Tubing method – Control and results

Jean-Marc Jacquin, Météo-France

jean-marc.jacquin@meteo.fr

1. Summary

In a first part the tubing method is presented after four years of use. We can go now toward medium-range forecast with a more probabilistic presentation.

The second part presents the results of the verification of the first years (1999 and 2000).

The third part describes the experiment about the Extreme Forecast Index at Météo-France.

2. About the tubing method

To get a good description of the method, the best is to refer to the Sixth Workshop Proceedings, which includes the lecture "Medium-range forecasting with ensemble prediction products", by Frédéric Atger. The practical aspects and the products are described in the Seventh Workshop Proceedings: "Medium range forecasts at Météo-France with the EPS", by Jean-Marc Jacquin.

In a nutshell we can say that: for general public, Météo-France provides a deterministic medium-range forecast (from Day+4 to Day+7), based on the interpretation of the ensemble mean, or the central cluster, which is supposed to be the most likely meteorological future. We join to this a confidence index, based on the spread of the ensemble, and in theory, on the number of tubes given by the EPS each day. In order to use a large synoptic scale, we group Day+4 with Day+5, and Day+6 with Day+7. The confidence index goes from 1 (bad confidence) to 5 (good confidence), but in practice, 1 and 5 are not used, or very carefully.

After four years, what are the lessons to learn from the method?

Many cases show us that we can get unsatisfied with the confidence index. For instance when there is a low above a part of France, the EPS presents generally a large spread and it is also inconsistent, and we know that the location of the low is much more uncertain than anything else. So that the index is often low because it is applied to the synoptic point of view. But on the contrary in these cases we could be more confident about a rainy weather or cold temperature or both of them.

Generally speaking, we do not choose the confidence index completely according to the number of tubes, or the number of possible variants. It depends more subjectively on the weather type, the consistency of the EPS, and also on the precision of the forecast. The less the medium-range forecast is detailed, the higher the IC is. And verification confirms this.

The main drawback of the confidence index is that the same index applies as well to a general forecast, as to a regional forecast, and to all kind of bulletin.

We noticed a change in the tubing since the EPS spread correction on the 6th of February. In fact, as we look at one tube by the fields of the extreme run which designs it, we can see that it is now very far from the central cluster mean, and sometimes quite not realistic and impossible to interpret. Even for Day 3 range, the extreme runs are from times to times different from every deterministic model. So we have to develop a new visualisation of the tubing in order to understand better the possible meteorological variants of the forecast, which is one of the goal of the tubing method.

In a same idea, we can ask ourselves what the consequences of an EPS with 100 runs could be in the future. Probably we will have to increase the radius of the central cluster in order not to get too many tubes. So that the central cluster mean would be smoother than now, and more difficult to read.

3. Control and results

Daily, the forecasters control in a subjective way their forecast. We compare the synoptic fields of the central cluster mean, with the analysis above France and around. The marks go from A (very good synoptic scenario) to D (very bad). As the forecasts group two days, the comparison takes the evolution into account.

Here is the result for 1999 and 2000:

Subjective mark of EPS central cluster around France:

Distribution of the marks:

%	Very good	Good	Bad	Very bad
D+4 and D+5	35%	39%	21%	5%
D+6 and D+7	13%	36%	36%	15%

The central cluster is very good or good on nearly ¾ of the days for Day+4 and Day+5, and nearly half for Day+6 and Day+7.

Which quality for each index?

D+4 and D+5

Confidence index:	1	2	3	4	5
Very good	50%	24%	34%	38%	0%
Good	50%	38%	37%	42%	100%
Bad	0%	35%	23%	16%	0%
Very bad	0%	3%	6%	4%	0%
Total:	**100%**	**100%**	**100%**	**100%**	**100%**

D+6 and D+7

Confidence index:	1	2	3	4	5
Very good	0%	8%	14%	21%	0%
Good	67%	31%	37%	43%	0%
Bad	33%	44,5%	34%	27,5%	0%
Very bad	0%	16,5%	15%	8,5%	0%

We can notice that a good index (4) is more often associated with a good forecast than a low index (2). But the difference is not so large. And a low index is more often with a good forecast than with a bad one.

At this moment, we can already wonder whether a mark B for a synoptic scenario is sufficient to get a good weather forecast. Also, we tried during a period another control method. We produced each day a chart with the real weather during the two last days, and we compared it with the medium-range forecast for the same period. As this method is very tedious, we replaced it at the beginning of 2001 with a direct control of each weather parameter, and a global mark at the end.

Subjective mark of real weather forecast on France: 1999 and 2000 result

Distribution of the marks

%	Very good	Good	Bad	Very bad
D+4 and D+5	18%	50%	27%	5%
D+6 and D+7	17%	38%	35%	10%

The distribution is close to the former. But the very good cases are less numerous, and unfortunately days with a good synoptic frame are not the same as one with a good real forecast.

Which quality for each index?

D+4 and D+5

Confidence index:	1	2	3	4	5
Very good mark	0%	5%	20%	18%	0%
Good mark	100%	30%	49%	54%	100%
Bad mark	0%	45%	26%	25%	0%
Very bad mark	0%	20%	5%	3%	0%
Total:	**100%**	**100%**	**100%**	**100%**	**100%**

D+6 and D+7

Confidence index:	1	2	3	4	5
Very good mark	0	12%	17%	21%	0%
Good mark	0	37%	39%	37%	0%
Bad mark	0	39%	34%	32%	0%
Very bad mark	0	12%	10%	10%	0%

To sum up, we could assess that the confidence index was a first step towards a probabilistic forecast. But it was a slippery step. We have now to manage a more accurate presentation for general public. Professional users can already get more specific probabilistic products, according to their need.

4. Subjective verification of the Extreme Forecast Index

What are the motivations?
Detection of extreme events in the early medium-range by the EFI method.
Subjective verification (by a forecaster).
Evaluation of the optimal EFI value in relationship with False Alarm and Non-Detection ratios.

What is an extreme event?
An event unusually forecast (model point of view)
An event unusually observed (user point of view)
The aim of the method is to join the two points of view.

Experimentation
Period: 1st of December 2000 to 31st of March 2001 (split into 2 periods: before and after the 6th of February, because of the change in the EPS spread).
2 parameters: Wind speed and 24h accumulated precipitation.
2 ranges: D2 and D3.
Thresholds (depending on areas in France):
24RR: between 40 and 120mm
FF: between 100 and 130km/h

Conclusion
High False Alarm ratio for each EFI value.
Encouraging result for Detection ratio.
Poor performance for the highest EFI values.
No large discrimination between low EFI values.
Better for rain than for wind speed.
Better at D2 than at D3
No difference between before and after 6th of February.
Different thresholds from model and for alarm bulletins.
No very severe events during the period of experiment.
It must go on.