P8.2

DEFINITIONS OF DETERMINISM

Brian J. Gaudet* Pennsylvania State University, University Park, PA 16802

1. BACKGROUND – MEAN SQUARED ERROR

A common meteorological method of assessing the skill of a forecast model F at predicting a set of observations O is the mean-squared error (MSE), or its square root, the root-mean-square (RMS) error. The MSE may be decomposed into (e.g., Murphy and Epstein 1989):

$$\frac{1}{N}\sum(F-O)^2$$
$$=\frac{1}{N}\left[\sum(\bar{F}-\bar{O})^2+\sigma_F^2+\sigma_O^2-2\sigma_O\sigma_F r_{FO}\right]. \quad (1)$$

Here $\bar{F}=\sum F/N$ is the sample mean of F over N data values, and $\bar{O}$ is the sample mean of O. The term σ_F^2 is the sample variance of the forecast, $\sum F'^2/N$ where $F'=F-\bar{F}$, and σ_O^2 is the observational variance. Finally, the correlation r_{FO} is defined as:

$$r_{FO}=\frac{\sum F'O'}{\sigma_O\sigma_F}. \quad (2)$$

For a given set of observations, the MSE is a function of $\bar{F}$, σ_F, and r_{FO}. If the corresponding statistics $\bar{O}$ and σ_O of the observations are stable and known, it is easy to construct a function whose statistics $\bar{F}$ and σ_F tend to minimize the MSE, if this is the only goal (e.g., in linear regression). In contrast, r_{FO} measures the match between individual F and O values and generally cannot be improved without increasing the skill of F in predicting O.

The overall minimum MSE score is zero, and is achieved with $\bar{F}=\bar{O}$, $\sigma_F=\sigma_O$, and $r_{FO}=1$, which occurs if and only if F and O are identical. It is also easy to show that as $\bar{F}$ approaches $\bar{O}$, and as r_{FO} approaches 1, the MSE uniformly decreases for fixed σ_F. However, for a fixed $\bar{F}$ and $r_{FO}<1$, there is a specific value of $\sigma_F<\sigma_O$ that minimizes the MSE (namely, $r_{FO}\sigma_O$). Paradoxically, further increasing the value of σ_F towards the observational value σ_O makes the MSE worse. This can be traced to the fact that squaring the error penalizes outward excursions from the true value more than it rewards inward excursions of the same amount, so the MSE tends to favor forecasts with small variability magnitude versus ones with large variability. We find this occurring when, for example, smoothing a forecast time series reduces its RMS error, even though no new information has been provided by the model.

The extreme case occurs when $r_{FO}=0$, meaning that forecast perturbations are completely uncorrelated with observational perturbations. Here the minimum MSE is achieved with $\bar{F}=\bar{O}$ and $\sigma_F=0$, or, in other words, a constant forecast of $\bar{O}$. Any additional forecast variability will only worsen the MSE by an amount σ_F^2. The only skill information remaining in the MSE value is $(\bar{F}-\bar{O})^2$, which is the square of the bias. It may also be argued that in a physical model (such as WRF-ARW) it is worthwhile to compare the magnitude of forecast statistical variability with that of the observations, and to expect that the two be as close as possible. This is the situation if the observations correspond to a second-order stationary random variable, where the mean and variance of the observations have well-defined stationary values, but the actual value at a particular time is assumed to be random. A common example is when the model's forecast precipitation field correctly predicts a field of isolated convective cells, but cannot predict their precise locations.

2. TYPES OF NON-DETERMINISM

We hypothesize that in a high-resolution (e.g., one minute sampling period) time series of wind spectra the low-frequency components are deterministic and can be verified with traditional measures such as the MSE, while the high-frequency components are non-deterministic, and should be verified statistically. This leads to two questions: a) What do we mean by determinism? b) How do we determine a time scale that separates the two components?

Invoking determinism or the lack thereof occurs in a variety of contexts. In atmospheric dynamics we generally make the classical physics assumption that if a flow field is known precisely its future evolution is theoretically deterministic. However, proving that the incompressible Navier-Stokes equations have smooth global solutions in 3D given any initial velocity is one of the unsolved millennium problems of the Clay Mathematics Institute. Next, there is the phenomenon called 'intrinsic predictability' (Lorenz 1969; Zhang et al. 2006). For certain nonlinear systems the physical evolution is completely deterministic (Lorenz (1963)), but because the evolution dynamics exhibit sensitivity to initial conditions and bounded trajectories, infinitesimal uncertainties in an initial state tend to lead to qualitatively different states at later times than the one for which the initial state is known exactly. This type of unpredictability leads to essential nondeterminism after time scales determined by the error exponential growth rate of the evolution equations. This unpredictability can be seen as a property of the system rather than the observer, because if the system possesses certain properties, only measurements with infinite precision can avoid qualitative unpredictability at finite times. In effect the most important parameter is the error growth rate which is determined by the system. For timescales much less than the error growth timescale, though, the system

* *Corresponding author address:* Brian J. Gaudet, Penn State Univ., Dept. of Meteorology, University Park, PA 16802; e-mail: bjg20@met.psu.edu

is said to be predictable.

