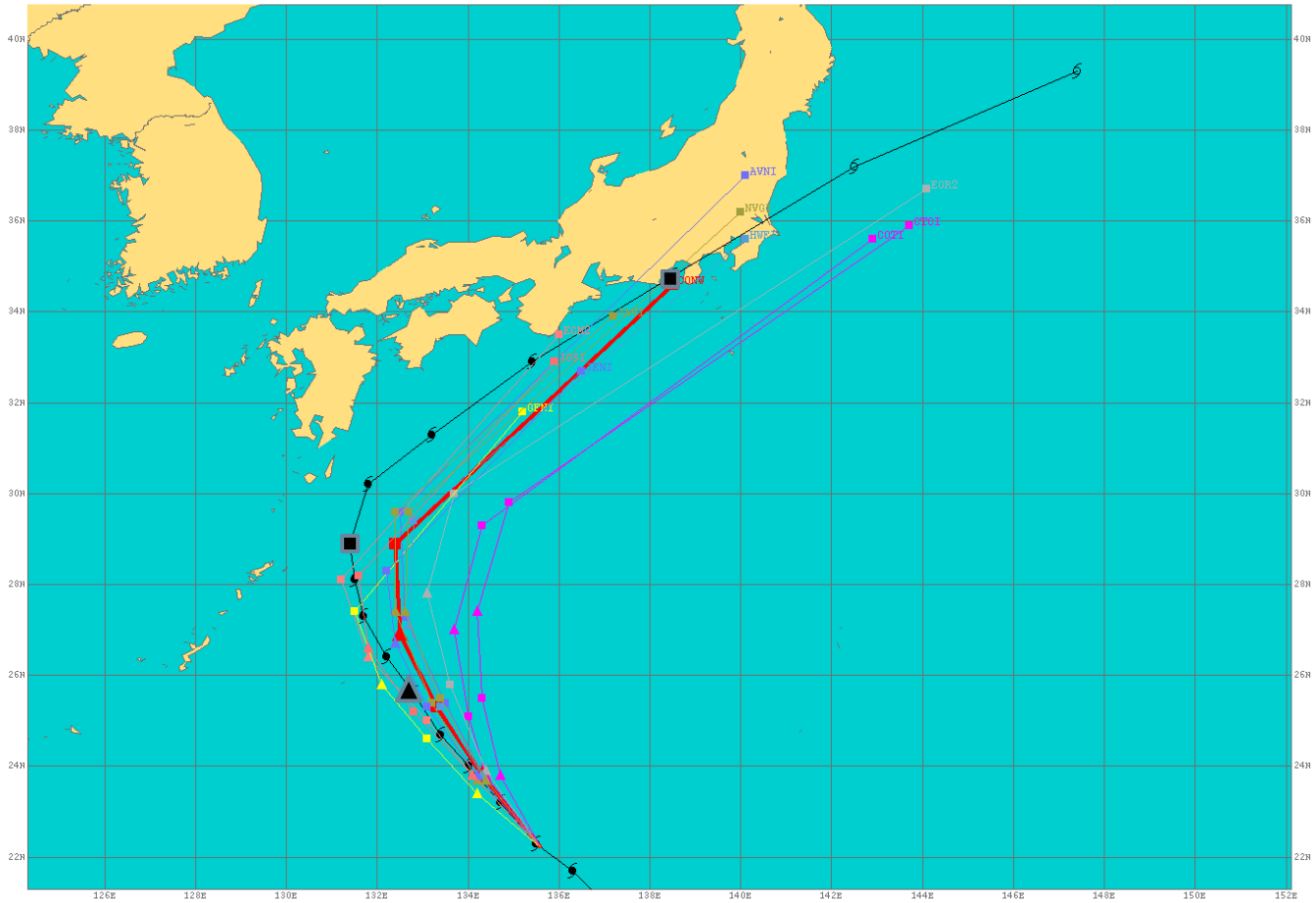


DeMaria et al. (2014)

