# ECMWF Feature article

METEOROLOGY

# Increasing trust in medium-range weather forecasts

# Increasing trust in medium-range weather forecasts

## Anders Persson, Bob Riddaway

The fifth edition '*User guide to ECMWF forecast products*' has been published and can be found at: http://www.ecmwf.int/products/forecasts/guide/. The following is a record of a conversation between Anders Persson, the author of the Guide, and Bob Riddaway, the Editor of the *ECMWF Newsletter* on the best use of the ECMWF products and how to really 'add value' to them.



Anders Persson



Bob Riddaway

## Contents of the User Guide

**Bob Riddaway (BR)** *You have just finished a new edition of the 'User guide to ECMWF forecast products'. Could you briefly summarize its contents?*

**Anders Persson (AP)** I hope it will facilitate the use of the numerical weather prediction (NWP) products from the ECMWF forecasting system, in particular the medium-range forecast products. The way forecasters deal with medium-range NWP output differs in many ways from how they deal with the short-range NWP on one hand, and monthly and seasonal NWP on the other.

**BR** *The ECMWF medium-range forecasts are the best in the world, skilful up to a week or more. Do we really need a 'User Guide'; can't the forecasts just be used straight away?*

**AP** This might be true for ordinary weather a few days ahead, but not for unusual, dangerous or extreme weather events. Decisions about postponing operations or evacuations cannot be made just by blindly following computer output. Consider what happens with a modern aircraft. The more they become equipped with advanced instruments and sophisticated techniques the thicker the manuals and the more demanding is the training on how to operate them. Although there are autopilots we all prefer to have trained pilots in the cockpit!

**BR** *So I assume the Guide is primarily written for weather forecasters. But what about other groups of meteorologists, and even non-meteorologists?*

**AP** It is intended that scientists and NWP modellers will find it a convenient reference to the ECMWF forecasting systems, as might non-meteorological users of our products. To this end, I have put in a lot of effort into maintaining an informal and non-technical style.

**BR** *Very little has been written on the use of NWP in practical forecasting.*

**AP** There are good textbooks and courses in NWP and dynamic meteorology but, as far as I know, nothing has been documented about what the forecasters should do when, for example, the short-range NWP conflicts with more recent observations. What to do when successive runs from the same NWP model differ widely? Or when various NWP models offer different solutions? Most forecast experience of the synoptic-scale weather systems is derived from the short range, where the forecasters modify their positions and intensities, and do not have to question their coming into existence as often is the case in the medium-range.

**BR** *That being so, on which sources have you based your advice in the Guide?*

**AP**  I have drawn on my own experiences, and discussions over the years with forecasters, scientists and statistically-interested meteorologists. It has helped me find out which forecast rules of thumb seem to work and which don't, as well which new rules could be suggested on theoretical grounds. As one of my Swedish professors (Tor Bergeron) once told me: "*Behind good forecast practice is often hidden good theories; equally, good theories should provide a basis for good forecast practice*". That's why it is important the take notice of how experienced forecasters work. What I became aware of during the work with the Guide, and found particularly interesting, is the relationship between how experienced forecasters work and statistical concepts.

**BR** *So that is why you have put a lot of emphasis on statistics in the Guide?*

**AP**  To a large extent operational weather forecasting involves assessing, combining and correcting information, to which you must apply a statistical approach. Forecasters collect and evaluate their forecast experience of NWP. *That's statistics*. They look for systematic errors or typical characteristics in the NWP. *That's statistics*. They update or modify their preliminary forecast in light of new information. *That's statistics*; in particular what is called Bayesian statistics. An intuitive use of statistical know-how seems to count as much as synoptic experience, in particular in the medium range. The Guide is also intended to encourage a statistical approach by showing it has a scientific basis.

**BR** *When we think about statistics in relation to weather forecasting we normally think about verification and statistical interpretation.*

**AP**  Forecasters use forecast verification both for product control and as a way to improve the forecasts. Two chapters of the Guide provide information on how to validate and verify the deterministic and probabilistic forecasts. This can be quite counterintuitive. "*What looks bad might be good, what looks good might be bad*" is one of the recurring themes in the Guide. Statistical interpretation is just given a short introduction, but much more could be written about it.

