Review of ensemble prediction fundamentals

Tom Hamill
NOAA ESRL, Physical Sciences Division
tom.hamill@noaa.gov
“Ensemble weather prediction”
Topics

• Chaos theory & ensembles
• Desired properties of ensembles
• Initializing ensembles
• Dealing with model error
  – mostly in Carolyn Reynolds’ talk
• Ensembles & hurricanes
• Some product ideas
“Chaos”

The Lorenz (1963) model

\[
\begin{align*}
\frac{dx}{dt} &= \sigma(y - x) \\
\frac{dy}{dt} &= x(\rho - z) - y \\
\frac{dz}{dt} &= xy - \beta z
\end{align*}
\]

\(\sigma, \rho, \beta\) are fixed.

Errors grow more quickly for some initial conditions than others.

from Tim Palmer’s 2006 book chapter
Initial conditions for “Lothar” ensemble forecasts

<table>
<thead>
<tr>
<th>Analysis 1</th>
<th>Analysis 2</th>
<th>Analysis 3</th>
<th>Analysis 4</th>
<th>Analysis 5</th>
<th>Analysis 6</th>
<th>Analysis 7</th>
<th>Analysis 8</th>
<th>Analysis 9</th>
<th>Analysis 10</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Analysis 11</td>
<td>Analysis 12</td>
<td>Analysis 13</td>
<td>Analysis 14</td>
<td>Analysis 15</td>
<td>Analysis 16</td>
<td>Analysis 17</td>
<td>Analysis 18</td>
<td>Analysis 19</td>
<td>Analysis 20</td>
</tr>
<tr>
<td>Analysis 21</td>
<td>Analysis 22</td>
<td>Analysis 23</td>
<td>Analysis 24</td>
<td>Analysis 25</td>
<td>Analysis 26</td>
<td>Analysis 27</td>
<td>Analysis 28</td>
<td>Analysis 29</td>
<td>Analysis 30</td>
</tr>
<tr>
<td>Analysis 31</td>
<td>Analysis 32</td>
<td>Analysis 33</td>
<td>Analysis 34</td>
<td>Analysis 35</td>
<td>Analysis 36</td>
<td>Analysis 37</td>
<td>Analysis 38</td>
<td>Analysis 39</td>
<td>Analysis 40</td>
</tr>
<tr>
<td>Analysis 41</td>
<td>Analysis 42</td>
<td>Analysis 43</td>
<td>Analysis 44</td>
<td>Analysis 45</td>
<td>Analysis 46</td>
<td>Analysis 47</td>
<td>Analysis 48</td>
<td>Analysis 49</td>
<td>Analysis 50</td>
</tr>
</tbody>
</table>

Ensemble forecast of the French / German storms (surface pressure)
Start date 24 December 1999 : Forecast time T+0 hours
Lothar 42-h MSLP forecasts

deterministic forecast totally misses damaging storm over France; some ensemble members forecast it well.

Probabilities commonly estimated from frequency of event in the ensemble.

Question: what constitutes a “good” ensemble forecast?

Here, the observed is outside of the range of the ensemble, which was sampled from the pdf shown. Is this a sign of a poor ensemble forecast?
Ensembles and truth should be draws from the same distribution

We need lots of samples from many situations to evaluate the characteristics of the ensemble.

- **OK**: Happens when observed is indistinguishable from any other member of the ensemble.
- **High Bias**: Happens when observed too commonly is lower than the ensemble members.
- **Too Little Spread**: Happens when there are either some low and some high biases, or when the ensemble doesn’t spread out enough.

ref: Hamill, MWR, March 2001
Such ensemble forecasts will be “reliable” (if there are enough members)

In a reliable forecast, the event occurs at the same relative frequency as the probability you forecast.

But even if the ensemble and the truth are drawn from the same distribution, with a small ensemble you won’t get reliable probabilities due to sampling error (Richardson, QJRMS, 2001)

BSS = 0.354
What other characteristics of ensemble forecasts are important?

“Sharpness” measures the specificity of the probabilistic forecast. Given two reliable forecast systems, the one producing the sharper forecasts is preferable.

But: don’t want sharp if not reliable. Implies unrealistic confidence.
“Spread-skill” relationship

Small-spread ensemble forecasts should have less ensemble-mean error than large-spread forecasts.

- **Best**: Ensemble-mean error from a sample of this pdf on avg. should be low.
- **Better**: Ensemble-mean error should be moderate on avg.
- **Nearly Useless**: Ensemble-mean error should be large on avg.
- **Climatology**:
Why run forecasts from many initial conditions?
Why run forecasts from many initial conditions?

these also have errors, and observations aren’t available everywhere
Why run forecasts from many initial conditions?

these also have errors, and observations aren’t available everywhere

this will inevitably have some errors, else why assimilate new observations?
Why run forecasts from many initial conditions?

