KONINKLIJK NEDERLANDSCH METEOROLOGISCH INSTITUUT

DE BILT (NEDERLAND)

No. 125

MEDEDEELINGEN EN VERHANDELINGEN

SERIE B

DEEL I, No. 2

★

DR W. BLEEKER

THE VERIFICATION OF WEATHERFORECASTS

INDEX DECIMALIS 551.509

★

TE VERKRIJGEN BIJ DE: TO BE PURCHASED FROM:

RIJKSUITGEVERIJ / 'S-GRAVENHAGE

1946

PRIJS F 1,—* PRICE F 1,—*

INTRODUCTION

The problem of the verification of weather forecasts has often been dealt with in meteorological literature. Almost every meteorological Service now seems to have its own checking method. The Dutch checking system was elaborated by Van Everdingen about 1905 (46), it was revised about 1910 (47) and remained in use unchanged up to 1940 [1]). During the last war, when the Netherlands weather service was stopped, the question has arisen as to whether, when the normal synoptic activities are resumed, this checking method should be maintained or whether it should be substituted by a different one, the scheme of which was laid down in 1939. This question led to a careful study of the literature on this subject, and the results of this study may be of general interest.

In examining the checking methods described below, we constantly come up against the problems which concern the relation between the public and the meteorologist and numbers of other matters, as the use of terminology, concerning which, therefore, we have some words to say.

A. Objectivity and subjectivity

The relation between the public and the forecaster is often disturbed by the contrast between objectivity and subjectivity (Schmausz, 30). This relation is governed by the terminology used by the forecaster and the interpretation applied by the public to certain terms in the forecast. The forecaster tries to formulate his forecasts in terms as *objective* as possible, but these terms are interpreted *subjectively* by the man in the street. Only a very small proportion of the public tries to form an objective image of the intentions of the forecaster, and it is disappointing to find, even in this minority, how widely the conceptions differ concerning terms in common use. For instance in Dutch the term "heavily cloudy" expresses a cloudiness of 7/10—9/10, while a cloudiness of 10/10 is indicated by "overcast". These officially established definitions prove not to correspond to the ideas of the general public, many of whom imagine that the term "heavily cloudy" means a greater degree of cloudiness than "overcast". It proves, moreover, that the international expression "partly cloudy" is interpreted by some "cloudy for part of the day, clear for the rest of the time", while for others it means "half cloudy for the whole of the day" (Emmons, 10).

But let us assume that the forecaster succeeds in composing an objective forecast which cannot give rise to any misunderstanding; even than difficulties will be caused by the subjective interpretation of the public. The forecast usually contains a statement about the future value of five meteorological elements, but the public wants to know, what the *weather* is going to be like. The general idea "weather" is an extremely complicated matter; it is namely, the integral of the influence of various meteorological factors upon our senses "With the eye", says Schmausz (4), "we observe the colours of the sky, the clouds, the electric discharges, the precipitation. The ear tells us something of the sounds of wind and thunder, our feeling-sense something about the temperature and our skin about the moisture".

Every meteorologist will moreover agree with Schmausz that it is very difficult to form a picture of the weather on a past day by combining the elements published in a year-book. This can only be done in certain cases, for instance for fine days. The reader of the forecasts inevitably has the same difficulty when he tries to form an idea about the future weather from the *objectively* given elements of a forecast which he will experience *subjectively*. This subjectivity is a great handicap for the forecaster, especially if he is serving different groups of the public. Seamen, farmers, townpeople, etc., all are interested in different elements of the forecast; they apply different standards and critisize from different points of view.

---

[1]) The numbers in brackets refer to the bibliography at the end of this article.

Under these circumstances the forecaster is often tempted to throw away all objective terms and to ask in angry mood if subjective phrases as "fine weather to-morrow too" would not best satisfy the majority of the public. Other subjective expressions are: cloudy, rainy, oppressive, raw, changeable, or steady. But it seems that in the long run the public does not altogether appreciate these familiar terms either (F r e y b e, 11).

If, therefore, we come to the conclusion that an objective terminology is preferable to a subjective one, it is necessary that the forecasters should define the terms they use very distinctly. Definitions of this kind have been made by many Weather Services (the American Signal Service in 1888, the Hamburger Seewarte in 1878 and 1903 (5), the Royal Dutch Meteorological Institute about 1910 (7), etc. The great drawback of such definitions is that they bind the forecasts in a kind of straight-jacket; year in year out, the nomenclature remains the same and newly found terms cannot be used as long as their significance has not officially been made known (F e r r a z, 48). It may be remarked that some meteorologists (E m m o n s, 10) would like to have the meaning of the objective terms printed daily in the papers, so that the public would always have the "code" to hand. We once made an experiment in this direction in cooperation with one of the important Netherlands daily papers; a "translation" was given every day beside the forecast. At first it was appreciated, but after some time it became tiresome and finally it appeared rather ridiculous. The public could not bear so much objectivity! It might be possible to find a medium way and publish a list of terms once a week.

In defining objective terms we should consider whether the meaning, attached to them, will be recognised by the public (F r e y b e 26, L a n d s b e r g 15), and here, with every effort at objectivity, subjectivity must intrude, giving rise to all sorts of difficulties. For instance, a scale in which the significance of the indications for windforce, light, moderate, fresh, strong and gale can be expressed in m/sec is required. It proves however, that the seamen and landmen mark the limits of the various terms at completely different values. The forecaster reserves the term "overcast" for conditions in which the cloudiness is 10/10, but for the public it makes great difference if the clouds are of the Cs- or the Ns-type. If the precipitation conditions are expressed in terms such as "dry", "light rain", "moderate rain" and "heavy rain", where do we draw the limits? Should they be chosen variable in connection with the season, as D i n i e s does (42)? In temperature-forecasts the question arises when the terms "not much change", "warmer" and "colder" should be applied and whether the chosen limits should be dependent upon the time of the year (H a z e n, 27). Moreover there is the question of whether these terms are to be applied to the temperature over the whole 24 hours or to the day time only; some consider that only the change in maximum-temperature should be indicated by such expressions. There are some services which do not give the daily changes, but the deviations to be expected from the normal temperature, and here again, of course, the problem of limits arises.

F r e y b e (11) and L a n d s b e r g (15) made some interesting experiments. After they had established a very acceptable terminology, they asked various groups of persons to underline in a questionnaire the terms which applied to the present weather conditions. We give L a n d s b e r g's interesting specified results in the following table:

Element	Percent of answers corresponding to proposed terminology
Sky aspect	62
Wind	40
Weather	57
Temperature	69

The low figure for the wind is remarkable. A striking case was that in which only 2 % of the persons asked gave the correct change of temperature. The subject was a clear morning with bright sunshine on which the temperature was 5,1° F lower than the day before. Most of the people underlined "warmer", drawing a cheque upon the future, and it turned out that the maximum was actually 12° F higher.

The results of these investigations are not very encouraging. Landsberg choose for his experiments freshman students in English Composition. It would certainly be worth while to repeat these experiments with other groups of people.

Moreover, even when we have decided to make use of objective terms, we may indulge in long discussions upon a less or greater latitude in the formulation of the forecasts. One meteorologist, for instance, will forecast "moderate to fresh southwesterly to northwesterly wind", while another will express the same situation with "moderate southwesterly wind".

A much discussed subject is the use of „doubt-terms" (possibly, probably, likely). Some are of opinion that they cannot be avoided, while others defend the opinion that their use should be avoided as much as possible as it is the business of the forecaster to estimate the chances, not of the public.

