We meteorologists cannot escape probabilities!

ANDERS PERSSON, SWEDISH METEOROLOGICAL SOCIETY

Anders Persson was employed at the SMHI 1967-90 and 2002-07, at ECMWF as a Staff Member at the Meteorological Operations Section, 1991-2001 and consultant at the UK Met Office 2008-10. During 1983-87 and 2003-11 he worked as an on-and-off consultant at ECMWF dealing with the problems about the best use of the ECMWF medium range forecasts.

Mureau, Rivera and many others might be worried, but I think that a lot of other colleagues are silently quite pleased: -Why bother with these awkward probabilities when the public doesn't want them?

It reflects an attitude that probability forecasting is something that since 1992 has been forced upon the meteorological community through the launch of the ensemble prediction system (ENS). But probability forecasting has old roots in forecasting meteorology since the early 1900's when it was championed by the American Clement Abbe and the Swede Anders Ångström among others. It does not work in isolation from deterministic forecasting, but is linked to it in a mathematically, operationally and psychologically consistent way.

In this article I will discuss why the meteorological community obviously has problems with probabilities. It has led to a situation where many meteorologists on one hand, knowing what is "politically correct" endorses the use of probabilities ,but on the other hand try to evade the use of it by different forms of quasi-deterministic approaches.

I will in this article suggest, from my own experiences, that problems with probabilities in forecasting meteorology is part of a wider problem of accommodating statistical concepts. The conclusion is that the meteorological community cannot "wish away" probabilities or forecast uncertainties. To remedy the situation I will at the end suggest a layout to a statistical "training course" for meteorologists.

1. My development as a weather forecaster

When I came into weather forecasting in the late 1960's we relied on "synoptic experience" underpinned by the "Bergen School" frontal theory and to some degree the "Chicago School" theories about planetary waves. We were supposed to tell the public and special customers in great detail in time and space "how the weather is going to be".

When after 5-6 years as an "Assistant Forecaster" I became a "Senior Forecaster" I started to question the prevailing deterministic forecast culture. During a visit to the UK Met Office in 1975 I learned that one way to deal with forecast uncertainty was just to limit the forecast information. It made the message easier to understand, remember and interpret. As I later found out, filtering out less predictable details also improved the forecast quality and minimized the forecast "jumpiness".

These less predictable scales might, however, contain important information. By leaving those out we run the risk of throwing out the baby with the water. The solution was to use probabilities.

As aviation forecaster at Gothenburg and Malmö airports I had become acquainted with the Terminal Aerodrome Forecast (TAF). Its code with deterministic parts, complemented with probabilistic parts, provided an excellent and balanced way to communicate weather forecast information.

Probabilities were, however, at that time among meteorologists mainly regarded as a way to "cover one's back". My colleagues at Malmö airport used to tease one of our colleagues, Torsten, for doing "elastic forecasts" with TEMPO or PROBs for stratus and/or bad visibility in situations when they went for CAVOK (Ceiling and visibility OK). But it didn't take long to realise that on a flatland 72 m above the sea level and close to the Danish sounds, TEMPO and PROB for stratus or low visibility was the name of the game. The pilots didn't mind Torsten's forecasts, on the contrary; when he could confidently promise CAVOK they could really trust him.

Scepticism about probabilities also prevailed at the Central Forecast Office at SMHI. However, one of my colleagues with wide views, Erik Liljas, thought otherwise and brought in the American meteorologists and statistician Alan Murphy as a consultant. From his lectures and papers we learned that uncertainty forecasts, in particular expressed as probabilities, served as valuable additional forecast information. Instead of pretending that we made exact and perfect forecasts, it was a better strategy to admit our forecast uncertainty and try to estimate the degree of uncertainty.

But if uncertainty information, and in particular probabilities, are valuable additional information, why don't people ask for them? And why are many meteorologists not unhappy about the current state of affairs? To answer these questions we must look into a) the attitude of the public to uncertainty information, b) the role of statistics in forecasting meteorology and c) the history of ensemble forecasting.

