# Model assessment using a multi-metric ranking technique

*Pat Fitzpatrick and Yee Lau, Mississippi State University at Stennis*
*Ghassan Alaka and Frank Marks, NOAA AOML Hurricane Research Division*

Philosophy

Weighting multiple metrics and techniques provides clearer product validation comparison

It also allows unequal weighting if certain metrics are deemed more important. Currently, equal weighting is used.

Results are tentative. Focus more on strategy and technique for discussions.

Metrics are combinations of two general categories: statistical and outliers

- Concept generated from frustration to distinguish top-tier ocean models. Early ranking concept shown below for drifter direction, but also computed for moored buoys, different seasons, current speed, different vertical levels, etc.

Table 6. As in Table 4, but for drifter current direction in four summer months.

| | First | Second | Third | Fourth | | Weighted Sum | Rank |
|---|---|---|---|---|---|---|---|
| NCOM AMSEAS | 9 | 11 | 3 | 1 | | 52 | 1 |
| NCOM Global | 3 | 1 | 8 | 12 | | 19 | 4 |
| HYCOM GOM | 3 | 7 | 7 | 7 | | 30 | 3 |
| HYCOM Global | 9 | 5 | 6 | 4 | | 43 | 2 |

- Also wanted a parameter easy for non-technical decision-makers to use
- Today's presentation focuses on 2016 and 2017 Atlantic tropical cyclone forecast products, but can be applied to any set of forecast guidances needing comparisons.

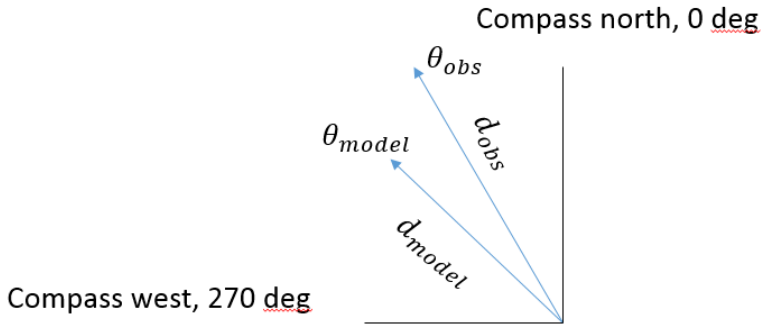
***Ocean model validation report available at:***
***http://www.ngi.msstate.edu/about/documents/progressReport2013.pdf***

# Tropical cyclone product comparisons

# Track by magnitude and direction

- Error distance did not allow proper usage of all the metrics, as its scalar and positive

Validation is done in a vector fashion:




- Validation is done in a vector fashion:
  1. By magnitude (nm)
  2. By direction (compass deg relative to observed track).
- Distance computed using Haversine formula. West and East quadrant discontinuities handled by adding or subtracting 360 in a consistent fashion.

# Track also by along- and cross-track

- However, many metrics are not possible, generally limited to absolute error

# Intensity

- $\Delta V_{max}$

# Structure (Radius of 34-knot winds)

- R34, right-front quadrant
- However, fewer products to examine

# Outlier metrics

- Absolute error percentage –
  - For track: percentage within 35 nm and within 7 deg for 48-h.
  - For track: percentage within 50 nm and within 12 deg for 72-h. (0 to 100%, 100% best)
  - For intensity: percentage within 20 kts for 48-h, and 25 kts for 72 h
  - For R34: percentage within 25 nm for 48-h, and 50 nm for 72 h

- Outlier metrics outside these thresholds (>=0, 0 best in all cases)
  1) Positive outlier percentage
  2) Negative outlier percentage
  3) Percentage within tolerance
  4) Number of occurrences with consecutive positive outliers
  5) Number of occurrences with consecutive negative outliers
  6) Maximum duration of consecutive positive outliers
  7) Maximum duration of consecutive negative outliers

Difference metrics computed as forecast minus observed

# Statistical metrics

- Statistical metrics –
  1) Model efficiency factor (<= +1, +1 best)
  2) Pearson correlation coefficient (-1 to +1, +1 best)
  3) Kendall's Tau (-1 to +1, +1 best)
  4) Reliability index (>= +1, +1 best)
  5) Multiplicative gross error (>= 0, 0 best)
  6) Root mean squared difference (>= 0, 0 best)
  7) Bias (any number, 0 best)
  8) Absolute error (>= 0, 0 best)

Difference metrics computed as forecast minus observed

# Track (2016)

**Only 72-hr shown, but other forecast intervals have been computed**

# Track products evaluated

- AEMN (GFS ensemble mean)
- GFDL
- GHMI (previous cycle GFDL, “early” interpolated)
- HWRF
- HWFI (previous cycle HWRF, “early” interpolated)
- Basin-scale HWRF (HB16)
- LBAR
- NHC official
- TVCA (consensus of variants for at least two of the following:  UKMET, HWRF, ECMWF, GFS, GFDL)