## The main messages in the Guide

**BR** *Another recurring theme in the Guide is a warning not to over-interpret the details of the NWP output.*

**AP**  What are considered the strong aspects of the high-resolution NWP models are also their weakest. Although a 10-day forecast of mean-sea-level pressure looks as realistic as an analysed map, most of the synoptic details are usually misleading: for example, features might have the wrong intensity, the wrong position or perhaps not come into existence at all. If these smaller non-predictable synoptic details were disregarded or removed, the remaining large-scale flow pattern would provide enough information to form a basis for useful forecasts.

**BR** *You make it almost sound as if increasing the model resolution would be detrimental to the developments of more skilful NWP models!*

**AP**  On the contrary, the Guide makes clear that the higher the resolution the better the forecasting system's ability to simulate atmospheric processes, facilitate the assimilation of observations and enable a better representation of mountains and coastlines, and their effect on the large-scale flow. Even if the smallest atmospheric features that can be resolved by the high-resolution might only be predictable for a few hours, their representation is nevertheless necessary for the energetic exchange between different scales. But just because the forecasts are *calculated* at very high resolution doesn't mean that they have to be *presented* at that resolution.

**BR** *But these smallest scales are not just numerical noise. There are numerous cases when small-scale developments, often involving extreme weather, have been correctly forecast several days in advance.*

**AP**  But the average predictive skill for the smallest scales is much lower than for the larger scales. Consequently this forecast information should generally not be presented in a categorical way – rather a probabilistic approach should be taken. This approach has found its best realization in the EPS (Ensemble Prediction System); the ensemble mean represents the most predictable part of the flow and the probabilities are based on the less predictable synoptic features that are not seen in the ensemble mean.

## Probabilistic forecasts

**BR** *A common opinion seems to be that probability forecasting is just a cheap means of escape for the forecaster!*

**AP**  It is true that it is more satisfying for the forecasters to express forecasts for severe events in probabilistic terms. But they do not do that to 'cover their backs'. The main advantage with probabilities lies with the users of the forecast. Since we, as everybody knows, cannot make perfect forecasts, users are better served if we acknowledge our uncertainty rather than pretend there is none. Verification of manually made probability forecasts has shown that the forecasters generally underestimate their uncertainty.

**BR** *It is often said that people do not understand probabilities.*

**AP**  I think they do. If someone is told there is a 30% probability of rain he or she will ignore the warning if it is just a matter of hanging out the washing, but not if it is about an outdoor exhibition of expensive art.

**BR** *But how can we make the public and other users appreciate probability forecasts?*

**AP**  It might sound preposterous but I think that the public and other users should, in some gentle way, be made aware that part of the responsibility for 'bad forecasts' rests on them! They do not always ask the right questions. After "What will the weather be?" they should follow up with: "How sure are you about this? What else could happen?" Unfortunately by not doing this they are in effect asking the meteorologists to make the decisions for them. This had the 'advantage' that any criticism of decisions made can be diverted to the meteorologists. We meteorologists appear partly to be paid to act as scapegoats!

**BR** *Another argument against probability forecasts is that they by definition are never right or wrong, except when 0% or 100% have been forecast.*

**AP**  The real value of weather forecasting is not primarily about getting it 'right' or 'wrong' but to support decision making. This is best achieved through the use of probabilities. Although they cannot be verified individually, their overall performance can. It has been shown that such forecasts have skill of great value to users who know how to interpret probabilistic information. In the Guide I quote Ed Lorenz who once said that what the critics of probability forecasting fail to recognize, or else are reluctant to acknowledge, is that a forecaster is paid not for exhibiting skill but for providing information to the public. A probability forecast conveys more information than a simple categorical forecast of rain or no rain.

**BR** *In the Guide you introduce the usefulness of probabilities by showing that that in some uncertain weather situations it is actually better for most users not to be given a forecast at all rather than one with a misleading high certainty. What is your thinking behind that rather unusual approach?*

**AP**  I wanted to challenge the common view of members of the public that in very uncertain weather situations, it is "better that the forecasters make the guesses than the laymen". I would argue that in such situations nobody should make any 'guesses' but take note of the uncertainty itself – information that only the forecasters can judge. Fittingly, another recurring theme in the Guide is the value of knowing one's uncertainty – it is better to know that we do not know than to believe that we know when we don't. Users can then interpret this 'total uncertainty' according to their own interest. From this thought-provoking example it follows that if the forecasters express their uncertainty in more elaborate ways, for example by probabilities, the usefulness of their forecasts will further increase.

