AGU PUBLICATIONS

Journal of Advances in Modeling Earth Systems

COMMENTARY

10.1002/2017MS000999

Special Section:
Earth and Space Science is Essential for Society

The primacy of doubt: Evolution of numerical weather prediction from determinism to probability

Tim Palmer[1]

[1]Department of Physics, University of Oxford, Oxford, UK

Key Points:
- Probability forecasts based on ensembles of integrations provide an alternative to more conventional deterministic prediction techniques
- Ensemble prediction is suggesting a new approach ("more accuracy with less precision") to solving the equations of motion
- Despite this, there is still some way to go before ensembles become the primary technique for predicting weather

Correspondence to:
T. Palmer,
Tim.Palmer@physics.ox.ac.uk

Citation:
Palmer, T. (2017), The primacy of doubt: Evolution of numerical weather prediction from determinism to probability, *J. Adv. Model. Earth Syst.*, *9*, 730–734, doi:10.1002/2017MS000999.

Received 2 APR 2017
Accepted 6 APR 2017
Accepted article online 20 APR 2017
Published online 20 APR 2017

Abstract Over the last 25 years, the focus of operational numerical weather prediction has evolved from that of estimating the most likely evolution of weather to that of estimating probability distributions of future weather associated with inevitable uncertainties in both initial conditions and model equations. This evolution from determinism to uncertainty has not only increased the scientific rigor of weather prediction, it has also increased the value of weather forecasts for users. In addition, it has opened up a new approach to solving the equations of motion, likely to be of importance for both weather and climate prediction in an age where high-performance computing is limited by power consumption. However, despite all this, the numerical weather prediction community has yet to embrace fully the concept of the *primacy* of doubt. It is now time to take the final step in this direction.

The evolution of numerical weather forecasting, from its formulation as a scientific initial value problem at the beginning of the 20th century, to today's complex numerical models has been described in the accurately titled review: "The quiet revolution of numerical weather prediction" [*Bauer et al*., 2015]. Here it is shown how the skill of deterministic "best-guess" weather forecasts in the range from 3 to 10 days ahead has improved by about a day a decade: today's 6 day forecast being as skillful as a 5 day forecast 10 years ago. There are several reasons for this increase in skill. First, the numerical models themselves have become more accurate representations of the underlying equations of motion, partly because advances in supercomputer technology have allowed the basic resolution of the numerical grids to become finer, and partly because parameterized representations of unresolved processes (such as clouds, radiation, and small-scale orographic drag) have improved. In addition, the advent of satellite data in the last decades of the 20th century has enabled much more accurate estimates of global initial conditions to be made, but only after 4-D variational data assimilation schemes which can assimilate satellite radiances directly into the models had been developed.

Although these developments are extremely impressive and certainly justify the investments in data and modeling over the years, by themselves they leave exposed a critical "Achilles Heel" in deterministic weather prediction—the chaotic unpredictability of weather. That is to say, inevitable uncertainties in initial conditions and in the numerical models can, on a day when the atmospheric flow is particularly unstable, lead to a best-guess forecast becoming completely unreliability in a few days or less. This problem has dogged weather prediction since its earliest days (leading the first U.K. Met Office Director, Robert Fitzroy, to an early grave). It remains a problem to the present day. In March 2017, a Washington Post article criticized the U.S. National Weather Service for predicting 12–18 inches of snow for New York City when in reality only 7 inches fell (https://www.washingtonpost.com/news/capital-weather-gang/wp/2017/03/16/we-succeeded-weather-service-director-defends-noreaster-forecast/?utm_term=.e392c6d92cb8). In particular, the article questioned whether the Weather Service was as effective in communicating the uncertainty for this event as it needed to be—in other words, whether the Weather Service has fully embraced the notion of *primacy* of doubt. (The phrase "primacy of doubt" is taken from James Gleick's biography [*Gleick*, 1992] of the great theoretical physicist Richard Feynman: "He believed in the primacy of doubt: not as a blemish on our ability to know, but as the essence of knowing.") The answer is, in common with all operational weather centers at the present time, that it has not. I hope and expect this to change in the coming years.

The ability to quantify uncertainty is not a "bolt-on" extra, but rather a *sine qua non*. Quantifying uncertainty is less about predicting the skill of some "central" prediction, as estimating probability distributions of future weather. Does that mean abandoning all the developments that have led to the improvements in deterministic weather prediction mentioned above, and rewriting the equations of meteorology as Liouville or Fokker-Planck equations, where the basic prognostic variables are probability distributions? No—this is quite impracticable for a number of reasons [*Ehrendorfer*, 2006]. Instead, an alternative approach to the operational prediction of uncertainty has been developed: the ensemble prediction system. Here an ensemble of (say 50) weather forecasts is made perturbing initial conditions and model equations, and the probability of a future weather event is determined by the fraction of ensemble member predicting the event.