In contrast to 'intrinsic predictability' is 'practical predictability', whose absence can lead to nondeterminism even in linear systems. Limits to practical predictability for weather predictions are related to finite errors in the initial or boundary conditions, whereas intrinsic predictability is limited by infinitesimal errors.

Finite errors do not necessarily imply a failure in the design or physics of numerical forecast models, although they can. Most mesoscale models incorporate initial and boundary conditions from coarser-resolution models; hence features that would normally be resolved in the mesoscale model domain are often absent initially or away from boundaries. These absences are potential sources of error. At the other extreme of scale, mesoscale model variables must be viewed as grid-cell averages, so subgrid fluctuations are not explicitly predicted and cannot be deterministically forecasted by them. The Reynolds-averaged effects of these fluctuations can be predicted using turbulence parameterizations, but ultimately closure assumptions must be made, which could introduce errors on the averaged scale as well. Also possible are surface features (terrain, land-use) which may exist at scales less than the grid spacing but may have an important effect on observations. Finally, even for features technically represented in the mesoscale domain, if their scale is less than a certain threshold, the numerical solution will differ appreciably from the true solution, and will not be a reliable forecast. Skamarock (2004) refers to this as the model's 'effective resolution'.

3. HEURISTIC NONDETERMINISM

For practical NWP applications we must first decide whether an observation set O is deterministic. This is stated as follows: no forecast F is consistently correlated with O, unless F is explicitly a function of O. Thus for all $F \neq f(O)$ we expect $r_{FO} = 0$ if O is nondeterministic. A stochastic (random) variable is nondeterministic, and is characterized only by its statistical moments (e.g., mean, variance) or its probability distribution (which could be derivable from a large number of statistical moments). However, as noted above, if $r_{FO} = 0$ the MSE actually penalizes forecasts with a variance $\sigma_F^2 < \sigma_O^2$ if σ_F^2 is increased toward σ_O^2. Thus the MSE is not appropriate for assessing the forecast value of a stochastic variable, and measures of the closeness of fit of statistical moments should be used instead.

By introducing 'heuristic nondeterminism' we are acknowledging that the real world is not divided into observed variables which show clear MSE-based skill and other clearly stochastic variables for which r_{FO} is always zero. However, we still would like to implement some reasonable criterion for separating variables for which the MSE score alone is appropriate from those for which statistical measures should be used, based on limits of practical predictability. It is always possible, of course, that poor correlations are the result of poor or at least technologically primitive forecast models rather than an intractable feature of the observations. We may still find it useful, though, to classify phenomena that resist predictability by state-of-the-art models as heuristically nondeterministic, until proven otherwise.

One criterion for heuristic nondeterminism can be stated as follows. In the special case where $\sigma_F = \sigma_O$ and $\bar{F} = \bar{O}$, it can be shown from (1) that when:

$$r_{FO} > 0.5, \tag{3}$$

the MSE of F against O exceeds that of the 'climatological' forecast $F_C \equiv \bar{O}$, or equivalently, the fluctuating forecast component F' has a better MSE score against O' than the trivial $F' = 0$. When (3) is applied to general NWP forecast models, it indicates that the variance-scaled fluctuating forecast component, $\sigma_O F'/\sigma_F$, has a better MSE score than $F' = 0$. Furthermore, when (3) is true it can be shown that any multiple of F' between 0 and σ_O/σ_F has a better MSE score than $F' = 0$.

By contrast, if (3) is violated, then the forecast model, after bias-correction and variance-scaling, has a worse MSE score than $F_C \equiv \bar{O}$. In fact, $\sigma_O F'/\sigma_F$ would have a worse MSE score than cF' for any $0 \leq c < \sigma_O/\sigma_F$. If this situation persists over many realizations, we take this as an indication that statistical-moment verification is more appropriate and hence the MSE is inappropriate.