Recent Track Consensus Results

- Track forecasts are considered mature, statistical methods are no longer used, and consensus forecasting has been successful in reducing errors since the 1990s (i.e., when NWP track models started to be the most skillful)
- Analysis Caveats
  - No bias correction performed
  - Errors for individual aids are non-homogeneous
  - Consensus forecasts is made up from available aids (variable consensus)

Example: wp182014 (Phanfone)

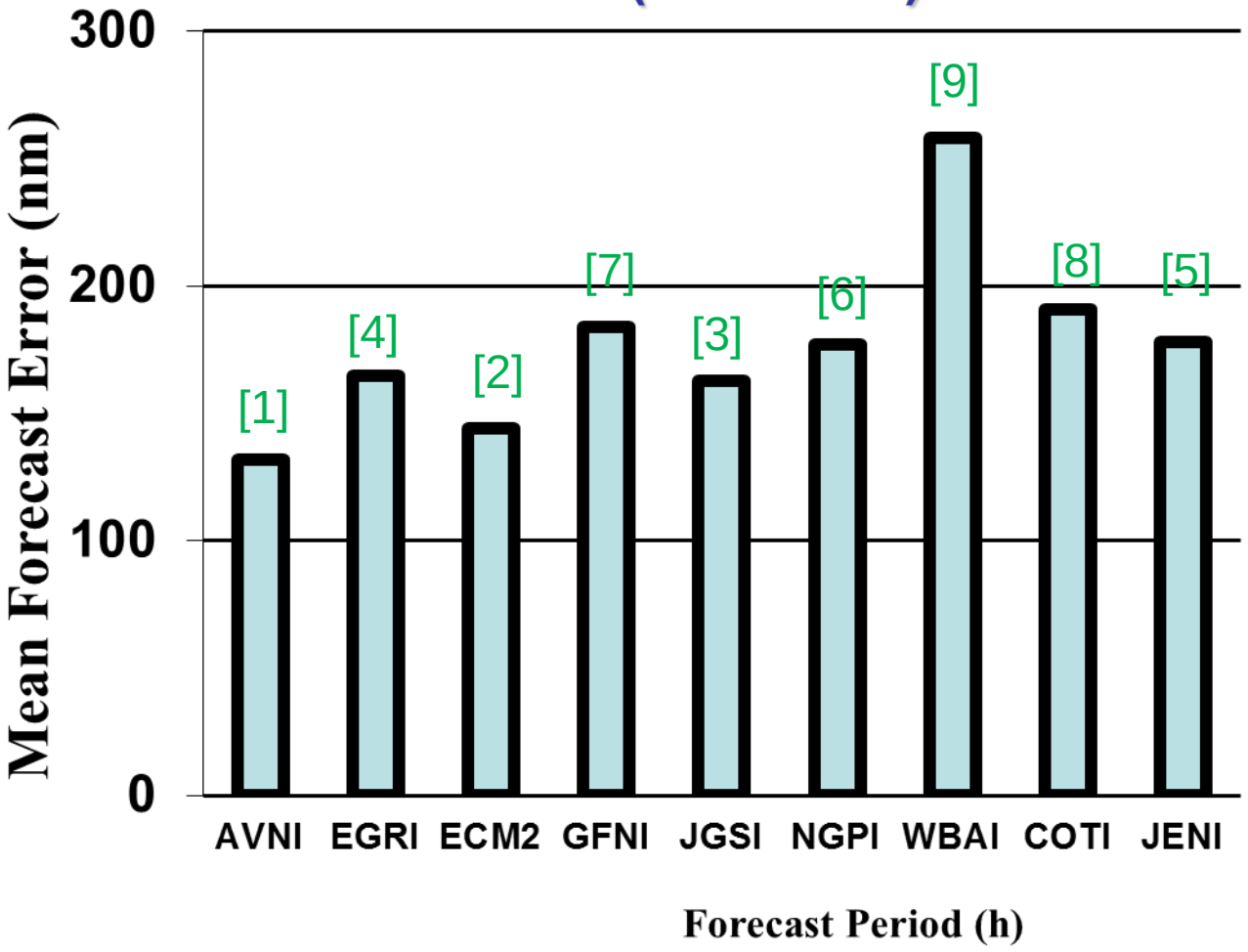


Aid Definitions

- NVGI/NGPI= NAVGEM
