TECHNICAL MEMORANDUM

17

On some predictability problems

C. Girard and M. Jarraud

Research Department

May 1980

European Centre for Medium-Range Weather Forecasts
Europäisches Zentrum für mittelfristige Wettervorhersage
Centre européen pour les prévisions météorologiques à moyen

Our purpose in this small paper is to warn against the danger of dealing with predictability problems without sufficient care.

We first study the problem of mean predictability, the differences in predictability and eventually we make some proposals.

1. PREDICTABILITY

One has first to agree on a definition of predictability: it can be for example the time needed by standard deviations to reach the norm, or by anomaly correlations to drop under 60%, or .... In fact it doesn't really matter for our purpose.

We assume now that we have run an ensemble E of n cases. For each of them we can define a predictability P_i, i=1,...n then, the mean predictability should be defined as

$$P_E = \frac{1}{n} \sum_{i=1}^{n} P_i \tag{1}$$

At present the "mean predictability" is defined as the predictability of a "mean" run which would have the same large scale scores (standard deviation, or anomaly correlations) as the mean of the scores of the individual cases.

This "predictability" P_m is always different and usually smaller than P_E.

We now give a proof of this for 2 cases and for anomaly correlations. It would be easy to extend the results to n cases or to standard deviations.

We consider 2 cases where anomaly correlations are defined by 2 function $f_1(t)$ and $f_2(t)$ and

$$g_1(t) = \frac{df_1(t)}{dt} \qquad g_2(t) = \frac{df_2(t)}{dt}$$

For cases 1 and 2 the predictabilities P_1 and P_2 are defined by

$$P_1 = \inf\ (t : f_1(t) = a) \tag{2}$$

$$P_2 = \inf\ (t : f_2(t) = a) \tag{3}$$

a being some fixed constant (it can be 60%, 50%,....)

now if $f_m = \frac{1}{2}(f_1 + f_2)$ (4)

then $P_m = \inf(t : f_m(t) = a)$

$$P_E = \frac{1}{2}(P_1 + P_2)$$

$$f_m(P_E) = \frac{1}{2}\left[f_1\left(\frac{P_1 + P_2}{2}\right) + f_2\left(\frac{P_1 + P_2}{2}\right)\right]$$

$$f_1\left(\frac{P_1 + P_2}{2}\right) = f_1(P_1) + \frac{P_2 - P_1}{2} g_1(t_1), t_1 \in [P_1, \frac{P_2 + P_1}{2}]$$

$$f_2\left(\frac{P_1 + P_2}{2}\right) = f_2(P_2) + \frac{P_1 - P_2}{2} g_2(t_2), t_2 \in [P_2, \frac{P_1 + P_2}{2}]$$

$$\Rightarrow f_m(P_E) = \frac{1}{2}[a + a + \frac{P_2 - P_1}{2}(g_1(t_1) - g_2(t_2))]$$

$$f_m(P_E) = a + \frac{P_2 - P_1}{4}(g_1(t_1) - g_2(t_2)) \quad (5)$$

Therefore $P_E = P_m \Rightarrow f_m(P_E) = a$

$$\Rightarrow P_2 = P_1 \text{ or } g_1(t_1) = g_2(t_2)$$

<u>both events have a mathematical probability of 0</u> if cases 1 and 2 are not identical

$$\Rightarrow \boxed{P_E \neq P_m} \quad (6)$$

Now we try to explain why usually $P_E > P_m$ (although the reverse might happen in theory).

If we have 2 cases, 1 and 2, we can always assume that $P_1 < P_2$.

Therefore (5) $\rightarrow f_m(P_E) = a + \alpha(g_1(t_1) - g_2(t_2))$ $\alpha > 0$
if we assume now

$g_1(t) < 0$

$g_2(t) < 0$

which implies f_1 and f_2 to be 2 decreasing functions of time, then f_m is also a decreasing function of time.