Operational probabilistic forecasting based on ensemble prediction has its roots in monthly weather forecasting [e.g., *Murphy and Palmer*, 1986]; on these timescales chaotic unpredictability is so pervasive that trying to make deterministic predictions of the weather is a complete nonstarter. However, 25 years ago, ensemble forecasts began to be made operationally over the medium-range horizon of 10 days, previously considered to be within the domain of deterministic prediction [*Palmer et al*., 1992; *Toth and Kalnay*, 1993; *Molteni et al*., 1996; *Toth and Kalnay*, 1997]. Since then, the use of ensemble prediction has become universal on all timescales: from short-range predictions a few hours ahead, to seasonal and decadal predictions and climate-change projections a century ahead.

A reason for this evolution is that an ensemble forecast has greater value than a corresponding deterministic forecast. To understand this, recall that there are many circumstances where the decisions we make can be influenced by factors that are far from certain, but which, if they did occur, would prove disastrous. We would not board a plane if there were a 10% chance of the wing dropping off midflight, nor of undergoing inessential surgery if there was a 10% chance of it leading to permanent paralysis (a situation which occurred to me). Similarly, for example, it may be quite rational to take protective action in the light of some severe weather event even if it far from certain that the event will occur. A way to see this quantitatively is to consider an idealized decision-theoretic model [*Murphy*, 1966]. Consider a weather-forecast user whose business is sensitive to some weather event E (freezing temperatures, strong winds, heavy rain and so on). If E occurs and the user has taken no protective measures, then he or she loses an amount L. However, based on the weather forecast, the user can protect against this loss, but at a cost C. (We will assume $C < L$; otherwise, there is no point taking protective action.)

A deterministic forecast provides the user with the simple decision strategy: take protective action when the event E is forecast. By contrast, a probabilistic forecast provides a more refined strategy: take protective action when the risk of the event $pL > C$. Here p is an estimate of the probability of E, based on the ensemble forecast. For example, according to this strategy, the user should almost always take protective action if $C \ll L$. However, when C is comparable with L, he or she should only take protective action when it is almost certain that E will occur. That an ensemble based prediction system provides greater value than a corresponding deterministic prediction system across a range of user values C/L can be seen in Figure 1, based on recent operational ECMWF forecasts.

Ensemble prediction systems are an integral part of the "quiet revolution," having been built from, and therefore relying on, developments in numerical methods, in physical parameterizations and in data assimilation. However, there is still a little way to go before ensemble prediction has become completely integrated into numerical weather prediction. That is to say, there is more to be done before the notion of *primacy* of doubt has been embedded into the *modus operandi* of numerical weather prediction. Let me explain why [*Palmer*, 2012].

Forecast uncertainty arises not only from imperfect knowledge of initial conditions, but also from imperfections in the forecast models themselves. This is particularly important in the tropics where unresolved diabatic processes play such an important role in shaping the resolved-scale circulations. Because of this, parameterization ansätze should be generalized from traditional deterministic schemes to more inherently stochastic schemes [*Buizza et al*., 1999; *Palmer*, 2001]. The word "stochastic" implies the addition of noise into our best-guess parameterizations. It may seem counterintuitive to some that adding noise to a model yields better forecasts. However, the notion that unresolved processes can be represented by a relatively simple deterministic formula (driven by resolved-scale fields) is, in many circumstances, incorrect. Stochastic schemes attempt to correct for this erroneous assumption, and in so doing provide more accurate representations of the subgrid processes. Indeed, having worked for many years at the European Centre for Medium-Range Weather Forecasts, I can say from experience that no change to the model has resulted in a greater increase in forecast scores than the introduction of the Stochastically Perturbed Parameterization

