KONINKLIJK NEDERLANDS METEOROLOGISCH INSTITUUT

MEDEDELINGEN EN VERHANDELINGEN

No. 81

A. W. HANSSEN and W. J. A. KUIPERS

ON THE RELATIONSHIP BETWEEN THE FREQUENCY OF RAIN AND VARIOUS METEOROLOGICAL PARAMETERS

(With reference to the problem of objective forecasting)

1965

F. 10,—

STAATSDRUKKERIJ- EN UITGEVERIJBEDRIJF / 'S-GRAVENHAGE

ON THE RELATIONSHIP BETWEEN THE FREQUENCY OF RAIN AND VARIOUS METEOROLOGICAL PARAMETERS

(With reference to the problem of objective forecasting)

PUBLIKATIE K.N.M.I. 102-81

U.D.C.: 551.509.324.2
551.509.314
551.577.36

KONINKLIJK NEDERLANDS METEOROLOGISCH INSTITUUT
MEDEDELINGEN EN VERHANDELINGEN
No. 81

ON THE RELATIONSHIP BETWEEN THE FREQUENCY OF RAIN AND VARIOUS METEOROLOGICAL PARAMETERS

(With reference to the problem of objective forecasting)

A. W. HANSSEN and W. J. A. KUIPERS

1965

STAATSDRUKKERIJ- EN UITGEVERIJBEDRIJF 'S-GRAVENHAGE
1965

PREFACE

In the present Mededeling en Verhandeling of the Royal Netherlands Meteorological Institute a statistical method for the prediction of precipitation, developed by the authors, is described.

The method has been applied at the Royal Netherlands Meteorological Institute during a number of years now and the results are satisfactory. In spite of the fact that numerical values of the various parameters that are used relate to Netherlands climatological circumstances, the method as such seems to be of sufficient general interest to justify publication.

The Director in Chief
Royal Netherlands Meteorological Institute
C. J. WARNERS

VOORWOORD

In de onderhavige Mededeling en Verhandeling van het K.N.M.I. wordt een overzicht gegeven van een statistische methode voor het voorspellen van neerslag, welke door de schrijvers is ontwikkeld.

De methode wordt sedert een aantal jaren met bevredigende resultaten in de Weerdienst in Nederland toegepast. Hoewel de waarden van de verschillende parameters, die daarbij worden gebruikt, uiteraard door het Nederlandse klimaat worden bepaald, lijkt de ontwikkelde gedachtengang van voldoend algemeen belang om publikatie te rechtvaardigen.

De Hoofddirecteur
Koninklijk Nederlands Meteorologisch Instituut
C. J. WARNERS

CONTENTS

0. INTRODUCTION

Before attempting to develop a method for rain-forecasting, the obvious thing to do is to carry out an investigation into the physical processes which produce rain. Without going into details as regards rain theories put forward up to now, it can be stated that the process of the forming of precipitation in general is determined mainly by the humidity, the vertical temperature distribution and the vertical motion in the atmosphere. It should in theory be possible to make a forecast of these elements, but, in practice, difficulties are encountered, owing to the complexity of the equations of motion to be solved and to lack of knowledge of the initial state. Consequently the conventional routine forecast is based not so much on theoretical considerations as on an estimate of observed rainbelts or weather systems associated with rain, such as fronts, depressions or troughs. The displacement is usually estimated by a freehand extrapolation, whereas modifications, if any, concerning variation of intensity or velocity of the displacement are based on qualitative considerations. The personal factor needs must play an important role in this procedure, and consequently a given initial state of the atmosphere does not give just one forecast. For this reason the procedure is qualified as subjective. This fact is merely stated here; it is beyond the scope of the present paper to enter into a discussion of the pros and cons of subjective forecasting. Comment will be restricted to the remark that it would seem preferable to eliminate the personal factor as much as possible. The first step in pursuing this object is to attempt to record the experience of the forecaster. Broadly speaking, this experience concerns a relationship between precipitation and weather systems. The weather system can, in turn, be characterized by the value of some weather elements such as pressure and wind direction at a number of places. For the sake of simplicity the number will be restricted to one only.

The first problem then, is to discover a relationship between precipitation on the one hand and some meteorological variable on the other. If such a relationship proves to exist it will be true to say that, in principle, the variable concerned has predictive value and it will be called a predictor of the precipitation phenomenon (predictand). A number of variables will be tested for their predictive value and, so that it will be possible to select the best ones, a means of measuring this property must be developed.

Next a procedure to evaluate the combined effect of more than one predictor will be described. It must be stressed that so far use has been made of simultaneous relationships between predictand and predictors. Consequently, application of the results to real "forecasts" necessitates the making of an estimate of the future value of a predictor. It is assumed, however, that the variability of the predictor is relatively small, so that a linear extrapolation in time is justified.

The principle underlying this investigation is the dependence of one meteorological variable on another. So it is sensible to start with a treatise on dependence in general. A definition of this concept and the development of a standard to measure it quantitatively will be the subject of the following chapter.

1. DEPENDENCE OF TWO METEOROLOGICAL VARIABLES

A typical feature of meteorological variables is the fact that they obey a probability distribution. This will be expressed by calling them variates. Considering two variates x and y, a correspondence may be fixed, namely by subjoining the value of x at a certain time to the simultaneous value of y. In view of the subsequent application, the variate y will be denoted as predictand and x as predictor. Starting with the simplest case of variates which can take only two values denoted as 0 and 1, four possibilities can be distinguished with regard to the occurrence of the values x and y. Let E_{ij} denote the event of the joint occurrence specified as $y=y_i$ and $x=x_j$, and let p_{ij} be the probability of this event, the following table can then be composed:

y_i \ x_j	0	1
0	p_{00}	p_{01}
1	p_{10}	p_{11}

table 0.1

Next, the concept of conditional probability is introduced, defined as the probability of the event $y=y_i$ when $x=x_j$. This probability is denoted by $p(y_i|x_j)$. From table 0.1 it follows:

$$p(0|0) = \frac{p_{00}}{p_{00} + p_{10}}. \quad (1) \qquad p(0|1) = \frac{p_{01}}{p_{01} + p_{11}}. \quad (2)$$

Now y is said to depend on x if, and only if, those probabilities are unequal. In this case the conditional probability of y apparently depends on the value of x. Taking this fact as our starting-point, a generalization can be made for the case where the variates can take more than two values.

Definition: A predictand is dependent on a predictor, if, and only if, the probability distribution of the variate describing the predictand depends on the value describing the predictor.

In practice, the population to which the double dichotomy represented by the table 0.1 above refers, is not completely known. In general only estimates of the probabilities obtained as relative frequencies from a sample are available. In order to test a hypothesis concerning a dependence on the basis of such a sample, a test quantity is needed, which must be a function I of the relative frequencies mentioned above.

So far, dependence has been defined only qualitatively and, consequently, it is not possible to decide whether a predictand is more dependent on one predictor than on another. For this purpose, a standard against which to measure dependence is required. An indication on how this problem may be solved is provided by the function I mentioned above, as will be demonstrated.

Let $\mathscr{I}=I(p)$ be the value of that function when substituting population values; then, by a proper choice of the function, we may obtain $\mathscr{I}=0$ in the case of independence. Reversely, dependence does exist if $\mathscr{I}\neq 0$. It is therefore wise to try to identify $\mathscr{I}$ with the standard against which dependence is measured.

1.0 DEPENDENCY-INDEX

At first sight there seems to be an infinite number of possibilities as regards the form of the function I. It will be shown, however, that it is possible to narrow down this range considerably by imposing certain conditions on I.

Dependence has been defined as $I(p)$, the value of I when substituting population values. When only sample values of p are available, an estimation of the dependence is made by substituting those sample values f in the function I. Now it is logical to require that the expectation of the estimate $I(f)$, thus defined, should equal the estimated quantity, based on the population or, expressed as a formula:

$$E\{I(f)\} = \mathscr{I}. \tag{3}$$

Owing to the relation:

$$E(f) = p \tag{4}$$

this condition is equivalent to:

$$E\{I(f)\} = I\{E(f)\}.$$

In general, the relation is valid if, and only if, I is a linear function. Introducing constants w_{ij}, the general form of $I(p)$ can be written as:

$$I(p) = \sum_i \sum_j w_{ij} p_{ij} . \tag{5}$$

In addition:

$$I(f) = \sum_i \sum_j w_{ij} f_{ij} \tag{6}$$

where f_{ij} is the sample value of p_{ij}.
According to the convention made above, $\mathscr{I}$ equals zero in the case of independence, which may be expressed by the relation:

$$\frac{p_{00}}{p_{00} + p_{10}} = \frac{p_{01}}{p_{01} + p_{11}} . \tag{7}$$

$$\text{Let: } \left.\begin{array}{l} p_0 = p_{00} + p_{01} \\ p_1 = p_{10} + p_{11} \end{array}\right\} \tag{8}$$

where p_0 is the apriori probability of the event $y=0$, i.e. the probability irrespective of the value of x, and p_1 is the a priori probability of $y=1$. From the definitions it follows that

$$p_{00} + p_{01} + p_{10} + p_{11} = p_0 + p_1 = 1 . \tag{9}$$

Applying rules of proportion, the relation (7) may be converted into:

$$\frac{p_{00}}{p_{10}} = \frac{p_{01}}{p_{11}} = \frac{p_0}{p_1} . \tag{10}$$

Substitution in (5) yields:

$$I(p) = \left(w_{00} \frac{p_0}{p_1} + w_{10} \right) p_{10} + \left(w_{01} \frac{p_0}{p_1} + w_{11} \right) p_{11} .$$

So that the right-hand member may identically be zero, a necessary and sufficient condition is that the forms between brackets vanish, which condition is equivalent to:

$$\frac{w_{00}}{w_{10}} = \frac{w_{01}}{w_{11}} = - \frac{p_1}{p_0} . \tag{11}$$

In view of the fact that I will be used as a standard it is desirable, in accordance with common use, that I equals unity in the case of extreme dependence. The latter is said to occur if the value of y is determined completely by the value of x, i.e.

$$\begin{aligned} &\text{if } x = 0, \quad \text{then} \quad y = 0 \\ &\text{if } x = 1, \quad \text{then} \quad y = 1 . \end{aligned}$$

From table 0.1 it follows that in this case the following relations hold:

$$\left.\begin{aligned} p_{01} &= p_{10} = 0 \\ p_{00} &= p_0 \\ p_{11} &= p_1 \end{aligned}\right\} . \tag{12}$$

Substitution of (12) in (5) yields the following additional condition:

$$I(p) = w_{00}p_0 + w_{11}p_1 = 1 . \tag{13}$$

Summarizing, it may be observed that 3 equations have been derived, namely (11) and (13), involving 4 unknowns, so that one degree of freedom still remains. In introducing an additional parameter t, equation (13) may be written as:

$$w_{00} = \frac{t}{p_0}; \quad w_{11} = \frac{1-t}{p_1} .$$

Substitution in (11) yields:

$$w_{10} = -\frac{t}{p_1}; \quad w_{01} = \frac{t-1}{p_0} .$$

The result is summarized in table 0.2.

$w_{ij} =$ (14)

$i \backslash j$	0	1
0	$\frac{t}{p_0}$	$\frac{t-1}{p_0}$
1	$\frac{-t}{p_1}$	$\frac{1-t}{p_1}$

table 0.2

It is note worthy that $I(p)$ proves to be independent of the parameter t, as will be demonstrated.
Substitution of (14) in (5) yields, in virtue of (8):

$$I(p) = \frac{t}{p_0} p_{00} + \frac{t-1}{p_0}(p_0 - p_{00}) + \frac{-t}{p_1}(p_1 - p_{11}) + \frac{1-t}{p_1} p_{11}$$

$$I(p) = \frac{p_{00}}{p_0} + \frac{p_{11}}{p_1} - 1. \tag{15}$$

Apparently the value of $I(p)$ is determined uniquely by imposing three conditions on it, viz.

1. I is linear
2. $I=0$ in case of independence (16)
3. $I=1$ in case of extreme dependence.

The conditions (16) being satisfied, the index $I(p)$ depends exclusively on the probabilities p_{ij}. As has been noted already, only estimates of those probabilities are actually known for which reason formula (6) is used as an estimator of $I(p)$. In this formula the values of table 0.2 are substituted as weighting-functions w_{ij}. It is true that these values are functions of probabilities which, strictly speaking, are not known. However, the estimates which are available originate in most cases from a sample of such a size that it is permissible to regard them as identical with the population values. The following example is given as an illustration:

y \ x	0	1	
0	0.20	0.10	0.30
1	0.10	0.60	0.70

$I=0.20/0.30+0.60/0.70-1=0.53.$

Obviously the conditional probability of $y=0$ equals $p_c=0.20/0.30=0.67$ if $x=0$; and $p_c=0.1/0.70=0.14$ if $x=1$. Consequently, for $x=0$, the conditional probability is much greater than the a priori probability (being 0.3). The dependence may be used to "predict" the value of y, when x is given. The predicted value will be denoted as y_p. In the above example the following course should be adopted:

$$y_p = 0 \quad \text{if} \quad x = 0 \qquad \text{and} \qquad y_p = 1 \quad \text{if} \quad x = 1.$$

The result of these predictions are summarized in a contingency-table which is identical with the table used to determine the dependence of y and x, as far as the cell data are concerned. Consequently the latter table may be interpreted as representing the results of the subsequent prediction. When used in this way the headings of the entries are modified as follows:

forecast

	$y \backslash y_p$	0	1
observed	0		
	1		

1.1 VARIANCE OF $I(f)$

In 1.0 it was shown that the parameter t does not interfere with the expectation of $I(f)$, but that the value of $I(f)$ itself does depend on t. This dependence is revealed when investigating the variance of $I(f)$. In order to deduce an expression for the variance a statistical model will be introduced by means of which an imaginary experiment may be performed simulating the process leading to the double dichotomy described in the foregoing section.

Suppose there is a box containing four kinds of tickets marked with the numbers w_{ij}, and the ratio of occurrence of the category involved equals p_{ij}. Now N times in succession a ticket is drawn and replaced. The numbers w_{ij} are added together giving a total amount S. So, if n_{ij} tickets marked w_{ij} are drawn:

$$S = \sum_{i,j=0}^{1} w_{ij} n_{ij} \qquad \text{or, defining:} \qquad f_{ij} = \frac{n_{ij}}{N}, \quad S = N \sum w_{ij} f_{ij}$$

or

$$\frac{S}{N} = \sum w_{ij} f_{ij} \qquad \text{(cf. (6))}.$$

The expectation of S/N equals:

$$E(S/N) = \sum w_{ij} p_{ij}.$$

So there seems to be accordance between the experiment producing the ratio S/N and the process giving rise to the index $I(f)$. However, the analogy is not complete, since the events E_{ij} are not mutually independent if the predictand shows *persistence*. Nevertheless, it is useful to formulate an expression for the variance in spite of its restricted validity due to persistence, since it will show the dependence of the variance on the parameter t. The expectation of S equals $E(S)=N\sum w_{ij}p_{ij}$; the generating function of S will be defined by $G(t)=\{\sum p_{ij}t^{w_{ij}}\}^N$. If we put $t=e^\alpha$ in $G(t)$ we obtain the moment generating function of S:

$$M(\alpha)=\left\{\sum p_{ij}e^{\alpha w_{ij}}\right\}^N \tag{17}$$

according to the theory on generating functions [8].
As will be known, the coefficient of $\alpha^r/r!$ in the expansion of M as a series of α, equals μ_r', the r^{th} moment of S. Expanding the exponential function, (17) may be reduced to:

$$M(\alpha)=\left\{\sum_{i,j=0}^{1} p_{ij}\sum_{k=0}^{\infty}\frac{(\alpha w_{ij})^k}{k!}\right\}^N=\left\{\sum_{k=0}^{\infty}\frac{\alpha^k}{k!}\sum_{i,j=0}^{1} p_{ij}w_{ij}^k\right\}^N.$$

Let: $$A_k=\sum_{i,j=0}^{1} p_{ij}w_{ij}^k \quad \text{then:} \quad M(\alpha)=\left\{\sum_{k=0}^{\infty}A_k\frac{\alpha^k}{k!}\right\}^N \tag{18}$$

Since in particular: $A_0=\sum p_{ij}=1$, equation (18) may be written as:

$$M(\alpha)=\left\{1+A_1\alpha+A_2\frac{\alpha^2}{2}+A_3\frac{\alpha^3}{3}+\dots\right\}^N$$

or, applying polynomial expansion:

$$M(\alpha)=1+N\left(A_1\alpha+A_2\frac{\alpha^2}{2}+\dots\right)$$
$$+\frac{N(N-1)}{2}(A_1\alpha+\dots)^2+\dots$$

From this it follows:

$$\mu_1'=NA_1=N\sum w_{ij}p_{ij}=NI(p)$$
$$\mu_2'=NA_2+N(N-1)A_1^2.$$

The second moment with respect to the mean value of S equals:

$$\mu_2 = \mu_2' - (\mu_1')^2 =$$
$$= N A_2 + N(N-1) A_1^2 - N^2 A_1^2 =$$
$$= N(A_2 - A_1^2).$$

Now the variance of $I(f)$ equals the variance of S/N, so

$$\operatorname{var} I(f) = \operatorname{var} S/N = \frac{1}{N^2}\operatorname{var} S = \frac{\mu_2}{N^2} = \frac{A_2 - A_1^2}{N}$$
$$= \frac{\sum p_{ij} w_{ij}^2 - (\sum p_{ij} w_{ij})^2}{N} \quad \text{or}$$

$$\operatorname{var} I(f) = \frac{p_{00}\dfrac{t^2}{p_0^2} + p_{01}\dfrac{(1-t)^2}{p_0^2} + p_{10}\dfrac{t^2}{p_1^2} + p_{11}\dfrac{(1-t)^2}{p_1^2} - I(p)^2}{N} \tag{19}$$

Let:

$$B = \frac{p_{00}}{p_0^2} + \frac{p_{10}}{p_1^2}$$
$$C = \frac{p_{01}}{p_0^2} + \frac{p_{11}}{p_1^2} \tag{20}$$

so that, by reason of (8):

$$B + C = \frac{1}{p_0} + \frac{1}{p_1} = \frac{p_0 + p_1}{p_0 p_1} = \frac{1}{p_0 p_1} \tag{21}$$

then:

$$\operatorname{var} I(f) = \frac{Bt^2 + C(1-t)^2 - \{I(p)\}^2}{N}. \tag{22}$$

Additionally, let:

$$D = Bt^2 + C(1-t)^2$$

then:

$$\operatorname{var} I(f) = \frac{D - \{I(p)\}^2}{N}. \tag{23}$$

The expression D can be written as follows:

$$D = C - 2Ct + t^2(B + C)$$

or, in view of (21):

$$D = C - 2Ct + \frac{t^2}{p_0 p_1}$$

$$D = \frac{1}{p_0 p_1}(t - Cp_0 p_1)^2 + C - C^2 p_0 p_1 . \tag{24}$$

Now, as $I(p)$ is independent of t, the variance has a minimum if D has a minimum. From (24) it follows that this is the case if the quadratic term vanishes, i.e. if:

$$t = Cp_0 p_1 . \tag{25}$$

The minimum of D equals:

$$D_{\min} = C - C^2 p_0 p_1 = Cp_0 p_1 \left(\frac{1}{p_0 p_1} - C \right) =$$
$$= Cp_0 p_1 (B + C - C) = BCp_0 p_1 .$$

Finally, the minimum of the variance equals:

$$\min \operatorname{var} I(f) = \frac{BCp_0 p_1 - \{I(p)\}^2}{N}. \tag{26}$$

So far, there has been no preference for any value of the parameter t, but at this point it seems reasonable to let it satisfy equation (25) so that the variance may be a minimum. On the other hand, it is inconvenient that in this case both t and the variance should depend on the probabilities p_{ij}, which, strictly speaking, are not known. From (22) it follows, in connection with (21), that the dependence mentioned vanishes if $t = \frac{1}{2}$. Substitution of this value in (22) yields:

$$\operatorname{var} I(f) = \frac{\dfrac{1}{4p_0 p_1} - \{I(p)\}^2}{N}. \tag{27}$$

In practice it appears that in many cases the difference between the value of $Cp_0 p_1$, as evaluated from estimated values of p_{ij}, and the constant value $t = \frac{1}{2}$, is rather small. This means that when using the value $t = \frac{1}{2}$, the variance is practically a minimum.