i.e. $g_m(t) = \frac{df_m(t)}{dt} < 0$ (7)

and

$t_1 \in [P_1, P_E]$

$t_2 \in [P_E, P_2]$

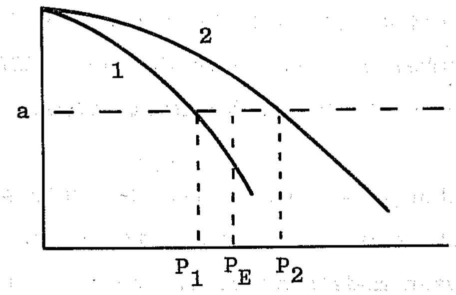


Fig. 1

and usually f_1 is decreasing more rapidly after having crossed the a-line than f_2 before having crossed the same line (because g_1 and g_2 are also 2 functions rapidly decreasing with time, as a mean). $\Rightarrow$ $g_1(t_1) < g_2(t_2)$

$$\Rightarrow f_m(P_E) = a + \alpha\beta \quad \alpha > 0 \quad \beta < 0 \quad \Rightarrow f_m(P_E) < a \tag{8}$$

$$(7) \text{ and } (8) \Rightarrow \boxed{P_E > P_m} \tag{9}$$

We can expect the difference to be smaller when the sums in E are not too different and therefore when a is large.

We can now give an order of the magnitude obtained in practice, in the quasi operational runs performed by the spectral model since September 1979.

Table 1. Mean predictability P_m and P_E computed from anomaly correlations (in hours)

		Sept.79 (5 cases)		OCT (4 cases)		NOV (5 cases)		DEC (4 cases)		JAN (5 cases)	
		P_E	P_m	P_E	P_m	P_E	P_m	P_E	P_m	P_E	P_m
a = 80%	ϕ1000	54	49	58	57	69	68	67	66	67	67
	ϕ500	74	68	85	75	83	83	85	82	89	84
a = 60%	ϕ1000	95	91	127	111	131	129	109	104	119	104
	ϕ500	111	107	125	119	133	131	127	124	134	132

The figures in this table show if necessary that the hypothesis we made to prove that $P_E > P_m$ are usually far from strong, although we cannot exclude the possibility of finding some months with $P_E < P_m$ in the future.

We must finally point out that the difference between P_E and P_m does not tend to zero when the number of cases n tends to ∞. The amplitude of $P_E - P_m$ is more or less connected with the differences between the best and worse cases.

This problem although serious is not dramatic since the differences are of the order of 0 to 15% between P_E and P_m. It can become critical when looking at differences between models as we try to explain in the next paragraph.

2. PREDICTABILITY DIFFERENCES BETWEEN TWO MODELS

We consider again an ensemble of n cases, but run in parallel by 2 models A and B. We can then define 2 sets of predictabilities.

A_i , i=1, n for model A

B_i , i=1, n for model B

$$A_E = \frac{1}{n} \sum_{i=1}^{n} A_i \qquad B_E = \frac{1}{n} \sum_{i=1}^{n} B_i$$

The differences between the 2 models can be evaluated for each case by

$$C_i = A_i - B_i$$

then $C_E = \frac{1}{n} \sum_{i=1}^{n} C_i = A_E - B_E$

Until now due to the wrong interpretation of mean predictability they were evaluated as

$C_m = A_m - B_m$ and small differences between

A_E and A_m , B_E and B_m can produce relatively considerable differences

between C_E and C_m.

This is due to the fact that

1. The differences between A_E and A_m or B_E and B_m are not constant in amplitude as seen from Table 1.

2. The amplitude of these differences is of the same order as the differences we are studying between 2 models.

If models A and B are considerably different it does not really produce large differences to the results, but if models A and B are fairly close (that is the case at present, and we can expect to be looking at closer and closer models in the future), looking at C_m instead of C_E can produce a completely wrong impression.