- EGRI= UK Met Office
- JGSI= Japanese Global
- GFNI= GFDL model
- AVNI= GFS
- HWFI= H-WRF
- ECM2= ECMWF
- CTCI=COAMPS-TC
- JENI=Japanese Ensemble
- AEMI=GFS ensemble mean
- ACEI=TC-ACCESS (Aussie Global)
- CONW= 2013 Consensus (average)
- COXX= Consensus Reruns
- WBAR= Weber barotropic model (removed 2013)

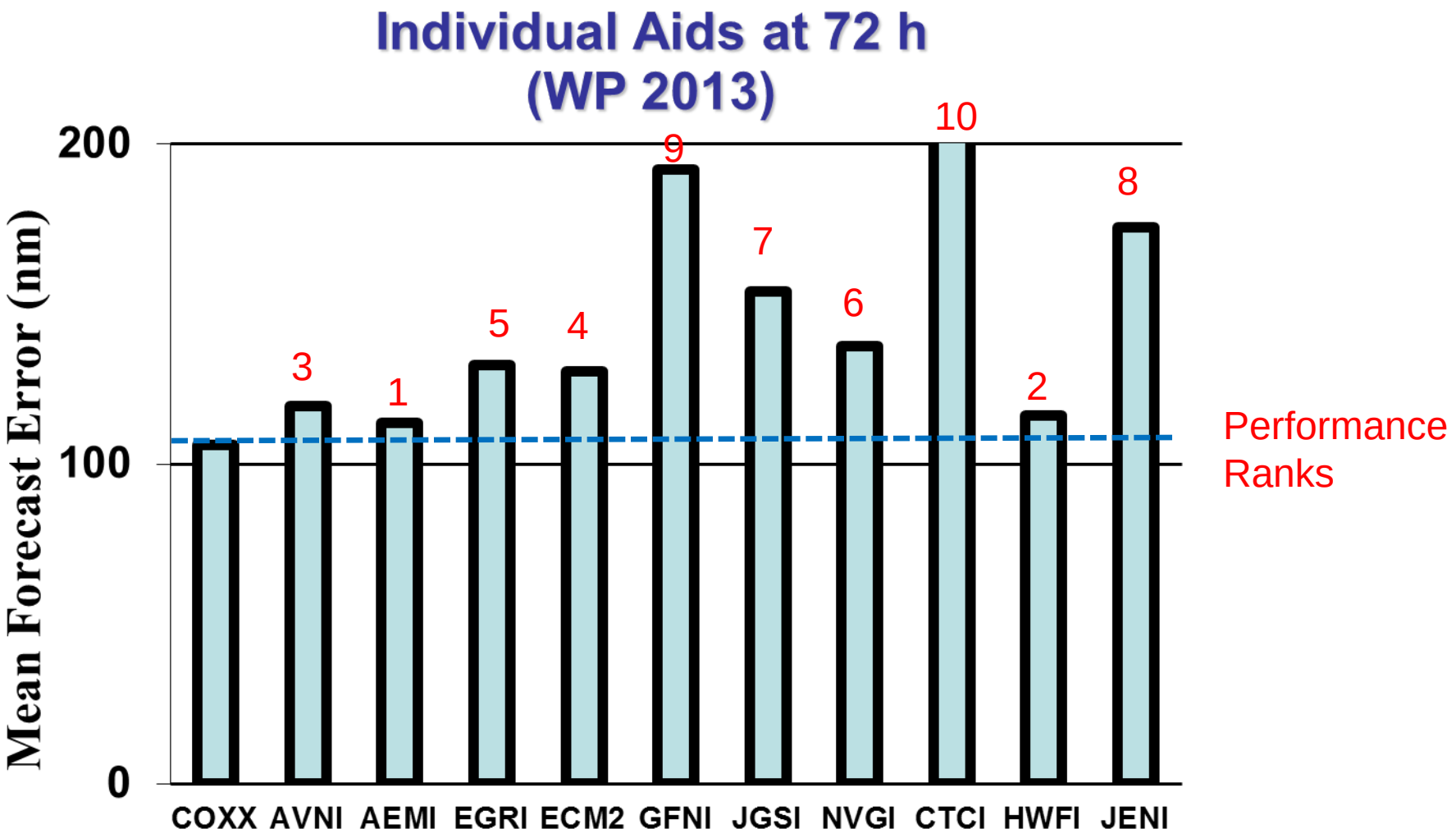
- All aids are “early” meaning they are there to use during the forecast process
- “I” in 4th character indicates 6-h interpolation
- “2” in 4th character indicates 12-h interpolation (ECM2)

