

HWT visit

Jamie was at HWT for the last week (5) of the testbed (31 May - 3 June)

Notes from Day 1 overview:

- CAPS radar assimilation ended up not running so there will be no GSD vs. CAPS comparison
- **Plan to continue controlled experiment (CLUE) in future HWT** - has been successful but also had a few "lessons learned" in this inaugural year that can be improved upon next year
- CAM hail size evaluation a focus for this year
 - Three hail algorithms: HAILCAST, direct output from the mp scheme (developed by G. Thompson), and machine learning (statistical) technique (Gagne)
- Ensemble sensitivity (Texas Tech work - Brian Ancell)
 - Features in flow early in the forecast that impact the ensemble response later (predictability)
- So far they have noticed if the CAMs don't handle overnight convection well they have problems the next day
 - There are a lot of solutions between the CAMs in this type of pattern
- Week 5 had a more of weaker shear/multi-cell storm pattern - this is a challenge for CAMs
- This blog entry is a good overview of what we did each day (<http://springexperiment.blogspot.com/2016/05/data-driven.html#more>)

Notes from sitting with forecasters each day

- opHRRR tends to have
 - PBL too warm/dry
 - too much convection
- parallel HRRR (going operational in early July)
 - has been decent during HWT
- 5-day MPAS
 - performance is region dependent
 - strongly forced systems easier
 - general temporal/spatial coverage OK but not specific storm location
- Thompson mp
 - less aggressive cold pools to slow propagation (this was an intentional design based on feedback from previous experiments)
 - see the result of this in the statistics
- **For verification (subjective during the experiment) they used LSRs (local storm reports), WFO warnings (especially in rural areas where no reports are received), and MESH (maximum estimated size of hail - MRMS hail product)**
 - Forecasters generally like the MESH - seems to be pretty accurate
- If we draw a 5% poly we would want 5 reports for each 100 grid box (at 80 km resolution) area
- General comment from HWT coordinators over the past 4 weeks
 - ARW (HRRR) generally has (incrementally) better performance than NAMRR - but on cases when NAMRR is better, it tends to be much better
- **When evaluating probabilities of 40 dBZ or greater they used reflectivity > 40 dBZ as the comparison field**
- In operations, NAM has poor sounding structure near convective initiation
- Forecasters need to be aware of the CWAs they issue
 - Don't want to change their poly just enough to include a CWA if it wasn't in there previously (unless warranted)
 - They joke that they could put so-and-so's house in a slight risk!
- **There is no reward to the forecaster for keeping the poly smaller (to reduce FAR) but they are punished if the area is too small and they miss reports**
 - Every bust makes them draw larger poly's
 - Only need a handful of reports to verify
- Don't care so much about FARs
- Hard to decrease probabilities once they are issued to the public ("Thou shalt not downgrade...")
 - They tend to err on the side of too low early on to avoid this problem
- How do you evaluate hail forecast if storms are in the wrong spot?!
- To start eh SFE2016 this post talks a bit about CLUE (<http://springexperiment.blogspot.com/2016/05/the-2016-spring-forecasting-experiment.html>); the final blog entry to wrapup SFE2016 is here (<http://springexperiment.blogspot.com/2016/06/sfe-2016-wrap-up.html>)

A few of the days they took ~ 2-5 minutes to show some objective statistics from the experiment

- **Aggregated ROC** for SFE2016 to-date (3-hrly ROC area by forecast lead time)
 - Assess mixed core vs. single core - In general, mixed (ARW+NMMB) beats core beats any single (ARW or NMMB) core; for single core, ARW generally beats NMMB
- When looking at **FSS**: mixed (ARW+NMMB) beats core beats any single (ARW or NMMB) core; for single core, NMMB generally beats ARW at shorter lead times and ARW beats NMMB at longer lead times
 - **When they compute FSS they do the following:**
 - Make obs 0/1 and apply smoother to get continuous values between 0-1 in obs
 - Apply a 40 km radius to forecast field
 - Difference forecast probabilities from the observations and look at the squared difference
- Does influence of DA extend longer when looking at probabilities rather than deterministic?
- **They compared PQPF to observations by using the same threshold for a single case**
- **This blog entry has an example of the ROC curves and PQPF comparison that we looked** (<http://springexperiment.blogspot.com/2016/05/clue-comparisons.html#more>). I can't seem to find a link to these plots on the testbed webpage (http://hwt.nssl.noaa.gov/Spring_2016/), however.