Murphy and Epstein (1989) used essentially the same criterion as (3) to determine if a forecast model has skill in the special case of $\sigma_F \approx \sigma_O$, and also noted the possibility of degrading the MSE score while improving the forecast of σ_O.

4. APPLICATION – AVERAGING METHOD

In view of the above, we hypothesize that because limits exist on practical predictability, there is a frequency within O such that features in its spectra at higher frequencies are heuristically nondeterministic, while features at lower frequencies are deterministic. There are a number of reasons why this might be the case. Consider O to be a time series of wind observations at a point. The observations might include turbulence that occurs at higher frequencies than the scales resolved in NWP models. In the very stable boundary layer, mesogamma scale or 'sub-mesocale' variability (Mahrt et al. 2008) occurs at longer timescales than turbulence (i.e., greater than tens of seconds) but still may not be well resolved by numerical models. Even if they can be resolved on the finest model grid, features at this scale probably will not be present in the initial or boundary conditions, or they may be influenced by topographic features that are not resolved. We will use the criterion (3), composited over synoptically similar cases,to determine the appropriate timescale of separation between deterministic and nondeterministic components of O.

If we perform an orthogonal, complete decomposition of the forecast and model time series (i.e., a discrete Fourier transform), then we can apply the MSE equation (1) to each orthogonal mode individually. If our hypothesis is correct then all modes above a particular frequency will have $r_{FO} < 0.5$.

A first attempt at doing this is shown in Figure 1. We examine the time series of u and v (with respect to model grid coordinates) using model output from Domain 4 of the WRF-ARW configuration of Seaman et al. (2008) and local tower observations for 07 Oct 2007, at approximately 9 m above the surface. The time series of one-minute averaged data covers the period 0100 - 1200 UTC. The method for imposing a cutoff filter at a particular frequency is to perform a running average over a time window corresponding to that frequency, and then storing the residual as its high-frequency component. This method is convenient but does not possess a perfect wave cutoff frequency, as it is rounded in Fourier space and introduces some high-frequency sidelobes. Also, we examine only a single case here to demonstrate the method, but an ensemble of cases is needed to establish confidence in a derived cutoff frequency.

Despite these caveats, for the running-averaged u field we see that the maximum predictability occurs around 2.5 hrs in the figure, where the correlation is about 0.8. Somewhere around 45 minutes is the point at which the correlation of the averaged u field falls below 0.5 for this particular case.

The averaged v field suggests a heuristic deterministic cutoff of about 1.5 hrs, with a correlation coefficient of 0.6. The skill for both fields begin to decline for averaging windows of 2 h and greater. Meanwhile, the residual field correlations oscillate but generally maintain low values within ± 0.1, as expected for high nondeterministic frequencies. At the largest averaging windows the residual components begin to exceed this range, however, suggesting that we are beginning to incorporate deterministic modes into the residual. Hence for our demonstration case the use of a filter scale of 2 hours is appropriate.

5. CONCLUSION

Different methods of determining the validity of forecast models may be required depending on whether the forecast variable is presumed to be random, or to be generally predictable in its occurrence. We have used a heuristic method to determine the time scale of high-resolution wind fields separating deterministic and non-deterministic components. In the process we have also tried to clarify some of the implications of predictability on the use of meteorological verification measures.

For the sample case of wind speed in a stable boundary layer, we find the relevant time scale is about 2 hours. Naturally, defining a broadly applicable timescale requires application of the methodology to a larger number of cases, which is work that is currently in progress.

6. ACKNOWLEDGMENTS

This work was sponsored by the Defense Threat Reduction Agency under contract no. W911NF-06-1-0439-MOD-P00001. The author has benefited from discussions with Dr. Nelson Seaman, Dr. Larry Mahrt, and Dr. David Stauffer on this topic.

REFERENCES

Gaudet, B.J., N.L. Seaman, D.R. Stauffer, S. Richardson, L. Mahrt, and J.C. Wyngaard, 2008: Verification of WRF-predicted mesogamma-scale spectra in the SBL using a high-frequency filter decomposition. 9th WRF Users' Workshop, Boulder, 23-27 June, 4 pp.

Lorenz, E.N. 1963: Deterministic nonperiodic flow. *J. Atmos. Sci.*, **20**, 130-141.

Lorenz, E.N., 1969: The predictability of a flow which possesses many scales of motion. *Tellus*, **21**, 289-307.

Mahrt, L., C. Thomas, and J.H. Prueger, 2008: Space-time structure of mesoscale modes in the stable boundary layer. Submitted to *Quart. J. Roy. Meteor. Soc.*.

