Tropical Cyclone Verification Code Intercomparison Project

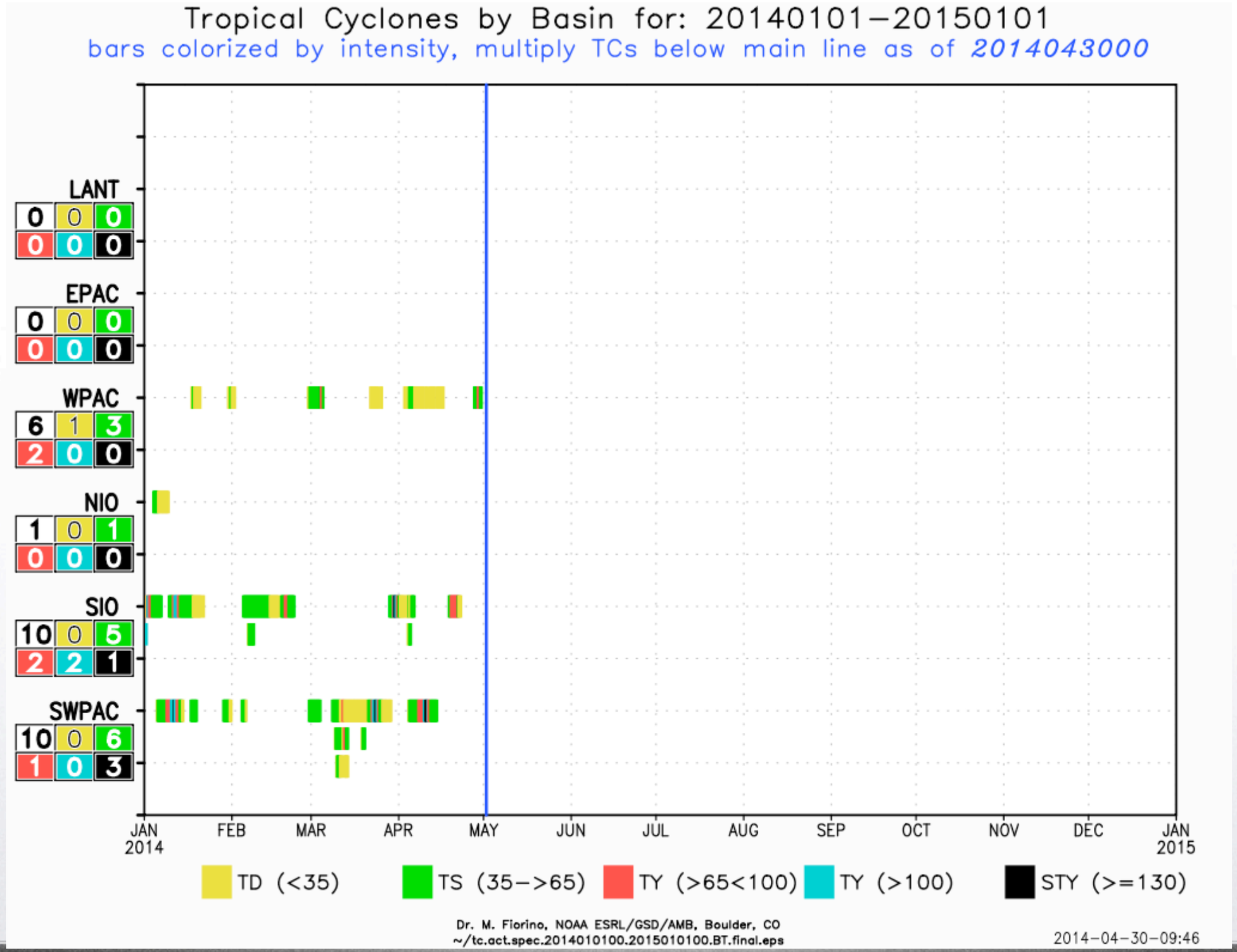
TCVCIP

pronounced "tee see vee sip"

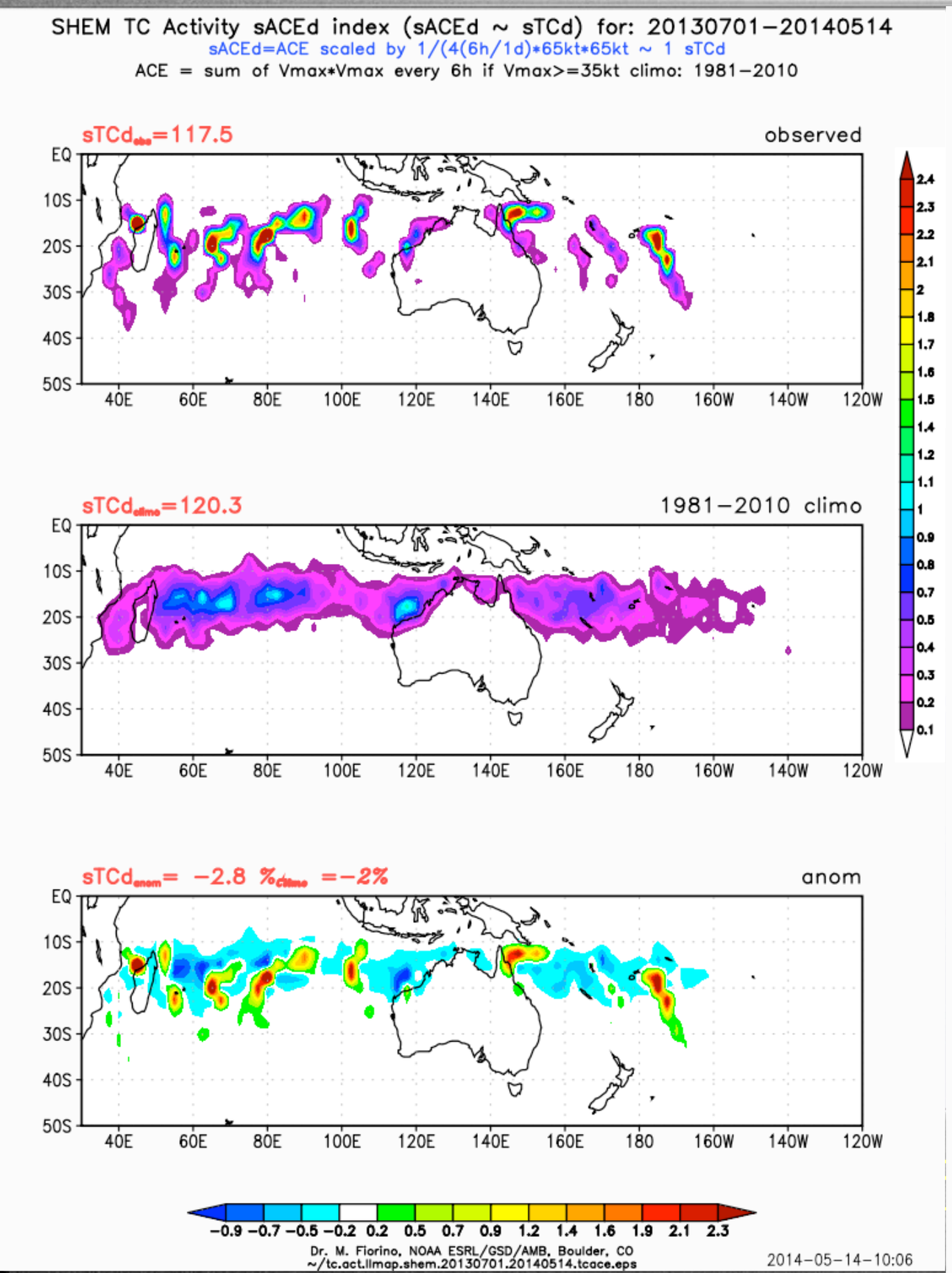
Mike Fiorino
michael.fiorino@noaa.gov
NOAA ESRL Boulder CO
14 May 2014

