**HWRF + COAMPS-TC + GFDL combined ensemble retrospective test**

Jon Moskaitis, Alex Reinecke, Jim Doyle: NRL-Monterey

**COAMPS-TC ensemble sample:**
- 185 cases from 10 TCs:
  1. Earl, 2010/07L
  2. Igor, 2010/11L
  3. Irene, 2011/09L
  4. Katia, 2011/12L
  5. Ernesto, 2012/05L
  6. Isaac, 2012/09L
  7. Nadine, 2012/14L
  8. Sandy, 2012/18L
  9. Gabrielle, 2013/07L
  10. Humberto, 2013/09L

**GFDL ensemble sample:**
- 716 cases from 2011-2013 HFIP stream 1.5 retro sample in Atlantic basins

**HWRF ensemble sample:**
- 520 cases from 2011-2013 HFIP stream 1.5 retro sample in Atlantic and Eastern Pacific basins

**HWRF + COAMPS-TC + GFDL sample:**
- 133 cases from 8 TCs:
  1. Irene, 2011/09L
  2. Katia, 2011/12L
  3. Ernesto, 2012/05L
  4. Isaac, 2012/09L
  5. Nadine, 2012/14L
  6. Sandy, 2012/18L
  7. Gabrielle, 2013/07L
  8. Humberto, 2013/09L
COAMPS-TC ensemble design

- 27-, 9-, 3-km horizontal grid spacing
- 1 control + 10 members with initial and boundary condition perturbations
- No physics perturbations
- No data assimilation

Control forecast:
- Initialized from the GFS analysis
- Vortex initialized with a Rankine vortex based on TC vitals

Ensemble members IC’s perturbed about the control:
- Synoptic perturbations drawn from static covariance (e.g. WRFVAR cv3)
- Vortex IC’s based on perturbed TC vitals
COAMPS-TC ensemble design

- **Synoptic perturbations:**
  - Perturbation drawn from a static covariance
  - Perturb the boundaries
  - Use WRFVAR cv3

- **Vortex scale perturbations:**
  - Vortex position, max wind, and RMW.

- Perturbation variance from:
  - Torn and Snyder 2012
  - Landsea and Franklin 2013

- Max wind and RMW covariance derived from 2001-2013 best track data.

- Variances and covariances depend on TC-vital max wind speed.
HWRF + COAMPS-TC + GFDL
Ensemble products

Example shown for particular Ernesto forecast, graphics for all 133 cases are available
Track forecast plot: Lines

HWRF + COAMPS-TC + GFDL, Ernesto (05L, 2012)

Number of ensemble members present at given lead time (out of possible 42)
Typically, COAMPS-TC has much greater track spread (2-3x) than GFDL or HWRF.
Track forecast plot: Ellipses

**Ellipses**

- **Inner and outer ellipses** contain 1/3 and 2/3 of ensemble members, respectively.

- Ellipses are centered on ensemble mean, with length and orientation of axes according to eigenvectors of ensemble position covariance matrix.

---

**TC = al052012, DTG = 2012080500**

- Black lines indicate the best track.
- Grey lines represent different model control runs.
- Blue ellipses represent HWRF forecasts.
- Green ellipses represent COAMPS-TC forecasts.
- Orange ellipses represent GFDL forecasts.

---

**Legend:**
- **Best track**
- **Ens. members**
- **HWRF control**
- **CTCX control**
- **GFDL control**
- **Ens. mean**

- **Colors and Time Stamps:**
  - Purple: 0 h (42)
  - Red: 24 h (42)
  - Orange: 48 h (42)
  - Green: 72 h (42)
  - Cyan: 96 h (42)
  - Blue: 120 h (42)

---

**Inner and outer ellipses contain 1/3 and 2/3 of ensemble members, respectively.**

**Ellipses are centered on ensemble mean, with length and orientation of axes according to eigenvectors of ensemble position covariance matrix.**
Typically, COAMPS-TC has much greater track spread (2-3x) than GFDL or HWRF.
Ellipse contains 2/3 of ensemble members, current lead time in black and older lead times in gray
Intensity and Min SLP

HWRF + COAMPS-TC + GFDL, Ernesto (05L, 2012)

Ensemble mean
Typically, COAMPS-TC has greater intensity spread than GFDL or HWRF.
Typically, COAMPS-TC has greater intensity spread than GFDL or HWRF.
Deterministic forecast validation

Homogeneous comparison of:
• COAMPS-TC control
• COAMPS-TC ensemble mean (9+ members)
• HWRF control
• HWRF ensemble mean (17+ members)
• GFDL control
• GFDL ensemble mean (8+ members)
• Combined ensemble (34+ members)
Track mean absolute error

HWRF, COAMPS-TC, and GFDL ensemble means perform similarly to their respective controls.

Combined ensemble mean has lower MAE than HWRF ensemble mean, COAMPS-TC ensemble mean and GFDL ensemble mean at all lead times.

Forecasts run every 12 h
Ensemble means have basically the same biases as the control members. It appears that the ensemble members are primarily making the same type of errors as the control. Forecasts run every 12 h.
The MAE and ME of the ensemble mean is generally lower than that of the control run.

The MAE of the combined ensemble mean is superior to the HWRF, COAMPS-TC, and GFDL ensemble means at nearly all lead times.
Probabilistic forecast diagnostics
**Average forecast spread**

- **Track**
  - COAMPS-TC
  - HWRF
  - GFDL
  - Combination

- **Intensity**
  - COAMPS-TC
  - HWRF
  - GFDL
  - Combination

**Number of members required for ensemble mean calculation**

- COAMPS-TC: 9 of 11
- HWRF: 17 of 21
- GFDL: 8 of 10
- Combination: 34 of 42
Ensemble spread vs. Error of ensemble mean

**COAMPS-TC Track**

- Average error of the ensemble mean
- Average spread of ensemble about its mean

**COAMPS-TC Intensity**

- Average error of the ensemble mean
- Average spread of ensemble about its mean

Forecasts run every 12 h
Ensemble spread vs. Error of ensemble mean

HWRF Track

HWRF Intensity

Forecasts run every 12 h
Ensemble spread vs. Error of ensemble mean

GFDL Track

GFDL Intensity

Sample size

Forecasts run every 12 h
Ensemble spread vs. Error of ensemble mean

Combined Track

Combined Intensity

Forecasts run every 12 h
Intensity rank histograms

Tau = 6 to 24

COAMPS-TC

GFDL

HWRF

Combined
Intensity rank histograms

COAMPS-TC

GFDL

HWRF

Combined
Intensity rank histograms

COAMPS-TC

GFDL

HWRF

Combined

Tau = 102 to 120
HWRF/COAMPS-TC/GFDL combined ensemble mean outperforms its component single model ensemble means in track MAE and intensity MAE.

For track, combined ensemble spread is similar to the MAE of the ensemble mean. However, for intensity, spread-skill plots and rank histograms show underdispersion.

Work is ongoing at NRL to apply additional probabilistic verification approaches to the individual and combined ensemble forecasts:

**Track reliability**: Does the observed position fall within the 2/3 probability ellipse for 2/3 of the forecast cases?

**Summary scores for intensity**: Continuous Ranked Probability Score, which facilitates comparison of probabilistic and deterministic forecasts.