Object-based verification of NWP model performance regarding the wind structure of Super Typhoon Hato

Wai-Kin WONG¹, WC Woo¹, KC Au-Yeung ¹ and YC Wong²
¹Hong Kong Observatory
²The Chinese University of Hong Kong

ESCAP/WMO Typhoon Committee Technical Conference (TECO)
26 – 27 February 2018
Hanoi, Viet Nam
Common TC Verification Metrics

**Position**

- All basins
- Northern hemisphere
- Southern hemisphere
- Western North Pacific

**Intensity**

- BoM
- CMA
- CMC
- ECMWF
- Météo-France
- JMA
- KMA
- NCEP
- NRL
- UK Met Office

**Intenity biases in TC forecasts.** Scatter plots of analysed (best track) versus 3-day forecast minimum sea level pressure, indicating how the different models share similar TC intensity biases especially for TCs analysed to have very low minimum pressure. The verification period is 2012–2014.

Source: Working Group on Numerical Experimentation (WGNE), jointly established by the World Climate Research Programme (WCRP) Joint Scientific Committee (JSC) and the World Meteorological Organization’s Commission for Atmospheric Sciences (CAS)
But TC comes in all sizes...
Typhoon Damrey approaching Vietnam on 3 Nov 2017
Impact Based Warnings

• Multi-levels Warnings based on local wind strength, for example:
  – No. 1 / 3 / 8 / 9 / 10 in Hong Kong & Macao
  – Blue / Yellow / Orange / Red in Mainland China
  – No. 1 - 5 in the Philippines
  – Watch / Warning in Australia
  – Stage 1 - 4 in India
  – Emergency Warning / Warning / Advisory in Japan

• Local wind strength depends not only on location and intensity of TC, but also wind structures
Object-based verification is to provide:

- **Performance for each attribute** (e.g. position, angle, intensity, etc.)
- **An overall score** considering all attributes with weighting

### Table

<table>
<thead>
<tr>
<th></th>
<th>Subjectivity</th>
<th>CSI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fcst A</td>
<td>Inaccurate Position</td>
<td>0</td>
</tr>
<tr>
<td>Fcst B</td>
<td>Inaccurate Size</td>
<td>~0.5</td>
</tr>
<tr>
<td>Fcst C</td>
<td>Inaccurate Axis Angle</td>
<td>~0.5</td>
</tr>
<tr>
<td>Fcst D</td>
<td>Miss</td>
<td>0</td>
</tr>
</tbody>
</table>
**MODE (Method for Object-Based Diagnostic Evaluation)**

- NCAR - Developmental Testbed Center (DTC)
  - MET *(Model Evaluation Tools)* verification package
  - MODE
  - https://dtcenter.org/met/users/

![Raw field](image1)

Object identification

![Object field](image2)

Attribute calculation

Object analysis

Object 1 Attributes:
- Centroid Position
- Area
- Axis Angle
- 90-percentile Intensity

Object 2 Attributes:
- ...
- ...
- ...
**MODE** *(Method for Object-Based Diagnostic Evaluation)*

- **Forecast object analysis**
  - Object 1 Attributes
  - Object 2 Attributes
  - ...

- **Observation object analysis**
  - Object 1 Attributes
  - Object 2 Attributes
  - ...

For each of all combinations of Forecast Objects and Observation Objects, calculate **Scoring Curves** for each attribute.

**Weighting**

**Score (Interest)**
Object Identification

Raw Field -> Convolution (Parameter: Convolution Radius) -> Convolved Field

Convolved Field -> Mask Creation (Parameter: Threshold) -> Mask

Mask -> Masking -> Objects
Why Convolution?

How many pieces of bread here?
3. Normally we don’t count the crumbs.

How many Strong Wind (or above) objects?
3
Why Convolution?

Raw Field

ECMWF 08-21 12Z t+24

ECMWF 08-22 12Z Analysis

Strong Wind Objects
(convolution radius = 5)
Why Convolution?