# 72-h ranking, magnitude

Sample size 243

| magnitude | AEMN_Mag | GFDL_Mag | GHMI_Mag | HB16_Mag | HWFI_Mag | HWRF_Mag | LBAR_Mag | OFCL_Mag | TVCA_Mag |
|---|---|---|---|---|---|---|---|---|---|
| modelEfficiency | 3 | 7 | 8 | 4 | 6 | 5 | 9 | 2 | 1 |
| pearson | 3 | 7 | 8 | 4 | 6 | 5 | 9 | 2 | 1 |
| kendall | 3 | 7 | 8 | 4 | 6 | 5 | 9 | 2 | 1 |
| reliabilityIndex | 5 | 7 | 8 | 2 | 6 | 4 | 9 | 3 | 1 |
| multiplicativeGrossErr | 3 | 7 | 8 | 4 | 6 | 5 | 9 | 2 | 1 |
| rmsd | 3 | 7 | 8 | 4 | 6 | 5 | 9 | 2 | 1 |
| Bias Error | 4 | 1 | 3 | 6 | 7 | 8 | 9 | 5 | 2 |
| Absolute Error | 3 | 7 | 8 | 4 | 6 | 5 | 9 | 2 | 1 |
| % > +tolerance | 3 | 4 | 5 | 6 | 9 | 7.5 | 7.5 | 1 | 2 |
| % < -tolerance | 8 | 5 | 6 | 1 | 3 | 2 | 9 | 7 | 4 |
| % within tolerance | 4 | 7 | 8 | 3 | 6 | 5 | 9 | 1 | 2 |
| # Consecutive > +tolerance | 2 | 5 | 5 | 7 | 8.5 | 5 | 8.5 | 1 | 3 |
| # Consecutive < -tolerance | 8 | 9 | 5.5 | 2.5 | 1 | 2.5 | 5.5 | 7 | 4 |
| max duration of consecutive > +tolerance | 2.5 | 6.5 | 6.5 | 6.5 | 9 | 6.5 | 4 | 1 | 2.5 |
| max duration of consecutive < -tolerance | 9 | 2.5 | 7 | 1 | 5.5 | 2.5 | 4 | 5.5 | 8 |
| total | 63.5 | 89 | 102 | 59 | 91 | 73 | 119.5 | 43.5 | 34.5 |
| average | 4.233333 | 5.933333 | 6.8 | 3.933333 | 6.066667 | 4.866667 | 7.966667 | 2.9 | 2.3 |
| overall rank | 4 | 6 | 8 | 3 | 7 | 5 | 9 | 2 | 1 |

# 72-h ranking, direction

Sample size 243

| direction | AEMN_Dir | GFDL_Dir | GHMI_Dir | HB16_Dir | HWFI_Dir | HWRF_Dir | LBAR_Dir | OFCL_Dir | TVCA_Dir |
|---|---|---|---|---|---|---|---|---|---|
| modelEfficiency | 5 | 7 | 8 | 2 | 6 | 4 | 9 | 3 | 1 |
| pearson | 5.5 | 7 | 8 | 2.5 | 5.5 | 4 | 9 | 2.5 | 1 |
| kendall | 2 | 7 | 8 | 4 | 6 | 5 | 9 | 3 | 1 |
| reliabilityIndex | 6 | 7 | 3 | 5 | 4 | 8 | 9 | 1 | 2 |
| multiplicativeGrossErr | 3 | 7 | 2 | 6 | 5 | 8 | 9 | 1 | 4 |
| rmsd | 5 | 7 | 8 | 2 | 6 | 4 | 9 | 3 | 1 |
| Bias Error | 1 | 3 | 5 | 4 | 8 | 2 | 9 | 6 | 7 |
| Absolute Error | 1 | 7 | 8 | 4 | 6 | 5 | 9 | 3 | 2 |
| % > +tolerance | 1 | 7 | 8 | 3.5 | 6 | 3.5 | 9 | 2 | 5 |
| % < -tolerance | 4 | 7 | 8 | 5.5 | 3 | 5.5 | 9 | 2 | 1 |
| % within tolerance | 2 | 7 | 8 | 4.5 | 6 | 4.5 | 9 | 1 | 3 |
| # Consecutive > +tolerance | 3 | 8 | 8 | 5.5 | 5.5 | 3 | 8 | 1 | 3 |
| # Consecutive < -tolerance | 6.5 | 8 | 6.5 | 3 | 4.5 | 4.5 | 9 | 1.5 | 1.5 |
| max duration of consecutive > +tolerance | 4 | 6.5 | 6.5 | 1.5 | 4 | 1.5 | 9 | 4 | 8 |
| max duration of consecutive < -tolerance | 2 | 7 | 4 | 9 | 7 | 7 | 4 | 1 | 4 |
| total | 51 | 102.5 | 99 | 62 | 82.5 | 69.5 | 129 | 35 | 44.5 |
| average | 3.4 | 6.833333 | 6.6 | 4.133333 | 5.5 | 4.633333 | 8.6 | 2.333333 | 2.966667 |
| overall rank | 3 | 8 | 7 | 4 | 6 | 5 | 9 | 1 | 2 |