In the beginning it was noted that $I(f)$ may be used as a test statistic. A special case is present when the reality of a supposed dependence must be tested. In this case the null-hypothesis states: $I(p) = 0$.

The expression for the variance proves to be very simple now, viz.:

$$\text{var}\, I(f) = \frac{1}{4p_0p_1N}. \tag{28}$$

Up to now, the number of classes has been restricted to two. In order to make an extension of this number possible, an attempt will be made to derive a general expression for w_{ij} on the basis of the value $t=\frac{1}{2}$. Substitution in table 2 yields:

$$w_{00} = \frac{1}{2p_0} = \frac{1-p_0}{2p_0p_1}$$

$$w_{01} = \frac{-1}{2p_0} = \frac{-p_1}{2p_0p_1}$$

$$w_{10} = \frac{-1}{2p_1} = \frac{-p_0}{2p_0p_1}$$

$$w_{11} = \frac{1}{2p_1} = \frac{1-p_1}{2p_0p_1}.$$

Introducing Kronecker's delta-notation, these equations may be summarized as follows:

$$w_{ij} = \frac{\delta_{ij} - p_j}{2p_0p_1}.$$

From the identity $(p_0+p_1)^2=1$,

it follows: $p_0^2 + p_1^2 + 2p_0p_1 = 1$

or: $2p_0p_1 = 1 - \sum_r p_r^2$.

Substitution yields:

$$w_{ij} = \frac{\delta_{ij} - p_j}{1 - \sum_r p_r^2} \tag{29}$$

1.2 EXTENSION OF THE DEFINITION OF DEPENDENCE

The formulas (5) and (29) raise no difficulties when the summation over the indices i, j and r is extended to more than two terms. Therefore the combination of these formulas can be used as a definition of the dependency-index, provided the conditions 2 and 3 mentioned in (16) prove to be satisfied.

Let:
$$\mathscr{I} = \frac{\sum_j \sum_i (\delta_{ij} - p_j) p_{ij}}{1 - \sum_r p_r^2} = \frac{\sum_j p_{jj} - \sum_j p_j \sum_i p_{ij}}{1 - \sum_r p_r^2}. \tag{30}$$

In conformity with (10) independence is defined by:

$$\frac{p_{ij}}{p_{jj}} = \frac{p_i}{p_j}.$$

Application of a rule of proportion yields:

$$\frac{\sum_i p_{ij}}{p_{jj}} = \frac{\sum_i p_i}{p_j} = \frac{1}{p_j}.$$

Substitution in (30) yields:

$$\mathscr{I} = \frac{\sum_j p_{jj} - \sum_j p_{jj}}{1 - \sum_r p_r^2} = 0.$$

On the other hand extreme dependence is defined (cf. (12)) as:

$$p_{ij} = p_{jj} \delta_{ij}.$$

From this definition it follows:

$$p_{jj} = p_j$$

and:
$$\sum_i p_{ij} = p_j \sum_i \delta_{ij} = p_j$$

Substitution in (30) yields:

$$\mathscr{I} = \frac{\sum_j p_j - \sum_j p_j p_j}{1 - \sum_r p_r^2} = 1.$$

Although this system has been developed primarily to define the dependence between a predictand and a predictor, it may also serve as a verification system for forecasts. The way in which it works in that capacity can be illustrated by comparing it with a sweepstake. For this purpose, with each prediction a stake will be associated, the magnitude of which equals p_j, the a priori probability that the prediction will be true. If the forecaster makes a hit, then this "prize" will be unity. So his net gain equals $1 - p_j$. On the contrary, if the prediction is wrong he gets no prize and he loses his stake completely, so his net gain is negative, viz. $-p_j$. The total net gain of a set of forecasts is compared with the total which would have been obtained if all forecasts had been hits. The ratio of both net gains will be used as a standard for the quality of the forecasts.

1.3 DEPENDENCE IN THE CASE OF A CONTINUOUS PREDICTOR

In the preceding sections dependence has been defined for cases where both the predictand and the predictor could take equal numbers of values. Actually, it often happens that the predictor is described by a variate x which can take more values than the predictand or even a continuous scale of values. In order to be able to apply the foregoing theory, it will be necessary to fix a correspondence between the predicted value y_p and x.

The question arises as to what factors will have to be considered in formulating that correspondence. It is to be expected that the value of the resulting dependency-index will be dependent on the assumed correspondence, so that the obvious course is to have it determined by the condition that $\mathscr{I}$ becomes a maximum. Let the population with which we are concerned be divided up into a number of groups, according to the value of x. On account of its form the dependency-index may be regarded as the weighted mean of the contributions of each group. These contributions being mutually independent, the index becomes a maximum if each contribution, separately, is a maximum.

Let the value $y_p = s$ be assigned to the group with code number k, i.e. $j = s$ if $x = k$, then in the corresponding contingency-table all columns except the one for which $j = s$ are empty, hence:

$$p_{ij} = 0 \quad \text{if} \quad j \neq s$$

Let furthermore $p_i^{(k)}$ be the conditional probability of the event $y = y_i$ if $x = k$, then:

$$p_{is} = p_i^{(k)}$$

whereas:

$$\sum_i p_i^{(k)} = 1 .$$

Both relations may be summarized in:

$$p_{ij} = \delta_{js} p_i^{(k)} . \tag{31}$$

From (31) it follows:

$$\sum_j p_{jj} = \sum_j \delta_{js} p_j^{(k)} = p_s^{(k)} . \tag{32}$$

Furthermore:

$$\sum_j p_j \sum_i p_{ij} = \sum_j \sum_i p_j p_{ij} = \sum_i \sum_j p_j p_{ij}$$
$$= \sum_i \sum_j p_j \delta_{js} p_i^{(k)} = \sum_i p_s p_i^{(k)}$$
$$= p_s \sum_i p_i^{(k)} = p_s$$

Substitution of (31) and (32) in (30) yields:

$$\mathcal{I} = \frac{p_s^{(k)} - p_s}{1 - \sum_r p_r^2} \tag{33}$$

At this point it is possible to define the mentioned correspondence, applying the following directory:

Directory

Let k be an arbitrary value of a predictor x, then the conditional probability $p(y_i|k) = p_i^{(k)}$ of the event $y = y_i$ is evaluated as well as the difference $p_i^{(k)} - p_i$, which will be called "probability-excess". Let s be that value of i, for which the probability-excess is a maximum, then the value s is assigned to y_p.
In case of a predictand with two classes, this directory can be simplified. Let $p_0^{(x)}$ be the conditional probability of $y=0$, then the probability of $y=1$ equals $p_1^{(x)} = 1 - p_0^{(x)}$, whereas $p_1 = 1 - p_0$. A comparison must now be made between the probability-excess $(p_0^{(x)} - p_0)$ and $(1 - p_0^{(x)}) - (1 - p_0) = p_0 - p_0^{(x)}$. The first one is larger if $p_0^{(x)} > p_0$. This result can be summarized as follows:
Rule: Let a predictand which is dependent on a continuous variable x be described by a variable y, which can assume only two values viz. 0 and 1. Additionally, p_0 be the a priori probability of the event $y=0$ and $p_0^{(x)}$ the conditional probability of that event, then:

$$y_p = 0 \quad \text{if} \quad p_0^{(x)} > p_0$$
$$y_p = 1 \quad \text{if} \quad p_0^{(x)} < p_0 ,$$

2. AVAILABLE DATA AND PROCESS OF GROWTH OF THE INVESTIGATION

2.0 INVESTIGATION CONCERNING SURFACE DATA

In the first phase the probability of rain at only one station (De Bilt) was investigated as a function of surface data. These data are available back to 1901 and consist of hourly observations of

a) atmospheric pressure*
b) wind (direction and speed)
c) amounts of rain.

The amounts of rain in the periods 06–18 GMT and 18–06 GMT are found by integration of the hourly amounts. They are related to the pressure and wind observations at 12 and 00 GMT respectively, thus to observations in the middle of the periods 06–18 and 18–06 GMT.

2.1 INVESTIGATION CONCERNING AEROLOGICAL DATA

Aerological data are much fewer than surface data. Apart from incidental aircraft observations at Soesterberg and early kite-observations, there are few data available before 1947. In that year a fairly homogeneous series of radio-soundings was started. After a preliminary investigation it was decided to punch the most important aerological data. In the punch cards the following are some of the quantities to be found: temperature, dew-point and saturation deficit at the 1000, 850, 700 and 500 levels (TT, T_dT_d and Δ); the temperature difference between the 500 and 850 mb levels. The following surface data are also punched: air pressure and direction of the wind at De Bilt (pp, dd) and the amounts of rain (RR) at 10 stations in the Netherlands (fig. 1).

2.2 SELECTION OF THE VARIABLES

After the investigation of the partial influence of the several variables, a

* Since the bulk of pressure data consisted of readings expressed in units of mm mercury, this unit is used in this paper for the sake of convenience.

selection of these variables has been made. By a proper combination of the selected variables the resulting probabilities of rain should have maximum dispersion. With respect to the surface variables the result is given in fig. 14 and 15. The diagrams given here are independent of the season. With a view to its

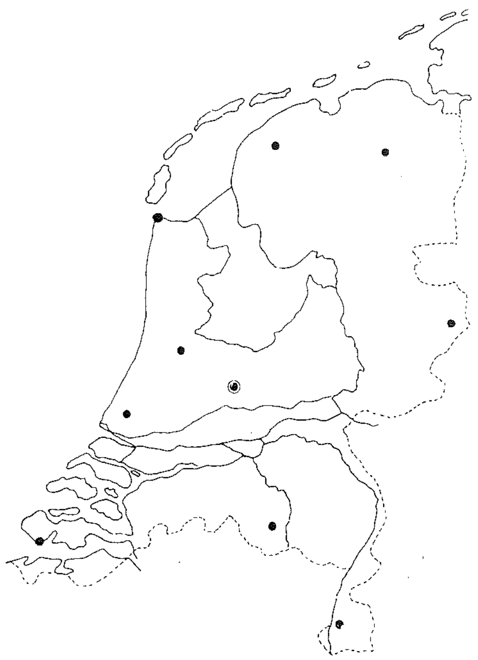


Figure 1. Distribution of reporting stations.

application in the field of rain-forecasting a division has been made into the following 4 groups:

March April and May	spring
June, July and August	summer
September, October and November	autumn
December, January and February	winter.

2.3 DISTRIBUTION OF THE RAINFALL

Supposing the values of the predictive variables to be known with sufficient accuracy in the middle of the forecasting period, it is then possible to determine the probability of rain at De Bilt. Referring to 1.3 it is also possible to forecast rain (R) or dry (D) at De Bilt. However, in practice an area forecast is made instead of a forecast for a fixed place. This is done by giving the fraction of the number of stations expected to have rain during the period of forecasting. The fractions are transposed into distributional terms and it is thus possible to make area forecasts. In this connection an investigation is made into the relation between the distribution of rainfall and the predictive variables.

2.4 APPLICATION AND TRIAL

During a test period the forecasting rules derived were applied to independent data. The observed values of the predictive variables were used in this case. For the purpose of comparison forecasts were also made in the conventional way by the meteorologists during this period (weather-service). It turned out that the dependency-index of the category of objective forecasts was a little higher than the dependency-index of the subjective forecasts made by the weather-service. However, it is difficult to draw a parallel, for as stated above, the observed values of the predictive variables were used and these values are not known at the beginning of the forecasting period. For an accurate comparison it is necessary, therefore, to estimate the values of the predictive variables in a conventional way. Therefore the resulting forecasts are not quite objective but semi-objective. These semi-objective forecasts have been compared with the subjective forecasts made by the weather-service during a second test-period. In the first trial the forecasting rules were concealed from the meteorologists so that there would be no bias. However, they had to be acquainted with these rules for the execution of the second trial, consequently, the trials had to be kept strictly apart.

3. SYSTEMATIC INVESTIGATION OF DEPENDENCY OF RAIN ON ONE VARIABLE

3.0 DEFINITION OF RAIN

There are many forms of hydrometeors, but in this investigation specification is avoided and all forms of precipitation are called "rain". Rain at a definite place and at a definite time may be characterized by its intensity. This variable is continuous and its use suggests that rain should be classified as a quantitative phenomenon. However, in statistics, a variable is called quantitative if its frequency distribution is continuous. However, the intensity of rain has not a continuous frequency distribution as zero-intensity has a finite frequency, whereas the frequencies of the remaining intensities are infinitesimal. On account of these considerations rain will be regarded as a qualitative variable with two classes, viz.: intensity 0 and intensity >0. The momentary intensity as a function of place and of time is extremely variable and therefore insufficiently representative. Moreover, its measurement presents difficulties from a technical point of view. That is why the sum of the intensity over a finite interval of time will be taken, viz. the amount of rain during a 12 hour period, denoted as RR. As it is not the intention to establish rules for forecasting precipitation originating from dew or fog, it is necessary to establish a convention by which the majority of such cases will be excluded. This has been done by fixing a minimum RR $=0.3$ mm. Rain, then, has two classes, RR <0.3 mm and RR $\geqslant 0.3$ mm, denoted by D and R respectively.

3.1 GENERAL PROCEDURE

In this section the probability of rain will be related to a number of meteorological quantities. For this purpose the dependency-index will be evaluated for each of those potential predictors separately. To ensure a clear understanding of the matter a description of the procedure to be followed will be given beforehand.

Referring to the preceding section, the predictand is described by a variate y which can assume two values, denoted as R and D, where *R* stands for: RR $\geqslant 0.3$ mm, and *D* for *RR* <0.3 mm. The predicted value y_p can also assume two values, likewise denoted as R and D. Finally, the correspondence between x on the one side and y on the other, is determined by:

$$y_p = R \quad \text{if} \quad p_R^{(x)} > p_R$$
$$y_p = D \quad \text{if} \quad p_R^{(x)} < p_R .$$

It turns out that this rule can be conveniently formulated in the following way: *rain (R) is predicted if the conditional probability, on account of the variate x, is greater than the climatological probability.* Clearly, climatological probability stands for the concept hitherto denoted as a priori probability.

Application of this rule gives rise to a contingency-table consisting of four numbers representing the numbers of cases involved plus two marginal totals.

		forecast		
		R	D	
observed	R	n_{00}	n_{01}	t_0
	D	n_{10}	n_{11}	t_1

The corresponding ratios n_{ij}/N, where $N = \sum n_{ij}$, may be identified with the frequencies f_{ij} defined in the foregoing.

As regards the estimator of the dependency-index, formula (6) will be chosen, in which the weighting factors w_{ij} given by (29) are inserted. At this point a difficulty is encountered, as p_j is not known. From sheer necessity an estimate of p_j must be accepted and as such the ratio t_j/N will be used. Accordingly, the expression for I becomes:

$$I = \frac{1}{2}\left(\frac{n_{00}}{t_0} - \frac{n_{01}}{t_0} - \frac{n_{10}}{t_1} + \frac{n_{11}}{t_1}\right).$$

On account of the relations:

$$n_{00} + n_{01} = t_0$$
$$n_{10} + n_{11} = t_1$$

it is possible to reduce it to:

$$I = \frac{n_{00}}{t_0} + \frac{n_{11}}{t_1} - 1\,. \tag{34}$$

Owing to the compromise of using this special estimate of p_j the formula proves to be very handy in practice. On the other hand, however, one should bear in mind that its use has certain implications for the validity of the derivation of the formula for the variance of I (27). This effect is very complicated, and an attempt to evaluate it has therefore been abandoned, especially since it becomes smaller and smaller with increasing N. For this reason formula (27) simply will be used; moreover, the same estimate for p_j is used in this formula as well.