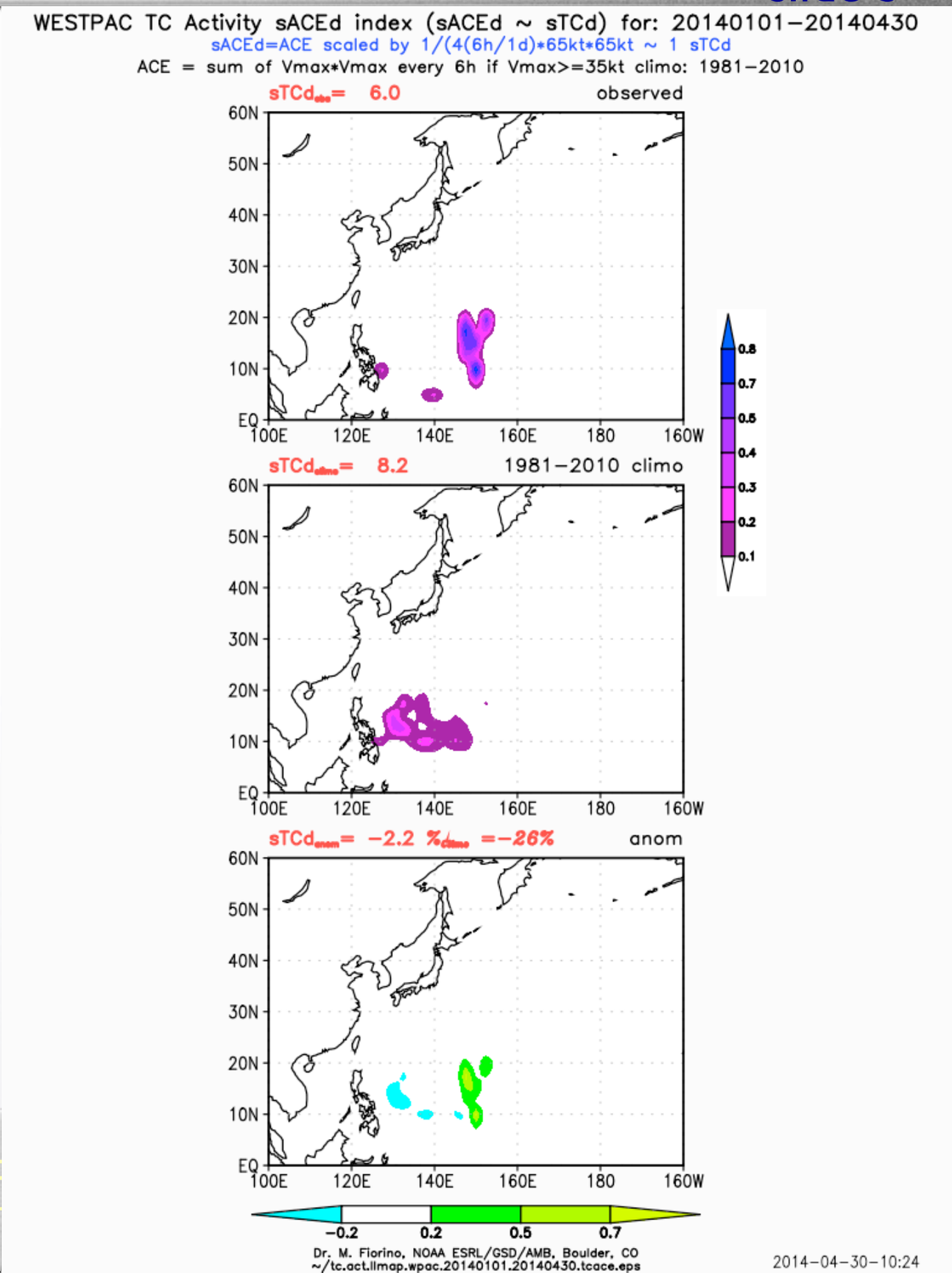
- motivations…
- why me? – doing TC verification code since 1977 – AMIP I&II verification
- stat diffs between ESRL v EMC v NHC v JTWC…
- TCVCIP basic case(s) – PE using working and final best tracks
- forecast error = FE = f(PE,IE) – not today…

TC activity 2014



SHEM 2014 season – WPAC 2014 so far…





Motivations – I

prior to joining NHC in May 2006...on active duty at FNMOC

"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" slide 5

- 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" http://www.wmo.int/pages/prog/arep/wwrp/new/documents/TC_verification_Final_11Nov13.pdf
  - not really…

