Tropical Cyclone Verification Code Intercomparison Project

TCVCIP
pronounced “tee see vee sip”

Mike Fiorino
michael.fiorino@noaa.gov
NOAA ESRL Boulder CO
14 May 2014

- motivations…
- why me? – doing TC verification code since 1977 – AMIP I&II verification
- stat diffs between ESRL v EMC v NHC v JTWC…
- TCVCIP basic case(s) – PE using working and final best tracks
- forecast error = FE = f(PE,IE) – not today…
TC activity 2014

Tropical Cyclones by Basin for: 20140101–20150101
bars colorized by intensity, multiply TCs below main line as of 2014043000

Jan 2014 Feb Mar Apr May Jun Jul Aug Sep Oct Nov Dec Jan 2015

- LANT
  - TD (<35)
  - TS (35–>65)
  - TY (>65<100)
  - TY (>100)
  - STY (>130)

- EPAC
  - TD (<35)
  - TS (35–>65)
  - TY (>65<100)
  - TY (>100)
  - STY (>130)

- WPAC
  - TD (<35)
  - TS (35–>65)
  - TY (>65<100)
  - TY (>100)
  - STY (>130)

- NIO
  - TD (<35)
  - TS (35–>65)
  - TY (>65<100)
  - TY (>100)
  - STY (>130)

- SIO
  - TD (<35)
  - TS (35–>65)
  - TY (>65<100)
  - TY (>100)
  - STY (>130)

- SWPAC
  - TD (<35)
  - TS (35–>65)
  - TY (>65<100)
  - TY (>100)
  - STY (>130)

Dr. W. Fiorino, NOAA ESRL/GDS/AMR, Boulder, CO
~\tc.oclspec.2014010100.2015010100.BT.final.png
2014-04-30 09:46
SHEM 2014 season – WPAC 2014 so far…

**SHEM TC Activity sACE index (sACEd ~ sT Cd) for: 20130701–20140514**

sACEd=ACE scaled by 1/(4(6h/1d)+65kt+5kt ~ 1 sT Cd
ACE = sum of Vmax+Vmax every 6h if Vmax>=35kt climo: 1981–2010

**WESTPAC TC Activity sACE index (sACEd ~ sT Cd) for: 20140101–20140430**

sACEd=ACE scaled by 1/(4(6h/1d)+65kt+5kt ~ 1 sT Cd
ACE = sum of Vmax+Vmax every 6h if Vmax>=35kt climo: 1981–2010
“You’re only as good as what you measure”

CAPT Vic Addison USN(ret), FLENUMMETOCCEN
departing officer Captain’s call in May 2006
Motivations – 2

as Roseanne Roseannadanna reminds us “...it's always something – if it ain't one thing, it's another”

• informal intercomparison of stats/errors with NHC & JTWC & EMC revealed:
  ‣ bugs in the codes...thank you James
  ‣ 0.1-1.0 nmi diff in mean depending on position error calc
  ‣ case selection can make a 5-20% diff in the mean
  ‣ hidden/implied filters
  ‣ significant diffs in tracker POD

• WMO 485 standard as with NWP field verification?
  ‣ “Verification Methods For Tropical Cyclone Forecasts”
  ‣ not really...
the bug you find today…

is ALWAYS the penultimate bug…

some small words of wisdom for son #2
rising sophomore
computer engineering major
Gonzaga U
### ESRL TCVC

processed ALL the NHC/JTWC/ECMWF adecks – the most complete set of TC trackers anywhere

<table>
<thead>
<tr>
<th>when</th>
<th>where</th>
<th>what</th>
</tr>
</thead>
<tbody>
<tr>
<td>1976-77</td>
<td>PSU</td>
<td>.f – TC NWP forecasts with MM0.0</td>
</tr>
<tr>
<td>1980-87</td>
<td>‘Monterey’ = NRL, FNMOC, NPS</td>
<td>.f – TC operational and research models</td>
</tr>
<tr>
<td>1998-1999</td>
<td>ECWMF</td>
<td>.pl .gs – ERA40 &amp; HRES</td>
</tr>
<tr>
<td>2000-2005</td>
<td>JTWC</td>
<td>.py flat-file DB – operational and in-house trackers</td>
</tr>
<tr>
<td>2006-2008</td>
<td>NHC</td>
<td>.py flat-file DB – operational and in-house trackers</td>
</tr>
<tr>
<td>2009-2014</td>
<td>ESRL</td>
<td>.py .obj hash DB – operational and in-house trackers</td>
</tr>
</tbody>
</table>

#### data types – ATCF

<table>
<thead>
<tr>
<th>adeck</th>
<th>bdeck</th>
<th>mdeck.py</th>
<th>vdeck.py</th>
<th>adeck.py</th>
</tr>
</thead>
<tbody>
<tr>
<td>forecast aid trackers</td>
<td>best track</td>
<td>merge deck combines TC info in adeck &amp; bdeck into one place – all storm info, e.g., TDO/HS initials</td>
<td>verification vars PE, IE, CTE, ATE, NICK, track length… key is model_storm</td>
<td>.py obj form of adeck includes mdeck.py input to making vdeck.py</td>
</tr>
<tr>
<td>posit + R? + …</td>
<td>working or final</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CARQ = TCvitals or ‘compute’ – initialize trackers</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

#### data sets

<table>
<thead>
<tr>
<th>NHC</th>
<th>JTWC</th>
<th>NHC/JTWC 9X</th>
<th>ECMWF all since 2006 + tracking ERA-40 fc</th>
<th>local trackers for all global models since 2006</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
sources of differences...and there are more...