3.2 SURFACE VARIABLES

3.2.0 *Persistence*

Many meteorological variables show persistence, i.e. the value at a given moment depends on the value they had at a preceding moment. It is known from experience that rain demonstrates persistence as well. This can be shown by using the amount of precipitation in the preceding period RR_{-1} as a predictor. Let:

$$y_p = R \quad \text{if} \quad RR_{-1} \geqslant 0.3\,mm$$

$$y_p = D \quad \text{if} \quad RR_{-1} < 0.3\,mm\,.$$

Splitting up the available data according to x, the following contingency-table is obtained:

		forecast R	forecast D	
observed	R	6819	5127	11946
observed	D	5099	20023	25122

table 3.0

From this table it follows, by substitution of the values concerned in (34):

$$I = \frac{6819}{11946} + \frac{20023}{25122} - 1 = 0.37 .$$

Additionally, it follows:

$$p_0 = \frac{11946}{37068} = 0.32 \text{ (the a priori probability)}$$

$$p_1 = \frac{25122}{37068} = 0.68 .$$

Consequently: $4p_0p_1 = 0.87$; $I^2 = 0.14$.
Substitution in (27) yields:

$$\text{var } I = \frac{1.15 - 0.14}{37068} = 0.0000272 .$$

So: $$\sigma_I = 0.005 .$$

The value of the standard deviation of the estimate I will be added, between brackets, as an indication of the accuracy of the estimate.

3.2.1 *Atmospheric pressure pp.*

A study of the literature on objective methods of rain-forecasting shows that atmospheric pressure as a predictive variable has gradually declined in importance. After 1945 authors no longer use pressure as a substantive predictor, but pressure differences, isobaric patterns etc. As far as is known, the predictive value of pressure has not been investigated in the synoptic field since 1917. (See references). This fact is especially striking as it will be shown that probability of rain is strongly dependent on this predictor.
Fig. 2 represents the conditional probability of rain as a function of pressure as measured in the middle of the 12-hour period over which RR is determined. As has been mentioned before, the a priori probability equals $p_0 = 0.32$. From the graph it follows that:

$$p_{pp} > p_0 \quad \text{if} \quad pp < 761\, mm$$

Consequently: $$y_p = R \quad \text{if} \quad pp < 761\, mm$$
$$y_p = D \quad \text{if} \quad pp \geqslant 761\, mm .$$

According to the rule stated in 3.1 this gives rise to the following contingency-table:

		forecast *R*	forecast *D*	
observed	*R*	8922	3024	11946
observed	*D*	8240	16882	25122

table 3.1

$$I = \frac{8922}{11946} + \frac{16882}{25122} - 1 = 0.42(0.005).$$

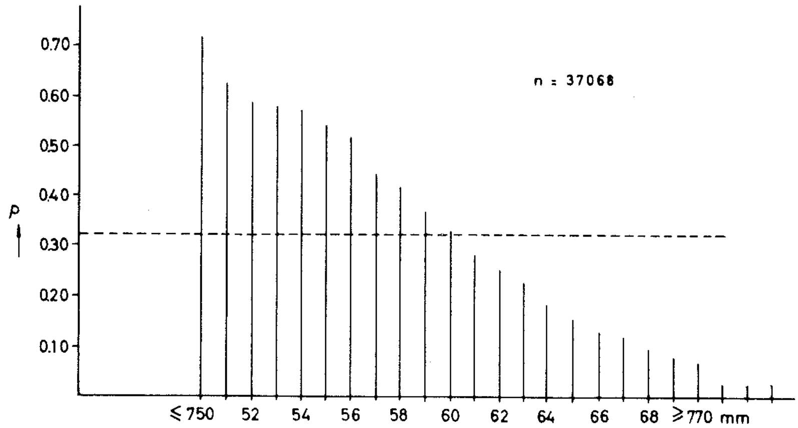


Figure 2. Probability of rain as a function of atmospheric pressure *pp*

Just as in this example the predictor is in general the momentary value of some variable. So a decision has to be made as to the time of observation to be chosen for that variable. It is evident that this should be done in such a way that the corresponding value of I is a maximum. For this purpose, the dependency of rain on pressure has been determined as a function of t, the time of observation with respect to the middle of the 12-hour period to which RR refers. The result is as follows:

t	-9	-3	0	3
I	0.42	0.46	0.48	0.46

As might be expected, the dependency is strongest if the time to which the predictor refers coincides with the middle of the period to which the predictand refers. In view of this result, this policy will in future always be followed with regard to the time to which the predictor refers.

It should be noted that, in performing this experiment, a relatively small sample was used, which is the cause of the difference between the value $I=0.48$ for $t=0$ and $I=0.42$ in table 3.1. It may be expected however that this sampling-effect does not influence considerably the differences in table 3.2.

3.2.2. *Pressure-tendency Δp*

In order to investigate the influence of the pressure-tendency a relatively small sample consisting of 1797 cases has been considered. The predictor was the pressure-change during a 12-hour period, the end of which coincides with the

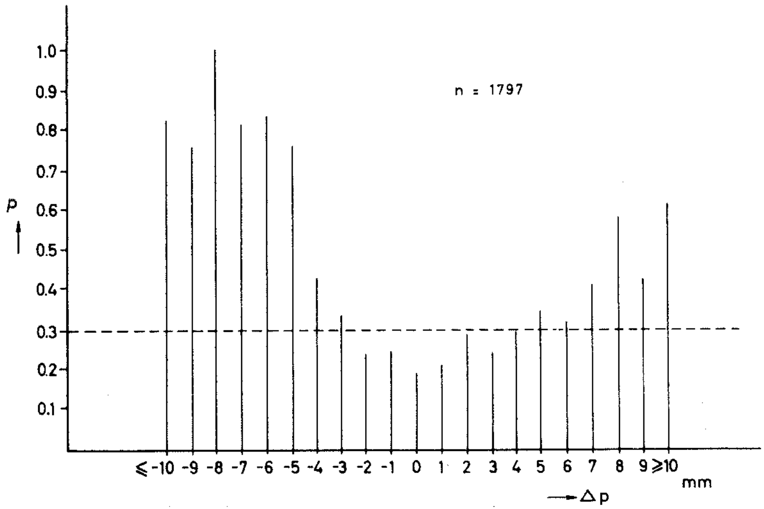


Figure 3. Probability of rain as a function of pressure-tendency Δp

middle of the period to which RR refers. The probability of rain as function of this predictor is shown in fig. 3. From this curve it can be derived that:

$$y_p = R \quad \text{if} \quad \begin{cases} \Delta p < -2 \\ \Delta p > 4 \end{cases}$$
$$y_p = D \quad \text{if} \quad -2 \leq p \leq 4.$$

This gives rise to the following table:

forecast

		R	D	
observed	R	225	315	540
	D	228	1029	1257

$$I = \frac{225}{540} + \frac{1029}{1257} - 1 = 0.23\,(0.025).$$

3.2.3 *Direction of the wind (dd)*

The compass-card is divided up into 16 directions. A separate class is reserved for low wind speed (*ff* < 3 kts) corresponding to variable wind. The probality of rain as a function of the wind direction is shown in fig. 4. From this graph

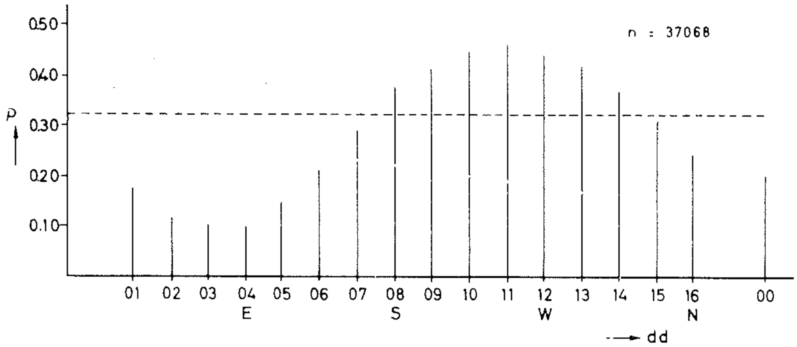


Figure 4. Probability of rain as a function of direction of surface wind dd.

it can be seen that the conditional probability is greater than the a priori probability in case of wind directions between south and northwest. This reflects the effect of the geographical situation of the Netherlands with respect to the North Sea. More especially it follows that:

$$y_p = R \quad \text{if} \quad dd = 08, 09, \quad \ldots 14$$
$$y_p = D \quad \text{if} \quad dd = 15, 16, 01 \ldots 07; 00.$$

This gives rise to the following contingency-table:

forecast

		R	*D*	
observed	*R*	8844	3102	11946
	D	12198	12924	25122

$$I = \frac{8844}{11946} + \frac{12924}{25122} - 1 = 0.25\,(0.005).$$

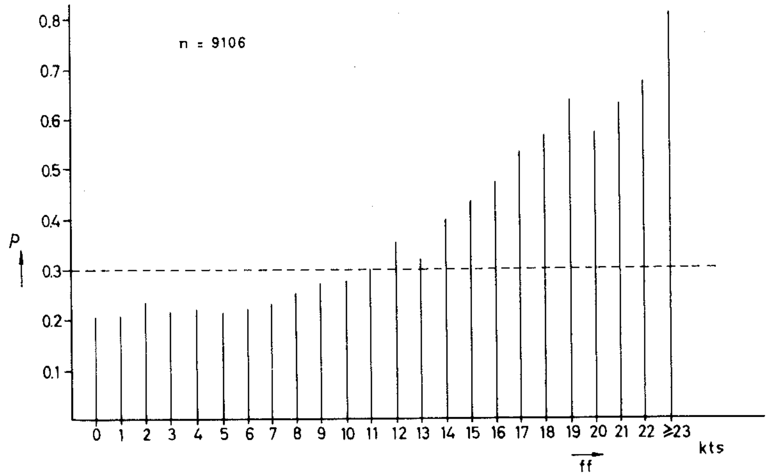


Figure 5. Probability of rain as a function of wind speed (surface)

3.2.4 *Wind speed*

The dependence of rain on wind speed (surface wind) has been investigated on the basis of 9106 cases. The result is represented by fig. 5. The critical value appears to be 10 knots. It leads to the following contingency-table:

forecast

		R	*D*	
observed	*R*	981	1655	2636
	D	1294	5176	6470

$I = 0.17$ (0.01)

table 3.5

A summary of the investigation on the dependency of 5 predictors referring to ground observations is given in the following table:

predictor	*I*
persistence	0.37
pressure	0.42
pressure-tendency	0.23
wind direction	0.25
wind speed	0.17

table 3.6

3.3 AEROLOGICAL VARIABLES

In the introduction it has been noted that one of the things by which the probability of rain is determined is the state of the upper air. In order to prove this assertion, it will be necessary to characterize this state by the value of a number of parameters at different levels. To this end, consideration will be given to humidity and temperature and probably also to the height of a level of constant pressure, the latter on account of the fact that this quantity may be regarded as equivalent to surface pressure, which proved to be correlated strongly with the probability of rain.

3.3.0 *Humidity*

As is probably known, the humidity at a certain level can be defined by various quantities of which we would mention:

1. relative humidity (U)
2. dewpoint depression ($T-T_d$)
3. saturation deficit (Δ)

The last quantity needs some explanation. The saturation deficit is defined as the mass of water which can evaporate in a unit volume of air before the air becomes saturated.

Let:
- w = mixing ratio
- w_m = maximum mixing ratio at a given temperature T
- e = vapour pressure
- e_m = maximum vapour pressure
- V = specific volume of the air
- p = pressure
- R = gas constant,

then the following relation holds:

$$\Delta = \frac{w_m - w}{V} \approx 0.62 \frac{(e_m - e)}{p} \frac{p}{RT} = 0.62 \frac{(e_m - e)}{RT}.$$

Apparently, the saturation deficit is practically independent of pressure.

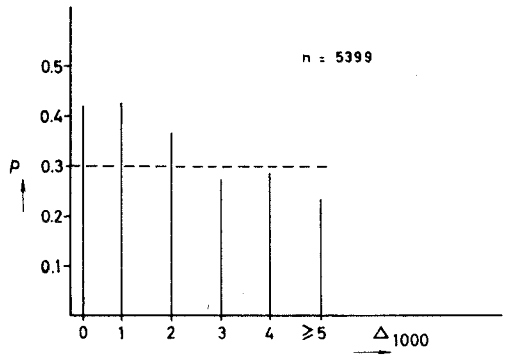


Figure 6. Probability of rain as a function of saturation deficit Δ at the 1000 mb level

For reasons to be explained in section 5.1, there is a preference for the use of this particular humidity indicator.

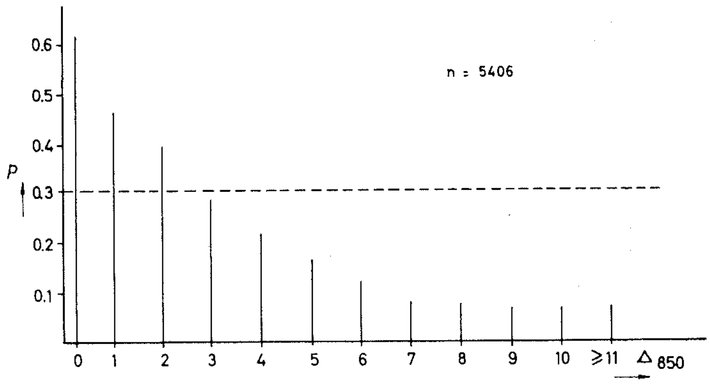


Figure 7. Probability of rain as a function of saturation deficit Δ at the 850 mb level

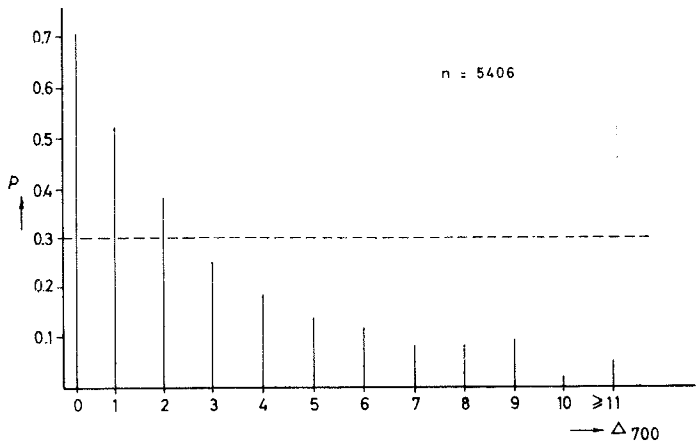


Figure 8. Probability of rain as a function of saturation deficit Δ at the 700 mb level.

The available data consisted of about 5000 radio soundings made during the period 1947–1952 inclusive. These data made it possible to evaluate the probability of rain as a function of Δ at four levels. The result is represented in the figures 6–9. In the light of these relationships the following contingency-tables can be composed:

forecast

observed		R	D	
	R	1101	536	1637
	D	1555	2207	3762

1000 mb

table 3.7

forecast

observed		R	D	
	R	1208	436	1644
	D	1324	2438	3762

850 mb

table 3.8

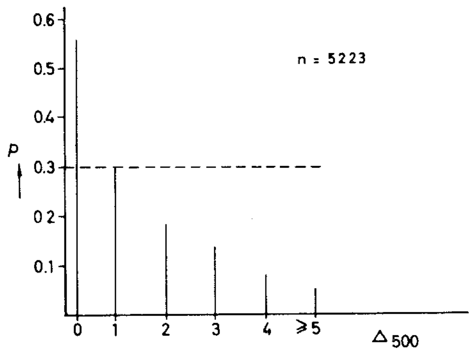


Figure 9. Probability of rain as a function of saturation deficit Δ at the 500 mb level.

forecast

		R	D	
observed	R	1152	493	1644
	D	1102	2660	3762

700 mb

table 3.9

forecast

		R	D	
observed	R	518	1067	1585
	D	428	3210	3638

500 mb

table 3.10

The number of cases is not the same in the four tables, owing to the fact that on the one hand not all soundings reached the highest levels, whereas on the other hand in some cases of low surface pressure the 1000 mb level was missing.

The corresponding dependency indices are shown in the next table:

level	I	$S(I)$
1000 mb	0.26	0.014
850 mb	0.38	0.014
700 mb	0.41	0.014
500 mb	0.21	0.014

table 3.11

An explanation of the small figure at the 1000 mb level may be that a mere cooling of the air during the night causes high humidity in the layers near the surface without involving an increase in the probability of rain. On the other hand, the low index at 500 mb may be due to the fact that processes producing rain in many cases take place at lower levels.

3.3.1 *Temperature*

There is no reason to assume that temperature itself has an influence on the probability of rain. However, vertical temperature differences may have some effect on account of their relation to thermodynamic stability. Fig. 10 shows the dependence of the probability rainfall on the difference $T_{850}-T_{500}$. The corresponding contingency-table is:

forecast

observed		*R*	*D*	
	R	243	221	464
	D	309	532	841

I=0.15 (0.03)

table 3.12

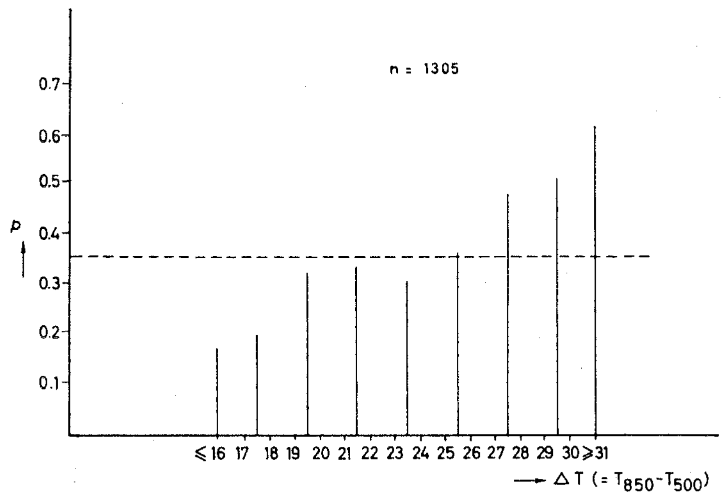


Figure 10. Probability of rain as a function of temperature difference $T_{850}–T_{500}$

The dependence appears to be rather weak, which is understandable when one considers that in the case of frontal rain the vertical structure is often stable. If it were possible to make a distinction between rain and showers *beforehand*, data could be divided up into two parts, one referring to the frontal rain and the other to showers. The latter would probably show a greater dependence on the temperature difference.

