NOAA’s multi-decadal global ensemble reforecast data set

Tom Hamill
NOAA ESRL, Physical Sciences Division
tom.hamill@noaa.gov

also: Jeff Whitaker, Gary Bates, Don Murray, Francisco Alvarez, Mike Fiorino, Tom Galarneau

Presentation to ensembles in geophysics conference, Toulouse, Nov 2012
Statistical post-processing for rare events is challenging without a large training sample.

Say you want to statistically post-process your model precipitation forecast to improve it. Heavy precipitation events like the one today are the ones you care about the most. How do you calibrate today’s forecast given past short sample of forecasts and observations?
2005 Rita official forecast (Houston, TX evacuated)

Is the model guidance for Rita biased too far east or west? Is the model guidance producing storms that are systematically less intense than they should be? Does the model spin up too many hurricanes in the Caribbean? How will you generate enough samples to know?
Regional simulations of past weather

Say you’d like to run a realistic, regional simulation of hurricane Rita at high resolution and not get the false skill that you will get if you force it with observed lateral boundary conditions. Where can you get lateral boundary conditions from a current-generation larger-scale forecast?
The atmosphere suppresses blocking subsequent to an active Indian-Ocean Madden-Julian Oscillation (MJO) during the Northern Hemisphere winter. Does the forecast model suppress blocking as well? How can one detect that with only a season or so of past forecasts and with both blocking and strong MJOs happening infrequently?
GEFS reforecast version 2 details

• Past forecasts using the currently operational GEFS, NOAA’s global ensemble forecast system.

• Each 00Z, 11-member forecast, 1 control + 10 perturbed.

• Reforecasts produced every day, for 1984120100 to current (actually, working on finishing late 2012 now).

• CFSR (NCEP’s Climate Forecast System Reanalysis) initial conditions (3D-Var) + ETR perturbations (cycled with 10 perturbed members). After ~ 22 May 2012, initial conditions from hybrid EnKF/3D-Var.

• Resolution: T254L42 to day 8, T190L42 from days 7.5 to day 16.

• Fast data archive at ESRL of 99 variables, 28 of which stored at original ~1/2-degree resolution during week 1. All stored at 1 degree. Also: mean and spread to be stored.

• Full archive at DOE/Lawrence Berkeley Lab, where data set was created under DOE grant.
Status

• 00Z reforecasts 1985-Sep 2012 completed and publicly available.

• Within a month or two, we will be pulling real-time GEFS data over from NCEP and putting it in our archive (hopefully within 12 h).

• Web sites are open to you now:
  – NOAA/ESRL site: fast access, limited data (99 fields).
  – US Department of Energy: slow access, but full data set

• Soon: experimental probabilistic precipitation forecast graphics over the US in real time.
Data that is readily available from ESRL

Table 1: Reforecast variables available for selected mandatory and other vertical levels. Φ indicates geopotential height, and an X indicates that this variable is available from the reforecast data set at 1-degree resolution; a Y indicates that the variable is available at the native ~0.5 degree resolution. AGL indicates “above ground level.”

<table>
<thead>
<tr>
<th>Vertical Level</th>
<th>U</th>
<th>V</th>
<th>T</th>
<th>Φ</th>
<th>q</th>
<th>Wind Power</th>
</tr>
</thead>
<tbody>
<tr>
<td>10 hPa</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>50 hPa</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>100 hPa</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>200 hPa</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>250 hPa</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>300 hPa</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>500 hPa</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>700 hPa</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>850 hPa</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>925 hPa</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>1000 hPa</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>σ = 0.996</td>
<td>X</td>
<td>X</td>
<td></td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>σ = 0.987</td>
<td>X</td>
<td>X</td>
<td></td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>σ = 0.977</td>
<td>X</td>
<td>X</td>
<td></td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>σ = 0.965</td>
<td>X</td>
<td>X</td>
<td></td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>80m AGL</td>
<td>X,Y</td>
<td>X,Y</td>
<td></td>
<td></td>
<td></td>
<td>X,Y</td>
</tr>
</tbody>
</table>

Also: hurricane track files
Data to be readily available from ESRL

Table 2: Single-level reforecast variables archived (and their units). Where an [Y] is displayed, this indicates that this variable is available at the native ~0.5-degree resolution as well as the 1-degree resolution.

