The occurrence of „twin" storms from the North West on the Dutch coast

by P. J. Rijkoort *) *and J. Hemelrijk* **)

Samenvatting

Het optreden van „tweelingstormen" aan de Nederlandse kust.

Enige dagen voor Kerstmis 1954 volgden twee vrij zware stormen dicht op elkaar (21 en 23 December). Dit deed de vraag rijzen of wellicht zulke „tweelingstormen" vaker voorkomen dan te rijmen valt met een aselecte verdeling van stormen over de dagen van een jaar. Bij een dergelijk onderzoek — hier gebaseerd op windsnelheden gemeten in den Helder van 1904—1944 en van 1945—1953 —dient rekening gehouden te worden met seizoeneffecten. De „jaarlijkse gang" der stormfrequentie werd onderzocht en een sinus-vormige kromme werd aangepast (fig. 3). De waargenomen en de theoretisch verwachte frequentie — de laatste onder de hypothese van aselectheid — van de tijdsintervallen tussen op elkaar volgende stormen werden berekend en uitgezet in fig. 2. Een toets, gebaseerd op combinatie van 2 × 2-tabellen, werd toegepast om deze hypothese te toetsen.

Daarbij bleek, dat tweeling-stormen inderdaad vaker voorkomen dan op grond van aselecte verdeling der stormen te verwachten is en wel ongeveer twee à drie maal zo vaak. De kans op een storm is twee en drie dagen na een storm groter dan anders. Dit feit kan van belang zijn voor de keuze van de hoogte van zeedijken. Een meteorologische verklaring van dit verschijnsel wordt te dezer plaatse niet gegeven.

Introduction and summary

Some days before Christmas 1954 two rather heavy storms occurred in the North Sea area. The first one reached its peak on 21 December, the second one on the 23rd. The strikingly short time-interval between the storms drew the attention of the meteorologists working on the problem of storm-surges, a problem which is closely connected with the prevention of floods. The question arose, whether the time-intervals between successive storms are in agreement with a random distribution of storm over the days of a year (seasonal effects being taken into account) or not. If twin storms occur oftener than is to be expected under the hypothesis of randomness, or in other words if the probability of the occurrence of a storm is raised

*) Royal Netherlands Meteorological Institute (K.N.M.I.), De Bilt.
**) Mathematical Centre, Amsterdam (Report SP57).

by the occurrence of a storm some days before, this fact may be of importance for the determination of dike heights.

This paper describes the statistical analysis of the data of 48 years. A histogram of the length of intervals between successive storms is compared with a theoretical distribution fitted to these data under the assumption of randomness of the occurrence of storms, seasonal fluctuations being taken into account. This assumption is then tested by means of a combination of 2×2 contigency tables, and rejected. The conditional probability of the occurrence of a storm two days after a storm is about twice as large as the unconditional probability. No attempt is made here to give a meteorological explanation of this phenomenon.

1.0. *The data*

The data consist of the wind-observations made by means of a Dines anemograph at den Helder from July 1904 to June 1944 and from July 1945 to June 1953. In 1922 the position of the anemograph was changed, which seems to have affected the values of the observations. Therefore the data were divided into two parts, the first one (*period I*) consisting of the data up to June 1922, the second one (*period II*) comprising the rest.

A *storm day* was defined as follows:

Let m be the mean wind-speed over a whole period (for period I: $m = 5.2$ and for period II: $m = 6.6$ metres/second). A storm-period is then a period during which the mean hourly wind-speed is uninterruptedly greater than $\frac{3}{2}$ m. A day is called a storm-day if:

A. the maximum wind-speed on that day exceeds $\frac{5}{2}$ m and is also the maximum wind-speed of a storm-period,

B. the direction of wind during the storm-period is between West and North for at least one hour.

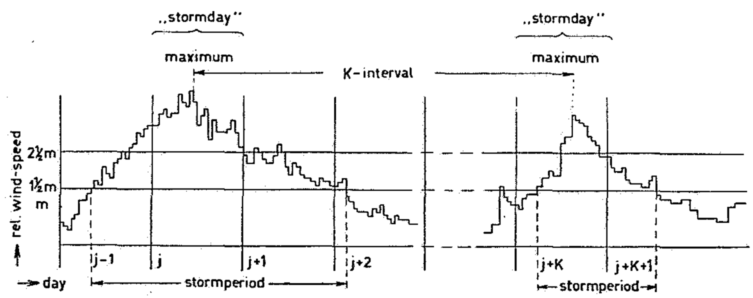

Fig. 1. Definition of storm-day and the interval between two consecutive storm-days.