3.3.2 *Height of the* 500 *mb level* (h_{500})

Fig. 11 shows the dependence of the probability of rain on the height of the 500 mb level. The data consist of a sample of 1030 cases during summer months in the period 1947–1952. The corresponding contingency-table is:

		forecast		
		R	*D*	
observed	*R*	242	44	286
	D	266	478	744

table 3.13

I=0.49 (0.03)

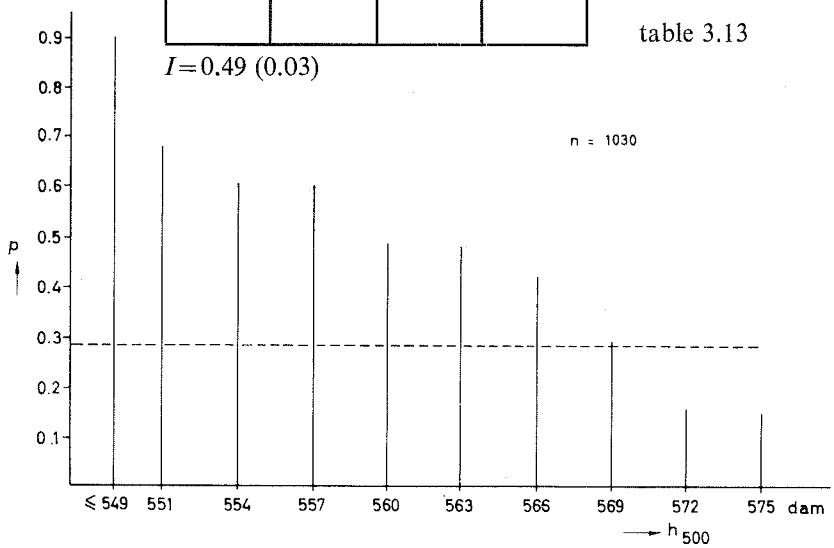


Figure 11. Probability of rain as a function of height of the 500 mb level

4. DEPENDENCE OF RAIN ON TWO OR MORE VARIABLES

The dependence on a number of variables V_1, V_2, V_3 etc. can be reduced to dependence on one variable x, by defining x as the probability of the occurrence y for given values of V_1, V_2 etc.:

$$x = p(y|V_1, V_2 \ldots)$$

According to the arrangement made in section 1.3, a correspondence is fixed between the predictor y_p and the variable x in such a way, that:

$$y_p = 0 \quad \text{if} \quad x > p_0$$
$$y_p = 1 \quad \text{if} \quad x > p_1$$

where p_i is the a priori probability of $y = y_i$.
From the foregoing it follows that the problems of determining the dependency-index consists substantially in evaluating the probability of the occurrence of y as a function of the variables V_i. Actually, the probability is estimated by splitting up the data multi-dimensionally into $n_1 . n_2 \ldots n_i$ groups, according to n_i classes of the variable V_i. Especially if the number of variables and the number of classes become large the estimation is unreliable as the number of cases becomes too small. In order to avoid as far as possible the error, due to this fact, a procedure has been developed, which can be demonstrated best by means of an example. The example relates to the influence of the saturation-deficit at the 700 and 850 mb levels. In diagr. 1 the entry of each cell consists of a fraction; the numerator stands for the number of cases of rain, whereas the denominator stands for the total number of cases at the values V concerned. The relative frequency of rain is plotted in diagr. 2. The diagrams are analysed by drawing isopleths of probability. The principle underlying the procedure is the assumption that the probability p is a monotonic decreasing function of V_1 and V_2. It appears, however, that the relative frequencies do not satisfy this assumption; for instance, in the uppermost horizontal line, p increases from 0.37 to 0.38 if Δ_{700} increases from 5 to 6. Such discrepancies will be considered attributable to the fact that the samples in some of the classes are too small. Although this effect causes difficulties in the drawing of isopleths, it seems nevertheless to be possible to perform the analysis in an objective and unique way, as will be shown hereunder. Starting with the isopleth $p = 0.50$, this line should, according to the assumption made above, be drawn in such a way that only values greater than 0.50 are

Δ_{850} ↓ \ Δ_{700} →	0	1	2	3	4	5	6	7	8	9	10	11	12	13
0	239/311	98/162	42/77	20/48	17/45	10/27	6/16	6/19	1/7	4/12	2			
1	90/123	157/265	87/192	34/105	12/60	12/54	12/40	3/23	3/20	2/7	2	1/2	1	
2	22/41	123/217	102/231	48/147	23/92	14/75	13/54	2/28	2/11	1/8	1/5	1	1/2	
3	12/24	29/85	47/135	35/130	23/89	11/51	5/38	3/31	2/13	2/15	2	2	1	
4	6/13	22/61	24/93	21/117	17/89	9/58	7/50	4/19	2/18	2	6	1	1/1	1
5	2/9	10/25	15/62	12/74	12/74	7/53	1/22	1/26	1/17	7	4		3	
6	3/5	4/12	7/26	5/47	6/57	2/38	1/36	1/23	2/10	8	2	2	2	
7	0/5	1/14	3/17	4/34	4/52	3/42	3/30	29	1/10	1/14	7	1		
8		3	10	7/27	3/32	39	22	3/21	8	1/3	5	1	2	
9		1/4	1/10	3/16	1/14	24	24	1/17	19	9	1	3		
10	0/1	1/1	3	4/9	1/17	1/14	28	22	7	4	4	2	1	
11			2/5	1/5	6	2/12	13	16	7	8	4	1	1/1	
12		1/1	5	7	4	13	8	2/21	9	11	4	2	1	
13		1	3	1/3	7	6	1/8	13	9	1	5	3	2	2
14		1	1	1	3	4	3	5	8	1/5	6	2	1	
15				3	4	3	5	5	2	5	3	1/2		
16				1	1	2		2	2	8	3	1		
17					1	2	1		1	7	2	3	3	
18				2	0	1	4	3	1	2	2	1	1	
19				1		1		1/1	2		1	2	1	1
20						1		1	5	1	2	1	4	2

Probability of rain as a function of two variables:
Δ_{850} and Δ_{700} (defined in 3.3.0).
In each cell the numerator of the fraction represents the cases of rain occurrence and the denominator stands for the total number of cases.

Diagram 1

Δ_{850} ↓ \ Δ_{700} →	0	1	2	3	4	5	6	7	8	9	10	11	12	13
0	0.77	0.60	0.55	0.42	0.38	0.37	0.38	0.32	0.14	0.33	0.00			
1	0.73	0.59	0.45	0.32	0.20	0.22	0.30	0.13	0.15	0.28	0.00	0.50	0.00	
2	0.54	0.56	0.44	0.33	0.25	0.19	0.24	0.07	0.18	0.12	0.20	0.00	0.50	
3	0.50	0.36	0.35	0.27	0.26	0.22	0.09	0.10	0.15	0.13	0.00	0.00	0.00	
4	0.46	0.36	0.26	0.18	0.19	0.16	0.14	0.21	0.11	0.00	0.00	0.00	0.00	0.00
5	0.22	0.40	0.24	0.16	0.16	0.13	0.05	0.04	0.06	0.00	0.00	0.00	0.00	
6	0.60	0.33	0.27	0.11	0.11	0.05	0.03	0.04	0.20	0.00	0.00	0.00	0.00	
7	0.00	0.08	0.18	0.12	0.08	0.07	0.09	0.00	0.10	0.07	0.00	0.00		
8		0.00	0.00	0.26	0.10	0.00	0.00	0.14	0.00	0.33	0.00	0.00	0.00	
9		0.25	0.10	0.19	0.07	0.00	0.00	0.06	0.00	0.00	0.00	0.00		
10	0.00	1.00	0.00	0.44	0.06	0.07	0.00	0.00	0.00	0.00	0.00	0.00	0.00	
11			0.40	0.20	0.00	0.08	0.00	0.00	0.00	0.00	0.00	0.00	1.00	
12		1.00	0.00	0.00	0.00	0.00	0.00	0.09	0.00	0.00	0.00	0.00	0.00	
13		0.00	0.00	0.33	0.00	0.00	0.12	0.00	0.00	0.00	0.00	0.00	0.00	0.00
14		0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.20	0.00	0.00	0.00	
15				0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.50		
16				0.00	0.00	0.00		0.00	0.00	0.00	0.00	0.00		
17					0.00	0.00	0.00		0.00	0.00	0.00	0.00	0.00	
18				0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	
19				0.00		0.00		1.00	0.00		0.00	0.00	0.00	0.00
20						0.00		0.00	0.00	0.00	0.00	0.00	0.00	0.00

a = 0.50-isoline
b = 0.20-isoline
c = 0.10-isoline
d = 0.05-isoline
e = 0.02-isoline

Diagram 2. Probability of rain as a function of Δ_{850} and Δ_{700}

found in the area to the upper left of this line. At first sight then, the 0.50 isopleth has been drawn as shown in diagr. 2. However, on closer inspection, it turns out that the cell $\{^{V_1=6}_{V_2=0}$ with $p=0.6$, does not fit in with this arrangement. The alternative is to include this cell in the area where $p>0.50$, but in such a case the interjacent cells (5.0) and (4.0) would have to be included as well. The total relative frequency in those three cells amounts to $\frac{6+2+3}{13+9+5}\frac{11}{27}=0.41$ and is, therefore, smaller than the frequency of the area to which it would be added. For this reason the inclusion is abandoned. If, on the other hand, the frequency had exceeded the fixed limit, the course of the isopleth would have been modified accordingly.

A start is usually made with a preliminary delimitation of an area where $p>p_0$. If there appear to be isolated cells outside this area in which also

$p > p_0$, an investigation is made into whether an extension in the direction of the lower-right (corresponding with the assumed decreasing probability) would be justified. The criterion for such an extension is that, for the area included, as a whole the condition $p > p_0$ must hold good.

RR₋₁ ↓ / RR₋₂ →	< 0.3	0.3–1.4	1.5–2.5	2.6–3.4	3.5–4.5	4.6–9.4	9.5–19.4	19.5–49.4	49.5–99.4
< 0.3	841/5251	168/643	77/214	27/97	31/88	47/126	14/40	0/2	
0.3– 1.4	259/559	110/266	59/111	34/71	35/63	41/103	12/22	2/3	
1.5– 2.5	133/222	57/91	29/57	16/33	10/23	19/45	4/10		0/1
2.6– 3.4	78/123	38/55	13/25	9/15	11/14	11/22	3/5		
3.5– 4.5	56/95	25/38	13/25	8/16	6/11	16/22	6/10	1/2	
4.6– 9.4	107/155	58/88	23/39	12/23	9/16	28/39	7/11	1/2	1/1
9.5–19.4	29/43	11/17	7/11	3/5	4/6	6/15	4/6	0/1	
19.5–49.4	4/5	1/1	1/1			1/1	1/2		
49.5–99.4		2/2							

The numerators of the fractions indicate the number of rain occurrences, the denominators represent the total number.

Diagram 3a. Probability of rain as a function of the amount of precipitation in the preceding 12-hour periods. The numerators of the fractions indicate the number of occurrences, the denominators represent the total number.

RR₋₁ ↓ / RR₋₂ →	< 0.3	0.3–1.4	1.5–2.5	2.6–3.4	3.5–4.5	4.6–9.4	9.5–19.4	19.5–49.4	49.5–99.4
< 0.3	0.16	0.26	0.36	0.28	0.35	0.37	0.35	0.00	
0.3– 1.4	0.46	0.41	0.53	0.48	0.56	0.40	0.55	0.67	
1.5– 2.5	0.60	0.63	0.51	0.48	0.44	0.42	0.40		0.00
2.6– 3.4	0.63	0.69	0.52	0.60	0.79	0.50	0.60		
3.5– 4.5	0.59	0.66	0.52	0.50	0.54	0.73	0.60	0.50	
4.6– 9.4	0.69	0.66	0.59	0.52	0.56	0.72	0.64	0.50	1.00
9.5–19.4	0.68	0.65	0.64	0.60	0.67	0.40	0.67	0.00	
19.5–49.4	0.80	1.00	1.00			1.00	0.50		
49.5–99.4		1.00							

Diagram 3b. Relation of probability of rain and persistence.

Although this procedure reduces the risk of drawing the isopleths incorrectly, nevertheless sampling effects cannot be avoided completely, as can be gathered from the rather whimsical course of the isopleths. For this reason it is advisable in any case to prevent as far as possible the occurrence of cells containing a small number of cases. This can be done by limiting the number of variables and the number of classes of those variables.
The procedure outlined above will be applied in what follows.

4.0 PERSISTENCE

In section 3.3.0 it has been shown that the probability of rain depends on the amount of precipitation in the preceding 12-hour period, this amount being specified in two classes. The question arises whether the amount of precipitation of the period previous to the preceding period, denoted as RR_{-2}, has some influence as well. When investigating this assumption, the specification of RR will moreover be extended by the introduction of 9 classes. The investigation was performed by using punch cards. The result is shown in diagr. 3a. The corresponding relative frequencies are represented in diagr. 3b. From the latter table it follows that:

$$y_p = R \quad \text{if} \quad 1.\ RR_{-1} \geqslant 0.3\,mm.$$
$$2.\ RR_{-1} < 0.3\,mm \text{ while } RR_{-2} \geqslant 1.5 \text{ mm.}$$
$$y_p = D \quad \text{if} \quad RR_{-1} < 0.3\,\text{mm while } RR_{-2} < 1.5 \text{ mm.}$$

This policy leads to the following contingency-table:

forecast

observed		R	D	
	R	1630	1009	2639
	D	1585	4885	6470

I=0.38 (0.01)

table 4.0

When using only the variable RR_{-1} the result would have been in the same sample:

forecast

		R	D	
observed	R	1434	1205	2639
	D	1214	5256	6470

$I=0.35$ (0.01)

table 4.1

It is true that a small, though significant, improvement is obtained. However, the use of the variable RR_{-2} would complicate the further investigation to such a degree that the minor gain does not warrant it.

Note: Although the concept persistence refers to the effect of the variable RR_{-1}, this variable itself will also be denoted as such, for the sake of convenience.

There is another aspect regarding the use of the variable RR_{-1} that deserves to receive some attention. The amount of precipitation can be regarded as a reflection of the state of the atmosphere in the recent past. However, it is known that in the case of an unstable atmosphere the rainfall may have a local character, i.e. the amount of precipitation is more or less subject to chance. In order to lessen the effect of chance it would be preferable to have the state of the atmosphere described by the amount of rain at a number of stations. For this purpose 10 key-stations have been selected (see fig. 1). Let n be the number of those 10 which report R during the preceding 12-hour period, then n may be used as a variable representing the persistence. This investigation has not yet been carried out owing to the fact that sufficiently long series of adequate precipitation data are not available for many stations.

4.1 ATMOSPHERIC PRESSURE AND PERSISTENCE

Where one of the variables is specified in only two or three classes, it is not necessary to utilize a diagram like the one described before. It is more efficient

in such a case to draw a graph of the probability of rain as a function of that variable which is specified in many classes, for each value of the other one. As the variable $P = RR_{-1}$ has two classes, this procedure may be applied here.
Fig. 12a and b represent the probability of rain as a function of *pp* for respectively $RR_{-1} \geqslant 0.3$ mm (a) and $RR_{-1} < 0.3$ mm (b). According to the rule stated in 3.1, the following policy must be observed:

$$y_p = R \quad \text{if} \quad pp \leq 758 \text{ mm while } RR_{-1} < 0.3 \text{ mm, or } P = D$$
$$y_p = R \quad \text{if} \quad pp \leq 765 \text{ mm while } RR_{-1} \geqslant 0.3 \text{ mm, or } P = R.$$

The result is:

forecast

		R	*D*	
observed	*R*	9260	2686	11946
	D	7780	17342	25122

$I = 0.47$ (0.005)

table 4.2

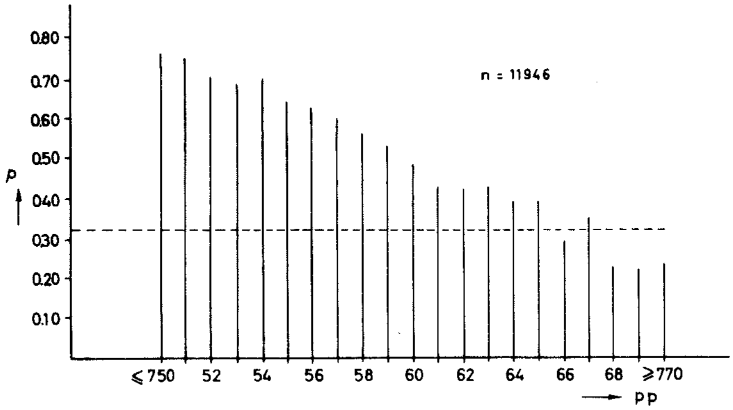


Figure 12a. Probability of rain as a function of atmospheric pressure for $RR_{-1} \geqslant 0.3$ mm

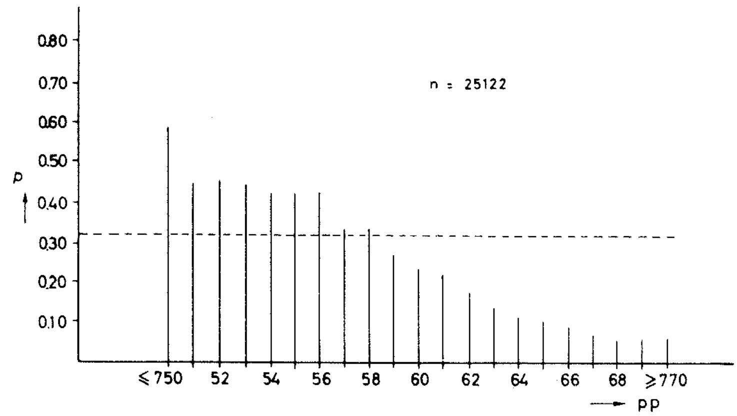


Figure 12b. Probability of rain as a function of atmospheric pressure for $RR_{-1} < 0.3$ mm

4.2 DIRECTION OF THE WIND AND PERSISTENCE

From fig. 13a it follows that in the case of rain in the preceding period ($P=R$), the probability is greater than the a priori probability, irrespective of the wind direction. On the contrary, if $P=D$, (fig. 13b) then the probability of rain is smaller than the a priori probability, again irrespective of the wind direction. It seems that the effect of persistence is so strong that it overshadows the effect of wind direction.