<table>
<thead>
<tr>
<th>Variable (units)</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean sea-level pressure (Pa)</td>
<td>[Y]</td>
</tr>
<tr>
<td>Skin temperature (K)</td>
<td>[Y]</td>
</tr>
<tr>
<td>Soil temperature, 0.0 to 0.1 m depth (K)</td>
<td>[Y]</td>
</tr>
<tr>
<td>Volumetric soil moisture content 0.0 to 0.1 m depth (fraction between wilting and saturation)</td>
<td>[Y]</td>
</tr>
<tr>
<td>Water equivalent of accumulated snow depth (kg m⁻², i.e., mm)</td>
<td>[Y]</td>
</tr>
<tr>
<td>2-meter temperature (K)</td>
<td>[Y]</td>
</tr>
<tr>
<td>2-meter specific humidity (kg kg⁻¹ dry air)</td>
<td>[Y]</td>
</tr>
<tr>
<td>Maximum temperature (K)</td>
<td>[Y]</td>
</tr>
<tr>
<td>Minimum temperature (K)</td>
<td>[Y]</td>
</tr>
<tr>
<td>10-m u wind component (ms⁻¹)</td>
<td>[Y]</td>
</tr>
<tr>
<td>10-m v wind component (ms⁻¹)</td>
<td>[Y]</td>
</tr>
<tr>
<td>Total precipitation (kg m⁻², i.e., mm)</td>
<td>[Y]</td>
</tr>
<tr>
<td>Water runoff (kg m⁻², i.e., mm)</td>
<td>[Y]</td>
</tr>
<tr>
<td>Average surface latent heat net flux (W m⁻²)</td>
<td>[Y]</td>
</tr>
<tr>
<td>Average sensible heat net flux (W m⁻²)</td>
<td>[Y]</td>
</tr>
<tr>
<td>Average ground heat net flux (W m⁻²)</td>
<td>[Y]</td>
</tr>
<tr>
<td>Sunshine</td>
<td></td>
</tr>
<tr>
<td>Convective available potential energy (J kg⁻¹)</td>
<td>[Y]</td>
</tr>
<tr>
<td>Convective inhibition (J kg⁻¹)</td>
<td>[Y]</td>
</tr>
<tr>
<td>Precipitable water (kg m⁻², i.e., mm)</td>
<td>[Y]</td>
</tr>
<tr>
<td>Total-column integrated condensate (kg m⁻², i.e., mm)</td>
<td>[Y]</td>
</tr>
<tr>
<td>Total cloud cover (%)</td>
<td></td>
</tr>
<tr>
<td>Downward short-wave radiation flux at the surface (W m⁻²)</td>
<td>[Y]</td>
</tr>
<tr>
<td>Downward long-wave radiation flux at the surface (W m⁻²)</td>
<td>[Y]</td>
</tr>
<tr>
<td>Upward short-wave radiation flux at the surface (W m⁻²)</td>
<td>[Y]</td>
</tr>
<tr>
<td>Upward long-wave radiation flux at the surface (W m⁻²)</td>
<td>[Y]</td>
</tr>
<tr>
<td>Potential vorticity on θ = 320K isentropic surface (K m² kg⁻¹ s⁻¹)</td>
<td></td>
</tr>
<tr>
<td>U component on 2 PVU (1 PVU = 1 K m² kg⁻¹ s⁻¹) isentropic surface (ms⁻¹)</td>
<td></td>
</tr>
<tr>
<td>V component on 2 PVU isentropic surface (ms⁻¹)</td>
<td></td>
</tr>
<tr>
<td>Temperature on 2 PVU isentropic surface</td>
<td></td>
</tr>
<tr>
<td>Pressure on 2 PVU isentropic surface</td>
<td></td>
</tr>
</tbody>
</table>
esrl.noaa.gov/psd/forecasts/reforecast2/download.html

Produces netCDF files.

Also: direct ftp access to allow you to read the raw grib files.
This DOE site will be ready for access to tape storage of full data (slower).

Use this to access full model state.
Characteristics of the unprocessed GEFS reforecasts.