Individual Aids at 72 h (WP 2012)



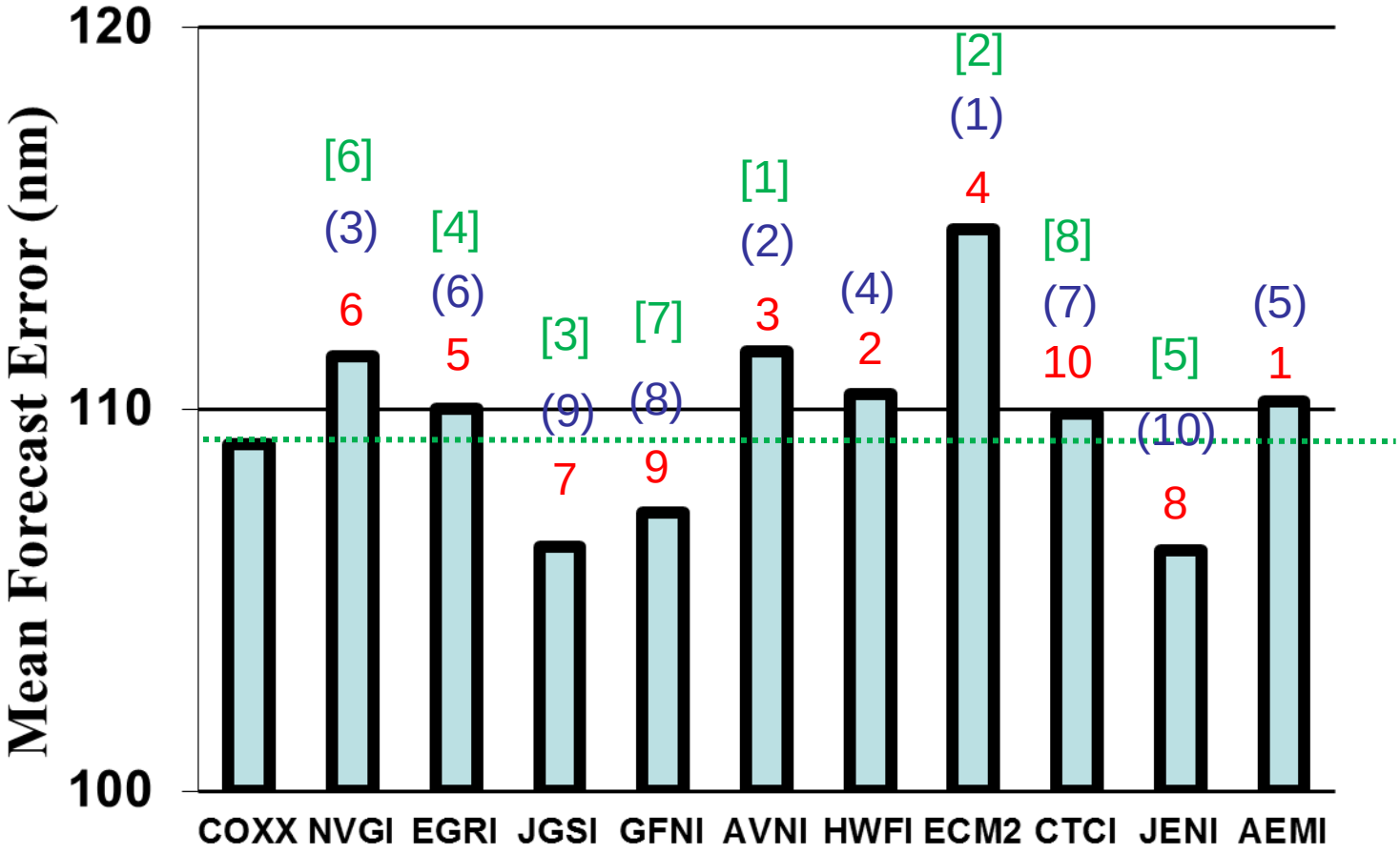
Performance Ranks

Real-time aid performance (inhomogeneous) for WP 2012 season. WBAI is considerably less skillful at this point. We have added two new members (COTI and JENI) in the last year. Is it time to retire WBAI from the track consensus?



Inhomogeneous aid track performance for CONW members and CONW reruns (COXX). COXX leader, HWFI, AVNI and AEMI very close.

WP Sensitivity Tests for 72 h (WP 2013)



Here we remove one aid (listed at bottom of each bar) from COXX to get contribution of each aid. Higher error indicates more contribution. ECM2 largest contributor, CTCI contributes even though it had the highest errors. WP 2013 season (283 cases).

Concept disproven

Current Performance

Past Performance