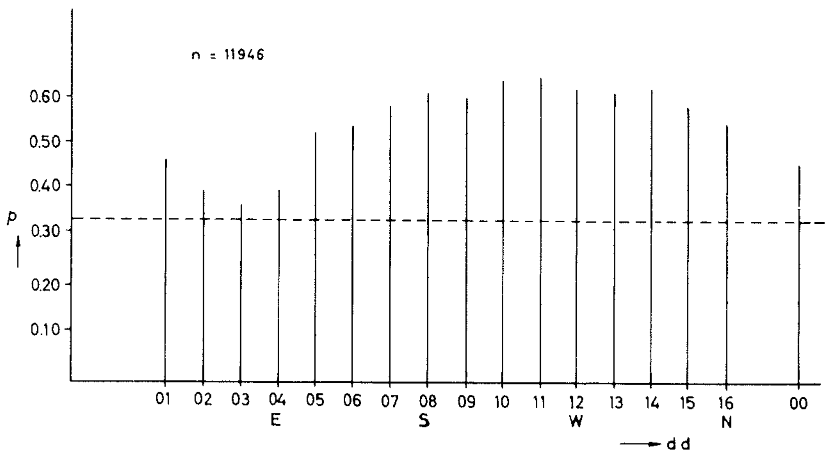


Figure 13a. Probability of rain as a function of direction of wind for $RR_{-1} < 0.3$ mm

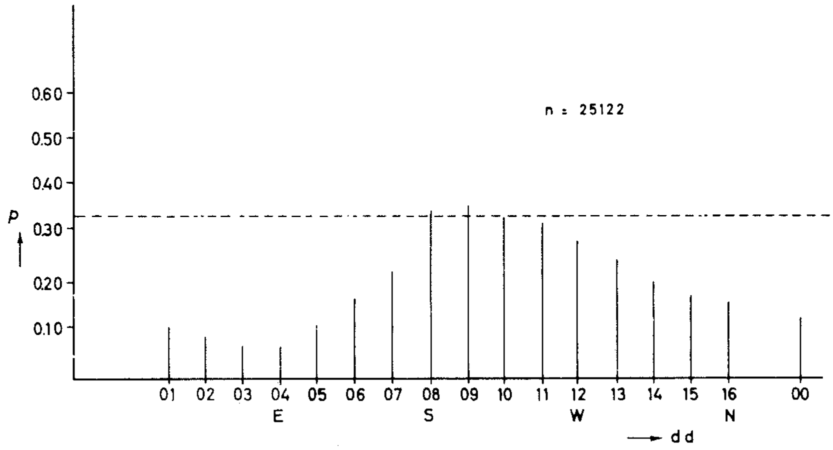


Figure 13b. Probability of rain as a function of direction of wind for $RR_{-1} \geqslant 0.3$ mm

4.3 WIND DIRECTION AND PRESSURE

In fig. 14 the isopleths of probability, as a function of *dd* and *pp*, have been drawn. The course of the 0.30 isopleth is important, as this isopleth practically coincides with the 0.32 isopleth, which corresponds with the a priori probability p_0. The 0.30 isopleth divides the diagram into two parts, of which the upper part refers to cases with $p > p_0$ and the lower part to cases with $p < p_0$
This division gives rise to the following table:

forecast

		R	*D*	
observed	*R*	9189	2757	11946
	D	7775	17347	25122

table 4.3

$I = 0.46$ (0.005)

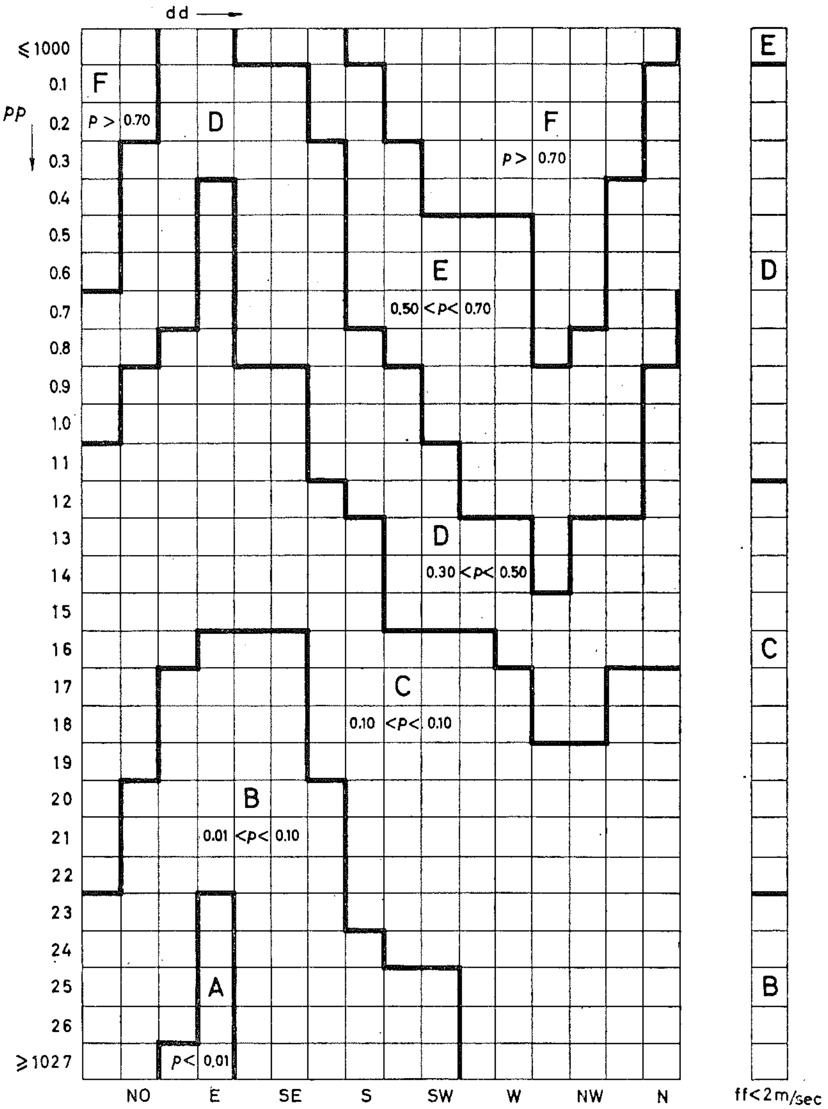


Figure 14. Probability of rain as a function of atmospheric pressure and direction of surface wind.

$p=0.50$ $p=0.30$ $p=0.10$ $p=0.05$ $p=0.02$

h_{500} ↓ \ pp →	00	02	04	06	08	10	12	14	16	18	20	22	24	26	28	30
⩽ 550	8/8	4/4	4/5	3/4	2/4	4/5										
552	1/2	1/1	2/2	1/2	2/2	2/4	1/1									
554	3/3	0/1	3/3	0/3	1/3	5/7	1/1	0/1			0/1					
556	5/6	2/2	8/9	0/1	1/1	2/5	2/3	1/2			1/1					
558	3/3	1/2	2/2	3/4	2/6	3/4	2/5	0/2	0/2		0/1					
560			2/2	4/6	4/5	2/4	6/9	1/7	0/4	0/2	1/5	0/2				
562	2/2	3/3	3/5	5/6	5/10	5/7	5/12	1/5	2/5	1/3	2/3	0/1				
564			1/1	2/2	5/6	9/14	4/12	3/7	4/8	1/4	0/5	0/3	0/1			
566				2/2	0/1	10/11	5/9	2/10	1/5	4/7	0/4	1/3	0/1			
568					4/4	7/13	6/12	4/6	6/21	1/12	2/8	1/3	1/5	0/1		0/1
570			1/1	2/4	1/1	4/8	3/9	0/8	2/7	3/14	0/8	0/2	0/7	0/1	1/1	
572					1/4	2/5	3/5	1/9	1/12	2/9	0/5	0/5	0/6	0/2	0/1	
574				0/1	1/3	2/5	0/4	2/6	3/14	2/16	2/12	1/11	0/4	0/1		0/1
576				0/1			4/9	0/10	1/13	1/16	1/20	2/8	2/3	0/2	0/1	
578				0/1			1/1	0/6	0/9	0/7	1/13	0/6	0/10	0/1		
580						0/1	0/3	0/8	0/6	1/16	0/28	0/19	0/13	1/6	0/1	0/1
582							1/3	0/4	0/3	0/6	0/10	0/10	0/6	0/4	0/1	0/1
584								0/2	0/2	1/4	0/2	1/6	1/8	0/1	0/1	
586								0/1	0/3	1/4	0/7	0/3	0/2		0/1	0/1
⩾ 588							0/1	1/3	0/2	0/9	0/8	0/9	0/7	0/4	0/1	0/1

Diagram 4. Probability of rain as a function of atmospheric pressure and height of 500 mb level

4.4 WIND DIRECTION, PRESSURE AND PERSISTENCE

In order to evaluate the combined effect of wind direction, pressure and persistence, the material has been divided up into two parts according to persistence P. Next, for each of these parts a diagram as described in the preceding section has been constructed. See fig. 15a and 15b. The corresponding contingency-table is:

forecast

		R	*D*	
observed	*R*	9700	2246	11946
	D	7833	17289	25122

$I=0.50$ (0.005)

table 4.4

4.5 PRESSURE AND HEIGHT OF 500 MB LEVEL

The analysis of the material with respect to pressure and height of 500 mb level is shown in diagr. 4. It results in the following contingency-table:

forecast

		R	*D*	
observed	*R*	227	59	286
	D	175	569	744

$I=0.56$ (0.03)

table 4.5

However, when using only pressure as predictive variable in the same sample, the table becomes:

forecast

observed		R	D	
	R	210	76	286
	D	136	608	744

$I = 0.55$

table 4.6

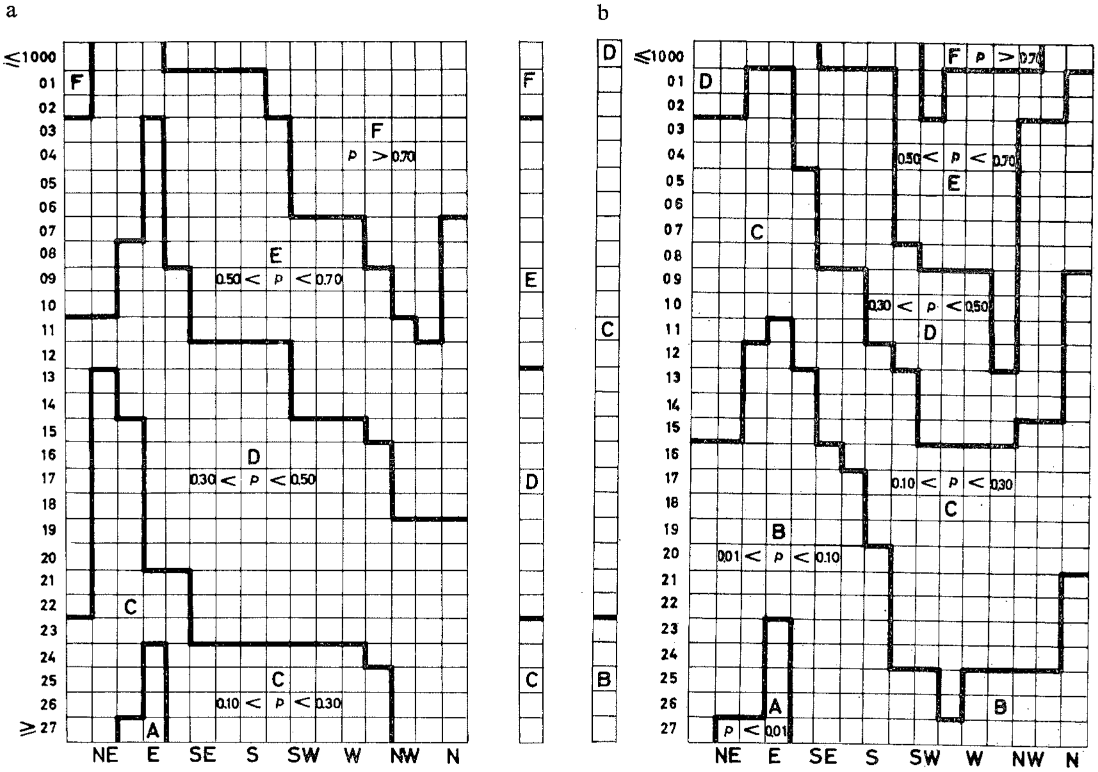


Figure 15a. Probability of rain as a function of atmospheric pressure and direction of surface wind for $RR_{-1} \geqslant 0.3$ mm

Figure 15b. Probability of rain as a function of atmospheric pressure and direction of surface wind for $RR_{-1} < 0.3$ mm

Consequently, the gain afforded by the variable h_{500} – if any – is very small. For this reason this variable has not been used as a predictor.

Note: It should be noted, by the way, that the values of I refer only to small samples taken during summer seasons. Accordingly, they are not representative of the dependence in general, in respect of all seasons. This experiment was merely meant as a comparison of the interdependence of h_{500}, pp and its combination.

A summary of the dependencyindices corresponding with the different combinations is given in the next table:

variables	I	s_I
RR_{-1}, RR_{-2}	.38	.010
pp, P	.47	.005
dd, P	.37	.005
pp, dd	.46	.005
pp, dd, P	.50	.005
H_{500}, pp	.56	.030

table 4.7

5. COMPOSED VARIABLES

Returning to the considerations put forward at the beginning of section 4, it should be noted that the isopleths of probability have yet another meaning; for an isopleth may also be interpreted as the locus of the points for which the composing variables V_1 and V_2 satisfy some relation. If, for example, an isopleth runs parallel to a diagonal from the lower left to the upper right in a diagram in which both axes are subdivided according to the same scale, it

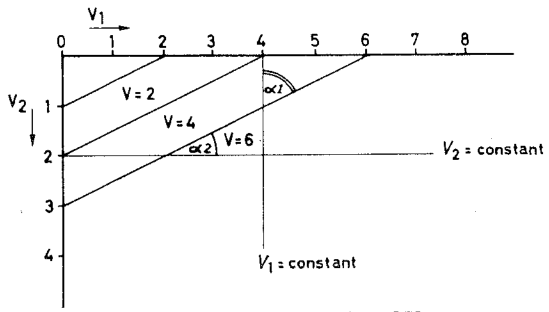


Figure 16. Illustration of composed variable $V = V_1 + 2V_2$

represents the locus of the points for which the sum of the variables is constant. It should also be noted that, owing to the finite increments of the variables, such a line is actually staircase-shaped. If all isopleths of probability were parallel to that diagonal, the problem of determining the dependence on two variables would be reduced to dependence on one variable, by introduction of the composed variable $V = V_1 + V_2$. This statement probably needs some elucidation.

Let the isopleth $p = p_0$ coincide with the isopleth $V = V_0$, then the value 0 is assigned to y_p in the area where $p > p_0$, and the value 1 in the complementary area where $p < p_0$. The material is thus split up into two parts according to the value of p, but the same division is accomplished when discriminating according to the value of V. Consequently, the division which is necessary to determine the dependency on the combination of V_1 and V_2 is produced by splitting up the material according to the value of the composed variable V.

Proceeding from the above-mentioned example, it can be stated quite generally that the isopleths of probability may also be regarded as isopleths of a variable which is a function of the components:

$$V = f(V_1, V_2).$$

It must be stressed that in doing so, the composed variable is merely defined formally, viz. as a quantity of which the isopleths coincide with the isopleths of probability of the predictand. Therefore, in general, V has no physical meaning, although it may so happen that it has. These two possibilities will be dealt with separately.