## Deterministic and categorical forecasts

**BR** *Turning for a while to the operational, deterministic model, I gather that the Guide doesn't recommend the use of this by itself, particularly at longer ranges.*

**AP** It is pointed out several times that the errors of the deterministic NWP forecasts increase to rather high levels, ultimately towards the level of a persistence forecast or a guess. These are significantly higher than the error level of using a climatological average as a forecast. So a forecast practice which blindly follows the deterministic output far out in the medium range might at some stage provide worse guidance than use of climatological information (Figure 1).

**BR** *Another difficulty faced by forecasters is 'jumpiness' in the forecasts.*

**AP** An NWP model must by necessity be 'jumpy'; some changes between forecast runs are necessary to enable the forecasts to improve. There is, however, no reason to convey this 'jumpiness' to the end-users. Nothing degrades the public's confidence more than 'jumpy' forecasts, in particular in connection with anomalous or extreme weather events. A bad five-day forecast will be identified as such only after five days; a 'jumpy' forecast will be identified immediately. Users also tend to draw the logical conclusion that the previous, disowned forecast must have been wrong. I think that many complaints about 'bad' forecasts actually refer to 'jumpy' forecasts.

**BR** *Does the Guide give any advice how to tackle this 'jumpiness' problem?*

**AP** Of course, the ultimate solution lies in the ensemble technique. This approach can also be applied to the deterministic forecasts. There is significant skill in previous forecasts. Just because the latest is on *average* best, doesn't mean that it is *always* so (Figure 2). A 'mini-ensemble' or 'consensus forecast' of the latest 3 or 4 deterministic forecasts displays similar characteristics as the EPS. When averaged, their mean has an increased accuracy and a higher degree of day-to-day consistency than the latest forecast. As with the EPS, the relative reduction of forecast 'jumpiness' in the 'mini-ensembel' is about three times larger than the relative reduction in forecast error. It is suggested in the Guide that forming a mini-ensemble is a fruitful way of combining output from the deterministic model and EPS.
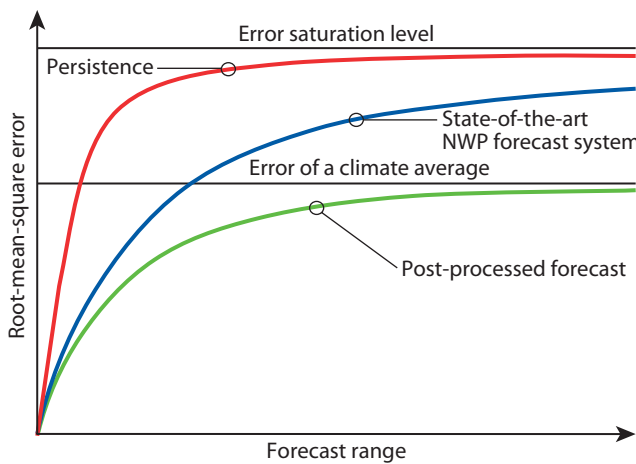


**Figure 1** A schematic illustration of the forecast error development of a state-of-art NWP forecast system and persistence, whose errors asymptotically approach a higher error level (the so called 'error saturation level' which is the average error of a guess) than post-processed forecasts, which approach a lower level defined by the 'error of a climate average'.
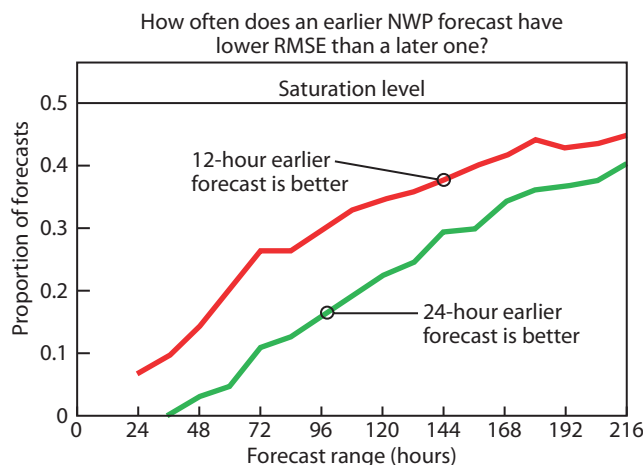


**Figure 2** The likelihood that a 12-hour or 24-hour earlier forecast is 'better' (in terms of root-mean-square error) than today's forecast. The parameter is the mean-sea-level pressure for Northern Europe for October 2009 to September 2010. The result is almost identical if the anomaly correlation coefficient is used as the verification measure. For very long forecast lead times both forecasts are almost without skill and the chance that a previous forecast is better becomes almost random and the likelihood approaches the 50% level (also called the 'saturation level').