This definition, which is illustrated by fig. 1, makes it possible to select all storm-days from the data mentioned above.

The random variable under consideration is the *length of the time-interval between consecutive storm-days.*

This variable will be denoted by $\underline{k}$, the underlining indicating the random

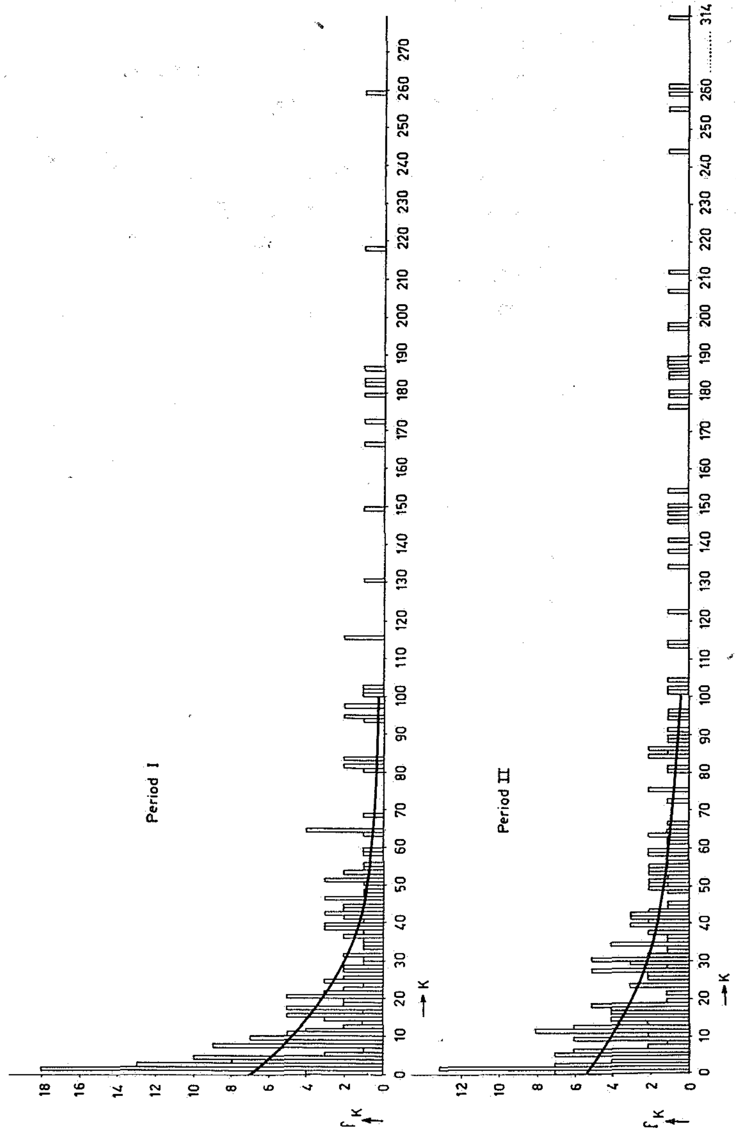

Fig. 2. Histograms of time-intervals k (in days) between successive N.W. storms observed at den Helder.

character of the variable; values assumed by $\underline{k}$ will sometimes be denoted by k, without underlining.

By definition $\underline{k}$ will assume the value k if the j^{th} day is a storm-day and the $(j+k)^{\text{th}}$ day is the next one ($j = 1, 2, \ldots$).