5.0 COMPOSED VARIABLE WITH PHYSICAL MEANING

Let, in the above-mentioned example, V_1 and V_2 stand for the saturation deficit at two levels, then $V = V_1 + V_2$ is a measure for the mean saturation-deficit averaged over the two levels to which V_1 and V_2 refer, giving each level the same weighting-factor. The same is true if V is a linear function of V_1 and V_2:

$$V = aV_1 + bV_2$$

in which case the levels have unequal weighting-factors. Indeed in some cases a weighted mean of some quantity may be interpreted physically.
In fig. 16 the isopleths have been drawn of the composed variable:

$$V = V_1 + 2V_2.$$

Let α_1 be the angle between the isopleth of V and that of V_1. From the figure it follows that $\alpha_2 < \alpha_1$, hence the smaller angle corresponds with the greater weighting factor. This result may be generalized for the case of curved isopleths:
RULE: *If V is a composed variable, then the greater weight corresponds to that variable of which the isopleth intersects the isopleth of probability under the smaller angle. In particular, if the isopleth of a variable intersects the isopleth of probability perpendicularly, the probability is independent of that variable.*
Up to now the drawing of isopleths has been primary, the lines having been interpreted as isopleths of a composed variable. This sequence may be reversed so that first the isopleths of a composed variable are drawn, followed by the hypothesis that these isopleths are identical with isopleths of probability. The essential difference between the two methods is that in the former case the isopleths of probability are drawn empirically, which involves the use of a great number of degrees of freedom, in the latter case, however, this number is much smaller, owing to the fact that an analytical function is used to describe the relation between V and its components. In the example of the linear combi-

nation for instance, this number is actually one, viz. the ratio of the weighting-factors. As is probably known, this number of degrees of freedom is closely related to the reliability of the result. For this reason, preference should be given to the latter sequence.

The component variables need not necessarily be similar quantities referring to different layers, as can be seen from the following example. Let $V_1 = T_{850}$ (temperature at 850 mb level) and $V_2 = T_{d_{850}}$ (the corresponding dew point). Referring back to 3.3.0, it will be remembered that the humidity at a specified level may be described by several indicators. Each of them is uniquely determined by T and T_d, p being constant. It is true that the relationship may be rather complicated, but it does exist, expressed either in the form of tabulations or in an analytical form. This circumstance has been used to select that indicator which is most closely related to the probability of rain.

In fig. 17, isopleths have been drawn of the probability of rain in dependence on T_{850} and $T_{d_{850}}$ as well as of this probability in dependence of the humidity indicators u (relative humidity) and Δ (saturation deficit). Although the former ones are rather irregular, owing to the smallness of the samples used, they clearly converge in the lower right part of the figure to line $T - T_d$, as do the isopleths of the saturation deficit. In contrast to this the isopleths of u diverge in that direction, whereas those of $T - T_d$ (which have not actually been drawn) are parallel. Consequently, the Δ-isopleths intersect the isopleths of probability under the smallest angle. Consequently, on account of the rule laid down in this section, the probability of rain depends on this humidity indicator more than on the other ones.

A second application of this rule is made in the examination of the combination of h_{500} and the atmospheric pressure pp. From diagr. 4 it can be seen that of the isopleths of the two components those of pp intersect the isopleths of probability under the smallest angles. This means that if a linear combination of the components were used as a composed variable, the weighting factor of h_{500} would be smaller than that of pp. Therefore, as only one of them will be used, the variable pp has been chosen.

Yet another conclusion may be drawn from the same diagram. It can be proved that the isopleths of the thickness: $h_{500} - h_{1000} = h_{500} - 0.8\ (pp - 1000)$ (h expressed in units of 10 m) are parallels, running from the upper left to the lower right. These parallels intersect the isopleths of probability practically perpendicularly. It can therefore be concluded that the probability of rain is almost independent of the mean temperature in the layer 1000–500 mb, for this temperature is proportional to the thickness.

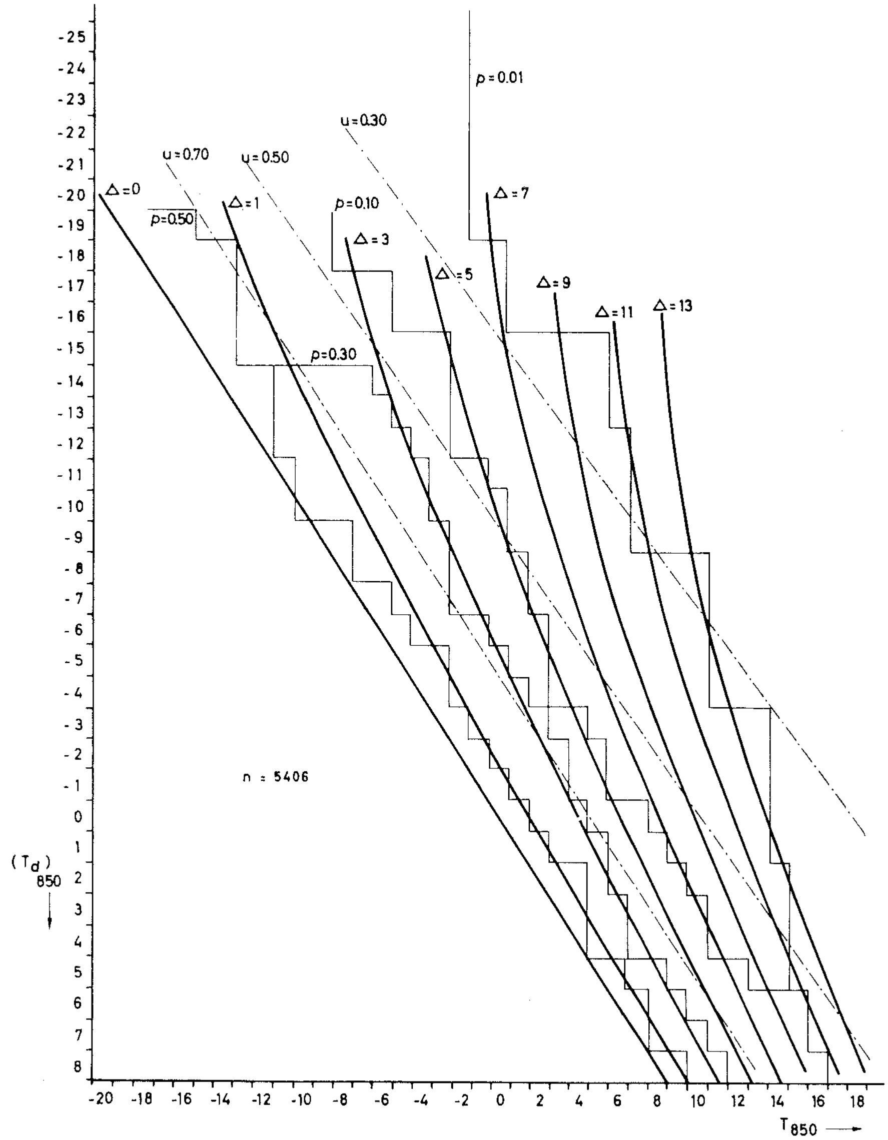


Figure 17. Probability of rain as a function of humidity indicators u (dash-dot) lines and Δ (heavy lines) in the 850 mb level

Finally, the method of linear combination with equal weight factors has been used with regard to the saturation deficit at different levels. The result is represented in the next table:

V	I	$S(I)$
$\Delta_{1000}+\Delta_{700}$	0.41	0.014
$\Delta_{850}+\Delta_{700}$	0.44	0.014
$\Delta_{500}+\Delta_{700}$	0.41	0.014

table 5.0

A comparison with table 3.11 shows that only the combination of Δ_{700} with Δ_{850} yields an improvement.

5.1 METHOD OF THE COMBINATION OF PROBABILITIES

The introduction of composed variables outlined in 5.0 leads to a method which was first used by BESSON in 1905 [1]. Later on, in 1946, it was re-introduced by BRIER [3].

As mentioned before, the composed variable is defined by making its value equal to the probability of the occurrence y at given value of V_1 and V_2: $V= p(y|V_1, V_2)$.

Let V_3 be a third variable, then the probability as a function of V and V_3 can be determined: $p=p(y|V, V_3)$. Next, in the case of the predictand being rain announced in terms of R and D according to the rule of 3.1, R is predicted if $p>p_R$ and D if $p<p_R$. The method appears to work, but the result cannot be expected to be optimal. In proof of this, the fact can be pointed to that the probability as defined above is not usually equal to the probability at given V_1, V_2 and V_3; or $p(y|V, V_3) \neq p(y|V_1, V_2, V_3)$. Consequently, it may happen that in some cases R will be predicted whereas the latter probability is smaller than the a priori probability of rain, and vice versa. The dependency-index thus obtained will be smaller than the one which would have resulted had the material been split up 3-dimensionally according to the values of the three variables.

Furthermore, it is understandable that I depends on the sequence of composition of the variables. This can be done in two additional ways, viz.: $V=p(y|V_2, V_3)$ followed by $p=p(y|V, V_1)$ and $V=p(y|V_1, V_3)$ followed by $p=p(y|V, V_2)$.

Understandably, all combinations have therefore to be investigated and the one that yields the highest index is selected. However, in doing so, one should always be aware of the fact that the results refer to a sample, which is only partly representative for the whole population. Consequently, the differences observed in I may not be real, but due to the fact that, by chance, one combination yields a better result than the others. This phenomenon is called selection-effect. The risk of becoming a victim of this selection-effect becomes greater as the number of possible combinations increases. Although the criterion just mentioned for selecting a combination should not be rejected (it cannot be replaced by a better one) one should realize that the result may be flattered and that it may fail when applied to independent data. Further, it should be emphasized once again that it is necessary to exercise caution in drawing isopleths of probability. From literature one gets the impression that these lines are often drawn to give as favourable a picture as possible without much

p_1 ↑ \ $\Delta_{700}+\Delta_{850}$ →	00	01	02	03	04	05	06	07	08	09	10	11	12	13	14	15	16	≥17
A			.20			.07		.08	.00	.00	.04	.00	.00	.00		.00	.00	
B	.50		.14		.12	.08	.13	.03	.03	.00	.02	.05	.04	.03	.03	.00	.00	.02
C	.61	.57	.43	.40	.27	.17	.20	.17	.19	.06	.07	.12	.08	.09	.07	.00	.00	.04
D	.76	.66	.70	.53	.44	.37	.39	.26	.25	.27	.19	.10	.17	.08	.11			
E	.84	.79	.67	.58	.48	.54	.43	.40	.41	.41	.21	.37	.25		.33			
F	.95	.91	.88	.72	.85	.65	.57	.56	.57									

$p_2=0.30$ $p_2=0.10$ $p_2=0.01$ $p_2=0.50$ $p_2=0.70$ $p_2=0.70$

Diagram 5. Probability of rain as a function of atmospheric pressure direction of surface wind and saturation deficit.

regard to the reality of the analysis thus obtained. It can safely be stated that a whimsical course of the isopleths is caused by the important role of chance. To what excesses failure to recognize this role may lead is shown in fig. 18, a reproduction of a diagram used for forecasting thunderstorms taken from [7]. Noting the contortions of the isopleth in order to include the yes-cases and to exclude the no-cases, it is hard to believe that the result of this paper will be

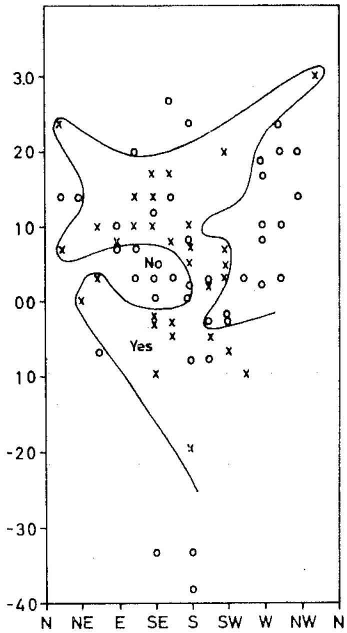


Figure 18. Unjustified discrimination of occurrence and non-occurrence (see text)

of much significance. Methods which have been developed in such a way are sure to fail when applied to independent data, and they consequently bring objective forecasting in general into disrepute.

On account of the risks involved in the application of composed variables, this procedure has to be avoided as much as possible. However, if the number of variables becomes greater than three, the method used so far of splitting the data multi-dimensionally fails as well, owing to the fact that the number of cases contained in one cell is too small to estimate with sufficient reliability the probability of rain under the relative circumstances. At this stage, one is forced to use the method of composed variables, at least if one wants to include more variables.

From the table 4.7 it appears that, with respect to surface variables, the best results are obtained by using atmospheric pressure, wind direction and persistence. On the other hand, with respect to the aerological variables, the combination $\Delta_{850}+\Delta_{700}$ proved to give the best relation to the probability of rain. Starting from this selection, the method of composed variables will be applied. On account of the fact that the variable *P*, which stands for persistence, is specified in only two classes, the data have been split up into two groups: $P=R$ and $P=D$. Accordingly, two diagrams have been constructed representing p_1 = the probability of rain as a function of *pp* and *dd*. These diagrams are analyzed by drawing isopleths for five values of p_1. As a result, the whole field is divided up into 6 regions. The "value" of p_1 in those regions is denoted by the letters A, B ... F (see figures 14 and 15).

region	*probability*
A	$p_1 \leqslant 0.01$
B	$0.01 < p_1 \leqslant 0.1$
C	$0.1 < p_1 \leqslant 0.3$
D	$0.3 < p_1 \leqslant 0.5$
E	$0.5 < p_1 \leqslant 0.7$
F	$0.7 < p_1$

Next, all cases of each class separately, are pooled regardless of the values of *pp*, *dd* or *P*, followed by an evaluation of the probability of rain as a function of $\Delta_{850}+\Delta_{700}$. The result is given in diagr. 5. On account of the final division into two parts with respective probabilities greater and smaller than the a priori probability of rain the following contingency-table can be constructed:

forecast

		R	*D*	
observed	*R*	1079	232	1311
	D	768	2103	2871

$I=0.56$ (0.015)

table 5.1

6. LOCAL CHARACTER OF RAIN

Up to now, precipitation at one place only has been considered. In practice, however, it is usual to give a forecast for an area and owing to the fact that rain is often restricted to a limited number of places, the form of the forecast has to be changed. It is usually impossible to indicate at what places it will rain and at what places it will not, but it may be possible to give a forecast referring to the areal coverage. This coverage is theoretically defined as the ratio of the area receiving rain to the whole area concerned. In practice, it is measured by the ratio, F, of the number of stations reporting rain to the total number of stations. To apply this procedure to the Netherlands 10 key stations have been selected (see fig. 1) whereas in accordance with the terminology used in domestic forecasts, the scale of F, covering the range 0 to 10 inclusive, is subdivided into the following classes:

			a priori probability
D	nowhere rain	$F=0$	0.39
MD	no-rain in most places	$F=0, 1, 2$	0.58
V	scattered rain	$F=1, 2$	0.19
P	local rain	$F=3, 4, 5, 6, 7$	0.24
MR	rain in most places	$F=8, 9, 10$	0.18

The procedure to be followed to forecast F is analogous to that described in the foregoing. The modification consists in substituting p (the probability of rain at a given location) by F, the areal coverage. Unfortunately, this can only be done partly, since, for most years used before, the ten stations are not available. However, this difficulty can be avoided at least partly. The composed variable p_1 introduced in 5.1 may formally be regarded as equivalent to other variables which have a physical meaning. Hence the dependence of F and p_1 may be determined and, as an extension, the dependence of F on the combination of p_1 and $\Delta=\Delta_{850}+\Delta_{700}$. The result is a probability distribution of the occurrence of the different classes of F as a function of p_1 and Δ. It should be stressed again that this probability distribution is not directly determined as a function of *pp*, *dd*, *P* and Δ although these variables have, in fact, been used during the procedure. This could only be done by splitting up the data multi-dimensionally according to those variables. In brief, the method consists in determining the probability distribution of F under "certain circumstances". Those circumstances are specified either by the value of the variable, as far as Δ is concerned, or, with regard to *pp*, *dd* and *P*, by the condition that these three variables satisfy some relation. The fact that p_1 refers to the probability of rain at a fixed location is not considered to constitute a restriction in this context.

After determining the probability distribution of F, a value of y_p is assigned to each combination of p_1 and Δ according to the directory drawn up in 1.3. This will be illustrated by an example. Let, at given values of p_1 and Δ, the conditional probability distribution be given in the second column of the next table:

class of F	*cond. prob.*	*a priori prob.*	*excess*
D	0.03	0.39	−0.36
V	0.10	0.19	−0.09
P	0.40	0.24	0.16
MR	0.37	0.18	0.19

A comparison with the a priori probabilities teaches that the probability-excess is largest for y_p=MR and consequently the forecast will be MR, regardless of the fact that, in an absolute sense, the probability of P is greater. If, in this way, the value of the predictor is assigned, the following contingency-table can be constructed:

forecast

observed	D	MD	V	P	MR	
D	1139	474	100	233	38	1984
V	122	221	142	274	96	855
P	65	137	89	643	287	1221
MR	6	23	47	296	530	902
	1332	855	378	1446	951	4962

I=0.43

table 6.0

The evaluation of the dependency-index is made by using the analogy described in 1.2.

According to the scale of F, the a priori probability of the occurrence D equals 0.39, hence the total stake for the 1332 predictions D that have been issued equals 1332 0.39 = 520. The following table shows the evaluation of the stakes for the various forecasts:

Prediction	*number*	*stake per prediction*	*subtotal*	
D	1332	0.39	520	
MD	855	0.58	496	
V	378	0.19	72	
P	1446	0.24	347	
MR	951	0.18	171	Total 1606

Number of hits = 1139 + 474 + 221 + 142 + 643 + 530 = 3149.
Hence net = 3149 − 1606 = 1543.
When carrying out the same calculation for perfect predictions, one should bear in mind that in such a case there would be no need for the overlapping class MD.

Prediction	*number*	*stake per prediction*	*subtotal*
D	1984	0.39	773
V	855	0.19	162
P	1221	0.24	293
MR	902	0.18	162
			1390

Number of hits = 4962
Net = 4962 − 1390 = 3572, finally $I = \frac{1543}{3572} = 0.43$

7. PRACTICAL APPLICATION

In the preceding sections an investigation was made to determine which meteorological variables can be considered to be predictors for rain. On account of climatological differences, which proved to exist dependent on the time of year, the data were split up according to the four seasons (section 2.2). Next, for each separate season, the data were split up into two classes according to persistence, viz. R: rain in the preceding period, and D: no-rain, in the preceding period. The following step was to draw isopleths of probability of rain as a function of atmospheric pressure and wind direction. This gives rise to a subdivision of the diagram in a number of classes denoted A, B, C, D, E and F. Finally, such a class, in its capacity as composed variable, was used in combination with the saturation deficit as an entry in a second diagram. This diagram was subdivided into areas in such a way that the excess of probability as a function of the variables concerned is maximal for y_p concerned.