How stationary are the errors?
500 hPa Z anomaly correlation
(from deterministic control)

500 hPa geopotential height anomaly correlation from reforecasts

Lines w/o filled colors for second–generation reforecast (2012, T254)


Perhaps a 1.5-2.5 day improvement.
Tropical cyclone track error and spread

Reforecast Ens. Mean Error and Spread

Ensemble–mean Position Error & Spread (km)

Forecast Lead (Days)
Statistical post-processing using reforecasts
Statistical post-processing of precipitation forecasts

Reliability, Day +3 10.0mm

This is data from Jul-Oct 2010, when the GEFS was T190.

Probabilities directly estimated from ensemble prediction systems are often unreliable.

Can we statistically post-process the current GEFS using reforecasts and improve reliability and skill?
Almost perfect reliability possible with very simple calibration algorithm.

Statistical post-processing method used was “rank analog” technique discussed in Whitaker and Hamill (2006 MWR) and Hamill et al. (2012, BAMS, submitted).
Skill of calibrated precipitation forecasts
(over US, 1985-2010, “rank analog” calibration method)

Verification here against 32-km North American Regional Reanalysis (tougher).
Verification in previous plot against 1-degree NCEP precipitation analysis (easier).
Other statistical post-processing work in progress

(a) Mean 120 to 240-h forecast wind speed, VT=2010011100

(b) Quantile of mean 120 to 240-h forecast wind speed, VT=2010011100

(c) CFSR analyzed average wind speed, VT=2010011100

(d) Quantile of analyzed wind speed, VT=2010011100

Say you don’t have observational or analysis data widely available for statistical post-processing. How can you leverage reforecasts to tell you whether or not today’s weather is unusual?

Here’s an example quantifying how unusual the forecast wind speed is relative to past model forecasts of wind speed for a similar time of the year.

This might be useful for making decisions for wind energy, for example.
Application: extended-range tornado forecasting

4/11/1996 Forecast, 204-hour through 276-hour leadtime
Using 3 PCs of 0-6 km Shear, log(CAPE) & Conv.Precip. as Predictors for Logistic Regression
Probability of tornado (>EF0) event

Francisco Alvarez, St. Louis University, is working with me and others on using the reforecasts to make extended-range predictions of tornado probabilities.

Ph.D. work, in progress.
Application: Improved 6-10 day and week-2 forecast guidance

Dan Collins of NCEP Climate Prediction Center leading this effort.
Regional reforecast initialization

Here, Hurricane WRF for Rita.

c/o Tom Galarneau
A synthetic example of using reforecasts to make track error bias corrections

72-h Forecast Verifying 1200 UTC 9 September

Red 🎈: mean forecast position
Blue dot: forecast positions of +72-h forecast analogs
End of red tail ___ : observed positions at +72 h

c/o Tom Galarneau, NCAR.
Diagnosis of model errors associated with infrequent phenomena

Example: atmospheric blocking, the Madden-Julian Oscillation, and their interaction
Under-forecasting of Atlantic block frequency after day +3
Change in blocking frequency under strong Indian Ocean MJO

Shaded areas are confidence 5/95% confidence intervals.

Suppression of blocking frequency in the east Pacific and Atlantic under strong MJO. Day +6 GEFS nicely replicates this suppression.
Conclusion

• GEFS reforecast data is now freely available for your use.
  – Fast archive of common variables.
  – Slower tape archive of full model state.

• We hope this will spur research in advanced statistical post-processing and facilitate understanding of GEFS model errors, facilitate regional reforecasts.
Supplementary slides
Basic analog technique for statistical downscaling (here, v1)

Today’s ens. mean forecast + subsequent analyzed precipitation

On the left are old forecasts similar to today’s ensemble-mean forecast. For making probabilistic forecasts, form an ensemble from the accompanying analyzed weather on the right-hand side.
Analog technique for statistical downscaling

Form an ensemble from these
Problem with basic analog reforecast technique

Say today’s forecast is for 20 mm. There are more forecasts slightly less than 20 mm than slightly more than 20 mm.