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 slide 7

when	where	what
1976-77	PSU	.f – TC NWP forecasts with MM0.0
1980-87	'Monterey' = NRL, FNMOC, NPS	.f – TC operational and research models
1988-1995	JTWC	.f – ATCF & BAM model
1998-1999	ECWMF	.pl .gs – ERA40 & HRES
2000-2005	JTWC	.py flat-file DB – operational and in-house trackers
2006-2008	NHC	.py flat-file DB – operational and in-house trackers
2009-2014	ESRL	.py .obj hash DB – operational and in-house trackers

data types – ATCF

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

data sets

NHC	JTWC	NHC/JTWC 9X	ECMWF all since 2006 + tracking ERA-40 fc	local trackers for all global models since 2006	

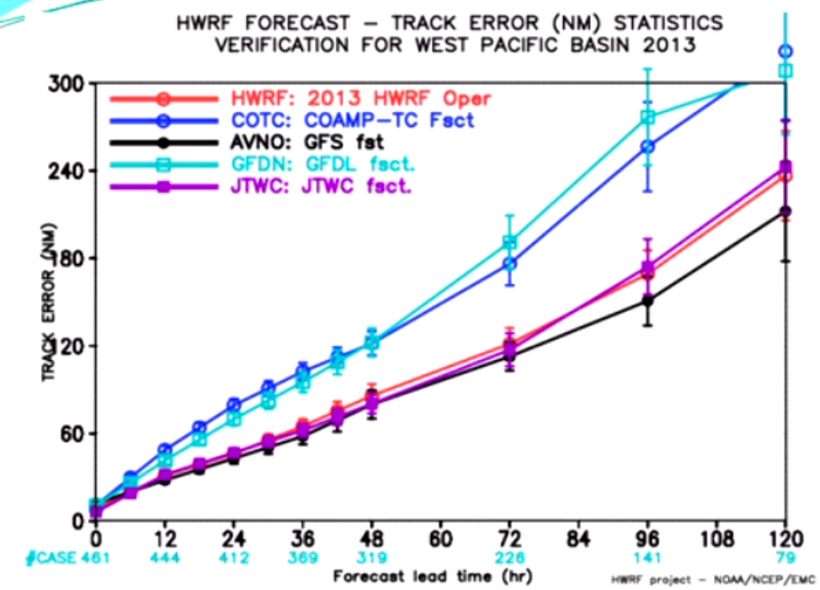
sources of differences…and there are more…

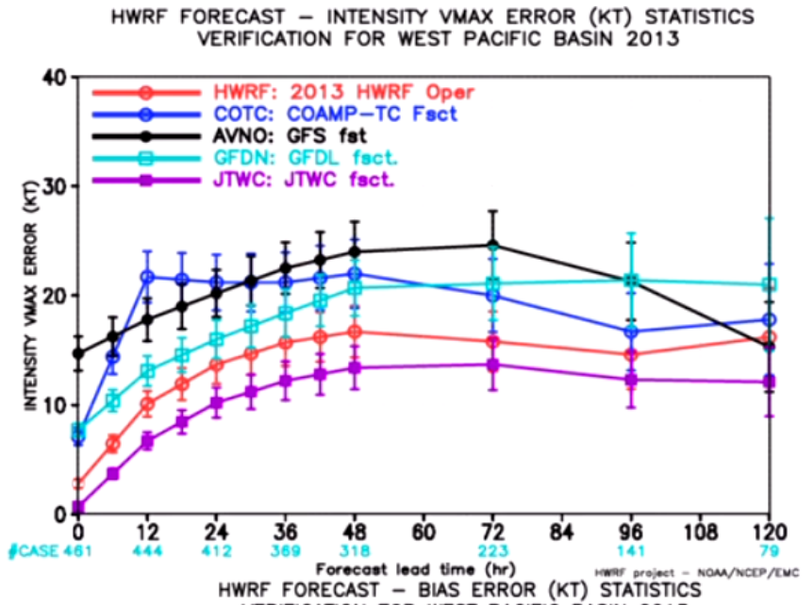
- tracker and tracker settings
  - **TIM** tracker (**T**C **I**n **M**odels) – compiler/machine sensitivity
  - input model fields – grid resolution – ECMWF tracker (full res) v EMX (1 deg grid)
  - TCvitals source – initialization sensitivity
  - tracking weak systems & dissipation
  - tracker POD – 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

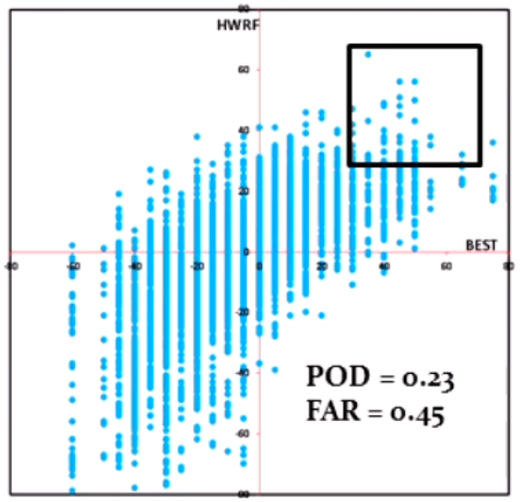
Vijay Tallapragada, NOAA/NWS/NCEP/EMC, College Park, MD; and S. Trahan, Y. C. Kwon, Z. Zhang, C. Kieu, Q. Liu, W. Wang, M. Tong, D. Sheinin, E. Liu, B. Zhang, S. Gopalakrishnan, X. Zhang, L. R. Bernardet, R. M. Yablonsky, J. W. Bao, R. J. Pasch, J. L. Franklin, D. A. Zelinsky, B. Strahl, W. Lapenta, R. L. Gall, and F. Toepfer