These diagrams (p. 60 A-D) can be used for objective rain-forecasting, provided an estimate is made of the value of the predictors concerned, in respect of the middle of the forecast period. These estimates can be made, at the latest, at the beginning of the forecast period, thus, at least 6 hours in advance.

In order to get an idea of the usefulness in practice, it is necessary to answer the following questions:

1) What is the difference between the dependency-index of the completely objective forecasts, which can only be made afterwards, and that of the conventional subjective forecast?

2) To what extent will the dependency-index decrease owing to the use of estimated values of the predictors instead of observed values?

In order to answer these questions, two experiments were performed:

1) During the period July 1956–February 1957 inclusive, 474 forecasts in total were made and expressed in locality terms. It should be noted that the objective method was not known to the forecasters producing the subjective forecasts. This precaution was made intentionally, so as to ensure that the objective method would not influence the forecasters. Accordingly, the forecasts made by them may be regarded as representative of subjective forecasting. The result is shown in table 7.0.

FOR THE DIAGRAMS(p. 60A-D) SEE THE ANNEX
AT THE END OF THIS BOOK

forecast

		objective						subjective				
		D	MD	V	P	MR		D	MD	V	P	MR
observed	D	79	38	12	8	2	139	83	32	15	9	0
	V	14	26	15	25	7	87	10	28	29	17	3
	P	5	17	8	70	33	133	5	19	28	56	25
	MR	1	6	6	34	68	115	1	6	6	37	65
		99	87	41	110	137	474	99	85	78	119	93

$I = 0.42$ (objective) $I = 0.41$ (subjective)

table 7.0

It appears, then ,that the objective method was equivalent to the subjective one.

2) On account of the results of the first experiment it was decided, in the middle of 1957, to introduce a semi-objective method for routine short-range forecasting by estimating future values of the predictors and using these estimated values as entries in the diagrams. This method was used, together with the completely objective method, for a year and the result obtained was as follows: (cf. table 7.1)

Contrary to expectations the results of the semi-objective forecasts were somewhat better than those of the objective ones. It is true that this conclusion is drawn from a rather small sample, but in any case it seems justifiable to assume that the effect of estimation errors in the predictors is generally small. In addition, and as an explanation of the good results, a survey will be given of the accuracy of estimation of the several predictors:

1) *Direction of the wind*

The mean absolute error amounts to one unit, i.e. 360/16=22.5°.

2) *Atmospheric pressure*

The mean error equals −1.1 mb, i.e. the pressure has been underestimated. The mean absolute error amounts to 1.7 mb.

forecast

		objective						semi-objective				
		D	MD	V	P	MR		D	MD	V	P	MR
observed	D	79	27	7	9	2	124	57	39	14	12	2
	V	9	11	9	25	14	68	4	8	14	29	13
	P	6	10	11	29	28	84	2	10	7	40	25
	MR	5	2	2	19	43	71	4	2	3	17	45
		99	50	29	82	87	347	67	59	38	98	85

$I=0.36$ $I=0.38$

table 7.1

3) *Saturation deficit*

This variable was underestimated as well: to the extent of 0.6 units (see 3.3.0). The mean absolute error amounts to 2.6 units. The estimation is performed by using the value of the deficit at a location which will be reached by going upstream the mean flow of the 850 and 700 mb levels, starting from the centre of the Netherlands and proceeding over the distance covered during 6 hours at the wind speed concerned.

8. SOME FURTHER RESULTS

In the foregoing section the adequacy of the semi-objective method was established by means of two experiments. In view of these results, the method has been in use in the daily weather service since the end of 1958. The results it showed over 4 years, from 1959 to 1962 inclusive, are given in this section. For technical reasons only the daytime periods (06–18 GMT) have been considered.

For the sake of clearness the forecasting procedure is first briefly recapitulated. For each of the two classes of persistence a set of diagrams is constructed. The rainfall at De Bilt in the previous 12-hour period determines which set of diagrams will be considered. Next, the atmospheric pressure and direction of the surface wind at De Bilt at 12 GMT are estimated. These data determine one class of the pressure/wind direction diagram. The future value of the saturation deficit at De Bilt at 12 GMT is also estimated. This value, in combination with the class of the pressure/wind direction diagram, defines the rain forecast in accordance with the rain classes of section 6. These data are entered on a form and, in the case where the views of the forecaster and the method differ, the subjective forecast is entered by the forecaster as well.

During the four years concerned 1085 forms were filled in properly, and the method was not used or not used correctly in 376 cases. For some reason or other, then, 1 out of 4 cases could not be verified and the data suggest that the omissions were not random but biassed (e.g. during dry spells). It may be assumed accordingly that the result has been influenced in an unfavourable way. In 258 cases the forecaster doubted the result of the method and this was made known by the fact that he filled in his own forecast. About 52% of these disagreements with the method proved correct, 24% proved incorrect and the remaining 24% did not improve the initial forecast. The improvements are more frequent than the deteriorations, so one is justified in disregarding the method in special cases. This matter will be elucidated later on. The relationship between the forecast and the observed classes when making use of the *method without alterations* can be derived from the following contingency-table:

forecast

observed		D	MD	V	P	MR		
	D	252	124	31	45	8	460	(.42)
	V	27	52	17	62	23	181	(.17)
	P	6	24	23	116	89	258	(.24)
	MR	2	5	6	68	105	186	(.17)
		287	205	77	291	225	1085	

table 8.0

$I = 0.40$

Application of the method *with due observance of the alterations* yields the following contingency-table:

forecast

observed		D	MD	V	P	MR	
	D	279	105	34	37	5	460
	V	25	66	24	61	5	181
	P	4	24	32	139	59	258
	MR	1	4	5	63	113	186
		309	199	95	300	182	1085

table 8.1

$I = 0.46$

It can be shown that this improvement is significant in a statistical sense. It is due to 258 divergent forecasts, involving 136 new hits and 58 new misses with 64 forecasts ineffectual. The contributions of the different meteorologists are given in the following table:

forecaster	A	*B*	*C*	*D*	*E*	*F*	*G*
number of forecasts	172	150	181	234	76	62	210
disagreements	42	35	38	74	12	14	43
improvement	25	18	22	39	5	7	20
deterioration	9	5	10	19	1	4	10
without effect	8	12	6	16	6	3	13
improvement as a percentage	9	9	7	9	5	5	5

table 8.2

Next, we may ask for the result when making use of the observed values of the predictors. This result is given in the following contingency-table of the *completely objective method:*

forecast **VB**

observed		D	MD	V	P	MR	
	D	289	102	30	33	6	460
	V	33	40	26	64	18	181
	P	17	23	24	120	74	258
	MR	3	6	3	72	102	186
		342	171	83	289	200	1085

table 8.3

The corresponding dependency-index amounts to 0.37. Thus it is found that the use of estimated values of the predictors gives good results just as in the case of the second experiment mentioned in section 7. Although the difference between the dependency-indices $I=0.40$ and $I=0.37$ is not significant, it is plausible to assume that an improvement will be effected when estimated values are used instead of observed ones.

The explanation of this phenomenon is not obvious and it cannot be gathered from the rather small differences between estimated and observed values of atmospheric pressure and wind direction. Table 8.4 shows some features of these differences (forecast minus observation):

	mean	mean deviation
atmospheric pressure pp	−1.1	1.8 (mb)
wind direction dd	0	1 (22.5°)
saturation deficit	−0.8	2.6 ($\frac{1}{2}$ gr/m^3)

table 8.4

The way in which the saturation deficit is estimated is probably the cause of the improvement. According to section 7 (p. 61) the estimation is performed by using the deficit at a location which will be reached by going upstream according to the mean flow of the 850 and 700 mb levels. The mean value of the saturation deficit of the source region is used to estimate the future deficit at De Bilt. If the mean of the source region should be correlated higher to the probability of rain than the observed deficit at De Bilt, then the rain forecast improves. On the other hand, the estimation of the saturation deficit is sometimes the very cause of difficulties as a consequence of the inhomogeneity of the distribution of saturation deficits of the source region.

It can be concluded from the foregoing that the forecast method herein described may contribute to the improvement of the daily weather forecasts, for one thing because greater attention is being given to the element rain. Attempts will be made to investigate the cases of divergence between method and forecaster. These divergences are expected to be important in connection with a potential improvement of the method. In this respect it is worth noting that tentative investigations have been made using geostrophic vorticity as a predictor.

SUMMARY

A relation between the frequency of rain and various meteorological parameters has been investigated with the aid of surface data of the period 1901–1951 and aerological data of the period 1945–1952. Owing to a necessary restriction of the number of parameters, a selection has been made on the basis of a quantitative criterion which has been developed for this purpose. Three parameters viz. persistence, atmospheric pressure and direction of surface wind are related to the local frequency of rain. By combining this frequency with a humidity parameter referring to the 850 and 700 mb levels the areal distribution of the occurrence of rain can be determined.

This procedure makes it possible to forecast rain, provided that the values of the predicted parameters in the midpoint of the forecast period are known. As the variability of the trend of the parameters is small, it is to be expected that a linear extrapolation will give sufficiently accurate results.

Use of this semi-objective method over a period of about two years has shown that only a highly skilled forecaster is able to rival it.

SAMENVATTING

In deze verhandeling worden de resultaten weergegeven van een onderzoek naar het verband tussen regenfrekwentie enerzijds en verschillende andere meteorologische variabelen anderzijds. Het basismateriaal bestaat uit synoptische waarnemingen uit de periode 1901–1951 en aerologische waarnemingen uit de periode 1945–1952.

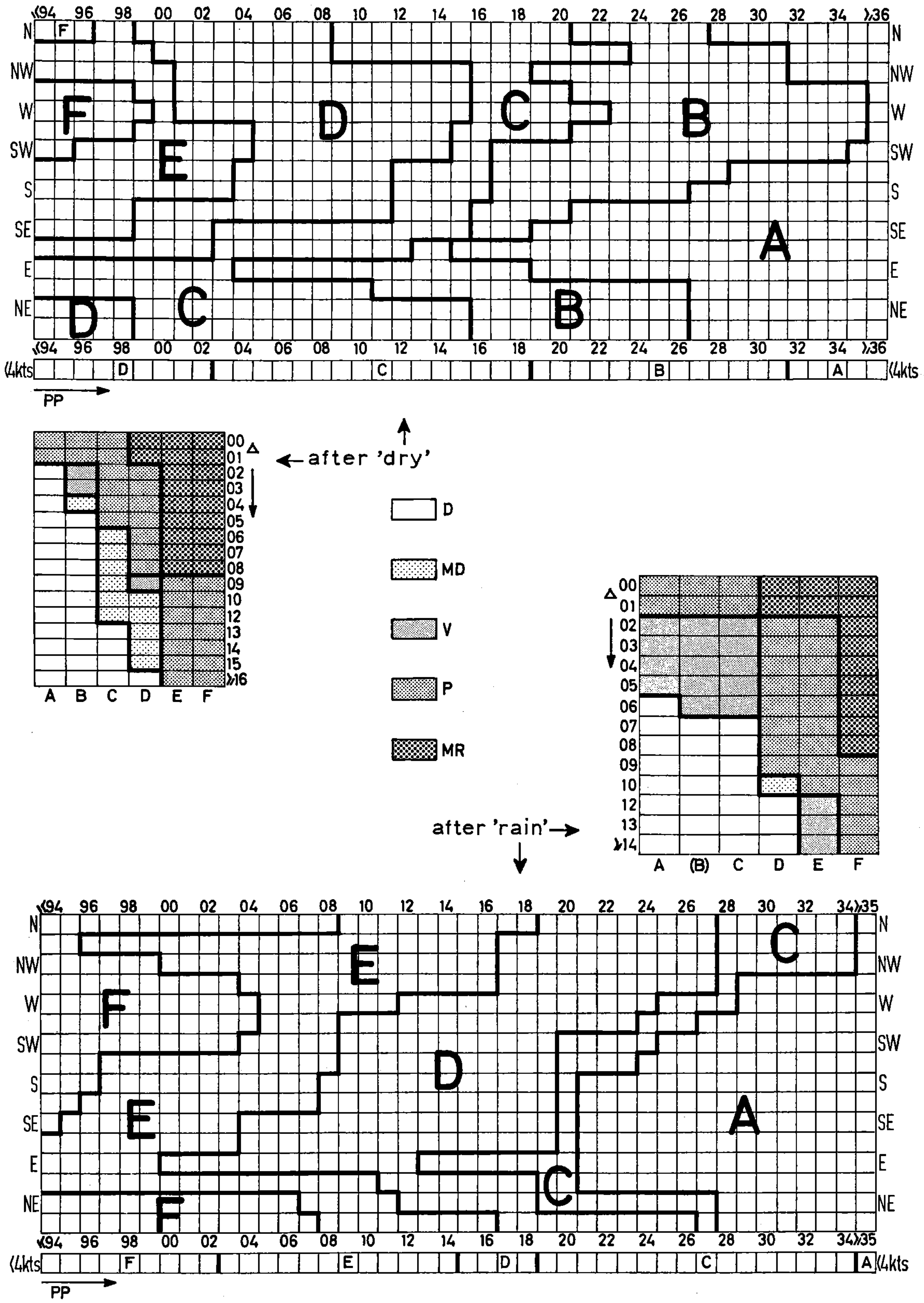
Met het oog op een noodzakelijke beperking van het aantal variabelen is een keuze gemaakt aan de hand van een daartoe ontworpen kriterium. Drie variabelen, nl. persistentie, luchtdruk en windrichting aan het aardoppervlak zijn in verband gebracht met de regenfrekwentie in het centrum van het land. Door deze regenfrekwentie te combineren met de vierde variabele i.c. de vochtigheid op de 850- en 700 mb vlakken, wordt de plaatselijkheid van de neerslag bepaald.

Deze opzet maakt het mogelijk regenvoorspellingen te maken onder voorwaarde dat de waarden van de vier gebruikte variabelen op het middentijdstip van de verwachtingsperiode bekend zijn. In de praktijk blijkt, dat deze waarden door lineaire extrapolatie met voldoende nauwkeurigheid kunnen worden geschat. Uit de toepassing van deze semi-objektieve methode over een ruime periode blijkt, dat de resultaten gelijkwaardig zijn met die van ervaren meteorologen.

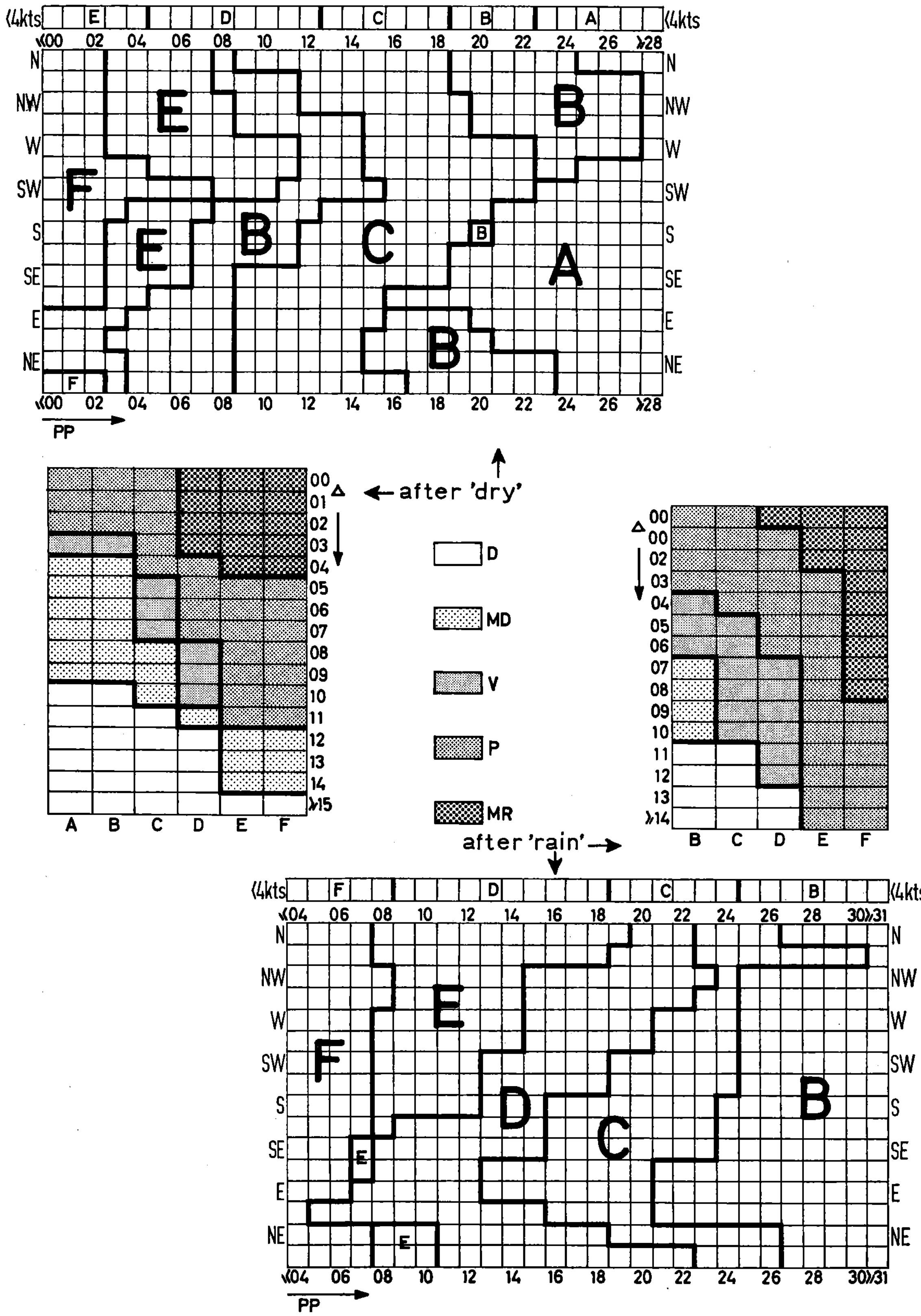
REFERENCES

1. BESSON, L. Essai de prévision méthodique du temps. Paris, Annales de l'Observatoire Municipal, 1905, 475–495.
2. ROLF, B. Probabilité et pronosties des pluies d'été. Upsala, 1917.
3. BRIER, G. W. A study of quantitative precipitation forecasting in the TVA basin. U.S. Wheater Bureau, Res. Paper, **26**, 1946.
4. FLOHN, H and HUTTARY, J. Zur Kenntnis der Struktur der Niederschlagsverteilung. Z.f. Meteor., **6**, 1952, 304–309.
5. BEEBE, R. C. The distribution of summer showers over a small area. Mon. Weather Rev., **80**, 1952, 95–98.
6. PANOFSKY, H. A. Some applications of statistics to meteorlogy. Pennsylvania, 1958.
7. ARDIS, C. V. An objective method for forecasting thunderstorms. Bull. Am. Meteor. Soc., **42**, 1962, 173.
8. AITKEN, A. C. Statistical Mathematics, p. 21. Edingburgh, 1939.