# Rankings, along- and cross-track

48-h
Sample size 302

| Avg Abs Err ranl | 4 | 7 | 8 | 3 | 6 | 5 | 9 | 2 | 1 |
|---|---|---|---|---|---|---|---|---|---|
| | AEMN_Cross | GFDL_Cross | GHMI_Cross | HB16_Cross | HWFI_Cross | HWRF_Cross | LBAR_Cross | OFCL_Cross | TVCA_Cross |

| Avg Abs Err ranl | 1 | 7 | 8 | 4 | 6 | 5 | 9 | 3 | 2 |
|---|---|---|---|---|---|---|---|---|---|
| | AEMN_Along | GFDL_Along | GHMI_Along | HB16_Along | HWFI_Along | HWRF_Along | LBAR_Along | OFCL_Along | TVCA_Along |

72-h
Sample size 243

| Avg Abs Err ranl | 3 | 7 | 8 | 4 | 6 | 5 | 9 | 2 | 1 |
|---|---|---|---|---|---|---|---|---|---|
| | AEMN_Cross | GFDL_Cross | GHMI_Cross | HB16_Cross | HWFI_Cross | HWRF_Cross | LBAR_Cross | OFCL_Cross | TVCA_Cross |

| Avg Abs Err ranl | 2 | 7 | 8 | 4 | 6 | 5 | 9 | 1 | 3 |
|---|---|---|---|---|---|---|---|---|---|
| | AEMN_Along | GFDL_Along | GHMI_Along | HB16_Along | HWFI_Along | HWRF_Along | LBAR_Along | OFCL_Along | TVCA_Along |

# Intensity (2016)

# Intensity products evaluated

- AEMN (GFS ensemble mean)
- DSHIPS
- GFDL
- GHMI (previous cycle GFDL)
- HWRF
- HWFI (previous cycle HWRF)
- Basin-scale (HB16)
- IVCN
- LGEM
- NHC official
- SHIPS

# 72-h ranking

Sample size 243

| ranking | AEMN_Vmax | DSHP_Vmax | GFDL_Vmax | GHMI_Vmax | HB16_Vmax | HWFI_Vmax | HWRF_Vmax | IVCN_Vmax | LGEM_Vmax | OFCL_Vmax | SHIP_Vmax |
|---|---|---|---|---|---|---|---|---|---|---|---|
| modelEfficiency | 11 | 6 | 8 | 5 | 10 | 3 | 4 | 1 | 9 | 2 | 7 |
| pearson | 11 | 8 | 9 | 6 | 3 | 5 | 4 | 1 | 10 | 2 | 7 |
| kendall | 11 | 8 | 9 | 6 | 3 | 5 | 4 | 1 | 10 | 2 | 7 |
| reliabilityIndex | 11 | 9 | 10 | 3 | 5 | 4 | 2 | 7 | 6 | 1 | 8 |
| rmsd | 11 | 6 | 8 | 5 | 10 | 3 | 4 | 1 | 9 | 2 | 7 |
| BiasErr | 10 | 1 | 7 | 6 | 11 | 3 | 4 | 2 | 8 | 5 | 9 |
| AbsErr | 11 | 7 | 8 | 5 | 9 | 4 | 3 | 1 | 6 | 2 | 10 |
| Pos% | 6 | 6 | 10 | 8.5 | 11 | 4 | 6 | 2 | 1 | 3 | 8.5 |
| Neg% | 11 | 8.5 | 8.5 | 6.5 | 1.5 | 4.5 | 6.5 | 4.5 | 10 | 1.5 | 3 |
| AcptAbsErr% | 9 | 6.5 | 11 | 8 | 10 | 3 | 5 | 2 | 4 | 1 | 6.5 |
| #CsctPos | 7 | 4.5 | 10 | 7 | 11 | 4.5 | 2.5 | 2.5 | 1 | 7 | 9 |
| #CsctNeg | 9.5 | 9.5 | 6.5 | 6.5 | 1 | 4.5 | 9.5 | 4.5 | 9.5 | 2.5 | 2.5 |
| MaxDurPos | 7 | 9 | 4 | 11 | 9 | 6 | 4 | 2 | 1 | 4 | 9 |
| MaxDurNeg | 9 | 9 | 6 | 2 | 6 | 4 | 2 | 6 | 11 | 2 | 9 |
| total | 134.5 | 98 | 115 | 85.5 | 100.5 | 57.5 | 60.5 | 37.5 | 95.5 | 37 | 102.5 |
| average | 9.607142857 | 7 | 8.214285714 | 6.107142857 | 7.178571429 | 4.107142857 | 4.321428571 | 2.678571429 | 6.821428571 | 2.642857143 | 7.321428571 |
| overall rank | 11 | 7 | 10 | 5 | 8 | 3 | 4 | 2 | 6 | 1 | 9 |
| | AEMN | DSHP | GFDL | GHMI | HB16 | HWFI | HWRF | IVCN | LGEM | OFCL | SHIP |