Performance of Operational HWRF for the 2013 Western Pacific Basin



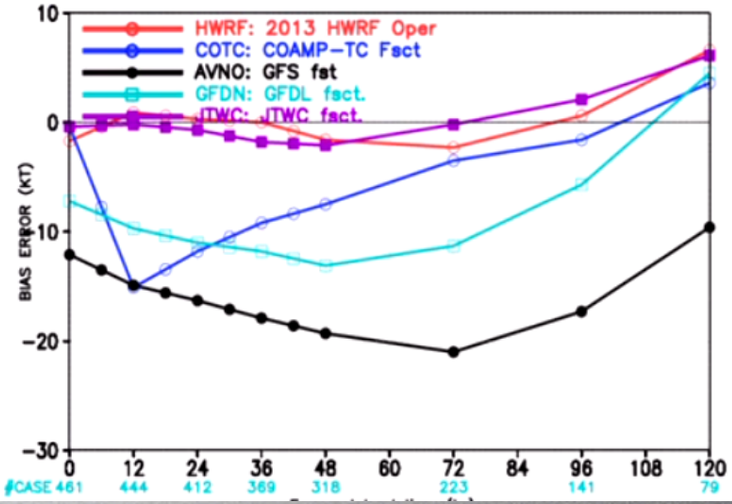


RI verification (VMAX change > 30 kt/24h)



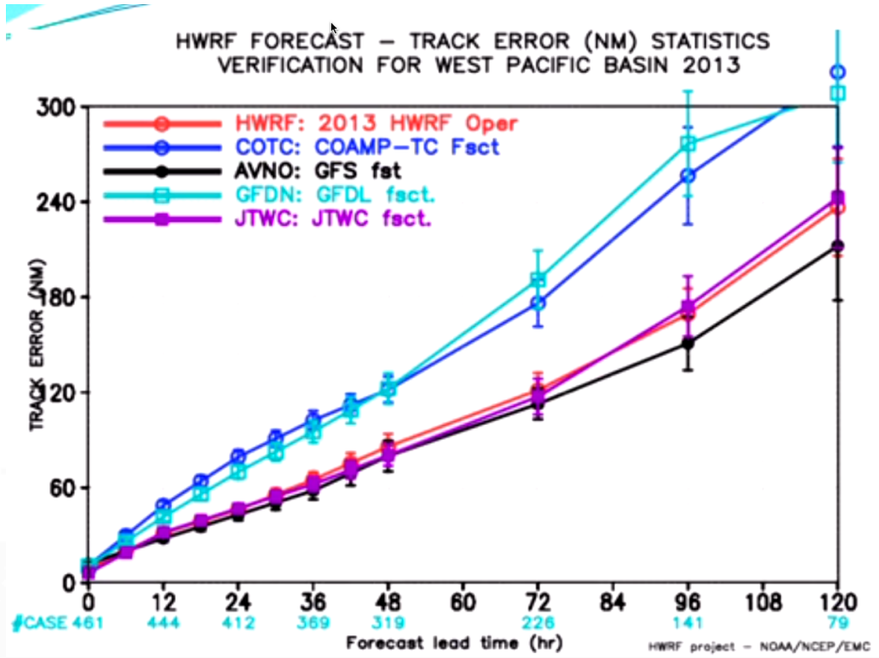
HWRF RI POD skill is ~ 23 % and by far has higher POD index as compared to other models and in other basins ***(previous analysis of RI for WPAC from 2012 HWRF showed <10% skill).***

