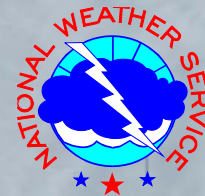# HFIP Performance Measures

James L. Franklin

HFIP Annual Meeting
November 2018
Miami, FL

Background

* Next phase of HFIP has identified three main focus areas (from *Weather Research and Forecasting Innovation Act of 2017*):

    * Improving forecasts of TC track and rapid intensification

    * Improving forecasts and communication of TC storm surge

    * Creating more effective watch and warning products through the use of risk communication research

New HFIP Goals

* Goal 1. Reduce track, intensity, and structure forecast guidance errors by 50% relative to a 2017 baseline. Reduce intensity forecast guidance errors by 50% for RI events. Metrics: 5.

* Goal 2. Produce 7-day track and intensity forecast guidance as accurate as a 2017 5-day baseline. Metrics: 2.

* Goal 3. Improve forecast guidance on pre-genesis disturbances, for track, intensity, and the timing of genesis, by 20% relative to a 2017 baseline. Metrics: 3.

* Goal 4. Improve hazard guidance and risk communication for all of the TC hazards (wind, surge, rainfall, and tornadoes) at actionable lead times through the application of social and behavioral sciences, resulting in a modernized suite of TC products, information, and services.
  * Goal 4 has 11(!) sub-goals. Metrics: 6, Milestones 14.

Goal 1 Metrics

Reduce track, intensity, and structure forecast guidance errors by 50% relative to a 2017 baseline. Reduce intensity forecast guidance errors by 50% for RI events.	
Metric 1.1	Mean absolute error (MAE) of TVCN track consensus
Baseline	TVCN MAE, Atlantic basin 2015-17, at 12, 24, 36, 48, 72, 96, and 120 hr: 23.1, 35.6, 47.6, 63.1, 101.5, 146.4, 195.9 n mi
Target	11.6, 17.8, 23.8, 31.6, 50.8, 73.2, 98.0 n mi

* Advantages:
  * TVCN provides good representation of overall state of the track guidance.

* Disadvantages:
  * TVCN MAE does not account for forecast difficulty – TVCN skill would have been better.
  * Composition of TVCN can vary from year to year.
  * Slow to respond to advances in individual models.

Goal 1 Metrics

Reduce track, intensity, and structure forecast guidance errors by 50% relative to a 2017 baseline. Reduce intensity forecast guidance errors by 50% for RI events.	
Metric 1.2	MAE of IVCN intensity consensus
Baseline	IVCN MAE, Atlantic basin 2015-17, at 12, 24, 36, 48, 72, 96, and 120 hr: 6.1, 8.5, 10.2, 11.4, 12.6, 14.4, 17.1 kt
Target	3.1, 4.3, 5.1, 5.7, 6.3, 7.2, 8.6 kt

* Advantages:
    * IVCN provides good representation of overall state of the intensity guidance.

* Disadvantages:
    * IVCN MAE does not account for forecast difficulty.
    * Composition of IVCN can vary from year to year.
    * Slow to respond to advances in individual models.
    * Some of the targets are lower than inherent ability to measure intensity.

Goal 1 Metrics

Reduce track, intensity, and structure forecast guidance errors by 50% relative to a 2017 baseline. Reduce intensity forecast guidance errors by 50% for RI events.	
Metric 1.3	RI metric TBD
Baseline	TBD
Target	TBD

* Option 1. Mean Atlantic basin IVCN errors for the subset of forecasts in which a 30-kt increase in intensity occurred during any 24-h period prior to the verification time.
    * Large sample will produce more robust results.
    * While attaining a 50% improvement in IVCN errors over all forecasts might be an unrealistic goal, attaining a 50% improvement for the RI cases is both detectible in principle and potentially of much greater value to users.
    * Can give misleading results; a favorable verification even if the RI event that occurred earlier was not well forecast.

Goal 1 Metrics

Reduce track, intensity, and structure forecast guidance errors by 50% relative to a 2017 baseline. Reduce intensity forecast guidance errors by 50% for RI events.	
Metric 1.3	RI metric TBD
Baseline	TBD
Target	TBD

* Option 2. Mean forecast error, of IVCN or a designated HFIP model, evaluated only at verification times when RI is ongoing.
    * Directly addresses the stated goal, but ignores cases where RI was forecast but did not occur. Models that are overly aggressive will do well using this metric. Small sample.