We can give here an example

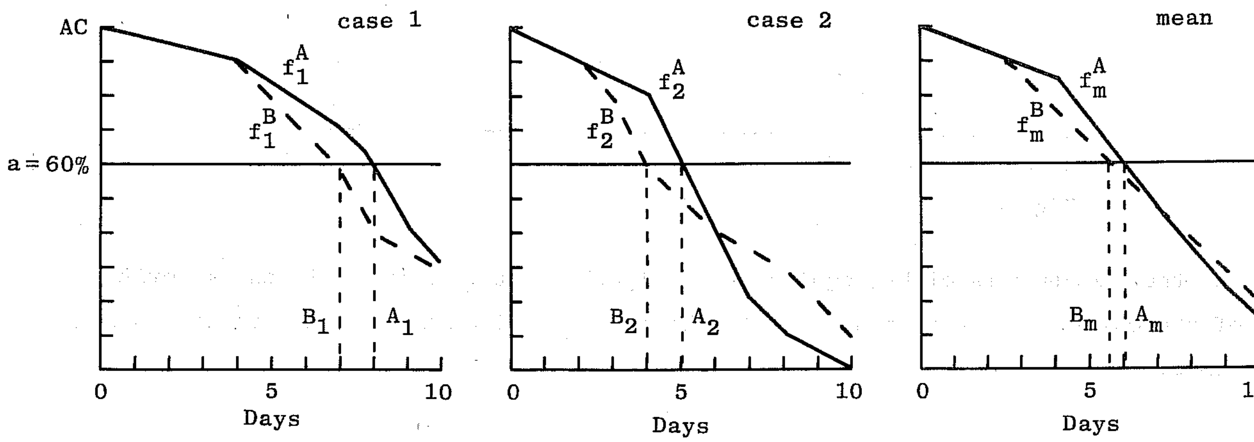


Fig. 2 Anomaly correlations as a function of time

In these 2 cases we have

$A_1 = 8$	$A_2 = 5$	$A_m = 6$	$A_E = 6.5$
$B_1 = 7$	$B_2 = 4$	$B_m = 5.5$	$B_E = 5.5$
$C_1 = 1$	$C_2 = 1$	$C_m = .5$	$C_E = 1.$

So here the mean of 2 improvements of 1 day is an improvement of ½ a day! Such a difference is far from exceptional as seen from our experience.

In practice, if models A and B are sufficiently close together on ensemble E, C_E might well be > 0 and C_m < 0, which means that a model which is slightly better might appear as slightly worse!

On the other hand, one can see that if the variance for model A is bigger than for model B on ensemble E, $C_E - C_m$ is likely to be bigger (and > 0) since we pointed out in Section 1 that the larger the variance, the more A_E was bigger than A_m.

This is usually the case in practice when model A is significantly better than model B.

The usual repartition is as in Fig. 3.

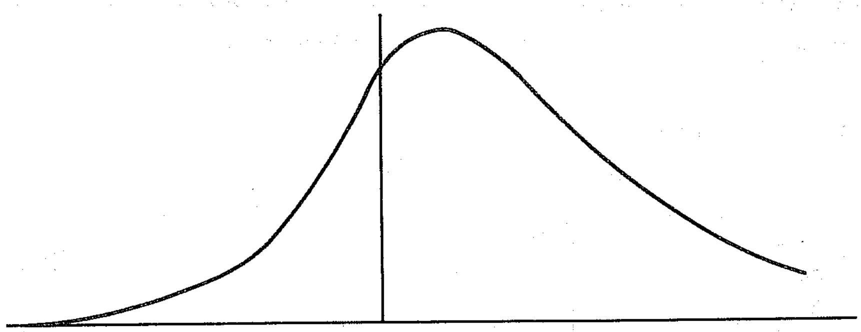


Fig. 3

In other words A is either equivalent, slightly better or worse in the majority of the cases, and it is significantly better more often than it is significantly worse.

Therefore taking P_m instead of P_E produces a spurious smoothing of the differences between models. In the next section we try to make some proposals in order to get more correct information on these differences

3. PROPOSALS

We first consider an ideal situation and we then try to propose ideas to adapt this ideal situation to real cases and to remove the problems arising from them.

3.1 Ideal situation

Let us consider again our ensemble E of n cases and anomaly correlations defined by

$$\{C_i = C_i(t) \quad , \; i = 1, \; n\}$$

if we assume $g_i(t) = \dfrac{dC_i(t)}{dt} < 0$ (10)

then all the C_i functions can be inverted.

Let $h_i = C_i^{-1}$

Instead of drawing the curve $C_i = C_i(t)$ we then draw the curve

$$h_i = h_i \; (c)$$

and $P_i = h_i$ (a) according to the choice of Section 1.

If now, instead of averaging the C_i to get $C_m = \frac{1}{n} \sum_{i=1}^{n} C_i(t)$ we average the

h_i to get $h_m = \frac{1}{n} \sum_{i=1}^{n} h_i(c)$

we get immediately P_E by

$$\boxed{P_E = h_m \quad (a)} \tag{11}$$

and in the case of 2 models A and B, we get $C_E = A_E - B_E$ by just looking at the graphs of h_m^A and h_m^B.

Unfortunately a lot of problems arise in practice.

3.2 Real situations

The first difficulty is that when t varies from 0 to 240 h C_i doesn't vary from 1. to -1., but from 1 to m_i with m_i varying from run to run between 60 and -40%.

That means that the area of definition of the h_i is not $[1, -1]$, but $[1, m_i]$, and we have to restrict ourselves to $\bigcap_{i=1}^{n} [1, m_i] = [1, m]$. Unfortunately (or fortunately from another point of view) m can be for some months bigger than 50% or even than 60%. Hopefully these cases are exceptional but they can be expected to be less exceptional with FGGE data, or in the (remote) future. Therefore we will propose a solution to avoid to reject these cases without modifying the statistics too much.