# Structure (2016)

**Only 4 products available**

# 72-h ranking

Sample size 164

| ranking | HB16_R34 | HWFI_R34 | HWRF_R34 | OFCL_R34 |
|---|---|---|---|---|
| modelEfficiency | 4 | 3 | 2 | 1 |
| pearson | 3 | 4 | 2 | 1 |
| kendall | 2 | 3 | 1 | 4 |
| reliabilityIndex | 3 | 4 | 2 | 1 |
| multiplicativeGrossErr | 4 | 3 | 2 | 1 |
| rmsd | 4 | 3 | 2 | 1 |
| BiasErr | 4 | 2 | 3 | 1 |
| AbsErr | 4 | 3 | 2 | 1 |
| Pos% | 4 | 2 | 3 | 1 |
| Neg% | 1.5 | 3.5 | 1.5 | 3.5 |
| AcptAbsErr% | 4 | 3 | 2 | 1 |
| #CsctPos | 4 | 2 | 3 | 1 |
| #CsctNeg | 1.5 | 4 | 1.5 | 3 |
| MaxDurPos | 4 | 1 | 3 | 2 |
| MaxDurNeg | 1 | 2.5 | 2.5 | 4 |
| total | 48 | 43 | 32.5 | 26.5 |
| average | 3.2 | 2.86666667 | 2.16666667 | 1.76666667 |
| overall rank | 4 | 3 | 2 | 1 |
| | HB16 | HWFI | HWRF | OFCL |

# 2017 hurricane season tentative results

# Track magnitude, 72 hrs

| | | | | | | |
|---|---|---|---|---|---|---|
| average | 3.5 | 2.93333333 | 5.06666667 | 2.93333333 | 2.96666667 | 3.6 |
| overall rank | 4 | 1.5 | 6 | 1.5 | 3 | 5 |
| magnitude | AEMN | HB17 | HWFI | HWRF | OFCL | TVCA |

Sample size 214

# Track direction, 72 hrs

| | | | | | | |
|---|---|---|---|---|---|---|
| average | 3.86666667 | 2.63333333 | 5.33333333 | 4.36666667 | 3.06666667 | 1.7333333 |
| overall rank | 4 | 2 | 6 | 5 | 3 | 1 |
| direction | AEMN | HB17 | HWFI | HWRF | OFCL | TVCA |

# Intensity, 72 hrs

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| average | 7.32142857 | 5.53571429 | 4.78571429 | 2.32142857 | 2.64285714 | 4.53571429 | 2.57142857 | 6.28571429 |
| overall rank | 8 | 6 | 5 | 1 | 3 | 4 | 2 | 7 |
| intensity | AEMN | DSHP | HB17 | HWFI | HWRF | LGEM | OFCL | SHIP |

# Structure (R34), 72 hrs

| | | | | |
|---|---|---|---|---|
| average | 3.71428571 | 1.75 | 2.07142857 | 2.4642857 |
| overall rank | 4 | 1 | 2 | 3 |
| R34 | HB17 | HWFI | HWRF | OFCL |

# What if interpolated "early" version of basin-scale HWRF [HB7I] included? (smaller sample size)

Sample size 156

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| average | 3.8 | 4.3666667 | 5.1 | 5.4 | 4.3 | 1.9 | 3.1333333 |
| overall rank | 3 | 5 | 6 | 7 | 4 | 1 | 2 |
| magnitude | AEMN | HB17 | HB7I | HWFI | HWRF | OFCL | TVCA |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| average | 3.8333333 | 3.3666667 | 5.0333333 | 6.1 | 5.2333333 | 2.6 | 1.8333333 |
| overall rank | 4 | 3 | 5 | 7 | 6 | 2 | 1 |
| direction | AEMN | HB17 | HB7I | HWFI | HWRF | OFCL | TVCA |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| average | 7.964285714 | 5.464285714 | 5.5 | 5.428571429 | 3.321428571 | 3.535714286 | 4.392857143 | 2.5 | 6.892857143 |
| overall rank | 9 | 6 | 7 | 5 | 2 | 3 | 4 | 1 | 8 |
| intensity | AEMN | DSHP | HB17 | HB7I | HWFI | HWRF | LGEM | OFCL | SHIP |