2. The public's urge for categorical answers

It is almost an uncontested belief in the meteorological community that "the public doesn't understand probabilities". I am not so sure. When we launched the ENS in 1992 at ECMWF the most devoted supporter of probability forecast was our "chef du cuisine" in the staff restaurant. He was daily commuting in his car between Reading and Basingstoke and listened in to Radio Berkshire's probability forecasts of slippery roads. He had over the years learnt to relate the probabilities to his own experiences of the road conditions.

But it is true; a majority of the public and many customers want deterministic forecasts. The main reason for this is, I believe, that they want us meteorologists to make the decision for them. A probability forecast such as "25% probability of >10 mm/12h" would force them to make their own decision and there would be nobody to blame if it was wrong. In cases of bad forecasts, we serve as scapegoats and are portrayed as incompetent idiots. But it is just a ceremonial to redirect the blame from the decision makers. If we really were incompetent, then we would of course have been replaced. Instead, are silently "rehabilitated" to serve as sacrifice lambs for the next serious weather event.

But are we left to choose between a Scylla of easily understood popular deterministic forecasting and a Charybdis of tricky and unpopular probability forecasting"? In summer 1986, July 5, when I watched the BBC weather a Third Way seemed to open up.

On the screen was the Chief BBC Forecaster, Bill Giles, discussing the uncertainty of a developing cyclone over the Bay of Biscay. He told us why the developing storm could move into southern England, but also why it could take a track over northern France. It struck me when I watched this "how clever he is". He had managed to turn a potential weakness, not knowing exactly what would happen, into strength. If we forecasters have the opportunity to present uncertain weather situations in honest ways, we get a chance to demonstrate our meteorological knowledge and experience.

I have since then during my years in Britain seen this "trick" demonstrated by the BBC forecasters several times. The most drastic case occurred in mid-December 2011. The approaching Atlantic storm was very unpredictable with large spread in the ensembles and the deterministic NWP models changing drastically from one run to the next. In one computer run there was no storm at all! All in all, nobody really knew if and where Britain would be hit by flooding, snow or gales.

Instead of taking a lay-back attitude and blame the "unreliable computer" the BBC forecasters, guided by the Met Office, acknowledged the great uncertainties and presented the situation as it evolved in honest professional ways. This highly impressed not only the British public but rendered the forecasters appreciation from the Met Office and BBC managements and even from the UK Government. Their presentations so cleverly turned their weakness into strength that few noticed that the public never were given a categorical "what will happen" weather forecast until the very last moment!

Uncertainties, presented in the right professional way will offer opportunities to the forecasters to enhance their professional standing – providing they break with some ingrained attitudes to statistical thinking.

3. The role of statistics in forecasting meteorology

Statistical methods were for long times seen as equally powerful as dynamical methods in weather forecasting. A comparison between the two approaches, led by Edward Lorenz in the late 1950's, was essentially a draw. However, with the increasing computer power the dynamic methods displayed more potential for improvement than the statistical. Statistics would thereafter serve as a secondary, but still important support to NWP.

Sweden was at that time, together with the US and UK, pioneering numerical weather prediction (NWP) but the forecast quality was still fairly low. These were the days of the quasi-geostrophic models and their construction restricted their ability to forecast extreme weather conditions. But we could see how the NWP slowly got better and better.

3.1 FORECAST VERIFICATION

With the increasing success of the NWP the modellers became excessively confident and started to spread the word that "in 5-10 years we do not need any forecasters"[1]. This, of course, created an animosity between them and the forecasters, which has lasted to our days. In their infighting both sides used verification statistics to argue their case. The forecast verifications at the meteorological institutes were therefore rarely conducted by an independent unit.

However, both sides had problems with the interpretations of the verifications which rarely showed their forecasts to be as "good" as perceived. Attempts to "sex up" poor statistical results were compensated by their inabilities to make correct interpretations. The mathematical simplicity of most verification scores may give an impression that the interpre-

[1] When I heard this for the first time in 1966, at the start of my meteorological education, I decided not to believe in it. Much later, during research into the history of NWP, I found evidence that it had been a political ploy to secure funding. The money providing authorities were not as much impressed by promises of "better" weather forecasts, as with "cheaper" forecasts.

tations will also be simple. That is not generally the case. During my time at ECMWF we coined the phrase: “What looks good might be bad, what looks bad might be good”.