1.1. *The probability distribution of* $\underline{k}$

1.1.0. *The frequency distribution of* $\underline{k}$

The observed frequency distribution of $\underline{k}$, for period I and II separately, and for all seasons together, may be found in fig. 2 in the form of a histogram. Both periods show a strong maximum at $k = 2$.

1.1.1. *The seasonal fluctuation of the storm-day frequency*

In fitting a theoretical distribution to these histograms the seasonal fluctuation, called "annual course" by meteorologists, of the probability that a certain day will be a storm-day has to be taken into account.

The days of each year were numbered 1, 2, ..., 365 (omitting Febr. 29

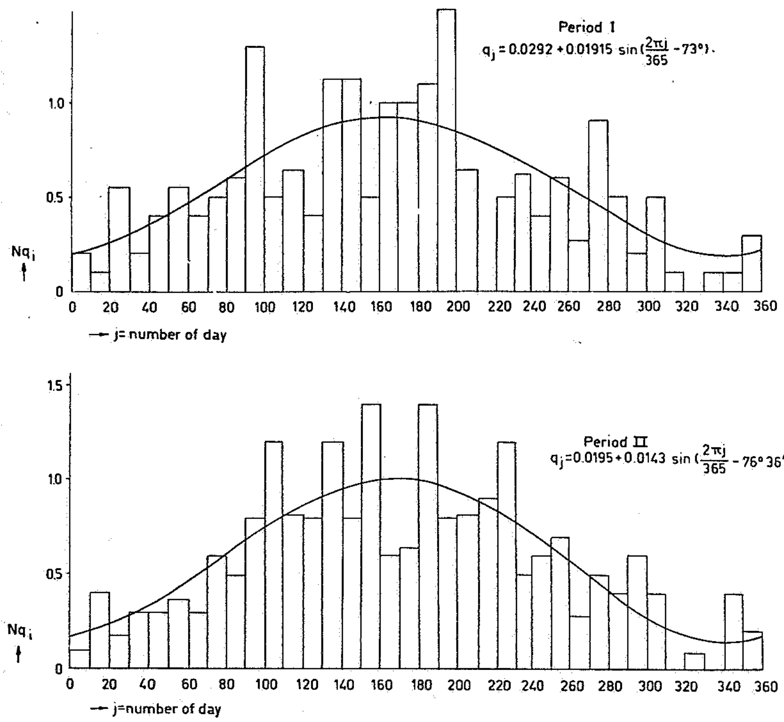

Fig. 3. The seasonal fluctuation of the storm-day frequency.

if present), starting with July 1st as no 1. Each month was divided into three decades of which the first two contain 10 days each. For each decade the number af storm-days during the period under consideration was counted. The result is shown in fig. 3.

Denoting by

j : the number of the day,

q_j : the probability that day j is a storm-day, a curve of the form

$$q_j = \alpha + \beta \sin\left(\frac{2\pi j}{365} + \varphi\right) \qquad (1.1.1)$$

was fitted to these data by means of the method of least squares. This led to the following results:

$$\begin{array}{l} \text{Period I: } \alpha = 0.0292; \; \beta = 0.01915; \; \varphi = -73^\circ, \\ \text{Period II: } \alpha = 0.0195; \; \beta = 0.0143\;; \; \varphi = -76^\circ 36'. \end{array} \qquad (1.1.1')$$

The fitted curves are also sketched in fig. 3. The χ^2-test for goodness of fit, pooling some decades with small expected frequencies, gave the following results:

Period I: $\chi^2 = 27.13$; $\nu = 23$ [1]; $P \approx 0.25$,
Period II: $\chi^2 = 22.07$; $\nu = 24$; $P \approx 0.60$.

The fitting seems to be reasonable.