First Guess → Data Assimilation

Observations

Forecast Model

Analysis

these also have errors, and observations aren’t available everywhere

hence the “initial condition” will inevitably have some error; it will inherit some characteristics of the forecast error and the analysis error.

this will inevitably have some errors, else why assimilate new observations?
Why run forecasts from many initial conditions?

these also have errors, and observations aren’t available everywhere

hence the “initial condition” will inevitably have some error; it will inherit some characteristics of the forecast error and the analysis error.

this will inevitably have some errors, else why assimilate new observations?

and of course errors tend to grow with time, so it’d be helpful to have a sense of the diversity of possible outcomes
EnKF (Ensemble Kalman Filter) naturally simulates uncertainty in observations, prior forecast.
Desirable properties for ensembles of initial conditions

(1) true model state and ensemble are random draws from the same distribution (same as before).
   - i.e., ensemble samples “analysis uncertainty.”
   - implies what you’d expect: larger differences between members in data voids, or where prior forecast differences were growing.

(2) differences between subsequent forecasts ought to grow quickly enough that ensemble-spread consistent with ensemble mean error.
   - with perfect forecast model, (2) will happen naturally if you take care of (1)
Spread should grow as quickly as error; part of spread growth from manner in which initial conditions are generated, some due to the model (e.g., stochastic physics, higher resolution increases spread growth). Focus on initial-condition aspect.
For most models, spread of hurricane tracks is smaller than track ensemble-mean error.
Global ensemble forecast models have systematic under-estimation of maximum wind speed. Lesson: we’re far from conquering model error in NWP and ensembles.
“Model error”

• Imperfections in the forecast model, due to:
  – inadequate resolution
  – unduly simple physical parameterizations
    • deterministic may be inappropriate
  – coding bugs
  – lack of coupling, e.g., ocean-atmosphere
  – use of limited-area nested model
    • boundary-condition imperfections
    • one-way nesting of outer domain, lack of ability for resolved scales to interact with planetary scales
  – etc.
Treating model error

• Improve your model
• Incorporate stochastic parameterizations where appropriate
• Multi-parameterization
• Multi-model

(Carolyn Reynolds will review further)
Ensemble products
(what we’re here to discuss)

• For the fields where we are starting to have some confidence in ensemble guidance, what can we do to convey that information in useful ways to the forecaster and to the public?
Initialized 00 UTC 5 August 2009.

* indicates observed best-track position.

Bi-variate normal distribution fit to ensemble member positions; contour encloses 90% of fitted probability.

GEFS/EnKF a bit north and too fast.

NCEP has northward & westward bias, few members track.

ECMWF tracks decent up to Taiwan landfall

CMC has very large spread, esp. after landfall.

UKMO too north, too fast.

Example: Typhoon Morakot
Example:

hurricane Jimena

Initialized 00 UTC 30 Aug 2009

all models have westward bias; none of the forecasts particularly good.
Example: Hurricane Bill

Initialized 00 UTC 19 August 2009.

All models slow, to varying extents.

GEFS/EnKF and ECMWF tracks decent.

UKMO, CMC have westward bias.

NCEP, FIM decent.
An experimental multi-model product

Dot area is proportional to the weighting applied to that member

• = ens. mean position
* = observed position
Multi-model error (GEFS/EnKF, ECMWF, FIM, UKMO, CMC, NCEP)

Not much improvement from multi-model. Why?
Multi-model error (GEFS/EnKF, ECMWF only)

Now some improvement, ~ 6 - 9 hours lead.
Questions

• Mine:
  – Are ellipses, colors useful way of conveying ensemble information?
  – Are products like the multi-model synthesis shown here potentially useful to forecasters?

• Yours
Acknowledgments

– NOAA HFIP program for 2009 summer test support
– NOAA THORPEX, ESRL base for EnKF development
– U. Texas NSF supercomputer for CPU cycles
– Phil Pegion for programming support, analysis.
– Mike Fiorino (tracker files, software sanity check)
– Other ESRLites, e.g., Chris Harrop for scripting, Stan Benjamin for FIM development and management, Paula McCaslin for display software
– Tim Marchok (tracker files)
– Daryl Kleist (GFS parallel runs)
– Chris Snyder, Jeff Anderson for ensemble bull sessions
– and you, for your constructive feedback
How the EnKF works:
2-D example

Start with a random sample from bimodal distribution used in previous Bayesian data assimilation example. Contours reflect the Gaussian distribution fitted to ensemble data.
Review of Atlantic Basin activity

Atlantic Basin Storms, 31 Jul 2009 to 03 Oct 2009

- Hurricane
- Tropical Storm
- Tropical Depression

Stations:
- Claudette
- Danny
- Bill
- Erika
- Ana
- Fred
Review of Eastern-Pacific activity

Eastern Pacific Storms, 31 Jul 2009 to 03 Oct 2009

- Hurricane
- Tropical Storm
- Tropical Depression

Locations:
- Guillermo
- Felicia
- Lana
- Hilda
- Linda
- Enrique
- Ignacio
- Olaf
- Marty
- Nora
- Kevin
- Jimena
Review of Western-Pacific activity

Western Pacific Storms, 31 Jul 2009 to 03 Oct 2009

- Super Typhoon
- Typhoon or Above
- Tropical Storm
- Tropical Depression