3.2 STATISTICAL INTERPRETATION

From forecast verification there is only a small step to correction of systematic forecast errors. During the 1970’s the model independent Perfect Prog Method (PPM) was replaced by the model dependent linear regression Model Output Statistics (MOS). One of the best overviews of these methods is found in the proceedings of the September 1982 ECMWF Seminar on Interpretation of Numerical Weather Prediction Products, http://www.ecmwf.int/publications/library/do/references/list/1616

When I arrived at the ECMWF half a year later I found a great animosity against statistical interpretation, in particular MOS, probably because it was seen as a competitor since it improved the forecasts as much as about ten years of model development. Even worse from the modellers’ point of view, to maintain a sufficiently large statistical sample of forecasts for the regression analysis the MOS meteorologists wanted to freeze the development of NWP, or at least maintain old version of NWP models.

A way to circumvent this problem was to apply adaptive statistical interpretation schemes often with a Kalman filter as the basic tool. They were accepted by the NWP modellers but scared the forecasters because these automatic systems appeared to replicate experienced forecasters.

3.3 PROBABILITY FORECASTING

Probability forecasting has by tradition been most successfully applied in North America where it figures in TV broadcasts as well as in newspapers. What is regarded as “difficult” in Europe does not seem to cause major problems “over there”. I do not think that Americans are more intelligent or better educated than Europeans, the only explanation must be that American public has been exposed to probability forecasts for longer times than the Europeans and, like the “chef du cuisine” at ECMWF, from their experience learned to link these probabilistic weather forecasts with what subsequently occurred.

Instead of putting the blame on the public, “not understanding probabilities” I think we must ask ourselves if the problem with probabilities rather lies within the meteorological community.

3.4 THE DETERMINISTIC TRADITION IN METEOROLOGY

In a highly deterministic conjecture in”Essai philosophique sur les probabilités” Pierre Simon Laplace invites us to regard the present state of the universe as the effect of its past and the cause of its future:

« Une intelligence qui, à un instant donné, connaîtrait toutes les forces dont la nature est animée et la situation respective des êtres qui la composent, si d’ailleurs elle était suffisamment vaste pour soumettre ces données à l’analyse, embrasserait dans la même formule les mouvements des plus grands corps de l’univers et ceux du plus léger atome ; rien ne serait incertain pour elle, et l’avenir, comme le passé, serait présent à ses yeux.»[2]

This conjecture, often called “Laplace’s Demon”, has had a great resonance in classical mechanics of which dynamic meteorology is a branch. It has served as a great inspiration to many great Scandinavian meteorologists, from Vilhelm Bjerknes to Lennart Bengtsson.

And why not? With an improvement of the ECMWF forecasts by about one day per decade, and no levelling out, the predictabilities looks like extending towards eternity. The progress is as exciting to follow as the voyage of one of the space probes leaving our solar system and heading for the nearest stars. In this sense the ECMWF great technological success story is comparable to other technological institutions like CERN, NASA and the Hubble Space telescope. Whereas these institutions explore the limits of Micro cosmos or Macro cosmos the ECMWF explores the limits of determinism and the possible existence of Laplace’s Demon.

The back side has been that the detailed and confident looking ECMWF 10-day forecasts appeared to make ideas about uncertainty and probabilities redundant. Forecasters soon found that they could safely blindly follow the computer output: when it was correct they could take the credit, if it was wrong, they could put the blame on the ECMWF, which after all had the best and fastest computers!

4. The history of ensemble forecasting

There has, however, always been one major “problem” with the ECMWF 10-day forecasts, their “jumpiness”. While it took five days to realise that a 5-day forecast was wrong, and by then most people had forgotten about it, within 12 hours a previous forecast would appear to be wrong being contradicted by a later one. This problem was brought to the fore by a group of Dutch meteorologists during a workshop at ECMWF in 1986:

“The increasing sophistication of our professional customers makes it harder and harder to maintain our professional reputation. Our clients discover time and again that medium-range forecasts suffer significant changes from one day to the next. …often enough a medium a medium-range forecast is not consistent with the one issued the day before. Do we really want the public to become as cynical about wea-