We suggest proceeding in this way.

1. We restrict ourself to the interval $[1, .5]$ since we can consider that there is very little information left afterwards, especially if we deal with a few cases.

2. On this interval we invert the C_i functions and we draw the following graph of the h_i functions. When h_i is not defined on $[1, m_i]$ $m_i > .5$ we shall propose something later. At this point we have another problem: we assumed that $g_i = \frac{dC_i}{dt} < 0$, but this is not always the case. In fact on $t \in [0,240]$ it is almost never the case, since there is always places where the curve increases for a short while.

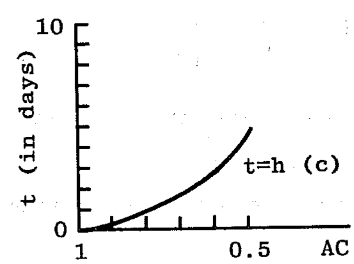


Fig. 4

We can easily get round this problem by defining h_i in the following way:

$$\forall c \in [1, .5]\ t = h_i(c) \text{ is defined as } \inf \{t_j : t_j = h_i(c)\}$$

This gives a unique definition for h_i and we already use it in practice when we consider as suspicious scores getting better after a period of decrease as shown in Fig. 5

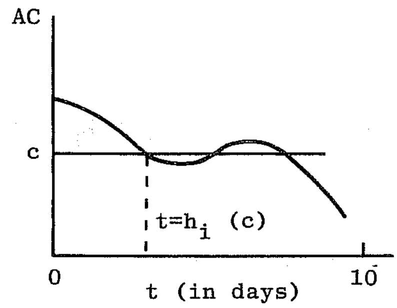


Fig. 5

The inversion of the C_i functions is then easy in theory, but it may be less easy in practice.

A straightforward method would be to make a scan with lines 1, .95, .90,......... .60, .55, .50, and for each of them find the intersection point with the usual curve, and then to join these points by straight lines, which would probably not be too expensive.

We come now to the problem of cases where C_i is still bigger than 50% (or more) at the end of the period of 10 days.

The first possibility is of course to run for 11 days if possible and we will get rid of the problem in almost all cases, but it is not feasible in many cases.

So we propose to assume after day 10 a kind of average decrease (linear, quadratic, ...?) which should not modify the means if there is a sufficient number of cases, and will not affect at all the resulting curve on $[1, m]$, which is the area of definition common to all the h_i

$$([1, m] = \bigcap_{i=1}^{n} [1, m_i]) .$$

Anyway, the same problem arises at present when you want to compare the increase in predictability of 2 cases which never reach the 60% after 10 days! The only clear comparison is what happens before.

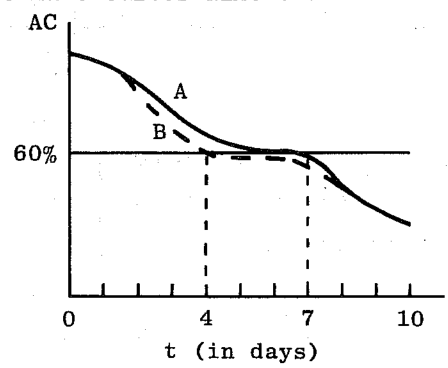
Therefore, even if an extrapolation method is used as proposed above, we suggest printing $m = \inf_{i=1,n}(m_i)$ on the graph.

The problem is of course more important for the long waves than for the total, or for the medium waves.

Once all this has been done, we can average the h_i curves and we get a mean curve which is a kind of predictability curve.

When taking 60% or 55% or 50% as a limit of predictability, one could object that it does not represent much if we have curves like the one in Fig. 6

Fig. 6



although runs A and B are very similar, a difference of almost 3 days in predictability appears between them, which is obviously a large overestimate. We can reply :

1. that happens on rare occasions

2. even with only 4 or 5 cases this is smoothed and it is likely over a large number of cases that it happens alternatively in favour of model A and B.

But all this disappears if we use our curves h_i.

If such a case is to influence $h_m(c)$ at the particular point C = 60%, it won't influence h_m at 65% and 55%. Therefore one can immediately see from the complete curve if a mean difference of 12 h between 2 models A and B is just due to some particular lucky or unlucky case or if it is consistent with the neighbour points.