Aid	2012	2013	Contribution (2013)
ECM2	2	4	1
NVGI/NGPI	6	6	3
GFNI	7	9	8
EGRI	4	5	6
JGSI	3	7	9
JENI	5	8	10
AVNI	1	3	2
HWFI	N/A	2	4
AEMI	N/A	1	5
CTCI	8	10	7
WBAR	9	REMOVED	REMOVED

Colors indicate models that share components (physics, initial conditions, etc.) – Typically one of the group provides big contributions while the others contribute considerably less… This is due to their relative independence…

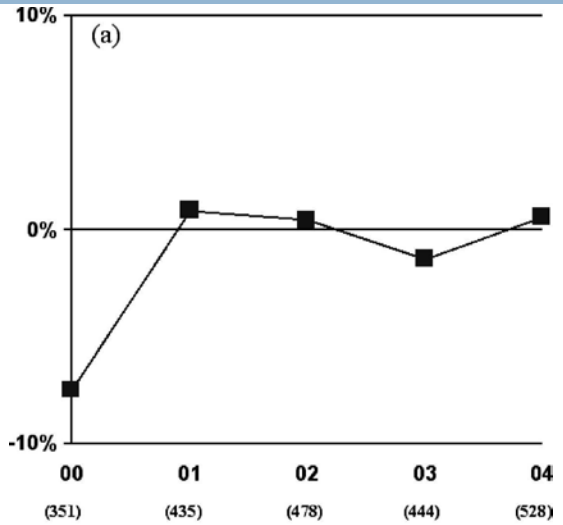
Comment on independence

- Model forecasts can be correlated.
- Those serial correlations result in a reduction of information or degrees freedom
- One can test
  - The number of effective degrees of freedom
    - See Sampson et al. (2008)
  - Past co-variation
  - If the number of aids is large one can combine strongly correlated models (e.g., a mean of an ensemble system)