Van Everdingen (46) who investigated the results of forecasts with doubt-terms demonstrated that predictions of rain with "possibly" were 60 % correct, while those with "probably" were 80 %. Landsberg (15) defined "possible" as 33 %, "probable" as 67 % and "likely" as 80 %. It is interesting to note that American military forecasters serving in France during 1914—1918, added notes to their forecasts on their reliability (8). Hallenbeck (12) tells that the farmers in his district requested him to give indications of the reliability of his precipitation forecasts; he complied with this request by adding to the forecasts the estimated chance of rain in a numerical value.

Schmausz, who considers the form in which the forecast is given to be of the utmost importance (4,30) makes some interesting comments on these question, which we give below in a free translation. "A weather forecast is always a hypothesis and not the result of computations. In forming a hypothesis people involuntarily fall into two groups; those who consider to greatest possible caution necessary, because they consider that they do more harm to the authority of the weather-service if they publish incorrect forecasts than if they express themselves with reserve, and those who don't mind at all if they occasionally make complete mistakes if in other cases they obtain a brilliant success by their boldness".

By analogy with Ostwald's classification of natural scientists, Schmausz calls the people who belong to the first group the "classicals", those who make up the second group the "romantics". In our opinion these denominations are not very happy. We should call them the "cautious" and the "bold". It is easily understood (Chromow, 2) that even with objective criticism the forecasts of the "cautious" gain higher marks than those of the "bold", who despise "rubber-forecasts" (Schreiber, 20). It is therefore not impossible that a checking system may to some extent force the hand of the forecaster.

The above classification raises the question whether all forecasters in a team should not be of a uniform type, that is all "cautious" or all "bold"? And further if both the air field and the military forecaster should not always be of the bold type?

It would certainly be worth while to examine these and other questions thoroughly, but it is not the purpose of this paper to expatiate upon the terminology and the psychology of weather-forecasts and weather-forecasters.

B. Is there any advantage in checking the weatherforecasts?

The question may be put: "Should the weather-forecast be checked or not?" Many meteorologists (Defant and others, 3) consider that the calculation of testing results is a waste of time; the hours devoted to it might be more usefully employed in a thorough study of the synoptic situations which led to the erroneous forecasts. Others again are great champions of checking, for which they conceived more or less effective systems, which they propagate on various occasions (Heidke, 52—60).

At any rate it was natural that during the latter half of the last century the necessity was felt for systematic checking of the forecasts. In the first place the official institutes had to compete with persons who more or less blindly issued forecasts upon astrological or other grounds which often found greater acceptance with the public than the forecasts of the official services. The meteorological institutes could not be better justified than by comparing the checking-results of official and blindfold [1]) forecasts. A comparison was often not so easy because of the terminology used, which was partly subjective and partly objective.

The difficulty could be overcome with the help of a group of people who judged all forecasts subjectively. But it soon became clear that a good check was only possible when the forecasts were expressed in objective terms. The comparison of the objectively expressed official forecast with the subjectively expressed blindfold forecasts was abandonned, the group of "jurymen" became superfluous, it was replaced by instruments.

A second consideration that led to the construction of objective tests was the strife that arose during 1875—1900 between the champions of centrally issued forecasts and the adherents of the system of local forecasts. In America and Germany this struggle led to very unedifying polemics.

It appears that it was in America that the first efforts were made to estimate the results of the forecasts in figures. At first these figures appear to refer to the entire forecast, later to the various components (van Bebber, 21). In Germany, where the forecasts and the gale-warnings of the Deutsche Seewarte were often unfavourably criticised, the usefulness of a good *objective* check was first realized by Köppen (62). He defended his opinion in a speech at the foundation meeting of the Deutsche Meteorologische Gesellschaft (Hamburg, 1884). About 20 years later he made his further ideas upon this subject known (63). Köppen's ideas were developed by Heidke (52—60), according to whom the testing of weather-forecasts is desirable because:

a. The meteorological service can justify its existence to the Government by means of figures.

b. The meteorological service can justify its existence to the public by figures.

c. The results are of use in fixing the limits of forecasting districts.

d. By means of the testing figures it could be seen whether the weather conditions at a station are subject to local influences.

e. By means of the testing results the earliest moment at which a forecast can be issued with a reasonable chance of success could be estimated.

f. The value of new theories can be tested by the checking results.

g. The forecaster will take a more critical view of his own work.

This list seems rather exaggerated; at present, for instance, it is no longer necessary to demonstrate the value of the weather service by high checking results (Schmausz, 31 and Wagemann, 33). As remarked by Shaw during a discussion at the Royal Meteorological Society (Dobson, 43) low checking results do not necessarily prove that the weather service is a useless institution; it is not impossible, for instance, that an important change of weather can be forecast with certainty and that a forecast of this kind may amply compensate for a number of cases in which small changes were given erroneously (see Clayton, 25). Moreover a forecast which may prove correct according to the accepted testing system is not necessarily the most useful forecast (Mahrt, 16). In this connection we may observe that in some checking systems (Heidke, 52—60) different weights are given to different kinds of forecasts in connection with their value to the public. The fixation of these weights is again a question of subjectivity and therefore leads to great difficulties.

[1]) It should be observed that the non-official forecasts are by no means all completely blindfold (Heidke, 54; Thomas 68). The nonofficial forecasts are often based upon climatological data. But for the sake of simplicity we shall include all forecasts which are not prepared on synoptic grounds as blindfold forecasts.

The limits of forecasting districts by no means need to be defined by the testing results; they can also be determined by simple climatological determinations (von Bebber, 1). Moreover, every experienced meteorologist will know from his daily practice which stations are under local influences.

At the same time, speaking generally, it cannot be denied that Heidke's remarks are of some value. We too consider, especially as regards point *g* of Heidke's list, that it is useful to check the forecasts objectively, while keeping in mind the disadvantages which adhere to that checking. At the same time we regard as a great hindrance the straight-jacket into which the forecasts are forced on account of the checking too, while the rivalry between the forecasters will contribute to make the cautious to use terms which will yield to highest checking results but will give the most inadequate information to the public.

In order to avoid these difficulties, in a number of institutes (Heidke 28, Thomas 68) double forecasts are always compiled, one intended for the public which is circulated in the usual way, the other is intended for testing and is inscribed in a special book for this purpose. The two forecasts must, of course, never be in contradiction with each other. By following this method, therefore, the forecaster has complete liberty in the choice of his terms in the officially circulated forecast, while forecasts intended for internal use, which may be expressed in very guarded terms are used for verification purposes. The system of double forecasts was introduced in Holland in 1939 for the rainfall forecasts, to the great satisfaction of the acting meteorologists.

C. Objective checking systems

If we come to the conclusion that it is desirable to check the weather-forecasts on objective lines the question follows as to what system to employ. It is clear that each element must be judged separately and that moreover the exact meaning of the terms used for the various elements must be fixed. Further, the time of validity of the forecast must be accurately defined and the results will only be of value when they are based upon a large number of "observations".

In setting up a sytem of verification we must not want the utmost perfection. Clayton (25) once remarked that in a checking system there ought really to be included the nature of the phenomenon, time of appearance, duration, intensity, importance to the public and moment of compilation of the forecast. These demands are too high; for the present we shall have to be content with more modest systems.

The checking systems may be divided into several groups, namely:

I. The objective systems constructed for meteorological elements for which in the forecasting terminology only two alternatives are laid down.

II. The objective systems which are suitable for meteorological elements for which three or more forecasting terms are available.

III. The objective systems which are constructed for meteorological elements which are expressed in numerical values in the forecasts.