• tracker and tracker settings
  ‣ **TIM** tracker (**TC In** Models) – compiler/machine sensitivity
  ‣ input model fields – grid resolution – ECMWF tracker (full res) v EMX (1 deg grid)
  ‣ TCvitals source – initialization sensitivity
  ‣ tracking weak systems & dissipation
  ‣ tracker POD – does tracker ‘cover’ all verifying posits in the best track

• calculation of errors
  ‣ great circle distance – radius of earth & formulae
  ‣ primary rule in NWP – remove ALL known errors regardless of size...

• bugs in tracker/verification code...

• **verification rule/conditions**
  ‣ NHC/JTWC – if it’s a TC initially and at the verifying forecast tau – **VERIFY**
  ‣ filter options
    • Vmax >= 25 kt?
    • if speed > 50 kt in tropics do not calc errors?
    • TC in a warning/advisory status?
Example of the problem...comp with EMC


Performance of Operational HWRF for the 2013 Western Pacific Basin

HWRF RI POD skill is ~23\% and by far has higher POD index as compared to other models and in other basins (previous analysis of RI for WPAC from 2012 HWRF showed <30% skill).

-The POD index is much higher (43\%) if one simply considers the intensity change tendency, say 6-h change of VMAX > 5 kt.
Example of the problem...comp with EMC

**working best track – case selection**

- **NHC/JT rule**
  - N0:476 N120:93

- **remove posits over land > 24 h**
  - N0:474 N120:77

- **only verify posits in WARN status**
  - N0:476 N120:48
Example of the problem…comp with EMC

**working best track – case selection – COMMENTS**

- closest inter-model relationship with EMC stats is for the NHC/JT rule…but not in terms of # of cases
  - NHC/JT rule is if it’s a TC initial and a TC at the forecast tau – verify

- two filtering options:
  - remove all land points > 24 h after landfall
  - only verify possits in a ‘WARNING’ status (advisory @ NHC), i.e., is an operationally significant TC – doing homogeneous comps with JTWC only gets some of the effect
    - JT is making forecasts for non-significant possits…

- why diffs?
  - implicit and/or unstated filtering in the EMC code?
  - different adecks (ATCF-speak for forecast aids) and/or bdecks (best track)?
    - ESRL has direct access
    - the bdecks (and sometimes) the adecks do change
    - bdeck processing – detection of TC state
    - errors in data files? they happen more than one would hope for…QC?
TCVCIP – basic protocol

• TCVCIP will provide common a- and b-decks
• set case-selection rule
  ‣ NHC/JT rule
  ‣ TCVC must be able to detect/determine if a posit is a TC
• state/document implicit/explicit filtering rules
  ‣ ask participants to submit code for documentation purposes
Example of the problem...comp with EMC
effect on % improvement over HFIP baseline

Remove posits over land > 24 h
N0:474 N120:77

NHC/JT rule
N0:476 N120:93

Only verify posits in WARN status
N0:476 N120:48
• very big impact when verifying WARN posits v TC posits
  ‣ 10-12 % change at tau 72 h!!!

• moral of the story? there’s a BIG diff between the JT working best track and the final best track
  ‣ applying the WARN filter in EPAC/LANT did not cause the big changes seen in WPAC
Example of the problem...comp with EMC
do not verify USN models – effect on % improve

WARN only
N72:196 N120:60

WARN only
N72:170 N120:48

WARN only
proper comp to JTWC
N72:183 N120:52

TCVCIP – TC Verification Code Intercomparison Project
Mike Fiorino ESRL – HFIP telcon 20140514
• whenever JTWC/OFCL are verified against models, the model tracker MUST be interpolated in time for a VALID comparison, especially when assessing FORECAST value
  ‣ blog describes the issue in detail – in preparation for submission to WAF – includes stand-alone .py that works with standard ATCF adecks and outputs standard ATCF adecks…no excuses…
    http://wxmapstertc.blogspot.com/2013/12/dynamical-model-tc-verification.html

• models verified has a large effect on # of cases and thereby the means…

• 6-h interpolation of GFS and HWRF changes the comparison against JTWC from “models beating JT” to “JT beating the models”
• TCVCIP will provide common a- and b-decks
  ‣ from both JTWC and NHC
    • real data
    • WPAC/EPAC/LANT/IO/SHEM
  ‣ working and final best tracks

• set case-selection rule
  ‣ NHC/JT rule
  ‣ TCVC must be able to detect/determine if a posit is a TC

• state/document implicit/explicit filtering rules
  ‣ ask participants to submit code for documentation purposes

• a-b-decks will NOT have errors for basic test…

• ask to output position and intensity errors on a storm-by-storm basis and the means
TCVCIP next steps…

• set up the a-b-decks
  › only have a few aids in the adecks and will NOT include JTWC/OFCL
  › 2014 SHEM
  › 2013 in WPAC/EPAC/IO
  › 2012 in LANT

• run the ESRL TCVC to provide a baseline
  › ‘ls’ – listing utilities to dig into the details
  › tables of stats
    • by-storm basis
    • season

• invite participants
  › JTWC, NHC, EMC, GFDL, DTC, ECMWF, JMA, BOM.oz …others?
  › UKMO is considering

• ESRL TCVC: [http://sourceforge.net/projects/wxmap2/](http://sourceforge.net/projects/wxmap2/)