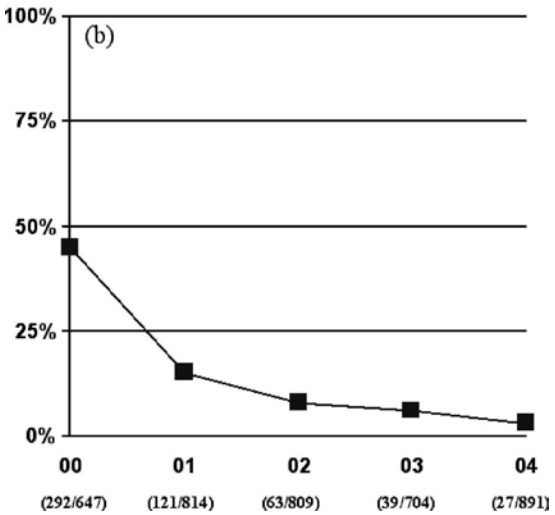
Comment on user selected consensus

- Selected consensus was a hot idea for track forecasting in the early-2000's.
- Conceptually it seems a user could select and remove a bad forecast
- However, verification of the methods reveled that user selected consensus under performed the all member equally weighted methods
- Forecasters confirmed this result as use of selected consensus fell of over time

Sampson et al. (2007)



Relative Skill



Frequency of use

Take away points

- Past and concurrent aid performance does not always indicate future performance.
- Past and concurrent performance does not always indicate how aid inclusion will impact the consensus forecast errors
- New aids become available and older aids become unavailable (argument for a variable approach)
- Independent modelling efforts seem to contribute more to the improvement of consensus forecasts (argument for many different efforts)
  - Multi-model consensus guidance out performed ensemble system means.
- User selected consensus did not provide additional improvements, but required extra work
- **Over the course of a season individual aids typically perform worst than the simple consensus, but not always**

Simple example: Initial Intensity Consensus Results (Sampson et al. 2008)

At the time (2006):

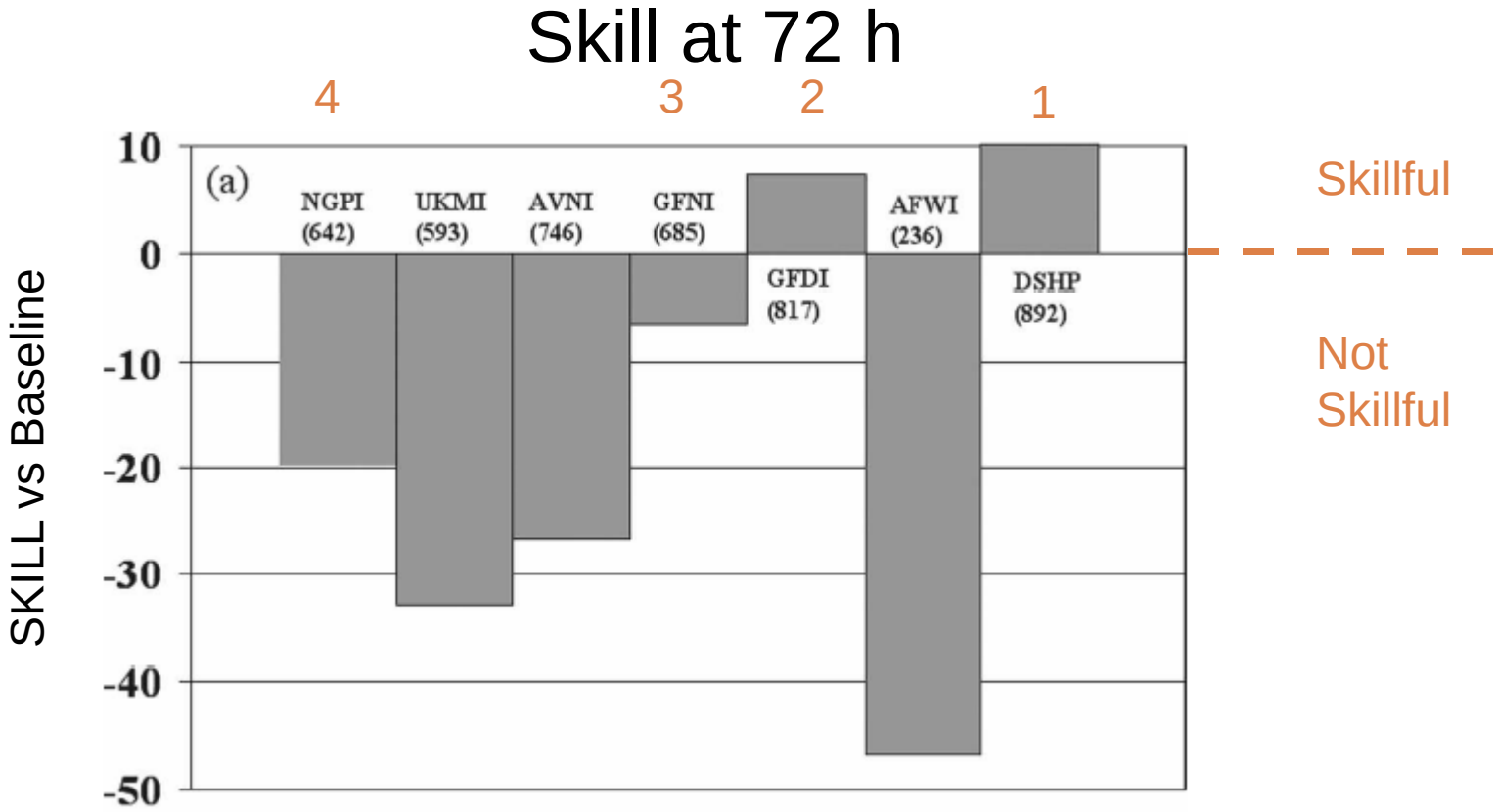
- Intensity forecasts had only recently become skillful (just a few models), & statistical methods were most skillful
- Few skillful models (2 or 3)
- Offers some insight to other practical issues
  - Availability
  - Variable vs. Fixed
- Analysis Caveats
  - No bias correction performed
  - Errors for individual aids are non-homogeneous