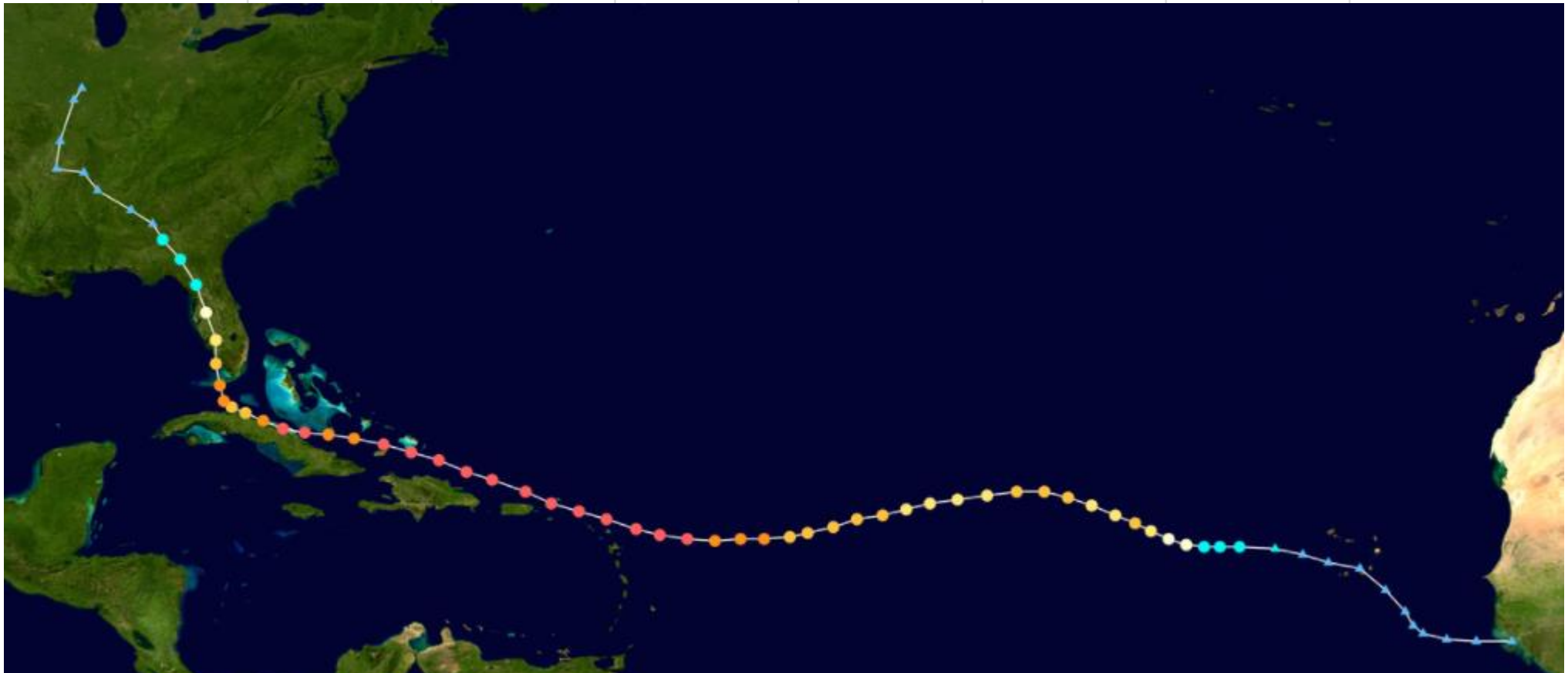
| | | | | | |
|---|---|---|---|---|---|
| average | 3.607142857 | 4.535714286 | 2.142857143 | 2.321428571 | 2.392857143 |
| overall rank | 4 | 5 | 1 | 2 | 3 |
| R34 | HB17 | HB7I | HWFI | HWRF | OFCL |

# Validation for Hurricane Irma

Sample size 27

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| average | 5.36666667 | 3.4 | 4.3 | 4.03333333 | 2.16666667 | 3.4 | 5.33333333 |
| overall rank | 7 | 2.5 | 5 | 4 | 1 | 2.5 | 6 |
| magnitude | AEMN | HB17 | HB7I | HWFI | HWRF | OFCL | TVCA |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| average | 3.1 | 3.8 | 5.5 | 4.86666667 | 4.96666667 | 2.4 | 3.36666667 |
| overall rank | 2 | 4 | 7 | 5 | 6 | 1 | 3 |
| direction | AEMN | HB17 | HB7I | HWFI | HWRF | OFCL | TVCA |



# Conclusions and Future work

Conclusions

- Multi-metric ranking is possibly an effective, flexible technique to compare forecasts products, including measures of extreme outlier events and consistency, into a single quantity
- Designed to drive model separation in validation comparisons

Tropical cyclone product comparisons

- Track – Consensus often best, but the experimental basin-scale HWRF is showing more skill than operational HWRF at 72 hrs. In 2017, GFS ensemble mean skill lagged more than in 2016.
- Intensity – HWRF began exceeding statistical schemes in 2016, and continued in 2017.
- Structure – In 2017, HWRF products exceeded NHC's empirical schemes for the first time.

Future

- Looking for operational center partners
- Validate with vector correlations. See:

Zaron, E. D., P. J. Fitzpatrick, S. L. Cross, J. M. Harding, F. L. Bub, J. D. Wiggert, D. Ko, Y. Lau, K. Woodard, and C. N. K. Mooers, 2015: Initial evaluations of a U. S. Navy rapidly relocatable Gulf of Mexico/Caribbean ocean forecast system in the context of the Deepwater Horizon oil spill disaster. *Frontiers of Earth Science*, **9**, 1-32. DOI: 10.1007/s11707-014-0508-x: 1-32.