Assuming correlation between forecast and observations, analogs will be biased toward lower precipitation amounts.
“Rank” analog procedure

• Convert precipitation forecast time series to ranks:

\[ x = [0, 0, 7, 15, 1, 3, 6, 4, 1, 2, 12, 5, 6, 8] \]

\[ x(r) = [1.5, 1.5, 11, 14, 3.5, 6, 9, 7, 3.5, 5, 13, 8, 10, 12] \]
“Rank” analog procedure

• Convert precipitation forecast time series to ranks:

\[
x = [0, 0, 7, 15, 1, 3, 6, 4, 1, 2, 12, 5, 6, 8]
\]

\[
x(r) = [1.5, 1.5, 11, 14, 3.5, 6, 9, 7, 3.5, 5, 13, 8, 10, 12]
\]

with standard analog, these would be the two forecasts with the closest values.
“Rank” analog procedure

• Convert precipitation forecast time series to ranks:

\[
x = [0, 0, 7, 15, 1, 3, 6, 4, 1, 2, 12, 5, 6, 8]
\]
\[
x(r) = [1.5, 1.5, 11, 14, 3.5, 6, 9, 7, 3.5, 5, 13, 8, 10, 12]
\]

with rank analog, these would be the two forecasts with the closest ranks.
Rank analog calibration details

• 24-h accumulated precipitation, validated on NARR grid (~32 km) over CONUS, 1985-2009.

• Rank analog approach: at each grid point in CONUS, using that grid point and +/- 3 surrounding grid points in N-S, E-W direction, find dates of 75 past forecasts that are closest in average precipitation rank of ensemble mean forecast. Make probabilistic forecasts from analyzed NARR precipitation data on dates of those 75 analogs.

• (Conventionally calculated) Brier Skill Scores, reliability diagrams, etc. NARR again used for verification.
Define BSS for evaluating blocking skill

• The blocking Brier Skill score is calculated after summing forecast and climatological Brier scores over the relevant longitudes in either the Pacific or Atlantic basins, respectively, then averaged. For example (Pac):

\[
BSS = 1.0 - \frac{BS_{\text{forecast}}}{BS_{\text{climo}}}
\]

\[
BS_{\text{forecast}} = \sum_{l_p=1}^{\text{nlon}} \sum_{i=1}^{\text{ndate}} \left( p_i^{\text{forecast}}(l_p) - o_i(l_p) \right)^2
\]

\[
BS_{\text{climo}} = \sum_{l_p=1}^{\text{nlon}} \sum_{i=1}^{\text{ndate}} \left( p_i^{\text{climo}}(l_p) - o_i(l_p) \right)^2
\]

\[
o_i(l_p) = \begin{cases} 1 & \text{if blocked} \\ 0 & \text{if unblocked} \end{cases}
\]

\[
p_i^{\text{forecast}}(l_p) = \text{ensemble – based probability of block for this longitude}
\]

\[
p_i^{\text{climo}}(l_p) = \text{climatological probability of block for this longitude}
\]
Computing the CRPSS of GEFS RMM1 and RMM2 forecasts

- CRPSS = 1 − CRPS(\text{forecast}) / CRPS(\text{climatology})

\[
\text{CRPS(\text{forecast})} = \sum_{i=1}^{\text{ndates}} \sum_{j=1}^{n\text{cats}} \frac{1}{n\text{cats}} \left( \Phi_{\text{forecast}}(i, x(j)) - \Phi_{\text{analyzed}}(i, x(j)) \right)^2
\]

\[
\text{CRPS(\text{climo})} = \sum_{i=1}^{\text{ndates}} \sum_{j=1}^{n\text{cats}} \frac{1}{n\text{cats}} \left( \Phi_{\text{climo}}(i, x(j)) - \Phi_{\text{analyzed}}(i, x(j)) \right)^2
\]

\[x(1) = -5.0, \quad x(2) = -4.9, \quad \ldots, \quad x(n\text{cats}) = +5.0\]

\[\Phi(\cdot) = \text{cumulative distribution function for either RMM1 or RMM2}\]