Potential members (in 2006)

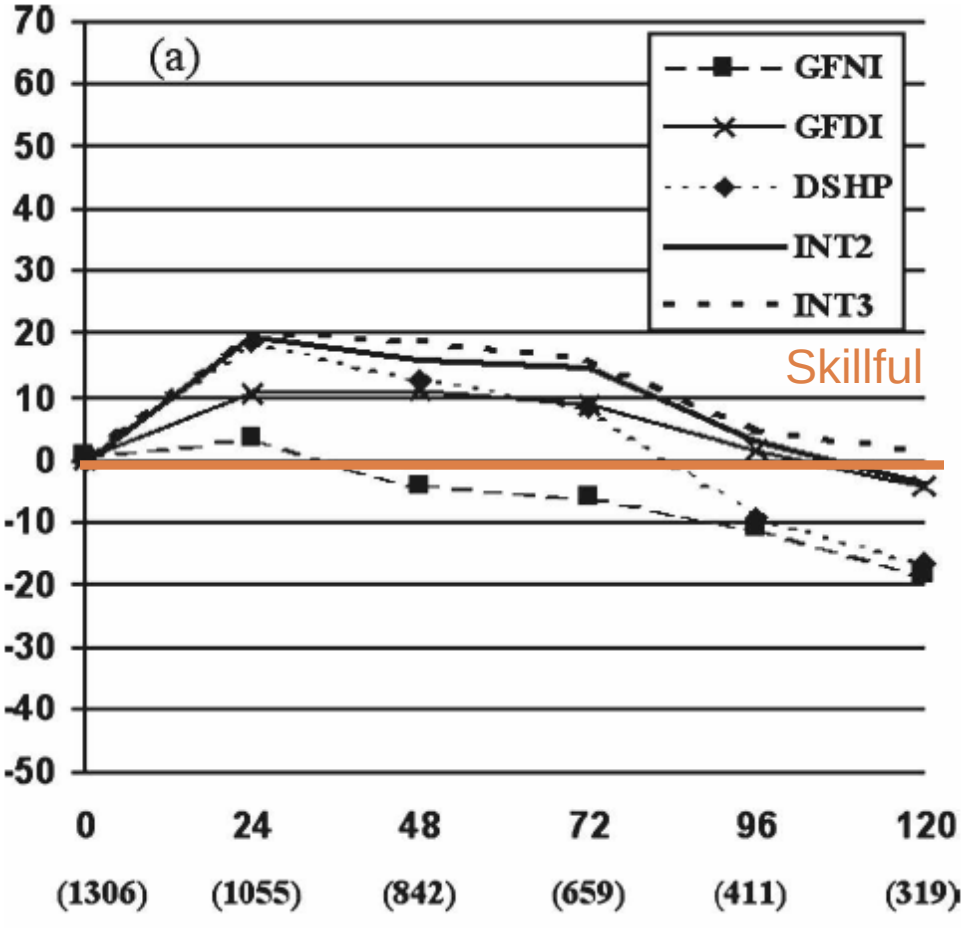
- NGPI – Navy NOGAPS model
- UKMI – UK Met Office
- AVNI – NCEP GFS model
- GFNI – GFLD Regional Hurricane Model (NOGAPS)
- GFDI –GFDL Regional Hurricane Model (GFS)
- AFWI – Air Force Regional Hurricane Model
- DSHP – Statistical Hurricane Intensity Prediction Scheme, accounting for decay over land.