IV. Semi-objective methods.

I. The objective systems constructed for meteorological elements for which only two alternatives are provided in the forecasting terminology

These systems are connected with elements, for which the forecaster has to chose either one term or the other, for example:

gale	—	no gale
dull (cloudiness 6/10—10/10)	—	sunny (cloudiness 0/10—5/10)
precipitation	—	dry

thunder	—	no thunder
rising temperature	—	falling temperature
temperature above normal	—	temperature below normal
nightfrost	—	no nightfrost.

If the forecast is in accordance with the left-hand column we may call it a positive prediction, if in accordance with the right-hand column it will be called a negative forecast or a non-prediction. With gale, thunder and nightfrost the non-prediction consists of simply not mentioning the phenomenon.

If the phenomenon appears in the manner indicated in the left-hand column we call it a positive occurrence; if it appears as indicated on the right we call it a negative occurrence or a non-occurrence.

In a series of s cases by means of which the meteorological predictions are to be checked, we distinguish:

p' positive predictions, $(s - p')$ non-predictions,

o' positive occurrences, $(s - o')$ non-occurrences.

The total number of cases in which a positive prediction is fulfilled being taken as c' (number of positive coincidences), we have (fig. 1) after Gilbert (50) four groups:

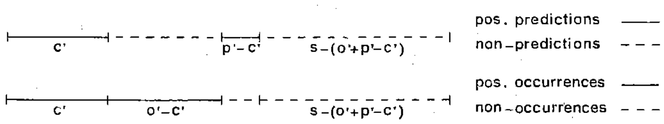

Figure 1.

c'	cases in which a positive occurrence was indicated by a positive prediction (positive coincidences),
$(o' - c')$	cases in which a positive occurrence was preceeded by a non-prediction,
$(p' - c')$	cases in which a non-occurrence was preceeded by a positive prediction,
$s - (o' + p' - c')$	cases in which a non-occurrence was preceeded by a non-prediction (negative coincidences).

It is important to keep these groups well in mind while following the discussion of various systems.

We now will investigate several checking systems, putting the following questions:

1. What value does the formula, deduced for the checking system, yield if all the predictions are correct, that is when $p' - c' = 0$ and $o' - c' = 0$, thus $p' = o' = c'$? In a good checking system this value should be independent of $\frac{o'}{s}$.

2. What value does the formula, deduced for the checking system, yield if all predictions are erroneous, that is when $c' = 0$ and $s - (o' + p' - c') = 0$, thus when $c' = 0$ and $s = o' + p'$? This value must also be independent of $\frac{o'}{s}$.

3. How high is the most probable value which the checking system yields if the p' positive predictions and the $(s - p')$ non-predictions are scattered by chance over the o' occurrences and the $(s - o')$ non-occurrences, thus when $c' = \frac{o'p'}{s}$ holds for the number of positive coincidences? An ideal testing system should, in our opinion, yield a value 0 (zero) in such cases.

4. For every checking system we should further inquire (Gilbert, 50), if it is capable of inversion, that is, whether the coefficient of verification found can bear exchange of positive occurrences and non-occurrences and therefore also of positive predictions and non-predictions. Making a positive prediction (e.g. rising temperature) costs the forecaster just as much trouble as making a non-prediction (e.g. falling temperature) and the result for the coefficient of verification should be the same, independent of whether we try to

forecast rising temperature correctly or falling temperature. If in the inversed system the number of positive predictions is given by p'', the number of positive occurrences by o'' and the number of positive coincidences by c'', than the following formula's hold:

$$c'' = s - (o' + p' - c')$$
$$p'' = s - p'$$
$$o'' = s - o'.$$

In our opinion an ideal checking system must be capable of inversion, that is to say the coefficient of verification with single dotted symbols (o', p', c') must deliver the same value as the coefficient with double dotted symbols (o'', p'', c'').

5. Further, for each checking system we shall enquire whether it gives the "cautious" forecaster opportunity to set up a series of predictions with which he can attain a high coefficient of verification, but by which he may supply the public with inadequate information. Such checking systems will be entirely repudiated by the "bold" forecaster and they will never stimulate synoptic meteorology to a further development. Moreover they furnish no opportunity for forming an opinion concerning the limits of the forecasting zones, the local influences and the earliest time, at which the forecasts can be issued. In a good checking system it should never be possible, independent of the value of $\frac{o'}{s}$, to gain high figures by tricky making of forecasts, which give high checking figures but a poor information to the public.

6. In the table on page 23 we give a scheme of a number of checking systems. The table includes the results which were calculated from a series of tornado predictions issued by Finley (49). On 100 out of the 2803 occasions during a certain period these predictions have been issued; during the 2803 forecasting periods 51 heavy storms occurred and it proved that 28 warnings were fulfilled. Therefore $s = 2803$, $p' = 100$, $o' = 51$ and $c' = 28$.

a. Finley's method

Finley calculated the "hit-percentage" of a series of meteorological predictions from the quotient of the sum of successful positive and non-predictions and the total number of predictions.

His coefficient of verification is therefore:

$$v'_F = \frac{c' + \{s - (o' + p' - c')\}}{s}$$

The value of v'_F varies between 1 for complete success and 0 for complete failure. This method is capable of inversion, as

$$v''_F = \frac{c'' + \{s - (o'' + p'' - c'')\}}{s} = \frac{c' + \{s - (o' + p' - c')\}}{s}$$

There are serious objections to this method. Gilbert (50) has pointed out that v'_F is à priori high, if the number of non-occurrences is high and the "cautious" forecaster keeps the number of non-predictions high, as the term $s - (o' + p' - c')$ has than a high value. The value of v'_F will be also à priori high, if the number of occurrences is high and the "cautious" forecaster issues many positive predictions; in such cases c' has a high value.

In the first case (e.g. with gale warnings) a high value for v'_F can be attained by issuing some positive predictions at random moments. If a positive prediction is never issued $v'_F = 1 - \frac{o'}{s}$ and this nearly equals 1, if $\frac{o'}{s}$ is small.

In the second case (e.g. with rain forecasts during the wet season in the tropics) a high value for v'_F can be obtained by giving frequent positive predictions. If a positive prediction is constantly issued, $v'_F = \frac{o'}{s}$ and this nearly equals 1. It is quite true that

Finley's method really does give the hit-percentage of successful positive and negative predictions. But the figures give too high an impression of the capabilities of the forecaster. The following confirms this.

The most probable value for v'_F which will be attained by a fortuitous issue of p' positive and $(s - p')$ non-predictions is:

$$v'_F = 2\frac{o'}{s}\frac{p'}{s} - \frac{o'}{s} - \frac{p'}{s} + 1$$

This value may be fairly high; for the American tornado-predictions it is 0,947. It would be better if the value of the coefficient of verification for such randomly issued forecasts was 0.

The idea rises to improve Finley's method by giving certain weight to the successful positive predictions and non-predictions according to the ratio of non-occurrences to positive occurrences (for instance in the ratio $\frac{s - o'}{s} : \frac{o'}{s}$). We then find:

$$v'_{FW} = \frac{c'\frac{s - o'}{s} + \frac{o'}{s}\{s - (o' + p' - c')\}}{c'\frac{s - o'}{s} + (p' - c') + (o' - c') + \frac{o'}{s}\{s - (o' + p' - c')\}}$$

The value of v'_{FW} lies between 1 and 0; the checking method is invertable and the "cautious" forecaster is unable to abuse the checking method. But the difficulty remains that in the random issue of p' positive predictions and $(s - p')$ non-predictions the most probable value of v'_{FW} differs from zero and (for the american tornado-prediction it is even 0,253).

b. Heidke's checking system (I)

Heidke (52—60), like many other writers, has pointed out that a coefficient of verification does not represent the skill of the meteorologist. He would like to determine the quality Q of the forecasts by

$$Q = \frac{E - B}{1 - B}.$$

Here E is the coefficient of verification computed by means of some checking instruction, B the coefficient of verification according, the same instruction calculated for a series of blindfold forecasts. Heidke chooses for B the result of a series of forecasts made on the principle "the weather to-morrow will be the same as to day".