[2] ”An intellect which at a certain moment would know all forces that set nature in motion, and all positions of all items of which nature is composed, if this intellect were also vast enough to submit these data to analysis, it would embrace in a single formula the movements of the greatest bodies of the universe and those of the tiniest atom; for such an intellect nothing would be uncertain and the future just like the past would be present before its eyes.”

ther forecasts as many of us are about economic forecasts?" (Tennekes et al, 1986)

The Dutch group argued that "no forecast is complete without a forecast of forecast skill" and that the time had come "to make skill forecasting one of the central goals of predictability research…it is a question of credibility. "

The problem of "forecast forecast skill" was subject to a lot of work at the ECMWF in the 1980's. The main approach was to find a statistical relation between the skill of the most recent forecast "jumpiness" and different predictors related to the flow type and "jumpiness". The aim was that a forecaster one day would be able to confidently base his/her outlook on the ECMWF high-resolution deterministic 10-day forecast system (HRES) up to say five days, the next day perhaps up to seven days.

The "forecast forecast skill" work didn't yield any useful results, because the problem was incorrectly posed. Any information in the "jumpiness", i.e. the spread of forecasts, should rather relate to the skill of the mean of these forecasts, than to any individual forecast.

With increasing computer capacity it was possible in 1992 to apply a modified Monte Carlo technique to the problem. In brief: the initial analysis was randomly perturbed into an ensemble of equally possible initial analyses. Each of these was run on a slightly coarser model version creating an ensemble of equally possible 10-day forecasts. This Ensemble Prediction System (ENS), has since the upgrade 1996 met all demands for effective medium range forecasts:

1 **As a source of deterministic forecasts:** The mean of the ensemble (EM) or the median of all the members would provide the least error estimation of a deterministic forecast. The spread around the mean would indicate its accuracy. A further advantage, and an important one, was that the EM was much less "jumpy" than the HRES.

2 **As a source of probabilistic forecasts:** The proportion of ensemble forecasts within certain intervals or exceeding certain thresholds could be used to calculate probabilities for these intervals or threshold. Calculations of combined probabilities would also be possible (e.g. for heavy rain and gale force winds). By relating the probabilities to their climatological values further insight could be reached.

3 **Synoptic overview:** By grouping together similar ensemble members into averages (clustering) the forecasters wouldn't have to look at all the ensemble members individually. With a limited number of clusters (≤ 6) the forecasters could easily relate the probabilities to the forecast synoptic flow regimes.

This is, however, not how the ENS mainly has been used by operational meteorologists or promoted by the scientists and modellers.

1. THE EM (OR MEDIAN) AS A PROVIDER OF DETERMINISTIC FORECASTS:

1.1 1.1 The EM was from the start met with some apprehension by the ECMWF modellers. Knowing that its forecasts would be more accurate than the HRES and much more stable it was of course seen as an unwanted competitor.

1.2 The probabilistic modellers feared that if the forecasters used the EM they would feel so relieved from uncertainties and "jumpiness" they would loose interest in probabilities.

1.3 Since it only needed a dozen members to compute a fair EM, not 30-50, the computer minded modellers saw a danger that an increased use of the EM could make it more difficult to motivate upgrades of the computer system.

1.4 The forecasters' main education and experience in the short range, where frontal zones and baroclinic instabilities play important roles, made them unaccustomed to use the EM where these synoptic features had been smoothed out when the spread in the ensemble had grown to a certain extent.[3]

1.5 The groups held these misgivings for themselves and united around the argument that the EM could not serve as a forecast since it, as an average, did not represent a physically possible state of the atmosphere. This is true, but irrelevant since what we want is the most accurate deterministic forecast, not a nice illustration to a meteorological textbook.

1.6 Another argument was that in cases of bi-modality, i.e. two distinct flow regimes, the EM would be completely unrealistic, "between two stools". Bimodality turned out to much less frequent than expected (see 3.1), but even in those cases it can be shown that the EM provides the "best estimate" for a deterministic forecast.

1.7 It was also said that the EM averaging wrecked the physical consistency between different weather parameters. That is also true, but has since long been the case also in deterministic models with respect to precipita-

[3] Before 1980 medium range forecasters were quite accustomed to work with temporally and spatially averaged maps. It was on such maps that C G Rossby in the late 1930's discovered "his" waves

We meteorologists cannot escape probabilities!

tion forecasts (accumulated values over a time interval) and for example cloud forecasts (instantaneous values).