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

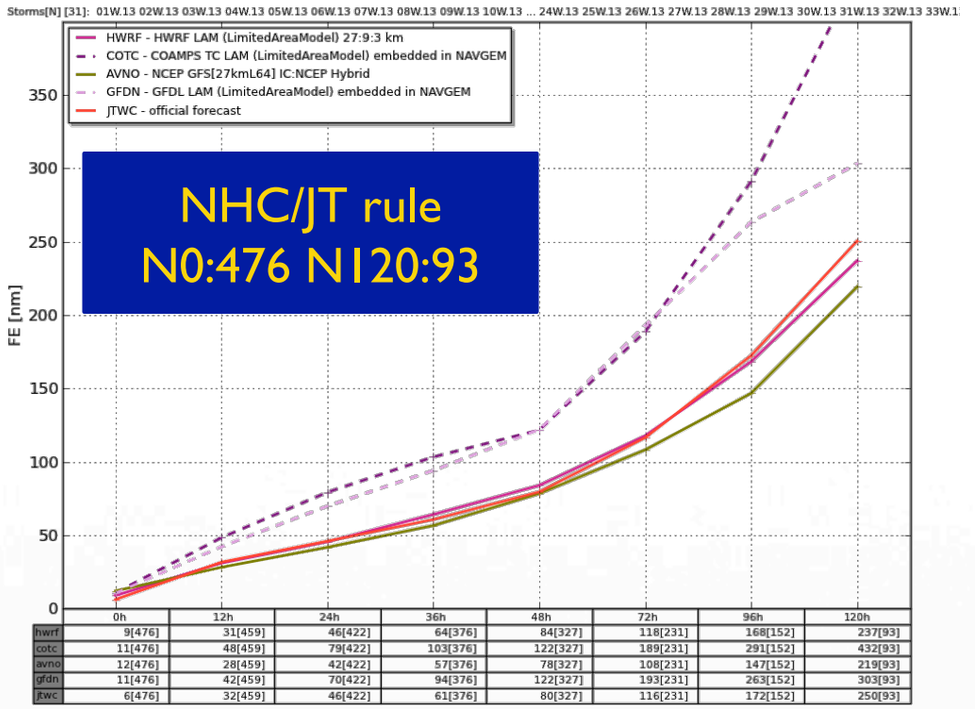


Example of the problem…comp with EMC

working best track – case selection

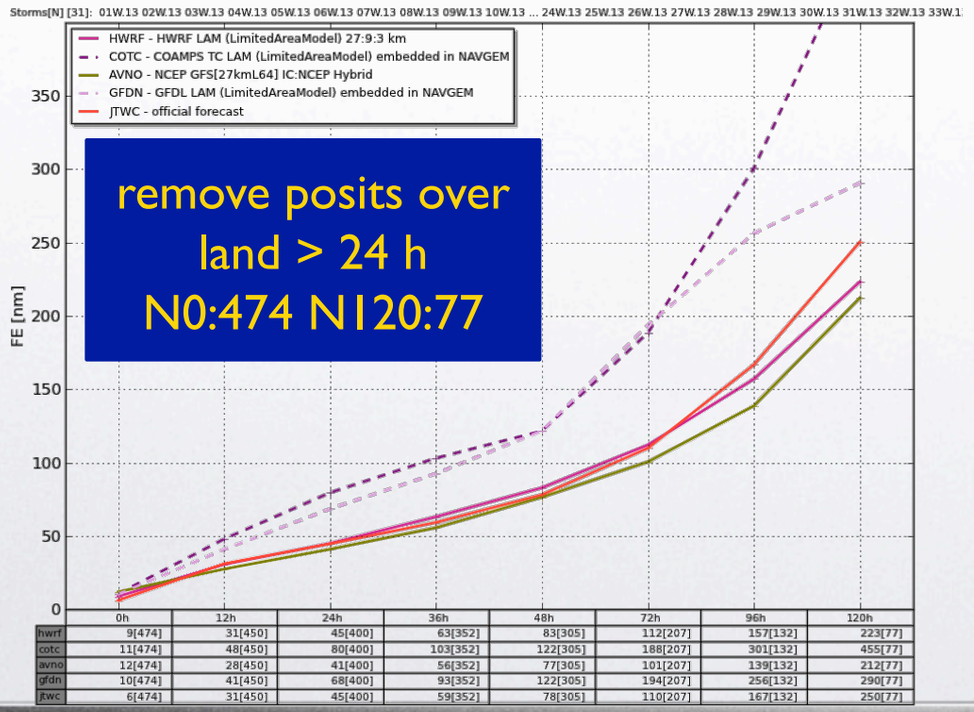


WPAC 2013 - ESRL TCVC - Working Best Track - NHC/JTWC rule



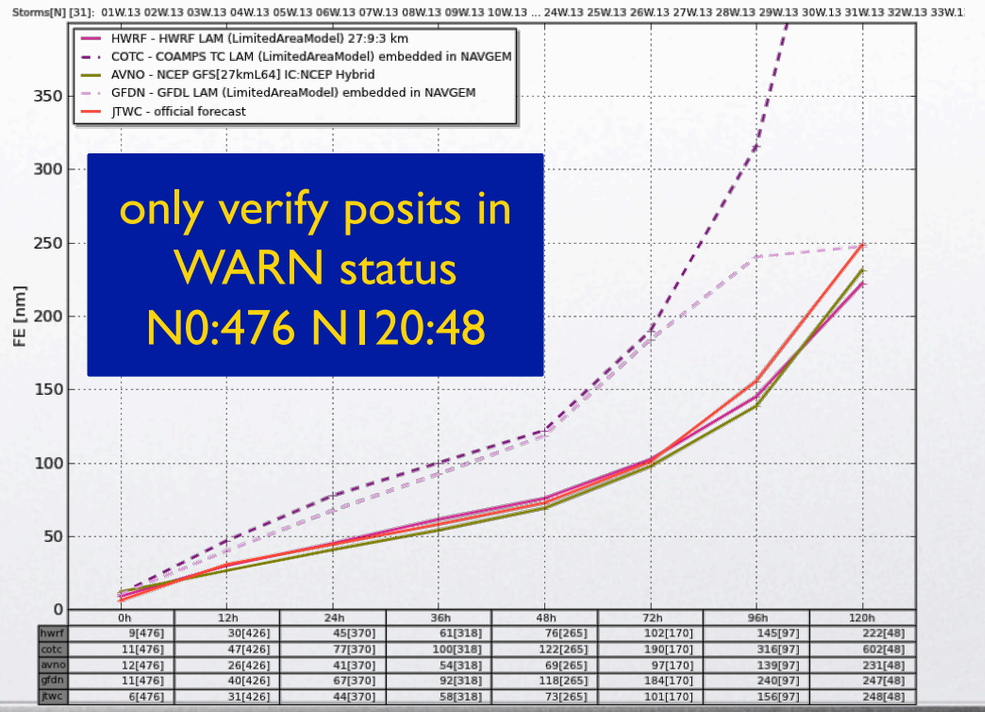
	0h	12h	24h	36h	48h	72h	96h	120h
hwrf	9[476]	31[459]	46[422]	64[376]	84[327]	118[231]	168[152]	237[93]
cotc	11[476]	48[459]	79[422]	103[376]	122[327]	189[231]	291[152]	432[93]
avno	12[476]	28[459]	42[422]	57[376]	78[327]	108[231]	147[152]	219[93]
gfdn	11[476]	42[459]	70[422]	94[376]	122[327]	193[231]	263[152]	303[93]
jtwc	6[476]	32[459]	46[422]	61[376]	80[327]	116[231]	172[152]	250[93]

WPAC 2013 - ESRL TCVC - Working Best Track - land < 24 h



	0h	12h	24h	36h	48h	72h	96h	120h
hwrf	9[474]	31[450]	45[400]	63[352]	83[305]	112[207]	157[132]	223[77]
cotc	11[474]	48[450]	80[400]	103[352]	122[305]	188[207]	301[132]	455[77]
avno	12[474]	28[450]	41[400]	56[352]	77[305]	101[207]	139[132]	212[77]
gfdn	10[474]	41[450]	68[400]	93[352]	122[305]	194[207]	256[132]	290[77]
jtwc	6[474]	31[450]	45[400]	59[352]	78[305]	110[207]	167[132]	250[77]

WPAC 2013 - ESRL TCVC - Working Best Track - WARN rule



	0h	12h	24h	36h	48h	72h	96h	120h
hwrf	9[476]	30[426]	45[370]	61[318]	76[265]	102[170]	145[97]	222[48]
cotc	11[476]	47[426]	77[370]	100[318]	122[265]	190[170]	316[97]	602[48]
avno	12[476]	26[426]	41[370]	54[318]	69[265]	97[170]	139[97]	231[48]
gfdn	11[476]	40[426]	67[370]	92[318]	118[265]	184[170]	240[97]	247[48]
jtwc	6[476]	31[426]	44[370]	58[318]	73[265]	101[170]	156[97]	248[48]