-All models are early except for DSHP

Step 1: Identify potential candidates



Step 2: Test



INT2=DSHP+GFDI
INT3=DSHP+GFDI+GFNI

Observations:

1. Adding the unskillful GFNI improved the consensus forecasts

Step 3: Practical considerations

- A three member variable consensus is available more often (in our case 91% vs 83%)
- Variable consensus forecasts can be more challenging for forecasters to interpret
- Fixed consensus forecasts are available for fewer forecast periods

Lessons learned from our experiences

- Consensus forecasts made by equally weighted aids work as well, if not better, than other approaches
- Increasing the number of more independent aids makes for better consensus guidance
- Aid independence is as important as aid skill when constructing consensus guidance
- That forming a consensus from skillful members was not sufficient to reduce the consensus mean error. The members must also demonstrate independence from each other.
- Selective consensus efforts thus far have not been worth the extra effort
- Trimming the poorest performing models is often a good strategy for building superior consensus guidance

From Economics & Finance

“….combination methods have gained even more ground in the forecasting literature, largely because of the strength of the empirical evidence suggesting that these methods systematically perform better than alternatives based on forecasts from a single model. Stable, equal weights have so far been the workhorse of the combination literature and have set a benchmark that has proved surprisingly difficult to beat. This is surprising since−on theoretical grounds−one would not expect any particular combination scheme to be dominant, since the various methods incorporate restrictions on the covariance matrix that are designed to trade off bias against reduced parameter estimation error. The optimal bias can be expected to vary across applications, and the scheme that provides the best trade-off is expected to depend on the sample size, the number of forecasting models involved, the ratio of the variance of individual models’ forecast errors as well as their correlations and the degree of instability in the underlying data generating process.” - Timmermann (2005)

"At least since the publication of “The Combination of Forecasts” (Bates and Granger [1969]), economists have known that combining forecasts from different sources can both improve accuracy and reduce forecaster error. In the intervening years, numerous studies have confirmed these conclusions, outlined conditions under which forecast combinations are most effective, and tried to explain why simple equal weights work so well relative to more sophisticated statistical techniques.” – Jones (2014)

References:

- Bates, J. M., and C. W. J. Granger (1969). The Combination of Forecasts," Operational Research Quarterly 20, 451-68.
- DeMaria, M., C.R. Sampson, J.A. Knaff and K.D. Musgrave, (2014). Is tropical cyclone intensity guidance improving? Bull. Amer. Meteor. Soc., 95, 387-398.
- Jones, Robert C. (2014). "Making Better Investment Decisions". The Journal of Portfolio Management 40 (2): 128–143.
- Joint Typhoon Warning Center (2014). 2013 Annual Tropical Cyclone Report: Western Pacific (Report). United States Navy, United States Air Force.
- Sampson, C.R., J.L. Franklin, J.A. Knaff, and M. DeMaria (2008). Experiments with a Simple Tropical Cyclone Intensity Consensus. Weather and Forecasting, 23, 304–312.