* Option 3. Mean forecast error, of IVCN or a designated HFIP model, evaluated only at verification times when RI is either ongoing or was forecast.
    * Not precisely responsive to the stated goal, but because it includes the false alarms, provides forecasters with better sense of the credibility of a model forecast of RI.

* Option 4. Probability of detection / false alarm rate (POD/FAR) for 0-24 h, 24-48 h, 48-72 h, etc., for either IVCN or a designated HFIP model.

Goal 1 Metrics

Reduce track, intensity, and structure forecast guidance errors by 50% relative to a 2017 baseline. Reduce intensity forecast guidance errors by 50% for RI events.	
Metric 1.4	MAE of 34-kt RVCN radii consensus
Baseline	34-kt RVCN MAE, of Atlantic basin 2016-17, at 12, 24, 36, 48, 72, 96, and 120 hr: 21, 22, 22, 22, 25, 28, 36 n mi
Target	10.5, 11, 11, 11, 12.5, 14, 18 n mi

* Advantages:
  * RVCN provides good representation of overall state of the TC size guidance.

* Disadvantages:
  * RVCN availability begins in 2016, so baseline sample smaller than desired.
  * Errors in the best-track radii large relative to the current errors, let alone relative to the target errors, making the targets essentially unreachable.

Goal 1 Metrics

Reduce track, intensity, and structure forecast guidance errors by 50% relative to a 2017 baseline. Reduce intensity forecast guidance errors by 50% for RI events.	
Metric 1.5	MAE of dynamical model (HWRF or follow-on) RMW forecasts, evaluated when an aircraft reconnaissance fix was made within 6 h of the initial and verifying times.
Baseline	C-RMW MAE, Atlantic basin reconnaissance-restricted sample for 2015-17, at 12, 24, 36, 48, 72, 96, and 120 h: 18.4, 18.7, 19.0, 19.0, 20.0, 20.5, 20.9 n mi
Target	9.2, 9.4, 9.5, 9.5, 10.0, 10.3, 10.5 n mi

* Advantages:
   * Baseline is available because it was needed to support WSP model.

* Disadvantages:
    * RMW not best-tracked; operational data will be noisy. Baseline and forecast models are different (i.e., "improvements over baseline" do not represent improvements over current capabilities.)

Goal 2 Metrics

Produce 7-day track and intensity forecast guidance as accurate as a 2017 5-day baseline.	
Metric 2.1	MAE of 7-day TVCN intensity consensus
Baseline	TVCN MAE, Atlantic basin 2015-17, at 178 h: 222.9 n mi
Target	120-h TVCN 2015-17: 195.9 n mi

* If NHC extends OFCL to 7 days, the inevitable handful of atrocious 7-day forecasts could undermine public confidence in track forecasting generally.

* Implementation decisions by NCEP over the past several years have largely disregarded observed degradations in the GFS/GFS ensemble 7-day forecasts.
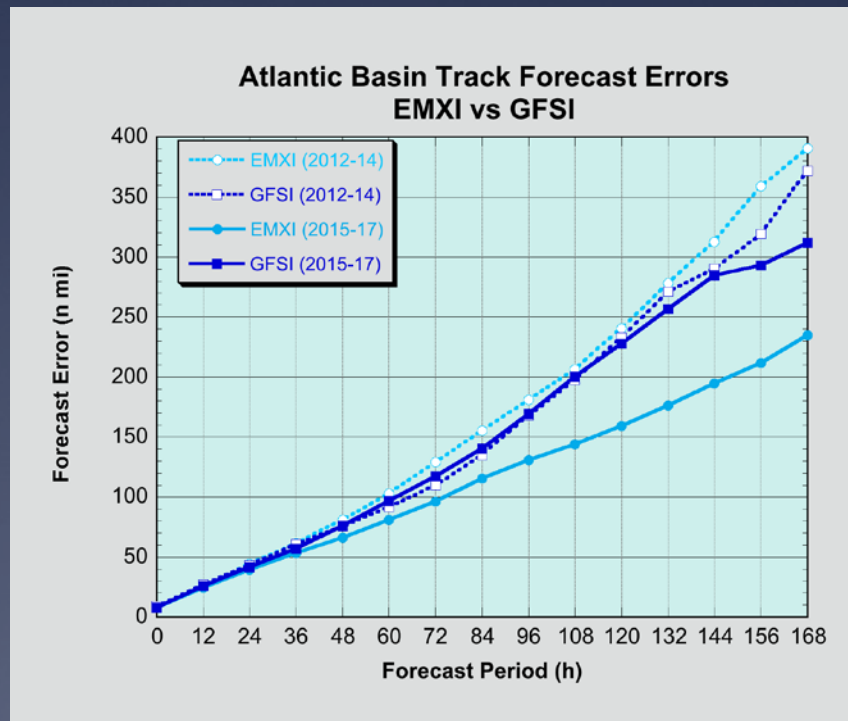
Atlantic Basin Track Forecast Errors EMXI vs GFSI