1.8 Those who favoured HRES and not EM as the provider of deterministic forecasts information used the argument that HRES was a "better model" thanks to its higher vertical and horizontal resolutions. This was indeed an issue the first 4-5 years but since then these model differences have mattered less and less. If the HRES output today was secretly replaced by output from the deterministic ENS Control few would notice any difference.

1.9 The promotion of HRES to EM sometimes took parodical appearances. If the HRES forecasts scored better than some other external model, it was said that "the ECMWF model is the best". When the EM scored better than HRES it was said to be "better at minimizing the RMSE (root mean square error)".

2. PROBABILITIES AND SPREAD:

Nor were probabilities high on the agenda when the ENS started in 1992. The main use of the ENS was seen to be in "forecasting the forecast skill", the expected skill of the current deterministic HRES forecast. But not everybody was pleased with that. A manager from a meteorological institution put forward his view that "what we want is better forecasts, not some b***y index telling us how wrong they are!"

2.1 The "forecast forecast skill" approach is as problematic when applied on ENS as is with lagged forecasts because the spread is related to the expected error of the EM, not to any individual forecast, whether from the HRES, ENS Control or any individual member.

2.2 The HRES-ENS model differences, which were emphasized in the provision of deterministic forecasts (see 1.6 above) were now downplayed, to motivate the use of the ENS for measuring the a priori skill of the HRES.

2.3 Even if we regard HRES as a "better" model this is, at best, only true in the short range and its role as a simulator of atmosphere motion. The HRES's ability to forecasts extremes slightly better than ENS Control is irrelevant due to both models' inability to provide certainty estimations. For the medium range forecasting, for which purpose the ECMWF was set up the ENS is therefore superior to both HRES and ENS Control.

2.4 The safest way to use the ENS spread to estimate the a priori skill of the HRES is to do it only in cases with high predictability, when there is small spread and an overwhelming majority of the members support the HRES. If the chance is 80% that a major hurricane will strike New York everybody knows what has to be done, but can everybody feel safe when the risk is "only" 20%?

2.5 So far the ENS spread. Probabilities were, as mentioned above, not high on the agenda in 1992 and became officially recognized by the Technical Advisory Committee (TAC) only in 1995, perhaps to motivate the 1996 upgrade from 32 to 50 members, an important requisite to calculate more reliable probabilities. Until then only the Predictability and Met Ops sections had promoted the use of probabilities on training courses and Member Sate visits..

2.6 One reason the TAC, and particular its chairman, was sceptical was because in their view probabilities could only be derived from MOS systems. The idea that probabilities could be calculated from the proportion of ENS members was questioned because how did we know that all members were equally likely, an argument already Laplace had to deal with. These misgivings often reflected controversy between frequentist and Bayesian statistical schools, going back 200 years (although none of us realised this at the time).

2.7 To use the spread information from a probabilistic system to measure the accuracy, not of the EM, but of a single deterministic model, can lead to severe inconsistencies such as a HRES forecast of 17 m/s and from the ENS a 30% probability for gales, or a 75% gale probability from the ENS when the HRES forecast has only 7 m/s.

In statistical theory the arithmetic mean is called the "first moment", the variance (spread) the "second moment" and the skewness the "third moment". The ENS probabilities therefore relate naturally to the EM - not to HRES or any individual model.

If the first moment is swept under the carpet, it is of course difficult to become acquainted with the last two moments, which define probabilities. Consequently, we can discuss at lengths the usefulness of probabilities, but as long as the ensemble mean or median is not in the picture, the probabilities will, so to say, "hang freely in the air".

The difficult to understand probabilities in forecasting meteorology is therefore made more difficult if fundamental relations in statistics are unknown or thwarted to suite political group interests.

3. CLUSTERS

Clusters can be made in an unlimited number of ways, every one with advantages and disadvantages. The one ECMWF decided on in 1992 made sure that the members within a cluster were synoptically similar during the three days from D+4 to D+7. This clustering, as with any other algorithm, was originally meant only to be a way to get an overview of the ENS but early on it took a perhaps too important role.