## Relationship between categorical and probabilistic forecasts

**BR** *What does the Guide say about the relation between categorical and probabilistic forecasts?*

**AP** The provision of categorical and probabilistic forecasts to the end-users support and complement each other. Issuing skilful categorical weather forecasts during normal weather conditions builds up trust with the users. If users have confidence in the weather service's ability to 'get it right' in normal weather conditions they will of course more likely trust its forecasts, even probabilistic ones, in cases of extreme weather.

**BR** *But sometimes deterministic NWP forecasts might conflict with the probabilistic guidance from the EPS.*

**AP** This is a very challenging problem and it is discussed in the Guide, but perhaps not as extensively as I would have liked because this topic requires more investigation. In principle, when the EPS spread is small one would expect the deterministic forecast to be within the ensemble, not display any 'jumpiness' and have good skill. But the opposite might not necessarily be true. At present there is a good overall agreement between the behaviour of the deterministic forecast and the EPS spread. But there is still room for improvement of the EPS, in particular with respect to its own 'jumpiness'.

**BR** *Why do you think there is some reluctance amongst forecasters to make use of the ensemble mean, the arithmetic average of all the ensemble members?*

**AP** It is said that such average fields do not constitute genuine dynamically three-dimensional representation of the atmosphere, cannot represent bi-modal solutions, and is not able to represent extreme or anomalous weather events. This is quite true, but if forecasts of extreme events *are* fully reliable they will not be smoothed out in the ensemble mean. If they are not reliable they will instead appear as probabilities. Bi-modal solutions , when the synoptic evolution sets off in two rather different directions, are quite rare so that is not a significant problem. Apart from providing deterministic weather forecasts with optimal accuracy and consistency the ensemble mean of, for example, the mean-sea-level pressure serves as a useful background for displaying probabilistic weather information, since it puts the probabilities into a synoptic context (Figure 3).

**BR** *It is also said that ensemble averaging might lead to inconsistencies between different weather parameters.*

**AP** Also true, but this problem can also be found in deterministic output, for example between forecasts of instantaneous cloud and 6-hour accumulated precipitation. The popularity of the EPSgrams is not lessened by its inability to show the interconnections between the four different weather parameters or bi-modal solutions.
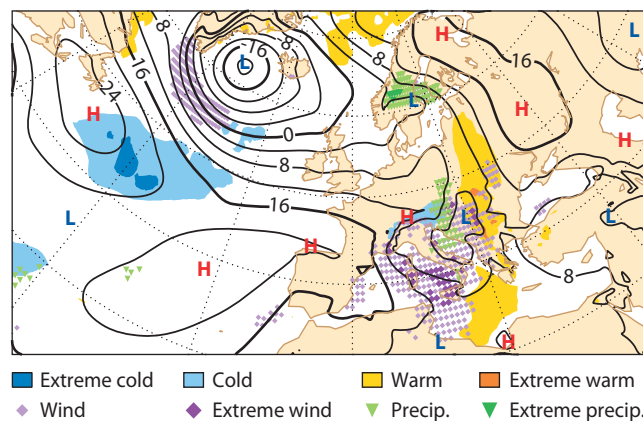


**Figure 3** The 'Anomalous Weather' or 'Interactive EFI' chart from 72 to +96 hours based on the EPS from 00 UTC on 16 September 2011. It shows the geographical distribution of the extreme forecast index (EFI) for maximum wind gust, 24-hour precipitation and 2-metre temperature, overlaid with the ensemble mean of the 1000 hPa geopotential height.

| ■ Extreme cold | ■ Cold | ■ Warm | ■ Extreme warm |
| ◆ Wind | ◆ Extreme wind | ▼ Precip. | ▼ Extreme precip. |

## Increase the trust in medium-range forecasts

**BR** *Estimations at the time the ECMWF was founded indicated that large economic gains would benefit society thanks to medium-range weather forecasts. At present, forecasts are, on average, synoptically useful for up to a week or more. What can be done to further increase the use of medium-range weather forecasts?*

**AP** One way is of course to further improve the overall skill of the deterministic forecasting system. It has increased by about a day each decade and it is likely that this trend will continue thanks to planned improvements (increased resolution, improved physics and even better use of observations) in the next decade.