Goal 2 Metrics

Produce 7-day track and intensity forecast guidance as accurate as a 2017 5-day baseline.	
Metric 2.2	MAE of 7-day IVCN intensity consensus
Baseline	AVNI MAE, Atlantic basin 2015-17, at 178 h: 22.2 kt
Target	120-h IVCN 2015-17: 17.3 kt

Goal 3 Metrics

Improve forecast guidance on pre-genesis disturbances, for track, intensity, and the timing of genesis, by 20% relative to a 2017 baseline.	
Metric 3.1	MAE of TVCN track consensus for invests and PTCs at analysis time
Baseline	TVCN MAE, Atlantic basin 2015-17 for invests and PTCs at analysis time, at 12, 24, 36, 48, 72, 96, and 120 hr: [x] n mi
Target	0.8*Baseline

∗ Guidance not run on all pre-genesis disturbances (only invests and PTCs).

∗ No post-storm best track for invests, or for entire lifetime of PTCs. Operational "working" best tracks will have to be used.

∗ Requires software development to compute baseline or evaluate metric.

Goal 3 Metrics

Improve forecast guidance on pre-genesis disturbances, for track, intensity, and the timing of genesis, by 20% relative to a 2017 baseline.	
Metric 3.2	MAE of IVCN intensity consensus for invests and PTCs at analysis time
Baseline	IVCN MAE, Atlantic basin 2015-17 for invests and PTCs at analysis time, at 12, 24, 36, 48, 72, 96, and 120 hr: [y] kt
Target	0.8*Baseline

* Guidance not run on all pre-genesis disturbances (only invests and PTCs).

* No post-storm best track for invests, or for entire lifetime of PTCs. Operational "working" best tracks will have to be used.

* Requires software development to compute baseline or evaluate metric.

Goal 3 Metrics

Improve forecast guidance on pre-genesis disturbances, for track, intensity, and the timing of genesis, by 20% relative to a 2017 baseline.	
Metric 3.3	MAE of the predicted time of tropical (or subtropical) cyclone genesis
Baseline	MAE of (sub-)tropical cyclone genesis time from GFDL tracker applied to the GFS for 2015-17 Atlantic basin invests and PTCs at analysis time: [z] h
Target	0.8*Baseline

∗ Requires substantial development effort. Unclear what the value of this measure is, given that status as a TC no longer has any bearing on what warnings are issued or products disseminated.

Goal 4 Metrics

Improve hazard guidance and risk communication for all of the TC hazards (wind, surge, rainfall, and tornadoes) at actionable lead times through the application of social and behavioral sciences, resulting in a modernized suite of TC products, information, and services.

Goal 4.1	**Increase lead time of real-time storm surge products and services from two days to three days with no loss of skill.**
Metric 4.1	Mean three-day cross-track error used by P-surge, unadjusted for initial intensity (five-yr mean NHC-official cross-track error for all Atlantic basin forecasts that initiate and verify between 10-45N and 60-100W).
Baseline	Mean three-day NHC-official cross-track error for all Atlantic basin forecasts that initiate and verify between 10-45N and 60-100W for the period 2013-17: 57 n mi
Target	Mean three-day cross-track errors less than or equal to the mean two-day NHC-official cross-track error for all Atlantic basin forecasts that initiate and verify between 10-45N and 60-100W for the period 2013-17: 35 n mi

∗   Cross-track error is a large driver of P-surge output.

Goal 4 Metrics

Improve hazard guidance and risk communication for all of the TC hazards (wind, surge, rainfall, and tornadoes) at actionable lead times through the application of social and behavioral sciences, resulting in a modernized suite of TC products, information, and services.