One can also better evaluate the difference between 2 models in the first days of the runs. In Fig. 7 we present as an example the curves for the 7 February 76 cases run by the grid point model (ECM) and 2 reruns of the spectral model (T63 and T40).

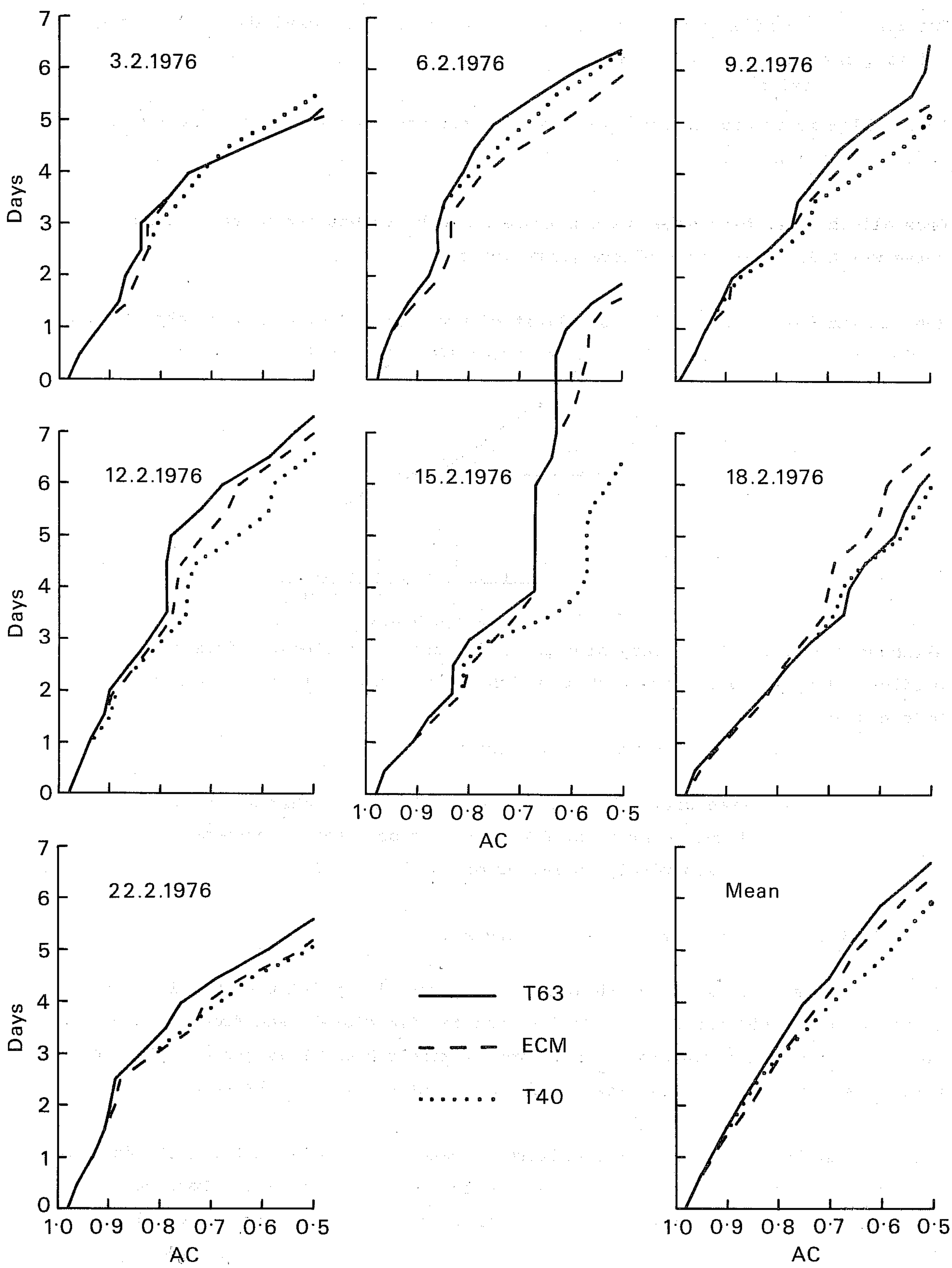


Fig. 7 Time needed to cross a certain level of anomaly correlation for the 1000 mb height field.

4. CONCLUSIONS

After having defined more precisely what we think should be the mean predictability and predictability differences between 2 models, we have made some proposals in Section 3. Of course this later section is more subjective and open to criticism. But its purpose is mainly to provide some elements for a discussion which could, we hope, emerge on an agreement on better solutions.