**BR**  *But is this enough? Although the decision makers always have had high confidence in the skill of the ECMWF forecasts in general, there seems to be less confidence about the skill of specific forecasts for a particular event, for example when extreme weather is likely.*

**AP**  I think a keyword here is 'trust'. Good forecasts that are not trusted have little value, irrespective of how well they verify. The EPS system is able to pin-point which deterministic forecasts can be relied on and thereby increase the willingness of users to make decisions based on these good forecasts.

**BR**  *Occasionally the EPS spread, and perhaps forecast 'jumpiness', tells us that the weather situation is difficult to predict. Can useful forecasts be provided in these circumstances?*

**AP**  It is my impression is that at present many users think that only very skilful medium-range forecasts, like those we have in the short range, can be used for decisions in extreme weather situations. But one of the most important points in the Guide is that useful decisions can also be made when there is low predictability and the deterministic forecast is likely to be wrong – *provided that the forecasters are aware of this uncertainty!*

## Role of medium-range forecasters

**BR**  *That brings us to the question of the role of the human forecasters.*
*What is their role given the steadily increasing skill of the deterministic forecasts and the EPS?*

**AP**  Almost since the start of NWP the demise of weather forecasters "within 5 to 10 years" has been announced. But the forecasters obviously serve a purpose, because today there are more carrying out operational duties than ever before, and increasingly so in the commercial sector. Of course automatic forecasts on the Internet and elsewhere might satisfy the need of weather information during normal conditions, but in situations with extreme or high-impact weather nobody would base their decisions just on automated output. In addition, in these critical situations, there will always be more than one source of automatic NWP information and they are unlikely to give a consistent message.

**BR**  *Do you think that most forecasters 'can add value' to NWP products in terms of root-mean-square error or any other verification score?*

**AP**  I do not want to interpret 'adding value' in this way. Competing with the computer in terms of verification scores is an internal meteorological contest and without much relevance outside the meteorological community. On the other hand, if there really were a contest, NWP models would be easily 'beaten', in particular in the medium range, because NWP modellers and forecasters play different games, like cricket versus baseball or football versus handball. NWP modellers strive to provide forecast systems with optimum accuracy with the over-riding condition that the atmospheric motions contains all scales – irrespective of whether they are predictable or not (Figure 4). Weather forecasters do the opposite, they disregard and damp unpredictable features in order to improve the accuracy and reduce the 'jumpiness' of their categorical forecasts (Figure 5). This means there is no contest between NWP modellers and forecasters! So I do not think it is constructive to define 'added value' in terms like that since weather forecasting is to support decision making.
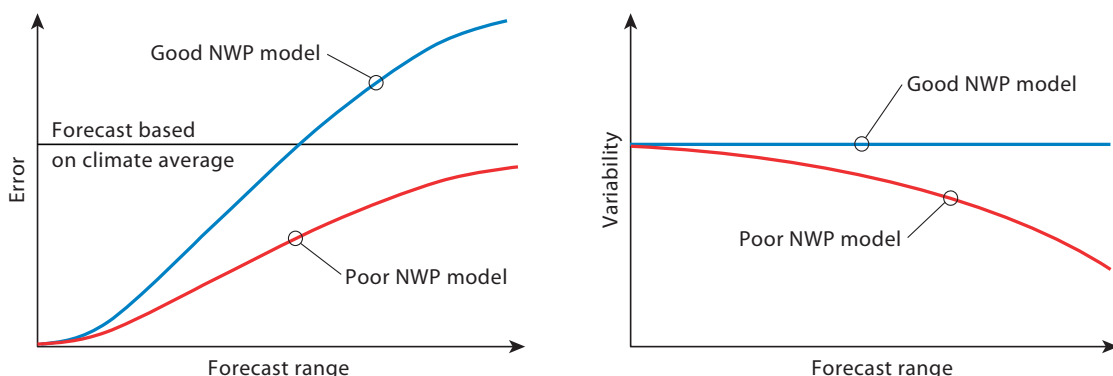