Goal 4.3	Improve the accuracy of the 34-, 50-, and 64-kt wind speed probabilities at two and five days by 50%.
Metric 4.3	The threat score of the 34-, 50-, and 64-kt wind speed probabilities, evaluated on a regular grid covering 15-50 °N and 60-100 °W, at days two and five.
Baseline	The threat score from the 2017 version of the wind speed probability model, run on the Atlantic 2015-2017 sample over the domain 15-50 °N, 60-100 °W, for the 34-, 50- and 64-kt thresholds: Day 2 [x, y, z]; Day 5 [xx, yy, zz]
Target	Day 2 1.5*[x, y, z] Day 5 1.5*[xx, yy, zz]

Goal 4 Metrics

Improve hazard guidance and risk communication for all of the TC hazards (wind, surge, rainfall, and tornadoes) at actionable lead times through the application of social and behavioral sciences, resulting in a modernized suite of TC products, information, and services.

Goal 4.4	Perform a systematic evaluation of dynamical model forecasts for wind gusts associated with TCs.
Metric(?) 4.4	Publication or technical report.
Baseline	N/A
Target	Publication or technical report.

Goal 4 Metrics

Improve hazard guidance and risk communication for all of the TC hazards (wind, surge, rainfall, and tornadoes) at actionable lead times through the application of social and behavioral sciences, resulting in a modernized suite of TC products, information, and services.

Goal 4.5	**Improve the accuracy and lead time of the WPC Excessive Rainfall Outlook for TCs.**
Metric 4.5	Brier Score of Day-3 Excessive Rainfall Outlook for landfalling Atlantic basin TCs
Baseline	Current Brier Score of Day-Three Excessive Rainfall Outlook, 2015-7 Atlantic basin CONUS-landfalling TCs: [x]
Target	Current Brier Score of Day-Two Excessive Rainfall Outlook: [y]

Goal 4 Metrics

Improve hazard guidance and risk communication for all of the TC hazards (wind, surge, rainfall, and tornadoes) at actionable lead times through the application of social and behavioral sciences, resulting in a modernized suite of TC products, information, and services.

Goal 4.6	**Improve skill of Quantitative Precipitation Forecasts (QPF) for landfalling TCs.**
Metric 4.6	QPF Brier Score for TCs affecting CONUS, Puerto Rico, and U.S. Virgin Islands
Baseline	QPF Brier Score for TCs affecting CONUS, Puerto Rico, and U.S. Virgin Islands during 2015-17: [x]
Target	10% improvement over baseline

Goal 4 Metrics

Improve hazard guidance and risk communication for all of the TC hazards (wind, surge, rainfall, and tornadoes) at actionable lead times through the application of social and behavioral sciences, resulting in a modernized suite of TC products, information, and services.

Goal 4.8	Improve the SPC one-day probabilistic forecast for tornadoes by 10%.
Metric 4.8	The resolution of the SPC one-day tornado probabilities associated with landfalling tropical cyclones, as measured by the Brier Score.
Baseline	Brier score of the SPC one-day tornado probabilities associated with landfalling tropical cyclones from 2015-2017.
Target	Baseline * 1.1

Goal 4 Metrics

Improve hazard guidance and risk communication for all of the TC hazards (wind, surge, rainfall, and tornadoes) at actionable lead times through the application of social and behavioral sciences, resulting in a modernized suite of TC products, information, and services.

Goal 4.9	**Assess how the current TC product suite is used across America's Weather Enterprise, by NWS partners, and by end users.**
Milestone 4.9.1	Documentation of the current TC product suite and completion of a baseline assessment of TC product use across the NWS.
Milestone 4.9.2	Completion of a baseline assessment of TC product use across America's Weather Industry, with an emphasis on identification of high- versus low-use products, including an assessment of how the product suite is modified and dissemination to public audiences.
Milestone 4.9.3	Completion of a baseline assessment of NWS core partners' and end-users' use and understanding of the current TC product suite, with particular emphasis on their numeracy skills and understanding of probabilities.
Target	Complete the milestones.

Goal 4 Metrics

Improve hazard guidance and risk communication for all of the TC hazards (wind, surge, rainfall, and tornadoes) at actionable lead times through the application of social and behavioral sciences, resulting in a modernized suite of TC products, information, and services.