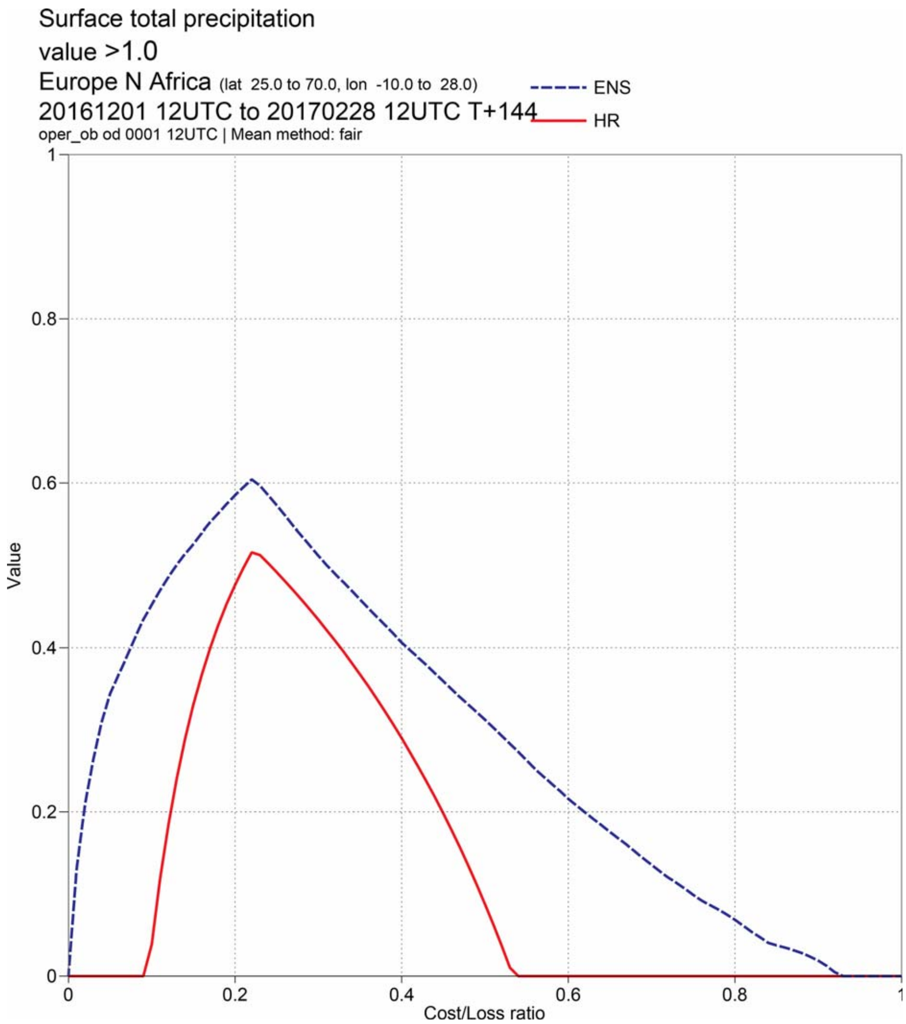

Figure 1. Potential economic value as a function of user cost-loss ratio, of the Day 6 high-resolution ECMWF forecast (solid red) and the ECMWF ensemble forecast (dashed blue) in predicting precipitation amounts greater than 1 mm over Europe and North Africa for the period December 2016 to February 2017. Potential economic value is normalized so that a perfect deterministic oracle has a value of unity and climatological probability has a value of zero. The event under study here is clearly not severe; however, the sample size is sufficiently large that the difference in value between deterministic and ensemble forecasts is statistically robust.

Scheme (SPPT) [*Palmer et al*, 2009]—increasing probabilistic skill in forecasting precipitation in the tropics by about 4 days. This scheme stochastically perturbs total diabatic tendencies in proportion to the magnitude of these tendencies. It is based on the physically based notion that individual parameterization schemes (e.g., for the boundary layer, clouds, radiation, and convection) have to be in mutual balance to produce realistic net tendencies—and therefore does not disturb this overall balance.

Currently at ECMWF (and indeed at all centers which have similar stochastic parameterization schemes), SPPT is added to the model after the model has been tuned in deterministic mode. This procedure is far from ideal. Since the underlying equations of motion are nonlinear, stochastic noise can create a mean offset on the climatology of the model [e.g., *Weisheimer et al*., 2014]. If the model has been tuned in deterministic mode, stochastic parameterization has the potential to detune it.

Why not simply tune the model with the stochastic scheme included? Two arguments why not to do so are sometimes raised. The first is that the deterministically tuned model is needed for the best-guess deterministic forecast (which is typically performed at higher resolution). However, as discussed, the aim of ensemble prediction is not to put error bars around a single, potentially unreliable, deterministic forecast. In addition, precisely because this model is run at higher resolution, the ensemble and the single high-resolution forecast will be inconsistent at grid points where resolution matters, e.g., due to orography. These inconsistencies can lead forecasters to give undue weight to the single high-resolution deterministic forecast, and to discount information provided by the ensemble. This can have very negative consequences in terms of the public perception of weather prediction when the high-resolution deterministic prediction goes wrong. The second argument is that tuning becomes more difficult in models with stochastic parameterization schemes; not least the characteristics of the stochastic noise are described by extra parameters which will need to be considered in the tuning mix. This can be countered by referring to the explosive growth of machine-learning technology, which allows tuning to be performed objectively with much larger parameter spaces than has hitherto been possible.