- Weighted averaging? Other metrics?

# Extra slides

- *mean absolute error*: $\mathrm{MAE} = \frac{1}{n}\sum_{t=1}^{n}|y_t - \hat{y}_t|$
  is comparable to the *median absolute deviation*: $\mathrm{MAD}(y_t) := \frac{1}{n}\sum_{t=1}^{n}|y_t - median(y_t)|$

- *root mean squared error*: $\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{t=1}^{n}(y_t - \hat{y}_t)^2}$
  is comparable to the *standard derivation*: $sd(y_t)$

- *root centered mean squared error*: $\mathrm{RCMSE} = \sqrt{\frac{1}{n-1}\sum_{t=1}^{n}(y_t - \hat{y}_t - \mathrm{mean}(y_t - \hat{y}_t))^2}$
  is comparable to $sd(y_t)$ and to differences of $y_t$-values.

- *root mean squared logarithmic error*: $\mathrm{RMSLE} = \sqrt{\frac{1}{n}\sum_{t=1}^{n}\log(y_t/\hat{y}_t)^2}$
  is comparable to $\mathrm{sd}(\log(y_t))$ and to log-ratios of values of $y_t$.

Reliability index is exp(RMSLE)

- The Pearson correlation coefficient:

$$r = \frac{\sum_{t=1}^{n}(x_t - \bar{x})(y_t - \bar{y})}{\sqrt{\sum_{t=1}^{n}(x_t - \bar{x})^2}\sqrt{\sum_{t=1}^{n}(y_t - \bar{y})^2}},$$

- The *rank correlation coefficient* according to Spearman is similar to Pearson's correlation. However, instead of the data itself, the ranks are used.

Symbolically, Spearman's rank correlation coefficient is denoted by $r_s$ . It is given by the following formula:

$r_s = 1 - (6\sum d_i^2) / (n(n^2-1))$

**Here $d_i$ represents the difference in the ranks given to the values of the variable for each item of the particular data*

This formula is applied in cases when there are no tied ranks. However, in the case of fewer numbers of tied ranks, this approximation of Spearman's rank correlation coefficient provides sufficiently good approximations.

- The *efficiency factor* of Nash and Sutcliffe (1970) is also a dimensionless statistical measure, which directly relates model predictions to observed data:

$$\begin{aligned} \text{EF} &= 1 - \frac{(SS \text{ over } y = \hat{y})}{(\text{improved } SS \text{ over } y)} \\ &= 1 - \frac{\sum_{t=1}^{n}(y_t - \hat{y}_t)^2}{\sum_{t=1}^{n}(y_t - \overline{y})^2}, \end{aligned}$$

where $SS$ is the sum of squares. The efficiency factor takes values between minus infinity and 1, whereas negative values do not recommend for the model and values near 1 indicate good model performance.

$$\textit{Multiplicative gross error} = \frac{100}{n}\sum \frac{|y - \hat{y}|}{\hat{y}}$$

Kendall's Tau: usually smaller values than Spearman's rho correlation. Calculations based on concordant and discordant pairs. Insensitive to error. P values are more accurate with smaller sample sizes.

Spearman's rho: usually have larger values than Kendall's Tau. Calculations based on deviations. Much more sensitive to error and discrepancies in data.

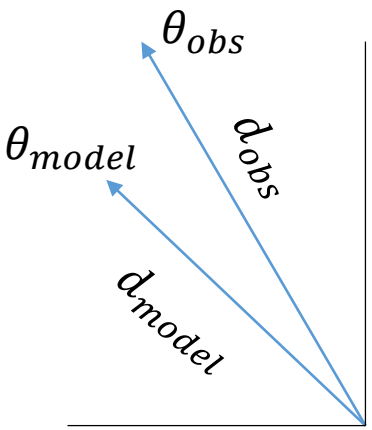
The main advantages of using Kendall's tau are as follows:

The distribution of Kendall's tau has better statistical properties.
The interpretation of Kendall's tau in terms of the probabilities of observing the agreeable (concordant) and non-agreeable (discordant) pairs is very direct.
In most of the situations, the interpretations of Kendall's tau and Spearman's rank correlation coefficient are very similar and thus invariably lead to the same inferences.