Goal 4.10	Identify requirements for a modernized TC product suite.
Milestone 4.10.1	Completion of a baseline assessment of NWS partners' and user TC information needs. Such an assessment should consider potential time scales from minutes to weeks in advance of an event, as well as characteristics of the needed information such as risk, confidence, uncertainty, formats, interactivity, methods of delivery, etc.
Milestone 4.10.2	Identification of intended communication objectives for social and behavioral sciences researchers.
Milestone 4.10.3	Completion of a baseline social and behavioral science analysis of efficacy of current TC products to support key decision-making by NWS partners and users to: 1) meet intended communication objectives and ensure partners' and users' information needs are met.
Milestone 4.10.4	Synthesis of baseline assessments and review by the HFIP Socio-Economic Working Group to determine operational viability of identified needs. Development of product prioritization, identifying which products/services should be streamlined, expanded, or discontinued.
Target	Complete the milestones.

Goal 4 Metrics

Improve hazard guidance and risk communication for all of the TC hazards (wind, surge, rainfall, and tornadoes) at actionable lead times through the application of social and behavioral sciences, resulting in a modernized suite of TC products, information, and services.	
Goal 4.11	**Develop and disseminate a modernized TC product suite that is informed by probabilistic information to better convey risk and uncertainty, and through which enables enhanced risk assessment and timely preparedness actions on the part of users, partners, and stakeholders to reduce loss of life and property, and which includes other weather-forecast-related organizations in our shared responsibility in the effective conveyance of risk and uncertainty.**
Milestone 4.11.1	Creation of working groups, guided by the HFIP Socio-Economic Working Group, representing interdisciplinary expertise (e.g., social, behavioral, and physical science researchers, operational forecasters) necessary to manage, develop, implement, and disseminate the proposed product changes.
Milestone 4.11.2	Gathering of NWS partner and user feedback on proposed product changes through the use of partner/user engagement, as well as social and behavioral science methodologies.
Milestone 4.11.3	Development of an NWS partnership with other weather-forecast-related organizations that disseminates the modernized NWS TC product suite, and that empowers the other weather information providers to develop theirs own TC products to better convey risk and uncertainty.
Milestone 4.11.4	Development of an NWS partnership with key agencies and professional organizations (e.g., FEMA, NEMA, IAEM, AMS, NWA, etc.) that ensures effective training and outreach is developed and available to key partners and users of the modernized TC product suite.
Target	Complete the milestones.

Summary of Metrics

	Metric	Responsible Agency (Proposed)	Capability In Place
1.1	Mean absolute error (MAE) of TVCN track consensus	NHC	Yes
1.2	MAE of IVCN track consensus	NHC	Yes
1.3	RI metric TBD	NHC	No
1.4	MAE of 34-kt RVCN radii consensus	NHC	Yes
1.5	MAE of dynamical model (HWRF or follow-on) RMW forecasts, evaluated when an aircraft reconnaissance fix was made within 6 h of the initial and verifying times.	Unknown	No
2.1	MAE of 7-day TVCN track consensus	NHC	Yes
2.2	MAE of 7-day IVCN track consensus	NHC	Yes
3.1	MAE of TVCN track consensus for invests and PTCs at analysis time	NHC	No
3.2	MAE of IVCN intensity consensus for invests and PTCs at analysis time	NHC	No
3.3	MAE of the predicted time of tropical (or subtropical) cyclone genesis	Unknown	No

Summary of Metrics

	Metric	Responsible Agency (Proposed)	Capability In Place
4.1	Mean three-day cross-track error used by P-surge, unadjusted for initial intensity (five-yr mean NHC-official cross-track error for all Atlantic basin forecasts that initiate and verify between 10-45N and 60-100W).	NHC	Yes
4.2	Number of surge products and services that have been expanded to entire NWS area of responsibility vulnerable to TC storm surge	NHC	Yes
4.3	The threat score of the 34-, 50-, and 64-kt wind speed probabilities, evaluated on a regular grid covering 15-50 °N and 60-100 °W, at days two and five.	NHC?	No
4.5	Brier Score of Day-3 Excessive Rainfall Outlook for landfalling Atlantic basin TCs	WPC?	No
4.6	QPF Brier Score for TCs affecting CONUS, Puerto Rico, and U.S. Virgin Islands	WPC?	No
4.8	The resolution of the SPC one-day tornado probabilities associated with landfalling tropical cyclones, as measured by the Brier Score.	SPC?	No