Example of the problem…comp with EMC

working best track – case selection -- COMMENTS

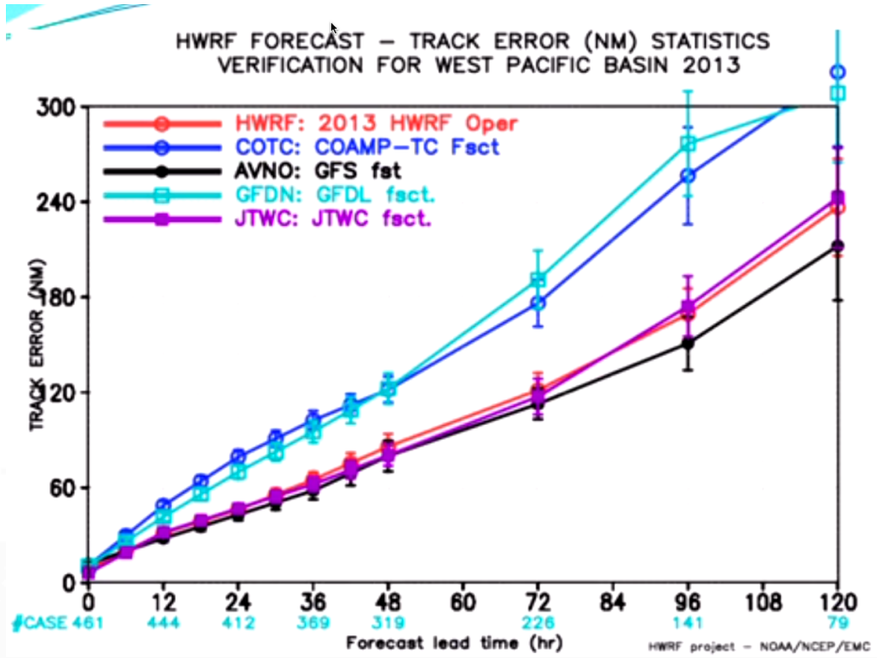
- closest inter-model relationship with EMC stats is for the NHC/JT rule…but not in terms of # of cases
  - NHC/JT rule is if it's a TC initial and a TC at the forecast tau – verify
- two filtering options:
  - remove all land points > 24 h after landfall
  - only verify posits in a 'WARNING' status (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 posits…
- 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

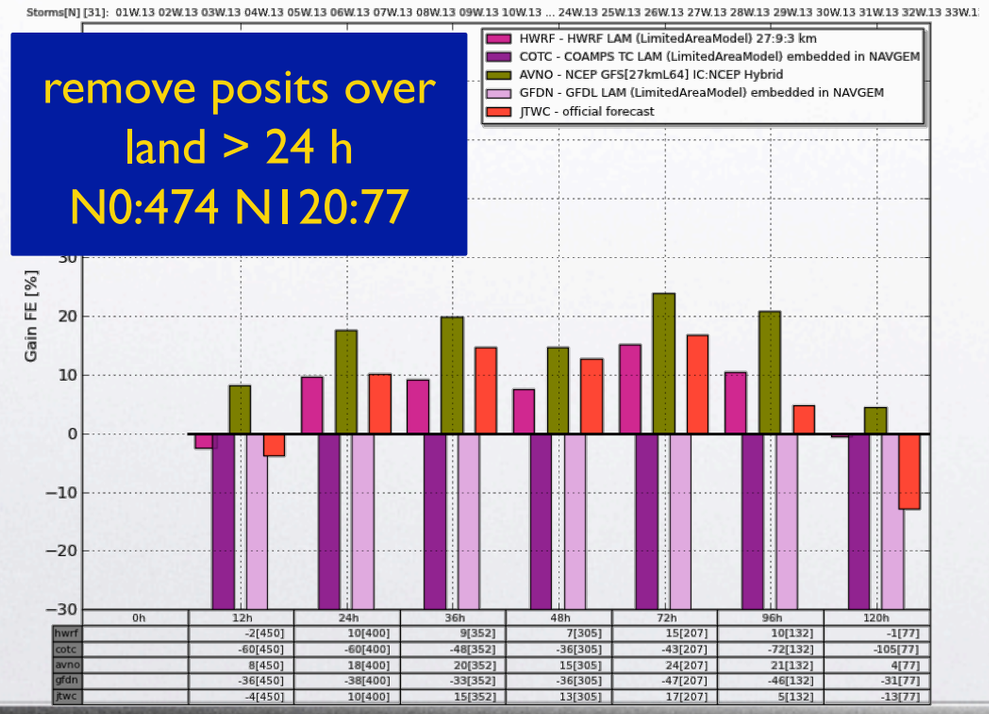
- TCVCIP will provide common a- and b-decks
- set case-selection rule
  - NHC/JT rule
  - TCVC must be able to detect/determine if a posit is a TC
- state/document implicit/explicit filtering rules
  - ask participants to submit code for documentation purposes

Example of the problem…comp with EMC

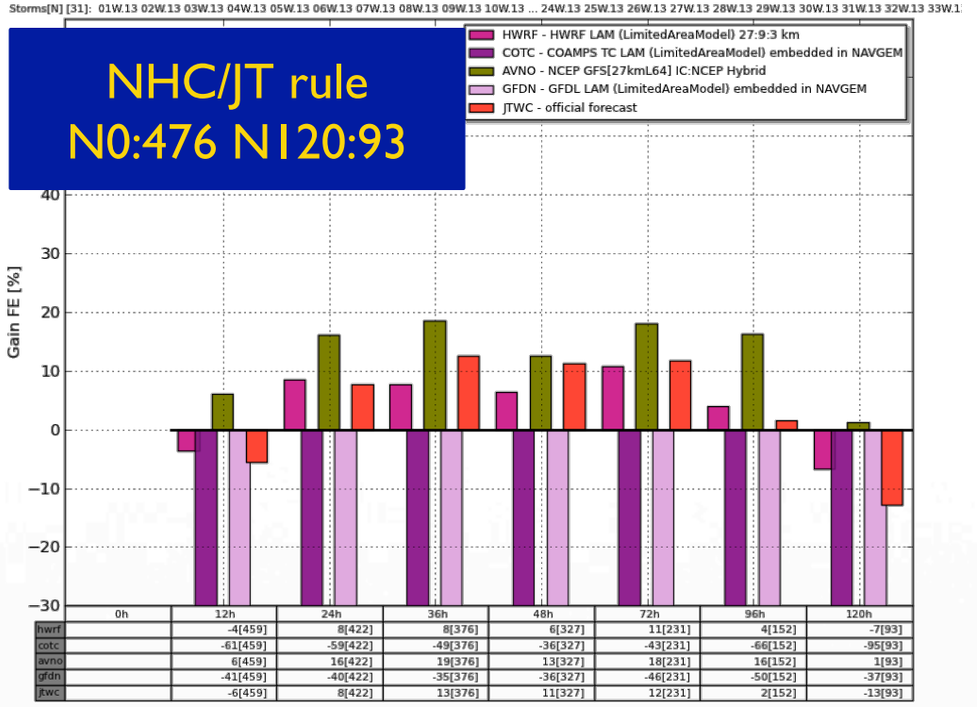
effect on % improvement over HFIP baseline