Murphy, A.H., and E.S. Epstein, 1989: Skill scores and correlation coefficients in model verification. *Mon. Wea. Rev.*, **117**, 572-582.

Skamarock, W.C., 2004: Evaluating mesoscale NWP models using kinetic energy spectra. *Mon. Wea. Rev.*, **132**, 3019-3032.

Seaman, N.L., B. Gaudet, A. Deng, S. Richardson, D.R. Stauffer, J.C. Wyngaard, and Larry Mahrt, 2008a: Evaluation of meander-like wind variance in high-resolution WRF model simulations of the stable nocturnal boundary layer. 10th AMS Conf. on Atmos. Chem., New Orleans, 21-24 Jan., 9 pp.

Seaman, N.L., B. Gaudet, J.C. Wyngaard, Larry Mahrt, Scott Richardson, and D.R. Stauffer, 2008b: Numerical prediction of mesogamma-scale wind meandering in the nocturnal stable boundary layer. 9th WRF Users' Workshop, Boulder, 23-27 June, 7pp.

Zhang, F., A.M. Odins, and J. W. Nielsen-Gammon, 2006: Mesoscale predictability of an extreme warm-season precipitation event. *Wea. Forecasting*, **21**, 149-166.

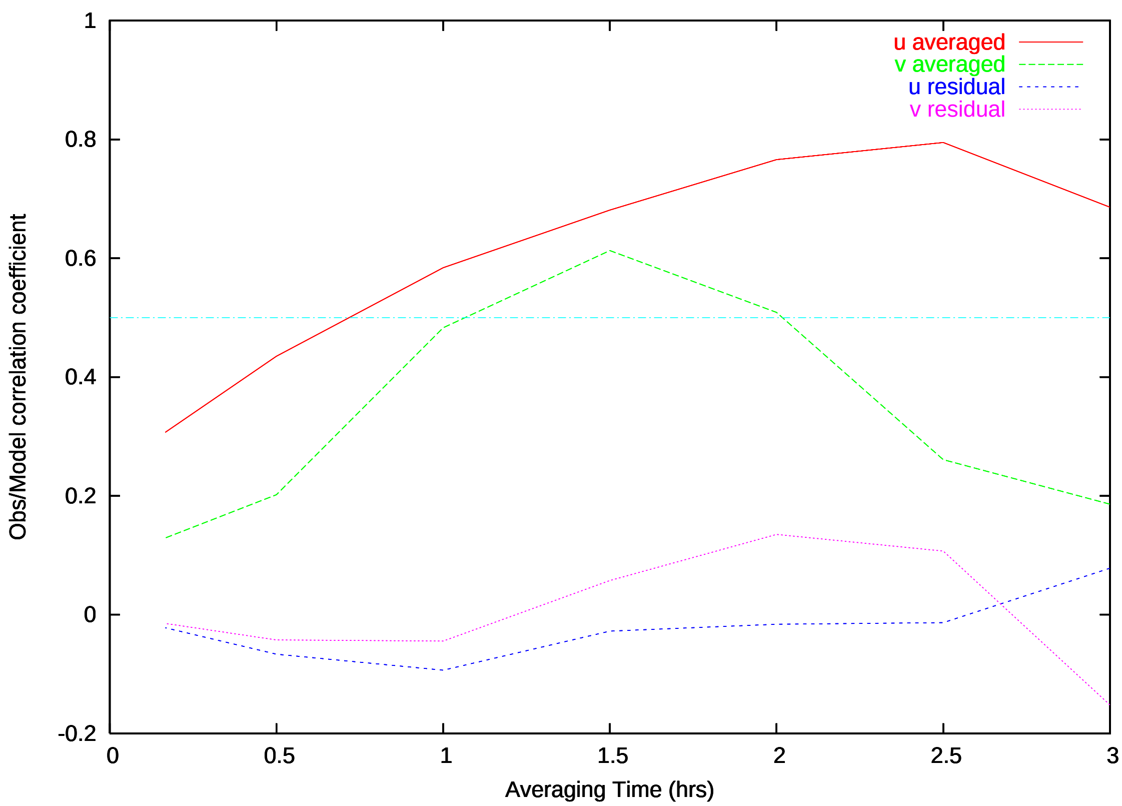

Figure 1: Correlation coefficients between WRF-ARW simulations on a 0.444-km grid at 9-m elevation and 9-m tower data, for case of 07 Oct 2007 case. Comparison is between both filtered and residual components of observations with model output for the time period 0100 - 1200 UTC.