Heidke's method might be made suitable for phenomena with two alternatives in the following way. E might represent the coefficient of verification according to Finley, B the most probable value for this coefficient if the p' positive and the $(s - p')$ non-predictions were issued at random, then i'_H, the ratio of success in inference or "skill score", can be represented as:

$$i'_H = \frac{\frac{c' + s - (o' + p' - c')}{s} - (2\frac{o'}{s}\frac{p'}{s} - \frac{o'}{s} - \frac{p'}{s} + 1)}{-2\frac{o'}{s}\frac{p'}{s} + \frac{o'}{s} + \frac{p'}{s}} = \frac{2(c's - o'p')}{o'(s - p') + p'(s - o')}.$$

The value of i'_H varies between 1 with complete success and $-\frac{2o'(s - o')}{o'^2 + (s - o')^2}$ with complete failure. It is undoubtedly a disadvantage that this ratio of success in inference has a changeable "zero-point". On the other hand i'_H stands inversion, the "cautious" forecaster cannot get high figures by tricky methods based on the checking system, and the most probable value of i'_H equals 0 with a random issue of p' positive predictions and $(s - p')$ non-predictions.

c. Gilbert's method (I)

Gilbert (50) considers that the ratio of verification should be calculated in the following manner. The number of successful positive predictions is c', the number of unsuccessful positive predictions is $(p' - c')$, the number of failures where a positive occurrence was not preceeded by a positive prediction is $(o' - c')$. The ratio of verification v'_G is now formed by the quotient of the number of successful positive predictions and the sum of this number and the number of failures, thus

$$v'_G = \frac{c'}{c' + (p' - c') + (o' - c')} = \frac{c'}{p' + o' - c'}.$$

The value of v'_G varies from 1 to 0. The method is not capable of inversion for

$$v''_G = \frac{c''}{p'' + o'' - c''} = \frac{s - (o' + p' - c')}{s - c'} \neq v'_G.$$

A forecaster, who always issues positive predictions obtains a value $v'_G = \frac{o'}{s}$. If $\frac{o'}{s} \gg \frac{1}{2}$ the "cautious" forecaster, who always gives positive predictions, will attain a high value. There are conditions thus under which the forecaster can abuse this method.

A further objection is that the most probable value for v'_G with random issue of p' positive and $(s - p')$ non-predictions is different to zero, namely

$$v'_G = \frac{o'p'}{s\,(p' + o') - o'p'}.$$

d. Gilbert's method (II)

Gilbert (50) has further remarked that the skill of the forecaster is not determined by the number of positive coincidences c' but by the difference of this number and the number of positive coincidences $\frac{o'p'}{s}$ that may be reached by the random issue of p' positive predictions and $(s - p')$ non-predictions. If then, to some extent in analogy with the method of Gilbert described above, we again represent the ratio of success in inference i'_G by the quotient of what the forecaster attains by skill and the sum of this and the failures, then:

$$i'_G = \frac{c' - \frac{o'p'}{s}}{(c' - \frac{o'p'}{s}) + (p' - c') + (o' - c')} = \frac{c's - o'p'}{(o' + p' - c')\,s - o'p'}.$$

The value of i'_G varies between 1 and $\frac{-o'(s - o')}{s^2 - o's + o'^2}$.

The method is capable of inversion, furthermore it does not give the "cautious" forecaster an opportunity to gain high figures by tricks. Finally i'_G by a random issue of p' positive and $(s - p')$ non-predictions has the value 0 as the most probable value. However the method is not to be recommended, as it is a great drawback that the "zero-point" is dependent upon $\frac{o'}{s}$.

e. The Dutch method

In verifying the Dutch gale warnings a method has been followed so far, in which the ratio of verification of the positive predictions was computed; thus

$$v'_p = \frac{c'}{p'}.$$

At the same time it was stated how many gales had occurred without a previously warning being given, that is how great $(o' - c')$ had been in the discussed period.

f. Doolittle's method (I)

Doolittle (44) was the first to call attention to the fact that the reliability of a series of forecasts can be expressed as the product of the ratio of verification for the positive occurrences v'_o and the ratio of verification for the positive predictions v'_p. The ratio of verification is therefore according to Doolittle (I):

$$v'_D = v'_o \times v'_p = \frac{c'}{o'} \times \frac{c'}{p'}.$$

The value of v'_D varies between 1 and 0. Doolittle's method (I) proves to be incapable of inversion.

If $\frac{o'}{s} \gg \frac{1}{2}$ the "cautious" forecaster can gain high figures by constantly giving positive predictions, for then (owing to $p' = s$ and $o' = c'$) $v'_D = \frac{o'}{s}$.

By a fortuitous issue of p' positive and $(s - p')$ non-predictions the most probable value for v'_D will be equal to $\frac{o'p'}{s^2}$. Doolittle's method (I) is therefore not to be recommended.

It should be remarked that Grossmann (4), probably ignorant of Doolittle's paper, recommended the use of v'_o and v'_p separately for verifying weather forecasts. According to Grossmann the forecaster should endeavour to attain high values for v'_o as well as for v'_p. Grossmann also pointed out that v'_o and v'_p are linked together in a peculiar way. When positive predictions are issued too frequently v'_o will be high but v'_p low, while, when too few positive predictions are given, v'_o will be low but v'_p in any case higher than when many positive predictions are given.

One disadvantage of this method of Grossmann is that the reliability of a series of forecasts cannot be expressed in one figure.

g. Doolittle's method (II)

Considering that with an arbitrary issue of p' positive and $(s - p')$ non-predictions the number of coincidences is $\frac{o'p'}{s}$, the *skill* of the forecaster after Doolittle (44) cannot be expressed in the product of $\frac{c'}{o'}$ and $\frac{c'}{p'}$, but the ratio of success in inference i'_D must be laid down as the product of

$$\frac{c' - \frac{o'p'}{s}}{o' - \frac{o'p'}{s}} \text{ and } \frac{c' - \frac{o'p'}{s}}{p' - \frac{o'p'}{s}}.$$

Therefore is:

$$i'_D = \frac{(c's - o'p')^2}{o'p'\,(s - o')\,(s - p')}.$$

The value of i'_D varies between 1 and 0; the value of 1 is reached as well in complete success as in complete failure. According to Doolittle this is logical, because it is quite as difficult to compile forecasts which are always right as forecasts which are always wrong. The value 0 is attained as the most probable value by the forecaster who issues p' positive and $(s - p')$ non-prediction at random. Doolittle's method (II) is moreover capable of inversion; and does not offer any opportunity to the "cautious" forecaster to gain a high value for i'_D by tricks.

This method of Doolittle is very serviceable in our opinion; the only slight drawback is that we cannot see directly from i'_D whether many correct or many incorrect predictions have been made. This can only be done by determining the value (and thus the sign) of $(c's - o'p')$.

h. Lacour's method

Lacour (64) thinks the coefficient of verification should be calculated from the quotient of the probability on a positive occurrence in making a positive prediction and the probability on a positive occurrence in making a non-prediction. Therefore:

$$v'_L = \frac{c'}{p'} : \frac{o' - c'}{(o' - c') + \{s - (o' + p' - c')\}} = \frac{c'}{p'} \times \frac{s - p'}{o' - c'}.$$

The value of v'_L may vary between ∞ and 0; the method is not capable of inversion. Moreover the method has a precarious side for "cautious" forecasters, as has been pointed out by Heidke (52). The "cautious" can namely attain the highest value for v'_L by giving a non-prediction on a day that he can say with great certainty that a non-occurrence will take place, and by issuing positive predictions for all other days. In that case $s - (o' + p' - c') = 1$, while $c' = o'$ and $p' = s - 1$, therefore $v'_L = \infty$.