SPRING

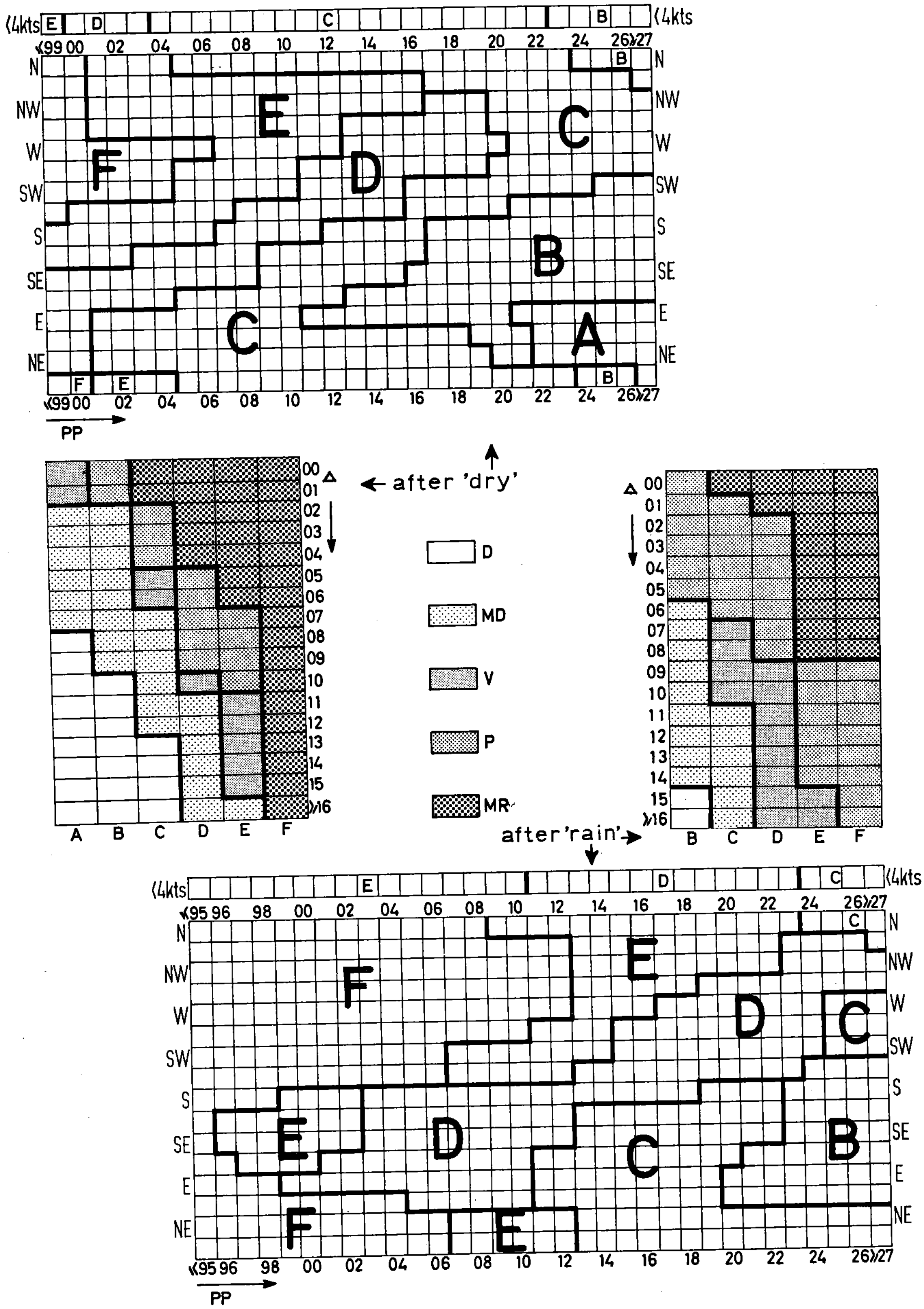


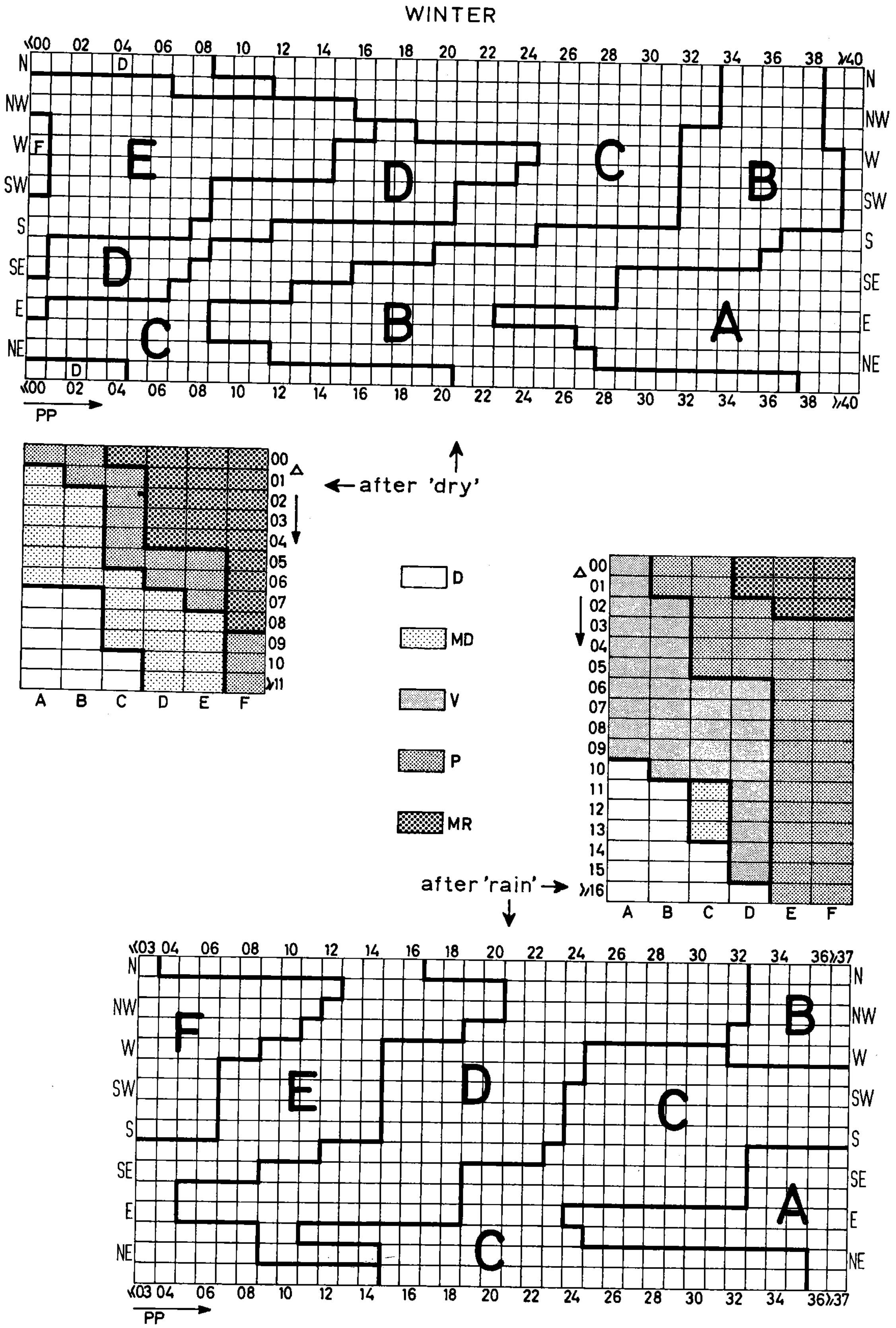
SUMMER



〈4kts
〈4kts
N
NW
W
SW
S
SE
E
NE
PP
E
F
B
C
A
D
after 'dry'
after 'rain'
D
MD
V
P
MR
00
01
02
03
04
05
06
07
08
09
10
11
12
13
14
〉15
〉14
〉28
〉31

AUTUMN




WINTER
«00 02 04 06 08 10 12 14 16 18 20 22 24 26 28 30 32 34 36 38 »40
N NW W SW S SE E NE
F E D C B D C B A
PP
00 01 02 03 04 05 06 07 08 09 10 »11
A B C D E F
after 'dry'
D
MD
V
P
MR
00 01 02 03 04 05 06 07 08 09 10 11 12 13 14 15 »16
A B C D E F
after 'rain'
«03 04 06 08 10 12 14 16 18 20 22 24 26 28 30 32 34 36 »37
N NW W SW S SE E NE
F E D C B A C
PP

Van de reeks Mededelingen en Verhandelingen *zijn bij het Staatsdrukkerij- en Uitgeverijbedrijf nog verkrijgbaar de volgende nummers:*

23, 25, 26, 27, 29*b*, 30, 31, 34*b*, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48,

alsmede

49. A. Labrijn. Het klimaat van Nederland gedurende de laatste twee en een halve eeuw. — The climate of the Netherlands during the last two and a half centuries. 1946. (114 blz. met 6 fig. en 1 kaart) . . . . *f* 1,15

50. J. P. M. Woudenberg. Het verband tussen het weer en de opbrengst van wintertarwe in Nederland. — The correlation between weather and yield of wheat in the Netherlands. 1946. (43 blz. met 6 fig.) 0,70

51. S. W. Visser. Weersverwachtingen op langen termijn in Nederland. — Long range weather forecasts in the Netherlands. 1946. (143 blz. met 25 fig.) 2,05

52. R. J. v. d. Linde en J. P. M. Woudenberg. Een methode ter bepaling van de breedte van een schaduw in verband met den tijd van een jaar en de oriëntatie van het beschaduwde object. — A method for determining the daily variation in width of a shadow in connection with the time of the year and the orientation of the overshadowing object. 1946. (6 blz. met 2 fig. en 2 kaarten) 0,40

53. A. Labrijn. Het klimaat van Nederland. Temperatuur, neerslag en wind — The climate of the Netherlands. Temperature, precipitations and wind. 1946. (71 blz. met 1 kaart) 2,50

54. C. Kramer. Electrische ladingen aan berijpte oppervlakten. — Electric charges on rime-covered surfaces. 1948. (128 blz. met 17 fig. en 1 afb.) 3,00

55. J. J. Post. Statistisch onderzoek naar de samenhang tussen het weer, de grasproduktie en de melkaanvoer. — Statistical research on the correlation between the weather, grass production and milk supply. 1949. (119 blz. met 25 fig. en 6 tab.) . . . 3,00

56. R. J. v. d. Linde and J. P. M. Woudenberg. On the microclimatic properties of sheltered areas. The oak-coppice sheltered area. — Over de microklimatologische eigenschappen van beschutte gebieden. Het landschap met eikenhakhoutwallen. 1950. 151 blz. met 52 fig.) 3,00

57. C. Kramer, J. J. Post en W. Wilten. Klimaat en brouwgerstteelt in Nederland. — Climate and growing of malting-barley in the Netherlands. 1952. (149 blz. met 27 fig.) 2,25

58. W. van der Bijl. Toepassing van statistische methoden in de klimatologie. — Application of statistical methods in climatology. 1952. (197 blz. met 19 fig.) 7,60

59. Tien wetenschappelijke bijdragen, uitgegeven bij het 100-jarig bestaan van het K.N.M.I. — English summary. 1954. (198 blz. met 53 fig.) 12,50

60. C. Kramer, J. J. Post en J. P. M. Woudenberg. Nauwkeurigheid en betrouwbaarheid van temperatuur- en vochtigheidsbepalingen in buitenlucht met behulp van kwikthermometers. 1954. (60 blz. met 11 fig.) 3,50

62. C. Levert. Regens. Een statistische studie. 1954. (246 blz. met 67 fig. en 143 tab.) 10,00

63. P. Groen. On the behaviour of gravity waves in a turbulent medium, with application to the decay and apparent period increase of swell. 1954. (23 blz.) 1,50

64. H. M. de Jong. Theoretical aspects of aeronavigation and its application in aviation meteorology. 1956. (124 blz. met 80 fig., 9 krt. en 3 tab.) 3,50

65. J. G. J. Scholte. On seismic waves in a spherical earth. 1956. (55 blz. met 24 fig.) 5,00

66. G. Verploegh. The equivalent velocities for the Beaufort estimates of the wind force at sea. 1956. (38 blz. met 17 tab.) 1,75

67. G. Verploegh. Klimatologische gegevens van de Nederlandse lichtschepen over de periode 1910—1940.

Deel I: Stormstatistieken. — Climatological data of the Netherlands light-vessels over the period 1910—1940. P. I.: Statistics of gales. 1956. (68 blz. met tabellen.) . . . . *f* 3,50

Deel II: Luchtdruk en wind; zeegang. — Climatological data of the Netherlands light-vessels over the period 1910—1940. D.II: Air pressure and wind; state of the sea. 1958. (91 blz. met tabellen.) 7,50

Deel III: Temperaturen en hydrometeoren; onweer. — Climatological data of the Netherlands light-vessels over the period 1910—1940. P. III: Temperatures and hydrometeors; thunderstorms. 1959. (146 blz. met tabellen.) 8,00

68. F. H. Schmidt. On the diffusion of stack gases in the atmosphere. 1957. (60 blz., 12 fig. en tab.) 5,00

69. H. P. Berlage. Fluctuations of the general atmospheric circulation of more than one year; their nature and prognostic value. 1957 7,50

70. C. Kramer. Berekening van de gemiddelde grootte van de verdamping voor verschillende delen van Nederland volgens de methode van Penman. 1957. (85 blz., fig. en tab.) 7,00

71. H. C. Bijvoet. A new overlay for the determination of the surface wind over sea from surface weather charts. 1957. (35 blz., fig. en tab.) 2,50

72. J. G. J. Scholte. Rayleigh waves in isotropic and anisotropic elastic media. 1958. (43 blz., fig. en tab.) 3,00

73. M. P. H. Weenink. A theory and method of calculation of wind effects on sea levels in a partly-enclosed sea, with special application to the southern coast of the North Sea. 1958. (111 blz. met 28 fig. en tab.) 8,00

74. H. M. de Jong. Geostrophic flow. Geostrophic approximation in the upper air flow with application to aeronavigation and air trajectories. 1959. (100 blz. met 17 fig., 14 krt. en 2 tab.) 5,00

75. S. W. Visser. A new analysis of some features of the 11-year and 27-day cycles in solar activity and their reflection in geophysical phenomena. 1959. (65 blz. met 16 fig. en 12 tab.) 3,50

76. A. R. Ritsema and J. Veldkamp. Fault plane mechanisms of South East Asian earthquakes. 1960. (63 blz. met 26 fig. en 11 tab.) 4,00

77. G. Verploegh. On the annual variation of climatic elements of the Indian Ocean. P. I.: text. P. II: charts. 1960. (64 blz. 15 fig., 28 krt.) 6,00

78. J. A. As. Instruments and measuring methods in paleomagnetic research. 1960. (56 blz., 20 fig.) 2,50

79. D. J. Bouman. Consistency of approximations in discontinuous fields of motion in the atmosphere with an introduction to the use of generalized functions or distributions in meteorology. 1961. (94 blz., 6 fig.) 6,50

80. H. Timmerman. The influence of topography and orography on the precipitation patterns in the Netherlands. 1963. (49 blz., 37 fig. en 5 tab.) 6,50

81. A. W. Hanssen and W. J. A. Kuipers. On the relationship between the frequency of rain and various meteorological parameters. 1965. (80 blz., fig. en tab.) . . . 10,—

82. G. A. de Weille. Forecasting crop infection by the potato blight fungus. 1964. (144 blz., 37 fig. en 37 tab.) 14,50

83. H. J. de Fluiter, P. H. van de Pol, J. P. M. Woudenberg (redactie) e.a., Fenologisch en faunistisch onderzoek over boomgaardinsekten. 1964. (231 blz., 84 fig. en 59 tab.) 9,25

84. D. J. Bouman and H. M. de Jong. Generalized theory of adjustment of observations with applications in meteorology. 1964. (89 blz., 8 fig. en 1 tab.). . . . . 11,—

85. L. Otto. Results of current observations at the Netherlands lightvessels over the period 1910—1939. Part I: Tidal analysis and the mean residual currents. 1964. (56 blz. en 8 tab.) 6,25

86. F. H. Schmidt. An analysis of dust measurements in three cities in the Netherlands. 1964. (68 blz., 14 fig. en 22 tab.). 5,50

520608*-140