If we embrace the concept of *primacy* of doubt, then the answer to this dilemma is simple: abandon the separate high-resolution deterministic forecast altogether and base all operational output on the ensemble forecast. The most likely forecast will then be, exclusively, the mode (or modes) of the ensemble distribution, and the rest of the distribution becomes completely consistent with this modal estimate.

Two consequences of embracing the *primacy* of doubt can be mentioned. The first is in the communication of uncertainty. At present there is a tendency for forecasters, in media-based forecasts of high-impact weather, to "err on the side of caution"—that is to say to be more certain than is justified by the ensemble output. This can occur for a number of reasons. For example, as discussed above, the high-resolution forecast may predict an event not unanimously supported by the ensemble. In these circumstances, forecasters may give undue weight to the former over the latter. In addition, the media themselves can put pressure on the forecasters for their forecasts to be communicated in a more deterministic way than is justified by the ensembles. By embracing the *primacy* of doubt, forecasters should resist these temptations and should instead communicate the uncertainty in a way consistent with the ensemble: they should tell it as it is! It is up to emergency response organizations to decide how to respond to this information: as the cost\loss model above indicates, it is perfectly rational for emergency response organizations to shut down subways, airports, and highways on the basis of a sufficiently high probability of a high-impact weather event. However, it must be made clear to all that these decisions are based on appropriate risk assessments by the response organizations, and not on the fact that, under pressure, the forecasters said that the weather event *will* occur, when in fact there was no rational reason for such unbridled certainty. A second consequence of embracing the *primacy* of doubt is more relevant to the numerical weather prediction community. Weather forecast centers and international committees like the Working Group on Numerical Experimentation should start promoting probabilistic ensemble-based scores (such as the Continuous Ranked Probability Skill Score which is strongly linked to Potential Economic Value [*Palmer and Richardson*, 2014]) as their *primary* means of intercomparison, rather than comparing forecasts using the more traditional deterministic 500 hPa rms error or anomaly correlation coefficient, whose relation to forecast value is much more tenuous. It is time for these latter scores, although they have a long and venerable history, and have served the community well, to quietly slip into a much more subservient and secondary position as the key intercomparison metrics of numerical weather prediction.

Of course, none of this detracts from the goal of producing higher-resolution (ultimately convectively resolved) global models. Indeed, this goal may be achievable more quickly if we fully embrace the notion of stochasticity in our weather and climate models. In particular, with explicit stochastically based estimates of subgrid parameterization uncertainty, one can now assess whether the numerical algorithms used in weather and climate models are unnecessarily complex or precise. For example, the weather and climate modeling community has for many years assumed by default that all floating-point variables should be coded with 64 bit (double-precision) representations. Motivated by the role of stochasticity in the ensemble prediction system, it has recently been shown that (virtually) the whole ECMWF forecast model can be recoded with 32 bit representations, leaving probabilistic skill scores unchanged [*Váňa et al.*, 2017]. This result is important because power consumption has become the constraining factor in high-performance computing. The single most important determinant of power consumption is the number of bits that the computer has to move from processor to processor, or from processor to memory. Reducing precision in this way will have a substantial impact on the efficient use of available energy.

It will be interesting to examine whether the data assimilation code can be similarly recoded with 32 bit representation without losing probabilistic skill. If it can, then the savings on the entire forecast system in moving to 32 bits would be comparable with a typical supercomputer upgrade. But why stop here? The massive growth in artificial intelligence research has led to the demand for ultrafast 16 and 8 bit chips, and raises the profound question of what the real information content is in the billions of bits that comprise a contemporary weather and climate model. For example, some of the Earth-System processes in weather and climate models are so uncertain that perhaps they could be represented with 16 bit representations without degrading forecast accuracy. (Sometimes it is argued that higher not lower precision is needed to represent such processes in climate models, where increments are tiny compared with the states onto which the increments are added, potentially leading to increments being systematically rounded to zero. However, this is not a sound argument. It is possible to use a procedure I will call the "Düben Trick" [*Düben et al*., 2017] whereby low precision is used to calculate the increments but where increments are added to the state at higher precision. In this way, computations and data movements—the parts that consume most energy—can be performed at low precision without loss of accuracy.) If we perform computations and only move the bits that contain real information, then we can redeploy the energy resources saved to, for example, build weather and climate models with higher resolution [*Palmer et al*., 2014]. This could be crucial as we plan how to exploit next-generation exascale computers efficiently.