In the arbitrary issue of p' positive predictions and $(s - p')$ non-predictions, the most probable value for v'_L is:

$$v'_L = \frac{s - p'}{s - o'}.$$

There are, thus, sufficient reasons for rejecting Lacour's method.

i. Clayton's method

Clayton (40,41) expresses the forecaster's skill in the difference of the probability on a positive occurrence in making a positive prediction and the chance on a positive occurrence after a non-prediction. Clayton's coefficient of verification v'_C is thus:

$$v'_C = \frac{c'}{p'} - \frac{o' - c'}{(o' - c') + \{s - (o' + p' - c')\}} = \frac{c's - o'p'}{p'(s - p')}.$$

Pierce (67) deduces the same formula on a different manner; Clayton however gives it a wider application.

The value of v'_C varies between $+1$ and -1. A forecaster who arbitrarily issues p' positive predictions and $(s - p')$ non-predictions will get 0 as the most probable value for v'_C. The method is capable of inversion.

When $\frac{o'}{s} \ll \frac{1}{2}$ the "cautious" forecaster can reach a high figure for v'_C by giving one positive prediction at a moment when he is fairly certain that a positive occurrence will follow and for the rest giving non-predictions. Then $p' = c' = 1$, thus

$$v'_C = \frac{s - o'}{s - 1} = \frac{s}{s - 1} - \frac{o'}{s - 1},$$

and this value lies close to 1. For the tornado predictions one would have found in such case $v'_C = 0{,}982$.

If $\frac{o'}{s} \gg \frac{1}{2}$ and the forecaster issues one non-prediction, when he is pretty certain that a non-occurrence will follow and further issues positive predictions, then $s - (o' + p' - c') = 1$, $o' = c'$, thus $v'_C = \frac{o'}{s - 1}$, and this value also lies close to 1.

For this reason Clayton's method should be rejected.

It can be seen at once that Clayton's method does not represent the forecaster's skill correctly. Clayton only considers the positive predictions and starts from the assumption that in a series of good predictions $\frac{c'}{p'}$ must lie as near as possible to 1 and $\frac{o' - c'}{s - p'}$ as close as possible to zero. But it is necessary to consider the positive occurrence too.

For a series of good forecasts $\frac{c'}{o'}$ must also be as near to 1 als possible and $\frac{p'-c'}{s-o'}$ as close to zero as possible.

The forecasters skill is therefore determined by $(\frac{c'}{p'}-\frac{o'-c'}{s-p'})$ and by $(\frac{c'}{o'}-\frac{p'-c'}{s-o'})$ or by their product, which proves to be equal to the ratio of success of inference according to Doolittle (II):

$$i'_D = (\frac{c'}{p'}-\frac{o'-c'}{s-p'}) \times (\frac{c'}{o'}-\frac{p'-c'}{s-o'}) = \frac{(c's-o'p')^2}{o'p'(s-o')(s-p')}.$$

The value of Doolittle's formula increases once more by this derivation.

j. *Wallén's method*

Wallén (70) seems to be the first to suggest that the calculation of the coefficient of verification should be based upon the determination of a correlation coefficient. It is natural that there should be a relation between the predicted weather and that which actually occurs; it seems logical to express this connection in a correlation coefficient.

The following scheme expresses the relation between predictions and occurrences:

Predictions \ Occurrences	Positive	Negative	Total
Positive	c'	$p'-c'$	p'
Negative	$o'-c'$	$s-(o'+p'-c')$	$s-p'$
Total	o'	$s-o'$	s

The correlation coefficient and therefore the coefficient of verification v'_W is:

$$v'_W = \frac{c's-o'p'}{\sqrt{p'(s-p') \times o'(s-o')}}.$$

The value thus found is the root of the value found by formula Doolittle II. Wallén's method, therefore, possesses all the advantages of the Doolittle II method; moreover v'_W indicated directly how the phenomena are correlated.

In our opinion Wallén's system is a very useful one; it can also be used in group II of verifying systems.

II. Systems, which are suitable for meteorological elements for which three or more prediction terms are fixed

For most of the meteorological elements three or more prediction terms are fixed. The wind, for instance, is expressed by the scale: calm, light, moderate, fresh, strong, gale; the cloudiness by: clear, fair, partly cloudy, cloudy, overcast; the temperature by: colder, not much change, warmer or even more detailed scales.

It has not proved possible to transform any of the formulas described under I so that they give in a simple way one figure as a coefficient of verification or as a ratio of success in inference. It is possible to calculate for *each term* a coefficient of verification or a ratio of success of inference, by taking the forecasts in which these terms are used as positive predictions and further regarding all predictions in which other terms are used as non-predictions, and finally to take the phenomena as positive occurrences in which the limits of the established terms are not exceeded while the rest are regarded as non-occurrences.

In this way a coefficient of verification and a ratio of success in inference can be calculated for terms as: colder, somewhat colder, not much change, somewhat warmer, warmer, etc. It should be remarked that it is even permissible that the meanings of the terms overlap each other, so that, for instance, colder indicates a change of 2°—5° C; somewhat colder a change of 1°—3° C and not much change a variation of less than 2° C. The coefficient of verification or the ratio of success in inference for a meteorological element might then be defined as the product or as the mean of the values calculated for the separate terms.

Amongst the methods described under I we should again prefer Wallén's system, but it must not be forgotten that other formulas cannot be abused here so easily by "cautious" forecasters as when there are only two alternatives. This is the case, for instance, with Clayton's method (40,41). We borrow the following example (refering to 90 temperature predictions) from Clayton:

Temperature Number of days		Occurrences		
		Warm	Slight change	Cold
		%	%	%
Warmer forecasted	35	69	—	—
Slight change forecasted	30	—	63	—
Colder forecasted	25	—	—	84
Warmer not forecasted	55	29	—	—
Slight change not forecasted	60	—	35	—
Colder not forecasted	65	—	—	22
Forecaster's skill		40	28	62
Mean			43 %	

A few further systems given in the literature will be more completely described.

a. Köppen's method

Köppen (62, 63) and other writers (Clayton, 24) recommend judging the results of a series of predictions by a table in which predicted and actually occurring weather are reproduced. Köppen's intentions can be best demonstrated by the following example regarding temperature predictions in the summer of 1884 issued by the Hamburger Seewarte.

Prediction \ Occurrence	Colder	Slight change	Warmer
After 15 × colder followed	7	3	5
After 18 × slight change followed	4	10	4
After 22 × warmer followed	1	11	10

The maxima should all lie on the diagonal which runs from left above to right below. The better the predictions the better the maxima stand out. This method may be regarded as a first effort to form an idea of the correlation between predictions and the actual following weather. It would be possible to calculate a correlation coefficient from this table, Köppen has not done this; the correlation-theory was in its infancy at the time when his publications appeared.

As to the meaning of the terms used by the Seewarte: colder indicated a fall of temperature greater than 1° C, slight change a change less than 1° C and warmer a rise greater than 1° C.

Köppen's method cannot be used when the meanings of the terms overlap.

b. Ekholm's method

Ekholm (45) recommends a method which is no other than a slight extension of Köppen's method. He calculates the ratio's of verification from a table containing predictions and occurrences. An example borrowed from a series of Swedish wind and gale warnings will serve as illustration (the warnings were issued in three degrees, namely "not dangerous" for force 0—6 Beaufort, "be careful" for 7, 8 and 9 Beaufort, and "danger" for 10, 11 and 12 Beaufort.