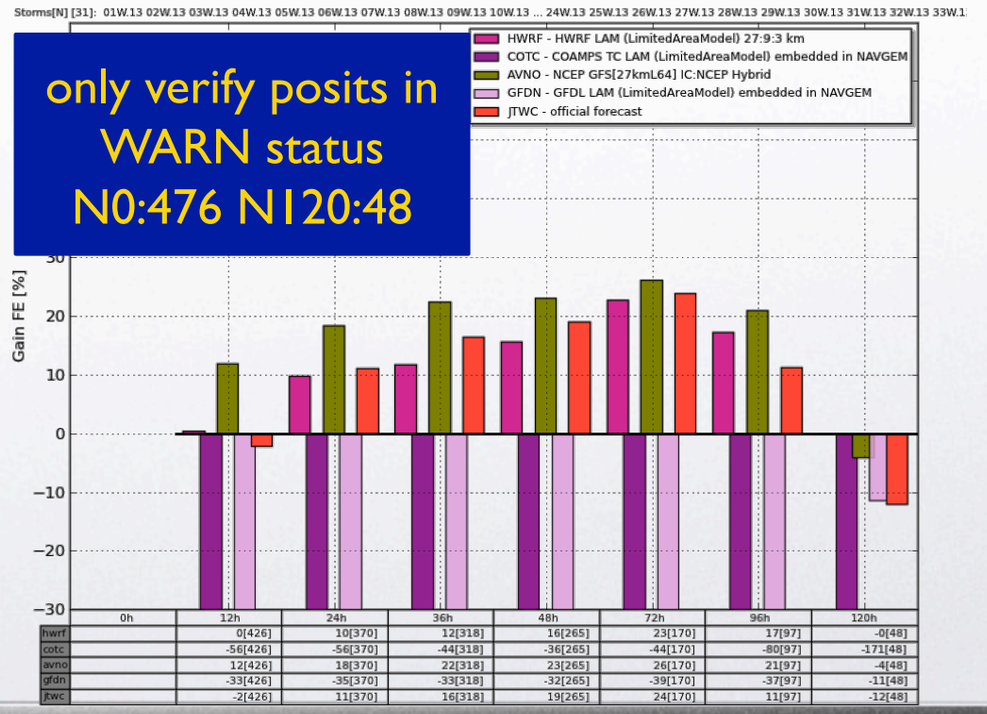
WPAC 2013 - ESRL TCVC - Working Best Track - land < 24 h



WPAC 2013 - ESRL TCVC - Working Best Track - NHC/JTWC rule



WPAC 2013 - ESRL TCVC - Working Best Track - WARN rule



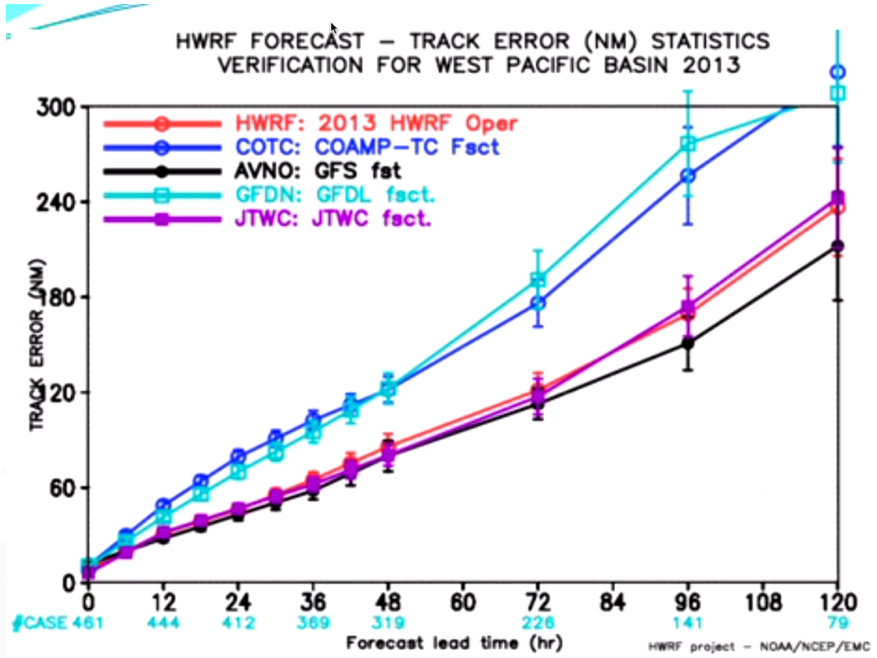
Example of the problem…comp with EMC

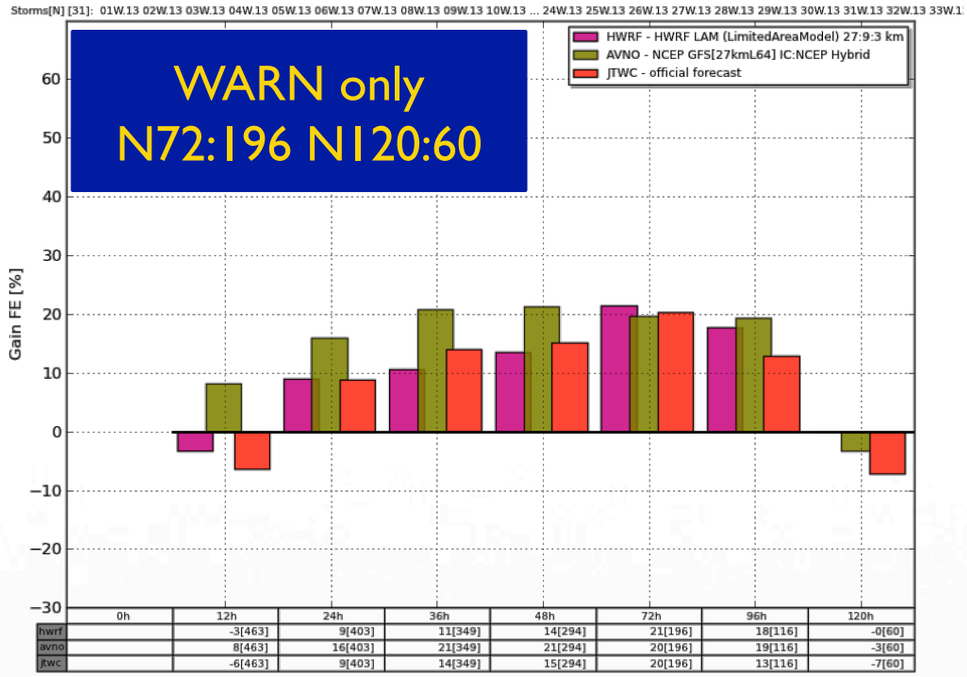
effect on % improvement over HFIP baseline – COMMENTS