**Figure 4**  The root-mean-square error and variability of two NWP systems: one a good state-of-the-art high-resolution NWP model, the other a poor NWP model due to excessive diffusion or coarse numerical resolution. The errors of the good model (blue line in the left panel) approach a high error level because the model is able to represent the whole spectrum of resolvable atmospheric scales throughout the forecast (blue line in the right panel), while the errors of the poor model approach a lower level (red line in the left panel) because the model suffers from a gradual reduction of the scales and thereby the variability (red line in the right panel).
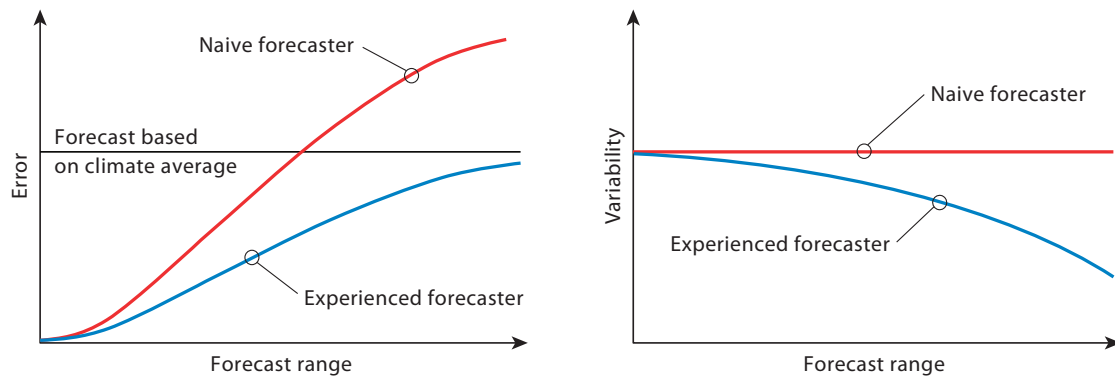
**Figure 5** The root-mean-square error and variability of two forecast practices: one experienced forecaster who disregards or damps less likely synoptic features, while the other naively reads off the raw output from a state-of-the-art NWP model. The errors of the experienced forecaster (blue line in the left panel) approach a low error level because of a gradual reduction of the less predictable scales and thereby the forecast variability (blue line in the right panel), while the errors of the naive forecaster approach a higher error level (red line in the left panel) because the forecasts maintain the whole spectrum of resolvable atmospheric scales for all lead times irrespective of whether they are predictable or not (red line in the right).

**BR** *Am I right in thinking that throughout the Guide you advise the forecasters not to try to imitate an NWP system?*

**AP** When I summarized the recommendations in the Guide they boiled down to the general advice to "do the opposite to computers":

Deterministic NWP output provides highly-detailed synoptic scenarios, irrespective of their predictability; forecasters are advised, with increasing forecast range, to avoid adding detailed information to the NWP, but rather remove information.

Deterministic NWP must change run by run; forecasters are advised to dampen any forecast 'jumpiness' in order to increase the confidence in the deterministic forecast.

Deterministic NWP gives an impression of very high certainty; forecasters are advised to make positive use of uncertainty and thereby increase the usefulness of the forecasts.

**BR** *So in your view there is a role for the forecasters if they take on the challenge to use the forecast uncertainty in a positive, productive way?*

**AP** Yes, in this respect the meteorologists' forecast experience coupled with knowledge of modern NWP remains crucial. What distinguishes the professionals is their ability to make use of uncertainty. This is quite obvious to anybody who follows a weather briefing, for example on radio or TV. In highly predictable weather situations anybody can confidently tell "what it is going to be". But in difficult weather situations it is only the professionals who can clearly express uncertainty: "There is a chance that this cold air might be delayed...The rain might exceed 50 mm in places...We cannot exclude hurricane force wind gusts." Knowing the uncertainty enables those who have to make decisions to give the weather information the weight it deserves and thereby in an optimal way take other non-weather related factors into consideration. It is the uncertainty information that really adds 'extra value' to the forecasts.