3.1 University scientists hoped through the clustering to find support to a popular hypothesis about "multiple equilibria" in the atmospheric flow. But only occasionally did the clustering show bi- or tri-modality, i.e. synoptically two or three distinct flow patterns. Mostly there was a rather continuous change from one cluster to the next.

3.2 The ECMWF clustering came early on under criticism when individual members, at a specific lead time, were found to be "outside" the cluster. In such cases a day-by-day clustering would have been better, but on the other hand presented other problems.

3.3 Some pedagogic images of the ENS easily gave the impression that the verifying Truth would always be found by the largest cluster. This might be the case when it contained > 80% of the members, but not when it was < 20%.

3.4 Even if a more populated cluster was closer to the Truth it was smoother due to the averaging process. To the forecasters it therefore appeared less realistic than less populated clusters which for the opposite reason preserved more details and appeared "more realistic", a problem similar to 1.4 above.

3.5 To remedy this problem, the ECMWF has in their new clustering algorithm made the most typical member represent the cluster instead of the cluster average. This has, however, opened up for the misunderstanding, know from the psychological literature, that what looks "typical" also is more probable..

3.6 In a speech at ECMWF in December 2000 about the future of ECMWF the President of the Council and Director of the German Weather Service, Udo Gärtner wanted us to find out which one of the 50 ensemble members was "the best". This is an application of the common "Model of the Day" approach where forecasters faced with a multitude of NWP models try to find out which one to follow.

3.7 A similar approach was already discussed in 1992, to use the clusters to judge which one of two successive, but "jumpy" HRES forecasts was the most likely – if any of them at all. Again, this is yet another attempt to try to escape probabilities and find a deterministic forecast in the ENS.

To summarize: the use of probabilities from the ECMWF ensemble system has been hampered not only for the commonly quoted reason that "probabilities are difficult to understand" but also due to internal meteorological agendas and poor understanding of basic statistical concepts.

5. What to do?

Although there is, and has always been, a strong element of statistics in weather forecasting this has rarely been reflected in the curriculum. Here is an outline of a weekly course in probability theory, which starts from the three different definitions of probabilities:

MONDAY: The classical definition of probability. This starts with the familiar illustration of tossing coins and dice, but move on to combinatoric which, among other things, will tell about the problems to add or divide probabilities.

TUESDAY: The frequentist definition of probabilities. This involves verification of probability forecasts, mean and variance and significance. Why is it not possible to deceive the "proper" Brier score? How does it seem to "know" my true opinion?

WEDNESDAY: The subjective definition of probabilities. From conditional probabilities (the chance of having rain if it is windy is normally not the same to have windy conditions if it rains. From Bayes' Rule we move into several applications well known in forecasting meteorology.

THURSDAY: Decisions from probabilities. From the elementary cost-loss model, we move into more complex models such as the Kahneman-Twersky Propect theory. The value of total uncertainty will also be explored.

FRIDAY: The psychology of probabilities. The communication of uncertainty does not totally depend on numerical values of probabilities. Verbal statements can function equally well as well as intervals with tacit probabilities.

The mathematics of probabilities is fairly simple, so the emphasis will be on the conceptual understanding.

6. Summary

The introduction of the ENS in 1992 was seen as a paradigm shift in operational weather forecasting by introducing probability forecasts. But probability forecasting had been around for almost 100 years, so perhaps the real paradigm shift was rather the introduction around 1980 of the ECMWF high-resolution deterministic 10-day forecast system (HRES) which gave the illusion about exact, confident and detailed weather forecasts.

The weather forecasting problem remains, however, a probabilistic matter, whether we "like" it or not. The question is to find ways to handle this problem waiting for the realization of "Laplace Demon" in a very distant future.

In the meantime we have to acknowledge uncertainty. It has been argued in this article that forecast uncertainty information is a win-win product: it increases the value of the forecast and it gives the forecasters opportunities to display their skill, experience and knowledge, which make people trust the forecasts. Because, as Alan Murphy used to emphasize, even if a forecasts is very accurate with no uncertainties, if it is not trusted, it will serve no purpose.