Occurrence / Prediction	0—6	7	8	9	10	11	12	Total	Coincidences [1])	Ratio of verification on predictions
Not dangerous . .	739	39	36	18	7	0	0	839	739	88,1
Careful	80	60	61	37	27	6	0	271	191	70,5
Dangerous	5	7	6	26	25	8	1	78	73	93,6
Total	824	106	103	81	59	14	1	1 188	1 003	84,4
Coincidences [2]). . .	739	67	67	63	52	14	1	1 003	—	—
Ratio of verification on occurrences	89,7	63,2	65,0	77,8	88,1	100,0	100,0	84,4	—	—

The values of the ratios of verification are not combined to one number by Ekholm, who considers it to be an advantage of his system that it can be seen from the figures how the reliability of the gale warnings rises with increasing danger.

c. Klein's method

Klein (29) calculates for the verification of the forecasts the coefficient of the number of predictions that are actually completely successful and the total number of forecasts. Klein therefore only attends to the sum of the figures which occur in Köppen's system from left above to right below. It is evident that this method of verification yields a smaller result according as more terms are fixed for an element. This led to the construction of the half-objective methods, described under IV, in which a certain value is given to predictions which were not completely successful.

III. The systems constructed for meteorological elements which are expressed in figures in the predictions

It is, generally speaking, not usual to express the elements in figures in the normal predictions, although it is often done in the aeronautical predictions.

But when, as mentioned on page 7, separate predictions are prepared for the public and for verification it is advisable to give the latter in figures. In the various weather-services where this takes place verification-systems have been developed which are discussed below.

a. Thomas' method

Thomas (68) notes the predicted values for wind-velocity, cloudiness and temperature for 8 o'clock at the following day. Then a frequency curve is constructed from the differences

[1]) As coincidences all cases are taken in which after the predictions „be careful" and „danger" a force 7—12 Beaufort occurred.

[2]) All cases in which force 7—12 Beaufort was preceeded by a prediction „be careful" or „danger" count as coincidences.

between predicted and observed values. The form of this curve indicates the reliability of the predictions.

The reliability is not expressed by a figure; this could certainly best be done by the construction of the standard deviation $\delta = \sqrt{\frac{\Sigma\delta^2}{n}}$, in which δ represents the deviation and n the number of predictions.

b. Dobson's method

D o b s o n (43) recommends the following method for verifying upper wind predictions. For each prediction the difference is determined between the predicted and the observed velocity. These differences (no account is taken of sign) are arranged in order of magnitude. Each value is then plotted as a dot in a diagram; the ordinate of this dot is given by the magnitude, the abscissa by its numerical position in the above series. A smooth line is then drawn through the dots (*A* in fig. 2). In place of the numerical positions a percentage scale of 0—100 can be plotted along the abscissa (see fig. 2); by means of curve *A* than it can be read, how often (in percentage of the number occasions) the differences were smaller than a given value.

D o b s o n draws a second analogous curve (*B* in fig. 2) on his diagram, which gives the same relation for predictions in which the wind would have been forecast to remain constant. This curve *B* thus gives an idea on the variation of the wind during the validity period of the forecast.

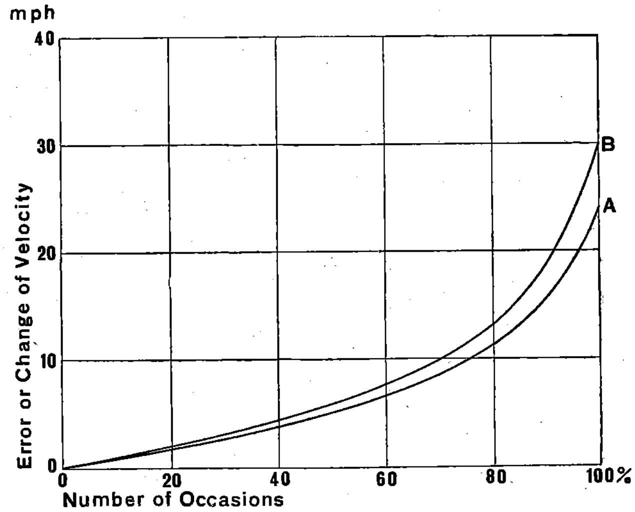

Figure 2.

The quality of the prediction is now gauged by the position which curve *A* occupies between the abscissa-axis and curve *B*. If *A* lies near the abscissa-axis the predictions have been good; if *A* lies near *B* the predictions were bad.

D o b s o n apparently made no effort to express the results of the verification in one figure.

c. Heidke's method (II)

H e i d k e (56—60) expresses the value V of the predictions by the formula:

$$V = \frac{\Sigma|\delta'| - \Sigma|\delta|}{\Sigma|\delta'|},$$

where δ and δ' are the deviations which possess the actually occurring phenomena from the weather service predictions and from predictions of the form "to-morrow the weather will be the same as to-day". H e i d k e mentions that he uses this verification method also for vectors (wind).

For complete success $V = 1$; V is small as the predictions are worse, for very bad predictions V may even be negative.

Like another checking system by H e i d k e, to be dealt with later, the disadvantage of this system is that it is so sensitive to the variability of the weather, which governs the factor $\Sigma|\delta'|$. This weakness has been pointed out by W a g e m a n n (33).

d. Methods of correlation

Wallén's idea of using the correlation coefficient as coefficient of verification has been mentioned by Clayton (40) and, after Chromow (2), it seems that Russian meteorologists have worked out the idea for predictions in which the meteorological elements are expressed in numerical values.

The correlation method, that could be used for wind-velocity, cloudiness, temperature and precipitation seems very useful, although the question would have to be further considered whether the normal correlation formulas should be used considering that the frequency-curve of some elements (e.g. cloudiness) is not a Gausz-curve.

IV. Half-objective methods

Under half-objective methods of verification we shall include those methods in which the checking of the predictions takes place entirely objectively, but in which the checking system itself contains valuations that are to a certain extent a question of taste. The first half-objective method to be found in literature seems to be owed to Köppen (62). It was applied to temperature predictions. To predictions colder, little change and warmer 100 was assigned when they were actually followed by the change predicted. To predictions of colder and warmer to which little change followed, and to predictions little change which were followed by colder or warmer, 50 was assigned. Predictions where colder was followed by warmer and warmer by colder 0 was given. The mean results yielded a useful valuation figure, which was however erroneously interpreted by many meteorologists as has been pointed out by Chromow (2). The percentage sign (%) was often added to this valuation figure, and so gave the impression that actually so much percent of the predictions had been correct. This is not right; in Köppen's system, for instance, a valuation figure of 50 may actually mean that 50 of 100 predictions were correct, but it may just as well mean that all the 100 predictions were incorrect, but by the charitable system of verification were appreciated to some extent.

Extensive half-objective checking systems are given by van Everdingen (47) and Dinies (42), less complete by Thomas (68). The remarkable thing in Dinies' system is, that the limits of the various terms in the rainfall predictions (dry, insignificant, light, moderate and heavy) are chosen in dependance upon the time of year.

A few verification methods, which we consider not to be entirely objective, will be discussed below.

a. Heidke's method (III).

Heidke has described the following system in a number of publications (52—60). First a series of official and a series of blindfold predictions is criticised with some half-objective method. For the official predictions a result E is obtained, for the blindfold predictions B. Heidke defines as quality Q:

$$Q = \frac{E - B}{1 - B}.$$

For the determination of E and B Heidke considers it very important to assign weights to different combinations of prediction and occurrence. He remarks, for instance, that it is often more difficult to forecast a change of weather than constancy. The successful predictions in which a change of weather is announced, therefore, should have more weight than the successful predictions in which no change is announced. The assignation of these weights, however, is in some respect a question of feeling.