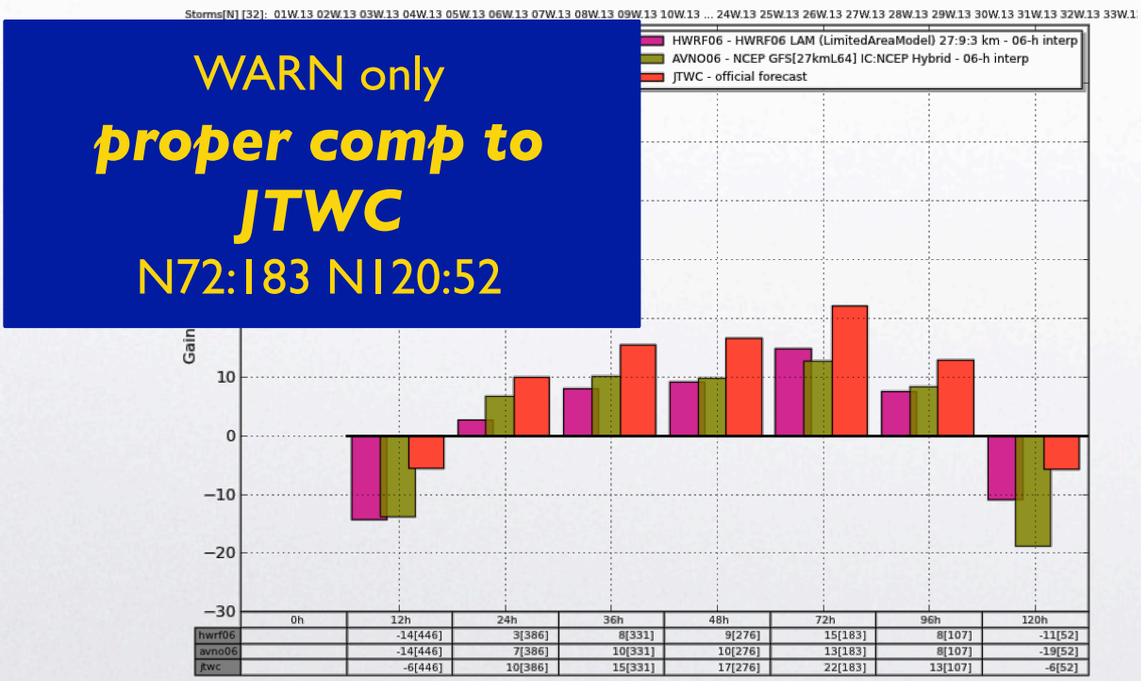
- very big impact when verifying WARN posits v TC posits
  - 10-12 % change at tau 72 h!!!
- moral of the story? there's a BIG diff between the JT working best track and the final best track
  - applying the WARN filter in EPAC/LANT did not cause the big changes seen in WPAC

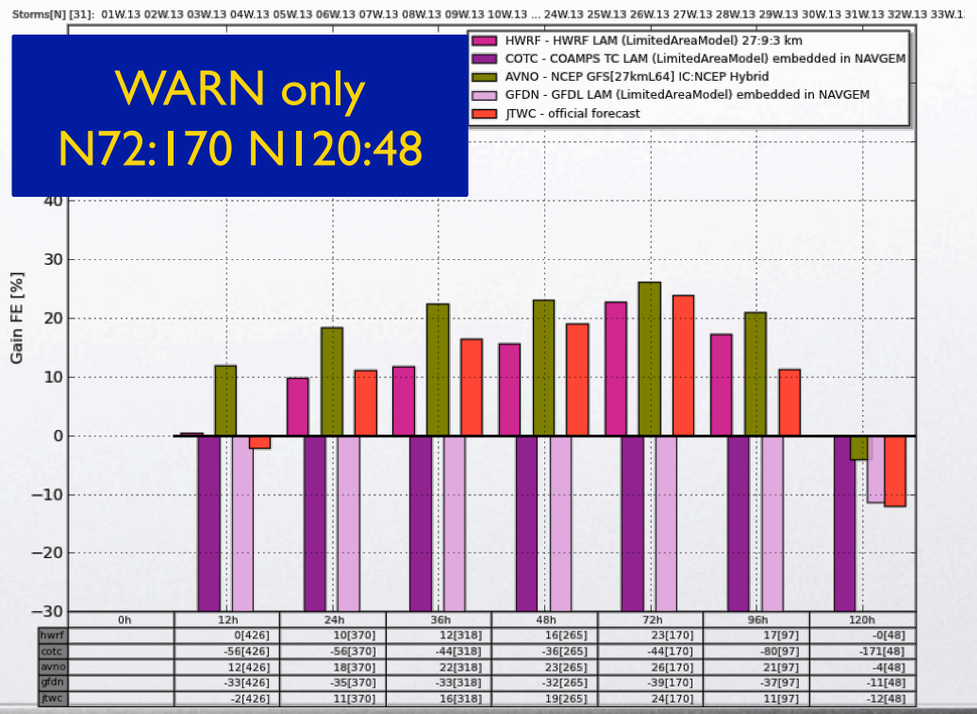
Example of the problem…comp with EMC

do not verify USN models – effect on % improve









Example of the problem…comp with EMC

do not verify USN models – effect on % improve -- COMMENTS

- 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 – test0

- 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/**