1.1.2. *The distribution of $\underline{k}$, the seasonal fluctuation being taken into account*

On the basis of the fitted seasonal pattern and under the hypothesis that the storms occur independently the following approximate formula for the expected values of the frequencies of values k of $\underline{k}$ may be derived (a short proof is given in the next section).

$$\begin{aligned} E\underline{f}_k \approx 365\, N\, (1-\alpha)^{k-1} \{\alpha^2 + \tfrac{1}{2}\beta^2 \cos\frac{2\pi k}{365} J_o\,(\mathrm{i}B) + \\ + 2\alpha\beta \cos\frac{\pi k}{365} \,.\, \mathrm{i}J_1\,(iB) - \tfrac{1}{2}\beta^2 J_2\,(iB)\} \end{aligned} \qquad (1.1.2),$$

where $B = \dfrac{365\gamma}{\pi} \sin \dfrac{(k-1)\,\pi}{365}$,

$\gamma = \dfrac{\beta}{1-\alpha}$,

α and β = constants from (1.1.1'),

[1]) Number of degrees of freedom.

N = number of years of the period,
J_h = Bessel function of order h (h = 0,1,2),
$i = \sqrt{-1}$.

These expected frequencies have been plotted in fig. 2 as a smooth line. The observed frequencies for $k = 2$ (and also for $k = 3$) are much larger than the expected frequencies. The significance of this phenomenon cannot be tested by means of the χ^2-test because of the seasonal fluctuation of the probability of a storm. It will, however, be tested in another way in section 2.

1.1.3. *Sketch of the derivation of formula* (1.1.2)

According to (1.1.1) the probability, $q_{k,j}$, that day j and day $j+k$ are successive storm-days can be written as

$$q_{k,j} = (\alpha + \beta S_j')(\alpha + \beta S'_{j+k}) \prod_{r=1}^{k-1} (1 - \alpha - \beta S'_{j+r}),$$

with $S_j' = \sin\left(\frac{2\pi j}{365} + \varphi\right)$.

Substituting $j + \frac{365\varphi}{2\pi}$ for j and putting $\gamma = \frac{\beta}{1-\alpha}$ we find for the expected frequency Ef_k of the value k in a given year

$$Ef_k = (1-\alpha)^{k-1} \sum_{j=1+\frac{365\varphi}{2\pi}}^{365+\frac{365\varphi}{2\pi}} [\{\alpha^2 + \alpha\beta(S_j + S_{j+k}) + \beta^2 S_j S_{j+k}\} \prod_{r=1}^{k-1} (1 - \gamma S_{j+r})], \tag{1.1.3.1}$$

with $S_j = \sin \frac{2\pi j}{365}$.

As $\gamma S_{j+k} < 1$ we use $e^{-\gamma S_{j+r}}$ as an approximation for $1 - \gamma S_{j+r}$. It can be shown that the error in the final result is smaller than 2% even for $k \approx 100$ and of course much smaller for lower values of k.

With $j = j' - \frac{1}{2}k$ and using the formula

$$\sum_{i=\varphi}^{\sigma} \sin ai = \frac{\sin \frac{\sigma + \varphi}{2} a \sin \frac{1 + \sigma - \varphi}{2} a}{\sin \frac{1}{2} a},$$

(1.1.3.1.) may then be written

$$\underline{E}f_k \approx (1-\alpha)^{k-1} \sum_{j'=1+\frac{1}{2}k+\frac{365\varphi}{2\pi}}^{365+\frac{1}{2}k+\frac{365\varphi}{2\pi}} \left[\left(\alpha^2 - \beta^2 \sin\frac{2\pi k}{365}\right) + 2\alpha\beta \cos\frac{\pi k}{365} S_j' + \right.$$

$$\left. + \beta^2 S^2_{j'}\right) \exp\left(-\frac{365\gamma}{\pi} \sin\frac{k-1}{365}\pi \, . \, S_{j'}\right)\Bigg] . \qquad (1.1.3.2)$$

Approximating this sum by means of an integral and using the notation

$$B = \frac{365\gamma}{\pi} \sin\frac{k-1}{365}\pi,$$

$$C_1 = \frac{365\,(1-\alpha)^{k-1}}{2\pi}\left(\alpha^2 - \beta^2 \sin\frac{2\pi k}{365}\right),$$

$$C_2 = \frac{365\,(1-\alpha)^{k-1}}{\pi}\,\alpha\beta \cos\frac{\pi k}{365},$$

$$C_3 = \frac{365\,(1-\alpha)^{k-1}}{2\pi}\,\beta^2,$$

(1.1.3.2) changes into

$$\underline{E}f_k \approx C_1 \int_0^{2\pi} e^{-B\sin x}\, dx + C_2 \int_0^{2\pi} \sin x \, . \, e^{-B\sin x}\, dx + \qquad (1.1.3.3)$$

$$+ \int_0^{2\pi} \sin^2 x \, . \, e^{-B\sin x}\, dx.$$

Writing the integrals in terms of Besselfunctions and multiplying by N, the number of years, we finally arrive at formula (1.1.2).