As blindfold forecasts Heidke always choose "weather to-morrow will be the weather of to-day". The calculation of E and B always delivers values smaller than 1. Q may then vary between $-\infty$ and 1; for $E = B = 1$, Q is indefinite.

The disadvantage of this system is the sensitivity to the variability of the weather, which governs the factor B.

b. Verification of rainfall predictions which are given for a certain area.

Much has been written about the terminology that may be used for rainfall predictions given for particular districts. The term "local rain", for instance, has been much discussed. According to Klein (13) there is no sense in using this term, according to others (von Bebber, 1) it is certainly permissible, as investigations has shown that in a forecasting district it is not always dry everywhere nor wet everywhere.

Van Everdingen (47) has not included the term "local rain" in his half-objective system. But he notes two values for each forecasting period, namely the usual valuation-figure and the highest figure which could have been reached for the whole district with one single forecasting term. The quotient of the sums of the two series gives the final result for a certain number of forecasts.

In latter years in Holland the term "local rain" has been used, beside the terms "dry", "no rain of importance" and "rain". Then it was fixed how many percent of the verifying stations must or might have rain according to various predictions, and how much this rain must or might amount to.

The fixing of these limits is, however, purely subjective, and an further subjective element was introduced by the fact that to unsuccessful forecasts not always 0 was awarded but according to the degree of unsuccesfulness a value between 0 and 1.

Addendum

This paper has been written in winter 1943, during war-time, when Dutch scientists had no relations with the allied world. After the liberation it turned out that somewhat similar work as published here had been done by R. H. Muller (Verification of short-range weather forecasts; a survey of the literature, Bull. Am. Met. Soc. **25**; 18, 47 and 88; 1944). Our study however definitely has another character; it is not descriptive but more critical. Our survey contains some publications not mentioned by Muller; on the other hand some of the latest publications which appeared during the war are even now not known to us. Since it may be expected that these publications will not reach Holland in near future, it was decided to publish our work in the original form.

De Bilt, Spring 1946.

List of Literature

Text-Books

1. B e b b e r, W. J. v a n; Handbuch der ausübende Wetterungskunde II. Chapters on „Prüfung der Wetterprognosen und Sturmwarnungen und Resultate" and "Räumliche Verteilung gleichzeitiger Niederschläge, Prognosen-Bezirke". Stuttgart, 1886.
2. C h r o m o w, S. P.; Einführung in die synoptische Wetteranalyse. Vienna, 1940.
3. D e f a n t, A.; Wetter und Wettervorhersage. Chapter "Erfolge und Kritik der Wettervorhersagen". Leipzig, Wien, 1918.
4. S c h m a u s z, A.; Das Problem der Wettervorhersage. Leipzig, 1942.

Terminology

5. A n o n y m u s; Erklärung der in den Wetterungsberichten und Wetteraussichten der Seewarte angewandte Ausdrücke. Ann. Hydr. **6,** 220, 1878 and Das Wetter **20,** 49, 1903.
6. — Verification at the Signal Office, Am. Met. J. **6,** 19, 1889—'90.
7. — Aanwijzingen tot het uitwerken en publiceeren van het volledig en het verkort telegrafisch Weerbericht, K.N.M.I., nr 100.
8. — The signal corps meteorological Service A. E. F., M. Weather Review **47,** 870, 1919.
9. B u y n i t k i; The meaning of the word fair in meteorology. M. Weather Review **43,** 613, 1915.
10. E m m o n s, G.; Suggestions for improved presentation of weather information. Bull. Am. Met. Soc. **21,** 311, 1940.
11. F r e y b e; Ein Landwirtschaftlicher Wetterdienst; Das Wetter **20,** 2 etc., 1903.
12. H a l l e n b e c k, C.; Forecasting precipitation in percentages of probability; M. Weather Review **48,** 645, 1920.
13. K l e i n, H. J.; Ueber die lokale und allgemeine Wettervoraussage unter besonderer Berücksichtigung der Prognose "Stellenweise Niederschlag". Wochenschrift für Astronomie, etc. **29,** 28, 1886.
14. L a m m e r t, W.; Der Volkstümliche Wetterbericht. Das Wetter **53,** 235, 1936.
15. L a n d s b e r g, H.; Weather forecasting terms. Bull. Am. Met. Soc. **21,** 317, 1940.
16. M a h r t, W.; Die Aufgaben der synoptische Meteorologie im Reichswetterdienst; Met. Zs. **52,** 399, 1935.
17. M e h r i n g; Ausdrücke und Verständnis; Das Wetter, **28,** 71, 1911.
18. M y r b a c h, O. v o n; Verläszlichkeit der Wetterprognosen (ein Vorschlag); Met. Zs. **30,** 496, 1913.
19. S c h m a u s z, A.; Die Sprache der Meteorologen; Das Wetter **52,** 309, 1935.
20. S c h r e i b e r, K.; Bericht über die Erfahrungen beim Flugberatungsdienst 1927—1928. Erf. Ber. deutschen Flugwetterdienstes, 3. Folge, nr. 13.

General remarks on usefulness of checking.

21. B e b b e r, W. J. v o n; Entgegnung auf die Auslassungen des Herrn dr. H e r m a n n J. K l e i n über Wetterprognosen und Sturmwarnungen; Das Wetter **8,** 193, 1891.
22. B e b b e r, W. J. v a n; Die Prüfung der Wetterprognose und die Erhaltungstendenz der Witterung; Das Wetter, **8,** 268, 1891.
23. B o r n s t e i n; Der Prophetentag, Das Wetter **28,** 23, 1911.
24. C l a y t o n, H. H.; Verification of Weather Forecasts; Am. Met. Journal **6,** 211, 1889—'90.
25. — Verification of Weather Forecasts; Am. Met. Journal **8,** 369, 1891—'92.
26. F r e y b e; Die Treffsicherheit von Prognosen; Das Wetter **28,** 117, 1911.
27. H a z e n, H. A.; The verification of weather forecasts; Am. Met. Journal **8,** 392, 1891—'92.
28. H e i d k e, P.; Ueber die Zweckmässigkeit einer objektiven Bestimmung von Erfolg und Güte der Wettervorhersagen. Met. Zs. **44,** 219, 1927.
29. K l e i n, H. J.; Ergebnisse rationeller Prüfungen von Wetterprognosen und deren Bedeutung für die Praxis; Wochenschrift für Astronomie, etc. **28,** 9, 1885.
30. S c h m a u s z, A.; Die Treffsicherheit der Prognosen; Das Wetter **28,** 68, 1911 and **28,** 167, 1911.

31. Schmausz, A.; Defant "Wetter und Wettervorhersage"; Met. Zs. **43**, 434, 1926.

32. Treibich; Welche Mittel sind geeignet die Treffsicherheit der Wetterprognosen und die Wertschätzung des Wetterdienstes in der Öffentlichkeit zu heben; Das Wetter **44**, 17, 1927.

33. Wagemann, H.; Prognosenprüfung; Ann. Hydr. **63**, 270, 1935.

34. Mars, S.; De weersverwachtingen van het K.N.M.I., de Bilt en van de Filiaalinrichting te Amsterdam en de weersvoorspellingen van Jules Capré. Hemel en Dampkring **1**, 73, 1903.

35. Meiszner, O.; Ueber die Zuverlässigkeit der "Wettervoraussichte für mehrere Tage"; Das Wetter **31**, 250, 1914.

36. — Ueber die Zuverlässigkeit langfristiger Wettervoraussage; Das Wetter **35**, 48, 1918.

37. — Weitere Bemerkungen über die Zuverlässigkeit langfristiger Wettervoraussagen; Das Wetter **42**, 272,1925.