ECMWF 08-21 12Z t+24

ECMWF 08-22 12Z Analysis

Raw Field

Strong Wind Objects
(no convolution)
**Attributes**

**Single Object Attributes**
- Centroid Position
- Area
- Axis Angle
- Intensity Percentile
- and more …

**Object Pair Attributes**
- Centroid Distance
- Area Ratio \((Smaller/Larger)\)
- Axis Angle Difference
- Intensity Percentile Ratio \((Smaller/Larger)\)
- Boundary Distance
- Interception Area Ratio \((Interception Area / Smaller of the \{Fctst,Obs\} Area)\)
Scoring Curves & Weighting (Defaults in MODE)

- **Centroid Distance (Grid)**
  - Weighting: 2

- **Boundary Distance (Grid)**
  - Weighting: 4

- **Axis Angle Difference (degree)**
  - Weighting: 1

- **Area Ratio**
  - Weighting: 1

- **Interception Area Ratio**
  - Weighting: 2
**MMI (Median of Maximum Interest)**

- A score for the whole forecast considering all objects

*Example*

**Forecast Objects:** 1, 2, 3  
**Observation Objects:** A, B

![Diagram showing the calculation of MMI](image)
Data adopted in this study:

**Forecast**

*Outputs from NWP models (ECMWF, JMA, NCEP, UKMO)*

**Observation**

*NOAA Multi-Platform Tropical Cyclone Wind Analysis*
Base Time: 2017-08-21 12Z, 36-hour f/c
Valid Time: 2017-08-23 00Z
SuperT Hato / *Gale Wind* or above
Base Time: 2017-08-21 12Z, 36-hour f/c
Valid Time: 2017-08-23 00Z
Super T Hato / **Gale Wind** or above
Base Time: 2017-08-21 12Z, 36-hour f/c
Valid Time: 2017-08-23 00Z
SuperT Hato / **Gale Wind** or above

ECMWF IFS
SuperT Hato / **Gale Wind** or above
SuperT Hato 36-hour forecast
Based at 2017/08/21 12Z Gale

WHICH IS BETTER?
SuperT Hato 36-hour forecast
Based at 2017/08/21 12Z Gale

UKMO: MMI 0.93

NCEP: MMI 0.87

ECMWF: MMI 0.96

JMA: MMI 0.95
Experiment

• Goal:
  • See whether performance of models as ranked by MODE MMI are consistent with subjective rankings by forecasters.

• Setup:
  • Without being shown the above object-based verification, 9 forecasters were asked to rank the performance of the TC structure by different models subjectively.

• Result …
Subjective Ranking of Hato:

<table>
<thead>
<tr>
<th>Forecaster #</th>
<th>Best</th>
<th>Better</th>
<th>Worse</th>
<th>Worst</th>
</tr>
</thead>
<tbody>
<tr>
<td>Forecaster #1</td>
<td>UKMO</td>
<td>ECMWF</td>
<td>NCEP</td>
<td>JMA</td>
</tr>
<tr>
<td>Forecaster #2</td>
<td>ECMWF</td>
<td>UKMO</td>
<td>NCEP</td>
<td>JMA</td>
</tr>
<tr>
<td>Forecaster #3</td>
<td>ECMWF</td>
<td>JMA</td>
<td>UKMO</td>
<td>NCEP</td>
</tr>
<tr>
<td>Forecaster #4</td>
<td>ECMWF</td>
<td>JMA</td>
<td>NCEP</td>
<td>UKMO</td>
</tr>
<tr>
<td>Forecaster #5</td>
<td>ECMWF</td>
<td>UKMO</td>
<td>JMA</td>
<td>NCEP</td>
</tr>
<tr>
<td>Forecaster #6</td>
<td>ECMWF</td>
<td>UKMO</td>
<td>JMA</td>
<td>NCEP</td>
</tr>
<tr>
<td>Forecaster #7</td>
<td>ECMWF</td>
<td>JMA</td>
<td>UKMO</td>
<td>NCEP</td>
</tr>
<tr>
<td>Forecaster #8</td>
<td>ECMWF</td>
<td>UKMO</td>
<td>JMA</td>
<td>NCEP</td>
</tr>
<tr>
<td>Forecaster #9</td>
<td>UKMO</td>
<td>ECMWF</td>
<td>NCEP</td>
<td>JMA</td>
</tr>
<tr>
<td>MODE MMI</td>
<td>ECMWF  (0.96)</td>
<td>JMA (0.95)</td>
<td>UKMO (0.93)</td>
<td>NCEP (0.87)</td>
</tr>
</tbody>
</table>
Discussions