2. *Test of the independance of the occurrence of storms*

2.0. *The test method*

The hypothesis to be tested is, that the probability of a storm on a certain day does not depend on the date of the previous one. Denoting by

A_j the fact that day j is a storm-day, and by
$\bar{A}_j$ the fact that this is not the case,

the hypothesis to be tested is

$$H_0 : \mathrm{P}\,[A_{j+k} \mid A_j] = \mathrm{P}\,[A_{j+k} \mid \bar{A}_j] \qquad (2.0.1)$$

for all j and k.

For any given j and k a 2×2-table may be set up giving the frequencies, over the N years of the period under consideration, of the four possible combinations; cf. table 2.0.

TABLE 2.0.

2 × 2-table for given values of j and k.

	A_j	$\bar{A}_j$	*total*
A_{j+k}	a_j	b_j	t_j
$\bar{A}_{j+k}$	c_j	d_j	u_j
total	r_j	s_j	N

In this table a_j is the number of years in which the j^{th} and the $(j+k)^{\text{th}}$ day were both storm-day, etc. We are interested in the validity of (2.0.1) for given k irrespective of j. For one value of k, however, $365-k$ 2×2-tables of this type can be written down. In each of these, the variable $\underline{a}_j$ has, if H_0 is true and for given values of the marginal frequencies r_j, s_j, t_j and u_j, a hypergeometric distribution. If the $\underline{a}_j$ are, for $j = 1, 2, \ldots, 365-k$, independently distributed — and this will be taken care of — then the random variable

$$\underline{x} = \frac{\sum_j \left(\underline{a}_j - \frac{r_j t_j}{N_j}\right)}{\sqrt{\sum_j \frac{r_j s_j t_j u_j}{N_j^2 (N_j - 1)}}} \tag{2.0.2}$$

will, if H_0 is true and conditionally for fixed margins of the 2×2-tables, be approximately normally distributed with zero mean and unit variance. The overall test for the whole set of 2×2-tables can then be based on $\underline{x}$. In formula (2.0.2) the symbol N has been given an index j, because it will prove to be necessary — in order to ensure the independence of the $\underline{a}_j$ — to omit certain pairs of days from some of the 2×2-tables. The method of combining 2×2-tables by means of (2.0.2), which is also valid if the probability of storms is different for different values of j, has been described by Constance van Eeden[1]).

Large positive and negative values of $\underline{x}$ respectively imply that

$$\sum_j \{\mathrm{P}\,[A_{j+k} \mid A_j] - \mathrm{P}\,[A_{j+k} \mid \bar{A}_j]\} > 0 \text{ and } < 0 \text{ respectively.}$$

2.1. *Application of the test*

For every period of years separately a table was prepared of the form of table 2.1, in which the storm-days have been indicated by crosses.

[1]) Constance van Eeden, Methoden voor het vergelijken, toetsen en schatten van onbekende kansen, *Statistica* 7 (1953) 141—16, cf. p. 149.

TABLE 2.1.1.

Part of a table of storm-days

year \ day	1	2	3	4	5	6	7	...
1	.	×	.	.	.	×	.	
2	.	.	×	.	×	.	.	
3	×	.	.	.	.	×	.	
4	.	.	.	×	.	.	.	
5	.	.	×	.	.	×	.	
.								
.								
.								

In this (part of the) table we find e.g. for $j = 3$ and $k = 2$: $a_j = 1$, $b_j = 0$, $c_j = 1$, $d_j = 3$ ($r_j = 2$, $s_j = 1$, $t_j = 1$, $u_j = 4$, $N_j = 5$).