38. Schmidt, K.; Die Treffsicherheit der Karlsruher Wettervorhersage. Deutsches Met. Jahrbuch 1928—1931 (Badische Landenwetterwarte) p. 75.

Objective methods

39. Anonymus.; Eine rationelle Methode zur Prüfung der Wetterprognosen. Publication of Department III, Deutsche Seewarte, 1884.

40. Clayton, H. H.; Rating Waether forecasts, Bull. Am. Met. Soc. **15**, 279, 1934.

41. — On verifying Weather Forecasts, Bull. Am. Met. Soc. **19**, 341, 1938.

42. Dinies; Ueber Prüfungsmethoden von Wettervorhersagen; Erf. Ber. deutschen Flugwetterdienstes 5. Sonderband.

43. Dobson, G. M. B.; Causes of errors in forecasting pressure gradients and winds. Q. J. **47**, 261, 1921.

44. Doolittle, M. H.; The verification of predictions. Bulletin of the Philosophical Society of Washington **7**, 122, 1884.

45. Ekholm, N.; Ueber die unperiodischen Luftdruckschwankungen und einige damit zusammenhängende Erscheinungen. Met. Zs. **24**, 145, 1907.

46. Everdingen, E. van; De uitkomsten der weersverwachtingen van het Kon. Ned. Met. Inst. in 1904 en 1905 en in den zomer van 1907. Mededeelingen en Verhandelingen, nr. 6 K.N.M.I. nr. 102.

47. — De methode van beoordeeling der weersverwachtingen van het Kon. Ned. Met. Inst. en de uitkomsten in den zomer van 1911. Mededeelingen en Verhandelingen, nr. 14 K.N.M.I. nr. 102.

48. Ferraz, S.; Verification of daily weather forecasts. Circular 19, Meeting Directors etc., Utrecht 1923.

49. Finley, J. P.; Tornado-predictions. Am. Met. Journal 1, 85, 1884—'85.

50. Gilbert, G.K.; Finley's tornado-predictions; Am. Meteor. Journal **1**, 166.

51. Grossmann, L.; Die Stürme und die Sturmwarnungen an die deutsche Küste; Aus dem Archiv der deutschen Seewarte **21;** nr. IV, 1898.

52. Heidke, P.; Berechnung des Erfolges und der Güte der Windstärkevorhersagen im Sturmwarnungsdienst. Geogr. Ann. **8**, 301, 1926.

53. — Zur Berechnung des Erfolges und der Güte der Wetterdienstvorhersagen; Met. Zs. **43**, 489, 1926.

54. — Die Berechnung des Erfolges und der Güte der Wetterdienstvorhersagen; Das Wetter **44**, 169, 1927.

55. — Erfolg und Güte örtlicher Vorhersagen im täglichen Wetterdienst.Aus dem Archiv der deutschen Seewarte **45**, nr. 1, 1928.

56. — Zur objektiven Prüfung örtlicher Vorhersagen im täglichen Wetterdienst; Btr. Ph. fr. Atm. **15**, 202, 1929.

57. — Zur objektiven Prüfung von Wettervorhersagen; Ann. Hydr. **58**, 134, 1930.

58. — Erfolg und Güte von Hamburger und Köningsberger Frostvorhersagen für die nächste Nacht. Ann. Hydr. **63**, 40, 1935.

59. — Ergebnisse einer objektiven Prüfung von Wettervorhersagen; Met. Zs. **52**, 487, 1935.

60. — Vorschläge zur objektiven Prüfung von Wetterdienstvorhersagen der Temperatur in Graden und der Bewölkung in Zehnteln für den nächsten Tag. Ann. Hydr. **69**, 222, 1941.

61. Klein, H. J.;Ergebnisse rationeller Prüfungen von Wetterprognosen und deren Bedeutung für die Praxis, Wochenschrift für Astronomie etc. **39**, 9, 1885..

62. Köppen, W.; Met. Zs. **1**, 39. 1884 and Met. Zs. **1**, 397, 1884.

63. — Wie erkennt man Blindlingsprognosen, Met. Zs. (Hannband), 347, 1906.

64. Lacour, D.; Archiv for Sövaesen, 1906.

65. Loris-Melikoff, M. A.; Sur la question de l'appréciation des prévisions du temps. Leningrad, Bull. Obs. Geoph. Centr., No. 1, 1927.

66. Nichols, W. S.; The mathematical elements in the estimation of signal service reports; Am. meteor. Jrnal 6, 386, 1889—'90.

67. Pierce, C. S.; The numerical measure of the success of predictions; Science **4**, 453, 1884.

68. Thomas, H.; Zur Prüfung der synoptischen Vorhersagen. Met. Zs. **50.** 29, 1933.

69. Toth, G.; Ein Versuch zur Vorhersage von Temperaturwerten; Az Idöjárás **37**, 29, 1933.

70. Walle, A.; Sur le contrôle des annonces de tempêtes, Geograf. Ann. **3**, 267, 1921.

Schedule of some objective systems of verification, constructed for meteorological elements for which the prediction terms provide only two alternatives

Method	Formula	Results on complete success	Results on complete failure	Results for Finley's tornado-predictions: direct	Results for Finley's tornado-predictions: invers	Results for random issues of p' positive and $(o-p')$ non-predictions: General	Results for random issues: For tornado predictions	Capable of inversion	Subject to abuse
FINLEY	$v'_F = \frac{c' + s - (o' + p' - c')}{s}$	1	0	0,966	0,966	$2\frac{o'}{s}\frac{p'}{s} - \frac{o'}{s} - \frac{p'}{s} + 1$	0,947	yes	yes
FINLEY (im proved).	$v'_{FW} = \frac{c'\frac{s-o'}{s} + \frac{o'}{s}s - (o' + p' - c')}{c'\frac{s-o'}{s} + (p-c') + (o-c') + \frac{o'}{s}s - (o' + p' - c')}$	1	0	0,446	0,466	$\frac{o'(s-o')}{s(p+2o') - o'(2p+o')}$	0,253	yes	no
HEIDKE (I). . .	$i'_H = \frac{2(c's - o'p')}{o'(s-p') + p'(s-o')}$	1	$-\frac{2o'(s-p')}{o'^2 + (s-o')^2}$	0,355	0,355	0	0	yes	no
GILBERT (I) . .	$v'_G = \frac{c'}{p' + o' - c'}$	1	0	0,230	0,968	$\frac{o'p'}{s(p'+o') - o'p'}$	0,012	no	yes
GILBERT (II) . .	$i'_G = \frac{c's - o'p'}{(o' + p' - c')s - o'p'}$	1	$-\frac{o'(s-o')}{s^2 - o'p' + o'^2}$	0,216	0,216	0	0	yes	no
DOOLITTLE (I) .	$v'_D = \frac{c'^2}{o'p'}$	1	0	0,154	0,965	$\frac{o'p'}{s^2}$	0,005	no	yes
DOOLITTLE (II)	$i'_D = \frac{(c's - o'p')^2}{o'p'(s-o')(s-p')}$	1	1	0,142	0,142	0	0	yes	no
LACOUR	$v'_L = \frac{c'}{p'} \times \frac{s-o'}{o'-c'}$		0	32,9	1,38	$\frac{s-p'}{s-o'}$	0,987	no	yes
CLAYTON. . . .	$v'_C = \frac{c's - o'p'}{p'(s-p')}$	1	—1	0,272	0,272	0	0	yes	yes
WALLÉN	$v'_W = \frac{c's - o'p'}{\sqrt{p'(s-p') \times o'(s-o')}}$	1	—1	0,377	0,377	0	0	yes	no