Richard Feynman believed in the *primacy* of doubt. While the numerical weather prediction community has increasingly realized the *importance* of representing doubt in our operational weather forecasts, it now needs to take the final leap and embrace its *primacy*—for the sake of meteorological science and its crucial impact on society.

References

Bauer, P., A. Thorpe, and G. Brunet (2015), The quiet revolution of numerical weather prediction, *Nature*, *525*, 47–55.

Buizza, R., M. Miller, and T. N. Palmer (1999), Stochastic representation of model uncertainties in the ECMWF ensemble prediction system, *Q. J. R. Meteorol. Soc.*, *125*, 2887–2908, doi:10.1002/qj.49712556006.

Düben, P. D., A. Subramanian, A. Dawson, and T. N. Palmer (2017), A study of reduced numerical precision to make superparameterization more competitive using a hardware emulator in the OpenIFS model, *J. Adv. Model. Earth Syst.*, *9*, doi:10.1002/2016MS000862.

Ehrendorfer, M. (2006), The Liouville equation and atmospheric predictability, in *Predictability of Weather and Climate*, edited by T. N. Palmer and R. Hagedorn, Cambridge Univ. Press, Cambridge, U. K.

Gleick, J. (1992), *Genius: The Life and Science of Richard Feynman*, 489 pp., Open Road Integr. Media, New York.

Molteni, F., R. Buizza, T. N. Palmer, and T. Petroliagis (1996), The ECMWF ensemble prediction system: Methodology and validation, *Q. J. R. Meteorol. Soc.*, *122*, 73–119.

Murphy, A. H. (1966), A note on the utility of probabilistic predictions and the probability score in the cost-loss ration decision situation, *J. Appl. Meteorol.*, *5*, 534–537.

Murphy, J., and T. N. Palmer (1986), A real time extended-range forecast by an ensemble of numerical integrations, *Meteorol. Mag.*, *115*, 337–348.

Palmer, T., P. Düben, and H. McNamara (2014), Stochastic modelling and energy-efficient computing for weather and climate prediction, *Philos. Trans. R. Soc. A*, *372*, 1–2, doi:10.1098/rsta.2014.0118.

Palmer, T. N. (2001), A nonlinear dynamical perspective on model error: A proposal for nonlocal stochastic-dynamic parametrisation in weather and climate prediction models, *Q. J. R. Meteorol. Soc.*, *127*, 279–304.

Palmer, T. N. (2012), Towards the probabilistic Earth-system simulator: A vision for the future of climate and weather prediction, *Q. J. R. Meteorol. Soc.*, *138*, 841–861, doi:10.1002/qj.1923.

Palmer, T. N., and D. Richardson (2014), Decisions, decisions!, *ECMWF Newsl.*, *141*, 12–14.

Palmer, T. N., F. Molteni, R. Mureau, R. Buizza, P. Chapelet, and J. Tribbia (1992), Ensemble prediction, *ECWMF Tech. Memo.*, *188*, 45 pp.

Palmer, T. N., R. Buizza, F. Doblas-Reyes, T. Jung, M. Leutbecher, G. J. Shutts, M. Steinheimer, and A. Weisheimer (2009), Stochastic parametrization and model uncertainty, *ECMWF Tech. Memo.*, *598*, 1–42.

Toth, Z., and E. Kalnay (1993), Ensemble forecasting at NMC: The generation of perturbations, *Bull. Am. Meteorol. Soc.*, *74*, 2317–2330.

Toth, Z., and E. Kalnay (1997), Ensemble forecasting at NCEP and the breeding method, *Mon. Weather Rev.*, *125*, 3297–3319.

Váňa, F., P. Düben, S. Lang, T. Palmer, M. Leutbecher, D. Salmond, and G. Carver (2017), Single precision in weather forecasting models: An evaluation with the IFS, *Mon. Weather Rev.*, *145*, 495–502, doi:10.1175/MWR-D-16-0228.1.

Weisheimer, A., S. Corti, T. Palmer, and F. Vitart (2014), Addressing model error through atmospheric stochastic physical parametrizations: Impact on the coupled ECMWF seasonal forecasting system, *Philos. Trans. R. Soc. A*, *372*, 20130290.