For $k = 1$ the hypothesis H_0 is certainly not satisfied. This is a consequence of the definition of a storm-day as the day with greatest windspeed of a storm-period, for a storm-period usually lasts more than one day and very often the day after a storm-day belongs to the same storm-period and cannot be a storm-day itself. Thus the probability therefore that a day following a storm-day is itself a storm-day must be small, but here this is of no interest. In accordance with the aforegoing the same holds for a day before a storm-day.

For $k = 2$ this effect may still be present, but then it could never account for the large frequency of this value as indicated in fig. 2.

In this case, therefore, it is of interest to test H_0. But the effect indicated (for $k = 1$) in the previous paragraph has to be eliminated. If now j, $j+1$ and $j+2$ are the numbers of three consecutive days and $j+1$ is a storm-day, there is a large probability that the days j and $j+2$ are not. At any rate whether day j is a storm-day or not, the probability that $j+2$ is one is decreased by the fact that $j+1$ is one, and thus this probability differs from those for day $j+2$, in other years, for which $j+1$ is not a storm-day. If, however, all pairs $(j, j+2)$ with $j+1$ a storm-day are not counted for the 2×2-table with number j, then this effect is clearly eliminated and the resulting $\underline{a}_j$ are indeed stochastically independent if H_0 is true. This elimination is a very simple procedure, which can easily be carried out; the total number N_j of pairs of days $(j, j+2)$ may now be different for different values of j.

For $k = 3$ a similar reasoning holds. If we consider the pair $(j, j+3)$, then day $j+2$ should not be a storm-day, for this lowers the probability for day $j+3$ to be one. Also, if day $j+1$ is a storm-day and if H_0 has to be rejected for $k = 2$, the probability for day $j+3$ depends on whether day $j+1$ was a storm-day or not. In general, for $k = k_0$, and if H_0 has been rejected for

$k = 1, 2, \ldots, k_1$, then only those pairs $(j, j+k_0)$ where none of the k_1 days previous to $j+k_0$ are storm-days should be retained.

With this proviso the test of section 2.0 may be applied successively for $k = 2, 3, \ldots$ The resulting values of the test statistic $\underline{x}$ from (2.0.2), although not stochastically independent because they are based on the same set of data, may give a clear impression of the dependence of the conditional probabilities P $[A_{j+k} \mid A_j]$ on k.

One more precaution was taken to ensure that, if H_0 is true, no difficulties will arise from different probabilities of storms for the same value of j. Such a difference would be present if the probability mentioned is not the same for all years, i.e. if the frequency of storms differs widely for different years. The years of both periods I and II together (for this test there was no reason to keep up the distinctions between these periods) were now divided into three groups with different numbers of storms in the whole year. These groups were

group A, with less than 7 storms per year,
group B, with 7, 8 or 9 storms per year,
group C, with more than 9 storms per year.

For every value of k considered, the 2×2-tables were written down separately for these three groups (and for all values of j) and then combined by means of formula (2.0.2). The results, in the form of values of $\underline{x}$, are given in table 2.1.2.

TABLE 2.1.2.

Results of the test for H_0, values of $\underline{x}$.

group \ k	2	3	4	5	6	7	10	15
A	+1.42	−0.80	−0.89	−0,81	−0,37	+0,87	+1.15	−0.17
B	+3.75	+2.12	+0.38	+1.69	+0.22	+1.09	−1.26	+1.27
C	+1.94	+3.49	−0.18	+0.10	−0.25	+0.81	+1.17	−0.58
combined	+4.00	+3.70	−0.19	+0.81	−0.21	+1.52	+0.56	+0.15

The values of $\underline{x}$ found for $k = 2$ and 3 are highly significant. For $k \geqslant 4$ no large values of $| \underline{x} |$ occur. We may, therefore, conclude that the probability of a storm-day is larger 2 and 3 days after a storm-day than otherwise. From fig. 2 this probability may be estimated to be more than twice as large as it would be if storms were distributed at random. Twin storms occur more often then would be expected on the basis of the frequency of storms only.