# Track products evaluated

- AEMN - GFS Ensemble Mean
- GFDL - GFDL Hurricane Model
- GHMI - Previous cycle GFDL, adjusted using a variable intensity offset correction that is a function of forecast time. Note that for track, GHMI and GFDI are identical.
- HWFI - Previous cycle HWRF, adjusted
- HWRF - HWRF Hurricane Model
- LBAR - Limited Area Barotropic Model
- NHC official – National Hurricane Center official forecast
- TVCA – consensus; average of at least two of EMXI, GFSI, EGRI, GHMI, and HWRI
    1. EMXI – Previous cycle of ECMWF, adjusted
    2. GFSI – Previous cycle of GFS, adjusted
    3. EGRI – Previous cycle of UKMET with subjective QC for tracker, adjusted
    4. HWRI – Previous cycle of HWRF, adjusted
- HEAD – HEDAS, all data assimilated. HEDAS is  HWRF Hurricane Ensemble Data Assimilation System

Compass north, 0 deg




Compass west, 270 deg

# 48-h ranking, magnitude

Sample size 302

| magnitude | AEMN_Mag | GFDL_Mag | GHMI_Mag | HB16_Mag | HWFI_Mag | HWRF_Mag | LBAR_Mag | OFCL_Mag | TVCA_Mag |
|---|---|---|---|---|---|---|---|---|---|
| modelEfficiency | 3 | 5 | 8 | 4 | 7 | 6 | 9 | 2 | 1 |
| pearson | 3.5 | 6 | 8 | 3.5 | 7 | 5 | 9 | 2 | 1 |
| kendall | 4 | 6 | 8 | 3 | 7 | 5 | 9 | 2 | 1 |
| reliabilityIndex | 6 | 7 | 8 | 2 | 4 | 5 | 9 | 3 | 1 |
| multiplicativeGrossErr | 3 | 7 | 8 | 4 | 6 | 5 | 9 | 1 | 2 |
| rmsd | 3 | 5 | 8 | 4 | 7 | 6 | 9 | 2 | 1 |
| Bias Error | 5 | 3 | 1 | 8 | 6 | 9 | 7 | 4 | 2 |
| Absolute Error | 3 | 4 | 8 | 5 | 7 | 6 | 9 | 2 | 1 |
| % > +tolerance | 4 | 3 | 5 | 7 | 6 | 8 | 9 | 2 | 1 |
| % < -tolerance | 6 | 7 | 8 | 1 | 3 | 2 | 9 | 5 | 4 |
| % within tolerance | 3.5 | 3.5 | 6 | 5 | 8 | 7 | 9 | 2 | 1 |
| # Consecutive > +tolerance | 3 | 5 | 4 | 9 | 6.5 | 8 | 6.5 | 1 | 2 |
| # Consecutive < -tolerance | 5 | 4 | 9 | 2 | 7 | 1 | 7 | 3 | 7 |
| max duration of consecutive > +tolerance | 3 | 3 | 5 | 6 | 8.5 | 8.5 | 7 | 3 | 1 |
| max duration of consecutive < -tolerance | 5.5 | 2.5 | 5.5 | 1 | 2.5 | 5.5 | 8.5 | 8.5 | 5.5 |
| total | 60.5 | 71 | 99.5 | 64.5 | 92.5 | 87 | 126 | 42.5 | 31.5 |
| average | 4.03333333 | 4.73333333 | 6.63333333 | 4.3 | 6.16666667 | 5.8 | 8.4 | 2.83333333 | 2.1 |
| overall rank | 3 | 5 | 8 | 4 | 7 | 6 | 9 | 2 | 1 |

# 48-h ranking, direction

Sample size 302

| direction | AEMN_Dir | GFDL_Dir | GHMI_Dir | HB16_Dir | HWFI_Dir | HWRF_Dir | LBAR_Dir | OFCL_Dir | TVCA_Dir |
|---|---|---|---|---|---|---|---|---|---|
| modelEfficiency | 2 | 7 | 8 | 3 | 4.5 | 4.5 | 9 | 6 | 1 |
| pearson | 2 | 7 | 8 | 3 | 5.5 | 4 | 9 | 5.5 | 1 |
| kendall | 1 | 7 | 8 | 3 | 6 | 5 | 9 | 4 | 2 |
| reliabilityIndex | 3 | 7 | 8 | 4 | 6 | 5 | 9 | 2 | 1 |
| multiplicativeGrossErr | 2 | 7 | 8 | 4 | 6 | 5 | 9 | 3 | 1 |
| rmsd | 2 | 7 | 8 | 3 | 4 | 5 | 9 | 6 | 1 |
| Bias Error | 6 | 8 | 7 | 4 | 5 | 2 | 9 | 3 | 1 |
| Absolute Error | 2 | 7 | 8 | 3 | 6 | 5 | 9 | 4 | 1 |
| % > +tolerance | 2 | 7 | 8 | 5 | 6 | 4 | 9 | 1 | 3 |
| % < -tolerance | 6 | 7 | 8 | 2 | 4 | 5 | 9 | 3 | 1 |
| % within tolerance | 4 | 7 | 8 | 3 | 6 | 5 | 9 | 2 | 1 |
| # Consecutive > +tolerance | 3 | 7 | 8 | 5 | 5 | 5 | 9 | 1 | 2 |
| # Consecutive < -tolerance | 5.5 | 7.5 | 7.5 | 2 | 4 | 5.5 | 9 | 3 | 1 |
| max duration of consecutive > +tolerance | 3 | 7 | 8 | 3 | 6 | 3 | 9 | 3 | 3 |
| max duration of consecutive < -tolerance | 3.5 | 8 | 6 | 6 | 1.5 | 3.5 | 9 | 1.5 | 6 |
| total | 47 | 107.5 | 116.5 | 53 | 75.5 | 66.5 | 135 | 48 | 26 |
| average | 3.13333333 | 7.16666667 | 7.76666667 | 3.53333333 | 5.03333333 | 4.43333333 | 9 | 3.2 | 1.73333333 |
| overall rank | 2 | 7 | 8 | 4 | 6 | 5 | 9 | 3 | 1 |