- \(\Phi(\cdot)\) estimated from normal distribution fit to sample mean and standard deviation.
The procedure we have applied is as follows: the 500 hPa field is firstly evaluated on a 4° by 4° regular latitude-longitude grid covering the Northern Hemisphere. Then the geopotential height gradients \( \text{GHGS} \) and \( \text{GHGN} \) (referring to middle and high latitudes respectively) are computed for each longitude point of the grid:

\[
\text{GHGS} = \frac{Z(\phi_o) - Z(\phi_s)}{(\phi_o - \phi_s)},
\]

\[
\text{GHGN} = \frac{Z(\phi_n) - Z(\phi_o)}{(\phi_n - \phi_o)},
\]

where

\[
\phi_n = 80° \text{N} + \Delta,
\]

\[
\phi_o = 60° \text{N} + \Delta,
\]

\[
\phi_s = 40° \text{N} + \Delta,
\]

\[
\Delta = 4°, 0° \text{ or } 4°.
\]

A given longitude is then defined as “blocked” at a specific instant in time if the following conditions are satisfied for at least one value of \( \Delta \):

1. \( \text{GHGS} > 0 \),
2. \( \text{GHGN} < -10 \text{ m/deg lat} \).

There are alternatives, such as PV-based index by Pelly and Hoskins. While these may have some advantages, this old standard used hereafter.
MJO deterministic verification metrics

\[
\text{COR}(\tau) = \frac{\sum_{i=1}^{N} [a_{1i}(t)b_{1i}(t) + a_{2i}(t)b_{2i}(t)]}{\sqrt{\sum_{i=1}^{N} [a_{1i}^2(t) + a_{2i}^2(t)]} \sqrt{\sum_{i=1}^{N} [b_{1i}^2(t) + b_{2i}^2(t)]}},
\]

where \( a_{1i}(t) \) and \( a_{2i}(t) \) are the observed RMM1 and RMM2 at day \( t \), and \( b_{1i}(t) \) and \( b_{2i}(t) \) are their respective forecasts, for the \( i \)th forecast with a \( \tau \)-day lead. Here, \( N \) is the number of forecasts.

\( \text{COR}(\tau) \) measures the skill in forecasting the phase of the MJO, which is insensitive to amplitude errors. \( \text{COR}(\tau) \) is equivalent to a spatial pattern correlation between the observations and the forecasts when they are expressed by the two leading combined EOFs.

\[
\text{RMSE}(\tau) = \sqrt{\frac{1}{N} \sum_{i=1}^{N} \left[ [a_{1i}(t) - b_{1i}(t)]^2 + [a_{2i}(t) - b_{2i}(t)]^2 \right]}.
\]

from Lin et al., Nov 2008 MWR.
Demo: Regional reforecast with WRF ARW v3.4 using global reforecast for initial, boundary conditions

• 2-way nested simulation 36-, 12- and 4-km with 36 vertical levels
  – 12- and 4-km moving nests
• Time step: 180, 60, and 20 s
• Initial and boundary condition: GFS reforecast ensemble member
• Tiedtke cumulus scheme on 36 and 12 km; explicit on 4 km
• YSU PBL scheme
• HYCOM ocean analysis
• WSM6 microphysics
• Noah land surface
• 2D Smagorinsky turbulence scheme
• Goddard shortwave radiation
• RRTM longwave radiation
• Second order diffusion
• Positive definite scalar advection
• Donelan wind-dependent drag formulation
• Garratt wind-dependent enthalpy surface fluxes

c/o Tom Galarneau, NCAR & NOAA HFIP grant
Bi-variate MJO RMM1 and RMM2 correlation and RMSE by half decade

The first 10 years are much less skillful than the subsequent 16.
(a) Composite 500 hPa geopotential height under block at Lon = 180E

(b) Composite 500 geopotential height under no block at Lon = 180E

Dec-Jan-Feb 1985-2010 CFSR data. Blocks defined here by Tibaldi/Molteni algorithm.
Blocking frequency
Dec-Jan-Feb

Grey bands define Euro/Atlantic and Pacific blocking sectors in subsequent plots.

Blocking frequency

Longitude (degrees)

0 30 60 90 120 150 180 210 240 270 300 330 360
GEFS blocking skill by half decade

Blocking is evaluated using Tibaldi-Molteni algorithm for every longitude, every day. Skill of the ensemble in predicting blocking for each longitude is then evaluated.

Decreased Atlantic sector skill in 1985-1989 period stands out.