• Most forecasters ranked ECMWF as the best and half of them ranked NCEP as the worst, generally in line with MODE MMI ranking.

• Many forecasters ranked JMA poorer than its MODE MMI rank (second best), probably as a result of large area of misses, which subjectively are “wronger”.

• Different weighting to attributes or fine-tune the calibration curves are to be designed
STS Pakhar 24-hour forecast
Based at 2017/08/26 00Z Gale
STS Pakhar 24-hour forecast
Based at 2017/08/26 00Z Gale

NCEP: MMI 0.43

ECMWF: MMI 0.85

JMA: MMI 0.00

UKMO: MMI 0.74
## Subjective Ranking of Pakhar

<table>
<thead>
<tr>
<th>Forecaster #</th>
<th>Best</th>
<th>Better</th>
<th>Worse</th>
<th>Worst</th>
</tr>
</thead>
<tbody>
<tr>
<td>Forecaster #1</td>
<td>ECMWF</td>
<td>UKMO</td>
<td>NCEP</td>
<td>JMA</td>
</tr>
<tr>
<td>Forecaster #2</td>
<td>ECMWF</td>
<td>UKMO</td>
<td>NCEP</td>
<td>JMA</td>
</tr>
<tr>
<td>Forecaster #3</td>
<td>ECMWF</td>
<td>UKMO</td>
<td>NCEP</td>
<td>JMA</td>
</tr>
<tr>
<td>Forecaster #4</td>
<td>ECMWF</td>
<td>UKMO</td>
<td>NCEP</td>
<td>JMA</td>
</tr>
<tr>
<td>Forecaster #5</td>
<td>ECMWF</td>
<td>UKMO</td>
<td>NCEP</td>
<td>JMA</td>
</tr>
<tr>
<td>Forecaster #6</td>
<td>ECMWF</td>
<td>UKMO</td>
<td>NCEP</td>
<td>JMA</td>
</tr>
<tr>
<td>Forecaster #7</td>
<td>ECMWF</td>
<td>UKMO</td>
<td>NCEP</td>
<td>JMA</td>
</tr>
<tr>
<td>Forecaster #8</td>
<td>ECMWF</td>
<td>UKMO</td>
<td>NCEP</td>
<td>JMA</td>
</tr>
<tr>
<td>Forecaster #9</td>
<td>ECMWF</td>
<td>UKMO</td>
<td>NCEP</td>
<td>JMA</td>
</tr>
<tr>
<td>MODE MMI</td>
<td>ECMWF</td>
<td>UKMO</td>
<td>NCEP</td>
<td>JMA</td>
</tr>
</tbody>
</table>
Discussion

• MODE MMI ranking matches perfectly with forecasters’ subjective ranking in this case.

• Likely due to significant difference in performance and clear-cut false alarms
Comments on performance of TC structure by NWPs

- ECMWF’s 36-hour forecast captured the gale wind structures of SuperT Hato quite well, revealing the usefulness of dynamical model guidance in this case.

- Most global NWP’s 24-hour forecast under-estimated gale winds of STS Pakhar, however.
Conclusions

• To support impact-based warnings, verification of TC wind structures are necessary.

• MODE provides an object-based verification method. MODE MMI seems to be useful for ranking forecast performance of TC wind structures.

• Tuning of the scoring curves and weightings to be explored to better reflect “what” forecasters or users are concerned.
Thank you very much