# 48-h ranking

Sample size 301

| ranking | AEMN_Vmax | DSHP_Vmax | GFDL_Vmax | GHMI_Vmax | HB16_Vmax | HWFI_Vmax | HWRF_Vmax | IVCN_Vmax | LGEM_Vmax | OFCL_Vmax | SHIP_Vmax |
|---|---|---|---|---|---|---|---|---|---|---|---|
| modelEfficiency | 11 | 5 | 6 | 7 | 8 | 3 | 4 | 2 | 9 | 1 | 10 |
| pearson | 11 | 6 | 8 | 7 | 3 | 5 | 4 | 2 | 9 | 1 | 10 |
| kendall | 11 | 5.5 | 7 | 4 | 3 | 8 | 5.5 | 2 | 9 | 1 | 10 |
| reliabilityIndex | 11 | 8 | 5 | 7 | 6 | 9 | 3.5 | 2 | 3.5 | 1 | 10 |
| rmsd | 11 | 5 | 6 | 7 | 8 | 3 | 4 | 2 | 9 | 1 | 10 |
| BiasErr | 11 | 4 | 7 | 6 | 10 | 5 | 1 | 2 | 9 | 3 | 8 |
| AbsErr | 9 | 6 | 10 | 7 | 8 | 3 | 4 | 2 | 5 | 1 | 11 |
| Pos% | 5.5 | 4 | 9.5 | 5.5 | 11 | 7 | 8 | 3 | 1.5 | 1.5 | 9.5 |
| Neg% | 11 | 7 | 9 | 8 | 2 | 3 | 4 | 5 | 10 | 1 | 6 |
| AcptAbsErr% | 10 | 3.5 | 11 | 7 | 8 | 3.5 | 6 | 2 | 5 | 1 | 9 |
| #CsctPos | 6.5 | 6.5 | 9 | 3 | 10.5 | 6.5 | 6.5 | 3 | 1 | 3 | 10.5 |
| #CsctNeg | 6.5 | 8 | 9 | 10.5 | 2 | 2 | 4.5 | 4.5 | 10.5 | 2 | 6.5 |
| MaxDurPos | 10.5 | 7.5 | 4.5 | 7.5 | 7.5 | 7.5 | 4.5 | 2.5 | 1 | 2.5 | 10.5 |
| MaxDurNeg | 8 | 8 | 3.5 | 1.5 | 3.5 | 8 | 8 | 8 | 8 | 1.5 | 8 |
| total | 133 | 84 | 104.5 | 88 | 90.5 | 73.5 | 67.5 | 42 | 90.5 | 21.5 | 129 |
| average | 9.5 | 6 | 7.464285714 | 6.285714286 | 6.464285714 | 5.25 | 4.821428571 | 3 | 6.464285714 | 1.535714286 | 9.214285714 |
| overall rank | 11 | 5 | 9 | 6 | 7.5 | 4 | 3 | 2 | 7.5 | 1 | 10 |
| | AEMN | DSHP | GFDL | GHMI | HB16 | HWFI | HWRF | IVCN | LGEM | OFCL | SHIP |

# 48-h ranking

Sample size 206

| ranking | HB16_R34 | HWFI_R34 | HWRF_R34 | OFCL_R34 |
|---|---|---|---|---|
| modelEfficiency | 3 | 4 | 2 | 1 |
| pearson | 3 | 4 | 2 | 1 |
| kendall | 3 | 4 | 2 | 1 |
| reliabilityIndex | 3 | 4 | 2 | 1 |
| multiplicativeGrossErr | 4 | 3 | 2 | 1 |
| rmsd | 3 | 4 | 2 | 1 |
| BiasErr | 4 | 2 | 3 | 1 |
| AbsErr | 3 | 4 | 2 | 1 |
| Pos% | 4 | 2.5 | 2.5 | 1 |
| Neg% | 1 | 4 | 2 | 3 |
| AcptAbsErr% | 3 | 4 | 2 | 1 |
| #CsctPos | 4 | 3 | 2 | 1 |
| #CsctNeg | 1 | 4 | 2 | 3 |
| MaxDurPos | 3 | 1 | 2 | 4 |
| MaxDurNeg | 1 | 2.5 | 2.5 | 4 |
| total | 43 | 50 | 32 | 25 |
| average | 2.866666667 | 3.333333333 | 2.133333333 | 1.666666667 |
| overall rank | 3 | 4 | 2 | 1 |
| | HB16 | HWFI | HWRF | OFCL |